

Técnicas de Visualização para Dados Multivariados

SCC5836 – Visualização Computacional

Prof. Fernando V. Paulovich

<http://www.icmc.usp.br/~paulovic>
paulovic@icmc.usp.br

Instituto de Ciências Matemáticas e de Computação (ICMC)
Universidade de São Paulo (USP)

- 1 Introdução
- 2 Técnicas baseadas em Ponto
 - Scatterplots
 - Projeção Multidimensional
 - Outras Técnicas
- 3 Técnicas baseadas em Linhas
- 4 Técnicas baseadas em Região
- 5 Combinação de Técnicas
- 6 Referências

- Técnicas que serão discutidas servem para a **visualização de listas e tabelas de dados** que **não** apresentam **atributo espacial**

- 1 Introdução
- 2 **Técnicas baseadas em Ponto**
 - Scatterplots
 - Projeção Multidimensional
 - Outras Técnicas
- 3 Técnicas baseadas em Linhas
- 4 Técnicas baseadas em Região
- 5 Combinação de Técnicas
- 6 Referências

Técnicas baseadas em Ponto

- Visualizações baseados em pontos são **projeções** de instâncias m -dimensionais para pontos em um **espaço visual** p -dimensional
 - Um *glifo* é associado a cada ponto p -dimensional

- 1 Introdução
- 2 **Técnicas baseadas em Ponto**
 - **Scatterplots**
 - Projeção Multidimensional
 - Outras Técnicas
- 3 Técnicas baseadas em Linhas
- 4 Técnicas baseadas em Região
- 5 Combinação de Técnicas
- 6 Referências

Scatterplots ou Matrizes de Scatterplots

- **Scatterplots** são a visualização mais comuns, mas conforme a **dimensionalidade aumenta**, é preciso buscar estratégias alternativas

Scatterplots ou Matrizes de Scatterplots

- **Scatterplots** são a visualização mais comuns, mas conforme a **dimensionalidade aumenta**, é preciso buscar estratégias alternativas
 - **Escolha das dimensões:** manualmente ou por algum algoritmo

Scatterplots ou Matrizes de Scatterplots

- **Scatterplots** são a visualização mais comuns, mas conforme a **dimensionalidade aumenta**, é preciso buscar estratégias alternativas
 - **Escolha das dimensões**: manualmente ou por algum algoritmo
 - **Redução do número de dimensões**: usando técnicas como PCA

Scatterplots ou Matrizes de Scatterplots

- **Scatterplots** são a visualização mais comuns, mas conforme a **dimensionalidade aumenta**, é preciso buscar estratégias alternativas
 - **Escolha das dimensões**: manualmente ou por algum algoritmo
 - **Redução do número de dimensões**: usando técnicas como PCA
 - **Incorporação das dimensões**: mapear outras dimensões para atributos gráficos

Scatterplots ou Matrizes de Scatterplots

- **Scatterplots** são a visualização mais comuns, mas conforme a **dimensionalidade aumenta**, é preciso buscar estratégias alternativas
 - **Escolha das dimensões**: manualmente ou por algum algoritmo
 - **Redução do número de dimensões**: usando técnicas como PCA
 - **Incorporação das dimensões**: mapear outras dimensões para atributos gráficos
 - **Múltiplos *displays***: mostrar usando sobreposição ou justaposição

Scatterplots ou Matrizes de Scatterplots

- **Matriz de scatterplots** define uma grade de m^2 células, sendo m o número de dimensões

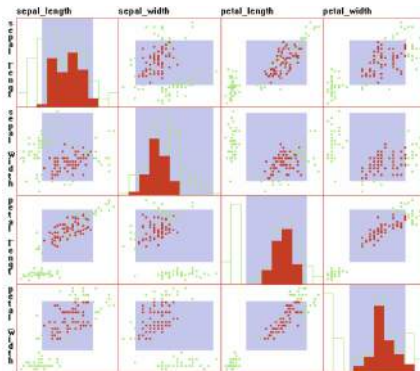


Figura: Pontos em vermelho são instâncias selecionadas.

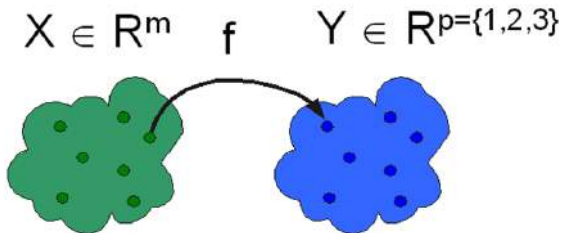
- 1 Introdução
- 2 **Técnicas baseadas em Ponto**
 - Scatterplots
 - **Projeção Multidimensional**
 - Outras Técnicas
- 3 Técnicas baseadas em Linhas
- 4 Técnicas baseadas em Região
- 5 Combinação de Técnicas
- 6 Referências

- O princípio dos **métodos de projeção** é tentar **preservar no espaço projetado p -dimensional as relações** presentes no espaço m -dimensional original
 - Por exemplo, relações de **similaridades, distâncias, vizinhanças**, etc.

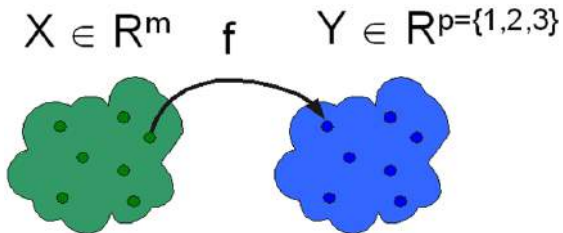
Projeção Multidimensional

- O princípio dos **métodos de projeção** é tentar **preservar no espaço projetado p -dimensional as relações** presentes no espaço m -dimensional original
 - Por exemplo, relações de **similaridades, distâncias, vizinhanças**, etc.
- Se $p = 2$ o resultado é um conjunto de **pontos no plano**
 - Pontos próximos indicam instâncias relacionadas, pontos distantes indicam instâncias não-relacionadas

Projeção Multidimensional baseada em Distância

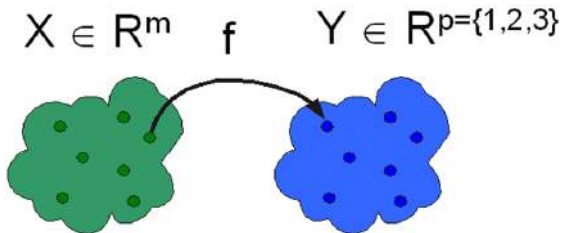


Projeção Multidimensional baseada em Distância



- $\delta : x_i, x_j \rightarrow \mathbb{R}, x_i, x_j \in X$

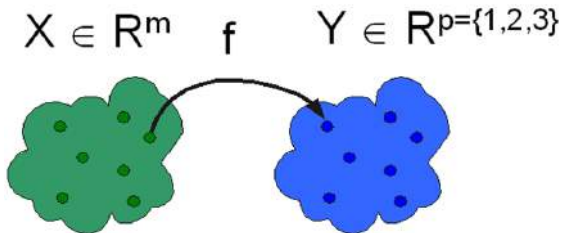
Projeção Multidimensional baseada em Distância



- $\delta : x_i, x_j \rightarrow \mathbb{R}, x_i, x_j \in X$

- $d : y_i, y_j \rightarrow \mathbb{R}, y_i, y_j \in Y$

Projeção Multidimensional baseada em Distância



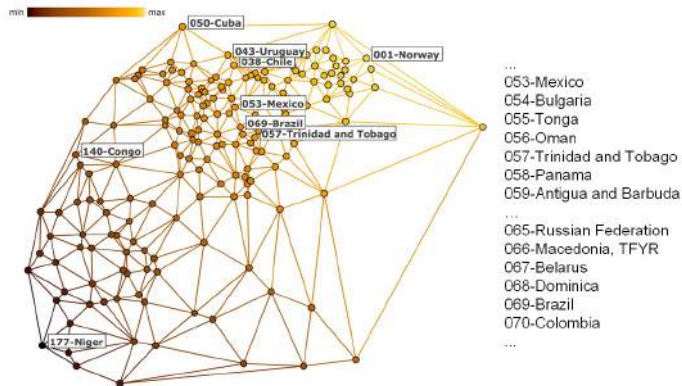
- $\delta : x_i, x_j \rightarrow \mathbb{R}, x_i, x_j \in X$

