



Programa Interdisciplinar de Pós-Graduação em

Computação Aplicada

Mestrado Acadêmico

Jonathan Machado

Um Método para Análise e Visualização de Dados
Georreferenciados Relacionados ao Trânsito de Veículos

São Leopoldo, 2017

UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA INTERDISCIPLINAR DE PÓS-GRADUAÇÃO
EM COMPUTAÇÃO APLICADA
NÍVEL MESTRADO

JONATHAN MACHADO

**UM MÉTODO PARA ANÁLISE E VISUALIZAÇÃO DE DADOS
GEORREFERENCIADOS RELACIONADOS AO TRÂNSITO DE
VEÍCULOS**

São Leopoldo
2017

Jonathan Machado

**UM MÉTODO PARA ANÁLISE E VISUALIZAÇÃO DE DADOS
GEORREFERENCIADOS RELACIONADOS AO TRÂNSITO DE
VEÍCULOS**

Dissertação apresentada como requisito parcial
para a obtenção do título de Mestre, pelo
Programa Interdisciplinar de Pós-Graduação
em Computação Aplicada da Universidade do
Vale do Rio dos Sinos – UNISINOS

Orientador: Prof. Dra. Marta Becker Villamil

Co-orientador: Prof. Dr. Luiz Paulo Luna de Oliveira

São Leopoldo

2017

M149m Machado, Jonathan.

Um método para análise e visualização de dados georreferenciados relacionados ao trânsito de veículos / Jonathan Machado. – 2017.

71 f. : il. ; 30 cm.

Dissertação (mestrado) – Universidade do Vale do Rio dos Sinos, Programa Interdisciplinar de Pós-Graduação em Computação Aplicada, 2017.

“Orientador: Prof. Dra. Marta Becker Villamil ; co-orientador: Prof. Dr. Luiz Paulo Luna de Oliveira.”

Dados Internacionais de Catalogação na Publicação (CIP)

(Bibliotecário: Flávio Nunes – CRB 10/1298)

Jonathan Machado

Um método para análise e visualização de dados georreferenciados
relacionados ao trânsito de veículos

Dissertação apresentada à Universidade do
Vale do Rio dos Sinos – Unisinos, como
requisito parcial para obtenção do título de
Mestre em Computação Aplicada.

Aprovado em 30 de março de 2017

BANCA EXAMINADORA

Marta Becker Villamil - UNISINOS

Nome do Componente da Banca Examinadora – Instituição a que pertence

Isabel Cristina Siqueira da Silva - UNIRITTER

Nome do Componente da Banca Examinadora – Instituição a que pertence

José Vicente Canto dos Santos - UNISINOS

Nome do Componente da Banca Examinadora – Instituição a que pertence

Prof. Dr. Marta Becker Villamil (Orientador)

Visto e permitida a impressão

São Leopoldo,

Prof. Dr. Sandro José Rigo

Coordenador PPG em Computação Aplicada

AGRADECIMENTOS

Agradeço a todos que, de alguma forma, contribuíram para a realização deste trabalho.

À minha orientadora, Professora Doutora Marta Becker Villamil, e ao Professor Doutor Luiz Paulo Luna de Oliveira, pela orientação, conhecimento, ideias, dedicação e apoio que tiveram durante o desenvolvimento deste trabalho.

A todos os professores e colegas do PIPCA pelo compartilhamento de conhecimentos, ideias e auxílio na realização dos trabalhos durante todo o mestrado.

A todos os órgãos, ferramentas e projetos que divulgam dados abertos, e que, consequentemente, contribuem com as atividades acadêmicas.

Por fim, agradeço a todos os familiares e amigos, em especial à minha esposa, Larissa, pelo apoio e compreensão durante o período dedicado a este estudo.

RESUMO

Os acidentes de trânsito de veículos são uma das maiores causas de mortes na população jovem mundial, e existe uma tendência ao crescimento no número de casos dos mesmos nos próximos anos. A ocorrência dos acidentes é influenciada por diversos fatores, tais como condições das vias, condições climáticas, fiscalização de leis por órgãos governamentais, dentre outros. Seria interessante conhecer de maneira mais detalhada quais destes fatores detém maior influência. Na internet, existe uma quantidade imensa de dados gerados pelos mais diversos órgãos e empresas, porém grande parte desta informação não é analisada por ninguém, seja por falta de acesso, ou porque os dados não estão estruturados de uma maneira que permita seu entendimento. A disponibilização de dados vem aumentando, seja por conta de políticas de dados abertos implantadas pelo governo ou através de ferramentas colaborativas da web, que possibilitam o registro de informações por parte da população, e que posteriormente disponibilizam seus dados. Este trabalho propõe um método de agrupamento de dados georreferenciados oriundos de diversas fontes, para realização de uma análise estatística utilizando a técnica de Análise de Componentes Principais, que poderá identificar de forma georreferenciada quais características influenciam mais na ocorrência de acidentes de trânsito de veículos. Após a análise, é explorada uma nova metodologia de visualização dos resultados, plotados sobre mapas, que podem servir de auxílio para órgãos do governo e tomadores de decisão que realizam ações para diminuir os acidentes de trânsito.

Palavras-Chave: Dados georreferenciados. Diagramas de Voronoi. Análise de Componentes Principais. Acidentes de Trânsito. Dados abertos e colaborativos.

ABSTRACT

Traffic accidents of vehicles are one of the biggest causes of deaths in the world's young population, and there is a tendency to increase this number in the next years. The occurrence of accidents is influenced by several factors, such as road conditions, climatic conditions, law enforcement by government agencies, among others. It would be interesting to know in more detail which of these factors has the greatest influence. On the internet, there is an immense amount of data generated by diverse agencies and companies, but much of this information is not analyzed, either because of lack of access, or because the data is not structured in a way that allows its understanding. The availability of data is increasing, either through open data policies implemented by the government, or through collaborative web tools, which make it possible record information by population, and subsequently make their data available. This work proposes a method of grouping georeferenced data from several sources, to perform a statistical analysis using the technique of Principal Components Analysis, which can identify in a georeferenced way which characteristics influence more in the occurrence of traffic accidents of vehicles. After the analysis, a new methodology for visualizing results, plotted on maps, is explored, which can serve as an aid to government agencies and decision makers who take actions to reduce traffic accidents.

Keyword: Georeferenced data. Voronoi Diagrams. Principal Component Analysis. Traffic-accidents. Open and collaborative data.

LISTA DE FIGURAS

Figura 1: Funcionamento de latitude e longitude.....	30
Figura 2: Casos de cólera e poços de água na análise de cólera de John Snow.....	31
Figura 3: Mapa de calor dos acidentes de trânsito de Porto Alegre em 2013.....	32
Figura 4: Diagrama Voronoi para oito pontos no plano	33
Figura 5: Adaptação do mapa dos casos de cólera de Snow, com Diagrama de Voronoi.....	34
Figura 6: Movimentação de eixos da PCA conforme distribuição dos dados.	35
Figura 7: Matriz de dados com n indivíduos e p variáveis.	35
Figura 8: Esquema de aplicação da análise de componentes principais	36
Figura 9: Índice de Acidentes na Zona por Extensão da Malha Viária	38
Figura 10: Mapa de calor dos crimes de Nova York, retirado do <i>site</i> Trulia.....	40
Figura 11: Tela inicial do <i>site</i> Data Portals	43
Figura 12: Mapa com dados do <i>site</i> Mapa Radar.....	46
Figura 13: Diagrama que representa as etapas do método proposto no estudo	47
Figura 14: Exemplos de dados georreferenciados utilizados.....	50
Figura 15: Digrama de Voronoi formado com a localização dos acidentes	52
Figura 16: Exemplo de visualização do resultado da análise.....	57
Figura 17: Percentual de influência das variáveis estudadas nos acidentes de trânsito	59
Figura 18: : Percentual de influência das variáveis estudadas nos acidentes de trânsito, para cada hospital	60
Figura 19: Visualização do resultado da análise para o Hospital das Clínicas	61
Figura 20: Visualização do resultado da análise para o Hospital Moinhos de Vento.....	62
Figura 21: Visualização alternativa do resultado da análise para o Hospital das Clínicas	63
Figura 22: Visualização alternativa do resultado da análise para o Hospital Moinhos de vento....	64

LISTA DE TABELAS

Tabela 1: Dados disponíveis para análise	48
Tabela 2: Exemplo de normalização do dia da semana/hora do acidente.....	54
Tabela 3: Matriz de dados com exemplo de indivíduos e variáveis	55

LISTA DE ABREVIATURAS

Gov Governo

LISTA DE SIGLAS

API	Application Programming Interface (Interface de Programação de Aplicativo)
CSV	Comma-Separated Values (Valores Separados por Vírgula)
HTML	HyperText Markup Language (Linguagem de Marcação de Hipertexto)
JSON	JavaScript Object Notation (Notação de Objetos JavaScript)
OMS	Organização Mundial da Saúde
ONG	Organização Não Governamental
PCA	Principal Component Analysis (Análise de Componentes Principais)
PIPCA	Programa Interdisciplinar de Pós-Graduação em Computação Aplicada
XML	Extensible Markup Language (Linguagem Extensível de Marcação)

SUMÁRIO

1 INTRODUÇÃO	25
1.1 Motivação	25
1.2 Problema	25
1.3 Objetivos	27
1.4 Organização do trabalho	27
2 CONCEITOS BÁSICOS.....	29
2.1 Dados abertos.....	29
2.1.1 Dados abertos governamentais	29
2.2 Visualização Georreferenciada e Geoprocessamento	30
2.3 Diagramas de Voronoi	33
2.4 Análise de componentes principais	34
3 TRABALHOS RELACIONADOS	37
3.1 Análise de dados de trânsito georreferenciados	37
3.2 Visualização de dados georreferenciados.....	40
4 DADOS DISPONÍVEIS NA WEB.....	43
4.1 Portais de dados abertos do governo.....	44
4.2 Ferramentas colaborativas com dados abertos.....	45
5 MÉTODO.....	47
5.1 Coleta de dados e definição das informações utilizadas.....	47
5.2 Normalização e relacionamento dos dados	50
5.2.1 Acidentes de trânsito.....	51
5.2.2 Precipitação pluviométrica	52
5.2.3 Radares.....	53
5.2.4 Dia da semana / hora.....	53
5.2.5 Abordagem balada segura.....	54
5.2.6 Buracos nas vias.....	55
5.3 Processamento dos dados (PCA).....	55
6 RESULTADOS.....	59
6.1 Visualização dos resultados	61
7 CONCLUSÃO	65
7.1 Trabalhos Futuros	66
REFERÊNCIAS.....	67

1 INTRODUÇÃO

1.1 Motivação

Estudos apontam que, em 2013, o universo digital era composto por 4,4 trilhões de *gigabytes* de dados e, que até 2020, deve atingir 44 trilhões. Dos 4,4 trilhões de *gigabytes* existentes em 2013, somente 22% poderiam ser analisados (EMC, 2014).

Grande parte dos dados existentes no mundo não está disponível ao público, nem estruturados a ponto de permitir sua compreensão. Com isso, a extração de conhecimento que poderia ser útil à sociedade, acaba não sendo possível. A fim de mudar esse cenário, diversas organizações, institutos de pesquisa, empresas e governo têm realizado ações com o objetivo de desenvolver tecnologias e meios para disponibilização, agregação de valor e consumo de dados que possam servir para análise e descoberta de outros conhecimentos (ISOTANI; BITTENCOURT, 2016).

Dentro desse contexto, desde 2009 o governo brasileiro vem desenvolvendo políticas de disseminação de informações governamentais para uso livre pela sociedade. Esse processo inclui a implantação do portal *dados.gov.br*, que disponibiliza dados públicos do Brasil, em um formato que possibilite sua reutilização para desenvolvimento de aplicativos para toda a sociedade (SLTI, 2015).

Paralelamente a isso, devido à popularização da internet e de *smartphones* com sensores GPS, a população das grandes cidades está cada vez mais capturando e compartilhando informações a respeito de eventos e fatos ocorridos nas cidades. Como exemplo, têm-se os dados de trânsito do aplicativo Waze (WAZE, 2016), os eventos marcados na ferramenta Guga (SCHÜNKE; OLIVEIRA; VILLAMIL, 2013) ou na ferramenta *colab.re* (COLAB.RE, 2015), as marcações de buracos nas vias por meio do Buracos Monitor (KLEIN; BORGES, 2015), as indicações de radares de trânsito do Mapa Radar (RADAR, 2016), além de aplicativos como *Facebook* e *Twitter*, que não têm esse fim específico, mas de forma indireta recebem inúmeros dados de eventos georreferenciados, enviados pelos seus usuários. Ademais, alguns destes aplicativos disponibilizam sua base de dados para que seja utilizada em outros projetos.

Avaliando esse contexto, pode-se notar que existe acesso a diversas fontes de dados, o que, por sua vez, permite obter informações relacionadas a diferentes variáveis a respeito das mesmas cidades, ou mesmas localizações geográficas. Estes dados podem ser agrupados com o intuito de que sejam aplicadas metodologias de análise de dados, a fim de tirar conclusões sobre os fatos ocorridos em uma cidade. Isso pode auxiliar a tomada de decisões por parte de governos, autoridades, membros de ONGs ou até mesmo empresários, além da própria população. Em suma, essa capacidade de combinar diferentes conjuntos de dados faz com que possam ser desenvolvidos mais e melhores produtos e serviços (OKFN, 2012).

1.2 Problema

Segundo dados do Relatório Global sobre o Estado da Segurança Viária de 2015, emitido pela Organização Mundial da Saúde (OMS, 2015), a maior causa de mortes entre

jovens de 15 a 29 anos, em 2012, foram os acidentes de trânsito de veículos, superando causas como suicídios e homicídios. Em adição, segundo dados levantados pelo Observatório Nacional de Segurança, os acidentes de trânsito devem se tornar a sétima maior causa de óbitos no mundo em 2030, matando mais que doenças como diabetes e hipertensão (ONS, 2014).

Com base nos dados apresentados, pode-se perceber o quanto é importante entender a causa da grande ocorrência de acidentes de trânsito (e das mortes e feridos que eles geram), bem como os fatores que mais influenciam nesses acontecimentos. Desta forma, os órgãos competentes ou interessados poderiam, dentre outras ações possíveis, agir sobre os fatores de influência, diminuindo o número de acidentes.

Nessa linha, pode-se dizer que em uma cidade há fatos que ocorrem de forma isolada, assim como há aqueles que estão relacionados a outras ocorrências ou características urbanas daquela mesma região. Por exemplo o grande número de acidentes de trânsito ocorridos numa rua ou bairro, pode estar diretamente ligado ao fato de não ocorrerem blitz policiais naquela área, ou pelo fato de haver muitos buracos na via. Todavia, para confirmar essa possibilidade e buscar entendimento, é necessário realizar a análise da relação entre os fatores que podem influenciar na ocorrência de tal fato.

Conforme citado, pode-se ter acesso a diversas fontes de dados, seja do governo ou de aplicativos que a população utiliza. Essas bases de dados contém grande quantidade de informações a respeito de eventos georreferenciados ocorridos nas cidades, o que dificulta a análise de forma manual, bem como a extração de relações e conclusões, caso não seja utilizado um método de análise adequado. Além disso, normalmente as informações são encontradas de forma isolada uma da outra, o que dificulta o seu agrupamento para a obtenção de conclusões.

Baseando-se neste contexto, este trabalho visa aplicar uma metodologia específica de coleta e agrupamento, análise e visualização de dados. Pretende-se que os dados relacionados ao trânsito, oriundos de diferentes fontes, possam não só ser relacionados, mas também sumarizados. Do mesmo modo, busca-se identificar as suas relações e exibir os resultados por meio de uma visualização específica, de forma a subsidiar a tomada de decisões a quem interessar.

Nessa perspectiva, Boni et al. (2010) realizaram um trabalho em que se estudou a relação entre acidentes de trânsito e bares na cidade de Porto Alegre. Uma das primeiras percepções foi que havia diversas fontes de dados relacionados ao trânsito, provenientes de diversos órgãos. Contudo essas bases não eram interligadas, e não havia campos em comum que pudessem servir para relacionar as informações, o que gerou grande dificuldade de interligação dos dados. Em vista disso, um dos objetivos do presente trabalho é utilizar uma metodologia capaz de superar essa dificuldade, utilizando-se de características georreferenciadas e temporais para estabelecer a relação entre os dados oriundos de diferentes fontes. Nesse sentido, o georreferenciamento irá servir não só para análise espacial, mas também para o vínculo dos dados.

Por fim, as principais contribuições do presente trabalho são: (i) encontrar as variáveis (dentre as estudadas) que mais influenciam na ocorrência de acidentes de trânsito na cidade de Porto Alegre, servindo como informação útil para tomadores de decisão, assim como para a população em geral; (ii) a elaboração de uma metodologia de agrupamento, análise de correlações e visualização dos resultados, que leva em conta o georreferenciamento dos dados, e que posteriormente poderá ser utilizada por outros pesquisadores ou órgãos, a fim de identificar fatores que influenciam em acidentes de trânsito em outras cidades, ou, usando

outras fontes de dados, a fim de identificar fatores que influenciam na ocorrência de outros tipos de fatos (assaltos, violência, doenças, etc).

1.3 Objetivos

O objetivo deste trabalho é propor um método para análise da correlação de dados georreferenciados dos fatores que podem influenciar na ocorrência de acidentes de trânsito de veículos, bem como a visualização georreferenciada destas correlações. Para que o objetivo geral do trabalho seja atingido, destacam-se alguns objetivos específicos:

- a) Coletar informações georreferenciadas, possivelmente relacionadas e oriundas de diferentes fontes.
- b) Elaborar um método para agrupamento e análise das informações usando suas características de georreferenciamento.
- c) Elaborar um método para visualização dos resultados obtidos, plotados sobre mapas, de forma georreferenciada.
- d) Aplicar o método de análise e visualização em dados relacionados ao trânsito, a fim de demonstrar a aplicação do trabalho.

