**Visualização de séries temporais multidimensionais: uma
análise sobre a qualidade do ar em São Carlos–SP**

Evandro Scudeleti Ortigossa

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências
de Computação e Matemática Computacional (PPG-CCMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Evandro Scudeleti Ortigossa

Visualização de séries temporais multidimensionais: uma análise sobre a qualidade do ar em São Carlos–SP

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para a obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de concentração: Ciências de Computação e Matemática Computacional.

Orientador: Prof. Dr. Luis Gustavo Nonato.

USP – São Carlos
Dezembro de 2018

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

077v   Ortigossa, Evandro Scudeleti
            Visualização de séries temporais
            multidimensionais: uma análise sobre a qualidade do
            ar em São Carlos-SP / Evandro Scudeleti Ortigossa;
            orientador Luis Gustavo Nonato. -- São Carlos, 2018.
            120 p.

            Dissertação (Mestrado - Programa de Pós-Graduação
            em Ciências de Computação e Matemática
            Computacional) -- Instituto de Ciências Matemáticas
            e de Computação, Universidade de São Paulo, 2018.

            1. Visualização de Informações. 2. Séries
            Temporais. 3. Poluição do Ar. 4. Material
            Particulado. I. Nonato, Luis Gustavo, orient. II.
            Título.

Evandro Scudeleti Ortigossa

**Multidimensional time series visualization: an analysis of
the air quality in São Carlos–SP**

Master dissertation submitted to the Institute of
Mathematical and Computer Sciences – ICMC-USP,
in partial fulfillment of the requirements for the
degree of the Master Program in Computer Science
and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and
Computational Mathematics.

Advisor: Prof. Dr. Luis Gustavo Nonato.

**USP – São Carlos
December 2018**

Dedicado à memória de Adevaldo Colonize. Um verdadeiro educador que inspirava seus alunos a lutarem para transformar sonhos em realidades.

Agradecimentos

À Universidade de São Paulo e ao Instituto de Ciências Matemáticas e de Computação, por oferecerem ensino superior do mais alto nível e pelo ambiente favorável ao conhecimento que proporcionam.

Aos professores que me acompanharam desde a graduação até o mestrado, em especial ao Prof. Dr. Luis Gustavo Nonato, pela orientação, atenção, por todas as ideias, sugestões e confiança na elaboração e desenvolvimento desta dissertação. Também ao Prof. Dr. Afonso Paiva Neto pelas dicas valiosas.

Ao grupo de estudos em Controle Ambiental e Monitoramento da Qualidade do Ar em Ambientes Externos, à Profa. Dra. Mônica Lopes Aguiar e ao Eduardo Alexandrina, do Departamento de Engenharia Química da Universidade Federal de São Carlos (UFSCar). Aos colegas do Visual and Geometry Processing Group (VGPG) e do Visualization, Imaging and Computer Graphics (VICG).

Aos meus pais que sempre deram todo o suporte às minhas escolhas e não mediram esforços para eu chegar até esta etapa da minha vida. Vocês são fundamentais.

Aos tios, primos e familiares que me incentivaram, em especial aos tios Marco e Mali que fizeram parte desta trajetória desde o primeiro dia de universidade.

A todos os amigos que de algum modo contribuíram nos estudos e trabalhos para que conseguíssemos entrar, sobreviver e passar pela universidade. Entre eles: Isabela, Vanessa, Ivan, Matheus, Tamires, Gabriel, Lucas, Fábio, Diego.

À Jéssica, por me encorajar durante boa parte desta pesquisa.

Ao System of a Down, Foo Fighters e tantas outras bandas, por suas músicas terem me mantido acordado e criativo nas longas madrugadas de programação e escrita.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo apoio financeiro nesses anos.

Por fim, a quem me apoiou durante o final da graduação e me motivou a entrar para a vida de pós-graduando.

“O amor é mestre, mas é preciso saber adquiri-lo, porque se adquire dificilmente, ao preço de um esforço prolongado; é preciso amar, de fato, não por um instante, mas até o fim.”

Dostoiévski

Resumo

ORTIGOSSA, E. S. Visualização de séries temporais multidimensionais: uma análise sobre a qualidade do ar em São Carlos-SP. 2018. 120 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos-SP, 2018.

Séries temporais permeiam os mais variados domínios de aplicação, sendo a análise e a compreensão da dinâmica de séries temporais um problema interessante e bastante complexo. Técnicas de visualização têm sido aplicadas com sucesso nesse contexto, pois tiram proveito das habilidades perceptuais humanas ao transformar dados abstratos em informação visual. Nesta dissertação, foi proposta uma metodologia de análise visual para séries temporais multidimensionais. A finalidade dessa metodologia é a exploração e o entendimento do comportamento de material particulado presente na atmosfera de São Carlos-SP, para melhor compreender a sinergia entre a qualidade do ar, o clima e as doenças respiratórias. As características do material particulado variam consideravelmente no decorrer do tempo, assumindo diferentes padrões, a depender das condições climáticas. Isso atesta a importância de ferramentas analíticas que sejam capazes de explorar as mudanças temporais dos dados. Duas ferramentas de visualização interativas foram desenvolvidas, utilizando principalmente a linguagem de programação *JavaScript* com a biblioteca D3. As ferramentas são portáteis em basicamente todos os navegadores web modernos, não exigindo a instalação de *softwares* específicos para que os usuários acessem-nas e façam as suas pesquisas. Os resultados obtidos a partir das análises dos dados amostrados demonstram que houve uma redução nos níveis de poluentes atmosféricos ao longo dos anos. A efetividade e a utilidade das tecnologias propostas são demonstradas em estudos de caso envolvendo dados reais, validados por especialistas do domínio de interesse.

Palavras chave: visualização de informações; séries temporais; poluição do ar; material particulado; questões de saúde.

Abstract

ORTIGOSSA, E. S. Multidimensional time series visualization: an analysis of the air quality in São Carlos–SP. 2018. 120 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos–SP, 2018.

Time series data are present in a wide variety of applications, and the explorations, analysis, and understanding of time series dynamics are interesting although it is a quite complex problems. Information visualization techniques have been successfully applied in this context taking advantage of human perceptual abilities when transforming abstract data into visual information. In this master dissertation, we present a new visual analytic methodology for multidimensional time series analysis. The aim of the proposed approach is to allow the exploration of the behavior of particulate matter present in the urban atmosphere of São Carlos–SP, in order to understand the synergy between air quality, climate, and respiratory diseases. Moreover, the features related to particulate matter vary considerably over time, assuming different patterns according to the weather conditions. This dynamic shows the importance of analytical tools that enable the exploration of temporal changes of the data. We developed two interactive visualization tools, mainly in JavaScript programming language and D3 library. The tools are portable and can be used in all the modern web browsers, not requiring any software installation. The data analysis results show a decrease in atmospheric pollutant level over the years. The effectiveness and usefulness of the proposed technologies are demonstrated in case studies which involve real data and it is validated by domain experts.

Keywords: information visualization; time series; air pollution; particulate matter; health issues.

Lista de Figuras

2.1	Cubo de dados formado por séries temporais multidimensionais resultantes de um procedimento de simulação paramétrica.	9
2.2	Diagrama esquemático da metodologia interativa de visualização de informações.	13
2.3	Representação evolutiva das múltiplas variáveis envolvendo a campanha de Napoleão contra a Rússia, decorrida em 1812.	15
2.4	Comparação da legibilidade individual ao se analisar múltiplas séries temporais em uma metáfora simples de linhas.	16
2.5	Exemplo de alguns dos princípios <i>Gestalt</i> de agrupamento de objetos.	19
2.6	Visualizações de espaço compartilhado e espaço dividido aplicadas sobre o mesmo conjunto de dados.	22
2.7	Sistema de visualização composto por múltiplas <i>views</i> distintas e interligadas.	24
2.8	Recursos interativos em ferramenta de visualização de informações para dispositivos com <i>displays</i> sensíveis ao toque.	29
3.1	Comparação entre duas metáforas aplicadas a séries temporais que contém eventos periódicos.	37
3.2	<i>KronoMiner</i> é composto por uma barra de ferramentas (a), estrutura radial para apresentar os múltiplos níveis de detalhes dos dados (b), visão geral (c), área de comparação (d), lente para uma visão detalhada (e), um <i>tooltip</i> (f), e um menu de contexto (g).	38
3.3	Distribuição dos padrões de frequência dos funcionários em um laboratório durante o ano de 1997.	39
3.4	<i>Flowstrates</i> apresentando informações espaço-temporais de fluxos migratórios com origem na Etiópia.	40
3.5	Aplicação de mapas 3D sobre as informações de desemprego no estado da Carolina do Norte, Estados Unidos, de 1999 a 2008.	41
3.6	Painel temporal em “pseudo-perspectiva”, conectando dois eixos adjacentes do sistema de coordenadas paralelas.	42
3.7	<i>ThemeRiver</i> e os temas abordados com maior frequência por Fidel Castro em seus discursos, de 1959 a 1961.	44

3.8	Receitas, em bilheterias, dos filmes em cartaz nos cinemas norte-americanos ao longo de 2007.	45
3.9	<i>HierarchicalTopics</i> visualizando notícias textuais, com <i>stream graphs</i> apresentando as tendências evolutivas dos tópicos selecionados.	47
3.10	<i>EvoRiver</i> e a evolução dos temas discutidos no <i>Twitter</i> durante a eleição presidencial de 2012, nos Estados Unidos.	48
3.11	<i>TimeNotes</i> apresentando seus vários níveis de <i>zooming</i> , visualizando informações de monitoramento animal em áudio.	50
3.12	Níveis de poluição do ar coletados diariamente em regiões de Jinan, na China.	52
4.1	Representação esquemática de um AGV-MP ₁₀	57
4.2	Visão inicial da interface com os dados de maio de 2015 até novembro de 2017 (superior) e os dados de abril de 2014 até abril de 2015 (inferior).	64
4.3	Ao escolher os dados de 1997 a 2006, a interface é atualizada para exibir em sua parte inferior os padrões históricos de concentração em massa do MP ₁₀	66
4.4	Informações detalhadas podem ser acessadas simplesmente movendo o <i>mouse</i> sobre um ponto de interesse.	68
4.5	Exemplo de aplicação da ferramenta de seleção interativa de dados, que utiliza o conceito de <i>zooming</i>	69
4.6	Gráficos auxiliares apresentando as variáveis meteorológicas dos dados de maio de 2015 até novembro de 2017.	71
4.7	Diferença de valores nas exportações e importações da Inglaterra com a Dinamarca e a Noruega, entre os anos de 1700 e 1780.	73
4.8	Diferença de valores entre os dados de 2015–2017 (azul) e 2014–2015 (vermelho).	74
4.9	Intervalo deslizante entre os conjuntos de dados <i>A</i> e <i>B</i>	75
4.10	Exemplo comparativo entre interpolação de dados linear e interpolação utilizando <i>splines</i> cúbicas, sobre os mesmos pontos de dados.	76
4.11	Comparação entre dois intervalos, dos conjuntos <i>A</i> e <i>B</i> , que apresentam altas concentrações de MP ₁₀	78
4.12	Visão inicial da interface, apresentando as ocorrências de “pneumonia” durante o ano de 2014 (superior) e os casos registrados em 1998 da mesma doença (inferior).	80
4.13	As ferramentas de interação apresentam informações do dado sob o cursor, ao navegar entre as barras agrupadas (a), no <i>stacked graph</i> (b) ou nos grupos auxiliares (c).	83
5.1	Análise dos picos de concentração de MP ₁₀ para os dados aferido de maio de 2015 até novembro de 2017.	89

5.2	Perfil descrito pelos dados coletados entre janeiro de 2016 e novembro de 2017.	91
5.3	Dados coletados para os intervalos de maio de 2015 até maio de 2016 (superior) e de maio de 2014 até maio de 2015 (inferior).	92
5.4	Evolução das médias anuais para a concentração do MP ₁₀ aferidas nas campanhas de 1997–2005 e 2014–2017, face aos limites médios de referência.	94
5.5	Índices gerais acumulados ao longo dos anos, para as principais doenças relacionadas ao aparelho respiratório.	95
5.6	Evolução dos casos de pneumonia, divididos em grupos etários, ao longo dos meses de 2004.	96
5.7	Evolução anual dos grupos etários diagnosticados com pneumonia (<i>stacked graph</i>), sob os valores médios do MP ₁₀ para o mesmo intervalo (gráfico de área).	97
5.8	Evolução anual dos grupos etários diagnosticados com bronquite (<i>stacked graph</i>), sob os valores médios do MP ₁₀ para o mesmo intervalo (gráfico de área).	98

Lista de Tabelas

4.1	Características básicas dos conjuntos de dados ambientais.	58
5.1	Valores médios limite para a concentração do MP_{10} atmosférico recomendados pela OMS (internacional) e CETESB (nacional).	93
5.2	Valores médios limite para a concentração do MP_{10} atmosférico recomendados pelo CONAMA, em vigor durante a campanha de 1997 a 2005.	94

Lista de Abreviaturas e Siglas

AGV-MP ₁₀	Amostrador de Grande Volume para partículas de até 10 µm
CETESB	Companhia Ambiental do Estado de São Paulo
CONAMA	Conselho Nacional do Meio Ambiente
CSS	<i>Cascading Style Sheets</i>
CSV	<i>Comma Separated Values</i>
D3	<i>Data-Driven Documents</i>
DOM	<i>Document Object Model</i>
Full HD	<i>Full High Definition</i>
HTML5	<i>Hypertext Markup Language</i> , versão 5
KDD	<i>Knowledge Discovery in Databases</i>
LCD	<i>Liquid Crystal Display</i>
LIFO	<i>Last In, First Out</i>
MP ₁₀	Material Particulado com diâmetro aerodinâmico de até 10 µm
OMS	Organização Mundial da Saúde
SGBD	Sistema de Gerenciamento de Banco de Dados
SUS	Sistema Único de Saúde
SVG	<i>Scalable Vector Graphics</i>
TSV	<i>Tab Separated Values</i>
UFSCar	Universidade Federal de São Carlos
W3C	<i>World Wide Web Consortium</i>

Sumário

1	Introdução	1
1.1	Contextualização	1
1.2	Objetivos e Contribuições Esperadas	3
1.3	Organização do Texto	5
2	Conceitos e Fundamentos	7
2.1	Considerações Iniciais	7
2.2	Séries Temporais	7
2.2.1	Multidimensionalidade	8
2.2.2	Técnicas de Análise e Mineração de Dados	10
2.3	Visualização de Informações e Séries Temporais	12
2.3.1	Desafios em Visualização de Séries Temporais	15
2.3.2	Percepção e Propriedades Gráficas	17
2.3.3	Categorização de Metáforas e Elementos Visuais	20
2.3.4	Técnicas Exploratórias e Interativas	25
2.4	Tecnologias e Ferramentas Utilizadas	30
2.5	Considerações Finais	33
3	Trabalhos Relacionados	35
3.1	Considerações Iniciais	35
3.2	Estado da Arte em Visualização de Séries Temporais	35
3.2.1	Metáforas Espirais e Radiais	37
3.2.2	Calendários	39
3.2.3	Tempo, Espaço e Mapas	40
3.2.4	Coordenadas Paralelas	42
3.2.5	As Metáforas Empilhadas	43
3.2.6	Grafos, Documentos Textuais e Tópicos	46
3.2.7	Séries Temporais Financeiras	48
3.2.8	Sistemas Multi-foco	50
3.2.9	Monitoramento da Qualidade do Ar	51
3.3	Considerações Finais	53

4	Desenvolvimento da Proposta	55
4.1	Considerações Iniciais	55
4.2	Conjuntos de Dados	56
4.2.1	Monitoramento da Qualidade do Ar	56
4.2.2	Doenças do Aparelho Respiratório	59
4.3	Tarefas Analíticas	60
4.4	O Sistema de Análise Visual dos Dados Ambientais	62
4.4.1	Interface Gráfica e seus Componentes	62
4.4.2	Ferramentas Interativas	67
4.4.3	Análise Comparativa entre Diferentes Intervalos	72
4.5	O Sistema de Análise Visual dos Índices de Saúde	79
4.5.1	Interface e seus Componentes	79
4.5.2	Ferramentas Interativas	83
4.6	Considerações Finais	85
5	Resultados	87
5.1	Considerações Iniciais	87
5.2	Estudos de Caso	87
5.2.1	Comparando Padrões dos Dados Ambientais	88
5.2.2	Influências Antropogênicas Sobre os Poluentes	91
5.2.3	Evolução das Doenças Respiratórias	95
5.3	Discussões	100
5.4	Considerações Finais	102
6	Conclusão	103
6.1	Limitações e Trabalhos Futuros	103
6.2	Considerações Finais	105
	Referências Bibliográficas	107

Capítulo 1

Introdução

1.1 Contextualização

O material particulado (MP) é um poluente atmosférico composto por partículas sólidas e líquidas suspensas no ar, que podem apresentar as mais variadas formas físicas e químicas (WHO, 2006). Diversos estudos já relataram a associação entre os elevados níveis de poluição do ar com efeitos nocivos sobre a saúde humana (MOHAMMED; KARANI; MITCHELL, 2017), tornando o poluente MP um assunto de notável preocupação para os órgãos sanitários e ligados à saúde pública. Pesquisas mostram que a exposição e inalação prolongadas de partículas com diâmetro aerodinâmico de até 10 μm (MP₁₀) estão relacionadas com o desenvolvimento de doenças agudas e crônicas do aparelho respiratório e cardiovascular, por ser esta a fração inalável que pode se acumular no organismo (BROOK, 2008; DU *et al.*, 2016).

A poluição atmosférica em centros urbanos é oriunda de fontes antropogênicas e naturais (SHEN *et al.*, 2018). Neste cenário, a coleta e a análise de dados ambientais são de fundamental importância para se compreender os fatores precursores da poluição, em termos locais e globais. Investigar a origem e o movimento sazonal dos poluentes é imprescindível para que as agências públicas gestoras tomem as adequadas decisões, visando reduzir os impactos da poluição atmosférica sobre a saúde humana.

Entretanto, existe uma considerável distância entre a capacidade dos dispositivos em obter e armazenar dados, e a habilidade das ferramentas analíticas atuais de processar, organizar e exibir adequadamente esses dados. Procedimento este que tem como finalidade auxiliar os analistas na identificação de padrões e anomalias presentes nos dados e, conseqüentemente, reunir características que possam servir como uma valiosa fonte de informações para estudos (MCLACHLAN *et al.*, 2008).

Conjuntos de dados que têm seus valores aferidos ao longo do tempo são denominados séries temporais. Uma série temporal pode ser formalmente definida como uma coleção de dados em que cada unidade discreta de informação p é caracterizada por um par ordenado $p = (t, v)$, com t sendo o instante de tempo em que v foi anotado. A variável dependente v pode ser unidimensional, ou mesmo um subconjunto com número arbitrário de variáveis.

É frequente encontrar séries temporais que têm suas instâncias compostas por múltiplas variáveis, dependentes e independentes entre si. Por exemplo, uma série temporal com dados sobre poluentes atmosféricos, em uma determinada região, pode ter suas instâncias constituídas também por um subconjunto de variáveis meteorológicas relativas àquele ambiente, que foram aferidas e registradas em diferentes intervalos de tempo. Essas múltiplas variáveis, também referenciadas pelo termo “dimensões”, descrevem os conjuntos de dados multidimensionais, isto é, dados cujos elementos representam pontos no espaço m -dimensional (SHNEIDERMAN, 1996).

Com aproximadamente metade dos tecidos neurais direta ou indiretamente ligados à visão, o sistema cognitivo humano possui uma sofisticada habilidade em reconhecer padrões em objetos, quando estimulado visualmente. Isso significa que o esforço necessário para identificar, interpretar e extrair características de interesse em conjuntos de dados é reduzido quando estes são exibidos dentro de elementos gráficos (HOFFMAN, 1998; WARD; GRINSTEIN; KEIM, 2010). Neste contexto, faz-se necessário o desenvolvimento de métodos capazes de sumarizar, selecionar e apresentar as informações contidas nos dados, de modo a viabilizar uma análise visual sobre as mesmas. No caso particular de séries temporais, intensas pesquisas têm sido desenvolvidas na direção de criar ferramentas de visualização capazes de revelar informações e padrões relevantes contidos nas séries. Gráficos de linha e gráficos de barras são largamente utilizados na representação de dados temporais nas mais variadas aplicações. No entanto, ainda que metáforas de linha tratem de maneira efetiva algumas séries temporais, elas não escalam bem para tratar múltiplas séries, resultando em problemas como a desordem e o congestionamento visual (HEER; KONG; AGRAWALA, 2009).

Representar séries temporais multidimensionais de modo eficiente e compreensivo, permitindo uma análise exploratória por parte dos usuários, está claramente além das capacidades de simples gráficos de linhas e de barras. A complexidade inerente a estas estruturas e a consequente desordem visual produzida quando metáforas simples são utilizadas para apresentar grandes volumes de informações em um ambiente limitado como as telas de computador, tornam os requisitos de *design* dos sistemas de exploração visual modernos ainda mais desafiadores. Quanto mais dados são reunidos, mais difícil se torna analisá-los, chegando um momento em que a percepção gráfica se torna impraticável devido

à quantidade de informações exibidas concorrentemente dentro do mesmo espaço (JAVED; MCDONNEL; ELMQVIST, 2010).

A dimensionalidade dos dados pode introduzir dificuldades adicionais de representação, mesmo em sistemas desenhados especificamente para explorar dados multidimensionais. Acompanhar visualmente a evolução de múltiplas variáveis ao longo do tempo pode ser uma tarefa bastante complexa. Neste contexto, um aspecto inovador da analítica visual está na utilização de interfaces gráficas interativas como parte do processo de extração de informações (HEER; SHNEIDERMAN, 2012). Essas ferramentas tiram proveito da habilidade visual humana, com o objetivo de gerar entendimento sobre grandes quantidades de dados a partir de visualizações que permitem a descoberta de padrões, tendências, anomalias e correlações, de modo interativo e não direcionado (BADAM *et al.*, 2016).

Assim, nesta pesquisa de Mestrado foram desenvolvidas novas ferramentas analíticas apoiadas por visualização, interativas e baseadas em web, que foram customizadas para dar suporte à análise exploratória do comportamento evolutivo da concentração de material particulado (MP₁₀), amostrado em uma região da cidade de São Carlos–SP, além de um conjunto de variáveis meteorológicas e índices de saúde locais.

Os sistemas, duas interfaces, foram construídos por meio da composição de variadas metáforas gráficas interligadas, aliado a mecanismos de interação com usuário. Ambos os sistemas permitem a identificação de padrões e a verificação de como esses padrões variam com o tempo, além de suportarem a comparação entre diferentes intervalos amostrais, exibindo informações detalhadas e sob demanda. Com as ferramentas, é possível identificar tendências gerais dos poluentes atmosféricos entre os diversos registros, com a flexibilidade de permitir ao usuário explorar interativamente aspectos de seu interesse, utilizando métodos que facilitam a navegação no espaço dos dados em níveis distintos de detalhamento.

1.2 Objetivos e Contribuições Esperadas

A análise de séries temporais é largamente estudada para a modelagem dos fenômenos que geraram esses dados, sendo objeto de investigação em muitos ramos da ciência. A visualização de informações é um meio de comunicação que representa os dados em formato gráfico, visando simplificar a sua apresentação e interpretação. O desafio de transformar as séries temporais em informações compreensivas é considerável e inclui uma ampla variedade de tarefas distintas.

O objetivo desta pesquisa foi desenvolver e validar uma metodologia de visualização que tem por finalidade permitir a exploração e a identificação de padrões e tendências em séries temporais multidimensionais relacionadas ao monitoramento da qualidade do

ar em ambientes urbanos. Para tal fim, são construídas ferramentas analíticas baseadas em visualização e integradas com técnicas interativas que tornam os sistemas propostos capazes de reorganizar as informações apresentadas de acordo com a escolha do usuário, reduzindo, com isso, os problemas habitualmente relacionados à escalabilidade dos dados. As informações resultantes são exibidas em layouts acessíveis de alta qualidade gráfica, fundamentados em conceitos e propriedades de percepção visual.

Em resumo, as principais contribuições desta dissertação são:

- Um novo ferramental de visualização que permite a identificação das variações temporais de poluentes atmosféricos, não apenas com base nos parâmetros amostrados, mas também por meio de comparações entre diferentes intervalos de tempo;
- Sistemas de visualização de informações que integram técnicas exploratórias a um conjunto de visões ligadas, as quais representam as informações como um todo e também a partir de elementos selecionados;
- Disponibilização *online* e com livre acesso de ferramentas dedicadas ao monitoramento e análise visual de informações meteorológicas, ambientais e relacionadas aos índices de saúde apurados na região amostral;
- Um conjunto de estudos de caso que revelam características interessantes a respeito da dinâmica ambiental de São Carlos–SP, discutindo o comportamento evolutivo, as sazonalidades e as possíveis causas dos fenômenos observados.

A visualização se posiciona como uma das etapas de um amplo processo de tratamento, exploração e análise dos dados, transformando dados abstratos em informação visual. Claramente, as prioridades estéticas e considerações utilitárias em uma ferramenta de visualização dependem de um profundo estudo das características de seu domínio de aplicação (BYRON; WATTENBERG, 2008), sendo que os esforços empreendidos aqui, podem não ser generalistas o suficiente para serem aplicados a quaisquer conjuntos de dados variantes no tempo, mas são bastante eficientes para os fins aos quais se propõem.

Os trabalhos descritos aqui transcorreram em estreita colaboração com o grupo de pesquisas em Controle Ambiental e Monitoramento da Qualidade do Ar em Ambientes Externos, do Departamento de Engenharia Química da Universidade Federal de São Carlos (UFSCar). Foi submetido um artigo relacionado a esta temática, que está em revisão:

- ALEXANDRINA, E. C.; ORTIGOSSA, E. S.; LUI, E. S.; GONÇALVES, J. A. S.; CORRÊA, N. A.; NONATO, L. G.; AGUIAR, M. L. Analysis and visualization of multidimensional time series: Particulate matter (PM₁₀) from São Carlos-SP (Brazil), *Atmospheric Pollution Research*, Elsevier, 2018.

1.3 Organização do Texto

Para documentar e exibir o que foi produzido neste trabalho de Mestrado, este documento está estruturado da seguinte maneira:

- No Capítulo 2 são descritos em profundidade os conceitos teóricos que alicerçam o desenvolvimento desta pesquisa, em especial os elementos estilísticos e de *design* aplicados na representação gráfica de séries temporais. O Capítulo prepara o leitor com um esclarecimento sobre os fundamentos desta dissertação;
- No Capítulo 3 é apresentada uma revisão bibliográfica sobre alguns dos trabalhos do estado da arte na área de visualização de informações, principalmente aqueles que abordam ferramentas analíticas de visualização dedicadas à exploração interativa dos múltiplos atributos de dados com dependência temporal. Por meio de um estudo comparativo, as características dessas técnicas foram identificadas, com destaque em suas vantagens e limitações;
- No Capítulo 4 são especificados os processos de construção dos sistemas de visualização projetados durante esta pesquisa, revelando ao leitor as suas capacidades e as suas funcionalidades com detalhes sobre as escolhas gráficas e os métodos dedicados à extração de informações úteis no contexto das séries temporais multidimensionais;
- No Capítulo 5 são descritos os resultados obtidos. Foram realizados estudos de caso com as ferramentas desenvolvidas para demonstrar sua utilidade e o modo como estes sistemas podem ser empregados para representar as particularidades e tendências encontradas nos dados. Também são discutidas as possíveis causas dos cenários revelados pela análise visual empreendida;
- Finalizando, no Capítulo 6 são delineadas as conclusões desta produção, com as suas limitações, uma discussão dos possíveis trabalhos futuros, além de um resumo de suas principais contribuições para o processo de análise visual de séries temporais multidimensionais.

Capítulo 2

Conceitos e Fundamentos

2.1 Considerações Iniciais

Antes de mergulhar no processo de desenvolvimento, deve ser estabelecida a base conceitual necessária para empreender este esforço, desde uma análise sobre as características dos dados, dos sistemas gráficos, bem como o modo em que a teoria se aplica na prática. Neste Capítulo são apresentados e detalhados os fundamentos teóricos necessários para atingir os objetivos e contribuições esperadas desta dissertação: a construção de ferramentas analíticas de visualização, desenhadas para suportar a representação e exploração visual de dados multidimensionais variantes no tempo.

Na Seção 2.2 são identificadas as terminologias e tarefas que abrangem o ambiente de análise das séries temporais multidimensionais, bem como suas propriedades e técnicas exploratórias. Um histórico sobre visualização de séries temporais, conceitos e propriedades das metáforas utilizadas para representar estes conjuntos de dados e os desafios na área de visualização estão apresentados na Seção 2.3. Já na Seção 2.4, são descritas as tecnologias computacionais utilizadas para desenvolver os métodos e as metáforas propostas nesta pesquisa. Por fim, as considerações finais deste Capítulo estão expostas na Seção 2.5.

2.2 Séries Temporais

Séries temporais são conjuntos de registros que têm seus valores aferidos ao longo do tempo (FUA; WARD; RUNDENSTEINER, 1999). As séries temporais são produzidas continuamente e podem ser encontradas nos mais diversos domínios de aplicação como, por exemplo, finanças, saúde, meteorologia, astronomia, notícias, monitoramento e sensoriamento remoto, simulações científicas de experimentos, coleções de imagens, até

no planejamento ou gerenciamento de projetos. De um ponto de vista prático, cada um desses tipos de dado requer ferramentas de tratamento específicas (AIGNER *et al.*, 2007). Formalmente, um conjunto de séries temporais é expresso por:

$$D = \{(t_1, v_1), (t_2, v_2), \dots, (t_n, v_n)\} \quad (2.1)$$

com

$$v_i = f\langle t_i \rangle \quad (2.2)$$

A variável dependente, v_i , pode configurar propriedades dos mais variados tipos, desde valores inteiros, reais, categóricos, entre outros (WEBER; ALEXA; MÜLLER, 2001). Quando v_i é caracterizado por um único atributo, qualitativo ou quantitativo, o conjunto é chamado de bidimensional, pois a evolução temporal de seus pares ordenados (t_i, v_i) pode ser representada no espaço de um plano cartesiano, por exemplo. No entanto, cada instância v_i também pode ser individualmente composta por um subconjunto contendo um número arbitrário de elementos, $v_i = \{v_{i1}, v_{i2}, \dots, v_{ik}\}$. Estes k atributos descrevem o espaço de características das variáveis dependentes do conjunto de dados. Neste caso, tal série temporal será denominada de multidimensional, pois suas instâncias são interpretadas como pontos no espaço m -dimensional, em que m é o número total de elementos em (t_i, v_i) . Assim, no espaço de exploração desses pontos são produzidas famílias de curvas variáveis temporalmente para cada atributo da variável dependente (KONYHA *et al.*, 2012).

2.2.1 Multidimensionalidade

Antes de introduzir em profundidade a problemática envolvendo a representação de séries temporais, uma importante definição deve ser esclarecida. Embora seja possível encontrar os termos “multidimensional”, “multivariado” ou “multivalorado” para definir conjuntos de dados cujos elementos individuais são compostos por uma coleção de parâmetros, estes termos não são empregados de modo consistente na literatura (FUA; WARD; RUNDENSTEINER, 1999). Ao longo deste documento, por convenção, será adotado o termo multidimensional para denotar os conjuntos de séries temporais formados por registros contendo múltiplas variáveis, sejam elas dependentes e/ou independentes entre si.

Consequente a alta capacidade de coleta, armazenamento e transmissão de conteúdo dos dispositivos e sensores, as séries temporais tendem a ser grandes conjuntos de dados, não apenas quanto ao total de instâncias ou ao intervalo de captura, mas também quanto ao número de atributos associados a cada instância, o que acaba tornando a sua análise uma tarefa complexa e massante (AIGNER *et al.*, 2008; JAVED; ELMQVIST, 2010).

Essa complexidade de exploração não está relacionada apenas com a diversidade das fontes de dados, a variedade de tipos ou mesmo o tamanho dos conjuntos, definido pelo número de instâncias. A dificuldade de interpretação das séries temporais está significativamente associada com a dimensionalidade dos dados, ou seja, em como representar adequadamente o elevado número de elementos vinculados a cada intervalo de tempo. A Figura 2.1 ilustra a estrutura de um conjunto de séries temporais multidimensionais.

Figura 2.1 – Cubo de dados formado por séries temporais multidimensionais resultantes de um procedimento de simulação paramétrica.

Fonte: Gruendl *et al.* (2016).

Considere um experimento hipotético de simulação envolvendo otimização paramétrica. Neste contexto, um conjunto de parâmetros é dado como entrada do método de simulação (colunas vermelhas da Figura 2.1). Cada parâmetro é recebido e processado por diversos mecanismos que, ao final, produzirão um conjunto de instâncias de resposta (colunas azuis da Figura 2.1).

Para verificar o comportamento desse sistema quando submetido a variações nos parâmetros iniciais e, com isso, encontrar quais os valores que melhor se adequam aos cenários em estudo, a evolução do procedimento de simulação tem suas respectivas entradas e saídas registradas ao longo do tempo. Após sucessivas iterações, com diversos parâmetros, além de possíveis ajustes nos métodos de simulação, é obtido um conjunto de dados que contém todos os estímulos e respostas processados durante a simulação, em cada instante. Estes dados, quando organizados e vistos sequencialmente, formam uma estrutura semelhante a um “cubo” de dados temporais.

À medida que a dimensionalidade dos cubos temporais aumenta, os desafios de representação e interpretação de seu conteúdo se intensificam, adicionando, por consequência, mais complexidade sobre as tarefas analíticas correlatas a extração de informações úteis desses dados. A análise de tendências e a verificação de interdependência entre os parâmetros de séries temporais, a cada instante, configuram a essência da exploração de dados multidimensionais variantes no tempo (GRUENDL *et al.*, 2016).

2.2.2 Técnicas de Análise e Mineração de Dados

A primeira etapa da análise de dados consiste na identificação do problema, algo que requer um estudo do domínio de aplicação para que sejam definidos os objetivos de estudo e as necessidades de descoberta. Delimitado o escopo exploratório, os analistas podem partir para a extração de informações. Porém, as técnicas de análise de dados mais utilizadas, no geral, não se mostram escaláveis quanto a complexidade dos dados. Além disso, o fator tempo, que rege o perfil comportamental das variáveis dependentes, é frequentemente negligenciado pelas técnicas de análise habituais, sendo considerado apenas como mais um parâmetro quantitativo entre os demais, deixando nas mãos do usuário o seu controle (MULLER; SCHUMANN, 2003; AIGNER *et al.*, 2007).

Tradicionalmente, conjuntos de séries temporais são organizados dentro de planilhas eletrônicas. Essas planilhas suportam uma variada gama de operações em linha e coluna, sendo a inspeção de planilhas um método de análise bastante utilizado e que oferece aos analistas uma descrição detalhada do conteúdo dos dados. Entretanto, além de não haver um modo eficaz de organização que catalogue e priorize informações dispostas em múltiplas planilhas, sua inspeção não é uma metodologia muito eficiente em transmitir ao usuário uma visão clara da evolução de tendências, bem como as correlações entre os pontos de dado ao longo do intervalo em análise (CHANG *et al.*, 2007).

A mineração de dados é uma área de pesquisas voltada ao processo de extração de informações previamente desconhecidas em bases de dados. Por meio da aplicação de técnicas e modelos computacionais, a mineração auxilia o pesquisador a extrair padrões válidos, novos e potencialmente úteis a partir dos dados (REZENDE, 2003). Esses métodos são derivados principalmente de áreas como matemática, estatística e aprendizado de máquina. Neste sentido, a mineração de séries temporais busca modelar e compreender o fenômeno que governa o processo evolutivo no qual estes dados estão imersos, desde a identificação de características precursoras de anomalias até, muitas vezes, a realização de previsões. Mas a mineração é apenas parte de um processo amplo e mais genérico de exploração, denominado KDD (*Knowledge Discovery in Databases*) (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996), que é pouco útil caso o usuário não disponha de meios adequados para conseguir

inspecionar os padrões extraídos. Uma ferramenta completa de análise, que converta dados temporais não processados em informação útil, deve conter ao menos três componentes principais: sumarização de dados, apresentação e seleção/manipulação de características de interesse (FUA; WARD; RUNDENSTEINER, 1999).

