

AI & CHATBOT

Aula 15 – Visualização de Dados
com Python

Prof. Henrique Ferreira

FIAP
GRADUAÇÃO

Visualização de Dados

Uma introdução

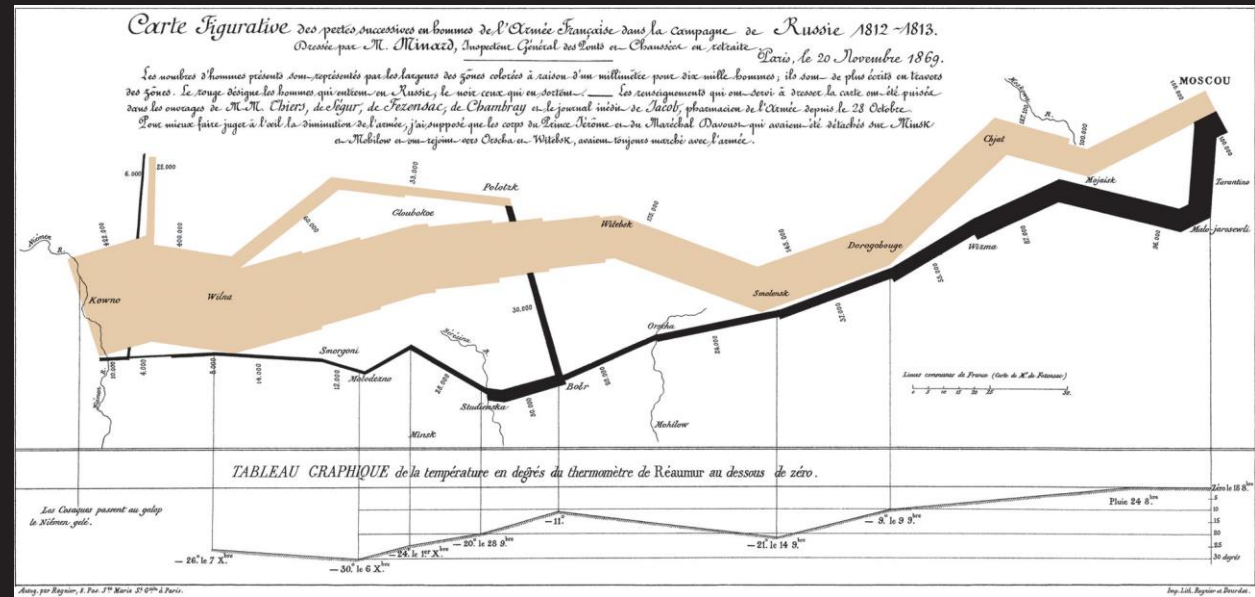
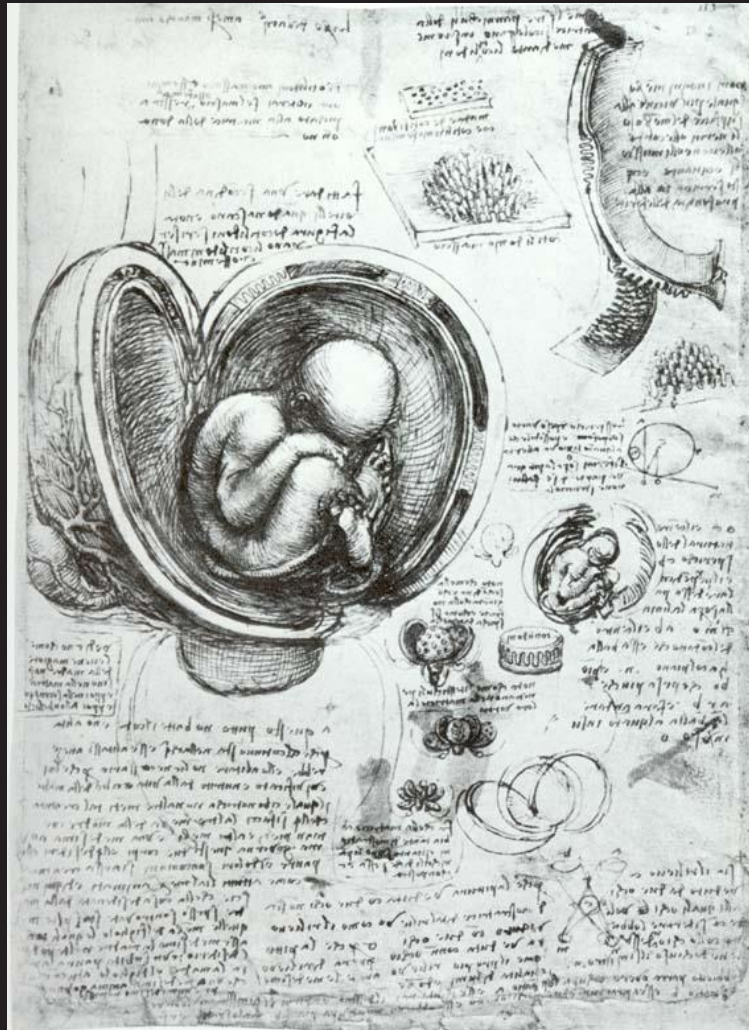
Visualização de Dados

- A visualização de dados é toda uma área do conhecimento que envolve pensar na melhor maneira de representar informações abstratas com objetivo de:
 1. Facilitar a comunicação;
 2. Gerar interpretações e ganho de compreensão;
- É preciso se preocupar em colocar todas as informações necessárias para se compreender o gráfico e também com a qualidade estética do resultado.

Gráficos

Alguns tipos de gráficos, diagramas e infografatos

Infografos



VISUAL COMMUNICATION

Writing with Pictures (Scott McCloud) WEBSTOCK 2011

Cognitive load-time: a little spunky beachball... making comics since 1984

CARTOONING	COMICS
A type of simplified image	Images in deliberate sequence
CLARITY	
• SPEED	• GUIDED FOCUS
• MODULARITY	• PARTICIPATION
• GRAPHIC IMPACT	• CREATIVE ALTERNATIVES
• IDENTIFICATION	• DYNAMISM
• Human Calligraphy	• WORD-PICTURE TO EDUCATION/MONEY

OPTIMISATION
 - Few important details

COMPEL TO PROPEL MO-MENTUM

TELLING STORIES WITH PICTURES

Resolve contours, watch out for non-ambiguous symbols.

"Attention map" needs messages

On fruits + vegetables

IF someone says "that's a terrible elephant" then you've succeeded

Now, we have to say: a snowman is not a mammal.

as you move through space you're moving through time

SEQUENCE

BEAT BEAT BEAT

SILENT BEATS

gacha, the eye blinks between panels.

gimac

FLAT

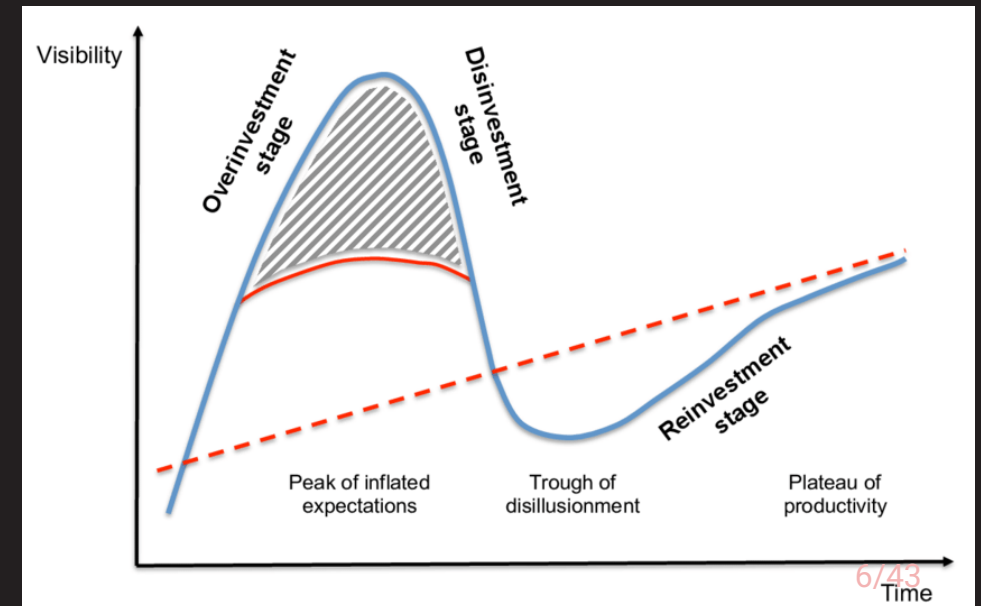
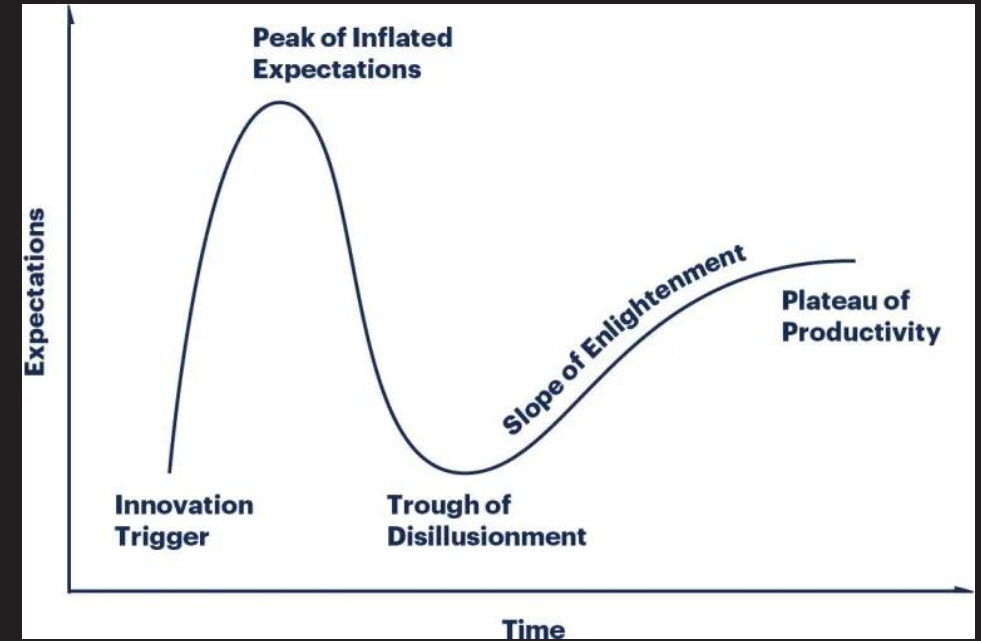
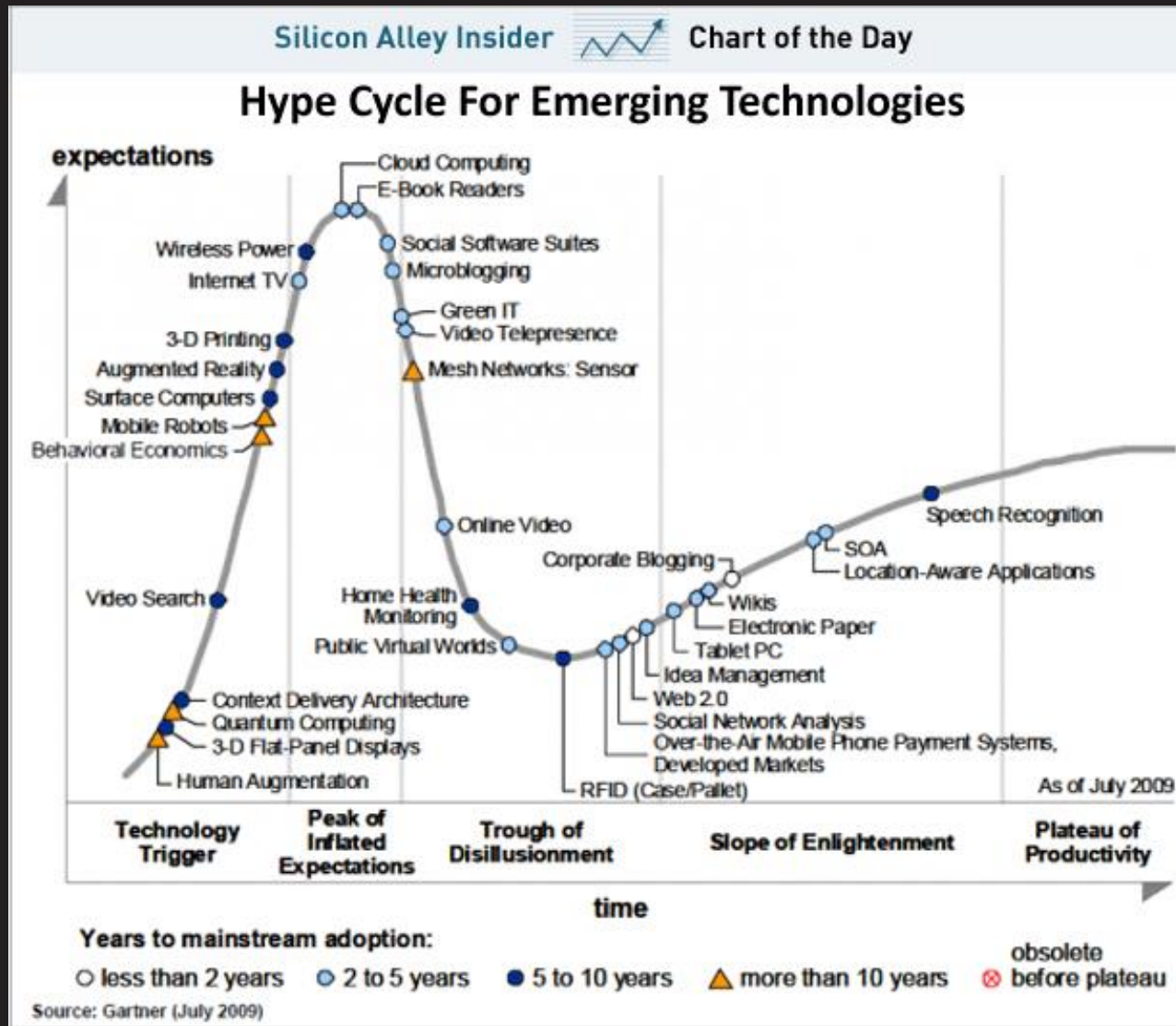
NINE LANTERNS

HANS ROHLING

Tan Gauld (Coke)

ARCADE FIRE

Infografos (Hype cycle)



Infografos (Hype cycle)

Hype Cycle for Emerging Technologies, 2018

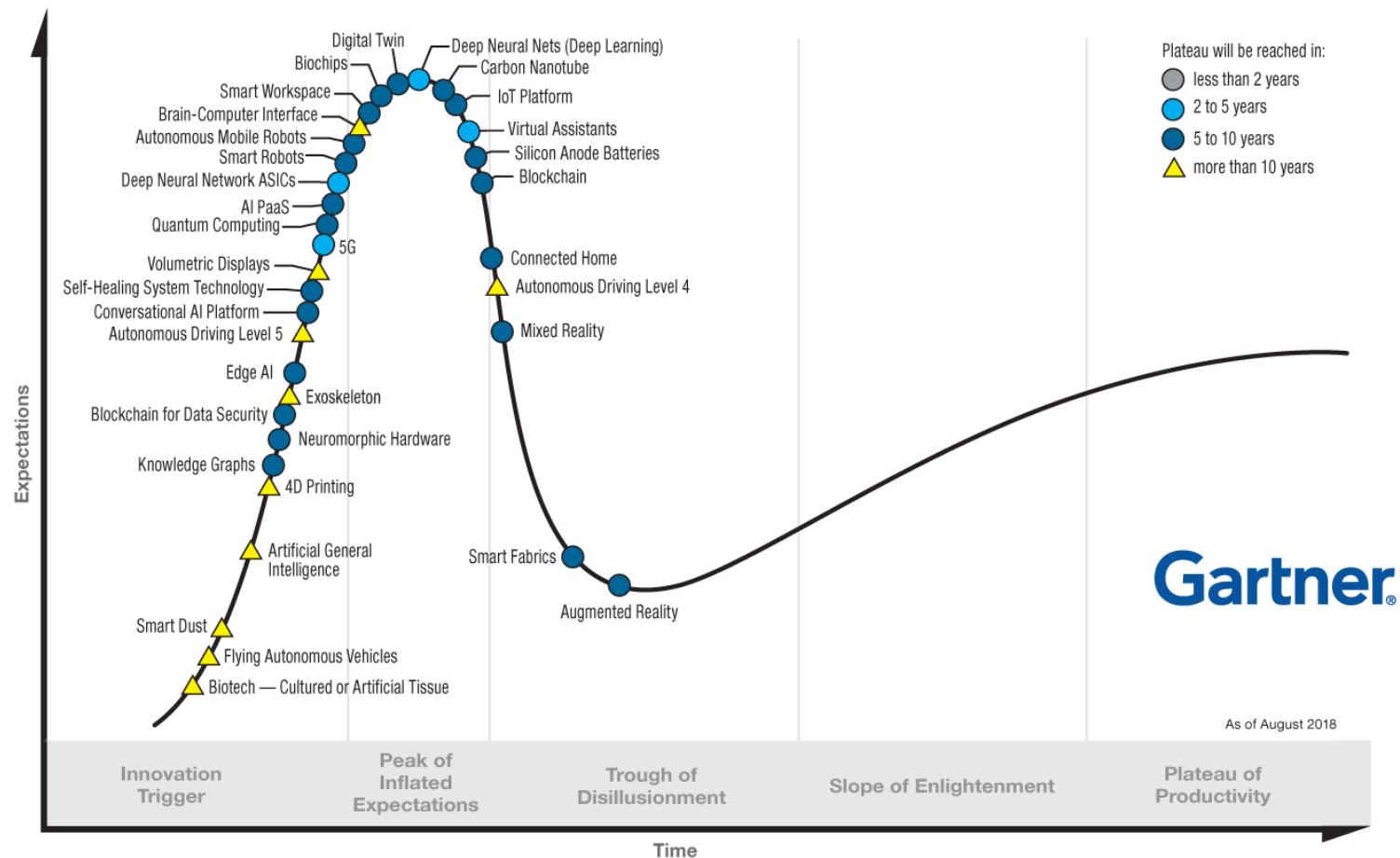


Gráfico de Barras/Colunas (Bar chart)