1.4 Organização do trabalho

Este trabalho está organizado da seguinte forma: no Capítulo 2 consta o detalhamento de alguns conceitos básicos que serão utilizados para execução do trabalho, e que são de importante entendimento; no Capítulo 3, há uma revisão de trabalhos relacionados ao tema deste estudo; já o Capítulo 4 apresenta fontes de dados abertos que poderiam ser utilizadas nesta pesquisa ou em estudos semelhantes; enquanto no Capítulo 5, é exibida a construção do método proposto para manipulação e análise dos dados. No Capítulo 6 são exibidos os resultados deste trabalho e exploradas as novas metodologias de visualização dos resultados, e por fim, o Capítulo 7 trata das conclusões e trabalhos futuros.

2 CONCEITOS BÁSICOS

Neste capítulo, serão detalhados alguns conceitos básicos de temas que serão abordados ao longo do desenvolvimento deste trabalho, e que são de grande importância no entendimento da dissertação. São abordados assuntos como: dados abertos, visualização georreferenciada e geoprocessamento, Diagramas de Voronoi e Análise de Componentes Principais.

2.1 Dados abertos

O termo “Dados Abertos” apareceu pela primeira vez em um documento de uma agência científica norte-americana, que pregava a troca de informações científicas entre diversos países (CHIGNARD, 2013). Segundo o Manual de Dados Abertos, elaborado pelo *Open Knowledge Internacional*, de uma forma geral, Dados Abertos são dados que podem ser livremente usados, reutilizados e distribuídos, tendo como única exigência a atribuição da fonte e possibilidade de novo compartilhamento com as mesmas regras (OKFN, 2012). A cartilha ainda aborda algumas características que os dados precisam ter para serem considerados abertos:

- a) Os dados devem estar disponíveis preferencialmente pela internet, em um formato conveniente e modificável e, além disso, por um custo baixo;
- b) Os dados devem ser disponibilizados sob regras que permitam sua reutilização, combinação com outros dados e posterior redistribuição;
- c) Qualquer pessoa deve ser capaz de ter acesso ao uso dos dados, não podendo haver restrições quanto ao acesso e quanto à posterior utilização.

2.1.1 Dados abertos governamentais

Na literatura em geral, é comum associar o termo “Dados Abertos” a “Dados Abertos Governamentais”, já que é um assunto que surgiu especificamente para estimular e ajudar os governos a divulgarem dados. Um dos grandes marcos dos Dados Abertos governamentais, deu-se em 2007, quando Barack Obama assumiu a presidência dos EUA e assinou dois memorandos relacionados à adesão do seu governo aos Dados Abertos (CHIGNARD, 2013). Além disso, umas das primeiras iniciativas de governos a abrir seus dados foi o *site* do governo dos EUA (www.data.gov), em operação desde 2008 (DATA.GOV, 2016).

No mundo, existem diversos governos que têm políticas de Dados Abertos, alguns impulsionados pelo *Open Government Partnership*, uma parceria iniciada em 2011 pelo governo de 8 países, dentre eles o Brasil - e que, em 2015, já contava com 69 países (OGP, 2015). Dentre os portais de dados abertos de alguns países, além dos EUA, podem-se citar os *sites* de dados do governo da Alemanha (www.offenedaten.de), da Grã-Bretanha (www.data.gov.uk), da Nova Zelândia (www.data.govt.nz), da Austrália (www.data.gov.au), entre outros.

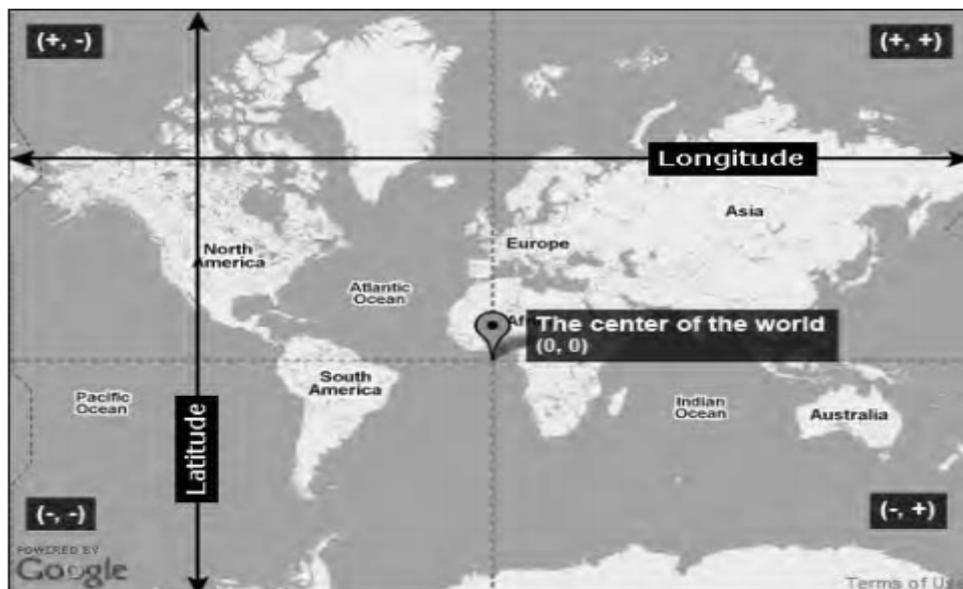
No Brasil, existem diversos movimentos de disponibilização de dados por parte do Governo Federal e dos governos estaduais e municipais. Conforme já citado, o governo brasileiro mantém o Portal Brasileiro de Dados Abertos (www.dados.gov.br) para disponibilização de dados em esfera federal. Outras iniciativas, como Dados RS (www.dados.rs.gov.br/), com informações do Estado do Rio Grande do Sul, e o Portal de Dados Abertos da Cidade de Porto Alegre (www.datapoa.com.br/) também estão surgindo, em diversos estados e cidades. Ainda neste contexto, a Administração Pública Federal, por meio da Lei de Acesso a Informação, reconheceu a necessidade de disponibilização dos dados governamentais em formato aberto (Lei nº 12.527/2011).

2.2 Visualização Georreferenciada e Geoprocessamento

A visualização georreferenciada e o geoprocessamento são dois termos intimamente ligados. O primeiro, também chamado de visualização geográfica, é definido por Ramos (2005) como o uso de representações visuais concretas, por meio do uso de papel ou computador, para tornar visíveis os problemas espaciais tratados, explorando habilidades humanas ligadas à visão. Por seu turno, o geoprocessamento pode ser definido como a análise espacial de dados geográficos, que tem como objetivo mensurar propriedades e relacionamentos entre fenômenos, levando em consideração a localização geográfica. Em outras palavras, geoprocessamento é a inclusão do espaço à uma análise realizada (DRUCK et al., 2004).

Nessa linha, segundo CONCAR (2010), os dados georreferenciados possuem características espaciais que os identificam quanto ao seu posicionamento na Terra e ao seu espaço próximo, em um determinado período de tempo. Neste trabalho, dados georreferenciados são tratados como aqueles que possuem informações de coordenadas de latitude e longitude, conforme Figura 1, ou de endereço (nome da cidade, da rua, e número do imóvel). Esta última, que pode ser facilmente convertida em latitude e longitude com o uso de serviços como a API Google Maps (DEVELOPERS GOOGLE, 2016).

Figura 1: Funcionamento de latitude e longitude.



Fonte: (SVENNERBERG, 2010).

Um dos primeiros trabalhos relacionados ao geoprocessamento, foi uma análise realizada em 1854 por John Snow. Durante uma epidemia de cólera, após cerca de 500 pessoas de uma mesma cidade morrerem em um período de dez dias, o autor marcou em um mapa do centro da cidade os locais das mortes e dos poços de água. Com a espacialização dos dados, representada na Figura 2, ele percebeu que a maioria dos casos estava concentrada em torno de um poço específico, que poderia ser a causa das mortes. Ele então ordenou que o poço fosse lacrado e ajudou a conter a epidemia (KOCH; DENIKE, 2009).

Com efeito, o trabalho de Snow entrou para a história como uma das primeiras análises de dados georreferenciados e é um bom exemplo para demonstrar como a relação espacial entre os dados pode contribuir para o melhor entendimento de um fenômeno (DRUCK et al., 2004). Trata-se de uma situação semelhante à proposta deste trabalho, que propõe a análise da relação entre diversas variáveis para entender o fenômeno “acidente de trânsito”, levando em consideração a localização geográfica dos eventos.

Figura 2: Casos de cólera e poços de água na análise de cólera de John Snow.



Fonte: (LOPES, 2001).

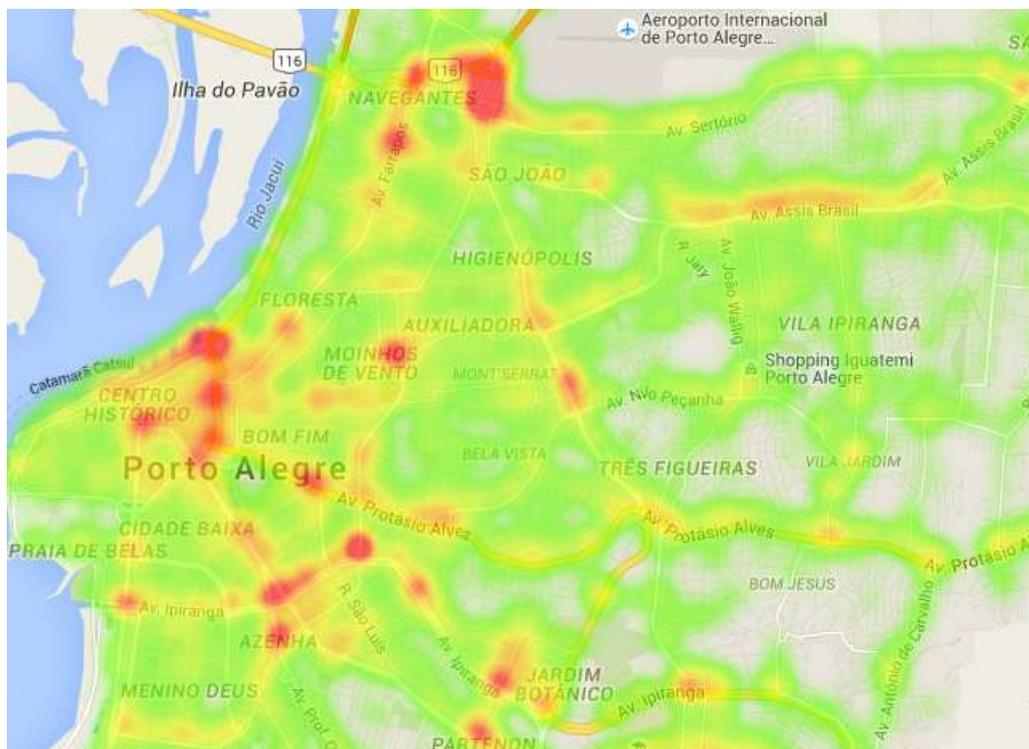
De maneira geral, Sistemas de Informação Geográfica (SIG) são *softwares* que implementam e facilitam tarefas de geoprocessamento. Esses sistemas vêm sendo utilizados frequentemente para determinação geográfica de distribuição de doenças, análise espacial e temporal de tendências, mapeamento de populações sob risco, alocação de recursos e planejamento de intervenções ao longo do tempo (BONI et al., 2010). São exemplos de *softwares* SIG conhecidos: QGIS, ArcGIS, GEOMEDIA, GRASS, gvSIG e TerraView.

Um recurso bastante utilizado para visualização de dados georreferenciados são os mapas de calor (*heat maps*), que podem ser definidos como a representação de dados por

meio de cores em uma imagem (GREENBERG; XU; KUMAR, 2013). Normalmente, os mapas de calor são utilizados para visualização da concentração de eventos sobre uma área específica, onde os dados são representados por cores que indicam a intensidade do dado, de acordo com uma escala de cores (PEREZ-LLAMAS; LOPEZ-BIGAS, 2011).

Nessa perspectiva, tem-se, na Figura 3 um exemplo de mapa de calor, que foi construído com a informação dos locais onde ocorreram acidentes de trânsito em Porto Alegre, no ano de 2013. Para construção do mapa de calor, foi utilizada a API do Google Maps, que traz funções para fácil integração com outras ferramentas. Todos os locais onde ocorreram acidentes de trânsito são representados pela cor verde, enquanto locais com maior concentração de acidentes são representados pela cor amarelo, podendo variar até o vermelho conforme a intensidade. Visualmente, torna-se mais fácil identificar onde ocorrem mais acidentes de trânsito na cidade de Porto Alegre.

Figura 3: Mapa de calor dos acidentes de trânsito de Porto Alegre em 2013



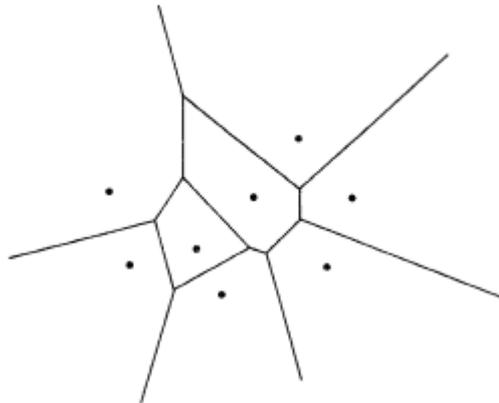
Fonte: Elaborado pelo autor.

Atualmente, aplicativos que utilizam visualização georreferenciada e geoprocessamento estão se tornando cada vez mais comuns, sobretudo em função da popularização e miniaturização de equipamentos de sistema GPS. Também contribuiu para esse processo o surgimento de ferramentas como o Google Maps (GOOGLE, 2016), que consiste em uma ferramenta web para visualização, pesquisa e navegação em mapas, com interface visual para usuários. O Google Maps também possui uma API - que pode ser acessada por desenvolvedores de *software* via *Javascript*, além de outras tecnologias, e que contém diversos métodos para pesquisa e manipulação de informações dentro dos mapas. A API possui uma documentação completa de todos os seus métodos e exemplos de utilização (DEVELOPERS GOOGLE, 2016).

2.3 Diagramas de Voronoi

Segundo Moura (2003), o princípio do Diagrama de Voronoi ou de Thiessen é de que, tendo-se um plano com diversas fontes geradoras, todos os pontos desse plano relacionam-se a uma única fonte geradora, que é a fonte mais próxima de si. Desse modo, são traçadas linhas que dividem os pontos relacionados a cada fonte geradora, e o resultado é uma malha de polígonos cujas distâncias entre a fonte e a borda de cada polígono são as menores possíveis. Um exemplo de Diagrama Voronoi gerado para oito pontos distribuídos em um plano, é apresentado na Figura 4.

Figura 4: Diagrama Voronoi para oito pontos no plano



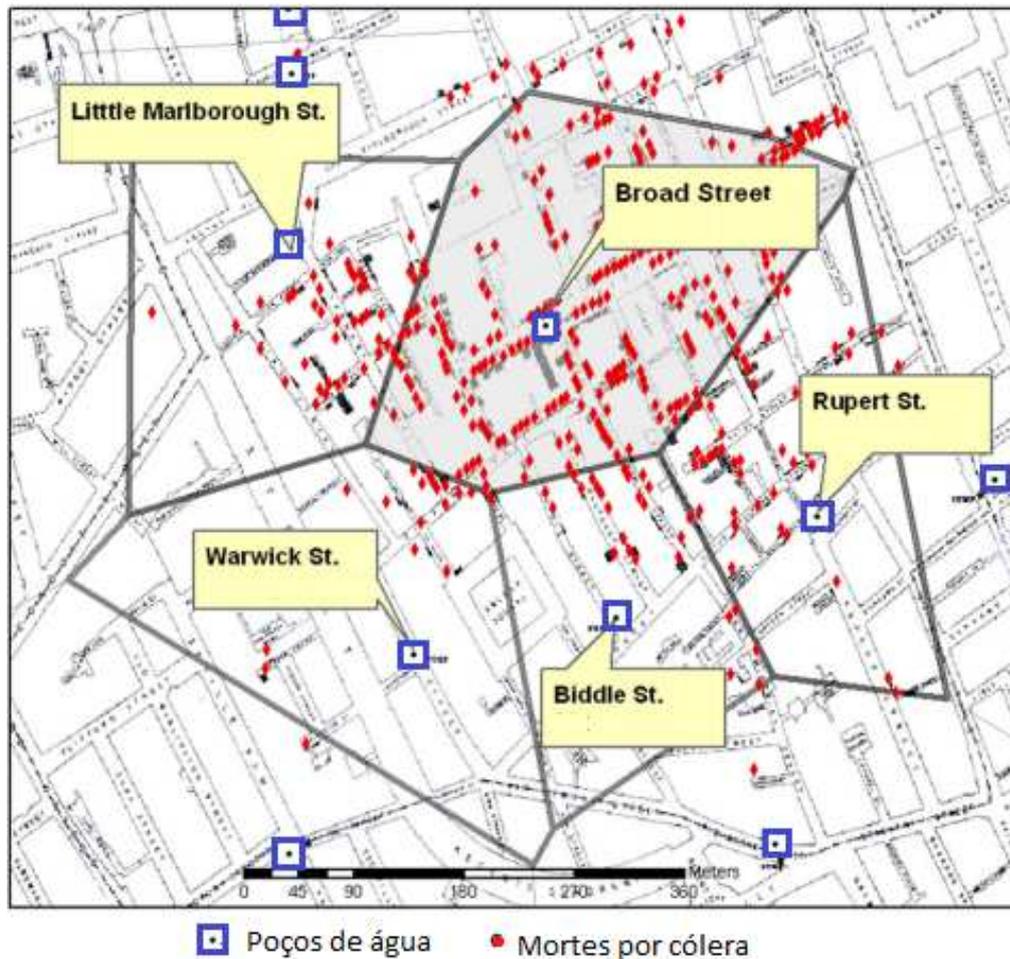
Fonte: (AURENHAMMER, 1991).

