

ESCOLA POLITÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
MESTRADO EM CIÊNCIA DA COMPUTAÇÃO

WAGNER GONÇALVES SIGNORETTI

**MODELO PARA ANÁLISE VISUAL DE DADOS MULTIVARIADOS NO CONTEXTO
GEOESPACIAL**

Porto Alegre
2018

PÓS-GRADUAÇÃO - *STRICTO SENSU*



Pontifícia Universidade Católica
do Rio Grande do Sul

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
ESCOLA POLITÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**MODELO PARA ANÁLISE
VISUAL DE DADOS
MULTIVARIADOS NO
CONTEXTO GEOESPACIAL**

WAGNER GONÇALVES SIGNORETTI

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Ciência da Computação na Pontifícia Universidade Católica do Rio Grande do Sul.

Orientador: Prof. Isabel Harb Manssour

**Porto Alegre
2018**

Ficha Catalográfica

S578m Signoretti, Wagner Gonçalves

Modelo para Análise Visual de Dados Multivariados no Contexto Geoespacial / Wagner Gonçalves Signoretti . – 2018.

90 p.

Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Computação, PUCRS.

Orientadora: Profa. Dra. Isabel Harb Manssour.

1. Visualização de dados. 2. Dados geoespaciais. 3. Dados multivariados. I. Manssour, Isabel Harb. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da PUCRS
com os dados fornecidos pelo(a) autor(a).
Bibliotecária responsável: Salete Maria Sartori CRB-10/1363

WAGNER GONÇALVES SIGNORETTI

**MODELO PARA ANÁLISE VISUAL DE DADOS MULTIVARIADOS NO CONTEXTO
GEOESPACIAL**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Ciência da Computação do Programa de Pós-Graduação em Ciência da Computação, Escola Politécnica da Pontifícia Universidade Católica do Rio Grande do Sul.

Porto Alegre, 27 de agosto de 2018.

BANCA EXAMINADORA:

Profa. Dra. Soraia Raupp Musse (PUCRS)

Prof. Dr. João Luiz Dihl Comba (UFRGS)

Profa. Dra. Isabel Harb Manssour (PUCRS) – Orientadora

DEDICATÓRIA

Dedico este trabalho aos meus pais, minha esposa e todos aqueles que se fizeram presentes durante minha trajetória.

“Se eu tivesse perguntado aos meus clientes o que queriam, teriam dito um cavalo mais rápido.”

(Henry Ford)

AGRADECIMENTOS

Agradeço a Deus, por abençoar todas as fases da minha vida. Se não fosse pela vontade dEle, nada disto seria possível. Agradeço aos meus pais, e, a todos que, de alguma forma, acreditaram em mim e me apoiaram nesta formação, meu muito obrigado!

À minha esposa, Ingrid, que participou comigo de noites mal dormidas, estresse, ansiedade e todos sintomas que apareceram durante o desenvolvimento da dissertação. Agradeço por todo o incentivo.

MODELO PARA ANÁLISE VISUAL DE DADOS MULTIVARIADOS NO CONTEXTO GEOESPACIAL

RESUMO

Com a popularidade dos telefones celulares, GPS e outras tecnologias, observamos um aumento na quantidade de dados geoespaciais que são gerados diariamente. Podemos encontrar conjuntos de dados com diversos atributos que estão associados a uma determinada localização geográfica, como dados abertos do governo, dados meteorológicos e dados de redes sociais. A visualização de dados geoespaciais facilita na análise de como diferentes atributos se correlacionam geograficamente, ao apresentar estes valores no mapa. No entanto, a visualização de dados multivariados no contexto geoespacial não é uma tarefa trivial. Assim, precisamos de alternativas para contribuir com a análise de dados e com a descoberta de conhecimento. O principal objetivo deste trabalho é apresentar um novo modelo para análise visual de dados multivariados no contexto geoespacial, que integra um mapa com outras representações visuais, apresentando múltiplos atributos de forma simultânea. O modelo oferece uma maneira interativa para aplicação de filtros e analisar os atributos presentes na visualização. Sua principal contribuição é a forma como os dados são apresentados e a possibilidade de aplicar as mesmas técnicas em conjuntos de dados diferentes.

Palavras-Chave: Visualização de dados, Dados geoespaciais, Dados multivariados.

MODEL FOR VISUAL ANALYSIS OF MULTIVARIATE DATA IN THE GEOSPATIAL CONTEXT

ABSTRACT

With the popularity of mobile phones, GPS and other technologies, we observe an increase in the amount of geospatial data that are daily generated. We can find datasets with several attributes associated to specific geographic coordinates, such as open government, meteorological and social network data. The visualization of geospatial data facilitates to communicate how different variables correlate to geographical locations by layering these variables over maps. However, the visualization of several variables associated with a specific location is not a trivial task. Thus, we need alternatives to contribute to the data analysis and knowledge discovery. The main goal of this work is to present a new model for visual analysis of geospatial multivariate data, which integrates a map with clustered charts and coordinated multiple views representing the values of each different attribute. This model offers a way to interactively apply different filters and analyze several attributes, updating all the related visual representations. Our main contribution is this new model proposition and its application to different open government datasets.

Keywords: Data visualization, Geospatial data, Multivariate data.

LISTA DE FIGURAS

Figura 2.1 – Mapa <i>Choropleth</i> apresentando os resultados das eleições presidenciais em 2008 nos Estados Unidos.	20
Figura 2.2 – <i>Heatmap</i> apresentando os locais onde mais fotografias foram registradas na ilha de Mainau.	21
Figura 2.3 – <i>Dot map</i> representando eventos em diferentes localidades dos Estados Unidos.	22
Figura 2.4 – Exemplo da aplicação de <i>cluster markers</i> para resolver a sobreposição de marcadores.	23
Figura 2.5 – Utilização de <i>bubble map</i> e a diferença da representação dos dados de forma absoluta e de forma relativa.	24
Figura 2.6 – Representação de um cartograma contínuo.	25
Figura 2.7 – <i>Flow map</i> representando o fluxo de turistas de Berlim nas rotas percorridas para visitar os pontos turísticos.	25
Figura 3.1 – Um <i>framework</i> de visualização analítica para dados geoespaciais. . .	35
Figura 3.2 – Visualização cartográfica de áreas vulneráveis à catástrofes naturais. . .	36
Figura 3.3 – Análise exploratória de dados geoespaciais em 3D.	37
Figura 3.4 – Kit de ferramentas para a visualização de dados geoespaciais na web. . .	38
Figura 3.5 – Ferramenta colaborativa para análise de dados geoespaciais.	39
Figura 3.6 – Análise visual de dados geoespaciais e <i>storytelling</i> utilizando HTML5. . .	40
Figura 3.7 – Demonstração do gráfico proposto como <i>Attribute signatures</i>	40
Figura 3.8 – <i>Attribute signatures</i> : técnica de visualização para dados geográficos multivariados.	41
Figura 3.9 – Visualização de dados espaço-temporais sobre gerenciamento urbano.	42
Figura 3.10 – Análise de dados estatísticos baseados em técnicas de visualização mistas.	42
Figura 3.11 – Análise visual para conjuntos de dados de incidentes de trânsito. . . .	44

Figura 3.12 – Visualização do impacto das variações geográficas em <i>clusters</i> multivariados.	44
Figura 3.13 – GazaPlot: Visualização interativa de dados geocientíficos.	45
Figura 3.14 – Visualização analítica para monitorar a qualidade do ar.	46
Figura 3.15 – Um <i>framework</i> para visualização analítica para dados de incidentes relacionados à desastres naturais.	47
Figura 3.16 – COPE: Exploração interativa de padrões espaço-temporais em diferentes localizações.	48
Figura 4.1 – Conjunto de etapas necessárias para geração da visualização.	54
Figura 4.2 – Estrutura do modelo proposto com seus componentes visuais (mapa e gráficos).	55
Figura 5.1 – Marcadores agrupados e representados através de gráficos de barras, no qual cada barra equivale a um tipo de acidente; (a) gráficos de barras agrupando os marcadores; (b) legenda dos tipos de acidente representados nos gráficos.	58
Figura 5.2 – Apresentação da interação existente entre os gráficos: (a) seleção da opção “DIA” no gráfico de pizza; (b) resultado da aplicação do filtro nos demais gráficos.	59
Figura 5.3 – Primeira versão implementada do modelo.	59
Figura 5.4 – Implementação final do modelo.	63
Figura 5.5 – Implementação final do modelo com filtros aplicados. São exibidos apenas acidentes ocorridos na zona leste.	64
Figura 5.6 – Implementação final do modelo com filtros aplicados através da movimentação do mapa. Neste caso, os gráficos estão considerando um único registro.	64
Figura 5.7 – Exemplos de como os marcadores podem ser agrupados. (a) Agrupados em gráficos de <i>donuts</i> e (b) agrupados em gráficos de barras.	65
Figura 5.8 – Interface gráfica que permite a definição de como os atributos serão representados.	66
Figura 5.9 – Tecnologias utilizadas na implementação do modelo.	68

Figura 6.1 – Estudos de caso com dados de criminalidade dos municípios do Rio Grande do Sul. 71

Figura 6.2 – Estudos de caso com dados de criminalidade de Bethesda, Estados Unidos. 72

LISTA DE TABELAS

Tabela 3.1 – Resultados dos processos de extração e seleção da RSL.	32
Tabela 3.2 – Trabalhos selecionados na RSL.	33
Tabela 3.3 – Trabalhos encontrados no <i>snowballing</i>	34
Tabela 3.4 – Técnicas de visualização e interação encontradas na RSL e no <i>snowballing</i>	50
Tabela 5.1 – Perfil dos participantes do grupo focal	60
Tabela 6.1 – Perfil dos participantes da avaliação final	72

SUMÁRIO

1	INTRODUÇÃO	16
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	TÉCNICAS DE VISUALIZAÇÃO PARA DADOS GEOESPACIAIS	19
2.1.1	<i>CHOROPLETH MAP</i>	19
2.1.2	<i>HEATMAP</i>	21
2.1.3	<i>DOT MAP</i>	21
2.1.4	<i>CLUSTER MARKERS</i>	22
2.1.5	<i>BUBBLE MAP</i>	23
2.1.6	CARTOGRAMA	24
2.1.7	<i>FLOW MAP</i>	24
2.2	OPERADORES DE INTERAÇÃO	26
2.2.1	NAVEGAÇÃO	26
2.2.2	SELEÇÃO	27
2.2.3	FILTRAGEM	27
2.2.4	CONEXÃO	28
3	TRABALHOS RELACIONADOS	29
3.1	OBJETIVO E QUESTÕES DE PESQUISA	30
3.2	TERMO DE BUSCA	31
3.3	CLASSIFICAÇÃO PRELIMINAR DOS RESULTADOS	31
3.4	EXTRAÇÃO DOS RESULTADOS	32
3.5	<i>SNOWBALLING</i>	33
3.6	TRABALHOS SELECIONADOS	34
3.7	ANÁLISE DOS RESULTADOS	48
4	DESCRIÇÃO DO MODELO PROPOSTO	53
4.1	ENTRADA DE DADOS	53

4.2	FUNCIONAMENTO DO MODELO	54
4.3	COMPONENTES VISUAIS	56
5	IMPLEMENTAÇÃO DO MODELO	57
5.1	IMPLEMENTAÇÃO PRELIMINAR	57
5.2	AVALIAÇÃO PRELIMINAR	60
5.3	IMPLEMENTAÇÃO FINAL	62
5.3.1	INTERFACE PARA ENTRADA DE DADOS	65
5.3.2	TECNOLOGIAS	67
6	TESTES E AVALIAÇÃO	69
6.1	ESTUDOS DE CASO	69
6.1.1	ACIDENTES DE TRÂNSITO NA CIDADE DE PORTO ALEGRE	69
6.1.2	CRIMES OCORRIDOS NO RIO GRANDE DO SUL AGRUPADOS POR MUNI- CÍPIO	70
6.1.3	CRIMES OCORRIDOS EM BETHESDA, MARYLAND	71
6.2	AVALIAÇÃO COM USUÁRIOS	72
7	CONCLUSÕES	77
	REFERÊNCIAS	79
	APÊNDICE A – Strings de busca	87
A.1	IEEEEXPLORE DIGITAL LIBRARY	87
A.2	SCOPUS	87
A.3	ESLEVIER SCIENCEDIRECT	87
A.4	ACM DIGITAL LIBRARY	88
	APÊNDICE B – Termos de Consentimento	89
B.1	GRUPO FOCAL	89
B.2	ENTREVISTAS	90

1. INTRODUÇÃO

De acordo com Ward et al. [69], “a visualização de dados pode ser definida como a comunicação da informação através de representações gráficas. Imagens sempre foram utilizadas como um mecanismo de comunicação, já que uma imagem pode conter várias informações e ser interpretada pelos seres humanos em menos tempo que uma página repleta de palavras, por exemplo. Isto ocorre por que o sistema de percepção humana consegue interpretar imagens de forma paralela, enquanto a velocidade que demoramos para ler um texto é limitada pelo processamento sequencial de leitura”.

A visualização de dados geoespaciais, por sua vez, é uma das formas mais comuns de visualização, pois muitas das informações disponíveis hoje em dia possuem uma localização específica. Dados geoespaciais são diferentes de outros tipos de dados por possuírem esta localização no mundo real. Informações geográficas podem ser utilizadas praticamente por qualquer tipo de aplicação, como máquinas de pagamento de cartão de crédito, interações em redes sociais e censos demográficos. Mapas correspondem a uma representação do mundo, sobre a qual as informações são exibidas através de pontos, linhas e áreas associadas uma localização específica. Os parâmetros de visualização, incluindo tamanho, forma, valor, textura, cor e orientação, mostram informações adicionais sobre os objetos em questão [69].

Não há um consenso sobre o que caracteriza um conjunto de dados como multivariado. Liu et al. [44] definem que conjuntos de dados multivariados devem apresentar cerca de 3 atributos que não sejam de natureza espacial ou temporal, enquanto Oliveira e Levkowitz [14] definem que estes conjuntos devem apresentar mais de 4 atributos relacionados. Considerando estas definições, os trabalhos que apresentam conjuntos de dados nestas condições são aqui classificados como multivariados. Existem diversas técnicas de visualização que possibilitam a representação de dados multivariados, tais como coordenadas paralelas [23], *table lens* [53], gráfico de dispersão [37], entre outras [69].

De acordo com Cook e Thomas [13], a análise visual (ou *Visual Analytics*) é a ciência do raciocínio analítico facilitada por interfaces visuais interativas. Estas interfaces são capazes de apresentar informações complexas e permitem que decisões sejam tomadas com base no que é apresentado [36]. Com isto, especialistas de diferentes domínios

podem direcionar seus esforços para compreender as visualizações e utilizar o seu conhecimento no domínio em questão para analisá-las [32, 29].

A quantidade de dados produzidos atualmente cresce de forma considerável. A capacidade de coletar e armazenar estes dados está aumentando a uma taxa maior que a de processá-los [13]. Com isto, existe a necessidade de visualizações que possibilitem a representação e a manipulação destes dados. Grande parte das visualizações encontradas na literatura foi concebida para um conjunto de dados específico, sem considerar que outros dados possam ser representados a partir das suas visualizações [49, 42, 74, 77, 41, 76]. De acordo com Fisher [18], a maioria das ferramentas de visualização geoespacial consegue comparar apenas variáveis específicas em um determinado momento.

Motivado pelas necessidades encontradas na literatura, este trabalho tem como objetivo facilitar a análise visual de dados multivariados no contexto geoespacial, de forma que tomadas de decisão possam ser feitas a partir das suas representações visuais. Com o intuito de encontrar quais técnicas de visualização e interação estão sendo apresentadas e propostas, na literatura, para representar dados multivariados no contexto geoespacial, é preciso fazer um estudo do estado da arte. Para isto, foi realizada uma revisão sistemática da literatura [33] e um *snowballing forward* [72] que considerou artigos que apresentaram técnicas de visualização relacionadas a mapas com o objetivo de auxiliar a análise e percepção visual de dados associados a uma localização geográfica. Através da análise destes trabalhos, pôde-se encontrar oportunidades que não foram abordadas na literatura.

A partir das oportunidades encontradas, é apresentado um modelo para visualização de dados multivariados que apresenta marcadores no mapa através de gráficos e combina este mapa com outros gráficos, agregando mais informações sobre os dados em questão. Suas principais contribuições incluem: (1) a possibilidade de representar diversos atributos no mapa de forma dinâmica; (2) uma abordagem interativa baseada nas técnicas *brushing and linking* [5], integradas à *coordinated multiple views* (CMV) [55, 56, 6] para auxiliar na análise visual; e (3) a utilização do modelo proposto para analisar dados abertos do governo. A contribuição está diretamente associada à descoberta de conhecimento, considerando que diversas técnicas são utilizadas com o propósito de estimular uma interação ativa do usuário final.

Este trabalho está estruturado da seguinte maneira: o Capítulo 2 fundamenta as técnicas de visualização e interação que são abordadas no decorrer do trabalho. O Capí-

tulo 3 define e explica o protocolo utilizado na realização da revisão sistemática da literatura, apresentando os trabalhos encontrados e incluídos na pesquisa. Este capítulo engloba uma análise dos resultados encontrados, relacionando os trabalhos com as técnicas presentes na literatura. Os Capítulos 4, 5 e 6 descrevem o modelo, suas implementações e a forma como foi avaliado. Por fim, o Capítulo 7 apresenta as conclusões do trabalho, seguido pelas referências bibliográficas.

2. FUNDAMENTAÇÃO TEÓRICA

Existem diversas técnicas que possibilitam a análise e a exploração de dados geoespaciais. A Seção 2.1 apresenta definições e aplicações das técnicas de visualização voltadas para a representação de dados geoespaciais, enquanto a Seção 2.2 apresenta os principais operadores de interação utilizados na área de visualização de dados.

Existem diversas técnicas de visualização e interação que permitem a visualização e análise de dados. A Seção 2.1 apresenta definições e aplicações das técnicas de visualização voltadas para a representação de dados geoespaciais, enquanto a Seção 2.2 apresenta os principais operadores de interação utilizados na área de visualização de dados.

2.1 Técnicas de Visualização para dados geoespaciais

Considerando a grande quantidade de dados que é produzida atualmente, torna-se muito difícil analisá-los em sua forma bruta. Por isso, diferentes técnicas de representações visuais destes dados são usadas como uma forma de auxiliar nesta análise. Quando muitas destas informações estão relacionadas a uma localização geográfica [69], são usadas, então, técnicas de visualização de dados geoespaciais.

É possível representar informações associadas a uma localização específica de um mapa através de pontos, linhas ou regiões, que, de acordo com a informação, podem ter diferentes propriedades como tamanho, forma, espessura, textura e cor. Os mapas podem ser subdivididos de acordo com as propriedades dos dados (qualitativo ou quantitativo, discreto ou contínuo) e das variáveis geográficas (pontos, linhas, superfícies e volumes). A seguir são apresentados exemplos das principais técnicas de visualização aplicadas à dados geoespaciais.

2.1.1 *Choropleth Map*

Mapas temáticos são a principal abordagem para se representar um determinado fenômeno. Existem diversos tipo de mapas temáticos, o mais popular deles é o mapa

choropleth. O mapa *choropleth* demarca suas regiões com diferentes cores e padrões de sombreamento de acordo com o valor definido para cada região. Mapas que utilizam esta técnica de visualização são utilizados para enfatizar a distribuição espacial de um ou mais atributos geográficos [50]. Na Figura 2.1, esta técnica é utilizada para apresentar os resultados das eleições presidenciais em 2008 nos Estados Unidos. A cor azul representa as regiões onde o candidato Obama teve mais votos, e a cor vermelha representa as regiões onde o candidato McCain teve mais votos.

Um problema presente nestes mapas é que os valores mais interessantes para a visualização podem estar concentrados em áreas pequenas e com uma grande concentração populacional, enquanto os demais valores podem estar concentrados em áreas maiores, porém, com uma concentração populacional menor. Nestes casos, os mapas *choropleth* tendem a destacar as regiões maiores, mesmo que não sejam valores tão relevantes para a visualização [69]. No censo demográfico dos Estados Unidos, por exemplo, os mapas *choropleth* tendem a destacar padrões em áreas com uma menor concentração populacional.

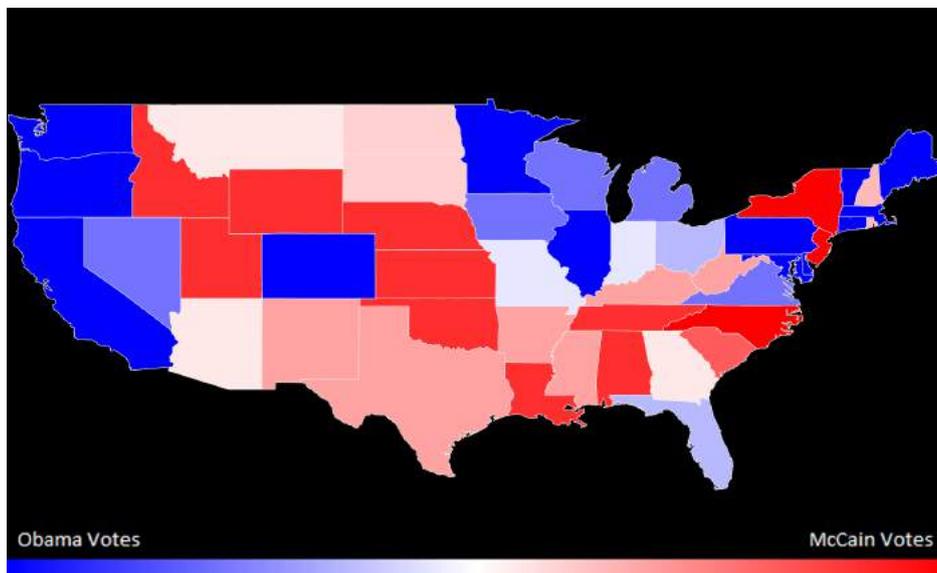


Figura 2.1 – Mapa *Choropleth* apresentando os resultados das eleições presidenciais em 2008 nos Estados Unidos [69].

2.1.2 Heatmap

De acordo com Sneath [62], os *heatmaps* se diferenciam dos mapas *choropleth*, apresentados anteriormente, porque o limite das cores não está associado aos limites geográficos, como cidades ou estados.

Este tipo de mapa é útil para a identificação de padrões, especialmente de pontos de concentração ou regiões com uma alta densidade de um determinado atributo. Na Figura 2.2, podemos ver, através de um *heatmap*, a quantidade de fotografias registradas na ilha de Mainau. Nesta figura as cores variam entre amarelo, vermelho e preto, sendo que as marcas amarelas representam os locais mais fotografados. Na Figura 2.2, os autores apresentam um problema presente no *heatmap*: a sobreposição de áreas, onde a região demarcada ultrapassa fronteiras geográficas [69].

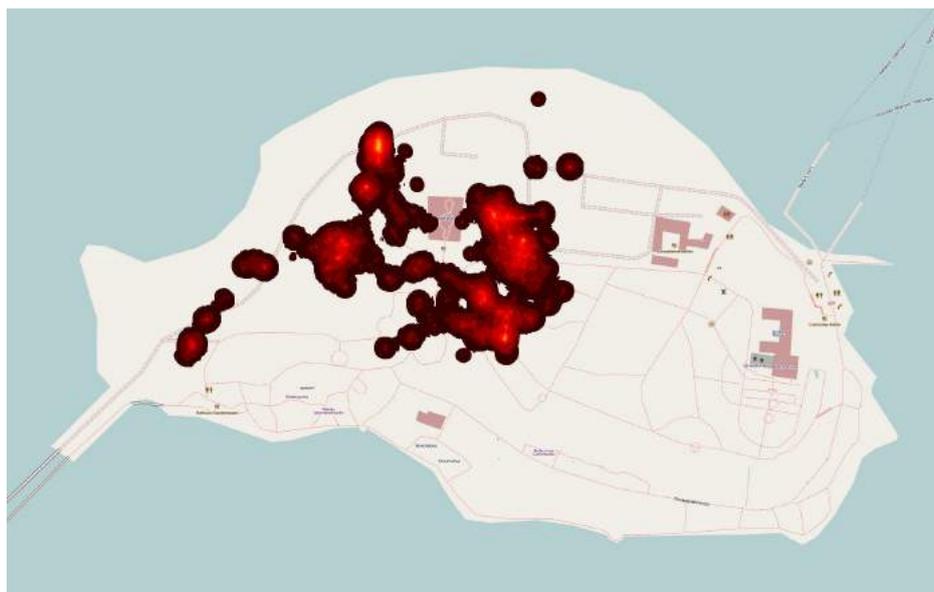


Figura 2.2 – *Heatmap* apresentando os locais onde mais fotografias foram registradas na ilha de Mainau [69].

2.1.3 Dot Map

Registros que apresentam informações geográficas podem ser representados no mapa através de pontos, de acordo com a suas coordenadas geográficas. Esta técnica de visualização simples é chamada *dot map*. O tamanho e a cor destes pontos pode ser

ajustado de acordo com o valor de cada atributo, com o objetivo de diferenciar os registros no mapa [31].

O uso desta técnica associada a um mapa possibilita uma representação precisa do valor de cada variável em uma determinada localização. Entretanto, é importante que o conjunto de dados disponibilize a latitude e a longitude na qual cada ponto deve ser adicionado. Na Figura 2.3 esta técnica foi utilizada por Keim et al. [31] para apresentar a localização espacial de eventos sobre o mapa dos Estados Unidos.

