



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de São José do Rio Preto

IBILCE - Instituto de Biociências, Letras e Ciências Exatas

Renan Augusto Pupin de Oliveira

Uma abordagem de visualização de informações multivaloradas
em um contexto geoespacial

Presidente Prudente
2025

Renan Augusto Pupin de Oliveira

Uma abordagem de visualização de informações multivaloradas
em um contexto geoespacial

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto

Orientador: Prof. Dr. Danilo Medeiros Eler

Presidente Prudente
2025

O48a

Oliveira, Renan Augusto Pupin de

Uma abordagem de visualização de informações multivaloradas em um contexto geoespacial / Renan Augusto Pupin de Oliveira. --

Presidente Prudente, 2025

71 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Tecnologia, Presidente Prudente

Orientador: Danilo Medeiros Eler

1. Visualização de Informações. 2. Mineração de dados. 3. Dados geoespaciais. I. Título.

IMPACTO POTENCIAL DESTA PESQUISA

O presente trabalho de mestrado possui um impacto no campo da Visualização de Informações e Mineração de Dados Geoespaciais ao propor uma ferramenta inovadora para a análise exploratória de grandes volumes de dados multivalorados e georreferenciados. O modelo desenvolvido amplia a capacidade de interpretação e abstração de conhecimento, contribuindo para a evolução dos Sistemas de Informação Geográfica (SIG) e para o desenvolvimento de novas abordagens metodológicas.

No aspecto técnico, a ferramenta GeoVis Explorer representa um avanço em termos de usabilidade e funcionalidade, permitindo a coordenação de múltiplas visualizações em um ambiente georreferenciado. Seu impacto se reflete na otimização do processo de tomada de decisão, facilitando a identificação de padrões e tendências em dados complexos.

Socialmente e economicamente, a pesquisa tem potencial para apoiar gestores, urbanistas, analistas de dados e pesquisadores em diferentes domínios, como planejamento urbano, monitoramento ambiental e análise socioeconômica. Ao fornecer uma abordagem mais intuitiva para a exploração de dados, a ferramenta pode ser utilizada para melhorar a gestão de cidades inteligentes, auxiliando na elaboração de políticas públicas baseadas em dados. Empresas e instituições públicas podem adotar essa solução para otimizar processos e tomar decisões baseadas em evidências, resultando em ganhos financeiros e melhor aproveitamento de recursos.

Por fim, o trabalho expande o conhecimento na área de visualização de informações, consolidando-se como uma contribuição relevante para pesquisas futuras e o aprimoramento de abordagens exploratórias de dados geoespaciais.

POTENTIAL IMPACT OF THIS RESEARCH

This Master's thesis has an impact on the field of Information Visualization and Geospatial Data Mining by proposing an innovative tool for the exploratory analysis of large volumes of multivalued and georeferenced data. The developed model expands the capacity for interpretation and knowledge abstraction, contributing to the evolution of

Geographic Information Systems (GIS) and to the development of new methodological approaches.

From a technical perspective, the GeoVis Explorer tool represents an advance in terms of usability and functionality, allowing the coordination of multiple visualizations in a georeferenced environment. Its impact is reflected in the optimization of the decision-making process, facilitating the identification of patterns and trends in complex data.

Socially, the research has the potential to support managers, urban planners, data analysts and researchers in different domains, such as urban planning, environmental monitoring and socioeconomic analysis. By providing a more intuitive approach to data exploration, the tool can be used to improve the management of smart cities, assisting in the development of data-driven public policies. Companies and public institutions can adopt this solution to optimize processes and make evidence-based decisions, resulting in financial gains and better use of resources.

Finally, the work expands knowledge in the area of information visualization, consolidating itself as a relevant contribution to future research and the improvement of exploratory approaches to geospatial data.



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Presidente Prudente

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: Uma abordagem de visualização de informações multivaloradas em um contexto geoespacial

AUTOR: RENAN AUGUSTO PUPIN DE OLIVEIRA

ORIENTADOR: DANILO MEDEIROS ELER

Aprovado como parte das exigências para obtenção do Título de Mestre em Ciência da Computação, área: Computação Aplicada pela Comissão Examinadora:

Prof. Dr. DANILO MEDEIROS ELER (Participação Virtual)
Departamento de Matematica e Computacao / Faculdade de Ciencias e Tecnologia de Presidente Prudente

Prof. Dr. MILTON HIROKAZU SHIMABUKURO (Participação Virtual)
Departamento de Matematica e Computacao / Faculdade de Ciencias e Tecnologia de Presidente Prudente - FCT/Unesp

Prof. Dr. DANILO BARBOSA COIMBRA (Participação Virtual)
Departamento de Ciências da Computação / Universidade Federal da Bahia - UFBA

Presidente Prudente, 14 de maio de 2025


unesp
Danilo Medeiros Eler

2025.06.05
11:11:30
-03'00'

AGRADECIMENTOS

A conclusão deste mestrado foi uma jornada desafiadora e enriquecedora, que não teria sido possível sem o apoio e incentivo de diversas pessoas.

Agradeço a todos meus professores, mas especialmente ao meu orientador, Prof. Dr. Danilo Medeiros Eler, por sua paciência, dedicação e orientação ao longo de todo o processo. Seu conhecimento e incentivo foram fundamentais para o desenvolvimento deste trabalho de pesquisa.

Aos meus familiares e amigos, que sempre estiveram ao meu lado, oferecendo apoio e palavras de incentivo e motivação nos momentos mais desafiadores. Aos amigos e colegas de pesquisa, que tornaram essa caminhada mais leve com suas trocas de ideias, debates enriquecedores e apoio mútuo.

Também expresso minha gratidão à instituição UNESP e ao programa de pós-graduação, pelo suporte acadêmico e pelas oportunidades de aprendizado que me proporcionaram ao longo deste percurso, além das contribuições para o desenvolvimento e evolução da Ciência no país.

Por fim, agradeço a todos que, de alguma forma, contribuíram para a realização deste trabalho. Cada palavra de incentivo, cada conselho e cada gesto de apoio foram essenciais para que eu chegasse até aqui.

Muito obrigado(a)!

RESUMO

As técnicas de visualização de informações têm sido cada vez mais utilizadas para enfrentar os desafios de realizar estudos com grandes volumes de dados, exigindo análises detalhadas e ferramentas cada vez mais robustas para extrair conclusões relevantes. Nesse contexto, as ferramentas de visualização surgem como um valioso recurso para simplificar a exploração desses dados, reduzindo a necessidade de iterações e simplificando a identificação de padrões e tendências das informações. Além dos dados multivalorados, os dados geoespaciais também exigem ferramentas específicas para explorar essas informações, pois muitas vezes é necessário realizar análises de correlações espaciais e extrair respostas desses dados. Apesar de existirem ferramentas que implementam as técnicas de visualização de dados geoespaciais, muitas delas enfrentam limitações que impactam na capacidade de explorar as informações, como, por exemplo, a baixa quantidade de técnicas que suportam para o processo exploratório, limitações relacionadas à capacidade de combinação e coordenação entre técnicas e, mais importante, por serem criadas em domínios específicos que não permitem serem utilizadas para outros conjuntos de dados. Diante destes desafios, este trabalho de mestrado desenvolveu uma nova ferramenta, bem como um modelo de análise de dados multivalorados e geoespaciais, permitindo a combinação de representações visuais em um ambiente georreferenciado. A ferramenta possibilita a análise exploratória integrada de informações, com foco na coordenação de múltiplas visualizações em um contexto georreferenciado, estabelecendo avanço nas áreas de visualização de informações e análise de dados. As conclusões deste trabalho apresentam os benefícios da ferramenta desenvolvida, como a melhoria na percepção intuitiva dos dados, a facilitação da análise exploratória e a ampliação das possibilidades de abstração de conhecimento para a tomada de decisão. Os testes com usuários indicaram que a abordagem adotada reduz a complexidade da exploração de dados, destacando-se como uma solução versátil para estudos em áreas que exigem análise visual interativa.

Palavras-chave: Mineração de dados, visualização de informações, dados geoespaciais, cidades inteligentes, Sistema de Informação Geográfica, Múltiplas visões coordenadas.

ABSTRACT

Information visualization techniques have been increasingly used to address the challenges of conducting studies with large volumes of data, requiring detailed analyses and increasingly robust tools to extract relevant conclusions. In this context, visualization tools emerge as a valuable resource to simplify the exploration of this data, reducing the need for iterations and simplifying the identification of patterns and trends in the information. In addition to multivalued data, geospatial data also require specific tools to explore this information, as it is often necessary to perform spatial correlation analyses and extract answers from this data. Although there are tools that implement geospatial data visualization techniques, many of them face limitations that impact the ability to explore information, such as, for example, the low number of techniques they support for the exploratory process, limitations related to the ability to combine and coordinate between techniques and, most importantly, because they are created in specific domains that do not allow them to be used for other data sets. In view of these challenges, this master's thesis developed a new tool, as well as a model for analyzing multivalued and geospatial data, allowing the combination of visual representations in a georeferenced environment. The tool enables integrated exploratory analysis of information, focusing on the coordination of multiple views in a georeferenced context, establishing progress in the areas of information visualization and data analysis. The conclusions of this work present the benefits of the developed tool, such as the improvement in the intuitive perception of data, the facilitation of exploratory analysis and the expansion of the possibilities of knowledge abstraction for decision-making. User tests indicated that the adopted approach reduces the complexity of data exploration, standing out as a versatile solution for studies in areas that require interactive visual analysis.

Keywords: Data mining, information visualization, geospatial data, smart cities, Geographic Information System, Multiple coordinated views.

LISTA DE FIGURAS

Figura 1 – Modelo de coleta de informações e suporte de decisões para cidades inteligentes (adaptada de Harrison et al. 2010)

Figura 2 – Visibilidade de pontos de interesse por meio de sensores instalados na cidade de Santander (adaptada de Sanchez et al. 2014)

Figura 3 – Processo de descoberta de conhecimento KDD (adaptada de: Camilo e Silva, 2009)

Figura 4 – Modelo de referência de visualização de Card et al. (adaptada de Freitas et al. 2001)

Figura 5 – Estrutura geral dos Sistemas de Informação Geográfica (adaptada de Câmara et al., 2001)

Figura 6 – Modelo para análise visual de dados multivalorados no contexto geoespacial (adaptada de Signoretti, 2018)

Figura 7 – Ferramenta para análise interativa de dados multivalorados no contexto espacial (adaptada de Zhang, Zhiyuan, et al, 2013)

Figura 8 – Visualização da análise de composição de partículas de um voo gerada pelos pesquisadores climáticos (Zhang, Zhiyuan, et al. 2013)

Figura 9 – Mapa socioeconômico e gráficos dinâmicos gerados no Ceará em mapas interativos. (De Medeiros, et al, 2011)

Figura 10 – Metodologia de análise interativa (Autor)

Figura 11 – Estrutura das funcionalidades da ferramenta

Figura 12 – Componente de modal com informações da instância de dados. (Autor)

Figura 13 – Interface da ferramenta GeoVis demonstrando as tecnologias utilizadas.

Figura 14 – Etapa de carregamento de dados na plataforma GeoVis. (Autor)

Figura 15 – Seleção dos atributos de geolocalização. (Autor)

Figura 16 – Escolha do campo a ser filtrado (Autor)

Figura 17 – Interface para a inserção do valor a ser filtrado (Autor)

Figura 18 – Demonstração de instâncias filtradas pela funcionalidade de filtros da ferramenta. (Autor)

Figura 19 – Demonstração de instâncias filtradas de modo geoespacial. (Autor)

Figura 20 – Técnica mapas de calor. (Autor)

Figura 21 – Técnica de gráfico em marcadores com o gráfico polar. (Autor)

Figura 22 – Técnica de gráfico em marcadores com o gráfico de setores. (Autor)

Figura 23 – Técnica de gráfico em marcadores com o gráfico de colunas. (Autor)

Figura 24 – Técnica de clusterização de marcadores. (Autor)

Figura 25 – Técnica de clusterização de marcadores com gráfico de setores nos clusters. (Autor)

Figura 26 – Técnica de clusterização de marcadores e gráfico de setores. (Autor)

Figura 27 – Técnica de cores em marcadores do mapa. (Autor)

Figura 28 – Técnica de coordenadas paralelas aplicando filtros no mapa. (Autor)

Figura 29 – Gráficos com dados das instâncias visíveis do mapa. (Autor)

Figura 30 – Mapa e Gráfico de barras categorizado por estados. (Autor)

Figura 31 – Instâncias filtradas pelo filtro interativo do gráfico. (Autor)

Figura 32 – Instâncias filtradas pelo filtro interativo das coordenadas paralelas. (Autor)

Figura 33 – Diferentes técnicas de visualização coordenadas na ferramenta. (Autor)

Figura 34 – Representação dos municípios com índice de escolaridade abaixo da média. (Autor)

Figura 35 – Representação dos municípios com índice de escolaridade abaixo da média no Gráfico de Setores agrupados por estado. (Autor)

Figura 36 – Representação gráfica dos municípios em relação à frequência escolar. (Autor)

Figura 37 – Representação gráfica do mapa com o filtro aplicado pela técnica Coordenadas Paralelas. (Autor)

LISTA DE TABELAS

Tabela 1 – Comparativo de funcionalidade entre estudos relacionados ao tema de visualização de informações georreferenciadas

LISTA DE APÊNDICES

Apêndice A – QUESTIONÁRIO PARA AS ENTREVISTAS.

LISTA DE ABREVIATURAS

CMV - Múltiplas visões coordenadas.

SUMÁRIO

1. INTRODUÇÃO	13
2. FUNDAMENTAÇÃO TEÓRICA	15
2.1. Cidades Inteligentes	15
2.2. Análise de dados e mineração de dados	17
2.3. Visualização de Informações	18
2.3.1. Visualização de Informações Geoespaciais	19
3. TRABALHOS RELACIONADOS	21
3.1. Revisão da literatura	21
3.2. Comparativo entre estudos	25
4. ABORDAGEM DE VISUALIZAÇÃO DE CONJUNTOS GEOESPACIAIS	29
4.1. A ferramenta GeoVis Explorer	30
4.1.1. Tecnologias	32
4.1.2. Entrada de dados	33
4.1.3. Filtros	35
4.1.4. Técnicas de representações gráficas	38
4.1.4.1. Mapas de calor	38
4.1.4.2. Gráficos em marcadores	39
4.1.4.3. Clusterização de marcadores	42
4.1.4.4. Cores em marcadores	45
4.1.4.5. Coordenadas paralelas	45
4.1.4.6. Representações gráficas	46
4.1.5. Exploração interativa	47
4.2. Etapas para avaliação do modelo	50
5. RESULTADOS DA ANÁLISE VISUAL	52
5.1. Resultados	52
5.1.1. Análise exploratória	53

5.2. Entrevistas para Avaliação da GeoVis Explorer	57
5.2.1. Condução das entrevistas	57
5.2.2. Pareceres das entrevistas	58
6. CONCLUSÕES E TRABALHOS FUTUROS	61
6.1. Conclusões	61
6.2. Trabalhos futuros	62
REFERÊNCIAS	
APÊNDICE	

1. INTRODUÇÃO

A crescente evolução tecnológica na área da Ciência da Computação tem gerado desafios significativos no que diz respeito à manipulação e compreensão de grandes volumes de dados (Vieira, *et al.*, 2012). À medida que a necessidade de análises detalhadas desses dados se torna necessária, deparamo-nos com a complexidade de conduzir numerosas iterações para extrair informações conclusivas sobre informações que estes dados representam (Freitas, 2000). Nesse contexto, surge a necessidade de desenvolver uma abordagem que simplifique a exploração desses vastos conjuntos de dados, por exemplo, por meio de visualizações (Card, 1999). As representações visuais dos dados se destacam como uma ferramenta valiosa para este tipo de análise, reduzindo a necessidade de interações intensivas e facilitando a identificação de padrões e tendências, tornando os dados mais informativos (Khan e Shah, 2011).

Com a popularização de dispositivos móveis, GPS e sensores IOTs, a quantidade de dados relacionados à localização, conhecidos como dados geoespaciais, tem sofrido um aumento significativo (Cook, 2005). A visualização desses dados geoespaciais em mapas oferece uma maneira eficaz de analisar como diferentes atributos se correlacionam espacialmente, possibilitando o desenvolvimento de modelos de análise visual no contexto geoespacial (Signoretti, 2018). Esse enfoque apresenta oportunidades únicas para aprimorar a gestão urbana e tornar as cidades mais inteligentes (Batty *et al.*, 2012), através do desenvolvimento de ferramentas que simplifiquem a exploração e a interpretação desses dados geoespaciais (Medeiros *et al.*, 2010).

No contexto deste trabalho, foram analisados trabalhos relacionados para observar contribuições e avanços apresentados por outros autores. Muitos destes trabalhos, descritos nas seções subsequentes, concluíram que a visualização de mapas coordenados com outras técnicas de visualização potencializam o processo exploratório de estudo dos dados. No entanto, as ferramentas desenvolvidas até então apresentam limitações na diversidade de conjuntos de dados que conseguem representar e, em alguns casos, na variedade de técnicas de visualização e exploração. (ver Tabela 1 na Seção 3.2 para uma comparação detalhada). Outro fator importante é que, ao analisar o contexto de cidades inteligentes e considerar o fato de estarem cada vez mais conectadas, torna-se desejável dispor de métodos para a análise de dados de diversos

domínios (Roberts, 1998), sendo esta também uma das grandes incapacidades de ferramentas existentes de suportar diante a necessidade da análise exploratória.

Esta dissertação apresenta um modelo e ferramenta para enfrentar desafios atualmente apresentados na literatura para analisar informações georreferenciadas e espaciais. Assim, possibilitando a combinação de representações visuais em um ambiente georreferenciado, sob a forma de mapas interativos. Consequentemente, este trabalho permitirá que usuários simplifiquem a exploração de informações e a extração de conhecimento a partir dos dados em estudo, fornecendo uma perspectiva integrada das representações das informações em uma única visualização e resultando em um ambiente dinâmico e flexível para a análise exploratória de informações. De forma específica, foi desenvolvida uma plataforma de visualização para análise de dados georreferenciados, que oferece visualizações capazes de facilitar a compreensão dos dados associados, e a criação de um ambiente coordenado para aprimorar e facilitar a análise.

A principal contribuição deste trabalho é o desenvolvimento do modelo e a ferramenta GeoVis Explorer, que possibilita a análise visual interativa de dados geoespaciais e multivalorados em um modelo de múltiplas visões coordenadas. A ferramenta propõe a capacidade da combinação de representações gráficas e espaciais, com mecanismos interativos de exploração, seleção e filtragem de dados, superando as limitações de algumas abordagens tradicionais existentes. Os resultados obtidos demonstram um impacto da ferramenta na identificação de padrões e relações espaciais, proporcionando ganhos cognitivos no processo de análise exploratória. Além disso, a validação com dados reais evidenciou sua flexibilidade e aplicabilidade em diferentes domínios, destacando-se como uma solução capaz de potencializar e auxiliar a tomada de decisão baseada em dados georreferenciados e apoiada por representações visuais.

2. FUNDAMENTAÇÃO TEÓRICA

Com o objetivo de contextualizar os tópicos abordados neste trabalho, esta seção introduz a motivação relacionada a fundamentação teórica sobre cada um dos temas relacionados ao estudo. Os temas que fundamentam este trabalho relacionam Técnicas de Visualização de Informações, Dados Geoespaciais e Sistemas de Informações Geográficas para exploração de dados no contexto de Cidades Inteligentes.

2.1. Cidades Inteligentes

Com a expansão populacional e a migração do homem do campo para as cidades, a infraestrutura, benefícios e recursos disponíveis não são suficientes para suportar as demandas e o crescimento da população (Kon e Santana, 2016). Uma maneira de lidar com este problema, é tornar as cidades mais sustentáveis, otimizar recursos e analisar dados que podem ajudar os governos a tomarem medidas que tornarão cidades mais inteligentes. Cidades inteligentes, ou em inglês *Smart Cities*, é um conceito que envolve o uso da tecnologia da informação nas cidades, a fim de melhorar o gerenciamento de seus serviços, sua infraestrutura, seus recursos e a qualidade de vida dos cidadãos (Harrison *et al.*, 2010). Buscando este objetivo, tecnologias de análise de dados podem ser aplicadas para analisar informações armazenadas em grandes bases de dados. No entanto, realizar a análise destes grandes volumes de dados de forma eficiente não é uma tarefa simples. Sendo assim, aplicações que auxiliam a exploração, mineração e visualização humana destas informações são essenciais para que este processo se torne viável. Uma das áreas das cidades inteligentes que pode se beneficiar deste tipo de aplicação, é a área da saúde pública, onde eventos são registrados pelos órgãos competentes junto aos dados de localidade da região envolvida (Bastos, 2020). Estes dados de localização podem ser utilizados para realizar estudos e fornecer suporte para que a gestão pública tenha ferramentas que auxiliem a tomada de decisão.

Harrison *et al.* (2010) traz um modelo, apresentado na Figura 1, com princípios para sistemas de informação que dão suporte para a tomada de decisão de cidades inteligentes. Os autores propõem uma abordagem que utiliza o conceito de instrumentação de sensores interconectados para a inclusão de análises complexas de

modelagem, otimização e visualização para que se possa tomar decisões mais inteligentes.

