



UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE
RIBEIRÃO PRETO
DEPARTAMENTO DE ADMINISTRAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO DE ORGANIZAÇÕES

**Análise visual aplicada em processos de judicialização de doenças raras do estado de São
Paulo**

Pâmela de Souza Dias

Orientador: Prof. Dr. Evandro Marcos Saidel Ribeiro

RIBEIRÃO PRETO

2022

Versão Corrigida. A original encontra-se disponível na FEA-RP/USP.

Prof. Dr. Carlos Gilberto Carlotti Junior
Reitor da Universidade de São Paulo

Prof. Dr. André Lucirton Costa
Diretor da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto

Prof. Dr. Jorge Henrique Caldeira de Oliveira
Chefe do Departamento de Administração



UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE
RIBEIRÃO PRETO
DEPARTAMENTO DE ADMINISTRAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO DE ORGANIZAÇÕES

Análise visual aplicada em processos de judicialização de doenças raras do estado de São Paulo

Pâmela de Souza Dias

Tese apresentada ao curso de Doutorado em Administração de Organizações do Programa de Pós-Graduação em Administração de Organizações da Universidade de São Paulo, da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, como requisito para a obtenção do título de Doutora.

Linha de pesquisa: Geração de Valor

Orientador: Prof. Dr. Evandro Marcos Saidel Ribeiro

RIBEIRÃO PRETO

2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na Publicação

Faculdade de Economia Administração e Contabilidade de Ribeirão Preto
Universidade de São Paulo
Programa de Pós-Graduação em Administração de Organizações

Dias, Pâmela de Souza

Análise visual aplicada em processos de judicialização de doenças raras do estado de São Paulo

Ribeirão Preto, 2022.

100 p.: il. ; 30 cm

Tese de Doutorado, apresentada à Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto/USP. Área de concentração: Administração de Organizações.

Orientador: Evandro Marcos Saidel Ribeiro.

1. visualização de dados. 2. mineração de dados. 3. judicialização da saúde. 4. saúde em São Paulo. 5. doenças raras.

FOLHA DE APROVAÇÃO

Pâmela de Souza Dias

Análise visual aplicada em processos de judicialização de doenças raras do estado de São Paulo

Tese apresentada ao curso de Doutorado em Administração de Organizações do Programa de Pós-Graduação em Administração de Organizações da Universidade de São Paulo, da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, como requisito para a obtenção do título de Doutora.

Linha de pesquisa: Geração de Valor

Orientador: Prof. Dr. Evandro Marcos Saidel Ribeiro

Aprovada em: ___/___/_____

Banca Examinadora

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

DEDICATÓRIA

“Dedico aos meus pais, os quais nunca mediram esforços para que eu alcançasse meus objetivos.”

AGRADECIMENTOS

A Deus, pelo dom da vida. Aos meus pais, Ivani Camilo Dias e Izabel de Souza Dias, pelo apoio incondicional. Ao meu orientador, Professor Doutor Evandro Marcos Saidel Ribeiro, por seu apoio, dedicação, competência e atenção especial na construção e revisão do trabalho, os quais foram fundamentais para a execução e conclusão dessa tese e para meu crescimento como discente, docente e pesquisadora.

Ao Departamento de Administração de Organizações, em especial ao Programa de Pós-Graduação em Administração de Organizações da Universidade de São Paulo (USP), pelo apoio à minha trajetória durante o doutorado e a todos os professores que, de alguma forma, contribuíram para minha formação.

Aos Professores Doutores Carlos Alberto Grespan Bonacim e Ildeberto Aparecido Rodello, por destinarem parte de seu precioso tempo para participarem da banca de qualificação e final desta tese, cuja contribuição foi decisiva para o amadurecimento do trabalho. Aos demais professores doutores participantes da banca, Andrea Queiróz Ungari, Luciana Romano Morilas e Plínio Rafael Reis Monteiro por suas preciosíssimas contribuições.

Ao grupo de pesquisa Habeas Data - Centro de Estudos e Pesquisa em Jurimetria e todos os seus membros, principalmente ao Janio Gustavo Barbosa, Pedro Augusto Gregorini e Luan Cesar Cardoso que contribuíram de forma direta na coleta dessa tese.

Aos meus ex-orientadores professores doutores Américo Pierangeli Costa e novamente Plínio Rafael Reis Monteiro, por contribuírem para minha evolução como pesquisadora. Também aos meus amigos professores Higor Vinícius Rodrigues Spineli Silva, Tatiana Aparecida Ribeiro dos Santos Benfica, Meryene de Carvalho Teixeira e Luciana Rodas Vera, os quais também contribuíram com minha trajetória acadêmica.

A todos os amigos da FEARP, especialmente à Mariana Silva Lovo, Rodrigo Ruas de Jesus Silva, Flavio Pinheiro Martins e Lucas Fernandes da Costa, os quais, cada um à sua maneira, também colaboraram para minha formação de doutora, entre outros colegas.

Por fim, agradeço o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001" (Portaria Nº 206, de 04/09/2018).

RESUMO

DIAS, P. S. **Análise visual aplicada em processos de judicialização de doenças raras do estado de São Paulo**, 2022. 100 f. Tese (Doutorado) – Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2022.

Ferramentas de *visual* e *business analytics*, se tornaram aliadas para melhorias no campo da saúde, pois, facilitam o fluxo, armazenamento e compreensão dos dados. Com o intuito de buscar novas abordagens que reduzam o tempo e o custo, mas mantém a qualidade, os achados podem surgir de diversos dados, estruturados e não estruturados. Judicialização é um termo que se refere à intervenção do poder judiciário na garantia dos direitos da população. No contexto da saúde, existe uma tendência para o aumento da judicialização na região metropolitana de São Paulo, que possui a maior população do país, assim como em todo Brasil. Diante disso, o presente estudo identificou, padrões em dados públicos de saúde do estado de São Paulo, a fim de descrever as características mais relevantes encontradas em processos de doenças raras, utilizando técnicas de *visual analytics*. Utilizou uma abordagem quantitativa, onde os dados públicos de judicialização da saúde foram analisados a partir de técnicas de *visual analytics* e mineração de dados a fim de identificar padrões a respeito das demandas de portadores de doenças raras e extrair as informações mais relevantes de um grande volume de dados. Concluiu-se que as características mais recorrentes nos processos públicos de doenças raras do estado de São Paulo, são autores, em sua maioria, do sexo masculino e pessoas físicas, enquanto os réus são a maioria pessoa jurídica, mas também predomina o sexo masculino. Quanto aos padrões de tratamento, predominou o tratamento médico-hospitalar e/ou fornecimento de medicamentos, fornecimento de medicamentos e planos de saúde.

Palavras chaves: visualização de dados, mineração de dados, judicialização da saúde, saúde em São Paulo, doenças raras

ABSTRACT

DIAS, P. S. **Visual analytics applied in judicialization processes of rare diseases in São Paulo state.** 2022. 100 f. Thesis (PHD) - School of Economics, Administration and Accounting of Ribeirão Preto, University of São Paulo, Ribeirão Preto, 2022.

Visual and business analytics tools have become allies for improvements in the health field, as they facilitate the flow, storage and understanding of data. In order to seek new approaches that reduce time and cost, while maintaining quality, findings can arise from a variety of structured and unstructured data. Judicialization is a term that refers to the intervention of the judiciary in guaranteeing the rights of the population. In the context of health, there is a trend towards an increase in judicialization in the metropolitan region of São Paulo, which has the largest population in the country, as well as in all of Brazil. Therefore, the present study identified patterns in public health data from the state of São Paulo, in order to describe the most relevant characteristics found in rare disease processes, using visual analytics techniques. It used a quantitative approach, where public health judicialization data were analyzed using visual analytics and data mining techniques in order to identify patterns regarding the demands of rare disease patients and extract the most relevant information. It was concluded that the recurrent characteristics in the public processes of rare diseases in the state of São Paulo, are mostly author of the process male and individuals, while the defendants are mostly legal entities, but the male sex is also predominant. As for the standards of treatment, predominated medical-hospital treatment and/or supply of medicines, supply of medicines and health plans.

Keywords: data visualization, data mining, judicialization in the health, health of São Paulo, rare diseases

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1 - Exemplo de mineração de processos com o software Disco | 29 |
| Figura 2 - Exemplo de visualização: Características dos processos da DRS XII Registro | 39 |
| Figura 3 – Dashboard de estatísticas de frequência dos dados gerais | 59 |
| Figura 4 - Frequência de processos por assunto | 61 |
| Figura 5 - <i>Boxplots</i> de valores das ações | 64 |
| Figura 6 - Valores dos processos da classe procedimento comum | 65 |
| Figura 7 - Valores dos processos da classe mandado de segurança..... | 66 |
| Figura 8 - Valores dos processos da classe procedimento comum excluída a ação de maior valor | 66 |
| Figura 9 - Frequência de processos por DRS | 69 |
| Figura 10 - Frequência de população por DRS | 69 |
| Figura 11 - Gráfico de dispersão entre porcentagens de processos e de população por DRS .. | 71 |
| Figura 12 - Gráfico de dispersão entre porcentagens de população e unidade SUS por DRS | 73 |
| Figura 13 - Gráfico de dispersão entre População e PIB per capita | 74 |
| Figura 14 – Exemplo de dashboard interativo do Tableau | 77 |
| Figura 15 - Contagem de processos por ano | 78 |
| Figura 16 - Dashboard da soma dos valores mensais dos processos de 2013 | 79 |
| Figura 17 - Dashboard da soma dos valores mensais dos processos de 2015 | 79 |
| Figura 18 - Características dos processos da DRS XII Registro | 80 |
| Figura 19 - Sequência de frequência dos movimentos dos processos..... | 81 |
| Figura 20 - Performance média de movimentos dos processos..... | 83 |
| Figura 21 - Performance de movimentos do processo de maior tempo | 84 |
| Figura 22 - Performance de movimentos do processo de menor tempo | 85 |
| Figura 23 - Performance de movimentos dos processos de 09/06/2013 a 01/09/2014 | 87 |
| Figura 24 - Performance de movimentos dos processos de 09/11/2015 a 01/03/2017 | 89 |

LISTA DE QUADROS

| | |
|----------------------------|----|
| Quadro 1 – Abordagens..... | 51 |
|----------------------------|----|

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1 - Divisão dos processos por classes | 61 |
| Tabela 2 - Comparação dos valores da soma de ações por classe | 65 |
| Tabela 3 - Frequência das sentenças classificadas manualmente | 67 |
| Tabela 4 - Frequência das sentenças classificadas automaticamente | 68 |
| Tabela 5 - Comparação entre porcentagens de processos e população por DRS | 70 |
| Tabela 6 - Características sociodemográficas das DRSs | 72 |
| Tabela 7 - Valores mínimos e máximos dos processos por cada tipo de sentença | 75 |
| Tabela 8 - Distribuição das atividades | 82 |
| Tabela 9 - Informações de performance média dos movimentos dos processos | 83 |
| Tabela 10 - Informações de performance média de movimentos dos processos de 09/06/2013 a 01/09/2014 | 87 |
| Tabela 11 - Características dos movimentos dos processos de 09/06/2013 a 01/09/2014 | 88 |
| Tabela 12 - Informações de performance média de movimentos dos processos de 09/06/2013 a 01/09/2014 | 89 |

SUMÁRIO

| | |
|--|-----------|
| 1. INTRODUÇÃO | 13 |
| 1.1 Objetivos | 18 |
| 1.1.1 Objetivo Geral | 18 |
| 1.1.2 Objetivos Específicos | 18 |
| 1.2 Justificativa | 19 |
| 1.3 Estrutura do trabalho..... | 21 |
| 2 FUNDAMENTAÇÃO TEÓRICA..... | 22 |
| 2.1 ESTATÍSTICA E <i>BUSINESS INTELLIGENCE</i> | 22 |
| 2.1.1 Estatísticas descritivas | 22 |
| 2.1.2 Tarefas de mineração de processo | 24 |
| 2.1.3 <i>Business Intelligence</i> | 29 |
| 2.2 <i>ANALYTICS: Business Analytics</i> | 32 |
| 2.3 CONCEITOS DE <i>VISUAL ANALYTICS</i> | 36 |
| 2.4 ASPECTOS SOBRE JUDICIALIZAÇÃO NA SAÚDE | 40 |
| 2.4.1 Judicialização na saúde..... | 40 |
| 2.4.2 Big data no campo da saúde | 43 |
| 2.4.3 Saúde no Brasil e no estado de São Paulo, e doenças raras..... | 45 |
| 3 PROCEDIMENTOS METODOLÓGICOS | 50 |
| 3.1 Caracterização da pesquisa | 50 |
| 3.2 Coleta dos dados e amostra..... | 52 |
| 3.3 Análise de dados | 54 |
| 3.4 Dados ausentes..... | 57 |
| 4 ANÁLISE DOS RESULTADOS E DISCUSSÃO | 59 |
| 4.1 Características gerais da base de dados | 59 |
| 4.2 Frequências estatísticas entre as categorias dos dados | 64 |
| 4.3 Resultados de análises por regionais | 68 |
| 4.4 Apresentação de visualizações dinâmicas | 75 |
| 4.5 Valores obtidos com mineração dos processos..... | 81 |
| 5 CONSIDERAÇÕES FINAIS..... | 92 |
| REFERÊNCIAS BIBLIOGRÁFICAS | 96 |

1. INTRODUÇÃO

Comumente a era atual é chamada de a era dos dados. Tal fato se deve ao aumento gradual de usuários de computadores e smartphones, o que, conseqüentemente, gera uma grande quantidade de dados, a partir de vários domínios, tais como, da saúde e negócios. Nesse sentido, várias ferramentas relacionadas aos negócios, são usadas para analisar e visualizar diferentes tipos de dados que são gerados diariamente (Jena, 2019).

Por essa razão, surgiram diversas ferramentas para análise de grandes volumes de dados, as quais trouxeram alternativas mais viáveis para soluções em diversos campos, entre eles, o da saúde. Com o uso de variadas técnicas é possível facilitar o fluxo e armazenamento de bases de dados nos mais variados setores, sendo utilizados dados estruturados e não estruturados para estudos de novas abordagens. Os novos modelos buscam reduzir o tempo e o custo de pesquisa ou processamento, porém, mantendo a qualidade (Galetsi, Katsaliaki, & Kumar, 2020).

Nesse sentido, as pesquisas em *Business Intelligence* (BI), termo popularmente conhecido como inteligência empresarial, tem se tornado uma opção viável para tal contexto. O princípio do BI se baseia em recursos e metodologias avançadas, para apoiar o processo de tomada de decisão, orientada por dados. Suas aplicações apresentam grande expressividade no serviço de saúde, provavelmente, pela grande disponibilidade de dados de assistência médica e de hospitais. Embora seja um campo de interesse multidisciplinar, a ciência da computação e os sistemas de informações gerenciais, são áreas centrais para o impulsionamento dos estudos sobre a temática (Liu, 2018).

Portanto, os sistemas de BI possuem a capacidade de fornecer informações complexas e competitivas para o processo de tomada de decisão. Apesar dos crescentes investimentos nesses sistemas, ainda é preciso direcionar uma maior atenção, pois diversas organizações ainda são incapazes de alcançar o sucesso desejado, enfrentando sub utilização e uso ineficaz dos modelos empregados (Ul-Ain, Giovanni, DeLone, & Waheed, 2019).

Nesse sentido, assim como o BI, o *Business Analytics* (BA) também se refere a tecnologias que dão suporte a tomadas de decisões a partir de análise de dados. Porém, as soluções de BI, em geral, têm suas implicações para a tomada de decisão e controle. Por outro lado, o BA possui como diferencial, o foco em modelagem preditiva, buscando resultados mais assertivos. A modelagem preditiva é utilizada para identificar questões mais lucrativas, com maior potencial, e também os pontos propensos a prejuízos (Liu, 2018; Wang, Yeoh, Richards, Wong, & Chang, 2019).

Logo, as competências de BA devem ser implementadas de modo que possam ser aproveitadas para melhorar a assimilação das análises de BI. Portanto, a expectativa é que se tenham ferramentas relativamente fáceis de usar, dados de alta qualidade e acessíveis aos envolvidos no projeto (Wang et al., 2019).

Outra possibilidade relevante, é utilizar dados abertos e compartilhá-los com outras instituições de fins semelhantes. Porém, se houver pouca interatividade entre os cientistas de dados e os analistas dos modelos de negócios, possivelmente ocorrerá sub aproveitamento, com modelos mapeados mas não implementados em análises (Hindle & Vidgen, 2018).

Portanto, para auxiliar a demanda de análises dessas áreas, são necessários o desenvolvimento e a frequente atualização de softwares com tais propósitos, por exemplo, ferramentas como Tableau, Microsoft Power BI e, RStudio. Em especial, o RStudio se destaca por ser uma interface do R, o qual é gratuito e receber contribuições de pesquisadores e especialistas de vários campos da ciência. Além disso, pode ser redistribuído e/ou modificado continuamente, tendo como um dos objetivos incentivar os usuários a desenvolverem sua própria programação. Existem milhares de pacotes disponíveis de vários autores em várias áreas, os quais são sempre atualizados quando novas contribuições são recebidas (Sousa, Rodrigues, Lima, & Chagas, 2020).

O desenvolvimento de sistemas analíticos avançados contrapõe com algumas dificuldades, tais como, determinar as necessidades certas, selecionar os algoritmos corretos e alinhar as iniciativas de análise com os objetivos desejados. Além disso, a falta de experiência em aprendizado das técnicas também é um empecilho para o avanço (Nalchigar & Yu, 2018).

Dessa forma, as técnicas de *visual analytics*, referida como análise visual em sua tradução, contribuem trazendo uma facilitação na compreensão das informações contidas, em um determinado volume de dados, comumente, grandes volumes. Portanto, a teoria dessa área implica em um conjunto de afirmações plausíveis que descrevem, com relativa precisão, como várias formas gráficas transmitem informações quantitativas, fornecendo uma diretriz para a construção e interpretação dos gráficos. Porém, trata-se de um conjunto de diretrizes que devem ser usadas com julgamento, as quais são aplicadas a uma variedade de gráficos e mapas estatísticos (Cleveland & McGill, 1984).

Dessa forma, considerando o contexto da saúde, o acesso a dados de qualidade gera pesquisas e cuidados médicos melhores. Isso requer uma abordagem de governança sistêmica para a coleta dos dados e seu uso em laboratórios, clínicas e instituições. O intuito é evitar desperdícios, possibilitar a prevenção e aprimorar a precisão da eficácia dos tratamentos (Knoppers & Thorogood, 2017).

Portanto, a questão da proteção e propriedade dos dados torna-se mais crucial do que nunca. Não é rebuscado supor que muitas famílias e pacientes teriam uma resposta semelhante ao diagnóstico ou tratamento baseado em sistemas inteligentes. Esse argumento também apoia a noção de que as informações sobre o uso de dados do paciente por terceiros, devem fazer parte do formulário de consentimento informado. Formulários de consentimento livre e esclarecido publicamente disponíveis, podem fornecer benefícios substanciais e proteção legal (Hasani et al., 2022).

Por essa razão, os cientistas mantêm esforços contínuos para promover a infraestrutura necessária a fim de desenvolver métodos computacionais capazes de melhorar o serviço, mitigar riscos, reduzir custos e aproveitar novas oportunidades. Novos campos analíticos, por exemplo, a análise de dados coletados de mídias sociais ou dados recuperados de aplicativos móveis, provavelmente, levarão a novos sistemas de informação úteis para a saúde (Galetsi et al., 2020).

Dessa forma, a mineração de dados contribui com seu potencial para extração de conhecimento útil de grande quantidade de dados para tomada de decisão, pois permite combinar a capacidade humana de exploração com a capacidade de processamento analítico. Além disso, um exemplo de sua aplicação na área da saúde, é auxiliando médicos que combatem infecções nosocomiais em unidades de terapia intensiva (Ltifi, Benmohamed, Kolski, & Ben Ayed, 2016), além de diversas outras possibilidades, considerando a variedade de dados da área.

O sistema de saúde brasileiro atual possui uma variedade de dados, os quais podem ser melhor aproveitados a partir de técnicas analíticas. O Sistema Único de Saúde (SUS), órgão de gestão pública, atua em todo o país e atende às necessidades de milhões de pessoas. Normalmente, possuem demandas por serviços que permanecem superiores às ofertas nas unidades de saúde e de pessoal (Younger, 2016).

O setor privado de saúde sem fins lucrativos, trata-se de instituições que necessitam de expressivas e permanentes transferências do fundo público para seu financiamento. São organizações privadas tidas como de interesse público, como as organizações sociais da saúde. Porém, ao mesmo tempo, têm se beneficiado das brechas e facilidades concedidas pela lei e apresentado uma evidente expansão. Nota-se um movimento recente de tais organizações pela busca concomitante da condição de entidades filantrópicas, assegurando múltiplas oportunidades de captação de recursos e de benefícios fiscais (Morais, Albuquerque, Oliveira, Cazuzu, & Silva, 2018).

Ademais, recentemente, em 2016, foi aprovada a Emenda Constitucional 95/2016, instituindo um novo regime fiscal com um limite de recursos públicos aplicados as despesas primárias dos próximos vinte anos. A implementação na área da saúde adotou 2017 como ano-base, com início em 2018. Portanto, nos próximos anos será feita a atualização dos limites com base na inflação, calculada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), referente ao último ano apurado em junho do ano anterior, ao qual se relaciona a lei orçamentária (Vairão Junior & Alves, 2017).

Os casos que afetam a saúde pública, relacionam com necessidades variadas. Para essa pesquisa, por exemplo, tratou das doenças raras que, de acordo com o Ministério da Saúde, são consideradas como tal quando afetam até 65 pessoas em cada 100.000 indivíduos, ou seja, 1,3 pessoas para cada 2.000. Estima-se que existam entre 6.000 a 8.000 tipos diferentes de doenças raras no mundo.

As principais causas das doenças raras são: fatores genéticos/hereditários; infecções bacterianas ou virais; infecções alérgicas e ambientais. O Ministério da Saúde possui 17 centros especializados para atendimento em Doenças Raras, distribuídos em diversos estados do Brasil. No estado de São Paulo são o Ambulatório de Especialidade da Faculdade de Medicina Abc, em Santo André, o HC da Unicamp em Campinas e o HC de Ribeirão Preto, em Ribeirão Preto (Ministério da Saúde, n.d.).

Embora as doenças raras afetem um pequeno número de pessoas, são um grande problema de saúde pública em todo o mundo. Quando comparado com o número total de doenças raras, este número é bastante elevado. Portanto, é necessário que os países incorporem as doenças raras em suas políticas de saúde, financiem pesquisas epidemiológicas e cubram os medicamentos órfãos com seguros de saúde (İnce & Güre, 2021).

Nesse sentido, indo de encontro com o propósito dessa pesquisa, Hasani et al., (2022) afirmam que a Inteligência Artificial (IA) pode abrir portas para compreensão de uma grande quantidade de dados de pacientes, desempenhando um papel na tomada de decisões clínicas. Uma possibilidade, é a descoberta de sinais precoces de doenças raras, sendo detectados em estudos de imagem com o auxílio de algoritmos de reconhecimento. Porém, se tal detecção não for proveniente de um pedido formal do paciente, revelar os dados que não pretendiam ser encontrados geram implicações éticas e psicológicas que precisam ser abordadas.

De acordo com a estimativa do IBGE (2018) o estado de São Paulo tem a maior população do país e é o segundo em rendimento nominal mensal domiciliar per capita, ficando atrás apenas do Distrito Federal. Não surpreendentemente, vários acontecimentos de importância nacional para a saúde ocorreram no estado. Exemplo disso, foi o processo de descentralização da

Secretaria de Estado da Saúde, a qual ocorreu por influência do movimento de reforma da administração pública paulista, realizado ao longo do governo de Franco Montoro, no período de 1983 a 1986 (Mendes et al., 2015).

Portanto, diante dos diversos desafios que compreendem o sistema de saúde atual, a judicialização se tornou mais frequente na área. Lopes e Mello (2018) apontam que o termo é compreendido como a intervenção do poder judiciário na garantia dos direitos da população. A existência de um volume significativo de processos de judicialização, sugere que a saúde precisa de uma maior atenção e que as políticas e planejamentos orçamentários municipais, estaduais e da união são insuficientes para atender a necessidade pública.

Nesse sentido, a judicialização da saúde ocasiona a criação de processos administrativos adicionais para atender às demandas judiciais. Assim, os custos operacionais do sistema de saúde aumentam e a gestão torna-se mais complexa, gerando ineficiência. Como resultado, o acesso oportuno a cuidados de saúde por parte das pessoas que não utilizam a via judicial fica comprometido, sendo esse um impacto negativo (Vargas-Pelaez et al., 2019).

Por exemplo, no estudo realizado pelo INSPER (2019) identificou-se que ocorreu um crescimento de aproximadamente 130% no número de demandas, ano a ano, de primeira instância relativas ao direito à saúde de 2008 para 2017. Apenas de 2016 para 2017 o aumento foi de 61.910 para 95.752. Tais dados se referem à processos de 1º instância obtidos pela Lei de Acesso à Informação (LAI). Os tribunais dos quais foram obtidos esses dados em formato passível de análise foram: TJRJ, TJMG, TJPI, TJAL, TJPE, TJSP, TJMA, TJMS, TJES, TJAC, TJCE, TJRO, TJRN, TJDF, TJMT, TJSC, TJTO.

Reforçando a argumentação dos autores anteriores, Barreiro e Furtado (2015) apontam que a ocorrência de processos de judicialização implica em falhas nas políticas públicas, seja na implementação, ineficiência dos parâmetros estabelecidos ou por sua ausência. A judicialização envolve condições políticas, sociais e jurídicas com uma crescente atuação do poder judiciário. Portanto, ao cessar as possibilidades nas instâncias político-representativas, o cidadão recorre a via judicial.

De maneira geral, existe uma tendência para o aumento da judicialização da saúde na região metropolitana de São Paulo, assim como em todo país. Tal fato pode significar a não observância dos princípios do SUS de universalidade, equidade e integralidade, com falhas na gestão federal, estadual e municipal. Porém, também podem haver interesses de grupos empresariais envolvidos, tais como, as indústrias farmacêuticas (Lopes & Mello, 2018).

Os juízes passam por tensões quando reforçam o direito à saúde, ordenando a entrega de um caro produto farmacêutico biotecnológico que os contribuintes de um país de renda média

devem pagar (Lamprea, 2017). Vale ressaltar que o Art. 196 da Constituição da República Federativa do Brasil de 1988 afirma que:

Saúde é direito de todos e dever do Estado, garantido mediante políticas sociais e econômicas que visem à redução do risco de doença e de outros agravos e ao acesso universal e igualitário às ações e serviços para sua promoção, proteção e recuperação.
(p.73)

Nesse sentido, foi proposto o seguinte problema de pesquisa:

Com o intuito de entender as novas dinâmicas do mercado e desenvolver métodos que levem à geração de valor no campo da saúde, quais padrões podem ser encontrados em processos de saúde do estado de São Paulo, eles demonstram as características mais relevantes dos processos de doenças raras e podem ser identificados a partir de técnicas de *visual analytics*?

A partir dessa análise, buscou-se demonstrar as possibilidades de contribuição que as técnicas de *visual analytics* podem trazer para geração de valor e tomada de decisão. Para tal, foram propostos métodos que facilitem a investigação de informações importantes, em um grande número de dados. Dessa forma, profissionais diversos, tais como, da saúde, que buscam compreender o cenário de judicialização de doenças raras, do direito, buscando conhecer técnicas que facilitem a compreensão das informações gerais em uma grande quantidade dados, e estudiosos da área de administração, replicando as técnicas e gerando valor para outros contextos, poderão ser beneficiados das informações geradas.

1.1 Objetivos

1.1.1 Objetivo Geral

Identificar padrões em dados públicos de saúde do estado de São Paulo, a fim de descrever as características mais relevantes, encontradas em processos de doenças raras, utilizando técnicas de *visual analytics*.

1.1.2 Objetivos Específicos

- A. Identificar as características mais recorrentes nos processos públicos de doenças raras do estado de São Paulo.

- B. Obter informações, a partir de mineração de dados, e localizar os principais padrões de movimentações nos processos públicos de portadores de doenças raras.
- C. Apresentar, a partir de técnicas de análise e visualização de dados, as informações mais relevantes contidas nos processos de forma simplificada, a fim de permitir a compreensão do público geral.

1.2 Justificativa

A realização dessa pesquisa contribui em perspectivas multidisciplinares. Agrega como um estudo para análises de grandes volumes de dados, em especial a análise visual, e para a área de políticas públicas de saúde, demonstrando diversos aspectos do contexto que envolve processos de judicialização na saúde de doenças raras do estado de São Paulo. A judicialização na saúde ocorre tanto em instituições públicas como privadas, porém, torna-se uma questão de políticas públicas de saúde quando há a intervenção do poder judiciário na garantia dos direitos da população.

Além disso, foram gerados conhecimentos significativos a respeito de como as técnicas de *business analytics* podem contribuir para a tomada de decisão na área da saúde. A escolha do contexto do estado de São Paulo se justifica por ser, de acordo com as estimativas do IBGE (2018), o estado com a maior população do país e, portanto, favorável para obter uma diversidade de fatores. Ademais, as abordagens de *visual analytics* ajudam os interessados a compreender, ou mesmo prevenir, as possibilidades para tomada de decisão.

Younger (2016) afirma que os usuários de planos de saúde privados relatam ter melhor acesso a serviços preventivos e maiores taxas de uso de cuidados de saúde do que aqueles sem tais planos. Essa premissa reforça a necessidade de mais estudos que investiguem possíveis melhorias nas políticas públicas de saúde. Dessa forma, se evidencia a necessidade de melhorias nos investimentos, os quais são precisos tanto na área da saúde como em uma abordagem de políticas integradas (Dantas et al., 2017).

Diante de tais desafios, a judicialização se tornou mais frequente na saúde, sendo, de maneira geral, uma tendência de aumento na região metropolitana de São Paulo, como em todo país (Lopes & Mello, 2018). Por exemplo, além do crescimento aproximado de 130% no período entre 2008 e 2017 já mencionado, nos processos de segunda instância, entre 2009 e 2017, ocorreu um aumento de cerca de 85% do número de demandas, também relativas ao

direito à saúde, conforme informado pelos tribunais por meio de dados levantados via (LAI) (INSPER, 2019).

Logo, o presente estudo contribuiu investigando, apoiando-se na literatura pertinente e nos dados coletados, os possíveis fatores, externos a área da saúde, que também contribuem para incidência de judicialização na área, tais como, características sociodemográficas.

Embora já existam alguns estudos envolvendo a judicialização com técnicas de visualização de dados e mineração de processos na área saúde, tais como, os trabalhos de Lopes e Mello (2018), Barreiro e Furtado (2015), Silva Junior e Dias (2016), e o de Borges e Ugá (2010), não foram encontrados trabalhos que investiguem, especificamente, como as técnicas mencionadas podem contribuir com a judicialização na saúde no escopo de doenças raras.

Porém, existem pesquisas que mencionam a judicialização e propõe uma otimização dos serviços públicos de saúde por meio de estruturas analíticas, por exemplo, o de Oliva e Felipe (2018). Tal fato demonstra que se trata de uma área que ainda necessita de mais contribuições, suportando a justificativa acadêmica do estudo, a qual demonstra a análise de grandes dados no campo da Administração, além de possuir uma significativa multidisciplinariedade com as áreas jurídicas e da saúde.