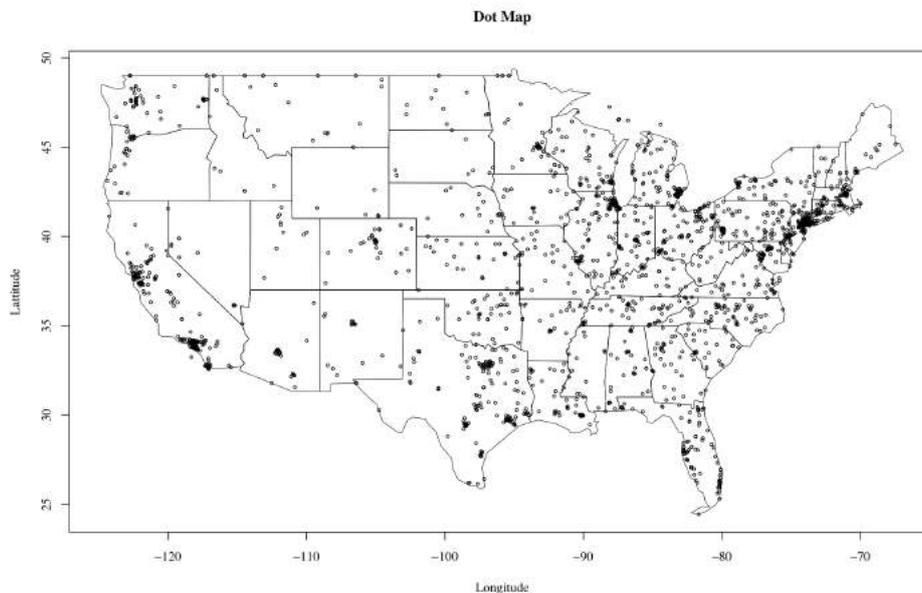


Figura 2.3 – *Dot map* representando eventos em diferentes localidades dos Estados Unidos [31].

2.1.4 *Cluster Markers*

De acordo com Rayson [54], *cluster markers* (marcadores agrupados) auxiliam a agrupar e representar um conjunto de dados sobre o mapa. Cada conjunto é representado por um círculo que tem seu tamanho ajustado de acordo com a quantidade de dados que estão sendo nele agrupados. De acordo com Delort [15], quando a densidade de marcadores é muito alta em uma região específica, estes marcadores são substituídos por um único objeto.

Cluster maps são ideais para mapas interativos, nos quais o usuário pode fazer *zoom in* e visualizar os pontos que representam os dados individuais que estão contidos

em cada *cluster*. Assim, os *cluster markers* auxiliam a reduzir a quantidade de marcadores exibidos no mapa, agrupando-os por proximidade e evitando a sobreposição caso vários marcadores sejam exibidos de forma simultânea.

A Figura 2.4a apresenta os marcadores antes de serem agrupados, sendo sobrepostos uns aos outros. Na Figura 2.4b, apenas um dos marcadores é colocado no mapa, o qual é responsável por agrupar os marcadores próximos a ele [47]. Neste exemplo, o *zoom* funciona como um mecanismo para o usuário especificar a quantidade de detalhes que ele espera da visualização.

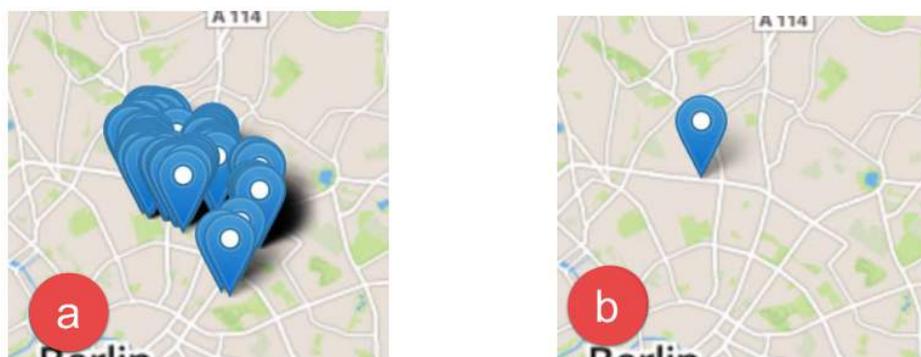


Figura 2.4 – Exemplo da aplicação de *cluster markers* para resolver a sobreposição de marcadores. Fonte: Meier e Heidmann [47]

2.1.5 *Bubble Map*

Bubble maps, ou mapas de bolha, podem representar duas variáveis de forma simultânea na mesma visualização: uma associada ao tamanho de cada bolha e outra associada à cor desta bolha. Assim, mapas de bolha auxiliam os usuários a diferenciar três parâmetros apenas olhando para o mapa, que correspondem a localização, cor e tamanho. Entretanto, podem surgir problemas quando um mapa tiver muitas localizações geográficas pequenas e próximas, pois poderá haver sobreposição de bolhas [7].

A Figura 2.5 apresenta um exemplo de *bubble map*, no qual podemos observar a diferença da utilização de mapeamento absoluto e mapeamento relativo. Os números absolutos são apresentados na Figura 2.5a, enquanto na Figura 2.5b os números são apresentados de forma relativa aos números populacionais. Podemos observar que devido à grande diferença populacional entre as regiões, o tamanho das bolhas acaba se invertendo.

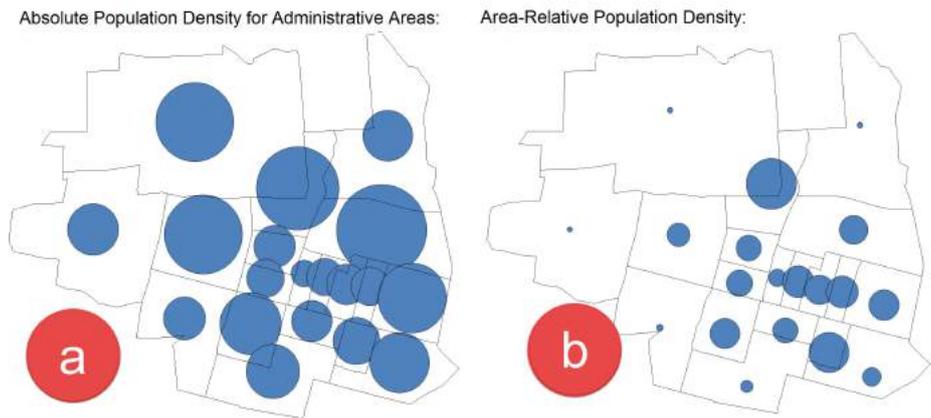


Figura 2.5 – Utilização de *bubble map* e a diferença da representação dos dados de forma absoluta (a) e de forma relativa (b). Fonte: Ward et al. [69]

2.1.6 Cartograma

Em um cartograma, a variável que está sendo mapeada é utilizada para fazer distorções no mapa, com o objetivo de ressaltá-la [64]. As alterações feitas na projeção das regiões devem preservar o aspecto da área real das regiões, conforme o conhecimento do usuário. Uma boa prática adotada pelos autores que utilizam esta técnica de visualização é apresentar o mapa original antes, para que o usuário acompanhe as alterações que estão sendo feitas [69].

A Figura 2.6 apresenta um cartograma contínuo, no qual a topologia do mapa se mantém, ou seja, as regiões continuam conectadas, contudo, podemos observar distorções na forma destas regiões.

2.1.7 *Flow Map*

De acordo com Ward et al. [69], a técnica de visualização *Flow map* é inspirada em algoritmos utilizados para desenhar grafos que minimizam a quantidade de vértices que se cruzam. A Figura 2.7 apresenta um exemplo de visualização do fluxo de turistas em Berlim nas rotas percorridas para visitar os pontos turísticos. Na representação apresentada podemos observar, além de quais vias são mais utilizadas pelos turistas, em quais direções estes turistas seguem. Desta forma, torna-se possível inferir informações como as regiões

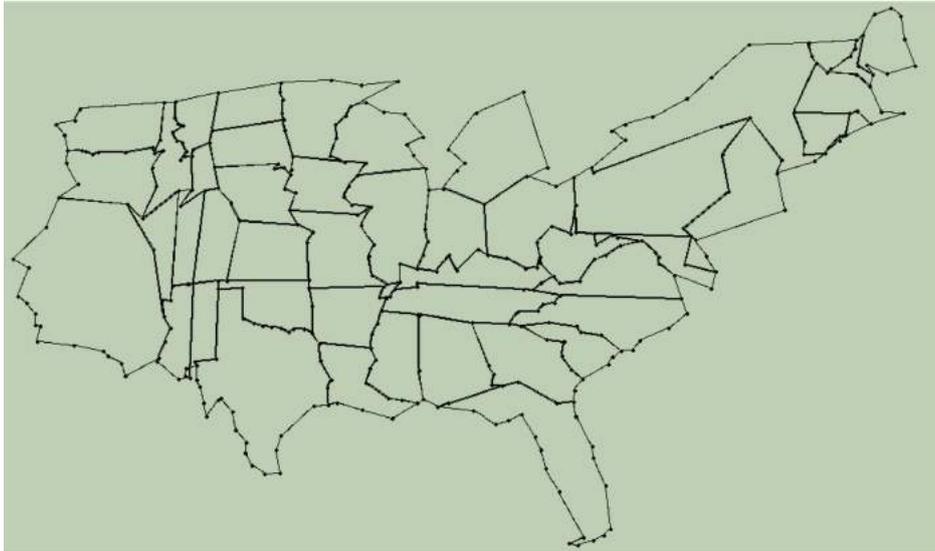


Figura 2.6 – Representação de um cartograma contínuo. Fonte: Keim et al. [30]

que os turistas se hospedam, locais com um maior fluxo em horário de almoço, entre outras que podem auxiliar em tomadas de decisão.

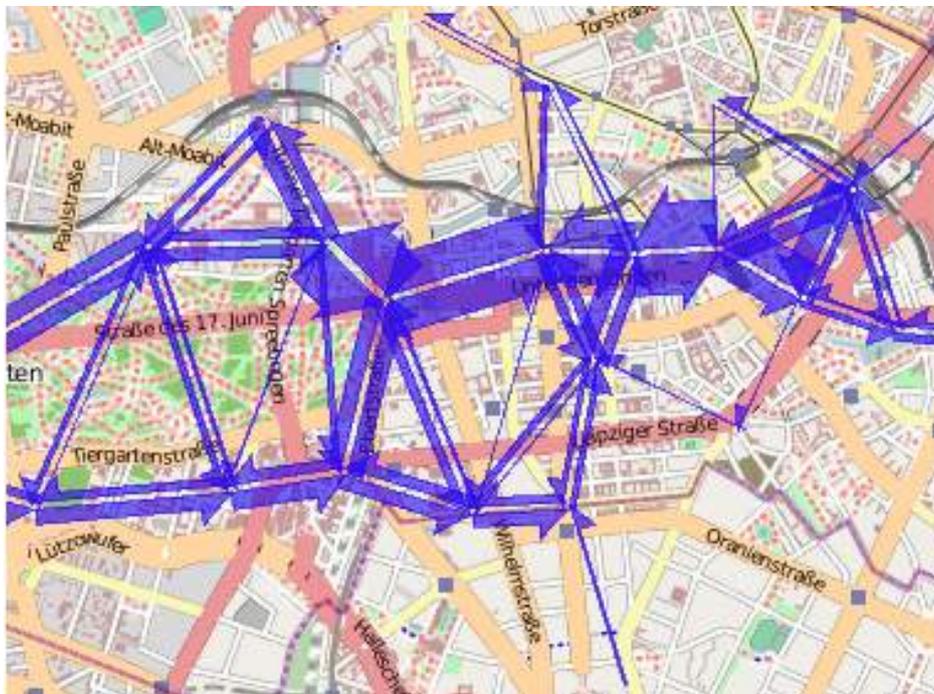


Figura 2.7 – *Flow map* representando o fluxo de turistas de Berlim nas rotas percorridas para visitar os pontos turísticos. Fonte: Phan et al. [51]

2.2 Operadores de Interação

A interação dentro do contexto de visualização de dados funciona como um mecanismo para modificar tanto o que é apresentado aos usuários, como a forma de apresentação [70]. Embora gráficos estáticos sejam muito úteis, eles utilizam pouco do imenso poder computacional disponível nos dias de hoje. Interação na visualização intensifica a exploração dos dados e a descoberta de conhecimento, além de permitir a redução da quantidade de informações sendo exibidas, propiciando visualizações mais claras [17]. De acordo com Shneiderman [61], primeiro deve ser apresentada uma perspectiva geral para que o usuário saiba onde encontrar os dados, porém, é necessário interagir com os gráficos para ver os detalhes. Por exemplo, quando se lida com conjuntos de dados com muitos registros é importante poder selecionar um subconjunto de registros para analisar, de forma que detalhes possam ser encontrados pelo usuário conforme necessário.

Existem diversas classes de interação [75], tais como: (1) navegação, que permite ao usuário controlar o posicionamento da câmera e redimensionar o que é exibido na tela através da movimentação, rotação e *zoom*; (2) seleção, que permite que operações como realçar, deletar ou modificar sejam aplicadas a objetos ou regiões de interesse; (3) filtros, que permitem que a quantidade de informação sendo exibida na tela seja reduzida ou aumentada através da eliminação ou inclusão de registros ou dimensões; e (4) conexão, que possibilita que diferentes componentes sejam conectados com o objetivo de demonstrar o relacionamento entre seus atributos.

Estas classes de interação são definidas e descritas como operadores de interação. Muitas vezes os operadores de interação estão relacionados às classes de interação, e praticamente todos os operadores podem ser executados de forma interativa ou automática dentro de cada visualização [69]. As subseções a seguir descrevem os operadores de interação mais comuns em visualizações de dados.

2.2.1 Navegação

A navegação, que algumas vezes é definida como exploração, é utilizada para percorrer um conjunto de dados a fim de encontrar um subconjunto de interesse. Os operado-

res de navegação podem funcionar com coordenadas absolutas ou relativas dentro de um determinado espaço. Estes operadores permitem que o usuário controle e inspecione uma projeção específica da visualização, oferecendo detalhes em diferentes granularidades, de acordo com o nível de detalhamento desejado [69].

Por exemplo, a técnica de interação de foco e contexto permite que a quantidade de detalhes seja mostrada de forma seletiva na área contextual. Baseia-se em 3 premissas: permitir uma perspectiva geral (contexto) e uma visão detalhada (foco) simultaneamente; as informações apresentadas na perspectiva geral podem ser diferentes do que é necessário mostrar na visão detalhada; e as opções de foco e contexto devem ser combinadas na mesma visualização de forma dinâmica, similar à visão humana [11]. *Panning* possibilita a navegação na visualização, sendo que o usuário pode movimentar a visualização em qualquer direção [21].

2.2.2 Seleção

Os operadores de seleção permitem selecionar parte dos dados que estão sendo exibidos em uma visualização para fazer alguma alteração sobre os mesmos, tal como realce, remoção e alteração de dados, ou para exibi-los de diferentes maneiras. A seleção dos dados pode ocorrer de diversas formas, como através de um clique em elementos da visualização para selecionar alguns deles ou todos elementos de uma área delimitada. Seleções também podem ser realizadas de forma indireta, quando os elementos são selecionados automaticamente se estão de acordo com um determinado critério [71].

2.2.3 Filtragem

Como o próprio nome sugere, os filtros reduzem o volume dos dados exibidos em uma visualização com base em determinados critérios. Estes critérios visam definir quais valores serão preservados ou removidos da visualização. Por exemplo, um *slider* pode ser utilizado para especificar um intervalo de interesse, e a visualização é automaticamente atualizada, refletindo a seleção feita pelo usuário [60]. O filtro por intervalos é apenas uma das formas de filtragem, pois também é possível selecionar itens em um conjunto ou lista

para que os dados sejam preservados ou escondidos, tal como na operação de esconder coluna disponível no Excel.

A diferença entre os operadores de seleção e filtragem é sutil, porém, importante. Filtros são geralmente aplicados de maneira indireta, pois, por exemplo, a especificação de valores para o filtro não é realizada na visualização de dados em si, e, sim, por meio de uma interface gráfica auxiliar. Geralmente os filtros são aplicados antes dos dados serem visualizados, para evitar que os mesmos sejam sobrepostos na tela. A seleção ocorre, na maioria das vezes, de maneira direta, feita através da interação do mouse na visualização em si.

2.2.4 Conexão

Operações de conexão são usados para vincular os componentes visuais que compõe a visualização. *Linked selection*, ou *linking*, é provavelmente a forma mais comum de comunicação entre componentes encontrada em ferramentas de visualização analítica [69], já que ilustra o relacionamento existente entre diferentes atributos da visualização.

Quando os componentes visuais podem ser manipulados de maneira interativa, o operador é chamado *brushing*. Através deste operador, as interações feitas em um determinado gráfico são automaticamente refletidas nos demais gráficos.

A união destas técnicas, chamada *brushing and linking* [5], combina diferentes métodos de visualização com o objetivo de suprir as deficiências de cada técnica em particular. Ao conectar diferentes técnicas de visualização através de *brushing and linking*, mais informações podem ser fornecidas aos usuários [28]. Gráficos de dispersão e coordenadas paralelas podem utilizar *brushing and linking* para que mais informações sejam reveladas à medida que ocorre interação [68].

O termo *Coordinated Multiple Views* (CMV), por sua vez, é utilizado para descrever sistemas com múltiplas visualizações que funcionam de forma integrada [55, 56, 6]. A integração destas múltiplas visualizações funciona com operadores de conexão, como *brushing and linking*, por exemplo.

3. TRABALHOS RELACIONADOS

De acordo com Kitchenham [33], Revisão Sistemática da Literatura (RSL) é uma técnica que possibilita a avaliação das pesquisas disponíveis na literatura, com o objetivo de identificar quais estudos são relevantes para uma determinada questão de pesquisa, uma área temática ou um fenômeno de interesse. A RSL teve início na literatura médica e, ao mostrar sua eficiência de forma geral, passou a ser utilizada também em outras áreas do conhecimento, sendo uma delas a ciência da computação [4]. Recentemente, a RSL tem se mostrado uma ótima forma de conhecer uma determinada área. Booth et al. [8] sugerem que sem uma RSL não somos capazes de entender plenamente o tópico de pesquisa escolhido, de identificar o que já foi pesquisado e o que ainda pode ser explorado.

Uma revisão sistemática da literatura reúne, de forma organizada, o resultado de vários estudos em uma determinada área do conhecimento. Além disso, responde a uma pergunta de pesquisa utilizando métodos sistemáticos e explícitos para identificar, selecionar e avaliar os estudos de forma criteriosa.

A RSL descrita neste documento tem como objetivo identificar na literatura científica de conferências e revistas, quais são as principais técnicas de visualização e interação utilizadas para representar dados multivariados no contexto geoespacial, seguindo as diretrizes propostas por Kitchenham [33]. Para complementar a busca por trabalhos relacionados foi feito um *snowballing forward*, esta técnica permite que trabalhos que não retornaram nas buscas da RSL sejam encontrados, além de adicionar trabalhos mais recentes ao estudo [72].

As seções a seguir estão organizadas da seguinte maneira: a Seção 3.1 apresenta o objetivo e as questões de pesquisa utilizadas para nortear a RSL. As Seções 3.2 e 3.3 apresentam o termo de busca utilizado para delimitar os resultados das bases de dados e os critérios de inclusão e exclusão utilizados para classificar os trabalhos. A Seção 3.4 apresenta os 12 trabalhos que atendem aos critérios de inclusão, apresentando-os de acordo com as questões de pesquisa, seguido pela Seção 3.5, que complementa os resultados da RSL adicionando 3 artigos novos aos resultados. Por fim, as Seções 3.6 e 3.7 apresentam estes trabalhos e trazem uma análise dos resultados encontrados, relacionando os trabalhos de acordo com as técnicas apresentadas.

3.1 Objetivo e Questões de Pesquisa

O objetivo desta RSL é conhecer o estado da arte sobre as técnicas de visualização e interação utilizadas para representar dados multivariados no contexto geoespacial. Neste caso, para serem considerados relevantes para a pesquisa, os artigos devem atender uma série de critérios, definidos na RSL.

Uma questão de pesquisa (QP) adequada, ou seja, bem formulada, permite ao pesquisador encontrar a definição correta e qual informação é necessária para alcançar seu objetivo [34]. Com a realização desta RSL, procura-se responder as seguintes questões de pesquisa:

- QP1: Quais técnicas de visualização geoespacial estão sendo utilizadas para apresentar dados multivariados?
- QP2: Quais técnicas de interação estão sendo utilizadas com as técnicas de visualização identificadas na QP1?
- QP3: As técnicas de visualização e interação apresentadas são aplicáveis a diferentes contextos?

Para estruturar a RSL, as questões de pesquisa foram definidas utilizando a classificação PICO, buscando delimitar o escopo de cada questão. De acordo com Sackett [63], PICO representa um acrônimo de População, Intervenção, Comparação e *Outcomes* (Resultados). Estes quatro componentes são elementos essenciais de qualquer questão de pesquisa e são definidos a seguir:

- População: Artigos científicos que apresentam técnicas de visualização de dados multivariados no contexto geoespacial;
- Intervenção: Técnicas de visualização mais utilizadas para apresentar mapas;
- Comparação: Não se aplica. Este estudo não tem por objetivo comparar os trabalhos encontrados com outras abordagens;
- *Outcomes* (Resultados): Contexto no qual as técnicas foram aplicadas, se foram desenvolvidas para um contexto específico ou podem ser aplicadas de forma genérica.

3.2 Termo de busca

A classificação PICO tem como objetivo a identificação de palavras-chave e a formulação de *strings* de busca a partir das questões de pesquisa. A partir das palavras-chave, sinônimos devem ser identificados e combinados utilizando conectores lógicos [33]. Para o contexto desta RSL foram identificadas as seguintes palavras-chave: *Geovisualization* e *Multivariate*. A partir destas palavras-chave, foi criada a seguinte *string* de busca:

((((GEOVISUALIZATION OR GEOSPATIAL OR GEOGRAPHIC* OR GEOVIS OR GEOSCIENCE) AND (VISUALIZ*) AND (MULTIVARIATE OR MULTI-VARIATE OR "MULTIPLE-PARAMETERS"))))

Para a realização das buscas por artigos, a *string* de busca foi aplicada nos motores de busca da *IEEEExplore digital library*¹, da *Scopus*², da *Elsevier ScienceDirect*³ e da *ACM digital library*⁴. Considerando que as buscas por meio das bases de artigos funcionam com diferentes sintaxes, o Apêndice A apresenta qual *string* de busca foi utilizada em cada base.

3.3 Classificação Preliminar dos Resultados

Para aprimorar os resultados encontrados na RSL, os trabalhos devem respeitar critérios de inclusão e exclusão. Estes critérios foram definidos buscando refinar a quantidade de trabalhos e garantir que estamos analisando artigos relevantes para a área de pesquisa desejada. Portanto, foram utilizados os seguintes critérios de inclusão: ter sido publicado nos últimos dez anos; estar escrito em Inglês; e disponível de forma integral.

Para que os artigos sejam descartados, os mesmos devem se enquadrar em pelo menos um dos seguintes critérios de exclusão: título ou resumo não relacionado ao tema; artigos com menos de quatro páginas; *keynote* de eventos; resumos; trabalhos que não

¹ieeexplore.ieee.org/Xplore/home.jsp

²www.scopus.com/home.uri

³www.sciencedirect.com/

⁴dl.acm.org/

apresentam técnicas de visualização de dados geoespaciais; e trabalhos que não apresentam técnicas de interação.

De acordo com a Tabela 3.1, foram encontrados 111 artigos após executar a *string* de busca nas bases de dados: 76 artigos foram encontrados na *Scopus*, 32 artigos encontrados através da *IEEEExplore digital library*, 2 artigos encontrados através da *ACM digital library* e 1 destes artigos encontrado através da *Elsevier ScienceDirect*.

3.4 Extração dos Resultados

Após a fase de classificação preliminar dos resultados, inicia-se a extração. Inicialmente, verificamos que do total de 111 artigos apresentados na Tabela 3.1, 33 deles retornaram em mais de uma base de dados. Por este motivo estes artigos foram classificados como duplicados e foram excluídos. Na sequência, todos artigos foram acessados e classificados de acordo com as questões de pesquisa, considerando que para ser classificado como um artigo aceito na RSL, este deve apresentar técnicas de interação e visualização de dados multivariados no contexto geoespacial. Assim, após todos os artigos terem sido analisados, apenas 12 deles foram selecionados por estarem de acordo com as questões de pesquisa. Os artigos selecionados estão listados na Tabela 3.2 e mais informações sobre cada um deles são apresentadas na Seção 3.6.

Base de dados	Extraídos	Aceitos
<i>Scopus</i>	76	12
<i>IEEEExplore digital library</i>	32	4
<i>ACM digital library</i>	2	0
<i>Elsevier ScienceDirect</i>	1	0
Total duplicados:	33	4
Total:	111	12

Tabela 3.1 – Resultados dos processos de extração e seleção da RSL.

Referência	Título do trabalho	Ano	Citações
Zhang et al. [78]	<i>An interactive visual analytics framework for multi-field data in a geo-spatial context</i>	2013	16
Opach e Rød [49]	<i>Cartographic visualization of vulnerability to natural hazards</i>	2013	23
Van Ho et al. [66]	<i>A web-enabled visualization toolkit for geo-visual analytics</i>	2012	23
Turkay et al. [65]	<i>Attribute signatures: Dynamic visual summaries for analyzing multivariate geographical data</i>	2014	35
Liono et al. [42]	<i>Visualization oriented spatiotemporal urban data management and retrieval</i>	2015	3
Liu et al. [43]	<i>Analysis of statistics data based on mixed visualization techniques</i>	2012	2
Jern [25]	<i>Collaborative Web-Enabled GeoAnalytics Applied to OECD Regional Data</i>	2009	19
Lundblad e Jern [45]	<i>Geovisual analytics and storytelling using HTML5</i>	2013	5
Wongsuphasawat et al. [74]	<i>Visual analytics for transportation incident data sets</i>	2009	16
Zhang et al. [77]	<i>Visualizing the Impact of Geographical Variations on Multivariate Clustering</i>	2016	8
Ho e Jern [22]	<i>Exploratory 3D geovisual analytics</i>	2008	10
Li et al. [40]	<i>An Interactive visualization approach to the overview of geoscience data</i>	2013	3

Tabela 3.2 – Trabalhos selecionados na RSL.

