

Mestrado Data Science

Data Visualization in Python



Conteúdo

Sessão nº 1

01

Motivação

02

Tipos de gráficos

03

Más práticas

04

Avaliação

05

Outras ferramentas

06

Atividade de grupo

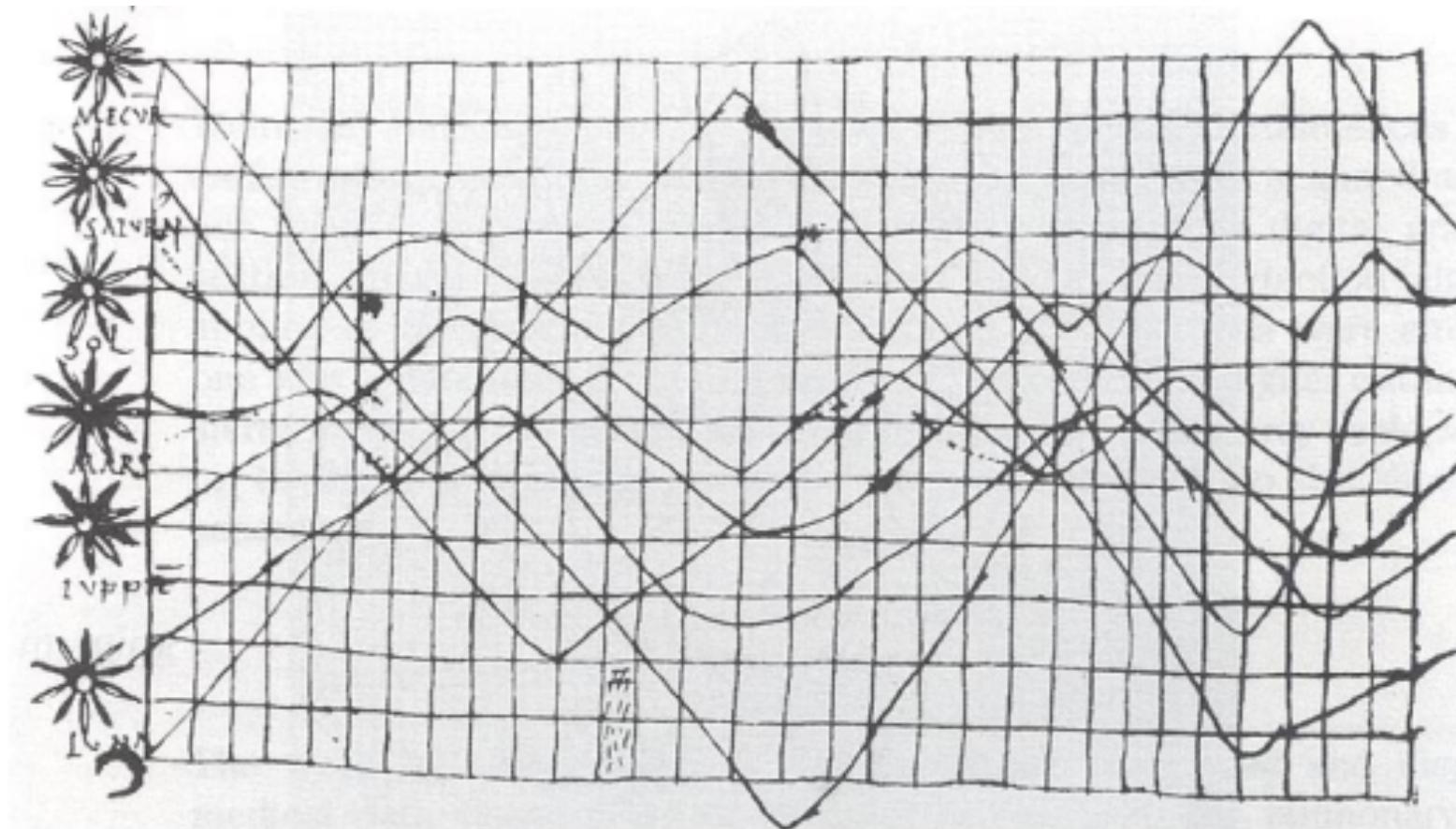
01

Motivação

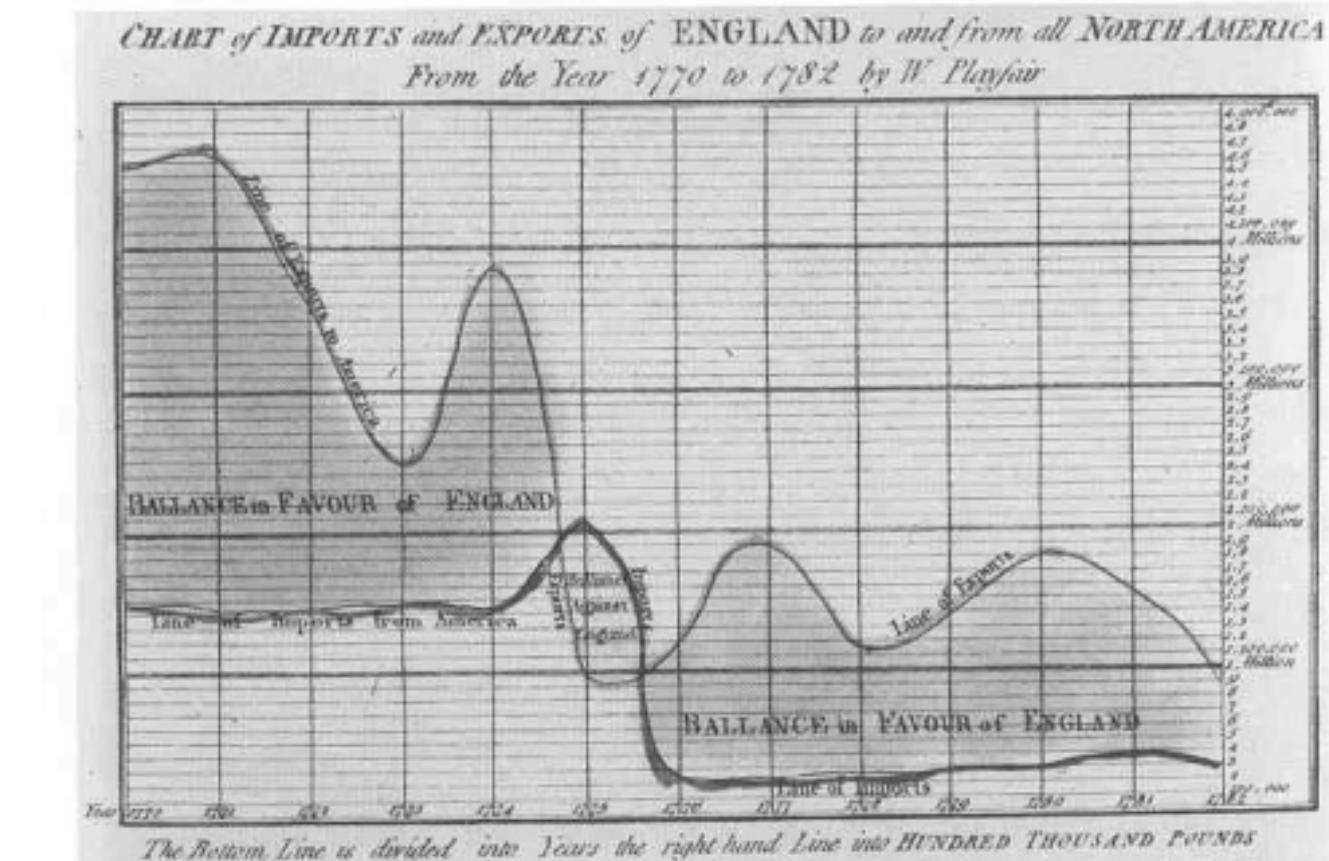
Onde começaram as primeiras visualizações de dados?



Alguns exemplos



Uma das mais antigas - Inclinação das órbitas ao longo do tempo ≈ Século X

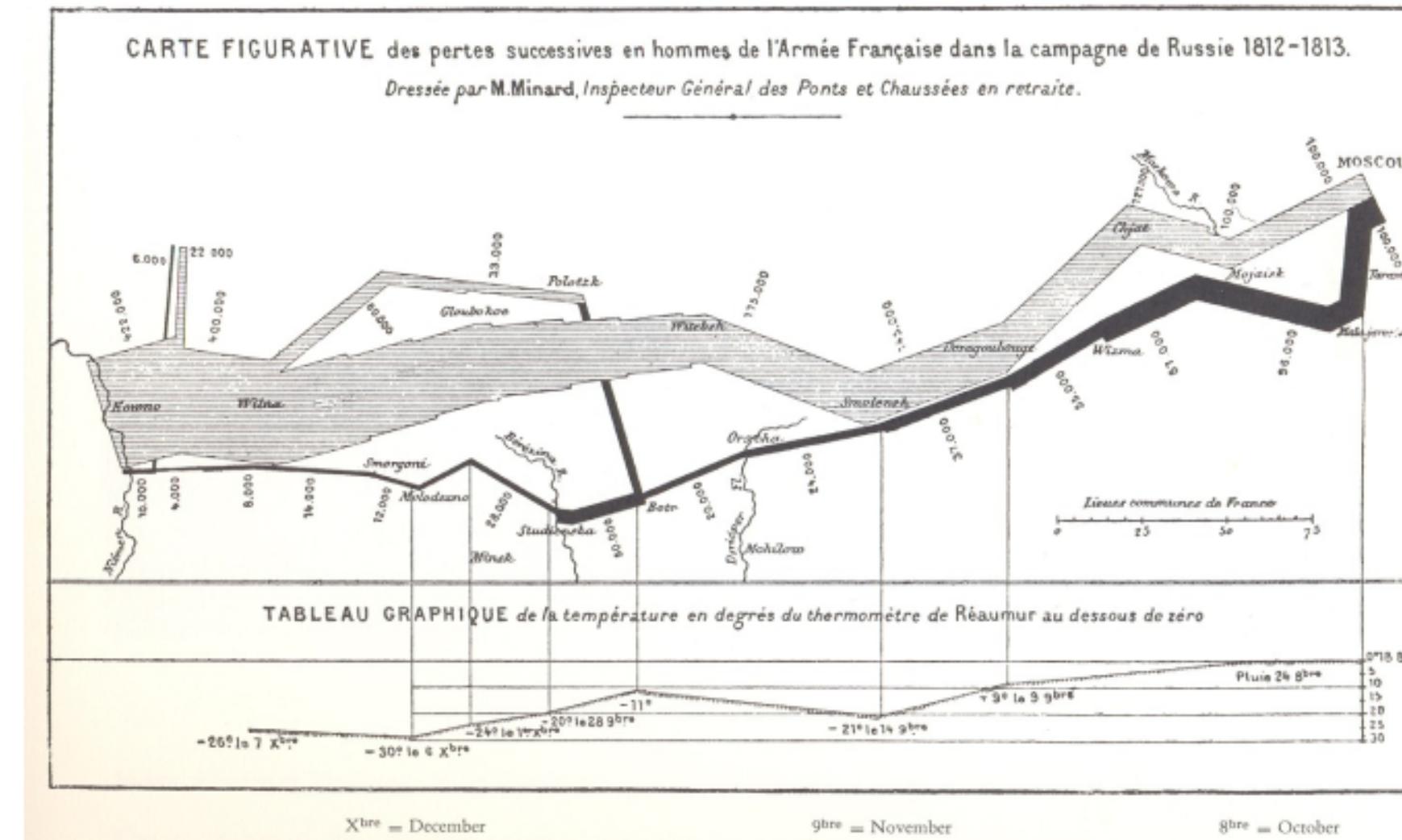


Importações e exportações entre 1770 e 1782

O uso do computador tornou a visualização mais prática!