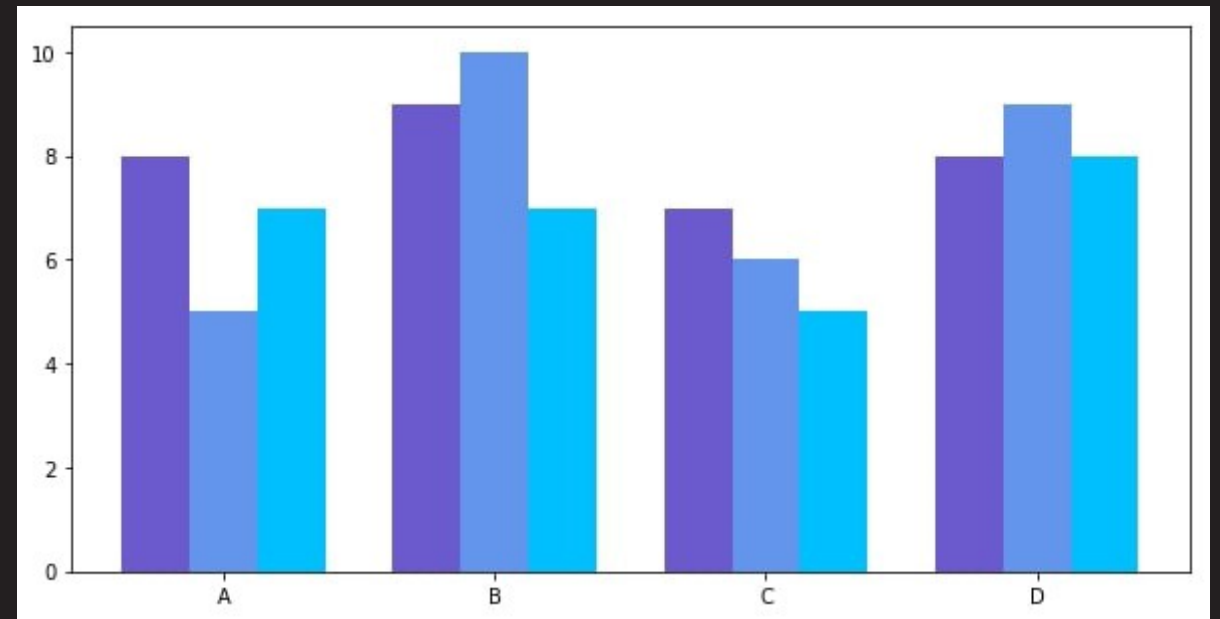
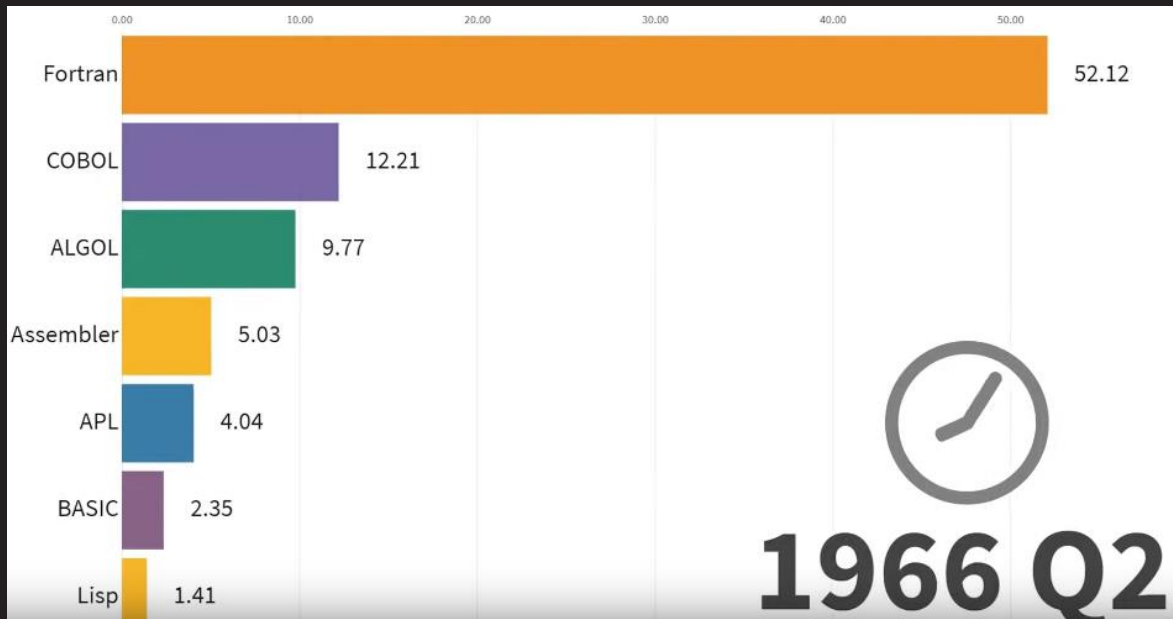
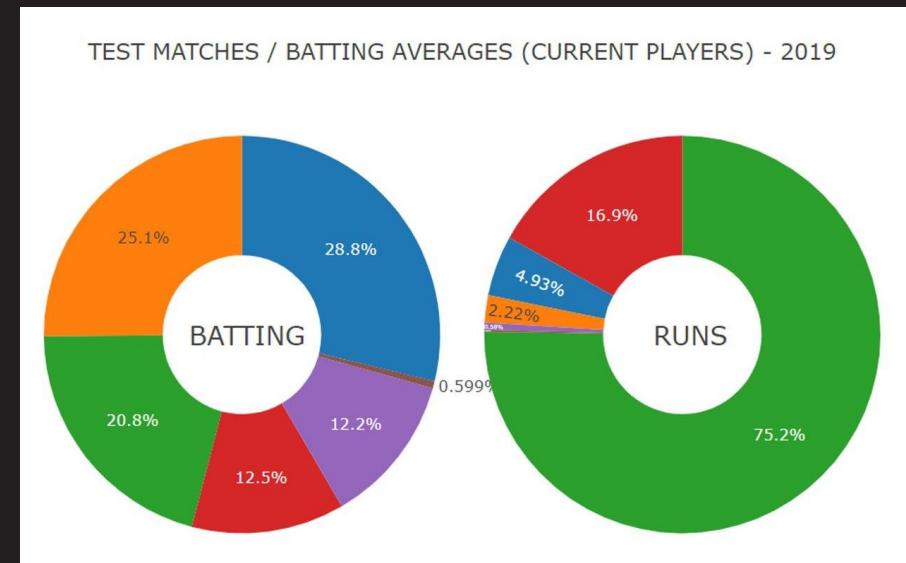
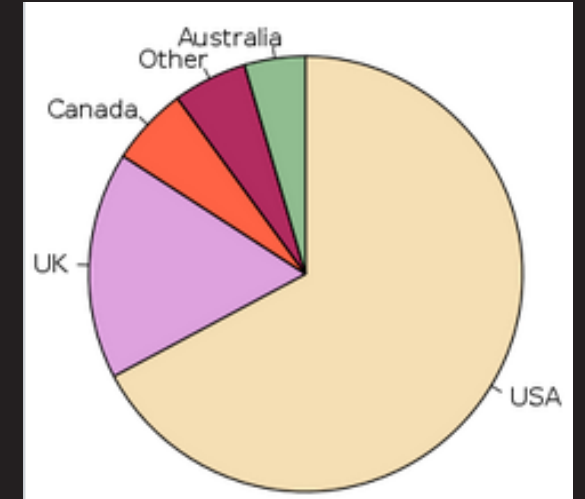
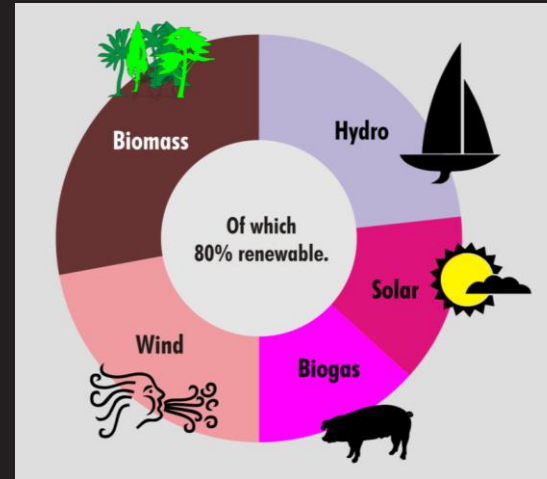
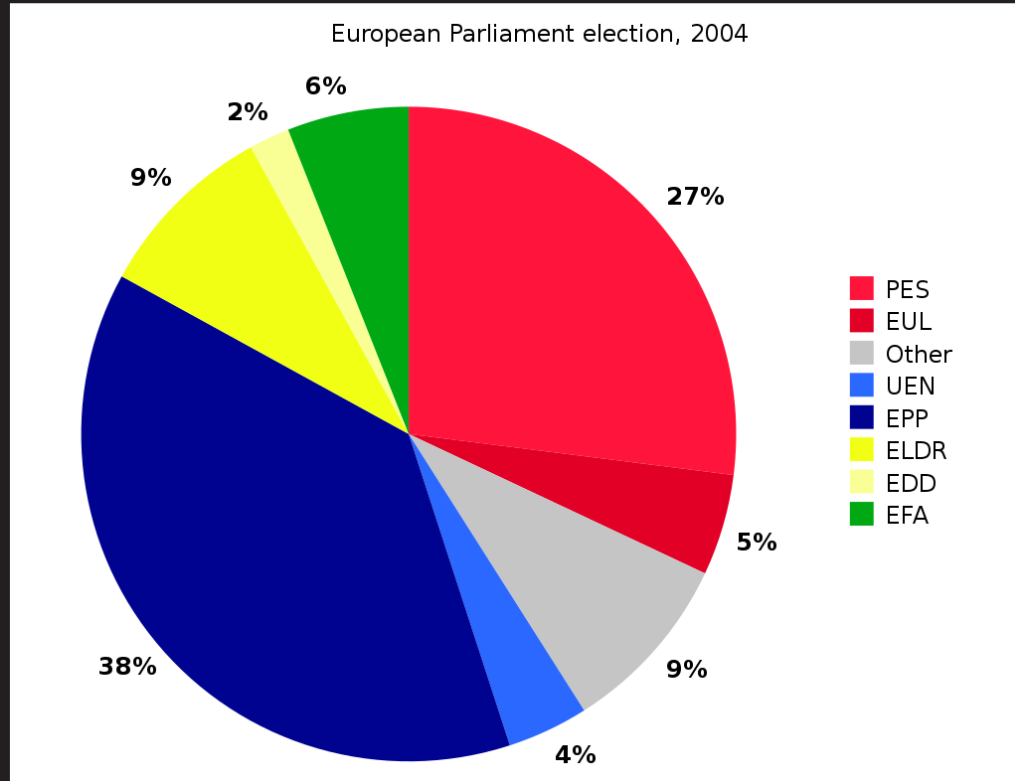
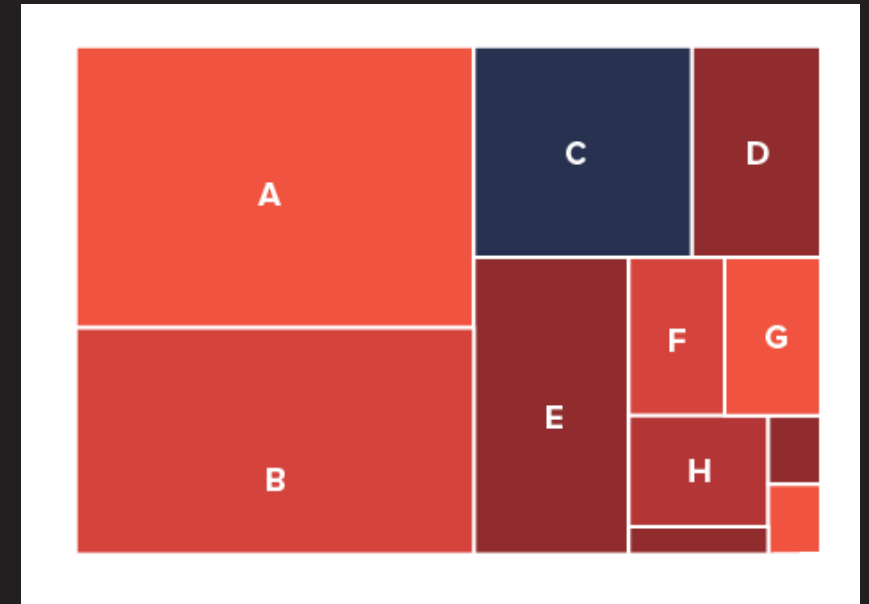
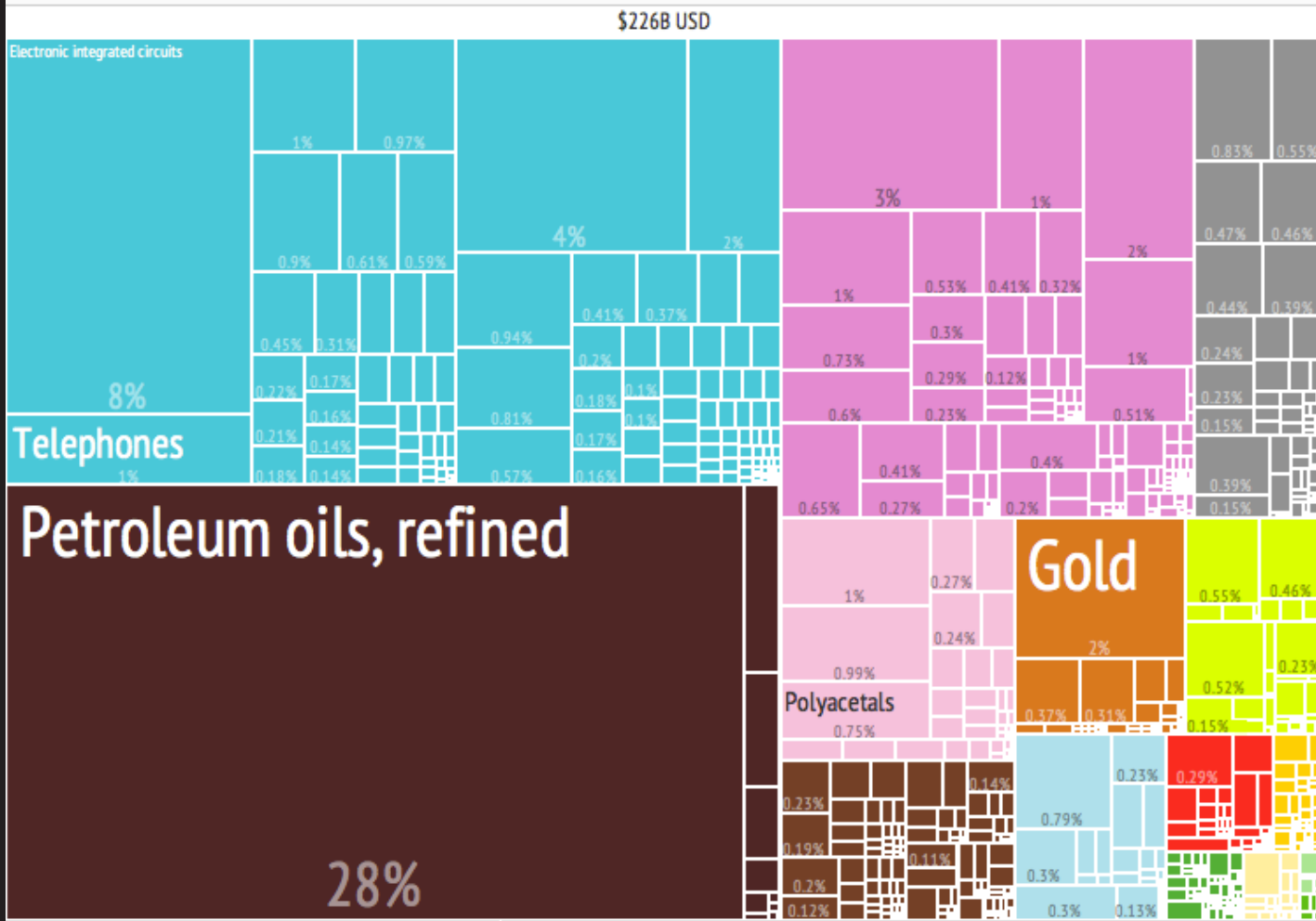


Gráfico de Pizza (Pie chart)

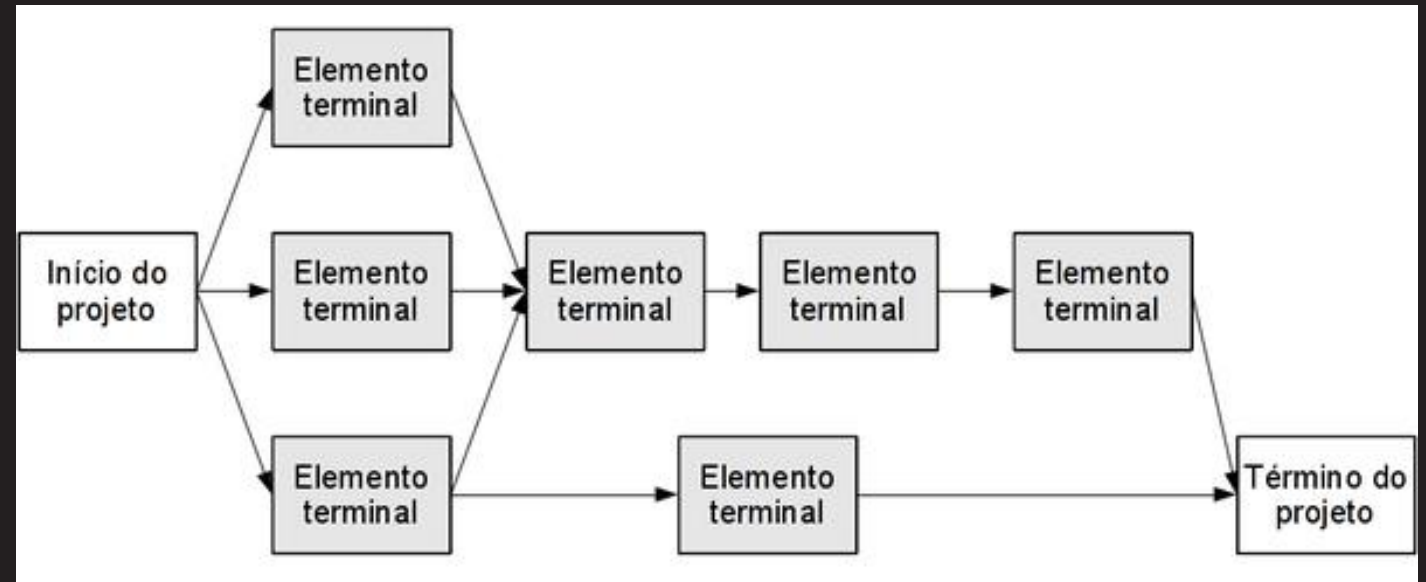
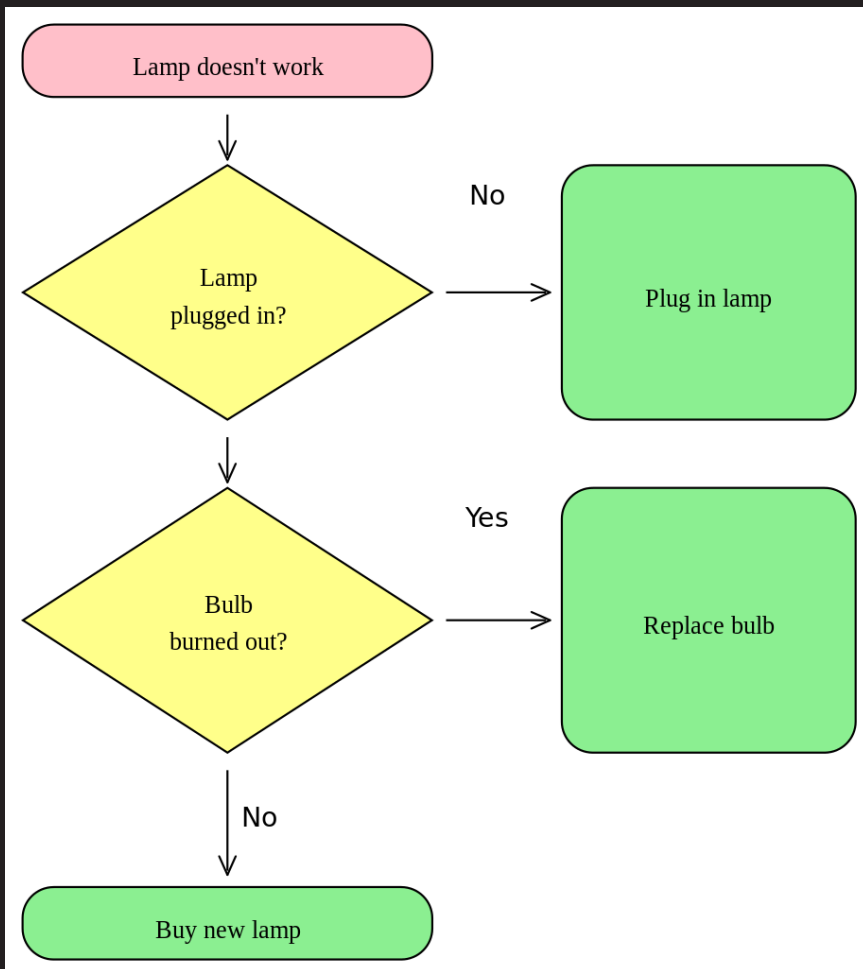


Treemap

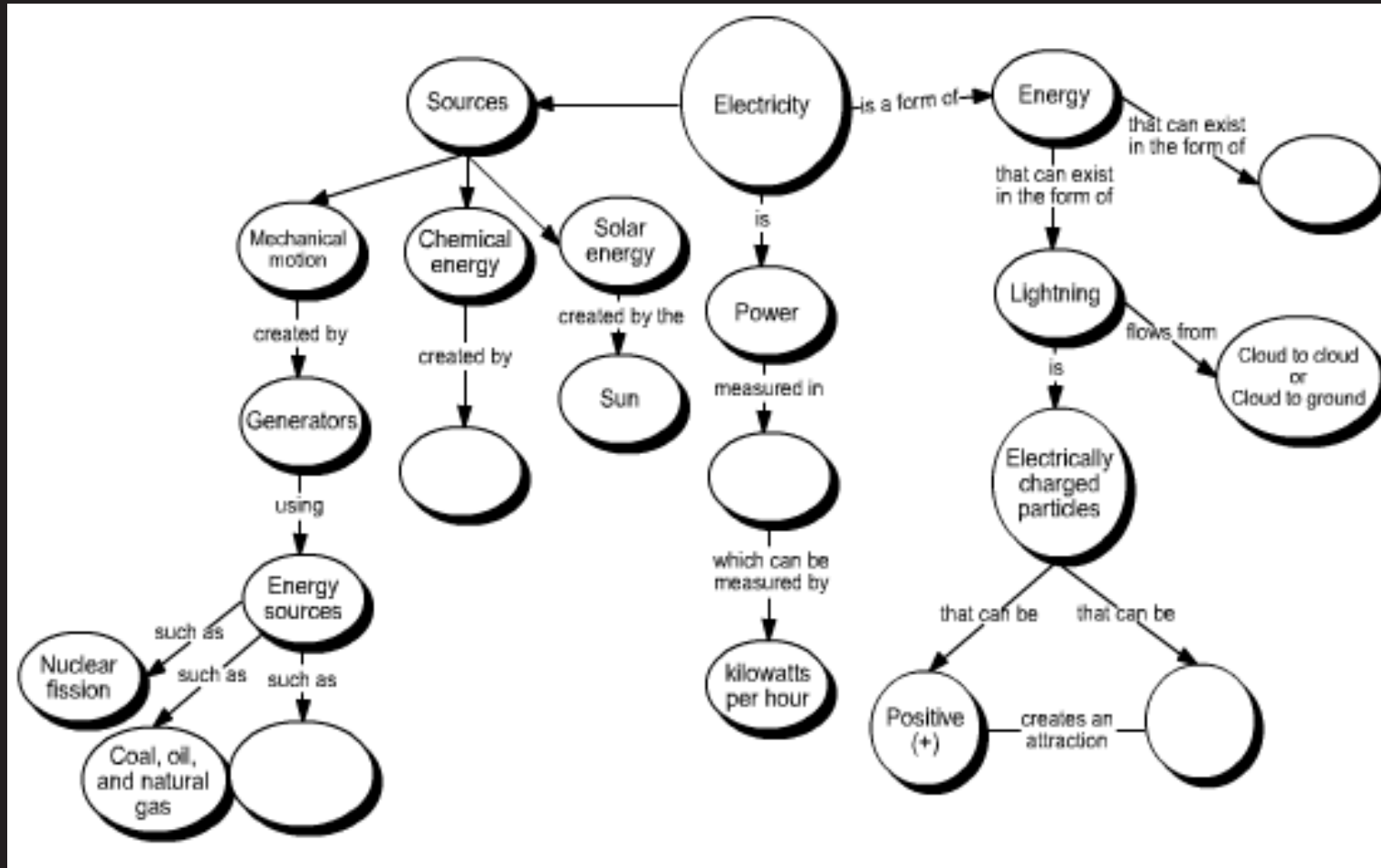
What did Singapore export in 2012?



Fluxograma (Flow chart)



Mapas conceituais (Flow chart)



Dendogramas

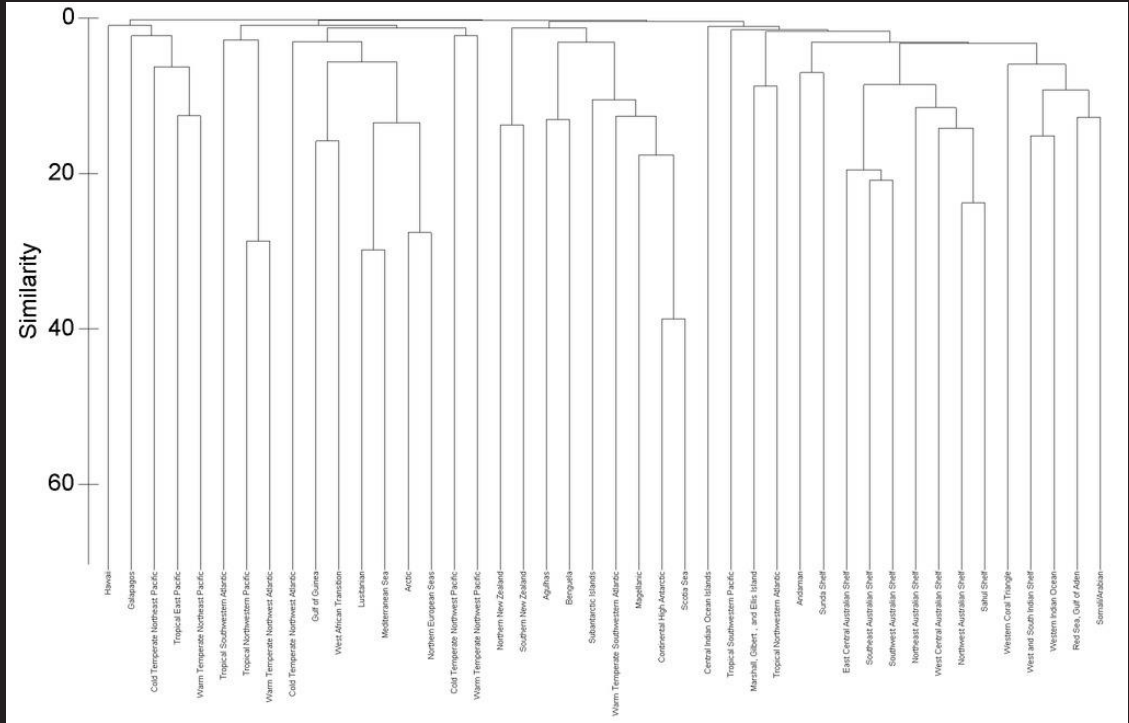
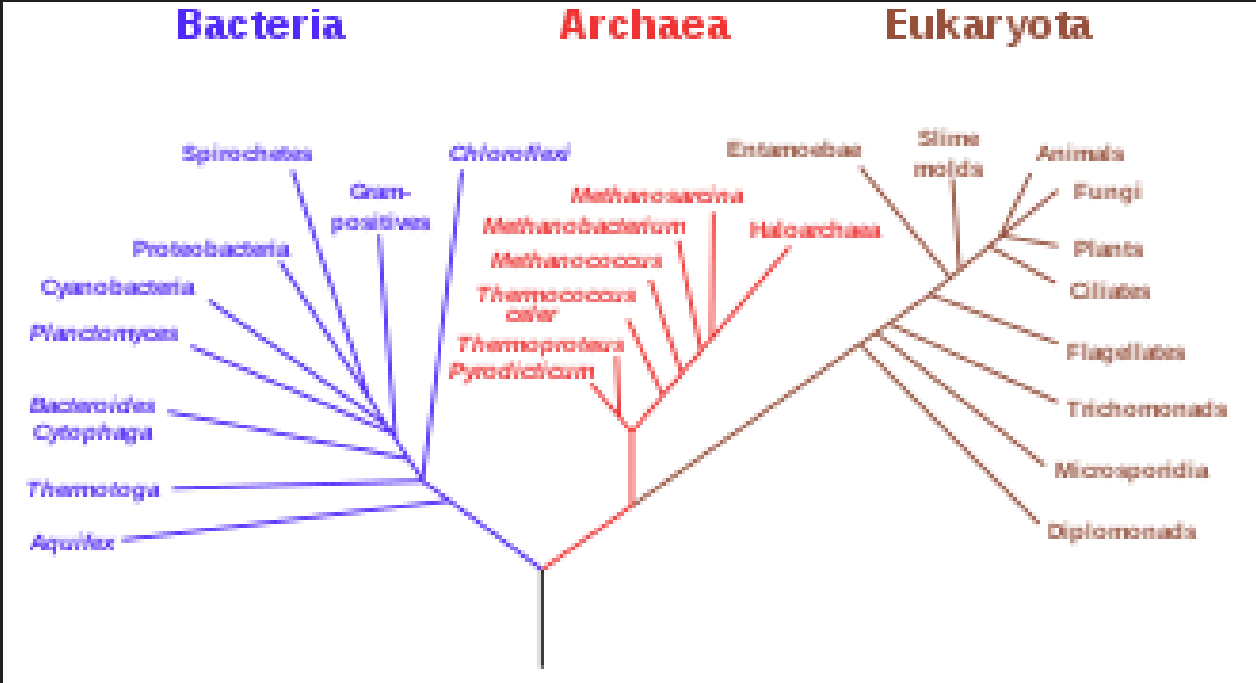