3.5 Snowballing

O termo *snowballing* [73] se refere ao uso das referências bibliográficas ou das citações de um artigo para identificar trabalhos adicionais. Seu procedimento funciona da seguinte maneira: é necessário ter um conjunto inicial de artigos, caracterizado, por exemplo, pelos artigos aceitos na fase de extração dos resultados da RSL; A partir deste conjunto de artigos, são feitas interações, procurando por trabalhos que os referenciam ou que são referenciados; Os trabalhos encontrados são submetidos aos mesmos critérios de inclusão e exclusão da RSL. *Snowballing* é particularmente útil para estender uma RSL, considerando que estudos novos devem citar pelo menos um estudo relevante na área [72].

De acordo com Wohlin [73], o *snowballing* possui dois processos de interações para busca de trabalhos: *backward* (para trás) e *forward* (para frente). No processo *backward* são analisadas as referências bibliográficas do conjunto inicial de artigos, enquanto no pro-

cesso *forward*, a ideia é utilizar o conjunto inicial para procurar por todos os artigos que os referenciaram [72].

Como entre os trabalhos selecionados na RSL não haviam publicações dos últimos dois anos, utilizamos estes trabalhos como o conjunto inicial de artigos e aplicamos o processo de interação *forward* do *snowballing*. Assim, para cada um dos 12 artigos resultantes da RSL foi aplicado o seguinte processo: busca pelo título do artigo no Google Acadêmico⁵; identificação de todos os artigos que o citaram; aplicação dos critérios de inclusão e exclusão da RSL. No total, foram encontrados e analisados 163 artigos, dos quais apenas 3 atendiam aos critérios de inclusão e exclusão e respondiam às questões de pesquisa. Estes três artigos são apresentados na Tabela 3.3.

Sumarizando, para este estudo foram considerados 15 artigos, sendo 12 artigos resultantes da RSL e 3 artigos resultantes do *snowballing*.

Referência	Título	Ano
Liao et al. [41]	<i>A web-based visual analytics system for air quality monitoring data</i>	2014
Zhang et al. [76]	<i>A visual analytics framework for microblog data analysis at multiple scales of aggregation</i>	2016
Li et al. [39]	<i>COPE: Interactive Exploration of Co-occurrence Patterns in Spatial Time Series</i>	2018

Tabela 3.3 – Trabalhos encontrados no *snowballing*.

3.6 Trabalhos selecionados

Zhang et al. [78] apresentam um *framework* interativo para análise visual de dados multivariados no contexto geoespacial. Foi desenvolvida uma extensão para o *Google Earth* que permite a aplicação de técnicas de interação, como *brushing* e filtros, e de visualização, como mapa com marcadores. A partir desta extensão, eles permitiram uma interação bidirecional entre os gráficos e o mapa (Figura 3.1). Os pesquisadores apresentam gráficos de pizza como marcadores localizados em regiões de interesse sobre o mapa. Estes gráficos de pizza apresentam mais informações sobre a região em questão no momento em que são selecionados. A viabilidade do *framework* foi demonstrada por meio de um estudo sobre o efeito de aerossóis no aquecimento global. As visualizações apresentadas são referentes

⁵Ferramenta de pesquisa do Google que permite pesquisas em trabalhos acadêmicos, literatura escolar, jornais de universidades e artigos variados.

à este contexto específico e os tipos de gráficos não podem ser alterados dinamicamente. Através deste *framework*, quatro atributos podem ser visualizados de forma simultânea.

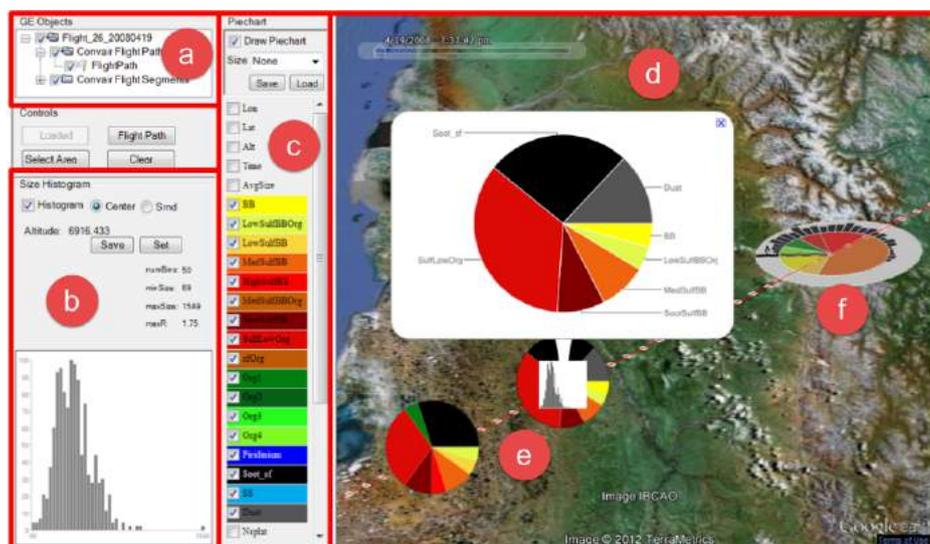


Figura 3.1 – Um *framework* de visualização analítica para dados geoespaciais. Fonte: Zhang et al. [78]

Opach e Rød [49] apresentam um estudo de caso sobre áreas de vulnerabilidade através do *ViewExposed*, uma ferramenta para análise de dados multivariados. A sua criação foi motivada pela necessidade de encontrar quais regiões se tornam mais vulneráveis em decorrência de alterações climáticas e quais são as causas desta vulnerabilidade. Por meio de validações com usuários, os pesquisadores constataram que a vinculação das múltiplas visualizações é a maior vantagem da ferramenta. Para que isso ocorra, os vértices dos objetos são destacados conforme ocorre interação com o mouse. Foi apresentado um estudo de caso com os municípios da Noruega (Figura 3.2). Ao clicar nos municípios, destacados no mapa *choropleth*, informações referentes àquele município são exibidas em um gráfico de coordenadas paralelas. Ao todo, doze atributos podem ser visualizados de forma simultânea. As informações apresentadas no *ViewExposed* são provenientes de conjuntos de dados em diferentes formatos: GeoJSON para definir o que é apresentado no mapa, CSV (*Comma-Separated Values*) para as informações do gráfico de coordenadas paralelas e JSON (*JavaScript Object Notation*) para as informações da tabela. Os componentes presentes na visualização são fixos: um mapa *choropleth*, um gráfico de coordenadas paralelas e uma tabela. Embora o estudo de caso apresentado seja sobre uma região específica, os pesquisadores acreditam que as técnicas de visualização presentes no *ViewExposed* podem ser aplicadas a dados de vulnerabilidade de qualquer localidade.

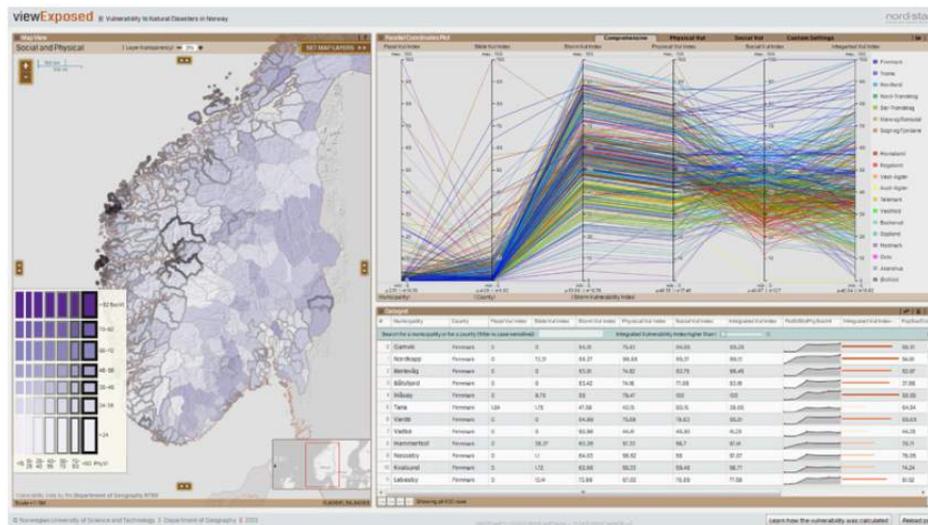


Figura 3.2 – Visualização cartográfica de áreas vulneráveis à catástrofes naturais. Fonte: Opach e Rød [49]

De acordo com Ho e Jern [22] dados multivariados que contêm informações de espaço e tempo são difíceis de serem analisados através de um único gráfico. Para que padrões possam ser encontrados, é conveniente que os dados sejam visualizados por intermédio de múltiplos métodos de representação visual, no qual cada um é responsável por realçar diferentes informações. Para isto, eles apresentam novas funcionalidades desenvolvidas para o GAV (*GeoAnalytics Visualization*) [26]⁶ baseadas nos princípios de visualização analítica, para suportar eventos espaço-temporais que possuem localização geoespacial e dados multivariados. O objetivo é utilizar técnicas de *coordinated multiple views* (CMV) para permitir a manipulação de dados através da interface e permitir que o processo de visualização analítica se torne mais interativo e efetivo.

A Figura 3.3 apresenta um estudo de caso com dados de temperatura e salinidade das água do oceano fornecidos pelo NOAA (*National Oceanic and Atmospheric Administration*-EUA). Cinco representações gráficas são apresentadas, permitindo comparar informações de temperatura e salinidade da água através de um *heatmap*, um gráfico dispersão e dois gráficos de área que exibem informações de profundidade e temperatura dos oceanos. Neste exemplo, o usuário selecionou uma área de interesse no mapa (norte da América do Sul) e as informações referentes a esta localidade são exibidas simultaneamente nos demais gráficos. A correlação entre temperatura e salinidade para um período de tempo específico pode ser analisada através do gráfico de dispersão e dos gráficos de área. As

⁶Ferramenta que oferece um conjunto de componentes para a visualização de dados analítica. Foi criada em 2007.

técnicas de visualização e interação podem ser aplicadas a conjuntos de dados que apresentem as mesmas características do estudo de caso, com informações de espaço e tempo, e permitem que até 4 atributos sejam visualizados de forma simultânea.

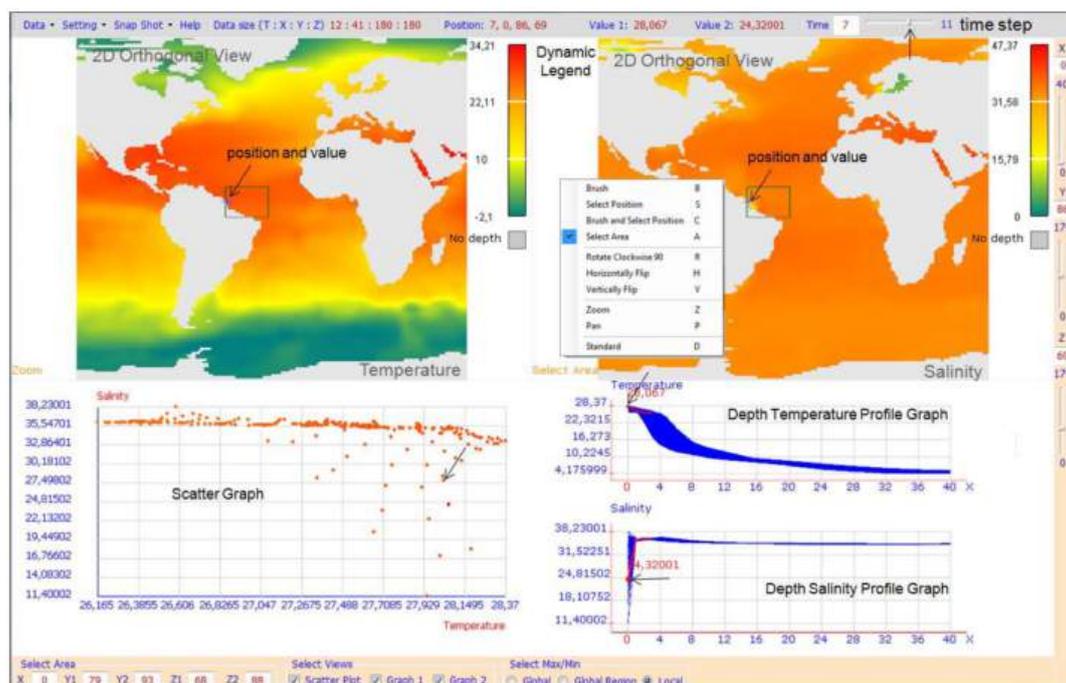


Figura 3.3 – Análise exploratória de dados geoespaciais em 3D. Fonte: Ho e Jern [22]

GAV Flash é um *framework* desenvolvido por Van Ho et al. [66] a partir do GAV [26]. O GAV Flash foi elaborado com o objetivo de diminuir o tempo levado para a customização de visualizações analíticas. A sua arquitetura de componentes em camadas oferece diversas representações geográficas interativas. Entre as técnicas de interação estão filtros, seleção de áreas no mapa, *brushing*, *zoom* e a movimentação do mapa. Os pesquisadores apresentam o uso do *framework* através de um estudo de caso com dados econômicos do OECD *Organization for Economic Co-operation and Development* (Figura 3.4). A visualização criada para exibir este conjunto de dados conta com um mapa *choropleth*, um gráfico de *scatterplot* e um gráfico de coordenadas paralelas. Estes gráficos são separados por divisórias que permitem que seu tamanho seja ajustado conforme a preferência do usuário. O *framework* permite a exibição de até treze atributos de forma simultânea e pode ser aplicado a diversos contextos, por exemplo: dados públicos, ambientais, logísticos, meteorológicos, ciências sociais e educacionais. A entrada de dados se dá por arquivos no formato CSV ou Excel, o qual é carregado para um banco de dados interno que é utilizado no tempo de execução. Ele foi projetado para lidar com dados em três dimensões, os quais são representados por atributo, espaço e tempo.

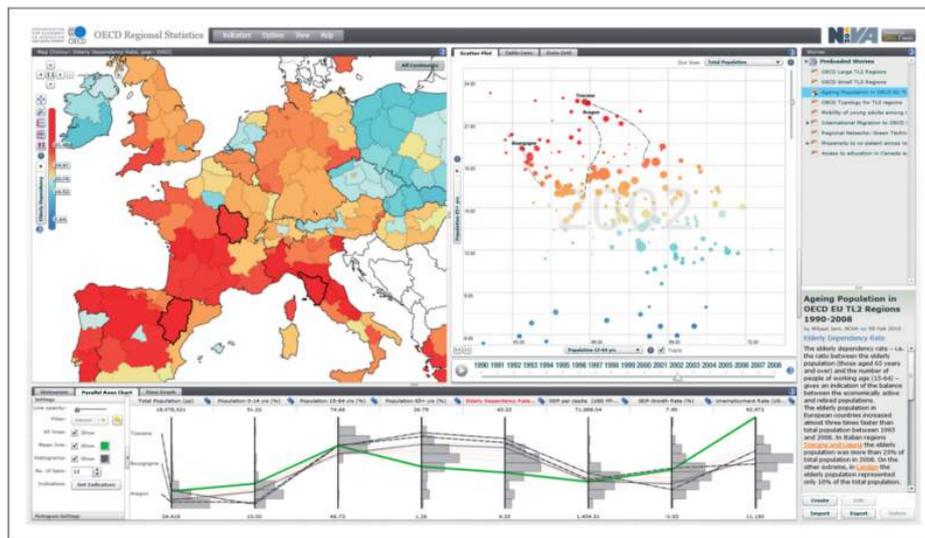


Figura 3.4 – Kit de ferramentas para a visualização de dados geoespaciais na web. Fonte: Van Ho et al. [66]

Jern [25] demonstra, através de um estudo de caso, que o GAV Flash [66] facilita a visualização de informações sócio-econômicas através de uma visualização geoespacial detalhada, que permite a análise das diferenças regionais e do desempenho dentro de um país, além de uma comparação de diferentes áreas do mesmo país.

O estudo de caso, apresentado na Figura 3.5, utiliza dados de ensino da União Europeia. O indicador de cores representa o percentual de trabalhadores com ensino superior completo. As regiões da Espanha, Bélgica e Noruega têm o nível mais alto de educação entre os trabalhadores. Estas regiões podem ser comparadas através de um gráfico de coordenadas paralelas. A Noruega apresenta a maior taxa de desemprego e uma baixa taxa de educação básica. Espanha tem o menor PIB. O gráfico de dispersão mostra a correlação dos gastos com pesquisa e desenvolvimento e o percentual de trabalhadores com ensino superior. Através do estudo de caso, 4 atributos podem ser visualizados de forma simultânea e o próprio estudo de caso sugere que as técnicas de visualização e interação presentes no GAV Flash [66] podem ser aplicadas a diversos contextos.

Lundblad e Jern [45] criaram uma versão do GAV Flash [25] em HTML5, com o objetivo de disponibilizar as técnicas de visualização presentes no GAV para mais usuários através da web. A estrutura do GAV inclui técnicas de visualização como: mapa *choropleth*, gráfico de dispersão, gráfico de barras, gráfico de linhas, gráfico de coordenadas paralelas e *treemap*. Esta versão do GAV, desenvolvida em HTML5, também suporta técnicas de interação como *brushing and linking* e demarcação de áreas no mapa. Para o estudo de

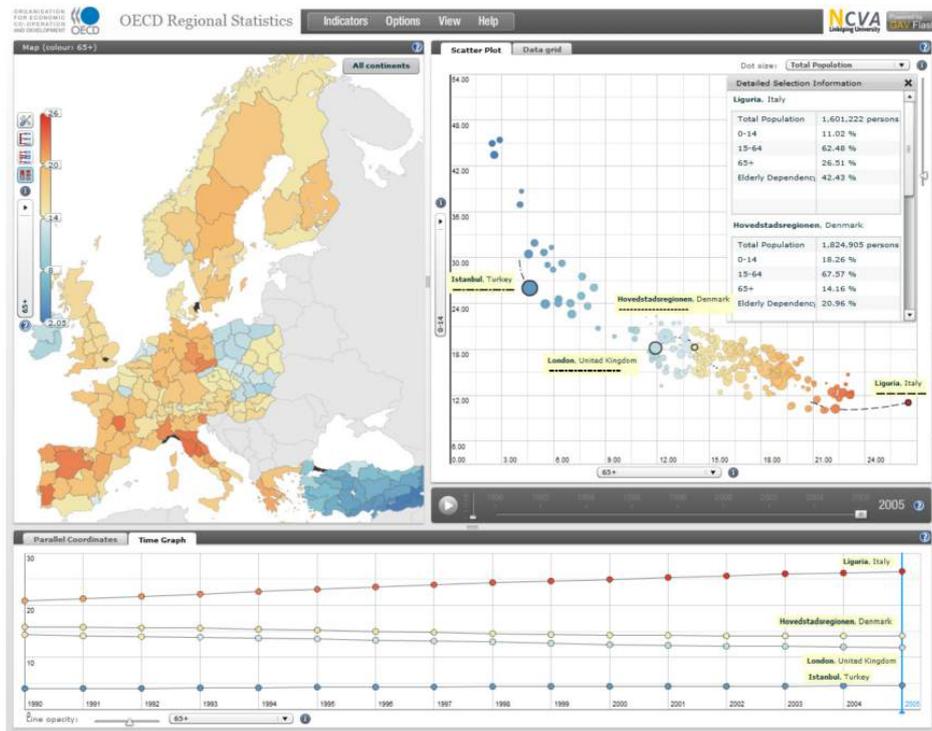


Figura 3.5 – Ferramenta colaborativa para análise de dados geoespaciais. Fonte: Jern [25]

caso, apresentado na Figura 3.6, foram utilizados dados estatísticos do Banco Mundial com o objetivo de demonstrar os problemas do envelhecimento da população entre os anos de 1960 e 2010. No mapa *choropleth*, os países são coloridos de acordo com o percentual da população com mais de 65 anos. Dentro de cada país foi adicionado um gráfico de pizza comparando a quantidade de pessoas entre 0-14 anos de idade com a quantidade de pessoas com idades acima de 65 anos. No lado direito, um gráfico *treemap* ilustra a população total dos países a partir do tamanho do retângulo e a quantidade de pessoas acima de 65 anos através das cores destes retângulos. As técnicas de visualização e de interação apresentadas podem ser aplicadas a diferentes contextos com dados multivariados, assim como o GAV [26], e permite a visualização de até 5 atributos de forma simultânea.

Turkay et al. [65] criaram um *framework* que tem como objetivo permitir a visualização da variação de múltiplos atributos geográficos de acordo com o tempo. Para isto, eles apresentam um gráfico chamado de *attribute signatures* (Figura 3.7). Este gráfico representa mudanças em um ou mais atributos ao longo do eixo de variação (eixo “x”), o qual corresponde ao aspecto geográfico (localização, extensão ou resolução) e é definido interativamente pelo usuário. As mudanças deste atributo ao longo do tempo são calculadas e representadas no eixo “y”. Podem ser exibidos quarenta e dois atributos de forma simultânea. A visualização é composta por um mapa, no qual podem ser empregadas as técnicas



Figura 3.6 – Análise visual de dados geoespaciais e *storytelling* utilizando HTML5. Fonte: Lundblad e Jern [45]

heatmap e *choropleth*; e outros quarenta e dois gráficos denominados *attribute signatures* (Figura 3.8). Os gráficos de *attribute signatures* podem ser alterados dinamicamente por gráficos de *sparkline*, *multiple sparkline*, barras e *multi-bar*. Como técnica de interação, o *framework* oferece *brushing and linking* entre o mapa e os gráficos. As técnicas de visualização presentes no *framework* podem ser aplicadas a diversos contextos com dados multivariados.

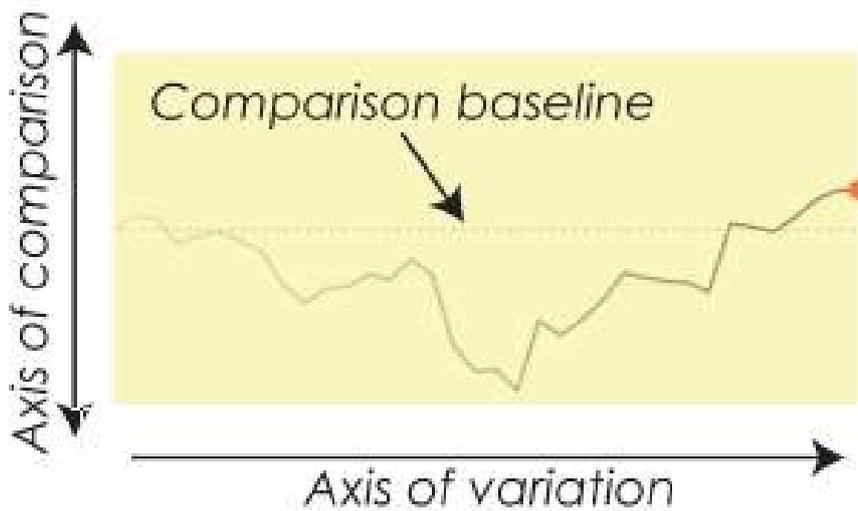


Figura 3.7 – Demonstração do gráfico proposto como *Attribute signatures*. Fonte: Turkyay et al. [65]

Liono et al. [42] apresentam uma estratégia de agregação de dados espaço-temporais, com o objetivo de permitir a exploração de dados multivariados sobre planejamento urbano. A solução proposta pelos pesquisadores permite que planejadores urbanos explorem da-

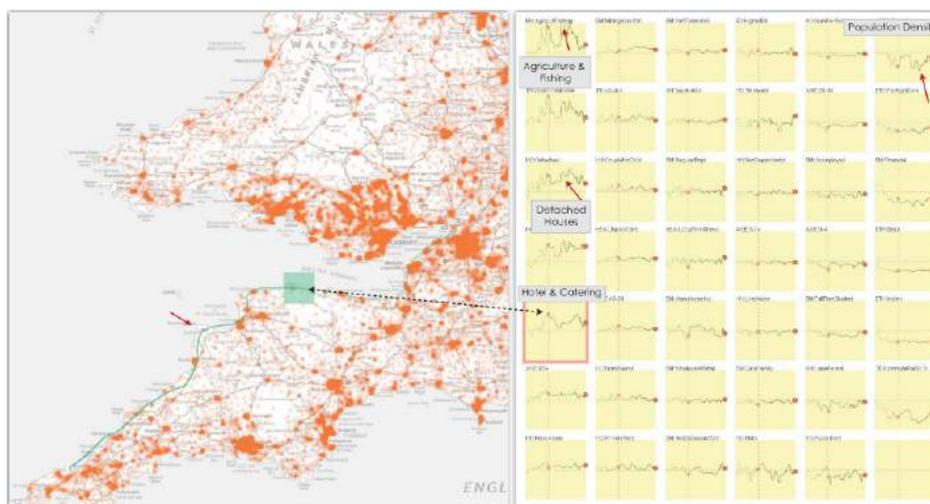


Figura 3.8 – *Attribute signatures*: técnica de visualização para dados geográficos multivariados. Fonte: Turkay et al. [65]

dos de maneira interativa, através da comparação visual e agregação de dados urbanos em um mapa *choropleth*, combinado a *volumetric extrusion*. *Extrusion* é uma técnica comum na visualização geoespacial 3D e já foi combinada com o *choropleth* anteriormente [57]. A altura do relevo varia de acordo com as propriedades definidas no conjunto de dados.

A entrada de dados se dá por conjuntos de dados em dois formatos: um arquivo *Shapefile* com as delimitações geográficas do sul da Austrália e outro arquivo no formato GTFS (*Google Transit Feed Specification*) com informações temporais sobre o trânsito desta localidade. As técnicas de visualização apresentadas no mapa são *heatmap*, *choropleth*, extrusão volumétrica, observação temporal (através de um *slider*) e agregação e dados (Figura 3.9). As técnicas de interação apresentadas são *zoom*, rotação do mapa e aplicação de filtros. Foi desenvolvido especificamente para o contexto de controle de dados urbanos e permite a visualização de três atributos de forma simultânea.