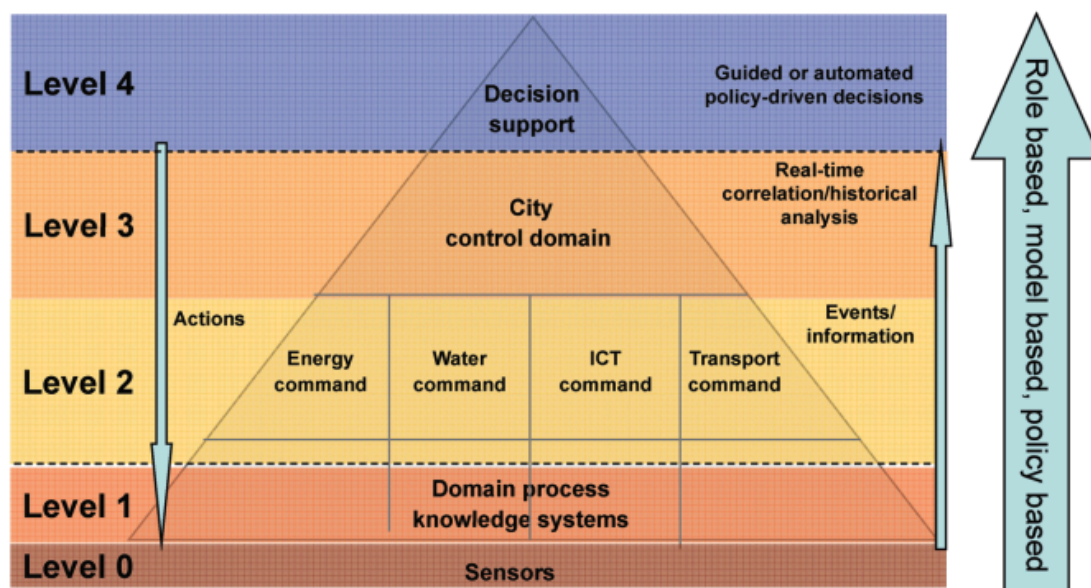


Figura 1: Modelo de coleta de informações e suporte de decisões para cidades inteligentes (adaptada de Harrison et al. 2010).

O modelo apresentado busca definir camadas estruturadas de comunicação entre diferentes domínios que partem da coleta de informações brutas (nível mais baixo), para sistemas de suporte à decisão (níveis mais altos). Um exemplo de aplicação de cidades inteligentes, foi o projeto na cidade de Santander na Espanha. Foram instalados mais de 20 mil sensores para coletar informações da cidade, como temperatura, espaços livres em estacionamento, identificadores de ponto de interesse e localidade de ônibus e táxis. A plataforma SmartSantander trouxe a visibilidade destes sensores em um mapa, onde foi possível analisar a utilização destes lugares e observar pontos de interesse na cidade, como é possível visualizar na Figura 2 (Sanchez *et al.*, 2014).

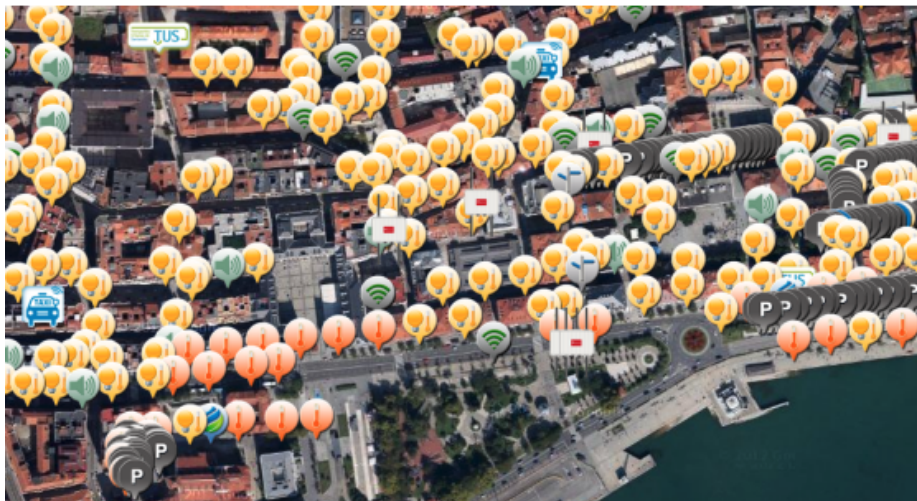


Figura 2: Visibilidade de pontos de interesse por meio de sensores instalados na cidade de Santander (adaptada de Sanchez et al. 2014).

2.2. Análise de dados e mineração de dados

Nos últimos anos pode ser notado um crescente interesse pela área de Ciência de Dados. Na era do Big Data, a informação tem se tornado cada vez mais valiosa, pois permite um entendimento mais profundo sobre comportamentos e ganhos diversos ao ser explorada. Com esse objetivo, esta Ciência busca obter conhecimento das informações de forma sistemática, fornecendo métodos, processos e modelos para organizar o conhecimento armazenado (Amaral, 2016).

A Mineração de Dados surgiu no final da década de 1980 com a necessidade de se explorar os dados de maneira eficiente. Com o volume de dados armazenados crescendo diariamente, técnicas tradicionais de exploração de dados não são mais adequadas para analisar grandes conjuntos de dados. A Mineração dos dados visa prover técnicas de exploração que dão sentido aos dados, além de viabilizar esta coleta e normalização dos mesmos para que sejam possíveis para o estudo humano (Camilo e Silva, 2009).

Fayyad *et al.* (1996), propõe que, até então, o modelo para transformação de dados em informações e conhecimento, é um processamento manual das informações em relatórios a serem analisados por especialistas. No entanto, o KDD (*Knowledge Discovery in Databases* ou Descoberta de Conhecimento nas Bases de Dados) surge

como uma tentativa para solucionar o problema da análise de dados em grande escala. A proposta do autor segue o conceito de dividir a análise em fases interativas, nas quais as informações vão sendo selecionadas em fases e transformadas para que seja feita uma análise exploratória mais precisa, como demonstrado na Figura 3.

A proposta do autor foi definir um modelo que compreendesse todas as etapas da descoberta do conhecimento, envolvendo etapas de entendimento, seleção e transformação de dados antes das etapas da descoberta de conhecimento.

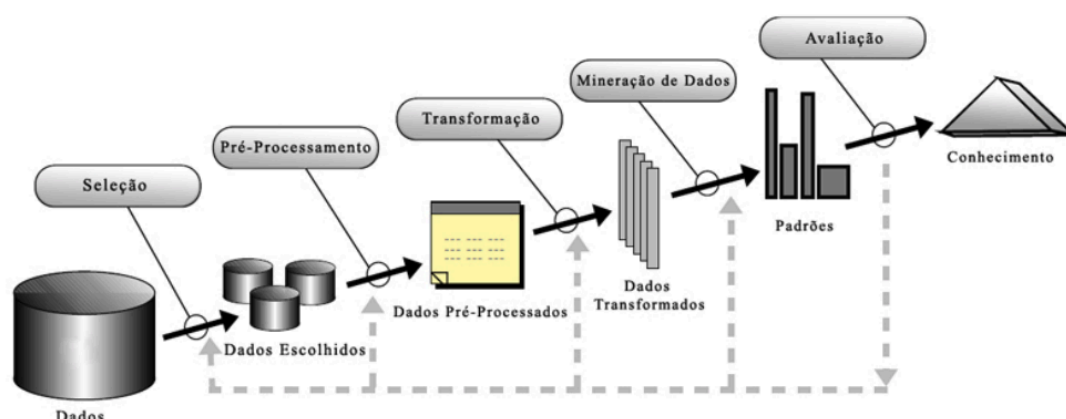


Figura 3: Processo de descoberta de conhecimento KDD (adaptado e traduzido de: Fayyad, 1996).

2.3. Visualização de Informações

Mesmo com técnicas que simplificam o entendimento e a descoberta dos dados, em certos casos, a Mineração de Dados resulta em informações que não são trivialmente analisáveis, sendo necessário aplicar um grande esforço para entender o que os dados representam. Assim, surge a necessidade de se utilizar abordagens alternativas para o estudo e descoberta de conhecimento, podendo utilizar outras técnicas que viabilizem a análise e compreensão dos dados. A Visualização de Informações é uma área que estuda métodos e técnicas de representações gráficas, com o objetivo de auxiliar a exploração dos dados por meios visuais. Estas técnicas trazem ganhos relacionados à cognição e melhoria no nível de compreensão dos dados (Freitas *et al.*, 2001).

Card *et al.*, 1999, descrevem a visualização como sendo uma sequência de mapeamentos ajustáveis para uma representação visual, que permite a interação do

usuário com o espaço de informação (cristalização do conhecimento) (Card, 1999). Este modelo aborda o conceito de expor os dados brutos em tabelas relacionais, que são mapeadas em estruturas visuais ou representações, como imagens ou visualizações. As operações sobre essas visualizações são chamadas de transformações visuais, que possuem o objetivo de mostrar as informações do conjunto de dados por meio da interação e manipulação de diferentes pontos de observação sobre o conjunto de dados de interesse.

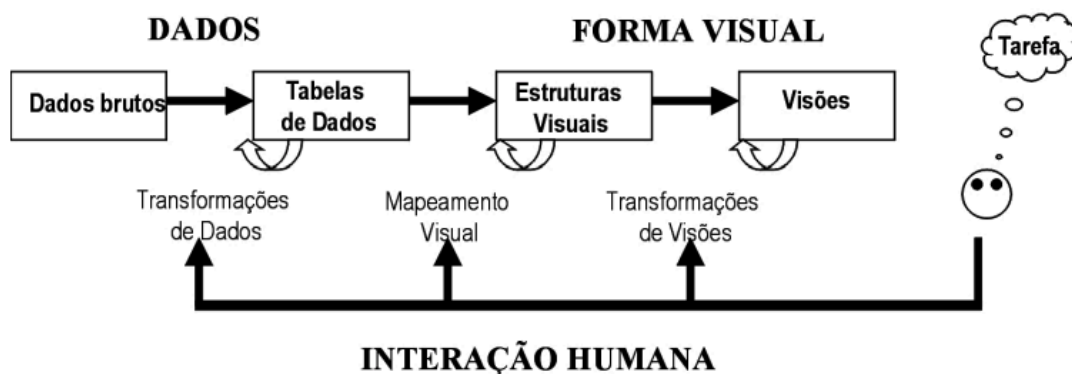


Figura 4: Modelo de referência de visualização de Card et al. (adaptada de Freitas et al. 2001).

2.3.1. Visualização de Informações Geoespaciais

Os dados geoespaciais são representações de informações que possuem um componente espacial. Estes dados podem representar informações geográficas e espaciais sobre a superfície terrestre, ou podem apenas representar um objeto em um determinado espaço (Câmara, 2001). Com este tipo de informação, é possível que sejam realizadas análises do espaço geográfico, pois agregam informações como localização, topologia, terreno, elevação e sistemas de coordenadas. Os dados geoespaciais de modo bruto, não são de fácil entendimento pela percepção humana, por isso torna-se necessário a utilização de ferramentas que auxiliam o processo de análise e exploração destes tipos de dados, gerando uma representação visual das informações (Aronoff, 1991). Os mapas surgem então como sendo visualizações que facilitam a análise destes dados geoespaciais, e seguindo o avanço tecnológico, eles também foram representados nas plataformas tecnológicas (Oliveira e Eler, 2016). Estas ferramentas recebem o nome de SIG (Sistema de Informações Geográficas), e surgiram com o objetivo de definir e

facilitar a análise da relação espacial entre os elementos (Kleiner e Meneguette, 2002). Além de prover um ambiente de mapas georreferenciado para a exploração dos dados, elas permitem que sejam adicionadas novas camadas de sobreposição com objetos vetoriais, como pontos, linhas, polígonos ou até dados matriciais, como imagens que podem ser armazenados em bancos de dados geoespaciais.

Segundo Câmara (2001), um SIG é um sistema especializado em representar estas análises, pois recuperam informações relacionadas ao domínio espacial e também as suas características alfanuméricas. Assim, o sistema representa determinada informação em sua interface quanto a sua localização terrestre por meio do atributo georreferenciado. Estes sistemas utilizam componentes que se relacionam de forma hierárquica, permitindo o usuário interagir com a interface para realizar as consultas ao banco de dados geográfico, como demonstrado na Figura 5.

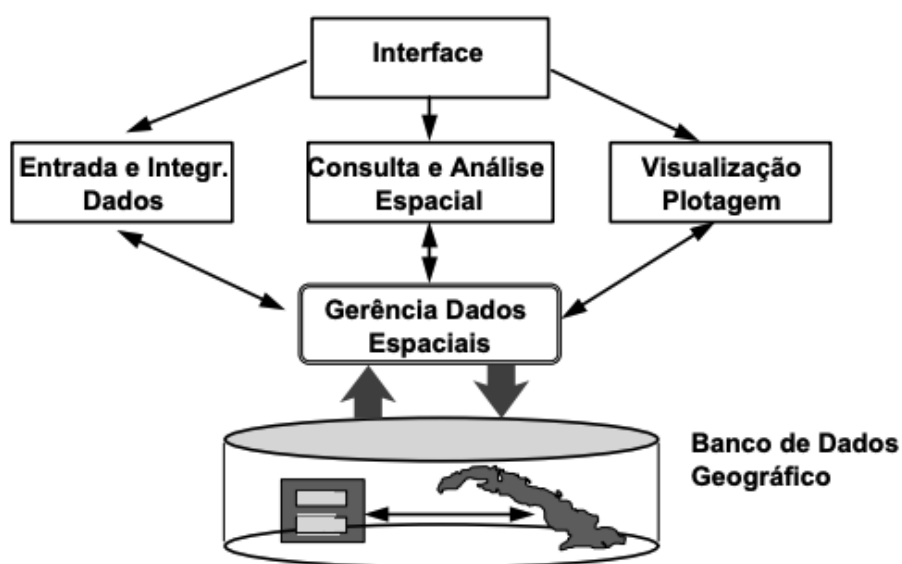


Figura 5: Estrutura geral dos Sistemas de Informação Geográfica (adaptada de Câmara et al., 2001)

Entre este modelo de consulta, temos os componentes de armazenamento e recuperação de dados (banco de dados geográfico), assim como funções de processamento da consulta e análise para a visualização final em uma interface.

3. TRABALHOS RELACIONADOS

Com o objetivo de identificar oportunidades de desenvolvimento de um novo modelo de visualizações coordenadas em um ambiente georreferenciado, esta seção busca analisar trabalhos relacionados de um âmbito comparativo em relação às suas funcionalidades. Nesta análise, foram levadas em consideração funcionalidades das ferramentas propostas em cada um dos estudos, como também a oportunidade de se aplicar a respectiva ferramenta em bases de dados genéricas.

3.1. Revisão da literatura

No artigo "Modelo para análise visual de informações multivaloradas no contexto espacial em múltiplas visões", Signoretti, 2018, traz uma abordagem de análise de informações geoespaciais em um mapa interativo. O trabalho propõe um novo modelo para análise visual de dados multivalorados no ambiente geoespacial. O autor integrou um mapa com representações gráficas, apresentando múltiplos atributos de forma simultânea e integradas à *CMV* (Múltiplas visões coordenadas) para auxiliar na análise visual (Roberts, 1998). Além de permitir uma maneira interativa para aplicação de filtros e analisar os atributos presentes na visualização, com a possibilidade de aplicar as mesmas técnicas em conjuntos de dados diferentes, como demonstrado na Figura 6.

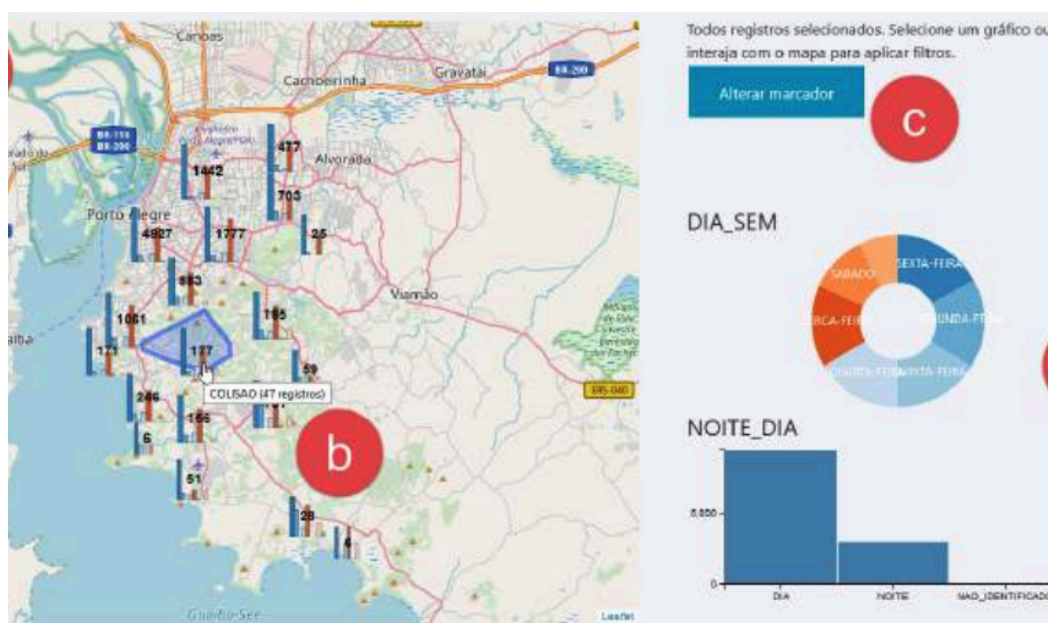


Figura 6: Modelo para análise visual de dados multivalorados no contexto geoespacial (adaptada de Signoretti, 2018).

No contexto de análise de dados públicos de Cidades Inteligentes, o autor fez a análise gráfica sobre dados de criminalidade, e analisou a relação entre os atributos do conjunto. Assim tornou-se possível constatar alguns fatos realizando apenas a análise gráfica: Os dias da semana com mais ocorrência de acidentes. Os tipos de acidentes mais comuns. As regiões mais propensas a determinados tipos de acidentes. O aumento de acidentes nos fins de semana. E o engavetamento como sendo o tipo de acidente mais comum. Para avaliar a eficácia do modelo, o estudo realizou uma avaliação com usuários por meio da técnica Entrevista em Profundidade (Boyce e Neale, 2006). Assim foram realizadas entrevistas individuais com alguns especialistas do assunto com o objetivo de explorar as suas perspectivas sobre uma determinada ideia ou situação. Seguindo a análise dos resultados dos experimentos e a entrevista, o modelo proposto confirmou, por meio da validação, que as técnicas nele presentes junto às visualizações e interações utilizadas ajudaram os participantes na análise de dados e na descoberta de conhecimento, sendo possível assim utilizar o modelo para analisar outros dados abertos do governo e ser aplicado para realizar estudos em Cidades Inteligentes. Já no artigo "Uma estrutura de análise visual interativa para dados multi-campos em um contexto geoespacial", Zhang, Zhiyuan, et al (2013) propõe uma ferramenta para análise de dados multi-campo no contexto espacial utilizando a plataforma Google Earth. Os autores buscaram demonstrar em um mesmo ambiente tanto as informações geográficas no mapa, como também as informações com múltiplos valores nos gráficos interativos, que por fim foi utilizado por pesquisadores para realizar um estudo climático utilizando a ferramenta.

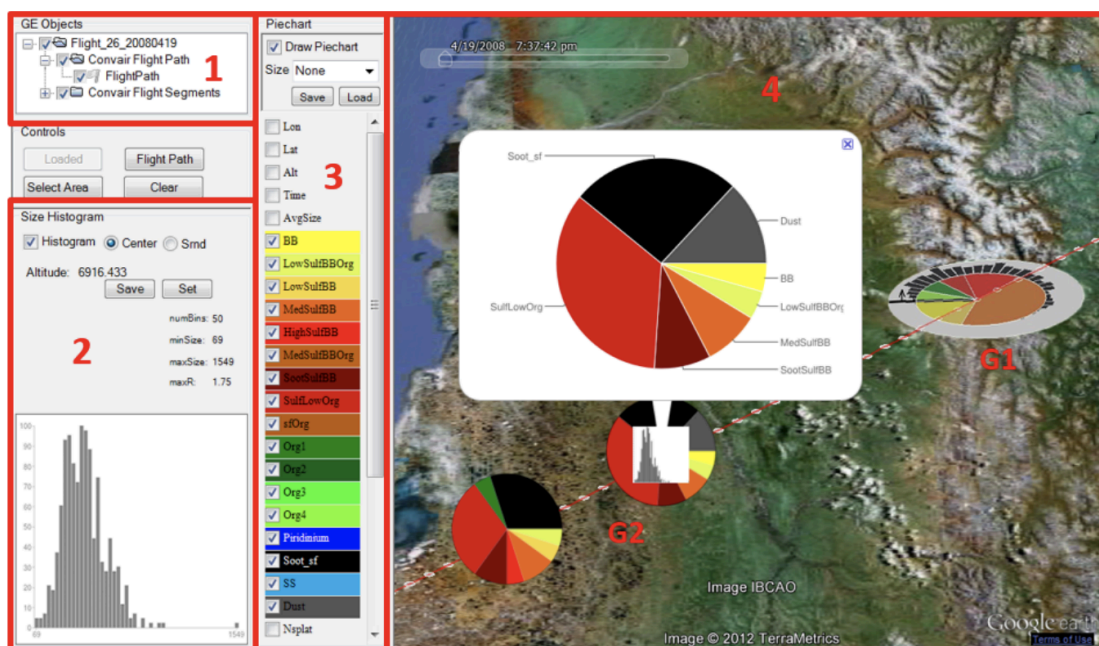


Figura 7: Ferramenta para análise interativa de dados multivalorados no contexto espacial (adaptada de Zhang, Zhiyuan, et al, 2013).