Gráfico de Gantt (Gantt chart)

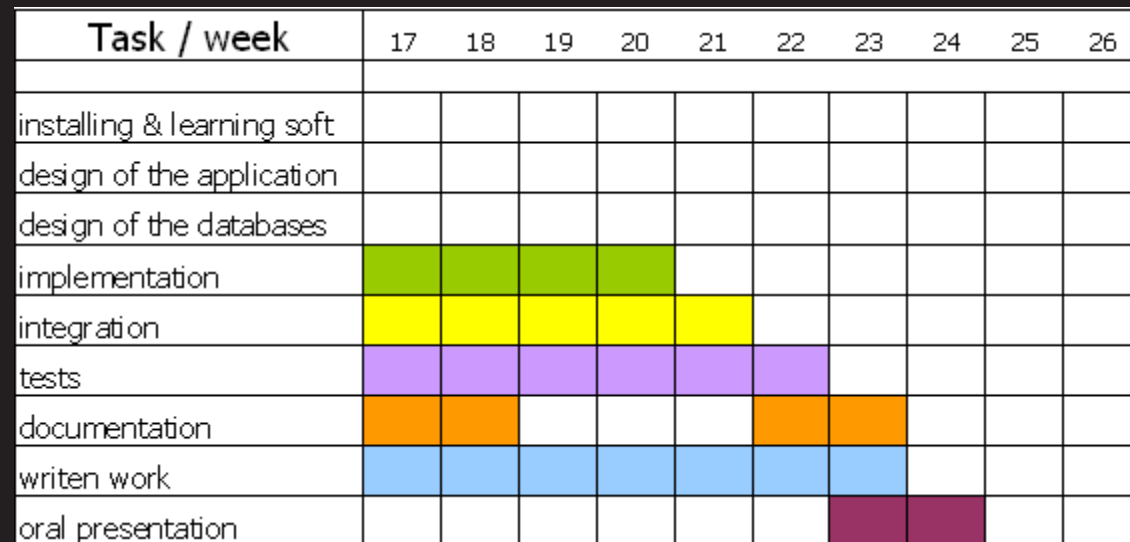
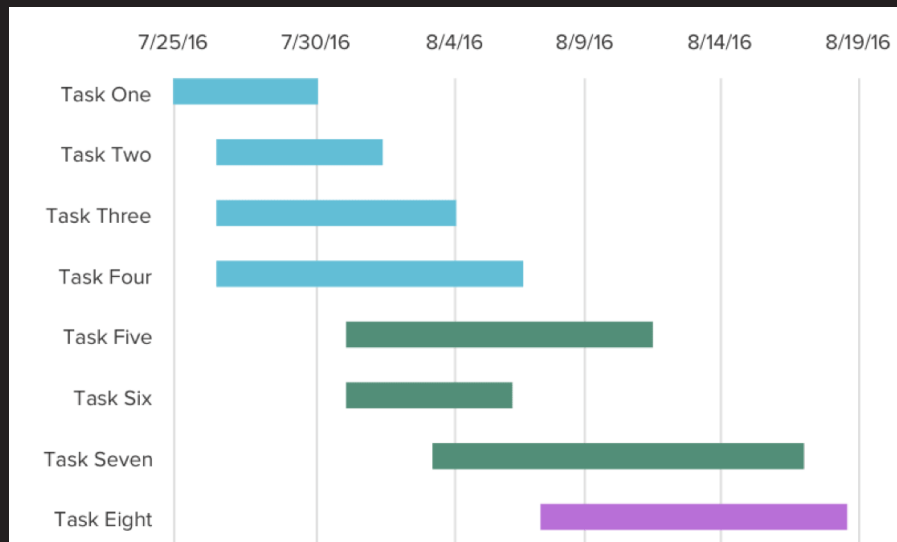
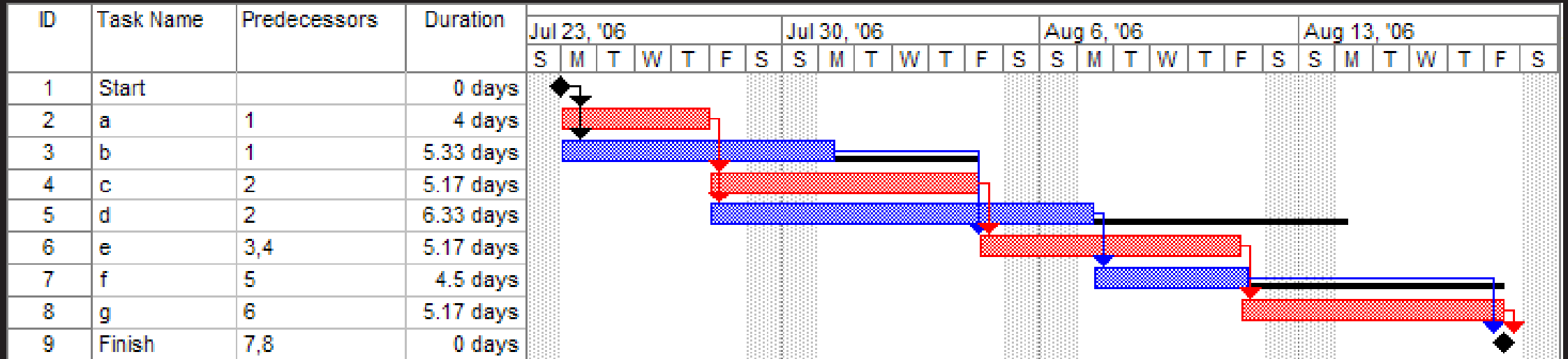
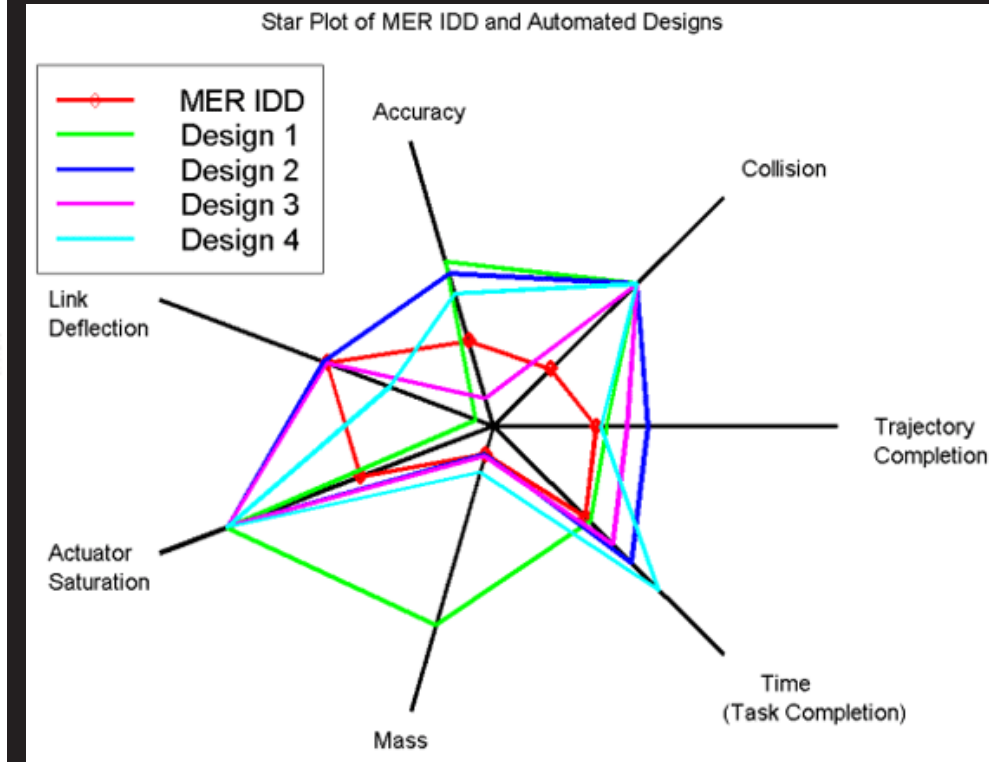
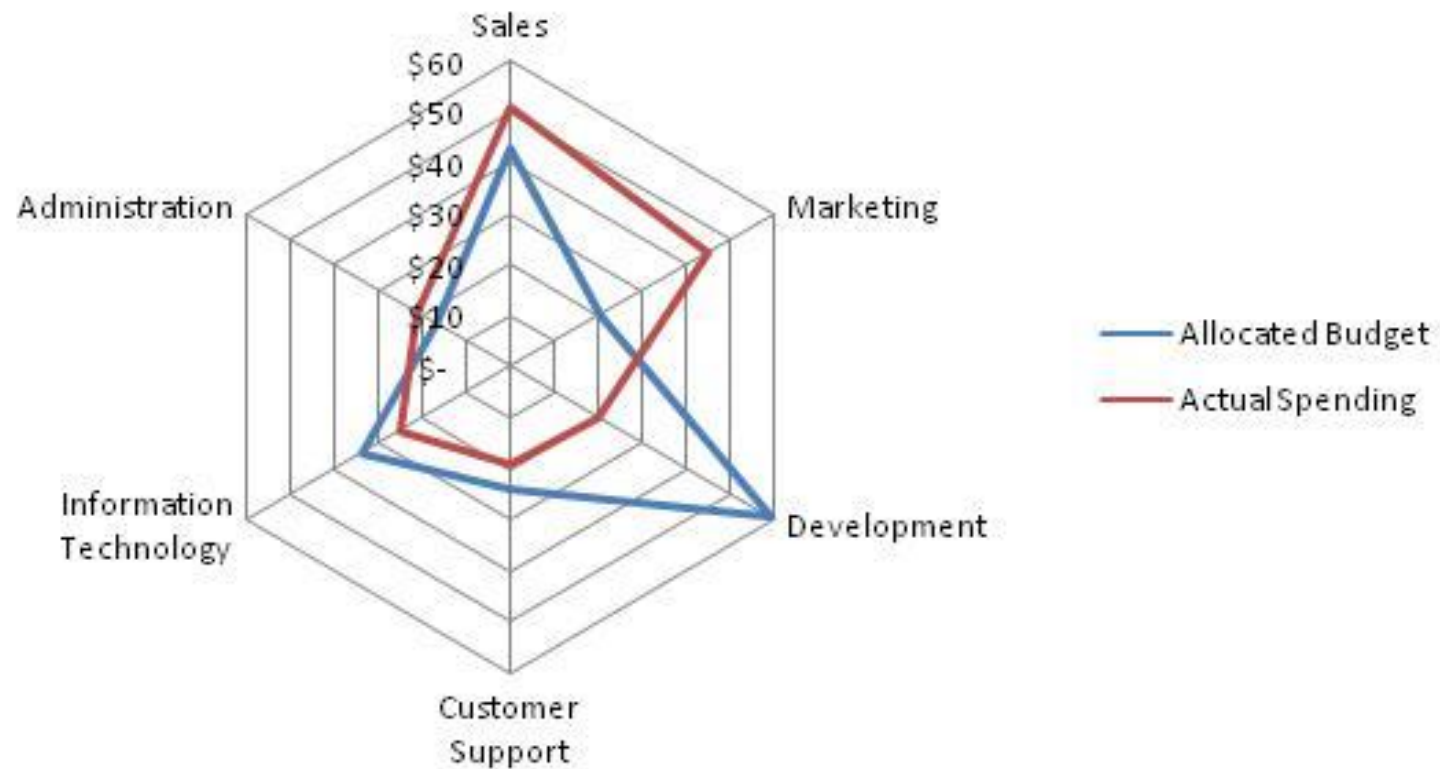
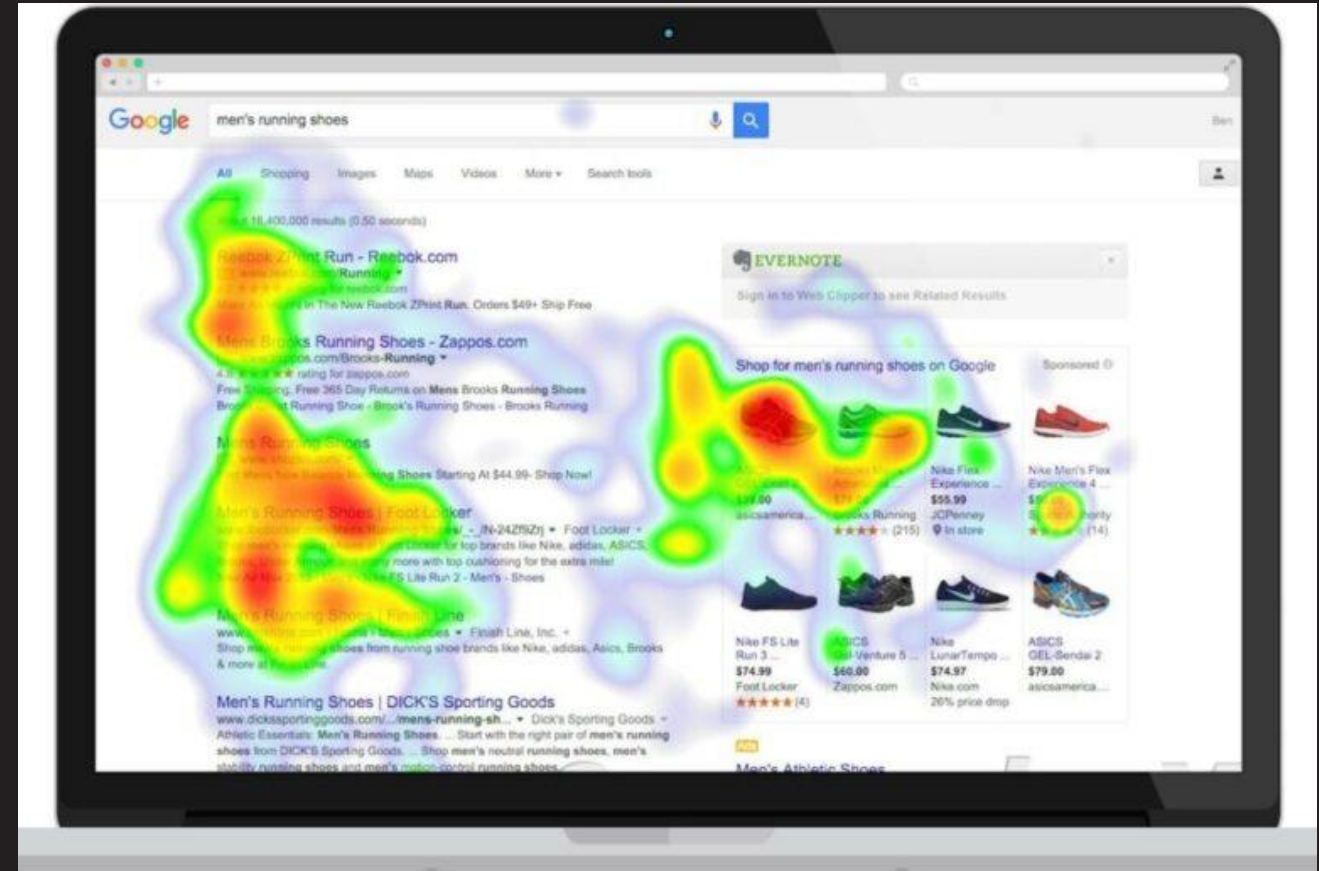
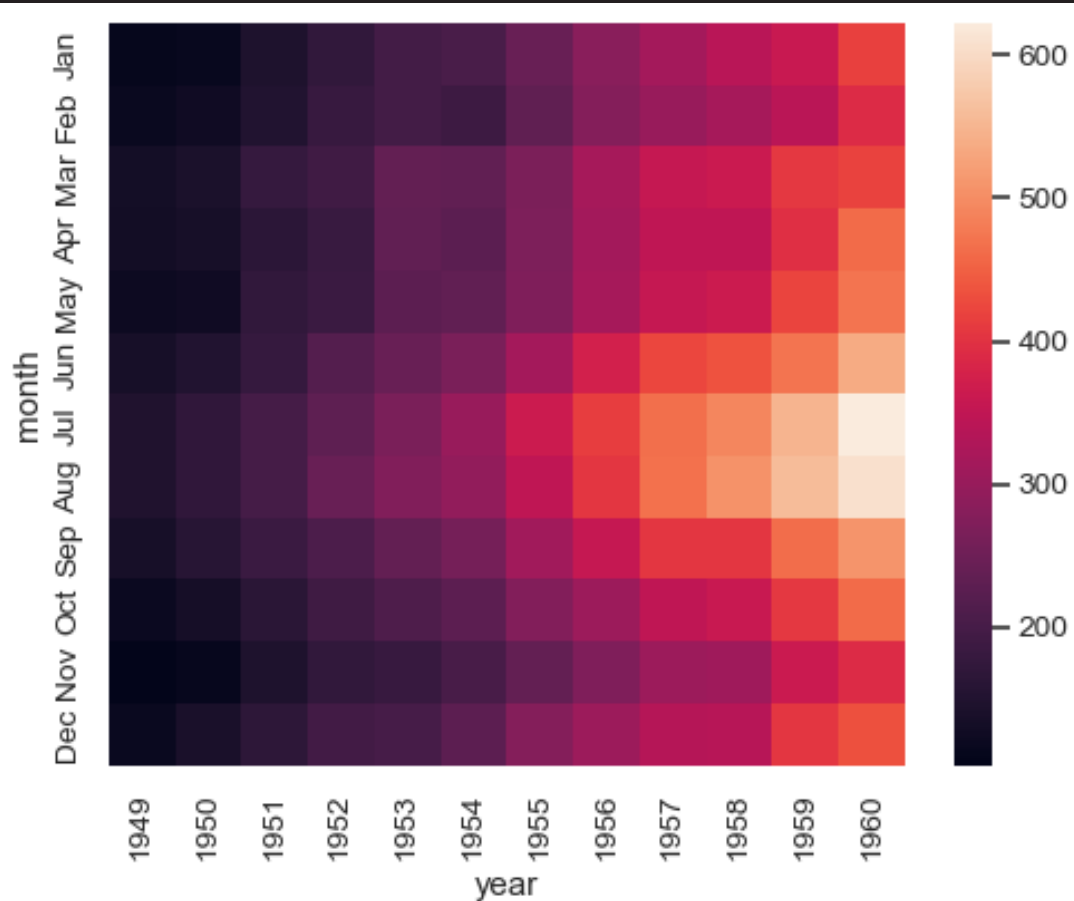


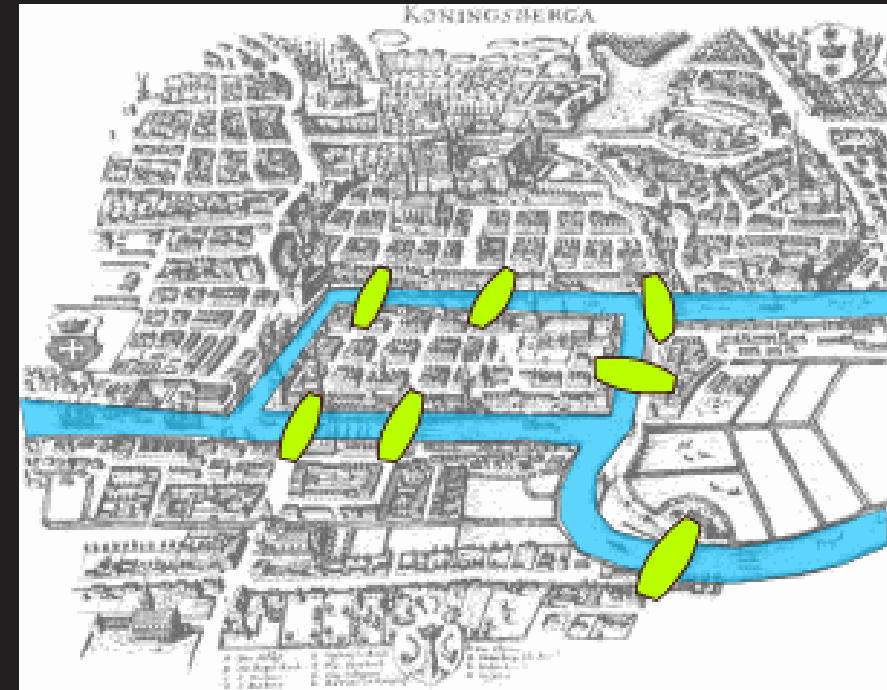
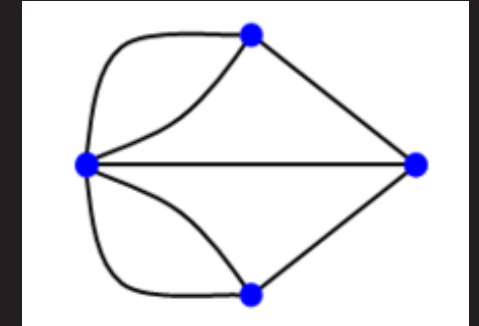
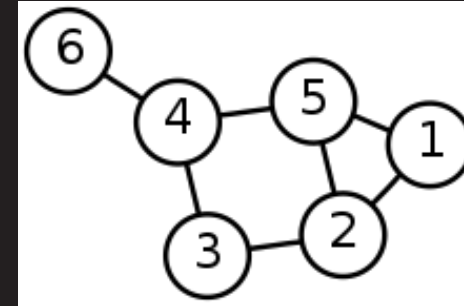
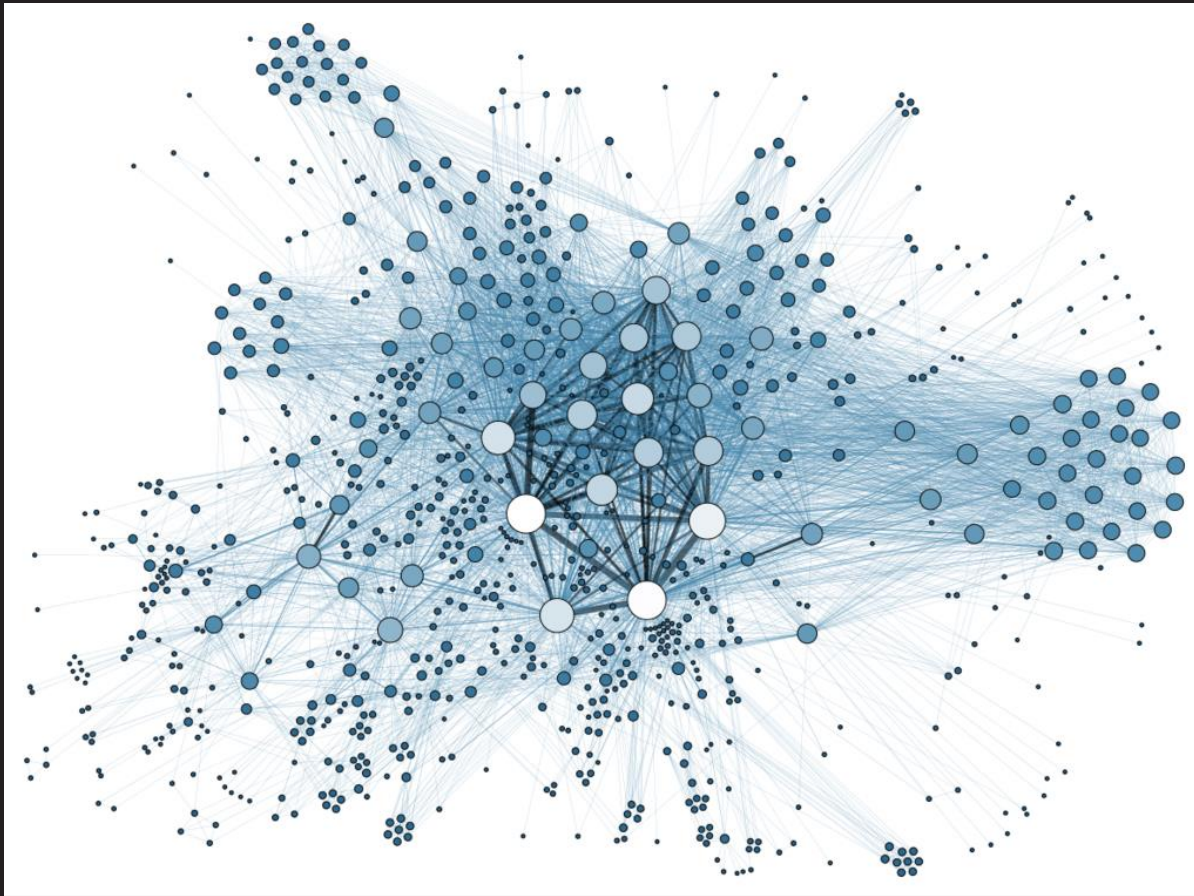
Gráfico de Radar (Radar plot)