Ao analisar dados provenientes de aplicações e dispositivos de mundo real, os valores registrados frequentemente contém os mais variados tipos de distorções. Por isso, como fase inicial do processo de análise, os dados devem ser organizados e estruturados. Entende-se por sumarização os métodos estatísticos de pré-processamento que são aplicados sobre as instâncias de dado e suas variáveis. Dados originados de múltiplas fontes podem manifestar diferentes características e padrões, que devem ser consolidados de acordo com os formatos aceitos pelos algoritmos que serão utilizados nos próximos componentes. Essa etapa envolve passos de transformação e “limpeza” nos dados, escalando valores e removendo ruídos indesejados, com a finalidade de organizar seu conteúdo, ressaltando propriedades nos atributos das múltiplas variáveis (CUI *et al.*, 2006).

É fundamental garantir a consistência entre as informações presentes nos dados brutos e as informações remodeladas nos dados pré-processados. Ou seja, os métodos de transformação aplicados nessa fase devem preservar ao máximo as características originais das séries temporais. Com isso, as propriedades importantes se mantêm, enquanto são tratados apenas os dados redundantes ou irrelevantes. A sumarização deve estar cuidadosamente alinhada com um profundo conhecimento das tarefas que serão consideradas pelas ferramentas de exploração, pois direciona os dados ao cumprimento das metas de análise.

Uma das principais finalidades da apresentação de séries temporais é o mapeamento da evolução dos dados de maneira consistente e compacta. Em outras palavras, deve ser possível identificar como os valores individuais das múltiplas séries se modificam no decorrer do tempo, enquanto são descritas características importantes dos dados originais, como seu perfil geral. A aquisição de informações úteis e o reconhecimento de particularidades nos dados podem ser simplificados quando estes são apresentados graficamente por meio de símbolos visuais. Assim, outro método exploratório muito utilizado é a inspeção de gráficos univariados que apresentam a evolução temporal da variável dependente.

Inspecionar muitos gráficos simultaneamente resulta apenas em uma visão fragmentada das informações (MCLACHLAN *et al.*, 2008), dificultando a identificação de padrões e a extração de informações que sejam de fato proveitosas. Deve ser levado em consideração que a memória de trabalho humana é limitada e a síntese de uma visão geral da evolução das informações pode ser prejudicada durante a tarefa de análise em múltiplas séries, visto que suas correlações podem não ser evidentemente pronunciadas.

Um importante aspecto do processo de exploração de dados é o oferecimento de

métodos de seleção que permitam a manipulação de seu conteúdo sob demanda (ALBERS; CORRELL; GLEICHER, 2014). Quando aplicadas, essas técnicas interativas atualizam o sistema de representação, exibindo diferentes níveis de detalhamento. Isso viabiliza ao analista realizar comparações entre subconjuntos de seu interesse, em diferentes porções dos dados. Mediante a seleção de parâmetros e a pesquisa de termos “chave”, por exemplo, a identificação de informações, que a princípio seriam dificilmente reveladas entre as múltiplas séries temporais, é facilitada. A combinação de técnicas de seleção de dados interativas altera a densidade de detalhes na representação e direciona a extração de características durante a exploração, permitindo uma verificação mais aprofundada entre as (dis)similaridades dos dados (MCLACHLAN *et al.*, 2008).

Séries temporais multidimensionais são encontradas nos mais variados formatos de dado, podendo ser aferidas com diferentes taxas de amostragem, dentro do mesmo conjunto. Não é incomum encontrar valores ausentes ou mesmo ruídos diversos entre suas instâncias. Isso faz da extração de informações nessas séries uma tarefa complexa que deve ser realizada meticulosamente, caso contrário, todo o esforço dedicado na coleta e no armazenamento de tais dados será desperdiçado (KEIM, 2002).

Quanto mais instâncias são reunidas, mais difícil se torna analisá-las, ressaltando características significativas entre suas muitas variáveis. Os analistas desejam não apenas interpretar a evolução desses dados ao longo do tempo, mas também examinar suas alterações durante todo o processo de coleta, para que seja possível ter uma visão geral da natureza dos dados e de seu comportamento. Logo, uma questão crítica é como melhor transformar esses dados em algo que as pessoas possam compreender, otimizando, com isso, o processo de tomada de decisões (WARE, 2004).

2.3 Visualização de Informações e Séries Temporais

Paralelamente ao desenvolvimento da área de mineração de dados, tem crescido as pesquisas em visualização de informações. Com finalidades complementares, ambas as áreas buscam desenvolver soluções para aprimorar o processo de interpretação dos dados. A visualização é mais especificamente direcionada a apoiar esse processo, por meio da criação de novos modelos gráficos de representação visual, a partir dos quais a aquisição de informações úteis dos dados é facilitada por ferramentas de manipulação e interação com o usuário (OLIVEIRA; LEVKOWITZ, 2003). A mineração visual, ou simplesmente visualização, pode ser vista como a união de técnicas originadas em ambas as áreas, inserindo estratégias visuais durante a mineração e oferecendo meios de avaliação dos resultados obtidos no processo de descoberta de conhecimento (WONG, 1999).

A Figura 2.2 ilustra as etapas da visualização. O primeiro estágio corresponde à obtenção dos dados, que são amostrados ou gerados por algum método de sintetização. Inicialmente, os dados são apenas registrados e dispostos em sistemas de armazenamento, como as já citadas planilhas. Essa é a etapa mais longa do processo, pois envolve o ambiente, ou o “mundo real”. Assim, o ambiente físico é a fonte dos dados, com o ambiente social determinando as convenções e modos como o ambiente físico deve ser interpretado (WARE, 2004). A princípio, os dados brutos nada têm a ensinar sobre o ambiente de origem devido a sua complexidade de interpretação. É a partir do primeiro nível de tratamento que os dados são arranjados em tabelas. Com isso, os dados são preparados para serem mais facilmente modelados em algo passível de entendimento. Neste estágio são aplicadas as técnicas de mineração, pré-processando e transformação. Também aqui, os atributos de interesse devem ser identificados, sendo que a extração dos mesmos pode ser feita por seleção manual ou com a assistência de métodos automatizados.

Figura 2.2 – Diagrama esquemático da metodologia interativa de visualização de informações.

Fonte: Adaptada de Ware (2004).

Em seguida, são desenvolvidos os algoritmos gráficos que modelarão, por meio de símbolos e propriedades visuais, as informações extraídas na etapa anterior. A escolha desses algoritmos depende do domínio de aplicação dos dados e deve estar alinhada com as necessidades dos analistas. Então, os atributos identificados são quantificados de acordo com os critérios específicos do modelo de visualização e a sua evolução é graficamente acompanhada em função do tempo, no caso das séries temporais. Card, Mackinlay e Shneiderman (1999) definiram a etapa de modelagem gráfica como uma sequência de mapeamentos que gera uma representação visual ajustável, visto que deve permitir a interação do analista com a informação. É essencial que a visualização reflita claramente as mudanças decorridas no período de amostragem dos dados.

No último estágio está o usuário, que assimilará a informação recebida, manipulando o conteúdo e a profundidade de detalhamento da visualização. Utilizando um dispositivo como o *mouse*, o analista pode selecionar intervalos que lhe sejam mais interessantes,

por exemplo. Essa interação tem por finalidade explorar as informações apresentadas, estabelecendo comparações entre os diferentes pontos de vista dos dados, o que facilitaria o processo de conversão da informação em conhecimento pelo usuário. Segundo Dimenstein e Cortella (2015), a distinção entre informação e conhecimento é que a informação é cumulativa, já o conhecimento é seletivo. Não basta ter informação, é preciso saber o que fazer com ela. No entanto, a informação é a base do conhecimento, ao passo que, a partir da informação acumulada e à disposição, pode ser selecionado o que se precisa.

A conexão do analista com o espaço de informações, apoiada por uma interface de visualização interativa, impulsiona o processo cognitivo de descoberta de conhecimento. Isso favorece a percepção de características antes indistintas, permitindo a captura de informações potencialmente úteis, que serão posteriormente analisadas e transformadas em conhecimento. A partir do conhecimento, é possível compreender e, caso necessário, modificar o ambiente de origem dos dados.

Em meados do século XVIII, um engenheiro escocês chamado William Playfair, foi o pioneiro em analisar dados de séries temporais financeiras utilizando gráficos de linha e de barras. Apesar de historicamente não ter sido o inventor dessas metáforas, Playfair foi quem popularizou o método de representação gráfica de informações quantitativas. Seu trabalho teve significativa relevância para a ciência contemporânea, sendo ele considerado um dos “pais” dos gráficos estatísticos (DER; EVERITT, 2014). Para Playfair, gráficos são ferramentas preferíveis a tabelas, pois esses possuem uma grande capacidade de comunicação ao representar a “forma” dos dados, permitindo ao leitor realizar análises comparativas (TUFTE, 2007). No entanto, apesar do valor informativo, suas avaliações eram limitadas à confrontação de uma ou duas séries temporais no mesmo espaço.

O célebre mapa elaborado em 1869 pelo engenheiro francês Charles Joseph Minard, apresentado na Figura 2.3, é considerado uma das primeiras representações efetivas de séries temporais multidimensionais. Minard utilizou um mapa de dados para descrever o fluxo de soldados do exército francês durante a desastrosa campanha de Napoleão contra a Rússia, em 1812.

À esquerda da Figura 2.3, a linha colorida de bronze indica a posição inicial dos franceses, estacionados na Polônia, quando, em junho, Napoleão invadiu a Rússia com seu poderoso exército de aproximadamente 450 mil combatentes, talvez a maior força armada já constituída até o momento (SUTHERLAND, 1985). A largura da linha é proporcional ao número de soldados franceses em cada uma das regiões do mapa. Ou seja, além do tamanho do exército, essa linha apresenta as rotas traçadas pelos franceses ao longo da campanha. Quando chegou a Moscou, em setembro, Napoleão se viu a frente de um exército enfraquecido, justamente no início do rigoroso inverno russo. Empregando táticas

Figura 2.3 – Representação evolutiva das múltiplas variáveis envolvendo a campanha de Napoleão contra a Rússia, decorrida em 1812.

Fonte: Tufte (2007).

de guerrilha, além do conhecimento de seu próprio território, os russos resistiram. Em outubro, os franceses foram obrigados a recuar, com seu exército destruído. Estima-se que cerca de 10 mil soldados retornaram à França, numa das maiores derrotas militares da história (AUSTIN, 1996). A linha de cor preta descreve o trajeto de volta dos combatentes.

Segundo Tufte (2007), o gráfico de Minard é, possivelmente, a melhor representação estatística já feita. Explorando espaço e tempo simultaneamente, a metáfora sintetiza a evolução do conflito, detalhando o número de soldados, sua localização, sua direção e a temperatura ambiente em cada instante do evento.

2.3.1 Desafios em Visualização de Séries Temporais

Nos últimos anos, muitos estudos têm sido direcionados para a representação gráfica de séries temporais multidimensionais (FUA; WARD; RUNDENSTEINER, 1999), sendo este um tema recorrente na comunidade de visualização de informações. Em sua análise sistemática sobre visualização de dados temporais, Aigner *et al.* (2007) definiram que as ferramentas modernas de visualização dedicadas a este tema devem viabilizar a percepção da evolução temporal dos dados, ressaltando padrões e tendências, com o objetivo de gerar informações que aprimorem a compreensão dos fenômenos analisados.

No entanto, as ferramentas de exploração visual modernas ainda estão fortemente baseadas em modelos de visualização tradicionais, como gráficos de linhas, histogramas, gráficos de setores circulares e gráficos de barras (HEER; BOSTOCK; OGIEVETSKY, 2010). De fato, os gráficos de linhas são um dos mais simples e populares métodos de

se visualizar dados variantes no tempo, demonstrando grandes capacidades de síntese de informações. Porém, enquanto as metáforas de linhas representam de maneira efetiva algumas séries temporais, em diversos experimentos, mesmo os corriqueiros, pode haver a necessidade de se monitorar o comportamento de mais de uma variável. Dentro do contexto de dados multidimensionais, é tarefa comum explorar a evolução de múltiplas séries em simultâneo (JAVED; MCDONNEL; ELMQVIST, 2010). A sequência de imagens apresentadas na Figura 2.4 ilustra essa questão.

Figura 2.4 – Comparação da legibilidade individual ao se analisar múltiplas séries temporais em uma metáfora simples de linhas.

Fonte: Elaborada pelo autor.

Ao incrementar a quantidade de curvas representadas dentro do mesmo espaço, a capacidade de transmissão de informações individualizadas se reduz drasticamente. É importante estudar os relacionamentos entre as séries individualmente. Mais especificamente, deve ser possível reconhecer as contribuições individuais, ou seja, quais as séries que contribuem positiva ou negativamente para a evolução do conjunto como um todo (WANG *et al.*, 2016). Logo, explorar múltiplas curvas de modo compreensivo em um gráfico de linhas mostra-se difícil, principalmente devido à limitação de espaço e resolução, que resulta em problemas de escalabilidade, como sobreposições e desordem visual (PERIN; VERNIER; FEKETE, 2013). Outras metáforas, por exemplo, as de barras, podem sumarizar grandes

volumes de dados, mas necessitam de muito espaço por elemento único. Além de verificar a correlação entre as famílias de curvas, os analistas precisam entender a relação entre as séries temporais, individualizando-as em diferentes escalas e intervalos de tempo.

A escalabilidade visual é um fator que sempre deve ser levado em consideração no desenvolvimento de sistemas gráficos. Define-se por escalabilidade visual, a capacidade dos sistemas e ferramentas de visualização em efetivamente representar conjuntos de dados com suporte a volumes crescentes, tanto em termos do montante de instâncias quanto do número de dimensões (EICK; KARR, 2002). A oclusão, que é a obstrução de símbolos gráficos devido a sobreposição, é um problema de escalabilidade recorrente quando se manipula grandes conjuntos. Neste contexto, representar milhares de registros no limitado espaço visual das telas de computadores, é um dos grandes desafios na área de visualização de informações (SHNEIDERMAN, 2008). Outro importante aspecto dos dados variantes no tempo é a dinâmica do seu fluxo de informações, que pode exigir a atualização periódica da visualização (KRSTAJIC; BERTINI; KEIM, 2011).

Quanto maiores os conjuntos, maiores serão os tempos necessários para processamento e geração dos gráficos. Um bom sistema de visualização também deve ser escalável computacionalmente, ou seja, suas ferramentas devem apresentar agilidade de resposta, mesmo ao manipular grandes quantidades de dados. Logo, a construção de novas metáforas capazes de integrar a representação de séries temporais com o controle do processo exploratório, transmitindo uma visão clara das informações, é uma questão fundamental.

2.3.2 Percepção e Propriedades Gráficas

Cleveland e McGill (1984) definiram percepção gráfica como a habilidade dos usuários em interpretar as codificações visuais, decodificando e compreendendo as informações apresentadas graficamente. A maior parcela da informação recebida pelo sistema cognitivo humano é sensorialmente obtida pela visão. A luz é refratada através dos componentes anatômicos dos olhos até atingir a retina, tecido composto por milhões de células fotossensíveis que transformam estímulos luminosos em impulsos elétricos. Porém, a informação não é interpretada pela retina. Os impulsos são enviados até o córtex visual, na parte traseira do cérebro, região responsável pela decodificação desses estímulos, que classifica, ordena e analisa as informações, fazendo uso ativo de comparações do conteúdo observado com aquele armazenado na memória visual (HARRISON; TONG, 2009).

Logo, a percepção de cenas e objetos que envolvem a transmissão de informações, não é um processo de observação meramente passivo. A pesquisa em percepção gráfica, com o objetivo de reduzir o esforço cognitivo na visualização de dados, é feita desde muito

antes dos computadores transformarem gráficos em ferramentas interativas. A partir do estudo de leis fundamentais da psicologia comportamental sobre problemas elementares de percepção, surgiu no início do século XX, uma das mais famosas doutrinas no que se refere a organização e ao agrupamento de elementos visuais: a teoria *Gestalt*.

De origem alemã, a palavra *Gestalt* pode ser traduzida como forma ou padrão. Uma das principais aplicações desse conceito está na compreensão e explicação do modo que o sistema cognitivo humano consegue reconhecer os estímulos sensitivos, em pontos que possuem características em comum (padrões), agrupando estes, em um novo e maior objeto visual (CHANG; NESBITT; WILKINS, 2007). A teoria *Gestalt* se baseia na premissa de que a percepção do padrão global de uma cena não pode ser explicada simplesmente pela soma de suas partes (WAGEMANS *et al.*, 2012a). Em outras palavras, enquanto um elemento único apresenta um contexto específico, quando em conjunto com outros elementos, estes passam a representar características e contextos distintos.

Para a teoria *Gestalt* o cérebro é um detector de padrões e discrepâncias que, dentre os vários objetos em seu campo visual, possui habilidades para identificar, distinguir e agrupar os elementos, de acordo com as variações de suas características em comum, formando conjuntos maiores e mais complexos. Explorar essa capacidade de agrupamento é de grande interesse para os desenvolvedores dedicados ao *design* de ferramentas, que integrarão sistemas de comunicação mais poderosos no sentido de permitirem aos usuários que encontrem padrões no meio de grandes quantidades de informações. Buscando orientar uma “configuração” conveniente do cenário diante dos observadores, tornando os dados apresentados mais assimiláveis, foram formulados princípios dentro da teoria *Gestalt*, com alguns deles ilustrados na Figura 2.5 e resumidos pelos seguintes itens:

- Princípio da proximidade – Elementos que estão espacialmente próximos são pré-atentivamente processados como um grupo, uma unidade que compartilha objetivos similares, mesmo se agrupados de modo não necessariamente explícito;
- Princípio da semelhança – Elementos representados por marcadores que compartilham características de forma semelhantes (tamanho, cor, orientação, simetria, paralelismo), são perceptualmente agrupados;
- Princípio do “destino comum” – Elementos que passam por transformações visuais semelhantes tendem a ser mentalmente agrupados. O dinamismo de movimento auxilia o espectador a perceber quais objetos estão relacionados à mesma ação;
- Princípio da continuidade – A silhueta de curvas e elementos suavemente arredondados é melhor reconhecida como parte de um grupo contínuo do que a união entre pontos por segmentos angulares;

- Princípio do fechamento – Elementos delimitados dentro de áreas com contornos nítidos, ainda que não totalmente contínuos, tendem a ser visualmente agrupados.

Figura 2.5 – Exemplo de alguns dos princípios *Gestalt* de agrupamento de objetos.

Fonte: Wagemans *et al.* (2012a).

Uma revisão abrangente sobre as origens históricas da teoria *Gestalt*, bem como uma discussão detalhada sobre organização e percepção por agrupamento de objetos, podem ser encontradas em Wagemans *et al.* (2012a). Wagemans *et al.* (2012b) apresentaram um estudo sobre as recentes formulações conceituais envolvendo os princípios *Gestalt*, desde a estruturação eficiente de informações em sistemas visuais até pesquisas neurológicas sobre os complexos estados perceptuais.

Segundo Aigner, Rind e Hoffmann (2012), as ferramentas de visualização que adequadamente tiram proveito das propriedades gráficas, apresentam melhor capacidade de transmissão de informações quando aplicadas a tarefas complexas que envolvem múltiplas análises, se comparadas às técnicas tradicionais. Logo, os sistemas de visualização modernos devem explorar aspectos de percepção visual, entre eles: o gerenciamento de espaço, o uso de símbolos e formas familiares ao domínio de aplicação, a orientação dos pontos de dado, a identidade visual entre parâmetros distintos e a capacidade de lidar com a desordem visual gerada por grandes quantidades de informações exibidas simultaneamente (LOHSE, 1991; JAVED; MCDONNEL; ELMQVIST, 2010).

Hao *et al.* (2005) analisaram a influência de variáveis de aspecto e alinhamento na visualização de múltiplas séries temporais. Os autores afirmaram que os elementos posicionados na parte superior da tela são cognitivamente classificados como mais importantes. De modo semelhante, são considerados importantes os elementos à esquerda, local em que usualmente estão os rótulos e demais informações úteis a respeito da escala vertical em gráficos. O posicionamento espacial ainda revela a ordem dos dados, sendo considerado um dos mais expressivos princípios perceptuais em visualização (MCLACHLAN *et al.*, 2008). As cores também são das mais sensíveis propriedades gráficas, uma vez que

contrastes bem definidos facilitam a leitura. Ou seja, dois objetos adjacentes serão mais facilmente discriminados quanto maiores forem seus contrastes, ao passo que o esforço necessário para se realizar tal individualização será maior caso a diferença de contraste seja suave. A dimensão também é um canal de informação, sendo que elementos maiores são interpretados como mais importantes, em relação aos menores.

O estudo da percepção cognitiva em sistemas de visualização de séries temporais tem avaliado como as diferentes propriedades gráficas influenciam nas habilidades dos analistas em comparar de maneira efetiva as múltiplas séries (ALBERS; CORRELL; GLEICHER, 2014). A partir de considerações cuidadosas das propriedades gráficas, os desenvolvedores dos sistemas de representação podem definir adequadamente o modo como os dados, a visualização e as tarefas analíticas interagem, simplificando o processo de comunicação, para que o usuário absorva a essência das informações. Pesquisas abrangentes sobre percepção gráfica e boas escolhas de *design* para sistemas de visualização podem ser encontradas em Adnan, Just e Baillie (2016) e Albers, Correll e Gleicher (2014). Javed, McDonnell e Elmqvist (2010) estudaram sistematicamente a performance de diversas representações, elencando importantes fatores de percepção gráfica, além de recomendações para a escolha adequada de métodos para a visualização de múltiplas séries temporais.

O objetivo da visualização é a comunicação. Nesta pesquisa, tanto a forma quanto a função dos elementos gráficos desenvolvidos têm como base as *guidelines* da teoria *Gestalt* e os recentes estudos em organização de sistemas gráficos. *Guidelines* são recomendações sucintas e úteis que descrevem abordagens para problemas, baseadas no conhecimento obtido por especialistas no domínio que sumarizaram sua experiência prática solucionando questões semelhantes (CRAFT; CAIRNS, 2005). Com isso, espera-se que a complexidade dos dados seja reduzida, facilitando a comunicação visual e a compreensão das informações ao apresentar os objetos de modo a explorar as habilidades perceptuais dos analistas.

Porém, as habilidades perceptuais humanas são limitadas, chegando um momento em que a percepção gráfica se torna impraticável devido a grande quantidade de informações exibidas concorrentemente, gerando sobrecarga e desordem visual (JAVED; MCDONNELL; ELMQVIST, 2010). Nesses casos, os sistemas de visualização devem se unir com métodos de exploração e de reorganização das informações em tela.

2.3.3 Categorização de Metáforas e Elementos Visuais

Por mais que a tecnologia atual tenha avançado, os *displays* gráficos dos dispositivos utilizados para apresentação de informações pertencem ao domínio bidimensional. Para contornar essa limitação física, os pesquisadores em visualização têm criado e explorado

ferramentas que viabilizem a representação de dados multidimensionais. Dados estes compostos por conjuntos de atributos que não podem ser simplesmente mapeados em planos cartesianos ou mesmo eixos espaciais tridimensionais (CHALMERS, 1996). Analogamente, a literatura a respeito de técnicas de análise, modelagem e visualização de séries temporais é extensa. Aigner *et al.* (2008) apresentaram um *framework* para categorização das diferentes abordagens de análise visual sobre dados temporais, sendo que discussões aprofundadas podem ser encontradas em pesquisas (SILVA; CATARCI, 2000; AIGNER *et al.*, 2007) e livros (AIGNER *et al.*, 2011; SHUMWAY; STOFFER, 2011) dedicados ao tema.

A representação de séries temporais multidimensionais é um problema recorrente. Javed, McDonnel e Elmqvist (2010) classificaram as técnicas de visualização de múltiplas séries em duas categorias: as de espaço compartilhado e as de espaço dividido. Nas técnicas de espaço compartilhado, as séries temporais são desenhadas dentro da mesma área gráfica (PERIN; VERNIER; FEKETE, 2013). Uma visualização em que os conjuntos de dados têm suas curvas sobrepostas nos mesmos eixos pode facilitar a comparação direta entre os valores, uma vez que o resultado é a sumarização visual dos múltiplos valores em uma única visão agregada das séries temporais (HEER; KONG; AGRAWALA, 2009). Porém, a quantidade de informações que podem ser representadas simultaneamente, sem ocasionar desordem visual, é limitada. Exibir múltiplos dados diferentes no mesmo espaço (mais de quatro) pode dificultar a legibilidade individual das informações, gerando sobrecarga cognitiva (cf. Figura 2.4d), principalmente quando o número de dimensões do dado motivar a necessidade de se utilizar codificações de cores com contrastes similares (JAVED; MCDONNEL; ELMQVIST, 2010).

Já nas técnicas de espaço dividido, cada uma das séries ocupa o seu próprio espaço, geralmente unidades (*views*) de tamanho reduzido. É preferível dividir as dimensões de séries multidimensionais que possuem tipos de dados diferentes entre si, sendo que as metáforas gráficas utilizadas para isso costumam não diferir, dependendo do contexto. Mostrar cada série lado a lado em seu próprio gráfico pode resultar em um ganho prático no tempo de estimativa das informações, em termos da minimização das sobrecargas ocasionadas por uma única e complexa visualização dos dados, que pode ser cognitivamente massante (MCLACHLAN *et al.*, 2008). Por outro lado, a precisão das estimativas também pode diminuir, ao passo que exibir múltiplos gráficos simultâneos exige que estes sejam construídos em porções de espaço individuais cada vez menores, podendo levar a oclusão de detalhes (KRSTAJIC; BERTINI; KEIM, 2011).

A Figura 2.6 exibe uma comparação entre dois sistemas de visualização, um de espaço compartilhado e o outro de espaço dividido. Os gráficos apresentam dados sobre os índices de desemprego nos Estados Unidos, com as variáveis correspondendo aos setores

da indústria entre os anos de 2000 e 2010. Na Figura 2.6a, foi utilizado um *Stream graph*, que é uma técnica de espaço compartilhado que desenha as séries temporais empilhado-as uma sobre a outra ao longo dos instantes de tempo (BYRON; WATTENBERG, 2008). Essa metáfora será abordada detalhadamente na Subseção 3.2.5. Já na Figura 2.6b, cada variável ocupa seu próprio espaço em pequenos gráficos de área. Neste caso, o autor optou por utilizar a mesma metáfora para todas as séries. Note que a razão de aspecto das metáforas se mantém homogênea (HAO *et al.*, 2005).

Figura 2.6 – Visualizações de espaço compartilhado e espaço dividido aplicadas sobre o mesmo conjunto de dados.

(a) *Stream graph* compartilha o espaço ao empilhar uma série temporal sobre a outra.

(b) Múltiplas *views* dividem as séries temporais em gráficos individualizados.

Fonte: Heer, Bostock e Ogievetsky (2010).

Já Wattenberg (2005b) identificou a categoria de técnicas baseadas em preenchimento espacial. Normalmente, a área que de fato caracteriza os dados em um gráfico ocupa apenas uma porção da área total entre os eixos do gráfico. As estratégias de

preenchimento espacial buscam o aproveitamento de todo (ou virtualmente todo) o espaço de tela disponível, sendo, por isso, consideradas eficientes em termos de aproveitamento de espaço (JUGEL *et al.*, 2014). Embora boa parte das técnicas de preenchimento espacial sejam voltadas a visualização de dados arranjados em estruturas hierárquicas (HEER; BOSTOCK; OGIEVETSKY, 2010; STASKO; ZHANG, 2000), a demanda por visualização de séries temporais multidimensionais em espaços limitados também tem motivado o desenvolvimento de métodos que utilizam o conceito de preenchimento espacial (PERIN; VERNIER; FEKETE, 2013; HAO *et al.*, 2005).

Card, Mackinlay e Shneiderman (1999) classificaram como híbridas as técnicas que interconectam múltiplas abordagens de visualização para investigar uma entidade conceitual complexa e/ou de natureza heterogênea. Essas técnicas podem utilizar duas ou mais *views* distintas, com o objetivo de aprimorar a expressividade e a capacidade de transmissão de informações das visualizações (OLIVEIRA; LEVKOWITZ, 2003). Neste contexto, são consideradas *views* distintas quando elas enfatizam, e permitem ao usuário compreender, diferentes aspectos de um dado (BALDONADO; WOODRUFF; KUCHINSKY, 2000).

Devido a baixa qualidade de resolução dos dispositivos utilizados até meados dos anos 2000, que praticamente impedia a apresentação detalhada de grandes quantidades de informações com uma única técnica, os pesquisadores introduziram os métodos multi-resolução. Nestes sistemas, as informações são exibidas em metáforas conectadas e com níveis crescentes de detalhamento, o que permite a exploração dos dados selecionados em maior profundidade (LAM; MUNZNER; KINCAID, 2007). Mesmo com a significativa melhoria na qualidade dos *displays*, a crescente complexidade nos dados, em termos de volume e dimensionalidade, impulsionou os desenvolvedores a evoluírem os ambientes multi-resolução para as interfaces híbridas com múltiplas *views* conjugadas, posicionando estas como alternativas na visualização de dados multifacetados (CHANG *et al.*, 2007).

Os sistemas modernos de múltiplas *views* devem ser construídos para tirar o máximo de proveito da diversidade dos dados, utilizando diferentes unidades e elementos gráficos para estabelecer conexão visual direta entre as variáveis e o eixo do tempo, suportando, com isso, a exploração por meio da navegação interativa entre as séries temporais, o que é essencial para uma análise visual bem-sucedida (AIGNER *et al.*, 2007). Ao particionar os dados em múltiplas *views* interligadas, aplicam-se estratégias exploratórias baseadas no conceito de “dividir para conquistar”, gerando porções de informação gerenciáveis que reduzem o esforço de memória necessário para o usuário compreender a informação em tela (BALDONADO; WOODRUFF; KUCHINSKY, 2000).

A Figura 2.7 exemplifica uma interface híbrida composta por quatro grupos de informação: um conjunto de parâmetros de configuração (Figura 2.7a); um painel de

coordenadas paralelas (Figura 2.7b, esta metáfora será discutida na Subseção 3.2.4); um repositório de séries temporais (Figura 2.7c); e um grupo com múltiplos gráficos, apresentando variáveis selecionadas interativamente pelo usuário (Figura 2.7d). O arranjo espacial entre os elementos da interface deve ser consistente, visando minimizar a necessidade do usuário trocar frequentemente de contexto (BALDONADO; WOODRUFF; KUCHINSKY, 2000) ou mesmo navegar repetidamente entre *views* que estejam distantes entre si, garantindo, com isso, que a atenção do analista esteja nos lugares corretos.

Figura 2.7 – Sistema de visualização composto por múltiplas *views* distintas e interligadas.

Fonte: Gruendl *et al.* (2016).

Para que a exploração de dados seja efetiva, a interação com o usuário é tão importante quanto a própria apresentação dos dados (WIJK; SELOW, 1999). Isto posto, os sistemas híbridos são profundamente dependentes de técnicas de interação e mecanismos de acoplamento que coordenem a navegação entre os elementos do sistema. Em outras palavras, as ações aplicadas em uma *view* devem ser mapeadas e automaticamente propagadas, com fluidez, para as outras *views*. Esse tipo de tarefa requer métodos de gerenciamento de dados sofisticados, que tornam a construção dos sistemas híbridos ainda mais desafiadora (BALDONADO; WOODRUFF; KUCHINSKY, 2000). Logo, os *designers* têm que dedicar cuidado extra com os erros de codificação, que podem ser difíceis de corrigir devido ao forte intrincamento dos elementos desse tipo de ambiente.

Apresentar múltiplos níveis de detalhamento simultaneamente, fornece ao usuário um contexto de alta densidade de informações, propriedade que aumenta o poder analítico sobre os dados para permitir a exploração sob diferentes perspectivas (SADANA; STASKO,

2016). Neste sentido, os sistemas multivisão aprimoram o desempenho exploratório ao reduzir o tempo de acesso sobre dados estruturalmente complexos, haja vista que é cognitivamente mais simples realizar múltiplas comparações visualmente, do que baseadas na memória de trabalho do usuário. Essa característica favorece a descoberta de relações imprevistas dentro de um ambiente de trabalho unificado e coordenado, em que uma informação complexa é dividida em múltiplas *views* simplificadas (NORTH; SHNEIDERMAN, 1997). As *views* devem ser posicionadas conforme os princípios de organização perceptual, descritos na Subseção 2.3.2, e interconectadas por ferramentas de interação. Assim, sistemas de visualização híbridos com *views* integradas surgem como uma solução adequada para explorar visualmente séries temporais multidimensionais.

As metáforas gráficas dedicadas às séries temporais são numerosas e, por isso, outros autores agruparam as técnicas de representação em categorias alternativas, considerando a geometria de seus elementos, a iconografia, o acesso à informação, entre outras características de estruturação visual (*layout*). Estudos compreensivos sobre taxonomias de técnicas de visualização podem ser encontrados em Oliveira e Levkowitz (2003) e Grinstein, Trutschl e Cvek (2001). Heer, Kong e Agrawala (2009) realizaram uma análise acerca de métodos espaço-eficientes para se visualizar séries temporais multidimensionais, enquanto Roberts (2007) e Baldonado, Woodruff e Kuchinsky (2000) sumarizaram uma visão geral sobre os fundamentos e os requisitos de *design* para sistemas multivisão.

2.3.4 Técnicas Exploratórias e Interativas

Os sistemas de visualização modernos são compostos essencialmente por dois componentes principais: a representação visual em si, que se refere ao mapeamento dos dados em modo gráfico, e as técnicas de interação, que são os meios de comunicação entre os usuários e os sistemas (YI *et al.*, 2007). No entanto, poucos estudos em visualização se dedicaram a avaliar os impactos sobre a percepção gráfica ocasionados pela utilização de técnicas de interação em representações visuais, com boa parte dos trabalhos abrangendo o ponto de vista puramente de classificação ou de desenvolvimento (ELMQVIST *et al.*, 2011). Historicamente, os estudos em visualização de informações concentraram seus esforços em desenvolver métodos para a representação visual dos dados. Ao passo que a interação e o *design* de componentes que proporcionem a experiência do usuário em navegar pelos diferentes aspectos da visualização ficaram a cargo da comunidade de pesquisas correlata em interação humano-computador (ADNAN; JUST; BAILLIE, 2016).

Enquanto os conjuntos de dados se tornam cada vez mais densos, as abordagens tradicionais para a visualização de séries temporais tendem a sofrer com a desordem e a oclusão de detalhes, já discutidos na Subseção 2.3.1, além de dependerem de ações de

rolagem (*scrolling*) para que seja possível alcançar todas as informações contidas nos pontos de dado (HAO *et al.*, 2005). Esta operação fragmenta a visão geral dos dados, sendo dependente da memória de trabalho do analista.

Com a evolução dos dispositivos, em termos de qualidade gráfica e poder computacional, que proporcionam a criação dos sistemas multivisão, as técnicas de interação com o usuário se mostram indispensáveis para coordenar e integrar o contexto das informações em tela. É por meio dos mecanismos de interação que o analista seleciona pontos em um gráfico, que são destacados nas demais *views* relacionadas. Com isso, os múltiplos níveis de abstração e detalhamento oferecem suporte a exploração dinâmica de padrões nos dados (HEER; BOSTOCK; OGIEVETSKY, 2010). Neste sentido, a capacidade de interagir diretamente com a visualização pode reduzir as limitações das ferramentas gráficas, uma vez que estas passam a oferecer meios para se aliviar a complexidade dos dados e controlar o fluxo das informações em análise.

Shneiderman (1996) introduziu os conceitos teóricos daquela que é reconhecida como uma das mais notáveis estratégias de exploração visual de dados: “*overview first, zoom and filter, then details-on-demand*” (visão geral primeiro, *zoom* e filtro, então os detalhes sob demanda). A abordagem parte da premissa de que certos padrões e tendências são observados apenas quando se analisa os dados a partir de um ponto de vista que compreenda o seu formato geral. Por exemplo, regiões do conjunto de dados contendo picos de máximos/mínimos locais podem facilmente ser confundidas com o ponto de máximo/mínimo global, induzindo o analista a tomar decisões baseadas em uma informação aparentemente adequada, porém equivocada.

Assim, em uma aplicação de visualização de informações, a visão geral apresenta o conjunto de dados sob a ótica de um contexto amplo, descrevendo uma imagem de toda a entidade conceitual que a ferramenta de visualização representa. Entretanto, o número de itens pode escalar aos milhares em grandes bases de dados, como as séries temporais multidimensionais, tornando impraticável fornecer detalhes sobre todos os itens apresentados (CRAFT; CAIRNS, 2005).

