

CIÊNCIA DE DADOS - o6

Prof. Júlio Cesar Nievola

PPGla – PUCPR

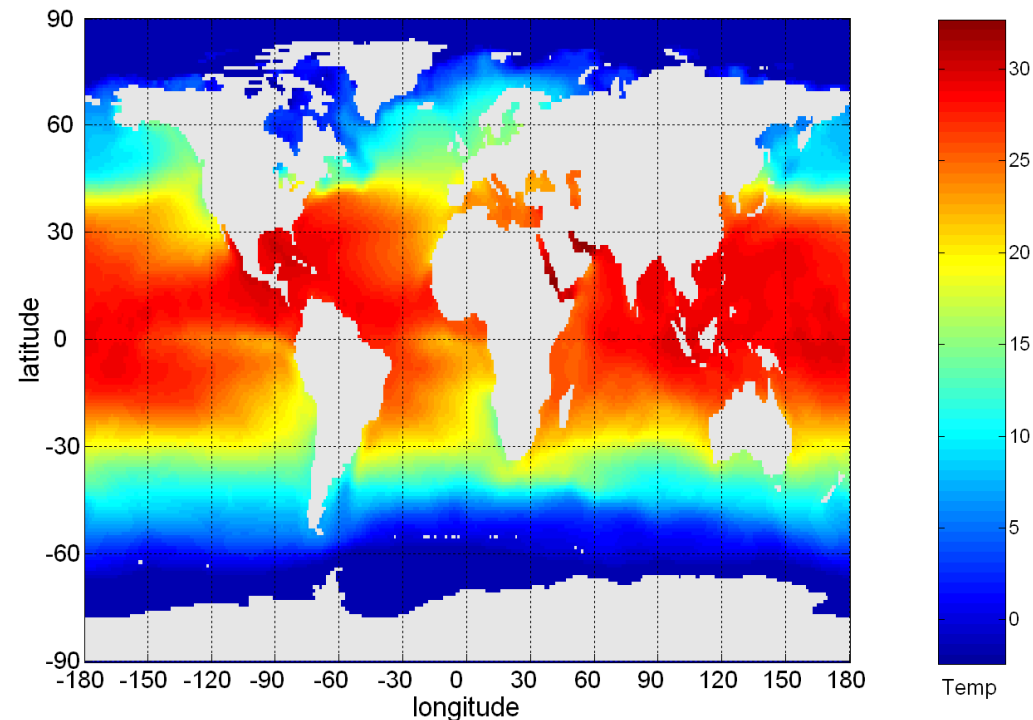
29/junho/2019

Visualização

- Visualização é a conversão dos dados para um formato visual ou tabular de tal forma que as características dos dados e os relacionamentos entre itens de dados ou atributos possa ser analisada ou reportada.
- Visualização de dados é uma das técnicas de maior apelo e poder para exploração de dados.
 - Os seres humanos tem uma habilidade bem desenvolvida de analisar grandes quantidades de informação que seja apresentada visualmente
 - Pode detectar padrões gerais e tendências
 - Pode detectar outliers e padrões não usuais

Temperatura da Superfície do Mar

- Abaixo tem-se a Temperatura da Superfície do Mar em Julho de 1982
 - Dezenas de milhares de pontos de dados estão sumarizados em uma única figura



Representação

- É o mapeamento da informação em um formato visual
- Objetos de dados, seus atributos e as relações entre objetos de dados são traduzidos em elementos gráficos tais como pontos, linhas, formatos e cores.
- Exemplo:
 - Objetos são representados frequentemente como pontos
 - Seus valores de atributo podem ser representados como a posição dos pontos ou as características dos pontos, e.g., cor, tamanho e formato
 - Se a posição é usada, então os relacionamentos entre os pontos, i.e., se eles formam grupos ou um ponto é um outlier, são facilmente percebidos.

Arranjo

- É a colocação de elementos visuais dentro de um display
- Pode fazer uma grande diferença no quão fácil é para compreender os dados
- Exemplo:

	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	1	0	0	1
3	0	1	0	1	1	0
4	1	0	1	0	0	1
5	0	1	0	1	1	0
6	1	0	1	0	0	1
7	0	1	0	1	1	0
8	1	0	1	0	0	1
9	0	1	0	1	1	0

