

Association for Information Systems

AIS Electronic Library (AISeL)

CONF-IRM 2021 Proceedings

International Conference on Information
Resources Management (CONF-IRM)

Summer 2021

Ferramenta de Visualização de Dados Públicos da Saúde Disponibilizados pelo DATASUS

Paula Monteavaro Franceschini

Josiane Brietzke Porto

Rafael Kunst

Follow this and additional works at: <https://aisel.aisnet.org/confirm2021>

This material is brought to you by the International Conference on Information Resources Management (CONF-IRM) at AIS Electronic Library (AISeL). It has been accepted for inclusion in CONF-IRM 2021 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Ferramenta de Visualização de Dados Públicos da Saúde Disponibilizados pelo DATASUS

Paula Monteavaro Franceschini
Universidade do Vale do Rio dos Sinos
paula.franceschini@edu.unisinos.br

Josiane Brietzke Porto
Universidade do Vale do Rio dos Sinos
josibrietzke@unisinos.br

Rafael Kunst
Universidade do Vale do Rio dos Sinos
rafaelkunst@unisinos.br

Abstract

For a long time, health data visualization techniques have been an essential and determining field for generating knowledge and directing public policies. In Brazil, the Informatics Department of the Unified Health System (DATASUS), is main responsible for storing brazilian public health-related data. Nowadays, data is considered a valuable item, thus, access to public information must be facilitated in order to be explored and understood by a greatest amount of people. This article presents a visualization tool for the Brazilian public health hospitalization system dataset, provided by DATASUS. The developed tool uses different data visualization techniques to support the analysis of epidemiological data. With the tool developed through the Design Science Research, it is possible to analyze the behavior of the causes of SUS hospital admissions, by gender, age or time of year. It is also possible to compare reasons for hospitalization over the past ten years and the main causes of hospitalizations in all states of Brazil.

Keywords: Big Data, Health Data Visualization, DATASUS, Covid-19, Design Science Research.

1. Introdução

A Lei de Acesso à Informação (Lei 12.527/2011) regulamenta o acesso e permite consulta aos dados de diversos órgãos governamentais pela população brasileira (Brasil, 2011). Em consonância com essa lei, o portal do Departamento de Informática do Sistema Único de Saúde (DATASUS), que já disponibilizava diversas bases de dados desde sua criação, se tornou importante ferramenta para consulta de dados abertos da saúde pública no Brasil (Ventura, 2013)

Nesse portal há dois aplicativos para consulta de dados abertos: o *TabNet* e o *TabWin*. Entretanto, os dados disponibilizados não são padronizados nem integrados e não são usadas técnicas e ferramentas avançadas para visualização, tornando a obtenção de informações uma tarefa extremamente árdua (Pires et al., 2011). Por esse motivo, dificuldades podem ser encontradas para explorá-los, além de ser uma experiência complexa para um cidadão sanar dúvidas a respeito da saúde no Brasil, uma vez que a visualização dessas informações não é intuitiva nem trivial no portal.

A visualização de dados é uma forma de apresentar histórias sobre dados para humanos, que são mais orientados visualmente (Qin et al., 2020). Desta forma, técnicas de visualização de dados aplicadas aos dados abertos do Sistema Único de Saúde (SUS) podem fazer com que mais pessoas consigam conhecer, compreender e explorar estes dados. Diante deste cenário surgiu a seguinte questão norteadora da presente pesquisa: Como facilitar e tornar acessível

informações sobre saúde pública brasileira, a partir de dados abertos disponibilizados pelo DATASUS?

Nesta direção, em busca de uma possível resposta para tal questão de pesquisa, esse estudo teve por objetivo facilitar o acesso e a exploração dos dados do DATASUS, usando conceitos de *Big Data* e técnicas de Inteligência Artificial, para implementação de uma ferramenta web de visualização desses dados, sob o método *Design Science Research*. Para isso, foi necessário o alcance dos seguintes objetivos específicos: (i) obter e analisar os dados disponibilizados pelo DATASUS; (ii) definir visualizações para exploração desses dados, de acordo com suas características; (iii) desenvolver uma ferramenta de visualização de dados, a partir desses dados abertos.

Justifica-se por prover transparência aos cidadãos brasileiros e por permitir acompanhamento social acerca de saúde pública em contexto nacional, com base em dados de fontes públicas e confiáveis. Torna-se importante para fins de monitoramento periódico de informações de interesse público, especialmente, em tempos de pandemia como a de Covid-19, em que se observa desinformação e diferentes visões de fatos, em múltiplos canais de comunicação. Essa ferramenta é relevante, inclusive, após essa pandemia, pois pode contribuir para uma melhor gestão do conhecimento e dos recursos, geralmente, escassos e dispendiosos, nas esferas de governo municipal, estadual e federal.

Por meio de emprego de tecnologias emergentes, a arquitetura da solução e a ferramenta construída nessa pesquisa podem auxiliar no tratamento dos desafios referentes à transformação digital nos processos de saúde pública, a partir do suporte a uma análise de dados abertos de forma visual, confiável e eficiente, possibilitando a descoberta de conhecimentos relevantes, não facilmente identificados, sem certo tratamento e preparação, via tecnologias especializadas. Ademais, os resultados dessa pesquisa facilitam iniciativas de educação para cidadania e para a saúde coletiva, oportunizando uma melhor experiência para todas as partes interessadas nessas informações.

Além dessa introdução, o artigo está organizado em: seção 2 traz conceitos de *big data* e visualização de dados em saúde pública; seção 3 explica os dados do DATASUS e traz trabalhos relacionados; seção 4 explica como foi desenvolvida a ferramenta; a seção 5 mostra os resultados obtidos e a seção 6 apresenta as considerações finais.