Como é possível analisar na Figura 7, o autor definiu áreas da aplicação para aplicar ferramentas interativas que permitem ao usuário configurar a visualização. Na Área 1 o autor trouxe o painel de controle para mostrar ou esconder objetos da visualização. Na Área 2 foi adicionado um histograma, permitindo a análise de distribuição dos dados. Já na Área 3 os autores listam os atributos do objeto representado e também possibilitam a seleção dos atributos da visualização. Por fim, a Área 4 apresenta o contexto espacial de interação, permitindo que o usuário interaja com os gráficos e expanda as visualizações gráficas desejadas (Figura 8).

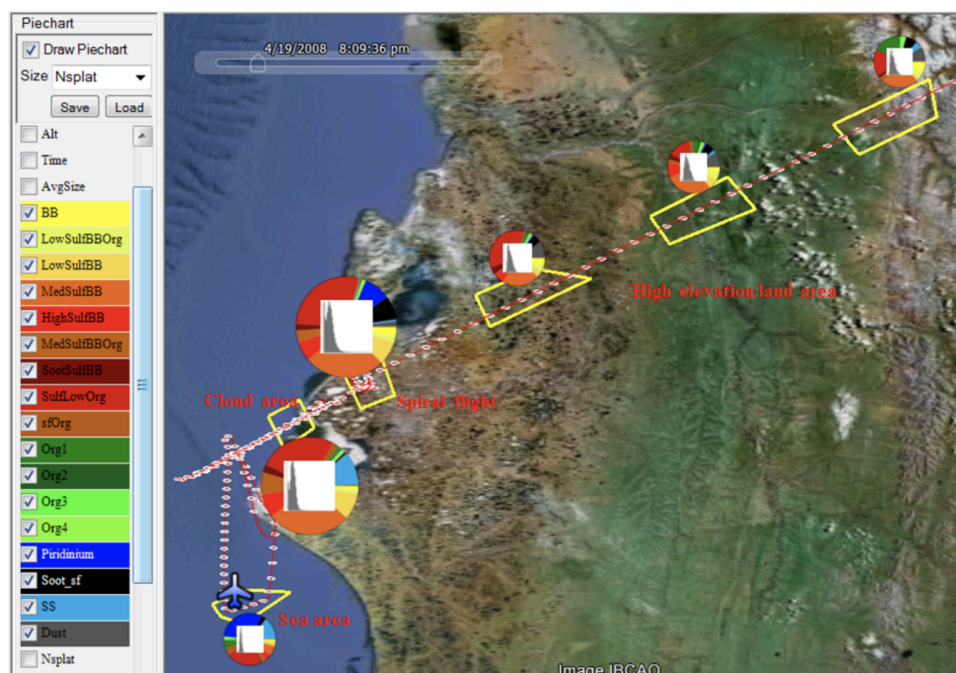


Figura 8: Visualização da análise de composição de partículas de um voo gerada pelos pesquisadores climáticos (Zhang, Zhiyuan, et al. 2013).

Por fim, o autor realizou um estudo focado na análise e pesquisa de informações climáticas, permitindo assim que cientistas deste tema visualizassem e realizassem análises interativas dos dados multi-campos geoespaciais na estrutura de duplo domínio consistindo em uma interface de análise geográfica e multivariada. Como conclusão, o trabalho pode ser validado pelo time de pesquisadores que puderam analisar as informações de uma perspectiva diferente das quais estavam acostumados a fazer. O software pode fornecer maneiras de analisar a distribuição de partículas de uma maneira visual em que facilitasse a descoberta de padrões relacionados ao voo de análise, provando assim que a ferramenta de análise interativa possibilita novas descobertas no campo de estudo em questão.

No artigo "Utilização de SIG-WEB usando software livre para disponibilização de dados georreferenciados na internet: Caso do Sistema Ceará em Mapas Interativos", o autor buscou utilizar o Sistema de Informação Geográfica (SIG) disponibilizado pelo IPECE (Instituto de Pesquisa e Estratégia Econômica do Ceará) em 2009 para realizar análises interativas sobre informações socioeconômicas, meio ambiente, imagens de satélite, recursos naturais e sistemas de transporte. Os mapas disponibilizados pelo sistema permitiam esconder camadas relacionadas aos diferentes domínios do mapa e a

aplicar filtros geoespaciais, possibilitando assim que fossem realizadas análises interativas sobre os dados (Figura 9). Assim, o estudo buscou verificar se a análise de indicadores gerados pela visualização, poderiam dar suporte aos membros públicos quanto à tomada de decisões, atuando de modo assistencial às políticas do estado do Ceará e podendo contribuir para estratégias do plano de desenvolvimento.

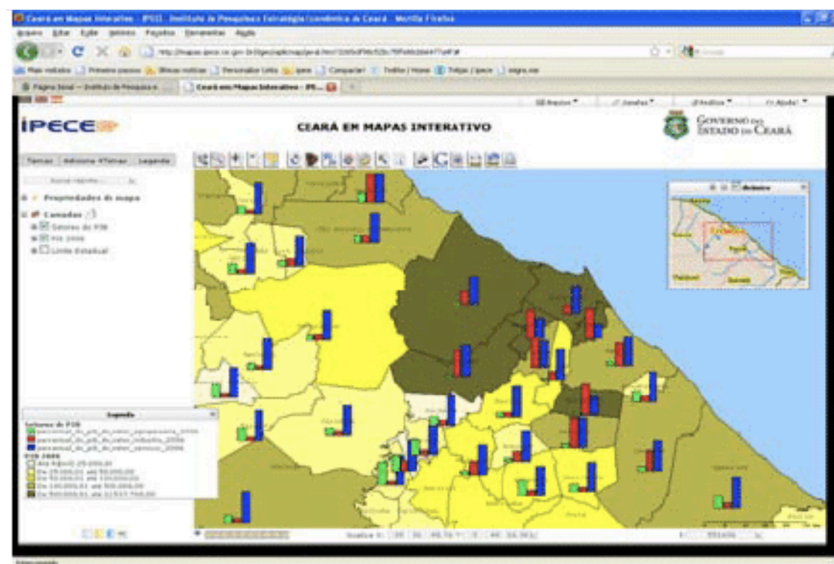


Figura 9: Mapa socioeconômico e gráficos dinâmicos gerados no Ceará em mapas interativos.
(De Medeiros, et al, 2011).

Com a análise exploratória da plataforma o autor pode identificar possíveis benefícios para estes tipos de aplicação. O contexto do estudo deste tema, favorece professores, pesquisadores e estudantes a ter acesso a informação atualizada para realizar pesquisas e estudos sobre os dados socioeconômicos e cartográficos atualizados. Além destes, os gestores públicos podem utilizar a ferramenta para executar análises que auxiliem a tomada de decisão e melhorem a gestão do território com ações governamentais.

3.2. Comparativo entre estudos

Para compreender-se similaridade e limitações entre abordagens de trabalhos na mesma área, foram relacionados na Tabela 1, estudos similares que utilizam técnicas de visualização em mapas para avaliar um ganho cognitivo em relação ao estudo de dados espaciais. Logo em seguida é realizada uma análise sobre temas deste estudo, além de buscar as limitações e pontos relevantes de cada um dos trabalhos.

Trabalho	Mapas interativos e geoinformação?	Mineração espacial?	Múltiplas visões coordenadas (CMV)?	Lista de visualizações
1 - Modelo para análise visual de dados multivariados no contexto geoespacial (Signoretti, 2018)	Sim	Sim	Sim	Gráfico de dispersão, linhas, barras e setores. Mapas interativos. Gráficos em marcadores.
2 - Utilização de SIG-WEB usando software livre para disponibilização de dados georreferenciados na internet: Caso do Sistema Ceará em Mapas Interativos (Medeiros, <i>et al.</i> , 2011)	Sim	Sim	Limitado	Gráfico de linhas e barras. Mapas interativos. Gráficos em marcadores.
3 - Design Study: Using Multiple Coordinated Views to Analyze Geo-referenced High-dimensional Datasets (Brodbeck, <i>et al.</i> , 2003)	Não	Limitado	Sim	Mapa de similaridade. Coordenadas paralelas. Mapa não interativo.
4 - Geospatial Digital Dashboard for Exploratory Visual Analytics (Sjöbergh e Tanaka, 2013)	Sim	Limitado	Sim	Gráfico de dispersão, linhas e barras. Mapas interativos. Clusterização.
5 - Interactive Dashboards: Using Visual Analytics for knowledge Transfer and Decision Support (Al-Hajj, Fisher e Pike, 2013)	Sim	Limitado	Limitado	Gráfico de linhas e barras. Mapas interativos. Clusterização. Gráficos em marcadores.
6 - Malaria-VisAnalytics: a tool for visual exploratory analysis of Brazilian public malaria data (Sironi <i>et al.</i> , 2022)	Sim	Limitado	Limitado	Mapa de calor. Representações de setores no mapa. Mapas interativos.
7 - Uma estrutura de análise visual interativa para dados multi-campos em um contexto geoespacial (Zhang, Zhiyuan, <i>et al.</i> 2013)	Sim	Limitado	Limitado	Mapas interativos. Gráfico de linhas e barras.

Tabela 1: Comparativo de funcionalidade entre estudos relacionados ao tema de visualização de informações georreferenciadas.

No comparativo descrito na Tabela 1, foram avaliados critérios relacionados ao suporte de cada uma destas ferramentas em relação à Mapas Interativos e geoinformação, capacidade de realizar a Mineração Espacial, possibilidade de utilizar Múltiplas Visões Coordenadas (*CMV*) e por último, as visualizações suportadas nestas ferramentas. Na seleção destes trabalhos, foram levadas em consideração aqueles que utilizam informações públicas relacionadas ao conceito de Cidades Inteligentes para criar as visualizações e análise dos dados referentes a informações públicas geradas pela população. É nítido que nestas ferramentas, suas respectivas aplicações foram desenvolvidas para suportarem um contexto específico de dados, como por exemplo Signoretti, 2018, buscar analisar dados de acidentes de trânsito na cidade de Porto Alegre, RS, Brasil, no ano de 2016 (fonte DataPOA). O primeiro estudo propõe um modelo para analisar visualmente dados multivariados no contexto geoespacial, que busca proporcionar uma compreensão mais profunda dos padrões complexos em conjuntos de dados georreferenciados, usando técnicas avançadas de visualização. O autor conclui que a integração de dados multivariados e informações geoespaciais pode fornecer *insights* valiosos em áreas como urbanismo e epidemiologia, no entanto dependem também da qualidade e tratamento dos dados em análise. Já o segundo estudo (Medeiros, *et al.*, 2011) destaca a implementação de um Sistema de Informações Geográficas (SIG) para web com um foco específico no Sistema de Dados do Ceará. O estudo concluiu que o uso de software livre para desenvolver SIGs web é uma alternativa eficaz, reduzindo custos e promovendo a acessibilidade a informações geoespaciais, no entanto o autor traz que a maior limitação está na necessidade de haver uma capacitação técnica para utilizando do sistema, e também para implementar e manter sistemas semelhantes. No terceiro estudo, o autor adota uma abordagem de estudo de *CMVs* para a análise de dados geoespaciais de alta dimensionalidade. Concluiu-se que há a eficácia dessa estratégia na facilitação da compreensão de padrões e relações espaciais. Contudo, a limitação reside na complexidade da implementação e na demanda por recursos computacionais para lidar com grandes conjuntos de dados. O quarto estudo (Sjöbergh e Tanaka, 2013) propõe um painel digital para análise visual exploratória geoespacial, onde as informações ficam dispostas em uma dashboard de análise. Como resultado, o estudo constatou a eficácia dessa ferramenta na facilitação da análise de dados georreferenciados, mas existe a necessidade de personalização para

diferentes contextos e expressa também a preocupação com a possibilidade de sobrecarregar usuários com excesso de informações. Já o estudo cinco (Al-Hajj, Fisher e Pike, 2013), explora o uso de painéis interativos para transferência de conhecimento e suporte à decisão, concluindo-se que, a abordagem destes painéis interativos que permitem a interação com o usuário, é valiosa para se analisar dados complexos em itens acionáveis. O autor também constata que a grande limitação destes painéis, podem estar na dependência da habilidade do usuário em saber interpretar corretamente as visualizações. Por último, o sexto estudo (Sironi et al., 2022) apresenta a ferramenta "Malaria-VisAnalytics", com o objetivo de permitir a análise visual exploratória de dados públicos e espaciais de malária no Brasil. No estudo os autores destacam na conclusão a utilidade dessa ferramenta para entender padrões epidemiológicos, mas citam também que é necessário cuidado com a generalização dos resultados obtidos para diferentes regiões e também na necessidade de atualização constante dos dados para relevância contínua dos dados em estudo. Ao compararmos estes estudos sumarizados na Tabela 1, podemos definir que embora existam ferramentas que buscam trazer uma proposta de visualização e exploração no ambiente de mapas interativos, grande parte delas possuem limitações quanto ao número de técnicas que disponibilizam, assim como limitação em implementar um modelo *CMV* interativo, assim como a incapacidade de realizar-se uma mineração espacial interativa com o processo exploratório. Outro ponto relevante é também relacionado a limitação em que as ferramentas possuem em representar conjuntos de dados multivalorados distintos, sendo as mesmas capazes de apenas criar um modelo de visualização para conjuntos de dados de domínio conhecidos. Diante do contexto de que as cidades possuem cada vez mais integrações tecnológicas e geram informações de diversos domínios de aplicação, ferramentas já desenvolvidas não suportam uma análise exploratória em todos os possíveis cenários de aplicação. Sendo assim, torna-se desejável uma ferramenta agnóstica quanto ao conjunto de dados, mas que permita a visualização de informações de diferentes conjuntos de dados, que possua uma análise exploratória em mapas interativos e que permita a exploração de informações utilizando diferentes visões exploratórias coordenadas.

4. ABORDAGEM DE VISUALIZAÇÃO DE CONJUNTOS GEOESPACIAIS

Este capítulo apresenta os métodos e ferramentas utilizados para a construção de uma metodologia para a realização da análise visual dos dados geoespaciais e multivalorados em um modelo de exploração que utiliza múltiplas visões coordenadas (CMV, do inglês *Coordinated Multiple Views*).

A metodologia proposta, busca implementar a coordenação entre visualizações gráficas e espaciais, assim como ferramentas de interação, seleção e exploração dos dados. Conforme apresentado na Figura 10, o modelo proposto realiza uma análise exploratória estruturada em etapas, na qual o usuário inicia com a adição dos dados de entrada, em seguida faz a configuração selecionando os atributos de localização, latitude e longitude. Após a configuração é possível aplicar filtros nos dados de maneira numérica ou espacial, adicionar visualizações gráficas ou geoespaciais dentro do mapa. Por último, pode-se interagir com as representações visuais para a realização da análise exploratória e descoberta de conhecimento.

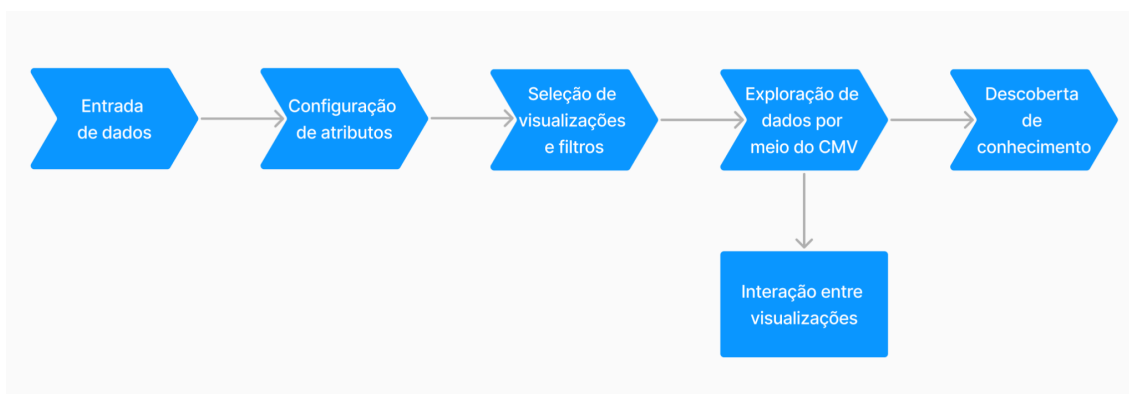


Figura 10: Metodologia de análise interativa (Autor).

Embora ferramentas de análise de dados geoespaciais existentes se proponham simplificar a análise de dados, elas enfrentam desafios (Signoretti, 2018) quanto à diversidade de conjuntos de dados que conseguem representar, assim como a falta de variedade de técnicas de visualização. Diante disso, foi desenvolvido uma nova ferramenta (GeoVis Explorer) para que possa superar essas limitações, permitindo a combinação de representações visuais genéricas em um ambiente georreferenciado e implementando o modelo proposto. Além desse propósito, é interessante proporcionar

aos usuários uma forma mais eficaz de explorar os dados e extrair conhecimento de conjuntos de dados complexos, especialmente em comparação com abordagens tradicionais para dados geoespaciais, oferecendo uma perspectiva integrada das representações das informações.

O objetivo principal foi desenvolver um modelo e uma ferramenta que possibilite a combinação de representações visuais em um contexto georreferenciado, por meio de mapas interativos. Com o ambiente interativo criado pela ferramenta, foram avaliados os ganhos na facilidade de exploração das informações e a extração de conhecimento desses dados, proporcionando uma visualização integrada das informações em um ambiente de visualização coordenada (destacado na Seção 4.1.5).

Apesar de existirem ferramentas que se propõem a analisar informações no contexto espacial, elas são pouco flexíveis quanto à capacidade de suportar a variedade dos dados em estudo, sendo limitadas a temas específicos e operando apenas com técnicas de visualização dependentes do cenário em que foram propostas (domínio fixo). Outro ponto relevante é que estas ferramentas oferecem pouco suporte para customizações das técnicas visuais secundárias, como a exploração interativa dos mapas e dados, que podem potencializar a visualização e ampliar a capacidade cognitiva na compreensão dos dados. Estas limitações serão descritas e exemplificadas em detalhes no Capítulo 5.

Por fim, este estudo também propõe a realização de testes para validar a eficácia deste ambiente dinâmico e flexível, diante da análise exploratória de dados espaciais e multivalorados, contribuindo com a evolução do conhecimento nas áreas de visualização de informações e assim permitir a análise, por exemplo, de dados originados por Cidades Inteligentes.

4.1. A Ferramenta GeoVis Explorer

A ferramenta foi construída para suportar conjuntos de dados georreferenciados e multivalorados. Para tanto, implementou-se técnicas visuais gráficas e mecanismos de interação, por exemplo, nos dados públicos abertos de diversos domínios. Assim, essa ferramenta foi utilizada para a realização de estudos exploratórios, com o objetivo de

aumentar a assertividade na tomada de decisão pelo usuário. Além disso, busca proporcionar ganhos cognitivos no processo exploratório, permitindo a personalização de diferentes tipos de visualizações interativas. Dessa forma, possibilita uma maior agilidade na identificação de padrões nos dados, contribuindo para uma análise mais precisa e confiável.

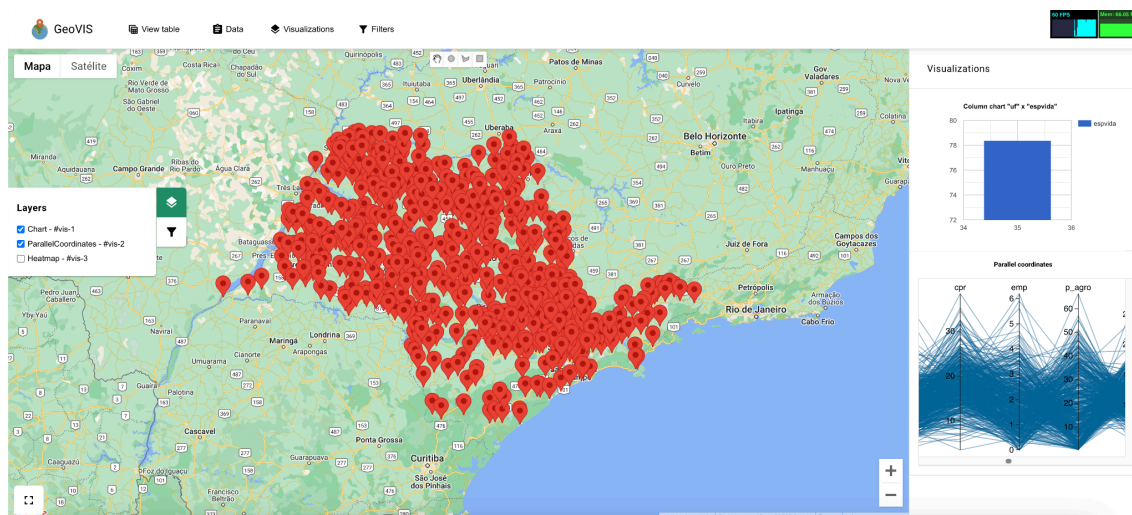


Figura 11: Estrutura das funcionalidades da ferramenta.

Como é possível visualizar na Figura 11, a ferramenta possui duas principais áreas de visualização: a visualização de mapas (esquerda) e a visualização de gráficos (direita). A customização de cada um destes dois ambientes pode ser feita pelo usuário por meio da configuração de visualizações e filtros, em que também é possível mostrar ou esconder cada um destas visualizações ou filtros por meio do menu lateral à esquerda da imagem, permitindo assim uma interatividade dinâmica com os dados e simplificando a customização do ambiente coordenado.

