



UNIVERSIDADE FEDERAL DO PARÁ
CENTRO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

RODRIGO AUGUSTO DE MORAES LOURENÇO

Explorando, Comparando e Coordenando
Múltiplas Fontes de Dados em uma
Ferramenta de Visualização de Informação

Prof. Dr. Bianchi Serique Meiguins
Orientador

Rodrigo Augusto de Moraes Lourenço

**Explorando, Comparando e Coordenando
Múltiplas Fontes de Dados em uma
Ferramenta de Visualização de Informação**

Dissertação de Mestrado apresentada
para obtenção do grau de Mestre em
Ciência da Computação.
Programa de Pós Graduação em
Ciência da Computação.
Instituto de Ciências Exatas e Naturais.
Universidade Federal do Pará.
Orientador Prof. Dr. Bianchi Serique
Meiguins

Lourenço, Rodrigo Augusto de Moraes

Explorando, Comparando e Coordenando Múltiplas Fontes de Dados em uma Ferramenta de Visualização de Informação / (Rodrigo Augusto de Moraes Lourenço); orientador, Bianchi Serique Meiguins. - 2012.

77 f. il. 28 cm

Dissertação (Mestrado) – Universidade Federal do Pará. Instituto de Ciências Exatas e Naturais. Programa de Pós-Graduação em Ciência da Computação. Belém, 2012.

1. Tecnologia da Informação. 2. Visualização de Informações. 3. Bancode Dados
I. Meiguins, Bianchi Serique, orient. II. Universidade Federal do Pará, Instituto de Ciências Exatas e Naturais, Programa de Pós-Graduação em Ciência da Computação.
III. Título.

CDD 22. ed. 004

Rodrigo Augusto de Moraes Lourenço

**Explorando, Comparando e Coordenando
Múltiplas Fontes de Dados em uma
Ferramenta de Visualização de Informação**

Dissertação de Mestrado apresentada
para obtenção do grau de Mestre em
Ciência da Computação.
Programa de Pós Graduação em
Ciência da Computação.
Instituto de Ciências Exatas e Naturais.
Universidade Federal do Pará.

Data da aprovação: Belém-PA. 02/07/2012

Banca Examinadora

Prof. Dr. Bianchi Serique Meiguins
Instituto de Computação – UFPA – Orientador

Prof. Dr. Eloi Luiz Favero
Instituto de Computação – UFPA – Membro

Prof. Dr. Josivaldo de Souza Araújo
Instituto de Computação – UFPA – Membro

Prof. Dr. Marcelo de Paiva Guimarães
Universidade Federal de São Paulo – UNIFESP – Membro

Visto:

Prof. Sandro Ronaldo Bezerra Oliveira, Dr. (UFPA)

Coordenador do PPGCC – UFPA

Dedico este trabalho a minha esposa
Roberta pelo amor, paciência, e
companheirismo, aos meus pais pelo apoio,
aos meus mestres e, todos aqueles que
participaram dessa conquista.

Agradecimento

Agradeço a Deus por iluminar minha caminhada nestes prazerosos e desafiadores anos de pesquisa, que ele ilumine a todos que buscam o conhecimento.

A minha esposa Roberta Lima Spindler, pelo apoio, amor, incentivo e exemplo de persistência. Te amo.

Aos meus pais Manuel David Lourenço e Maria da Conceição de Moraes Lourenço por me mostrarem garra no trabalho, amor e persistência, e todos os diversos ensinamentos que levarei para a toda a vida.

Aos professores do segundo grau, que me mostraram talento e a paixão por ensinar. Verdadeiros heróis em sala de aula que, com paciência e persistência, transmitem conhecimento e enfrentam os desafios da sala de aula.

Aos professores do ensino superior, estes que contribuíram diretamente na minha formação profissional, abrindo diversas portas com novas oportunidades.

Em especial a M.Sc. Aruanda Meiguins e ao Prof. Dr. Bianchi Meiguins que tenho grande carinho e admiração, sempre deram apoio, orientação e acreditaram no meu trabalho. Marcaram os estes últimos anos da melhor maneira possível, e hoje são grandes amigos que levo.

Agradeço a empresa Rede Informática LTDA que foi o palco destes últimos anos, como ambiente de trabalho, pesquisa e confraternização. Aonde conheci grandes amigos, aprendi muito, trabalhei bastante e dei muitas risadas. A todos os colegas de empresa e de pesquisa, pelos bons anos de trabalho, convivência, companheirismo, aprendizados e pizzas.

Por fim a todos aqueles que contribuíram direta ou indiretamente para a conclusão deste trabalho.

SUMÁRIO

LISTA DE FIGURAS	9
PUBLICAÇÕES	11
RESUMO	12
ABSTRACT	13
1 INTRODUÇÃO	14
1.1 Objetivos.....	15
1.2 Organização da Dissertação	15
2 VISUALIZAÇÃO DE INFORMAÇÃO	16
2.1 Definição de Visualização	16
2.2 Histórico de visualização.....	17
2.3 Processo de Visualização.....	24
2.3.1 Pipeline de Visualização	26
2.4 Papel da Percepção e Papel do Usuário.....	27
2.5 Tipos de Técnicas de Visualização de Informação	29
2.5.1 Técnicas de Visualização de Informação.....	30
2.6 Múltiplas Visões Coordenadas	33
2.7 PRISMA	35
3 TRABALHOS RELACIONADOS	38
3.1 Ferramentas com Coordenação Flexível	38
3.1.1 Snap-Together.....	38
3.1.2 Improvise	39
3.2 Ferramentas com Coordenação Não-Flexível.....	40
3.2.1 VisGets.....	40
3.2.2 HotMap	41
3.2.3 Prisma	41
4 IMPLEMENTAÇÃO DA FERRAMENTA	43
4.1 Principais Características	43
4.2 Considerações no uso de múltiplas bases de dados	43
4.2.1 Carregamento de múltiplas bases de dados	43
4.2.2 Coordenação entre múltiplas bases de dados.....	43
4.3 Evolução PRISMA.....	44
4.3.1 Singletons.....	45
4.3.2 Falta de padronização	45
4.3.3 Abstração para múltiplas bases.....	46

4.4	Arquitetura e Modelagem.....	46
4.4.1	ModuloVis	46
4.4.2	Núcleo.....	49
4.4.3	Apresentação.....	50
4.4.4	Controle de Coordenação.....	50
5	A FERRAMENTA E SUAS FUNCIONALIDADES	52
5.1	Aspectos Gerais da Interface	52
5.2	Controles de Configuração	52
5.2.1	Configurações gerais de um MóduloVis.....	53
5.2.2	Dispersão de Dados	54
5.2.3	Treemap	55
5.2.4	Coordenadas Paralelas	55
5.3	Configuração das Cores	56
5.3.1	Cores Discretas ou Categóricas	56
5.3.2	Cores Contínuas	56
5.4	Consultas Dinâmicas (Filtros)	57
5.4.1	Filtros Categóricos	57
5.4.2	Filtros Contínuos.....	57
5.5	Seleção e Brushing.....	58
5.5.1	Realizando uma Seleção	58
5.5.2	Realizando um Brushing.....	59
5.6	Detalhes Sob Demanda.....	60
5.7	Coordenação Entre Visões.....	61
5.8	Coordenação Entre Bases de Dados	61
5.8.1	Painel de configuração.....	61
5.8.2	Organização das visualizações.....	62
5.8.3	Relacionamento visual entre visão e configuração.....	63
5.8.4	Desacoplamento de visões	63
6	TESTES DE USABILIDADE	65
6.1	Visão Geral.....	65
6.2	Tarefas Propostas	66
6.3	Avaliação Pós-Teste (NASA-TLX).....	66
6.4	Resultados	67
7	CONSIDERAÇÕES FINAIS	70
7.1	Desafios Encontrados	71
7.2	Trabalhos Futuros	71
	REFERÊNCIAS.....	72

LISTA DE FIGURAS

Figura 1: Exemplos de pintura das cavernas (http://www.lascaux.culture.fr/).	17
Figura 2: Placa de calcário Kish da Mesopotâmia.	17
Figura 3: Mapa de Peutinger. (LENDERING, 2012).....	18
Figura 4: Mapa de John Snow de epidemia de cólera em Londres	19
Figura 5: Mapa de Minard - Representação gráfica da marcha de Napoleão à Rússia (SPENCE, 2007).....	19
Figura 6: (a) Representação circular das fases da lua; (b) Movimento planetário.	20
Figura 7: Playfair - Uso de valores abstratos nos eixos.....	20
Figura 8: Gráfico de série temporal sobre a balança comercial.	20
Figura 9: Visualização de eventos históricos.	21
Figura 10: Visualização de Florence Nightingale envolvendo comportamento ao longo do tempo e comparações.....	21
Figura 11: Imagens médicas de Leonardo da Vinci.	22
Figura 12: Mapa do metro de Londres inspirado no diagramas de circuitos elétricos... ..	22
Figura 13: Representações geográficas reais do metro de Londres.....	23
Figura 14: Exemplo de Eletrocardiograma: (a) normal – esquerda, (b) anormal - direita	23
Figura 15: Representações de vasos sanguíneos (http://www.bodywolds.com).	24
Figura 16: Estudo do fluxo de ar na decolagem de aeronaves. (http://quest.nasa.gov/aero/background/tools).....	24
Figura 17: Ferramenta para estudo de genes, para mapeamento de relacionamentos e comportamentos dos mesmos. (http://www.genedata.com/products/selector/modules.html).....	24
Figura 18: Etapas do processo de visualização. (WARD, GRINSTEIN e KEIM, 2010)25	25
Figura 19: Pipeline de visualização (WARD, GRINSTEIN e KEIM, 2010).....	26
Figura 20: Exemplo de ambiguidade em uma imagem.	27
Figura 21: identificação rápida por (a) cor, (b) orientação, e (c) exemplo de não uso de primitivas de pré-atenção.....	27
Figura 22: Recomendações de Bertin para utilização de atributos visuais bidimensionais	29
Figura 23: Exemplo de Técnica de Dispersão de Dados multidimensional.	31
Figura 24: Planos cartesianos X-Y, Y-Z e X-Z representando os pontos (3, 2, 0), (4, 1, 4), (7, 5, 3), (6, 1, 5)	31
Figura 25: Pontos representados através de Coordenadas Paralelas	31
Figura 26: Exemplo de aplicação da técnica de Coordenadas Paralelas	32
Figura 27: Treemap com alto aspecto relacional e Treemap com baixo aspecto relacional - cluster e squarified respectivamente	32
Figura 28: Organização hierárquica dos dados no Treemap	33
Figura 29: ferramenta PRISMA, exemplo de utilização de múltiplas visões de dados.. ..	33

Figura 30: PRISMA com exemplo de coordenação entre as diversas visualizações.	36
Figura 31: Esquema de coordenação no PRISMA.	36
Figura 32: Tipos de filtro no PRISMA.	37
Figura 33: PRISMA - detalhes sob demana.	37
Figura 34 Ferramenta Snap-Together (NORTH, 2000)	39
Figura 35 Ferramenta Improvise (WEAVER, 2006).....	39
Figura 36 Ferramenta VisGets (DÖRK, CARPENDALE, <i>et al.</i> , 2008)	40
Figura 37 Ferramenta HotMap (HOEBER e YANG, 2006)	41
Figura 38 Ferramenta Prisma	42
Figura 39 Prisma apresentando múltiplas bases de dados.....	42
Figura 40 Estação de trabalho com múltiplos displays	44
Figura 41 Exemplo de Singleton	45
Figura 42 Reestruturação dos Módulos Principais.....	46
Figura 43 MóduloVis	47
Figura 44 Núcleo	49
Figura 45 Apresentação.....	50
Figura 46 Visão Detalhada da Arquitetura.....	51
Figura 47 Interface principal.....	52
Figura 48 Abas de Configurações.....	53
Figura 49 Abas de Configurações, representação alternativa.....	53
Figura 50 Controle de Hierarquia.....	54
Figura 51 Configuração da técnica Dispersão de Dados.....	54
Figura 52 Configurações Treemap.....	55
Figura 53 Configurações Coordenadas Paralelas.....	55
Figura 54 Exemplo de cores para atributos categóricos.....	56
Figura 55 Exemplo de cores para atributos numéricos.....	56
Figura 56 Seleção de atributos para filtragem.....	57
Figura 57 Exemplo de configuração de filtro para atributos categóricos.....	57
Figura 58 Exemplo de filtro para atributos numéricos.....	58
Figura 59 Realizando seleção I.....	58
Figura 60 Realizando seleção II.....	59
Figura 61 Brushing Treemap.....	59
Figura 62 Brushing Coordenadas Paralelas.....	60
Figura 63 Brushing Disperão.....	60
Figura 64 Detalhes Sob Demanda.....	61
Figura 65 Painel de Coordenação.....	62
Figura 66 Organização de interface gráfica de usuário para 3 bases de dados.....	62
Figura 67 Relacionamento de visão controle 1.....	63
Figura 68 Relacionamento visão controle 2.....	63
Figura 69 Visões desacopladas da janela principal da ferramenta.....	64
Figura 70 Média de acertos por grupo e tarefa.....	67
Figura 71 Média de tempo em segundos gasto em cada tarefa.....	68
Figura 72 Médias das complexidades por tarefa e grupo.....	68
Figura 73 Carga de Trabalho final entre os grupos.....	69

PUBLICAÇÕES

Lourenço, R., Guimarães, R., Carneiro, N., Franco, R., Meiguins, A., & Meiguins, B. (2012). Exploring, Comparing and Coordinating Multiple Datasets in an Information Visualization Tool. *IV2012*. Montpellier: IEEE Press.

Lourenço, R., Guimarães, R., Carneiro, N., Meiguins, A., & Meiguins, B. (2010). Using Multiple Datasets in Information Visualization Tool. In: T. Sobh, & K. Elleithy, *Innovations in Computing Sciences and Software Engineering* (pp. 601-606). Springer.

RESUMO

A quantidade de dados eletrônicos disponíveis para análise cresce a cada dia. É fato que esses dados encontram-se dispersos em diversas plataformas e formatos, e que há um esforço grande de pré-processamento deles para convergir, normalmente, em única base de dados multidimensional para realização da análise dos mesmos em busca de novas informações e conhecimentos. Com o objetivo de diminuir este esforço e tempo de pré-processamento, propõe-se que o pré-processamento dos dados seja realizado de forma local, naquele conjunto de dados. Posterior a isso, através de modelos de coordenação de dados, implementados em uma ferramenta de visualização, seriam definidos dados comuns entre as bases de dados envolvidas para que elas pudessem ser relacionadas. A manipulação e visualização dos dados seriam realizados através de ferramenta de visualização de informação, ações básicas de uma ferramenta de visualização de informação, tais como: aplicar filtros e configurar cores seriam aplicados integralmente nas bases coordenadas como se fossem únicas, facilitando a percepção do usuário sobre os dados e seus relacionamentos. Como estudo de caso, foi utilizada a ferramenta de visualização de informação PRISMA, que tem como características principais a facilidade de incluir novas técnicas de visualização de informação em seu framework, e a possibilidade de apresentar múltiplas técnicas de visualização de informação sobre a mesma base de dados. A evolução do PRISMA para múltiplas bases coordenadas tem como objetivo facilitar a descoberta de informações obtidas pelo relacionamento entre duas ou mais bases de dados. A obtenção destas novas funcionalidades ocorreu em duas etapas distintas: A primeira etapa tratou da avaliação e reformulação da arquitetura da ferramenta para permitir a interação por diversas bases de dados sem perda de funcionalidade, em relação as atuais do PRISMA. Já a segunda etapa teve como foco a implementação do módulo de coordenação entre bases de dados e como apresentar ao usuário estes novos controles.

ABSTRACT

The amount of electronic data available for analysis grows every day. It is a fact that these data are spread across multiple platforms and formats, and there is a great effort to pre-process these data to converge normally into a single dataset to perform multidimensional analysis of the dataset searching new information and knowledge. In order to reduce this effort and time spent on pre-processing, it is proposed that the pre-processing of data is done locally, on the dataset. Subsequent to this, through coordination of data models, implemented in a visualization tool, would be defined common data between the databases involved so that they could be related. The manipulation and visualization of data would be achieved through an information visualization tool, the basic actions of a information visualization tool, such as apply filters and set colors would be applied in full at the coordinated datasets as if they are only one, facilitating user perception about the data and their relationships. As a case study, we used the PRISMA information visualization tool, which has as its main features the ease of visualization techniques include new information in its framework, and the possibility of multiple visualization techniques for information on the same dataset. The evolution of PRISMA for multiple coordinated datasets aims to facilitate the discovery of information obtained by the relationship between two or more datasets. Achieving these new features occurred in two stages: The first stage dealt with the evaluation and redesign of the architecture of the tool to allow interaction by several datasets without loss of functionality in relation to the current PRISMA. The second step focuses on the implementation of the module of coordination between databases and how to present the user with these new controls.

1 INTRODUÇÃO

Business Intelligence (BI) é uma área que durante os últimos anos tem sido alvo de crescente interesse de empresas de tamanho variado e no mundo do mercado. Este interesse vem das necessidades das empresas de ganhar conhecimento sobre o seu próprio negócio para tomar decisões que podem aumentar os lucros, melhorar a competitividade, alcançar metas, entre outros.

Esta área se faz valer hoje, neste momento em que o arsenal tecnológico cresce rapidamente em termos de produtos eletrônicos, máquinas com alto poder de processamento e armazenamento, sistemas de informação que recolhem e armazenam os dados gerados a cada segundo em todo tipo de transação, tais como cartões de crédito de compra ou transferências bancárias. Com isso, a quantidade de dados que pode e é de fato armazenada em repositórios hoje em dia é enorme, e na maioria dos casos são apenas armazenados, tornando-se apenas lixo digital.

Assim, fazendo escolhas oportunas e apropriadas de negócios diante de enormes quantidades de dados torna-se extremamente difícil sem o apoio de uma ferramenta. Em muitos casos, os dados são analisados por meio de uma planilha ou tabular com gráficos simples. Em algumas situações, estas folhas são muito detalhadas (com várias colunas e grande volume de páginas), fazendo com que a pessoa que está à procura, incapaz de visualizar todos os dados de uma só vez, e perde a informação ali contida. Por outro lado, gerar uma planilha sintetizada (com poucas colunas e páginas) pode impedir o analista de perceber detalhes que poderiam ser relevantes para a tomada de decisão.

Neste contexto, a área de Visualização de Informação (VI) tem aumentado nas últimas décadas (SPENCE, 2007), com o desenvolvimento de ferramentas computacionais que auxiliam a tomada de decisões. AVI é destinada a auxiliar a análise e compreensão dos dados (KERREN, EBERT e MEYER, 2007) de forma rápida, intuitiva e interativa, utilizando a capacidade cognitiva humana para extrair informações de dados através de representações visuais dos dados (SPENCE, 2007). O sentido da visão humana, quando operado corretamente torna-se um recurso muito eficaz para esta tarefa, uma vez que este efeito tem grande facilidade e velocidade de detecção de alterações no ambiente.

Além dessas necessidades, muitas vezes têm-se a necessidade de estabelecer uma relação entre duas ou mais fontes de dados. Uma das maneiras de enfrentar este problema é unir os conjuntos de dados em um só, mas esta tarefa pode exigir um esforço árduo e demorado de ser realizado, para que por fim se possa apresentar este novo conjunto de dados para uma ferramenta VI.

Dado este cenário, este trabalho apresenta uma ferramenta de visualização de informação que permite a análise de conjunto de dados múltiplos e coordenação utilizando técnicas de visualização (como Dispersão (SPENCE, 2007), Coordenadas

Paralelas(INSELBERG e DIMSDALE, 1990) e Treemap (SHNEIDERMAN, 1992)) e recursos como visão geral dos dados, seleção, brushing, filtro e detalhes sob demanda, VI "mantra"(SHNEIDERMAN, 1996).

1.1 Objetivos

O objetivo principal desta dissertação é apresentar uma ferramenta computacional de visualização de informação utilizando múltiplas visões coordenadas e coordenação entre bases de dados. Para alcançar este objetivo maior, pode-se destacar os seguintes objetivos específicos:

- Analisar, propor e desenvolver meios de coordenação entre bases de dados para ferramentas de VI;
- Desenvolver meios para fácil identificação do contexto visualizado;
- Analisar, reformular e expandir a arquitetura atual da ferramenta Prisma;
- Realizar testes com usuários focados às novas funcionalidades;
- Propor uma arquitetura para ferramentas de visualização de informação com coordenação entre bases de dados;
- Propor meios de interação entre múltiplas bases de dados;

1.2 Organização da Dissertação

O restante do texto da dissertação está organizado da seguinte forma:

O Capítulo 2 discorre sobre a área de visualização de informações, abordando os conceitos, características, técnicas de visualização utilizadas neste trabalho, o uso de múltiplas visões coordenadas e a utilização de coordenação entre bases de dados.

O Capítulo 3 relaciona os principais trabalhos relacionados que contribuíram para o desenvolvimento dessa dissertação.

O Capítulo 4 apresenta os principais aspectos do desenvolvimento da ferramenta, como: características e limitações, tecnologias, arquitetura, modelagem.

O Capítulo 5 detalha todas as funcionalidades da ferramenta, abordando as configurações das visões, cores, filtros, detalhes sob demanda, a realização de seleção e brushing e a coordenação entre bases de dados.

O Capítulo 6 trata dos testes de usabilidade realizados na ferramenta e seus resultados.

O Capítulo 7 conclui a dissertação apresentando as considerações finais e trabalhos futuros.

2 VISUALIZAÇÃO DE INFORMAÇÃO

2.1 Definição de Visualização

Pode-se definir visualização como um tipo de comunicação da informação que utiliza representações gráficas (WARD, GRINSTEIN e KEIM, 2010). Imagens têm sido utilizadas como forma de comunicação muito antes da linguagem escrita. Uma imagem simples pode conter muitas informações simultaneamente, e pode ser processada muito mais rapidamente por um ser humano quando comparada com uma página de texto (palavras). A razão disso está no sistema de percepção humana que processa as diversas informações da imagem em paralelo, enquanto a interpretação de um texto é limitada pelo processo sequencial de leitura do mesmo. Imagens podem também ser independentes de uma linguagem local, como um gráfico ou um mapa que podem ser entendidos por um conjunto de pessoas sem uma língua comum (WARE, 2004).

De acordo com Few (2009), de maneira geral, o termo Visualização é aplicado para representação visual da informação, e pode ser associado a três palavras principais, criando três termos com significados um pouco diferentes, são eles: Visualização de Dados, Visualização da Informação e Visualização Científica.

Para Few (2009), o termo Visualização de Dados pode ser utilizado como um termo guarda-chuva para cobrir todos os tipos de representações visuais que suportam a exploração, análise e comunicação de dados. Independente da representação, contanto que seja visual, e independente do que ela representa, contanto que seja uma informação, ela constitui-se uma Visualização de Dados. Os termos Visualização da Informação e Visualização Científica são subconjuntos de Visualização de Dados. Eles se referem a tipos particulares de representações visuais com propósitos específicos.

De acordo com Card et al (1999) existe diferença entre os termos Visualização de Informação e Visualização Científica. Visualização de Informação é o uso de um ambiente computacional interativo, que possibilita representação visual de dados abstratos para amplificar a cognição. Ambiente computacional interativo uma visualização é visualizada em um computador, onde a esta pode ser manipulada de forma simples, direta e livre, incluindo ações como filtro de dados, navegação de dados, etc. A representação visual da informação é utilizada para representar atributos de dados abstratos tais como localização, tamanho, forma, cor para formar a imagem que permita a visualização de padrões, tendências e exceções. Dados abstratos dizem respeito a informações de dados quantitativos, processos e relacionamentos entre os dados. E por fim, o conceito de cognição ampliada procura, através da visualização e interação com os dados, estender a habilidade do usuário compreender de forma mais fácil os dados e seus relacionamentos. A Visualização Científica é definida como representação visual de dados científicos que são geralmente de natureza física, em vez de abstratos (SPENCE, 2007). Por exemplo, uma imagem de ressonância magnética e uma imagem de raio-x são considerados Visualização Científica porque eles apresentam dados que possuem

forma física, buscando representar fielmente esta forma de uma maneira que seja fácil de visualizar, reconhecer e compreender.

A visualização, e suas diversas variantes, podem ser aplicadas em várias áreas do conhecimento, alguns exemplos(MAZZA , 2009):

- Análise de mercado financeiro;
- Sistemas de análise de estruturas de engenharia;
- Imagem de Ressonância Magnética para apoio a diagnósticos médicos;
- Análise de comportamento da população humana;
- Simulação de processos complexos;
- Análise da simulação de um sistema físico;
- Análise de movimentação motora do corpo humano para eficiência em esportes.

Em cada caso, a visualização fornece uma alternativa para suplementar a informação textual ou verbal, e tem como objetivo fornecer uma descrição mais rica de uma informação do que sua contraparte em texto.

2.2 Histórico de visualização

Talvez a primeira técnica de registro gráfico e representação da informação seja a utilizada pelo homem primitivo, quando da realização de pinturas nas paredes das cavernas. Um exemplo é a caverna Chauvet-Pont-d'Arc (Figura 1), localizada no sudoeste da França, com mais de 250 pinturas, feitas há mais de 30 mil anos.



Figura 1: Exemplos de pintura das cavernas (<http://www.lascaux.culture.fr/>).

O sistema de escrita mais antiga usava imagens para codificar símbolos e todas as palavras, este sistema é chamado de logogramas. A placa de calcário Kish (Figura 2) é considerada o documento escrito mais antigo. Outros sistemas de escrita antigos, como o Egípcio, surgiram posteriormente com maiores possibilidades de expressão.