- $d : y_i, y_j \rightarrow \mathbb{R}, y_i, y_j \in Y$

- $f : X \rightarrow Y, |\delta(x_i, x_j) - d(f(x_i), f(x_j))| \approx 0, \forall x_i, x_j \in X$

Projeção Multidimensional

- IDH/ONU de 2006 (<http://hdr.undp.org/hdr2006/statistics/>)
(<http://hdr.undp.org/en/data>)



- <http://infoserver.lcad.icmc.usp.br/>

Projeção Multidimensional

Símbolo	Significado
X	conjunto de objetos no espaço original m -dimensional.
m	dimensionalidade do espaço original.
x_i	i -ésimo objeto no espaço original. Quando esse admite uma representação vetorial, suas coordenadas são dadas por $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{im})$.
Y	conjunto de pontos no espaço projetado p -dimensional.
p	dimensionalidade do espaço projetado.
y_i	i -ésimo ponto do espaço projetado. Quando esse admite uma representação vetorial, suas coordenadas são dadas por $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{ip})$.
n	número de objetos no espaço original e de pontos no projetado.
$\delta(\mathbf{x}_i, \mathbf{x}_j)$	dissimilaridade entre os objetos i e j no espaço original.
$d(\mathbf{y}_i, \mathbf{y}_j)$	distância entre os pontos i e j no espaço projetado.

Tabela: Símbolos mais frequentes e seus significados.

- Podemos **categorizar** as técnicas de projeção nos seguintes grupos

- Podemos **categorizar** as técnicas de projeção nos seguintes grupos
 - Técnicas baseadas em força (*Force-Direct Placement (FDP)*)

- Podemos **categorizar** as técnicas de projeção nos seguintes grupos
 - Técnicas baseadas em força (*Force-Direct Placement (FDP)*)
 - Técnicas de decomposição espectral

- Podemos **categorizar** as técnicas de projeção nos seguintes grupos
 - Técnicas baseadas em força (*Force-Direct Placement (FDP)*)
 - Técnicas de decomposição espectral
 - Técnicas de redução de dimensionalidade

Técnicas baseadas em Força

- As técnicas de projeção baseadas no **Modelo de Molas** são as mais simples
 - Tentam levar um sistema de objetos conectados por molas a um **estado de equilíbrio**
 - Sistema iterativo que, a partir de uma configuração inicial atualiza as forças associadas às molas para puxar ou empurrar os objetos até atingir uma posição de equilíbrio
 - As **forças** no sistema são calculadas proporcionais à diferença entre as dissimilaridades $\delta(\mathbf{x}_i, \mathbf{x}_j)$ e as distâncias $d(\mathbf{y}_i, \mathbf{y}_j)$

- Os **objetos** são modelados como **partículas ponto-massa** ligadas entre si por meio de **molas**
 - Conjunto de partículas está sujeito às leis de *Newton*

$$f = m \times a \quad (1)$$

Modelo de Molas

- Os **objetos** são modelados como **partículas ponto-massa** ligadas entre si por meio de **molas**
 - Conjunto de partículas está sujeito às leis de *Newton*

$$f = m \times a \quad (1)$$

- Resolve-se um sistema de equação diferenciais ordinárias de segunda ordem. Problema: o custo computacional é alto, $O(n^2)$ a **cada iteração**

$$\begin{cases} v' = a = f/m \\ p' = v \end{cases} \quad (2)$$

Force Scheme

- **Force Scheme** [Tejada et al., 2003] é uma outra abordagem, mas ainda com custo computacional alto, $O(n^2)$

- 1: **para** iter=1 **até** k **faça**
- 2: **para todo** $y_i \in Y$ **faça**
- 3: **para todo** $y_j \in Y$ com $y_j \neq y_i$ **faça**
- 4: Calcular \vec{v} como sendo o vetor de y_i para y_j
- 5: Mover y_j em direção de \vec{v} uma fração de Δ
- 6: **fim para**
- 7: **fim para**
- 8: Normalizar as coordenadas da projeção no intervalo $[0, 1]$ em ambas as dimensões
- 9: **fim para**

$$\Delta = \frac{\delta(\mathbf{x}_i, \mathbf{x}_j) - \delta_{min}}{\delta_{max} - \delta_{min}} - d(\mathbf{y}_i, \mathbf{y}_j) \quad (3)$$

Sammon's Mapping

- A **Sammon's Mapping** [Sammon, 1969] minimiza uma função de perda usando o método de gradientes descendentes

$$S = \frac{1}{\sum_{i < j} \delta(\mathbf{x}_i, \mathbf{x}_j)} \sum_{i < j} \frac{(d(\mathbf{y}_i, \mathbf{y}_j) - \delta(\mathbf{x}_i, \mathbf{x}_j))^2}{\delta(\mathbf{x}_i, \mathbf{x}_j)} \quad (4)$$

- A m -ésima iteração desse método é definida como

$$y_{pq}(m+1) = y_{pq}(m) - MF \times \Delta_{pq}(m) \quad (5)$$

- Onde y_{pq} denota a coordenada q do ponto p e

$$\Delta_{pq}(m) = \frac{\partial S(m)}{\partial y_{pq}(m)} \bigg/ \left| \frac{\partial^2 S(m)}{\partial y_{pq}^2(m)} \right| \quad (6)$$

Aproximação de Pekalska

- A técnica de **Sammon's Mapping** é $O(n^3)$. A fim de reduzir essa complexidade, Pekalska et al [Pekalska et al., 1999] definiram uma abordagem em que apenas uma amostra inicial dos pontos é projetada, e as posições dos demais elementos são interpoladas a partir do posicionamento dos pontos projetados.

- Para essa interpolação, o operador V é encontrado resolvendo o sistema linear

$$D \times V = Y$$

- onde D é a matriz de distâncias $\delta(\mathbf{x}_i, \mathbf{x}_j)$ entre os elementos da amostra e Y é a projeção da amostra

Classical Scaling (MDS)

- A **Classical Scaling (MDS)** [Borg and Groenen, 2005] foi uma das primeiras técnicas de projeção (definida em 1930)

- Sejam \mathbf{x}_i ($i = 1, \dots, n$) as coordenadas de n pontos em um espaço Euclidiano m -dimensional, onde $\mathbf{x}_i = (x_{i1}, \dots, x_{im})^T$, e seja \mathbf{B} a matrix do produto interno entre vetores,

$$[\mathbf{B}]_{ij} = b_{ij} = \mathbf{x}_i^T \mathbf{x}_j$$

- Com distância Euclidiana entre os pontos i e j dada por

$$\hat{\delta}(\mathbf{x}_i, \mathbf{x}_j)^2 = (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j) \quad (7)$$

- A partir de uma matriz da distâncias ($\hat{\delta}(\mathbf{x}_i, \mathbf{x}_j)^2$), encontra a matriz do produto interno \mathbf{B} , e a partir de \mathbf{B} calcula as coordenadas dos pontos

Classical Scaling (MDS)

- A matrix \mathbf{B} pode ser reescrita como

$$\mathbf{B} = -\frac{1}{2}\mathbf{H}\mathbf{A}\mathbf{H} \quad (8)$$

- \mathbf{A} é a matrix $[\mathbf{A}]_{ij} = a_{ij} = \hat{\delta}(\mathbf{x}_i, \mathbf{x}_j)^2$
- \mathbf{H} é a matrix de centragem

- Usando decomposição espectral, \mathbf{B} pode ser reescrita como

$$\mathbf{B} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$$

- Como $\mathbf{B} = \mathbf{X}\mathbf{X}^T$, a matrix de coordenadas \mathbf{X} é dada por

$$\mathbf{X} = \mathbf{V}_1\mathbf{\Lambda}_1^{-\frac{1}{2}}$$

Definição

- Técnicas para **redução de dimensionalidade** buscam encontrar uma representação em um espaço de **menor dimensão** que capture o **conteúdo original** dos dados segundo algum critério

Principal Component Analysis (PCA)

- **PCA** [Jolliffe, 1986] é uma das técnicas de redução de dimensionalidade mais utilizadas
 - Consegue **capturar** a maior parte da **variabilidade** com poucas dimensões
 - **Elimina** grande parte do “**ruído**” existente, etc.