Os diagramas de Voronoi foram abordados pela primeira vez em 1840, por Gauss (GAUSS, 1840). Depois disso, sua ideia foi explorada por Dirichlet (DIRICHLET, 1850) e, em 1908, o russo Georgy Voronoy (VORONOVY, 1908), cujo nome foi atribuído ao Diagrama, fez um trabalho explorando os mesmos conceitos dos autores anteriores, porém explorando um maior número de dimensões.

Por sua vez, Berg et al. (2000) definem formalmente os Diagramas de Voronoi da seguinte maneira: considera-se um conjunto de n pontos em um plano P , onde $P = \{p_1, p_2, \dots, p_n\}$. Esses pontos são chamados de locais (pontos de referência). Define-se o diagrama de Voronoi do conjunto P como a divisão do plano em n regiões, uma para cada local em P . Um ponto q pertence à célula correspondente ao local p_i , se, e somente se, $\text{dist}(q, p_i) < \text{dist}(q, p_j)$ para cada $p_j \in P$ com $j \neq i$ ($\text{dist}(a, b)$ é a distância euclidiana).

Nessa linha, na Figura 5 pode-se ver uma imagem proveniente do trabalho de Koch e Denike (2009), que é uma adequação do mapa que John Snow fez em 1854 para estudar os casos de cólera em uma cidade. Os pontos azuis representam os poços de água da cidade, e que serão os pontos geradores para o Diagrama de Voronoi. Todos os outros pontos do mapa serão considerados para o diagrama, mas os pontos em vermelho, que são os locais dos casos de cólera, são os que interessam. Assim, são criadas delimitações entre cada um dos poços de água, que ficam exatamente entre dois poços vizinhos, formando polígonos. Cada caso de cólera fica dentro de um polígono, e, automaticamente, é associado ao poço de água que está dentro dessa área.

Figura 5: Adaptação do mapa dos casos de cólera de Snow, com Diagrama de Voronoi



Fonte: adaptado de KOCH e DENIKE (2001).

Maiores informações a respeito da história e da implementação de algoritmos de Diagramas de Voronoi, podem ser obtidas em (AURENHAMMER, 1991). Vale salientar que existem *softwares* que geram automaticamente Diagramas Voronoi, como é o caso do Matlab (MATHWORKS, 2016) e do ArcGIS (ESRI, 2016).

2.4 Análise de componentes principais

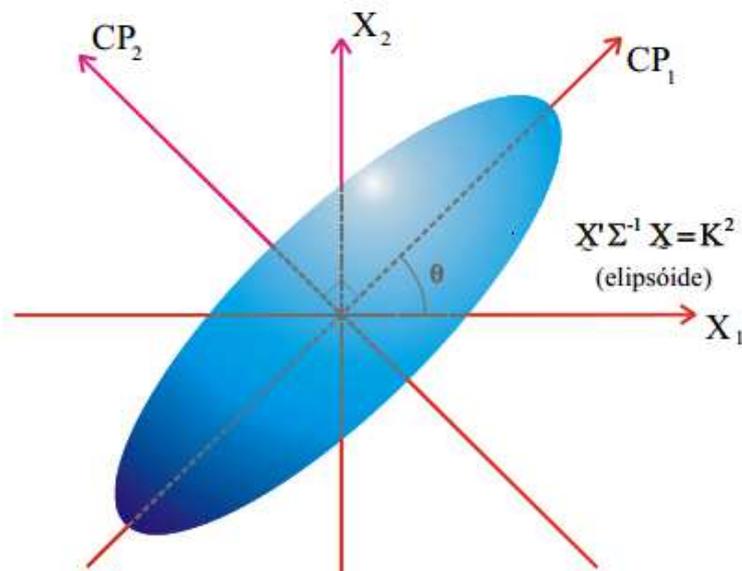
A PCA (*Principal Component Analysis*, ou Análise de Componentes Principais) surgiu em um trabalho de Pearson (1901), com o objetivo de encontrar linha e planos que melhor se ajustassem a um conjunto de pontos em um espaço *p-dimensional*. Entretanto, ela somente foi denominada “Principal Component Analysis” e explorada pela primeira vez, da maneira como é conhecida hoje, por Hotelling (1933).

Essa técnica é bastante útil para o presente trabalho, visto que há aqui uma grande quantidade de dados a serem analisados. De fato, segundo Vicini (2005), a PCA é uma técnica matemática de análise multivariada, que possibilita investigações com um grande número de dados disponíveis. A ideia principal da análise é reduzir o conjunto de dados a serem analisados, principalmente quando o conjunto de dados tem uma grande quantidade de variáveis inter-relacionadas.

Não obstante, a técnica é capaz de transformar um conjunto de dados em outro conjunto com dimensões equivalentes, sendo este último chamado de “componentes principais” (CP). Essa transformação acontece com a menor perda de informação possível, porém, busca eliminar algumas variáveis originais que contêm pouca informação, de maneira que esta eliminação só é possível se as variáveis iniciais possuírem correlação entre si. Em linhas gerais, a PCA determina as variáveis de maior influência na formação de cada componente (VICINI, 2005).

Como exemplo, suponha-se que é utilizada a PCA para análise de duas variáveis - X_1 e X_2 . Na Figura 6, a elipse representa a variação dos dados, considerando os eixos X_1 e X_2 . Após a execução da PCA, os eixos X_1 e X_2 são movimentados de forma a ficarem na direção da maior variabilidade dos dados, originando os eixos representados por CP1 (componente principal 1) e CP2 (componente principal 2) (JOHNSON; WICHERN, 1992).

Figura 6: Movimentação de eixos da PCA conforme distribuição dos dados.



Fonte: (LOPES, 2001).

Para o cálculo da PCA, é necessário ter uma matriz (M) de dimensão $n \times p$ conforme Figura 7. Observa-se que as linhas (n) representam os indivíduos, ao passo que as colunas (p) representam as variáveis ou características dos indivíduos.

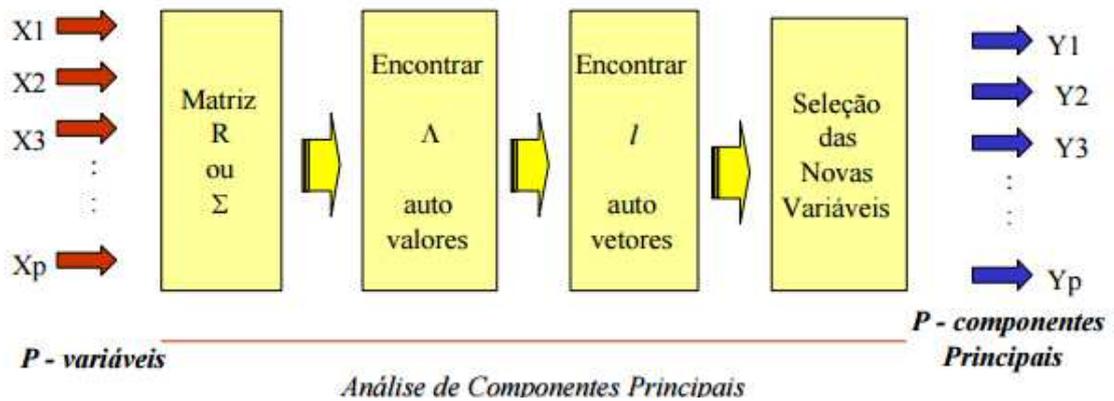
Figura 7: Matriz de dados com n indivíduos e p variáveis.

Indivíduos	Variáveis						
	X_1	X_2	X_3	...	X_j	...	X_p
1	X_{11}	X_{12}	X_{13}	...	X_{1j}	...	X_{1p}
2	X_{21}	X_{22}	X_{23}	...	X_{2j}	...	X_{2p}
3	X_{31}	X_{32}	X_{33}	...	X_{3j}	...	X_{3p}
...
i	X_{i1}	X_{i2}	X_{i3}	...	X_{ij}	...	X_{ip}
...
n	X_{n1}	X_{n2}	X_{n3}	...	X_{nj}	...	X_{np}

Fonte: Elaborado pelo autor.

O esquema representado na Figura 8 é um resumo do passo a passo da aplicação da PCA, para o qual a matriz M deve ser submetida. Um maior detalhamento dos cálculos e do passo a passo necessário para aplicação do método PCA podem ser encontrados em (JACKSON, 1991).

Figura 8: Esquema de aplicação da análise de componentes principais



Fonte: (SOUZA, 2000).

Com efeito, existem diversos *softwares* que realizam automaticamente os cálculos necessários para execução da PCA. Normalmente, é necessário informar um arquivo de texto ou planilha eletrônica que contenha os dados, em um formato que seja aceito pelo *software*, e ele então processará as informações, como é o caso do Matlab (MATWORKS, 2016) e R (BORCARD, 2011).

3 TRABALHOS RELACIONADOS

Em verdade, diversos autores já realizaram trabalhos de temáticas semelhantes, buscando analisar e explicar as características ligadas aos acidentes de trânsito. No entanto o presente estudo apresenta características distintas, que não são encontradas em outros trabalhos pesquisados: (i) propõe uma metodologia para analisar dados oriundos de diversas fontes, sejam eles fornecidos por órgãos públicos, empresas, aplicativos/portais colaborativos, entre outros, desde que sejam georreferenciados; (ii) foca principalmente em características urbanas, incluindo também sociais e meteorológicas, e não em características dos próprios acidentes, como dados do carros ou condutores envolvidos nos acidentes; (iii) tem por objetivo encontrar as características que mais influenciam na concentração dos acidentes de trânsito em uma área urbana, por intermédio da análise da área que cada acidente ‘ocupa’ com relação aos outros; (iv) é simples e objetiva, o que poderá ser útil para tomadores de decisão e para a população em geral, sem a necessidade de compreender e analisar grandes tabelas, gráficos ou dados resultantes de análises estatísticas; (v) propõe dois métodos inovadores para visualização de forma georreferenciada dos resultados da análise, buscando facilitar o entendimento por parte de quem está buscando a informação. Além disso, apesar da metodologia poder ser aplicada para qualquer localização geográfica, este trabalho contribui com a análise dos dados para a cidade de Porto Alegre, possibilitando que autoridades locais utilizem os resultados para tomar ações visando à redução do número de acidentes.

3.1 Análise de dados de trânsito georreferenciados

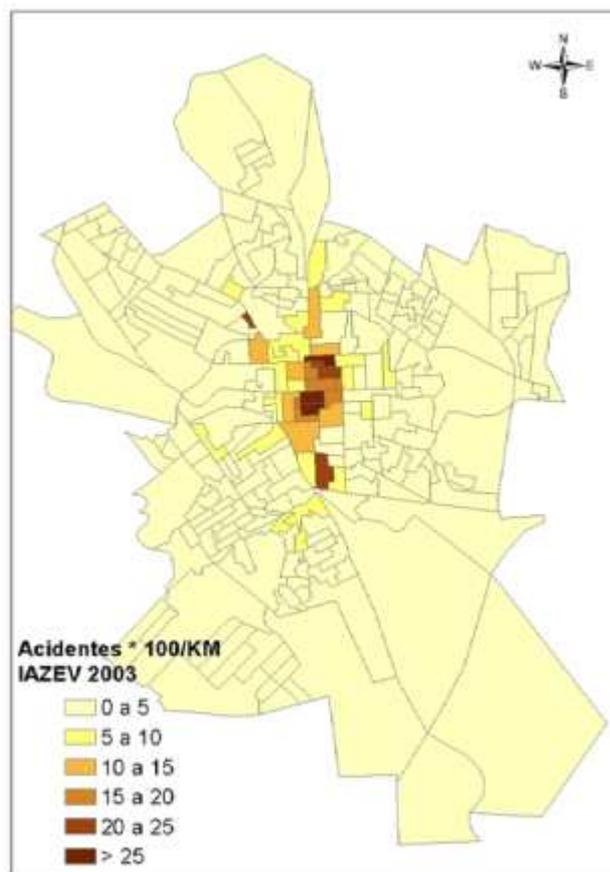
Existem diversos trabalhos que buscaram analisar dados georreferenciados relacionados aos acidentes de trânsito, utilizando as mais diversas técnicas. Segundo Jacob e Sayer (1983), na Europa e América do Norte a preocupação em entender o trânsito e os motivos dos acidentes de trânsito surge já nos anos 1940. Posteriormente, no início da década de 70, foram os países de terceiro mundo que notaram que os problemas de segurança rodoviária vinham crescendo e surgiram então as primeiras pesquisas realizadas nestes países, com o objetivo de determinar a natureza e extensão dos problemas dos acidentes de trânsito.

Queiroz (2003), por exemplo, propôs analisar geograficamente os acidentes de trânsito da cidade de Fortaleza (CE), a partir de uma base de dados da ocorrência destes, chamada de SIAT-FOR, que é alimentada por nove entidades, dentre órgãos municipais, estaduais e institutos independentes. Nesse contexto, o autor fez uma revisão de métodos de análise estatística espacial na área de segurança de tráfego, explicando suas características e diferenças. Sem utilizar algum um método específico, ele realizou algumas validações na base de dados utilizando o SIG TransCAD 3.0, tais como quantidade de acidentes que tem informação de geolocalização por severidade do acidente, quantidade de acidentes em meio de quadra e em esquinas e, para as esquinas, a quantidade de acidentes em locais com semáforo e sem semáforo. Por último, criou alguns mapas temáticos georreferenciados utilizando os dados disponíveis, como o local dos acidentes com óbitos, cruzamentos com mais acidentes, e quantidade de acidentes por bairro.

Já Santos (2006) realizou uma análise espacial dos acidentes de trânsito no município de São Carlos (SP), fazendo uso de SIG associado a ferramentas de estatística espacial. O

autor utilizou uma base de dados de Boletins de Ocorrência de acidentes de trânsito ocorridos no período de 2001 a 2003. Utilizou o *software* Crimestat para identificar pontos com maior ocorrência de acidentes e também para fazer uma distribuição espacial dos acidentes por tipo de ocorrência. Também identificou zonas críticas com maior ocorrência de acidentes, utilizando dois métodos: IAZEV (Índice de Acidentes na Zona por Extensão da Malha Viária), um índice que leva em conta a quantidade de acidentes em uma determinada área e a quantidade de quilômetros de vias nessa mesma área; e IUPSZ (Índice de Unidade Padrão de Severidade por Zona), um índice por meio do qual é atribuído um peso para cada tipo de severidade de acidente, valores estes que são somados e posteriormente divididos pela quantidade de quilômetros de vias em determinada área. Além disso, o autor ainda identificou tendências de deslocamento dos acidentes com o passar do tempo. A Figura 9 exibe um dos resultados da análise do autor.

Figura 9: Índice de Acidentes na Zona por Extensão da Malha Viária



Fonte: (SANTOS, 2006).

Já Sehaber (2013) analisou em torno de 17 mil acidentes de trânsito ocorridos no trecho paranaense da rodovia BR-376, entre 2009 e 2012, a partir de boletins de ocorrência fornecidos pela Polícia Rodoviária Federal do Paraná. Entre os dados utilizados, estão características dos acidentes, dos veículos e dos condutores envolvidos. O autor dividiu os dados em dois grandes grupos, cada um agrupando aproximadamente metade do trecho da rodovia que foi utilizado. Para cada um dos grupos, normalizou os dados e utilizou o método de Análise de Componentes Principais para avaliar quais os locais mais perigosos da rodovia e quais as variáveis que mais influenciam na ocorrência de acidentes.

Alguns autores utilizam alguns artifícios para coletar dados relacionados ao trânsito em diferentes ferramentas, com o objetivo de aumentar e/ou melhorar sua base de dados de análise. Em seu trabalho, Barbosa (2015) propõe uma metodologia para avaliar o impacto que as ocorrências de acidentes de trânsito causam na velocidade em que os demais veículos estão andando na via. Para execução do trabalho, o autor utilizou dados de trânsito da cidade do Rio de Janeiro (RJ): (i) malha viária da cidade, obtida por meio da API da ferramenta Google Maps; (ii) informação de velocidade média dos locais de interesse, obtidos a partir de uma ferramenta *online* de uma empresa que faz medições de velocidade nas vias por intermédio de equipamentos fixos instalados; (iii) extração de dados em postagens do *Twitter* a respeito de localização de acidentes de trânsito, utilizando uma metodologia para a filtragem das informações. Assim o autor cruzou estes dados e fez uma análise para identificar o quanto um acidente influencia na velocidade dos outros veículos.

Por sua vez, Redlich (2013) buscou desenvolver um modelo para auxiliar na análise de eventos ocorridos no trânsito de grandes cidades, usando uma grande massa de dados publicados na *internet*. A autora utilizou conceitos de ontologia para extrair dados da ferramenta *Twitter*, publicados por usuários comuns em linguagem natural.

Andrade et al. (2014) mapeou 379 acidentes de trânsito ocorridos na rodovia BR 277 entre 2007 e 2009, com o objetivo de determinar quais os fatores que afetam nas mortes de trânsito. Foram utilizadas quatro técnicas: análise espacial por pontos pela aplicação do estimador de densidade de Kernel e análise de Wavelet para identificar regiões com maior concentração de ocorrências, análise de ambiente construído e PCA para verificar padrões que contribuem para ocorrência de acidentes de trânsito. A principal conclusão da análise, é que as variáveis tamanho da estrada em área urbana, iluminação limitada, via dupla e poucas vias auxiliares, foram associados a maior ocorrência de acidentes de trânsito.

Um trabalho interessante que também busca utilizar dados abertos, porém somente governamentais, para realizar uma análise estatística a fim de identificar quais são os fatores que mais influenciam na ocorrência dos casos de dengue. Neste trabalho Kaieski, Oliveira e Villamil (2016) coletam dados de temperatura e pluviosidade de todo o território brasileiro, e fazendo uso da técnica de PCA, obtendo como conclusão, a porcentagem que cada um dos dados influencia na ocorrência dos casos de dengue em cada estado brasileiro. Adicionalmente, os autores propõem uma forma de visualização georreferenciada de um mapa de calor temporal dos casos de dengue utilizados no trabalho, que pode auxiliar na análise da distribuição dos casos ao longo do tempo.