Liu et al. [43] projetaram uma ferramenta de visualização híbrida baseada em sistemas analíticos que tem como objetivo exibir dados estatísticos integrados com gráficos de coordenadas paralelas, mapa, gráfico de dispersão e *treemap*. Esta ferramenta possibilita a exploração de dados nas perspectivas de espaço, tempo e dimensão. É apresentado um estudo de caso com dados fornecidos pelo *Guotai'an CSMAR*⁷, contendo dados econômicos de 31 províncias e 4 municípios entre os anos de 1990 e 2009. Na Figura 3.10, são exibidos 14 indicadores sócio-econômicos através do método de visualização proposto. Este estudo de caso mostra que o sistema pode exibir e analisar dados estatísticos para encontrar pa-

⁷us.gtarsc.com/



Figura 3.9 – Visualização de dados espaço-temporais sobre gerenciamento urbano. Fonte: Liono et al. [42]

drões de desenvolvimento, e permite que os usuários compreendam o desenvolvimento da economia nacional.

A visualização é composta por um mapa *choropleth*, que apresenta dados estatísticos; um gráfico de coordenadas paralelas, que apresenta atributos multivariados; um gráfico *treemap*, que apresenta o relacionamento hierárquico de dados econômicos; e um gráfico de dispersão, que mostra a tendência de variação dos valores ao longo do tempo. A ferramenta proposta combina as técnicas de visualização com *coordinated multiple views* (CMV) e permite interação através da seleção dos dados. As técnicas de visualização apresentadas podem ser aplicadas a diversos contextos geoespaciais e permitem que até dezesseis atributos sejam apresentados de forma simultânea.

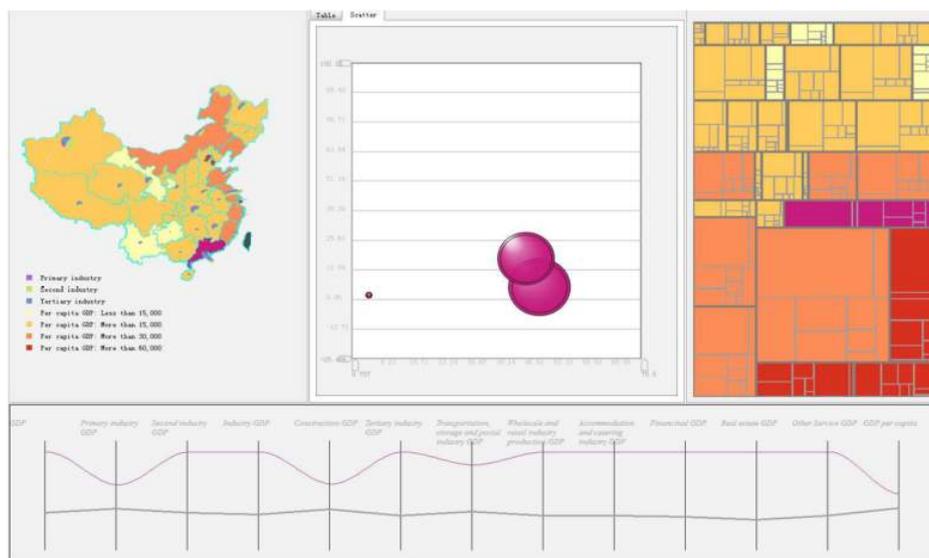


Figura 3.10 – Análise de dados estatísticos baseados em técnicas de visualização mistas. Fonte: Liu et al. [43]

Wongsuphasawat et al. [74] propõem uma ferramenta de análise visual de dados de incidentes de trânsito. A ferramenta, chamada Fervor, fornece ao usuário um conjunto de funcionalidades, que inclui filtros, funções de classificação estatística e recursos de exploração de dados multidimensionais. Na Figura 3.11 podemos ver a tela dividida em 3 seções. No lado esquerdo fica o painel de controle, que oferece opções de filtros para limitar a quantidade de informação exibida. O painel de controle também oferece a opção de classificação de variáveis, na qual os usuários podem ordenar as variáveis ou os relacionamentos entre as variáveis a partir de diversos critérios. A área principal da visualização, no lado direito, é dividida de forma horizontal em duas seções redimensionáveis. A seção superior direita é dedicada ao mapa e oferece duas opções: os usuários podem escolher entre ícones, sendo cada incidente exibido como um círculo no mapa, ou *heatmap* para colorir as regiões no mapa. A seção inferior direita oferece uma visão tabular tradicional dos dados e três componentes de visualização: histograma, plotagem bidimensional (2D) e gráfico de coordenadas paralelas. O histograma mostra a distribuição da variável selecionada de acordo com o conjunto de dados. O gráfico 2D possibilita a comparação de duas variáveis através de um gráfico de dispersão e como um *treemap*. O gráfico de coordenadas paralelas fornece uma visão geral das relações de múltiplas variáveis.

A seleção de um determinado incidente, ou de um conjunto de incidentes, em um destes componentes aplica a mesma seleção aos demais gráficos. O estudo de caso apresentado utiliza dados urbanos do estado de *Maryland*-EUA e possibilita a visualização de até sete atributos de forma simultânea. As técnicas de visualização e interação apresentadas são específicas para dados de transporte.

Zhang et al. [77] desenvolveram um *framework* analítico que permite explorar e comparar o impacto de variações geográficas entre locais e escalas para *clusters* de dados multivariados. Marcadores são projetados sobre um mapa *choropleth* com o objetivo de permitir a análise de dependências visuais entre atributos multivariados. Como estudo de caso, foram utilizados dados sócio-econômicos do Censo de Bureau, EUA para demonstrar como as técnicas apresentadas auxiliam na análise e compreensão de dados multivariados, possibilitando a descoberta de padrões.

A Figura 3.12 apresenta um exemplo da exploração dos *clusters*. A Figura 3.12a apresenta um mapa *choropleth*, e os critérios utilizados para o agrupamento dos atributos são exibidos em uma miniatura quando o mouse é posicionado sobre a região. A

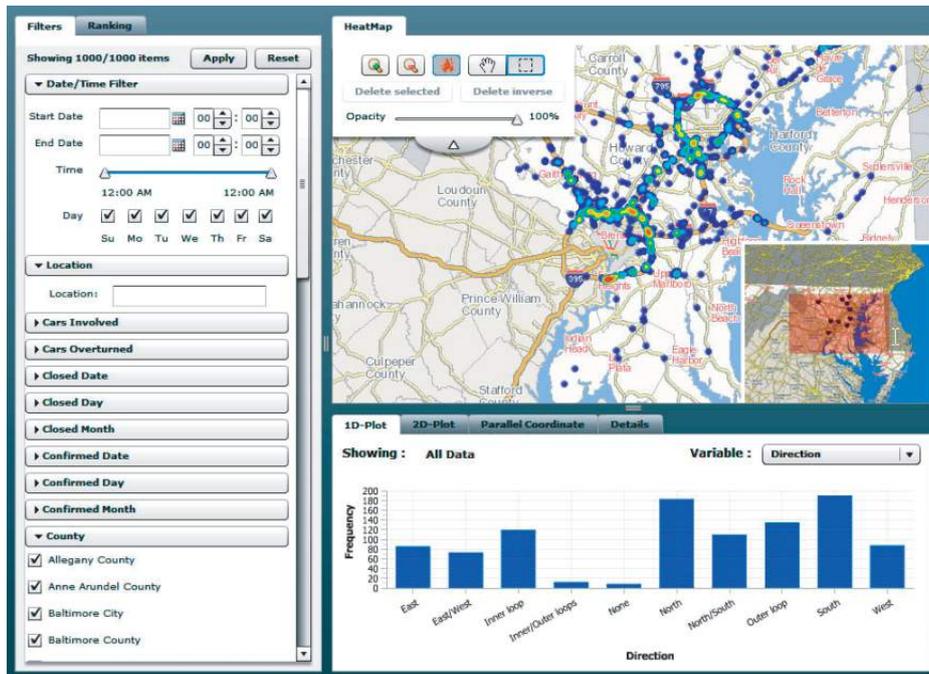


Figura 3.11 – Análise visual para conjuntos de dados de incidentes de trânsito. Fonte: Wongsuphasawat et al. [74]

Figura 3.12b apresenta um gráfico de dispersão dos atributos que são representados no mapa, e a Figura 3.12c mostra um gráfico de coordenadas paralelas com informações dos estados selecionados. A Figura 3.12d oferece um segundo gráfico de coordenadas paralelas com informações de todos os estados, e não apenas dos que foram selecionados. A Figura 3.12e oferece um terceiro gráfico de coordenadas paralelas com informações dos estados agrupados, e a Figura 3.12f apresenta um gráfico de radar com informações dos demais componentes.

As técnicas de visualização e interação presentes no *framework* podem ser aplicadas a outros contextos que possuem dados geoespaciais e permitem a visualização de até sete atributos ao mesmo tempo.

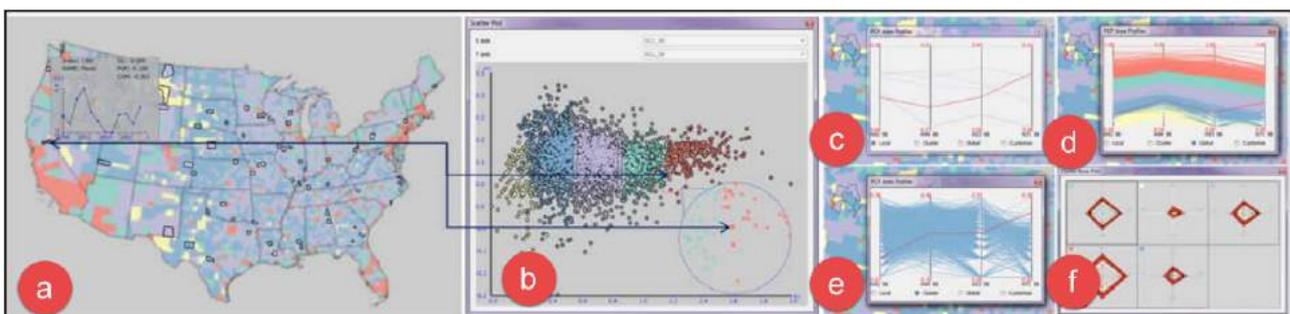


Figura 3.12 – Visualização do impacto das variações geográficas em *clusters* multivariados. Fonte: Zhang et al. [77]

Li et al. [40] apresentam o *GazaPlot*, um *framework* para visualização de dados analítica que proporciona uma visão detalhada de variações temporais. Os autores acreditam que uma visão completa e simultânea de padrões espaço-temporais é extremamente importante, por este motivo eles apresentam ambos na mesma visualização. Este *framework* consiste em um mapa *choropleth* no centro da visualização, um anel no entorno do mapa apresentando informações espaço-temporais (*ring-band*) e marcadores agrupados no entorno deste anel (*cluster ring*).

O mapa é apresentado no centro da visualização enquanto atributos temporais são associados à ele. Mecanismos interativos são utilizados para realçar os recursos selecionados, ajustando a área de exibição e criando uma visualização em *fisheye*. Na Figura 3.13 foram representadas as alterações climáticas que aconteceram entre os anos 2001 e 2012. Ao representar as estações climáticas, padrões climáticos ao longo do tempo puderam ser encontrados.

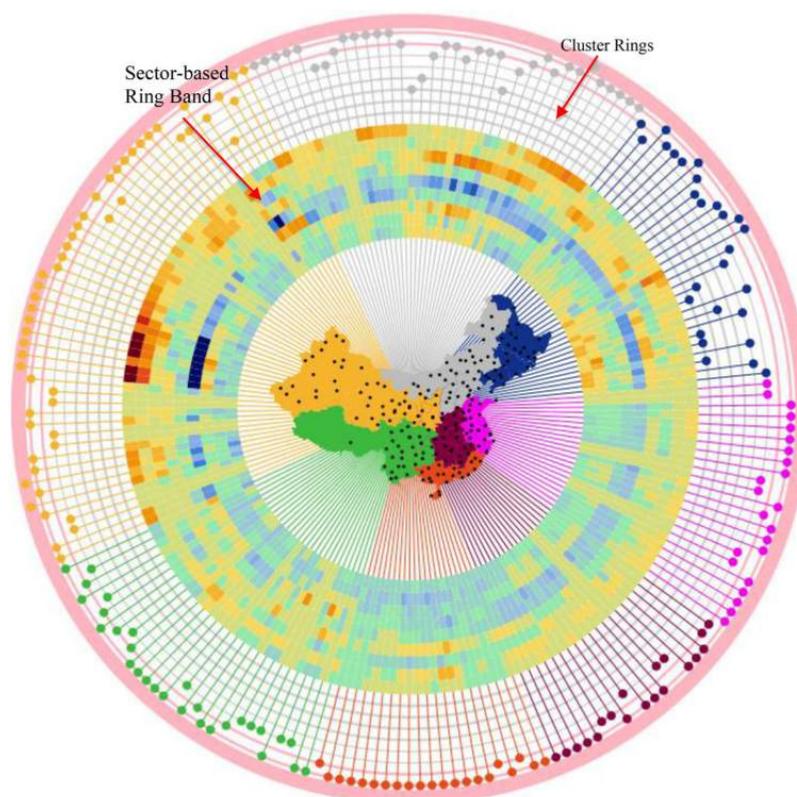


Figura 3.13 – GazaPlot: Visualização interativa de dados geocientíficos. Fonte: Li et al. [40]

Os pesquisadores Liao et al. [41] desenvolveram um sistema que permite a análise de dados climáticos, tendo como objetivo analisar a qualidade do ar. Para isto, eles utilizaram dados coletados em 8 estações de medição na região de Pequim, China. A Figura 3.14 ilustra as técnicas de visualização utilizadas para apresentar estes dados, na qual

cada estação é representada através de um gráfico de pizza (Figura 3.14a), posicionado no mapa de acordo com a sua localização e exibindo informações sobre os níveis de poluição em cada estação. Para representar informações temporais, foi utilizado um gráfico de dispersão (Figura 3.14b), exibindo a evolução dos índices de poluição ao longo do tempo. Para possibilitar a exibição de dados multidimensionais, os pesquisadores combinaram as duas visualizações anteriores com um gráfico de coordenadas paralelas (Figura 3.14c). O gráfico de coordenadas paralelas foi escolhido por permitir a correlação das informações de poluentes e onde eles estão concentrados. Ao todo, 8 atributos podem ser exibidos e analisados de forma simultânea, e a visualização foi projetada especificamente para apresentar informações sobre a qualidade do ar.

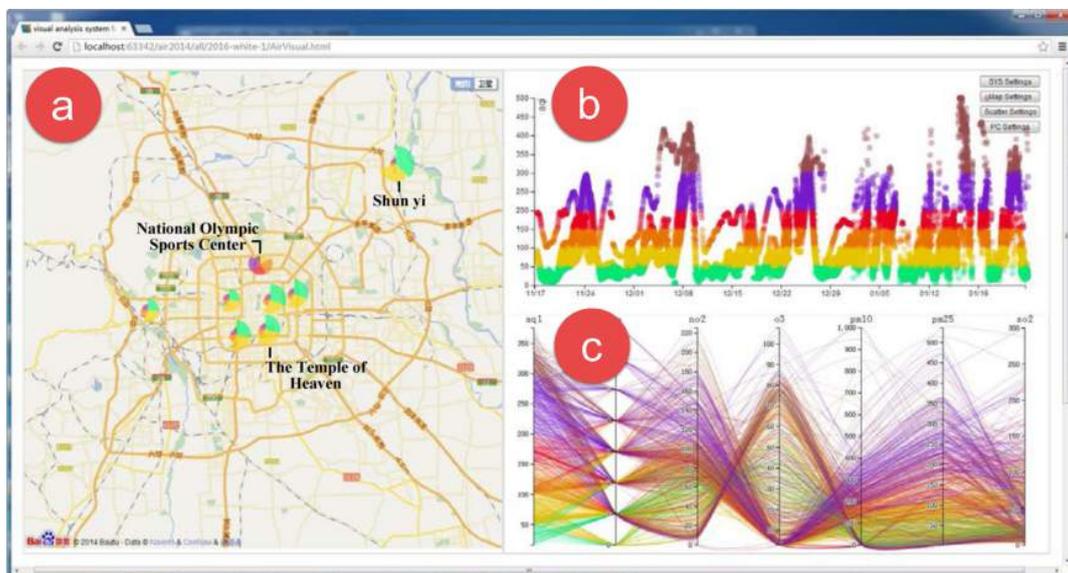


Figura 3.14 – Visualização analítica para monitorar a qualidade do ar. Fonte: Liao et al. [41]

Zhang et al. [76] apresentam um ambiente para visualização analítica que permite a exploração de dados de incidentes relacionados à desastres naturais em tempo real. Estas informações são divididas em categorias e visualizadas através de marcadores interativos, os quais preservam o contexto a medida que ocorre interação com estes marcadores. Na Figura 3.15, podemos ver o painel de controle (a), onde algumas opções do que está sendo visualizado pode ser alterado; um gráfico de linhas exibindo informações temporais (b); uma árvore de categorias (c); uma tabela (d); e o marcador posicionado no mapa (e). O marcador é um gráfico de pizza e, conforme o mouse é posicionado sobre este marcador, é exibida uma lente, a qual é segmentada em setores uniformemente espaçados que correspondem às categorias exibidas no gráfico de pizza. Permite a representação de 8

atributos de forma simultânea e foi desenvolvida especificamente para conjuntos de dados sobre desastres naturais.

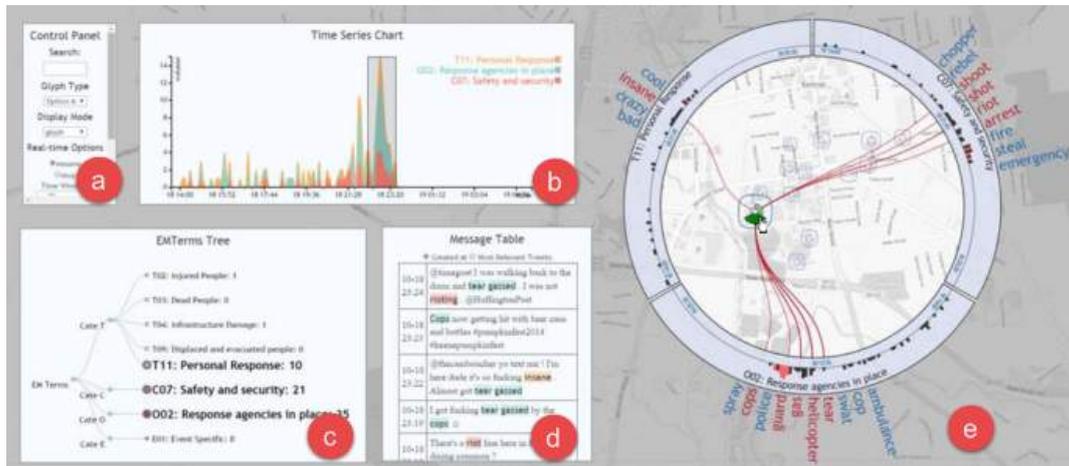


Figura 3.15 – Um *framework* para visualização analítica para dados de incidentes relacionados à desastres naturais. Fonte: Zhang et al. [76]

Por fim, outro *framework* analítico é apresentado por Li et al. [39]. Conforme ilustra a Figura 3.16, este *framework* consiste em 3 componentes: um que apresenta a visualização dos registros de acordo com a localização e o tempo que eles ocorreram (Figura 3.16b); Gráficos de barras que permitem a seleção destes valores de acordo com o tempo e os atributos que foram selecionados (Figura 3.16a); e outros gráficos de barras que permitem ver a distribuição dos valores por país (Figura 3.16c). O objetivo dos pesquisadores é possibilitar a exploração de acontecimentos em diferentes localizações de forma simultânea, permitindo ver a distribuição dos eventos ao longo do tempo. Ao todo, cinco atributos diferentes podem ser visualizados no mapa de forma simultânea e as técnicas apresentadas podem ser aplicadas a outros contextos.

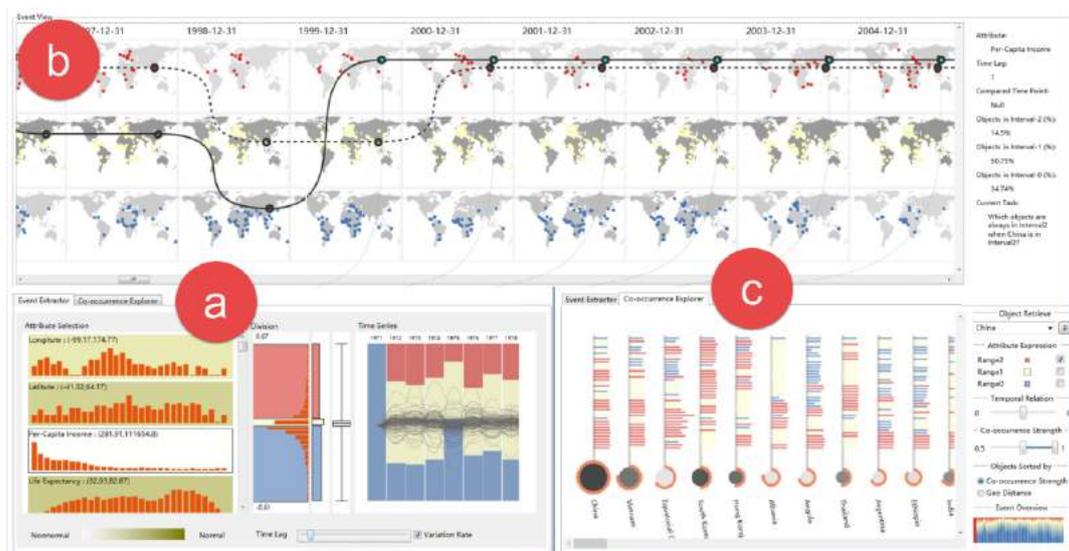


Figura 3.16 – COPE: Exploração interativa de padrões espaço-temporais em diferentes localizações. Fonte: Li et al. [39]

3.7 Análise dos Resultados

Considerando os trabalhos analisados e as questões de pesquisa definidas no Capítulo 3.1, observa-se que diversas técnicas de visualização e interação têm sido utilizadas para representar dados multivariados em mapas. Foram considerados os 12 trabalhos incluídos na RSL e os 3 trabalhos selecionados no *snowballing*, totalizando 15 trabalhos. A Tabela 3.4 apresenta as técnicas de visualização e interação presentes nos artigos selecionados, ou seja, em quantos trabalhos cada técnica apareceu, respondendo, assim, à primeira questão de pesquisa (QP1 - “Quais técnicas de visualização geoespacial estão sendo utilizadas para apresentar dados multivariados?”).

Podemos ver que a técnica mais utilizada é o *choropleth*, que aparece em 7 dos 15 trabalhos. Os autores Opach e Rød 3.2 defendem que o mapa *choropleth* deve ser utilizado para apresentar dados relativos através do mapa [24]. Porém, embora permitam a representação de dados relativos, apenas um atributo pode ser visualizado do mapa. Para tornar possível a representação de dados multivariados, Leonowicz [38] sugere a combinação dos mapas *choropleth* com outras técnicas de visualização. Para representar dados multivariados, Opach e Rød 3.2 combinaram o mapa com um gráfico de coordenadas paralelas. Andrienko e Andrienko [3] ressaltam os benefícios da combinação de mapas *choropleth* com gráficos de coordenadas paralelas.

Para possibilitar a representação de dados multivariados, os autores Jern [25], Lundblad e Jern [45] e Van Ho et al. [66], combinaram o mapa *choropleth* com gráficos de dispersão, coordenadas paralelas e histograma.

Liu et al. [43] também combinaram técnicas de visualização com o objetivo de apresentar dados multivariados. O sistema projetado por eles possui um mapa *choropleth* associado à gráficos *treemap*, dispersão e coordenadas paralelas. O objetivo desta combinação é proporcionar uma análise visual a partir de diferentes perspectivas visuais, de forma que dados temporais, espaciais e multivariados são apresentados na mesma visualização. Também com o objetivo de apresentar dados multivariados de forma simultânea, Li et al. [40] combinaram o mapa *choropleth* com um gráfico *treemap* no formato circular, onde as informações são realçadas de acordo com a interação do mouse.

Outra técnica de visualização geoespacial mencionada é o *heatmap*, encontrada em 3 dos 15 trabalhos selecionados. Wongsuphasawat et al. [74] utilizaram *heatmap* para representar dados de transporte, e acreditam que esta técnica auxilia a revelar a quantidade de ocorrências em regiões específicas do mapa. Embora o *heatmap* apresente informações de apenas uma atributo, os pesquisadores combinaram a sua utilização com gráficos de histograma, dispersão e coordenadas paralelas para possibilitar a visualização simultânea de dados multivariados.

A técnica de *marker cluster* não foi encontrada no trabalhos selecionados. Apenas 5 dos 15 trabalhos utilizaram marcadores no mapa para representar a localização específica de um determinado evento ou dado, porém, em nenhum deles estes marcadores são agrupados conforme ocorre interação com o mapa. No trabalho de Lundblad e Jern [45] podemos ver que há sobreposição dos marcadores e, este problema também iria ocorrer nos demais trabalhos [43, 78, 39, 41] caso a densidade de marcadores no mapa fosse muito alta [47]. Nestes casos, a técnica de *marker cluster* ajudaria a mitigar o problema de sobreposição dos marcadores.

As técnicas de interação presentes na Tabela 3.4 respondem à segunda questão de pesquisa (QP2 - “Quais técnicas de interação estão sendo utilizadas com as técnicas de visualização identificadas na QP1?”). A técnica de interação que os autores mais utilizaram é *brushing and linking*, presente em todos os trabalhos.