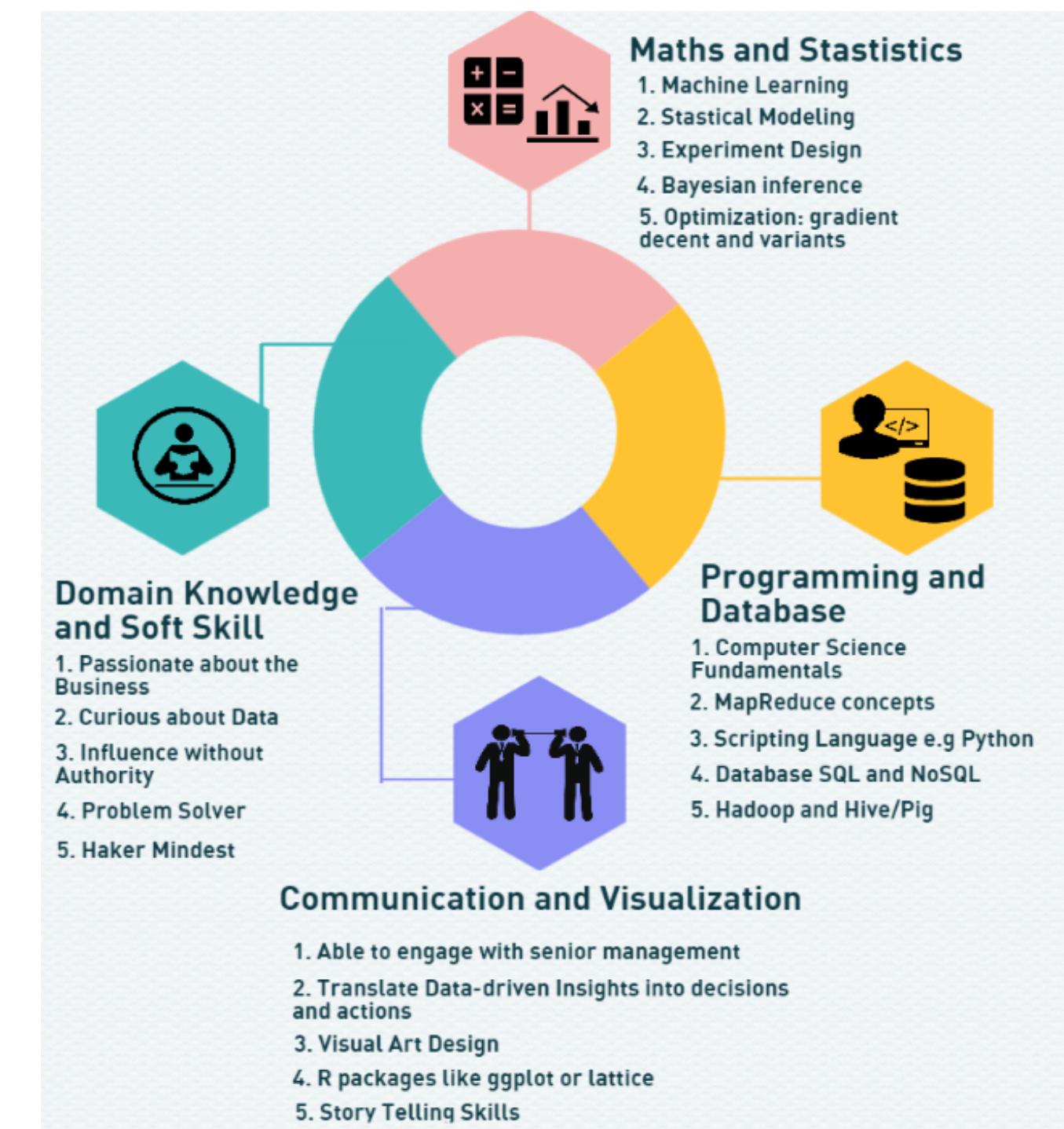
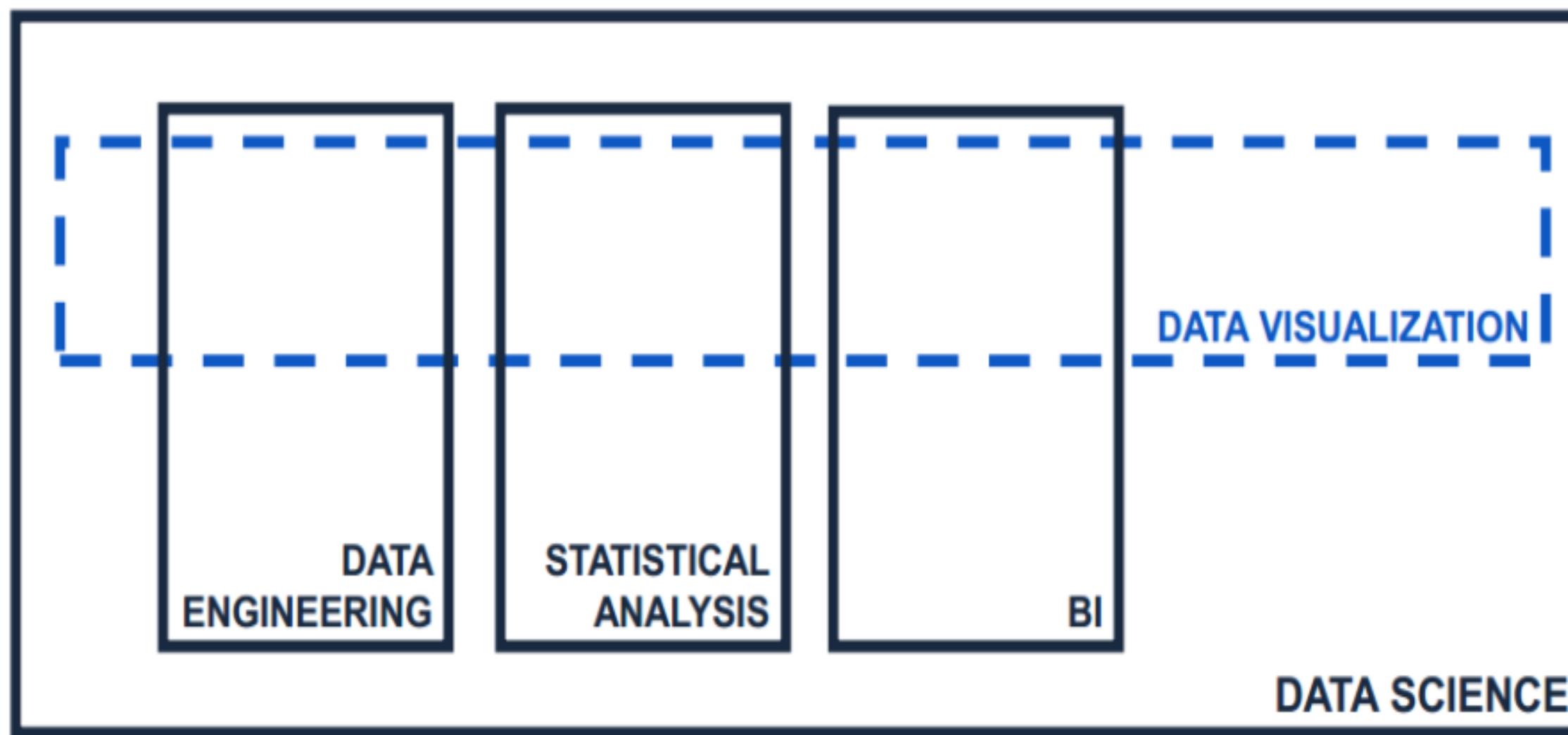
Visualização multidimensional

6 dimensões: localidade (2), nº de homens e sua localização, data e temperatura.



Exército 1861

Quão importante é a visualização de dados no mundo analítico?



02

Tipos de gráficos

Quais os tipos de gráficos que vocês conhecem?



