

# Uma Ferramenta para exploração de conjuntos de dados em um ambiente georreferenciado utilizando uma aplicação de mapas na web

Renan Augusto Pupin de Oliveira, Danilo Medeiros Eler

Curso de Bacharelado em Ciência da Computação – Universidade Estadual “Júlio de Mesquita Filho” (UNESP) – Campus de Presidente Prudente  
CEP 19060-900 – Presidente Prudente – SP – Brasil

renan.pupin@gmail.com, danilo@fct.unesp.br

**Abstract.** *Given the large amount of available information, the data is stored and generated on a large scale. The need to conduct studies on these data, there is some difficulty with the effort to understand such information. From this paradigm, you need something that facilitates the process of exploration of large data sets. The visual representation of approaches has been extremely valuable tools for this type of study. To this end, the use of representations through charts and maps are of great importance for this application, as well as enabling the user to have a canonical perspective representation of the data also allows it to have a geographical perspective in a georeferenced environment easy interaction with the data. Using the combination of these features, the tool is not only able to facilitate the discovery of patterns and survey study of assumptions about the data set, but it also has extreme ability to facilitate abstraction and converting information into relevant knowledge about the data study.*

**Resumo.** *Diante da numerosa quantidade de informações disponíveis, os dados são armazenados e gerados em grande escala. Na necessidade de realizar estudos sobre estes dados, há uma certa dificuldade em relação ao esforço para o entendimento de tais informações. Partindo deste paradigma, é necessário algo que facilite o processo de exploração de grandes conjuntos dados. As abordagens de representação visuais tem se mostrado ferramentas extremamente valorosas para este tipo de estudo. Com este objetivo, o emprego de representações por meio de gráficos e mapas são de grande importância para esta aplicação, pois além de possibilitar o usuário de ter uma perspectiva canônica da representação dos dados, permite também que ele tenha uma perspectiva geográfica em um ambiente georreferenciado de fácil interação com os dados. Utilizando a combinação destes recursos, a ferramenta não só é capaz de facilitar a descoberta de padrões e o levantamento de hipóteses de estudo sobre o conjunto de dados, mas também possui extrema capacidade de facilitar a abstração e a conversão de informações em conhecimentos relevantes sobre os dados de estudo.*

## 1. Introdução

Atualmente, muitas aplicações disponibilizam dados em grande escala, enquanto que a tecnologia para análise e descoberta do conhecimento a partir destes dados é ainda limitada. Pesquisas tem sido desenvolvidas para lidar tanto com a grande quantidade de elementos quanto a grande quantidade de variáveis em conjuntos de dados e fornecer mecanismos mais efetivos para a análise destes dados. Assim, a integração destas áreas para a exploração e análise de conjuntos de dados é conhecida como analítica visual [Wong e Thomas, 2004]. Esta analítica visual engloba a estatística e é feita de maneira mais eficiente por meio da aplicação de técnicas apropriadas de visualizações de informações em conjuntos de dados multidimensionais. Além disso, lidar com grandes volumes de dados exige muito trabalho, causando uma ineficiência no estudo e em certos cenários, pode não ser uma tarefa trivial.

Abordagens que utilizam técnicas de visualização são largamente empregadas com o objetivo de dar sentido a esses dados, por meio de representações visuais e de formas de interação [Card et al., 1999]. Com o surgimento de técnicas de visualização, é possível obter uma melhoria na abstração e entendimento dos dados, além de permitir ao usuário interagir sobre o conjunto, descobrir padrões, identificar anomalias, levantar hipóteses de estudo e facilitar a tomada de decisões estratégicas. Por meio das formas de representação, o surgimento e avanço desta área na computação possibilita que o próprio computador detecte relações que sejam úteis para estudos pelo conceito chamado de mineração de dados. A mineração de dados surge então tanto para a pesquisa científica, quanto para aplicações diversas, como estimular a competitividade entre empresas diante a necessidade da tomada de decisões.

Embora as técnicas de visualização possuam um poder maior em dar sentido aos dados, toda técnica possui sua limitação. Geralmente, uma técnica consegue apresentar uma visualização focada no conjunto de atributos ou focar sua representação no relacionamento entre as instâncias. Para isso, na tentativa de suavizar os pontos fracos de uma técnica, diferentes técnicas podem ser combinadas de forma complementar, possibilitando que o usuário tenha diferentes perspectivas sobre um conjunto de dados.

Estudos realizados desde a antiguidade buscam também representar visualizações utilizando mapas. Estes estudos mostraram que mapas georreferenciados são de extrema relevância para análise de certos comportamentos em áreas delimitadas [Furtado et al., 2008]. Levando isso em consideração, o desenvolvimento de uma ferramenta para complementar esta exploração, representando informações em um ambiente geográfico, facilitará a realização de estudos, permitindo novas conclusões relativas a áreas e regiões de interesse. Tendo como foco esta abordagem, o objetivo deste trabalho foi integrar diversas formas de visualização como mapas de calor, gráficos e linhas de relacionamento para criar representações que facilitem a exploração de conjuntos de dados em mapas. Assim o processo de estudo e mineração dos dados do conjunto permite a investigação à procura de informações úteis, possibilitando a

descoberta de padrões, que possivelmente influencie decisões com um maior grau de confiança.

Neste trabalho as seções definem a organização em que as informações estão dispostas. A Seção 1 apresenta a introdução sobre o assunto que será argumentado ao longo do trabalho; a Seção 2 apresenta uma fundamentação teórica sobre os conceitos utilizados neste estudo; a Seção 3 cita os trabalhos relacionados e o impacto dos benefícios resultantes de cada estudo; a Seção 4 discorre sobre a ferramenta desenvolvida; a Seção 5 mostra um estudo de caso utilizando o conjunto de dados na ferramenta; já na Seção 6 são demonstradas as conclusões obtidas pelo desenvolvimento deste trabalho.

## **2. Fundamentação Teórica**

Os mapas foram criados diante a necessidade humana se localizar, conhecer áreas, rotas e representar seu espaço [Martinelli, 2005]. Eles são representados por meio de uma projeção cartográfica, onde esta projeção transforma coordenadas polares (devido a curvatura e o formato físico da terra) em coordenadas que podem ser representadas em um plano do mapa. Assim, um mapa é uma representação visual bidimensional de um espaço tridimensional. Ele provê a visualização de dados espaciais buscando representar os aspectos geográficos, delimitados por elementos físicos e administrativos.

Mapas modernos contam com a presença de escalas, que permitem determinar a medida real correspondente de certa distância, área ou forma baseada no quociente representado no mapa. Graças às propriedades métricas, os mapas permitem extrair medidas de superfícies, ângulos e distâncias com uma precisão praticamente exata. Por isso, eles constituem uma importante fonte de informação e permitem o desenvolvimento de diversas atividades utilizando a visualização proporcionada. Com a tecnologia atual, as técnicas baseadas em fotografias por satélite permitem o conhecimento do contorno exato de uma região, além de fornecer dados relacionados a hidrografia e relevo.

Assim baseado nas capacidades que os mapas proporcionam, é possível desenvolver uma vasta quantidade de aplicações utilizando dados georreferenciados levando em consideração áreas geográficas e obter conclusões acerca de estudos destas aplicações. Neste trabalho foram estudadas abordagens que permitam a abstração de informações relevantes de conjuntos de dados georreferenciados, por exemplo, dados de agricultura, climatologia, dados urbanos, entre outros. Estes dados poderão trazer qualquer tipo de informação, visto que o foco deste trabalho foi proporcionar a capacidade de iteração e representação com conjuntos de dados, aplicando uma abordagem estatística e permitindo assim a facilidade de abstração de informações relevantes, levando em consideração áreas, delimitações e regiões cartográficas.

Ao estudar técnicas renomadas de representações em ambientes georreferenciados, chegamos ao conceito de mapas de calor. Um Heatmap, ou simplesmente “mapa de calor”, é um tipo de representação em mapas que mapeia uma elevada concentração de atividade, destacando a densidade ou magnitude de informações de pontos relacionados.

Como muitos estudos mostraram, os mapas de calor são uma das melhores técnicas de visualização de dados quando o objetivo é destacar concentrações de atividades [Khoo et al., 2014]. O gradiente de cor padrão configura o menor valor no mapa como azul escuro e o maior valor como vermelho intenso, representando com uma transição entre estes extremos os valores intermediários. Por meio desta técnica de representação, “pontos quentes” podem ser identificados sem muito esforço.

Nos mapas de calor, o gradiente das cores é gerado a partir da adição de valores de intensidade na escala de cores (Figura 1). Para dar uma breve ideia de como as cores são representadas, é possível observar a sobreposição das cores na escala das cores primárias vermelho, verde e azul (RGB). Ao realizar a sobreposição de pontos no mapa, os aglomerados de maior intensidade tendem a ficarem mais “quentes”, variando os tons de cores até os extremos.



Figura 1: Representando a escala de cores dos heatmaps, da esquerda para a direita os valores de intensidade aumentam.

Na Figura 2, é apresentada uma aplicação da técnica de mapas de calor, que pode ser uma técnica de extrema relevância na visualização de dados em mapas georreferenciados. Nesta figura, é possível notar “áreas quentes” em relação à quantidade de cliques realizados pelos usuários, onde a grande concentração de intensidade está no acesso aos três primeiros links da página de busca.