Figura 2: Placa de calcário Kish da Mesopotâmia.

As primeiras formas de visualização surgiram em função da necessidade de viagens, comércio, religião e comunicação. Mapas fornecem suporte para viajantes, onde o planejamento e a sobrevivência eram os objetivos principais do uso dos mesmos. O mapa de Peutinger (Figura 3) (LENDERING, 2012) ou cartograma, foi um dos mapas rodoviários mais antigos representando aproximadamente 70.000 milhas de estradas de uma parte do Império Romano. No mapa há símbolos especiais que possibilitam identificar larguras de rios e florestas, ou distância aproximada das estradas, por exemplo. Um aspecto interessante deste mapa é que as distâncias estão distorcidas. A escala leste-oeste é maior que a escala norte-sul. Assim Roma aparece tão próxima a Cartagenas quanto Nápoles esta de Pompéia. Uma das razões para isso é que as posições relativas eram mais importantes que a precisão, e em alguns casos as distorções foram devido ao “meio de impressão” que estava sendo utilizado, rolo de papiro, mas largo do que alto.

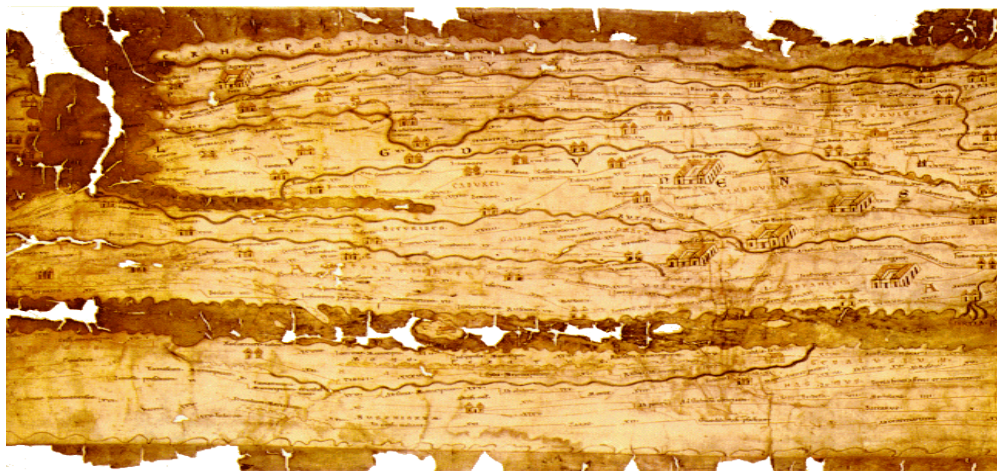


Figura 3: Mapa de Peutinger. (LENDERING, 2012)

Outros mapas de destaque na história da visualização são o mapa de John Snow, sobre uma epidemia de cólera em Londres, e o mapa de Minard sobre a marcha de Napoleão á Rússia, os dois serão apresentados a seguir.

A Figura 4 mostra uma parte do mapa de John Snow de Londres sobre mortes resultantes da cólera ocorrida em Londres em 1663. Cada item da pilha representa uma morte individual por cólera. Uma visão geral sob o mapa demonstra uma alta concentração de morte por cólera em torno da bomba de água da Broad Street. Com o fechamento da referida bomba de água a epidemia de cólera cessou, causando aproximadamente 500 mortes. Contudo, o mapa pode ser ainda melhor explorado com aspectos geográficos e temporais da doença, respondendo perguntas como: O que causou essa concentração ? Por que houve mortes longe do centro?

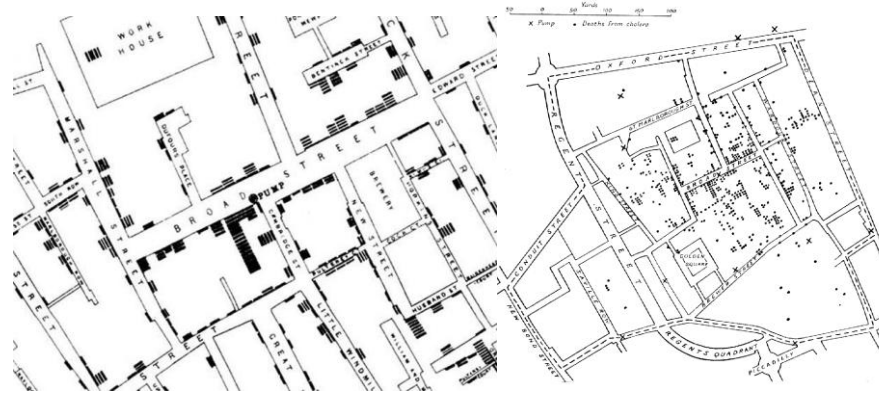


Figura 4: Mapa de John Snow de epidemia de cólera em Londres

A representação da marcha Napoleônica até a Rússia por Minard (Figura 5) se destaca por associar dados geográficos com dados de series temporais sobre uma representação estática. Este é um dos últimos mapas de Minard sobre marcha, cujo grande apelo esta na ênfase da grande perda de tropas do exército francês durante essa expedição de guerra, de 400.000 para 10.000. O mapa pode ser lido da seguinte forma: a largura da linha representa o tamanho do exercito naquela localização, a cor indica a direção da movimentação das tropas (ideia de fluxo), a temperatura é registrada em diferentes pontos do mapa associado com um gráfico de linha na parte inferior do mapa.

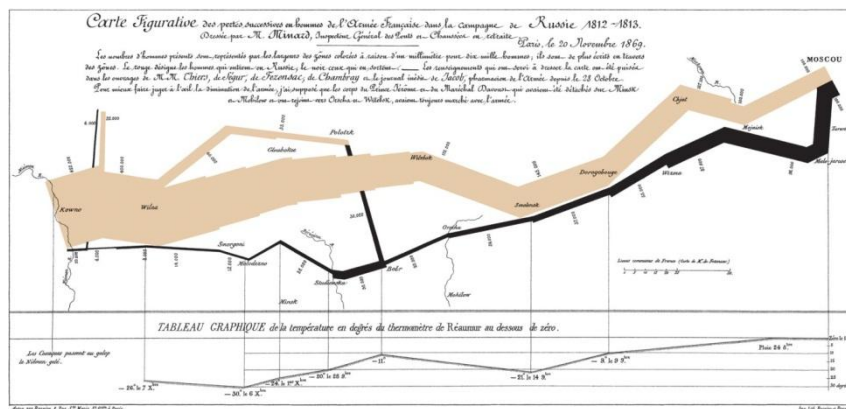
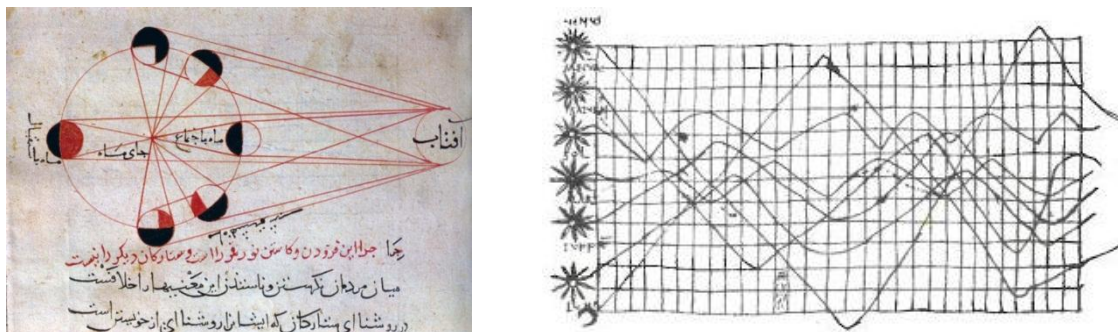


Figura 5: Mapa de Minard - Representação gráfica da marcha de Napoleão à Russia(SPENCE, 2007).

Em relação a visualizações voltadas a séries temporais, as primeiras a se destacarem foram as representação circular das fases da lua por Abu Rayhan Muhammad ibn Ahmad al Biruni, e representação de movimentação planetária, ambas do mesma época, por volta do ano 1030(Figura 6).

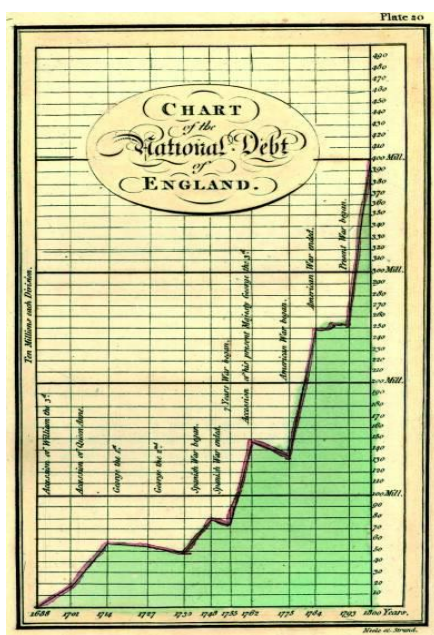


(a)

(b)

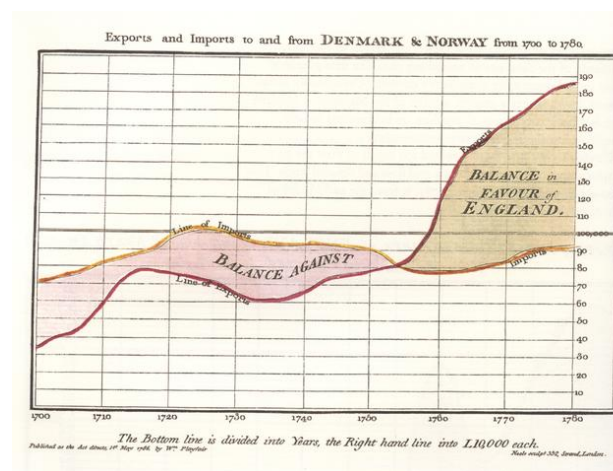
Figura 6: (a) Representação circular das fases da lua; (b) Movimento planetário.

Um avanço relevante para visualização da informação foi a representação abstrata para eixos, permitindo que outros dados pudessem ser utilizados como coordenadas. Sem a interpretação matemática abstrata, só seriam utilizadas interpretações geoespaciais planares. Assim pode-se ter densidade em um dos eixos e temperatura em outro. Por exemplo, a (Figura 7), desenvolvida por William Playfair, apresenta o debito de uma nação ao longo do tempo. Playfair é um dos pioneiros da área de visualização da informação. Outro exemplo de Playfair inclui o gráfico de série temporal sobre a balança comercial entre Inglaterra e Noruega/Dinamarca ao longo de alguns anos(Figura 8).



(a)

Figura 7: Playfair - Uso de valores abstratos nos eixos.



(b)

Figura 8: Gráfico de série temporal sobre a balança comercial.

Outro exemplo é a visualização Joseph Priestley sobre o tempo de vida de pessoas famosas (Figura 9). Priestley (1733-1804) foi um dos primeiros a tentar retratar eventos históricos graficamente. Esta imagem, de 1765, é uma amostra de visualização que apresenta os tempos de vida de 2000 pessoas famosas, a partir de 1200 AC a 1750 DC, estas pessoas foram categorizadas em: homens de saber e estadistas, Ele convencionou o uso para indicar as data de nascimento e mortes incertas.

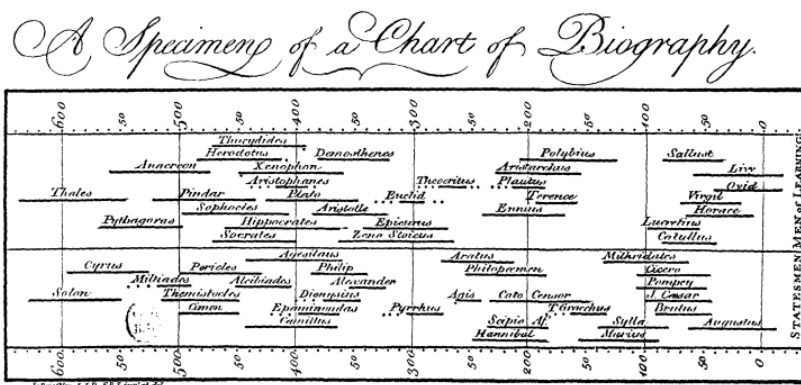


Figura 9: Visualização de eventos históricos.

Florence Nightingale, uma enfermeira inglesa, resolveu usar estatísticas sobre as mortes de soldados na guerra da Crimeia(1853-1856), que aconteceu no mar Negro, entre a Rússia e uma coligação envolvendo Inglaterra, França e Império Otomano. A ideia principal era a representação e comparação das mortes dos soldados por batalha com outras causas.O diagrama revelou que a maioria dos soldados morria nos leitos de hospitais, e não nos campos de batalha - eram 10 vezes mais mortes causadas por tifo, cólera e disenteria do que por ferimentos de batalha. A falta de ar fresco, luz e higiene nos hospitais provocava milhares de mortes desnecessárias. Era a primeira vez que se via fatalidades militares com números - e o diagrama era tão dramático que o governo inglês resolveu melhorar as condições sanitárias dos hospitais militares. E, assim, reduziu a mortalidade de soldados de 42% para 2,2%. O gráfico destaca 3 categorias: azul para mortes infecciosas evitáveis; em vermelho, ferimentos de batalhas; em preto, demais causas. Vendo que os soldados morriam mais nos leitos que em batalhas.

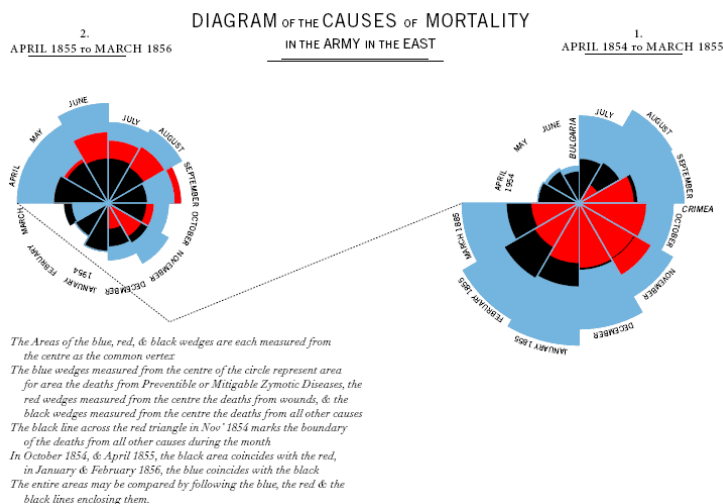


Figura 10: Visualização de Florence Nightingale envolvendo comportamento ao longo do tempo e comparações.

Visualizações médica foram bastantes populares, particularmente para treinamento de novos médicos. Existem inúmeros exemplos, contudo poucos são mais famosos os que impressionantes desenhos de Leonardo da Vinci sobre a anatomia humana (Figura 11).

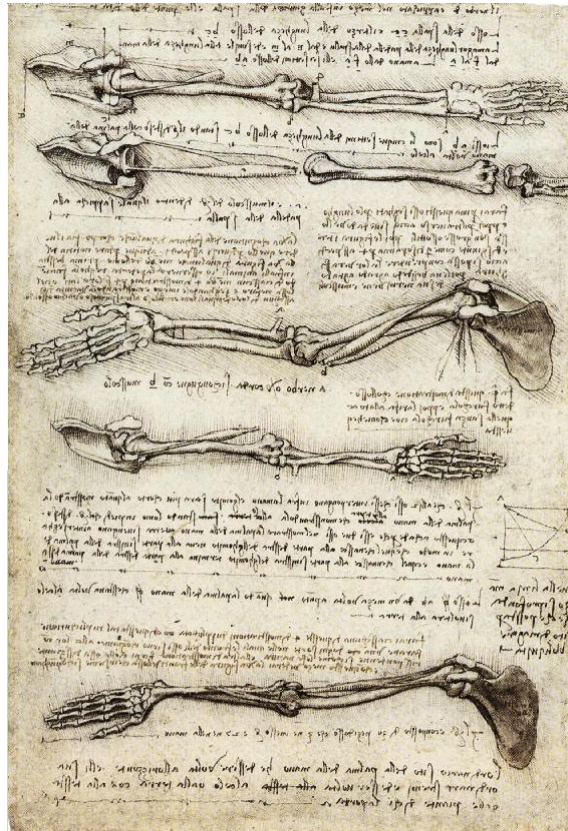


Figura 11: Imagens médicas de Leonardo da Vinci.

As visualizações mais modernas têm como objetivo fornecer diferentes de níveis de qualidade e quantidade de visões das informações a serem comunicadas. Por exemplo, a Figura 12 apresenta o mapa de metro de Londres.



Figura 12: Mapa do metro de Londres inspirado no diagramas de circuitos elétricos.

Em 1933, Harry Beck projetou o mapa de metrô de Londres (Figura 12). Antes do mapa do metro de Beck, várias linhas de metrô eram representadas geograficamente, muitas vezes se sobrepondo ao mapa das ruas (Figura 13). Beck percebeu que a localização geográfica era uma informação supérflua para os usuários de metrô, eles queriam saber apenas a ordem e relação das estações entre si, para decidir onde mudar

de estação. Inspirado na simplicidade de diagramas de engenharia elétrica, Beck projetou o mapa que seria o paradigma para todos os mapas de transporte público que vieram em seguida.

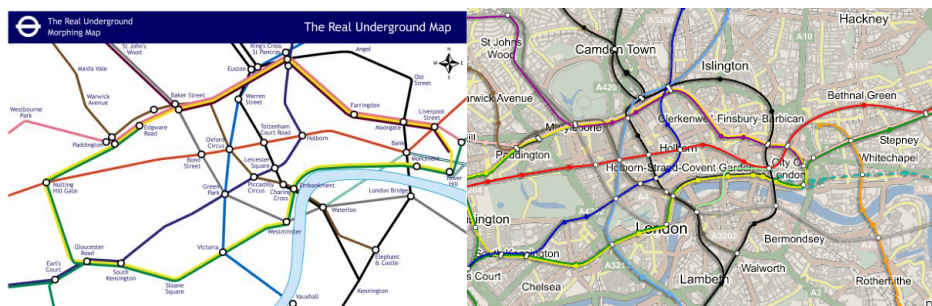


Figura 13: Representações geográficas reais do metro de Londres.

É possível utilizar visualização para comparar padrões e identificar anomalias, por exemplo, a Figura 14 (a) apresenta uma Eletrocardiograma normal, enquanto que a Figura 14 (b) apresenta um eletrocardiograma de um homem de 83 anos com hipertrofia do ventrículo esquerdo e arterial. O diagnóstico é baseado nas diferenças entre o eletrocardiograma normal e o de estudo para identificação de doenças cardiovasculares.

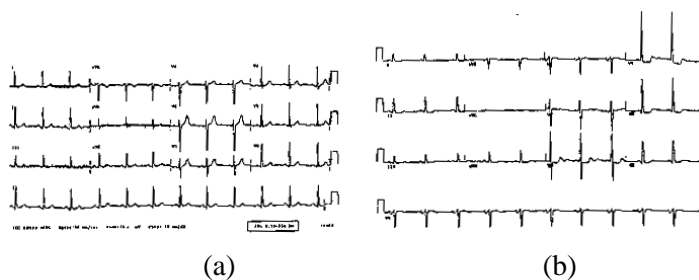


Figura 14: Exemplo de Eletrocardiograma: (a) normal – esquerda, (b) anormal - direita

As possibilidades de aplicações de visualizações são inúmeras, no dia a dia já é possível encontrar comumente mapas, grafos e gráficos simples, utilizados para as mais diversas tarefas, como por exemplo: traçado de roteiros, análises econômicas, estatísticas, tendências de mercado, etc. Os tipos de objetos que podem ser visualizados são os mais diversos: dados, algoritmos, resultados de cálculos, processos, entre outros. A ideia principal é que o usuário possa obter prontamente informação da representação visual, mas que ela seja interativa, gerando novas informações e percepções do objeto de estudo. A seguir serão apresentadas um conjunto de exemplos de visualizações modernas, suportadas por computador, de uma variedade de aplicações, incluindo reconstruções médicas, simulação aeronáutica, e bioinformática. Figura 15 apresenta os vasos sanguíneos sobrepostos a um crânio e um corpo de tubarão. A Figura 16 apresenta o fluxo de ar gerado por um jato durante a decolagem. Figura 17 apresenta um fluxo para análise de genes, suas ligações e comportamento, sobre determinadas condições de experimento (hipótese).

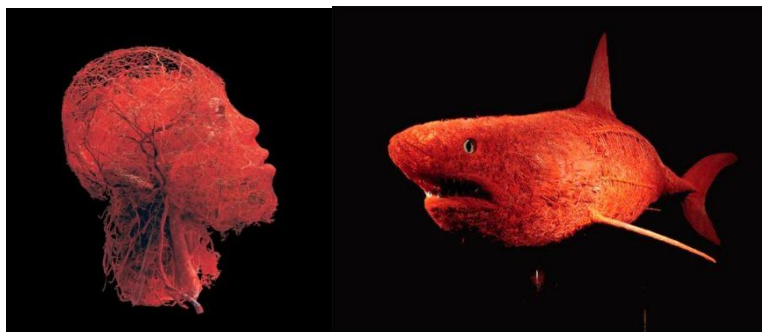


Figura 15: Representações de vasos sanguíneos (<http://www.bodywolds.com>).

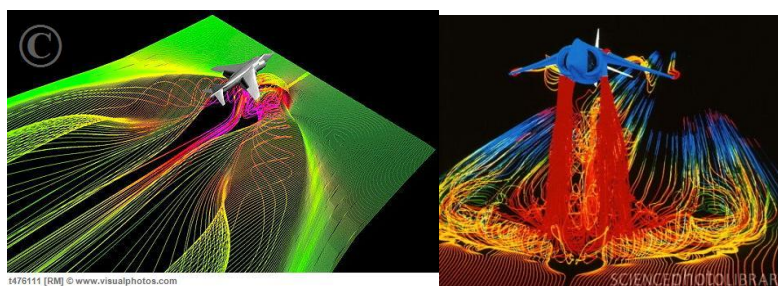


Figura 16: Estudo do fluxo de ar na decolagem de aeronaves. (<http://quest.nasa.gov/aero/background/tools>)

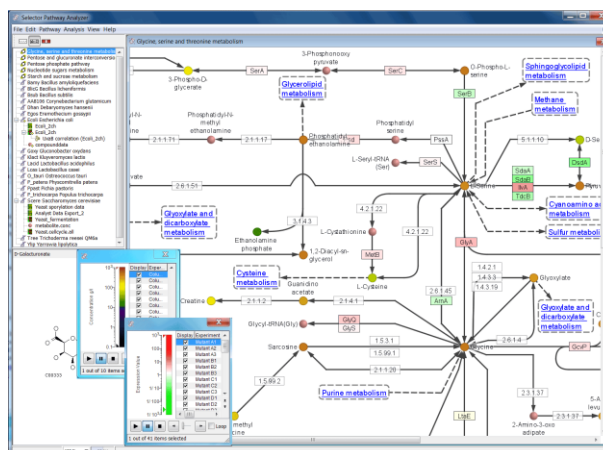


Figura 17: Ferramenta para estudo de genes, para mapeamento de relacionamentos e comportamentos dos mesmos. (<http://www.genedata.com/products/selector/modules.html>)

2.3 Processo de Visualização

Em um processo de visualização (Figura 18), normalmente o projetista concentra seus esforços iniciais nos tipos de dados que estão disponíveis para visualização e que tipo de informação o usuário espera extrair da visualização que será proposta. Os dados podem ter origem diversa, tais como: consultas a banco de dados relacional, arquivos XML (*Extended Markup Language*), arquivos TXT (texto) ou CSV (*Comma-Separated Values*), e podem apresentar estruturas simples ou complexas. O usuário deseja utilizar a visualização para explorar, procurar por algo interessante no conjunto de dados (*outliers*, padrões de grupos, ou tendências) para confirmar hipóteses, ou apresentar o resultado de uma análise para uma plateia.

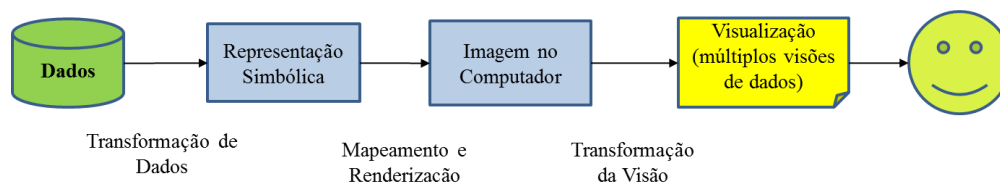


Figura 18: Etapas do processo de visualização. (WARD, GRINSTEIN e KEIM, 2010)

Um passo importante para efetiva visualização de dados é definição de como os dados serão representados visualmente, e como essas representações serão inseridas no contexto da técnica de visualização selecionada (Representação Simbólica). Os atributos dos dados são utilizados para definir os objetos gráficos que serão visualizados, tais como: pontos, linhas, e formas, bem como as características destes objetos, tais como: tamanho, posição, orientação e cor. Desta forma, considerando um gráfico X-Y e uma base de dados, pode-se imaginar um atributo numérico contínuo para eixo X, outro atributo numérico contínuo para o eixo Y, um atributo categórico para forma, outro atributo categórico para tamanho, e outro atributo categórico ou numérico para cor. Alguns autores apresentam outras formas de representação, tais como: textura, transparência, animação, som, etc. No final do processo ocorre a apresentação dos resultados ao usuário, a exibição de múltiplas visões de dados.