2. *Big Data* e Visualização de Dados na Saúde Pública

Os dados relacionados à saúde pública geralmente podem ser caracterizados como *big data*, pois são altamente complexos, em função do grande volume, várias fontes, velocidade em que são gerados e, algumas vezes, baixo grau de veracidade (Sedig, 2014). Neste contexto, técnicas de visualização de dados são extremamente úteis e podem ajudar a examinar as várias dimensões desses dados (temporal, espacial, entre outras), além da análise estatística (Joshi et al., 2017).

Além disso, a visualização de dados é uma ferramenta crítica e poderosa para auxiliar no descobrimento e entendimento dos dados da saúde, auxiliando na tomada de decisões e na comunicação entre os interessados (Martinez et al., 2016). Para tanto, os dados devem ser coletados, explorados, armazenados, analisados e finalmente convertidos em informações, sendo importante saber como mostrar essa informação de forma adequada e compreensível, para que todos os destinatários possam entendê-la. (Few, 2007).

1. Diante de uma pandemia, como a da Covid-19, a visualização de informações mostra-se essencial para pesquisas e descoberta de conhecimentos a respeito da evolução da doença e para comunicar melhor a população sobre riscos e evoluções do cenário pandêmico. Nas diferentes esferas, como por exemplo, na federal, o governo criou o “Painel Coronavírus” (Ministério da Saúde, 2020a), que apresenta um *dashboard* com diversas informações sobre a doença no Brasil. Em contexto internacional, também existem outras ferramentas de visualização de dados na saúde como o portal *HealthData.org*, que disponibiliza diversas visualizações organizadas por estudos.

3. Dados do Sistema Único de Saúde Brasileiro

O SUS é um dos maiores e mais complexos sistemas de saúde pública do mundo, garantindo acesso integral, universal e gratuito para toda a população do Brasil (Ministério da Saúde, 2020b). O DATASUS tem como responsabilidade prover sistemas de informação e suporte de informática aos órgãos do SUS, além de ser o principal responsável pelo armazenamento de dados acerca da saúde da população brasileira (Ministério da Saúde, 2019.).

Através do seu portal, o DATASUS disponibiliza conjuntos de dados de alguns sistemas sob sua gestão, além de dois programas para consulta e exploração dos dados: o *TabWin* e *TabNet*. Porém, ambos possuem limitações, como compatibilidade com apenas um sistema operacional (*Windows*), no caso do *TabWin* e o fato de só permitir tabulação de dados, sem explorar diferentes formas de visualização, no caso do *Tabnet* (Silva, 2009). Além destes programas, o DATASUS disponibiliza as bases de dados de alguns sistemas para *download*, em formato *Data Base Container* (DBC), no qual os dados são armazenados de modo compactado. Porém, pelo formato próprio, somente pode ser lido pelo *TabWin* e por uma ferramenta de linha de comando, disponibilizada pelo DATASUS (Petruzalek, 2016).

Estudos anteriores também criaram ferramentas para auxiliar na análise ou processamento dos dados do DATASUS. Para *R*, por exemplo, existem os pacotes *Microdatasus*, com função de *download* e pré-processamento desses dados (Saldanha et al., 2019) e o *read.dbc*, que lê arquivos de dados do formato “.DBC”, facilitando o processamento (Petruzalek, 2016). Outro exemplo é uma ferramenta em *Java*, para tratar os dados do DATASUS e exportá-los para um banco de dados relacional (Mendes et al., 2020). No presente trabalho foi utilizado o *PySus* (Coelho, 2020), biblioteca *Python* que disponibiliza diversas funções para auxiliar na obtenção e na análise de dados do DATASUS, inclusive, descompactando os arquivos “.DBC”.

4. Metodologia de Pesquisa

O objetivo deste trabalho foi criar uma ferramenta de visualização de dados públicos de saúde disponibilizados pelo DATASUS, facilitando análise e entendimento das informações por qualquer cidadão interessado em conhecê-los. Foi desenvolvida sob o paradigma *Design Science*, operacionalizado pelo método *Design Science Research* (DSR), visando projetar, desenvolver e avaliar um artefato de visualização de dados, conforme mostra a Figura 1 (Dresch; Lacerda; Antunes, 2015). Atualmente, as etapas Avaliação e Conclusão de DSR estão em andamento, portanto, destacadas em cinza nessa figura.

Conscientização	<ul style="list-style-type: none"> • Revisão da literatura e de ferramentas existentes • Identificação do problema • Definição do escopo
Sugestão	<ul style="list-style-type: none"> • Artefato: Ferramenta de visualização de dados do DATASUS
Desenvolvimento	<ul style="list-style-type: none"> • Escolha de uma das bases do DATASUS e de técnicas de visualização • Projeto e arquitetura da ferramenta • Desenvolvimento da versão inicial da ferramenta
Avaliação	<ul style="list-style-type: none"> • Entrevistas com representantes da Quádrupla Hélice • Revisão e refinamento do artefato
Conclusão	<ul style="list-style-type: none"> • Discussão e divulgação dos resultados parciais • Discussão e divulgação dos resultados finais • Publicação do artefato

Figura 1: Etapas da pesquisa sob o método *Design Science Research*

Em Conscientização, a partir da literatura e de ferramentas existentes, o problema identificado foi a dificuldade de exploração e de consulta de dados disponibilizados pelo DATASUS. A partir disso definiu-se o escopo e como solução proposta, o desenvolvimento de ferramenta *web*, com diferentes possibilidades de visualização e de exploração dos dados, de acordo com suas respectivas características, em Sugestão. Para projeto e construção desse artefato, na etapa Desenvolvimento foi escolhida uma das bases do DATASUS, explorada e categorizada, conforme os tipos de dados contidos nela. Após foram escolhidos dados relevantes e técnicas de visualização para exibi-los de forma a extrair informações e facilitar o entendimento.

