

Análise Visual Interativa: Utilizando Dados Abertos de Acidentes de Trânsito

Camila Moser de Oliveira, Lucas Antunes de Freitas, Isabel Harb Manssour

PUCRS, Escola Politécnica

Porto Alegre, RS, Brasil

{camila.moser,lucas.freitas.003}@acad.pucrs.br, isabel.manssour@pucrs.br

Resumo—Com a popularização da Internet e o avanço da tecnologia da informação, tem aumentado o volume de informações geradas diariamente. Neste contexto, sites governamentais têm disponibilizado dados de diversas áreas, como saúde, educação e mobilidade, para uso e redistribuição por qualquer pessoa. Porém, há o grande desafio de apresentar estas informações de forma clara e de fácil compreensão, para que possam ser analisadas. O objetivo deste trabalho é apresentar uma aplicação *web* que permite fazer uma análise visual de dados referentes a acidentes de trânsito da cidade de Porto Alegre/RS. São apresentados gráficos interativos com o intuito de ajudar a encontrar padrões que gerem conhecimento e auxiliem o usuário desta aplicação em possíveis e futuras tomadas de decisões. Utilizamos uma combinação de *Heat Map* animado, que permite fazer uma análise temporal, e três gráficos associados como técnicas de visualização. Além disso, disponibilizamos uma série de filtros que permitem fazer diferentes combinações de dados para serem analisados. Nossa principal contribuição é a possibilidade de fazer uma análise visual de dados abertos do governo sobre acidentes de trânsito.

Palavras-Chave: Dados Abertos, Visualização de Informações, Acidentes de Trânsito, *Heat Map*.

Abstract—With the popularization of the Internet and the advancement of information technology, increase the volume of information daily generated. In this context, government sites have provided data from several areas, such as health, education and mobility, for use and redistribution by anyone. However, there is the great challenge of presenting this information in a clear and easy-to-understand way, so that they can be analyzed. The main goal of this work is to present a web application that allows making a visual analysis of data related to traffic accidents in the city of Porto Alegre/RS. Interactive graphics are presented to help find patterns that generate knowledge and assist the user of this application in possible and future decision making. We use a combination of animated Heat Map, allowing a temporal analysis, and three associated graphs as visualization techniques. In addition, we provide a series of filters that allow different combinations of data to be analyzed. Our main contribution is the possibility of making a visual analysis of the government's open data on traffic accidents.

Keywords: Open Data, Information Visualization, Traffic Accidents, Heat Map.

I. INTRODUÇÃO

O avanço da tecnologia da informação, dos equipamentos de aquisição de imagens, sinais e dados em geral, e o crescente desenvolvimento da internet, tem disponibilizado um volume cada vez maior de informações. A regra nos últimos anos é

tentar tornar todo e qualquer dado possível em digital [1]. Assim, surge o desafio de apresentar as informações com maior relevância e representá-las de forma mais compreensível [2].

Uma forma de contornar as dificuldades de selecionar e analisar informações relevantes entre os grandes volumes de dados disponíveis atualmente é a utilização das técnicas de visualização da informação. Através destas técnicas é possível obter uma representação visual que, se por um lado abstrai detalhes de um conjunto de informações, por outro, propicia uma organização desse conjunto seguindo algum critério [3]. O uso de técnicas de visualização e interação possibilita realizar uma análise visual, permitindo explorar as informações, extrair conhecimentos e apresentar resultados [3].

Uma grande fonte de informações disponível hoje em dia são os *sites* que os governos federais, estaduais e municipais oferecem com dados de diversas áreas como saúde, educação, mobilidade, turismo, entre outros. Através desta iniciativa, todo dado disponibilizado pode ser livremente usado, reusado e redistribuído por qualquer pessoa [4]. Por exemplo, dados de acidente de trânsito são disponibilizados nestes sites e trazem informações que podem ser usadas para análise e estatística, pois a violência no trânsito gera altos custos financeiros e um forte abalo emocional para a sociedade [5]. Segundo pesquisas feitas pelo CPES (Centro de pesquisa e Economia do Seguro) [6], há uma estimativa que, em 2016, os custos decorrentes de acidentes de trânsito no país geraram um prejuízo de R\$ 146,8 bilhões, e neste valor não está incluído o custo com gastos hospitalares e reabilitações.