Ao identificar características consideradas interessantes, o usuário pode selecioná-las para verificar suas particularidades. Detalhes sob demanda significa selecionar um objeto, ou um grupo de objetos, e apresentar seus detalhes ponto a ponto, com o analista seletivamente ajustando a profundidade de detalhamento, ocultando ou revelando dados de seu interesse (SHNEIDERMAN, 1996). *Zoom* e filtro são métodos de exploração que envolvem operações interativas de redução na complexidade da visualização, removendo da interface as informações não essenciais e reorganizando a representação em tela (CRAFT; CAIRNS, 2005). Então, com a coleção inicial de elementos reduzida, torna-se mais fácil

navegar individualmente entre os detalhes dos dados destacados (SHNEIDERMAN, 1996).

Apesar de estreitamente correlacionados, *zoom* e filtro são abordagens distintas. Craft e Cairns (2005) classificaram o *zoom* como um método de filtragem de dados por navegação, que controla o conteúdo do *display* sem a necessidade de alterar o contexto representacional da visualização, eliminando os itens indesejados ao agir sobre o ponto de vista da representação. Ainda segundo os mesmos autores, a dinâmica da estratégia de detalhes sob demanda é um processo de abstração que segue o Princípio da Omissão Seletiva. O sistema cognitivo processa as informações sensoriais em estágios de organização, ou seja, dados complexos tendem a ser reduzidos para quantidades gerenciáveis, por organização e/ou simplificação de seu formato. Com isso, as informações podem ser mentalmente processadas, com rapidez suficiente para que as mudanças no ambiente sejam respondidas em tempo hábil (RESNIKOFF, 1989).

O termo “filtro” é comumente empregado pela comunidade de visualização para nomear o processo de delimitação dos dados apresentados, de acordo com alguma metodologia de seleção, seja por pesquisas interativas, ajuste de parâmetros de configuração do sistema ou mesmo pela escolha dos pontos de dado, selecionando-os diretamente com o cursor do *mouse*, por exemplo. Porém, “filtro” e suas derivações são termos amplamente utilizados em outras áreas de pesquisas, apresentando significados específicos. Para evitar possíveis ambiguidades, aqueles processos que envolvem a delimitação interativa do espaço de dados, nomeados em visualização de informações como filtragem, serão caracterizados nesta dissertação por termos como “seleção de dados” ou simplesmente “seleção”.

Ao trabalhar com séries temporais multidimensionais, cada ponto de dado carrega diversas informações que devem ser apresentadas para análise direta pelo usuário. Uma interface gráfica eficiente garante que as informações importantes estejam sempre visíveis, mesmo em regiões altamente densas. Esta é uma definição simples, que representa um grande desafio de desenvolvimento (MCLACHLAN *et al.*, 2008). Neste sentido, a estratégia de visão geral com detalhes sob demanda se encaixa bem quando aplicada aos sistemas de múltiplas *views*, pois busca contornar a complexidade visual e revelar os detalhes dos itens selecionados sem alterar o contexto representacional da visualização. Após o analista selecionar um grupo de elementos ou um intervalo de tempo específico, as unidades gráficas da interface são atualizadas para apresentar os detalhes desses itens. Esta ação é particularmente importante ao favorecer a descoberta de ligações e sazonalidades entre os atributos de dois (ou mais) elementos, relacionando as informações descobertas com restante do conjunto de dados (CRAFT; CAIRNS, 2005).

Aigner *et al.* (2007) definiram a coordenação de múltiplas *views*, como a correta propagação de uma interação originada de uma *view* para todas as outras *views* que são

coordenadas. Uma questão importante surge quanto à coordenação entre as múltiplas *views* dos sistemas de visualização. O acompanhamento preciso das alterações é dificultado quando há demora (*delay*) entre a interação e o efeito de transição, que ajustará o conteúdo resultante de tal interação na interface (HEER; ROBERTSON, 2007). Em outras palavras, as ações do usuário têm que receber resposta imediata do sistema, atualizando a representação e permitindo que o processo de exploração de dados seja de fato interativo. No entanto, as mudanças visuais não devem ser abruptas (MCLACHLAN *et al.*, 2008). Isto é, a ferramenta gráfica deve proceder interações fluidas.

Elmqvist *et al.* (2011) descreveram a fluidez nos sistemas de visualização como um conceito intangível, mas caracterizado por efeitos suaves e contínuos, aplicados sobre gráficos que respondem aos comandos do usuário rapidamente. Devido a essa intangibilidade, construir ferramentas com *design* fluido não é uma tarefa trivial. Elmqvist *et al.* (2011) ainda defenderam que esta é uma abstração poderosa, que pode transformar o processo exploratório em uma experiência abrangente e agradável, ajudando o analista a compor um modelo mental do fluxo evolutivo das informações. Os autores também elaboraram em seu estudo um conjunto de *guidelines* voltadas ao desenvolvimento de sistemas que suportem interações responsivas e integradas às representações visuais.

São numerosas as técnicas interativas já desenvolvidas para aprimorar a navegação em sistemas de visualização. Estas vão desde as mais simples, como os destacamentos, que atuam no controle visual atraindo a atenção do usuário para algum subconjunto de dados específico (ADNAN; JUST; BAILLIE, 2016), até as técnicas mais sofisticadas, que utilizam animações para orientar os espectadores durante os eventos de transição.

Heer e Robertson (2007) afirmaram que a animação é uma abordagem promissora para facilitar a percepção de mudanças entre gráficos que contém informações relacionadas. Por meio de experimentos supervisionados, os autores avaliaram as vantagens das transições animadas em operações para reduzir a oclusão, redimensionando eixos e reescalando valores. Porém, esses efeitos não garantem um bom desempenho exploratório, com as conclusões da pesquisa desestimulando o uso de transições muito complexas, além de alertar que a introdução de animações em gráficos estatísticos exige um estudo cuidadoso, pois elas podem induzir os analistas ao erro, caso os seus efeitos violem a semântica visual dos dados. Ainda segundo Heer e Robertson (2007), animações simples são preferíveis para suavizar e tornar mais agradáveis as atualizações de tela necessárias durante as tarefas interativas, sugerindo que estas devem apresentar um tempo de resposta entre 1 e 3 segundos.

Sadana e Stasko (2016) introduziram um sistema interativo de múltiplas *views* coordenadas, arquitetado especialmente para *tablets*, que são dispositivos móveis com telas sensíveis ao toque normalmente sem periféricos de entrada, como teclados ou *mouses*.

Apesar de serem aparelhos com boas capacidades computacionais, *tablets* e *smartphones* apresentam restrições de espaço, visto que, em geral, suas telas têm de 4 a 13 polegadas. A interação nestes dispositivos ocorre diretamente com o toque das mãos do usuário, logo, imprecisões podem ocorrer durante tarefas que exijam a seleção de informações, por exemplo. Entretanto, esses aparelhos são muito populares atualmente, de modo que os sistemas de visualização capazes de explorar as capacidades de *tablets* e *smartphones* terão alcance a um público-alvo consideravelmente maior. Para isso, os autores realizaram uma análise sobre os requisitos de desenvolvimento para interfaces multivisão, aliadas a técnicas de interação consideradas apropriadas para selecionar dados e coordenar as *views* apresentadas nesses dispositivos.

A Figura 2.8 apresenta alguns comandos gestuais para interagir no sistema proposto por Sadana e Stasko (2016). Observe que o usuário deve pressionar a tela com dois dedos, afastando-os, para obter um *zooming* (Figura 2.8a-b). Para definir uma área de seleção de informações, o usuário toca em algum ponto da visualização e arrasta horizontalmente até que a área compreenda os itens de seu interesse (Figura 2.8c).

Figura 2.8 – Recursos interativos em ferramenta de visualização de informações para dispositivos com *displays* sensíveis ao toque.

Fonte: Adaptada de Sadana e Stasko (2016).

Atualmente, a manipulação interativa de recursos integra o processo de raciocínio analítico, com virtualmente todas as ferramentas de visualização modernas suportando algum nível de exploração de dados por meio do mapeamento dinâmico de seus atributos (OLIVEIRA; LEVKOWITZ, 2003). Neste contexto, Yi *et al.* (2007) definiram uma taxonomia sobre os objetivos da interatividade em visualização de informação. McLachlan *et al.* (2008) descreveram princípios de desenvolvimento para sistemas de representação baseados em exploração interativa aplicados a múltiplas *views*, em que as séries temporais podem ser comparadas lado a lado, dentro da mesma metáfora visual e em diferentes níveis de detalhamento. Perin, Vernier e Fekete (2013) compararam gráficos estáticos com abordagens interativas para avaliar quantitativamente a eficácia de representações visuais alternativas aplicadas a conjuntos de múltiplas séries temporais. Segundo estes autores,

seu trabalho mostra que técnicas interativas consideradas simples unificam diferentes *views*, aprimorando substancialmente as tarefas exploratórias.

Pike *et al.* (2009) identificaram em seu estudo sobre a interatividade como ciência, os desafios em desenvolver boas ferramentas interativas, defendendo que estas devem atuar com naturalidade, parecendo ações óbvias e sem chamar a atenção do usuário para seus mecanismos de operação, por mais aprimorados que sejam, pois isso poderia gerar confusão. Para Pike *et al.* (2009), o bom *design* é aquele que apenas funciona. Os analistas necessitam reconhecer com precisão os elementos nas diferentes *views* da interface gráfica para compreender as relações entre seus estados atual e anterior, durante as interações inerentes ao processo exploratório (HEER; ROBERTSON, 2007).

Como a análise visual concentra seus esforços na relação entre as representações e a cognição humana, o simples desenvolvimento de novas metáforas gráficas raramente é suficiente para estimular a percepção (PIKE *et al.*, 2009). Assim, nesta pesquisa, o desenvolvimento gráfico foi integrado às técnicas interativas, com o objetivo de auxiliar o usuário na identificação de detalhes sob demanda de padrões e tendências em séries temporais multidimensionais. As transições entre os estados da visualização são suavemente animadas, obtendo, com isso, um ambiente de análise orientado e contínuo.

2.4 Tecnologias e Ferramentas Utilizadas

A evolução da web como é conhecida hoje está intimamente relacionada com o desenvolvimento das linguagens de programação, entre elas, o HTML (*Hypertext Markup Language*). Mais especificamente, o HTML é uma linguagem de marcação, ou seja, suas instruções, chamadas de *tags*, definem características da estrutura de um texto, desde tamanhos e tipos de fontes, até cores, espaçamentos e posicionamentos (W3C, 2017c). Diferentes *tags* podem estruturar diferentes trechos ou mesmo todo o texto. Com sua especificação consolidada, o HTML se tornou a linguagem de marcação mais utilizada para formatar o conteúdo de páginas web, sendo suportado por basicamente todos os navegadores (*browsers*) modernos.

No entanto, até meados da década de 1990, cada navegador interpretava e apresentava ao usuário o conteúdo das páginas web de acordo com suas próprias especificações. Não havia uma padronização mínima entre os diversos *browsers*, sendo frequente encontrar instabilidades ao se acessar uma página que havia sido desenvolvida seguindo as definições de um navegador diferente àquele que se estava utilizando. No final dos anos de 1990, havia uma “guerra” entre os navegadores, em que cada um tentava liderar o mercado desses softwares e estabelecer seu modelo como padrão, com as empresas desenvolvedoras

dos *browsers* chegando inclusive ao embate judicial. Essas disparidades criavam barreiras tecnológicas que reduziam a experiência de navegação dos usuários, desestimulando o acesso e, conseqüentemente, reduzindo o potencial da web.

Com o objetivo de mitigar os problemas de compatibilidade, a W3C (*World Wide Web Consortium*), que é a organização responsável pelo estabelecimento de especificações na *World Wide Web*, começou a trabalhar para convencionar um modelo de desenvolvimento multiplataforma para a representação e interação em documentos HTML (e outras linguagens de marcação). Foi nesse cenário que surgiu o DOM (*Document Object Model*), como um esforço para padronizar o desenvolvimento web.

DOM é uma interface de programação de aplicativos que especifica o modo como os documentos em HTML devem ser representados, baseando-se no bem estabelecido paradigma de programação orientada a objetos (W3C, 2000). O modelo define os relacionamentos lógicos do documento, criando uma estrutura hierárquica de programação que os objetos da interface devem seguir. Com isso, as incompatibilidades entre navegadores foram reduzidas drasticamente, uma vez que todos suportam o DOM (MURRAY, 2013). Ainda assim, as páginas web eram, em sua grande maioria, interfaces de conteúdo estático. Qualquer interação que os usuários desejassem realizar em seu conteúdo, para ser efetivada, a página deveria ser atualizada e recarregada pelo navegador, o que empobrecia a fluidez do processo. Esse cenário começou a mudar em 2008, com a introdução do HTML5 e o desenvolvimento dos métodos de processamento assíncrono (HALES, 2012), que propiciou o surgimento das páginas dinâmicas.

Neste sentido, a especificação do DOM permite ainda a portabilidade para a execução de *scripts* e programas, além da aplicação de elementos estilísticos dentro da estrutura HTML por meio de diversas linguagens de programação, o que enriquece o conteúdo das páginas web. *JavaScript* é a linguagem aceita como padrão pela maioria dos *browsers* modernos (CROCKFORD, 2008; HALES, 2012) para criar, acessar, navegar e gerenciar interativa e dinamicamente os objetos da interface. Por ser independente de plataforma ou sistema operacional, seus *scripts* podem ser executados nos mais variados dispositivos, incluindo os chamados dispositivos móveis. Já o CSS (*Cascading Style Sheets*) é uma linguagem declarativa voltada a aplicação de refinamentos estéticos (cores, estilo de fonte, posicionamento) nos elementos gráficos (W3C, 2017a). CSS oferece mecanismos simples e flexíveis de resposta a eventos baseados em *JavaScript*.

Todo esse conjunto de esforços empreendidos pela comunidade web contribuíram para fazer da Internet a ferramenta de comunicação poderosa que é hoje. Paralelamente a essa evolução, surgiu a necessidade de transmitir informações em sistemas visuais de alta qualidade. Aliado aos aprimoramentos no HTML5, foi desenvolvido o elemento *Canvas*,

dedicado à codificação gráfica interativa (W3C, 2017b). No entanto, as imagens criadas em *Canvas* são dependentes de resolução. Ou seja, após definidas, podem sofrer distorções quando exibidas em diferentes escalas. Isso pode ser problemático, uma vez que as telas dos dispositivos possuem as mais variadas dimensões.

Essa limitação foi superada com o SVG (*Scalable Vector Graphics*), que é uma linguagem modularizada para desenhar vetores bidimensionais (W3C, 2011). Os objetos criados em SVG vão desde formas geométricas simples, texto, até imagens mais complexas, que podem ser agrupadas e transformadas em conjunto. Além de também ser um dos padrões definidos pela W3C e, assim como os demais paradigmas, ter uma vasta documentação disponível (BOSTOCK; OGIEVETSKY; HEER, 2011). O diferencial em se utilizar SVG para desenvolvimentos gráficos reside no fato da linhagem criar “vetores escaláveis”. Com isso, ao ampliar ou reduzir uma figura em SVG, transformações matemáticas são executadas com o propósito de escalar a imagem, fazendo com que esta não sofra distorções significativas. Todos os atributos e propriedades dos elementos SVG podem ser acessados interativamente via DOM, recorrendo a métodos escritos em *JavaScript*.

Como pôde ser visto, a paleta de tecnologias dedicadas ao desenvolvimento web é vasta. No entanto, o grande sortimento de ferramentas gera ao menos um problema claro: para construir interfaces de alta qualidade, os desenvolvedores devem dominar profundamente as especificações dessas múltiplas ferramentas. Isso implica dispendear razoáveis quantidades de tempo em treinamentos até que os profissionais tenham fluência nas linguagens. De fato, o conceito de bibliotecas não é algo recente em programação. Bibliotecas são desenvolvidas e utilizadas há décadas para reunir coleções modularizadas de métodos e funções complexas, provenientes de uma ou mais linguagens, com a finalidade de simplificar a vida dos programadores.

Nesse contexto, surgiram abordagens como Processing (REAS; FRY; MAEDA, 2007), Protovis (BOSTOCK; HEER, 2009), entre outras, que oferecem aos desenvolvedores, coleções de módulos voltados à construção de sistemas de visualização dinâmicos. Apesar de terem contribuído com o desenvolvimento gráfico, ambas as bibliotecas têm limitações. Protovis sofre com a escalabilidade computacional, apresentando baixa performance quando lidando com grandes quantidades de dados (BOSTOCK; OGIEVETSKY; HEER, 2011). Enquanto o Processing utiliza um padrão de renderização proprietário e não especificado pela W3C, o que pode resultar em instabilidades devido a falta de suporte em alguns navegadores, algo que não ocorreria utilizando o SVG, por exemplo.

Já a D3.js (BOSTOCK; OGIEVETSKY; HEER, 2011), ou simplesmente D3 (*Data-Driven Documents*), é uma biblioteca escrita em *JavaScript* para a construção de aplicações, em que os dados são dinamicamente mapeados e associados a representações gráficas por

meio de objetos SVG (MURRAY, 2013). D3 é capaz de carregar informações numéricas e categóricas de diversos tipos, utilizando métodos computacionais de manuseio de dados (*data handlers*) simplificados. Com estes métodos, é possível ler os padrões de arquivos mais utilizados para o armazenamento de informações, entre eles o CSV (*Comma-Separated Values*), TSV (*Tab-Separated Values*), inclusive páginas HTML também podem servir como dados de entrada. Assim, por aliar alta qualidade visual, recursos interativos, eficiência computacional e uma boa documentação, D3 foi a biblioteca gráfica escolhida para o desenvolvimento web desta pesquisa.

Um dos principais motivos para se oferecer um sistema de visualização de dados *online* está na abrangência mundial da web, além de sua facilidade de utilização. O público-alvo não precisará instalar softwares específicos pra acessar a aplicação e explorar as informações. Porém, apesar dos *browsers* modernos terem incluído sofisticadas ferramentas gráficas, este ainda não é um ambiente otimizado em termos de memória e, principalmente, capacidade de processamento (BOSTOCK; OGIEVETSKY; HEER, 2011). Mesmo com a D3 apresentando boas avaliações, a performance dos navegadores pode ser um obstáculo para tratar quantidades acentuadas de operações sobre dados.

Foi escolhido o *Python* para desenvolver localmente os *scripts* que executem operações que sejam dependentes de processamento mais intenso. *Python* é uma linguagem de programação de propósito geral, que tem se tornado bastante popular na computação científica (PEDREGOSA *et al.*, 2011). *Python* oferece ao desenvolvedor um ambiente rico em bibliotecas de código aberto, dedicadas ao processo de mineração de dados, transformações matemáticas algébricas otimizadas, além da geração de gráficos (MCKINNEY, 2012). Assim como a biblioteca D3, a documentação e os exemplos disponíveis de aplicações em *Python* são vastos, reforçando a sua escolha para a análise de dados.

2.5 Considerações Finais

A visualização tem o poder de transformar fenômenos em elementos visíveis e compreensivos (SHNEIDERMAN, 1996). Neste Capítulo foram discutidos os conceitos que envolvem os sistemas de visualização, que têm por objetivo tornar mais explícitas as informações que estão implícitas em meio aos dados brutos. Conforme exposto, são muitas as questões que devem ser consideradas e que podem influenciar na capacidade das ferramentas visuais em transmitir efetivamente as informações, desde a escolha adequada das metáforas, o uso de técnicas interativas que facilitem a manipulação dos dados, um *design* que explore as habilidades cognitivas, até mesmo a qualidade e o tamanho dos dispositivos utilizados para acessar o sistema (THOMAS; COOK, 2005).

A adequada apresentação de uma série temporal multidimensional é fundamental para a sua interpretação, pois a estrutura desses dados é complexa de se “manipular” em seu estado bruto. Da maneira em que esses dados geralmente são disponibilizados (planilhas) não é rápido, ou mesmo fácil, encontrar características relevantes. Por isso, o desenvolvimento de técnicas de visualização é essencial para se obter entendimento sobre as séries temporais, sumarizando e dispondo as informações mais intuitivamente, e oferecendo métodos que permitem aos analistas examinar os dados desde a sua estrutura.

Comumente, um comportamento anômalo é aquele que desvia substancialmente do que é esperado ou “normal”. Mas o que constitui uma anomalia pode variar dependendo da tarefa em mãos, transformando a identificação de padrões desconhecidos em um problema particularmente difícil (LIN *et al.*, 2004). Segundo a teoria *Gestalt*, há uma tendência cognitiva em se completar a informação implícita faltante. Isso significa que uma parte daquilo que se vê é complementada pelo cérebro, de acordo com o conteúdo apreendido na memória visual. Esse conteúdo pode diferir bastante a cada espectador, sofrendo grande influência de seu conhecimento e experiências pessoais.

Assim, um bom sistema de visualização explora mecanismos gráficos que estimulem as capacidades ativas, atenuando o ruído existente entre emissor e espectador, engajando este durante o processo de detecção de padrões e anomalias. A exploração visual de séries temporais, quando unida a paradigmas de interação com o usuário, permitem a realização de testes de hipóteses que se mostram eficientes na solução de problemas que incluem a descoberta de padrões (LIN *et al.*, 2004).

Capítulo 3

Trabalhos Relacionados

3.1 Considerações Iniciais

Segundo Albers, Correll e Gleicher (2014), nenhum padrão de desenvolvimento gráfico dominará em todas as tarefas de visualização de informações, quando se trata de séries temporais. Por isso, as ferramentas devem ser fruto de um profundo estudo do domínio de aplicação, das necessidades analíticas dos usuários, bem como das funcionalidades a serem desempenhadas pelos sistemas visuais que essas ferramentas integram.

A construção de sistemas de representação que sejam capazes de revelar padrões e o modo como estes padrões evoluem com o tempo, em grandes quantidades de dados, não é uma tarefa trivial. Explorar visualmente séries temporais multidimensionais requer ferramentas que suportem uma navegação cuidadosa, além de participativa por parte do analista. Neste Capítulo será apresentada uma revisão da literatura, recorrendo a uma análise sobre as características e funcionalidades desenvolvidas em algumas das pesquisas mais relevantes em visualização e analítica visual.

Na Seção 3.2 são destacados alguns trabalhos que aplicam a visualização, especificamente aqueles estudos que desenvolveram sistemas e/ou técnicas analíticas assistidas graficamente e dedicados ao tratamento de séries temporais multidimensionais, discutindo sua metodologia e suas principais vantagens e fragilidades. Na Seção 3.3 estão algumas considerações sobre este Capítulo.

3.2 Estado da Arte em Visualização de Séries Temporais

Uma rica variedade de ferramentas e técnicas dedicadas à análise de dados têm sido desenvolvida para auxiliar na exploração de séries temporais (AIGNER *et al.*, 2008).

Discussões abrangentes podem ser encontradas em pesquisas (KEHRER; HAUSER, 2013; SILVA; CATARCI, 2000) e livros (AIGNER *et al.*, 2011). Porém, análises exploratórias efetivas, que extraíam informações úteis a partir de dados com múltiplas entidades variantes no tempo, podem compreender tarefas bastante desafiadoras (THAKUR; HANSON, 2010). É neste contexto que as ferramentas de visualização constituem a sua importância prática, ao transformar dados abstratos em informação visual, que é mais facilmente assimilável. Visualizações eficientes devem apoiar o entendimento das informações transmitidas, ou seja, as interfaces devem oferecer meios para auxiliar os usuários na interpretação dos dados, de acordo com a área de conhecimento em que atuam (FEDERICO *et al.*, 2014).

Visualizar similaridades em séries temporais é uma questão importante em muitas aplicações (FU, 2011). Algumas ferramentas constroem visualizações que organizam os dados de acordo com suas similaridades globais, dependendo de métodos de agrupamento (ZIEGLER *et al.*, 2010) e *layouts* com grades adaptativas (HAO *et al.*, 2007) para comparar visualmente as diferentes séries. Os dados temporais também podem ser recursivamente mapeados em subconjuntos ordenados dentro de partições do ambiente gráfico, com base em sua importância global (HAO *et al.*, 2005), ou transformados em miniaturas, que são utilizadas para comparações (KUMAR *et al.*, 2005).

Conforme apresentado na Subseção 2.3.4, a interatividade é parte dos sistemas gráficos modernos. Adnan, Just e Baillie (2016) investigaram os efeitos dos *tooltips*, que são mensagens de texto apresentadas ao usuário quando o cursor do *mouse* é posicionado sobre um elemento da interface, facilitando a identificação de características e valores quantitativos dos objetos representados. Os autores afirmam que essa técnica de interação aprimora consideravelmente a experiência de navegação sobre a visualização de séries temporais, sem perdas significativas de eficiência e precisão analítica. Os resultados da pesquisa sugerem que alguns usuários tendem a preferir instruções textuais simples e claras sobre os pontos de seu interesse, ante a alternativas de interação mais complexas que demandem esforço de aprendizado.

Goethem *et al.* (2017) utilizaram seleção de informações com *brushing* em seu sistema para facilitar a detecção de tendências em séries temporais. O *brushing* é um método de seleção de subconjuntos de dados mediante uma interação direta do usuário com a representação gráfica. Normalmente, para definir uma área de seleção com *brushing*, o usuário deve clicar com o *mouse* na região da interface que contém o ponto inicial das informações a serem selecionadas. Então, com o botão pressionado, ele arrasta o cursor até que a área de seleção compreenda todos os pontos de seu interesse. Quando o botão do *mouse* é liberado, os dados compreendidos dentro da área definida são destacados e podem ser enviados como entrada para algum método de verificação (cf. Figura 2.8c).

3.2.1 Metáforas Espirais e Radiais

Visualizações espirais são alternativas em tarefas que buscam identificar características cíclicas em séries temporais. Weber, Alexa e Müller (2001) apresentaram os *spiral graphs*, metáforas com o propósito de dar suporte à detecção de comportamentos periódicos previamente desconhecidos, por meio da estruturação do eixo temporal em espiral (AIGNER *et al.*, 2008). A Figura 3.1 ilustra duas representações do mesmo conjunto de dados, que registra casos de gripe ao longo de um intervalo de três anos, utilizando barras (Figura 3.1a) e o gráfico em espiral (Figura 3.1b). Os *spiral graphs*, entretanto, requerem que os dados sejam adequadamente parametrizados, algo usualmente difícil de ser obtido sem que haja um conhecimento prévio de seu conteúdo (ANDRIENKO *et al.*, 2010), além de não serem vantajosos para séries que tenham ciclo periódico longo.

Figura 3.1 – Comparação entre duas metáforas aplicadas a séries temporais que contém eventos periódicos.

Fonte: Aigner *et al.* (2008).

Métodos de visualização orientados a comparação entre séries temporais também aplicam mecanismos interativos para assistir na identificação de similaridades locais. Esses métodos variam em termos das métricas e metáforas visuais empregadas para comparar os períodos. Köthür *et al.* (2015) utilizaram janelas de correlação cruzada combinadas com quadros visuais, enquanto Krstajic, Bertini e Keim (2011) aplicaram a estratégia de detalhes sob demanda em uma ferramenta compacta para análise visual de múltiplos eventos em dados temporais. Para permitir a verificação aprofundada em áreas densas, estes autores utilizaram distorções de área por meio de lentes de aumento, que são técnicas interativas desenvolvidas para inspecionar pontos de dado, alargando-os. O usuário pode

definir o nível de ampliação e as dimensões da região sob a lente, operando de modo semelhante ao *zoom*, porém, com atuação local e podendo ser reposicionada com o *mouse*.

Zhao, Chevalier e Balakrishnan (2011) apresentaram *KronoMiner*, uma ferramenta para a exploração dinâmica de séries temporais, com a proposta de ser flexível o suficiente para apoiar tarefas analíticas complexas com independência de domínio. Baseado em um *display* de informações radial, a ferramenta permite a manipulação direta dos dados mediante seleções que podem destacar e rearranjar diferentes partes das séries temporais, simultaneamente. Os detalhes dos dados são obtidos com uma técnica interativa nomeada pelos autores como *MagicAnalytics Lens*, um tipo de lente que, além da capacidade de aumento, compara pares de séries temporais em tempo real. *KronoMiner* trata os dados de maneira bastante generalista, podendo lidar com uma grande variedade de informações ao custo de perder alguma capacidade analítica mais aprofundada. A Figura 3.2 ilustra a ferramenta e seus componentes, aplicados sobre séries temporais do mercado financeiro.

Figura 3.2 – *KronoMiner* é composto por uma barra de ferramentas (a), estrutura radial para apresentar os múltiplos níveis de detalhes dos dados (b), visão geral (c), área de comparação (d), lente para uma visão detalhada (e), um *tooltip* (f), e um menu de contexto (g).

Fonte: Zhao, Chevalier e Balakrishnan (2011).

Seguindo um conceito similar, Zhao *et al.* (2011) desenvolveram *ChronoLenses*, que estende as capacidades das lentes interativas de *KronoMiner* para a exploração de dados temporais, ao permitir que um número arbitrário de séries sirvam como entrada para as suas lentes. Os autores ainda incluíram outras ferramentas interativas na interface, por meio das quais os usuários conseguem comparar e analisar estatisticamente as regiões de interesse em uma ou múltiplas séries, em diferentes estágios. Isto é, os resultados das transformações entre dois conjuntos temporais em uma lente podem servir como informações de entrada para outra lente, por exemplo. Vários trabalhos utilizam a abstração das lentes entre

as suas técnicas interativas. Tominski *et al.* (2017) revisaram a literatura sobre o tema, identificando os principais estudos relacionados a esta técnica e discutindo suas abordagens e propriedades em um contexto geral de visualização.

3.2.2 Calendários

Wijk e Selow (1999) desenvolveram um sistema baseado em calendário para explorar similaridades periódicas em séries temporais. Os dados são inicialmente subdivididos em sequências diárias e, então, é aplicado um algoritmo de agrupamento para classificar as séries que apresentarem padrões semelhantes. A ferramenta gráfica é composta por dois componentes principais: um calendário combinado a um gráfico de múltiplas linhas. No calendário, cada dia é colorido de acordo com o grupo ao qual foi classificado. Já o gráfico de linhas apresenta as curvas que caracterizam cada uma das classes de séries temporais, definidas pelo algoritmo de agrupamento empregado.

A Figura 3.3 apresenta o sistema, analisando padrões diários de frequência dos funcionários de um laboratório de pesquisas na Holanda, presentes ao longo do dia, dia após dia, durante um ano. A ferramenta oferece uma boa visão geral das informações, porém, sua aplicação está condicionada apenas a dados com características periódicas que podem ser formatadas em um calendário, ou seja, tem pouca utilidade para séries que não apresentem padrões diários ou semanais (LIN; KEOGH; LONARDI, 2005).

Figura 3.3 – Distribuição dos padrões de frequência dos funcionários em um laboratório durante o ano de 1997.

Fonte: Wijk e Selow (1999).

3.2.3 Tempo, Espaço e Mapas

Boyandin *et al.* (2011) apresentaram *Flowstrates*, uma interface de visualização dedicada a análise das variações temporais em conjuntos georreferenciados, que contém dados de movimentações origem-destino de entidades entre diversas localidades. A ferramenta utiliza dois mapas laterais e um *heat map* (GRINSTEIN; TRUTSCHL; CVEK, 2001) central para representar separadamente a dimensão espacial da dimensão temporal, permitindo a manipulação individual das informações. A Figura 3.4 ilustra a visão geral do sistema *Flowstrates*. Cada linha do *heat map* central define um par origem-destino, com a coloração das suas células descrevendo a magnitude do fluxo em cada intervalo de tempo. Na parte esquerda, está posicionado o mapa de origem dos dados. À direita, está o mapa dos destinos. Ambos os mapas são ligados ao tempo por meio de *links*.

O exemplo da Figura 3.4 foi aplicado sobre dados referentes ao deslocamento de refugiados, partindo da Etiópia, cujo fluxo mais expressivo foi com a Somália. Note que a partir de 1988, a migração de etíopes com destino à Somália foi reduzida drasticamente, chegando mesmo à inversão do fluxo migratório. Esse dado coincide e pode ser explicado pelo início, em 1986, das tensões que levaram a Somália a deflagrar uma violenta guerra civil, evento que teve seu auge¹ entre os anos de 1991 e 1993, quando os principais destinos para os etíopes passaram a ser o Sudão e o Quênia. Apesar de sumarizar bem as ligações entre os componentes visuais, *Flowstrates* sofre com a oclusão, além de impedir a percepção espacial direta da distância percorrida pelas entidades em seu deslocamento.

Figura 3.4 – *Flowstrates* apresentando informações espaço-temporais de fluxos migratórios com origem na Etiópia.

Fonte: Boyandin *et al.* (2011).

¹O conflito armado na Somália permanece sem solução até o momento desta escrita.

Ainda no contexto de séries temporais multidimensionais contendo informações variantes também espacialmente, Thakur e Hanson (2010) propuseram um sistema de visualização utilizando mapas interativos tridimensionais. A ferramenta empilha verticalmente os dados temporais em colunas com discos poligonais, que são geograficamente distribuídas sobre os mapas 3D. Cada disco define um intervalo de tempo, sendo a evolução dos valores traduzida mediante o uso de cores e com o diâmetro dos discos, que são escalados de acordos com os valores normalizados dos dados.

Embora seja bastante interativo, o sistema sofre com a oclusão entre os elementos gráficos, algo que é relativamente comum em sistemas de visualização 3D. No entanto, os autores incluíram uma metáfora 2D para representar as classificações regionais e, com isso, contornar essa limitação. A Figura 3.5 ilustra a aplicação da ferramenta, apresentando os índices de desemprego no estado norte-americano da Carolina do Norte. As pilhas verticais descrevem as taxas para cada região do estado (condados), podendo ser interativamente selecionadas para inspeção e comparação, individualmente ou em grupos.

Figura 3.5 – Aplicação de mapas 3D sobre as informações de desemprego no estado da Carolina do Norte, Estados Unidos, de 1999 a 2008.

Fonte: Thakur e Hanson (2010).

Sistemas de visualização tridimensionais, como o da Figura 3.5, costumam utilizar o conceito de cubo espaço-temporal (*space-time cube*) para caracterizar a evolução espacial dos objetos ao longo do tempo. Na abordagem de cubo espaço-temporal, o componente espacial é representado no plano horizontal, enquanto o domínio temporal é apresentado na vertical, estabelecendo, com isso, um volume de informações. Em Bach *et al.* (2014) pode ser encontrada uma revisão sobre a visualização de dados temporais, bem como um conjunto de operações relativas ao cubo espaço-temporal. Esse estudo foi recentemente estendido por Bach *et al.* (2017). Já Andrienko e Andrienko (2013) realizaram uma extensa revisão sobre as ferramentas e os métodos dedicados a visualização espaço-temporal, sob a perspectiva de deslocamentos e movimentações.

3.2.4 Coordenadas Paralelas

Figura 3.6 – Painel temporal em “pseudo-perspectiva”, conectando dois eixos adjacentes do sistema de coordenadas paralelas.

Fonte: Gruendl *et al.* (2016).

A visualização de dados multidimensionais em coordenadas paralelas foi inicialmente desenvolvida em meados dos anos 1980, com a metáfora se notabilizando principalmente após os trabalhos de Inselberg e Dimsdale (1990). Desde então, a ferramenta é aplicada em várias pesquisas que envolvem a representação de informações multidimensionais. As coordenadas paralelas quebram o paradigma de eixos coordenados ortogonais ao dispor horizontalmente múltiplos eixos verticais (paralelos entre si) ao longo da área gráfica. Cada eixo define uma dimensão do conjunto de dados, com pares de eixos interligados por segmentos geométricos que os cruzam nos pontos correspondentes às magnitudes da variável nesta dimensão. O perfil de cada segmento ao longo de todos os eixos descreve o comportamento de cada entrada de dado nas múltiplas dimensões do conjunto.

As coordenadas paralelas oferecem vantagens para comparar variáveis de magnitudes distintas, além de facilitar a verificação de padrões e tendências gerais dos dados. Teoricamente, a quantidade de dimensões representadas simultaneamente não é um grande problema, porém, na prática, visualizar densos conjuntos de dados em coordenadas paralelas pode resultar em sobreposições de elementos, dificultando a percepção de correlações entre as múltiplas dimensões (FUA; WARD; RUNDENSTEINER, 1999).

