

FCT/Unesp – Presidente Prudente
Departamento de Matemática e Computação

Visualização de Informação: Introdução e Aplicações em Análise de Dados

Prof. Danilo Medeiros Eler
danilo.eler@unesp.br

Sumário

- O que é Visualização?
- Porque Estudar Visualização?
- Exemplos Introdutórios de Aplicação em Análise de Dados

O que é Visualização?

- Do dicionário Michaelis (Moderno Dicionário da Língua Portuguesa, Editora Melhoramentos, 1998, São Paulo):
 - Visualizar
 - Tornar visual ou visível
 - Ver uma imagem mental; figurar mentalmente
 - Visualização
 - Transformação de conceitos abstratos em imagens reais ou mentalmente visíveis
 - Conversão de números ou dados para um formato gráfico, que pode ser facilmente entendido

O que é Visualização?

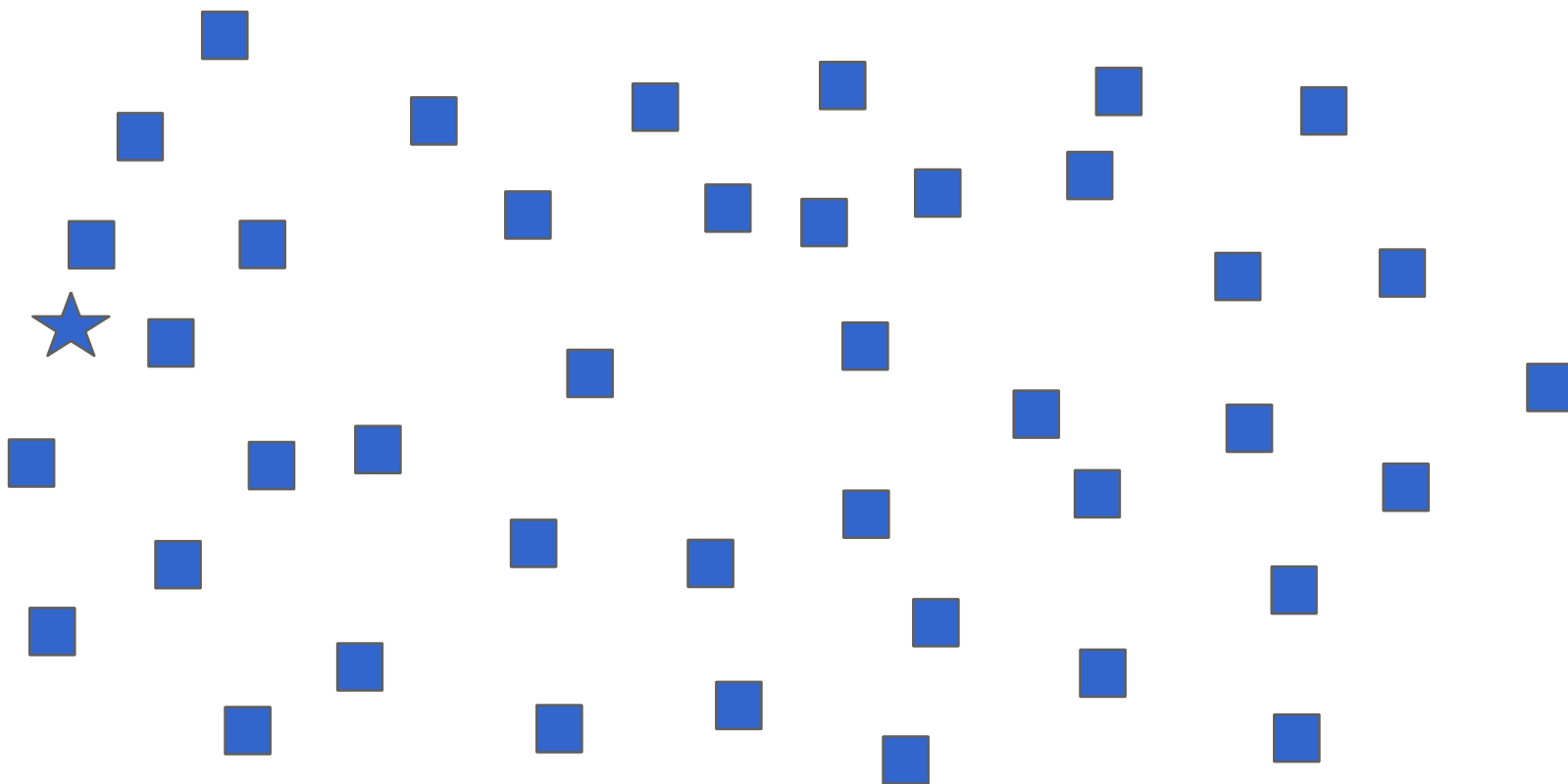
- Visualização não está relacionada ao computador
- Ben Shneiderman (1999)
 - O propósito da visualização é o *insight*, não as imagens
 - Principais objetivos desse *insight*
 - Realizar descobertas
 - Verificação de hipóteses
 - Tomada de decisões
 - Explicação de questões concretas

O que é Visualização?

- A Visualização está relacionada com a cognição do ser humano
- Visualizar é algo que fazemos naturalmente
- O sistema visual humano é:
 - ❑ O sentido com maior capacidade de captação de informações por unidade de tempo
 - ❑ Rápido e paralelo
 - ❑ Treinado para reconhecer padrões

O que é Visualização?

Onde está a estrela?



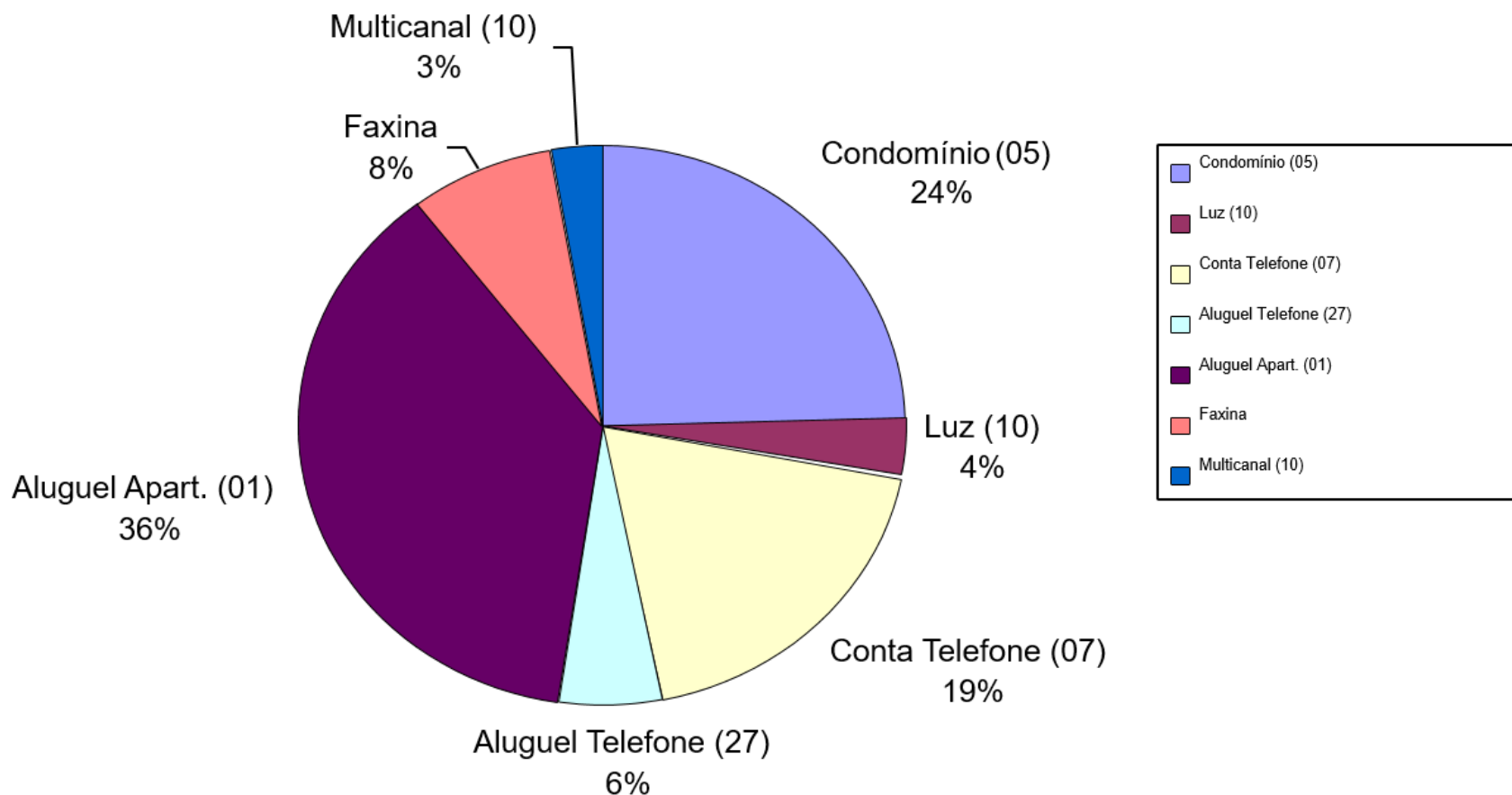
Exemplos

Despesas de casa

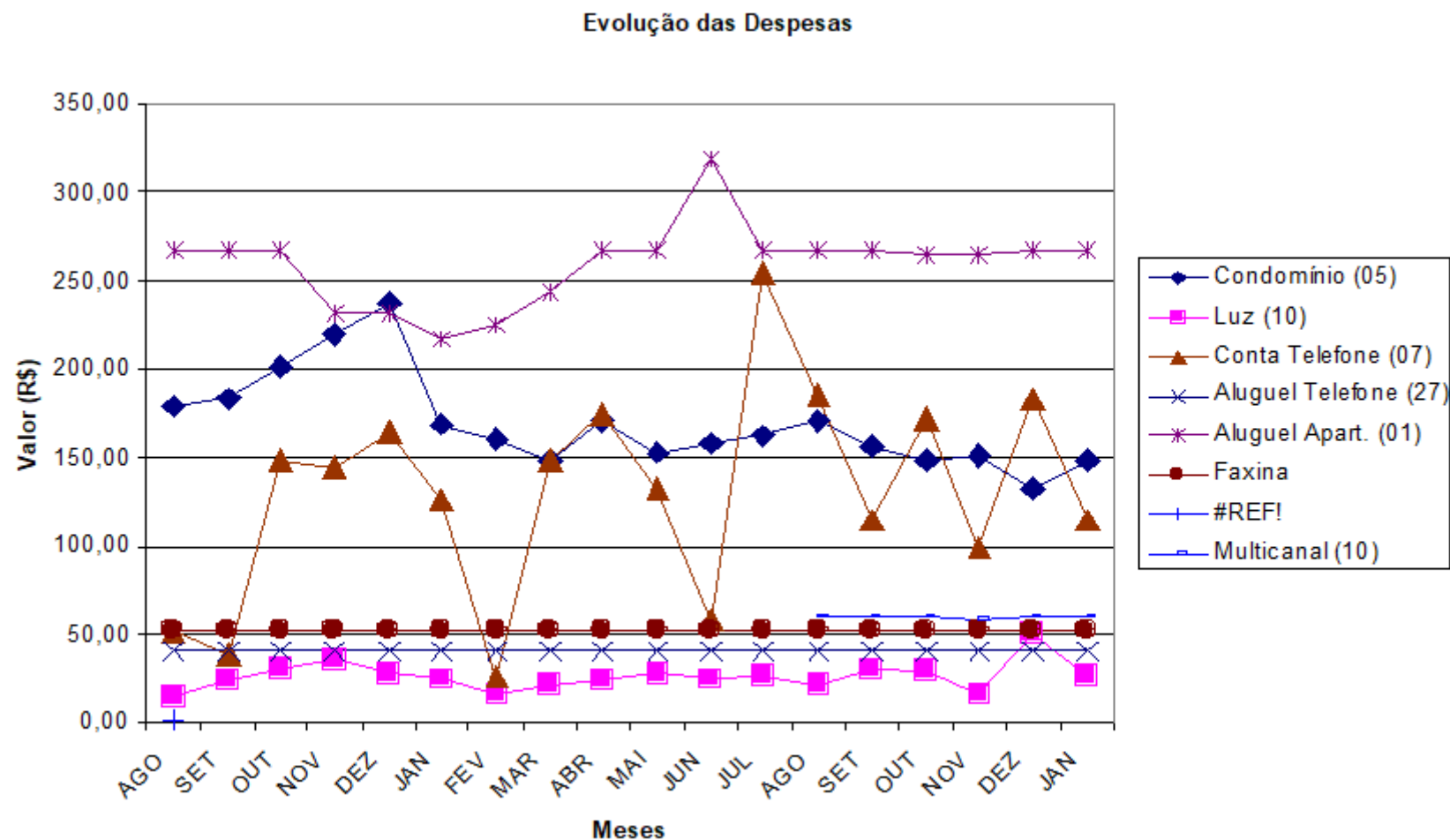
5	Contas	AGO	SET	OUT	NOV	DEZ	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ	JAN
6	Condomínio (05)	179,61	183,81	201,21	219,73	238,10	168,90	160,10	148,00	170,35	152,55	157,70	162,25	171,25	155,85	148,90	150,35	132,20	148,32
7	Luz (10)	14,58	23,50	30,24	35,94	27,30	24,19	15,89	21,60	23,84	27,13	24,19	26,09	21,25	29,55	28,68	15,38	49,77	26,44
8	Conta Telefone (07)	51,40	38,35	149,00	143,95	164,10	126,68	25,49	148,88	174,76	132,51	56,90	254,52	185,74	114,42	171,74	98,16	183,39	114,57
9	Aluguel Telefone (27)	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00
10	Aluguel Apart. (01)	267,08	267,08	267,08	232,08	232,08	217,08	225,00	243,55	267,08	267,08	319,00	267,08	267,08	267,08	265,00	265,00	267,08	267,08
11	Faxina	52,40	52,40	52,40	52,40	52,40	52,40	52,40	52,40	52,40	52,40	52,40	52,40	52,40	52,40	52,40	52,40	52,40	52,40
12	Multicanal (10)													59,90	59,90	59,90	57,90	59,90	59,90
13	Total	605,07	605,14	739,93	724,10	753,98	629,25	518,88	654,43	728,43	671,67	650,19	802,34	797,62	719,20	766,62	679,19	784,74	708,71

Exemplos

Despesas de Casa



Exemplos



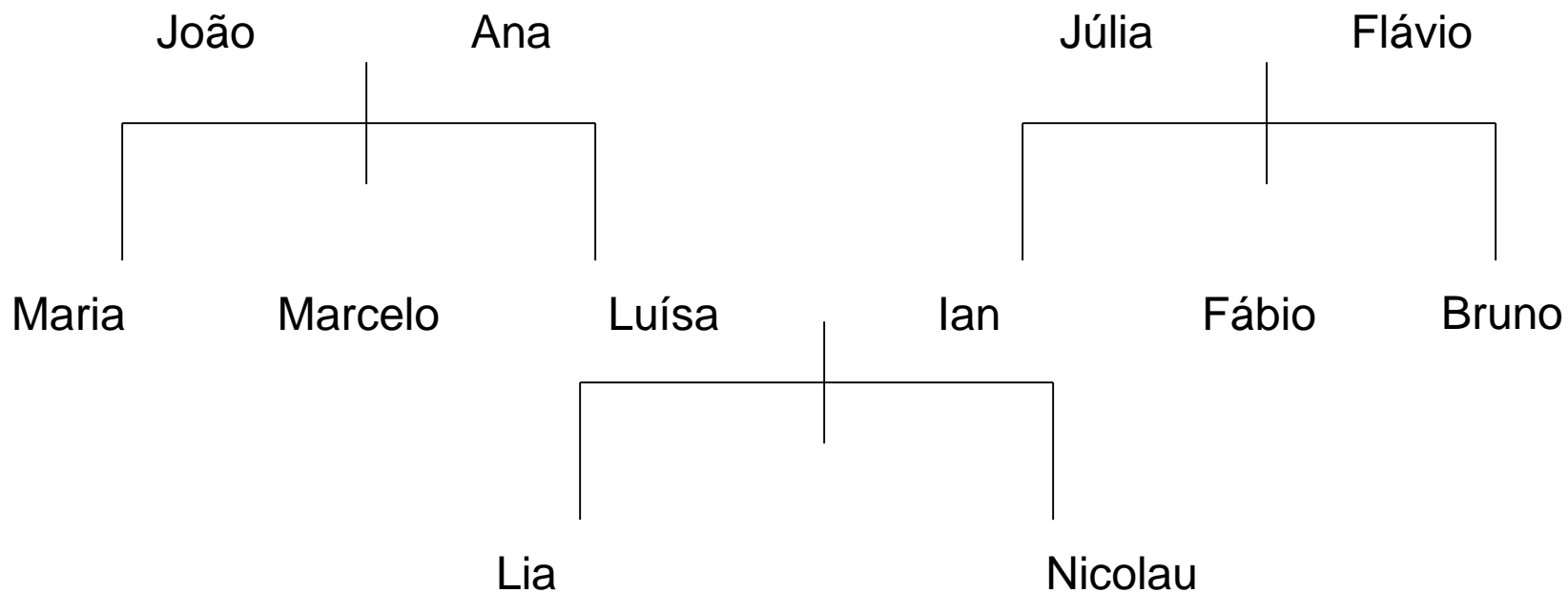
Exemplos

Pessoa1	Pai-P1	Mãe-P1	Pessoa2	Pai-P2	Mãe-P2
Maria	João	Ana	Marcelo	João	Ana
Marcelo	João	Ana	Luísa	João	Ana
Luísa	João	Ana	Maria	João	Ana
Ian	Júlia	Flávio	Fábio	Júlia	Flávio
Fábio	Júlia	Flávio	Bruno	Júlia	Flávio
Bruno	Júlia	Flávio	Ian	Júlia	Flávio
Lia	Luísa	Ian	Nicolau	Luísa	Ian

Qual o nome dos avós de Lia?