Dados como estes de acidentes de trânsito são difíceis de serem explorados, por serem de natureza multivariada [7]. Neste caso, o uso de técnicas de visualização e interação possibilita a realização de uma análise visual, a fim de obter conclusões para tomar melhores decisões [8].

Por exemplo, uma análise visual de dados de acidentes de trânsito permitiria mapear o número de ocorrências, e categorizá-las por período, tipo de veículo, situação climática, grau de acidente, entre outros. Estas informações podem gerar conhecimento de grande valor para a sociedade, possibilitando auxiliar na identificação de medidas a serem tomadas. Diferentes tipos de gráficos podem ser gerados, como, por exemplo, um *Heat Map* aplicado a um mapa geográfico para mostrar a região que apresenta um elevado número de acidentes envolvendo motos em dias de chuva. Este tipo de representação visual auxiliaria órgãos públicos de mobilidade urbana

a tomarem conhecimento deste fato, para, então, melhorarem a malha viária urbana onde ocorre este tipo de acidente.

Considerando este contexto, este trabalho tem como objetivo apresentar uma aplicação que fornece diferentes formas de visualização dos dados abertos sobre acidentes de trânsito da cidade de Porto Alegre/RS. Estes dados são disponibilizados através do *site* DataPOA¹. Assim, espera-se auxiliar, por exemplo, a EPTC (Empresa Pública de Transporte e Circulação) e o DETRAN/RS (Departamento Estadual de Trânsito do Rio Grande do Sul) em futuras tomadas de decisões, sobre pontos e horários em que o trânsito da cidade precisa de melhorias ou de maior fiscalização para diminuir um pouco o prejuízo que os acidentes de trânsito causam para os cofres públicos.

O restante deste documento está organizado da seguinte maneira: a Seção II apresenta alguns trabalhos relacionados que abordam a visualização de informações de acidentes de trânsito e um trabalho que aborda a visualização de informações de crimes. A aplicação desenvolvida, incluindo suas técnicas de visualização e interação, está descrita na Seção III, e os resultados obtidos são apresentados na Seção IV. A última seção apresenta as conclusões e trabalhos futuros.

II. TRABALHOS RELACIONADOS

Encontramos um grande número de trabalhos relacionados ao tema que utilizam dados abertos disponibilizados por órgãos governamentais, dados de mídias sociais, dados obtidos pelo próprio autor ou dados sobre o tráfego nas cidades [9], [10]. Também observamos que o uso de visualização geoespacial está presente em grande parte dos trabalhos relacionados, apresentando técnicas como o *Heat Map* e mapa com ícones para auxiliar na análise dos dados. A seguir apresentamos os trabalhos mais semelhantes à nossa abordagem (ver Seção III).

Um trabalho que também utiliza os dados do DataPOA², faz a previsão de risco de ferimentos por acidentes de trânsito em Porto Alegre [11]. Para representação dos dados geoespaciais foi utilizado o *Framework Django GIS*³ Brasil, um projeto *Open Source* que tem como colaborador o mesmo autor do artigo [11]. Neste trabalho, as áreas onde ocorreram mais acidentes de trânsito são apresentadas através de um *Heat Map*, mas não são apresentadas outras formas de visualização para representar os dados.

O mapa de crime da cidade de Halifax⁴ no Canadá, é uma aplicação *web* de dados abertos sobre crimes. Ele utiliza um *Heat Map* animado para a visualização destes dados, com opção de *pause* e *play*, além de possuir filtros para selecionar um ano (2013 a 2018) e o tipo de crime (roubo de veículos, furto de veículos, invasão e roubo, assalto, furto). Neste caso, é atribuída uma cor diferente para cada tipo de crime, pois quando os *pins* são mostrados no *Heat Map* é possível identificá-los só pela cor. Há também a opção de retirar os *pins* quando estiver na função *pause*. Entretanto,

não são apresentados outros gráficos para representar os tipos de crimes.

