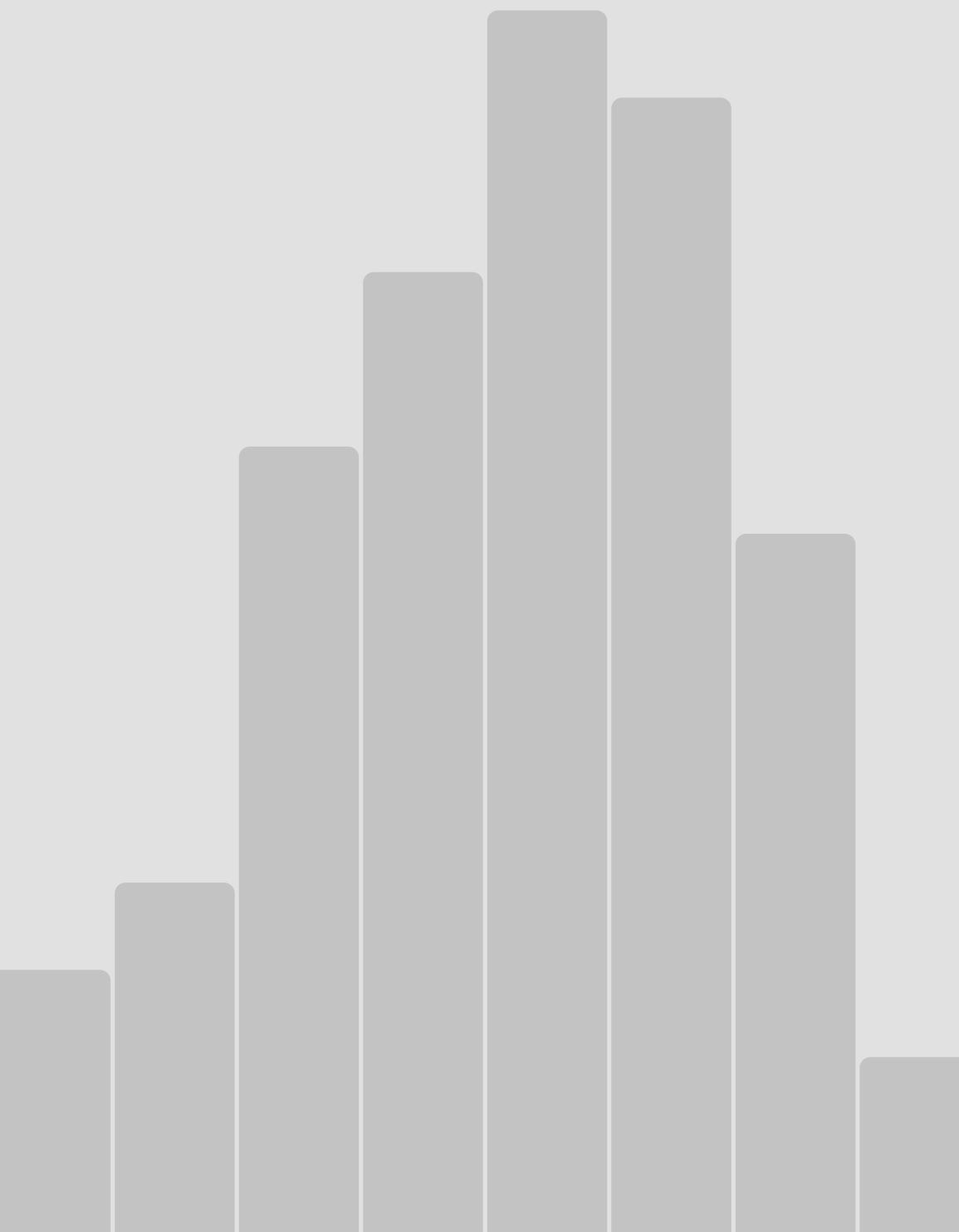


# Visualização de dados para apuração

---

Como jornalistas podem extrair o maior potencial de ferramentas de visualização de dados?