Neste contexto, Fua, Ward e Rundensteiner (1999) aprimoraram a metáfora aplicando técnicas interativas de agregação com o objetivo de estruturar os dados em diferentes níveis de abstração, diminuindo, com isso, a densidade e a desordem visual. Apesar de ser uma ferramenta gráfica bem estabelecida, o sistema de coordenadas paralelas se mostra pouco útil em revelar as variações de dados dependentes do tempo. Esta questão foi

inicialmente enfrentada por Tominski, Abello e Schumann (2004), que organizaram as coordenadas radialmente. Os autores posicionaram o eixo temporal ao centro da metáfora, com as demais variáveis dispostas circularmente, ao seu redor. Entretanto, a abordagem sofre com a sobreposição, além de não facilitar uma visão geral do conjunto.

Gruendl *et al.* (2016) introduziram um engenhoso esquema de “pseudo-perspectiva” nas coordenadas paralelas, por meio de um painel temporal que se abre entre dois, ou mais, eixos paralelos, e que pode ser movido para frente e para trás. O painel exibe a evolução das séries temporais, tratando o tempo como uma extensão em profundidade da tela, algo que permite ao analista verificar as flutuações dos atributos com a dependência do tempo. A Figura 3.6 ilustra a atuação do painel temporal, que é parte integrante de uma interface de visualização ampla, composta por um conjunto de ferramentas interativas destinadas a comparar séries individuais ou grupos de séries temporais (cf. Figura 2.7).

3.2.5 As Metáforas Empilhadas

Os *stacked graphs* (gráficos empilhados) são abordagens de espaço compartilhado, projetados para a visualização de dados multidimensionais, que literalmente empilham um atributo sobre o outro. O resultado é a sumarização visual dos valores das instâncias de dado, o que provê uma visualização agregada do conjunto de dados (HEER; KONG; AGRAWALA, 2009). Diferentes metáforas podem ser consideradas na abordagem empilhada: as barras são indicadas para representar múltiplas variáveis categóricas, enquanto as linhas, ou áreas, são mais apropriadas para as séries temporais. Assim, os *stacked graphs* se posicionam como alternativas interessantes para serem utilizados quando a informação da soma ou da agregação dos valores dos múltiplos atributos é tão ou mais importante quanto a própria individualização desses valores.

Wattenberg (2005a) desenvolveu *NameVoyager*, uma ferramenta que utiliza *stacked graphs* animados para explorar dados demográficos. Baseado em web, o sistema representa as tendências históricas em nomeação de bebês, com seu eixo horizontal definindo a variação temporal, enquanto o eixo vertical descreve a quantidade de crianças que receberam um certo nome no período correspondente (ELMQVIST *et al.*, 2011). A interface é interativa, reagindo diretamente a cada tecla inserida durante as pesquisas textuais. Segundo o autor, esta instantaneidade poupa o usuário do trabalho de clicar em um botão “submit”, além de auxiliar na demonstração de como os dados são minerados. As variáveis (nomes) do conjunto são representadas em faixas coloridas, organizadas alfabeticamente, com as pilhas de nomes podendo também ser selecionadas com o *mouse*, entre animações suaves e com alto grau de fluidez (ELMQVIST *et al.*, 2011). A visualização se tornou popular por apresentar características interativas que lembram um jogo (WATTENBERG, 2005a).

Havre, Hetzler e Nowell (2000) apresentaram o protótipo *ThemeRiver*, uma das primeiras aplicações voltadas especificamente para visualizar a evolução de séries temporais multidimensionais. Aprimorado em Havre *et al.* (2002), a estratégia consiste em empilhar uma série sobre a outra, construindo um *layout* de aparência fluida, mediante curvas com perfil suave. Entretanto, as camadas de informação não são empilhadas a partir do eixo horizontal, mas sim, de modo simétrico e centralizado no eixo temporal, diferindo a metáfora das abordagens clássicas dos *stacked graphs* (BYRON; WATTENBERG, 2008).

A Figura 3.7 apresenta o *ThemeRiver* aplicado sobre coleções de discursos pronunciados por Fidel Castro, entre 1959 e 1961, com destaque para alguns dos temas discutidos com maior frequência por Castro. Na parte superior da Figura 3.7, encontra-se uma sequência cronológica de eventos relacionados ao período histórico. Logo, a metáfora tem por finalidade representar o fluxo geral de tendências e mudanças temporais nas séries, atuando como uma “linha do tempo”, em vez de apenas realçar *outliers* ou comportamentos atípicos isolados (CHANG *et al.*, 2007).

Figura 3.7 – *ThemeRiver* e os temas abordados com maior frequência por Fidel Castro em seus discursos, de 1959 a 1961.

Fonte: Havre *et al.* (2002).

Posteriormente, o *ThemeRiver* foi refinado por Byron e Wattenberg (2008), sendo então renomeado para *stream graph* (gráfico de fluxo). A visualização ganhou notoriedade ao estampar uma das páginas do *The New York Times*, apresentando um contraste entre os rendimentos em bilheteria de cinemas dos filmes lançados em 2007, frente às indicações ao

Oscar daquele mesmo ano. No canto superior esquerdo da Figura 3.8, há um destaque para *La Vie En Rose*, filme premiado com o *Oscar* na categoria “melhor atriz”, que apresentou uma audiência modesta quando comparado a grandes sucessos, como *Shrek the Third*.

Com base nessa visualização, observa-se que o público e a academia do cinema nem sempre têm opiniões convergentes. Vale ressaltar que, neste exemplo, os autores rotacionaram a metáfora em 90 graus, com a orientação vertical definindo a escala temporal, ou seja, a longevidade em cartaz de cada filme, enquanto a largura de cada listra representa o rendimento em bilheteria. Byron e Wattenberg (2008) justificaram que esta escolha de *design* rotacionado foi tomada pra que o *stream graph* se encaixasse adequadamente no formato em colunas das páginas de jornal. Existe uma versão *online* da ferramenta em Bloch *et al.* (2008), esta com o eixo do tempo disposto na horizontal.

Figura 3.8 – Receitas, em bilheteria, dos filmes em cartaz nos cinemas norte-americanos ao longo de 2007.

Fonte: Byron e Wattenberg (2008).

Embora Byron e Wattenberg (2008) tenham realizado uma discussão matemática em seu trabalho, os algoritmos de posicionamento das camadas e de suavização das curvas permaneceram um tanto nebulosos ou não completamente desenvolvidos. Estas questões foram esclarecidas apenas recentemente por Bartolomeo e Hu (2016), que receberam, por suas contribuições, a premiação de *Best Paper* na *18th EG/VGTC Conference on*

Visualization (EuroVis 2016), um dos mais renomados eventos na área de visualização, demonstrando, com isso, o interesse da comunidade de pesquisas por esta metáfora.

No entanto, *stacked graphs* e suas variações apresentam limitações, como a capacidade de manusear um pequeno conjunto de séries temporais simultaneamente. Ainda que aparentemente efetivos em representar padrões agregados, são pouco hábeis para a comparação de séries individuais. O empilhamento visual não é informativo para alguns tipos de dado ou para valores negativos, por exemplo (HEER; KONG; AGRAWALA, 2009). Além disso, os analistas podem interpretar inadequadamente o espaço entre as curvas. Byron e Wattenberg (2008) sugeriram a ordenação das pilhas de informações para reduzir esse problema, mas isso ainda não o elimina (HEER; KONG; AGRAWALA, 2009).

Outra deficiência das metáforas empilhadas está na falta de mecanismos interativos. Assim, para tirar proveito de sua habilidade em transmitir as evoluções temporais, permitindo a comparação dinâmica entre diferentes grupos de informação, Shi *et al.* (2012) desenvolveram *RankExplorer*. Na interface, múltiplos itens são ranqueados e sumarizados em um *stacked graph*, que permite ao usuário comparar o fluxo das mudanças, ao longo do tempo, com o auxílio de marcadores que identificam trocas de posição; além de gráficos de barras, que exibem oscilações nas classificações; e linhas, para informações estatísticas.

Dou *et al.* (2013) e Cui *et al.* (2014) incorporaram o *stream graph* em sistemas para exploração de coleções textuais. A Figura 3.9 mostra a ferramenta *HierarchicalTopics*, desenvolvida por Dou *et al.* (2013), apresentando uma análise sobre conjuntos de notícias. Inicialmente, os textos são organizados em categorias principais, que podem se ramificar em tópicos, estabelecendo uma estrutura hierárquica em formato de árvore (BURCH; BECK; DIEHL, 2008), com os *stream graphs* (centro-direita) descrevendo a evolução dos temas selecionados pelo usuário, como se fossem espectros ou “linhas temporais” destes.

3.2.6 Grafos, Documentos Textuais e Tópicos

Documentos textuais podem ser examinados e correlacionados de acordo com a similaridade de seu conteúdo, citações ou ligações diretas. As páginas web são documentos textuais conectados por ligações chamadas *hyperlinks*. Um conjunto de páginas densamente relacionadas representa um tópico da web. Se novas páginas são criadas ou as já existentes são alteradas ao longo do tempo, essas conexões mudam.

Neste cenário, Toyoda e Kitsuregawa (2005) introduziram *WebRelievo*, um sistema para visualizar a evolução estrutural de páginas web. A interface representa as séries temporais por meio de grafos (HERMAN; MELANCON; MARSHALL, 2000), com os nós definindo as páginas e as arestas descrevendo os *hyperlinks*. Os usuários podem analisar

Figura 3.9 – *HierarchicalTopics* visualizando notícias textuais, com *stream graphs* apresentando as tendências evolutivas dos tópicos selecionados.

Fonte: Dou *et al.* (2013).

interativamente o espaço de informações de *sites* e seus relacionamentos, verificando como os tópicos são aglutinados ou divididos. O sistema exibe a variação temporal em gráficos separados, um pra cada intervalo de tempo, com a quantidade de gráficos sendo limitada pelo espaço disponível em tela. A ferramenta é um tanto confusa a primeira vista, além de sofrer com a oclusão ao explorar comunidades web muito densas. Para contornar essas questões, Itoh, Toyoda e Kitsuregawa (2010) apresentaram uma alternativa com propósito semelhante, agora baseada em ambiente tridimensional. A evolução dos elementos é obtida com a técnica de plano deslizante, que “fatia” o eixo do tempo e exibe o grafo web de acordo com a posição desse plano, algo que pode resultar em oclusões.

A exploração de textos originados em redes sociais apresenta desafios que combinam a detecção de relevância e o interesse sobre os temas. Considerando a dependência temporal das publicações, o nível de complexidade em sua análise aumenta consideravelmente. Alguns trabalhos se concentraram em desenvolver metáforas para visualizar o processo de distribuição espaço-temporal da informação (CAO *et al.*, 2012). Mas esses métodos não facilitam a visualização do relacionamento entre os múltiplos temas com a dinâmica de sua propagação. Sun *et al.* (2014) apresentaram uma abordagem inovadora, na qual os usuários podem visualizar a evolução de comportamentos, como o poder de cooperação e de competição entre tópicos ao longo do tempo. Os autores se basearam nos *stacked graphs*, aprimorando o conceito, pois a competição entre variáveis resulta da eventual adição de valores positivos com negativos, algo que não geraria resultados visuais lógicos em *stacked graphs* tradicionais (HEER; KONG; AGRAWALA, 2009; SUN *et al.*, 2014).

O *design* resultante, nomeado *EvoRiver*, é uma composição visual em duas partes: a camada superior, que contém as linhas de tópicos que representam os comportamentos

classificados como positivos; e a camada inferior, na qual estão os tópicos que apresentam sentimentos negativos. Quando, em algum momento, ocorrer uma mudança no poder de competição de um tópico (do negativo para o positivo, ou vice-versa), este alternará de camada. Instantes de tempo podem ser interativamente selecionados para investigar, com detalhes sob demanda, os relacionamentos entre pares de tópicos, que são indicados por arcos conectados (DANG; PENDAR; FORBES, 2016).

A Figura 3.10 ilustra o *EvoRiver* (Figura 3.10a) e os demais elementos que integram o sistema: um gráfico bivariado para apresentar as variações no poder de um tópico (Figura 3.10b); uma representação de arcos (Figura 3.10c); e uma nuvem que exibe as palavras-chave relativas ao tema selecionado, com cada palavra dimensionada de acordo com sua frequência (Figura 3.10d). Os pesquisadores utilizaram dados do *Twitter* para analisar a evolução dos temas discutidos durante a eleição presidencial de 2012, nos Estados Unidos. Os resultados indicaram que “Governo” e “Política” estão frequentemente vinculados. Porém, com a influência de “Questões Internacionais”, os temas “Governo” e “Política” alternam de competidores para colaboradores. Com isso, os analistas identificaram os complexos relacionamentos entre as questões mais importantes abordadas naquele processo político (CHEN; LIN; YUAN, 2017). No entanto, o sistema depende de um classificador de sentimentos baseado não apenas em análise estatística, mas também em sugestões de especialistas de domínio, o que pode introduzir um viés de subjetividade ao método.

Figura 3.10 – *EvoRiver* e a evolução dos temas discutidos no *Twitter* durante a eleição presidencial de 2012, nos Estados Unidos.

Fonte: Sun *et al.* (2014).

3.2.7 Séries Temporais Financeiras

Métodos de visualização de informação na área de finanças e negócios são motivados pelas grandes quantidades de dados numéricos produzidos pelas organizações, sejam elas

públicas ou privadas. Lei e Zhang (2010) projetaram um sistema para análise visual de ativos em bolsas de valores. As ações são agrupadas por meio de uma variação do algoritmo *k-means* e representadas em um *ringmap* (ZHAO; FORER; HARVEY, 2008). A abordagem proposta pelos autores pode ser vista como uma intersecção entre as técnicas *spiral graph* e *heatmap*, porém, construída com um número variável de anéis concêntricos, cada um dos quais apresentando múltiplos ativos dentro de um intervalo. A ferramenta conta com interatividade por *zooming* para análise exploratória dos detalhes.

Já em Fu *et al.* (2005), os pesquisadores utilizaram uma estrutura de árvore binária especializada, conhecida como *SB-Tree*, para armazenar as séries temporais contendo a evolução dos preços de ações. A alta volatilidade e a multidimensionalidade são características dos dados originados do mercado financeiro, logo, os sistemas neste contexto necessitam processar e incrementar periodicamente seu conteúdo, algo que pode ser computacionalmente custoso. Então, o sistema processa os pontos de dado e calcula as suas importâncias, com a *SB-Tree* sendo gerada pelos pontos classificados como perceptualmente importantes. A representação é feita de modo incremental, adaptada a atualizar com maior frequência as informações classificadas como mais importantes.

Chang *et al.* (2007) apresentam *WireVis*, uma abordagem multivisão para auxiliar analistas a explorar séries temporais contendo registros diários de transações de grandes bancos, rastreando as possíveis fraudes financeiras. O sistema apresenta o relacionamento entre contas, tempo e as palavras-chave dentro das transações, em quatro janelas sincronizadas e dedicadas a visualização de tarefas específicas (KO *et al.*, 2016). A ferramenta combina um *heatmap*, uma rede de avaliação de palavras, uma ferramenta de pesquisas por amostragem e um descritor de transações. Com isso, os analistas obtêm a visão geral de todas as transações, que podem ser manualmente individualizadas, para o exame mais refinado sobre detalhes específicos, atualizando imediatamente a visualização a cada interação. Testes com usuários demonstraram que a interface é eficiente em descobrir padrões suspeitos em grandes conjuntos de séries temporais (CHANG *et al.*, 2007). Entretanto, as múltiplas janelas introduzem complexidade ao sistema, aumentando a carga cognitiva e perceptual necessária em tarefas analíticas que exijam atenção a mais de uma *view* simultaneamente (GREEN; RIBARSKY; FISHER, 2009).

Construir sistemas para a análise exploratória de informações do mercado financeiro é algo complexo, devido ao grande volume, diversidade e dinâmica desse tipo de dado. Porém, a importância das aplicações financeiras tem estimulado esforços consideráveis entre os pesquisadores. Green, Ribarsky e Fisher (2009) realizaram um extenso estudo sobre os métodos de análise visual no domínio financeiro, categorizando as diversas ferramentas existentes, bem como suas metodologias e tarefas analíticas.

3.2.8 Sistemas Multi-foco

Figura 3.11 – *TimeNotes* apresentando seus vários níveis de *zooming*, visualizando informações de monitoramento animal em áudio.

Fonte: Walker, Borgo e Jones (2016).

Javed e Elmqvist (2010) introduziram uma técnica interativa para visualização de séries temporais por meio de empilhamento de *zooming*. O método subdivide o espaço de modo hierárquico, permitindo o foco em múltiplas regiões da série temporal, em diferentes níveis de detalhes (WANG *et al.*, 2018). O usuário define uma área de seleção de dados no gráfico e, então, é criada uma nova região com as informações ampliadas. Esta região é empilhada logo abaixo do gráfico de origem, que é redimensionado para acomodá-la. Embora não seja propriamente uma técnica de interação multi-foco, o *stack zooming* possibilita a construção de *displays* com níveis reduzidos de distorção, mantendo a representação original da informação, como sua visão geral (JAVED; ELMQVIST, 2010). Porém, a técnica foi projetada para ser utilizada em espaços unidimensionais, sendo possível, mas complexa, a sua adaptação para 2D (JAVED; GHANI; ELMQVIST, 2012).

O *zoom* hierárquico contínuo desempenha um avanço significativo sobre a navegação convencional passo a passo. Sem o *zoom* contínuo, as informações de interesse devem ser selecionadas em cada nível, até que o trecho desejado seja delimitado, usualmente exigindo várias interações (ANDREWS *et al.*, 2002).

Neste contexto e inspirados pela abordagem de *stack zooming*, Walker, Borgo e Jones (2016) desenvolveram *TimeNotes*. O sistema suporta a navegação hierárquica, explorando interativamente múltiplos segmentos de séries temporais, que são selecionados para análise detalhada. Estes podem ser comparados com outros trechos já destacados,

permitindo ao analista agrupar e rotular as amostras contendo tendências ou comportamentos similares (ANDRIENKO *et al.*, 2018). Segundo os autores, o método de navegação é eficiente, pois permite ao usuário dividir o espaço de informações e construir a interface apenas com o dado considerado relevante, na granularidade requerida, agindo como um histórico gráfico de suas tarefas analíticas. A Figura 3.11 apresenta o sistema *TimeNotes* de Walker, Borgo e Jones (2016), visualizando um áudio de monitoramento animal com 15 minutos, obtido por sensores implantados em uma ave.

Ainda seguindo a temática de visualização de dados temporais em métodos baseados em organização hierárquica, Gomi e Itoh (2010) criaram *MIAOW*, um navegador para imagens inspirado na metáfora gráfica de preenchimento espacial *Treemap* (JOHNSON; SHNEIDERMAN, 1991). Trata-se de um visualizador tridimensional para análise de fotografias, que classifica e agrupa as imagens hierarquicamente de acordo com a localização geográfica (latitudes e longitudes) e o instante em que foram capturadas. Por meio do modelo de cubo espaço-temporal, os grupos de fotos são distribuídos no plano horizontal em regiões retangulares aninhadas, com a variação temporal sendo definida pelo eixo vertical. O sistema apresenta muitas ferramentas interativas, mas também algumas limitações. Não são todos os equipamentos fotográficos que têm a capacidade de georreferenciar as fotos que registram, reduzindo a utilidade do navegador, além do fato que agrupamentos com pequenas quantidades de imagens podem não ser adequadamente representados.

3.2.9 Monitoramento da Qualidade do Ar

Li *et al.* (2013) desenvolveram ferramentas interativas para a visualização de séries temporais, obtidas ao longo de 4 anos, a partir da amostragem diária da concentração dos poluentes atmosféricos de algumas cidades chinesas. Os autores basearam seu sistema em uma estrutura de árvore do tipo *node-link* (BURCH; WEISKOPF, 2011) para organizar as cidades, que podem então ser subdivididas de acordo com as suas respectivas áreas de monitoramento. Com isso, o usuário pode verificar e comparar a situação do ar nas diferentes estações de coleta de dados, ao longo do tempo. A Figura 3.12 ilustra o sistema, apresentando a condição do ar na cidade de Jinan, província de Shandong, China. Embora disponha de boas contribuições na disposição dos elementos visuais, a técnica sofre com o limitado espaço em tela, não sendo escalável o suficiente para representar estruturas complexas e grandes conjuntos de séries temporais.

Para visualizar os dados de mais localidades em simultâneo, Li *et al.* (2014) aprimoraram a ferramenta anterior. As cidades monitoradas são dispostas em um painel estruturado em árvore hiperbólica (LAMPING; RAO; PIROLI, 1995), que pode ser expandido em módulos com discos temporais, para a verificação da qualidade do ar entre

Figura 3.12 – Níveis de poluição do ar coletados diariamente em regiões de Jinan, na China.

Fonte: Li *et al.* (2013).

os pontos de coleta da cidade selecionada. Ambos os trabalhos hierarquizaram as séries temporais, comparando-as por meio de sua subdivisão em camadas. Então, os autores aplicaram as visualizações sobre informações referentes ao monitoramento do ar para aprimorar a percepção do comportamento das variáveis e discriminar as fontes de poluentes (RODRIGUEZ; MENDEZ, 2015). Apesar de sumariarem grandes quantidades de dados ambientais, as interfaces facilitam a comparação de regiões dentro da mesma cidade, mas não viabilizam o cruzamento de informações entre duas, ou mais, cidades.

Yoon, Xu e Tourassi (2016) realizaram uma análise sobre séries temporais que registram os valores dos poluentes do ar, coletados entre os anos de 1998 e 2007 pela agência ambiental dos Estados Unidos (EPA, 2015). Os pesquisadores se concentraram em correlacionar a ocorrência de casos de câncer de pulmão (NCI, 2015), em determinadas regiões do território norte-americano, com a concentração do material particulado destas localidades. Ou seja, o objetivo do trabalho foi investigar se existe alguma relação entre a incidência regional de câncer de pulmão, com o histórico de exposição aos poluentes. Apesar de ressaltarem que o material particulado não é o único fator de risco, os autores confirmaram com seus resultados que existem maiores chances de incidência de câncer em localidades expostas a altos índices de poluição no ar.

Embora tenha relações com os estudos de Yoon, Xu e Tourassi (2016), quanto à análise de séries temporais que descrevem os níveis de poluentes no ar, esta dissertação de Mestrado difere sensivelmente quanto às abordagens visuais. Aqui, o objetivo não é apenas

analítico, mas sim o de construir ferramentas que viabilizem a exploração interativa dos dados. Como demonstrado, visualizar múltiplas séries temporais em um espaço reduzido tem guiado os pesquisadores a desenvolver ferramentas e técnicas de representação que permitem analisar os relacionamentos entre dados de diferentes categorias (KRSTAJIC; BERTINI; KEIM, 2011). Infelizmente, a maioria das técnicas não é capaz de mostrar uma visão geral e intuitiva das múltiplas séries (CHO *et al.*, 2014), sendo a questão da escalabilidade ainda um problema central. As ferramentas de visualização projetadas aqui reúnem alguns dos pontos fortes dos diversos sistemas descritos na literatura, buscando superar as limitações inerentes à exploração de séries temporais multidimensionais.

3.3 Considerações Finais

Neste Capítulo foram apontados alguns dos trabalhos do estado da arte em visualização de informações, especialmente as metáforas dedicadas às séries temporais multidimensionais. Este estudo mostrou que a representação dos dados variantes no tempo é um campo potencialmente fértil, com as mais variadas estratégias sendo empregadas, de acordo com o contexto e o comportamento de interdependência dos dados em análise. As ferramentas mais modernas e abrangentes oferecem mecanismos de interação com o usuário, visando reduzir, ao menos parcialmente, o problema da dimensionalidade. Porém, os trabalhos apresentam limitações. A maior parte das técnicas tem por objetivo visualizar os caminhos entre os dados, mas são incapazes de rastrear as informações ou conduzir uma análise sistemática de detecção de padrões (SHI *et al.*, 2014).

Quanto maior a dimensionalidade dos dados, maiores os desafios em construir metáforas capazes de sumarizar a evolução das múltiplas variáveis em um cenário compacto e intuitivo, mesmo com a composição de diferentes elementos gráficos e técnicas exploratórias. Assim, esta pesquisa difere dos trabalhos revisados, em vários aspectos. Aqui, serão construídas ferramentas analíticas de visualização com o objetivo de facilitar a exploração interativa de séries temporais multidimensionais, contendo informações sobre a qualidade do ar na cidade de São Carlos–SP, por usuários comuns e especialistas de domínio, que utilizarão uma ampla variedade de dispositivos como meios de acesso.

Vale ressaltar que, neste Capítulo, foram discutidos os trabalhos em visualização de informações e analítica visual, considerados correlatos ao tema desta dissertação ou que contribuam com o esclarecimento de alguma ferramenta desenvolvida. A comunidade de pesquisas em visualização é altamente produtiva. Todos os anos são publicados novos trabalhos em diversos periódicos, congressos e eventos especializados nos mais variados domínios de aplicação, contendo métodos inovadores e cada vez mais eficientes em preencher

as lacunas ainda existentes. Mesmo com os consideráveis esforços empreendidos, seria inviável proceder aqui uma análise sobre todas as técnicas relevantes em visualização de dados temporais multidimensionais. Porém, semelhante ao realizado no Capítulo anterior, sempre que possível houve a indicação de *surveys* e *reviews* para que o leitor interessado em algum domínio específico possa consultar estudos já realizados na literatura, sobre as ferramentas para análise visual de informações relacionadas àquela área de aplicação.

Capítulo 4

Desenvolvimento da Proposta

4.1 Considerações Iniciais

As técnicas de visualização tradicionais apresentam limitações quando aplicadas a conjuntos de séries temporais. Essas facilitam apenas a inspeção de uma ou de algumas variáveis em um instante de tempo, característica que dificulta ou mesmo impede o seu uso em séries multidimensionais (MAATEN, 2014). As soluções de visualização modernas devem viabilizar uma análise exploratória mais abrangente, com funcionalidades e recursos que tornem os dados mais compreensíveis às mentes humanas e guiem o analista na descoberta de informações não óbvias e potencialmente importantes.

Neste contexto, há o desafio de se construir novas ferramentas para visualização de séries temporais, que sejam flexíveis o suficiente para suportar tarefas interativas, enquanto promovem uma exploração que resulte em ganho no poder analítico sobre a dinâmica temporal dos dados (ZHAO; CHEVALIER; BALAKRISHNAN, 2011). A metodologia desenvolvida e apresentada neste Capítulo visa preencher essas lacunas por meio da arquitetura de sistemas gráficos dedicados a visualização e exploração interativa de padrões e tendências em conjuntos multidimensionais variantes no tempo.

Na Seção 4.2 são descritas as características dos conjuntos de séries temporais que serão analisadas, bem como as estratégias de pré-processamento adotadas para “limpar” e padronizar seus valores, formatando-os para serem adequadamente trabalhados. Já na Seção 4.3 foram definidas algumas tarefas analíticas relacionadas ao entendimento da evolução temporal desses dados. Nas Seções 4.4 e 4.5, são abordados os detalhes de projeto das ferramentas de visualização construídas: duas interfaces gráficas interativas que permitem ao analista conduzir uma exploração de dados sob demanda, guiando-o na descoberta de informações de seu interesse. O sistema descrito na Seção 4.4 apresenta

amostras do ar de São Carlos–SP, ao passo que a ferramenta da Seção 4.5 trata índices de doenças do aparelho respiratório humano, também registrados em São Carlos–SP. Por fim, na Seção 4.6, estão algumas considerações finais deste Capítulo.

4.2 Conjuntos de Dados

De modo a realizar as análises apresentadas neste Capítulo, cinco conjuntos diferentes de séries temporais foram selecionados, cada um apresentando características específicas. Os dados foram disponibilizados pelo grupo de estudos em Controle Ambiental e Monitoramento da Qualidade do Ar em Ambientes Externos, do Departamento de Engenharia Química da Universidade Federal de São Carlos (UFSCar). Este trabalho foi desenvolvido em colaboração com os pesquisadores do referido grupo.

Os conjuntos de séries temporais se dividem em dois grupos: o primeiro contém informações ambientais obtidas por amostragem de poluentes oriundos do ar atmosférico; e o segundo grupo de dados apresenta índices hospitalares de doenças relacionadas ao aparelho respiratório. Todos os dados foram coletados na cidade de São Carlos–SP.

4.2.1 Monitoramento da Qualidade do Ar

São Carlos está localizada próxima ao centro geográfico do Estado de São Paulo, possuindo uma população de aproximadamente 246 mil habitantes, distribuídos por uma área de 1.137 km². A temperatura média anual é de 19,6 graus Célsius, com a umidade relativa do ar média oscilando entre 76% no verão e 54% no inverno (IBGE, 2017). De acordo com Marques *et al.* (2001), a Praça dos Voluntários, no centro da cidade, é o ponto adequado para o monitoramento do material particulado atmosférico (MP), pois concentra a circulação diária de um elevado número de pessoas, serviços e veículos.

As coletas do MP foram realizadas individualmente por um período de 23 horas e 30 minutos cada, utilizando um amostrador de grandes volumes para partículas com diâmetro aerodinâmico de até 10 µm (AGV-MP₁₀), ilustrado na Figura 4.1. O equipamento conta com uma bomba de sucção à vazão constante de 1,13 m³ min⁻¹ (em condições reais de temperatura e pressão), além de um cabeçote com separador inercial de um estágio, que retém o material com diâmetro superior a 10 µm. Com este corte inicial, o ar prossegue por bocais até encontrar o elemento filtrante posicionado na base do cabeçote, onde o MP com diâmetro menor ou igual a 10 µm (MP₁₀) é coletado (ENERGETICA, 2016). As características físico-químicas presentes no ar foram aferidas posteriormente, mediante o estudo laboratorial do elemento filtrante empregado em cada evento de coleta.

Diversos cálculos e procedimentos específicos de análise química foram utilizados para a obtenção dessas informações, porém os métodos empregados na aferição dos dados estão além do escopo desta dissertação, podendo ser consultados em pesquisas dedicadas ao tema (MARQUES *et al.*, 2001; ALEXANDRINA, 2015).

Figura 4.1 – Representação esquemática de um AGV-MP₁₀.

Fonte: Marques *et al.* (2001).

Seguindo os procedimentos de coleta descritos, foram registradas a data e a concentração em massa do material particulado coletado do ar atmosférico local, sendo a unidade de massa expressa em $\mu\text{g m}^{-3}$. Além disso, também foram registradas as variáveis meteorológicas do ponto de amostragem, como a precipitação, em milímetros; a velocidade do vento, em metros por segundo; a direção do vento, em graus; a temperatura média, em graus Celsius; e a taxa percentual de umidade relativa do ar; para cada evento diário de coleta (INMET, 2014). Então, foram delimitados três conjuntos de séries temporais ambientais.

O primeiro conjunto é formado por amostragens realizadas em dias alternados, entre maio de 2015 e novembro de 2017. O padrão de aquisição de dados em dias alternados, com ciclo de 23 horas e 30 minutos por elemento filtrante, segue os protocolos da *International Organization for Standardization* (ISO 13271, 2012), que definem as metodologias de aferição da concentração em massa do material particulado em meio fluido gasoso, sendo este o padrão atualmente adotado pelo grupo de pesquisas responsável pelo projeto de monitoramento do ar em São Carlos-SP. O segundo conjunto compreende as informações coletadas diariamente entre abril de 2014 e abril de 2015. Também foi utilizado um terceiro conjunto de séries temporais, contendo os dados aferidos entre setembro de 1997

e fevereiro de 2006. Vale destacar que este terceiro conjunto foi aferido em trabalhos de monitoramento do passado, quando ainda não havia um padrão bem estabelecido para os ciclos de coleta (se diários ou em dias alternados), com as amostragens ocorrendo com uma frequência média de duas vezes por semana.

A separação desses dados em três bases distintas, originadas de duas campanhas de coleta recentes e uma campanha com registros históricos datados de cerca de 20 anos atrás (considerando a data de escrita deste texto), foi motivada por diversas questões, entre elas, a grande mudança nas legislações ambientais ao longo dos últimos anos, atualmente mais rígidas. Além das alterações nos padrões de aquisição dos dados, antes diário, atualmente em dias alternados. Assim, pretende-se verificar os reflexos que tais mudanças tiveram sobre as concentrações do MP_{10} . A Tabela 4.1 sumariza as principais características dos conjuntos de séries temporais ambientais.

Tabela 4.1 – Características básicas dos conjuntos de dados ambientais.

Período de amostragem	Padrão de coletas	N.º de instâncias
Maio-2015 a novembro-2017	Dias alternados	276
Abril-2014 a abril-2015	Diariamente	309
Setembro-1997 a fevereiro-2006	Duas vezes por semana	683

Fonte: Dados da pesquisa.

Inicialmente os dados foram organizados em planilhas eletrônicas, com cada coluna armazenando as informações de uma das variáveis ambientais. Por sua vez, cada uma das linhas da planilha registra uma instância de dado, ou seja, o conjunto das variáveis ambientais aferidas em um dia de amostragem. Essas planilhas foram posteriormente convertidas em tabelas no formato TSV, que é um formato de arquivo de texto plano, em que os campos (variáveis) são delimitados por tabulação. Em seguida, os dados foram tratados e tiveram seu separador decimal convertidos para o ponto, que é o padrão norte-americano adotado pela maioria dos métodos computacionais de *data handler*, como padrão de leitura e escrita de valores numéricos. Isso foi necessário porque o padrão de separador decimal utilizado nas planilhas, em português do Brasil, foi a vírgula.

Durante os ciclos de coleta ocorreram algumas falhas, para quaisquer um dos três conjuntos de dados do monitoramento ambiental. Essas falhas estão relacionadas com problemas no equipamento de filtragem, que acarretaram em uma interrupção durante o ciclo de coleta, como quedas de energia ou ainda alguma falha de natureza mecânica, por exemplo. Havendo uma falha, a amostragem em curso foi considerada corrompida e descartada. Para esses eventos de falha, os dados meteorológicos registrados também foram descartados, ou seja, toda a instância referente a um ciclo de amostragem corrompido foi descartada. Essa decisão de pré-processamento mantém a consistência e a completude dos

dados e, inicialmente, não causa prejuízos aos resultados, visto que as falhas do equipamento foram ocorrências pontuais que representaram menos de 0,5% do total amostrado. Por isso, essas instâncias corrompidas podem ser consideradas numericamente insignificantes.

4.2.2 Doenças do Aparelho Respiratório

Os outros dois conjuntos de séries temporais foram obtidos junto ao Sistema de Informações Hospitalares do SUS (Sistema Único de Saúde), que é mantido pelo Ministério da Saúde. O SUS realiza a coleta de dezenas de variáveis relacionadas às internações hospitalares registradas em instituições públicas e privadas, em todo o território brasileiro, gerando, com isso, um extenso banco de dados que fornece um panorama da saúde no Brasil. Esses dados são compilados e disponibilizados de acordo com o mês e o ano de ocorrência, podendo ser acessados em <http://www2.datasus.gov.br/DATASUS/>.

O sistema de gerenciamento de dados do SUS permite ao usuário realizar pesquisas variadas em seus registros. Por meio dessas pesquisas, é possível solicitar o *download* de tabelas contendo dados pré formatados. Então, foram selecionados os registros de São Carlos-SP, contendo os casos de doenças relacionadas ao aparelho respiratório, que são as enfermidades comumente associadas à qualidade do ar. As tabelas fornecidas pelo SUS são abrangentes, com atributos categorizando as faixas etárias dos pacientes, o tipo de doença diagnosticada, o sexo, o caráter de urgência do atendimento, os valores das despesas médicas, além de outras variáveis.

Vale destacar que o SUS fornece estatísticas de outras ocorrências médicas, dados estes que podem ser potencialmente valiosos para diversas aplicações de visualização. Como mencionado, para este trabalho foram delimitadas apenas as doenças relacionadas ao aparelho respiratório e, entre as principais, estão a laringite, pneumonia, bronquite, amigdalite e asma. Existem outras doenças ligadas ao sistema respiratório, no entanto, duas situações foram observadas nos dados: a primeira tem a ver com a esparsidade de ocorrência de algumas doenças, que não apresentaram registros durante longos intervalos, com poucos eventos registrados, normalmente nos meses de inverno; a segunda situação está ligada aos registros em si, que não são precisos em categorizar certas doenças, classificando-as utilizando termos genéricos como “outras doenças do aparelho respiratório”.