Exemplos

Exemplo - Árvore Genealógica

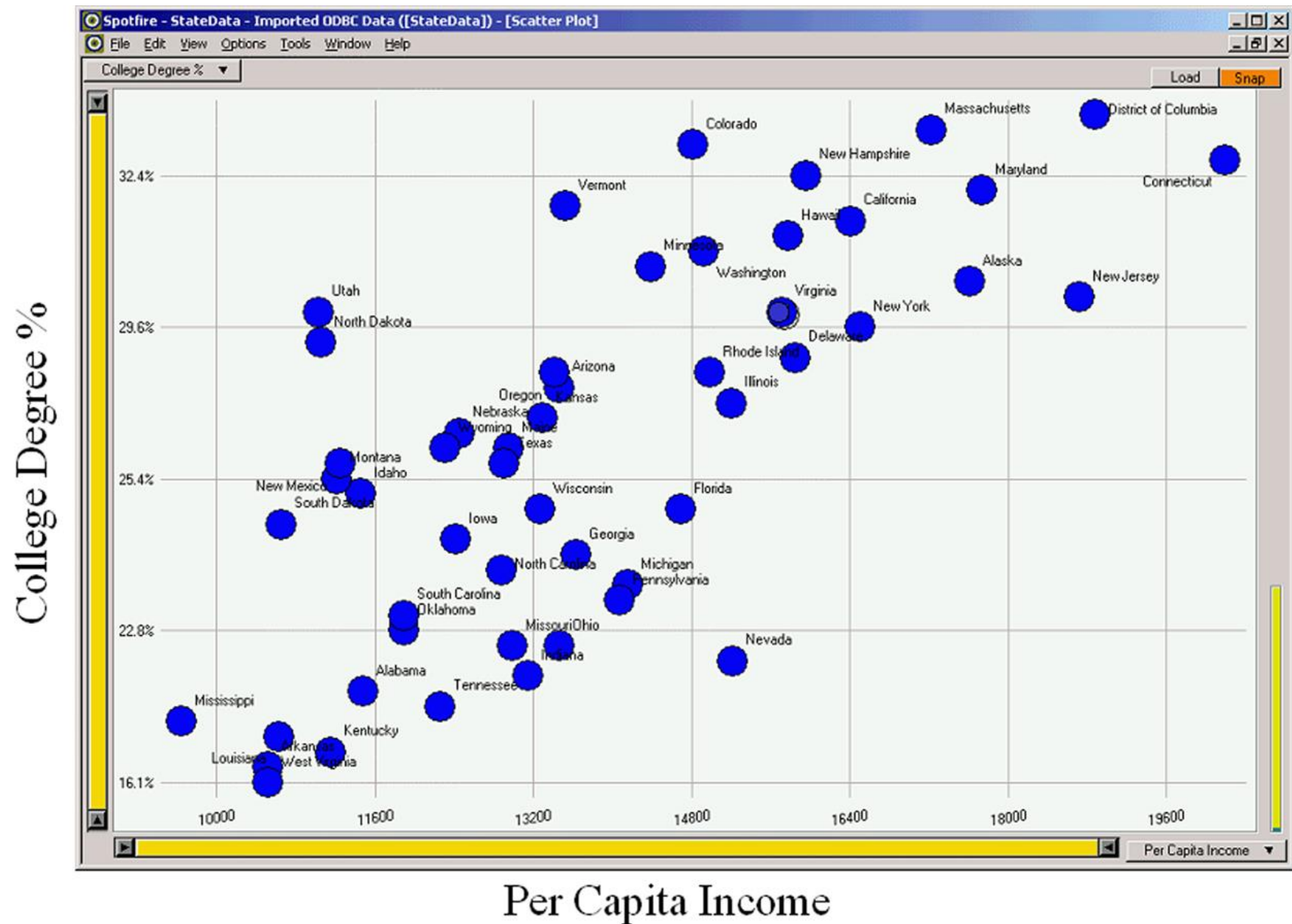


Exemplos

- Qual é o estado com maior renda?
- Há alguma relação entre a renda e a escolaridade?
- Existem *outliers* (discrepantes, exceções)?

Table - StateData ()			Load	Snap	
State	College Degree %	Per Capita Income			
Alabama	20.6%	11486	Minnesota	30.4%	14389
Alaska	30.3%	17610	Mississippi	19.9%	9648
Arizona	27.1%	13461	Missouri	22.3%	12989
Arkansas	17.0%	10520	Montana	25.4%	11213
California	31.3%	16409	Nebraska	26.0%	12452
Colorado	33.9%	14821	Nevada	21.5%	15214
Connecticut	33.8%	20189	New Hampshire	32.4%	15959
Delaware	27.9%	15854	New Jersey	30.1%	18714
District of Columbia	36.4%	18881	New Mexico	25.5%	11246
Florida	24.9%	14698	New York	29.6%	16501
Georgia	24.3%	13631	North Carolina	24.2%	12885
Hawaii	31.2%	15770	North Dakota	28.1%	11051
Idaho	25.2%	11457	Ohio	22.3%	13461
Illinois	26.8%	15201	Oklahoma	22.8%	11893
Indiana	20.9%	13149	Oregon	27.5%	13418
Iowa	24.5%	12422	Pennsylvania	23.2%	14068
Kansas	26.5%	13300	Rhode Island	27.5%	14981
Kentucky	17.7%	11153	South Carolina	23.0%	11897
Louisiana	19.4%	10635	South Dakota	24.6%	10661
Maine	25.7%	12957	Tennessee	20.1%	12255
Maryland	31.7%	17730	Texas	25.5%	12904
Massachusetts	34.5%	17224	Utah	30.0%	11029
Michigan	24.1%	14154	Vermont	31.5%	13527
Minnesota	30.4%	14389	Virginia	30.0%	15713
			Washington	30.9%	14923
			West Virginia	16.1%	10520
			Wisconsin	24.9%	13276
			Wyoming	25.7%	12311

Exemplos



O que é Visualização?

- É a representação de um domínio utilizando gráficos, imagens, animações, interações para apresentar dados, estruturas e o comportamento de um conjunto de dados
- Técnicas são utilizadas para compreender os dados e extrair conhecimento

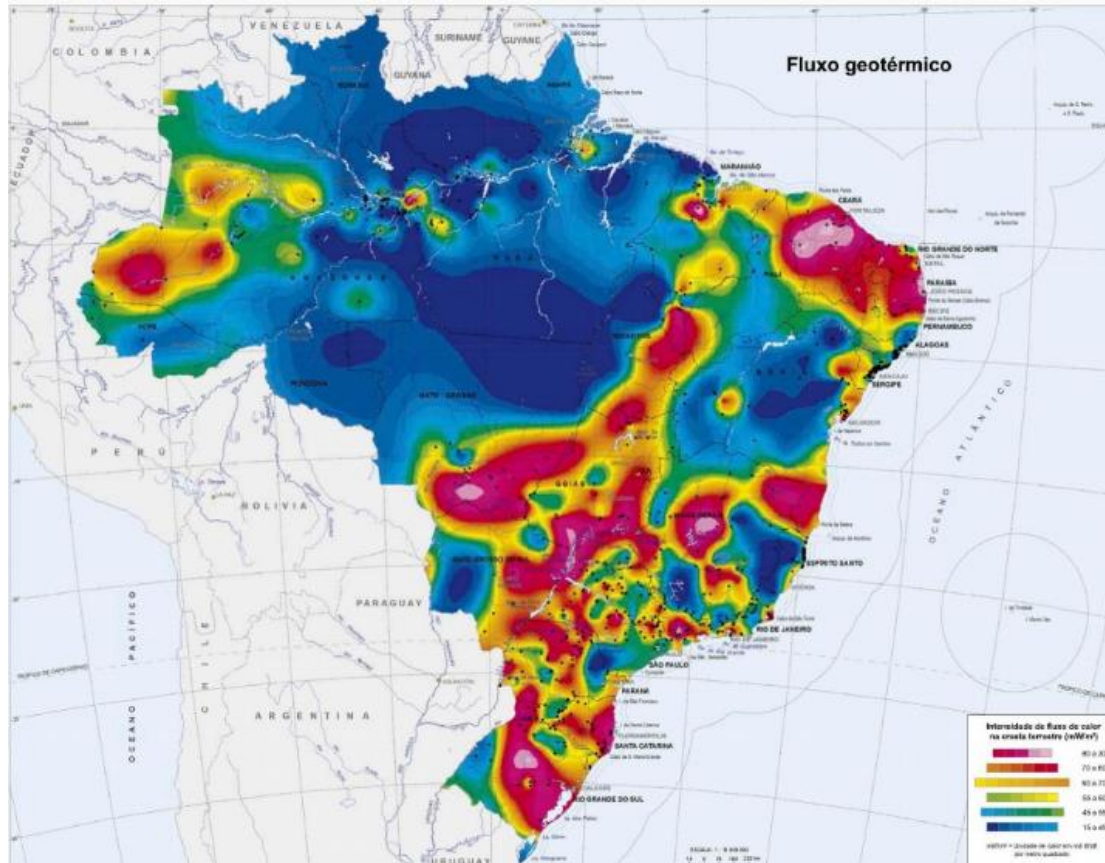
Telea (2015)
Williams et al. (1995)

O que é Visualização?

- Ward et al (2010) definem Visualização como a comunicação da informação utilizando representações gráficas
 - Imagens podem ser utilizadas como mecanismos de comunicação
 - Uma única imagem pode conter muita informação e pode ser processada mais rapidamente quando comparada com uma página com palavras ou números

Visualização no dia-a-dia

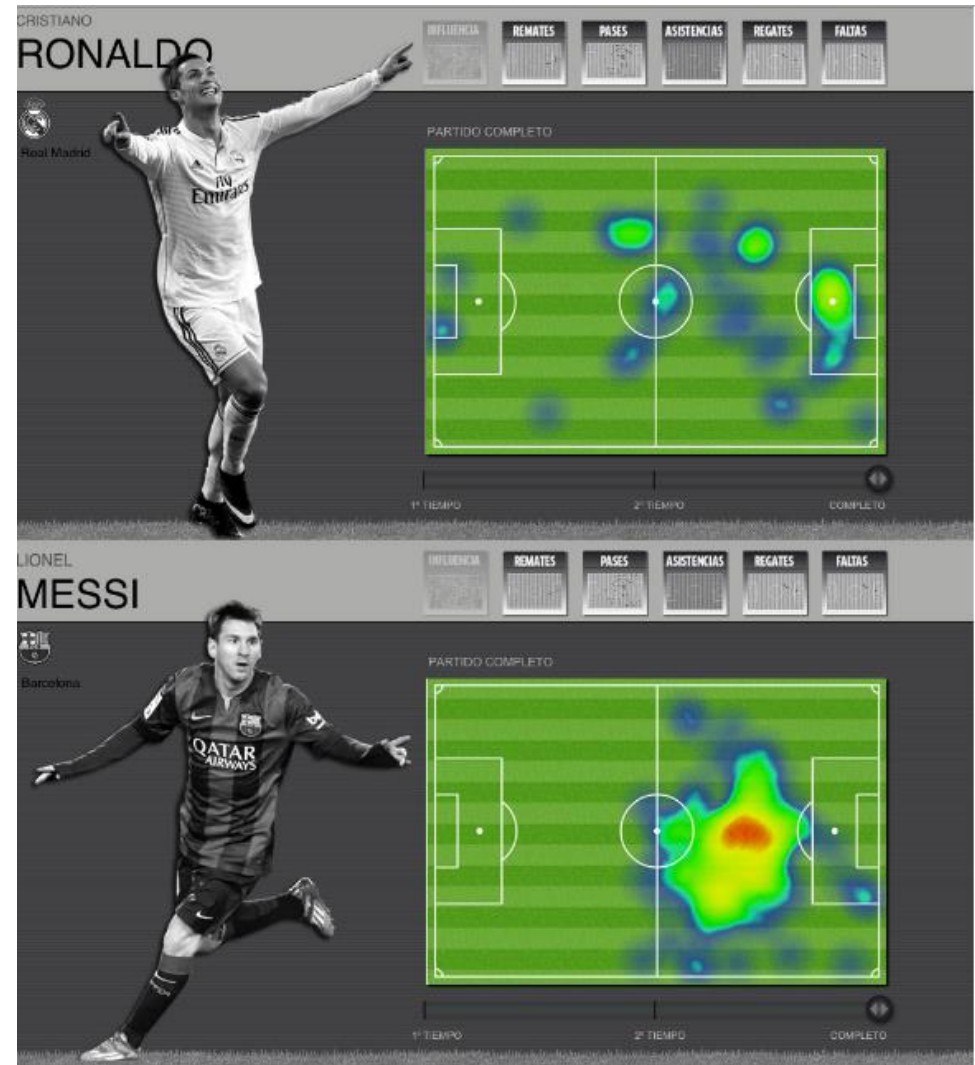
- Imagem contendo a intensidade de fluxo de calor no Brasil



http://1.bp.blogspot.com/-RXLzm9dsX6w/Ufem_gobJaI/AAAAAAAAAHog/5-aqGi-4zCU/s1600/geot%C3%A9rmico.jpg

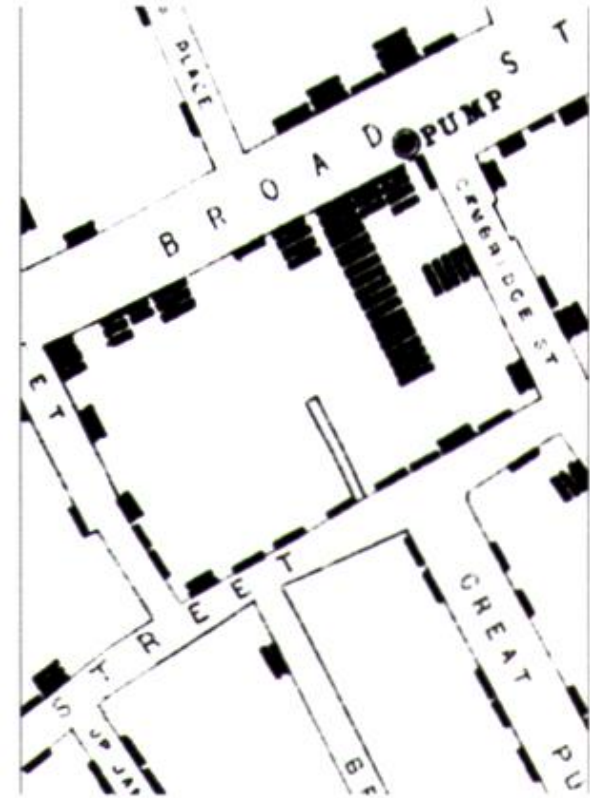
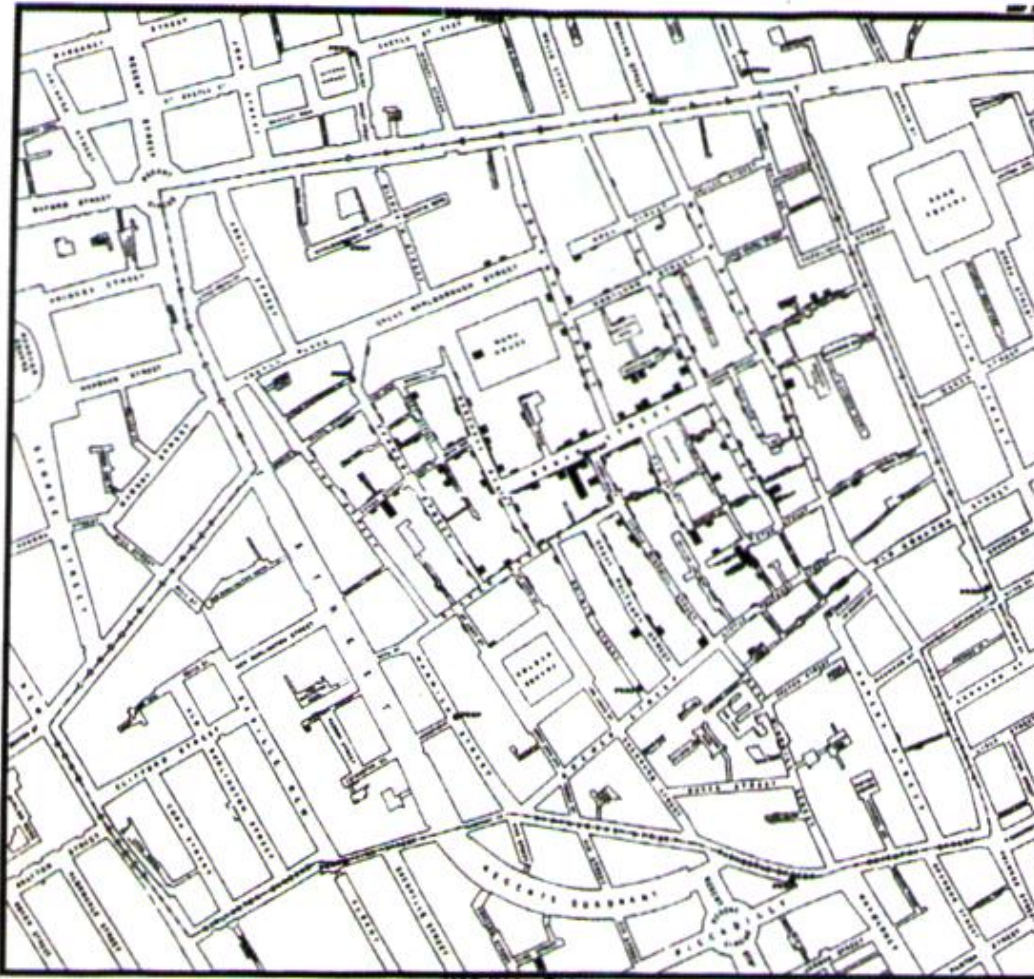
Visualização no dia-a-dia

- Mapas de calor indicando o posicionamento de jogadores em campo



História da Visualização

- Em 1663, foi utilizada a visualização para evidenciar o local da origem do surto de cólera em Londres
 - Foi observado a quantidade de mortos perto da bomba d'água

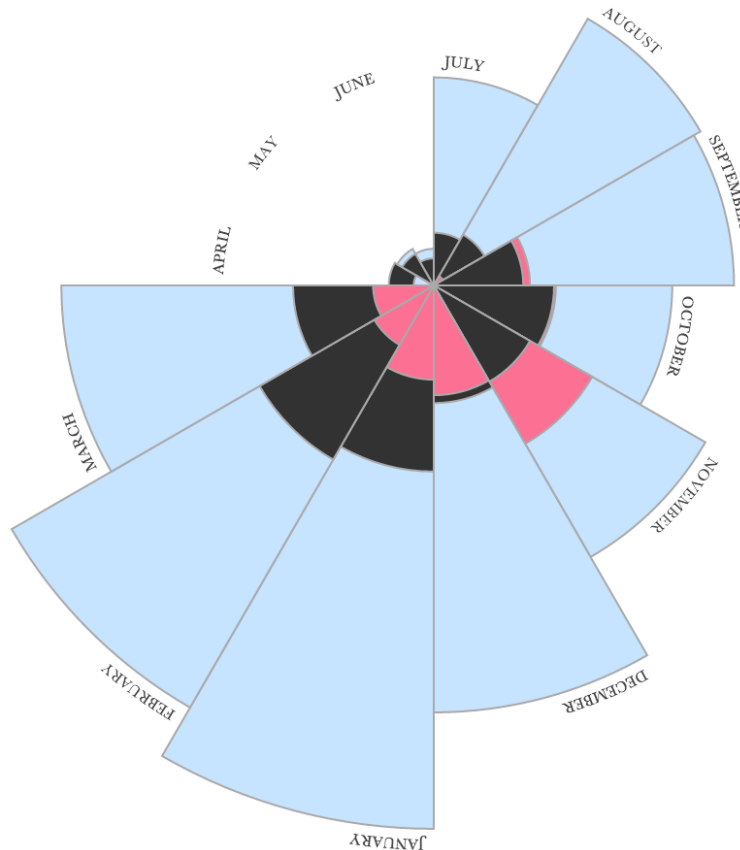


História da Visualização

- Florence Nightingale, enfermeira britânica, apresentou mensalmente as mortes no exército

Abril/1854 – Março 1855

- Morte por doenças
- Morte ferimentos em batalha
- Outros tipos de morte



Guerra da Crimeia
1853 a 1856

O que é visualização?