# Nas aulas anteriores

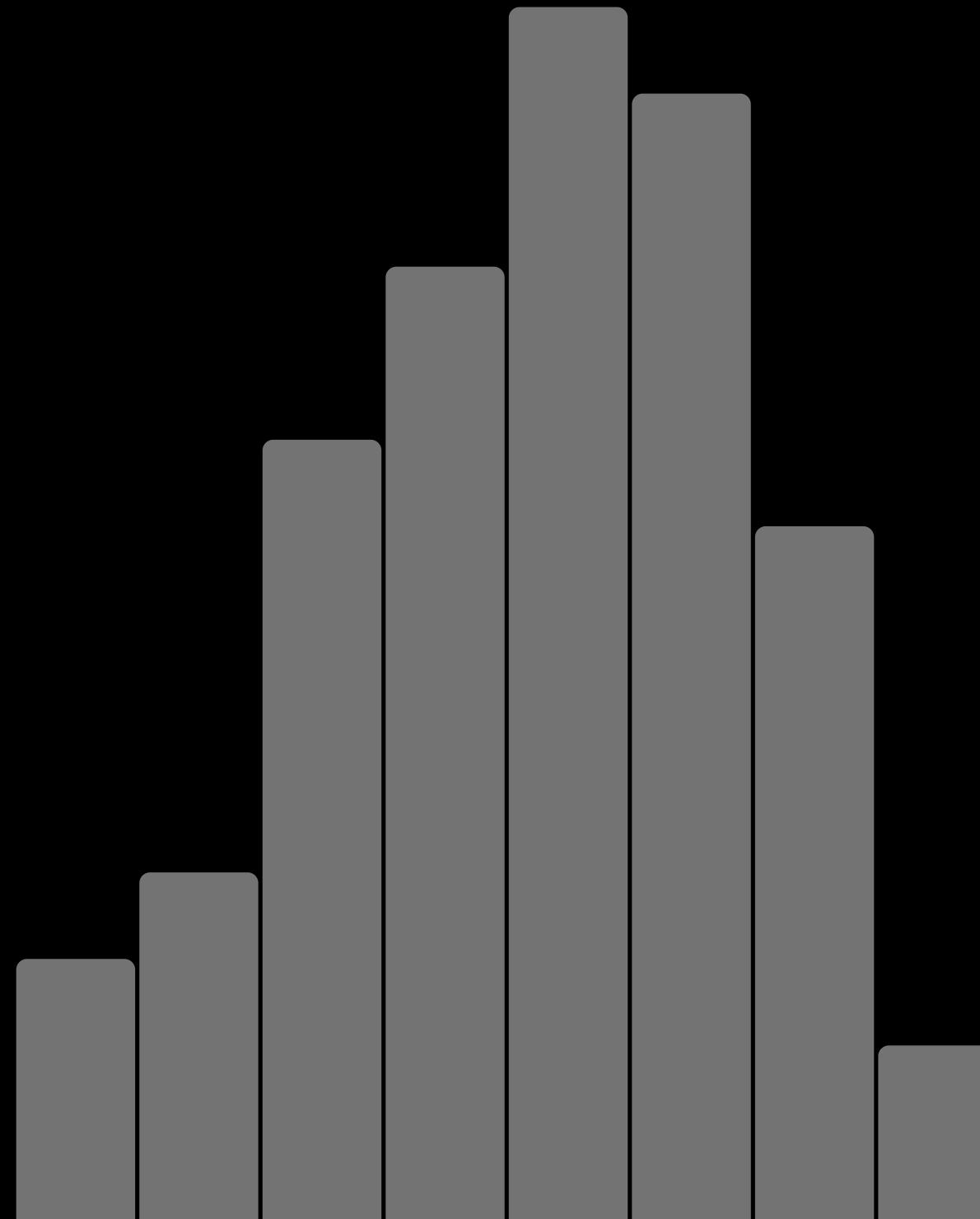


**Fundamentos,  
crítica e redesign**  
Profa. Júlia Giannella

**Introdução à  
visualização  
para apuração**  
Prof. Vinicius Soeiro

- O que é visualização para apuração?
- Revisão de medidas estatísticas descritivas
- Gráficos para apuração

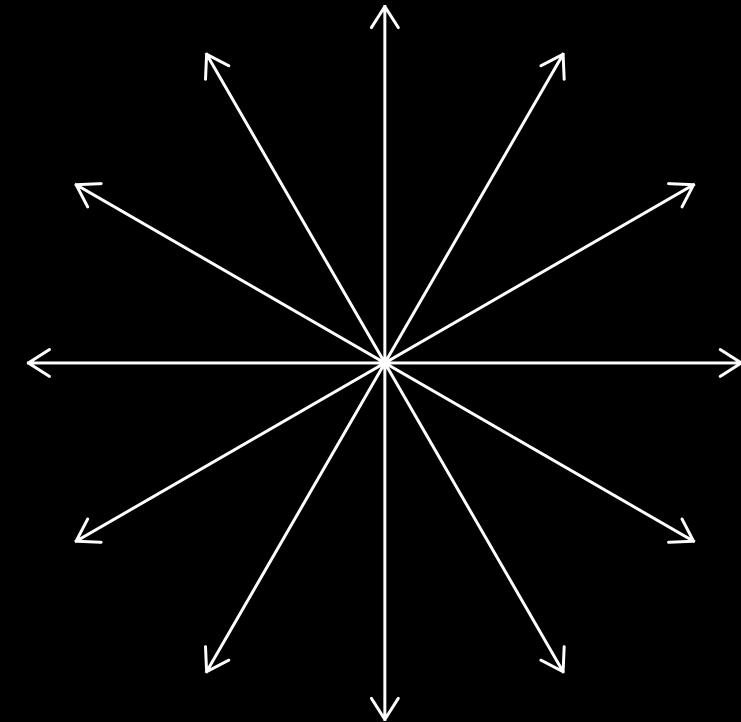
# Hoje e amanhã



1. Análise  
Exploratória de  
Dados para apuração

2. Ferramentas para  
apuração

# Análise exploratória de dados



## Tipos de variáveis

Com que variáveis estamos lidando:  
qualitativas ou quantitativas?  
contínuas ou discretas? Nominais ou  
ordinais? Espaciais? Temporais?

## Valores ausentes

Existem dados faltantes na sua amostra?

## Medidas Descritivas

Medidas de posição, medidas de dispersão, medidas de forma, medidas de correlação.

# Medidas descriptivas

## Posição

média aritmética, mediana, moda, média geométrica; quartis, decis, percentis, máximo e mínimo

## Forma

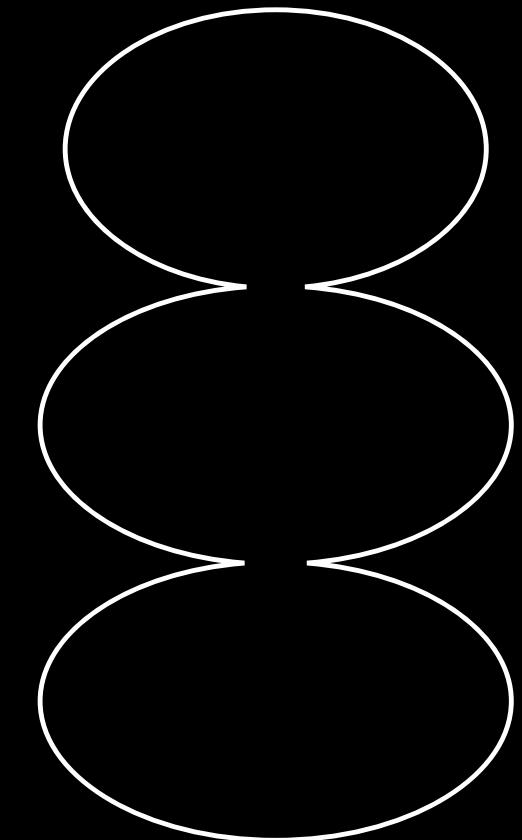
coeficiente de assimetria e curtose

## Dispersão

amplitude total, variância, desvio-padrão, desvio absoluto, coeficiente de variação.