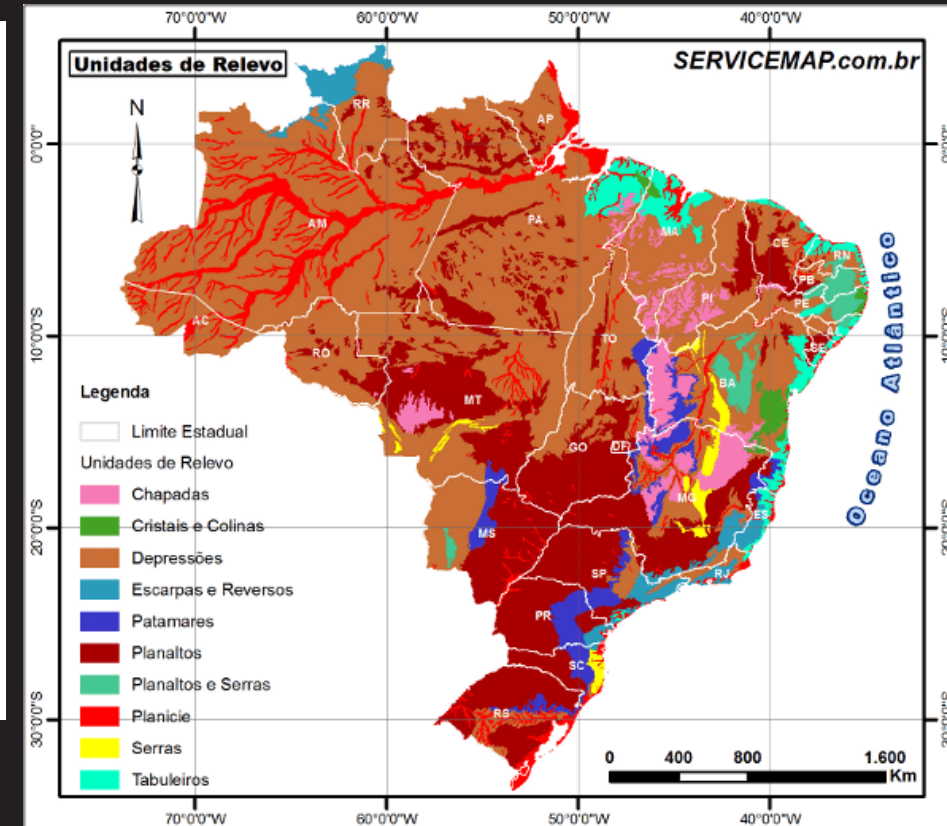
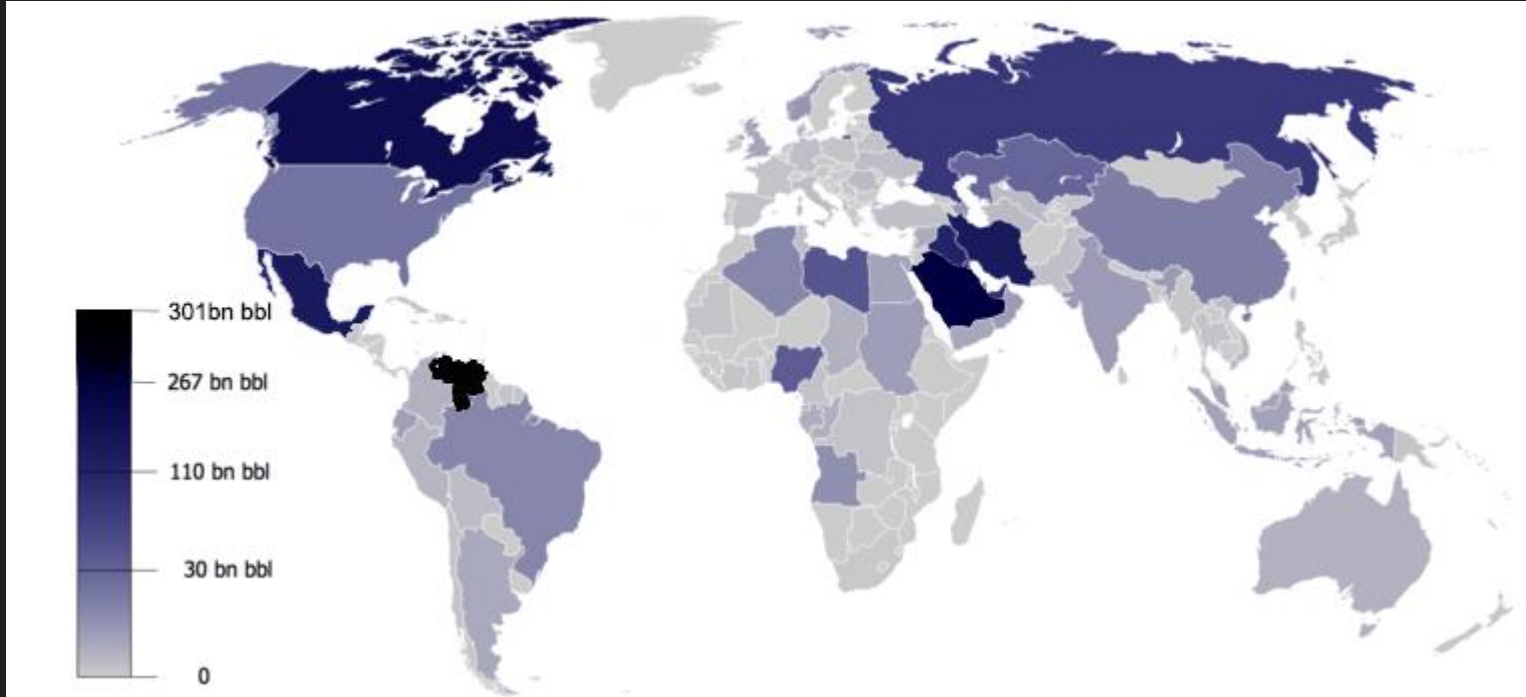
Mapa de calor (Heatmap)



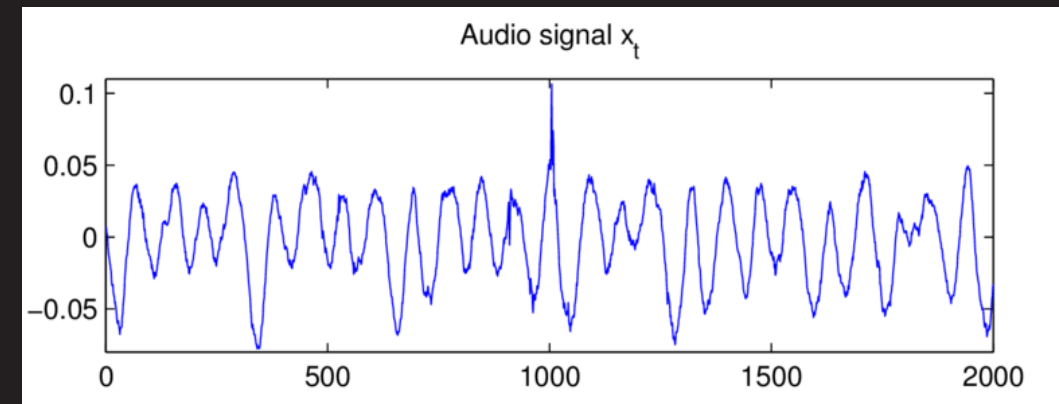
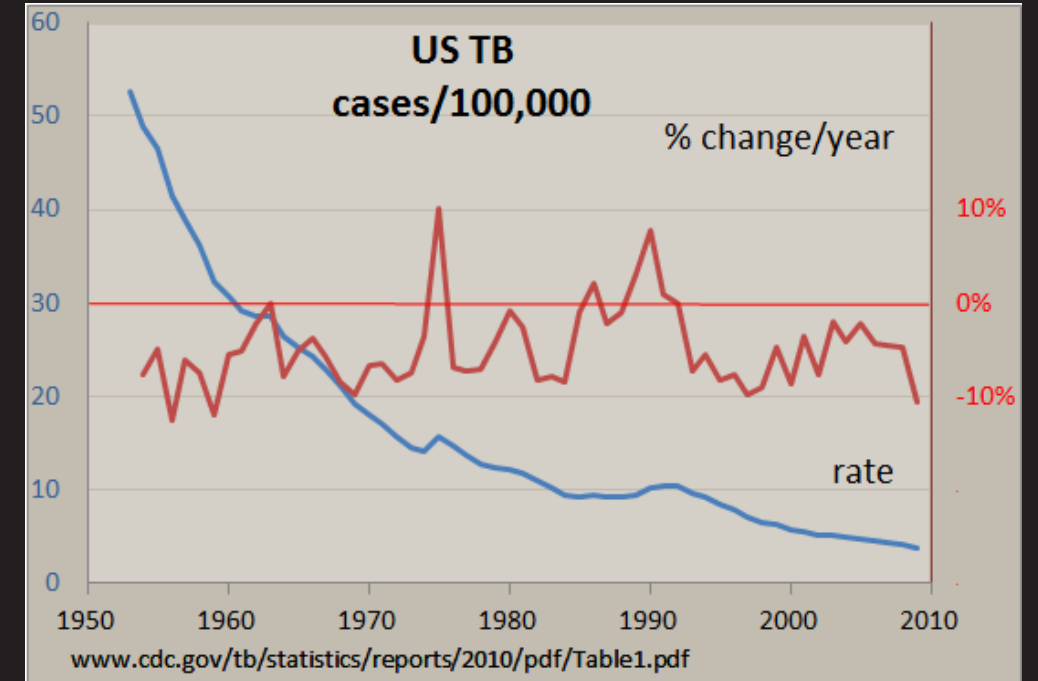
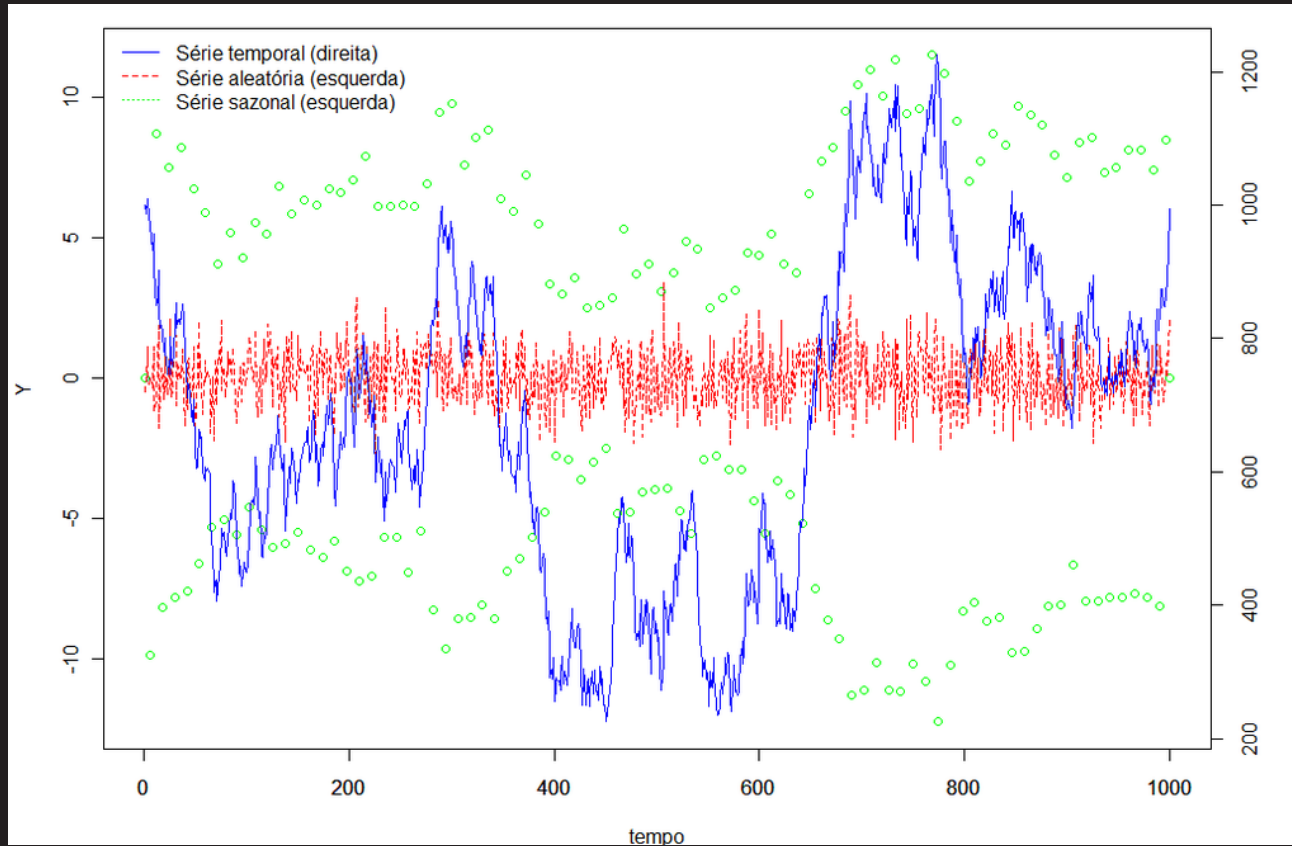
Grafos ou árvores ('Network' plot)



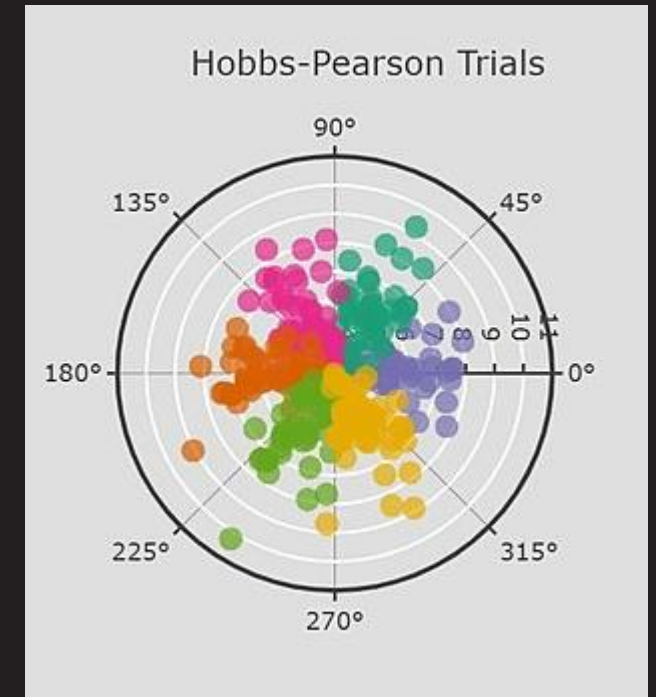
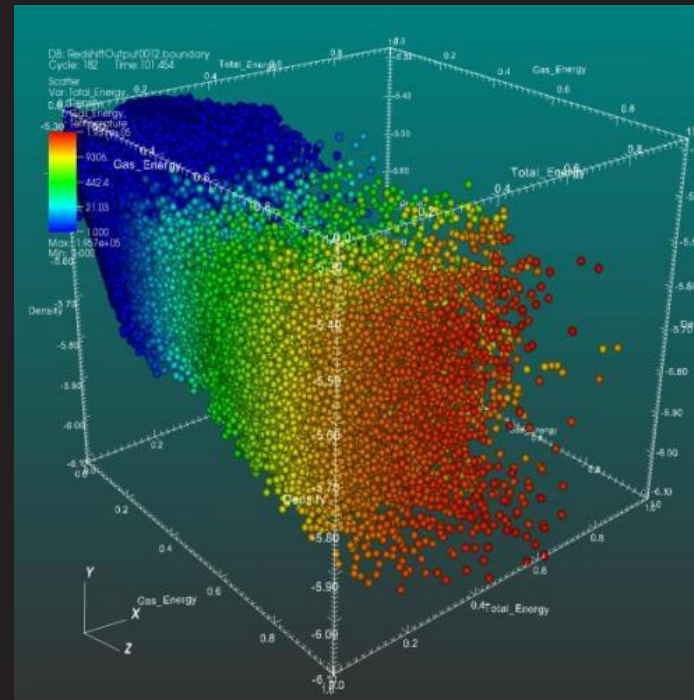
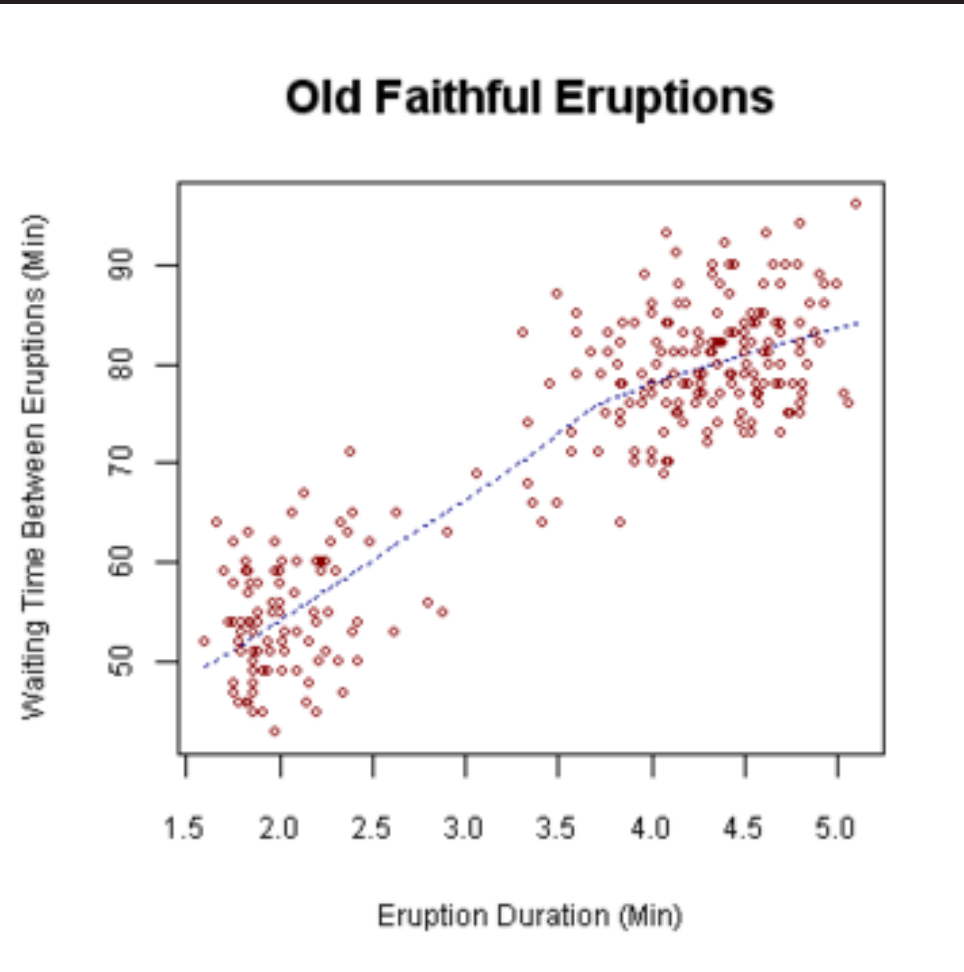
Mapas



Série Temporal



Dispersão (Scatter plot)



Dispersão (Scatter plot)

- Scatter plots permitem identificar correlações e grupos:

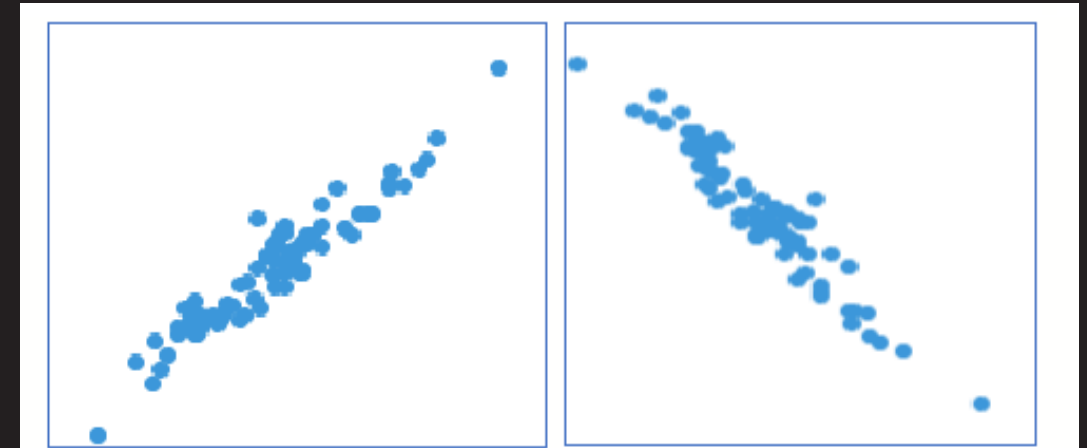
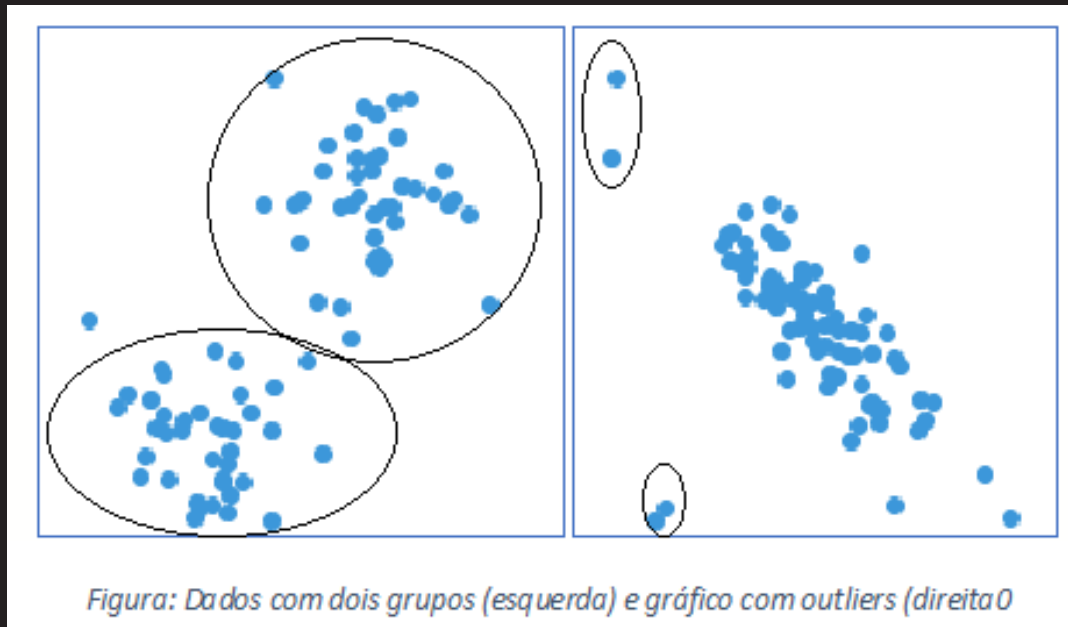


