

**UNIVERSIDADE DE CAXIAS DO SUL  
CENTRO DE COMPUTAÇÃO E TECNOLOGIA DA INFORMAÇÃO  
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**GLAUBER CINI**

**ESTUDO SOBRE A VISUALIZAÇÃO DE INFORMAÇÕES  
NO CONTEXTO EDUCACIONAL**

**Caxias do Sul  
2013**

**GLAUBER CINI**

**ESTUDO SOBRE A VISUALIZAÇÃO DE INFORMAÇÕES  
NO CONTEXTO EDUCACIONAL**

Trabalho de Conclusão de Curso para a  
obtenção do Grau em Bacharel em  
Ciência da Computação da Universidade  
de Caxias do Sul

Orientadora: Carine Geltrudes Webber

**Caxias do Sul  
2013**

## **AGRADECIMENTOS**

Agradeço a minha mãe Clacir e meu pai Roberto, pelo incentivo ao estudo durante todas as etapas da minha vida. A minha irmã Gláucia pelo alicerce psicológico e exemplo que ela representa. E a minha orientadora, Carine, pela paciência, tempo e conhecimento dedicados durante todas as etapas desse processo.

*“Deves aprender as regras do jogo. E depois deves jogar melhor que todo mundo.”*

**Albert Einstein**

## RESUMO

Este trabalho apresenta estudos realizados sobre a Visualização de Informação, abordando um modelo de construção e suas principais técnicas para a geração de visualizações (visões). A visualização de informação caracteriza-se por englobar diversas técnicas com a finalidade de organizar, encontrar e descobrir informações em bases de dados altamente dimensionadas. O objetivo principal deste trabalho é avaliar as técnicas de visualização de informação determinando quais são adequadas para serem utilizadas em dados educacionais. O trabalho apresenta estudos realizados sobre as principais técnicas e seus respectivos resultados gráficos. Para atingir o objetivo proposto utilizou-se dados disponibilizados publicamente sobre a educação, todos referentes à avaliação de alunos e situados em três áreas distintas, o primeiro é o *Geometry* com dados de um curso de geometria, o segundo é o conjunto de dados da Língua Chinesa com dados de quem está aprendendo uma nova língua, e por último o Álgebra I composto por resolução de equações. Também, utilizou-se para a criação das visões a ferramenta GGobi e o conjunto de aplicações de Hoffman (MIVAC e RadViz3D). Selecionou-se ambas as ferramentas por apresentarem uma melhor iteração entre usuário-computador com as visões selecionadas (*Scatterplot Matrix*, RadViz, Coordenadas Paralelas e *Survey Plots*). Por fim, analisou-se as amostras dos conjuntos selecionados, utilizando as visões propostas. Obtidas as análises, criou-se três cenários. Os quais foram utilizados para entrevistar professores. A fim de discutir e avaliar a usabilidade das técnicas selecionadas.

**Palavras-chave:** visualização de informação, gráficos, visões, dados educacionais, ferramentas de visualização.

## ABSTRACT

This work presents studies about Information Visualization, approaching a building model and its key techniques to generate visualizations (views). Information visualization is characterized by encompass several techniques in order to organize, find and discover information in highly scaled databases. The main objective of this work is to evaluate the information visualization techniques determining which proposes the best decision in a pedagogical context. The large volume of educational data available hampers its manual analysis, so the techniques of information visualization can aid in this analysis. This work presents studies on the key techniques and their graphical results. To reach that goal we used publicly available data on education, all referring to the evaluation of students and situated in three distinct areas, the first is the Geometry with data from a geometry course, the second is the data set of Chinese Language data with those learning a new language, and lastly the Algebra I composed of solving equations. Was also used to create the views GGobi tool and the Hoffman set applications (MIVAC and RadViz3D). These tools were chosen because they have a better iteration between user-computer with the proposed visions (Scatterplot Matrix, RadViz, Parallel Coordinates and Survey Plots). Finally, we analyzed samples from selected sets using the proposed vision. Once analyzes was obtained, we created three scenarios. Which were used to interview professors. In order to discuss and evaluate the usability of selected techniques.

**Keywords:** information visualization, graphics, visions, educational data, visualization tools.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Um diagrama esquemático do processo de visualização .....	17
Figura 2 - Modelo de Referência de Visualização .....	18
Figura 3 - Primeira adaptação do modelo de Card para a transformação de dados.	19
Figura 4 - Segunda adaptação do modelo de Card para a transformação de dados. .....	23
Figura 5 - Adaptação do dicionário de Glifos.....	24
Figura 6 - Adaptação das Propriedades Visuais .....	24
Figura 7 - Terceira adaptação do modelo de Card para a Transformação das Visões .....	25
Figura 8 - Coordenadas Paralelas exemplo com carros em quatro dimensões .....	28
Figura 9 - Aplicação de cores em coordenadas paralelas.....	29
Figura 10 - Exemplo de Matriz de <i>Scatterplots</i> .....	30
Figura 11 - Exemplo da técnica RadViz .....	31
Figura 12 - Coordenadas Paralelas Circulares e PolyViz.....	32
Figura 13 - Exemplo de <i>Survey Plots</i> para quatro dimensões .....	32
Figura 14 - Esboço da técnica Chernoff <i>faces</i> .....	33
Figura 15 - Adaptação das Faces de Chernoff.....	34
Figura 16 - Adaptação para <i>Star Glyphs</i> com sete dimensões .....	35
Figura 17 - Conjunto de ícones gerados pela técnica de <i>Stick Figure</i> .....	35
Figura 18 - <i>Stick Figure</i> utilizando uma amostra populacional .....	36
Figura 19 – Dimensões adicionais adaptadas para a Figura 18.....	37
Figura 20 - Construção de telas para seis dimensões .....	38
Figura 21 – Cem ações previstas no Índice de Ações de Frankfurt .....	39
Figura 22 - Construção de segmentos para seis dimensões .....	39
Figura 23 - Segmentos de círculo para oito dimensões .....	40
Figura 24 - Cinquenta ações previstas no Índice de Ações de Frankfurt .....	41
Figura 25 - Duas diferentes maneiras de visualizar os dados.....	42
Figura 26 – Exemplo de <i>Dimensional Stacking</i> .....	42
Figura 27 - Exemplo de árvore representada por uma <i>Treemap</i> .....	43
Figura 28 - <i>n-Vision</i> , exemplo para cinco dimensões.....	44
Figura 29 - Exemplo de grafo com distribuição otimizada e acíclico .....	45

Figura 30 - Exemplo de grafo ortogonal e otimizado para simetria .....	46
Figura 31 - Exemplo de grafos em três dimensões .....	46
Figura 32 - Parvis, exemplo de interface .....	59
Figura 33 - Aplicação MIVAC de Hoffman, utilizando Segmentos de Círculo .....	60
Figura 34 - GGobi - exemplo de visões geradas .....	61
Figura 35 - <i>Dimensional Stacking</i> utilizando a XmdvTool e dados do IrisDB .....	62
Figura 36 - InfoVis <i>Toolkit</i> - Exemplo de <i>Treemap</i> .....	63
Figura 37 – Adaptação do Modelo de Card et al. (1999).....	66
Figura 38 – <i>Layout</i> de importação de <i>xls</i> .....	67
Figura 39 – Exibição dos dados contidos no arquivo .....	68
Figura 40 – Seleção de colunas para transformação. ....	68
Figura 41 – Opções de transformação (exportação) .....	69
Figura 42 – Ilustração do problema contido na amostra .....	70
Figura 43 – <i>Brush</i> aplicado à dimensão Duração.....	76
Figura 44 – <i>Brush</i> aplicado à dimensão Resultado da Etapa.....	76
Figura 45 – <i>Brush</i> aplicado à dimensão Etapa a Resolver.....	77
Figura 46 – PolyViz sobre amostra de dados Geometria .....	78
Figura 47 – RadViz3D sobre amostra de dados Geometria .....	79
Figura 48 – RadViz sobre a amostra de dados Geometria .....	79
Figura 49 – Visão geral da amostra nas Coordenadas Paralelas .....	81
Figura 50 – Utilização de <i>brush</i> no primeiro estudante .....	81
Figura 51 – Utilização de <i>brush</i> na duração.....	82
Figura 52 – Utilização de <i>brush</i> na Área de Conhecimento Avaliada.....	82
Figura 53 – Utilização de <i>brush</i> na dimensão Vezes na Etapa .....	82
Figura 54 – <i>Survey Plots</i> , exemplo de funcionamento .....	83
Figura 55 – Análise realizada no <i>Survey Plots</i> .....	84
Figura 56 – <i>Brush</i> aplicado à dimensão Id Estudante .....	92
Figura 57 – Análise sobre o plano Número de Repetições por Id Estudante .....	92
Figura 58 – Análise sobre o plano Id Estudante por Resposta.....	92
Figura 59 – <i>Brush</i> aplicado à dimensão Resposta .....	93
Figura 60 – <i>Brush</i> aplicado ao plano Resposta por Passo.....	93
Figura 61 – Análise sobre o plano Número de Repetições por Id Estudante .....	93
Figura 62 – <i>Brush</i> aplicado à dimensão Resposta .....	94
Figura 63 – Análise feita sobre quatro planos diferentes .....	94

Figura 64 – PolyViz sobre amostra de dados da Língua Chinesa.....	95
Figura 65 – RadViz3D sobre amostra de dados da Língua Chinesa.....	96
Figura 66 – RadViz sobre a amostra de dados da Língua Chinesa .....	96
Figura 67 – Utilização de <i>brush</i> no quarto passo .....	97
Figura 68 – <i>Brush</i> na terceira Seção.....	98
Figura 69 – Visualização da amostra utilizando Coordenadas Paralelas.....	98
Figura 70 – <i>Brush</i> aplicado à dimensão Resposta .....	105
Figura 71 – Análise do plano Resposta por Número da Tentativa .....	105
Figura 72 – Análise sobre o plano Resposta por Passo.....	106
Figura 73 – Análise sobre o plano Passo por Duração e pontos correspondentes .	106
Figura 74 – Análise sobre o plano Seção por Passo e pontos correspondentes no plano Resposta .....	107
Figura 75 – PolyViz sobre amostra de dados Álgebra I .....	108
Figura 76 – RadViz3D sobre amostra de dados Álgebra I .....	108
Figura 77 – RadViz sobre amostra de dados Álgebra I.....	109
Figura 78 – <i>Brush</i> sobre a dimensão Duração .....	110
Figura 79 – <i>Brush</i> na dimensão Número da Tentativa .....	110
Figura 80 – <i>Brush</i> em outros pontos da dimensão Número da Tentativa .....	111
Figura 81 – <i>Brush</i> na dimensão Resposta .....	111
Figura 82 – Distorção por <i>brush</i> para trocas de níveis de dicas.....	112
Figura 83 – Visão simulada no primeiro cenário .....	114
Figura 84 – Visão simulada no segundo cenário.....	116
Figura 85 – Visão simulada no terceiro cenário .....	117

## LISTA DE TABELAS

Tabela 1 - Classificação de informações.....	20
Tabela 2 - Tipos de dados sugeridos .....	21
Tabela 3 - Amostra de dados do conjunto <i>Geometry</i> .....	49
Tabela 4 - Fatores de Dificuldade .....	51
Tabela 5 - Amostra de dados da Língua Chinesa .....	52
Tabela 6 - Amostra de dados agrupados por um problema .....	54
Tabela 7 – Relação de ferramentas para Visualização de Informação .....	56
Tabela 8 – Relação de respostas do problema <i>dog on a rope</i> .....	71
Tabela 9 – Amostra selecionada do conjunto Geometria.....	71
Tabela 10 - Resultado da preparação da amostra Geometria.....	74
Tabela 11 – Amostra selecionada do conjunto de dados da Língua Chinesa .....	85
Tabela 12 – Resultado da preparação da amostra do conjunto de dados da Língua Chinesa.....	89
Tabela 13 – Amostra selecionada do conjunto Álgebra I .....	100
Tabela 14 – Resultado da preparação da amostra Álgebra I .....	102

## LISTA DE SIGLAS

CMU	<i>Carnegie Mellon University</i>
CSV	<i>Comma-Separated Values</i>
CWCTC	Central Westmoreland Carreira e Centro de Tecnologia
DA	<i>Dimensional Anchors</i>
ETL	<i>Extract, Transform and Load</i>
InfoVis	<i>Information Visualization</i>
IVTK	<i>Information Visualization ToolKit</i>
PSLC	<i>Pittsburgh Science of Learning Center</i>
BI	<i>Business Intelligence</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>14</b>
1.1 OBJETIVOS .....	15
1.2 ESTRUTURA DO TRABALHO .....	16
<b>2 VISUALIZAÇÃO DE INFORMAÇÃO</b> .....	<b>17</b>
2.1 TRANSFORMAÇÃO DOS DADOS .....	19
2.2 MAPEAMENTO VISUAL .....	22
2.3 TRANSFORMAÇÃO DAS VISÕES .....	25
2.4 CONSIDERAÇÕES FINAIS .....	26
<b>3 TÉCNICAS PARA VISUALIZAÇÃO EXPLORATÓRIA DE DADOS</b> .....	<b>27</b>
3.1 TÉCNICAS DE PROJEÇÃO GEOMÉTRICA .....	27
3.2 TÉCNICAS BASEADAS EM ÍCONES .....	33
3.3 TÉCNICAS ORIENTADAS A <i>PIXEL</i> .....	37
3.4 TÉCNICAS HIERÁRQUICAS .....	41
3.5 TÉCNICAS BASEADAS EM GRAFOS .....	45
3.6 CONSIDERAÇÕES FINAIS .....	46
<b>4 CONJUNTOS DE DADOS DO EXPERIMENTO</b> .....	<b>48</b>
4.1 CONJUNTO DE DADOS GEOMETRIA .....	48
4.2 CONJUNTO DE DADOS DA LÍNGUA CHINESA .....	51
4.3 CONJUNTO DE DADOS ÁLGEBRA .....	53
4.4 CONSIDERAÇÕES FINAIS .....	54
<b>5 FERRAMENTAS</b> .....	<b>56</b>
5.1 CRITÉRIOS .....	56
5.2 PARVIS .....	58
5.3 APLICAÇÕES DE HOFFMAN .....	59
5.4 GGOBI .....	60
5.5 XMDVTOOL .....	61
5.6 INFOVIS <i>TOOLKIT</i> .....	63
5.7 CONSIDERAÇÕES FINAIS .....	64
<b>6 EXPERIMENTO</b> .....	<b>65</b>
6.1 MODELO .....	65

6.2 DIMENSIONADOR.....	66
6.3 GEOMETRIA.....	69
<b>6.3.1 Scatterplot Matrix.....</b>	<b>75</b>
<b>6.3.2 RadViz, RadViz3D e PolyViz.....</b>	<b>77</b>
<b>6.3.3 COORDENADAS PARALELAS.....</b>	<b>80</b>
<b>6.3.4 Survey Plots.....</b>	<b>83</b>
6.4 LÍNGUA CHINESA.....	84
<b>6.4.1 Scatterplot Matrix.....</b>	<b>91</b>
<b>6.4.2 RadViz, RadViz3D e PolyViz.....</b>	<b>95</b>
<b>6.4.3 Coordenadas Paralelas.....</b>	<b>97</b>
<b>6.4.4 Survey Plots.....</b>	<b>99</b>
6.5 ÁLBEBRA I.....	99
<b>6.5.1 Scatterplot Matrix.....</b>	<b>104</b>
<b>6.5.2 RADVIZ, RADVIZ3D E POLYVIZ.....</b>	<b>107</b>
<b>6.5.3 Coordenadas Paralelas.....</b>	<b>110</b>
<b>6.5.4 Survey Plots.....</b>	<b>112</b>
6.6 TESTES BASEADOS EM CENÁRIOS SIMPLIFICADOS.....	113
<b>6.6.1 Primeiro cenário.....</b>	<b>113</b>
<b>6.6.2 Segundo cenário.....</b>	<b>115</b>
<b>6.6.3 Terceiro cenário.....</b>	<b>117</b>
<b>7 CONCLUSÃO.....</b>	<b>119</b>
7.1 SÍNTESE DO TRABALHO.....	119
7.2 RESULTADOS E CONTRIBUIÇÕES.....	120
7.3 TRABALHOS FUTUROS.....	122
<b>REFERÊNCIAS.....</b>	<b>123</b>

## 1 INTRODUÇÃO

A visualização de informações, definida por Card et al. (1990), é a representação visual de dados abstratos de forma interativa, baseada no computador para amplificar a cognição. Atualmente pode-se gerar rapidamente uma representação visual com o uso do processamento do computador e, assim, oferecer ao usuário a habilidade de não somente obter uma apresentação clara e compreensível, mas também de modificá-la, a fim de permitir o destaque dos pontos de seu interesse (Savitraz, 2010).

Sendo assim, esta representação deve expressar as suas propriedades mais relevantes, mostrando como os diferentes itens – dimensões - estão relacionados entre si. A projeção de uma representação que funcione para grandes conjuntos de dados é um problema frequente (Cava, 2002).

Para facilitar a análise de toda essa informação, artefatos que diminuem a densidade dos dados e tentam aumentar a compreensão das informações como um todo são utilizados. Keim (2002) cita exemplos visuais como: gráficos de coordenadas x-y, gráficos de linhas e histogramas. Já Shneiderman (2001) cita exemplos numéricos como: medianas, desvios-padrão e intervalos.

Contudo esses meios podem esconder determinadas informações, ou não apresentar de forma clara como ela pode estar relacionada, evitando assim a subjetividade humana.

Long (2009) cita também o problema do limitado tamanho das telas, onde o número de pixels muitas vezes excede o da resolução. Logo, não é possível que todos os valores sejam apresentados simultaneamente. Um segundo ponto destacado por ele é quando o número de dimensões excede de três, destacando a problemática das ferramentas ainda quando tratando de apenas duas dimensões.

Adendo a isso, Keim (2002) expõe o problema da proporção, onde cita que avanços tecnológicos possibilitaram que computadores pudessem armazenar imensas quantidades de dados. Desta forma, a capacidade de coletar e armazenar essas mesmas informações estão aumentando a uma velocidade superior do que se consegue analisá-las.

Nesta mesma direção, a visualização de informações tem como objetivo auxiliar no entendimento de determinados assuntos, o qual, sem uma representação visual, exigiria um maior esforço para a compreensão (Gondim; Machado; Santos, 2008).

A visualização de informação também é utilizada na área da educação. Merceron e Yacef (2004) dizem que, estatísticas simples, consultas ou algoritmos de visualização são úteis para dar aos professores uma visão geral do andamento da classe. Gueraud (2004) usa cenários pedagógicos para controlar objetos de aprendizagem interativos. Nestes cenários, registros são usados para construir gráficos que mostram exatamente onde cada aluno está na sequência de aprendizagem, oferecendo assim um acompanhamento de longa distância para o professor.

Da mesma forma em Duval (2005), as respostas dos exercícios feitos pelos alunos são gravadas. Simples consultas permitem mostrar gráficos aos professores de todos os alunos com os exercícios que eles resolveram ou não com sucesso, tornando os professores mais conscientes de como anda o progresso de cada um durante o curso. Porém técnicas de visualização avançadas para explorar esses dados são sugeridas por Mazza e Dimitrova (2003) onde é possível externalizar dados de alunos e gerar representações pictóricas exploratórias para os instrutores do curso.

Com a capacidade cognitiva humana auxiliada por um modelo de visualização de informação focado na área da educação, surge uma maneira diferente de se avaliar dados com maior excelência, com maior ênfase na descoberta. Desta forma é possível extrair-se um maior número de informações com maior relevância.

## 1.1 OBJETIVOS

O objetivo deste trabalho de conclusão de curso é realizar um estudo exploratório dos tipos de visualização de informação que podem ser utilizados em bases de dados educacionais e dos tipos de ferramentas existentes que possam ser utilizadas nestes casos.

A fim de atingir o objetivo geral, este trabalho está dividido em quatro objetivos específicos.

- Pesquisar como a Visualização de Informação pode ser aplicada a um conjunto de dados, buscando novos resultados e maneiras de tornar a visualização mais clara;
- Identificar os tipos de visualização mais apropriados para uso em Bases de Dados Educacionais;
- Pesquisar ferramentas ou bibliotecas que permitam a visualização da informação em Bases de Dados Educacionais;
- Aplicar as ferramentas selecionadas aos dados educacionais.

## 1.2 ESTRUTURA DO TRABALHO

O trabalho está estruturado em cinco capítulos, conforme descrito a seguir.

O primeiro capítulo procura introduzir o leitor ao assunto abordado, apresentando a estrutura do trabalho e seus objetivos.

O capítulo 2 visa, através de modelos, conceituar o processo de Visualização de Informação.

O capítulo 3 apresenta a formulação das principais técnicas Visualização de Informação, apresentando sua construção e suas visões resultantes.

Na continuação, o capítulo 4 apresenta três conjuntos de dados, os quais as técnicas de visualização serão empregadas.

O capítulo 5 apresenta as principais ferramentas que podem auxiliar na geração das visões desejadas.

O capítulo 6 apresenta cenários simplificados utilizando algumas técnicas selecionadas e aplicadas sobre amostras dos conjuntos de dados. Os cenários foram demonstrados a professores, e o resultado desta entrevista também consta no capítulo.

Por fim, o capítulo 7 apresenta as conclusões finais do trabalho.

## 2 VISUALIZAÇÃO DE INFORMAÇÃO

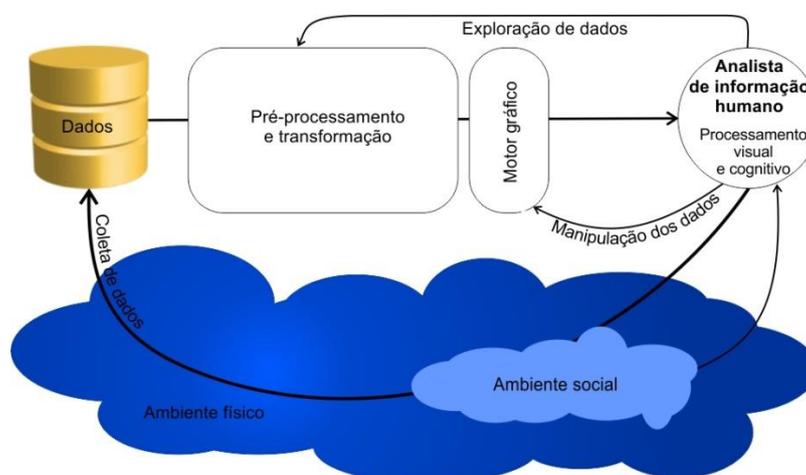
Representações gráficas têm sido usadas como instrumento de comunicação desde os primórdios da humanidade. Exemplo disso são os hieróglifos, os quais provavelmente formam o sistema organizado de escrita mais antigo do mundo. Com o advento da ciência, as representações gráficas passaram a embutir um significado cada vez mais regido por convenções, a exemplo de gráficos matemáticos e cartas cartográficas.

Normalmente, essas representações têm como propósito comunicar uma ideia já existente. Todavia, tendo como propósito aproveitar as especiais características da percepção visual humana – cognição - para a resolução de problemas lógicos, uma segunda abordagem possível consiste em utilizar as representações gráficas para criar ou descobrir a própria ideia (Savitraz, 2010).

A Visualização de Informação é definida por Keim et al. (2006), como a comunicação de dados abstratos relevantes em termos de ação através do uso de interfaces visuais interativas, ou ainda como um método que combina aspectos da visualização científica, interação humano computador, *data mining*, imagem e gráficos, entre outros (Börner, 2001).

Para Ware (2012) o processo de visualização está definido em quatro etapas básicas conforme ilustra a Figura 1.

Figura 1 - Um diagrama esquemático do processo de visualização



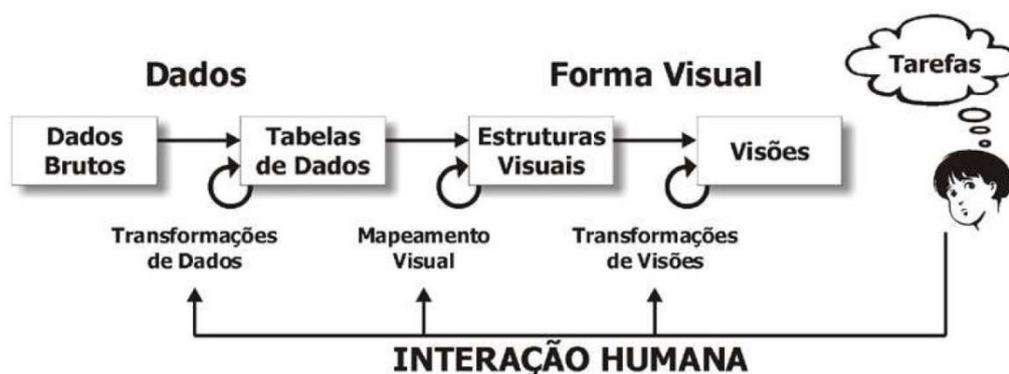
Fonte: Ware (2012).

A primeira etapa consiste em coletar e armazenar os dados. A segunda em um pré-processamento induzido, transformando os dados em algo logicamente organizado. A terceira é composta pelo *hardware* e o algoritmo gráfico que produz uma visualização. A última etapa é composta pelo perceptual humano e o sistema cognitivo, que terá o entendimento dos dados facilitado, melhorando assim a tomada de decisões e a formação de hipóteses.

Dentre as interações existentes na Figura 1, Ware (2012) salienta duas, a primeira a respeito da importância da coleta de dados, onde sempre que a informação se apresentar relevante, o analista pode decidir por adicionar ou remover conjuntos de dados. A segunda se refere à exploração de dados, iniciada no pré-processamento computacional que ocorre antes da visualização. O analista, sabendo que nesse momento os dados já foram processados, tem a possibilidade de consultá-los, assim ele conhece o conjunto e caso uma visualização errada seja gerada, ele conseguirá distinguir mais facilmente, não desistindo do significado dos dados.

Card et al. (1999) sugere um modelo (Figura 2) semelhante ao de Ware (2012), porém, este evidencia outros dois fatores. O primeiro corresponde aos procedimentos que ocorrem entre as etapas. O segundo é a existência de uma iteração no processo em si.

Figura 2 - Modelo de Referência de Visualização



Fonte: Card et al. (1999).

Neste modelo a informação segue um fluxo que vai dos Dados Brutos se propaga até o humano. A cada etapa, uma série de transformações são aplicadas. Já as interações do usuário controlam os parâmetros dessas transformações, que,

por exemplo, podem restringir a visão para certos intervalos de dados, ou trocar a natureza da transformação.

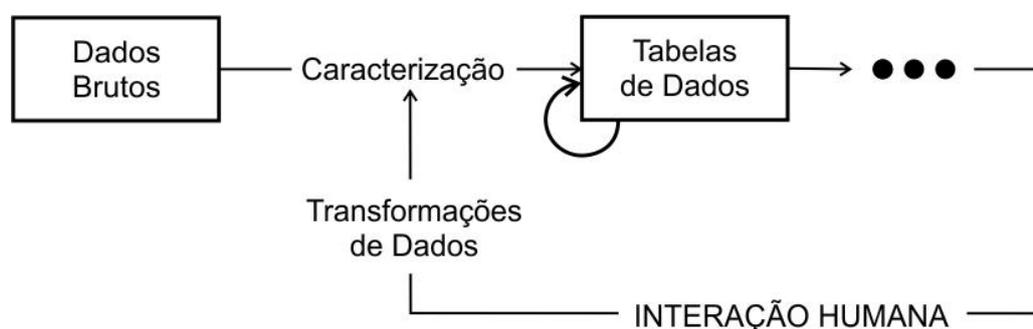
A transformação de dados mapeia os Dados Brutos para as Tabelas de Dados, onde relações podem ser criadas para estender a descrição e significado destes dados. O Mapeamento Visual transforma as Tabelas de Dados em Estruturas Visuais. Estas estruturas combinam suportes espaciais, marcas, propriedades gráficas, dimensões e seleções. Finalmente, as Transformações Visuais criam as Visões das Estruturas Visuais especificando parâmetros gráficos como posição, dimensão e seleção.

Ao longo das próximas seções o modelo sugerido por Card et al. (1999), é detalhado.

## 2.1 TRANSFORMAÇÃO DOS DADOS

Os Dados Brutos são dados que não sofreram nenhum processo computacional. Eles podem ser coletados de diversas fontes como: sensores, planilhas, formulários e textos. Após a coleta, os dados seguirão o fluxo, onde sofrerão o processo de caracterização (Figura 3).

Figura 3 - Primeira adaptação do modelo de Card para a transformação de dados.



Fonte: Card et al (1990).

A caracterização dos dados, por definir a classe – tipo – e a dimensão de cada conjunto de dados, acaba sendo uma das considerações iniciais na escolha de

uma técnica de visualização. Freitas et al. (2001) apresentam uma proposta de caracterização baseada em critérios como: classe da informação, tipos dos valores, natureza e dimensão do domínio (conforme resumo Tabela 1).

Tabela 1 - Classificação de informações

Critério	Classe	Exemplo
Classe da informação	Característica	Gênero
	Escalar	Temperatura
	Vetor	Grandeza física associada a um fluido
	Tensor	Grandeza física associada a um fluido
	Relacionamento	<i>Link</i> num hiperdocumento
Tipo dos valores	Alfanumérico	Gênero
	Númérico	Temperatura
	Símbolo	<i>Link</i> num hiperdocumento
Natureza do domínio	Discreto	Marcas de automóveis
	Contínuo	Superfície de um terreno
	Contínuo-Discretizado	Anos (tempo discretizado)
Dimensão do domínio	1D	Medida de uma grandeza no tempo
	2D	Superfície de um terreno
	3D	Volume de dados médicos
	n-D	Dados de uma população

Fonte: FREITAS et al., 2001. p. 3.

De acordo com o critério classe de informação, os atributos podem representar: a) uma característica, como uma categoria, atributo nominal ou ordinal (Ware, 2012). O que simboliza valores pertencentes a um grupo restrito de elementos, podendo ou não apresentar uma ordem natural entre seus elementos (Cart et al., 1999); b) uma propriedade com valores escalares, vetoriais ou tensoriais, que dentro de um intervalo podem assumir valores como inteiros ou reais; ou c) a existência de relacionamentos (hierarquia ou ligação) entre entidades.

Quanto ao tipo, os atributos podem assumir valores alfanuméricos ou numéricos, dentre uma enumeração finita ou infinita, ou podem ser compostos de valores simbólicos, representando a identificação de uma entidade ou fenômeno relacionado (Freitas, 2005).

A natureza do domínio indica se um dado está definido num domínio discreto, restrito a um conjunto finito ou infinito de elementos, num domínio contínuo ou num domínio contínuo-discretizado. Em um domínio contínuo, os valores podem

estar definidos continuamente ou para superfícies. Por exemplo, uma amostragem no tempo (que constitui um domínio contínuo) quando feita mensalmente indica um domínio contínuo-discretizado.

Os dados podem, também, ser categorizados segundo o critério dimensão (Luzzardi, 2003). A dimensionalidade é usada para referir tanto o número de dimensões do espaço-domínio onde determinado objeto está definido, podendo ser um espaço unidimensional (1D), bidimensional (2D) ou tridimensional (3D), quanto o número de atributos de uma entidade, geralmente, referenciado como n-dimensional (nD), multidimensional ou multivariável. Assim, visualizações multidimensionais referem-se à visualização de informações que contêm múltiplas variáveis a serem codificadas em uma única estrutura visual (Card et al., 1999).

De maneira geral, os valores assumidos pelas variáveis no momento da caracterização podem ser enquadrados seguindo os três tipos básicos sugeridos por Card et al. (1999) que constam na Tabela 2.

Tabela 2 - Tipos de dados sugeridos

Tipo da Variável	Relação	Exemplo
Nominais	São somente iguais (=) ou diferentes ( $\neq$ ) dos outros valores.	Qualquer conjunto desordenado, como, por exemplo, títulos de filmes {Guerra nas Estrelas, Instinto, Carandiru}.
Ordinais	Obedecem a relações de ordem como maior (>) ou menor (<)	Qualquer conjunto ordenado, como, por exemplo, uma avaliação {Ruim, Médio, Bom, Muito Bom}.
Quantitativos	Podem sofrer em seus elementos operações aritméticas.	Quantidade [0, n] Tamanho [0, 10] Duração [0, 360]

Fonte: Card et al. (1999).

Após a caracterização, como mostra a Figura 3, os dados podem sofrer transformações, onde são preparados para serem armazenados nas Tabelas de Dados.

Tweedie (1997) afirma que o processo de transformar dados brutos em tabelas de dados envolve normalmente o ganho ou perda de dados. Por exemplo, dados brutos podem conter erros em alguns pontos, ou, simplesmente ter valores faltando. Ambos devem ser completados ou ajustados antes de serem enviados para

as tabelas de dados. Outro exemplo citado por ele são cálculos estatísticos, que podem derivar outro conjunto de informações.

Para facilitar o entendimento, Tweedie (1997) propõe quatro tipos de transformações:

- Valores para valores derivados: cálculos estáticos, transformações numéricas.
- Estruturas para estruturas derivadas: ordenação, classificação ou promoção.
- Valores para estruturas derivadas: um filme que nos dados brutos tem a informação de tempo gravada num intervalo de  $[0, 360]$  pode ser transformado para uma nova estrutura separando, por exemplo, em [Curto, Médio, Longo].
- Estruturas para valores derivados: uma estrutura que contém, por exemplo, todos os gastos de uma pessoa, podem ser transformados para um único valor quando somado.

Após definida as transformações, essas informações devem ser armazenadas em locais logicamente organizados chamados de Tabelas de Dados.

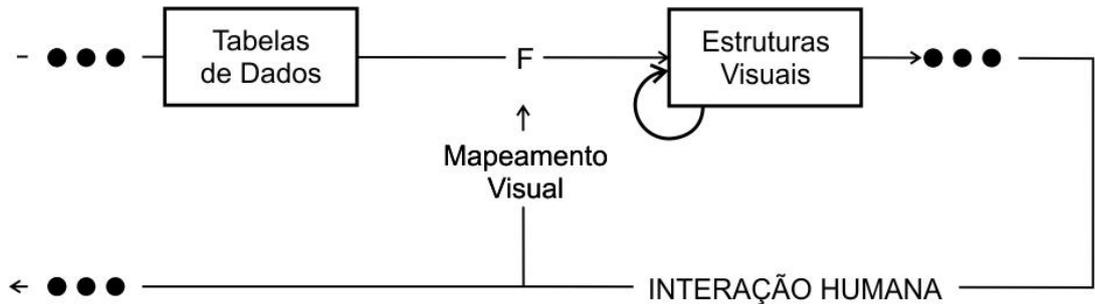
As Tabelas de Dados podem ser definidas simplesmente como dados organizados. Nas tabelas, os dados são dispostos de maneira lógica em um formato de linha-e-coluna semelhante ao de uma planilha. Cada linha representa um registro exclusivo e cada coluna representa um campo no registro (Microsoft, 2013).

## 2.2 MAPEAMENTO VISUAL

O mapeamento visual é o ponto central da Visualização de Informação. Sua função é de mapear as Tabelas de Dados em Estruturas Visuais tendo como objetivo comunicar informações do computador para o ser humano, utilizando uma representação visual como meio de comunicação (Figura 4).

Para isso um conjunto de dados é computacionalmente mapeado em forma visual por alguma função  $F$ , que leva o conjunto de dados como entrada e gera a representação visual como saída.

Figura 4 - Segunda adaptação do modelo de Card para a transformação de dados.



Fonte: Card et al. (1990).