A partir de pesquisa na literatura e de exemplos de ferramentas existentes de visualização de dados da saúde, as técnicas de visualização escolhidas foram: *Dashboard*, Linha do tempo, Pirâmide de Resultados, Evolução CID, Ocorrência por UF e Visualizações Combinadas. Em especial, Preim et al. (2019) fez pesquisa sobre visualizações analíticas em dados de saúde pública, destacando técnicas comumente usadas, visando dar suporte aos profissionais dessa área e, conseqüentemente, às medidas de prevenção. Entre as técnicas estão: os *dashboards* e visualizações combinadas, que apresentam diversas informações relevantes sobre determinado processo ou tarefa, de maneira integrada; os gráficos de linha do tempo, que se sustentam no fato de dados da saúde pública serem contínuos, geralmente (por exemplo, o número de casos de uma doença é continuamente monitorado); e a pirâmide de resultados, que mostra a frequência de determinada doença de acordo com idade e gênero da população. O *Global Burden of Disease* (GBD) é o estudo epidemiológico observacional mundial mais abrangente até o momento e examina tendências desde 1990, incluindo dados sobre mortalidade e morbidade de diversos países, informando médicos, pesquisadores e formuladores de políticas (University of Washington, 2013). As visualizações de “Evolução CID” e “Ocorrência por UF” foram inspiradas neste estudo anterior.

A construção das visualizações levou em consideração a Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde (CID). A hierarquia da CID possui até quatro níveis, sendo cada subnível mais específico em relação ao anterior. Para a construção da ferramenta considerou-se apenas o primeiro nível de hierarquia da classificação CID-10. Quanto à base dados e pipeline de processamento dos dados, para implementação das visualizações foi escolhida a base de dados do Sistema de Internações Hospitalares (SIH) do

SUS, que contém dados de grande parte das internações hospitalares no Brasil. Esses dados são disponibilizados em arquivos “.DBC”, separados por estado, mês e ano. Os arquivos de dados mais recentes possuem 113 colunas, com várias informações sobre cada internação. O processamento de dados divide-se em 4 módulos, cada um com objetivos e tarefas específicos, conforme Figura 2. A seguir, a arquitetura e o funcionamento de cada um deles são descritos.

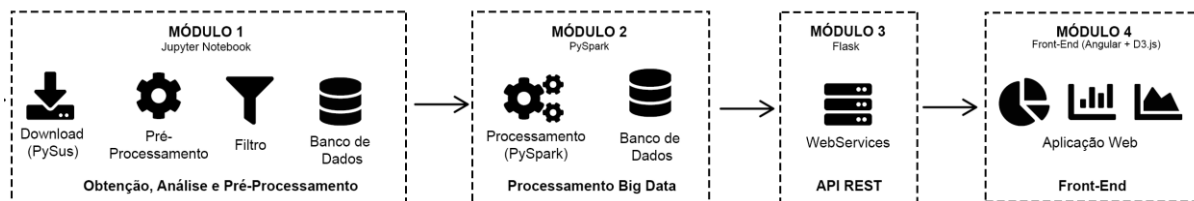


Figura 2: Módulos da ferramenta desenvolvida

O **Módulo 1** foi desenvolvido em *Python*, usando *Jupyter Notebook*, para realizar obtenção, análise e pré-processamento dos dados e o armazenamento em um banco de dados. A obtenção dos dados foi feita por meio da biblioteca *PySUS*, que disponibiliza função para buscar arquivos de dados do SIH por FTP, retornando o arquivo já descompactado e transformado em um *dataframe* da biblioteca *Python* de processamento de dados *pandas* (McKinney, 2010). Após análise foram escolhidas colunas referentes à UF, onde ocorreu a internação, sexo e idade do paciente, CID de diagnóstico principal e indicativo se houve morte, para compor as visualizações da ferramenta implementada. Com esses atributos, a obtenção dos dados foi feita da seguinte forma: Obtenção do arquivo de dados para cada estado do Brasil e para cada mês e ano do intervalo escolhido (2010 - 2020), utilizando a biblioteca *PySUS*; Filtro para manter apenas as colunas desejadas; Transformação dos tipos de campos de acordo com o tipo de dado; Gravação dos dados em um banco de dados *PostgreSQL*, sendo uma tabela para cada estado.

No **Módulo 2** (*PySpark*), após obtenção e pré-processamento dos dados, foram definidas as técnicas de visualização implementadas (ver seção 5) e o formato dos dados necessário para cada uma delas. Trata-se de outro *Jupyter Notebook*, desenvolvido para preparar os dados a fim de facilitar desenvolvimento destas técnicas de visualização. Neste módulo, dados obtidos foram processados e armazenados de acordo com as técnicas de visualização implementadas, com o objetivo de diminuir o tempo de processamento no momento da visualização. Foi criada função para cada uma das visualizações escolhidas e foi utilizado o *PySpark* para processamento dos grandes conjuntos de dados. Por fim, cada tabela criada no Módulo 1 foi processada por cada uma dessas funções e os dados gerados armazenados em tabela específica da visualização. Por exemplo, para visualização de dados que mostrará apenas quantidade de casos por CID de acordo com o gênero foi criada uma tabela, contendo o ano, estado e código CID de referência, juntamente com o total de casos para cada gênero. Desta forma, o cálculo já fica armazenado nesta tabela e não é necessário processar grande volume de dados em tempo real, diminuindo consideravelmente o tempo de espera para renderização dos gráficos.