Principal Component Analysis (PCA)

- **PCA** [Jolliffe, 1986] é uma das técnicas de redução de dimensionalidade mais utilizadas
 - Consegue **capturar** a maior parte da **variabilidade** com poucas dimensões
 - **Elimina** grande parte do “**ruído**” existente, etc.

- Primeiro é criada a matriz de covariância $\mathbf{C}_{m \times m}$ dos atributos
- Em seguida obtém a decomposição espectral, encontrando

$$\mathbf{C} = \mathbf{U}\mathbf{A}\mathbf{U}^T \quad (9)$$

- Por fim é obtido o espaço reduzido fazendo

$$\mathbf{S} = \mathbf{X} \times [u_1, u_2, \dots, u_p] \quad (10)$$

Least-Square Projection (LSP)

- Na **Least-Square Projection (LSP)** [Paulovich et al., 2008] primeiro é selecionada uma **amostra** das instâncias multidimensionais, chamados de “pontos de controle”. Esses pontos são projetados em um espaço \mathbb{R}^p
- Para cada objeto é definido um conjunto de vizinhos no \mathbb{R}^m .
- Fazendo uso das **relações de vizinhança** definidas no \mathbb{R}^m é construído um sistema linear cuja solução é uma projeção das instâncias restantes no **fecho convexo** de seus respectivos k vizinhos mais próximos

Least-Square Projection (LSP)

- Seja $V_i = \{p_{i_1}, \dots, p_{i_{k_i}}\}$ um conjunto de k_i pontos em uma vizinhança de um ponto p_i e \tilde{p}_i sejam as coordenadas de p_i no \mathbb{R}^d

- Suponha que as coordenadas \tilde{p}_i sejam dadas pela seguinte equação

$$\begin{aligned}\tilde{p}_i - \sum_{p_j \in V_i} \alpha_{ij} \tilde{p}_j &= 0 \\ 0 \leq \alpha_{ij} \leq 1; \sum \alpha_{ij} &= 1\end{aligned}\tag{11}$$

- Quando $\alpha_{ij} = \frac{1}{k_i}$ teremos p_i no centróide dos pontos em V_i

Least-Square Projection (LSP)

- A partir dessa equação é definido um conjunto de sistemas lineares

$$L\mathbf{x}_1 = 0, L\mathbf{x}_2 = 0, \dots, L\mathbf{x}_d = 0 \quad (12)$$

- Onde $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_d$ são os vetores contendo as coordenadas cartesianas dos pontos e L é a matriz $n \times n$ cujas entradas são dadas por

$$l_{ij} = \begin{cases} 1 & i = j \\ -\alpha_{ij} & p_j \in V_i \\ 0 & \text{caso contrário} \end{cases}$$

Least-Square Projection (LSP)

- Os pontos de controle são inseridos no sistema como novas linhas na matriz
- Assim, dado um conjunto de pontos de controle $S_c = \{p_{c_1}, \dots, p_{c_{nc}}\}$, é possível re-escrever o sistema

$$A\mathbf{x} = \mathbf{b} \quad (13)$$

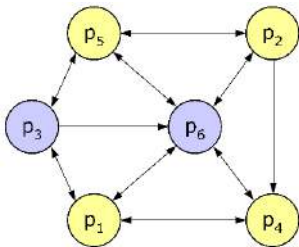
- Em que A é uma matrix retangular $(n + nc) \times n$ dada por

$$A = \begin{pmatrix} L \\ C \end{pmatrix}, \quad c_{ij} = \begin{cases} 1 & x_j \text{ é um ponto de controle} \\ 0 & \text{caso contrário} \end{cases}$$

- E \mathbf{b} é o vetor:

$$b_i = \begin{cases} 0 & i \leq n \\ x_{i_c} & n < i \leq n + nc \end{cases}$$

Least-Square Projection (LSP)



(a) Grafo de vizinhança.

$$A = \begin{array}{|cccccc|} \hline 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \hline \end{array} \begin{array}{l} \leftarrow L \\ \leftarrow C \end{array}$$

(b) Matrix Laplaciana.

Least-Square Projection (LSP)

- O **sistema linear** com os pontos de controle apresenta **rank-completo** e pode ser resolvido aplicando-se **mínimos quadrados**
 - Encontrar \mathbf{x} que minimize $\|A\mathbf{x} - \mathbf{b}\|^2$, isto é, $\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}$

Exemplos de Projeção

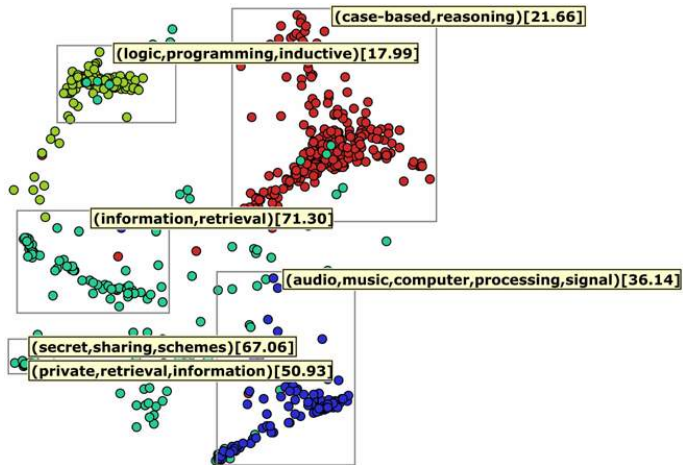


Figura: Projeção de coleção com artigos científicos (LSP).

Exemplos de Projeção

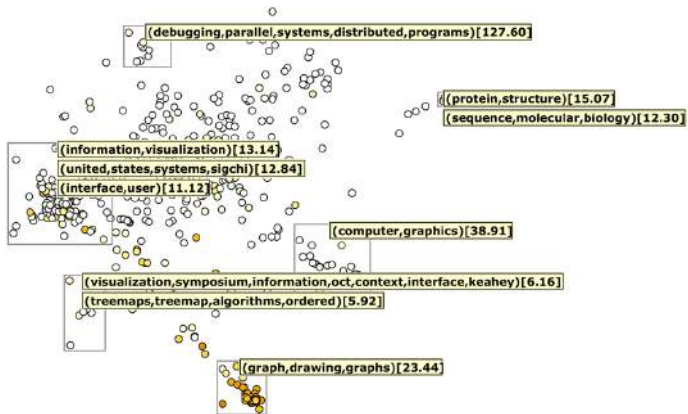


Figura: Projeção de coleção com artigos científicos (LSP).

Exemplos de Projeção

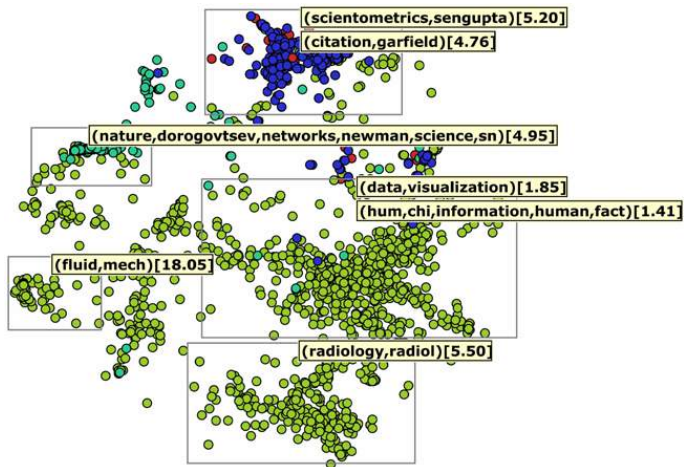


Figura: Projeção de coleção com artigos científicos (LSP).

Exemplos de Projeção

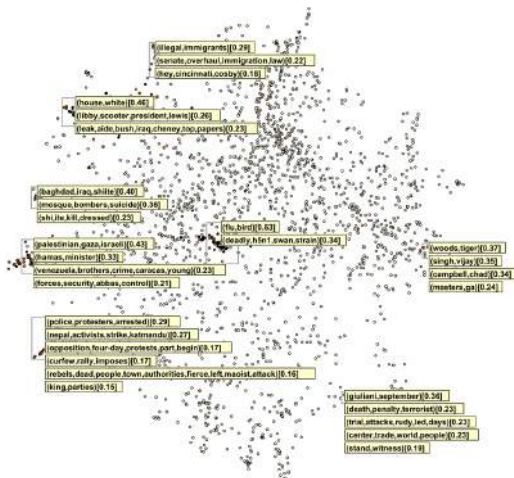


Figura: Projeção de coleção notícias curtas de jornal RSS (LSP).