Figura: Relação linear positiva (esquerda) e negativa (direita)

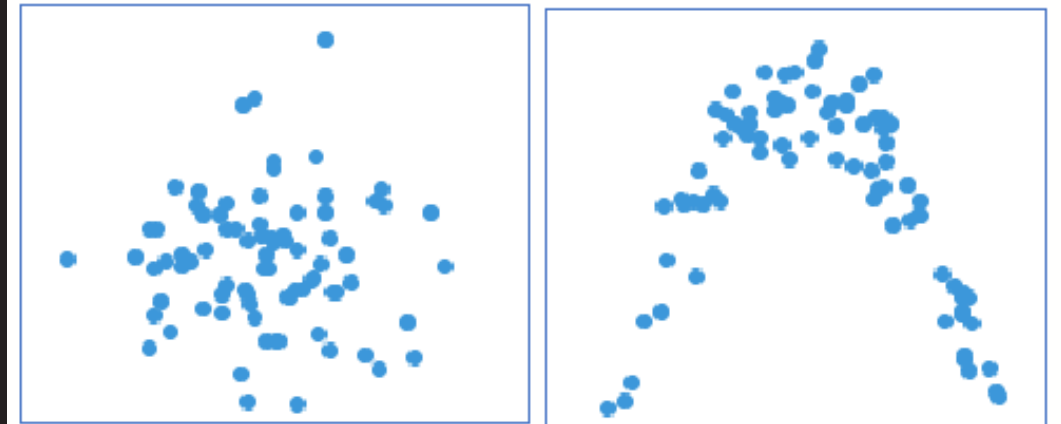


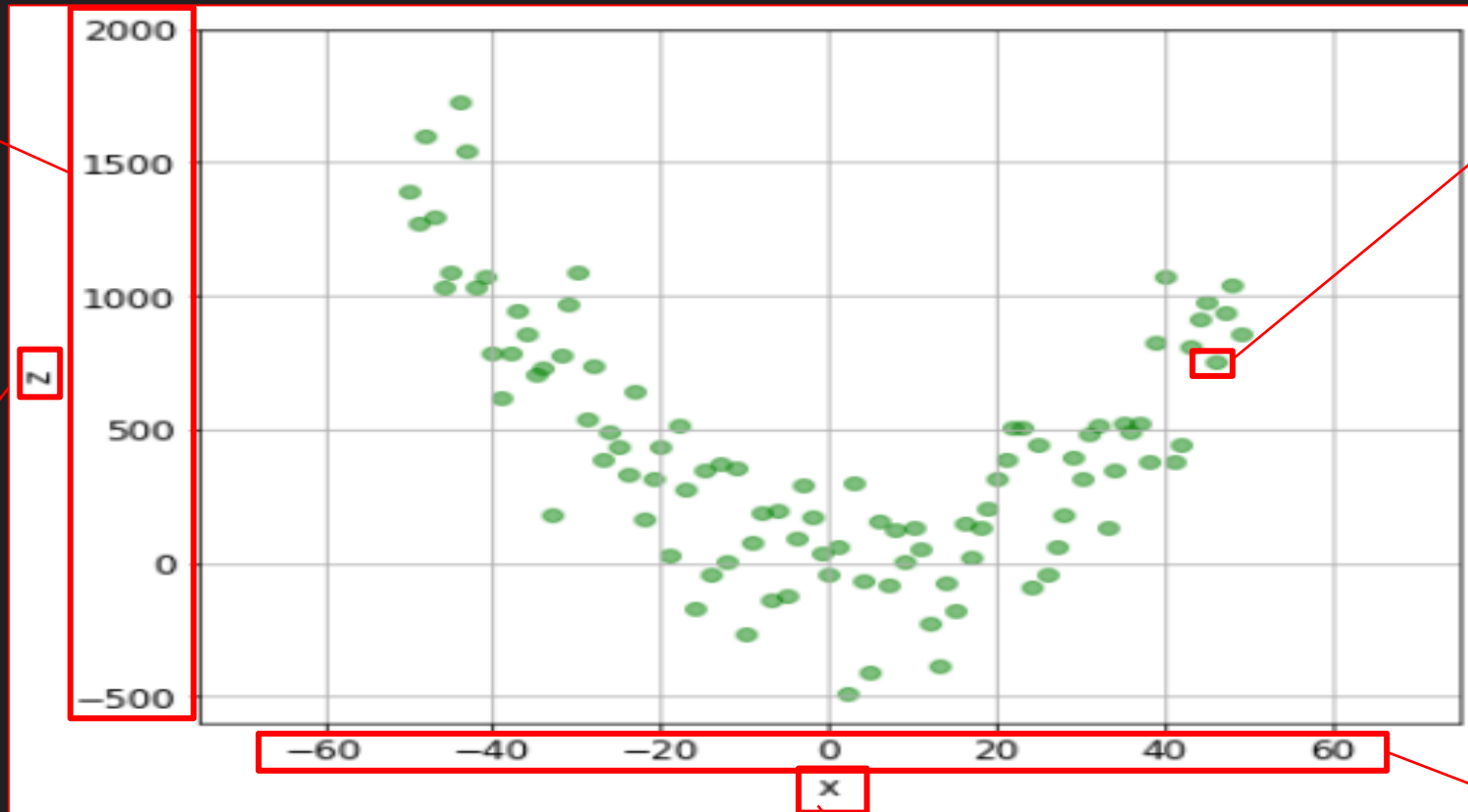
Figura: Relação neutra (esquerda) e não linear (direita)

Dispersão (Scatter plot)

Valores para a variável

Um item qualquer do conjunto de dados

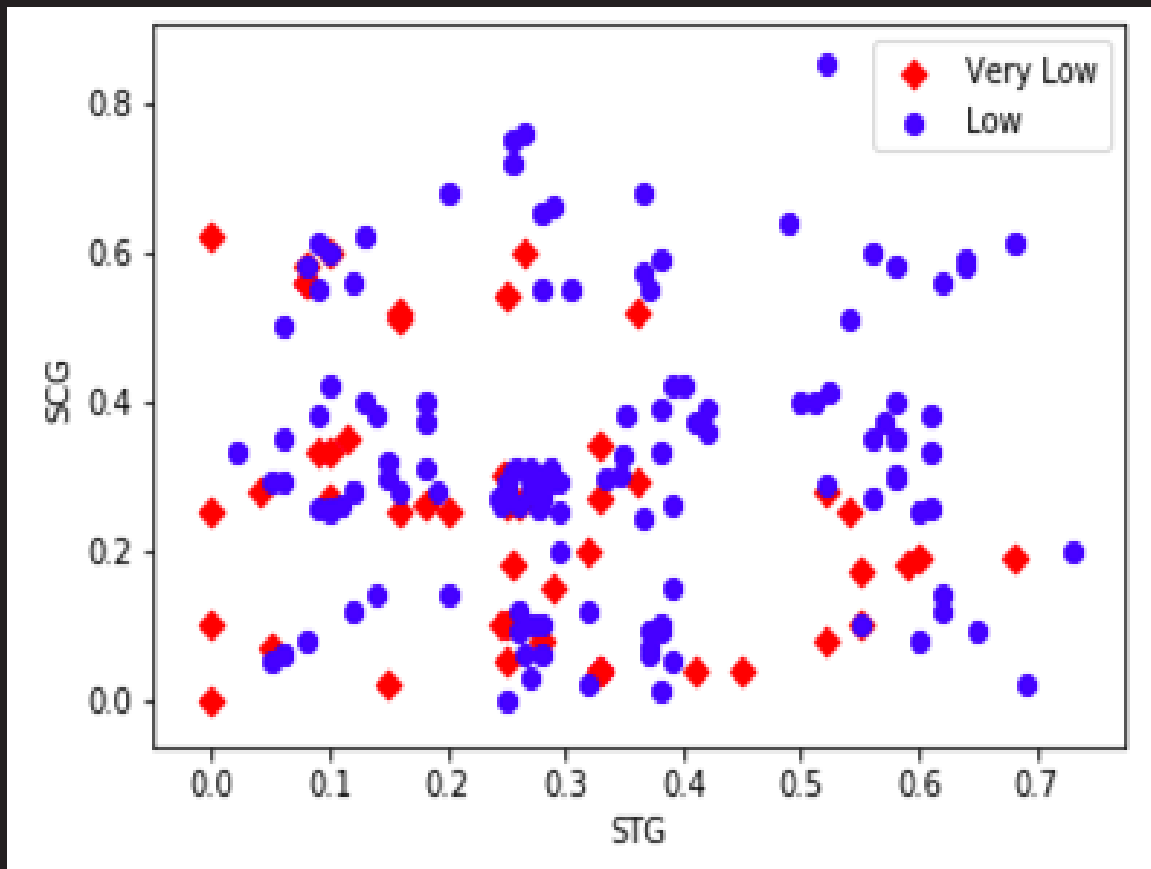
Nome da variável no eixo vertical



Valores para a variável

Nome da variável no eixo horizontal

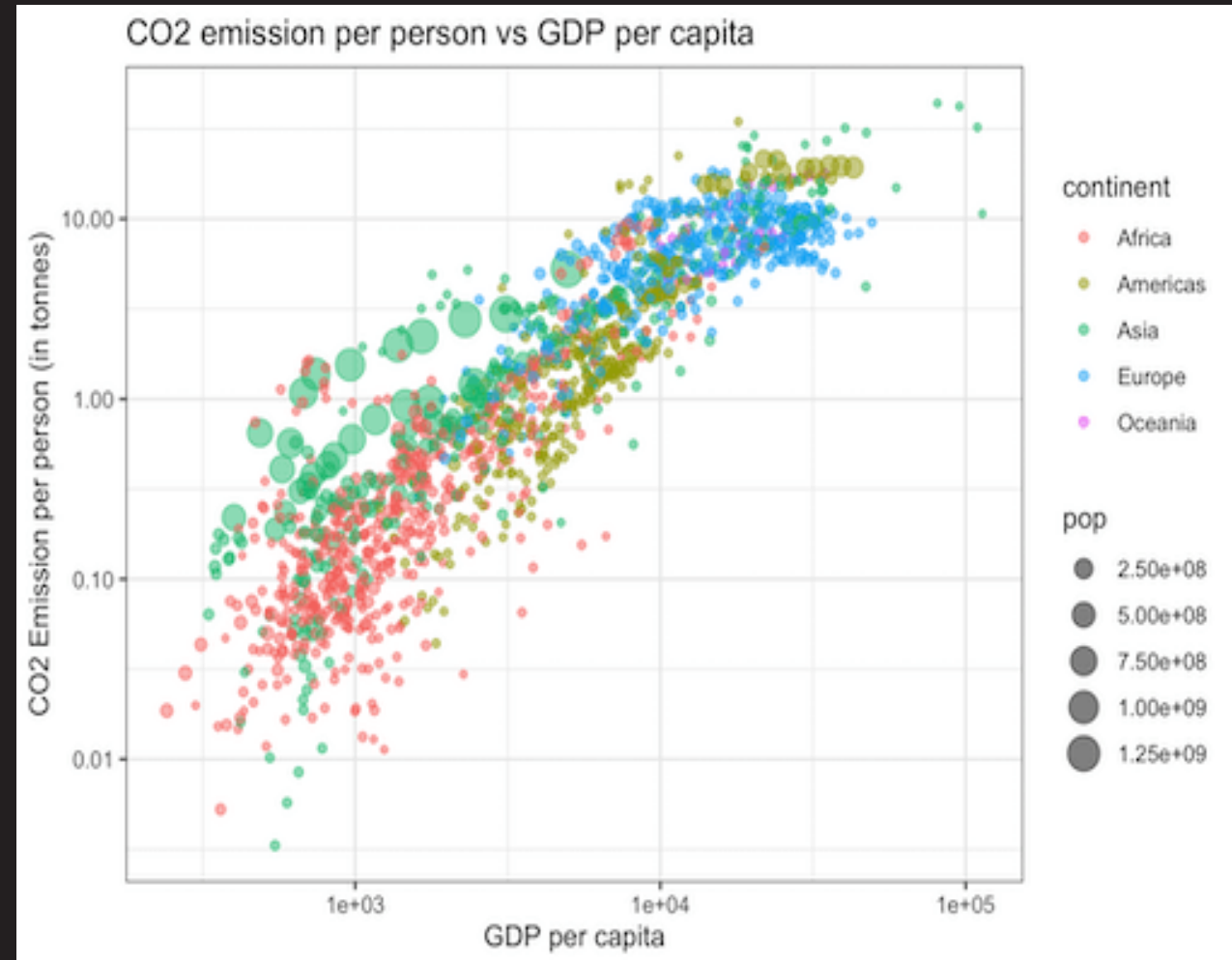
Dispersão (Scatter plot)



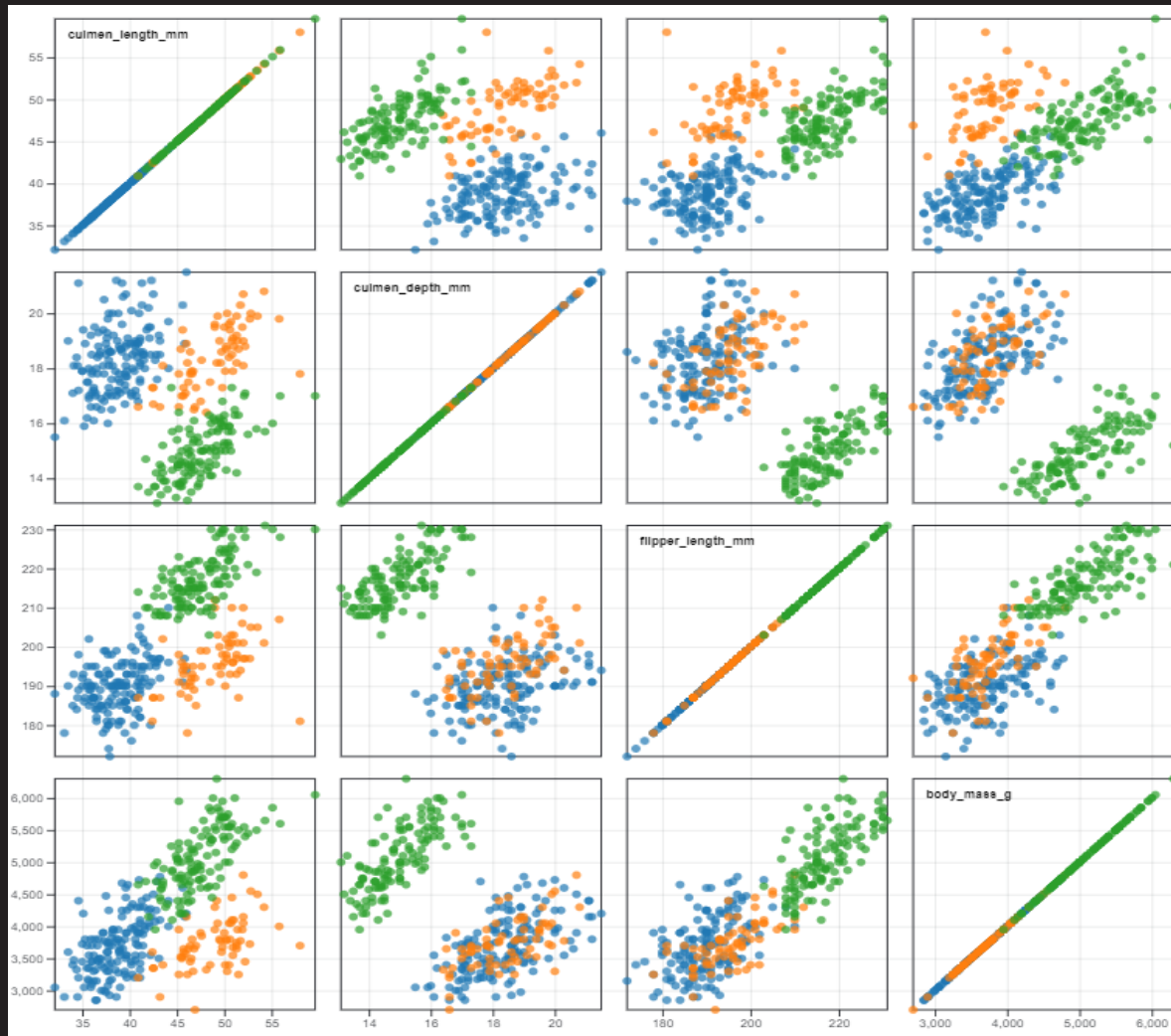
Temos 2 dimensões: x e y . Uma terceira dimensão pode ser colocada: a cor do ponto ou seu estilo pode representar uma classe (categoria) diferente à que ele pertence. Uma legenda costuma ser colocada nesse caso para indicar o esquema de cores.

Dispersão (Scatter plot)

Outra dimensão que pode existir: o tamanho dos pontos pode conter informação também. Os tamanhos são criados proporcionalmente a algum atributo numérico. Uma legenda pode ser utilizada para mostrar o significado de diferentes tamanhos.



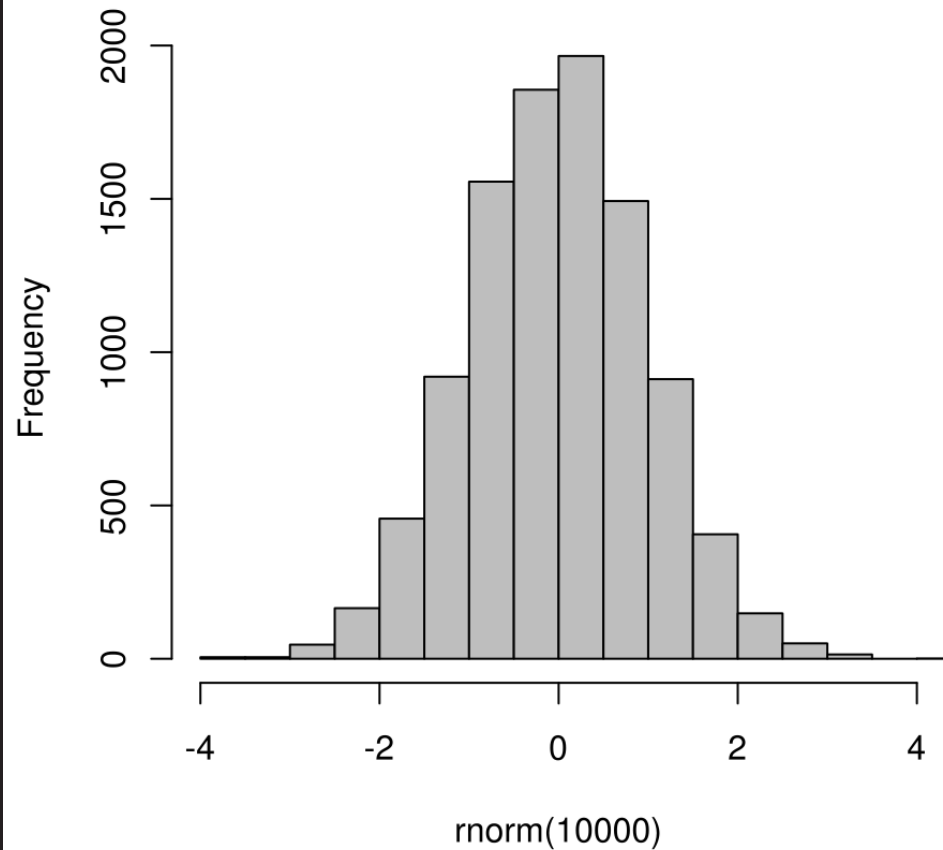
Dispersão (Scatter plot)



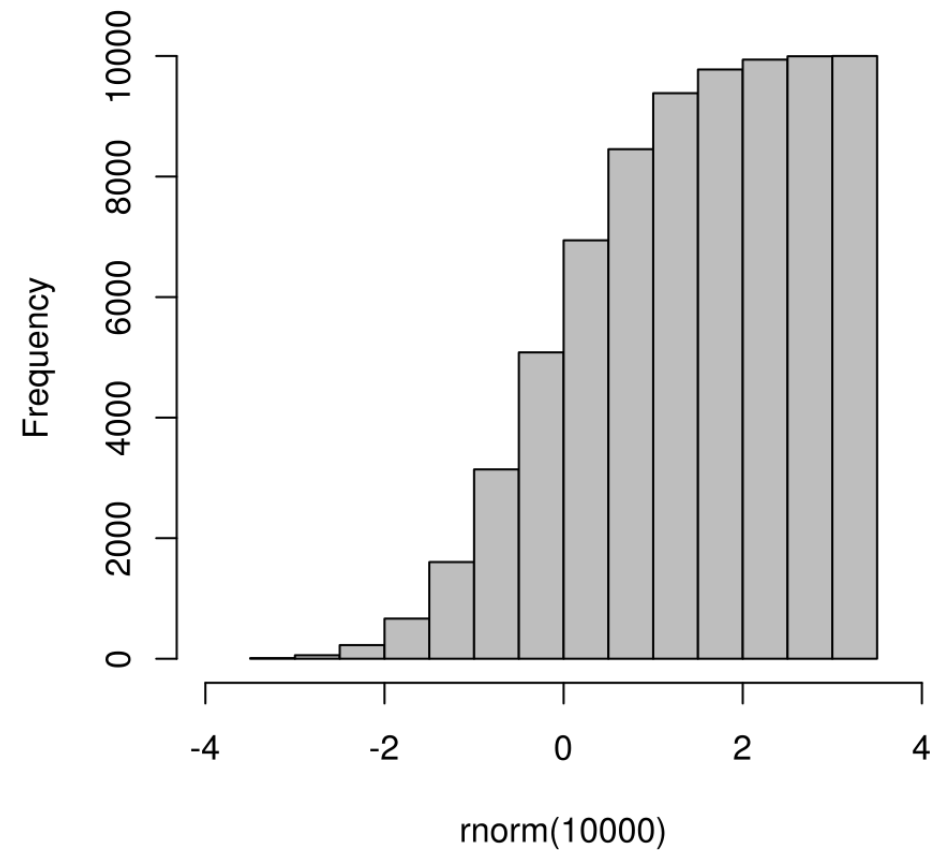
Uma matriz de *scatter plots* pode ser gerada para mostrar todos os possíveis pares de atributos, quando a base de dados possui vários atributos.