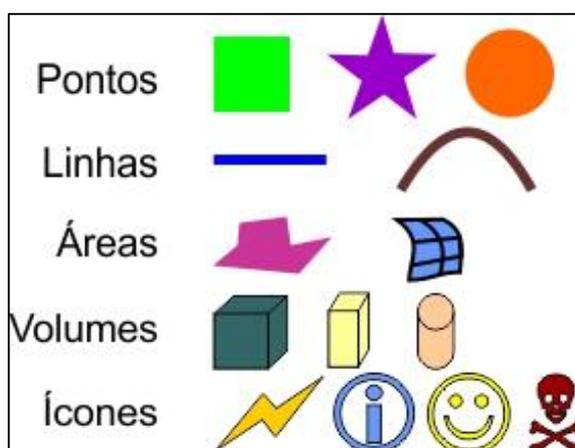
North (2005) cita quatro características importantes da função  $F$ :

- Computável:  $F$  é uma função matemática que pode ser executada por um algoritmo. Embora haja um espaço significativo para a criatividade na concepção dessas funções, a execução delas deve ser algorítmica;
- Invertível: Deve ser possível a utilização de  $F^{-1}$ , o inverso da função  $F$ , para reconstruir os dados da representação visual a um grau de precisão desejado. Se isso não for possível, a visualização será ambígua, enganosa, ou não interpretável;
- Comunicável:  $F$  (ou, de preferência  $F^{-1}$ ) deve ser conhecida pelo usuário para que ele possa interpretar a representação visual. Esta função pode ser comunicada junto à visualização, ou já conhecida por ele através de experiências anteriores. Em termos de usabilidade, esta é uma questão de capacidade de aprendizado;
- Cognoscível:  $F^{-1}$  deve minimizar a carga cognitiva para decodificar a representação visual. Esta é uma questão de percepção e desempenho humano.

Além da função  $F$ , o mapeamento visual é composto por espaços visuais (áreas que serão utilizadas para a plotagem), glifos e propriedades visuais.

Os Glifos, também comumente chamados de marcas, mapeiam cada entidade de dado em uma entidade visual, seguindo um dicionário de Glifos conforme mostra a Figura 5.

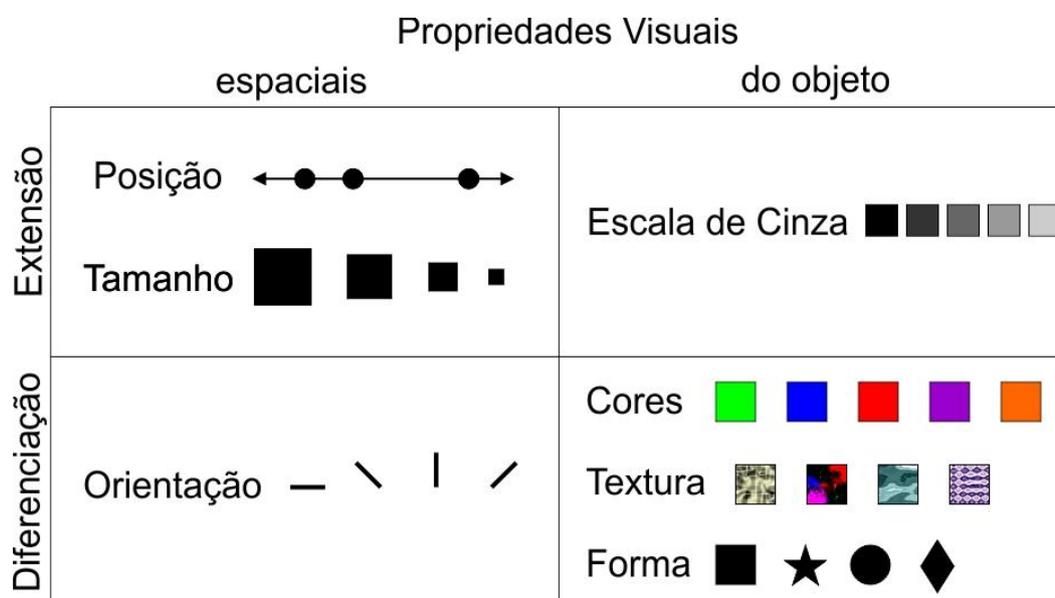
Figura 5 - Adaptação do dicionário de Glifos



Fonte: North (2005).

Após mapear as entidades em glifos, deve-se definir o comportamento delas no espaço, para isso são utilizadas “Propriedades Visuais”. Essas propriedades incluem posição espacial, tamanho, cor, orientação e forma, as quais irão mapear os valores dos atributos de cada entidade (Figura 6).

Figura 6 - Adaptação das Propriedades Visuais



Fonte: Card et al. (1999) e North (2005).

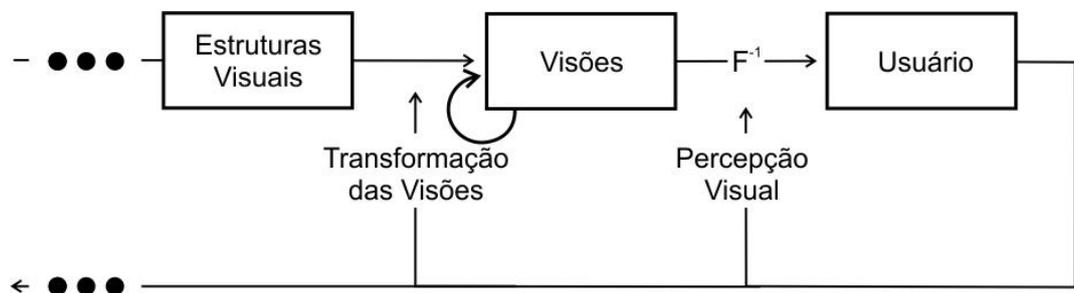
North (2005) também cita outras propriedades visuais, incluindo em sua relação: movimento, *blink* (piscar), densidade e transparência.

O resultado desses mapeamentos para ser efetivo, deve proporcionar uma rápida interpretação ao ser humano, transmitindo um número maior de distinções ou um menor número de erros do que outros mapeamentos.

### 2.3 TRANSFORMAÇÃO DAS VISÕES

As Transformações das Visões representam a última etapa humano-computacional no processo de Visualização de Informação (Figura 7). Elas são responsáveis por tornar desenhos estáticos vindos das Estruturas Visuais em Visões. Assim os usuários tem a possibilidade de selecionar parâmetros para que estas sejam modificadas utilizando uma mesma estrutura.

Figura 7 - Terceira adaptação do modelo de Card para a Transformação das Visões



Fonte: Card et al. (1999)

Existem três transformações comuns sobre as Visões, são elas:

- Testes de localização permitem obter informações adicionais sobre um item da tabela de dados;
- Controles de ponto de vista, para *zoom* e troca da imagem, de modo a oferecer diferentes visões. Uma técnica chamada visão geral+detalhe (*overview + detail*) também é um tipo de controle de ponto de vista (Shneiderman, 1996);
- Distorção - distorções da imagem visando criar visualizações do tipo foco+contexto (Card et al., 1999).

## 2.4 CONSIDERAÇÕES FINAIS

Neste capítulo, procurou-se apresentar um modelo de referência para a área de visualização de informação, apresentando as etapas de transformações aplicadas aos dados brutos até se tornarem visões. Também se apresentou pontos onde o usuário pode interagir durante todas as fases desse processo.

Com o modelo referenciado neste capítulo, os dados citados no Capítulo 4 foram adaptados respeitando todas as etapas de transformação até que fosse possível a geração das visões (Capítulo 3). Seguir o modelo proposto de Card et al. (1999) é de suma importância, caso contrário o resultado pode ser a dificuldade ou distorção da construção das dimensões e conseqüentemente, a geração de visões de baixa credibilidade.

No próximo capítulo, serão demonstradas como estas visões devem ser construídas, apresentando os grupos de técnicas mais comumente citados nas bibliografias.

### 3 TÉCNICAS PARA VISUALIZAÇÃO EXPLORATÓRIA DE DADOS

Desde que os computadores começaram a ser usados para criar visualizações, novas técnicas têm sido desenvolvidas. Essas foram estendidas para trabalhar com conjuntos de dados maiores e possibilitando a interatividade.

Para a maior parte dos dados armazenados em uma base, no entanto, não existe qualquer padrão de mapeamento para o sistema de coordenadas cartesiano, uma vez que os dados não são inerentes à semântica de duas ou três dimensões. Em geral, as Tabelas de Dados organizadas de forma lógica – também chamadas de bases de dados relacionais –, formam conjuntos de dados multidimensionais junto aos atributos que correspondentes às dimensões (Keim; Kriegel, 1996).

Existem conjuntos de técnicas amplamente conhecidas para visualização de bancos de dados multidimensionais, tais como (Keim; Kriegel, 1996) (Hoffman, 1999):

- Técnicas de projeção geométrica;
- Técnicas iconográficas;
- Técnicas orientadas a pixel;
- Técnicas hierárquicas;
- Técnicas baseada em grafos;
- Híbridas.

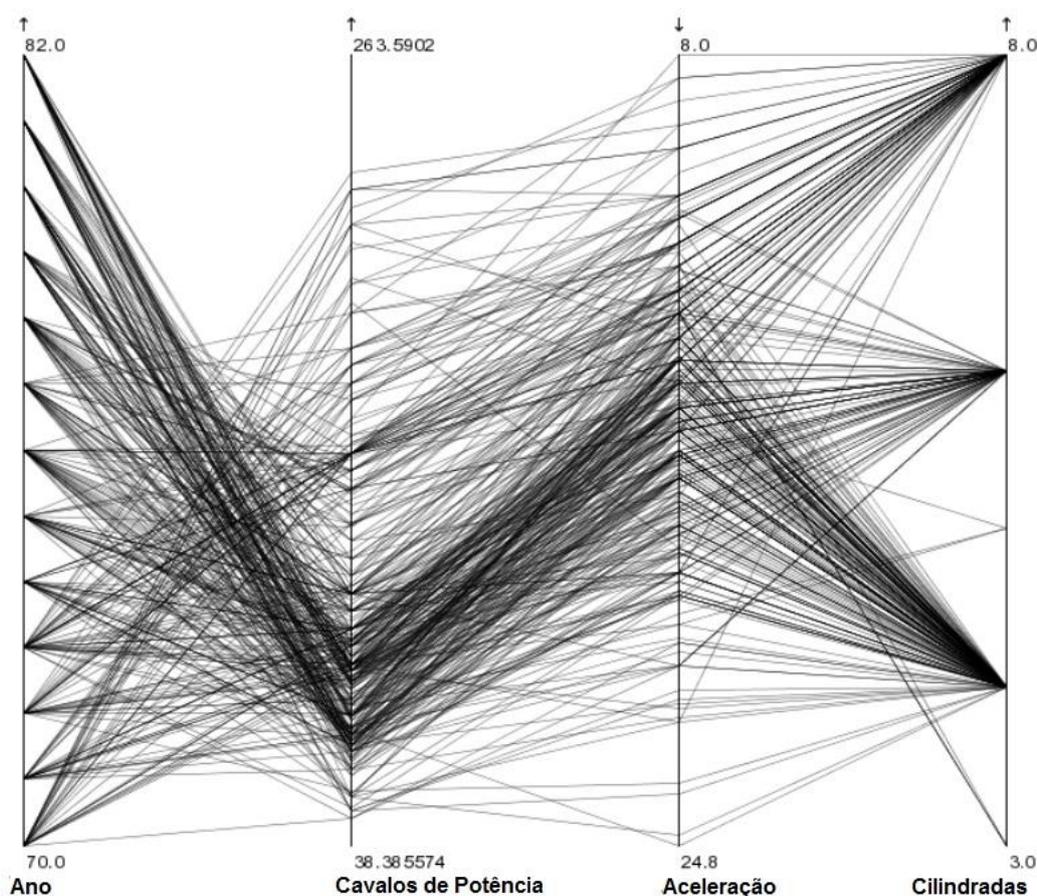
#### 3.1 TÉCNICAS DE PROJEÇÃO GEOMÉTRICA

As técnicas de projeção geométrica visam encontrar projeções “interessantes” em um conjunto de dados multidimensional. Em particular citam-se:

- Coordenadas Paralelas - *Parallel Coordinates* (Inselberg, 1985);
- Matriz de *Scatterplots* (Furnas, 1994);
- *RadViz* (Hoffman et al., 1997);
- Coordenadas Paralelas Circulares (Hoffman, 1999);
- *PolyViz* (Hoffman, 1999) (Grinstein et al., 2001);
- *Survey Plots* – (Lohninger, 1994) (Grinstein et al., 2001).

A técnica de Coordenadas Paralelas mapeia um espaço de  $k$ -dimensões para um dos eixos do *display*  $x$  ou  $y$ . Os eixos correspondem às dimensões, as quais estão linearmente dimensionadas a partir de um valor mínimo ao valor máximo da dimensão correspondente. Cada item é apresentado como uma linha poligonal, transpassando cada um dos eixos no ponto que corresponde ao valor da dimensão considerada (Figura 8).

Figura 8 - Coordenadas Paralelas exemplo com carros em quatro dimensões



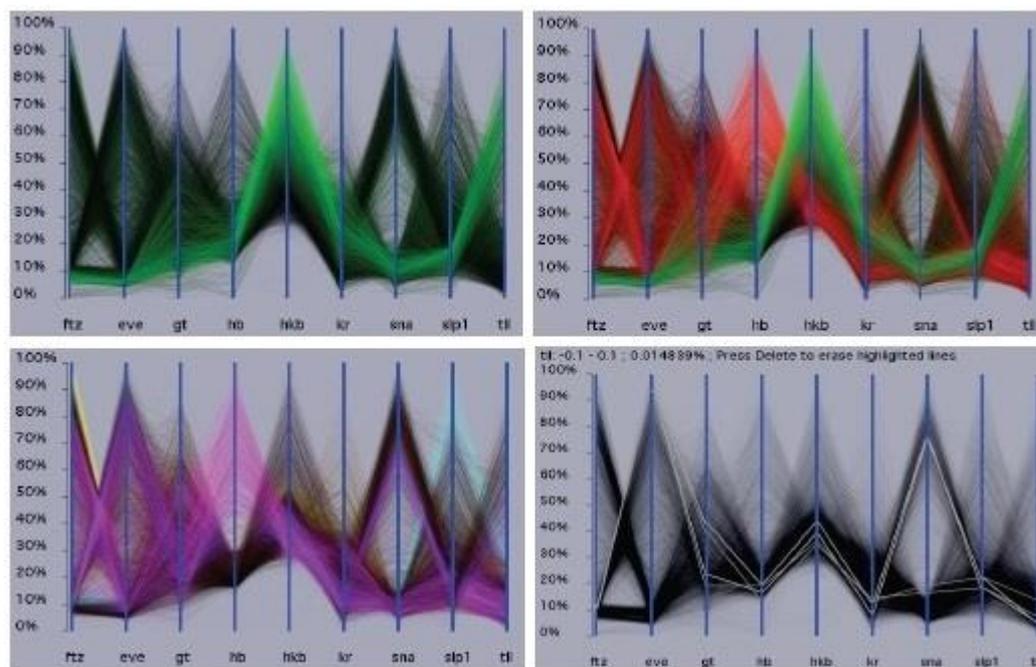
Fonte: Hauser, Ledermann e Doleisch (2002).

Embora o princípio da técnica de visualização de coordenadas paralela seja bastante simples, ela é muito eficaz na revelação de uma vasta gama de características dos dados, tais como suas distribuições e dependências funcionais.

No entanto, uma vez que as linhas poligonais podem sobrepor-se, o número de itens que podem ser visualizado no *display* ao mesmo tempo, é limitado a cerca de 1.000 itens (Keim; Kriegel, 1996). Para reduzir esse problema ou melhorar a

carga visual, algumas funcionalidades podem ser adicionadas a técnica. Um exemplo disto é a atribuição de cores para destacar determinados itens (Figura 9).

Figura 9 - Aplicação de cores em coordenadas paralelas



Fonte: Rübél (2006).

Outras abordagens incluindo recursos iterativos podem ser adequadas à técnica, dentre eles cita-se:

- Visibilidade: esconder itens ou dimensões existentes;
- Permutabilidade: alterar a ordem entre as dimensões;
- Escalabilidade: alterar valor mínimo e máximo de cada dimensão;
- *Zoom*: Melhorar visualização de uma faixa específica de dados;
- Informação sob demanda: adicionar ou remover faixas de dados;
- Múltiplas visualizações;
- Auxílio e adição de outras técnicas de visualização.

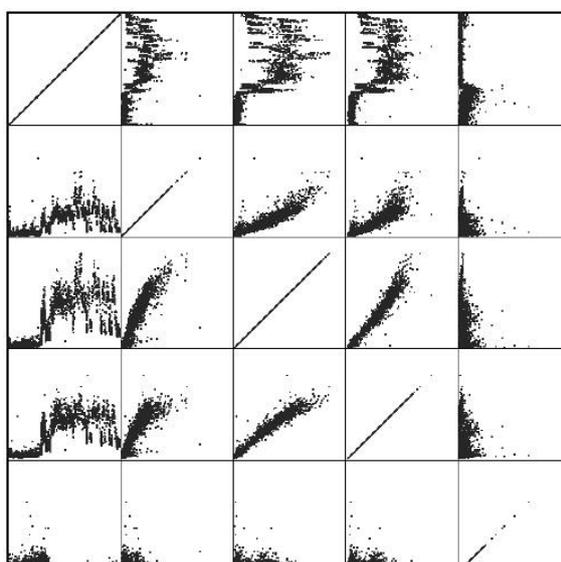
De maneira geral essa técnica permite transformar a busca por relações entre os atributos um processo de reconhecimento de padrões mais intuitivo.

A técnica de *Scatterplots* consiste no desenho de pontos em coordenadas x e y, se assemelhando muito a gráficos bidimensionais. Já a Matriz de *Scatterplots* é uma extensão dessa técnica, onde múltiplas projeções com diferentes pares de

dimensões são desenhadas, facilitando a interpretação das correlações entre esses pares (Figura 10).

Ao manter uma relação entre dimensões e projeções, um grande conjunto de dimensões irá resultar em uma projeção de tamanho proporcional, assim, sendo necessárias técnicas iterativas como o *zooming* – ampliação – para a análise dos resultados. Outras variações das Matrizes de *Scatterplots* são *hyperbox* (Alpern; Carter, 1991) e *hyperslice* (Wijk; Liere, 1993).

Figura 10 - Exemplo de Matriz de *Scatterplots*



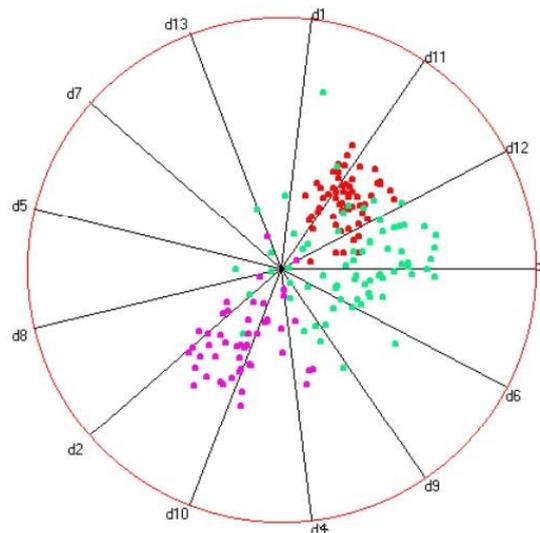
Fonte: Wong e Bergeron (1997).

*RadViz*, também chamada de coordenadas radiais tem um princípio semelhante ao das Coordenadas Paralelas. Organizada num formato circular, as linhas - *Dimensional Anchors* (DAs) - que formam os segmentos correspondem às dimensões, estas se originam no centro e vão até o perímetro com espaçamentos de mesmo tamanho (Hoffman et al., 1997).

Os dados são colocados entre as DAs de acordo com a “força” que possuem entre as dimensões até que a soma delas seja zero (Figura 11). Hoffman (1999) explica analogamente que, é como se os dados estivessem presos a molas imaginárias entre as DAs e a posição em que o dado deve ser desenhado corresponde ao equilíbrio da forças das molas associadas às dimensões. Outras características dessa técnica são (Hoffman, 1999) (Fayyad et al., 2002):

- Vários pontos são mapeados na mesma posição;
- Valores aproximados são posicionados mais ao centro do círculo;
- Valores aproximados que estão associados a dimensões que se encontram em eixos opostos também se posicionam mais ao centro do círculo;
- Valores maiores projetam o ponto desenhado para regiões mais próximas as dimensões (eixos) associadas.

Figura 11 - Exemplo da técnica *RadViz*



Fonte: Hoffman (1999).

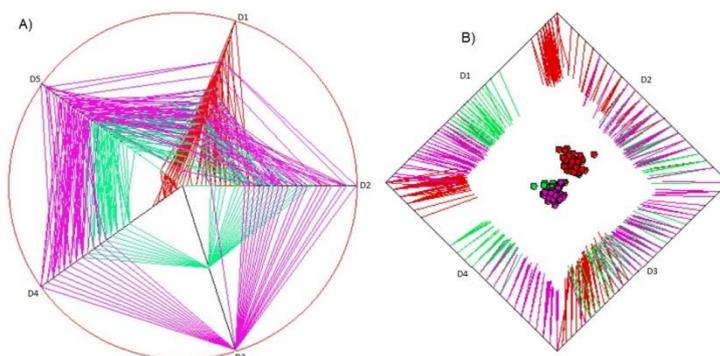
Hoffman (1999) cita outras duas técnicas, a primeira chamada de Coordenadas Paralelas Circulares – também conhecida como Gráficos de Estrela Sobrepostos – e a segunda denominada *PolyViz*.

A técnica de Coordenadas Paralelas Circulares é uma adaptação da já citada técnicas de Coordenadas Paralelas. Neste formato circular, os maiores valores são mapeados na parte mais externa do círculo, enquanto os menores são mapeados mais ao centro (Figura 12, a). Segundo Hoffman (1999) a assimetria resultante dessa técnica facilita a detecção de padrões.

A *PolyViz* estende o a técnica de *RadViz* utilizando elementos das Coordenadas Geométricas (Circulares), na *Polyviz* as dimensões são dispostas como uma linha e não apenas de um ponto. As mola “imaginárias” são utilizadas ao longo da DAs (linhas), que corresponde a todos os valores existentes na dimensão.

Cada atributo está posicionado como na *RadViz* (Grinstein et al, 2001). Assim pode-se aproveitar de melhor forma a capacidade de exibição de agrupamentos, contidas na *Radviz*, com a habilidade de visualizar a distribuição de dados em cada dimensão, das Coordenadas Paralelas (Figura 12, b).

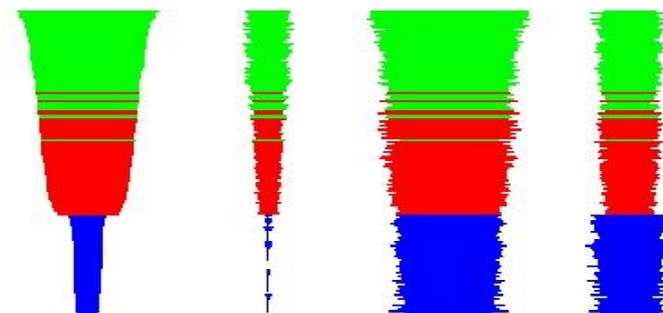
Figura 12 - Coordenadas Paralelas Circulares e *PolyViz*



Fonte: Hoffman (1999) e Grinstein et al. (2001).

*Survey Plot* consiste em  $n$  áreas retangulares, onde cada uma delas representa uma dimensão. Os dados são mapeados para linhas (ou retângulos dependendo do número de registros e tamanho da área de saída) estendidas de um ponto central, onde o comprimento de cada linha corresponde ao valor a ser representado (Grinstein et al., 2001) (Figura 13). O método permite visualizar a correlação entre quaisquer duas variáveis (especialmente quando os dados são classificados por uma dimensão) e pode encontrar regras exatas em um conjunto de dados de aprendizado de máquina.

Figura 13 - Exemplo de *Survey Plots* para quatro dimensões

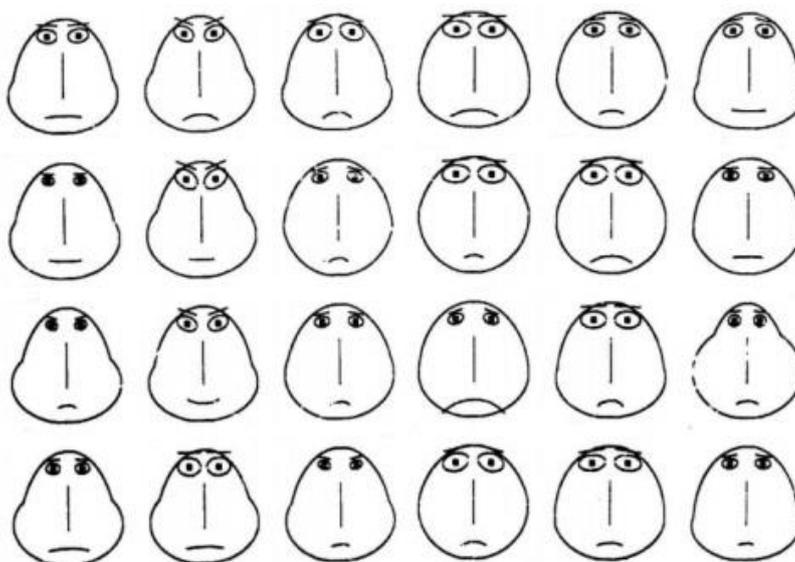


Fonte: Orange (2013).

### 3.2 TÉCNICAS BASEADAS EM ÍCONES

As técnicas baseadas em ícones consistem em mapear cada atributo multidimensional para um ícone. Keim e Kriegel (1996) citam como uma das primeiras abordagens a técnica das Faces de Chernoff, proposta por Chernoff em 1973. Essa técnica utiliza para as duas primeiras dimensões do atributo as duas dimensões do *display* (x e y), para as demais são utilizadas propriedades de um ícone que simboliza um rosto, como por exemplo: a forma do nariz, boca, olhos, e a forma do rosto em si (Figura 14).

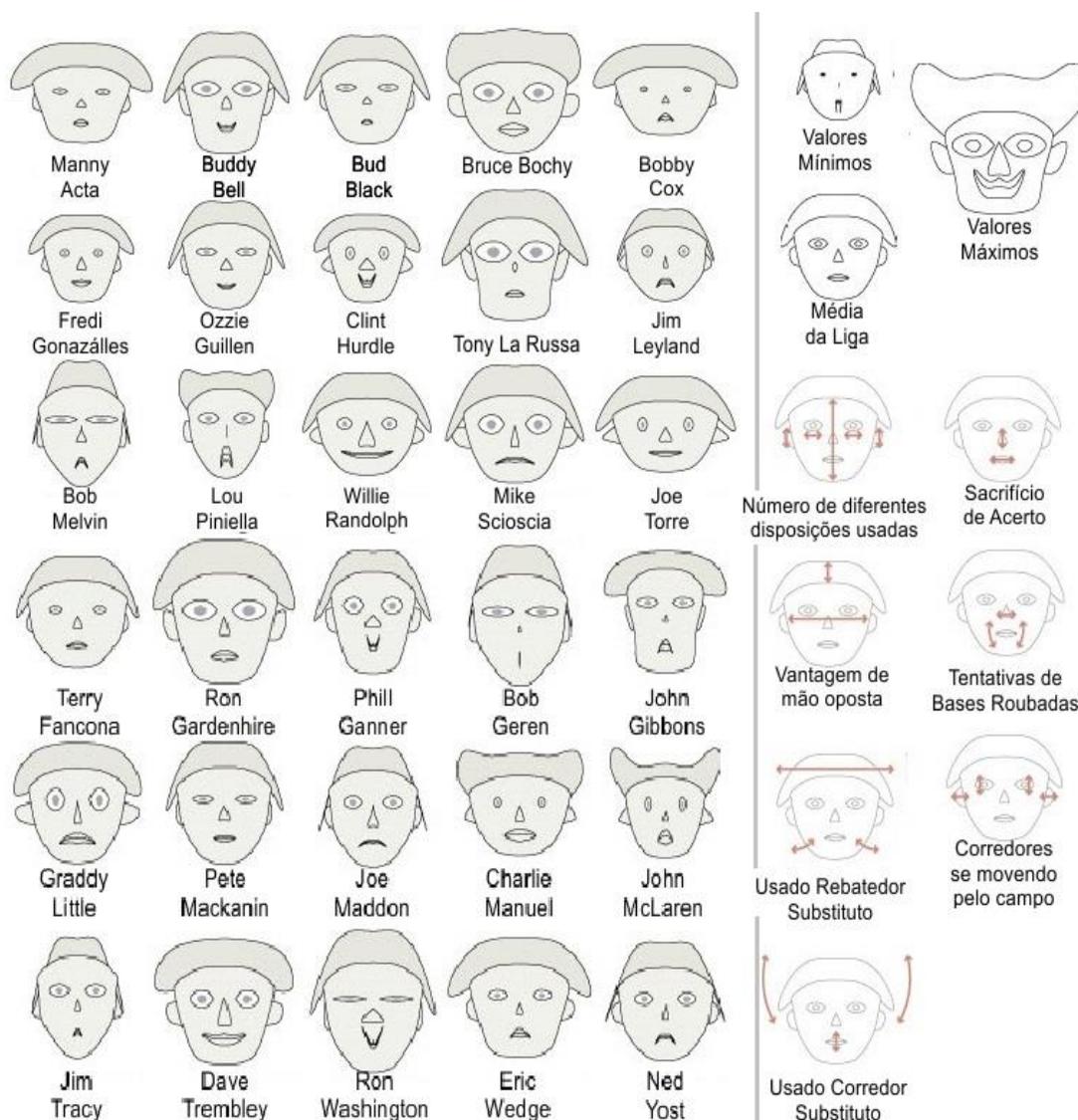
Figura 14 - Esboço da técnica Chernoff faces



Fonte: Chernoff (1973).

Outro exemplo da aplicação dessa técnica foi demonstrado por Wang C. Steve em 2008, onde suas faces foram temas de uma matéria do *The New York Times* escrita por Alan Schwarz. Steve mapeou o desempenho de alguns técnicos de baseball dos Estados Unidos conforme a Figura 15 onde à esquerda estão os técnicos e seus respectivos nomes e a direita a legenda das dimensões consideradas pelo cientista.

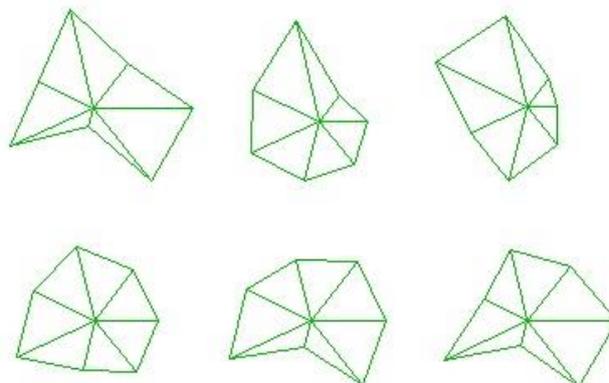
Figura 15 - Adaptação das Faces de Chernoff



Fonte: Schwarz (2013).

Semelhante a técnica das Faces de Chernoff, Ward (1994) utiliza outro ícone para fazer a representação das dimensões, baseado nas citações de Siegel et al., (1972), a técnica *Star Glyphs* mapeia os dados em formas de estrela, onde cada dimensão corresponde a uma de suas pontas e o tamanho é proporcional ao valor que o atributo representa.

Figura 16 - Adaptação para Star Glyphs com sete dimensões

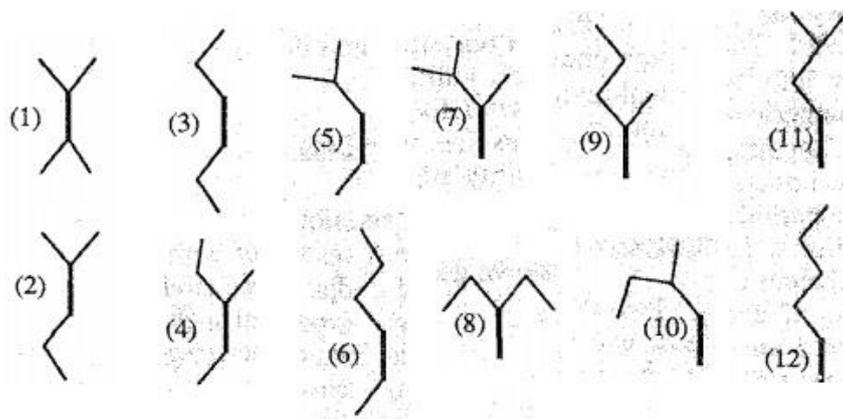


Fonte: Ward (1994).

As técnicas das Faces de Chernoff e a *Star Glyphs* focam na sensibilidade humana às faces (ou glifos) e suas características, porém o número de atributos que podem ser visualizados usando essa técnica, no entanto, é bastante limitado.

Uma técnica mais adequada para visualizar maiores quantidades de dados é a chamada “*stick figure*”, apresentada pelos autores Pickett e Grinstein em 1988. Ela consiste em ícones formados por pequenas linhas, onde, novamente as duas principais dimensões serão mapeadas para x e y, e adicionalmente as demais serão mapeadas para os ângulos e / ou comprimentos dos membros. Na Figura 17 pode ser visualizada uma família de *stick figures*, onde todos os ícones possuem quatro membros.

Figura 17 - Conjunto de ícones gerados pela técnica de *Stick Figure*

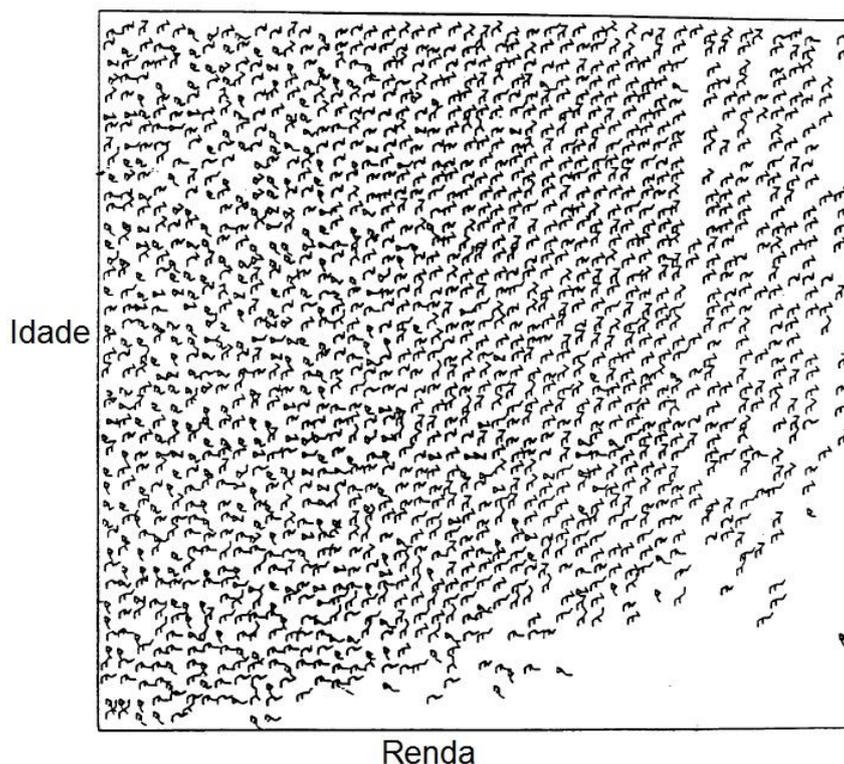


Fonte: Pickett e Grinstein (1988).

Caso o conjunto de dados seja relativamente denso em relação às dimensões que serão exibidas, a visualização resultante irá apresentar padrões de texturas, os quais variam de acordo com as características dos atributos. E, portanto, são detectáveis pela percepção *preattentive*. Em visualização, o termo *preattentive* é traduzido como a noção da velocidade e de facilidade com que certas propriedades são identificadas nas imagens visualizadas.