Jern [25] utilizou *brushing and linking* e filtros estáticos com o objetivo de auxiliar na análise das informações apresentadas na visualização. Wongsuphasawat et al. [74] utili-

Técnica		Total
Visualização de dados geoespaciais	<i>Choropleth</i>	7
	<i>Heatmap</i>	3
	<i>Dotmap</i>	4
	Marcadores	5
Visualização associada ao mapa	Coordenadas paralelas	5
	Gráfico de dispersão	6
	Gráfico de barras	4
	Gráfico de linhas	6
	<i>Cluster ring</i>	1
	<i>Ring-band</i>	1
Interação	<i>Treemap</i>	1
	<i>Brushing and linking</i>	13
	<i>Multiple views</i>	13
	<i>Zoom</i>	7
	Filtros estáticos	6
	Movimentação no mapa	10
Foco+Contexto	3	

Tabela 3.4 – Técnicas de visualização e interação encontradas na RSL e no *snowballing*.

zaram para que os elementos selecionados no gráfico de coordenadas paralelas sejam refletidos no mapa e vice-versa. Van Ho et al. [66] escolheram esta técnica por ela representar o relacionamento entre os diferentes elementos da visualização de forma interativa. Ainda, Zhang et al. [78] utilizaram *brushing and linking* para integrar dois componentes visuais. Desta forma, as operações realizadas em um destes componentes são automaticamente refletidas no outro.

A técnica CMV foi bastante utilizada, aparecendo em 13 dos 15 trabalhos. Opach e Rød [49] a utilizaram para vincular um mapa *choropleth* com um gráfico de coordenadas paralelas. Os autores classificam a utilização de múltiplas visualizações como bem sucedida, uma vez que os usuários podem analisar os dados no contexto geográfico ou no contexto de um atributo específico. Independentemente de qual visualização o usuário preferir, as visualizações irão proporcionar respostas simultâneas. Liao et al. [41] também utilizaram múltiplas visualizações, sendo elas um mapa com marcadores, um gráfico de coordenadas paralelas e um gráfico de dispersão. Para permitir interação entre os componentes, foram utilizadas as técnicas *brushing and linking*.

Todos os trabalhos apresentam interações partindo dos gráficos para o mapa, isto é, os operadores de filtros e seleções são aplicados nos gráficos e são apenas refletidos nos mapas. A única interação existente que tem como ponto de partida o mapa é a seleção de uma região específica, resultando em filtros nos demais gráficos de acordo com a região

selecionada no mapa. Embora 10 trabalhos ofereçam a movimentação (*panning*) [46] do mapa como técnica de interação (Tabela 3.4), nenhum dos trabalhos a utilizou para aplicar filtros a partir das interações com o mapa.

Em resposta à terceira questão de pesquisa (QP3 - “As técnicas de visualização e interação apresentadas são aplicáveis a diferentes contextos?”), observamos que os trabalhos foram concebidos com diferentes propósitos. Entre os trabalhos analisados, 6 buscaram as melhores técnicas de visualização para representar um contexto específico [49, 42, 74, 77, 41, 76], enquanto em 9 deles foram criadas visualizações genéricas que podem ser aplicadas a diferentes contextos.

Entre estes 9 trabalhos, metade está relacionado ao GAV (*GeoAnalytics visualization*), proposto em 2007 por Jern e Johansson [26]: Ho e Jern [22] implementaram o suporte à *coordinated multiple views* (CMV); Van Ho et al. [66] implementaram uma versão utilizando a tecnologia Flash; Jern [25] apresentou um estudo de caso utilizando o GAV Flash [66]; e Lundblad e Jern [45] criaram uma nova versão utilizando a tecnologia HTML5. Os quatro trabalhos estão relacionados à mesma ideia original (GAV [26]) e a entrada de dados não foi especificada. Porém, como diferentes tipos de gráficos são oferecidos, o usuário pode escolher quais técnicas de visualização melhor representam os dados e, então, gerar a visualização.

Nos demais trabalhos [78, 43, 65, 40, 39], são apresentadas ferramentas analíticas que possibilitam que dados de múltiplos contextos sejam representados a partir de suas visualizações. Embora múltiplas técnicas de visualização tenham sido utilizadas, elas são estáticas e não podem ser alteradas com base no atributo que será exibido.

Apenas 5 trabalhos especificaram como funciona a entrada ou a leitura dos dados utilizados nas visualizações, e se os usuários podem, através de alguma interface, definir quais propriedades desejam utilizar para gerar a visualização. O GAV [26] utiliza conjuntos de dados no formato XML (*Extensible Markup Language*) para gerar as visualizações. Logo, os trabalhos encontrados que utilizaram esta ferramenta [22, 66, 25, 45] trabalham com este formato arquivo. Entretanto, não foram mencionadas quais informações precisam constar neste conjunto de dados para que seja possível gerar a visualização. Para customizar a visualização os usuários devem estender a implementação do *framework* e definir qual base de dados desejam utilizar.

No *ViewExposed*, apresentado por Opach e Rød [49], a entrada de dados acontece através de 3 arquivos diferentes. Não foram especificadas as informações que cada arquivo deve conter para que a visualização seja criada.

Embora a maior parte dos trabalhos tenha sido concebida para que dados de diferentes contextos possam ser representados, nenhum deles define quais informações devem constar nos arquivos de entrada ou qual o limite de registros suportado.

Apesar das técnicas de visualização permitirem a visualização dos dados multivariados através de diferentes abordagens, a escolha da representação gráfica correta interfere diretamente na compreensão dos usuários à respeito do que é apresentado. De acordo com Korpela [35], para determinar qual tipo de gráfico será utilizado na visualização devemos considerar informações como: a quantidade de variáveis que serão exibidas, a quantidade de informações presente no conjunto de dados de entrada, se serão exibidas informações temporais, etc. Gráficos de barras são bons para fins comparativos, enquanto gráficos de linha possibilitam a visualização de padrões. Gráficos de pizza são bons para composições simples e não devem ser utilizados para representar comparações entre muitas variáveis. Considerando estes fatores, seria interessante ter a opção de alterar qual tipo de gráfico será utilizado para representar cada atributo.

Ainda existem limitações na área de visualização de dados. Embora diversas técnicas tenham sido abordadas na literatura, questões como: a representação de dados multivariados em mapas; aplicação de filtros a partir da movimentação do mapa; e a possibilidade de escolher o tipo de gráfico que irá representar cada atributo; ainda podem ser aprimoradas.

4. DESCRIÇÃO DO MODELO PROPOSTO

O Dicionário Aurélio [16] define o termo “modelo” como algo “que serve ou pode servir de modelo ou de exemplo”. De acordo com Justi [27], “atualmente, o ponto de vista mais aceito é que um modelo é uma representação de uma ideia, objeto, acontecimento, processo ou sistema, criado com um objetivo específico”. Considerando estas definições, o modelo proposto neste trabalho foi elaborado com o objetivo de oferecer uma alternativa para a análise visual de dados multivariados, possibilitando que suas visualizações possam ser empregadas a conjuntos de dados de diferentes contextos.

Como o foco é o uso de dados multivariados, foi necessário pensar em uma estratégia de visualização que represente e correlacione a maior quantidade possível de informação de forma simultânea, bem como oferecer alternativas para manipular estes dados. Além disso, também existe o desafio de oferecer mais informações conforme ocorrem as interações com o mapa ou com os gráficos associados à ele.

O modelo desenvolvido permite explorar e analisar dados geoespaciais com o propósito de auxiliar em tomadas de decisão, através da combinação de diferentes técnicas de visualização e interação. Interações feitas em qualquer uma das representações visuais são automaticamente refletidas nas demais representações visuais. O modelo também foi projetado para que as suas visualizações e o seu comportamento possam ser empregados a diferentes conjuntos de dados multivariados associados a uma localização geográfica.

As seções a seguir apresentam o projeto do modelo proposto, incluindo a descrição da sua entrada de dados, do seu funcionamento e dos seus componentes visuais.

4.1 Entrada de Dados

O modelo foi projetado de forma que conjuntos de dados provenientes de diferentes contextos possam ser submetidos às mesmas técnicas de visualização e interação. Para isto, optamos por uma entrada de dados no formato tabular, no qual as duas primeiras colunas contêm uma posição geográfica (latitude e longitude) e as demais colunas contêm uma propriedade associada a esta localização. Quanto ao formato, consideramos que estes dados podem estar disponíveis em arquivos no formato CSV (*Comma-Separated Values*),

que foi escolhido por ser amplamente utilizado [59]. Por exemplo, a maior parte dos dados abertos do governo brasileiro¹ disponibilizados atualmente e utilizados como estudo de caso neste trabalho, estão disponíveis neste formato.

Além de estarem no formato csv, os dados de entrada devem ter seus atributos separados por vírgulas e as duas primeiras colunas, obrigatoriamente, devem informar as coordenadas geográficas (latitude e longitude) de cada registro. As demais colunas podem conter propriedades quantitativas e qualitativas de cada registro, as quais serão utilizadas na composição dos gráficos, conforme descrito na Seção 5.3.1.

4.2 Funcionamento do Modelo

Inicialmente, definimos um conjunto de etapas que devem ser seguidas para gerar a visualização de um conjunto de dados de entrada. Estas etapas estão ilustradas na Figura 4.1.

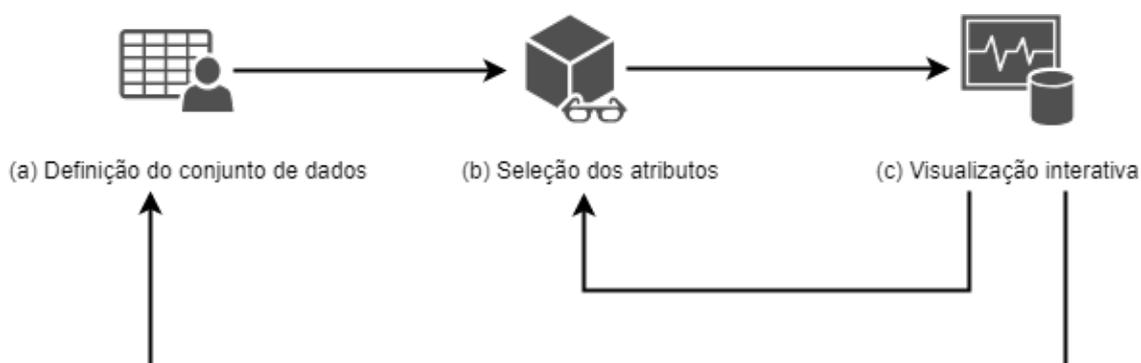


Figura 4.1 – Conjunto de etapas necessárias para geração da visualização.

A primeira etapa (Figura 4.1a) consiste na especificação do conjunto de dados que será utilizado. O modelo aceita conjuntos de dados multivariados, com informações geoespaciais e organizados em um formato tabular, conforme descrito na seção 4.1. Considerando que o modelo permite a representação de múltiplos atributos ao mesmo tempo, conjuntos de dados com uma grande quantidade de atributos são ideais para serem explorados através do modelo. A segunda etapa (Figura 4.1b) engloba a definição dos atributos que serão analisados e como cada um destes atributos será representado visualmente. A última etapa (Figura 4.1c) consiste na geração da visualização de acordo com as especifica-

¹dados.gov.br

ções feitas nas etapas anteriores. De acordo com as definições feitas na segunda etapa, um ou mais atributos serão agrupados no mapa enquanto outros atributos serão apresentados no demais gráficos. Por conseguinte, as definições de conjunto de dados (a), atributos que serão representados (b) e a geração visualização (c) são ubíquas e podem ser alteradas à qualquer momento.

O modelo foi projetado para suportar a visualização de dados multivariados, possibilitando a representação de até 18 atributos simultaneamente. Estes atributos podem ser dispostos da seguintes maneira: até 13 podem ser representados através dos gráficos que agrupam marcadores no mapa, e mais 5 podem ser representados em gráficos associados ao mapa. A Figura 4.2 exemplifica a disposição do mapa e dos gráficos no modelo. Neste exemplo, o mapa está localizado no lado esquerdo da janela de visualização, ocupando 50% do espaço total. No lado direito, os outros 50% do espaço são divididos em 6 blocos: o primeiro bloco exibe a quantidade total de registros sendo apresentados na visualização e um botão que permite alterar o tipo de gráfico utilizado para agrupar os marcadores no mapa. Os outros 5 blocos são espaços reservados para representar graficamente 1 atributo cada um. Estes gráficos podem ser, por exemplo, gráficos de barras, de pizza, *donut* ou de linhas.

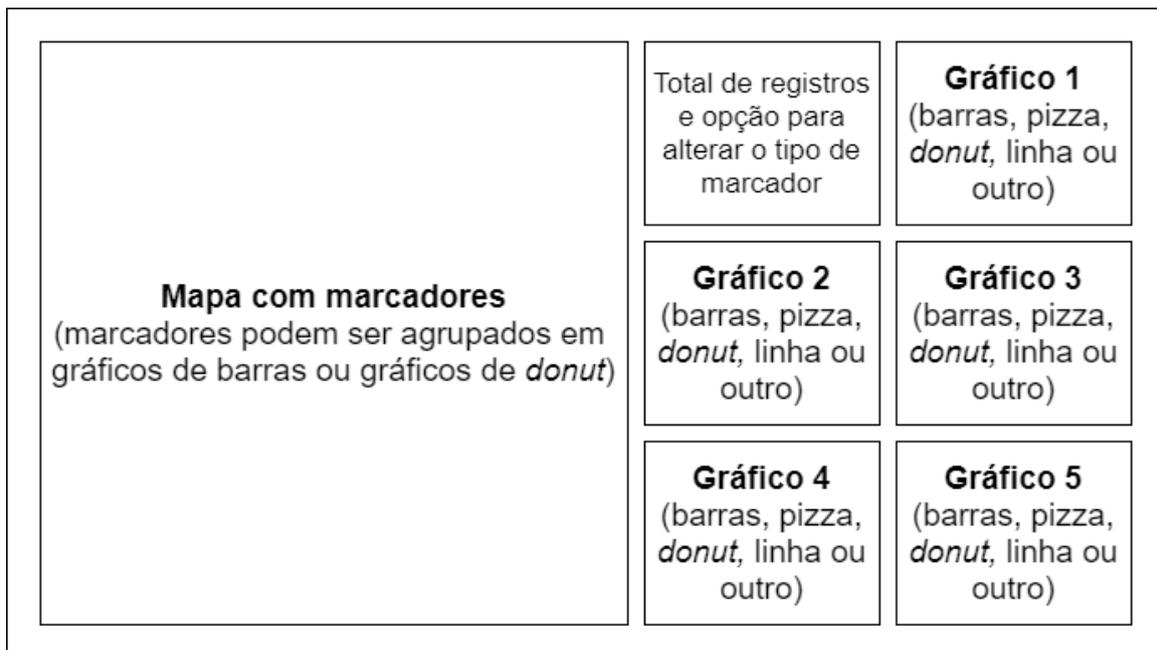


Figura 4.2 – Estrutura do modelo proposto com seus componentes visuais (mapa e gráficos).

4.3 Componentes Visuais

O modelo foi projetado para ser interativo e para possibilitar análise visual de dados multivariados. Portanto, ele deve suportar as técnicas de interação *brushing and linking* integradas à CMV. Desta forma, qualquer um dos gráficos presentes no modelo pode fazer o papel de um filtro para os demais, incluindo o mapa. De maneira análoga, o mapa também pode ser um filtro para os demais gráficos: a interação através do mapa utilizando as técnicas de *panning* e *zoom* funciona como a aplicação de um filtro nos demais gráficos, isto é, apenas os registros referentes ao perímetro que está sendo exibido no mapa são considerados para compor os gráficos. Quando o perímetro que está sendo exibido no mapa é alterado, todos os gráficos são atualizados de acordo com os dados presentes no mapa.

Considerando as limitações existentes na exibição de muitos marcadores no mapa, abordados na Subseção 2.1.4, os marcadores devem ser agrupados de acordo com a sua proximidade no modelo proposto. Assim, para possibilitar a exibição de dados multivariados, estes marcadores são exibidos através de gráficos de barras ou *donuts*. A escolha pelo gráfico de barras é porque, de acordo com os estatísticos Cleveland e McGill [12], este tipo de gráfico é eficaz para a comunicação de dados quantitativos e fornece uma estimativa precisa para a comparação de valores. Já os gráficos *donuts*, embora não apresentem os dados de maneira tão precisa [12], atraem mais a atenção dos usuários segundo Oshin et al. [67]. Por atraírem a atenção dos usuários, Gelman e Unwin [19] acreditam que é uma técnica eficaz para a visualização de informações.

No espaço reservado para a representação dos gráficos 1 a 5 no modelo (Figura 4.2), a escolha de qual tipo de gráfico será utilizado deve ser feita de acordo com o atributo em questão. Por exemplo, gráficos de pizza são bons para ilustrar percentuais relativos a um todo. Já o gráfico de linhas é a melhor escolha para representar dados temporais. Da forma como foi projetado, outros gráficos também poderiam ser usados no modelo, como, por exemplo, gráficos de coordenadas paralelas que permitiriam a representação de múltiplos atributos no mesmo espaço [23].

5. IMPLEMENTAÇÃO DO MODELO

Este capítulo apresenta a primeira implementação do modelo, seguida pela descrição da sua avaliação preliminar. A sua implementação final e a demonstração da sua aplicabilidade, através de 3 estudos de caso, também são apresentados.

5.1 Implementação Preliminar

Com o objetivo de validar o modelo proposto, foi feita uma primeira implementação para ele. A organização dos componentes visuais (mapa e gráficos) é a proposta na sua definição (Figura 4.2). O objetivo desta implementação é ter uma prova de conceito para ser validada com um grupo focal, descrita na próxima seção. Nesta etapa, o mapa suporta a representação de apenas um atributo com propriedades qualitativas, o qual é agrupado de acordo com a sua localização e representado através de gráficos de barras (Figura 5.3b). As opções de gráficos disponíveis são gráficos de barras, pizza e linhas (Figura 5.3a e c). Nesta etapa do trabalho, as técnicas de interação entre o mapa e os gráficos não foram completamente implementadas.

Embora fosse possível filtrar valores nos gráficos e ver os demais gráficos se comportando de acordo, interações feitas com o mapa não eram refletidas nos gráficos. Para validar esta implementação, foi utilizado um conjunto de dados abertos do governo sobre os acidentes de trânsito que aconteceram na cidade de Porto Alegre durante o ano de 2016 (veja a seção 6.1.1 para mais detalhes). Neste caso, os marcadores foram exibidos como gráficos de barras e ilustram a quantidade total de acidentes de acordo com a sua classificação (colisão, capotagem, engavetamento, etc.).

Na Figura 5.1a é apresentada a utilização da técnica *marker cluster* no mapa, no qual os marcadores são agrupados em gráficos de barras. Cada barra representa um tipo de acidente, de acordo com a legenda (Figura 5.1b). A quantidade de marcadores agrupados é indicada através de um número no centro de cada gráfico.

À medida que ocorre uma interação em um dos gráficos associados ao mapa, os outros gráficos são automaticamente redimensionados de acordo com a seleção feita. Sendo assim, uma vantagem desta forma de interação é a possibilidade de filtrar a infor-

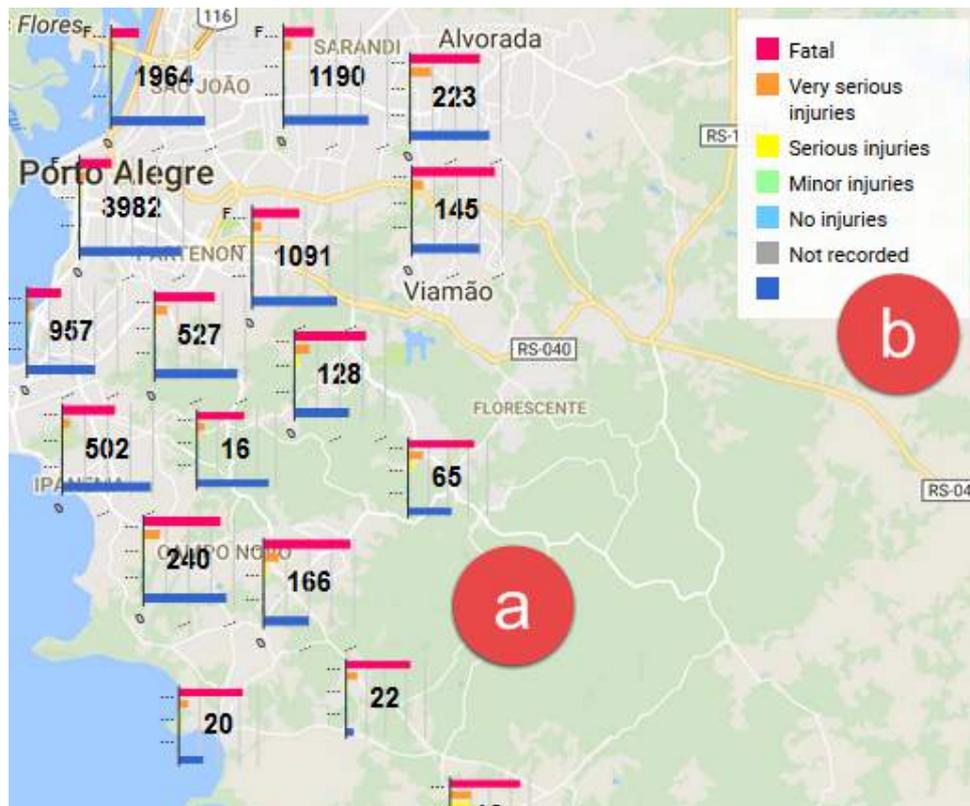


Figura 5.1 – Marcadores agrupados e representados através de gráficos de barras, no qual cada barra equivale a um tipo de acidente; (a) gráficos de barras agrupando os marcadores; (b) legenda dos tipos de acidente representados nos gráficos.

mação. A Figura 5.2 exemplifica a interação existente entre os gráficos. É demonstrado como os gráficos estavam antes da opção “DIA” ser selecionada no gráfico de pizza (Figura 5.2a). Após esta seleção, os demais gráficos foram redimensionados (Figura 5.2b). Através destes filtros, a visualização pode ser reduzida à intervalos ou regiões específicas.

A Figura 5.3 apresenta a aplicação desta implementação do modelo no mapa de Porto Alegre, também com o conjunto de dados de acidentes de trânsito (Seção 6.1.1). Os atributos tipo de acidente, dia da semana, horário e localização foram utilizados para compor os gráficos presentes na Figura 5.3a, resultando em dois gráficos de barras e dois gráficos de pizza, respectivamente. A data de acontecimento destes acidentes foi utilizada para criação do gráfico de barras presente na Figura 5.3c. Os atributos gravidade e localização foram utilizados para a exibição dos gráficos e marcadores no mapa, apresentados na Figura 5.3b.

As bibliotecas *Google Maps API*¹ e *D3.js* [9] foram utilizadas para a representação do mapa e gráficos de barras no mapa. Para permitir que os marcadores fossem agrupados

¹developers.google.com/maps/

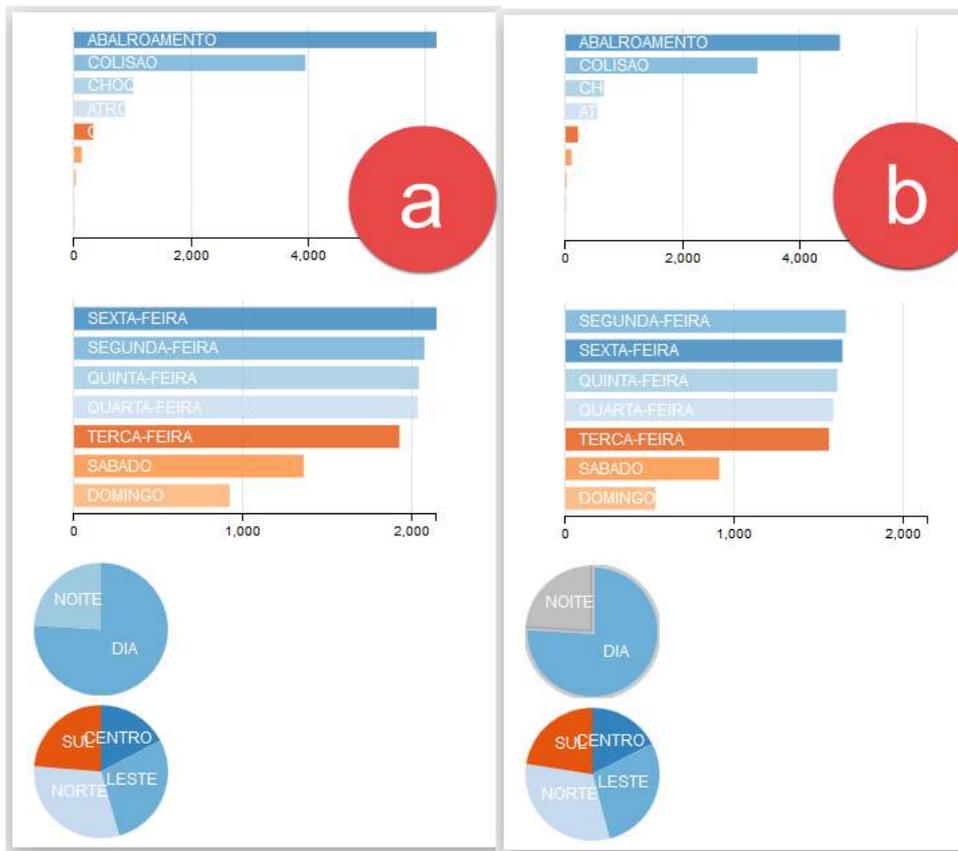


Figura 5.2 – Apresentação da interação existente entre os gráficos: (a) seleção da opção “DIA” no gráfico de pizza; (b) resultado da aplicação do filtro nos demais gráficos.