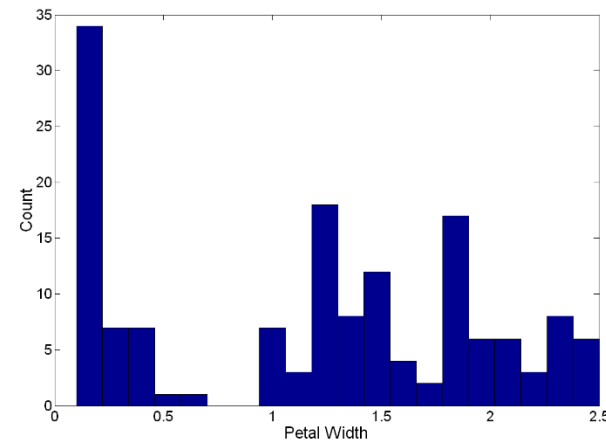
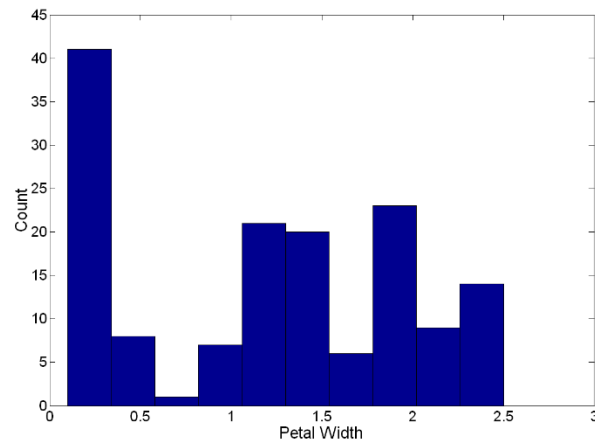
	6	1	3	2	5	4
4	1	1	1	0	0	0
2	1	1	1	0	0	0
6	1	1	1	0	0	0
8	1	1	1	0	0	0
5	0	0	0	1	1	1
3	0	0	0	1	1	1
9	0	0	0	1	1	1
1	0	0	0	1	1	1
7	0	0	0	1	1	1

Seleção

- É a eliminação ou a tirada de ênfase de certos objetos e atributos
- Seleção pode envolver a escolha de um subconjunto de atributos
 - Redução de dimensionalidade é muito usada para reduzir o número de dimensões para duas ou três
 - De outra forma, pares de atributos podem ser usados
- Seleção também pode envolver escolher um subconjunto de objetos
 - Uma região da tela mostra um número fixo de pontos
 - Pode-se amostrar, mas querer preservar pontos em áreas esparsas

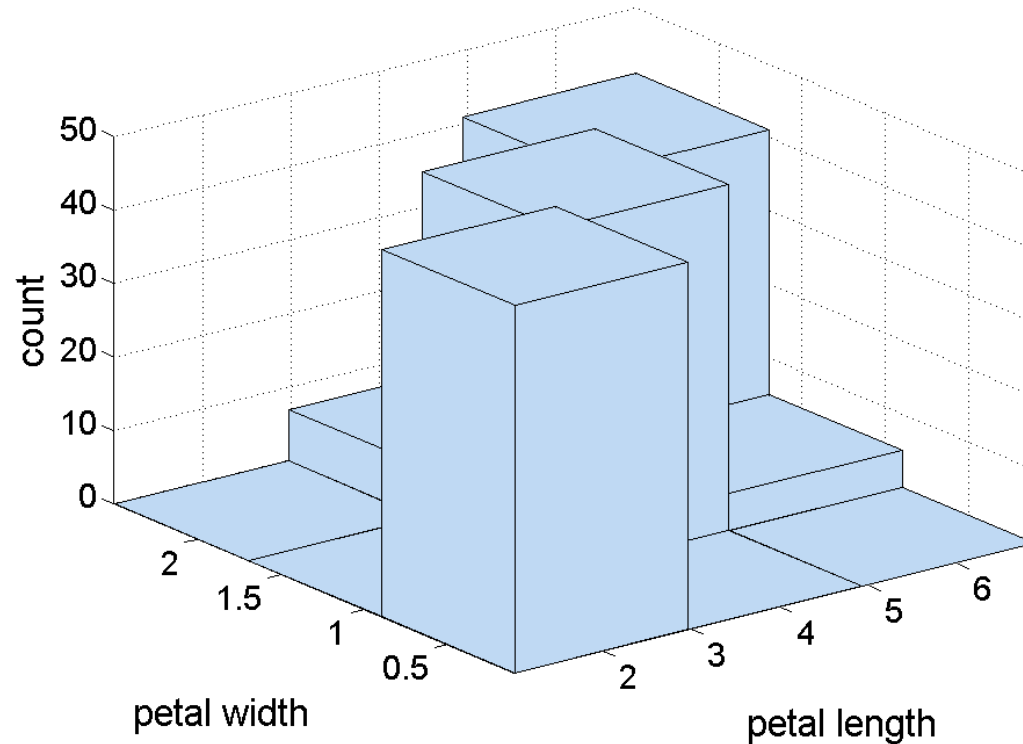
Técnicas de Visualização: Histogramas

- Histograma
 - Usualmente mostra distribuição de valores de uma variável
 - Divide os valores em faixas e mostra um gráfico de barras do número de objetos em cada faixa.
 - A altura de cada barra indica o número de objetos
 - Formato do histograma depende do número de faixas
- Exemplo: Petal Width (10 e 20 faixas, respectivamente)



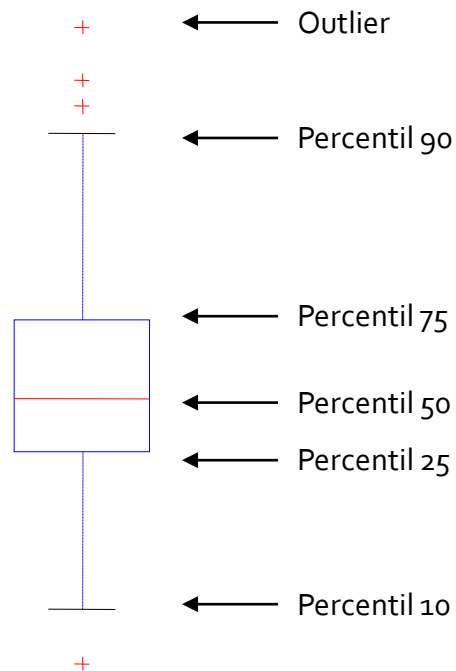
Histogramas Bi-Dimensionais

- Mostra a distribuição conjunta dos valores de dois atributos
- Exemplo: petal width e petal length
 - O que isto nos mostra?