Outro trabalho que permite uma análise visual para dados de acidentes de trânsito é o ICE (*Incident Cluster Explorer*) [7]. Esta aplicação apresenta funcionalidades como: mapas interativos, histogramas, exploração de dados multidimensionais e coordenadas paralelas. Possui um grande conjunto de filtros e o usuário pode escolher qual tipo de dados será filtrado. Uma forma de visualização interativa disponibilizada é um gráfico de histograma associado com um mapa. Cada barra do histograma pode ser representada por um tipo de acidente, tal como: colisão fatal, colisão com lesão corporal, colisão com danos materiais, etc. O usuário, ao clicar em qualquer uma das barras do histograma, obtém a frequência de cada tipo de acidente. Além disso, é possível visualizar no mapa os pontos onde ocorreram estes acidentes. O usuário também tem a opção de selecionar um subconjunto dos acidentes a partir de um período de tempo informado. Entretanto, neste trabalho não é apresentado um *Heat Map* e não são utilizados algoritmos de mineração sobre os dados coletados.

O analisador de acidentes de trânsito, *software* desenvolvido na linguagem de programação C# utilizando o *Google Maps* [12], também tem como objetivo analisar acidentes de trânsito para auxiliar na identificação de problemas e na descoberta de soluções. A fonte de dados utilizada para este trabalho foi obtida através do portal de dados abertos da cidade de Nova Iorque (*NYC OpenData*⁵) referente ao ano de 2015. As técnicas de visualização disponibilizadas permitem identificar o número total de mortes em acidentes de trânsito em cada rua. Para visualização dos dados, utiliza um mapa da cidade com *pins* que representam os locais onde ocorreram mortes por acidentes de trânsito, e um gráfico de barras para mostrar a quantidade de acidentes dependendo do tipo de informação que se deseja obter, como, por exemplo: a quantidade de acidentes nos bairros; e um padrão dos tipos de carros que se envolvem em acidentes. Entretanto neste trabalho também não é disponibilizado um *Heat Map*.

III. PROJETO DA ANÁLISE VISUAL INTERATIVA

Para alcançar os objetivos apresentados na Seção I, o desenvolvimento deste trabalho foi dividido em etapas seguindo os princípios do KDD (*Knowledge-Discovery in Databases*), ou seja, é composto de diferentes atividades que vão desde a procura por dados abertos do governo relativos a acidentes de trânsito do município de Porto Alegre/RS, até a visualização gráfica das informações extraídas destes dados. Estas etapas são descritas a seguir.

Inicialmente, coletamos e pré-processamos os dados. Nesta etapa, para ajudar na limpeza e padronização dos dados, já que os mesmos apresentavam diferentes formatações dependendo do ano, utilizamos o *OpenRefine*⁶.

Com o término da primeira etapa, começamos o processo de desenvolvimento da arquitetura da aplicação. Primeiro foi

¹<http://datapoa.com.br/>

²<http://datapoa.com.br/dataset/acidentes-de-transito>

³<https://django-gis-brasil.readthedocs.io/en/v0.3/#django-gis-bra>

⁴<http://www.crimeheatmap.ca/>

⁵<https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95/>

⁶<http://openrefine.org/>

criada a **Camada de Persistência** utilizando a tecnologia *Elasticsearch*⁷. Para implementar a **Camada do Servidor** utilizamos a tecnologia *Node.js*⁸, na qual foram criadas as *queries* de consulta aos dados armazenados e a lógica de tratamento dos dados, transformando-os em formato JSON (*JavaScript Object Notation*) e viabilizando, assim, a manipulação dos mesmos pela camada de apresentação.

A **Camada de Apresentação** foi desenvolvida utilizando o *framework AngularJS*⁹. Nela foram implementadas as funcionalidades para apresentar os dados tratados pelo servidor na forma de gráficos e a possibilidade de interagir com eles de forma dinâmica a partir de filtros.

Com a arquitetura criada, começamos a desenvolver as representações visuais da aplicação. Para o *Heat Map* animado, que se modifica com o passar do tempo conforme o número de acidentes encontrados a partir da aplicação dos filtros disponíveis, foi usada a API (*Application Programming Interface*) do *Google Maps* e consultada a solução do *Halifax Crime Map* disponível no *GitHub*¹⁰.

