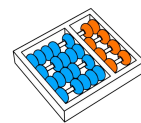


VISUALIZAÇÃO DE DADOS

Prof. Dr. Julio Cesar dos Reis

Instituto de Computação - Universidade Estadual Campinas



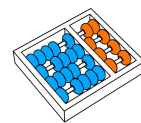
VISUALIZAÇÃO DE DADOS

Visualização de dados é uma das formas mais poderosas e intuitivas de exploração de dados

- Temos boa habilidade para analisar grandes quantidades de dados apresentados visualmente
- Podemos detectar padrões e tendências
- Ajuda a detectar outliers e padrões incomuns

Passo essencial em análise de dados

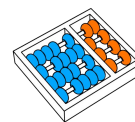




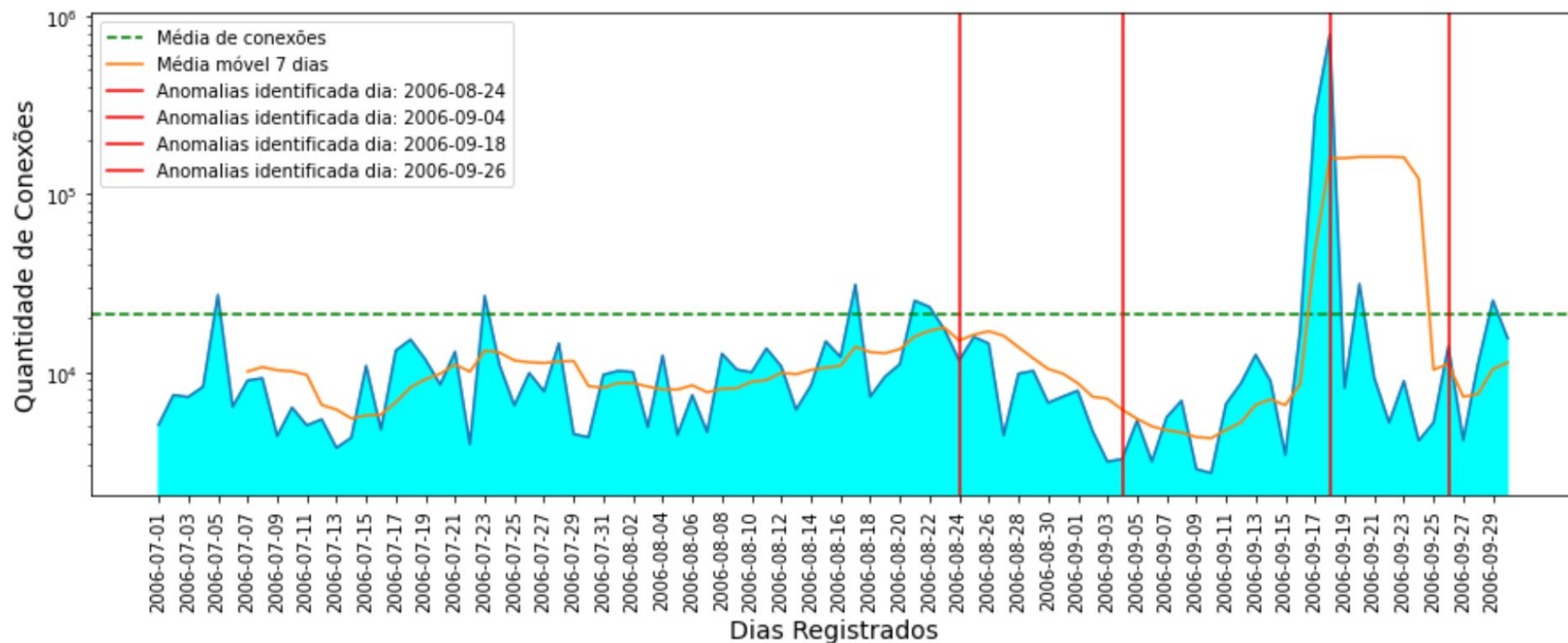
EXEMPLO SEM VISUALIZAÇÃO

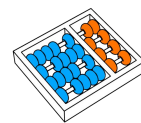
— — —

	data	ip local	numero conexoes	id sistema autonomo	indice ip
0	2006-07-01	192.168.1.2	1	701	0
1	2006-07-01	192.168.1.2	1	714	0
2	2006-07-01	192.168.1.2	1	1239	0
3	2006-07-01	192.168.1.2	1	1680	0
4	2006-07-01	192.168.1.2	1	2514	0
...
95	2006-07-01	192.168.1.7	6	11908	5
96	2006-07-01	192.168.1.7	2	13867	5
97	2006-07-01	192.168.1.7	8	14366	5
98	2006-07-01	192.168.1.7	1	14501	5
99	2006-07-01	192.168.1.7	2	15290	5