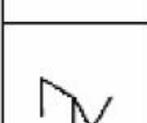
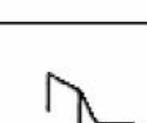
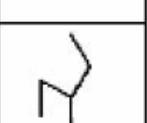
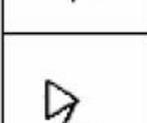
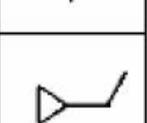
A Figura 18, por exemplo, utiliza uma amostra populacional para gerar a visualização utilizando a técnica de *stick figure*. Nela além das dimensões idade e renda, outras como ocupação, nível educacional, estado civil e sexo também foram mapeadas para propriedades das figuras (Figura 19). Os padrões na textura indicando a dependência funcional dos atributos de renda e idade.

Figura 18 - *Stick Figure* utilizando uma amostra populacional



Fonte: Pickett e Grinstein (1988).

Figura 19 – Dimensões adicionais adaptadas para a Figura 18

	trabalho inferior casado	trabalho superior casado	trabalho inferior solteiro	trabalho superior solteiro
homem baixa escolaridade				
homem alta escolaridade				
mulher baixa escolaridade				
mulher alta escolaridade				

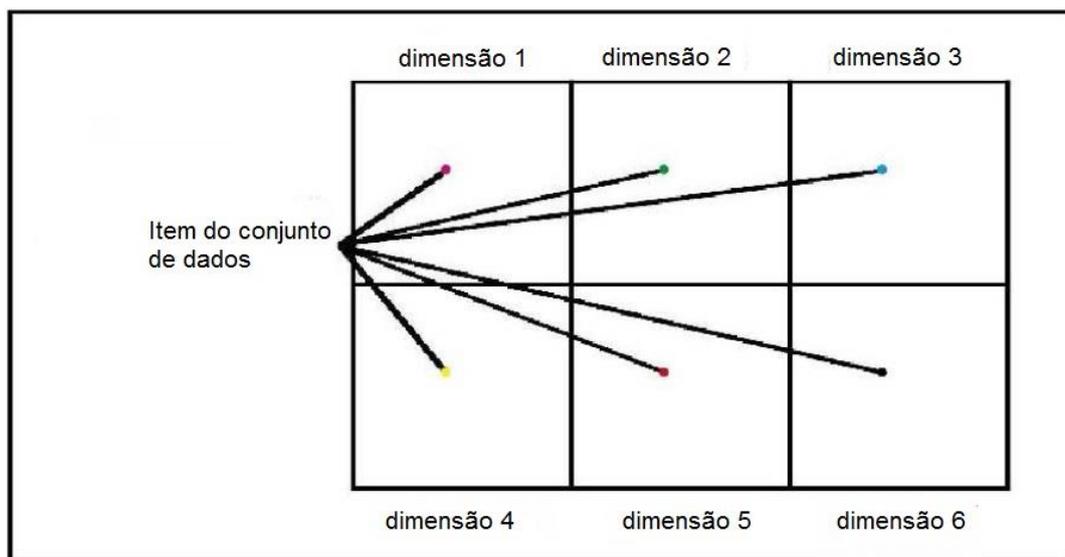
Fonte: Pickett e Grinstein (1988).

Nota-se que, em ambas as técnicas, *stick figure* e Faces de Chernoff, o número de dimensões que podem ser visualizadas é limitada.

### 3.3 TÉCNICAS ORIENTADAS A *PIXEL*

A ideia básica das técnicas orientadas a *pixels* é mapear cada atributo a um *pixel* colorido e apresentar suas dimensões em janelas separadas, ou seja, para cada conjunto que possuir  $n$  dimensões, a tela será dividida em  $n$  partes (Figura 20). Uma vez que, em geral, essa técnica usa apenas um *pixel* por atributo, ela nos permite visualizar uma maior quantidade de dados (Keim, 1996) (Keim, 2002).

Figura 20 - Construção de telas para seis dimensões



Fonte: Keim (1996).

Se cada item é representado por um *pixel*, a questão principal é como organizar os *pixels* na tela. Essas técnicas usam diferentes arranjos para diferentes fins. Ao organizar os *pixels* de forma adequada, a visualização resultante fornece informações detalhadas sobre as correlações locais, dependências e pontos chaves. Exemplos bem conhecidos são: a técnica de padrão recursivo e a técnica de segmentos de círculo.

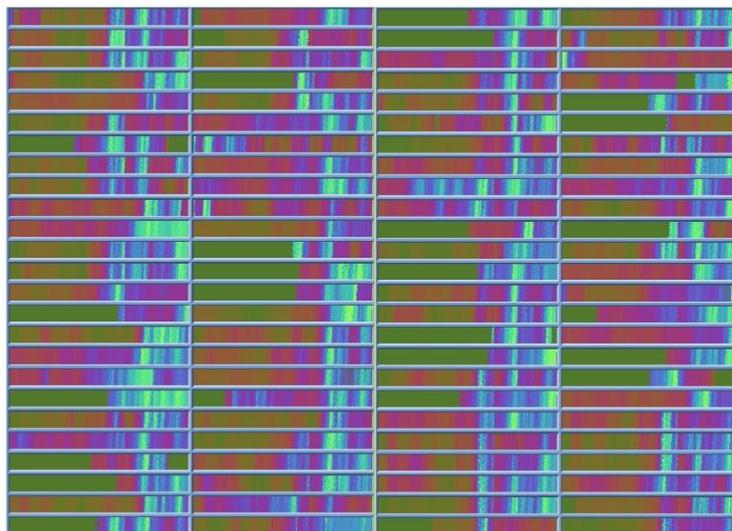
A técnica de padrão recursivo organiza os *pixels* num formato recursivo e genérico de “vai e vem”, sendo destinada em particular, para conjuntos de dados onde um atributo represente uma ordem natural (por exemplo, séries cronológicas).

O usuário pode especificar parâmetros para cada nível da recursão, e assim, controla a distribuição dos *pixels* para formar subestruturas semanticamente significativas. O elemento base em cada nível da recursão é um conjunto de altura  $h_i$  e largura  $w_i$  conforme especificado pelo usuário. Primeiramente, os elementos correspondem a *pixels* individuais que são distribuídos dentro de um retângulo de altura  $h_1$  e largura  $w_1$  da esquerda para a direita, em seguida, para baixo para trás da direita para a esquerda, após novamente para frente, da esquerda para a direita, e assim por diante.

A mesma distribuição básica é feita em todos os níveis de recursão, com a única diferença que os elementos básicos que estão dispostos no nível  $i$  são

padrões de distribuições resultantes do nível  $(i - 1)$ . Na Figura 21, por exemplo, a visualização por padrões recursivos de dados financeiros é mostrada. A visualização mostra 20 anos (janeiro de 1974 a abril de 1995) dos preços diários de 100 ações previstas Índice de Ações de Frankfurt.

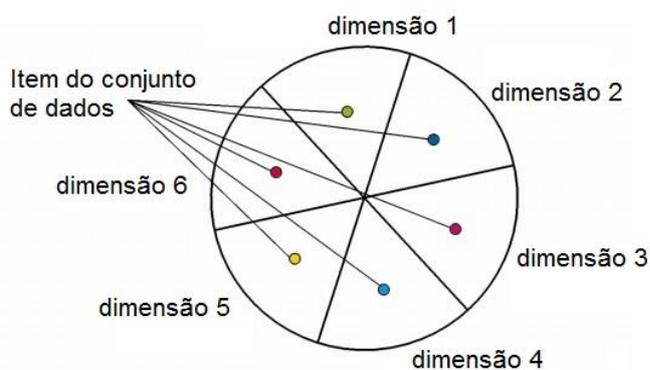
Figura 21 – Cem ações previstas no Índice de Ações de Frankfurt



Fonte: Keim (2002).

A ideia da técnica de segmentos de círculo é a de representar os dados em um círculo dividido em segmentos, um para cada dimensão. Dentro das dimensões, cada item é novamente representado por um único *pixel* colorido (Figura 22).

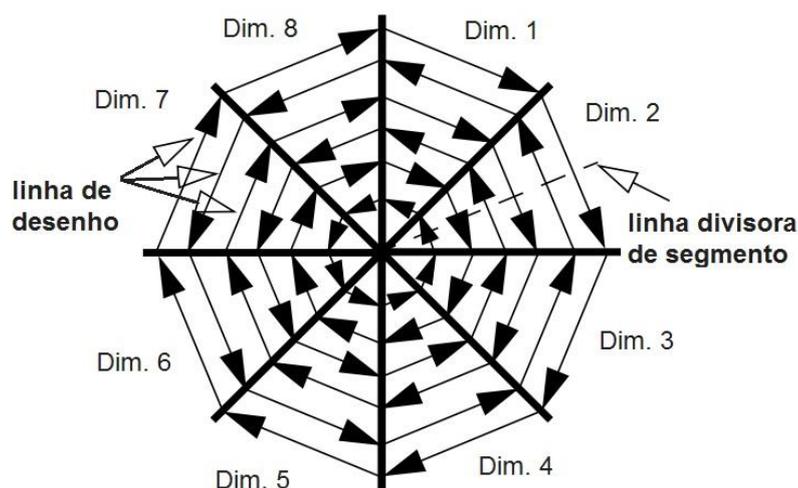
Figura 22 - Construção de segmentos para seis dimensões



Fonte: Keim (2000).

Assim se o conjunto de dados for formado por  $k$  dimensões, o círculo será dividido em  $k$  segmentos. Os elementos são dispostos dentro do segmento de um lado a outro, ao longo da chamada “linha de desenho” que é ortogonal às linhas que constituem as fronteiras do segmento (Figura 23).

Figura 23 - Segmentos de círculo para oito dimensões

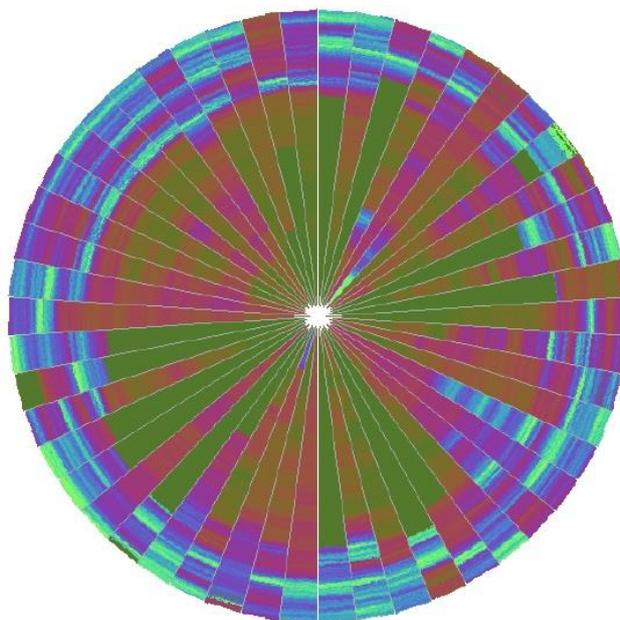


Fonte: Keim et al. (1996).

A “linha de desenho” começa no centro do círculo e desenha os *pixels* de uma linha da fronteira do segmento para outra. Quando a “linha de desenho” atinge uma dessas linhas, ela é movida paralelamente ao longo da “linha divisora de segmento” para o lado de fora do círculo e o sentido da “linha de desenho” muda. Este processo é repetido até que todos os elementos de uma dimensão sejam visualizados e, em seguida, todo o processo é reiniciado para as dimensões restantes.

Sendo assim, perto do centro, todos os atributos estão perto uns dos outros aumentando a comparação visual dos seus valores. A Figura 24 mostra um exemplo com os mesmo dados utilizados no exemplo anterior, considerando somente cinquenta ações e usando a técnica de segmentos de círculo.

Figura 24 - Cinquenta ações previstas no Índice de Ações de Frankfurt



Fonte: Keim (2002).

### 3.4 TÉCNICAS HIERÁRQUICAS

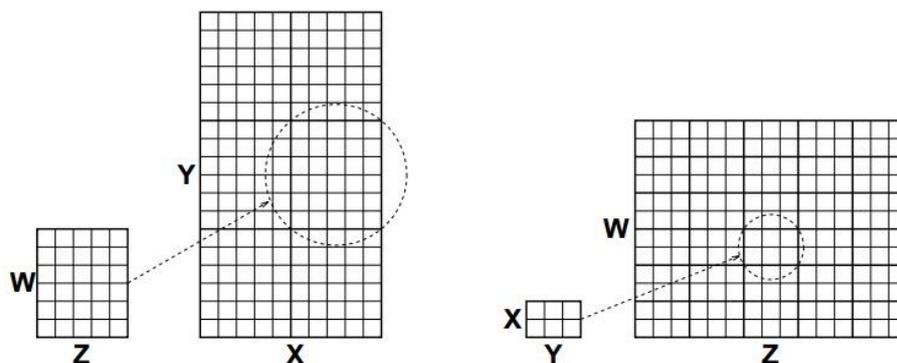
Keim e Kriegel (1996) dizem que as Técnicas Hierárquicas subdividem um espaço em  $k$ -dimensões e apresenta esses subespaços de forma hierárquica. Ele cita como mais expressivas as técnicas:

- *Dimensional Stacking* por Leblanc et al. (1990).
- *Treemaps* por Shneiderman (1999);
- *n-Vision*, também conhecida como “mundos dentro de mundos” por Feiner e Beshers (1990);

Wong e Bergeron (1997) citam a técnica de *Dimensional Stacking* como mais vantajosa dentre as outras por não precisar de funções ou regras adicionais para que os dados possam ser plotados na tela. Keim (2002) define a ideia básica dessa técnica como: aplicar um sistema de coordenadas dentro de um sistema de coordenadas. E exemplifica dizendo, que, duas dimensões formam os eixos externos, outras duas dimensões serão colocadas dentro dessas dimensões e assim por diante. Para melhorar a visualização, Keim (2002) diz que as dimensões de

maior importância devem ser escolhidas primeiro. A formulação dessa técnica está na Figura 25.

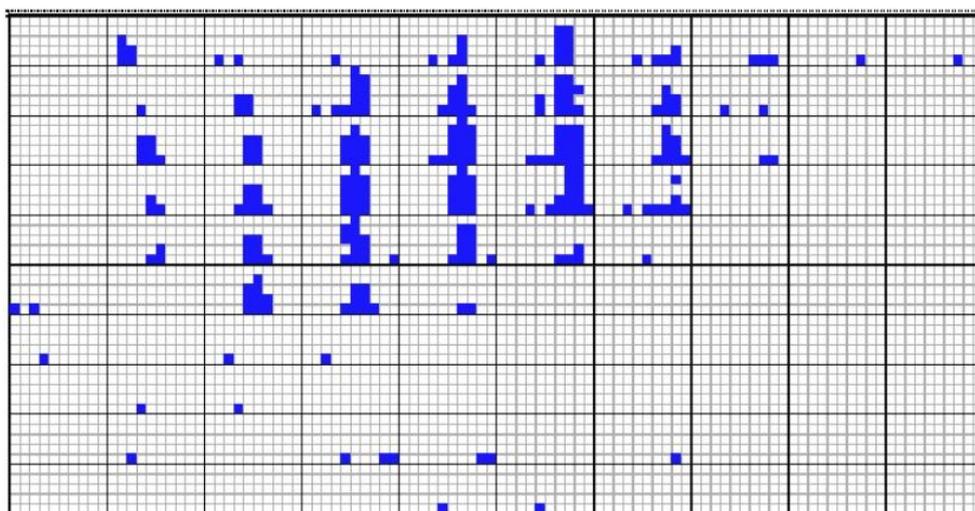
Figura 25 - Duas diferentes maneiras de visualizar os dados



Fonte: Wong e Bergeron (1997).

Seguindo a formulação proposta por Wong e Bergeron (1997), Keim (2002) exemplifica seu funcionamento na Figura 26, onde a visualização mostra quatro dimensões, as duas mais externas são a latitude e longitude de uma exploração de óleo qualquer, as duas internas são o grau de minério e a profundidade.

Figura 26 – Exemplo de *Dimensional Stacking*



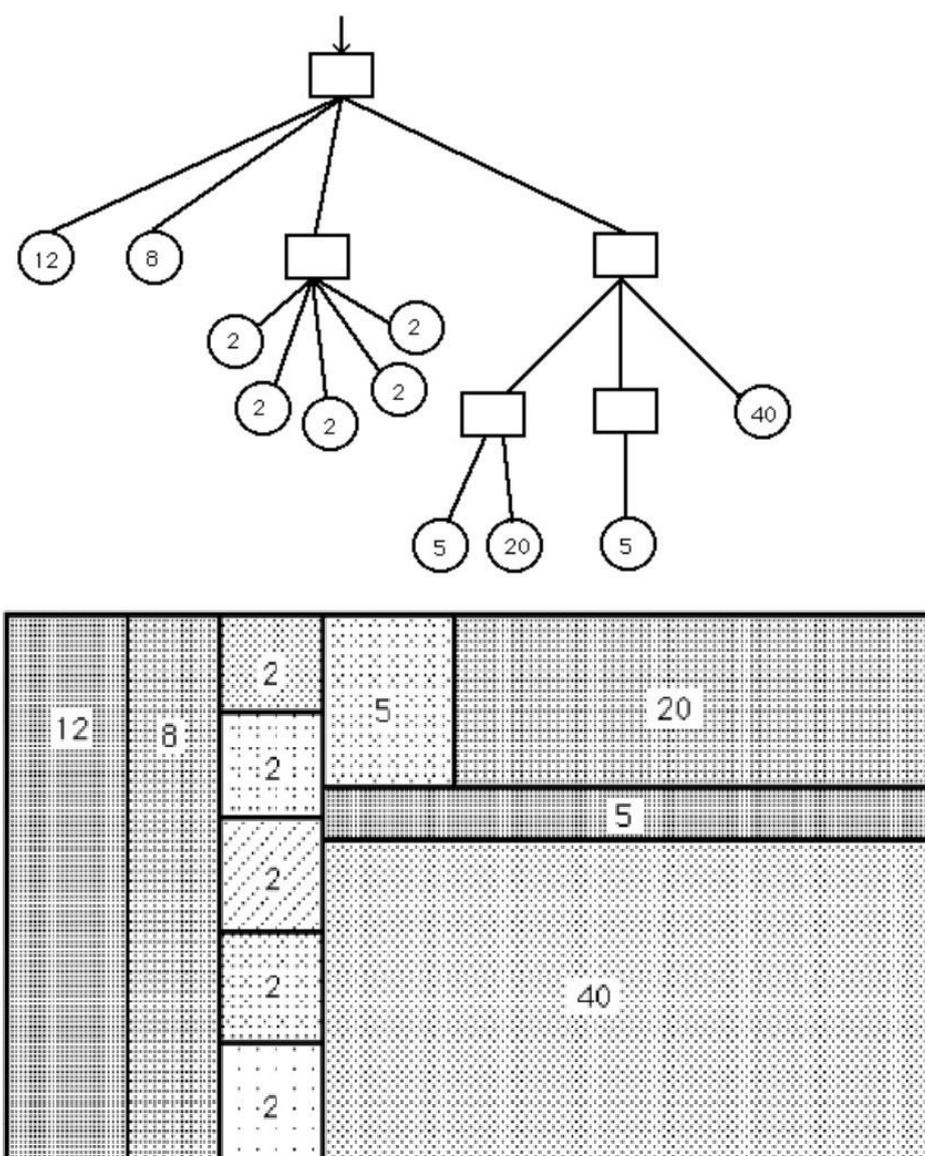
Fonte: Keim (2002).

Além da *Dimensional Stacking*, outra técnica chamada *Treemap* é usada para mostrar de forma hierárquica as informações.

A técnica de *Treemap* consiste em mapear as informações para espaços retangulares (2D) dentro de um *display*, sendo que este *display* terá 100% do seu espaço preenchido. Controles interativos permitem que os usuários alterem a maneira de como os blocos retangulares estão dispostos, bem como também alterar o seu conteúdo.

A técnica particiona o *display* de forma que uma coleção de blocos seja desenhada e represente a estrutura de uma árvore (Figura 27).

Figura 27 - Exemplo de árvore representada por uma *Treemap*

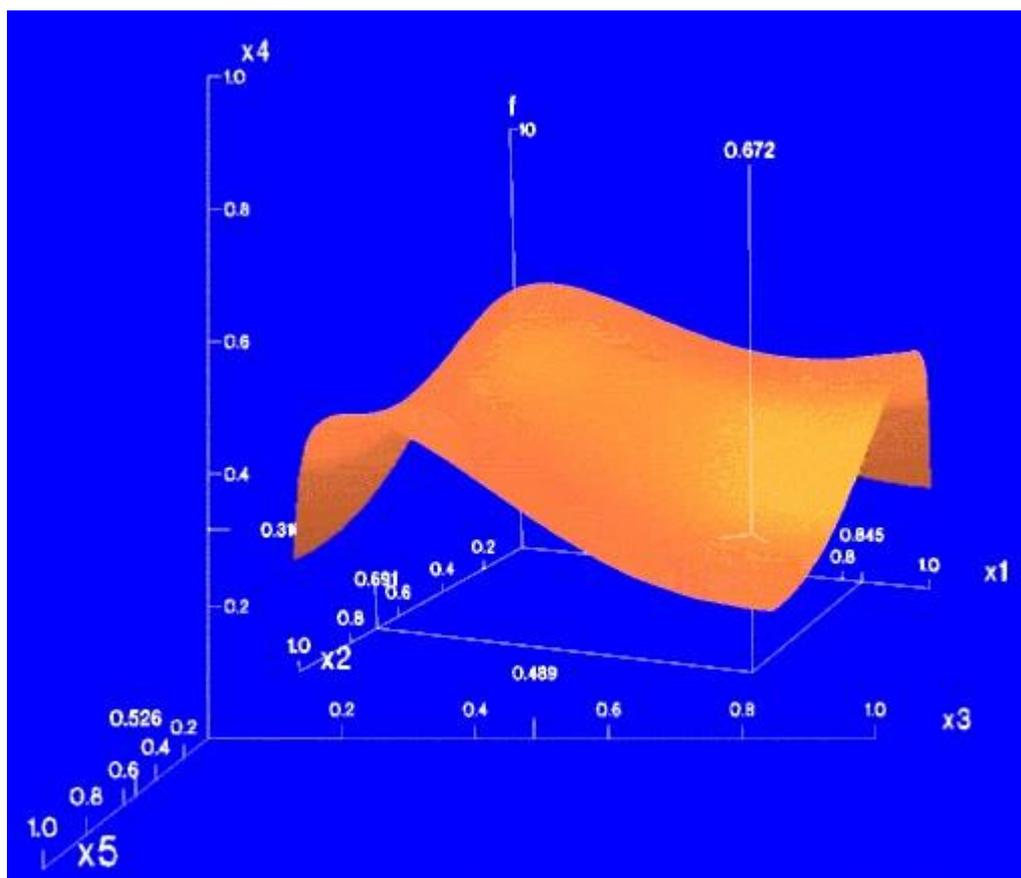


Fonte: Shneiderman (1999).

A terceira técnica citada por Keim e Kriegel (1996) é a chamada *n-Vision*, também conhecida como “mundos dentro de mundos”. A abordagem simples desta técnica consiste em separar uma ou mais dimensões, a fim de reduzir a dimensionalidade do conjunto de dados para até três dimensões, assim, a fatia resultante do conjunto pode ser visualizada e manipulada em qualquer equipamento gráfico convencional 3D.

As dimensões que não foram selecionadas podem ser adicionadas posteriormente, ou seja, uma ou mais novas dimensões podem ser escolhidas para formar um novo modelo (mundo) 3D que será incorporado ao primeiro modelo (Figura 28). Este processo pode então ser repetido recursivamente (Bescher; Feriner, 2013).

Figura 28 - *n-Vision*, exemplo para cinco dimensões



Fonte: Bescher e Feriner (2013).

### 3.5 TÉCNICAS BASEADAS EM GRAFOS

Sendo comumente anexadas as técnicas hierárquicas em diversas bibliografias, as técnicas baseadas em grafos tem sua abordagem separada apenas pela ênfase sobre informações hierárquicas complexas, altamente interconectadas e que possam conter propriedades inerentes à teoria dos grafos.

No caso de grafos 2D, as propriedades podem ser:

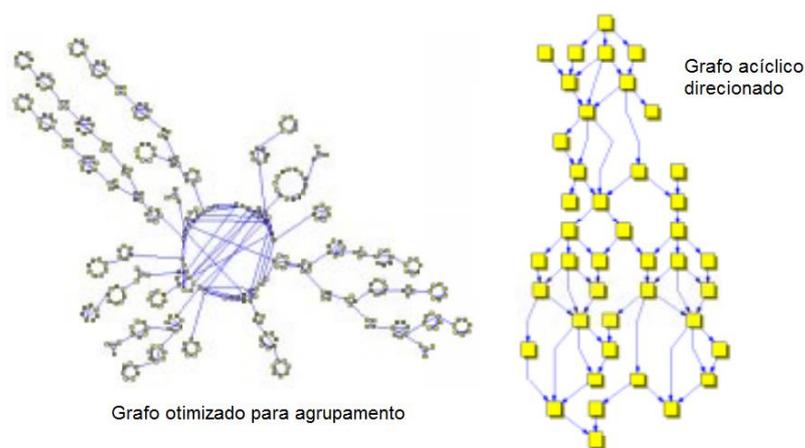
- Planaridade: não há arestas que se cruzem;
- Ortogonalidade: somente linhas ortogonais;
- Propriedade de grade: As coordenadas dos vértices são números inteiros.

As visualizações ainda podem apresentar propriedades estéticas, Keim (1997) cita essas propriedades como “Metas de Otimização”, citando as seguintes:

- Número mínimo de cruzamento entre as arestas;
- Exibição ótima das simetrias;
- Exibição ótima dos *clusters* – agrupamentos;
- Número mínimo de curvas em gráficos poligonais;
- Distribuição uniforme dos vértices;
- Comprimento uniforme das arestas.

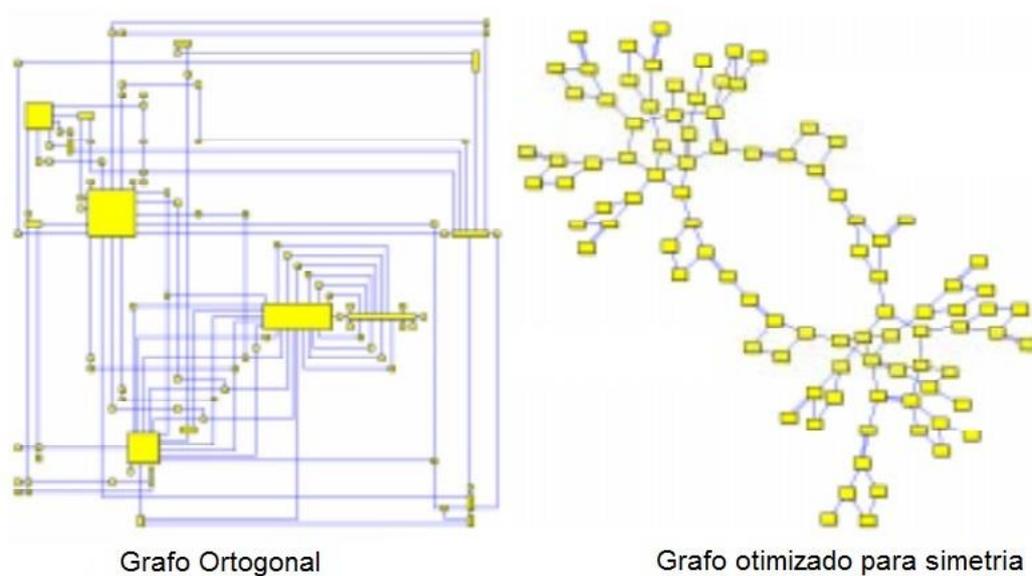
As Figura 29 e Figura 30 mostram exemplos de grafos em 2D. Já a Figura 31 mostra exemplos de grafos em 3D.

Figura 29 - Exemplo de grafo com distribuição otimizada e acíclico



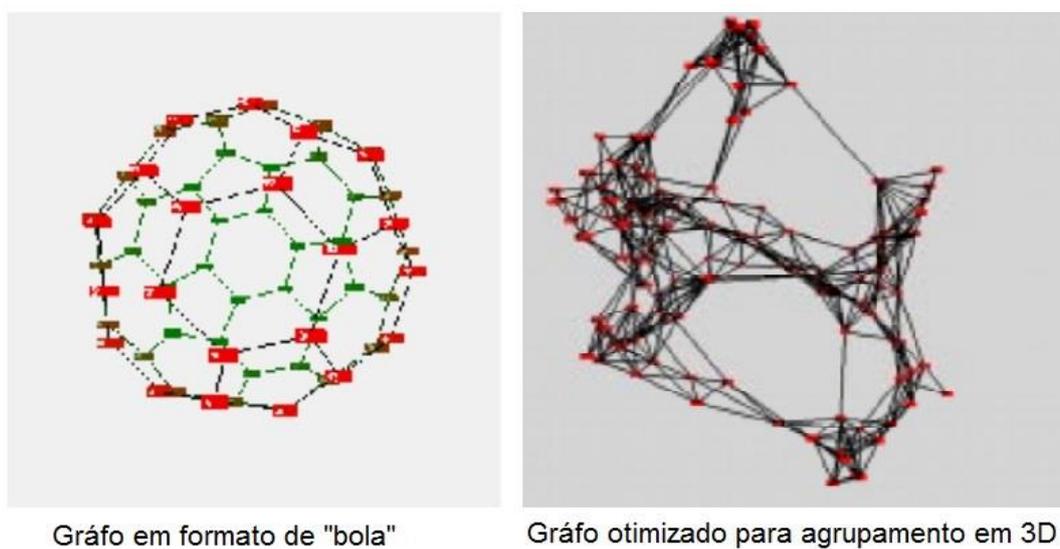
Fonte: Keim (1997).

Figura 30 - Exemplo de grafo ortogonal e otimizado para simetria



Fonte: Keim (1997).

Figura 31 - Exemplo de grafos em três dimensões



Fonte: Keim (1997).

### 3.6 CONSIDERAÇÕES FINAIS

Neste capítulo, procurou-se demonstrar as definições e as principais técnicas para visualização de informação, apresentando modelos de referência

comumente citados nas bibliografias. Apresentadas em grupos, as técnicas demonstram maneiras de organizar as dimensões de acordo com a informação representada e a quantidade de dados existentes em determinados conjuntos.

Atualmente novas propostas para a visualização de informação, não envolvem a descoberta de novas visões e sim a extensão e implementação de novos recursos nas visões já existentes, muitas vezes com a utilização de técnicas híbridas, ou seja, envolver mais de uma visualização simultaneamente.

Com o estudo teórico das técnicas contidas neste capítulo e com o estudo das estruturas dos dados propostos no Capítulo 4 será possível avaliar quais técnicas podem representar da melhor forma os cenários existentes.

## 4 CONJUNTOS DE DADOS DO EXPERIMENTO

Os conjuntos de dados que serão utilizados para com as técnicas de visualização de informação foram retirados da PSLC - *Pittsburgh Science of Learning Center - DataShop*, onde já existe uma grande incidência de utilização dos mesmo, garantindo assim, a validade e integridade dos dados contidos nesse repositório.

Dentre os conjuntos existentes, três foram selecionados, sendo eles:

- Conjunto de Dados Geometria;
- Conjunto de Dados da Língua Chinesa;
- Conjunto de Dados Álgebra.

Os conjuntos serão descritos mais amplamente nas próximas sessões.

### 4.1 CONJUNTO DE DADOS GEOMETRIA

O *Geometry*, como é chamado este conjunto, é composto por dados coletados no período entre os anos de 1996 e 1997 de um curso de Geometria, que teve como principal investigador Kenneth Koedinger. Esse conjunto foi formado a partir de dados que continham o passo a passo dos estudantes, sendo que os tempos apresentados foram simulados (Koedinger et al., 2010).

Para geração do arquivo que representa o conjunto de dados, foram escolhidos todos os intervalos disponíveis de datas e modelos de conhecimento. O arquivo resultante é formado por dados de 6.778 instâncias, dividido em 29 modelos de conhecimento, onde cada modelo possui as suas áreas de conhecimento subdivididas em níveis (Zat, 2012).

Com a escolha de algumas colunas aleatórias, uma amostra da organização estrutural dos dados e das informações contidas no *dataset*, pode ser conferida na Tabela 3.

Tabela 3 - Amostra de dados do conjunto *Geometry*

Linha	Id do Estudante	Problema	KC (Original)	Resultado
294	Stu_0a8e3638e3c0deb4e5e49c72286a5b83	CIRCLE_O	ALT:CIRCLE-RADIUS	CORRECT
295	Stu_0a8e3638e3c0deb4e5e49c72286a5b83	CIRCLE_O	ALT:CIRCLE-RADIUS	CORRECT
296	Stu_0a8e3638e3c0deb4e5e49c72286a5b83	CIRCLE_O	ALT:CIRCLE-DIAMETER	CORRECT
297	Stu_0a8e3638e3c0deb4e5e49c72286a5b83	COVERING_PO OL	ALT:CIRCLE-AREA	CORRECT
298	Stu_0a8e3638e3c0deb4e5e49c72286a5b83	COVERING_PO OL	ALT:CIRCLE-RADIUS	INCORRECT
301	Stu_0c90d4341f3dcf90745bf28fac694592	TRAPEZOID_A BCD	ALT:TRAPEZOID-AREA	INCORRECT
305	Stu_0c90d4341f3dcf90745bf28fac694592	TRAPEZOID_A BCD	ALT:TRAPEZOID-AREA	CORRECT
314	Stu_16faef9fdc3e509fc03b6aab061f3eec	BUILDING_A_SI DEWALK	ALT:COMPOSE-BY-ADDITION	CORRECT
315	Stu_16faef9fdc3e509fc03b6aab061f3eec	BUILDING_A_SI DEWALK	ALT:PARALLELOGRAM-AREA	CORRECT
316	Stu_16faef9fdc3e509fc03b6aab061f3eec	BUILDING_A_SI DEWALK	ALT:PARALLELOGRAM-AREA	CORRECT
538	Stu_1a265adcd4233ea0bfd9b009d4db0aca	PAINTING_THE _WALL	ALT:PARALLELOGRAM-SIDE	CORRECT
539	Stu_1a265adcd4233ea0bfd9b009d4db0aca	PAINTING_THE _WALL	ALT:PARALLELOGRAM-AREA	CORRECT
540	Stu_1a265adcd4233ea0bfd9b009d4db0aca	PAINTING_THE _WALL	ALT:PARALLELOGRAM-AREA	INCORRECT

Fonte: <<http://pslccdatashop.org>> (Koedinger et al., 2010).

Neste conjunto, um total de quinze habilidades – que se encontram na coluna *KC (Original)* – foram avaliadas, denominadas como regra de produção elas são descritas na relação a seguir (Cen et al., 2006) (Zat, 2012):

- *Circle-area*: A partir do raio deve-se encontrar a área do círculo;
- *Circle-circumference*: A partir do diâmetro deve-se encontrar a circunferência de um círculo;
- *Circle-diameter*: A partir do raio ou circunferência deve-se encontrar o diâmetro de um círculo;
- *Circle-radius*: Encontre o raio através da área, circunferência ou diâmetro;

- *Compose-by-addition*: Em  $a+b=c$ , a partir de dois valores  $a$ ,  $b$  ou  $c$ , encontrar o terceiro;
- *Compose-by-multiplication*: Em  $a*b=c$ , a partir de dois valores  $a$ ,  $b$  ou  $c$ , encontrar o terceiro;
- *Parallelogram-area*: A partir da base e altura deve-se encontrar a área do paralelogramo;
- *Parallelogram-side*: A partir da área e altura (ou base) deve-se encontrar a base (ou altura);
- *Pentagon-area*: A partir de um lado e do apótema deve-se encontrar a área de um pentágono;
- *Pentagon-side*: A partir da área e do apótema deve-se encontrar o lado (ou o apótema);
- *Trapezoid-area*: A partir da altura e duas bases deve-se encontrar a área de um trapézio;
- *Trapezoid-base*: A partir da área e altura deve-se encontrar a base de um trapézio;
- *Trapezoid-height*: A partir da área e da base deve-se encontrar a altura de um trapézio;
- *Triangle-area*: A partir da base e da altura deve-se encontrar a área de um triângulo;
- *Triangle-side*: A partir da base e do lado deve-se encontrar a altura de um triângulo.

