

PGCOMP - Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal da Bahia (UFBA)  
Av. Milton Santos, s/n - Ondina  
Salvador, BA, Brasil, 40170-110

<https://pgcomp.ufba.br>  
[pgcomp@ufba.br](mailto:pgcomp@ufba.br)

Visualizações multidimensionais são representações gráficas que atuam na apresentação de conjuntos de dados multidimensionais ou multivariados. Entretanto, muitas dessas técnicas apresentam pouca escalabilidade visual à medida que as dimensões aumentam, exigindo um espaço visual cada vez maior para acomodar todo o conjunto de dados. Uma forma de contornar esse problema é utilizar técnicas de Projeções Multidimensionais que realizam redução de dimensionalidade e buscam preservar os padrões dos dados originais no espaço projetado. Durante esse processo, é altamente provável que erros e distorções estejam presentes nos *layouts* de projeção. Um desafio na pesquisa desse tema é mensurar e visualizar os erros e distorções incorporados nos mapeamentos das projeções. Nesse sentido, este trabalho desenvolveu o MultiVisD3, uma abordagem de *Visual Analytics* para analisar a qualidade de Projeções Multidimensionais. Esta abordagem compreende um *dashboard* como um painel visual interativo contendo múltiplas visualizações coordenadas para facilitar a compreensão e comparação de métricas de erro entre duas projeções. Outras técnicas de visualização como *Treemap* e visão tabular apoiam a extração de informações sobre a qualidade das projeções. Por fim, a abordagem proposta foi submetida a uma avaliação com usuário utilizando conjuntos de dados conhecidos para analisar aspectos de funcionalidade e usabilidade. Os resultados mostraram que os usuários avaliaram positivamente a abordagem proposta neste trabalho. Mais de 90% dos participantes executaram corretamente oito das nove tarefas que lhes foram solicitadas e mais de 90% deles aprovaram os aspectos de usabilidade relacionados às interações e disposição das visualizações.

Palavras-chave: visualização de dados; projeção multidimensional; múltiplas visões coordenadas; métricas de qualidade de projeção.

# MultiVisD3: uma abordagem *Visual Analytics* para Análise da Qualidade de Projeções Multidimensionais Utilizando Múltiplas Visões Coordenadas

José Augusto Duarte Gomes

Dissertação de Mestrado

Universidade Federal da Bahia

Programa de Pós-Graduação em  
Ciência da Computação

Setembro | 2023

MSC | 167 | 2023

MultiVisD3: uma abordagem *Visual Analytics* para Análise da Qualidade de Projeções Multidimensionais Utilizando Múltiplas Visões Coordenadas

José Augusto Duarte Gomes

UFBA







Universidade Federal da Bahia  
Instituto de Computação

Programa de Pós-Graduação em Ciência da Computação

**MULTIVISD3: UMA ABORDAGEM *VISUAL*  
*ANALYTICS* PARA ANÁLISE DA  
QUALIDADE DE PROJEÇÕES  
MULTIDIMENSIONAIS UTILIZANDO  
MÚLTIPLAS VISÕES COORDENADAS**

José Augusto Duarte Gomes

DISSERTAÇÃO DE MESTRADO

Salvador  
29 de setembro de 2023



JOSÉ AUGUSTO DUARTE GOMES

**MULTIVISD3: UMA ABORDAGEM *VISUAL ANALYTICS* PARA  
ANÁLISE DA QUALIDADE DE PROJEÇÕES  
MULTIDIMENSIONAIS UTILIZANDO MÚLTIPLAS VISÕES  
COORDENADAS**

Esta Dissertação de Mestrado foi apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Orientador: Danilo Barbosa Coimbra

Salvador

29 de setembro de 2023

Sistema de Bibliotecas - UFBA

G633 Gomes, José Augusto Duarte.  
MultiVisD3: uma abordagem *Visual Analytics* para Análise da Qualidade de Projeções Multidimensionais Utilizando Múltiplas Visões Coordenadas / José Augusto Duarte Gomes – Salvador, 2023.  
102 f.

Orientador: Prof. Dr. Danilo Barbosa Coimbra.  
Dissertação (Mestrado) – Universidade Federal da Bahia, Instituto de Computação, 2023.

1. Sistema de Dados. 2. Computação.. I. Coimbra, Danilo Barbosa. II. Universidade Federal da Bahia. Instituto de Computação. III Título.

CDU – 004



---

*“MultiVisD3: uma abordagem Visual Analytics para Análise da Qualidade de Projeções Multidimensionais Utilizando Múltiplas Visões Coordenadas”*

José Augusto Duarte Gomes

Dissertação apresentada ao Colegiado do Programa de Pós-Graduação em Ciência da Computação na Universidade Federal da Bahia, como requisito parcial para obtenção do Título de Mestre em Ciência da Computação.

**Banca Examinadora**

---

Prof. Dr. Danilo Barbosa Coimbra (Orientador)



Documento assinado digitalmente

BRUNO PEREIRA DOS SANTOS

Data: 15/10/2023 22:21:38-0300

Verifique em <https://validar.iti.gov.br>

---

Prof. Dr. Bruno Pereira dos Santos - PGCOMP

---

Prof. Dr. Rafael Messias Martins - (Linnaeus University)

---



## AGRADECIMENTOS

Agradeço a Deus por ser o meu Alpha e meu Ômega nesse percurso!

Agradeço a Universidade Federal da Bahia (UFBA), ao Instituto de Computação (IC) e ao Programa de Pós-Graduação em Ciência da Computação (PGCOMP) por me proporcionar essa oportunidade única no âmbito acadêmico e profissional.

Agradeço ao Prof. Dr. Danilo Barbosa Coimbra por cumprir com excelência a missão de orientar, pelos sábios e incansáveis direcionamentos e condução ao longo de toda trajetória do mestrado.

Agradeço aos Professores do PGCOMP pelos compromissos com a qualidade da docência e da ciência.

Agradeço aos colegas do Grupo de Pesquisa de Visualização pelo compartilhamento de saberes.

Agradeço aos colegas do Laboratório de Pesquisa de Pós-Graduação do PGCOMP pelas experiências enriquecedoras nesse processo.

Agradeço aos colegas Robson Silva, Marcus Freire, Fernanda Eustáquio, Francisco Fatore, Sávio Santana, Thiago Lima, Janara Nascimento, Sérgio Andrade e Elias Fonseca pelo incentivo e apoio em momentos fundamentais dessa caminhada.

Agradeço ao Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS), em especial aos colegas do Núcleo de Produção de Dados.

Agradeço a minha família pelo apoio incondicional e motivação durante essa imersão, nas pessoas de minha mãe Ilza, minha irmã Michele, seu esposo e filhas, minha irmã Olávia, meu filho Levi.

Agradeço minha namorada Umelda Fagundes pelo carinho e cuidado nos altos e baixos desse trajeto.



## RESUMO

Visualizações multidimensionais são representações gráficas que atuam na apresentação de conjuntos de dados multidimensionais ou multivariados. Entretanto, muitas dessas técnicas apresentam pouca escalabilidade visual à medida que as dimensões aumentam, exigindo um espaço visual cada vez maior para acomodar todo o conjunto de dados. Uma forma de contornar esse problema é utilizar técnicas de Projeções Multidimensionais que realizam redução de dimensionalidade e buscam preservar os padrões dos dados originais no espaço projetado. Durante esse processo, é altamente provável que erros e distorções estejam presentes nos *layouts* de projeção. Um desafio na pesquisa desse tema é mensurar e visualizar os erros e distorções incorporados nos mapeamentos das projeções. Nesse sentido, este trabalho desenvolveu o MultiVisD3, uma abordagem de *Visual Analytics* para analisar a qualidade de Projeções Multidimensionais. Esta abordagem compreende um *dashboard* como um painel visual interativo contendo múltiplas visualizações coordenadas para facilitar a compreensão e comparação de métricas de erro entre duas projeções. Outras técnicas de visualização como *Treemap* e visão tabular apoiam a extração de informações sobre a qualidade das projeções. Por fim, a abordagem proposta foi submetida a uma avaliação com usuário utilizando conjuntos de dados conhecidos para analisar aspectos de funcionalidade e usabilidade. Os resultados mostraram que os usuários avaliaram positivamente a abordagem proposta neste trabalho. Mais de 90% dos participantes executaram corretamente oito das nove tarefas que lhes foram solicitadas e mais de 90% deles aprovaram os aspectos de usabilidade relacionados às interações e disposição das visualizações.

### **Palavras-chave:**

visualização de dados; projeção multidimensional; múltiplas visões coordenadas; métricas de qualidade de projeção.



## ABSTRACT

Multidimensional visualizations are graphical representations that act in the presentation of multidimensional or multivariate datasets. However, many of these techniques exhibit poor visual scalability as dimensions increase, requiring an increasingly larger visual space to accommodate the entire dataset. One way to get around this problem is to use Multidimensional Projections techniques that perform dimensionality reduction and seek to preserve the patterns of the original data in the projected space. During this process, it is highly likely that errors and distortions will be present in the projection layouts. A challenge in researching this topic is measuring and visualizing the errors and distortions incorporated in the projection mappings. In this sense, this work developed MultiVisD3 a Visual Analytics approach to analyze the quality of Multidimensional Projections. This approach comprises a dashboard as an interactive visual panel containing multiple coordinated views to facilitate understanding and comparison of error metrics between two projections. Other visualization techniques such as Treemap and tabular view support the extraction of information about the quality of projections. Finally, the proposed approach was subjected to a user evaluation using known datasets to analyze aspects of functionality and usability. The results showed that users positively evaluated the approach proposed in this work. More than 90% of participants correctly performed eight of the nine tasks asked for them and more than 90% of them approved the usability aspects related to interactions and visualization arrangement.

**Keywords:**

data visualization; multidimensional projection; multiple coordinated views; projection quality metrics.



# SUMÁRIO

<b>Capítulo 1—Introdução</b>	<b>1</b>
1.1 Contextualização . . . . .	1
1.2 Motivação . . . . .	2
1.3 Problema de Pesquisa . . . . .	2
1.4 Objetivos . . . . .	2
1.5 Questões de Pesquisa . . . . .	3
1.6 Estrutura do Trabalho . . . . .	3
<b>Capítulo 2—Visualizações Multidimensionais</b>	<b>5</b>
2.1 Considerações Iniciais . . . . .	5
2.2 Visualização de Dados . . . . .	5
2.2.1 Visualização Multidimensional . . . . .	6
2.2.1.1 Projeções Multidimensionais . . . . .	8
2.2.1.2 Métricas de Erro . . . . .	16
2.2.2 Visualização Hierárquica . . . . .	17
2.3 Considerações Finais . . . . .	18
<b>Capítulo 3—Múltiplas Visões Coordenadas</b>	<b>20</b>
3.1 Considerações Iniciais . . . . .	20
3.2 Múltiplas Visões Coordenadas . . . . .	20
3.3 Múltiplas Visões . . . . .	21
3.4 Coordenação . . . . .	23
3.5 Interação . . . . .	28
3.5.1 Classes de Interação . . . . .	28
3.5.2 Estilos de Interação . . . . .	31
3.5.3 Exemplos de Interação em Múltiplas Visões Coordenadas . . . . .	31
3.6 Considerações Finais . . . . .	33
<b>Capítulo 4—Trabalhos Relacionados</b>	<b>35</b>
4.1 Considerações Iniciais . . . . .	35
4.2 Visualização Multidimensional . . . . .	35
4.3 Domínio de Aplicação de Dados . . . . .	36
4.4 Discussão . . . . .	42
4.5 Considerações Finais . . . . .	49

<b>Capítulo 5—MultiVisD3: <i>Projection Error Analyzer</i></b>	51
5.1 Considerações Iniciais . . . . .	51
5.2 Metodologia . . . . .	51
5.3 Tecnologias Utilizadas . . . . .	55
5.4 Descrição do MultiVisD3: <i>Projection Error Analyzer</i> . . . . .	57
5.4.1 Layout . . . . .	58
5.4.2 Recursos Interativos e Coordenações entre as Visões . . . . .	62
5.5 Considerações Finais . . . . .	64
<b>Capítulo 6—Avaliação</b>	67
6.1 Considerações Iniciais . . . . .	67
6.2 Cenário de Uso . . . . .	67
6.3 Avaliação com Usuários . . . . .	72
6.3.1 Metodologia . . . . .	72
6.3.1.1 Dispositivos de Coleta de Dados . . . . .	73
6.3.1.2 Participantes da Pesquisa . . . . .	74
6.3.1.3 Objetivos do Processo de Avaliação e Tarefas de Exploração . . . . .	74
6.3.1.4 Questões de Usabilidade . . . . .	75
6.3.2 Resultados . . . . .	76
6.3.3 Discussão . . . . .	81
6.4 Considerações Finais . . . . .	82
<b>Capítulo 7—Conclusão</b>	84
7.1 Contribuições . . . . .	84
7.2 Limitações . . . . .	85
7.3 Trabalhos Futuros . . . . .	85
<b>Referências Bibliográficas</b>	88
<b>Apêndice A—Formulário sobre Conhecimento Prévio das Técnicas de Visualização</b>	95
<b>Apêndice B—Formulário com Tarefas e Questões de Usabilidade</b>	111
<b>Apêndice C—Aprovação do Conselho de Ética em Pesquisa da UFBA</b>	137
<b>Apêndice D—Aprovação do Conselho de Ética em Pesquisa da FIOCRUZ-BA</b>	145



## LISTA DE FIGURAS

2.1	Visualizações Multidimensionais (A - Coordenadas Paralelas; B - Matriz de Gráficos de Dispersão; C - Gráfico de Radar). . . . .	8
2.2	Mapa de documentos produzido a partir de Projeção Multidimensional. Adaptado de Paulovich, Oliveira e Minghim (2007). . . . .	9
2.3	Visualizações Hierárquicas (A - Diagrama de Árvore; B - <i>Treemap</i> ; C - <i>Sunburst</i> ). . . . .	18
3.1	Justaposição de dois gráficos de dispersão A e B, lado a lado. Adaptado de Munzner (2014). . . . .	22
3.2	Sobreposição de duas visualizações em camadas. Dois gráficos de dispersão A e B empilhados, resultando em uma única visualização C. Adaptado de Munzner (2014). . . . .	22
3.3	Combinação da coordenação Selecionar itens de forma bidirecional. Ao selecionar itens em qualquer uma visão, os itens da outra visão também são selecionados. Adaptado de North e Shneiderman (1997). . . . .	24
3.4	Combinação da coordenação Navegar na visualização de forma bidirecional. Ao movimentar qualquer uma das visões, o deslocamento para a mesma direção será refletido na outra visão. Adaptado de (NORTH; SHNEIDERMAN, 1997). . . . .	25
3.5	Combinação da coordenação selecionar itens de forma bidirecional. Os usuários podem selecionar itens no contexto de uma visão para localizar e visualizar informações correspondentes em outra visão. Adaptado de North e Shneiderman (1997). . . . .	25
3.6	Exemplo de múltiplas visualizações compartilhando a codificação visual. Neste caso, o canal de cor das visões está conectado. Adaptado de Munzner (2014). . . . .	26
3.7	Exemplo de três pares de visualizações compartilhando: A e B toda base de dados; C e D mesmo subconjunto do <i>dataset</i> ; E e F subconjuntos diferentes da coleção de dados. Adaptado de Munzner (2014). . . . .	27
3.8	Exemplo de navegação compartilhada entre MVC. Adaptado de Munzner (2014). . . . .	27
3.9	Exemplos de possíveis combinações de compartilhamento de codificação visual e de dados. Adaptado de Munzner (2014). . . . .	27
4.1	Exemplo de interface com Múltiplas Visões Coordenadas. Adaptado de Godinho et al. (2007). . . . .	36

4.2	Tela de um cenário de exploração da aplicação WebPrisma. Adaptado de Almeida et al. (2009). . . . .	37
4.3	Interface da aplicação PRISMA explorando a base de dados carros com Múltiplas Visões Coordenadas. Adaptado de Godinho et al. (2007). . . . .	38
4.4	Lado esquerdo, Múltiplas Visões Coordenadas na mesma tela. Lado direito, visualizações coordenadas em telas separadas. Adaptado de Maciel et al. (2008). . . . .	39
4.5	Interface da aplicação ComVis com Múltiplas Visões Coordenadas. Adaptado de Matkovic et al. (2008). . . . .	40
4.6	Múltiplas Visões Coordenadas da ferramenta PRISMA explorando dados de tráfego de rede. Adaptado de Godinho et al. (2007). . . . .	41
4.7	Múltiplas Visões Coordenadas aplicadas a geo-visualização. Adaptado de Jern et al. (2007). . . . .	41
4.8	Múltiplas Visões Coordenadas representando graficamente duas bases de dados do domínio de indicadores sociais e econômicos. Adaptado de Guimarães et al. (2009). . . . .	42
4.9	Projeção Multidimensional (lado esquerdo) exibindo dados de sensores e Coordenadas Paralelas (lado direito) exibindo os mesmos dados sem redução de dimensionalidade. Adaptado de Aoki et al. (2013). . . . .	43
4.10	Mapas de similaridade coordenados usando Projeção Multidimensional. Adaptado de Eler (2011). . . . .	44
4.11	Exemplo da técnica de interação <i>linking &amp; brushing</i> realçando itens de dados nas visões coordenadas: Lado esquerdo, Gráfico de Dispersão, <i>Treemap</i> e Coordenadas Paralelas. Adaptado de Almeida et al. (2009); Lado direito, Matriz de Gráficos de Dispersão, RadViz e Coordenadas Paralelas. Adaptado de Pillat e Freitas (2006). . . . .	46
4.12	Interface da aplicação mVis, um visualizador de dados multidimensionais, que combina Múltiplas Visões Coordenadas com recursos de Aprendizado de Máquina. Adaptado de Chegini et al. (2019). . . . .	48
4.13	Exemplo de execução do Vistiles em diferentes dispositivos móveis. Adaptado de Langner, Horak e Dachsel (2018). . . . .	49
5.1	Pipeline da MultiVisD3. . . . .	52
5.2	Representação dos valores de erro normalizados por meio de um diagrama de árvore e um <i>Treemap</i> . . . . .	55
5.3	Tecnologias Utilizadas no MultiVisD3: <i>Projection Error Analyzer</i> . . . . .	56
5.4	<i>Layout</i> da Visão ‘Erro Agregado’. . . . .	58
5.5	Botão ‘Menu’ e suas Opções; Botão ‘Ajuda’. . . . .	60
5.6	<i>Layout</i> da Visão ‘Classe’. . . . .	61
5.7	<i>Layout</i> da Visão ‘Preservação de Vizinhança’. . . . .	62
5.8	Menu da Visão ‘Preservação de Vizinhança’. . . . .	62
6.1	Identificação de similaridade visual de um par de instâncias entre técnicas de projeção multidimensional. . . . .	68

6.2	Similaridade visual de grupos de instâncias entre técnicas de PM. . . . .	69
6.3	Verificação das instâncias por intervalos de erro agregado normalizados. .	69
6.4	Verificação das instâncias por valor de erro agregado. . . . .	70
6.5	Exploração do Erro Agregado de Projeções Multidimensionais na perspectiva de Classe. . . . .	71
6.6	Verificação da preservação de vizinhança nas PMs. . . . .	71
6.7	Fluxo da Avaliação com Usuário. . . . .	73
6.8	Perfil dos Participantes. . . . .	77
6.9	Conhecimentos Prévios das Técnicas de Visualização Utilizadas. . . . .	78
6.10	Respostas das Tarefas de Exploração. . . . .	79
6.11	Nível de Complexidade das Tarefas de Exploração. . . . .	80
6.12	Nível de Complexidade para Manusear o MultiVisD3: <i>Projection Error Analyzer</i> . . . . .	80
6.13	Percepção de Usabilidade do Usuário. . . . .	81



## **LISTA DE TABELAS**

6.1	Objetivos do Processo de Avaliação, Tarefas de Exploração e Suas Descrições.	75
-----	--	----



## LISTA DE SIGLAS

<b>CP</b> Coordenadas Paralelas . . . . .	7
<b>PMs</b> Projeções Multidimensionais . . . . .	8
<b>MDS</b> <i>Metric Multidimensional Scaling</i> . . . . .	12
<b>PCA</b> <i>Principal Component Analysis</i> . . . . .	12
<b>T-SNE</b> <i>t-Stochastic Neighborhood Embedding</i> . . . . .	12
<b>UMAP</b> <i>Uniform Manifold Approximation and Projection</i> . . . . .	12
<b>LAMP</b> <i>Local Affine Multidimensional Projection</i> . . . . .	12
<b>LSP</b> <i>Least Square Projection</i> . . . . .	12
<b>VA</b> Vizinho Ausente . . . . .	16
<b>VF</b> Vizinho Falso . . . . .	16
<b>VV</b> Vizinhos Verdadeiros . . . . .	16
<b>IBGE</b> Instituto Brasileiro de Geografia e Estatística . . . . .	41
<b>IDH</b> Índice de Desenvolvimento Humano . . . . .	41
<b>IC</b> Instituto de Computação . . . . .	74

<b>CIDACS</b> Centro de Integração de Dados e Conhecimentos para Saúde . . . . .	74
<b>FIOCRUZ</b> Fundação Osvaldo Cruz . . . . .	74
<b>CEPs</b> Comitês de Ética em Pesquisa . . . . .	72



## INTRODUÇÃO

### 1.1 CONTEXTUALIZAÇÃO

A utilização massiva de recursos provenientes dos constantes avanços tecnológicos, resulta em coleções de dados com características heterogêneas e volumosas. Esse fenômeno impacta não apenas na quantidade, mas também na complexidade dos dados produzidos, pois inclui formas não estruturadas, novas e diferentes. Isso potencializa a necessidade de representar visualmente informações para que pessoas e organizações possam visualizar dados, obter conhecimento e utilizá-los em tomadas de decisão.

Nesse sentido, diversas áreas de conhecimento como estatística, inteligência artificial, *Visual Analytics*, visualização, entre outras, estão contribuindo no intuito de contornar esse desafio de analisar e converter dados em informações úteis (CUI, 2019). Contudo, a área de visualização tem a vantagem de usar algumas propriedades gráficas, como formas, cor, tamanho, posição etc, que são processadas de maneira rápida e eficiente pela capacidade da percepção visual humana (WARD; GRINSTEIN; KEIM, 2015).

Um dos desafios da visualização é representar conjuntos de dados com alta dimensionalidade. Mesmo que a literatura apresente técnicas mais adequadas, chamadas de visualizações multidimensionais, estas geralmente apresentam pouca escalabilidade visual, ou seja, a medida que aumenta a quantidade de dimensões, é necessário um espaço visual cada vez maior para comportar todo o conjunto de dados. Entretanto, o grupo de técnicas que realizam redução de dimensionalidade conseguem contornar esse problema, sendo estas chamadas de Projeções Multidimensionais. Elas buscam preservar os padrões da base original no espaço projetado, posicionando instâncias mais próximas ou mais distantes na tela caso tenham maior ou menor similaridade, respectivamente (THIJSSSEN et al., 2023).

Diante disso, este estudo propõe o MultiVisD3, que é uma abordagem de *Visual Analytics* que utiliza múltiplas visualizações interativas para analisar a qualidade de duas Projeções Multidimensionais, por meio de métricas que auxiliam na tarefa de identificar erros e distorções durante o processo de mapeamento dos dados.

## 1.2 MOTIVAÇÃO

As técnicas de visualização são ferramentas importantes para apoiar o processo de exploração visual, através de representações gráficas que traduzem volumes de dados em formas visuais para auxiliar a percepção humana, e tornar a análise mais eficiente. No entanto, as técnicas de visualização possuem limitações físicas considerando o espaço 2D das telas. Uma das abordagens para lidar com as restrições da bidimensionalidade é vincular as visualizações através das Múltiplas Visões Coordenadas. Isto é, algum procedimento que estabeleça algum tipo de relação entre duas ou mais representações visuais (WILHELM, 2008).

Quando usadas de maneira adequada, as Múltiplas Visões Coordenadas podem minimizar a sobrecarga cognitiva do usuário em comparação com uma única visualização (SCHERR, 2008). O conceito de coordenar múltiplas visões pode contribuir com a superação da complexidade visual. A ideia é, em vez de criar uma única visualização complexa, cria-se várias visualizações mais simples e as vincula, para que, ao usuário interagir com uma visualização, as outras visões atualizem e mostrem os resultados dessa interação. Como resultado o analista usa múltiplas visões que requerem menos interpretação e usam combinações específicas de dados (WILLS, 2008).

## 1.3 PROBLEMA DE PESQUISA

As Projeções Multidimensionais possuem escalabilidade visual para mapear os crescentes volumes de dados da atualidade, todavia, essas técnicas contém erros e distorções intrínsecas em seus *layouts* quando realizam esse mapeamento. Identificar quais técnicas possuem menos distorções ou mesmo onde e como ocorre essas distorções é um desafio na literatura científica da área (NONATO; AUPETIT, 2019).

Diante disso, a problemática deste estudo é que os trabalhos desenvolvidos nessa linha de pesquisa não abordam o uso de MVC para explorar e comparar a qualidade de diferentes técnicas de projeção multidimensional justapostas, e também não consideram características hierárquicas. Por um lado, autores utilizam uma técnica de visualização multidimensional não escalável para explorar erros, ou quando utilizam mais de uma, não consideram diferentes tipos de erros ou representações hierárquicas desses erros.

## 1.4 OBJETIVOS

Esta pesquisa tem como objetivo geral propor uma abordagem de visualização para analisar a qualidade das Projeções Multidimensionais utilizando Múltiplas Visualizações Coordenadas. Como objetivos específicos desta pesquisa, destacam-se:

- Investigar quais técnicas de projeções multidimensionais estão associadas com o estado-da-arte da literatura.
- Identificar na literatura métricas de qualidade de projeções multidimensionais.
- Desenvolver uma abordagem visual que coordene múltiplas visualizações por meio de um conjunto de interações.

- Implementar as métricas de qualidade na abordagem visual desenvolvida, bem como outras visualizações que forneçam suporte a melhor compreensão das projeções multidimensionais.
- Selecionar conjuntos de dados amplamente utilizados na literatura da área para testar a abordagem visual proposta.
- Avaliar a abordagem de visualização proposta realizando testes com usuários.

## 1.5 QUESTÕES DE PESQUISA

O desenvolvimento deste trabalho foi norteado pelas Questões de Pesquisa (QP) listadas abaixo.

- **QP1** É possível comparar a qualidade de duas projeções multidimensionais distintas?
- **QP2** Como diferentes métricas de erros podem ser visualizadas em diferentes projeções multidimensionais?
- **QP3** É possível associar o valor de erros das projeções multidimensionais com representações hierárquicas ?
- **QP4** O quão útil é associar MVC como apoio à compreensão de diferentes tipos de erros em diferentes PMs utilizando representação hierárquica destes erros?

## 1.6 ESTRUTURA DO TRABALHO

Além deste capítulo, o presente trabalho está organizado da seguinte forma: no Capítulo 2 são apresentados conceitos fundamentais da área de visualização, com exemplos de visualizações tradicionais, multidimensionais e hierárquicas. O Capítulo 3 descreve os conceitos relacionados as Múltiplas Visões Coordenadas, bem como as estratégias para vincular visualizações, além de abordar sobre diferentes tipos de interações. O Capítulo 4 apresenta os trabalhos relacionados que utilizam Múltiplas Visões Coordenadas para extrair informações em diferentes tipos de domínios de dados através de técnicas de visualização multidimensional.

No Capítulo 5 é detalhada o trabalho de pesquisa, com a metodologia, tecnologias utilizadas e a descrição do MultiVisD3: *Projection Error Analyzer*. O Capítulo 6 apresenta as avaliações realizadas no MultiVisD3, contendo a descrição de um cenário de uso para destacar as funcionalidades da aplicação e, uma avaliação experimental realizada com usuários para explorar elementos de funcionalidade e usabilidade.



## VISUALIZAÇÕES MULTIDIMENSIONAIS

### 2.1 CONSIDERAÇÕES INICIAIS

A medida que a quantidade e complexidade dos dados coletados e armazenados aumenta, maior é a necessidade de integrar métodos que auxiliem no processo de extração de informação como ocorre com as técnicas de visualização de dados. Em específico, as técnicas de visualização multidimensional que auxiliam a contornar o desafio de representar visualmente todo o conjunto de dados. Este capítulo inicia apresentando conceitos da área de visualização e, em seguida, apresenta as técnicas para o contexto específico de multidimensionalidade.

### 2.2 VISUALIZAÇÃO DE DADOS

Visualização pode ser definida como a comunicação de informação usando representações gráficas. Uma imagem, por exemplo, pode conter muitas informações e pode ser processada de maneira mais rápida do que uma página de texto. Isso acontece porque a interpretação da imagem é realizada em paralelo dentro do sistema perceptivo humano, enquanto que a velocidade da análise do texto é limitada pelo processo sequencial de leitura. O entendimento de uma imagem também pode acontecer independente da uma linguagem local, pois um gráfico ou mapa pode ser entendido por um grupo de pessoas sem um idioma comum (WARD; GRINSTEIN; KEIM, 2015).

As visualizações podem ser exploratória ou explanatória. A visualização exploratória é aplicada para análise e compreensão dos dados, ou seja, descobrir novos conhecimentos e identificar estruturas, padrões, anomalias, tendências e relacionamentos (WARD; GRINSTEIN; KEIM, 2015). A visualização explanatória é usada para transmitir informações, apresentação de resultados e explicação de conclusões (KNAFLIC, 2018).

A necessidade obter informações úteis dos volumes de dados coletados provocou uma ampliação da atividade de pesquisa no campo da análise de dados e a *Visual Analytics* é uma área que realiza pesquisas multidisciplinares envolvendo visualização, mineração de dados, gerenciamento de dados, fusão de dados, estatística, interação humano-computador, análise de dados, estatística, percepção, cognição e raciocínio analítico. *Visual Analytics* combina técnicas de análise automatizada com visualizações interativas

para entender, raciocinar e tomar decisões eficazes a partir de grandes e complexas coleções de dados. Esta área contribui para o desenvolvimento de ferramentas e técnicas para habilitar pessoas a sintetizar informações e obter conhecimento a partir de grandes volumes de dados, dinâmicos, ambíguos e muitas vezes conflitantes (KEIM et al., 2008; CUI, 2019).

Um fator relevante para o processo de explorar visualmente dados é a ação do usuário em interagir e manipular as representações gráficas. Portanto, é essencial permitir que os usuários personalizem, modifiquem e refinem de forma interativa as visualizações até que sintam que atingiram seus objetivos. A combinação de visualização e análise de dados contribui para a construção de um modelo visual que aproxime o analista dos dados (WARD; GRINSTEIN; KEIM, 2015).

Ao utilizar representações gráficas é importante saber que os dados são mapeados para diferentes atributos visuais como posição, forma ou marcador, tamanho, brilho, cor, orientação, textura e movimento. Esse último pode ser associado a qualquer um dos outros atributos visuais e está ligado a transmissão de informações enquanto uma variável muda ao longo do tempo. Esses oito atributos visuais auxiliam a percepção humana, pois podem codificar informações e podem ser combinados e ajustados conforme necessário para maximizar a eficácia de uma visualização (WARD; GRINSTEIN; KEIM, 2015).

Em relação aos tipos de visualização, as tradicionais<sup>1</sup>, isto é, aquelas que mapeiam conjuntos de dados simples, com duas ou três dimensões são mais comuns de serem utilizadas, seja pela facilidade de interpretação ou pela familiaridade com determinados padrões. Contudo, as visualizações mais tradicionais não são adequadas para representar dados multidimensionais. Para tal, as visualizações multidimensionais são mais escaláveis, conseguindo mapear conjuntos de dados com um número maior de dimensões, sendo descritas a seguir.

### 2.2.1 Visualização Multidimensional

As visualizações multidimensionais são representações gráficas que codificam dados com  $n$ -dimensões para estruturas visuais. Desta forma, elas mapeiam conjuntos de dados de um espaço  $n$ -dimensional compostos por muitas dimensões e instâncias. Essas visualizações representam as dimensões dos dados através de elementos visuais como formas, tamanho ou cor, por exemplo (PILLAT, 2006). Keim e Kriegel (1996) classificam as visualizações multidimensionais como:

- Projeções geométricas, que visam mapear os atributos dos dados para uma representação gráfica, através de projeções geométricas como linhas e pontos, por exemplo;
- Inconográficas ou técnica baseada em ícones. Em que a ideia é mapear cada item de dado multidimensional para um ícone;

---

<sup>1</sup>As visualizações tradicionais são comumente denominadas de explanatórias, pois apresentam resultados de um modo mais direto e são interpretadas mais facilmente. Alguns exemplos são os Gráfico de Colunas ou Barras, Histograma, *Box Plot* ou Diagrama de Caixa, Gráfico de Linhas, Gráfico de Dispersão ou *Scatter Plot*, Gráfico de Pizza etc.

- Orientadas a pixel, cujo propósito é mapear cada valor de dado para um pixel colorido, e apresentar os valores de dados pertencentes a um atributo em partes separadas.

Para compreender melhor como ocorre a representação de dados multidimensionais, serão apresentadas e discutidas, a seguir, três técnicas de visualização multidimensional muito populares, sendo elas: Coordenadas Paralelas, Matriz de Gráficos de Dispersão e Gráfico de Radar.

**Coordenadas Paralelas (CP)** - Foram introduzidas pela primeira vez por Inselberg em 1985 (INSELBERG, 1985). Essa, é uma técnica de visualização de projeção geométrica amplamente utilizada para exploração de dados multidimensionais. Nela, cada atributo ou dimensão é representado por uma linha vertical, chamada de eixo. O valor de cada item de dado dos atributos são normalizados, convertidos para pontos e mapeados para os eixos através de polilinhas horizontais (HEINRICH; WEISKOPF, 2013).

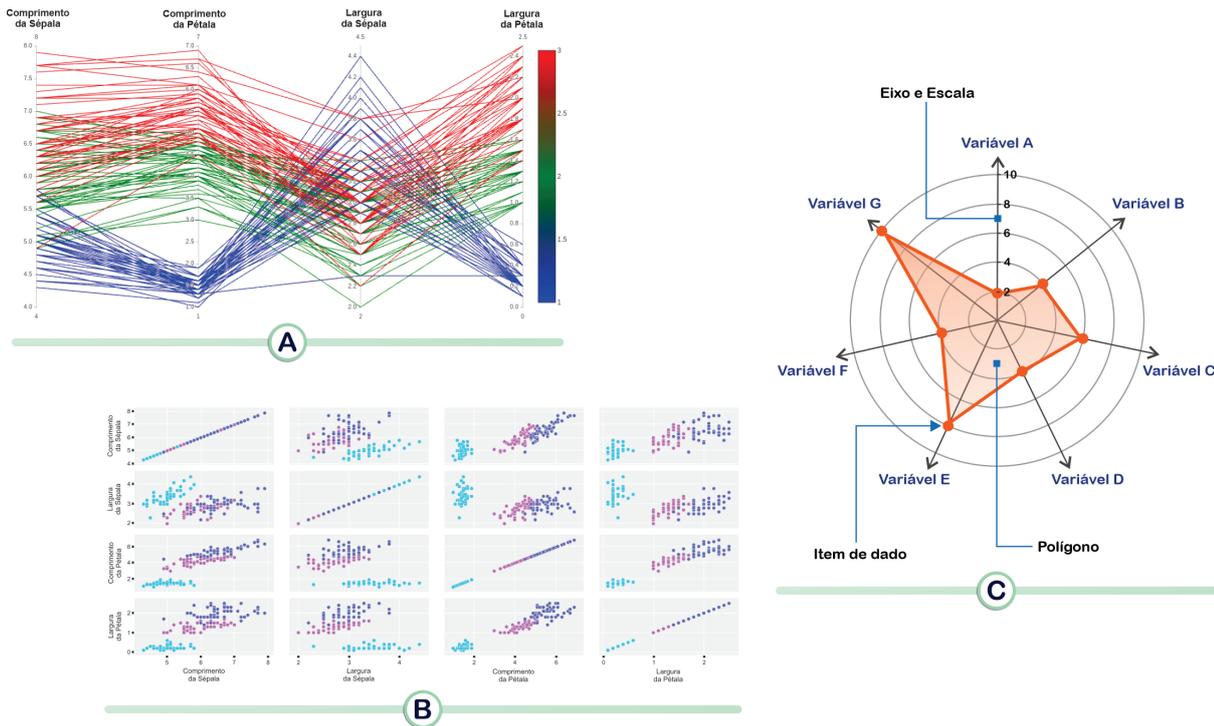
As CPs auxiliam na análise dos dados em busca de agrupamentos de linhas semelhantes (indicando correlação parcial entre pares de dimensões), pontos de cruzamento semelhantes (indicando correlações negativas parciais) e linhas que são isoladas ou têm uma inclinação significativamente diferente de seus vizinhos (indicando *outliers*). A combinação dessa representação gráfica com técnicas de interação, amplia o potencial de revelar relações entre as dimensões de dados (WARD; GRINSTEIN; KEIM, 2015).

A Figura 2.1 item 'A' exemplifica o uso de CPs para visualizar o conjunto de dados de flores Iris. As cores das polilinhas estão mapeando as espécies, assim, é possível observar através do mapa de cor, que as instâncias se concentram em três grupos: azul (setosa), verde (versicolor) e vermelho (virginica).

**Matriz de Gráficos de Dispersão** - Ou Matriz de *Scatter Plots* é uma técnica de visualização de dados multidimensional classificada como geométrica. Ela é uma extensão do gráfico de dispersão para representar dados multidimensionais. Dessa forma, múltiplos gráficos de dispersão são exibidos em forma de matriz, e mapeiam diferentes pares de atributos do conjunto de dados. Nessa representação gráfica, os itens de dados são mapeados para pontos ou marcadores visuais, e são plotados em coordenadas (x, y) (PILLAT, 2006).

Há diversas variações da Matriz de *Scatter Plots*, algumas com suporte a recursos de interação ou com histogramas nas diagonal da matriz. A Figura 2.1 item 'B' ilustra um exemplo de uma matriz com dezesseis gráficos de dispersão representando a base de dados Iris. Os atributos largura da pétala, largura da sépala, comprimento da pétala e comprimento da sépala são mapeados par a par, gerando uma matriz  $n \times n$ . A cor dos pontos representa o atributo que identifica a espécie, sendo ciano (setosa), roxo (versicolor) e azul (virginica).

**Gráfico de Radar** - É uma visualização geométrica projetada para dados multidimensionais, utiliza uma forma de diagrama radial também conhecido como Diagrama de Kiviat, Gráfico Aranha, Gráfico de Teia, Gráfico Polar e Gráfico Estelar (Figura 2.1 item 'C'). Os gráficos de radar são utilizados para comparar múltiplas variáveis quantitativas. Isso os torna úteis para ver quais variáveis têm valores semelhantes, mais altos, mais baixos ou se há valores discrepantes entre cada variável (DRAPER; LIVNAT; RIESEN-



**Figura 2.1** Visualizações Multidimensionais (A - Coordenadas Paralelas; B - Matriz de Gráficos de Dispersão; C - Gráfico de Radar).

FELD, 2009).

Um Gráfico de Radar é construído mapeando cada variável para um dos eixos que irradiam de um ponto central comum. Os eixos são espaçados igualmente entre si e o intervalo de cada variável é dimensionado de forma que cada eixo tenha o mesmo comprimento. Os pontos de dados são desenhados em seus locais apropriados nos eixos e linhas são desenhadas conectando-os (DRAPER; LIVNAT; RIESENFELD, 2009).

Os Gráfico de Radar podem ser plotados de diferentes formas: *i*) os eixos são desenhados, mas a circunferência não; *ii*) a circunferência, mas omitem os eixos; *iii*) uma combinação desses dois últimos. Além disso, múltiplos Gráfico de Radar podem ser sobrepostos para permitir comparações entre vários conjuntos de dados distintos com nomes de campos comuns (DRAPER; LIVNAT; RIESENFELD, 2009).

Conforme os conjuntos de dados crescem em quantidade tanto de colunas quanto de registros, as técnicas de visualização multidimensionais supracitadas apresentam limitações quanto a escalabilidade visual, tornando a representação gráfica sobrecarregada e de difícil compreensão. Contudo, outras técnicas conseguem contornar esse problema, como por exemplo, as projeções multidimensionais.

### 2.2.1.1 Projeções Multidimensionais

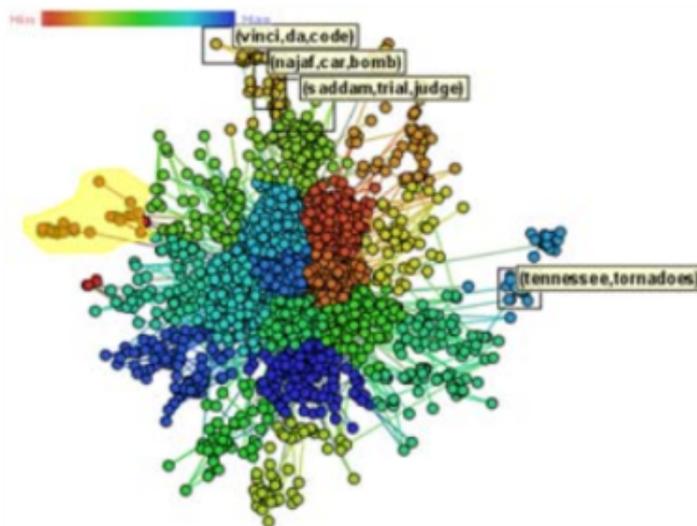
As Projeções Multidimensionais (PMs) são caracterizadas como uma família de mapeamentos capazes de produzir pontos que preservam a similaridade e vizinhança em

espaços multidimensionais, com *layouts* baseados em gráficos de dispersão. Uma medida de similaridade calcula o quão parecidos são dois itens de dados. Por outro lado, dissimilaridade é uma medida numérica de como dois itens de dados são diferentes. Quanto maior a medida, mais distintos são os itens. Os valores altos de similaridade codificam itens quase idênticos, enquanto baixa similaridade significa que os itens são diferentes (NONATO; AUPETIT, 2019).

As Projeções Multidimensionais auxiliam na interpretação de grandes volumes de dados com alta dimensionalidade. Contribuindo para a extração de informação, através de um conjunto de pontos no plano, em uma reta ou em um volume (PAULOVICH, 2008; NEVES et al., 2015). Uma técnica de projeção multidimensional pode ser descrita como uma função injetiva  $f : \mathbb{R}^m \rightarrow \mathbb{R}^p$  que mapeia dados de um espaço  $m$ -dimensional em um espaço  $p$ -dimensional com  $p = \{1, 2, 3\}$  e  $p > m$ , preservando, no espaço projetado, algumas informações sobre as relações de distância entre os itens de dados no espaço  $m$ -dimensional.

Formalmente, tem-se que  $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$  seja um conjunto de dados  $m$ -dimensional em  $\mathbb{R}^m$  com  $\delta : \mathbb{R}^m, \mathbb{R}^m \rightarrow \mathbb{R}$  uma medida de dissimilaridade entre duas instâncias de dados  $m$ -dimensionais, e  $d : \mathbb{R}^p, \mathbb{R}^p \rightarrow \mathbb{R}$  uma distância (normalmente Euclidiana) entre dois pontos no espaço de projeção. Uma projeção multidimensional visa tornar  $|\delta(x_i, x_j) - d(f(x_i), f(x_j))|$  o mais próximo possível de zero,  $\forall x_i, x_j \in \mathbf{X}$  (TEJADA; MINGHIM; NONATO, 2003; PAULOVICH et al., 2008; NEVES et al., 2015; PAULOVICH, 2008).

A Figura 2.2 representa visualmente uma coleção de 2.684 notícias da RSS, coletadas de quatro agências de notícias *online* (BBC, CNN, *Reuters* e *Associated Press*) por dois dias em abril de 2006. Após o pré-processamento, 2.242 dimensões de dados foram analisadas. Nesta Projeção Multidimensional as cores dos pontos representam os grupos de temas das notícias mais frequentes. A partir da interação do usuário, foram exibidos mais detalhes das regiões selecionadas através de rótulos flutuantes sobre os pontos.



**Figura 2.2** Mapa de documentos produzido a partir de Projeção Multidimensional. Adaptado de Paulovich, Oliveira e Minghim (2007).

Existem diferentes taxonomias que classificam as técnicas de Projeção Multidimensional de acordo com objetivos específicos. Nonato e Aupetit (2019), por exemplo, apresentam uma categorização abrangente considerando oito características, sendo elas:

- Tipos de Dados - Os principais tipos de dados manipulados em aplicações de visualização são dissimilaridade, ordinal, cartesiano, vizinhança e categóricos. Esses dados geralmente estão em uma matriz de dados com características identificadas (atributos ou dimensões) associadas a cada instância de dados. Ou uma matriz de dissimilaridade, onde cada entrada é responsável pela similaridade ou dissimilaridade entre os itens. Algumas técnicas de Projeção Multidimensional são mais flexíveis com relação aos tipos de dados que são capazes de lidar, mas nenhuma é capaz de manipular dados de todas as categorias.
- Linearidade - Uma projeção pode ser linear ou não linear. Ambos os tipos são bem abordados na literatura e igualmente importantes na prática. As projeções lineares são fáceis de entender e usar, mas não conseguem capturar bem as distribuições de amostras espalhadas em variedades complexas em  $nD$ . As projeções não lineares são melhores para esses conjuntos de dados, mas são mais difíceis de controlar em termos de parâmetros (ESPADOTO et al., 2021).
- Flexibilidade de Supervisão - Métodos supervisionados de Projeção Multidimensional levam em conta informações de rótulo de classe para realizar o mapeamento, com o objetivo de colocar as instâncias pertencentes a uma mesma classe próximas umas das outras no espaço visual. Algumas técnicas modificam as distâncias de acordo com as informações da classe, tornando as instâncias da mesma classe mais próximas, ao mesmo tempo em que separa instâncias em classes distintas.
- Capacidade de lidar com estruturas multinível - Mecanismos hierárquicos têm sido uma das alternativas usadas para lidar com grandes conjuntos de dados, permitindo diminuir a carga computacional e a questão de gerar *layouts* desordenados. A maioria das técnicas hierárquicas de Projeção Multidimensional trabalham em duas etapas, primeiro construindo uma representação hierárquica dos dados e, em seguida, mapeando os dados de níveis específicos da hierarquia para o espaço visual. Nem todas as técnicas podem ser adaptadas para lidar com dados hierárquicos. Pois enfrentam o desafio de como “sincronizar” o mapeamento de diferentes níveis para preservar as estruturas de contexto e vizinhança.
- Localidade - O conceito de localidade é usado na literatura para caracterizar duas propriedades diferentes dos métodos de Projeções Multidimensionais:
  - Modelagem local - São métodos que visam extrair informações geométricas e/ou topológicas locais dos dados. Fazendo um mapeamento global que preserva essas informações locais em uma segunda etapa do processo. Em outras palavras, os métodos locais de modelagem dependem das informações de vizinhança de cada instância de dados para construir um único mapa que projeta os dados enquanto preserva as informações de vizinhança obtidas.