EXEMPLO COM VISUALIZAÇÃO



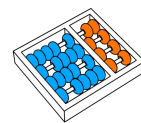


OBJETIVO DAS VISUALIZAÇÕES

Podemos gerar visualizações (gráficos) para diferentes objetivos, incluindo:

- Comparações
- Tendências
- Dispersões
- Relações
- Composições
- Categorias





OBJETIVO DAS VISUALIZAÇÕES

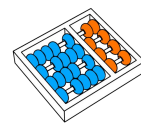
Podemos gerar visualizações (gráficos) para diferentes objetivos, incluindo:

- Comparações
- Tendências
- Dispersões
- Relações
- Composições
- Categorias

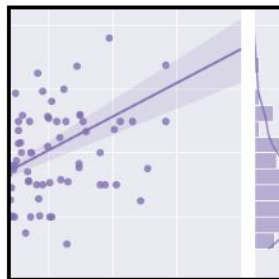
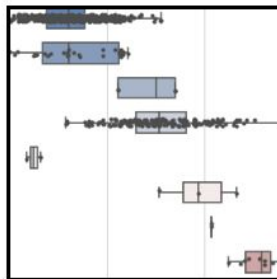
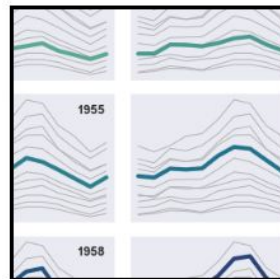
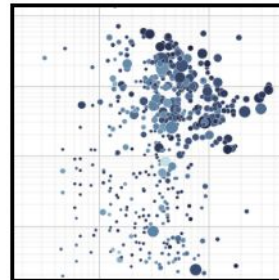
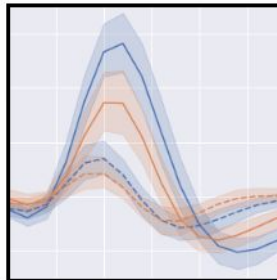
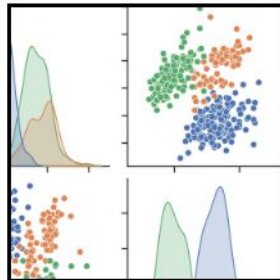


Para cada objetivo, temos um tipo de gráfico que melhor expressa o objetivo desejado!

TIPOS DE GRÁFICOS



- Barras
- Linhas
- Pizza
- Histogramas
- Boxplot
- Heatmaps
- Scatter plots
- Violinplot
- ...



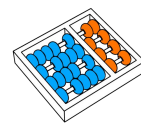


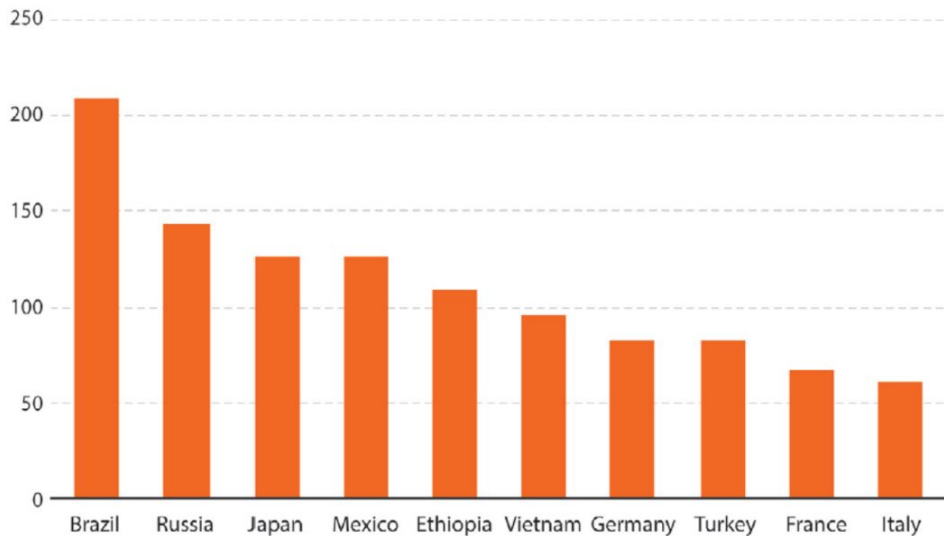
GRAFICO DE BARRAS

Ajudam a comparar dados entre diferentes itens. Também é utilizado para mostrar mudança ao longo do tempo