Exemplos de Projeção

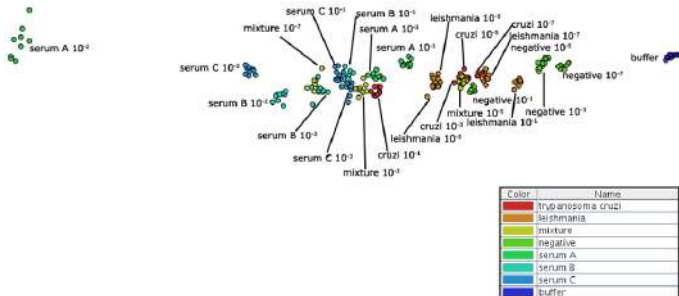


Figura: Projeção de dados de sensores (PCA).

Exemplos de Projeção

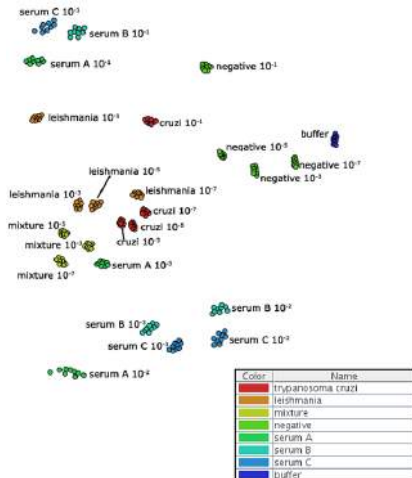


Figura: Projeção de dados de sensores (Force Scheme).

Exemplos de Projeção



Color	Name
Magenta	Phytic Acid 10-2
Dark Blue	Phytic Acid 10-3
Light Pink	Phytic Acid 10-4
Teal	Phytic Acid 10-5
Yellow	Phytic Acid 10-6
Green	Buffer

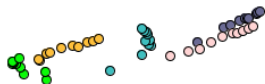


Figura: Projeção de dados de sensores (Force Scheme). Escolhendo o melhor sensor.

Exemplos de Projeção

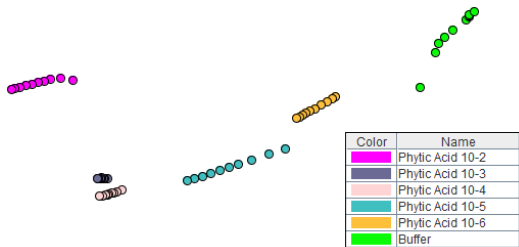


Figura: Projeção de dados de sensores (Force Scheme). Escolhendo o melhor sensor.

Exemplos de Projeção

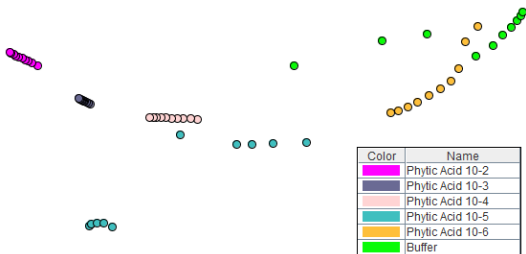


Figura: Projeção de dados de sensores (Force Scheme). Escolhendo o melhor sensor.

Exemplos de Projeção



Figura: Projeção de dados de sensores (Force Scheme). Escolhendo o melhor sensor.

- 1 Introdução
- 2 **Técnicas baseadas em Ponto**
 - Scatterplots
 - Projeção Multidimensional
 - **Outras Técnicas**
- 3 Técnicas baseadas em Linhas
- 4 Técnicas baseadas em Região
- 5 Combinação de Técnicas
- 6 Referências

- Outra técnica baseada em **pontos** bastante conhecida é a **RadViz**
 - Para um conjunto m -dimensional, m âncoras são definidas e posicionadas uniformemente ao longo de uma circunferência

Técnicas baseadas em Pontos

- Outra técnica baseada em **pontos** bastante conhecida é a **RadViz**
 - Para um conjunto m -dimensional, m âncoras são definidas e posicionadas uniformemente ao longo de uma circunferência
- Uma instância de dado é mapeada na posição v considerando uma circunferência de raio 1 centrada na origem, fazendo
 - Seja $\mathbf{x} = (x_0, x_1, \dots, x_{m-1})$ o vetor normalizado representando uma instância \mathbf{x} , e um conjunto de vetores unitários A , em que A_j representa a $j^{\text{ésima}}$ âncora no plano, o vetor v dá a posição de equilíbrio do ponto:

$$\sum_{j=0}^{m-1} (A_j - v)x_j = 0$$

$$v = \frac{\sum_{j=0}^{m-1} (A_j x_j)}{\sum_{j=0}^{m-1} x_j}$$

- É importante considerar que **ordenações diferentes** das âncoras levarão a **resultados diferentes**
 - **Interação** pode ajudar o usuário
 - Técnicas que automaticamente **maximizam o espalhamento** dos pontos são boas alternativas

RadViz

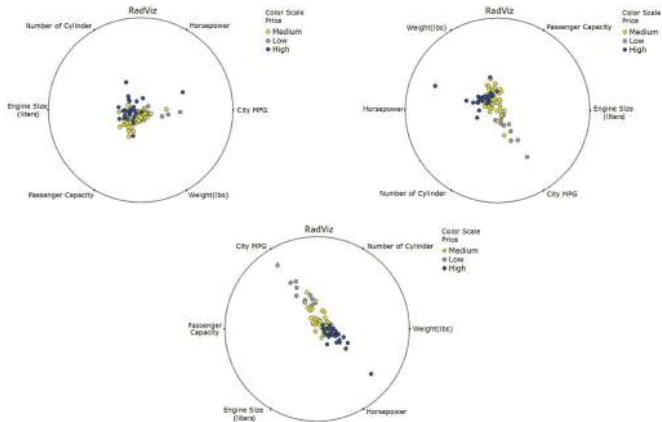


Figura: Diferentes ordenações das âncoras no RadViz. Preço do carro é mapeado para a cor do glifo.

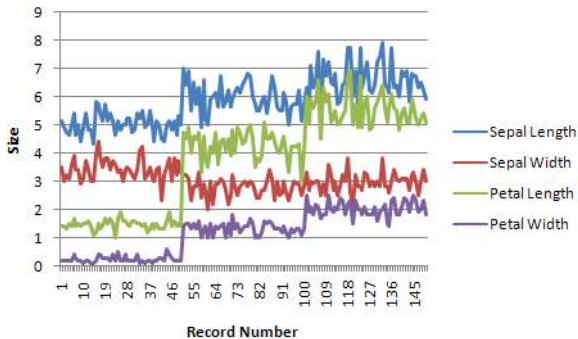
- 1 Introdução
- 2 Técnicas baseadas em Ponto
 - Scatterplots
 - Projeção Multidimensional
 - Outras Técnicas
- 3 Técnicas baseadas em Linhas**
- 4 Técnicas baseadas em Região
- 5 Combinação de Técnicas
- 6 Referências

Técnicas Baseadas em Linhas

- Nas representações usando **linha**, características dos dados podem ser percebidas por meio de **cruzamentos**, **curvatura**, etc.

Gráficos de Linhas

- Um gráfico de linha representa no eixo **vertical a faixa de valores** de diferentes atributos das instâncias e no **horizontal alguma ordenação** dessas instâncias
 - Visualização univariada para **poucas dimensões (atributos)**



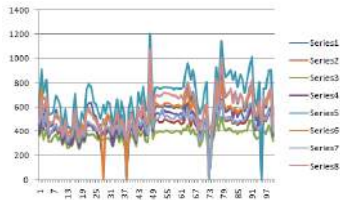
- **Sobreposição** não deve ser empregada com muitas dimensões por problemas de **oclusão**

Gráficos de Linhas

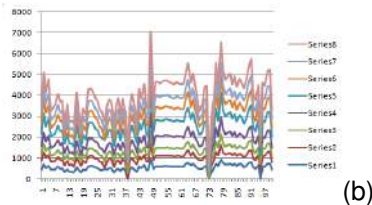
- **Sobreposição** não deve ser empregada com muitas dimensões por problemas de **oclusão**
- Possíveis soluções
 - **Empilhar** as linhas usando como base a dimensão anterior (difícil de avaliar o valor)

- **Sobreposição** não deve ser empregada com muitas dimensões por problemas de **oclusão**
- Possíveis soluções
 - **Empilhar** as linhas usando como base a dimensão anterior (difícil de avaliar o valor)
 - **Ordenar** as instâncias com base em uma dimensão

Gráficos de Linhas



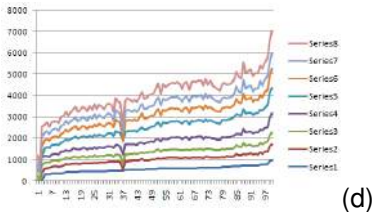
(a)



(b)



(c)



(d)

Figura: Gráficos representando salários de professores em 100 universidades. (a) sobreposição (b) empilhamento (c) sobreposição ordenada (d) empilhamento ordenado.