Outra importante etapa do processo de visualização é a definição dos controles de interação para visualização do conjunto de dados e seus relacionamentos. Os controles de interação são importantes porque o processo de visualização apoiado por computador é bastante dinâmico, na qual o usuário pode interferir em todas as etapas do processo, restringindo o conjunto de dados de análise, analisando em detalhes um item de dados específico, encontrar itens similares de dados, etc.

Não há fórmula para uma visualização bem sucedida, e uma visualização útil é dependente do usuário que a está utilizando. Por esse aspecto, é interessante que uma ferramenta de visualização seja flexível o suficiente para permitir adaptar-se aos diversos perfis de usuário: experientes, iniciantes, com mais ou menos habilidades no domínio do problema, até os mesmos se sentirem confortáveis a atingir seus objetos em relação a visualização proposta.

Frequentemente a visualização é parte de um processo maior, que pode ser análise exploratória de dados, descoberta de conhecimento, ou análise visual. Neste processo de descoberta, a preparação dos dados é importante, para evitar ou minimizar a manipulação ou análise de dados errôneos ou com ruídos. A visualização no processo de exploração de dados é usada para transmitir informação, descobrir novos conhecimentos, e identificar padrões, anomalias, tendências e relacionamento no conjunto de dados de análise.

O processo começa com dado e a geração de uma imagem, a visualização, ou um modelo via computador é tradicionalmente descrito como um pipeline, uma sequência de estágios que podem ser estudados independentemente em termos de algoritmos, estrutura de dados, e sistemas de coordenadas. Estes processos ou pipelines são diferentes para computação gráfica, visualização e descoberta de conhecimento, mas se sobrepõem em muitas partes.

2.3.1 Pipeline de Visualização

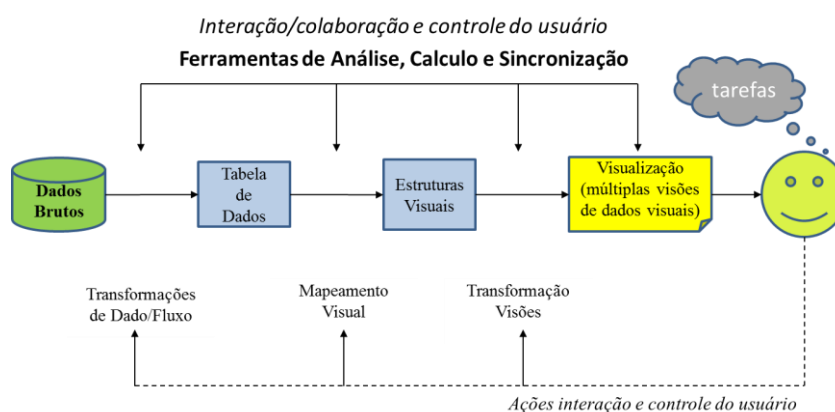


Figura 19: Pipeline de visualização(WARD, GRINSTEIN e KEIM, 2010)

A interação com uma ferramenta de visualização ocorre em diversos níveis e gera diversos produtos intermediários, os estágios do pipeline de visualização são os seguintes (Figura 19):

- Modelagem de dados: o dado para ser visualizado, de um arquivo ou banco de dados, tem que estar estruturado para facilitar a sua visualização. O nome, tipo, faixa de valor, e a semântica para cada atributo ou campo de um dado gravado devem estar disponíveis no formato que garanta rápido acesso e fácil modificação.
- Seleção de Dados: a seleção de dados envolve identificar um subconjunto dos dados potências que serão visualizados. Isto pode ocorrer com total controle do usuário ou via métodos algorítmicos, por exemplo, detectar automaticamente características de interesse do usuário.
- Mapeamento Visual do dado: um dos estágios mais importantes do pipeline de visualização é execução do mapeamento visual do dado gravado, que pode ser mapeado para característica de tamanho, posição, cor, forma, etc. Esse mapeamento envolve a aplicação de técnicas de computação gráfica, tais como: transformações de escala, rotação, translação, etc.
- Configuração de parâmetros da Cena (Transformação da Visão): como um gráfico tradicional, o usuário deve especificar vários atributos da visualização que são relativamente independentes dos dados. Alguns exemplos: mapeamento de cor (para muitos domínios de problema, certas cores já estão claramente definidas), mapeamento de som (quando houver informações a serem transmitidas por canal de áudio, por exemplo, a completude correta de uma tarefa), especificações de luz (para visualizações tridimensionais), etc.
- Renderização ou geração da visualização: a projeção específica ou renderização da visualização varia de acordo com o mapeamento que esta sendo utilizado, que podem envolver técnicas de mapeamento de sombra e textura, embora muitas técnicas de visualização necessitam somente desenhar linhas e polígonos com sombreado uniforme. Além disso, a maioria das visualizações incluem

informações complementares para facilitar a interpretação dos dados, tais como: eixos, anotações, imagens, etc.

2.4 Papel da Percepção e Papel do Usuário

Em qualquer visualização, um aspecto crítico relacionado ao usuário é seu sistema visão, com suas habilidades e limitações. Se a meta de uma visualização é a precisão na transmissão da informação com imagens, então é essencial que as habilidades para percepção sejam consideradas. Uma imagem pode estimular o usuário em várias ideias, mas se o objetivo é apresentar uma conclusão, não pode haver ambiguidades, como o elefante de muitas pernas da Figura 20.

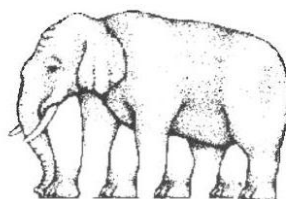


Figura 20: Exemplo de ambiguidade em uma imagem.

Usuário interage com visualizações de acordo com o que ele vê e interpreta. Entender como os usuários veem deveria ajudar a produzir melhores visualizações, ou no mínimo evitar visualizações ruins. O cérebro humano processa a entrada visual de forma paralela e contínua. Textura, cor e movimento são exemplos de atributos primitivos que os seres humanos conseguem perceber. O sistema de percepção humana recebe as entradas e as processa de várias maneiras. O primeiro processo é a pré-atenção, um sistema de altodesempenho que identifica rapidamente diferenças nas cores e textura. Há outras características que o sistema de percepção humana trata, tais como orientação de linhas, tamanho, largura, e tamanho de um objeto, curvatura, agrupamento e movimento. As Figura 21(a) e Figura 21(b) apresentam exemplos de fácil identificação, utilizando diferenciação de cor e orientação. A Figura 21(c) apresenta um exemplo que não faz uso do processo de pré-atenção, por exemplo, para identificar um “R” correto.

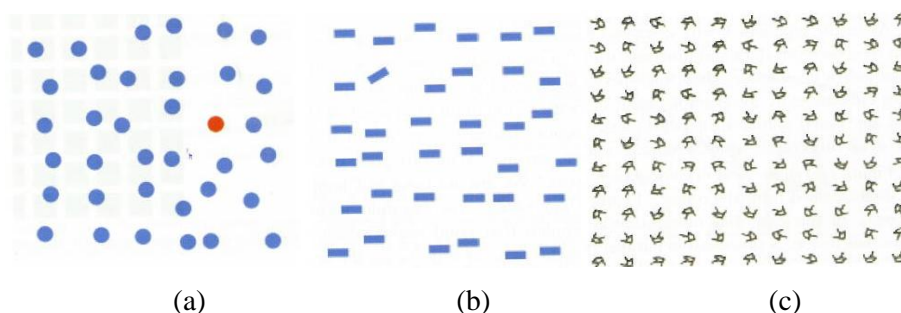


Figura 21: identificação rápida por (a) cor, (b) orientação, e (c) exemplo de não uso de primitivas de pré-atenção.

Um trabalho pioneiro no relacionamento do uso de atributos visuais para utilizar o sistema de percepção humana de forma mais eficiente para representações gráficas 2D foi realizado por Jacques Bertin (BERTIN, 1983), e definiu diretrizes para seu uso.

As variáveis visuais definidas por Bertin são:

Duas dimensões planas (x, y): esta é a superfície onde a informação é gerada. A informação pode ser representada em qualquer dos três tipos:

- Ponto: representa uma localização no plano que não tem um comprimento teórico ou área;
- Linha: tem comprimento mensurável no plano, mas nenhuma área;
- Área: significa tamanho mensurável. O significado se aplica a toda a área coberta pela marca (ícone, desenho, etc).

Além das dimensões planares, a marca no plano também por transmitir informações por:

- Tamanho: tal como a altura de uma coluna de área, de um sinal, o número de sinais de igual;
- Brilho (valor): os vários graus entre o preto eo branco;
- Textura: a variação dos itens que compõe a textura, tais como largura, tamanho, composição, etc, de uma área com um determinado valor;
- Cor (saturação): as variações dentro de uma mesma saturação;
- Orientação: a orientação das linhas ou padrões que variam da vertical para a horizontal;
- Forma: variações na forma com uma área constante (por exemplo, quadrado, círculo, estrela, etc.).

Bertin, em seguida, analisa mapeamentos de informações para a representação visual de quatro tipos diferentes de percepções (perguntas):

- Quantitativa: identificar a relação entre dois elementos em uma dimensão particular, ou todos os elementos do grupo com quantidades semelhantes;
- Seletiva: é para identificar todos os elementos de uma determinada categoria;
- Ordenada: é para identificar a ordenação de uma dimensão particular.

Bertin considera os mapeamentos de cada um dos elementos visuais, para cada tipo de percepção de cada tipo de representação planar; identificar combinações que são eficazes daqueles que não são. E resume seus resultados em uma tabela de mapeamentos de eficácia entre as percepções e as representações visuais(BERTIN, 1983).

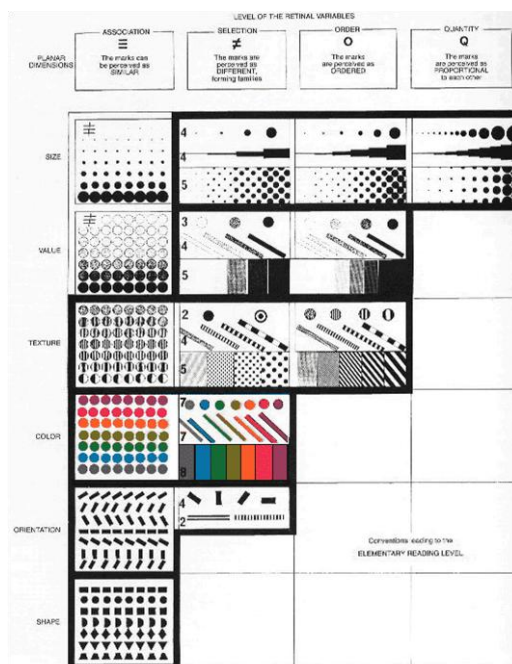


Figura 22: Recomendações de Bertin para utilização de atributos visuais bidimensionais

Apesar de não haver diretrizes padronizadas para a construção de visualização eficiente, algumas regras são adotadas para minimizar ou resolver esses problemas. Segundo Brath (1999), um gráfico de informação eficiente deve ter as seguintes metas:

- Induzir o espectador a pensar no que é mais importante,
- Apresentar muitos números em um pequeno espaço,
- Fazer com que grande quantidade de dados se torne coerente,
- Encorajar comparações de diferentes áreas dos dados,
- Revelar dados em vários níveis de detalhe.

2.5 Tipos de Técnicas de Visualização de Informação

O formato visual com que as informações são disponibilizadas ao usuário tem influência direta na tarefa de extração do conhecimento de um determinado sistema de informação, e é dependente dos dados. Shneiderman (1996) classificou os dados em 7 (sete) tipos diferentes, e para cada um deles pode-se descrever uma visualização diferente, são eles:

- 1D: Tipo de dado representado por textos;
- 2D: Este tipo de dado é usado para representar mapas geográficos, plantas de engenharia, plano cartesiano e etc;
- 3D: Este tipo de dado possui as características dos dados bidimensionais incorporando a informação de volume;
- Multidimensional: Dados multidimensionais descrevem um item com mais de três atributos. Técnicas de consultas dinâmicas e gráficos de dispersão são utilizados para auxiliar a visualização desse tipo de dado.
- Temporal: Tipo de dado onde o atributo tempo é incorporado aos tipos anteriormente descritos;

- Hierárquico: Tipo de dado que possui uma estrutura hierárquica onde cada item tem um nó pai;
- Rede de dados: são nós conectados por links previamente definidos. Esses links podem ser organizados em árvores ou em hierarquias, e a melhor maneira de manipulação é permitindo mudar o foco sobre os nós.

Shneiderman(1996) sugeriu as características de uma boa ferramenta de visualização são definidas em função das ações que o usuário poderá realizar na ferramenta, sendo elas:

- Visão geral: o usuário precisa ganhar uma noção sobre todos os dados que serão analisados. Esta noção está baseada nos parâmetros que o usuário escolheu para a visualização, nos limites do dispositivo gráfico usado e de sua percepção. Os atributos gráficos mais usados são posição, cor, tipo de representação e tamanho.
- Zoom: a técnica de zoom é importante porque permite focar em um certo subconjunto dos dados para análise, ou seja, analisar um determinado contexto. Além disso, conforme se vai aplicando o zoom, mais detalhes sobre uma determinada visão dos dados são mostrados, o que se chama de zoom semântico.
- Filtro: usuários freqüentemente precisam reduzir o tamanho do conjunto de dados, eliminando itens com base em seus atributos. Uma das maneiras mais eficientes é o uso de Consultas Dinâmicas, que permitem a realização de consultas em uma base de dados sem a necessidade de linhas de comandos, possibilitando inclusive a descoberta e formulação de hipóteses sobre os dados da base de dados.
- Detalhes sob demanda: quando os usuários estão explorando um conjunto de dados, eles necessitarão ver detalhes sobre um item em particular. Isto é normalmente feito usando o click do mouse. As informações adicionais podem aparecer em uma janela auxiliar, ou na própria visão dos dados.

Adicionalmente, podem-se incluir mais duas características (CARR, 1999):

- Relacionamento: se o usuário descobre um item de interesse, ele pode precisar saber sobre outros itens com atributos similares, a ferramenta então poderia apontar esses itens similares.
- Histórico: o usuário precisa de suporte para desfazer uma ação, mostrar os passos percorridos até aquele ponto.

2.5.1 Técnicas de Visualização de Informação

Algumas técnicas serão mencionadas ao longo do texto e aparecem com maior frequência, entre elas destacam-se: Dispersão de Dados (gráfico X-Y), Coordenadas Paralelas e Treemap.

2.5.1.1 Dispersão de Dados

A técnica de dispersão apresentada, também chamada de Gráfico de Dispersão 2D, possibilita visualizar uma relação (correlação) entre duas variáveis X e Y (por exemplo, peso e altura). São representados pontos de dados individuais no espaço bidimensional onde os eixos representam as variáveis (X no eixo horizontal e Y no eixo vertical). A técnica de Dispersão de Dados possibilita uma análise multidimensional a medida que são disponibilizadas em uma ferramenta de visualização de informação configurações de eixos, forma, tamanho, cor, etc.

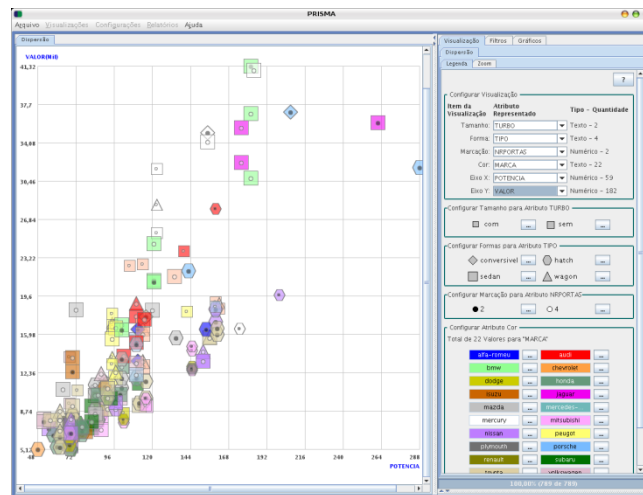


Figura 23: Exemplo de Técnica de Dispersão de Dados multidimensional.

2.5.1.2 Coordenadas Paralelas

Considere a representação dos quatro pontos tridimensionais $(3, 2, 0)$, $(4, 1, 4)$, $(7, 5, 3)$, $(6, 1, 5)$ no plano cartesiano, conforme mostra a Figura 24. Em coordenadas paralelas, têm-se os eixos X, Y e Z paralelos (Figura 25). Dessa forma, cada ponto do plano cartesiano torna-se uma linha, conforme a Figura 40.

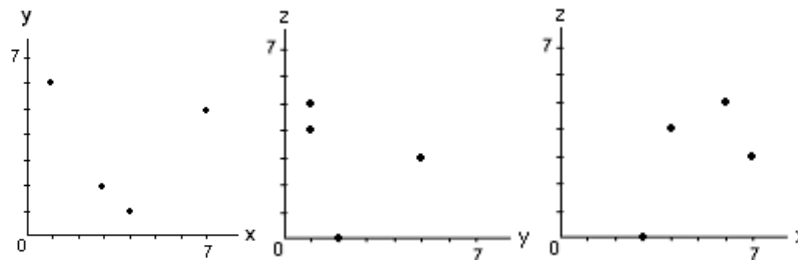


Figura 24: Planos cartesianos X-Y, Y-Z e X-Z representando os pontos $(3, 2, 0)$, $(4, 1, 4)$, $(7, 5, 3)$, $(6, 1, 5)$

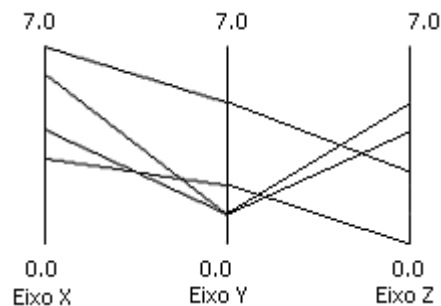


Figura 25: Pontos representados através de Coordenadas Paralelas

A técnica consiste em interligar os eixos aos seus adjacentes através de linhas retas (Figura 26). Essas linhas são traçadas de acordo com os registros da base de dados. Cada eixo representa uma escala que contém todos os valores possíveis para um determinado atributo (INSELBERG, 2009).

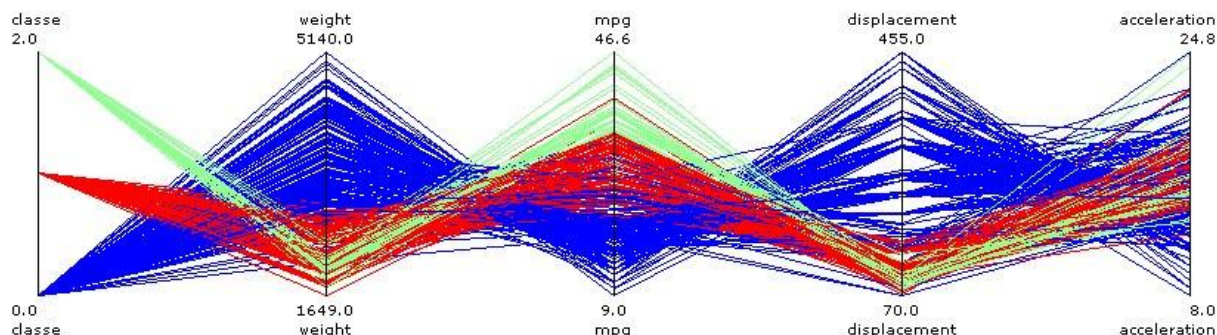


Figura 26: Exemplo de aplicação da técnica de Coordenadas Paralelas

2.5.1.3 Treemap

Treemaps são métodos de visualização de preenchimento de espaço, capazes de representar grandes coleções hierárquicas de dados quantitativos (SHNEIDERMAN e WATTENBERG, 2001) (SHNEIDERMAN, 2009). A função de um Treemap basicamente é dividir um espaço em vários retângulos, onde a área de cada retângulo se relaciona ao valor de um atributo em um registro específico de um conjunto de dados. Portanto, quanto maior o valor deste atributo maior será a área do retângulo que representa um item de dado ou um conjunto de dados. Os algoritmos utilizados para a criação das sequências de retângulos no espaço de preenchimento são cluster e squarified.

Os algoritmos cluster e squarified têm um apelo visual muito bom, em parte por causa dos retângulos que devem ficar o mais parecido possível com quadrados, evitando assim os retângulos finos (com alto aspecto relacional entre largura e altura) que dificultam visualizações, comparações e seleções. Por outro lado, o apelo visual sacrifica a ordem dos nós ou elementos. Além disso, caso haja uma alteração no tamanho dos nós, a posição dos retângulos no layout pode mudar drasticamente, causando desconforto e confusão aos usuários que interagem com uma aplicação (Figura 28).

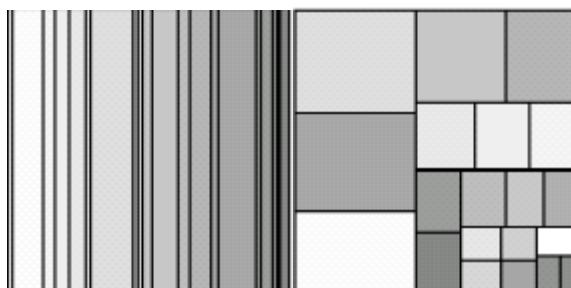


Figura 27: Treemap com alto aspecto relacional e Treemap com baixo aspecto relacional - cluster e squarified respectivamente

Além de ser uma técnica de ocupação de espaços, o Treemap possibilita a organização dos dados hierarquicamente (Figura 28), e desta forma, permitindo realizar comparações entre grupos de itens.



Figura 28: Organização hierárquica dos dados no Treemap

2.6 Múltiplas Visões Coordenadas

Sistemas de múltiplas visões usam duas ou mais representações visuais distintas para auxiliar o processo de investigação de uma única entidade conceitual (BALDONADO, KUCHINSKY e WOODRUFF, 2000). Uma visão é considerada distinta das outras se permitir ao usuário aprender sobre diferentes aspectos da entidade conceitual, ou pela apresentação de informações diferentes, ou enfatizando diferentes aspectos da mesma informação, por exemplo, utilizando representações diferentes ou técnicas de visualização diferentes(Figura 29).

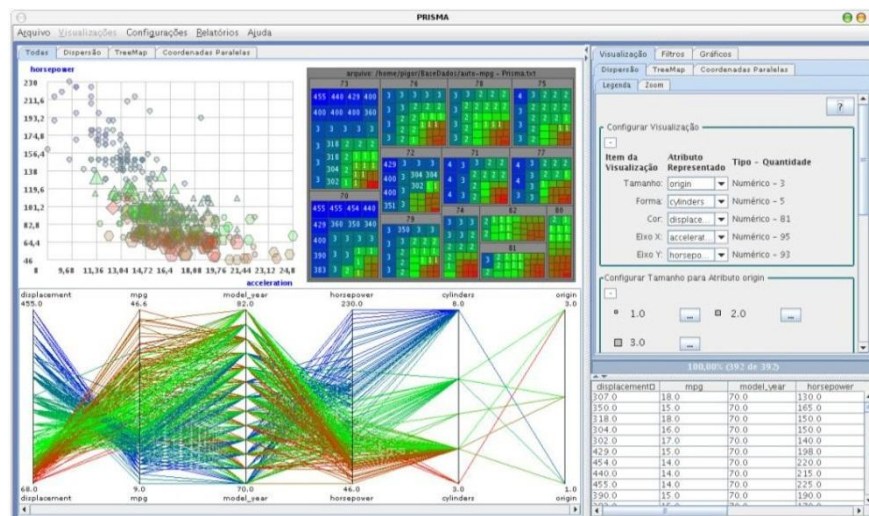


Figura 29: ferramenta PRISMA, exemplo de utilização de múltiplas visões de dados.

De acordo com North et al (2000) e Baldonado et al (2000), o uso de sistemas de múltiplas visões coordenadas apresenta algumas vantagens na análise, entre elas destacam-se: melhora do desempenho do usuário na percepção dos dados, facilita a descoberta de relacionamentos não triviais entre os dados, minimiza o overhead cognitivo de uma única visão ou de uma visão mais complexa, entre outras.

Os sistemas de visualização de informação que utilizam múltiplas visões coordenadas podem ser classificados por níveis de flexibilidade em relação aos dados, visões e coordenação.

- Dados: usuários podem utilizar diferentes conjunto de dados em suas visualizações.
- Visões: usuários podem escolher diferentes conjuntos de visualização para determinado conjunto de dados.
- Coordenação: usuário poderá escolher diferentes tipos de coordenação entre pares de visões para auxiliar sua necessidade de exploração dos relacionamentos entre os dados.

Para o desenvolvimento de sistemas de visualização de informação com múltiplas visões coordenadas, as recomendações mais frequentes são (BALDONADO, KUCHINSKY e WOODRUFF, 2000):

- Quando há uma diversidade de atributos, modelos, perfis de usuário, níveis de abstração ou gênero.
- Quando as visões diferentes destacam correlações ou disparidades.
- Quando há necessidade de diminuir a complexidade do conjunto de dados, utilizando múltiplas visões mais simples.
- Usar múltiplas visões minimamente, justificar o uso de múltiplas visões versus custo de aprendizado do usuário e espaço de visualização.

Pillat et al (2005) destaca como principais possibilidades de coordenação de múltiplas visões:

- Seleção: Itens de dados selecionados em uma visão são destacados em outras visões.
- Filtro: Reduzir o conjunto de dados para análise em todas as visões.
- Cor, Transparência e Tamanho: Características visuais para representar a variação de valores de um dado atributo dos dados em todas as visões.
- Ordenação: Valores de um atributo definem a ordem das representações visuais dos dados.
- Rótulo: Determina que conteúdo os rótulos exibirão para cada item de dados das visões.
- Manipulação de Atributos: Permite ao usuário adicionar / remover atributos das visões de dados.