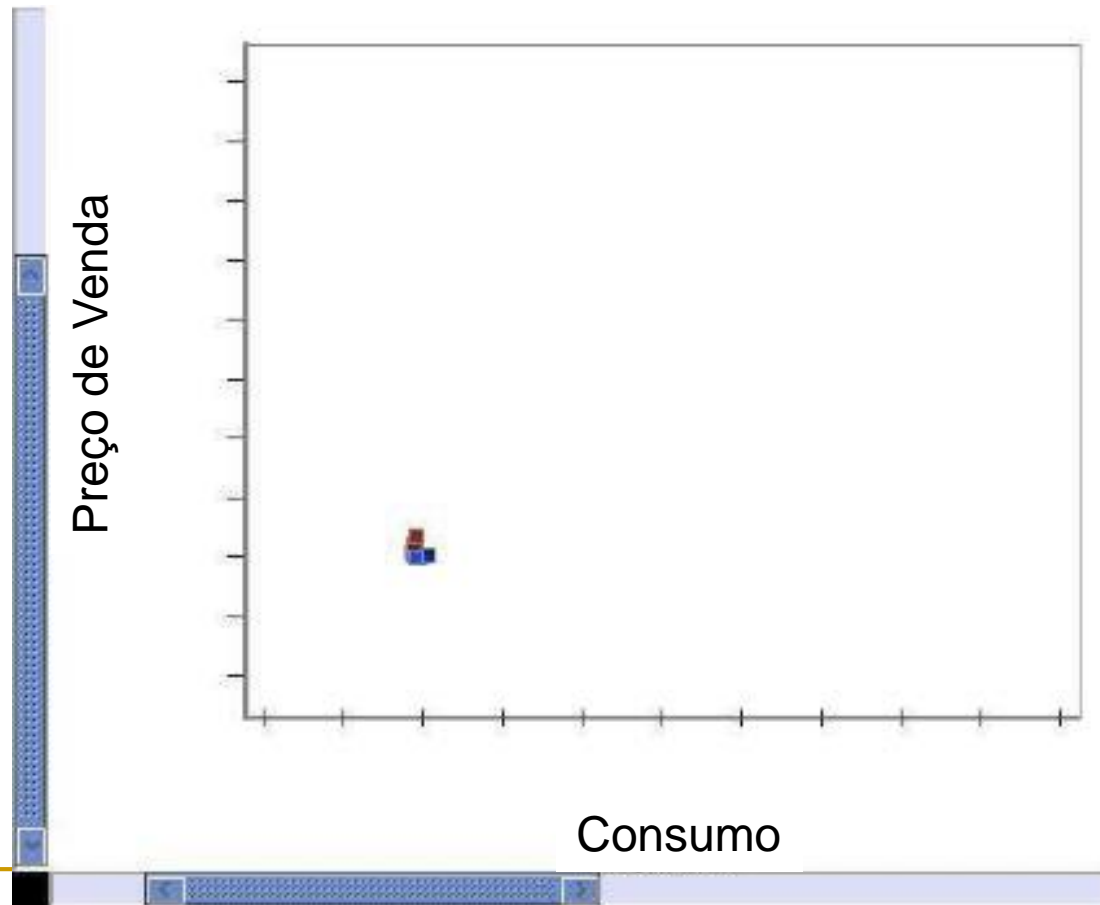
O propósito da visualização é obter 'insight' por meio de representações gráficas interativas, considerando vários aspectos relacionados a algum processo no qual estamos interessados



Adaptado de (Robert Spence, 2007)

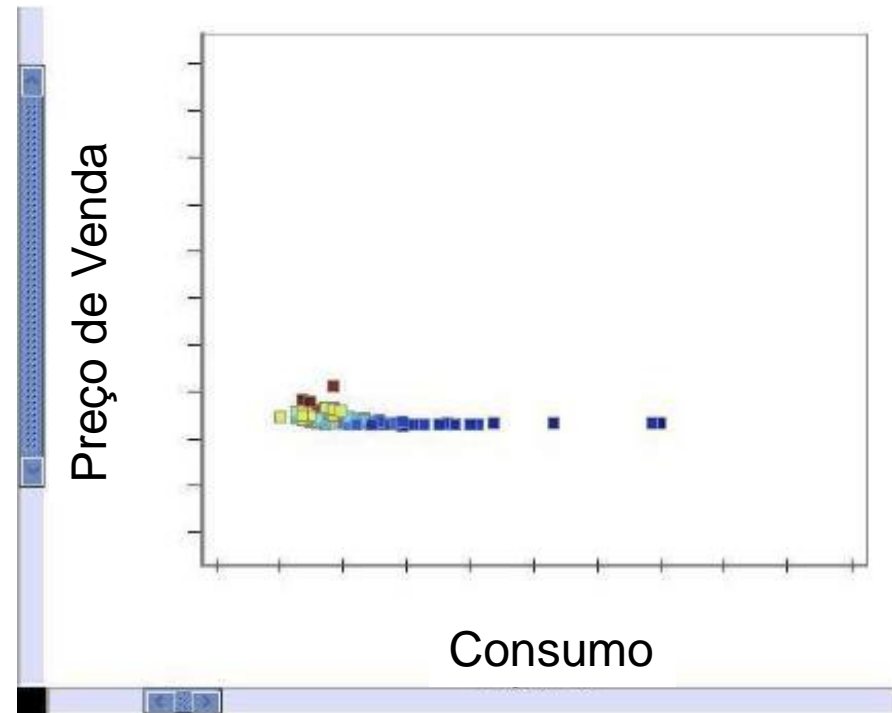
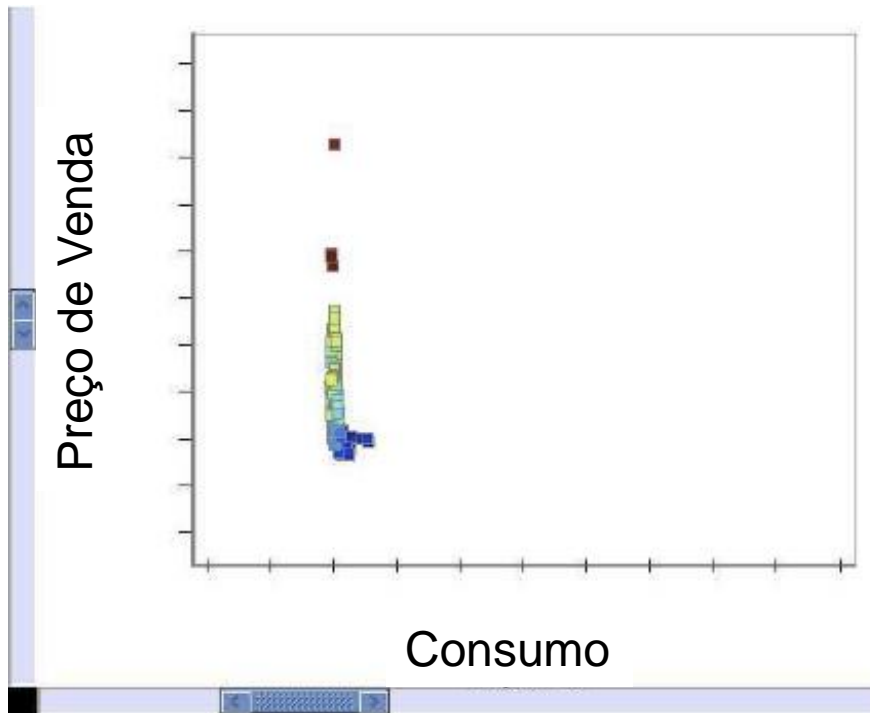
Por que estudar visualização?

- Nessa figura os dados são apresentados em uma escala uniforme em ambos os eixos



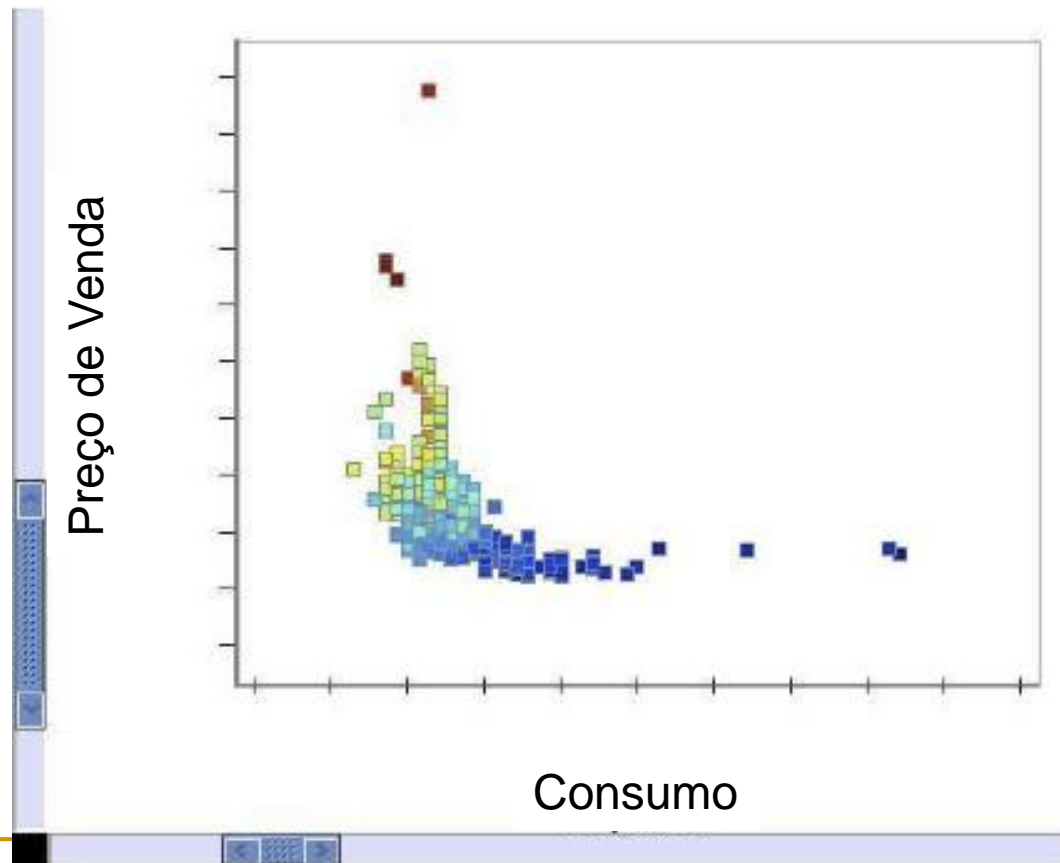
Por que estudar visualização?

- Nessas figuras a escala foi alternada
 - Agrupamentos são criados na direção da escala



Por que estudar visualização?

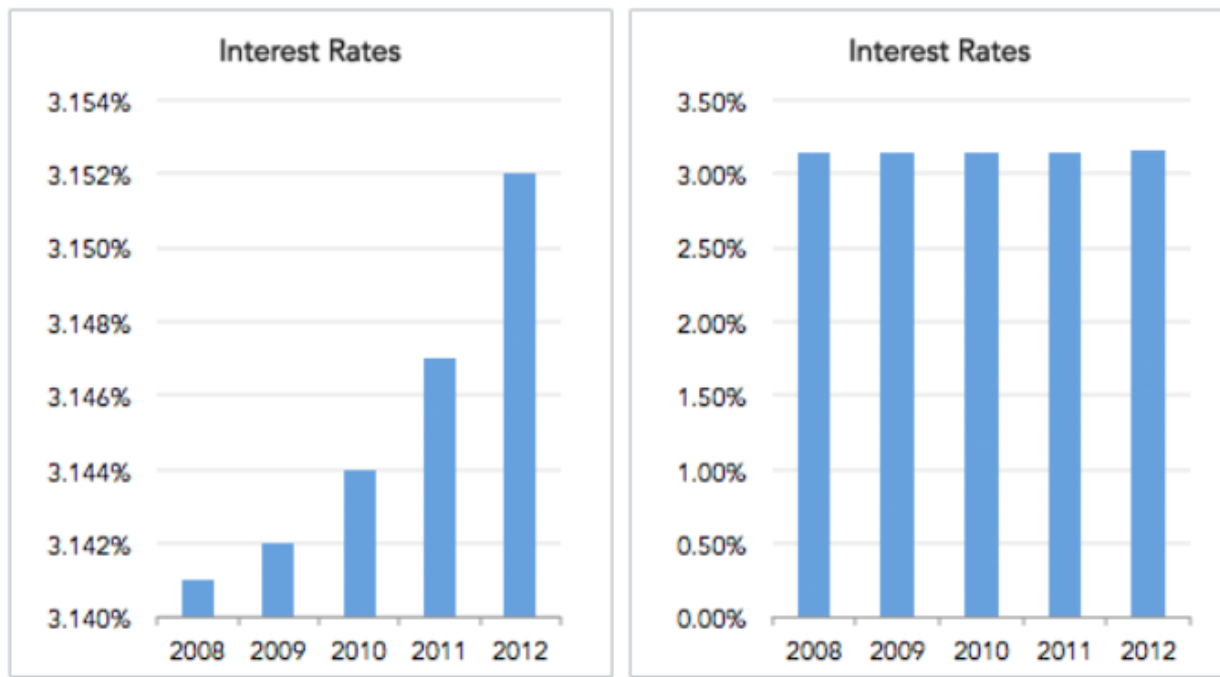
- Nessa figura são utilizados os valores originais dos dados, ou seja, os valores de mínimo e máximo dos eixos X e Y



Por que estudar visualização?