## Associação

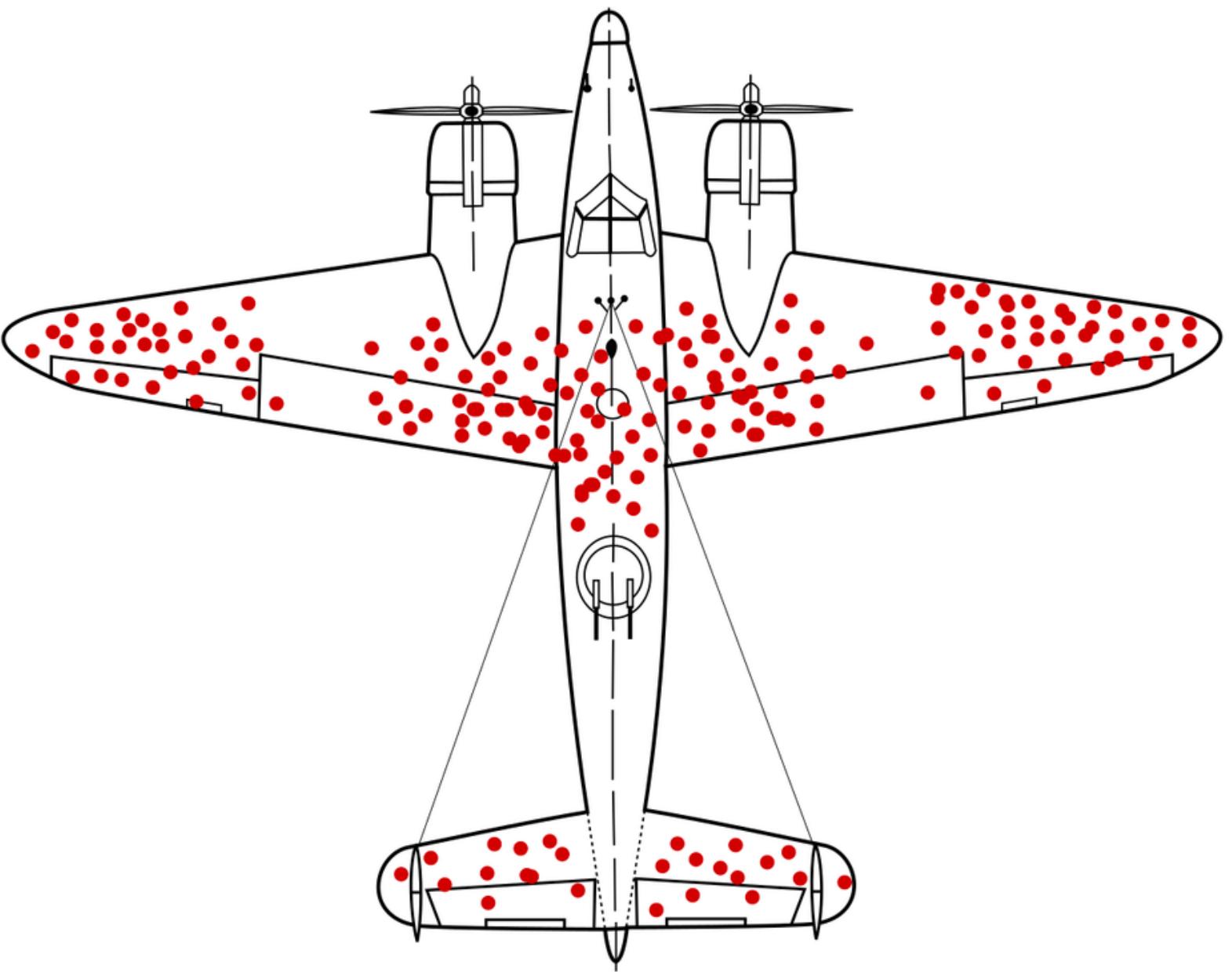
correlação de pearson, correlação de spearman



# Pontos de atenção na apuração



# Viés de sobreviêvncia



# Consistência

CORONAVÍRUS < □

## Prefeituras negam aplicação de vacina vencida e falam em erro de registro no sistema

Segundo registros oficiais, cerca de 26 mil doses da AstraZeneca fora da validade foram aplicadas em 1.532 municípios

2.jul.2021 às 21h37

CORONAVÍRUS · TEXTOS LIBERADOS

Folha errou ao não afirmar que dados sobre vacinas vencidas poderiam decorrer de falhas do sistema; texto foi alterado

Reportagem apontou problemas no processo de vacinação e registro de informações; quem recebeu AstraZeneca deve conferir lote e validade no cartão

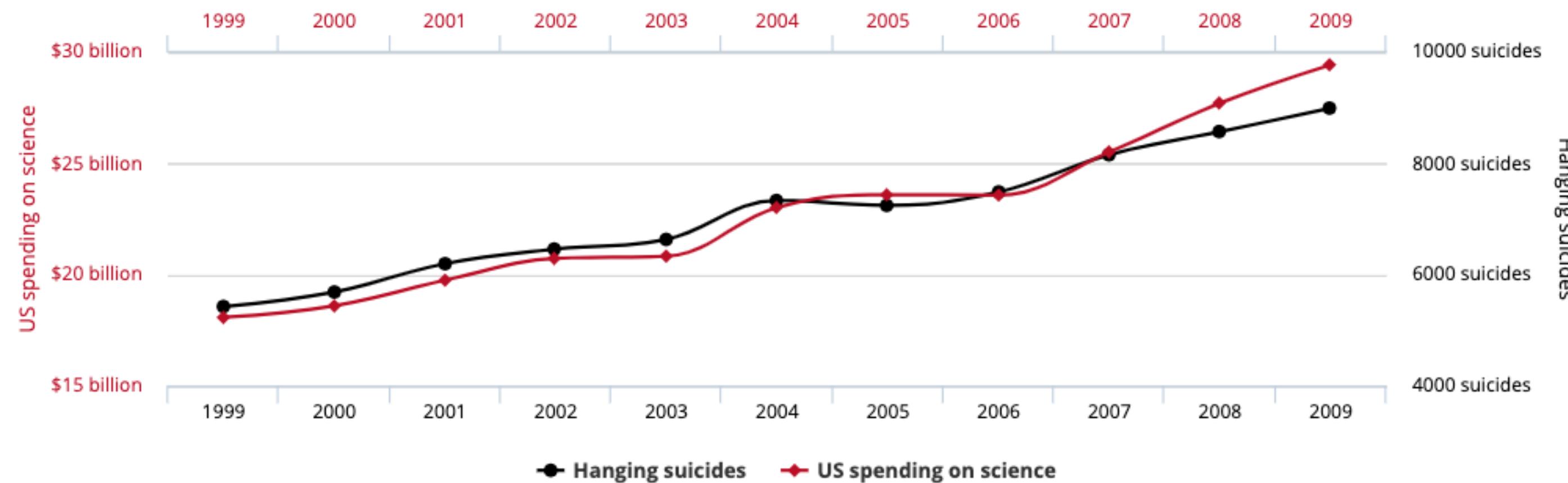
Correlação não é causalidade

## US spending on science, space, and technology

correlates with

## Suicides by hanging, strangulation and suffocation

Correlation: 99.79% ( $r=0.99789126$ )



Fontes de dados: Escritório de Administração e Orçamento dos EUA e Centros para Controle e Prevenção de Doenças

tylervigen.com

[www.tylervigen.com](http://www.tylervigen.com)

causalidade

X → Y

# Relação Espúria

---

É uma relação estatística existente entre duas variáveis, mas onde não existe nenhuma relação causa-efeito entre elas. Essa relação estatística pode ocorrer por pura coincidência ou por causa de uma terceira variável (fator de confusão).

# Prática 1

Análise Exploratória de dados do ENEM

com Pandas e Seaborn

[Abrir código no Colab](#)

[Baixar dados do ENEM](#)



[pandas.pydata.org](https://pandas.pydata.org)

---

Pandas é uma ferramenta de análise e manipulação de dados de código aberto rápida, poderosa, flexível e fácil de usar, construída sobre a linguagem de programação Python.