Com os dados prontos foi construído o **Módulo 3**, uma *API REST* usando o framework *Flask*, que disponibiliza dados para consulta por *webservice*. Nesta API foi criado um *webservice* para cada tipo de visualização, que recebe como parâmetros os filtros informados na ferramenta, busca e filtra os dados na tabela específica do banco de dados e retorna um arquivo *JSON*. Por fim, o **Módulo 4**, um *front-end* da aplicação web foi construído em *Angular*, com a biblioteca *D3.js*, para construção dos gráficos. Para cada tipo de visualização foi criada uma página com um componente de filtros, que podem mudar de acordo com o tipo de visualização. Ao alterar os filtros e/ou atualizar a página o módulo de *front-end* chama o *webservice* referente

àquela visualização, informando os filtros selecionados, recebe o *JSON* com os dados filtrados e renderiza o gráfico. As visualizações desenvolvidas estão descritas na próxima seção.

5. Resultados

Na ferramenta *web* de visualização de dados do SIH do SUS é possível manipular e filtrar de acordo com o estado (UF), ano (entre 2010 e 2020) e o código do CID do grupo de classificação da doença, conforme visão geral das visualizações na ferramenta, a seguir.

5.1 Dashboard com Dados do SIH - SUS

Um *dashboard* é uma exibição visual das informações mais importantes, necessárias para atingir um ou mais objetivos, consolidadas em uma única tela, de forma que possam ser monitoradas e entendidas à primeira vista (Few, 2006). A visualização de *dashboard* da ferramenta criada visou demonstrar visão geral sobre os dados de cada UF em cada ano, em uma tela. A Figura 3(a) mostra a tela com os dados de 2020, no estado do Rio Grande do Sul (RS), por exemplo. Nela são exibidos primeiramente 4 indicadores numéricos, que apresentam: o número total de internações, a média de internações por dia, o número de altas e o número de mortes. Depois são exibidos dois gráficos: um de pizza, que mostra a divisão das internações entre homens e mulheres e outro com total de novas internações, ao longo do tempo.

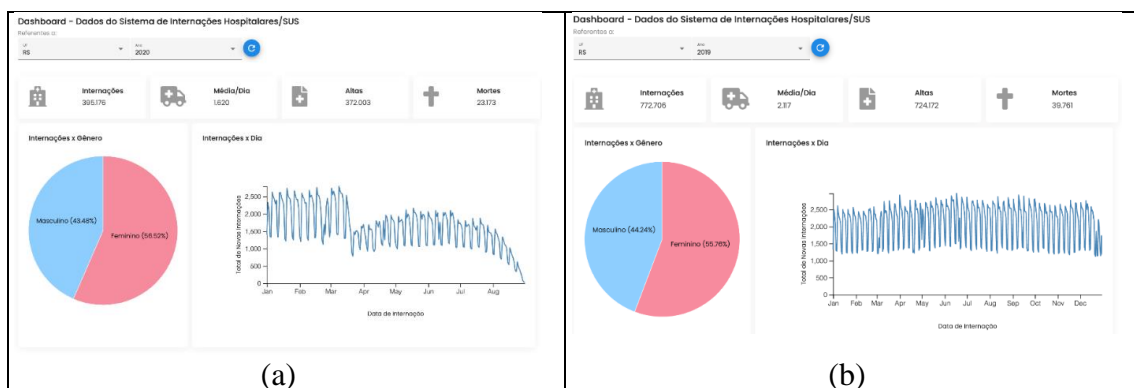


Figura 3: *Dashboard* com dados do SIH - SUS no RS nos anos de 2020 e 2019

Analisando este *dashboard* se tem panorama geral das internações pelos indicativos numéricos. Também é possível ver no primeiro gráfico, que o número de mulheres internadas foi um pouco maior do que o de homens. E, no gráfico da linha do tempo, percebe-se que houve queda no número de internações, a partir do meio de Março de 2020, quando iniciou a quarentena da Covid-19, na maior parte dos estados brasileiros. Esse dado fica mais evidente, quando comparado ao *dashboard* do RS de 2019 na Figura 3(b), em que o número de internações manteve mesmo comportamento, ao longo de todo ano.

5.2 Linha do Tempo

Geralmente, dados de saúde pública são contínuos, por serem agregados e se referirem às populações. Esse é o caso dos dados de internações, que servem para monitorar o número de casos de internação, por uma dada doença ao longo do tempo. A “Linha do Tempo” mostra essa característica dos dados, sendo possível ver se o número de casos aumentou ou não em uma época do ano e comparar o número de casos, entre diferentes grupos de doenças.