Outra ação presente na ferramenta é a inspeção dos dados de uma instância no mapa. Para isso o usuário pode acionar o evento de clique sob um marcador no mapa, e assim um componente gráfico de modal será aberto representando os valores dos dados pertencentes a instâncias selecionadas dentro deste componente (ver Figura 12).

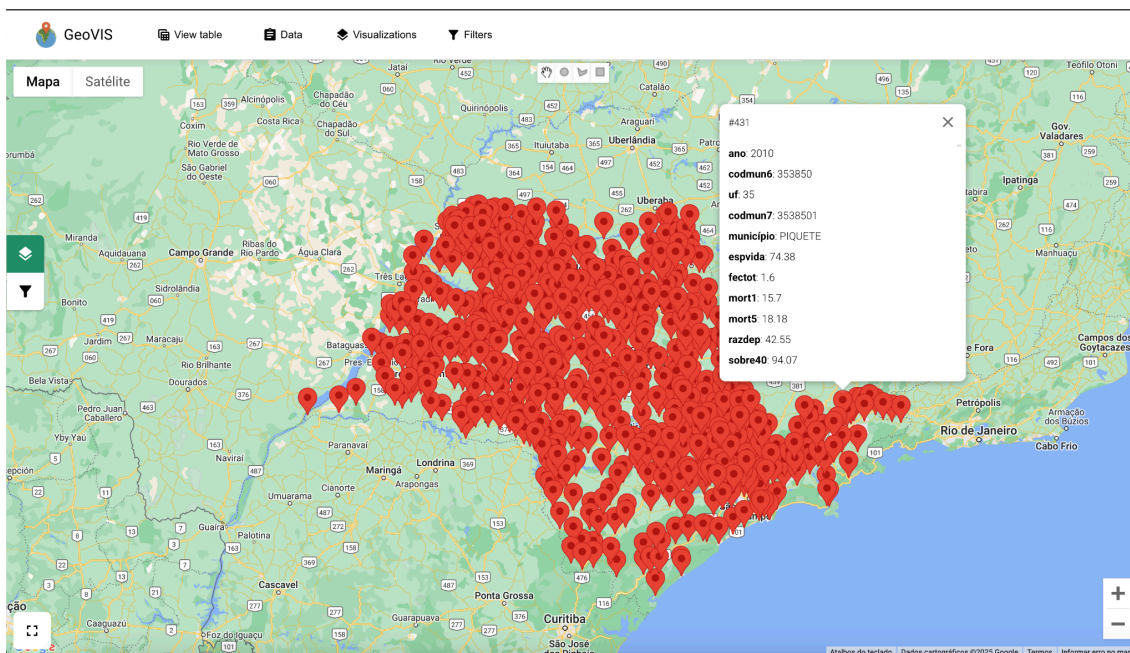


Figura 12: Componente de modal com informações da instância de dados (Autor).

A interação dinâmica entre as diferentes visualizações gráficas da ferramenta, favorece a descoberta de correlações e padrões que poderiam passar despercebidos em abordagens estáticas. Assim, a combinação dessas técnicas transforma a visualização geoespacial em um recurso poderoso para tomada de decisão em processos de descoberta de conhecimento, como áreas de planejamento urbano, monitoramento ambiental, análise de mercado, análise socioeconômica, etc.

4.1.1. Tecnologias

Para o desenvolvimento da ferramenta, foi utilizada a linguagem javascript e tecnologias para a plataforma web com suporte para mapas, tais como: biblioteca de interfaces react, a plataforma de mapas Google Maps, biblioteca de gráficos Google Charts, a biblioteca d3 para a visualização de coordenadas paralelas, a biblioteca de clusterização gráfica Google Marker Clusters e também a técnica Google Heatmap para permitir a adição de visualizações de mapa de calor no ambiente de mapas.

A Figura 13 ilustra a representação das tecnologias utilizadas para a construção das visualizações na ferramenta.

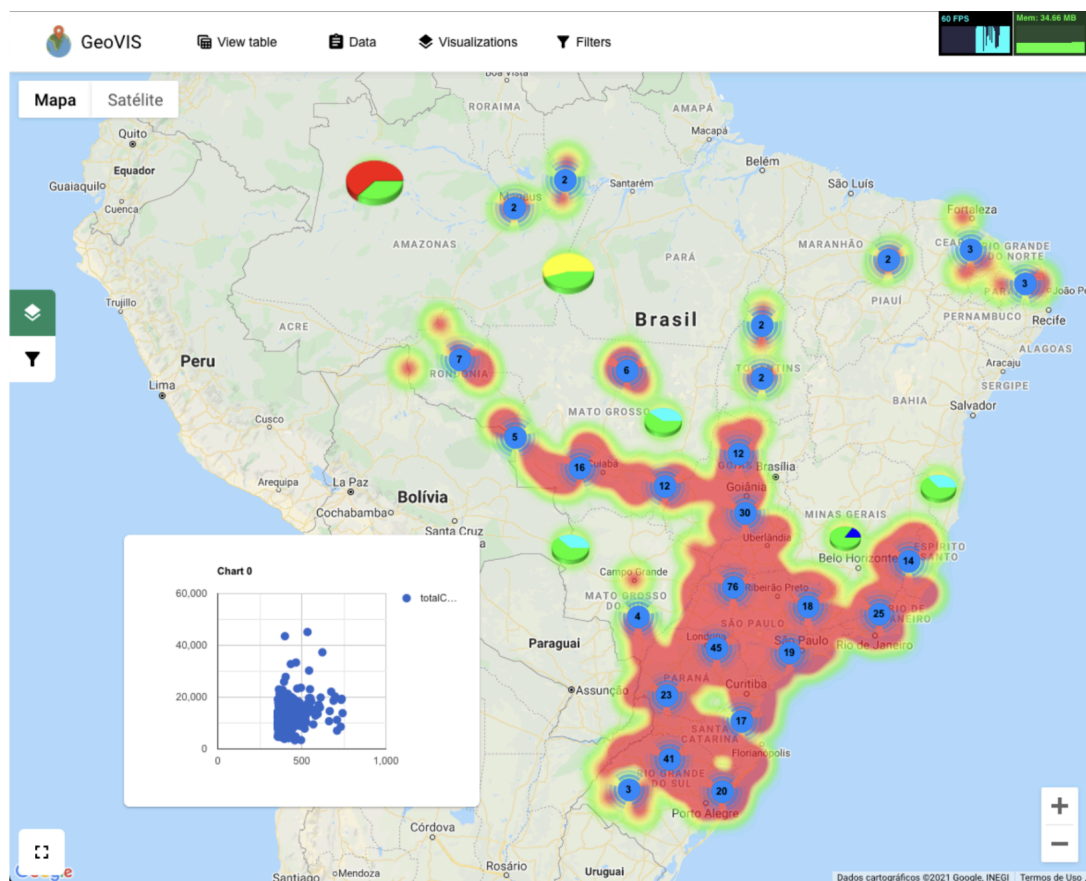


Figura 13: Interface da ferramenta GeoVis demonstrando as tecnologias utilizadas.

4.1.2. Entrada de dados

Para realizar a interação com a ferramenta, é necessário carregar um conjunto de dados. Neste processo são aceitos arquivos no formato “csv”, que poderão ter atributos em qualquer padrão de nome, mas deverão ter obrigatoriamente dois atributos para a representação georreferenciada, os campos de latitude e longitude.

Inicialmente o usuário pode selecionar um arquivo local e carregar na interface da ferramenta, conforme demonstrado na Figura 14.

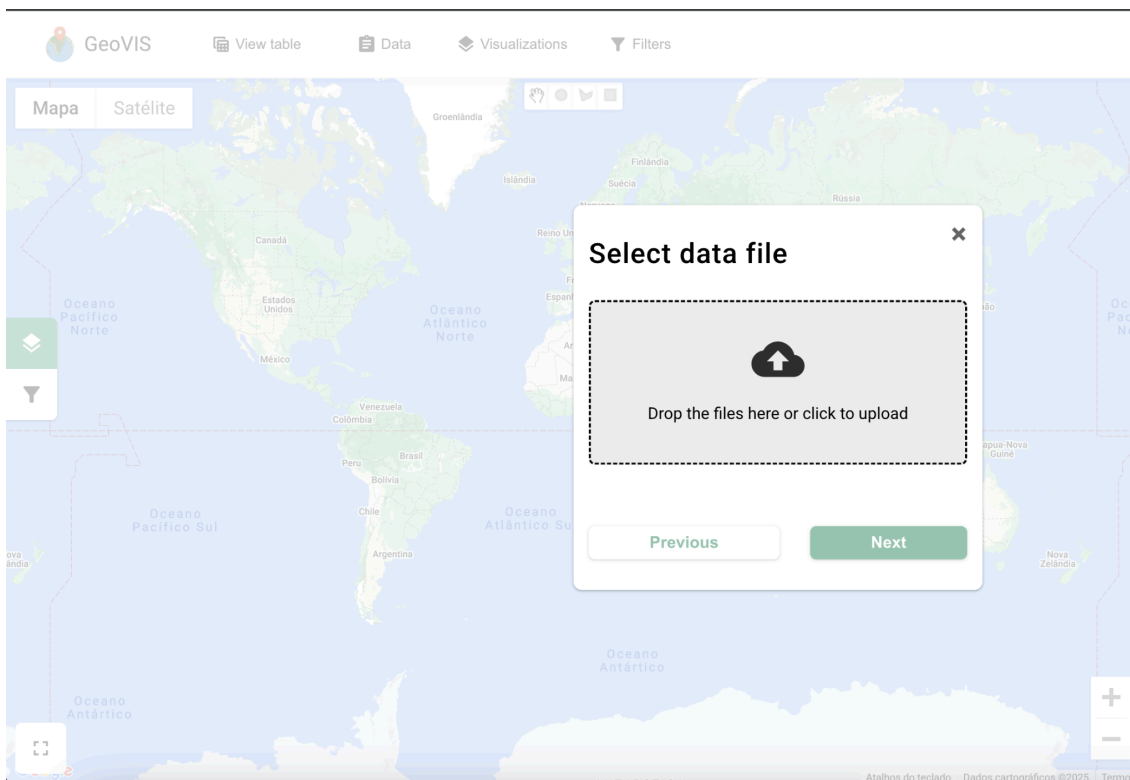


Figura 14: Etapa de carregamento de dados na plataforma GeoVis (Autor).

Após a seleção do conjunto de dados é possível escolher os atributos referentes aos dados de latitude e longitude das instâncias (Figura 15), permitindo a seleção dentre uma lista de atributos automaticamente extraídos do conjunto de dados. Logo a ferramenta faz a extração automática destes metadados durante a fase de carregamento dos dados, permitindo que a ferramenta se adapte e seja customizada a qualquer conjunto de dados.

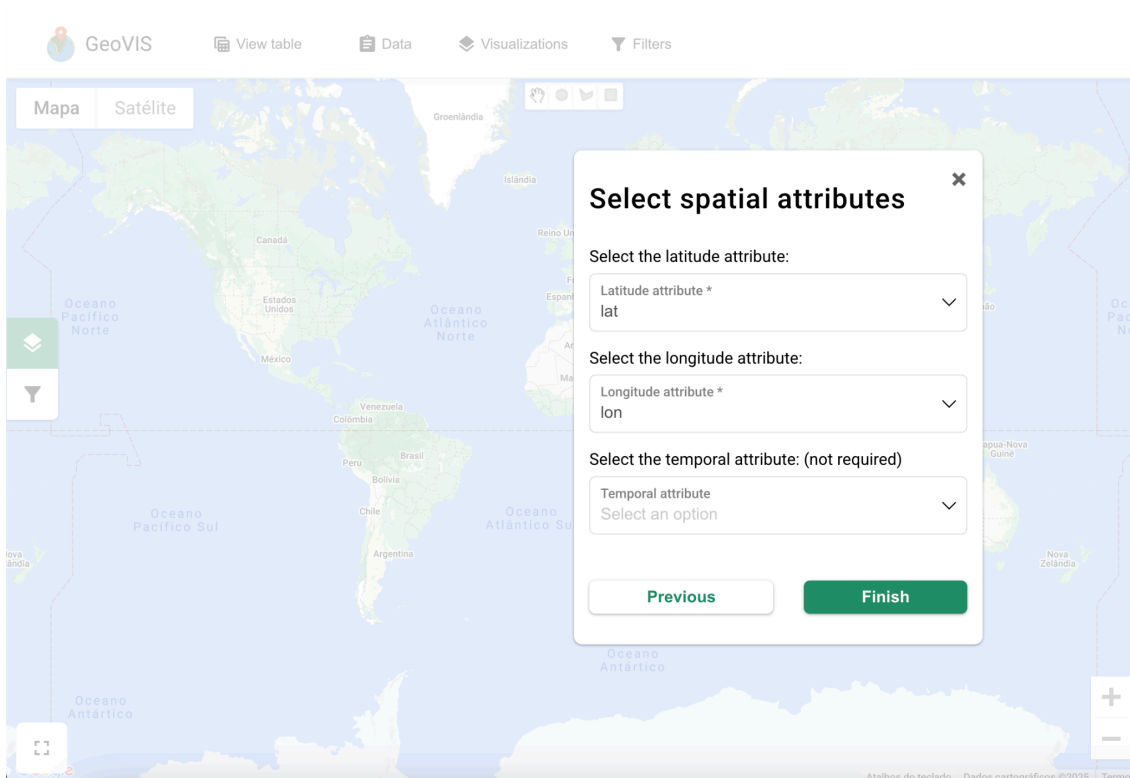


Figura 15: Seleção dos atributos de geolocalização (Autor).

4.1.3. Filtros

Foi implementado na ferramenta a funcionalidade de filtros, permitindo a seleção de instâncias de dados baseado no valor de seus atributos. A interface permite a adição e remoção de filtros genéricos, baseados em condições relacionadas a valores numéricos, booleanos ou texto dos atributos, ou até informações de latitude por meio de filtros espaciais.

Conforme apresentado nas Figuras 16 e 17, o usuário possui uma interface para selecionar o atributo desejado, as condições de comparação deste atributo e por último um campo para inserir o valores desejado para o filtro.

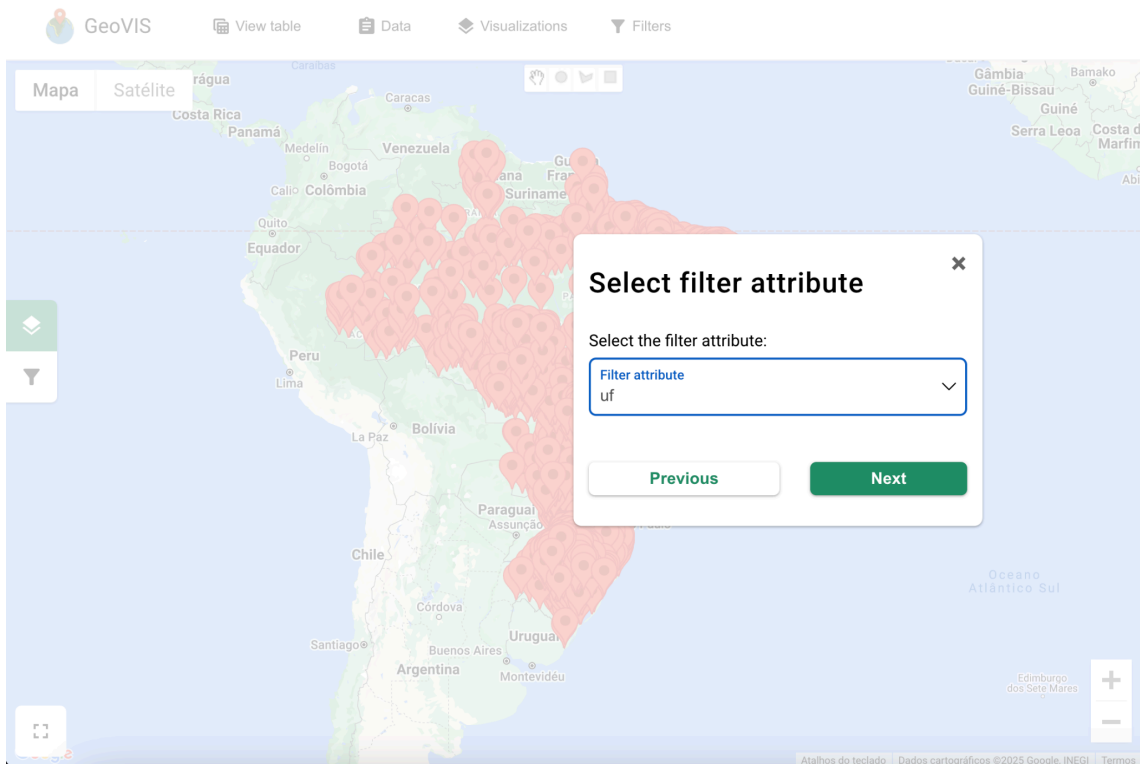


Figura 16: Escolha do campo a ser filtrado (Autor).

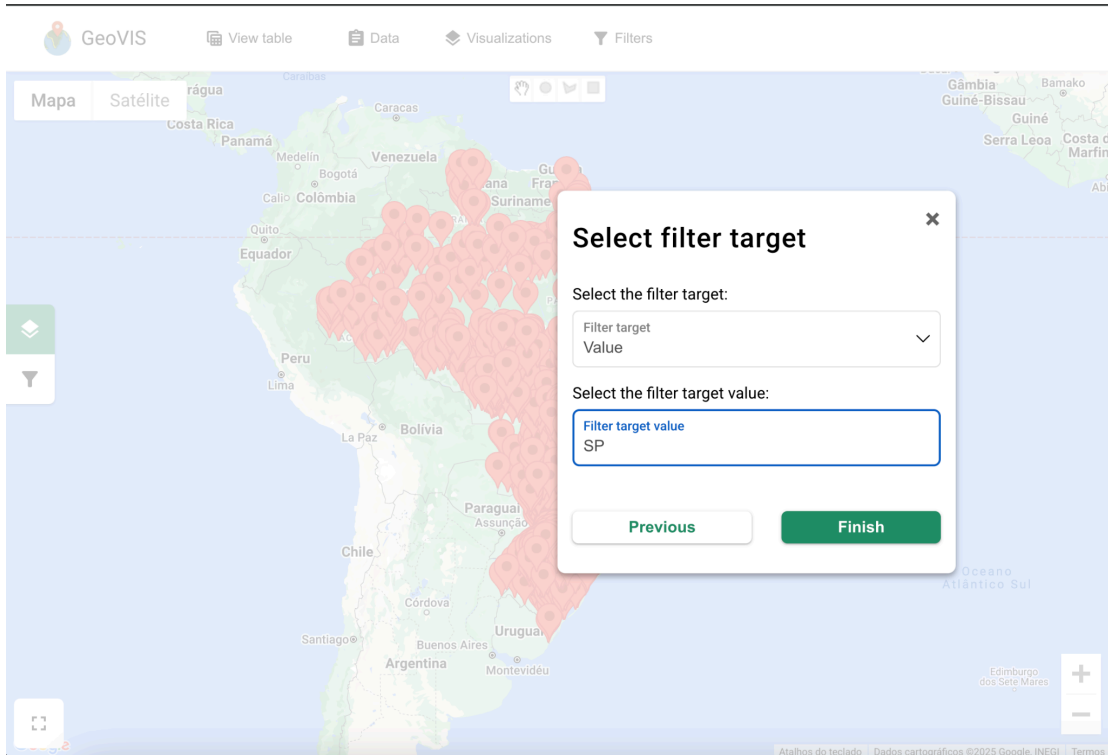


Figura 17: Interface para a inserção do valor a ser filtrado (Autor).

Uma vez que o filtro foi adicionado, a ferramenta verifica todas as instâncias do conjunto de dados e exibe somente aquelas em que satisfazem os critérios adicionados pelo filtro desejado. Como demonstrado na Figura 18, apenas as instâncias de dados do conjunto que satisfazem o critério informado foram exibidas.