Portanto, a principal lacuna que se pretendeu investigar, a partir do desenvolvimento de métodos que levem à geração de valor nas instituições de saúde públicas e privadas, foi identificar as características mais recorrentes nos casos de judicialização na saúde. A partir dos resultados obtidos, foram disponibilizadas informações para profissionais e demais interessados na área, a respeito das características dos processos dessa natureza por meio dos resultados de mineração de processos e análise visual. Tal fato suporta a justificativa gerencial, considerando a utilidade da pesquisa para o contexto de saúde pública e privada, demonstrando possibilidades de melhorias nos serviços que podem gerar benefícios para a população e trabalhadores envolvidos. Também vale mencionar a relevância de se compreender os impactos causados pela judicialização na saúde para a população, principalmente pela literatura apontar indícios que pessoas com acesso a melhores condições sociodemográficas, possuem mais informações a respeito das possibilidades judiciais para se conseguir tratamentos e/ou medicamentos (Vargas-Pelaez et al., 2019).

Logo, o estudo pretendeu apontar as principais semelhanças nos processos mais recorrentes, tais como, características sociodemográficas e padrões de tratamentos. Os resultados atingidos aplicando técnicas de mineração de dados, a qual utiliza o processamento analítico a fim de extrair conhecimento útil de grande quantidade de dados (Ltifi et al., 2016),

apontaram os momentos em que os processos têm uma tramitação lenta, o que dificulta a fluidez para julgamento das ações.

Por sua vez, as técnicas de BA contribuíram com sua capacidade de análise preditiva, dando suporte a tomadas de decisões, buscando resultados mais assertivos (Wang et al., 2019), e, assim, satisfazendo o intuito de oferecer informações a partir dos resultados obtidos. A análise facilita a visualização das informações e, conseqüentemente, possibilita a idealização de melhores estratégias e direcionamento de recursos.

1.3 Estrutura do trabalho

O presente trabalho é composto por cinco capítulos, incluindo essa Introdução, que compreende a limitação do tema, o problema de pesquisa, os objetivos e a justificativa. A fundamentação teórica apresenta abordagens de autores que desenvolveram estudos significativos no contexto da saúde no estado de São Paulo, em especial no que se refere a judicialização. Aborda técnicas estatísticas com potencial para a tomada de decisão na área da saúde, tais como, mineração de dados, *business intelligence* e *business analytics*.

Os procedimentos metodológicos descrevem as técnicas que foram necessárias para a realização dessa pesquisa, além de apresentar suas justificativas e limitações. Os métodos empregados foram estabelecidos com base nos propósitos do estudo e são apresentados nas seguintes seções: caracterização da pesquisa, coleta de dados e amostra, análise de dados, e dados ausentes. No capítulo seguinte apresenta-se a análise dos principais resultados e sua discussão.

Por fim, no último capítulo, considerações finais, são comentados os resultados e debatidos com as diversas partes deste trabalho propondo, ainda, sugestões de futuros trabalhos e apresentadas as limitações enfrentadas nessa pesquisa.

2 FUNDAMENTAÇÃO TEÓRICA

O presente capítulo apresenta a fundamentação teórica sobre os principais assuntos abordados no trabalho. Neste ponto, é exposta a revisão de literatura que embasa os objetivos investigados.

2.1 ESTATÍSTICA E *BUSINESS INTELLIGENCE*

2.1.1 Estatísticas descritivas

Os dados são uma coleção de elementos para calcular, raciocinar ou medir. Podem ser coletados, armazenados ou processados, mas não colocados em um contexto no qual qualquer significado possa ser inferido, pois as informações precisam ser organizadas de maneira a estabelecer conhecimento. O conhecimento é o conceito de entendimento de informações com base em padrões reconhecidos de maneira a fornecer respostas úteis (Loshin, 2003).

Portanto, a análise estatística é uma ferramenta que modifica a quantidade de informações envolvidas nos dados de forma resumida, permitindo, assim, obter significado. Os níveis de medição são importantes, pois determinam o procedimento estatístico a ser usado e o tamanho da amostra. Considerando que o erro estatístico de procedimentos é bastante comum, não apenas os profissionais da área devem ter interesse no assunto, mas sim todos os aqueles precisam estar cientes do nível em que as variáveis são medidas e as suas respectivas técnicas. Ademais, as análises descritivas são, provavelmente, as estatísticas mais simples de executar e interpretar, pois fornecem uma estratégia útil para resumir dados e fornecer uma descrição da amostra, embora não forneça informações para análise causal (Fisher & Marshall, 2009).

Logo, as estatísticas descritivas são um conjunto de estatísticas que resumem as características e a distribuição de um conjunto de dados em uma série de valores que incluem mínimo, máximo, intervalo, percentil, média, mediana, moda, desvio médio, desvio padrão, assimetria e curtose. Representam informações que podem ser usadas como base para comparar como as séries de dados se diferem. Em estatística espacial (técnicas que usam propriedades topológicas, geométricas ou geográficas), a estatística descritiva é formulada modificando a estatística clássica para acomodar a natureza bidimensional do espaço geográfico (Lee, 2020).

O nível de mensuração é o mecanismo pelo qual uma variável é pontuada. Geralmente, existem três níveis amplos de medição: nominal, ordinal e contínuo (intervalo/razão). O nível nominal de medida (ou categórico) é a pontuação em categorias amplas. Por exemplo,

na categoria sexo, os homens e as mulheres da amostra precisaram ser alocado em uma das duas, masculino ou feminino. Para dados de nível nominal deve haver categorias suficientes para todos os casos pertencerem e excluírem mutuamente, onde cada caso pertence a apenas uma categoria. Não há hierarquia entre as categorias, ou seja, não há diferença numérica (Fisher & Marshall, 2009).

O nível ordinal de medição é a pontuação dos objetos da pesquisa em categorias ordenadas hierarquicamente. Por exemplo, é usado para variáveis que não podem ser medidas diretamente, como dor, satisfação ou ansiedade. Tomando o exemplo da medição dos níveis de ansiedade, um participante pode ser solicitado a indicar o nível de concordância com o qual uma situação específica provoca ansiedade usando uma escala de 0 a 10, onde 0 é igual a nenhuma ansiedade e 10 é ansiedade total. Como os dados nominais, as categorias ordinais devem ser exaustivas e mutuamente exclusivas, mas, diferentemente dos dados nominais, existe uma hierarquia numérica entre as categorias. Um ponto importante é que, embora os valores numéricos estejam associados às categorias ordinais, esses valores existem para fornecer uma hierarquia em vez de uma medida absoluta (Fisher & Marshall, 2009).

Os dados de nível contínuo são geralmente medidos diretamente usando escalas infinitas, a qual possuem divisões de igual distância, tais como, peso em gramas, pressão em mmHg e volume em mililitros. Os dados contínuos podem ser divididos em níveis de intervalo e razão de medição. Os dados medidos no nível da razão têm um zero absoluto, significando que uma pontuação zero é igual a uma ausência de existência. Por exemplo, zero em comprimento, altura ou peso, significa que não existe. A medição de intervalo não possui um zero absoluto. Exemplificando com a medição de temperatura, zero não significa que estamos ausentes de temperatura (Fisher & Marshall, 2009).

A estatística descritiva pode ajudar pesquisadores e estudantes a compreenderem rapidamente o conteúdo dos dados e avaliar quais podem ser mais adequados aos seus objetivos. Possui apresentação visual na forma de gráficos bidimensionais, os quais são mais fáceis para fins educacionais e de pesquisa do que dados ou texto brutos (Brownstein, Adolfsson, & Ackerman, 2019).

O uso da estatística tem sido feito para estabelecer uma relação de causalidade e investigar fatores que influenciam os comportamentos dos agentes jurídicos. Por essa razão, a jurimetria tem sido aplicada como instrumento na coleta e análise de dados pelo Conselho Nacional de Justiça, mediante a parcerias para elaborações de estudos específicos. Isso se justifica por ser necessário identificar e quantificar o problema para planejar soluções, sempre levando em consideração que as demandas representam interesses de indivíduos com particularidades as

quais não podem ter suas pretensões analisadas de forma genérica e impessoal. Portanto, é preciso garantir humanização, a qual jamais poderá ser substituída por fórmulas matemáticas ou softwares (Couto & Oliveira, 2016).

Não surpreendentemente, a jurimetria tem se tornado um importante instrumento para auxiliar à otimização judicial. Traz contribuições pela perspectiva empírica, com o emprego de metodologia estatística em questões jurídicas, especialmente no que tange à perspectiva temporal. Busca descrever os interesses concretos dos agentes jurídicos, seus conflitos e as soluções proferidas pelos julgadores, com o intuito de auxiliar na compreensão das necessidades dos cidadãos, oferecendo, dessa forma, subsídios para definição de leis mais coerentes com a realidade social (Couto & Oliveira, 2016).

Apesar disso, a gestão da justiça com o gerenciamento de processos e mapeada pela jurimetria, demonstrou que a estatística aplicada ao Direito corresponde a uma ferramenta relevante na condução das peculiaridades da tramitação processual. Logo, constata-se a possibilidade de estabelecer um ponto em comum entre o empirismo e o realismo jurídico correlacionando a jurimetria (Couto & Oliveira, 2016).

Dessa forma, as estatísticas descritivas aliadas a técnicas mais sofisticadas, contribuem para fácil identificação de informações que foram extraídas de dados difíceis de serem compreendidos em seu conteúdo original. Estima-se que a quantidade de dados armazenados no mundo duplique a cada 20. Portanto, com a ocorrência desse crescimento e a disponibilidade de máquinas que podem realizar pesquisas se tornando comuns, aumentam as oportunidades de mineração de dados. À medida que o mundo cresce em complexidade, a mineração de dados se torna mais importante para elucidar padrões ocultos. A capacidade de analisar dados de forma inteligente tem se tornado um recurso valioso, pois pode levar a novas ideias, melhorar a tomada de decisão e gerar vantagens competitivas (Witten, Frank, Hall, & Pal, 2017).

2.1.2 Tarefas de mineração de processo

A mineração de processos é útil para implantar novos modelos de negócios, bem como para analisar e melhorar os existentes. Enquanto muitos pesquisadores estão desenvolvendo novas técnicas de mineração de processos e os fornecedores as incorporam em seus softwares, ainda precisam ser testadas em processos da vida real, principalmente as técnicas mais avançadas. A maior parte da pesquisa de mineração de processos tem se concentrado na descoberta de processos (criação de modelos de processos a partir de dados brutos) e técnicas de repetição para verificar a conformidade e analisar gargalos. Essas técnicas ajudaram as

organizações a lidar com problemas de conformidade e desempenho (De Leoni, Van Der Aalst, & Dees, 2016).

As técnicas de mineração de processos usam dados de eventos para mostrar o que pessoas, máquinas e organizações estão realmente fazendo. Fornece novas percepções que podem ser usados para identificar e resolver problemas de desempenho e conformidade. Nos últimos anos, a adoção da mineração de processos na prática aumentou rapidamente. O campo está amadurecendo e isso reflete não só na aceitação da indústria, mas também no sucesso de conferências sobre o tema (Van Der Aalst, 2019).

Em geral, as tecnologias de mineração de dados podem ser classificadas em duas categorias: tecnologias de mineração de dados supervisionadas e tecnologias de mineração de dados não supervisionadas. As tecnologias supervisionadas visam aprender o relacionamento complicado entre várias variáveis. Geralmente é utilizado para tarefas de previsão ou tarefas de classificação. As tecnologias não supervisionadas são boas para descobrir estruturas intrínsecas, relações e interconectividade de dados (Zhao, Zhang, Zhang, Wang, & Li, 2020).

A mineração busca encontrar padrões nos dados que forneçam conhecimento ou permitam a tomada de decisão. A técnica ocorre com armazenados dos dados eletronicamente e pesquisa automatizada, ou pelo menos aumentada, por computador. Porém, isso não é particularmente novo, economistas, estatísticos, analistas e engenheiros de comunicação há muito tempo trabalham com a ideia de que padrões nos dados podem ser buscados automaticamente, identificados, validados e usados para previsão. A novidade é o aumento impressionante de oportunidades para encontrar padrões nos dados. Nos últimos anos ocorreu um crescimento significativo de bancos de dados, principalmente os de atividades cotidianas (Witten et al., 2017).

Na mineração de processos, as medidas de precisão são usadas para quantificar o quanto um modelo de processo super aproxima o comportamento visto em um log de eventos. Os logs de eventos consistem em sequências de eventos em que cada evento representa de uma atividade de processo (Tax, Lu, Sidorova, Fahland, & van der Aalst, 2018).

Tais técnicas ajudam a melhorar os processos usando dados de eventos, as principais perspectivas de mineração de processos, incluindo fluxo de controle, tempo, caso e perspectivas organizacionais. Esses dados estão amplamente disponíveis em sistemas de informação. No entanto, eles geralmente contêm informações altamente confidenciais. Por exemplo, os sistemas de informações de saúde registram dados de eventos que podem ser utilizados por técnicas de mineração de processos para melhorar tratamentos, reduzir o tempo de espera do paciente, melhorar a produtividade dos recursos entre outros. No entanto, os dados registrados incluem

informações altamente confidenciais relacionadas às atividades de tratamento. Os profissionais responsáveis pela manipulação e tratamento dos dados devem fornecer percepções sobre os processos subjacentes, mas, ao mesmo tempo, não deve revelar informações confidenciais (Rafiei & van der Aalst, 2021).

No entanto, para uma análise mais refinada, é essencial correlacionar diferentes características do processo. Tais como, o quanto desvios do processo normativo causam atrasos e custos adicionais, como casos rejeitados são tratados de forma diferente nas fases iniciais do processo, a influência da experiência do médico no processo de tratamento, entre outras questões (De Leoni et al., 2016).

Os padrões descobertos com a mineração de dados devem ser significativos para que possam resultar em vantagens. Algumas das suas aplicações mais ativas são em domínios nos quais se possuem grandes volumes de dados gravados com precisão. Os bancos foram os primeiros a adotar a tecnologia de mineração de dados, devido ao uso do aprendizado de máquina para avaliação de crédito (Witten et al., 2017).

As possibilidades de se criar soluções com mineração de processos são diversas. Por exemplo, suponha que a lealdade inconstante dos consumidores em um mercado altamente competitivo seja um problema. Um banco de dados com os perfis dos consumidores pode fornecer uma solução, pois os padrões de comportamento dos antigos clientes podem ser analisados para identificar características distintas entre os que provavelmente trocaram para concorrentes daqueles que permaneceram fiéis. Uma vez encontradas essas características, os consumidores que provavelmente não são leais, podem ser direcionados para um tratamento especial, o qual seria inviável para ser aplicado como um todo. Considerando que a economia atual é altamente competitiva, os dados são a matéria-prima para sustentar o crescimento dos negócios (Witten et al., 2017).

Portanto, nota-se que a mineração de processos é útil para implantar novos modelos de negócios, bem como para analisar e melhorar os existentes. Além disso, os serviços financeiros e de seguros começaram a usar análises de dados nos anos 1800 para precificar seguros de vida e gerar relatórios. Ao longo dos anos, as histórias de sucesso foram numerosas e as instituições financeiras que atualmente não usam análises preditivas são vistas como obsoletas e fadadas ao fracasso. Desde então, continuaram inovando, passando por melhorias, incluindo o uso de redes neurais, a introdução de pontuações de crédito e a aplicação de vários conjuntos de dados comportamentais (seja de terceiros ou de mídia social) para complementar suas previsões futuras (Isson & Harriott, 2016).

Ademais, atualmente a técnica também está sendo usada para detectar alterações nos padrões bancários individuais que podem anunciar, por exemplo, uma mudança de banco ou mudança para outra cidade que pode resultar na escolha de um banco diferente. O método pode determinar grupos para os quais novos serviços são adequados, como um cluster de usuários confiáveis e lucrativos que raramente se envolve com inadimplência (Witten et al., 2017).

Técnicas de mineração de processos podem ser usadas para analisar processos de negócios usando os dados registrados durante sua execução. Essas técnicas são alavancadas em uma ampla gama de domínios, incluindo saúde, onde se concentra principalmente na análise de diagnóstico, tratamento e processos organizacionais. Apesar da enorme quantidade de dados gerados em hospitais por funcionários e maquinários envolvidos em processos de saúde, não há evidências de uma adoção sistemática de mineração de processos além de estudos de caso direcionados em um contexto de pesquisa. Ao desenvolver e usar a mineração de processos na área da saúde, as características distintivas desses processos, como sua variabilidade e foco centrado no paciente, exigem atenção direcionada (Munoz-Gama et al., 2022).

As diretrizes clínicas visam melhorar a qualidade dos processos de atendimento por meio de percepções baseados em evidências. No entanto, pode haver boas razões para se desviar de tais diretrizes ou as mesmas podem fornecer suporte insuficiente, pois não são adaptadas a um ambiente específico (por exemplo, política do hospital ou características do grupo de pacientes). Assim, as técnicas de mineração de processos podem ser usadas para mediar entre dados de eventos que refletem a realidade clínica e diretrizes clínicas que descrevem as melhores práticas em medicina. As técnicas de mineração de processos também podem ser usadas para verificar a conformidade, analisar desvios e enriquecer os modelos com diagnósticos relacionados à conformidade (Rovani, Maggi, De Leoni, & Van Der Aalst, 2015).

Nesse contexto, a ciência de dados orientada a processos na área da saúde deve propagar a pesquisa e a aplicação de técnicas voltadas para a melhoria na área. Para enfrentar os principais desafios, é necessária uma colaboração estreita entre especialistas de vários domínios. A esse respeito, a simbiose entre o conhecimento relacionado a dados e processos, e o conhecimento clínico, é essencial. Dentro de um contexto de pesquisa, é necessária uma interação próxima entre campos, incluindo medicina, informática médica, ciência da computação e gerenciamento de processos de negócios (Munoz-Gama et al., 2022).

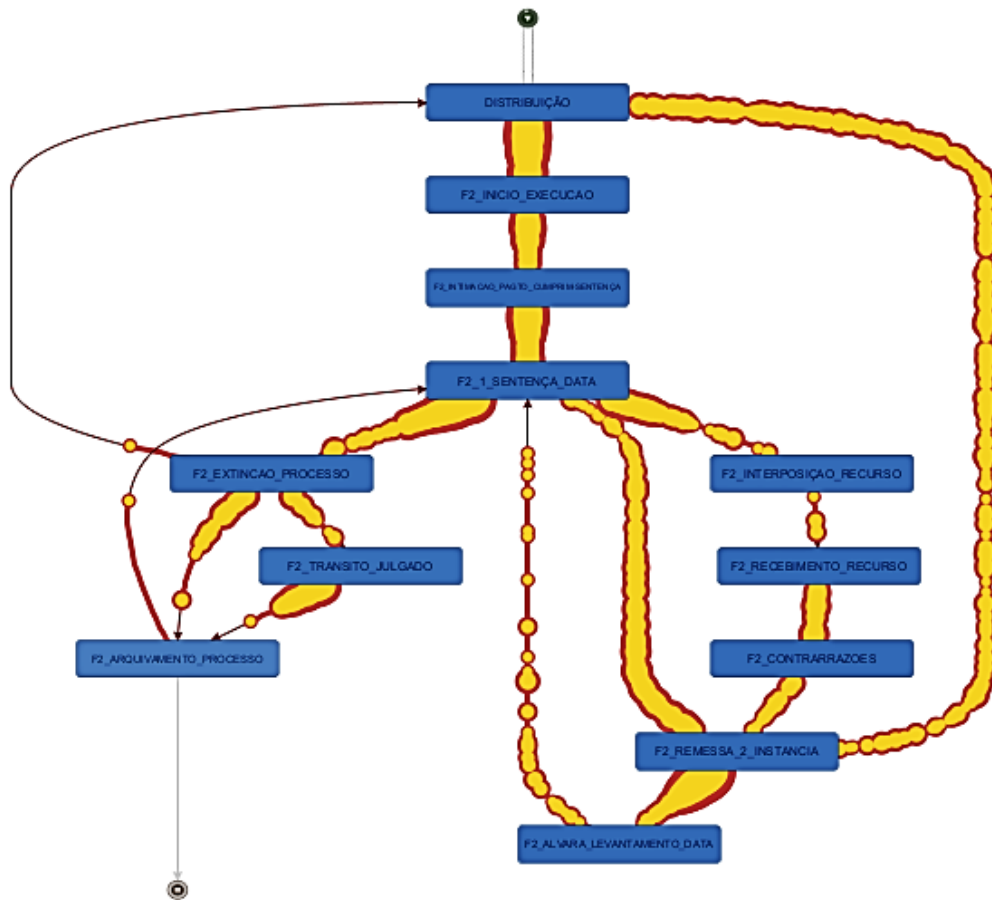
O domínio da mineração de processos na saúde tem como principais desafios as características distintivas que aumentam a complexidade do uso da técnica. Tais desafios precisam ser enfrentados pela comunidade, incluindo pesquisadores e profissionais, a fim de estabelecer estruturalmente a mineração de processos na área da saúde, como um instrumento

poderoso para apoiar a análise e a melhoria de processos baseados em evidências. Para fazer com que isso ocorra, pesquisadores dessas diversas áreas precisam unir forças em um espírito de cocriação. Dessa forma, se podem desenvolver métodos inovadores de mineração de processos que apoiem ao máximo os médicos e tomadores de decisão para gerenciar e melhorar os processos de saúde do mundo real de maneira baseada em evidências (Munoz-Gama et al., 2022).

Como a mineração de processos voltada para a saúde trata-se de um campo de pesquisa relativamente jovem, os seus *stakeholders* devem se conectar ativamente e aprender com campos mais maduros, tais como, inteligência artificial e aprendizado de máquina. Os pontos fortes encontrados serão baseados na combinação inteligente de percepções de várias áreas de pesquisa. Essa interseção interdisciplinar envolverá, cada vez mais, pesquisas sobre big data e Internet das Coisas (Munoz-Gama et al., 2022).

Nesse sentido, esse estudo traz evidências para o campo de saúde através de mineração de processos. A Figura 1 apresenta um exemplo de visualização animada feita com o software Fluxicon Disco. Nela é possível identificar quais são os caminhos que os processos percorrem, acompanhando a direção das setas, e também onde os trâmites são mais lentos, apresentando linhas mais grossas. As áreas amarelas representam os processos agrupados em movimento.

Figura 1 - Exemplo de mineração de processos com o software Disco



Fonte: dados da pesquisa

2.1.3 Business Intelligence

O *Business Intelligence* evoluiu pela quantidade de dados gerados pela internet e pelo aumento exponencial de dispositivos inteligentes, alterando a maneira como as organizações e os indivíduos usam as informações. O fenômeno do Big Data, o volume, a variedade e a velocidade dos dados, impactaram o BI e o uso da informação. Novas tendências, como as análises rápidas e a ciência de dados, surgiram a partir disso (Larson & Chang, 2016).

De acordo com o *Data Warehousing Institute*, o BI é definido como processos, tecnologias e ferramentas necessárias para transformar dados em informações, as informações em conhecimento e o conhecimento em planos que impulsionam ações comerciais lucrativas. Abrange armazenamento de dados, ferramentas de BA e gerenciamento de conteúdo/conhecimento. Tal definição demonstra que há uma hierarquia imposta aos diferentes escopos de inteligência (Loshin, 2003).

Corroborando com o autor anterior, Marinheiro e Bernardino (2015) afirmam que o BI é uma ampla categoria de aplicativos e tecnologias para coletar, armazenar, analisar e fornecer

dados de acesso para ajudar os usuários a tomar melhores decisões de negócios. Enquanto no passado o mercado era estritamente dominado por ferramentas comerciais e de código fechado (com restrição a manipulação do código fonte), os últimos anos foram caracterizados pelo nascimento de soluções de BI com código aberto (código fonte disponibilizado para estudar e modificar).

Entre as principais opções de código aberto estão JasperSoft, Palo, Pentaho, SpagoBI e Vanilla. Porém, o acesso à informação através de *tablets* e *smartphones*, o tratamento de enormes quantidades de dados, e a disponibilização da informação em tempo real ainda são paradigmas e um desafio a ser alcançado sem que os custos organizacionais aumentem (Marinheiro & Bernardino, 2015).

A pesquisa acadêmica relacionada ao BI também tem prosperado. O número de trabalhos tem aumentando rapidamente com tópicos que variam de conceitos, metodologias, aplicativos e gerenciamentos. Tais pesquisas estão frequentemente focadas no gerenciamento e no apoio à decisão e os artigos são comumente encontrados em revistas de sistemas de gestão e informação (Liu, 2018).

Segundo Ul-Ain et al. (2019), a maioria das pesquisas sobre BI de 2000 a 2019 foram a respeito de organizações que deveriam enfatizar o desenvolvimento de capacidades e competências específicas (de usuários) para alcançar o sucesso organizacional. Além disso, uma cultura corporativa baseada em fatos, conhecimento e aprendizado pode motivar os envolvidos a aplicarem as informações oferecidas pelos sistemas de BI.

Ademais, é importante destacar que a prática é mais do que apenas uma coleção de ferramentas, de forma que sem os processos e as pessoas certas, os resultados do BI são de pouco valor. Logo, se o conhecimento que pode ser usado para ações lucrativas for ignorado, a prática terá pouco valor. Infelizmente, frequentemente os dados e informações são usados de forma intercambiável. Com isso, tem-se o risco de colidir com a compreensão de qualquer indivíduo dos termos, dados, informações e conhecimento (Loshin, 2003).

No entanto, o sucesso do sistema de BI precisa ser fundamentado também em teorias de decisão para ajudar a entender como os sistemas e as informações resultantes podem beneficiar o processo de decisão. O Modelo de Sucesso de DeLone & McLean (1992), o Modelo de Aceitação de Tecnologia de Davis (1989) e a Teoria de Difusão da Inovação proposta por Rogers Everett (1995), tem sido muito utilizados nas pesquisas em associação ao BI (Ul-Ain et al., 2019).

O processo de transformar dados em informações pode ser resumido como a determinação de quais dados devem ser coletados e gerenciados e em qual contexto.

Transformar informações em conhecimento envolve componentes analíticos, como *data warehousing*, processamento analítico, a qualidade dos dados, *business rule analysis* e mineração de dados. Ser capaz de agir com base nas ferramentas aprendida é o ponto principal de qualquer estratégia de BI, gerando retorno dos investimentos em tecnologia da informação (Loshin, 2003).

Medir o valor do investimento em BI continua sendo um desafio para as organizações, principalmente devido ao fato dos esforços em tecnologia serem determinados com base em quanto tempo o retorno dos investimentos ocorre. Tais iniciativas exigem um investimento inicial significativo e contínuo; a expectativa é de que a organização se torne mais inteligente, trabalhe mais e tome melhores decisões através do uso de informações (Larson & Chang, 2016).

Portanto, é válido ressaltar as várias etapas que um sistema BI passa. Inicialmente, na fase de descoberta, as expectativas dos projetos não são claras o suficiente para as partes interessadas. Os *stakeholders* começam com o conhecimento de que são necessários recursos de informação e análises. Já os profissionais de TI estão prontos para atender aos requisitos sem um ponto de partida bem definido. Começam com a definição de questões comerciais que fornecem informações sobre fontes de dados, dimensões e fatos necessários (Larson & Chang, 2016).

Em seguida, no design, concentra-se fortemente na modelagem, mas pode começar com o estabelecimento da arquitetura do sistema. A arquitetura em BI é mais do que infraestrutura de hardware, inclui componentes comerciais, técnicos, de processo, dados e projeto. Concentra-se na definição dos drivers, objetivos e estratégia da organização que direcionam as necessidades de informações. Além disso, descreve a metodologia do processo que inclui a estrutura para aquisição e apresentação de dados. Se a arquitetura for estabelecida, o design se concentrará na modelagem de dados e processos para dar suporte às necessidades de informações (Larson & Chang, 2016).

Na fase do desenvolvimento, pode-se incluir uma ampla variedade de atividades. O foco principal é produzir um sistema de trabalho que aplique o contexto de negócios aos dados e apresente informações de uma maneira que permita aos usuários finais analisar informações acessíveis. As atividades podem incluir a codificação de ETL (extração, transformação, carregamento), configurações ou a criação de scripts para agendamento de tarefas. O escopo do desenvolvimento pode envolver a aquisição de dados para preparação, apresentação, acesso e entrega de informações (Larson & Chang, 2016).

Os sistemas de BI têm muitos componentes independentes que exigem integração e são impactados por mudanças contínuas. Devido a essa complexidade, a fase de implantação é formal e controlada. Concentra-se na integração de novas funcionalidades e recursos na produção e em testes de regressão para verificar se a funcionalidade em funcionamento anteriormente não é afetada. Foca na introdução de novos componentes e na manutenção da estabilidade do sistema (Larson & Chang, 2016).

Por fim, na etapa de entrega de valor, inclui a estabilização, manutenção, gerenciamento de mudanças e feedback para o usuário final. Quando bem-sucedidos, geralmente, os sistemas têm uma vida útil longa e exigem gerenciamento de programas para lidar com as mudanças e manter o valor contínuo. Devido à mudança contínua e aos usos dinâmicos das informações, as mudanças que afetam o BI podem ser iniciadas a partir de sistemas de origem, processos de negócios, atualizações de software, nova integração de dados e estratégia organizacional. O feedback do usuário final fornece uma compreensão de como as informações são usadas e do valor geral obtido (Larson & Chang, 2016).

2.2 ANALYTICS: Business Analytics

O sucesso das organizações está fortemente relacionado com o quão bem elas podem analisar os dados disponíveis, extrair conhecimento significativo e otimizar processos organizacionais. Nesse contexto, o *Business Analytics* oferece a oportunidade de aproveitar o conhecimento e o valor ocultos nos sistemas de informações empresariais para inovar, aprimorar o gerenciamento e a produção, direcionando com precisão os esforços de marketing, além de desenvolver e gerenciar serviços lucrativos de pós-venda (Omar, Minoufekar, & Plapper, 2019).

Ao contrário dos relatórios padrões, que relatam sobre a mecânica do funcionamento de um negócio, o BA fornece informações sobre como melhorá-lo. Pode-se aprimorar a utilidade de um sistema BI com a conversão de dados em uma interface gráfica para que os usuários possam visualizar e interagir. Os aplicativos extraem informações do banco de dados e formulam um processo analítico interativo, fornecendo acesso a diferentes aspectos. Por exemplo, separando por um subconjunto dos dados do banco principal ou gerenciando suas próprias visualizações dos dados internamente. Dessa forma, permite a análise de diversas situações, tais como, do comportamento do consumidor, da produtividade, do canal de vendas e da cadeia de suprimentos (Loshin, 2003).

Ainda a respeito do comportamento de consumidores, pode ser aplicado para a criação de perfil do cliente, *call center* eficientes, alcance do público alvo, personalização de apresentações, e para estimar o valor da vida útil e a lealdade do consumidor. Já para produtividade, apoia com análises de métricas de otimização e eficácia na tentativa de entender quando e onde a melhor produtividade é alcançada dentro de uma organização. Também são feitas análises de defeitos, planejamento de capacidade, relatórios financeiros, gerenciamento de riscos e planejamento de recursos (Loshin, 2003).

Para canais de vendas, inclui a criação e análise de campanhas de marketing, avaliando o desempenho das vendas e a eficácia dos canais. Na cadeia de suprimentos é útil para caracterizar e comparar canais de fornecimento de vários vendedores e fornecedores, incluindo o seu gerenciamento, eficácia, controle de estoque e análise da rede de distribuição. Por fim, é feita a avaliação de tendências de atividades como uma maneira de identificar padrões interessantes ou preditivos, detecção de fraudes e análise de redes sociais (Loshin, 2003).

Tratando-se do escopo dessa pesquisa, doenças raras (DRs), tem-se a expectativa de que a eficácia de ferramentas de BA, treinadas e avaliadas em condições comuns, seja capaz de ser usada para tais doenças. Logo, se faz necessário estimular a inclusão das DRs no estágio de treinamento e desenvolvimento de novos sistemas, com aplicação do algoritmo para o diagnóstico ou gerenciamento de pessoas portadores de tais enfermidades, pois o viés de amostragem ganha informações com ausência de generalização. Portanto, a implantação de BA que exclui um subgrupo de indivíduos diverge de qualquer progresso em direção ao objetivo de maximizar os benefícios clínicos e abordar as iniquidades de saúde em populações historicamente negligenciadas (Hasani et al., 2022).