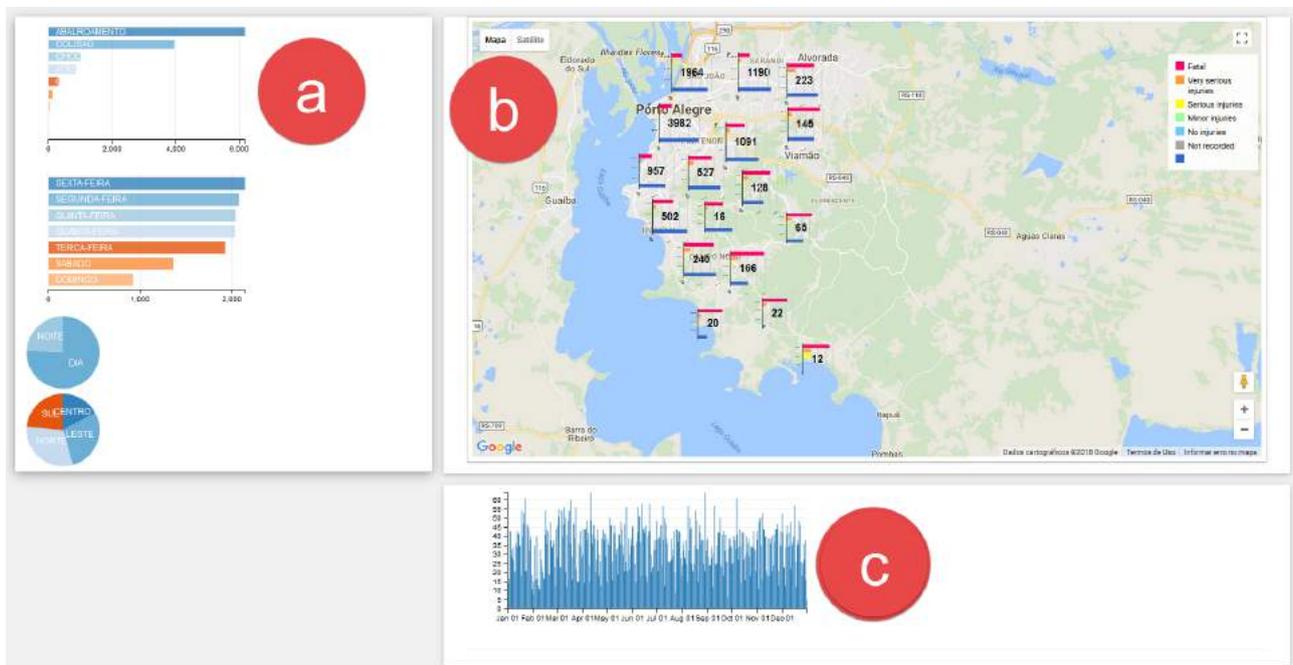


Figura 5.3 – Primeira versão implementada do modelo.

e representados no mapa através de um gráfico de barras, foi utilizada a biblioteca *Marker*

*Cluster*², fornecida pelo Google. A biblioteca *Crossfilter*³ foi utilizada para apresentar múltiplos gráficos encadeados e permitir que filtros sejam aplicados à todos eles de forma simultânea. Esta biblioteca permite que dados multivariados sejam filtrados e os gráficos sejam redimensionados.

5.2 Avaliação Preliminar

Visando avaliar quais técnicas de visualização podem auxiliar na compreensão dos dados, bem como validar o modelo proposto, foi realizado um grupo focal. Grupo focal é uma técnica de pesquisa qualitativa que, por meio de entrevistas em grupo, busca perspectivas diferentes sobre o mesmo assunto, gerando pluralidade de ideias [52, 48]. Neste caso, é importante que um moderador conduza a discussão entre os participantes, encorajando o engajamento de todos os participantes de forma equilibrada [58]. O perfil dos participantes, o roteiro que foi utilizado para conduzir o grupo focal e uma análise dos resultados obtidos são descritos a seguir.

Conforme mostra a Tabela 5.1, o grupo foi formado por 7 participantes, com idades entre 22 e 43 anos de idade, com diferentes perfis, mas que poderiam contribuir com diferentes pontos de vista para a discussão. Também participaram do grupo focal dois moderadores, pesquisadores da área de visualização de dados.

#	Idade	Gênero	Área de Formação	Foco de Pesquisa
P1	34	Masculino	Computação	Visualização de dados abertos
P2	35	Masculino	Geografia	Análise geoespacial e ciência de dados
P3	29	Masculino	Estatística e Economia	Estatística e dados abertos
P4	28	Masculino	Computação	Processamento paralelo e distribuído
P5	22	Feminino	Computação	Interação humano-computador
P6	43	Masculino	Computação	<i>Deep Learning</i>
P7	37	Masculino	Administração	Visualização de dados

Tabela 5.1 – Perfil dos participantes do grupo focal

Considerando o perfil dos participantes quanto a frequência de uso de ferramentas de visualização, destacam-se os participantes P1, P2, P6 e P7, que usam diariamente este tipo de ferramenta, sendo que o uso de mapas, especificamente, também é feito diariamente por P2 e P7. Estes participantes são também os que demonstraram um uso mais

²developers.google.com/maps/documentation/android-api/utility/marker-clustering

³square.github.io/crossfilter/

diversificado de ferramentas. P1 utiliza o Tableau⁴ e a D3.js para o desenvolvimento de visualizações, bem como a API do Google Maps⁵ e o Open Street Map⁶. O P7 também utiliza algumas ferramentas, sendo o Cognos Analytics e o Tableau para visualização, e o MapBox e MapInfo para mapas. Já P2 e P3 utilizam uma variedade maior de ferramentas. O P2 também utiliza o Tableau, além de PowerBI, Python e R para visualização, e diversas ferramentas para sensoriamento remoto e aeroespacial. Já o P3 utiliza para visualização Excel, Qlik Sense⁷, R/R Studio⁸, Google Mapas e Google Street View, tendo acrescentado ainda o Leaflet⁹ na lista de ferramentas para uso de mapas. Os usuários com menor frequência de uso de ferramentas (P4 e P5), restringem-se ao uso do Excel e do Google Mapas.

Quanto aos procedimentos para a realização do grupo focal, primeiramente, os participantes assinaram o termo de consentimento livre e esclarecido (Apêndice B) e preencheram o formulário sobre o seu perfil. Após isso, foram realizadas três atividades para introduzir o assunto: na primeira, os participantes tiveram que escolher uma técnica de visualização geoespacial que, na sua opinião, melhor representasse uma determinada informação; na segunda atividade os participantes fizeram uma análise de três ferramentas de visualização analítica (Qlik Sense, Oracle BI¹⁰ e Microsoft Power BI¹¹) visando identificar os detalhes mais significativos de cada uma delas; na terceira atividade os participantes foram divididos em duas duplas e um trio e ilustraram como representariam um determinado cenário, desenhando com canetas coloridas em um mapa impresso. Após a realização das atividades para contextualização sobre a visualização de dados geoespaciais, foi feita uma apresentação da implementação preliminar do modelo aos participantes, para, a partir disso, eles sugerirem melhorias e debaterem sobre as funcionalidades, interações e técnicas de visualização presentes implementadas. O objetivo desta demonstração foi coletar opiniões, sugestões e considerações para possíveis melhorias, que são descritas a seguir.

Todos participantes mencionaram identificar com facilidade as funcionalidades presentes no modelo, e todos concordaram que o modelo pode ser aplicado a conjuntos de dados relacionados à diferentes contextos. Dentre os possíveis contextos que o modelo

⁴www.tableau.com

⁵www.google.com/maps

⁶www.openstreetmap.org

⁷www.qlik.com/us

⁸Linguagem de programação e ambiente de desenvolvimento integrado para cálculos estatísticos e gráficos.

⁹<http://leafletjs.com>

¹⁰www.oracle.com/solutions/business-analytics/business-intelligence/index.html

¹¹powerbi.microsoft.com

poderia ser aplicado, três foram mencionados: (1) ocorrência de crimes por cidade, adicionando informações sócio-econômicas e proporcionando a comparação destes fatores entre diferentes áreas; (2) nível de educação no país, permitindo comparar o impacto da distribuição de recursos federais e; (3) a comparação de dados de manutenção pública com índices de inundação.

Um dos participantes mencionou que “praticamente qualquer conjunto de dados que contenha dados quantitativos e qualitativos poderia ser aplicado ao modelo”. Todos os participantes constataram que o modelo auxilia na análise dos dados. A integração da filtragem entre os gráficos e o mapa e a possibilidade de utilização da movimentação do mapa para aplicar filtros nos gráficos foram destacadas como os pontos mais relevantes do modelo.

Entre as sugestões de melhoria estão: (1) ajustar automaticamente a escala de eixo vertical ou horizontal dos gráficos; (2) incluir títulos dinamicamente, para que seja possível identificar facilmente o que cada gráfico representa; (3) aplicar filtros nos gráficos através da movimentação no mapa; (4) usar *heatmap* para representar atributos categóricos no mapa; e (4) substituir o gráfico de barras que agrupa os marcadores por um gráfico de *donut*, para que o espaço vazado no centro do gráfico permita visualizar o mapa atrás do gráfico.

5.3 Implementação final

Após coletar as opiniões e sugestões fornecidas pelo grupo focal descritas na seção anterior, foram aplicadas melhorias à implementação do modelo. A Figura 5.4 apresenta sua implementação final, composta por um mapa e mais cinco gráficos, dispostos de acordo com a Figura 4.2. Neste exemplo, foi utilizado como entrada o conjunto de dados abertos do governo descrito na Seção 6.1.1.

No exemplo da Figura 5.4, os marcadores são agrupados em gráficos de barras (b), representando tipos de acidente de acordo com a localização de cada registro. Quando o mouse é posicionado sobre os gráficos, o perímetro correspondente aos registros que estão sendo agrupados é destacado no mapa (a). Ao lado esquerdo do mapa (a), uma legenda identifica as barras de acordo com suas cores. Operações como *panning* e *zoom* estão disponíveis no mapa através da utilização do mouse.

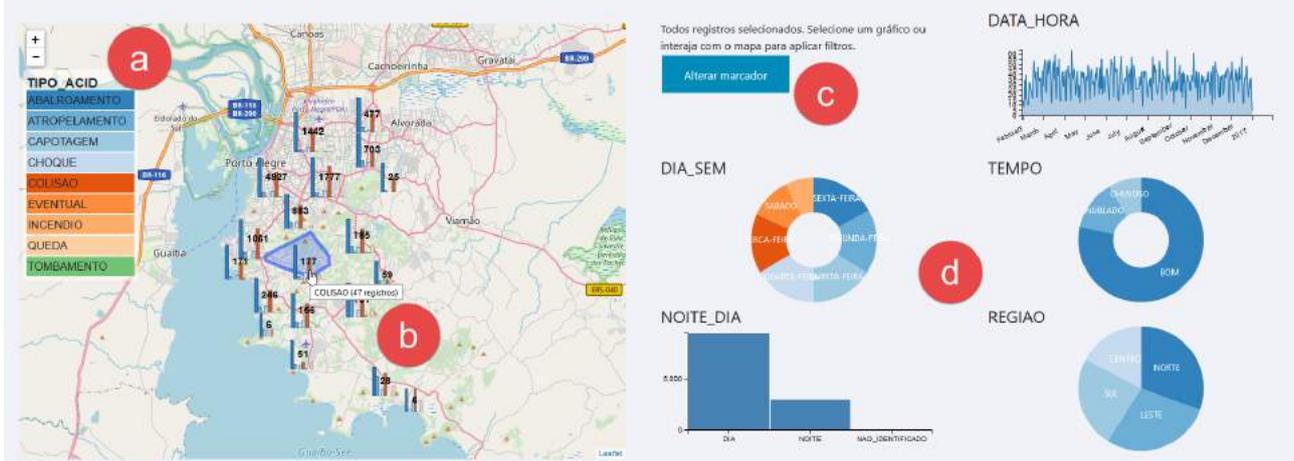


Figura 5.4 – Implementação final do modelo.

Considerando a implementação inicial do modelo (Seção 5.1), as principais alterações realizadas para atender às sugestões de melhoria provenientes do grupo focal foram: (I) todos os gráficos possuem um título/legenda (Figura 5.4d) descrevendo o que ele está representando (esta informação é composta pelo nome da coluna no arquivo de entrada); (II) a paleta de cores foi atualizada para utilizar cores neutras (Figura 5.4a), evitando associações enganosas entre atributos que podem não estar relacionados; (III) inclusão de um gráfico de *donuts* como alternativa para agrupar os marcadores (Figura 5.7a); e (IV) possibilidade de escolher o posicionamento e o tipo de gráfico que será utilizado de acordo com o atributo desejado (Figura 5.8). Além destas mudanças, as implementações que estavam pendentes na primeira versão foram concluídas, incluindo: o uso da movimentação do mapa com um filtro, de forma que apenas registros que atendem ao perímetro sendo exibido no mapa são considerados para compor os demais gráficos; aplicação de *coordinated multiple views* (CMV), de forma que cada componente visual (mapa ou gráficos), quando manipulado, ocasiona a aplicação de filtros nos gráficos e no mapa; e o perímetro utilizado para compor cada gráfico no mapa é destacado com uma linha azul quando o mouse é colocado sobre cada gráfico (Figura 5.4b).

A Figura 5.5 ilustra o funcionamento dos filtros na visualização. Neste exemplo, 4 filtros diferentes estão sendo aplicados de forma simultânea: sextas-feiras (Gráfico DIA_SEM); dias chuvosos (Gráfico TEMPO); noite (Gráfico NOITE_DIA); e zona norte (Gráfico REGIÃO). A quantidade de registros é reduzida à 10 e as diferenças podem ser observadas entre as Figuras 5.4 e 5.5.

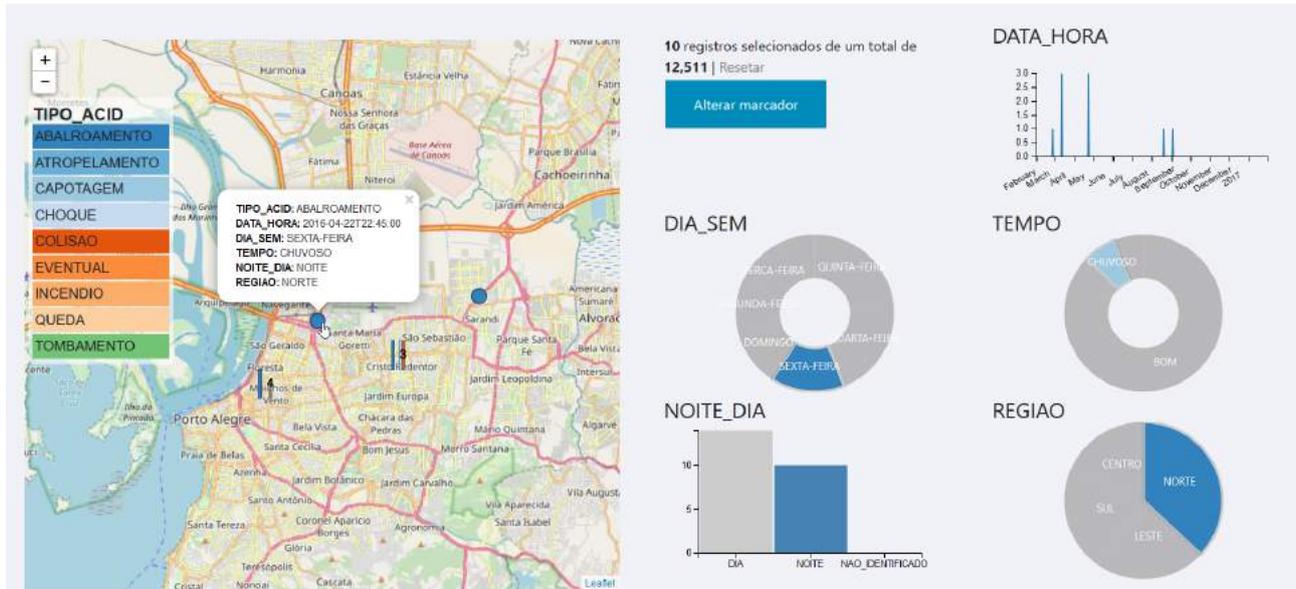


Figura 5.5 – Implementação final do modelo com filtros aplicados. São exibidos apenas acidentes ocorridos na zona leste.

A aplicação de filtros através da movimentação do mapa (*panning*) é demonstrada na Figura 5.6. Neste exemplo, o campo de visão é restringido para uma região que possui apenas um marcador. Pode-se observar, então, que os demais componentes da visualização são reajustados e redimensionados para corresponder à este marcador.

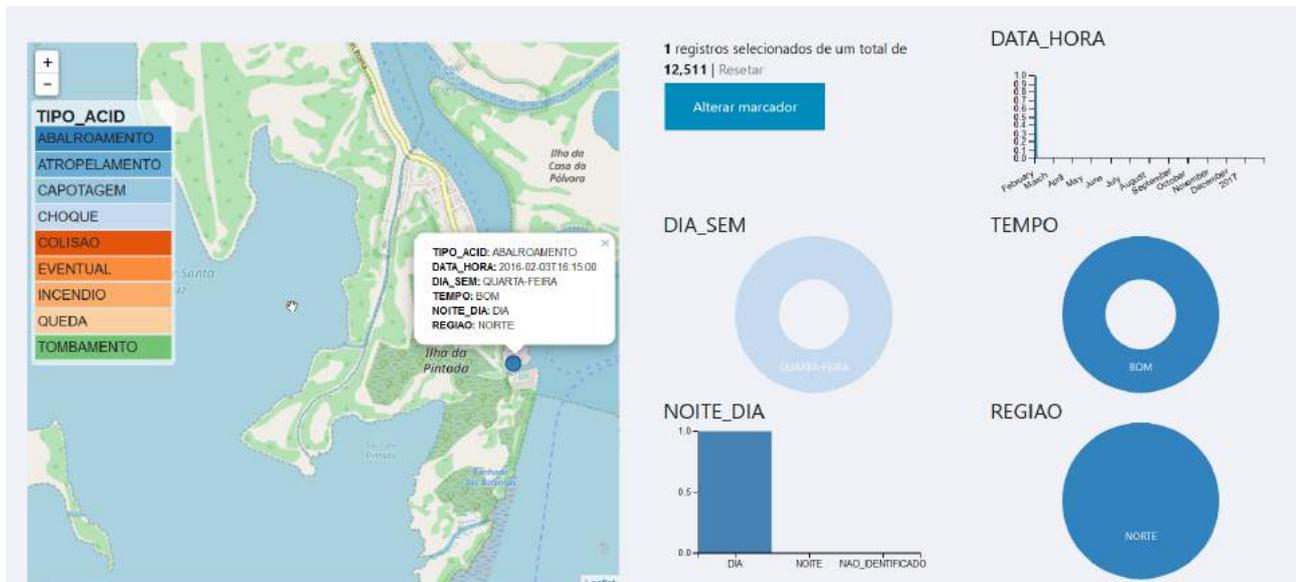


Figura 5.6 – Implementação final do modelo com filtros aplicados através da movimentação do mapa. Neste caso, os gráficos estão considerando um único registro.

A Figura 5.7 apresenta os dois tipos de gráficos que podem agrupar os marcadores no mapa, ambos exemplificados com os dados descritos na Seção 6.1.1. A primeira

opção (Figura 5.7a) agrupa os marcadores em um gráfico de *donuts*, categorizando os tipos de acidente por cores e exibindo o total de marcadores agrupados no centro do gráfico. A imagem também exemplifica o comportamento do mouse sobre os gráficos, destacando em azul a região considerada no gráfico. A segunda opção, ilustrada na Figura 5.7b, agrupa os marcadores através de um gráfico de barras, também exibindo o total de marcadores no centro do gráfico. Neste exemplo, são apresentados cinco atributos diferentes de forma simultânea, no qual cada barra representa a soma dos valores encontrados em cada marcador (carros envolvidos, vítimas fatais, mortes posteriores, feridos e motocicletas envolvidas nos acidentes). Ambas as alternativas permitem a visualização de um único atributo com dados categóricos ou múltiplos atributos com dados quantitativos. Esta restrição existe, pois valores quantitativos podem ser somados e representados de forma concorrente em um gráfico, enquanto valores textuais não podem ser associados da mesma maneira.

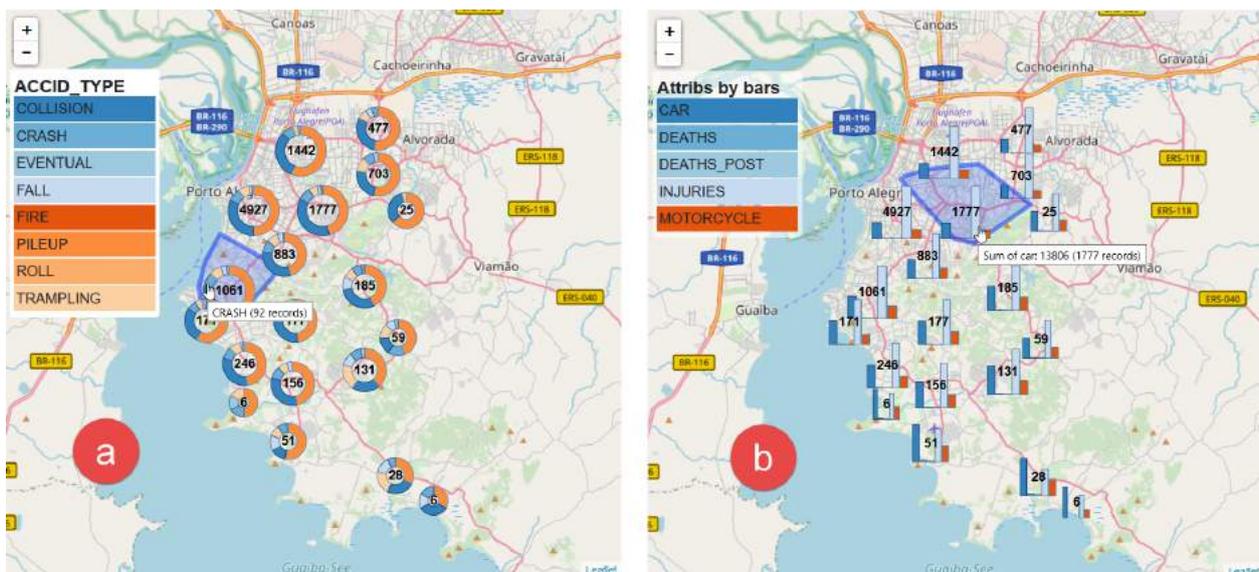


Figura 5.7 – Exemplos de como os marcadores podem ser agrupados. (a) Agrupados em gráficos de *donuts* e (b) agrupados em gráficos de barras.

5.3.1 Interface para Entrada de Dados

Considerando que as técnicas de visualização e interação podem ser aplicadas a diferentes conjuntos, conforme definido na Seção 4.1, é importante possibilitar a escolha de qual gráfico irá representar cada atributo. Para isto, foi desenvolvida a interface gráfica apresentada na Figura 5.8, que permite definir: os atributos que serão exibidos no mapa; os

atributos que irão compor os outros cinco gráficos (Figura 5.4d); e o tipo de representação visual para cada atributo (gráficos de barras, linhas, pizza ou *donut*).

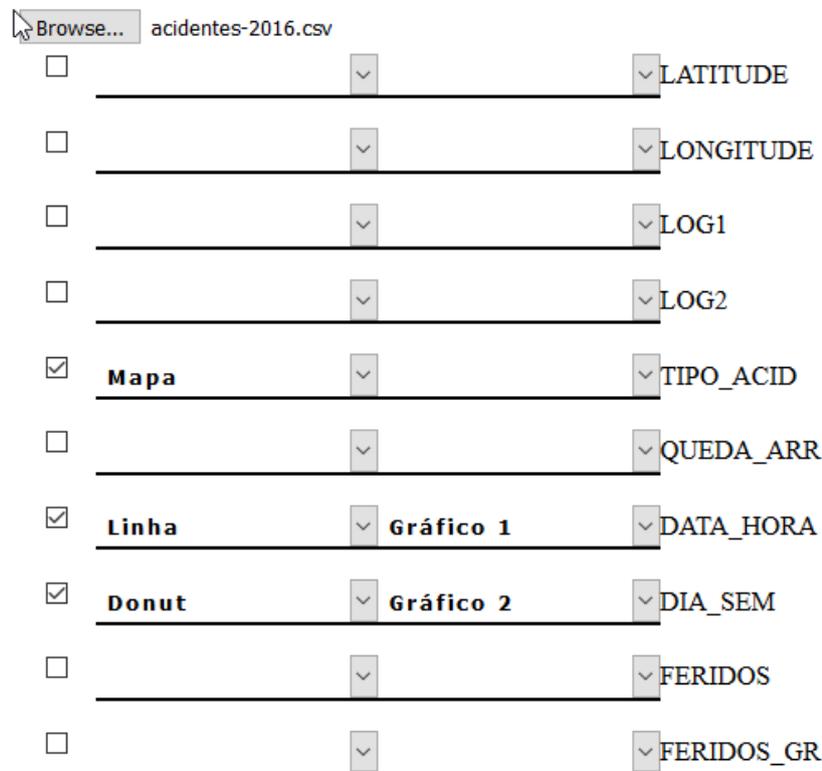


Figura 5.8 – Interface gráfica que permite a definição de como os atributos serão representados.

A interface gráfica implementada também permite especificar o conjunto de dados no formato CSV que será utilizado. Neste caso, cada linha apresentada na interface corresponde a uma coluna do conjunto de dados. Os atributos latitude e longitude são reconhecidos automaticamente como informações de localização geográfica. Na primeira coluna de opções da interface é possível definir como o atributo será representado, se é no mapa ou através de gráficos de pizza, barra, linhas ou *donut*. Na segunda coluna de opções é possível definir a localização do atributo na visualização, podendo ser no próprio mapa ou nos espaços reservados para os gráficos (neste caso, é necessário indicar o gráfico). Apenas as linhas selecionadas serão consideradas para gerar a visualização. Se mais de um atributo for definido como “Mapa”, cada atributo será representado por uma cor diferente no gráfico responsável por agrupar os marcadores, totalizando a soma dos valores encontrados em cada marcador. Através desta interface, o conjunto de dados, os atributos selecionados, a forma de representação visual e o posicionamento na visualização podem ser alterados e redefinidos a qualquer momento, de acordo com a Figura 4.1.