Pontos importantes antes de escolher um gráfico

1. Natureza do problema

- Comunicar
- Explorar
- Confirmar

3. Número de dimensões

- Unidimensional
- Bidimensional
- etc

2. Natureza dos dados

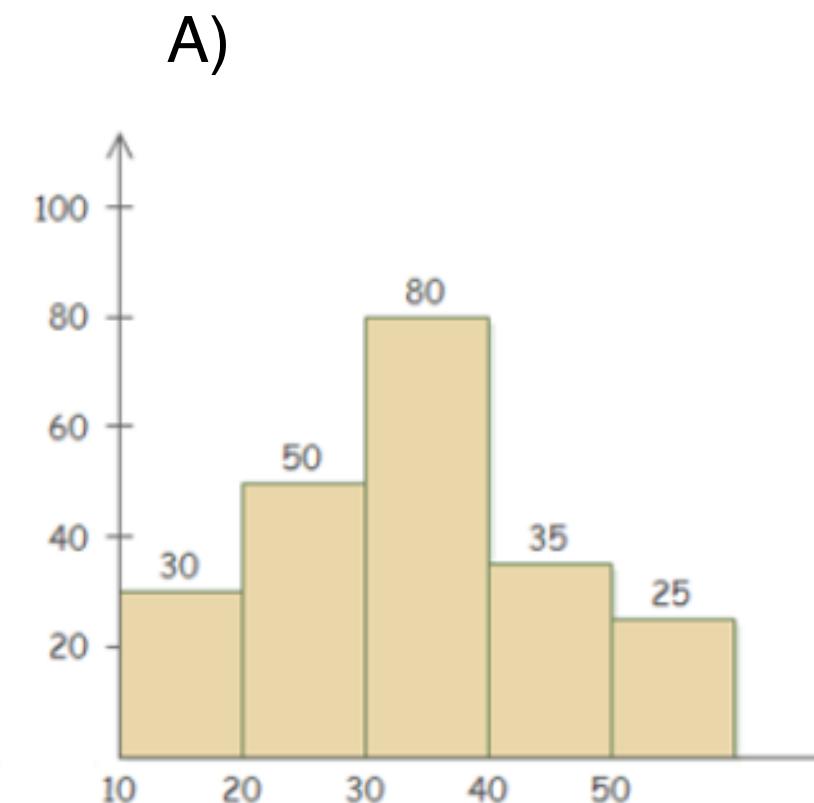
- Categóricos
- Numéricos

4. Estrutura dos dados

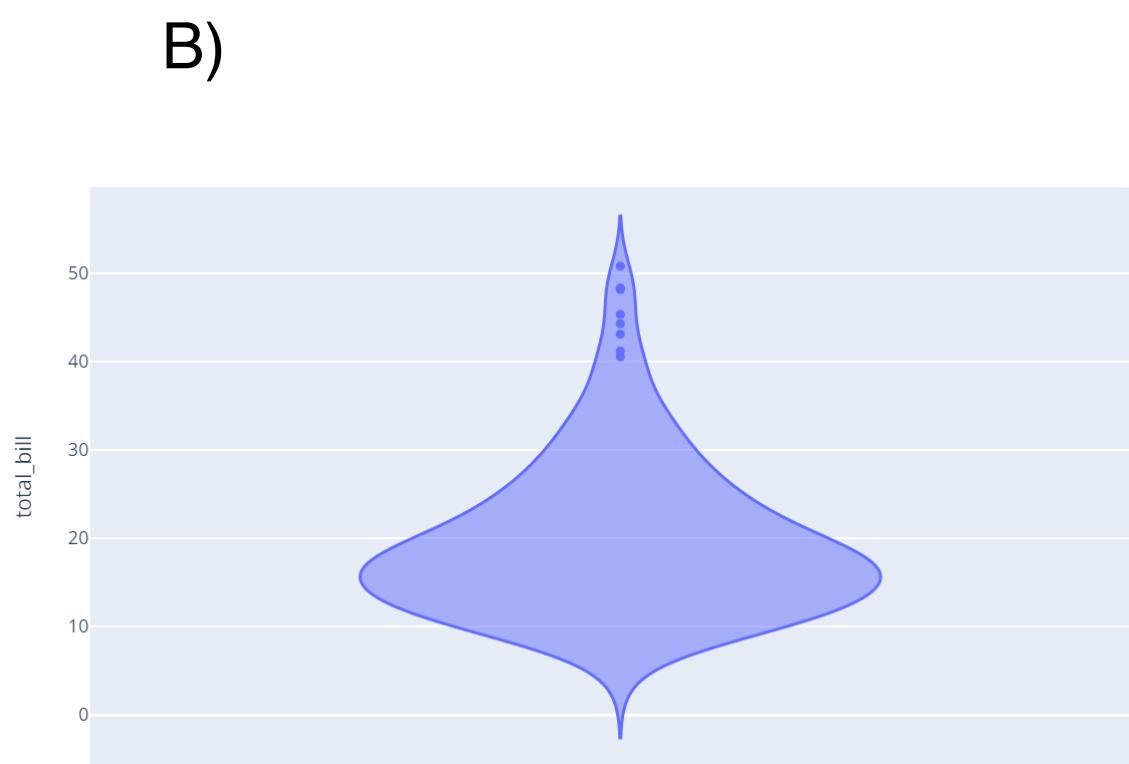
- Linear
- Temporal
- Espacial
- Hierárquico
- Rede