Dessa forma, poderá alcançar uma melhora e robustez dos sistemas de BA e sua utilidade em uma variedade de configurações clínicas, incluindo as DRs e sua diversidade de dados. O desempenho aprimorado de tais sistemas contribui para um mercado mais diversos, portanto, mais lucrativo. Logo, as vantagens financeiras e sociais de tais dispositivos médicos precisos e bem-sucedidos, seriam benéficas para a comunidade e para os desenvolvedores de sistemas de inteligência artificial. Além disso, estudar e coletar dados sobre doenças pode fornecer percepções sobre os mecanismos biológicos subjacentes que podem ser aproveitados pelos desenvolvedores para melhorar ainda mais o desempenho e a robustez de sistemas futuros (Hasani et al., 2022).

Entretanto, Omar et al. (2019) afirmam que embora apresente variadas vantagens, a adoção do BA muitas vezes é feita de forma esporádica e concentrada nos esforços departamentais, em vez de no empreendimento da organização. Tal fato pode ser associado com

diferentes desafios, por exemplo, dados isolados em sistemas de TI incompatíveis, comunicação interdepartamental deficiente e falta de envolvimento de executivos seniores, os quais são essenciais para liderar uma transição para a tomada de decisões e ações baseadas em evidências.

Ademais, a liderança de mercado através do uso eficaz de BA é o resultado da superação de três barreiras. A primeira é uma barreira tecnológica, associada à padronização da coleta, agregação e armazenamento de dados. A segunda aborda a cultura empresarial em relação à tomada de decisões e ações baseadas em evidências e compartilhamento de informações. Também visa obter interoperabilidade contínua dentro e além dos limites organizacionais. Isso cria o ecossistema perfeito para responder a perguntas de negócios por meio da BA que oferecem resultados e recomendações acionáveis. Por fim, a terceira barreira está relacionada à monetização de dados por meio da inovação do modelo de negócios para criar valor e capturar ações em um mercado anteriormente inexplorado. Isso permite a diferenciação em ofertas equivalentes e uma liderança de mercado segura (Omar et al., 2019).

A implementação do BA nas organizações passa por três etapas: iniciação, adoção e assimilação. Fatores organizacionais, obstáculos gerenciais e centralização analítica influenciam os estágios mais avançados como adoção e assimilação, enquanto o fator ambiental, intensidade da competição, afeta positivamente apenas a iniciação. Aspectos de contexto tecnológico, a infraestrutura e a qualidade dos dados são os pontos mais importantes em todas as fases. O impacto dos fatores organizacionais aumenta à medida que o estágio de adoção avança. Portanto, conclui-se que os efeitos dos fatores em cada estágio são diferentes (Nam, Lee, & Lee, 2019).

Logo, é preciso elaborar diferentes planos gerenciais para cada estágio. Sugerindo que o objetivo seja conquistar competitividade no mercado, deve-se priorizar a preparação da infraestrutura relacionada a dados, investindo em equipamentos e infraestrutura de TI interno que permita a coleta de dados estruturados e não estruturados de transações comerciais e fontes externas e a reintegração desses dados quando eles forem armazenados. Esse fator contextual bem estabelecido relacionado a dados é a base de uma aplicação de BA bem-sucedida.

Dada as expectativas com big data, seria vantajoso ter incentivos governamentais para a utilização do BA no setor privado. Parte-se do princípio que a iniciativa é mais eficaz do que esperar que as organizações avancem sozinhas em direção à inovação. É preciso mudar os rumos das políticas e programas, entendendo o que os setores privados esperam de apoio do governo, fornecendo educação gratuita para a tecnologia em BA e big data, e executado outros programas de suporte (Nam et al., 2019).

Além disso, outra técnica que vem recebendo destaque é o *People Analytics* (PA). Também chamado de *Talent Analytics*, se baseia nos princípios da analítica tradicional para avaliar opções de recursos humanos (RH), levando em consideração fatores como custo, deslocamento, facilidade e conveniência. Uma das melhores maneiras de lidar com a migração do BA para o PA, é inspirando na maneira que setor de marketing adotou o *analytics*, uma vez que a área tem aspectos semelhantes com o RH. Procurando estrategicamente segmentar, atrair, crescer e reter consumidores, o marketing se abriu para a análise, a fim de entender completamente sua base de clientes, segmentar dados demográficos, otimizar a compreensão do ciclo de vida do cliente e o gerenciamento do relacionamento (Isson & Harriott, 2016).

No caso do RH, a busca seria por funcionários de ponta, os quais tendem a criar os melhores produtos, gerando receita de forma mais eficiente. O uso de PA resulta em ótimos locais de trabalho, satisfação para os consumidores, os quais influenciam outros para consumir na mesma organização. Encontrar, gerenciar e manter os melhores talentos é fundamental para o crescimento, principalmente pela capacidade de se adaptarem às mudanças nas condições de negócios (Isson & Harriott, 2016).

Na prática, pode-se dizer que o PA funciona a partir de algumas características. Os dados podem ser coletados de diversas fontes, tais como, redes sociais, metadados (dados sobre dados), avaliações de usuários, tendências de vendas, entre outros. Esses dados são organizados e apresentados de maneira que seja possível reconhecer opções de soluções para o propósito em questão (Tursunbayeva, Di Lauro, & Pagliari, 2018).

Felizmente, o mercado de ferramentas e serviços comerciais de PA é diversificado, oferecendo inúmeros benefícios funcionais e estratégicos. As empresas que comercializam tais sistemas e serviços enfatizam mais os benefícios para os empregadores do que para os funcionários. Portanto, torna-se uma direção para o desenvolvimento de novas inovações e demandas de mercado (Tursunbayeva et al., 2018).

Normalmente, as fontes de dados podem ser divididas em três categorias: dados de talentos, da empresa e de mercado. Os dados de talentos incluem itens como custos indiretos, para o departamento de RH, e para recrutamento. Além disso, também é preciso considerar estruturas organizacionais, controle de talentos, qualidade de contratação, desempenho e engajamento dos funcionários, remuneração e benefícios, produtividade e liderança, aprendizado e desenvolvimento de carreira, planejamento de sucessão, rotatividade, histórico de desempenho, resultados dos testes de seleção e bem-estar geral. Dependendo da função, a presença em mídias sociais, com dados disponíveis publicamente, fornecem um conjunto complementar aos dados de talentos (Isson & Harriott, 2016).

Quanto aos dados da empresa, incluem itens como desempenho de vendas, receitas associadas, bases de clientes, crescimento de ações, diversidade de produtos, lealdade, rotatividade, tráfego e conversão. Por último, os dados de mercado incluem estatísticas a partir de folha de pagamento, de taxas de emprego e desemprego, de produto interno bruto, de taxas de rotatividade, de vagas e demissões, e de salários e ordenados. Todos esses dados podem ser detalhados por setor, tamanho da empresa, ocupação, estado e cidade. Ao identificar esses diferentes fluxos de dados e vinculá-los à estratégia de gerenciamento de talentos, será possível saber quais são os pontos essenciais para criar um plano acionável (Isson & Harriott, 2016).

Contudo, Tursunbayeva et al., (2018) afirmam existir uma acentuada ausência de considerações éticas em relação às práticas de PA, algumas das quais são encobertas ou ultrapassam os limites das próprias organizações. Os exemplos incluem o monitoramento de mídias sociais pessoais ou atividades de e-mail, que têm implicações de privacidade, o uso de decisões algorítmicas para recrutamento ou promoção com potencial para introduzir viés e discriminação, o que levanta questões de transparência e consentimento.

Várias organizações de variados setores adotaram a análise de dados para criar valor comercial em diferentes estágios, dependendo do nível de maturidade analítica e dos objetivos comerciais mais prementes. Portanto, a análise preditiva, e conseqüentemente o BA, está se tornando um diferenciador generalizado e competitivo em diferentes funções de negócios. Por exemplo, a Target, uma das maiores varejistas dos Estados Unidos, consegue prever quando o filho de uma mulher grávida deve nascer com base nos produtos que ela compra, simplesmente combinando os dados do seu cartão de fidelidade com as informações das mídias sociais. Assim, detectando e aproveitando as mudanças nos padrões de compra, permite que a empresa direcione promoções para produtos relacionados a fase do bebê (Isson & Harriott, 2016).

Outro exemplo, foi a Google que conseguiu prever a epidemia de gripe de 2009 duas semanas antes dos Centros de Controle de Doenças, simplesmente aproveitando as tendências de pesquisa on-line (por exemplo, relacionadas a sintomas). No campo da saúde os hospitais pediátricos estão aplicando análise de dados às transmissões ao vivo dos batimentos cardíacos de um bebê para identificar padrões e, com base na análise, o sistema pode detectar infecções 24 horas antes do bebê começar a apresentar sintomas, o que permite uma intervenção precoce (Isson & Harriott, 2016).

2.3 CONCEITOS DE VISUAL ANALYTICS

Gráficos e visualização de informações são uma mistura de conceitos, métodos e procedimentos que abrange muitas áreas, por exemplo, os princípios do design de mapas (da cartografia), diretrizes sobre como exibir melhor os dados em um gráfico (da estatística), recomendações práticas sobre o uso de layout e paletas de cores (do design gráfico), princípios de estilo de escrita (do jornalismo), uma ampla variedade de ferramentas de software (da computação) entre outros. A visualização deve ser vista como uma tecnologia, pois a tecnologia pode ser, em primeiro lugar, qualquer objeto, processo ou método criado para ajudar em uma tarefa, sendo um meio para cumprir um propósito humano. Os gráficos de informações também são tecnologias, sendo ferramentas cujo objetivo é ajudar a completar tarefas (Cairo, 2013).

No que se refere a conceitos, a teoria de métodos gráficos lida com uma parte pequena, mas importante, de todo o processo de percepção gráfica. Primeiramente, participa da identificação de um conjunto de tarefas elementares que são realizadas quando as pessoas extraem informações quantitativas dos gráficos. A segunda parte, é uma ordenação das tarefas com base na precisão com que as pessoas as realizam. Os elementos da teoria são testados por experimentação, na qual os sujeitos registram seus julgamentos das informações extraídas dos gráficos. Os experimentos validam esses elementos, mas também sugerem que o conjunto de tarefas elementares pode ser expandido (Cleveland & McGill, 1984).

É necessária uma melhoria significativa nos gráficos populares e, nesse sentido, o *visual analytics* (análise visual) oferece ferramentas que se pode usar para explorar e analisar um conjunto de informações. Dessa forma, pode-se dizer que são ferramentas que ajudam os leitores a descobrirem as histórias dos dados sozinhos. Felizmente o design dos gráficos atuais são mais democráticos do que era uma década atrás, período que os dados eram escassos e os softwares caros e difíceis de usar. Além disso, também temos o potencial de usar gráficos de forma mais inteligente, porque os princípios que informam seu design nunca foram tão claros (Cairo, 2013).

Contudo, um desafio substancial na realização de experimentos gráficos perceptivos é justamente a forma que as pessoas realizam julgamentos, o qual pode ser enviesado. Os sujeitos tentarão obter a resposta certa e poderão executar tarefas muito mais cognitivas do que as tarefas perceptivas básicas que realizam na vida real (Cleveland & McGill, 1984).

Nesse sentido, faz-se necessário manter a coerência com os padrões que são encontrados nos resultados. A palavra "padrão" aparece frequentemente na literatura de visualização e análise *analytics*. Andrienko, Andrienko, Miksch, Schumann e Wrobel (2021) sugerem que existe a necessidade de estudos empíricos sobre a percepção de exibições visuais que se concentrem especificamente na percepção de relacionamentos, em vez de julgamento de valores

absolutos. Tais estudos precisam apontar perspectivas sobre a facilidade de um usuário identificar relações particulares entre os elementos da base, e se também é possível identificar relacionamentos inexistentes.

Buscando evitar que os gráficos de informações sejam interpretados de formas equivocadas, a análise visual atua como a ciência do raciocínio analítico facilitado por interfaces visuais interativas, simplificando a compreensão do desempenho de atividades cognitivas envolvendo big data. O termo *analytics* refere-se ao uso sistemático de dados combinado com análises quantitativas e qualitativas para tomar decisões, podendo ser aplicado para melhorias de qualidade e desempenho, como avaliação preditiva de risco, suporte a decisões e gerenciamento de recursos (Simpao, Ahumada, & Rehman, 2015).

Tratando-se do campo da saúde, os avanços na tecnologia de computadores, sistemas de monitoramento de pacientes e sistemas de registros virtuais, permitiram o rápido acúmulo de dados de pacientes em formato eletrônico. A integração contínua de big data e análises de cuidados de saúde, aumentou a demanda por profissionais que possam preencher a lacuna entre as ciências médicas e da informação (Simpao et al., 2015).

A sofisticação e a onipresença crescentes dos sistemas de monitoramento eletrônico tanto no hospital quanto na comunidade, alimentam quantidades, cada vez maiores, de busca por profissionais que tragam respostas de assistência médica para médicos, administradores e pesquisadores. A expectativa é de que os métodos de análise tragam ferramentas promissoras para melhorar o atendimento ao paciente, a avaliação de qualidade, a gestão financeira e outras áreas que compõe o campo da saúde. A análise visual pode facilitar a exploração de big data para gerar hipóteses e orientar a seleção e o uso de métodos de análise avançada (Simpao et al., 2015).

A análise visual oferece uma nova capacidade para analisar sistemas de saúde e apoiar uma melhor tomada de decisão. Embora a técnica continue a evoluir rapidamente, é necessário garantir que essa evolução reflita as necessidades práticas dos *stakeholders*. A disponibilidade de dados de saúde cresceu exponencialmente, tanto em quantidade quanto em complexidade. A velocidade dessa evolução gerou novos desafios para traduzir dados complexos em políticas efetivas baseadas em evidências, sendo, dessa forma, necessário que as técnicas de análise visual evoluam juntas (Chung et al., 2020).

Em seu estudo, Guo et al., (2020) demonstraram a eficácia de um sistema a partir de um estudo de caso, usando um conjunto de dados da unidade de terapia intensiva neonatal da UC Davis, em Sacramento, CA, EUA. Foi desenvolvido um sistema de análise visual para apoiar estudos comparativos de registros de pacientes. O intuito foi facilitar a comparação de

sequências com diferentes comprimentos a partir de um método de alinhamento de sequências. Por meio de uma interface interativa, o usuário pode identificar rapidamente os pacientes de interesse e revisar convenientemente os aspectos temporais e multivariados dos prontuários.

Dessa forma, o método de incorporação e sistema de análise visual elaborado pelos autores, foi útil para analisar os resultados dos exames médicos de recém-nascidos. Porém, os métodos são genéricos o suficiente para serem aplicados a outros registros médicos que podem ser representados como sequências temporais multivariadas sem grandes modificações. Além disso, podem ser utilizados para outras aplicações, como suporte à decisão de carreira, análise financeira e sistemas de segurança cibernética (Guo et al., 2020). Portanto, nota-se o potencial generalizado que pode ser obtido ao desenvolver ferramentas de análise visual.

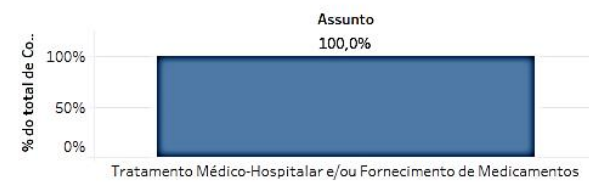
Por fim, Chung et al., (2020) aponta que existe uma escassez de ferramentas de visualização que forneçam informações abstratas para dados altamente complexos, sendo essa uma grande lacuna entre os métodos de visualização existentes. Portanto, deve-se continuar a desenvolver abordagens de análise visual, envolvendo especialistas de domínio, para entender melhor o papel das técnicas e garantir seu impacto genuíno na melhor tomada de decisão.

Na Figura 2 é demonstrado um exemplo de análise visual utilizando o software Tableau. O gráfico é interativo, apresentando as frequências de assunto, sentença e data da DRS XII Registro. Porém, é uma visualização facilmente alterável, podendo apresentar os dados de outra DRS com apenas um clique.

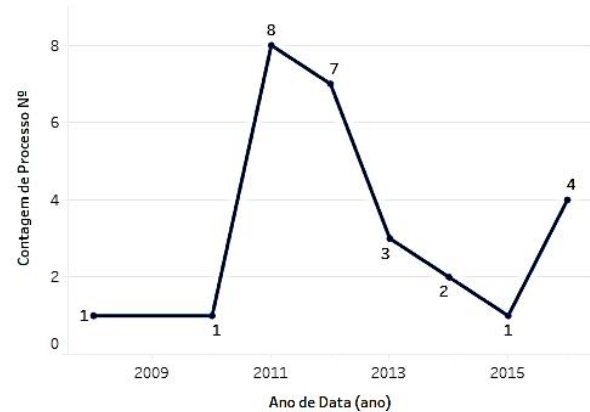
Figura 2 - Exemplo de visualização: Características dos processos da DRS XII Registro



Assunto



Data



sentença



Fonte: dados da pesquisa

2.4 ASPECTOS SOBRE JUDICIALIZAÇÃO NA SAÚDE

2.4.1 Judicialização na saúde

O termo judicialização refere-se à intervenção do poder judiciário a favor dos direitos da população. Pode ser visto como a possibilidade de efetivação do direito à saúde, entretanto, também gera um custo no Poder Judiciário, pois movimentada toda estrutura desse órgão, tais como, juízes, defensores públicos, promotores de justiça, serventuários e despesas judiciais, sendo a maioria tramitada pela justiça gratuita. Os tribunais superiores enfrentam desafios quando experimentam recursos judiciais que poderiam desestabilizar suas práticas tradicionais de julgamento (Lamprea, 2017; Lopes & Mello, 2018). Portanto, com a intervenção do poder judiciário, a judicialização se tornou uma questão de políticas públicas, mesmo que o processo tenha origem de instituições privadas.

Trata-se de um fenômeno complexo que envolve aspectos científicos, jurídicos e sociais. Pode ser resultado de diferentes fatores, tais como, deficiências do sistema de saúde, interesses

da indústria farmacêutica e/ou empoderamento do cidadão. Na literatura são encontradas duas abordagens para o tema. A normativa, tendo a judicialização como interferência do Poder Judiciário no Poder Executivo. E a social, como forma de participação cidadã (Vargas-Peláez, Rover, Leite, Rossi Buenaventura, & Farias, 2014).

Além disso, pode-se perceber ainda que o modelo processual tradicional não comporta o fenômeno da judicialização no que se refere a desigualdade social. Isso ocorre porque as demandas individuais favorecem direitos aos cidadãos que possuem conhecimento judicial, deixando aqueles com desconhecimento desse instrumento refém de falhas da implementação da política pública de saúde (Barreiro & Furtado, 2015).

Mesmo que não tenham um profundo conhecimento de política e justiça, a judicialização permitiu que a população falasse a linguagem dos direitos. Prover benefícios individuais afeta toda a estrutura de gestão de um sistema coletivo de saúde. A judicialização pode garantir esse direito como também pode não corroborar com a implantação dos princípios da universalidade, equidade e integralidade. A revisão judicial tornou-se um fenômeno global, levando alguns cientistas sociais a falar de uma "judicialização" ou "juridicalização" da política (Lopes & Mello, 2018; Merriam & Grossman, 2015).

Não é incomum o uso de judicialização com alta frequência em processos para obter terapias de alto custo para pacientes. Desde a aprovação da Constituição Brasileira de 1988, presume-se que a saúde é um dever do Estado, entretanto, o que ocorre é que não há recursos suficientes para oferecer a todos tudo o que precisam. Por outro lado, o sub financiamento e a gestão inadequada da saúde não devem moldar tais práticas como genuínas. A acessibilidade e sustentabilidade dos sistemas de saúde são uma responsabilidade compartilhada da sociedade, políticos, provedores de saúde, médicos e todos os outros múltiplos envolvidos. (Pereira da Veiga, Drummond-Lage, Da Veiga, & Wainstein, 2019).

Dentro do escopo da saúde, a judicialização pode ser utilizada para diferentes fins. Por exemplo, embora entre 2000 e 2014 o Brasil tenha tomado medidas com o objetivo de ampliar a cobertura de seus sistemas de saúde e o acesso a medicamentos, entre agosto e dezembro de 2014 a judicialização de medicamentos foi considerada um fenômeno generalizado no país. O fato surgiu independentemente da proteção constitucional ou da cobertura populacional do sistema de saúde. Na ocasião, as causas citadas foram as dificuldades em garantir o acesso aos medicamentos cobertos e a influência do marketing farmacêutico na avaliação das necessidades e comportamentos de prescrição. Além disso, também se apontou o financiamento de medicamentos sem evidências de eficácia ou segurança (Vargas-Pelaez et al., 2019). Nesse trabalho, foi abordado o ambiente de judicialização de doenças raras.

De acordo com Vargas-Pelaez et al. (2019), no contexto brasileiro, um dos impactos positivos da judicialização foi a criação de equipes técnicas para auxiliar na tomada de decisão dos juízes. A influência das políticas econômicas e sociais nacionais na judicialização era pouco conhecida. Dessa forma, indícios sugerem que a judicialização acaba beneficiando as pessoas com maior renda e/ou mais capacitadas. Portanto, nota-se que a cobertura populacional é fragmentada e depende das condições socioeconômicas do indivíduo, principalmente relacionado à sua situação de trabalho, estado de saúde e capacidade de pagamento. O beneficiário de uma ação pula a fila e passa por cima dos que seguem a via administrativa.

Embora a Constituição Brasileira estabeleça que o Estado deve garantir assistência à saúde integral, universal e gratuita, porém, com a judicialização, as decisões legais determinam que o Ministério da Saúde trate pacientes fora das diretrizes gerais, prestando tratamentos estabelecidos por critérios individuais. Dessa forma, o financiamento desse modelo é impactado por decisões judiciais que exigem o fornecimento de medicamentos e/ou tratamentos. O conhecimento do impacto dessas decisões judiciais é essencial para o desenvolvimento de estratégias que garantam a oferta de tratamentos baseados em evidências e protejam a sustentabilidade do sistema (Pereira, Miranda Kretzschmar, Tonini, Xavier, & Barros, 2020).

Por exemplo, no estudo realizado por Pereira et al., (2020), concluiu-se que nas distribuições do medicamento Miglustate no ano de 2019 as decisões judiciais representaram 38% dos gastos do Ministério da Saúde com Miglustato. Além dos custos inesperados, também se tem o risco de submeter pacientes a tratamentos sem orientação sistemática em protocolos não oficiais. A qualificação técnica do judiciário é exigida na perspectiva da medicina pública e baseada em evidências.

A judicialização apresenta características imprevisíveis se tratando de questões orçamentais e financeiras e, conseqüentemente, traz problemas para as políticas públicas de saúde. Siena, Carvalho, Netto e Silva (2021) observaram em seu estudo que, embora tenha ocorrido uma redução na quantidade de ações no estado de Rondônia nos anos de 2015 a 2018, os custos anuais da judicialização na saúde aumentaram.

Portanto, nota-se que uma grande quantidade de ações não, necessariamente, reflete em um custo elevado à saúde pública. Uma possível explicação para tal ocorrência seria demandas por medicamentos mais caros e fora da tabela SUS ou para procedimentos de maior complexidade. Logo, observa-se uma contradição ao princípio da universalidade do direito à saúde, pois, conforme já destacado anteriormente, recursos que estavam planejados para atender a população geral, passa a privilegiar demandas individuais, caracterizando como uma das principais desvantagens da judicialização para população como um todo (Siena et al., 2021).

Além disso, a complexidade dessa discussão precisa considerar a rigidez do processo orçamentário brasileiro, levando em conta questões como flexibilidade do orçamento público, o qual já possui uma destinação pré-determinada e outra parcela comprometida com despesas obrigatórias, valores que não podem ser alterados dentro da anualidade orçamentária. Portanto, os valores que o administrador público tem autonomia para aplicar as políticas públicas de saúde são limitados, pode-se afirmar apenas que as demandas judiciais têm uma participação expressiva no orçamento que está disponível para o gestor (Siena et al., 2021).

Porém, no médio e longo prazo as despesas de judicialização da saúde podem comprometer o orçamento público, pois os recursos repassados anualmente já são insuficientes para as despesas totais, necessitando de frequentemente de contemplações. Assim os indícios apontam que, considerando o orçamento total da saúde, a judicialização não traz impactos negativos, mas em termos de valores destinados aos gestores administrativos, o impacto é relevante (Siena et al., 2021).

Nesse sentido, vale destacar o impacto do setor privado nessa temática, tais como, inovações tecnológicas para aumento da expectativa e qualidade de vida e interesses do mercado farmacêutico, os quais não podem sobrepor o bem-estar social. Estudos apontam que existe um interesse da indústria farmacêutica na judicialização da saúde, mas, nem sempre, os medicamentos fornecidos atendem da melhor maneira as necessidades dos pacientes. Dessa forma, associações entre saúde pública com instituições pautas em bases científicas, é um caminho eficaz para o equilíbrio entre custo e oportunidade (Bretas, Ferreira Junior, & Lucena Ruas Riani, 2021).

Ainda sobre o sistema privado, pacientes com acesso a planos particulares afirmam ter acesso a melhores condições de atendimento de saúde. Porém, tais fornecedoras de assistência médica suplementar tem sido alvo de investigações e inspeções, por questões como reclamações de valores abusivos, tratamentos insatisfatórios e até recusar de atendimento. Antes de optar pela judicialização, os clientes podem optar por canais como o Programa de Proteção e Defesa do Consumidor (Procon), a Agência Nacional de Saúde Suplementar (ANS) e as próprias operadoras dos planos de saúde. Os autores concluíram que quando se trata de procedimentos médicos, os tribunais asseguram imediatamente a cobertura procurada por meio de uma medida cautelar, deixando o mérito da causa para outro momento (Ryngelblum, Oliveira, Telles, & Costa, 2019).

2.4.2 Big data no campo da saúde

Os serviços de saúde pública surgiram no Brasil com a criação de uma Direção Geral de Saúde no final do século XIX. A partir disso, iniciaram movimentos em prol de uma reforma na saúde brasileira envolvendo diversos segmentos da sociedade, tais como, intelectuais, pesquisadores, organizações de trabalhadores e partidos políticos, baseados no conceito de universalidade e igualdade de direitos (Younger, 2016).

Atualmente, pode-se dizer que a área enfrenta problemas relacionados ao desequilíbrio de ofertas de serviços e as necessidades da população. Diante disso, intelectuais, pesquisadores e demais profissionais da área, passaram a buscar alternativas de melhores atendimentos diante de poucos recursos. Nesse sentido, a aplicação de técnicas de big data, a qual tem crescido em todas as áreas da ciência nos últimos anos, pode contribuir com importantes ganhos em termos de dinheiro, tempo e vidas. Por exemplo, na medicina de precisão, em prontuários eletrônicos de pacientes, e com a utilização da internet das coisas, encontram-se oportunidades promissoras para o uso de Big Data (Chiavegatto Filho, 2015).

No que se refere a medicina de precisão, o objetivo é identificar, precisamente, o tratamento adequado de forma individual, ou seja, em vez de prescrever o mesmo medicamento para todos que possuem o mesmo sintoma, espera-se indicá-lo apenas para quem o medicamento de fato funcionará. Quanto aos prontuários eletrônicos de pacientes, no Brasil, ainda são utilizados prontuários digitalizados específicos para cada unidade. Em muitos casos, utiliza-se em papel, o que dificulta a transferência, a atualização e a compreensão das informações. O fato dos prontuários não poderem ser acessados por outros centros de saúde traz perdas de tempo, dinheiro e vidas. Por fim, a internet das coisas promete que a maioria dos objetos de uso diário estarão, de alguma forma, conectados à internet. As possibilidades de uso na saúde são imensas. Por exemplo, o uso de *wearables* conectados ao corpo que poderão identificar um risco de infartos e acidentes vasculares antes mesmo da própria pessoa (Chiavegatto Filho, 2015).

Dessa forma, Galetsi et al. (2020) listou entre técnicas de big data quais são as mais empregadas para pesquisas em saúde. Em seu estudo, identificou a modelagem como a mais utilizada. Trata-se de métodos de análise matemática rápidos e econômicos com relações aproximadas entre variáveis. Sequencialmente classificou o aprendizado de máquina, uma inteligência artificial destinada a projetar algoritmos que permitem que os computadores desenvolvam comportamentos baseados em dados empíricos. Posteriormente o autor listou a mineração de dados, técnica que será empregada no presente estudo, seguido das abordagens de visualização, técnicas usadas para criar tabelas, imagens, diagramas e outras formas de

exibição intuitiva para entender os dados. Vale afirmar que entre as linguagens de programação mais utilizadas em big data, R[®] e Python[®] têm se destacado (Chiavegatto Filho, 2015).

Além disso, espera-se que os aprimoramentos das técnicas permitam que os pacientes possam gerenciar a sua própria saúde de forma autônoma através do uso de aplicativos. Tal premissa se sustenta pela existência de plataformas e software baseados em nuvem, por exemplo, para compartilhamento e recuperação de imagens médicas e registros de saúde a partir de dados estruturados (Galetsi et al., 2020).

A possibilidade de algoritmos responderem e se adaptarem automaticamente aos dados sem a necessidade de intervenção humana contínua, é um desejo desde a Revolução Industrial. Organizar e interpretar dados aplicando técnicas estatísticas, usar simulação para análise quantitativa de um sistema em um cenário randômico, além de métodos de otimização que melhoram a precisão dos problemas de previsão seguindo estratégias computacionais, entre outros, são formas de aproximarmos dessa expectativa. Porém, é válido lembrar que os avanços também trazem preocupações. Nesse caso, o grande volume de dados dificulta diferenciar entre a importância e qualidade diante de muita quantidade. Ademais, outro desafio do uso de big data trata-se da privacidade, com risco de vazamento de dados sigilosos por descuido de cientistas ou por invasões criminosas (Chiavegatto Filho, 2015; Galetsi et al., 2020).

2.4.3 Saúde no Brasil e no estado de São Paulo, e doenças raras

Ao longo dos anos, os serviços de saúde brasileiro passaram por mudanças e ajustes. A Constituição de 1988 garante o direito e a responsabilidade do Estado de fornecer o SUS como rede de serviços de saúde em cada nível de governo. Também estabelece novas fontes de receita para a seguridade social, por meio de contribuições obrigatórias vinculadas à receita bruta e ao lucro líquido das empresas (Younger, 2016).

Considerando tais mudanças ocorridas, a Lei Eloy Chaves, criada em 1923, destaca-se por ter proporcionado um sistema de seguridade social para os trabalhadores urbanos empregados no setor privado. No entanto, o acesso aos serviços não eram o principal objetivo do sistema de saúde, em vez disso, havia um sistema de cidadania regulada pelo qual os direitos sociais, incluindo aposentadorias e cobertura médica, eram restritos aos trabalhadores do setor privado que recebiam salários regulares. (Younger, 2016).

Mais adiante, ocorreram movimentos em prol de uma reforma na saúde do país, por exemplo, a assinatura da Constituição da Organização Mundial da Saúde em 1946. Além disso, o sistema passou a ser dividido em dois modelos básicos de assistência. A medicina liberal ou

privada, operando no mercado e a medicina administrada pelo governo em hospitais públicos e clínicas. Na década de 1970, o primeiro passo da reforma do sistema de saúde foi estender a cobertura para determinados serviços, começando com cuidados de urgência e emergência (Younger, 2016).