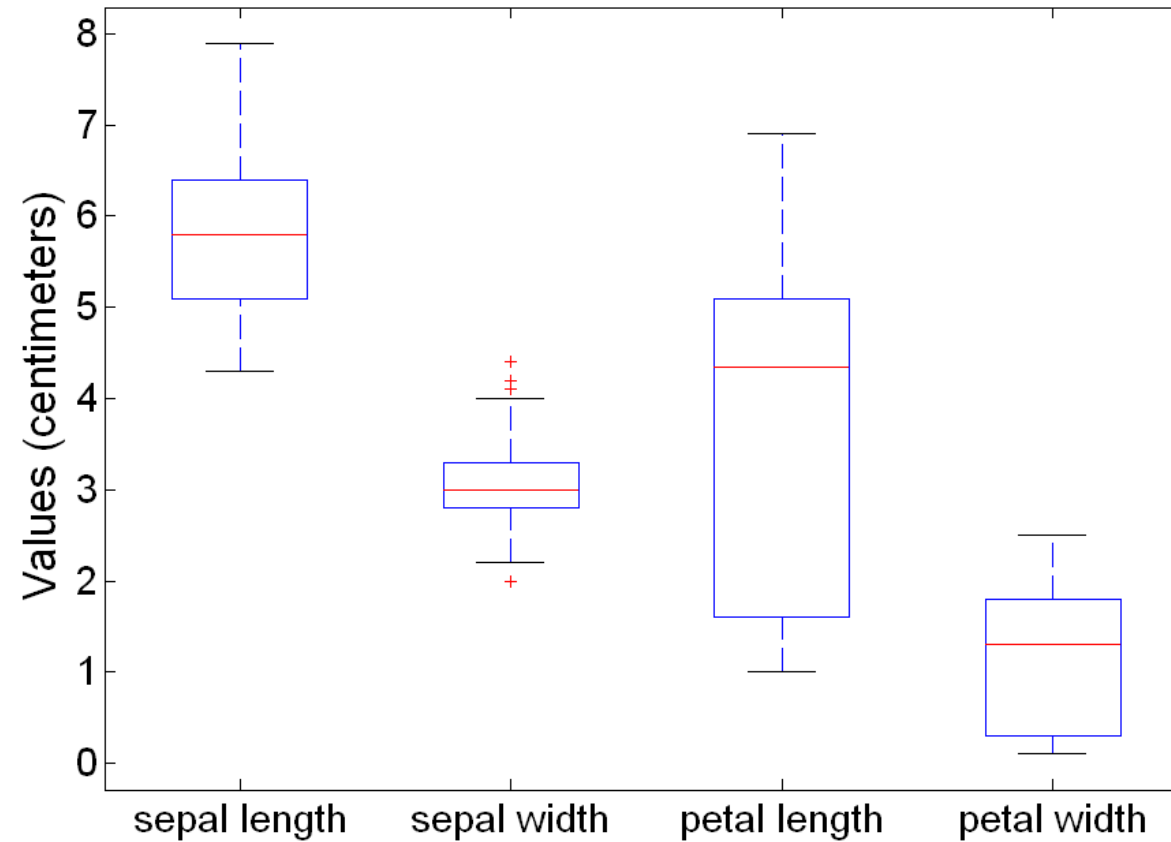
Técnicas de Visualização: Gráficos de Caixa

- Gráficos de Caixa
 - Inventados por J. Tukey
 - Outra forma de indicar a distribuição dos dados
 - Figura mostra a parte básica de um gráfico de caixa