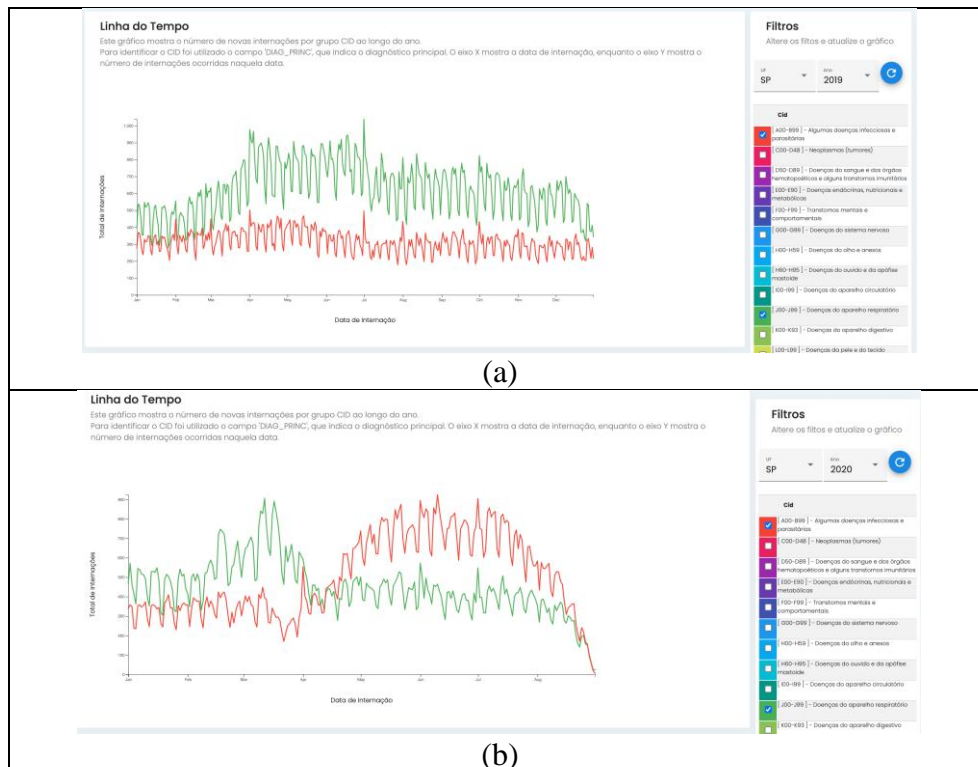


Figura 4: Linha do tempo com dados das internações no estado de SP em 2019 e 2020

Por exemplo, a Figura 4 mostra a comparação entre o número de internações por “Algumas Doenças Infecciosas e Parasitárias” e “Doenças do Aparelho Respiratório”, em São Paulo (SP). No lado direito é possível ver o componente de filtro usado na maioria das visualizações da ferramenta. Com ele, dependendo do tipo de visualização, é possível escolher um ou mais grupo CID para filtrar os dados. Além ver que, para cada grupo CID foi atribuída uma cor, com o objetivo de facilitar a identificação ao visualizar os gráficos gerados pela ferramenta. Na Figura 4(a) pode-se ver que o total de internações de “Algumas Doenças Infecciosas e Parasitárias” manteve comportamento parecido ao longo do ano, enquanto, as internações pelo outro grupo tiveram um pico no meio de 2019. Ao passar o mouse pela linha referente ao CID, essa linha fica em destaque e o sistema apresenta uma *tooltip*, com informações detalhadas sobre aquela data e de uma das linhas do gráfico. Comparando os dados de internações dos mesmos capítulos CID, entre os anos de 2019, na Figura 4(a) e de 2020, na Figura 4(b), se pode ver que nesse último ano houve inversão em SP: a partir de Março, as internações por “Algumas Doenças Infecciosas e Parasitárias” aumentaram, enquanto, as internações por “Doenças do Sistema Respiratório” diminuíram.

5.3 Pirâmide de Resultados

Uma pirâmide de resultados é baseada em uma pirâmide populacional, que é um tipo de gráfico comumente usado para descrever composição por sexo e idade de uma população. As pirâmides populacionais possuem dois histogramas justapostos verticalmente, um para pessoas do sexo masculino e outro para pessoas do sexo feminino, com um eixo vertical comum para idade, que geralmente é representada por um ano único ou categorias de intervalos de 5 anos. Quanto maior a barra do eixo vertical, maior a proporção de indivíduos daquela idade e o eixo horizontal representa o número de casos de uma determinada doença (Chui et al., 2011). Ao observar uma pirâmide de resultados, é possível perceber, pelo seu formato, se um grupo de idade e/ou gênero teve um número maior de internações por aquele motivo.