Uma vez definido o conjunto de dados, inserido na aplicação (Figura 5.8) e definidos os atributos que irão compor a visualização (Figura 4.1b), é utilizada a biblioteca *csv2geojson*¹² para converter o arquivo de entrada para o formato GeoJSON (*Geographic JavaScript Object Notation*), que é o formato usado na implementação do modelo. O formato GeoJSON possibilita a atribuição de propriedades geoespaciais para cada registro (coordenadas, polígonos, linhas, etc.) e, por este motivo, facilita manipulações com estes dados [2].

A quantidade de atributos ou colunas existentes nos dados de entrada não é limitada, porém, podem ser selecionados até 18 atributos para serem representados na visualização. Uma limitação presente na implementação do modelo é a inserção de conjuntos de dados com muitos registros, uma vez que isto é feito em tempo de execução através do navegador. Por este motivo, o conjunto de dados de entrada deve ser limitado em 150000 registros.

5.3.2 Tecnologias

Para possibilitar as visualizações interativas, foram adotadas diferentes tecnologias em relação à implementação inicial. A API do Google Maps foi trocada pela API do Leaflet (Figura 5.9b). Esta escolha foi motivada pela flexibilidade que o Leaflet proporciona na criação de marcadores customizados para o mapa. Para criar estes ícones como gráficos de barras e gráficos de *donuts*, a biblioteca D3.js está sendo utilizada (Figura 5.9c). Para agrupar os marcadores de acordo com sua proximidade, foi utilizada a biblioteca Marker Cluster¹³, fornecida pelo Leaflet. O Crossfilter¹⁴ está sendo utilizado para coordenar as interações entre os diferentes componentes visuais, conectando os gráficos e os movimentos do mapa ao resto da visualização. A biblioteca dc.js¹⁵ está sendo utilizada para criar os gráficos associados (Figura 5.4d). A biblioteca Turf.js¹⁶ está sendo empregada para permitir aplicar filtros nos gráficos (Figura 5.4d) a partir da movimentação do mapa (Figura 5.4b).

¹²github.com/mapbox/csv2geojson

¹³github.com/Leaflet/Leaflet.markercluster

¹⁴square.github.io/crossfilter/

¹⁵dc-js.github.io/dc.js/

¹⁶turfjs.org/

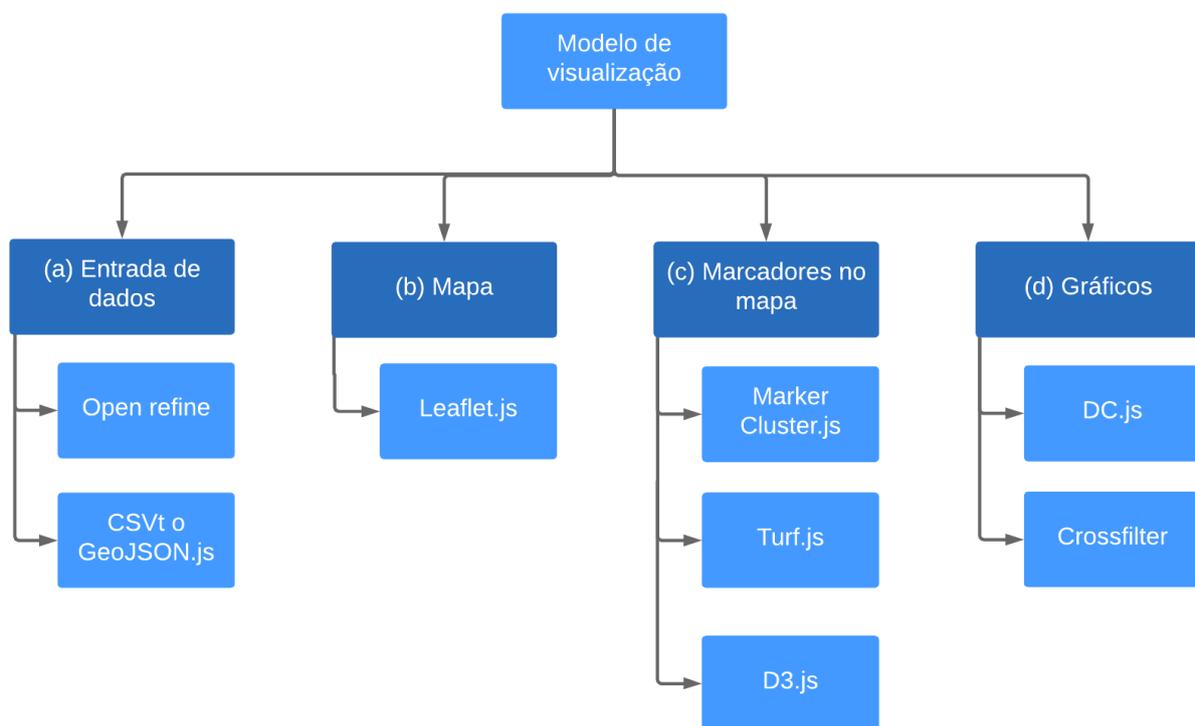


Figura 5.9 – Tecnologias utilizadas na implementação do modelo.

6. TESTES E AVALIAÇÃO

O conteúdo a seguir está organizado da seguinte maneira: as Subseções 6.1.1, 6.1.2 e 6.1.3 apresentam 3 estudos de caso utilizando o modelo; e a Seção 6.2 explica a etapa de validação final do modelo, realizada por meio de entrevistas semi-estruturadas com especialistas da área.

6.1 Estudos de caso

Com o objetivo de demonstrar que a implementação do modelo é aplicável a diferentes conjuntos de dados, as subseções a seguir mostram seu uso com três conjuntos de dados diferentes. Esses conjuntos de dados foram escolhidos por serem multivariados, isto é, por possuírem um grande número de atributos. Assim, é possível analisar as técnicas de visualização e interação propostas para o modelo, a fim de verificar o quanto auxiliam na análise do dados.

Para que os dados utilizados nos estudos de caso estivessem de acordo com o formato suportado pelo modelo (Seção 4.1), eles precisaram ser tratados. Para isto, os dados foram editados e normalizados utilizando a ferramenta *Open Refine*¹, a fim de remover os registros inválidos e padronizar os demais (Figura 5.9a).

6.1.1 Acidentes de Trânsito na Cidade de Porto Alegre

O primeiro estudo de caso utiliza os dados de acidentes de trânsito ocorridos na cidade de Porto Alegre, RS, Brasil, no ano de 2016. O DataPOA², portal de dados abertos do governo de Porto Alegre, mantém registros de acidentes de trânsito ocorridos entre os anos de 2000 e 2016. Esses registros geram grandes conjuntos de dados multivariados, com informações sobre a localização (latitude e longitude), tipo, data, hora, quantidade de feridos, quantidade de mortos, tipo de veículo e condições climáticas de cada acidente.

¹openrefine.org/

²datapoa.com.br/dataset/acidentes-de-transito

A Figura 5.4 apresenta os dados de acidentes de trânsito aplicados ao modelo. Os atributos data, dia da semana, condições climáticas, período dia e região são representados, respectivamente, nos seguintes gráficos (Figura 5.4d): um gráfico de linha, dois gráficos de *donuts*, um gráfico de barras e um gráfico de pizza. Os atributos latitude, longitude e tipo de acidente foram usados para posicionar e agrupar os marcadores no mapa.

Ao representar os dados através do modelo, é possível encontrar a relação entre os atributos, auxiliando na análise visual. A combinação dos atributos se torna clara, como os dias com mais ocorrências de acidentes, os tipos de acidentes mais comuns e as regiões da cidade mais propensas à acidentes. O cruzamento dos dados nos permite ver algumas relações, como o aumento do número de acidentes nos finais de semana e o engarrafamento sendo o tipo de acidente mais comum.

A Figura 5.7 mostra os marcadores agrupados no mapa através de diferentes gráficos, conforme descrito na Seção 5.3. Ao explorar a combinação de representações visuais do mapa e dos gráficos (Figuras 5.4 e 5.7), podemos observar que a quantidade de feridos é geralmente maior do que a quantidade de carros envolvidos nos acidentes, dando indícios que a maior parte dos acidentes não são fatais, e se concentram na zona norte da cidade sob boas condições climáticas.

6.1.2 Crimes ocorridos no Rio Grande do Sul agrupados por município

No segundo estudo de caso foram utilizados dados criminais registrados nos municípios do Rio Grande do Sul, Brasil. Esses dados, fornecidos anualmente pela Secretaria de Segurança Pública do RS (SSP-RS) e disponibilizados pela FEE (Fundação de Economia e Estatística)³, possuem diversos indicadores quantitativos associados a cada município. Na Figura 6.1, 13 atributos são apresentados de forma simultânea no mapa através de gráficos de *donuts*.

Os atributos destes dados criminais são: Delitos relacionados à Corrupção, Estelionato, Extorsão, Extorsão Mediante Sequestro, Furto, Furto de Veículos, Homicídio Doloso, Latrocínio, Posse de Entorpecentes, Tráfico de Entorpecentes, Roubo, Roubo de Veículos e Delitos Relacionados à Armas e Munições. O mapa possibilita identificar os municípios ou regiões com mais crimes e as categorias destes crimes. Através do modelo, podemos

³visualiza.fee.tche.br/crime/

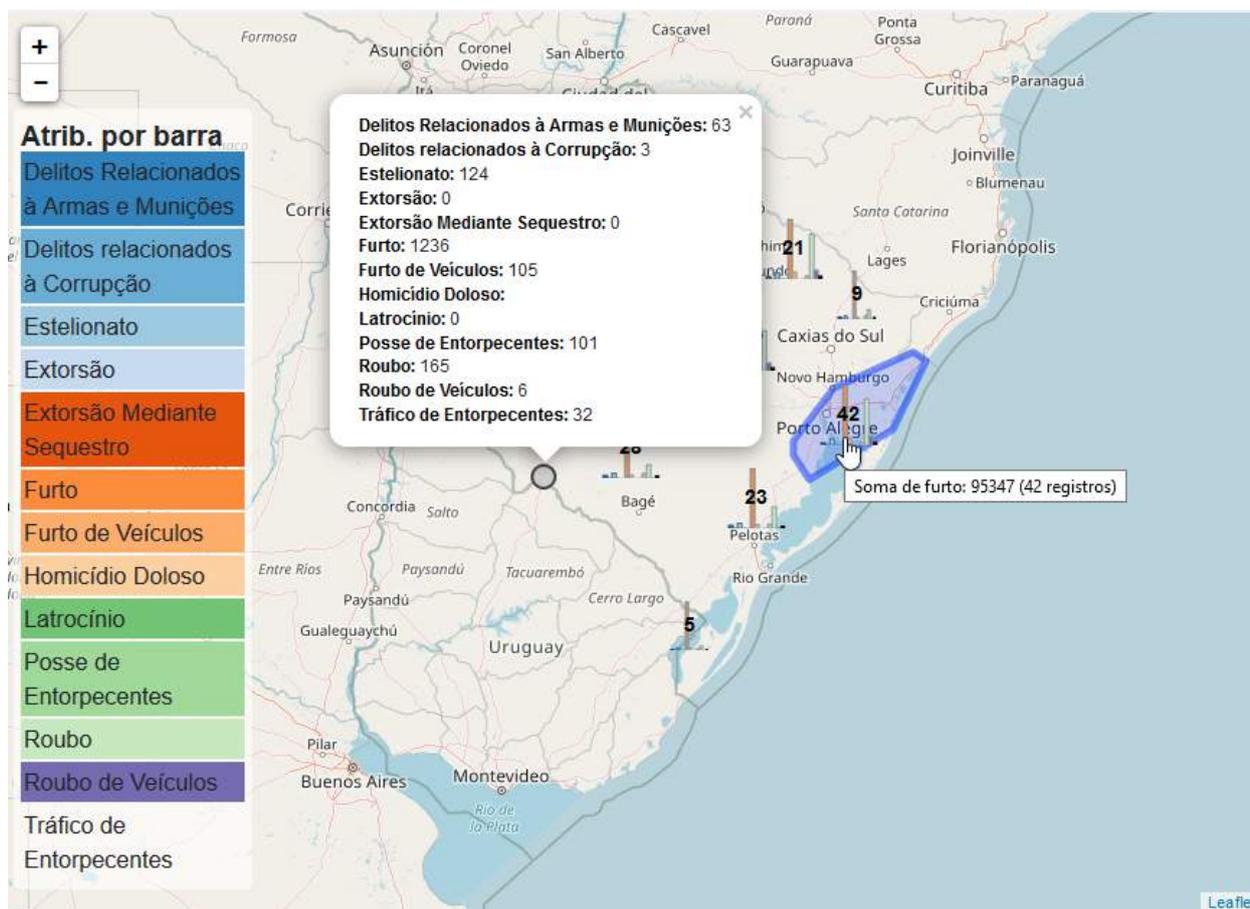


Figura 6.1 – Estudos de caso com dados de criminalidade dos municípios do Rio Grande do Sul.

ver que a categoria de crime com mais registros no Rio Grande do Sul é furto. Ao clicar no marcador final, um *pop-up*, contendo as informações utilizadas para compor o gráfico, é exibido (Figura 6.1).

6.1.3 Crimes ocorridos em Bethesda, Maryland

Um conjunto de dados criminais, semelhante ao utilizado na Seção 6.1.2, foi usado no terceiro estudo de caso. Os registros são da cidade de Bethesda, MD, EUA, e são disponibilizados pelo governo dos Estados Unidos⁴. A principal diferença em relação ao estudo de caso apresentado na Seção 6.1.2, é que os dados estão associados a localizações geográficas dentro da área de uma cidade, sendo, assim, agrupados por regiões da cidade (Figura 6.2).

⁴catalog.data.gov/dataset/crime

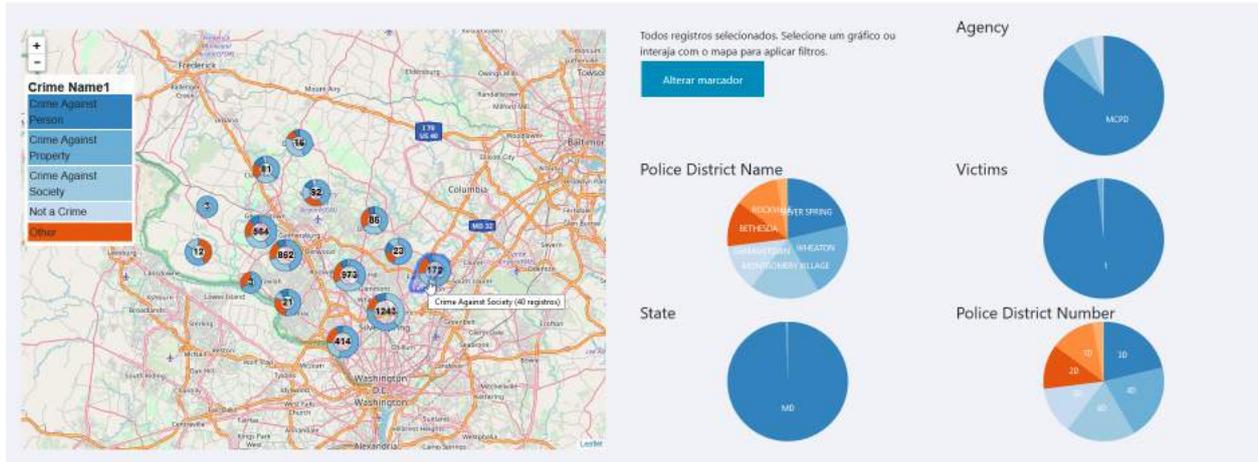


Figura 6.2 – Estudos de caso com dados de criminalidade de Bethesda, Estados Unidos.

6.2 Avaliação com usuários

De acordo com Boyce e Neale [10], entrevista em profundidade é uma técnica de pesquisa qualitativa que envolve a realização de entrevistas individuais com uma quantidade pequena de pessoas. O objetivo é explorar as suas perspectivas sobre uma determinada ideia ou situação. Para avaliar a eficácia do modelo, foram conduzidas entrevistas em profundidade com 5 especialistas no domínio, com idades entre 25 e 45 anos. Os entrevistados foram selecionados de acordo com a sua experiência nas áreas de visualização de dados e dados abertos do governo, como mostra a Tabela 6.1. Foram conduzidas entrevistas semi-estruturadas [1] com questões norteadoras, permitindo que depoimentos detalhados possam ser coletados.

#	Idade	Gênero	Área de Formação	Foco de Pesquisa
E1	37	Masculino	Computação	Visualização de dados
E2	40	Masculino	Administração	Dados abertos do governo
E3	25	Masculino	Computação	Visualização de dados geoespaciais
E4	43	Feminino	Administração	Dados abertos do governo
E5	45	Feminino	Computação	Dados abertos do governo

Tabela 6.1 – Perfil dos participantes da avaliação final

O primeiro entrevistado (E1) possui formação em Ciência da Computação e atua profissionalmente como gerente de arquitetura de TI, onde utiliza softwares analíticos, como o Tableau e o Cognos Analytics; o segundo entrevistado (E2) é professor universitário no curso de Ciência da Computação e realiza pesquisas sobre dados abertos do governo; o

terceiro entrevistado (E3) está cursando mestrado em Ciência da Computação, com pesquisas na área de visualização de dados geoespaciais, familiarizado com as tecnologias Tableau e Highcharts⁵; a quarta entrevistada (E4) é professora universitária no curso de Administração, no qual leciona e realiza pesquisas sobre dados abertos do governo; e a quinta entrevistada (E5) é professora universitária no curso de Administração, no qual leciona e realiza pesquisas sobre dados abertos do governo, utilizando ferramentas como Knowledge Foundation⁶, Open Data Barometer⁷ e Doing Business⁸.

As entrevistas duraram cerca de 30-40 minutos e tiveram seus áudios gravados com o consentimento do participante. Primeiramente, o propósito da entrevista era explicado aos entrevistados e eles assinavam o termo de consentimento livre e esclarecido (Apêndice ??). Na sequência, o objetivo e o funcionamento do modelo eram explicados, bem como a natureza dos conjuntos de dados utilizados no estudos de caso. Os entrevistados tiveram 1-2 minutos para manipular o modelo e conhecer suas funcionalidades. Em seguida, eles foram solicitados a executar as seguintes tarefas:

- Tarefa 1: utilizar o conjunto de dados de acidentes de trânsito na cidade de Porto Alegre (estudo de caso 1, apresentado na Seção 6.1.1) para encontrar qual região apresenta mais acidentes, aos sábados, em dias chuvosos;
- Tarefa 2: encontrar qual região apresenta mais acidentes no período da noite;
- Tarefa 3: alterar a forma como os marcadores são representados no mapa, alternando entre gráficos de barras e gráficos de *donuts*;
- Tarefa 4: gerar uma nova visualização utilizando os dados de criminalidade no RS (estudo de caso 2, apresentado na Seção 6.1.2);
- Tarefa 5: encontrar qual o tipo de crime mais recorrente no Rio Grande do Sul;
- Tarefa 6: encontrar qual tipo de crime apresenta mais registros na região metropolitana de Porto Alegre;

Os 5 entrevistados conseguiram realizar as tarefas e encontrar as respostas sem dificuldades. As perguntas feitas aos participantes com o objetivo de coletar suas opiniões

⁵www.highcharts.com/

⁶www.knowledgefoundation.com/

⁷opendatabarometer.org/

⁸www.doingbusiness.org/

e sugestões a respeito do modelo são apresentadas a seguir, junto com uma síntese de suas respostas.

- Pergunta 1: “De modo geral, as visualizações e interações fornecidas auxiliam na análise e interpretação dos dados?”

Todos os participantes responderam que acreditam que a forma como as informações são representadas e a interação existente entre os gráficos auxiliam na análise dos dados. O E2 mencionou que “a visualização e as interações ajudam na análise dos dados” e o E3 observou que “a visualização de vários atributos ao mesmo tempo é uma das forças do modelo” e que “os filtros aplicados através da interação ajudam a entender o relacionamento dos atributos”. O E5 acredita que a combinação de filtros entre os gráficos possibilita uma compreensão do relacionamento entre os atributos. O E4 comentou que o uso de uma interface semelhante a outras aplicações utilizadas no cotidiano, como Uber⁹, Google Maps e gráficos simples (Excel), diminui a curva de aprendizagem.

- Pergunta 2: “Qual a sua opinião sobre cada uma das funcionalidades do modelo (*zoom*; movimentação do mapa; interação dos gráficos em si e com o mapa; e alteração do tipo de gráfico)?”

Na resposta para esta pergunta, o E1 mencionou que “o uso do mapa como filtro é um diferencial do modelo, e a maneira como o *zoom* funciona, adaptando o tamanho dos *clusters* e a maneira como os *clusters* são redimensionados de acordo com a interação do *zoom*, é boa”. E3 também mencionou a aplicação de filtros a partir da movimentação do mapa como um diferencial do modelo. O E4 destacou os gráficos agrupando marcadores como um ponto positivo, já que evita a sobreposição dos marcadores. E5 observou que “o *zoom* poderia funcionar de forma mais suave” e disse preferir os gráficos de *donuts* para agrupar os marcadores no mapa.

- Pergunta 3: “Na sua opinião, existe a necessidade de incluir outros gráficos ou remover algum(ns) dos gráficos existentes?”

Ao responder esta pergunta, o participante E1 sugeriu: usar apenas o gráfico de *donut*, eliminando a opção do gráfico de pizza para os gráficos associados ao mapa,

⁹www.uber.com

por considerar as duas opções próximas demais; incluir gráficos de coordenadas paralelas e de radar no espaço reservado para os gráficos 1 a 5; e mover o botão que altera os marcadores entre gráfico de *donut* e gráfico de barras para dentro do mapa, próximo a legenda - desta forma ficaria mais claro para os usuários que ele está relacionado ao mapa. O E2 também sugeriu que o gráfico de radar e de dispersão sejam adicionados como alternativa (Figura 5.4d), possibilitando a visualização de mais de um atributo de forma simultânea.

- Pergunta 4: “É possível identificar diferentes contextos em que o modelo pode ser usado? Quais?”

A aplicabilidade do modelo a diferentes contextos foi destacada por todos entrevistados. E2 sugeriu a utilização do modelo para associar informações sócio-econômicas com a ocorrência de doenças como malária e tuberculose, pois “isto demonstraria se populações mais pobres estão mais suscetíveis a determinadas doenças”. O E3 mencionou a flexibilidade do modelo como um fator positivo, pois qualquer conjunto de dados com localização geoespacial pode ser visualizado por meio de sua interface. O E4 sugeriu a utilização do modelo com dados de serviços públicos, podendo apresentar horários de funcionamento de postos de saúde, vagas em escolas e obras de infra-estrutura.

- Pergunta 5: “Indique os pontos positivos do modelo de visualização e sugestões de melhoria”

Os entrevistados E1, E3, E4 e E5 sugeriram que o botão para alterar os marcadores entre gráficos de barras/*donuts* poderia ser colocado dentro do mapa, próximo à legenda. De acordo com E1, o posicionamento deste botão na visualização não proporciona uma sensação de *affordance* [20]: “não dá a ideia que acontecerá alguma coisa no mapa”. O E1 também sugeriu que a opção “*resetar*” seja alterada por um ícone. Os participantes E1, E2, E3 e E5 sugeriram a utilização de uma paleta de cores neutra, visando evitar a sensação de relacionamento entre as cores. O E2 observou que, quando muitos valores são exibidos nos gráficos de barras, as barras ficam muito finas, o que dificulta a identificação de qual atributo elas estão representando. Já o E3 sugeriu que as barras sejam ordenadas de acordo com seus valores nos gráficos de barras.

As tarefas realizadas, as perguntas respondidas e os comentários dos participantes durante a realização das entrevistas demonstram que as técnicas de visualização e interação presentes no modelo auxiliam na análise e exploração dos dados. Conforme mencionado pelos entrevistados, a curva de aprendizagem para o modelo é menor pela familiaridade por grande parte da população com o uso de mapas e gráficos.

7. CONCLUSÕES

A crescente disponibilidade de dados multivariados geoespaciais e as limitações presentes na análise destes dados trazem a necessidade de alternativas que auxiliem nesta análise [13]. A visualização destes tipos de dados, os quais representam o relacionamento de diversos atributos junto com sua localização espacial, não é uma tarefa trivial.

Por meio do estudo apresentado no Capítulo 3, foi possível identificar quais técnicas de visualização e interação tem sido utilizadas para visualização e exploração de dados multivariados no contexto geoespacial. Além disso, também foi possível verificar a existência de limitações para lidar com alguns cenários, como a representação de múltiplos marcadores e ainda prover informações detalhadas à respeito de cada item, conforme apresentado por Meier e Heidmann [47].

Assim, visando melhorar a forma como estas informações são apresentadas, foi proposto e desenvolvido um modelo para visualização e exploração de dados multivariados no contexto geoespacial. Neste modelo, diversas técnicas de visualização de dados e de interação são integradas com o objetivo de possibilitar análises visuais e auxiliar na tomada de decisão a partir de diferentes conjuntos de dados. O desafio, neste caso, era apresentar a maior quantidade possível de informações de forma simultânea e ainda assim se manter compreensível. O modelo foi implementado utilizando diferentes tecnologias, com o propósito de confirmar que as técnicas nele presentes podem ser aplicadas a diferentes conjuntos de dados. Para isto, são apresentados três estudos de caso com dados abertos do governo.