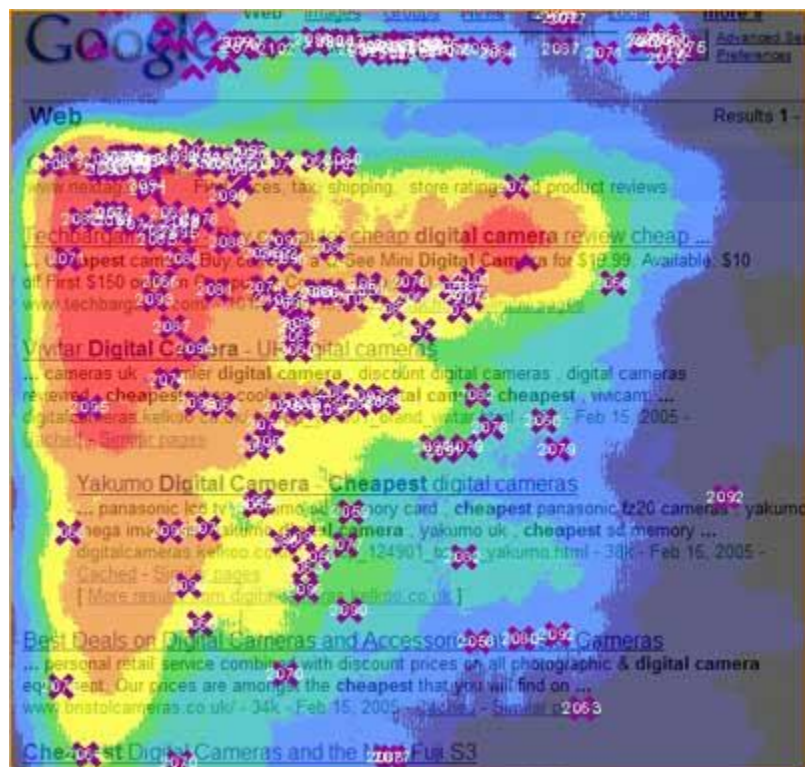


Figura 2: Áreas de calor representando pontos mais clicados em uma pesquisa realizada no Google. Imagem extraída de:  
<https://smellypunks.files.wordpress.com/2009/09/google-serp-heat-map.jpg>

Ao representar dados georreferenciados em mapas, certamente uma grande capacidade de análise de conhecimento é intensificada por meio da visualização de informações nesse ambiente. Porém, a representação de tais dados pode ser ainda mais potencializada ao traçar linhas entre as instâncias das informações representadas baseando-se em critérios de similaridade, é possível intensificar ainda mais a capacidade de abstração. A ligação, relacionamento ou simplesmente associação, utiliza mecanismos para estabelecer relacionamento entre as instâncias de objetos. Uma ligação é uma conexão conceitual entre duas instâncias tomando como critério algum atributo com estrutura e semântica comuns. Por exemplo, podemos associar duas instâncias pela relação em que possuem em algum de seus atributos, como realizar a associação entre elas onde um determinado atributo é igual, ou até mesmo próximo para ambas. Por exemplo, na Figura 3, relações entre instâncias foram demonstradas ao plotar linhas entre estados com atributos relacionadas a aquele em que foi selecionado.

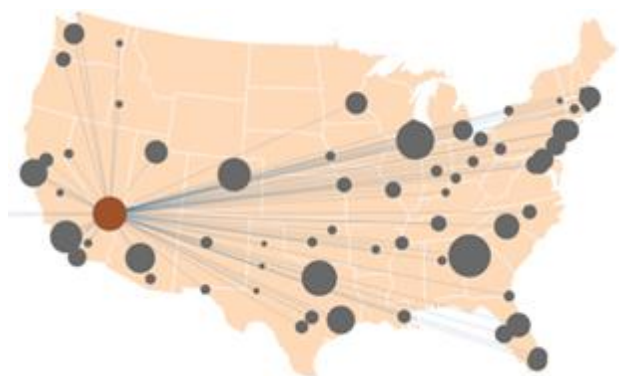


Figura 3 - Representação de semelhança entre instâncias em um mapa online. Figura obtida em: <http://leekuaneyew.straitstimes.com/ST/recap/index.html>

Em uma aplicação de mapas online, é possível combinar as diferentes técnicas de visualização a fim de obter-se uma abstração gráfica melhorada em relação aos valores dos atributos e realizar a ligação entre instâncias que se relacionam. A Figura 4 é um exemplo disso, onde, além de fornecer uma ligação entre instâncias similares, apresenta informações estatísticas em gráficos baseando-se no conjunto de instâncias tracejadas no mapa.

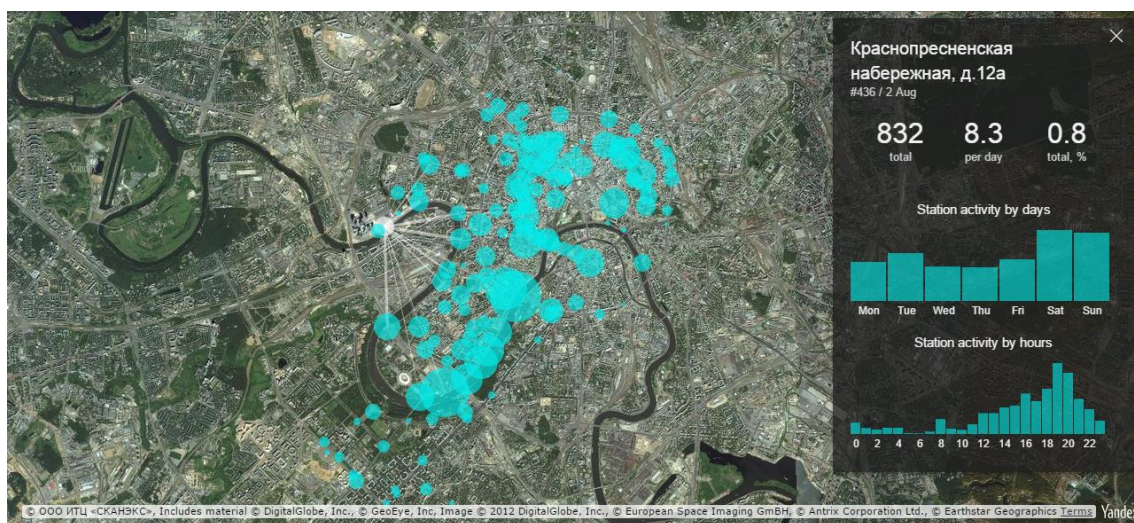


Figura 4 – Representações em mapas integrados a gráficos e destaque de relacionamento entre instâncias. Imagem extraída de <http://urbica.co/velo/>

Por meio destas representações combinadas, podemos definir o conceito de técnicas de visualização híbridas. As técnicas de visualização híbridas têm sido largamente utilizadas na exploração de conjuntos de dados, pois buscam dar sentido aos grandes volumes de dados por meio de representações visuais. Alguns objetos podem conter um número grande de atributos, e devido à grande dimensão destes objetos e volume de dados, a representação destes conjuntos se torna uma tarefa complicada. Com isso a

representação se torna um desafio, visto que a capacidade humana de analisar e identificar todas as informações relevantes disponíveis permanecem a mesma [SALAZAR 2012].

Para auxiliar nesse processo de exploração, uma estratégia comum é utilizar diferentes técnicas, dando ao usuário diferentes perspectivas de um mesmo conjunto de dados. O principal objetivo da visualização de dados é demonstrar a informação de maneira mais clara e objetiva utilizando meios gráficos. Assim é possível realizar a compreensão de um complexo conjunto de dados comunicando seus principais aspectos de forma mais intuitiva.

Diante de várias possibilidades de representação dos dados, muitas das maneiras abordadas conseguem representar os dados fornecendo diferentes perspectivas de representação. Como mostrado em estudos atuais, ao combinar várias formas de representação, é possível tirar conclusões acerca de grande volume de dados de uma maneira quase instantânea.

### **3. Trabalhos Relacionados**

Em 2011, Geisa Bugs realizou estudos sobre mapas interativos na web relativos a participação popular no planejamento urbano. Nesse trabalho, a autora fundamentou-se na falta de interesse dos cidadãos diante a participação do planejamento urbano. No estudo de fundamentação, o trabalho indica motivos que influenciam a falta de interesse da população na participação desta área. A autora argumenta que o principal motivo de desinteresse é a falta de igualdade nos diferentes setores da sociedade assim como a falta de acesso a meios de contribuição. Diante de tais circunstâncias, o estudo foi realizado tentando facilitar este processo de interação por meio da utilização de mapas, aumentando assim a interatividade entre usuários e sistemas de coletas de informação. Os serviços de mapas online estão cada vez mais próximos do público, e permitem uma rápida navegação para a visualização de locais em múltiplas escalas permitindo assim a massificação de seu uso e a criação de conteúdo georreferenciado de acesso público [BUGS, 2011].

A cognição da informação envolve a percepção humana por meio da representação geográfica, sendo extremamente relevante para assuntos como coleta de dados e representações gráficas. Assim a pesquisa de aspectos cognitivos que envolvem uso de informações e ferramentas geográficas pelo público auxilia no entendimento e compreensão de tais informações, que foram representadas na forma de mapas. Contudo considerando a representação das informações em mapas, foram realizados estudos de cartografia acerca do tema abordado, fornecendo assim suporte para uma melhor participação pública no tema de planejamento urbano utilizando o conceito da ciência cognitiva.