A biblioteca *Angular-nvd3*¹¹ foi usada para implementar os demais gráficos que exibem outros dados consultados a partir da API de consulta no servidor. Neste caso, foi necessário formatar o resultado da consulta desses dados para poder ser interpretado pela biblioteca e apresentado através de gráficos de barra e de setores.

A interface da aplicação está apresentada na Figura 1. Na parte superior desta interface está o mapa que possui a visualização do *Heat Map* animado. Para controlar a animação há um botão para pausá-la (*pause*) e outro para ativá-la (*play*). Logo abaixo estão os três gráficos que mudam de acordo com os atributos selecionados. O primeiro gráfico é um *Historical Bar Chart* que mostra a quantidade de acidentes em faixas de horários (0h a 23h). O segundo é um *Pie Chart* que mostra as regiões da cidade de Porto Alegre (Norte, Sul, Leste e Centro) nas quais mais ocorreram acidentes. O terceiro gráfico é um *Bar Chart* horizontal que mostra os tipos de acidentes e o total de acidentes de cada tipo. Abaixo dos gráficos estão os *checkboxes* que possibilitam selecionar o(s) ano(s) que se deseja analisar. Por ser uma quantidade muito grande de dados, além dos *checkboxes*, é possível selecionar os seguintes filtros em um menu lateral que pode ser exibido: Tipo de Veículo, Faixa de Horário e Condição do Tempo.

IV. RESULTADOS OBTIDOS

Para realizar a análise dos resultado apresentados nos gráficos, optamos por fazer um estudo envolvendo acidentes com bicicletas entre os anos 2009 a 2012, e entre os anos de 2013 a 2016. Essa escolha foi devido ao aumento do número de ciclovias na cidade de Porto Alegre durante estes anos [13]. Assim, seria possível analisar se com o aumento de ciclovias houve uma redução no número de acidentes.

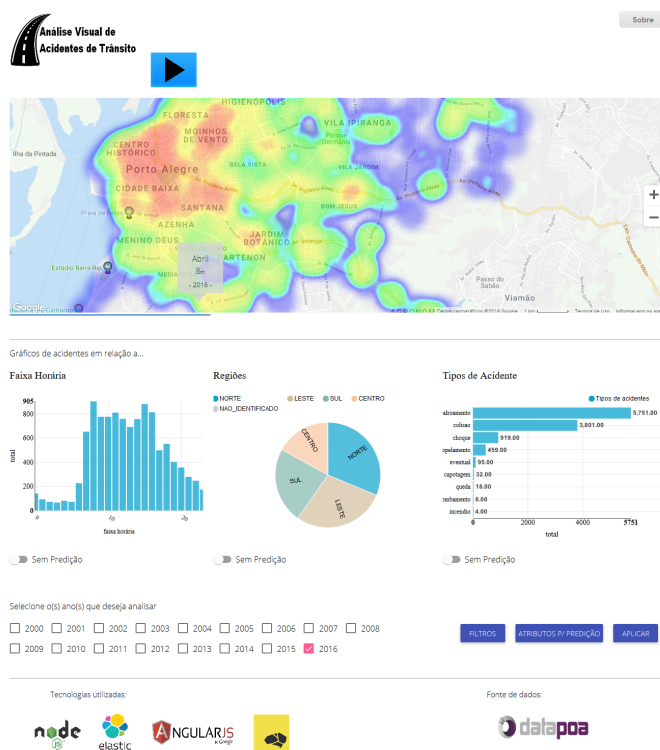


Figura 1. Tela principal da aplicação

Definimos que a avaliação se dará da seguinte forma: primeiro analisamos os anos anteriores ao aumento das ciclovias, que consiste em visualizar os dados dos anos 2009, 2010, 2011 e 2012; a segunda análise é relativa aos anos posteriores, quando foram visualizados os dados dos anos 2013, 2014, 2015 e 2016.

A primeira análise, apresentada na Figura 2, o gráfico de **Faixa Horário**, revelou que as faixas de horário com maior ocorrência de acidentes envolvendo bicicletas eram nos horários das 17 horas (82 acidentes), 18 horas (101 acidentes) e 19 horas (100 acidentes). O horário de maior ocorrência foi às 18 horas, com uma diferença ínfima em relação ao horário das 19 horas.