[seaborn.pydata.org](https://seaborn.pydata.org)

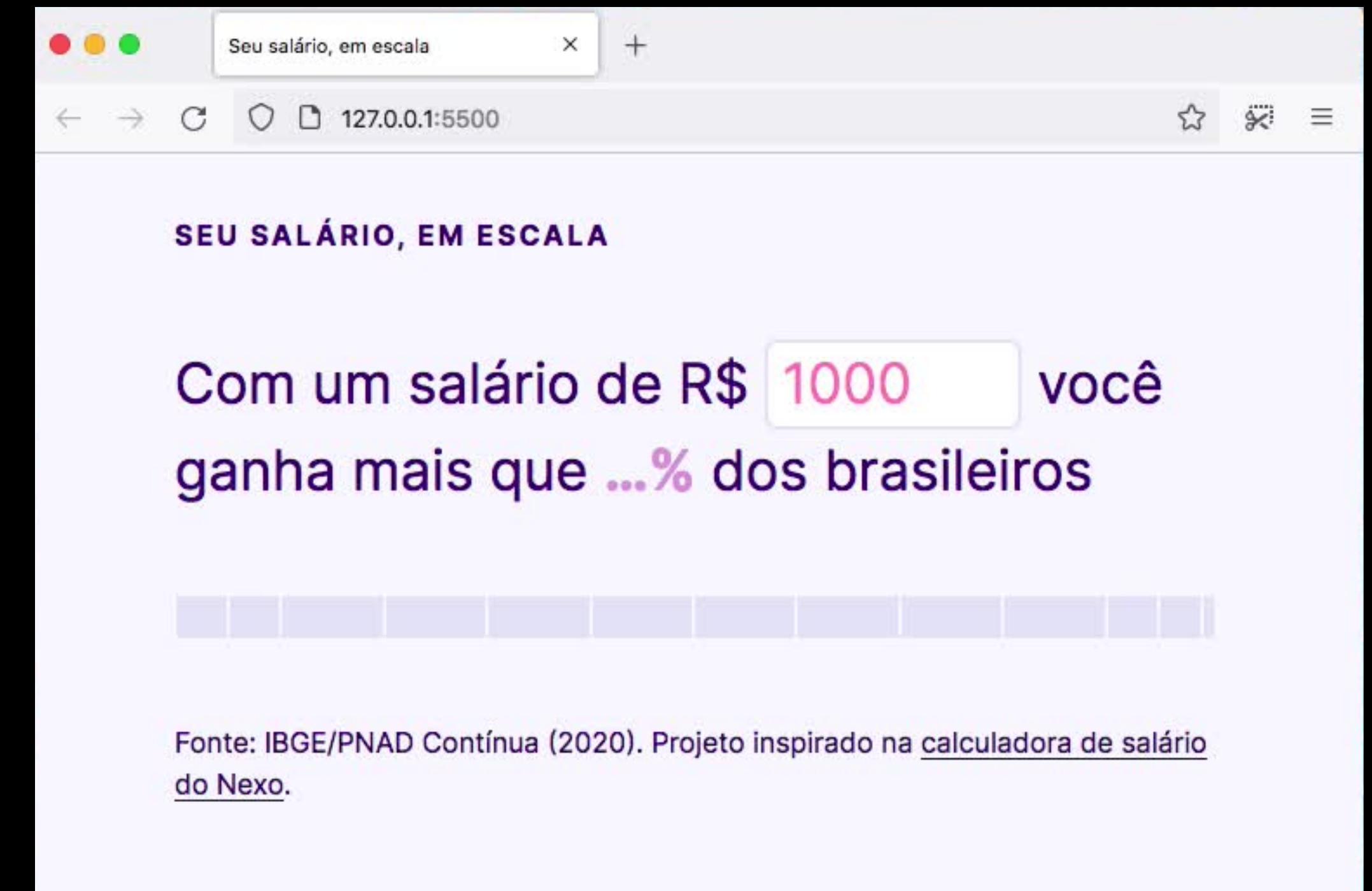
---

Seaborn é uma biblioteca de visualização de dados Python baseada em matplotlib. Ela fornece uma interface de alto nível para desenhar gráficos estatísticos atraentes e informativos.

## Prática 2

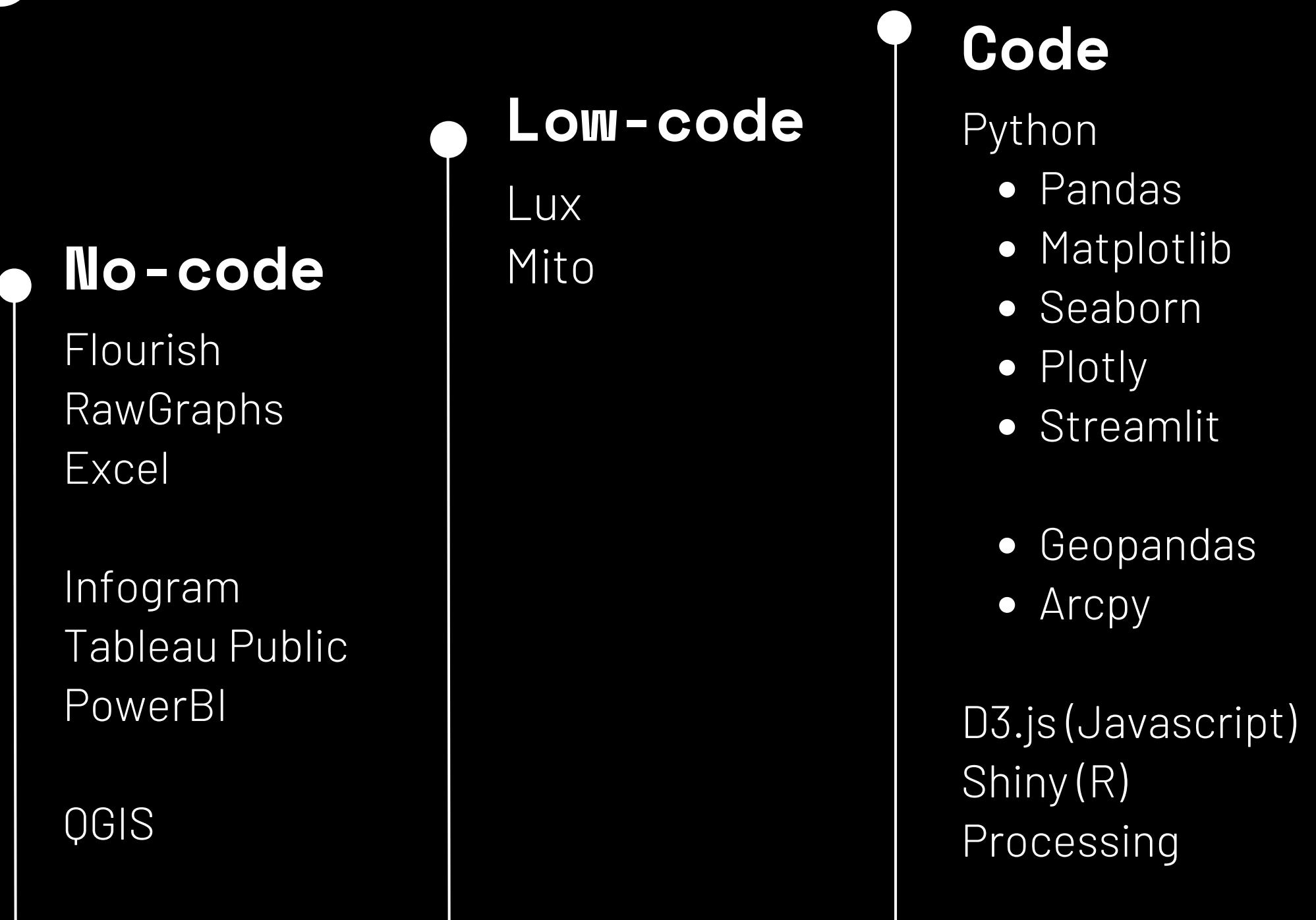
Extraindo e processando  
dados para calculadora

# Projeto Calculadora de Salário



O que é PNAD-IBGE?