O Wiki Crimes permite que os usuários visualizem, pesquisem e abstraíam informações representadas por gráficos e a utilização da técnicas Heatmap para indicar a



aglomeração de incidência de crimes. A página é iterativa e exibe estatísticas automaticamente sobre as contribuições.

Já em outro trabalho, foram realizados estudos da gestão de dados e da geoinformação. Neste caso foi conclusivo que o processo de gestão do acesso a geoinformação e a representação em mapas, permite que pessoas tomem decisões mais eficientes em relação a solução de problemas sociais, urbanos e ambientais. Foram constatados aspectos de organização de acesso a informação, que quanto mais eficiente for a organização dos dados geoespaciais, mais aperfeiçoada será a interpretação e a representação gerada [Hubner, 2008].

Como descrito no estudo, um geoportal é uma importante aplicação que disponibiliza informações geoespaciais na web. A organização dos dados nesse caso é por meio da utilização de metadados (“dados sobre dados”), que descrevem especificações em relação as informações disponibilizadas pela aplicação, como localização, projeções nativas e a explicação das informações disponíveis.

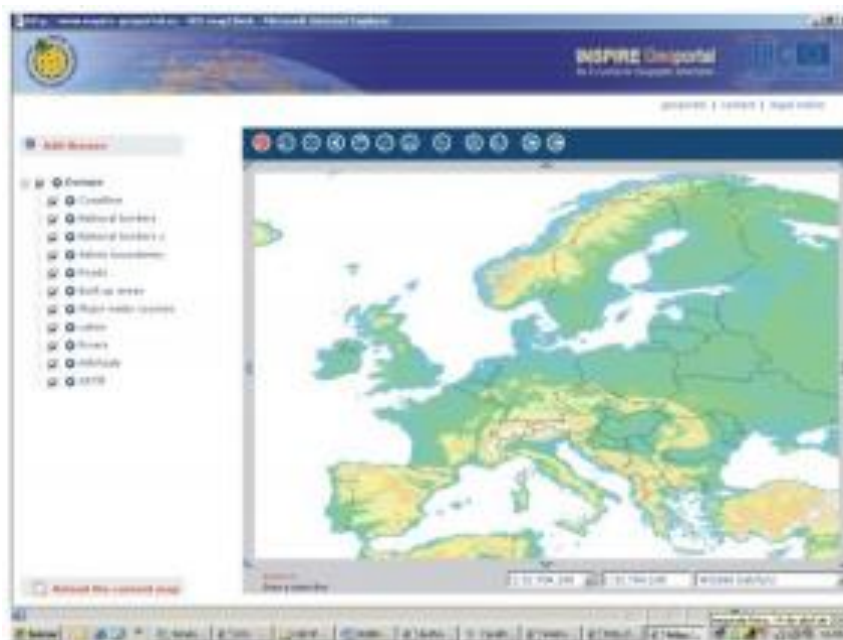


Figura 5: “GeoConnections Discovery Portal” demonstrando a recuperação de informações em aplicações web. Figura extraída de [Hubner, 2008].

Em outro estudo realizado por Medeiros e Alves em 2010, a disponibilização e análise de informações georreferenciadas utilizando SIG's WEB (Sistema de Informações Georreferenciadas WEB) foram analisadas de maneira significativa. Os autores estudam a importância de informações representadas em mapas interativos, para auxiliar gestores na aplicação do planejamento de políticas públicas e tomadas de decisão. A interface do mapa em questão permitia a representação de imagens de satélite, do meio ambiente, de recursos hídricos, de sistemas de transportes, de equipamentos públicos e privados,



e de indicadores socioeconômicos. Ao analisar tais representações, a ferramenta de mapas permitia ao usuário adicionar camadas nos mapas, localizar objetos geográficos e aplicar filtros baseado em regras de seleção entre os atributos do banco de dados. [Medeiros e Alves, 2010].

Nas Figuras 6 e 7, o autor demonstra a facilidade que uma ferramenta de mapa web pode proporcionar ao permitir a interação do usuário com os dados brutos por meio de uma forma de representação da informação. A aplicação possibilita a utilização de filtros, geração de gráficos dinâmicos, seleção de áreas geográficas para a localização de pontos de desejo próximos, medição de distâncias entre pontos de estudo, a possibilidade de consulta às informações e a visualização de detalhes do terreno.

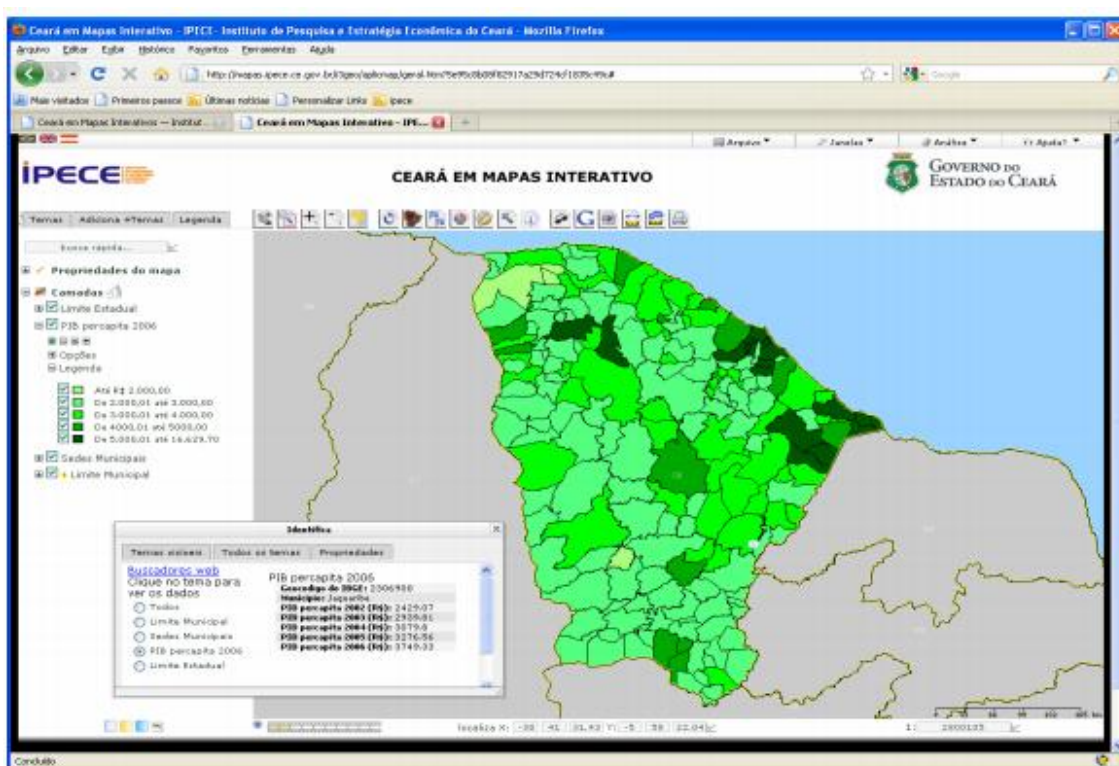


Figura 6: Mapa do PIB per capita dos municípios cearenses em 2006 com a inclusão de um filtro fornecido pela ferramenta. Figura extraída de [Medeiros e Alves, 2010].

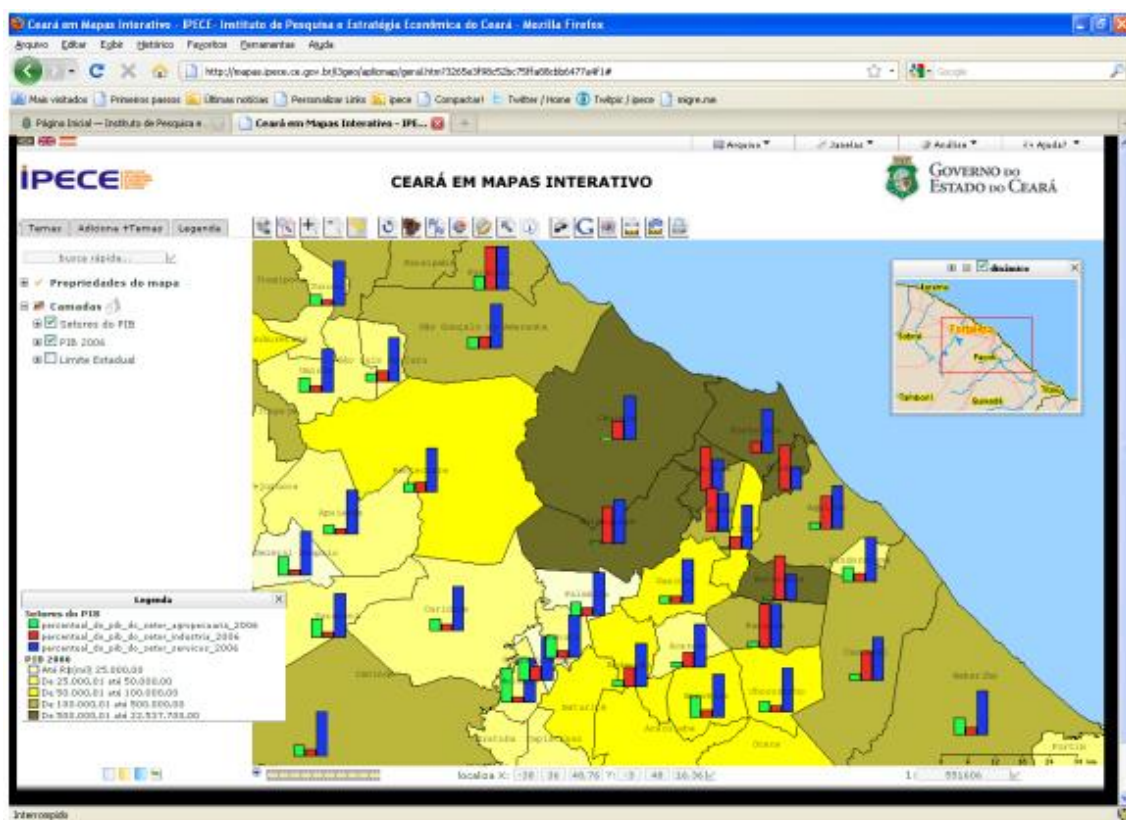


Figura 7: Mapa socioeconômico e gráficos dinâmicos gerados pelo mapa interativo referente ao estado do Ceará. Figura extraída de [Medeiros e Alves, 2010].