Exemplo de Gráfico de Caixa

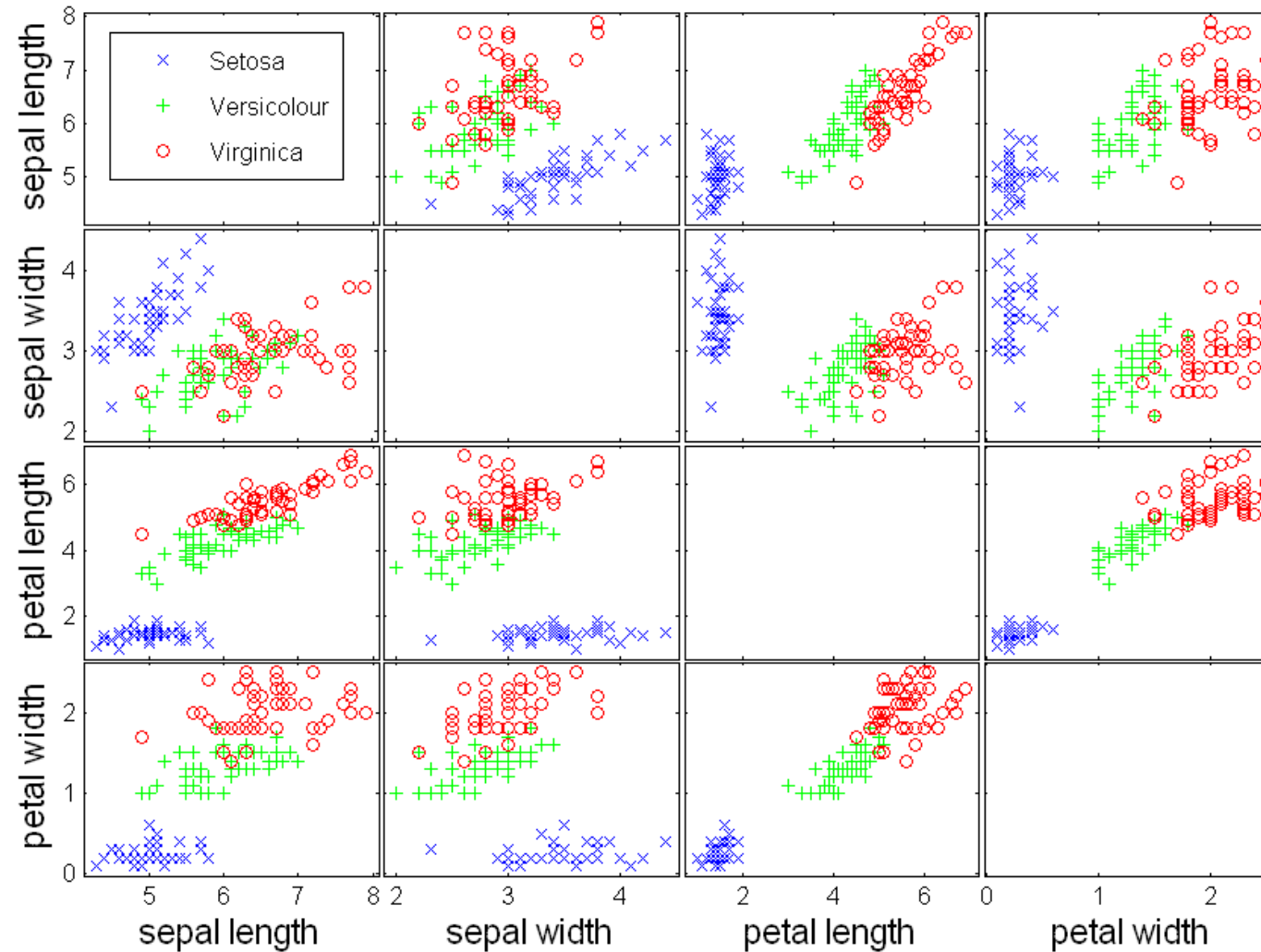
- Gráficos de Caixa podem ser usados para comparar atributos



Técnicas de Visualização: Gráficos de Dispersão

- Gráficos de Dispersão
 - Valores dos atributos determinam a posição
 - Gráficos de dispersão bidimensionais são mais comuns, mas também há gráficos tridimensionais
 - Frequentemente atributos adicionais podem ser mostrados usando tamanho, forma e cor dos marcadores que representam os objetos
 - É útil ter arranjos de gráficos de dispersão para sumarizar de maneira compacta os relacionamentos de vários pares de atributos
 - Exemplo a seguir