# Ferramentas para apuração



# No-code

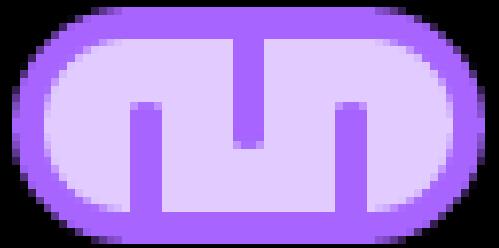


+tableau<sup>+</sup>+public



QGIS

# Low-code



[trymito.io](https://trymito.io)

---

Mito é uma planilha que mora dentro de seus notebooks JupyterLab. Ele permite que você edite dataframes do Pandas como um arquivo do Excel e gera código Python que corresponde a cada uma de suas edições.



[github.com/lux-org/lux](https://github.com/lux-org/lux)

---

Lux é uma biblioteca Python que facilita a exploração de dados rápida e fácil, automatizando o processo de visualização e análise de dados.

# Code



# Qual a melhor ferramenta?

1.1052	.90484	.10017	1.0050	.09967
1.2214	.81873	.20134	1.0201	.19738
1.3499	.74082	.30452	1.0453	.29131
1.4918	.67032	.41075	1.0811	.37995
1.6487	.60653	.52110	1.1276	.46212
1.8221	.54881	.63665	1.1855	.53705
2.0138	.49659	.75858	1.2552	.60437
2.2255	.44933	.88032	1.3374	.66404
2.4596	.40657	.98111	1.4331	.71630
2.7183	.36381	1.1752	1.5181	.76159
3.0042	.32356	1.3356	1.6881	.80050
3.3201	.28695	1.5095	1.8536	.83365
3.6693	.25353	1.6984	1.9787	.86172
4.0552	.22660	1.9043	2.1530	.88535
4.4817	.197313	2.1293	2.3522	.90515
4.9530	.170190	2.3756	2.5774	.92167
5.4739	.148268	2.6456	2.8284	.93541
6.0496	.126530	2.9422	3.1074	.94681
6.6859	.104957	3.2682	3.4174	.95624
7.3891	.08534	3.6269	3.7774	.96403
8.1662	.0686	4.0219	4.1774	.97045
9.0250	.05250	4.4571	4.6074	.97574
9.9742	.039370	4.9370	5.0874	.98010
11.023	.027569	5.69	5.8474	.98367
12.182	.017632	6.1323	6.2823	.98661
13.464	.01147	6.7690	6.9190	.98903
14.880	.008063	7.4735	7.6235	.99101
16.445	.0051919	8.2527	8.4027	.99263
18.174	.0030596	9.1146	9.2646	.99396
20.086	.0010.018	10.068	10.218	.99505
22.198	.0011.076	11.122	11.272	.99595
24.533	.0012.246	12.287	12.437	.99668
27.113	.0013.538	13.575	13.725	.99728
29.964	.0014.965	14.999	15.149	.99777
33.115	.0016.543	16.573	16.723	.99818
36.598	.0018.285	18.313	18.463	.99851
40.447	.0020.211	20.236	20.386	.99878

Qual a melhor  
ferramenta?

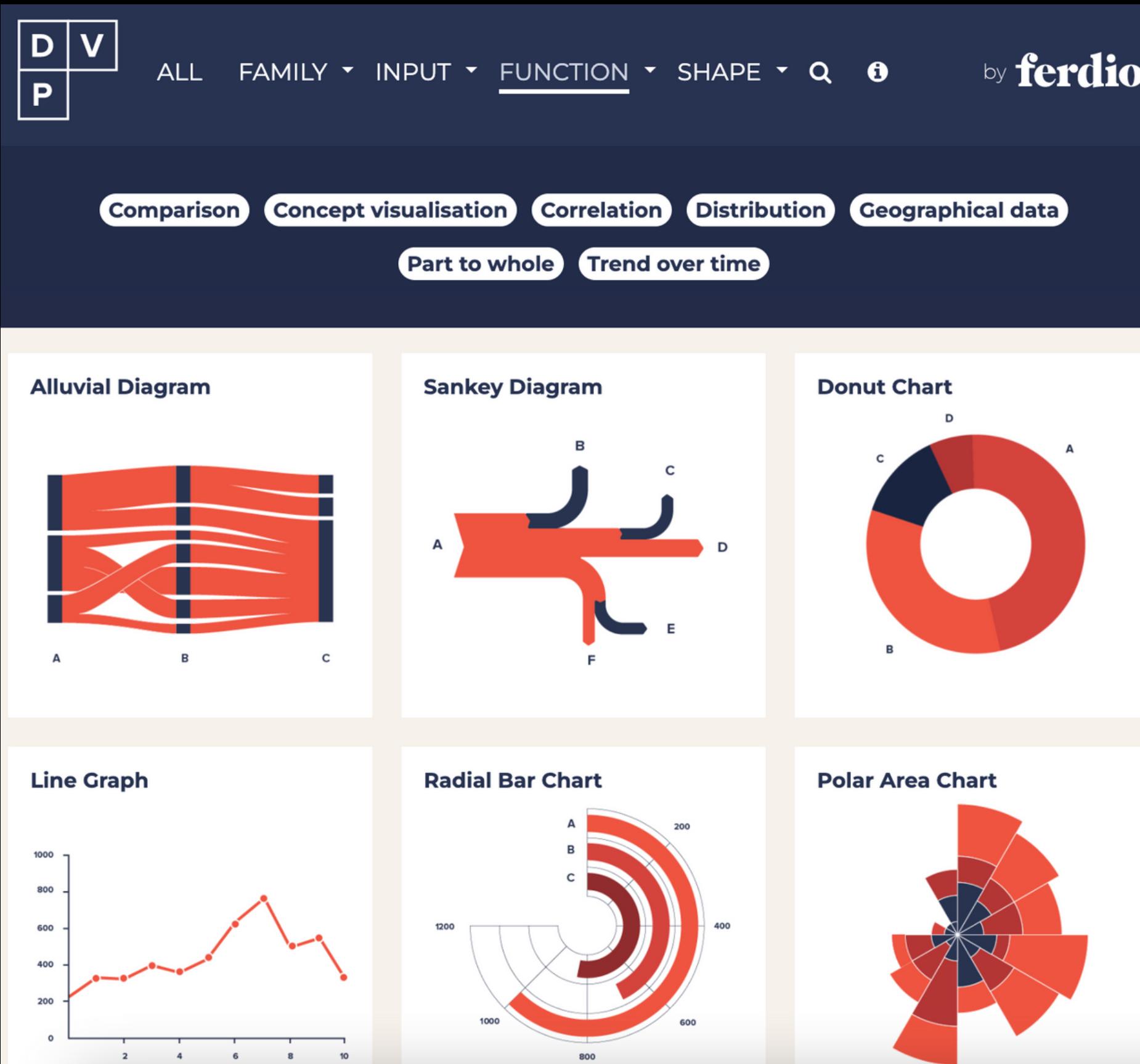
Aquela que você sabe utilizar :)