Por meio de tais representações e facilidades proporcionadas por sistemas interativos, é possível chegar a conclusões de alta relevância, no caso a capacidade de melhora no planejamento de políticas públicas e assim a melhoria na qualidade de vida da população residente na área de estudo. Mais uma vez foi constatado o enorme potencial de geoprocessamento por meio da coleta, padronização e análise de informações georreferenciadas. Com isso o sistema desenvolvido enriquece o valor das informações e aumenta sua importância por meio da utilização de representações em mapas interativos, pois melhoram a precisão de acerto na tomada de decisões em regiões que necessitam de melhorias socioeconômicas. [Medeiros e Alves, 2010].

#### 4. A Ferramenta Geovis Explorer

A ferramenta desenvolvida foi nomeada “Geovis Explorer” e está disponível na plataforma de código open source “Github”. Ela está sob a licença MIT e seu código pode ser acessado em: <https://github.com/renanpupin/geovis>. Uma demonstração de seu funcionamento está disponível por meio do link: <http://renanpupin.github.io/geovis/>.

A ferramenta é capaz de carregar conjuntos de dados no formato JSON (Javascript Object Notation), onde a escolha deste formato para o armazenamento dos dados em relação

ao XML é devido ao formato JSON oferecer uma performance de parse mais rápida que o parse do XML.

Na formatação das informações do conjunto de dados, cada representação de uma instância deve possuir um identificador numérico único, assim como dois atributos de geolocalização contendo as coordenadas de latitude e longitude para representação geográfica. Já no vetor denominado “infodata”, todos os atributos que representam as informações para o estudo do conjunto de dados deverão ser incluídas. Cada um destes atributos deverá possuir um nome e um valor de tipo numérico, textos ou booleanos, seguindo o formato semelhante ao exemplo da Figura 8.

```
1  [
2  .....{
3  .....    "id": 1,
4  .....    "geodata":
5  .....    {
6  .....        "lat": -22.075,
7  .....        "lon": -51.425
8  .....    },
9  .....    "infodata": [
10 .....        {
11 .....            "name": "nome",
12 .....            "value": "feature2"
13 .....        },
14 .....        {
15 .....            "name": "categoria",
16 .....            "value": "cat2"
17 .....        },
18 .....        {
19 .....            "name": "valor",
20 .....            "value": 5
21 .....        },
22 .....        {
23 .....            "name": "ativo",
24 .....            "value": false
25 .....        }
26 .....    ]
27 .....}
28 ]
```

Figura 8: Exemplo de uma instância pertencente ao conjunto de dados no formato aceito pela ferramenta.

A ferramenta busca remover a complexidade de interação do usuário com seus dados, fornecendo uma série de visualizações e filtros para que o entendimento das informações seja simples, intuitivo e facilitado. A Figura 9 apresenta a barra de navegação da ferramenta em sua versão 1.0. Nela estão contidos os links para as funcionalidades que em que o usuário poderá interagir, permitindo adicionar dados, visualizações, filtros, salvar o ambiente com as configurações atuais e carregar um ambiente de visualizações já configurado a partir de um arquivo previamente salvo.



Figura 9: Menu da ferramenta Geovis Explorer

Além do menu de interação com as funcionalidades, a ferramenta também possui um menu lateral onde são listados as camadas e os filtros da aplicação. Com a aba Camadas selecionada, o usuário é capaz de configurar a visibilidades das camadas de sobreposição no mapa por meio da marcação ou não do campo checkbox para a camada correspondente, conforme é demonstrado na Figura 10. Esta opção permite que os usuários montem um ambiente totalmente configurável, deixando visível apenas a combinação de visualizações que julgar mais relevante, facilitando a percepção de aspectos de seu estudo.

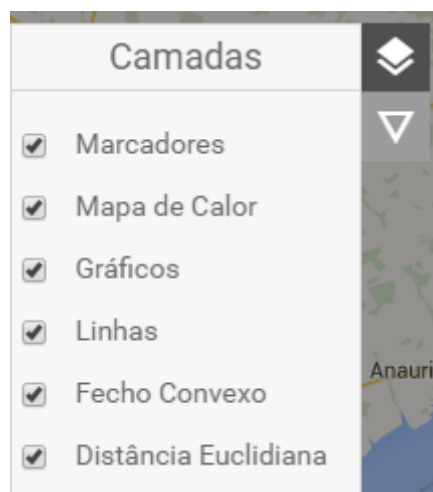


Figura 10: Menu lateral contendo as camadas de sobreposição do mapa.

No mesmo menu, com a segunda aba selecionada, é possível visualizar os filtros de dados que estão ativos no mapa e a quantidade de marcadores correspondentes que estão visíveis dentre o total de instâncias que passaram pelo filtro. Esta capacidade permite ao usuário ter uma noção estatística sobre as regras aplicadas por meio dos filtros em seu estudo e também na facilidade em lembrar quais regras estão ativas na visualização atual, conforme demonstrado na Figura 11.

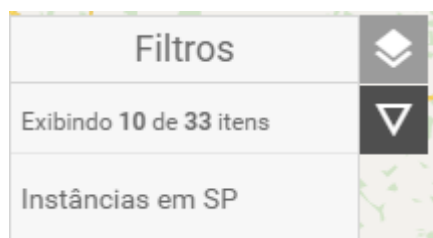


Figura 11: Menu lateral contendo os filtros ativos no mapa e a quantidade de instâncias que estão visíveis com as configurações atuais.

#### 4.1. Visualizações

No menu de funcionalidades, por meio do item “Visualizações” o usuário é capaz de adicionar ou remover visualizações em seu ambiente georreferenciado. Tais visualizações podem ser Mapas de Calor, Gráficos, Linhas de Similaridade, Linhas de Distância Euclidiana e o Fecho Convexo. Uma abordagem mais detalhada sobre cada um destes tipos de visualização está descrita nas subseções abaixo.

##### 4.1.1 Mapas de Calor

Na visualização de mapas de calor, a ferramenta adiciona sobreposições de cores para definir a localidade de concentração e aglomeração dos marcadores no mapa. A definição das áreas de alta concentração é representada por cores quentes (vermelho) e é gerada levando em consideração a localização geográfica de cada uma das instâncias. Utilizando esta visualização, o usuário é capaz de encontrar “pontos quentes” entre a localidade do conjunto de instâncias de seus dados, como demonstrado na Figura 12.

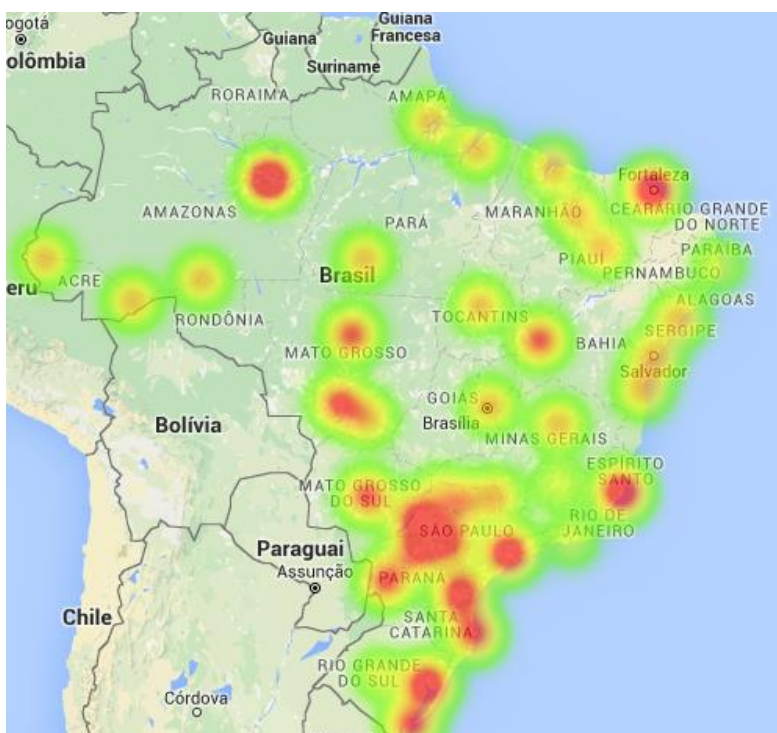


Figura 12: Visualização de mapas de calor gerada pela ferramenta.