Após a definição das regras de produção foram definidos quatro fatores de dificuldade (Tabela 4) para avaliar os problemas (Cen et al., 2006) (Zat, 2012):

- *Embed*: Indica se uma forma é incorporada em outro formato, isto é, dois problemas exigem a mesma regra de produção *Circle-area* em algum dos seus passos. Em um dos problemas o círculo é incorporado a um quadrado, já no outro problema o círculo é apresentado isoladamente. Isso significa que os estudantes podem encontrar maior dificuldade ao calcular a área do círculo quando ela está inserida em outra figura;
- *Backward*: Refere-se à apresentação da fórmula, isto é, a ordem em que ela aparece. A forma no sentido para frente do *Compose-by-addition* é  $S = S_1 + S_2$ , já a sua forma no sentido inverso é  $S_1 = S - S_2$ ;

- *Repeat*: Indica se a regra de produção foi utilizada mais de uma vez na solução do mesmo problema;
- *FigurePart*: Indica qual parte da figura geométrica será calculada.

Tabela 4 - Fatores de Dificuldade

Nome do Fator	Valor do Fator
<i>Embed</i>	<i>alone, embed</i>
<i>Backward</i>	<i>forward, backward</i>
<i>Repeat</i>	<i>initial, repeat</i>
<i>FigurePart</i>	<i>area, area-difference, area-combination, diameter, circumference, radius, side, segment, base, height, apothem</i>

Fonte: Cen et al (2006).

As regras de produção e os fatores de dificuldade podem ser considerados os dados com maior relevância no conjunto *Geometry* (Zat, 2012), sendo assim é possível que esses dois conjuntos de informações tornem-se duas dimensões de grande importância na análise por Visualização de Informação.

## 4.2 CONJUNTO DE DADOS DA LÍNGUA CHINESA

O *Chinese Tone Study*, como é chamado o projeto detentor deste conjunto, possui informações sobre estudantes que estão aprendendo a língua Chinesa. Investigado por Ying Liu, esses dados foram coletados em um período entre os anos de 2005 e 2006 na escola de nome CMU.

Segundo Koedinger et al. (2010), o aprendizado dessa língua trabalha com o recurso de tons, que representa um dos maiores desafios para quem deseja aprendê-la.

Sobre um pressuposto que diz que, caso as características críticas do contorno tonal forem atendidas o aprendizado torna-se mais fácil. Assim, para que hipóteses pudessem ser testadas um experimento foi projetado e dividido em três condições, nelas os estudantes poderiam perceber a diferença entre os tons. A cada

condição os estudantes receberam um *feedback* referente ao seu desempenho (Koedinger et al., 2010).

Para geração do arquivo que representa o conjunto de dados, foram escolhidos todos os intervalos de datas e modelos de conhecimento disponíveis. O arquivo resultante é formado por dados de 48.443 instâncias, dividido em três modelos de conhecimento, onde cada modelo possui as suas áreas de conhecimento subdivididas em níveis (Zat, 2012). Os dados foram retirados no dia 08 de junho de 2012 e o projeto ainda estava em andamento, sendo assim modificações nas informações podem ocorrer nesse conjunto de dados, porém não constarão neste trabalho (Koedinger et al., 2010).

Na tabela apresentada (Tabela 5) onde há uma amostra dos dados, as colunas “Probl.” e “P” referem-se ao problema resolvido pelo aluno e os passos utilizados pelo mesmo até a solução, respectivamente.

Tabela 5 - Amostra de dados da Língua Chinesa

Linha	Código do Estudante	Probl.	P	Resultado	Feedback
5949	Stu_3a1e655328599b6 a1208115f2271c942	I01w01	1	INCORRECT	No, this is not correct. Please click on "hint" in the upper right corner for help.
5950	Stu_3a1e655328599b6 a1208115f2271c942	I01w01	2	CORRECT	Correct! Now click on Done to go to the next one.
5951	Stu_3a1e655328599b6 a1208115f2271c942	I01w01	1	HINT	Focus on the pitch change and listen to the sound again.
5952	Stu_3a1e655328599b6 a1208115f2271c942	I01w01	2	CORRECT	Correct! Now click on Done to go to the next one.
5953	Stu_3a1e655328599b6 a1208115f2271c942	I01w01	1	CORRECT	No Hint or Other Message Text
5954	Stu_3a1e655328599b6 a1208115f2271c942	I01w02	1	INCORRECT	No, this is not correct. Please click on "hint" in the top right corner for help.
5955	Stu_3a1e655328599b6 a1208115f2271c942	I01w02	2	INCORRECT	No, this is not correct. Please click on "hint" in the top right corner for help.
5956	Stu_3a1e655328599b6 a1208115f2271c942	I01w02	3	CORRECT	CORRECT Correct! Now click on Done to go to the next one.
5957	Stu_3a1e655328599b6 a1208115f2271c942	I01w02	1	CORRECT	CORRECT No Hint or Other Message Text

Fonte: <<http://pslcsdatashop.org>> (Koedinger et al., 2010).

Essa tabela também busca exemplificar uma situação, tendo como base a linha de número 5954, pode-se observar o seguinte: o estudante respondeu o problema 101w02 e foi comunicado que a resposta estava incorreta e poderia solicitar ajuda. Porém, mesmo com a opção de ajuda, ele preferiu assinalar novamente uma resposta sendo comunicado que a sua escolha estava incorreta. O estudante acertou a questão na terceira tentativa, sem recorrer à opção de ajuda disponível (Zat, 2012).

#### 4.3 CONJUNTO DE DADOS ÁLGEBRA

O Álgebra I 2006, como é chamado este conjunto, é composto por dados coletados no período entre os meses de outubro e dezembro de 2006 aplicados à *Central Westmoreland Carreira e Centro de Tecnologia (CWCTC)*, sendo investigado principalmente por Lisa Anthony. Esse conjunto é em sua maioria composta por números e equações matemáticas, o que aumenta a complexidade dos dados.

O *A Multimodal Interface for Solving Equations*, como é chamado o projeto detentor deste conjunto, compara padrões cognitivos de um tutor de Álgebra que trabalha com exemplos, tendo como ideia principal incentivar o estudante a praticar um tipo de problema levando-o a observar o seu aprendizado e desempenho com o passar do tempo.

Para geração do arquivo que representa o conjunto de dados, foram escolhidos todos os intervalos de datas e modelos de conhecimento disponíveis. O arquivo resultante é formado por dados de 12.568 instâncias, dividido em quatro modelos de conhecimento, onde aqui, somente dois modelos possuem dados com subdivisões (Zat, 2012).

Por ser um conjunto de dados consistido basicamente por números e expressões, como pode ser notado na amostra contida na Tabela 6, as dimensões e como elas podem ser estruturadas serão eleitas mediante a sua possibilidade de criação.

Tabela 6 - Amostra de dados agrupados por um problema

L.	Cod. Est.	Probl.	Nome Etapa	P	R.	Seleção	Ação	Ent.
15	Student 09tSMme Session 1	Eg40 -6x+8=8	-6x+8 = 8	1	OK	-6x+8=8 strategic	subtract	8
16	Student 09tSMme Session 1	Eg40 -6x+8=8	-6x+8-8=8-8	1	OK	-6x+8-8=8- 8 strategic	Clit	-6x+8=8
17	Student 09tSMme Session 1	Eg40 -6x+8=8	-6x = 8-8	1	OK	-6x=8-8 strategic	Clit	8-8
18	Student 09tSMme Session 1	Eg40 -6x+8=8	-6x = 0	1	OK	-6x=0 strategic	divide	-6
19	Student 09tSMme Session 1	Eg40 -6x+8=8	-6x/-6 = 0/-6	1	OK	-6x/-6 = 0/- 6 strategic	Rf	-6x/-6
20	Student 09tSMme Session 1	Eg40 -6x+8=8	x = 0/-6	1	OK	x = 0/-6 strategic	Rf	0/-6

Fonte: <<http://pslclatashop.org>> (Koedinger et al., 2010).

#### 4.4 CONSIDERAÇÕES FINAIS

Por se tratarem de dados inerentes ao ensino, a criação das dimensões pode se tornar um dos principais desafios, uma vez que os dados foram medidos em níveis, passos ou métodos, sem a definição de intervalos numéricos, o que leva-se a avaliar quais informações devem sofrer uma análise qualitativa e quantitativa para que um intervalo significativo seja definido.

Diante de uma análise prévia dos três conjuntos de dados, pode-se constatar que, o conjunto *Geometry* apresenta uma grande quantidade de informações coletadas, sendo dividido em mais de setenta variáveis (atributos), porém dentre estas muitas se eliminam durante um processo de ETL por não apresentarem uma informação relevante ou por se tratar de um dado oblíquo. Contudo o conjunto ainda se mostrou ainda bastante multivariado após os procedimentos, onde se acredita

que as técnicas geométricas com suporte a construções circulares das dimensões, possam ser a melhor escolha para este conjunto.

Já os conjuntos da Língua Chinesa e Álgebra I, apresentam um número bem inferior de variáveis comparados ao *Geometry*, não ultrapassando de vinte, onde novamente as técnicas geométricas – circulares ou não – ou a técnica de *dimensional stacking* – do conjunto de técnicas hierárquicas – podem apresentar uma melhor avaliação.

As técnicas baseadas em ícones apresentam uma construção mais ilustrativa, onde as dimensões talvez dependam de processos manuais de desenho para serem geradas, sendo assim, mais difícil de serem utilizadas.

Já as técnicas orientadas a grafos se destinam a informações com extremas relações entre si, o que inicialmente não é o caso dos *datasets* escolhidos.

Ao final pode-se classificar todos os *dataset* junto à Tabela 1 da seguinte forma: A classe de informação é composta por características e informações escalares. Os valores são alfanuméricos (exemplo ids dos alunos) e numéricos (como a duração). A natureza do domínio é Contínua-Discretizada por apresentar uma ordem em passos. E a dimensão do domínio como visto, é  $nD$ , ou seja, multidimensionado.

## 5 FERRAMENTAS

Para atender o conjunto de técnicas citadas por este trabalho e demais citadas por Keim e Kriegel (1996), uma série de ferramentas pode ser utilizada para a geração de visualizações.

### 5.1 CRITÉRIOS

Para a escolha das ferramentas foi utilizado como critério a possibilidade de geração das visões citadas neste trabalho. Tratando-se de uma disciplina em expansão, as aplicações existentes se restringem a trabalhos de graduação ou ferramentas de laboratório aplicadas a uma atividade específica, não havendo “produtos prontos” para que uma avaliação fosse feita entre os escolhidos.

Algumas ferramentas apresentam as mesmas técnicas de visualização, pois em algumas situações como será mencionado nas seções a seguir, apresentaram fechamentos inesperados, logo mais de uma foi escolhida para garantir a geração das visões.

Na Tabela 7 uma relação das ferramentas, visualizações suportadas, formatos de entrada e saída e atributos técnicos, pode ser conferida.

Tabela 7 – Relação de ferramentas para Visualização de Informação

(continua)

Ferramenta	Tipos de Visualização	Requerimentos de Sistema	Versão	Referências	Entrada de Dados	Formatos de Saída	Licença
Parvis	Coordenadas Paralelas	Sistema compatível com o Java 6.0 ou superior	0.3.1b	<a href="http://www.ediavirus.org/parvis/">http://www.ediavirus.org/parvis/</a>	STF	Somente Interface	GPL

Tabela 7 – Relação de ferramentas para Visualização de Informação

(conclusão)

Ferramenta	Tipos de Visualização	Requerimentos de Sistema	Versão	Referências	Entrada de Dados	Formatos de Saída	Licença
XmdvTool	Coordenadas Paralelas; Matriz de <i>Scatterplots</i> ; <i>Star Glyphs</i> ; <i>Display</i> orientado a <i>pixel</i> ; <i>Dimensional Stacking</i>	Windows ou Linux	8.0	<a href="http://davis.wpi.edu/xmdv/">http://davis.wpi.edu/xmdv/</a>	OKC; CSV	BMP; JPG; PNG	Gratuita
InfoVis Toolkit (IVTK)	Coordenadas Paralelas; <i>Treemap</i>	Toolkit	0.9beta 2	<a href="http://ivtk.sourceforge.net/">http://ivtk.sourceforge.net/</a>	--	--	ISC; MIT
Aplicações de Hoffman	<i>RadViz</i> ; <i>Survey Plots</i> ; Coordenadas Paralelas; Coordenadas Paralelas Circulares; Matriz de <i>Scatterplots</i> ; Segmentos de Círculo; <i>PolyViz</i> ; <i>RadViz 3D</i>	Sistema compatível com o Java 6.0 ou superior	--	<a href="http://www.cs.uml.edu/~phoffman/">http://www.cs.uml.edu/~phoffman/</a>	CSV.	Somente Interface	Gratuita
Ggobi	Matriz de <i>Scatterplots</i> ; Coordenadas Paralelas	Windows, Linux ou Mac OS X 10.5	2.1.10a	<a href="http://www.ggobi.org/">http://www.ggobi.org/</a>	XML; CSV	Somente Interface	CPL; GPL; BSD

## 5.2 PARVIS

Parvis (Ledermann, 2013) é uma ferramenta destinada à geração de visões utilizando a técnica geométrica de coordenadas paralelas. Tem como objetivo principal proporcionar um sistema flexível e uma interface reutilizável através de componentes e normas compatíveis com o Java Swing (Eckstein et al., 1998), (Fowler, 1998) e Java Beans (Oracle, 2013), com o intuito de apresentar visualizações sempre no estado da arte e prover uma interação visual com o conjunto de dados.

Essa ferramenta possui sua última versão com data de 19/03/2003 e apresenta alguns botões com certos problemas de funcionamento e o menu de ajuda parece ainda não ter sido criado, porém, ainda desempenha o seu papel de gerador de visões de coordenadas paralelas muito bem. Além de apresentar quatro tipos de interações: ordenação das dimensões, *zoom* – possibilidade de expandir os limites de uma dimensão -, translação e *brush* – destacar determinadas informações.

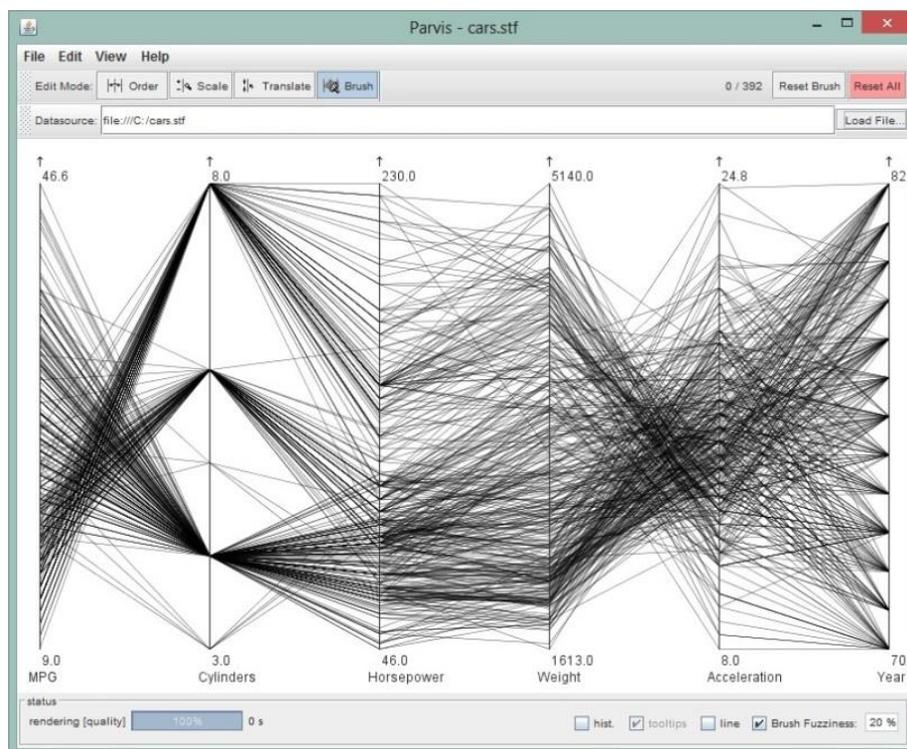
Por se tratar de uma aplicação *opensource* o Parvis foi utilizado pela *The Pennsylvania State University*, na criação de uma ferramenta chamada GeoViz Toolkit (GeoVISTA, 2013), que incluiu as coordenadas paralelas como parte de seu programa, fazendo pequenas correções e adaptações sobre a aplicação original.

Sua entrada de dados acontece por um arquivo de texto puro com extensão STF, sendo dividido da seguinte forma: a primeira linha deve apenas conter o número de dimensões que o conjunto contém, as linhas seguintes devem conter somente o nome da dimensão e o tipo de dado que ela representa [inteiro, real ou *string*], e por último as linhas que representem os valores das dimensões, onde seus valores devem separar-se por espaços.

Observa-se que esta ferramenta mesmo dando suporte ao uso de dimensões do tipo *string* no arquivo de entrada dos dados, os mesmos não são apresentados nas visões por não se tratarem de intervalos numéricos.

Para a Figura 32 foi utilizado um conjunto de dados denominado *cars*, cujo acompanha a ferramenta. Este conjunto possui sete dimensões [*MPG*, *Cylinders*, *Horsepower*, *Weight*, *Acceleration*, *Year*, *Origin*], sendo todas do tipo real exceto a última que possui o tipo *string*. Nesta figura a ferramenta apresenta uma relação entre as 400 instâncias e as seis dimensões numéricas.

Figura 32 - Parvis, exemplo de interface



Fonte: Ledermann (2013).

### 5.3 APLICAÇÕES DE HOFFMAN

Patrick Hoffman (Hoffman, 2013) desenvolveu uma série de aplicações em conjunto com sua tese de doutorado em 1999. Essas aplicações englobam uma série de técnicas, entre as mais referenciadas cita-se: *Radviz* – Coordenadas Circulares (Hoffman et al., 1997), Coordenadas Paralelas e Segmentos de Círculo.

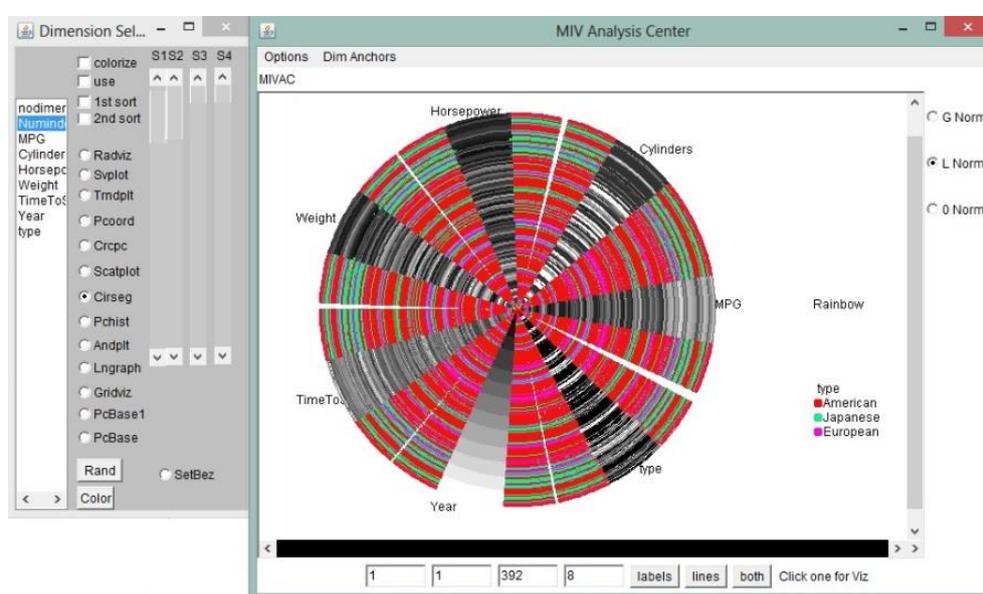
As ferramentas em conjunto, apresentam-se bem completas, por outro lado, demonstram problemas de usabilidade e às vezes param de funcionar, necessitando seu fechamento e reabertura. A versão encontrada na página pessoal de Hoffman tem como data de *upload* 05/09/2001 e a maior parte das funcionalidades se encontram numa aplicação de nome *MIVAC*.

Os dados utilizados para a geração das visões são carregados de arquivos no formato CSV, onde a primeira linha deve conter o nome das dimensões e as

demais seus respectivos valores. Observa-se que a ferramenta não faz distinções de tipos no arquivo CSV, porém dados não numéricos não são carregados.

Na Figura 33 foi utilizada a técnica de segmentos de círculo e um conjunto de dados que acompanha a ferramenta, possuindo oito dimensões [*Numindex1*, *MPG*, *Cylinders*, *Horsepower*, *Weight*, *TimeToSixty*, *Year*, *type*] e 392 instâncias. A primeira dimensão por se tratar de um número sequencial e a última por estar em formato de texto foram desconsideradas na geração da visão.

Figura 33 - Aplicação MIVAC de Hoffman, utilizando Segmentos de Círculo



Fonte: Hoffman (2013).

#### 5.4 GGOBI

GGobi (GGobi, 2013) é mantido pela *The GGobi Foundation, Inc* uma fundação sem fins lucrativos de apoio a pesquisa de gráficos estatísticos. É uma aplicação para a geração de visualizações, tendo como principais visões geradas: as coordenadas paralelas e gráficos de dispersão – *scatterplots matrices*.

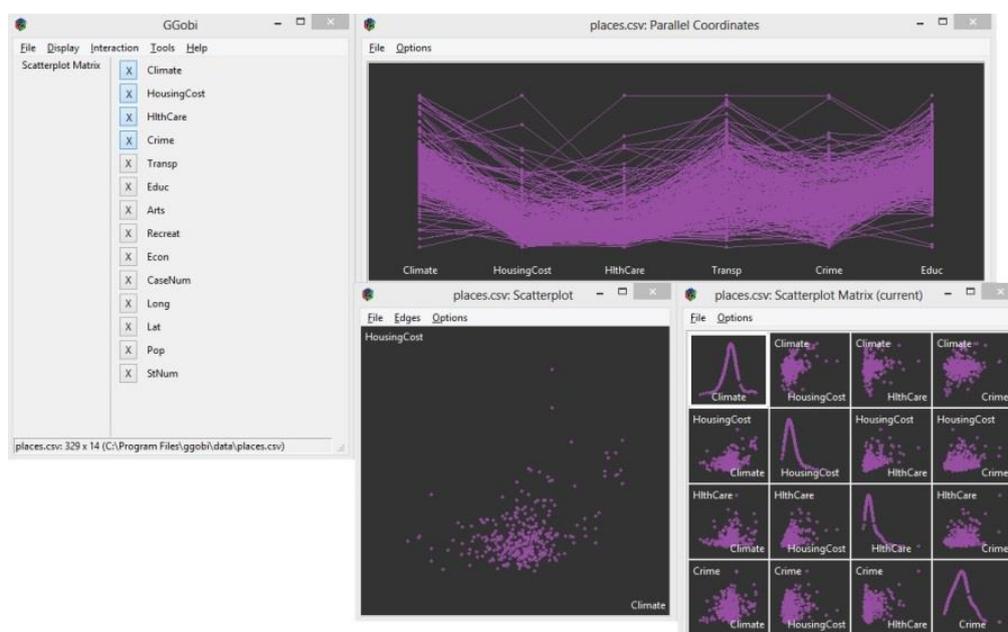
Esta aplicação se demonstrou mais moderna, com uma interface mais agradável possuindo: opção de fácil seleção de quais dimensões devem ser exibidas, interação por *brush*, possibilidade de ordenação das dimensões e

alteração dos dados em tempo real. A versão encontrada no site tem como data 12/03/2010 e uma atualização para 64 bits em 10/06/2012.

A ferramenta disponibiliza para a importação dos dados dois formatos, o XML e o CSV. A estrutura do XML é definida pelo arquivo que acompanha a ferramenta “ggobi.dtd”, já o CSV é construído pela primeira linha que representa o nome das dimensões e as demais linhas seguem com seus respectivos valores.

Para a Figura 34 foi utilizado um conjunto de dados que acompanha a ferramenta denominado *places*, que possui 14 dimensões das quais seis foram escolhidas para a formulação do exemplo.

Figura 34 - GGobi - exemplo de visões geradas



Fonte: GGOBI (2013).

## 5.5 XMDVTOOL

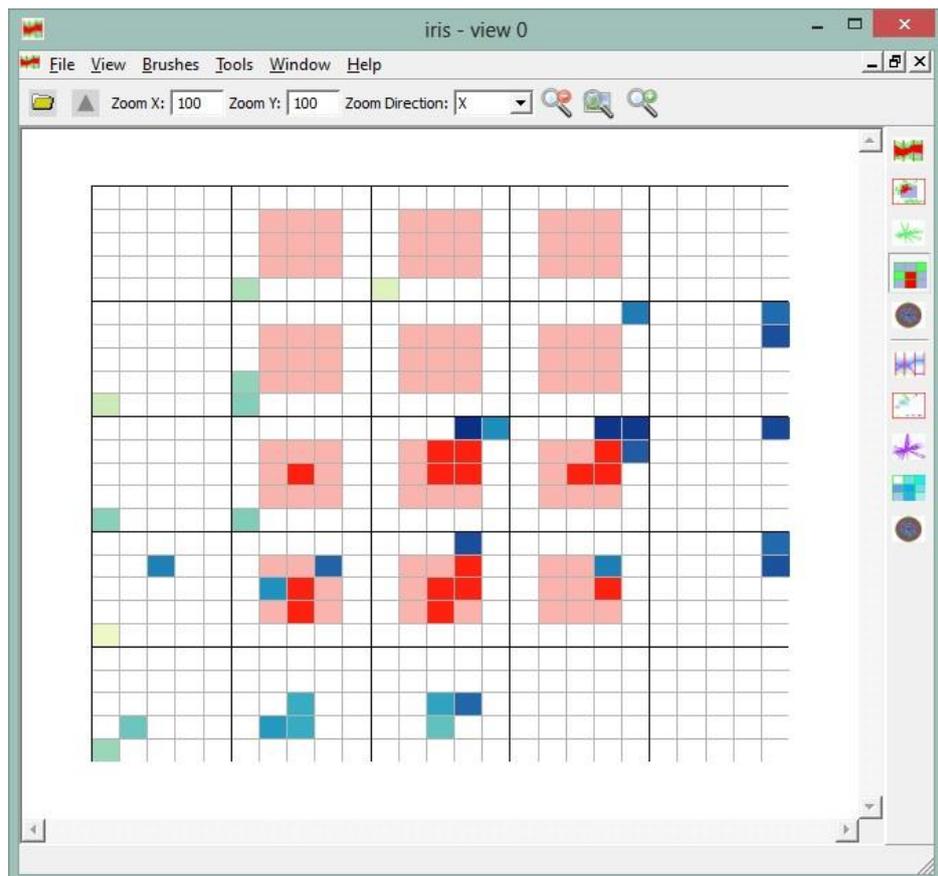
XmdvTool é um *software* de domínio público para a exploração visual interativa de conjuntos de dados multivariados, inicialmente criado por Ward (1994). Disponível em todas as principais plataformas, como UNIX, Linux, Mac e Windows a XmdvTool é desenvolvida usando *Qt* e *Eclipse CDT*, suportando cinco técnicas de

visualização de informação: *Scatterplots*, *Star Glyphs*, *Coordenadas Paralelas*, *Dimensional Stacking* e *displays* baseados em *pixel*.

A ferramenta está na versão 8.0 com data de 20/10/2010, apresentando uma interface bastante intuitiva e suportando somente interatividades como *brush* e a escolha dinâmica das dimensões. Ela também possui integração com dois formatos de arquivos, um nativo chamado de “okc” e com manual da estrutura disponível em <http://davis.wpi.edu/xmdv/fileformats.html>, como também arquivos no formato CSV.

A Figura 35 mostra um exemplo da técnica de *Dimensional Stacking* utilizando dados do *IrisDB* que acompanham a aplicação, este conjunto possui apenas quatro dimensões [*sepal\_length*, *sepal\_width*, *petal\_length*, *petal\_width*].

Figura 35 - *Dimensional Stacking* utilizando a XmdvTool e dados do *IrisDB*



Fonte: Ward (1994)

## 5.6 INFOVIS TOOLKIT

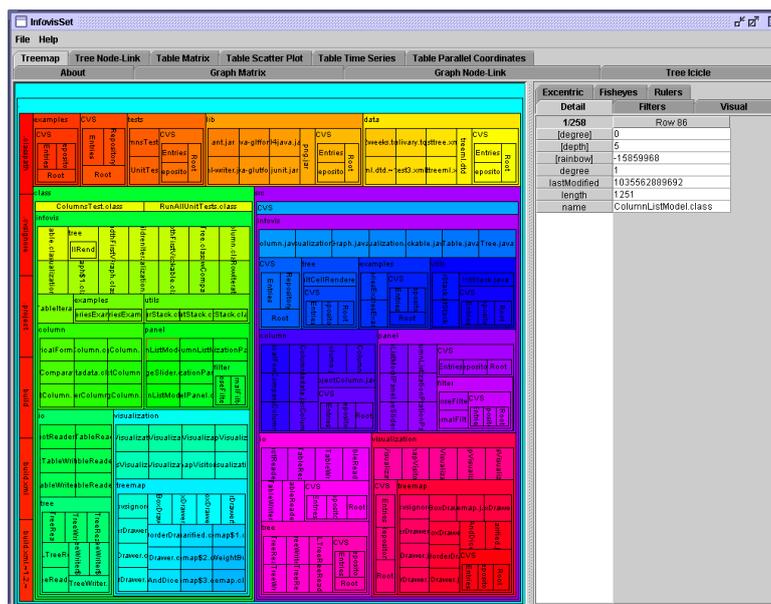
O InfoVis *Toolkit* – IVTK – é um conjunto de ferramentas gráficas interativas escrito em Java/Swing, com o intuito de facilitar o desenvolvimento de aplicações e componentes de Visualização de Informação e é mantido por Jean-Daniel Fekete e colaboradores (Fekete, 2013).

O *toolkit* foi encontrado na versão 0.9beta2 em um repositório do *sourceforge* <<http://sourceforge.net/projects/ivtk/>> e sua última atualização foi em 25/04/2013, divergindo da versão 0.9beta de 2005 encontrada na página pessoal de Jean-Daniel Fekete, o que demonstra que aparentemente o projeto tem uma continuidade mesmo sem o fundador.

Não sendo uma aplicação, esse conjunto de ferramentas pode auxiliar na construção de outros softwares que tenham como objetivo a geração de visões, sendo que nativamente o IVTK já possui suporte para nove tipos de visualizações, mas com estruturas que possibilitam a implementação de novas funcionalidades.

Um exemplo de visualização encontrada no site do desenvolvedor é a *Treemap* na Figura 36.

Figura 36 - InfoVis *Toolkit* - Exemplo de *Treemap*



Fonte: Fekete (2013).

## 5.7 CONSIDERAÇÕES FINAIS

Este capítulo procurou demonstrar uma série de ferramentas que podem ser utilizadas para a geração de visões. Todas que aqui constam, foram escolhidas por serem – de maneira mais frequente – citadas em bibliografias.

Pode-se notar que mesmo com certa continuidade as ferramentas não apresentam uma aplicação completa, sendo que muitas delas contêm problemas de usabilidade e funcionalidade, ou apresenta um número limitado de visões que podem ser geradas. No caso das aplicações de Hoffman, onde um conjunto bastante completo de geradores de visualizações é encontrado, sofre por estar desatualizado desde 2001. Já a XmdvTool bastante utilizada nas gerações das visões propostas por Keim (1997) o programa apresenta fechamentos inesperados em determinadas tarefas.

Quanto à entrada de dados, todas as aplicações oferecem suporte a arquivos no formato CSV ou equivalente. Neste ponto, destaca-se a importância do desenvolvimento ou escolha de uma ferramenta de ETL, para a criação quantitativa e qualitativa das dimensões que serão usadas pelas ferramentas de InfoVis.

## 6 EXPERIMENTO

O experimento a ser desenvolvido utilizará amostras dos três conjuntos de dados públicos retirados do *PSCL DataShop*, onde a integridade dos dados é garantida possuindo diversas publicações com o uso desse repositório. As amostras tiveram suas instâncias e atributos selecionados. Foram removidos atributos que apresentavam o mesmo significado, em branco, incompletos ou que continham algum controle do *software* utilizado na criação do *dataset*. Após as seleções as amostras foram transformadas para um formato numérico utilizando uma ferramenta citada na segunda seção deste capítulo. Por fim os dados das amostras foram apresentados utilizando as técnicas de visualização Scatterplot Matrix, Coordenadas Paralelas, RadViz e *Survey Plots*. Onde, nas seções três, quatro e cinco, as visões geradas são acompanhadas de sua análise, bem como alguns dos passos utilizados nas ferramentas, os quais possibilitaram a análise e a descoberta de informação. O processo completo e o modelo utilizado são demonstrados na primeira seção.

As visualizações *Dimensional Stack* e *Circle Segments* não são utilizadas neste experimento. A primeira visualização é limitada a quatro dimensões, não sendo válida para os *datasets* analisados. Quanto à segunda forma, ela é específica para séries temporais, não sendo, portanto útil neste experimento.

### 6.1 MODELO

Para que os dados contidos nos conjuntos fossem apresentados na forma de visões, o experimento utilizou-se do modelo proposto por Card et al. (1999) e contido no Capítulo 2.

A fim de demonstrar como as etapas do modelo de Card et al. (1999), criou-se uma adaptação do modelo, o qual está contido na Figura 37, seguida de sua explicação.

Figura 37 – Adaptação do Modelo de Card et al. (1999)



O item 1 apontado no modelo é referente aos dados fornecidos pelo *PSCL DataShop*, os quais estão no formato de texto. Em 2 converteu-se os dados no formato texto para dados estruturados em tabelas (planilhas) (Zat, 2012). No item 3, que representa uma iteração na mesma etapa do processo, coletou-se uma amostra mantendo a mesma estrutura tabular. Em 4, baseando-se nas ferramentas selecionadas, criou-se uma aplicação denominada *Dimensionador*, a qual teve a função de converter os dados contidos nas planilhas para um formato compatível com as ferramentas de visualização, substituindo assim a necessidade de uma ferramenta de ETL. O item 5, que consta três vezes na adaptação, se refere às estruturas visuais e processos adotados conforme a ferramenta selecionada. Por fim, em 6 referem-se à criação das visões e em 7 as interações possíveis em cada visão, como por exemplo, o *brush*, a translação e o *zoom*.