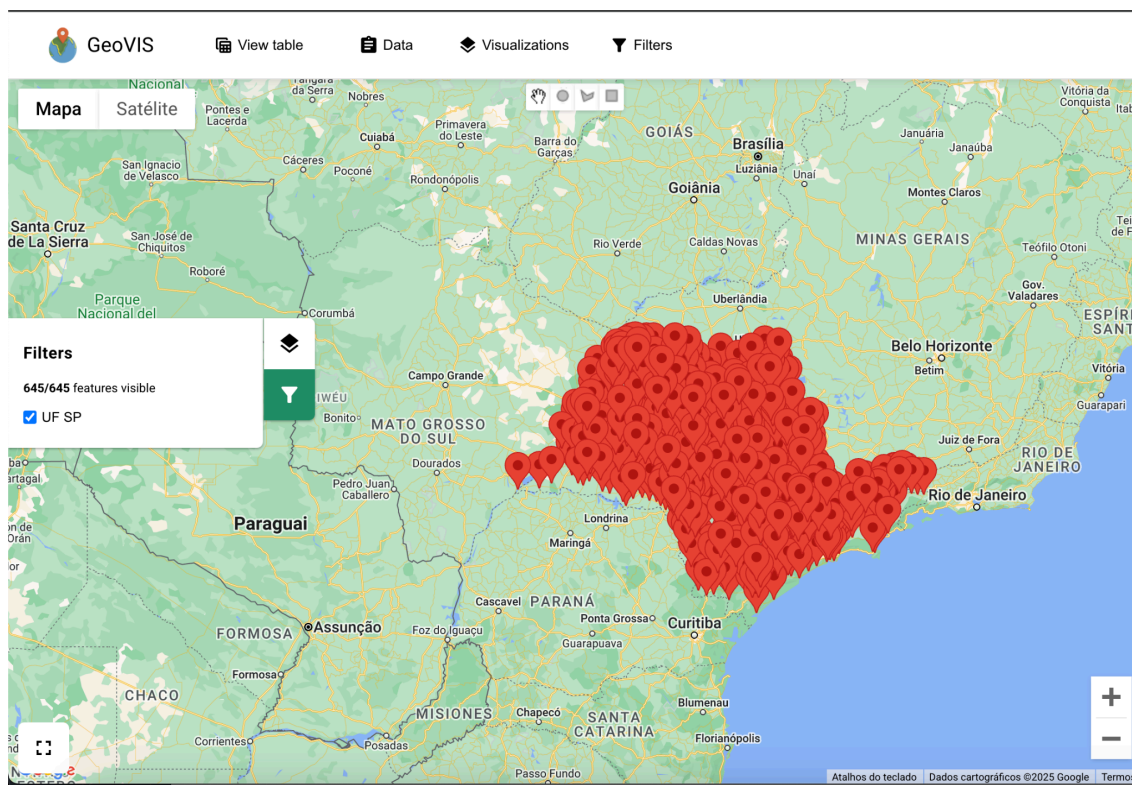


Figura 18: Demonstração de instâncias filtradas pela funcionalidade de filtros da ferramenta.

(Autor).

Outra funcionalidade de filtros que a ferramenta fornece, é o filtro geoespacial. Neste caso o usuário pode escolher dentre as opções, selecionar instâncias em um raio, em um retângulo ou até desenhar um polígono para a seleção de instâncias baseadas em uma determinada localização. Este tipo de filtro também pode ser combinado com filtros de atributos numéricos e permite assim a visualização de instâncias contidas em determinado espaço geográfico, conforme demonstrado na Figura 19.

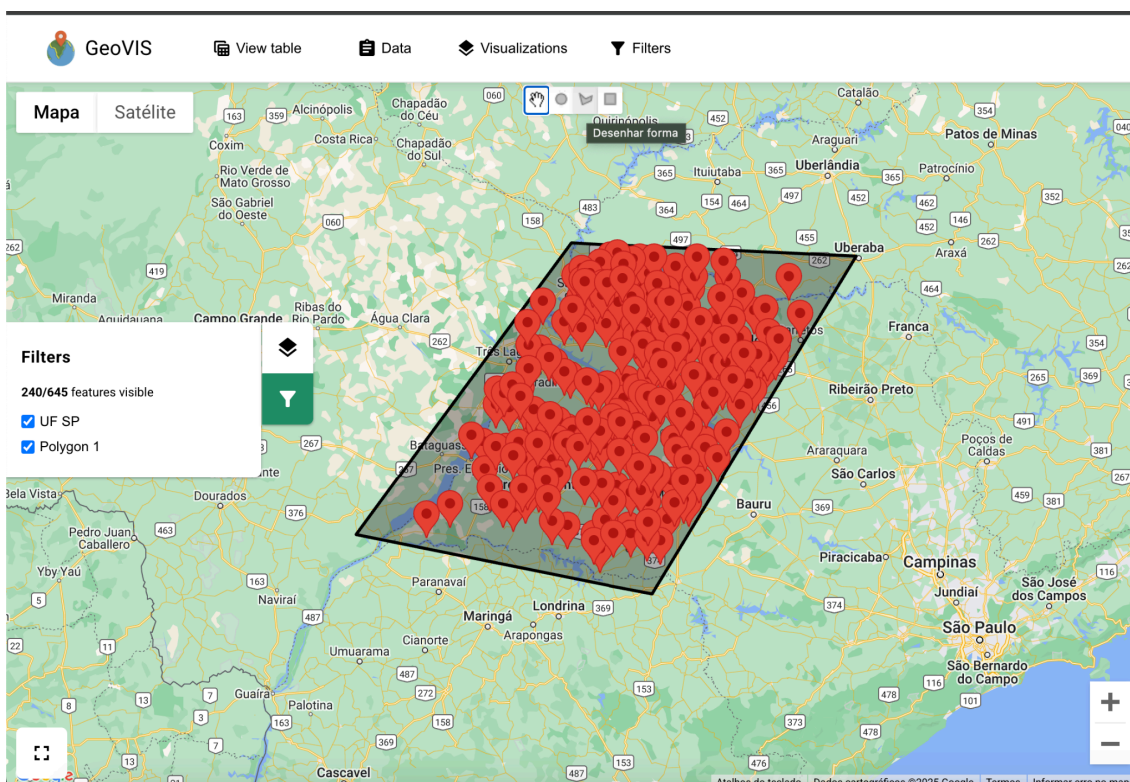


Figura 19: Demonstração de instâncias filtradas de modo geoespacial (Autor).

A ferramenta permite a adição de múltiplos filtros e a combinação entre eles, permitindo também mostrar ou esconder conforme a necessidade de exploração por meio do menu lateral exibido na Figura 19.

4.1.4. Técnicas de representações gráficas

Para permitir a exploração de grandes conjuntos de dados, foram implementadas algumas técnicas de visualizações gráficas para representar os dados. Dentre essas técnicas, foram escolhidas tanto representações que permitem a análise de dados multivalorados, quanto representações para os dados espaciais no ambiente de mapas. A seguir as técnicas implementadas serão descritas individualmente.

4.1.4.1. Mapas de calor

A técnica mapas de calor foi implementada na ferramenta para permitir a identificação de “zonas quentes” em relação a densidade de instâncias representadas em determinada localidade ou proximidade. Na Figura 20 podemos visualizar que a técnica

representou por meio de cores quentes, o agrupamento de instâncias com determinada proximidade.

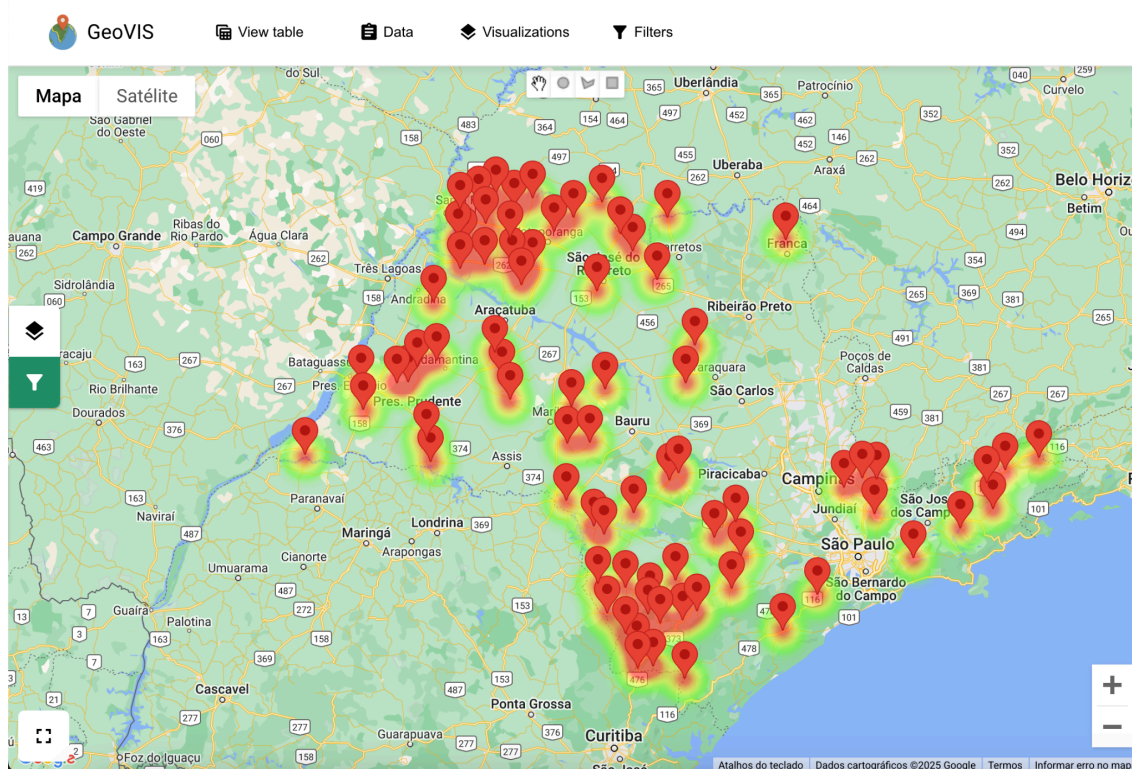


Figura 20: Técnica mapas de calor (Autor).

4.1.4.2. Gráficos em marcadores

Conforme apresentado na seção Trabalhos Relacionados deste estudo, em alguns estudos os autores adicionaram gráficos para representar informações numéricas em marcadores do mapa com o objetivo de simplificar o processo exploratório deixando evidente padrões em informações de dados multivalorados em determinadas regiões do mapa. A ferramenta desenvolvida neste trabalho também implementa técnicas de representações gráficas nas instâncias dos marcadores, como gráfico de linhas, radar, coluna, setores, polar e de radar. A Figura 21 demonstra um exemplo desta técnica.

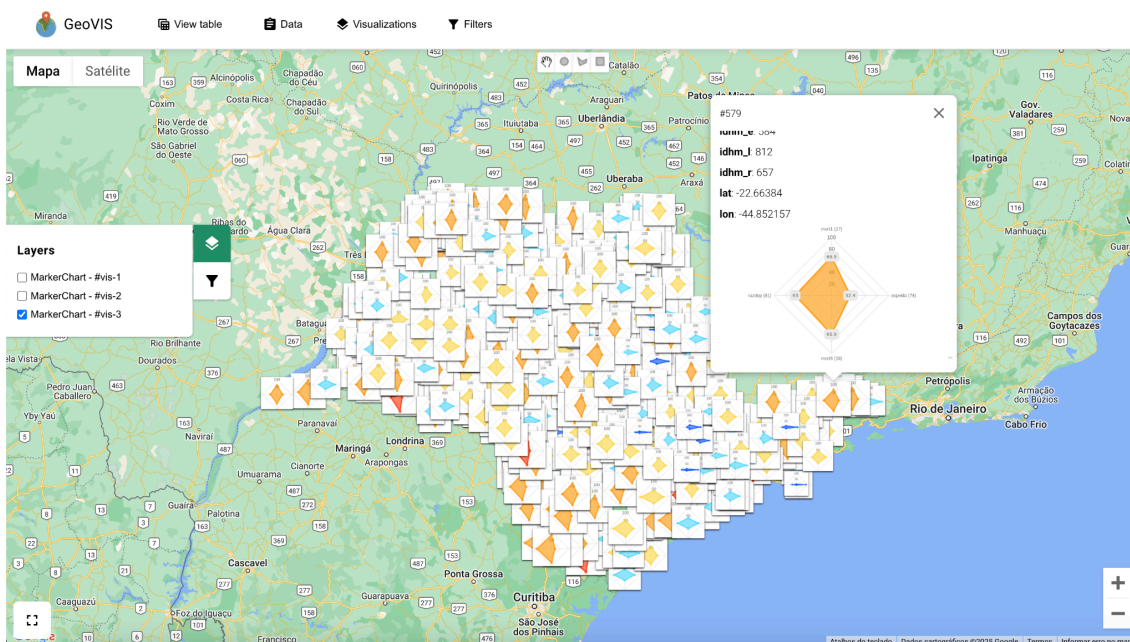


Figura 21: Técnica de gráfico em marcadores com o gráfico polar (Autor).

Na interface da ferramenta, é possível selecionar os atributos que serão adicionados no gráfico. Assim a representação do marcador no mapa, se transforma em um gráfico com a representação dos atributos de sua instância. As cores escolhidas para cada um destes gráficos, é baseado nos valores dos atributos escolhidos, onde os atributos de maior intensidade em relação aos valores de máximo e mínimo do conjunto possuem uma cor mais quente (vermelho), já os valores de menor intensidade possuem uma cor mais fria (azul).

Por exemplo na Figura 22, foi adicionado um gráfico de setor nos marcadores, onde o atributo representado foi a expectativa de vida do conjunto de dados do Censo disponível pelo portal de dados abertos do governo federal. Logo é possível identificar na imagem quais as instâncias com valores de maior intensidade do conjunto de dados pelas cores vermelhas.

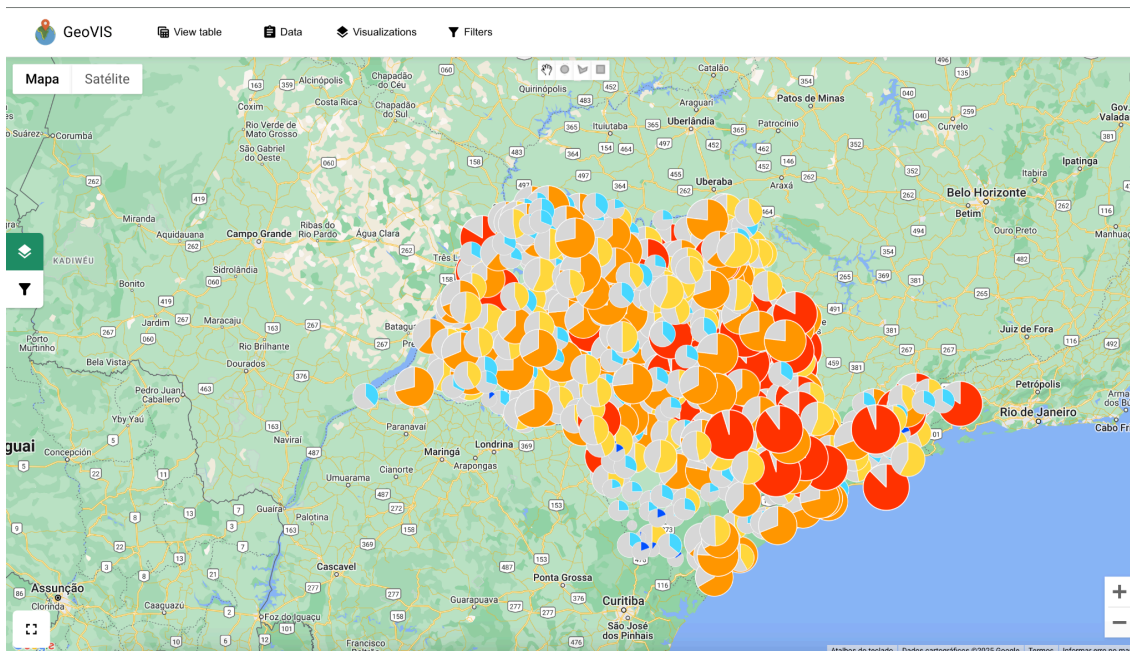


Figura 22: Técnica de gráfico em marcadores com o gráfico de setores (Autor).

Analisando a visualização gerada, é possível encontrar áreas geográficas onde a expectativa de vida é maior e outras áreas com expectativa de vida menores, permitindo assim uma análise exploratória baseada em ganhos na cognição a partir da visualização.

Já na Figura 23, é possível visualizar o gráfico de colunas relacionando dois atributos, o atributo de índice de desenvolvimento humano municipal (idhm) e a taxa de escolaridade. Como representado na imagem, é possível também ver detalhes do gráfico por meio do evento de clique na instância selecionada, possibilitando o entendimento dos valores numéricos do gráfico e um estudo aprofundado de determinados dados.

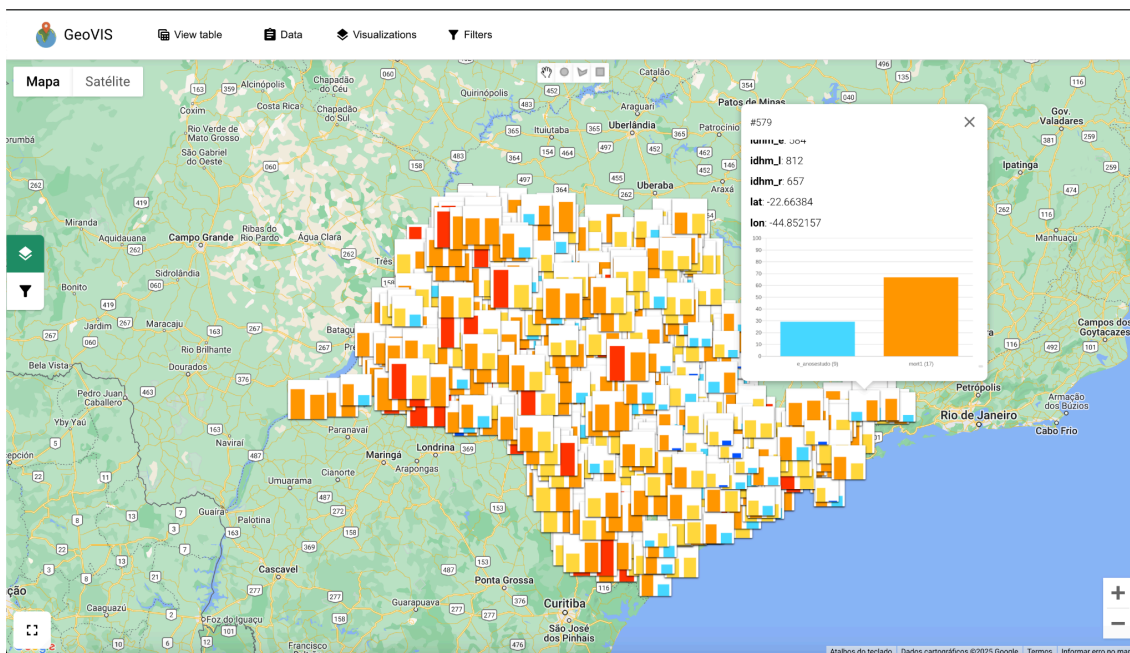


Figura 23: Técnica de gráfico em marcadores com o gráfico de colunas (Autor).

4.1.4.3. Clusterização de marcadores

Outra técnica comumente usada para lidar com problemas de alta dimensionalidade de dados, é a técnica de clusterização de marcadores. Na ferramenta GeoVis, foi implementado esta técnica dentro do ambiente de mapas para agrupar os marcadores baseado em proximidade para reduzir a quantidade de instâncias renderizadas em uma curta localidade (Figura 24). Nesta técnica os marcadores são agrupados com base em sua proximidade e são representados por um círculo visual, no qual possuem um contador renderizado dentro do cluster. Além do contador, a técnica de Clusterização de marcadores implementa a também transição entre cores, utilizando cores quentes como o rosa e vermelho para indicar clusters com maior quantidades de instâncias agrupadas, assim como as cores no tom de azul (mais frias) para indicar a menor incidência de marcadores próximos dentro de um cluster.

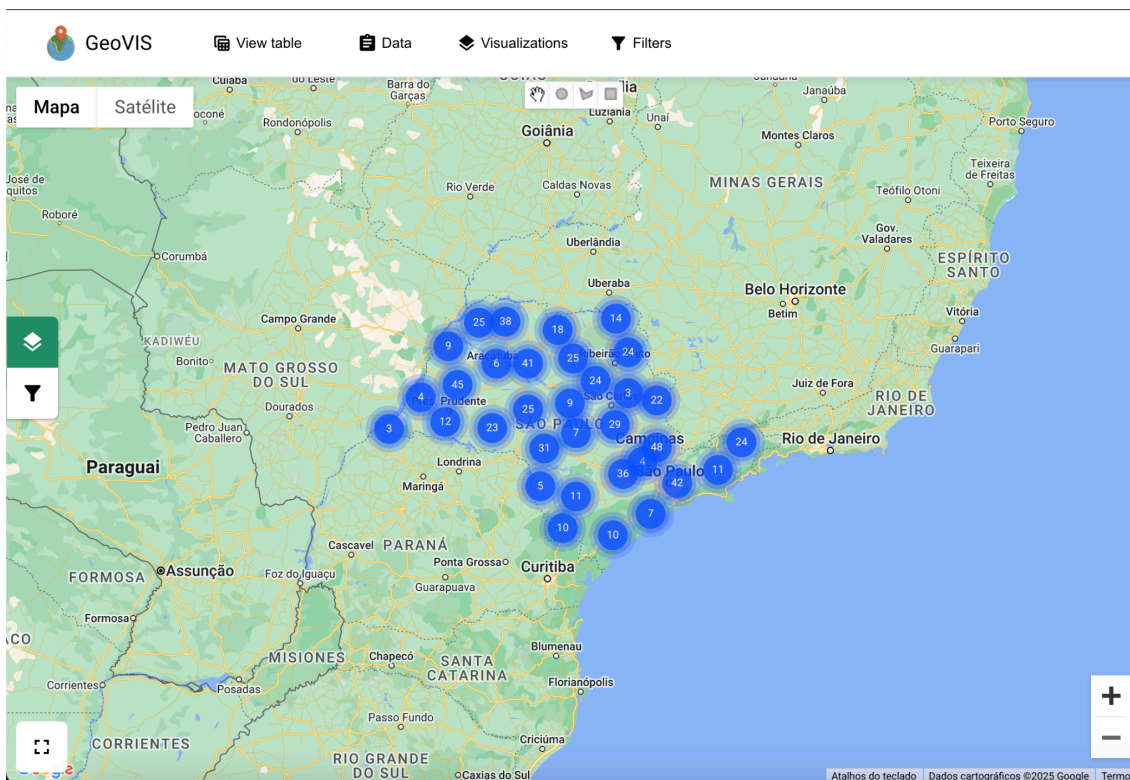


Figura 24: Técnica de clusterização de marcadores (Autor).