De acordo com a estimativa do IBGE (2018), o estado de São Paulo possui a maior população do país. Estados populosos convivem com diversos riscos para saúde, tais como, lesões de trânsito, alta poluição do ar e comprometimento de níveis básicos de atividade física. Porém, o campo da saúde enfrenta desafios em todo Brasil. Por exemplo, com a aprovação da Emenda Constitucional 95/2016, iniciado em 2018 no campo da saúde, se fixou para os próximos vinte anos um limite de recursos públicos para despesas primárias, ou seja, o montante das despesas totais antes do pagamento dos juros da dívida. No período de 2010 a 2015 a saúde foi uma das funções que mais consumiu tais recursos (Mariano, 2017; Sá et al., 2017; Vairão Junior & Alves, 2017).

Dessa forma, não surpreendentemente, São Paulo possui um histórico de participação em momentos relevantes na área da saúde no país. O estado paulista participou da centralização e descentralização que antecederam a criação do principal sistema de saúde brasileiro, o SUS. Ademais, evidencia-se a importância que a dimensão territorial vem tomando no processo de regionalização de políticas de saúde. É necessário que seja prioridade disponibilizar tais serviços para todos os cidadãos, em especial nos municípios que compõem regiões em que grande parte das necessidades não são atendidas, exigindo-se a presença de uma quantidade mínima de ações e serviços (Mendes et al., 2015; Younger, 2016).

Em 2001, uma emenda constitucional reverteu o sistema de financiamento do setor da saúde às receitas gerais e o governo federal precisou alocar e gastar um montante equivalente ao orçamento do ano anterior ajustado para o produto nacional bruto. O crescimento médio foi de 2,4%, usando o orçamento de 1999 como base. Desde a Emenda Constitucional (EMC) 29/2000, os municípios têm sido legalmente obrigados a investir 15% de todo o seu orçamento em saúde (Dantas et al., 2017; Younger, 2016). A EMC 29/2000 define que os recursos mínimos aplicados nas ações e serviços públicos de saúde devem ser equivalentes, no caso dos estados e do distrito federal, a doze por cento do produto da arrecadação dos impostos a que se refere o art. 155 e no caso dos municípios e do distrito federal, a quinze por cento do produto da arrecadação dos impostos a que se refere o art. 156.

Porém, a esfera estadual de poder ainda apresenta dificuldades. Nesse sentido, a construção histórica de apoios técnicos regionais podem potencializar a associação solidária dos gestores na construção das regiões que não tem sido fortalecida no estado de São Paulo

(Mendes et al., 2015). Mendes et al., (2015) realizaram um estudo com depoimentos de gestores da saúde em relação ao governo do Estado. Em sua maioria, os relatos descreveram como um órgão distante, autoritário, puramente burocrático-administrativo, isso quando não dificulta o processo dos serviços hospitalares.

Além disso, conforme mencionado, a existência de uma diversidade de processos de judicialização, indica a necessidade de mais atenção à saúde. Portanto, é preciso melhores investimentos, os quais são necessários tanto na área da saúde como em uma abordagem de políticas integradas. Por exemplo, as políticas de meio ambiente podem melhorar o desempenho em saúde, especialmente quanto a saneamento, medidas preventivas correlacionadas a doenças infecciosas e até condições degenerativas crônicas que sofrem influência significativa do meio ambiente (Dantas et al., 2017; Lopes & Mello, 2018).

Outro ponto a se considerar, é a frequência que a população faz uso dos serviços públicos de saúde, o que influencia não apenas tais serviços, como também os que são ofertados de forma privada. Embora a área da saúde necessite de mais atenção, diversos programas são eficazes e a sua utilização estimula o aumento de investimentos. Um deles é a Estratégia Saúde da Família (ESF) do Ministério da saúde que busca promover a qualidade de vida da população. O programa tem contribuído para o declínio da mortalidade nos níveis nacional e regional, inclusive em São Paulo. Visitas domiciliares pós-natal fazem parte da ESF no Brasil, além das bases comunitárias de prestação de serviços de saúde que também podem ser uma maneira altamente eficaz de reduzir a mortalidade neonatal, principalmente em áreas urbanas com baixa e média renda (Brentani et al., 2016).

O Programa de Saúde da Família (PFS) foi implantado em 1994 sendo um modelo de saúde do governo para reestruturar a atenção primária no SUS, hoje é conhecido como ESF. Uma característica inovadora do modelo foi a reorganização de clínicas primárias que enfocam nas famílias e comunidades, e a integração de cuidados médicos, com ações de promoção da saúde pública. A equipe de saúde da família é composta por médico, enfermeiro, auxiliar de enfermagem e agentes comunitários de saúde (Younger, 2016).

Corroborando com Brentani et al. (2016), Guimarães, Lucas e Timms (2019) apontam que a falta de acessibilidade pode sustentar as desigualdades na saúde com ônus, geralmente, carregado por grupos socialmente desfavorecidos que experimentam níveis mais altos de doenças e, conseqüentemente, têm vidas mais curtas. As principais barreiras à acessibilidade envolvem a distância, segurança, serviços de transporte público e qualidade dos serviços de saúde. Na cidade de São Paulo, mesmo sob severas restrições financeiras e de tempo, as pessoas optam por viajar mais tempo para acessar instalações consideradas mais adequadas. Porém,

outros pontos também são considerados como desafiadores, por exemplo, a má qualidade do relacionamento com o paciente e os longos tempos de espera para atendimento médico. O combate às desigualdades na saúde exige a implementação de políticas integradas de transporte e saúde, levando em consideração a adequação e a qualidade desses serviços.

Além disso, outras questões relacionadas com consequências de residir em estados populosos, tais como, estresse e violência, resultando em altos níveis de ansiedade e depressão, merecem destaque. Fillenbaum (et al., 2019) realizaram um estudo na cidade de São Paulo para analisar se áreas propensas à violência fazem maior uso de serviços de saúde mental. Entretanto, o uso foi maior entre os moradores de áreas com menor índice de homicídio e não foi associado à experiência traumática pessoal, sugerindo uma maior necessidade de informações e divulgação a respeito da disponibilidade e benefícios de tais serviços.

Além disso, outras áreas da saúde também precisam de maior atenção e divulgação de informações, como as doenças raras. O reconhecimento de um grupo de doenças como “doenças raras” é uma tentativa para facilitar a inclusão em uma categoria apropriada quando for necessário. Tais necessidades se referem ao reconhecimento das diferenças individuais e a defesa de uma medicina personalizada, na tentativa de permitir a individualidade dos pacientes (Hasani et al., 2022).

As doenças raras afetam menos de uma em cada 2.000 pessoas, o que significa que são muito menos visíveis na população em geral do que as doenças comuns. Tem várias consequências negativas para os pacientes e para os sistemas de saúde. Por causa desse efeito negativo, os programas e políticas de saúde devem ser bem estruturados. No entanto, os esforços dos países nesta área devem ser avaliados para garantir a eficácia (İnce & Güre, 2021).

Portanto, diante da necessidade de avanços na área de DRs, os métodos de inteligência artificial têm se destacado. Porém, para alcançar bons resultados, é necessário que as partes interessadas estejam amplamente envolvidas, tais como, pacientes, médicos e todos os provedores relevantes, como os sistemas de saúde, agências regulatórias, sociedade e profissionais. É necessário o envolvimento inclusivo, transparente e ativo dessas partes em todo o desenvolvimento, projeto, vigilância e pós-implantação de sistemas de inteligência artificial na área (Hasani et al., 2022).

İnce e Güre (2021) avaliaram o desempenho das políticas de 18 países da Organização para a Cooperação e Desenvolvimento Econômico (OECD) em termos de doenças raras, usando indicadores de políticas de saúde. Identificaram que a Alemanha, França e Holanda estão entre os países com as políticas mais fortes. Letônia, Grécia e Turquia estão entre os países com pior desempenho na classificação. Portanto, são necessários estudos de avaliação de desempenho

para a prestação de serviços e arranjos de políticas no campo das doenças raras e isso inclui pesquisas para o Brasil.

Atualmente já existem softwares avançados para soluções médicas. Porém, embora tais iniciativas sejam o primeiro passo essencial para melhorias na DRs, a responsabilidade não termina aí. Deve-se incluir a prevenção de abusos em potencial estando ciente dos potenciais efeitos a jusante e implicações éticas da pós-implantação. Apesar de sistemas que permitem o monitoramento e detecção contínua e precoce das doenças seja clinicamente benéfico, essa informação pode ser, intencional ou não, mal utilizada no contexto de generalizações excessivas por sistemas de saúde. Ou seja, pode causar transtornos com os pacientes ao oferecer informações que não eram previamente esperadas (Hasani et al., 2022).

3 PROCEDIMENTOS METODOLÓGICOS

Nesse capítulo são descritos os procedimentos metodológicos, especificamente a respeito da caracterização da pesquisa, coleta e amostra, e as técnicas utilizadas para análise dos dados obtidos.

3.1 Caracterização da pesquisa

Trata-se de um estudo quantitativo de natureza exploratória e descritiva. A pesquisa analisou processos de doenças raras do estado de São Paulo, em busca de padrões e relações entre os atributos encontrados. O intuito dessa escolha se deve ao fato do estado ter a maior população do país e ser o segundo em rendimento nominal mensal domiciliar per capita (IBGE, 2018). Desta forma, o método de pesquisa exploratória foi importante por fornecer informações descritivas sobre as características contidas nesses documentos. A escolha dessa abordagem justifica-se pelo fato da pesquisa exploratória ter como objetivo principal, fornecer informações iniciais e compreensão sobre o problema estudado (Malhotra, 2008).

Apesar do estudo não apresentar hipóteses específicas e essa ser uma característica da abordagem descritiva, enquadra-se na técnica por apresentar um enunciado claro do problema e necessidade detalhada de informações (Malhotra, 2008). Portanto, justifica-se a escolha da abordagem quantitativa descritiva. Além disso, seria inviável, considerando que se trata de 35.430 processos, a obtenção de informações em profundidade, sendo essa uma limitação da abordagem de pesquisa utilizada (Hair, Wolfinbarger, Ortinau, & Bush, 2008; Malhotra, 2012).

Ademais, encontrou-se muitos resultados numéricos e pontuais, justificando a natureza quantitativa do estudo. Embora as generalizações em pesquisas exploratórias possam ser comprometidas, tal método ajuda a decifrar, examinar e interpretar os padrões mais importantes que surgem a partir dos dados (Malhotra, 2008). A fim de construir um trabalho que não privilegie generalizações abstratas nem especificidades singulares, optou por conciliar os dois métodos.

Essa pesquisa associou a estrutura dos padrões encontrados nos documentos, a diferentes processos com intuito de adquirir medicamentos e/ou tratamentos variados. Porém, todos se enquadram como judicialização na saúde, fornecendo informações a partir dos dados do estado de São Paulo, os quais auxiliam na compreensão das relações dessa área, sendo de interesse para acadêmicos e profissionais de campos variados. O Quadro 1 apresenta as principais abordagens da pesquisa.

Quadro 1 – Abordagens

| OBJETIVO GERAL | OBJETIVOS ESPECÍFICOS | PROPOSIÇÃO | TEORIAS APLICADAS | TÉCNICAS DE ANÁLISE |
|---|---|---|--|--|
| Identificar padrões em dados públicos de saúde do estado de São Paulo, a fim de descrever as características mais relevantes, encontradas em processos de doenças raras, utilizando técnicas de <i>visual analytics</i> . | Identificar as características mais recorrentes nos processos públicos de doenças raras do estado de São Paulo. | As características mais recorrentes nos processos públicos de doenças raras do estado de São Paulo, são semelhanças que levam a quais informações relevantes? | Mineração de texto (Benchimol, Kazinnik, & Saadon, 2022) | Extração das informações principais contidas nos processos. |
| | Obter informações, a partir de mineração de dados, e localizar os principais padrões de movimentações nos processos públicos de portadores de doenças raras. | Os processos públicos de portadores de doenças raras possuem semelhanças suficientes para segmentá-los em padrões? | Mineração de dados (Van Der Aalst, 2019) | Extração de informações de movimentações contidas nos processos, a partir de padrões e repetições. |
| | Apresentar, a partir de técnicas de análise e visualização de dados, as informações mais relevantes contidas nos processos de forma simplificada, a fim de permitir a | As informações organizadas em <i>dashboards</i> , irão suprir a expectativa de fornecer informações relevantes para profissionais e demais interessados da área da saúde? | <i>Visual analytics</i> (Cleveland & McGill, 1984) | Adequação das informações resultantes em <i>dashboards</i> , a partir de técnicas de <i>visual analytics</i> . |

| | | | | |
|--|-------------------------------|--|--|--|
| | compreensão do público geral. | | | |
|--|-------------------------------|--|--|--|

Fonte: Elaborado pela autora

3.2 Coleta dos dados e amostra

Para os fins desse estudo, foram utilizados dados secundários de domínio público, coletados e extraídos em uma única etapa, referentes ao período de janeiro de 2013 a dezembro de 2017, extraídos dos Repositórios de Jurisprudência e dos Diários Oficiais da Justiça, e armazenados eletronicamente, resultando em um *dataset* para essa pesquisa. O intuito dessa escolha se deu pela credibilidade de tais fontes, pois os Repositórios de Jurisprudência possuem o registro dos repositórios autorizados e credenciados da jurisprudência do Superior Tribunal de Justiça (STJ) e os Diários Oficiais da Justiça, são os diários oficiais da União, dos Estados e dos Municípios.

Buscando garantir que a Lei Geral de Proteção de Dados Pessoais (LGPD), Lei nº 13.709/2018, a qual refere-se à legislação brasileira que regula as atividades de tratamento de dados pessoais, fosse respeitada nessa pesquisa, garantiu-se que a coleta seria restrita a dados públicos e que nomes de pessoas físicas não seriam divulgados.

Tais documentos foram obtidos durante a pesquisa financiada pelo Conselho Nacional de Justiça do grupo de pesquisa Habeas Data, interno da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto (FEA-RP/USP) o qual participo. O grupo é composto por diversos pesquisadores, os quais desempenham papéis com propósitos diferentes dentro das necessidades de pesquisa. Dessa forma, os dados foram coletados por outro pesquisador, por pesquisa automatizada, e fornecidos para o desenvolvimento desse trabalho.

Os documentos em questão, correspondem a processos da área de saúde do estado de São Paulo. Para alcançar o objetivo de analisar as demandas de portadores de doenças raras, foram listadas mais de 200 enfermidades classificadas como doenças raras. Para identificar tais doenças, foram consultadas algumas fontes e optou-se pela lista mais completa dentre os materiais encontrados, disponibilizado pela Organização Não Governamental (ONG) Health On the Net. Dessa forma, criou-se uma lista de palavras-chaves para identificação e extração dos processos de interesse. Por meio de técnicas de linguagem natural, a pesquisa verificou a presença das palavras chaves nos assuntos dos documentos.

Os documentos foram armazenados no software MongoDB, o qual se trata de um banco de dados de código aberto, escrito na linguagem C++. Por ser classificado como um programa de banco de dados NoSQL, posteriormente foi utilizado o SQLite, uma biblioteca em linguagem C que implementa banco de dados SQL embutido. Assim como o MongoDB, programas que usam a biblioteca SQLite podem ter acesso a banco de dados SQL sem executar um processo de sistema de gerenciamento de banco de dados separado. Por fim, a coleta resultou em 48.599 processos dentro dos filtros mencionados. Os documentos foram salvos no formato CSV.

Foi adotada uma amostra por conveniência, um tipo de amostragem em que a probabilidade de seleção de cada objeto é desconhecida. Nesse método, são selecionados os objetos mais acessíveis ao pesquisador. A técnica apresenta vantagens, como a possibilidade de conseguir a amostra em pouco tempo, mas também possui desvantagens, como o risco de não atingir uma representatividade (Hair et al., 2008). A amostra do estudo possui 315 cidades e não foram encontradas outras localidades com as palavras-chaves pré-determinadas na coleta.

Além dessa etapa, foi necessário a identificação dos movimentos dos processos, para extrair informações a partir de mineração de dados. Os movimentos referem-se a datas em que os processos executaram alguma ação, por exemplo, da distribuição do processo para remeça de 2ª instância. Para alcançar esse objetivo, foi realizada uma nova coleta, nas mesmas fontes, com o intuito de buscar essas informações adicionais. Para isso, utilizou-se a técnica de regex para extrair o número dos processos e, assim, criar uma lista dos processos de interesse. Regex é uma abreviação de *regular expression*, traduzido como expressão regular, cujo objetivo é identificar um conjunto de caracteres de interesse ou padrões específicos de forma automatizada, concisa e flexível (Dewi, Meiliana, & Chandra, 2019).

Dessa forma, a partir da lista dos processos foi refeita a coleta, incluindo as informações de movimento e também de valor das ações. Esse procedimento ocorreu em outubro de 2021 e, por essa razão, a base contém informações do ano de 2021, embora a coleta inicial seja de janeiro de 2013 a dezembro de 2017. De posse dessas informações e com o auxílio do software Fluxicon Disco, foi possível realizar a análise por mineração de processos e, dessa forma, atender o objetivo específico B: obter informações, a partir de mineração de dados, e localizar os principais padrões de movimentações nos processos públicos de portadores de doenças raras.

O Fluxicon Disco é uma ferramenta inovadora que permite gerar percepções visuais e acionáveis por mineração de processos a partir de dados brutos em poucos minutos, criando mapas visuais. A ferramenta foi projetada para adequar ao fluxo de trabalho dos profissionais, considerando, o desempenho, o controle, desvios ou variações de forma inteligente. O Disco contém os algoritmos de mineração de processos mais rápidos e a estrutura de gerenciamento

e filtragem de logs mais eficiente. Além das razões já mencionadas, o software foi escolhido principalmente por proporcionar uma experiência de usuário simplificada, permitindo modificações rapidamente (“Fluxicon Disco.,” n.d.)

A classificação das sentenças e a respectiva porcentagem de probabilidade de acerto, foram concluídas na coleta inicial de forma automatizada. Na segunda etapa de coleta, foram mantidas as sentenças publicadas e, assim, foi possível avaliar a eficiência da ferramenta automatizada para classificação de sentenças. Por fim, foi criada uma nova coluna com os nomes das doenças identificadas de forma automatizada por membros do Habeas Data. Portanto, resultou nas seguintes divisões finais: Assunto, Autor Gênero, Autor Nome, Autor Pessoa, Classe, Conteúdo, Dados do fórum, Data, Nº Processo, Réu Nome, Réu Pessoa, Réu Gênero, Sentença, Classificação da Sentença, Probabilidade das Sentenças, local, DRSs, Valor, Movimentos, Doenças e Id.

3.3 Análise de dados

Conforme desenvolveu-se a análise, foram identificadas algumas necessidades de mudanças nos padrões dos dados. A princípio, tratava-se de 316 arquivos CSV., onde cada um representava uma quantidade de processos variados de uma cidade específica. Porém, um dos arquivos era um compilado de processos, cuja as cidades não foram encontradas, esses dados foram excluídos. Para facilitar a análise, foi feita uma padronização no software Tableau de todas as informações dos processos, resultando, inicialmente, nas seguintes divisões: Assunto, Autor Gênero, Autor Nome, Autor Pessoa, Classe, Conteúdo, Dados do fórum, Data, Nº Processo, Réu Nome, Réu Gênero, Réu Pessoa, Sentença, Probabilidade das Sentenças, e id.

O Tableau é uma ferramenta que contribui aprimorando a tomada de decisão, aumentando a conscientização operacional e o desempenho em toda a organização. Possui diferentes alternativas para o campo de *business intelligence*, sendo uma escolha amplamente usada para fins de visualização de dados em diferentes organizações. Porém, não suporta aspectos de automatização de relatórios com a ajuda de agendamento. Portanto, sempre há algum esforço manual necessário quando os usuários precisam atualizar os dados e os parâmetros, além de não ser uma ferramenta totalmente aberta (Jena, 2019).

Em sequência, foram adicionadas a cada um dos arquivos uma nova coluna com o nome da cidade e, em seguida, com o software RStudio, realizando o comando merge, criou-se um único arquivo, o qual continha todas as cidades. O software R[®], está disponível para aquisição gratuita e possui milhares de pacotes de vários autores em diversas áreas, atualizados quando

novas contribuições são recebidas e acompanhados de um tutorial. Atualmente, o R possui mais de 12.700 pacotes acessíveis em vários repositórios. No entanto, essa magnitude de opções faz com que muitos usuários não estejam cientes dos pacotes específicos disponíveis para sua área de interesse (Sousa et al., 2020). Portanto, torna-se atraente para utilização em variadas pesquisas e além disso, também tem sido utilizado para analisar conteúdo textual, identificando termos de grandes ocorrências em diferentes documentos e as co-ocorrências (termos com outros termos). Para analisar um determinado texto, envolve o uso de várias funções buscando eliminar caracteres e termos indesejáveis, tais como, pontuação, espaços em branco e palavras de conexão. Algumas plotagens e gráficos podem ser usadas para aprimorar visualmente o conteúdo analítico (Nguyen, 2013). Tais capacidades, vão de encontro com as necessidades dessa pesquisa.

Considerando que era um total de 315 cidades, foi necessário incluir uma nova divisão para facilitar a compreensão dos resultados nos *dashboards* de visualizações. Dessa forma, optou-se pela divisão dos Departamentos Regionais de Saúde (DRS) do estado de São Paulo, os quais estão disponíveis no site estadual, do Governo do Estado de São Paulo, da Secretaria de Estado da Saúde (Governo do Estado de São Paulo).

As cidades são divididas em 17 regionais, alguns distritos não possuem seu nome incluídos na divisão, assim, foram classificados na DRS da cidade que pertence. O banco de dados possuía informações de pelo menos uma cidade pertencente a cada regional, portanto, foram incluídas as 17 DRS na pesquisa.

A fim de evitar possíveis distorções nos resultados, devido a uma maior população concentrada em determinadas DRSs, incluiu-se nos dados a variável população, a qual representa a soma populacional das cidades coletadas incluídas em cada DRS. A variável foi definida a partir da listagem da população de todas as cidades do estado de São Paulo. De forma semelhante, também foram coletadas informações sociodemográficas, a fonte de ambas foi o IBGE (“Censo IBGE estado de São Paulo,” n.d.), a fim de atender o objetivo específico A, cujos detalhes estão especificados no próximo capítulo.

Com essa lista e o auxílio do editor de planilhas Microsoft Excel, utilizou a funcionalidade PROCV, a qual permite localizar dados em linhas de uma tabela ou de um intervalo específico. Dessa forma, ordenando alfabeticamente e somando os valores populacionais, obteve-se a informação necessária para cada DRS. Por fim, optou pela utilização dos dados em porcentagem, convertido automaticamente pelo Tableau, sendo considerado 100% a soma das populações de todas DRSs.

Os procedimentos descritos até aqui foram focados nas análises a partir de técnicas de BA, ou seja, tecnologias que dão suporte a tomadas de decisões a partir de análise de dados, possuindo como diferencial a modelagem preditiva (Wang et al., 2019). Nesse sentido, a análise pode seguir por 3 focos diferentes: análise descritiva, identificando os pontos principais nos dados, análise preditiva, a fim de encontrar indícios de probabilidade de resultados futuros, e uma análise prescritiva, propor a partir do conhecimento atual recomendações para situações semelhantes no futuro (Zhao et al., 2020).

Portanto, os procedimentos descritos até esse ponto buscaram atender ao objetivo específico A: identificar as características mais recorrentes nos processos públicos de doenças raras do estado de São Paulo. Por exemplo, semelhanças em características sociodemográficas, e padrões de tratamentos.

Considerando que o presente estudo empregou a utilização de mineração de dados, pois buscou encontrar padrões, as técnicas de BA voltadas para análises descritivas são mais favoráveis. Além disso, justifica-se essa escolha por, em geral, a mineração de dados ser classificada em duas categorias: mineração de dados supervisionadas (tarefas de previsão ou de classificação) e mineração de dados não supervisionadas (busca relações e interconectividade de dados) (Witten et al., 2017; Zhao et al., 2020). Nesse sentido, foram necessárias ambas abordagens.

O tratamento do banco de dados caracteriza-se como uma mineração de texto, a qual tem como objetivo principal, capturar e analisar todos os significados possíveis incorporados no texto, nesse caso, nos processos. Adotou uma abordagem quantitativa que, em geral, é um campo de pesquisa que estuda a capacidade de decodificar dados da linguagem natural com ferramentas computacionais (Benchimol et al., 2022).

A análise quantitativa de texto tem raízes em um conjunto de métodos computacionais simples, focados em quantificar a presença de certas palavras-chave ou conceitos em um texto. A abordagem comum na mineração de texto quantitativa é encontrar relações entre conceitos, gerando o que é conhecido como rede semântica. A análise de redes semânticas é caracterizada por sua capacidade de ilustrar as relações entre palavras dentro de um texto, fornecendo percepções sobre sua estrutura e significado. Como resultado, a análise da rede semântica permite que o significado seja revelado considerando as relações entre os conceitos (Benchimol et al., 2022).

Inicialmente, foram identificadas as ocorrências de categorias referentes ao acontecimento de doenças raras nos documentos, levando em consideração sua pertinência. Em seguida, foram definidas as categorias com suas descrições. As informações expressas nos

documentos foram classificadas buscando identificar a frequência, o vocabulário e os conceitos das palavras para interpretar os significados das informações.

Posteriormente, foram organizadas as informações contidas no banco de dados resultante e avaliado se já estava no momento de partir para construção dos resultados ou se ainda era necessário coletar mais alguma informação. Conforme descrito anteriormente, foram incluídas, após a coleta inicial, informações a respeito das movimentações, valores das ações, doenças, DRSs, população e características sociodemográficas. Por fim, foi feita a interpretação dos dados, indicando como os dados estão relacionados com a temática de doenças raras.

Dessa forma, concluiu-se o último objetivo específico C: Apresentar, a partir de técnicas de análise e visualização de dados, as informações mais relevantes contidas nos processos de forma simplificada, a fim de permitir a compreensão do público geral.

3.4 Dados ausentes

As observações válidas coletadas passaram por algumas análises com o intuito de exclusão dos casos inválidos e obtenção de resultados adequados para a pesquisa. Nesse tópico, serão apresentadas as características da pesquisa no que se refere à dados ausentes. Dados ausentes, correspondem a informações específicas que não estão disponíveis para um determinado caso. Tais dados, tratam-se de uma realidade em pesquisas com análises multivariadas, sendo necessário compreender os processos que conduziram a essas falhas para selecionar a ação apropriada para tal. Quando a razão da ocorrência de dados ausentes é desconhecida, deve-se tentar identificar padrões nos dados para caracterizá-los (Hair, Joseph F., Black, Jr, William C. Babin, Barry J. & Anderson, 2005).

Nesse sentido, como se trata de um estudo com processos, o mais provável é que as ausências sejam consequências de erros na ferramenta de coleta ou indisponibilidade da informação, por exemplo, dado em segredo de justiça. Inicialmente a base possuía 48.599 processos, porém, após incluída as datas de movimentações, foram excluídos os processos que não possuía essas informações, restando 35.430. Dentre essas exclusões, 207 (0,42%) processos não foram encontrados e 3.206 (6,60%) estavam em segredo de justiça. Nos 9.756 (20,07%) restantes, alguns são erro de coleta ou o número do processo correspondia a segunda instância e alguns estavam com linhas vazias. Observou-se que os dados perdidos foram pontuais, totalizando aproximadamente 27% da base original.

Portanto, após a identificação dos dados ausentes, buscou-se verificar se havia algum padrão nesses dados. Observou-se que os dados perdidos foram pontuais. Dessa forma, não foi

possível utilizar os processos com informações incompletas para um tratamento de dados ausentes e uma posterior contribuição com os resultados da pesquisa, pois para atingir os resultados esperados, era necessário que todos os processos possuíssem informações sobre movimentações.

4 ANÁLISE DOS RESULTADOS E DISCUSSÃO

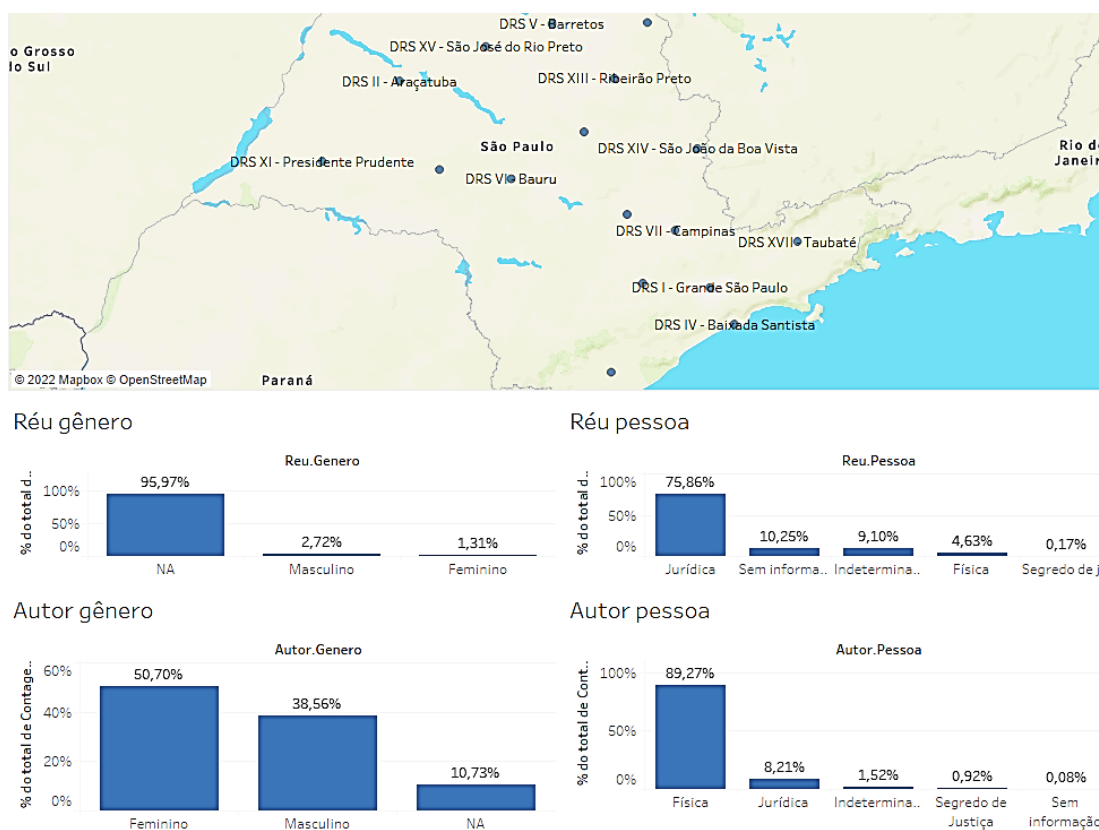
Neste capítulo são apresentados os resultados que foram obtidos com a análise da base de dados, além da discussão com foco em suas implicações teóricas.

4.1 Características gerais da base de dados

No que se diz respeito às características gerais, a base possui 35.430 processos distribuídos em 315 cidades paulistas. Os autores desses processos são a maioria, 17.736, do sexo masculino, 13.686 são do sexo feminino e 4.008 NA (não informado). Além disso, 31.422 processos foram abertos por pessoa física, enquanto 3.094 foram por pessoa jurídica, 334 estavam em segredo de justiça, 558 indeterminados e em 32 a informação não foi encontrada.

Já no que se refere aos réus, são a maioria, 1.189, do sexo masculino, 617 são do sexo feminino e 46.793 NA. Ademais, apenas 2.152 são pessoas físicas, enquanto 34.143 são pessoa jurídica, o que contribui para o grande número de processos com gênero não informado. Do restante, 1.994 estavam em segredo de justiça, 4.504 indeterminados e em 5806 a informação não foi encontrada. A Figura 3 apresenta as informações mencionadas.