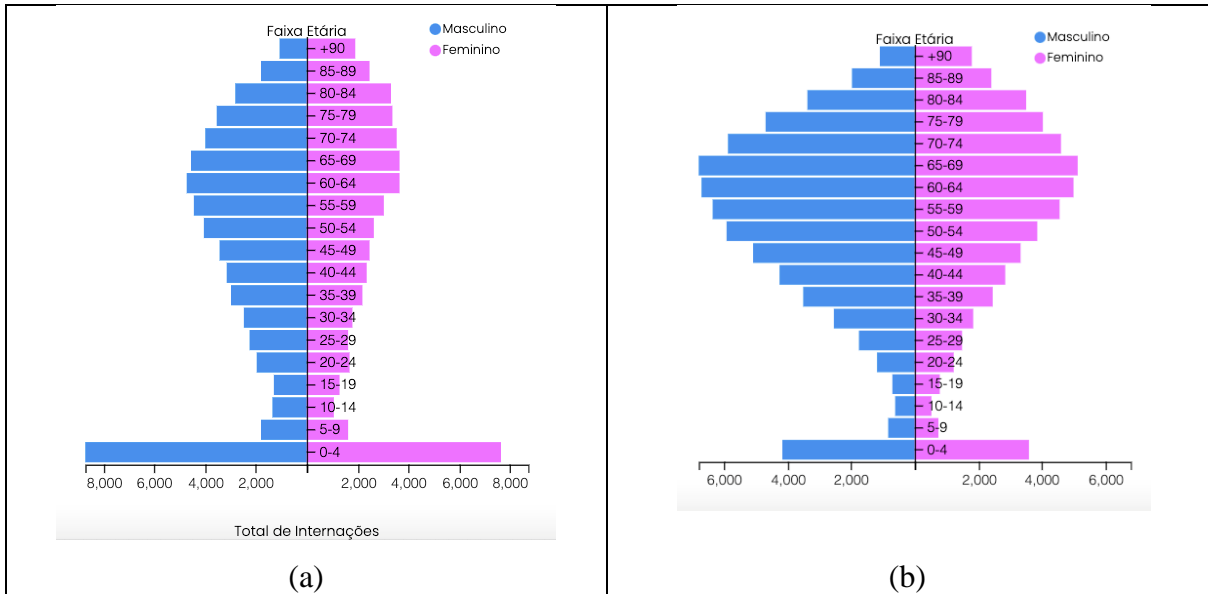


Figura 5: Pirâmides de “Algumas doenças infecciosas e parasitárias” em 2019 e 2020 em SP

A Figura 5 compara a incidência de “Algumas doenças infecciosas e parasitárias” em 2019 em SP (à esquerda) com a incidência em 2020 (à direita), por exemplo. Enquanto, em 2019, a maioria das internações desse grupo era de crianças, em 2020 houve aumento significativo da internação de adultos, o que corrobora com as características de faixa etária dos principais afetados pelo vírus Covid-19.

5.4 Diagnóstico por CID

Com objetivo de explorar o comportamento das doenças entre homens e mulheres, esta visualização compara a incidência da doença em cada gênero. Para montar o gráfico é possível escolher um ou mais grupos CID. A Figura 6 mostra a incidência de todos os grupos CIDs, nas internações em 2020 no RS, divididas por gênero. Este gráfico mostra, por exemplo, que a principal causa de internações entre indivíduos do sexo masculino no ano de 2020 no RS foi “(S00 – T98) Lesões, envenenamentos e algumas outras consequências de causas externas”, sendo também o grupo de maior diferença entre os gêneros, exceto, por “(O00 – O99) Gravidez, parto e puerpério”.

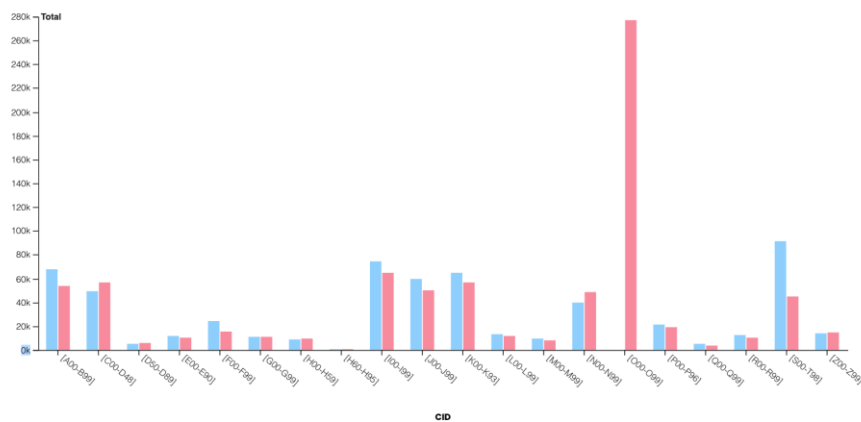


Figura 6: Incidência de internações por cada grupo CID no RS em 2020 por gênero

5.5 Evolução CID

Esta visualização foi inspirada no estudo GBD e mostra mudanças no *ranking* de internações, quando se compara um intervalo de anos. A Figura 7 mostra as dez primeiras posições dos grupos CID por internações, nos anos de 2010 e 2020, no RS. Na comparação entre esses anos houve mudanças significativas, como a queda das internações por “Doenças do aparelho respiratório”, que era a primeira posição em 2010 e foi para a sétima posição em 2020 e o aumento das internações por “Algumas afecções originadas no período perinatal”, entre outras.



Figura 7: Dez primeiras posições de grupo CID no *ranking* de internações em 2010 e 2020

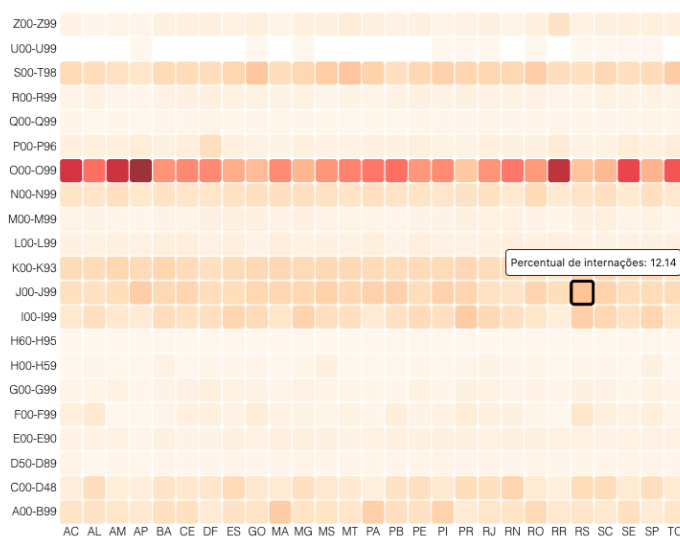


Figura 8: Mapa de calor com incidência de cada grupo CID em cada estado brasileiro

5.6 Ocorrência por UF

Inspirada em visualizações do GBD, a Figura 8 acima mostra mapa de calor entre UFs e capítulos CID. A cor de cada quadrado indica percentual de internações para aquele grupo CID naquela região. Quanto mais próximo do vermelho, maior percentual de internações e quanto mais próximo do branco, menor número de internações. Analisando-se dados de 2019 de todos estados brasileiros, o principal motivo de internações foi “(O00 – O99) Gravidez, Parto e Puerpério”, sendo maior em Amapá (AP) e Roraima (RR) e menor, no Paraná (PR) e RS.