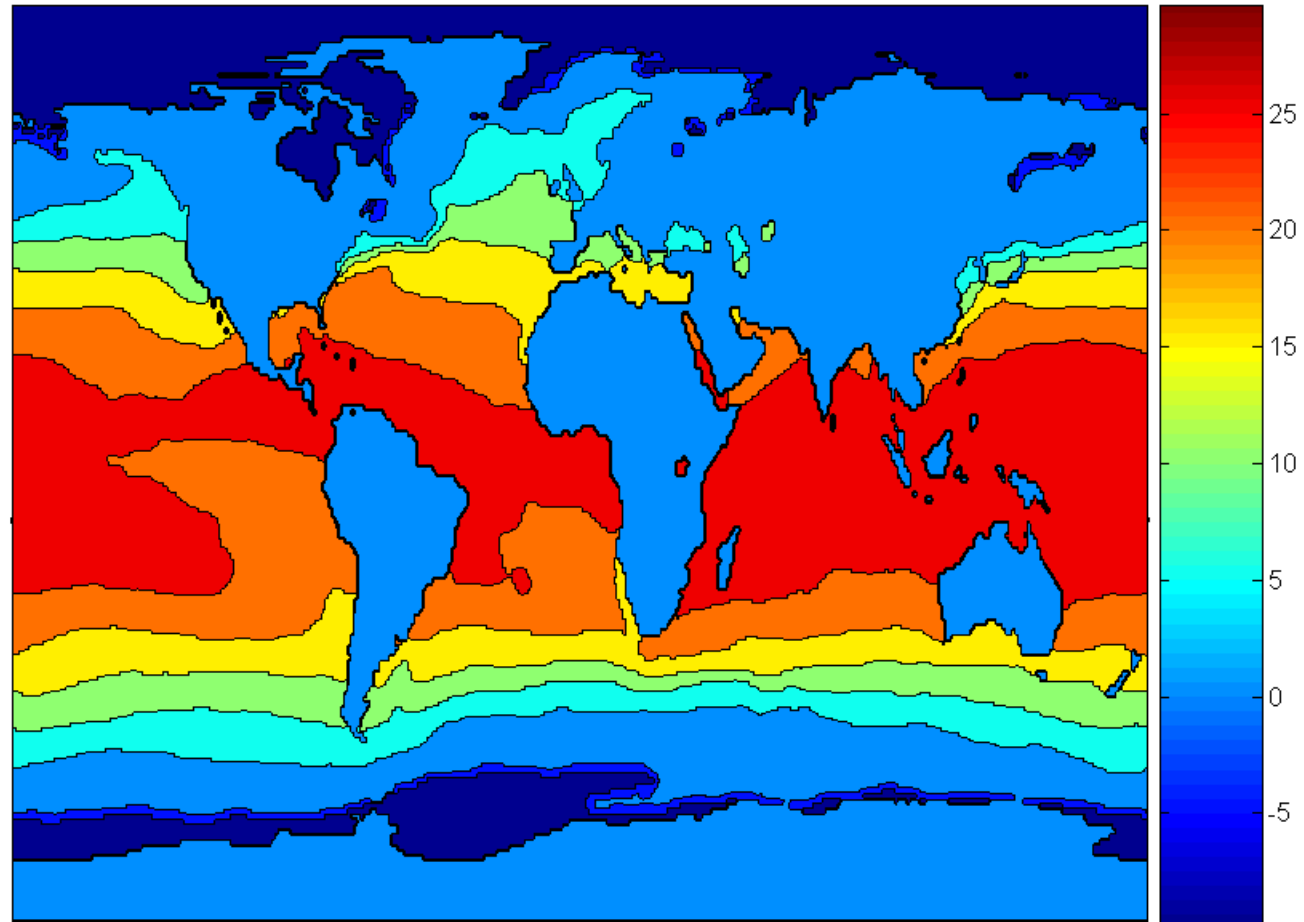
Arranjo de Gráficos de Dispersão para Atributos Iris



Técnicas de Visualização: Gráficos de Contorno

- Gráficos de Contorno
 - Útil quando um atributo contínuo é medido em uma grade espacial
 - Particionam o plano em regiões de valores similares
 - Linhas de contorno que formam os limites destas regiões conectam pontos com valores iguais
 - O exemplo mais comum são os mapas de contorno de elevação
 - Também pode indicar temperatura, precipitação, pressão do ar, etc.
 - Exemplo para Temperatura da Superfície do mar a seguir

