

MÉTODOS PARA VISUALIZAÇÃO DE BIG DATA APLICADOS AO CONTEXTO DE REDUÇÃO DE MORTALIDADE NEONATAL

Vinicius E. Torrezin¹, Luciana Alves², Carlos E. Beluzo³, Tiago Carvalho⁴

¹ Graduando em Tecnologia de Análise e Desenvolvimento de Sistemas, Bolsista CNPq, IFSP, Câmpus Campinas, vinicius.torrezin@aluno.ifsp.edu.br.

² Docente do Departamento de Demografia do IFCH e pesquisadora do NEPO da UNICAMP, Pesquisadora Colaboradora, luciana@nepo.unicamp.br.

³ Professor no Instituto Federal de São Paulo, Câmpus Campinas, Pesquisador Colaborador, beluzo@ifsp.edu.br.

⁴ Professor no Instituto Federal de São Paulo, Orientador, Câmpus Campinas, tiagojc@ifsp.edu.br.

Área de conhecimento (Tabela CNPq): 1.03.03.04-9 Sistemas de Informação

Apresentado no

10º Congresso de Inovação, Ciência e Tecnologia do IFSP ou no 4º Congresso de Pós-Graduação do IFSP

27 e 28 de novembro de 2019- Sorocaba-SP, Brasil

RESUMO: A redução da mortalidade neonatal é de extrema importância. O período neonatal é o mais vulnerável para a sobrevivência da criança e é influenciado pelas condições de vida da população e saúde. No Brasil mesmo contendo diferentes fontes de dados como o IBGE e DataSUS, a maioria não tem formas intuitivas de visualizações desses dados. Deste modo é proposto neste trabalho a utilização de diferentes ferramentas de *visual analytics* para *big data* de modo a possibilitar um entendimento maior de dados públicos ligados à mortalidade infantil através do uso visualizações interativas. Ao longo do desenvolvimento do trabalho, foi utilizada a biblioteca D3.js para desenvolver as visualizações interativas, de modo que a informação seja de fácil entendimento para o público em geral, principalmente para gestores públicos. Conclui-se que as visualizações interativas permitem a criação de novos *insights* e extração novas informações sobre as bases de dados, de uma forma mais intuitiva do que observar dados tabulares. Além disso, permite também que agentes da área da saúde extraiam dos dados mais informações relevantes associadas aos fatores de risco neonatal.

PALAVRAS-CHAVE: mortalidade neonatal; *visual analytics*; *big data*.

Big Data Visualization Methods Applied to the Neonatal Mortality Reduction Context

ABSTRACT: *Reducing neonatal mortality is of utmost importance. The neonatal period is the most vulnerable to child survival and is influenced by the population's living conditions and health. In Brazil even containing different data sources such as IBGE and DataSUS, most do not have intuitive ways of visualizing this data. Thus, it is proposed in this work the use of different visual analytics tools for big data in order to enable a better understanding of public data related to child mortality through the use of interactive visualizations. Throughout the work, we use the D3.js library to develop interactive visualizations so that the information is easily understood by the general public, especially public managers. It is concluded that interactive visualizations allow the creation of new insights and extract new information about databases, in a more intuitive way than observing tabular data. In addition, it also allows health workers to extract from the data more relevant information associated with neonatal risk factors.*

KEYWORDS: *neonatal mortality; visual analytics; big data;*

INTRODUÇÃO

A redução da mortalidade neonatal é de extrema importância para o mundo todo. O período neonatal representa o momento mais vulnerável para a sobrevivência de uma criança e é fortemente influenciado pelas condições de vida da população e da atenção à saúde. Em particular, o Brasil conta com níveis preocupantes de mortalidade neonatal e infantil.

Apesar de existirem diferentes fontes de dados públicos que permitam a definição e exploração de indicadores relacionados a este tipo de problema, tais como o IBGE e o DataSUS, apenas contar com um grande número de bases de dados, sem ferramentas corretas para sua exploração, torna-se ineficaz. Bases de dados como as do DataSUS, são um bom exemplo de volumes massivos de dados (*big data*), e apesar de muitas vezes conterem informações extremamente relevantes, são subutilizadas devido à dificuldade de extração das informações, por conta da ausência de ferramentas que permitam uma exploração efetiva e eficaz da informação nela contida.

Desta forma, neste artigo, é proposto o desenvolvimento de formas de visualização simples, interativas e de fácil entendimento para os usuários através da aplicação do *framework* D3.js. Segundo BOSTOCK (2019) o D3 auxilia a dar “vida” aos dados, utilizando HTML, SVG e CSS, adicionando comportamentos dinâmicos para a interação do usuário e animações.