A decisão de pré-processamento tomada para tratar esses registros pouco claros foi criar uma categoria “outras doenças”, que agrega os números das ocorrências esparsas e mal categorizadas. Essa decisão organiza os dados e evita confusões para o usuário do sistema de visualização, uma vez que reduz a quantidade de variáveis que poderiam poluir o *layout*. Assim, foram selecionadas para análise visual as seguintes variáveis: “laringite”,

“pneumonia”, “bronquite”, “amigdalite”, “asma” e “outras doenças”. A evolução temporal de cada uma dessas enfermidades foi agrupada de acordo com a faixa etária dos pacientes, dentro de 8 intervalos de idade: menores de 5 anos, 5 a 9 anos, 10 a 19 anos, 20 a 29 anos, 30 a 39 anos, 40 a 49 anos, 50 a 59 anos e 60 anos ou mais.

Para contrastar com as informações do monitoramento ambiental, os dados obtidos junto ao SUS compreendem os anos de 1998 a 2006 e os anos de 2014 a 2017, sendo que o conteúdo de 2014 inicia no mês de abril (cf. Subseção 4.2.1). Os arquivos referentes aos anos de 1996 e 1997 não foram encontrados no sistema do SUS. Também não estão disponíveis os registros para anos anteriores a esses. Note que aqui não houve uma separação de conteúdo em três conjuntos, como decidido no cenário para os dados ambientais. Naquela situação, ocorreu uma mudança metodológica no padrão de amostragem do ar durante o período de 2014 a 2017, algo que motivou a separação das amostras nos conjuntos para 2014–2015 e 2015–2017. Já no caso dos dados de saúde, essa separação não faria sentido, uma vez que não há informações sobre mudanças nos métodos de registro do SUS.

Os demais passos de pré-processamento seguem o descrito em 4.2.1, ou seja, os dados foram convertidos para tabelas no formato TSV e a vírgula do separador decimal foi substituída pelo ponto, quando necessário. Por fim, foi adicionada uma coluna extra contendo os valores médios mensais da concentração do material particulado em ambos os conjuntos de séries temporais dos índices de saúde. Como o SUS fornece suas tabelas com números agregados mensalmente, o MP_{10} também teve seus níveis agregados seguindo o padrão mensal para corresponder com os intervalos de registro das enfermidades.

4.3 Tarefas Analíticas

Após adequar e estruturar os dados, foram realizadas entrevistas com os pesquisadores especialistas no domínio para identificar os principais desafios que estes enfrentavam ao analisar os dados. Ferramentas com a capacidade de simplificar a representação dos dados coletados e que viabilizassem a identificação, seleção e a comparação de informações entre diferentes períodos, tendências e sazonalidades, foram os principais requisitos dos especialistas. Existia a necessidade por recursos que agilizassem o processo de investigação da dinâmica de dispersão dos poluentes sobre a cidade, correlacionando essas informações com a evolução temporal das demais variáveis registradas. A partir das demandas dos especialistas, foi definida a seguinte lista de tarefas analíticas:

- **Seleções interativas (T1):** Como selecionar interativamente subconjuntos de pontos de interesse? É possível construir métodos de seleção de dados que permitam verificar de modo interativo, subconjuntos pertencentes a períodos distintos?

- **Padrões sazonais (T2):** Quais são os padrões (se existirem padrões) sazonais de distribuição de poluentes? Como os padrões mudam ao longo dos meses e quais as possíveis influências das estações do ano?
- **Tendência de concentração dos poluentes (T3):** É possível verificar alguma tendência de estabilidade, aumento ou diminuição nas concentrações do MP₁₀ coletado em São Carlos-SP, com o passar dos anos? Mais especificamente, as recentes alterações no controle ambiental tiveram algum efeito sobre a qualidade do ar?
- **Dinâmica temporal das doenças respiratórias (T4):** Quem são as pessoas mais afetadas pelas doenças relacionadas ao aparelho respiratório? Como os índices dessas doenças têm mudado ao longo do tempo?
- **Padrões climáticos e de doenças (T5):** A incidência de doenças respiratórias acompanha as tendências de concentração dos poluentes? É possível verificar alguma correlação entre essas informações?

Métodos de análise de propósito geral, aliados a técnicas de visualização tradicionais e estáticas, acabam por limitar a capacidade exploratória durante a verificação dos dados, chegando mesmo a, eventualmente, reduzir a sua significância. Por isso, foi definido como primeiro passo de desenvolvimento, oferecer aos usuários recursos visual-analíticos mais robustos, como os mecanismos interativos de seleção que permitem não apenas selecionar dados de interesse, mas também interagir e comparar os diferentes aspectos das informações selecionadas, por meio de visões interconectadas. Esse é o objetivo de T1.

As tarefas T2 e T3 estão relacionadas com o problema do entendimento da evolução temporal dos poluentes dispersos no ar da cidade. Já as tarefas T4 e T5 estão associadas com o problema de analisar dados discretizados em diferentes categorias, contrastando suas tendências globais e também locais. É importante viabilizar o aprofundamento dos níveis de sumarização das informações, partindo de visões gerais dos conjuntos de séries temporais para análises individualizadas de instâncias de dado.

Detectar as semelhanças de informações pertencentes a diferentes escalas também é essencial para compreender o modo como os padrões variam ao longo do tempo. Por exemplo, o padrão de concentração dos poluentes pode variar no decorrer dos anos ou dos meses de um ano, sob a influência de fatores como os meteorológicos. Esse fato está relacionado com T3 e, para ser respondido, demanda recursos visuais que permitam seleções para explorar e comparar os múltiplos aspectos dos diferentes períodos de amostragem.

Entre os objetivos de construção dos sistemas propostos aqui, estão a integração de métodos de seleção com ferramentas gráficas que possibilitem ao usuário pesquisar visualmente nos conjuntos de dados mediante seleções interativas, algo que vem para

simplificar o ambiente das análises exploratórias. Além do mais, nem todos os especialistas em domínio são treinados nas ciências de computação, logo, essas ferramentas devem ser simples e intuitivas. Com uma arquitetura que equilibre simplicidade e expressividade, os sistemas de visualização podem suportar análises com menor exigência de treinamento e, por consequência, mais eficiência, em diferentes escalas e níveis de abstração, enquanto facilitam a descoberta de tendências temporais nem sempre óbvias.

4.4 O Sistema de Análise Visual dos Dados Ambientais

Um sistema eficiente de visualização de informações deve ser construído visando minimizar o esforço cognitivo envolvido em analisar os dados (GOETHEM *et al.*, 2017). Para dar suporte à exploração dos dados amostrados, foi construída uma interface gráfica baseada em web, que está disponível *online* em <https://evortigosa.github.io/pollution/>, com livre acesso. A interface foi elaborada utilizando principalmente a linguagem de programação *JavaScript*, com *D3.js*. A biblioteca *D3* simplifica a manipulação de tabelas ao permitir um grande controle sobre a construção de elementos gráficos interativos e de alta qualidade em *SVG*, o que faz da *D3* uma escolha superior quando comparada a alternativas como o *HTML5 Canvas*, por exemplo. A vantagem de se utilizar *JavaScript* e *D3* está no fato de serem portáteis por praticamente todos os navegadores web modernos. Isto é, não é necessário instalar o “software *D3*” para que usuários, em qualquer parte do mundo, operando os mais variados tipos de dispositivos, tenham acesso ao sistema e consigam fazer pesquisas nos dados disponibilizados. Basta ter conexão com a Internet.

Embora o espaço disponível ainda seja uma questão central quando se trata de sistemas visuais, com os recentes avanços tecnológicos, as telas dos dispositivos eletrônicos têm apresentado qualidade gráfica cada vez maior. Logo, já não é incomum encontrar monitores com grandes densidades de pixels e com capacidade de reproduzir imagens de alta definição. Neste sentido, a interface foi projetada para oferecer melhor experiência de navegação em *displays* com resolução de 1920×1080 , *Full HD*, ou superior. Mas também é possível navegar entre as funcionalidades do sistema a partir de dispositivos com resoluções inferiores, apenas ajustando as configurações de *zoom* do *browser*. Em computadores, esse ajuste é usualmente feito por meio dos comandos *Ctrl+* e *Ctrl-*. Já em *smartphones* e *tablets*, devem ser aplicados movimentos com os dedos, como ilustrado na Figura 2.8.

4.4.1 Interface Gráfica e seus Componentes

A metáfora gráfica escolhida para exibir os índices de concentração do material particulado foi o *scatter plot* (BECKER; CLEVELAND, 1987). Basicamente, um *scatter*

plot mapeia no plano cartesiano os valores de uma variável utilizando símbolos. Por exemplo, seja $S \in \mathbb{R}^2$ um conjunto de séries temporais cujos elementos s registram o valor em massa da concentração do MP_{10} , coletados ao longo do intervalo de tempo t . Para cada instante de tempo i , a posição relativa ao eixo das abscissas do i -ésimo símbolo é horizontalmente determinada por meio do valor de t_i em que s_i foi registrado. Ao passo que a posição relativa ao eixo das ordenadas desse i -ésimo símbolo é verticalmente determinada pelo valor de s_i , apurado em t_i . Variações do *scatter plot* básico, com diferentes símbolos, de diferentes tamanhos, formatos e cores, aprimoram a capacidade de interpretação da representação gráfica.

Outras metáforas podem ser consideradas, porém *scatter plots* têm como vantagem o fato de consumirem apenas alguns pixels para representar pontos únicos de dado, destacando-se como gráficos espaço-eficientes. Além disso, são metáforas bem estabelecidas e comumente utilizadas no universo de análises químicas. Segundo McLachlan *et al.* (2008), para se obter um rápido aprendizado sobre as ferramentas de um sistema visual, deve-se explorar a experiência e a intuição do usuário. Desse modo, as representações devem ser familiares quando for apropriado. Entretanto, *scatter plots* sofrem com a escalabilidade visual. A oclusão acontece à medida que a quantidade de pontos se torna maior do que a resolução da tela (JUGEL *et al.*, 2014), já quando existem pontos muito próximos uns dos outros, podem ocorrer sobreposições que obscurecem a legibilidade individual dos símbolos. Assim, é necessário oferecer alguns mecanismos para mitigar essas desvantagens. A Figura 4.2 ilustra a visão inicial da interface.

Na parte superior (Figura 4.2a-b), encontra-se a representação do conjunto de séries temporais registrado em dias alternados, entre maio de 2015 e novembro de 2017. Este módulo se subdivide em dois grupos. À esquerda (Figura 4.2a) se posiciona o *scatter plot* que exibe a evolução temporal da concentração do material particulado no ar. A magnitude do MP_{10} , para cada dia de amostragem, é definida por um ponto disposto verticalmente de acordo com a escala em $\mu\text{g m}^{-3}$. Dentro do *scatter plot* há uma linha horizontal denotando os níveis aceitáveis de poluentes para o período amostral de 24 horas. Essa linha referencial reflete as normas ambientais em vigor durante as campanhas de monitoramento do MP_{10} , normas estas que serão abordadas a seguir, na Subseção 5.2.2. Logo à direita (Figura 4.2b), verificam-se quatro gráficos de área, agrupados verticalmente, que exibem a evolução temporal da precipitação, temperatura, velocidade do vento e umidade relativa do ar. Essa configuração segue o conceito de *small multiples* (ZIEGLER *et al.*, 2010), que é um conjunto de gráficos auxiliares a uma visualização principal, similares entre si, com a finalidade de acompanhar e comparar as diferentes variáveis do conjunto de dados, em porções de espaço reduzidos.

Figura 4.2 – Visão inicial da interface com os dados de maio de 2015 até novembro de 2017 (superior) e os dados de abril de 2014 até abril de 2015 (inferior).

Fonte: Elaborada pelo autor.

Os grupos de *views* foram distribuídos espacialmente desse modo para permitir que o usuário analise a evolução da concentração do MP₁₀, em um determinado período, enquanto verifica o comportamento de quatro variáveis meteorológicas registradas ao longo do mesmo período. Vale observar que as variáveis meteorológicas possuem diferentes escalas e unidades de medida. Entretanto, a razão de aspecto dos gráficos auxiliares é homogênea, ou seja, todos têm as mesmas dimensões.

Analogamente, na parte inferior da interface (Figura 4.2e-f) é encontrado o segundo módulo de *views*, com as informações coletadas diariamente, para o período de abril de 2014 até abril de 2015. Essa composição de *layout* segue a configuração classificada por Javed e Elmqvist (2012) como justaposição. Em *layouts* justapostos, as visualizações são posicionadas lado a lado (ou uma sobre a outra) na interface, permanecendo totalmente visíveis aos usuários. Devido ao formato, em grande parte, retangular, dos monitores de computador, com o lado horizontal normalmente mais extenso, foi decidido adotar o estilo dos grupos justapostos em grade, empilhados na vertical. O benefício da organização de *layout* em grade está no resultado compacto da visualização, com melhor aproveitamento do espaço disponível (SADANA; STASKO, 2016).

Os pontos são caracterizados em ambos os *scatter plots* de acordo com suas respectivas estações, que podem ser “verão”, “outono”, “inverno” e “primavera”. Essa informação é refletida por meio de cores, que são decodificadas pela legenda posicionada na parte

inferior esquerda da interface (Figura 4.2g). Todas as cores utilizadas no sistema foram selecionadas tendo por base o mapa divergente sugerido pelo *ColorBrewer* (HARROWER; BREWER, 2003), um guia bem conhecido que auxilia seus usuários na escolha apropriada de esquemas de cores para sistemas visuais, em conformidade com propriedades de percepção gráfica.

O esquema de cores escolhido é particularmente adequado (*friendly*) para exibição em telas de dispositivos eletrônicos (LCD), impressão e fotocópia, considerando inclusive os eventuais leitores portadores de daltonismo. A opacidade dos círculos não é total e foi configurada assim para permitir que o usuário consiga identificar aqueles pontos que sofrem com algum grau de oclusão. A cor de fundo de toda a interface é o branco, enquanto os símbolos gráficos apresentam cores vívidas. Segundo Ware (2004), áreas extensas devem apresentar cores de baixa saturação, sendo que o alto contraste entre as cores de fundo e os objetos do primeiro plano aprimora a legibilidade das informações.

Além disso, para reforçar a transmissão visual da magnitude da concentração do MP_{10} , cada um dos círculos do *scatter plot* teve seu diâmetro calculado em escala. Todos os conjuntos de dados são mapeados durante a inicialização da interface, com os valores mínimo e máximo de concentração do MP_{10} sendo armazenados. Foi criado um método para escalar o diâmetro de cada círculo, dimensionando o tamanho do círculo a ser desenhado de acordo com os valores limite de concentração do MP_{10} . O método recebe os valores de mínimo e máximo para definir seu domínio, e então o raio de escala estará dentro do intervalo de imagem do método entre 3 e 16 pixels (definidos empiricamente para garantir o bom aspecto geral da interface). Logo, quanto maior a concentração de material registrada em um ponto, maior será o diâmetro de seu respectivo círculo no *scatter plot*. Essa escolha segue um importante princípio de *design*: elementos maiores são perceptualmente interpretados como mais significantes, estratégia também observada na icônica visualização da IPO do *Facebook* (NYTIMES, 2012).

No centro da interface (Figura 4.2c) está posicionado o gráfico bivariado. Trata-se de um gráfico de linha dupla, uma linha para cada variável, em que a área entre as duas linhas é preenchida com a cor da linha que representa a variável de maior valor. Essa metáfora, que será descrita detalhadamente na Subseção 4.4.3, foi desenvolvida com o objetivo de esboçar a diferença quantitativa entre as concentrações do MP_{10} nos módulos superior e inferior. O gráfico bivariado possui a mesma extensão horizontal dos *scatter plots* e serve como um referencial. A predominância da cor que representa uma das variáveis indica que esta variável apresenta maior quantidade de valores altos, no intervalo.

Logo à direita do gráfico bivariado (Figura 4.2d), posiciona-se um *combo box* no qual é possível alterar o conjunto de dados em análise no módulo inferior da interface. Ao clicar

na ferramenta, será exibida a opção de alternar para o conjunto registrado entre os anos de 1997 e 2006, além dos dados de abril de 2014 e abril de 2015, previamente carregados. Acima do *combo box*, encontra-se a informação da concentração média do MP_{10} referente ao intervalo de dados representado no módulo superior. De modo semelhante, logo em baixo do *combo box* se destaca a concentração média do MP_{10} , neste caso, referente ao intervalo apresentado no módulo inferior. A legenda, vista na Figura 4.2h, define o padrão de cores adotado no gráfico bivariado.

A Figura 4.3 mostra a visualização do MP_{10} aferido entre 1997 e 2006, representando o perfil geral da concentração de poluentes para cada ano, em cada estação. Esse posicionamento dos módulos foi escolhido para que o usuário pudesse comparar as informações mais atuais (2015–2017), coletadas principalmente em dias alternados e exibidas no módulo superior, com os dados amostrados diariamente em campanhas anteriores (1997–2006 e 2014–2015), exibidos nos gráficos do módulo inferior. Com isso, o usuário pode verificar a evolução temporal dos poluentes atmosféricos, enquanto compara os períodos atual e passado, que possuem suas respectivas e diferentes particularidades, como as leis ambientais vigentes e a metodologia de amostragem, por exemplo.

Figura 4.3 – Ao escolher os dados de 1997 a 2006, a interface é atualizada para exibir em sua parte inferior os padrões históricos de concentração em massa do MP_{10} .

Fonte: Elaborada pelo autor.

Note que apenas o módulo inferior é atualizado quando se alterna entre os conjuntos de dados disponíveis no *combo box*. O usuário deve ser capaz de identificar, distinguir e rastrear as porções corretas da informação que sofre uma ação de atualização. Uma interação que implica na atualização de todos os elementos da página, incluindo aqueles não diretamente vinculados à interação em si, quebra a fluidez do processo exploratório. A interface foi concebida para, inicialmente, apresentar uma visão geral de dois conjuntos de séries temporais, viabilizando ao analista a comparação visual direta entre os valores desses conjuntos. Alguns padrões são melhor evidenciados apenas quando o observador se posiciona de um ponto de vista que compreenda todo o conjunto, ou seja, a partir de uma perspectiva que transmita ao usuário um contexto geral das informações. Logo, características de interesse que poderiam não ser legíveis por outro ângulo, pronunciam-se

desde o início para que sejam selecionadas e aproveitadas como entrada para métodos de exploração futuros ou para que sejam o ponto de partida de uma atualização da perspectiva atual do sistema de visualização (CRAFT; CAIRNS, 2005).

Exibir o contexto geral dos dados e permitir que o usuário desenvolva sua análise por meio da identificação e seleção das informações que lhe sejam relevantes, utilizando ferramentas de interação com a interface, em vez de apenas consumir informações estáticas e previamente apuradas, pode ser de maior utilidade para o processo exploratório.

4.4.2 Ferramentas Interativas

Um sistema de visualização deve mapear um número de propriedades nos dados sem que haja a necessidade de interação, ou seja, deve oferecer meios para a identificação de características pela simples navegação no espaço de informações, sem que pesquisas sejam formalizadas para isso. Além do mais, a interface deve oferecer canais adicionais de informação, por meio dos quais o usuário consiga identificar propriedades de acordo com o seu interesse ou objetivo de estudo (ANDREWS *et al.*, 2002). O modo como essas novas informações serão exibidas, em simultâneo com as que já estão em tela, deve ser cuidadosamente estudado para que o sistema permaneça intuitivo, pois quanto mais acessível for a visualização, maior será o seu público e impacto (MURRAY, 2013).

A Figura 4.4 mostra uma das ferramentas de interação de usuário com a interface. Cada ponto de dado dentro dos *scatter plots* registra um subconjunto de informações sobre o momento em que a coleta foi realizada. Ao mover o *mouse* sobre quaisquer um dos pontos, é exibida uma caixa de diálogo, ou “*tooltip*”, contendo as informações detalhadas do respectivo ponto. No exemplo da Figura 4.4, observa-se que para o dia de 17 de outubro de 2014, a concentração de MP_{10} amostrado foi de $91.76 \mu g m^{-3}$, não houve precipitação, a velocidade média do vento foi de 0.67 m/s, soprando a 3.33° . A temperatura média durante o dia foi de 31.8° Célsius, com 41% de umidade relativa do ar. Todo o estilo de coloração, transparências, arredondamento de bordas e posicionamento desse *tooltip* foi definido via CSS. Suas ações foram programadas dentro dos métodos de tratamento dos eventos de *mouse* oferecidos pela D3.

Após verificar o perfil geral descrito pelas concentrações de MP_{10} , o usuário pode investigar mais detalhadamente o comportamento de apenas uma parte dos dados, seja no *scatter plot* superior ou no inferior, ou mesmo em ambos. Para tal, foi desenvolvida uma ferramenta de seleção de dados com *zooming*. Essa técnica envolve o ajuste do tamanho e/ou posição dos elementos para reduzir a complexidade na representação dos dados, removendo as informações não interessantes da visualização (CRAFT; CAIRNS, 2005).

Figura 4.4 – Informações detalhadas podem ser acessadas simplesmente movendo o *mouse* sobre um ponto de interesse.

Fonte: Elaborada pelo autor.

Aplicar “*zooming-in*” em uma região amplia sua área e reposiciona seus elementos, revelando algumas informações que anteriormente estavam em oclusão devido às aglomerações que eventualmente se formam em espaços de grande densidade de pontos. Já o “*zooming out*” diminui o espaço entre os elementos, possibilitando representar maiores quantidades de informação. Com essas ferramentas, o ponto de vista da interface é atualizado para representar apenas as regiões que o usuário selecionar para analisar detalhadamente. A Figura 4.5 ilustra o funcionamento do *zooming*.

Para delimitar uma região, basta clicar com o botão esquerdo do *mouse* na parte do *scatter plot* que contém a porção inicial dos dados de interesse. Com o botão pressionado, o usuário deve arrastar o cursor para definir uma área de seleção, até que essa área compreenda todos os pontos que deseja analisar. Como visto na Figura 4.5a, é gerado um retângulo de cor cinza, com opacidade reduzida, que acompanhará o movimento de arrasto do cursor, enquanto o botão esquerdo estiver pressionado. Este elemento serve para auxiliar o analista a identificar quais os pontos que efetivamente estão entrando na seleção. Quando os pontos de interesse forem selecionados, basta soltar o botão esquerdo do *mouse* e um *zooming-in* irá automaticamente entrar em operação, alargando suavemente a região selecionada até preencher toda a área de seu respectivo *scatter plot*.

Quando muitos pontos de dado são representados no limitado espaço da tela, prover

Figura 4.5 – Exemplo de aplicação da ferramenta de seleção interativa de dados, que utiliza o conceito de *zooming*.

(a) Partindo do conjunto de dados coletado entre 1997 e 2006, apenas o intervalo correspondente ao ano de 1999 foi selecionado.

(b) *Zooming-in* reorganiza a visualização para representar apenas o intervalo selecionado, ou seja, o ano de 1999.

Fonte: Elaborada pelo autor.

um detalhamento mais aprofundado para todos os itens é impraticável (CRAFT; CAIRNS, 2005). No exemplo da Figura 4.5a, foi selecionada a área referente ao ano de 1999, apenas. A Figura 4.5b mostra o resultado desta ação. Agora é possível verificar o período escolhido em detalhes, com menor grau de sobreposição entre os elementos, uma vez que a coleção de itens representados foi delimitada de várias centenas para algumas dezenas. Clicando e arrastando dentro de um dos *scatter plot*, o usuário seleciona e expande uma região de seu interesse, logo, o *zooming-in* é aplicado como meio para facilitar a navegação nos detalhes do conjunto de dados ou mesmo em itens individuais (SHNEIDERMAN, 1996).

Caso houvesse a necessidade de verificar, por exemplo, a concentração de MP_{10} somente durante o inverno de 1999, bastaria fazer uma nova seleção contendo os pontos

entre os meses de junho a setembro de 1999, seguindo o procedimento descrito acima. Não há limite de profundidade para os níveis de detalhamento da visualização. Ao aplicar o *zooming-in* sucessivamente, é possível selecionar elementos até que reste apenas um único. Para reverter os efeitos dessas operações, basta efetuar um duplo clique com o botão esquerdo do *mouse* em qualquer local dentro do *scatter plot* e um *zooming out* restaurará suavemente a metáfora ao estado anterior, até que seja alcançado o seu ponto inicial.

Segundo Craft e Cairns (2005), uma interface otimizada deve permitir que se retorne a um estado anterior do processo exploratório por meio de um histórico. Então, quando um *zooming-in* é aplicado em algum dos módulos, todas as variáveis que definem o estado vigente deste módulo são armazenadas em uma “pilha”. O termo pilha refere-se aqui a um tipo de estrutura de dados que aplica a estratégia LIFO (*Last In, First Out*), isto é, o último elemento a entrar na pilha, será o primeiro a sair. Esse clássico conceito computacional foi aplicado para garantir o preciso controle do histórico do *zooming*, permitindo que o analista retorne sempre a um estado anterior em sua pesquisa, independente do caminho que tenha tomado para chegar no estado atual.

Com sucessivos duplos cliques (*zooming out*), os estados anteriores serão desempilhados do histórico, um a um, retornando gradativamente até o ponto de partida da representação. Além de facilitar na recuperação de erros ou no refinamento de uma pesquisa, o histórico permite a comparação do estado atual com os anteriores, algo que pode contribuir com o processo de entendimento dos dados (CRAFT; CAIRNS, 2005).

O efeito de uma operação de *zooming*, seja *zooming-in* ou *zooming out*, é local e independente. Ou seja, o ato de selecionar objetos em um dos *scatter plots* não afetará na disposição dos objetos do outro *scatter plot*. A transição entre os estados de *zooming* é suave para que, com isso, o usuário preserve seu senso de posição e contexto durante a navegação (SHNEIDERMAN, 1996; WAGEMANS *et al.*, 2012a). Note que o termo “*zooming*” é utilizado com frequência na literatura de maneira geométrica, referindo-se genericamente apenas às mudanças escalares na representação dos elementos em tela (CRAFT; CAIRNS, 2005). Em vez disso, o termo *zooming* é empregado de modo semântico no contexto desta dissertação, para descrever as mudanças relativas ao ponto de vista em que os dados são representados em suas metáforas (MCLACHLAN *et al.*, 2008).

Os gráficos auxiliares que contém as variáveis meteorológicas, por serem metáforas compactas, apresentam em seus eixos coordenados somente o essencial em termos de rotulação. Os eixos verticais exibem o nome da variável a qual cada gráfico se refere, junto de seus respectivos valores de máximo e mínimo. Já os eixos horizontais refletem o mesmo intervalo sob análise no *scatter plot* com a concentração de MP₁₀, anexo ao grupo auxiliar. Assim, foi decidido que a rotulagem da escala temporal não seria impressa porque essa

informação é a mesma da visualização do MP_{10} anexa, tendo como único efeito prático o de poluir a interface com uma informação redundante.

Figura 4.6 – Gráficos auxiliares apresentando as variáveis meteorológicas dos dados de maio de 2015 até novembro de 2017.

Fonte: Elaborada pelo autor.

Ao navegar pelo espaço interno de um dos gráficos auxiliares, uma linha vertical de referência é ativada, revelando as informações daquelas variáveis nos pontos correspondentes à posição do *mouse*. Como indicado na Figura 4.6, a data de amostragem é posicionada na base do gráfico sob o qual se paira o cursor, ajudando o analista a manter sua atenção sempre próxima da região em que observa. Ao mover o cursor ao longo dos gráficos, as informações são atualizadas com suavidade. Essas funcionalidades estão disponíveis em todos os gráficos auxiliares e as suas ações dinâmicas também foram desenvolvidas a partir dos métodos de tratamento dos eventos de *mouse* da biblioteca D3.

Vale destacar que o conteúdo das metáforas auxiliares é atualizado em escala temporal sempre que for aplicado um *zooming-in* ou *zooming out* no *scatter plot* anexo a este grupo. Ou seja, as escalas horizontais se mantêm coerentes em todos os gráficos do módulo, após quaisquer operações exploratórias na representação do MP_{10} , o que reforça a justificativa de se omitir as rotulações dos eixos horizontais dos grupos auxiliares.

Quando se aplica um *zooming-in*, o método responsável por executar essa ação no *scatter plot* retorna as datas inicial e final do intervalo destacado pelo usuário. Essas informações são enviadas como dados de entrada a outro método, que selecionará automa-

ticamente apenas as informações referentes a este intervalo, dentro das séries temporais meteorológicas. Então, as metáforas auxiliares são atualizadas para exibir somente esse intervalo de dados. Analogamente, após um *zooming out* no *scatter plot*, os gráficos auxiliares são gradualmente atualizados até que se alcance o estado de dados inicial. Todas as metáforas de um grupo auxiliar são alteradas ao mesmo tempo, em sincronia com a atuação do *zooming*.

Note que o *tooltip* (Figura 4.4) e os gráficos auxiliares (Figura 4.6) apresentam informações semelhantes, porém, em níveis contextuais diferentes. O usuário pode inspecionar todas as informações de seu interesse ativando o *tooltip* ao pairar o cursor sobre os pontos dos *scatter plots*. Entretanto, quando há centenas ou milhares de pontos de dado para serem verificados, a estratégia de análise ponto a ponto pode se tornar tediosa, até mesmo impraticável. Sendo a memória de trabalho humana limitada, é importante oferecer ao usuário uma visão geral das múltiplas variáveis meteorológicas, suportada pelos gráficos auxiliares, no sentido de orientar o usuário durante o processo exploratório. Assim, tanto o *tooltip* quanto os gráficos auxiliares atuam em conjunto para apresentar informações semelhantes, mas dentro de contextos e com níveis granulares distintos, não se tratando de uma mera duplicação de informações na interface.

4.4.3 Análise Comparativa entre Diferentes Intervalos

Ao analisar séries temporais, os usuários normalmente estão interessados em inspecionar a evolução do dado ao longo do tempo, por meio da comparação de informações localizadas em diferentes posições do eixo temporal (AIGNER *et al.*, 2007). Para aprimorar a percepção visual sobre a diferença entre os intervalos de dados nos módulos superior e inferior, foi desenvolvido o gráfico posicionado no centro da interface (Figura 4.2c). Trata-se de uma variação do gráfico de área bivariado, que é composto por duas linhas interligadas e sobrepostas no mesmo espaço. Para destacar a diferença entre as linhas, o espaço entre elas é preenchido com diferentes cores, dependendo de qual das linhas apresenta o maior valor a cada ponto, ao longo do eixo horizontal.

Ao evidenciar aumentos ou diminuições periódicas da força de uma variável em determinados intervalos, este gráfico também viabiliza o reconhecimento dos padrões temporais nos dados. Neste sentido, o objetivo da metáfora bivariada é destacar a evolução das diferenças entre os dados, expressa diretamente pela área compreendida entre suas curvas, e não somente os valores numéricos que cada uma das duas variáveis assume pontualmente ao longo do tempo. Logo, esta metáfora guarda semelhanças com o *stream graph*, discutido na Subseção 3.2.5, ao facilitar a determinação empírica do relacionamento entre as duas variáveis, contrastando-as simultaneamente no mesmo espaço.

Esse tipo de gráfico se baseia no princípio estatístico de análise bivariada, técnica utilizada para realizar estudos de tendências entre duas variáveis (BABBIE, 2011). Na clássica obra publicada em 1786, *The Commercial and Political Atlas*, William Playfair representou os valores da balança comercial da Inglaterra com vários outros países. O autor utilizou diferentes metáforas em seu trabalho, entre elas, a combinação de linhas e áreas coloridas, semelhante ao gráfico bivariado desenvolvido aqui.

A Figura 4.7 apresenta uma das análises comerciais empreendidas por Playfair. No exemplo, a linha vermelha descreve os valores das exportações realizadas pelos ingleses com a Dinamarca e a Noruega, em conjunto, enquanto que a linha amarela representa as importações, durante o mesmo período e ao longo dos anos de 1700 a 1780. Quando as importações superaram as exportações, ou seja, a balança comercial foi desfavorável aos ingleses, a área entre as linhas foi preenchida com a cor vermelha, que é o padrão de cor adotado para a linha das exportações. Analogamente, nos períodos em que a balança comercial foi favorável à Inglaterra, a área entre as linhas foi preenchida com a cor referente às exportações, o amarelo.

Figura 4.7 – Diferença de valores nas exportações e importações da Inglaterra com a Dinamarca e a Noruega, entre os anos de 1700 e 1780.

Fonte: Tufte (2007).

Como se observa na Figura 4.8, seguindo o padrão definido pela legenda, em azul estão representadas as variações da concentração do material particulado para o

período de 2015 a 2017 (*A*), e em vermelho está a evolução da concentração do MP_{10} entre 2014 e 2015 (*B*). Então, o gráfico bivariado representa a diferença relativa entre os dois conjuntos mediante a variação de amplitude de suas áreas coloridas. Isto é, para os intervalos formados por pontos de dado originados do conjunto *A*, que apresentem valores superiores em relação às suas contrapartes, pontos originados de *B*, a área entre as linhas será preenchida com a cor azul. Ao passo que nos intervalos em que os pontos de *B* apresentarem valores maiores, face aos de *A*, a área será preenchida em vermelho.

Figura 4.8 – Diferença de valores entre os dados de 2015–2017 (azul) e 2014–2015 (vermelho).

Fonte: Elaborada pelo autor.

Note que, diferindo do exemplo visto na Figura 4.7, para o gráfico bivariado desenvolvido aqui foi adotado como modelo de coloração das seções de área entre as linhas, utilizar a cor da linha que apresentar os maiores valores dentro da seção. Essa escolha de *design* foi feita para que houvesse maior destaque do conjunto de dados que dá origem aos maiores índices de MP_{10} , e também revele quais e quantos são os intervalos em que cada conjunto possui valores predominantes ao longo do gráfico.

Ainda que a inspeção minuciosa para descobrir valores ponto a ponto não seja a finalidade da metáfora bivariada, foi adicionada uma ferramenta de interação com o usuário, como é possível notar nas proximidades do centro da Figura 4.8. Semelhante ao que foi desenvolvido para os gráficos auxiliares, apresentados na Figura 4.6, trata-se de uma linha vertical referencial que segue a posição do cursor do *mouse*, ao longo da navegação pela área interna da metáfora. Dois marcadores com os valores das concentrações de MP_{10} acompanham simultaneamente a silhueta de ambas as curvas. Esses marcadores têm seus valores suavemente atualizados conforme o usuário navega com o cursor pelo gráfico, servindo como uma referência contextual.

Para tirar o máximo de proveito das capacidades do gráfico bivariado, o ideal é fazer comparações entre intervalos de tempo correspondentes, ou seja, a metáfora deve refletir a diferença entre duas estações climáticas equivalentes. Porém, no que diz respeito aos dados de entrada, algumas questões se tornaram evidentes durante a construção dessa metáfora e exigiram tratamentos específicos.

A primeira adversidade está relacionada com a grande diferença de tamanho entre as três bases de dados sob análise. Essa diferença não se dá apenas quanto ao número total de instâncias de cada conjunto, o que já era esperado, uma vez que um dos conjuntos é composto por informações coletadas em dias alternados, enquanto o outro é formado

por amostragens diárias. Mas, principalmente, existe uma grande diferença de amplitude temporal entre os intervalos de coleta, que abrangem 1 ano, 3 anos e 10 anos, em diferentes momentos, conforme exposto na Subseção 4.2.1. Não faria sentido e nem seria útil, ou mesmo lógico, representar no gráfico bivariado a diferença entre duas bases de dados com diferentes amplitudes temporais, sem algum tipo de ajuste para que as estações ao menos coincidisse. Não proceder esse ajuste, causaria a perda de uma das principais motivações da metáfora: a identificação de características sazonais a partir do contraste visual entre duas variáveis. Para mitigar esse problema, foi desenvolvida e aplicada a estratégia de interpolação com intervalo de dados deslizante.

A Figura 4.9 ilustra o conceito de intervalo deslizante sobre os conjuntos A e B , com amplitudes temporais de 3 anos e 1 ano, respectivamente. Nessa abordagem, um subconjunto menor será comparado ponto a ponto com um subconjunto maior. Ao se alcançar o final do subconjunto menor, este é reiniciado, sucessivamente, até que se atinja o final do subconjunto maior. Esse movimento é semelhante ao deslizamento de um intervalo, ou janela de dados (KEOGH; LIN; TRUPPEL, 2003). Entende-se aqui, que os termos maior e menor se referem estritamente à extensão temporal e não ao número de instâncias de dado de cada conjunto. Com essa estratégia, é assegurada a comparação de meses e estações correspondentes, mesmo dispondo de bases de dados que compreendam intervalos de tempo arbitrários. Note que a estratégia de deslizamento gera ciclos, que tem por consequência, dentro do processo de comparação, a não linearidade temporal da base menor.