- Mapeamento local - Essas técnicas contam com informações de vizinhança, mas, em contraste com os métodos locais de modelagem, criam uma família de transformações locais para projetar os dados, ou seja, as instâncias não são projetadas por uma única transformação global, mas por um conjunto de mapeamentos locais. Cada mapeamento é responsável por projetar um subconjunto dos dados, preservando as estruturas locais correspondentes.
- Direção - É responsável por conduzir o processo de projeção com base em subconjuntos específicos de instâncias. Um recurso usado a muito tempo são os mecanismos de ponderação. Permitindo que certas instâncias tenham maior influência no processo de mapeamento. Ajustar pesos para dirigir projeções não é uma tarefa fácil e pode se tornar oneroso em termos computacionais. Outro recurso empregado são os pontos de controle que orientam as projeções de acordo com a interação do usuário. Existem também os conceitos de pivôs (ou pontos de referência) que orientam a precisão do processo de mapeamento com base nas regiões de interesse.
- Estabilidade - Refere-se a como responder a variações nos dados de entrada de uma técnica de Projeção Multidimensional, onde a palavra “variações” é responsável por qualquer alteração nos dados, desde pequenas perturbações a um aumento ou diminuição no número de instâncias a serem mapeado. Se o mapeamento for estável, pequenas variações nos dados devem levar a pequenas variações na projeção. Além disso, um método estável não deve modificar a posição de instâncias projetadas quando novos dados são manipulados.
- Capacidade de lidar com dados fora do núcleo - Significa que a técnica de Projeção Multidimensional é capaz de mapear novos dados preservando sua relação com os dados projetados anteriormente, ou seja, o método pode trabalhar de forma contínua. O exemplo mais direto são as técnicas lineares, que podem naturalmente projetar novos dados usando a mesma transformação linear computada para projetar dados anteriores.

Espadoto et al. (2021) propõem uma Taxonomia onde as técnicas são agrupadas considerando dois atributos, linearidade (linear ou não linear) e localidade ou tipo de vizinhança (local ou global). Esta classificação é mais simplificada e tem o intuito de ajudar os usuários não especialistas a selecionarem as técnicas mais apropriadas para o seu problema. Os quatro grupos resultantes dessa Taxonomia são:

- Linear e Global - Técnicas que utilizam apenas transformações lineares e consideram todas as amostras de uma vez.
- Não Linear e Local - Técnicas que utilizam funções não lineares e buscam preservar a vizinhança local para cada amostra.
- Não Linear e Global - Técnicas que utilizam funções não lineares e consideram todas as amostras de uma vez.

- Linear e Local - Técnicas que utilizam apenas transformações lineares para reduzir a dimensionalidade em pequenas vizinhanças separadas.

Os trabalhos que discutem as taxonomias supracitadas apresentam e analisam diferentes técnicas de Projeções Multidimensionais como *Metric Multidimensional Scaling* (MDS) (TORGERSON, 1958), *Principal Component Analysis* (PCA) (JOLLIFFE, 1986), *t-Stochastic Neighborhood Embedding* (T-SNE) (MAATEN; HINTON, 2008) , *Uniform Manifold Approximation and Projection* (UMAP) (MCINNES; HEALY; MELVILLE, 2018), *Local Affine Multidimensional Projection* (LAMP) (JOIA et al., 2011), *Least Square Projection* (LSP) (PAULOVICH et al., 2008) etc.

Essas pesquisas realizaram experimentos e investigaram vários aspectos como escalabilidade, performance e capacidade de manter as características dos dados no espaço de baixa dimensão, etc. Espadoto et al. (2021) utilizaram conjuntos de dados distintos em seus testes, incluindo *datasets* tabulares, e as técnicas LAMP, t-SNE e UMAP se destacaram com resultados altos ou acima da média. Nonato e Aupetit (2019) também fazem boas recomendações a essas técnicas. A seguir é apresentada de maneira sumarizada a abordagem de transformação utilizada por cada uma das três.

## LAMP

*Local Affine Multidimensional Projection* é resultado de uma combinação de propriedades que a destaca dentre os métodos de projeção multidimensional local. Além de ser econômica e altamente precisa em relação a outras técnicas de projeção, a LAMP não depende de gráficos de vizinhança e sua formulação matemática admite um subconjunto bastante reduzido de instâncias amostrais como entrada, o que favorecem a robustez ao lidar com aplicações altamente interativas que envolvem grandes conjuntos de dados (JOIA et al., 2011).

Segundo Joia et al. (2011), a técnica LAMP faz uso de um subconjunto de amostras, denominados de pontos de controle, e sua localização no espaço visual. A informação obtida dos pontos de controle é utilizada para construir uma família de mapeamentos ortogonais afins, um para cada instância a ser projetada. A LAMP apresenta baixa distorção mesmo quando é utilizado um número reduzido de pontos de controle. Para realizar a projeção de uma instância  $x$ , a técnica LAMP encontra uma função linear no sentido de mapeá-la para o espaço visual através da transformação  $f_x(p) = pM + t$  que minimiza:

$$\sum_i \alpha_i \|f_x(x_i) - y_i\|^2, \quad \text{sujeito a } M^T M = I \quad , \quad (2.1)$$

onde a matriz  $M$  e o vetor  $t$  são as variáveis,  $I$  é a matriz identidade e  $\alpha_i$  são pesos escalares definidos como:

$$\alpha_i = \frac{1}{\|x_i - x\|^2} \quad . \quad (2.2)$$

Para minimização da Equação 2.1, é necessário reformular sua escrita e transcrevê-la em forma matricial. No entanto, esse processo recai nas restrições da ortogonalidade,

discutidas detalhadamente por (GOWER; DIJKSTERHUIS, 2004). A solução é tomar as derivadas parciais em relação a  $t$  igual a zero, e escrever  $t$  em termos de  $M$  como:

$$t = \tilde{y} - \tilde{x}M, \quad \tilde{x} = \frac{\sum_i \alpha_i x_i}{\alpha}, \tilde{y} = \frac{\sum_i \alpha_i y_i}{\alpha}, \quad (2.3)$$

onde  $\alpha = \sum_i \alpha_i$ . Logo, a minimização da Equação 2.1 pode ser reescrita como:

$$\sum_i \alpha_i \|\hat{x}_i M - \hat{y}_i\|^2, \quad \text{sujeito a } M^T M = I, \quad (2.4)$$

onde  $\hat{x}_i = x_i - \tilde{x}$  e  $\hat{y}_i = y_i - \tilde{y}$ . O problema de minimização da Equação 2.4 pode ser expresso na forma matricial

$$\|AM - B\|_F, \quad \text{sujeito a } M^T M = I, \quad (2.5)$$

onde  $\|\cdot\|_F$  denota a norma de *Frobenius* e as matrizes  $A$  e  $B$  são dadas por

$$A = \begin{bmatrix} \sqrt{\alpha_1} \hat{x}_1 \\ \sqrt{\alpha_2} \hat{x}_2 \\ \vdots \\ \sqrt{\alpha_k} \hat{x}_k \end{bmatrix}, \quad B = \begin{bmatrix} \sqrt{\alpha_1} \hat{y}_1 \\ \sqrt{\alpha_2} \hat{y}_2 \\ \vdots \\ \sqrt{\alpha_k} \hat{y}_k \end{bmatrix}. \quad (2.6)$$

O problema de minimização da Equação 2.5 é um exemplo típico do chamado Problema Ortogonal de *Procruste* (GOWER; DIJKSTERHUIS, 2004), cuja solução é conhecida por ser

$$M = UV, \quad A^T B = UDV, \quad (2.7)$$

onde  $UDV$  corresponde à decomposição em valores singulares (do inglês, *Singular Value Decomposition*) de  $A^T B$ . Uma vez calculado  $M$ , a projeção  $y$  de  $x$  é realizada por

$$y = fx(x) = (x - \tilde{x})M + \tilde{y}. \quad (2.8)$$

A LAMP além de resultar em mapeamentos altamente precisos, sua construção matemática descrita acima revela-se também competitiva em termos de tempos computacionais. Pois,  $A^T B$  é de fato uma matriz de apenas duas colunas, portanto pode ser decomposta muito rapidamente, resultando em um algoritmo com complexidade computacional igual a  $O(kn)$  (JOIA et al., 2011).

## t-SNE

A técnica denominada de *t-Distributed Stochastic Neighbor Embedding* é uma variação do método *Stochastic Neighbor Embedding*. A t-SNE pode ser otimizada e produz visualizações melhores, reduzindo a tendência de aglomerar pontos no centro do mapeamento. Além disso, é capaz de capturar muito bem grande parte da estrutura local dos dados de alta dimensão, ao mesmo tempo que revela a estrutura global, como a presença de *clusters* em diversas escalas. A t-SNE utiliza métodos probabilísticos para converter um

conjunto de dados de alta dimensão em uma matriz de similaridades de pares e em seguida representar esses dados no espaço visual de baixa dimensão (MAATEN; HINTON, 2008).

O mapeamento realizado pela t-SNE ocorre em duas etapas principais, sendo a primeira, uma distribuição utilizando métodos probabilísticos com pares de instâncias similares no espaço multidimensional de maneira que as instâncias similares tenham alta probabilidade de serem vizinhas, ao passo que instâncias dissimilares tenham probabilidade baixa (NEVES, 2016). Essa probabilidade é dada por:

$$P_{ij} = \frac{P_{i|j} + P_{j|i}}{2n} \quad , \quad (2.9)$$

onde,

$$P_{j|i} = \frac{\exp\left(-\frac{\delta(x_i, x_j)^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(-\frac{\delta(x_i, x_k)^2}{2\sigma_i^2}\right)} \quad , \quad (2.10)$$

os valores de  $\sigma_i$  controlam indiretamente o tamanho da vizinhança em  $X$ .

Na segunda etapa, também é realizada uma distribuição probabilística dos pontos no espaço transformado, definida como:

$$q_{ji} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|y_k - y_l\|)^{-1}} \quad . \quad (2.11)$$

A posição das instâncias no espaço projetado é obtido pela minimização da divergência de *Kullback-Leibler* (KULLBACK; LEIBLER, 1951), onde

$$KL(P||Q) = \sum_{j \neq i} P_{ij} \log \frac{P_{ij}}{q_{ij}} \quad . \quad (2.12)$$

Embora a t-SNE apresente contribuições e bons resultados para a visualização de dados, ela possui pontos de atenção, como, por exemplo, ter menos sucesso se for aplicado em conjuntos de dados com uma dimensionalidade intrínseca muito elevada (MAATEN; HINTON, 2008).

## UMAP

*Uniform Manifold Approximation and Projection* usa múltiplas aproximações locais e suas representações de conjuntos simpliciais difusos locais para construir uma representação topológica dos dados de alta dimensão. Desta forma, o *layout* da representação de dados no espaço de baixa dimensão é otimizado, para minimizar a entropia cruzada entre as duas representações topológicas (MCINNES; HEALY; MELVILLE, 2018). A representação visual produzida pela UMAP acontece em duas fases principais (WANG et al., 2021):

1. Construção do gráfico para o espaço de alta dimensão:

- (a) Para cada observação, encontre os  $k$  vizinhos mais próximos para uma determinada métrica de distância  $\text{Distância}_{i,j}$  (geralmente Euclidiana).
- (b) Para cada observação, calcule  $p_i$ , a distância positiva mínima da observação  $i$  para um vizinho.
- (c) Calcular um parâmetro de escala  $\sigma_i$  para cada observação  $i$  resolvendo a equação:

$$\log_2(k) = \sum_{j=1}^k \exp\left(\frac{-\max\{0, \text{Distância}_{i,j} - p_i\}}{\sigma_i}\right) . \quad (2.13)$$

Intuitivamente,  $i$  é usado para normalizar as distâncias entre cada observação e seus vizinhos para preservar as proximidades relativas de alta dimensão.

- (d) Defina a função peso:

$$w(x_i, x_j) \stackrel{def}{=} \exp\left(\frac{-\max\{0, \text{Distância}_{i,j} - p_i\}}{\sigma_i}\right) . \quad (2.14)$$

- (e) Defina um grafo ponderado  $G$  cujos vértices são observações e onde para cada aresta  $(i, j)$  afirma-se que  $x_i$  é o vizinho mais próximo de  $x_j$ , ou vice-versa. O peso (simétrico) de uma aresta  $(i, j)$  é igual a

$$w_{i,j} = w(x_i, x_j) + w(x_j, x_i) - w(x_i, x_j).w(x_j, x_i) . \quad (2.15)$$

## 2. Otimização do layout do gráfico de baixa dimensão:

- (a) Inicialização: inicialize  $Y$  usando incorporação espectral.
- (b) Intere sobre o Gráfico  $G$  aplicando os passos do gradiente descendente nas observações  $i$  em cada aresta  $(i, j)$  da seguinte forma:
  - i. Aplique uma força atrativa (no espaço de baixa dimensão) na observação  $i$  em direção à observação  $j$  (a força é uma aproximação do gradiente):

$$y_i = y_i + \alpha \cdot \frac{-2ab\|y_i - y_j\|_2^{2(b-1)}}{1 + a(\|y_i - y_j\|_2^2)^b} \bar{w}_{i,j}(y_i - y_j) . \quad (2.16)$$

- ii. Os hiper-parâmetros  $a$  e  $b$  são ajustados usando os dados da função

$$(1 + a(\|y_i - y_j\|_2^2)^b)^{-1}$$

para a função de peso não normalizada

$$\exp(-\max\{0, \text{Distância}_{i,j} - p_i\})$$

com o objetivo de criar uma aproximação suave. O hiper-parâmetro  $\alpha$  indica a taxa de aprendizagem.

- iii. Aplique uma força repulsiva à observação  $i$ , afastando-a da observação  $k$ . Aqui,  $k$  é escolhido aleatoriamente de uma amostra de vértices não vizinhos, ou seja, uma aresta  $(i, k)$  que não está em  $G$ . Essa força repulsiva é a seguinte:

$$y_i = y_i + \alpha \cdot \frac{b}{(\epsilon + \|y_i - y_k\|_2^2)(1 + a(\|y_i - y_k\|_2^2)^b)} (1 - \bar{w}_{i,k})(y_i - y_k) \quad , \quad (2.17)$$

onde  $\epsilon$  é uma pequena constante.

A descrição fornecida aqui baseia-se no trabalho de (MCINNES; HEALY; MELVILLE, 2018) e inclui pequenas modificações nos denominadores conforme a implementação do UMAP (MCINNES et al., 2018) discutidas por (WANG et al., 2021).

### 2.2.1.2 Métricas de Erro

Os mapeamentos de dados do espaços de alta dimensão para espaços de menor dimensão ocorrem sob condições muito particulares, por isso erros e distorções são altamente propensos a estarem presentes nos *layouts* das PM. As estruturas de vizinhança observadas no espaço visual podem não ser exatamente as mesmas existentes no espaço multidimensional original (SACHA et al., 2017). Uma das grandes preocupações no contexto de visualização é mensurar os erros e distorções incorporados nos mapeamentos dos métodos de PM.

As distorções ou erros são causadas basicamente por dois fenômenos diferentes que afetam as estruturas de vizinhança no espaço visual: O primeiro chamamos de Vizinho Ausente (VA), acontece quando instâncias vizinhas no espaço original  $D$  são mapeadas distantes umas das outras no espaço visual  $M$ ; O segundo é denominado de Vizinho Falso (VF) e ocorre no sentido inverso, isto é, um ou mais itens que não são vizinhos no espaço original são mapeados próximos uns dos outros no espaço visual. Existem ainda os Vizinhos Verdadeiros (VV), que representam o *layout* ideal em PM, ou seja, as instâncias que são vizinhas tanto no espaço original quanto no visual (NONATO; AUPETIT, 2019).

No geral, os fenômenos VA e VF resultam em estruturas de vizinhança enganosas e, uma vez que é altamente provável que essas distorções ocorram, as incertezas estão presentes no processo de análise visual. O grau de incerteza está diretamente relacionado à intensidade com que ocorrem os fenômenos VA e VF. Portanto, medir a intensidade dos fenômenos VA e VF é fundamental para avaliar a incerteza contida nos *layouts* do PM. Muitos métodos quantitativos foram propostos para avaliar a intensidade das distorções nos mapeamentos de PM.

Existem diferentes estudos que fornecem revisões detalhadas sobre as métricas de qualidade existentes (ESPADOTO et al., 2021; NONATO; AUPETIT, 2019). O trabalho de Espadoto et al. (2021) realiza um *benchmark* com métricas de qualidade usadas para avaliar técnicas de PM, enquanto Nonato e Aupetit (2019) abordam diferentes tipos de erros ou distorções que podem resultar dos mapeamentos do PM e analisa os mecanismos existentes para avaliar quantitativamente tais distorções.

Para esta pesquisa considerou-se o trabalho de Martins et al. (2014) para propor uma abordagem de visualização que permitisse analisar os dados por diferentes perspectivas: *i*) preservação da vizinhança; *ii*) erro agregado de cada instância *i* para todas as outras instâncias; *iii*) e erro agregado do agrupamento de pontos considerando a variável rótulo da base. Para quantificar os problemas de preservação da vizinhança, primeiro foi definido o erro de projeção do ponto *i* versos um ponto  $j \neq i$  como

$$e_{ij} = \frac{d^m(q_i, q_j)}{\max_{i,j} d^m(q_i, q_j)} - \frac{d^n(p_i, p_j)}{\max_{i,j} d^n(p_i, p_j)} \quad (2.18)$$

Tem-se que  $e_{ij} \in [-1, 1]$ . Erros negativos indicam pontos cujas projeções estão muito próximas (vizinhos falsos). Erros positivos representam pontos projetados que estão mais distantes (vizinhos ausentes). Valores zero indicam projeções “boas”, que se aproximam de maneira ideal das distâncias em  $D^n$ .

Em seguida é fornecida uma visão geral de como o erro de projeção se espalha por todo um conjunto de dados, calculando para cada ponto *i* o erro agregado

$$e_i^{aggr} = \sum_{j \neq i} |e_{ij}| \quad (2.19)$$

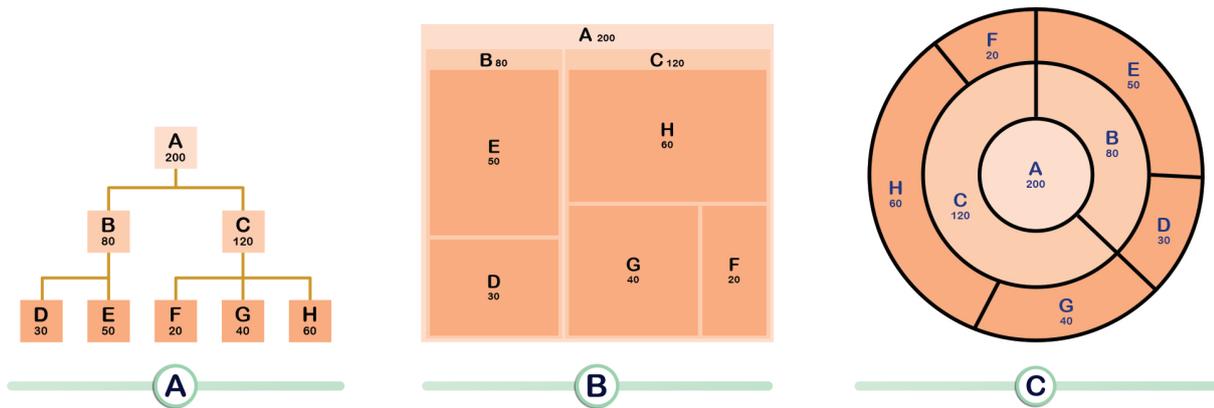
O valor de  $e_i^{aggr}$  fornece o erro de projeção do ponto *i* em relação a todos os outros pontos. Valores menores de  $e^{aggr}$  mostram pontos cujas projeções podem ser comparadas de forma confiável com a maioria das outras projeções em termos de avaliação de similaridade. Valores maiores de  $e^{aggr}$  mostram pontos que estão mal posicionados em relação à maioria dos outros pontos.

### 2.2.2 Visualização Hierárquica

Há diferentes visualizações que se propõem a representar dados hierárquicos, das quais *Treemap* e *Sunburst*, são muito populares, e mapeiam um espaço multidimensional em um unidimensional por meio do aninhamento e preenchimento do espaço com retângulos, quadrados ou polígonos para descrever a hierarquia das informações (CHEN et al., 2015). A Figura 2.3 item ‘A’, exemplifica uma estrutura de dados hierárquicos sendo representada no tradicional Diagrama de Árvore.

**Treemap** - É projetada para representar conjuntos de dados quantitativos e hierárquicos (Figura 2.3 item ‘B’). No entanto, tem se mostrado muito eficaz para análise de dados multidimensionais, em cenários envolvendo múltiplas visões coordenadas e quando usada em conjunto com técnicas multidimensionais. Nesta abordagem, os dados não são representados em estruturas de árvore. Onde o nó raiz fica na parte superior e os nós filhos abaixo do nó pai com linhas conectando-os. Ao invés disso, na visualização *Treemap* é definida uma hierarquia de atributos, e os dados de cada dimensão são mapeados para uma sequência aninhada de retângulos. Esse tipo de representação otimiza o preenchimento e a utilização do espaço de telas bidimensionais (2D) (SHNEIDERMAN, 1998; BEDERSON; SHNEIDERMAN; WATTENBERG, 2002).

**Sunburst** - É uma combinação de visualização com preenchimento radial, técnicas de navegação e interação para fornecer um sistema útil para representar hierarquias de



**Figura 2.3** Visualizações Hierárquicas (A - Diagrama de Árvore; B - *Treemap*; C - *Sunburst*).

informações (STASKO; ZHANG, 2000). Ele usa uma série de anéis concêntricos para expressar a estrutura hierárquica (Figura 2.3 item 'C'). Cada anel é dividido em vários setores que demonstram nós. A área do setor é proporcional ao valor do atributo do nó. E uma série de setores ao longo do raio radial refletem as relações pais-filhos (CHEN et al., 2015).

## 2.3 CONSIDERAÇÕES FINAIS

A área de visualização de dados fornece subsídios e ferramentas computacionais essenciais para as tarefas de análise de dados. Foram apresentados os principais conceitos que abrangem essa área. Em um segundo momento, apresentou-se exemplos de técnicas específicas para tratar a questão de bases de dados multidimensionais, ou ainda as técnicas capazes de representar dados hierárquicos.

As técnicas de visualização mais tradicionais geralmente mapeiam informações com poucas dimensões, dentre elas estão os gráficos de barra, de linhas e *box plot* etc. Elas têm características mais explanatórias, possibilitando inferir resultados facilmente, muito em razão da base de dados ser de baixa complexidade, com poucas dimensões.

As representações gráficas multidimensionais são mais adequadas para tratar a questão de multidimensionalidade. Dentre elas foram apresentadas as Coordenadas Paralelas, Gráfico de Radar e Matriz de Gráficos de Dispersão. Estas possuem características de visualizações exploratórias e mapeiam as dimensões/atributos em eixos, no entanto, possui limitações de escalabilidade visual para representar dimensões a partir da ordem das dezenas.

Nesse sentido, as projeções multidimensionais auxiliam nessa lacuna, pois como o plano projetado é euclidiano, não há representação explícita dos eixos, além de apoiarem na interpretação de grandes volumes de dados com alta dimensionalidade. Por fim, foram apresentadas as técnicas hierárquicas de visualização. A ideia é que elas possam servir de suporte às PMs na extração de informação, conforme o tema deste trabalho.



## MÚLTIPLAS VISÕES COORDENADAS

### 3.1 CONSIDERAÇÕES INICIAIS

Devido o crescimento e a diversidade das bases de dados, é comum recorrer a mais de uma técnica de visualização no processo de compreender os dados. Porém, usar múltiplas visualizações separadas podem sobrecarregar visualmente e cognitivamente o usuário, além de esbarrar nas limitações que as técnicas apresentam em relação ao tipo de dado e ao tamanho do *dataset* explorado. Um dos desafios de extrair conhecimento de coleções de dados complexas e grandes é, por exemplo, vincular e relacionar as informações de uma visualização com outra visualização. Esse vínculo, aumenta a capacidade de análise e permite uma maior interação do usuário com os dados (ROBERTS, 2005).

Este capítulo apresenta conceitos relacionados a Múltiplas Visões Coordenadas. Em seguida, serão descritas as formas de particionar o espaço de tela em múltiplas visões. Após, serão abordadas as estratégias de coordenação para vincular visualizações. E por fim, será tratado sobre as interações, ou seja, as formas que o usuário altera as visualizações durante a análise exploratória, suas classes, seus estilos e alguns exemplos.

### 3.2 MÚLTIPLAS VISÕES COORDENADAS

Para Roberts (2007) e Scherr (2008), MVC é uma técnica de visualização exploratória que possibilita ao usuário conhecer seus dados de forma interativa. Essas interações são fundamentais na exploração dos dados e percepção de padrões, pois permitem ao usuário selecionar e/ou filtrar parte dos dados, modificar a posição de objetos nas representações gráficas, etc. e essas interações são refletidas nas demais visualizações.

Para que esse conceito funcione na prática, é necessário que um esquema de comunicação seja estabelecido entre as visualizações, de modo que haja um elemento emissor/ativo e um ou mais elementos receptores/passivos. O gráfico ativo envia uma mensagem para todos os gráficos passivos, que executam a operação contida na mensagem (CHEN; HÄRDLE; UNWIN, 2007). As visualizações coordenadas possibilitam ao usuário especialista, nesse caso também chamado de analista, explorar visualmente o conjunto de dados através

de representações visuais que compartilham ou trocam informações entre si (WILHELM, 2008).

As Visões coordenadas podem facilitar de forma significativa a assimilação do usuário (ALMEIDA et al., 2009). Permite que o observador execute a correlação entre duas ou mais visualizações diferentes do mesmo conjunto de dados (MACIEL et al., 2008). E para auxiliar os desenvolvedores em seus projetos de visualizações coordenadas e ajudar os usuários na escolha de *softwares* com essas características, Baldonado, Woodruff e Kuchinsky (2000) sugerem que para um maior custo/benefício os sistemas de visualização com MVC sejam desenvolvidos considerando quatro regras:

- Regra da Diversidade - quando houver uma variedade de atributos, modelos, perfis de usuário, níveis de abstração ou gênero;
- Regra da Complementaridade - quando diferentes visualizações destacam correlações ou disparidades;
- Regra da Decomposição - quando os dados são complexos, recomenda-se particionar dados em várias visualizações para criar visões de dados mais simples e gerenciáveis, além de promover mais entendimento durante a interação com as dimensões dos dados;
- Regra da Parcimônia - utilizar o mínimo de múltiplas visões, se o uso de MVC gerar um custo de aprendizado muito elevado para o usuário.

As principais vantagens das Múltiplas Visões Coordenadas são: *i*) a facilidade das exibições gráficas ocuparem o mesmo espaço de tela; *ii*) a velocidade e a flexibilidade com as quais diferentes aspectos dos dados podem ser mostrados; *iii*) a aplicabilidade em estruturas de dados complexas. Essas vantagens são características essenciais no estágio exploratório da análise de dados. Elas auxiliam na investigação da natureza da distribuição dos dados, na busca por comportamento incomum ou inesperado e ajudam a detectar relações, estruturas e padrões (CHEN; HÄRDLE; UNWIN, 2007; WILHELM, 2008).

### 3.3 MÚLTIPLAS VISÕES

Uma visão dos dados ou uma única visualização serve como ferramenta para oferecer ao usuário uma forma de examinar os dados e obter *insights*. Apesar da utilidade inquestionável dessa abordagem de uma única visão dos dados, ela apresenta alguns problemas que a impede de ser uma solução completa. Por exemplo, os gráficos estão se tornando cada vez mais complexos e mais difíceis de interpretar. Desse modo, apesar da fácil interpretabilidade de gráficos 1D e 2D, visões que contêm muitos atributos (dimensões) se tornam menos intuitivas (WILLS, 2008).

Neste contexto de múltiplas visões, frequentemente o termo *'facet'* é utilizado e significa dividir ou particionar. Ele está relacionado, por exemplo, a divisão ou o particionamento do espaço de tela em várias visualizações ou em várias camadas de visualizações. Essa divisão do espaço de tela em múltiplas visões, tem a ver com o particionamento dos

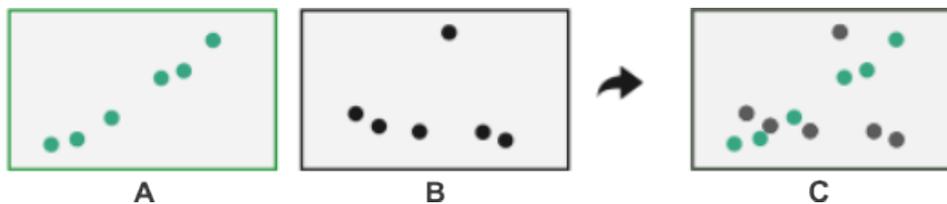
dados entre as visualizações para diminuir a complexidade visual (MUNZNER, 2014). E isso pode ocorrer das seguintes formas:

- Justaposição de visualizações, sendo uma ao lado da outra. Cabe ao desenvolvedor escolher quando mostrar cada visualização e como organizá-las na tela. Em geral, a justaposição de visões é utilizada em um cenário onde as visualizações são permanentemente visíveis lado a lado para que os usuários possam olhar entre elas. Porém, existe a opção de ter uma exibição que apareça temporariamente, em resposta a alguma ação do usuário. As visualizações justapostas oferecem muitas opções de criar visualizações coordenadas ou vinculadas (MUNZNER, 2014). Um exemplo de justaposição de visões lado a lado é mostrado na Figura 3.2.



**Figura 3.1** Justaposição de dois gráficos de dispersão A e B, lado a lado. Adaptado de Munzner (2014).

- Sobreposição de visualizações em camadas, através da combinação de várias camadas de visualização, empilhadas uma sobre a outra, compõe-se uma única visão. A Figura 3.2 apresenta a sobreposição de dois gráficos de dispersão.



**Figura 3.2** Sobreposição de duas visualizações em camadas. Dois gráficos de dispersão A e B empilhados, resultando em uma única visualização C. Adaptado de Munzner (2014).

O desafio das visualizações justapostas é a área de tela necessária para mostrar essas várias visões lado a lado. A área de exibição é um recurso escasso. Em contrapartida, a sobreposição de camadas não requer mais espaço na tela. Uma limitação séria da sobreposição é o número de camadas que podem ser sobrepostas uma sobre a outra para que a desordem visual não se torne grave: duas camadas são muito viáveis, três são possíveis, mas, com cuidado, e mais do que isso seria difícil. Em contraste, a justaposição pode acomodar um número maior de visualizações, onde várias são simples e até algumas dezenas são viáveis, porém, com cuidado (MUNZNER, 2014).

No contexto de múltiplas visões, Roberts (2007) comenta que sistemas que usam apenas duas visualizações justapostas (lado a lado), geralmente são chamados de sistemas de visão dupla. E Scherr (2008) cita algumas variações desse tipo de sistema:

- Visualizações com visão geral e detalhes - usam uma visualização para exibir todo ou uma parte maior do conjunto de dados e outra visão para mostrar uma parte menor dos dados com mais detalhes.
- Visualizações de foco e contexto - é outra metodologia para dar detalhes em uma área e mostrar o contexto dessa informação em outra.
- Visualizações de Diferença - são usadas para mostrar ou destacar as diferenças nos dados observados. Geralmente isto é feito através da fusão de duas ou mais visões.
- Pequenos Múltiplos - são várias visualizações em miniatura, organizadas e representam dados de atributos.

Nas múltiplas visões pode-se utilizar representações de dois tipos: 1) uma única forma, ou seja, várias visualizações do mesmo tipo na mesma tela, ou 2) multiforme, quando se utiliza dois ou mais tipos diferentes de visualizações na mesma exibição. Através de múltiplas visões multiforme o usuário pode facilmente comparar os dados entre duas ou mais representações (ROBERTS, 2007). As múltiplas visões não são exibições separadas ou isoladas sobre os dados. Mas, são ferramentas poderosas que geralmente compartilham algum relacionamento coordenado entre si (SCHERR, 2008). Conforme discutiremos na próxima seção.

### 3.4 COORDENAÇÃO

No contexto de MVC, o termo ‘coordenadas’ significa que as interações ou operações realizadas em uma visualização são refletidas nas outras. Portanto, as alterações feitas em uma visão são propagadas para as demais, mantendo os dados analisados consistentes (ROBERTS, 2007; ALMEIDA et al., 2009). A coordenação em múltiplas visões é um mecanismo implícito ou escondido, e se manifesta quando o usuário exerce alguma interação com a visualização. A partir disso, as ações são replicadas em outras visões (SCHERR, 2008). Para isso acontecer, as representações gráficas precisam estar conectadas de alguma forma.

North e Shneiderman (2000) mencionam que MVC podem ser classificadas segundo seu nível de flexibilidade em três características essenciais: 1) dados - permitir ao usuário utilizar mais de um conjunto de dados; 2) visualização - possibilitar ao usuário selecionar o tipo de representação visual apropriado para o conjunto de dados; 3) coordenação - oferecer ao usuário a opção de escolher o tipo de coordenação (recursos de interação) entre cada par de visões.

Pillat (2006) explica que o relacionamento ou a coordenação entre as múltiplas visões podem acontecer em dois sentidos: 1) unidirecional - levando em consideração duas visualizações conectadas, apenas uma visão pode refletir as interações do usuário para a outra visão, ou seja, apenas um sentido; e 2) bidirecional ( $\Leftrightarrow$ ) - ou duplo sentido, as alterações feitas em qualquer uma visualização reflete na outra.

North e Shneiderman (1997) apresentaram a taxonomia 2x3 que organiza as estratégias de coordenação para conectar as múltiplas visualizações. As visões são compostas

por representações visuais e um conjunto de dados, e ambos estão associados a alguma das seguintes ações do usuário:

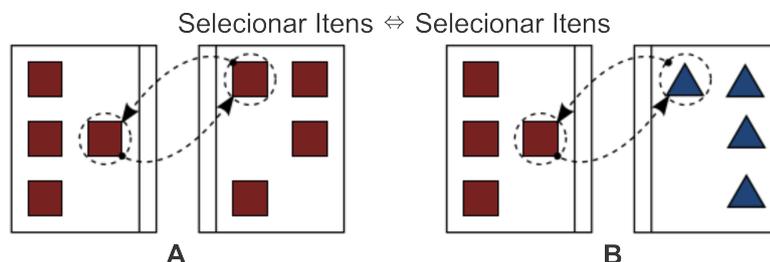
- Selecionar itens - ao usuário manipular diretamente uma técnica de visualização, uma ação fundamental nesse processo é a seleção de itens mostrados na tela;
- Navegar pelas visualizações - a visualização pode representar uma parte dos dados e permitir que os usuários naveguem na exibição para ver outras partes.

North e Shneiderman (1997) consideram que os itens de dados representados nas visualizações coordenadas podem ser iguais ou diferentes. Por exemplo:

- Mesmo conjunto de itens de dados - a coordenação de múltiplas visões da mesma coleção de dados permite a navegação simultânea em vários aspectos diferentes das informações. Nesse caso, existe uma relação correspondente e implícita entre os dados de cada visualização;
- Conjuntos diferentes de itens de dados - as coleções de dados representados em cada visualização não são iguais, porém são relacionadas. Por exemplo, um documento de texto e um dicionário proporcional.

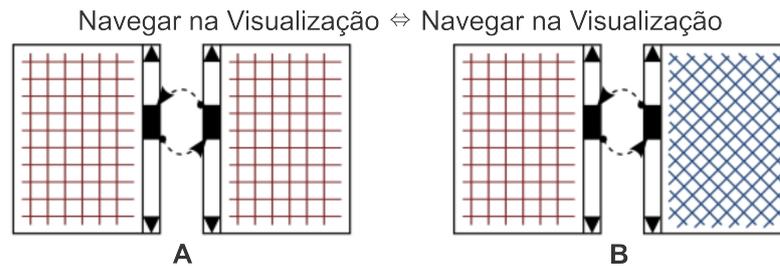
Levando em consideração um par de visualizações, a taxonomia 2x3 de North e Shneiderman (1997) também sugere a possibilidade de realizar as seguintes combinações de estratégias de coordenação de maneira bidirecional:

- Selecionar itens  $\Leftrightarrow$  Selecionar itens - essa coordenação conecta a seleção de itens em várias visualizações. Um exemplo dessa combinação de coordenação é mostrado na Figura 3.3. Esta imagem mostra dois pares de visões coordenadas. Do lado esquerdo 'A' com duas visões do mesmo conjuntos de dados e do lado direito 'B' com um par de visualizações de conjuntos de dados diferentes, porém, com alguma relação entre eles. Em ambos os casos, ao usuário selecionar um item em uma visão, os itens correspondentes na visão vinculada também são selecionados.



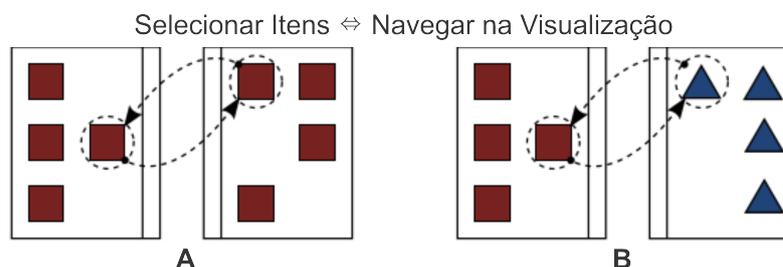
**Figura 3.3** Combinação da coordenação Selecionar itens de forma bidirecional. Ao selecionar itens em qualquer uma visão, os itens da outra visão também são selecionados. Adaptado de North e Shneiderman (1997).

- Navegar na visualização  $\Leftrightarrow$  Navegar na visualização - essa coordenação vincula a navegação de uma visualização à navegação de outra visualização simultaneamente. A Figura 3.4 ilustra esse tipo de combinação de coordenação. Essa imagem mostra dois pares de visões coordenadas. Do lado esquerdo 'A' com duas visões do mesmo conjunto de dados e do lado direito 'B' com um par de visualizações de conjuntos de dados diferentes, porém, com alguma relação entre eles. Em ambos os casos, ao mover a barra de rolagem de uma visão, para navegar pela representação dos dados, a outra visão acompanhará o movimento automaticamente.



**Figura 3.4** Combinação da coordenação Navegar na visualização de forma bidirecional. Ao movimentar qualquer uma das visões, o deslocamento para a mesma direção será refletido na outra visão. Adaptado de (NORTH; SHNEIDERMAN, 1997).

- Selecionar itens  $\Leftrightarrow$  Navegar na visualização - essa coordenação combina a seleção de itens em uma janela com a navegação em outra janela ou vice-versa. Uma amostra desse tipo de combinação é apresentado na Figura 3.5. Essa imagem mostra dois pares de visões coordenadas. Do lado esquerdo 'A' com duas visões do mesmo conjunto de dados e do lado direito 'B' com um par de visualizações de conjuntos de dados diferentes, porém, com alguma relação entre eles. Em ambos os casos, ao selecionar um item em qualquer uma visão, os itens correspondentes serão selecionados na outra visão, e o usuário poderá navegar entre eles.



**Figura 3.5** Combinação da coordenação selecionar itens de forma bidirecional. Os usuários podem selecionar itens no contexto de uma visão para localizar e visualizar informações correspondentes em outra visão. Adaptado de North e Shneiderman (1997).

Neste sentido, Munzner (2014) também apresentou uma taxonomia das principais formas para criar conexão entre múltiplas visualizações:

- Codificação visual compartilhada: ocorre quando um canal visual é usado da mesma maneira em várias visões. Um exemplo é uma visualização multiforme, quando duas exibições usam diferentes técnicas de visualização, mas com o mesmo mapeamento de cores. Essa forma de vinculação não exige necessariamente interatividade e pode ser feita com visualizações completamente estáticas.

A Figura 3.6 mostra duas visualizações coordenadas que exemplificam o compartilhamento de codificação visual. Quando o usuário selecionar alguma parte do gráfico de barras (A), os pontos do gráfico de dispersão (B) que mapeiam os mesmos dados serão destacados com a cor vermelha. O mesmo acontece para os gráficos de dispersão C e D da mesma imagem, porém, utilizando a cor verde.



**Figura 3.6** Exemplo de múltiplas visualizações compartilhando a codificação visual. Neste caso, o canal de cor das visões está conectado. Adaptado de Munzner (2014).

Ward, Grinstein e Keim (2015) afirmam que outras formas de codificação visual incluem aquelas que modificam o mapa de cores, o tamanho das entidades gráficas e suas formas. Estas podem ser usadas para enfatizar ou revelar características de interesse do analista. Mesmo as limitações de algumas técnicas de visualização podem ser superadas usando essas variações.

- Quantidade de dados compartilhados: existem três alternativas, supondo que nosso sistema tenha duas visões com dados compartilhados: 1) ambas as exibições podem mostrar todos os dados; 2) com visão geral-detalhes, onde uma visão poderia mostrar um subconjunto do que está na outra; 3) com pequenas múltiplas visões, em que as visões poderiam mostrar partições diferentes do conjunto de dados.

A Figura 3.7 exibe três pares de visões coordenadas que exemplificam como particionar dados entre as visualizações. No primeiro par de visões ‘A’ representa um gráfico de dispersão e ‘B’ um gráfico de barras, ambos, visualizando todo o conjunto de dados. O segundo par de visões ‘C’ e ‘D’ mostram dois gráficos de dispersão visualizando partes do mesmo subconjunto dos dados. O par de representações gráficas ‘E’ e ‘F’ exibem dois gráficos de dispersão visualizando partes diferentes da base de dados.

- Navegação compartilhada: essa conexão permite movimentar uma visualização, e esse movimento ser sincronizado para as outras visualizações. A Figura 3.8 apresenta um par de visualizações multiforme, onde, ao movimentar o gráfico de dispersão (A) para a direita, o gráfico de barras (B) se movimenta para a mesma direção automaticamente.



**Figura 3.7** Exemplo de três pares de visualizações compartilhando: A e B toda base de dados; C e D mesmo subconjunto do *dataset*; E e F subconjuntos diferentes da coleção de dados. Adaptado de Munzner (2014).



**Figura 3.8** Exemplo de navegação compartilhada entre MVC. Adaptado de Munzner (2014).

- **Combinação:** em um sistema de visualização, pelo menos um par de visões pode compartilhar a codificação visual e os dados. A Figura 3.9 exhibe as possibilidades de combinar o compartilhamento de codificação visual com o compartilhamento dos dados. A imagem tem dois quadrantes que destacam situações em que: *i*) há redundância na conexão entre as visões. Técnicas iguais para visualizar partes iguais do conjunto de dados é redundante; e *ii*) não há conexão entre as visões. Técnicas diferentes para visualizar partes diferentes do conjunto de dados sem compartilhar nenhuma codificação visual, não cria conexão entre as múltiplas visualizações.

		Dados		
		Todos	Subconjunto	Nenhum
Codificação	Igual	Redundante	Visão Geral Detalhes	Pequenos Múltiplos
	Diferente	Multiforme	Visão Geral Detalhes	Não Conectado

**Figura 3.9** Exemplos de possíveis combinações de compartilhamento de codificação visual e de dados. Adaptado de Munzner (2014).

Existe uma concordância entre os autores sobre as formas de criar algum tipo de

conexão entre as múltiplas visões e torná-las coordenadas. Coordenação é uma ferramenta que provê mecanismos interativos que possibilitam ao usuário manipular com facilidade as representações gráficas, com o objetivo de perceber características e as relações do conjunto de dados (ALMEIDA et al., 2009). Muitas limitações de visualizações, como sincronização, correlação de informações visuais ou não visuais, oclusão, explosão de visão e multitarefa, podem ser corrigidas ou amenizadas por meio de coordenação entre as visões (BOUKHELIFA; ROBERTS; RODGERS, 2003).

Estudos realizados por North e Shneiderman (2000) identificaram que utilizar coordenação em visualização aumenta a performance da exploração do analista em até 80% de acordo a tarefa realizada. E Yi et al. (2007) acrescenta que o componente de interação beneficia o diálogo entre o usuário e o conjunto de dados durante a análise, pois além de alguns limites das representações serem superados, a cognição do usuário pode ser ampliada. A próxima seção reúne mais informações sobre estes recursos de interação, incluindo suas classes e estilos.

### 3.5 INTERAÇÃO

Os sistemas de visualização fornecem uma variedade de interações que permitem ao usuário: *i*) selecionar a parte dos itens de dados que deseja analisar, e *ii*) personalizar o modo como a informação visual é apresentada. Quando múltiplas visualizações dos mesmos dados são apresentadas juntas, a coordenação entre as visões pode ser controlada através de estruturas visuais interativas ou técnicas de interação. Alguns exemplos são: *i*) a alteração da perspectiva de elementos gráficos; *ii*) zoom com diferentes níveis de detalhe; *iii*) e a manipulação de outros parâmetros da exibição para obter as informações necessárias para entender os comportamentos das coleções de dados (GAO et al., 2011).

As interações dentro do contexto de visualização de informação são instrumentos para modificar o que os usuários veem e como eles o veem (WARD; GRINSTEIN; KEIM, 2015). A interação é uma parte essencial da visualização de informação, pois sem ela, uma técnica ou sistema de visualização produz uma imagem estática ou imagens animadas. E apesar da visualização ter seu valor analítico e expressivo, sua utilidade se torna mais limitada à medida que o conjunto de dados que ela representa, cresce. A interação desencadeia mudanças na representação que favorecem a descoberta de conhecimento (YI et al., 2007).

Na próxima seção apresentamos duas classificações das técnicas de interação, a primeira proposta por Ward, Grinstein e Keim (2015) que baseia-se nas funcionalidades das interações. E a outra, apresentada por Yi et al. (2007) com foco na intenção do usuário ao realizar uma operação interativa.

#### 3.5.1 Classes de Interação

Existe uma variedade de técnicas para o usuário realizar interações nas ferramentas de visualização. Apesar disso, os recursos interativos servem para um número reduzido de propósitos que podem ser agrupados em categorias ou classes (YI et al., 2007). A seguir, vamos apresentar duas classificações sobre esses vários mecanismos de interação, uma baseada na funcionalidade das técnicas de interação, e outra baseada na intenção do

usuário ao utilizar a interação. Ward, Grinstein e Keim (2015), agruparam as técnicas de interação em oito classes distintas, segundo as funcionalidades dos recursos interativos nas visualizações:

- Navegação - controles para que o usuário altere a posição da câmera e dimensione a visualização, ou seja, o que é mapeado para a tela. Exemplos de técnicas de interação desta classe são panorâmica, rotação e zoom.
- Seleção - permite ao usuário identificar um objeto, uma coleção de objetos ou regiões de interesse a serem alvos de alguma operação, como destacar, excluir e modificar.
- Filtrar - possibilita o usuário reduzir o tamanho dos dados que estão sendo mapeados para a tela, eliminando da visualização registros, dimensões ou ambos.
- Reconfigurar - esta classe de técnicas de interação oferece controles para o usuário alterar a maneira como os dados são mapeados para os elementos gráficos, como reordenar os dados ou *layouts*, viabilizando uma maneira diferente de visualizar um subconjunto dos dados.
- Codificação - controles para o usuário alterar os atributos visuais, como o tamanho do ponto ou cor da linha, para revelar recursos diferentes.
- Conexão - é possível vincular diferentes vistas ou objetos para mostrar itens relacionados.
- Abstrair/Elaborar - nesta classe o usuário pode modificar o nível de detalhe para mais ou para menos.
- Híbrido - Nesta categoria se enquadram as técnicas que permitem combinar várias classes em uma técnica, por exemplo, aumentando o espaço da tela de uma ou mais áreas de foco, para permitir que os usuários vejam mais detalhes, enquanto mostra as outras áreas de dados em um espaço menor que preserva o contexto.

Yi et al. (2007) realizou um estudo abrangente sobre as técnicas de interação usadas em várias áreas de aplicação. Foi identificado que para diferentes técnicas de representação, são usadas técnicas de interação distintas para executar uma tarefa semelhante ou atingir um objetivo semelhante. Assim, a pesquisa foi direcionada para o que os usuários alcançam usando as técnicas de interação, e não em como as técnicas disponíveis nos sistemas de visualização funcionavam. Como resultado do estudo, Yi et al. (2007), classificou as técnicas de interação em sete categorias, a partir do que o usuário deseja alcançar ao utilizar recursos interativos nas visualizações. As intenções detectadas foram:

- Selecionar - quando a intenção do usuário é marcar algo interessante. Nessas interações, o usuário pode selecionar um ou mais itens de dados de seu interesse para localizá-los ou acompanhá-los com facilidade, mesmo em um grande conjunto de dados ou com alterações nas representações visuais.