Em âmbito mundial, também existem diversos trabalhos relacionados. Ertunç, Çay e Mutluoglu (2016) fizeram uma análise dos acidentes de trânsito ocorridos na Turquia, onde cerca de 4.000 pessoas morrem a cada ano, vítimas dos acidentes. Foram analisadas informações georreferenciadas de um banco com dados dos acidentes (como o dia e hora, tipo do acidente, número de mortos e número de feridos) ocorridos nos anos de 2009 e 2010 na cidade de Antalya, e dentre os resultados o autor identificou quais os cruzamentos e pontos da cidade com maior concentração de acidentes em cada ano e também no período todo.

Mehmandar, Soori e Mehrabi (2016) realizaram um trabalho com o objetivo de prover a analisar a tendência de mortes por acidente de trânsito no Irã. Coletaram todas as informações relacionadas a acidentes de trânsito fatais dos anos de 2004 até 2013, disponíveis na base de dados da Organização de Medicina Legal do Irã. Para a análise, os autores utilizaram o modelo auto-regressivo integrado de média móvel (ARIMA), podendo notar como resultado, que a quantidade de mortes no trânsito vem diminuindo e deve continuar a diminuir, além disso, identificaram sazonalidade em um período do ano, onde a população

costuma viajar mais. Após as previsões para 2014 e 2015, compararam o resultado com o que ocorreu em 2014, e o resultado foi satisfatório, com pouca variação.

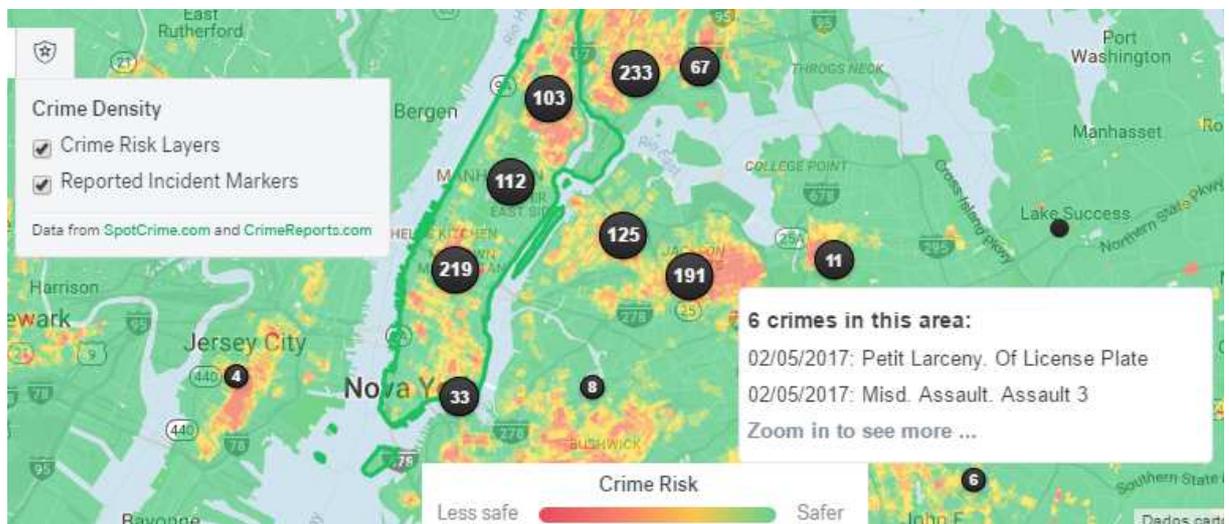
Enquanto Yalcin (2013) analisou os dados de acidentes de trânsito da cidade de Osmaniye, na Turquia. Ele utilizou o método qui-quadrado, levando em conta diferentes tipos de variáveis dos acidentes, como estação do ano, mês, dia da semana, horário, se é dia ou não, e número de veículos envolvidos. Os resultados mostraram quais as variáveis tem relação significativa com a ocorrência de acidentes. Alguns dos resultados foram que há horários de pico de acidentes, entre 8:00h e 9:00h da manhã, 13:00h e 14:00h, e 16:00h e 17:00h da tarde, enquanto que todos os finais de semana tem a mesma distribuição de acidentes. Além disso, notou que no outono há um maior pico de acidentes.

Outros autores, como Anastasopoulos et al. (2012) utilizam regressão de tobit para analisar acidentes de trânsito. O autor cita que há vários trabalhos que usam essa técnica, mas que nenhum leva em conta a gravidade das lesões causadas pelo acidente, situação que o mesmo leva em consideração. Desta forma, ele faz uma análise dos acidentes ocorridos durante um período de 5 anos nas rodovias de Whashington, nos Estados Unidos.

3.2 Visualização de dados georreferenciados

Tratando de trabalhos relacionados a visualização de dados georreferenciados, temos várias propostas práticas, com mapas temáticos que exploram os mais diferentes tipos de dados. Como exemplos práticos, o *site* Trulia (TRULIA, 2016) apresenta um mapa interativo onde constam informações dos crimes ocorridos e registrados nos departamentos de polícia em todo o território dos Estados Unidos. A ferramenta disponibiliza marcas georreferenciadas nos pontos onde ocorreram crimes, e ao clicar em um ponto, são exibidas informações referentes ao crime ocorrido naquele local. Além disso, no mesmo mapa, é possível ver um mapa de calor dos crimes, onde áreas pintadas na cor vermelha representam locais com maior incidência de crimes, e áreas na cor verde representam locais com nenhuma incidência de crimes, enquanto que as cores intermediárias representam locais com quantidades intermediárias de crimes. Na Figura 10 consta um mapa extraído do *site*, exibindo os crimes para a cidade de Nova York.

Figura 10: Mapa de calor dos crimes de Nova York, retirado do *site* Trulia



Fonte: (TRULIA, 2016).

Por sua vez, o *site* WikiCrimes (WIKICRIMES, 2016), é uma ferramenta colaborativa, onde a população pode reportar locais onde ocorreram crimes (violência, uso e venda de drogas, uso excessivo de álcool, dentre outros), informando o período do dia, os motivos pelo qual o usuário acredita que o fato ocorreu e uma descrição. A ferramenta, em sua página inicial, exibe um mapa de calor plotado sobre um mapa, criado a partir dos locais onde foram reportados crimes. Ainda na página inicial são exibidos, para o local do mapa que está sendo visualizado, o percentual de cada tipo de ocorrência e o percentual de motivos das ocorrências, enquanto em uma área específica é possível ver mais estatísticas a respeito de todas as ocorrências registradas na ferramenta, como o percentual por período do dia, mapa de calor das ocorrências por país, percentual de tipo de ocorrência no mundo todo, ou alguns gráficos de estatísticas por país.

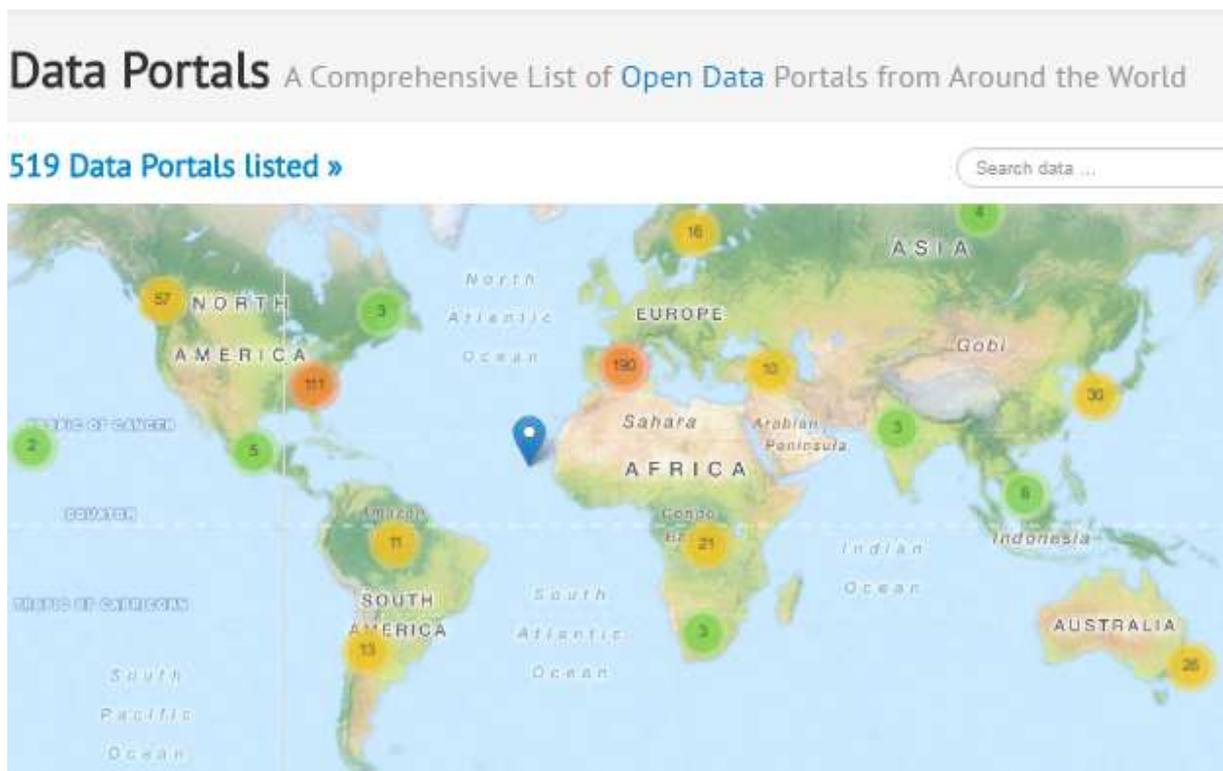
Ao longo deste estudo são vistos outros trabalhos e ferramentas que fazem uso de visualização georreferenciada, como é o caso do mapa dos casos de cólera de John Snow em 1854 (KOCH; DENIKE, 2009), um dos primeiros trabalhos relacionados a visualização de dados de forma georreferenciada, das ferramentas colaborativas Mapa Radar (RADAR, 2016) e Chega de Fiufiu (OLGA, 2016), que neste caso foram abordados como potenciais fontes de dados para análise.

4 DADOS DISPONÍVEIS NA WEB

Neste capítulo serão elencadas fontes de dados das mais diversas categorias, contanto que estar ligados ao trânsito, que é o objeto deste estudo. São apresentados somente dados que estejam disponíveis na *web*, e que sigam ao menos parte das recomendações do Manual de Dados Abertos, já detalhado neste trabalho. Assim, garante-se que os dados possam servir de fonte para estudos este

No *site Data Portals*, são catalogados diversos *sites* que servem para disseminação de dados, mantidos pelos mais variados órgãos. Em sua página inicial, é possível conferir um mapa, no qual há marcas georreferenciadas nos locais do mundo onde há portais de disponibilização de dados, conforme ilustrado na Figura 11. Clicando sobre a marca, são exibidas informações sobre o *site* e um atalho que redireciona para o mesmo. Segundo o próprio *site*, o *Data Portals* é a lista mais abrangente de portais de dados abertos do mundo, sendo mantido por representantes de governos, organizações internacionais e ONG's (DATAPORTALS, 2016).

Figura 11: Tela inicial do *site Data Portals*



Fonte: (DATAPORTALS, 2016).

Existem também ferramentas online que auxiliam governo, empresas e outros órgãos na disponibilização de dados em um portal, como é o caso do CKAN. Este consiste em um sistema de gerenciamento de dados, voltado para governos nacionais e regionais, empresas e organizações, tornando os dados acessíveis, fornecendo ferramentas para agilizar a publicação e o compartilhamento. A ferramenta tem uma API que facilita o acesso aos dados, contando com integração com sistemas CMS, além de outras características que tornam a ferramenta interessante. Além disso, no *site* há um catálogo de portais de dados abertos que usam a

própria ferramenta, de modo que podem ser encontradas fontes de dados dos mais diversos tipos e locais do mundo (CKAN, 2016).

4.1 Portais de dados abertos do governo

Em função da grande demanda da população por transparência do governo e por disponibilização de dados, diversos órgãos do governo têm aberto os seus dados em portais específicos para esse fim. Essas ações vêm sendo incentivadas por medidas como a Lei de Acesso a Informação (Lei nº 12.527/2011).

Nesse sentido, uma das principais iniciativas do governo brasileiro é o Portal Brasileiro de Dados Abertos (disponível em <http://dados.gov.br/>). Esse portal agrupa mais de mil conjuntos de dados, de categorias como governo e política, saúde, educação, segurança, transporte e trânsito, economia e cultura. A maioria dos dados está disponível para *download* em formato HTML ou CSV, e alguns podem ser consultados *online*. Além dos dados, o *site* tem uma seção com uma lista de aplicativos - desenvolvidos por organizações ou pela população - que utilizam essas informações (DADOS.GOV.BR, 2016).

Em nível estadual, existem diversos governos que possuem seus próprios portais de disponibilização de dados abertos, como é o caso do Dados RS (disponível em <http://dados.rs.gov.br/>), que é mantido pelo Governo do Estado do Rio Grande do Sul. Segundo o órgão, tem como objetivo fomentar o desenvolvimento colaborativo de soluções tecnológicas que tenham impacto na vida da população, além da possibilidade de criação de empregos e empresas fomentados pela utilização dos dados. No portal podem ser encontradas informações relativas a assuntos como, por exemplo, educação, governo, finanças, trânsito e saúde. Alguns dados têm atualização diária e diversos deles possuem informações com georreferenciamento (DADOS.RS, 2016).

Outra iniciativa do Estado do Rio Grande do Sul é o Dados TCE, que agrupa dados do Tribunal de Contas do Estado do Rio Grande do Sul. São aí disponibilizados mais de 25 mil conjuntos de dados, agrupados em categorias como despesas, receitas, previdência, gestão fiscal, dentre outros, que estão disponíveis em formatos como CSV, JSON, HTML e XML (DADOS.TCE, 2016).

Do mesmo modo, em nível municipal, têm-se inúmeros exemplos de portais similares. Um deles é o Data POA, que agrupa conjuntos de dados dos mais diversos temas, relacionados à cidade de Porto Alegre (RS). O *site* tem como objetivo a disponibilização de dados abertos, que podem servir para *hackers*, estudantes, desenvolvedores, jornalistas, pesquisadores e empresários que tenham interesse em criar, de forma colaborativa, serviços para a comunidade. Além da disponibilização dos dados, o *site* conta com um catálogo de aplicativos já desenvolvidos por pessoas ou entidades, que tenham usado os dados do *site* para tal (DATAPOA, 2015).

Existem diversas outras iniciativas para disponibilização de dados por parte de órgãos do Governo Brasileiro. A grande maioria das prefeituras estaduais e das capitais brasileiras, além de outros órgãos estaduais e municipais, mantêm portais de dados abertos, como é o caso do Dados Recife (<http://dados.recife.pe.gov.br/>), Alagoas em Dados e Informações (<http://dados.al.gov.br/>), Governo Aberto SP (<http://www.governoaberto.sp.gov.br/>), Dados Abertos Prefeitura de São Paulo (<http://dados.prefeitura.sp.gov.br/>), Data Rio (<http://data.rio/>), entre diversos outros.

Além dos portais específicos para disponibilização de dados, existem órgãos específicos que disponibilizam informações da sua área de atuação, diretamente em seus *sites*. Por exemplo, o Centro Integrado de Comando da Prefeitura de Porto Alegre disponibiliza dados de diversas estações meteorológicas da cidade. Englobando indicadores como temperatura, sensação térmica, umidade, índice de chuva (ou precipitação pluviométrica), pressão e vento, que podem ser consultados por trimestre, período de datas ou mês em um período de anos (CEIC, 2016).

Igualmente, o *site* do Departamento de Trânsito do Rio Grande do Sul (DETRAN-RS) disponibiliza dados de estudos e pesquisas realizadas pelo órgão, que podem ser visualizados no formato de cartilhas, separadas por tema e período (DETRANRS, 2016).

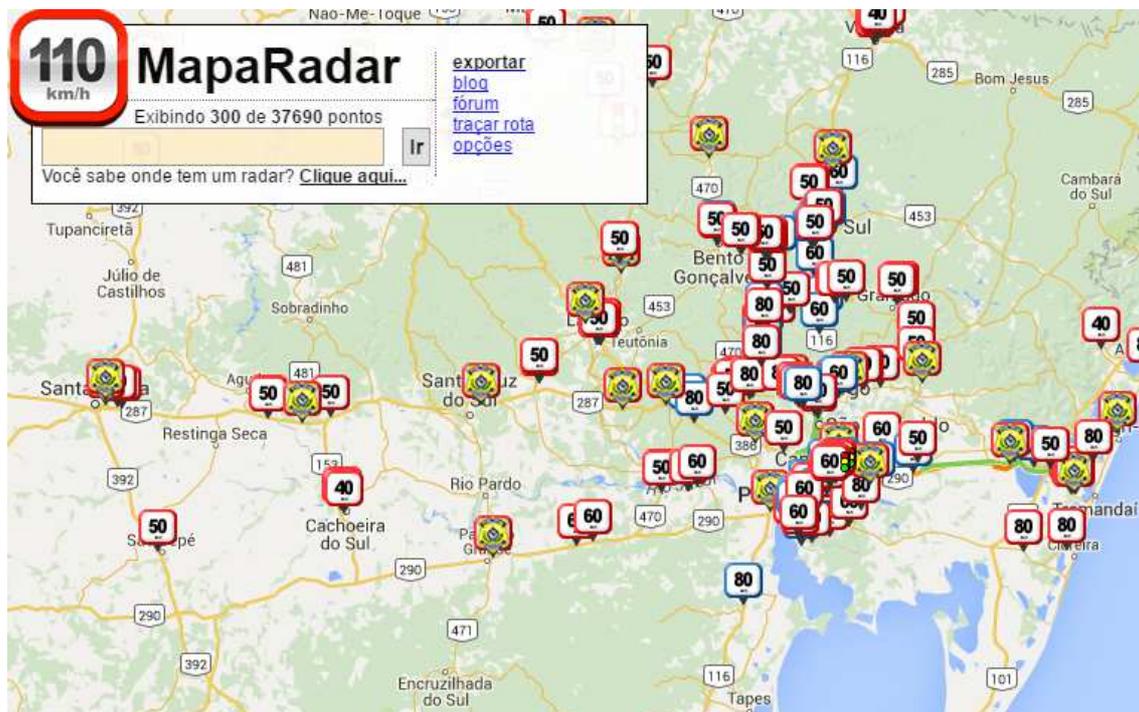
Por fim, uma ferramenta que é voltada para disseminação de dados abertos de governo, dentre outras funcionalidades voltadas para governos, é o Socrata. Trata-se de uma solução baseada na internet, que permite que organizações ligadas ao governo possam disponibilizar seus dados em um formato de catálogo centralizado, e também que os consumidores de dados possam encontrá-los. Como uma de suas facilidades, possui uma API para consumo dos dados (SOCRATA, 2015).