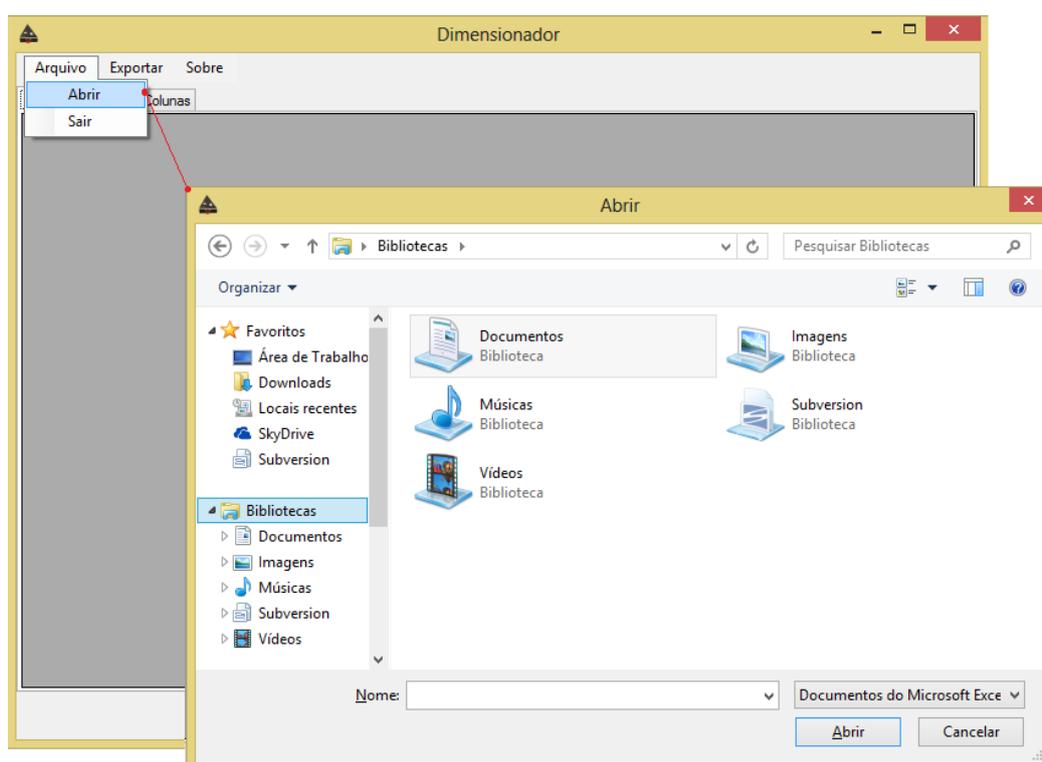
## 6.2 DIMENSIONADOR

As amostras utilizadas se apresentam no formato de planilhas, onde os atributos são representados pelas colunas e seus valores representados por intervalos numéricos ou por uma sequência de caracteres. Uma vez em que as ferramentas mencionadas no Capítulo 5 restringem os atributos a formatos numéricos, foi necessário compatibilizar os conjuntos de dados. Assim surgiu a necessidade de criar uma ferramenta que avaliasse os valores contidos e os compatibilizasse no formato exigido por cada ferramenta.

Para suprir essa necessidade foi criada a aplicação de nome Dimensionador. Que tem suporte para – somente – conjuntos de dados contidos em arquivos no formato *x/s* (Microsoft Excel 97 – 2003) como fonte de dados.

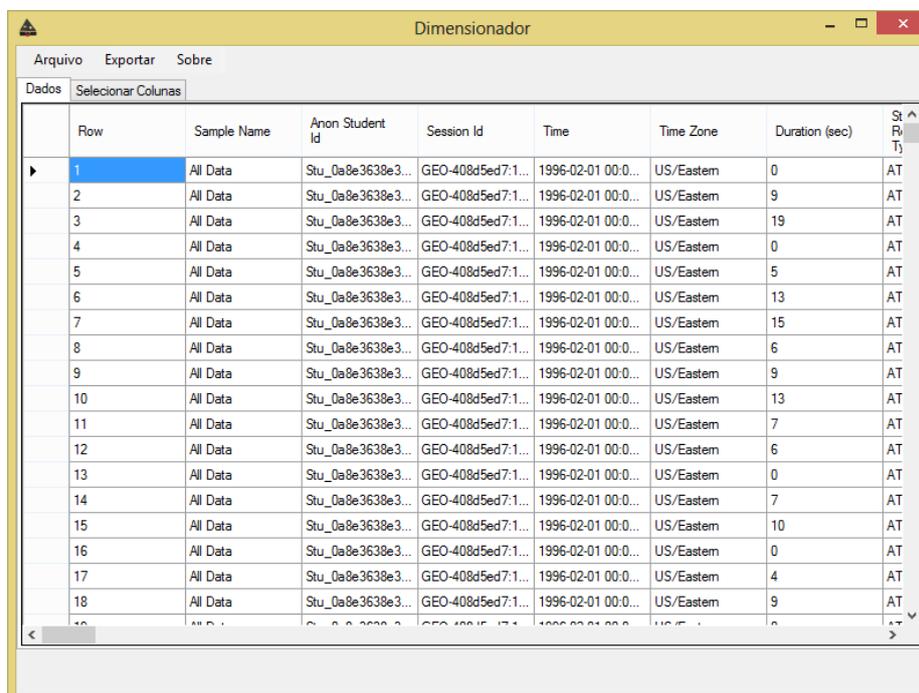
As próximas figuras demonstram o processo que deve ser executado para compatibilizar estes arquivos para os formatos exigidos pelas demais ferramentas.

Figura 38 – *Layout* de importação de *x/s*



A seleção do arquivo – fonte – de dados é apresentada na Figura 38, onde o usuário pode navegar por diretórios até selecionar a planilha desejada.

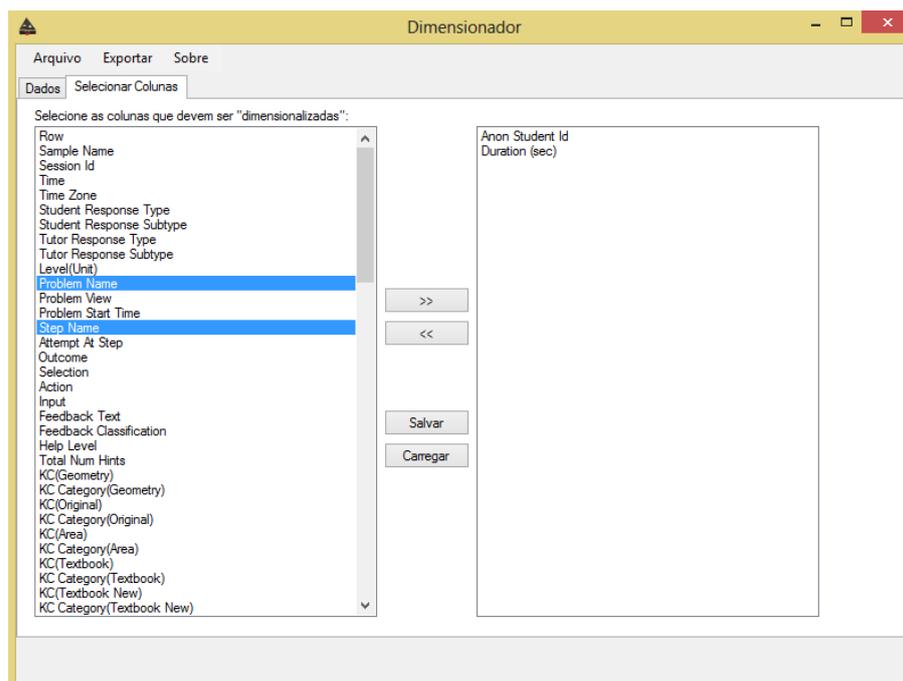
Figura 39 – Exibição dos dados contidos no arquivo



Row	Sample Name	Anon Student Id	Session Id	Time	Time Zone	Duration (sec)	St R T
1	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	0	AT
2	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	9	AT
3	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	19	AT
4	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	0	AT
5	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	5	AT
6	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	13	AT
7	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	15	AT
8	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	6	AT
9	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	9	AT
10	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	13	AT
11	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	7	AT
12	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	6	AT
13	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	0	AT
14	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	7	AT
15	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	10	AT
16	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	0	AT
17	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	4	AT
18	All Data	Stu_0a8e3638e3...	GEO-408d5ed7:1...	1996-02-01 00:0...	US/Eastern	9	AT

Ao selecionar o arquivo desejado, a aplicação irá exibir os dados na aba Dados conforme mostra a Figura 39.

Figura 40 – Seleção de colunas para transformação.



Seleção de colunas para transformação:

Selecione as colunas que devem ser "dimensionalizadas":

- Row
- Sample Name
- Session Id
- Time
- Time Zone
- Student Response Type
- Student Response Subtype
- Tutor Response Type
- Tutor Response Subtype
- Level(Unit)
- Problem Name**
- Problem View
- Problem Start Time
- Step Name**
- Attempt At Step
- Outcome
- Selection
- Action
- Input
- Feedback Text
- Feedback Classification
- Help Level
- Total Num Hints
- KC(Geometry)
- KC Category(Geometry)
- KC(Original)
- KC Category(Original)
- KC(Area)
- KC Category(Area)
- KC(Textbook)
- KC Category(Textbook)
- KC(Textbook New)
- KC Category(Textbook New)

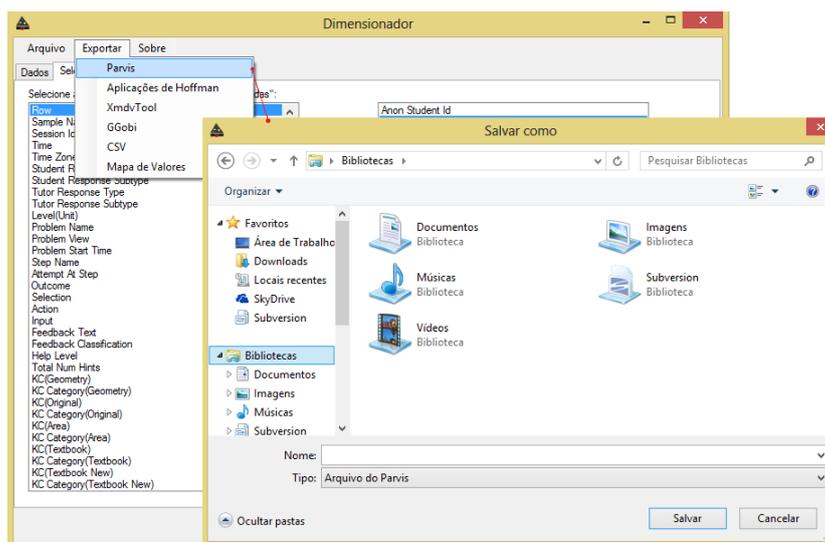
Colunas selecionadas:

- Anon Student Id
- Duration (sec)

Botões: >>, <<, Salvar, Carregar

Com o arquivo carregado na aplicação é possível fazer a seleção de quais colunas (atributos) serão dimensionadas (transformadas), esta opção se encontra na aba Selecionar Colunas (Figura 40), que permite também salvar ou carregar a atual seleção.

Figura 41 – Opções de transformação (exportação)

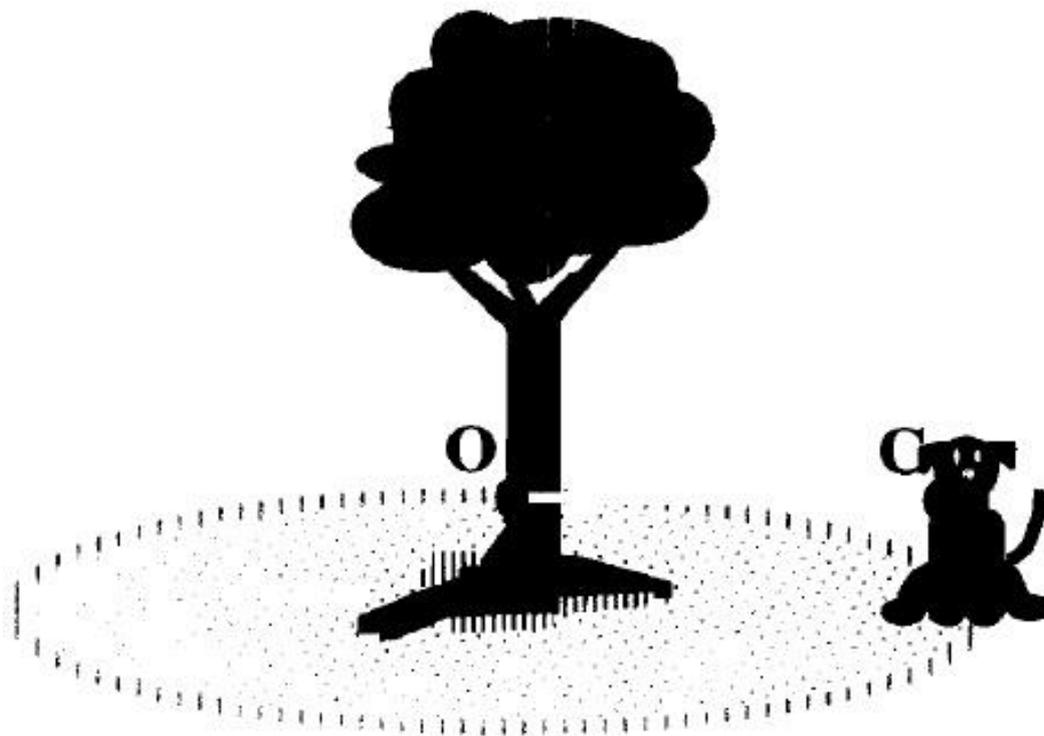


Com a seleção de colunas concluída, pode-se exportar (preparar) as informações para uma das quatro aplicações compatíveis, Parvis, Aplicações de Hoffman, XmdvTool, GGobi ou também para o formato CSV. Além de exportar os dados compatibilizados, a ferramenta possui um exportador do que chamamos de “mapa”, é um arquivo também no formato CSV que contém os valores originais de cada coluna e o valor para o qual ele foi “traduzido”. Uma amostra desse procedimento é mostrado na Figura 41.

### 6.3 GEOMETRIA

O *dataset* de Geometria contém dados de resolução problemas por alunos utilizando o software *Geometry Tutor* (Anderson et al., 1985). A fim de projetar as visualizações foi selecionada uma amostra de quatro alunos resolvendo um mesmo problema. O problema denominado *dog on a rope* (Figura 42) é descrito a seguir.

Figura 42 – Ilustração do problema contido na amostra



“Suponha que a cachorrinha Cacau esteja amarrada a uma árvore no centro do seu quintal. A corda tem 43 pés de comprimento. No entanto, Cacau continua vagando em torno de sua árvore, conseqüentemente enrolando sua corda cada vez mais em torno da árvore. Eventualmente, Cacau fica enrolada tão perto da árvore que ela não pode andar por aí. O proprietário de Cacau, Chris, quer construir uma cerca para que ele possa acabar com a corda por completo.

Ajude Chris a descobrir a área a ser cercada que daria a Cacau espaço equivalente para se movimentar ao qual a corda originalmente lhe deu.

Uma vez que a área é determinada, Chris precisa comprar a cerca. Qual é o comprimento da cerca que ele vai precisar?”

A resposta ao problema está na Tabela 8. A solução é composta por 3 etapas principais, sendo elas: a) converter a unidade de medida do comprimento da corda de pés para metros, b) calcular a área que o cão possui para andar e c) calcular o perímetro da circunferência.

Tabela 8 – Relação de respostas do problema *dog on a rope*

Unidades	Comprimento da Corda (Raio) – Pés   Metros	Área que o cão possui para andar (Área) - Metros Quadrados	Comprimento de cerca (Circunferência) - Metros
Questão	43   13	530,93	81,67

Junto aos quatro alunos e ao problema outras dimensões foram analisadas, são elas: Duração (ao resolver uma etapa), número de soluções de um problema, etapa que está sendo resolvida, quantidade de vezes que a etapa foi repetida, área do conhecimento avaliada, se fórmula de área foi utilizada e o resultado da etapa. Todas estas dimensões estão contidas em uma amostra de dezessete instâncias do conjunto de dados.

Este conjunto é composto por informação no formato texto, impossibilitando a utilização das mesmas nas ferramentas aqui citadas. Para que estes dados (contidos na Tabela 9) pudessem ser utilizados foi necessária a produção e utilização de outra ferramenta, o Dimensionador – citado na seção anterior –, que é responsável por “preparar” as colunas que se encontram em um formato não compatível, em um formato numérico equivalente. Além da preparação também buscou-se informar que tipo de dado cada atributo representa (classificação encontrada na Tabela 2). A Tabela 10 mostra o resultado desta preparação, contendo os valores originais do conjunto, sua codificação equivalente e tipo.

Tabela 9 – Amostra selecionada do conjunto Geometria

(continua)

Estudante	Duração	ProbNome	ProbN oSolu coes	EtapaA Resol ver	Etap aVe zes	AreaCo nhecim entoAva liada	Utiliza caoFo rmula Area	Resultado Etapa
Stu_0a8e3638e3c0deb4e5e49c72286a5b83	0	DOG_ON_A_ROPE	1	(AREA QUESTION1)	1	ALT:CIRCLE-AREA	Non-area formula	CORRECT
Stu_0a8e3638e3c0deb4e5e49c72286a5b83	19	DOG_ON_A_ROPE	1	(CIRCUMFERENCE QUESTION1)	1	ALT:CIRCLE-CIRCUMFERENCE	Area formula	CORRECT

Tabela 9 – Amostra selecionado do conjunto Geometria

(continuação)

Estudante	Duracao	ProbNome	ProbNoSolucoes	EtapaAResolver	EtapasVezez	AreaConhecimentoAvaliada	UtilizacaoFormulaArea	ResultadoEtapa
Stu_0a8e3638e3c0deb4e5e49c72286a5b83	0	DOG_ON_A_ROPE	2	(AREA QUESTION1)	1	ALT:CIR CLE- AREA	Non-area formula	CORRECT
Stu_0a8e3638e3c0deb4e5e49c72286a5b83	10	DOG_ON_A_ROPE	2	(CIRCUMFERENCE QUESTION1)	1	ALT:CIR CLE- CIRCUM FERENC E	Area formula	CORRECT
Stu_2b6f028b112758c7ce53d84207eab22c	0	DOG_ON_A_ROPE	1	(AREA QUESTION1)	1	ALT:CIR CLE- AREA	Non-area formula	INCORRECT
Stu_2b6f028b112758c7ce53d84207eab22c	1	DOG_ON_A_ROPE	1	(AREA QUESTION1)	2	ALT:CIR CLE- AREA	Non-area formula	CORRECT
Stu_2b6f028b112758c7ce53d84207eab22c	80	DOG_ON_A_ROPE	1	(CIRCUMFERENCE QUESTION1)	1	ALT:CIR CLE- CIRCUM FERENC E	Area formula	CORRECT
Stu_8fb744b51a1f14e5e8cda4e4aec68e2f	0	DOG_ON_A_ROPE	1	(AREA QUESTION1)	1	ALT:CIR CLE- AREA	Non-area formula	INCORRECT
Stu_2b6f028b112758c7ce53d84207eab22c	1	DOG_ON_A_ROPE	1	(AREA QUESTION1)	2	ALT:CIR CLE- AREA	Non-area formula	CORRECT

Tabela 9 – Amostra selecionado do conjunto Geometria

(conclusão)

Estudante	Duracao	ProbNome	ProbNoSolucoes	EtapaAResolver	Etapavezes	AreaConhecimentodaAvaliada	UtilizacaoFormulaArea	ResultadoEtapa
Stu_8fb744b51a1f14e5e8cda4e4aec68e2f	11	DOG_ON_A_ROPE	1	(CIRCUMFERENCE QUESTION1)	1	ALT:CIRCLE-CIRCUMFERENCE	Area formula	INCORRECT
Stu_8fb744b51a1f14e5e8cda4e4aec68e2f	1	DOG_ON_A_ROPE	1	(CIRCUMFERENCE QUESTION1)	2	ALT:CIRCLE-CIRCUMFERENCE	Area formula	CORRECT
Stu_cda7a650c5856cf2f6738072447d7825	0	DOG_ON_A_ROPE	1	(AREA QUESTION1)	1	ALT:CIRCLE-AREA	Non-area formula	CORRECT
Stu_cda7a650c5856cf2f6738072447d7825	10	DOG_ON_A_ROPE	1	(CIRCUMFERENCE QUESTION1)	1	ALT:CIRCLE-CIRCUMFERENCE	Area formula	INCORRECT
Stu_cda7a650c5856cf2f6738072447d7825	1	DOG_ON_A_ROPE	1	(CIRCUMFERENCE QUESTION1)	2	ALT:CIRCLE-CIRCUMFERENCE	Area formula	CORRECT
Stu_cda7a650c5856cf2f6738072447d7825	0	DOG_ON_A_ROPE	2	(AREA QUESTION1)	1	ALT:CIRCLE-AREA	Non-area formula	CORRECT
Stu_cda7a650c5856cf2f6738072447d7825	11	DOG_ON_A_ROPE	2	(CIRCUMFERENCE QUESTION1)	1	ALT:CIRCLE-CIRCUMFERENCE	Area formula	INCORRECT
Stu_cda7a650c5856cf2f6738072447d7825	1	DOG_ON_A_ROPE	2	(CIRCUMFERENCE QUESTION1)	2	ALT:CIRCLE-CIRCUMFERENCE	Area formula	CORRECT

Tabela 10 - Resultado da preparação da amostra Geometria

(continua)

Atributo	Conversão	Tipo
Coluna: Estudante	Valor	Nominal
Stu_0a8e3638e3c0deb4e5e49c72286a5b83	0	
Stu_2b6f028b112758c7ce53d84207eab22c	1	
Stu_8fb744b51a1f14e5e8cda4e4aec68e2f	2	
Stu_cda7a650c5856cf2f6738072447d7825	3	
Duracao	Valor	Quantitativo
0	0	
19	19	
10	10	
1	1	
80	80	
11	11	
Atributo	Conversão	Nominal
ProbNome	Valor	
DOG_ON_A_ROPE	0	
ProbNoSolucoes	Valor	Quantitativo
1	1	
2	2	
EtapaAResolver	Valor	Nominal
(AREA QUESTION1)	0	
(CIRCUMFERENCE QUESTION1)	1	
EtapaVezes	Valor	Quantitativo
1	1	
2	2	
AreaConhecimentoAvaliada	Valor	Nominal
ALT:CIRCLE-AREA	0	
ALT:CIRCLE-CIRCUMFERENCE	1	

Tabela 10 – Resultado da preparação da amostra Geometria

(conclusão)

Atributo	Conversão	
UtilizacaoFormulaArea	Valor	Nominal
Non-area formula	0	
Area formula	1	
ResultadoEtapa	Valor	Ordinal
CORRECT	0	
INCORRECT	1	

As próximas sessões apresentam o resultado obtido na geração das visualizações. Não foi seguida nenhuma ordem específica neste processo. As cores das visualizações foram atribuídas pelo último atributo, o Resultado da Etapa.

### 6.3.1 *Scatterplot Matrix*

Para esta primeira visualização utilizou-se as dimensões: Estudante, Duração, Nome do Problema, Número de Soluções Possíveis, Etapa a Resolver e Resultado da Etapa. A necessidade de selecionar um conjunto menor de dimensões surgiu por uma limitação gráfica, por se tratar de um plano cartesiano de gráficos, a *scatterplot matrix* apresentou gráficos muito pequenos, impossibilitando a análise. Assim, dentro das dimensões existentes, seis consideradas de maior importância foram selecionadas. Outras conclusões poderiam surgir ao adicionar ou trocar as dimensões escolhidas, já que todas elas são importantes no contexto.

Para geração das próximas figuras, foi utilizada uma interação chamada de *brush*, ela é responsável por prover uma área selecionável com o intuito de destacar certas informações. Dentro de um dos gráficos gerados, é possível selecionar um determinado número de pontos, assim, o *brush* colore em todos os outros gráficos os pontos onde a mesma informação se encontra. As análises podem ser conferidas nas Figuras 43, 44 e 45.

Figura 43 – *Brush* aplicado à dimensão Duração

Os estudantes resolveram as etapas com durações bem próximas, com exceção do aluno “1”, que em uma das etapas teve uma demora maior, mesmo assim a respondeu corretamente.

Figura 44 – *Brush* aplicado à dimensão Resultado da Etapa

Um total de cinco passos foram respondidos incorretamente, exceto o aluno “0” que respondeu todos corretamente.

Figura 45 – *Brush* aplicado à dimensão Etapa a Resolver

Quando o passo se tratava de cálculo de área, todos os alunos respondiam quase que instantaneamente.

### 6.3.2 RadViz, RadViz3D e PolyViz

O conceito destas três visualizações, PolyViz, RadViz3D e RadViz é muito semelhante. Uma vez que, a PolyViz e a RadViz3D auxiliam ou completam a compreensão da RadViz. Nestas visualizações as dimensões exercem uma espécie de força sobre os dados, e o conjunto de forças entre as dimensões definirão em que ponto a informação será desenhada.

A RadViz por se tratar de uma visualização plana (2D) pode apresentar mais de uma informação desenhada no mesmo ponto, o que dificulta o reconhecimento do volume de informação em determinada parte da visualização. Para auxiliar neste problema a PolyViz apresenta em cada dimensão pequenos traços em direção ao ponto em que a informação foi desenhada, já a RadViz3D usa um ambiente 3D para dar uma percepção melhor do volume.

Estas visões utilizam uma das dimensões para colorir os pontos, esta mesma dimensão não será incluída na formação dos segmentos (eixos). Neste

experimento a dimensão Resultado da Etapa foi selecionada para esta classificação. As cores foram atribuídas automaticamente pelas ferramentas, onde nas duas primeiras figuras a cor vermelha representa respostas incorretas e a cor rosa as corretas, já na terceira figura – a RadViz – a cor amarela representa as respostas incorretas e o azul as corretas.

As visões têm como objetivo apresentar a informação formando “grupos”. O resultado obtido utilizando os dados da amostra é exibido nas Figuras 46, 47 e 48.

Figura 46 – PolyViz sobre amostra de dados Geometria

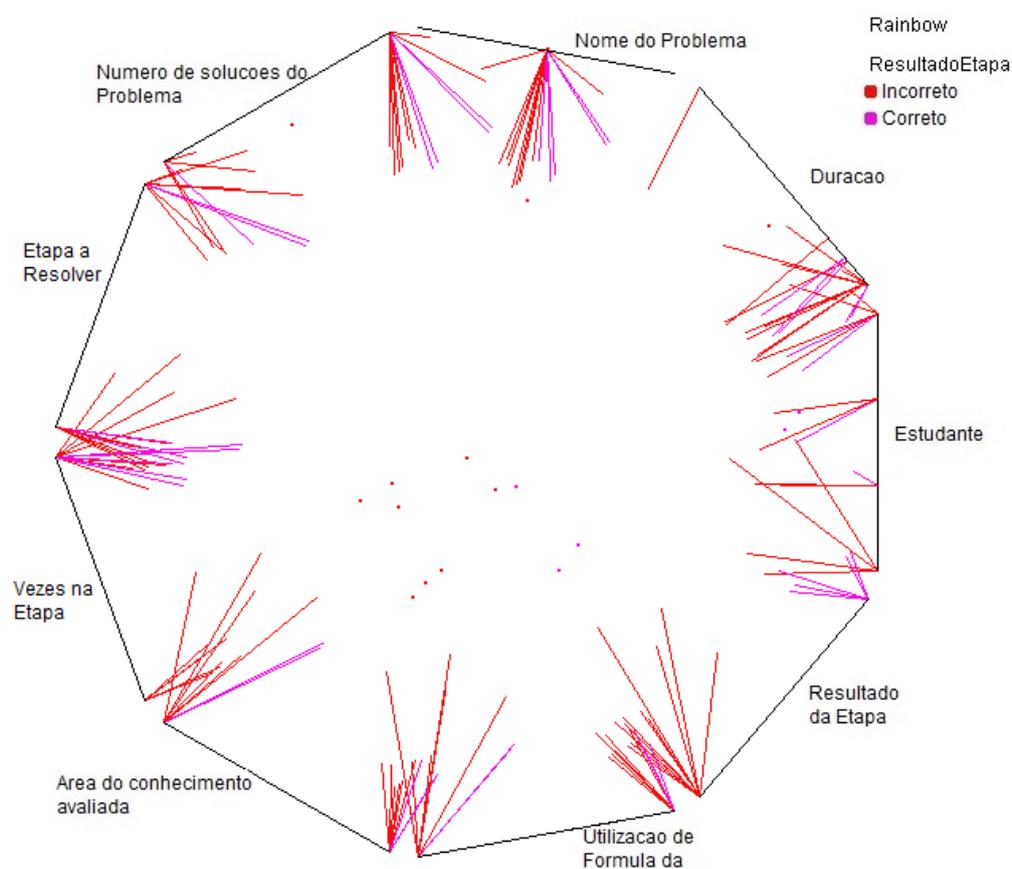


Figura 47 – RadViz3D sobre amostra de dados Geometria

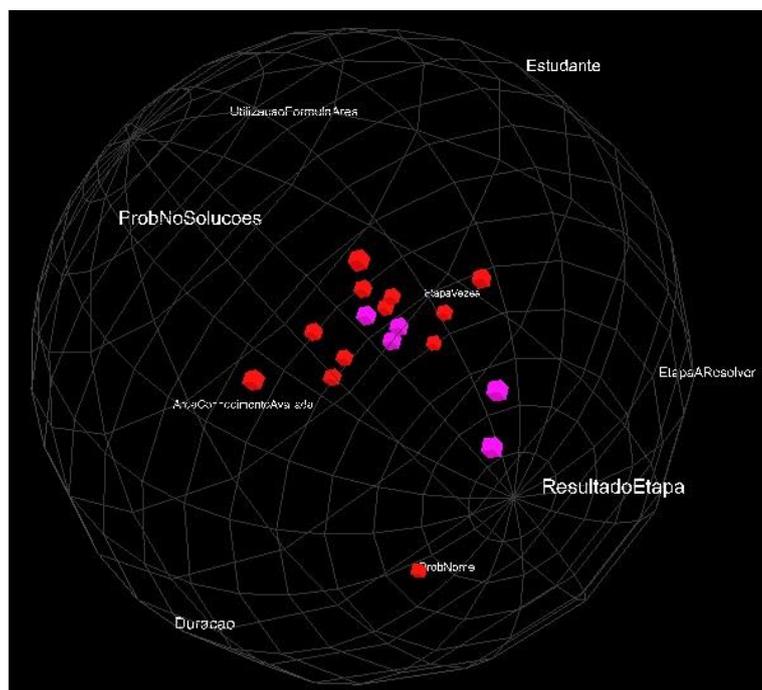
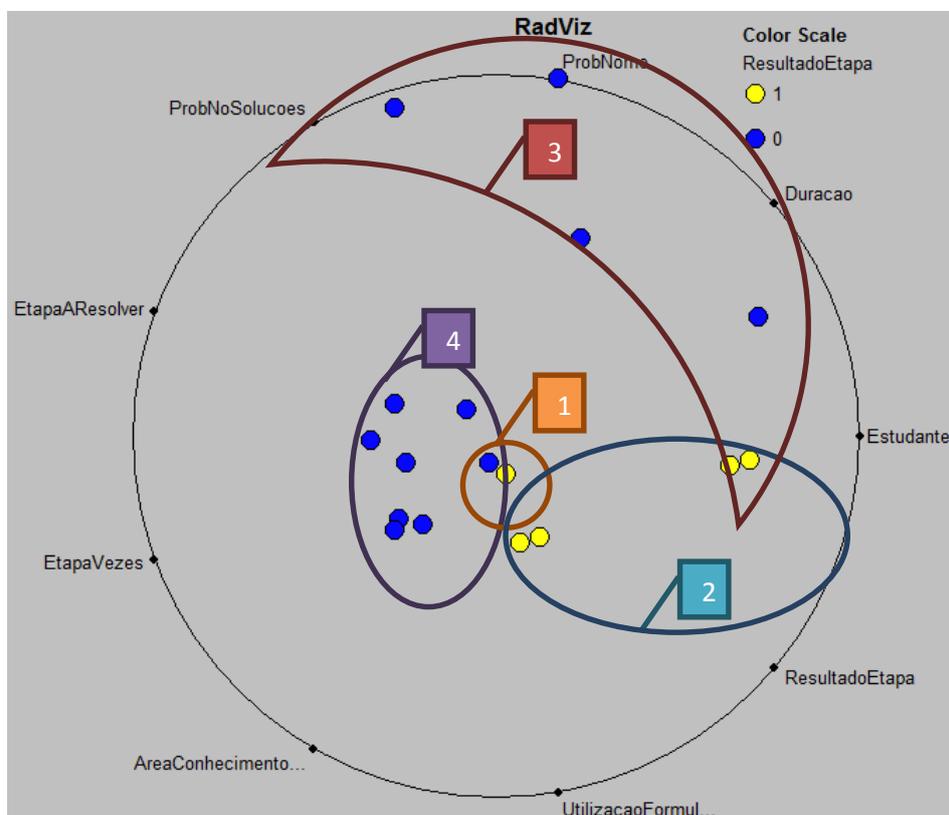


Figura 48 – RadViz sobre a amostra de dados Geometria



As três visões precedentes foram geradas pelas Aplicações de Hoffman, sendo que, como em outras ferramentas, sendo que como em outras ferramentas, nestas ao selecionar um dos pontos é possível visualizar qual instância originou o ponto desenhado. Cada ponto representa um passo na construção da solução. Assim, a partir da Figura 48, conclui-se que:

A região marcada na figura com o número 1 consiste na seleção de dois pontos, os quais representam duas instâncias da amostra. O primeiro ponto colorido com a cor azul (azul representa a resposta correta) refere-se a um aluno resolvendo a primeira etapa do problema, sem repetições e com e referente à área de conhecimento *ALT:CIRCLE-CIRCUMFERENCE*. O segundo ponto de cor amarela (resposta incorreta) possui as mesmas informações do primeiro ponto. Assim percebe-se que este grupo foi formado por instâncias semelhantes que resultarão em resultados opostos.

A região marcada pelo número 2 ao verificar porque a concentração de erros – resultado da etapa incorreta (cor amarela) – se concentrava mais ao lado direito do círculo, descobriu-se que todos estes alunos erraram a primeira etapa.

A região marcada pelo número 3 os pontos que ficaram com uma distância maior do centro, representam todas as etapas respondidas pelos 4 alunos no tempo de 0 (zero) segundos.

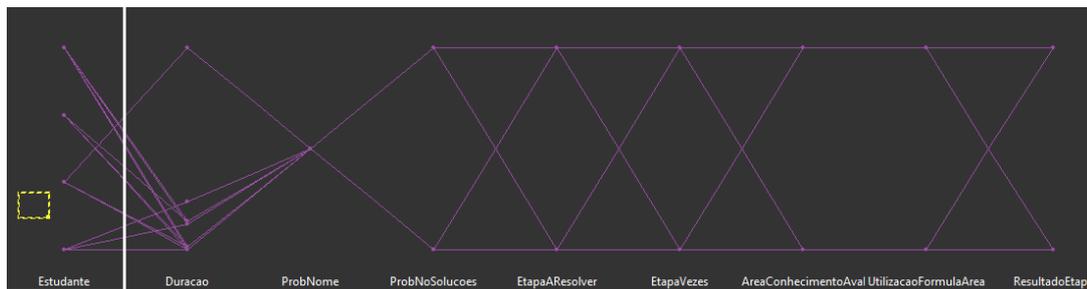
A região marcada pelo número 4 assinala uma concentração de acertos mais ao centro do círculo. Descobriu-se um padrão, existem dois acertos para cada aluno da amostra (quatro alunos ao total), sendo que nenhum destes acertos foi instantâneo (duração igual à zero).

### 6.3.3 COORDENADAS PARALELAS

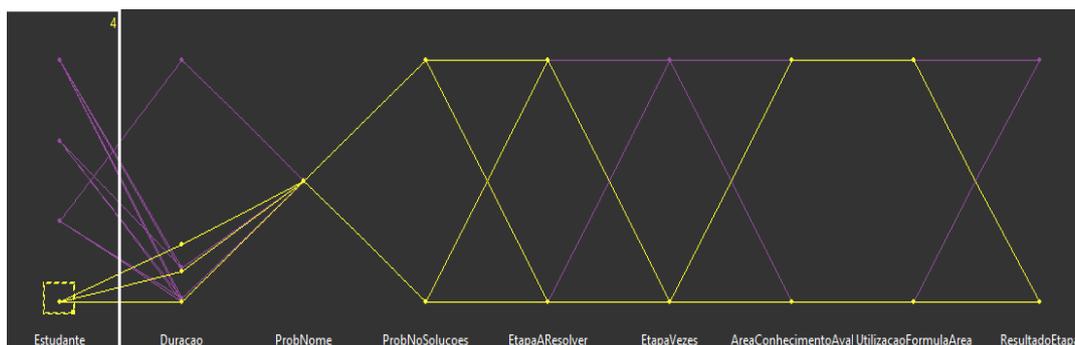
As visualizações resultantes da técnica de coordenadas paralelas formam uma espécie de gráfico x e y de linhas, onde ao término de uma dimensão Y uma segunda dimensão começa, e o eixo X é o “percurso” que a linha faz entre as dimensões.

Utilizando-se desta técnica e também da interação por *brush* gerou-se as Figuras 49, 50, 51, 52 e 53.