- Explorar - quando a intenção do usuário é ver algo mais. As técnicas desta classe permitem que um subconjunto diferente de dados seja examinado. As interações do tipo explorar não fazem alterações completas nos dados visualizados, alguns novos itens de dados são incluídos na visualização quando outros são removidos. Um exemplo é a técnica panorâmica, quando o usuário prende a cena e a move com um mouse ou simplesmente altera a visualização por meio de barras de rolagem.
- Reconfigurar - quando a intenção do usuário é ver por uma perspectiva diferente. Essa classe altera o arranjo espacial das representações. Ela permite ao usuário alterar a maneira como os itens de dados são organizados ou mudar o alinhamento dos itens de dados, a fim de fornecer perspectivas diferentes sobre o conjunto de dados.
- Codificar - quando a intenção do usuário é ver por uma representação diferente. Esse tipo de técnica permite que os usuários alterem os atributos visuais como cor, tamanho, orientação e forma de cada elemento de dados. Os elementos visuais são importantes e ajudam a cognição do analista e impactam na maneira como o observador entende os relacionamentos e as distribuições dos itens de dados.
- Resumir/Elaborar - quando a intenção do usuário é ver uma quantidade maior ou menor de detalhes. Esse tipo de interação permite ajustar o nível de abstração de uma representação de dados. O usuário pode mudar de uma visão geral para uma visão com mais detalhes dos dados. Exemplos desse tipo de técnica de interação são: ‘detalhes sob demanda’, que fornecem informações detalhadas quando o cursor do mouse passa sobre um item de dados; e zoom.
- Filtrar - quando a intenção do usuário é ver algo a partir de alguma condição. Em interações com esta característica o usuário não altera a perspectiva dos dados, apenas especifica um intervalo ou condição, de modo que apenas os elementos de dados que atendem a esses critérios sejam apresentados. Os itens de dados fora do intervalo ou que não satisfazem a condição são ocultados ou mostrados de maneira diferente.
- Conectar - quando a intenção do usuário é ver itens relacionados. Esse grupo de técnicas se aplica à i) destacar associações e relacionamentos entre itens de dados que já estão representados e ii) mostrar itens de dados ocultos que são relevantes para um item especificado. Um exemplo é o *Linking & Brushing*, usado para destacar um item de dado selecionado nas outras visões.

Nesta seção apresentamos duas classificações das técnicas de interação a partir de dois pontos de vista diferentes, uma delas proposta por Ward, Grinstein e Keim (2015), que se baseia no funcionamento das técnicas de interação para agrupá-las em classes. A outra classificação foi baseada no estudo de Yi et al. (2007) que divide as técnicas em categorias que refletem a intenção do usuário ao utilizá-las. Apesar de mudar algumas nomenclaturas e a posição em que elas são mostradas, os autores concordam em suas classificações. Ward, Grinstein e Keim (2015), ainda acrescentam uma categoria chamada Híbrido, em que eles identificaram técnicas que são a combinação de outras técnicas.

### 3.5.2 Estilos de Interação

As técnicas de interação permitem que o usuário manipule a visualização e realize alterações segundo seus objetivos de exploração dos dados. Manipulação é a ação do usuário ao manusear os recursos interativos nas visualizações. Existem dois estilos de interação, através de manipulação indireta ou manipulação direta das visões (ROBERTS, 2007). E esses estilos definem a forma como são manuseadas as representações gráficas durante a análise e exploração dos dados.

**Interação indireta** - As interações realizadas através de manipulação indireta são feitas por meio de componentes integrados nas visualizações como controles deslizantes (*sliders*), menus e botões. A partir desses componentes é possível filtrar os dados e mudar como as informações são exibidas. Um exemplo de manipulação indireta é a técnica de interação consulta dinâmica (ROBERTS, 2007). Neste estilo de interação, o analista manipula as visualizações indiretamente através de controles predefinidos na interface do usuário. Por exemplo, controles relacionados a entrada de dados, filtragem, agrupamento e outras transformações; e definições de mapeamentos visuais como cor, forma e relacionamentos de mapeamento (GAO et al., 2011).

**Interação direta** - As técnicas que usam manipulação direta permitem ao usuário selecionar elementos diretamente na visualização, através de manipuladores ou *widjets*. A técnica de *Linking & Brushing* é um exemplo, onde os elementos são selecionados e realçados em uma tela, e simultaneamente as mesmas informações também são destacadas em qualquer outra tela vinculada (ROBERTS, 2007). Outras formas de interação direta incluem clicar em um componente visual para apresentar uma visão mais detalhada, a exibição panorâmica, selecionar objetos gráficos e reposicioná-los. (GAO et al., 2011).

### 3.5.3 Exemplos de Interação em Múltiplas Visões Coordenadas

Para Boukhelifa, Roberts e Rodgers (2003), a essência de se utilizar visualização em exploração de dados é a busca por informação visual. Keim (1997) complementa que fazer isso de maneira dinâmica ou interativa, adicionando técnicas de interação nas visualizações torna o processo exploratório mais efetivo. As técnicas de interação são os mecanismos que o usuário pode manipular para realizar mudanças nas representações visuais. Nesse contexto, Keim (2002) diferencia os termos dinâmico e interativo. Dinâmico tem a ver com as alterações feitas automaticamente nas visualizações, enquanto que interativo se refere as alterações feitas manualmente nas visualizações, por interação do usuário.

Para implementar visualizações coordenadas, é necessário um ambiente interativo com mais de uma visualização. A coordenação é um mecanismo de interação e, portanto, requer um sistema onde o usuário possa interagir com representações gráficas de dados (WILLS, 2008). Munzner (2014) enfatiza que os recursos de interação liberam o poder total das visualizações coordenadas. Para Baldonado, Woodruff e Kuchinsky (2000) a coordenação pode ser usada com o objetivo de propagar as interações realizadas pelo usuário para todas as visualizações vinculadas.

Técnicas que possam contribuir no processo de exploração de dados com alta dimensionalidade é um desafio. E dentre as interações que podem ser usadas em múltiplas visões coordenadas, Pillat (2006) lista quatro categorias de interações (neste contexto, também

chamada de coordenações) que se aplicam a técnicas de visualização multidimensional:

- Mapeamento interativo de dados para visualização - mapeamento dinâmico ou interativo dos atributos (dimensões) de dados para parâmetros de uma visualização. Esses parâmetros são os eixos (x, y e z) e os elementos utilizados nas representações visuais como cor, tamanho, orientação, transparência, entre outros. Esses atributos visuais ampliam o número de dimensões de dados que podem ser mapeados em uma visão (KEIM, 1997; PILLAT, 2006).
- Reordenação de Dimensões - as técnicas de visualização multidimensional suportam alguma forma de reordenação de dimensões (atributos). Esse tipo de interação (coordenação) permite que um usuário possa mudar a disposição em que atributos de dados aparecem nas visualizações (PILLAT, 2006).
- Filtragem - ao investigar grandes conjuntos de dados, é necessário criar partições de forma interativa e analisar nos subconjuntos relevantes. Isso pode ser feito por meio de seleção direta do subconjunto desejado (navegação) ou especificando propriedades do subconjunto desejado (consulta). No caso da coleção dos dados ser muito grande a navegação se torna difícil e a consulta pode não produzir os resultados esperados. Para suprimir essas limitações foram criadas outras técnicas para melhorar a filtragem interativa na exploração dos dados (KEIM, 2002).
- Consulta Dinâmica - um exemplo de técnica de filtragem interativa aplicada em MVC que trouxe um novo conceito para interagir com bases de dados. Este recurso permite aos usuários realizarem consultas a partir do manuseio de *widgets* gráficos, como botões deslizantes (*sliders*). O resultado da consulta ajuda o usuário a encontrar tendências e exceções com facilidade (AHLBERG; WILLIAMSON; SHNEIDERMAN, 1992).
- *Linking & Brushing* ou Seleção - A ideia é combinar diferentes métodos de visualização dos mesmos dados para superar as deficiências das técnicas individuais. Os pontos selecionados em uma visão são destacados em todas as visualizações, possibilitando a detecção de dependências e correlações. Aplicar *Linking & Brushing* interativo em múltiplas visões coordenadas fornece mais informações do que considerar as visualizações dos dados de forma independente ou separadas (KEIM, 2002).

O Zoom é uma técnica de interação muito conhecida e utilizada. Ao lidar com a visualização de grande quantidade de dados é necessário apresentar os dados em uma forma reduzida para fornecer uma visão geral e ao mesmo tempo permitir que a exibição seja modificada para diferentes resoluções. O zoom não significa apenas exibir os objetos de dados em tamanhos maiores, mas também significa que a representação dos dados mude para apresentar mais detalhes em níveis de zoom mais altos (KEIM, 2002).

Outra interação aplicada a dados multidimensionais utilizada em MVC é a técnica de detalhes sob demanda. Que permite obter interativamente mais detalhes dos dados visualizados. Esses detalhes são, por exemplo, os valores ou informações adicionais dos

atributos correspondentes a um item de dado. Essa técnica é uma alternativa rápida, onde o usuário pode a qualquer momento, durante o processo de exploração, selecionar um item na visualização, e os detalhes relacionados ao elemento selecionado são exibidos em um *frame* ou *pop-up* (KEIM, 1997; PILLAT, 2006; ALMEIDA et al., 2009).

### 3.6 CONSIDERAÇÕES FINAIS

Este capítulo apresentou como Múltiplas Visões Coordenadas podem lidar com o desafio de representar visualmente e explorar bases de dados complexas e volumosas. As múltiplas visões coordenadas permitem criar uma conexão entre as visualizações, e sugerem formas de particionar os dados entre as visões. Nesse sentido, foram apresentados diversos conceitos e termos relevantes. Além disso, vincular ou coordenar visualizações também contribui para superar as restrições bidimensionais das telas. Nessa perspectiva, foram descritas maneiras de particionar o espaço de tela para justapor ou sobrepor as visões.

Conforme foi abordado, a interação torna o processo exploratório mais efetivo. Através das técnicas de interação o usuário manipula as representações visuais. Dentre a diversidade de técnicas de interação, duas taxonomias foram apresentadas: uma segundo a funcionalidade da técnica e outra de acordo a intenção do usuário ao utilizá-la.

Por fim, também foi citado sobre os estilos dos mecanismos interativos das visões. As manipulações que o usuário opera nas visualizações podem ser feitas de duas formas: *i*) selecionando itens de dados diretamente nas visualizações, ou *ii*) por meio da manipulação indireta, através de botões nas interfaces das visões para filtrar dados.



## TRABALHOS RELACIONADOS

### 4.1 CONSIDERAÇÕES INICIAIS

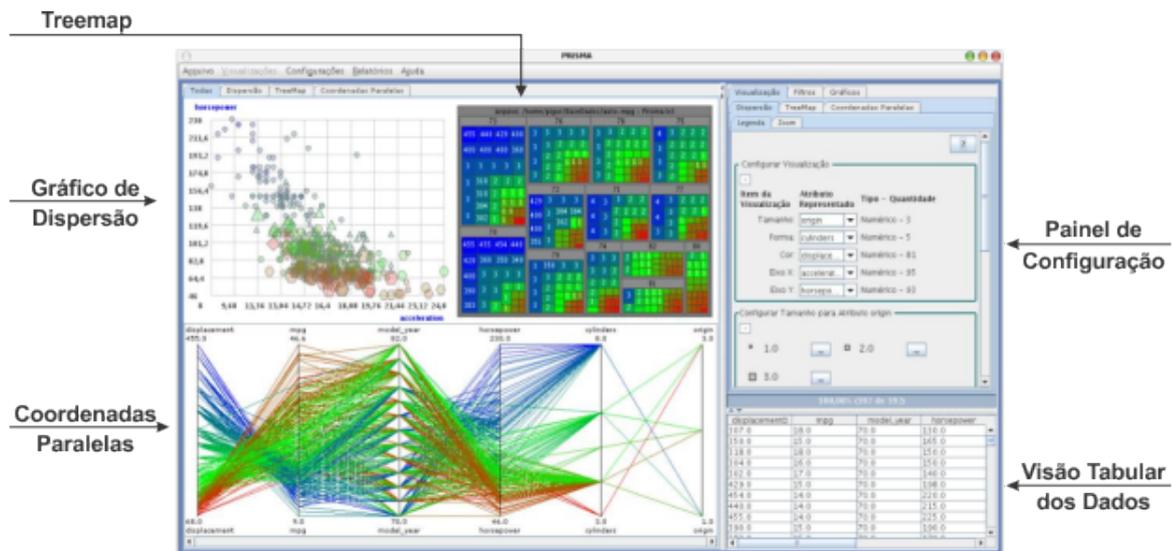
No capítulo anterior vimos que visualizações podem conter Múltiplas Visões Coordenadas, sendo estas uma importante solução para visualização exploratória de dados. Essa abordagem visa utilizar várias visões para um mesmo ou diferentes conjuntos de dados. Além disso, podem estar dispostas de diferentes modos e conter vários tipos de interações capazes de serem refletidas nas outras visões.

Nessa estratégia, diversas técnicas de visualização podem ser empregadas como visões. Dependendo do tipo de informação que se deseja extrair e, segundo as características como tipos dos dados e quantidade de dimensões, por exemplo. Nesse sentido, podem ser utilizadas técnicas tradicionais de visualização de dados, que geralmente mapeiam poucas dimensões, ou técnicas de visualizações multidimensionais, que suportam um número maior de dimensões. Neste capítulo serão apresentados trabalhos que desenvolveram técnicas com Múltiplas Visões Coordenadas para extrair informações nos mais diferentes tipos de domínios de dados.

### 4.2 VISUALIZAÇÃO MULTIDIMENSIONAL

Para melhor entender a ideia de Múltiplas Visões Coordenadas que contenham visualização multidimensional, a Figura 4.1 apresenta a tela principal da aplicação PRISMA (GODINHO et al., 2007). Nesta, é possível observar diversos componentes que fazem parte de um sistema de visualizações coordenadas, dentre eles: *i)* o painel de configuração (lado direito), que permite alterar diversos parâmetros nas visualizações como cor, tamanho, forma e eixos (x e y) em que os atributos serão mapeados. Além de ajustes na legenda, zoom, filtros e os gráficos auxiliares que irão representar seus resultados; *ii)* uma visão tabular dos dados (lado direito), que mostra os valores dos itens de dados; *iii)* e, por fim, no lado esquerdo, as técnicas de visualização multidimensionais *Treemap* e *Coordenadas Paralelas*, e um Gráfico de Dispersão tradicional com duas dimensões.

As visualizações multidimensionais quando utilizadas em Múltiplas Visões Coordenadas, auxiliam na extração de informação em diferentes áreas de domínios de dados, tais



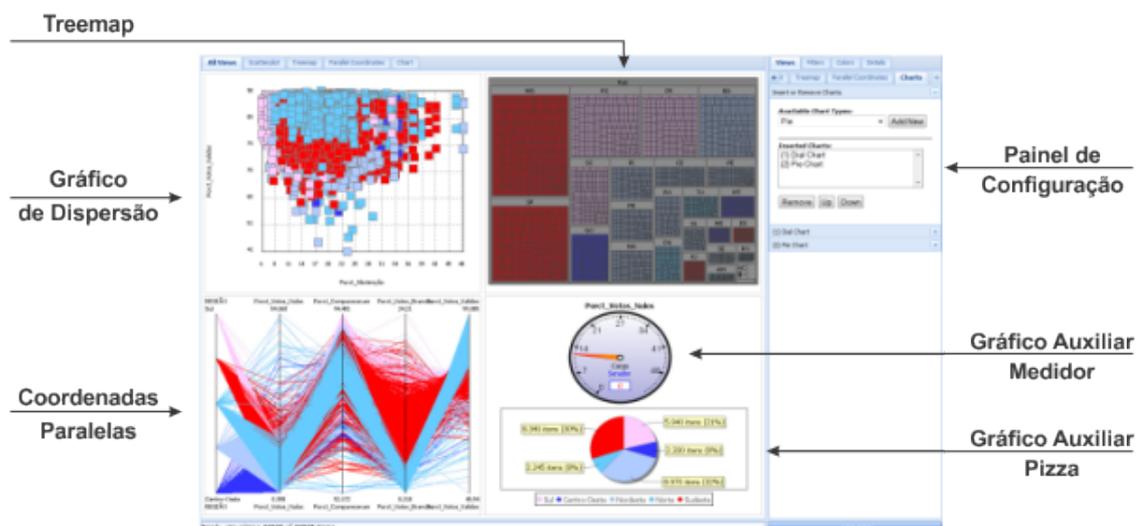
**Figura 4.1** Exemplo de interface com Múltiplas Visões Coordenadas. Adaptado de Godinho et al. (2007).

como: política (ALMEIDA et al., 2009), futebol e vendas online (MACIEL et al., 2008), indicadores sociais e econômicos (GUIMARÃES et al., 2009), morfometria craniana (PILLAT, 2006), fornecimento e uso de energia (JERN et al., 2007), veículos (GODINHO et al., 2007; PILLAT; FREITAS, 2006), tráfico de rede (KAUER et al., 2008) e dados meteorológicos (MATKOVIC et al., 2008). A seguir, esses trabalhos serão apresentados e detalhados de acordo com seus domínios de aplicação.

### 4.3 DOMÍNIO DE APLICAÇÃO DE DADOS

**Dados políticos** - Almeida et al. (2009) aplicou Múltiplas Visões Coordenadas na ferramenta chamada WebPrisma. Ela foi utilizada para explorar e extrair informações de uma coleção de dados multidimensionais das eleições de 2006 no Brasil. A técnica de visualização multidimensional Coordenadas Paralelas foi aplicada em conjunto com outras visões contendo *Treemap* e Gráfico de Dispersão. A aplicação também dispõe de uma visão para exibir características adicionais dos dados através de gráficos auxiliares como barra, pizza, linha e medidores. A Figura 4.2 exibe a interface do WebPrisma plotando Múltiplas Visões Coordenadas.

A base de dados explorada é composta por 27.825 itens e 20 atributos, sendo 8 categóricos e 12 numéricos. Os atributos incluem, por exemplo, cidade, região, cargo político e o número de votos (válidos, em branco e inválidos). Durante a análise dos dados foi percebido que o percentual de votos inválidos para o cargo de Deputado Estadual, por exemplo, foi maior nos estados Bahia (Nordeste), Maranhão (Nordeste) e Goiás (Centro-Oeste). E também foi observado que as regiões com o maior comparecimento dos eleitores nas urnas foram Sul, Sudeste, Nordeste, Centro-Oeste e Norte, respectivamente.



**Figura 4.2** Tela de um cenário de exploração da aplicação WebPrisma. Adaptado de Almeida et al. (2009).

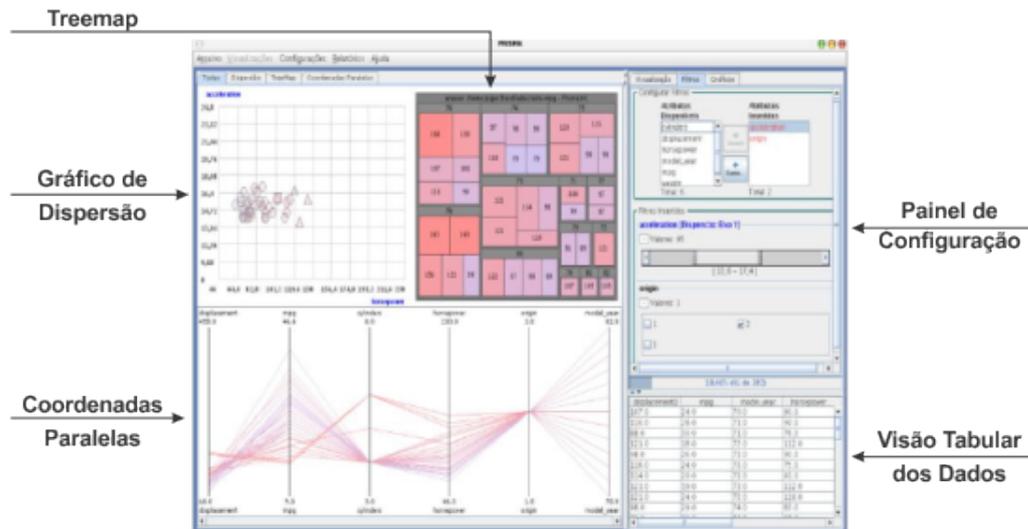
**Dados de veículos** - Godinho et al. (2007) e Pillat e Freitas (2006) utilizaram um *dataset* multidimensional contendo características de carros como peso, potência, origem (Americano, Japonês, Europeu), ano (1970 a 1982), entre outras, totalizando 8 dimensões e 392 instâncias.

A Figura 4.3 mostra um cenário construído na aplicação PRISMA de Godinho et al. (2007), e destaca as partes principais da interface: *i*) as técnicas de visualização Gráfico de Dispersão (lado esquerdo), Coordenadas Paralelas (parte inferior esquerda) e *Treemap* (parte superior); *ii*) um painel de configuração (lado direito); *iii*) e uma visão tabular dos dados para mostrar os valores dos itens de dados selecionados.

Nesse cenário, foram aplicados dois mecanismos de filtro: *i*) seleção de atributo - onde o usuário selecionou dois atributos, aceleração e origem; *ii*) conjunto de valores ou intervalo - foi definido um intervalo de valores a serem filtrados. Para o atributo aceleração o intervalo 13,6 - 17,4. E para o atributo origem todos os itens de dados com valor igual a dois. Na interface, os itens de dados com filtro ativo foram destacados em vermelho. Esse recurso ajuda a reduzir a sobrecarga visual para o usuário, quando o conjunto de dados possui muitos atributos e/ou instâncias.

**Dados de futebol e vendas online** - Maciel et al. (2008) utilizou a ferramenta PRISMA (GODINHO et al., 2007), que inclui as técnicas Coordenadas Paralelas, *Treemap* e Gráfico de Dispersão. Os conjuntos de dados visualizados foram *FIFA World Cup* (futebol) e outro sobre vendas de produtos eletrônicos (vendas online), ambos com dados categóricos e numéricos. No trabalho não consta informações adicionais sobre a base de dados, como número de atributos, instâncias, por exemplo.

O estudo analisou duas abordagens de Múltiplas Visões Coordenadas: *i*) visualizações mostradas em telas separadas e *ii*) visualizações mostradas na mesma tela. Ao comparar



**Figura 4.3** Interface da aplicação PRISMA explorando a base de dados carros com Múltiplas Visões Coordenadas. Adaptado de Godinho et al. (2007).

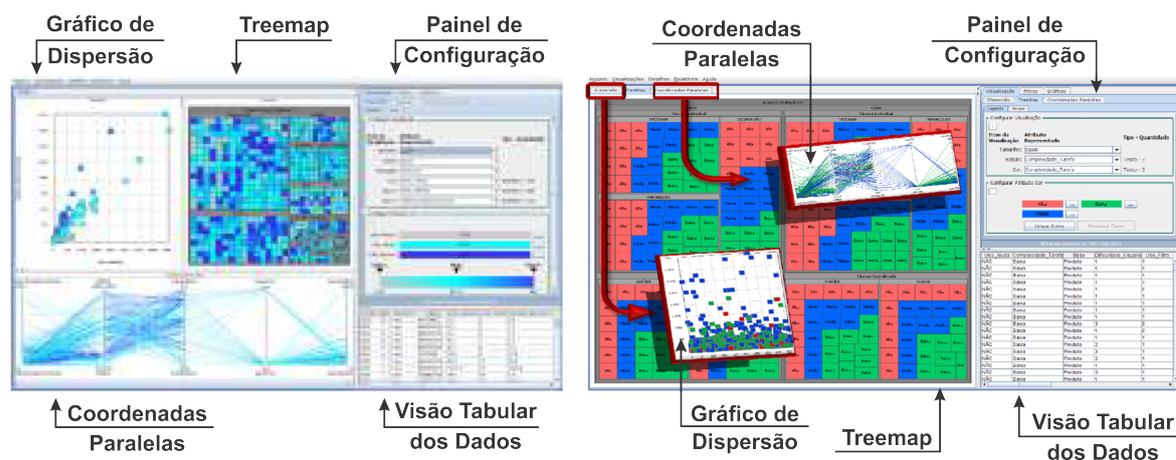
os acertos dos usuários em relação as tarefas propostas, a taxa foi maior quando as visualizações foram apresentadas na mesma tela. Os experimentos realizados tinham o objetivo de avaliar o impacto que as visualizações coordenadas causam na análise e entendimento dos dados.

Foram propostas tarefas de baixa, média e alta complexidade para os usuários realizarem, exigindo explorar as visualizações, extrair conhecimento dos dados e responder questões. Cada tarefa envolvia analisar uma quantidade de atributos, segundo seu nível de complexidade: um ou dois atributos (baixa), três ou quatro (média) cinco ou mais (alta). O estudo concluiu que para exploração e análise de bases de dados com cinco atributos ou mais, as Múltiplas Visões Coordenadas são mais úteis.

A Figura 4.4 exhibe as duas abordagens de Múltiplas Visões Coordenadas utilizada por Maciel et al. (2008). Do lado esquerdo, as visualizações coordenadas são exibidas na mesma tela, o que facilita a comparação dos resultados das interações do usuário. Enquanto que do lado direito, as visões coordenadas ficam em telas diferentes, exigindo que o analista alterne entre as janelas para ter acesso a técnica que deseja visualizar.

**Dados de morfometria craniana** - Pillat (2006) desenvolveu uma aplicação baseada no *framework toolkit InfoVis*, que suporta coordenar mecanismos interativos entre as visualizações. Esta ferramenta foi aplicada em um estudo com usuários para avaliar o uso e contribuições de Múltiplas Visões Coordenadas na exploração de dados multidimensionais. O conjunto de dados utilizado contém 20 atributos e 157 instâncias, organizados em: *i*) 17 atributos numéricos, correspondentes a medidas do crânio e o número de cromossomos; *ii*) três atributos categóricos como código, local e sexo.

Os usuários especialistas no domínio de dados morfométricos, utilizaram a aplicação com Múltiplas Visões Coordenadas no intuito de obter conhecimento sobre a forma de



**Figura 4.4** Lado esquerdo, Múltiplas Visões Coordenadas na mesma tela. Lado direito, visualizações coordenadas em telas separadas. Adaptado de Maciel et al. (2008).

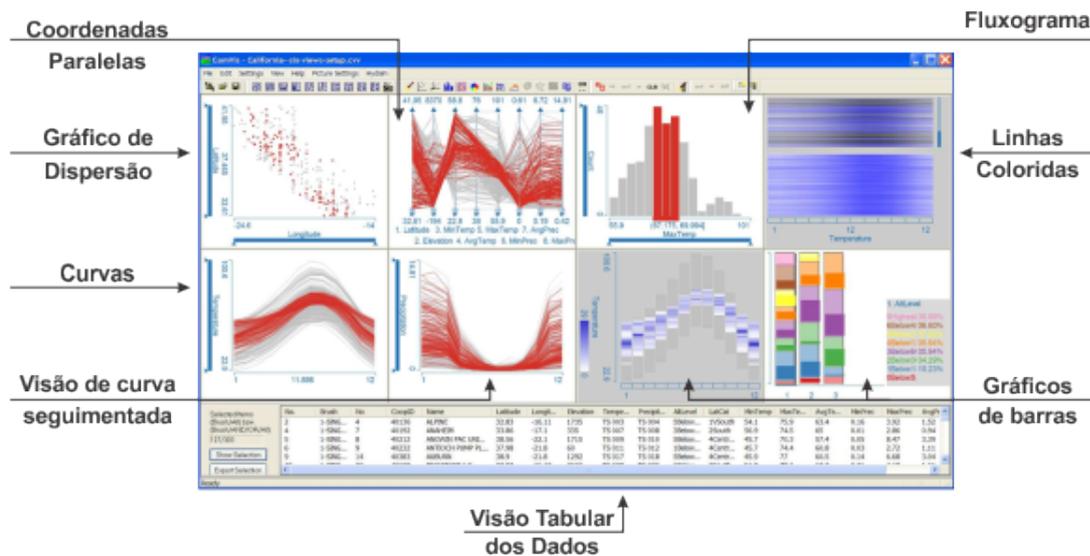
determinadas estruturas, por meio das medidas de crânios de animais. Três visualizações multidimensionais foram disponibilizadas na ferramenta: Coordenadas Paralelas, RadViz e matriz de gráficos de dispersão. Com base na análise dos dados, os usuários conseguiram identificar: crânios com medidas fora do padrão; variações nos padrões de medidas de diferentes localidades; variações nas medidas de animais com número de cromossomos diferente; e variações nas medidas de animais de sexo diferente.

Os participantes dos experimentos destacaram que a técnica de visualização Coordenadas Paralelas foi extremamente útil para a identificação rápida de crânios com medidas fora do padrão e/ou medidas com diferenças significativas entre as amostras. Essas ocorrências também podem evidenciar erros na coleta das medidas, por isso, detectá-las de forma rápida é muito importante.

**Dados meteorológicos** - Matkovic et al. (2008) criou uma ferramenta de pesquisa chamada ComVis que permite a criação de protótipos para testes de novas técnicas e paradigmas de análise visual interativa. Esta ferramenta suporta Múltiplas Visões Coordenadas contendo dados multidimensionais. O conjunto de dados explorado contém atributos categóricos e numéricos de 303 estações meteorológicas da Califórnia. Atributos como longitude, latitude, e temperaturas médias de cada estação foram analisados. Não foram mencionados os números de atributos e instâncias utilizados.

A Figura 4.5 mostra a interface da ferramenta ComVis representando dados de estações meteorológicas. As representações visuais utilizadas foram Gráfico de Dispersão, Coordenadas Paralelas, Fluxograma, Gráfico de Linhas, Visão de Curvas Segmentadas. As Visualizações Coordenadas ajudaram os usuários especialistas a observar a curva de temperatura das estações meteorológicas.

**Dados de tráfego de rede** - Kauer et al. (2008) utilizou a aplicação PRISMA, baseada em Múltiplas Visões Coordenadas para avaliar as transações em um conjunto de logs de um servidor de rede. As visualizações utilizadas foram Coordenadas Paralelas, *Treemap*



**Figura 4.5** Interface da aplicação ComVis com Múltiplas Visões Coordenadas. Adaptado de Matkovic et al. (2008).

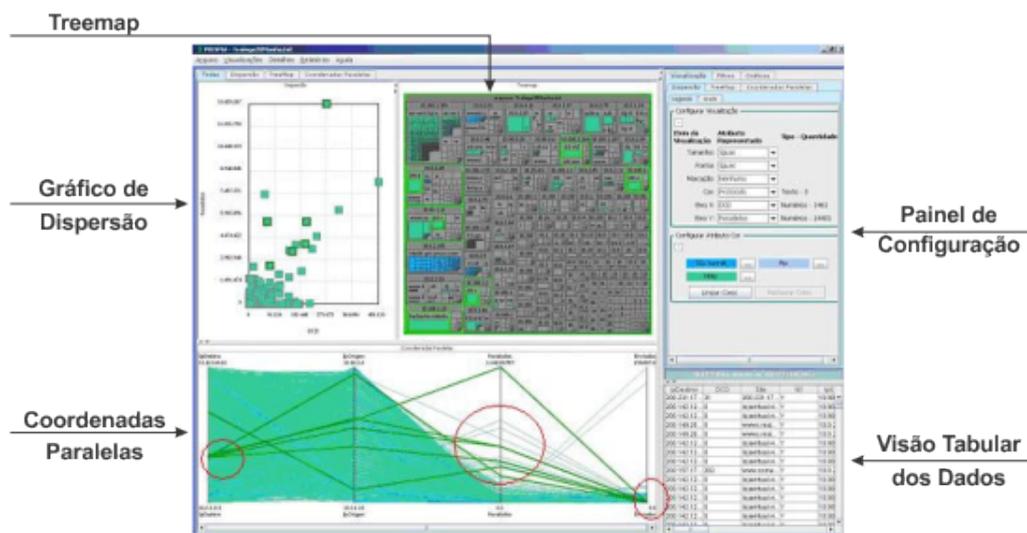
e Gráfico de Dispersão. Foram analisados 12 atributos e cerca de 105.000 registros. Contendo informações sobre IP de origem e destino, usuário, bytes enviados e recebidos, entre outros.

A Figura 4.6 mostra um dos cenários de exploração do PRISMA identificando grupos de sites que geram muito tráfego. A partir da análise da base de dados foi possível conhecer o comportamento dos usuários conectados a rede. Esta estudo permitiu criar mecanismos para melhorar a consciência das pessoas que utilizam os serviços de rede e aumentar a segurança das informações da empresa.

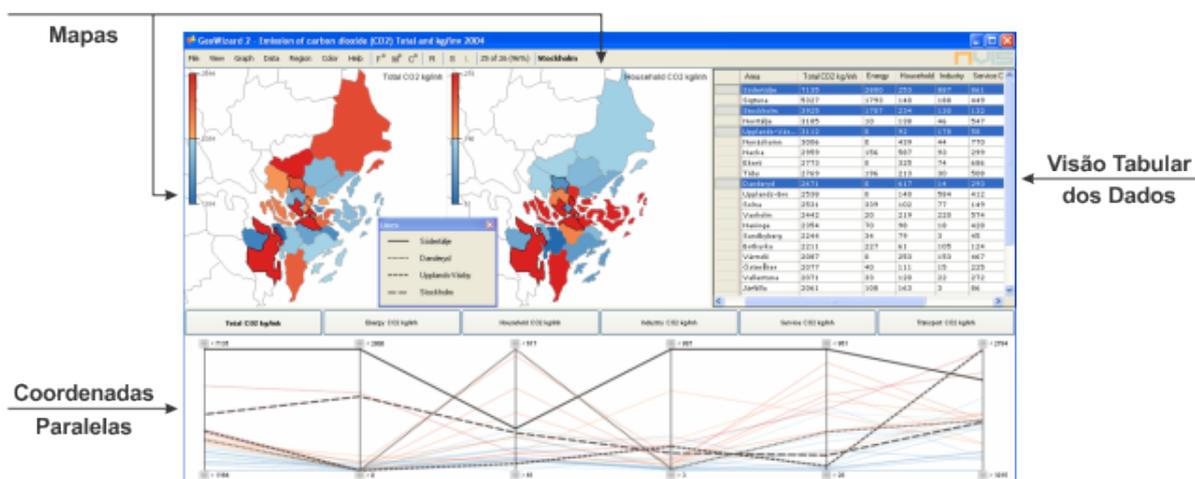
**Dados de geo-visualização** - Jern et al. (2007) desenvolveu uma ferramenta aplicada a geo-visualização, chamada GeoWizard. Essa aplicação foi utilizada para analisar dados multidimensionais do domínio de fornecimento e uso de energia em municípios suecos, em específico para controle da emissão de dióxido de carbono entre 290 regiões suecas durante o período de 1990 a 2004.

A Figura 4.7 apresenta a tela do GeoWizard contendo visões coordenadas que auxiliam o usuário a acompanhar as características de emissão de  $CO_2$  de uma localidade através das visualizações de Mapas (lado esquerdo e parte superior), visão tabular dos dados (lado direito superior) e Coordenadas Paralelas (parte inferior).

A interface do GeoWizard apresentada na Figura 4.7, compara a emissão total de  $CO_2$  (mapa à esquerda) com a emissão de  $CO_2$  doméstico (mapa à direita). A análise dos dados, através das visualizações, revelou informações importantes sobre quatro municípios específicos, que são destacados em todas as visões: 1) *Södertälje* tem mais problemas, com um total elevado de emissões de  $CO_2$  na indústria; 2) *Danderyd*, uma área de alta renda, tem problemas com as emissões de  $CO_2$  nas residências; 3) *Upplands-Väsby* mostra uma maior concentração relacionada com transporte; 4) e *Stockholm*, o maior município



**Figura 4.6** Múltiplas Visões Coordenadas da ferramenta PRISMA explorando dados de tráfego de rede. Adaptado de Godinho et al. (2007).



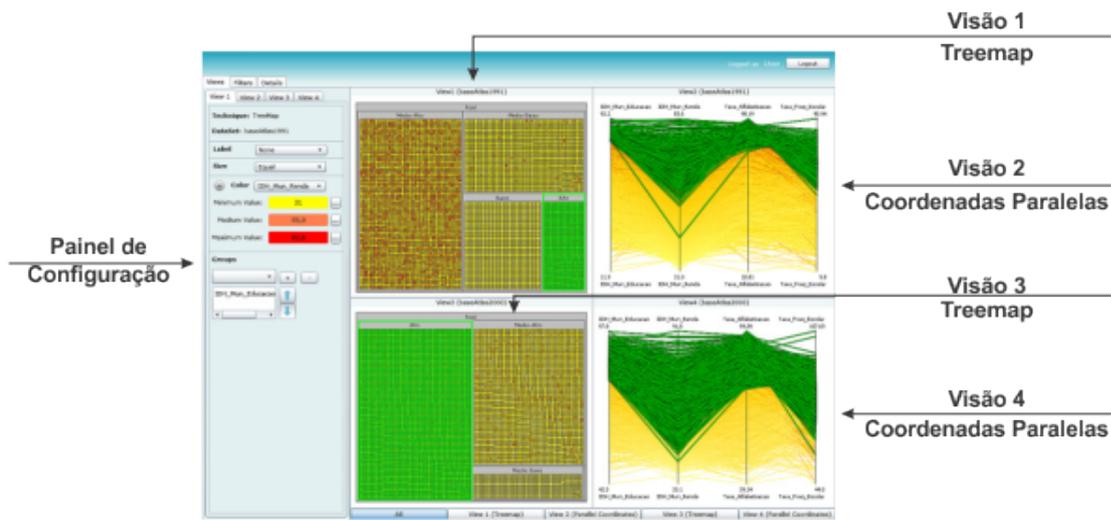
**Figura 4.7** Múltiplas Visões Coordenadas aplicadas a geo-visualização. Adaptado de Jern et al. (2007).

de Estocolmo, mostra emissões médias.

**Dados de indicadores sociais e econômicos** - (GUIMARÃES et al., 2009) utilizou dados referentes a um estudo realizado no Brasil a cada dez anos pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Nos experimentos, duas bases de dados diferentes foram utilizadas, porém com os mesmos atributos: *i*) baseAtlas1991 e *ii*) baseAtlas2000. A partir da exploração dos indicadores sociais e econômicos de 1991 e 2000, foi percebido, por exemplo, que apesar do considerável aumento do Índice de Desenvolvimento Humano

(IDH) educacional brasileiro de 1991 a 2000, o IDH de Renda não aumentou na mesma proporção neste período.

A Figura 4.8 mostra um cenário com Múltiplas Visões Coordenadas criado na aplicação de Guimarães et al. (2009) para explorar mais de um conjunto de dados simultaneamente. A imagem exibe as visões 1 (*Treemap*) e 2 (Coordenadas Paralelas) representando a base de dados baseAtlas1991, e as visões 3 (*Treemap*) e 4 (Coordenadas Paralelas) visualizando o conjunto de dados baseAtlas2000.



**Figura 4.8** Múltiplas Visões Coordenadas representando graficamente duas bases de dados do domínio de indicadores sociais e econômicos. Adaptado de Guimarães et al. (2009).

#### 4.4 DISCUSSÃO

Por meio dos trabalhos da literatura que utilizam MVC pode-se constatar que é comum o uso de diferentes técnicas de visualização para realizar análise exploratória de dados, incluindo as multidimensionais, como as Coordenadas Paralelas, Matriz de Scatter plots. Além disso, são utilizadas as Projeções Multidimensionais, *Treemap* etc. Desse modo, a seguir serão discutidas vantagens e desvantagens das técnicas à luz dos trabalhos analisados.

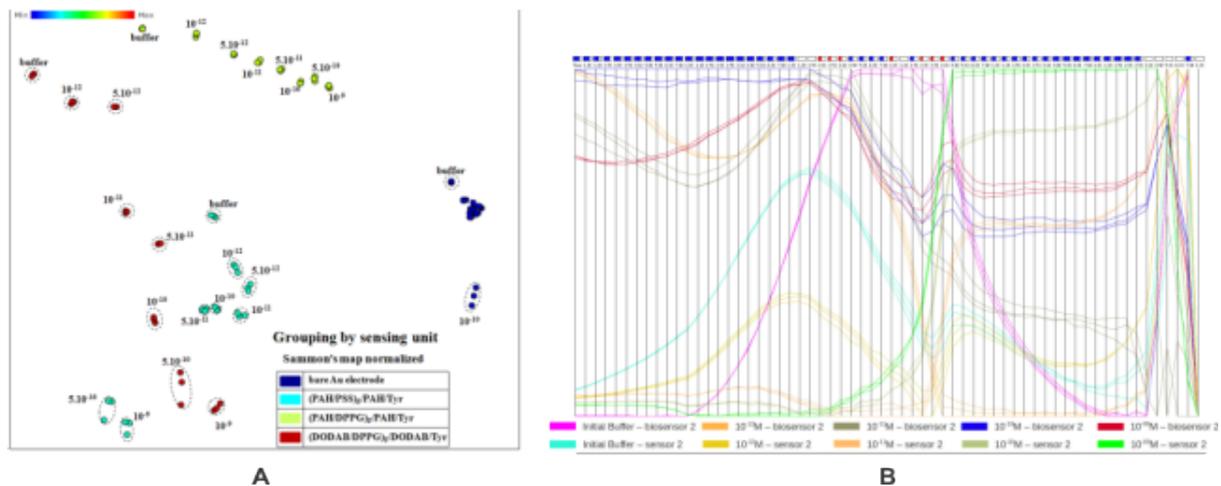
**RadViz:** o uso da técnica RadViz provê algumas vantagens como representar um número maior de dimensões que as visualizações tradicionais e identificar *outliers* e *clusters* de dados (PILLAT, 2006) e (PILLAT; FREITAS, 2006). Apesar das vantagens desta visualização, no estudo de Pillat (2006), esta técnica não permitiu uma boa representação de informações quantitativas e com valores muito semelhantes. Os itens de dados ficaram sobrepostos e seu mapeamento foi considerado complicado e de difícil compreensão pelos especialistas de domínio.

**Matriz de Gráficos de Dispersão:** é muito útil para a percepção de *outliers*, *clusters* e correlações entre dimensões dos dados. Pillat (2006), utilizou a matriz de *scatter plots* para identificar variações nas medidas de crânios em função da localização geográfica.

As interações mais utilizadas na Matriz de Gráficos de Dispersão são: seleção ou *linking & brushing*, filtros, mapeamento de atributo por cor e remoção de atributo das visões. A Matriz de Gráficos de Dispersão pode ser avaliada como uma técnica de fácil aprendizagem, no entanto, ela possui limitações quanto ao número de dimensões e itens de dados que podem ser representados (PILLAT, 2006; PILLAT; FREITAS, 2006).

No estudo de caso realizado por Pillat (2006) com exploração de dados morfométricos, os usuários utilizaram parcialmente a Matriz de Gráficos de dispersão, ou seja, somente em algumas determinadas tarefas. As correlações obtidas por meio da Matriz mostraram claramente que as medidas dos crânios analisados eram alométricas, isto é, quando uma aumenta, as demais também aumentam.

**Projeções Multidimensionais:** A Figura 4.9 mostra uma Projeção Multidimensional e as Coordenadas Paralelas exibindo dados provenientes de Sensores e Biossensores. Na parte A da Figura 4.9 é exibida uma projeção multidimensional baseada na técnica de *Sammon's Mapping*. Os pontos representam os dados de impedância de diferentes biossensores, a cor está mapeando as substâncias em análise, e a proximidade dos pontos indicam concentração dessas substâncias. Ainda na parte B da Figura 4.9, foram utilizadas as Coordenadas Paralelas no intuito de comparar dados de sensores, as colunas (eixos) mapeiam os dados de impedância (frequência) de diferentes biossensores, as linhas estão mapeando as substâncias em análise, permitindo identificar a frequência da concentração das substâncias.

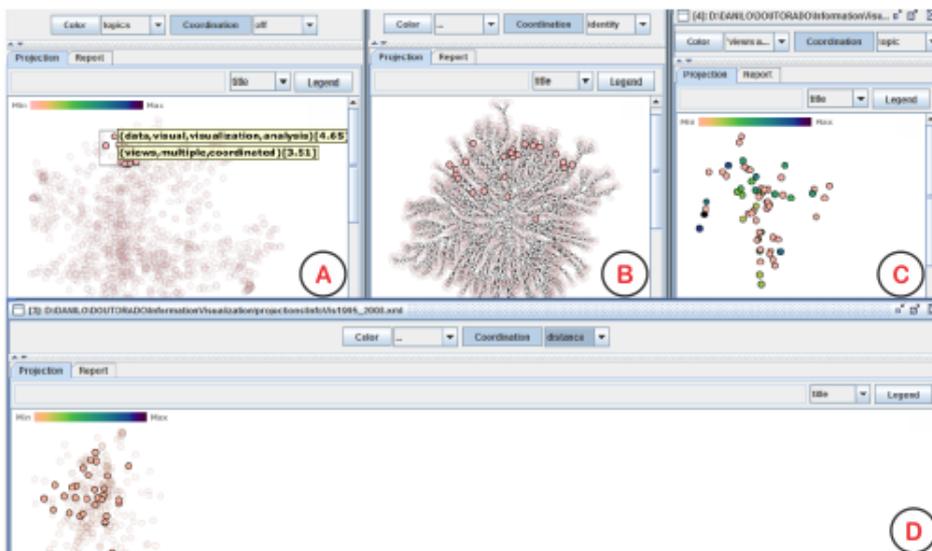


**Figura 4.9** Projeção Multidimensional (lado esquerdo) exibindo dados de sensores e Coordenadas Paralelas (lado direito) exibindo os mesmos dados sem redução de dimensionalidade. Adaptado de Aoki et al. (2013).

Foi utilizada a técnica *Sammon's Mapping* do método *Least Square Scaling*, que dá mais ênfase as pequenas dissimilaridades entre as distâncias, sendo capaz de desdobrar

dados coletados com alta dimensionalidade. A base de dados utilizada foi resultado de um processo de quantização de sinais elétricos de sensores e biossensores que captam elementos sensoriais e geram um subconjunto de dados. A base de dados continha 62 dimensões, classificada em dados de sensor2 e biossensor2.

Eler (2011) desenvolveu um protótipo baseado na ferramenta PEX (*Projection Explorer*)<sup>1</sup> onde o usuário pode criar mapas de similaridade com técnicas de Projeção Multidimensional e explorar os dados com Múltiplas Visões Coordenadas. A Figura 4.10 mostra quatro mapas de similaridade coordenados: o mapa identificado com a letra **A** representa uma coleção de artigos científicos da conferência *Information Visualization Conference - IV* (1997 a 2008). Este foi gerado pela técnica *Least Squares Projection - LSP*; o mapa identificado pela letra **B**, representa os mesmos dados do mapa **A** e foi gerado pela técnica *Neighbour-Joining - NJ*; o mapa identificado pela letra **C** representa uma coleção de artigos da conferência *Coordinated & Multiple Views* (2003 A 2008), e foi gerado pela técnica LSP; e o mapa indentificado com a letra **D**, que representa a coleção de artigos da conferência *IEEE Symposium on Information Visualization (IEEE InfoVis)*, e foi gerado com a técnica LSP.



**Figura 4.10** Mapas de similaridade coordenados usando Projeção Multidimensional. Adaptado de Eler (2011).

No exemplo de Múltiplas Visões Coordenadas de mapas de similaridade mostrado na Figura 4.10 mostra que itens de dados foram selecionados no mapa **A** e através da interação *linking & brushing* os itens de dados correspondentes foram realçados nas demais visões.