4.2 Ferramentas colaborativas com dados abertos

Alguns aplicativos colaborativos disponibilizam seus dados para que possam ser consultados pela população em geral. Uma ferramenta bastante conhecida é o *Twitter*, que disponibiliza seus dados através de uma API que pode ser acessada por *softwares*, disponibilizando o conteúdo das publicações realizadas nos últimos sete dias pelos usuários (TWITTER, 2016). Segundo Barbosa (2015) cada publicação feita por um usuário no *Twitter* armazena um código identificador da publicação, a identificação do usuário, o texto postado, o idioma, o local (cidade), as coordenadas geográficas de quem postou e a hora da criação da postagem, sendo possível consultar todos esses dados na API. Em um trabalho do autor, ele cria um programa que percorre postagens de terminados usuários, procurando por palavras chave como “acidente” ou “colisão”, no campo de texto, e usa os atributos da postagem para identificar locais onde possivelmente ocorreram acidentes de trânsito.

Nessa mesma linha, temos a ferramenta Mapa Radar. Esta consiste em um mapa online, por meio do qual é possível visualizar uma marca georreferenciada em todos os locais onde há um radar de trânsito, conforme exibido na Figura 12. Clicando sobre a marca é possível visualizar informações detalhadas, como a latitude e a longitude, velocidade máxima permitida no radar e a data de criação do registro. A inserção dos radares é feita de forma colaborativa pelos usuários do *site*, porém deve ser solicitada em um fórum que está disponível no próprio *website*. Em um link, é possível fazer o *download* de todos os dados de radares marcados do *site*, sendo necessário somente cadastrar-se antes (RADAR, 2016).

Figura 12: Mapa com dados do *site* Mapa Radar



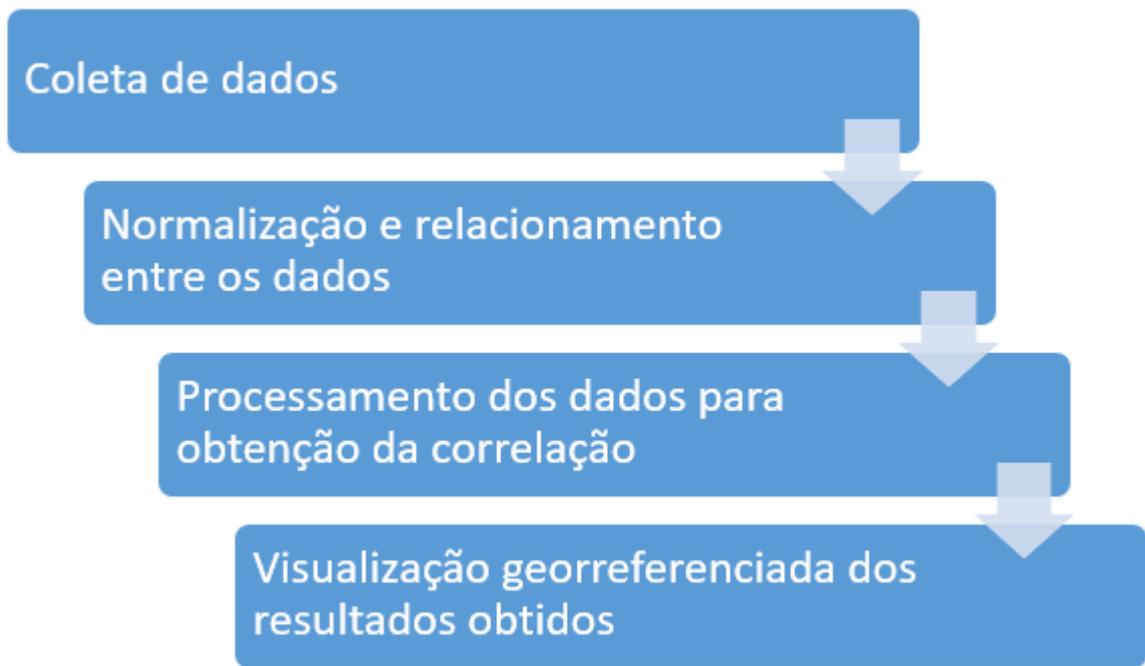
Fonte: (RADAR, 2016).

O *site* Chega de FiuFiu serve para mapear lugares perigosos para mulheres no Brasil. É uma ferramenta por meio da qual mulheres que tenham sofrido assédio ou algum tipo de violência, ou pessoas que saibam de alguém que tenha passado por isso, podem marcar o local onde tal fato aconteceu e descrever em detalhes o ocorrido. A página exibe um mapa que permite a consulta por cidade, e são exibidas marcas georreferenciadas nos locais marcados pelos usuários. Clicando sobre essas marcas, é possível ver os detalhes enviados pelo usuário. Ademais, todas as informações enviadas pelos usuários para o *site* estão disponíveis para download em formato de planilha eletrônica, contendo inclusive a latitude e a longitude da ocorrência (OLGA, 2016).

5 MÉTODO

Neste capítulo é detalhado o método proposto neste trabalho. Para isso, são apresentadas informações relativas a cada etapa, sendo elas, coleta de dados, normalização e relacionamento entre os dados, processamento da correlação dos dados através da técnica PCA (capítulo 2, seção 2.4) e a visualização georreferenciada dos resultados obtidos. A Figura 13 apresenta um diagrama que ilustra cada uma das etapas do método proposto no presente estudo.

Figura 13: Diagrama que representa as etapas do método proposto no estudo



Fonte: Elaborado pelo autor.

5.1 Coleta de dados e definição das informações utilizadas

Como caso de uso para aplicação do método, decidiu-se analisar dados e fatos que possam estar ligados a ocorrências de acidentes de trânsito da cidade de Porto Alegre, capital do Estado do Rio Grande do Sul, com população estimada de 1.476.867 habitantes e área territorial de 496,6 km² em 2015 (IBGE, 2016).

Inicialmente, foram exploradas possíveis fontes de dados e coletados quaisquer dados que pudessem estar relacionados aos acidentes de trânsito na cidade de Porto Alegre. Foram priorizados dados mais recentes, preferencialmente a partir do ano de 2010, evitando, assim, trabalhar com indicadores defasados, tendo em vista as alterações que comumente ocorrem no trânsito das cidades. Ao proceder desta maneira, buscou-se garantir que não tenham sido alteradas as principais variáveis que influenciam nos acidentes de trânsito. Na Tabela 1 é possível verificar um resumo dos dados que foram coletados e o ano para o qual estão disponíveis.

Tabela 1: Dados disponíveis para análise

Dados	2010	2011	2012	2013	2014	2015	2016
Acidente de trânsito	X	X	X	X	X	X	
Precipitação pluviométrica	X	X	X	X	X	X	P
Radar							A
Dia da semana / hora	X	X	X	X	X	X	X
Abordagens balada segura	X	X	X	X	X	X	P
Buracos				P	X	X	

X: disponível

A: atualizado em

P: parcial

Analisando os dados obtidos, chegou-se à conclusão de que um estudo levando em conta dados do ano de 2015 é o ideal, visto que é um período recente, onde há disponibilidade de todos os dados, já que alguns não se encontram disponíveis em outros períodos. No caso dos radares, considera-se que, como os dados foram atualizados no início de 2016, pode-se inferir que, em 2015, estavam com configurações (quantidade e localização) semelhantes ou iguais.

Esses dados foram processados para extração e geração de novas informações, as quais foram utilizadas para a análise realizada neste trabalho. Mais detalhes do processamento e das informações utilizadas são encontrados nas próximas seções.

No que diz respeito às informações, a base de dados com a listagem de todos os acidentes de trânsito ocorridos em Porto Alegre foi obtida em formato de planilha eletrônica, que apresentava um total de 19.183 acidentes no ano de 2015. Essa base possui diversos dados a respeito dos acidentes, como a localização (endereço, latitude e longitude), tipo de acidente, data e horário (informação que irá gerar um dado a ser utilizado na análise), número de feridos, número de mortes, tipo de veículo envolvido (carro, moto, ônibus, táxi, bicicleta, entre outros) e se chovia ou não. Essa base de dados foi obtida no portal Data POA, fornecida pela Empresa Pública de Transportes e Circulação (EPTC) (DATAPOA, 2016).

A informação de precipitação pluviométrica (quantidade de chuva) foi obtida a partir do *site* do Centro Integrado de Comando (CEIC) da cidade de Porto Alegre, que disponibiliza os dados históricos do ano de 2008 até 2016 (CEIC, 2016), que precisaram ser extraídos mês a mês, e informam a quantidade de chuva em milímetros em cada uma das 21 estações meteorológicas disponíveis, para cada dia do mês em questão.

As informações dos radares de trânsito de todo o Brasil puderam ser extraídas do *site* colaborativo Mapa Radar, onde está disponível uma planilha eletrônica com a localização (latitude e longitude) do radar, a velocidade máxima e a data de criação do registro do equipamento. Neste trabalho, trata-se como radar qualquer equipamento fixo que sirva para medir a velocidade dos veículos em uma via, estando ele sozinho ou acompanhado de um semáforo (RADAR, 2016).

Na mesma base de dados em que consta a localização geográfica de cada acidente de trânsito, obtida no portal Data POA e fornecida pela EPTC (DATAPOA, 2016), consta a data e hora da ocorrência de cada um dos acidentes, informação esta, que também será utilizada para a análise realizada neste trabalho.

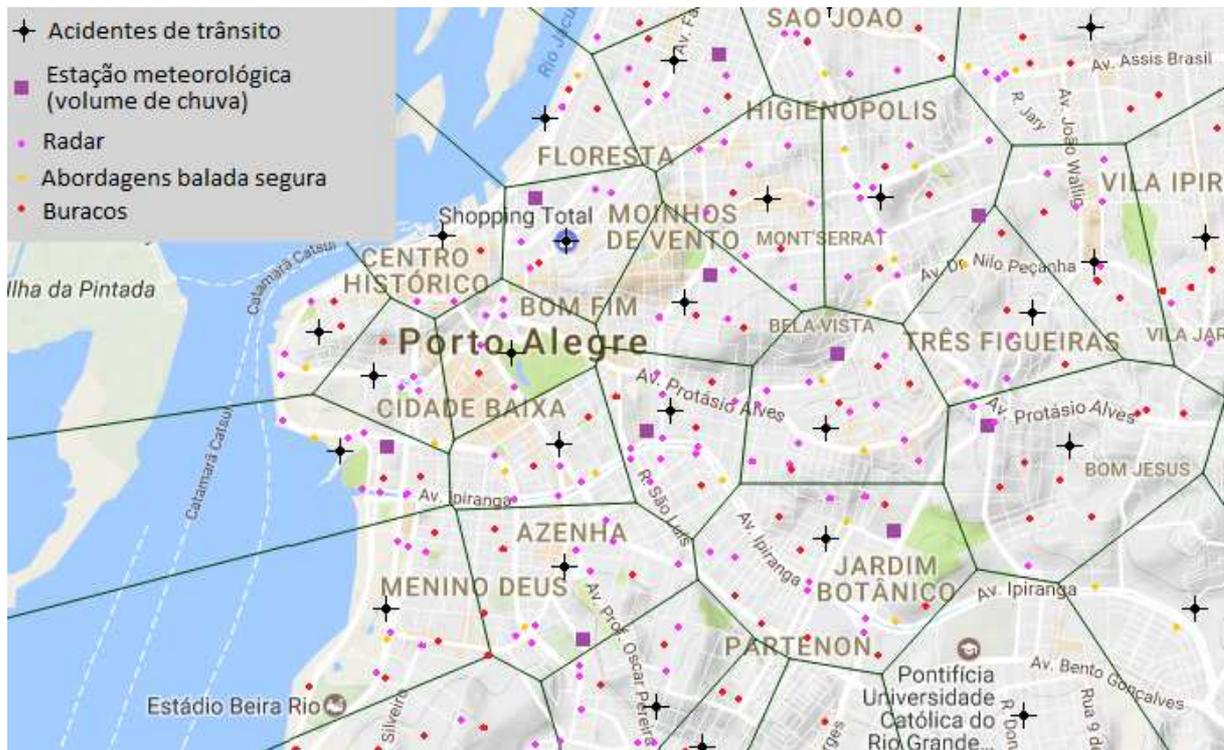
Por sua vez, a Balada Segura é uma política pública do Estado do Rio Grande do Sul para segurança no trânsito, que reúne parcerias a fim de salvar vidas no trânsito. Tem como objetivo ações de educação e fiscalização, em especial relacionadas ao combate ao consumo de álcool no trânsito, em locais onde há maior número de acidentes. São realizadas blitzes com policiais em locais de grande concentração de bares e casas noturnas, e em roteiros normalmente utilizados para deslocamentos até festas e eventos (BALADA, 2016). Foi possível extrair as informações dos locais de realização de blitz da Balada Segura de todo o Estado do Rio Grande do Sul no portal Dados RS, onde consta o endereço da blitz (município, logradouro e número), entre outras informações. Há dados do ano de 2011 até o ano de 2016 (a informação é atualizada diariamente), fornecidos pelo DETRAN-RS (DADOS.RS, 2016).

Igualmente útil, Buracos Monitor é uma ferramenta colaborativa que possui *site* e aplicativo para *smartphones*, por meio do qual é possível que a população informe locais onde há buracos nas vias. Pode-se visualizar um mapa que contém marcadores georreferenciados nos locais onde os usuários reportaram buracos. Clicando sobre os marcadores é possível ver a descrição informada e ainda consultar se outras pessoas confirmaram a existência ou não do buraco (BURACOS, 2016). Essa ferramenta não disponibiliza seus dados para *download* através do *site*, porém por meio de uma área específica, foi possível o envio de uma mensagem para os responsáveis, que posteriormente disponibilizaram os dados exclusivamente para a execução deste trabalho. Foram fornecidos os dados dos buracos marcados do ano de 2014 até 2016, contendo identificação do buraco, localização geográfica (latitude e longitude), data e hora de marcação e a quantidade de pessoas que confirmaram ou anularam a existência do buraco.

Nesse sentido, a Figura 14 exibe um mapa em que uma base de dados de exemplo está plotada de forma georreferenciada sobre o mapa da cidade de Porto Alegre. Os dados são fictícios, pois a base de dados real contém muitos indicadores, o que inviabiliza sua visualização (um dos problemas que este trabalho visa resolver).

No mapa, constam todos os tipos de dados que estarão presentes na análise deste trabalho: locais dos acidentes de trânsito (a área do seu polígono do Diagrama de Voronoi e a data/hora de ocorrência), locais das estações meteorológicas que contêm as informações de precipitação pluviométrica, locais dos radares, locais onde ocorreram blitzes da balada segura, e por fim, locais onde há buracos nas vias.

Figura 14: Exemplos de dados georreferenciados utilizados.



Fonte: Elaborado pelo autor.

5.2 Normalização e relacionamento dos dados

Como o objetivo deste trabalho é analisar o conjunto de dados e identificar o quanto cada uma das características influencia na concentração dos acidentes de trânsito, utilizando a técnica PCA, necessita-se de uma base de dados composta por amostras e diversas características de cada uma das amostras. Cada amostra é relacionada a um acidente de trânsito e as características desses acidentes são os outros dados já citados na última seção: área que representa a concentração de acidentes, quantidade de chuva, radares, dia/hora do acidente, abordagem da Balada Segura e buracos nas vias.

Além disso, para que essa base de dados que será criada possa ser submetida à PCA, é preciso que ela esteja em formato de tabela (planilha eletrônica), de modo que cada linha representará uma amostra (também chamada de batelada em algumas bibliografias relacionadas à PCA) e cada coluna representará uma característica dessa amostra. Também é necessário que os valores de cada uma das características de cada acidente sejam numéricos.

Assim, para chegar a uma base de dados que obedeça às características explicadas até o momento neste capítulo, foi necessário tratar, normalizar e relacionar os dados. Para isso, todas as informações foram carregadas para um Banco de Dados Relacional Oracle, que por meio da linguagem de programação PL/SQL permite manipulação dos dados e implementação de estruturas de programação, como repetições e testes, além da utilização de *web services*, como o da API do Google Maps, a qual foi utilizada neste trabalho.

Além disso, como alguns dados utilizados não possuíam latitude e longitude e sim o endereço por extenso (nome da rua, número do imóvel, bairro, cidade, etc), foi utilizado em

todos os casos o método de Geocodificação da API do Google Maps para obtenção da latitude e longitude. Da mesma forma, para alguns dados em que era conhecida a latitude e longitude - porém necessitava-se saber o endereço -, foi utilizado o método de Geocodificação Reversa da mesma API. A API do Google Maps foi acessada utilizando *web service*, e as instruções de como utilizar os métodos foram obtidas da documentação oficial (GOOGLE, 2016).