A ferramenta desenvolvida ao longo deste trabalho também permite a combinação de técnicas de representações visuais com a técnica de clusterização de marcadores. Por exemplo, é possível utilizar a técnica Gráficos em marcadores combinada com a técnica de clusterização, neste caso ao renderizar um cluster, a ferramenta calcula os valores normalizados de cada um dos atributos das instâncias pertencentes ao conjunto do cluster, e com isso, utiliza a média ou mediana dos valores dos atributos para representar o novo gráfico dentro do marcador. Assim a visualização combinada entre as duas técnicas permite uma ágil visualização sobre a representação de um grupo de instâncias agrupadas de acordo com os atributos em análise adicionados dentro da visualização gráfica. Por exemplo, na Figura 25 é possível visualizar a técnica de clusterização de marcadores junto a técnica de visualização gráfica, permitindo uma análise de representações gráficas baseada no agrupamento geográfico de instâncias próximas. Já a Figura 26, traz a mesma representação, mas com a variante de não mostrar os gráficos no agrupamento de clusters.

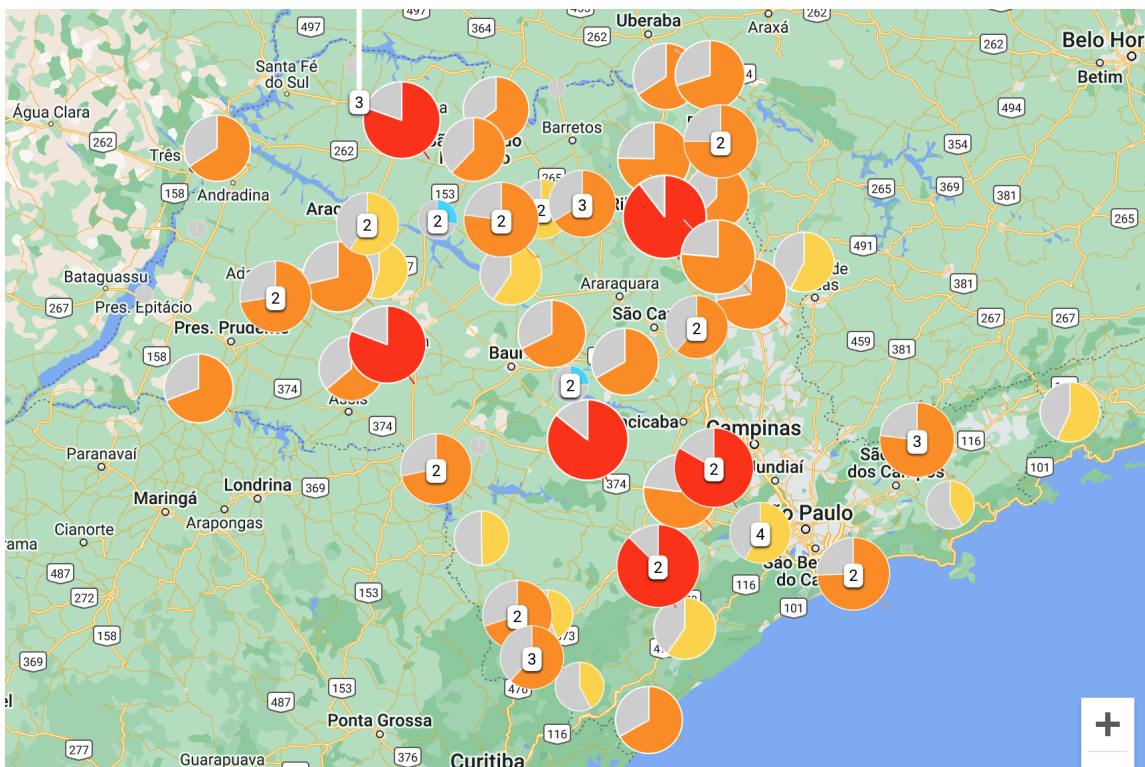


Figura 25: Técnica de clusterização de marcadores com gráfico de setores nos clusters (Autor).

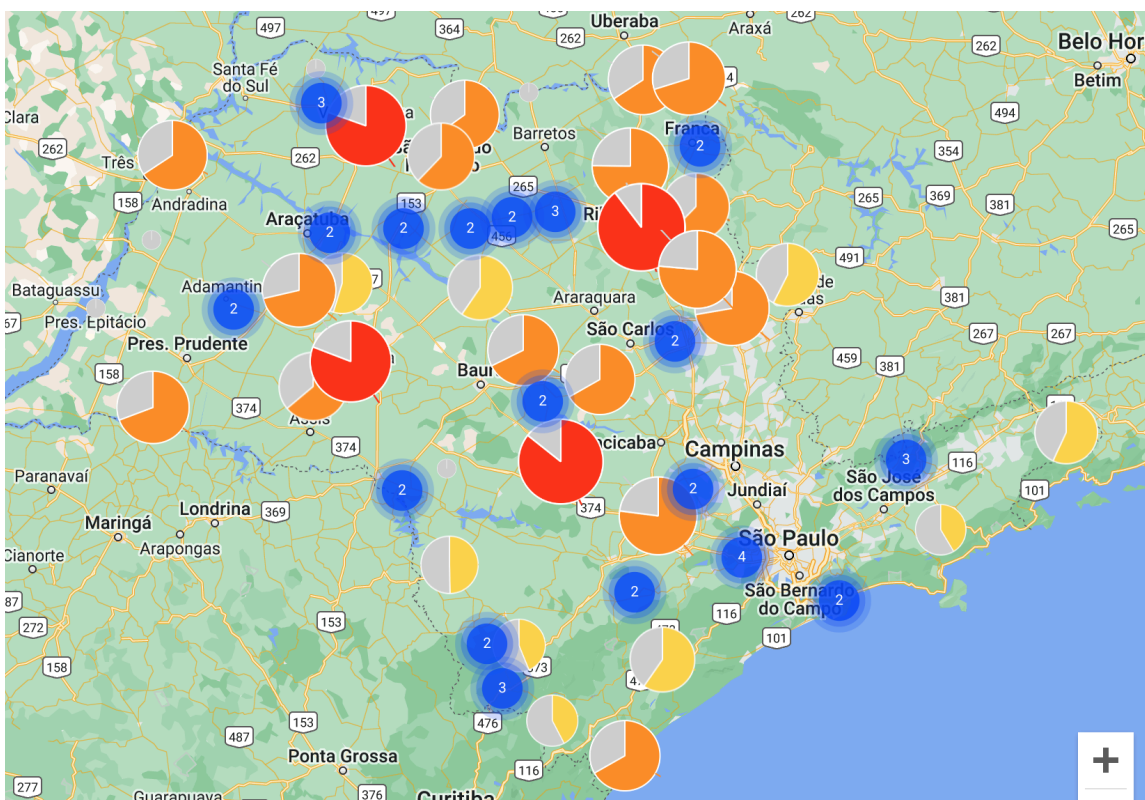


Figura 26: Técnica de clusterização de marcadores e gráfico de setores (Autor).

4.1.4.4. Cores em marcadores

A técnica de cores em marcadores de mapa (mapeamento cromático), é amplamente utilizada para transmitir informações de forma visual baseada em cores. A ferramenta permite utilizar esta técnica para diferenciar instâncias de marcadores baseadas em seus atributos. Esta técnica permite simplificar a visualização entre diferentes categorias dos dados, possibilitando destacar características específicas, como densidade populacional, altitudes, temperaturas, ou outros dados relevantes.

Como demonstrado na Figura 27, foi adicionado a técnica baseada no atributo UF (unidade federativa) e assim possível visualizar com base na cor as instâncias com determinado valor de atributo.

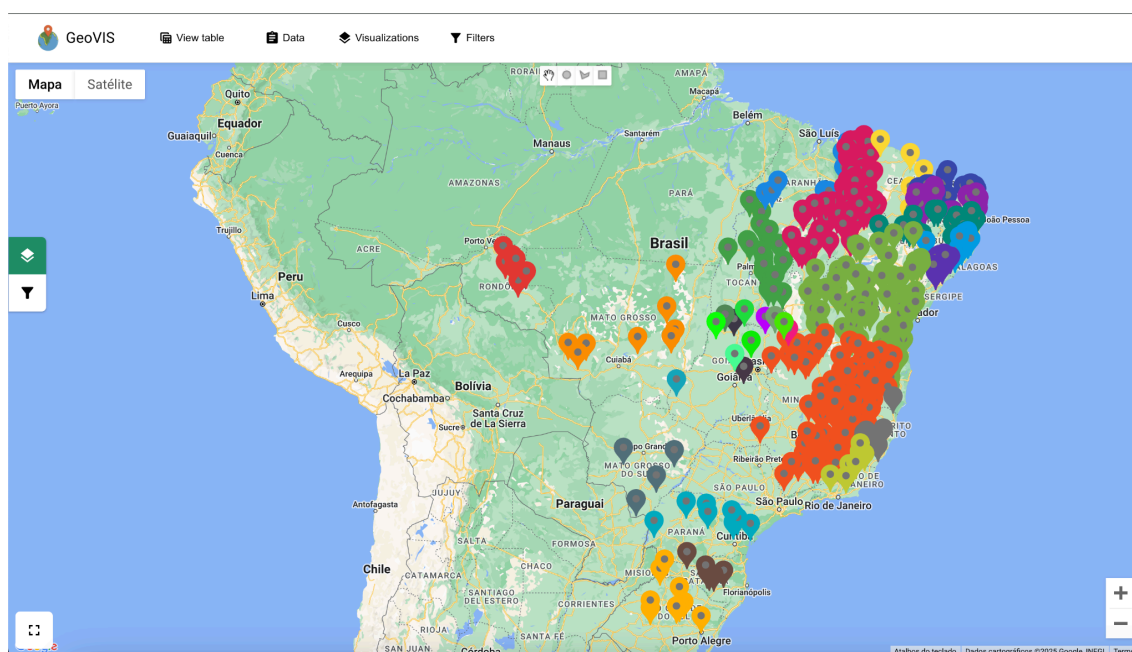


Figura 27: Técnica de cores em marcadores do mapa (Autor).

4.1.4.5. Coordenadas paralelas

A técnica de coordenadas paralelas é um método utilizado para representar e analisar dados multidimensionais de forma visual. Ao analisar conjuntos de dados com uma alta dimensionalidade de atributos, em vez de usar um sistema cartesiano tradicional, onde cada variável é representada em um eixo perpendicular, às coordenadas paralelas dispõem múltiplos eixos paralelos, cada um correspondendo a

uma variável e permitindo uma análise conjunta de todos os atributos do conjunto. A ferramenta GeoVis Explorer permite implementar esta técnica, e realizar observações de modo interativo sobre linhas que cruzam os eixos dos atributos, permitindo identificar padrões nos dados e até filtrar instâncias do mapa com base nos atributos destacados na visualização de Coordenadas Paralelas. (Figura 28)

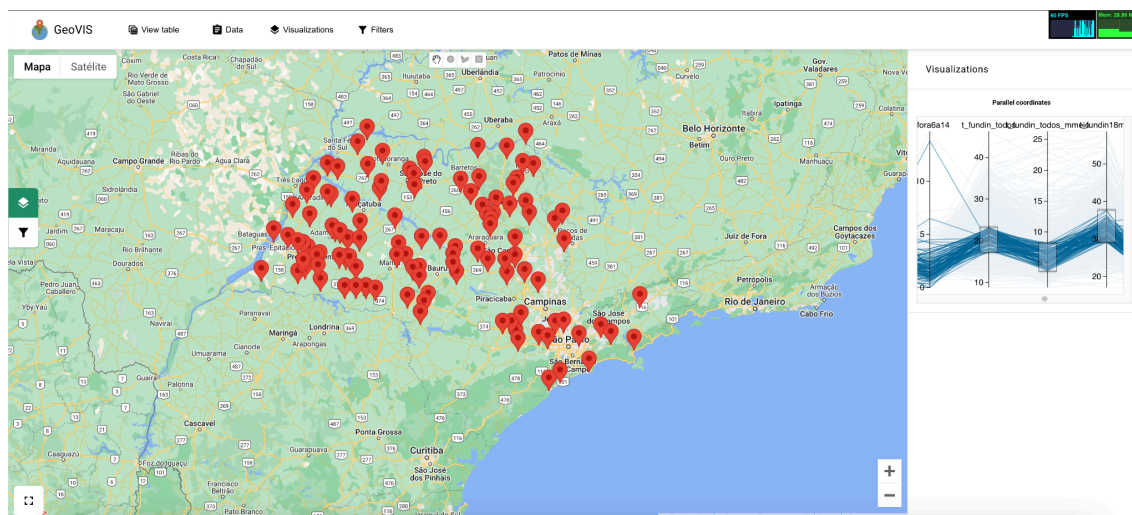


Figura 28: Técnica de coordenadas paralelas aplicando filtros no mapa (Autor).

4.1.4.6. Representações gráficas

A técnica de Representações Gráficas é uma poderosa ferramenta visual utilizada para representar dados e simplificar a análise de grandes quantidades de informações. A ferramenta implementou os gráficos de diversas formas, como barras, linhas, setores área, histograma ou dispersão, assim o usuário também pode customizar suas representações utilizando este recurso e interagindo com estes gráficos para exercerem filtros nos atributos representados na visualização. Logo, os gráficos podem coordenar as instâncias de dados do mapa, exibindo ou escondendo as instâncias que satisfazem o critério selecionado a partir da interação de clique do usuário em determinado setor do gráfico.

Por exemplo, a Figura 29 demonstra gráficos adicionados na ferramenta a partir de uma seleção prévia de atributos pelo usuário com os dados visíveis dentro da visualização do mapa.

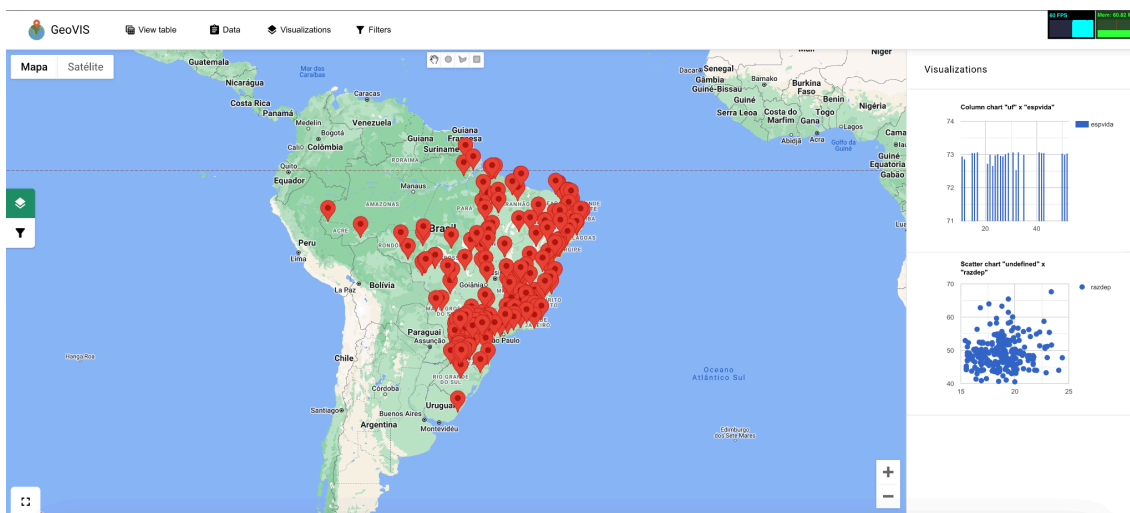


Figura 29: Gráficos com dados das instâncias visíveis do mapa (Autor).

4.1.5. Exploração interativa

Como descrito nos Capítulos 2 e 3, a exploração interativa de dados permite a análise dinâmica e detalhada de subgrupos de um conjunto de dados. Neste caso a interação do usuário com os dados permite que análises exploratórias customizadas e mais específicas sejam feitas nos dados, sejam elas pela seleção de subgrupos de informações visíveis dentro de uma visualização, como também a aplicação de filtros entre técnicas de visualizações que coordenam novas formas de representação entre si.

As representações gráficas combinadas com interatividade permitem que os usuários naveguem pelos dados de interesse de forma intuitiva, ajustando visualizações conforme a necessidade da análise desejada. No contexto geoespacial, a inclusão de filtros espaciais e de dados numéricos aprimora a capacidade exploratória da ferramenta, permitindo que os usuários refinem as visualizações com base em intervalos de valores ou regiões de interesse.

Foi implementado na ferramenta GeoVis Explorar estas funcionalidades dentro de suas visualizações suportadas, assim o usuário tem a capacidade de interagir com a ferramenta por meio de ações de clique, zoom, seleção de informações visíveis para potencializar sua capacidade exploratória de dados. Por meio deste modelo de interação, é possível coordenar as visualizações para representar dados de acordo com os critérios

resultantes da interação e assim possibilitar o destaque de padrões de concentração e variação espacial, simplificando a identificação de conhecimento sobre os dados.

Um exemplo pode ser visto na Figura 30, foi gerada uma visualização a partir de dados de interesse e adicionado um gráfico de colunas relacionando os presentes nos marcadores do junto. Logo em seguida (Figura 31) foi selecionado uma coluna do gráfico por meio do evento de clique, em seguida o mapa atualizou as instâncias visíveis no mapa de acordo com o filtro interativo aplicado pelo usuário por meio do clique, disponibilizando assim uma exploração coordenada entre as técnicas neste ambiente georeferenciado.

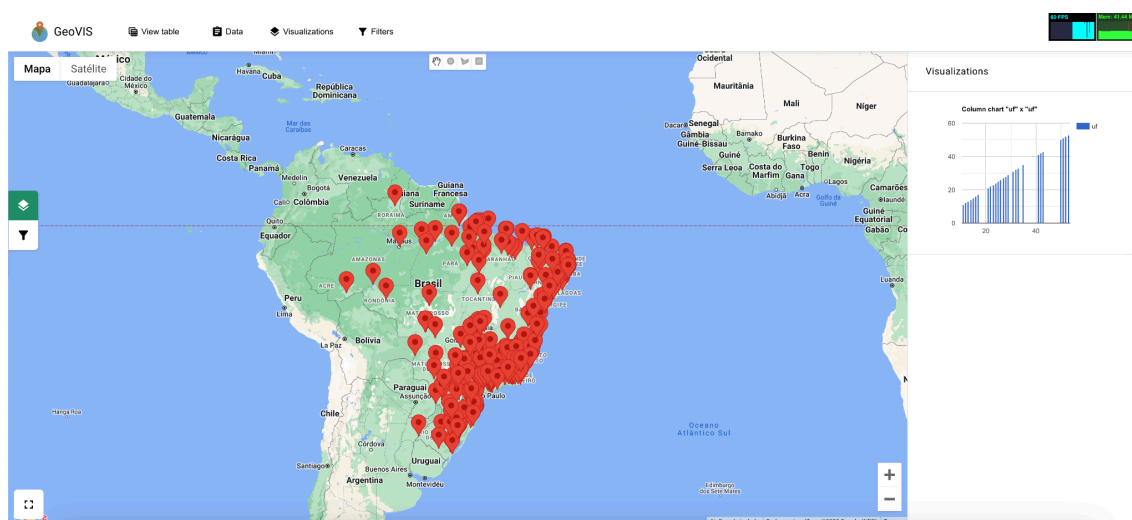


Figura 30: Mapa e Gráfico de barras categorizado por estados (Autor).

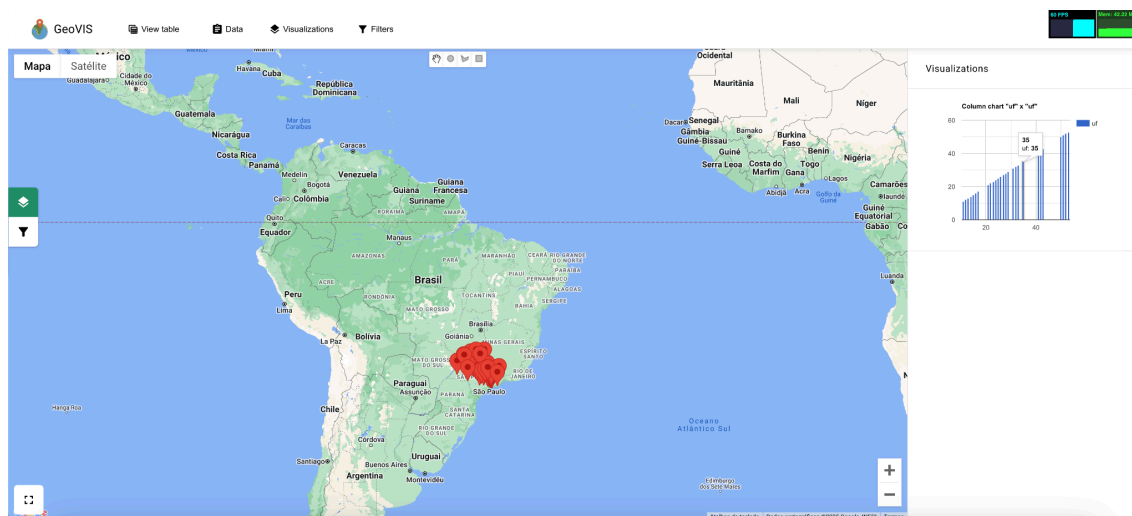


Figura 31: Instâncias filtradas pelo filtro interativo do gráfico (Autor).