Espera-se como contribuições e produtos deste projeto, além de um estudo aprofundado de ferramentas de visualização para *big data*, a produção de ferramentas visuais que agreguem diferentes tipos de indicadores, os quais possuem grande utilidades para gestores públicos, entre outros.

MATERIAL E MÉTODOS

A coleta de dados de nascimentos e mortalidade é uma tarefa importante, sendo que no Brasil foi adotado o Sistema de Informações sobre Mortalidade (SIM) e Sistema de Informações sobre Nascidos Vivos (SINASC) para a obtenção de dados em relação a mortalidade e nascimentos, respectivamente. É avaliado que o SIM e SINASC possuem uma grande melhora desde sua criação, conseguindo uma grande cobertura e qualidade dos seus dados, mesmo que em alguns aspectos deixe a desejar. Desta forma mostra que esses dados são confiáveis (JORGE et al., 2007). Além do SIM e SINASC, foram utilizadas outras bases de dados que o DataSUS disponibiliza e inclusive base de dados do IBGE, estas bases nos fornecem indicadores que são utilizados para as visualizações, sendo elas o Produto Interno Bruto (PIB), a cobertura de vacinação, o Índice de Desenvolvimento Humano (IDH), a taxa de desemprego, a cobertura de esgoto, a proporção de óbitos neonatais por causa, a cobertura de consultas pré-natal, entre diversos outros indicadores.

No contexto de Big Data, as ferramentas de *Visual Analytics* são utilizadas de diferentes formas, e aplicadas em diferentes áreas. Abousalh-Neto e Kazgan (2012) utilizam a ferramenta *SAS® Visual Analytics Explorer* para realizar uma análise exploratória em uma base de dados contendo 158 milhões de entradas de *logs* em uma rede bancária fictícia, obtendo diferentes formas de visualização dos dados e extraíndo informações relevantes dos mesmos.

Os autores Barkwell et al. (2018) propuseram um conjunto de ferramentas de *visual analytics* para o contexto de extração de informações em dados musicais. Tais ferramentas são utilizadas para descobrir e explorar diferentes padrões de dados musicais, especialmente quando aplicados para extrair informações de dados líricos apresentados de forma textual. O Sistema desenvolvido pelos autores, nomeado de *Hue-Saturation-Value based visualiser* (HSVVis), fornece para os usuários uma representação circular, de forma interativa, dos padrões de dados, permitindo a exploração e percepção de padrões recentes.

Já no contexto de cuidados com a saúde, diferentes trabalhos discutem os ganhos que técnicas de visualizações de informação podem trazer. Os autores Caban e Gotz (2015), deixam claro em seu trabalho que os dados relacionados a cuidados com a saúde está crescendo em uma velocidade sem precedentes, e que diferentes desafios estão surgindo simultaneamente com a produção de dados:

- (1) Clínicas tentam combinar dados por diferentes tipos de ferramentas (clínicos, exames, etc.);
- (2) Gestores tentam tomar decisões baseados em dados;
- (3) Pesquisadores tentam entender as diferenças entre as populações, dentre outros.

Segundo os autores, de forma a facilitar o entendimento e interpretação de dados complexos, as técnicas de *visual analytics* combinam conceitos de exploração de dados, aprendizado de máquina,

cognição humana e interação humano-computador. Assim, como o volume de dados que podem ser utilizados no contexto de saúde cresce em taxas sem precedentes, técnicas de *visual analytics* podem ser utilizadas nesse contexto para se extrair o melhor que os dados podem oferecer.

A metodologia proposta para este trabalho consiste em implementar possíveis formas de visualização da literatura de *Visual Analytics* para permitir uma visualização mais inteligível para os dados relacionados à mortalidade, fazendo o acesso à informação simples e acessível a usuários comuns: A proposta é que os usuários possam:

- Selecionar e visualizar dados sugeridos pela aplicação como relevantes, como por exemplo: dados familiares, tempo de gravidez, informações biológicas da mãe, etc.,
- Definir o tipo de correlação desejada, sendo neste projeto exclusivamente o risco de morte neonatal, bem como os fatores de causa e influência que se deseja avaliar;

RESULTADOS E DISCUSSÃO

Tem-se a ferramenta D3.js que é referida como “*Data-Driven Documents*”, sendo uma biblioteca do JavaScript, combinando HTML, SVG e CSS, resultando em uma ferramenta de *Data Visualization* (CHEN; ZHOU, 2016). A visualização D3 é provida por regras de mapeamento customizáveis pelo usuário, como a cor, o tamanho, entre outros (FAN BAO; JIA CHEN, 2014). O D3 permite atualizar visualizações, criando novas visualizações primitivas para novos dados e assim removendo os que não são mais utilizados (LI et al., 2015). Adotando o mapa do Brasil no formato JSON, convertido a partir de um arquivo em *shapefile*. Desta forma o objetivo é mostrar de forma visual e interativa o mapa do Brasil dividido em três níveis de detalhes, sendo eles: Estados, Microrregiões e Municípios. Foi criado um mapa de calor, como visto na Fig. 1, usando os dados da taxa de mortalidade precoce, ocorrida entre 0 a 6 dias, e a tardia, ocorrida entre 7 a 27 dias, dividido em um período de tempo entre 2000 e 2016, sendo esta a principal visualização na plataforma.