Dado o contexto de múltiplas visões coordenadas, destacam-se os principais desafios no desenvolvimento de sistemas de múltiplas visões coordenadas (PILLAT,

VALIATI e FREITAS, 2005)(BALDONADO, KUCHINSKY e WOODRUFF, 2000)(NORTH e SHNEIDERMAN, 2000):

- Os mecanismos de coordenação.
- Requisitos computacionais para renderização das visões.
- Disposição da interface – layout, com espaço normalmente muito reduzido para novas visões.
- Interação do usuário entre as diversas formas de visualização.
- Aspectos cognitivos relacionados ao uso de sistemas de múltiplas visões coordenadas:
 - Tempo e esforço necessário para o aprendizado do sistema.
 - Sobrecarga de informações na memória de trabalho do usuário.
 - Esforço necessário para comparação.
 - Esforço necessário para troca de contexto.

2.7 PRISMA

O PRISMA (Figura 30) é uma ferramenta de visualização de informação baseada em múltiplas visões coordenadas para explorar dados multidimensionais usando as técnicas: treemap, dispersão de dados e coordenadas paralelas. As principais características do PRISMA serão apresentadas, pois o resultado deste projeto de dissertação é uma evolução e variante do PRISMA para manipular simultaneamente múltiplas bases de dados.

Principais aspectos que guiaram a concepção do Prisma (GODINHO, MEIGUINS, *et al.*, 2007):

- Suporte para requisições de qualquer base de dados ou técnica de visualização.
- Desenvolvimento de uma interface fácil e usável para tarefas de usuário.
- Suporte para visualização e interação com dados em mais de uma técnica de visualização, simultaneamente ou individualmente.
- Mecanismos de coordenação entre as técnicas de visualização de informação desenvolvidas e os mecanismos de interação de dados.
- Implementação de três técnicas de visualização: dispersão de dados, treemap e coordenadas paralelas;
- Portabilidade para diferentes plataformas provida pelo uso do JAVA como linguagem de programação;
- Suporte para diferentes recursos de dados, a partir de arquivos de texto pré-formatados para arquivos XML e base de dados relacional.

O PRISMA nesta versão (Figura 30) possui quatro guias, três com cada uma das técnicas e uma com as três técnicas na mesma tela. A direita da tela se localizam as guias de configuração das técnicas e a tabela de dados, as quais permitem redimensionamento.

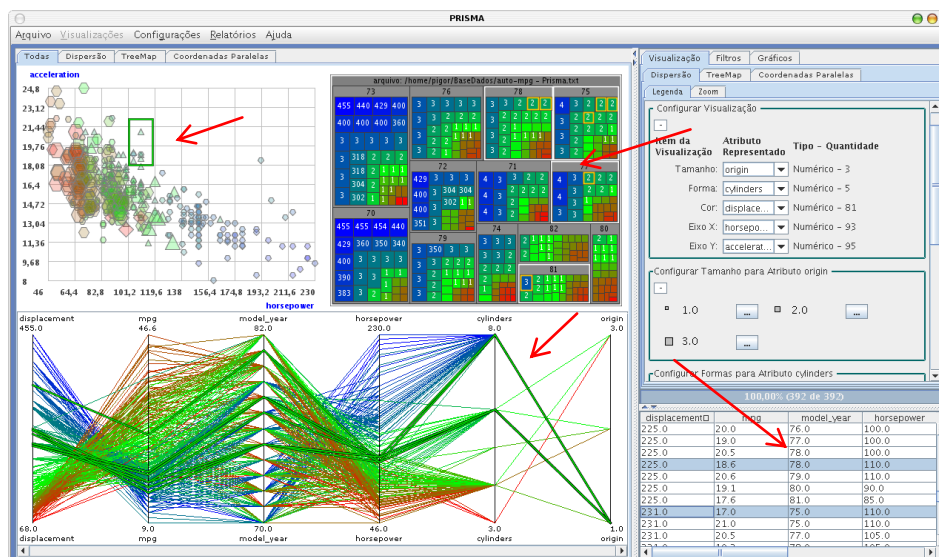


Figura 30: PRISMA com exemplo de coordenação entre as diversas visualizações.

A coordenação no PRISMA (Figura 30)(Figura 31) permite analisar os dados simultaneamente e individualmente. O atributo cor, por exemplo, é coordenado para todas as visões, já o atributo tamanho é coordenado somente para as técnicas de dispersão e treemap, entretanto deve ser especificado em cada visão. O filtro, a seleção e os detalhes-sob-demanda são outras tarefas coordenadas do PRISMA para todas as técnicas.

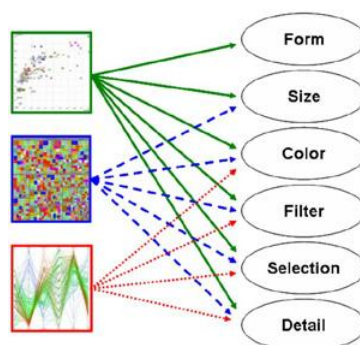


Figura 31: Esquema de coordenação no PRISMA.

O PRISMA utiliza dois tipos de mecanismos de filtro (Figura 32) para selecionar os itens de dados selecionados nas visões: RangeSlider e Checkbox. Os filtros são construídos de acordo com as características de cada atributo. A interface de filtro é dinâmica. O usuário seleciona o atributo a ser filtrado. Então o filtro do atributo correspondente surge na interface. A interface enfatiza em vermelho os atributos que estão ativos.

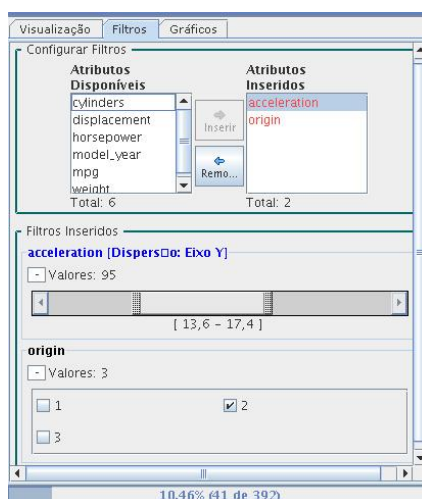


Figura 32: Tipos de filtro no PRISMA.

Os detalhes-sob-demanda (Figura 33) são configurados para todas as visualizações. O usuário seleciona os atributos que serão apresentados como informação adicional sobre cada item de dado usando o menu de opção correspondente.

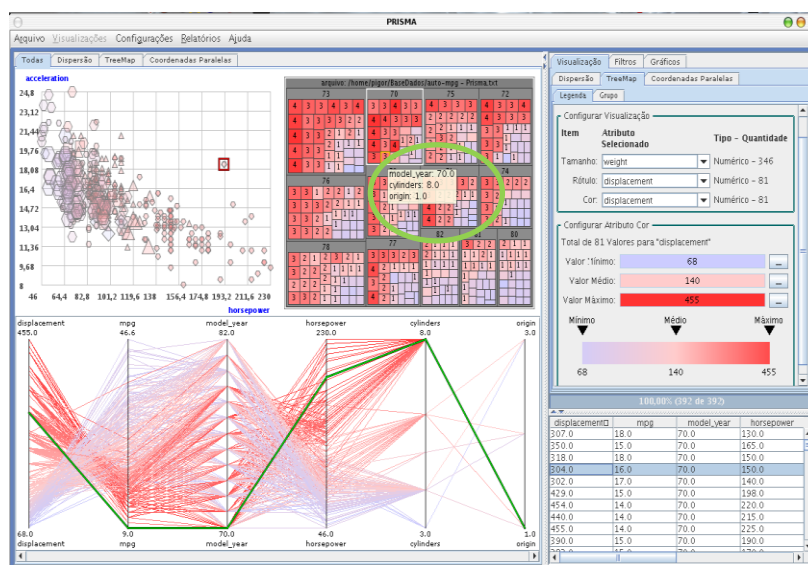


Figura 33: PRISMA - detalhes sob demanda.

O prisma também apresenta gráficos de pizza, barra e linha para apoiar a análise de dados. Os gráficos são coordenados com as visões, ou seja, qualquer filtro aplicado nas visões atualiza os gráficos.

3 TRABALHOS RELACIONADOS

Neste capítulo são apresentadas algumas ferramentas com funcionalidades semelhantes as propostas neste trabalho.

3.1 Ferramentas com Coordenação Flexível

Nesta seção são apresentados trabalhos relacionados que possuem mecanismos de coordenação flexíveis, ou seja, a coordenação pode ser configurada pelo usuário.

3.1.1 Snap-Together

Snap-Together (NORTH e SHNEIDERMAN, 2000) é uma ferramenta de visualização de informação que permite a criação de visualizações e coordenações de forma rápida e dinâmica sem a necessidade de programação (Figura 34). A ferramenta é flexível em dados, visualizações e coordenações. Além disso, é baseada no modelo de dados relacional.

Para explorar uma base de dados, o usuário primeiro deve carregar as relações (tabelas ou consultas) em visualizações. Em seguida, constrói as coordenações especificando ações para acoplar fortemente as visualizações. As ações de coordenação possíveis são seleção, navegação (*zoom* e *scroll*), destaque e carregamento (carrega uma tabela baseada na chave estrangeira de outra), que estão disponíveis dependendo do tipo de visualização escolhida. Apesar dessa flexibilidade, a ferramenta foi desenvolvida para desktop e é limitada a consultas SQL (*Structured Query Language*) via JDBC (*Java Database Connectivity*).

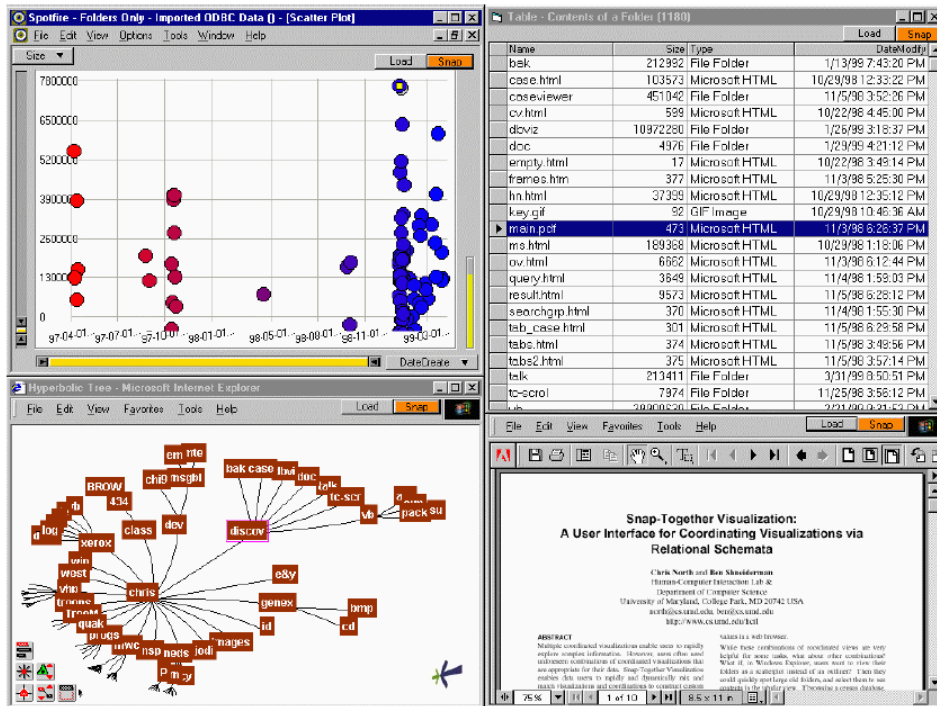


Figura 34 Ferramenta Snap-Together (NORTH, 2000)

3.1.2 Improvise

Improvise (WEAVER, 2004) é um ambiente que permite usuários construírem e explorarem visualizações de dados relacionais com múltiplas visões. Possibilita ao usuário carregar dados, criar visões, especificar abstrações visuais e estabelecer coordenações interativamente (Figura 35).

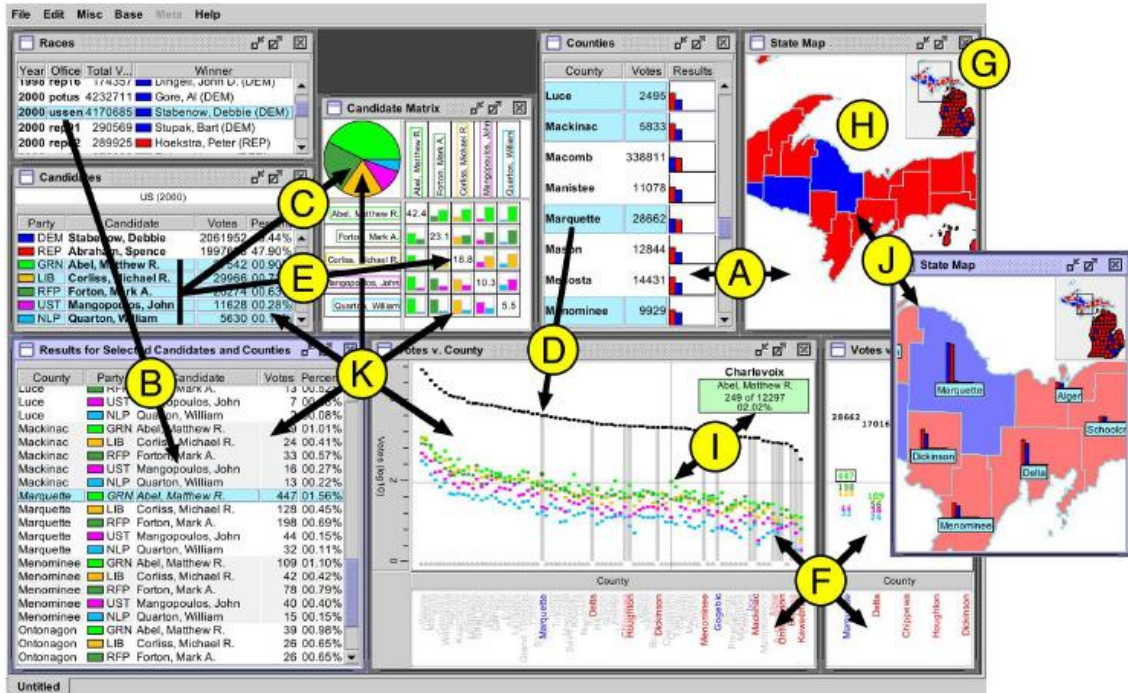


Figura 35 Ferramenta Improvise (WEAVER, 2006)

A ferramenta utiliza um mecanismo de coordenação baseado em compartilhamento de objetos simples (*live properties*) acoplados com uma linguagem de abstração visual baseada em expressões (*coordinated queries*). Essa combinação possibilita um melhor controle sobre a dependência entre a aparência e comportamento das visões, porém a ferramenta se torna muito complexa para um usuário final, necessitando ser operada por um especialista. Como o Snap-Together, o *Improvise* também foi escrito para plataforma desktop.

3.2 Ferramentas com Coordenação Não-Flexível

Nesta seção são apresentados trabalhos relacionados que não possuem mecanismos de coordenação flexíveis, ou seja, a coordenação não pode ser configurada pelo usuário.

3.2.1 VisGets

VisGets (DÖRK, CARPENDALE, *et al.*, 2008) propõe a exploração visual de informação na web. A ferramenta constitui-se de pequenos widgets (componentes acopláveis) de visualização de informação cuja manipulação gera uma consulta web (Figura 36). O objetivo é oferecer suporte ao usuário na tarefa de obter insight sobre a vasta quantidade de recursos e informações da web e um meio de filtrar visualmente os dados, através de consultas de busca dinâmicas, que combinam filtros de mais de um atributo. A ferramenta focou a interação em três atributos comuns que são mais difundidos: tempo, espaço e tópico (tag).

O VisGets estendeu e ajustou o conceito básico de consultas dinâmicas para aplicar em dados distribuídos e dinâmicos da web. Possui coordenação através do uso do mouse, que pode destacar itens relacionados, ou através das barras de seleção (*range sliders*), que são utilizadas para filtrar intervalos temporais. O VisGets utiliza a própria Web como fonte de dados para produzir as visualizações, o que o torna uma aplicação bastante específica.

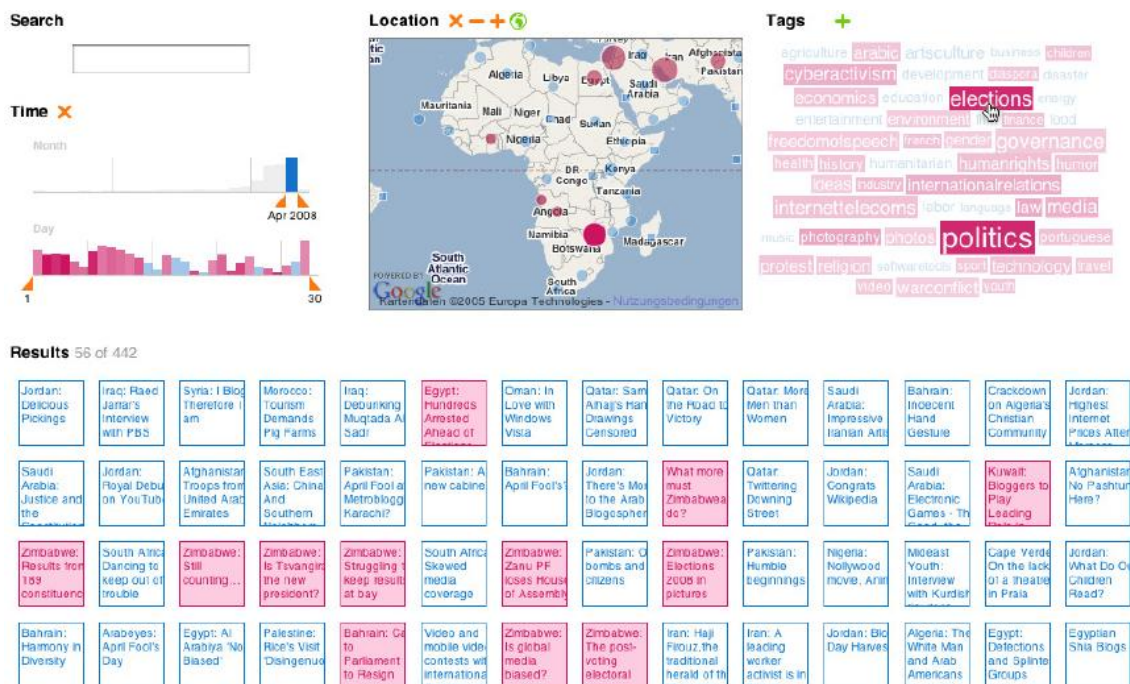


Figura 36 Ferramenta VisGets (DÖRK, CARPENDALE, *et al.*, 2008)

3.2.2 HotMap

Assim como o VisGets, o HotMap (HOEBER e YANG, 2006) também é voltado para exploração visual de resultados da web, através do sistema de busca do Google™. Possui duas visões coordenadas entre si: visão dos resultados detalhados, mostrando 20 resultados por vez e um mapa que exibe a visão geral compacta dos 100 primeiros resultados da busca (Figura 37).

A coordenação é possível através da visão geral compacta (região direita da Figura 37), que permite descobrir em que parte dos resultados se está – por exemplo, se no início dos resultados, ou se na metade – através do retângulo delimitado presente nessa visão. Além disso, ao clicar em um item da janela de detalhes, o retângulo automaticamente é deslocado para a região correspondente da busca no mapa da visão geral. Como no VisGets, a ferramenta é específica para resultados de buscas web, não sendo flexível para o usuário na escolha de outro tipo de dados.

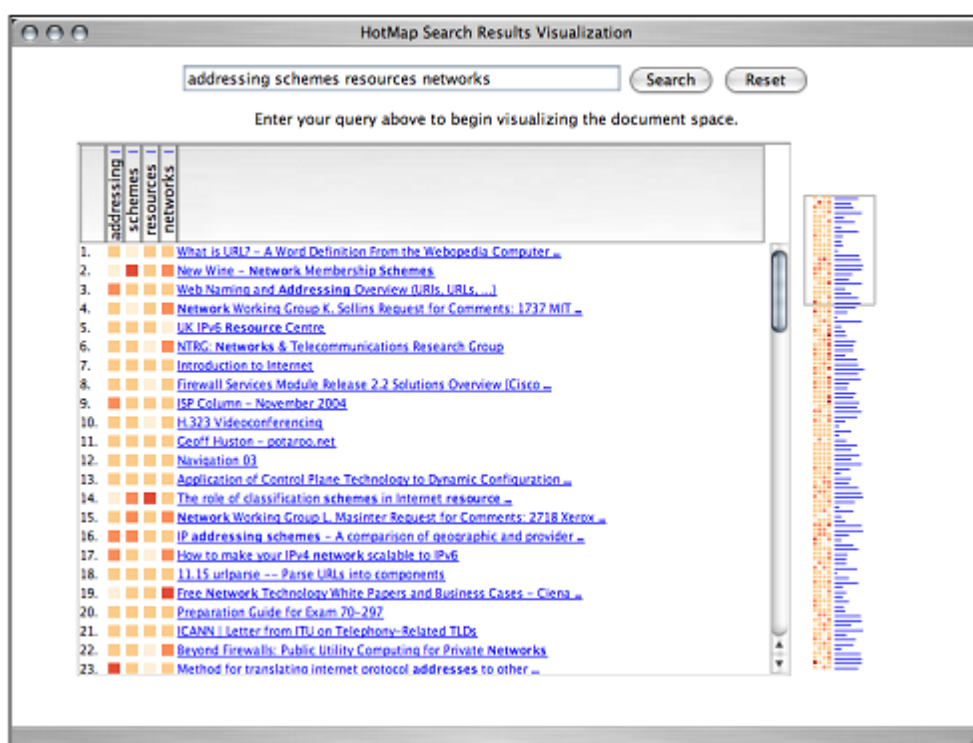


Figura 37 Ferramenta HotMap (HOEBER e YANG, 2006)

3.2.3 Prisma

O PRISMA (GODINHO, MEIGUINS, *et al.*, 2007) é uma ferramenta de visualização de informação com múltiplas visões coordenadas. Disponibiliza três técnicas de visualização: dispersão de dados, treemap e coordenadas paralelas, conforme ilustra a Figura 38. A ferramenta foi desenvolvida totalmente em Java, dando total portabilidade aos sistemas que executam Java. Além disso, também possui recursos importantes para uma boa ferramenta de VI, como seleção, filtros dinâmicos, zoom, configurações dos atributos, gráficos estatísticos, relatórios customizados, acesso a fontes de dados em arquivos texto pré-formatados e bancos de dados, dentre outros. O Prisma foi o trabalho relacionado que contribuiu diretamente para essa dissertação, uma vez que serviu de ponto de partida para o desenvolvimento da ferramenta. Na Figura 39

observa-se a versão da ferramenta resultado deste trabalho visualizando múltiplas bases de dados.