Outro exemplo pode ser visto na Figura 32, em que foi adicionada a técnica de mapas de calor e a técnica de visualização coordenadas paralelas para se analisar os dados do Censo do portal de dados aberto do governo. Em seguida um filtro interativo na técnica coordenadas paralelas foi aplicado, selecionando interativamente os valores abaixo de 0.5 para o atributo “índice de escolaridade”. Assim foi possível analisar por meio da combinação de técnicas, os pontos de maior concentração das cidades com menor índice de escolaridade no conjunto de dados, possibilitando um entendimento simplificado por meio da análise interativa.

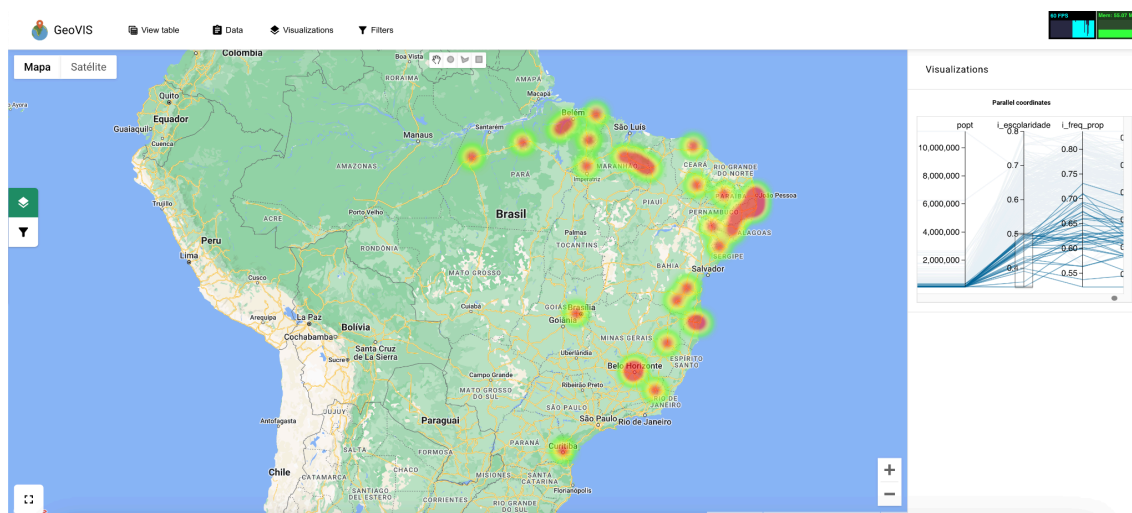


Figura 32: Instâncias filtradas pelo filtro interativo das coordenadas paralelas (Autor).

Assim diversas técnicas podem ser combinadas entre si para potencializar ainda mais o processo exploratório (Figura 33), como descrito no início deste capítulo, uma das grandes limitações presentes em ferramentas existentes é a falta de capacidade de customização do processo exploratório e de técnicas de visualizações que estejam aptas a serem utilizadas em diversos domínios, principalmente para cidades inteligentes. Logo as possibilidades de utilização desta ferramenta e do modelo coordenado de visualização são diversos, possibilitando estudos cada vez mais específicos e direcionados a informações de estudo.

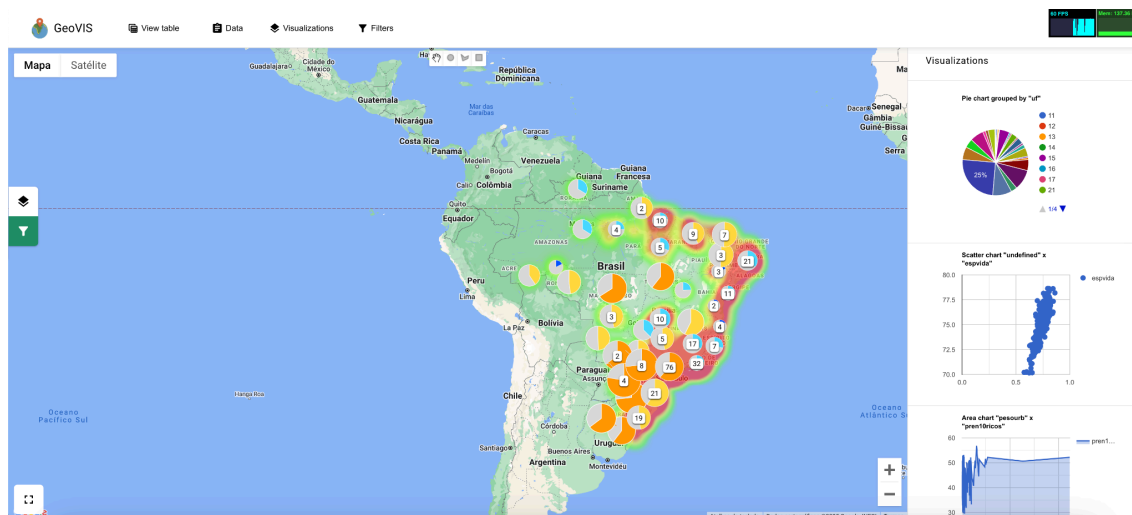


Figura 33: Diferentes técnicas de visualização coordenadas na ferramenta (Autor).

4.2. Etapas para a avaliação do modelo

A realização deste trabalho foi dividida em etapas com atividades específicas para a execução do projeto de pesquisa.

A primeira etapa foi na realização de pesquisas para encontrar e avaliar conjuntos de dados de domínio aberto, como os conjuntos disponibilizados nas plataformas de dados públicos, como por exemplo os dados disponibilizados no portal "Dados Abertos" do governo (<https://dados.gov.br/home>). Foram avaliados a qualidade dos dados e a possibilidade da aplicação de técnicas de modelagem e normalização (Min-Max) de dados, com o objetivo de realizar-se testes com cenários reais sobre possíveis conclusões resultantes da análise de dados. Nesta etapa também foram relacionadas as lacunas encontradas nas ferramentas existentes para se encontrar oportunidades na construção de uma nova ferramenta.

Na segunda etapa de construção da ferramenta GeoVis Explorer, foi desenvolvido uma interface para o carregamento dos conjuntos de dados de estudo, em seguida implementadas as visualizações gráficas, como mapas na plataforma web, mapas de calor, gráficos de setores, de barras, de linha, coordenadas paralelas e histograma. Em seguida foram implementados os filtros geoespaciais para a seleção de marcadores no ambiente de mapas, e também os filtros para atributos multivalorados baseado em valores dos atributos.

Na terceira etapa foi construída a coordenação de interações entre visualizações (*CMVs*), que possibilitou a interação do usuário com as visualizações e a coordenação entre as visualizações a partir da interação do usuário com as representações visuais que interagem com os dados.

5. RESULTADOS DA ANÁLISE VISUAL

Neste capítulo, serão apresentados os resultados da análise visual realizada na ferramenta GeoVis Explorer em relação à análise de informações de um conjunto de dados utilizado no estudo. Inicialmente serão descritas as evidências extraídas a partir da análise das representações geradas sobre dados, assim como os resultados obtidos pela percepção dos usuários por meio de entrevistas semiestruturadas.

O capítulo aborda as visualizações geradas a partir dos conjuntos de dados integrados, enfatizando como os diferentes tipos de representações visuais e a aplicação de filtros interativos permitiram a identificação de padrões e correlações relevantes para a compreensão espacial do conjunto de dados analisados. As representações gráficas, mapas georreferenciados, métodos de exploração espacial e a interatividade entre as visualizações foram fundamentais para revelar correlações em análises de atributos numéricos, evidenciando a importância destas técnicas na exploração e interpretação dos dados. Adicionalmente, a seção inclui uma avaliação qualitativa da ferramenta baseada nas entrevistas realizadas com usuários da área de Ciência de Dados e da Informação. As percepções dos usuários entrevistados forneceram confirmações valiosas sobre a usabilidade, a análise de informações de alta dimensionalidade e o potencial da ferramenta para apoiar a tomada de decisão em contextos geoespaciais.

A seguir, serão detalhados os resultados iniciais obtidos a partir deste modelo de análise e as principais percepções dos entrevistados, ressaltando tanto os pontos positivos quanto as oportunidades de aprimoramento identificadas durante o processo de avaliação. Dessa forma, este capítulo não apenas apresenta os resultados técnicos obtidos a partir da análise visual dos dados, mas também integra as perspectivas dos usuários, oferecendo uma visão abrangente sobre a aplicabilidade prática da ferramenta no ambiente geoespacial para se analisar informações de domínios variados.

5.1. Resultados

A ferramenta GeoVis Explorer foi validada com diferentes conjuntos de dados geoespaciais para avaliar sua eficácia na exploração interativa e na extração de conhecimento sobre os dados por meio das visualizações geradas. Os testes iniciais

foram conduzidos com dados públicos disponíveis no portal de dados aberto do governo federal, com conjuntos de dados contendo informações socioeconômicas, por exemplo, dados do Censo Demográfico e dados de tráfego urbano. O objetivo foi de validar a flexibilidade da ferramenta em suportar dados genéricos, e também a aplicabilidade da ferramenta na análise de informações em diferentes cenários de mineração de dados. Os resultados obtidos demonstraram ganhos significativos na descoberta de padrões e relações espaciais que poderiam passar despercebidos em abordagens convencionais.

5.1.1. Análise exploratória

Para exemplificar estes resultados, foi utilizado o conjunto de dados com informações do Censo Brasileiro disponível no portal de dados abertos (<https://dados.gov.br/>) contendo informações socioeconômicas dos municípios brasileiros e assim utilizada a ferramenta GeoVis Explorar para a realização das análises. Inicialmente os dados foram carregados na ferramenta e foi aplicada a técnica de filtro para selecionar quais os municípios que possuem o atributo “índice de escolaridade” (i_escolaridade) abaixo da média nacional, em seguida foi adicionado um mapa de calor para representar concentrações geográficas destes dados (Figura 34).

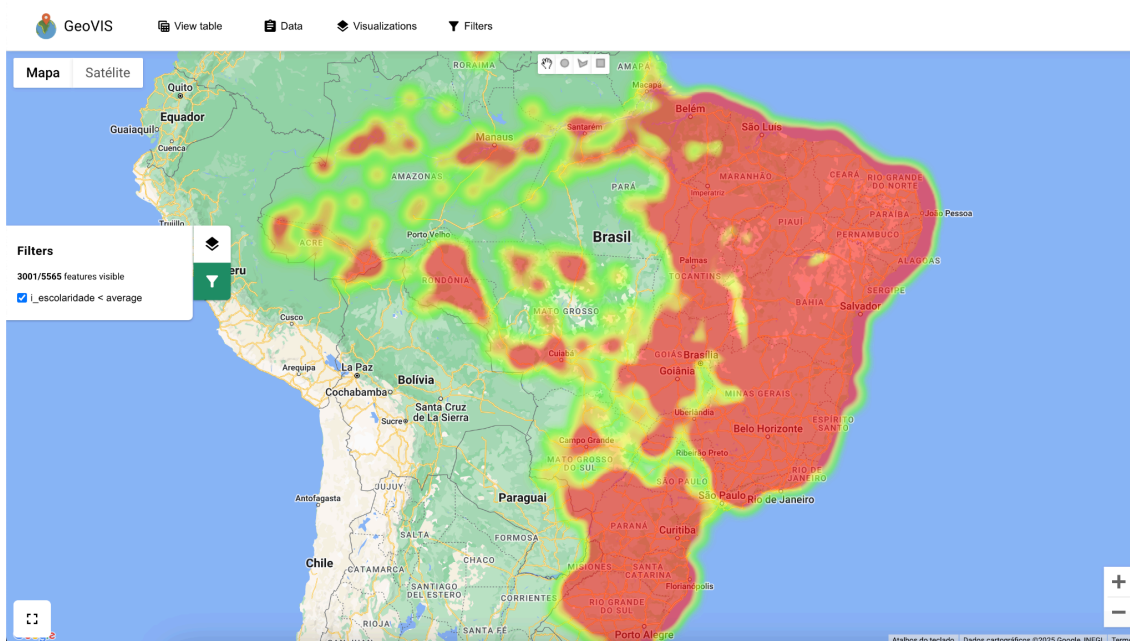


Figura 34: Representação dos municípios com índice de escolaridade abaixo da média (Autor).

Para se analisar as proporções relacionadas a unidades federativas, foi adicionado um gráfico de setores para representar a quantidade de instâncias em cada um dos estados, destacando estados com maior volume de municípios com este índice socioeconômico de escolaridade abaixo da média. Assim, ao clicar no setor correspondente ao estado de Minas Gerais (cor roxa) do gráfico de setores (evento de clique), a ferramenta realiza o filtro dos marcadores no mapa e ajusta a visualização para focar nas instâncias de dados resultantes desta ação. Como demonstrado na Figura 35, é possível visualizar que o estado de Minas Gerais possui o menor índice de escolaridade (18% do conjunto, como pode ser visto no gráfico de setores à direita da imagem), por meio da análise exploratória dos dados, onde também se tornam evidentes as microrregiões com a maior concentração desses municípios.

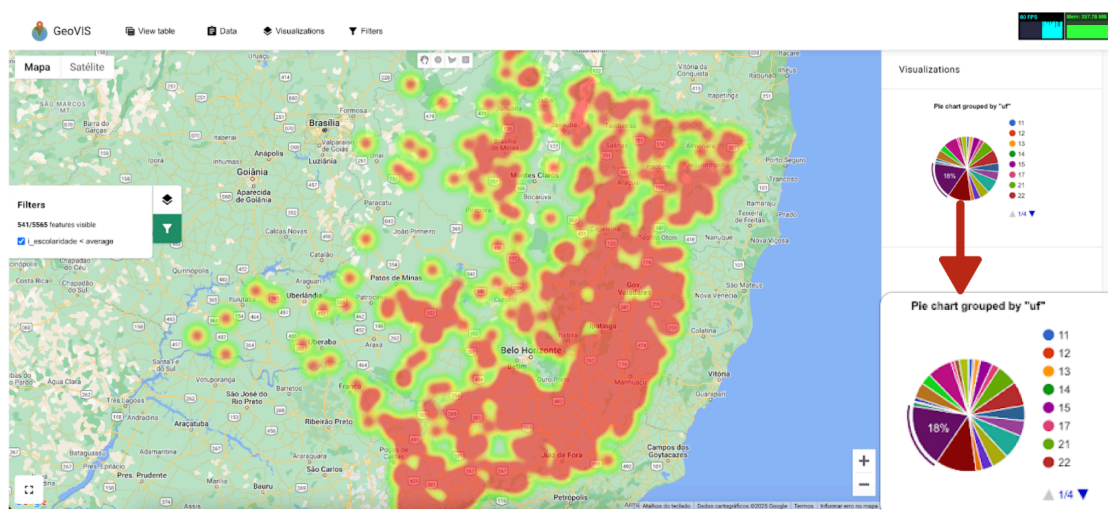


Figura 35: Representação dos municípios com índice de escolaridade abaixo da média no Gráfico de Setores agrupados por estado (Autor).

Outra possibilidade de exploração gráfica foi a adição da técnica de “Gráficos em marcadores” com o gráfico de setores no mapa para o atributo “Subíndice de frequência escolar” (i_freq_prop), facilitando, por exemplo, a evidência da frequência escolar destes municípios selecionados pelas visualizações anteriores, e permitindo entender informações de regionais sobre a frequência da população jovem às escolas. Como é possível visualizar na Figura 36, os gráficos representados nos marcadores das instâncias possuem as cores mais quentes na região sul do estado de Minas Gerais,

indicando em determinadas áreas a frequência escolar superior nestas regiões e, por outro lado, na parte nordeste do estado os municípios com menores taxas de frequência escolar.

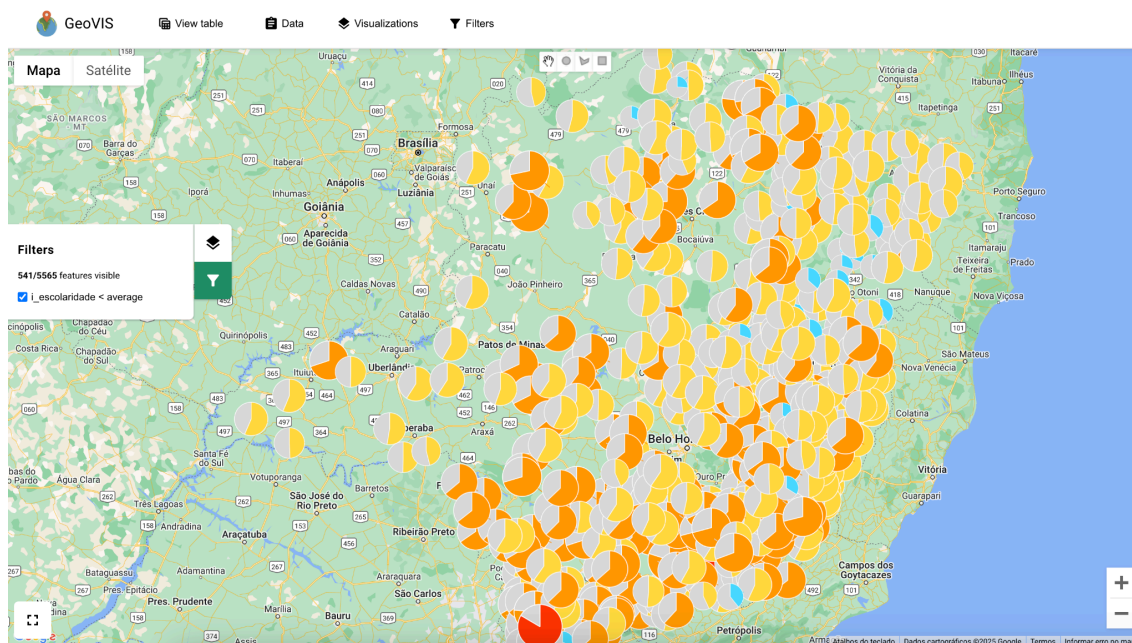


Figura 36: Representação gráfica dos municípios em relação à frequência escolar (Autor).

Outra visualização adicionada neste estudo de caso foi a técnica “Coordenadas Paralelas” junto a visualização de “Gráfico em marcadores” com as variáveis selecionadas para o gráfico: Índice de Desenvolvimento Humano Municipal (idhm), Subíndice de escolaridade - IDHM Educação (i_escolaridade) e Subíndice de frequência escolar (i_freq_prop). Como apresentado na Figura 37, esta nova visualização gerada pode representar nos marcadores do mapa os três atributos previamente selecionados (gráfico de radar) e com isso foi possível analisar a relação entre as variáveis renderizadas no gráfico. Em seguida foi adicionado uma seleção de dados com o filtro nas coordenadas paralelas para o atributo idhm contendo os valores acima de 75% do valor normalizado da coluna (como pode ser visto na coluna idhm da visualização). Assim as instâncias foram filtradas e representadas no mapa permitindo uma análise visual da correlação geográfica com os três atributos representados nos gráficos, possibilitando a compreensão de padrões nestes dados. Logo foi possível visualizar no mapa que: das instâncias com valores superiores para os atributos de idhm, as cores dos marcadores se mostraram em tons majoritariamente mais quente, indicando uma relação

entre municípios com maior idhm e escolaridade, possibilitando a compreensão que estes municípios de idhm mais alto, possuem em sua população a característica do índice de escolaridade mais alto e também uma maior frequência escolar pela população jovem.

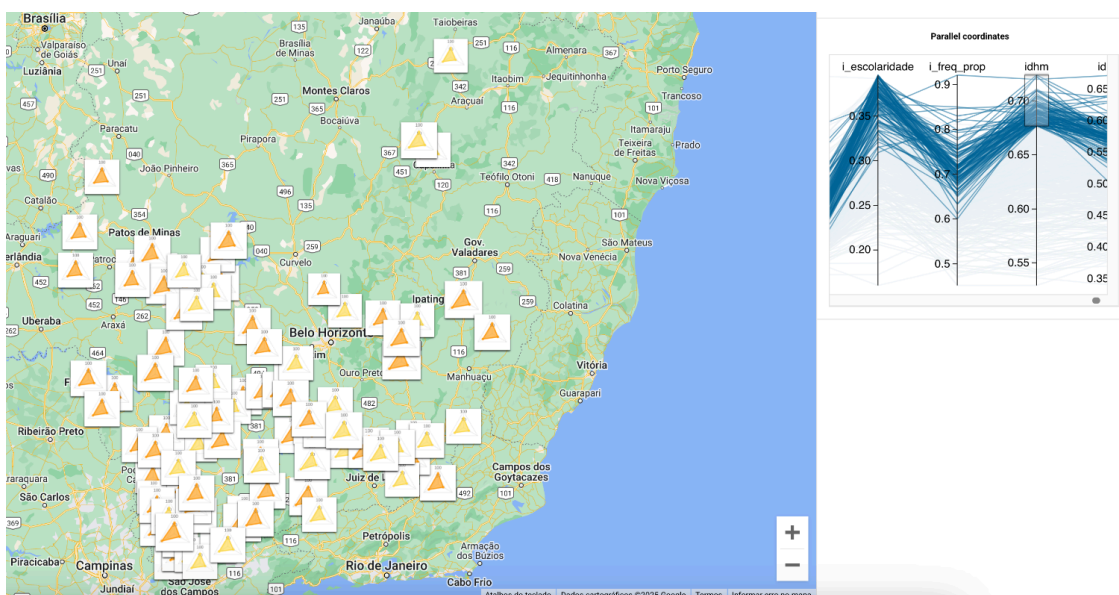


Figura 37: Representação gráfica do mapa com o filtro aplicado pela técnica Coordenadas Paralelas (Autor).