**Treemap:** Nos cenários de exploração de dados criados por Almeida et al. (2009), a representação hierárquica da visualização *Treemap* se mostrou muito eficiente. A Figura 4.1 exhibe uma análise do número de votos inválidos por estado nas eleições de 2006 no

<sup>1</sup>Projection Explorer - PEX: <http://vicg.icmc.usp.br/vicg/tool/1/projection-explorer-pex>

Brasil, onde a técnica *Treemap* representa visualmente uma hierarquia de dois níveis: Região e Estado. Na visão do Gráfico de Dispersão, os dados na parte superior foram destacados pelo usuário, representando maior número de votos inválidos. Com o auxílio do *linking & brushing*, os estados correspondentes a estes itens de dados foram realçados e facilmente identificados no *Treemap*. Os estados com maiores porcentagens de votos inválidos foram Bahia, Maranhão e Goiás.

No processo de exploração e extração de conhecimento dos dados com Múltiplas Visões Coordenadas, além das técnicas de visualização, diversas técnicas de interação são aplicadas. Alguns exemplos de mecanismos interativos são *linking & brushing* ou seleção, detalhes sob demanda, zoom, filtros, mapeamentos por cor, por tamanho e por forma. Esses mecanismos interativos tornam o processo de exploração dos dados mais efetivo, pois as manipulações do usuário em uma visão são refletidas simultaneamente em outras visões.

A interação *linking & brushing* é a mais comum em visualizações de MVC (PILLAT; FREITAS, 2006; ALMEIDA et al., 2009; MACIEL et al., 2008; GUIMARÃES et al., 2009; SADANA; STASKO, 2016; GODINHO et al., 2007; KAUER et al., 2008; MATKOVIC et al., 2008; PILLAT, 2006; JERN et al., 2007). A ideia desse artifício interativo é selecionar um ou mais itens de dados em um visualização coordenada, e então os itens de dados correspondentes sejam realçados nas demais visões.

A Figura 4.11 (lado esquerdo) mostra uma tela da aplicação WebPrisma (ALMEIDA et al., 2009), onde é possível perceber a técnica de interação *linking & brushing*. Quando itens selecionados no Gráfico de Dispersão são destacados em um tom de verde perceptível, realçando em verde seus itens correspondentes nas visões *Treemap* e Coordenadas Paralelas. A Figura 4.11 (lado direito), exhibe o resultado do recurso interativo *linking & brushing* na aplicação de Pillat (2006) e Pillat e Freitas (2006). Na interface apresentada um item de dado está selecionado em todas as visualizações. Esse item aparece destacado com uma linha vermelha nas Coordenadas Paralelas (lado direito inferior) e como um círculo vermelho nas técnicas RadViz (centro inferior) e Matriz de Gráficos de Dispersão (lado direito superior).

Outra funcionalidade do protótipo de Pillat e Freitas (2006) é o diagrama de coordenação (lado superior esquerdo). É uma estrutura que permite ao usuário adicionar e remover interações coordenadas entre as visualizações a qualquer momento, podendo sincronizar as ações que são importantes para sua tarefa. Estão disponíveis interações de mapeamento de dados para propriedades visuais como cor, tamanho, transparência, ordenação em profundidade e texto de rótulo; filtragem de dados e dimensões; reordenamento de dimensões; e seleção.

Outra interação presente em MVC é o zoom (ALMEIDA et al., 2009; KAUER et al., 2008). Onde, o usuário pode clicar e/ou arrastar sobre uma área da visualização e um subconjunto de itens serão selecionados com um nível de detalhes ampliado. Este recurso interativo é coordenado, logo, os itens correspondentes nas visões vinculadas serão destacados ou terão o nível de detalhes ampliado. Outro recurso de interação são os filtros, que permitem aos usuários fazerem várias alterações e customizações nas visualizações por meio de painéis de configuração e botões. (PILLAT, 2006; GUIMARÃES et al., 2009;



**Figura 4.11** Exemplo da técnica de interação *linking & brushing* realçando itens de dados nas visões coordenadas: Lado esquerdo, Gráfico de Dispersão, *Treemap* e Coordenadas Paralelas. Adaptado de Almeida et al. (2009); Lado direito, Matriz de Gráficos de Dispersão, RadViz e Coordenadas Paralelas. Adaptado de Pillat e Freitas (2006).

SADANA; STASKO, 2016; GODINHO et al., 2007; ALMEIDA et al., 2009; PILLAT; FREITAS, 2006).

Quando as interações que permitem filtrar dados são combinadas com as visualizações multidimensionais, permitem aos usuários extrair informações significativas. Em um estudo realizado por Pillat (2006), usuários especialistas aplicaram várias combinações de filtros para identificar variações nas medidas de crânios de animais em função da localização geográfica das amostras. Por meio de recursos de filtragem coordenados, conseguiram identificar: *i*) em Pelotas, as amostras apresentaram medidas mais altas; *ii*) na Estação Ecológica de Taim as medidas foram mais baixas; e *iii*) nos demais locais, as amostras apresentam um padrão intermediário entre Taim e Pelotas. Os filtros podem ser aplicados em atributos específicos e estreitar o intervalo de itens de dados exibidos nas representações.

Outros exemplos de recursos interativos em MVC e que contribuem para o analista fazer comparações mais específicas entre atributos de forma mais intuitiva envolvem o uso de atributos visuais, como cor (*Color by*), tamanho (*Size by*) e sobreposição (*Sort by*), para mapear informações nas visões.

A interação *Color by* é o mapeamento por cor para valores das instancias de atributos. Com este recurso o usuário pode escolher uma cor, ou mapa de cores, para representar os itens de dados de um atributo em todas as visões. É possível observar esse mapeamento nos trabalhos de: (GODINHO et al., 2007; ALMEIDA et al., 2009; MACIEL et al., 2008; MATKOVIC et al., 2008; JERN et al., 2007; PILLAT, 2006; PILLAT; FREITAS, 2006). No trabalho de Pillat e Freitas (2006) (Figura 4.11 lado direito), o recurso de interação de mapeamento por cor (*Color by*) foi usado no intuito de representar os valores do

atributo Cilindros (*Cylinders*) para níveis de intensidades da cor verde. As instancias desse atributo foram mapeadas em todas das visões utilizando tons de verde.

Outro mecanismo de interação utilizado em MVC é o *Size by* ou mapeamento por tamanho dos valores dos itens de dados de um atributo (PILLAT, 2006) e (PILLAT; FREITAS, 2006). Na Figura 4.11 (lado direito), as formas coloridas de verde escuro e com tamanho grande têm os valores mais altos para os atributos Cilindros (*Cylinders*) e Ano (*Year*), respectivamente.

*Sort by* ou *Sorting* é utilizado para mudar a ordem de sobreposição de pontos e linhas, por exemplo. Nas Coordenadas Paralelas ou Matriz de Gráficos de Dispersão, o *Sort by* faz com que as polilinhas ou os pontos que representam as instâncias de um atributo se sobreponham (“trazer para frente”) aos demais. É possível ver exemplos dessa interação nos seguintes trabalhos: (PILLAT, 2006; PILLAT; FREITAS, 2006; GODINHO et al., 2007; ALMEIDA et al., 2009; MATKOVIC et al., 2008; JERN et al., 2007).

Outra interação muito utilizada em MVC no processo de exploração de dados é o recurso detalhes sob demanda. Geralmente, os detalhes do item de dado selecionado aparecem como textos flutuantes ou em uma visão tabular dos dados. O objetivo deste mecanismo interativo é exibir mais informações sobre itens de dados. Em experimentos com usuários, Maciel et al. (2008) constatou que a técnica de interação detalhes sob demanda foi usada por 62% dos usuários participantes de seu estudo com Múltiplas Visões Coordenadas.

**MVC - Projeção Multidimensional - Aprendizado de Máquina:** Chegini et al. (2019) apresenta uma abordagem de análise visual combinando recursos de Aprendizado de Máquina com Múltiplas Visões Coordenadas, por meio da aplicação mVis, um visualizador de dados multidimensionais. O conjunto de técnicas disponíveis permite explorar os dados, dividi-los em partições rotuladas e construir um classificador. A Figura 4.12 exibe a tela do mVis representando um conjunto de dados de futebol que contém 42 atributos e 318 instâncias. Foram identificadas as partes principais da interface: *i*) quatro visualizações coordenadas: Matriz de Gráficos de Dispersão (A), Gráfico de Dispersão (B), Mapa de Similaridade ou Projeção Multidimensional (C) e Coordenadas Paralelas (E); *ii*) e um painel de configuração das partições criadas.

Através desta ferramenta, o usuário pode executar um fluxo de trabalho interativo que envolve a análise exploratória dos dados, rotulação de padrões interessantes ou *outliers* por um processo semi-supervisionado, aprendizado de máquina supervisionado para avaliar o grau de aprendizagem do classificador. As interações do usuário são usadas para orientar os algoritmos de aprendizado de máquina. No mVis, a visualização Coordenadas Paralelas suporta várias interações, incluindo *linking & brushing*, filtro, e reordenamento de eixos. Outras interações foram incluídas como os *widgets* e parâmetros mínimos para os algoritmos de Aprendizado de Máquina, no intuito de manter as interações intuitivas.

No cenário da Figura 4.12, a visão A mostra a Matriz de *Scatter Plots* representando as relações par a par entre as dimensões. Foi selecionado um Gráfico de Dispersão da matriz e mostrado em detalhes na visão B. A visão C é um mapa de similaridade com as partições agrupadas e codificadas por cores, conforme indicado no painel de partições na visão D. Se duas partições tiverem dimensões associadas (pela interação do usuário),

elas serão conectadas por uma linha. A visão E mostra as Coordenadas Paralelas que representam todas as dimensões do conjunto de dados. As dimensões que participam dos algoritmos de aprendizado de máquina são indicadas por uma faixa azul.



**Figura 4.12** Interface da aplicação mVis, um visualizador de dados multidimensionais, que combina Múltiplas Visões Coordenadas com recursos de Aprendizado de Máquina. Adaptado de Chegini et al. (2019).

**Múltiplas Visões Coordenadas distribuídas em dispositivos móveis:** Langner, Horak e Dachsel (2018) apresentam o *framework* conceitual VisTiles<sup>2</sup>, uma aplicação para distribuir e coordenar visualizações em um conjunto de dispositivos móveis para a exploração de dados multidimensionais. Diferente das interfaces de visualização baseadas em *desktop*, os dispositivos móveis oferecem uma interface dinâmica e definida pelo usuário, com suporte à exploração colaborativa de dados co-localizados com diferentes fluxos de trabalho individuais.

O VisTiles permite que os usuários interajam com Múltiplas Visualizações Coordenadas distribuídas em vários dispositivos móveis. Os principais componentes do *framework* são: *i)* *layouts* dinâmicos e flexíveis para MVC, com foco na distribuição de visões entre aparelhos diferentes e *ii)* um conceito de interação para adaptações inteligentes e combinações de visualizações utilizando arranjos lado a lado dos dispositivos.

A Figura 4.13 exhibe a aplicação em execução em diferentes cenários e dispositivos móveis, incluindo: (A) alinhamento de visualizações; (B) compartilhando a interface do usuário (painel de configuração) com outro dispositivo para redução da interferência visual; (C) extensão do espaço de exibição das visualizações; ou (D) extensão do espaço de exibição dos dados computados; combinações de dispositivos também podem ser usadas para (E) filtrar itens de dados; ou (F) manipular vistas de visão geral e de detalhes por movimentos espaciais.

<sup>2</sup>Protótipo VisTiles: <https://github.com/imldresden/vistiles>



**Figura 4.13** Exemplo de execução do Vistiles em diferentes dispositivos móveis. Adaptado de Langner, Horak e Dachsel (2018).

## 4.5 CONSIDERAÇÕES FINAIS

Uma alternativa relevante para explorar e analisar dados multidimensionais é o conceito de Múltiplas Visões Coordenadas. Neste Capítulo foram apresentados diversos trabalhos que aplicaram esta abordagem para ampliar a extração de conhecimento em vários domínios de dados, dentre os quais destacam-se: política, futebol, vendas online, indicadores sociais e econômicos, morfometria craniana, fornecimento e uso de energia, veículos, tráfego de rede e dados meteorológicos. Vimos que técnicas de visualização tradicionais e multidimensionais, compartilham recursos interativos, para tornar mais intuitiva a ação do usuário especialista ao manipular as representações gráficas em busca de informação. Ficou evidente que a técnica de Coordenadas Paralelas é comumente utilizada em aplicações que suportam MVC. Nesse sentido, as técnicas de interação *linking* & *brushing* e detalhes sob demanda, entre outras, são muito utilizadas no processo de exploração de dados.



## MULTIVISD3: *PROJECTION ERROR ANALIZER*

### 5.1 CONSIDERAÇÕES INICIAIS

Este capítulo descreve o processo de desenvolvimento do MultiVisD3: *Projection Error Analyzer*, uma aplicação Web para colaborar com a interpretação de erros em Projeções Multidimensionais. As seções estão organizadas da seguinte maneira: na Seção 5.2 é apresentada e discutida detalhadamente a metodologia que foi adotada para o desenvolvimento da abordagem de *Visual Analytics* proposta. Na Seção 5.4 são descritas as funcionalidades e os recursos interativos associados. Na Seção 5.5 são relatadas as considerações finais sobre este trabalho.

### 5.2 METODOLOGIA

A metodologia empregada no desenvolvimento do MultiVisD3 segue as etapas do *pipeline* de visualização ilustrado na Figura 5.1. Nesse fluxo o usuário é um elemento essencial, pois ele realiza ações de interação em vários estágios. O processo inicia com a seleção de uma base de dados na aplicação, em seguida, são aplicadas rotinas de pré-processamento, estruturação e conversão dos dados em elementos visuais, por fim, as visualizações são carregadas na tela e disponíveis para o usuário realizar a exploração visual. A seguir, serão detalhadas cada uma das etapas que compõem esse *pipeline*:

#### Dados

No estado atual do MultiVisD3, não permitir ao usuário inserir uma base de dados de sua escolha, pois a aplicação possui coleções de dados pré-carregados. Os conjuntos de dados utilizados foram extraídos do Repositório *UCI Machine Learning*<sup>1</sup>, a saber, *Image Segmentation*, *Wine*, *Íris* e *Wireless Indoor Localization*, os quais são comumente empregados em pesquisas relacionadas a projeções multidimensionais (JOIA et al., 2011; VENTOCILLA; RIVEIRO, 2020; XIA et al., 2022). As bases contêm arquivos de metadados em separado com os nomes das colunas e informações sobre a variável categórica

<sup>1</sup>Disponível em: <https://archive.ics.uci.edu>. Acessado em 07/07/2023.

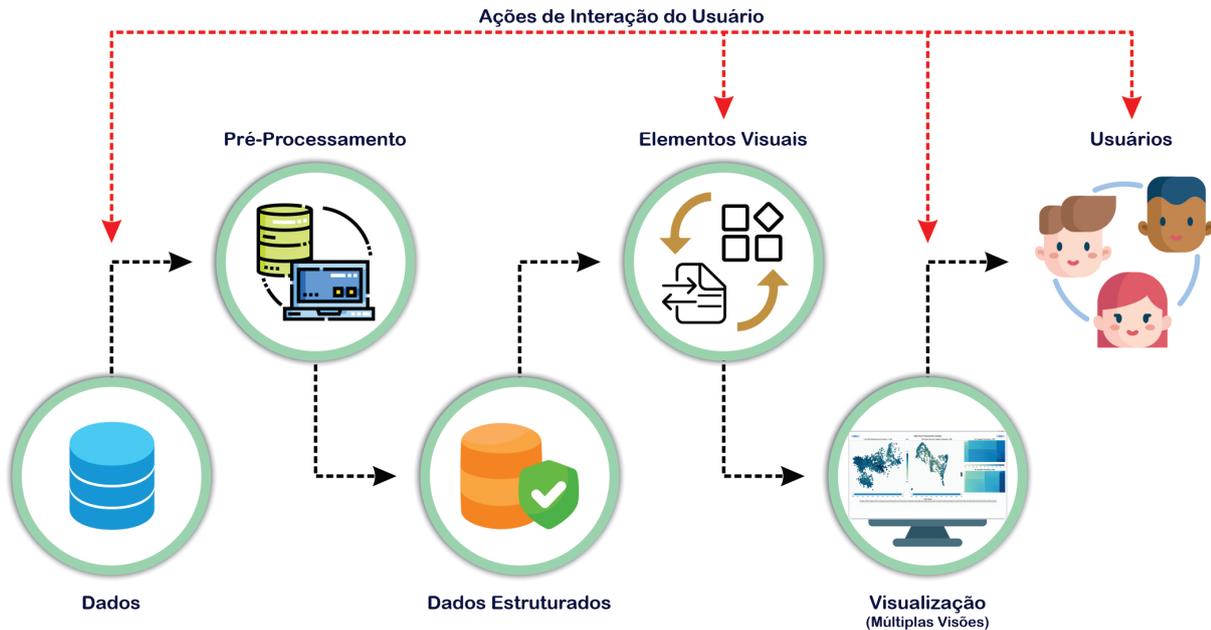


Figura 5.1 Pipeline da MultiVisD3.

que contém o rótulo da base. O formato dos arquivos é ‘CSV’ (*Comma Separated Values*) e serão descritos a seguir:

- *Wine*<sup>2</sup> - Essa base de dados é constituída pelo resultado de uma análise química de vinhos cultivados na mesma região da Itália, mas derivados de três cultivares diferentes. A análise determinou as quantidades de constituintes encontradas em cada um dos três tipos de vinhos. Esse conjunto de dados possui 14 dimensões e 178 instâncias.
- *Image Segmentation*<sup>3</sup> - Esse *dataset* é composto por dados de 7 imagens externas que foram segmentadas manualmente para criar uma classificação para cada pixel. Cada instância é uma região 3x3. Essa base de dados tem 21 dimensões e 2100 instâncias.
- *Íris*<sup>4</sup> - É uma base de dados conhecida na literatura de reconhecimento de padrões. Ela contém 3 classes, onde cada classe se refere a um tipo de planta íris. Esse *dataset* é composto por 6 dimensões e 147 instâncias.
- *Wireless Indoor Localization*<sup>5</sup> - Essa base de dados foi coletada em quatro salas diferentes observando as intensidades de sinal de sete redes WiFi visíveis em um *smartphone*. Esse conjunto de dados tem 9 dimensões e 2000 instâncias.

<sup>2</sup>Disponível em: <https://archive.ics.uci.edu/dataset/109/wine> Acessado em 07/07/2023.

<sup>3</sup>Disponível em: <https://archive.ics.uci.edu/dataset/50/image+segmentation>. Acessado em 07/07/2023.

<sup>4</sup>Disponível em: <https://archive.ics.uci.edu/dataset/53/iris>. Acessado em 07/07/2023.

<sup>5</sup>Disponível em: <https://archive.ics.uci.edu/dataset/422/wireless+indoor+localization>. Acessado em 07/07/2023.

## Pré-Processamento

Para que um conjunto de dados seja mapeado corretamente pelas técnicas de visualização é necessário que esteja em um formato e uma estrutura apropriados para esse propósito. Por isso, é fundamental realizar uma etapa de preparação desses dados para aprimorar sua qualidade e permitir que sejam convertidos corretamente em atributos visuais. Diante disso, os dados projetados no MultiVisD3 foram submetidos a duas etapas de pré-processamento que realizaram as seguintes rotinas:

### Padronização

- Verificar se os dados possuem valores nulos e não numéricos como caracteres especiais ou *strings*. Caso haja, os valores ausentes são substituídos pelo valor médio da coluna.
- Adicionar cabeçalho ao *dataset* se os nomes das colunas vierem em arquivos de metadados separados.
- Remover caracteres especiais dos nomes das colunas, pois esses podem dificultar ou até mesmo impedir a manipulação dos dados.
- Inserir uma coluna ‘id’ no conjunto de dados, se não houver.
- Padronizar o valor inicial da coluna ‘id’ em todas as bases para começar como o mesmo valor (um, por exemplo). Além disso, garantir que os valores sejam sequenciais e não negativos.
- Transformar os valores do tipo *string* da variável categórica que representa o rótulo da base em valores numéricos e renomear a coluna para ‘class’.
- Reordenar as colunas da base para que ‘id’ fique sempre no início e ‘class’ sempre no final.

### Normalização e Conversão

- Normalização Min-max - que realiza uma transformação linear nos dados preservando as relações entre os valores originais (HAN; KAMBER; PEI, 2012). Suponha que  $min_A$  e  $max_A$  sejam os valores mínimo e máximo de um atributo  $A$ . A Normalização Min-max mapeia um valor  $v_i$  de  $A$  para  $v'_i$  no intervalo  $[novo\_min_A, novo\_max_A]$  calculado por

$$v'_i = \frac{v_i - min_A}{max_A - min_A} (novo\_max_A - novo\_min_A) + novo\_min_A \quad . \quad (5.1)$$

- Conversão dos dados para o formato JSON<sup>6</sup> (*JavaScript Object Notation*).

---

<sup>6</sup>Disponível em: <https://www.json.org/json-en.html>. Acessado em 07/07/2023.

## Dados Estruturados

Nessa etapa os dados são estruturados para atender os requisitos de cada uma das técnicas de visualização. Sendo assim, são realizados diferentes procedimentos como: mapeamento de valores, a exemplo de intervalos, mínimos, máximos etc., para prover parâmetros para as funções de conversão; criação de escalas para dar suporte ao mapeamento de dados qualitativos (cores, por exemplo) e quantitativos (distâncias, por exemplo); criação de estruturas que representam algum tipo de hierarquia de dados; criação de objetos ou estruturas abstratas para apoiar a conversão dos dados em metáforas visuais.

## Elementos Visuais

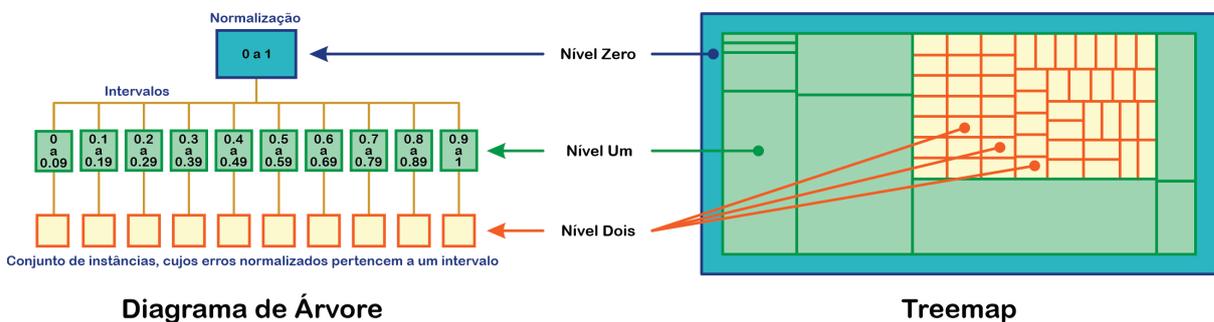
Este estágio do pipeline é realizado automaticamente e visa criar uma abstração visual dos dados, ou seja, representações baseadas nos atributos visuais (cor, tamanho, forma, entre outros), apresentados na Seção 2.2. Os valores são convertidos em entidades gráficas, e podem ser mapeados para o tamanho de um componente visual, enquanto outros podem representar a posição ou a cor do objeto.

Em relação a organização das visões e a divisão do espaço de tela, foi adotada a justaposição das visualizações, conforme é comumente utilizado nos trabalhos e ferramentas de Múltiplas Visões Coordenadas (Seção 3.2) para evitar perda de informações com a oclusão e confusão visual nas representações gráficas.

## Visualização

Esta etapa executa a renderização dos elementos visuais e gera o *front-end* do MultiVisD3 com as múltiplas visualizações, das quais as principais técnicas implementadas foram: as técnicas de Projeção Multidimensional LAMP e t-SNE (Seção 2.2.1.1), que são escaláveis e suportam alta dimensionalidade; o *Treemap* para colaborar com a tarefa de identificar quais instâncias foram projetadas com maior e menor valor de erro, segundo a métrica utilizada; e a Visão Tabular para apoiar com informações adicionais dos dados analisados. Além disso, mecanismos de configuração, como visão dos painéis, seleção de base de dados e customização dos elementos visuais foram também adicionados à tela da aplicação.

A visualização *Treemap* foi escolhida com base nos estudos discutidos na Seção 4 para auxiliar na comparação da qualidade das PMs. Criou-se uma estrutura hierárquica de dados da seguinte forma: a métrica de Erro Agregado calcula o valor de erro para cada instância ou ponto projetado; os valores de erro são normalizados entre zero e um; em seguida, são criados dez intervalos (0 - 0.09, 0.1 - 0.19, 0.2 - 0.29, 0.3 - 0.39, 0.4 - 0.49, 0.5 - 0.59, 0.6 - 0.69, 0.7 - 0.79, 0.8 - 0.89, 0.9 - 1); por fim, os valores de erro normalizados de cada instância são agrupados nos seus respectivos intervalos. A Figura 5.2 ilustra através de um diagrama de árvore e um *Treemap* como os dados foram estruturados para representar de forma hierárquica os valores dos erros normalizados. Em ambas as visualizações os dados estão divididos em três níveis: zero, um e dois.



**Figura 5.2** Representação dos valores de erro normalizados por meio de um diagrama de árvore e um *Treemap*.

No nível zero encontra-se o nó raiz, que no diagrama de árvore é representado por um nó sozinho no topo, e no *Treemap* pelo retângulo externo. Esse nível corresponde a normalização dos valores de erro no intervalo entre zero e um. Em seguida, no nível um o nó raiz é subdividido em dez ramos com novos intervalos, exibidos no diagrama de árvore, enquanto que no *Treemap* a área do retângulo externo é decomposta em dez subáreas proporcionais a quantidade de instâncias associadas a cada novo intervalo. O nível dois são as instâncias e como pode ser observado, não há sobreposição das instâncias. Portanto, é possível identificar mais facilmente quais intervalos possuem maior ou menor quantidade de erros agregados para ambas técnicas de projeção.

### Ações de Interação do Usuário

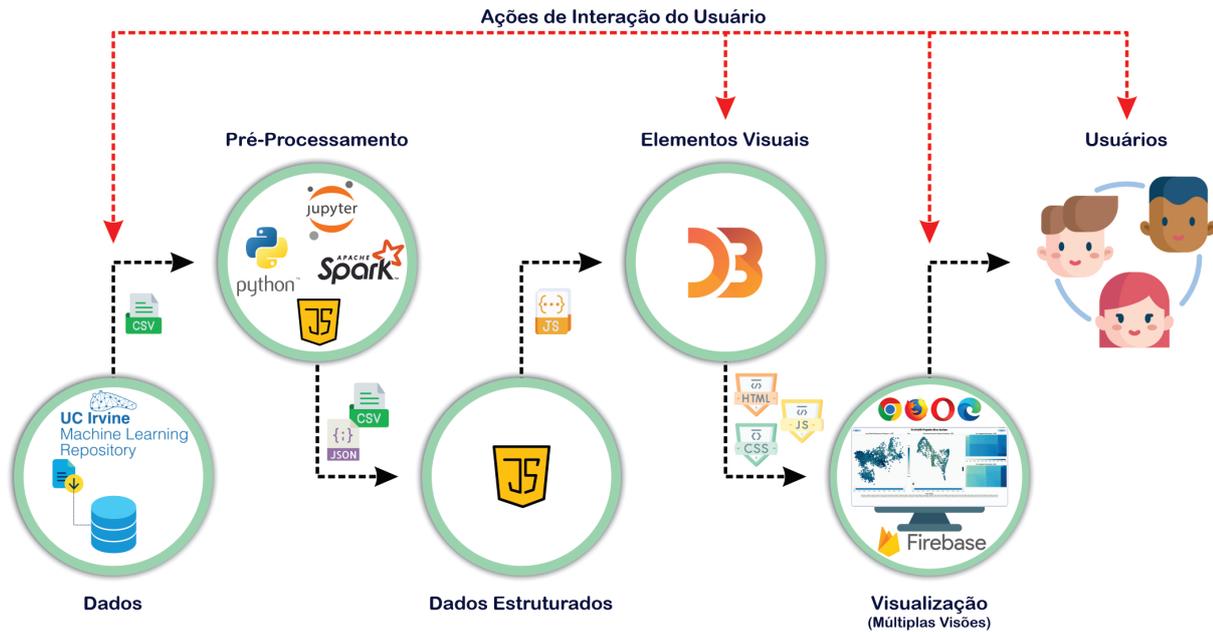
O usuário é um elemento central nesse fluxo (Figura 5.1), pois ao manusear os recursos interativos, as alterações nas visualizações vão ocorrendo e sendo propagadas para as outras visões vinculadas e os cenários de exploração dos dados vão sendo construídos. Para apoiar esse processo de extração visual de informação, foram implementados diferentes mecanismos, estratégias e abordagens para favorecer uma experiência mais interativa. Esses recursos são apresentados na Seção 5.4.2.

## 5.3 TECNOLOGIAS UTILIZADAS

As tecnologias utilizadas foram selecionadas com a premissa de que permitissem o desenvolvimento de uma abordagem de *Visual Analytics* que exigisse o mínimo ou nenhum esforço do usuário em termos de instalação e configuração. A Figura 5.3 destaca os recursos tecnológicos utilizados em cada etapa, bem como os tipos de dados gerados e consumidos pelo MultiVisD3 no fluxo de execução. Esses mecanismos serão descritas a seguir:

- Pré-Processamento

O pré-processamento, conforme detalhado na Seção 5.2, foi realizado em duas etapas: 1) Padronização; e 2) Normalização e Conversão. A seguir serão apresentadas



**Figura 5.3** Tecnologias Utilizadas no MultiVisD3: *Projection Error Analyzer*.

as tecnologias que deram suporte as rotinas de tratamento dos dados nessas etapa:

- Padronização - Essa etapa tem como entrada e saída arquivos no formato ‘CSV’. As tecnologias que possibilitaram as transformações nos dados foram: *Jupyter Notebook*<sup>7</sup>, utilizado para configurar o ambiente de criação e edição de códigos através de uma interface Web interativa e a linguagem de programação. *Python*<sup>8</sup> integrada com o mecanismo multilinguagem para executar a engenharia dos dados *Apache Spark*<sup>9</sup>.
- Normalização e Conversão - onde o *input* é um arquivo ‘CSV’ e seu *output* gera arquivos do tipo ‘JSON’. Para esses procedimentos foi utilizado JavaScript<sup>10</sup> (JS), uma linguagem de programação interpretada (ou compilada *just-in-time*) que implementa diversas funcionalidades.

- Dados Estruturados

Para realizar a estruturação dos dados (Seção 5.2) foi utilizada a linguagem JavaScript.

- Elementos Visuais

<sup>7</sup>Disponível em: <https://jupyter.org/> Acessado em: 06/08/2023

<sup>8</sup>Disponível em: <https://www.python.org/> Acessado em: 06/08/2023

<sup>9</sup>Disponível em: <https://spark.apache.org/> Acessado em: 06/08/2023

<sup>10</sup>Disponível em: <https://developer.mozilla.org/en-US/docs/Web/JavaScript> Acessado em: 06/08/2023

A abstração dos dados e conversão em atributos visuais foi realizada pelo D3.js<sup>11</sup>, uma biblioteca JavaScript gratuita e de código aberto para visualização de dados baseada em padrões da Web que oferece flexibilidade na criação de gráficos interativos, dinâmicos e orientados por dados.

- Visualização (Múltiplas Visões)

O MultiVisD3: *Projection Error Analyzer*<sup>12</sup> segue os padrões Web e utiliza HTML<sup>13</sup> (*HyperText Markup Language*) para construção do website e CSS<sup>14</sup> (*Cascading Style Sheets*) que descrevem a aparência ou apresentação para os navegadores. Para uma melhor experiência, recomenda-se o navegador Chrome<sup>15</sup>, no entanto, a aplicação também foi testada no Firefox<sup>16</sup>, Opera<sup>17</sup> e Edge<sup>18</sup>.

A abordagem de *Visual Analytics* desenvolvida faz uso do *Color Brewer*<sup>19</sup>, uma ferramenta para orientação na escolha de esquemas de cores para uso na web e encontra-se hospedada no *Firebase*<sup>20</sup>, uma plataforma que fornece um conjunto de ferramentas para desenvolvedores, dentre elas, o serviço *Hosting* que é um recurso de hospedagem de conteúdo da Web em nível de produção. Além disso, a aplicação encontra-se disponível no repositório ‘MultiVisD3’ do GitHub<sup>21</sup>.

## 5.4 DESCRIÇÃO DO MULTIVISD3: *PROJECTION ERROR ANALIZER*

Durante todo o processo de implementação dos *layouts* e seus recursos, a experiência do usuário foi considerada um elemento primordial e contribuiu nas decisões de *design* e organização dos itens visuais, a otimização do espaço em tela, a customização dos diferentes mecanismos interativos, entre outros.

A fim de facilitar a interpretação de métricas de qualidade em projeções multidimensionais, foi utilizado o conceito de múltiplas visões coordenadas, considerando seus diversos benefícios, apresentados na Seção 3.2. Em relação ao divisão do espaço em tela, as visualizações foram posicionadas de maneira justaposta (lado a lado), para auxiliar nas tarefas de análise e comparação dos dados. Essa abordagem minimizou a complexidade visual e favoreceu a criação de cenários de exploração mais simplificados que cooperam para a percepção das diferenças entre os dados observados, conforme discutido na Seção 3.3.

---

<sup>11</sup>Disponível em: <https://d3js.org/> Acessado em: 07/07/2023

<sup>12</sup>Disponível em: <https://multivisd3.web.app/> Acessado em: 06/09/2023

<sup>13</sup>Disponível em: <https://developer.mozilla.org/pt-BR/docs/Web/HTML> Acessado em: 06/08/2023

<sup>14</sup>Disponível em: <https://developer.mozilla.org/pt-BR/docs/Web/CSS> Acessado em: 06/08/2023

<sup>15</sup>Disponível em: <https://www.google.pt/intl/pt-PT/chrome/> Acessado em: 06/08/2023

<sup>16</sup>Disponível em: <https://www.opera.com/pt-br/download> Acessado em: 06/08/2023

<sup>17</sup>Disponível em: <https://www.mozilla.org/en-US/firefox/new/> Acessado em: 06/08/2023

<sup>18</sup>Disponível em: <https://www.microsoft.com/pt-br/edge/> Acessado em: 06/08/2023

<sup>19</sup>Disponível em: <https://colorbrewer2.org/> Acessado em: 06/08/2023

<sup>20</sup>Disponível em: <https://firebase.google.com> Acessado em: 06/08/2023

<sup>21</sup>Disponível em: <https://github.com/joseaugustoduarte/MultiVisD3/> Acessado em: 05/09/2023

### 5.4.1 Layout

O MultiVisD3: *Projection Error Analyzer* permite que o usuário explore o mesmo conjunto de dados a partir de três perspectivas diferentes, sendo elas: Erro Agregado, Classe e Preservação de Distância. Cada uma dessas perspectivas possibilitam uma visão dos dados e correspondem a um painel visual que utilizam mais de uma visualização do mesmo tipo para apoiar o processo de comparação dos dados entre as diferentes representações. O *layout* é flexível e se ajusta de acordo a perspectiva selecionada pelo usuário. Por padrão, o espaço em tela é preenchido, predominantemente, pelas Projeções Multidimensionais. Além disso, a depender da visão escolhida, outros componentes podem ser adicionados ou removidos, dentre eles estão os *Treemaps*, a visão tabular e os mecanismos de interação indireta como *sliders* e botões. Os três painéis visuais são apresentados a seguir:

#### Erro Agregado

Esse painel visual utiliza o atributo cor para mapear os valores da métrica de erro agregado de projeção nas visualizações. Sendo assim, as Projeções Multidimensionais e os *Treemaps* compartilham a mesma escala de cor sequencial para apoiar a identificação de instâncias com valores de erro agregado mais altos e mais baixos. A Figura 5.4 mostra essa perspectiva representando o conjunto de dados *Segmentation*, e através das legendas é possível perceber que as cores mais claras mapeiam valores de erros mais baixos e as cores mais escuras correspondem a valores de erros mais altos. Essa mesma imagem também destaca os principais componentes dessa visão que serão descritos a seguir, conforme os rótulos da Figura 5.4:

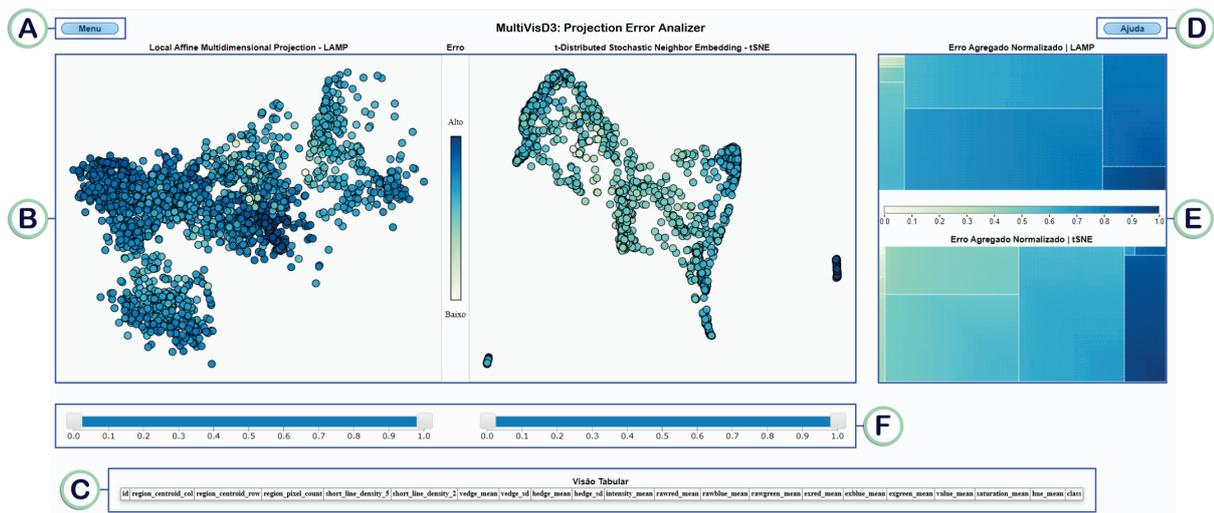


Figura 5.4 Layout da Visão ‘Erro Agregado’.

A) O botão ‘Menu’ fica no canto superior esquerdo e permite ao usuário acessar um conjunto de configurações para personalizar sua experiência de exploração. Essa

funcionalidade possibilita mais flexibilidade para definir como os atributos visuais das instâncias serão representados. Os itens do ‘Menu’, são exibidos na Figura 5.5, e suas funcionalidades serão explicadas a seguir:

- Idioma - Alterar o idioma da aplicação e escolher entre Inglês e Português Brasil.
- Base de Dados - Selecionar uma das bases de dados pré-carregadas e conferir alguns metadados como nome da base, número de registros e atributos.
- Visão - Escolher uma das três perspectivas para interpretar a qualidade das PMs.
- Tamanho do Círculo - Modificar o tamanho dos círculos das PMs.
- Cor de Fundo - Mudar o nível de luminosidade (brilho) da cor de fundo das PMs.
- Transparência do Círculo - Aumentar ou diminuir o nível de transparência dos círculos das PMs. Esse recurso é importante quando se tem oclusão visual causada pela sobreposição dos dados.
- Linha de Borda do Mapa de Árvore - Adicionar ou remover a linha de borda dos retângulos dos *Treemaps*.
- Rótulo ao Passar o *Mouse* - Habilitar ou desabilitar a exibição de textos informativos quando o usuário passar o ponteiro do *mouse* sobre alguma instância tanto nas PMs quanto nos *Treemaps*.
- Cor do Realce ao Passar o *Mouse* - Trocar a cor do realce que ocorre ao passar o ponteiro do *mouse* sobre algum elemento das visualizações.
- Paleta de Cores - Alterar a paleta de cores das PMs e dos *Treemaps*.

**B)** As duas técnicas de Projeção Multidimensional representam todo o conjunto de dados, sendo a LAMP e a t-SNE (discutidas na Seção 2.2.1.1). Entre elas tem a legenda que indica as cores que mapeiam os valores da métrica de erro agregado.

**C)** A ‘Visão Tabular’ tem a finalidade de mostrar as informações contidas na base de dados sobre uma ou mais instâncias. Ocorre que quando o usuário passa o ponteiro do *mouse* sobre os itens de dados nas visualizações (PM ou *Treemaps*), ou então, realiza a seleção dos mesmos, a visão tabular mostra o registro da base de dados, correspondente.

**D)** O botão ‘Ajuda’ apresenta explicações sucintas sobre como utilizar os principais recursos interativos.

**E)** Os dois *Treemaps* representam os valores da métrica de erro agregado normalizados para uma mesma escala (entre zero e um) e divididos em dez intervalos. Desta forma, foi possível agrupar as instâncias pertencentes a cada intervalo e ordená-las, a fim de possibilitar uma rápida identificação das instâncias com valores de erro mais altos e

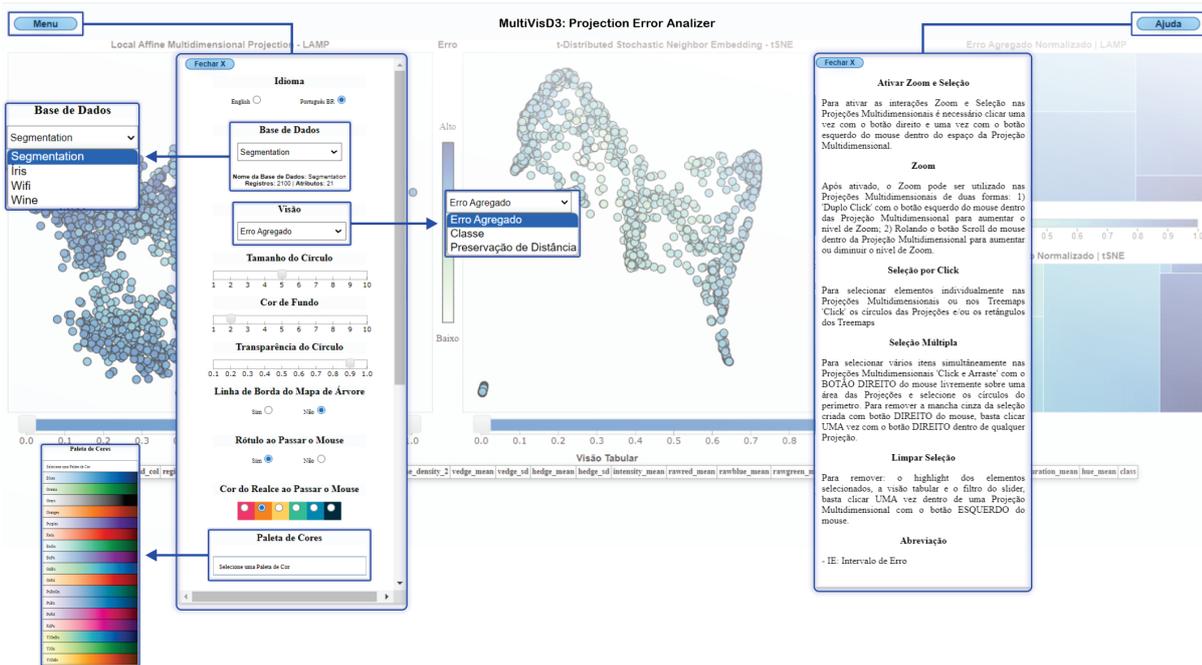


Figura 5.5 Botão ‘Menu’ e suas Opções; Botão ‘Ajuda’.

mais baixos, por exemplo. A legenda dos *Treemaps* utiliza a mesma paleta de cores das Projeções Multidimensionais.

F) Os *sliders* são mecanismos interativos que funcionam como filtros e realçam somente as instâncias do intervalo de erro selecionado pelo usuário.

## Classe

Nesta visão, o desenho do *layout* é bem semelhante ao anterior, incluindo as opções do ‘Menu’, pois preserva a maioria dos componentes, conforme apresentados nas Figuras 5.5 e 5.6. No entanto, elementos foram adaptados para mapear os dados categóricos do atributo ‘class’ que é o rótulo da base de dados. Dentre as modificações estão: remoção da legenda localizada entre os *Treemaps*; mudança do esquema de cores das visualizações para qualitativo.

Outra alteração foi a adaptação da legenda de cores (Figura 5.6 item G) que possui um conjunto de quadrados representando as classes do *dataset*. Além disso, possui um filtro interativo associado que possibilita investigar o erro agregado das instâncias de uma ou mais classes isoladamente. As categorias mapeadas na legenda funcionam como botões que permitem habilitar ou desabilitar a transparência nos itens correspondentes e deixar realçados somente os dados referentes as classes que se deseja analisar.

A Figura 5.6 mostra a base de dados *Segmentation* sendo representada visualmente. É possível ver que as cores estão mapeando as categorias do atributo ‘class’ nas visualizações. Na projeção LAMP, por exemplo, as instâncias das classes zero (vermelha) e

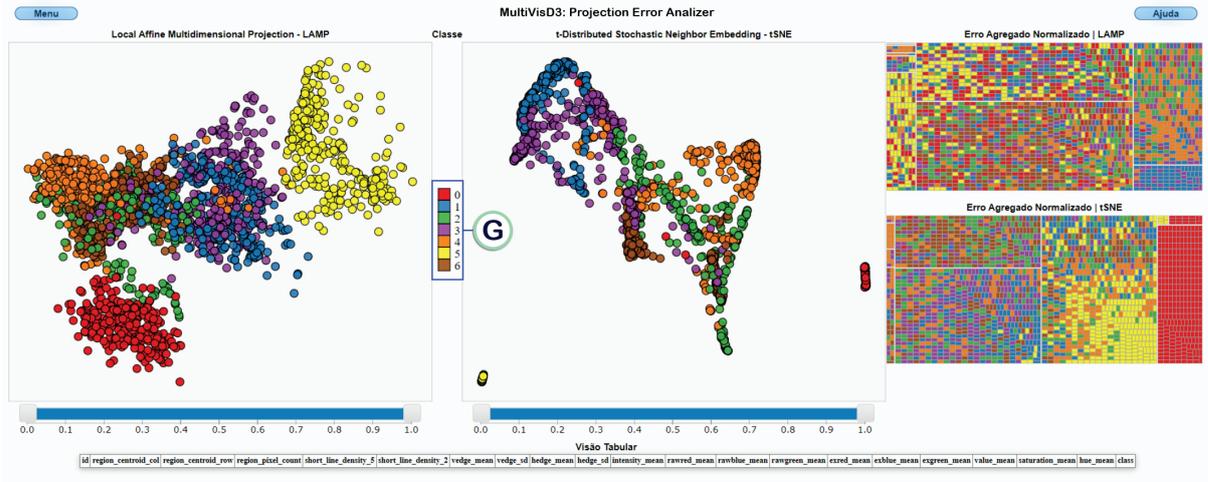


Figura 5.6 *Layout* da Visão ‘Classe’.

cinco (amarela) estão mais dispersas, enquanto que na técnica t-SNE os pontos desses grupos apresentam mais separação e sobreposição. Considerando essas classes, os *Tree-maps* mostram que os erros das classes zero e cinco estão mais distribuídos na LAMP, ao passo que na t-SNE as instâncias apresentam mais erros altos de projeção.

### Preservação de Distância

Essa visão tem como propósito auxiliar o usuário a identificar o nível de preservação de vizinhança das instâncias. O *layout* possui semelhanças com os anteriores, no entanto, foram removidos alguns itens como os *Treemaps* e os *Sliders*, conforme mostra a Figura 5.7. As instâncias mapeadas nas PMs dessa visão utilizam a cor branca e, caso o usuário passe o ponteiro do *mouse* ou selecione os pontos, as cores vão alterando, de acordo a legenda, para indicar os vizinhos ausentes e falsos. A codificação visual do atributo cor também foi alterada para um esquema divergente.