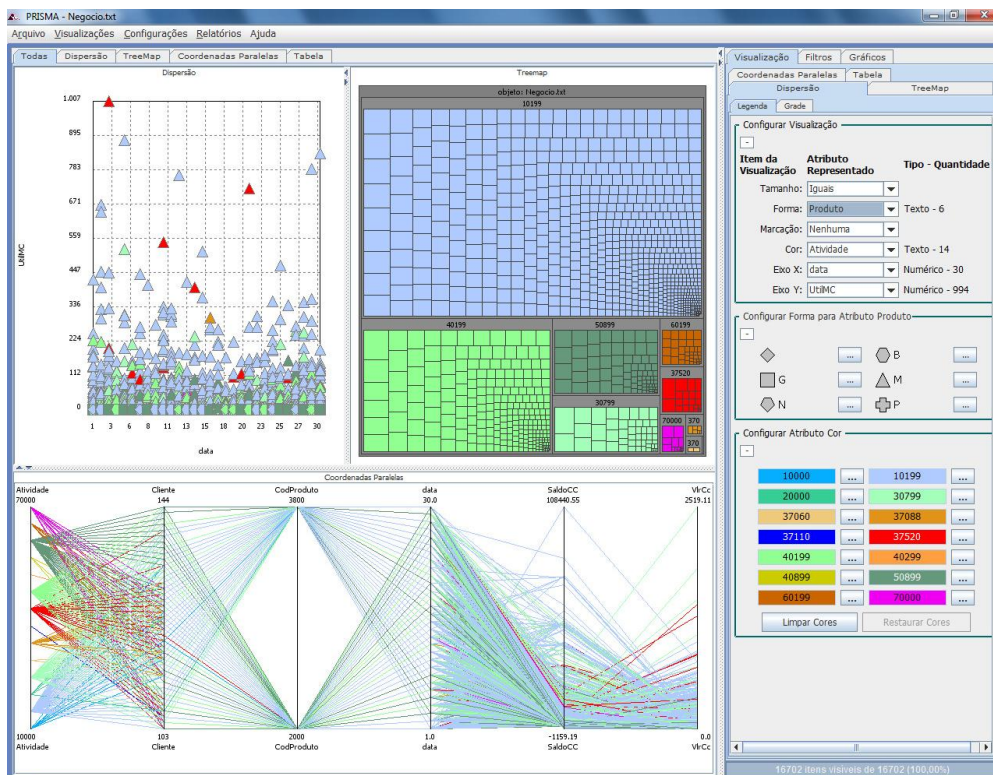


Figura 38 Ferramenta Prisma

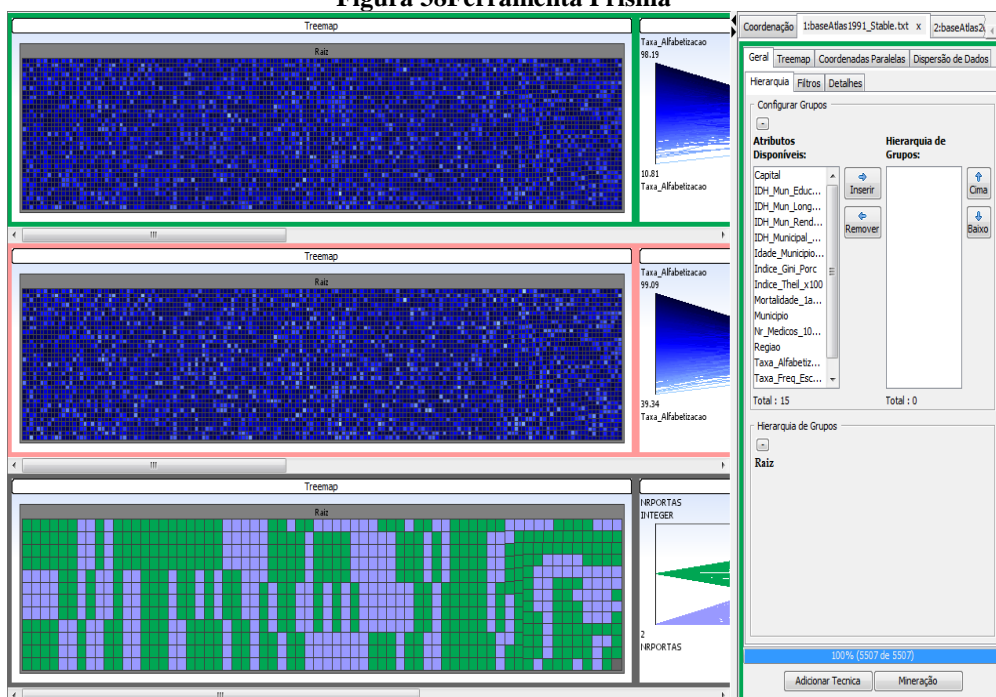


Figura 39 Prisma apresentando múltiplas bases de dados

4 IMPLEMENTAÇÃO DA FERRAMENTA

Neste capítulo é apresentada a evolução na arquitetura da ferramenta PRISMA, as principais limitações e como a nova arquitetura funciona e soluciona o problema de coordenação entre bases de dados e propõe maior flexibilidade.

4.1 Principais Características

A ferramenta foi desenvolvida paraplataforma desktop utilizando as linguagens Java, SQL e XML. O ponto de partida da implementação deste trabalho foi a ferramenta Prisma, a qual passou por um processo de reengenharia para garantir a possibilidade de carregar duas ou mais bases de dados simultaneamente e com isso poder implementar e aplicar novos conceitos como a coordenação entre bases de dados.

Outras características relacionadas a evolução da ferramenta incluem uma arquitetura de menor acoplamento, maior flexibilidade na criação de novas técnicas de visualização de informação e maior flexibilidade na utilização de novas camadas de apresentação.

4.2 Considerações no uso de múltiplas bases de dados

4.2.1 Carregamento de múltiplas bases de dados

Carregar múltiplas bases de dados em uma mesma ferramenta tem como objetivo permitir a análise comparativa ou complementar de bases de dados. Em um caso hipotético poderíamos ter uma base de dados com boletins de ocorrência criminais, outra base de dados metereológica e com isso verificar se os níveis de chuva influenciam na atividade criminal, ou se relacionarmos com uma terceira base de índice de escolaridade, desemprego e trabalho informal e verificar o impacto desses índices na ação criminal.

Outra possibilidade é carregar a mesma base de dados várias vezes na ferramenta para permitir a análise entre partes específicas da base de dados, como comparar visões de dois meses diferentes (uma base de dados com filtro configurado para Janeiro e a outra base de dados com o filtro configurado para Fevereiro).

4.2.2 Coordenação entre múltiplas bases de dados

Carregar múltiplas bases de dados somente não é suficiente, pois não trás vantagens significativas em relação a reabrir a ferramenta de visualização várias vezes e em cada uma carregar uma base diferente. É preciso tirar vantagem do ambiente formado, e para isso é necessário interligar as diversas bases de dados de alguma maneira.

O meio de relacionamento entre as bases de dados escolhidas para esse trabalho foi a coordenação de atributos (colunas) da base de dados. Esta abordagem foi escolhida por

estar no nível dos metadados e possibilitar maior liberdade na escolha das bases de dados.

A correlação é feita com o uso de agrupamentos de atributos, cada agrupamento permite apenas um atributo por base de dados. O agrupamento simboliza igualdade entre os atributos contidos, ou seja, todos os atributos de determinado grupo representam a mesma informação. Com isso, podem-se configurar filtros ou cores para um grupo e ter esta configuração disseminada por todas as bases de dados carregadas, o reflexo disto são suas devidas técnicas de visualização de informação coordenadas.

Ao carregar múltiplas bases de dados, a aplicação deve ter o cuidado de deixar claro ao usuário quais visões serão impactadas a cada configuração realizada, já que a multiplicação de controles é inevitável. Neste trabalho, optou-se pela utilização de bordas coloridas, cada base de dados aberta tem uma cor de borda específica que contorna tanto os controles quanto as visões com a mesma cor para fácil correlação entre visão e controle.

Para realizar a exibição de múltiplas bases de dados é necessária a exibição de pelo menos uma visão por base, em função disso, a análise de múltiplas bases é beneficiada pelo uso de displays de alta resolução e pelo uso de múltiplos displays. Para permitir uma exibição e organização flexível de visões em diversos monitores, é desejável que a ferramenta apresente um meio de exibir e distribuir as visões através de vários displays.

Neste trabalho, optou-se pela possibilidade de extrair a visão da janela principal da ferramenta para uma janela própria, com isso a distribuição das visões pelos displays fica a cargo do sistema operacional pelo posicionamento das janelas. Além de facilitar a distribuição em vários displays essa abordagem também permite o fácil redimensionamento da visão pelo simples redimensionamento da janela em que esta está contida. Na nova janela a visão mantém a borda com a cor relacionada a base de dados de exibição e ao fechar a janela a visão é reanexada à janela principal da aplicação.

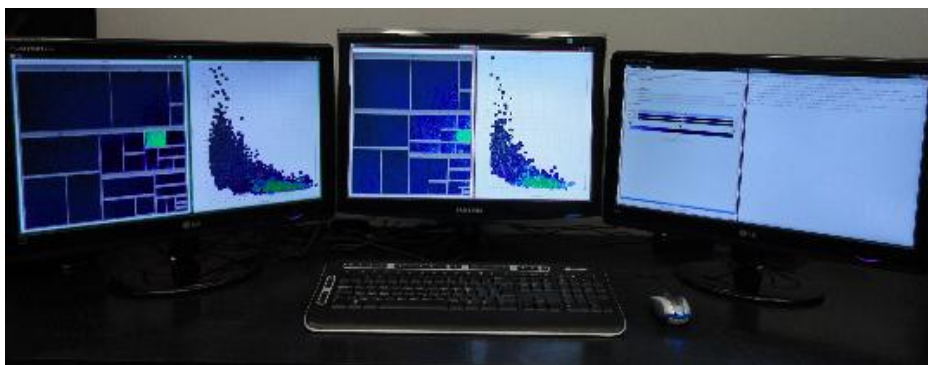


Figura 40 Estação de trabalho com múltiplos displays

4.3 Evolução PRISMA

Para permitir o carregamento e coordenação entre bases de dados a ferramenta passou por uma grande reformulação de sua arquitetura. Esta reformulação foi necessária, pois em diversas partes da implementação original, assumiu-se que a multiplicidade não seria necessária, com isso, diversos atalhos e implementações específicas para uma base de dados foram criadas.

4.3.1 Singletons

Aplicar o padrão de projeto Singleton é garantir que uma classe tem apenas uma instância e prover um ponto de acesso global a ela (ELISABETH FREEMAN, 2004). Os singletons utilizados na aplicação implementam o padrão de projeto bloqueando o acesso externo ao método construtor e disponibilizando a instância única por um método público e estático chamado "getInstance()". No PRISMA o padrão de projeto foi utilizado como solução de comunicação dentro da aplicação. Um exemplo de singleton pode ser visto na Figura 41.

```
public class Treemap {
    private static Treemap instance;
    private Treemap() { }
    public static Treemap getInstance() {
        if (instance == null) {
            instance = new Treemap();
        }
        return instance;
    }
}
```

Figura 41 Exemplo de Singleton

Os principais módulos e configurações da aplicação eram Singletons, o uso indiscriminado do padrão de projeto tornou extremamente difícil seguir o fluxo de execução da ferramenta, pois durante uma interação do usuário era comum observar a pilha de execução com 4 ou mais acessos desencadeados de Singletons. Esta utilização além de dificultar a depuração e organização do código fonte naturalmente inviabilizava a multiplicidade visada neste trabalho.

Para solucionar este problema foi criado o Núcleo, um módulo central, intermediador e controlador do sistema. Cada Singleton eliminado implicou em uma nova rota de acesso a informação por intermédio do Núcleo ou em uma nova abordagem de instanciação dos componentes para evitar chamadas desnecessárias. Esta etapa também serviu para eliminar alguns bugs gerados por estados inconsistentes nos componentes devido ao acesso indevido.

4.3.2 Falta de padronização

Um dos grandes desafios da ferramenta era a implementação de novos módulos, como um novo meio de acessar informação ou uma nova visualização, isto ocorria primeiramente pelo uso de Singletons e pela falta de padrões e regras a serem seguidas.

A solução para tal falta de padrões e regras foi a criação de interfaces de comunicação, estas interfaces abstraíram a implementação dos módulos que compõe a ferramenta e proporcionaram fácil integração. Após esta revisão, criar uma nova implementação de determinado módulo, como uma nova técnica de visualização de informação, é uma tarefa de implementar as necessidades impostas pela devida interface.

Este processo, assim como a remoção dos Singletons, também resultou em novas abordagens de instanciação dos componentes para evitar chamadas desnecessárias e propor maior desacoplamento dos componentes.

4.3.3 Abstração para múltiplas bases

Para atingir um dos objetivos deste trabalho, abrir múltiplas bases de dados, foi necessária a separação de responsabilidades encontradas no Núcleo. A responsabilidade de controlar e coordenar os diversos módulos necessários para geração de visões foi transferida para o MóduloVis. Nessa nova organização, o Núcleo é responsável pela comunicação com a camada de apresentação e é responsável pela criação de um MóduloVis por base de dados carregada.

4.4 Arquitetura e Modelagem

A arquitetura do software de visualização foi concebida com o objetivo de apresentar uma distribuição bem definida de responsabilidades, principalmente inspirada no modelo MVC. A arquitetura é dividida em três módulos principais (Figura 42): Núcleo, MóduloVis e Apresentação. Cada um destes módulos principais é subdividido em módulos especializados. Os módulos internos do MóduloVis e do Núcleo, contém modelos estruturais e estruturas de controle das suas funcionalidades específicas, por outro lado, os módulos internos da camada Apresentação são compostos principalmente por interfaces de interação com o usuário.

A Figura 43, Figura 44 e Figura 45 detalham o interior de cada uma das camadas principais, exibindo as principais classes, interfaces e o relacionamento entre elas para melhor entendimento do sistema.

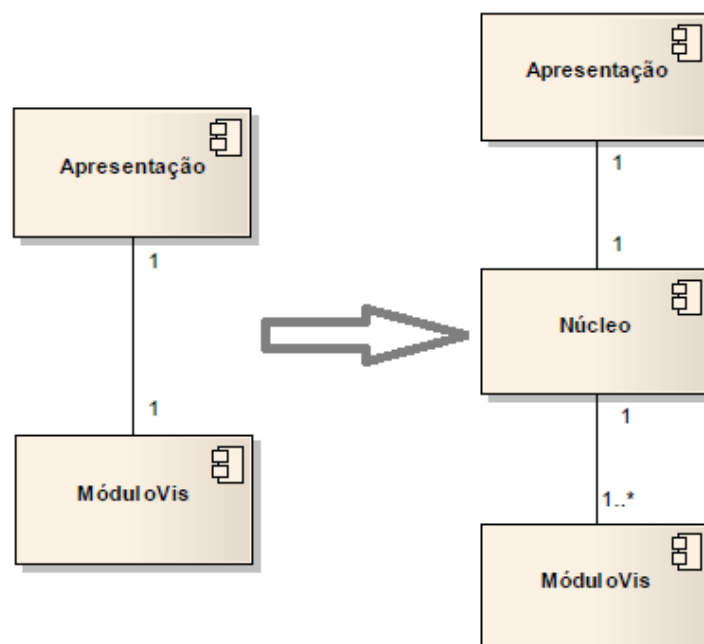


Figura 42 Reestruturação dos Módulos Principais

4.4.1 MóduloVis

O MóduloVis retém a responsabilidade de manter informações de configuração e gerar imagens com base nestas configurações. O MóduloVis é composto por:

- Uma fonte de dados (XML, CSV ou Banco de Dados SQL) para carga, seleção e organização dos registros;
- Um gerenciador de filtros;

- Um gerenciador de cores;
- Um gerenciador de detalhes sob demanda.
- Uma ou mais técnicas de visualização para geração de visões gráficas;

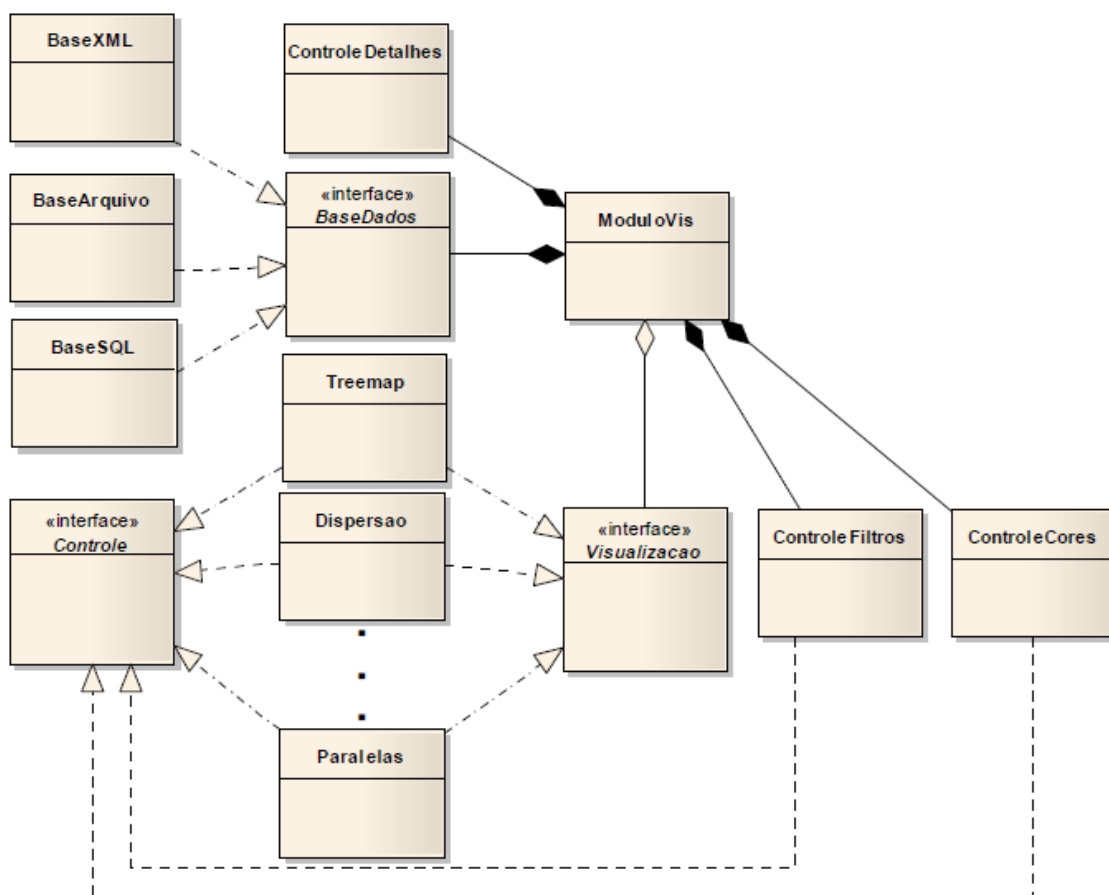


Figura 43 MóduloVis

4.4.1.1 Acesso a dados

A carga, o acesso e a gerencia dos dados é realizada de diversas maneiras, de acordo com a fonte dos dados, a única restrição é obedecer as regras impostas pela interface de comunicação com o MóduloVis. Isto possibilita implementações específicas e otimizadas para a estrutura em que os dados se encontram, como por exemplo a tradução do filtro em um comando em SQL.

Os principais métodos da interface BaseDados são:

- conectar, desconectar: estabelece e finaliza comunicação com a fonte de dados;
- cargaInicialDados: realiza operações necessárias apenas no carregamento dos dados, como a obtenção de metadados, limiares e totais dos atributos da fonte de dados;
- filtrar: submete à camada de dados os filtros configurados;
- getColunas: fornece o conjunto de atributos da fonte de dados e suas características, como tipo (numérico ou alfanumérico), menor valor, maior valor;

- `setHierarquia`: define a configuração da árvore em que os dados serão organizados, ex.: [País, Estado, Cidade];
- `setOrdenacao`: define como os registros devem ser ordenados;
- `getDados`: retorna os registros que devem ser exibidos em forma de lista;
- `getRaiz`: retorna os registros que devem ser exibidos em forma de árvore;

4.4.1.2 Gerenciador de filtros

O gerenciador de filtros tem o propósito de auxiliar na redução dos registros visualizados para um conjunto relevante e significativo na análise dos dados. Para isso é responsável por criar e gerenciar modelos de filtros, estes podem ser de três tipos de acordo com as características do atributo:

- Filtro contínuo múltiplo: para atributos numéricos, é composto por diversos intervalos. Ex.: Custo de vida dos municípios do Brasil;
- Filtro categórico de seleção múltipla: para atributos categóricos com pequeno número de valores possíveis, é composto de uma lista com os valores selecionados. Para este filtro escolhem-se individualmente quais valores serão selecionados para exibição. Ex.: Estados do Brasil;
- Filtro categórico De Até: para atributos categóricos com grande número de valores possíveis, é composto por duas letras ou palavras que servirão de limites de seleção. Ex.: Municípios do Brasil.

A cada interação na interface de filtragem o gerenciador de filtros informa ao `ModuloVis` o novo conjunto de condições de filtragem. Fica a cargo do `ModuloVis` aplicar os filtros imediatamente ou aplicar os filtros com atraso ou por intervenção do usuário, como um botão `Aplicar Filtros`.

4.4.1.3 Gerenciador de cores

Para gerenciar a coloração exibida nas visualizações utiliza-se uma abordagem similar a utilizada no gerenciador de filtros. De acordo com o tipo e características do atributo da base de dados um determinado modelo é utilizado. São dois possíveis casos:

- Coloração em degradê: para atributos numéricos, é composta de uma cor para o menor valor do atributo e uma cor para o maior valor do atributo. Com essas duas cores de referências são obtidas as cores para os valores intermediários do atributo.
- Coloração por valor: para atributos categóricos com pequeno número de valores possíveis, é composto de um mapeamento valor-cor que contempla todos os possíveis valores do atributo.

Os atributos categóricos com grande número de valores distintos não são elegíveis para atributo de coloração. Assim, como o gerenciador de filtros, cada alteração realizada dispara uma notificação ao `MóduloVis`.

4.4.1.4 Gerenciador de detalhes sob demanda

Este gerenciador guarda a lista de atributos que deve ser exibida quando a representação gráfica de um registro ou grupo de registros da base de dados sofre uma determinada interação ditada pela técnica de visualização, a interação mais comum é a ação de posicionar o mouse sobre a representação gráfica do registro.

O modelo deste gerenciador é disponibilizado as técnicas de visualização por intermédio do `MóduloVis` durante a etapa de desenho da técnica de visualização.

4.4.1.5 Técnicas de visualização de informação

As técnicas de visualização de informação têm com objetivo gerar representações gráficas interativas de determinados contextos. Estes contextos são compostos por:

- registros da base de dados que satisfazem as condições de filtragem;
- mapeamento de cores para os registros;
- atributos que devem ser exibidos em detalhes sob demanda;
- configurações específicas da visualização.

4.4.2 Núcleo

Núcleo é o módulo central do sistema, é composto de uma ligação com a camada de apresentação, um gerador de MóduloVis, um MóduloVis por base de dados aberta e um ControleCoordenação. É responsável por diversas atividades de controle, inicialização e encerramento da aplicação como:

- criação de um MóduloVis por base de dados carregada;
- inicialização da camada de apresentação;
- execução de determinadas chamadas provenientes da interface com o usuário, como abrir nova base de dados;
- entrelaçamento entre os componentes gráficos do módulo de Apresentação com os modelos contidos no MóduloVis e no ControleCoordenação;

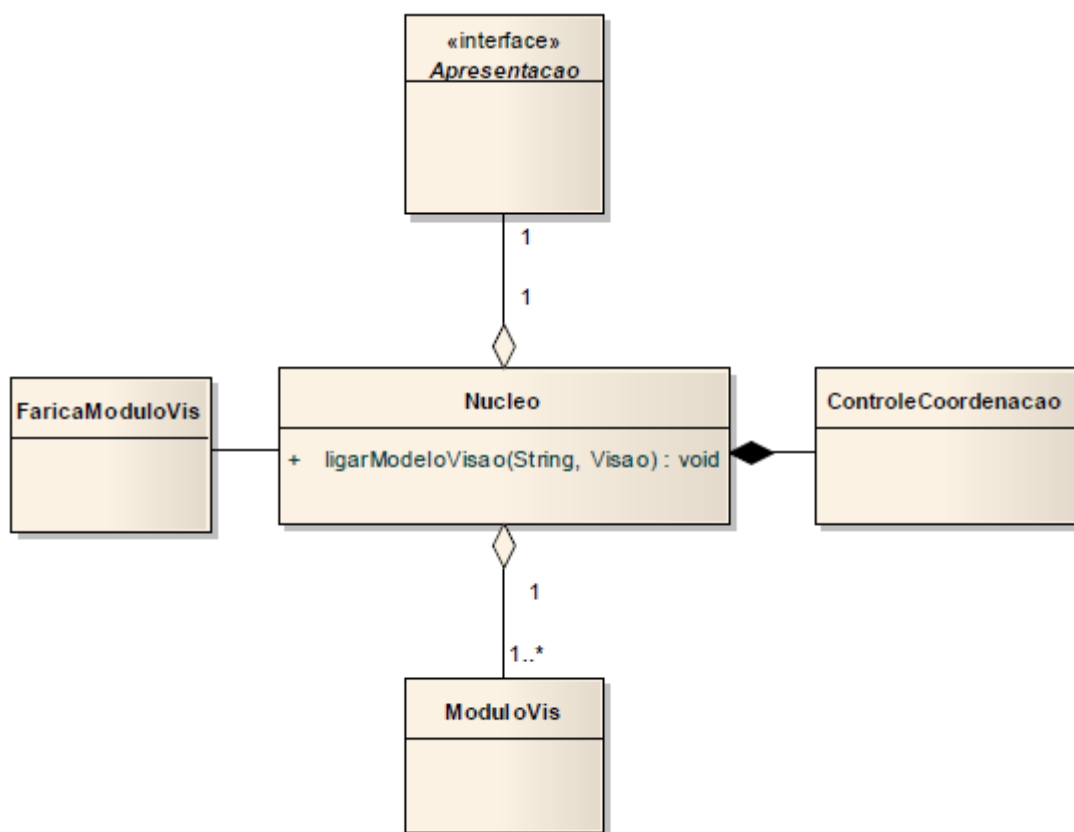


Figura 44 Núcleo

4.4.3 Apresentação

A Apresentação cria e gerencia o grupo de componentes de interface gráfica de usuário. Estes componentes são criados de acordo com a necessidade do usuário, adicionar uma nova fonte de dados para a aplicação requer a criação de novos painéis de configuração por exemplo. Apresentação tem um forte laço com o Núcleo, sempre que a Apresentação cria uma nova visão (seja esta um painel de configuração ou um painel de desenho) é feita uma requisição ao Núcleo para entrelaçar a visão recém criada com seu respectivo módulo.

Esta abordagem foi escolhida visando trabalhos futuros com interfaces web, em que o módulo de Apresentação poderá se tornar um Webservice provendo informações e um caminho de comunicação para qualquer plataforma compatível.