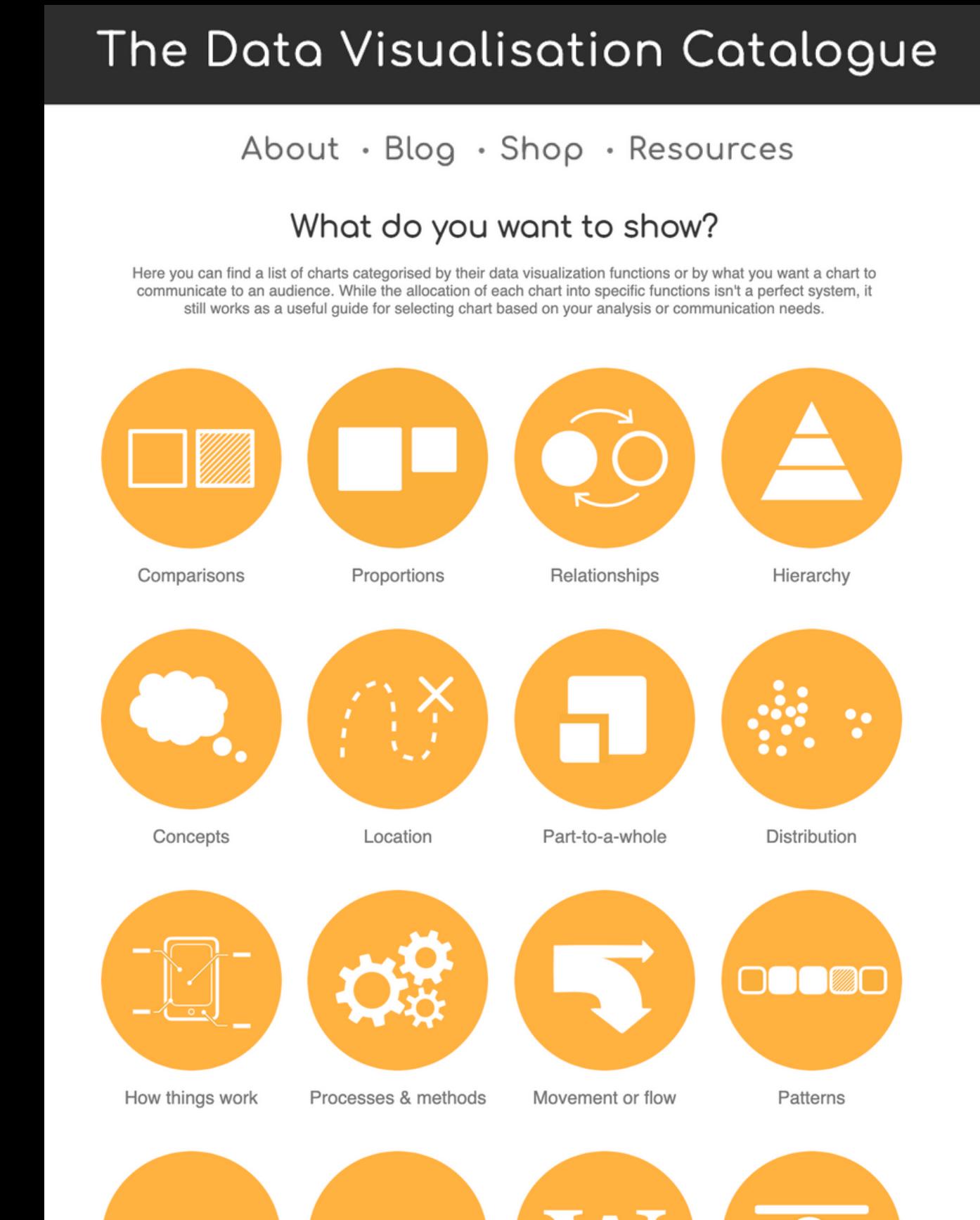
# Como escolher uma ferramenta para apuração?

- 1. Escopo**
- 2. Recursos**
- 3. Público**
- 4. Objetivo**
- 5. Entregáveis**
- 6. Escolhas visuais**

Como escolher um gráfico?