Fácil de identificar qual grupo (barra) possui maior valor

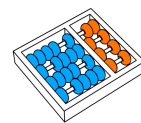
The total population in Brazil exceeds that of other countries

(Millions of people)



Source: The World Bank

GRÁFICO DE LINHAS



Mostram processo ou
tendências ao longo do tempo

Podem ser utilizados para
comparar diferentes grupos

Fácil de observar diferenças
entre os comportamentos

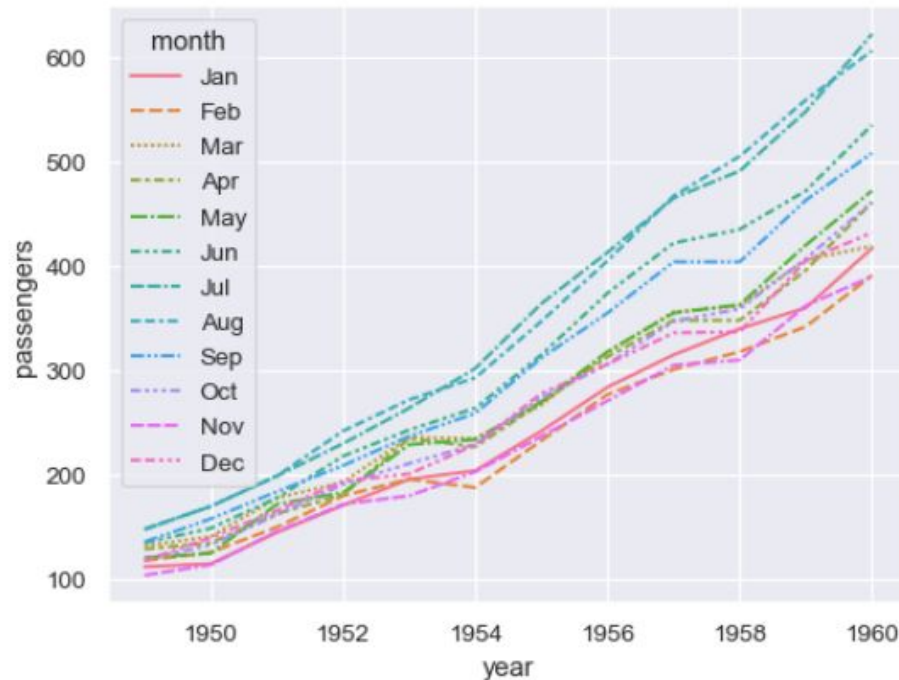


GRAFICO DE PIZZA

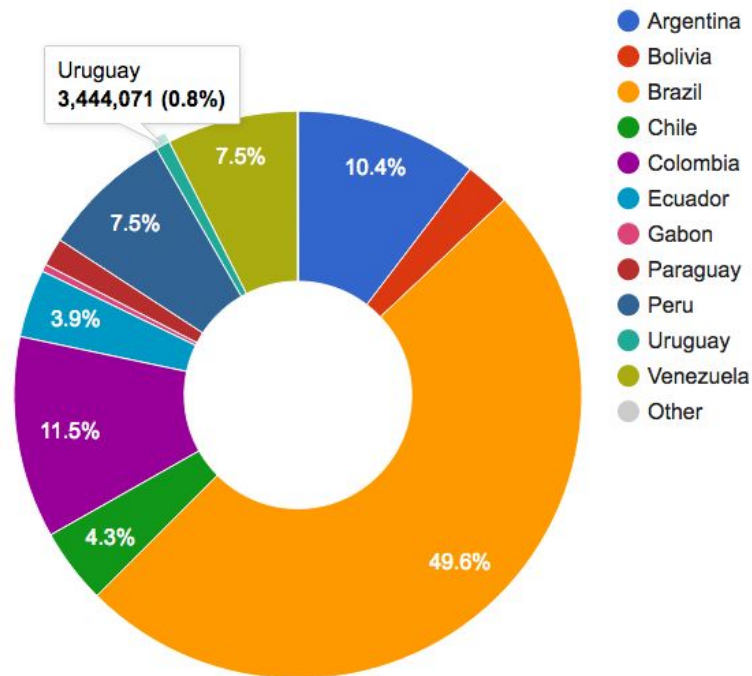
Mostram valores estáticos, geralmente utilizados para mostrar composições dos dados.