Figura 3 – Dashboard de estatísticas de frequência dos dados gerais

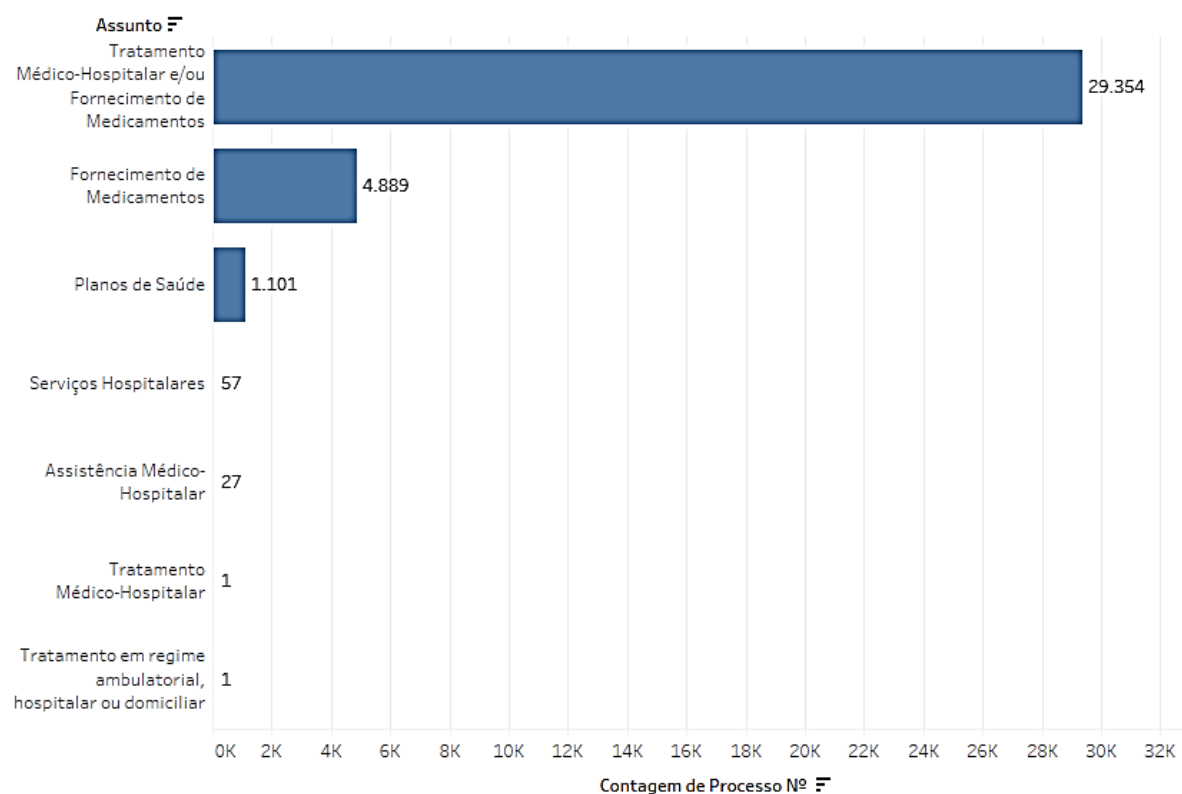


Fonte: dados da pesquisa

Lopes e Mello (2018) corroboram com esse resultado demonstrando que, embora exista uma tendência para o aumento da judicialização da saúde na região metropolitana de São Paulo, assim como em todo país, também podem haver interesses de grupos empresariais envolvidos, tais como, as indústrias farmacêuticas. Por sua vez, Vargas-Peláez et al., (2014) demonstram que a judicialização se tornou uma questão de políticas públicas, mesmo que o processo tenha origem de instituições privadas, influenciada por diferentes fatores, tais como, deficiências do sistema de saúde. Portanto, nota-se que tanto as instituições públicas como as privadas, se apresentam como réus de ações com tais características, sugerindo, dessa forma, a predominância de pessoa jurídica.

Também se buscou identificar qual doença estaria associada com cada processo, uma vez que o foco dessa pesquisa está em ações associadas a doenças raras. Foram encontrados 63 processos associados a paralisia, 2 a síndrome de rett, 1 a síndrome de sjogren, 1 a síndrome de cushing, 1 a síndrome de asperger e 35.362 sem informação. A partir da leitura do conteúdo de alguns processos foi possível identificar que muitas ações são motivadas pela ausência de auxílio a medicamentos, o que justifica um baixo número de doenças identificadas nesses textos. Vargas-Pelaez et al., (2019) apontam em seu trabalho que entre agosto e dezembro de 2014 a judicialização de medicamentos foi considerada um fenômeno generalizado no Brasil.

Os processos são distribuídos em 7 assuntos distintos. A maioria corresponde a Tratamento Médico-Hospitalar e/ou Fornecimento de Medicamentos (82,851%), seguido de Fornecimento de Medicamentos (13,799%), Planos de Saúde (3,108%), Serviços Hospitalares (0,161%), Assistência Médico Hospitalar (0,076%), Tratamento Médico-Hospitalar (0,003%) e, por fim, Tratamento em regime ambulatorial, hospitalar ou domiciliar (0,003%). Essas divisões foram criadas automaticamente a partir das informações contidas nos processos coletados, por meio de técnicas de linguagem natural. A seguir, na Figura 4, é possível ver a quantidade numérica de processos por assunto distinto.

Figura 4 - Frequência de processos por assunto

Fonte: dados da pesquisa

Os dados possuem 39 classes distintas, as quais foram divididas em 3 grupos: acima de 1.000 processos, entre 999 a 50 processos e 49 processos ou menos. Tal divisão foi adotada a fim de permitir a realização das análises considerando as classes mais relevantes. A Tabela 1 apresenta essas divisões, assim, é possível identificar que o grupo 1, composto por 5 classes, representa 93,27% dos processos da base de dados, o grupo 2 representa 6,32% e o grupo 3 apenas 0,42%.

Tabela 1 - Divisão dos processos por classes

| | Classe | Contagem de processos | Porcentagem de processos |
|----------------|--|-----------------------|--------------------------|
| Grupo 1 | Procedimento Ordinário | 10.567 | 29,825% |
| | Procedimento Comum | 8.789 | 24,807% |
| | Mandado de Segurança | 8.781 | 24,784% |
| | Procedimento do Juizado Especial Cível | 3.018 | 8,518% |
| | Ação Civil Pública | 1.890 | 5,334% |
| Grupo 2 | Requisição de Pequeno Valor | 961 | 2,712% |
| | Cumprimento de sentença | 726 | 2,049% |
| | Procedimento ordinário | 149 | 0,421% |

| | | | |
|---------------------|--|--------|--------|
| | Procedimento Sumário | 127 | 0,358% |
| | Cumprimento de Sentença contra a Fazenda Pública | 95 | 0,268% |
| | Cautelar Inominada | 95 | 0,268% |
| | Outras medidas provisionais | 85 | 0,240% |
| Grupo 3 | Execução Contra a Fazenda Pública | 44 | 0,124% |
| | Embargos à Execução | 14 | 0,040% |
| | Precatório | 13 | 0,037% |
| | Cumprimento Provisório de Sentença | 9 | 0,025% |
| | Providência | 7 | 0,020% |
| | Outros procedimentos de jurisdição voluntária | 7 | 0,020% |
| | Alvará Judicial | 7 | 0,020% |
| | Outros Feitos não Especificados | 5 | 0,014% |
| | Interdição | 5 | 0,014% |
| | Tutela Antecipada Antecedente | 4 | 0,011% |
| | Impugnação ao Valor da Causa | 4 | 0,011% |
| | Restauração de Autos | 3 | 0,008% |
| | Monitória | 3 | 0,008% |
| | Medidas de Proteção à Criança e Adolescente | 3 | 0,008% |
| | Cumprimento Provisório de Decisão | 3 | 0,008% |
| | Alvará Judicial Lei 6858/80 | 3 | 0,008% |
| | Tutela Cautelar Antecedente | 2 | 0,006% |
| | Impugnação ao Cumprimento de Sentença | 2 | 0,006% |
| | Procedimento do Juizado Especial cível | 1 | 0,003% |
| | Petição | 1 | 0,003% |
| | Impugnação de Assistência Judiciária | 1 | 0,003% |
| | Exibição | 1 | 0,003% |
| | Declaração de Ausência | 1 | 0,003% |
| | Carta Precatória Cível | 1 | 0,003% |
| | Autorização judicial | 1 | 0,003% |
| | Alienação Judicial de Bens | 1 | 0,003% |
| Ação Civil Coletiva | 1 | 0,003% | |

Fonte: dados da pesquisa

As classes processuais possuem algumas características adotadas no campo do Direito. Frequentemente são classificadas por códigos e siglas, definidos a partir de um fundamento legal. Em algumas situações, fica expresso que o pedido do autor(a) está evidenciado no próprio

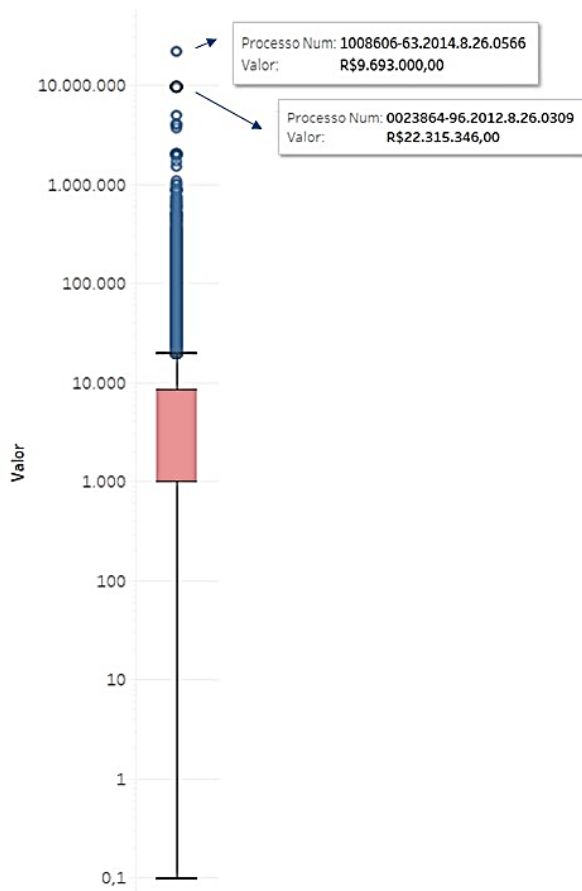
nome da classe, além de também ficar bem exposto no assunto do processo, embora também ocorram classes que não coincidem com o pedido devido ao seu grau de generalidade (Watanabe, 2015). Nota-se na Tabela 1 que as classes não coincidem com os assuntos dos processos. Para afirmar tal premissa com total segurança, basta comparar com a Figura 4, a qual consta os assuntos dos processos.

Com o auxílio do software Tableau, foi analisado o boxplot dos valores das ações, o qual representou graficamente a variação de dados observados de uma variável numérica, por meio de quartis. A Figura 5 apresenta os dados de valor de ação, sendo a média R\$ 11.477,00, o mínimo R\$ 0,10, o máximo R\$ 22.315.346,00 e a mediana R\$ 1.000,00.

Portanto, é possível observar que 2 ações se destacam, uma no valor de R\$ 22.315.346,00 que corresponde ao assunto Tratamento Médico-Hospitalar e/ou Fornecimento de medicamentos, cidade de Jundiaí, da classe Procedimento Comum Cível, movida por uma pessoa física contra o estado de São Paulo. A outra ação de destaque, no valor de R\$ 9.693.000, trata-se de um processo extinto, de tramitação prioritária, de pessoa física contra a Fazenda Pública do Estado de São Paulo, também do assunto Tratamento Médico-Hospitalar e/ou Fornecimento de medicamentos e da cidade de São Carlos.

Os processos mencionados são, visualmente, os dados mais discrepantes do restante. Excluindo esses dois valores, com o intuito de alcançar uma média mais homogênea, tem-se média de R\$ 10.574,00 e a mediana se mantém em R\$ 1.000,00. O comportamento desses dados demonstra que a base não segue uma distribuição normal, pois o valor da média acima da mediana representa uma característica de distribuição assimétrica.

Figura 5 - Boxplots de valores das ações



Fonte: dados da pesquisa

4.2 Frequências estatísticas entre as categorias dos dados

Considerando apenas os processos finalizados, ou seja, as ações com "transito em julgado", a classe procedimento ordinário predomina com 33,58% dos processos, mandado de segurança 23,59%, procedimento comum 19,64%, procedimento do juizado especial cível 11,17% e ação civil pública com 4,93%. Para esses dados, a ação de maior valor corresponde a R\$ 9.693.000,00. Porém, analisando as somas e as médias do conjunto de processos por classe, note-se na Figura 6 que a classe procedimento comum, possui um valor maior que mandado de segurança, representado na Figura 7, mesmo essa última possuindo mais processos.

Observando as médias na Tabela 2, percebe-se que a ocorrência se mantém, o que sugere que a classe contém processos com valores mais altos. Entretanto, a presença da ação no valor de R\$ 9.693.000,00 pode ser a razão dessa discrepância. Logo, excluiu a ação mencionada, conforme a Figura 8, alterando o valor total para R\$ 73.520.865,54 e a média para R\$ 16.751,17.

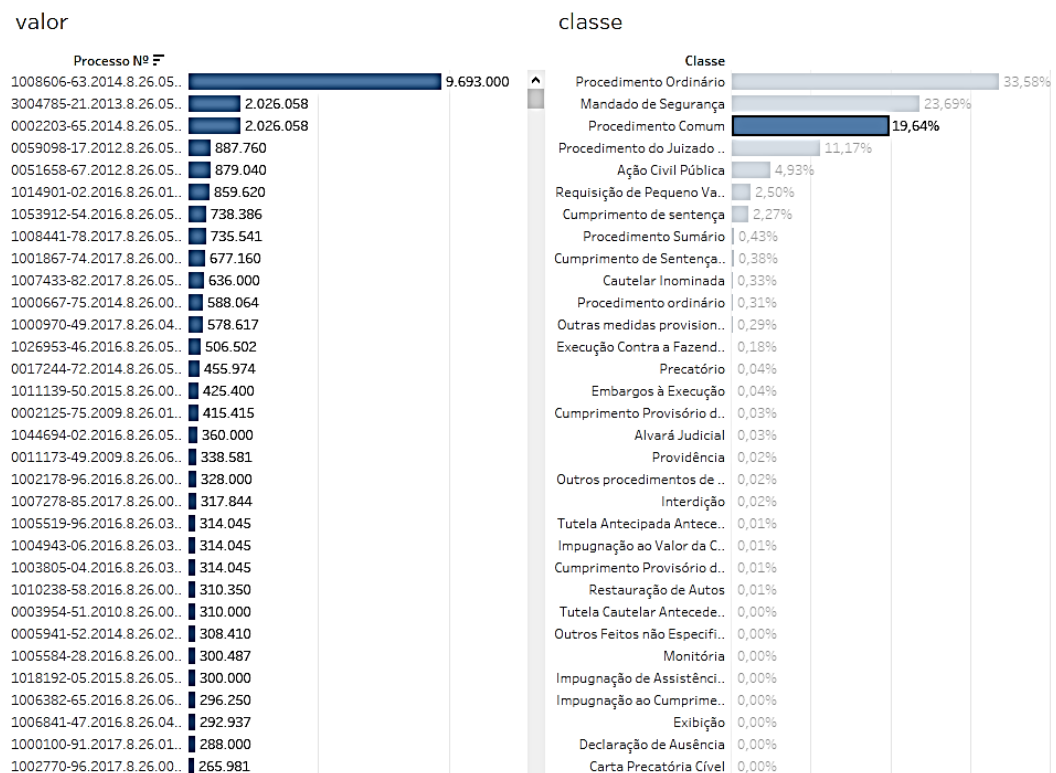
Portanto, sugere-se que as ações da classe procedimento comum são mais caras que as de mandado de segurança.

Tabela 2 - Comparação dos valores da soma de ações por classe

| Classe | Porcentagem de processos | Soma dos valores dos processos | Média dos valores dos processos |
|--|--------------------------|--------------------------------|---------------------------------|
| Procedimento ordinário | 33,58% | R\$ 98.083.681,20 | R\$ 13.067,37 |
| Mandado de segurança | 23,59% | R\$ 24.354.155,46 | R\$ 4.600,33 |
| Procedimento comum | 19,64% | R\$ 83.213.865,54 | R\$ 18.955,32 |
| Procedimento do juizado especial cível | 11,17% | R\$ 16.653.291,11 | R\$ 6.669,32 |
| Ação civil pública | 4,93% | R\$ 13.200.116,40 | R\$ 11.989,21 |

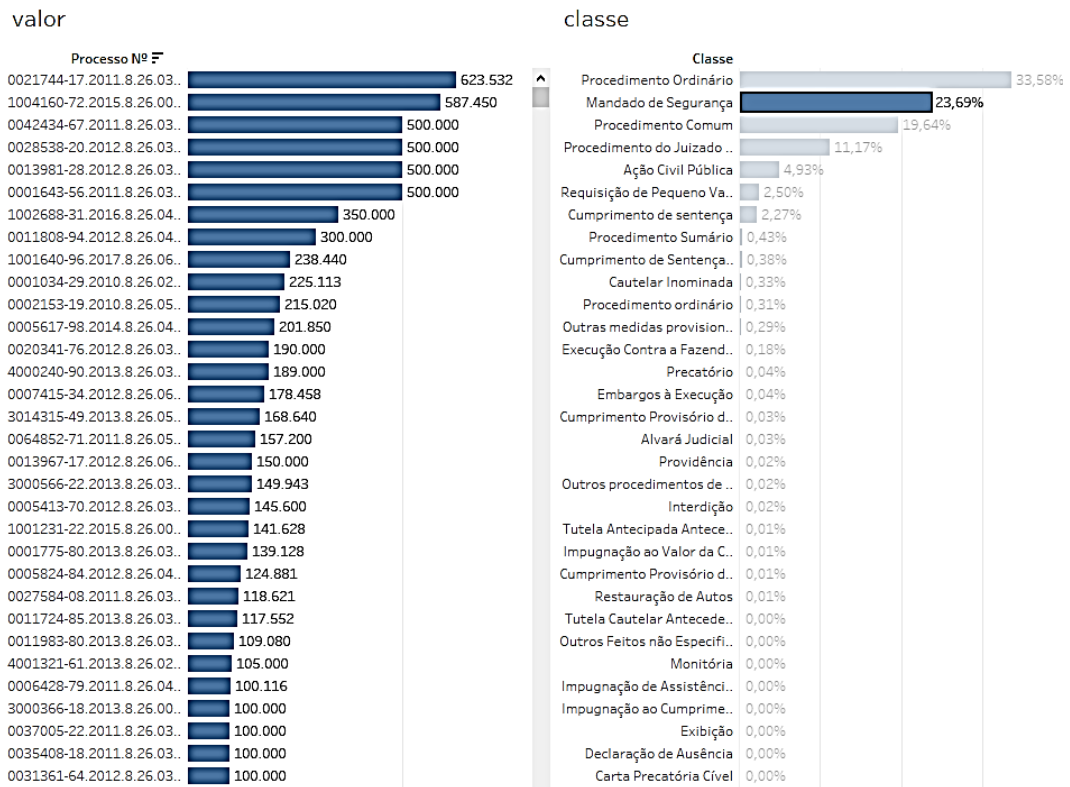
Fonte: dados da pesquisa

Figura 6 - Valores dos processos da classe procedimento comum



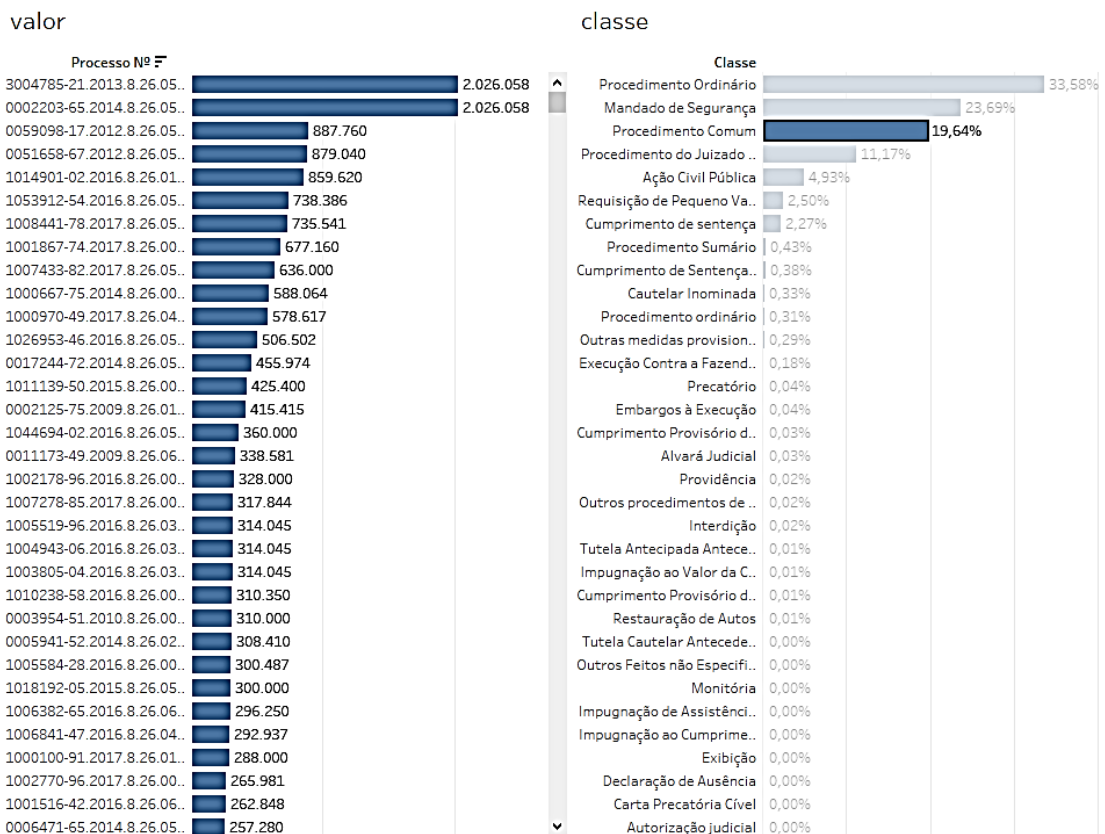
Fonte: dados da pesquisa

Figura 7 - Valores dos processos da classe mandado de segurança



Fonte: dados da pesquisa

Figura 8 - Valores dos processos da classe procedimento comum excluída a ação de maior valor



Fonte: dados da pesquisa

Em relação as sentenças desses processos, a base possui duas divisões principais. Na primeira estão as sentenças classificadas manualmente conforme a coleta. Já a segunda divisão, apresenta sentenças classificadas automaticamente e com a porcentagem de acerto para cada classificação. Na Tabela 3 é possível ver a quantidade de processos para cada sentença, classificada manualmente, e sua respectiva porcentagem. Dessa forma, nota-se que uma grande parte ainda não recebeu classificação de sentença até o momento da coleta da pesquisa (39,90%) e outra parte significativa foram classificadas como procedentes (39,58%), correspondendo juntas a quase 80% dos processos.

Tabela 3 - Frequência das sentenças classificadas manualmente

| Sentenças | Contagem de Processos | Porcentagem |
|--|------------------------------|--------------------|
| Sentenças não classificadas | 14.136 | 39,90% |
| Procedente | 14.022 | 39,58% |
| Extinção | 3.891 | 10,98% |
| Improcedente | 1.300 | 3,67% |
| Parcial procedente | 1.092 | 3,08% |
| Extinção da execução | 666 | 1,88% |
| Homologatória acordo | 196 | 0,55% |
| Erro coleta | 90 | 0,25% |
| Sentença de revelia | 23 | 0,06% |
| Reconhecimento decadência/prescrição | 8 | 0,02% |
| Homologatória renúncia/desistência autor | 6 | 0,02% |

Fonte: dados da pesquisa

Já no caso das sentenças classificadas automaticamente, cada processo recebeu um valor para probabilidade de acerto, o qual a média varia entre 0,6 e 0,7 (Extinção de Processo 0,7; Homologado 0,7; Improcedente 0,6; NC 0,6; Parcialmente Procedente 0,6; Procedente 0,6; Sem Mérito 0,6). Possui 7 divisões, ou seja, 3 a menos que as divisões das sentenças classificadas manualmente. Dessa forma, torna-se inviável fazer uma comparação entre as duas classificações de maneira geral, pois a quantidade de processos em cada classificação não poderá ser igual. Entretanto, é possível verificar para cada processo de forma individual, porém, essa análise não se enquadra ao foco dessa pesquisa.

Na Tabela 4 estão os valores de frequência das sentenças classificadas automaticamente. Assim como na classificação manual, nota-se que as sentenças não classificadas (NC) e as procedentes representam juntas aproximadamente 80% dos processos, 53,39% e 27,05%

respectivamente. Essa informação é favorável para confirmar se, de fato, as classificações automáticas são eficazes. Para as demais análises a partir daqui, será utilizada as sentenças classificadas manualmente.

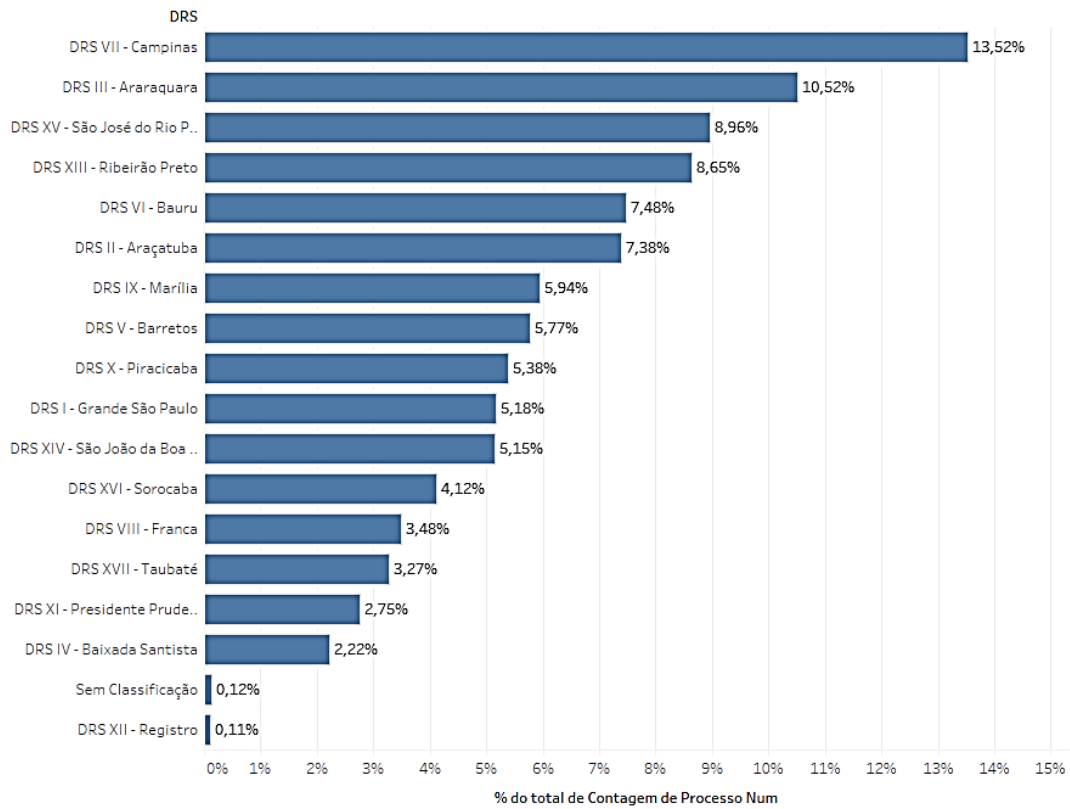
Tabela 4 - Frequência das sentenças classificadas automaticamente

| Classificação Sentenças | Contagem de Processos | Porcentagem |
|--------------------------------|------------------------------|--------------------|
| NC | 18.916 | 53,39% |
| Procedente | 9.584 | 27,05% |
| Sem Mérito | 3.800 | 10,73% |
| Parcialmente Procedente | 2.382 | 6,72% |
| Extinção do Processo | 498 | 1,41% |
| Improcedente | 235 | 0,66% |
| Homologado | 15 | 0,04% |

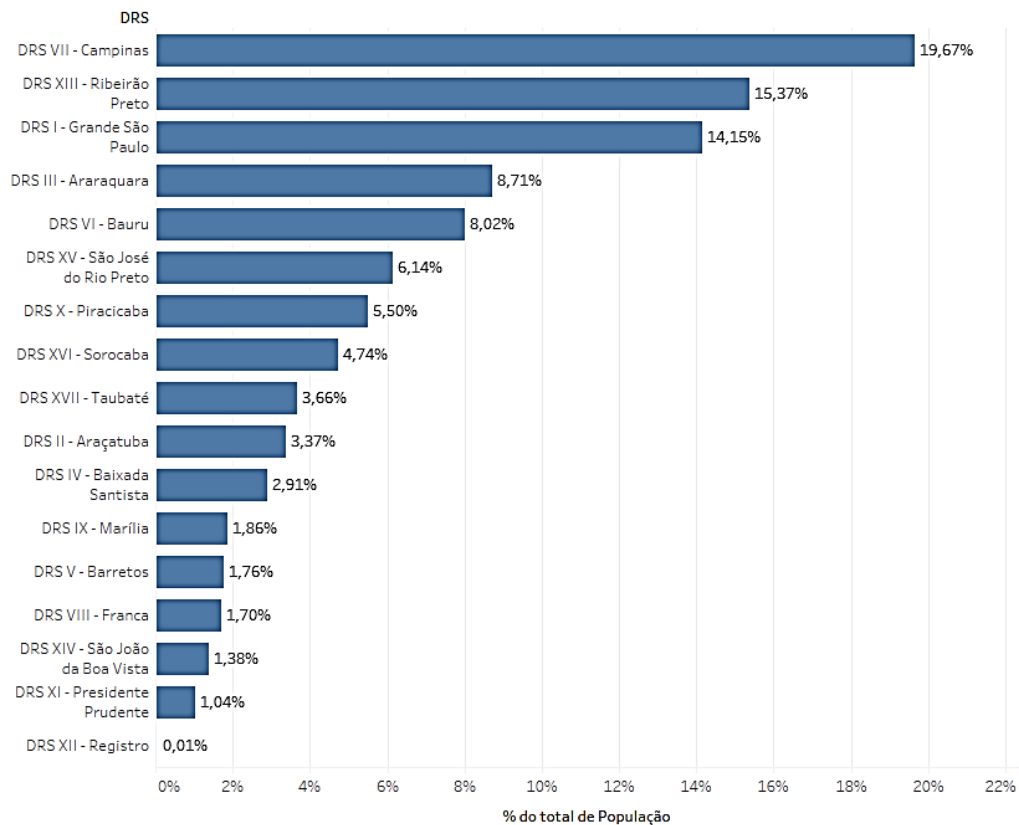
Fonte: dados da pesquisa

4.3 Resultados de análises por regionais

Conforme já mencionado no capítulo anterior, considerando que a base de dados possui 315 cidades distintas, foi necessário incluir uma nova divisão para facilitar as visualizações, adotando a divisão das 17 DRS do estado de São Paulo. Na Figura 9 é possível identificar que a regional com a maior quantidade de processos é a de Campinas, estando a regional da capital, São Paulo, em décimo lugar. Dessa forma, para concluir qual região de fato possui uma maior quantidade de processos, identificou-se a população contida em cada regional. Portanto, observando a Figura 10, identifica-se que a regional de Campinas também possui a maior população, o que se torna esperado, em termos de proporcionalidade, que tenha um maior número de processo. Entretanto, a capital São Paulo aparece como a terceira maior população, logo, realmente possui um menor número de processos se comparado, por exemplo, com a regional de Araraquara. A Tabela 5 apresenta as diferenças em porcentagens.