No *layout* de ‘Preservação de Distância’ o ‘Menu’ foi alterado e não constam os itens ‘Linha de Borda do Mapa de Árvore’ e a ‘Paleta de Cores’. No entanto, foi adicionada a opção ‘Vizinhos’, onde o usuário pode selecionar três alternativas: Falsos, Verdadeiros e Ausentes (Figura 5.8, item A). Nessa mesma Figura, os itens B, C e D mostram o resultado da seleção da mesma instância, habilitando apenas uma opção por vez.

Se o usuário selecionar apenas a opção ‘Falso’, quando ocorrer o *mouseover* ou a seleção das instâncias, somente esses vizinhos serão realçados (Figura 5.7, item B). Caso apenas a opção ‘Verdadeiro’ seja ativada, esses vizinhos unicamente serão destacados (Figura 5.8, item C). O resultado será semelhante se exclusivamente a opção ‘Ausente’ for selecionada, apenas os vizinhos correspondentes serão evidenciados (Figura 5.8, item D).

A Figura 5.7 e 5.8 retratam a base de dados *Segmentation* sendo representada na Visão de ‘Preservação de Distância’. Nela, o usuário pode escolher investigar todos os tipos de vizinhos das instâncias ou analisá-los isoladamente. De acordo com as interações

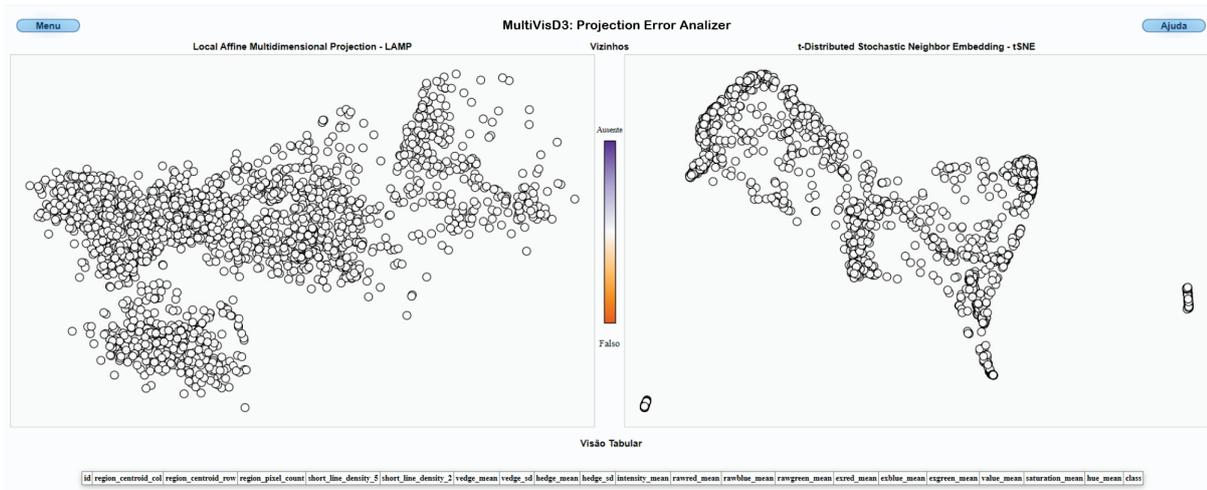


Figura 5.7 Layout da Visão ‘Preservação de Vizinhaça’.

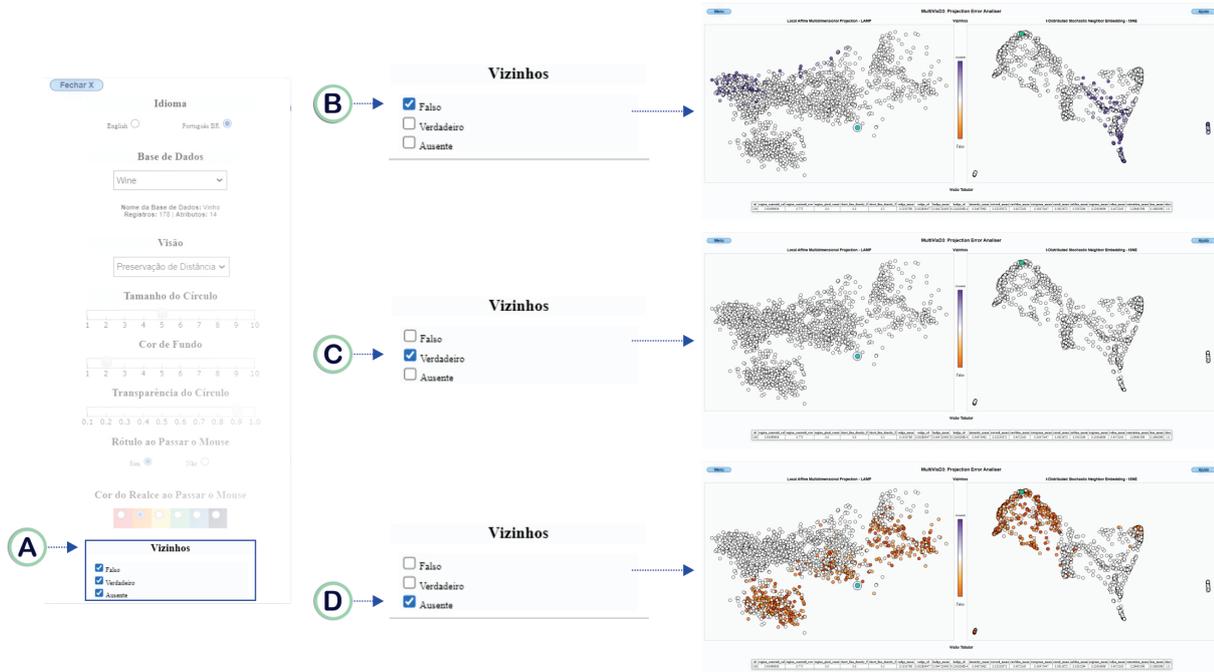


Figura 5.8 Menu da Visão ‘Preservação de Vizinhaça’.

do usuário as informações do quanto essa instância preserva sua vizinhança é mapeada na cor do ponto.

#### 5.4.2 Recursos Interativos e Coordenações entre as Visões

A abordagem proposta combina diferentes mecanismos de conexão entre as visualizações para torná-las coordenadas e refletirem as ações do usuário. Além disso, possui um

conjunto de classes e estilos de recursos interativos que possibilitam elevar o potencial analítico e expressivo das visualizações, favorecendo a descoberta de conhecimento, conforme discutidos na Seção 3.5. Os recursos interativos diretos ou indiretos utilizados, bem como as estratégias de coordenação incorporadas ao *layout*, serão apresentados a seguir:

### Recursos Interativos

As técnicas de interação foram implementadas de forma combinada para gerar mecanismos híbridos, reunindo uma ou mais funcionalidades interativas, para conectar as visualizações, conforme discutido na Seção 3.5.1.

- *Zoom* - Possibilita um maior nível de detalhes, através da aproximação e distanciamento dos itens utilizando o botão *scroll* do *mouse* nas Projeções Multidimensionais.
- Panorâmica - Ou navegação, permite que o usuário prenda a cena, clicando e mantendo pressionado o botão esquerdo do *mouse* nas Projeções Multidimensionais, em seguida movendo para qualquer direção.
- Seleção por *Click* - O usuário pode selecionar um item de dado por vez, através do *click* do *mouse*, tanto nas PMs quanto nos *Treemaps*. Essa seleção é refletida nos itens correspondentes em todas as visualizações, através do *Linking & Brushing*. Além disso, esse item será trazido para frente, ficando visível em ambas as técnicas de PMs e facilitará sua exploração. Esta interação é muito útil para lidar com a oclusão dos dados sobrepostos. Por fim, a visão tabular mostra os dados originais do registro conforme encontra-se na base.
- Seleção Livre ou Múltipla - Permite que o usuário crie uma seleção ao clicar e manter pressionado o botão direito do *mouse* e arrastar para desenhar, a mão livre, uma região ao redor dos pontos que deseja selecionar. Quando soltar o botão do *mouse* a seleção será concluída. Essa ação, permite a seleção de um ou vários itens simultaneamente nas PMs por meio movimentos arbitrários. Além disso, os itens selecionados serão trazidos para frente, ficando visíveis em ambas as técnicas de PM. A seleção dessas instâncias é refletida nas outras visualizações por meio do *Linking & Brushing*. Ao passo que a visão tabular mostra os dados originais do registro conforme encontra-se na base.
- *Mouseover* - Essa ação do usuário aciona algumas funcionalidades interativas como:
  - i) ‘Detalhes sob Demanda’, que ao passar o *mouse* sobre um item das PMs ou *Treemaps* é exibido um texto flutuante com mais informações, como o ‘id’, o valor do erro etc;
  - ii) a visão tabular mostra os dados originais do registro conforme encontra-se na base;*Linking & Brushing* - Ao passar o ponteiro do *mouse* sobre um item das visualizações ou mesmo selecionar um ou mais itens, essas instâncias são realçadas em todas as PMs e *Treemaps*.

- *Click Reset* - Após criar seu cenário de exploração dos dados o usuário pode voltar para configurações padrão de todas as visualizações, apenas clicando com o botão esquerdo do *mouse* em alguma área vazia de qualquer PM. Dessa forma é possível realizar outras investigações.
- Filtros - Há pelo menos dois tipos de filtros na aplicação: *i*) os *Sliders* de intervalos de erro agregado que realça somente as instâncias do intervalo selecionado e em paralelo a visão tabular mostra os dados originais dos registros conforme encontram-se na base; *ii*) as categorias da legenda de cores da visão ‘Classe’ que funcionam como botões e realçam as instâncias correspondentes à classe selecionada.

### Estratégias de Coordenação

As múltiplas visões coordenadas utilizam estratégias para conectar as visualizações. O MultiVisD3: *Projection Error Analyzer* combina as estratégias de coordenação para torná-las bidirecional, conformes discutidas nas Seção 3.4. As conexões utilizadas são descritas a seguir:

- Selecionar itens  $\Leftrightarrow$  Selecionar itens - Essa coordenação conecta a seleção das Projeções Multidimensionais com os *Treemaps* e vice-versa. Ao usuário selecionar um item em uma visualização, o item correspondente nas demais visualizações também são selecionados.
- Selecionar itens  $\Leftrightarrow$  Navegar itens - Essa estratégia combina a seleção de itens em uma Projeção Multidimensional com a navegação na outra projeção, e vice-versa.
- Selecionar itens  $\rightarrow$  Navegar itens - Essa estratégia combina a seleção de itens nos *Treemaps* com a navegação nas Projeções Multidimensionais.
- Codificação Visual Compartilhada - Diferentes canais visuais são compartilhados entre as representações gráficas, a exemplo do esquema de cores que é partilhado entre as PMs e os *Treemaps*. Enquanto que o tamanho e a transparência dos círculos, bem como a cor de fundo estão conectados somente entre as projeções. Por fim, o contorno ou linha de borda dos retângulos estão interligados apenas nos *Treemaps*.
- Quantidade de dados compartilhados - Na aplicação as PMs e os *Treemaps* mapeiam todo o conjunto de dados. Enquanto que a Visão Tabular utiliza a ‘detalhes sob demanda’ e projeta outros aspectos dos dados correspondentes aos itens selecionados nas outras visualizações.

## 5.5 CONSIDERAÇÕES FINAIS

Este capítulo apresentou diversos aspectos relacionados ao desenvolvimento da abordagem de *Visual Analytics* proposta, dentre os quais, o fluxo de *design*, as bases de dados

e as tecnologias utilizadas, bem como os componentes visuais do *layout*, as particularidades de cada visão ou perspectiva de exploração dos dados, as coordenações entre as visualizações e os recursos interativos.

Para cumprir com os objetivos deste estudo, o processo de desenvolvimento foi pautado em uma metodologia elaborada a partir do referencial teórico e alinhada com os trabalhos relacionados, em especial considerando o estado-da-arte do tema da pesquisa. Como contribuição deste trabalho, foi desenvolvido o MultiVisD3: *Projection Error Analyzer* para dados multidimensionais utilizando o conceito de Múltiplas Visões Coordenadas, para apoiar a interpretação da qualidade das Projeções Multidimensionais.



## AVALIAÇÃO

### 6.1 CONSIDERAÇÕES INICIAIS

Este capítulo apresenta as avaliações realizadas no MultiVisD3: *Projection Error Analyzer*. Inicialmente será apresentado um cenário de uso para destacar as funcionalidades da aplicação e, posteriormente, será apresentada uma avaliação experimental realizada com usuários para explorar elementos de funcionalidade e usabilidade. O propósito é compreender como essa abordagem pode contribuir para a interpretação de erros de Projeções Multidimensionais utilizando Múltiplas Visões Coordenadas.

### 6.2 CENÁRIO DE USO

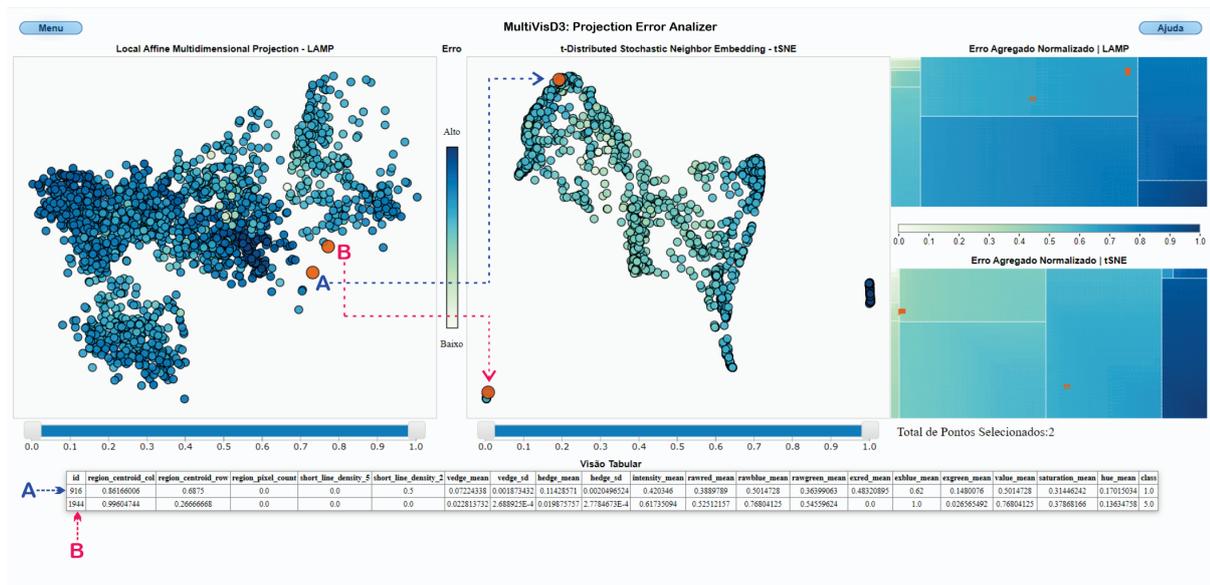
Esta seção descreve um cenário hipotético de caso de uso da abordagem de *Visual Analytics* desenvolvida, com o objetivo de apresentar as funcionalidades das técnicas de visualização, em conjunto com seus respectivos mecanismos de interação. O cenário de uso considera um usuário com conhecimento do domínio envolvendo projeções multidimensionais. Foi utilizada uma versão pré-processada do conjunto de dados *Image Segmentation*<sup>1</sup> com 21 dimensões e 2.100 instâncias, apresentada na Seção 5.2.

#### Comparação das técnicas de PM.

O MultiVisD3: *Projection Error Analyzer* segue o pipeline de visualização ilustrado na Figura 5.1. No início da exploração o usuário seleciona a base de dados *Image Segmentation* e, em seguida, são apresentadas as técnicas de visualizações, que correspondem as Projeções Multidimensionais e os *Treemaps*. Com a atenção visual voltada para as Projeções Multidimensionais, o usuário decide selecionar duas instâncias na técnica LAMP (pontos A e B da Figura 6.1). Essa seleção é refletida na projeção t-SNE e nas demais visualizações, incluindo a visão tabular, conforme apresentado na Figura 6.1.

---

<sup>1</sup>Image Segmentation. (1990). UCI Machine Learning Repository. <https://doi.org/10.24432/C5GP4N>.



**Figura 6.1** Identificação de similaridade visual de um par de instâncias entre técnicas de projeção multidimensional.

Ao observar o resultado da seleção na outra projeção (t-SNE), o usuário consegue perceber padrões visuais (de similaridade) distintos, onde, na projeção LAMP, os pontos apresentam mais proximidade, portanto, mais similaridade do que na projeção t-SNE. Ao verificar as informações da visão tabular, ele constata que os valores da base de dados que representam as instâncias selecionadas possuem diferenças significativas, o que pode justificar uma distancia maior entre os pontos.

### Análise do erro agregado nas PMs.

Dando continuidade ao processo de exploração visual, o usuário decide selecionar um conjunto de pontos (A) na t-SNE (Figura 6.2). Os mesmos pontos são realçados também na projeção LAMP. Ao comparar ambos padrões visuais ele nota que os mesmos pontos estão mais dispersos na Lamp e com menos sobreposição.

O usuário decide observar a qualidade das projeções multidimensionais na perspectiva do erro agregado mapeado na cor e considerando os intervalos normalizados de erro mapeados nos *sliders* e nos *Treemaps*. Para realizar essa exploração ele utiliza os *sliders* que realizam filtros e destacam apenas as instâncias que fazem parte de um determinado intervalo de erro normalizado. Essa interação reflete tanto no nível de transparência dos pontos nas projeções quanto nos retângulos dos *Treemaps*, conforme apresentado na Figura 6.3. O usuário escolheu o intervalo de erro agregado normalizado entre 0,7 e 0,8 em ambas projeções. Ao verificar o resultado obtido, fica evidente que a LAMP possui mais instâncias nesse intervalo de erro do que a t-SNE, não somente pela disposição dos pontos nas visões de projeção, mas, principalmente, nos *Treemaps*, que não contém sobreposição, por isso é mais fácil de perceber visualmente e comparar quantitativamente

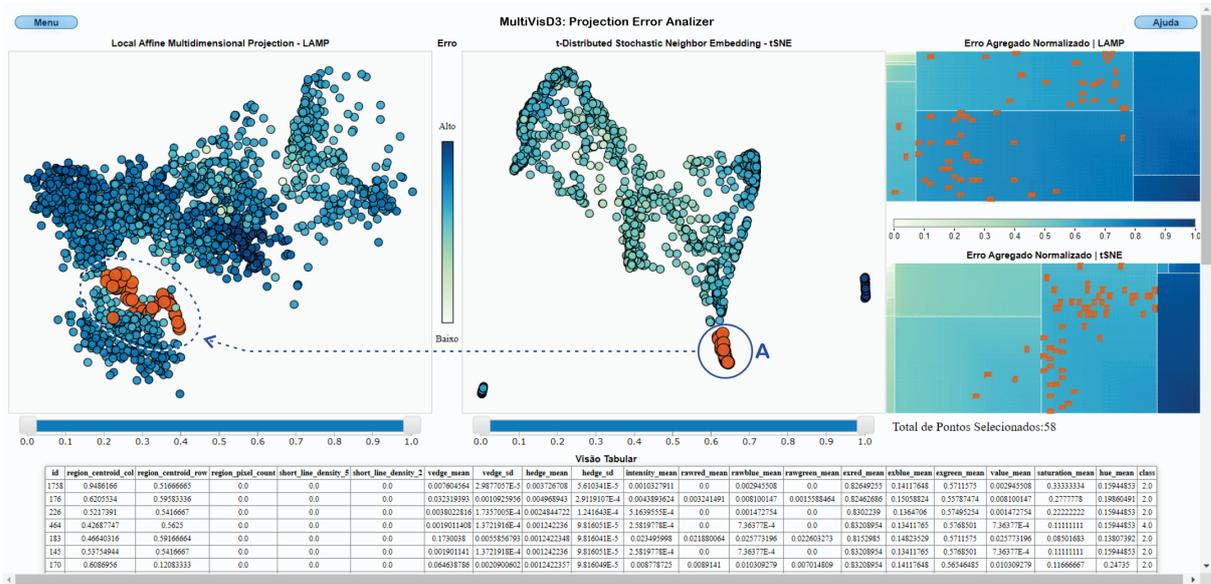


Figura 6.2 Similaridade visual de grupos de instâncias entre técnicas de PM.

as instâncias.

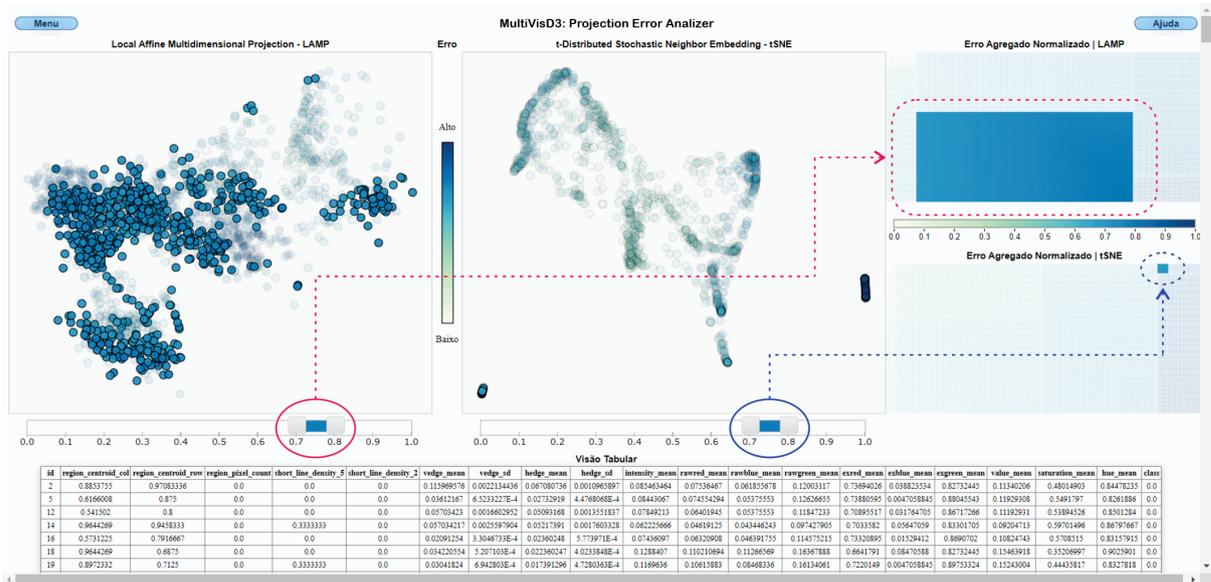


Figura 6.3 Verificação das instâncias por intervalos de erro agregado normalizados.

Após, o usuário decide comparar a qualidade de projeção de um grupo de instâncias (B) pela visão de erro agregado. Ele faz uma seleção múltipla desse conjunto de pontos na projeção t-SNE, e essa operação é refletida nas outras técnicas de visualização, como pode ser visto na Figura 6.4. Ao observar o resultado dessa seleção nos *Treemaps*, o usuário consegue interpretar que os valores de erro do conjunto dos pontos seleccionados são mais altos na técnica t-SNE, sugerindo que essas instâncias foram melhor projetadas

na técnica LAMP.

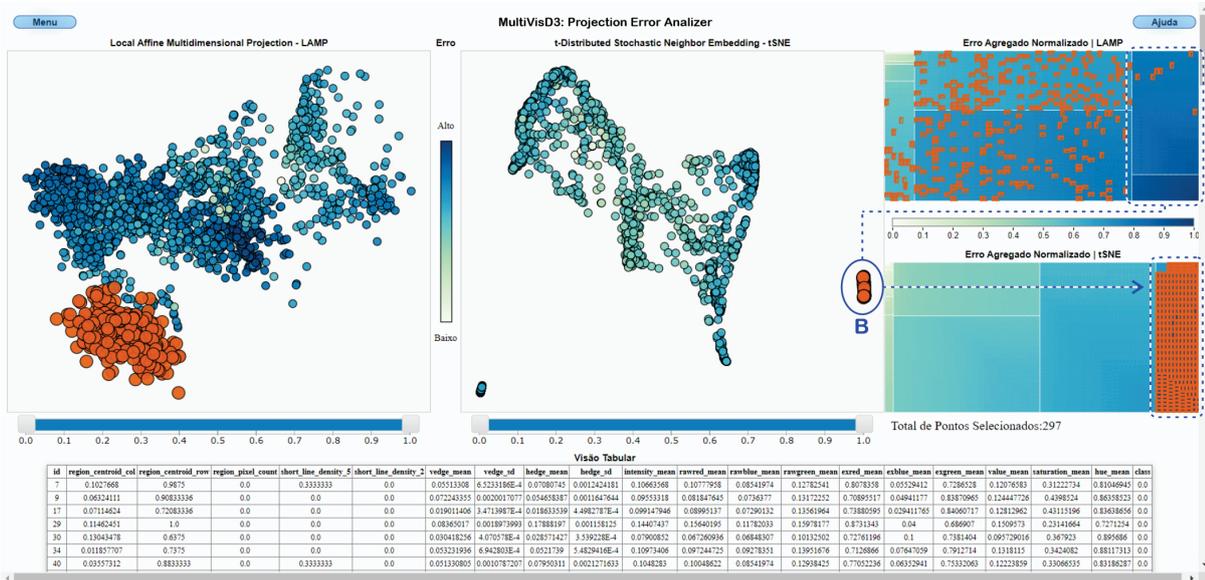


Figura 6.4 Verificação das instâncias por valor de erro agregado.

### Análise do erro agregado nas PMs com a perspectiva de classe.

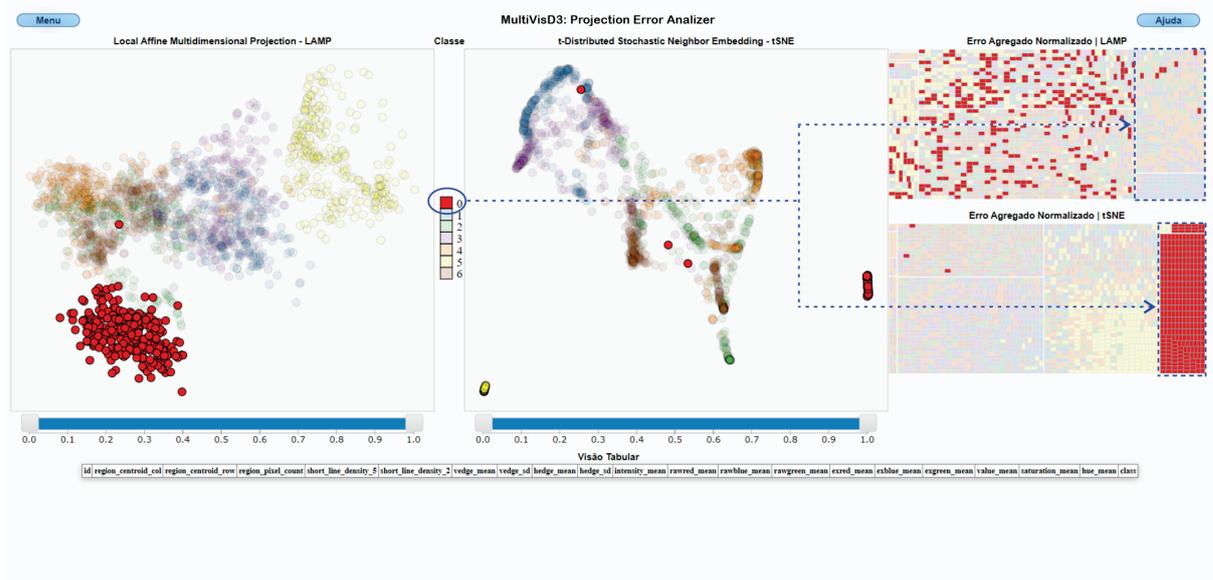
O usuário continua sua exploração acessando o menu e selecionando a visão ‘Classe’, passando a investigar como está a qualidade das projeções multidimensionais, mas desta vez considerando os rótulos da base de dados *Image Segmentation*. Lembrando que os rótulos estão associados ao atributo ‘class’ (classe), e a base possui sete rótulos: 0 (zero), 1 (um), 2 (dois), 3 (três), 4 (quatro), 5 (cinco) e 6 (seis).

O usuário utiliza o filtro interativo presente na legenda de cores para analisar o erro agregado das instâncias de cada classe. Ao selecionar a classe zero, o usuário nota através da visualização *Treemap* que existem mais pontos na projeção multidimensional t-SNE com valores mais altos de erro agregado, sinalizando uma menor qualidade na projeção dessa técnica comparado com a LAMP para essa classe específica.

Em mais detalhes, quase a totalidade das instâncias da classe zero na projeção t-SNE estão no intervalo de 0.9 a 1.0, ao passo que na LAMP as mesmas instâncias estão distribuídas em intervalos distintos, com nenhuma instância nesse intervalo. A Figura 6.5 apresenta esse cenário de uso.

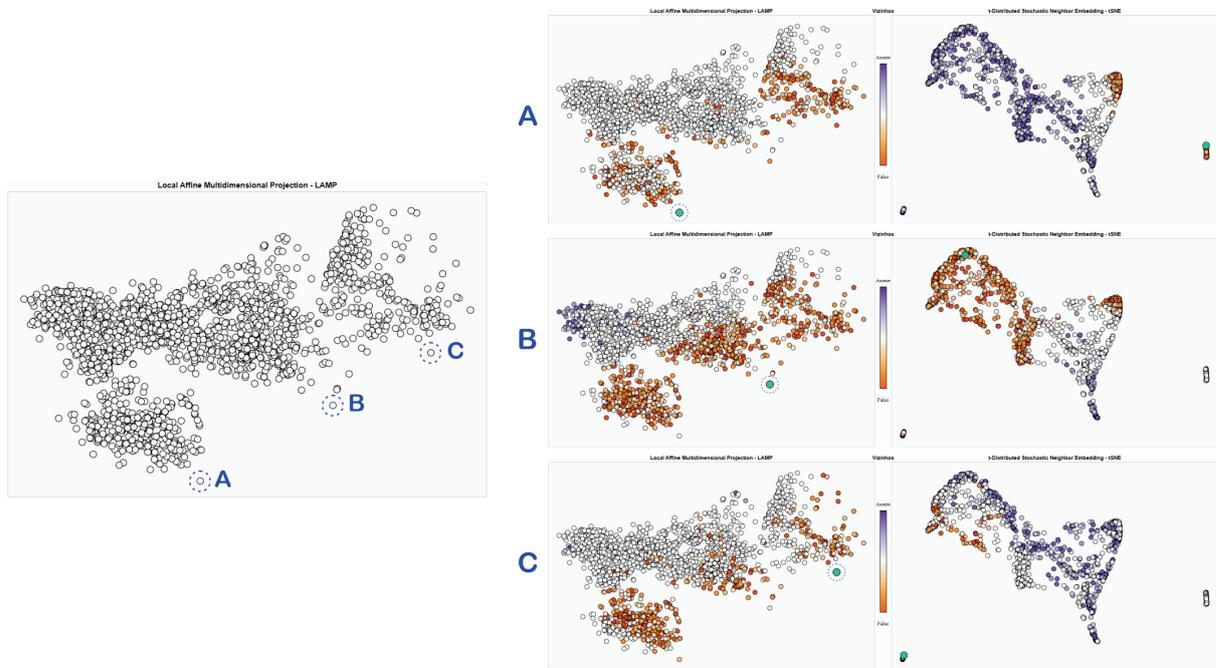
### Verificação da preservação de vizinhança nas PMs.

O usuário escolhe verificar outros aspectos dos dados e seleciona a visão ‘Preservação de Vizinhança’ no menu. Durante essa observação ele analisa três pontos (A, B e C) na técnica de projeção LAMP (Figura 6.6). Ao selecionar cada um dos pontos o usuário obtém as informações de vizinhos ausentes (cor roxo), falsos (cor laranja) e aqueles que



**Figura 6.5** Exploração do Erro Agregado de Projeções Multidimensionais na perspectiva de Classe.

preservaram a vizinhança (cor branca), considerando os dados originais. Ao comparar o resultado dos pontos selecionados em ambas as técnicas de projeção o usuário constata que a vizinhança do ponto ‘A’ na projeção LAMP foi melhor preservada, pois possui uma maior quantidade de vizinhos verdadeiros, além de não possuir vizinhos ausentes.



**Figura 6.6** Verificação da preservação de vizinhança nas PMs.

### 6.3 AVALIAÇÃO COM USUÁRIOS

As avaliações com usuários ocorreram de modo não presencial, utilizando vídeo chamadas online via plataforma *Google Meet*<sup>2</sup>. As vantagens abarcam a economia de tempo e de custo de deslocamento para realização do procedimento, possibilitando o recrutamento de uma quantidade maior de usuários uma vez que não é necessário a presença física.

Selecionou-se participantes (usuários) de duas instituições (UFBA e FIOCRUZ) para a avaliação de usabilidade. O projeto de pesquisa desta dissertação foi submetido à apreciação dos Comitês de Ética em Pesquisa (CEPs) da Universidade Federal da Bahia (Parecer N<sup>o</sup>: 5.757.369 - Apêndice C) e da Fundação Osvaldo Cruz - Bahia (Parecer N<sup>o</sup>: 5.842.393 - Apêndice D), pois a etapa de avaliação com usuários envolveu seres humanos. Então, somente após o parecer de aprovação em ambos os CEPs é que foram iniciadas as coletas de dados.

#### 6.3.1 Metodologia

O tempo médio de duração da seção de avaliação foi de aproximadamente 60 minutos, distribuídos em duas etapas principais e seus estágios, conforme apresentado na Figura 6.7. A condução da avaliação seguiu rigorosamente o fluxo descrito a seguir para todos os usuários a fim de garantir a uniformidade no processo.

A primeira etapa foi direcionada a apresentação do projeto e ao treinamento dos usuários que participaram do processo de avaliação. Esta foi dividida em quatro estágios, sendo eles:

**Apresentação** - Ao iniciar a seção de avaliação, foi apresentada brevemente o *MultiVisD3: Projection Error Analyzer*, informações do contexto da pesquisa, a aprovação pelo Conselho de Ética em Pesquisa dos procedimentos adotados, bem como o tempo de duração estimado e a quantidade de etapas. O Tempo previsto de duração deste estágio foi de 5 minutos.

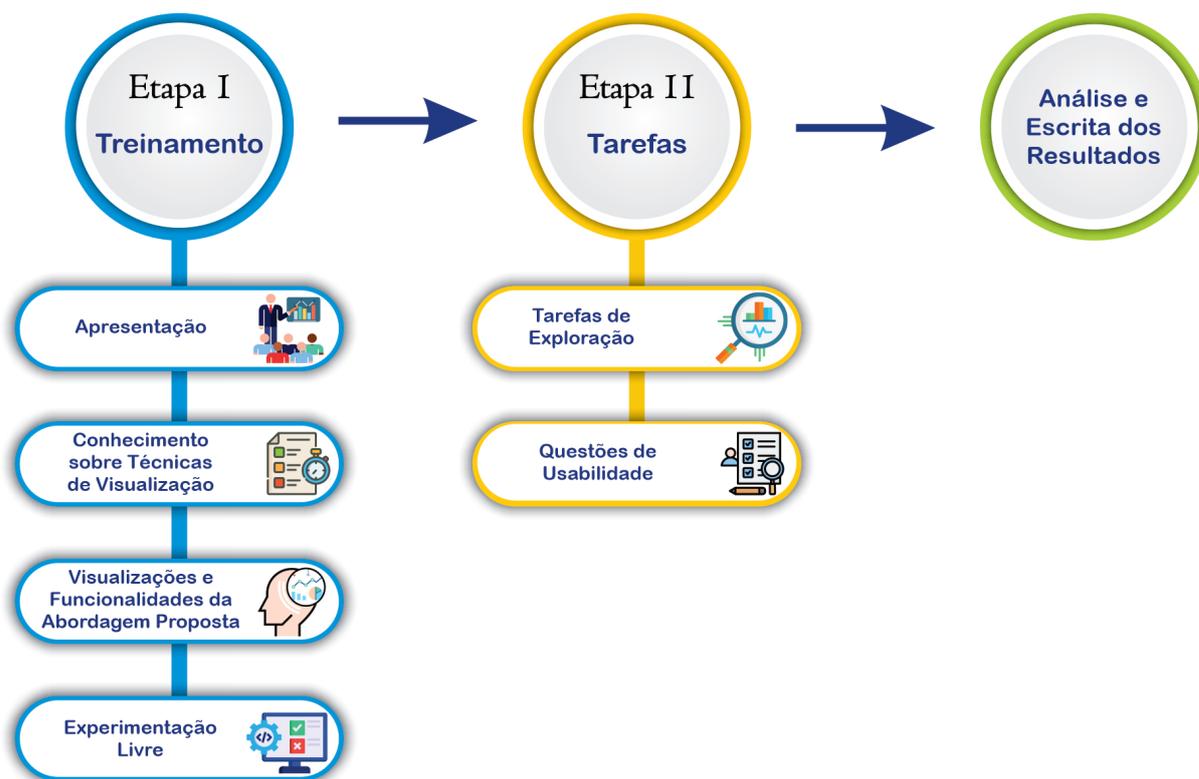
**Conhecimento Prévio das Técnicas de Visualização** - No segundo estágio, o(a) participante teve acesso a um formulário *online* (Apêndice A), onde na primeira parte continha perguntas sobre o perfil do usuário como faixa etária, formação e se o usuário tinha conhecimento sobre determinadas técnicas de visualização. Em seguida, eram explicados os principais conceitos das técnicas de Projeção Multidimensional e *Treemap*. Logo após era solicitado ao usuário que respondesse seis questões para verificar sua compreensão sobre os tópicos abordados. O Tempo previsto de duração deste estágio foi de 10 minutos.

**Visualizações e Funcionalidades do MultiVisD3** - Explicação sobre a disposição das técnicas de visualização utilizadas, os recursos das interfaces, os mecanismos de interação e a coordenação entre as visões. Além disso, foram demonstradas as principais funcionalidades da ferramenta para gerar cenários de exploração e obter conhecimento dos dados. O Tempo previsto de duração deste estágio foi de 10 minutos.

**Experimentação Livre** - O usuário pode experimentar livremente a interface da

---

<sup>2</sup>Google Meet - Videochamadas, reuniões e conferências *online* em tempo real. Disponível em: <https://meet.google.com>. Acessado em: 20/07/2023



**Figura 6.7** Fluxo da Avaliação com Usuário.

ferramenta e treinar a execução das interações e suas coordenações nas visualizações. O Tempo previsto de duração deste estágio foi de 10 minutos.

A segunda etapa do processo foi direcionada à funcionalidade contendo tarefas para avaliação da aplicação, além de perguntas relacionadas a sua usabilidade, conforme descrito nos dois estágios a seguir:

**Tarefas de Exploração** - Realização de tarefas, especificadas por questionário (Apêndice B), que envolveram comparação, identificação, localização, manipulação de diferentes tipos de mecanismos de interação como filtro, seleção, zoom, pan, seleção múltipla etc. para avaliar a qualidade das projeções multidimensionais. O Tempo previsto de duração deste estágio foi de 20 minutos.

**Questões de Usabilidade** - Após a conclusão das tarefas foi solicitado ao participante que registrasse suas percepções de usabilidade do MultiVisD3: *Projection Error Analyzer* sobre sua efetividade no processo de interpretar e avaliar a qualidade das projeções multidimensionais. O Tempo previsto de duração deste estágio foi de 5 minutos.

A última etapa deste processo avaliativo foi a análise: *i)* dos resultados das tarefas; *ii)* das informações registradas durante a observação; e *iii)* o resultado da tabulação das respostas dos questionários.

**6.3.1.1 Dispositivos de Coleta de Dados** Para coletar os dados, este estudo realizou uma avaliação com usuários para explorar elementos de funcionalidade e usabilidade

da abordagem de *Visual Analytics* desenvolvida. Os dispositivos de coleta de dados utilizados foram:

- Observação participante - para documentar as interações, comentários, dúvidas e sugestões em todas as etapas do processo de avaliação com usuários.
- Questionário estruturado - composto por questões fechadas, utilizando respostas pré-definidas com mensuração baseada na escala *Likert* com 5 níveis. Para viabilizar a aplicação do questionário foi utilizada a ferramenta *Google Forms*.

### 6.3.1.2 Participantes da Pesquisa

O perfil do participante foi definido considerando a experiência ou conhecimento sobre análise de dados e utilização de visualizações. Em seguida, foram selecionadas duas instituições que tivessem candidatos alinhados com o perfil desenhado e pudessem contribuir com a pesquisa, sendo elas:

- Instituto de Computação (IC) da Universidade Federal da Bahia (UFBA), localizado no Campus de Ondina, na Avenida Milton Santos, s/n, Ondina, Salvador-Bahia. O IC dispõe de cursos de Graduação e Pós-graduação;
- Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS) vinculado a Fundação Oswaldo Cruz (FIOCRUZ), lotado no Parque Tecnológico da Bahia: Rua Mundo, 121, Tecnocentro, Trogoy, Salvador, Bahia. O CIDACS realiza estudos e pesquisas com base em projetos interdisciplinares originados na vinculação de grandes volumes de dados na área de saúde.

Somando as pessoas selecionadas nas duas instituições, foram recrutados um total de catorze participantes, através de convite individual, por e-mail, dos quais: sete eram estudantes do IC da UFBA e membros do grupo de pesquisa de Visualização de Dados; e sete eram colaboradores do CIDACS associados aos Núcleos de Produção e Visualização de Dados. Em relação ao tamanho da amostra, essa escolha fundamenta-se nos trabalhos relacionados que envolvem pesquisa com usuários como Pillat (2006) e Godinho et al. (2007).

### 6.3.1.3 Objetivos do Processo de Avaliação e Tarefas de Exploração

O propósito da avaliação é mensurar o quão eficiente o usuário consegue ser quando realiza a exploração visual no MultiVisD3: *Projection Error Analyzer*, criando seus próprios cenários de análise para responder tarefas específicas. Para isso foram definidos objetivos que nortearam a avaliação, de modo a: organizar a etapa de treinamento de modo mais didático; idealizar cenários de uso hipotético; e elaborar as tarefas da avaliação com usuários. As tarefas de exploração propostas requerem a utilização dos diferentes mecanismos de interação da aplicação para interpretação de erros de Projeções Multidimensionais utilizando Múltiplas Visões Coordenadas.

Os objetivos que apoiaram a estruturação desse processo, bem como as tarefas serão descritos de forma sumarizada na Tabela 6.1, essas últimas são apresentadas por completo no questionário de avaliação (Apêndice B).

Objetivos	Tarefas	Descrição da Tarefa
1 - Avaliar a perspectiva de similaridade visual entre técnicas de PM distintas.	1 - Avaliar a similaridade visual de um par de pontos nas PMs.	Seleção de um par de pontos e avaliação da similaridade visual entre eles em ambas as técnicas de PM.
	2 - Avaliar a similaridade visual de um grupo de pontos nas PMs.	Seleção de um grupo de pontos, comparação do seu padrão visual em ambas as projeções e indicar em qual delas existe maior similaridade dos pontos selecionados.
2 - Verificar a qualidade da PM através da métrica de erro agregado.	3 - Verificar em cada PM a quantidade de instâncias com erro agregado para um determinado intervalo de valores normalizados.	Utilizar o recurso interativo <i>slider</i> para selecionar um intervalo de erro agregado em ambas as técnicas de projeção.
	4 - Analisar a qualidade da projeção de instâncias através da métrica de erro agregado.	Seleção de um conjunto de pontos em uma projeção, analisar visualmente as projeções e os <i>Treemaps</i> e identificar em qual projeção os valores de erro agregado dos pontos selecionados são mais altos.
	5 - Verificar a qualidade da PM por meio de intervalos normalizados dos valores da métrica de erro agregado.	Utilizar os <i>sliders</i> para filtrar um intervalo de erro agregado normalizado em ambas as projeções. Em seguida, observar o resultado nas PMs e <i>Treemaps</i> e responder em qual projeção tem uma maior quantidade de instâncias nesse intervalo contendo erros mais baixos.
3 - Analisar a qualidade da PM considerando a perspectiva do atributo ‘classe’ que caracteriza o rótulo do conjunto de dados.	6 - Avaliar a qualidade das PMs considerando o atributo rótulo da base de dados.	Na visão de ‘Classe’ utilizar o filtro interativo presente na legenda de cores para realçar as instâncias da categoria solicitada. Após, avaliar os intervalos de erro agregado nos <i>Treemaps</i> e identificar em qual técnica de PM as instâncias da classe selecionada apresentam valores mais altos de erro.
	7 - Analisar a qualidade da PM nas instâncias agrupadas na visão de Classe utilizando coordenação.	Na visão de ‘Classe’ utilizar: os <i>sliders</i> para filtrar um intervalo de erro nas duas PMs; e o filtro presente na legenda de cores para realçar as instâncias da classe requerida. Após, analisar o resultado, especialmente dos <i>Treemaps</i> , e indicar em qual técnica de PM a classe selecionada apresenta valores altos de erro agregado no intervalo determinado.
4 - Explorar a preservação de vizinhança dos pontos nas PMs.	8 - Analisar a qualidade das PMs na perspectiva de preservação de vizinhança.	Na visão de ‘Preservação de Distância’ mudar a cor do realce do <i>mouse</i> para verde e analisar a vizinhança de três pontos nas PMs.
	9 - Investigar a preservação de vizinhança de instâncias nas PMs.	Manter a configuração anterior, selecionar e comparar a vizinhança de uma instância nas PM. Por fim, indicar em qual de projeção o ponto possui mais e menos vizinhos ausentes e falsos.

**Tabela 6.1** Objetivos do Processo de Avaliação, Tarefas de Exploração e Suas Descrições.

#### 6.3.1.4 Questões de Usabilidade

Ao final da realização das tarefas de exploração, onde o usuário manuseou os recursos do MultiVisD3: *Projection Error Analyzer*, era solicitado que respondesse um conjunto de questões do tipo *likert*, relacionadas a usabilidade, com cinco níveis de concordância, sendo 1 para ‘discordo totalmente’ e 5 ‘concordo totalmente’.

Uma escala *Likert* consiste em uma ou mais afirmações ou perguntas acompanhadas por uma série de opções de resposta. Estas permitem aos pesquisadores coletar estimativas quantitativas de características subjetivas, produzindo dados numéricos que podem ser resumidos e visualizados de maneira semelhante a outros dados quantitativos coletados em uma avaliação (SOUTH et al., 2022). A seguir, serão elencadas as perguntas utilizadas nessa etapa:

- (P1) - Você concorda que a ferramenta foi útil para avaliar aspectos de qualidade das PMs?
- (P2) - Você concorda que os diferentes mecanismos de interação disponíveis na ferramenta auxiliaram na exploração dos dados e compreensão da informação?
- (P3) - Você concorda que utilizar diferentes visualizações na mesma tela com interações coordenadas entre elas colaboraram para melhor analisar os erros nas projeções?
- (P4) - Você concorda que a visualização *Treemap* auxilia na identificação dos intervalos de erros agregados dos pontos?
- (P5) - Você concorda que mapear os erros agregados em duas projeções justapostas (lado a lado) auxilia a compreender e comparar as suas respectivas qualidades.
- (P6) - Você concorda que a visão de classe contribui para avaliar a qualidade das projeções pela perspectiva dos erros em cada classe da base de dados?
- (P7) - Você concorda que a visão de ‘preservação de vizinhança’ contribui para entender como estão dispostos os vizinhos das instâncias no espaço original e no espaço projetado?