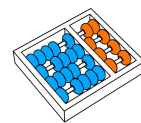
Representa números em porcentagens

Soma total de todos os segmentos precisa ser igual a 100%

Fácil de identificar o tamanho das partes em relação ao total

Population in South America



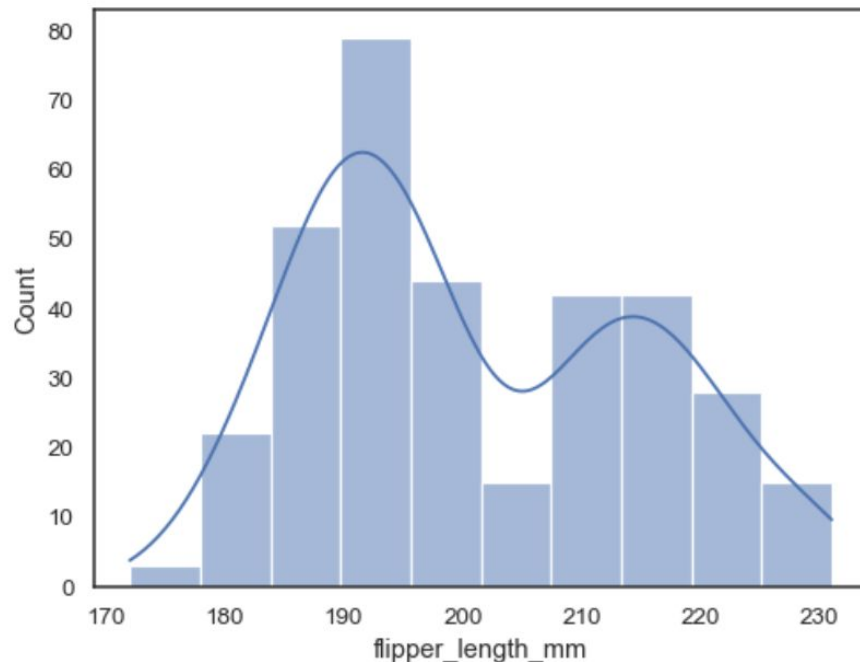


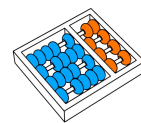
HISTOGRAMAS

É uma variação do gráfico de barras
me mostra uma distribuição de
frequência

É útil para resumir grandes
conjuntos de dados

Ideal para representar o
comportamento dos dados



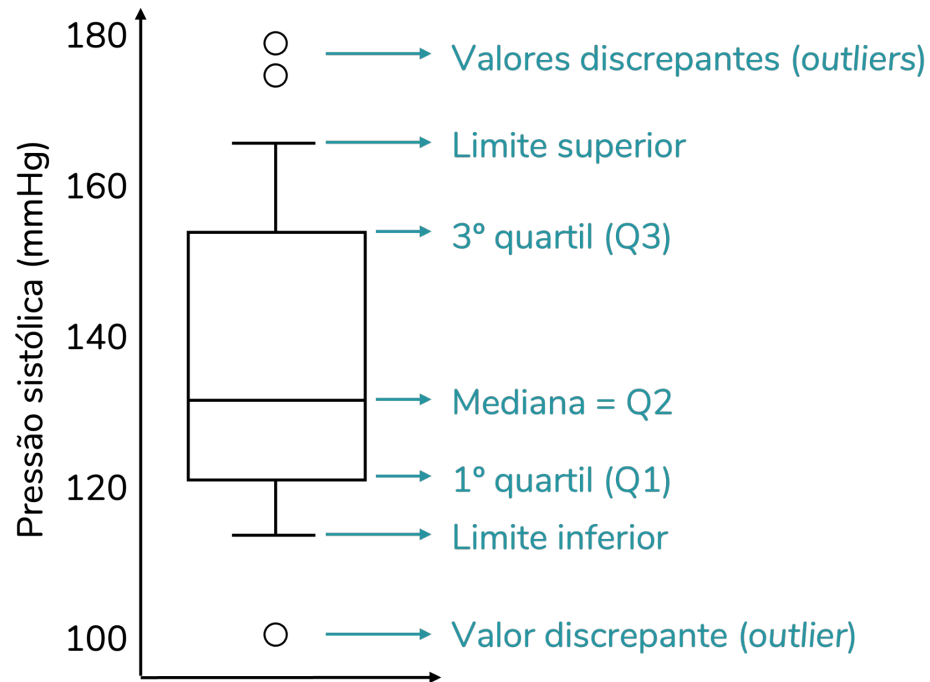


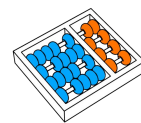
BOXPLOT

Ilustra o conjunto de dados apresentando

- Valor mínimo