Gráficos unidimensionais mais conhecidos

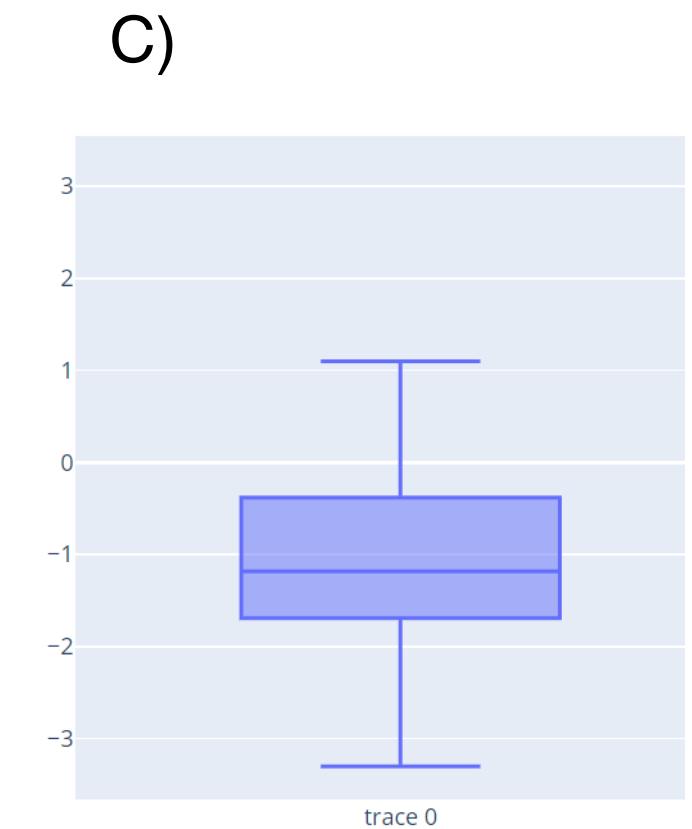
1. Pie chart



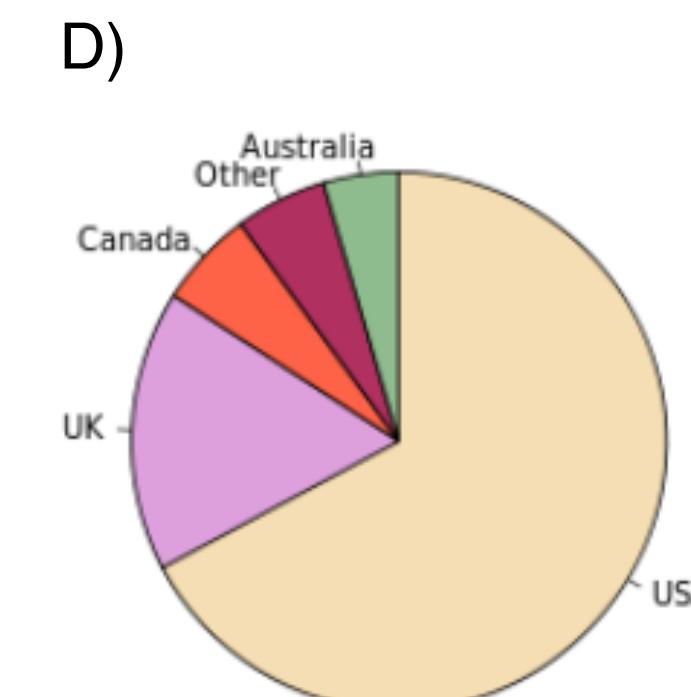
2. Gráfico de barras



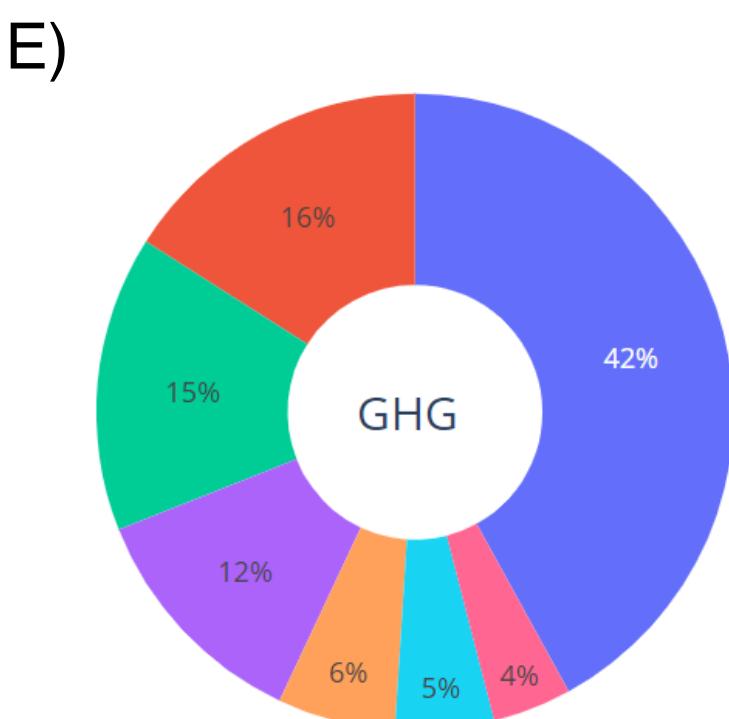
3. Box plot



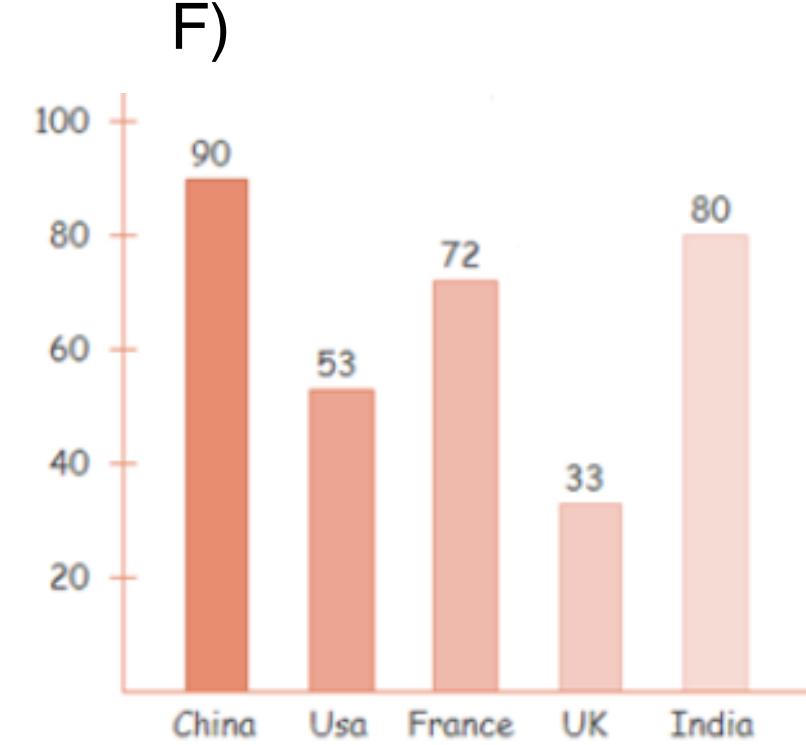
4. Histograma



5. Violin plot

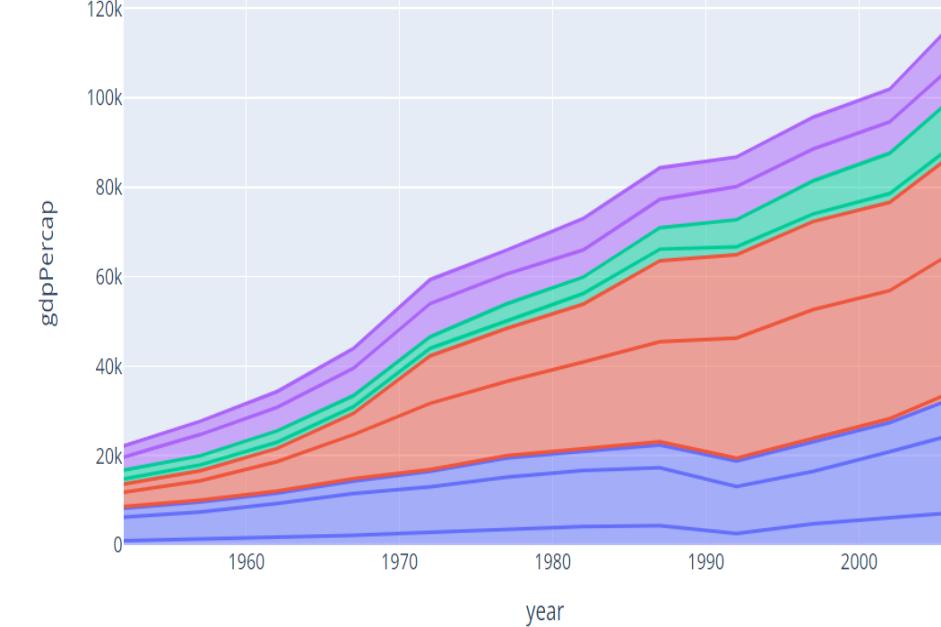


6. Donut plot

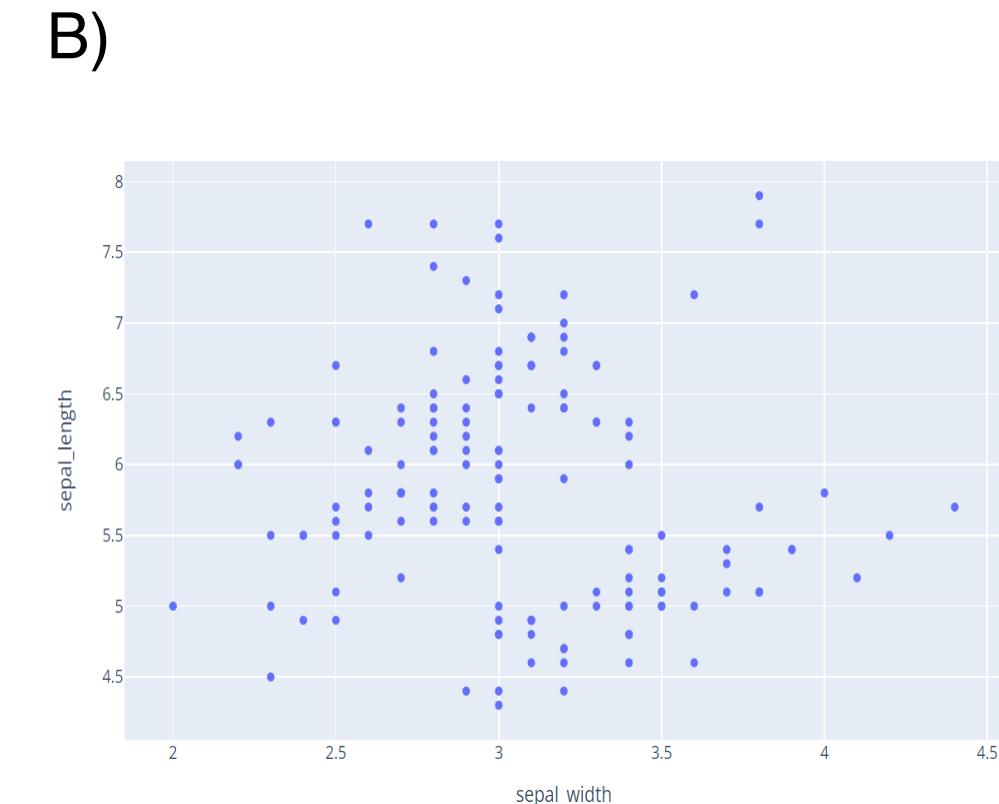


Gráficos multidimensionais mais conhecidos

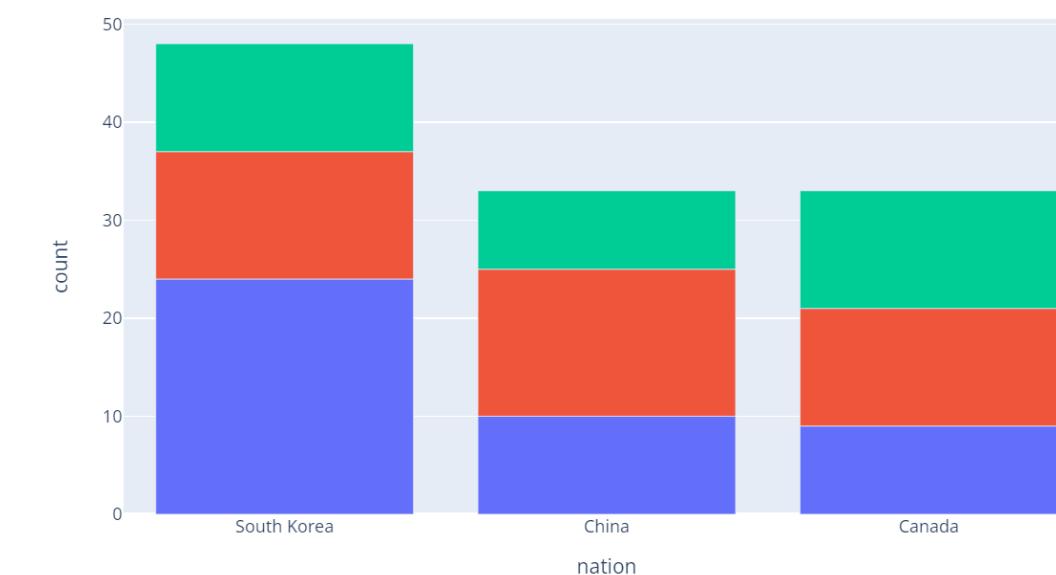
1. Gráfico de pontos



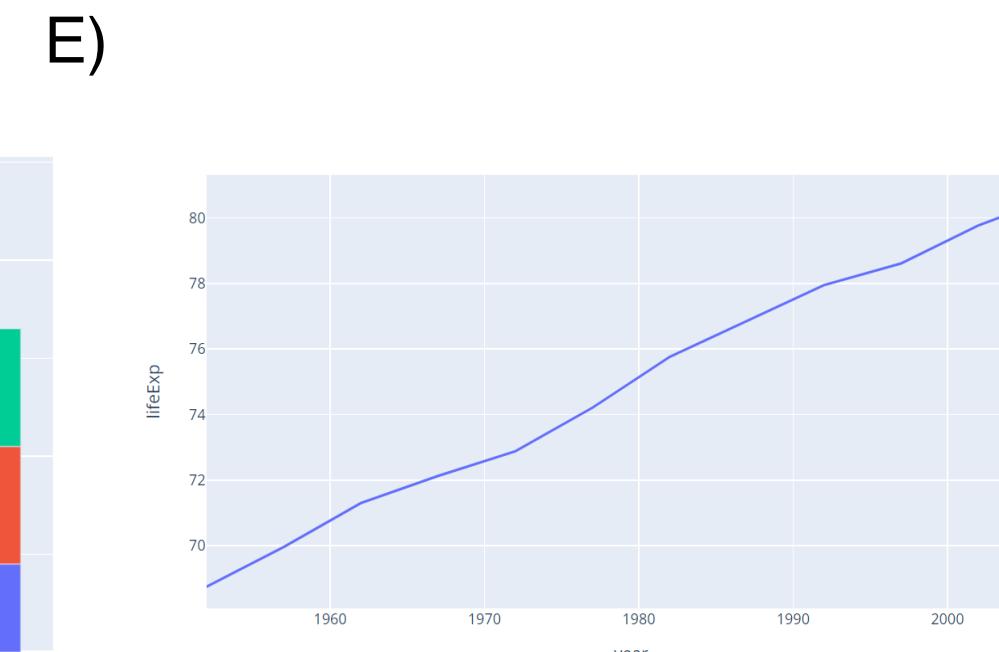
2. Gráfico de linhas



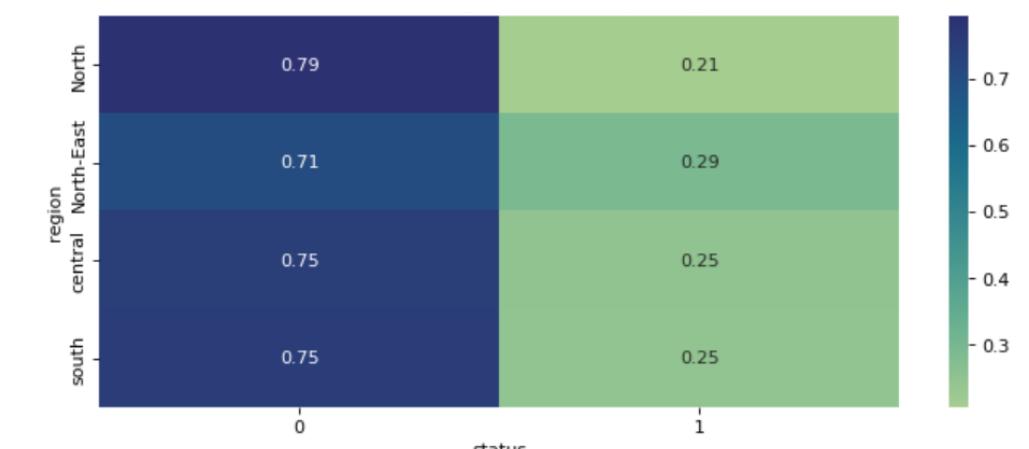
