

Tópicos em Informática

Visualização de dados

Profa. Dra. Kelly Lais Wiggers

Motivação

- A visualização de dados no formato tabular pode exigir muito esforço
 - Falhas de atenção
 - Sobrecarga de informação
- Capacidade cognitiva do ser humano
 - “O sistema visual humano tem um canal amplo para nossos cérebros” (Munzner, T. (2014). Visualization Analysis and Design. A K Peters/CRC Press)
 - Informações são perdidas durante o processo de sumarização estatística

O que veremos hoje:

- Bibliotecas gratuitas de visualização de dados
- Como entender gráficos

Distribuição



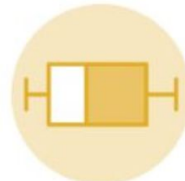
Violin



Density



Histogram



Boxplot



Ridgeline

1. Histograma

2. Boxplot

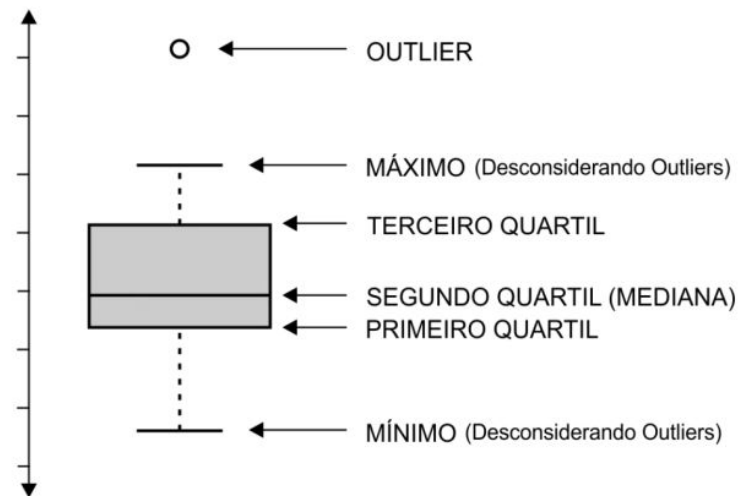
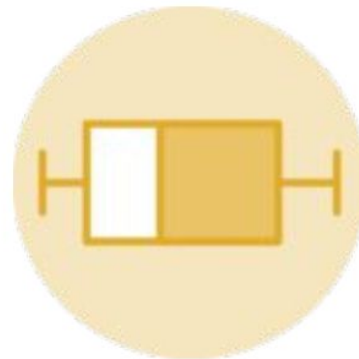
Histograma

- Distribuição de dados numéricos
- Bins
 - Explore valores de Bins
- Não é o mesmo que um gráfico de barras
- <https://www.data-to-viz.com/graph/histogram.html>
- <https://plotly.com/python/histograms/>



BoxPlot

- Representação estatística de variáveis através de seus quartis
 - IQR
- Não representa quantidade de observações
- Outliers



Fonte: abgconsultoria.com.br

Vejamos alguns exemplos

Correlação

- Gráficos de dispersão
 - Scatter plot
 - Matriz de Scatter Plot
- Mapa de calor (Heatmap)



Scatter



Heatmap



Correlogram



Bubble



Connected scatter



Density 2d

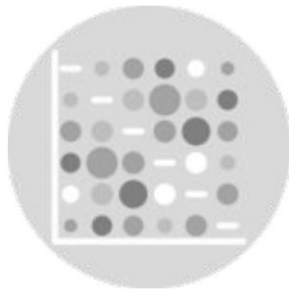
Gráficos de dispersão

- Scatter plot
 - Distribuição entre 2 variáveis numéricas
 - Pode ser enriquecido por distribuições marginais
 - Evite Overplotting
 - <https://www.data-to-viz.com/graph/scatter.html>
 - <https://plotly.com/python/line-and-scatter/>



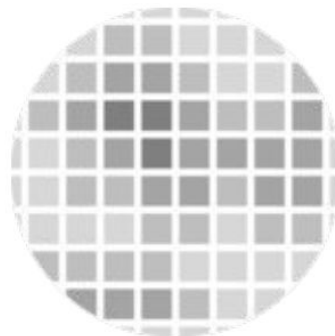
Gráficos de dispersão

- Matriz de Scatter plot (correlograma)
 - Útil para análise exploratória
 - Visualiza as relações entre as diversas variáveis do conjunto de dados
 - <https://www.data-to-viz.com/graph/correlogram.html>
 - <https://plotly.com/python/splom/>