Exemplo de Gráfico de Contorno: SST Dec 1998

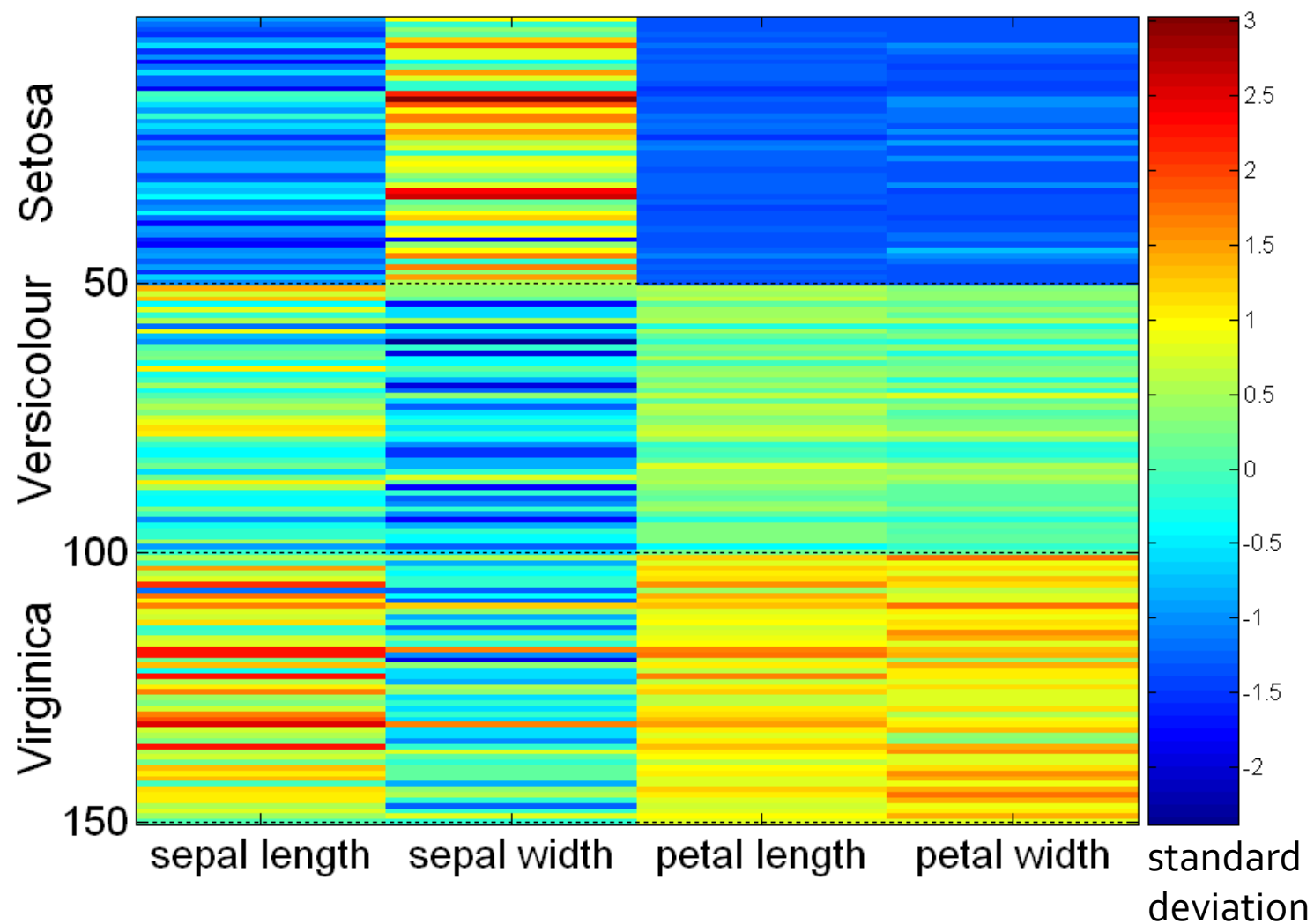


Celsius

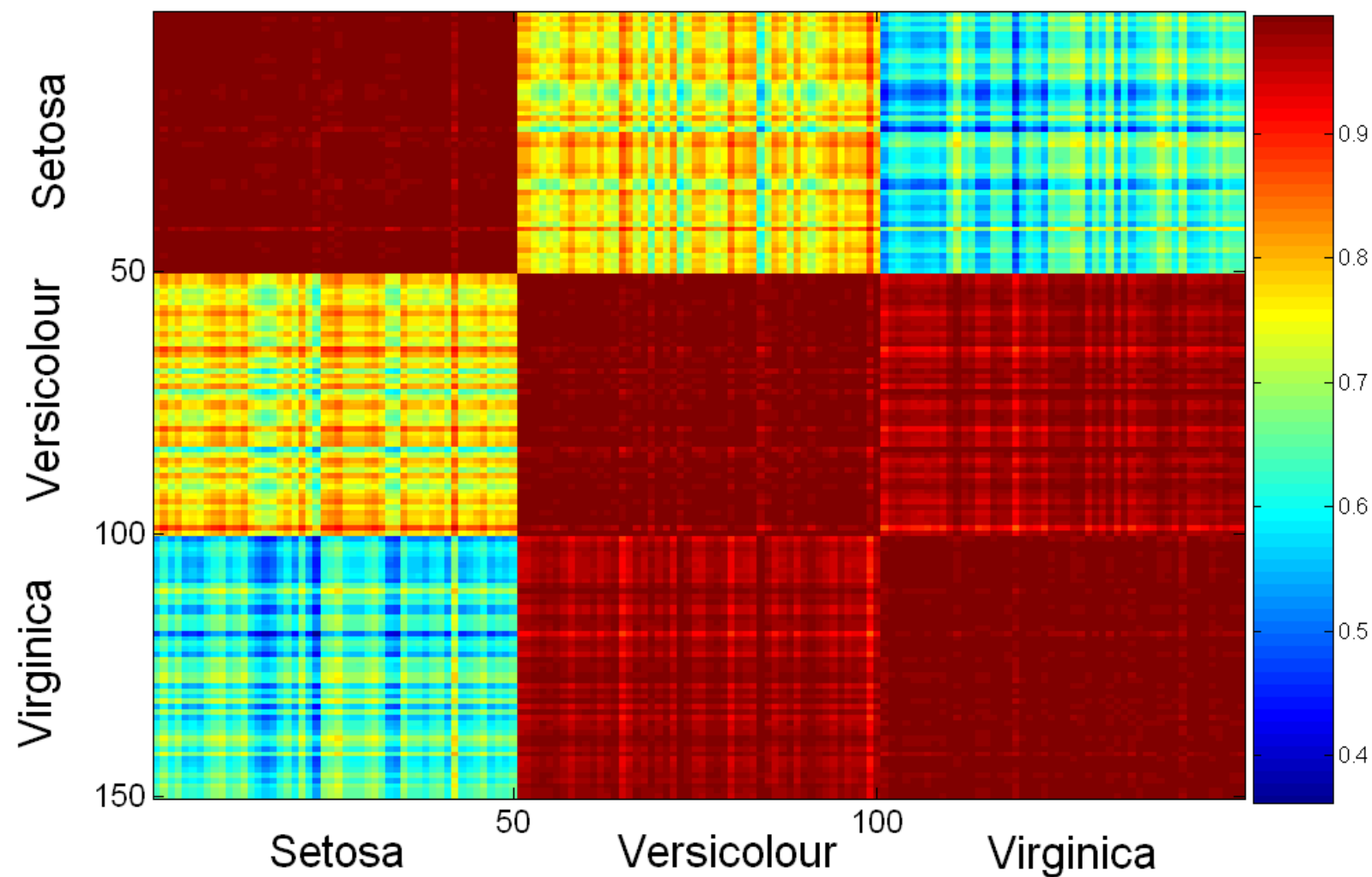
Técnicas de Visualização: Gráficos Matriciais