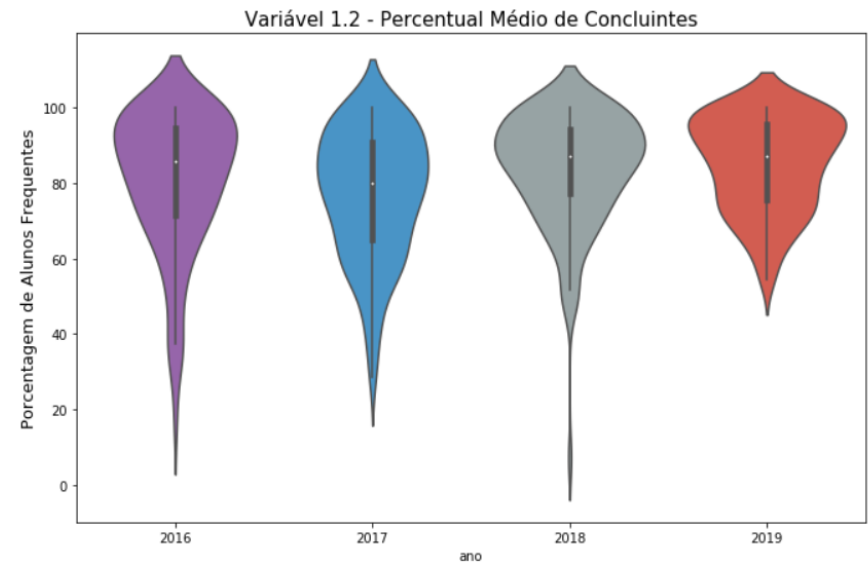
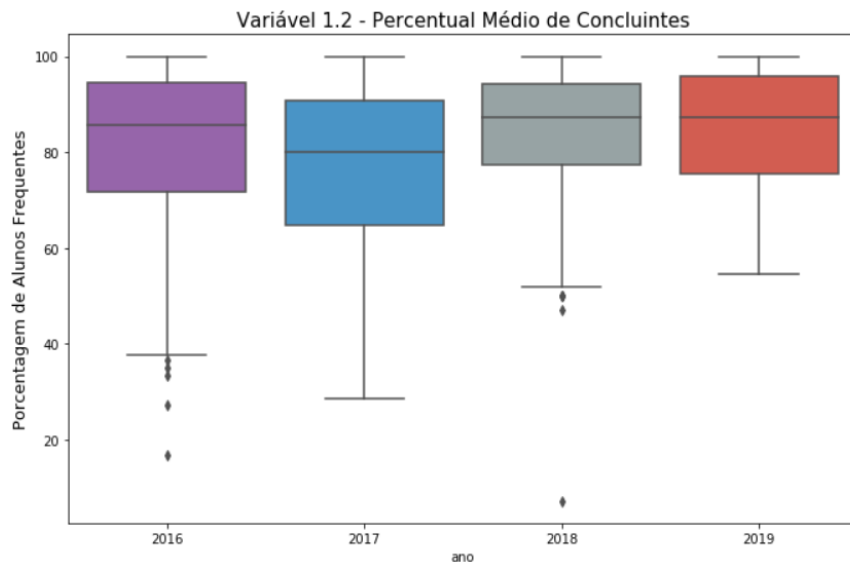
- A visualização pode ser utilizada para distorcer a verdade
 - Pode ser utilizada com esse propósito

Same Data, Different Y-Axis



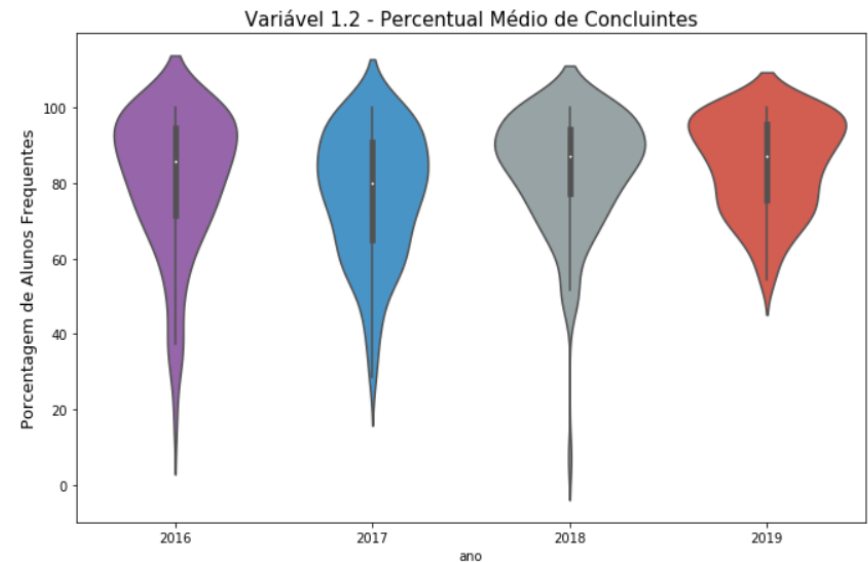
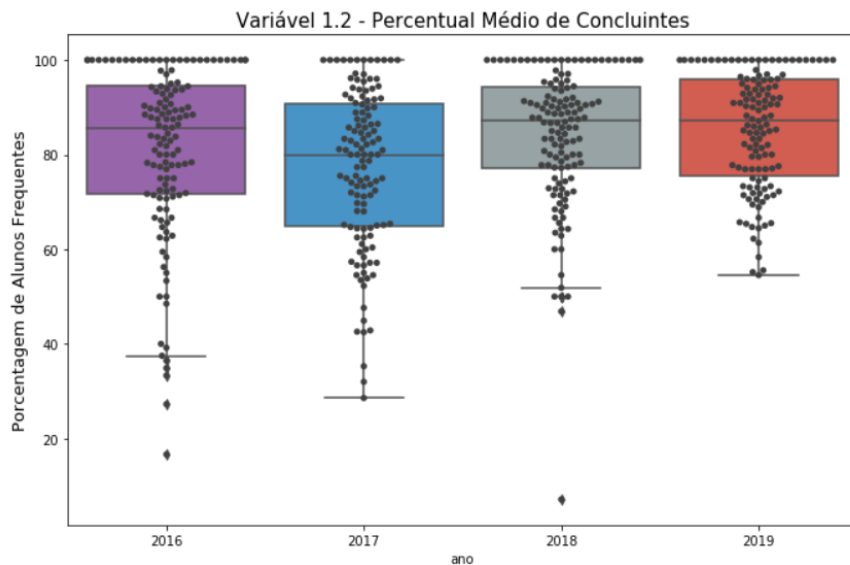
Por que estudar visualização?

- Podemos usar diferentes maneiras de representar os dados
 - Podem impactar o resultado obtido na tomada de decisões



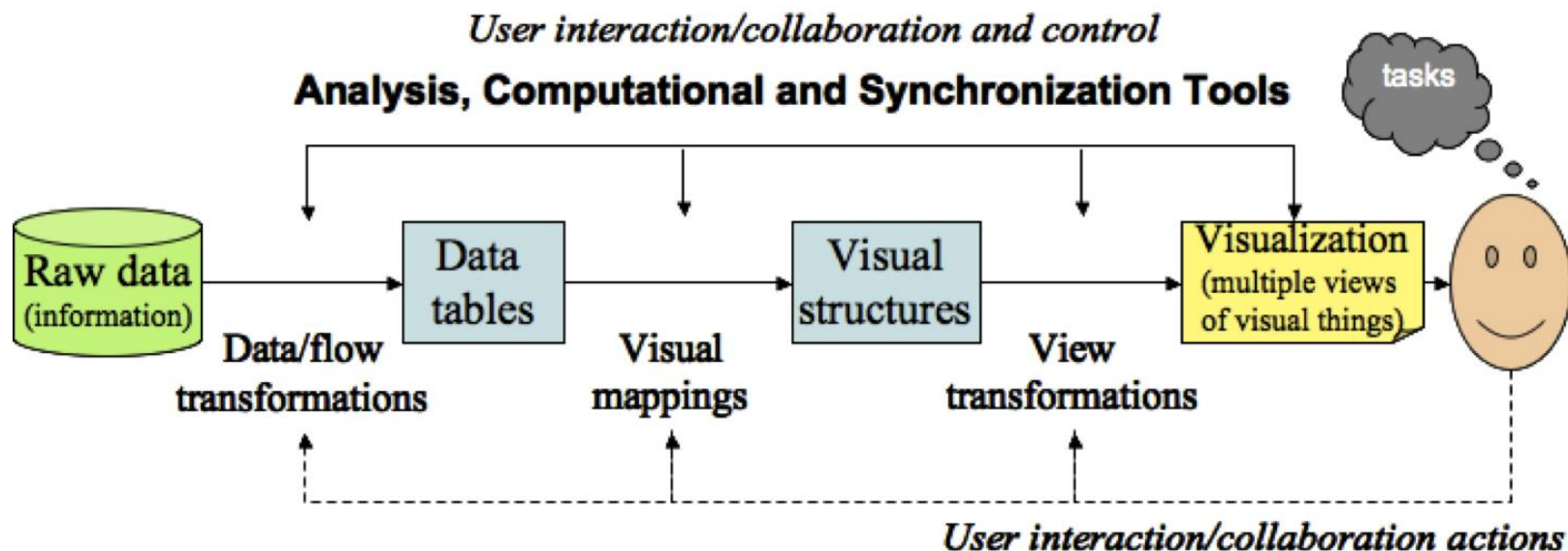
Por que estudar visualização?

- Podemos usar diferentes maneiras de representar os dados
 - Podem impactar o resultado obtido na tomada de decisões



Por que estudar visualização?

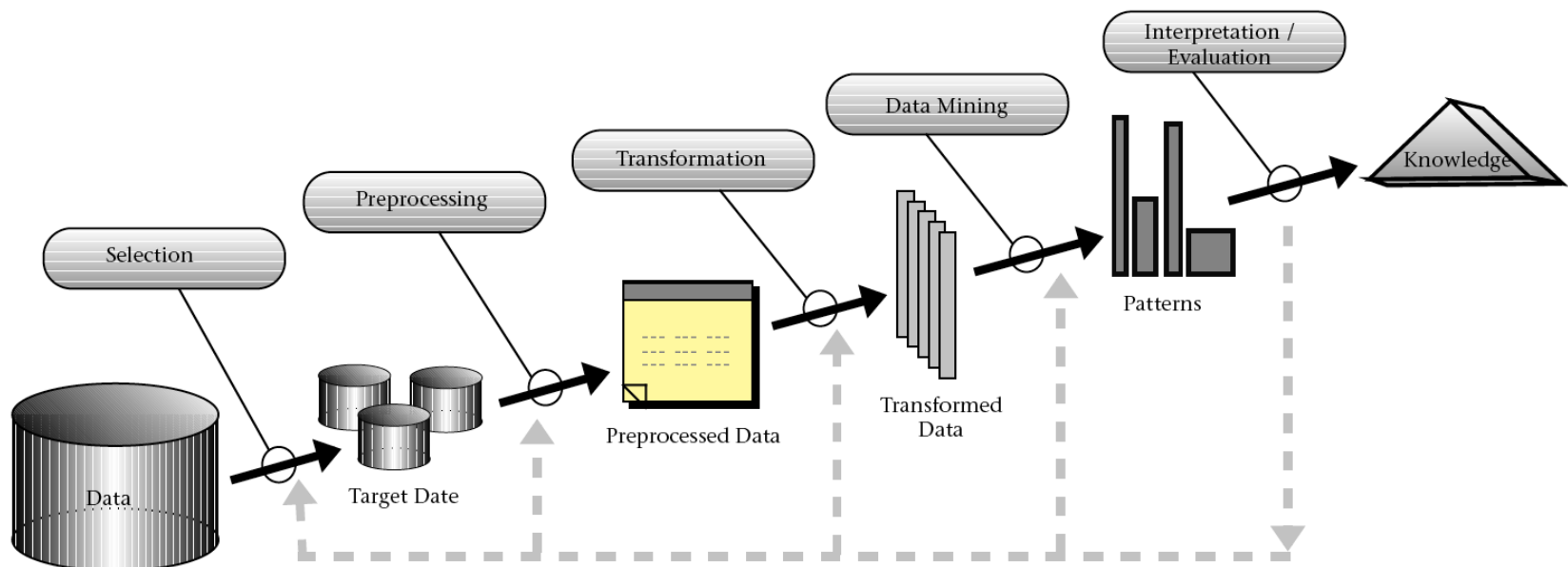
- Pipeline de visualização utilizado pela maioria dos sistemas



Exemplos de Aplicação em Análise de Dados

Análise de Dados

■ Processo de descoberta de conhecimento



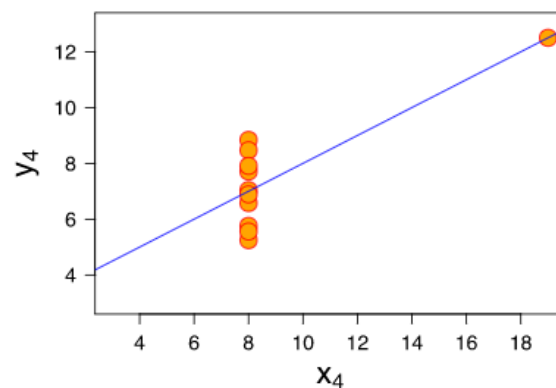
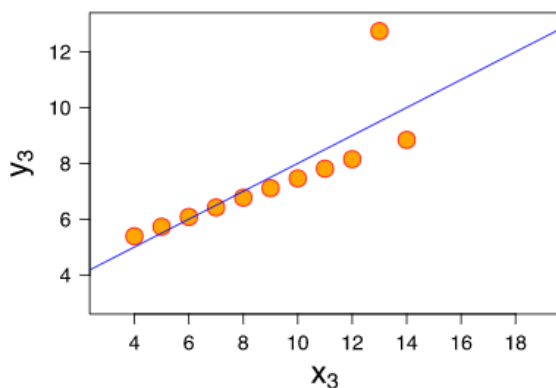
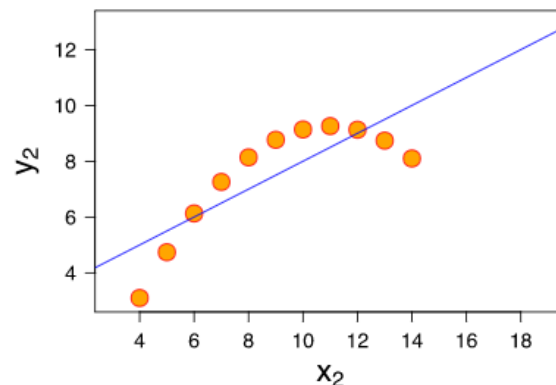
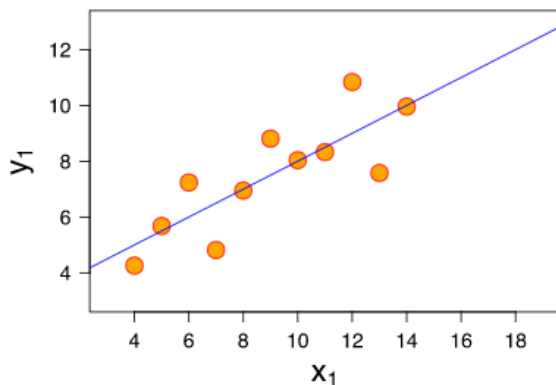
[Fayyad, Piatetsky-Shapiro, and Smyth 1996]

Análise de Dados

- Muito pode ser realizado com estatísticas, mineração de dados, aprendizado de máquina, entre outras técnicas
- A Visualização possibilita explorar questões que não são diretamente efetuadas
 - Auxilia na formulação de novas questões
 - Distribuições, correlações e tendências são melhor compreendidas quando visualizadas

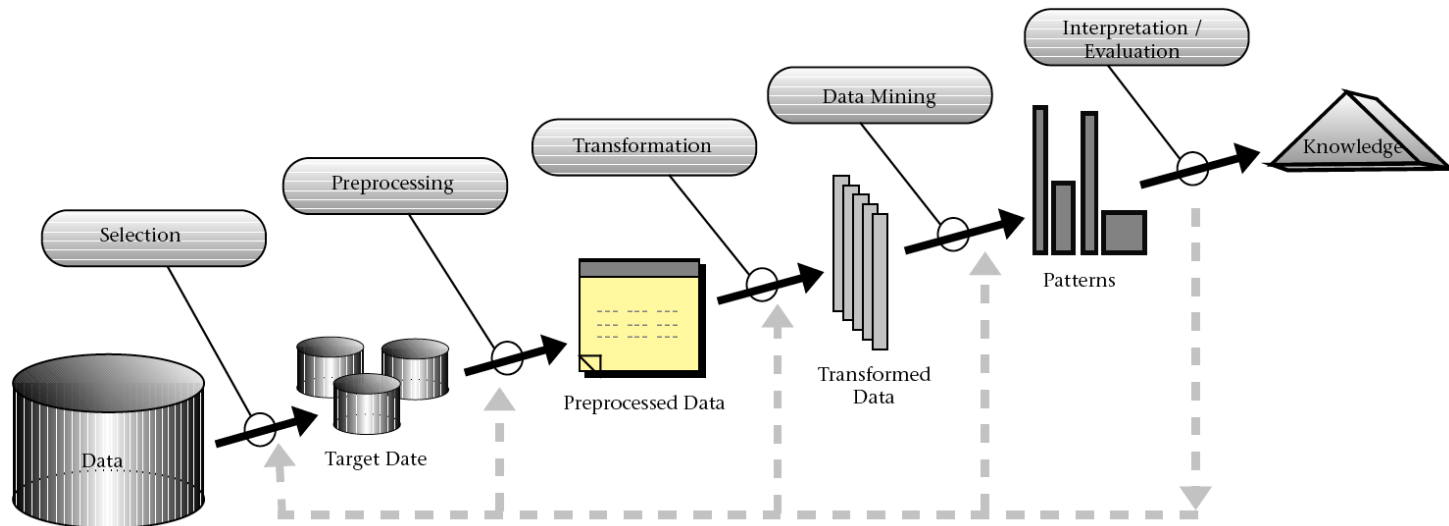
Análise de Dados

- **Quarteto de Anscombe:** Quatro conjuntos de dados que aparentam ser idênticos quando descritos por certas técnicas de estatística descritiva (como média, variância, correlação, regressão linear, entre outras)



Análise de Dados

- Por exemplo, a visualização pode apoiar
 - ❑ Exploração inicial do conjunto de dados
 - ❑ Entendimento das relações de similaridade entre instâncias
 - ❑ Entendimento do comportamento dos atributos
 - ❑ Interpretação dos modelos de aprendizagem de máquina



Visualização de Conjuntos de Dados

■ Dados estruturados

Identificador	Salário	Idade	Peso	Nível	Aprovado
P1	1500,89	30	87,6	1	0
P2	789,52	48	74,6	2	0
P3	1000,00	28	70,5	2	1
P4	589,36	39	90,5	3	1

■ Não Estruturados

□ Texto, Imagem, Som, Vídeo

Visualização de Conjuntos de Dados

■ Conjunto de dados Iris

- ❑ Contém 150 amostras de flores (iris) de três espécies: *setosa*, *versicolor*, e *virginica*
- ❑ Coletadas por Anderson (1935)
 - inclui 50 observações de cada espécie, sendo registrados 4 atributos para cada observação: *sepal length*, *sepal width*, *petal length*, *petal width* (em cm)
 - Três espécies = três classes



Iris setosa



Iris versicolor



Iris virginica

Visualização de Conjuntos de Dados

- 150 flores – 4 atributos – 3 classes

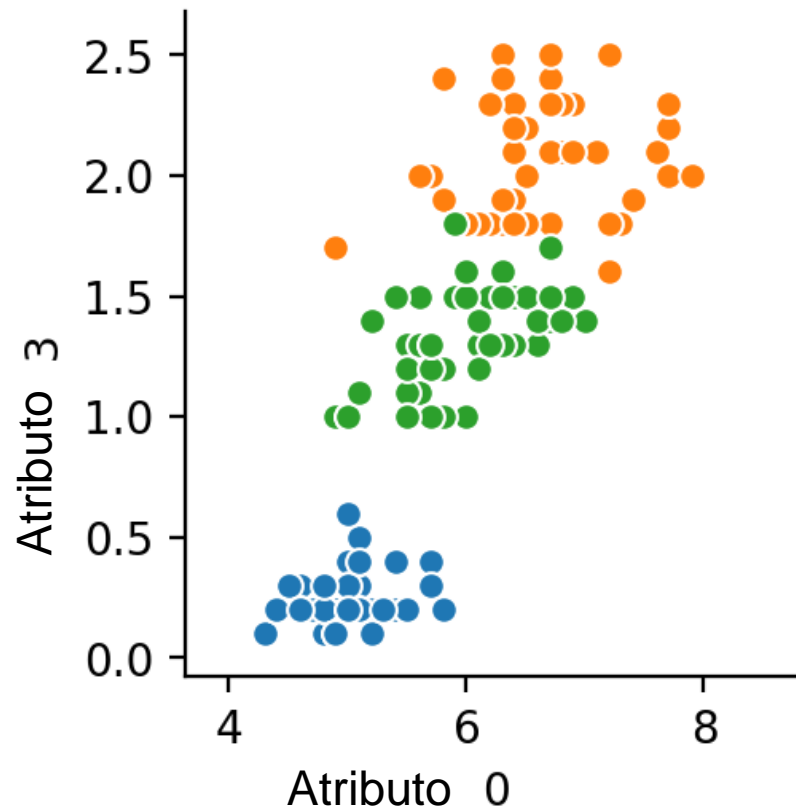
Instância	Comp Sépala	Larg Sépala	Comp Pétala	Larg Pétala	Classe
FLOR 0	5.1	3.5	1.4	0.2	1
FLOR 1	4.4	3	1.3	0.2	1
FLOR 2	6.5	2.8	4.6	1.5	2
FLOR 3	6.4	2.9	4.3	1.3	2
FLOR 4	6.8	2.8	4.8	1.4	2
FLOR 5	5.5	2.4	3.8	1.1	2
FLOR 6	6.4	3.2	5.3	2.3	3
FLOR 7	6.3	2.7	4.9	1.8	3