3. Paired bar chart



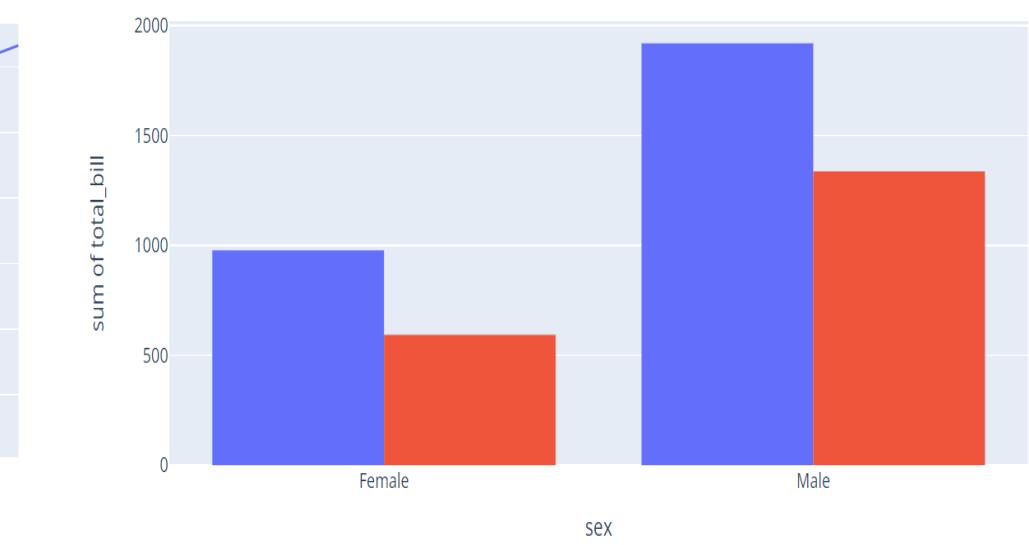
4. Stacked bar chart



5. Heatmap

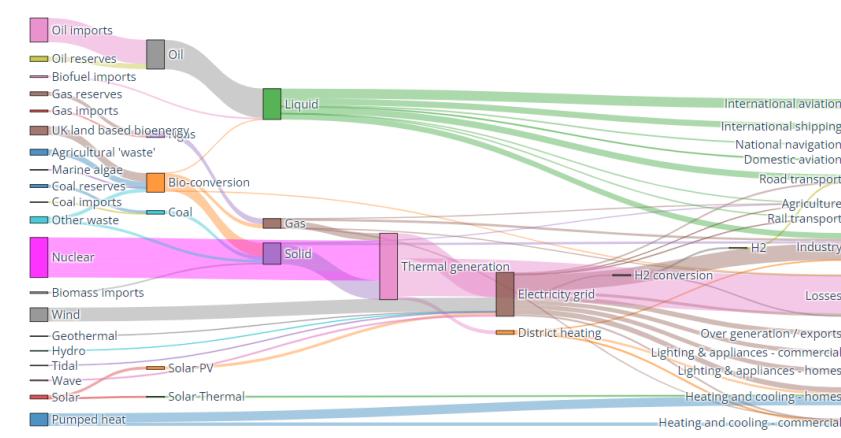


6. Gráfico de área

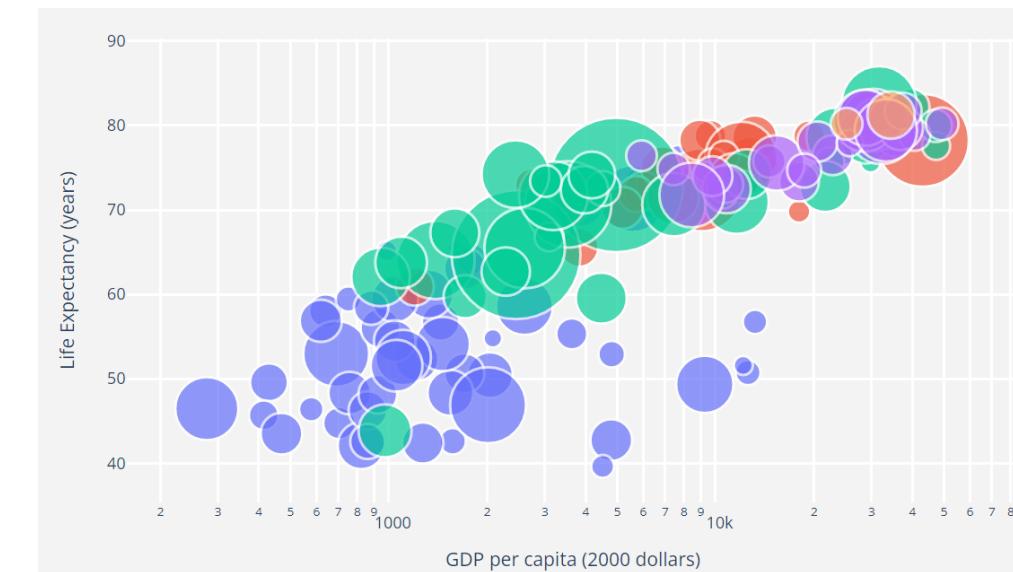


Gráficos multidimensionais menos conhecidos

1. Radar plot

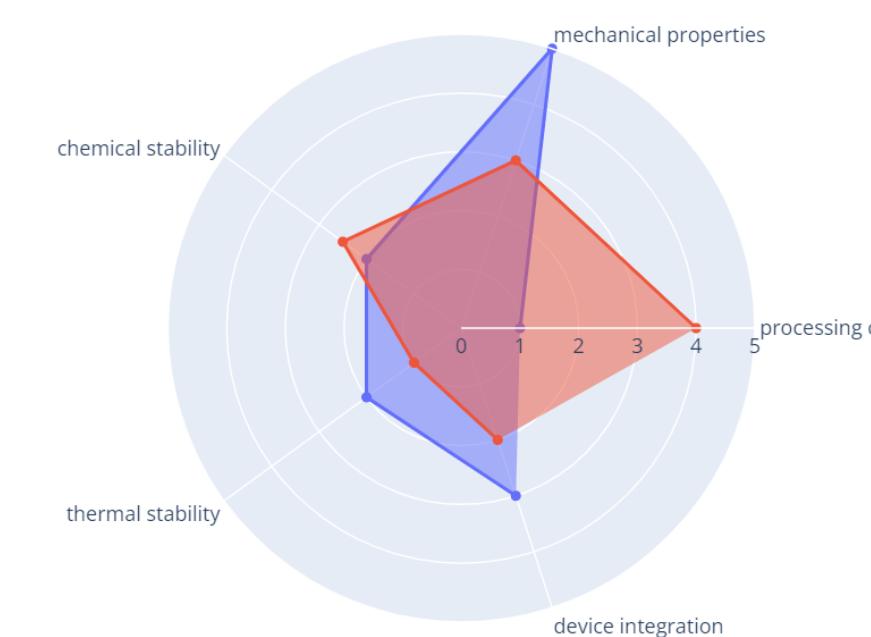


2. Coordinate plots



3. Bubble chart

4. Waterfall

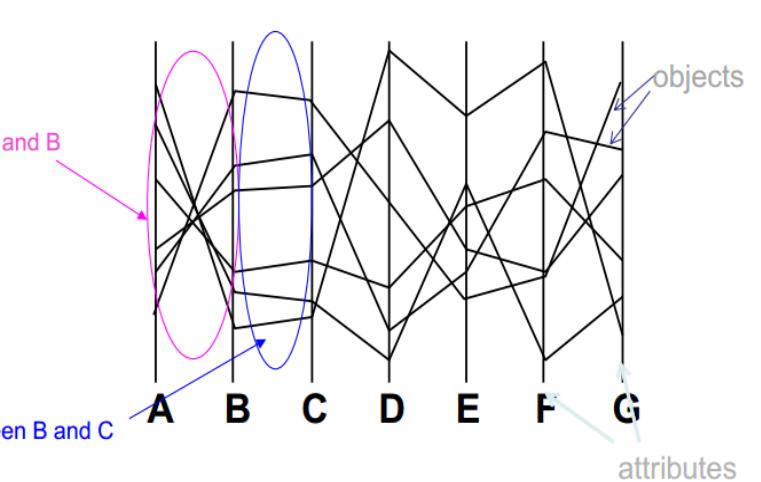


5. Sankey

E)



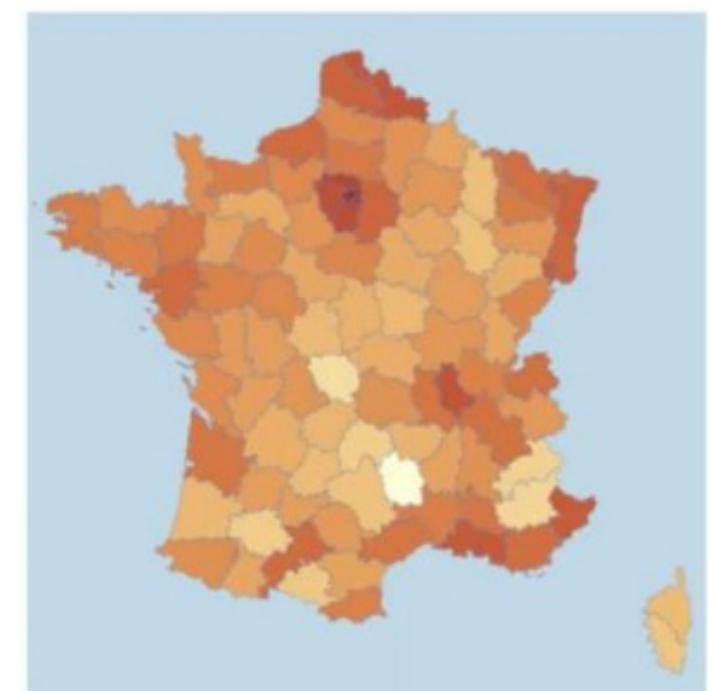
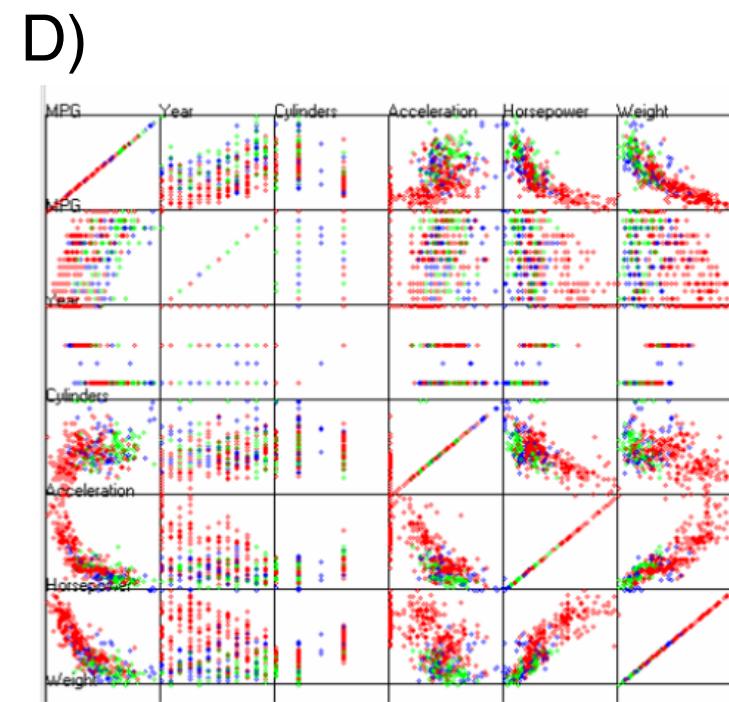
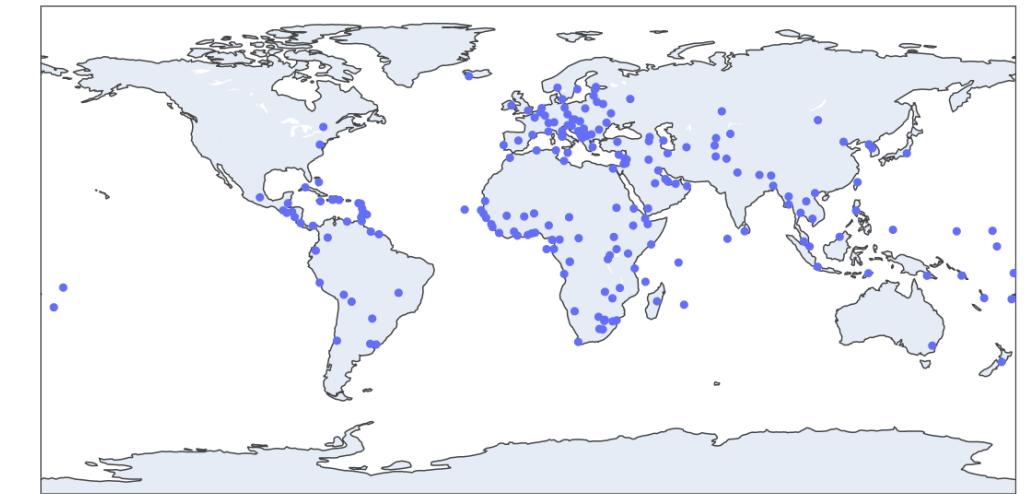
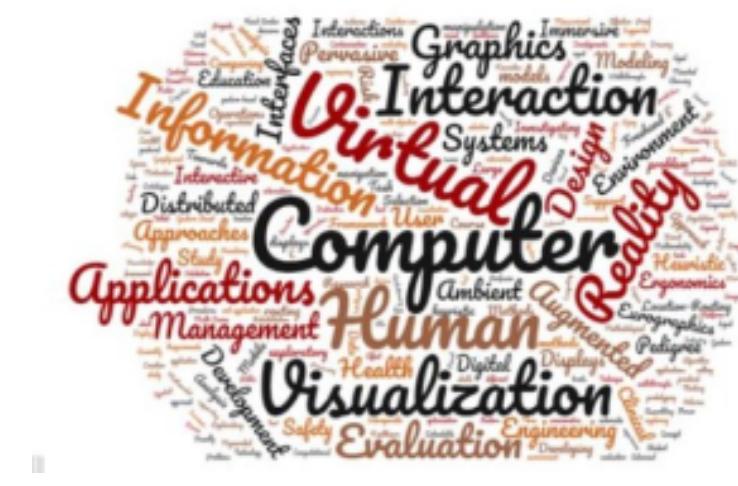
6. Treemap