Figura 9 - Frequência de processos por DRS

Fonte: dados da pesquisa

Figura 10 - Frequência de população por DRS

Fonte: dados da pesquisa

Tabela 5 - Comparação entre porcentagens de processos e população por DRS

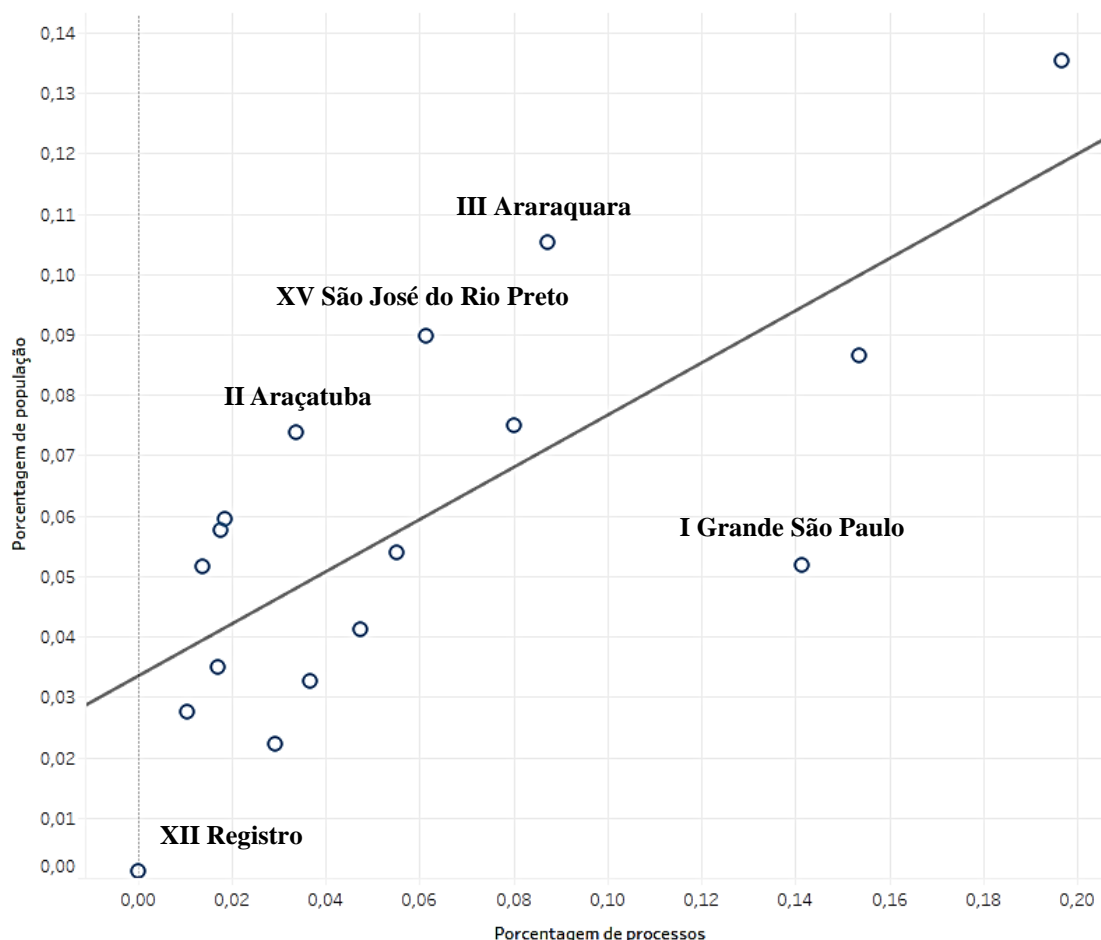
| DRS | Porcentagem de Processos | Porcentagem de população |
|---------------------------------|---------------------------------|---------------------------------|
| DRS I - Grande São Paulo | 5,19% | 14,15% |
| DRS II - Araçatuba | 7,39% | 3,37% |
| DRS III - Araraquara | 10,53% | 8,71% |
| DRS IV - Baixada Santista | 2,22% | 2,91% |
| DRS IX - Marília | 5,95% | 1,86% |
| DRS V - Barretos | 5,77% | 1,76% |
| DRS VI - Bauru | 7,49% | 8,02% |
| DRS VII - Campinas | 13,53% | 19,67% |
| DRS VIII - Franca | 3,49% | 1,70% |
| DRS X - Piracicaba | 5,39% | 5,50% |
| DRS XI - Presidente Prudente | 2,75% | 1,04% |
| DRS XII - Registro | 0,11% | 0,01% |
| DRS XIII - Ribeirão Preto | 8,66% | 15,37% |
| DRS XIV - São João da Boa Vista | 5,16% | 1,38% |
| DRS XV - São José do Rio Preto | 8,97% | 6,14% |
| DRS XVI - Sorocaba | 4,13% | 4,74% |
| DRS XVII - Taubaté | 3,27% | 3,66% |

Fonte: dados da pesquisa

Portanto, a fim de identificar a influência do tamanho da população, na quantidade de processos das regionais, optou-se pela realização de uma análise de correlação. Trata-se de uma estatística que aborda questões relativas à relação ou dependência entre as variáveis, nesse caso, a variável dependente é o número de processos e a independente o número da população. O R^2 determina o quão próximos os dados estão da linha de regressão ajustada. Como o valor foi de $R^2 = 0,54$, pode-se dizer que o modelo explica moderadamente a variabilidade dos dados de resposta ao redor de sua média (Scott, Flaherty, & Currall, 2013).

O gráfico de dispersão utiliza coordenadas cartesianas XY para exibir os valores de duas variáveis quantitativas. Fornecem evidências do relacionamento e sua natureza linear. Assim, permitem explorar as relações de pares de um conjunto de dados com potencialmente muitas variáveis (Scott et al., 2013). Nesse caso, o gráfico contribuiu para uma identificação equilibrada da quantidade de processos nas regionais. Assim, é possível observar na **Figura 11** que a DRS I Grande São Paulo e a XII Registro, realmente possuem uma quantidade menor de processos esperados, da mesma forma que as DRSs II Araçatuba, III Araraquara e XV São José do Rio Preto, possuem uma maior quantidade de processos.

Figura 11 - Gráfico de dispersão entre porcentagens de processos e de população por DRS



Fonte: dados da pesquisa

Por fim, buscando encontrar mais indícios a respeito desses resultados, foram identificadas algumas características sociodemográficas das DRSs. Adotou como fonte o site do Censo IBGE do estado de São Paulo, acessado em 2022, e optou pelas variáveis que possuíam alguma relação direta com qualidade de saúde. Os dados coletados foram apenas das cidades de DRSs que se destacaram, seja por uma menor ou maior quantidade de processos, optou-se por analisar as seguintes: I Grande São Paulo, II Araçatuba, III Araraquara, XII Registro e XV São José do Rio Preto. Devida a extensão dos nomes das variáveis, foram feitas substituições por siglas, conforme está na Tabela 6. A legenda abaixo especifica as substituições.

Tabela 6 - Características sociodemográficas das DRSs

| DRS | PO | RSM | TE | PIB | IDHM | SUS | ESA |
|----------------------------|-----------|------------|-----------|---------------|-------------|------------|------------|
| I - Grande São Paulo | 26,8% | 34,2% | 91,2% | R\$ 50.619,85 | 0,763 | 46 | 84,3% |
| II – Araçatuba | 24,2% | 28,9% | 98,3% | R\$ 28.878,33 | 0,769 | 12 | 93,5% |
| III – Araraquara | 28,1% | 30,3% | 81,4% | R\$ 32.868,79 | 0,759 | 18 | 96,2% |
| XII – Registro | 14,4% | 36% | 97,5% | R\$ 26.196,57 | 0,71 | 7 | 70,1% |
| XV - São José do Rio Preto | 25,6% | 28,2% | 98,2% | R\$ 29.002,63 | 0,758 | 11 | 90,8% |

Fonte: IBGE (acessado 2022)

Legenda:

População ocupada (2019) – PO

Percentual da população com rendimento nominal mensal per capita de até 1/2 salário mínimo (2010) - RSM

Taxa de escolarização de 6 a 14 anos de idade (2010) - TE

PIB per capita (2019) - PIB

Índice de Desenvolvimento Humano Municipal (IDHM) (2010) - IDHM

Estabelecimentos de Saúde SUS (2009) - SUS

Esgotamento sanitário adequado (2010) - ESA

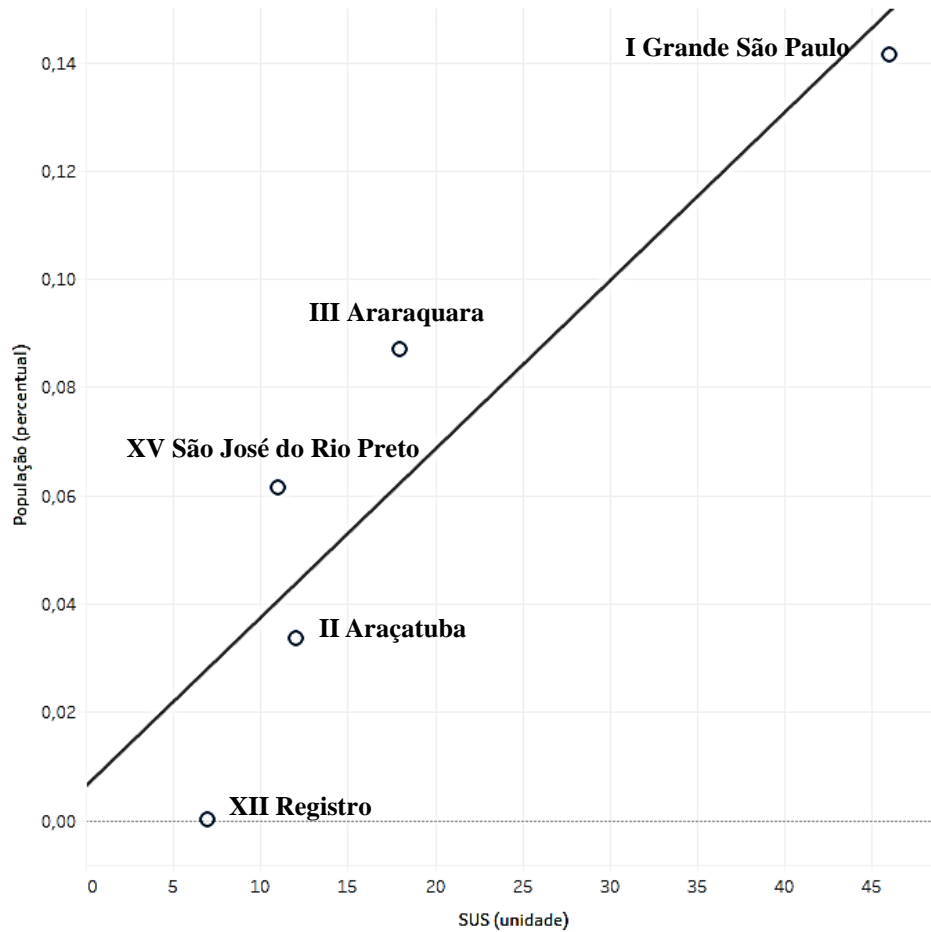
Os valores da Tabela 6 referem-se às médias das cidades de cada DRS. Nota-se que nos valores de Taxa de escolarização de 6 a 14 anos de idade, há uma variação de 16,9% entre II Araçatuba e III Araraquara, o Esgotamento sanitário adequado, variou 23,4%, entre XV São José do Rio Preto e XII Registro. Já o percentual da população com rendimento nominal mensal per capita de até 1/2 salário mínimo, XII Registro apresenta uma quantidade de 7,8% a mais que XV São José do Rio Preto, a qual possui o menor valor. A respeito da População ocupada, XV São José do Rio Preto apresenta 13,7% a mais que XII Registro que ocupa o último lugar. O Índice de Desenvolvimento Humano Municipal não deve diferenças expressivas.

Com base nessa análise inicial, nota-se que XV São José do Rio Preto tem um destaque positivo, assim como XII Registro tem um destaca negativo. Entretanto, XII Registro possui uma quantidade menor de processos, enquanto XV São José do Rio Preto possui uma quantidade maior de ações.

As variáveis Estabelecimentos de Saúde SUS e PIB per capita, são diretamente influenciadas pela população local, por essa razão, optou por analisá-las com gráficos de dispersão. Na Figura 12 é possível observar que XII Registro está afastado da linha de tendência em comparação as outras DRSs, portanto, possui uma quantidade de Estabelecimentos de Saúde

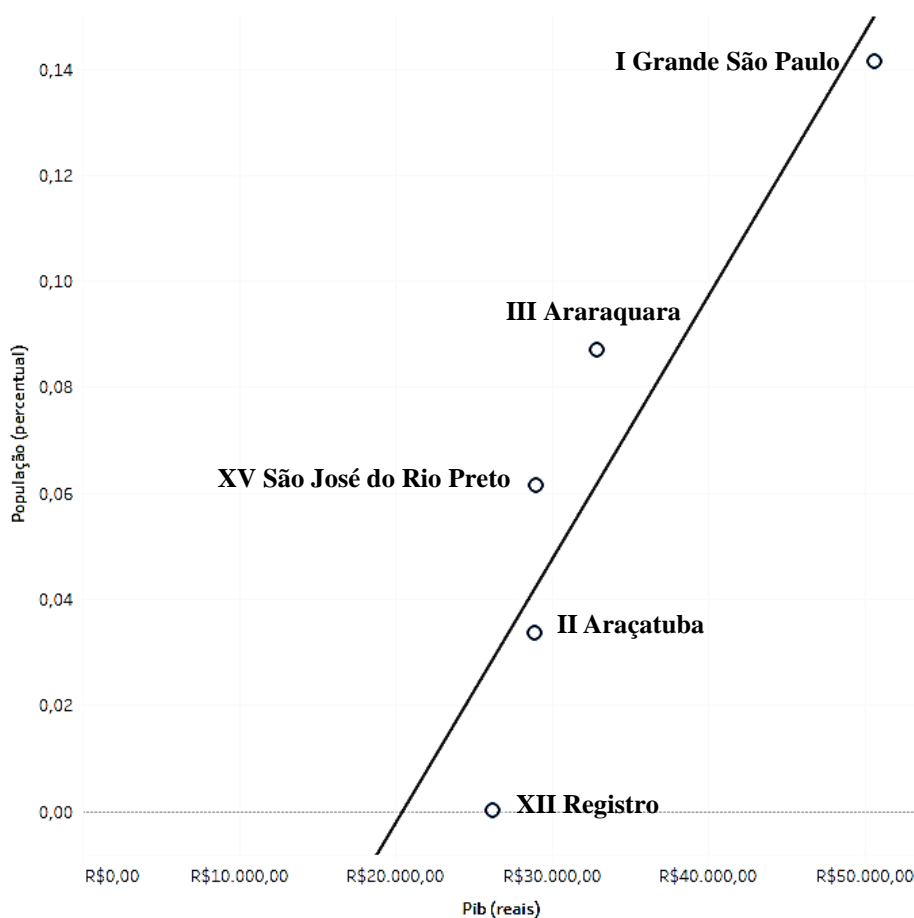
SUS desfavorável. Além disso, a DRS também aparece em uma posição desfavorável as demais na Figura 13, indicando um PIB per capita inferior.

Figura 12 - Gráfico de dispersão entre porcentagens de população e unidade SUS por DRS



Fonte: dados da pesquisa

Figura 13 - Gráfico de dispersão entre População e PIB per capita



Fonte: dados da pesquisa

Portanto, considerando que XII Registro foi apontada como desfavorável em seus dados sociodemográficos, mas apresenta uma quantidade de processos menor que o esperado, não foram encontrados indícios de relação entre as características sociodemográficas pesquisadas e quantidade de processos de judicialização da saúde em doenças raras.

Porém, embora Lopes e Mello (2018) apontem uma existência de uma tendência para o aumento da judicialização da saúde em todo país, Vargas-Pelaez et al. (2019) afirmam que existem indícios de que a judicialização acaba beneficiando as pessoas com maior renda e/ou mais capacitadas, pois o beneficiário de uma ação pula a fila e passa por cima dos que seguem a via administrativa. Portanto, esse argumento sugere que a falta de acesso a boas condições sociodemográficas, pode levar a população a desinformação a respeito da possibilidade de judicialização, o que seria coerente com as condições observadas na DRS XII Registro.

Conforme já mencionado no capítulo anterior, as ações possuem um valor mínimo de R\$ 0,10 e o valor máximo de R\$ 22.315.346,00. A Tabela 7 demonstra os valores mínimos e máximos dos processos por cada tipo de sentença. Assim, é possível identificar que tanto a ação de valor mínimo, como a ação de valor máximo, são sentenças procedentes que, portanto, movimentam uma quantia significativa de dinheiro.

Tabela 7 - Valores mínimos e máximos dos processos por cada tipo de sentença

| Sentença | Valor mínimo | Valor máximo |
|--|---------------------|---------------------|
| Sentenças não classificadas | R\$ 1,00 | R\$ 4.011.662,00 |
| Procedente | R\$ 0,10 | R\$ 22.315.346,00 |
| Extinção | R\$ 1,00 | R\$ 738.385,50 |
| Improcedente | R\$ 100,00 | R\$ 5.000.000,00 |
| Parcial procedente | R\$ 100,00 | R\$ 889.088,35 |
| Extinção da execução | R\$ 61,70 | R\$ 596.534,05 |
| Homologatória acordo | R\$ 100,00 | R\$ 1.970.129,50 |
| Erro coleta | R\$ 415,00 | R\$ 430.652,08 |
| Sentença de revelia | R\$ 100,00 | R\$ 18.000,00 |
| Reconhecimento decadência/prescrição | R\$ 724,00 | R\$ 8.360,85 |
| Homologatória renúncia/desistência autor | R\$ 678,00 | R\$ 45.480,00 |

Fonte: dados da pesquisa

A importância desses resultados é reforçada pelos trabalhos de Lamprea (2017) e Lopes & Mello (2018) os quais apontam que a existência de um volume significativo de processos de judicialização, sugere que a saúde precisa de uma maior atenção e políticas de planejamentos orçamentários para atender à necessidade pública. Além disso, gera um custo no Poder Judiciário, pois movimenta toda estrutura desse órgão, tais como, juízes, defensores públicos, promotores de justiça, serventuários e despesas judiciais, sendo a maioria tramitada pela justiça gratuita. Nesse sentido, reforça o benefício de se investir em técnicas de big data para otimizar a tomada de decisão no campo da saúde, a qual pode contribuir com importantes ganhos em termos de dinheiro, tempo e vidas (Chiavegatto Filho, 2015).

4.4 Apresentação de visualizações dinâmicas

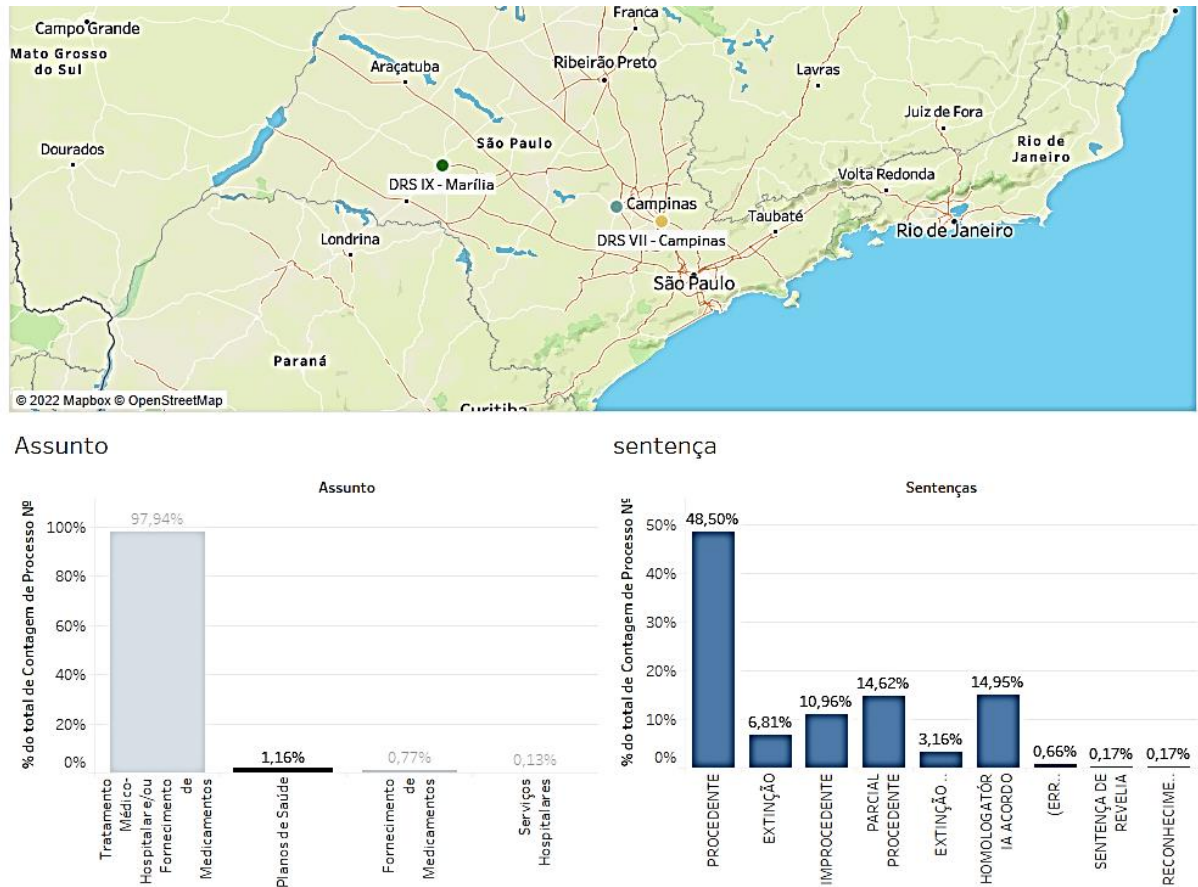
A presente pesquisa foi desenvolvida com o apoio de três ferramentas principais: R Studio, Tableau e Fluxicon Disco. O Tableau permite desenvolver visualizações dinâmicas com interação entre as diferentes variáveis dos dados. Com essa funcionalidade, os resultados podem

ser facilmente alterados, buscando responder variadas questões que podem surgir dos diversos *stakeholders* do campo da saúde.

Nesse sentido, a fim de demonstrar o potencial dessa funcionalidade, propôs-se responder algumas perguntas. Para elaborá-las, observou-se os gráficos individuais e optou por investigar pontos de destaque ou discrepância. As sentenças "não classificadas" foram removidas e manteve-se apenas os processos finalizados. Para tal, foram consideradas encerradas as ações "transito em julgado", momento em que não há mais recursos cabíveis para as partes e é feita a solicitação de execução da sentença. Com essa alteração, passou para 22.351 processos, diminuindo 36,9% da quantidade analisada inicialmente. Além disso, com essa exclusão, passaram-se para apenas 6 assuntos, sendo eliminada a categoria "tratamento médico-hospitalar".

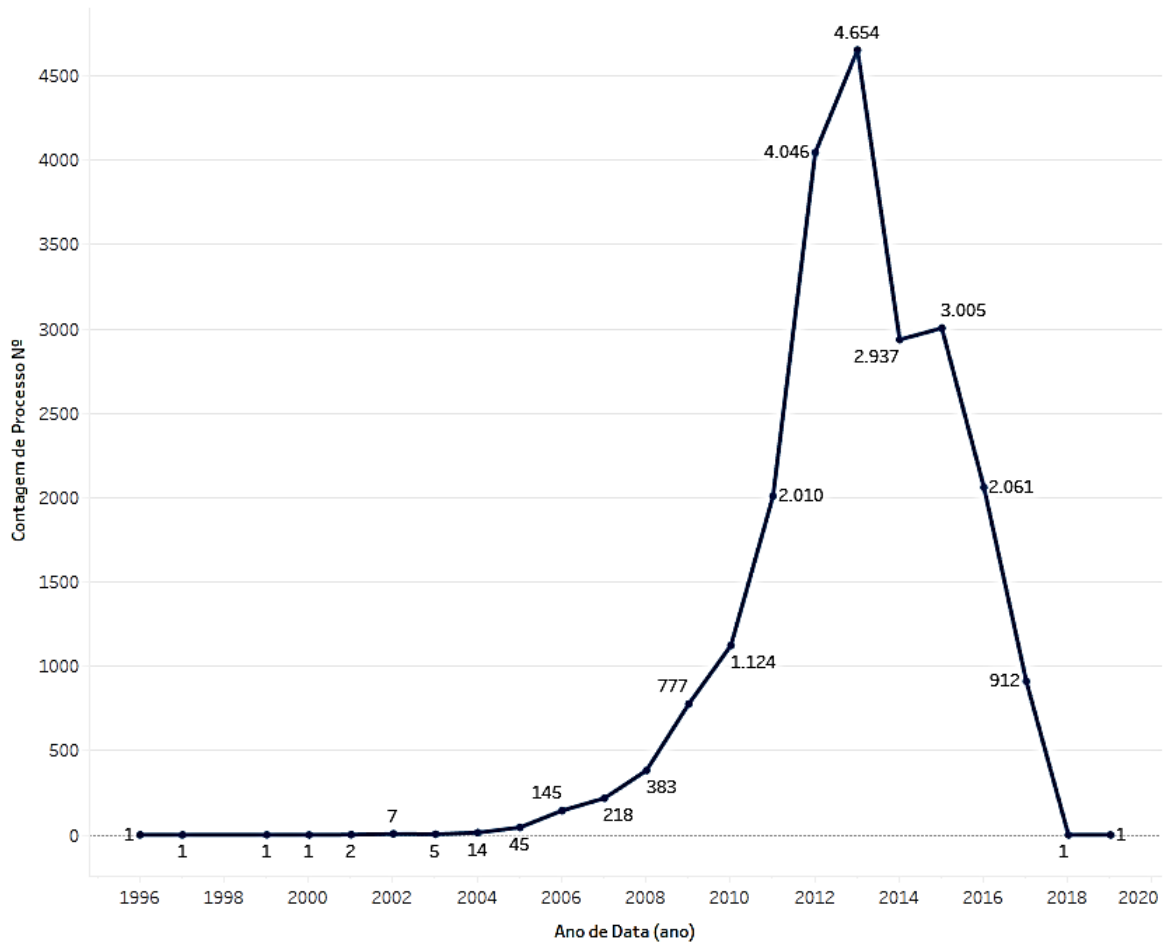
A Figura 14 propõe responder a seguinte pergunta: Quais são as regiões mais afetadas por processos relacionados a planos de saúde e qual tem sido a predominância das sentenças nesses casos? Para obter esse resultado, seleciona-se Plano de Saúde no gráfico de Assuntos e, automaticamente, temos as respostas para as outras variáveis. Nota-se observando o mapa que IX Marília e VII Campinas são as regiões mais afetadas e a predominância das sentenças é procedente.

Figura 14 – Exemplo de dashboard interativo do Tableau



Fonte: dados da pesquisa

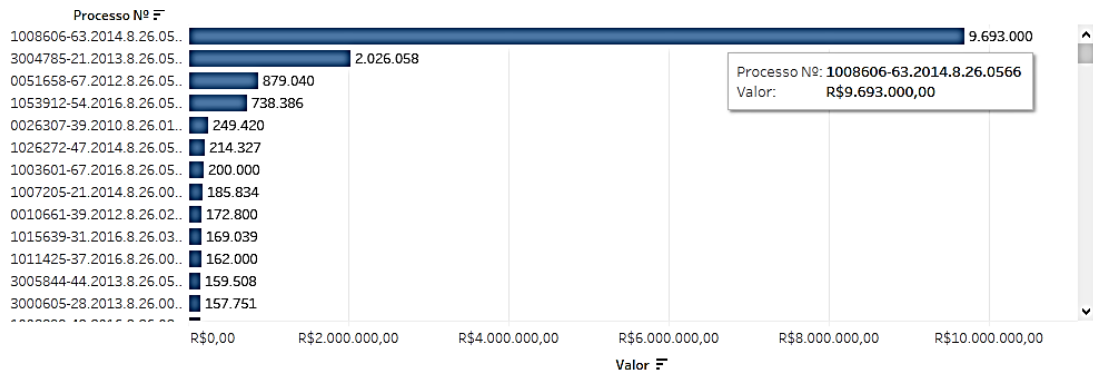
Observando a Figura 15, nota-se que a partir de 2010 inicia um crescimento na quantidade de processos, alcançando o pico em 2013 e voltando a ter outro aumento em 2015. Portanto, foram investigadas as características dos processos de 2013 e 2015.

Figura 15 - Contagem de processos por ano

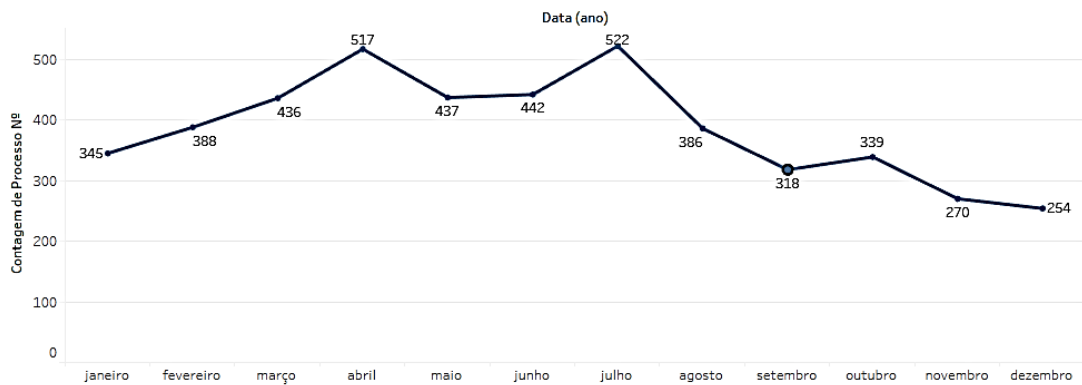
Fonte: dados da pesquisa

As Figura 16 e Figura 17 representam as respostas para a seguinte pergunta: Considerando que nos anos de 2013 e 2015 houveram aumentos nas quantidades de processos, quais são as características desses períodos que se diferenciam das características gerais da base de dados? Os dois períodos tiveram os mesmos resultados. Em termos de valores, embora se tenha a maior quantidade de processos em julho no caso de 2013, com 522 ações, e em novembro no caso de 2015, com 554 ações, a maior soma ocorreu em setembro, contabilizando R\$ 9.693.000,00 no total de 318 processos em 2013 e 329 em 2015. No caso das sentenças e dos assuntos, o padrão se manteve e as DRSs não apresentaram alterações significativas.