Distribuição

Ordinary histogram



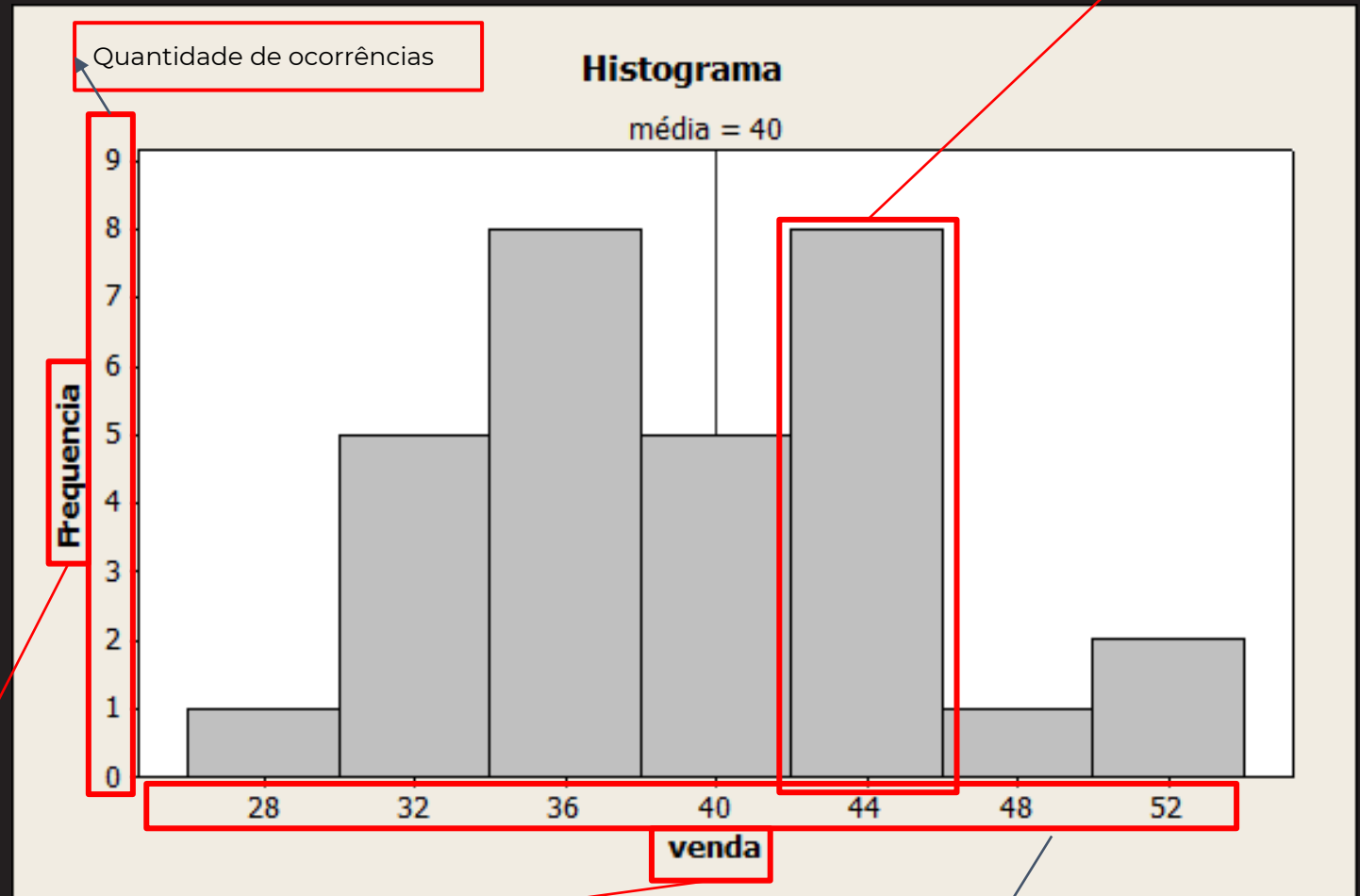
Cumulative histogram



Distribuição (Histogramas)

O eixo x tem valores de um atributo, e o eixo y tem a contagem de quantos dados possuem caem em cada valor.

- Se o eixo x for um atributo categórico, temos uma coluna para cada categoria (e podemos chamar de *gráfico de barras*).
- Se for um atributo numérico, dividimos o intervalo em n sub-intervalos (chamados *bins*), e contamos quantos dados caem em cada sub-intervalo.

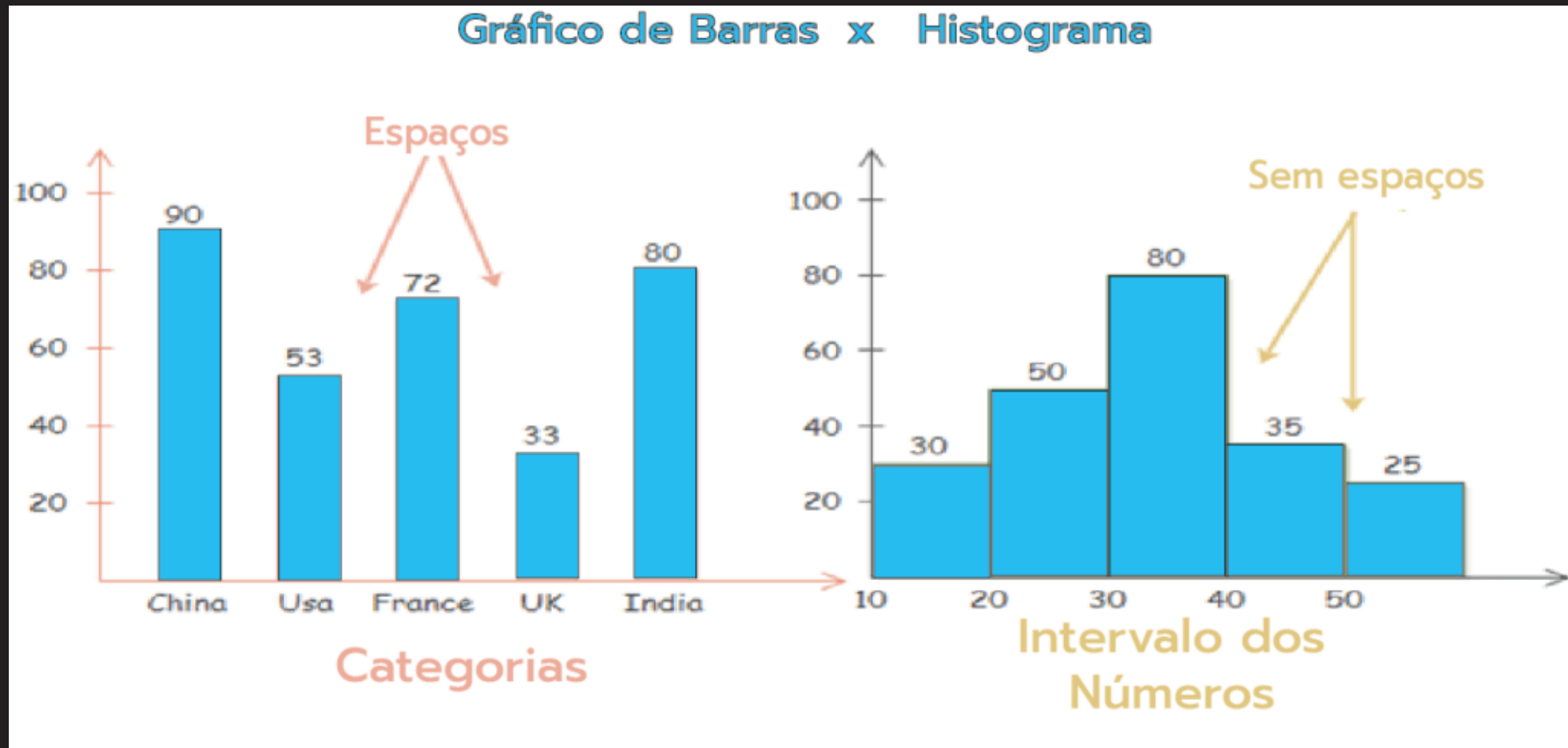


Frequência
(ou contagem)

Nome da variável no
eixo horizontal

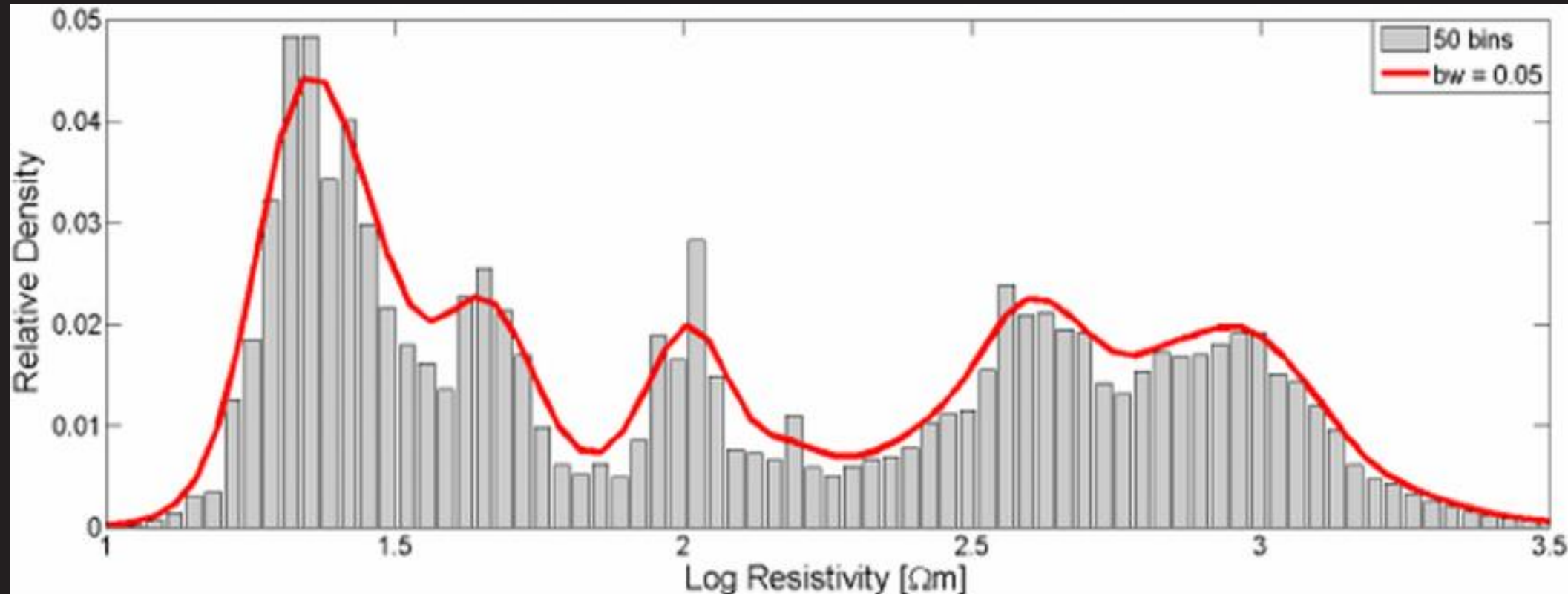
Valores para a
variável

Distribuição (Histogramas)

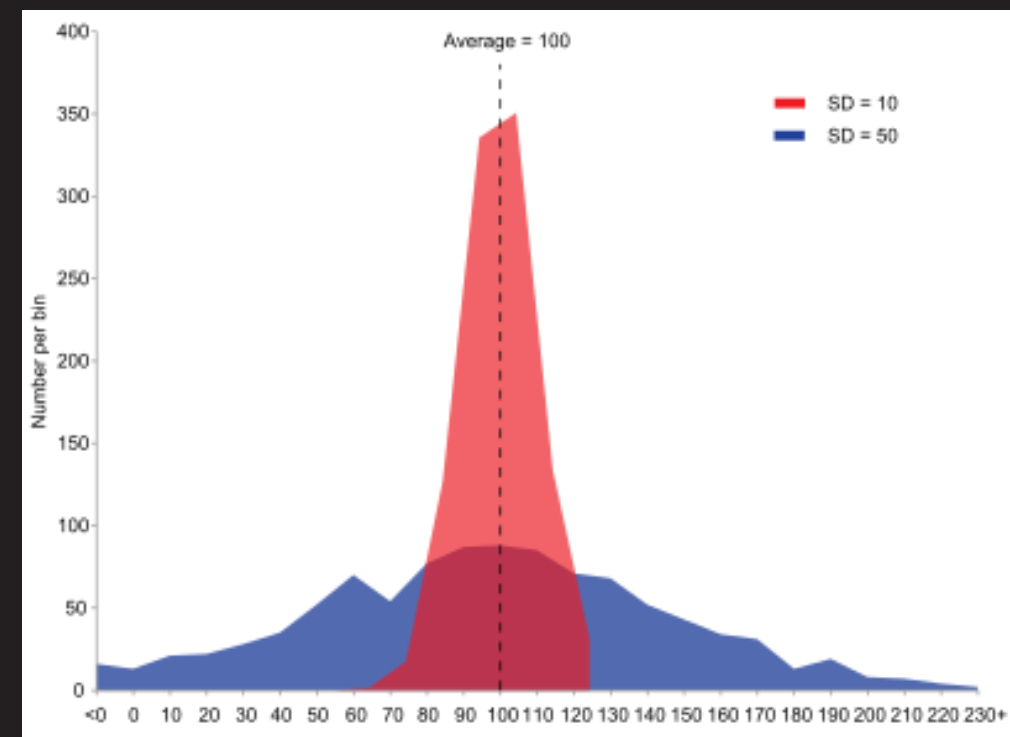
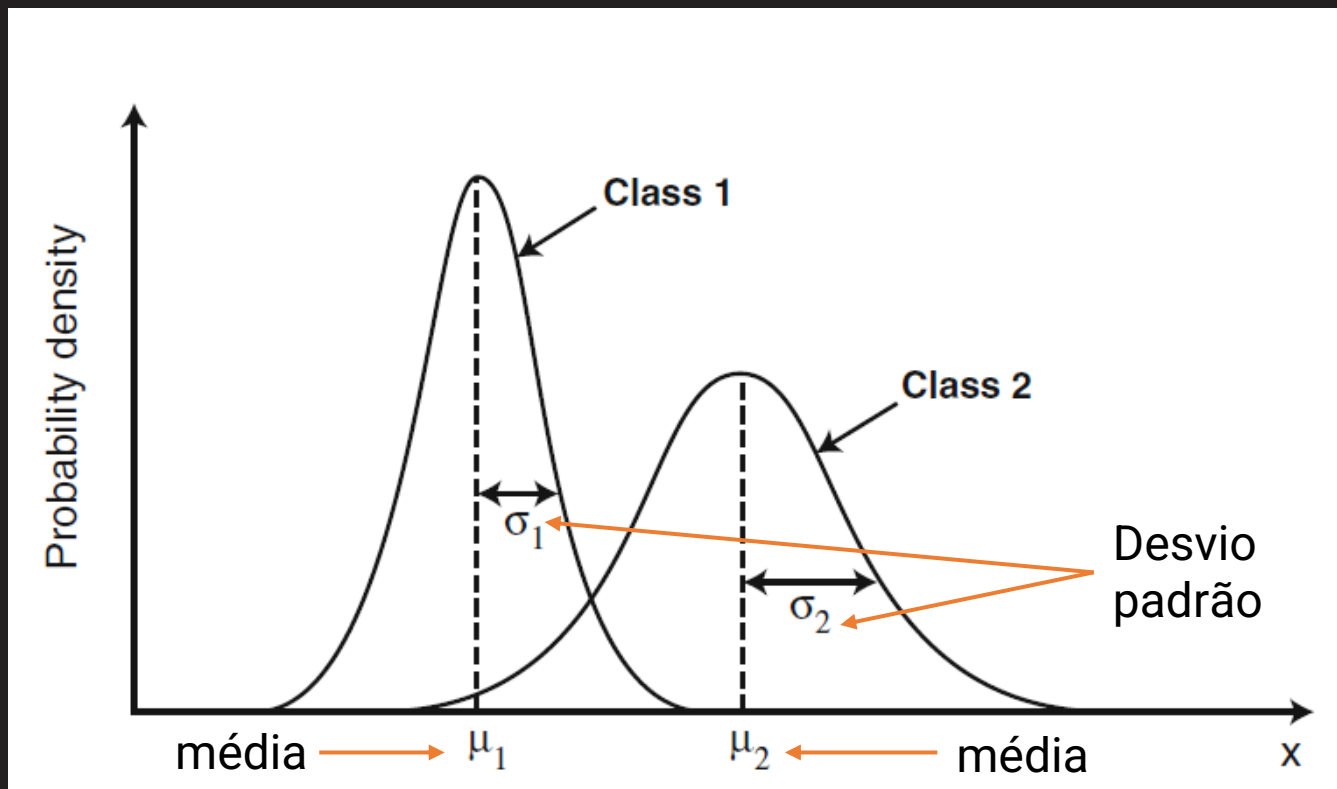


Distribuição (Histogramas)

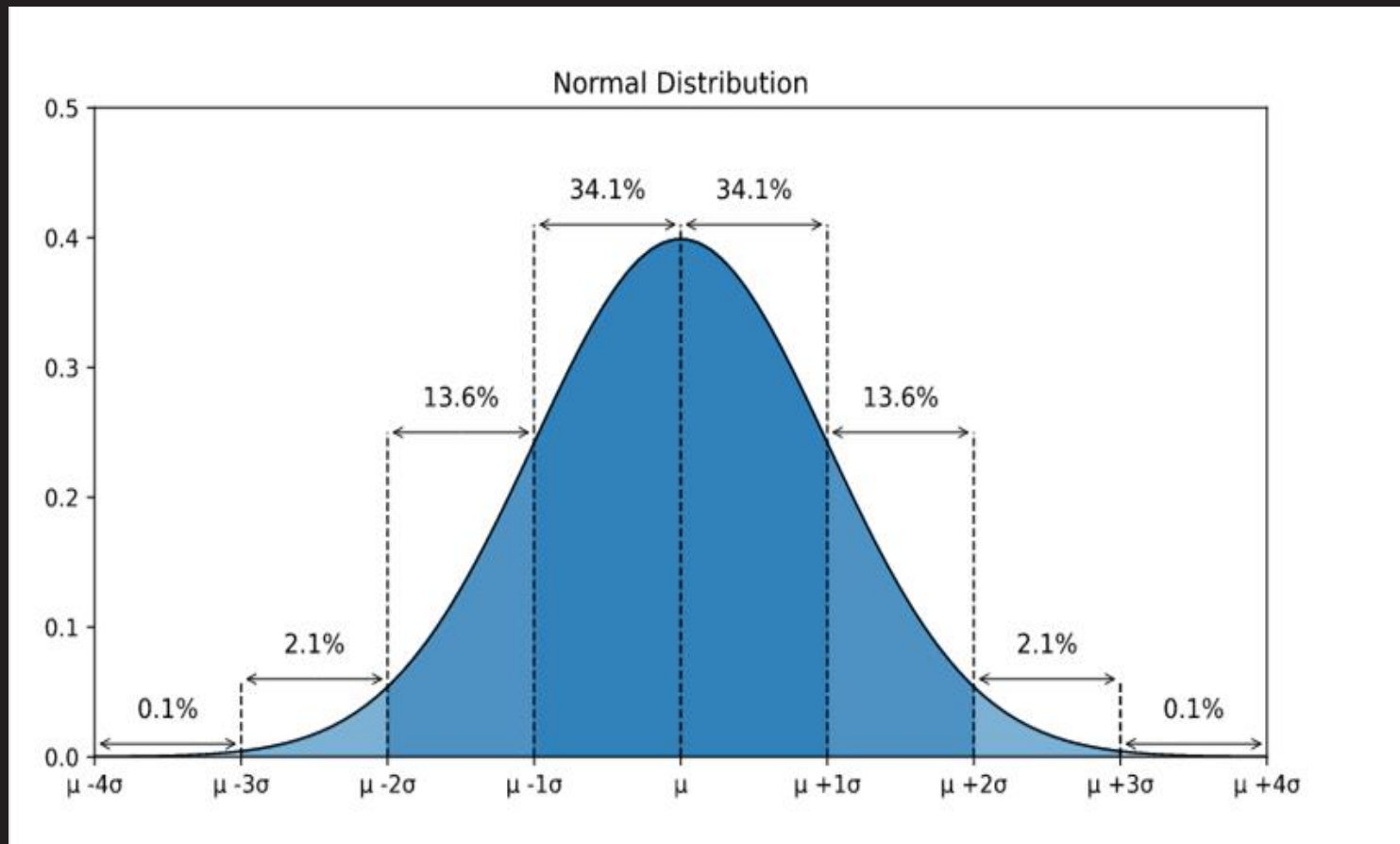
Podemos estimar uma curva para o histograma, geralmente chamada de *Kernel Density Estimation* (KDE), que representa a distribuição.



Distribuição



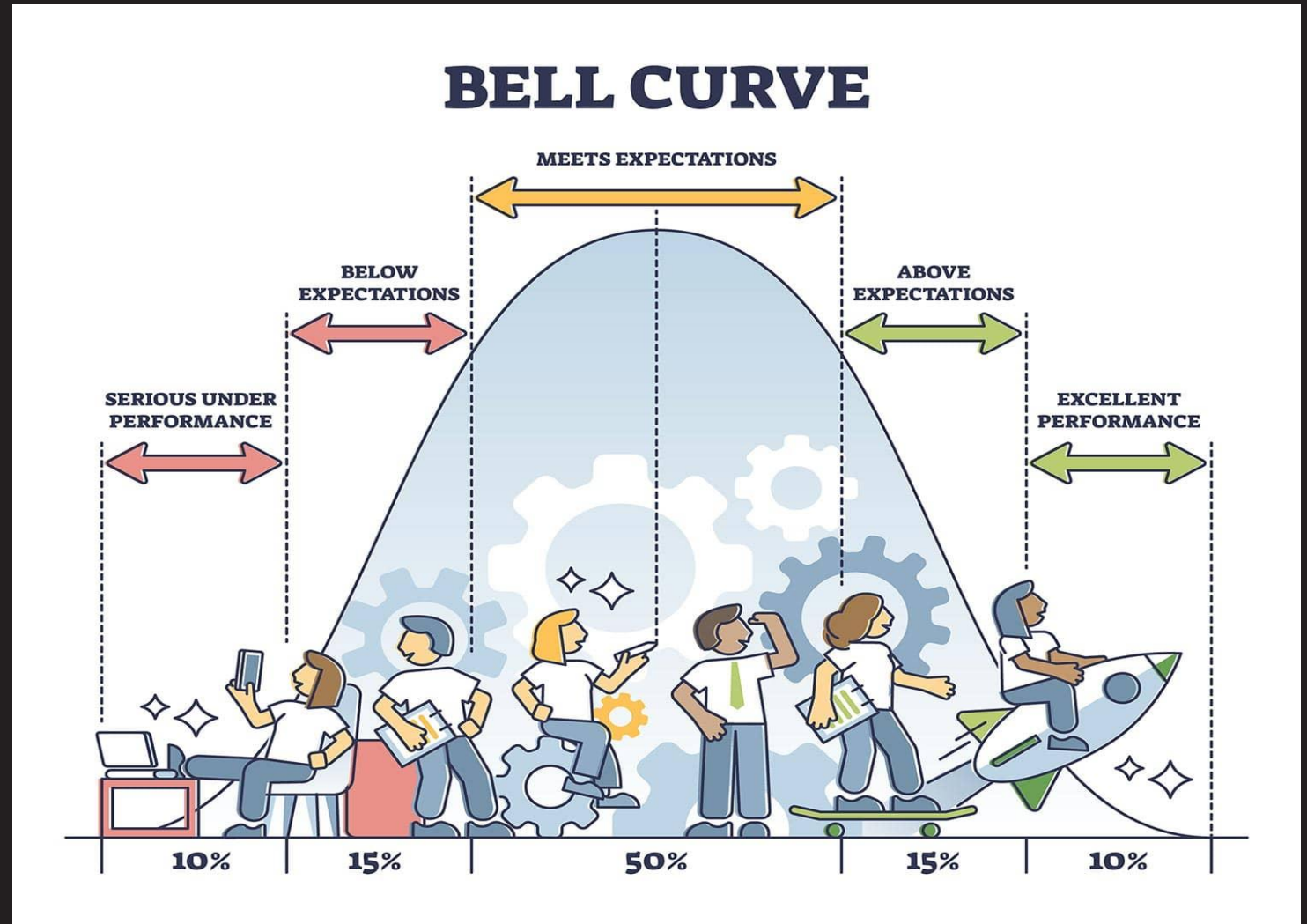
Distribuição



Existe um comportamento em histogramas e distribuições chamado de Distribuição normal ou gaussiana. Ela é uma distribuição com formato de “sino”.