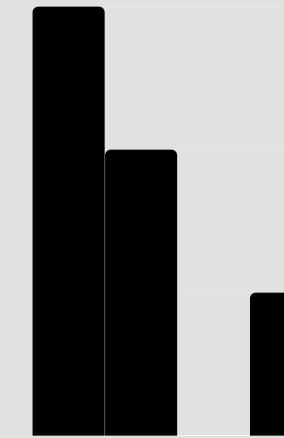


<https://datavizproject.com/#>

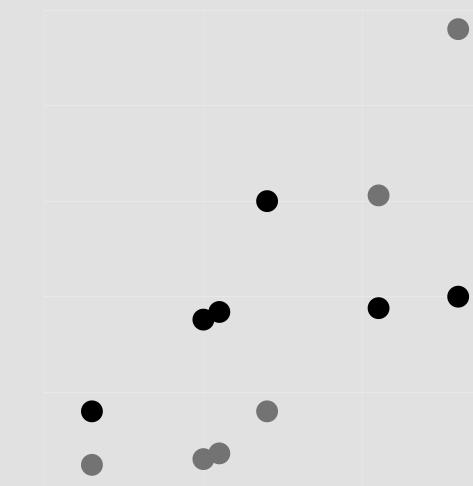


<https://datavizcatalogue.com/>

# Gráficos para apuração



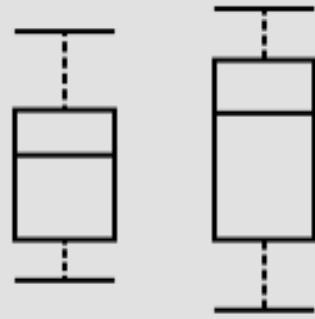
Histograma



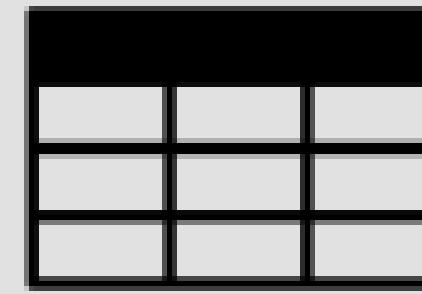
Dispersão



Linhas



BoxPlots



Tabelas



Mapas

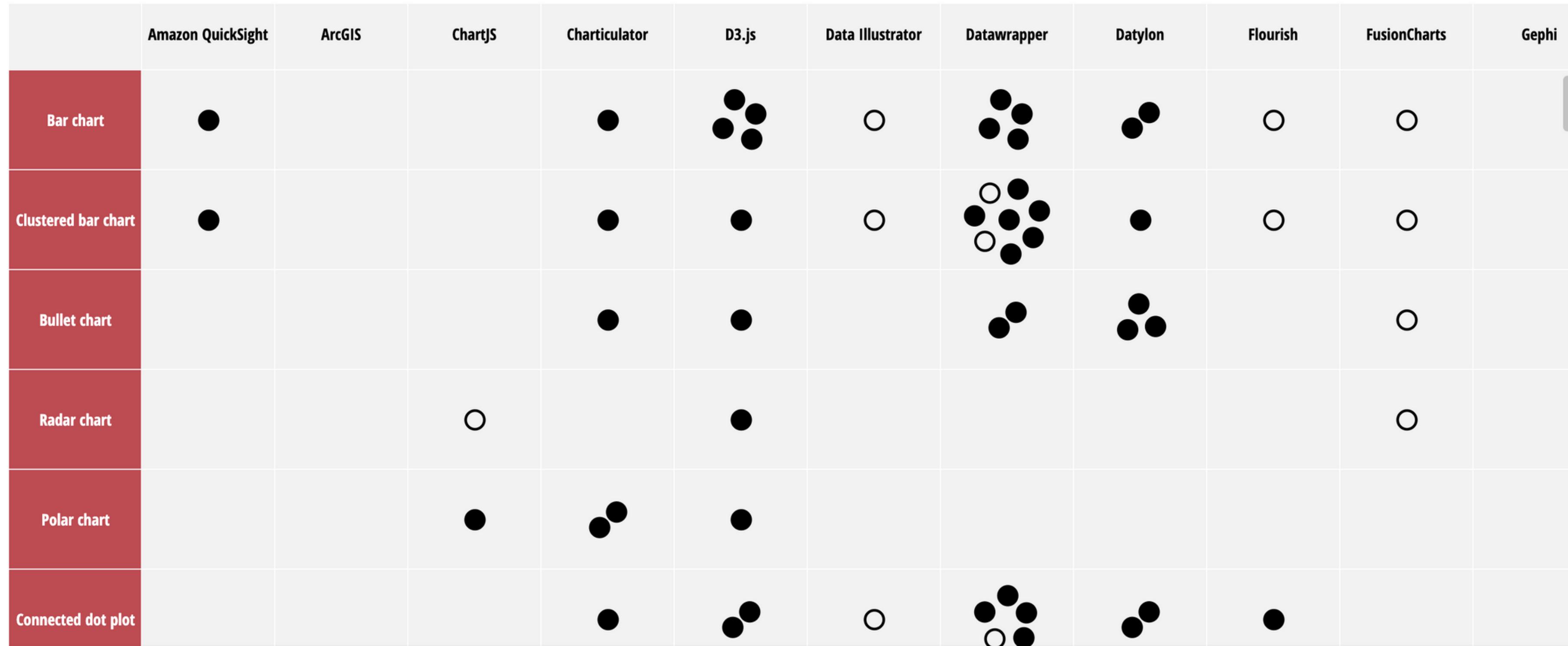
Como escolher  
uma ferramenta  
para um gráfico?

# ⊕ THE CHARTMAKER DIRECTORY

ABOUT

≡ Filter by chart name or AKA

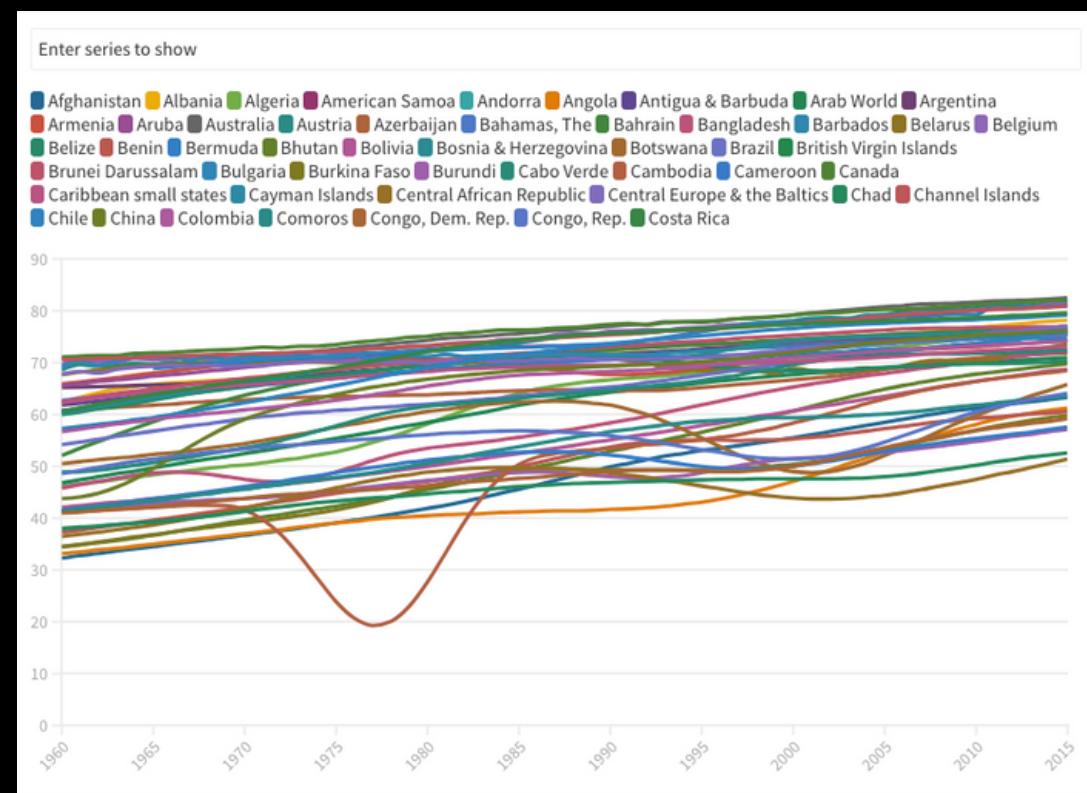
Reference Type: ● Example ○ Solution | Chart Families: ● Categorical ○ Hierarchical ● Relational ○ Temporal ● Spatial