Figura 4.9 – Intervalo deslizante entre os conjuntos de dados A e B .

Fonte: Elaborada pelo autor.

Além disso, partiu-se do princípio de que nessas três bases de dados, os intervalos entre dias subsequentes de amostragem também poderiam não coincidir. Por exemplo, o dia d , do mês m , da base A , pode não ter sua contraparte no conjunto B , seja pela alternância entre os dias de coleta, ou mesmo algum tipo de falha, algo que impediria uma comparação ponto a ponto entre os dados. Então, para equalizar os registros e garantir que haja correspondência entre os dias pertencentes a estações correspondentes, mas em conjuntos (intervalos) diferentes, utilizou-se a interpolação por *splines* cúbicas. Esta robusta técnica de aproximação, descrita detalhadamente por McKinley e Levine (1998), consiste em dividir o intervalo de interesse em uma série de subintervalos e interpolar estes subintervalos com polinômios cúbicos, obtendo uma curva de aparência suave. A Figura 4.10 apresenta

uma interpolação com *splines* cúbicas sobre pontos de dado arbitrários. O resultado das aproximações se traduz em uma curva com maior aspecto de continuidade entre os intervalos discretos, quando comparada à interpolação linear, por exemplo.

Figura 4.10 – Exemplo comparativo entre interpolação de dados linear e interpolação utilizando *splines* cúbicas, sobre os mesmos pontos de dados.

Fonte: Elaborada pelo autor.

Outra questão importante relacionada às bases de dados, tem a ver com os períodos prolongados em que, por alguma adversidade, não houve amostragem (cf. Seção 4.2.1). No geral, foram ocorrências raras que não causaram prejuízos nas demais partes da interface. Porém, em alguns casos há registros ausentes por períodos superiores a duas semanas. Embora preserve o perfil geral de uma sequência de tempo, a interpolação pode distorcer a velocidade com que o perfil da curva muda de valor, ocasionando saltos locais (POST; HARRISON, 2008). Grandes intervalos entre dois pontos de dado consecutivos tendem a provocar efeitos indesejados no método interpolador, que apresentará estimativas muito distantes da média, podendo atingir valores negativos, algo que não teria lógica neste contexto. Mesmo com o gráfico bivariado não tendo o compromisso intrínseco de ser preciso na descrição de valores pontuais, estimativas pouco razoáveis causariam prejuízos estéticos a todo o sistema. Para essas ocorrências de dados faltantes, existem estratégias de mineração de dados que podem ser adotadas, dentre as variadas técnicas de pré-processamento.

A imputação de dados faltantes é uma estratégia que substitui um valor ausente por um valor constante admissível (MYRTVEIT; STENSRUD; OLSSON, 2001). Os valores faltantes imputados são obtidos a partir do próprio conjunto de dados, por meio da média, da mediana ou da moda, que é aquele valor que ocorre com mais frequência entre os

valores de uma variável. Neste trabalho, os valores ausentes de MP_{10} foram preenchidos, somente em períodos com dados faltantes superiores a uma semana, com a média. A média tem por característica não influenciar de modo abrupto o conjunto de dados, por se tratar de uma tendência geral deste. Diferente dos demais módulos da interface, em que se pretende transmitir fielmente os valores aferidos e, por isso, optou-se por excluir as entradas corrompidas, conforme definido em 4.2.1, o gráfico bivariado representa um “esboço” da diferença entre duas variáveis. Logo, tanto a imputação do valor médio quanto a interpolação, justificaram-se para essa metáfora.

Ajustados os intervalos deslizantes e, quando necessário, os dados faltantes, as interpolações para gerar as novas bases de dados que alimentarão o gráfico bivariado podem ser executadas. Para isso, foi desenvolvido um *script* utilizando a linguagem *Python*, versão 3. Basicamente, o programa faz a leitura de dois conjuntos de dados, C_1 e C_2 , que serão chamados de conjuntos iniciais, e, para cada conjunto inicial, gera dois vetores, $\vec{x}$ e $\vec{y}$, contendo somente os valores de concentração do MP_{10} e suas correspondentes datas de coleta. Os pares de vetores são interpolados por *splines* cúbicas, uma interpolação I para cada par $\vec{x}$ e $\vec{y}$, de cada conjunto inicial, resultando em duas curvas definidas pelas funções interpoladoras, I_{C_1} e I_{C_2} . A seguir, gera-se um novo conjunto de dados contendo novas datas, que são obtidas pela concatenação única das datas dos conjuntos iniciais; os valores de concentração, resultantes da iteração da função interpoladora I_{C_1} , ponto a ponto, em cada uma das novas datas; e também os valores de concentração, obtidos de modo semelhante, a partir da função interpoladora I_{C_2} .

Ao final, dois novos arquivos foram gerados: um referente à diferença entre os intervalos 2015–2017 e 2014–2015, que é carregado automaticamente quando se inicializa a interface; e outro para a comparação dos intervalos de 2015–2017 e 1997–2006, que é carregado quando o usuário seleciona os dados coletados entre 1997 e 2006 no *combo-box* (Figura 4.2d). O processo de geração desses novos conjuntos, desde as leituras dos dados iniciais, até a interpolação dos valores e a escrita nos arquivos, não levou mais de 1 segundo cada (computador comum, com 24GB de RAM e processador Intel® Core™ i7-3770 de 3.40 GHz, em ambiente Linux 64-bits). O código fonte está disponível em um repositório *online* em <<https://goo.gl/CbDJAn>>, onde também se encontram os dados amostrados.

Assim como nos grupos auxiliares, ao se aplicar *zoom* nos *scatter plots*, o conteúdo do gráfico bivariado será atualizado para corresponder com as informações selecionadas pelo usuário, como apresentado na sequência de imagens da Figura 4.5. No entanto, deve ser destacado que essa atualização só tem efeito quando originada no *scatter plot* que exibe a base de dados de maior amplitude temporal. Essa escolha de *design* se deu por conta do efeito cíclico e da não linearidade temporal ocasionada pela técnica de janela deslizante.

Efeito esse que poderia causar confusão visual, dado que o conjunto de menor amplitude temporal pode apresentar algumas repetições de instâncias no gráfico bivariado, para o trecho de dado selecionado.

A Figura 4.11 apresenta um recorte da interface que ilustra o funcionamento do gráfico bivariado após alguns passos exploratórios. Inicialmente, a interface gráfica foi carregada com as bases de dados *A* e *B*. A partir disso, foi aplicado no *scatter plot* localizado na parte superior, que exibe o conjunto de dados *A*, um *zooming-in* na área correspondente ao período entre os meses de junho de 2017 e novembro de 2017. Esta operação destaca em *A* a estação do inverno, em que são revelados alguns pontos de alta concentração de PM_{10} . Com a finalidade de realizar uma comparação com as informações do material particulado contidas em *B*, representado no *scatter plot* inferior, em um intervalo semelhante, também foi aplicado um *zooming-in*, agora na área correspondente aos meses de junho de 2014 até novembro de 2014.

Figura 4.11 – Comparação entre dois intervalos, dos conjuntos *A* e *B*, que apresentam altas concentrações de PM_{10} .

Fonte: Elaborada pelo autor.

Quando se verifica o gráfico bivariado resultante dessas operações, ao centro da Figura 4.11, identificam-se vários picos nas curvas, ora referentes aos dados de *A*, ora referentes aos de *B*. Dois picos se destacam em azul, indicando como sua origem, dias

pertencentes ao inverno de 2017 que apresentaram valores de MP_{10} sensivelmente elevados (Figura 4.11a-b). Os demais picos estão majoritariamente coloridos em vermelho, indicando dias de alta concentração de material pertencentes ao ano de 2014 (Figura 4.11c-d-e).

O gráfico bivariado não possui rótulos no eixo horizontal, que representa o tempo. Essa decisão se justifica devido a própria construção deste gráfico, que interliga e sobrepõe duas séries temporais, registradas ao longo de diferentes anos. A cor das linhas exprime a qual dos conjuntos de dados cada curva pertence, relacionado-as aos seus respectivos módulos de análise, superior ou inferior, da interface. Seria um tanto redundante exibir essas marcações, visto que os *scatter plots* já exibem a rotulagem temporal em seus eixos horizontais. Novamente, essa metáfora tem o intuito de informar o fluxo evolutivo das duas variáveis e não características pontuais.

4.5 O Sistema de Análise Visual dos Índices de Saúde

Para viabilizar a análise visual dos números ligados à incidência de doenças relacionadas ao aparelho respiratório, registrados em São Carlos-SP, foi desenvolvida uma segunda interface com *JavaScript* e *D3*. Esta interface também está disponível *online*, podendo ser acessada em [<https://evortigosa.github.io/health2/>](https://evortigosa.github.io/health2/).

As configurações básicas deste sistema são semelhantes às aquelas definidas para o trabalho da Seção 4.4, ou seja, a resolução de 1920×1080 , *Full HD*, foi estabelecida como padrão. Devido a utilização de elementos gráficos escaláveis em *SVG*, é possível navegar em ambientes com menor (ou maior) resolução, desde que, para isso, o usuário ajuste o *zoom* de seu *browser*, seja em computadores ou dispositivos móveis.

4.5.1 Interface e seus Componentes

Esta interface é baseada em três metáforas: barras agrupadas, *stacked graphs* e gráficos de área. Os gráficos de barras são bastante conhecidos dentro das aplicações de visualização, porém, mostram-se pouco eficientes quanto ao aproveitamento de espaço, uma vez que cada instância de dado ocupa individualmente o espaço de toda uma barra, que pode se estender verticalmente da linha de base do gráfico até o seu topo.

Horizontalmente, essa limitação espacial influencia na quantidade de barras (instâncias de dado) que podem ser exibidas simultaneamente, pois quanto mais instâncias houver, mais barras serão necessárias para representá-las. Para que várias barras sejam exibidas ao mesmo tempo, em um espaço limitado, será necessário reduzir a espessura individual de cada barra, de modo que todas elas possam ser acomodadas dentro da área

interna do gráfico. No entanto, a espessura de uma barra não pode ser reduzida a tal ponto que dificulte a sua identificação e interpretação, algo que resultaria em confusão na leitura do gráfico. Então, para que uma aplicação visual utilize metáforas de barras e seja eficiente em transmitir informações, a quantidade de instâncias representadas não deve ser elevada ou devem ser desenvolvidos mecanismos interativos para selecionar os dados a serem exibidos.

A Figura 4.12 apresenta a visão inicial da interface. Note que este *layout* também se divide em dois módulos gráficos semelhantes entre si, um superior e um inferior, sendo que essa configuração foi escolhida visando facilitar a comparação ponto a ponto entre as informações representados em ambos os módulos, equivalente ao descrito em 4.4.1. No módulo de visualização superior, encontra-se disponível o conjunto de dados consolidado entre 2014 e 2017. Este módulo é inicializado com o ano de 2014 e com a enfermidade “pneumonia” selecionados (Figura 4.12d). Já no módulo posicionado abaixo, estão os índices das doenças respiratórias registrados entre 1998 e 2006. Por padrão, este módulo é inicializado com o ano de 1998 e “pneumonia” previamente selecionados.

Figura 4.12 – Visão inicial da interface, apresentando as ocorrências de “pneumonia” durante o ano de 2014 (superior) e os casos registrados em 1998 da mesma doença (inferior).

Fonte: Elaborada pelo autor.

Examinando o módulo superior, observa-se que à esquerda (Figura 4.12a) está posicionada a variação temporal dos casos da enfermidade escolhida pelo usuário, registrados ao longo do ano selecionado. Neste gráfico, são apresentados os números da patologia em questão, discretizados mensalmente, por meio de agrupamentos de barras. Isto é,

há um grupo com até 8 barras coloridas que exibem a quantidade de casos registrados daquela doença, sendo um grupo de barras destinado a cada um dos meses do ano. Cada uma das barras contidas em um grupo (mês) representa uma das 8 faixas etárias em que os pacientes se distribuem (cf. Subseção 4.2.2). Logo acima desta metáfora de barras agrupadas, encontra-se um gráfico de área que apresenta a variação (média) mensal do poluente MP_{10} , amostrado dentro do ano selecionado.

No centro do módulo (Figura 4.12b), posiciona-se um *stacked graph*. Para manter a consistência visual entre metáforas contendo informações semelhantes, foi configurado o *layout* em barras para este *stacked graph*. Essa metáfora agrega os números anuais de incidência da doença respiratória escolhida, com as barras empilhadas seguindo as classes etárias dos pacientes. Além de quebrar com o paradigma dos gráficos de barras tradicionais, ao introduzir mais informações por elemento barra individual, esta metáfora apresenta um resumo da evolução temporal da doença selecionada, ao longo de todos os anos contidos na série a qual o ano apresentado no gráfico de barras agrupadas pertence, atuando como uma referência contextual para o usuário. Este *stacked graph* sintetiza uma grande quantidade de informações e ocupa uma porção de espaço reduzida, mostrando-se como uma alternativa espaço-eficiente. Logo acima, localiza-se um gráfico de área com o panorama mensal do MP_{10} , abrangendo o mesmo período do *stacked graph*.

À direita (Figura 4.12c), destaca-se um grupo auxiliar que integra 5 gráficos de área dispostos verticalmente. Esses gráficos individualizam e apresentam o perfil geral das demais enfermidades (não mapeadas nas metáforas de barras do módulo) contidas na série temporal, ao longo de todo o intervalo. O grupo auxiliar, ou *small multiples*, também funciona como meio de transmissão de informações contextuais. Com isso, o usuário tem visualmente e ao seu alcance um resumo do comportamento geral das demais enfermidades. Vale destacar que não há divisões etárias dentro das variáveis exibidas nos grupos auxiliares. Os números de cada uma das enfermidades são agregados sem considerar as suas faixas etárias e, então, são representados nos gráficos de área do grupo auxiliar, um gráfico para cada variável. Este é basicamente o mesmo conceito funcional dos grupos auxiliares da interface dedicada aos dados ambientais (cf. Figura 4.2b).

Note que as curvas das metáforas de área desta interface foram suavizadas. Essa escolha foi tomada para evitar o efeito “serrilhado” que pode ocorrer nesse tipo de gráfico. Devido à discretização mensal dos índices das doenças, há menor quantidade de instâncias de dado nas séries, face a quantidade de pontos representados em intervalos de tempo semelhantes pelo sistema dedicado aos dados ambientais (granularidade em dias). Quando poucos pontos são interpolados em gráficos baseados em linhas, há a necessidade de se aumentar o espaçamento entre esses pontos para preencher o espaço disponível, resultando

em curvas angulares, nem sempre agradáveis visualmente. Essa suavização é realizada pelo método interpolador da D3, definindo uma variável de aspecto para “curva natural”.

A partir dos dois *combo boxes* posicionados em baixo do gráfico de barras agrupadas (Figura 4.12d), é possível selecionar e alterar o ano de registro dos dados, bem como a patologia mapeada nas representações das Figuras 4.12a-b. Ao se escolher uma nova doença, todas as metáforas do módulo são atualizadas. De modo coloquial, quando o usuário escolhe uma nova variável (doença), esta “sairá” do grupo auxiliar e será apresentada em detalhes nos gráficos de barras (agrupadas e empilhadas), ao passo que a variável que estava antes nas metáforas de barras “voltará” para o grupo auxiliar.

Tecnicamente, o método responsável pelo manuseio dos dados opera sobre a variável escolhida, separando as faixas etárias a serem representadas nas barras agrupadas, dentro do ano escolhido. Então, os valores mensais pertencentes a cada ano, para a variável selecionada, são agregados mantendo a separação de faixas etárias. Com isso, cria-se um subconjunto com valores agregados anualmente, que serão exibidos no *stacked graph*. Por fim, as demais enfermidades são agregadas em números gerais, isto é, sem levar em conta a separação etária, gerando subconjuntos para cada uma das variáveis, que serão apresentadas nas áreas do grupo auxiliar. Ao definir uma nova variável de interesse, o processo se repete e atualiza os mapeamentos do módulo.

Quando o usuário modifica o ano de interesse, a representação de barras agrupadas é alterada para exibir as informações daquele ano, para a enfermidade em análise. Os demais gráficos do módulo não são alterados. Essa organização das metáforas gráficas do módulo foi configurada para criar a intuição de fluxo de informações. Da esquerda para a direita, as variáveis de interesse são mapeadas em diferentes granularidades: da mais fina, no gráfico em barras agrupadas, que apresenta as informações mês a mês, para cada faixa etária acometida pela enfermidade selecionada; passando por um nível granular intermediário, no *stacked graph*, que mapeia os casos da doença selecionada discretizados anualmente, ainda com a diferenciação etária; até o nível granular mais grosso, no grupo auxiliar, que descreve a evolução das demais patologias em números gerais.

Há uma legenda na parte inferior direita do *layout* (Figura 4.12e), que informa ao usuário o padrão de cores que representa cada uma das classes etárias dos pacientes. O esquema de cores desta interface também foi obtido junto ao *ColorBrewer* (HARROWER; BREWER, 2003), no entanto, foi utilizado aqui um mapa qualitativo. Os esquemas de cores qualitativos são os mais adequados para serem aplicados em representações de dados categóricos. Em particular, o esquema adotado é conveniente para a exibição em telas de dispositivos eletrônicos (LCD) e para a impressão. Conforme descrito em 4.4.1, a cor de fundo de toda a interface é o branco.

4.5.2 Ferramentas Interativas

Todas as características e funcionalidades descritas para o módulo de visualização superior são válidas e replicáveis no módulo inferior. A Figura 4.13 apresenta algumas das ferramentas de interação com o usuário oferecidas pela interface. Ao pairar o cursor do *mouse* sobre qualquer barra, seja nos gráficos de barras agrupadas ou nos *stacked graphs*, de ambos os módulos, será exibido um *tooltip* com informações da instância de dado representada por aquele elemento. O *tooltip* acompanhará o movimento do *mouse* enquanto o cursor estiver dentro da barra. No exemplo da Figura 4.13a, observa-se que durante o mês de setembro de 2014, 30 pessoas com mais de 60 anos foram acometidas com pneumonia e deram entrada em alguma unidade médica da cidade.

Figura 4.13 – As ferramentas de interação apresentam informações do dado sob o cursor, ao navegar entre as barras agrupadas (a), no *stacked graph* (b) ou nos grupos auxiliares (c).

Fonte: Elaborada pelo autor.

Além do *tooltip*, a barra sobre a qual o usuário está movendo o cursor se destaca das demais, por meio do realce de seu contorno (Figura 4.13a). Essa funcionalidade está disponível apenas para os gráficos de barras agrupadas, pois, devido às características de construção desta metáfora, as barras de um mesmo grupo encontram-se bem próximas umas das outras. Ainda que o esquema de cores utilizado para categorizar as classes etárias diferencie suficientemente os elementos, o destacamento de contorno transmite ao usuário mais clareza de qual é a instância de dado sob investigação, mesmo que esta seja representada por uma barra relativamente pequena e/ou esteja entre várias outras.

Na Figura 4.13b, encontra-se o *stacked graph* com os números da doença selecionada acumulados anualmente. Neste exemplo, foram atendidos e diagnosticados com pneumonia, o total de 230 pacientes com 60 anos ou mais durante o ano de 2014. Vale reforçar que 2014 possui registros entre os meses de abril até dezembro (cf. Subseção 4.2.2). Os demais anos disponíveis no sistema contém informações de janeiro a dezembro.

Ao mover o *mouse* dentro de uma das metáforas de área, seja naquelas que mapeiam o perfil mensal do MP_{10} ou nos grupos auxiliares, será ativada uma linha vertical, acompanhada de marcadores, que informará ao usuário os valores pontuais das variáveis sob a linha. Os gráficos de área apresentam apenas o essencial em termos de rotulagem vertical, como o nome da variável que representam e seus valores de mínimo e máximo. Observe que apenas o gráfico da base do grupo auxiliar possui rótulos temporais no eixo horizontal (Figura 4.13c). Todas as variáveis apresentadas dentro do mesmo grupo auxiliar coincidem em escala temporal, enquanto as áreas que representam o MP_{10} coincidem com o intervalo das metáforas de barras as quais estão conectadas. Logo, não há necessidade de replicar a rotulagem temporal. Porém, nos grupos auxiliares, a data em que o dado foi registrado é apresentada sob a metáfora mais próxima do cursor.

Conforme apresentado na Figura 4.13c, 68 casos de pneumonia foram registrados em São Carlos-SP durante o mês de setembro de 2014. Destes, 30 pacientes tinham mais de 60 anos, como está ilustrado na Figura 4.13a. Os outros 38 pacientes estão distribuídos pelas 7 demais faixas etárias, com destaque para os menores de 5 anos, que registraram 17 ocorrências dessa doença em setembro de 2014. É interessante observar que essas duas faixas etárias, menores de 5 anos e maiores de 60 anos, destacam-se e apresentam uma certa tendência predominante de adoecimento quando se verifica o cenário conjuntural ilustrado pela Figura 4.13.

Por meio da interação direta com o sistema de visualização, o analista obtém informações de modo simplificado e sucinto. Ou seja, com os recursos interativos oferecidos nesta interface, o conteúdo detalhado daquelas instâncias que o usuário tem interesse é identificado rapidamente e com baixo esforço cognitivo, algo que pode resultar em ganhos

de tempo no processo analítico desses dados. Todas as funcionalidades interativas das ferramentas descritas aqui foram desenvolvidas dentro dos métodos de tratamento de eventos de *mouse* da D3, com definições de formatação e estilo feitas via CSS.

Note que as ferramentas interativas desta interface são menos sofisticadas, quando comparadas com as ferramentas detalhadas na Subseção 4.4.2. O intuito da representação visual dos índices das doenças relacionadas ao aparelho respiratório foi apresentar um panorama da evolução desses dados, ao longo do tempo. Neste contexto, este sistema gráfico se posiciona como um painel de informações interativo, ante ao sistema de análise e exploração visual dos dados ambientais, especificado na Seção 4.4.

Entretanto, vale ressaltar que os métodos de leitura e tratamento de dados desenvolvidos para esta interface são mais complexos e realizam maiores quantidades de operações de álgebra relacional, se comparados aos métodos equivalentes da interface dedicada aos dados ambientais. Os dois conjuntos de séries temporais analisados aqui contêm registros de 6 tipos de enfermidades que se subdividem, cada enfermidade, em 8 categorias etárias. Ou seja, além dessas duas séries possuírem instâncias com maior quantidade de atributos (dimensões), os valores de cada um desses atributos são desmembrados e/ou agregados de diferentes maneiras durante a navegação, dependendo das necessidades representacionais da metáfora que exibirá esses dados.

4.6 Considerações Finais

Neste Capítulo foram descritos os sistemas de visualização desenvolvidos para viabilizar análises sobre o comportamento evolutivo de séries temporais, amostradas em São Carlos–SP. As interfaces gráficas foram construídas a partir da demanda de especialistas no domínio de monitoramento ambiental e controle da qualidade do ar. Destaca-se que o processo de desenvolvimento das metáforas, bem como da metodologia exploratória, envolveu consultas aos especialistas durante vários estágios. Isso motivou a redefinição de técnicas, componentes e funcionalidades, com o objetivo de customizar e adequar os recursos, de acordo com as tarefas analíticas definidas por esses especialistas.

Embora a comunidade científica tenha se concentrado em aplicações de mineração visual de dados, o problema fundamental sobre o espaço de *design* dos sistemas de representação de séries temporais não foi plenamente explorado até o momento (BACH *et al.*, 2014). As ferramentas gráficas devem ser cuidadosamente projetadas para apresentar um espaço de comunicação conciso, pois o excesso de informação se transforma em ruído, sobrecarregando o usuário. No entanto, concisão não significa omissão de informações, que podem ser potencialmente importantes. Assim, o equilíbrio entre a quantidade de dados

exibidos em simultâneo e as boas escolhas de *design*, aliadas a recursos interativos que contornem as complexidades estruturais dos dados, são os desafios a serem enfrentados no desenvolvimento de ferramentas de visualização.

Cada passo da construção envolveu estudos em *design* gráfico, visando facilitar a experiência do usuário em analisar visualmente grandes quantidades de informações. As técnicas de navegação e interação foram incluídas para dar suporte à geração de diferentes níveis de detalhamento, reorganizando as representações conforme o analista se aprofunda nas particularidades dos dados.

Ambas as interfaces foram desenvolvidas em página única, reunindo metáforas compactas que apresentam simultaneamente a evolução das variáveis observadas ao longo do tempo. Essa decisão foi tomada, pois o processo de identificar e correlacionar detalhes entre itens de dado é afetado quando estes estão distribuídos em interfaces que requeiram acesso a páginas externas, ou mesmo demandem rolamentos dentro de uma página grande (MCLACHLAN *et al.*, 2008). Ou seja, os sistemas visuais devem ser projetados para dar suporte a comparações diretas, com o objetivo de determinar como as múltiplas variáveis afetam o comportamento geral dos dados.

Capítulo 5

Resultados

5.1 Considerações Iniciais

Em seu trabalho sobre validação de sistemas de visualização de informações, Plaisant (2004) argumentou que é prática comum na comunidade de visualização, a investigação da usabilidade das ferramentas recém desenvolvidas, mediante a condução de experimentos exploratórios qualitativos chamados “estudos de caso”. Os estudos de caso devem demonstrar tarefas reais ao usuário, enquanto descrevem e documentam todo o contexto situacional, bem como as variáveis envolvidas no processo. Plaisant (2004) ainda defendeu que os estudos de caso têm potencial para legitimar o sistema de visualização, uma vez que apresentam aos utilizadores as suas funcionalidades e capacidades.

Este Capítulo coloca em perspectiva alguns dos resultados obtidos com as ferramentas construídas nesta pesquisa. Para ilustrar a efetividade oferecida pelas interfaces projetadas, em abordar as tarefas analíticas apresentadas na Seção 4.3, serão apresentados três estudos de caso na Seção 5.2. Já na Seção 5.3 há uma discussão ampliada, levando em consideração não apenas a visualização em si, mas também as particularidades do meio ambiente em que os dados foram amostrados. Na Seção 5.4 estão algumas considerações finais sobre este Capítulo.

5.2 Estudos de Caso

O primeiro estudo de caso aborda as tarefas analíticas T1 e T2, enquanto o segundo está centrado na verificação de tendências gerais sobre as amostragens do MP₁₀, relacionando-se com T1 e T3. Já o terceiro estudo de caso foi elaborado com o objetivo de investigar quais os padrões de ocorrência das doenças respiratórias, além de buscar

paralelos entre essas informações e os níveis dos poluentes, relacionando-se com T4 e T5. Em todos os estudos de caso foram utilizados os conjuntos de dados descritos na Seção 4.2.

5.2.1 Comparando Padrões dos Dados Ambientais

O objetivo deste estudo de caso é analisar os padrões temporais dos poluentes, em diferentes intervalos de amostragem, para compreender como a dispersão desses materiais tem evoluído na cidade, de acordo com a influência das variáveis ambientais. Além do mais, também será investigada a dinâmica temporal das diferentes estações do ano.

Delimitando a análise apenas às informações disponíveis nos dados que compreendem a série de maio de 2015 até novembro de 2017, verificam-se três intervalos que apresentaram picos nos valores da concentração de material particulado (cf. Figura 4.2a). Esses três intervalos exibem algumas características em comum, que serão investigadas em maior profundidade, selecionando as informações de interesse com *zooming* (T1).

Inicialmente, observa-se que nas três situações houve uma tendência de aumento nas concentrações do MP_{10} , que começou a se manifestar em meados do mês de junho, fato este que coincide com o início do inverno, estação regionalmente marcada pelos baixos índices pluviométricos e pela queda nas temperaturas médias. Como ilustram os *scatter plots* da Figura 5.1, essa tendência de aumento atingiu os seus valores máximos entre os meses de agosto e outubro. Note que o ano de 2015 (Figura 5.1a) apresentou uma ligeira diferença com relação aos dois anos seguintes, com seus picos de concentração de MP_{10} registrados no período entre os meses de setembro e outubro, ou seja, entre o final do inverno e o início da primavera.

O comportamento verificado em 2015 se assemelha ao perfil descrito pelos dados amostrados durante o ano de 2014 (cf. Figura 4.2e), período que também apresentou altos níveis de MP_{10} durante o início da primavera. Nos dois anos seguintes, 2016 (Figura 5.1b) e 2017 (Figura 5.1c), os picos de concentração se limitaram ao inverno, principalmente entre os meses de agosto e setembro, com uma acentuada tendência de diminuição nos valores a partir dos meses de outubro (início da primavera).

Comparando os perfis descritos pelas variáveis meteorológicas, representadas nos gráficos auxiliares da Figura 5.1, nota-se que houve uma relativa estabilidade no clima da região amostral, para os três intervalos selecionados. Em geral, ocorreram longos períodos sem precipitação, com temperaturas médias em tornos dos 20 ou 25 graus Célsius. Os ventos foram suaves, com raras ocorrências acima dos 3 m/s. A umidade relativa do ar apresentou níveis próximos aos 50% em diversas ocasiões. Os mais altos níveis de concentração do material particulado, registrados entre os anos de 2015 e 2017, ocorreram

Figura 5.1 – Análise dos picos de concentração de PM_{10} para os dados aferido de maio de 2015 até novembro de 2017.

Fonte: Elaborada pelo autor.

em três dias do mês de setembro de 2017, com valores que ultrapassaram os $85 \mu\text{g m}^{-3}$. Vale destacar que no dia 26 de setembro de 2017 (início da primavera, Figura 5.1c), registrou-se o maior volume de MP_{10} da série 2015–2017, com $91,50 \mu\text{g m}^{-3}$.

Inspecionando o gráfico “Umidade” na Figura 5.1c, é interessante observar que existe uma certa “coincidência” entre o intervalo contendo esses pontos de máxima concentração do MP_{10} , com uma queda acentuada nas taxas de umidade relativa do ar. Essa tendência de queda na umidade relativa também ocorreu durante setembro de 2017, quando foram registrados, ao longo de vários dias consecutivos, níveis considerados preocupantes pela OMS, próximos aos 30%.

Obviamente, não seria suficiente cientificamente, ou mesmo prudente, gerar alguma conclusão a respeito da dinâmica ambiental da região de amostragem, tendo por base apenas o aparente comportamento inversamente coincidente dessas duas variáveis. Ainda que as ferramentas de visualização de informações sejam projetadas para guiar o processo de análise de dados, seria um tanto simplista esperar que apenas a análise visual pudesse responder questões tão complexas. No entanto, a partir das perspectivas verificadas por meio da exploração interativa entre centenas de instâncias de dado, é possível inferir algumas hipóteses e cenários que servirão de base a futuras investigações e testes estatísticos.

Um importante aspecto sobre a dinâmica de dispersão dos poluentes é a sazonalidade. Na Figura 4.3 foram evidenciadas variações sazonais nos índices de MP_{10} . A temperatura do ar afeta diretamente a dispersão dos poluentes atmosféricos. O ar frio é mais denso do que o ar quente, fato este que, por consequência, aumenta o arrasto aerodinâmico, que é um tipo de força de atrito que atua sobre as partículas suspensas, dificultando a sua dissipação. As taxas mais elevadas de umidade relativa geralmente acompanham o ar quente, com isso também afetando na concentração do material particulado. Durante as estações “secas” (abril a outubro), a concentração de MP_{10} é maior do que nas estações “chuvosas” (novembro a março). De acordo com Pozza *et al.* (2010), uma das razões dessa redução sazonal nas concentrações de MP_{10} entre novembro e março foi a deposição das partículas devido à incidência das chuvas.

A Figura 5.2, obtida por meio do *zooming-in*, destaca que as mais altas concentrações do MP_{10} foram registradas durante o inverno. Mesmo com as mudanças metodológicas nos padrões de coletas, de diárias para dias alternados (cf. Seção 4.2), não houve alterações significativas no perfil da concentração do material particulado, ou seja, as estações do ano que historicamente apresentavam os maiores e os menores valores do MP_{10} , não sofreram grandes mudanças em seu padrão recente de concentração de poluentes.

A inspeção visual das características climáticas do meio ambiente encontrado na região de amostragem, ofereceu indícios da possível influência dos elementos naturais sobre

Figura 5.2 – Perfil descrito pelos dados coletados entre janeiro de 2016 e novembro de 2017.

Fonte: Elaborada pelo autor.

as tendências anuais de aumento ou diminuição da concentração de poluentes atmosféricos. Neste sentido, um fator que aparenta ter grande interferência sobre a dinâmica de dispersão do MP_{10} é a sazonalidade climática imposta pelas estações do ano, uma vez que os meses secos de inverno costumam concentrar maiores volumes de partículas, enquanto nos meses chuvosos (primavera-verão), as concentrações diminuem. Essas informações devem ser tomadas pelos especialistas como um guia inicial para as suas pesquisas, uma vez que, como mencionado, somente as informações obtidas pela análise visual ainda não oferecem meios suficientes para explicar as condições da qualidade do ar de São Carlos-SP.

A discussão acima demonstra que as ferramentas visual-analíticas desenvolvidas são capazes de atender as demandas das tarefas T1 e T2. Como já mencionado, o perfil do gráfico que representa o MP_{10} amostrado em 2015, entre o inverno e a primavera (Figura 5.1a), apresentou um padrão mais assemelhado ao encontrado nestas estações em 2014 (Figura 4.2e), sendo esses intervalos ligeiramente diferentes dos dois anos seguintes. É interessante salientar que não foram registradas alterações expressivas nos padrões meteorológicos nesses períodos, ao longo dos anos. Isso sugere a possível incidência de eventos que estariam além dos fatores naturais, característicos do microclima da região. Logo, é fundamental considerar e discutir as diversas fontes antropogênicas de interação com o ambiente, e a relação entre os impactos causados pela ação do homem na emissão de poluentes com o local de pesquisas e todo o seu entorno.

5.2.2 Influências Antropogênicas Sobre os Poluentes

Este estudo de caso foi elaborado em conjunto com os especialistas de domínio, que estavam interessados em verificar os efeitos de longo prazo das ações dos agentes antropogênicos sobre o comportamento dos poluentes. Ampliando a análise anterior sobre as informações representadas na Figura 4.2, observam-se as já mencionadas tendências de aumento nas concentrações de MP_{10} , especialmente entre as estações outono-primavera de 2014 a 2015, outono-primavera de 2015 a 2016 e outono-inverno de 2016 a 2017.

Além das sazonalidades climáticas, essa tendência de crescimento se verificou porque foram registradas, com maior frequência, queimadas de vegetação em geral durante esses períodos, principalmente a queima da palha da cana-de-açúcar (INPE, 2018). A cidade de São Carlos-SP se encontra em uma região circundada por grandes produtores e usinas de beneficiamento de cana-de-açúcar. No entanto, essa tendência de aumento também pode estar relacionada com a realização de diversos eventos públicos nas proximidades do local das coletas, como feiras livres e festivais. Esses eventos acarretaram em aumento no fluxo de pessoas e veículos automotores circulando naquele ambiente, algo que, por consequência, provocou a ressuspensão de solos e um acréscimo na emissão de poluentes provenientes da queima de combustíveis fósseis, por parte dos veículos.

Para proceder a análise, os especialistas utilizaram o *zooming-in* para selecionar apenas os pontos pertencentes aos intervalos de maio de 2015 até maio de 2016 (módulo superior) e de maio de 2014 até maio de 2015 (módulo inferior), apresentados na Figura 5.3. Em ambos os módulos, as mais altas concentrações de MP_{10} foram registradas a partir do inverno, persistindo até meados da primavera, o que coincide com o auge da safra de cana-de-açúcar. De fato, é comum ocorrerem nas regiões próximas à cidade, muitas queimadas neste período que estão diretamente relacionadas ao manejo desta cultura.

Figura 5.3 – Dados coletados para os intervalos de maio de 2015 até maio de 2016 (superior) e de maio de 2014 até maio de 2015 (inferior).

Fonte: Elaborada pelo autor.