5.7 Visualizações Combinadas

Esta visualização baseia-se em estudo anterior (Chui et al., 2011), chamada de *Multiple Coordinated View* (MCV). São exibidos gráficos com eixos combinados, permitindo visão

geral sobre dados e visões específicas ao mesmo tempo, como na Figura 9(a), que exemplifica os dados de internações por “Algumas doenças infecciosas e parasitárias” em SP, em 2019.

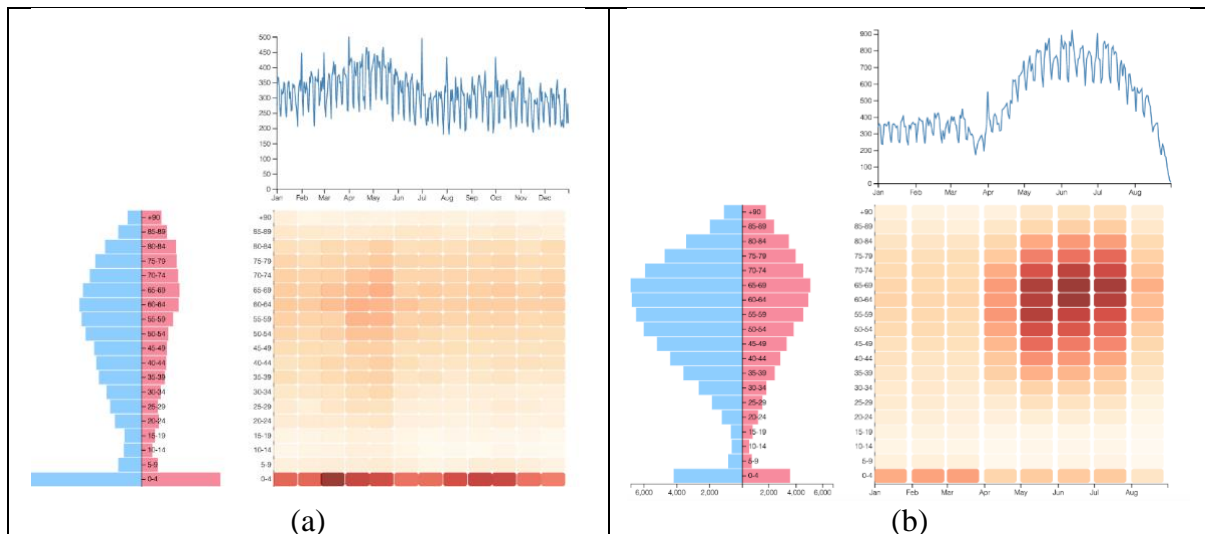


Figura 9: MCV de “Algumas doenças infecciosas e parasitárias” em SP em 2019 e 2020

Na Figura 9, o gráfico mais acima mostra a linha do tempo das internações, o mais à esquerda mostra a pirâmide de gênero e idade e o gráfico abaixo, à direita mostra um mapa de calor referente às internações ao longo do tempo e por grupo de idade. O gráfico de linha do tempo e o mapa de calor compartilham o mesmo eixo x , que indica os meses do ano. Desta forma, se vê que onde o mapa de calor apresenta cores mais escuras, a linha do tempo também apresenta valores mais altos. Além disso, o gráfico de pirâmide e o mapa de calor compartilham o mesmo eixo y . Mostra-se nos grupos de idade em que a incidência da doença é maior, o mapa de calor também possui cores mais escuras.

No exemplo da Figura 9(b) é possível ver em detalhes o que foi mostrado na Figura 5 sobre incidência de “Algumas doenças infecciosas e parasitárias” de SP em 2020 e, na Figura 4, que trouxe linha do tempo com internações nesse mesmo estado e ano. Tal visualização combinada mostra que no início de 2020, internações por esse grupo CID eram frequentes em crianças e, a partir do meio do ano, as internações por esse motivo cresceu na população adulta masculina.

6. Conclusão

Na área da saúde pública, técnicas de visualização de dados já servem como ferramentas para auxiliar nos estudos epidemiológicos há muito tempo. Com o advento da era da informação, onde dados são gerados, processados e armazenados a todo momento, a visualização de informações tornou-se ainda mais importante para auxiliar na descoberta e comunicação de conhecimentos. Sendo assim foi desenvolvida uma ferramenta focada em internações hospitalares do SUS, a partir de dados abertos do DATASUS, por meio de técnicas e ferramentas de processamento de *Big Data*.

O principal objetivo dessa ferramenta foi criar visualizações intuitivas e fáceis de serem exploradas, uma vez que as ferramentas disponibilizadas hoje pelo DATASUS são muitas vezes complexas de serem utilizadas por qualquer pessoa que não tenha conhecimentos especializados previamente, tanto sobre técnicas de visualização quanto sobre os dados abertos disponibilizados. As técnicas adotadas na criação da ferramenta foram escolhidas a partir de um estudo de ferramentas e artigos anteriores referentes à visualização de dados na área da

saúde pública. A maior parte delas tratou de explorar o comportamento epidemiológico dos grupos de doença, ao longo do tempo e/ou localização. A partir dessas visualizações criadas é possível comparar o perfil de internações em cada ano, em cada estado brasileiro. Alguns dos exemplos mostrados trazem dados do ano de 2020 e mostram as alterações no comportamento das internações em função da pandemia de Covid-19.