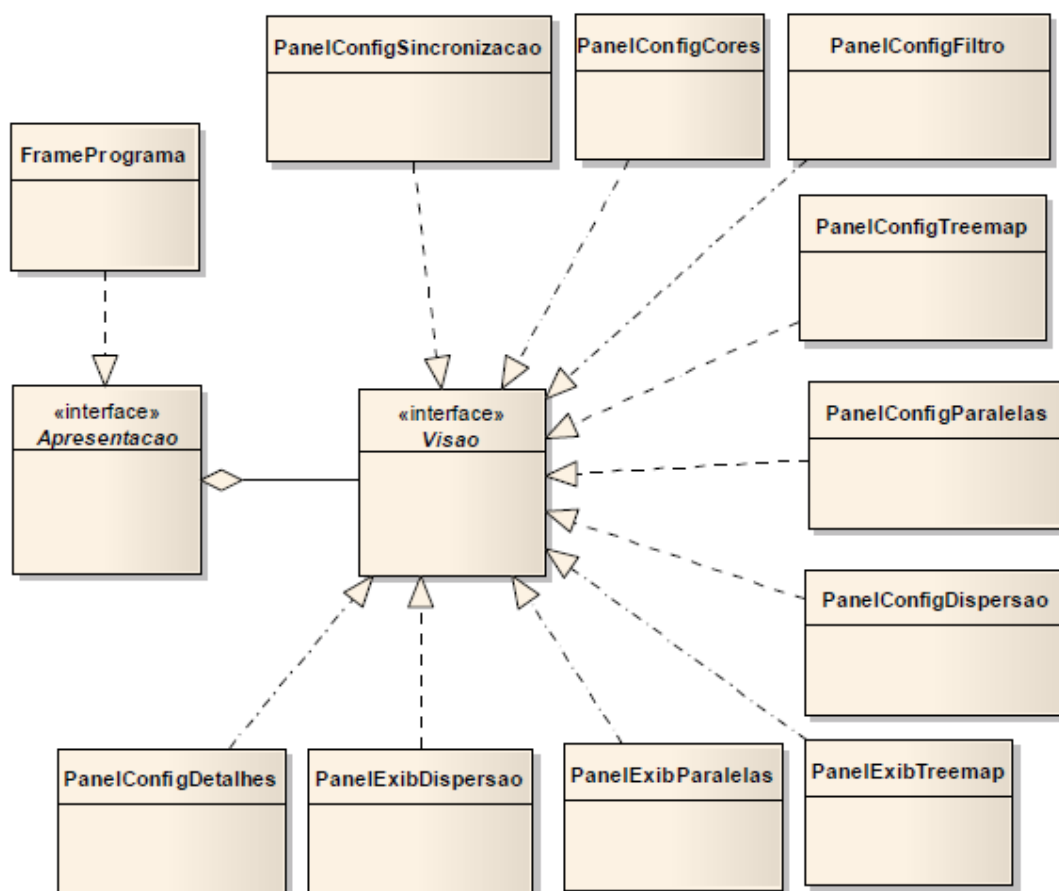


Figura 45 Apresentação

4.4.4 Controle de Coordenação

Buscando auxiliar na análise de múltiplas bases de dados, a ferramenta disponibiliza controles para coordenar os atributos de bases de dados distintas. Para alcançar este nível de coordenação utiliza-se o componente ControleCoordenacao. Este componente tem como objeto principal de funcionamento os grupos de atributos, estes grupos são utilizados para representar atributos de mesmo significado extraídos de diferentes bases de dados.

A criação de um grupo de coordenação afeta diretamente os controles de filtro e de cores, destes controles são extraídos os modelos relacionados aos atributos contidos nos grupos de coordenação. Estes modelos iniciais são guardados para caso o grupo de coordenação seja desfeito, e no lugar deles é colocado um novo modelo único para os atributos envolvidos.

Além de gerar modelos únicos para atributos de mesmo significado, este componentemonitora os eventos gerados pelos ModuloVis, e sempre que um controle de filtro ou de cor opera sobre um modelo gerado pela coordenação o estado é propagado para os outros MóduloVis.

A Figura 46 mostra a representação completa da arquitetura, de laranja a camada de apresentação, de verde a camada central, de vermelho o MóduloVis. No destaque de preto temos a comunicação entre ControleCoordenação e a interface de controles por ele monitorado.

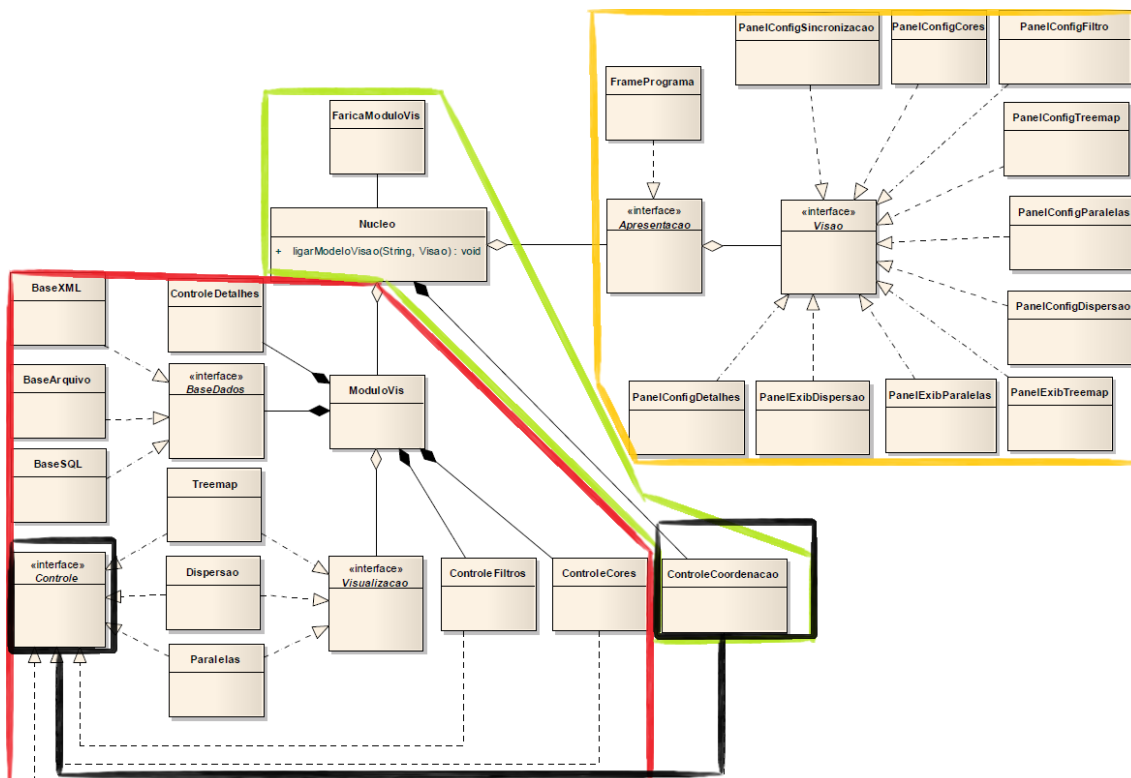


Figura 46 Visão Detalhada da Arquitetura

5 A FERRAMENTA E SUAS FUNCIONALIDADES

Neste capítulo é apresentada a ferramenta com foco na interface gráfica com o usuário, as diversas interações possíveis pelo usuário e as restrições de alguns controles.

5.1 Aspectos Gerais da Interface

A interface gráfica de interação da ferramenta é composta de três áreas gerais de interação:

1. Menu de aplicação: utilizado para carregar uma nova base de dados, salvar/restaurar configurações, exibir tópicos de ajuda, emitir relatórios e sair da aplicação;
2. Organizador de visões: exhibe e organiza as visões gráficas configuradas;
3. Controles: conjunto de controles relacionados a base de dados e técnicas de visualização carregadas, mais especificamente o lugar em que será configurado: filtro, detalhes sob demanda, cores, coordenação as técnicas de visualização (cada técnica com seu controle próprio).

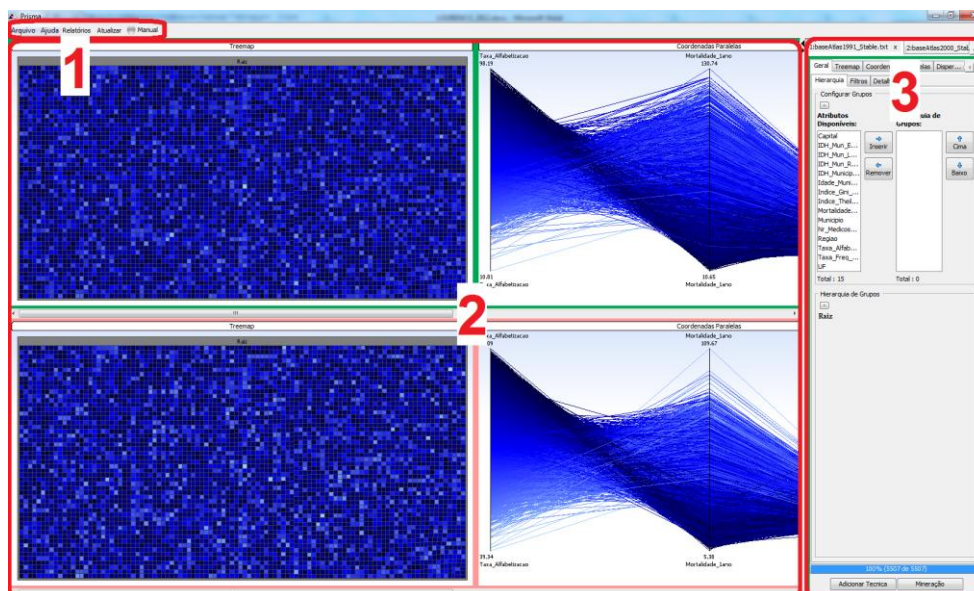


Figura 47 Interface principal

5.2 Controles de Configuração

Os controles de configuração estão organizados em uma árvore flexível de até 3 níveis de abas criadas dinamicamente, de acordo com o número de bases de dados e de

técnicas de visualização de informação utilizadas. Uma visão parcial das abas pode ser vista na Figura 48.

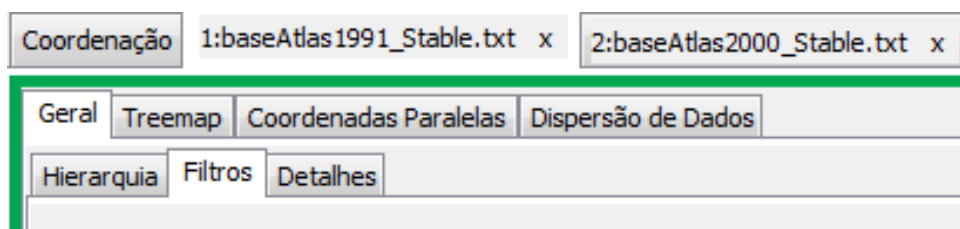


Figura 48 Abas de Configurações

O primeiro nível é o mais geral e contém uma aba de configurações de coordenação (sem sub-abas) entre bases de dados, e uma ou mais abas de configuração de MóduloVis, uma aba (com sub-abas) para cada base de dados carregada.

O segundo nível contém as abas pertinentes ao MóduloVis, destas abas, uma para configurações gerais (com sub-abas) e abas de visualização em que é criada uma aba (sem sub-abas) por técnica de visualização utilizada. Uma característica atípica desta aba é a presença de bordas coloridas, cada MóduloVis tem uma cor diferente, isto é utilizado para facilitar o relacionamento entre representação gráfica e configuração.

O último nível têm três abas fixas: Hierarquia, Filtros e Detalhes. Com o objetivo de facilitar a percepção das configurações de cores estas encontram-se replicadas em todas as abas de visualização. Outra representação das abas de controle encontra-se na Figura 49.

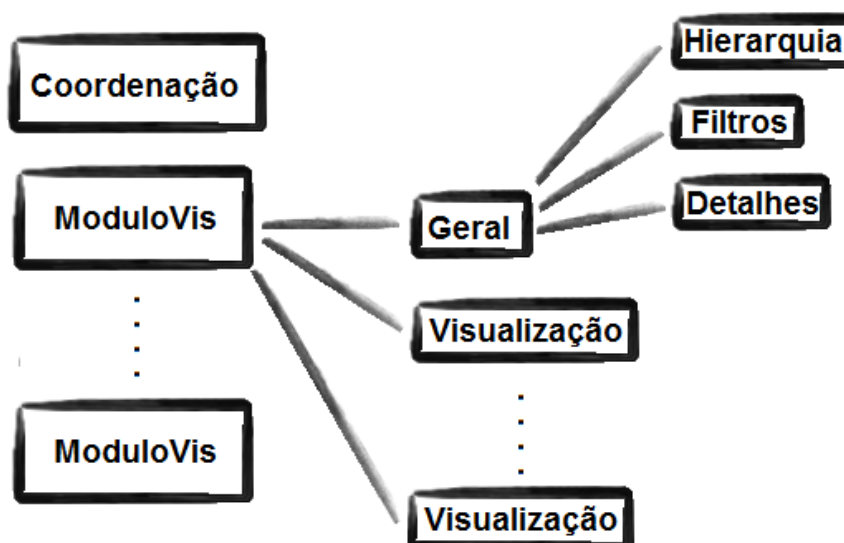


Figura 49 Abas de Configurações, representação alternativa

5.2.1 Configurações gerais de um MóduloVis

Algumas configurações são comuns a todas as visões geradas por um MóduloVis, estas configurações ficam agrupadas na sub-aba Geral.

A primeira aba (Figura 50) é a Hierarquia, esta representa o modo com que os dados serão agrupados e além de classificar os registros em grupos serve de parâmetro para obtenção de dados totalizados, como por exemplo, a média da mortalidade até um ano da região norte. O controle é composto de duas listas de atributos, uma contém os

atributos disponíveis para hierarquia e a outra com os atributos utilizados. A ordem dos atributos na segunda lista implica na ordem de agrupamento e na parte inferior do painel de controle visualiza-se a configuração realizada utilizando uma estrutura de árvore ao invés da lista. Para hierarquia usa-se apenas atributos categóricos.

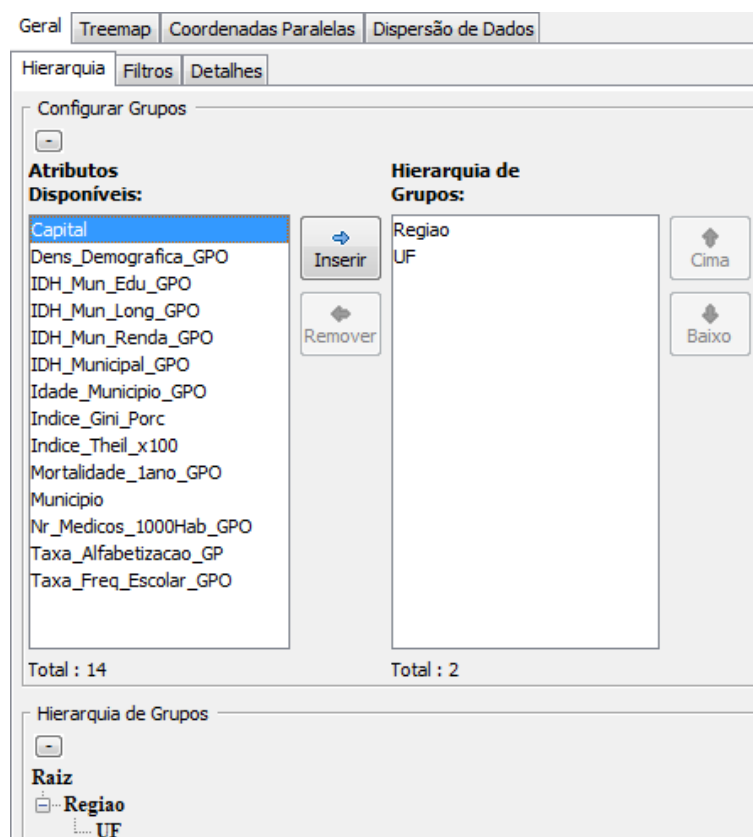


Figura 50 Controle de Hierarquia

5.2.2 Dispersão de Dados

Os controles para a técnica Dispersão de Dados são compostos de 5 configurações específicas, destas, 3 aceitam apenas atributos categóricos (Tamanho, Forma e Marcação) e outras duas (Eixo X e Eixo Y) apenas atributos numéricos.



Figura 51 Configuração da técnica Dispersão de Dados

5.2.3 Treemap

Para o Treemap configuram-se Tamanho e Rótulo (Figura 52) dos quadrados. Para tamanho o atributo deve ser numérico, já o rótulo pode ser de qualquer tipo de atributo, pois sua aplicação é voltada para a exibição da informação apenas, não está envolvido em operações específicas, apenas um texto informativo.

Item da Visualização	Atributo Representado	Tipo - Quantidade
Tamanho:	Nr_Medicos_1000Hab	Numérico - 235
Rótulo:	Nenhum	
Cor:	Taxa_Alfabetizacao	Numérico - 3492

Figura 52 Configurações Treemap

5.2.4 Coordenadas Paralelas

Para a técnica de Coordenadas Paralelas, os atributos da base de dados são representados em eixos verticais, na Figura 53, configura-se a lista de eixos que participarão da representação gráfica, sua ordem e a distância entre eles.

Mapear Atributos

Eixos Disponíveis	Eixos Inseridos
Ano_Instalacao	Taxa_Alfabetizacao
Capital	Mortalidade_1ano
Distancia_Capital	Renda_per_Capita
Domicilio_Agua_E	
Domicilio_Energia	
Domicilio_Telefon	
Domicilio_Urbano	
Esperanca_Vida_	

Total: 37 Total: 0

Configurações

Auto-Ajustar

100

Ajustar automaticamente escalas dos eixos

Figura 53 Configurações Coordenadas Paralelas

5.3 Configuração das Cores

Para configuração de cores da aplicação utilizam-se tanto atributos categóricos quanto numéricos. Cada um destes tipos implica em uma abordagem diferente na atribuição de cor aos dados. Apenas um atributo pode estar ativo como atributo de coloração, além disso, as interações nos controles de cor não são descartadas na troca de atributo de cor, ou seja, se alguma alteração for realizada no mapeamento de cor do atributo X e o atributo de cor passar a ser Y, as configurações de X são armazenadas para que se X voltar a ser o atributo de cor as interações anteriores poderão ser restauradas.

5.3.1 Cores Discretas ou Categóricas

Para os atributos categóricos, é realizado um mapeamento de cada possível valor do atributo com uma cor equivalente. Na tela de controle são exibidos os possíveis atributos, cada um inserido em uma caixa com sua cor equivalente, esta cor pode ser alterada no botão ao lado como pode ser visto na Figura 54.



Figura 54 Exemplo de cores para atributos categóricos

5.3.2 Cores Contínuas

A Figura 55 ilustra a segunda abordagem de coloração da ferramenta, em atributos contínuos, utiliza-se o mapeamento de intervalos de cores. Na implementação três valores são mapeados, o limite inferior do atributo, o ponto médio e o limite superior, a partir desse mapeamento pode-se inferir as cores para os valores intermediários.

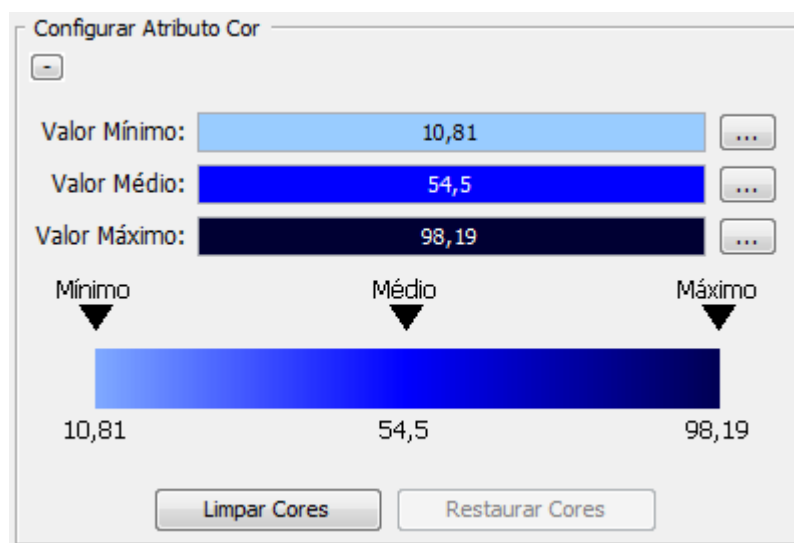


Figura 55 Exemplo de cores para atributos numéricos

5.4 Consultas Dinâmicas (Filtros)

Para realização de consultas dinâmicas na base de dados é necessário primeiro seleccionar quais atributos serão submetidos às regras de filtragem. Esta abordagem foi escolhida para evitar a exibição de controles de filtro não relevantes à análise desejada.

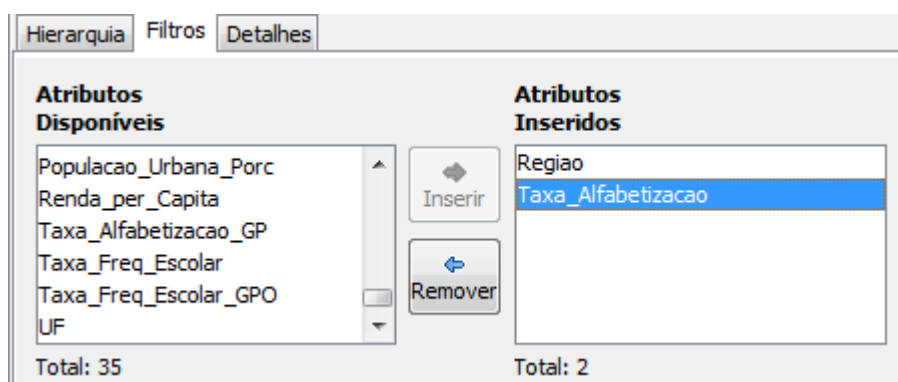


Figura 56 Seleção de atributos para filtragem

5.4.1 Filtros Categóricos

Para atributos categóricos os controles são compostos de caixas de seleção. Cada valor distinto do atributo tem sua própria caixa de seleção que, de acordo com seu estado de marcação, determinará a exibição ou oclusão dos registros representantes do determinado valor. Para facilitar seleções múltiplas pode-se escrever parte do nome do valor em uma caixa de texto, como pode ser visto na Figura 57.

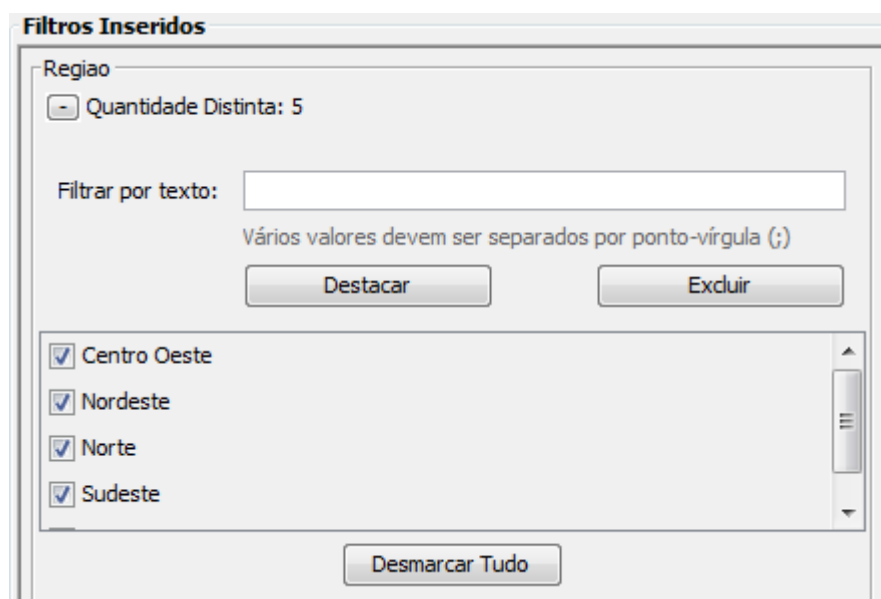


Figura 57 Exemplo de configuração de filtro para atributos categóricos

5.4.2 Filtros Contínuos

Para atributos contínuos utiliza-se um controle de seleção múltipla de intervalos, além dos ponteiros de seleção conta-se com um gráfico de barras ao fundo representando a concentração dos registros em relação ao atributo selecionado (Figura 58).

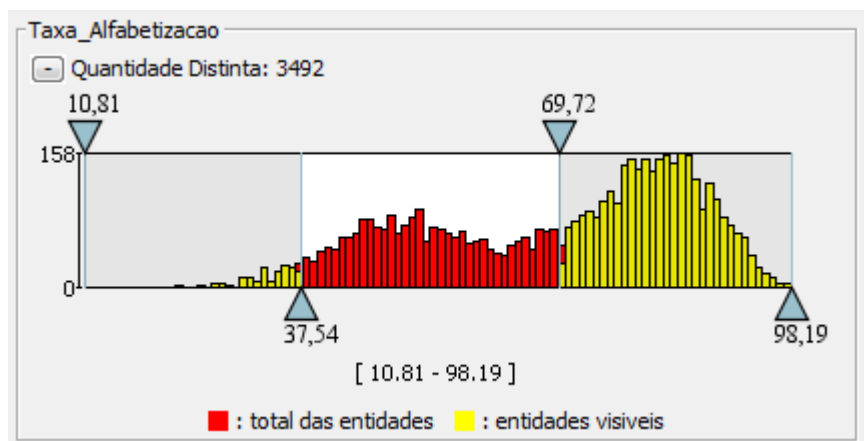


Figura 58 Exemplo de filtro para atributos numéricos

5.5 Seleção e Brushing

As ações de seleção e brushing não dispõem de um painel de controle lateral, pois estas são ações que tratam da discriminação de determinado grupo com base em uma região da visão gráfica. Esta interação pode variar entre as técnicas de visualização de informação.

5.5.1 Realizando uma Seleção

A Seleção ocorre em duas etapas distintas, na primeira etapa (Figura 59) uma área ou elemento gráfico de uma técnica de visualização deve ser selecionado com o ponteiro. A segunda etapa (Figura 60) é a oclusão dos registros fora da área de seleção e a propagação desta informação para as outras visões (coordenação).