Nas próximas seções, para cada um dos dados utilizados na análise, será explicado como este foi manipulado, tratado, normalizado e relacionado aos demais.

5.2.1 Acidentes de trânsito

Com a informação, por meio da latitude e longitude, dos pontos exatos onde ocorreram os acidentes de trânsito, foi criado um Diagrama de Voronoi, no qual cada ponto gera um polígono. Para cada polígono foi medida a quantidade de metros quadrados que ele representa em escala real, o que está diretamente ligado à concentração dos acidentes de trânsito. Essa é a informação que se deseja explicar neste trabalho, a correlação (influência) dos demais dados coletados, detalhados nesta seção, com a concentração de acidentes de trânsito.

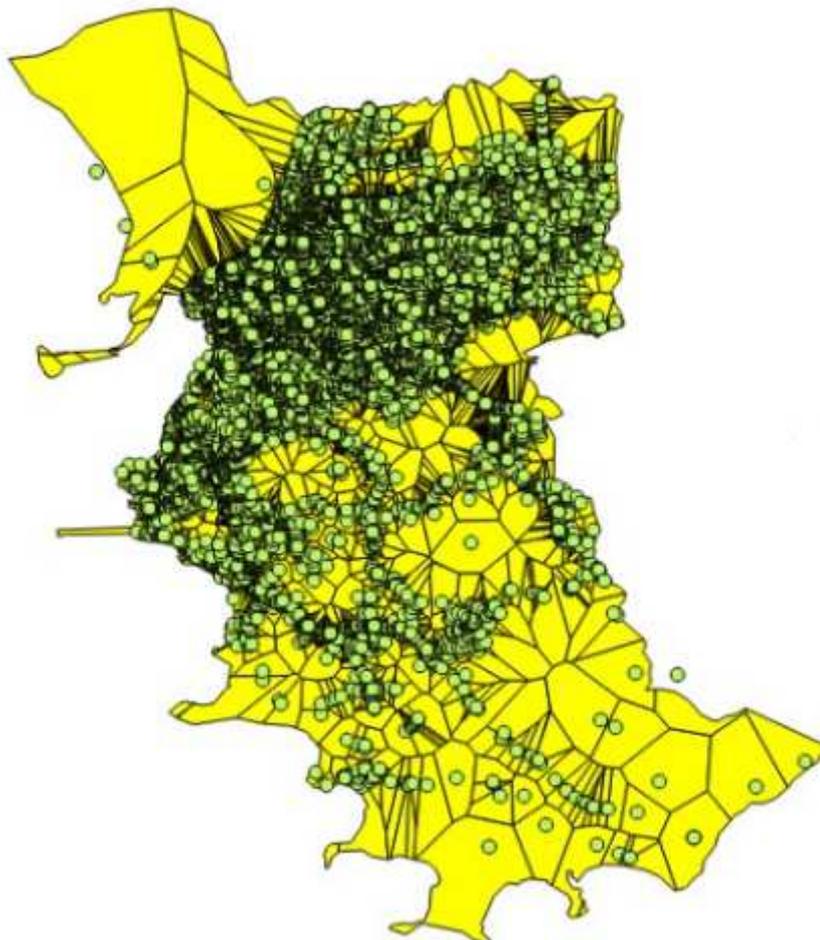
De fato, a planilha contendo os dados de acidentes de trânsito possui diversas informações a respeito de cada acidente. Contudo, como o foco do trabalho não é avaliar especificamente as características do acidente, mas as características urbanas envolvidas, informações como tipo do acidente, características do veículo, condutores e demais envolvidos, foram descartadas. Foi carregado para o banco de dados somente o código identificador do acidente, data e hora de ocorrência e latitude e longitude do local onde ocorreu o acidente.

Em posse dessas informações, criou-se um Diagrama de Voronoi, no qual os pontos em que ocorreu cada um dos acidentes foram utilizados como os pontos geradores para cada nodo do diagrama. Este foi criado utilizando o *software* QGIS (QGIS, 2016), e os dados foram submetidos ao *software* em formato de planilha eletrônica. Como área delimitadora do Diagrama de Voronoi foi utilizada a área da cidade de Porto Alegre, encontrada em formato KMZ (um formato de arquivos para representação de dados georreferenciados, que também é suportado pelo QGIS), obtida no *site* Gmapas, onde a fonte consta como sendo o IBGE (GMAPAS, 2016).

Com o Diagrama de Voronoi pronto, foi possível calcular a área de cada um dos polígonos do diagrama no próprio *software*. Essa informação foi extraída para formato de planilha eletrônica e carregada para o banco de dados de forma que pudesse ser relacionada aos acidentes de trânsito e demais dados. Para os locais onde há mais de um acidente de trânsito exatamente no mesmo local (comum em cruzamentos), o Diagrama de Voronoi foi gerado considerando que todos os pontos no mesmo local são um único ponto. Entretanto, após o cálculo da área de cada nodo do Diagrama de Voronoi, a área do nodo formado a partir de um desses casos, foi dividida pelo número de pontos e o resultado atribuído como área a cada um desses pontos.

Na Figura 15 é possível ver o Diagrama de Voronoi criado pelo QGIS com base na localização geográfica de cada um dos acidentes, que estão representados com pontos verdes.

Figura 15: Digrama de Voronoi formado com a localização dos acidentes



Fonte: Elaborado pelo autor.

5.2.2 Precipitação pluviométrica

As planilhas eletrônicas que contêm o volume de chuva por dia em cada uma das estações meteorológicas de Porto Alegre, e que estão divididas em arquivos separados por mês, foram carregadas para o banco de dados com a seguinte estrutura: data da medição do volume de chuva, identificador da estação meteorológica da medição e volume de chuva em milímetros. Adicionalmente, em uma estrutura separada, foi carregada para o banco de dados, a localização (latitude e longitude) de cada uma das estações meteorológicas, informação que está disponível no *site* do CEIC.

Por sua vez, este dado foi relacionado ao acidente de trânsito da seguinte maneira: foram percorridos os acidentes de trânsito, um a um, e a partir da sua localização geográfica foi identificada a estação meteorológica mais próxima. Para esta, foi buscado o valor de volume de chuva na data do acidente, e esse dado foi relacionado ao acidente. Como algumas estações meteorológicas não possuem dado de medição de chuva em alguns dias específicos (faltam 14,8% das 7.665 medições que deveriam existir, para 21 estações meteorológicas e

365 dias no ano), buscou-se o volume de chuva da segunda estação meteorológica mais próxima que tenha medição de volume de chuva na data em questão.

5.2.3 Radares

As planilhas que contêm latitude e longitude dos radares, e que estão separadas em arquivos com informações de radar fixo, radar acoplado a semáforo e semáforo com câmera, foram todas carregadas para o banco de dados. Como se têm dados de radares de todo o Brasil, em um primeiro momento foram percorridos todos eles, e, por meio do método de Geocodificação Reversa da API do Google Maps, foi obtido o endereço de cada um e descartados os radares que não pertencem à cidade de Porto Alegre.

Inicialmente, o objetivo era medir a distância até o radar mais próximo. Porém dessa forma, por exemplo, dois acidentes ocorridos um em uma via com somente um radar a dez metros do acidente, e outro em uma via com muitos radares, sendo o mais próximo também a dez metros do acidente, teriam o mesmo valor no que se refere à distância até o radar mais próximo. Então verificou-se que a informação dos locais com radar poderia ser utilizada de melhor maneira, e decidiu-se obter três tipos de informação:

- (i) Quantidade de radares em determinado raio: foi contada a quantidade de radares encontrados em um raio de 100 metros ao redor do acidente.
- (ii) Raio necessário para englobar determinada quantidade: foi identificado o raio necessário (em metros) ao redor do acidente para englobar cinco radares. O raio mínimo é de 500 metros e foi incrementado de 100 em 100 metros, com valor máximo de 3.000 metros.
- (iii) Distância até o radar mais próximo: foi medida a distância (em metros) até o radar mais próximo do acidente.

5.2.4 Dia da semana / hora

A intenção de utilizar essa informação, é identificar se dia da semana e hora influenciam na ocorrência dos acidentes. Desse modo, pode-se confirmar, por exemplo, se, em um determinado dia da semana e horário, como uma segunda-feira pela manhã, quando grande parte da população está indo ao trabalho, ocorre a maioria dos acidentes, ou se o dia da semana/hora não influencia de maneira alguma.

Essa informação está na mesma planilha eletrônica que contém os dados dos locais dos acidentes de trânsito, portanto, já estava no banco de dados e relaciona-se diretamente com cada um dos acidentes de trânsito, por isso não foi necessária uma lógica específica para vínculo entre os dados.

Como havia a necessidade de que todos os dados utilizados fossem numéricos, para que pudessem ser submetidos à PCA, foi necessária a normalização do dado, que foi feita da seguinte forma: a cada dia da semana foi atribuído um valor numérico de 1 a 7 (1=segunda-feira, 2=terça-feira, ..., 6=sábado, 7=domingo), e concatenado a esse número, foram usados a hora e minuto do acidente, cada um com dois dígitos. Dessa forma, foi possível converter a informação em um valor numérico, que mantém uma lógica, onde segunda-feira pela manhã é

representada por um número menor, aumentando o número ao longo de cada dia da semana/hora, até chegar ao domingo à noite com um valor bem maior. Exemplos de normalização do dado são exibidos na Tabela 2.

Tabela 2: Exemplo de normalização do dia da semana/hora do acidente.

Valor Original	Valor Normalizado
Segunda-feira às 08:30h	10830
Quarta-feira às 17:37h	31737
Sexta-feira às 12:31h	51231
Domingo às 21:55h	72155

5.2.5 Abordagem balada segura

Da planilha eletrônica que contém os dados das operações de Balada Segura, foram carregados para o banco de dados somente data, cidade, nome da rua e número do imóvel. Como a planilha contém dados das blitzes realizadas em todo o Rio Grande do Sul, foram descartados os registros de cidades diferentes de Porto Alegre. Posteriormente, percorreu-se cada um dos registros restantes, para fazer a conversão do endereço em latitude e longitude, com utilização do método de Geocodificação da API do Google Maps.

Inicialmente, o objetivo era medir a distância até a blitz de Balada Segura mais próxima. No entanto, desta forma, por exemplo, dois acidentes ocorridos um em uma via com somente uma blitz a dez metros do acidente e outro, em uma via com mais blitzes realizadas, sendo a mais próxima também a dez metros do acidente, teriam o mesmo valor no que se refere à distância até a blitz mais próxima. Identificou-se que a informação dos locais com Balada Segura poderia ser utilizada de melhor maneira e então se decidiu obter três tipos de informação:

- (i) Quantidade de blitzes em determinado raio: foi contada a quantidade de blitzes encontradas em um raio de 100 metros ao redor do acidente.
- (ii) Raio necessário para englobar determinada quantidade: foi identificado o raio necessário (em metros) ao redor do acidente para englobar cinco blitzes. O raio mínimo é de 500 metros, e foi incrementado de 100 em 100 metros, com valor máximo de 3.000 metros.
- (iii) Distância até a blitz mais próxima: foi medida a distância (em metros) até a blitz mais próxima do acidente.

Para todas as medições, foram usadas informações das operações realizadas desde 30 dias antes até o próprio dia do acidente. Como as blitzes são realizadas em locais de grande concentração de festas e eventos ou em vias normalmente usadas para deslocamentos até esses lugares (BALADA, 2016), chegou-se à conclusão de que mesmo uma blitz já realizada há alguns dias, poderia influenciar na ocorrência ou não de acidentes de trânsito, devido ao fato de condutores evitarem passar pelo local, por exemplo.

5.2.6 Buracos nas vias

A base de dados fornecida pela equipe do Buracos Monitor, em formato JSON, foi processada e as informações inseridas no banco de dados: identificador, latitude, longitude e a data de inserção do buraco no sistema.

Em um primeiro momento, o objetivo era medir a distância até o buraco mais próximo. Entretanto, desta forma, por exemplo, dois acidentes ocorridos um em uma via com somente um buraco a dez metros do acidente, e outro em uma via com muitos buracos, sendo o mais próximo também a 10 metros do acidente, teriam o mesmo valor no que se refere à distância até o buraco mais próximo. Assim, identificou-se que a informação dos locais dos buracos nas vias poderia ser utilizada de melhor maneira, e decidiu-se obter três tipos de informação:

- (i) Quantidade de buracos em determinado raio: foi contada a quantidade de buracos encontrados em um raio de 100 metros ao redor do acidente.
- (ii) Raio necessário para englobar determinada quantidade: foi identificado o raio necessário (em metros) ao redor do acidente para englobar cinco buracos. O raio mínimo é de 500 metros, e foi incrementado de 100 em 100 metros, com valor máximo de 3.000 metros.
- (iii) Distância até o buraco mais próximo: foi medida a distância (em metros) até o buraco mais próximo do acidente.

Para todas as medições foram considerados buracos que tenham sido reportados na ferramenta Buracos Monitor nos 30 dias que antecederam a data do acidente. Esse critério foi utilizado, tendo em vista que normalmente dentro de algumas semanas os buracos são reparados pelos órgãos responsáveis.

5.3 Processamento dos dados (PCA)

Ao fim da normalização e relacionamento dos dados, foi obtida uma matriz de amostras (indivíduos), na qual cada um é referente a um acidente de trânsito, com diversas variáveis (características) cada, que se referem aos tipos de dados detalhados anteriormente (capítulo 5, seção 5.2). Essa matriz foi extraída do banco de dados em formato de planilha eletrônica, e parte de seus dados estão exemplificados na Tabela 3.

Tabela 3: Matriz de dados com exemplo de indivíduos e variáveis

Indivíduos	Variáveis					
	Chuva	Radar*	Momento Semana	Balada Segura*	Buraco*	...
0,002489	6,2	0	62045	7400	0	...
0,001141	0	0	71640	6900	0	...
0,000939	0	0	11605	4400	10	...
0,000672	0	0	11800	9999	0	...

* Quantidade no raio

Após execução da PCA, foram encontradas as componentes principais, que foram utilizadas para calcular a correlação da variável que se quer explicar (área dos acidentes de trânsito), em função das outras variáveis. Para isso, foi utilizada a Equação 1:

$$\alpha_v(k) = \frac{\frac{n_v}{n_k}}{\sum_{\substack{v=1 \\ v \neq k}}^V \frac{n_v}{n_k}} = \frac{n_v}{\sum_{\substack{v=1 \\ v \neq k}}^V n_v} \quad (1)$$

Onde n é a componente principal utilizada e a é a correlação da variável k em função das v variáveis.

5.4 Visualização georreferenciada dos resultados

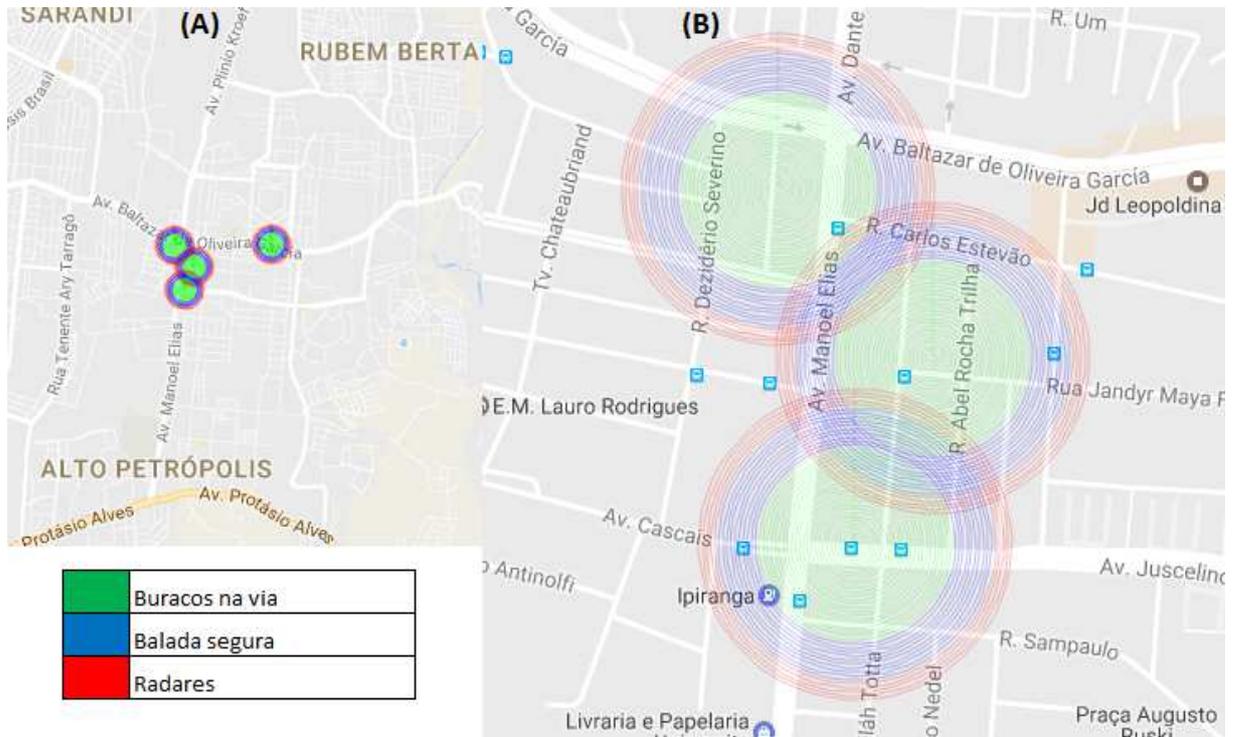
Um dos objetivos do estudo é elaborar uma forma de visualização dos resultados da análise. Buscou-se que essa visualização seja útil para tomadores de decisão e população em geral, de forma intuitiva, apresentando todas as informações possíveis em um único mapa, sem necessidade de ler grandes quantidades de dados. Nessa perspectiva, foram criados com os resultados, alguns protótipos de visualização dos dados, para avaliar aquela considerada a mais adequada.