⋮

FLOR 145	5.5	2.4	3.7	1	2
FLOR 146	5.5	2.6	4.4	1.2	2
FLOR 147	5.8	2.6	4	1.2	2
FLOR 148	5.7	2.8	4.1	1.3	2
FLOR 149	7.7	3.8	6.7	2.2	3

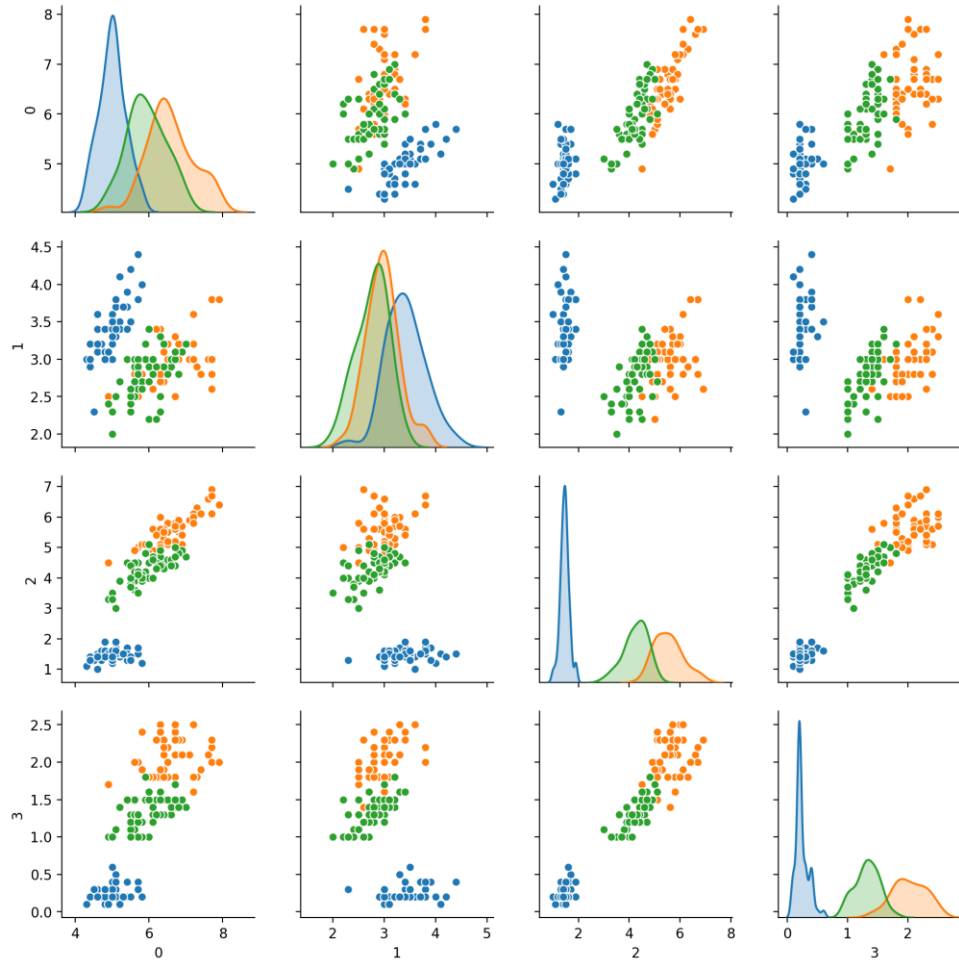
Visualização de Conjuntos de Dados

■ Gráfico de Dispersão (Scatter Plot)



Visualização de Conjuntos de Dados

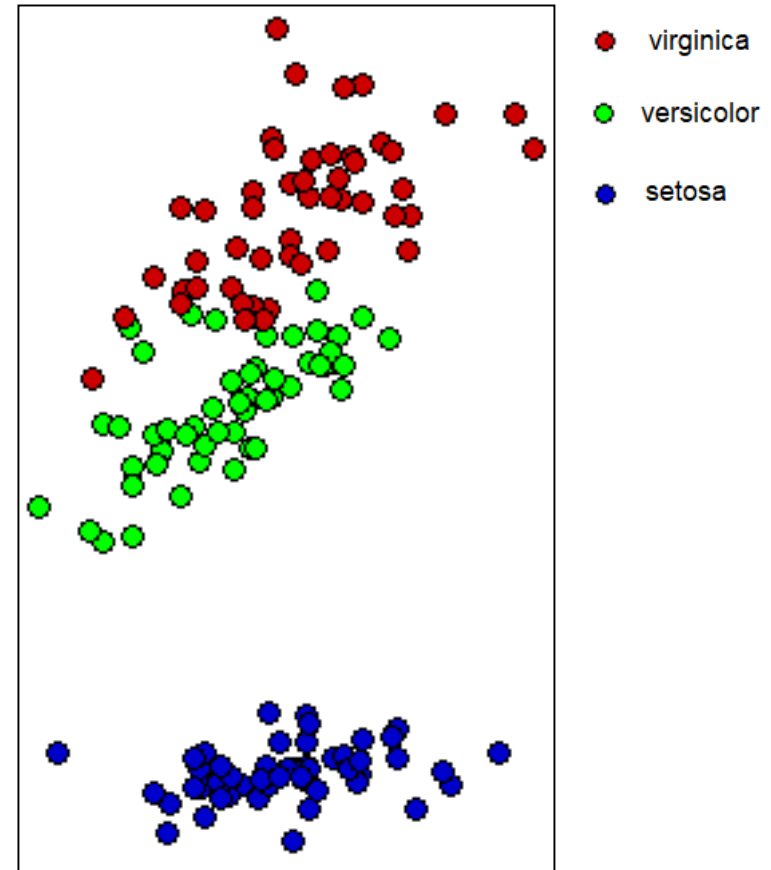
■ Matriz de Gráfico de Dispersão (Scatter Plot Matrix)



Projeção Multidimensional

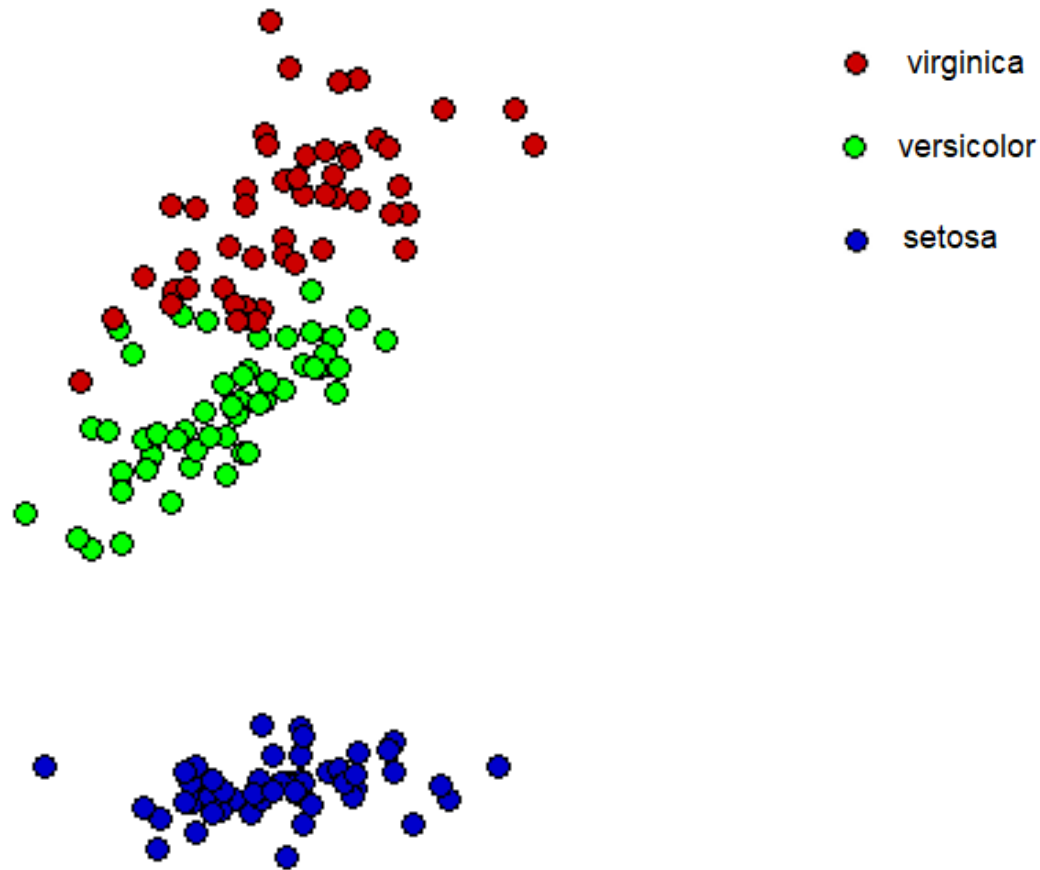
- Uma maneira de visualizar as relações de similaridade de um conjunto de dados é reduzir sua dimensionalidade

Instância	Comp Sépala	Larg Sépala	Comp Pétala	Larg Pétala	Classe
FLOR 0	5.1	3.5	1.4	0.2	1
FLOR 1	4.4	3	1.3	0.2	1
FLOR 2	6.5	2.8	4.6	1.5	2
FLOR 3	6.4	2.9	4.3	1.3	2
FLOR 4	6.8	2.8	4.8	1.4	2
FLOR 5	5.5	2.4	3.8	1.1	2
FLOR 6	6.4	3.2	5.3	2.3	3
FLOR 7	6.3	2.7	4.9	1.8	3
⋮					
FLOR 145	5.5	2.4	3.7	1	2
FLOR 146	5.5	2.6	4.4	1.2	2
FLOR 147	5.8	2.6	4	1.2	2
FLOR 148	5.7	2.8	4.1	1.3	2
FLOR 149	7.7	3.8	6.7	2.2	3



Projeção Multidimensional

- Exploração inicial do conjunto de dados
- Entendimento das relações de similaridade entre instâncias



Coordenadas Paralelas

■ Exemplo para o conjunto de dados Iris



Iris setosa

sepal length	sepal width	petal length	petal width
5.1	3.5	1.4	0.2
4.9	3	1.4	0.2
...
5.9	3	5.1	1.8