- Gráficos Matriciais
 - Podem plotar a matriz de dados
 - Pode ser útil quando os objetos são ordenados de acordo com a classe
 - Normalmente os atributos são normalizados para evitar que um atributo domine o gráfico
 - Gráficos de similaridade ou matrizes de distância também podem ser úteis para visualizar os relacionamentos entre objetos
 - Exemplos de gráficos matriciais estão a seguir

Visualização da Matriz de Dados Iris



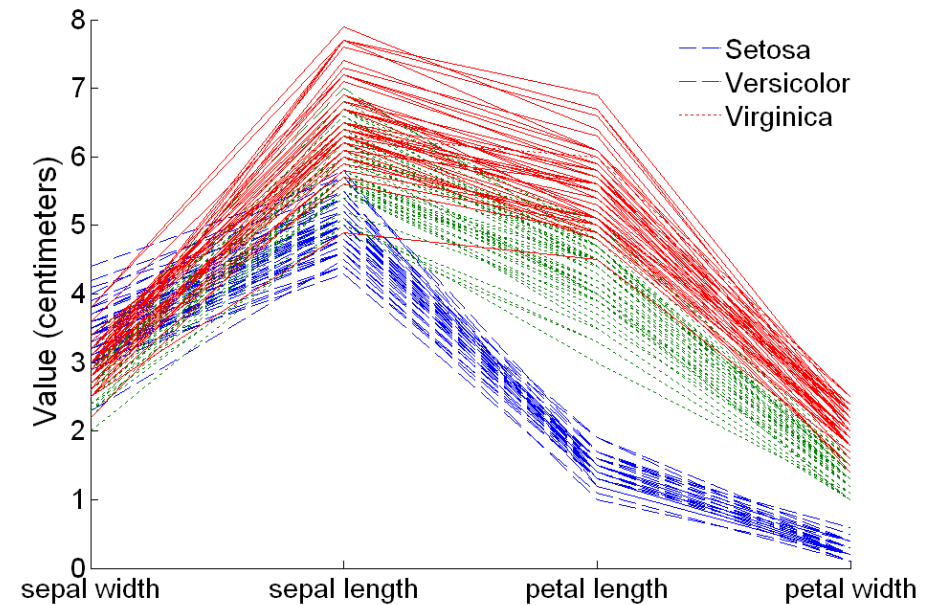
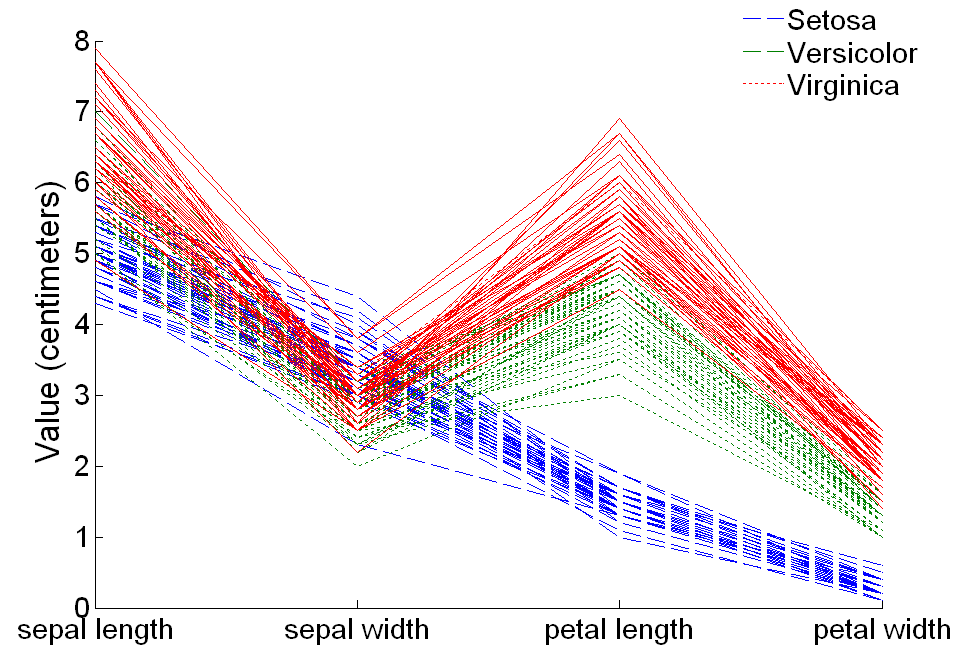
Visualização da Matriz de Correlação Iris