Gráficos de dispersão

- Mapa de calor (Heatmap)
 - Representa valores numa tabela por intensidades de cores
 - Dados normalizados
 - Séries temporais
 - <https://www.data-to-viz.com/graph/heatmap.html>
 - <https://plotly.com/python/heatmaps/>



Vejamos alguns exemplos

Ranqueamento

- Gráfico de barras
- Coordenadas paralelas
- Nuvem de palavras



Barplot



Spider / Radar



Wordcloud



Parallel



Lollipop



Circular Barplot

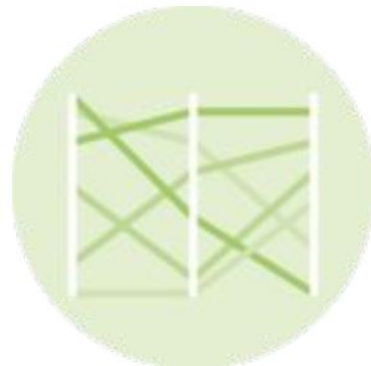
Gráfico de barras

- Relação entre uma variável categórica e uma métrica numérica
- Barras ordenadas são mais intuitivas
- **Não** é um histograma
- <https://www.data-to-viz.com/graph/barplot.html>
- <https://plotly.com/python/bar-charts/>



Coordenadas paralelas

- Comparação entre múltiplas variáveis
 - Podem ser heterogêneas
- Eixos verticais
- Relações entre variáveis
- Evite overplotting (Spaghetti plot)
- <https://www.data-to-viz.com/graph/parallel.html>
- <https://plotly.com/python/parallel-coordinates-plot/>



Nuvem de palavras

- Representação da relevância de palavras
 - Cor
 - Tamanho
- Máscaras (formas)
- <https://www.data-to-viz.com/graph/wordcloud.html>
- https://amueller.github.io/word_cloud/



Parte de um todo (hierarquias)

- Gráfico de setores
- Dendograma



Treemap



Venn diagram



Doughnut



Pie chart



Dendrogram



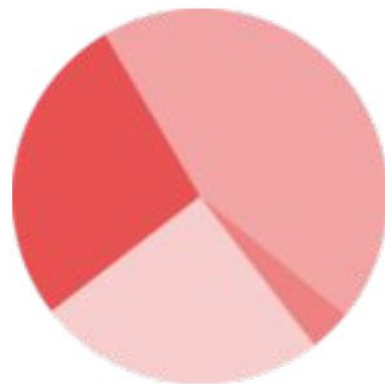
Circular packing



Sunburst

Gráfico de setores

- Gráfico de “pizza”
- Representa porções de um todo (%)
 - ○ Soma de todos os setores = 100%
- Não confundir com o Sunburst Plot
- <https://www.data-to-viz.com/caveat/pie.html>
- <https://plotly.com/python/pie-charts/>



Dendograma

- organiza determinados fatores e variáveis
- se emprega um método quantitativo que leva a agrupamentos
- os agrupamentos são formados em cada passo e para avaliar os níveis de similaridade (ou distância) dos agrupamentos
- <https://plotly.com/python/dendrogram/>



Evolução

- Gráfico de linhas
- Gráfico de área



Line plot



Area



Stacked area



Streamchart

Gráfico de linhas

- Representa a evolução de uma ou várias variáveis numéricas
- Também utilizado em Scatter plots para representar tendências e padrões
 - p.e. Linha de regressão
- Também sofre do problema de **Spaghetti plot**
- <https://www.data-to-viz.com/graph/line.html>
- <https://plotly.com/python/line-charts/>



Gráfico de área

- Representa a evolução de um conjunto de dados todo
- Grupos
 - Proporções relativas
- Valor relativo representado pela largura da “onda” no ponto x
- <https://www.data-to-viz.com/graph/stackedarea.html>
- <https://plotly.com/python/filled-area-plots/>



Vejamos alguns exemplos

Qual gráfico utilizar?

- Verificar o tipo de conjunto de dados: categórico, numérico, híbrido, etc.
- Acessar o catálogo: <https://www.data-to-viz.com/> e verificar as sugestões.