Figura 16 - Dashboard da soma dos valores mensais dos processos de 2013

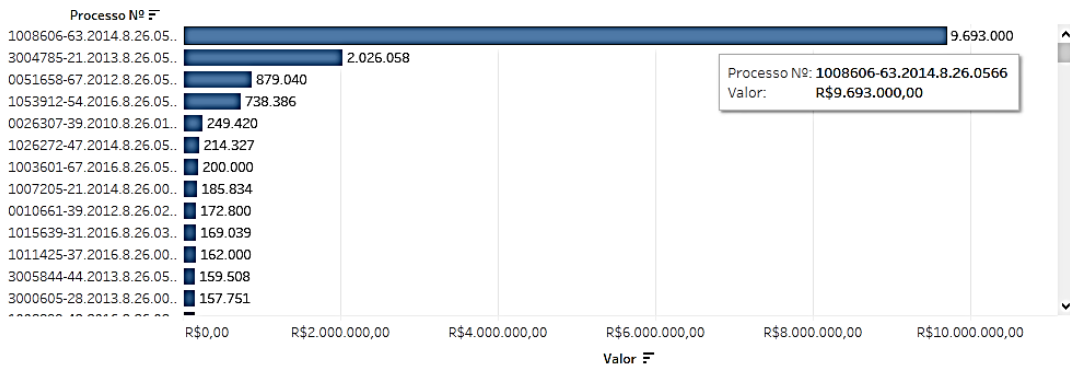


Data (2013)

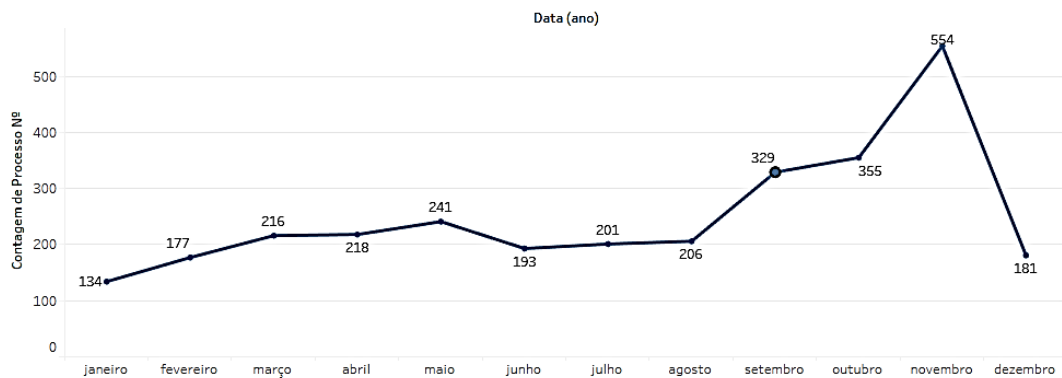


Fonte: dados da pesquisa

Figura 17 - Dashboard da soma dos valores mensais dos processos de 2015



Data (2015)



Fonte: dados da pesquisa

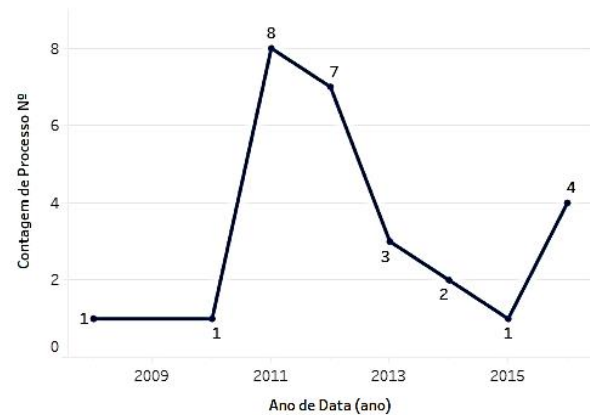
Figura 18 - Características dos processos da DRS XII Registro



Assunto



Data



sentença



Fonte: dados da pesquisa

Por fim, considerando que a DRS XII Registro apresentou resultados de destaque, tanto por possuir uma menor quantidade de processos do que o esperado, como por possuir dados sociodemográficos, de maneira geral, inferiores as demais DRSs, a Figura 18 buscou identificar as características dos processos dessa regional.

A princípio, nota-se que os picos de grandes volumes de processos que ocorreram nos dados gerais, não ocorrem nesse caso. O assunto predominou tratamento médico-hospitalar e/ou fornecimento de medicamentos e as sentenças mantiveram o padrão.

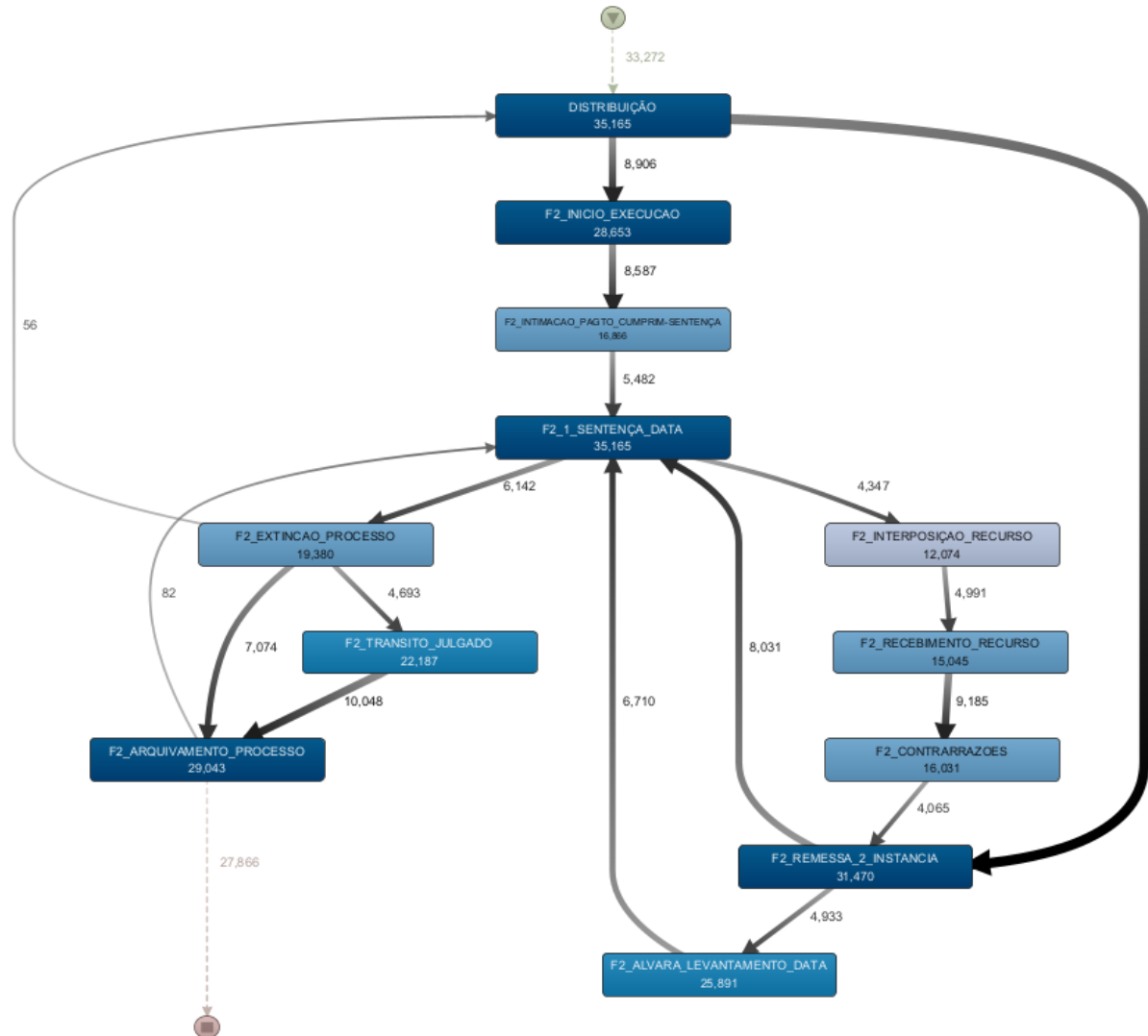
As informações representadas pelas Figuras até aqui, estão disponíveis no link:

https://public.tableau.com/app/profile/p.mela.dias2517/viz/Tableaupublic_16491051437650/Histrial#1

Porém, para ter acesso a informações interativas, é preciso contatar a autora do presente trabalho, pois não é permitida a edição pelo público.

4.5 Valores obtidos com mineração dos processos

Figura 19 - Sequência de frequência dos movimentos dos processos



Fonte: dados da pesquisa

A Figura 19 trata-se da representação da movimentação dos processos criado pelo software Fluxicon Disco. Ajustando a ferramenta ao mínimo de detalhes, se tem o comportamento principal dos processos. Assim, optando por manter todas as atividades, nota-se no início da imagem a entrada de 33.272 processos, finalizando com 27.866. Da distribuição, que são 35.165 casos, se tem 6.122 casos que não atingiram o ponto de arquivamento de processo. Para compreender a razão de alguns casos não terem participado da distribuição, seria necessária uma análise individual, o que não é o foco dessa pesquisa.

Em relação ao padrão das movimentações, apenas 58 casos percorrem 3 atividades, enquanto 6.993 percorrem por 8 atividades. Mais de 50% dos processos, 19.516, percorrem de

7 a 9 atividades. Na prática, isso significa que a maioria dos processos passaram por um maior número de atividades até atingir sua finalização. Observa-se que algumas atividades (nós), assim como algumas linhas (arestas) estão em tons mais escuros. Isso se deve ao fluxo de processos envolvidos. Quanto mais escuro, maior a movimentação. Foram identificadas 12 atividades, as quais estão distribuídas de acordo com a Tabela 8.

Tabela 8 - Distribuição das atividades

| Atividade | Número de processos |
|---|----------------------------|
| Distribuição | 35.165 |
| Início de execução | 28.653 |
| Intimação, pagamento, cumprimento de sentença | 16.866 |
| Sentença data | 35.165 |
| Interposição de recurso | 12.074 |
| Recebimento de recurso | 15.045 |
| Extinção de processo | 19.380 |
| Transito julgado | 22.187 |
| contrarrazões | 16.031 |
| Remessa 2ª instância | 31.470 |
| Alvará, levantamento, data | 25.891 |
| Arquivamento de processo | 29.043 |

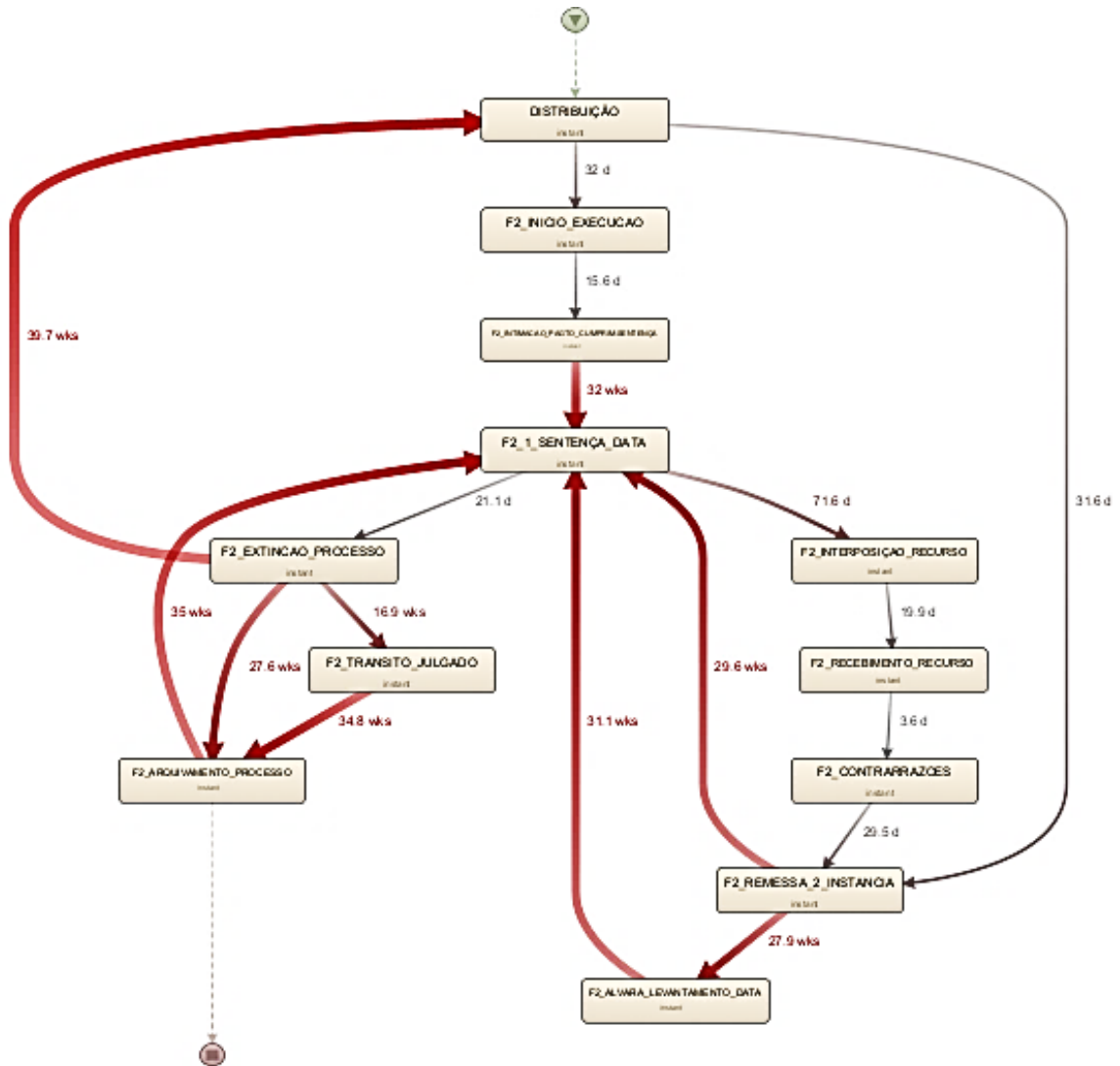
Fonte: dados da pesquisa

A média de duração dos casos é de 34,7 meses, ou seja, o tempo médio desde o início até o término de um processo, considerando o momento em que os dados foram coletados. Como a mediana é de 27,5 meses, existe uma diferença entre a média, então espera-se que tenha casos discrepantes em termos de duração de tempo. Observa-se que apenas 1 caso atinge a duração de 19 anos e 194 dias, enquanto para o maior número de casos, 952, a duração é de 1 ano e 212 dias.

Na Figura 20 são apresentadas a performance média dos movimentos dos processos. Quanto mais escura for a linha (aresta), representa que precisou de mais tempo para seguir de uma atividade para outra. Considerando as arestas em maior destaque, nota-se que de "Intimação, pagamento, cumprimento de sentença" para "Sentença data" teve duração média de 32 semanas. Já de "Remessa 2ª instância" para "Sentença data" a média foi de 29,6 semanas. De "Alvará, levantamento, data" para "Sentença data" ocorreram em média 31,1 semanas, "Extinção de processo" para "Arquivamento de processo" 27,6 semanas e de "Transito julgado" para "Arquivamento de processo" a média foi de 34,8 semanas. A duração máxima de uma

atividade para outra, considerando os 12 nós, foi de 14,2 anos, enquanto a duração mínima foi de transferência instantânea. Tais informações estão descritas na Tabela 9.

Figura 20 - Performance média de movimentos dos processos



Fonte: dados da pesquisa

Tabela 9 - Informações de performance média dos movimentos dos processos

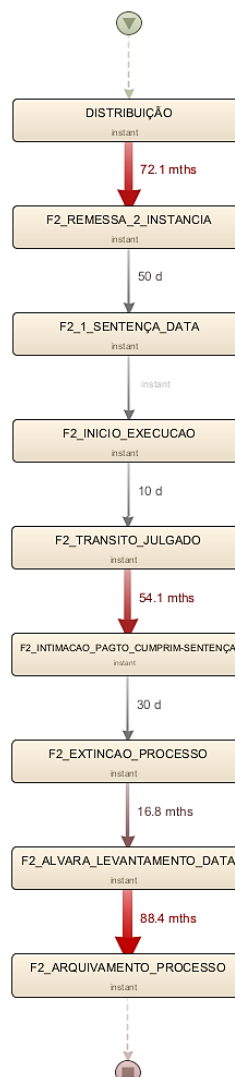
| Início | Fim | Média |
|---|--------------------------|--------------|
| Intimação, pagamento, cumprimento de sentença | "Sentença data" | 32 semanas |
| Remessa 2ª instância | Sentença data | 29,6 semanas |
| Alvará, levantamento, data | Sentença data | 31,1 semanas |
| Extinção de processo | Arquivamento de processo | 27,6 semanas |
| Transito julgado | Arquivamento de processo | 34,8 semanas |

Fonte: dados da pesquisa

A distribuição da duração dos casos demonstra com casos que foram resolvidos rapidamente, atingindo o ponto mais alto com 957 casos durando, distribuição assimétrica com 1 ano e 212 dias. O caso solucionados mais rapidamente durou 6 dias, enquanto o mais demorado levou 19 anos e 194 dias.

Portanto, conforme demonstra a Figura 21, foram adicionados filtros para investigar o comportamento dos casos descritos anteriormente. No caso da duração máxima, filtrados com acima de 19 anos, se tem 1 caso e 9 atividades, com duração de 19 anos e 194 dias. No mapa de performance é possível identificar os momentos em que o processo se manteve por mais tempo. Com destaque da "Distribuição" para "Remessa 2ª instância" em 72,1 meses, do "Transito julgado" para "Intimação, pagamento, cumprimento de sentença" com 54,1 meses e de "Alvará, levantamento, data" para "Arquivamento de processo " com 88,4 meses.

Figura 21 - Performance de movimentos do processo de maior tempo



Fonte: dados da pesquisa

Já se tratando da duração mínima, conforme a Figura 22, filtrados até 6 dias, se tem 1 caso, com 4 atividades. No mapa de performance é possível identificar os movimentos desse processo, sendo de "sentença, data" para "extinção de processo" a única movimentação que durou 4 dias e as demais duraram 24 horas.

Figura 22 - Performance de movimentos do processo de menor tempo

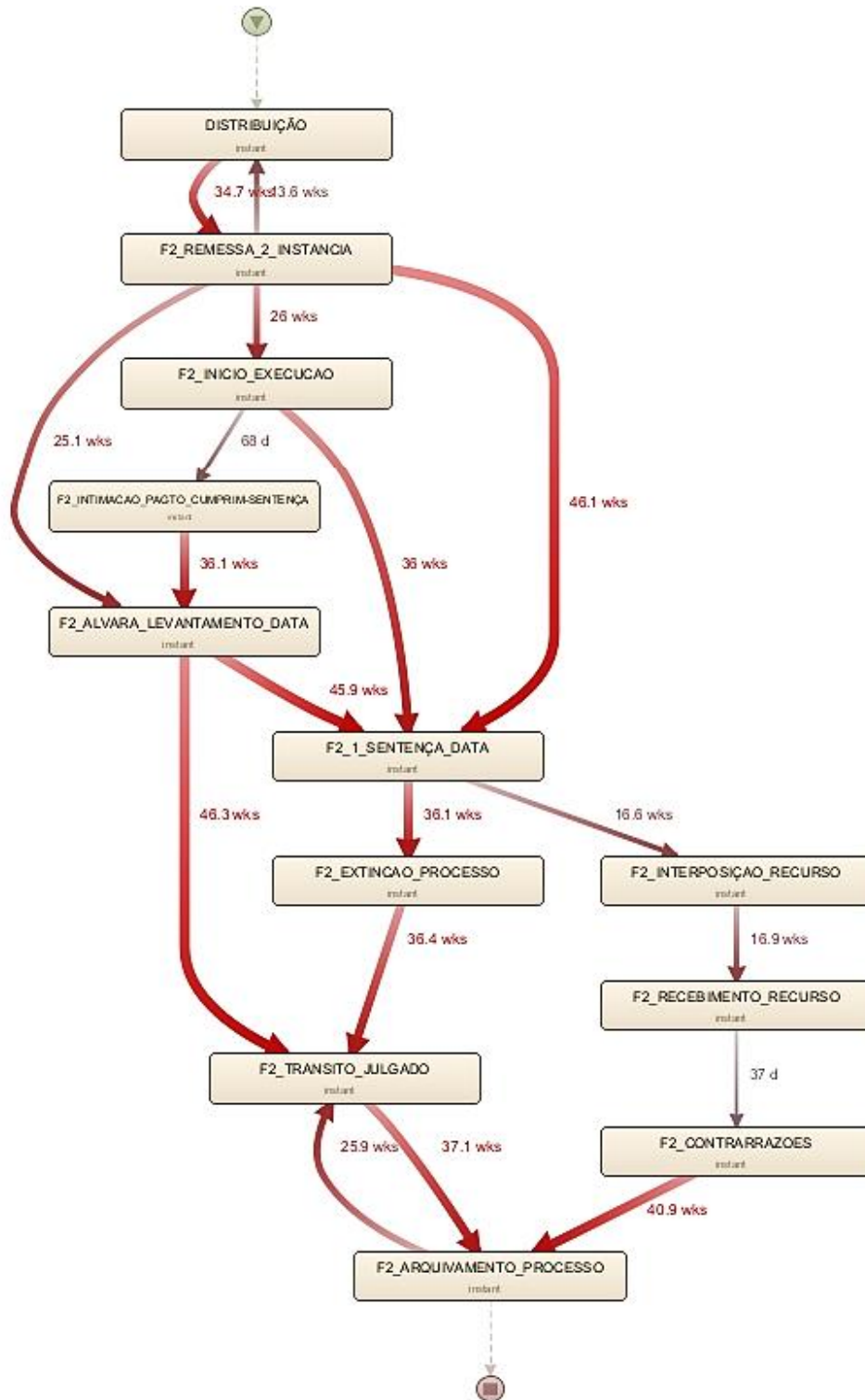


Fonte: dados da pesquisa

Na análise do volume de casos ativos ao longo do tempo, nota-se a presença de dois picos, dois momentos em que um grande número de casos estava ativo, em 2013 e 2015. Com o auxílio da aplicação de filtros, buscou-se investigar as características desses casos. Adotando o período de 09/06/2013 a 01/09/2014, se tem 702 casos, aproximadamente 1% da base total, com 12 atividades e duração média de 28, 8 semanas. Combinando os mapas de frequência e de performance, é possível identificar os momentos em que os processos transitaram de forma mais lenta e a quantidade de ações contidas em cada movimentação. Conforme nota-se na Figura 23, existem 5 momentos de destaque em termos de duração dos movimentos considerando a duração total de todos os processos. De "remessa 2ª instância" para "sentença

data" se tem intervalo mínimo de 24 horas, médio de 66 dias e máximo de 46,1 semanas, de "início execução" para "sentença data" o intervalo mínimo é de 24 horas, médio de 77,9 dias e máximo de 36 semanas, de "alvará, levantamento, data" para "sentença data" o intervalo mínimo é de 7 dias, médio de 13,6 semanas e máximo de 45,9 semanas. De "alvará, levantamento, data" para "transito julgado" o intervalo mínimo é de 3 dias, médio de 13,4 semanas e máximo de 46,3 semanas, por fim, de "extinção de processo" para "transito julgado" o intervalo mínimo é de 24 horas, médio de 65,7 dias e máximo de 36,4 semanas. Tais informações estão na Tabela 10.

Figura 23 - Performance de movimentos dos processos de 09/06/2013 a 01/09/2014



Fonte: dados da pesquisa

Tabela 10 - Informações de performance média de movimentos dos processos de 09/06/2013 a 01/09/2014

| Início | Fim | Mínimo | Média | Máximo |
|----------------------|---------------|----------|---------|--------------|
| remessa 2ª instância | sentença data | 24 horas | 66 dias | 46,1 semanas |

| | | | | |
|----------------------------|------------------|----------|--------------|--------------|
| início execução | sentença data | 24 horas | 77,9 dias | 36 semanas |
| alvará, levantamento, data | sentença data | 7 dias | 13,6 semanas | 45,9 semanas |
| alvará, levantamento, data | transito julgado | 3 dias | 13,4 semanas | 46,3 semanas |
| extinção de processo | transito julgado | 24 horas | 65,7 dias | 36,4 semanas |

Fonte: dados da pesquisa

Portanto, em média, a movimentação que dura maior tempo é de "alvará, levantamento, data" para "sentença data". Entretanto, embora siga um padrão semelhante, na análise do mapa de frequência nota-se um destaque de "sentença, data" para "extinção de processo" com 248 casos que não é observado no mapa de performance. Dessa forma, as análises trazem indícios de que as movimentações de "remeça 2ª instância", "início, execução" e "alvará, levantamento, data", as 3 para "sentença data" são os pontos que trazem mais empecilhos para uma rápida movimentação desses processos. No que se refere a outras características, são processos de natureza cível, com 40,31% da classe procedimento comum cível, 30,57% mandado de segurança e 21,62% procedimento do juizado especial cível. Predominaram as sentenças julgada procedentes (23,75%) e sentença completa com resolução de mérito (19,79%). Os assuntos predominantes foram tratamento médico-hospitalar e/ou fornecimento de medicamento (74,42%) e fornecimento de medicamento (24,38%). A Tabela 11 traz essas informações.

Tabela 11 - Características dos movimentos dos processos de 09/06/2013 a 01/09/2014

| Atributo | Característica | Porcentagem |
|-----------------|---|--------------------|
| Natureza | Cível | 100% |
| Classe | Procedimento comum cível | 40,31% |
| Classe | Mandado de segurança | 30,75% |
| Classe | Procedimento do juizado especial cível | 21,62% |
| Sentença | Procedentes | 23,75% |
| Sentença | Sentença completa com resolução de mérito | 19,79% |
| Assunto | Tratamento médico-hospitalar e/ou fornecimento de medicamento | 74,42% |
| Assunto | fornecimento de medicamento | 24,38% |

Fonte: dados da pesquisa

Semelhante ao recorte anterior, também foram filtrados os processos do período de 09/11/2015 a 01/03/2017, onde se tem 694 casos, apenas 8 a menos que os filtrados anteriormente, aproximadamente 1% da base total, com 12 atividades e duração média de 27, 3 semanas. Conforme nota-se na Figura 24, também existem 5 momentos de destaque em

termos de duração dos movimentos considerando a duração total de todos os processos. De "alvará, levantamento, data" para "sentença data" o intervalo mínimo é de 48 horas, médio de 87,5 dias e máximo de 41,1 semanas, de "sentença data" para "transito julgado" se tem intervalo mínimo instantâneo, médio de 46,9 dias e máximo de 32 semanas, de "extinção de processo" para "transito julgado" o intervalo mínimo de 24 horas, médio de 66,2 dias e máximo de 35,4 semanas. De "transito julgado" para "arquivamento de processo" o intervalo mínimo é instantâneo, médio de 38,5 dias e máximo de 48,9 semanas, por fim, de "remessa 2ª instância" para "sentença data" o intervalo mínimo é de 24 horas, médio de 69,9 dias e máximo de 49 semanas, como demonstra a Tabela 12.

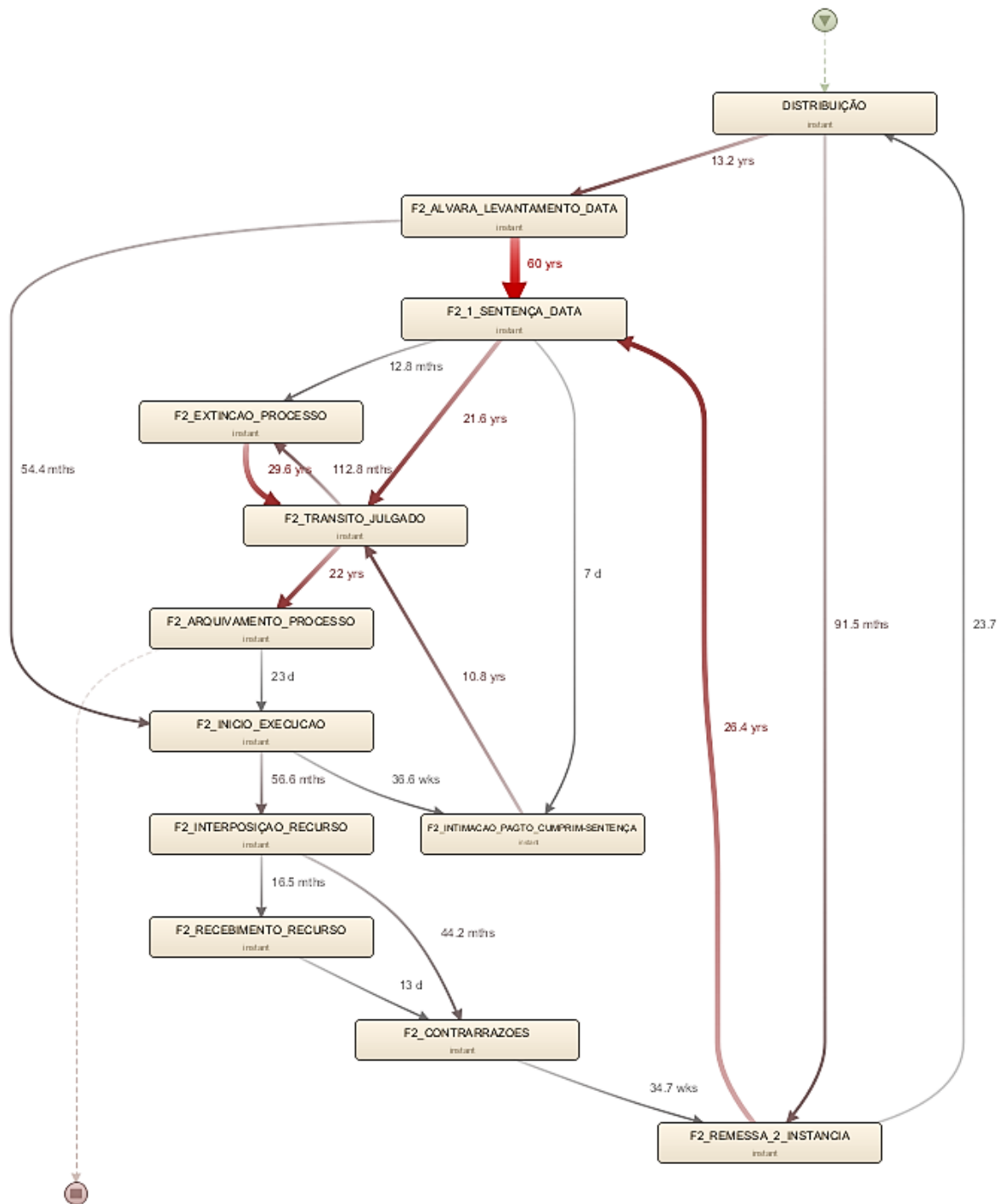
Tabela 12 - Informações de performance média de movimentos dos processos de 09/06/2013 a 01/09/2014

| Início | Fim | Mínimo | Média | Máximo |
|----------------------------|--------------------------|---------------|--------------|---------------|
| alvará, levantamento, data | sentença data | 48 horas | 87,5 dias | 41,1 semanas |
| sentença data | transito julgado | instantâneo | 46,9 dias | 32 semanas |
| extinção de processo | transito julgado | 24 horas | 66,2 dias | 35,4 semanas |
| transito julgado | arquivamento de processo | instantâneo | 38,5 dias | 48,9 semanas |
| remessa 2ª instância | sentença data | 24 horas | 69,9 dias | 49 semanas |

Fonte: dados da pesquisa

Portanto, em média, a movimentação que dura maior tempo, assim como no primeiro recorte, é de "alvará, levantamento, data" para "sentença data". Entretanto, na análise do mapa de frequência nota-se um destaque de "distribuição" para "remessa 2ª instância" com 252 casos que não é observado no mapa de performance, sendo esses indícios de que as movimentações mencionadas são pontos que trazem mais empecilhos para uma rápida movimentação desses processos. Embora as Figura 23 e Figura 24 tenham semelhanças em seus resultados, é notório que o padrão de movimentação é diferente o que indica que as razões pelas quais ocorreu o primeiro pico, não justifica de forma precisa a ocorrência do segundo pico.

Figura 24 - Performance de movimentos dos processos de 09/11/2015 a 01/03/2017



Fonte: dados da pesquisa

No que se refere a outras características, são processos de natureza cível, com 41,35% da classe procedimento do juizado especial cível, 35,58% procedimento comum cível e 9,65% mandado de segurança cível. Predominaram as sentenças julgada procedentes (56,24%) e sentença julgada procedente em parte (8,96%). Os assuntos predominantes foram tratamento

médico-hospitalar e/ou fornecimento de medicamento (62,3%) e fornecimento de medicamento (35,47%).