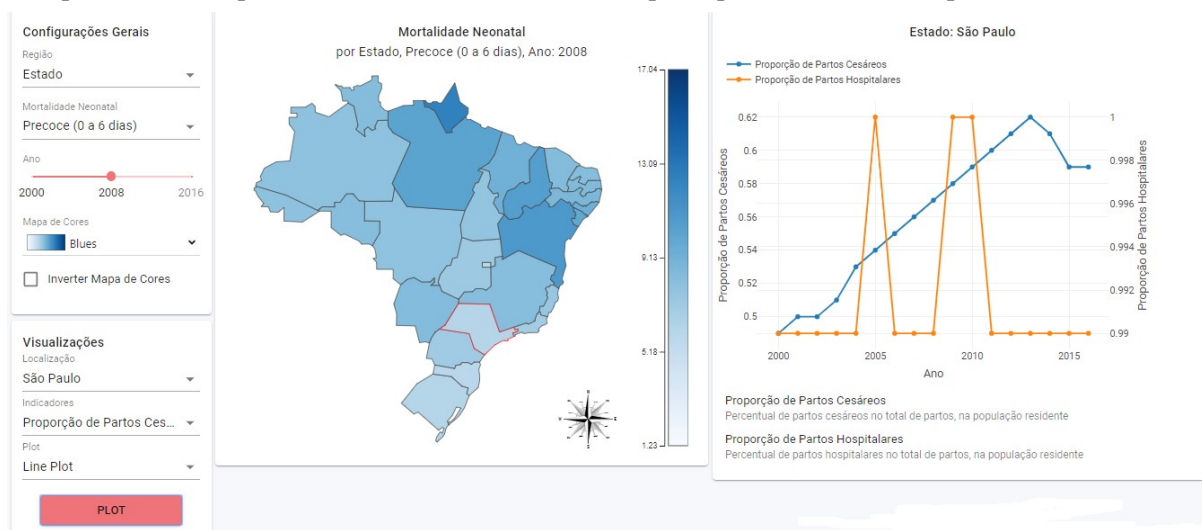


FIGURA 1. Plataforma com a visualização do mapa de calor em taxa de mortalidade neonatal precoce (0 a 6 dias) e um gráfico de linha comparando os indicadores de partos cesários e hospitalares.

Na abordagem para prover uma visualização mais simples, foi adotado gráficos mais simples e interativos, que ao clicar no local pelo mapa de calor, carrega os dados para aquele local. Como visto na Fig. 2 foram utilizados os indicadores de proporção de partos cesários e hospitalares, está visualização permiti observar cada um dos indicadores e como se comportam durante os anos, além de poder compará-los.

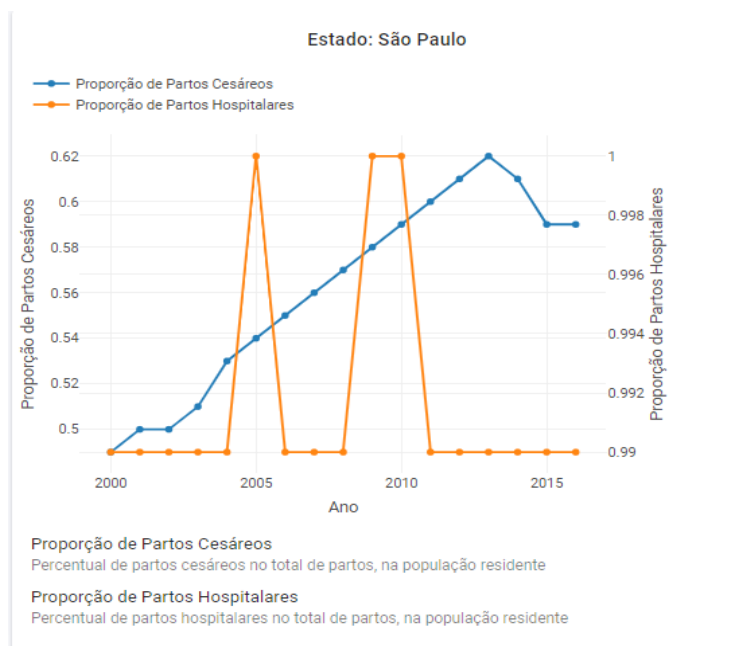


FIGURA 2. Gráfico de linha comparando os indicadores de partos cesários e hospitalares no estado de São Paulo.