Técnicas de Visualização: Coordenadas Paralelas

- Coordenadas Paralelas
 - Usadas para plotar os valores dos atributos de dados de alta dimensionalidade
 - Em lugar de eixos perpendiculares, usa-se um conjunto de eixos paralelos
 - Valores dos atributos de cada objeto são plotados como um ponto em cada um dos eixos coordenados correspondentes e os pontos são ligados por linhas
 - Então, cada objeto é representado como uma linha
 - Frequentemente linhas representam uma classe distinta de objetos agrupados, ao menos para alguns atributos
 - Ordenar atributos é importante para ver os grupos

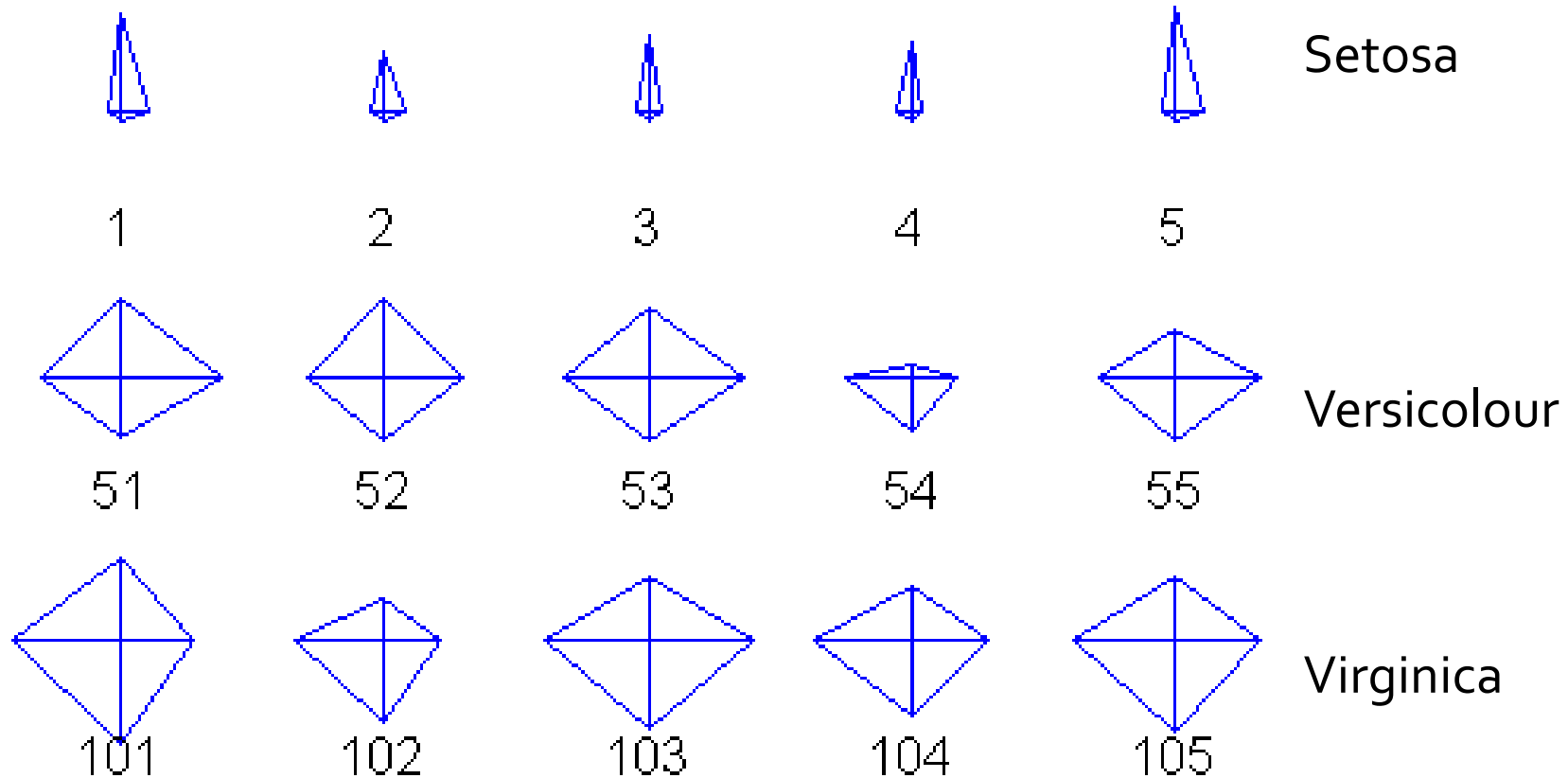
Gráfico de Coordenadas Paralelas para Iris



Outras Técnicas de Visualização

- Gráficos Estrela
 - Abordagem similar a coordenadas paralelas, mas eixos irradiam a partir de um ponto central
 - A linha conectando os valores de um objeto é um polígono
- Faces de Chernoff
 - Abordagem criada por Herman Chernoff
 - Esta abordagem associa cada atributo com a característica de uma face
 - Valores de cada atributo determinam a aparência da característica facial correspondente
 - Cada objeto torna-se uma face separada
 - Baseia-se na habilidade humana de distinguir faces

Gráficos Estrela para Iris



Faces de Chernoff para Iris



1



2



3



4



5

Setosa



51



52



53



54



55

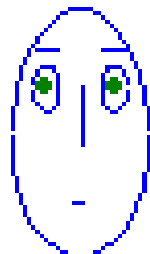
Versicolour



101



102



103



104



105

Virginica