### 6.3.2 Resultados

**Perfil do Participante** - A população desse estudo com usuários foi composta por 14 participantes, que incluiu pessoas com perfis distintos, seja em relação a formação acadêmica, experiência com o tema desse projeto e faixas etárias. A Figura 6.8 apresenta os dados compilados com todas essas características. Na faixa etária, a metade, ou seja, 50% (7), têm entre 30 e 39 anos, 35,71% (5) têm entre 18 e 29 anos e 14,29% (2) têm idade entre 40 e 50 anos. A formação acadêmica predominante é o mestrado com 35,71% (5), seguida de especialização com 28,57% (4), graduação com 21,43% (3), doutorado com 7,14% (1) e ensino médio com 7,14% (1). Em relação a familiaridade com visualização de dados, o percentual de familiaridade positiva ultrapassa 75%, sendo que 42,86% (6) declararam ter muita familiaridade, 35,71% (5) afirmaram ter total familiaridade, 14,29% (2) têm alguma familiaridade e apenas 7,14% (1) expressaram ter pouca familiaridade.

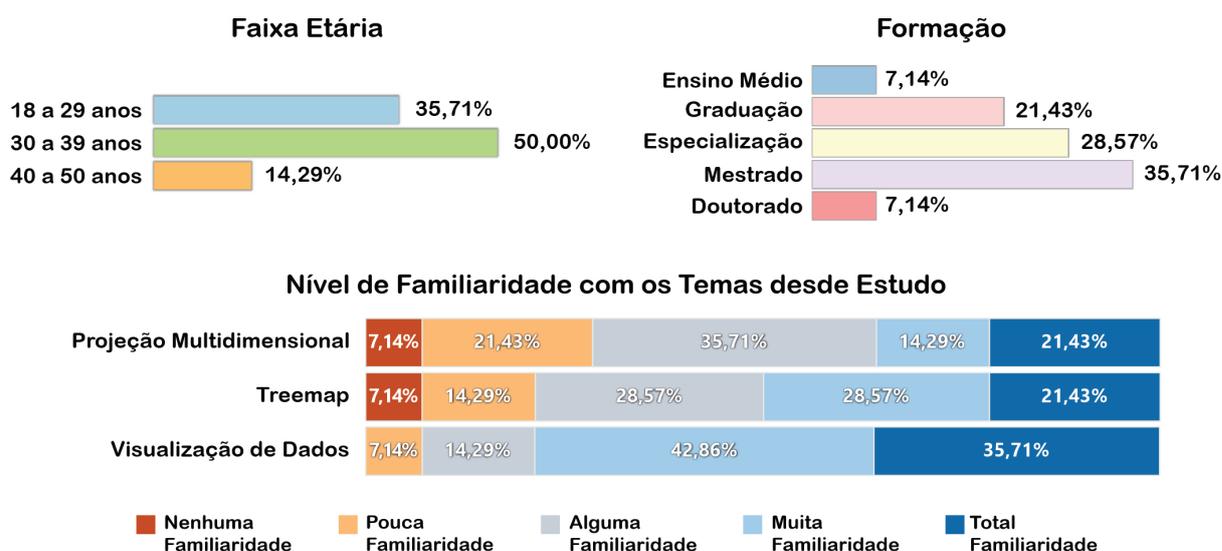


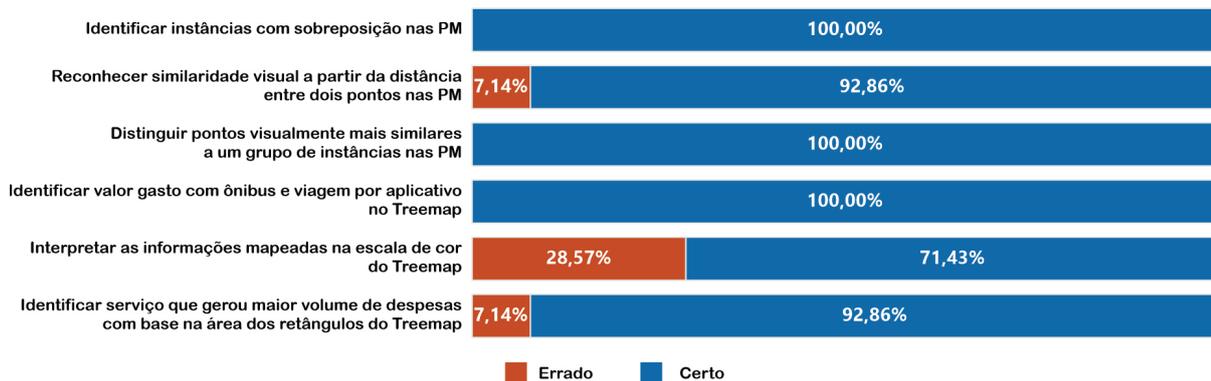
Figura 6.8 Perfil dos Participantes.

O estudo também verificou qual nível de familiaridade os usuários declaravam ter sobre as duas principais técnicas de visualização utilizadas na ferramenta, sendo elas: *i*) a Projeção Multidimensional, onde 35,72% dos usuários apresentam uma familiaridade positiva, sendo 21,43% (3) com total familiaridade e 14,29% (2) com muita familiaridade. Já um grupo de participantes correspondente a 35,71% (5) se identificou com uma familiaridade neutra, ou seja, responderam ter alguma familiaridade.

Por fim, 21,43% (3) expressaram ter pouca familiaridade e somente 7,14% (1) declararam não ter nenhuma familiaridade com a técnica; e *ii*) o *Treemap*, em que os usuários apresentaram uma maior familiaridade positiva, comparado à técnica anterior, considerando 21,43% (3) com total familiaridade e 28,57% (4) tendo muita familiaridade. Aqueles que declararam ter apenas alguma familiaridade corresponderam a 28,57% (4) e os que não possuíam familiaridade totalizaram 7,14% (1).

**Conhecimentos prévios das Técnicas de Visualização utilizadas** - Durante a etapa de treinamento, foram apresentados aos participantes os conceitos gerais utilizados pelas técnicas de Projeção Multidimensional e *Treemap*, em seguida foi solicitado que respondessem três tarefas envolvendo as Técnicas, considerando seus conhecimentos prévios (Apêndice A) e o resultado é mostrado na Figura 6.9. A análise das respostas revelou um alto nível de assimilação das definições explicadas. Das três tarefas que envolviam conhecimentos relacionados às Projeções Multidimensionais, em duas delas, envolvendo a identificação de instâncias com sobreposição e distinção de pontos visualmente mais similares a um determinado grupo de instâncias, 100% (14) dos participantes responderam corretamente. E na tarefa que tratava sobre o reconhecimento de similaridade visual a partir da distância entre dois pontos houve um alto percentual de acerto (92,86% (13)), e apenas 7,14% (1) dos usuários não acertaram.

Quanto as questões relacionadas a interpretação de informações na Técnica *Treemap*, quando a pergunta envolveu o mapeamento de valores das despesas mensais com serviços



**Figura 6.9** Conhecimentos Prévios das Técnicas de Visualização Utilizadas.

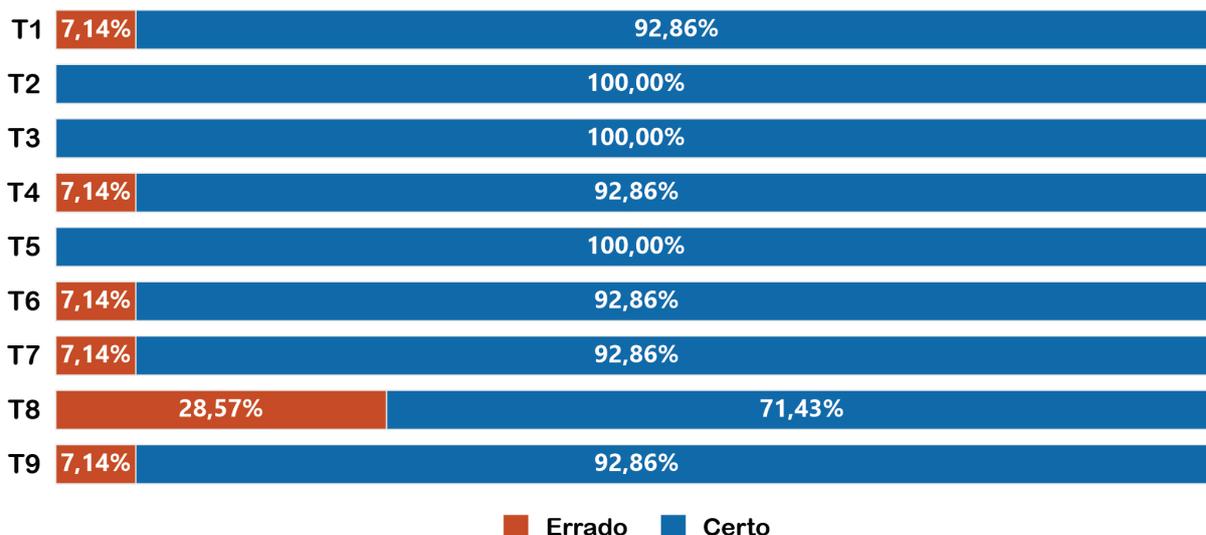
básicos no tamanho dos retângulos, 100% dos participantes conseguiram identificar o total de gastos com Ônibus e Viagens por Aplicativo. Na questão em que o atributo cor do *Treemap* mapeou os valores da métrica de erro agregado de projeção, 71,43% (10) dos usuários conseguiram associar corretamente a escala de cores aos valores correspondentes, enquanto 28,57% (4) não acertaram.

Na pergunta onde o usuário teve que considerar a área dos retângulos para identificar dentre os tipos de serviço de alimentação o que gerou o maior volume de despesa mensal, 92,86% (13) dos usuários acertaram e somente 7,14% (1) não marcaram a resposta correta. Vale ressaltar que, embora os usuários tenham expressado estar mais familiarizados com a *Treemap*, o resultado foi o oposto quando aplicadas determinadas tarefas para interpretação visual das duas técnicas.

**Resultado das Tarefas de Exploração** - Durante a execução das tarefas os usuários tiveram a oportunidade de combinar as funcionalidades do MultiVisD3: *Projection Error Analyzer* para criar os cenários de exploração e responder livremente as questões propostas. Considerando os resultados, apresentados na Figura 6.10, de maneira geral, o percentual de acerto em todas as tarefas superou 70%.

As tarefas T1 e T2 compõem o Objetivo 1 e estão relacionadas com a avaliação da similaridade visual de pontos isolados e grupos de pontos nas Projeções Multidimensionais, respectivamente. Na T1 92,86% (13) dos usuários conseguiram identificar corretamente a técnica de projeção em que os pontos selecionados tinham maior similaridade visual, e na T2 o percentual de assertividade foi ainda maior, 100% (14). Observou-se, portanto, que os usuários conseguiram mais eficácia em identificar similaridade visual em grupos de pontos.

Nas tarefas T3, T4 e T5, relacionadas ao Objetivo 2, o usuário teve que analisar a qualidade da projeção das instâncias a partir da métrica de erro agregado. Para realização dessas tarefas era essencial utilizar as informações mapeadas na técnica *Treemap*. Os resultados mostraram que a maioria dos participantes conseguiram compreender o mapeamento de erro agregado nas técnicas de Projeção Multidimensional. Para as tarefas T3 e T5, que necessitavam contar a quantidade de instâncias com erro, foi registrado 100% (14) de acertos, enquanto que a tarefa T4 foi de 92,86% (13). Isso indica que houve mais



**Figura 6.10** Respostas das Tarefas de Exploração.

dificuldade em compreender o valor do erro do que quais instâncias tinham erro.

As tarefas T6 e T7 estavam associadas ao Objetivo 3 e tiveram como foco a utilização do atributo rótulo da base de dados para apoiar na avaliação da qualidade das projeções. Em ambas, os usuários alcançaram o mesmo percentual de acertos, 92,86% (13). Por fim, nas tarefas T8 e T9, relacionadas ao Objetivo 4, o usuário realizou a análise da qualidade das Projeções Multidimensionais por meio da perspectiva de preservação de vizinhança. Na T8, 71,43% (10) os participantes conseguiram identificar corretamente vizinhos Falsos e Ausentes, enquanto que na T9 foram 92,86% (13) de acertos. A tarefa T8 foi a que obteve menor porcentagem de acerto dentre todas, e indica uma dificuldade dos usuários ao comparar cenários distintos por meio da seleção de pontos.

**Nível de Complexidade das Tarefas** - Ao concluir cada tarefa, era solicitado ao usuário que informasse o nível de complexidade da experiência. A Figura 6.11 sumariza as respostas coletadas. De forma geral, o nível de complexidade que predominou na maioria das tarefas foi muito baixa ou baixa. Houve ainda um percentual de participantes que variou entre 14,29% a 42,86%, que considerou a complexidade das tarefas como mediana. Nas tarefas T3, T5, T6 e T7, uma quantidade pequena de pessoas, 7,14% (1), afirmaram que a complexidade foi alta.

A tarefa T8, que solicitava uma análise da qualidade das PMs na perspectiva de preservação de vizinhança, foi considerada de alta complexidade por 21,43% (3) dos participantes. Sendo também a que obteve o maior percentual de respostas incorretas, 28,57% (4). Se compararmos as respostas dos resultados das tarefas com a complexidade, é possível perceber que a complexidade alta e tarefas com erro teve convergência somente para T6, T7 e T8. Isso indica que a visão de erro agregado não teve associação entre complexidade e eficácia.

**Nível de Complexidade do MultiVisD3** - A avaliação do nível de complexidade

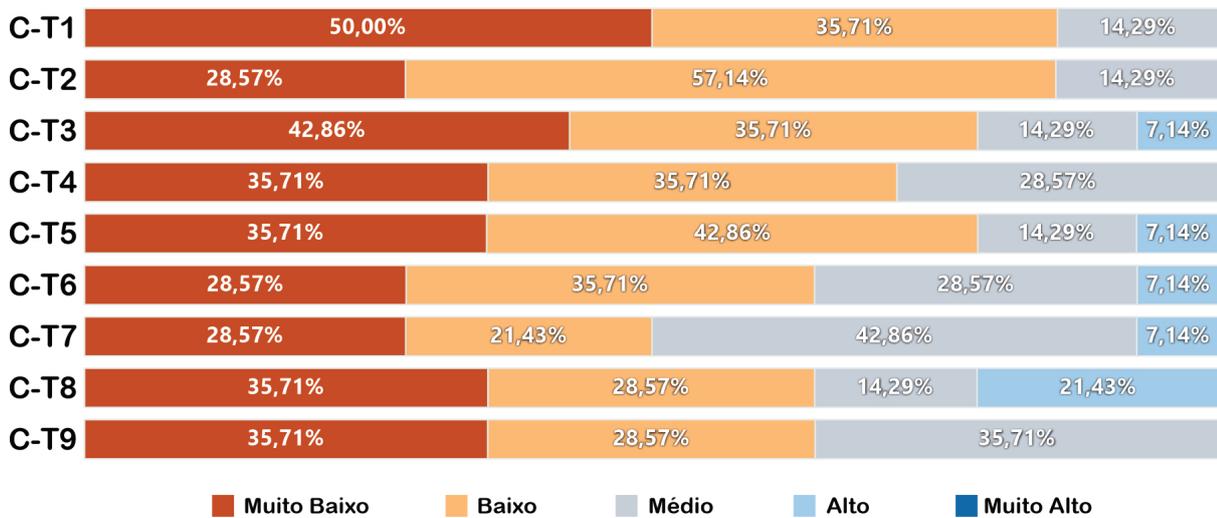


Figura 6.11 Nível de Complexidade das Tarefas de Exploração.

da aplicação, de modo geral, foi positiva e sugere uma baixa curva de aprendizagem. A maioria dos usuários (42,86% (6)) afirmaram que o nível de complexidade para manusear a ferramenta e seus recursos na realização das tarefas (Figuras 6.12) foi baixo e 28,57% (4) consideraram muito baixo. Para 14,29% (2) a ferramenta teve uma complexidade mediana. Ainda, para 14,28% (2) a ferramenta foi complexa, com 7,14% (1) classificando como complexidade alta e 7,14% (1) como muito alta.

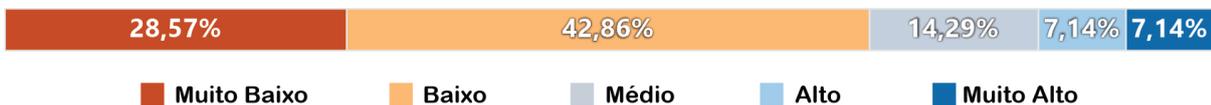


Figura 6.12 Nível de Complexidade para Manusear o MultiVisD3: *Projection Error Analyzer*.

**Questões de usabilidade** - O MultiVisD3: *Projection Error Analyzer* foi avaliado positivamente pelos participantes nas sete perguntas (de P1 a P7 na Figura 6.13) relacionadas a usabilidade. Na **P1** 78,57% (11) os usuários concordaram totalmente e 21,43% (3) concordaram que a ferramenta foi útil para avaliar aspectos de qualidade das Projeções Multidimensionais. Quando questionados na **P2** sobre os diferentes mecanismos de interação disponíveis na ferramenta, 64,29% (9) dos participantes concordaram totalmente e 35,71% (5) concordaram que esses auxiliaram na exploração dos dados e compreensão das informações.

A **P3** indagou sobre os benefícios de utilizar diferentes visualizações na mesma tela com interações coordenadas entre elas e a maior parte dos usuários (92,86% (13)) concordou totalmente que elas colaboraram para melhor analisar os erros nas projeções, enquanto que 7,14% (1) ficaram neutros, não concordaram e não discordaram. Na **P4** 57,14% (8) dos participantes concordaram totalmente e 42,86% (6) concordaram que a visualização *Treemap* auxiliou na identificação dos intervalos de erros agregados dos pontos. Na **P5** 92,86% (13) dos usuários concordaram totalmente e 7,14% (1) concordaram

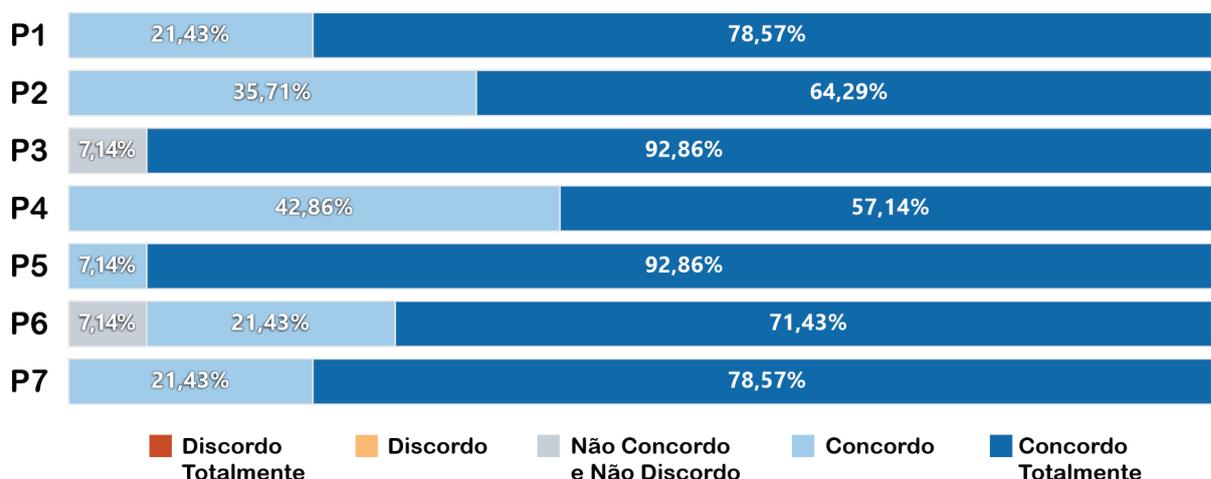


Figura 6.13 Percepção de Usabilidade do Usuário.

que mapear os erros agregados em duas projeções justapostas (lado a lado) auxiliou a compreender e comparar as suas respectivas qualidades.

Na **P6** 71,43% (10) dos participantes concordaram totalmente e 21,43% (3) concordaram que a visão de ‘Classe’ contribuiu para avaliar a qualidade das projeções. Os 7,14% (1) restantes ficaram neutros, não concordaram e não discordaram. Na **P7** 78,57% (11) dos usuários concordaram totalmente e 21,43% (3) concordaram que a visão de ‘Preservação de Vizinhança’ contribuiu para entender como estão dispostos os vizinhos das instâncias no espaço original e no espaço projetado.

### 6.3.3 Discussão

Ao analisar os resultados da avaliação foi possível perceber que os usuários declararam, inicialmente, ter uma maior familiaridade com visualização de dados, porém um menor domínio das técnicas de Projeção Multidimensional e *Treemap* (Figura 6.8). Isso reforçou a necessidade de realizar um processo de treinamento com didática suficiente para preencher essas lacunas de conhecimento e habilitar os participantes à utilizarem o MultiVisD3: *Projection Error Analyzer* para realizarem as tarefas. Outro aspecto observado foi que após as explicações dos principais conceitos das técnicas supracitadas os usuários responderam seis questões relacionadas, das quais conseguiram alcançar 100% de aproveitamento em três perguntas e mais de 70% nas outras três (Figura 6.9).

A abordagem de *Visual Analytics* desenvolvida possui um layout com várias visualizações na mesma tela e uma diversidade de recursos interativos, no entanto, a maioria dos usuários, mais de 70%, classificaram seu manuseio como de baixa ou muito baixa complexidade. Essa avaliação foi percebida como positiva e sugere que a ferramenta é intuitiva e possui uma curva de aprendizagem baixa. Outro dado que corrobora com essa percepção é o fato dos participantes conseguirem concluir oito das nove tarefas com um percentual de acerto acima de 90% (Figura 6.10).

Contudo, observou-se que os usuários tiveram mais dificuldades em realizar tarefas relacionadas análise da qualidade das PMs na perspectiva de preservação de vizinhança.

Verificando a complexidade das tarefas, a tarefa T8 por exemplo, foi apontada como a mais complexa por 21,43% (3) dos participantes. Sendo também essa tarefa a que obteve o maior percentual de respostas incorretas com 28,57% (4).

Em relação ao objetivo principal deste processo de avaliação, que era compreender se o MultiVisD3: *Projection Error Analyzer* pode contribuir para a interpretação de erros de Projeções Multidimensionais utilizando Múltiplas Visões Coordenadas, as respostas (Figura 6.13) referentes a experiência de usabilidade revelou que os usuários consideraram que:

- a ferramenta foi útil para avaliar aspectos de qualidade das Projeções Multidimensionais;
- os diferentes mecanismos de interação disponíveis na ferramenta auxiliou na exploração dos dados e compreensão das informações;
- a utilização de diferentes visualizações na mesma tela com interações coordenadas entre elas colaboraram para melhor analisar os erros nas projeções;
- a visualização *Treemap* auxiliou na identificação dos intervalos de erros agregados das projeções;
- o mapeamento da métrica de erro agregado em duas projeções justapostas (lado a lado) auxiliou a compreender e comparar as suas respectivas qualidades;
- a visão de ‘Classe’ contribuiu para avaliar a qualidade das projeções;
- a visão de ‘Preservação de Vizinhança’ contribuiu para entender como estão dispostos os vizinhos das instâncias no espaço original e no espaço projetado.

## 6.4 CONSIDERAÇÕES FINAIS

Este capítulo descreveu os procedimentos de avaliação realizados no MultiVisD3: *Projection Error Analyzer*, dentre eles, um cenário de uso hipotético para apresentar os principais recursos, além de um teste de exploração da aplicação com usuários para analisar as funcionalidades e usabilidade. Além disso, são apresentados e discutidos os resultados, a partir da análise dos dados coletados.



## CONCLUSÃO

Este capítulo apresenta a conclusão do presente estudo e destaca aspectos importantes dessa pesquisa com uma visão geral das contribuições desse trabalho, contendo as principais respostas das questões de pesquisa, as principais limitações do trabalho, e os direcionamentos para trabalhos futuros.

### 7.1 CONTRIBUIÇÕES

O MultiVisD3 é uma abordagem de *Visual Analytics* que pode ser executada diretamente no navegador, não sendo necessário realizar nenhum tipo de instalação ou configuração no dispositivo do usuário. Seu intuito é colaborar com a análise da qualidade das Projeções Multidimensionais por meio da exploração visual de seus erros/distorções. Este trabalho apresentou um painel interativo com múltiplas visões coordenadas, que contribui para a exploração visual de erros em PMs. A elaboração deste estudo foi pautado nas questões de pesquisa, apresentadas na Seção 1.5, que direcionaram o processo de desenvolvimento. As respostas a essas questões, serão apresentadas a seguir e representam as contribuições deste trabalho:

A reflexão da **QP1** foi sobre a possibilidade de comparar a qualidade de duas projeções multidimensionais distintas. Em retorno, o MultiVisD3: *Projection Error Analyzer* foi projetado para que um conjunto de dados fosse representado graficamente por duas técnicas de PMs diferentes na mesma tela. Essa posição justaposta em conjunto com a coordenação entre as duas visualizações possibilitam comparação imediata das projeções e suas respectivas métricas de erros.

Na **QP2** a indagação é de como diferentes métricas de erros poderiam ser visualizadas em diferentes projeções multidimensionais. Em resposta, foi desenvolvido um *layout* flexível que comporta perspectivas ou visões diferentes dos dados para diferentes métricas de erros. Essa contribuição permitiu que as métricas de erro agregado e preservação de vizinhança das diferentes técnicas de PMs pudessem ser analisadas.

A **QP3** retratou sobre a viabilidade de associar o valor dos erros das projeções multidimensionais com representações hierárquicas. Em resposta, foi criada uma estrutura

hierárquica contendo os valores de erro agregado normalizados e separados por intervalos. Isso favoreceu a utilização da técnica *Treemap*, que contribuiu para uma melhor interpretação da qualidade das PMs.

O questionamento da **QP4** foi sobre a eficiência e utilidade da abordagem desenvolvida. A implementação do MultiVisD3: *Projection Error Analyzer* combinou diferentes estratégias utilizando o conceito de Múltiplas Visões Coordenadas contendo mecanismos de interação coordenados entre diferentes visualizações. De acordo com os resultados da avaliação realizada pelos usuários, houve uma qualificação positiva dos aspectos relacionados a **QP4**. Dentre eles, destacam-se o percentual de 100% que avaliaram:

- A ferramenta como útil para analisar aspectos de qualidade das Projeções Multidimensionais;
- Os diferentes mecanismos de interação como fundamentais na exploração dos dados e compreensão das informações;
- O *Treemap* como relevante para identificar erros nas PM.

Além disso, mais de 90% dos participantes apontaram que utilizar diferentes visualizações na mesma tela com interações coordenadas entre elas colaborou para melhor analisar os erros nas projeções. Por outro lado, observou-se que os usuários tiveram mais dificuldades em realizar tarefas relacionadas análise da qualidade das PMs na perspectiva de preservação de vizinhança.

## 7.2 LIMITAÇÕES

Embora os resultados apresentados tenham sido positivos, algumas limitações podem ser observadas. No estado atual, é possível comparar somente duas técnicas de Projeções Multidimensionais. Do mesmo modo, é possível inspecionar apenas duas métricas de erro. A maior limitação técnica de acessibilidade está relacionada a falta de responsividade para o redimensionamento de janelas e a necessidade de adaptar o *layout* tanto para telas menores, como as de dispositivos móveis, quanto para telas sensíveis ao toque, onde, interações de seleção podem ser um desafio em decorrência do tamanho pequeno dos objetos visuais que representam as instâncias (pontos nas projeções ou retângulos no *Treemap*).

Outro aspecto limitante está em não permitir ao usuário inserir um conjunto de dados de sua escolha. No estado atual da abordagem, as rotinas de pré-processamento estão prontas e, portanto, seria necessário somente acoplar à ferramenta essa possibilidade. Contudo, a depender das características do conjunto de dados, as rotinas de cálculo das métricas de erro podem ocasionar um tempo de espera excessivo para o usuário.

## 7.3 TRABALHOS FUTUROS

Os estágios para a continuidade desse estudo vão ao encontro de soluções para contornar as limitações supracitadas, como escalar a quantidade tanto das técnicas de PM quanto das métricas de erro para ampliar a capacidade de comparação dos dados no espaço de

baixa dimensão, além de elevar a capacidade de mensuração de outros tipos de erros e distorções presentes nos mapeamentos das PMs.

Para o problema das limitações de responsividade, adaptação para telas menores e/ou sensíveis ao toque, uma solução é reescrever o *layout* do MultiVisD3: *Projection Error Analyzer* incluindo tecnologias responsivas, como *Bootstrap* por exemplo, de modo a comportar essas funcionalidades. Por fim, quanto aos conjuntos de dados, considerar as possibilidades de: *i*) adicionar bases de dados mais robustas e complexas, com centenas para milhares de atributos e dezenas para centenas de milhares de instâncias; e *ii*) adicionar a opção do usuário inserir seu próprio conjunto de dados, de modo a realizar as etapas de pré-processamento em tempo de execução.



## REFERÊNCIAS BIBLIOGRÁFICAS

- AHLBERG, C.; WILLIAMSON, C.; SHNEIDERMAN, B. Dynamic queries for information exploration: An implementation and evaluation. In: ACM. *Proceedings of the SIGCHI conference on Human factors in computing systems*. [S.l.], 1992. p. 619–626.
- ALMEIDA, L. H. et al. Webprisma: An interactive web-based tool for exploratory visualization using multiple coordinated views. In: IEEE. *Information Visualisation, 2009 13th International Conference*. Barcelona, Spain, 2009. p. 645–650.
- AOKI, P. H. et al. Molecularly designed layer-by-layer (lbl) films to detect catechol using information visualization methods. *Langmuir*, ACS Publications, v. 29, n. 24, p. 7542–7550, 2013.
- BALDONADO, M. Q. W.; WOODRUFF, A.; KUCHINSKY, A. Guidelines for using multiple views in information visualization. In: ACM. *Proceedings of the working conference on Advanced visual interfaces*. [S.l.], 2000. p. 110–119.
- BEDERSON, B. B.; SHNEIDERMAN, B.; WATTENBERG, M. Ordered and quantum treemaps: Making effective use of 2d space to display hierarchies. *ACM Trans. Graph.*, ACM, New York, NY, USA, v. 21, n. 4, p. 833–854, out. 2002. ISSN 0730-0301. Disponível em: <<http://doi-acm-org.ez10.periodicos.capes.gov.br/10.1145/571647.571649>>.
- BOUKHELIFA, N.; ROBERTS, J. C.; RODGERS, P. J. A coordination model for exploratory multiview visualization. In: IEEE. *Coordinated and Multiple Views in Exploratory Visualization, 2003. Proceedings. International Conference on*. [S.l.], 2003. p. 76–85.
- CHEGINI, M. et al. Interactive labelling of a multivariate dataset for supervised machine learning using linked visualisations, clustering, and active learning. *Visual Informatics*, Elsevier, 2019.
- CHEN, C.-h.; HÄRDLE, W. K.; UNWIN, A. *Handbook of data visualization*. [S.l.]: Springer Science & Business Media, 2007.
- CHEN, Y. et al. Sunburst with ordered nodes based on hierarchical clustering: a visual analyzing method for associated hierarchical pesticide residue data. *Journal of visualization*, Springer Berlin Heidelberg, Berlin/Heidelberg, v. 18, n. 2, p. 237–254, 2015. ISSN 1343-8875.
- CUI, W. Visual analytics: A comprehensive overview. *IEEE access*, IEEE, v. 7, p. 81555–81573, 2019.

DRAPER, G.; LIVNAT, Y.; RIESENFELD, R. A survey of radial methods for information visualization. *IEEE transactions on visualization and computer graphics*, IEEE, LOS ALAMITOS, v. 15, n. 5, p. 759–776, 2009. ISSN 1077-2626.

ELER, D. M. *Múltiplas visões coordenadas para exploração de mapas de similaridade*. Tese (Doutorado) — Universidade de São Paulo, 2011.

ESPADOTO, M. et al. Toward a quantitative survey of dimension reduction techniques. *IEEE transactions on visualization and computer graphics*, IEEE, United States, v. 27, n. 3, p. 2153–2173, 2021. ISSN 1077-2626.

GAO, Q. et al. Performance visualization for large-scale computing systems: A literature review. In: JACKO, J. A. (Ed.). *Human-Computer Interaction. Design and Development Approaches*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. p. 450–460. ISBN 978-3-642-21602-2.

GODINHO, P. I. A. et al. Prisma-a multidimensional information visualization tool using multiple coordinated views. In: IEEE. *2007 11th International Conference Information Visualization (IV'07)*. [S.l.], 2007. p. 23–32.

GOWER, J. C.; DIJKSTERHUIS, G. B. *Procrustes problems*. [S.l.]: OUP Oxford, 2004.

GUIMARÃES, R. V. et al. Using multiple coordinated views for multiple datasets analysis. In: IEEE. *Information Visualisation, 2009 13th International Conference*. [S.l.], 2009. p. 627–632.

HAN, J.; KAMBER, M.; PEI, J. Data mining concepts and techniques third edition. *University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University*, 2012.

HEINRICH, J.; WEISKOPF, D. State of the art of parallel coordinates. In: *Eurographics (STARs)*. [S.l.: s.n.], 2013. p. 95–116.

INSELBERG, A. The plane with parallel coordinates. *The visual computer*, Springer, v. 1, n. 2, p. 69–91, 1985.

JERN, M. et al. The gav toolkit for multiple linked views. In: IEEE. *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)*. [S.l.], 2007. p. 85–97.

JOIA, P. et al. Local affine multidimensional projection. *IEEE Transactions on Visualization and Computer Graphics*, v. 17, n. 12, p. 2563–2571, 2011.

JOLLIFFE, I. T. *Principal Component Analysis*. New York, NY: Springer, 1986. (Springer Series in Statistics). ISBN 9780387962696.

KAUER, A. L. da S. et al. An information visualization tool with multiple coordinated views for network traffic analysis. In: IEEE. *2008 12th International Conference Information Visualisation*. [S.l.], 2008. p. 151–156.

- KEIM, D. et al. Visual analytics: Definition, process, and challenges. In: *Information visualization*. Castelo de Dagstuhl, Alemanha: Springer, 2008. p. 154–175.
- KEIM, D. A. Visual techniques for exploring databases. 1997.
- KEIM, D. A. Information visualization and visual data mining. *IEEE Transactions on Visualization & Computer Graphics*, IEEE, n. 1, p. 1–8, 2002.
- KEIM, D. A.; KRIEGEL, H.-P. Visualization techniques for mining large databases: A comparison. *IEEE Transactions on Knowledge & Data Engineering*, IEEE, n. 6, p. 923–938, 1996.
- KNAFLIC, C. N. *Storytelling com Dados: um guia sobre visualização de dados para profissionais de negócio*. [S.l.]: Alta Books Editora, 2018.
- KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. *The annals of mathematical statistics*, JSTOR, v. 22, n. 1, p. 79–86, 1951.
- Langner, R.; Horak, T.; Dachselt, R. Vistiles: Coordinating and combining co-located mobile devices for visual data exploration. *IEEE Transactions on Visualization and Computer Graphics*, v. 24, n. 1, p. 626–636, Jan 2018. ISSN 1077-2626.
- MAATEN, L. Van der; HINTON, G. Visualizing data using t-sne. *Journal of machine learning research*, v. 9, n. 11, 2008.
- MACIEL, M. d. P. da S. et al. The impact of multiple coordinated views on the visual data exploration and analysis. In: IEEE. *Information Visualisation, 2008. IV'08. 12th International Conference*. [S.l.], 2008. p. 113–119.
- MARTINS, R. M. et al. Visual analysis of dimensionality reduction quality for parameterized projections. *Computers graphics*, Elsevier Ltd, OXFORD, v. 41, p. 26–42, 2014. ISSN 0097-8493.
- MATKOVIC, K. et al. Comvis: A coordinated multiple views system for prototyping new visualization technology. In: IEEE. *2008 12th International Conference Information Visualisation*. London, 2008. p. 215–220.
- MCINNES, L.; HEALY, J.; MELVILLE, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.
- MCINNES, L. et al. *UMAP*. 2018. Disponível em: <<https://github.com/lmcinnes/umap>>.
- MUNZNER, T. *Visualization analysis and design*. [S.l.]: AK Peters/CRC Press, 2014.
- NEVES, T. T. d. A. T. *Projeções multidimensionais para a análise de fluxos de dados*. Tese (Doutorado) — Universidade de São Paulo, 2016.

- NEVES, T. T. d. A. T. et al. Análise visual utilizando projeções multidimensionais. *Revista de Informática Teórica e Aplicada*, v. 22, n. 2, p. 258–288, 2015.
- NONATO, L. G.; AUPETIT, M. Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment. *IEEE transactions on visualization and computer graphics*, IEEE, United States, v. 25, n. 8, p. 2650–2673, 2019. ISSN 1077-2626.
- NORTH, C.; SHNEIDERMAN, B. A taxonomy of multiple window coordination. 1997.
- NORTH, C.; SHNEIDERMAN, B. Snap-together visualization: can users construct and operate coordinated visualizations? *International Journal of Human-Computer Studies*, Elsevier, v. 53, n. 5, p. 715–739, 2000.
- PAULOVICH, F. V. *Mapeamento de dados multi-dimensionais-integrando mineração e visualização*. Tese (Doutorado) — Universidade de São Paulo, São Carlos, São Paulo, 2008.
- PAULOVICH, F. V. et al. Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping. *IEEE Transactions on Visualization and Computer Graphics*, IEEE, v. 14, n. 3, p. 564–575, 2008.
- PAULOVICH, F. V.; OLIVEIRA, M. C. F.; MINGHIM, R. The projection explorer: A flexible tool for projection-based multidimensional visualization. In: IEEE. *XX Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2007)*. [S.l.], 2007. p. 27–36.
- PAULOVICH, F. V. et al. Pex-web: Content-based visualization of web search results. In: IEEE. *2008 12th International Conference Information Visualisation*. [S.l.], 2008. p. 208–214.
- PILLAT, R. M. Coordenação dinâmica de visualizações de dados multidimensionais. 2006.
- PILLAT, R. M.; FREITAS, C. M. Coordinating views in the infovis toolkit. In: ACM. *Proceedings of the working conference on Advanced visual interfaces*. [S.l.], 2006. p. 496–499.
- ROBERTS, J. C. Exploratory visualization with multiple linked views. In: *Exploring geovisualization*. [S.l.]: Elsevier, 2005. p. 159–180.
- ROBERTS, J. C. State of the art: Coordinated & multiple views in exploratory visualization. p. 61–71, 2007.
- SACHA, D. et al. Visual interaction with dimensionality reduction: A structured literature analysis. *IEEE transactions on visualization and computer graphics*, IEEE, United States, v. 23, n. 1, p. 241–250, 2017. ISSN 1077-2626.

SADANA, R.; STASKO, J. Designing multiple coordinated visualizations for tablets. In: WILEY ONLINE LIBRARY. *Computer Graphics Forum*. [S.l.], 2016. v. 35, n. 3, p. 261–270.

SCHERR, M. Multiple and coordinated views in information visualization. *Trends in Information Visualization*, v. 38, p. 1–8, 2008.

SHNEIDERMAN, B. *Tree visualization with tree-maps: A 2-d space-filling approach*. [S.l.], 1998.

SOUTH, L. et al. Effective use of likert scales in visualization evaluations: A systematic review. In: WILEY ONLINE LIBRARY. *Computer Graphics Forum*. [S.l.], 2022. v. 41, n. 3, p. 43–55.

STASKO, J.; ZHANG, E. Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In: *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*. [S.l.]: IEEE, 2000. p. 57–65. ISBN 0769508049. ISSN 1522-404X.

TEJADA, E.; MINGHIM, R.; NONATO, L. G. On improved projection techniques to support visual exploration of multi-dimensional data sets. *Information Visualization*, SAGE Publications Sage UK: London, England, v. 2, n. 4, p. 218–231, 2003.

THIJSEN, J. et al. Scaling up the explanation of multidimensional projections. In: THE EUROGRAPHICS ASSOCIATION. *EuroVis Workshop on Visual Analytics (EuroVA)*. [S.l.], 2023. p. 1–6.

TORGERSON, W. S. *Theory and methods of scaling*. Wiley, 1958.

VENTOCILLA, E.; RIVEIRO, M. A comparative user study of visualization techniques for cluster analysis of multidimensional data sets. *Information Visualization*, v. 19, n. 4, p. 318–338, 2020. Disponível em: <<https://doi.org/10.1177/1473871620922166>>.

WANG, Y. et al. Understanding how dimension reduction tools work: an empirical approach to deciphering t-sne, umap, trimap, and pacmap for data visualization. *The Journal of Machine Learning Research*, JMLRORG, v. 22, n. 1, p. 9129–9201, 2021.

WARD, M. O.; GRINSTEIN, G.; KEIM, D. *Interactive data visualization: foundations, techniques, and applications*. [S.l.]: AK Peters/CRC Press, 2015.

WILHELM, A. Linked views for visual exploration. In: \_\_\_\_\_. *Handbook of Data Visualization*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. p. 199–215. ISBN 978-3-540-33037-0. Disponível em: <[https://doi.org/10.1007/978-3-540-33037-0\\_9](https://doi.org/10.1007/978-3-540-33037-0_9)>.

WILLS, G. Linked data views. In: \_\_\_\_\_. *Handbook of Data Visualization*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. p. 217–241. ISBN 978-3-540-33037-0. Disponível em: <[https://doi.org/10.1007/978-3-540-33037-0\\_10](https://doi.org/10.1007/978-3-540-33037-0_10)>.

XIA, J. et al. Revisiting dimensionality reduction techniques for visual cluster analysis: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, v. 28, n. 1, p. 529–539, 2022.

YI, J. S. et al. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization & Computer Graphics*, IEEE, n. 6, 2007.



**Apêndice**

**A**

**FORMULÁRIO SOBRE CONHECIMENTO PRÉVIO  
DAS TÉCNICAS DE VISUALIZAÇÃO**

# Conhecimentos Relacionados às Técnicas de Visualizações

Este formulário tem como propósito apresentar elementos utilizados em representações visuais que apoiam o processo de exploração de dados com alta dimensionalidade. Esta pesquisa é efetuada junto ao Programa de Pós-Graduação em Ciência da Computação (PGCOMP) da Universidade Federal da Bahia – UFBA.

Meu nome é José Augusto Duarte Gomes, e esta avaliação com usuários é uma das etapas da minha pesquisa de mestrado, sob orientação do Profº Dr. Danilo Barbosa Coimbra. Para dúvidas e/ou contribuições, estou à disposição no e-mail [joseaugustoduarte@gmail.com](mailto:joseaugustoduarte@gmail.com).

As perguntas apresentadas a seguir pretendem fomentar a identificação de padrões visuais em visualizações de dados.

---

\* Indica uma pergunta obrigatória

## Termo de Consentimento Livre e Esclarecido

Prezado(a) Participante,

O(A) Senhor(a) está sendo convidado(a) a participar da pesquisa intitulada “Uma Abordagem de Visual Analytics para Análise de Erros de Projeções Multidimensionais Utilizando Múltiplas Visões Coordenadas”, desenvolvida por José Augusto Duarte Gomes, discente de Mestrado em Ciência da Computação da Universidade Federal da Bahia (UFBA), sob orientação do Professor Dr. Danilo Barbosa Coimbra.

Este estudo propõe o desenvolvimento de uma técnica de Visual Analytics implementada em uma aplicação Web. Ela utiliza diferentes visualizações justapostas na mesma tela que compartilham diversos mecanismos de interação. A ideia da proposta é que essa abordagem proporcione ao usuário explorar visualmente um conjunto de dados multidimensional e compreender a qualidade das visualizações multidimensionais apresentadas. Em específico, são utilizadas técnicas de interação, Treemaps e tabelas para melhor compreender a qualidade das técnicas de Projeções Multidimensionais, quando aplicadas a diferentes bases de dados com alta dimensionalidade.

O convite à sua participação se deve à realização de um teste de usabilidade da abordagem de Visual Analytics proposta, onde, os participantes deste estudo serão estudantes do Instituto de Computação da Universidade Federal da Bahia (UFBA), e colaboradores do Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS), que tenham alguma experiência ou conhecimento sobre análise de dados e utilização de visualizações.

Sua participação é voluntária, isto é, ela não é obrigatória, tendo plena autonomia para decidir se quer ou não participar, bem como retirar sua participação a qualquer momento. Não haverá penalidade de nenhuma maneira caso decida não consentir sua participação, ou desistir da mesma. Contudo, ela é muito importante para a execução da pesquisa.

Serão garantidas a confidencialidade e a privacidade das informações prestadas. Qualquer dado que possa identificá-lo(a) será omitido na divulgação dos resultados da pesquisa, e o material será armazenado em um acervo digital do pesquisador, prezando pela segurança da informação. A qualquer momento, durante a pesquisa, ou posteriormente, será possível solicitar ao pesquisador informações sobre sua participação e/ou sobre a pesquisa, o que poderá ser feito através dos meios de contato explicitados neste Termo.

A sua participação consistirá em atuar na avaliação de elementos de usabilidade da abordagem de Visual Analytics proposta. Os dispositivos de coleta de dados utilizados neste estudo serão: i) a observação participante, para documentar as interações, comentários, dúvidas e sugestões em todas as etapas do processo; ii) questionário semiestruturado online com questões de múltipla escolha e questões dissertativas/abertas.

O tempo estimado de duração do procedimento de avaliação da abordagem de Visual Analytics proposta será de aproximadamente 60 minutos, dividido em quatro etapas: 1) Apresentação dos objetivos da pesquisa e da abordagem de Visual Analytics proposta, com duração de dez minutos; 2) Responder questionário online sobre conhecimentos relacionados às técnicas de visualizações utilizadas, com duração de dez minutos; 3) Experimentação livre

da abordagem de Visual Analytics proposta e realização de tarefa teste, com duração de dez minutos. 4) Realização de tarefas de exploração, baseadas em heurística para avaliar a interação e os aspectos visuais das técnicas de visualização. Essas tarefas envolvem comparação, identificação, localização, manipulação de mecanismos de interação como filtro, seleção, zoom, pan, seleção múltipla etc. Após a conclusão de cada tarefa o participante irá registrar suas percepções de usabilidade em um questionário online. A etapa quatro terá duração de 30 minutos.

Os dados coletados tanto na Observação Participante quanto nos Questionários serão armazenados em arquivos digitais, mas somente terão acesso aos mesmos o pesquisador e seu orientador. Ao final da pesquisa, todo material será mantido em arquivo digital, por pelo menos 5 anos, conforme Resolução 466/12.

Os principais benefícios relacionados à sua participação nesta pesquisa são: contribuir para a construção de representações visuais mais dinâmicas e intuitivas, por meio de técnicas de visualização que utilizam diversos mecanismos de interação; apoiar o processo de análise exploratória de uma ampla variedade de dados complexos e com alta dimensionalidade; e avançar no estado da arte do tema relacionado a métricas de qualidade de projeções multidimensionais, por meio de uma melhor compreensão da qualidade das técnicas de Projeções Multidimensionais.

Toda pesquisa possui riscos e sua participação pode envolver algum desconforto, ainda que pequeno, relacionado ao tempo dedicado ao preenchimento dos questionários. Será feito todo esforço possível para minimizar eventuais desconfortos e não ocupar o tempo do participante desnecessariamente. Em ocorrências mais graves, referentes a estes riscos de origem intelectual, psíquica ou moral, a pessoa poderá ser encaminhada para atendimento psicológico. Recomenda-se o Serviço Médico Universitário da UFBA, localizado na Rua Padre Feijó, 240, 4o andar do Ambulatório Professor Magalhães Neto. Além disso, caso haja comprovação de despesas decorrentes de sua participação nessa pesquisa, você terá o direito ao ressarcimento, bem como, poderá ser indenizado por danos decorrentes a sua participação nessa pesquisa, nos termos da Lei.