Como descrito, estas e outras análises podem ser aplicadas na ferramenta para potencializar a descoberta de padrões e a extração de informações dos dados, a combinação entre técnicas interativas favorece o processo exploratório de descoberta de conhecimento. Logo os testes iniciais evidenciam que a ferramenta GeoVis Explorer proporciona um ambiente interativo e flexível para a exploração de dados geoespaciais e multivariados por meio de suas visualizações customizáveis. Entre os principais benefícios observados, destacam-se: a facilidade na identificação de padrões em dados espaciais, a possibilidade de exploração simplificada por meio das representações gráficas e filtros interativos, a capacidade de coordenar as visualizações do modelo visual de forma intuitiva.

Com base nesses resultados iniciais, a ferramenta foi avaliada em relação às suas capacidades analíticas com o objetivo de se comprovar como um recurso poderoso para especialistas ao realizarem análises de dados geoespaciais junto a informações multivariadas.

5.2. Entrevistas para Avaliação da GeoVis Explorer

Com o objetivo de avaliar a GeoVis Explorer quanto a usabilidade, facilidade de utilização e também sobre a eficácia nos ganhos cognitivos possibilitados pelo novo modelo, foi necessária a participação de usuários e pesquisadores das áreas de Visualização de Informações para identificar como o desenvolvimento de uma abordagem de análise visual de dados geoespaciais fornece ganhos para análises visuais nestes tipos de dados. Para isso, foram realizadas entrevistas e questionários de avaliação da ferramenta com usuários de diferentes experiências e tempo de trabalho.

Para a realização das entrevistas e das pesquisas com os usuários, foram utilizados os dados abertos coletados a partir do CENSO 2022, que possui informações socioeconômicas dos municípios brasileiros. Em seguida os dados presentes neste conjunto foram normalizados com as coordenadas geográficas de cada um dos municípios e aplicados na ferramenta para a realização de testes do modelo proposto.

5.2.1. Condução das entrevistas

Entrevistas semiestruturadas foram conduzidas de modo remoto com profissionais da área de Ciência de Dados e Computação que utilizam (ou já utilizaram) ferramentas de representações gráficas em seu trabalho. A princípio o foco das entrevistas foi de familiarizar o usuário com a ferramenta e os dados, e em seguida propor a execução de pequenas tarefas de exploração de dados por meio das visualizações disponíveis no modelo, e por fim registrar a opinião dos usuários em relação aos ganhos proporcionados pelo modelo.

Durante as entrevistas, os objetivos do trabalho foram apresentados, assim como uma breve introdução das técnicas de visualização implementadas e também das interações visuais disponíveis na ferramenta.

Os entrevistados foram convidados via e-mail para as entrevistas, que foram realizadas por meio da plataforma de vídeo chamada Google Meet. As entrevistas foram gravadas com a autorização dos entrevistados e variaram entre 45 a 60 minutos de duração. A experiência e a especialidade de cada um dos entrevistados, foram validadas previamente por meio do seu perfil profissional no LinkedIn e também por meio de seu Currículo Lattes.

Para cada uma das entrevistas, as respostas e considerações foram transcritas e posteriormente disponibilizadas em uma planilha para que em seguida, fossem utilizadas para possíveis correções ou melhorias na ferramenta GeoVis Explorer.

O questionário para as entrevistas é apresentado pelo Apêndice A e as principais perguntas realizadas são:

- Quais as ferramentas de visualização de informações que possui maior familiaridade?
- Quais técnicas de visualizações listadas possui domínio; quais as impressões sobre o modelo proposto pela ferramenta?
- Quais as dificuldades ao se operar a ferramenta?
- Quais os ganhos no processo exploratório?
- Considera que a ferramenta pode ser utilizada na exploração de dados públicos?

5.2.2. Pareceres das entrevistas

Durante a realização das entrevistas, foram observados e coletados os pareceres e percepções por parte dos participantes sobre a ferramenta GeoVis Explorer. Apesar de existirem alguns pontos de melhoria que serão destacados a seguir, os usuários conseguiram utilizar a ferramenta sem dificuldades técnicas e elogiaram a facilidade de configuração do ambiente de visualização.

Quanto ao nível de familiaridade com tecnologias similares a este trabalho, as entrevistas realizadas evidenciam distintos níveis entre os participantes, os quais possuem formações e atuações profissionais diversas. A maior parte dos entrevistados possuem experiência em desenvolvimento de sistemas e ciência de dados, onde declararam o uso recorrente dessas ferramentas em suas atividades diárias. Além disso, alguns participantes mencionaram a experiência no desenvolvimento de ferramentas próprias voltadas à visualização de informações, indicando um conhecimento aprofundado na área em estudo. Houve entre os entrevistados, profissionais com um menor contato prático com ferramentas de análise visual, sendo que suas atividades profissionais são mais específicas na utilização de ferramentas visuais gráficas e não no desenvolvimento delas. No que se refere às técnicas de visualização utilizadas, os entrevistados mencionaram conhecer e utilizar com frequência gráficos de pizza, histogramas e mapas de calor, enquanto a técnica Coordenadas Paralelas apenas alguns usuários já possuíam conhecimento prévio. O uso de mapas foi majoritariamente associado a aplicações pessoais do dia-a-dia, com exceção de casos específicos em que sua utilização esteve relacionada a análises profissionais práticas.

Na etapa inicial de demonstração da ferramenta, a experiência guiada foi bem recebida, o que proporcionou um aprendizado direto sobre as funcionalidades e auxiliou os participantes na compreensão do funcionamento da ferramenta. Embora alguns tenham sugerido melhorias na documentação para garantir uma curva de aprendizado menor, a maioria relatou que a interface funciona de forma simples e que sua operação é intuitiva, especialmente para aqueles já familiarizados com ferramentas de análise visual.

A interatividade e a possibilidade de customização rápida do ambiente de visualização da ferramenta foram frequentemente mencionados como pontos positivos, assim como a capacidade de se utilizar a solução em diferentes bases de dados. Outro ponto positivo relatado foi a possibilidade de explorar e configurar visualizações complementares, com destaque para a interação entre múltiplas técnicas de visualização.

Como parte dos pareceres dos usuários, algumas limitações foram apontadas. A falta de um recurso para customizar os nomes das camadas de visualizações criadas foi um dos pontos mais recorrentes, pois poderia facilitar a organização e identificação das diferentes visualizações, visto que a ferramenta nomeia somente utilizando qual a

técnica de visualização de informações utilizada. Além disso, houve sugestões para adicionar um atalho de “lixeira” para remover filtros e camadas de forma mais intuitiva a partir do menu lateral. Um outro ponto mencionado foi a necessidade de um recurso informativo (ícone) ao lado das opções de visualização para explicar melhor o que é cada técnica, principalmente para usuários que nunca viram uma técnica em específico.

A experiência guiada foi elogiada, pois facilitou o uso inicial da ferramenta. No entanto, alguns participantes questionaram se a ferramenta permaneceria intuitiva sem um guia ou manual, especialmente para usuários sem experiência prévia. Um dos participantes relatou que mesmo sem um treinamento inicial guiado, que a criação de uma documentação simples já seria suficiente para um usuário começar a utilizar a ferramenta sem dificuldades. Além disso, alguns relataram a necessidade de compreensão dos dados antes de utilizar a ferramenta, destacando a necessidade do usuário que estiver operando a ferramenta já ter conhecimento dos atributos do conjunto de dados em processo de exploração.

Em relação a melhorias e sugestões de novas funcionalidades, houve recomendações para a implementação de filtros na visualização em tabela, a inclusão de visualização de série temporal e a possibilidade de modificar a visualização da técnica “Mapas de calor” com base em atributos relativos à intensidade, como, por exemplo, os atributos de população ou área para determinar “zonas de calor” representadas no mapa.

Com base nos pareceres dos utilizadores, as entrevistas evidenciaram que a ferramenta possibilita a exploração de dados geoespaciais e multivalorados de forma intuitiva, onde os participantes elogiaram a interface simples de utilizar, podendo ser utilizadas por qualquer usuário. Assim, as limitações discutidas no Capítulo 4 a partir da proposta deste trabalho, foram comprovadas por meio das entrevistas, validando a ferramenta em relação a utilização de suas funcionalidades pelos usuários.

6. CONCLUSÕES E TRABALHOS FUTUROS

Neste capítulo, serão apresentados as conclusões e oportunidades para trabalhos futuros identificados até o momento. O subcapítulo a seguir detalha as conclusões obtidas durante a etapa de estudo e validação, com ênfase na ferramenta desenvolvida e no potencial impacto dos estudos na compreensão e visualização eficaz de dados geoespaciais. Por fim serão descritas as oportunidades de evolução da ferramenta no subcapítulo sobre trabalhos futuros.

6.1. Conclusões

Este trabalho propôs e avaliou a ferramenta GeoVis Explorer, uma abordagem dinâmica e flexível para análise visual interativa de dados geoespaciais e multivalorados. A pesquisa teve como principal motivação as limitações encontradas nas ferramentas existentes, as quais, frequentemente, apresentam falta de funcionalidades em combinar representações visuais, integrar diferentes técnicas de análise e proporcionar um ambiente multidomínio para a exploração de dados em múltiplos domínios.

A ferramenta desenvolvida foi projetada para permitir a coordenação de múltiplas visualizações, possibilitando a análise exploratória e facilitando a identificação de padrões, correlações e tendências nos dados. Dentre os principais recursos implementados, destacam-se mapas de calor, gráficos embutidos em marcadores, coordenadas paralelas e filtros espaciais interativos. A validação experimental foi conduzida por meio de entrevistas e testes com usuários, os quais indicaram que a GeoVis Explorer contribui para uma percepção dos dados de forma mais intuitiva e aprimorou o processo de abstração de conhecimento para a tomada de decisão.

Os resultados evidenciam que a ferramenta permite a visualização integrada de informações georreferenciadas e multivaloradas, reduzindo a complexidade inerente à análise desses dados. No entanto, algumas oportunidades de melhoria foram identificadas ao longo da pesquisa, possibilitando o desenvolvimento de trabalhos futuros.

Do ponto de vista científico, este estudo contribui para as áreas de Visualização de Informações, Mineração de Dados Geoespaciais e Sistemas de Informação Geográfica (SIG), fornecendo um modelo flexível para a análise exploratória de dados. Logo, conforme evidenciam as análises realizadas durante a pesquisa com usuários, a ferramenta proposta pode ser aplicada para a análise de dados de diferentes domínios, como análise socioeconômica, monitoramento ambiental e planejamento urbano, demonstrando sua versatilidade e potencial impacto em diversas áreas do conhecimento, como na análise de dados de cidades inteligentes.

6.2. Trabalhos futuros

Com base nas conclusões obtidas, diversas oportunidades para trabalhos futuros podem ser exploradas. Foram citados durante as entrevistas, possíveis otimizações que faria a ferramenta oferecer mais vantagens na utilização dos usuários, como por exemplo: a otimização de desempenho da ferramenta com grandes volumes de dados, adição de novas técnicas de visualização, além disso a usabilidade da interface pode ser aprimorada para atender usuários com diferentes níveis de familiaridade com visualização de dados. Há também a oportunidade de adição de integração com bancos de dados geoespaciais, permitindo a atualização dinâmica e consultas dos dados em tempo real.

Outra oportunidade seria implementar a ferramenta em projetos reais de análise urbana, mobilidade, infraestrutura e segurança pública, avaliando seu impacto na formulação de políticas públicas baseadas em dados gerados por cidades inteligentes. Assim é possível também incorporar técnicas de aprendizado de máquina e redes neurais para recomendar visualizações e identificar automaticamente anomalias e correlações não triviais nos dados utilizando técnicas de inteligência artificial.

REFERÊNCIAS

- [1] Marcos R. Vieira, et al., 2012, Bancos de Dados NoSQL: Conceitos, Ferramentas, Linguagens e Estudos de Casos no Contexto de Big Data Disponível
- [2] Henrique Freitas, 2000. Análise de dados qualitativos: aplicação e tendências mundiais em sistemas de informação. Revista de Administração-RAUSP, 35.4: 84-102.
- [3] Mackinlay Card, 1999. Readings in information visualization: using vision to think. Morgan Kaufmann.
- [4] Muzammil Khan; KHAN, Sarwar Shah, 2011. Data and information visualization methods, and interactive mechanisms: A survey. International Journal of Computer Applications, 34.1: 1-14.
- [5] Thomas Cook, 2005. "Illuminating the path: The research and development agenda for visual analytics", Relatório Técnico, Pacific Northwest National Lab., 190p.
- [6] Wagner Gonçalves Signoretti, 2018. Modelo para análise visual de dados multivariados no contexto geoespacial.
- [7] Michael Batty, et al, 2012. Smart cities of the future. The European Physical Journal Special Topics, 214.1: 481-518.
- [8] Cleyber N. de Medeiros, Mário Aragão, Daniel D. M. GOMES, 2010. Disponibilização e Análise de Informações Georreferenciadas utilizando um Sig-Web. Fortaleza: IPECE.
- [9] Fabio Kon, Eduardo Felipe Zambom Santana, 2016, "Cidades Inteligentes: Conceitos, plataformas e desafios"
- [10] Harrison, C., Eckman, B., Hamilton, R., Hartswick, P., Kalagnanam, J., Paraszczak, J., and Williams, P., 2010. Foundations for smarter cities. IBM Journal of Research and Development, 54(4):1-16.

- [11] Leonardo S. Bastos, et al., 2020. "COVID-19 e hospitalizações por SRAG no Brasil: uma comparação até a 12ª semana epidemiológica de 2020." *Cadernos de Saúde Pública* 36.
- [12] Fernando Amaral, 2016. *Introdução à Ciência de Dados: mineração de dados e big data*. Alta Books Editora.
- [13] Cássio O. Camilo, João Carlos da Silva, 2009. "Mineração de dados: Conceitos, tarefas, métodos e ferramentas." *Universidade Federal de Goiás (UFG)*: 1-29.
- [14] Carla M. D. S. Freitas, et al., 2001. "Introdução à visualização de informações." *Revista de informática teórica e aplicada*. Porto Alegre. Vol. 8, n. 2 (out. 2001), p. 143-158.
- [15] Gilberto Câmara, Clodoveu Davis, Antônio Miguel Vieira Monteiro, 2001. "Introdução à ciência da geoinformação."
- [16] Stan Aronoff, 1991. *Geographic information systems: a management perspective*. No. BOOK. WDL publications.
- [17] Ricard M. Kleiner, Arlete A. C. Meneguette, 2002. "A tecnologia dos Sistemas de Informações Geográficas (SIG) e a internet." *SIMPÓSIO BRASILEIRO DE GEOMÁTICA, II*: 357-364.
- [18] Jonathan Roberts, 1998. "On encouraging multiple views for visualization". In: *Proceedings of the IEEE Conference on Information Visualization*, pp. 8–14.
- [19] Carolyn Boyce, Pamela Neale, 2006, "Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input". Watertown, United States: Pathfinder International, 2 ed., 16p.
- [20] Renan Oliveira, Danilo Eler, 2016, *Geovis Explorer - Uma Ferramenta para exploração de conjuntos de dados em um ambiente georreferenciado utilizando uma aplicação de mapas na web*.

- [21] Marcello Martinelli, 2005, Os Mapas da Geografia, Disponível em: http://www2.fct.unesp.br/docentes/geo/raul/cartografia_tematica/leitura%20201-MAPAS%20DA%20GEOGRAFIA.pdf
- [22] Zhang, Zhiyuan, et al. "An interactive visual analytics framework for multi-field data in a geo-spatial context." *Tsinghua Science and Technology* 18.2 (2013): 111-124.
- [23] DE MEDEIROS, Cleyber Nascimento, et al. Utilização de SIG-WEB usando software livre para disponibilização de dados georreferenciados na internet: Caso do Sistema Ceará em Mapas Interativos. SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 2011, 15: 8740-8747.
- [24] MEIER, Sebastian; HEIDMANN, Frank. Too many markers, revisited: An empirical analysis of web-based methods for overcoming the problem of too many markers in zoomable mapping applications. In: 2014 14th International Conference on Computational Science and Its Applications. IEEE, 2014. p. 121-125.
- [25] SANCHEZ, Luis, et al. SmartSantander: IoT experimentation over a smart city testbed. *Computer Networks*, 2014, 61: 217-238.
- [26] FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. *AI magazine*, 1996, 17.3: 37-37.
- [27] Christófano, R. M., Júnior, W. E. M., & Eler, D. M. (2021). PlaceProfile: Employing Visual and Cluster Analysis to Profile Regions based on Points of Interest. In *ICEIS* (1) (pp. 506-514).
- [28] Brodbeck, D., & Girardin, L. (2003, July). Design study: Using multiple coordinated views to analyze geo-referenced high-dimensional datasets. In *Proceedings International Conference on Coordinated and Multiple Views in Exploratory Visualization-CMV 2003*- (pp. 104-111). IEEE.

- [29] Sjöbergh, J., & Tanaka, Y. (2013, September). Geospatial digital dashboard for exploratory visual analytics. In International Workshop on Information Search, Integration, and Personalization (pp. 3-17). Cham: Springer International Publishing.
- [30] Al-Hajj, S., Pike, I., & Fisher, B. (2013, November). Interactive dashboards: Using visual analytics for knowledge transfer and decision support. In Proceedings of the 2013 Workshop on Visual Analytics in Healthcare, Washington, DC, USA (Vol. 16).
- [31] Sironi, A. P., Bertoldo, J., Sampaio, V., Coimbra, D., Rasella, D., & Barreto, M. E. (2022). Malaria-VisAnalytics: a tool for visual exploratory analysis of Brazilian public malaria data. *Malaria Journal*, 21(1), 1-16.

DADOS CURRICULARES

Nome completo: Renan Augusto Pupin de Oliveira

Data de nascimento: 18/11/1993

Nacionalidade: Brasileira

Nome em citações bibliográficas:

- OLIVEIRA, Renan Augusto Pupin de
- Oliveira, R. A. P.

Formação Acadêmica: Bacharel em Ciência da Computação, 2016, Universidade Estadual Paulista Júlio de Mesquita Filho - UNESP

Currículo Lattes: <http://lattes.cnpq.br/6133608721897573>

APÊNDICE A – CONDUÇÃO PARA AS ENTREVISTAS

1 - Apresentação da metodologia da entrevista

- Esta metodologia tem o objetivo de validar um estudo de mestrado criado com objetivo de otimizar as análises de dados em um ambiente de visualização de mapas.
- A entrevista terá a duração de 45 minutos, com o propósito de coletar as impressões sobre as técnicas utilizadas, sobre a ferramenta e também sobre a possibilidade de se utilizar esta ferramenta para realizar análises de dados de domínio público.
- A entrevista será gravada e todo o conteúdo aqui será utilizado para comprovar a metodologia científica presente neste projeto.
- Assinatura dos termos de consentimento

2 - Contexto do trabalho

- O projeto iniciou-se durante a minha graduação com o apoio do professor Danilo Eler.
- Hoje existem ferramentas de visualizações de informações (BI) geográficas e espaciais, no entanto este estudo busca se aprofundar mais em técnicas e permitir uma interação coordenada do utilizador com os dados
- A ferramenta funciona em um ambiente georreferenciado da web e possui gráficos e técnicas de visualização para dar suporte a exploração dos dados
- Os dados que serão utilizados nesta entrevista, são relacionados ao Censo de 2010, onde existem informações socioeconômicas dos municípios do Brasil, com informações como população, PIB, taxa de escolaridade, etc.

3 - Perguntas iniciais

- Qual o seu nome e data de nascimento?
- Qual a sua formação acadêmica?
- Qual a sua profissão e local de trabalho?
- Você possui familiaridade com ferramentas de análises visuais? Quais ferramentas?

- Você utiliza ou já utilizou estes tipos de ferramentas no seu dia-a-dia?
- Quais das seguintes técnicas de visualização você domina?
 - Mapas
 - Gráfico de pizza, gráfico de setores, gráfico de barras, gráfico polar, gráfico de área
 - Histograma
 - Coordenadas paralelas
 - Mapas de calor

4 - Exploração guiada

- Apresentação da interface da ferramenta e modos de visualização
- Apresentar como carregar dados
- Apresentar o menu de filtros de dados
- Apresentar o filtro espacial
- Apresentar a adição de técnicas de visualização
- Apresentar a coordenação entre técnicas

5 - Exploração livre

- Agora você poderá utilizar as técnicas da ferramenta para executar algumas ações:
 - Encontrar quais cidades possuem o pib per capita acima da média
 - Gerar histograma para visualizar a distribuição de cidade dentro dos estados
 - Gerar um gráfico de setores para visualizar a distribuição de população destas cidades
 - Utilizar a CMV dentro do gráfico de pizza para selecionar determinada classificação de municípios

6 - Feedbacks

- Quais as impressões sobre a ferramenta?
- Considera que a ferramenta é intuitiva e simples de operar?
- Quais foram as dificuldades?
- A ferramenta ajuda no processo exploratório de análise dos dados?

- A ferramenta pode ser usada na exploração de dados públicos?
- Acredita que profissionais de diferentes domínios com conhecimento em ferramentas visuais possam operar a ferramenta sem dificuldades?
- Quais os principais pontos positivos ao utilizar a ferramenta?
- Quais os principais pontos negativos ao utilizar a ferramenta?
- Acredita que alguma outra técnica possa ser implementada na ferramenta?
- Gostaria de compartilhar algum feedback ou compartilhar alguma sugestão de melhoria?