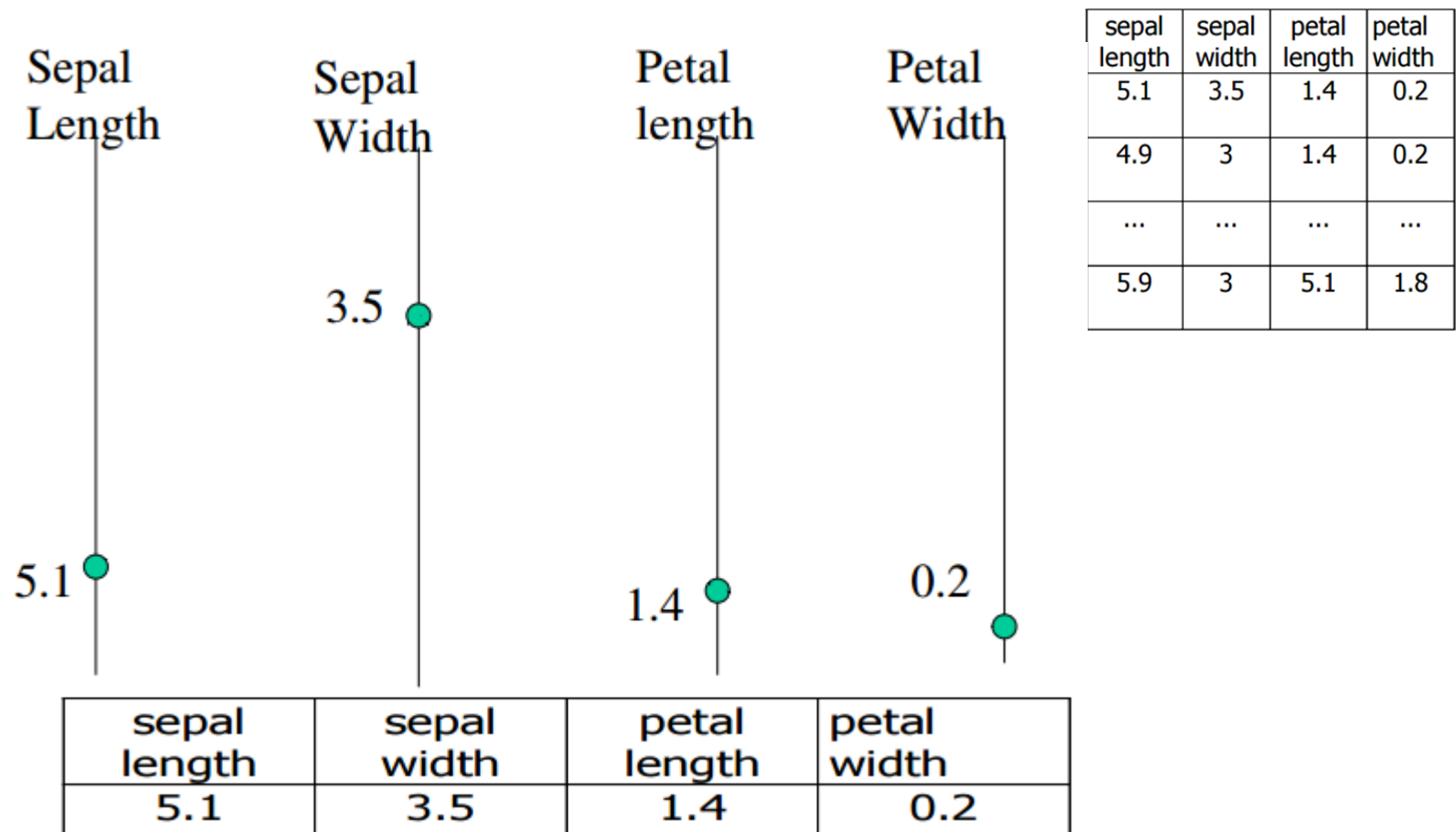
Iris versicolor



Iris virginica

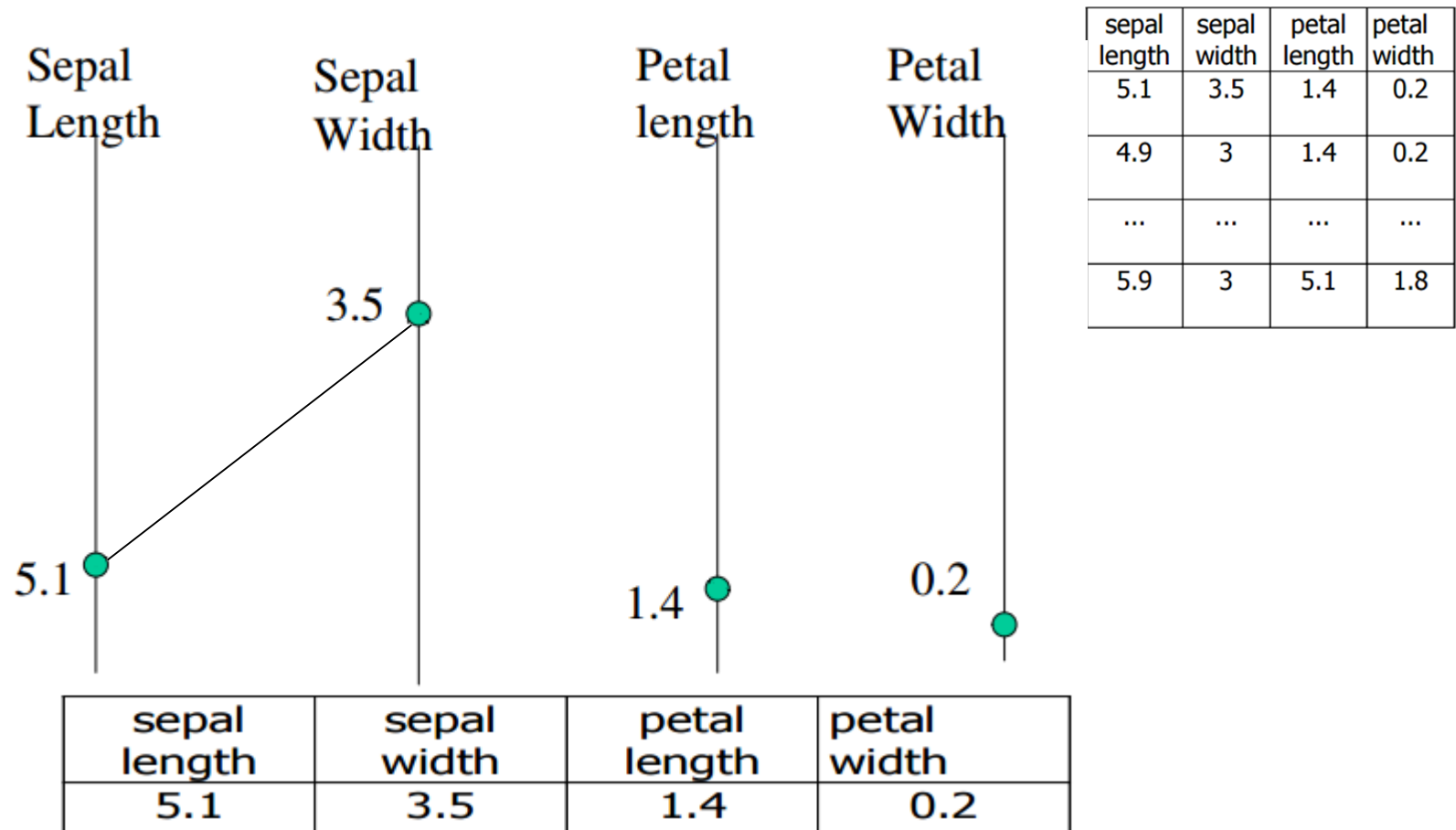
Coordenadas Paralelas

■ Exemplo do conjunto de dados Iris



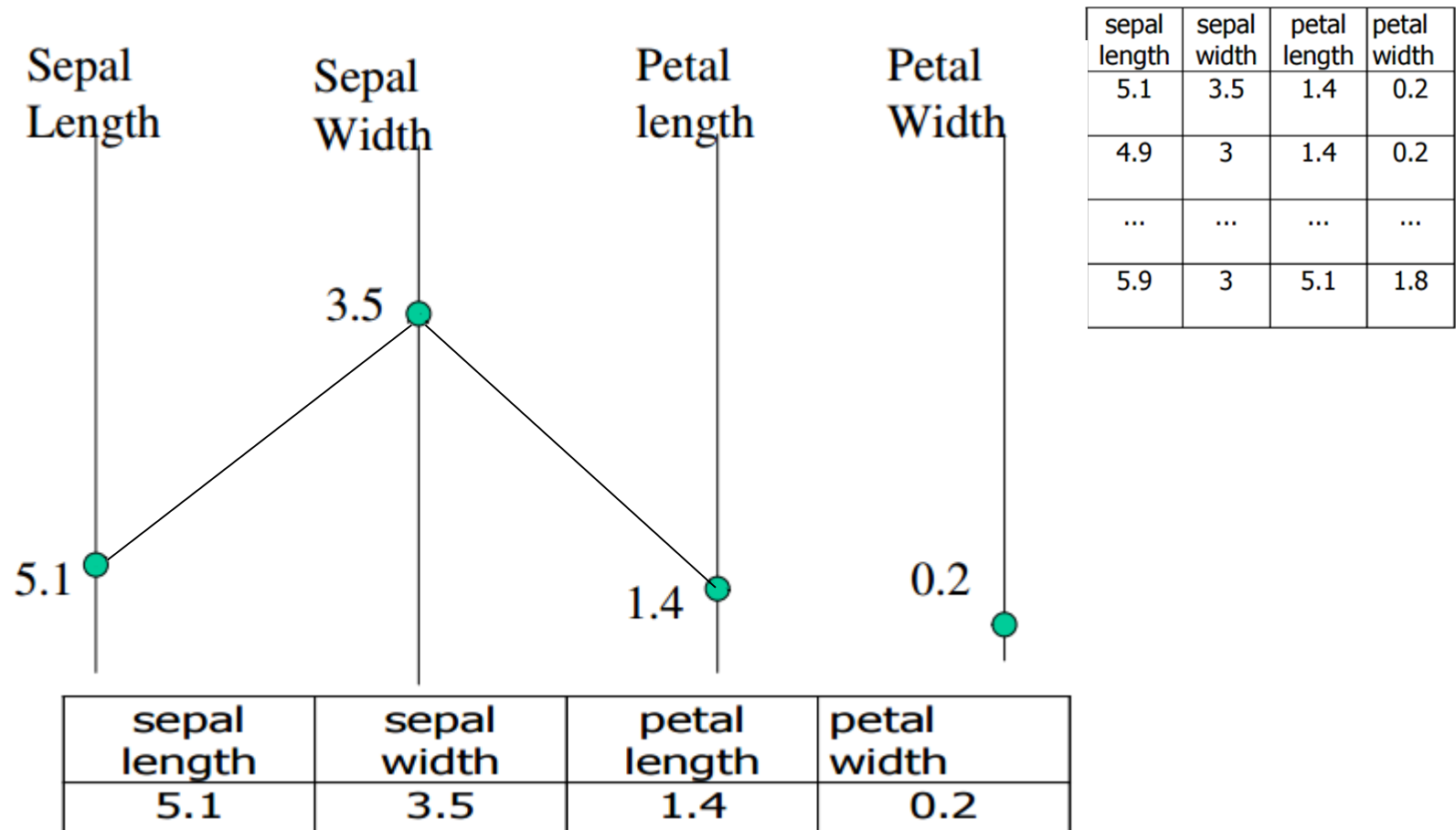
Coordenadas Paralelas

■ Exemplo do conjunto de dados Iris



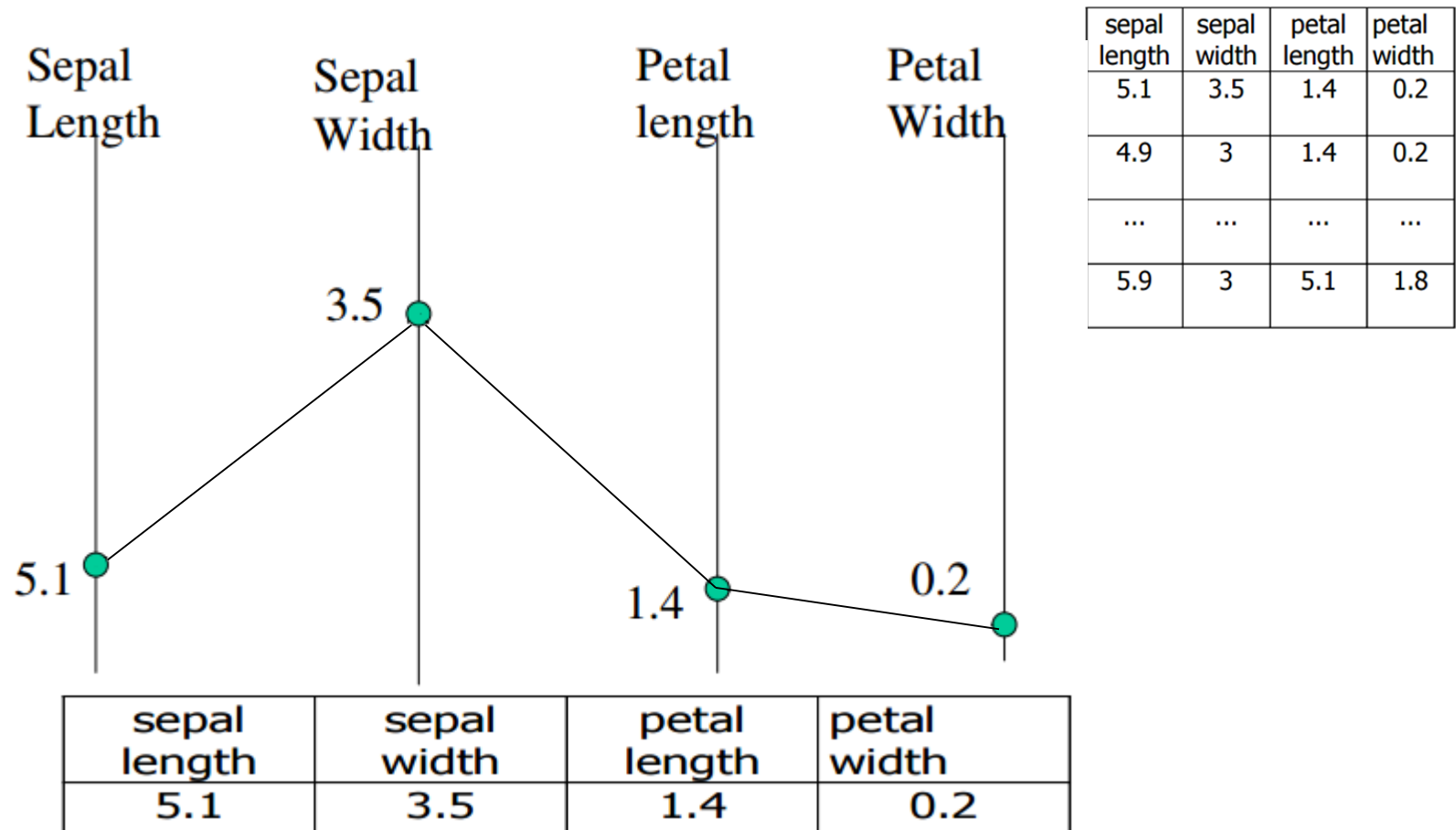
Coordenadas Paralelas

■ Exemplo do conjunto de dados Iris



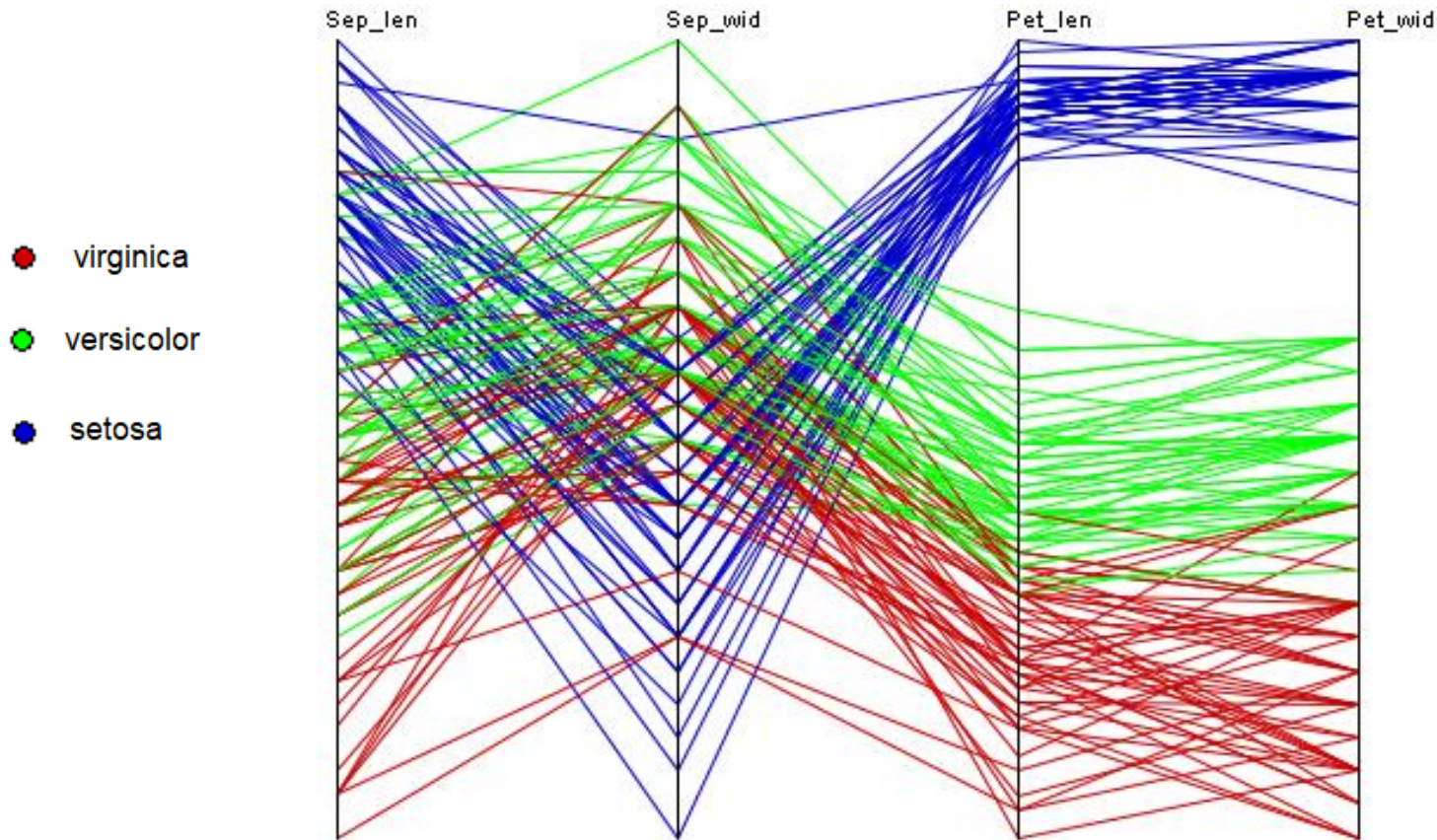
Coordenadas Paralelas

■ Exemplo do conjunto de dados Iris



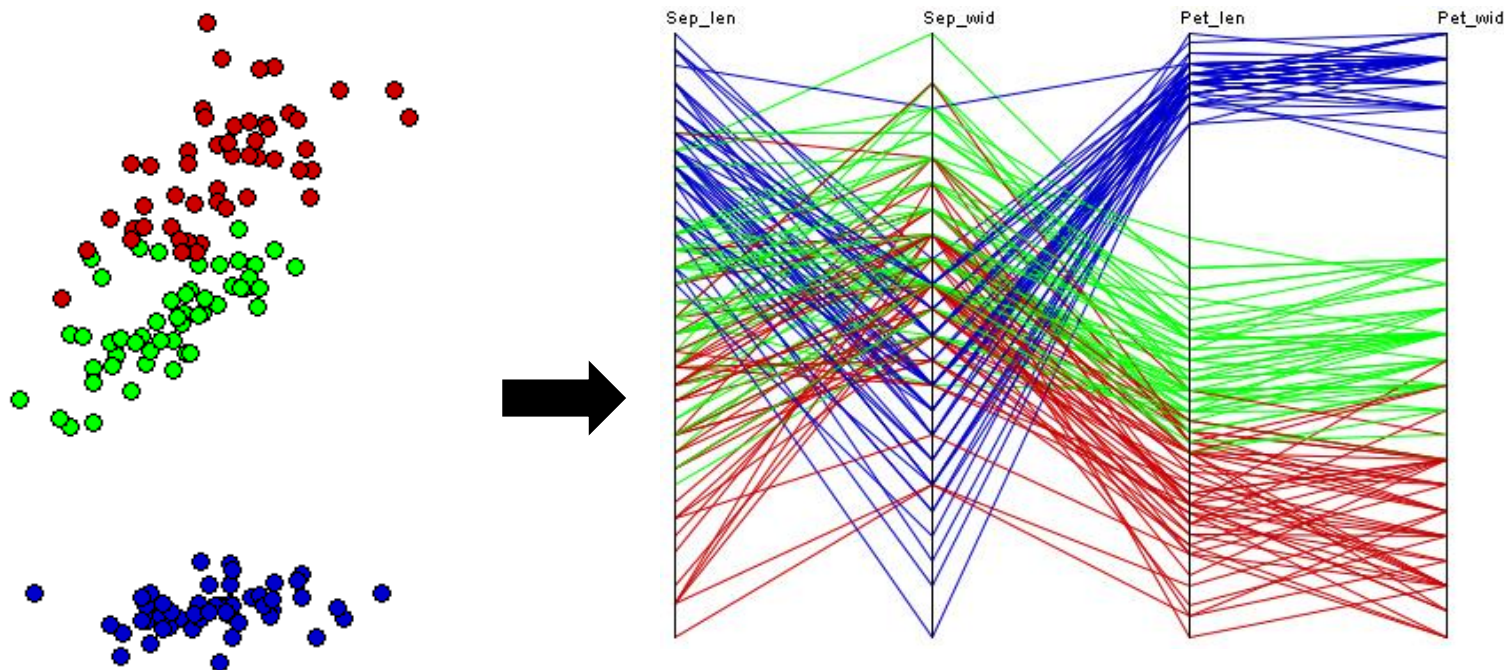
Coordenadas Paralelas

■ Exemplo do conjunto de dados Iris



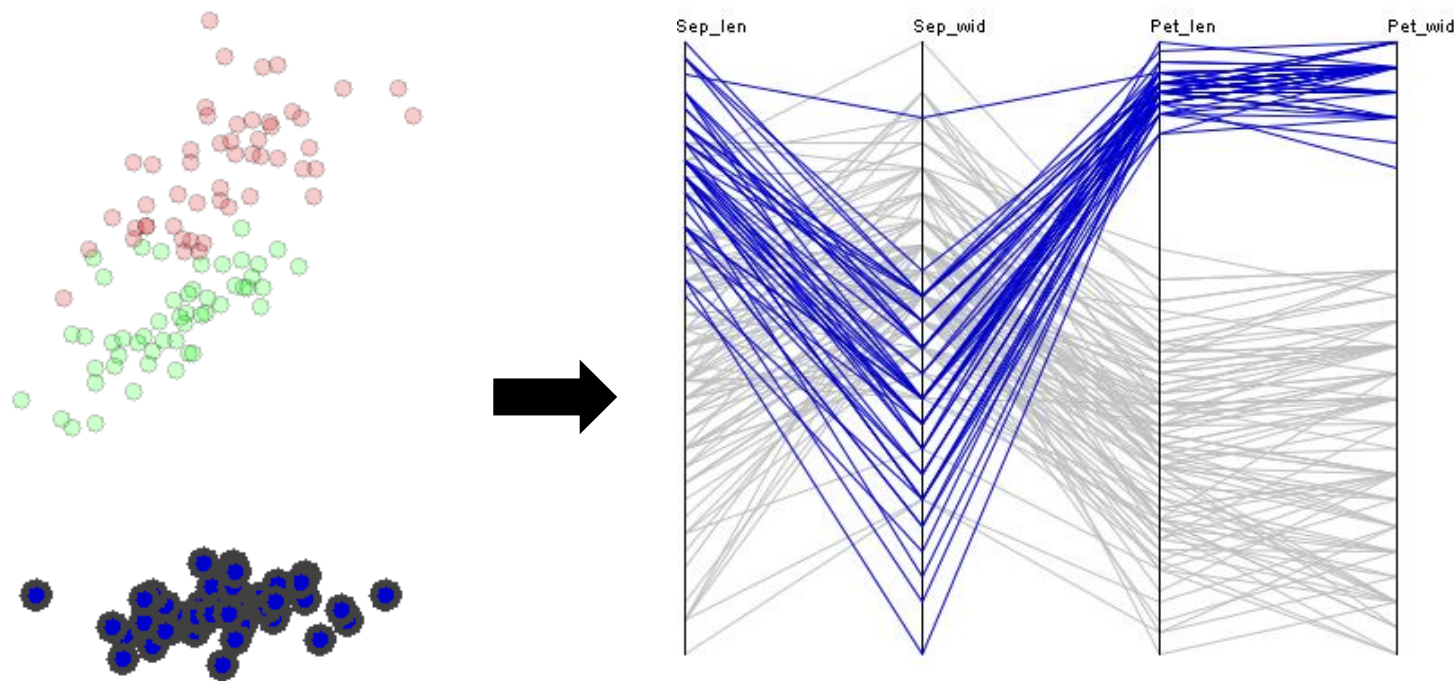
Projeção e Coordenadas Paralelas

- Coordenação entre diferentes técnicas de visualização
- Entendimento do comportamento dos atributos