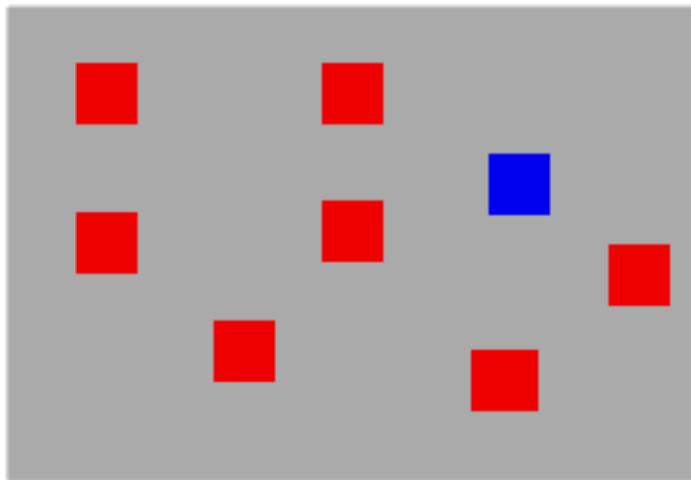
F)

Gráficos multidimensionais menos conhecidos

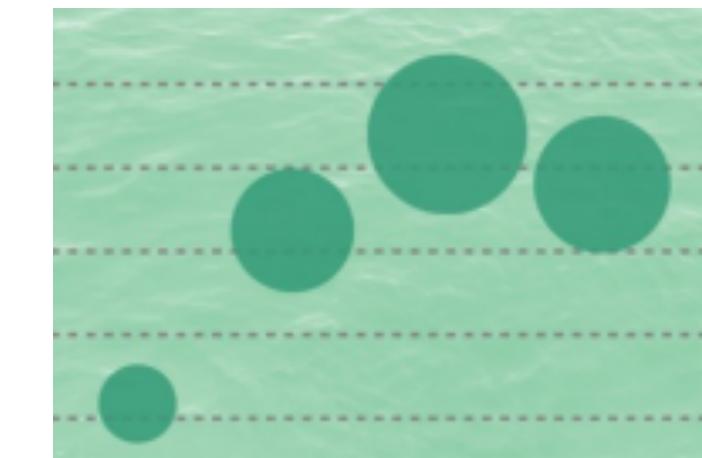
1. Choropleth Map
 2. Word Cloud
 3. Transit map
 4. Matrix scatterplot
 5. Scatterplot map
 6. Lines map



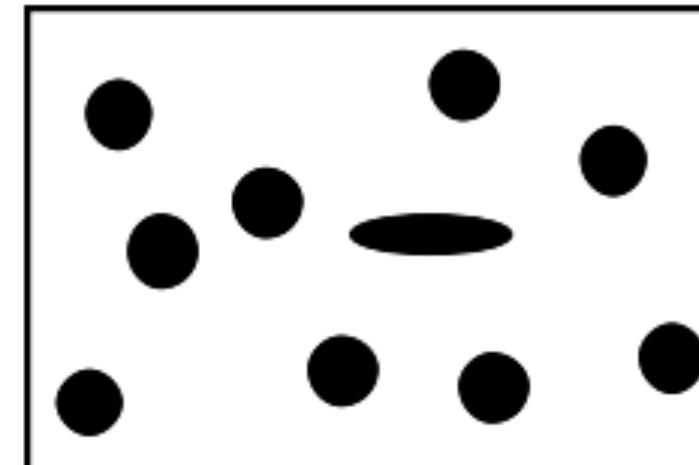
Algumas técnicas para criar mais dimensões



Colour



Size



Shape

03

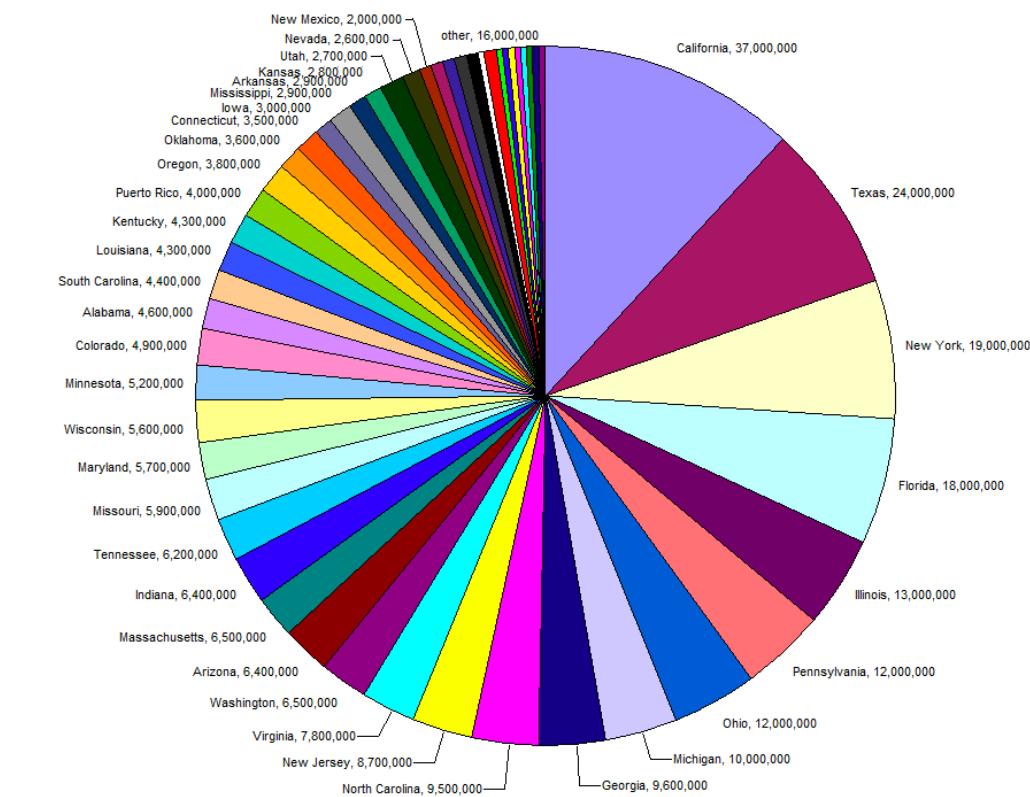
Más prácticas

Algumas sugestões?