- Essa técnica só **funciona** se existir um **relacionamento entre as unidades** das diferentes dimensões

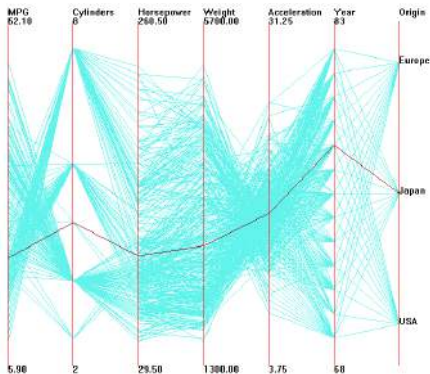
- Essa técnica só **funciona** se existir um **relacionamento entre as unidades** das diferentes dimensões
- Possíveis soluções
 - **Múltiplos eixos verticais** podem ser usados para as diferentes dimensões

Gráficos de Linhas

- Essa técnica só **funciona** se existir um **relacionamento entre as unidades** das diferentes dimensões
- Possíveis soluções
 - **Múltiplos eixos verticais** podem ser usados para as diferentes dimensões
 - Criar múltiplos **gráficos empilhados coordenados**

Coordenadas Paralelas

- Na técnica **Coordenadas Paralelas**, os eixos são paralelos (não ortogonais) espaçados uniformemente, e as instâncias de dados são polilinhas que interceptam esses eixos em uma posição determinada pelo valor do atributo associado ao eixo



- Na **interpretação** desse tipo de visualização podemos

Coordenadas Paralelas

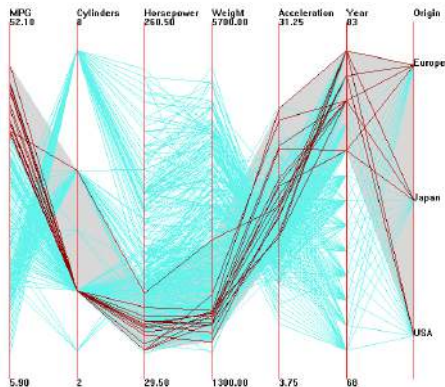
- Na **interpretação** desse tipo de visualização podemos
 - Localizar **grupos** de polilinhas **semelhantes**

- Na **interpretação** desse tipo de visualização podemos
 - Localizar **grupos** de polilinhas **semelhantes**
 - Localizar pontos de **cruzamento**

- Na **interpretação** desse tipo de visualização podemos
 - Localizar **grupos** de polilinhas **semelhantes**
 - Localizar pontos de **cruzamento**
 - Localizar polilinhas **diferentes**

Coordenadas Paralelas

- Interpretar relações entre **dimensões não consecutivas** pode ser **difícil**
 - **Interação** pode ajudar a superar esse problema



- Existem diversas variantes da técnica de coordenadas paralelas

- Existem diversas variantes da técnica de coordenadas paralelas
 - **Coordenadas paralelas hierárquicas** mostra agrupamentos e não os dados originais

Coordenadas Paralelas

- Existem diversas variantes da técnica de coordenadas paralelas
 - **Coordenadas paralelas hierárquicas** mostra agrupamentos e não os dados originais
 - **Semi-transparência** pode ser usada para mostrar grupos em grandes bases de dados

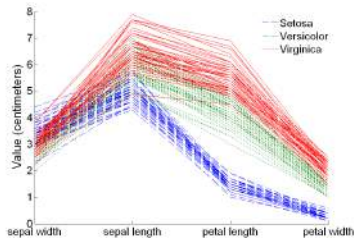
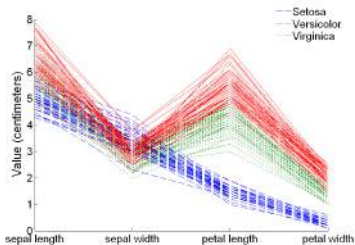
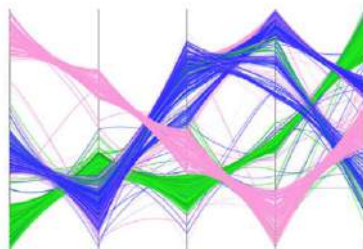
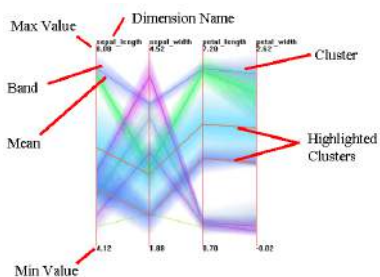
Coordenadas Paralelas

- Existem diversas variantes da técnica de coordenadas paralelas
 - **Coordenadas paralelas hierárquicas** mostra agrupamentos e não os dados originais
 - **Semi-transparência** pode ser usada para mostrar grupos em grandes bases de dados
 - Agrupamento, re-ordenação e diferentes espaçamentos baseadas em **correlação**

- Existem diversas variantes da técnica de coordenadas paralelas
 - **Coordenadas paralelas hierárquicas** mostra agrupamentos e não os dados originais
 - **Semi-transparência** pode ser usada para mostrar grupos em grandes bases de dados
 - Agrupamento, re-ordenação e diferentes espaçamentos baseadas em **correlação**
 - **Incorporação de histogramas** nos eixos para ajudar na interpretação das distribuições

- Existem diversas variantes da técnica de coordenadas paralelas
 - **Coordenadas paralelas hierárquicas** mostra agrupamentos e não os dados originais
 - **Semi-transparência** pode ser usada para mostrar grupos em grandes bases de dados
 - Agrupamento, re-ordenação e diferentes espaçamentos baseadas em **correlação**
 - **Incorporação de histogramas** nos eixos para ajudar na interpretação das distribuições
 - Usar **curvas nos pontos de intersecção** para melhorar a continuidade entre eixos

Coordenadas Paralelas



Curvas de Andrews

- Na técnica de **Curvas de Andrews**, cada ponto multivariado $D = (d_1, d_2, \dots, d_N)$ é usado para criar uma curva da forma

$$f(t) = \frac{d_1}{\sqrt{2}} + d_2 \sin(t) + d_3 \cos(t) + d_4 \sin(2t) + d_5 \cos(2t) + \dots$$

- Com $-\pi \leq t \leq \pi$

- Para número ímpar de dimensões o termo final é $d_N \cos(\frac{N-1}{2}t)$ e para número par é $d_N \sin(\frac{N}{2}t)$

Curvas de Andrews

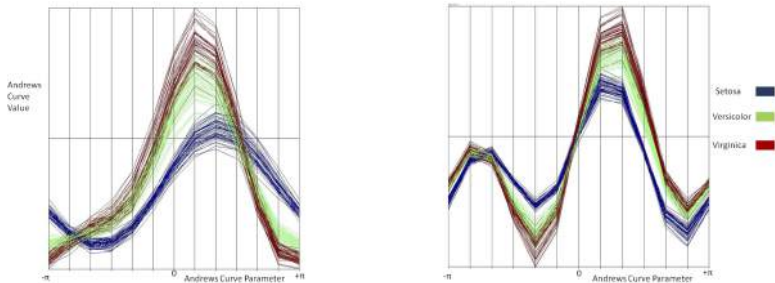
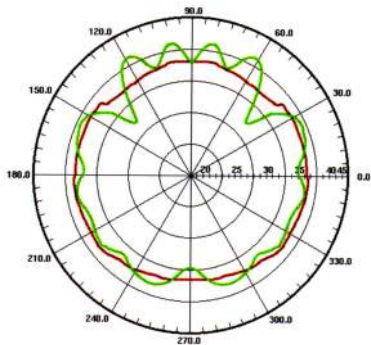


Figura: Resultados da técnica de Curvas de Andrews usando diferentes ordenações das dimensões (dados da Iris).