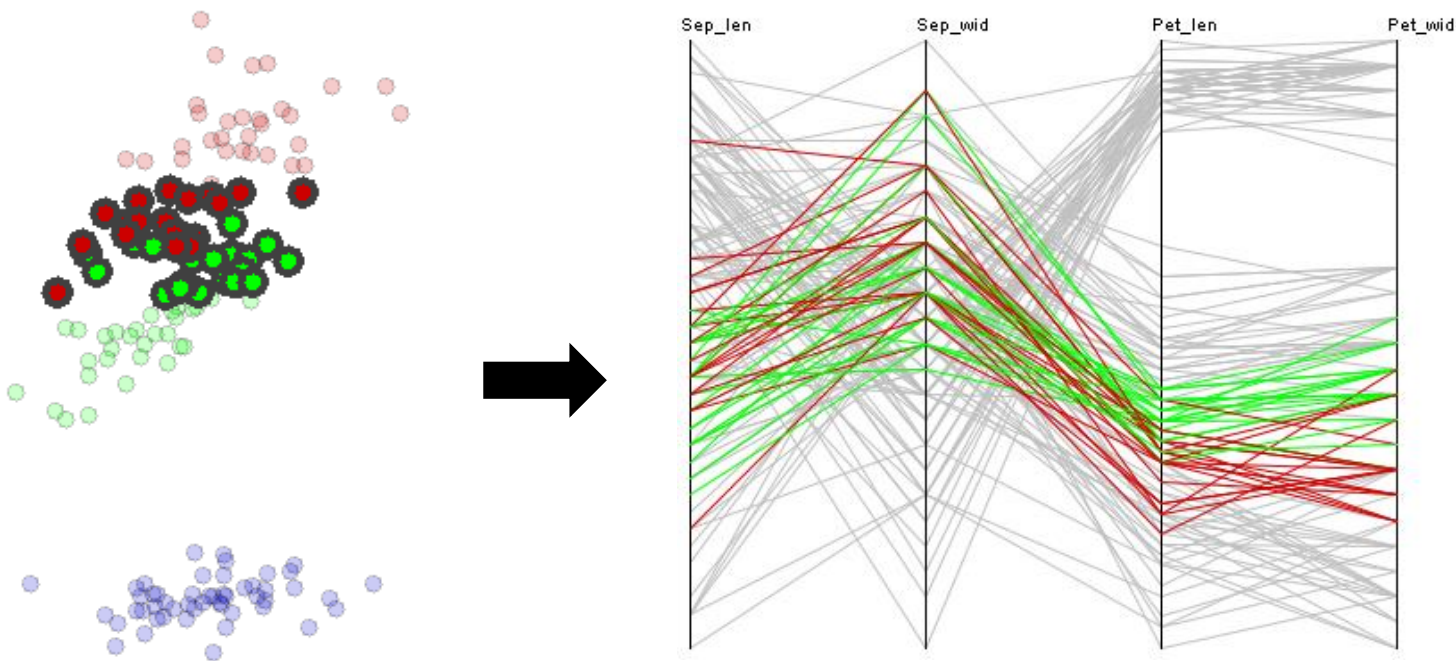
Projeção e Coordenadas Paralelas

- Coordenação entre diferentes técnicas de visualização
- Entendimento do comportamento dos atributos



Projeção e Coordenadas Paralelas

- Coordenação entre diferentes técnicas de visualização
- Entendimento do comportamento dos atributos



Análise de Modelos de Aprendizagem de Máquina

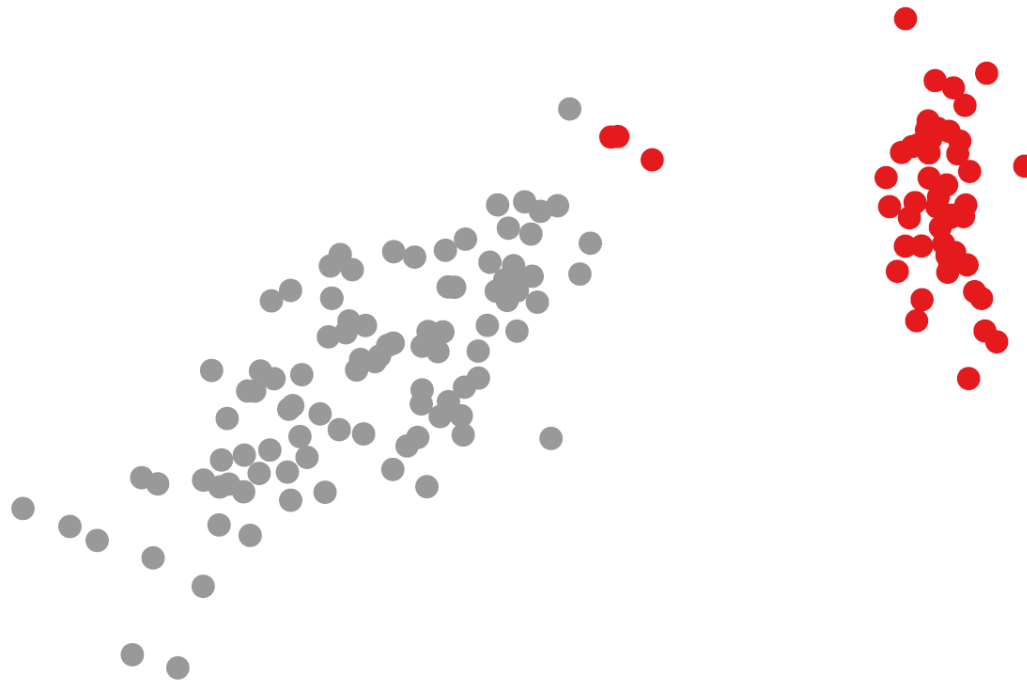
- Análise de algoritmos de agrupamento
 - Ex. k-means



Análise de Modelos de Aprendizagem de Máquina

- Análise de algoritmos de agrupamento
 - Ex. k-means

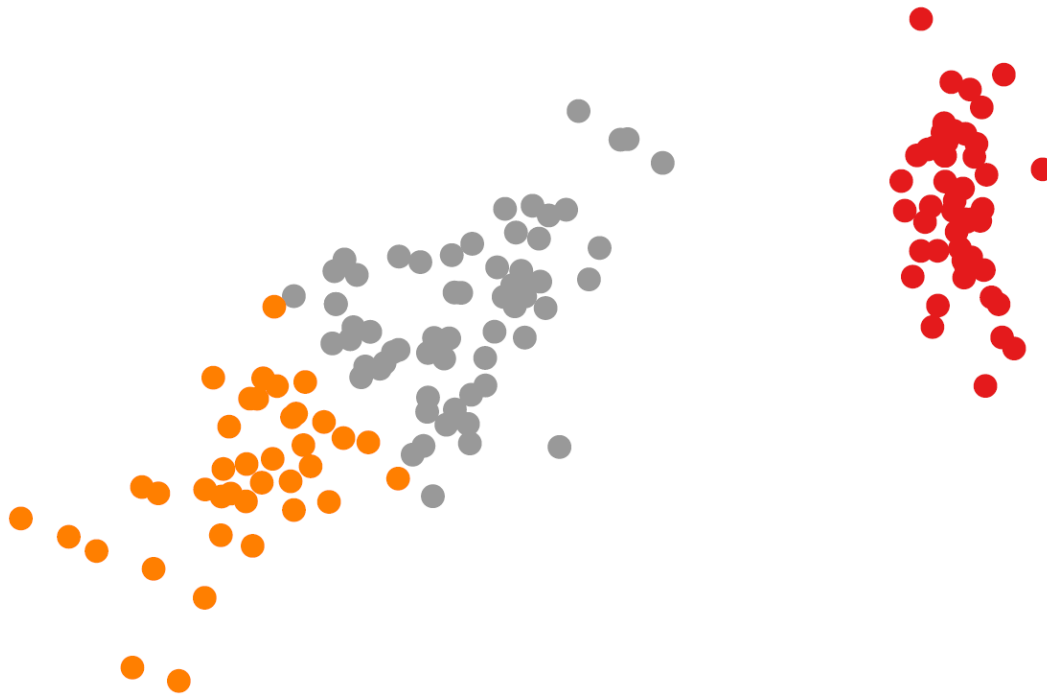
k=2



Análise de Modelos de Aprendizagem de Máquina

- Análise de algoritmos de agrupamento
 - Ex. k-means

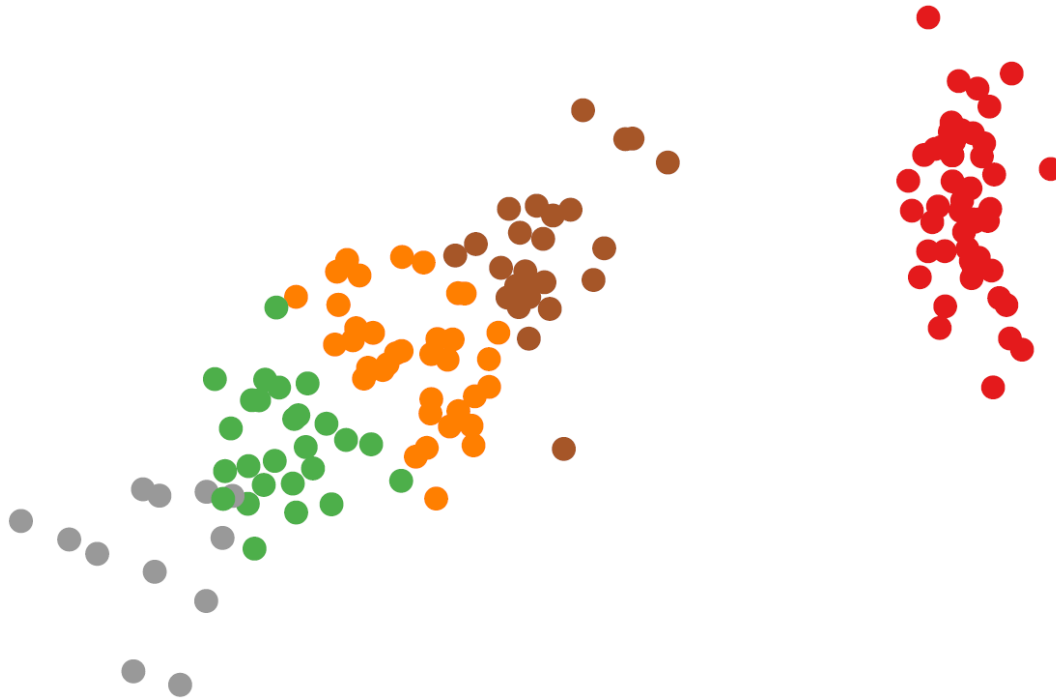
k=3



Análise de Modelos de Aprendizagem de Máquina

- Análise de algoritmos de agrupamento
 - Ex. k-means

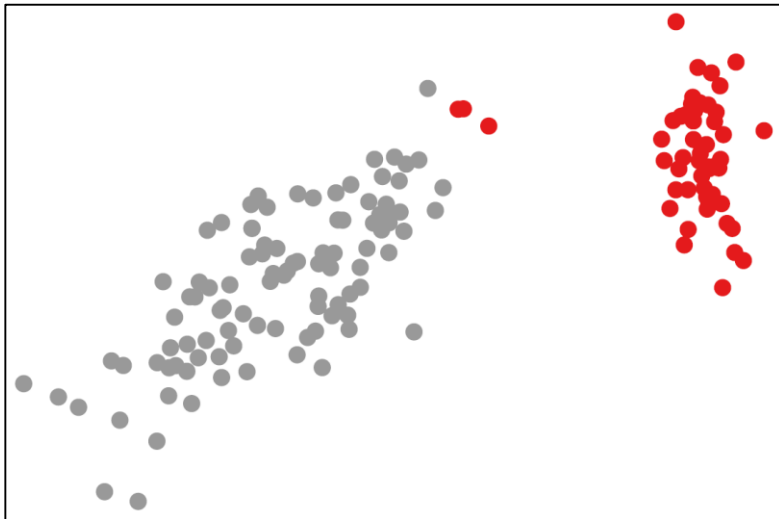
k=5



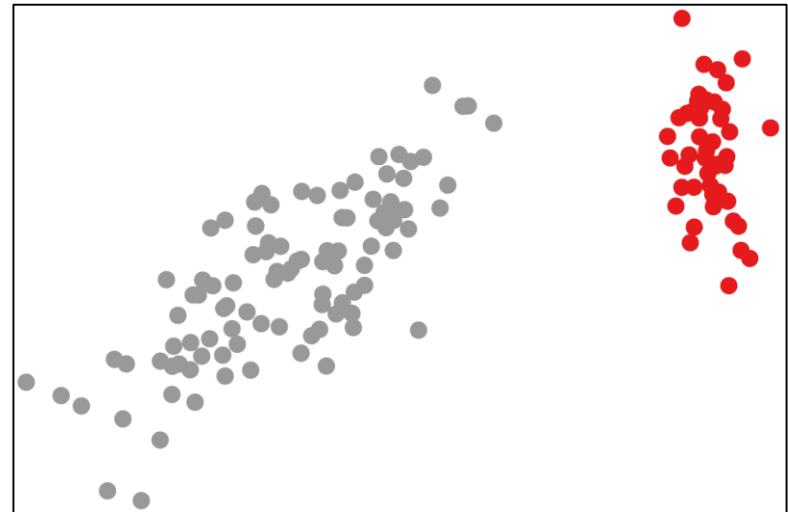
Análise de Modelos de Aprendizagem de Máquina