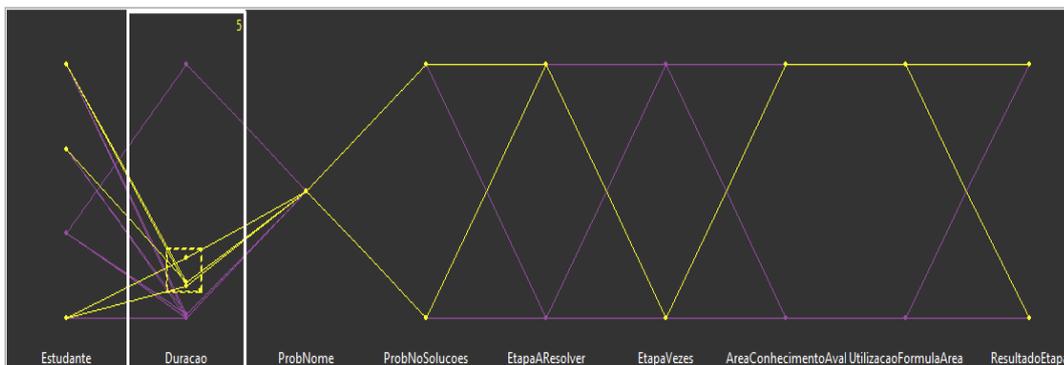
Figura 49 – Visão geral da amostra nas Coordenadas Paralelas



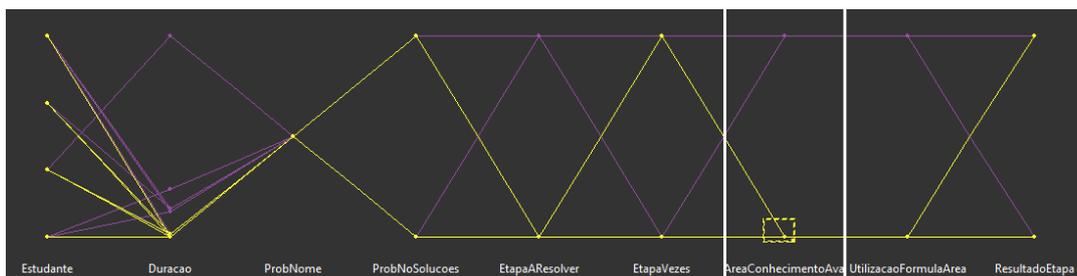
Na Figura 49 é possível ter uma visão geral do comportamento da turma e dos valores máximos e mínimos de todas as dimensões utilizadas, exemplo disso, são os quatro pontos iniciais que representam os quatro alunos, passando por durações diferenciadas e se encontrando em um único ponto “problema”.

Figura 50 – Utilização de *brush* no primeiro estudante

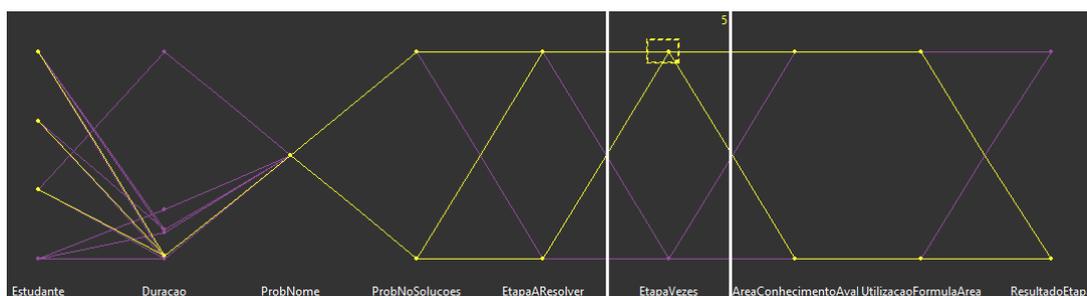
Na Figura 50 o primeiro aluno, por exemplo, não repetiu nenhuma etapa, e mesmo em áreas do conhecimento distintas sendo avaliadas, com ou não a utilização de fórmulas da área, ele respondeu sempre corretamente.

Figura 51 – Utilização de *brush* na duração

Na Figura 51 os alunos que responderam no intervalo de 10 a 20 segundos não tiveram a etapa *AREA QUESTION1* para responder.

Figura 52 – Utilização de *brush* na Área de Conhecimento Avaliada

Na Figura 52 duas coisas foram descobertas: a) A área do conhecimento avaliada, nesta amostra, está diretamente relacionada com a utilização ou não de fórmula. b) Sempre que a área de conhecimento avaliada fosse *ALT:CIRCLE-AREA* as respostas são quase que imediatas.

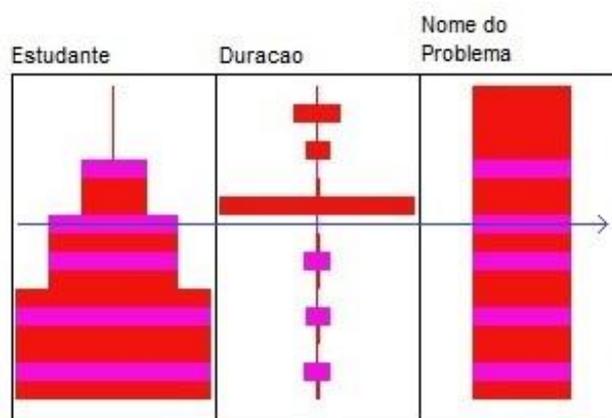
Figura 53 – Utilização de *brush* na dimensão Vezes na Etapa

Na Figura 53 sempre que a etapa foi executada duas vezes a resposta foi correta.

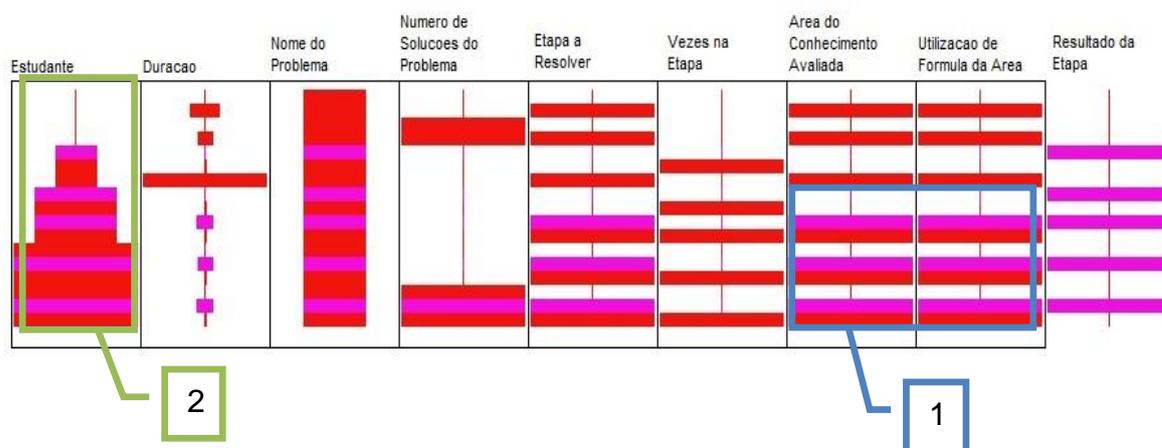
### 6.3.4 Survey Plots

O valor de cada dimensão é representado por uma barra horizontal com largura diferenciada para cada valor que ela possa representar. Na Figura 54 encontram-se as três primeiras dimensões da amostra. Cada dimensão é representada por um conjunto de barras horizontais, sendo que sua largura é dada conforme o valor que ela possui. Além disso, representado por uma seta azul, cada dimensão possui uma relação com a dimensão seguinte, onde, para determinado registro do conjunto de dados, todos os valores se encontraram na mesma altura. Porém cada um na sua respectiva dimensão.

Figura 54 – Survey Plots, exemplo de funcionamento



As barras horizontais foram coloridas pela ferramenta, que selecionou uma cor para cada valor da última dimensão – a dimensão Resultado –, onde a cor vermelha representa as respostas incorretas, e a cor rosa as corretas.

Figura 55 – Análise realizada no *Survey Plots*

Com a visualização da Figura 55 pode-se concluir que. Em “1”, como anteriormente, a área do conhecimento avaliada está diretamente relacionada com a utilização de fórmula, inclusive nos acertos e erros. Em “2” a última etapa foi acertada por todos.

#### 6.4 LÍNGUA CHINESA

Para realizar os testes sobre o conjunto de dados da Língua Chinesa, selecionou-se uma amostra de quatro alunos resolvendo um mesmo problema (identificado como I01w01), tendo um total de 89 instâncias.

Dentre os atributos existentes selecionou-se os seguintes: Id do Estudante, Lição do Curso, Nível da Seção, Problema, Passo, Componente responsável pela entrada de dados, Número de Repetições, Texto Retornado e Resposta.

Assim como no conjunto de dados Geometria, a amostra do conjunto da Língua Chinesa (Tabela 11) também necessitou ter os dados preparados na ferramenta Dimensionador para que a mesma pudesse ser utilizada nas ferramentas propostas.

Esta preparação incluiu uma conversão de dados originais para dados numéricos, e também a inclusão do tipo conforme apresenta a

Tabela 12.

Tabela 11 – Amostra selecionada do conjunto de dados da Língua Chinesa

(continua)

Problema	Passo	ComponenteEntradaDeDados	NumeroDeRepeticoes	TextoRetornado	Resposta
I01w01	1	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	1	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	2	radioGroup UpdateRadioButton	1	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	2	radioGroup UpdateRadioButton	2	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	2	radioGroup UpdateRadioButton	3	Focus on the pitch change and listen to the sound again.	HINT
I01w01	2	radioGroup UpdateRadioButton	4	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	2	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	3	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	3	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	1	radioGroup UpdateRadioButton	1	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	1	radioGroup UpdateRadioButton	2	Focus on the pitch change and listen to the sound again.	HINT
I01w01	1	radioGroup UpdateRadioButton	3	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	1	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	2	radioGroup UpdateRadioButton	1	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	2	radioGroup UpdateRadioButton	2	Focus on the pitch change and listen to the sound again.	HINT
I01w01	2	radioGroup UpdateRadioButton	3	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	2	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	4	radioGroup UpdateRadioButton	1	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	4	radioGroup UpdateRadioButton	2	Focus on the pitch change and listen to the sound again.	HINT

Tabela 11 – Amostra selecionada do conjunto de dados da Língua Chinesa

(continuação)

Problema	Passo	ComponenteEntradaDeDados	NumeroDeRepeticoes	TextoRetornado	Resposta
I01w01	4	radioGroup UpdateRadioButton	3	The sound pitch is going down, i.e., it is high-falling. Listen again.	HINT
I01w01	4	radioGroup UpdateRadioButton	4	This is the forth tone. Its pitch goes down. Please click the correct answer.	HINT
I01w01	4	radioGroup UpdateRadioButton	4	This is the forth tone. Its pitch goes down. Please click the correct answer.	HINT
I01w01	4	radioGroup UpdateRadioButton	5	The sound pitch is going down, i.e., it is high-falling. Listen again.	HINT
I01w01	4	radioGroup UpdateRadioButton	6	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	4	radioGroup UpdateRadioButton	7	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	4	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	5	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	5	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	1	radioGroup UpdateRadioButton	1	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	1	radioGroup UpdateRadioButton	2	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	1	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	2	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	2	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	6	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	6	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	7	radioGroup UpdateRadioButton	1	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	7	radioGroup UpdateRadioButton	2	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	7	radioGroup UpdateRadioButton	3	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	7	done ButtonPressed	1	No Hint or Other Message Text	CORRECT

Tabela 11 – Amostra selecionada do conjunto de dados da Língua Chinesa

(continuação)

Problema	Passo	ComponenteEntradaDeDados	NumeroDeRepeticoes	TextoRetornado	Resposta
I01w01	3	radioGroup UpdateRadioButton	1	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	3	radioGroup UpdateRadioButton	2	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	3	done ButtonPressed	1	No HInt or Other Message Text	CORRECT
I01w01	4	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	4	done ButtonPressed	1	No HInt or Other Message Text	CORRECT
I01w01	5	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	5	done ButtonPressed	1	No HInt or Other Message Text	CORRECT
I01w01	3	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	3	done ButtonPressed	1	No HInt or Other Message Text	CORRECT
I01w01	8	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	8	done ButtonPressed	1	No HInt or Other Message Text	CORRECT
I01w01	9	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	9	done ButtonPressed	1	No HInt or Other Message Text	CORRECT
I01w01	4	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	4	done ButtonPressed	1	No HInt or Other Message Text	CORRECT
I01w01	1	radioGroup UpdateRadioButton	1	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	1	radioGroup UpdateRadioButton	2	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	1	done ButtonPressed	1	No HInt or Other Message Text	CORRECT
I01w01	2	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	2	done ButtonPressed	1	No HInt or Other Message Text	CORRECT
I01w01	3	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	3	done ButtonPressed	1	No HInt or Other Message Text	CORRECT
I01w01	4	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	4	done ButtonPressed	1	No HInt or Other Message Text	CORRECT

Tabela 11 – Amostra selecionada do conjunto de dados da Língua Chinesa

(continuação)

Problema	Passo	ComponenteEntradaDeDados	NumeroDeRepeticoes	TextoRetornado	Resposta
I01w01	5	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	5	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	6	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	1	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	1	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	1	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	1	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	1	radioGroup UpdateRadioButton	1	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	1	radioGroup UpdateRadioButton	2	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	1	radioGroup UpdateRadioButton	3	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	1	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	2	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	2	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	3	radioGroup UpdateRadioButton	1	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	3	radioGroup UpdateRadioButton	2	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	3	radioGroup UpdateRadioButton	3	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	3	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	4	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT

Tabela 11 – Amostra selecionada do conjunto de dados da Língua Chinesa  
(conclusão)

Problema	Passo	ComponenteEntradaDeDados	NumeroDeRepeticoes	TextoRetornado	Resposta
I01w01	4	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	1	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	5	radioGroup UpdateRadioButton	1	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	5	radioGroup UpdateRadioButton	2	Focus on the pitch change and listen to the sound again.	HINT
I01w01	5	radioGroup UpdateRadioButton	3	No, this is not correct. Please click on "hint" in the upper right corner for help.	INCORRECT
I01w01	5	radioGroup UpdateRadioButton	4	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	5	done ButtonPressed	1	No Hint or Other Message Text	CORRECT
I01w01	6	radioGroup UpdateRadioButton	1	Correct! Now click on Done to go to the next one.	CORRECT
I01w01	6	done ButtonPressed	1	No Hint or Other Message Text	CORRECT

Tabela 12 – Resultado da preparação da amostra do conjunto de dados da Língua Chinesa

(continua)

Atributo	Conversão	Tipo
IdEstudante	Valor	Nominal
Stu_7792dd2aad4d5547e11d9317b03c32c6	0	
Stu_7ac0299cbdf2c39b2642a5311ee9c5bc	1	
Stu_80736f26d8cb7d6cea55f49176d23e4d	2	
Stu_ecee596a242de13b779391cdaa2c528d	3	
LicaoDoCurso	Valor	Ordinal
lesson1	0	

Tabela 12 – Resultado da preparação da amostra do conjunto de dados da  
Língua Chinesa

(continuação)

Atributo	Conversão	Tipo
NivelDaSecao	Valor	Ordinal
1	1	
2	2	
3	3	
Problema	Valor	Nominal
I01w01	0	
Passo	Valor	Ordinal
1	1	
2	2	
3	3	
4	4	
5	5	
6	6	
7	7	
8	8	
9	9	
ComponenteEntradaDeDados	Valor	Nominal
radioGroup UpdateRadioButton	0	
done ButtonPressed	1	
Atributo	Conversão	Tipo
NumeroDeRepeticoes	Valor	Quantitativo
1	1	
2	2	
3	3	
4	4	
5	5	
6	6	
7	7	

Tabela 12 – Resultado da preparação da amostra do conjunto de dados da Língua Chinesa

(conclusão)

Atributo	Conversão	Tipo
TextoRetornado	Valor	Nominal
Correct! Now click on Done to go to the next one.	0	
No Hint or Other Message Text	1	
No, this is not correct. Please click on "hint" in the upper right corner for help.	2	
Focus on the pitch change and listen to the sound again.	3	
The sound pitch is going down, i.e., it is high-falling. Listen again.	4	
This is the forth tone. Its pitch goes down. Please click the correct answer.	5	
Resposta	Valor	Ordinal
CORRECT	0	
INCORRECT	1	
HINT	2	

O atributo classificador utilizado nas visualizações desta amostra foi o Resposta.

#### 6.4.1 Scatterplot Matrix

Nesta visualização utilizou-se todos as dimensões. Exceto a dimensão Problema e a Lição do Curso, pois a análise foi desenvolvida sobre um único problema encontrado na amostra e em uma única lição do curso.

Na primeira análise da visualização, preocupou-se em descobrir informações sobre um único indivíduo, no caso, selecionou-se o segundo aluno. Os passos seguidos encontram-se nas Figuras 56, 57 e 58.

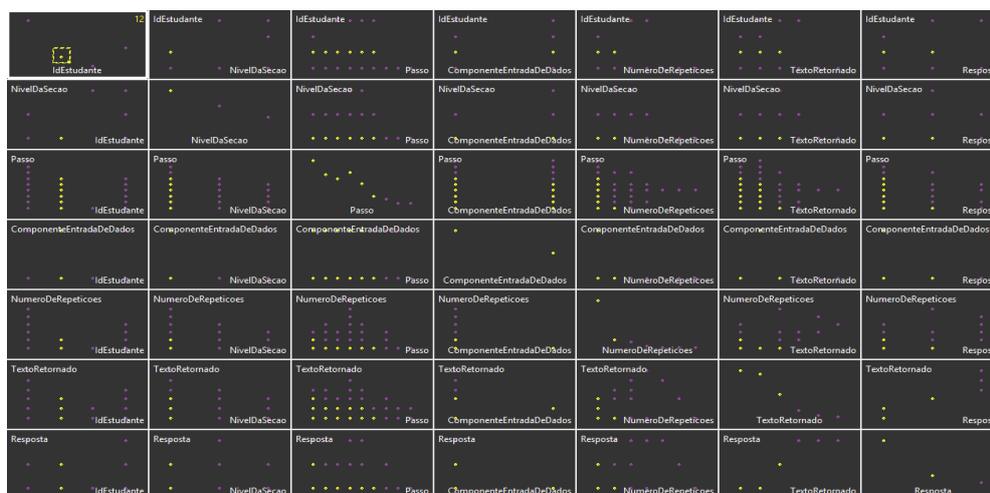
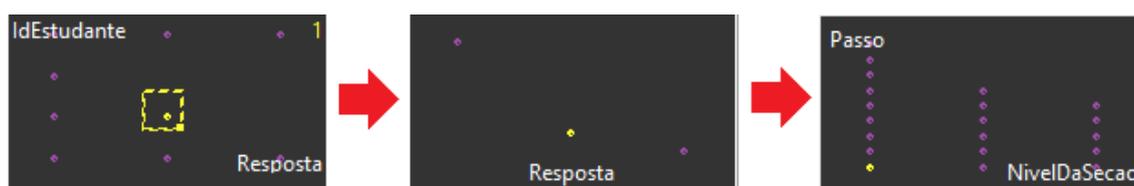
Figura 56 – *Brush* aplicado à dimensão Id Estudante

Figura 57 – Análise sobre o plano Número de Repetições por Id Estudante



Figura 58 – Análise sobre o plano Id Estudante por Resposta



Na Figura 56 utilizou-se o *brush* sobre o segundo aluno (número 1 da preparação) pode-se perceber que, ele não executou os últimos três passos, acertou e errou, mas não pediu ajuda (clicou no botão de dica) nenhuma vez. Notou-se também que ele precisou de no máximo duas repetições, bem abaixo do primeiro aluno (número 0 da preparação) que precisou de até sete repetições. Para entender melhor essa situação, foi utilizado o *brush* no plano Número de Repetições por Id Estudante (Figura 57), onde notou-se que ele só repetiu uma única vez o primeiro passo do primeiro nível da seção. Já na Figura 58 procurou-se descobrir o comportamento dos resultados desse aluno. Assim, utilizou-se o *brush* sobre o plano

Id Estudante por Resposta, onde descobriu-se que ele errara somente uma vez, e de acordo com a Figura 57, ele repetiu somente o exercício errado, acertando todos os outros sem qualquer problema.

Na segunda avaliação tentou-se descobrir algum comportamento sobre as questões respondidas incorretamente. Conforme mostram os passos retratados nas Figuras 59, 60 e 61.

Figura 59 – *Brush* aplicado à dimensão Resposta



Figura 60 – Brush aplicado ao plano Resposta por Passo



Figura 61 – Análise sobre o plano Número de Repetições por Id Estudante



Com um total de 17 respostas incorretas, conforme mostra a Figura 59, pode-se perceber com auxílio da Figura 60 que os passos 6, 8 e 9 não apresentaram erros. Já a Figura 61 mostra que dois dos quatro alunos erraram novamente após o primeiro erro.

Realizou-se uma última avaliação sobre o botão “dica” existente neste conjunto de dados conforme análise das Figuras 62 e 63.

Figura 62 – *Brush* aplicado à dimensão Resposta

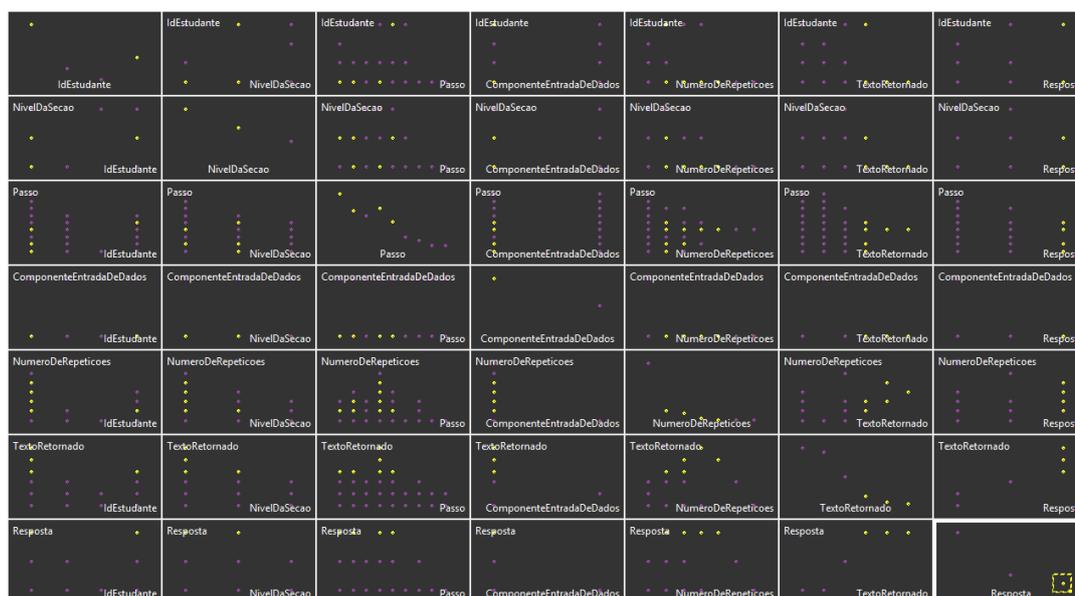
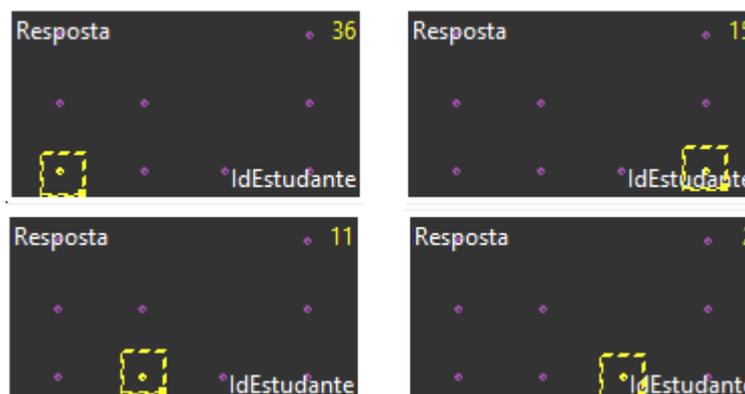


Figura 63 – Análise feita sobre quatro planos diferentes



Pode-se analisar que o botão “dica” foi pressionado oito vezes e devolveu um total de três mensagens diferentes para os alunos (Figura 62). Notou-se que, o segundo e o terceiro aluno não pediram dicas. Para o terceiro ou após o quinto

passo ninguém solicitou ajuda. Na Figura 63 buscou-se os resultados dos alunos que pediram dicas e contra os outros dois (segundo e terceiro alunos, cujos não pediram ajuda) eles apresentaram uma maior participação, bem como também um melhor desempenho.

#### 6.4.2 RadViz, RadViz3D e PolyViz

As Figuras 64 e 65 apresentam o PolyViz e o RadViz3D respectivamente, com o intuito de auxiliar o RadViz (Figura 66) na análise de informações, onde juntas apresentam de melhor forma o volume de dados com informações semelhantes.

Estas visões buscam apresentar a informação formando “grupos”. E, as ferramentas utilizadas fazem a seleção automática das cores, onde nas duas primeiras figuras o vermelho representa as respostas corretas, o verde as incorretas e o rosa a requisição de dica. Já na terceira visão a cor azul representa as respostas corretas, o cinza as incorretas e, por fim, o amarelo representa a requisição de dica.

Figura 64 – PolyViz sobre amostra de dados da Língua Chinesa

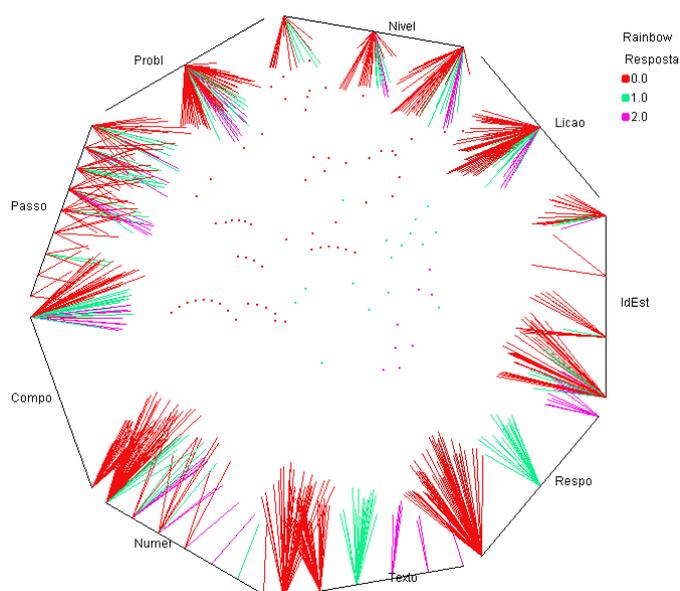


Figura 65 – RadViz3D sobre amostra de dados da Língua Chinesa

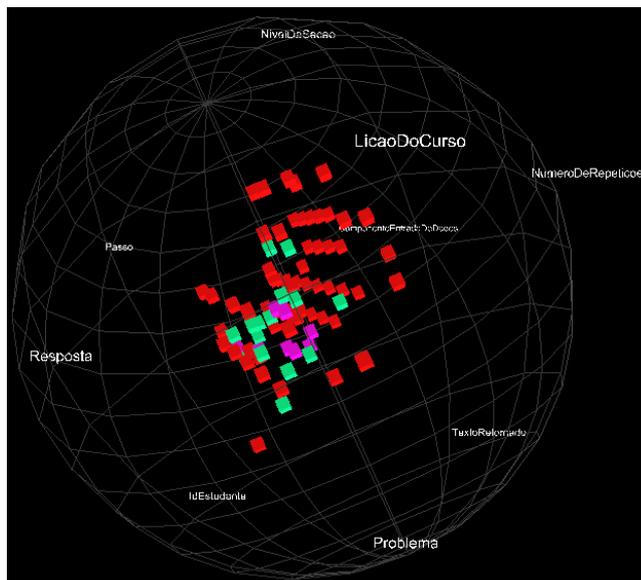
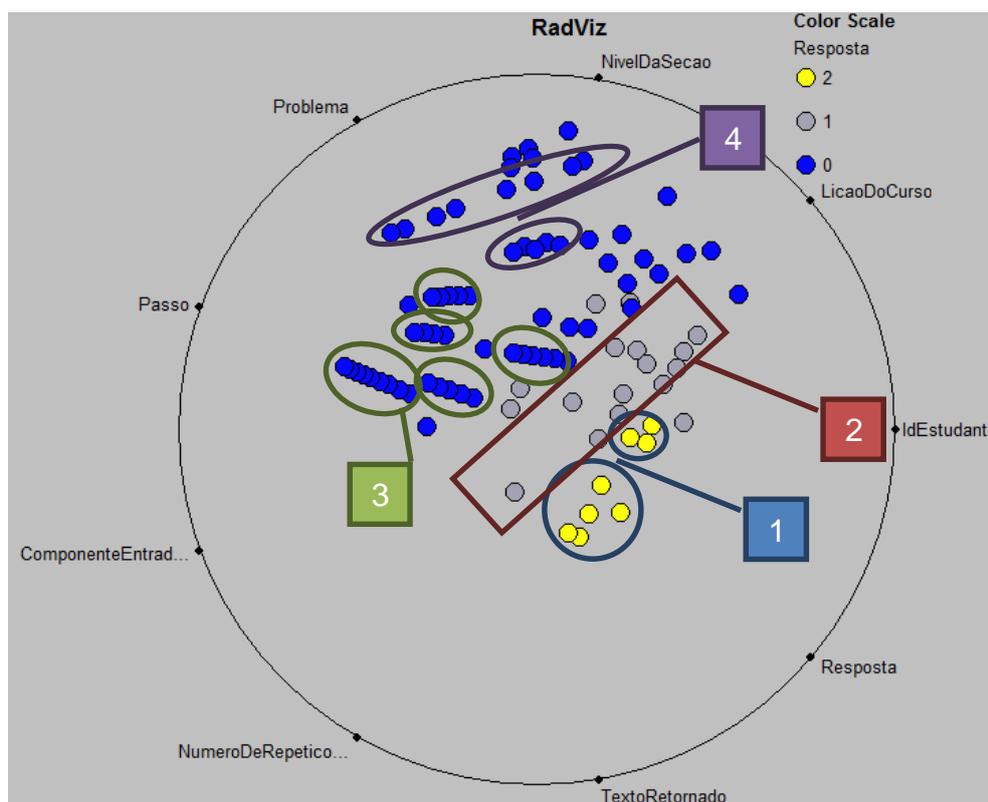


Figura 66 – RadViz sobre a amostra de dados da Língua Chinesa



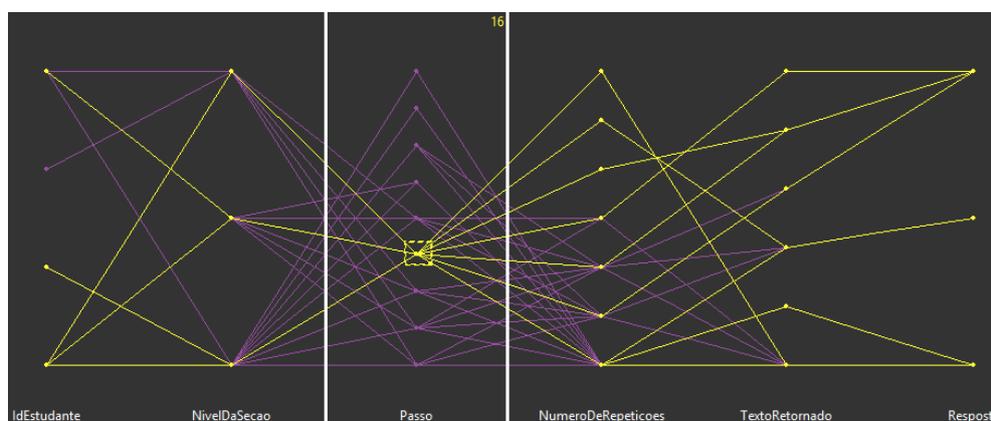
No item 1 da Figura 66 marcou-se dois grupos de Pedidos de Dicas, o primeiro com cinco elementos tem em comum o primeiro nível da seção, o segundo com três elementos tem em comum o segundo nível da seção, ou seja, o terceiro

nível não formou um grupo, concluindo-se que este nível não utilizou requisições de dicas. No item 2, onde existe uma maior concentração de erros, encontrou-se um padrão onde, somente o primeiro e o quarto aluno foram encontrados na seleção, sendo que, neste grupo a maioria das instâncias apresentou o número de repetições entre 1 e 2. Já no item 3, pode-se perceber uma sequência de pontos azuis, esses representam acertos “sequencias” dos alunos sem qualquer repetição. Pela quantidade de pontos percebe-se que até o passo 4, com exceção do terceiro aluno que tem uma quantidade baixa de instâncias, os acertos aconteceram independente da seção, ou qualquer outro fator. No item 4 acontece o mesmo, porém com um número máximo de até três repetições.

### 6.4.3 Coordenadas Paralelas

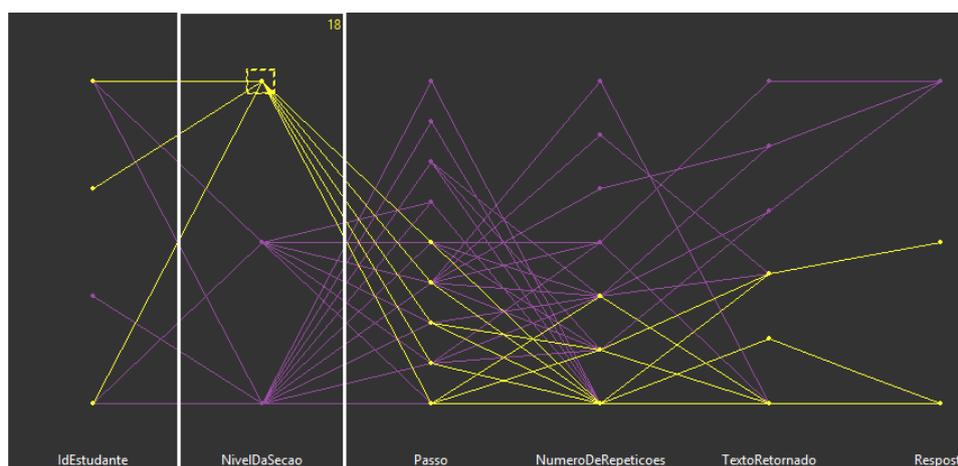
Notou-se que a dimensão Componente de Entrada de Dados não apresentou, ou não conseguiu-se extrair nenhuma informação nas visualizações antecedentes. Por este motivo esta dimensão foi desconsiderada na análise da visualização por coordenadas paralelas. As Figuras 67, 68 e 69 demonstram a análise feita com as dimensões restantes.

Figura 67 – Utilização de *brush* no quarto passo



Ao utilizar a técnica de *brush* sobre a dimensão Passo verificou-se que o passo 4, independentemente da Seção foi o que originou um maior número de repetições.

Figura 68 – *Brush* na terceira Seção



Percebeu-se ao selecionar o Nível de Seção igual a 3 que o mesmo teve um baixo número de repetições entre os alunos, e que mesmo com acertos e erros, nenhum dos alunos solicitou dicas.