A visualização mais interessante, explora conceitos dos mapas de calor, com áreas pintadas de forma mais intensa em locais onde há maior concentração de acidente, e além disso, explora a utilização de diferentes cores para pintar o mapa, de maneira que cada cor representa uma das variáveis de influência na concentração dos acidentes, e as cores que mais aparecem indicam que a variável correspondente influencia de forma mais intensa.

Na Figura 16 pode ser observado um exemplo dessa visualização somente com quatro acidentes, com dois níveis de zoom diferentes, onde (A) está mais distante e (B) mais próximo. Optou-se por utilizar somente as três variáveis principais (as de maior influência) na visualização, porque em testes preliminares a visualização com 11 cores diferentes - cada uma representando uma variável analisada - ficou muito confusa.

A visualização funciona da seguinte forma, a partir do local onde ocorreu cada um dos acidentes é criada uma série de círculos, cada um com o raio maior que o anterior. Os primeiros círculos (mais próximos do local do acidente) são da cor da variável que representa a característica que menos influencia na ocorrência dos acidentes de trânsito. A quantidade de círculos representa o percentual que essa variável influencia (exemplo: 18% = 18 círculos). Após os círculos da variável que menos influencia, vêm os círculos da segunda variável que menos influencia, em sua cor e quantidade correspondentes, e assim sucessivamente.

Figura 16: Exemplo de visualização do resultado da análise



Fonte: Elaborado pelo autor.

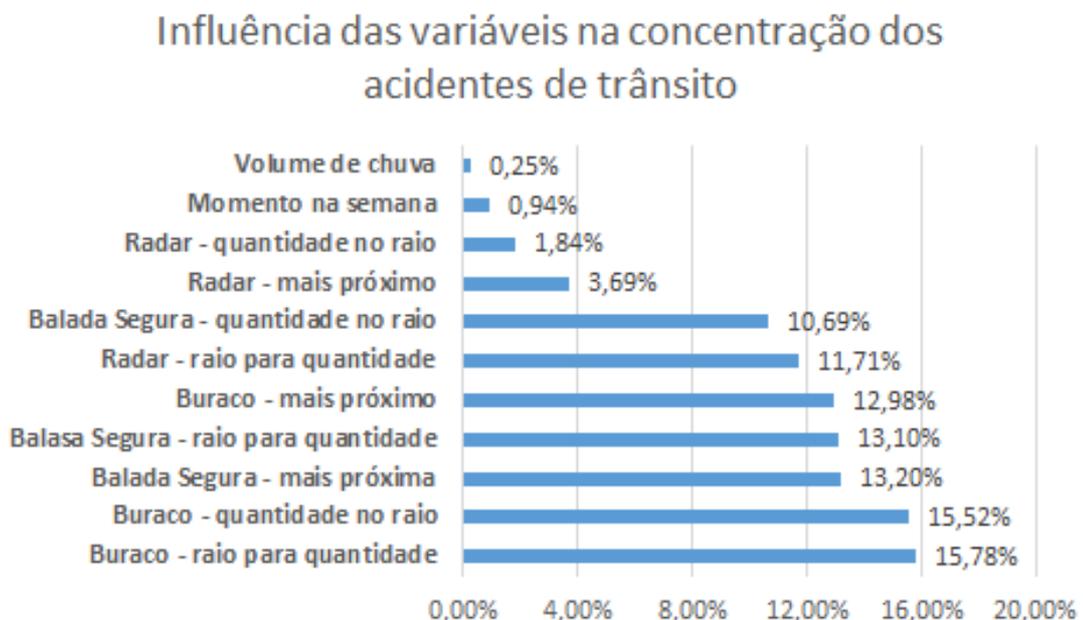
Nesse exemplo todos os acidentes têm a mesma configuração de quantidade e cor dos círculos, visto que todos pertencem a um mesmo conjunto de dados analisados. Em outros casos, poderiam ser selecionados diversos grupos de acidente para análise, de acordo com sua localização geográfica, e cada grupo de acidentes teria uma configuração diferente de quantidade e cor dos círculos, de acordo com o resultado da análise para aquele grupo.

6 RESULTADOS

O método proposto foi aplicado em três diferentes conjuntos de dados referentes ao ano de 2015: o primeiro conjunto abrange dados de todos os acidentes de trânsito da cidade de Porto Alegre no período; o segundo e terceiro conjunto, a título de exemplo, abrangem dados de acidentes no entorno de dois importantes hospitais de Porto Alegre.

Na Figura 17, pode-se observar o resultado da análise para todas as variáveis, considerando uma análise executada com dados de todos os acidentes de trânsito coletados neste trabalho. Como resultado, obteve-se o percentual que cada uma das variáveis tem de influência sobre a área dos acidentes (concentração de acidentes).

Figura 17: Percentual de influência das variáveis estudadas nos acidentes de trânsito



Fonte: Elaborado pelo autor.

Desta forma, foram apresentados os resultados da pesquisa, exibindo quais variáveis mais influenciam na concentração de acidentes de trânsito, levando em conta todos os acidentes de trânsito de Porto Alegre no período de interesse. Porém, um dos objetivos deste trabalho é propor uma metodologia de análise e visualização diferenciada, para que, dependendo das necessidades de quem está explorando os dados, a análise e exibição sejam feitas para regiões específicas da cidade, que podem apresentar características diferentes e, conseqüentemente, comportamentos diferentes com relação ao trânsito.

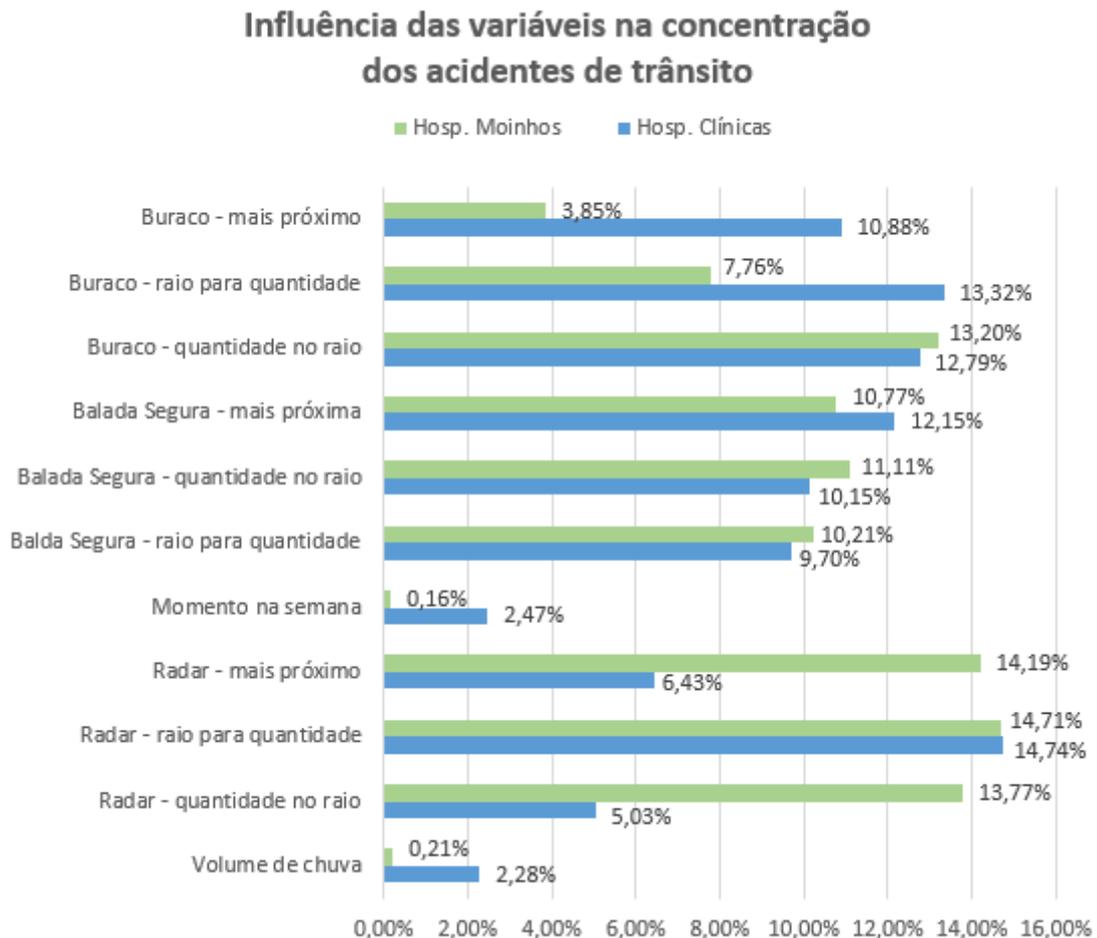
Como exemplo, foram exploradas a análise e a visualização das áreas em torno de dois dos maiores hospitais da cidade de Porto Alegre (Hospital das Clínicas e Hospital Moinhos de Vento). Por serem áreas de grande circulação de pessoas, provenientes tanto da própria cidade, quanto de outras, em busca de atendimento médico, é interessante averiguar quais variáveis estão influenciando na concentração dos acidentes de trânsito nesses locais. É importante ressaltar que os hospitais foram utilizados somente como exemplo, e que utilizando a mesma metodologia poderiam ter sido exploradas áreas em torno de pontos turísticos (informação que poderia ser útil para órgãos ligados ao turismo, por exemplo),

escolas, órgãos públicos, empresas, locais de realização de eventos ou até mesmo localizações genéricas, como o centro da cidade - desde que se tenha conhecimento da localização geográfica do local que se queira estudar.

Optou-se por analisar os dados considerando um raio de um quilômetro ao redor de cada um dos hospitais. Inicialmente, foi obtida a localização de cada uma das instituições e para cada uma delas foram identificados os acidentes que estavam a menos de um quilômetro. Assim foram criadas duas bases de dados independentes, que serão processadas com a PCA e cujas variáveis de influência na concentração de acidentes serão calculadas de forma independente.

Após executar o método de processamento dos dados com PCA, para cada uma das bases de dados, foram obtidos os resultados apresentados na Figura 18, onde consta o percentual de influência das variáveis, na concentração dos acidentes de trânsito de veículos, para cada um dos dois hospitais.

Figura 18: Percentual de influência das variáveis estudadas nos acidentes de trânsito, para cada hospital



Fonte: Elaborado pelo autor.

6.1 Visualização dos resultados

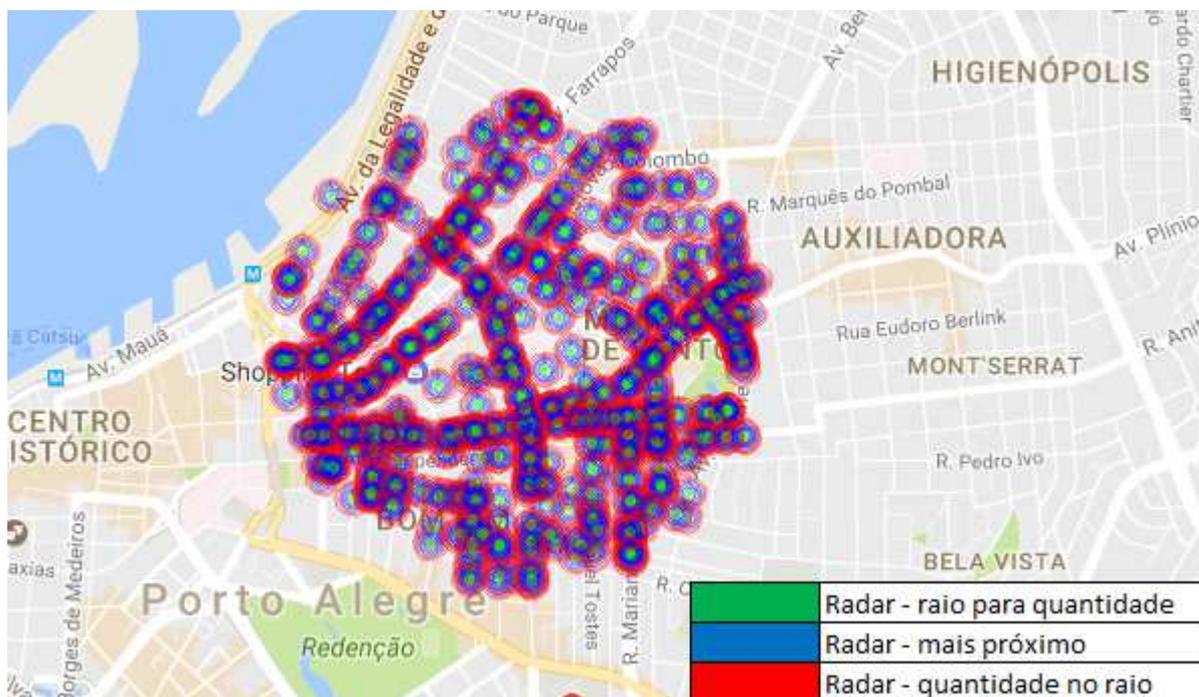
Seguindo a lógica de visualização georreferenciada explicada anteriormente, e com a informação das variáveis que mais influenciam na concentração de acidentes de trânsito nas proximidades de cada um dos hospitais, exibidas na Figura 18, foi criada uma visualização desses dados reais, plotados sobre o mapa da cidade de Porto Alegre, de forma georreferenciada. Na Figura 19 é apresentada a visualização dos resultados para os acidentes próximos ao Hospital das Clínicas, e na Figura 20 é apresentada a visualização dos resultados para os acidentes próximos ao Hospital Moinhos de Vento.

Figura 19: Visualização do resultado da análise para o Hospital das Clínicas



Fonte: Elaborado pelo autor.

Figura 20: Visualização do resultado da análise para o Hospital Moinhos de Vento



Fonte: Elaborado pelo autor.

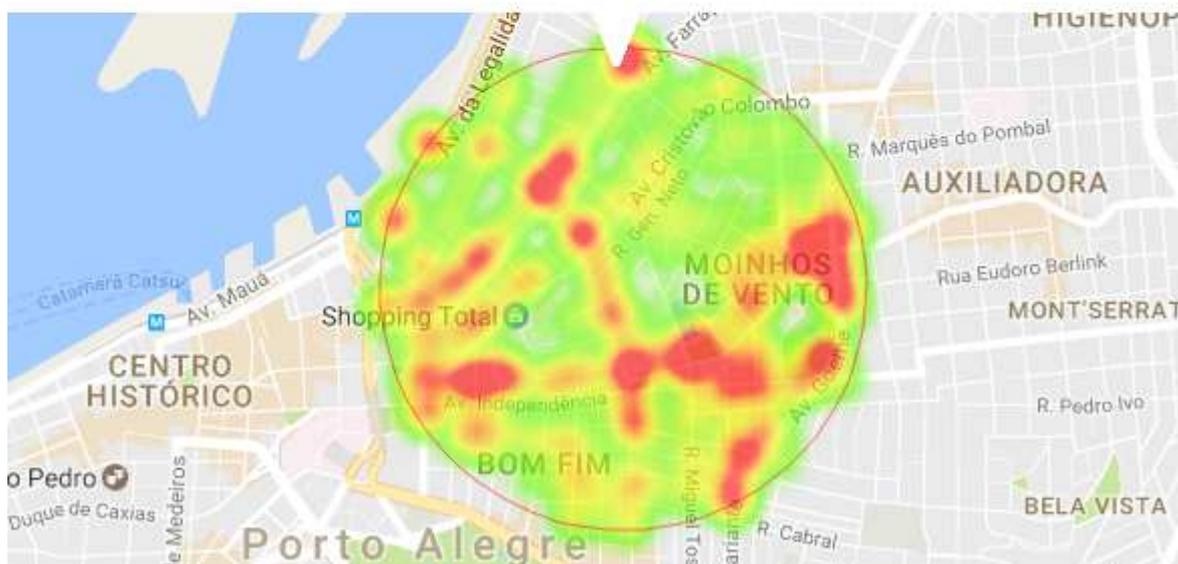
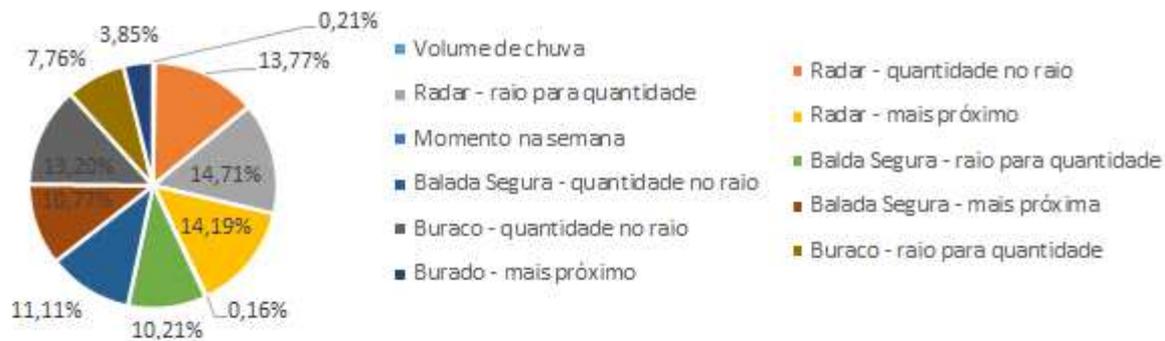
Avaliando essa visualização, encontrou-se facilidade em identificar os locais com maior concentração de acidentes de trânsito, porém houve dificuldade em identificar qual variável mais influencia na ocorrência destes. Por esse motivo foi explorado outro tipo de visualização de dados, que funciona de forma mais simplificada, mas pode facilitar a visualização de cada uma das variáveis de influência, além de possibilitar que todas as 11 variáveis apareçam na visualização. Essa outra possibilidade de visualização consta na Figura 21 para o resultado da análise com relação ao Hospital das Clínicas e, na Figura 22 com relação ao resultado da análise para o Hospital Moinhos de Vento.

Figura 21: Visualização alternativa do resultado da análise para o Hospital das Clínicas



Fonte: Elaborado pelo autor.