Para validar o modelo proposto foram utilizadas duas abordagens: inicialmente, foi realizado um grupo focal, no qual uma versão preliminar do modelo foi apresentada, e as sugestões e críticas foram coletadas para aperfeiçoamento do mesmo; depois, foram realizadas entrevistas individuais com especialistas na área de visualização de dados e análise de dados abertos do governo, nas quais a implementação final do modelo foi apresentada. Estas abordagens revelaram que as técnicas presentes no modelo ajudaram os participantes na análise de dados e na descoberta de conhecimento por meio das visualizações e interações fornecidas pelo modelo. Portanto, as contribuições deste trabalho foram: (I) a possibilidade de representar diversos atributos no mapa e associar este mapa com outros gráficos; (II) uma abordagem interativa, baseada na integração das técnicas *brushing and*

linking com *coordinated multiple views* (CMV), auxiliando na análise visual; e (III) a utilização deste modelo para analisar dados abertos do governo. Portanto, sua contribuição está diretamente relacionada com a descoberta de conhecimento, considerando que estas técnicas são utilizadas para estimular uma interação ativa dos usuários.

Como trabalhos futuros, as sugestões mencionadas nas entrevistas com especialistas podem ser aplicadas ao modelo. Entre estas sugestões estão: a possibilidade de inclusão de outros tipos de gráfico, como radar e coordenadas paralelas; melhorias da usabilidade da interface, como a localização do botão que altera a forma como os marcadores são agrupados e as cores utilizadas na legenda. Como o foco deste trabalho era a forma como os dados são representados e não o seu desempenho, também pretende-se reformular a maneira como os conjuntos de dados são inseridos e as operações realizadas com estes dados, visando deixar seu funcionamento mais rápido.

Além disso, pretende-se aplicar o modelo em novos estudos de caso, com o objetivo de aprofundar a investigação sobre sua utilidade, tentando resolver problemas enfrentados por analistas de dados e pesquisadores que devem lidar com uma grande quantidade de dados multivariados. O próximo caso a ser investigado é sobre a transparência e a percepção da corrupção no mundo.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Aghamanoukjan, A.; Buber, R.; Meyer, M. “Qualitative interviews”. In: *Qualitative Marktforschung*, Springer, 2009, cap. 26, pp. 415–436.
- [2] Andrews, C. “Emerging technology: Ajax and geojson”. Capturado em: <https://www.directionsmag.com/article/2639>, Setembro 2007.
- [3] Andrienko, G.; Andrienko, N. “Exploring spatial data with dominant attribute map and parallel coordinates”, *Computers, Environment and Urban Systems*, vol. 25–1, Janeiro 2001, pp. 5–15.
- [4] Atallah, A.; Castro, A. “Revisão sistemática da literatura e metanálise.” São Paulo, Brasil: Lemos-Editorial, 1998, pp. 42–48.
- [5] Baeza-Yates, R.; Ribeiro-Neto, B. “Modern information retrieval”. New York, United States: ACM press, 1999, 944p.
- [6] Baldonado, M.; Woodruff, A.; Kuchinsky, A. “Guidelines for using multiple views in information visualization”. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*, 2000, pp. 110–119.
- [7] Bederson, B. B. “Photomesa: a zoomable image browser using quantum treemaps and bubblemaps”. In: *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, 2001, pp. 71–80.
- [8] Booth, A.; Sutton, A.; Papaioannou, D. “Systematic approaches to a successful literature review”. London, England: Sage, 2016, 325p.
- [9] Bostock, M.; Ogievetsky, V.; Heer, J. “D³ data-driven documents”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 17–12, Setembro 2011, pp. 2301–2309.
- [10] Boyce, C.; Neale, P. “Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input”. Watertown, United States: Pathfinder International, 2006, 2 ed., 16p.
- [11] Card, M. “Readings in information visualization: using vision to think”. San Diego, United States: Morgan Kaufmann, 1999, 688p.

- [12] Cleveland, W.; McGill, R. “Graphical perception: Theory, experimentation, and application to the development of graphical methods”, *Journal of the American Statistical Association*, vol. 79–387, Setembro 1984, pp. 531–554.
- [13] Cook, K.; Thomas, J. “Illuminating the path: The research and development agenda for visual analytics”, Relatório Técnico, Pacific Northwest National Lab., 2005, 190p.
- [14] De Oliveira, M.; Levkowitz, H. “From visual data exploration to visual data mining: a survey”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 9–3, Julho 2003, pp. 378–394.
- [15] Delort, J.-Y. “Hierarchical cluster visualization in web mapping systems”. In: Proceedings of the 19th International Conference on World Wide Web, 2010, pp. 1241–1244.
- [16] Ferreira, A. “Novo Aurélio Século XXI: o dicionário da língua portuguesa”. Rio de Janeiro, Brasil: Nova Fronteira, 1999, 2128p.
- [17] Few, S. “Now you see it: simple visualization techniques for quantitative analysis”. Washington, United States: Analytics Press, 2009, 329p.
- [18] Fisher, E.; Bynum, J.; Skinner, J. “Slowing the growth of health care costs—lessons from regional variation”, *New England Journal of Medicine*, vol. 360–9, Fevereiro 2009, pp. 849–852.
- [19] Gelman, A.; Unwin, A. “Infovis and statistical graphics: different goals, different looks”, *Journal of Computational and Graphical Statistics*, vol. 22–1, Janeiro 2013, pp. 2–28.
- [20] Gibson, J. “The senses considered as perceptual systems.” Boston, United States: Houghton Mifflin Company, 1966, 335p.
- [21] Guiard, Y.; Bourgeois, F.; Mottet, D.; Beaudouin-Lafon, M. “Beyond the 10-bit barrier: Fitts’ law in multi-scale electronic worlds”. In: *People and Computers XV—Interaction without Frontiers*, Springer, 2001, cap. 12, pp. 573–587.
- [22] Ho, Q.; Jern, M. “Exploratory 3d geovisual analytics”. In: Proceedings of the IEEE International Conference on Research, Innovation and Vision for the Future in Computing and Communication Technologies, 2008, pp. 276–283.

- [23] Inselberg, A. “The plane with parallel coordinates”, *The Visual Computer*, vol. 1–2, Agosto 1985, pp. 69–91.
- [24] Jenks, G. “Contemporary statistical maps—evidence of spatial and graphic ignorance”, *The American Cartographer*, vol. 3–1, Janeiro 1976, pp. 11–19.
- [25] Jern, M. “Collaborative web-enabled geanalytics applied to oecd regional data”. In: *Proceedings of the Cooperative Design, Visualization, and Engineering*, 2009, pp. 32–43.
- [26] Jern, M.; Johansson, S.; Johansson, J.; Franzen, J. “The gav toolkit for multiple linked views”. In: *Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization*, 2007, pp. 85–97.
- [27] Justi, R. “La enseñanza de ciencias basada en la elaboración de modelos”, *Enseñanza de las Ciencias: Revista de Investigación y Experiencias Didácticas*, vol. 24–2, Novembro 2006, pp. 173–184.
- [28] Keim, D. “Information visualization and visual data mining”, *IEEE Transactions on Visualization and Computer Graphics*, Janeiro 2002, pp. 1–8.
- [29] Keim, D.; Mansmann, F.; Schneidewind, J.; Thomas, J.; Ziegler, H. “Visual analytics: Scope and challenges”. In: *Visual Data Mining*, Springer, 2008, cap. 2, pp. 76–90.
- [30] Keim, D.; Panse, C.; North, S. “Medial-axis-based cartograms”, *IEEE Computer Graphics and Applications*, vol. 25–3, Maio 2005, pp. 60–68.
- [31] Keim, D.; Panse, C.; Sips, M.; North, S. C. “Visual data mining in large geospatial point sets”, *IEEE Computer Graphics and Applications*, vol. 24–5, Setembro 2004, pp. 36–44.
- [32] Keim, E. D.; Kohlhammer, J.; Ellis, G. “Mastering the information age solving problems with visual analytics”. In: *Proceedings of the Eurographics*, 2010, pp. 5.
- [33] Kitchenham, B. “Procedures for performing systematic reviews”, Relatório Técnico, Keele University, 2004, 33p.
- [34] Kitchenham, B.; Budgen, D.; Brereton, P. “Evidence-Based Software engineering and systematic reviews”. Boca Raton, Flórida: CRC Press, 2015, 390p.

- [35] Korpela, K. "The art of data visualization", *Information Systems Audit and Control Association*, vol. 1, Fevereiro 2016, pp. 1–5.
- [36] Kovalerchuk, B.; Schwing, J. "Visual and spatial analysis: advances in data mining, reasoning, and problem solving". Washington, United States: Springer Science & Business Media, 2005, 582p.
- [37] Kruskal, J. "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis", *Psychometrika*, vol. 29–1, Março 1964, pp. 1–27.
- [38] Leonowicz, A. "Two-variable choropleth maps as a useful tool for visualization of geographical relationship", *Geografija*, vol. 42, Junho 2006, pp. 33–37.
- [39] Li, J.; Chen, S.; Zhang, K.; Andrienko, G.; Andrienko, N. "Cope: Interactive exploration of co-occurrence patterns in spatial time series", *IEEE Transactions on Visualization and Computer Graphics*, 2018, pp. 1–8.
- [40] Li, J.; Meng, Z.-P.; Huang, M.-L.; Zhang, K. "An interactive visualization approach to the overview of geoscience data", *Journal of Visualization*, vol. 20–3, Agosto 2017, pp. 433–451.
- [41] Liao, Z.; Peng, Y.; Li, Y.; Liang, X.; Zhao, Y. "A web-based visual analytics system for air quality monitoring data". In: Proceedings of the 22nd International Conference on Geoinformatics, 2014, pp. 1–6.
- [42] Liono, J.; Salim, F. D.; Subastian, I. F. "Visualization oriented spatiotemporal urban data management and retrieval". In: Proceedings of the ACM First International Workshop on Understanding the City with Urban Informatics, 2015, pp. 21–26.
- [43] Liu, F.; Tian, K.; Sun, Y.; Li, B. Y.; Lin, H. "Analysis of statistics data based on mixed visualization techniques". In: Proceedings of the Applied Mechanics and Materials, 2012, pp. 2479–2484.
- [44] Liu, S.; Maljovec, D.; Wang, B.; Bremer, P.-T.; Pascucci, V. "Visualizing high-dimensional data: Advances in the past decade". In: Proceedings of the Eurographics Conference on Visualization, 2015, pp. 115–127.

- [45] Lundblad, P.; Jern, M. "Geovisual analytics and storytelling using html5". In: Proceedings of the 17th International Conference Information Visualisation, 2013, pp. 263–271.
- [46] MacEachren, A. M.; Wachowicz, M.; Edsall, R.; Haug, D.; Masters, R. "Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with knowledge discovery in database methods", *International Journal of Geographical Information Science*, vol. 13–4, Junho 1999, pp. 311–334.
- [47] Meier, S.; Heidmann, F. "Too many markers, revisited: An empirical analysis of web-based methods for overcoming the problem of too many markers in zoomable mapping applications". In: Proceedings of the 14th International Conference on Computational Science and Its Applications, 2014, pp. 121–125.
- [48] Morgan, D. "The focus group guidebook". London, England: Sage publications, 1997, 108p.
- [49] Opach, T.; Rød, J. "Cartographic visualization of vulnerability to natural hazards", *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 48–2, Junho 2013, pp. 113–125.
- [50] Peterson, G. "GIS cartography: a guide to effective map design". Florida, United States: CRC Press, 2009, 246p.
- [51] Phan, D.; Xiao, L.; Yeh, R.; Hanrahan, P. "Flow map layout". In: Proceedings of the IEEE Symposium on Information Visualization, 2005, pp. 219–224.
- [52] Rabiee, F. "Focus-group interview and data analysis". In: Proceedings of the Nutrition Society, 2004, pp. 655–660.
- [53] Rao, R.; Card, S. K. "The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information". In: Proceedings of the Conference on Human Factors in Computing Systems, 1994, pp. 318–322.
- [54] Rayson, J. "Aggregate towers: Scale sensitive visualization and decluttering of geospatial data". In: Proceedings of the IEEE Symposium on the Information Visualization, 1999, pp. 92–99.

- [55] Roberts, J. “On encouraging multiple views for visualization”. In: Proceedings of the IEEE Conference on Information Visualization, 1998, pp. 8–14.
- [56] Roberts, J.; Wright, M. “Towards ubiquitous brushing for information visualization”. In: Proceedings of the Tenth International Conference on Information Visualization, 2006, pp. 151–156.
- [57] Sandvik, B. “Using kml for thematic mapping”, Dissertação de Mestrado, School of GeoSciences, University of Edinburgh, Edimburgo, Escócia, 2008, 20p.
- [58] Scrimshaw, S.; Hurtado, E. “Anthropological Approaches to Improving Programme Effectiveness”. Los Angeles, United States: UCLA Latin American Center Publications, 1987, pp. 944–944.
- [59] Shafranovich, Y. “Common format and mime type for comma-separated values files”. Capturado em: <https://tools.ietf.org/html/rfc4180>, Janeiro 2018.
- [60] Shneiderman, B. “Dynamic queries for visual information seeking”, *IEEE Software*, vol. 11–6, Novembro 1994, pp. 70–77.
- [61] Shneiderman, B. “The eyes have it: A task by data type taxonomy for information visualizations”. In: Proceedings of the IEEE Symposium on Visual Languages, 1996, pp. 336–343.
- [62] Sneath, P. “The application of computers to taxonomy”, *Journal of General Microbiology*, Agosto 1957, pp. 201–226.
- [63] Straus, S.; Richardson, W.; Glasziou, P.; Haynes, R. “Evidence-based medicine: how to practice and teach EBM”. Toronto, Canadá: Elsevier Health Sciences, 2005, 324p.
- [64] Tobler, W. “Cartograms and cartosplines”. In: Proceedings of the Workshop on Automated Cartography and Epidemiology, 1976, pp. 53–58.
- [65] Turkay, C.; Slingsby, A.; Hauser, H.; Wood, J.; Dykes, J. “Attribute signatures: Dynamic visual summaries for analyzing multivariate geographical data”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 20–12, Dezembro 2014, pp. 2033–2042.
- [66] Van Ho, Q.; Lundblad, P.; Åström, T.; Jern, M. “A web-enabled visualization toolkit for geovisual analytics”, *Information Visualization*, vol. 11–1, Janeiro 2012, pp. 22–42.

- [67] Vartanian, O.; Navarrete, G.; Chatterjee, A.; Fich, L. B.; Leder, H.; Modroño, C.; Nadal, M.; Rostrup, N.; Skov, M. “Impact of contour on aesthetic judgments and approach-avoidance decisions in architecture”. In: Proceedings of the National Academy of Sciences, 2013, pp. 10446–10453.
- [68] Voigt, R. “An extended scatterplot matrix and case studies in information visualization”, Dissertação de Mestrado, Magdeburg University of Applied Sciences, Germany, 2002, 113p.
- [69] Ward, M.; Grinstein, G.; Keim, D. “Interactive Data Visualization: foundations, techniques, and applications”. Florida, United States: CRC Press, 2015, 578p.
- [70] Ward, M.; Yang, J. “Interaction spaces in data and information visualization”. In: Proceedings of the Eurographics Conference on Visualization, 2004, pp. 137–145.
- [71] Wills, G. “Selection: 524,288 ways to say “this is interesting””. In: Proceedings of the IEEE Symposium on Information Visualization, 1996, pp. 54–60.
- [72] Wohlin, C. “Guidelines for snowballing in systematic literature studies and a replication in software engineering”. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, 2014, pp. 38.
- [73] Wohlin, C. “Second-generation systematic literature studies using snowballing”. In: Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering, 2016, pp. 15.
- [74] Wongsuphasawat, K.; Pack, M.; Filippova, D.; VanDaniker, M.; Olea, A. “Visual analytics for transportation incident data sets”, *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1–2138, Dezembro 2009, pp. 135–145.
- [75] Yang, J.; Ward, M. O.; Rundensteiner, E. A. “Interring: An interactive tool for visually navigating and manipulating hierarchical structures”. In: Proceedings of the IEEE Symposium on Information Visualization, 2002, pp. 77–84.
- [76] Zhang, J.; Ahlbrand, B.; Malik, A.; Chae, J.; Min, Z.; Ko, S.; Ebert, D. “A visual analytics framework for microblog data analysis at multiple scales of aggregation”. In: Proceedings of the Computer Graphics Forum, 2016, pp. 441–450.

- [77] Zhang, Y.; Luo, W.; Mack, E. A.; Maciejewski, R. “Visualizing the impact of geographical variations on multivariate clustering”. In: *Proceeding of the Computer Graphics Forum*, 2016, pp. 101–110.
- [78] Zhang, Z.; Tong, X.; McDonnell, K. T.; Zelenyuk, A.; Imre, D.; Mueller, K. “An interactive visual analytics framework for multi-field data in a geo-spatial context”, *Tsinghua Science and Technology*, vol. 18–2, Abril 2013, pp. 111–124.

APÊNDICE A – STRINGS DE BUSCA

A.1 IEEEExplore digital library

((((((geovisualiz OR geospatial OR geographic* OR geovis OR “spatial domains”)
AND (visualiz*) AND (analy*) AND (interacti*) AND (multivariate OR multi-variate OR multi-
parameter OR “different properties”))))))*

A.2 Scopus

TITLE-ABS-KEY ((((((geovisualiz OR geospatial OR geographic* OR geovis OR
“spatial domains”) AND (visualiz*) AND (analy*) AND (interacti*) AND (multivari-
ate OR multi-variate OR multi-parameter OR “different properties”)))))) AND (LIMIT-TO (DOCTYPE, “ar”) OR LIMIT-TO (DOCTYPE, “cp”)) AND (LIMIT-TO (PUBYEAR, 2018) OR
LIMIT-TO (PUBYEAR, 2017) OR LIMIT-TO (PUBYEAR, 2016) OR LIMIT-TO (PUBYEAR,
2015) OR LIMIT-TO (PUBYEAR, 2014) OR LIMIT-TO (PUBYEAR, 2013) OR LIMIT-TO (PUBYEAR, 2012) OR LIMIT-TO (PUBYEAR, 2011) OR LIMIT-TO (PUBYEAR, 2010) OR
LIMIT-TO (PUBYEAR, 2009) OR LIMIT-TO (PUBYEAR, 2008))*

A.3 Eslevier ScienceDirect

((((((Title-Abstr-Key (“geovisualiz”) OR Title-Abstr-Key (“geospatial”) OR Title-Abstr-
Key (“geographic”) OR Title-Abstr-Key (“geovis*”) OR Title-Abstr-Key (“spatial domains”)
AND Title-Abstr-Key (visualiz*) AND Title-Abstr-Key (“analy*”) AND Title-Abstr-Key (“inte-
racti”) AND (Title-Abstr-Key (“multivariate”) AND Title-Abstr-Key (“multi-variate”) AND Title-
Abstr-Key (“multi-parameter”) OR Title-Abstr-Key (“different properties”))))))*

A.4 ACM digital library

(acmdlTitle: (+geovisualiz) OR acmdlTitle: (+geospatial*) OR acmdlTitle: (+geographic*) OR acmdlTitle: (+geovis) OR acmdlTitle: (+spatial domains)) AND acmdlTitle: (+visualiz*) AND acmdlTitle: (+analy*) AND acmdlTitle: (+interacti*) AND (acmdlTitle: (+multivariate) OR acmdlTitle: (+multi-variate) OR acmdlTitle: (+multi-parameter) OR acmdlTitle: (+different properties)) OR ((recordAbstract: (+geovisualiz*) OR recordAbstract: (+geospatial*) OR recordAbstract:(+geographic*) OR recordAbstract: (+geovis) OR recordAbstract: (+spatial domains)) AND recordAbstract: (+visualiz*) AND recordAbstract: (+analy*) AND recordAbstract: (+interacti*) AND (recordAbstract: (+multivariate) OR recordAbstract: (+multi-variate) OR recordAbstract: (+multi-parameter) OR recordAbstract: (+different properties)))*

APÊNDICE B – TERMOS DE CONSENTIMENTO

B.1 Grupo focal

Termo de Consentimento Livre e Esclarecido

Você está sendo convidado a participar de um estudo sobre visualização de dados geoespaciais. Este trabalho faz parte da pesquisa de Mestrado do aluno Wagner Signoretti, sob orientação da professora Isabel Manssour, no Programa de Pós-Graduação em Ciência da Computação da PUCRS. Agradecemos a todos os participantes deste estudo realizado sob nossa responsabilidade pela inestimável contribuição que prestam para o avanço das pesquisas sobre Visualização de Dados.

Lembramos que o objetivo do estudo ora em realização não é avaliar o participante, mas, sim, avaliar os aspectos de visualização e interação humano-computador relacionados ao dispositivo computacional que está sendo analisado. Este estudo será registrado em papel e, também, por meio de gravações de áudio e vídeo de toda ou parte da entrevista. O uso que se faz dos registros efetuados durante o estudo é estritamente limitado a atividades de acadêmicas, garantindo-se para tanto que:

1. O anonimato dos participantes será preservado em todo e qualquer documento divulgado em foros científicos (tais como conferências, periódicos, livros e assemelhados) ou pedagógicos (tais como apostilas de cursos, *slides* de apresentações, e assemelhados).
2. Os áudios e vídeos gravados serão utilizados apenas para apoio à análise e não serão divulgados externamente.
3. Todo participante que se sentir constrangido ou incomodado durante uma situação pode interromper a entrevista e estará fazendo um favor à equipe se registrar por escrito as razões ou sensações que o levaram a esta atitude. A equipe fica obrigada a descartar a entrevista para fins da avaliação a que se destinaria.
4. Todo participante tem direito de expressar por escrito, na data do estudo, qualquer restrição ou condição adicional que lhe pareça aplicar-se aos itens acima enumerados (1 e 2). A equipe se compromete a observá-las com rigor e entende que, na ausência de tal manifestação, o participante concorda que rejam o comportamento ético da equipe somente as condições impressas no presente documento.
5. A equipe tem direito de utilizar os dados do estudo, mantidas as condições acima mencionadas, para quaisquer fins acadêmicos, pedagógicos e/ou de desenvolvimento contemplados por seus membros.

[a ser preenchido pelo pesquisador]

Pesquisador: Wagner Signoretti

Data: 03 / 01 / 2018

Participante: _____

Condições especiais (caso não haja condições especiais, escreva "nenhuma"):

continua no verso

Por favor, indique sua posição em relação aos termos acima:

Estou de pleno acordo com os termos acima.

Em anexo registro condições adicionais para este estudo.

Assinatura do participante

Assinatura do pesquisador

B.2 Entrevistas

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Nós, Wagner Gonçalves Signoretti (aluno de mestrado) e Isabel Harb Manssour (professora orientadora), responsáveis pela pesquisa: **técnicas de visualização e exploração de dados multivariados geoespaciais**, estamos fazendo um convite para você participar como voluntário nesse estudo.

Esta pesquisa pretende validar um modelo de visualização de dados para auxiliar na exploração de dados geoespaciais. Para isso precisamos mapear atividades e recursos de visualização de dados utilizados durante esse processo e a avaliação do modelo proposto. Não há benefícios a curto prazo para os participantes dessa pesquisa, contudo, ao término desse estudo é esperada a seguinte contribuição: avaliação do modelo e das técnicas de visualização e interação apresentadas.

Para a coleta dos dados será utilizada a técnica de entrevista seguindo roteiro semiestruturado e observação do desenvolvimento das atividades pré-estabelecidas no sistema sob análise. Entendemos que há riscos mínimos durante essas atividades como: divulgação de dados confidenciais (quebra de sigilo) e desconforto ou constrangimento durante gravações de áudio. Lembrando que o objetivo deste estudo não é avaliar o participante, mas, sim, avaliar os processos de trabalho e o sistema computacional que o participante estará usando durante o teste. O uso que se faz dos registros efetuados durante o teste é estritamente limitado a atividades acadêmicas e buscaremos garantir seu anonimato e confidencialidade.

Outras informações importantes:

- Durante todo o período da pesquisa você tem o direito de esclarecer qualquer dúvida ou pedir qualquer outro esclarecimento, bastando para isso entrar em contato com Wagner Signoretti através do telefone (51) 99507 9330.
- Você tem garantido o seu direito de não aceitar participar ou de retirar sua permissão, a qualquer momento, sem nenhum tipo de prejuízo ou retaliação, pela sua decisão.
- As informações desta pesquisa serão confidenciais, e serão divulgadas apenas em eventos ou publicações científicas, não havendo identificação dos participantes, a não ser entre os responsáveis pelo estudo, sendo assegurado o sigilo sobre sua participação.
- Ao assinar este termo de consentimento, você não abre mão de nenhum direito legal que teria de outra forma.
- Não assine este termo de consentimento a menos que tenha tido a oportunidade de fazer perguntas e tenha recebido respostas satisfatórias para suas dúvidas.
- Se você concordar em participar deste estudo, você rubricará todas as páginas e assinará e datará duas vias originais deste termo de consentimento. Você receberá uma das vias para seus registros e a outra será arquivada pelo responsável pelo estudo.

Eu, _____, após a leitura deste documento e de ter tido a oportunidade de conversar com o pesquisador responsável, para esclarecer todas as minhas dúvidas, acredito estar suficientemente informado, ficando claro para mim que minha participação é voluntária e que posso retirar este consentimento a qualquer momento sem penalidades ou perda de qualquer benefício. Estou ciente também dos objetivos da pesquisa, dos procedimentos aos quais serei submetido, dos possíveis danos ou riscos deles provenientes e da garantia de confidencialidade e esclarecimentos sempre que desejar.

Rubrica do participante

Rubrica dos pesquisadores responsáveis



Pontifícia Universidade Católica do Rio Grande do Sul
Pró-Reitoria de Graduação
Av. Ipiranga, 6681 - Prédio 1 - 3º. andar
Porto Alegre - RS - Brasil
Fone: (51) 3320-3500 - Fax: (51) 3339-1564
E-mail: prograd@pucrs.br
Site: www.pucrs.br