#### 4.1.2 Gráficos

Outro recurso da ferramenta são as visualizações de gráficos. Por meio delas, o usuário é capaz de adicionar representações utilizando gráficos de linhas, barras ou de setores (pizza). Como os conjuntos de dados de entrada podem ser diversamente variáveis, o usuário é capaz de selecionar o atributo no qual o gráfico deverá levar em consideração para o cálculo e representação das estatísticas que serão visualizadas. Um exemplo do resultado desta funcionalidade, pode ser analisado por meio da Figura 13.

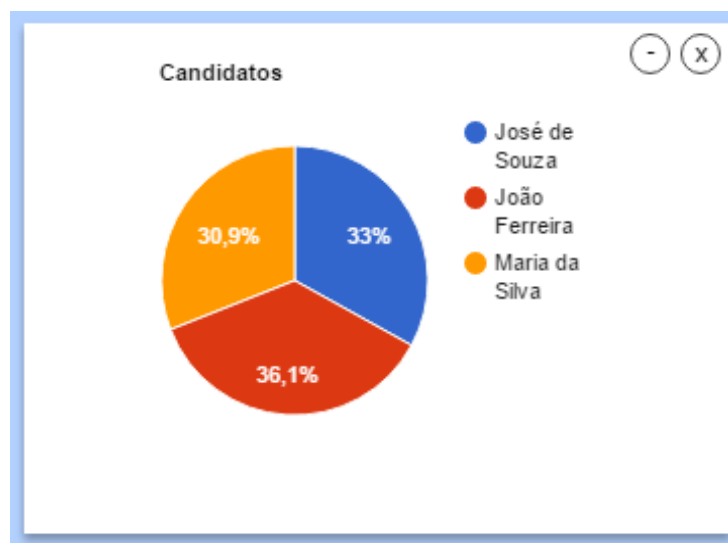


Figura 13: Gráfico de pizza representando a intensidade de votos nos candidatos a eleição.

#### 4.1.3 Linhas de Similaridade

Na visualização em linhas, a ferramenta é capaz de traçar linhas entre as instâncias similares com base em um atributo não numérico selecionado do conjunto de dados, facilitando a visualização de quais marcadores no mapa que se assemelham. Por exemplo na Figura 14, a visualização de linhas foi adicionada e o atributo escolhido foi estado. Assim quando o usuário clica em um marcador, a aplicação é capaz de encontrar e traçar linhas de similaridade entre todos os outros marcadores que identificam-se com o valor do atributo selecionado a partir da instância clicada.



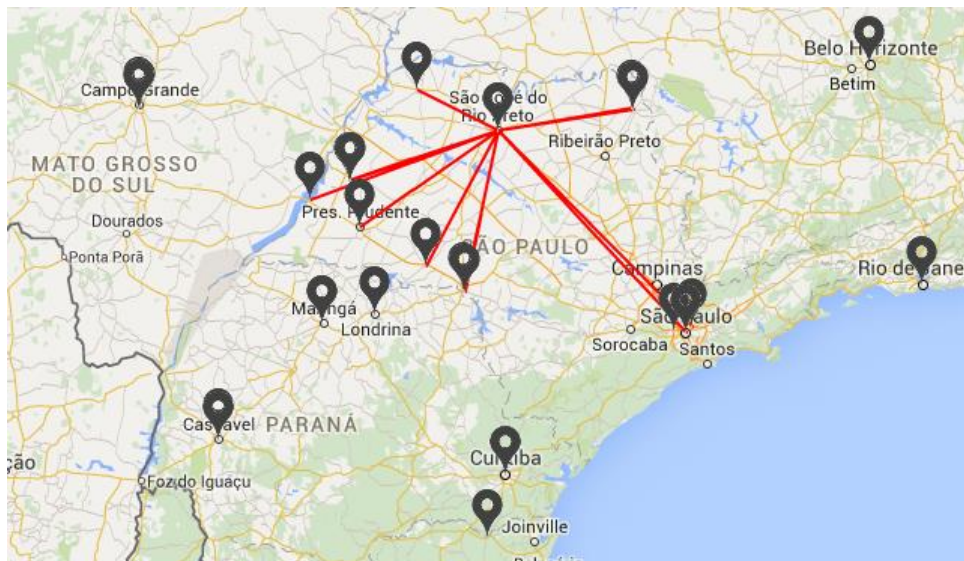


Figura 14: Linhas conectando instâncias baseando-se no atributo estado.

#### 4.1.4 Linhas de Distância Euclidiana

Na visualização da distância euclidiana, é possível traçar linhas entre as instâncias com maior proximidade baseado na aplicação do algoritmo da distância euclidiana para todos atributos numéricos presentes no conjunto de dados.

A distância euclidiana é de grande importância para aplicações na mineração de dados, pois facilita a seleção de pontos levando em consideração seus  $n$ -atributos. Ela representa a distância métrica entre dois pontos, podendo ser provada pela aplicação repetida do teorema de Pitágoras. Por exemplo, para os pontos  $n$ -dimensionais  $P = (p_1, p_2, \dots, p_n)$  e  $Q = (q_1, q_2, \dots, q_n)$ , a distância euclidiana é computada por meio da fórmula da Equação 1.

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (1)$$

Equação 1: Fórmula da distância euclidiana para dois pontos  $n$ -dimensionais.



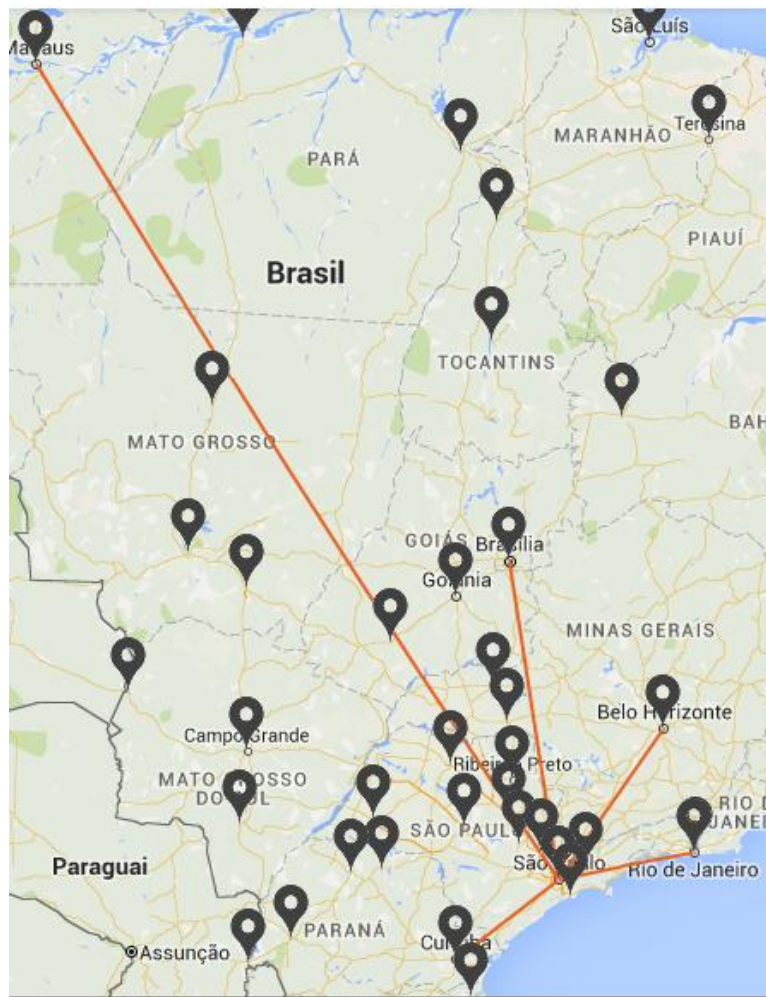


Figura 15: Linhas de distância euclidiana entre os cinco marcadores com maior proximidade da cidade de São Paulo.

Como é possível conferir na Figura 15, ao clicar na instância que representa a cidade de São Paulo, os marcadores que possuem seus atributos mais próximos (similares) da cidade de São Paulo por meio da aplicação do algoritmo da distância euclidiana são: Curitiba, Rio de Janeiro, Belo Horizonte, Brasília e Manaus. Os atributos levados em consideração foram: população, PIB e IDH.

#### 4.1.5 Fecho Convexo

Na visualização do fecho convexo, a ferramenta é capaz de delimitar uma área por meio de um desenho de um polígono que contém todas as instâncias visíveis do mapa, como demonstrado na Figura 16. Esta forma de visualização permite que o usuário tenha uma rápida percepção sobre a área que suas instâncias ocupam no espaço georreferenciado e facilita que ele delimite estas regiões geográficas.



Figura 16: Fecho convexo delimitando a área coberta dos marcadores.

## 4.2. Filtros

Por meio do item “Filtros” no menu de funcionalidades, a ferramenta permite adicionar critérios de seleção entre as instâncias do mapa. Para a aplicação desta funcionalidade, é necessário que o usuário configure alguns parâmetros como, por exemplo, o nome do filtro, a condição que o filtro levará em critério para seleção das instâncias, o atributo que será filtrado e também o valor comparativo para o atributo. Uma abordagem mais detalhada e exemplificada sobre o funcionamento dos filtros poderá ser analisada no Capítulo 5.