Caso você tenha alguma dúvida, necessite de qualquer esclarecimento ou ainda desejar retirar-se da pesquisa, por favor, entre em contato com o pesquisador a qualquer tempo. Os resultados da pesquisa estarão à sua disposição quando finalizada. Seu nome ou o material que indique sua participação não será liberado sem a sua permissão. Este termo será redigido em duas vias, sendo que uma será arquivada digitalmente pelo pesquisador responsável, e a outra será fornecida ao participante da pesquisa.

Informações de contato dos responsáveis pelo acompanhamento da pesquisa:

- Pesquisador Responsável: José Augusto Duarte Gomes. Rua Mestre Pastinha, 300, Federação, Salvador, Bahia. Cel. (71) 9 9315-4149. E-mail: joseaugustoduarte@gmail.com.
- Comitê de Ética em Pesquisa (CEP-IGM): Centro de Pesquisas Gonçalo Moniz, FIOCRUZ-BA. Rua Waldemar Falcão, 121, Candeal de Brotas, Salvador, Bahia. Tel.: (71) 3176-2285 / E-mail: cep.igm@fiocruz.br.
- Comitê de Ética em Pesquisa da Escola de Enfermagem da Universidade Federal da Bahia (UFBA). Rua Augusto Viana, s/n, 4º andar, sala 432-437, Canela, Salvador, Bahia, CEP: 40110-060. Telefone: (71) 3283-7615. E-mail: cepee.ufba@ufba.br.

1. **De posse das informações do Termo de Consentimento Livre e Esclarecido, você concorda em colaborar com esta pesquisa?** \*

*Marcar apenas uma oval.*

- Concordo
- Não Concordo *Pular para a seção 5 (Agradecimentos.)*

### **Perfil do Usuário**

Na primeira parte deste questionário, serão feitas algumas perguntas sobre idade, educação e seu nível de familiaridade com Visualização de Dados.

2. **Qual sua faixa etária? \***

*Marcar apenas uma oval.*

- Até 17 anos
- 18 a 29 anos
- 30 a 39 anos
- 40 a 50 anos
- 51 anos ou mais
- Prefiro não informar

3. **Qual seu nível completo de formação? \***

*Marcar apenas uma oval.*

- Pouca ou nenhuma educação formal
- Ensino Fundamental
- Ensino Médio
- Graduação
- Especialização
- Mestrado
- Doutorado
- Prefiro não informar

#### 4. Qual seu nível de familiaridade com Visualização de Dados? \*

Marcar apenas uma oval.

Nenhuma Familiaridade

1

2

3

4

5

Muita Familiaridade

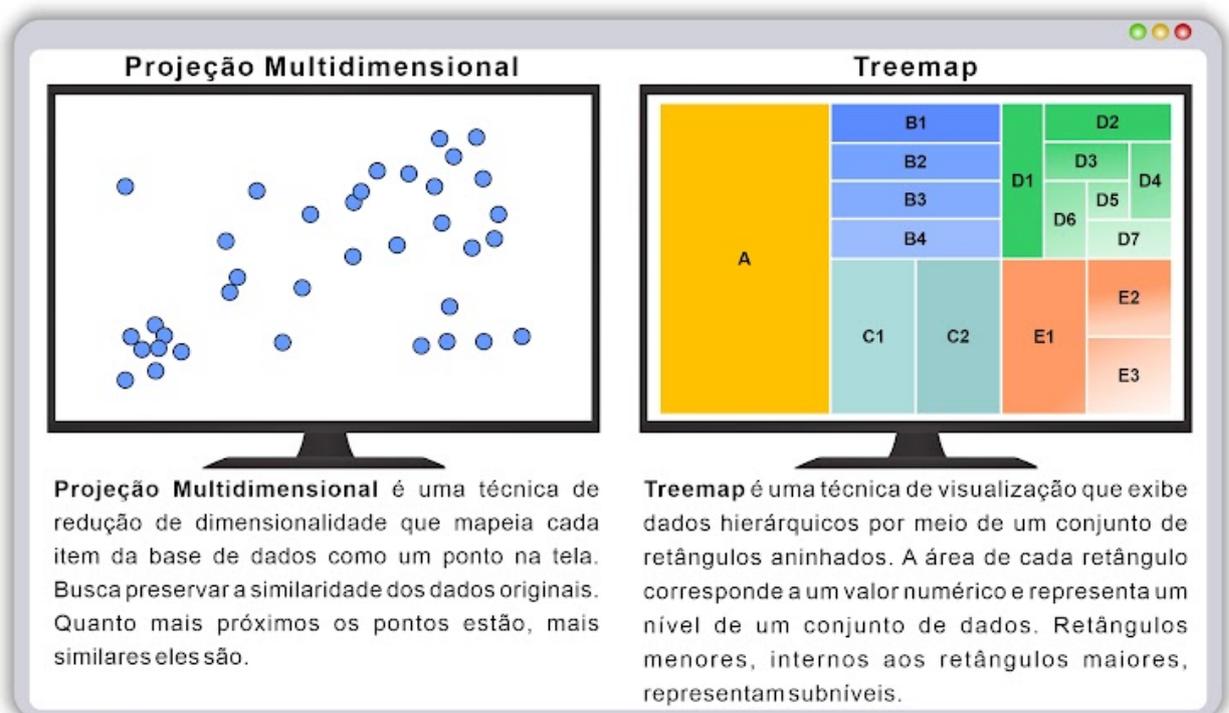


Figura 01 - Técnicas utilizadas em Visualização de Dados

**5. Qual seu nível de familiaridade com a técnica de Projeção Multidimensional? \***

Marcar apenas uma oval.

Nenhuma Familiaridade

---

1

---

2

---

3

---

4

---

5

---

Muita Familiaridade

---

**6. Qual seu nível de familiaridade com a técnica de Treemap? \***

Marcar apenas uma oval.

Nenhuma Familiaridade

---

1

---

2

---

3

---

4

---

5

---

Muita Familiaridade

---

## Compreensão de Padrões Visuais

Nesta seção, serão apresentados os atributos visuais das técnicas de Projeção Multidimensional e Treemap.

Ao utilizar representações gráficas é importante saber que os dados são mapeados para diferentes atributos visuais que podem auxiliar a percepção humana. Dentre os diferentes elementos visuais que podem codificar informações, a Figura 02 destaca: posição, forma, tamanho, brilho e cor. Esses atributos podem ser combinados e ajustados para converter dados em formatos visuais com o intuito de maximizar a eficácia de uma visualização.

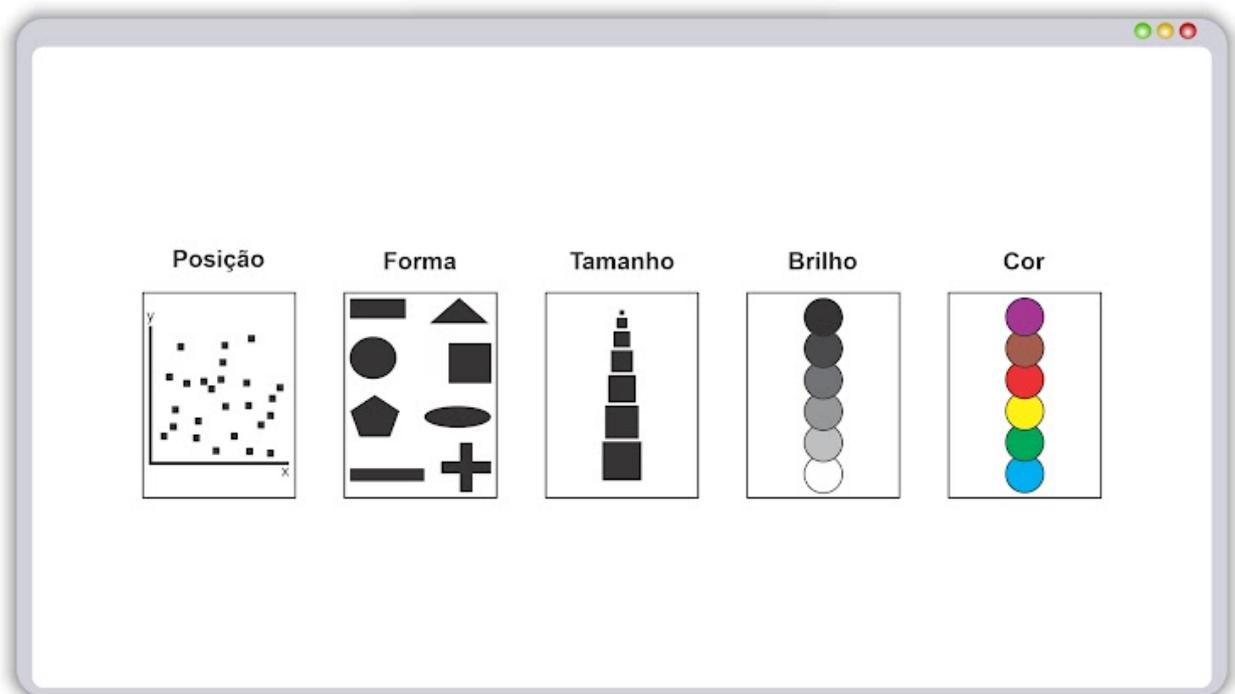


Figura 02 - Atributos visuais que codificam informações em representações gráficas

A parte superior da Figura 03 mostra a visão de uma base de dados em formato de tabela, contendo 'n' itens, onde, cada item corresponde a uma linha. Na parte inferior da imagem tem uma projeção multidimensional, e nela, cada item (linha) da base de dados é representado por um ponto.

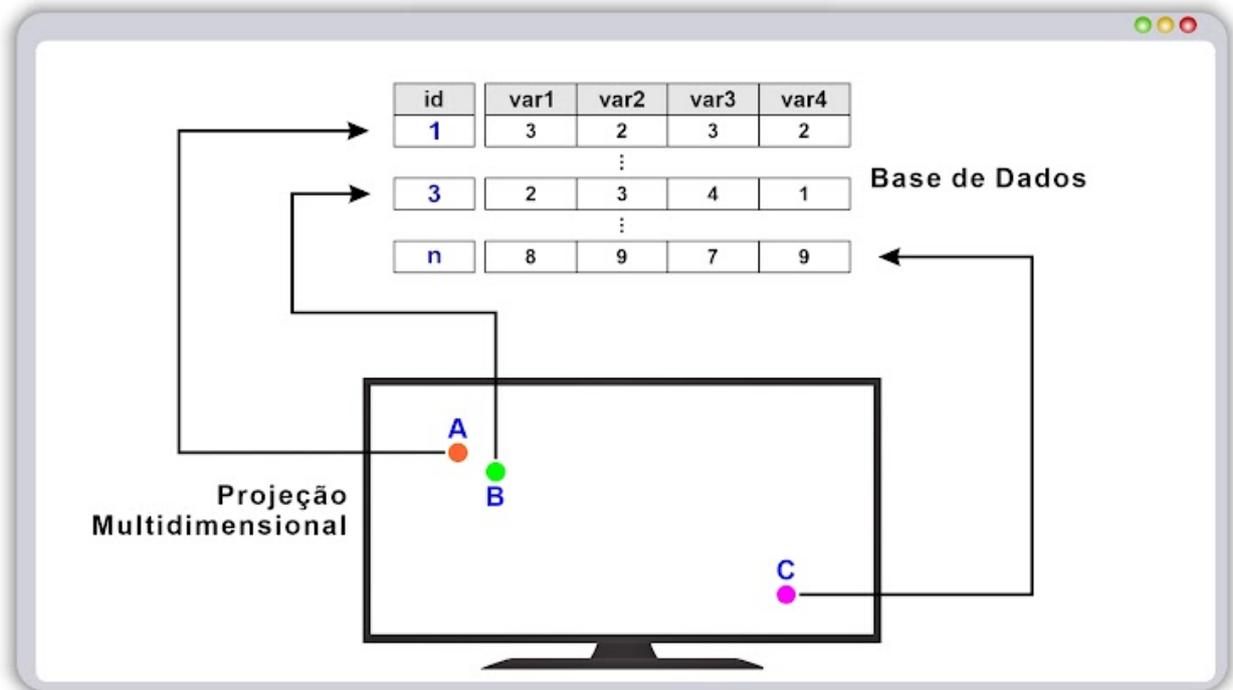


Figura 03 - Projeção Multidimensional de uma base de dados

Com base na Figura 03, é possível observar que os valores das linhas '1' e '3' são muito semelhantes. Logo, seus pontos correspondentes foram mapeados bem próximos na tela, para representar essa similaridade. A terceira linha da tabela possui valores bem diferentes das anteriores, e por essa razão, o ponto correspondente é projetado mais distante dos demais.

Na Projeção Multidimensional, de forma geral, quanto mais semelhantes os valores são na base de dados (espaço original), mais próximos eles são mapeados na tela (espaço projetado).

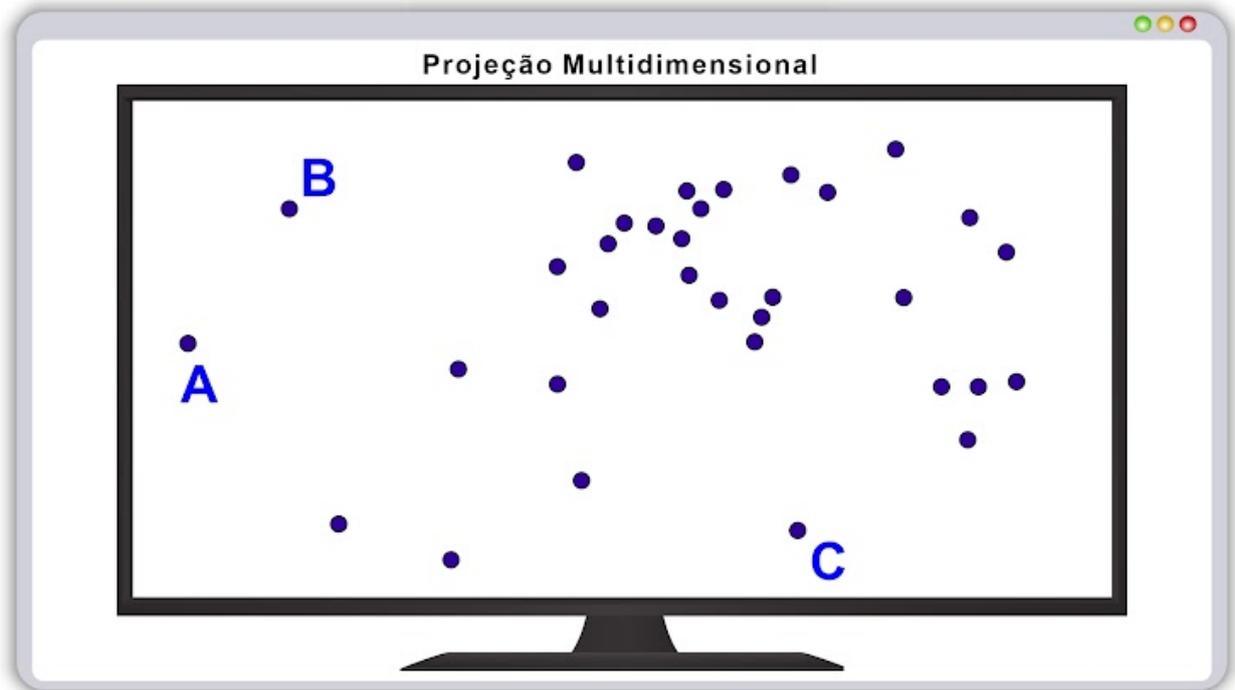


Figura 04 - Similaridade dos dados

7. Com base nas informações anteriores, qual par de pontos demonstra ter mais similaridade na Figura 04? \*

Marcar apenas uma oval.

- par 'A-B'
- par 'B-C'
- par 'C-A'
- Os pares de pontos 'A-B' e 'A-C' possuem o mesmo grau de similaridade

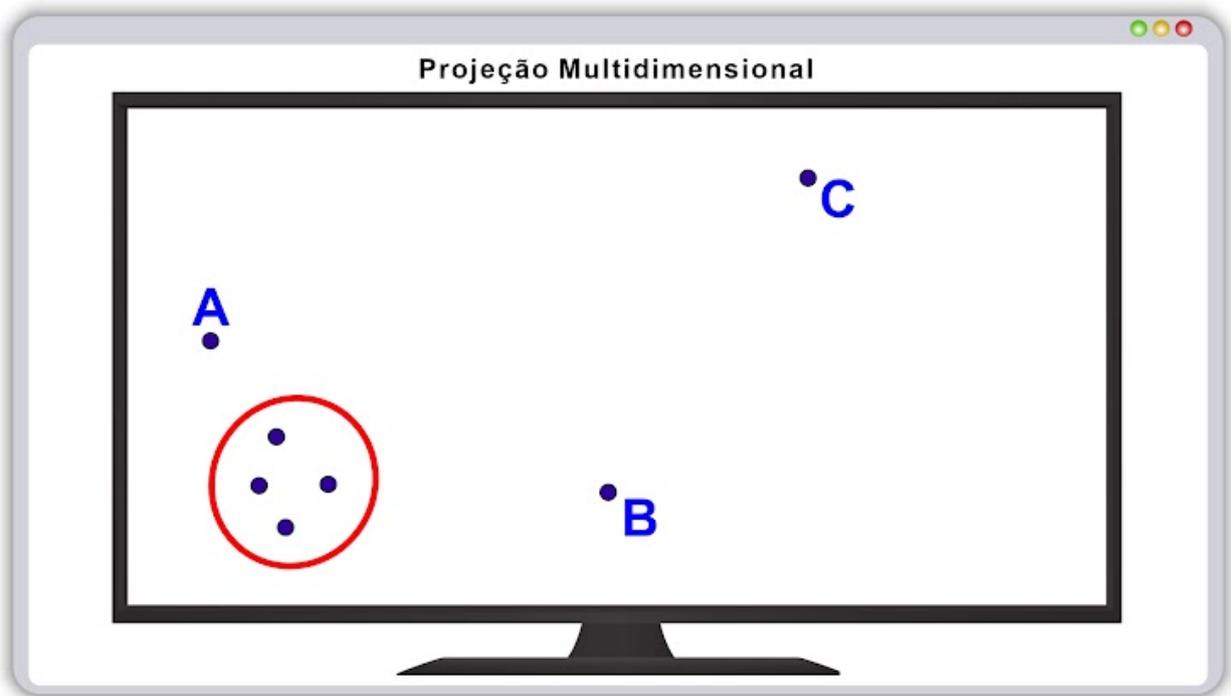


Figura 05 - Similaridade na projeção dos pontos

8. A Figura 05 mostra um conjunto de pontos, sendo quatro pontos circutados, e três pontos rotulados com as letras A, B e C. Qual dos pontos rotulados possui mais similaridade com o grupo de pontos circutados? \*

Marcar apenas uma oval.

- ponto A
- ponto B
- ponto C
- pontos B e C

Outro aspecto importante a ser considerado sobre as Projeções Multidimensionais é a sobreposição dos dados no espaço projetado. Quando há alta similaridade nos dados originais pode ocorrer do método de projeção mapeá-los para pontos muito próximos ou sobrepostos no espaço projetado.

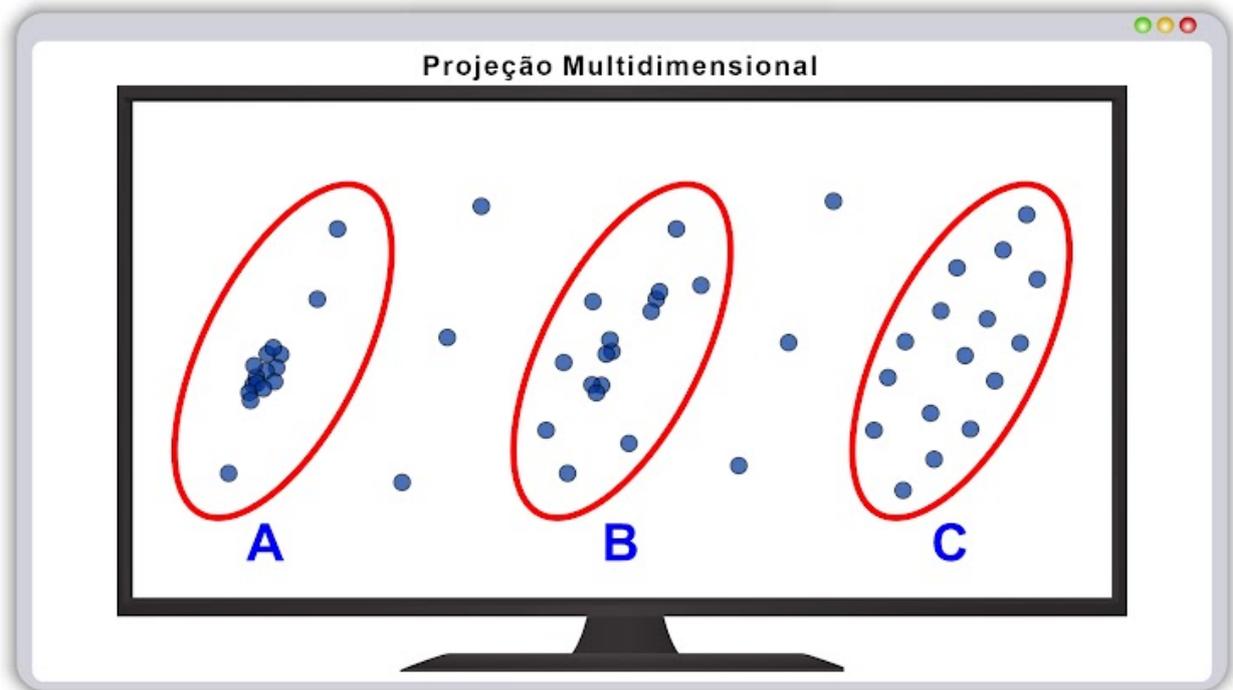


Figura 06 - Similaridade dos dados e sobreposição de pontos

9. Qual dos conjunto de pontos circulado na Figura 06 contém a maior concentração de pontos sobrepostos? \*

Marcar apenas uma oval.

- Conjunto A
- Conjunto B
- Conjunto C
- Conjuntos C e B

### Treemap

Nesta seção, você vai analisar o mapeamento de valores no tamanho dos retângulos do Treemap.

A Figura 07 retrata um Diagrama de Árvore e um Treemap representando um exemplo de hierarquia de dados, juntamente com suas respectivas explicações.

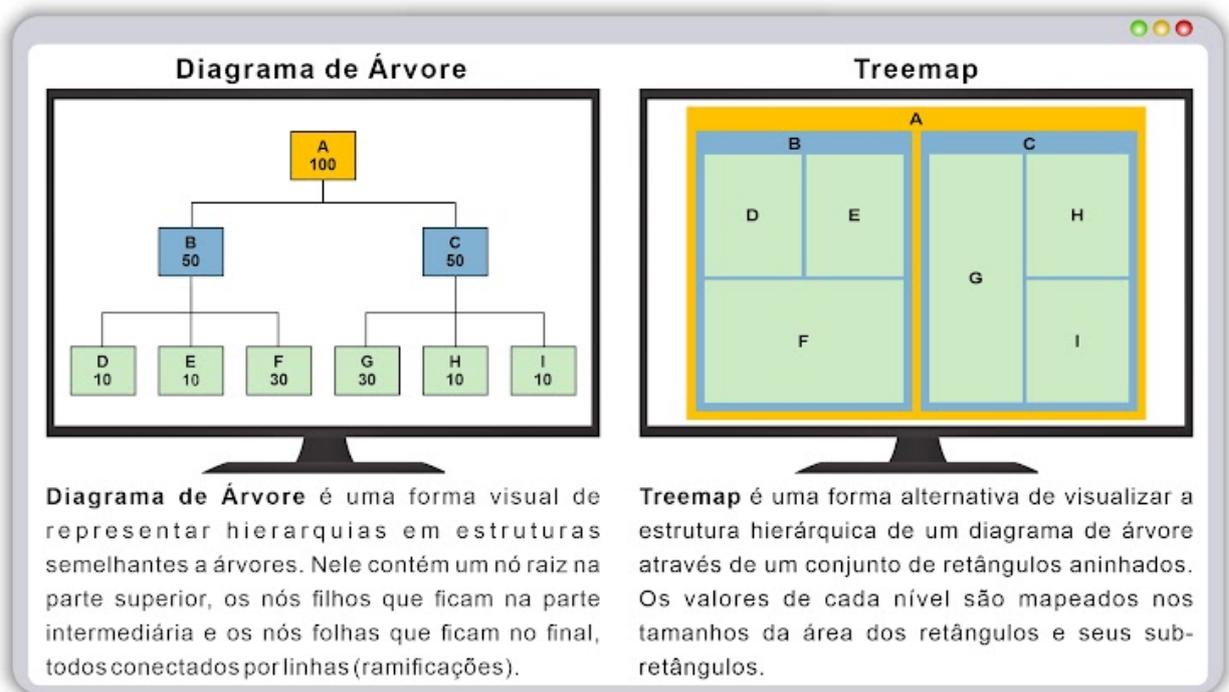


Figura 07 - Representação de valores no Treemap

O Treemap da Figura 08 está mapeando no tamanho de seus retângulos os valores de despesas mensais pagas em serviços como alimentação, transporte, água, energia e internet.

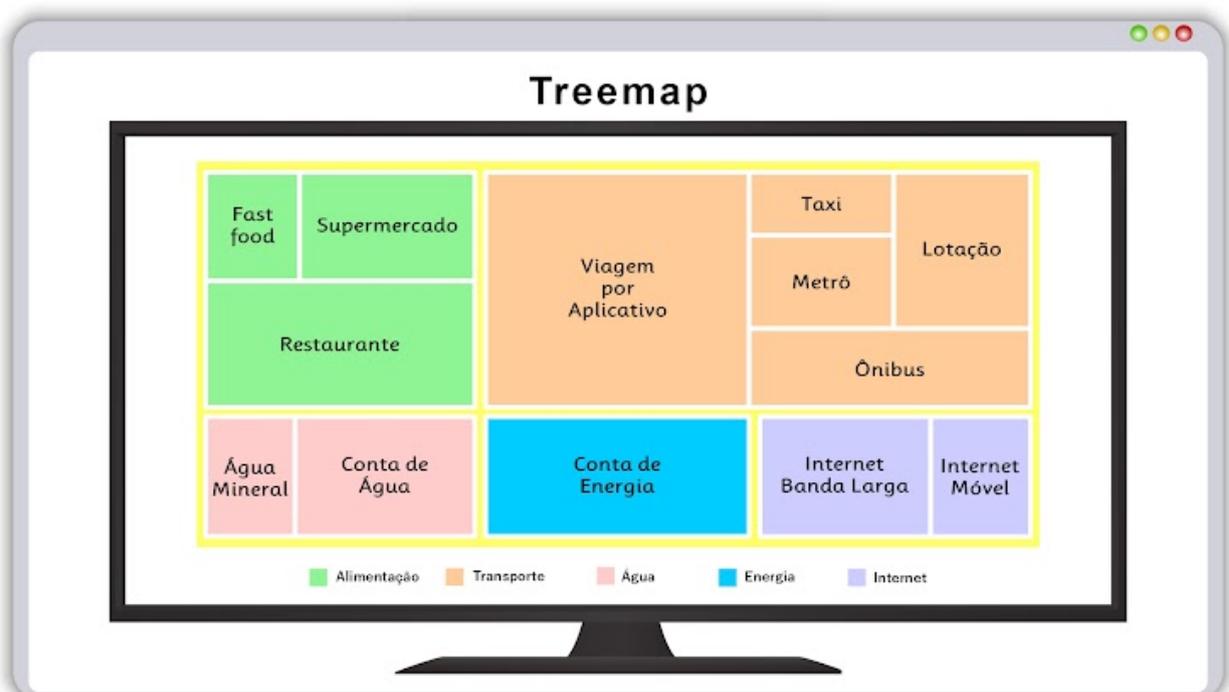


Figura 08 - Representação de despesas mensais com Treemap

## 10. Qual o tipo de alimentação gerou o maior volume de despesas? \*

Marcar apenas uma oval.

- Restaurante
- Supermercado
- Fast Food
- Taxi

O Treemap da Figura 09 está mapeando no tamanho de seus retângulos os valores relativos de despesas mensais pagas em serviços como alimentação, transporte, água, energia e internet.

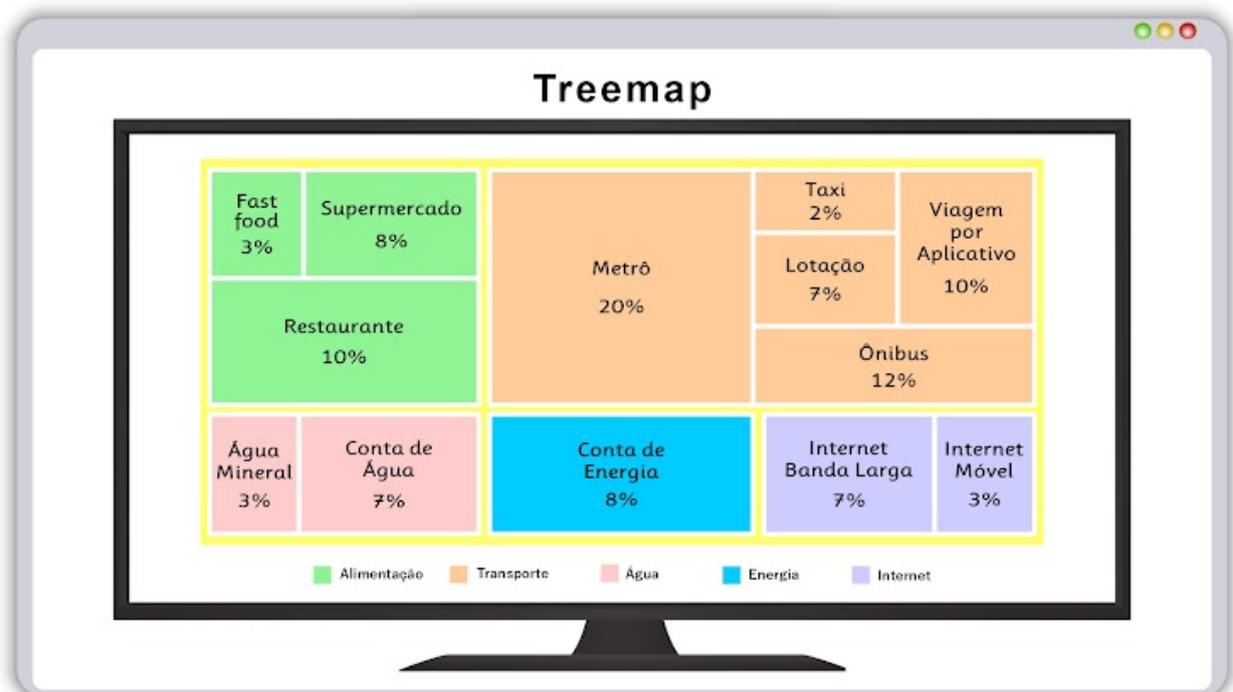


Figura 09 - Representação de valores relativos de despesas mensais com Treemap

## 11. Considerando a Figura 09, em relação as despesas com Transporte, qual a porcentagem gasta com Ônibus e Viagem por Aplicativo? \*

Marcar apenas uma oval.

- 32%
- 12%
- 22%
- Nenhuma das respostas anteriores

No Treemap da Figura 10 o atributo cor dos retângulos está mapeando os valores de erro de projeção de uma base de dados. A legenda logo abaixo do Treemap indica que quanto maior o valor, mais escuro fica o tom da cor.

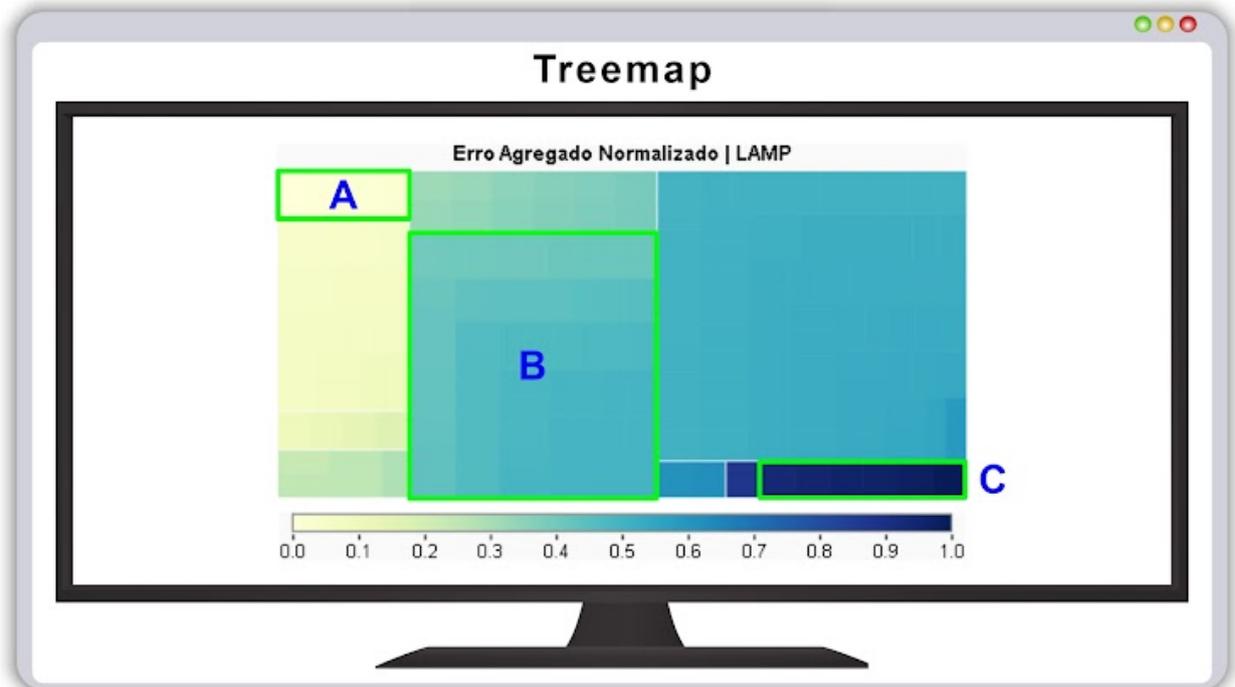


Figura 10 - Faixas de erros de projeção mapeadas no atributo tamanho

12. Considerando a escala de cor utilizada no Treemap da Figura 10, o retângulo 'B', com maior área, corresponde a qual intervalo de escala de cor ? \*

*Marcar apenas uma oval.*

- Intervalo 0.9 a 1.0
- Intervalo 0.0 a 0.1
- Intervalo 0.4 a 0.5
- Nenhum dos intervalos acima

Agradecimentos.

Obrigado por colaborar.

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários



**Apêndice**

**B**

**FORMULÁRIO COM TAREFAS E QUESTÕES DE  
USABILIDADE**

# Avaliação da Abordagem de Visual Analytics Proposta

Este formulário está vinculado a pesquisa de mestrado do aluno José Augusto do PGComp-UFBA e segue as mesmas normas descritas no formulário anterior. O objetivo desse formulário é registrar as respostas do uso da abordagem de Visual Analytics desenvolvida, por meio da realização de tarefas de exploração, baseadas em **heurística** para avaliar a interação, percepção e aspectos visuais da abordagem. Essas tarefas envolvem encontrar padrões visuais e utilizar mecanismos de interação para extração de informação.

\* Indica uma pergunta obrigatória

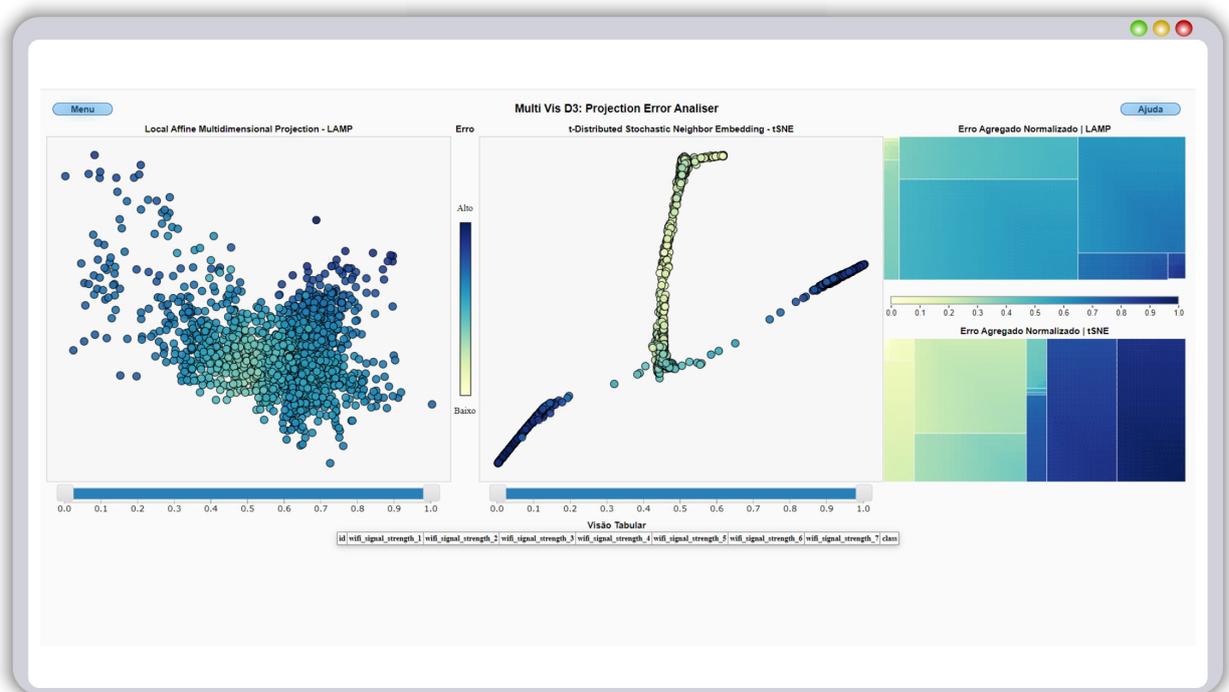


Figura 1 - Abordagem de Visual Analytics Proposta

## TAREFA 01

Avaliar a perspectiva de similaridade visual de um par de pontos nas Projeções Multidimensionais.

**Antes de iniciar essa Tarefa realize os seguintes ajustes:**

- Resetar a aplicação para as configurações padrão;
- Selecionar a base de dados 'Segmentation';
- Selecionar a visão 'Erro Agregado'.

Conforme mostra a Figura 2, selecione os pontos A (id: 916) e B (id: 1944) na Projeção Multidimensional LAMP.

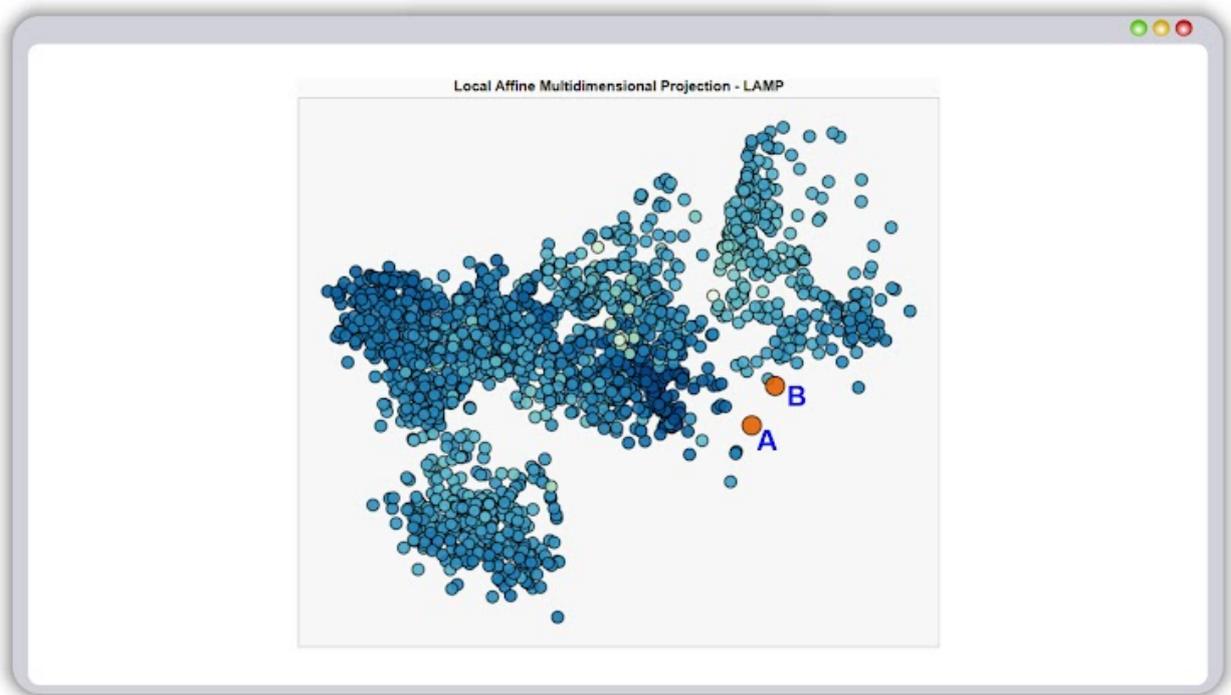


Figura 2 - Seleção de pontos na Projeção Multidimensional LAMP

1. Considerando os dois pontos selecionados, marque a afirmativa correta: \*

*Marcar apenas uma oval.*

- Há uma menor similaridade dos pontos A e B na Projeção Multidimensional LAMP e uma maior similaridade na Projeção Multidimensional tSNE.
- A similaridade dos pontos A e B na Projeção Multidimensional LAMP se mantém na Projeção Multidimensional tSNE.
- Há uma maior similaridade dos pontos A e B na Projeção Multidimensional LAMP e uma menor similaridade na Projeção Multidimensional tSNE.
- A similaridade dos pontos A e B na Projeção Multidimensional tSNE é maior que na Projeção Multidimensional LAMP.

## 2. Qual o nível de complexidade dessa tarefa? \*

Marcar apenas uma oval.

Muito baixo

1

2

3

4

5

Muito alto

### TAREFA 02

Avaliar a similaridade visual de um grupo de pontos nas nas Projeções Multidimensionais.

**Antes de iniciar essa Tarefa realize os seguintes ajustes:**

- Resetar a aplicação para as configurações padrão;
- Selecionar a base de dados 'Segmentation';
- Selecionar a visão 'Erro Agregado'.

Considere a Figura 3 e selecione o conjunto de pontos A na Projeção Multidimensional tSNE.

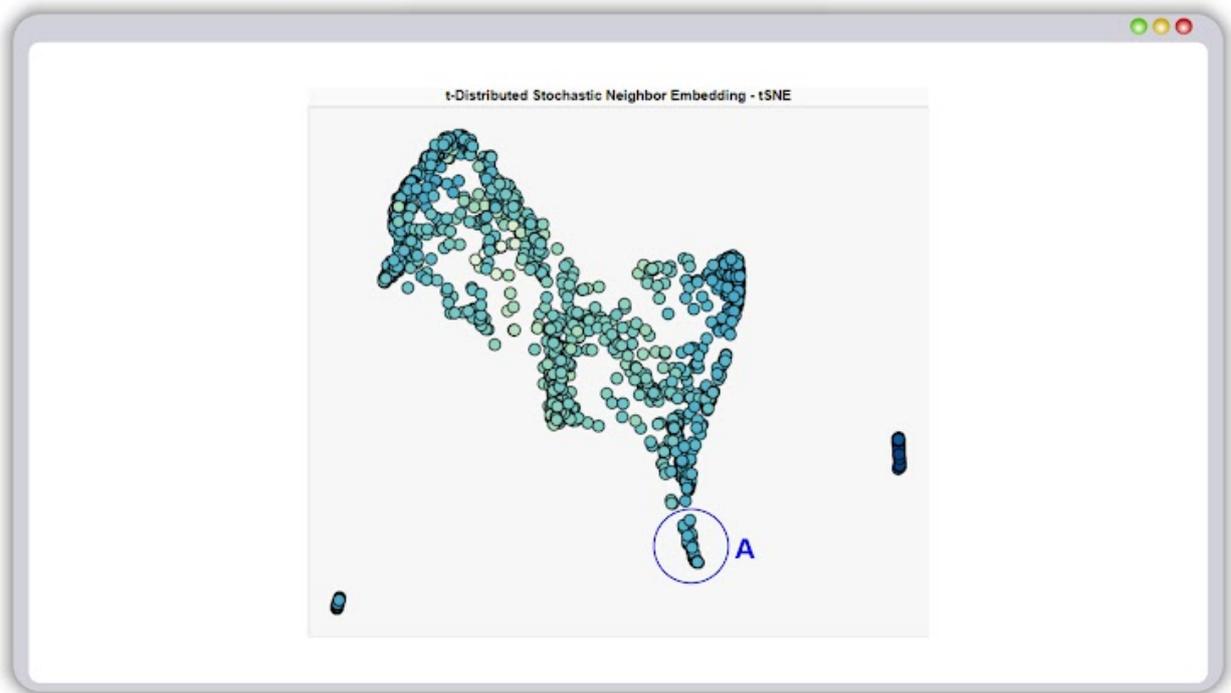


Figura 3 - Seleção de um conjunto de pontos na Projeção Multidimensional tSNE

3. Compare o padrão visual nas duas projeções e indique em qual técnica de Projeção Multidimensional existe uma maior similaridade entre os pontos selecionados. \*

Marcar apenas uma oval.

- LAMP
- tSNE
- Tanto a LAMP quanto a tSNE parecem possuir uma similaridade visual muito semelhante
- Nenhuma das técnicas

## 4. Qual o nível de complexidade dessa tarefa? \*

Marcar apenas uma oval.

Muito baixo

1

2

3

4

5

Muito alto

## TAREFA 03

Verificar em cada PM a quantidade de instâncias com erro agregado para um determinado intervalo de valores normalizados.

**Antes de iniciar essa Tarefa realize os seguintes ajustes:**

- Resetar a aplicação para as configurações padrão;
- Selecionar a base de dados 'Segmentation';
- Selecionar a visão 'Erro Agregado'.

5. Localize as instâncias do intervalo de erro agregado de 0.7 a 0.8 e indique qual das técnicas tem mais pontos/instâncias nesse intervalo. \*

*Marcar apenas uma oval.*

- LAMP
- tSNE
- Tanto a LAMP quanto a tSNE parecem possuir a mesma quantidade de pontos/instâncias
- Nenhuma das técnicas

6. Qual o nível de complexidade dessa tarefa? \*

*Marcar apenas uma oval.*

Muito baixo

1

2

3

4

5

Muito alto

#### TAREFA 04

Analisar a qualidade da projeção de instâncias através da métrica de erro agregado.

**Antes de iniciar essa Tarefa realize os seguintes ajustes:**

- Resetar a aplicação para as configurações padrão;
- Selecionar a base de dados 'Segmentation';
- Selecionar a visão 'Erro Agregado'.

**Considere a Figura 4 e selecione o conjunto de pontos B na projeção multidimensional tSNE.**



Figura 4 - Seleção de um conjunto de pontos na Projeção Multidimensional tSNE

7. **Qual a alternativa melhor representa o resultado da seleção do conjunto de pontos B?** \*

*Marcar apenas uma oval.*

- Os erros destes pontos são mais altos na tSNE.
- Os erros destes pontos são mais altos na LAMP.
- Os erros destes pontos contém valores altos em ambas as projeções.
- Os erros destes pontos não contém valores altos em nenhuma das projeções.

## 8. Qual o nível de complexidade dessa tarefa? \*

Marcar apenas uma ova.