CONCLUSÕES

Este trabalho é parte integrante do projeto intitulado “Plataforma de Apoio à Decisão para Políticas Públicas de Saúde Gestacional Baseada em Técnicas de Visualização de Informações e Aprendizado de Máquina” aprovado na chamada *Grand Challenge Explorations* da Fundação Bill & Melinda Gates, Ministério da Saúde e Fapesp.

Ao longo de seu desenvolvimento, construímos duas formas de visualização interativa de dados: a primeira baseada em um mapa de calor interativo que exibe as taxas de mortalidade neonatal ao longo dos anos para regiões do Brasil (estados, microrregiões e municípios); a segunda forma de visualização foca na exibição de dados (como o PIB) que são correlacionados à mortalidade infantil em uma visualização de gráfico de linha simples, que evolui temporalmente (exibe a evolução ao longo dos anos).

As visualizações desenvolvidas ao longo do trabalho têm se mostrado úteis para o entendimento do comportamento dos dados e sua relação com o fenômeno da mortalidade infantil. Desta forma, pretende-se como trabalhos futuros desenvolver novas visualizações para outras variáveis ligadas ao fenômeno de mortalidade infantil.

AGRADECIMENTOS

Agradecemos o apoio do grupo de pesquisa PICAp-IFSP Campinas; ao apoio financeiro da Fundação Bill & Melinda Gates (OPP1201970) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (443774/2018-8), implementado via projeto institucional IFSP (SUAP: 23305.012088.2018-11); ao Programa Institucional de Bolsas de Iniciação Científica e Tecnológica do Instituto Federal de São Paulo - PIBIFSP Campinas (SUAP 23305.012164.2018-98); e à NVIDIA pela doação de uma GPU XP Titan, a ser utilizada nesta pesquisa.

REFERÊNCIAS

ABOUSALH-NETO, N. A.; KAZGAN, S. Big data exploration through visual analytics. 2012 IEEE Conference on Visual Analytics Science and Technology (VAST). Anais... In: 2012 IEEE CONFERENCE ON VISUAL ANALYTICS SCIENCE AND TECHNOLOGY (VAST). Out. 2012.

BARKWELL, K. E. et al. Big Data Visualisation and Visual Analytics for Music Data Mining. 2018 22nd International Conference Information Visualisation (IV). Anais... In: 2018 22ND INTERNATIONAL CONFERENCE INFORMATION VISUALISATION (IV). Jul. 2018.

BOSTOCK, M. D3.js - Data-Driven Documents. Disponível em: <<https://d3js.org/>>. Acesso em: 25 maio. 2019.

CABAN, J. J.; GOTZ, D. Visual analytics in healthcare--opportunities and research challenges. Journal of the American Medical Informatics Association: JAMIA, v. 22, n. 2, p. 260–262, mar. 2015.

CHEN, L.; ZHOU, H. Research and application of dynamic and interactive data visualization based on D3. 2016 International Conference on Audio, Language and Image Processing (ICALIP). Anais... In: 2016 INTERNATIONAL CONFERENCE ON AUDIO, LANGUAGE AND IMAGE PROCESSING (ICALIP). Jul. 2016.

Departamento de Informática do SUS (DataSUS). Disponível em: <<http://datasus.saude.gov.br/>>. Acesso em: 2 jun. 2019.

FAN BAO; JIA CHEN. Visual framework for big data in d3.js. 2014 IEEE Workshop on Electronics, Computer and Applications. Anais... In: 2014 IEEE WORKSHOP ON ELECTRONICS, COMPUTER AND APPLICATIONS. Maio 2014.

Instituto Brasileiro de Geografia e Estatística (IBGE). Disponível em: <<https://www.ibge.gov.br/>>. Acesso em: 2 jun. 2019.

JORGE, M. H. P. DE M.; LAURENTI, R.; GOTLIEB, S. L. D. Análise da qualidade das estatísticas vitais brasileiras: a experiência de implantação do SIM e do SINASC. Ciência & Saúde Coletiva, v. 12, p. 643–654, jun. 2007.

LI, X. et al. Advanced aggregate computation for large data visualization. 2015 IEEE 5th Symposium on Large Data Analysis and Visualization (LDAV). Anais... In: 2015 IEEE 5TH SYMPOSIUM ON LARGE DATA ANALYSIS AND VISUALIZATION (LDAV). Out. 2015.