## 4.3. Dados

No menu de funcionalidades a opção “Dados” permite que o usuário selecione o arquivo que representa seu conjunto de dados. Uma vez que o conjunto seja selecionado e carregado, a aplicação adicionará as instâncias de marcadores no mapa e o usuário será capaz de iniciar o processo exploratório do estudo de seus dados.

## 4.4 Salvar

Uma vez que o conjunto de dados esteja carregado, e que o usuário tenha adicionado visualizações ou filtros no mapa, ele poderá salvar a representação atual por meio do item do menu de funcionalidades “Salvar”. Salvando a aplicação, ele não

necessariamente terá que repetir todo o processo de configuração para que seu ambiente esteja customizado. Ao salvar, um novo arquivo será gerado contendo uma estrutura no formato JSON, onde a aplicação deverá ser capaz de retornar o ambiente idêntico à sua última utilização.

#### **4.5 Carregar**

Já a opção “Carregar” no menu de funcionalidades, permite que o usuário selecione o conjunto de dados e o arquivo de configurações gerados. Assim a ferramenta deve ser capaz de retornar as visualizações e as representações anteriormente geradas no mapa.

### **5. Estudo de Caso**

Neste estudo, será analisado um conjunto de dados que contém atributos que representam informações socioeconômicas dos 120 municípios do Brasil com maior população. Além do atributo identificador e os atributos de coordenadas geográficas necessários para todas as instâncias do conjuntos de dados, o conjunto em questão possui os atributos: “cidade”, “estado” e “região”, sendo todos os três do tipo string; um atributo booleano chamado “capital” para indicar se a cidade é uma capital ou não; a “população” em milhões de pessoas, o “PIB” em bilhões de reais e o Índice de Desenvolvimento Humano (“IDH”), sendo estes do tipo numérico. As informações utilizadas neste conjunto de dados foram obtidas por meio do portal IBGE, onde as mesmas foram captadas durante o Censo Populacional em 2010.

```
{
  "id": 1,
  "geodata": {
    "lat": -23.550520,
    "lon": -46.633309
  },
  "infodata": [
    {
      "name": "cidade",
      "value": "São Paulo"
    },
    {
      "name": "estado",
      "value": "SP"
    },
    {
      "name": "região",
      "value": "SE"
    },
    {
      "name": "capital",
      "value": true
    },
    {
      "name": "população (em milhões de pessoas)",
      "value": 11.967
    },
    {
      "name": "PIB (em bilhões de reais)",
      "value": 570.706
    },
    {
      "name": "IDH",
      "value": 0.805
    }
  ]
},
```

Figura 17: Exemplo de uma instância pertencente ao conjunto de dados do estudo de caso.

Para exemplificar a eficiência da ferramenta, será demonstrado por meio de visualizações e filtros, sua capacidade de facilitar o estudo de grandes conjuntos de dados. Ao carregar o conjunto de dados na ferramenta, a representação obtida é demonstrada pela Figura 18.



Figura 18: Visualização contendo a marcação de todas as instâncias das cidades brasileiras pertencentes ao conjunto de dados.

Analisando a geolocalização dos marcadores, ainda não está tão visível onde está localizada a maior concentração dos maiores municípios no mapa. Para este fim, a ferramenta proporciona uma visualização bem interessante para tirar esse tipo de conclusão, os mapas de calor. Se adicionarmos a visualização de mapas de calor à aplicação e escondermos os marcadores no menu lateral, chegamos a uma representação de acordo com a Figura 19.







Figura 20: Representação gerada pelo filtro aplicado mostrando apenas as cidades do estado de São Paulo.

Com o filtro adicionado, caso o usuário deseje saber a quantidade de marcadores visíveis, ele pode expandir o menu lateral contendo as informações dos filtros. Assim é possível observar o texto com detalhes da quantidade de instâncias visíveis no mapa em relação ao total de instâncias do conjunto e também os filtros ativos na aplicação.

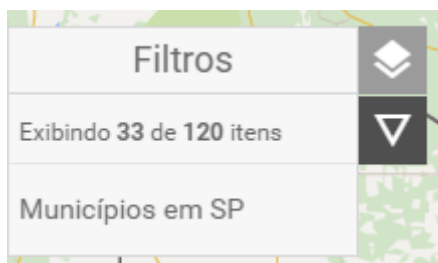


Figura 21: Lista de filtros no menu lateral e a informação da quantidade de instâncias visíveis.

Podemos também adicionar a visualização de ligação em linhas, levando em consideração o atributo "região". Neste caso quando o usuário clicar em um dos marcadores, serão traçadas linhas entre todas as instâncias que possuírem a região semelhante à região do município clicado.



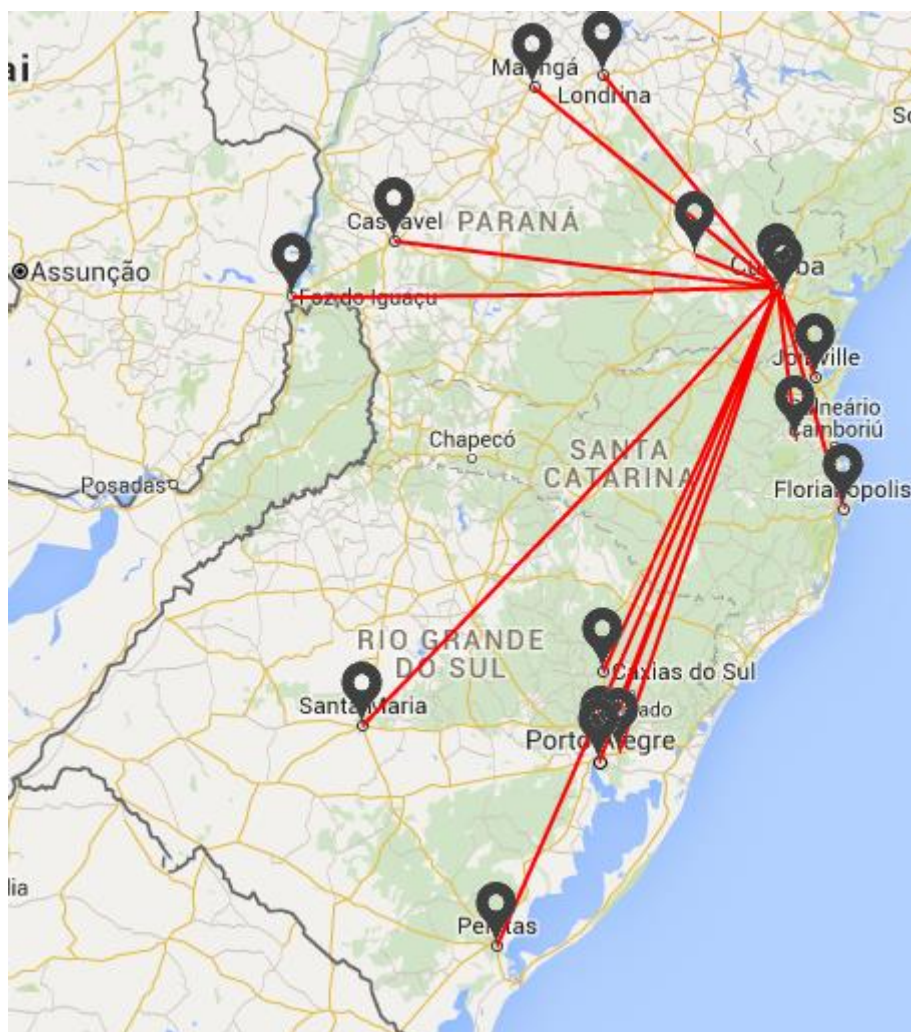


Figura 22: Linhas de ligação traçadas entre todas as instâncias da região sul após clicar sobre o marcador do município de Curitiba.

Caso o usuário julgue relevante ter uma abordagem estatística sobre os dados em estudo, ele poderá adicionar gráficos na visualização. Os gráficos neste caso, apresentam grande significância na análise das informações, pois permitem que certos tipos de conclusões sejam facilmente obtidas sem requerer muito esforço durante a análise. Para demonstrar tal facilidade, foram adicionados dois gráficos, um deles representado a quantidade de municípios pelo atributo “região” e o outro demonstrando a quantidade de municípios pelo atributo “estado”, conforme demonstrado pelas Figuras 23 e 24.



Figura 23: Gráfico de barras demonstrando a quantidade de ocorrências dos municípios levando em consideração o atributo “estado”. É notável que o estado que possui os principais municípios socioeconômicos é o estado de São Paulo, contendo trinta e três ocorrências.

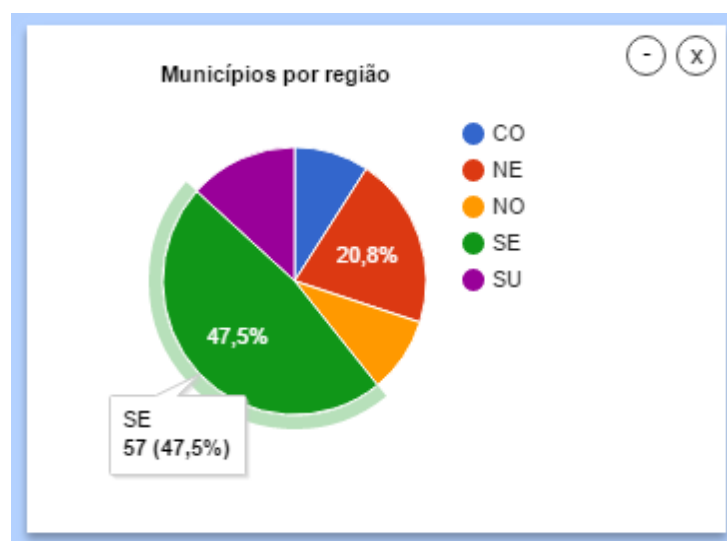


Figura 24: Gráfico de pizza ou setores demonstrando a porcentagem de municípios levando como critério o atributo “região”.