No gráfico **Regiões**, a região sul apresentava o maior número de ocorrências de acidentes, com o total de 379 acidentes. A diferença é grande em relação a segunda região com mais acidentes, a região norte, com o total de 299 acidentes. Já a região do centro da cidade obteve o menor número de acidentes, com um total de 101 acidentes.

Já no gráfico **Tipos de Acidente**, o abaloamento apareceu em primeiro lugar, com o total de 776 ocorrências de acidentes. Uma diferença muito grande em relação ao segundo tipo com mais ocorrências, a colisão, com 149 acidentes. Os demais tipos de acidentes obtiveram valores muito baixos em relação aos anteriores.

Na segunda análise em comparação com a primeira, o número de acidentes envolvendo bicicletas reduziu consideravelmente. Como é possível verificar no gráfico **Faixa de Horário**

⁷<https://www.elastic.co/products/elasticsearch>

⁸<https://nodejs.org/en/>

⁹<https://angular.io/>

¹⁰<https://github.com/davidhampgonsalves/crime-heatmaps>

¹¹<http://krispo.github.io/angular-nvd3/>

Gráficos de acidentes em relação a...

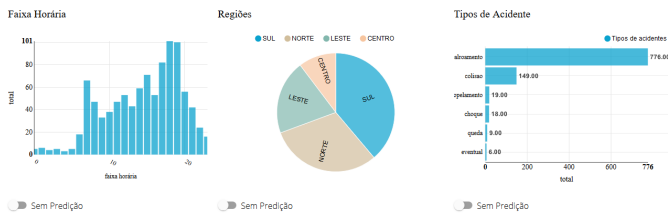


Figura 2. Gráficos com os resultados dos anos 2009 a 2012 referentes a acidentes com bicicletas

da Figura 3, o horário com maior ocorrência continuou sendo às 18 horas, mas com um total de 76 acidentes. Em relação a análise anterior, o número de acidentes reduziu em torno de 25% e as outras faixas também tiveram uma redução significativa.

O gráfico **Regiões** mostra que a região sul continua sendo a região com mais acidentes envolvendo bicicletas, com um total de 255 acidentes. Comparado com a análise dos anos 2009 a 2012, a quantidade de acidentes teve uma redução ao redor de 40% nos acidentes. Porém, a região do centro da cidade com 123 acidentes, teve um aumento de 21% nos acidentes. Este aumento pode estar relacionado com a forma como são registrados os dados na hora em que ocorre o acidente, não existe uma verificação do registro.

No gráfico **Tipos de Acidentes**, o tipo abaloamento, com um total de 549 acidentes, teve uma redução média de 30% em relação a análise anterior. Mas, outros tipos de acidentes como atropelamento, choque e queda obtiveram um aumento em relação a análise anterior.

Gráficos de acidentes em relação a...

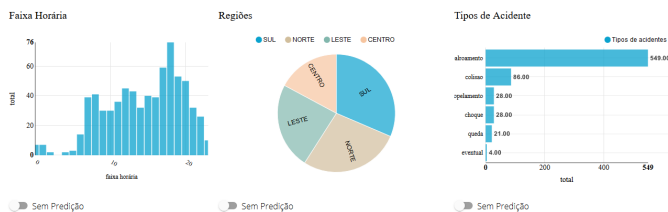


Figura 3. Gráficos com os resultados dos anos 2013 a 2016 referentes a acidentes com bicicletas

V. CONCLUSÕES E TRABALHOS FUTUROS

Através do estudo realizado neste trabalho foi possível verificar a importância da visualização de informações associada a técnicas de interação para grandes volumes de dados, pois facilitam a percepção visual do usuário, permitindo-o analisar e compreender as informações para auxiliar na tomada de decisões. Sendo assim, nossa contribuição foi o desenvolvimento de uma aplicação *web* que permite fazer uma análise visual interativa dos dados de acidentes de trânsito.