Técnicas de Eixos Radiais

- Para cada técnica em que os eixos são paralelos existe uma correspondente com os eixos **radiais**
 - **Circular line graph** é um exemplo, especialmente útil para eventos cíclicos



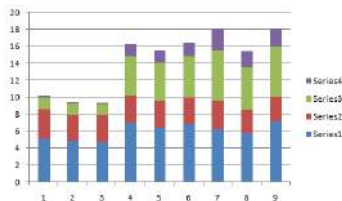
- 1 Introdução
- 2 Técnicas baseadas em Ponto
 - Scatterplots
 - Projeção Multidimensional
 - Outras Técnicas
- 3 Técnicas baseadas em Linhas
- 4 **Técnicas baseadas em Região**
- 5 Combinação de Técnicas
- 6 Referências

Técnicas baseadas em Região

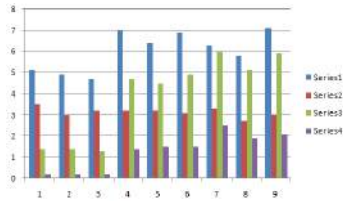
- Técnicas baseadas em **região** apresentam valores empregando **polígonos preenchidos** mapeados para **tamanho, cor, forma** ou outros atributos
 - Apesar das limitações cognitivas muitas técnicas vem sendo desenvolvidas

Gráficos de Barras e Histogramas

- Uma visualização muito empregada são os **gráficos de barras**



(a)



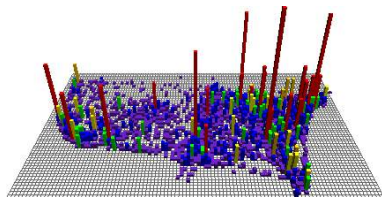
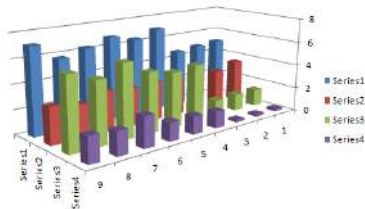
(b)

Figura: Exemplos de gráficos de barras. (a) empilhados (b) agrupados.

- Se a tarefa é estudar a **distribuição dos dados**, pode-se usar **histogramas**
 - Com valores **nominais** ou em um **intervalo de poucos inteiros** é simples – mesmo número de barras quanto de valores distintos

Gráficos de Barras e Histogramas

- A versão 3D do gráfico de barras é conhecido como **Cityscape**
 - Se todas as células forem populadas também recebe o nome de **Histograma 3D**
 - Técnica bastante popular para dados **geo-referenciados**

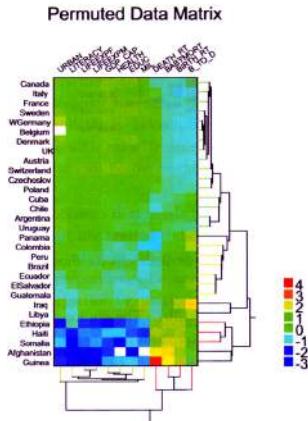


Displays Tabulares

- No caso dos **dados** estarem organizados como uma **tabela**, **visualizações podem ser facilmente geradas**

Displays Tabulares

- **Heatmaps** mapeiam os valores de uma tabela em cores, renderizando-os como quadrados/retângulos coloridos

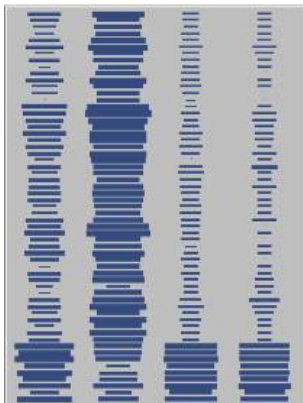


Displays Tabulares

- Pode aplicar **Permutações** e **Reordenações** de linhas ou colunas para **melhorar** a visualização
 - Maximizar a diagonal

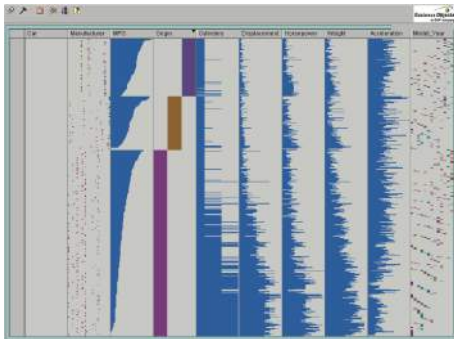
Displays Tabulares

- **Survey Plots** varia o **tamanho das células** ao invés de colorir
 - Evita problemas de percepção causados por interferência em cores adjacentes



Displays Tabulares

- **TableLens** combina todas essas ideias e adiciona mecanismos que permitem visualizar uma tabela como um todo, assim como ter uma **visão detalhada**
- Procure por *TableLens demo*



- 1 Introdução
- 2 Técnicas baseadas em Ponto
 - Scatterplots
 - Projeção Multidimensional
 - Outras Técnicas
- 3 Técnicas baseadas em Linhas
- 4 Técnicas baseadas em Região
- 5 Combinação de Técnicas**
- 6 Referências

Combinação de Técnicas

- Existe uma gama de técnicas que **combinam características** de duas ou mais classes de técnicas apresentadas
 - A criação de glifos

- Alguns exemplos de **glifos** são

- Alguns exemplos de **glifos** são
 - **Profiles** – altura e cor de barras

- Alguns exemplos de **glifos** são
 - **Profiles** – altura e cor de barras
 - **Stars** – tamanho de raios igualmente espaçados emanando do centro

- Alguns exemplos de **glifos** são
 - **Profiles** – altura e cor de barras
 - **Stars** – tamanho de raios igualmente espaçados emanando do centro
 - **Anderson/metroglyphs** – tamanho de raios

- Alguns exemplos de **glifos** são
 - **Profiles** – altura e cor de barras
 - **Stars** – tamanho de raios igualmente espaçados emanando do centro
 - **Anderson/metroglyphs** – tamanho de raios
 - **Stick figures** – tamanho, ângulo, cor do membro

- Alguns exemplos de **glifos** são
 - **Profiles** – altura e cor de barras
 - **Stars** – tamanho de raios igualmente espaçados emanando do centro
 - **Anderson/metroglyphs** – tamanho de raios
 - **Stick figures** – tamanho, ângulo, cor do membro
 - **Trees** – tamanho, espessura, ângulos dos ramos, etc.

- Alguns exemplos de **glifos** são
 - **Profiles** – altura e cor de barras
 - **Stars** – tamanho de raios igualmente espaçados emanando do centro
 - **Anderson/metroglyphs** – tamanho de raios
 - **Stick figures** – tamanho, ângulo, cor do membro
 - **Trees** – tamanho, espessura, ângulos dos ramos, etc.
 - **Autoglyph** – cor de caixas

- Alguns exemplos de **glifos** são
 - **Profiles** – altura e cor de barras
 - **Stars** – tamanho de raios igualmente espaçados emanando do centro
 - **Anderson/metroglyphs** – tamanho de raios
 - **Stick figures** – tamanho, ângulo, cor do membro
 - **Trees** – tamanho, espessura, ângulos dos ramos, etc.
 - **Autoglyph** – cor de caixas
 - **Boxes** – altura, comprimento, profundidade, etc.

- Alguns exemplos de **glifos** são
 - **Profiles** – altura e cor de barras
 - **Stars** – tamanho de raios igualmente espaçados emanando do centro
 - **Anderson/metroglyphs** – tamanho de raios
 - **Stick figures** – tamanho, ângulo, cor do membro
 - **Trees** – tamanho, espessura, ângulos dos ramos, etc.
 - **Autoglyph** – cor de caixas
 - **Boxes** – altura, comprimento, profundidade, etc.
 - **Hedgehogs** – picos em um campo vetorial, com variação em orientação e espessura

- Alguns exemplos de **glifos** são
 - **Profiles** – altura e cor de barras
 - **Stars** – tamanho de raios igualmente espaçados emanando do centro
 - **Anderson/metroglyphs** – tamanho de raios
 - **Stick figures** – tamanho, ângulo, cor do membro
 - **Trees** – tamanho, espessura, ângulos dos ramos, etc.
 - **Autoglyph** – cor de caixas
 - **Boxes** – altura, comprimento, profundidade, etc.
 - **Hedgehogs** – picos em um campo vetorial, com variação em orientação e espessura
 - **Faces** – tamanho e posição dos olhos, nariz, boca, etc.