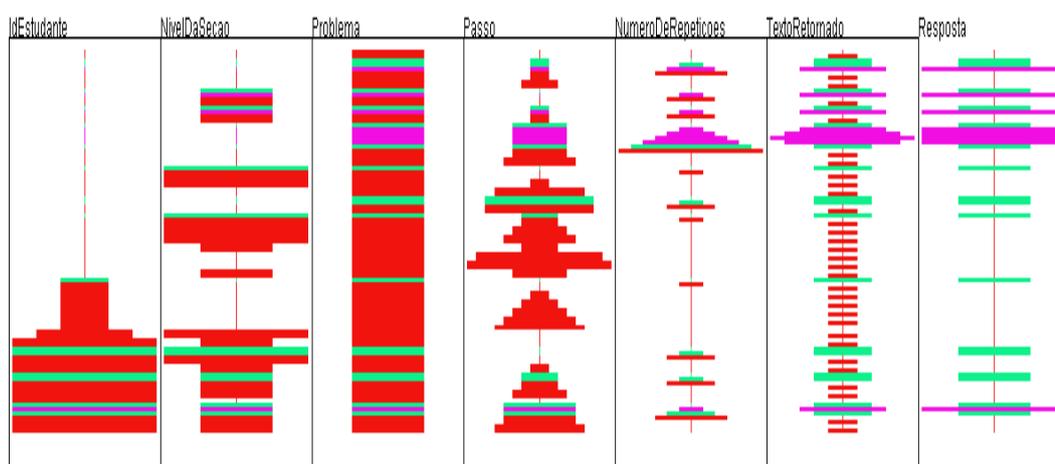
Figura 69 – Visualização da amostra utilizando Coordenadas Paralelas



A relação entre Texto Retornado e Resposta é direta, ou seja, não há respostas diferentes para um mesmo texto retornado.

#### 6.4.4 Survey Plots

Aqui, como na visão por PolyViz, as cores foram selecionadas automaticamente, sendo o vermelho as respostas corretas, o verde as incorretas e o rosa a requisição de dica.



Na visualização da amostra utilizando *Survey Plots* analisou-se a distribuição das barras como um todo. Onde percebeu-se que houve um esforço entre erros e dicas até que a questão fosse respondida com sucesso. Uma segunda distribuição na dimensão “Passo” demonstrou uma sequência de acertos, conclusão já obtida ao analisar a mesma amostra com utilizando a visualização por RadViz.

#### 6.5 ÁLBEBRA I

A fim de realizar testes sobre o conjunto de Álgebra I, foi utilizada uma amostra aleatória contendo instâncias sobre um aluno (identificado como 9YVP8j) resolvendo quatro problemas. Os problemas estão divididos em passo (tentativas utilizadas pelo aluno até chegar à resolução), sendo que o primeiro contém 21, o segundo 18, o terceiro 19 e por fim o quarto contendo 17 passos.

O *dataset* Álgebra I contém dados de 31 alunos descritos em 34 atributos. Dentre estes atributos, foram selecionados os seguintes: Id do Estudante (identificador único para cada aluno), Duração (em segundos da resolução do passo do exercício), Seção (que corresponde a divisão dos exercícios solucionados), Problema, Passo (ou etapa da resolução), Número da Tentativa e Resposta. Este *dataset* não contém respostas erradas, ou seja, o atributo Resposta contém dados como: dica inicial, mudança no nível de dicas e OK (quando a resposta do passo está correta). Há também o item *BUG* que indica erro no sistema e que não foi avaliado.

Assim como as demais amostras, a amostra do conjunto Álgebra necessitou ter os dados preparadas na aplicação Dimensionador para o uso das ferramentas propostas. Esta preparação inclui uma conversão de dados originais para dados numéricos, e a inclusão do tipo conforme apresenta a Tabela 14.

A Tabela 13 apresenta os dados selecionados para o experimento.

Tabela 13 – Amostra selecionada do conjunto Álgebra I

(continua)

IdEstudante	Duracao	Secao	Problema	Passo	NoDaTentativa	Resposta
9YVP8j	7	ES_02-5	Eg40 $-4x+9 = -1$	$-4x+9 = -1$	1	INITIAL_HINT
9YVP8j	1	ES_02-5	Eg40 $-4x+9 = -1$	$-4x+9 = -1$	2	HINT_LEVEL_CHANGE
9YVP8j	3	ES_02-5	Eg40 $-4x+9 = -1$	$-4x+9 = -1$	3	HINT_LEVEL_CHANGE
9YVP8j	7	ES_02-5	Eg40 $-4x+9 = -1$	$-4x+9 = -1$	4	OK
9YVP8j	4	ES_02-5	Eg40 $-4x+9 = -1$	$-4x+9-9 = -1-9$	1	INITIAL_HINT
9YVP8j	1	ES_02-5	Eg40 $-4x+9 = -1$	$-4x+9-9 = -1-9$	2	HINT_LEVEL_CHANGE
9YVP8j	6	ES_02-5	Eg40 $-4x+9 = -1$	$-4x+9-9 = -1-9$	3	OK
9YVP8j	4	ES_02-5	Eg40 $-4x+9 = -1$	$-4x = -1-9$	1	INITIAL_HINT
9YVP8j	1	ES_02-5	Eg40 $-4x+9 = -1$	$-4x = -1-9$	2	HINT_LEVEL_CHANGE
9YVP8j	3	ES_02-5	Eg40 $-4x+9 = -1$	$-4x = -1-9$	3	OK
9YVP8j	2	ES_02-5	Eg40 $-4x+9 = -1$	$-4x = -10$	1	INITIAL_HINT
9YVP8j	2	ES_02-5	Eg40 $-4x+9 = -1$	$-4x = -10$	2	HINT_LEVEL_CHANGE
9YVP8j	3	ES_02-5	Eg40 $-4x+9 = -1$	$-4x = -10$	3	HINT_LEVEL_CHANGE
9YVP8j	6	ES_02-5	Eg40 $-4x+9 = -1$	$-4x = -10$	4	BUG
9YVP8j	14	ES_02-5	Eg40 $-4x+9 = -1$	$-4x = -10$	5	OK
9YVP8j	11	ES_02-5	Eg40 $-4x+9 = -1$	$-4x/-4 = -10/-4$	1	INITIAL_HINT
9YVP8j	1	ES_02-5	Eg40 $-4x+9 = -1$	$-4x/-4 = -10/-4$	2	HINT_LEVEL_CHANGE
9YVP8j	9	ES_02-5	Eg40 $-4x+9 = -1$	$-4x/-4 = -10/-4$	3	OK

Tabela 13 – Amostra selecionada do conjunto Álgebra I

(continuação)

IdEstudante	Duracao	Secao	Problema	Passo	NoDaTentativa	Resposta
9YVP8j	3	ES_02-5	Eg40 $-4x+9 = -1$	$-4x/-4 = 5/2$	1	INITIAL_HINT
9YVP8j	1	ES_02-5	Eg40 $-4x+9 = -1$	$-4x/-4 = 5/2$	2	HINT_LEVEL_CHANGE
9YVP8j	4	ES_02-5	Eg40 $-4x+9 = -1$	$-4x/-4 = 5/2$	3	OK
9YVP8j	6	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-3 = 7y+(-2)$	1	INITIAL_HINT
9YVP8j	2	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-3 = 7y+(-2)$	2	HINT_LEVEL_CHANGE
9YVP8j	3	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-3 = 7y+(-2)$	3	OK
9YVP8j	44	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-3 = 7y-2$	1	INITIAL_HINT
9YVP8j	1	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-3 = 7y-2$	2	HINT_LEVEL_CHANGE
9YVP8j	4	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-3 = 7y-2$	3	HINT_LEVEL_CHANGE
9YVP8j	9	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-3 = 7y-2$	4	OK
9YVP8j	4	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-3+2 = 7y-2+2$	1	INITIAL_HINT
9YVP8j	1	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-3+2 = 7y-2+2$	2	HINT_LEVEL_CHANGE
9YVP8j	5	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-3+2 = 7y-2+2$	3	OK
9YVP8j	4	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-3+2 = 7y$	1	INITIAL_HINT
9YVP8j	1	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-3+2 = 7y$	2	HINT_LEVEL_CHANGE
9YVP8j	4	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-3+2 = 7y$	3	OK
9YVP8j	4	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-1 = 7y$	1	INITIAL_HINT
9YVP8j	1	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-1 = 7y$	2	HINT_LEVEL_CHANGE
9YVP8j	3	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-1 = 7y$	3	HINT_LEVEL_CHANGE
9YVP8j	8	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-1 = 7y$	4	OK
9YVP8j	4	ES_02-5	Eg40 $-3 = 7y+(-2)$	$-1/7 = 7y/7$	1	OK
9YVP8j	2	ES_02-5	Eg40 $2 = -7+(-6y)$	$2 = -7+(-6y)$	1	INITIAL_HINT
9YVP8j	1	ES_02-5	Eg40 $2 = -7+(-6y)$	$2 = -7+(-6y)$	2	HINT_LEVEL_CHANGE
9YVP8j	4	ES_02-5	Eg40 $2 = -7+(-6y)$	$2 = -7+(-6y)$	3	OK
9YVP8j	2	ES_02-5	Eg40 $2 = -7+(-6y)$	$2 = -7-6y$	1	INITIAL_HINT
9YVP8j	2	ES_02-5	Eg40 $2 = -7+(-6y)$	$2 = -7-6y$	2	HINT_LEVEL_CHANGE
9YVP8j	3	ES_02-5	Eg40 $2 = -7+(-6y)$	$2 = -7-6y$	3	HINT_LEVEL_CHANGE
9YVP8j	7	ES_02-5	Eg40 $2 = -7+(-6y)$	$2 = -7-6y$	4	OK
9YVP8j	1	ES_02-5	Eg40 $2 = -7+(-6y)$	$2+7 = -7-6y+7$	1	INITIAL_HINT
9YVP8j	1	ES_02-5	Eg40 $2 = -7+(-6y)$	$2+7 = -7-6y+7$	2	HINT_LEVEL_CHANGE
9YVP8j	6	ES_02-5	Eg40 $2 = -7+(-6y)$	$2+7 = -7-6y+7$	3	OK
9YVP8j	1	ES_02-5	Eg40 $2 = -7+(-6y)$	$9 = -7-6y+7$	1	OK
9YVP8j	2	ES_02-5	Eg40 $2 = -7+(-6y)$	$9 = -6y$	1	INITIAL_HINT
9YVP8j	1	ES_02-5	Eg40 $2 = -7+(-6y)$	$9 = -6y$	2	HINT_LEVEL_CHANGE
9YVP8j	3	ES_02-5	Eg40 $2 = -7+(-6y)$	$9 = -6y$	3	HINT_LEVEL_CHANGE
9YVP8j	5	ES_02-5	Eg40 $2 = -7+(-6y)$	$9 = -6y$	4	OK
9YVP8j	3	ES_02-5	Eg40 $2 = -7+(-6y)$	$9/-6 = -6y/-6$	1	INITIAL_HINT

Tabela 13 – Amostra selecionada do conjunto Álgebra I

(conclusão)

IdEstudante	Duracao	Secao	Problema	Passo	NoDaTentativa	Resposta
9YVP8j	1	ES_02-5	Eg40 2 = -7+(-6y)	9/-6 = -6y/-6	2	HINT_LEVEL_CHANGE
9YVP8j	3	ES_02-5	Eg40 2 = -7+(-6y)	9/-6 = -6y/-6	3	OK
9YVP8j	3	ES_02-5	Eg40 2 = -7+(-6y)	-3/2 = -6y/-6	1	OK
9YVP8j	7	ES_02-6	Eg40 0 = -2+2y	0 = -2+2y	1	INITIAL_HINT
9YVP8j	1	ES_02-6	Eg40 0 = -2+2y	0 = -2+2y	2	HINT_LEVEL_CHANGE
9YVP8j	3	ES_02-6	Eg40 0 = -2+2y	0 = -2+2y	3	HINT_LEVEL_CHANGE
9YVP8j	7	ES_02-6	Eg40 0 = -2+2y	0 = -2+2y	4	OK
9YVP8j	6	ES_02-6	Eg40 0 = -2+2y	0+2 = -2+2y+2	1	OK
9YVP8j	4	ES_02-6	Eg40 0 = -2+2y	2 = -2+2y+2	1	INITIAL_HINT
9YVP8j	1	ES_02-6	Eg40 0 = -2+2y	2 = -2+2y+2	2	HINT_LEVEL_CHANGE
9YVP8j	12	ES_02-6	Eg40 0 = -2+2y	2 = -2+2y+2	3	OK
9YVP8j	13	ES_02-6	Eg40 0 = -2+2y	2 = -2+2y+2	4	INITIAL_HINT
9YVP8j	3	ES_02-6	Eg40 0 = -2+2y	2 = -2+2y+2	5	HINT_LEVEL_CHANGE
9YVP8j	3	ES_02-6	Eg40 0 = -2+2y	2 = -2+2y+2	6	OK
9YVP8j	5	ES_02-6	Eg40 0 = -2+2y	y1R1	1	INITIAL_HINT
9YVP8j	5	ES_02-6	Eg40 0 = -2+2y	yR2	1	BUG
9YVP8j	14	ES_02-6	Eg40 0 = -2+2y	yR2	2	OK
9YVP8j	4	ES_02-6	Eg40 0 = -2+2y	y1R1	2	OK
9YVP8j	3	ES_02-6	Eg40 0 = -2+2y	constant terms2R1	1	OK
9YVP8j	1	ES_02-6	Eg40 0 = -2+2y	constant terms1R1	1	BUG

Tabela 14 – Resultado da preparação da amostra Álgebra I

(continua)

Atributo	Conversão	Tipo
IdEstudante	Valor	Nominal
9YVP8j	0	
Duracao	Valor	Quantitativo
7	7	
1	1	
3	3	
4	4	
6	6	
2	2	
14	14	

Tabela 14 – Resultado da preparação da amostra Álgebra I

(continuação)

Atributo	Conversão	Tipo
Duracao	Valor	Quantitativo
11	11	
9	9	
44	44	
5	5	
8	8	
12	12	
13	13	
Secao	Valor	Nominal
ES_02-5	0	
ES_02-6	1	
Problema	Valor	Nominal
Eg40 $-4x+9 = -1$	0	
Eg40 $-3 = 7y+(-2)$	1	
Eg40 $2 = -7+(-6y)$	2	
Eg40 $0 = -2+2y$	3	
Passo	Valor	Ordinal
$-4x+9 = -1$	0	
$-4x+9-9 = -1-9$	1	
$-4x = -1-9$	2	
$-4x = -10$	3	
$-4x/-4 = -10/-4$	4	
$-4x/-4 = 5/2$	5	
$-3 = 7y+(-2)$	6	
$-3 = 7y-2$	7	
$-3+2 = 7y-2+2$	8	
$-3+2 = 7y$	9	
$-1 = 7y$	10	
$-1/7 = 7y/7$	11	
$2 = -7+(-6y)$	12	
$2 = -7-6y$	13	
$2+7 = -7-6y+7$	14	
$9 = -7-6y+7$	15	
$9 = -6y$	16	

Tabela 14 – Resultado da preparação da amostra Álgebra I

(conclusão)

Atributo	Conversão	Tipo
Passo	Valor	Ordinal
$9/-6 = -6y/-6$	17	
$-3/2 = -6y/-6$	18	
$0 = -2+2y$	19	
$0+2 = -2+2y+2$	20	
$2 = -2+2y+2$	21	
$y1R1$	22	
$yR2$	23	
constant terms2R1	24	
constant terms1R1	25	
NoDaTentativa	Valor	Quantitativo
1	1	
2	2	
3	3	
4	4	
5	5	
6	6	
Resposta	Valor	Ordinal
INITIAL_HINT	0	
HINT_LEVEL_CHANGE	1	
OK	2	
BUG	3	

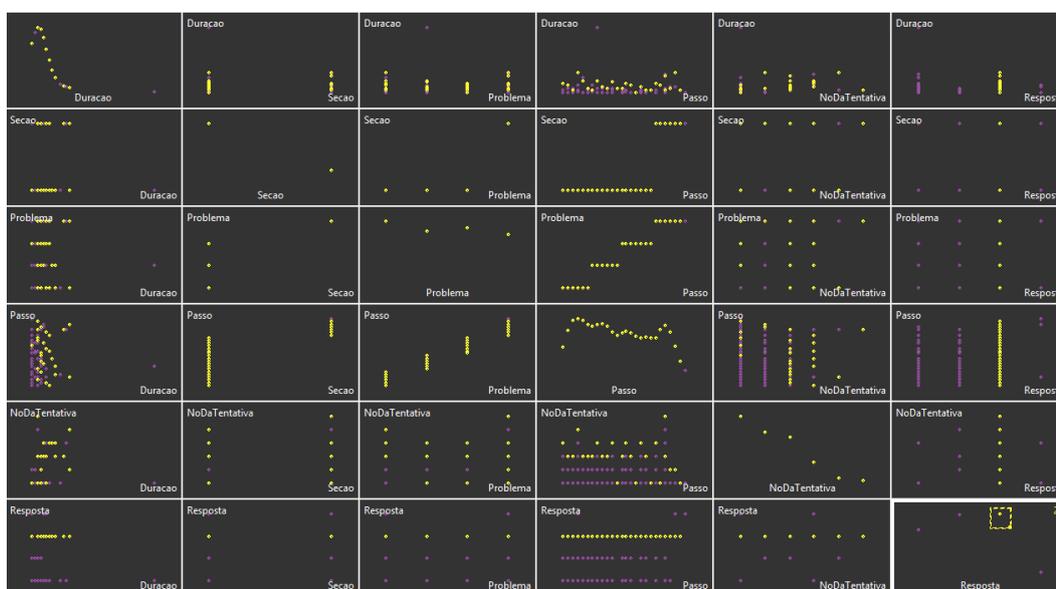
O atributo classificador utilizado nas visualizações desta amostra foi o Resposta.

### 6.5.1 Scatterplot Matrix

Dentre as dimensões selecionadas, a dimensão Id do Estudante foi omitida nesta visualização, pois a análise é desenvolvida sobre um único estudante.

A Figura 70 apresenta a visualização preliminar dos dados referentes às resoluções de problemas pelo aluno.

Figura 70 – *Brush* aplicado à dimensão Resposta



De imediato buscou-se descobrir padrões sobre os acertos deste estudante. Para isso, a distorção de *brush* foi utilizada no plano da dimensão Resposta, que informou um total de 26 acertos. Além dos acertos, notou-se que o plano Resposta por Número da Tentativa apresentou uma linha reta como pode ser visto na Figura 71.

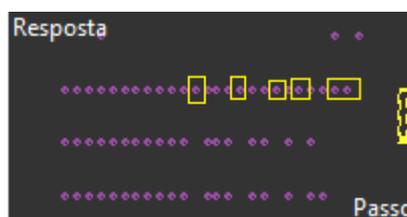
Figura 71 – Análise do plano Resposta por Número da Tentativa



Com a linha reta formada pelos pontos amarelos notou-se que o estudante respondeu corretamente independentemente da tentativa. Porém, em alguns

momentos, ele precisou realizar diversas tentativas de resolução do problema para acertar a resposta.

Figura 72 – Análise sobre o plano Resposta por Passo



Nota-se também que o estudante solicitou ajuda inicial 20 vezes. No conjunto existiam 26 respostas certas (OK). Percebeu-se então que, nas 6 vezes em que a ajuda inicial não foi solicitada, o estudante acertou a resposta. Isto sugere que o aluno teve certeza da solução destes passos.

Figura 73 – Análise sobre o plano Passo por Duração e pontos correspondentes



Na Figura 73 buscou-se no plano Duração por Passo todos os itens que estavam fora da concentração. Notou-se que os maiores tempos estão ou em pedir a primeira dica ou em responder a pergunta, porém isso não acontece nos últimos passos.

Através desta visualização é possível observar que o aluno necessita de mais tempo para a resolução dos primeiros passos do que dos passos finais. Isso pode atestar que o aluno tem dificuldades iniciais mas que graças às ajudas consegue compreender melhor o problema e resolvê-lo corretamente. De forma geral, pode-se fazer a hipótese de que houveram obstáculos epistemológicos iniciais mas o aluno foi capaz de avançar e possivelmente aprender com este exercício.

Figura 74 – Análise sobre o plano Seção por Passo e pontos correspondentes no plano Resposta



Conforme a Figura 74, nos últimos passos pediu no máximo a dica inicial ou já respondeu corretamente. Ele não precisou nenhuma vez de dicas adicionais.

### 6.5.2 RADVIZ, RADVIZ3D E POLYVIZ

As Figuras 75 e 76 apresentam o PolyViz e o RadViz3D respectivamente, com o intuito de auxiliar o RadViz (Figura 77) na análise de informações. Como já mencionado anteriormente, as imagens em conjunto possibilitam uma melhor visualização uma vez que as projeções em duas dimensões podem apresentar pontos sobrepostos. E, as cores dos pontos, atribuídas automaticamente de acordo com a dimensão Resposta são as seguintes: nas duas primeiras visões o vermelho representa o pedido de dica inicial, o verde representa a troca de nível do pedido de dica, o azul a resposta *OK* e o rosa a ocorrência de *BUG*.

Figura 75 – PolyViz sobre amostra de dados Álgebra I

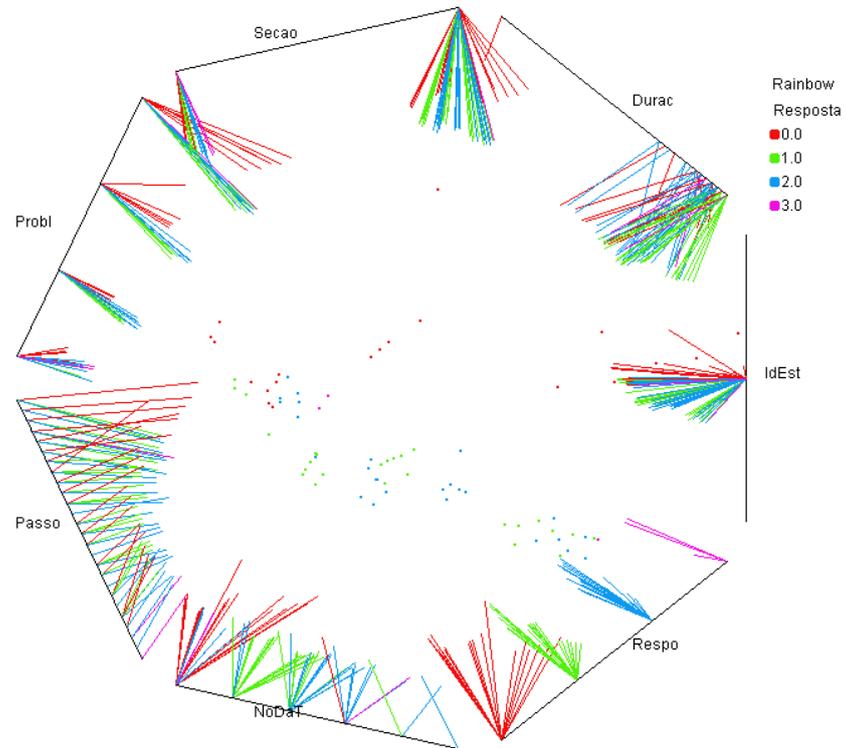


Figura 76 – RadViz3D sobre amostra de dados Álgebra I

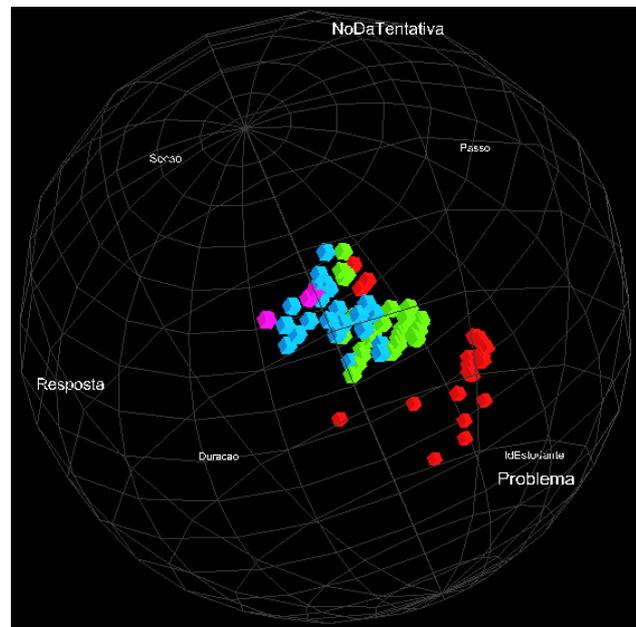
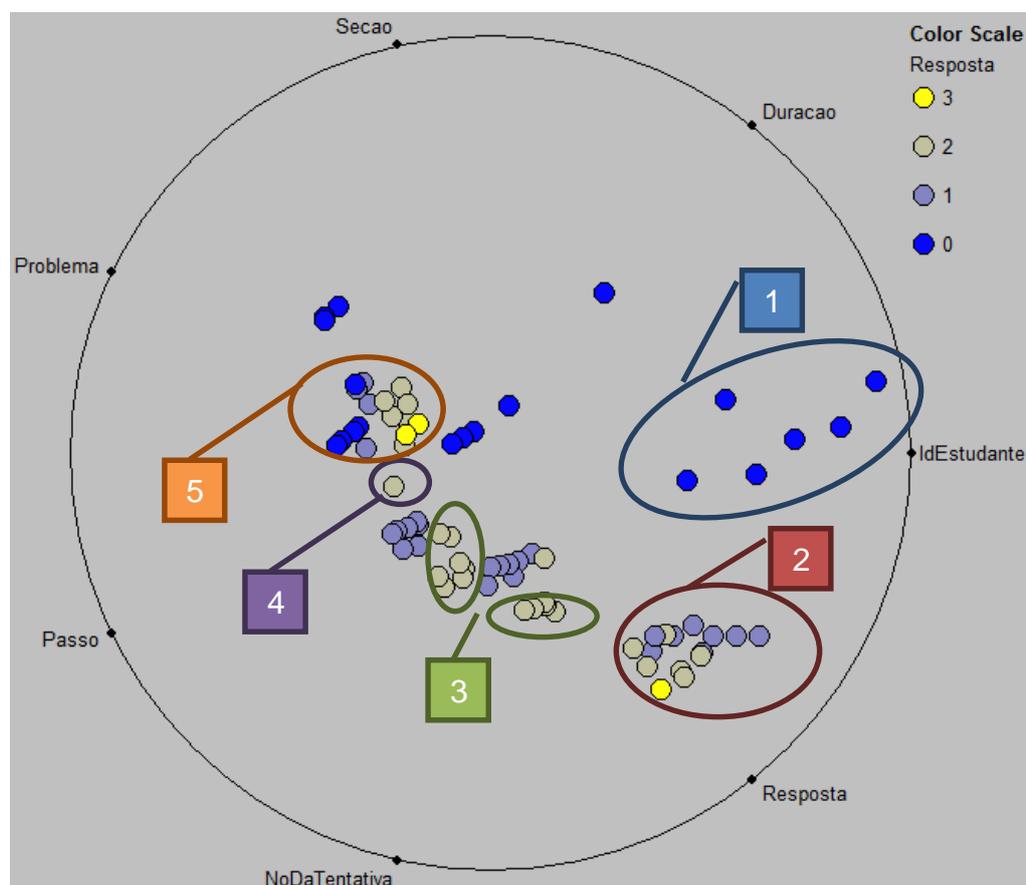


Figura 77 – RadViz sobre amostra de dados Álgebra I



O item 1 da Figura 77 destaca um grupo que apresenta um padrão. Os pontos deste grupo representam os passos (1 ao 5) do primeiro problema da primeira seção, onde o aluno pediu a dica inicial. No item 2 pôde-se encontrar a parte restante deste primeiro problema, ou seja, os pedidos de troca de nível de dica e as repostas.

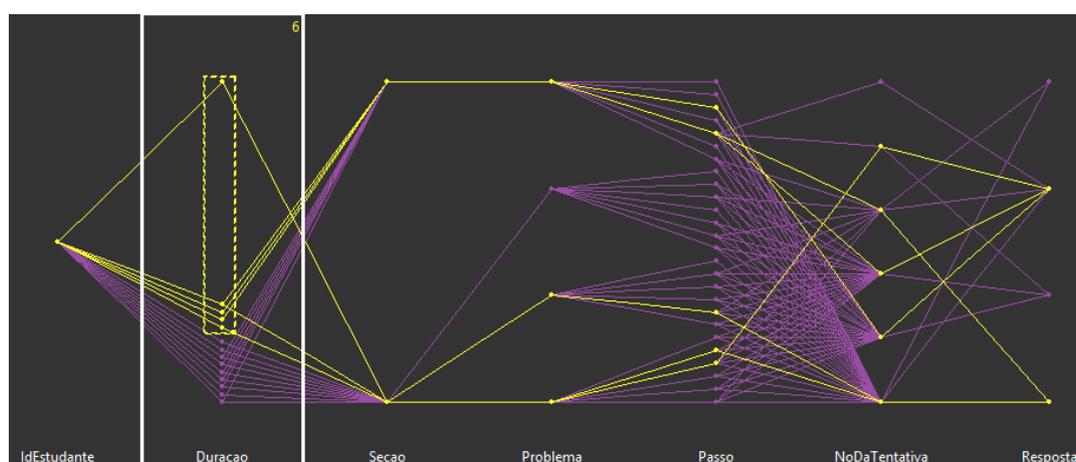
Já no item 3 percebeu-se dois grupos de acertos, estes são referentes ao segundo e ao terceiro problema da primeira seção onde o número de tentativas não passou de 4 e as durações tiveram tempos semelhantes.

Por apresentar uma distância maior dos outros pontos, marcou-se os pontos do item 4. Onde descobriu-se que ele é o único acerto que aconteceu na única vez em que foi necessária uma sexta tentativa, isso aconteceu no único problema da segunda seção. Já o item 5 apresentou uma resolução dos passos com uma ordem natural, ou seja, o estudante pediu a ajuda inicial, algumas vezes a trocou de nível e concluiu o exercício.

### 6.5.3 Coordenadas Paralelas

Seguindo o modelo dos com coordenadas paralelas encontradas nas seções anteriores, o resultados para a amostra de dados de Álgebra encontram-se nas Figuras 78, 79 e 80.

Figura 78 – *Brush* sobre a dimensão Duração



Na Figura 78 selecionou-se durações que apresentaram um intervalo maior, e como no teste por *Scatterplot Matrix*, estas são as instâncias de maior duração, onde os alunos apresentaram somente duas “Respostas”, ou escolheram somente a dica inicial ou responderam de prontidão. Pode-se notar também que isso não ocorreu no terceiro problema da primeira seção e, nos problemas que ocorreram, sempre foi num passo intermediário, nunca num inicial ou final.

Figura 79 – *Brush* na dimensão Número da Tentativa

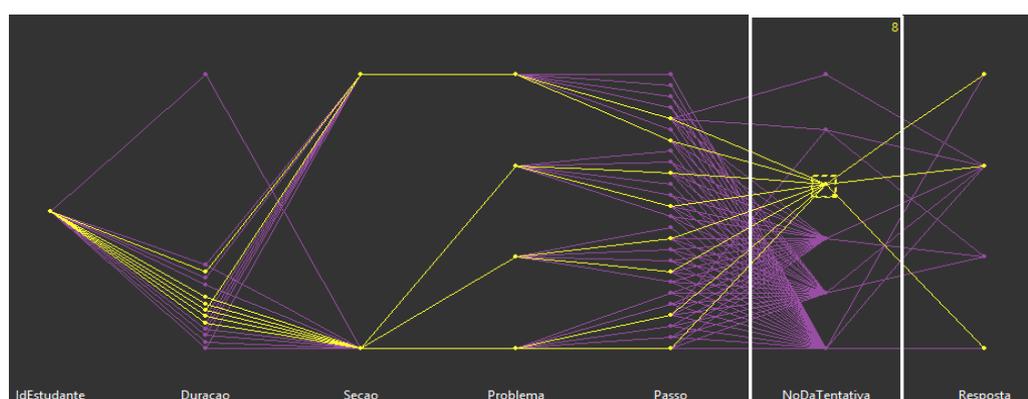
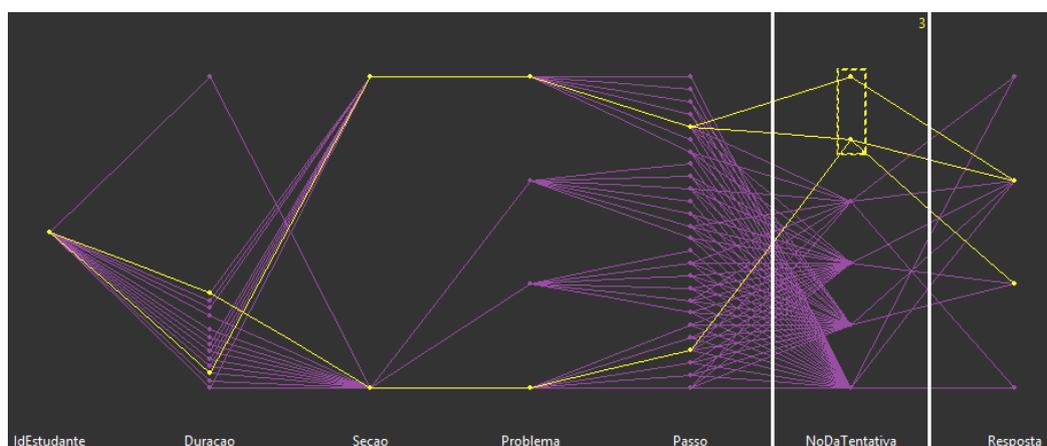
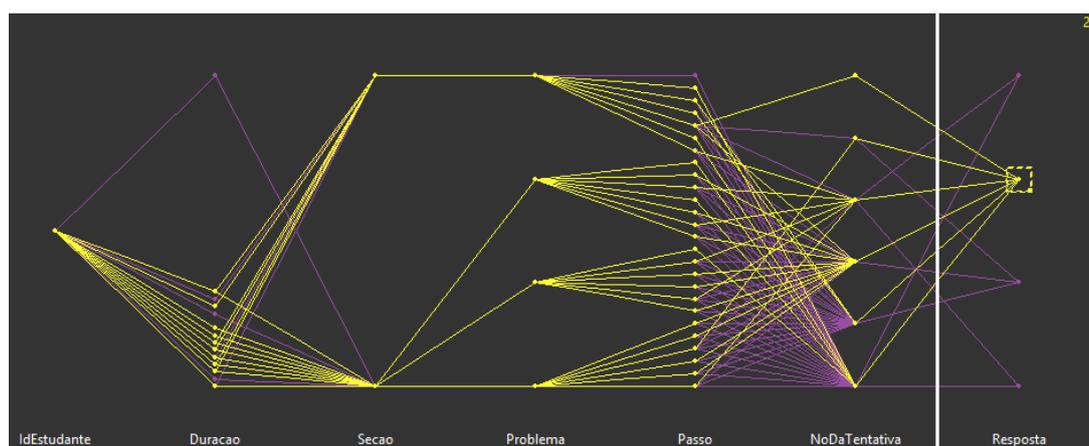
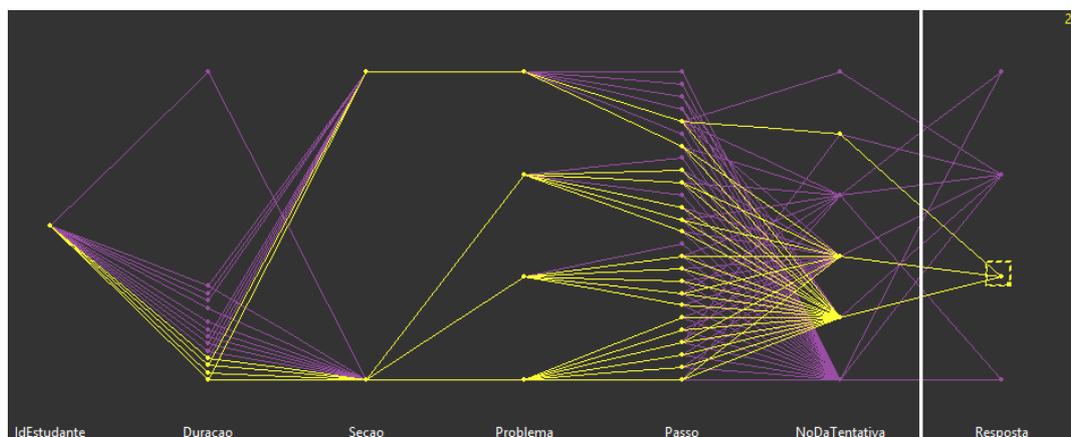


Figura 80 – *Brush* em outros pontos da dimensão Número da Tentativa

Buscando-se o número de tentativas como mostra a Figura 79, descobriu-se que a quarta tentativa foi a última a ter um maior número de instâncias, apresentando um total de oito. Verificando-se os pontos da quinta e a sexta tentativa, a distorção por *brush* foi utilizada para circular as linhas que representam as duas tentativas como mostra a Figura 80. E, pode-se notar que além das duas últimas tentativas apresentarem um total de 3 instâncias, logo, somente duas respostas finais foram dadas.