Esperamos que a aplicação desenvolvida ajude os usuários a conseguirem descobrir padrões nos dados, e que as técnicas de

visualização implementadas tragam conhecimento para auxiliar em tomadas de decisão dos órgãos públicos responsáveis pelo trânsito. O objetivo neste caso é identificar os locais, horários e até mesmo condições que levam ao aumento de ocorrência de acidentes, para conseguir auxiliar na redução de futuros acidentes de trânsito na cidade de Porto Alegre. Para isto, entraremos em contato com os órgãos responsáveis para apresentar a aplicação desenvolvida.

Como trabalhos futuros existe a possibilidade de incluir algoritmos de mineração, pois a aplicação também foi projetada com esta finalidade. Portanto, pretendemos fazer um estudo detalhado a respeito dos algoritmos de mineração, com o objetivo de fazer uma previsão, tomando como base o histórico de acidentes. Assim, através dos botões de seleção “Sem Predição” que podem ser vistos embaixo dos gráficos na Figura 1, será possível visualizar os resultados deste algoritmo e testar a aplicação com outros conjunto de dados diferentes. Além disso, faremos testes com usuários para avaliar se a aplicação desenvolvida é de fácil utilização, e identificar as melhorias que devem ser feitas em relação à sua usabilidade.

REFERÊNCIAS

- [1] “Estudo da emc prevê que volume de dados virtuais armazenados será seis vezes maior em 2020,” 2014. [Online]. Available: <https://goo.gl/B27Njg>
- [2] C. M. D. S. Freitas, O. M. Chubachi, P. R. G. Luzzardi, and R. A. Cava, “Introdução à visualização de informações,” *Revista de informática teórica e aplicada. Porto Alegre. Vol. 8, n. 2 (out. 2001)*, p. 143-158, 2001.
- [3] M. O. Ward, G. Grinstein, and D. Keim, *Interactive Data Visualization: Foundations, Techniques, and Applications, Second Edition*. New York, NY, USA: CRC Press Taylor Francis Group, September 2015.
- [4] B. C. Atkins, “Adaptive photo collection page layout,” in *International Conference on Image Processing, 2004. (ICIP'04)*, vol. 5. IEEE, 2004, pp. 2897–2900.
- [5] “Estimativa dos custos dos acidentes de trânsito no brasil com base na atualização simplificada das pesquisas anteriores do ipea,” 2016. [Online]. Available: http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=27755
- [6] “Acidentes no trânsito custam r\$ 146 bilhões ao país, aponta estudo,” 2017. [Online]. Available: <http://epocanegocios.globo.com/Brasil/noticia/2017/05/acidentes-no-transito-custam-r-146-bilhoes-ao-pais-aponta-estudo.html>
- [7] M. L. Pack, K. Wongsuphasawat, M. VanDaniker, and D. Filippova, “Ice—visual analytics for transportation incident datasets,” in *IEEE International Conference on Information Reuse & Integration, 2009. (IRI'09)*. IEEE, 2009, pp. 200–205.
- [8] D. A. Keim, F. Mansmann, J. Schneidewind, and H. Ziegler, “Challenges in visual data analysis,” in *Tenth International Conference on Information Visualization, 2006. (IV 2006)*. IEEE, 2006, pp. 9–16.
- [9] H. Guo, Z. Wang, B. Yu, H. Zhao, and X. Yuan, “Tripvista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection,” in *IEEE Pacific Visualization Symposium, 2011. (PacificVis)*. IEEE, 2011, pp. 163–170.
- [10] J. Pu, S. Liu, Y. Ding, H. Qu, and L. Ni, “T-watcher: A new visual analytic system for effective traffic surveillance,” in *IEEE 14th International Conference on Mobile Data Management, 2013. (MDM)*, vol. 1. IEEE, 2013, pp. 127–136.
- [11] C. S. Perone, “Injury risk prediction for traffic accidents in porto alegre/rs, brazil,” *CoRR*, vol. abs/1502.00245, 2015. [Online]. Available: <http://arxiv.org/abs/1502.00245>
- [12] E. Abdullah and A. Emam, “Traffic accidents analyzer using big data,” in *International Conference on Computational Science and Computational Intelligence, 2015. (CSCI)*. IEEE, 2015, pp. 392–397.
- [13] “Rede cicloviária de porto alegre aumentou 52% em 2013,” 2014. [Online]. Available: http://observapoa.com.br/default.php?reg=346&p_secao=17