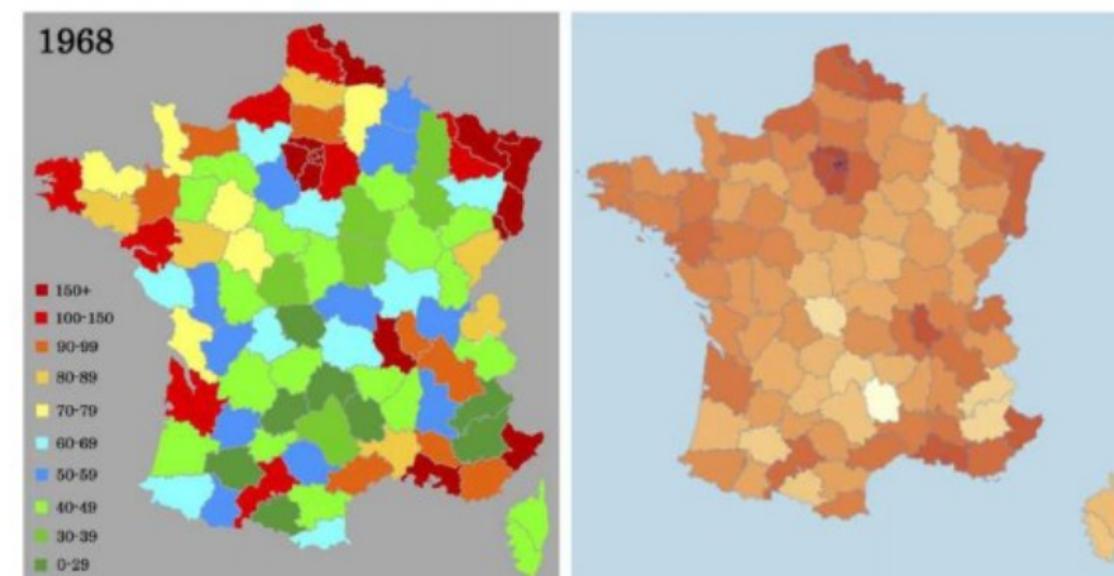


Alguns exemplos de más práticas

1. Temos boa visibilidade do que está a ser apresentado?

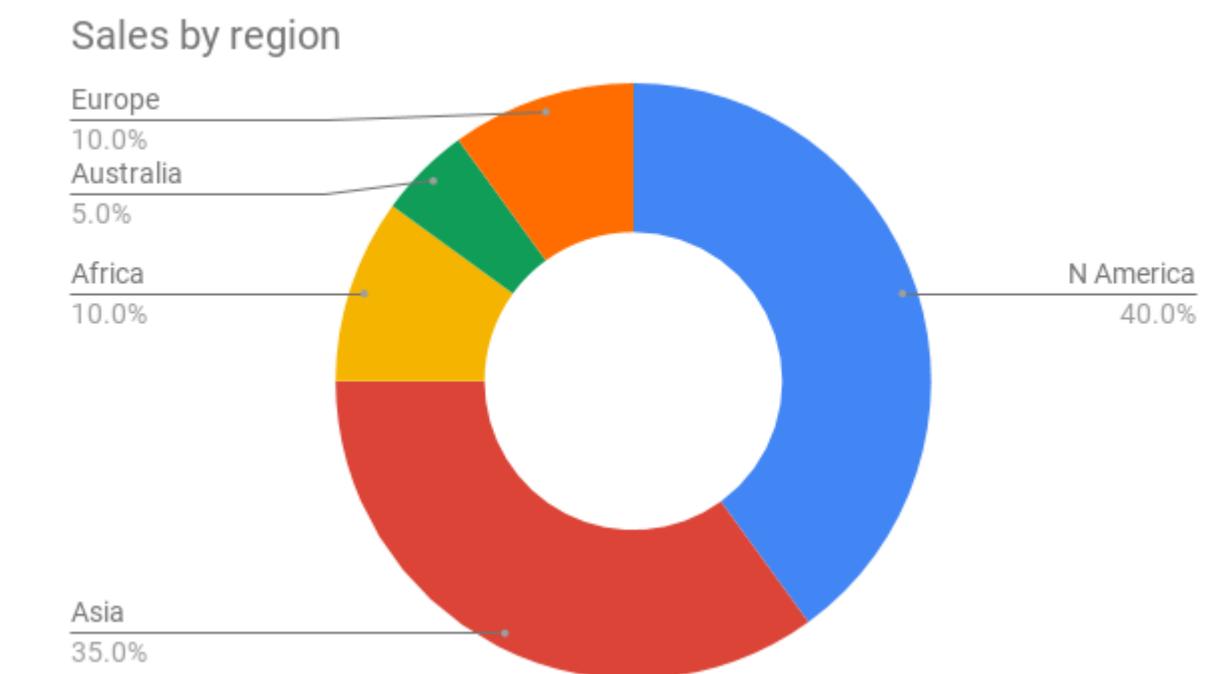


2. Certas cores têm significado predefinidos

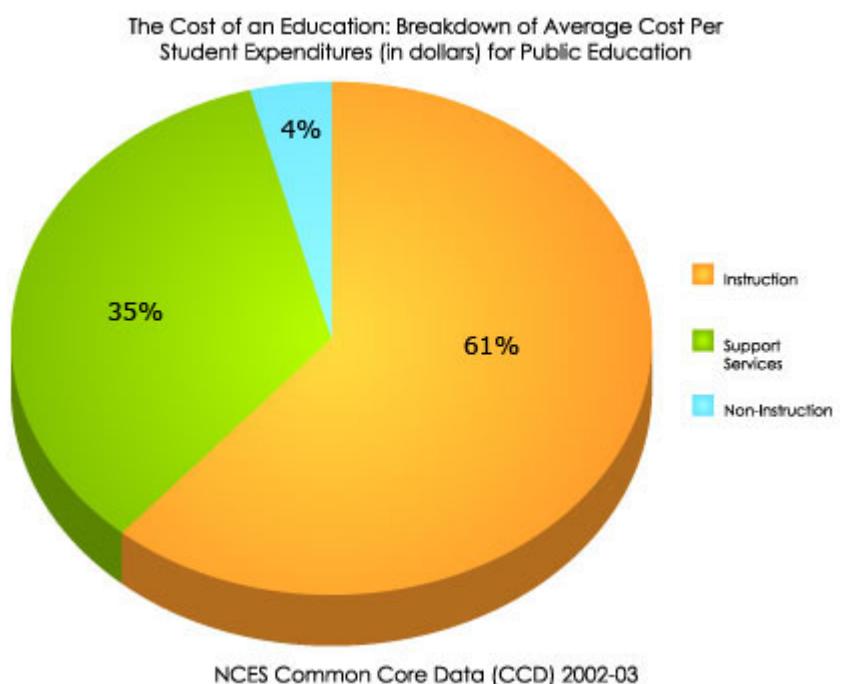


Alguns exemplos de más práticas

3. Não dar o contexto necessário

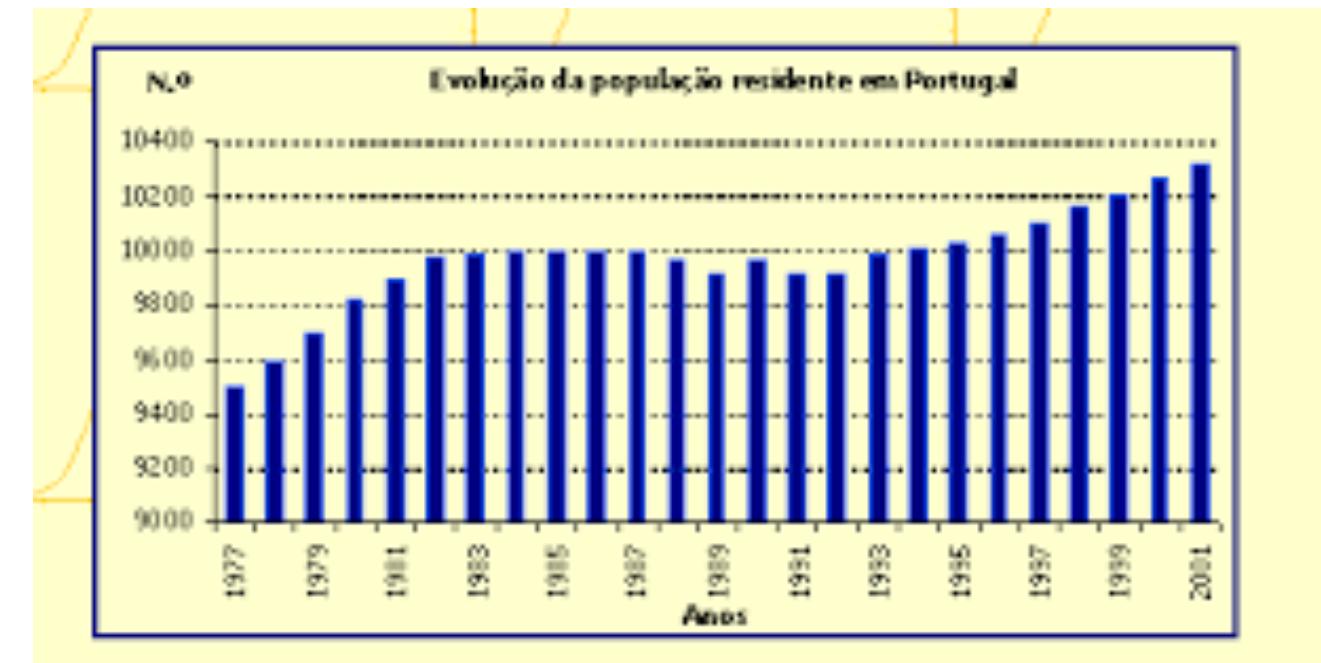


4. Colocar dimensões desnecessárias

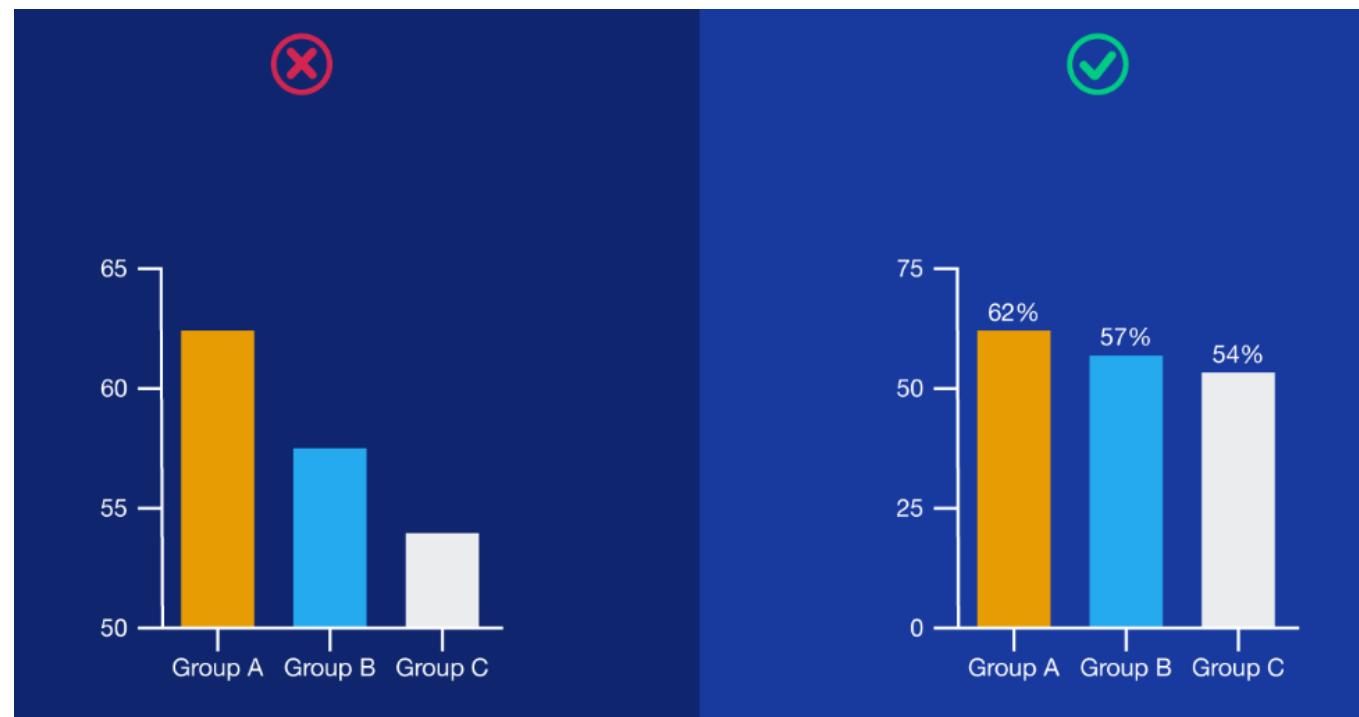


Alguns exemplos de más práticas

5. Usar gráficos de barras para mostrar evolução
(ou usar linhas quando não há uma evolução no eixo x)

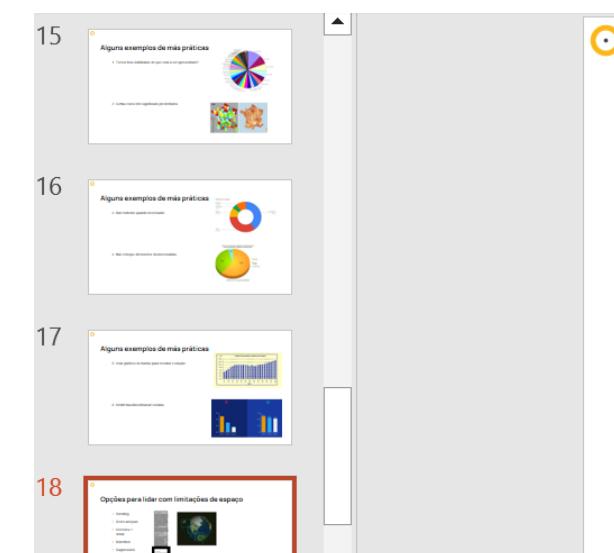
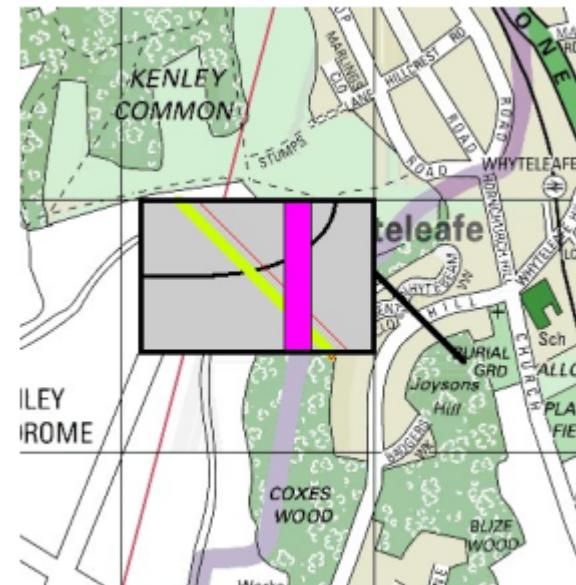
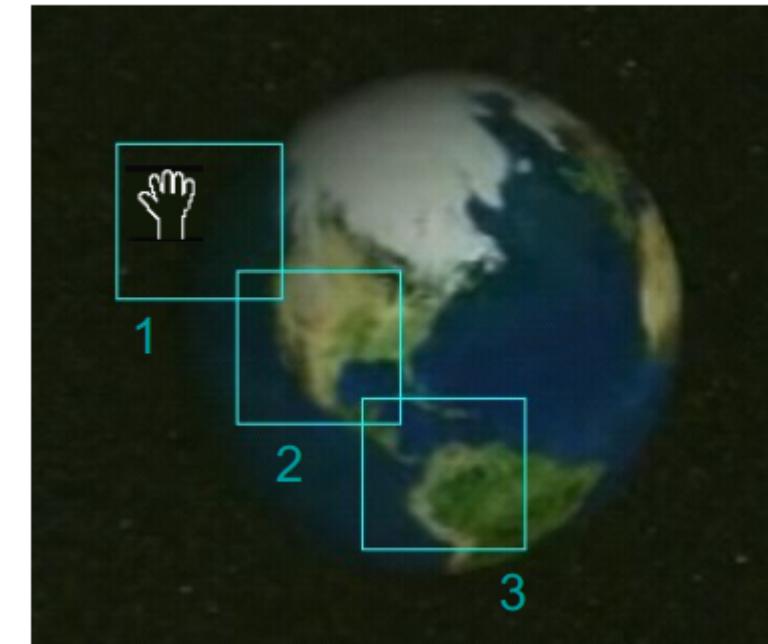
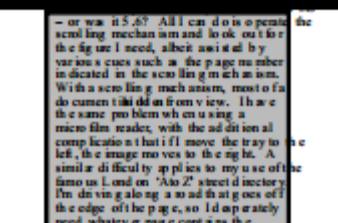
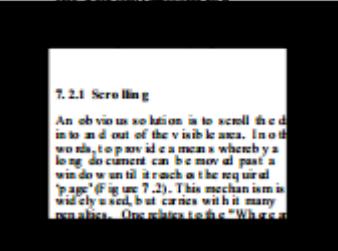
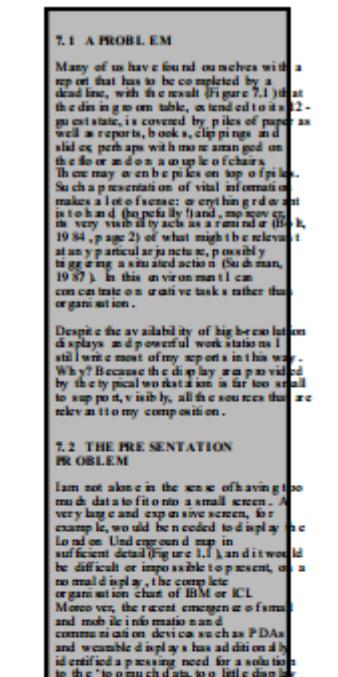


6. Omitir baselines/truncar escalas



Opcões para lidar com limitações de espaço

- Scrolling
 - Zoom and pan
 - Overview + detail
 - Distorção
 - ex: escala logarítmica
 - Supressão
 - ex: filtro



Opções para I

- Scrolling
 - Zoom and pan
 - Overview + detail
 - Distortion
 - Suppression

04

Avaliação

Como podemos avaliar o produto final?