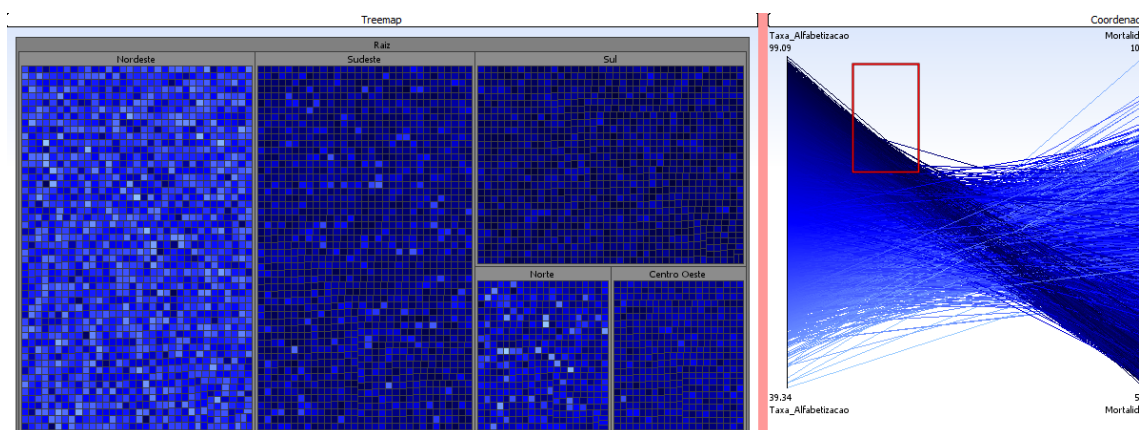


Figura 59 Realizando seleção I

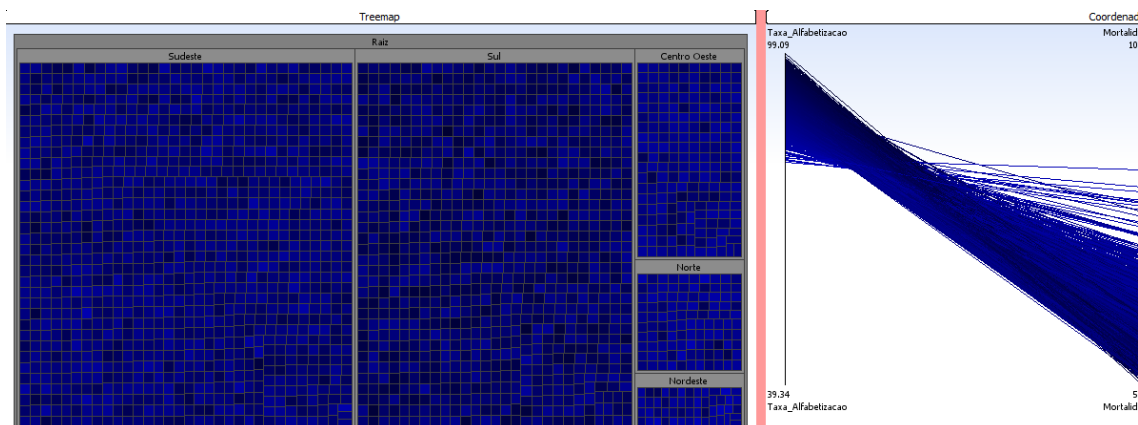


Figura 60 Realizando seleção II

5.5.2 Realizando um Brushing

O Brushing funciona de maneira análoga a Seleção, a diferença está que para realizar-se um Brushing deve-se manter a tecla CTRL pressionada durante a interação delimitadora. A outra diferença é que ao invés de remover da visualização os registros fora do grupo delimitador, a mudança ocorre nos registros selecionados que recebem destaque em forma de um contorno verde. Um exemplo de Brushing coordenado entre as técnicas pode ser visto em Figura 61, Figura 62 e Figura 63.

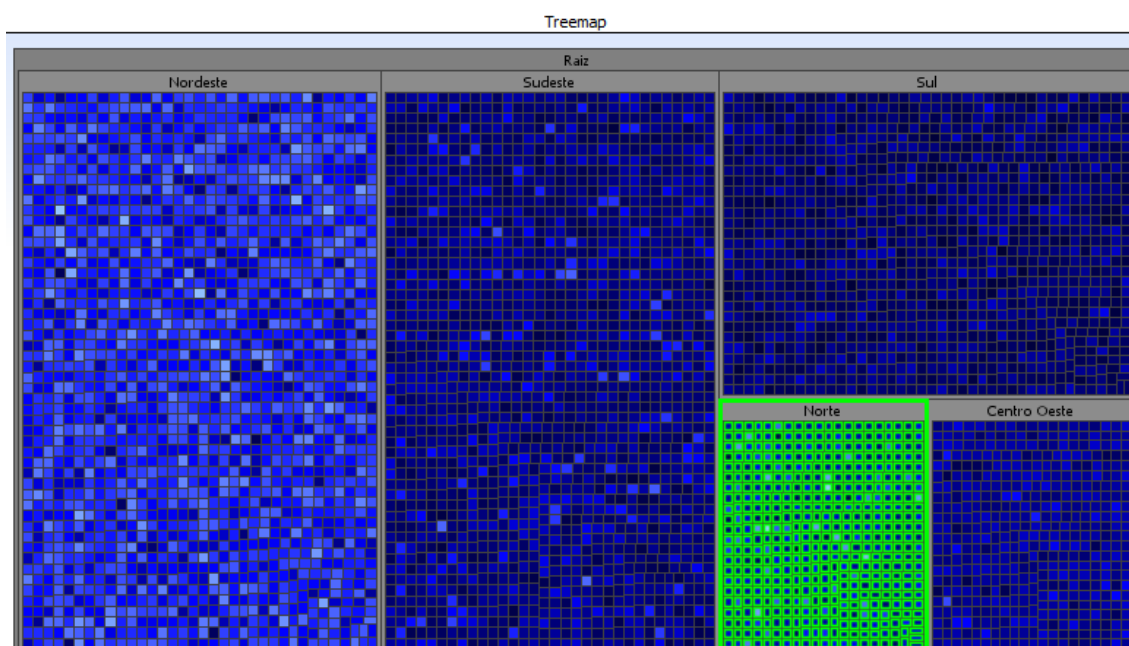


Figura 61 Brushing Treemap

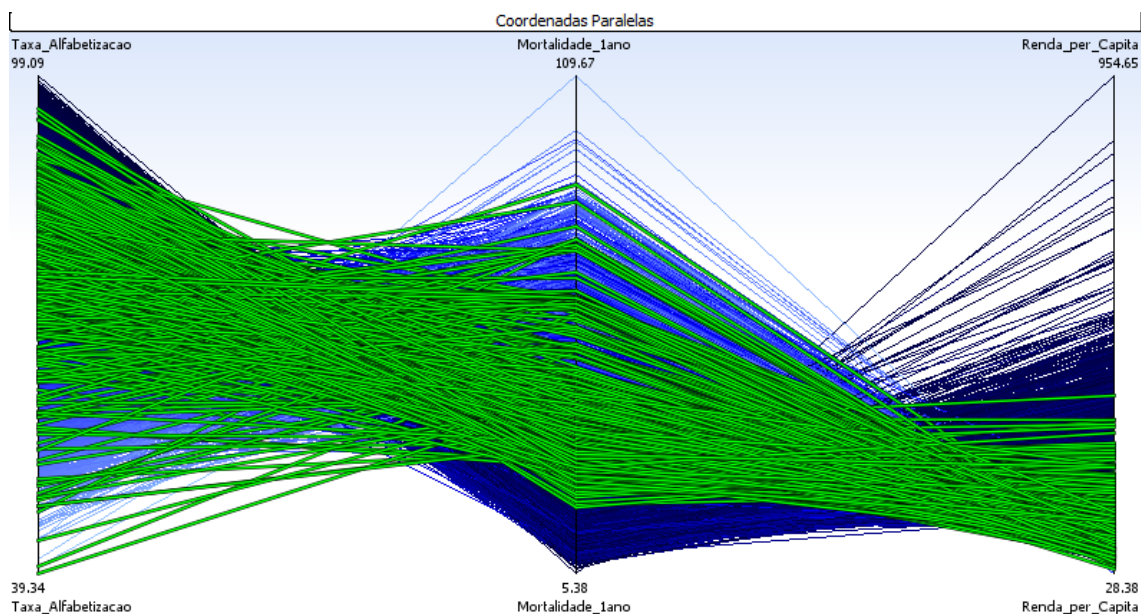


Figura 62 Brushing Coordenadas Paralelas

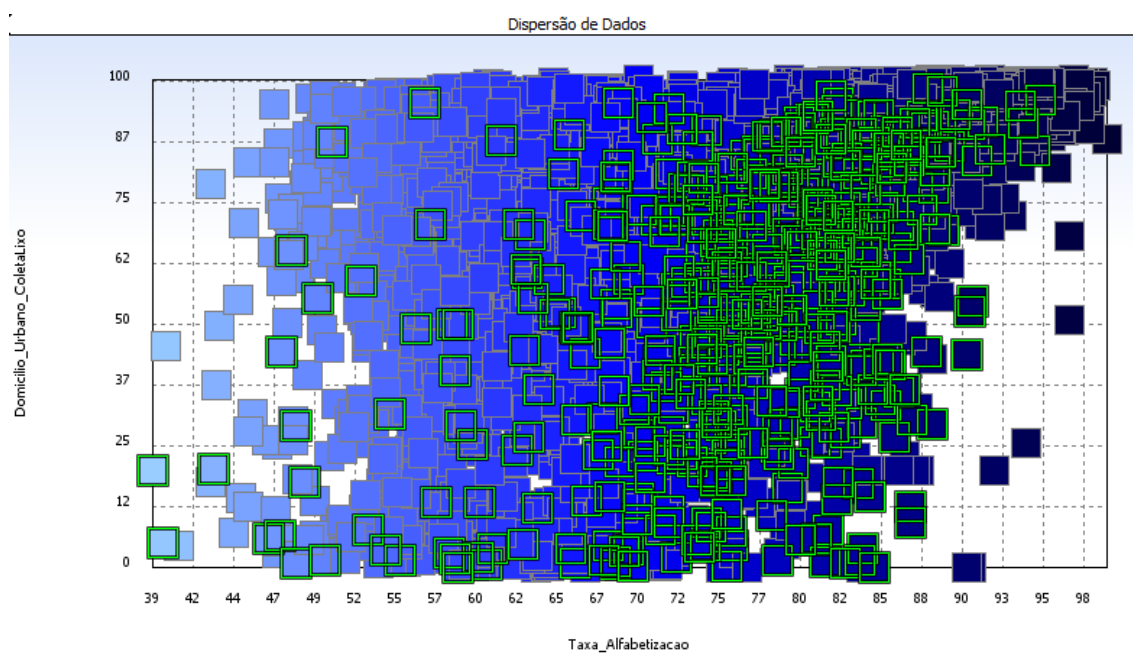


Figura 63 Brushing Dispersão

5.6 Detalhes Sob Demanda

A exibição de detalhes sob demanda de um determinado registro ou grupo de registros ocorre automaticamente ao passar do mouse sobre a representação gráfica do registro ou grupo. Na Figura 64 é mostrado um exemplo utilizando a técnica Treemap.

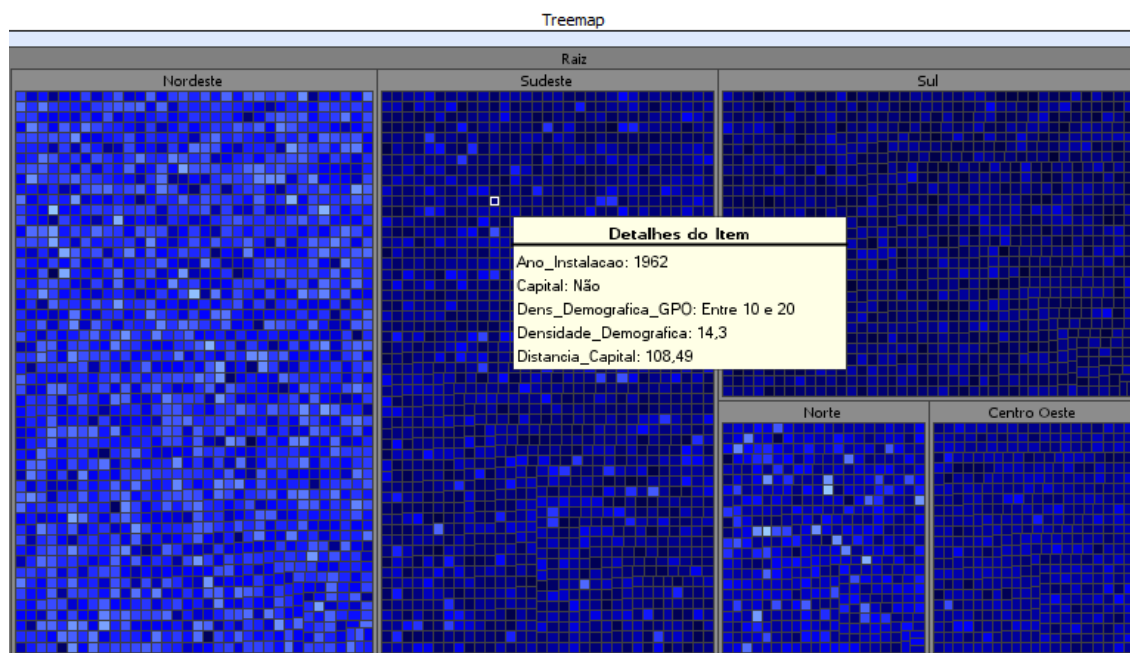


Figura 64 Detalhes Sob Demanda

5.7 Coordenação Entre Visões

A coordenação entre visões de uma mesma base de dados (mesmo MóduloVis) está presente e ocorre de forma automática ao configurar-se cor, filtro, seleção, brushing e a hierarquia dos dados.

5.8 Coordenação Entre Bases de Dados

Para coordenação entre bases de dados destaca-se na ferramenta quatro aspectos importantes: o painel de configuração, a organização das visões, e o relacionamento configuração e visão, e o desacoplamento das visões em janelas individuais.

5.8.1 Painel de configuração

Para a coordenação entre bases de dados configura-se grupos de coordenação, estes são listas de atributos que compõem o grupo especificado. Cada atributo nas listas é precedido pelo identificador único dado a base de dados carregada. Em um grupo só é possível adicionar um atributo por base de dados, e só é possível utilizar um determinado atributo uma única vez. A Figura 65 demonstra a adição de um novo atributo em um grupo de coordenação.

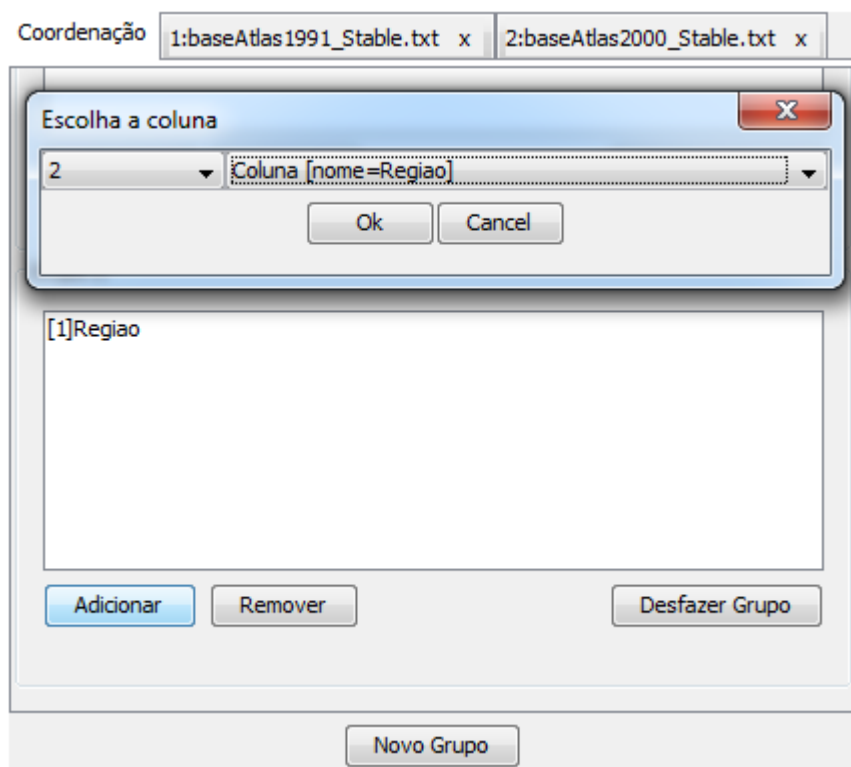


Figura 65 Painel de Coordenação

5.8.2 Organização das visualizações

A necessidade de abrir múltiplas bases de dados trás como requisito a flexibilidade na camada de apresentação de poder exibir uma ou várias visões. Para resolver esta questão a estratégia escolhida foi de exibir na horizontal as visões das diversas técnicas de uma determinada base de dados, e fazer na vertical a separação das bases de dados. Na Figura 66 pode-se visualizar o comportamento da ferramenta ao carregar 3 bases de dados.

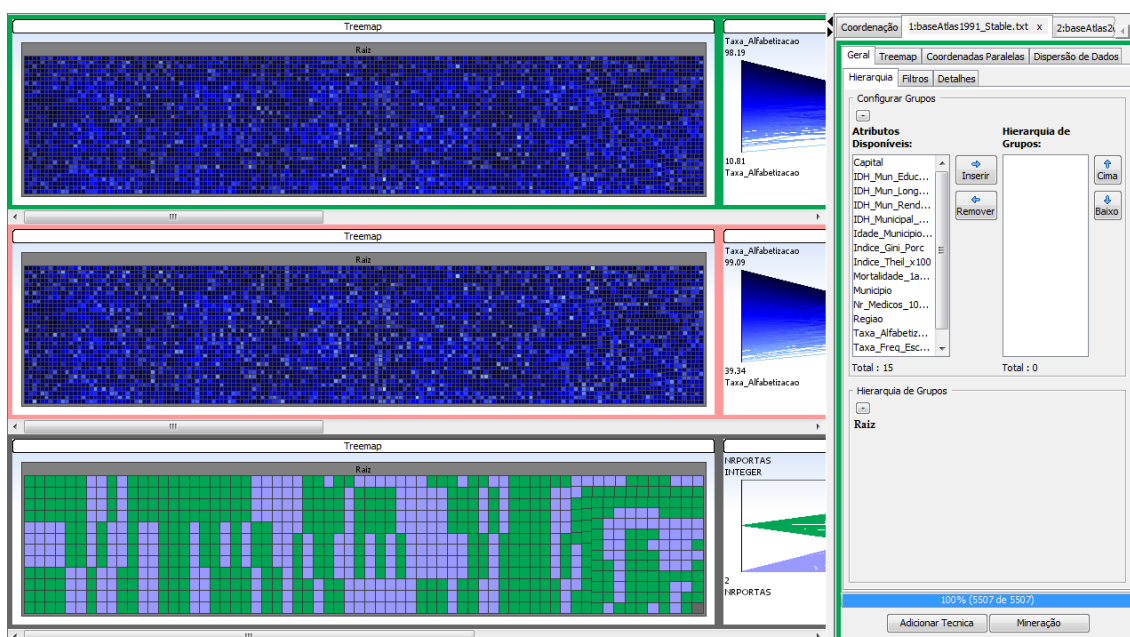


Figura 66 Organização de interface gráfica de usuário para 3 bases de dados

5.8.3 Relacionamento visual entre visão e configuração

Quando se trata de multiplicidade, e no caso deste trabalho temos múltiplas técnicas e múltiplas bases de dados em análise, é sempre um problema exibir uma grande quantidade de opções de configuração e deixar claro ao usuário aonde determinada configuração irá impactar. Para amenizar este problema optou-se pela utilização de bordas coloridas contornando tanto os painéis de configuração de bases de dados quanto as visões. Na Figura 67 e na Figura 68 pode-se visualizar o casamento do painel de configuração com das visões pelas bordas verde e vermelha.

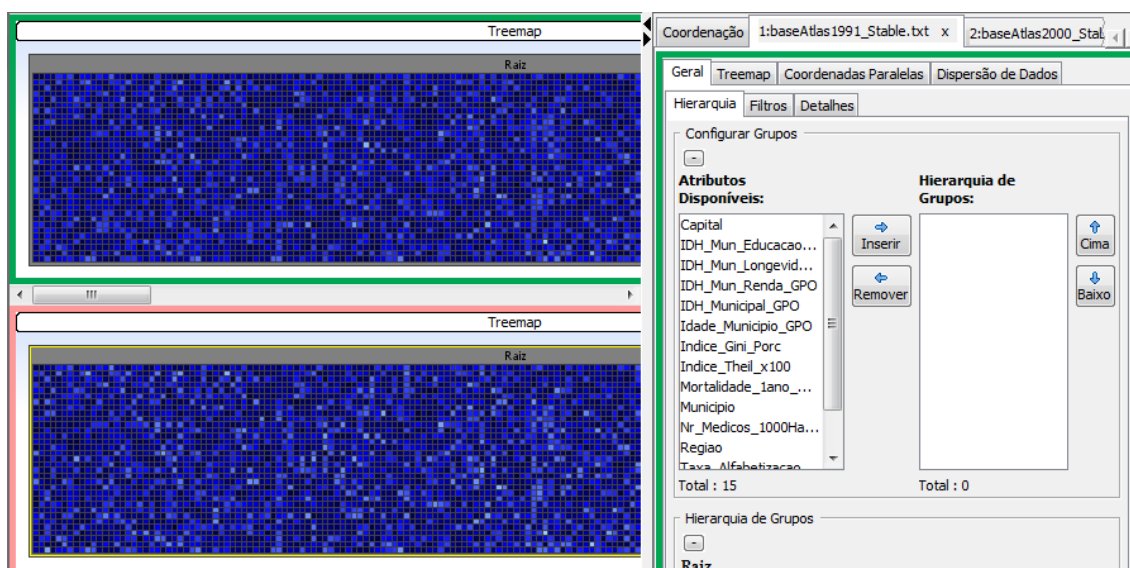


Figura 67 Relacionamento de visão controle 1

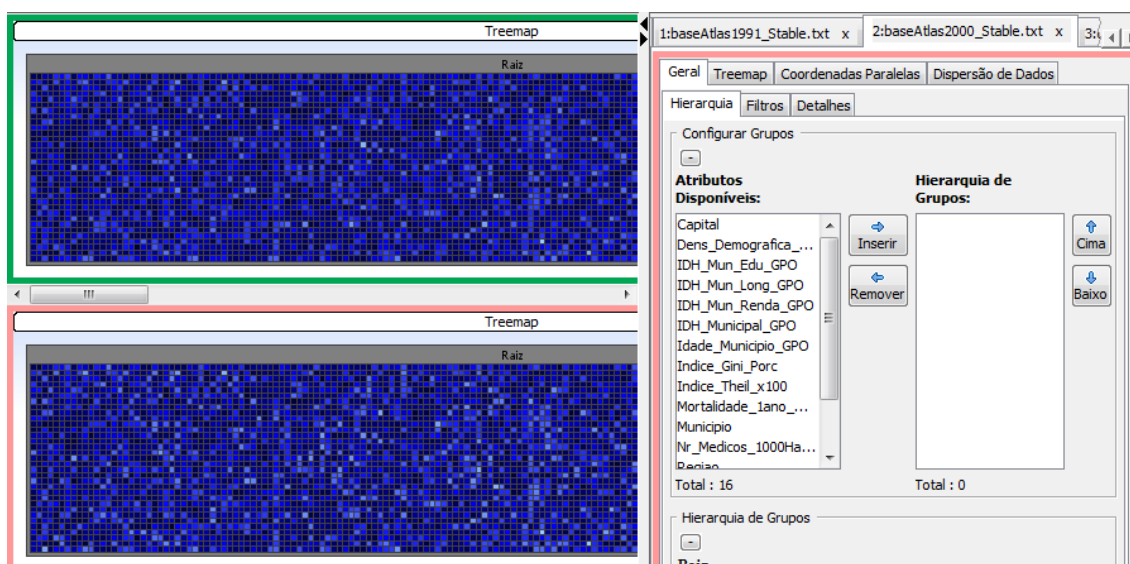


Figura 68 Relacionamento visão controle 2

5.8.4 Desacoplamento de visões

Durante uma análise na ferramenta as várias visões exibem diferentes faces dos dados, porém, como a utilização da ferramenta é um processo interativo, em muitos

momentos deseja-se analisar com maior detalhe ou garantir destaque a determinada visão dos dados. Para isso permite-se o desacoplamento da visão da janela principal da aplicação (Figura 69). Esta característica também facilita a interação em múltiplos displays pela facilidade de posicionar a janela de uma visão chave em um display e a configuração em outro, por exemplo.