Muito baixo

1

2

3

4

5

Muito alto

### TAREFA 05

Verificar a qualidade da PM por meio de intervalos normalizados dos valores da métrica de erro agregado.

**Antes de iniciar essa Tarefa realize os seguintes ajustes:**

- Resetar a aplicação para as configurações padrão;
- Selecionar a base de dados 'Segmentation';
- Selecionar a visão 'Erro Agregado'.

**Filtrar o intervalo de erro agregado normalizado de 0.1 a 0.2 nas duas projeções, LAMP e tSNE, conforme mostra a Figura 5.**

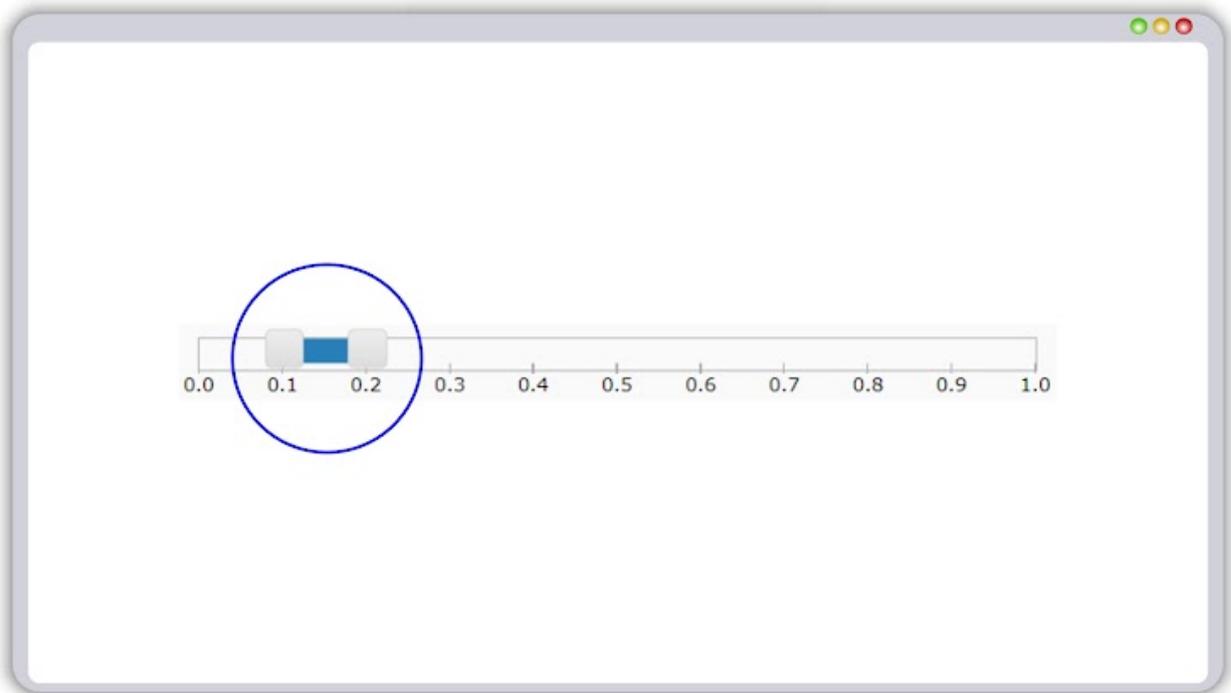


Figura 5 - Filtro de intervalo de erro utilizando o recurso interativo 'Slider'

9. **Qual projeção tem uma maior quantidade de pontos nesse intervalo contendo erros mais baixos? \***

*Marcar apenas uma oval.*

- LAMP
- tSNE
- Parecem ter a mesma quantidade de pontos na LAMP e na tSNE
- Nenhuma das técnicas

10. Qual o nível de complexidade dessa tarefa? \*

*Marcar apenas uma oval.*

Muito baixo

1

2

3

4

5

Muito alto

### TAREFA 06

Avaliar a qualidade das PMs considerando o atributo rótulo da base de dados.

**Antes de iniciar essa Tarefa realize os seguintes ajustes:**

- Resetar a aplicação para as configurações padrão;
- Selecionar a base de dados 'Segmentation';
- Selecionar a visão 'Classe'.

11. **Avalie em qual técnica de projeção multidimensional o conjunto de pontos pertencente a classe 0 (zero) apresenta valores mais altos de erro agregado?** \*

*Marcar apenas uma oval.*

- LAMP
- tSNE
- Tanto LAMP quanto tSNE apresentam valores altos de erro agregado para esse intervalo
- Nenhuma das técnicas

12. **Qual o nível de complexidade dessa tarefa? \***

*Marcar apenas uma oval.*

Muito baixo

1

2

3

4

5

Muito alto

## TAREFA 07

Analisar a qualidade da PM nas instâncias agrupadas na visão de Classe utilizando coordenação.

**Antes de iniciar essa Tarefa realize os seguintes ajustes:**

- Resetar a aplicação para as configurações padrão;
- Selecionar a base de dados 'Wifi';
- Selecionar a visão 'Classe'.

13. **Em qual técnica de projeção multidimensional o conjunto de pontos pertencente a classe 4 apresentam valores altos de erro agregado no intervalo de 0.9 a 1.0?** \*

*Marcar apenas uma oval.*

- LAMP
- tSNE
- Tanto LAMP quanto tSNE apresentam valores altos de erro agregado para esse intervalo
- Nenhuma das técnicas

14. **Qual o nível de complexidade dessa tarefa?** \*

*Marcar apenas uma oval.*

- Muito baixo
- 1
- 2
- 3
- 4
- 5
- Muito alto

## TAREFA 08

Analisar a qualidade das Projeções Multidimensionais na perspectiva de preservação de vizinhança.

**Antes de iniciar essa Tarefa realize os seguintes ajustes:**

- Resetar a aplicação para as configurações padrão;
- Selecionar a base de dados 'Segmentation';
- Selecionar a visão 'Preservação de Distância'.
- Alterar a opção 'cor do realce ao passar o mouse' para verde

**Considere a Figura 7 e analise a vizinhança dos pontos A, B e C na projeção multidimensional LAMP.**

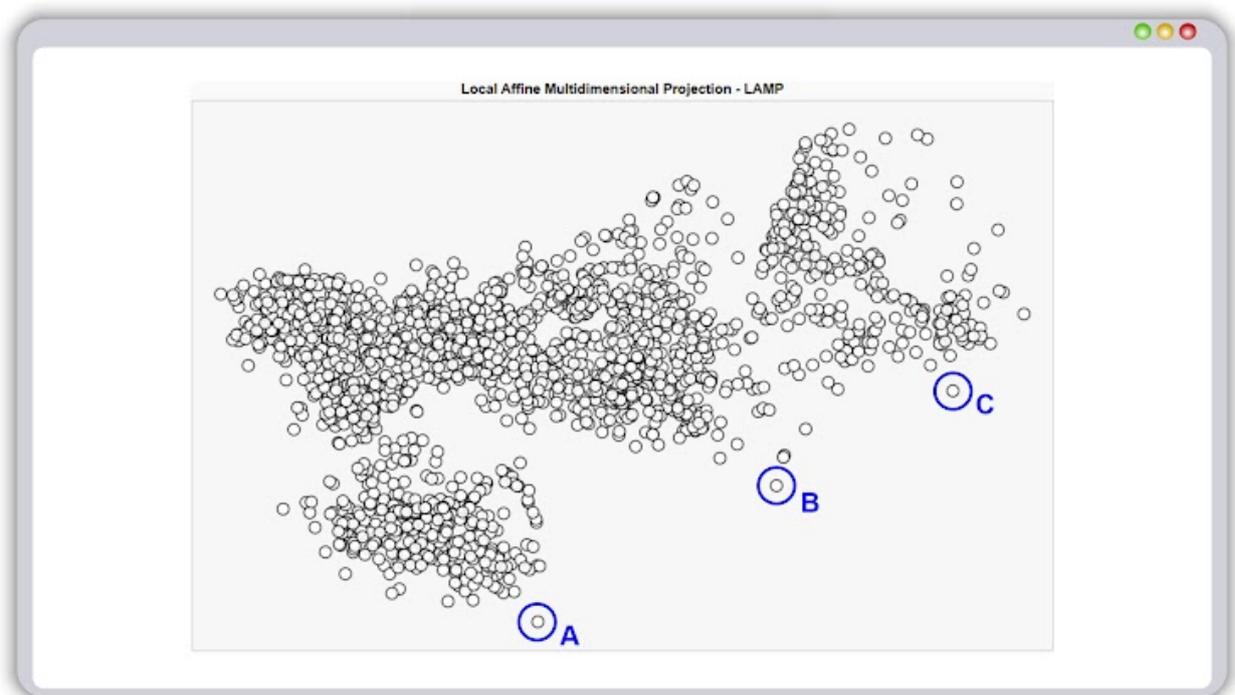


Figura 7 - Preservação de Vizinhança na Projeção Multidimensional LAMP

## 15. Qual dos pontos possuem somente vizinhos Falsos na projeção LAMP? \*

Marcar apenas uma oval.

- Ponto 'A'
- Ponto 'B'
- Ponto 'C'
- Nenhum dos pontos

## 16. Qual o nível de complexidade dessa tarefa? \*

Marcar apenas uma oval.

Muito baixo

1

2

3

4

5

Muito alto

## TAREFA 09

Investigar a preservação de vizinhança de instâncias nas Projeções Multidimensionais.

**Antes de iniciar essa Tarefa realize os seguintes ajustes:**

- Resetar a aplicação para as configurações padrão;
- Selecionar a base de dados 'Wifi';
- Selecionar a visão 'Preservação de Distância'.
- Alterar a opção 'cor do realce ao passar o mouse' para verde

Considere a Figura 6 e posicione o ponteiro do mouse sobre o ponto circulado (id: 1204) na projeção multidimensional tSNE.

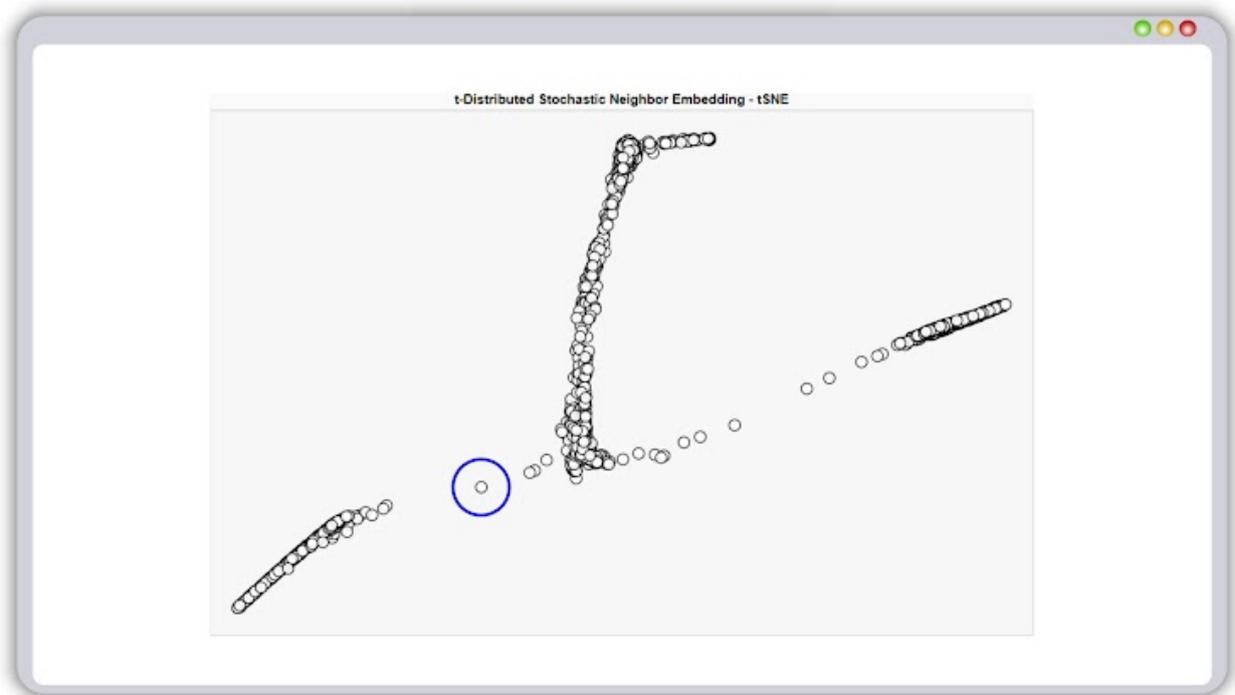


Figura 6 - Preservação de Vizinhaça na Projeção Multidimensional LAMP

17. Compare a vizinhaça dos pontos nas projeções e marque qual a alternativa correta. \*

Marcar apenas uma oval.

- Na técnica LAMP o ponto possui mais vizinhos ausentes e na tSNE mais vizinhos falsos.
- Na técnica tSNE o ponto possui mais vizinhos falsos e na LAMP menos vizinhos ausentes.
- Na técnica LAMP o ponto possui mais vizinhos falsos e na tSNE mais vizinhos ausentes.
- Ambas as técnicas possuem a mesma quantidade de vizinhos ausentes e falsos.

## 18. Qual o nível de complexidade dessa tarefa? \*

*Marcar apenas uma oval.*

Muito baixo

1

2

3

4

5

Muito alto

### Avaliação de Usabilidade

Após ter realizado as tarefas registre aqui suas percepções relacionadas a usabilidade.

19. Você concorda que a ferramenta foi útil para avaliar aspectos de qualidade das Projeções Multidimensionais? \*

*Marcar apenas uma oval.*

Discordo totalmente

1

2

3

4

5

Concordo totalmente

20. Você concorda que os diferentes mecanismos de interação disponíveis na ferramenta auxiliaram na exploração dos dados e compreensão da informação? \*

*Marcar apenas uma oval.*

Discordo totalmente

1

2

3

4

5

Concordo totalmente

21. Você concorda que utilizar diferentes visualizações na mesma tela com interações coordenadas entre elas colaboraram para melhor analisar os erros nas projeções? \*

*Marcar apenas uma oval.*

Discordo totalmente

1

2

3

4

5

Concordo totalmente

22. Qual o nível de complexidade para manusear a ferramenta na realização das tarefas? \*

*Marcar apenas uma oval.*

Muito Baixo

1

2

3

4

5

Muito Alto

23. Você concorda que a visualização Treemap auxilia na identificação dos intervalos de erros agregados dos pontos? \*

*Marcar apenas uma oval.*

Discordo totalmente

1

2

3

4

5

Concordo totalmente

24. Você concorda que mapear os erros agregados em duas projeções justapostas (lado a lado) auxilia a compreender e comparar as suas respectivas qualidades. \*

*Marcar apenas uma oval.*

Discordo totalmente

1

2

3

4

5

Concordo totalmente

25. Você concorda que a visão de classe contribui para avaliar a qualidade das projeções pela perspectiva dos erros em cada classe da base de dados? \*

*Marcar apenas uma oval.*

Discordo totalmente

1

2

3

4

5

Concordo totalmente

26. Você concorda que a visão de 'preservação de vizinhança' contribui para entender como estão dispostos os vizinhos das instâncias no espaço original e no espaço projetado? \*

Marcar apenas uma oval.

Discordo totalmente

1

2

3

4

5

Concordo totalmente

27. Você gostaria de ser informado caso os resultados da pesquisa sejam publicados em um artigo? \*

Marcar apenas uma oval.

Sim *Pular para a pergunta 28*

Não *Pular para a seção 13 (Agradecimentos)*

Informe seu endereço de e-mail

28. Informe um endereço de e-mail válido para receber os dados da publicação. \*

\_\_\_\_\_

Agradecimentos

Obrigado pela sua participação e contribuição!



**Apêndice**

**C**

**APROVAÇÃO DO CONSELHO DE ÉTICA EM  
PESQUISA DA UFBA**

**PARECER CONSUBSTANCIADO DO CEP**

**DADOS DO PROJETO DE PESQUISA**

**Título da Pesquisa:** Uma Abordagem de Visual Analytics para Análise de Erros de Projeções Multidimensionais Utilizando Múltiplas Visões Coordenadas

**Pesquisador:** JOSE AUGUSTO DUARTE GOMES

**Área Temática:**

**Versão:** 2

**CAAE:** 62907022.8.0000.5531

**Instituição Proponente:** Universidade Federal da Bahia - UFBA

**Patrocinador Principal:** Financiamento Próprio

**DADOS DO PARECER**

**Número do Parecer:** 5.757.369

**Apresentação do Projeto:**

Conforme descrito no Formulário de Informações Básicas da Plataforma Brasil:

Trata-se de um protocolo de pesquisa que abordará "O crescimento massivo e diversificado dos dados na era da big data, deixa cada vez mais em evidência o desafio de analisar dados com alta dimensionalidade. Nesse sentido, diferentes técnicas de projeção multidimensional foram desenvolvidas para gerar visualizações que contribuem para a análise exploratória de uma ampla variedade de dados complexos e com alta dimensão. O mapeamento realizado na projeção varia de acordo com a técnica aplicada. Esta pesquisa tem como propósito desenvolver uma abordagem de Visual Analytics para interpretação de qualidade de projeções multidimensionais baseada em múltiplas visualizações coordenadas. Nesse processo, a utilização de diferentes representações visuais na mesma tela e compartilhando mecanismos de interação pode tornar as tarefas de exploração e interpretação mais intuitivas. O percurso metodológico deste estudo se constituiu a partir de uma pesquisa aplicada e exploratória, na qual pretende realizar uma pesquisa de campo, onde ocorrerá uma sessão de avaliação com usuários de aproximadamente sessenta minutos para explorar elementos de usabilidade da abordagem de visualização proposta. Os participantes deste estudo serão estudantes do Instituto de Computação da Universidade Federal da Bahia (UFBA), e colaboradores do Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS), ambos devem ter alguma experiência ou conhecimento sobre análise de dados e utilização de

**Endereço:** Rua Augusto Viana S/N 3º Andar

**Bairro:** Canela

**CEP:** 41.110-060

**UF:** BA

**Município:** SALVADOR

**Telefone:** (71)3283-7615

**Fax:** (71)3283-7615

**E-mail:** cepee.ufba@ufba.br

Continuação do Parecer: 5.757.369

visualizações. Serão selecionados um total de vinte participantes, sendo dez de cada instituição. Com o intuito de enriquecer o processo de compreensão do objeto estudado, este estudo se propõe a utilizar dispositivos de coleta de dados qualitativos e quantitativos. Um dos instrumentos será a observação participante para documentar as interações, comentários, dúvidas e sugestões durante o processo de avaliação com os usuários. Esta, será realizada através de videochamada via a plataforma Google Meet. Outro dispositivo de coleta de dados será o questionário semiestruturado online com questões de múltipla escolha e questões abertas. Para analisar os dados serão utilizadas análise de conteúdo e estatística descritiva. Como resultado, pretende-se desenvolver uma visualização que contribua para a interpretação de métricas de qualidade em projeções multidimensionais."

**Objetivo da Pesquisa:**

Conforme descrito no Formulário de Informações Básicas da Plataforma Brasil:

Objetivo Primário:

"- Desenvolver uma abordagem de exploração visual com múltiplas visualizações coordenadas para interpretação de qualidade de técnicas de projeção multidimensional."

Objetivo Secundário:

- "- Elaborar uma solução de Visual Analytics que contenha múltiplas visualizações coordenadas;
- Implementar mecanismos interativos para explorar métricas de erros na visualização de projeções multidimensionais;
- Acoplar técnicas de visualização que auxiliem na exploração e interpretação de métricas de qualidade em projeções multidimensionais;
- Desenvolver uma abordagem de Visual Analytics que facilite a interpretação de métricas de qualidade em projeções multidimensionais;
- Avaliar a abordagem proposta realizando testes com usuários."

"Esta "Pesquisa de Avaliação com Usuário" não apresenta riscos físicos aos envolvidos. Contudo, sua participação pode envolver algum desconforto, ainda que pequeno, relacionado ao tempo

**Endereço:** Rua Augusto Viana S/N 3º Andar  
**Bairro:** Canela **CEP:** 41.110-060  
**UF:** BA **Município:** SALVADOR  
**Telefone:** (71)3283-7615 **Fax:** (71)3283-7615 **E-mail:** cepee.ufba@ufba.br

Continuação do Parecer: 5.757.369

dedicado ao preenchimento do questionário. Será feito o possível para minimizar possíveis desconfortos e não ocupar o tempo do participante desnecessariamente. A qualquer momento a pessoa poderá desistir de participar da pesquisa sem provocar nenhum prejuízo. As informações obtidas serão confidenciais e será assegurado o sigilo sobre a participação. Logo, os resultados deste estudo em documentos científicos e publicações, não permitirão nenhum tipo de identificação."

Benefícios:

- "- Construção de representações visuais mais dinâmicas e intuitivas, por meio de técnicas de visualização que utilizam diversos mecanismos de interação;
- Apoiar o processo de análise exploratória de uma ampla variedade de dados complexos e com alta dimensionalidade;
- Avançar no estado da arte do tema relacionado a métricas de qualidade de projeções multidimensionais, por meio de uma melhor compreensão da qualidade das técnicas de Projeções Multidimensionais."

**Avaliação dos Riscos e Benefícios:**

Conforme descrito no Formulário de Informações Básicas da Plataforma Brasil:

Riscos:

**Comentários e Considerações sobre a Pesquisa:**

Trata-se de um protocolo de pesquisa de Dissertação do Programa de Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia. Com financiamento próprio, esta pesquisa tem como propósito desenvolver uma abordagem de Visual Analytics para interpretação de qualidade de projeções multidimensionais baseada em múltiplas visualizações coordenadas. Uma abordagem são as visualizações multidimensionais, elas são representações gráficas que codificam dados com n-dimensões para estruturas visuais. Nesse processo, a utilização de diferentes representações visuais na mesma tela e compartilhando mecanismos de interação pode tornar as tarefas de exploração e interpretação mais intuitivas.

**Endereço:** Rua Augusto Viana S/N 3º Andar  
**Bairro:** Canela **CEP:** 41.110-060  
**UF:** BA **Município:** SALVADOR  
**Telefone:** (71)3283-7615 **Fax:** (71)3283-7615 **E-mail:** cepee.ufba@ufba.br

Continuação do Parecer: 5.757.369

O percurso metodológico deste estudo se constituiu a partir de uma pesquisa aplicada exploratória e pretende realizar uma experiência prática de avaliação de usabilidade com usuários. Com o intuito de enriquecer o processo de compreensão do objeto estudado, se propõe a utilizar dispositivos de coleta de dados qualitativos e quantitativos, como a observação participante (Esta etapa será realizada através de video chamadas via a plataforma Google Meet para aproximar o processo do local dos participantes, diminuindo ou eliminando possíveis desconfortos, como deslocamento para um lugar específico, ou ainda possibilitando maior número de usuários uma vez que não é necessário a presença física. No caso de os usuários concordarem, a video chamada será gravada, como forma de registro da observação participante) e questionário semiestruturados com questões abertas (participante pode responder livremente) e, também, questões fechadas (com respostas pré definidas com mensuração baseada na escala Likert. Para viabilizar a aplicação do questionário será utilizada a ferramenta Google Forms). Para analisar os dados qualitativos coletados na observação

participante e questões abertas dos formulários será adotado o procedimento de Análise de Conteúdo para estruturar os dados e extrair informações dos mesmos. Os dados quantitativos obtidos através das perguntas fechadas dos questionários serão analisados a partir de estatística descritiva, que auxiliará na sumarização de um conjunto de valores, possibilitando obter uma visão global dos dados através de tabelas, visualizações e medidas descritivas. Serão selecionados participantes das duas instituições, sendo: i) entre cinco e dez estudantes do Instituto de Computação (IC) da Universidade Federal da Bahia (UFBA); ii) entre cinco e dez colaboradores do Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS).

Critérios de inclusão: estudantes do Instituto de Computação da Universidade Federal da Bahia (UFBA), e colaboradores do Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS), que tenham alguma experiência ou conhecimento sobre análise de dados e utilização de visualizações. Critérios de exclusão: Os estudantes que não pertencem ao Instituto de Computação da Universidade Federal da Bahia (UFBA); Pessoas que não são colaboradores do Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS); Pessoas que não tenham alguma experiência ou conhecimento sobre análise de dados e utilização de visualizações.

Número previsto de participantes: 20;

**Endereço:** Rua Augusto Viana S/N 3º Andar

**Bairro:** Canela

**CEP:** 41.110-060

**UF:** BA

**Município:** SALVADOR

**Telefone:** (71)3283-7615

**Fax:** (71)3283-7615

**E-mail:** cepee.ufba@ufba.br

ESCOLA DE ENFERMAGEM DA  
UNIVERSIDADE FEDERAL DA  
BAHIA - UFBA



Continuação do Parecer: 5.757.369

Previsão de início da pesquisa: 14/11/2022;

Previsão de encerramento da pesquisa: 28/02/2023.

**Considerações sobre os Termos de apresentação obrigatória:**

Foram anexados 03 documentos ao protocolo de pesquisa na Plataforma Brasil. Conforme solicitado no parecer consubstanciado 5.710.412, houve adequações nos documentos apresentados.

**Recomendações:**

Apresentar, como notificação, via Plataforma Brasil, os relatórios parciais e final do projeto, contados a partir da data de aprovação do protocolo de pesquisa, conforme a Resolução CNS 466/2012, itens X.1.- 3.b. e XI.2.d.

**Conclusões ou Pendências e Lista de Inadequações:**

Protocolo de pesquisa atende aos preceitos éticos emanados das resoluções 466/2012 e 510/2016 do Conselho Nacional de Saúde. Assim, sugere-se parecer de aprovação.

**Considerações Finais a critério do CEP:**

Aprovação ad referendum, tendo em vista considerações prévias em reunião de Colegiado. Ressalta-se que, após realizar modificações atendendo às recomendações descritas no parecer consubstanciado anterior, esta segunda versão do projeto atende aos princípios éticos e bioéticos emanados das Resoluções n.466/2012 e n.510/2016 do Conselho Nacional de Saúde.

**Este parecer foi elaborado baseado nos documentos abaixo relacionados:**

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_2007947.pdf	25/10/2022 22:17:15		Aceito
Projeto Detalhado / Brochura Investigador	projeto_pesquisa_cep_jose_augusto.pdf	25/10/2022 22:16:19	JOSE AUGUSTO DUARTE GOMES	Aceito
Cronograma	cronograma.pdf	25/10/2022 22:04:58	JOSE AUGUSTO DUARTE GOMES	Aceito
Outros	termo_de_confidencialidade.pdf	01/09/2022 23:39:49	JOSE AUGUSTO DUARTE GOMES	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE.pdf	01/09/2022 23:37:53	JOSE AUGUSTO DUARTE GOMES	Aceito

**Endereço:** Rua Augusto Viana S/N 3º Andar

**Bairro:** Canela

**CEP:** 41.110-060

**UF:** BA

**Município:** SALVADOR

**Telefone:** (71)3283-7615

**Fax:** (71)3283-7615

**E-mail:** cepee.ufba@ufba.br

ESCOLA DE ENFERMAGEM DA  
UNIVERSIDADE FEDERAL DA  
BAHIA - UFBA



Continuação do Parecer: 5.757.369

Declaração de Instituição e Infraestrutura	termo_de_autorizacao_da_instituicao_coparticipante_ic_assinado.pdf	01/09/2022 23:37:40	JOSE AUGUSTO DUARTE GOMES	Aceito
Declaração de Instituição e Infraestrutura	termo_de_autorizacao_da_instituicao_coparticipante_cidacs.pdf	01/09/2022 23:37:30	JOSE AUGUSTO DUARTE GOMES	Aceito
Orçamento	orcamento.pdf	01/09/2022 23:31:08	JOSE AUGUSTO DUARTE GOMES	Aceito
Declaração de concordância	declaracao_orientador_concordancia_de_senvolvimento_projeto_pesquisa.pdf	01/09/2022 23:30:01	JOSE AUGUSTO DUARTE GOMES	Aceito
Declaração de Pesquisadores	termo_de_compromisso_do_pesquisador.pdf	01/09/2022 23:29:47	JOSE AUGUSTO DUARTE GOMES	Aceito
Declaração de Instituição e Infraestrutura	termo_de_autorizacao_institucional_da_propONENTE_ic_assinado.pdf	01/09/2022 23:26:16	JOSE AUGUSTO DUARTE GOMES	Aceito
Folha de Rosto	folha_de_rosto.pdf	01/09/2022 22:58:32	JOSE AUGUSTO DUARTE GOMES	Aceito

**Situação do Parecer:**

Aprovado

**Necessita Apreciação da CONEP:**

Não

SALVADOR, 14 de Novembro de 2022

---

**Assinado por:**  
**Anderson Reis de Sousa**  
**(Coordenador(a))**

**Endereço:** Rua Augusto Viana S/N 3º Andar

**Bairro:** Canela

**CEP:** 41.110-060

**UF:** BA

**Município:** SALVADOR

**Telefone:** (71)3283-7615

**Fax:** (71)3283-7615

**E-mail:** cepee.ufba@ufba.br



**Apêndice**

**D**

**APROVAÇÃO DO CONSELHO DE ÉTICA EM  
PESQUISA DA FIOCRUZ-BA**



CENTRO DE PESQUISAS  
GONÇALO MONIZ -  
FIOCRUZ/BA



## PARECER CONSUBSTANCIADO DO CEP

Elaborado pela Instituição Coparticipante

### DADOS DO PROJETO DE PESQUISA

**Título da Pesquisa:** Uma Abordagem de Visual Analytics para Análise de Erros de Projeções Multidimensionais Utilizando Múltiplas Visões Coordenadas

**Pesquisador:** JOSE AUGUSTO DUARTE GOMES

**Área Temática:**

**Versão:** 2

**CAAE:** 62907022.8.3001.0040

**Instituição Proponente:** Centro de Integração de Dados e Conhecimentos para Saúde - CIDACS

**Patrocinador Principal:** Financiamento Próprio

### DADOS DO PARECER

**Número do Parecer:** 5.842.393

#### Apresentação do Projeto:

Resposta as pendências do parecer consubstanciado anterior deste protocolo.

Informações contidas nos campos "Apresentação do Projeto", "Objetivo da Pesquisa" e "Avaliação dos Riscos e Benefícios" foram retirados do arquivo "PB\_INFORMAÇÕES\_BÁSICAS\_DO\_PROJETO\_2051018".

#### 1. Resumo:

O crescimento massivo e diversificado dos dados na era do big data, deixa cada vez mais em evidência o desafio de analisar dados com alta dimensionalidade. Nesse sentido, diferentes técnicas de projeção multidimensional foram desenvolvidas para gerar visualizações que contribuem para a análise exploratória de uma ampla variedade de dados complexos e com alta dimensão. O mapeamento realizado na projeção varia de acordo com a técnica aplicada. Esta pesquisa tem como propósito desenvolver uma abordagem de Visual Analytics para interpretação de qualidade de projeções multidimensionais baseada em múltiplas visualizações coordenadas. Nesse processo, a utilização de diferentes representações visuais na mesma tela e compartilhando mecanismos de interação pode tornar as tarefas de exploração e interpretação mais intuitivas. O percurso metodológico deste estudo se constituiu a partir de uma pesquisa aplicada exploratória e

**Endereço:** Rua Waldemar Falcão, 121 na área de Escritórios do 3º Pavimento do Edifício Garagem, sala CEP

**Bairro:** Candeal

**CEP:** 40.296-710

**UF:** BA

**Município:** SALVADOR

**Telefone:** (71)3176-2285

**Fax:** (71)98814-7051

**E-mail:** cep.igm@fiocruz.br

Continuação do Parecer: 5.842.393

pretende realizar uma experiência prática de avaliação de usabilidade com usuários. Com o intuito de enriquecer o processo de compreensão do objeto estudado, se propõe a utilizar dispositivos de coleta de dados qualitativos e quantitativos, como a observação participante e questionário. Para analisar os dados serão utilizadas análise de conteúdo e estatística descritiva. Como resultado, pretende-se desenvolver uma visualização que contribua para a interpretação de métricas de qualidade em projeções multidimensionais.

## 2. Hipótese:

Utilizar diferentes representações visuais na mesma tela, compartilhando mecanismos de interação, contribui para explorar de forma mais intuitiva bases de dados com alta dimensionalidade. Múltiplas Visualizações Coordenadas auxiliam na interpretação de qualidade de projeções multidimensionais.

## 3. Metodologia Proposta:

O percurso metodológico desta pesquisa iniciou-se com um levantamento bibliográfico em publicações diversas relacionadas à análise de dados com alta dimensionalidade e a utilização de múltiplas visões coordenadas. Este estudo está classificado, quanto a sua natureza, como pesquisa aplicada, pois propõe o desenvolvimento de uma Abordagem de Visual Analytics para Análise de Erros de Projeções Multidimensionais Utilizando Múltiplas Visões Coordenadas. Além disso, foi desenvolvida uma ferramenta e pretende-se realizar uma experiência prática de avaliação de usabilidade com usuários. Esta pesquisa caracteriza-se como exploratória, pois buscou-se identificar as lacunas apresentadas pelos trabalhos relacionados para direcionar a elaboração de hipóteses e a aproximação do problema, que é o desafio de analisar erros de projeções multidimensionais utilizando múltiplas visões coordenadas. Para Medeiros (2021), pesquisas exploratórias contribuem para uma série de vantagens propositivas, dentre as quais: obter uma visão geral do objeto pesquisado; formular hipóteses e se familiarizar com o problema; auxiliar no estabelecimento de critérios, métodos e técnicas para elaboração de uma pesquisa. Com o intuito de enriquecer o processo de compreensão do objeto estudado, a abordagem dessa pesquisa será qualiquantitativa, que é caracterizada pela utilização de instrumentos de coleta de dados tanto qualitativos, como a observação participante e/ou questões dissertativas em formulários, quanto

**Endereço:** Rua Waldemar Falcão, 121 na área de Escritórios do 3º Pavimento do Edifício Garagem, sala CEP  
**Bairro:** Candeal **CEP:** 40.296-710  
**UF:** BA **Município:** SALVADOR  
**Telefone:** (71)3176-2285 **Fax:** (71)98814-7051 **E-mail:** cep.igm@fiocruz.br

Continuação do Parecer: 5.842.393

quantitativos, como formulários com questões fechadas. Segundo Sampieri e Torres (2014), pesquisas que integram métodos quantitativos e qualitativos em um mesmo estudo, podem promover uma discussão mais ampla e maximizar o entendimento do fenômeno que se pretende estudar. Para coletar os dados, este estudo pretende realizar uma pesquisa de campo, onde ocorrerá uma avaliação com usuários para explorar elementos de usabilidade da abordagem de Visual Analytics proposta. Os dispositivos de coleta de dados utilizados serão:

- Observação participante, para documentar as interações, comentários, dúvidas e sugestões em todas as etapas do processo de avaliação com usuários. Esta etapa será realizada através de videochamadas via a plataforma Google Meet para aproximar o processo do local dos participantes, diminuindo ou eliminando possíveis desconfortos, como deslocamento para um lugar específico, ou ainda possibilitando maior número de usuários uma vez que não é necessário a presença física. No caso de os usuários concordarem, a videochamada será gravada, como forma de registro da observação participante.
- Questionário semiestruturado com questões abertas, ou seja, em que o participante pode responder livremente e, também, questões fechadas, com respostas pré-definidas com mensuração baseada na escala Likert. Para viabilizar a aplicação do questionário será utilizada a ferramenta Google Forms.

#### 4. Critério de Inclusão:

Os participantes deste estudo serão estudantes do Instituto de Computação da Universidade Federal da Bahia (UFBA), e colaboradores do Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS), que tenham alguma experiência ou conhecimento sobre análise de dados e utilização de visualizações.

#### 5. Critério de Exclusão:

Não serão participantes deste estudo:

Os estudantes que não pertencem ao Instituto de Computação da Universidade Federal da Bahia (UFBA); Pessoas que não são colaboradores do Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS); Pessoas que não tenham alguma experiência ou conhecimento sobre análise de dados e utilização de visualizações.

#### 6. Outras informações importantes:

**Endereço:** Rua Waldemar Falcão, 121 na área de Escritórios do 3º Pavimento do Edifício Garagem, sala CEP  
**Bairro:** Candeal **CEP:** 40.296-710  
**UF:** BA **Município:** SALVADOR  
**Telefone:** (71)3176-2285 **Fax:** (71)98814-7051 **E-mail:** cep.igm@fiocruz.br

Continuação do Parecer: 5.842.393

- a) Tamanho da Amostra no Brasil: 20;
- b) Não haverá uso de fontes secundárias de dados (prontuários, dados demográficos, etc);
- c) O Estudo não é multicêntrico no Brasil;
- d) Não propõe dispensa do TCLE;
- e) Não haverá retenção de amostras para armazenamento em banco;
- f) Início previsto da pesquisa: 14/11/2022;
- g) Orçamento: R\$ 1.200,00;
- h) Financiamento: "não se aplica". "Para a elaboração e realização deste projeto foi necessário que o pesquisador responsável custeasse com meios financeiros próprios a aquisição de recursos para o desenvolvimento da pesquisa.

**Objetivo da Pesquisa:**

1. Objetivo Primário:

Desenvolver uma abordagem de exploração visual com múltiplas visualizações coordenadas para interpretação de qualidade de técnicas de projeção multidimensional.

2. Objetivo Secundário:

- a) Elaborar uma solução de Visual Analytics que contenha múltiplas visualizações coordenadas;
- b) Implementar mecanismos interativos para explorar métricas de erros na visualização de projeções multidimensionais;
- c) Acoplar técnicas de visualização que auxiliem na exploração e interpretação de métricas de qualidade em projeções multidimensionais;
- d) Desenvolver uma abordagem de Visual Analytics que facilite a interpretação de métricas de qualidade em projeções multidimensionais;
- e) Avaliar a abordagem proposta realizando testes com usuários.

**Endereço:** Rua Waldemar Falcão, 121 na área de Escritórios do 3º Pavimento do Edifício Garagem, sala CEP  
**Bairro:** Candeal **CEP:** 40.296-710  
**UF:** BA **Município:** SALVADOR  
**Telefone:** (71)3176-2285 **Fax:** (71)98814-7051 **E-mail:** cep.igm@fiocruz.br

### **Avaliação dos Riscos e Benefícios:**

#### 1. Riscos:

Esta "Pesquisa de Avaliação com Usuário" não apresenta riscos físicos aos envolvidos. Contudo, sua participação pode envolver algum desconforto, ainda que pequeno, relacionado ao tempo dedicado ao preenchimento do questionário. Será feito o possível para minimizar possíveis desconfortos e não ocupar o tempo do participante desnecessariamente. A qualquer momento a pessoa poderá desistir de participar da pesquisa sem provocar nenhum prejuízo. As informações obtidas serão confidenciais e será assegurado o sigilo sobre a participação. Logo, os resultados deste estudo em documentos científicos e publicações, não permitirão nenhum tipo de identificação.

#### 2. Benefícios:

O benefício relacionado à participação nesta pesquisa é contribuir para: Construção de representações visuais mais dinâmicas e intuitivas, por meio de técnicas de visualização que utilizam diversos mecanismos de interação;

Apoiar o processo de análise exploratória de uma ampla variedade de dados complexos e com alta dimensionalidade;

Avançar no estado da arte do tema relacionado a métricas de qualidade de projeções multidimensionais, por meio de uma melhor compreensão da qualidade das técnicas de Projeções Multidimensionais.

### **Comentários e Considerações sobre a Pesquisa:**

A pesquisa apresentada é relevante e o proponente envolvido possui experiência na metodologia que será executada.

### **Considerações sobre os Termos de apresentação obrigatória:**

Vide lista de conclusões e pendências.

### **Recomendações:**

-

Continuação do Parecer: 5.842.393

**Conclusões ou Pendências e Lista de Inadequações:**

I. Análise das respostas às pendências apontadas no parecer consubstanciado anterior deste protocolo:

1. "Inserir Danilo Barbosa Coimbra como parte da equipe no formulário da Plataforma Brasil. RES. CNS Nº 510/2016, cap. I, art. 02-II". O membro da equipe não foi inserido no formulário, mas consta anuência do mesmo no "termo de confidencialidade". Justificativa aceita;

2. "Justificar o tamanho da amostra (participante da pesquisa). No projeto informa seleção "entre cinco e dez colaboradores de cada centro participante". CNS Nº 510/2016, cap. I, art. 02-X". Justificativa inserida;

3. "Substituir a frase "o estudo não apresenta riscos físicos aos envolvidos" por "Toda pesquisa possui risco ao participante da pesquisa, porém...". RES. CNS Nº 510/2016, cap. IV, art. 19". Correção realizada;

4. "Descrever na metodologia o cenário onde e como os participantes da pesquisa serão recrutados nos seus centros participantes do estudo. RES. CNS Nº 510/2016, cap. III, art. 06". Descrição realizada;

5. "No instrumento de coleta "questionário" excluir "Qual o seu gênero?", visto que esta pergunta não terá impacto no resultado do estudo". Item excluído;

6. No TCLE:

a) "Inserir numeração nas páginas no modelo 1/3, 2/3, etc.". Número de páginas inserido, mas não no formato solicitado. Formato acatado;

b) "Inserir que os participantes terão direito a ressarcimento de despesas e indenização decorrentes a danos decorrentes da sua participação na pesquisa, nos termos da Lei. RES. CNS Nº 510/2016, cap. III, art. 09". Informações inseridas;

c) "Inserir campo para rubrica do participante e pesquisador em todas as páginas do documento.

**Endereço:** Rua Waldemar Falcão, 121 na área de Escritórios do 3º Pavimento do Edifício Garagem, sala CEP  
**Bairro:** Candeal **CEP:** 40.296-710  
**UF:** BA **Município:** SALVADOR  
**Telefone:** (71)3176-2285 **Fax:** (71)98814-7051 **E-mail:** cep.igm@fiocruz.br

Continuação do Parecer: 5.842.393

RES. CNS Nº 510/2016, seção. II, art. 17-10.3". Informação inserida;

d) "Inserir no final endereços, número do telefone e e-mail dos CEPs participantes e endereço do pesquisador responsável (Comitê de Ética em Pesquisa (CEP-IGM): Centro de Pesquisas Gonçalo Moniz, FIOCRUZ-BA, R. Waldemar Falcão, 121, Candeal de Brotas Tel.: 71-3176-2285 / E-mail: cep.igm@fiocruz.b). RES. CNS Nº 510/2016, seção. II, art. 08 e 09". Informações inseridas;

e) "Toda pesquisa possui risco ao participante da pesquisa, mesmo que mínimos". Esclarecer os riscos submetidos ao participante da pesquisa e informar como estes serão minimizados. RES. CNS Nº 510/2016, cap. IV, art. 19". Informações inseridas;

f) "Substituir a frase "Recebi uma cópia deste termo de consentimento livre e esclarecido" por "Recebi uma via deste termo de consentimento livre e esclarecido. RES. CNS Nº 510/2016, seção. II, art. 17-10.3". Informação inserida;

g) "Inserir garantia de informação sobre a forma de acompanhamento da pesquisa, inclusive considerando de benefícios diretos, quando houver. RES. CNS Nº 510/2016, seção. II, art. 17-05". Informação inserida;

h) "Substituir o texto do TCLE contido no projeto pelo texto corrigido contido no arquivo "TCLE". Texto substituído.

Obs: Para que as informações contidas no "projeto" também possam constar no "Formulário da Plataforma Brasil", sugerimos que o proponente do estudo submeta uma emenda (Ec1) anexando uma carta explicando o motivo da emenda ("Atualização do formulário da Plataforma Brasil") e listando os itens que sofreram atualização. Assim poderá fazer todas as atualizações (metodologia, cenário de abordagem do participante, riscos, etc), inclusive o cronograma do estudo, incluindo data de início da pesquisa posterior a aprovação do CEP.

#### **Considerações Finais a critério do CEP:**

O prazo para a execução desse projeto é 28/02/2023 conforme cronograma especificado nas informações básicas do projeto na plataforma.

**Endereço:** Rua Waldemar Falcão, 121 na área de Escritórios do 3º Pavimento do Edifício Garagem, sala CEP  
**Bairro:** Candeal **CEP:** 40.296-710  
**UF:** BA **Município:** SALVADOR  
**Telefone:** (71)3176-2285 **Fax:** (71)98814-7051 **E-mail:** cep.igm@fiocruz.br

Continuação do Parecer: 5.842.393

Em cumprimento da Res. 466/2012 e Norma Complementar vigente, enviar relatórios parciais a cada seis meses e relatório final em até um mês após o término da vigência do projeto conforme cronograma aprovado neste protocolo.

Relatório final: 28/03/2023.

Caso haja inclusão de outra fonte de financiamento após esta aprovação, informar ao CEP como emenda ao protocolo incluindo o Termo de Outorga ou documento equivalente e realizando as alterações pertinentes na Folha de Rosto para indicar modificação do patrocinador principal.

The present study has been approved by the Comitê de Ética em Pesquisa do Centro de Pesquisas Gonçalo Moniz/FIOCRUZ (IORG0002090/OMB No. 0990-0279 valid until 01/27/2025). The protocol and procedures presented in the project are in full accordance with the Brazilian legislation regarding the ethical standards in conducting research involving human beings (Res. CNS 466/2012).

**Este parecer foi elaborado baseado nos documentos abaixo relacionados:**

Tipo Documento	Arquivo	Postagem	Autor	Situação
Cronograma	Cronograma_atualizado.pdf	31/12/2022 08:54:07	Jorge Clarêncio Souza Andrade	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	tcle_corrigido.pdf	31/12/2022 08:53:25	Jorge Clarêncio Souza Andrade	Aceito
Projeto Detalhado / Brochura Investigador	projeto_pesquisa_jose_augusto_corrigido.pdf	31/12/2022 08:52:23	Jorge Clarêncio Souza Andrade	Aceito
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_2051018.pdf	30/12/2022 16:04:19		Aceito
Outros	carta_resposta_cep.pdf	30/12/2022 16:01:44	JOSE AUGUSTO DUARTE GOMES	Aceito
Projeto Detalhado / Brochura Investigador	projeto_pesquisa_cep_jose_augusto.pdf	25/10/2022 22:16:19	JOSE AUGUSTO DUARTE GOMES	Aceito
Outros	termo_de_confidencialidade.pdf	01/09/2022 23:39:49	JOSE AUGUSTO DUARTE GOMES	Aceito
TCLE / Termos de Assentimento /	TCLE.pdf	01/09/2022 23:37:53	JOSE AUGUSTO DUARTE GOMES	Aceito

**Endereço:** Rua Waldemar Falcão, 121 na área de Escritórios do 3º Pavimento do Edifício Garagem, sala CEP  
**Bairro:** Candeal **CEP:** 40.296-710  
**UF:** BA **Município:** SALVADOR  
**Telefone:** (71)3176-2285 **Fax:** (71)98814-7051 **E-mail:** cep.igm@fiocruz.br

Continuação do Parecer: 5.842.393

Justificativa de Ausência	TCLE.pdf	01/09/2022 23:37:53	JOSE AUGUSTO DUARTE GOMES	Aceito
---------------------------	----------	------------------------	------------------------------	--------

**Situação do Parecer:**

Aprovado

**Necessita Apreciação da CONEP:**

Não

SALVADOR, 31 de Dezembro de 2022

---

**Assinado por:**  
**Jorge Clarêncio Souza Andrade**  
**(Coordenador(a))**

**Endereço:** Rua Waldemar Falcão, 121 na área de Escritórios do 3º Pavimento do Edifício Garagem, sala CEP  
**Bairro:** Candeal **CEP:** 40.296-710  
**UF:** BA **Município:** SALVADOR  
**Telefone:** (71)3176-2285 **Fax:** (71)98814-7051 **E-mail:** cep.igm@fiocruz.br