- Alguns exemplos de **glifos** são
 - **Profiles** – altura e cor de barras
 - **Stars** – tamanho de raios igualmente espaçados emanando do centro
 - **Anderson/metroglyphs** – tamanho de raios
 - **Stick figures** – tamanho, ângulo, cor do membro
 - **Trees** – tamanho, espessura, ângulos dos ramos, etc.
 - **Autoglyph** – cor de caixas
 - **Boxes** – altura, comprimento, profundidade, etc.
 - **Hedgehogs** – picos em um campo vetorial, com variação em orientação e espessura
 - **Faces** – tamanho e posição dos olhos, nariz, boca, etc.
 - **Arrows** – tamanho, comprimento, cor

- Alguns exemplos de **glifos** são
 - **Profiles** – altura e cor de barras
 - **Stars** – tamanho de raios igualmente espaçados emanando do centro
 - **Anderson/metroglyphs** – tamanho de raios
 - **Stick figures** – tamanho, ângulo, cor do membro
 - **Trees** – tamanho, espessura, ângulos dos ramos, etc.
 - **Autoglyph** – cor de caixas
 - **Boxes** – altura, comprimento, profundidade, etc.
 - **Hedgehogs** – picos em um campo vetorial, com variação em orientação e espessura
 - **Faces** – tamanho e posição dos olhos, nariz, boca, etc.
 - **Arrows** – tamanho, comprimento, cor
 - etc.

Glifos e Ícones



PROFILE GLYPHS



**STARS AND
METROGLYPHS**



STICKS AND TREES



AUTOGLYPH/BOX GLYPH



FACE GLYPHS



ARROWS/WEATHERVANES

Figura: Exemplos de glifos multivariados.

- Na **escolha do glifo** vários aspectos devem ser considerados

- Na **escolha do glifo** vários aspectos devem ser considerados
 - **Viés imposto** dependendo do atributo gráfico usado (linha pode ser melhor que orientação)

- Na **escolha do glifo** vários aspectos devem ser considerados
 - **Viés imposto** dependendo do atributo gráfico usado (linha pode ser melhor que orientação)
 - **Glifos adjacentes** serão mais **facilmente comparados** dos que os afastados

- Na **escolha do glifo** vários aspectos devem ser considerados
 - **Viés imposto** dependendo do atributo gráfico usado (linha pode ser melhor que orientação)
 - **Glifos adjacentes** serão mais **facilmente comparados** dos que os afastados

- Uma vez escolhido o glifo, existem $N!$ **possibilidades distintas** de se mapear as dimensões às propriedades visuais do glifo
 - Qual a melhor escolha de mapeamento?

- Algumas **estratégias** para escolher o **mapeamento**

- Algumas **estratégias** para escolher o **mapeamento**
 - **Dimensões correlacionadas** podem ser mapeadas para atributos visuais **adjacentes** – ajuda a revelar padrões e *outliers*

- Algumas **estratégias** para escolher o **mapeamento**
 - **Dimensões correlacionadas** podem ser mapeadas para atributos visuais **adjacentes** – ajuda a revelar padrões e *outliers*
 - Dimensões mapeadas de forma a compor **glifos simétricos** – glifos assimétricos serão facilmente reconhecidos

- Algumas **estratégias** para escolher o **mapeamento**
 - **Dimensões correlacionadas** podem ser mapeadas para atributos visuais **adjacentes** – ajuda a revelar padrões e *outliers*
 - Dimensões mapeadas de forma a compor **glifos simétricos** – glifos assimétricos serão facilmente reconhecidos
 - Dimensões podem ser mapeadas de **acordo com seus valores** em uma 'instância de referência'

- Algumas **estratégias** para escolher o **mapeamento**
 - **Dimensões correlacionadas** podem ser mapeadas para atributos visuais **adjacentes** – ajuda a revelar padrões e *outliers*
 - Dimensões mapeadas de forma a compor **glifos simétricos** – glifos assimétricos serão facilmente reconhecidos
 - Dimensões podem ser mapeadas de **acordo com seus valores** em uma 'instância de referência'
 - Dimensões podem ser mapeadas manualmente, de acordo com o **conhecimento do usuário**

- Existem três diferentes **estratégias de *layout*** (posicionamento)

- Existem três diferentes **estratégias de layout** (posicionamento)
 - **uniforme**: glifos são escalados e posicionados com espaçamento uniforme para ocupar a tela inteira – evita sobreposições

- Existem três diferentes **estratégias de layout** (posicionamento)
 - **uniforme**: glifos são escalados e posicionados com espaçamento uniforme para ocupar a tela inteira – evita sobreposições
 - **dirigido-a-dados**: os valores de dados determinam a posição dos glifos – podem ser posicionados usando *scatterplots* ou projeções multidimensionais

- Existem três diferentes **estratégias de layout** (posicionamento)
 - **uniforme**: glifos são escalados e posicionados com espaçamento uniforme para ocupar a tela inteira – evita sobreposições
 - **dirigido-a-dados**: os valores de dados determinam a posição dos glifos – podem ser posicionados usando *scatterplots* ou projeções multidimensionais
 - **dirigido-a-estrutura**: se existir alguma estrutura implícita, como cíclica ou hierárquica, essa pode ser usada para posicionar os glifos

Glifos e Ícones

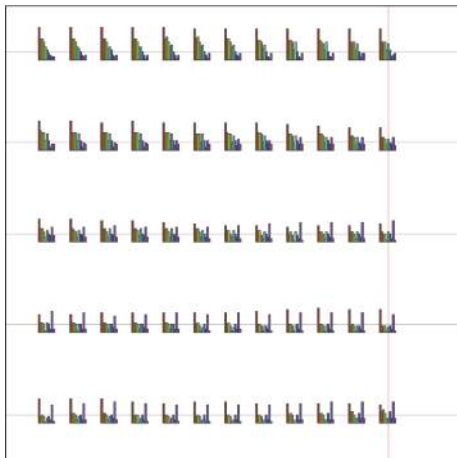


Figura: Exemplos de posicionamento de glifos uniforme.

Glifos e Ícones

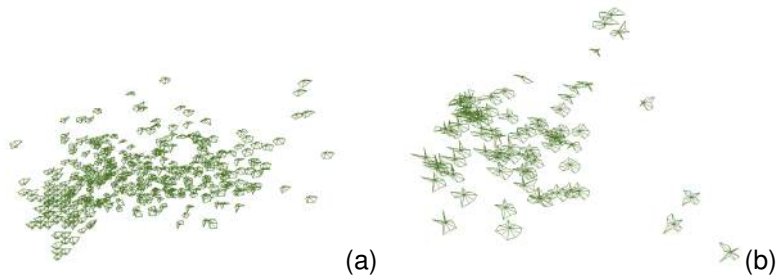


Figura: Exemplos de posicionamento de glifos dirigido-a-dados. (a) usando dimensões (b) usando PCA.

Glifos e Ícones

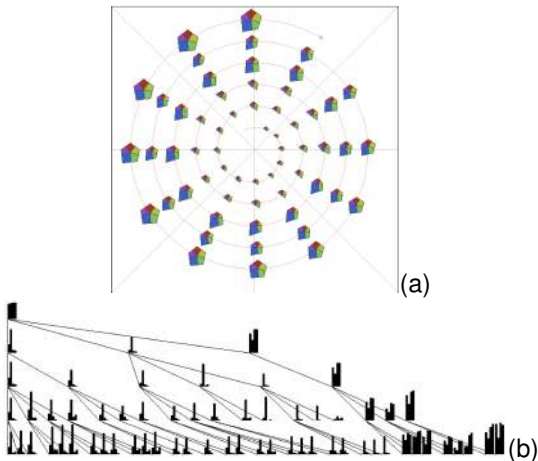


Figura: Exemplos de posicionamento de glifos dirigido-a-estrutura. (a) estrutura em espiral (b) estrutura hierárquica.