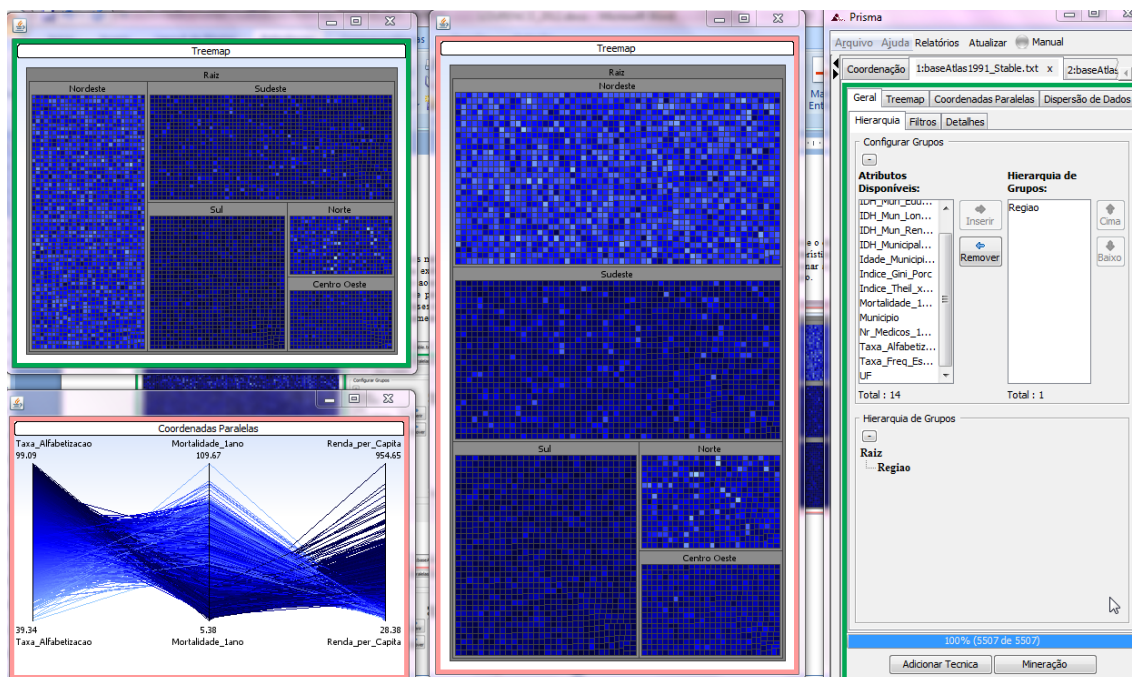


Figura 69 Visões desacopladas da janela principal da ferramenta

6 TESTES DE USABILIDADE

Este capítulo tem como objetivo descrever o processo de aplicação do teste de usabilidade e o perfil dos usuários que o realizaram, bem como os resultados concernentes sobre uso do protótipo em questão.

6.1 Visão Geral

Segundo Rubin(1994), testes com usuários têm por objetivo avaliar a usabilidade através da experiência de uso de usuários que possivelmente utilizarão o sistema. Essa metodologia de avaliação não se baseia apenas na coleta de dados quantitativos, mas também na observação do comportamento dos usuários, observando suas reações quanto a aspectos do sistema, especialmente problemas e dificuldades que só ocorreriam em um cenário real (BARBOSA e SILVA, 2010).

Rocha e Baranauskas(2003) ressaltam a importância de ter um objetivo bem definido antes da definição do teste, uma vez que estes devem ser realizados de forma a ressaltar os critérios de usabilidade a serem medidos. Neste trabalho procura-se ressaltar o ganho de tempo e a redução de esforço (físico, mental e cognitivo) na análise visual de dados dispostos em mais de uma base de dados, considerando-se utilizar uma única ferramenta com duas base de dados coordenadas ou utilizar uma base de dados em cada ferramenta de visualização de forma separada. O teste foi projetado utilizando o conceito de pursuit, definido por Spence(2007), onde o usuário realiza movimentos no espaço de informação buscando um objetivo único, como a resposta para uma pergunta, com o mínimo de referência a um modelo mental acumulado sobre o espaço de informação.

Utilizou-se no teste duas bases de dados do IBGE (2009), com dados referentes ao censo dos anos 1991 e 2000. Ambas as base de dados tem 5508 registros (número de municípios) e 34 atributos, sendo 12 categóricos, em que cada registro representa um município brasileiro. Para atender o objetivo do teste foram avaliados 10 usuários divididos em 2 grupos, ambos de 5 usuários. O primeiro grupo utilizou a funcionalidade de coordenação de dados, carregando as duas bases em uma mesma instância do PRISMA. O segundo grupo não utilizou a funcionalidade de coordenação e fez uso de uma instancia do PRISMA para cada base de dados. Cada grupo foi subdividido em 3 participantes que já tinham experiência na área de visualização de informação, e algum contato prévio com a ferramenta PRISMA, e 2 participantes que não tiveram nenhum contato com a ferramenta e pouca experiência em visualização. Para os participantes que não tinham experiência foi oferecida uma apresentação, de aproximadamente 10 minutos, sobre a ferramenta e todos os participantes tiveram 5 minutos de interação livre com a ferramenta antes da realização dos testes. Durante a realização das tarefas foi permitido aos usuários perguntar aos avaliadores questões referentes apenas a utilização da ferramenta, em nenhuma hipótese sobre como realizar as tarefas propostas.

As métricas consideradas foram o sucesso na realização da tarefa proposta e o tempo despendido nesta. Posteriormente as tarefas, pediu-se que os usuários realizassem o Nasa-TLX Test (HART e STAVENLAND, 1988), com o intuito de verificar se a utilização da coordenação entre bases efetivamente contribuía para a redução do esforço na utilização da ferramenta.

As tarefas propostas foram projetadas para serem de dificuldade crescente, da primeira para a última, envolvendo cada vez mais configurações (como filtros) específicas ou um número maior de atributos a analisar.

6.2 Tarefas Propostas

Para não tornar o teste extenuante nem interferir com a disposição do usuário em realizá-lo foram definidas apenas quatro tarefas. Estimou-se um tempo de 6 minutos para as 3 primeiras e 8 minutos para a última tarefa, resultando num tempo despendido inferior a 30 minutos para a realização das mesmas.

As duas primeiras tarefas requerem ao usuário que realize uma busca no espaço de informação para responder a pergunta proposta na tarefa. Optou-se por utilizar respostas objetivas (SIM ou NÃO) para que fosse mais fácil avaliar o sucesso dos participantes, além das respostas objetivas foi solicitado aos participantes que graduassem a tarefa em relação ao nível de dificuldade da mesma (Muito Fácil, Fácil, Normal, Difícil e Muito Difícil).

A primeira tarefa pede para que o usuário compare a região Norte nas duas bases e descubra se houve um aumento na taxa de alfabetização dos municípios desta região. A segunda tarefa requer que o usuário identifique um município específico e avalie se dois atributos correlacionados (frequência escolar e taxa de alfabetização) apresentam ou não comportamento discrepante. A terceira tarefa pede ao usuário que compare as regiões nas duas bases e identifique a região que possui a maior renda per capita em ambas. A última tarefa questiona o usuário se houve evolução nos índices de cinco atributos com relação ao município de Santos-SP, tais atributos são: população total, renda per capita, esperança de vida no nascimento, mortalidade no 1º ano e número de médicos a cada mil habitantes. Devido a comparação entre 5 atributos das duas bases, bem como as configurações propícias para a análise destes, esta tarefa foi classificada como complexa.

6.3 Avaliação Pós-Teste (NASA-TLX)

Utilizou-se o Nasa-TLX com o objetivo de analisar e comparar a carga de trabalho subjetiva gerada pelo mesmo conjunto de tarefas executadas sobre a mesma ferramenta, tendo como diferencial apenas a utilização ou não da coordenação entre múltiplas bases de dados. O TLX utiliza a média ponderada sobre as notas atribuídas pelo usuário para cada categoria (demanda mental, demanda física, demanda temporal, nível de esforço, nível de frustração e nível de desempenho) e da relevância de cada uma dessas categorias para o usuário.

De acordo com Hart e Stavenland (1988), as respostas variam de acordo com o grau que cada uma das categorias listadas contribuiu para a carga de trabalho em uma determinada tarefa, podendo ser determinada pela perspectiva das taxas. Tal taxa é apurada a partir de um valor numérico, sendo este determinado a partir de um formulário que contém as seis (6) demandas, sendo que cada uma delas apresenta uma escala

graduada sem valores numéricos, dessa forma, o participante assinala a magnitude que determinado fator contribuiu para a formação da carga em determinada tarefa.

6.4 Resultados

Percebe-se na Figura 70 que o desempenho dos usuários que usou a funcionalidade de coordenação teve melhor rendimento que o grupo sem esta funcionalidade, com relação ao fato de acertarem com precisão as respostas das tarefas. Este rendimento foi 24% maior. O gráfico abaixo mostra a média de acertos de cada grupo com relação às tarefas. Para melhor entendimento, a partir de agora o grupo com a funcionalidade de coordenação é identificada pela sigla CC_MB e o grupo sem a função de coordenação de dados pela sigla CC_N.

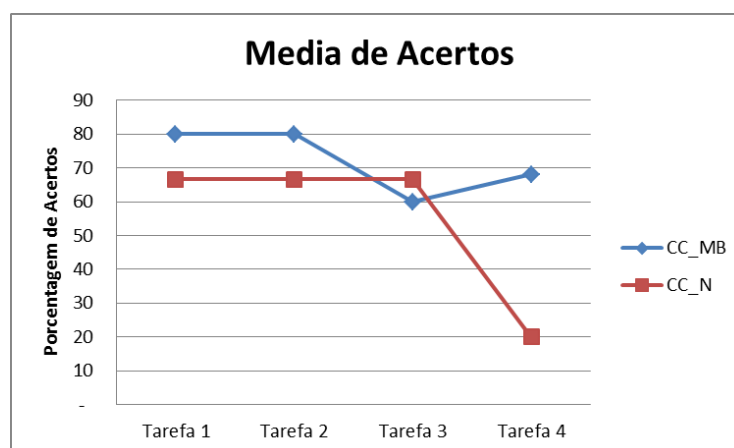


Figura 70 Média de acertos por grupo e tarefa

A partir das observações realizadas durante os testes, atribui-se esse rendimento no nível de tarefas realizadas com sucesso a dois fatores principais: a relevância das configurações realizadas, que refletidas sobre a outra base elucidada mais rapidamente a informação procurada, e a diminuição do efeito de desorientação causado por alternar entre duas instâncias da aplicação, uma vez que refletidas as configurações, o usuário depende menos da memória e do mapeamento mental feito sobre as ações previamente realizadas.

Com relação ao tempo despendido em cada tarefa, a média de tempo entre o grupo que utilizou a coordenação entre as bases de dados e o grupo que não fez uso do recurso foi semelhante. Apenas nas tarefas 2 e 4 este tempo foi ligeiramente superior a 1 minuto. Observa-se, no entanto, que os usuários que utilizaram a coordenação entre bases de dados despenderam muito tempo para configurar essa coordenação. Para tal configuração, a ferramenta oferece uma interface composta por um componente combo box para cada base em cada coordenação, listando todos os atributos da base (34, nas bases de testes) neste componente. Isso demanda que o usuário busque o atributo desejado em um componente que mostra poucos atributos por vez, sendo necessário rolar o componente para chegar ao atributo desejado. Fator este que evidencia a necessidade de reformulação do controle para tal configuração. A Figura 71 discrimina a média de tempo gasta por cada grupo de usuários em cada tarefa, seguindo a nomenclatura definida para a Figura 70.

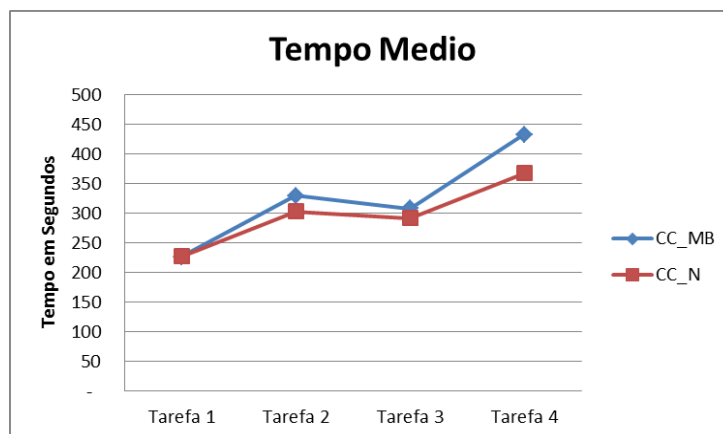


Figura 71 Média de tempo em segundos gasto em cada tarefa

As seguintes observações pertinentes a Figura 71 podem ser elaboradas com relação ao tempo gasto pelos usuários na realização das quatro tarefas. Sabe-se que as tarefas possuem teoricamente complexidade crescente. É natural inferir que o usuário precise de mais tempo para resolver uma tarefa mais complexa que outra. Por conseguinte, os testes realizados procuram encontrar naturalmente uma evolução no tempo. Porém, a Figura 71 apresenta uma pequena distorção na evolução do tempo quanto à tarefa três, onde o grupo CC_N e o grupo CC_MB, na média, conseguiram ser mais rápidos nesta tarefa do que na anterior. Em uma análise mais minuciosa da Figura 71, esta distorção foi mais acentuada para o grupo CC_MB, evidenciando que a ferramenta com coordenação de dados mostrou-se mais eficaz na análise comparativa de vários dados ao mesmo tempo. Outra observação em relação aos grupos sobre o tempo médio de resposta de todas as tarefas, o qual foi 24% superior para o grupo CC_MB, por outro lado a média de acertos desse grupo foi 8% superior, o que demonstra que a ferramenta permitiu que os usuários aumentassem a qualidade de suas respostas, porém a ferramenta diminuiu um pouco a agilidade dos usuários.

A análise dos usuários, com relação à complexidade, pode ser observada no gráfico da Figura 72. Foi questionado ao usuário qual a complexidade de cada tarefa realizada. Observa-se que a complexidade entre os grupos quase se iguala nas tarefas 1 e 3, porém a disparidade é absoluta com relação as tarefas 2 e 4, onde o grupo CC_MB, através do uso da ferramenta de coordenação dados, mostrou teoricamente melhor desempenho, pois atribuíram às tarefas graus de complexidade baixo, apesar da quarta tarefa ser considerada a mais complexa, como já dito anteriormente.

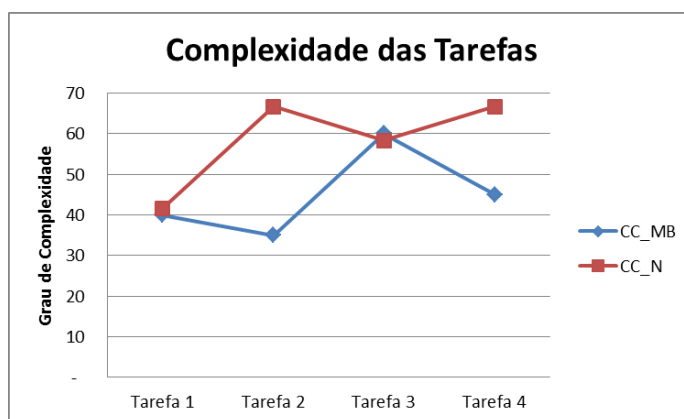


Figura 72 Médias das complexidades por tarefa e grupo

Continuando o processo de análise dos resultados, a próxima etapa a ser analisada é a aplicação do NASA-TLX. Depois de realizado os questionamentos devidos ao qual o NASA-TLX requer aos usuários, obteve-se o workload ou a carga de trabalho final sobre as seis demandas preconizadas pelo NASA-TLX. Este resultado pode ser visualizado na Figura 73abaixo.

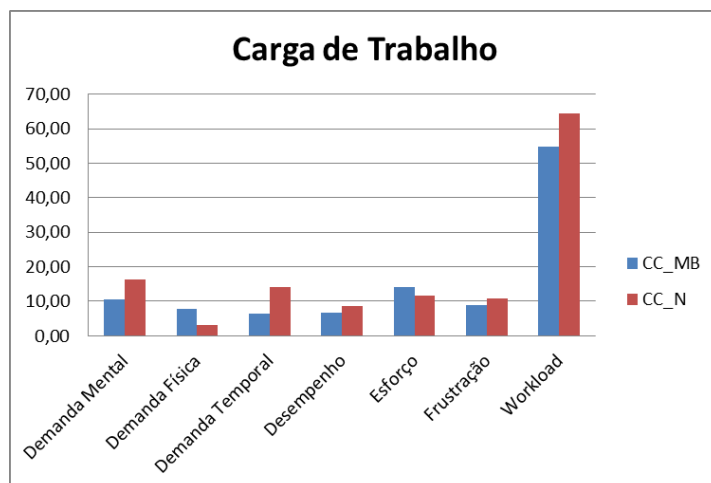


Figura 73 Carga de Trabalho final entre os grupos

Em relação à carga de trabalho geral, o grupo CC_MB teve uma carga de trabalho inferior ao grupo CC_N, isso demonstra que os usuários que utilizaram a ferramenta de coordenação de dados em múltiplas bases, no geral, conseguiram melhor explorar e analisar os dados para resolver os problemas propostos. Ainda com relação a Figura 73, percebem-se diferenças razoáveis com relação à demanda mental e a demanda temporal. Estas características enaltecem a qualidade da ferramenta em reduzir o tempo e a atividade mental e perceptiva do usuário na análise visual dos dados.

7 CONSIDERAÇÕES FINAIS

O trabalho tem como objetivo apresentar o potencial e as características de uma ferramenta de visualização de informação que pode lidar com diferentes conjuntos de dados e sincroniza-los através da relação entre atributos semelhantes entre os conjuntos de dados envolvidos. Os pontos de vista criados podem ser apresentados em várias técnicas, e as visões são coordenados entre si para melhorar a percepção do usuário em relação aos dados e seus relacionamentos.

Com isso, o uso de múltiplos conjuntos de dados relacionados estende o conceito de compreensão de dados e suas relações, como é feito ao longo de conjuntos de dados diferentes. Por exemplo, para procurar padrões semelhantes de informações entre os conjuntos de dados. No entanto, o uso não é limitado à utilização de conjuntos de dados diferentes. O mesmo conjunto de dados pode ser explorado com diferentes configurações de filtro, essas tarefas simples de analisar exibições de dados com diferentes configurações e filtros podem não estar disponíveis em algumas ferramentas IV.

Após um estudo minucioso do funcionamento da ferramenta PRISMA detectou-se os possíveis impedimentos de implementação, após a retirada desses impedimentos a nova arquitetura foi proposta e posteriormente validada na implementação da ferramenta.

A multiplicidade de bases de dados implicou em um número maior de elementos de interface gráfica, para melhorar o processo cognitivo do usuário e evitar configurações erradas (usuário poderia se confundir pela similaridade dos controles) foram utilizadas bordas coloridas para interligar as visões gráficas à seus devidos controles de interface de usuário.

Um conceito adicional apresentado implica que a ferramenta tem de ser extensível ao longo de um número de monitores, a fim de contornar a dificuldade dos limites da área de visualização, visto que, novos conjuntos de dados são analisados simultaneamente.

Os testes realizados proveram resultados satisfatórios, principalmente pelo fato da média de acerto ter sido maior para o grupo de usuários utilizando coordenação entre bases de dados (8%), além de evidenciar a necessidade de uma revisão nos componentes de interface gráfica relacionados à coordenação entre bases de dados já que os usuários que utilizaram esta funcionalidade demoraram mais tempo para resolver as questões (24% de tempo a mais). Dos testes também notou-se que os usuários que utilizaram a coordenação afirmaram ter tido uma menor carga de trabalho dos que não utilizaram a coordenação.

7.1 Desafios Encontrados

Dos desafios encontrados no desenvolvimento deste trabalho destaca-se o desafio técnico de entender completamente e reformular o funcionamento da ferramenta PRISMA sem perda de funcionalidade, a correção de bugs originários dos novos conceitos implementados, e encontrar bases de dados correlacionáveis para testes e análises.

7.2 Trabalhos Futuros

Destacam-se os seguintes potenciais para desenvolvimentos futuros:

- Pesquisar e propor novos meios de coordenação entre bases de dados;
- Reformulação de como a coordenação entre bases de dados é apresentada ao usuário;
- Aplicação de novos ensaios de usabilidade;
- Criar um módulo de apresentação para web;
- Estudar meios de interação pelo uso de dispositivos móveis;
- Expandir a arquitetura para ser capaz de lidar com vários usuários ao mesmo tempo;
- Estudar novos cenários de uso sobre conjuntos de dados mais distintos;
- Aplicar novos testes envolvendo o uso de múltiplos displays e o impacto destes no desempenho dos usuários.

REFERÊNCIAS

- BALDONADO, M.; KUCHINSKY, A.; WOODRUFF, A. **Guidelines for Using Multiple Views in Information Visualization**. Advanced Visual Interfaces. [S.l.]: [s.n.]. 2000. p. 110-119.
- BARBOSA, S. D. J.; SILVA, B. S. D. **Interação Humano Computador**. Rio de Janeiro: Campus, 2010.
- BERTIN, J. **Semiology of Graphics**. [S.l.]: [s.n.], 1983.
- BRATH, R. K. **Effective Information Visualization Guidelines and Metrics for 3D Interactive Representations of Business**. [S.l.]. 1999.
- CARD, S.; MACKINLAY, S.; SHNEIDERMAN, B. **Readings in Information Visualization: Using Vision to Think**. [S.l.]: [s.n.], 1999.
- CARR, D. A. **Guidelines for Designing Information Visualization Applications**. Proceedings of Ericsson Conference in Usability Engineering. [S.l.]: [s.n.]. 1999. p. 1-7.
- DÖRK, M. et al. **VisGets: Coordinated Visualizations for Web-based Information Exploration and Discovery**. Proceedings of IEEE Symposium on Information Visualization. [S.l.]: [s.n.]. 2008. p. 1205-1212.
- ELISABETH FREEMAN, E. F. B. B. K. S. E. R. **Design Patterns**. [S.l.]: O'Reilly Media, 2004.
- FEW, S. **Now You See It: Simple Visualization Techniques for Quantitative Analysis**. [S.l.]: [s.n.], 2009.
- GODINHO, P. et al. **PRISMA - A Multidimensional Information Visualization Tool using Multiple Coordinated Views**. 11th International Conference on Information Visualization. Zurich: [s.n.]. 2007. p. 23-32.
- HART, S. G.; STAENLAND, L. E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: HANCOCK, P.; MESHKATI, N. **Human mental workload**. Amsterdam: North Holland: Elsevier, 1988. p. 183-139.
- HOEBER, O.; YANG, X. D. **Exploring Web Search Results Using Coordinated Views**. Coordinated and Multiple Views in Exploratory Visualization. [S.l.]: [s.n.]. 2006. p. 3 - 13.
- IBGE, P. N. P. A. D. D. **Síntese de Indicadores**. IBGE. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2009.
- INSELBERG, G. **Parallel Coordinates: Visual Multidimensional Geometry and Its Applications**. [S.l.]: Springer, 2009.

INSELBERG, A.; DIMSDALE, B. Parallel Coordinates: A Tool for Visualizing Multidimensional Geometry. **Proceedings of IEEE Visualization**, p. 361-375, 1990.

KERREN, A.; EBERT, A.; MEYER, J. Human-Centered Visualization Environments. [S.l.]: Springer, 2007. p. 403.

LENDERING, J. Peutinger Map. **Peutinger Map**, 01 jun. 2012. Disponível em: <<http://www.livius.org/pen-pg/peutinger/map.html>>. Acesso em: 01 jun. 2012.

LOURENÇO, R. et al. Using Multiple Datasets in Information Visualization Tool. In: SOBH, T.; ELLEITHY, K. **Innovations in Computing Sciences and Software Engineering**. [S.l.]: Springer, 2010. p. 601-606.

LOURENÇO, R. et al. **Exploring, Comparing and Coordinating Multiple Datasets in an Information Visualization Tool**. IV2012. Montpellier: IEEE Press. 2012.

MAZZA, R. **Introduction to Information Visualization**. [S.l.]: [s.n.], 2009.

NORTH, C.; SHNEIDERMAN, B. **Snap-Together Visualization**: A User Interface for coordinating Visualizations via Relational Schemata. *Advanced visual interfaces*. [S.l.]: [s.n.]. 2000. p. 23-26.

PILLAT, R. M.; VALIATI, E. R.; FREITAS, C. M. D. **Experimental Study on Evaluation of Multidimensional Information Visualization Techniques**. CLIHC'05. Cuernavaca - Mexico: [s.n.]. 2005. p. 20-30.

ROCHA, H. V. D.; BARANAUSKAS, M. C. C. **Design e Avaliação de Interfaces Humano-Computador**. Campinas: Instituto de Computação - Universidade Estadual de Campinas, 2003.

RUBIN, J. **Handbook of usability testing**. New York: John Wiley & Sons, 1994.

SHNEIDERMAN, B. Tree visualization with tree-maps: 2-d space-filling approach. **ACM Transactions on Graphics**, 1992.

SHNEIDERMAN, B. **The eyes have it**: a task by data type taxonomy for information visualizations. *Proceedings of the IEEE Symposium on Visual Languages*. [S.l.]: [s.n.]. 1996. p. 336 -343.

SHNEIDERMAN, B. Treemaps for Space-Constrained Visualization of Hierarchies. **HCIL - Department of Computer Science - University of Maryland**, p. 1-10, 01 jan. 2009. Disponível em: <<http://www.cs.umd.edu/hcil/treemap-history>>. Acesso em: 01 maio 2012.

SHNEIDERMAN, B.; WATTENBERG, . **Ordered Treemap Layouts**. *IEEE Symposium on Information Visualization*. [S.l.]: [s.n.]. 2001. p. 73-78.

SPENCE, R. **Information Visualization: Design for Interaction**. Second Edition. ed. Barcelona: Acm Press, 2007.

WARD, M. O.; GRINSTEIN, G.; KEIM, . **Interactive Data Visualization: Foundations, Techniques, and Applications**. [S.l.]: [s.n.], 2010.

WARE, C. **Information Visualization: Perception for Design** (Second Edition). [S.l.]: [s.n.], 2004.

WEAVER, C. Building Highly-Coordinated Visualizations In *Improvise*. **Proceedings of the IEEE Symposium on Information Visualization**, Austin, TX, 2004.