Observa-se que há uma tendência geral de diminuição nos índices de concentração do MP_{10} . Essa tendência está evidenciada no gráfico bivariado apresentado no centro da

Figura 5.3. Há uma predominância de picos com o padrão de cor vermelho, que é a cor que caracteriza a curva da série 2014–2015, nesse gráfico bivariado. Esta informação atesta que dentro do intervalo de dados 2014–2015, há maior quantidade de valores que superam (numericamente e ponto a ponto) os valores provenientes do intervalo 2015–2016. De acordo com a Companhia Ambiental do Estado de São Paulo (CETESB), órgão responsável pelas regulamentações dos níveis de poluentes nas cidades do estado, são considerados níveis aceitáveis de MP_{10} até $120 \mu g m^{-3}$ para o período de 24 horas de exposição.

Ainda de acordo com a CETESB, a média anual não deve ultrapassar os $40 \mu g m^{-3}$. Verificou-se que este valor não foi atingido em qualquer um dos anos de monitoramento, sendo que de 2014 a 2015, a média obtida foi de $27,81 \mu g m^{-3}$; de 2015 a 2016, a média registrada foi de $22,27 \mu g m^{-3}$; enquanto de 2016 a 2017, a média obtida foi $20,82 \mu g m^{-3}$. Para a Organização Mundial da Saúde, no que se refere aos padrões de qualidade do ar, a nível internacional, a média aceitável deve ser de $20 \mu g m^{-3}$ (anual) e $50 \mu g m^{-3}$ (diária). Note que no intervalo de 2014 a 2015, foram vários os dias que ultrapassaram a média diária recomendada pela OMS. Já entre 2015 e 2017 foram poucos os dias que ultrapassaram a média diária. Considerando o valor médio anual estabelecido pela OMS, constata-se que a média foi ultrapassada em todos os anos (Figura 5.4).

A Tabela 5.1 sumariza os valores recomendados pela OMS e pela CETESB para a concentração do MP_{10} , de acordo com o intervalo de amostragem. Esses valores guiam as legislações ambientais atuais e referenciam o monitoramento realizado a partir de 2014.

Tabela 5.1 – Valores médios limite para a concentração do MP_{10} atmosférico recomendados pela OMS (internacional) e CETESB (nacional).

	Intervalo de amostragem	Valor médio limite ($\mu g m^{-3}$)
OMS	24 horas	50
	Anual	20
CETESB	24 horas	120
	Anual	40

Fonte: WHO (2006), CETESB (2013).

Em nenhum momento as concentrações diárias ultrapassaram os padrões estabelecidos pelo Conselho Nacional do Meio Ambiente (CONAMA), resolução número 003 de 28/06/1990 (BRASIL, 1990), que é de $150 \mu g m^{-3}$. Segundo a mesma resolução, a média aritmética anual também não deve ultrapassar os $50 \mu g m^{-3}$ (Tabela 5.2). Estes valores não foram atingidos em quaisquer um dos anos de monitoramento. As maiores médias anuais foram de $39,51$ e $33,99 \mu g m^{-3}$ nas temporadas de 1999 a 2000 e de 2000 a 2001, respectivamente. Nas demais temporadas de amostragem as médias aritméticas anuais variaram entre $26,65$ a $33,83 \mu g m^{-3}$, conforme apresenta a Figura 5.4.

Tabela 5.2 – Valores médios limite para a concentração do MP_{10} atmosférico recomendados pelo CONAMA, em vigor durante a campanha de 1997 a 2005.

	Intervalo de amostragem	Valor médio limite ($\mu\text{g m}^{-3}$)
CONAMA	24 horas	150
	Anual	50

Fonte: BRASIL (1990).

Figura 5.4 – Evolução das médias anuais para a concentração do MP_{10} aferidas nas campanhas de 1997–2005 e 2014–2017, face aos limites médios de referência.

Fonte: Elaborada pelo autor.

Os valores aferidos nas coletas de MP_{10} em São Carlos–SP, na campanha de monitoramento de 2014 a 2017, mostram claramente uma diminuição dos índices de MP_{10} , em comparação com a campanha de 1997 a 2005. Afirmação esta que é reforçada quando se verifica a predominância de picos e áreas coloridas em vermelho no gráfico bivariado da Figura 4.5a, que apresenta esta metáfora aplicada aos conjuntos de dados de 2015 a 2017 (módulo superior omitido na referida Figura por questão de espaço) e de 1997 a 2005. Essa tendência de diminuição nas concentrações de MP_{10} pode ter sido motivada, em grande parte, pelo aumento no rigor de aplicação da lei número 997 de 31/05/1976 (SÃO PAULO, 1976) e do decreto 54.487 de 26/06/2009 (SÃO PAULO, 2009), que cobram dos agentes antropogênicos o controle dos poluentes emitidos pelos mesmos.

Além disso, o tema das queimadas relacionadas a cultura da cana-de-açúcar está em pauta e gera inúmeras discussões há muitos anos. Ações recentes, ajuizadas pelo Ministério Público Federal e acatadas pela Justiça Federal, vêm aumentando gradativamente as proibições sobre a queima da palha da cana-de-açúcar no interior do estado de São Paulo, impondo pesadas sanções e multas pecuniárias aos responsáveis pelo descumprimento das determinações judiciais. Logo, essa tendência de diminuição nas concentrações de poluentes atmosféricos, que foi constatada na região de São Carlos–SP, aparenta ter se beneficiado da redução das contribuições oriundas das queimadas da cana-de-açúcar.

O gráfico bivariado, aliado ao *zooming* interativo, apresentaram-se como importantes ferramentas analíticas, permitindo a revelação de tendências dominantes entre as séries e viabilizando, com isso, respostas para as tarefas T1 e T3. Com isso, verificou-se que as séries temporais representadas no módulo inferior da interface dedicada aos dados ambientais, 1997–2005 e 2014–2015, quando contrastadas com os dados da campanha mais recente de monitoramento, 2015–2017, apresentados no módulo superior, demonstram que as concentrações de poluentes sofreram uma redução, com o passar dos anos.

5.2.3 Evolução das Doenças Respiratórias

Neste estudo de caso foram analisados como os índices das doenças respiratórias têm evoluído, bem como os grupos de maior vulnerabilidade a essas enfermidades.

A Figura 5.5 apresenta o perfil geral das principais doenças respiratórias, registradas ao longo dos anos de 1998 a 2006 (Figura 5.5a) e de 2014 a 2017 (Figura 5.5b). Note que, neste cenário, foi omitida a categoria “Outros”. Embora contenha números relevantes, esta categoria é formada pela agregação de vários tipos de doenças que também são ligadas ao aparelho respiratório, porém, com registros individuais esparsos (cf. Subseção 4.2.2). Assim, optou-se por destacar aqui apenas as cinco principais enfermidades que têm a sua tipificação bem definida.

Figura 5.5 – Índices gerais acumulados ao longo dos anos, para as principais doenças relacionadas ao aparelho respiratório.

Fonte: Elaborada pelo autor.

Os gráficos da Figura 5.5 deixam claro que pneumonia foi a doença respiratória que mais acometeu os cidadãos em São Carlos–SP, com bronquite em seguida. O pico de casos de pneumonia ocorreu no mês de maio de 2004 (evidenciado pelo marcador interativo, Figura 5.5a), quando 157 pessoas foram diagnosticadas com pneumonia nas unidades de saúde da cidade. Maio de 2004 também apresentou o pico dos casos de bronquite da série 1998–2006, com 31 ocorrências. É interessante notar a expressividade da asma (Figura 5.5a), porém, em números gerais, esta foi a terceira enfermidade mais frequente.

Para se investigar mais atentamente o perfil de ocorrências de pneumonia no mês de maio de 2004, devem ser selecionadas as opções “Pneumonia” e “2004” nos respectivos *boxes* do módulo inferior. Com isso, a interface será atualizada para apresentar o gráfico da Figura 5.6, que contém barras agrupadas de acordo com as faixas etárias acometidas pela doença, sob a discretização mensal. Observe que em maio, 70 casos de pneumonia foram registrados entre os menores de 5 anos, com outros 59 atendimentos para os maiores de 60 anos. Ou seja, cerca de 82% dos atendimentos médicos realizados em São Carlos–SP, de pacientes com pneumonia, em maio de 2004, estavam concentrados em pessoas pertencentes a apenas duas (entre as oito) faixas etárias.

Figura 5.6 – Evolução dos casos de pneumonia, divididos em grupos etários, ao longo dos meses de 2004.

Fonte: Elaborada pelo autor.

Embora maio seja o mês mais proeminente da Figura 5.6, verifica-se que há um padrão dentro do ano de 2004. As categorias etárias compreendendo os menores de 5 anos (crianças) e os maiores de 60 anos (idosos) ultrapassaram as outras faixas durante todos os meses, com, normalmente, as crianças menores de 5 anos representando o principal grupo de risco à doença, ainda que em alguns meses (fevereiro e agosto) o grupo dos idosos tenha apresentado números superiores.

Essa informação é reforçada quando são analisados os números de pneumonia

acumulados ao longo dos anos. A Figura 5.7 apresenta esses valores. Note que o ano de 2004 (Figura 5.7a) foi aquele em que mais se atendeu casos de pneumonia nas unidades de saúde da cidade, com mais de 900 registros da doença. Novamente há uma clara predominância das crianças com até 5 anos e dos idosos como os maiores grupos de risco, não apenas em 2004, mas em toda a série 1998–2006. Com perfil semelhante a 2004, 2003 foi o segundo ano de maior incidência de pneumonia. Os demais anos da série apresentaram números ligeiramente inferiores, entre os 700 e os 900 casos.

Figura 5.7 – Evolução anual dos grupos etários diagnosticados com pneumonia (*stacked graph*), sob os valores médios do MP_{10} para o mesmo intervalo (gráfico de área).

Fonte: Elaborada pelo autor.

Realizando a mesma verificação sobre os casos acumulados de bronquite, selecionada alterando a doença no respectivo *combo box* do módulo inferior, obtém-se o gráfico da Figura 5.8a. Os grupos de maior vulnerabilidade à bronquite são as crianças menores de 5 anos e os idosos, assim como se verificou no exemplo anterior, para a pneumonia (Figura 5.7a). No entanto, seguindo as informações da Figura 5.8a, há uma predominância dos idosos como o principal grupo de risco à bronquite, na série 1998–2006. Outra vez, o ano de 2004 foi o que apresentou mais registros, considerando todas as categorias etárias.

As Figuras 5.7b e 5.8b apresentam os números dos atendimentos médicos de pneumonia e bronquite, respectivamente, registrados e agregados por ano, pertencentes ao intervalo 2014–2017. Verifica-se, com base nesses dois gráficos, que o padrão de risco permanece semelhante ao encontrado no período 1998–2006, isto é, crianças de até 5 anos e idosos continuam sendo os grupos de maior vulnerabilidade às doenças em questão.

Figura 5.8 – Evolução anual dos grupos etários diagnosticados com bronquite (*stacked graph*), sob os valores médios do MP_{10} para o mesmo intervalo (gráfico de área).

Fonte: Elaborada pelo autor.

Deve ser ressaltada a importância desse tipo de análise, que pode contribuir com o direcionamento das políticas de gestão da saúde pública. Dados coletados ao longo de vários anos foram sumarizados e aplicados em um painel visual interativo, que apresenta de modo claro e direto a evolução dos atendimentos médicos na cidade, revelando os grupos mais sensíveis às doenças respiratórias, bem como as suas distribuições detalhadas. Tendo em mãos as informações geradas pelo sistema proposto, as entidades responsáveis podem realocar seus recursos, concentrando esforços nos grupos e períodos mais críticos.

Os especialistas de domínio também estavam particularmente interessados em analisar a dinâmica temporal desses dados como um todo. Ou seja, verificar como os padrões dos registros de doenças respiratórias têm evoluído ao longo dos anos até os dias atuais. Iniciando a análise a partir de uma visão geral dos dados, apresentada pelos gráficos auxiliares das Figuras 5.5a e 5.5b, comparam-se os números gerais de cada variável (doença) nas séries 1998–2006 e 2014–2017, respectivamente. Sem a necessidade de muitos esforços analíticos, é possível verificar que houve uma sensível redução nas curvas de todos os gráficos entre as Figuras 5.5a e 5.5b. Uma ressalva deve ser feita quanto a laringite, sobre a qual não foram encontrados registros para o intervalo 2014–2017.

Em números, pneumonia, bronquite e asma permaneceram como as doenças mais expressivas na série 2014–2017. O pico de ocorrências de pneumonia foi registrado em junho de 2017 (Figura 5.5b), com 93 casos, face aos 157 registrados em maio de 2004 (1998–2006).

O período com mais ocorrências de bronquite também foi em junho de 2017, com 33 registros, frente aos 31 casos de maio de 2004. Neste caso houve um aumento, porém, o perfil geral dos registros de bronquite da série 2014–2017 apresentou uma redução em relação a 1998–2006. Uma mudança significativa é vista na variável asma, que apresentou um perfil geral bastante atenuado no período 2014–2017, comparado com 1998–2006.

Centralizando a análise comparativa sobre a evolução das duas principais variáveis, pneumonia e bronquite, observam-se que, na Figura 5.7b, os números acumulados anualmente de pneumonia são inferiores aos apresentados pela Figura 5.7a. A oscilação de casos de pneumonia em 2014–2017 esteve entre 400 e 750 registros, frente ao verificado na série 1998–2006, que parte de um patamar em torno de 750 registros no ano com menor quantidade de ocorrências (2005). O cenário apresentado pela Figura 5.8b é semelhante aos casos de bronquite, que variaram entre 70 e 160 ocorrências dentro dos anos 2014–2017, enquanto no intervalo 1998–2006, registrou-se entre 140 e 200 casos. Mesmo em um cenário com os dois últimos anos do conjunto 2014–2017 apresentando números acumulados superiores aos vistos nos dois anos anteriores (Figura 5.8b).

Uma informação interessante que pode ser observada comparando os gráficos das Figuras 5.7 e 5.8, vem das metáforas auxiliares posicionadas acima dos *stacked graphs*. Essas metáforas de área complementam a visualização e apresentam a variação da concentração média mensal do MP_{10} , dentro do mesmo período destacado pelo *stacked graph* sobre o qual estão conectadas. Entre 1998 e 2006, a média mensal dos poluentes amostrados foi de até $80 \mu g m^{-3}$, permanecendo abaixo dos $50 \mu g m^{-3}$ no período de 2014 a 2017. Logo, houve uma redução nas concentrações de MP_{10} , informação esta que já era conhecida, de acordo com o apresentado nos estudos de caso anteriores. A nova informação, que foi revelada neste estudo de caso, mostra que também houve uma clara diminuição no número de registros de doenças respiratórias no decorrer dos anos.

Embora as informações apresentadas em ambos os sistemas de visualização tenham confirmado que houve uma tendência de melhora na qualidade do ar em São Carlos–SP, devido às menores quantidades de partículas suspensas, assim como também houve uma tendência de redução nos registros das doenças relacionadas ao aparelho respiratório, não há como defender, com certeza, que uma hipotética correlação entre esses dados seja verdadeira, apenas com base em informação visual. As análises visuais empreendidas aqui levantaram algumas evidências a respeito dos padrões e tendências contidos nos dados ambientais e nos índices de saúde, ajudando a responder as tarefas analíticas T3 e T4.

Já a tarefa T5 foi parcialmente respondida, uma vez que, sim, as ocorrências das doenças respiratórias registradas nos últimos anos têm acompanhado a tendência geral de redução verificada nas concentrações dos poluentes, no mesmo período. No entanto, para

afirmar categoricamente que essas informações não são apenas coincidentes, sendo elas, na verdade, correlatas, há de se realizar mais pesquisas e testes estatísticos específicos e com essa finalidade, algo que vai além do escopo desta dissertação.

Para serem representados nesta interface, foram selecionados subconjuntos de séries temporais contendo dados de atendimentos médicos de doenças respiratórias, pois são essas as enfermidades comumente relacionadas à poluição atmosférica, que é o interesse fundamental dos especialistas de domínio associados a este trabalho. Como foi citado na Subseção 4.2.2, o Sistema de Informações Hospitalares do SUS coleta uma ampla variedade de dados das unidades de saúde de todo o Brasil. Esta interface está apta a receber as variáveis de outras categorias disponibilizadas pela plataforma de acesso à informação do SUS, dependendo do interesse analítico dos especialistas que utilizarão a ferramenta.

5.3 Discussões

Basicamente, os *layouts* dos sistemas desenvolvidos e apresentados aqui, seguem uma configuração com dois módulos gráficos semelhantes, um posicionado acima do outro. Por meio desse arranjo, o analista seleciona interativamente trechos de duas séries temporais distintas para fazer a comparação visual direta entre elas, “sobrepondo-as”.

Especificamente, optou-se por representar no módulo superior da interface dos dados ambientais, as informações coletadas seguindo o padrão atualmente adotado pelos pesquisadores responsáveis pelo monitoramento do ar em São Carlos–SP, baseado nas referências metodológicas do ISO 13271 (2012) (cf. Subseção 4.2.1). A partir disso, o analista pode selecionar no módulo inferior, os dados amostrados nas campanhas passadas, para então comparar os padrões e tendências “atuais” com os “históricos”. Lembrando que a interface dedicada aos índices de saúde segue o modelo de disposição de elementos gráficos do sistema dedicado aos dados ambientais, explorando, com isso, a familiaridade dos usuários durante a utilização das ferramentas.

Outros *layout* foram sugeridos e considerados pelos especialistas de domínio. Entre as opções estava exibir em ambos os módulos gráficos um único conjunto de séries temporais, formado pela união dos dados contidos em todos os períodos de monitoramento, 1997–2006, 2014–2015 e 2015–2017 (atual). Então, o analista delimitaria os trechos de seu interesse para comparação, em ambos os módulos, por meio do *zoom*, no caso da interface dedicada aos dados ambientais. Embora essa alternativa pudesse conferir um nível acima no quesito interatividade, ao eliminar a necessidade do *combo box* para a seleção dos conjuntos de dados, deve ser observado que há um hiato expressivo, de quase 10 anos, entre a campanha histórica (1997–2006) e as campanhas recentes (2014–2015 e 2015–2017), quando não

foram realizados projetos de monitoramento do ar pelos pesquisadores responsáveis. Em testes realizados com essa abordagem, houve a formação de duas aglomerações de pontos de dado, separados por um grande vazio de informações, algo que resultou em um *layout* com aspecto visual pouco agradável.

Avaliar a eficiência e a efetividade das ferramentas de visualização é algo complexo. Embora as boas práticas de *design* e desenvolvimento, originadas em décadas de pesquisas comportamentais, sejam guias indispensáveis para explorar e tirar proveito das capacidades cognitivas dos usuários, essas teorias não garantem que as informações transmitidas sejam realmente compreendidas por seus espectadores. Os estudos de caso também são métodos úteis para demonstrar as funcionalidades dos sistemas visuais, entretanto, eles não permitem respostas objetivas para questões como “quanto a interface possibilitou ao usuário desempenhar melhor o seu trabalho?”. Quantificações do ganho de conhecimento e do fluxo cognitivo poderiam ser empregadas como meios de validação, mas são processos subjetivos e difíceis de serem avaliados. Assim, métricas padronizadas para testar e avaliar a efetividade de sistemas de representação de informações ainda são necessárias na área de visualização (PIKE *et al.*, 2009).

Porém, após aplicarem as ferramentas gráficas em seus procedimentos de análise, os especialistas de domínio retornaram comentários positivos. Esses analistas, engenheiros químicos do Departamento de Engenharia Química da UFSCar, classificaram as interfaces como ambientes “*práticos e de fácil entendimento*”. Ressaltaram a importância da interatividade na apresentação de informações detalhadas (valores, datas), simplesmente passando o *mouse* sobre os elementos gráficos, com esse mecanismo sendo considerado “*muito bom, principalmente para que o usuário comum possa acompanhar e compreender com mais facilidade o que está acontecendo com a qualidade do ar em sua cidade*”. Outro ponto destacado pelos especialistas foi o gráfico bivariado, que integra a interface dedicada aos dados ambientais (cf. Figura 4.2c), porque “*mostra nitidamente a diferença entre os dados quando são comparados dois intervalos, ou seja, os pontos em que estão registradas as maiores ou menores concentrações de MP_{10} , e isso é valioso*”.

Ambas as interfaces têm seus códigos de desenvolvimento abertos, sendo acessados diretamente pelo navegador, teclando *Ctrl+U* (comando que pode variar, dependendo do navegador). Vale lembrar que esses sistemas de visualização também são acessados por meio dos navegadores web de dispositivos móveis, como *smartphones* e *tablets*. Todas as funcionalidades descritas aqui são suportadas e operadas a partir desses dispositivos, de modo semelhante ao exemplo ilustrado pela Figura 2.8. Entretanto, o tamanho reduzido de suas telas e a utilização dos dedos, pelo usuário, podem gerar uma certa imprecisão na seleção e análise visual das informações.

Todas as informações exibidas pelas ferramentas de análise visual são processadas sob demanda e representadas em um web *browser*, com livre acesso e sem a necessidade de instalação de *softwares* específicos, como os sistemas de gerenciamento de banco de dados (SGBD). Essa característica amplia o alcance das ferramentas, visto que a sua utilização estará condicionada apenas ao acesso à Internet.

5.4 Considerações Finais

Neste Capítulo foram apresentadas as funcionalidades analíticas das ferramentas desenvolvidas, com o objetivo de explorar séries ambientais e índices de saúde amostrados em São Carlos–SP e buscar respostas às tarefas analíticas derivadas das demandas dos especialistas em domínio. As discussões acima mostraram que essas tarefas foram cumpridas de modo intuitivo. A flexibilidade oferecida pela combinação de métodos de seleção e comparação interativa entre conjuntos de séries temporais, facilita ao analista manusear e apurar blocos de informações destacadas em diferentes períodos. Com isso, um importante aspecto sobre a dinâmica temporal desses dados é evidenciado, a sazonalidade, permitindo ao usuário compreender padrões e mudanças nas variáveis monitoradas.

Ao permitir que o usuário tenha o controle do conteúdo do sistema de visualização, podendo manualmente alguns padrões dominantes, podem ser reveladas algumas características interessantes nos dados, ainda que secundárias (LIN *et al.*, 2004). Assim, foi adotada como estratégia de exploração das interfaces construídas aqui, oferecer uma visão geral dos dados para que o usuário analise os detalhes das informações de seu interesse sob demanda, por meio de seleções como o zoom, por exemplo, que tornam a navegação um processo mais ágil e focado (SHNEIDERMAN, 1996).

Capítulo 6

Conclusão

6.1 Limitações e Trabalhos Futuros

As técnicas analíticas baseadas em exploração interativa possibilitam ao usuário identificar com mais facilidade as informações de seu interesse como, neste exemplo, as características dinâmicas e sazonais da dispersão dos poluentes em ambientes urbanos, as tendências históricas de aumento ou diminuição do MP_{10} , ou mesmo quais os dias que ultrapassaram os valores máximos de concentração de partículas, com índices definidos por entidades como a OMS (WHO, 2006), além de órgãos nacionais de pesquisas e proteção ambiental. Porém, existem algumas limitações e possibilidades de extensão em trabalhos futuros que sejam relacionados a esta temática.

Uma limitação que foi levantada diz respeito às falhas que o equipamento de coleta apresentou. A princípio, por serem eventos raros, a decisão tomada foi remover todo o conteúdo do respectivo evento falho de seu respectivo conjunto. Como trabalho futuro, elenca-se o desenvolvimento e aplicação de alguns mecanismos de tolerância a essas falhas. Entre esses mecanismos, considera-se a introdução de técnicas de inteligência computacional, especialmente da classe dos algoritmos de aprendizado de máquina.

A partir dos dados históricos registrados, são determinadas funções de mapeamento entre os dados passados e os dados futuros, de tal modo que se torna possível utilizar a informação conhecida para estimar a informação desconhecida. Com isso, é possível preencher as lacunas formadas pelos dados faltantes, com as saídas geradas pelos métodos de predição. As redes neurais artificiais e suas variantes, como o *deep learning* (KORCZAK; HERNES, 2017), estão em estudo por apresentarem bons resultados preditivos e uma relativa flexibilidade de implementação. Entretanto, o conhecimento obtido pelo modelo fica embutido na ponderação dos nós da rede, dificultando a sua explicação.

Nesta pesquisa, o foco dos sistemas visual analíticos desenvolvidos foi verificar a dinâmica temporal dos dados amostrados ao longo de anos de estudos. Como descrito na Seção 4.2, as bases utilizadas são ricas no que diz respeito a situação ambiental de São Carlos-SP, sendo de grande relevância oferecer aos analistas um ferramental que viabilize a exploração e a correlação adequada dessas informações com a variação do MP₁₀.

Esse objetivo foi cumprido, com a representação dos valores de precipitação, temperatura, velocidade do vento e umidade relativa em *small multiples* interativos, na interface dedicada aos dados ambientais. No entanto, a direção do vento é exibida apenas pontualmente nos *tooltips*, quando o usuário inspeciona um ponto de seu interesse em algum dos *scatter plots*. Pretende-se desenvolver meios para que o usuário também acompanhe a evolução desse dado. Isso introduz alguns complicadores quanto ao *design* e a usabilidade do sistema, uma vez que adiciona mais informações dentro do mesmo espaço visual. Além disso, a direção do vento é expressa em graus, uma escala com interpretação relativamente complexa que não é comumente expressa em gráficos temporais.

Disponibilizar as informações amostradas em tempo real, ou seja, no instante em que os dados forem analisados em laboratório e registrados, poderá ser desenvolvido futuramente. Isso envolve a necessidade de atualizar as séries *online*, a cada análise, enquanto a interface seria automaticamente atualizada assim que a nova entrada de dado fosse incluída em seu respectivo conjunto. No curto prazo, essa modificação não traria algum tipo de problema. Porém, quando se considera o longo prazo, haverá a necessidade de lidar com um grande volume (crescente) de pontos que, em algum momento, poderão ultrapassar a quantidade horizontal de pixels da tela. Essa consequência, natural quando se trata de dados dinâmicos e crescentes, como as séries temporais, gera níveis consideráveis de oclusão e desordem visual, que devem ser levados em consideração no projeto de sistemas de visualização de informações para que sejam adequadamente mitigados. Oferecer diferentes níveis de agregação temporal pode ser uma estratégia particularmente interessante nesse contexto (AIGNER *et al.*, 2007).

Também é elencado como possível trabalho futuro, oferecer a opção para que o próprio analista faça o envio (*upload*) de seus conjuntos de séries, tornando as ferramentas ainda mais flexíveis na análise visual, embora, para isso, seja necessário desenvolver mecanismos para tratamento e adequação dos dados, que considerem os diversos padrões existentes. Ainda há espaço para o aprimoramento dos métodos interativos das interfaces, tal como uma maior interligação entre os eventos contextuais de *mouse*, que seriam ativados em todos os gráficos que simultaneamente representam as partes do mesmo ponto de dado. Com isso, os marcadores e *tooltips* apresentariam suas informações ao mesmo tempo, não importando em qual parte da representação do ponto o usuário paire o cursor.

Outro item que ainda poderá ser arquitetado é uma integração/comunicação entre a interface dos dados ambientais e a interface dos índices de saúde, transformando ambas as ferramentas em um único sistema visual analítico. Devido à grande diversidade encontrada entre os conjuntos de dados tratados por ambas as interfaces, essa será uma tarefa complexa que demandará mais pesquisas em projetos de *design* para que haja o adequado equilíbrio na apresentação das informações, que deve ser compacta e significativa.

6.2 Considerações Finais

A análise temporal busca estender os métodos de análise e mineração de dados convencionais, incorporando em suas tarefas o reconhecimento de características ligadas à variação do tempo (POST; HARRISON, 2008). O campo da exploração sobre as séries temporais é vasto, sendo composto por muitos problemas específicos que devem ser investigados por técnicas que se encaixem em seus domínios de aplicação. Logo, a necessidade da descoberta de padrões em grandes bases de séries temporais multidimensionais exige a construção de ferramentas voltadas a representação e análise cada vez mais sofisticadas, que proporcionem uma navegação interativa entre os diversos objetos, facilitando a interpretação de seus relacionamentos e de sua distribuição ao longo tempo.

Nesta dissertação foi introduzida uma nova metodologia para a exploração de dados, baseada na composição de metáforas e ferramentas visuais interativas que simplificam a interpretação e auxiliam a revelar características, tendências e padrões em séries temporais multidimensionais relacionadas ao monitoramento da qualidade do ar. Diferentes estratégias foram combinadas para criar sistemas gráficos que transformam milhares de pontos abstratos de dado em informação visual, guiando o usuário na investigação de dados que foram coletadas ao longo de vários anos.

Ao representar de modo compacto e sucinto as múltiplas variáveis de cada instância de dado, o usuário desenvolve uma navegação objetiva por características de seu próprio interesse. Mesmo com o centro das análises empreendidas nesta dissertação girando em torno de informações sobre os poluentes coletados na região central da cidade de São Carlos-SP, as técnicas para exploração de séries temporais desenvolvidas aqui são flexíveis o suficiente para serem reutilizadas na representação de dados de diferentes fontes, desde que feitas as devidas adequações de contexto. Entretanto, acredita-se que esse tipo de informação ambiental seja particularmente interessante e socialmente relevante.

A análise do comportamento do material particulado revelou que, ao longo das últimas décadas, houve uma melhora na qualidade do ar na cidade. A agroindústria, um dos grandes contribuintes para a poluição atmosférica, enfrentou e ainda enfrenta

restrições em certas práticas, anteriormente aceitas em seu processo produtivo, como as queimadas, por exemplo. Todavia, não se pode deixar de investigar constantemente as concentrações do material particulado disperso no ar, para que seja possível identificar com rapidez os períodos críticos, as entidades emissoras, os impactos causados e, então, alertar as autoridades responsáveis pelo devido controle ambiental para que estas exerçam as suas funções e assegurem a qualidade do ar que os cidadãos respiram.

Como foi demonstrado nos estudos de caso, as interfaces permitem uma rápida identificação e destaque dos períodos de maior concentração de poluentes, com a comparação destes dados com as informações climáticas do respectivo período destacado; além de identificar quais os grupos de pessoas com maior sensibilidade às doenças respiratórias. Isso possibilita aos analistas levantar hipóteses sobre as prováveis causas e efeitos das variações observadas, servindo como um ponto de partida para direcionar estudos mais aprofundados nessa complexa área de pesquisas que envolve o meio ambiente e a qualidade do ar em ambientes externos e urbanos.

A falta de ferramentas adequadas, que revelem a progressão temporal simultânea de múltiplos atributos, dificultam a evolução das pesquisas que buscam o entendimento dos resultados obtidos em análises amostrais. Assim, por serem ferramentas abertas e acessíveis em uma variada gama de dispositivos, os sistemas desenvolvidos aqui se mostram como uma alternativa no monitoramento urbano. A Internet é um poderoso instrumento de acesso, que amplia consideravelmente o alcance do público em geral à informação. Portanto, a integração entre sistemas de visualização de séries temporais multidimensionais que sejam de fato acessíveis, com técnicas que possibilitem a reorganização desses dados conforme são explorados pelos analistas, contribui com um conjunto fundamental de ferramentas dedicadas à análise visual de informações.

Os recursos tecnológicos que são viabilizados pela computação, buscam soluções para problemas teóricos e práticos, não necessariamente relacionados ao ambiente computacional. Por meio de algoritmos projetados com o objetivo de automatizar processos, transforma-se o manual em automático, mais eficiente e menos custoso.

Referências Bibliográficas

ADNAN, M.; JUST, M.; BAILLIE, L. Investigating time series visualisations to improve the user experience. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2016. (CHI '16), p. 5444–5455.

AIGNER, W.; MIKSCH, S.; MÜLLER, W.; SCHUMANN, H.; TOMINSKI, C. Visualizing time-oriented data—a systematic view. *Computers & Graphics*, Pergamon Press, Inc., Elmsford, NY, USA, v. 31, n. 3, p. 401–409, jun. 2007. ISSN 0097-8493.

_____. Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 14, n. 1, p. 47–60, jan. 2008. ISSN 1077-2626.

AIGNER, W.; MIKSCH, S.; SCHUMANN, H.; TOMINSKI, C. *Visualization of Time-Oriented Data*. 1st. ed. New York, NY, USA: Springer Publishing Company, Incorporated, 2011.

AIGNER, W.; RIND, A.; HOFFMANN, S. Comparative evaluation of an interactive time-series visualization that combines quantitative data with qualitative abstractions. *Computer Graphics Forum*, The Eurographs Association and John Wiley & Sons Ltd., Chichester, UK, v. 31, n. 3pt2, p. 995–1004, jun. 2012. ISSN 0167-7055.

ALBERS, D.; CORRELL, M.; GLEICHER, M. Task-driven evaluation of aggregation in time series visualization. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2014. (CHI '14), p. 551–560.

ALEXANDRINA, E. C. Caracterização e composição química do material particulado grosso (MP₁₀) no centro da cidade de São Carlos (SP). In: *Dissertação (Mestrado)*. São Carlos: Universidade Federal de São Carlos, UFSCar, 2015. p. 140 f.

ANDREWS, K.; KIENREICH, W.; SABOL, V.; BECKER, J.; DROSCHL, G.; KAPPE, F.; GRANITZER, M.; AUER, P.; TOCHTERMANN, K. The InfoSky visual explorer: Exploiting hierarchical structure and document similarities. *Information Visualization*, Palgrave Macmillan, v. 1, n. 3/4, p. 166–181, dez. 2002. ISSN 1473-8716.

ANDRIENKO, G.; ANDRIENKO, N.; DEMSAR, U.; DRANSCH, D.; DYKES, J.; FABRIKANT, S. I.; JERN, M.; KRAAK, M.-J.; SCHUMANN, H.; TOMINSKI, C. Space, time and visual analytics. *International Journal of Geographical Information Science*, Taylor & Francis, Inc., Bristol, PA, USA, v. 24, n. 10, p. 1577–1600, out. 2010. ISSN 1365-8816.

ANDRIENKO, N.; ANDRIENKO, G. Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, Palgrave Macmillan, v. 12, n. 1, p. 3–24, jan. 2013. ISSN 1473-8716.

- ANDRIENKO, N.; LAMMARSCH, T.; ANDRIENKO, G.; FUCHS, G.; KEIM, D. A.; MIKSCH, S.; RIND, A. Viewing visual analytics as model building. *Computer Graphics Forum*, The Eurographs Association and John Wiley & Sons, Ltd., Chichester, UK, Early View, n. 0, p. 1–25, 2018.
- AUSTIN, P. B. *1812: The Great Retreat*. 1st. ed. London, UK: Greenhill Books, 1996.
- BABBIE, E. *The Basics of Social Research*. 5th. ed. Belmont, CA, USA: Wadsworth Publishing Group, 2011.
- BACH, B.; DRAGICEVIC, P.; ARCHAMBAULT, D.; HURTER, C.; CARPENDALE, S. A review of temporal data visualizations based on space-time cube operations. In: BORGO, R.; MACIEJEWSKI, R.; VIOLA, I. (Ed.). *STAR – State of The Art Report, Eurographics Conference on Visualization (EuroVis)*. Swansea, Wales, United Kingdom: Eurographics Association, 2014. p. 23–41.
- _____. A descriptive framework for temporal data visualizations based on generalized space-time cubes. *Computer Graphics Forum*, The Eurographs Association and John Wiley & Sons, Ltd., Chichester, UK, v. 36, n. 6, p. 36–61, set. 2017. ISSN 0167-7055.
- BADAM, S. K.; ZHAO, J.; SEN, S.; ELMQVIST, N.; EBERT, D. TimeFork: Interactive prediction of time series. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2016. (CHI '16), p. 5409–5420.
- BALDONADO, M. Q. W.; WOODRUFF, A.; KUCHINSKY, A. Guidelines for using multiple views in information visualization. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. New York, NY, USA: ACM, 2000. (AVI '00), p. 110–119.
- BARTOLOMEO, M. D.; HU, Y. There is more to streamgraphs than movies: Better aesthetics via ordering and lassoing. In: *Proceedings of the Eurographics / IEEE VGTC Conference on Visualization*. Goslar Germany, Germany: Eurographics Association, 2016. (EuroVis '16), p. 341–350.
- BECKER, R. A.; CLEVELAND, W. S. Brushing scatterplots. *Technometrics*, v. 29, n. 2, p. 127–142, May 1987.
- BLOCH, M.; BYRON, L.; CARTER, S.; COX, A. *The Ebb and Flow of Movies: Box Office Receipts 1986-2007*. New York, NY, USA: The New York Times Company, 2008. <http://www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html>. Online; acessado em 3 de março de 2018.
- BOSTOCK, M.; HEER, J. Protovis: A graphical toolkit for visualization. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 15, n. 6, p. 1121–1128, nov. 2009. ISSN 1077-2626.
- BOSTOCK, M.; OGIEVETSKY, V.; HEER, J. D3: Data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 17, n. 12, p. 2301–2309, dez. 2011. ISSN 1077-2626.
- BOYANDIN, I.; BERTINI, E.; BAK, P.; LALANNE, D. Flowstrates: An approach for visual exploration of temporal origin-destination data. In: *Proceedings of the 13th Eurographics / IEEE - VGTC Conference on Visualization*. Chichester, UK: The Eurographs Association and John Wiley & Sons, Ltd., 2011. (EuroVis '11), p. 971–980.