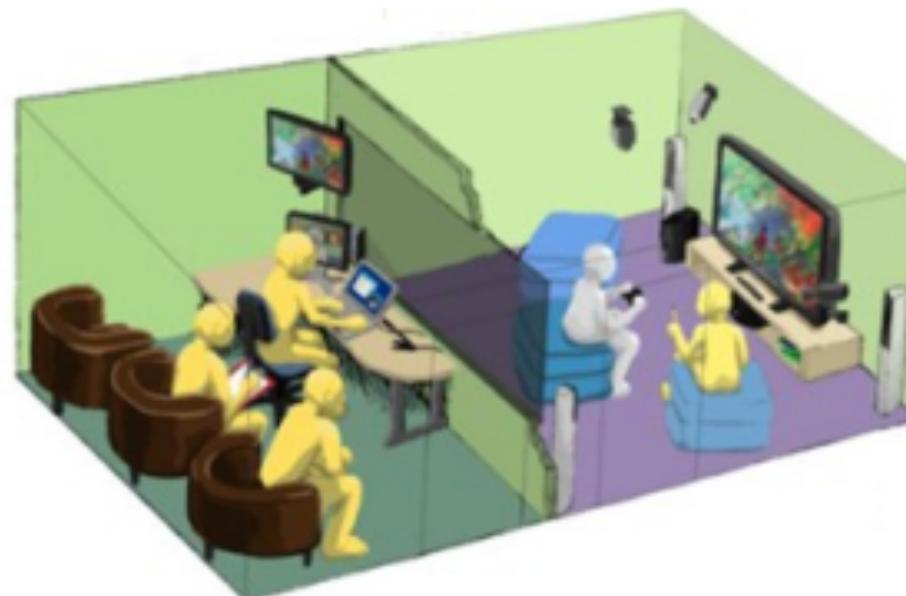
Tipos de métodos

1. Analíticos (não usa utilizadores)

- Heuristic evaluation
- Cognitive Walkthrough
- ...

2. Empíricos (usa utilizadores)

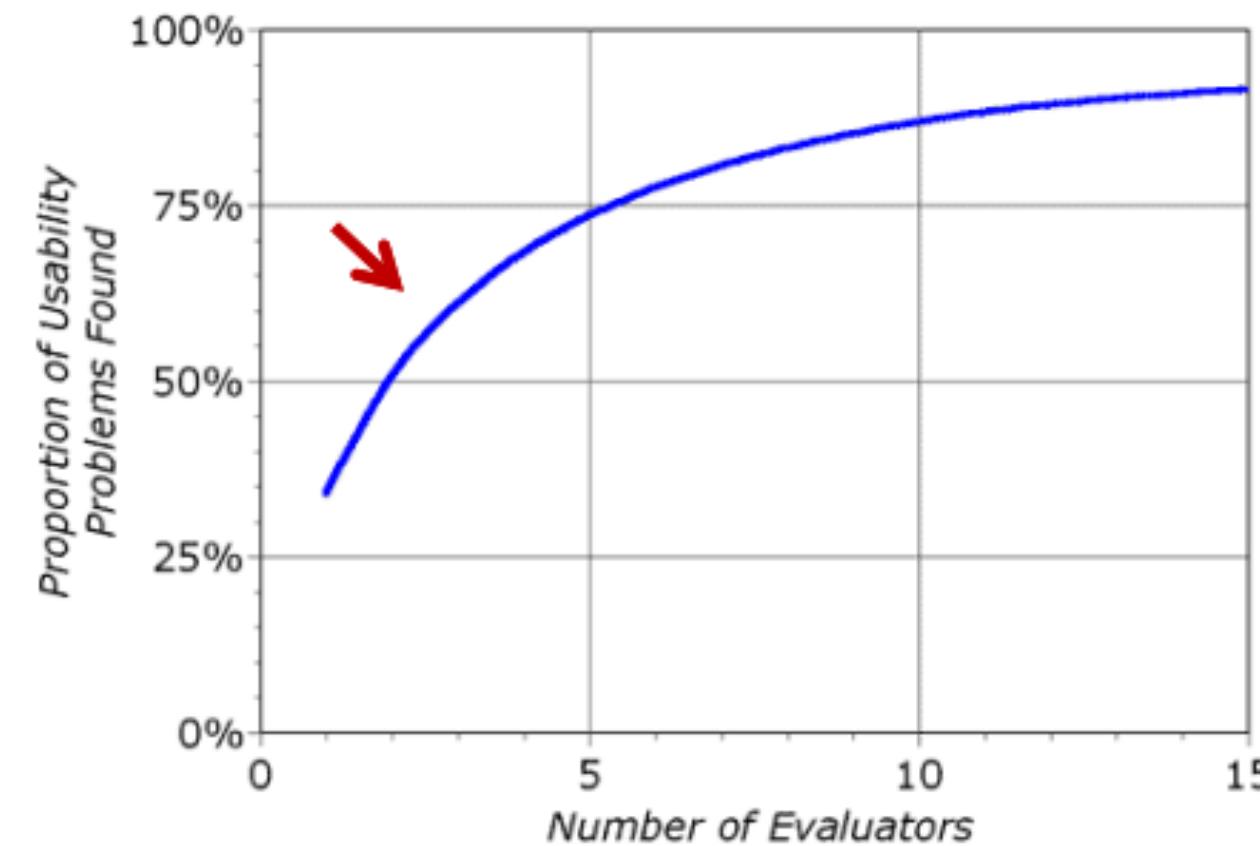
- Observação
- Query
- Experiencias controladas
-



Heuristic evaluation

- 1. Pode ser usada em várias fases do desenvolvimento**
- 2. Permite obter resultados úteis mesmo por analistas pouco experientes**
- 3. Rápido, barato e de fácil avaliação**
- 4. Avaliadores devem trabalhar de forma independente e devem descrever o problema e o seu grau de severidade**
- 5. Só no fim se devem agregar todos os pontos identificados**

Nº de avaliadores



De acordo com o artigo de Nielsen o número ideal será entre 3 a 5 avaliadores.

Para mais informações ver link abaixo.

<https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/>

Grau de severidade

Uma escala comum é a seguinte de 0 a 4:

- 0. Não vejo um problema**
- 1. Precisa ser corrigido se houver tempo extra**
- 2. Correção de baixa prioridade**
- 3. Correção de alta prioridade**
- 4. Correção imperativa (problema muito grave)**

<https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/>

Limitações

- **Vários pontos subjetivos**
- **Envolve várias pessoas**
- **Não é possível identificar todos os problemas**

Vamos agora conhecer a avaliação empírica!

Avaliação empírica

De uma forma geral:

- **Devemos pedir consentimento**
- **Confidencialidade**
- **Garantir segurança**
- **Liberdade (para desistir)**
- **Limitar o stress (estamos a avaliar o produto não o utilizador)**

Observação

Muitas abordagens, desde muito simples a muito complexas.

- **Direta:** observador tira notas
- **Indireta:** existe gravação de áudio/vídeo
- **Think Aloud:** utilizadores têm que ir explicando o que estão a fazer
- **Logging:** a atividade é gravada no sistema
- **Combinações das anteriores, etc**

Query

Principais variantes:

- Questionário
- Entrevista

Notas:

Devem ser preparados cuidadosamente

Dados recolhidos devem ser analisados cuidadosamente

SUS: System Usability Scale

- Criado para avaliar vários produtos ou serviços;
- Um standard na industria (com milhares de referências);
- Escala fácil de utilizar (10 questões com 5 respostas possíveis);
- Bons resultados mesmo com poucos utilizadores.

Questões

- 1. Gostaria de usar este sistema frequentemente?**
- 2. Achou o sistema desnecessariamente complexo?**
- 3. Achou o sistema fácil de usar?**
- 4. Necessitaria de suporte de um técnico para poder usar o sistema?**
- 5. Achou que as várias funções do sistema estavam bem interligadas?**
- 6. Achou que há muita inconsistência no sistema?**
- 7. Achou que a maioria das pessoas aprenderia a usar o sistema rapidamente?**
- 8. Achou o sistema complicado de usar?**
- 9. Sentiu-se confiante a usar o sistema?**
- 10. Precisava aprender coisas antes de usar o sistema?**

Escala

	Strongly disagree		Strongly agree		
	<input type="text"/>				
1. I think that I would like to use this system frequently	<input type="text"/>				
2. I found the system unnecessarily complex	<input type="text"/>				
3. I thought the system was easy to use	<input type="text"/>				
4. I think that I would need the support of a technical person to be able to use this system	<input type="text"/>				

...

Let $R(n)$ be the answer to Question n:

$$SUS = \left(\sum_{n=1}^5 R(n) - 1 + 5 - R(n*2) \right) * 2.5$$

0...100; SUS > 68 would be considered above average

05

Outras ferramentas

Para além do Python que outras ferramentas conhecem?



Mais conhecidas



Microsoft

The Tableau logo is displayed in a stylized font where each letter is composed of a different colored plus sign (+). The colors include orange, red, blue, and teal. The letters are arranged in a staggered, overlapping fashion.

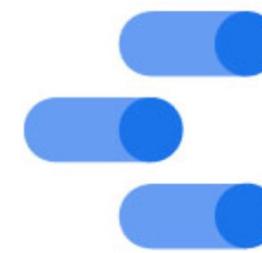


QlikView

Outras ferramentas



R software – ggplot2



Google
Data Studio



D3.js

<https://www.kdnuggets.com/2018/02/comparative-analysis-top-6-bi-data-visualization-tools-2018.html>

06

Atividade de grupo



Atividade

Vamos dividir a sala em diferentes grupos.

Cada grupo deve:

1. Selecionar uma visualização em: [https://www.tibco.com/products/tibco-spotfire/
learn/demos?Section=Interactive](https://www.tibco.com/products/tibco-spotfire/learn/demos?Section=Interactive)
2. Identificar problemas de usabilidade ou más práticas
3. Apresentar à turma os resultados (se não tivermos tempo, podem enviar-me por
email)

Terão 5 min para discussão no grupo e 2 para apresentação à turma.

**Muito obrigado pela
sua atenção.**

Fábio Ferreira

Endereço de e-mail:

fabio7jnferreira@gmail.com