Figura 22: Visualização alternativa do resultado da análise para o Hospital Moinhos de vento



Fonte: Elaborado pelo autor.

7 CONCLUSÃO

Por meio do presente trabalho foi possível apresentar uma proposta de metodologia para análise de dados georreferenciados oriundos de diversas fontes e que possam estar ligados aos acidentes de trânsito. Além disso, foi possível exibir uma visualização dos resultados plotados sobre mapas da cidade, além da aplicação de um caso de uso estudando as causas da concentração de acidentes na cidade de Porto Alegre/RS.

No que diz respeito ao primeiro item, foi apresentada a metodologia para agrupamento e análise dos dados provenientes de diversas fontes, o que foi bastante útil, já que possibilita que diversos dados, de fontes independentes, possam ser incorporados a análise conforme a necessidade.

Sobre a visualização dos dados, foram apresentadas duas propostas de visualização. A primeira mais inovadora, apresenta todas as conclusões diretamente no mapa, porém mostrou-se relativamente confusa. A segunda visualização, mais conservadora, continua apresentando todos os resultados em um único mapa, facilitando a leitura e o entendimento dos resultados.

Com efeito, buscou-se desenvolver essa metodologia de análise e visualização de maneira que futuramente pudesse ser construída uma ferramenta *online* para utilização por parte de órgãos e pessoas interessadas. Desse modo, permitir-se-ia o acesso a dados atualizados e realização de análises, selecionando locais da cidade conforme a necessidade, também possibilitando comparar locais diferentes, já que a visualização foi feita com a ferramenta do Google Maps, que possibilita interação com o mapa.

No que se refere ao estudo de caso para os dados da cidade de Porto Alegre, a pesquisa apresenta quais, dentre os dados utilizados na análise, mais influenciam na concentração dos acidentes de trânsito. Quando analisados todos os dados, concluiu-se que não há uma característica da cidade que influencie totalmente na concentração dos acidentes de trânsito, visto que os percentuais ficaram parecidos entre diversas variáveis (sete apresentaram valor entre 10% e 16%). Contudo, foi visível que algumas variáveis influenciam muito pouco (quatro apresentam valores entre 0,25% e 4%), o que pode indicar que os órgãos responsáveis talvez não necessitem se preocupar com tais fatores.

Ao analisar, somente a título de exemplo, as variáveis que influenciam na concentração dos acidentes nas proximidades de dois hospitais da cidade, obtiveram-se valores distintos entre os locais. Isso mostra que, possivelmente em locais diferentes da cidade, diferentes características influenciam mais na concentração dos acidentes de trânsito. Todavia, ao comparar as análises dos dois hospitais, juntamente com a análise dos dados de toda a cidade, pôde-se notar que os fatores “momento na semana” e “volume de chuva” são os que menos influenciam nas três análises.

Esse trabalho possui como principal contribuição, dentre outras, uma nova abordagem computacional que emprega PCA em agrupamentos de dados georreferenciados para identificação dos fatores que influenciam na ocorrência de trânsito, bem como a sua visualização sobre mapas.

7.1 Trabalhos Futuros

Como trabalhos futuros, identificou-se a possibilidade de se pesquisar em fontes de dados à procura de novas variáveis que possam ser incluídas nas análises. Posteriormente, podem-se consultar especialistas em trânsito para decidir quais variáveis novas usar, ou até mesmo para tomar conhecimento de novas fontes de dados que possam ser consultadas.

Outra possibilidades de melhoria, que poderia deixar os resultados da análise mais apurados, seria utilizar uma maneira de compensar os polígonos muito extensos em um único sentido que apareceram no Diagrama de Voronoi, e que talvez tenham feito com que a área de alguns acidentes ficasse com o valor muito alto de forma indevida. Esses polígonos extensos em um único sentido aparecem principalmente em vias nas quais há muitos acidentes, porém, quando ao redor daquela via não aparecem outras vias próximas, assim os polígonos ficam pequenos em um sentido (por causa dos outros acidentes próximos na mesma via) e grandes em outro (já que não há outra via próxima que possibilite a ocorrência de acidentes).

REFERÊNCIAS

ANASTASOPOULOS, Panagiotis Ch. SHANKAR, Venky N. HADDOCK, John E. MANNERING, Fred L. **Accident Analysis and Prevention**, v. 45, p. 110-119, 2012.

ANDRADE, Luciano de. VISSOCI, João Ricardo Nickenig. RODRIGUES, Clarissa Garcia. FINATO, Karen. CARVALHO, Elias. PIETROBON, Ricardo. SOUZA, Eniuce Menezes de. NIHEI, Oscar Kenji. LYNCH, Catherine. CARVALHO, Maria Dalva de Barros. **Brazilian Road Traffic Fatalities: A Spatial and Environmental Analysis**. Plos One. 2014. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0087244>>. Acesso em: 20 dez. 2016.

AURENHAMMER, Franz. **Voronoi Diagrams – A Survey of a Fundamental Geometric Data Structure**. ACM Computing Surveys, v. 23, No. 3, p. 345-405, 1991.

BALADA Segura. **A balada segura**. Disponível em:

<<http://www.baladasegura.rs.gov.br/conteudo/993/a-balada-segura>>. Acesso em: 12 junho 2016.

BARBOSA, Ivanildo. **Avaliação do impacto de acidentes de trânsito no tráfego de vias urbanas a partir de valores de velocidade**. 2015. Tese (Doutorado em Informática). Pontifca Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, 2015.

BERG, Mark de. KREVELD, Marc van. OVERMARS, Mark. SCHWARZKOPF, Otfried. **Computational Geometry – Algorithms and Applications**. Springer: Berlin, 2000.

BONI, Raquel de. BASTOS, Francisco Inácio. WEBER, Eliseu. HASENACK, Heinrich. PECHANSKY, Flavio. Geoprocessamento no estudo da relação entre acidentes de trânsito e bares em Porto Alegre. **Uso de bebidas alcoólicas e outras drogas nas rodovias brasileiras**, p.84-88, 2010.

BORCARD, Daniel. GILLET, François. LEGENDRE, Pierre. **Numerical Ecology with R**. Nova Iorque: Springer. 2011.

BURACOS Monitor. **Buracos Monitor**. Disponível em:

<<http://www.buracosmonitor.com.br/>>. Acesso em: 2 junho 2016.

CEIC – Centro Integrado de Comando. **Dados Históricos das Estações Meteorológicas**. Disponível em: <http://www2.portoalegre.rs.gov.br/ceic/default.php?p_secao=52>. Acesso em: 2 junho 2016.

CHIGNARD, Simon. **A brief history of Open Data**. 2013. Disponível em:

<<http://www.paristechreview.com/2013/03/29/brief-history-open-data/>>. Acesso em: 25 abril 2016.

CKAN. **The open source data portal software**. Disponível em: <<http://ckan.org/>>. Acesso em: 4 junho 2016.

COLAB.RE. **Colab.re**. Disponível em: <<http://www.colab.re/>>. Acesso em: 2 junho 2016.

CONCAR – Comissão Nacional de Cartografia. **Plano de ação para implantação da Infraestrutura Nacional de Dados Espaciais**. Disponível em:

< <http://www.concar.ibge.gov.br/arquivo/planodeacaoINDE.pdf>>. Acesso em 30 abril 2016.

DADOS.GOV.BR. **Portal Brasileiro de Dados Abertos**. Disponível em:

<<http://dados.gov.br/>>. Acesso em: 3 jun. 2016.

DATA.GOV. **The home of the U.S. Government's open data**. 2016. Disponível em:

<<https://www.data.gov/>>. Acesso em: 29 abril 2016.

DATAPOA. **Sobre**. Disponível em: <<http://datapoa.com.br/about>>. Acesso em: 4 junho 2016.

DADOS.RS. **Sobre**. Disponível em: <<http://dados.rs.gov.br/about>>. Acesso em: 3 junho 2016.

DADOS.TCE. **Sobre**. Disponível em: <<http://dados.tce.rs.gov.br/about>>. Acesso em: 3 junho 2016.

DATA PORTALS. **Sobre DataPortals.org**. Disponível em: <<http://dataportals.org/about>>. Acesso em: 3 junho 2016.

DETRAN RS. **Estudos e pesquisas**. Disponível em:

<<http://www.detran.rs.gov.br/lista/799/estudos-e-pesquisas>>. Acesso em: 3 junho 2016.

DEVELOPERS GOOGLE. **Google Maps API – Maps Javascript API**. Disponível em:

< <https://developers.google.com/maps/documentation/javascript/?hl=pt-br>>. Acesso em: 20 junho 2016.

DIRICHLET, Johann Peter Gustav Lejeune. “Über die Reduktion der positive quadratischen Formen mit drei unbestimmten ganzen Zahlen. **Journal für die Reine und Angewandte Mathematik**, v. 40, p. 209-227, 1850.

DRUCK, Suzana. CARVALHO, Marília Sá. CÂMARA, Gilberto. MONTEIRO, Antonio Miguel Monteiro. **Análise espacial de dados geográficos**. 2004. Disponível em:

<<http://www.dpi.inpe.br/gilberto/livro/analise/>>. Acesso em: 30 abril 2016.

EMC. **The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet Things**. 2014. Disponível em: <<http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>>. Acesso em: 20 abril 2016.

ESRI. **ArcGIS Pro – Create Thiessen Polygons**. Disponível em:

< <http://pro.arcgis.com/en/pro-app/tool-reference/analysis/create-thiessen-polygons.htm>>. Acesso em: 10 junho 2016.

GAUSS, Carl Friedrich. Recursion der Untersuchungen über die Eigenschaften der positive ternären quadratische Formen von Ludwig August Seeber. **Journal für die reine und angewandte Mathematik**, v. 20, p. 312-320, 1840.

GMAPAS. Arquivos KML e KMZ com municípios brasileiros. Disponível em:

<<http://www.gmapas.com/poligonos-ibge/municipios-do-brasil>>. Acesso em: 18 dez. 2016.

GOOGLE. **Google Maps geocoding API**. Disponível em:

<<https://developers.google.com/maps/documentation/geocoding/intro?hl=pt-br>>. Acesso em: 5 junho 2016.

GREENBERG, Ira. XU, Dianna. KUMAR, Deepak. **Processing – Creative coding and generative art in processing 2**. Apress, 2013.

HOTELLING, Harold. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, Washington D.C., v.24, p.417-41, 1933.

IBGE – Instituto Brasileiro de Estatística. **Rio Grande do Sul – Porto Alegre**. Disponível em: <<http://cidades.ibge.gov.br/xtras/perfil.php?lang=&codmun=431490&search=rio-grande-do-sul|porto-alegre|infograficos:-informacoes-completas>>. Acesso em: 10 junho 2016.

ISOTANI, Seiji. BITTENCOURT, Ig Ibert. **Dados Abertos Conectados**. 2016. Disponível em: <<http://ceweb.br/livros/dados-abertos-conectados/>>. Acesso em: 20 abril 2016.

JACKSON, J. Edward. **A User's Guide to Principal Component Analysis**. Chichester: John Wiley & Sons, 1991.

JACOBS, G. D. SAYER, I. Road accidents in developing countries. **Accident Analysis and Prevention**, v.15, n.5, p.337-353, 1983.

KAIESKI, Naira. OLIVEIRA, Luiz Paulo Luna de. VILLAMIL, Marta Becker. **Vis-health – exploratory analysis and visualization of dengue cases in Brazil**. IEEE 49th Hawaii International Conference on Systems Sciences, 2016.

KOCH, Tom. DENIKE, Kenneth. **Crediting his critics' concerns: Remaking John Snow's map of Broad Street cholera, 1854**. Social Science & Medicine, Vancouver. 2009. Disponível em: <<http://www.albany.edu/faculty/fboscoe/papers/koch2009.pdf>>. Acesso em: 30 abril 2016.

LOPES, Luis Felipe Dias. **Análise de componentes principais aplicados à confiabilidade de sistemas complexos**. 2001. Tese (Doutorado Engenharia de Produção). Universidade Federal de Santa Catarina, Florianópolis, SC, 2001.

MATHWORKS. **Documentation – pca**. Disponível em:

<<http://www.mathworks.com/help/stats/pca.html>>. Acesso em: 10 junho 2016.

MATHWORKS. **Documentarion – voroonoi**. Disponível em:

<<http://www.mathworks.com/help/matlab/ref/voronoi.html>>. Acesso em: 10 junho 2016.

MEHMANDAR, Mohammadreza. SOORI, Hamid. MEHRABI, Yadolah. **Predicting and analyzing the trend of traffic accidents deaths in Iran in 2014 and 2015**, 2016.

International Journal of Critical Illness and Injury Science. 2016. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4901831/>>. Acesso em: 20 dez. 2016.

MOURA, Ana Paula Mourão. **Geoprocessamento na gestão e planejamento urbano**. Belo Horizonte: Da Aurora. 2003.

OGP – Open Government Partnership. **2015 Annual Report**. 2015. Disponível em: <http://www.opengovpartnership.org/sites/default/files/OGPreport2015_0.pdf>. Acesso em 25 abril 2016.

OKFN – Open Knowledge International. **Open Data Handbook**. 2012. Disponível em: <<http://opendatahandbook.org/>>. Acesso em: 8 abril 2016.

OLGA. **Chega de fiu fiu**. Disponível em: <<http://chegadefiuuiu.com.br/>>. Acesso em: 5 junho 2016.

OMS – Organização Mundial da Saúde. **Relatório global sobre o estado da segurança viária 2015**. Disponível em: <http://www.who.int/violence_injury_prevention/road_safety_status/2015/Summary_GSRRS_2015_POR.pdf>. Acesso em 10 abril 2016.

ONS – Observatório Nacional de Segurança. **Retrato da segurança viária no Brasil**. Disponível em: <<http://iris.onsv.org.br/portaldados/downloads/retrato2014.pdf>>. Acesso em 10 abril 2016.

PEARSON, Karl. On lines and planes of closest fit to systems of points in the space. **Philosophical Magazine**, v. 2, n. 11, p.559-572, 1901.

PEREZ-LLAAMAS, Christian. LOPEZ-BIGAS, Nuria. **Gitools: Analysis and visualization of genomic data using interactive heat-maps**. Plos One. 2011. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0019541>>. Acesso em: 20 junho 2016.

QGIS. **A free and open source geographic information system**. Disponível em: <<http://www.qgis.org/en/site/>>. Acesso em: 12 dez. 2016.

QUEIROZ, Marcelo Pereira. **Análise espacial dos acidentes de trânsito do município de Fortaleza**. 2003. Dissertação (Mestrado em Engenharia de Transportes). Universidade Federal do Ceará, Fortaleza, CE, 2003.

RADAR Mapa. **MapaRadar**. Disponível em: <<http://maparadar.com/#contato/>>. Acesso em: 2 junho 2016.

RAMOS, Cristhiane da Silva. **Visualização cartográfica e cartografia multimídia**. São Paulo: UNESP, 2005.

REDLICH, Luciana Rosa. **Modelagem de eventos de trânsito com base em clipping de grandes massas de dados da Web**. 2013. Dissertação (Mestrado em Informática). Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, 2013.

SANTOS, Luciano dos. **Análise de acidentes de trânsito do município de São Carlos utilizando sistemas de informações geográficas – SIG e ferramentas de estatística espacial**. 2006. Dissertação (Pós-Graduação em Engenharia Urbana). Universidade Federal de São Carlos, São Carlos, SP, 2006.

SCHÜNKE, Luana Carine. OLIVEIRA, Luiz Paulo Luna. VILLAMIL, Marta Becker. **GUGA: a tool for participatory cities.** The Eighteenth IEEE Symposium on Computers and Communications (ISCC 2013), Split, Croácia, July 2013.

SEHABER, Vanessa Ferreira. **Análise estatística multivalorada dos acidentes de trânsito da BR-376 no período entre os anos de 2009 e 2012.** 2013. Dissertação (Pós-Graduação em Métodos Numéricos em Engenharia). Universidade Federal do Paraná, Curitiba, PR, 2013.

SLTI – Secretaria de Logística e Tecnologia da Informação. **Cartilha técnica para publicação de dados abertos no Brasil.** Disponível em: <<http://dados.gov.br/cartilha-publicacao-dados-abertos/>>. Acesso em 25 abril 2016.

SOCRATA. **Socrata Open Data.** Disponível em: <<https://www.socrata.com/products/open-data/>>. Acesso em: 12 junho 2016.

SOUZA, Adriano Mendonça. **Monitoração e ajuste de realimentação em processos produtivos multivariados.** 2000. 215f. Tese (Doutorado em Engenharia de Produção). Universidade Federal de Santa Catarina, Florianópolis, SC, 2000.

SVENNERBERG, Gabriel. **Beginning Google Maps API 3.** 2010. Disponível em: <http://maps.unomaha.edu/Peterson/GEOG8670_Spring15/Google_Maps_v3.pdf>. Acesso em: 30 abril 2016.

TRULIA. **Crime data in New York.** Disponível em: <https://www.trulia.com/real_estate/New_York-New_York/crime/>. Acesso em: 15 dez. 2016.

VICINI, Lorena. **Análise multivariada da teoria à prática.** 2005. 179f. Trabalho de Conclusão do Curso de Especialização (Especialização em Estatística e Modelagem Quantitativa no Departamento de Estatística). Universidade Federal de Santa Maria, Santa Maria, RS, 2005.

VORONOI Georges Feodosevich. **Nouvelles applications des parametres continus à la théorie des forms quadratiques.** Journal für die Reine und Angewandte Mathematik. v 134. p. 198-287, 1908.

WAZE. **Waze.** Disponível em: <<https://www.waze.com/pt-BR/about>>. Acesso em: 5 jun. 2016).

WIKICRIMES. **Mapeando crimes colaborativamente.** Disponível em: <<http://www.wikicrimes.org/main.html>>. Acesso em: 16 dez. 2016).

YALCIN, Guler. **Non-spatial analysis for the road traffic accidents.** Procedia – Social and behavioral sciences. v 92. p. 1033-1038, 2013.