- BRASIL. Resolução n. 003 de 28 de junho de 1990. Dispõe sobre padrões de qualidade do ar, previstos no PRONAR. Ministério do Meio Ambiente, Conselho Nacional do Meio Ambiente (CONAMA), 1990. <<http://www.mma.gov.br/port/conama/res/res90/res0390.html>>. Online; acessado em 3 de março de 2018.
- BROOK, R. D. Cardiovascular effects of air pollution. *Clinical Science*, Portland Press Limited, London, UK, v. 115, n. 6, p. 175–187, 2008. ISSN 0143-5221.
- BURCH, M.; BECK, F.; DIEHL, S. Timeline trees: Visualizing sequences of transactions in information hierarchies. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. New York, NY, USA: ACM, 2008. (AVI '08), p. 75–82.
- BURCH, M.; WEISKOPF, D. Visualizing dynamic quantitative data in hierarchies—TimeEdgeTrees: Attaching dynamic weights to tree edges. In: *Proceedings of International Conference on Visualization Theory and Applications*. New York, NY, USA: Springer, 2011. p. 177–186.
- BYRON, L.; WATTENBERG, M. Stacked graphs – geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 14, n. 6, p. 1245–1252, nov. 2008. ISSN 1077-2626.
- CAO, N.; LIN, Y. R.; SUN, X.; LAZER, D.; LIU, S.; QU, H. Whisper: Tracing the spatio-temporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Computer Society, Washington, DC, USA, v. 18, n. 12, p. 2649–2658, December 2012. ISSN 1077-2626.
- CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. (Ed.). *Readings in Information Visualization: Using Vision to Think*. 1st. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999.
- CETESB. *Companhia Ambiental do Estado de São Paulo*. São Paulo – Secretaria do Meio Ambiente do Estado de São Paulo (SMA), 2013. <<https://cetesb.sp.gov.br/ar/padroes-de-qualidade-do-ar/>>. Online; acessado em 3 de março de 2018.
- CHALMERS, M. A linear iteration time layout algorithm for visualising high-dimensional data. In: *Proceedings of the 7th Conference on Visualization '96*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1996. (VIS '96), p. 127–ff.
- CHANG, D.; NESBITT, K. V.; WILKINS, K. The Gestalt principles of similarity and proximity apply to both the haptic and visual grouping of elements. In: *Proceedings of the Eight Australasian Conference on User Interface - Volume 64*. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2007. (AUIC '07), p. 79–86.
- CHANG, R.; GHONIEM, M.; KOSARA, R.; RIBARSKY, W.; YANG, J.; SUMA, E.; ZIEMKIEWICZ, C.; KERN, D.; SUDJANTO, A. WireVis: Visualization of categorical, time-varying data from financial transactions. In: *Proceedings of the 2007 IEEE Symposium on Visual Analytics Science and Technology*. Washington, DC, USA: IEEE Computer Society, 2007. (VAST '07), p. 155–162.
- CHEN, S.; LIN, L.; YUAN, X. Social media visual analytics. *Computer Graphics Forum*, The Eurographs Association and John Wiley & Sons, Ltd., Chichester, UK, v. 36, n. 3, p. 563–587, jun. 2017. ISSN 0167-7055.

CHO, M.; KIM, B.; BAE, H.-J.; SEO, J. Stroscope: Multi-scale visualization of irregularly measured time-series data. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 20, n. 5, p. 808–821, maio 2014. ISSN 1077-2626.

CLEVELAND, W. S.; MCGILL, R. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, American Statistical Association, v. 79, n. 387, p. 531–554, 1984.

CRAFT, B.; CAIRNS, P. Beyond guidelines: What can we learn from the visual information seeking mantra? In: *Proceedings of the Ninth International Conference on Information Visualisation*. Washington, DC, USA: IEEE Computer Society, 2005. (IV '05), p. 110–118.

CROCKFORD, D. *JavaScript: The Good Parts*. Sebastopol, CA, EUA: O'Reilly Media, Inc., 2008.

CUI, Q.; WARD, M.; RUNDENSTEINER, E.; YANG, J. Measuring data abstraction quality in multiresolution visualizations. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 12, n. 5, p. 709–716, set. 2006. ISSN 1077-2626.

CUI, W.; LIU, S.; WU, Z.; WEI, H. How hierarchical topics evolve in large text corpora. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Computer Society, Washington, DC, USA, v. 20, n. 12, p. 2281–2290, dez. 2014. ISSN 1077-2626.

DANG, T. N.; PENDAR, N.; FORBES, A. G. TimeArcs: Visualizing fluctuations in dynamic networks. In: *Proceedings of the Eurographics / IEEE VGTC Conference on Visualization*. Goslar Germany, Germany: Eurographics Association, 2016. (EuroVis '16), p. 61–69.

DER, G.; EVERITT, B. *A Handbook of Statistical Graphics Using SAS ODS*. 1st. ed. Boca Raton, FL, USA: CRC Press, A Chapman & Hall Book, Taylor & Francis, 2014.

DIMENSTEIN, G.; CORTELLA, M. S. *A Era da Curadoria: O que importa é saber o que importa!* 1st. ed. Campinas, SP: Papirus 7 Mares, 2015.

DOU, W.; YU, L.; WANG, X.; MA, Z.; RIBARSKY, W. HierarchicalTopics: Visually exploring large text collections using topic hierarchies. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 19, n. 12, p. 2002–2011, dez. 2013. ISSN 1077-2626.

DU, Y.; XU, X.; CHU, M.; GUO, Y.; WANG, J. Air particulate matter and cardiovascular disease: the epidemiological, biomedical and clinical evidence. *Journal of Thoracic Disease*, AME Publishing Company, Hong Kong, v. 8, n. 1, p. E8–E19, 2016.

EICK, S. G.; KARR, A. F. Visual scalability. *Journal of Computational and Graphical Statistics*, Taylor & Francis, v. 11, n. 1, p. 22–43, 2002.

ELMQVIST, N.; MOERE, A. V.; JETTER, H.-C.; CERNEA, D.; REITERER, H.; JANKUN-KELLY, T. J. Fluid interaction for information visualization. *Information Visualization*, Palgrave Macmillan, v. 10, n. 4, p. 327–340, out. 2011. ISSN 1473-8716.

ENERGETICA. *Amostrador de Grande Volume (AGV) para Partículas de até 10 μm (MP_{10})*. ENERGÉTICA IND. E COM. LTDA., 2016. <http://www.energetica.ind.br/wp/env1/wp-content/uploads/2016/01/env1_manual-mp10_rev_11.pdf>. Online; acessado em 3 de dezembro de 2018.

EPA. *Environmental Protection Agency, AirData*. United States of America, 2015. <<http://www3.epa.gov/airdata/>>. Online; acessado em 10 de agosto de 2018.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. Advances in knowledge discovery and data mining. In: FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R. (Ed.). Menlo Park, CA, USA: American Association for Artificial Intelligence, 1996. cap. From Data Mining to Knowledge Discovery: An Overview, p. 1–34.

FEDERICO, P.; HOFFMANN, S.; RIND, A.; AIGNER, W.; MIKSCH, S. Qualizon graphs: Space-efficient time-series visualization with qualitative abstractions. In: *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. New York, NY, USA: ACM, 2014. (AVI '14), p. 273–280.

FU, T.-c. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, Pergamon Press, Inc., Tarrytown, NY, USA, v. 24, n. 1, p. 164–181, fev. 2011. ISSN 0952-1976.

FU, T.-c.; CHUNG, F.-l.; LAM, C.-f.; LUK, R.; NG, C.-m. Adaptive data delivery framework for financial time series visualization. In: *Proceedings of the 2005 International Conference on Mobile Business*. Washington, DC, USA: IEEE Computer Society, 2005. (ICMB '05), p. 267–273.

FUA, Y.-H.; WARD, M. O.; RUNDENSTEINER, E. A. Hierarchical parallel coordinates for exploration of large datasets. In: *Proceedings of the Conference on Visualization '99: Celebrating Ten Years*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1999. (VIS '99), p. 43–50. ISBN 0-7803-5897-X.

GOETHEM, A. V.; STAALS, F.; LOFFLER, M.; DYKES, J.; SPECKMANN, B. Multi-granular trend detection for time-series analysis. *IEEE Transactions on Visualization and Computer Graphics*, Institute of Electrical and Electronics Engineers, v. 23, n. 1, p. 661–670, 1 2017. ISSN 1077-2626.

GOMI, A.; ITOH, T. MIAOW: A 3D image browser applying a location- and time-based hierarchical data visualization technique. In: *Proceedings of the International Conference on Advanced Visual Interfaces*. New York, NY, USA: ACM, 2010. (AVI '10), p. 225–232.

GREEN, T. M.; RIBARSKY, W.; FISHER, B. Building and applying a human cognition model for visual analytics. *Information Visualization*, Palgrave Macmillan, v. 8, n. 1, p. 1–13, jan. 2009. ISSN 1473-8716.

GRINSTEIN, G.; TRUTSCHL, M.; CVEK, U. High-dimensional visualizations. In: *Proceedings of the 7th Data Mining Conference*. San Francisco, California, USA: ACM, 2001. (KDD Workshop), p. 7–19.

GRUENDL, H.; RIEHMANN, P.; PAUSCH, Y.; FROEHLICH, B. Time-series plots integrated in parallel-coordinates displays. In: *Proceedings of the Eurographics / IEEE VGTC Conference on Visualization*. Goslar Germany, Germany: Eurographics Association, 2016. (EuroVis '16), p. 321–330.

- HALES, W. *HTML5 and JavaScript Web Apps*. Sebastopol, CA, EUA: O'Reilly Media, Inc., 2012.
- HAO, M.; DAYAL, U.; KEIM, D.; SCHRECK, T. Multi-resolution techniques for visual exploration of large time-series data. In: *Proceedings of the 9th Joint Eurographics / IEEE VGTC Conference on Visualization*. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2007. (EUROVIS'07), p. 27–34.
- HAO, M. C.; DAYAL, U.; KEIM, D. A.; SCHRECK, T. Importance-driven visualization layouts for large time series data. In: *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*. Washington, DC, USA: IEEE Computer Society, 2005. (INFOVIS '05), p. 27–.
- HARRISON, S. A.; TONG, F. Decoding reveals the contents of visual working memory in early visual areas. *Nature Neuroscience*, Nature Publishing Group, v. 458, p. 632–635, 2009.
- HARROWER, M. A.; BREWER, C. A. ColorBrewer.org: An online tool for selecting color schemes for maps. *The Cartographic Journal*, Taylor & Francis, v. 40, n. 1, p. 27–37, 2003.
- HAVRE, S.; HETZLER, B.; NOWELL, L. ThemeRiver: Visualizing theme changes over time. In: *Proceedings of the IEEE Symposium on Information Visualization*. Washington, DC, USA: IEEE Computer Society, 2000. (INFOVIS '00), p. 115–.
- HAVRE, S.; HETZLER, E.; WHITNEY, P.; NOWELL, L. ThemeRiver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 8, n. 1, p. 9–20, jan. 2002. ISSN 1077-2626.
- HEER, J.; BOSTOCK, M.; OGIEVETSKY, V. A tour through the visualization zoo. *Communications of the ACM*, ACM, New York, NY, USA, v. 53, n. 6, p. 59–67, jun. 2010. ISSN 0001-0782.
- HEER, J.; KONG, N.; AGRAWALA, M. Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2009. (CHI '09), p. 1303–1312.
- HEER, J.; ROBERTSON, G. Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 13, n. 6, p. 1240–1247, nov. 2007. ISSN 1077-2626.
- HEER, J.; SHNEIDERMAN, B. Interactive dynamics for visual analysis. *Communications of the ACM*, ACM, New York, NY, USA, v. 55, n. 4, p. 45–54, abr. 2012. ISSN 0001-0782.
- HERMAN, I.; MELANCON, G.; MARSHALL, M. S. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 6, n. 1, p. 24–43, jan. 2000. ISSN 1077-2626.
- HOFFMAN, D. D. *Visual Intelligence: How We Create What We See*. 1st. ed. New York, NY, USA: W. W. Norton and Company, Inc., 1998.

IBGE. *Instituto Brasileiro de Geografia e Estatística*. Brasil – Ministério do Planejamento, Desenvolvimento e Gestão (MP), 2017. <<http://www.cidades.ibge.gov.br/v3/cidades/municipio/3548906>>. Online; acessado em 3 de março de 2018.

INMET. *Instituto Nacional de Meteorologia*. Brasil – Ministério da Agricultura, Pecuária e Abastecimento (MAPA), 2014. <<http://www.inmet.gov.br/portal/index.php?r=estacoes/estacoesAutomaticas>>. Online; acessado em 3 de março de 2018.

INPE. *INPE - Portal do Monitoramento de Queimadas e Incêndios*. INPE - Instituto Nacional de Pesquisas Espaciais, 2018. <<http://www.inpe.br/queimadas/bdqueimadas>>. Online; acessado em 3 de dezembro de 2018.

INSELBERG, A.; DIMSDALE, B. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In: *Proceedings of the 1st Conference on Visualization*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1990. (VIS '90), p. 361–378.

ISO 13271. *Stationary source emissions – Determination of PM₁₀/PM_{2.5} mass concentration in flue gas – Measurement at higher concentrations by use of virtual impactors*. International Organization for Standardization, 2012. <<https://www.iso.org/standard/53581.html>>. Online; acessado em 3 de março de 2018.

ITOH, M.; TOYODA, M.; KITSUREGAWA, M. An interactive visualization framework for time-series of web graphs in a 3D environment. In: *Proceedings of the 2010 14th International Conference Information Visualisation*. Washington, DC, USA: IEEE Computer Society, 2010. (IV '10), p. 54–60.

JAVED, W.; ELMQVIST, N. Stack zooming for multi-focus interaction in time-series data visualization. In: *2010 IEEE Pacific Visualization Symposium (PacificVis)*. Taipei, Taiwan: IEEE Computer Society, 2010. p. 33–40.

_____. Exploring the design space of composite visualization. In: *Proceedings of the 2012 IEEE Pacific Visualization Symposium*. Washington, DC, USA: IEEE Computer Society, 2012. (PACIFICVIS '12), p. 1–8.

JAVED, W.; GHANI, S.; ELMQVIST, N. Polyzoom: Multiscale and multifocus exploration in 2D visual spaces. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2012. (CHI '12), p. 287–296.

JAVED, W.; MCDONNELL, B.; ELMQVIST, N. Graphical perception of multiple time series. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 16, n. 6, p. 927–934, nov. 2010. ISSN 1077-2626.

JOHNSON, B.; SHNEIDERMAN, B. Tree-Maps: A space-filling approach to the visualization of hierarchical information structures. In: *Proceedings of the 2nd Conference on Visualization '91*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1991. (VIS '91), p. 284–291.

JUGEL, U.; JERZAK, Z.; HACKENBROICH, G.; MARKL, V. M4: A visualization-oriented time series data aggregation. *Proceedings of the Very Large Database Endowment*, Very Large Database Endowment, v. 7, n. 10, p. 797–808, jun. 2014. ISSN 2150-8097.

KEHRER, J.; HAUSER, H. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 19, n. 3, p. 495–513, mar. 2013. ISSN 1077-2626.

KEIM, D. A. Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 8, n. 1, p. 1–8, jan. 2002. ISSN 1077-2626.

KEOGH, E.; LIN, J.; TRUPPEL, W. Clustering of time series subsequences is meaningless: Implications for previous and future research. In: *Proceedings of the Third IEEE International Conference on Data Mining*. Washington, DC, USA: IEEE Computer Society, 2003. (ICDM '03), p. 115–. ISBN 0-7695-1978-4.

KO, S.; CHOR, I.; AFZAL, S.; YAU, C.; CHAE, J.; MALIK, A.; BECK, K.; JANG, Y.; RIBARSKY, W.; EBER, D. S. A survey on visual analysis approaches for financial data. In: *Proceedings of the Eurographics / IEEE VGTC Conference on Visualization: State of the Art Reports*. Goslar, Germany: Eurographics Association, 2016. (EuroVis '16), p. 599–617.

KONYHA, Z.; LEZ, A.; MATKOVIĆ, K.; JELOVIĆ, M.; HAUSER, H. Interactive visual analysis of families of curves using data aggregation and derivation. In: *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies*. New York, NY, USA: ACM, 2012. (i-KNOW '12), p. 24:1–24:8. ISBN 978-1-4503-1242-4.

KORCZAK, J.; HERNES, M. Deep learning for financial time series forecasting in A-Trader system. In: GANZHA, M.; MACIASZEK, L. A.; PAPRZYCKI, M. (Ed.). *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*. Prague, Czech Republic: IEEE, 2017. v. 11, p. 905–912.

KRSTAJIC, M.; BERTINI, E.; KEIM, D. A. CloudLines: Compact display of event episodes in multiple time-series. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Computer Society, Los Alamitos, CA, USA, v. 17, p. 2432–2439, 2011.

KUMAR, N.; LOLLA, V. N.; KEOGH, E.; LONARDI, S.; CHOTIRAT, A. R.; WEI, L. Time-series bitmaps: a practical visualization tool for working with large time series databases. In: *Proceedings of the 2005 SIAM International Conference on Data Mining*. Philadelphia, PA, USA: SIAM: Society for Industrial and Applied Mathematics, 2005. p. 531–535.

KöTHUR, P.; WITT, C.; SIPS, M.; MARWAN, N.; SCHINKEL, S.; DRANSCH, D. Visual analytics for correlation-based comparison of time series ensembles. *Computer Graphics Forum*, The Eurographs Association and John Wiley & Sons, Ltd., Chichester, UK, v. 34, n. 3, p. 411–420, jun. 2015. ISSN 0167-7055.

LAM, H.; MUNZNER, T.; KINCAID, R. Overview use in multiple visual information resolution interfaces. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Computer Society, Washington, DC, USA, v. 13, n. 6, p. 1278–1285, 2007. ISSN 1077-2626.

LAMPING, J.; RAO, R.; PIROLI, P. A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In: *Proceedings of the SIGCHI Conference on*

- Human Factors in Computing Systems*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1995. (CHI '95), p. 401–408.
- LEI, S. T.; ZHANG, K. A visual analytics system for financial time-series data. In: *Proceedings of the 3rd International Symposium on Visual Information Communication*. New York, NY, USA: ACM, 2010. (VINCI '10), p. 20:1–20:9. ISBN 978-1-4503-0436-8.
- LI, N.; JIANG, Z.; LIU, Z.; MENG, X. A method of hierarchical time-series data visualization. In: *Proceedings of the 6th International Symposium on Visual Information Communication and Interaction*. New York, NY, USA: ACM, 2013. (VINCI '13), p. 113–114.
- LI, N.; JIANG, Z.; LIU, Z.; SUN, H. Hyperbolic tree + time disc: Visualizing hierarchical time-series data. In: *Proceedings of the 7th International Symposium on Visual Information Communication and Interaction*. New York, NY, USA: ACM, 2014. (VINCI '14), p. 188–191.
- LIN, J.; KEOGH, E.; LONARDI, S. Visualizing and discovering non-trivial patterns in large time series databases. *Information Visualization*, Palgrave Macmillan, v. 4, n. 2, p. 61–82, jul. 2005. ISSN 1473-8716.
- LIN, J.; KEOGH, E.; LONARDI, S.; LANKFORD, J. P.; NYSTROM, D. M. Visually mining and monitoring massive time series. In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2004. (KDD '04), p. 460–469.
- LOHSE, J. A cognitive model for the perception and understanding of graphs. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 1991. (CHI '91), p. 137–144.
- MAATEN, L. V. D. Accelerating t-SNE using tree-based algorithms. *Journal of Machine Learning Research*, JMLR.org, v. 15, n. 1, p. 3221–3245, jan. 2014. ISSN 1532-4435.
- MARQUES, K. A.; CELLI, C. E.; PASSONI, J. H.; TEIXEIRA, D.; BACHIEGA, E.; VIDAL, E. S.; CARVALHO, W. M.; AGUIAR, M. L.; COURRY, J. R. Assessment of atmospheric particulate matter in São Carlos-SP, Brazil. *Environmental Management and Health*, v. 12, n. 1, p. 17–25, 2001.
- MCKINLEY, S.; LEVINE, M. Cubic spline interpolation. *College of the Redwoods*, v. 45, n. 1, p. 1049–1060, January 1998.
- MCKINNEY, W. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. 1st. ed. Sebastopol, CA, EUA: O'Reilly Media, Inc., 2012.
- MCLACHLAN, P.; MUNZNER, T.; KOUTSOFIOS, E.; NORTH, S. LiveRAC: Interactive visual exploration of system management time-series data. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2008. (CHI '08), p. 1483–1492.
- MOHAMMED, G.; KARANI, G.; MITCHELL, D. Trace elemental composition in PM₁₀ and PM_{2.5} collected in Cardiff, Wales. *8th International Conference on Sustainability in Energy and Buildings, SEB-16*, Energy Procedia, Turin, Italy, v. 111, p. 540–547, 2017.

- MULLER, W.; SCHUMANN, H. Visualization for modeling and simulation: Visualization methods for time-dependent data - an overview. In: *Proceedings of the 35th Conference on Winter Simulation: Driving Innovation*. New Orleans, Louisiana: Winter Simulation Conference, 2003. (WSC '03), p. 737–745. ISBN 0-7803-8132-7.
- MURRAY, S. *Interactive Data Visualization for the Web*. 1st. ed. 1005 Gravenstein Highway North, Sebastopol, CA, USA: O'Reilly Media, Inc., 2013.
- MYRTVEIT, I.; STENSRUD, E.; OLSSON, U. H. Analyzing data sets with missing data: An empirical evaluation of imputation methods and likelihood-based methods. *IEEE Transactions on Software Engineering - Special section on the seventh international software metrics symposium*, IEEE Press, Piscataway, NJ, USA, v. 27, n. 11, p. 999–1013, nov. 2001. ISSN 0098-5589.
- NCI. *National Cancer Institute, State Cancer Profiles*. United States of America, 2015. <<http://statecancerprofiles.cancer.gov/>>. Online; acessado em 10 de agosto de 2018.
- NORTH, C.; SHNEIDERMAN, B. A taxonomy of multiple window coordinations. *Technical Report CS-TR-3854*, Department of Computer Science, University of Maryland, College Park, MD, USA, p. 1–, 1997.
- NYTIMES. *The Facebook Offering: How It Compares?* The New York Times, 2012. <www.nytimes.com/interactive/2012/05/17/business/dealbook/how-the-facebook-offering-compares.html>. Online; acessado em 08 de julho de 2018.
- OLIVEIRA, M. C. F.; LEVKOWITZ, H. From visual data exploration to visual data mining: A survey. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 9, n. 3, p. 378–394, jul. 2003. ISSN 1077-2626.
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, JMLR.org, v. 12, p. 2825–2830, nov. 2011. ISSN 1532-4435.
- PERIN, C.; VERNIER, F.; FEKETE, J.-D. Interactive horizon graphs: Improving the compact visualization of multiple time series. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2013. (CHI '13), p. 3217–3226.
- PIKE, W. A.; TASKO, J.; CHANG, R.; O'CONNELL, T. A. The science of interaction. *Information Visualization*, Palgrave Macmillan, v. 8, n. 4, p. 263–274, dez. 2009. ISSN 1473-8716.
- PLAISANT, C. The challenge of information visualization evaluation. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. New York, NY, USA: ACM Press, 2004. (AVI '04), p. 109–116.
- POST, A. R.; HARRISON, J. H. Temporal data mining. *Clinics in Laboratory Medicine*, Elsevier Inc., v. 28, n. 1, p. 83–100, vii, 2008.

- POZZA, S. A.; LIMA, E. P.; COMIN, T. T.; GIMENES, M. L.; COURY, J. R. Time series analysis of PM_{2.5} and PM_{10-2.5} mass concentration in the city of Sao Carlos, Brazil. *International Journal of Environment and Pollution*, Inderscience Enterprises Ltd., v. 41, n. 1-2, p. 90–108, 2010.
- REAS, C.; FRY, B.; MAEDA, J. *Processing: A Programming Handbook for Visual Designers and Artists*. 1st. ed. Cambridge, MA, EUA: The MIT Press, 2007.
- RESNIKOFF, H. L. *The Illusion of Reality*. New York, NY, USA: Springer-Verlag New York, Inc., 1989.
- REZENDE, S. O. *Sistemas Inteligentes: Fundamentos e Aplicações*. 1st. ed. Barueri, SP: Editora Manole Ltda, 2003.
- ROBERTS, J. C. State of the art: Coordinated & multiple views in exploratory visualization. In: *Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization*. Washington, DC, USA: IEEE Computer Society, 2007. (CMV '07), p. 61–71.
- RODRIGUEZ, L.; MENDEZ, D. Pollution spots: A novel method for air pollution monitoring. In: *Air Pollution XXIII*. Southampton, UK: WIT Press, 2015. (WIT Transactions on Ecology and the Environment, v. 198), p. 129–140. ISSN 1743-3541.
- SADANA, R.; STASKO, J. Multiple coordinated visualizations for tablets. *Computer Graphics Forum*, The Eurographs Association and John Wiley & Sons Ltd., Chichester, UK, v. 35, n. 3, p. 261–270, jun. 2016.
- SHEN, R.; SCHÄFER, K.; SCHNELLE-KREIS, J.; SHAO, L.; NORRA, S.; KRAMAR, U.; MICHALKE, B.; ABBASZADE, G.; STREIBEL, T.; ZIMMERMANN, R.; EMEIS, S. Seasonal variability and source distribution of haze particles from a continuous one-year study in beijing. *Atmospheric Pollution Research*, Elsevier, v. 9, n. 4, p. 627–633, 2018.
- SHI, C.; CUI, W.; LIU, S.; XU, P.; CHEN, W.; QU, H. RankExplorer: Visualization of ranking changes in large time series data. *IEEE Transactions on Visualization and Computer Graphics*, IEEE, v. 18, n. 12, p. 2669–2678, October 2012.
- SHI, C.; WU, Y.; LIU, S.; ZHOU, H.; QU, H. LoyalTracker: Visualizing loyalty dynamics in search engines. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Computer Society, v. 20, n. 12, p. 1733–1742, 2014.
- SHNEIDERMAN, B. The eyes have it: A task by data type taxonomy for information visualizations. In: *Proceedings of the 1996 IEEE Symposium on Visual Languages*. Washington, DC, USA: IEEE Computer Society, 1996. (VL '96), p. 336–.
- _____. Extreme visualization: Squeezing a billion records into a million pixels. In: *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*. New York, NY, USA: ACM, 2008. (SIGMOD '08), p. 3–12.
- SHUMWAY, R. H.; STOFFER, D. S. *Time series analysis and its applications: with R examples*. 3rd. ed. New York, NY, USA: Springer Publishing Company, Incorporated, 2011.

SILVA, S. F.; CATARCI, T. Visualization of linear time-oriented data: A survey. In: *Proceedings of the First International Conference on Web Information Systems Engineering*. Washington, DC, USA: IEEE Computer Society, 2000. (WISE '00, v. 1), p. 310–.

STASKO, J.; ZHANG, E. Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In: *Proceedings of the IEEE Symposium on Information Visualization 2000*. Washington, DC, USA: IEEE Computer Society, 2000. (INFOVIS '00), p. 57–.

SUN, G.; WU, Y.; LIU, S.; PENG, T. Q.; ZHU, J. J. H.; LIANG, R. EvoRiver: Visual analysis of topic coopection on social media. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Computer Society, Washington, DC, USA, v. 20, n. 12, p. 1753–1762, December 2014. ISSN 1077-2626.

SUTHERLAND, D. M. G. *France 1789–1815: Revolution and Counterrevolution*. 1st. ed. Cary, NC, USA: Oxford University Press USA, 1985.

SÃO PAULO. *Lei n. 997, de 31 de maio de 1976. Dispõe sobre o controle da poluição do meio ambiente*. Assembleia Legislativa do Estado de São Paulo, 1976. <<https://www.al.sp.gov.br/repositorio/legislacao/lei/1976/alteracao-lei-997-31.05.1976.html>>. Online; acessado em 3 de março de 2018.

_____. *Decreto n. 54.487, de 26 de junho de 2009. Altera a redação e inclui dispositivos e anexos no Regulamento da Lei n. 997, de 31 de maio de 1976*. Assembleia Legislativa do Estado de São Paulo, 2009. <<https://www.al.sp.gov.br/repositorio/legislacao/decreto/2009/decreto-54487-26.06.2009.pdf>>. Online; acessado em 3 de março de 2018.

THAKUR, S.; HANSON, A. J. A 3D visualization of multiple time series on maps. In: *Proceedings of the 2010 14th International Conference Information Visualisation*. Washington, DC, USA: IEEE Computer Society, 2010. (IV '10), p. 336–343. ISSN 1550-6037.

THOMAS, J. J.; COOK, K. A. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. Los Alamitos, CA, USA: IEEE Computer Society Press, 2005.

TOMINSKI, C.; ABELLO, J.; SCHUMANN, H. Axes-based visualizations with radial layouts. In: *Proceedings of the ACM Symposium on Applied Computing*. New York, NY, USA: ACM, 2004. (SAC '04), p. 1242–1247. ISBN 1-58113-812-1.

TOMINSKI, C.; GLADISCH, S.; KISTER, U.; DACHSELT, R.; SCHUMANN, H. Interactive lenses for visualization: An extended survey. *Computer Graphics Forum*, The Eurographics Association and John Wiley & Sons, Ltd., Chichester, UK, v. 36, n. 6, p. 173–200, set. 2017. ISSN 0167-7055.

TOYODA, M.; KITSUREGAWA, M. A system for visualizing and analyzing the evolution of the web with a time series of graphs. In: *Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia*. New York, NY, USA: ACM, 2005. (HYPERTEXT '05), p. 151–160.

TUFTE, E. R. *The Visual Display of Quantitative Information*. 2nd. ed. Cheshire, Connecticut, USA: Graphics Press LLC, 2007.

W3C. *What is the Document Object Model?* World Wide Web Consortium, 2000. <<https://www.w3.org/TR/DOM-Level-2-Core/introduction.html>>. Online; acessado em 27 de março de 2018.

_____. *Scalable Vector Graphics (SVG) 1.1 (Second Edition)*. World Wide Web Consortium, 2011. <<http://www.w3.org/TR/SVG11/>>. Online; acessado em 27 de março de 2018.

_____. *CSS Snapshot 2017*. World Wide Web Consortium, 2017. <<https://www.w3.org/TR/CSS/>>. Online; acessado em 27 de março de 2018.

_____. *HTML 5.2: The canvas element*. World Wide Web Consortium, 2017. <<https://www.w3.org/TR/html5/scripting-1.html#the-canvas-element>>. Online; acessado em 27 de março de 2018.

_____. *HTML 5.2, W3C Recommendation*. World Wide Web Consortium, 2017. <<https://www.w3.org/TR/html/>>. Online; acessado em 27 de março de 2018.

WAGEMANS, J.; ELDER, J. H.; KUBOVY, M.; PALMER, S. E.; PETERSON, M. A.; SINGH, M.; HEYDT, R. V. D. A century of Gestalt psychology in visual perception: I. perceptual grouping and figure-ground organization. *Psychological Bulletin*, American Psychological Association, v. 138, n. 6, p. 1172–1217, 2012.

WAGEMANS, J.; FELDMAN, J.; GEPSHTEIN, S.; KIMCHI, R.; POMERANTZ, J. R.; HELM, P. A. V. D.; LEEUWEN, C. V. A century of Gestalt psychology in visual perception: II. conceptual and theoretical foundations. *Psychological Bulletin*, American Psychological Association, v. 138, n. 6, p. 1218–1252, 2012.

WALKER, J.; BORGO, R.; JONES, M. W. TimeNotes: A study on effective chart visualization and interaction techniques for time-series data. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Computer Society, Washington, DC, USA, v. 22, n. 1, p. 549–558, January 2016. ISSN 1077-2626.

WANG, Y.; HAN, F.; ZHU, L.; DEUSSEN, O.; CHEN, B. Line graph or scatter plot? automatic selection of methods for visualizing trends in time series. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Computer Society, Washington, DC, USA, v. 24, n. 2, p. 1141–1154, February 2018. ISSN 1077-2626.

WANG, Y.; WU, T.; CHEN, Z.; LUO, Q.; QU, H. STAC: Enhancing stacked graphs for time series analysis. In: HANSEN, C.; VIOLA, I.; YUAN, X. (Ed.). *2016 IEEE Pacific Visualization Symposium (PacificVis)*. Taipei, Taiwan: IEEE Computer Society, 2016. p. 234–238.

WARD, M.; GRINSTEIN, G.; KEIM, D. *Interactive Data Visualization: Foundations, Techniques, and Applications*. 1st. ed. Natick, MA, USA: A. K. Peters, Ltd., 2010.

WARE, C. *Information Visualization: Perception for Design*. 2nd. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2004.

WATTENBERG, M. Baby names, visualization, and social data analysis. In: *Proceedings of the 2005 IEEE Symposium on Information Visualization*. Washington, DC, USA: IEEE Computer Society, 2005. (INFOVIS '05), p. 1–.

- _____. A note on space-filling visualizations and space-filling curves. In: *Proceedings of the 2005 IEEE Symposium on Information Visualization*. Washington, DC, USA: IEEE Computer Society, 2005. (INFOVIS '05), p. 24–.
- WEBER, M.; ALEXA, M.; MÜLLER, W. Visualizing time-series on spirals. In: *Proceedings of the IEEE Symposium on Information Visualization*. Washington, DC, USA: IEEE Computer Society, 2001. (INFOVIS '01), p. 7–.
- WHO. *Air quality guidelines. Global update 2005. Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide*. 3rd. ed. Copenhagen, Denmark: World Health Organization, Regional Office for Europe, 2006.
- WIJK, J. J. V.; SELOW, E. R. V. Cluster and calendar based visualization of time series data. In: *Proceedings of the 1999 IEEE Symposium on Information Visualization*. Washington, DC, USA: IEEE Computer Society, 1999. (INFOVIS '99), p. 4–.
- WONG, P. C. Guest editor's introduction: Visual data mining. *IEEE Computer Graphics and Applications*, v. 19, n. 5, p. 20–21, 09 1999. ISSN 0272-1716.
- YI, J. S.; KANG, Y. A.; STASKO, J.; JACKO, J. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 13, n. 6, p. 1224–1231, nov. 2007. ISSN 1077-2626.
- YOON, H.-J.; XU, S.; TOURASSI, G. D. Predicting lung cancer incidence from air pollution exposures using shapelet-based time series analysis. In: *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. Piscataway, NJ, USA: IEEE, 2016. p. 565–568. ISSN 2168-2208.
- ZHAO, J.; CHEVALIER, F.; BALAKRISHNAN, R. KronoMiner: Using multi-foci navigation for the visual exploration of time-series data. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2011. (CHI '11), p. 1737–1746.
- ZHAO, J.; CHEVALIER, F.; PIETRIGA, E.; BALAKRISHNAN, R. Exploratory analysis of time-series with ChronoLenses. *IEEE Transactions on Visualization and Computer Graphics*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 17, n. 12, p. 2422–2431, dez. 2011. ISSN 1077-2626.
- ZHAO, J.; FORER, P.; HARVEY, A. S. Activities, ringmaps and geovisualization of large human movement fields. *Information Visualization*, Palgrave Macmillan, v. 7, n. 3, p. 198–209, jun. 2008. ISSN 1473-8716.
- ZIEGLER, H.; JENNY, M.; GRUSE, T.; KEIM, D. *et al.* Visual market sector analysis for financial time series data. In: IEEE. *Visual Analytics Science and Technology (VAST)*. Salt Lake City, UT, USA, 2010. p. 83–90.