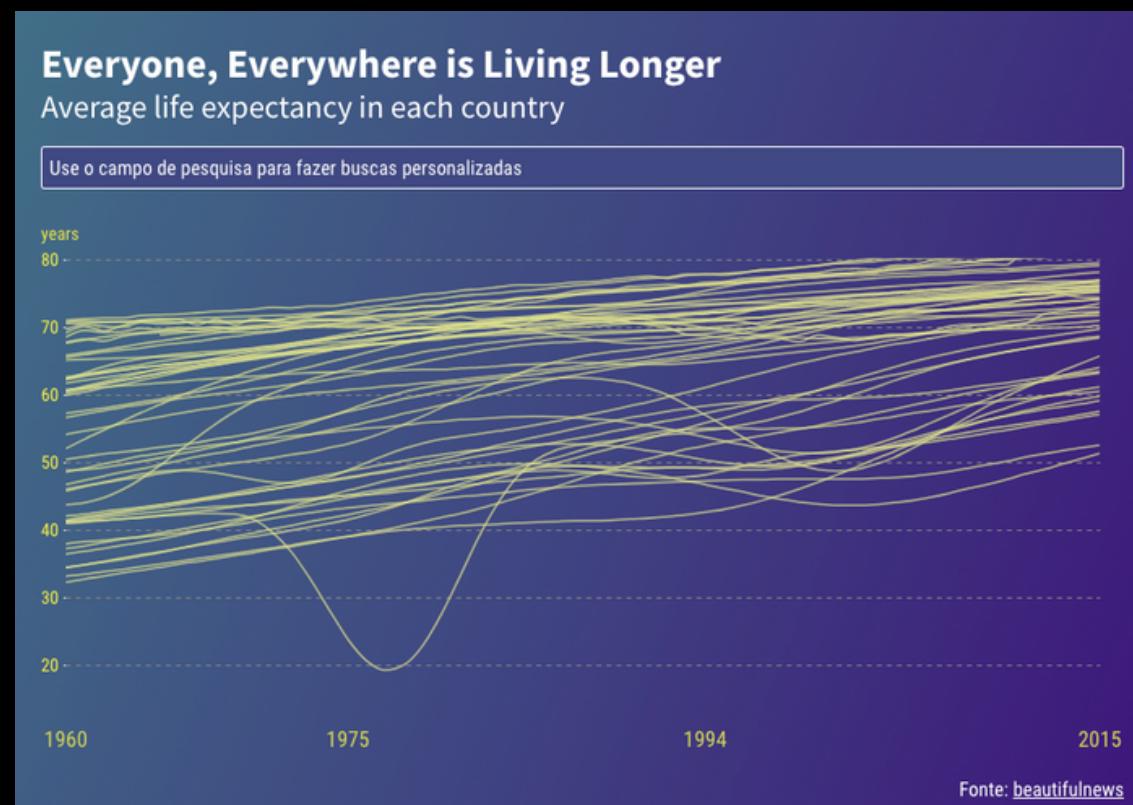
<http://chartmaker.visualisingdata.com/>

# Prática 3

## Personalização no Flourish



Antes



Depois

Ver no Flourish

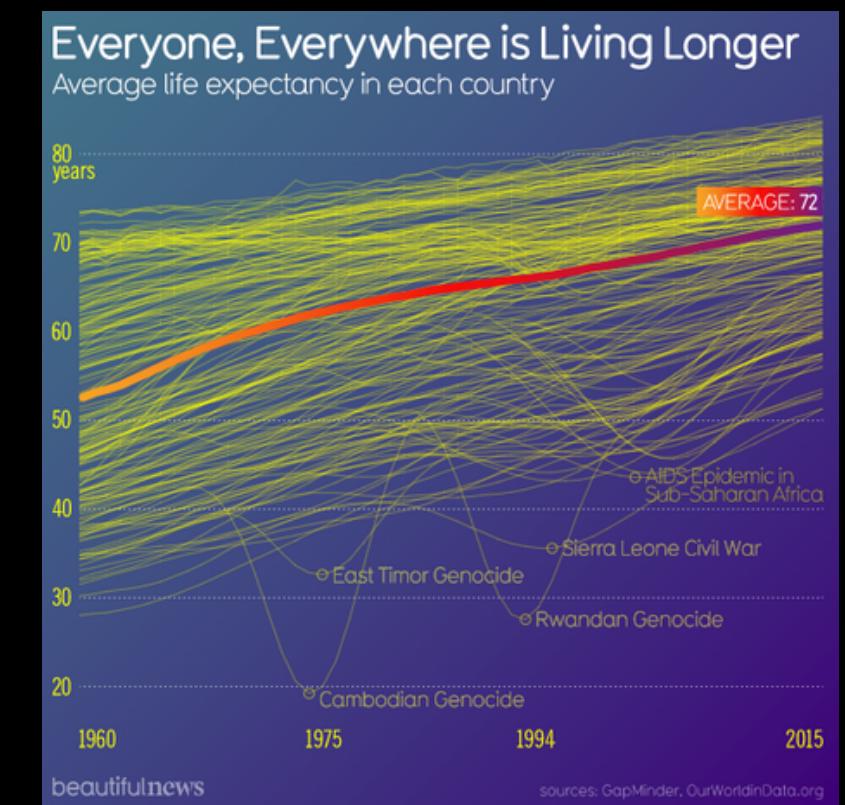


Gráfico original

Fonte de dados

Prática 4

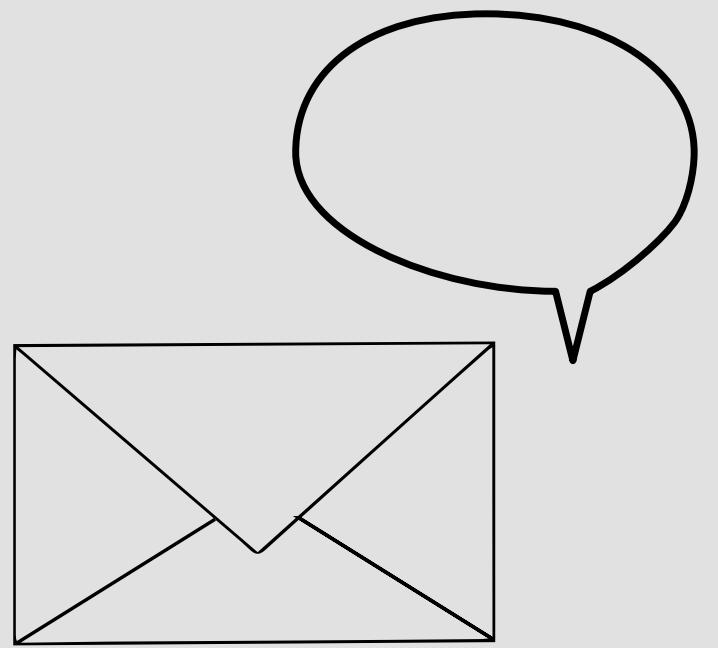
Matplotlib

[matplotlib.org](http://matplotlib.org)

# Prática 5

# Plotly

[plotly.com/python](https://plotly.com/python)



Entre em  
contato

**[youtube.com/programacaodinamica](https://youtube.com/programacaodinamica)**

Youtube

**kizzy@pgdinamica.com**

Email

**@kizzy\_terra | @pgdinamica**

Twitter, Instagram, LinkedIn