Essa versão inicial desenvolvida ainda será demonstrada e avaliada por representantes da Quádrupla Hélice, em ciclos da etapa Avaliação do método DSR, para fins de refinamento quanto aos objetivos da solução e aprimoramento do projeto da ferramenta proposta. Ademais, com o conhecimento aprofundado da base de dados, uma possível implementação futura seria a inclusão de *chatbot*, para auxiliar o cidadão a construir um *dashboard* personalizado de acordo com dados de interesse em visualizar, tirando dúvidas sobre as visualizações formadas, aumentando ainda mais a autonomia do usuário na análise dos dados disponibilizados.

Além de futuros trabalhos relacionados às visualizações por níveis mais específicos do CID, a análise de padrões a respeito das internações, podendo servir como insumo para a predição de uso de recursos públicos, inclusive, conforme a época do ano e/ou localização. Um estudo de correlação entre outras variáveis e a classificação de CID, a exibição de visualizações referentes aos outros campos da tabela SIH, a possibilidade de comparar visualizações de anos diferentes em uma mesma tela, entre outras pesquisas futuras.

Referências

- Brasil. (2011). *LEI Nº 12.527, DE 18 DE NOVEMBRO DE 2011*.
http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm
- Coelho, Flavio Codeco. (2020). *PySUS*. Disponível em: <https://pypi.org/project/PySUS/>
- Chui, K. K. H., Wenger, J. B., Cohen, S. A., & Naumova, E. N. (2011). Visual Analytics for Epidemiologists: Understanding the Interactions Between Age, Time, and Disease with Multi-Panel Graphs. *PLoS ONE*, 6(2), e14683.
<https://doi.org/10.1371/journal.pone.0014683>
- Saldanha, R. D. F., Bastos, R. R., & Barcellos, C. (2019). Microdatasus: a package for downloading and preprocessing microdata from Brazilian Health Informatics Department (DATASUS). *Cadernos de saude publica*, 35(9), e00032419-e00032419.
- Dresch, A., Lacerda, D. P., & Antunes, J. A. V. J. (2015). *Design science research: método de pesquisa para avanço da ciência e tecnologia*. Porto Alegre: Bookman.
- Few, S. (2006). *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly Media, Inc.
- Few, S., & Edge, P. (2007). *Data visualization: past, present, and future*. IBM Cognos Innovation Center.
- Joshi, A., Amadi, C., Katz, B., Kulkarni, S., e Nash, D. (2017). A Human-Centered Platform for HIV Infection Reduction in New York: Development and Usage Analysis of the Ending the Epidemic (ETE) Dashboard. *JMIR Public Health Surveill*, 3(4):e95.
- Martinez, R., Ordunez, P., Soliz, P. N., e Ballesteros, M. F. (2016). Data visualisation in surveillance for injury prevention and control: Conceptual bases and case studies. *Injury Prevention*, 22(Suppl 1):i27-i33.
- McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 56-61). <https://doi.org/10.25080/Majora-92bf1922-00a>
- Mendes, D., Lobato, F., & Jacob Jr, A. (2020). *Ferramenta de Pré-Processamento e*

- Visualização de dados do DATASUS*. 1–10. <https://doi.org/10.5753/wtrans.2019.6436>
- Ministério da Saúde. *DATASUS*. Disponível em: <http://datasus.saude.gov.br/>. Acessado em 28/10/2019
- Ministério da Saúde. (2020a). *Painel Coronavírus*. Disponível em <https://covid.saude.gov.br/> Acessado em 03/08/2020
- Ministério da Saúde. (2020b). *SUS*. Disponível em <https://www.saude.gov.br/sistema-unico-de-saude/1366-sus>. Acessado em 03/08/2020
- Petruzalek, D. (2016). READ. DBC: um pacote para importação de dados do datasus na linguagem R. In *Anais do XV Congresso Brasileiro de Informática em Saúde [Internet]* (pp. 27-30).
- Preim, B., & Lawonn, K. (2020, February). A survey of visual analytics for public health. In *Computer Graphics Forum* (Vol. 39, No. 1, pp. 543-580).
- Pires, F. A., Abrahão, M. T., Rebelo, M. S., Santos, R. S., Nobre, M. C., & Gutierrez, M. A. (2011). Ambiente para extração de informações de saúde a partir de bases de dados do SUS. *BIS. Boletim Do Instituto de Saúde (Impresso)*, 13(1), 39–45.
- Qin, X., Luo, Y., Tang, N., e Li, G. (2020). Making data visualization more efficient and effective: a survey. In *VLDB Journal*, volume 29, pages 93–117. Springer
- Sedig, K. e Ola, O. (2014). The Challenge of Big Data in Public Helth: An Opportunityfor Visual Analytics. *Online Journal of Public Health Informatics*, 5(3):223
- Silva, N. P. Da. (2009). *A utilização dos programas TABWIN e TABNET como ferramentas de apoio à disseminação das informações em saúde*. 98.
- Ventura, M. (2013). Lei de acesso à informação, privacidade e a pesquisa em saúde. *Cadernos de Saude Publica*, 29(4):636–638.
- University of Washington. (2013). *GBD Compare*. Institute for Health Metrics and Evaluation. Disponível em <http://viz.healthmetricsandevaluation.org/gbd-compare/>