Os resultados dessa pesquisa foram focados em considerações gerais, porém, assim como mencionado para o software Tableau, o Fluxicon Disco também permite investigar casos, variáveis e atividades específicas, excluindo informações que possam ser consideradas irrelevantes para o momento. Portanto, a ferramenta traz um potencial promissor para detectar gargalos específicos que podem estar dificultando a fluidez de um processo.

Nesse contexto, reforçando o que já foi mencionado anteriormente com as análises feitas pelo Tableau, os resultados contribuem a fim de otimizar o tempo de resolução desses processos e, conseqüentemente, ganhos para a tomada de decisão no campo da saúde. Conforme aponta Chiavegatto Filho (2015) o investimento em técnicas de big data otimiza o campo da saúde, podendo contribuir com importantes ganhos em termos de dinheiro, tempo e vidas.

5 CONSIDERAÇÕES FINAIS

Nessa quinta seção, formulam-se as considerações finais, apontam-se as limitações e indicam-se sugestões para estudos futuros. A princípio, no que se refere ao objetivo geral dessa pesquisa, nota-se que técnicas de visualização de dados, associadas ao *business analytics*, são eficazes para abrir caminhos para novas possibilidades a partir de uma base extensa de dados.

Tal premissa justifica-se pela quantidade de informações apresentadas nos resultados dessa pesquisa, além das possibilidades para serem exploradas futuramente, as quais estão como sugestões de estudos no final desse capítulo, além da notável facilidade de multidisciplinariedade, limitando-se nesse trabalho aos campos da Administração, Direito e Ciências da Saúde.

Os resultados apontam que as características mais recorrentes nos processos públicos de doenças raras do estado de São Paulo, são autores, em sua maioria, do sexo masculino e pessoas físicas, enquanto os réus são a maioria pessoa jurídica, mas também predomina o sexo masculino. A predominância de pessoa jurídica como réu é justificada pela natureza desses processos, uma vez que a literatura aponta que farmacêuticas e demais empresas e instituições, públicas e privadas, associadas à saúde são os principais alvos dessas ações.

Os valores das ações possuem média de R\$ 11.477,00 e a mediana de R\$ 1.000,00. Entretanto, excluindo os 2 valores mais discrepantes, a média altera-se para R\$ 10.574,00 e a mediana mantém o valor inicial. Portanto, um grande número de processos de judicialização da saúde, independentemente de estar relacionadas à portadores de doenças raras, traz um gasto significativo ao poder judiciário, além de sugerir a necessidade de uma maior atenção a políticas orçamentárias para saúde.

Dessa forma, são apreciadas iniciativas de estudos e propostas que busquem minimizar a quantidade de tais processos, inclusive do campo de big data, contribuindo com ganhos em termos de dinheiro, tempo e vidas. A diminuição de processos dessa natureza diminuiria um gasto significativo nos cofres públicos.

Ademais, os resultados trazem indícios de que as regionais DRS I Grande São Paulo e a XII Registro, possuem uma quantidade menor de processos se comparadas com restante da base de dados. De forma similar, as DRSs II Araçatuba, III Araraquara e XV São José do Rio Preto, possuem uma maior quantidade de processos. A literatura aponta que, de maneira geral, há uma tendência de aumento no número de processos no estado de São Paulo e os achados dessa pesquisa enriquece essa informação demonstrando regiões específicas, em particular referentes a doenças raras, que são afetadas por tais processos.

Buscando responder o objetivo específico A; identificar as características mais recorrentes nos processos públicos de doenças raras do estado de São Paulo. Por exemplo, semelhanças em características sociodemográficas e padrões de tratamentos; nota-se um destaque positivo para a DRS XV São José do Rio Preto, com melhores índices de esgotamento sanitário adequado, percentual da população com rendimento nominal mensal per capita de até 1/2 salário mínimo e população ocupada, enquanto XII Registro apresenta os piores valores nos mesmos quesitos, além de ter uma quantidade de Estabelecimentos de Saúde SUS desfavorável e uma posição desfavorável as demais em relação ao PIB per capita.

Entretanto, XII Registro apresenta uma quantidade menor de processos esperados, enquanto XV São José do Rio Preto possui uma maior quantidade de processos. Esse resultado corrobora com o indício de que pessoas com maior renda e/ou mais capacitadas, são beneficiadas por processos de judicialização, pois uma população sem meios para acesso a informações de qualidade, dificilmente saberá que há a opção judicial para obter tratamentos de saúde.

Quanto aos padrões de tratamento, predominou o tratamento médico-hospitalar e/ou fornecimento de medicamentos, fornecimento de medicamentos e planos de saúde. Não houve diferenças relevantes nas alterações entre as DRSs.

A partir da mineração de processos, foi possível identificar em quais atividades as ações levam mais tempo para serem solucionadas e, conseqüentemente, faz com que o andamento total do processo seja mais demorado. Portanto, a fim de atender o objetivo específico B; obter informações, a partir de mineração de dados, e localizar os principais padrões de movimentações nos processos públicos de portadores de doenças raras; foram identificados os gargalos dos processos.

Considerando a média de todos os processos, destaca-se com maior retardo a tramitação de " Intimação, pagamento, cumprimento de sentença" para "Sentença data"; "Remessa 2ª instância" para "Sentença data"; "Alvará, levantamento, data" para "Sentença data"; "Extinção de processo" para "Arquivamento de processo"; e de " Transito julgado" para "Arquivamento de processo". Entretanto, analisando o processo mais longo da base de dados, com duração de 19 anos e 194 dias, nota-se diferenças com a média geral de atrasos, sendo os períodos de transferência mais longos de "Distribuição" para "Remessa 2ª instância", do " Transito julgado" para "Intimação, pagamento, cumprimento de sentença" e de "Alvará, levantamento, data" para "Arquivamento de processo".

Similar aos resultados encontrados no Tableau, também foram identificados dois momentos em que um grande número de casos estava ativo, em 2013 e 2015. Investigando os

processos de 2013, encontrou-se 702 casos, aproximadamente 1% da base total, com 12 atividades e duração média de 28, 8 semanas. Além disso, houveram 5 momentos de destaque em termos de duração dos movimentos considerando a duração total de todos os processos: de "remessa 2ª instância" para "sentença data"; de "início execução" para "sentença data"; de "alvará, levantamento, data" para "sentença data"; de "alvará, levantamento, data" para "transito julgado" e, por fim, de "extinção de processo" para "transito julgado".

Os processos filtrados do ano de 2015 tiveram resultados similares ao recorte de 2013, com 694 casos, aproximadamente 1% da base total, 12 atividades, duração média de 27, 3 semanas e também 5 momentos de destaque em termos de duração dos movimentos considerando a duração total de todos os processos. Entretanto, foram apenas 3 movimentos iguais ao recorte anterior, de "alvará, levantamento, data" para "sentença data" de "extinção de processo" para "transito julgado" e de "remessa 2ª instância" para "sentença data". Diferenciado foram "sentença data" para "transito julgado" e "transito julgado" para "arquivamento de processo".

Portanto, nota-se tanto nos resultados de 2013, como nos de 2015, 2 semelhanças com os gargalos da média de todos os processos, "Remessa 2ª instância" para "Sentença data" e "Alvará, levantamento, data" para "Sentença data". Dessa forma, aumentam os indícios da necessidade de propor soluções para reduzir o tempo gasto para a concluir as 2 movimentações mencionadas. Vale ressaltar que, conforme mencionado no capítulo anterior, os padrões de movimentação dos processos dos 2 recortes são diferentes, o que indica que as razões pelas quais ocorreu o primeiro pico, não justifica, de forma precisa, a ocorrência do segundo pico.

Por fim, respondendo ao objetivo específico C; apresentar, a partir de técnicas de análise e visualização de dados, as informações mais relevantes contidas nos processos de forma simplificada, a fim de permitir a compreensão do público geral; foram respondidas com auxílio do Tableau algumas perguntas que podem surgir no contexto da judicialização da saúde.

Buscando compreender quais são as regiões mais afetadas por processos relacionados a planos de saúde e qual tem sido a predominância das sentenças nesses casos, os dashboards revelaram que IX Marília e VII Campinas são as regiões mais afetadas e a predominância das sentenças é procedente.

Além disso, considerando que nos anos de 2013 e 2015 houveram aumentos nas quantidades de processos, se propôs identificar as características desses períodos que se diferenciam das características gerais da base de dados. Os dois períodos tiveram os mesmos resultados. Em termos de valores, embora se tenha a maior quantidade de processos em julho no caso de 2013, com 522 ações, e em novembro no caso de 2015, com 554 ações, a maior

soma ocorreu em setembro, contabilizando R\$ 9.693.000,00 no total de 318 processos em 2013 e 329 em 2015. No caso das sentenças e dos assuntos, o padrão se manteve e as DRSs não apresentaram alterações significativas.

Considerando que a DRS XII Registro apresentou resultados de destaque, tanto por possuir uma menor quantidade de processos do que o esperado, como por possuir dados sociodemográficos, de maneira geral, inferiores as demais DRSs, também se utilizaram dashboards para investigar as características dos processos dessa regional. Concluiu-se que os picos de grandes volumes de processos que ocorreram nos dados gerais, não ocorrem para a DRS em questão. O assunto predominou tratamento médico-hospitalar e/ou fornecimento de medicamentos e as sentenças mantiveram o padrão.

Portanto, a presente pesquisa revela um significativo potencial gerencial. Tanto o software Tableau, como o Fluxicon Disco, permitem modificar a interação entre as diversas variáveis, em busca de encontrar respostas que possam se materializar como soluções práticas. Logo, reforça a premissa de que os recursos provenientes das técnicas de Big Data trazem ganhos em termos de tempo e, conseqüentemente, em termos de saúde.

As classes processuais envolvem estudos principalmente do campo do Direito. Dessa forma, uma limitação dessa pesquisa, enquadra-se na falta de aprofundamento dos resultados referentes a classes. Portanto, sugere-se a colaboração de estudiosos da área processual para o desenvolvimento de estudos interdisciplinares, explorando as informações contidas nas classes dessa pesquisa.

Além disso, considerando que foram listadas 200 doenças raras para coleta e foram identificadas nos conteúdos apenas 68 processos distribuídos em 5 doenças diferentes, sugere-se para pesquisas futuras a investigação de medicamentos associados a doenças raras. Os dados demonstraram que entre os 7 assuntos distintos distribuídos nos processos, a maioria corresponde a Tratamento Médico-Hospitalar e/ou Fornecimento de Medicamentos (82,85%), seguido de Fornecimento de Medicamentos (13,80%). Portanto, reforça a premissa de que estudos investigando medicamentos associados a doenças raras trará contribuições relevantes.

REFERÊNCIAS BIBLIOGRÁFICAS

- Andrienko, N., Andrienko, G., Miksch, S., Schumann, H., & Wrobel, S. (2021). A theoretical model for pattern discovery in visual analytics. *Visual Informatics*, 5(1), 23–42. <https://doi.org/10.1016/J.VISINF.2020.12.002>
- Barreiro, G. S. de S., & Furtado, R. P. M. (2015). Inserindo a judicialização no ciclo de políticas públicas. *Revista de Administração Pública*, 49(2), 293–314.
- Benchimol, J., Kazinnik, S., & Saadon, Y. (2022). Text mining methodologies with R: An application to central bank texts. *Machine Learning with Applications*, 8, 100286. <https://doi.org/10.1016/J.MLWA.2022.100286>
- Borges, D. da C. L., & Ugá, M. A. D. (2010). Conflitos e impasses da judicialização na obtenção de medicamentos: as decisões de 1ª instância nas ações individuais contra o Estado do Rio de Janeiro, Brasil, em 2005. *Cadernos de Saúde Pública*, 26(1), 59–69. <https://doi.org/10.1590/S0102-311X2010000100007>
- Brentani, A., Grisi, S. J. F. E., Taniguchi, M. T., Ferrer, A. P. S., de Moraes Bourroul, M. L., & Fink, G. (2016). Rollout of community-based family health strategy (programa de saúde de família) is associated with large reductions in neonatal mortality in São Paulo, Brazil. *SSM - Population Health*, 2, 55–61. <https://doi.org/10.1016/j.ssmph.2016.01.001>
- Bretas, J., Ferreira Junior, S., & Lucena Ruas Riani, J. (2021). INCORPORAÇÃO DE TECNOLOGIAS NO SUS E POSSÍVEIS RELAÇÕES COM A JUDICIALIZAÇÃO DA SAÚDE. *RAHIS- Revista de Administração Hospitalar e Inovação Em Saúde*, 18(1), 30–43. <https://doi.org/10.21450/RAHIS.V18I1.6667>
- Brownstein, N. C., Adolfsson, A., & Ackerman, M. (2019). Descriptive statistics and visualization of data from the R datasets package with implications for clusterability. *Data in Brief*, 25, 104004. <https://doi.org/10.1016/j.dib.2019.104004>
- Bueno, S. (2010). *Dicionário da Língua Portuguesa* (1st ed., Vol. 1). São Paulo: DCL.
- Burmeister, E., & Aitken, L. M. (2012). Sample size: how many is enough? *Australian Critical Care: Official Journal of the Confederation of Australian Critical Care Nurses*, 25(4), 271–274. <https://doi.org/10.1016/J.AUCC.2012.07.002>
- Cairo, A. (2013). *The Functional Art: An introduction to information graphics and visualization (Voices That* (1st ed.). Berkeley, CA: New Riders.
- Censo IBGE estado de São Paulo. (n.d.). Retrieved March 16, 2022, from <https://cidades.ibge.gov.br/brasil/sp/panorama>
- Chiavegatto Filho, A. D. P. (2015). Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. *Epidemiologia e Serviços de Saúde*, 24(2), 325–332. <https://doi.org/10.5123/s1679-49742015000200015>
- Chung, Y., Bagheri, N., Salinas-Perez, J. A., Smurthwaite, K., Walsh, E., Furst, M. A., ... Salvador-Carulla, L. (2020). Role of visual analytics in supporting mental healthcare systems research and policy: A systematic scoping review. *International Journal of Information Management*, 50, 17–27. <https://doi.org/10.1016/J.IJINFOMGT.2019.04.012>
- Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387), 531–554. <https://doi.org/10.1080/01621459.1984.10478080>
- Constituição da República Federativa do Brasil de 1988, Governo do Brasil § (1988). Brasília: Presidência da República.
- Couto, M. B., & Oliveira, S. P. de. (2016). Gestão da justiça e do conhecimento: a contribuição da jurimetria para a administração da justiça. *Revista Jurídica*, 2(43), 771–801. DOI: 10.6084/m9.figshare.4667898
- Dantas, M. K., Oliveira, L. R. de, Ferolla, L. M., Paschoalotto, M. A. C., Lopes, J. E. F., Passador, J. L., & Passador, C. S. (2017). Cross-sectoral assessment of public policies in

- health and the environment: Scenario of the municipalities in the state of Sao Paulo. *Evaluation and Program Planning*, 65, 30–39. <https://doi.org/10.1016/J.EVALPROGPLAN.2017.06.003>
- De Leoni, M., Van Der Aalst, W. M. P., & Dees, M. (2016). A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs. *Information Systems*, 56, 235–257. <https://doi.org/10.1016/J.IS.2015.07.003>
- Departamentos Regionais de Saúde - Secretaria da Saúde - Governo do Estado de São Paulo. (n.d.). Retrieved December 9, 2021, from <http://www.saude.sp.gov.br/ses/institucional/departamentos-regionais-de-saude/?page=1>
- Dewi, L. C., Meiliana, & Chandra, A. (2019). Social Media Web Scraping using Social Media Developers API and Regex. *Procedia Computer Science*, 157, 444–449. <https://doi.org/10.1016/J.PROCS.2019.08.237>
- EMC29. (2000). EMENDA CONSTITUCIONAL N° 29. Retrieved December 18, 2019, from http://www.planalto.gov.br/ccivil_03/constituicao/Emendas/Emc/emc29.htm
- Ferreira, J. C., & Patino, C. M. (2015). O que realmente significa o valor-p? *Jornal Brasileiro de Pneumologia*, 41(5), 485. <https://doi.org/10.1590/S1806-37132015000000215>
- Fillenbaum, G. G., Blay, S. L., Mello, M. F., Quintana, M. I., Mari, J. J., Bressan, R. A., & Andreoli, S. B. (2019). Use of mental health services by community-resident adults with DSM-IV anxiety and mood disorders in a violence-prone area: São Paulo, Brazil. *Journal of Affective Disorders*, 250, 145–152. <https://doi.org/10.1016/j.jad.2019.03.010>
- Fisher, M. J., & Marshall, A. P. (2009). Understanding descriptive statistics. *Australian Critical Care*, 22(2), 93–97. <https://doi.org/10.1016/j.aucc.2008.11.003>
- Galetsis, P., Katsaliaki, K., & Kumar, S. (2020). Big data analytics in health sector: Theoretical framework, techniques and prospects. *International Journal of Information Management*, 50, 206–216. <https://doi.org/10.1016/J.IJINFOMGT.2019.05.003>
- Guimarães, T., Lucas, K., & Timms, P. (2019). Understanding how low-income communities gain access to healthcare services: A qualitative study in São Paulo, Brazil. *Journal of Transport and Health*, 15, 1–17. <https://doi.org/10.1016/j.jth.2019.100658>
- Guo, R., Fujiwara, T., Li, Y., Lima, K. M., Sen, S., Tran, N. K., & Ma, K. L. (2020). Comparative visual analytics for assessing medical records with sequence embedding. *Visual Informatics*, 4(2), 72–85. <https://doi.org/10.1016/J.VISINF.2020.04.001>
- Hair, Joseph F. , Black, Jr, William C. Babin, Barry J. & Anderson, R. E. (2005). *Análise multivariada de dados*. (Bookman, Ed.), *Pearson New International Edition* (5th ed.).
- Hair, J. F., Wolfinbarger, M., Ortinau, D. J., & Bush, R. P. (2008). *Fundamentos de pesquisa de marketing*. (Bookman, Ed.). <https://doi.org/10.1017/CBO9781107415324.004>
- Hasani, N., Farhadi, F., Morris, M. A., Nikpanah, M., Rahmim, A., Xu, Y., ... Saboury, B. (2022). Artificial Intelligence in Medical Imaging and its Impact on the Rare Disease Community: Threats, Challenges and Opportunities. *PET Clinics*, 17(1), 13–29. <https://doi.org/10.1016/J.CPET.2021.09.009>
- Hindle, G. A., & Vidgen, R. (2018). Developing a business analytics methodology: A case study in the foodbank sector. *European Journal of Operational Research*, 268(3), 836–851. <https://doi.org/10.1016/J.EJOR.2017.06.031>
- IBGE. Brasil em Síntese. São Paulo. Panorama. (2018). Retrieved December 3, 2018, from <https://cidades.ibge.gov.br/brasil/sp/panorama>
- İnce, Ö., & Güre, M. D. P. (2021). Evaluation of rare diseases policy performance of oecd countries using mcdm methods. *Health Policy and Technology*, 10(3), 100537. <https://doi.org/10.1016/J.HLPT.2021.100537>
- INSPER. (2019). *JUDICIALIZAÇÃO DA SAÚDE NO BRASIL: perfil das demandas, causas e propostas de solução*. São Paulo - SP.
- Isson, J. P., & Harriott, J. (2016). *People analytics in the era of big data : changing the way*

- you attract, acquire, develop, and retain talent* (1st ed.). New Jersey: Wiley.
- Jena, B. (2019). An Approach for Forecast Prediction in Data Analytics Field by Tableau Software. *International Journal of Information Engineering and Electronic Business*, *11*(1), 19–26. <https://doi.org/10.5815/ijieeb.2019.01.03>
- Knoppers, B. M., & Thorogood, A. M. (2017). Ethics and Big Data in health. *Current Opinion in Systems Biology*, *4*, 53–57. <https://doi.org/10.1016/J.COISB.2017.07.001>
- Lamprea, E. (2017). The Judicialization of Health Care: A Global South Perspective. *Annual Review of Law and Social Science*, *13*(1), 431–449. <https://doi.org/10.1146/annurev-lawsocsci-110316-113303>
- Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management*, *36*(5), 700–710. <https://doi.org/10.1016/j.ijinfomgt.2016.04.013>
- Lee, J. (2020). Statistics, Descriptive. In *International Encyclopedia of Human Geography* (pp. 13–20). Elsevier. <https://doi.org/10.1016/b978-0-08-102295-5.10428-7>
- Leirião, L. F. L., & Miraglia, S. G. E. K. (2019). Environmental and health impacts due to the violation of Brazilian emissions control program standards in Sao Paulo Metropolitan Area. *Transportation Research Part D: Transport and Environment*, *70*, 70–76. <https://doi.org/10.1016/J.TRD.2019.03.006>
- Liu, Y.-H. (2018). Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study. *Expert Systems with Applications*, *111*, 2–10. <https://doi.org/10.1016/J.ESWA.2018.05.018>
- Lopes, F. D., & Mello, T. R. de C. (2018). A Judicialização e seu Papel na Efetivação do Direito à Saúde Pública. *Revista de Gestão Em Sistemas de Saúde*, *7*(3), 275. <https://doi.org/10.5585/rgss.v7i3.433>
- Loshin, D. (2003). *Business Intelligence: The Savvy Manager's Guide*. (M. Kaufmann, Ed.) (1st ed.). Elsevier Science.
- Ltifi, H., Benmohamed, E., Kolski, C., & Ben Ayed, M. (2016). Enhanced visual data mining process for dynamic decision-making. *Knowledge-Based Systems*, *112*, 166–181. <https://doi.org/10.1016/J.KNOSYS.2016.09.009>
- Malhotra, N. K. (2008). *Investigación de mercados*. (Pearson Educación, Ed.) (5th ed.). Mexico. Retrieved from https://www.academia.edu/15023888/Investigacion_de_Mercados_-_Naresh_Malhotra?auto=download
- Malhotra, N. K. (2012). *Pesquisa de Marketing: Uma Orientação Aplicada*. (A. Editora, Ed.) (6th ed.).
- Mariano, C. M. (2017). Emenda constitucional 95/2016 e o teto dos gastos públicos: Brasil de volta ao estado de exceção econômico e ao capitalismo do desastre. *Revista de Investigações Constitucionais*, *4*(1), 259. <https://doi.org/10.5380/rinc.v4i1.50289>
- Marinheiro, A., & Bernardino, J. (2015). Experimental Evaluation of Open Source Business Intelligence Suites using OpenBRR. *IEEE Latin America Transactions*, *13*(3), 810–817. <https://doi.org/10.1109/TLA.2015.7069109>
- Mendes, Á., Louvison, M. C. P., Ianni, A. M. Z., Leite, M. G., Feuerwerker, L. C. M., Tanaka, O. Y., ... Almeida, C. A. L. (2015). O processo de construção da gestão regional da saúde no estado de São Paulo: subsídios para a análise. *Saúde e Sociedade*, *24*(2), 423–437. <https://doi.org/10.1590/S0104-12902015000200003>
- Merriam, J., & Grossman, J. B. (2015). Judicial Review. *International Encyclopedia of the Social & Behavioral Sciences*, 897–901. <https://doi.org/10.1016/B978-0-08-097086-8.86041-9>
- MINISTÉRIO DA SAÚDE. (n.d.). Doenças raras. Retrieved November 8, 2021, from <https://antigo.saude.gov.br/saude-de-a-z/doencas-raras>

- Morais, H. M. M. de, Albuquerque, M. do S. V. de, Oliveira, R. S. de, Cazuzu, A. K. I., & Silva, N. A. F. da. (2018). Organizações Sociais da Saúde: uma expressão fenomênica da privatização da saúde no Brasil. *Cadernos de Saúde Pública*, 34(1). <https://doi.org/10.1590/0102-311x00194916>
- Munoz-Gama, J., Martin, N., Fernandez-Llatas, C., Johnson, O. A., Sepúlveda, M., Helm, E., ... Zerbato, F. (2022). Process mining for healthcare: Characteristics and challenges. *Journal of Biomedical Informatics*, 127. <https://doi.org/10.1016/J.JBI.2022.103994>
- Nalchigar, S., & Yu, E. (2018). Business-driven data analytics: A conceptual modeling framework. *Data & Knowledge Engineering*, 117, 359–372. <https://doi.org/10.1016/J.DATAK.2018.04.006>
- Nam, D., Lee, J., & Lee, H. (2019). Business analytics adoption process: An innovation diffusion perspective. *International Journal of Information Management*, 49, 411–423. <https://doi.org/10.1016/j.ijinfomgt.2019.07.017>
- Nguyen, E. (2013). Text Mining and Network Analysis of Digital Libraries in R. In *Data Mining Applications with R* (pp. 95–115). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-411511-8.00004-9>
- Ochoa, C. (2013). Qual é o tamanho da amostra que eu preciso? Retrieved October 21, 2021, from <https://www.netquest.com/blog/br/blog/br/qual-e-o-tamanho-de-amostra-que-preciso>
- Oliva, S. Z., & Felipe, J. C. (2018). Optimizing Public Healthcare Management Through a Data Warehousing Analytical Framework. *IFAC-PapersOnLine*, 51(27), 407–412. <https://doi.org/10.1016/J.IFACOL.2019.02.004>
- Omar, Y. M., Minoufekr, M., & Plapper, P. (2019). Business analytics in manufacturing: Current trends, challenges and pathway to market leadership. *Operations Research Perspectives*, 6, 100127. <https://doi.org/10.1016/j.orp.2019.100127>
- Pereira, A., Miranda Kretzschmar, A. K., Tonini, K., Xavier, L. C., & Barros, S. C. (2020). PDG6 HEALTH JUDICIALIZATION IN BRAZIL: impact of miglustate medicine obligation for patients without official indication of use, in 2019 in brazil. *Value in Health*, 23, S130. <https://doi.org/10.1016/J.JVAL.2020.04.297>
- Pereira da Veiga, C., Drummond-Lage, A. P., Da Veiga, C., & Wainstein, A. J. A. (2019). PCN39 access to medical interventions in Brazil: the role of clinical research and of health judicialization. *Value in Health Regional Issues*, 19, 20. <https://doi.org/10.1016/j.vhri.2019.08.107>
- Process Mining and Automated Process Discovery Software for Professionals - Fluxicon Disco. (n.d.). Retrieved January 19, 2022, from <https://fluxicon.com/disco/>
- Rafiei, M., & van der Aalst, W. M. P. (2021). Group-based privacy preservation techniques for process mining. *Data and Knowledge Engineering*, 134. <https://doi.org/10.1016/J.DATAK.2021.101908>
- Rovani, M., Maggi, F. M., De Leoni, M., & Van Der Aalst, W. M. P. (2015). Declarative process mining in healthcare. *Expert Systems with Applications*, 42(23), 9236–9251. <https://doi.org/10.1016/J.ESWA.2015.07.040>
- Ryngelblum, A. L., Oliveira, P. C. D. De, Telles, R., & Costa, M. C. (2019). Evasão estratégica das regulações: a judicialização do sistema privado de saúde brasileiro. *Cadernos EBAPB.BR*, 17(3), 623–638. <https://doi.org/10.1590/1679-395175094>
- Sá, T. H. de, Tainio, M., Goodman, A., Edwards, P., Haines, A., Gouveia, N., ... Woodcock, J. (2017). Health impact modelling of different travel patterns on physical activity, air pollution and road injuries for São Paulo, Brazil. *Environment International*, 108, 22–31. <https://doi.org/10.1016/J.ENVINT.2017.07.009>
- Scott, M., Flaherty, D., & Currall, J. (2013). Statistics: are we related? *Journal of Small Animal Practice*, 54(3), 124–128. <https://doi.org/10.1111/JSAP.12029>

- Siena, O., Carvalho, E. M., Netto, F. L. F., & Silva, R. C. (2021). Impactos Financeiros e Orçamentários da Judicialização da Saúde no Estado de Rondônia. *Amazônia, Organizações e Sustentabilidade*, 10(2), 266. <https://doi.org/10.17648/AOS.V10I2.2419>
- Silva Junior, G. B. da, & Dias, E. R. (2016). Avaliação da satisfação dos usuários de um serviço de saúde público-privado no nordeste do Brasil e a judicialização da saúde. *Revista de Direito Sanitário*, 17(2), 13. <https://doi.org/10.11606/issn.2316-9044.v17i2p13-29>
- Simpao, A. F., Ahumada, L. M., & Rehman, M. A. (2015). Big data and visual analytics in anaesthesia and health care. *British Journal of Anaesthesia*, 115(3), 350–356. <https://doi.org/10.1093/BJA/AEU552>
- Sousa, D. F. de, Rodrigues, S., Lima, H. V. de, & Chagas, L. T. (2020, January 1). R software packages as a tool for evaluating soil physical and hydraulic properties. *Computers and Electronics in Agriculture*. Elsevier B.V. <https://doi.org/10.1016/j.compag.2019.105077>
- Tax, N., Lu, X., Sidorova, N., Fahland, D., & van der Aalst, W. M. P. (2018). The imprecisions of precision measures in process mining. *Information Processing Letters*, 135, 1–8. <https://doi.org/10.1016/J.IPL.2018.01.013>
- Tursunbayeva, A., Di Lauro, S., & Pagliari, C. (2018, December 1). People analytics—A scoping review of conceptual boundaries and value propositions. *International Journal of Information Management*. Elsevier Ltd. <https://doi.org/10.1016/j.ijinfomgt.2018.08.002>
- Ul-Ain, N., Giovanni, V., DeLone, W. H., & Waheed, M. (2019). Two decades of research on business intelligence system adoption, utilization and success – A systematic literature review. *Decision Support Systems*, 113113. <https://doi.org/10.1016/J.DSS.2019.113113>
- Vairão Junior, N. S., & Alves, F. J. dos S. (2017). A EMENDA CONSTITUCIONAL 95 E SEUS EFEITOS. *Revista de Contabilidade Do Mestrado Em Ciências Contábeis Da UERJ*, 22(2), 54–75. <https://doi.org/10.12979/32365>
- Van Der Aalst, W. M. P. (2019). A practitioner's guide to process mining: Limitations of the directly-follows graph. *Procedia Computer Science*, 164, 321–328. <https://doi.org/10.1016/J.PROCS.2019.12.189>
- Vargas-Peláez, C. M., Rover, M. R. M., Leite, S. N., Rossi Buenaventura, F., & Farias, M. R. (2014, November 1). Right to health, essential medicines, and lawsuits for access to medicines - A scoping study. *Social Science and Medicine*. Elsevier Ltd. <https://doi.org/10.1016/j.socscimed.2014.08.042>
- Vargas-Pelaez, C. M., Rover, M. R. M., Soares, L., Blatt, C. R., Mantel-Teeuwisse, A. K., Rossi, F. A., ... Farias, M. R. (2019). Judicialization of access to medicines in four Latin American countries: a comparative qualitative analysis. *International Journal for Equity in Health*, 18(1), 1–14. Retrieved from <https://doi.org/10.1186/s12939-019-0960-z>
- Wang, S., Yeoh, W., Richards, G., Wong, S. F., & Chang, Y. (2019). Harnessing business analytics value through organizational absorptive capacity. *Information & Management*, 103152. <https://doi.org/10.1016/J.IM.2019.02.007>
- Watanabe, E. (2015). A arquitetura de metadados de processos judiciais cíveis. *Encontros Nacionais de Pesquisa e Pós-Graduação Em Ciência Da Informação*, 1128–1134. Retrieved from <http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/2670>
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). What's it all about? In *Data Mining* (pp. 3–41). Elsevier. <https://doi.org/10.1016/B978-0-12-804291-5.00001-5>
- Younger, D. S. (2016). Health Care in Brazil: Implications for Public Health and Epidemiology. *Neurologic Clinics*, 34(4), 1071–1083. <https://doi.org/10.1016/J.NCL.2016.06.002>
- Zhao, Y., Zhang, C., Zhang, Y., Wang, Z., & Li, J. (2020). A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis. *Energy and Built Environment*, 1(2), 149–164. <https://doi.org/10.1016/j.enbenv.2019.11.003>