- Análise de algoritmos de agrupamento
 - Ex. k-means e DBSCAN

k=2

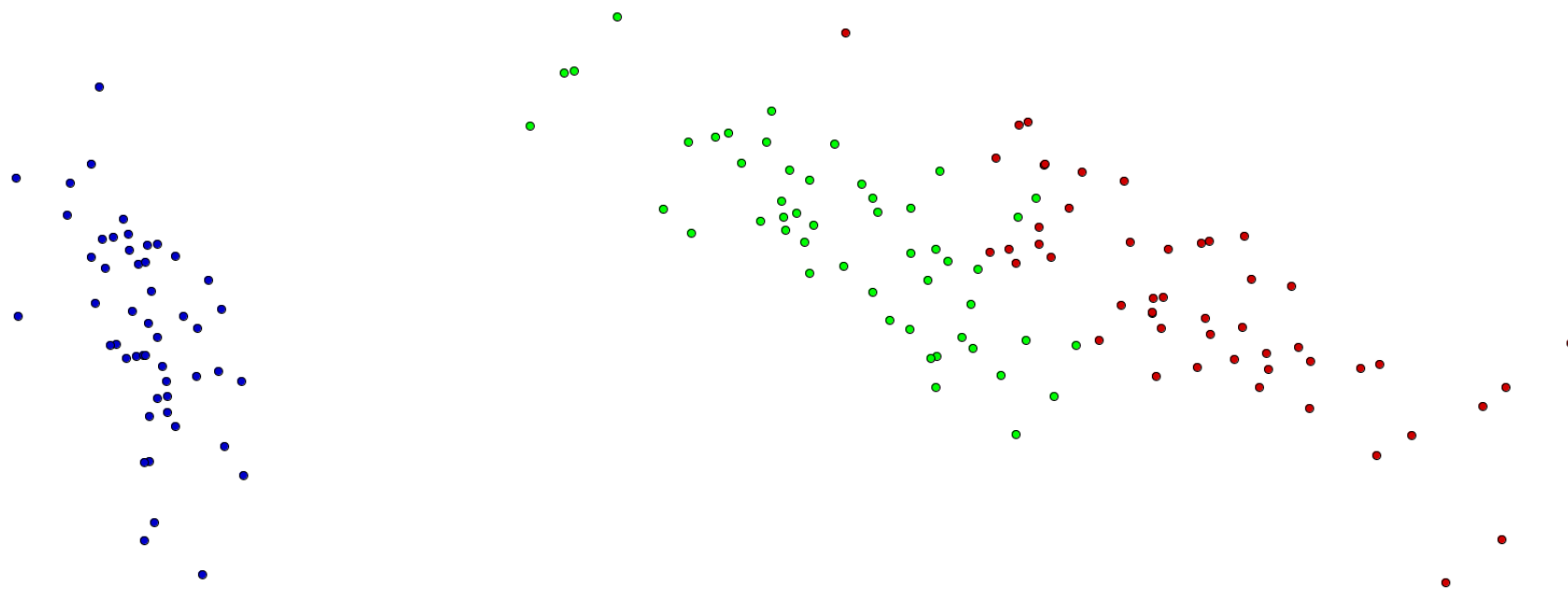


eps = 1
minPts = 5



Análise de Modelos de Aprendizagem de Máquina

- Análise de algoritmos de classificação
 - Ex. k-NN

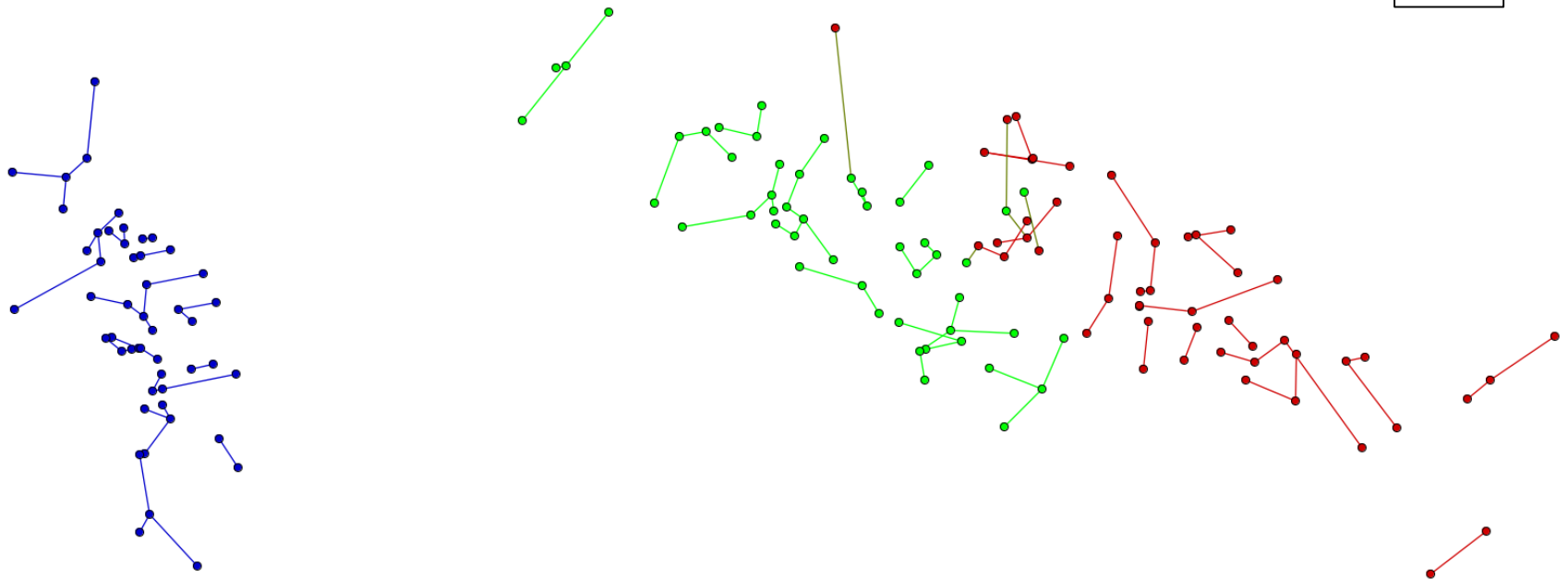


Análise de Modelos de Aprendizagem de Máquina

■ Análise de algoritmos de classificação

□ Ex. k-NN

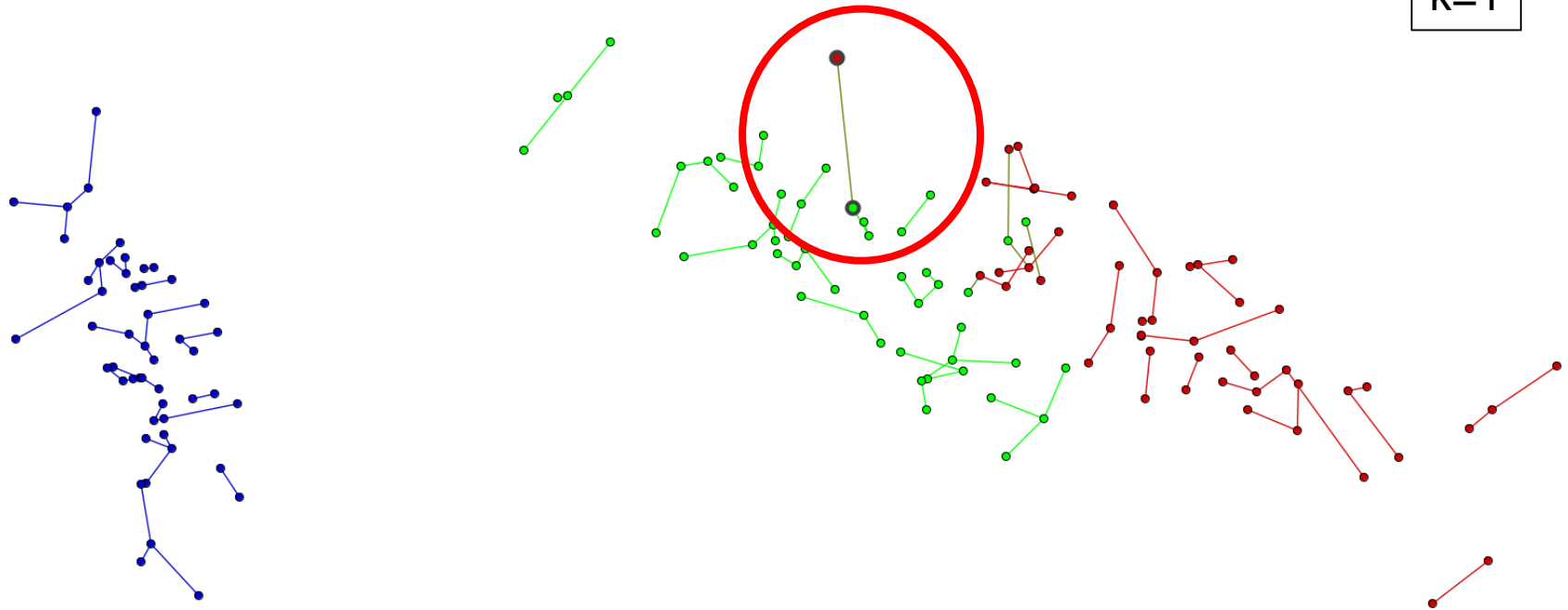
k=1



Análise de Modelos de Aprendizagem de Máquina

■ Análise de algoritmos de classificação

□ Ex. k-NN

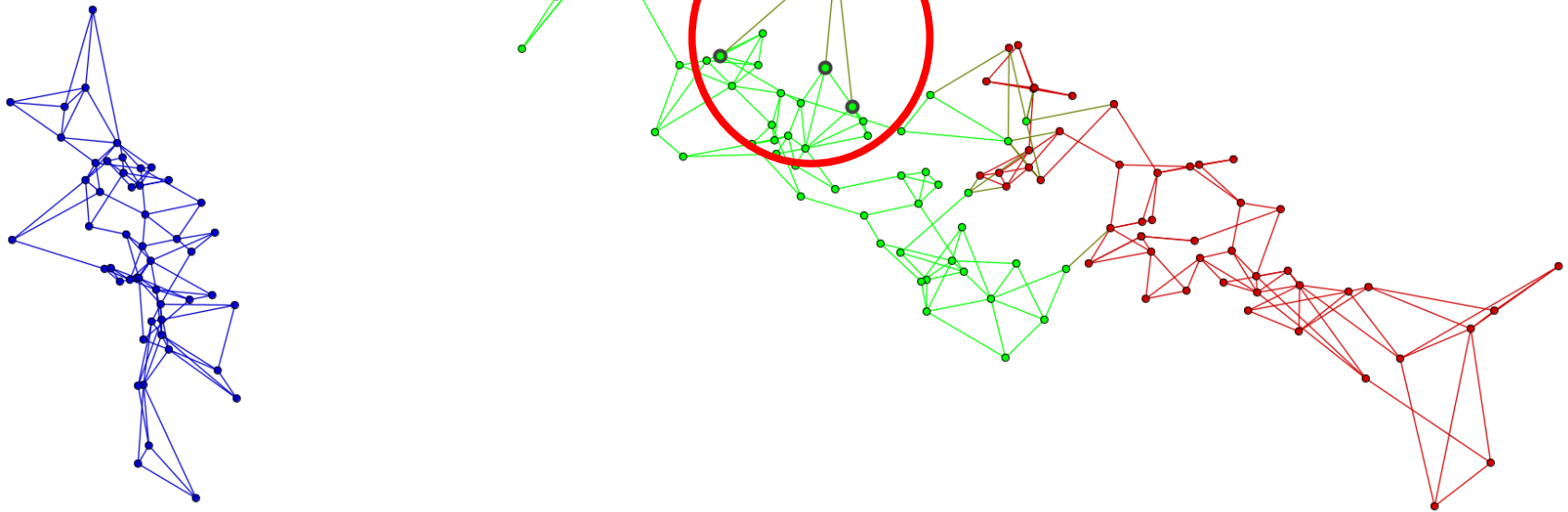


Análise de Modelos de Aprendizagem de Máquina

■ Análise de algoritmos de classificação

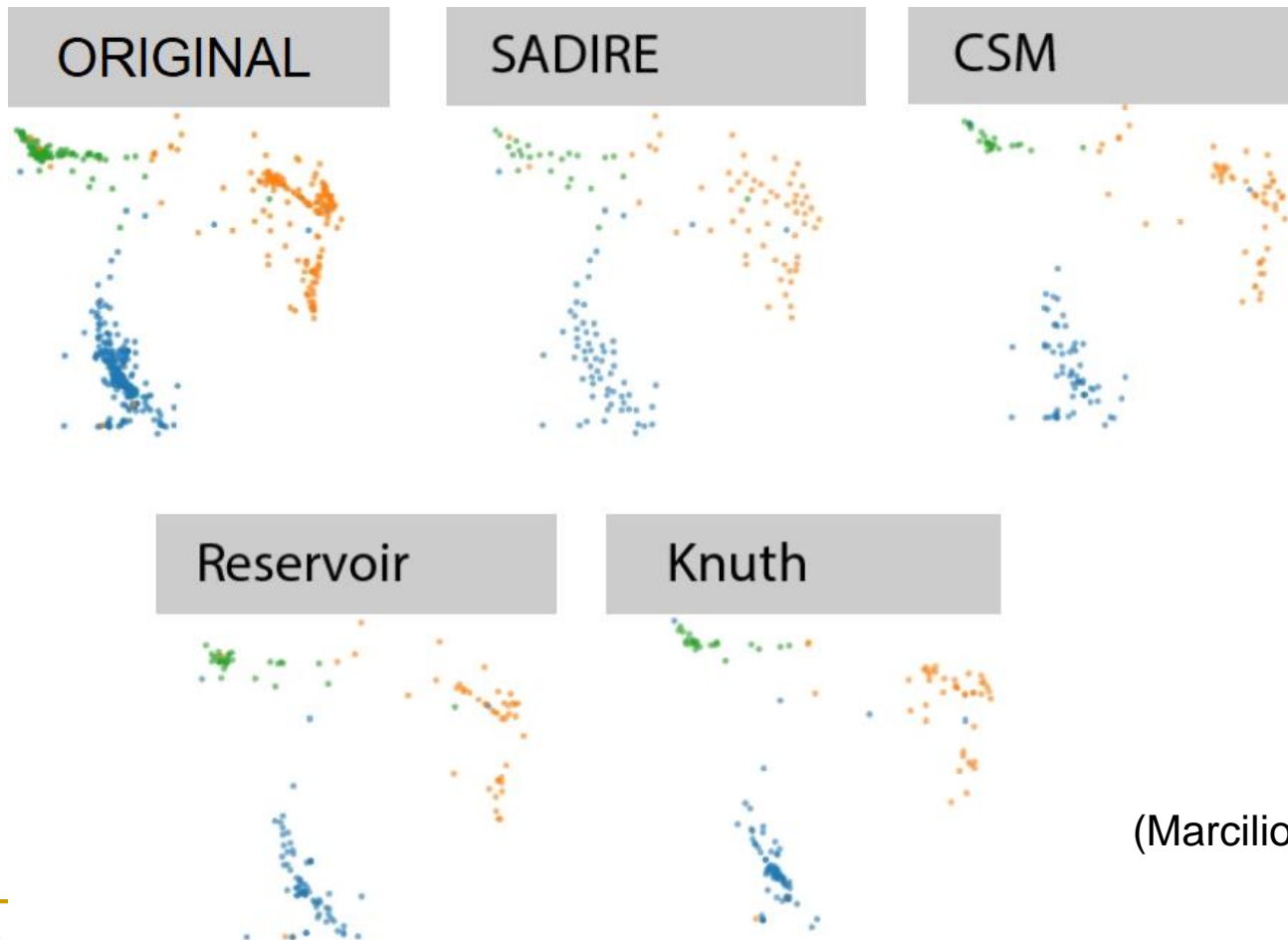
□ Ex. k-NN

k=3



Amostragem

- Comparação de diferentes técnicas de amostragem (seleção de representativos)



(Marcilio-Jr e Eler, 2020)

Conclusões

- A visualização colabora com a exploração de conjuntos de dados
- É preciso estudar os principais conceitos e as técnicas para saber aplicá-las
- É uma ferramenta útil para tarefas de análise de dados e de aprendizagem de máquina
- A visualização tem sido empregada em Explainable Artificial Intelligence (XAI)

Contato

- Página


- <https://daniloeler.github.io/>

- Email

- danilo.eler@unesp.br


Introdução à Visualização

■ <https://daniloeler.github.io/>

 **unesp**

Danilo Medeiros Eler

Pesquisas [Ensino](#)



Danilo Medeiros Eler
Assistant Professor

Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP)
Faculdade de Ciências e Tecnologia
Departamento de Matemática e Computação
[Currículo Lattes](#)

Education

Postdoc ICMC-USP - São Carlos, Brazil, 2012
Ph.D. ICMC-USP - São Carlos, Brazil, 2011
MsC. ICMC-USP - São Carlos, Brazil, 2006
B.Sc. FIPP-Unoeste - Presidente Prudente, Brazil, 2001

PESQUISAS (RESEARCH)

[Currículo Lattes](#)

[Videos](#)

[GitHub](#)

[RADAR COVID-19](#)

[GoogleMyCitations](#)

ENSINO (TEACHING)

[Visualização de Informação](#)
2o Semestre

[Projeto e Análise de Algoritmos](#)
2o Ano / 2o Semestre

[Introdução à Ciência da Computação](#)
1o Ano / 1o Semestre

[Programação Orientada a Objetos](#)
2o Ano / 1o Semestre

Introdução à Visualização

Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp)

Faculdade de Ciências e Tecnologia de Presidente Prudente (FCT/Unesp)

Departamento de Matemática e Computação (DMC)

Programa de Pós-Graduação em Ciência da Computação (PPGCC/Unesp)

Visualização de Informação -- Turma 2020

Prof. Dr. Danilo Medeiros Eler

Aulas

1. Introdução à Visualização
 - a. [Parte 1 - Visão Geral \(video\)](#)
 - b. [Parte 1 - Visão Geral - slides de aula \(pdf\)](#)
 - c. [Parte 2 - Processo de Visualização \(video\)](#)
 - d. [Parte 2 - Processo de Visualização - slides de aula \(pdf\)](#)
2. Fundamentos Sobre Dados
 - a. [Parte 1 - Tipos de Dados \(video\)](#)
 - b. [Parte 1 - Tipos de Dados - slides de aula \(pdf\)](#)
 - c. [Parte 2 - Estrutura dentro e entre instâncias \(video\)](#)
 - d. [Parte 2 - Estrutura dentro e entre instâncias - slides de aula \(pdf\)](#)
 - e. [Parte 3 - Processamento dos dados \(video\)](#)
 - f. [Parte 3 - Processamento dos dados - slides de aula \(pdf\)](#)
 - g. [Parte 3 - Processamento dos dados - exemplos com python \(html\)](#)
3. Fundamentos de Visualização
 - a. [Parte 1 - Variáveis Visuais \(video\)](#)
 - b. [Parte 1 - Variáveis Visuais - slides de aula \(pdf\)](#)
 - c. [Parte 2 - Taxonomia \(video\)](#)
 - d. [Parte 2 - Taxonomia - slides de aula \(pdf\)](#)
4. Técnicas de Visualização para Dados Multivariados
 - a. [Parte 1 - Técnicas Orientadas a Pixel e Técnicas Iconográficas \(video\)](#)
 - b. [Parte 1 - Técnicas Orientadas a Pixel e Técnicas Iconográficas - slides de aula \(pdf\)](#)
 - c. [Parte 2 - Técnicas Baseadas em Grafos e Técnicas Hierárquicas \(video\)](#)
 - d. [Parte 2 - Técnicas Baseadas em Grafos e Técnicas Hierárquicas - slides de aula \(pdf\)](#)
 - e. [Parte 3 - Técnicas de Projeção Geométrica \(video\)](#)
 - f. [Parte 3 - Técnicas de Projeção Geométrica - slides de aula \(pdf\)](#)
5. Visualização de Dados Não Estruturados
 - a. [Parte 1 - Coleções de Documentos \(video\)](#)
 - b. [Parte 1 - Coleções de Documentos - slides de aula \(pdf\)](#)
 - c. [Parte 2 - Coleções de Imagens \(video\)](#)
 - d. [Parte 2 - Coleções de Imagens - slides de aula \(pdf\)](#)
6. Conceitos de Interação
 - a. [Conceitos de Interação \(video\)](#)
 - b. [Conceitos de Interação - slides de aula \(pdf\)](#)

Referências

- Ward, M., Grinstein, G. G., Keim, D. Interactive data visualization foundations, techniques, and applications. Natick, Mass., A K Peters, 2010.
- Robert Spence. Information Visualization: Design for Interaction. 2nd Edition. Pearson : Prentice Hall, 2007
- Alexandru C Telea. Data visualization: principles and practice. Boca Raton : CRC Press, 2015.

Referências

- Peter Lyman and Hal R. Varian, How Much Information, 2003;
ww2.sims.berkeley.edu/research/projects/how-much-info/
- Michael Friendly's web site
 - <http://www.datavis.ca/>
- TAN. Text mining: The state of the art and challenges. 1999. Disponível em:
citeseer.ist.psu.edu/tan99text.html (acessado em 2006.08.10)

Referências

- Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. “The Kdd Process for Extracting Useful Knowledge from Volumes of Data.” Commun. ACM 39 (11). New York, NY, USA: ACM: 27–34, 1996
- J. G. Williams, K. M. Sochats, and E. Morse. “Visualization.” Annual Review of Information Science and Technology (ARIST) 30, 161–207, 1995
- Jennifer Tanaka. Drowning in Data, Newsweek, 4/28/1998, p. 85

Referências

- W. E. Marcilio-Jr, Danilo M. Eler. SADIRE: a context-preserving sampling technique for dimensionality reduction visualizations. Journal of Visualization 23 (6), 999-1013, 2020