- Primeiro quartil
- Mediana
- Terceiro quartil
- Valor máximo

Permite identificação de outliers (**anomalia**)



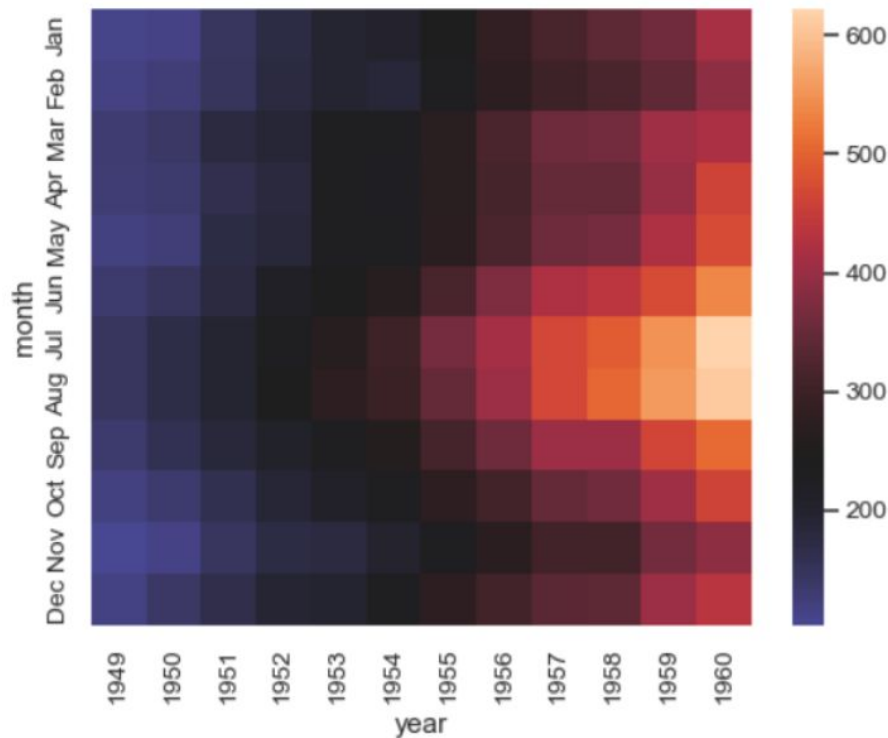


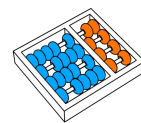
HEATMAPS

Mapas de calor

Usam cores e saturações para representar os valores dos dados

Simplificando, um mapa de calor é uma tabela com células codificadas por cores





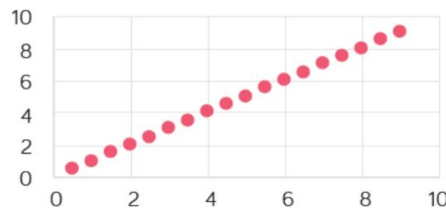
SCATTERPLOT - DISPERÇÃO

Utilizado para apresentar
correlações ou a **falta delas**

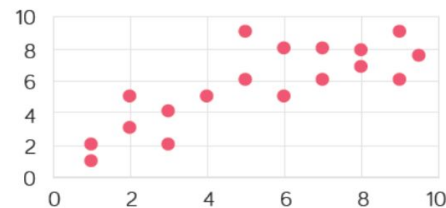
Uma variável é plotada ao longo
de um **eixo horizontal** e a outra
ao longo de um **eixo vertical**