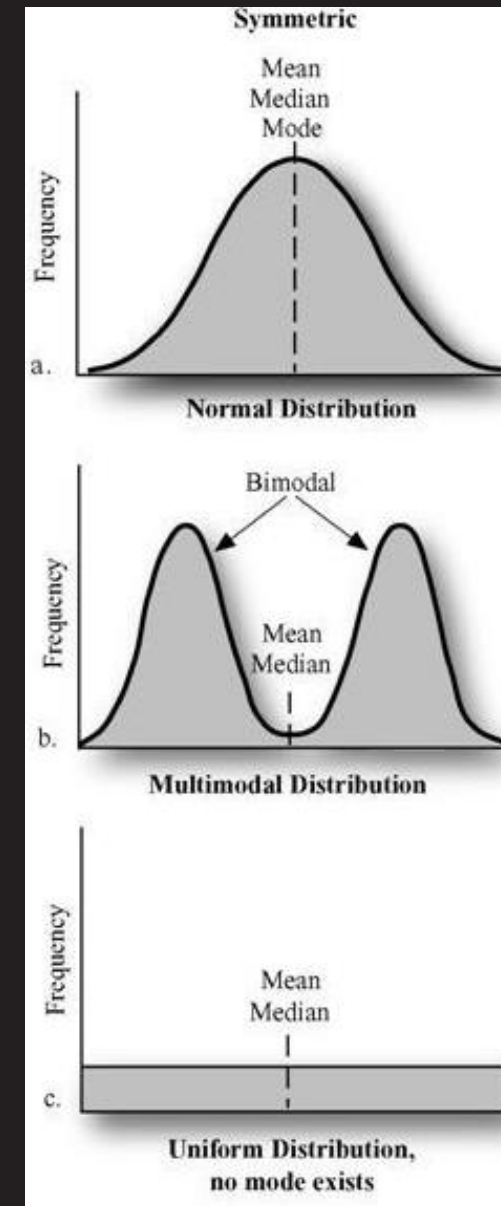
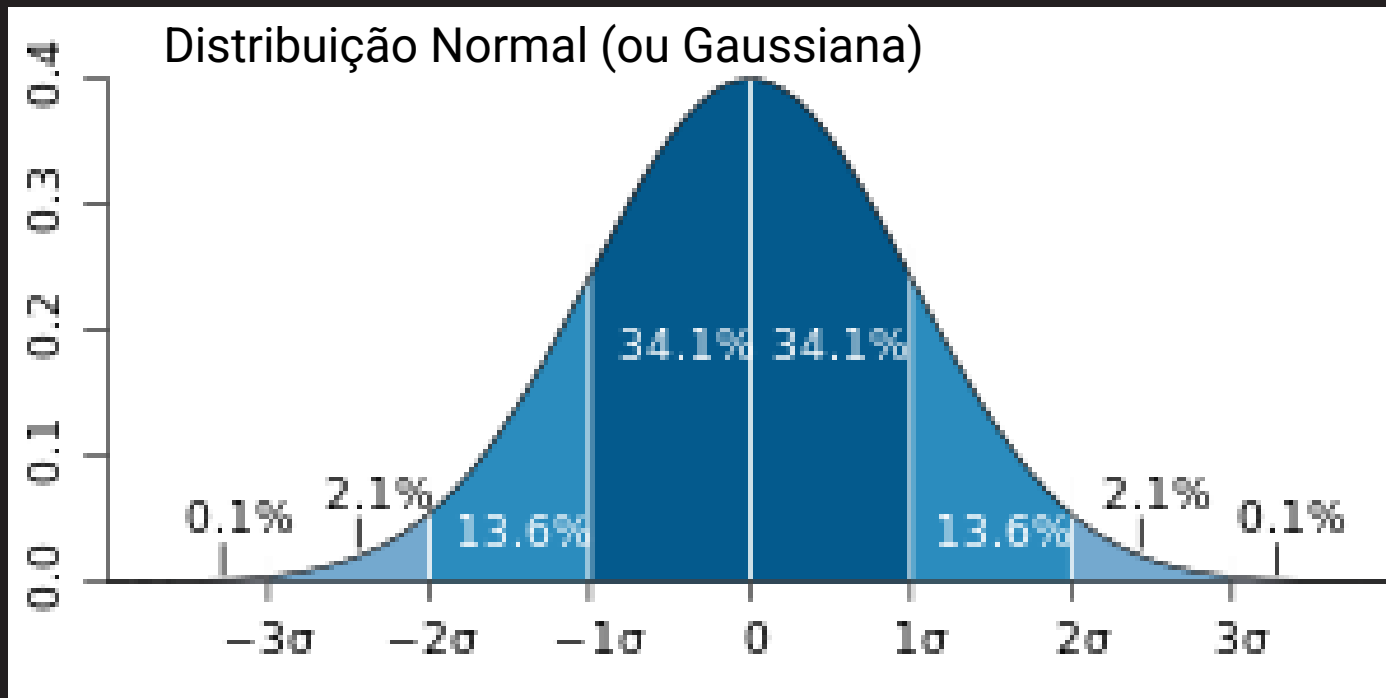
Distribuição

A distribuição Gaussiana é **centrada na média**, e ocorre em diversos dados observáveis do mundo real. A maioria dos dados está perto da média e, quanto mais distante, menor a frequência de encontrarmos um valor.



Distribuição

Na vida real, nem todos os processos são “normais” (gaussianos):



Distribuição

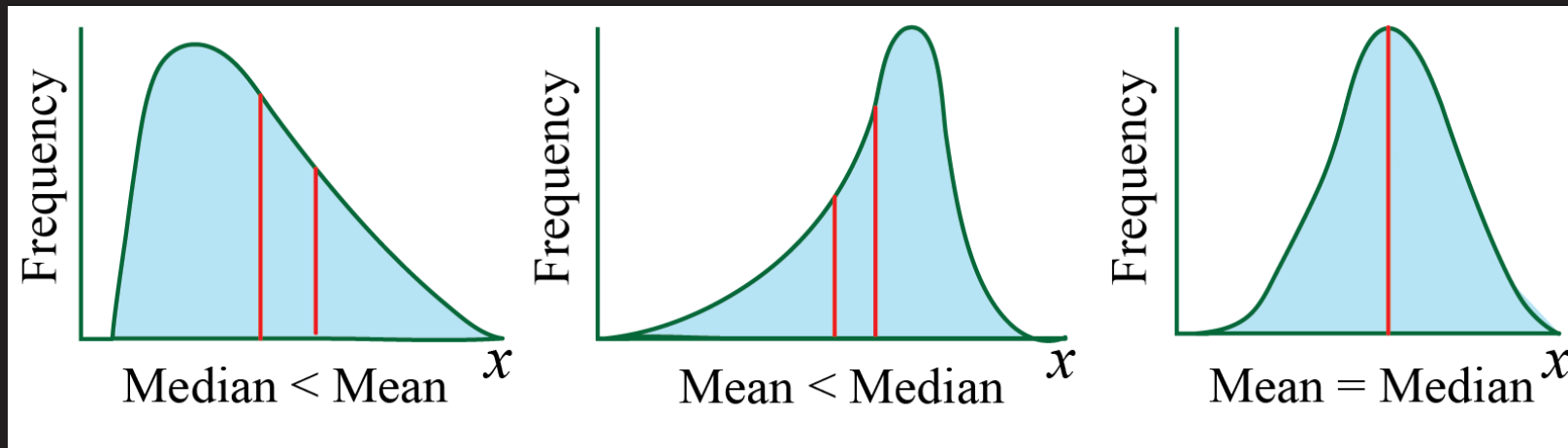
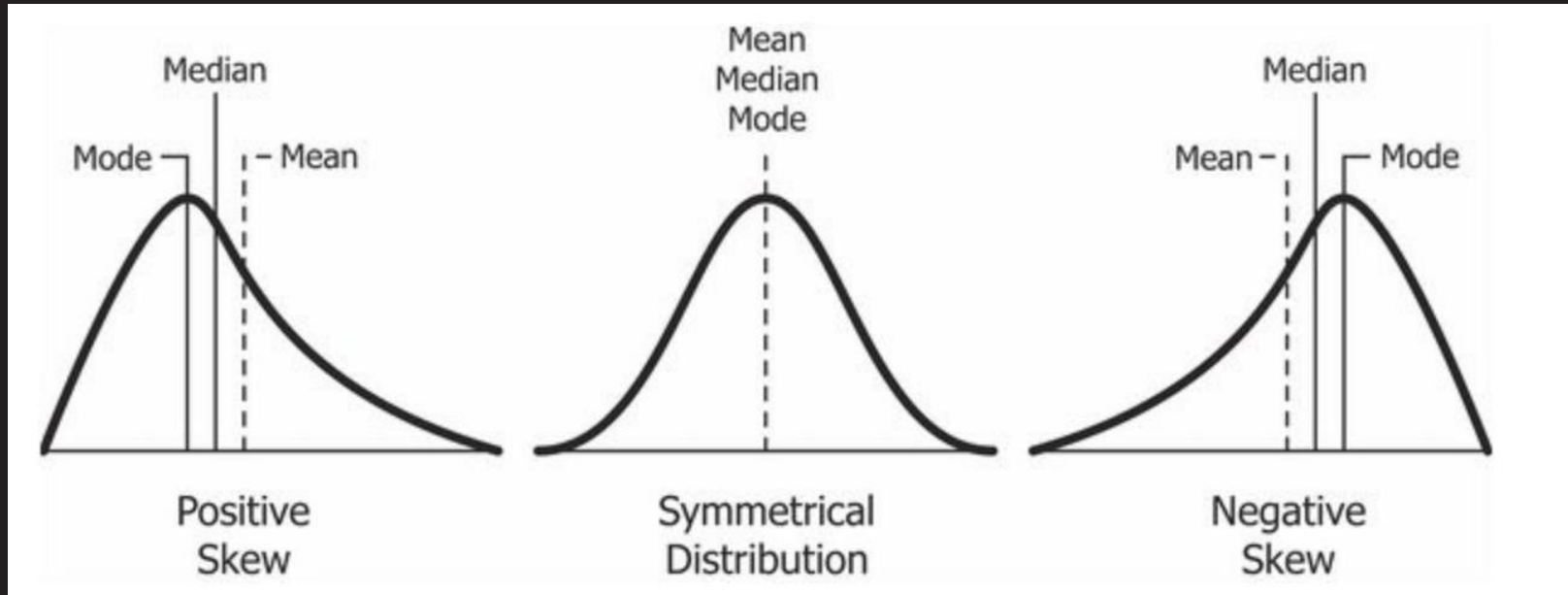


Diagrama de caixa (Boxplot)

E se quiséssemos mostrar um histograma (ou distribuição, num senso mais genérico) de forma compacta? Boxplots fornecem uma “vista superior” da distribuição:

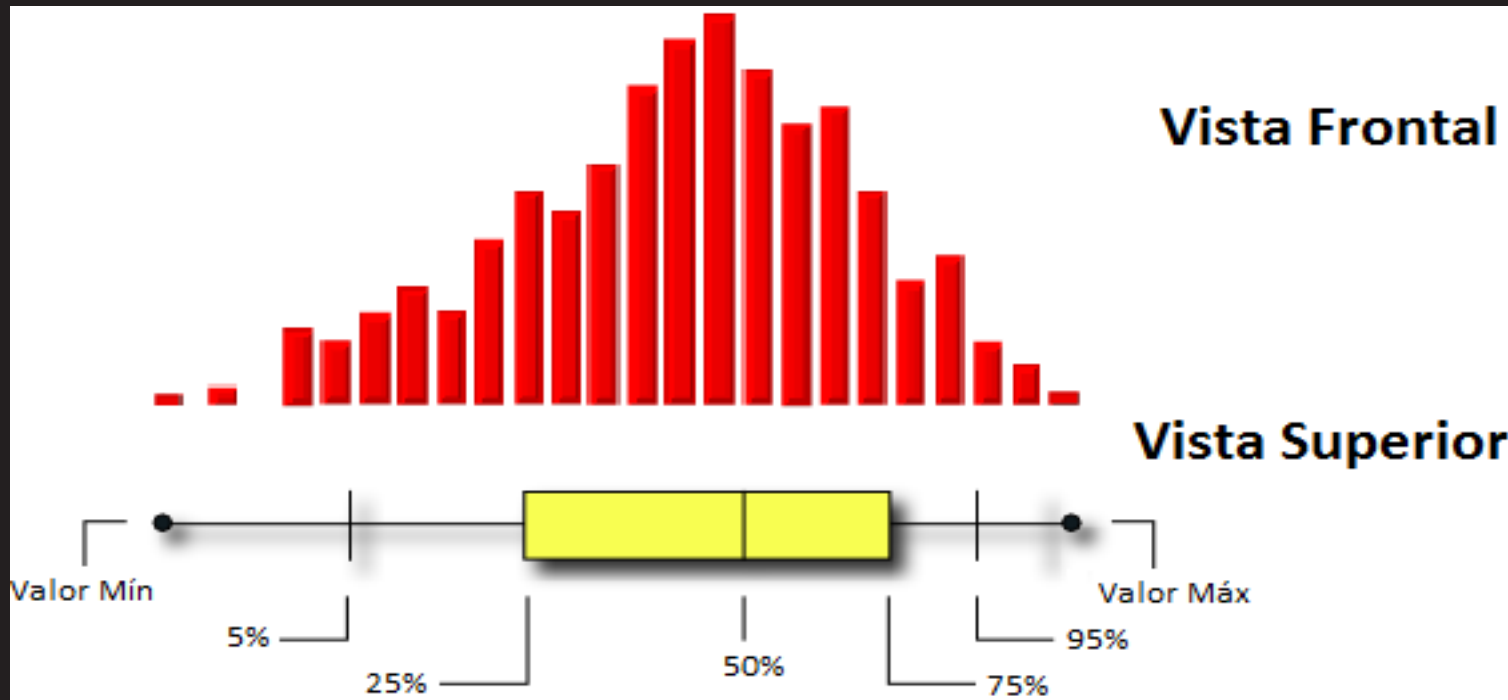
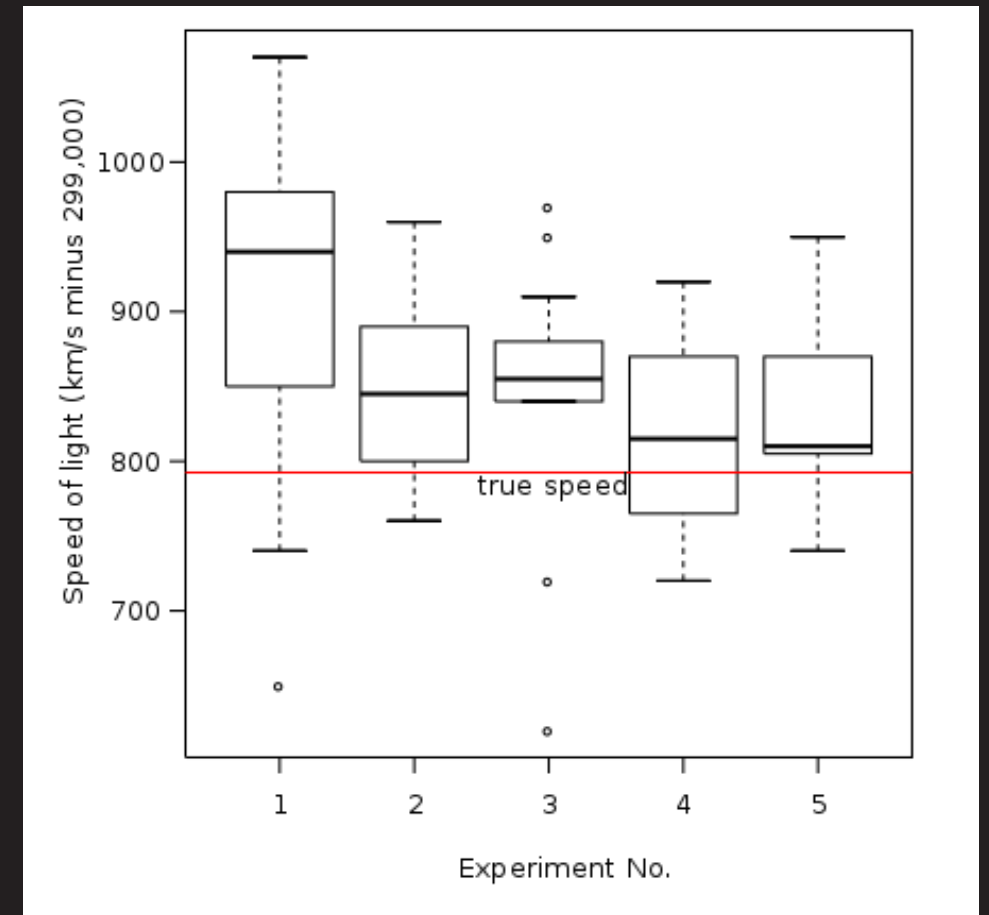
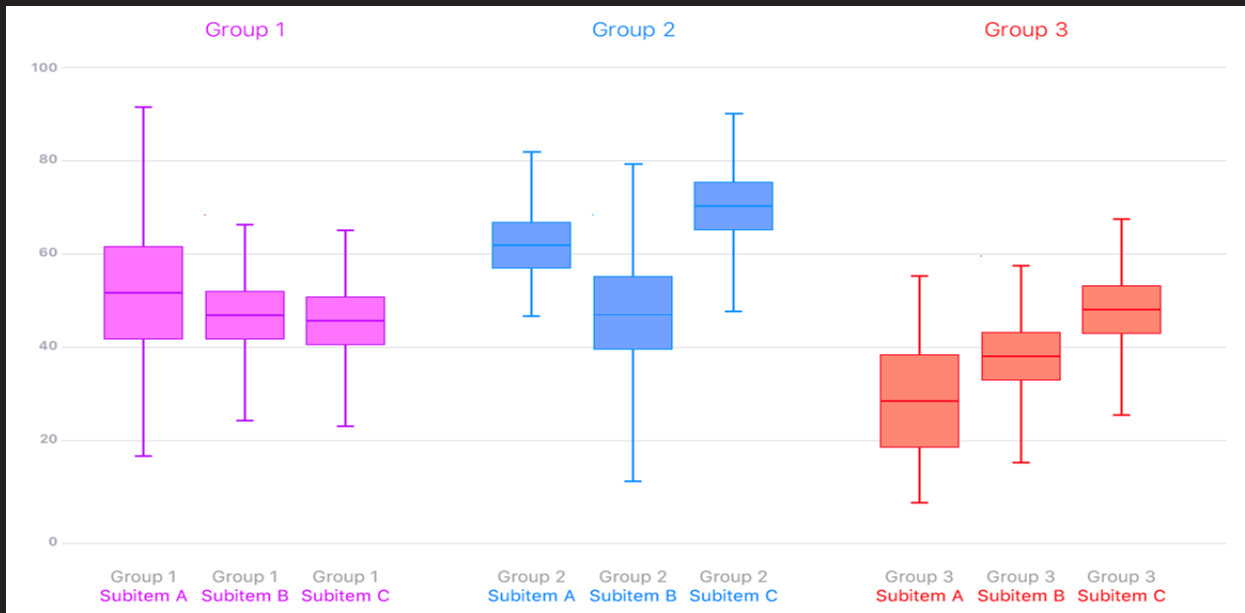
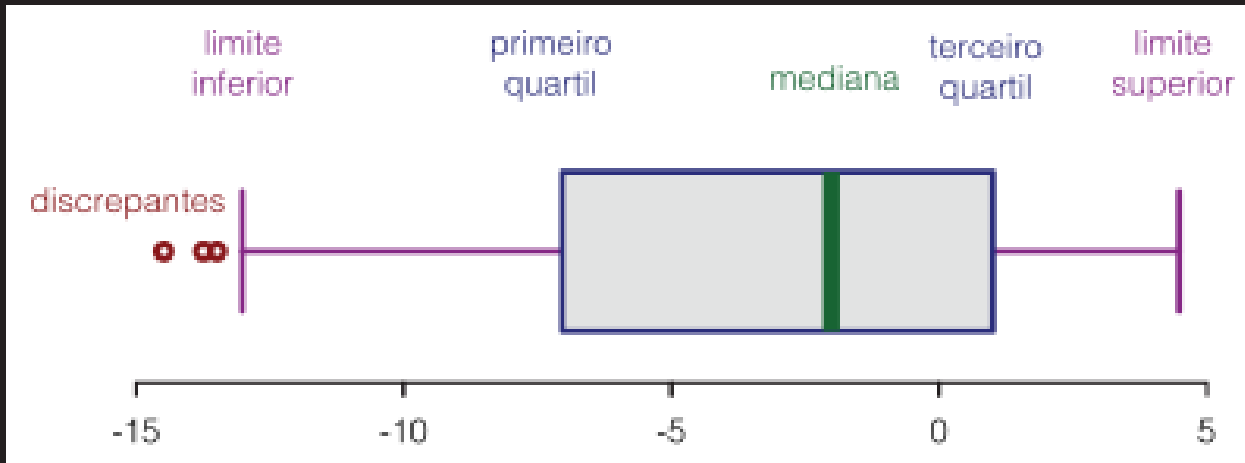
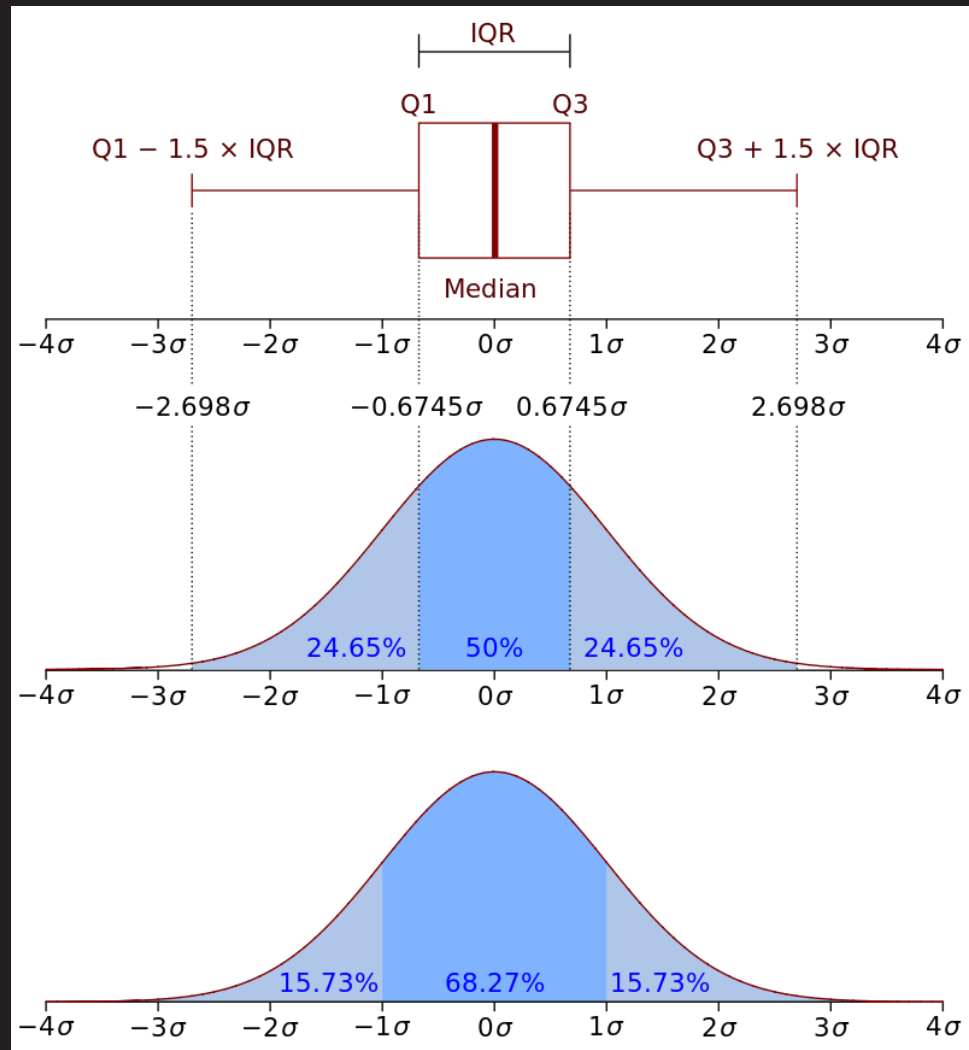