Display Denso de Pixels

- **Display Denso de Pixels** (técnicas orientadas-a-pixel) mapeiam cada **valor** para **pixels individuais** e criam um polígono preenchido para representar cada dimensão dos dados
 - Fazem uso **completo** do espaço de tela
 - Cada **valor** de dado define a **cor do pixel**

Display Denso de Pixels

- Na forma mais simples, cada **dimensão** irá gerar uma **imagem no display**

Display Denso de Pixels

- Na forma mais simples, cada **dimensão** irá gerar uma **imagem no display**
- Os elementos precisam ser desenhados de forma que **elementos próximos fiquem próximos**

Display Denso de Pixels

- Na forma mais simples, cada **dimensão** irá gerar uma **imagem no display**
- Os elementos precisam ser desenhados de forma que **elementos próximos fiquem próximos**
 - Pode-se criar uma imagem fazendo uma travessia da esquerda-para-direita da **direita-para-esquerda**

Display Denso de Pixels

- Na forma mais simples, cada **dimensão** irá gerar uma **imagem no display**
- Os elementos precisam ser desenhados de forma que **elementos próximos fiquem próximos**
 - Pode-se criar uma imagem fazendo uma travessia da esquerda-para-direita da **direita-para-esquerda**
 - Pode-se usar uma travessia em **espiral**

Display Denso de Pixels

- Na forma mais simples, cada **dimensão** irá gerar uma **imagem no display**
- Os elementos precisam ser desenhados de forma que **elementos próximos fiquem próximos**
 - Pode-se criar uma imagem fazendo uma travessia da esquerda-para-direita da **direita-para-esquerda**
 - Pode-se usar uma travessia em **espiral**
 - Pode-se usar uma **curva recursiva** para preenchimento do *layout* (pontos próximos na lista ficam próximos no display)

Display Denso de Pixels

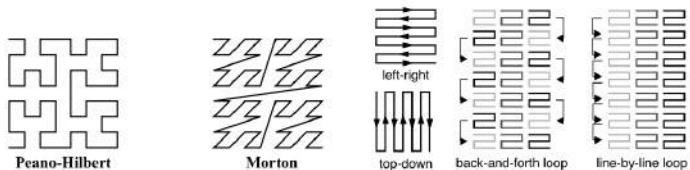
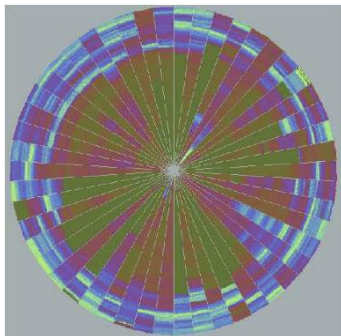
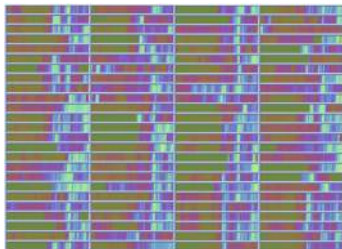


Figura: Exemplos de padrões de layout para displays densos de pixels.

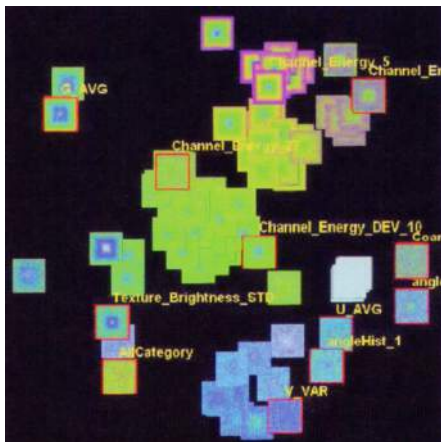
Display Denso de Pixels

- As **imagens** de cada dimensão podem ser **posicionadas na tela** de várias formas, como em um **grade** ou em **segmentos de círculo**



Display Denso de Pixels

- Outra possibilidade é **usar algum método de projeção** para posicionar as dimensões similares proximamente



Display Denso de Pixels

- A **ordenação** dos dados (quando possível) **pode ajudar** a identificação de padrões interessantes

Display Denso de Pixels

- A **ordenação** dos dados (quando possível) **pode ajudar** a identificação de padrões interessantes
 - **Ordenar** com base em uma **dimensão** (atributo) ajudar a revelar agrupamentos dentro dessa dimensão

Display Denso de Pixels

- A **ordenação** dos dados (quando possível) **pode ajudar** a identificação de padrões interessantes
 - **Ordenar** com base em uma **dimensão** (atributo) ajudar a revelar agrupamentos dentro dessa dimensão
 - **Ordenar** com base na **distância** a um vetor n -dimensional ajuda a revelar agrupamentos em várias dimensões

Display Denso de Pixels

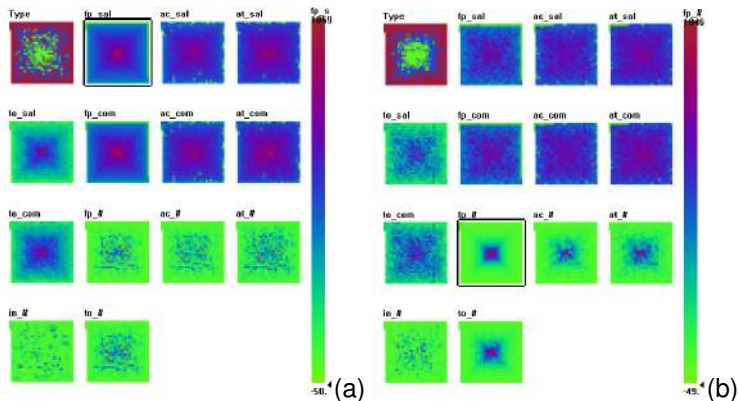


Figura: Exemplos de diferentes ordenações. (a) dados ordenados com base no salário dos professores. Existe pouca correlação com outros atributos. (b) dados ordenados com base em número de “full professors”. Somente relacionado com o número total de empregados.

Display Denso de Pixels

- Outra opção consiste em posicionar os pixels sobre gráficos de barras padrão

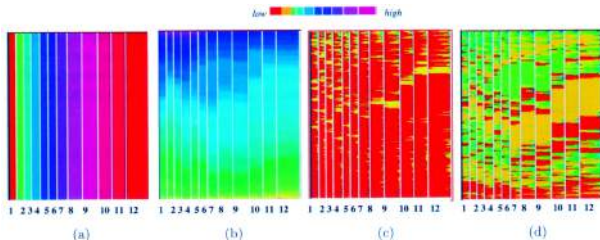


Figura: Exemplo de visualização de atividades de compra online por meio de 'pixel bar charts'. (a) cor representa o mês (b) cor representa o valor da compra (c) cor representa o número de visitas (d) cor representa a quantidade comprada.

- 1 Introdução
- 2 Técnicas baseadas em Ponto
 - Scatterplots
 - Projeção Multidimensional
 - Outras Técnicas
- 3 Técnicas baseadas em Linhas
- 4 Técnicas baseadas em Região
- 5 Combinação de Técnicas
- 6 Referências

- Ward, M., Grinstein, G. G., Keim, D. **Interactive data visualization foundations, techniques, and applications.** Natick, Mass., A K Peters, 2010.



Borg, I. and Groenen, P. J. F. (2005).

Modern Multidimensional Scaling: Theory and Applications.
Springer, second edition.



Jolliffe, I. T. (1986).

Principal Component Analysis.
Springer-Verlag.



Paulovich, F. V., Nonato, L. G., Minghim, R., and Levkowitz, H. (2008).

Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping.
IEEE Trans. Visual. Comp. Graph., 14(3):564–575.



Pekalska, E., de Ridder, D., Duin, R. P. W., and Kraaijveld, M. A. (1999).

A new method of generalizing Sammon mapping with application to algorithm speed-up.
In Boasson, M., Kaandorp, J. A., Tonino, J. F. M., and Vosselman, M. G., editors, *Annual Conf. Advanced School for Comput. Imag.*, pages 221–228.

Referências II



Sammon, J. W. (1969).

A nonlinear mapping for data structure analysis.

IEEE Trans. Comput., 18(5):401–409.



Tejada, E., Minghim, R., and Nonato, L. G. (2003).

On improved projection techniques to support visual exploration of multidimensional data sets.

Information Visualization, 2(4):218–231.