Também é utilizado para mostrar
tendências

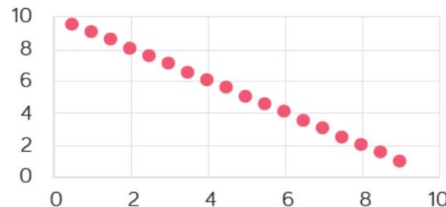
Perfectly Positively Correlated, $\rho=1$



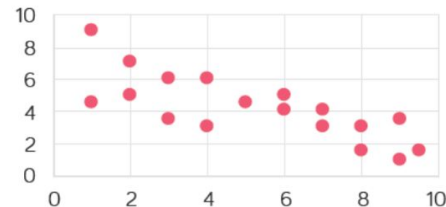
Strong Positive Correlation, $\rho=0.8$



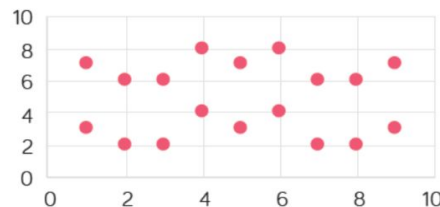
Perfectly Negatively Correlated, $\rho=-1$



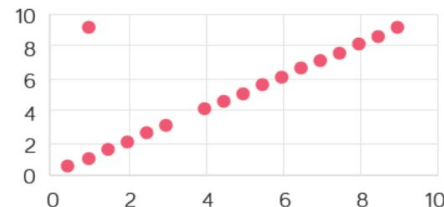
Strong Negative Correlation, $\rho=0.8$

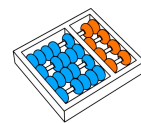


Zero Correlation, $\rho=0$



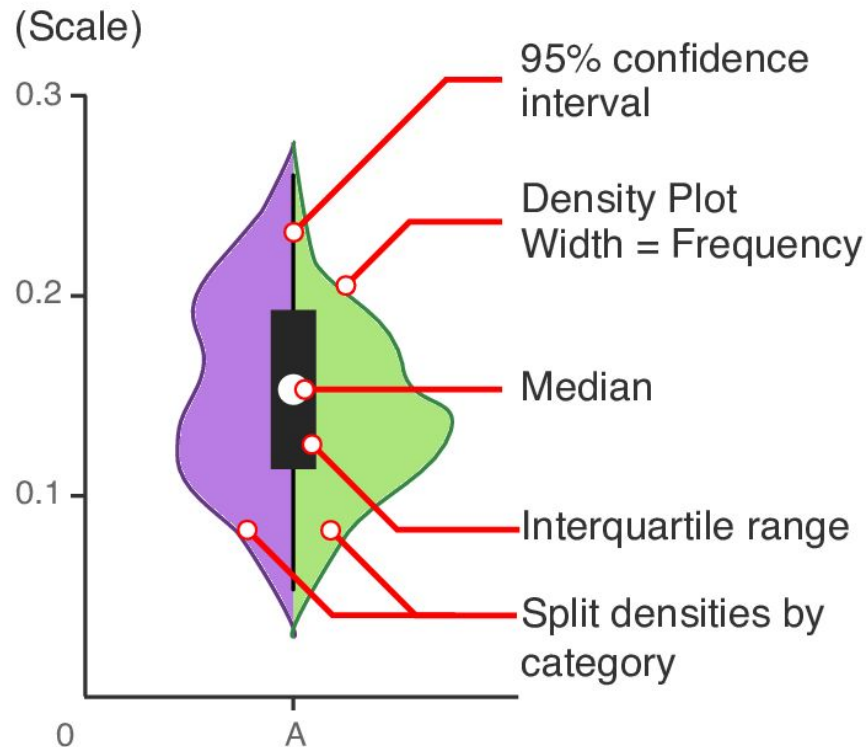
Positively Correlated with an Outlier, $\rho=0.8$

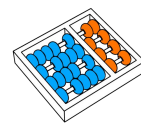




VIOLINPLOT

Combina as visualizações do boxplot com a do histograma para apresentar de forma mais detalhada o comportamento dos dados





FERRAMENTAS PARA VISUALIZAÇÃO

— — —

Neste curso estudaremos como gerar visualizações utilizando as seguintes ferramentas

- Pandas
- Matplotlib
- Seaborn

 pandas

 seaborn

matplotlib 

- Criação e configuração gráficos
- Pandas
- Matplotlib
- Seaborn