Outra espécie de gráfico, mas não menos importante, é o gráfico de linhas. Este gráfico é valioso para o estudo específico deste conjunto de dados, pois permite que o usuário encontre os municípios que possuem o Índice de Desenvolvimento Humano (IDH) abaixo da média, facilitando assim a tomada de decisões com um grau de confiança maior. A partir deste gráfico, que está demonstrado pela Figura 25, é possível combiná-lo com mais um filtro, permitindo uma melhor abstração do estudo e a chegada a conclusões bem relevantes.

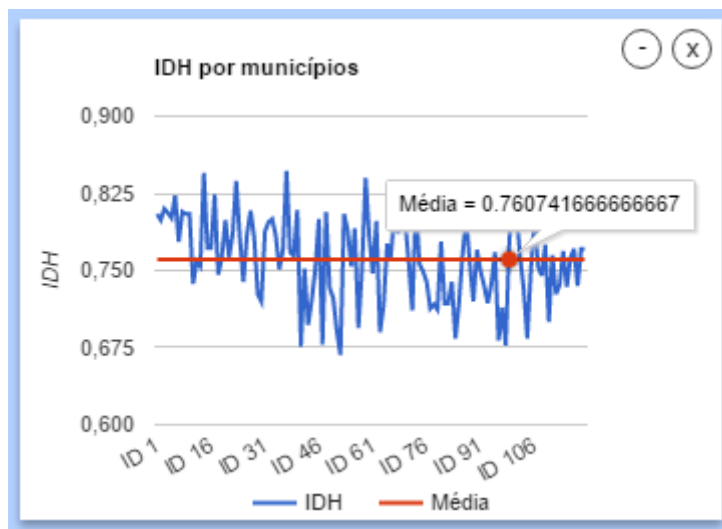


Figura 25: Gráfico de linhas representando o atributo IDH de cada município identificado pelo ID e a média do IDH de todos eles.

Sabendo que a média de IDH dos municípios do conjunto é de 0.760, podemos adicionar um filtro levando em consideração este atributo e este valor para identificar quais são os municípios que estão com o IDH abaixo da média. Mais uma vez utilizando os mapas de calor e combinando com o filtro, podemos facilmente analisar e identificar quais são estes municípios com o IDH abaixo da média, como demonstrado na Figura 26.

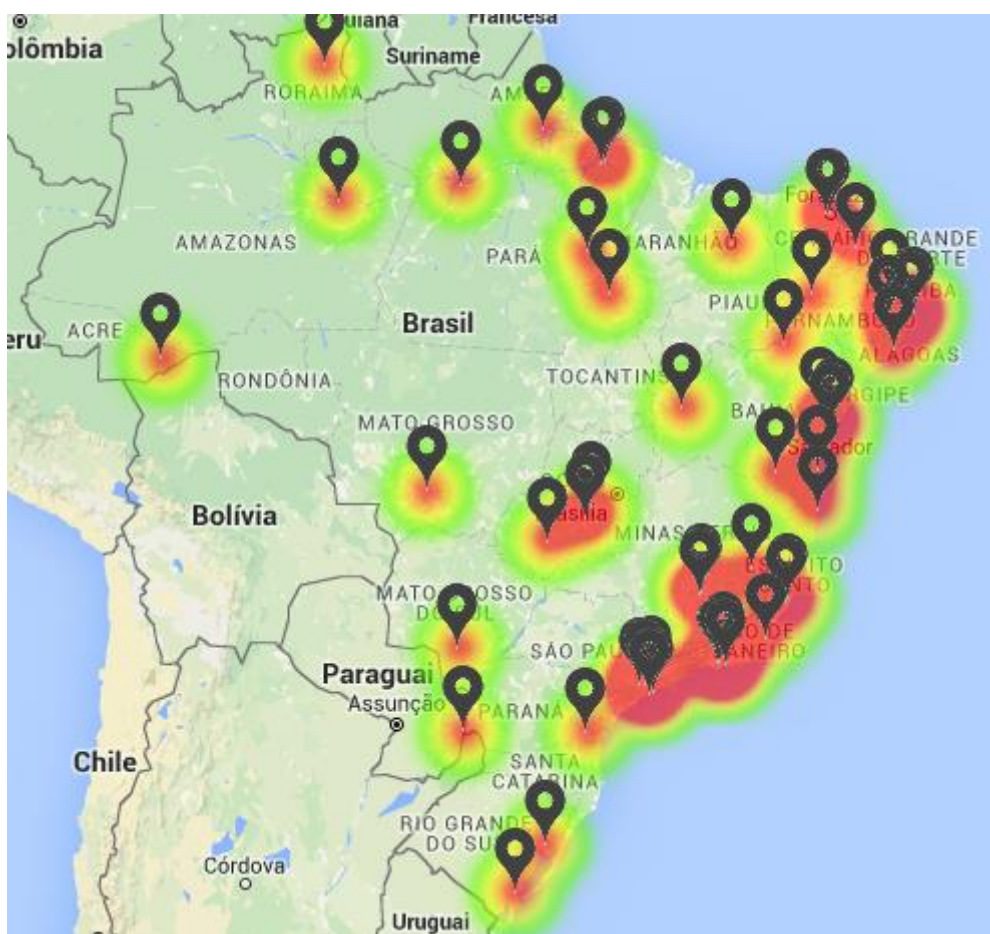


Figura 26: Visualização de mapas de calor combinada com o filtro de municípios que possuem o IDH abaixo da média entre os municípios brasileiros do conjunto de dados.

Este tipo de conclusão se mostra relevante para, por exemplo, órgãos governamentais, pois possibilitam uma conclusão rápida e simples sobre as principais áreas que necessitam de foco para que o índice de desenvolvimento humano evolua no território nacional.

Se desejarmos realizar um estudo ainda mais detalhado sobre, por exemplo, os municípios do estado de São Paulo que estão com o IDH abaixo da média, podemos adicionar um filtro para mostrar apenas os municípios deste estado. Com o objetivo de intensificar ainda mais a capacidade de abstração proporcionada pela ferramenta, adicionamos também a visualização do fecho convexo, que facilita a identificação da área que envolve tais municípios como demonstrado na Figura 27.

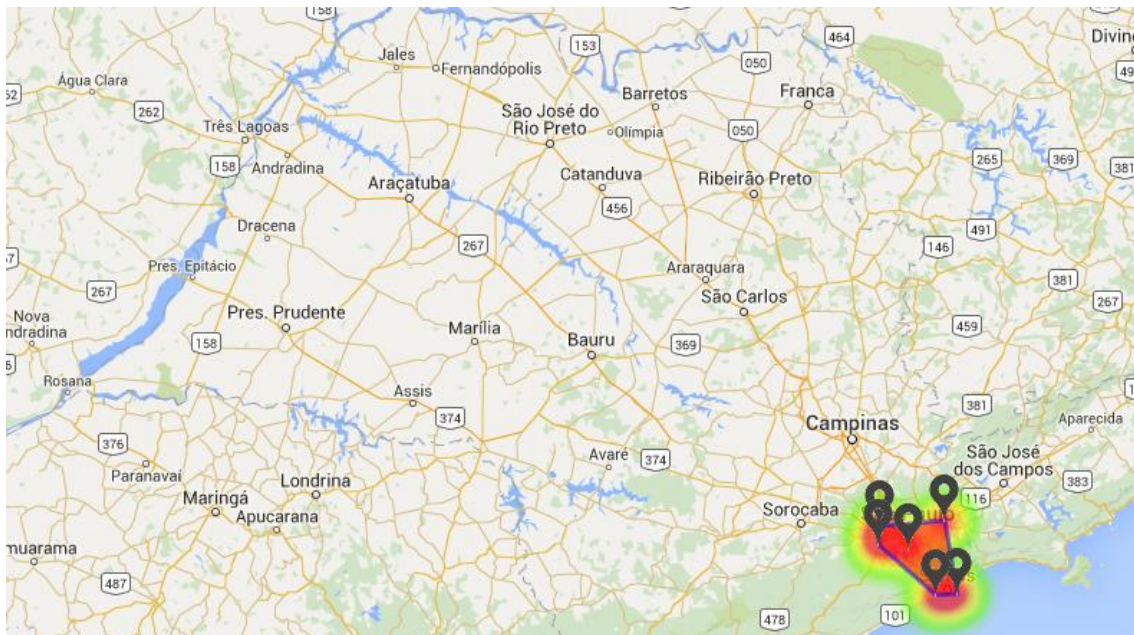


Figura 27: Área que estão contidos os municípios com IDH abaixo da média no estado de São Paulo.

Analisando a Figura 27, é nítido que dentre os municípios com maior população no estado de São Paulo, os que possuem o IDH abaixo da média estão na região litorânea do estado delimitada pelo polígono convexo.