Figura 81 – *Brush* na dimensão Resposta

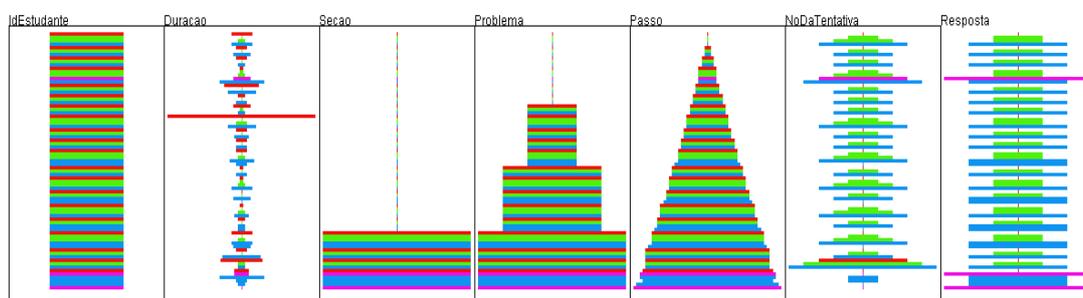
Utilizando-se do *brush* sobre as instâncias que representam as respostas OK pode-se notar que o último passo do último exercício não foi respondido.

Figura 82 – Distorção por *brush* para trocas de níveis de dicas

Selecionando as instâncias que representam as trocas dos níveis de dicas, percebeu-se que nos passos do último problema as trocas não aconteceram com a muita frequência, situação diferente dos demais.

#### 6.5.4 Survey Plots

A ferramenta não apresenta as legendas das cores, porém, utilizando outras funcionalidades deduziu-se que, a cor vermelha representa o pedido de dica inicial, o verde representa a troca de nível de dica, o azul as repostas *OK*, e a cor rosa corresponde as instâncias com a resposta *BUG*.



No geral encontram-se alguns padrões na distribuição tanto das formas como das cores. A dimensão Passo, por exemplo, praticamente forma um triângulo demonstrando que houve uma distribuição quase uniforme no número de tentativas,

com leve exceção nos últimos passos que possuem um formato um pouco mais retangular. Já nas cores pode-se perceber que sempre, ou quase sempre, há um pequeno retângulo vermelho no início de cada passo até um azul, representando bem a curva de dica inicial e resposta.

## 6.6 TESTES BASEADOS EM CENÁRIOS SIMPLIFICADOS

A fim de avaliar as ferramentas e visões propostas, uma entrevista foi formulada. Esta entrevista compreendeu a visualização de três cenários bem específicos. Observa-se que as visões que acompanharam as entrevistas foram produzidas nas ferramentas do estudo, o que possibilitava ao entrevistado a interação.

Os três cenários foram apresentados a três professores de áreas distintas do ensino. O primeiro é professor da área de redes de computador, o segundo professor da área de banco de dados e o terceiro professor de negócios com ênfase em B.I. (*Business Intelligence*).

Para cada cenário formulado além da simulação da visão em uma ferramenta, uma análise parcial – por escrito – também era fornecida. Esta análise tinha como objetivo focar o entrevistado a uma situação específica da amostra, diminuindo o escopo de todas as possíveis análises a somente uma. Assim, conseguiu-se observar se a análise que acompanhava os cenários poderia ser percebida nas visões pelo entrevistado.

Os cenários, entrevistas e conclusões são descritos nas seções seguintes.

### 6.6.1 Primeiro cenário

Para o primeiro cenário formulou-se um documento contendo uma visão utilizando a *Scatterplot Matrix* sobre a amostra de dados Geometria (Figura 83). Além da visão, a análise – por escrito – a seguir também constava no documento.

“Os estudantes resolveram as etapas com durações próximas, com exceção do aluno 1, o qual em uma das etapas levou mais tempo para responder e mesmo assim a respondeu corretamente.”

Figura 83 – Visão simulada no primeiro cenário



Formulado o cenário, passou-se a fase de entrevista. A seguir descreve-se as percepções obtidas, uma para cada professor.

O primeiro entrevistado interagiu com a *Scatterplot Matrix* – simulada conforme proposto, onde a visão e a ferramenta foram demonstradas. Após apresentou-se a análise preliminar. Aqui notou-se que após explicações de pontos específicos da visão e dos dados contidos na amostra, o profissional conseguiu entender. Porém, o entrevistado diante da visão – a qual continha todos os gráficos x e y (plano cartesiano entre todas as dimensões da amostra) – demonstrou uma certa dificuldade de focar somente no que era demonstrado na análise preliminar. Aqui duas observações podem ser feitas: a primeira é que as visões em sua essência são artifícios que especialistas em visões devem usar em dados os quais também são especialistas ou tenham um bom conhecimento. Logo é de se esperar que um primeiro contato com explicações específicas possam causar certas resistências no entendimento. A segunda observação é que, o desejo de interagir com a ferramenta e buscar informações ficou evidente.

Para o segundo profissional, buscou-se apresentar a visão e os dados antes de fazer qualquer pergunta ou apresentar a análise preliminar. O entrevistado interagiu alguns minutos com a ferramenta, uma espécie de adaptação, e após foi apresentada a análise preliminar. Semelhante ao primeiro, o segundo pediu alguns esclarecimentos – principalmente sobre os significados dos dados contidos na amostra –, mas foi capaz de compreender a análise. Cita-se aqui, um desconforto do entrevistado na visão simulada quanto à quantidade de gráficos gerados.

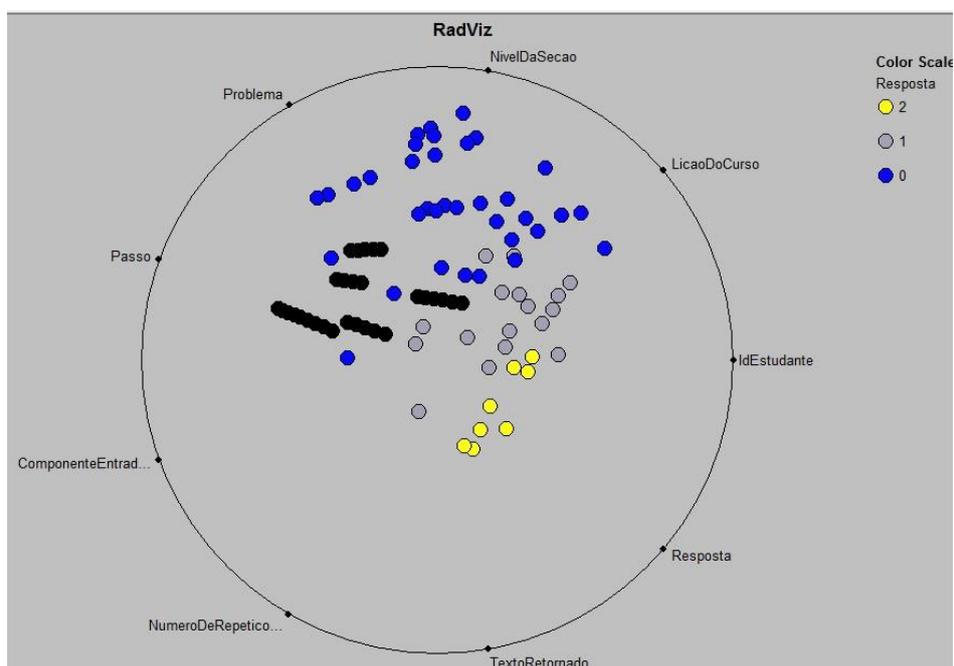
Por último, o terceiro entrevistado, além de ser um professor, possuía a qualificação de analista de BI (*Business Intelligence*). Seguindo um procedimento parecido com o segundo, a visão e os dados foram apresentados anteriormente à apresentação da análise preliminar. Este profissional demonstrou uma facilidade em interagir com a ferramenta apresentada. Fez alguns questionamentos pontuais. O primeiro a respeito da dispersão dos dados nas ferramentas, o segundo a respeito de como localizar na amostra um dado desenhado na visão. Após, foi apresentada a análise observada. O entrevistado – com o cenário reproduzido na ferramenta – conseguiu sem maiores dificuldades compreender o que lhe foi apresentado.

### **6.6.2 Segundo cenário**

Para o segundo cenário formulou-se um documento mantendo o mesmo formato encontrado no primeiro cenário. Nele, uma visão utilizando o RadViz sobre a amostra de dados da Língua Chinesa (Figura 84). Além da visão, a análise seguinte – por escrito – constava no documento.

“A sequências de pontos azuis (que ao selecionar ficam pretos) representam acertos sequenciais dos alunos sem qualquer repetição. Pela quantidade de pontos percebe-se que até o passo 4 (consultar dados), com exceção do terceiro aluno que tem uma quantidade baixa de instâncias, os acertos aconteceram independente da seção, ou qualquer outro fator.”

Figura 84 – Visão simulada no segundo cenário



As observações obtidas são descritas a seguir.

O cenário não foi demonstrado ao primeiro entrevistado por limitações de tempo.

Para o segundo profissional, utilizou-se o mesmo formato de apresentação, primeiro a visão e os dados antes de apresentar a análise preliminar. O entrevistado notou a formação de grupos (característica da visão), interagiu e questionou sobre como descobrir quais instâncias correspondem aos pontos selecionados ou plotados na visão. Após, observou os dados contidos na amostra (arquivo no formato de planilha) e procurou entender a relação entre ponto e instância. Assim, com o entendimento da técnica e entendida a amostra, foi apresentada a análise preliminar. A qual foi entendida sem dificuldades. Notou-se também que após o entendimento da análise, o entrevistado – por conta própria – começou a marcar outros grupos formados pela ferramenta e verificar quais instâncias eles representavam, entende-se que o mesmo já estava tentando fazer análises ou descobertas por si só.

Por último, realizou-se a entrevista para o terceiro entrevistado. Como no primeiro cenário, ele com poucas explicações pôde utilizar sem maiores dificuldades a ferramenta e a visão. Antes que a análise fosse apresentada, o entrevistado já

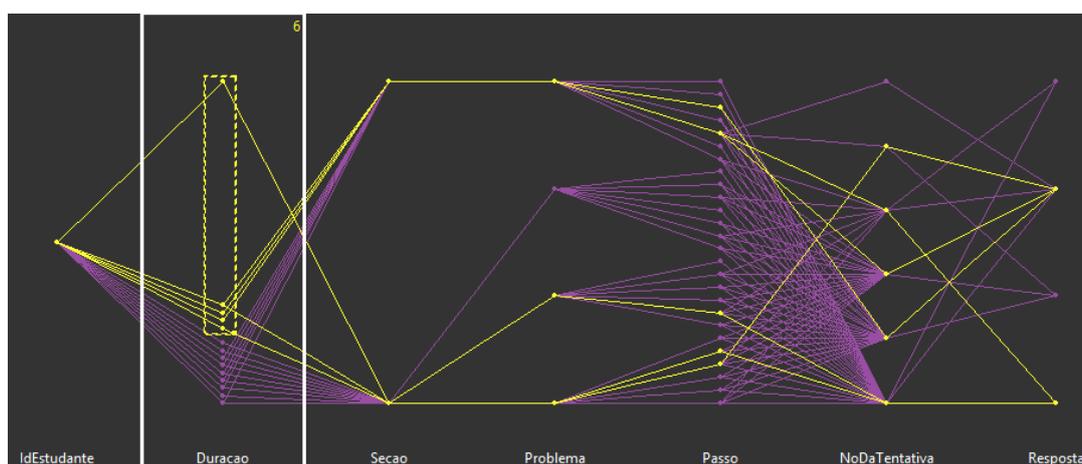
estava interagindo com as ferramentas e com os dados. Nota-se que o conceito de formação de grupos proposto pela visão ficou evidente para este entrevistado. Após, então, apresentou-se a análise, a qual não apresentou nenhuma dificuldade a ser entendida pelo entrevistado.

### 6.6.3 Terceiro cenário

No terceiro cenário formulou-se um documento contendo uma visão utilizando as Coordenadas Paralelas sobre a amostra de dados Álgebra. Além da visão, a análise seguinte – por escrito – constava no documento.

“Ao selecionar as durações de maior valor, percebe-se que os alunos apresentaram somente duas Respostas, ou escolheram somente a dica inicial ou responderam de prontidão. Pode-se notar também que isso não ocorreu no terceiro problema da primeira seção e, nos problemas que ocorreram, sempre foi num passo intermediário, nunca num inicial ou final.”

Figura 85 – Visão simulada no terceiro cenário



As observações obtidas são descritas a seguir.

O cenário não foi demonstrado ao primeiro entrevistado por limitações de tempo.

Os entrevistados (segundo e terceiro) demonstraram um entendimento comum ao utilizar as coordenadas paralelas. Ambos, que agora interagem com a

terceira visão, não apresentavam muita dificuldade em compreender as visões e as amostras de dados, mesmo o cenário sendo distinto dos outros dois. Aqui, onde ambos já conheciam o *brush* – seleção amarela que destaca a informação – o utilizavam de forma quase natural, ou seja, sabiam como aumentar o tamanho da caixa de seleção, sabiam que podiam destacar itens na visão e verificar na amostra quais instâncias representavam os itens destacados. Além disso, as coordenadas paralelas por se assemelharem a gráficos x e y, fez com que os entrevistados apresentassem um maior conforto no entendimento, uma vez que a visão somente relaciona as dimensões (atributos) na forma horizontal.

## 7 CONCLUSÃO

Este capítulo destina-se a apresentação de uma síntese das atividades desenvolvidas durante o desenvolvimento deste trabalho, descreve também os resultados analisados e ideias para trabalhos futuros.

### 7.1 SÍNTESE DO TRABALHO

A Visualização de Informação destina-se a reforçar a cognição humana utilizando-se de representações – interativas – visuais, apresentando uma interface gráfica capaz de manipular um grande número de atributos, também chamados neste estudo de dimensões. Possibilitando que pesquisadores façam descobertas, decisões ou esclarecimentos. Estes sobre padrões, grupos de itens, ou, itens individuais.

O principal objetivo deste trabalho foi pesquisar ferramentas e visões afim de atingir um processo capaz de avaliar dados como os da educação. Além do processo já estabelecido buscou-se também analisar dados em três áreas distintas.

Foram testadas ferramentas de visualização, tendo sido escolhidas as seguintes: GGobi para a criação das *Scatterplot Matrix* e Coordenadas Paralelas e as Aplicações de Hoffman (MIVAC e RadVizX) para a criação do RadViz (além do RadViz3D e do PolyViz) e *Survey Plots*.

Foi necessário desenvolver uma aplicação capaz de compatibilizar os dados selecionados de um *dataset* para as ferramentas de visualização selecionadas. Isso possibilitou que os *dataset* pudessem ser analisados nas ferramentas propostas.

Com base no conhecimento adquirido em Visualização de Informação, o processo, ferramentas pesquisadas e a ferramenta desenvolvida. Foi possível formar um pequeno ecossistema consistido das fases de: transformação dos dados brutos em dados estruturados, reorganização dos dados estruturados, conversão destes dados para glifos e a visão em si.

Obtidas as análises, três cenários distintos foram desenvolvidos. Estes cenários continham uma visão e uma análise preliminar cada. Eles foram utilizados na formulação de um entrevista, a qual foi aplicada a três professores.

## 7.2 RESULTADOS E CONTRIBUIÇÕES

Para possuir o perfil de cientista da computação, espera-se que um profissional consiga lidar com novas áreas de estudo de forma a compreendê-las. Fazendo assim que sua aplicação seja eficiente e proveitosa na resolução dos mais diversos problemas computacionais. Espera-se que o mesmo desenvolva cada vez mais sua capacidade intelectual e de pesquisa garantindo, desta maneira, a criação de novas metodologias e a evolução das já existentes a fim de suprir as necessidades que os usuários possuem nas mais diversas áreas de atuação.

Na formulação deste trabalho, foi proposta uma pesquisa ampla, que buscou encontrar diferentes maneiras de como visualizar informações. A Visualização de Informação é um problema humano-computacional que pode ou não prover uma interface capaz de auxiliar na descoberta de novas informações, mas é dependente do analista que a usar. Para que a ideia se concretizasse houve a necessidade de estender os estudos para áreas que transcendem às abordadas no currículo do curso de graduação.

No decorrer das análises realizadas, foi possível identificar que o uso das ferramentas e visões produziram conclusões coerentes com as amostras selecionadas, bem como forneceram um meio que possibilitasse a descoberta de novas. Aqui cita-se, por exemplo, a conclusão obtida da amostra do conjunto Álgebra I, onde foi possível perceber que, alunos tiveram obstáculos epistemológicos, mas que ao continuar respondendo o problema em questão, pode-se notar um possível aprendizado, demonstrando que o processo ao ser utilizado por um especialista, é capaz de prover um caminho capaz de analisar dados educacionais. Sendo válido afirmar, também, que o processo como citado pode ser proveitoso nos mais diversos conjuntos de dados, contribuindo para a análise e descoberta de informações que vão além do ensino.

Notou-se também uma distinção na usabilidade das visões. A *Scatterplot Matrix* e as Coordenadas Paralelas apresentaram uma maior eficiência na análise de instâncias do *dataset* (podendo ser etapas ou indivíduos). Para a percepção de grupo é necessário que o usuário faça a seleção adequada das instâncias. Já o RadViz e o *Survey Plots* apresentam um conceito que visa a criação de grupos, ou, a percepção dos mesmos. Elas ressaltam, de acordo com a sua concepção, indivíduos semelhantes.

Observou-se que, as visões utilizadas são capazes de representar os conjuntos escolhidos, sendo a melhor para representar o todo, a visão RadViz. Pois ela é capaz de relacionar todas as dimensões para todos os indivíduos de uma só vez, agrupando os que contenham informações semelhantes.

Porém, ao observar do ponto de vista dos professores, que buscam avaliar o desempenho. Afirma-se que, tanto para dados de alunos resolvendo um único problema, como para um aluno resolvendo diversos problemas a visão por Coordenadas Paralelas é a melhor para o acompanhamento do professor. Afirma-se também que as Coordenadas Paralelas: a) facilitaram a compreensão da relação entre as dimensões, ou seja, pode-se reunir uma série de fatores (dimensões) como, acerto, número de passos e tentativas e com isto verificar se existe um comportamento distinto ou que se ressalve dentro destas três dimensões. b) proveram uma interface de fácil manuseio. As Coordenadas Paralelas por serem semelhantes a um gráfica x-y facilita a compreensão inclusive no primeiro contato. Seu uso também não oferece dificuldades por apresentar as dimensões e desenhá-las de forma clara. c) As entrevistas demonstraram que os professores tiveram uma maior facilidade tanto na compreensão da análise preliminar como na utilização da visão.

As análises realizadas no formato de entrevista – que foi realizada utilizando três professores - também permitiu concluir que a utilização do processo proposto e dos artefatos utilizados é viável e possibilita tanto a percepção de padrões nos dados bem como a descoberta de novas informações.

Conclui-se, portanto, que o presente trabalho atingiu os objetivos especificados e deixou importantes contribuições para a Visualização de Informação, tais como o desenvolvimento de uma ferramenta integradora entre as aplicações responsáveis por apresentar as visões, as ferramentas de visualização e também

uma forma de analisar e descobrir informações. Onde novos conhecimentos foram adquiridos e agregados aos já obtidos, contribuindo, desta forma, para o enriquecimento de mais uma área da Ciência da Computação.

### 7.3 TRABALHOS FUTUROS

O presente trabalho abordou uma maneira de como utilizar as ferramentas de visualização bem como demonstrar uma possível análise das informações contidas nos conjuntos de dados.

Como sugestão para trabalhos futuros, a criação de uma ferramenta capaz de suprir mais etapas do processo de visualização, onde as pesquisadas neste trabalho não ofereceram o tratamento dos dados estruturados.

## REFERÊNCIAS

AHLBERG, C. et al. **Dynamic queries for information exploration: An implementation and evaluation**. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 1992. p. 619-626.

ALPERN, Bowen; CARTER, Larry. **The hyperbox**. In: Visualization, 1991. Visualization'91, Proceedings., IEEE Conference on. IEEE, 1991. p. 133-139, 418.

ANDERSON, John R.; BOYLE, C. Franklin; YOST, Gregg. **The Geometry Tutor**. In: IJCAI. 1985. p. 1-7.

ANDREWS, David F. Plots of high-dimensional data. **Biometrics**, p. 125-136, 1972.

ANKERST, Mihael; KEIM, Daniel A.; KRIEGEL, Hans-Peter. **'Circle Segments': A Technique for Visually Exploring Large Multidimensional Data Sets**. Dim, v. 1501, p. 1, 2011.

ANUPAM, V. et al. **Research report. DataSpace: 3-D visualizations of large databases**. In: Information Visualization, 1995. Proceedings. IEEE, 1995. p. 82-88.

BECKER, Richard A. et al. **Visualizing network data**. Visualization and Computer Graphics, IEEE Transactions on, v. 1, n. 1, p. 16-28, 1995.

BEDDOW, Jeff. **Shape coding of multidimensional data on a microcomputer display**. In: Proceedings of the 1st conference on Visualization'90. IEEE Computer Society Press, 1990. p. 238-246.

BESHERS, Clifford; FEINER, Steven. n-Vision and AutoVisual. Disponível em: <<http://monet.cs.columbia.edu/projects/AutoVisual/AutoVisual.html>>. Acesso em 02/06/2013.

BÖRNER, Katy; ZHOU, Yuezheng. **A Software Repository for Education and Research in Information Visualization**. In: Proceedings of the Fifth International Conference on Information Visualization, London, UK, p. 257-262, jul. 2001.

BUJA, Andreas et al. **Interactive data visualization using focusing and linking**. In: Visualization, 1991. Proceedings of IEEE Conference. IEEE, 1991. p. 156-163, 419.

CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. **Information Visualization. Readings in Information Visualization: Using Vision to Think**, Morgan Kaufmann Publishers, p. 1-34, San Francisco, California, USA, 1999.

CAVA, R. A. **Bifocal Tree: Uma Técnica para Visualização de Estruturas Hierárquicas**. Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, RS, Brasil, 2002.

CEN, H., KOEDINGER K., JUNKER B. **Learning Factors Analysis A General Method of Cognitive Model Evaluation and Improvement**. The 8th International Conference on Intelligent Tutoring Systems. 2006.

CHERNOFF, Herman. **The use of faces to represent points in k-dimensional space graphically**. Journal of the American Statistical Association, v. 68, n. 342, p. 361-368, 1973.

CLEVELAND, William S. **Visualizing data**. Hobart Press, 1993.

CONSENS, Mariano; MENDELZON, Alberto. **Hy+: A hygraph-based query and visualization system**. In: ACM SIGMOD Record. ACM, 1993. p. 511-516.

DUVAL, Patrick et al. **Empowering Learning Objects: an experiment with the Ganesha platform**. In: World Conference on Educational Multimedia, Hypermedia and Telecommunications. p. 4592-4599, 2005.

ECKSTEIN, Robert; LOY, Marc; WOOD, Dave. **Java swing**. O'Reilly & Associates, Inc., 1998.

FAYYAD, Usama M.; WIERSE, Andreas; GRINSTEIN, Georges G. **Information visualization in data mining and knowledge discovery**. Morgan Kaufmann Pub, 2002.

FEINER, S. K.; BESHERS, Clifford. **Visualizing n-dimensional virtual worlds with n-vision**. In: ACM SIGGRAPH Computer Graphics. ACM, 1990. p. 37-38.

FEKETE, Jean-Daniel. **The InfoVis Toolkit**. Disponível em <<http://ivtk.sourceforge.net/>>, acesso em 16/06/2013.

FOWLER, Amy. **A Swing architecture overview**. The Swing Connection Website, 1998.

FREITAS, C. M. D. S.; et al. **Data usability Issues in Information Visualization Applications**. In: WORKSHOP ON SPATIAL DATA USABILITY, v. 1, p. 1–1, Wageningen, Holanda, 2001. Center for Geo-Information.

FREITAS, Carla Maria Dal Sasso ; WAGNER, Flavio Rech . **Ferramentas de suporte às tarefas da análise exploratória visual**. RITA, v. 2, n. 1, p. 5-36, 1995.

FURNAS, George W.; BUJA, Andreas. **Prosection views: Dimensional inference through sections and projections**. Journal of Computational and Graphical Statistics, v. 3, n. 4, p. 323-353, 1994.

GEOVISTA, Penn State Geography. **GeoViz Toolkit**. Disponível em <<http://www.geovista.psu.edu/geoviztoolkit/>>, acesso em 24/06/2013.

GGOBi. **GGobi - Out of sight, out of mind**. Disponível em <<http://www.ggobi.org/>>, acesso em 16/06/2013.

GONDIM, R. P.; MACHADO, M. C. C.; SANTOS, M. S. **VIDEWeb: Um Ambiente para Visualização de Dados Educacionais na Plataforma WebPIDE**. Projeto de Graduação, Departamento de Computação e Estatística da Universidade Federal de Mato Grosso do Sul, UFMS, Campo Grande, MS, Brasil, 2008.

GRINSTEIN, Georges; PICKETT, Ronald; WILLIAMS, Marian G. **Exvis: An exploratory visualization environment**. In: Graphics Interface. 1989. p. 254-261.

GRINSTEIN, Georges; TRUTSCHL, Marjan; CVEK, Urrka. **High-dimensional visualizations**. In: Proceedings of Workshop on Visual Data Mining, ACM Conference on Knowledge Discovery and Data Mining. 2001. p. 1-14.

GUERAUD, Viviane et al. **Suivi à distance de classe virtuelle active**. In: Technologies de l'Information et de la Connaissance dans l'Enseignement Supérieur et de l'Industrie. p. 377-383, 2004.

HAUSER, Helwig; LEDERMANN, Florian; DOLEISCH, Helmut. **Angular brushing of extended parallel coordinates**. In: Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on. IEEE, 2002. p. 127-130.

HOFFMAN, P. E. **Table Visualization: A formal Model and Its Applications**. 1999. Doctoral dissertation, Computer Science Dept., University of Massachusetts at Lowell.

HOFFMAN, P. E.; GRINSTEIN, G. G.; MARX K.; GROSSE I.; STANLEY E. **DNA Visual and Analytic Data Mining**. In: IEEE Symposium on Information Visualization, INFOVIS, 1997. Proceedings... [S.I.]: IEEE, 1997. p. 437-441.

HOFFMAN, Patrick et al. **DNA visual and analytic data mining**. In: Proceedings of Visualization'97, IEEE, 1997. p. 437-441.

HOFFMAN, Patrick. **Patrick E. Hoffman Home Page**. Disponível em <<http://www.cs.uml.edu/~phoffman/>>, acesso em 16/06/2013.

HUBER, Peter J. **Projection pursuit**. The annals of Statistics, p. 435-475, 1985.

INSELBERG, Alfred. **Lecture Notes: Parallel Coordinates**. Senior Fellow San Diego SuperComputing Center, USA, 2004.

INSELBERG, Alfred. **The plane with parallel coordinates**. *The Visual Computer*, v. 1, n. 2, p. 69-91, 1985.

INSELBERG, Alfred; DIMSDALE, Bernard. **Parallel coordinates for visualizing multi-dimensional geometry**. Springer Japan, 1987.

KEIM, Daniel A.; SCHNEIDEWIND, Jörn; SIPS, Mike. **CircleView: a new approach for visualizing time-related multidimensional data sets**. In: Proceedings of the working conference on Advanced visual interfaces. p. 179-182, ACM, 2004.

KEIM, Daniel A. et al. **Challenges in Visual Data Analysis**. In: IEEE Tenth International Conference on Information Visualization, London, UK, p. 9-16, jul. 2006.

KEIM, Daniel A. **Information Visualization and Visual Data Mining**. IEEE Transaction on Visualization and Computer Graphics, v. 7, n. 1, p. 100-107, VOL. 7, NO. 1, 1 trim. 2002.

KEIM, Daniel A. **Visual exploration of large databases**. Magazine Communications of the ACM, New York, v. 44, n. 8, p. 38-44, aug. 2001.

KEIM, Daniel A. **Designing pixel-oriented visualization techniques**: Theory and applications. Visualization and Computer Graphics, IEEE Transactions on, v. 6, n. 1, p. 59-78, 2000.

KEIM, Daniel A. et al. **Recursive pattern: A technique for visualizing very large amounts of data**. In: Proceedings of the 6th conference on Visualization'95. IEEE Computer Society, 1995. p. 279.

KEIM, Daniel A. et al. **Supporting data mining of large databases by visual feedback queries**. In: Data Engineering, 1994. Proceedings. 10th International Conference. IEEE, 1994. p. 302-313.

KEIM, Daniel A. **Pixel-oriented database visualizations**. ACM Sigmod Record, v. 25, n. 4, p. 35-39, 1996.

KEIM, Daniel A. **Visual techniques for exploring databases**. Bibliothek der Universität Konstanz, 1997.

KEIM, Daniel A.; KRIEGEL, H.-P. **VisDB: Database exploration using multidimensional visualization**. Computer Graphics and Applications, IEEE, v. 14, n. 5, p. 40-49, 1994.

KEIM, Daniel A.; KRIEGEL, H.-P. **Visualization techniques for mining large databases: A comparison**. In: Knowledge and Data Engineering, IEEE Transactions, v. 8, n. 6, p. 923-938, 1996.

KOEDINGER, K.R., BAKER, R.S.J.d., CUNNINGHAM, K., SKOGSHOLM, A., LEBER, B., Stamper, J. **A Data Repository for the EDM community: The PSLC DataShop**. In Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (Eds.) Handbook of Educational Data Mining. Boca Raton, FL: CRC Press, 2010.

LEBLANC, Jeffrey et al. **Exploring n-dimensional databases**. In: Proceedings of the 1st conference on Visualization'90. IEEE Computer Society Press, 1990. p. 230-237.

LEDERMANN, Flo. **Parvis**. Disponível em <<http://www.mediavirus.org/parvis/>>, acesso em 16/06/2013.

LOHNINGER, H. **INSPECT: a program system to visualize and interpret chemical data**. Chemometrics and intelligent laboratory systems, v. 22, n. 1, p. 147-153, 1994.

LONG, T.V. **Visualizing High-density Clusters in Multidimensional Data**. A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Computer Science, dec. 2009.

LUZZARDI, P. R. G. **Critérios de Avaliação de Técnicas de Visualização de Informações Hierárquicas**. Tese de doutorado, UFRGS, Porto Alegre, 2003.

MAZZA, Riccardo; DIMITROVA, Vania. **CourseVis: Externalising student information to facilitate instructors in distance learning**. In: Proceedings of the International conference in Artificial Intelligence in Education. Sydney, Australia, p. 117-129, 2003.

MCFADDEN, Fred R.; PRESCOTT, Mary B.; HOFFER, Jeffrey A. **Modern database management**. Addison-Wesley Longman Publishing Co., Inc., 1998.

MERCERON, Agathe; YACEF, Kalina. **Educational data mining: a case study**. Artificial Intelligence in education: supporting learning through Socially Informed Technology. IOS Press, p. 467-474, 2005.

MICROSOFT, SQL Server 2012. **Microsoft Developer Network**. Disponível em <<http://msdn.microsoft.com/pt-br/library/ms189084.aspx>>, acesso em 14/04/2013.

NORMAN, Donald A. **Things that make us smart: Defending human attributes in the age of the machine**. 1993.

NORTH, Chris. **Information visualization**. Handbook of Human Factors and Ergonomics, Fourth Edition, p. 1209-1236, 2005.

ORACLE, JavaBeans™ Component API. **Java SE Documentation**. Disponível em <<http://docs.oracle.com/javase/6/docs/technotes/guides/beans/>>, acesso em 16/06/2013.

ORANGE. **Open source data visualization and analysis**. Disponível em <<http://orange.biolab.si/>>, acesso em 22/06/2013.

PICKETT, Ronald M.; GRINSTEIN, Georges G. **Iconographic displays for visualizing multidimensional data**. In: Proc. IEEE Conf. on Systems, Man and Cybernetics, IEEE Press, Piscataway, NJ. 1988. p. 519.

PLAISANT, Catherine. **The challenge of information visualization evaluation**. In: Proceedings of the working conference on Advanced visual interfaces. ACM, 2004. p. 109-116.

ROBERTSON, George G. et al. **Cone trees: animated 3D visualizations of hierarchical information**. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 1991. p. 189-194.

RÜBEL, Oliver et al. **PointCloudXplore: Visual analysis of 3D gene expression data using physical views and parallel coordinates**. In: Proceedings of the Eighth Joint Eurographics/IEEE VGTC conference on Visualization. Eurographics Association, 2006. p. 203-210.

SAVITRAZ, J.D. **Uma Abordagem de Integração e Exploração visual de Dados Educacionais na Plataforma Web-PIDE**. 2010. 87 f. Mestrado em Ciência da Computação – Faculdade de Computação, Universidade Federal de Mato Grosso do Sul, Mato Grosso do Sul. 2010.

SCHWARZ, Alan. **Scientist at Work | Steve C. Wang: Professor Puts a Face on the Performance of Baseball Managers**. Disponível em: <<http://www.nytimes.com/2008/04/01/science/01prof.html>>, acesso em 01/06/2013.

SHNEIDERMAN, Ben. **Inventing discovery tools: combining information visualization with data mining**. In: Keynote for Discovery Science 2001 Conference, Washington, DC, p. 5-6, nov. 2001.

SHNEIDERMAN, Ben. **The eyes have it: A task by data type taxonomy for information visualizations.** In: Visual Languages, 1996. Proceedings., IEEE Symposium on. IEEE, 1996. p. 336-343.

SHNEIDERMAN, Ben. **Tree visualization with tree-maps: 2-d space-filling approach.** ACM Transactions on graphics (TOG), v. 11, n. 1, p. 92-99, 1992.

SIEGEL, John H. et al. **The surgical implications of physiologic patterns in myocardial infarction shock.** Surgery, v. 72, n. 1, p. 126, 1972.

TAN, P., STEINBACH, M., KUMAR, V. **Introdução ao Data Mining.** Ciência Moderna. 900 f, 2009.

TWEEDIE, L. A. **Characterizing Interactive Externalizations.** In: Proceedings of CHI'97. ACM Conference on Human Factors in Computing Systems, 1997. p. 375-382.

VAN WIJK, Jarke J.; VAN LIERE, Robert. **HyperSlice: visualization of scalar functions of many variables.** In: Proceedings of the 4th conference on Visualization'93. IEEE Computer Society, 1993. p. 119-125.

VASUDEVAN, Venu. **Supporting high-bandwidth navigation in object-bases.** In: Data Engineering, 1994. Proceedings. 10th International Conference. IEEE, 1994. p. 294-301.

WARD, Matthew O. **Xmdvtool: Integrating multiple methods for visualizing multivariate data.** In: Proceedings of the Conference on Visualization'94. IEEE Computer Society Press, 1994. p. 326-333.

WARE, Colin. **Information visualization: perception for design.** Morgan Kaufmann Pub, 2012.

WONG, Pak Chung; BERGERON, R. Daniel. **30 years of multidimensional multivariate visualization.** Scientific Visualization, Overviews, Methodologies, and Techniques. Washington, DC, USA: IEEE Computer Society, p. 3-33, 1997.

WONG, Pak Chung; BERGERON, R. Daniel. **Multiresolution multidimensional wavelet brushing.** In: Visualization'96. Proceedings. IEEE, 1996. p. 141-148.

ZAT, Daline. **Estudo Comparativo de Algoritmos de Agrupamento para Mineração de Dados Educacionais**. Projeto de Graduação, Centro de Computação e Tecnologia da Informação da Universidade de Caxias do Sul, UCS, Caxias do Sul, RS, Brasil, 2012.