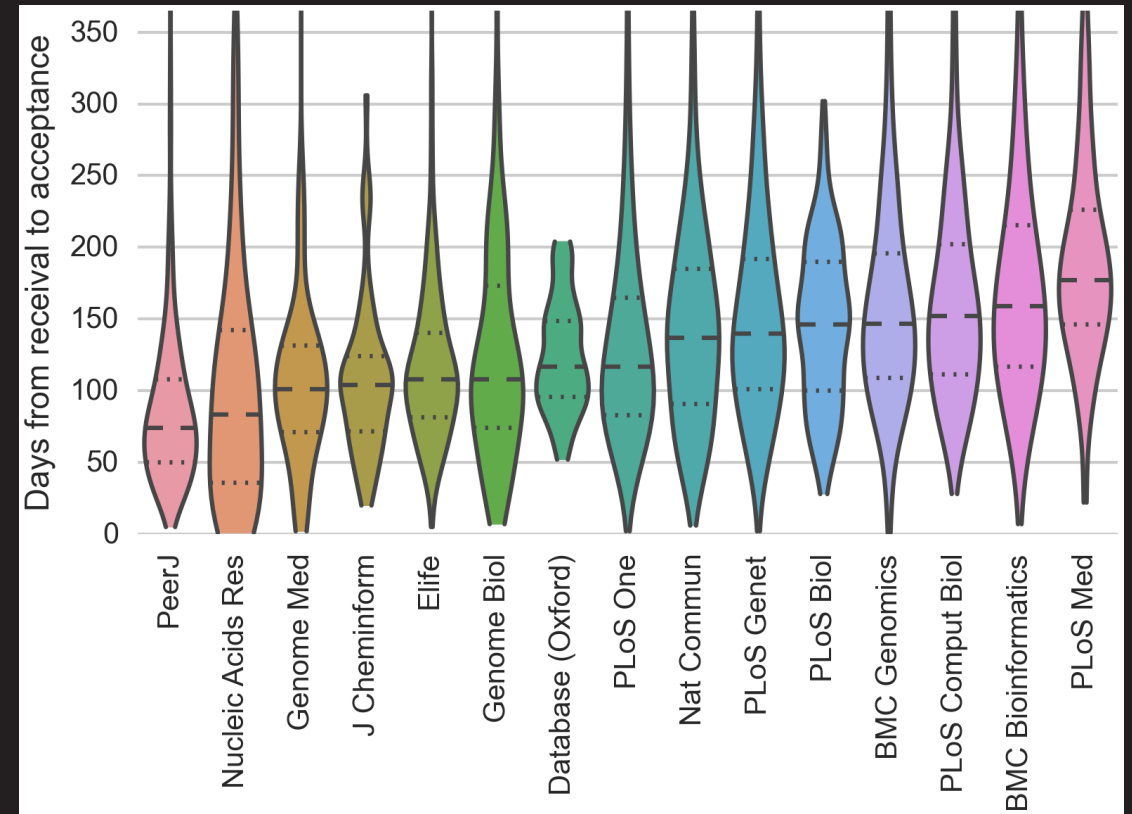
Diagrama de caixa (Boxplot)



Distribuição + Boxplot



Violin plot



Violinplot

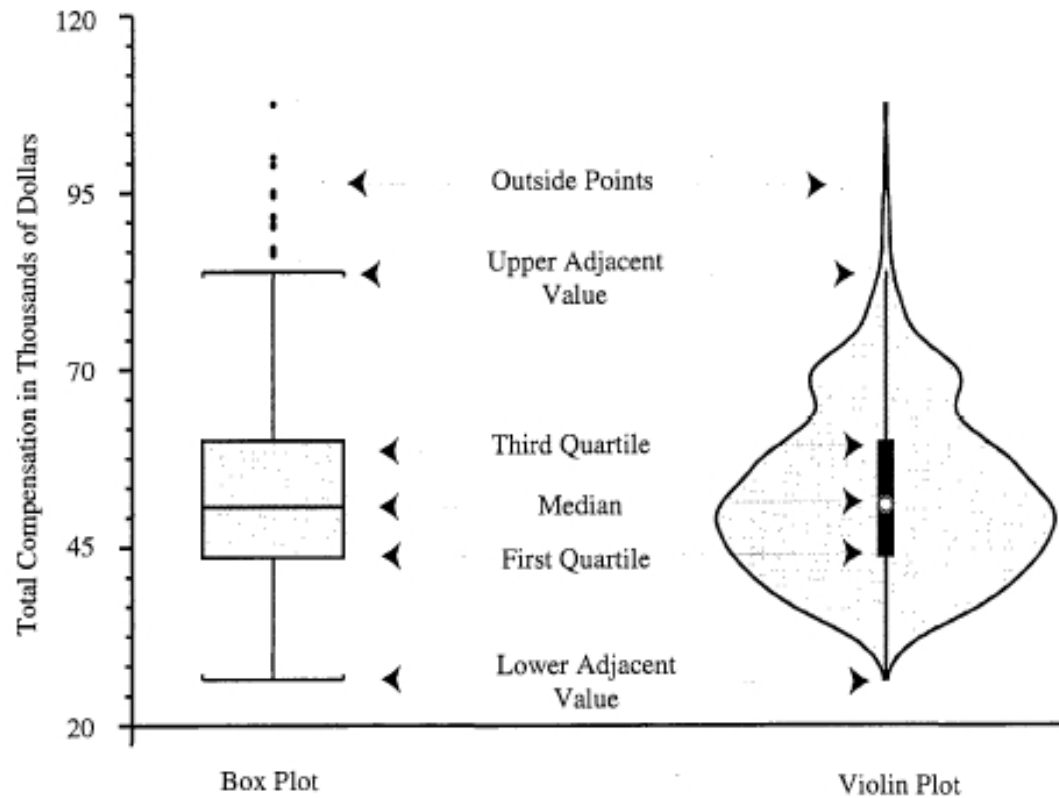
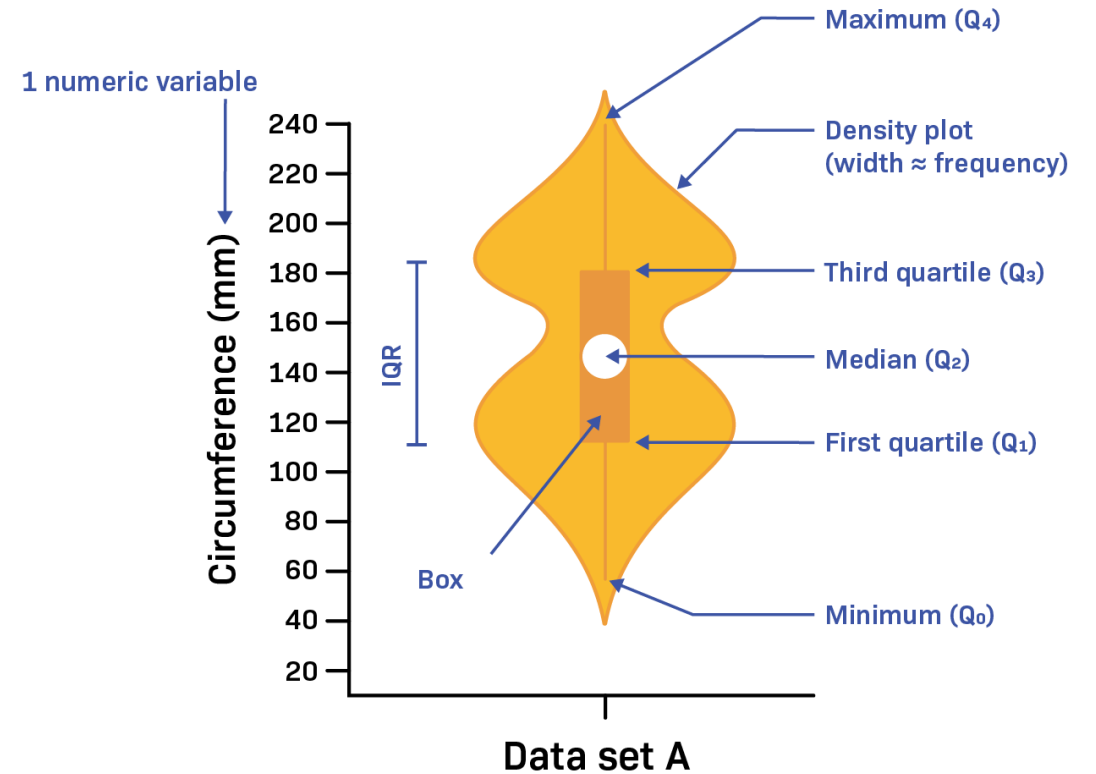
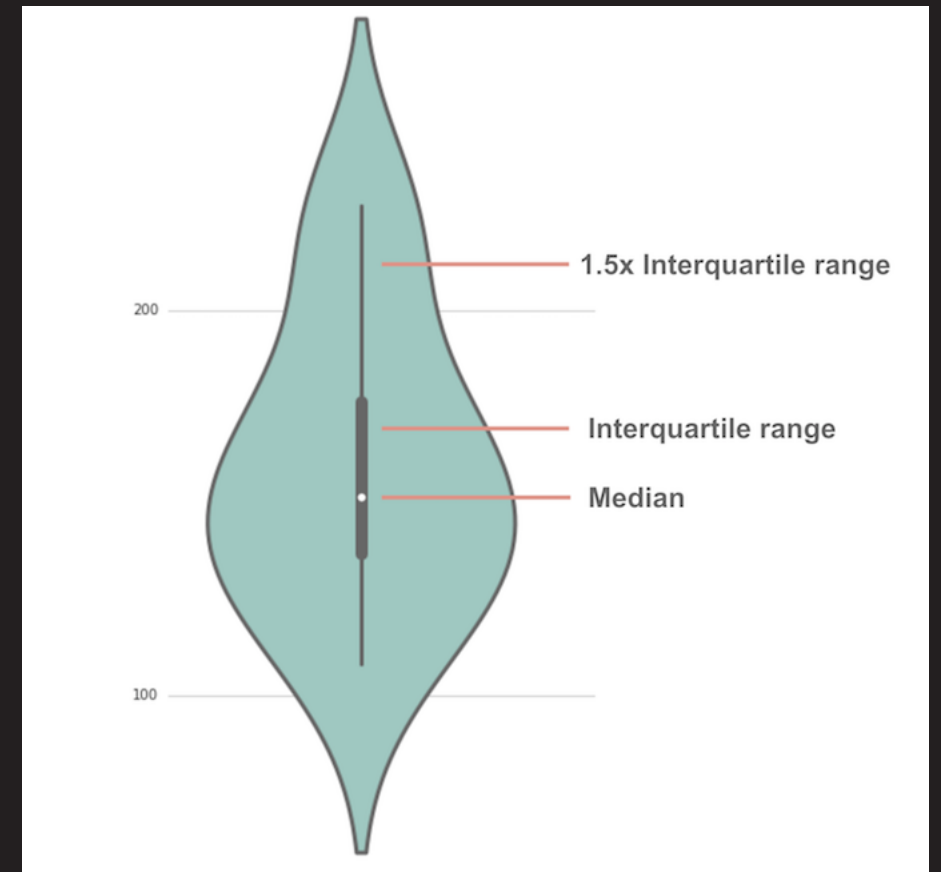
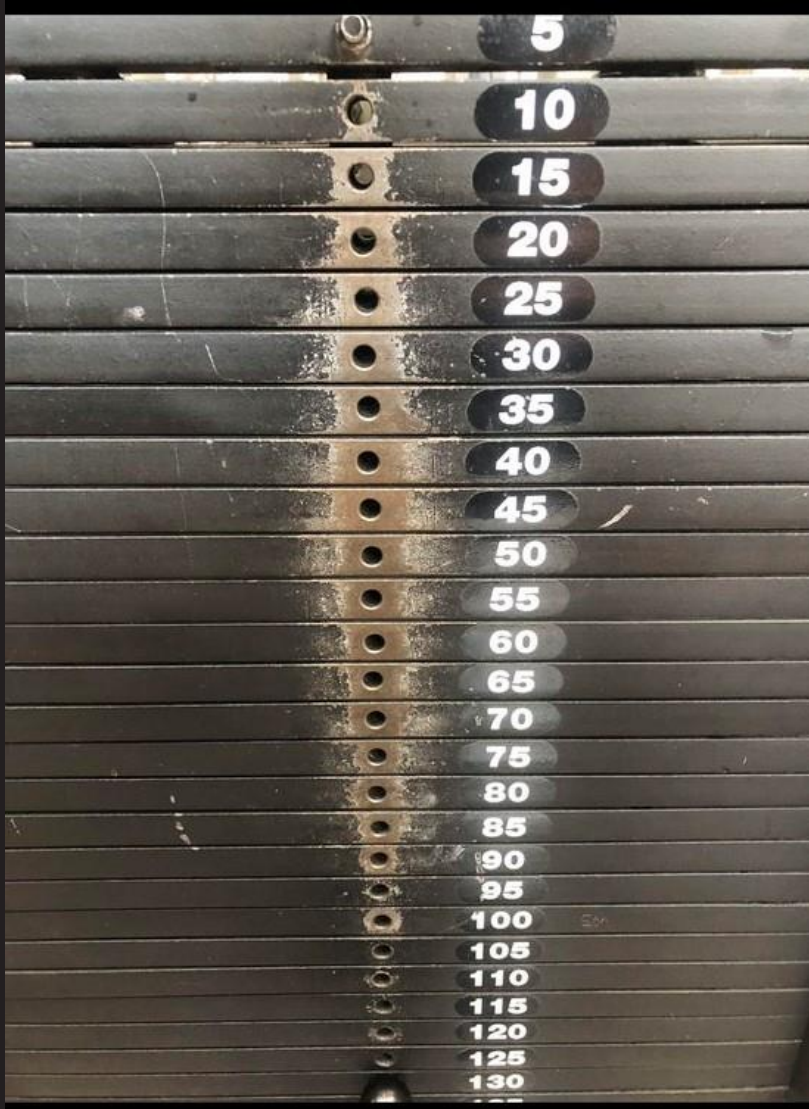


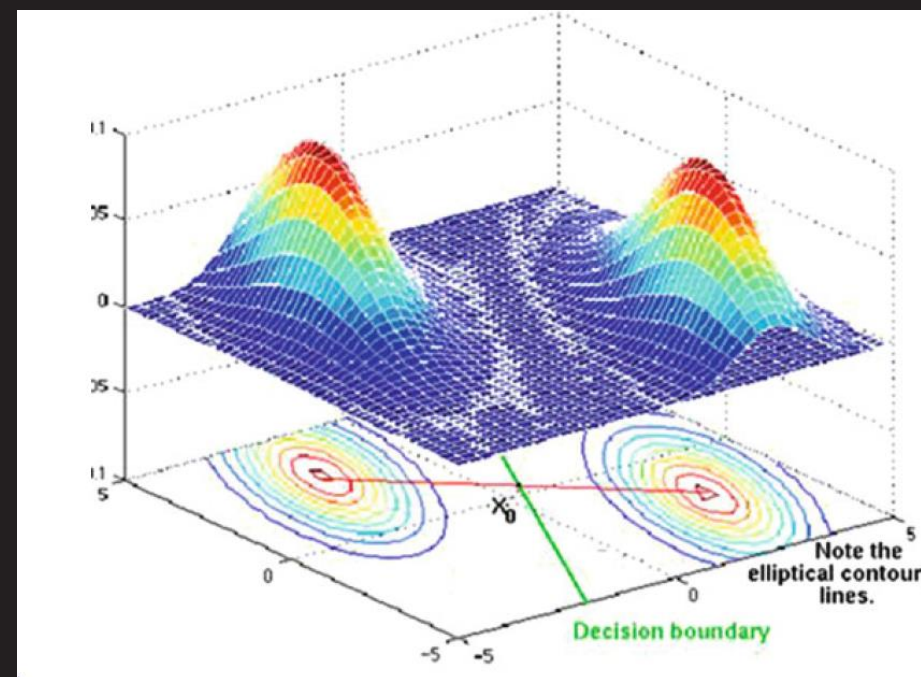
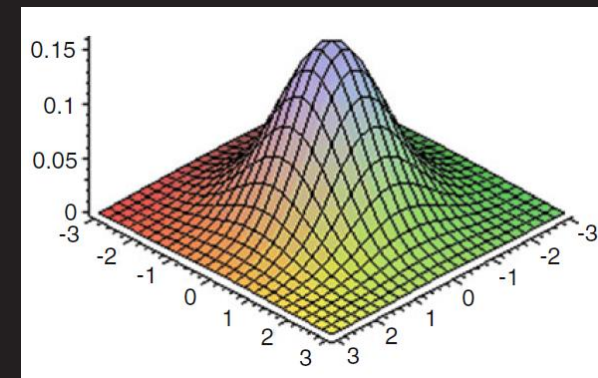
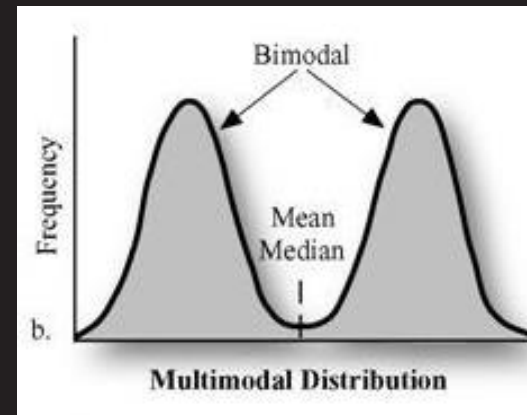
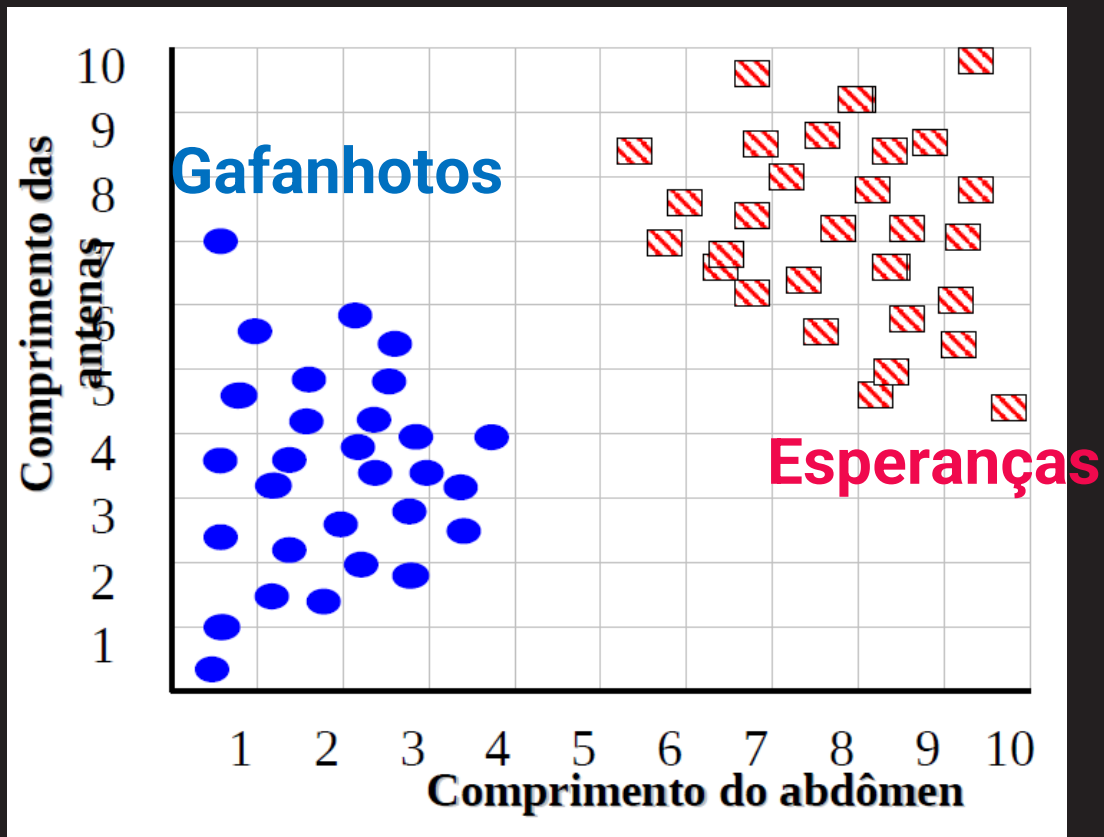
Figure 1. Common Components of Box Plot and Violin Plot. Total compensation for all academic ranks.



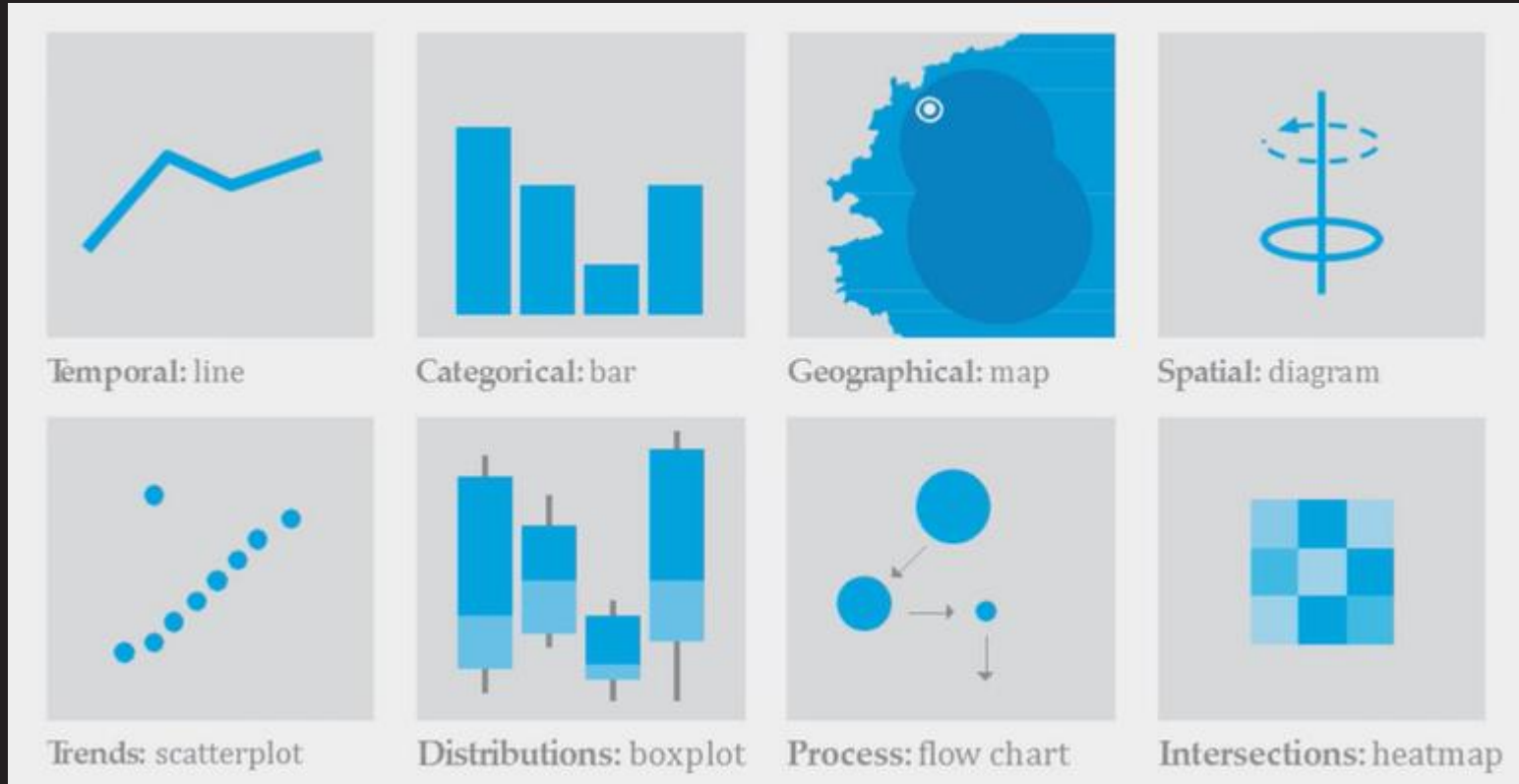
Exemplo: violinplot na academia



Distribuição + Dispersão



Melhores práticas



- 1) Estabeleça uma hierarquia da informação que quer apresentar;
- 2) Escolha o tipo de visualização de acordo com o conteúdo;
- 3) Faça um esboço na mão;
- 4) Produza a imagem (com o código adequado no Matplotlib/seaborn ou outra ferramenta de design).

<https://www.youtube.com/watch?v=Lb4uG4rlwPA>

Gráficos com Python

Bibliotecas

- Existem bibliotecas prontas para se gerar gráficos em Python;
- Boa parte do trabalho está em ler a documentação dessas bibliotecas para se entender como gerar o gráfico desejado;
- Duas muito conhecidas são:



<https://matplotlib.org/stable/index.html>



seaborn

<https://seaborn.pydata.org/index.html>

Próximos Passos

O que veremos na próxima aula

Nas próxima aulas...

- Ciência de Dados
- Aprendizado de Máquina Supervisionado de Classificação



Copyright © 2023

Slides do Prof. Henrique Ferreira - FIAP

Todos direitos reservados. Reprodução ou divulgação total ou parcial deste documento é expressamente proibido sem o consentimento formal, por escrito, do Professor (autor).