A ferramenta fornece também outra visualização importante para estudos de relacionamento e proximidade entre as instâncias, as linhas de distância euclidiana. Mantendo o mapa de calor e removendo os demais filtros e visualizações, podemos adicionar a visualização “Distância Euclidiana” preenchendo o número três no campo “Número de Instâncias” para que o algoritmo selecione as três instâncias mais próximas do marcador clicado. Ao selecionar um marcador como, por exemplo, o que representa a instância da cidade de São Paulo, são traçadas linhas entre as três instâncias numericamente mais próximas. Por meio destas linhas, podemos rapidamente analisar quais são as instâncias que possuem maior proximidade entre seus atributos numéricos, como demonstrado pela Figura 28.



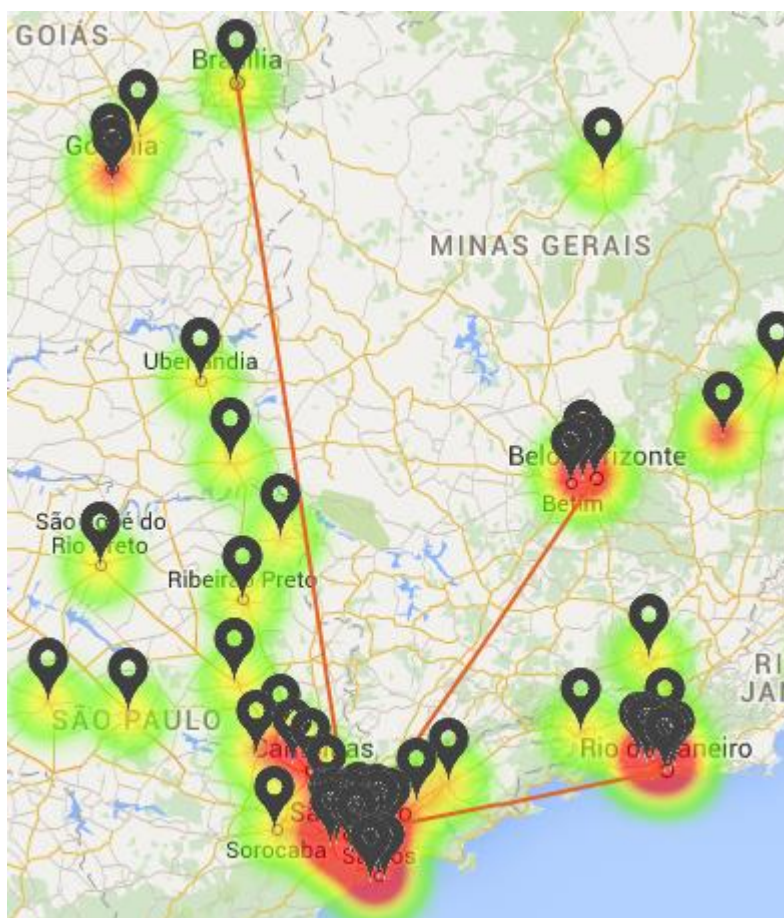


Figura 28: As três cidades com maior proximidade de São Paulo, utilizando a distância euclidiana.

Analisando as instâncias que foram selecionadas, podemos afirmar que para a cidade de São Paulo, as três cidades numericamente mais semelhantes à ela são: Brasília, Belo Horizonte e Rio de Janeiro. Para tal conclusão, os valores dos atributos numéricos do conjunto de dados População, PIB e IDH foram levados em consideração. Se fossemos analisar os dados brutos no formato texto para chegarmos a este conhecimento, certamente seria uma tarefa bem mais trabalhosa. Esta forma de visualização além de ter uma importante aplicação na mineração dos dados, tem grande poder em facilitar a percepção de conclusões em estudos sobre os conjuntos de dados com dados estatísticos.

## 6. Conclusões

Neste trabalho, foram apresentados a ferramenta e suas facilidades, com uma abordagem de estudo sobre os objetivos que ela propõe. O foco principal foi enfatizar a capacidade que a ferramenta proporciona no estudo aprofundado sobre grandes conjuntos de dados, fornecendo uma análise intuitiva para a detecção de padrões e levantamento de hipóteses sobre os dados. Dentre os recursos utilizados que facilitam a

realização de estudos foram, a integração de representações gráficas ao ambiente do mapa, a aplicação filtros de atributos do conjunto de dados, a implementação de técnicas de mineração de dados e a utilização de mapas georreferenciados. Como demonstrado na Seção 5, o principal objetivo da ferramenta foi possibilitar o estudo intuitivo sobre conjuntos de dados complexos, basicamente convertendo informações em conhecimento para o usuário.

A ferramenta Geovis Explorer já fornece várias técnicas e abordagens para facilitar o estudo sobre os dados, porém muitas melhorias ainda poderão ser desenvolvidas. Novas funcionalidades de busca e eficiência na performance poderão ser analisadas para trabalhos futuros, assim como uma melhoria na usabilidade da interface da ferramenta e a criação de uma versão responsiva a dispositivos mobile. Por se tratar de um projeto open source, outra possível melhoria seria a criação de uma documentação específica para usuários que desejam contribuir por meio da implementação de novas visualizações e a aplicação de melhorias no código da ferramenta.



## Referências

[Medeiros, Cleyber Nascimento; Aragão, Mário Cesar Alves] Disponibilização e análise de informações georreferenciadas utilizando um SIG-WEB

Disponível em: [http://www.ipece.ce.gov.br/publicacoes/textos\\_discussao/TD\\_78.pdf](http://www.ipece.ce.gov.br/publicacoes/textos_discussao/TD_78.pdf)

[Hübner, Cleice Edinara; Oliveira, Francisco Henrique], Gestão da Geoinformação em Implementações Multiusuários.

Disponível em:

[http://www.geolab.faed.udesc.br/publicacoes/Cleice/cobracos2008\\_1.pdf](http://www.geolab.faed.udesc.br/publicacoes/Cleice/cobracos2008_1.pdf)

[SALAZAR, F. 2012] Um estudo sobre o papel de medidas de similaridade em visualização de coleções de documentos. [S.l.], 2012.

[Bugs, Geisa; Reis, Antônio Tarcísio] Participação popular no planejamento urbano: mapas interativos e ferramentas SIG na internet e aspectos cognitivos.

Disponível em:

<http://unuhospedagem.com.br/revista/rbeur/index.php/anais/article/view/2843/2780>

[Card, S. K.; Mackinlay, J. D.; Shneiderman, B. Readings] in information visualization: using vision to think San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., p. 463-464, 1999.

[Henry, N.; Fekete, J.-D.; McGuffin, M. J.] Nodetrix: a hybrid visualization of social networks. IEEE Transactions on Visualization and Computer Graphics, v. 13, n. 6, p. 1302-1309, 2007.

Disponível em: <http://dx.doi.org/10.1109/TVCG.2007.70582>

[Munster, E.] A theme landscape for tagged data. In: Proceedings of the 2010 14th International Conference Information Visualisation, IV '10, Washington, DC, USA: IEEE Computer Society, 2010,p. 134-139 (IV '10, ).

Disponível em: <http://dx.doi.org/10.1109/IV.2010.29>

[Oliveira, M. C. F.;] Levkowitz, H. From Visual Data Exploration to Visual Data Mining: A Survey. IEEE Transactions on Visualization and Computer Graphics, v. 9, n. 3, p. 378-394, 2003.

[Medeiros, C. N; Aragão, M. C. A.; Gomes, D. D. M.; Albuquerque, E. L. S. ] Utilização de SIG-WEB usando software livre para disponibilização de dados georreferenciados na internet: Caso do Sistema Ceará em Mapas Interativos

Disponível em: [http://www.ipece.ce.gov.br/publicacoes/textos\\_discussao/TD\\_78.pdf](http://www.ipece.ce.gov.br/publicacoes/textos_discussao/TD_78.pdf)

[Michael Khoo, Lily Rozaklis, Catherine Hall, Diana Kusunoki, Michael Rehrig, 2014] Heat Map Visualizations of Seating Patterns in an Academic Library

Disponível em:

[https://www.ideals.illinois.edu/bitstream/handle/2142/47285/274\\_ready.pdf?sequence=2](https://www.ideals.illinois.edu/bitstream/handle/2142/47285/274_ready.pdf?sequence=2)

[Vasco Furtado, José Eurico, Leonardo Ayres, Rafael Alves, Marcos de Oliveira, 2008] WikiCrimes - Um Sistema Colaborativo para Mapeamento Criminal

Disponível em:

[http://www.infobrasil.inf.br/userfiles/WikiCrimes\\_Um%20Sistema%20Colaborativo%20para%20Mapeamento%20Criminal.pdf](http://www.infobrasil.inf.br/userfiles/WikiCrimes_Um%20Sistema%20Colaborativo%20para%20Mapeamento%20Criminal.pdf)

[Marcello Martinelli, 2005] Os Mapas da Geografia

Disponível em:

[http://www2.fct.unesp.br/docentes/geo/raul/cartografia\\_tematica/leitura%202/1-MAPAS%20DA%20GEOGRAFIA.pdf](http://www2.fct.unesp.br/docentes/geo/raul/cartografia_tematica/leitura%202/1-MAPAS%20DA%20GEOGRAFIA.pdf)