

ADMINISTRAÇÃO

IBM0112 DATA MINING

Visualização de dados

Cassius Figueiredo

Princípios de visualização de dados

Como transmitimos informações sobre dados?

- A forma mais tradicional e acessível de transmitir informações sobre dados é por meio de visualizações.
- A forma correta de transmitir dados é aquela que cumpre seus objetivos como gerador da informação e faz com que seu público alvo absorva a informação da forma desejada.

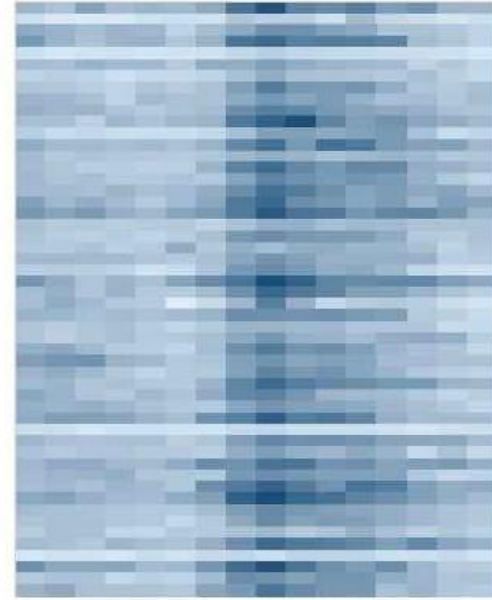
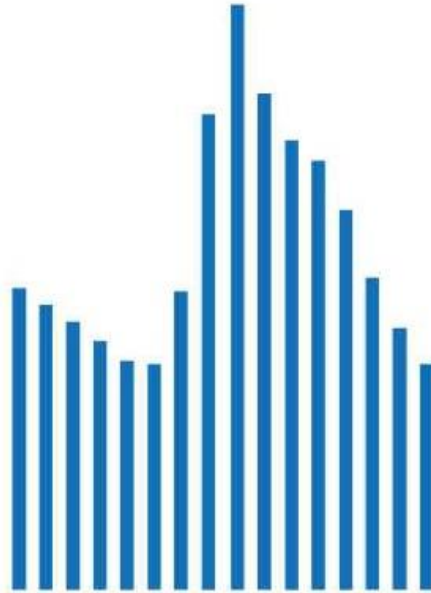
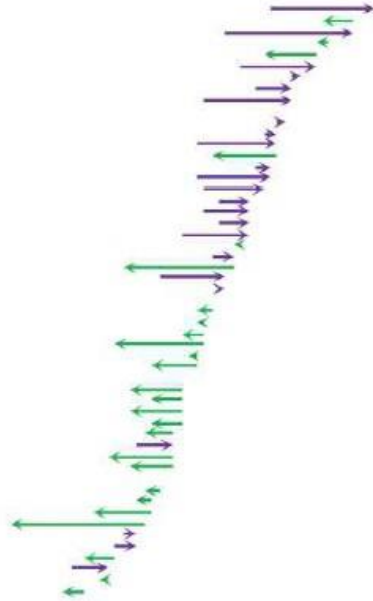
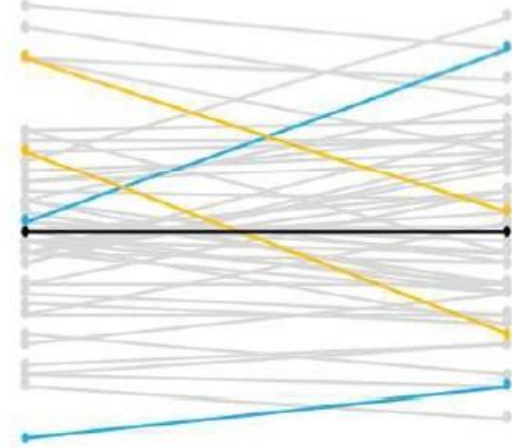
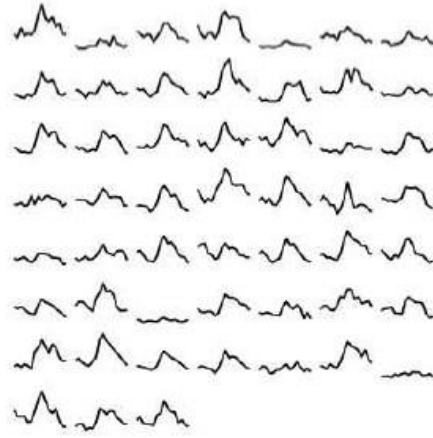
Fatores importantes

- Público-alvo;
- Quantidade de dados;
- Característica dos dados;
- Experiência do cientista de dados;
- Disponibilidade técnica;
- Meio de apresentação, etc.

Exemplo

- Vamos ver formas diferentes de apresentar a informação sobre a parte da população que ganha até um salário mínimo nos E.U.A., por estado.

Exemplo



A importância da visualização de dados, o Quarteto de Anscombe

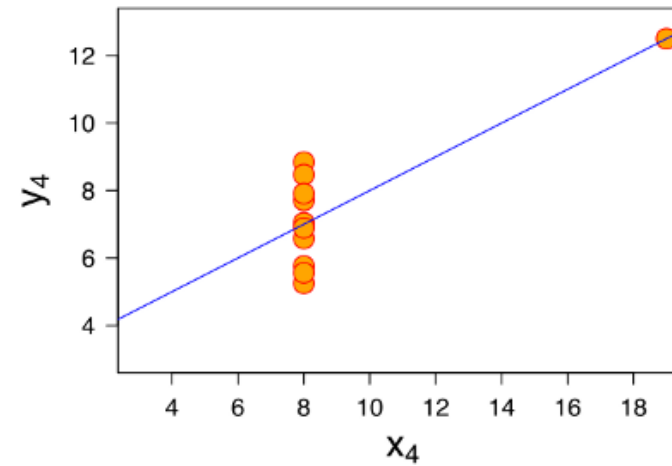
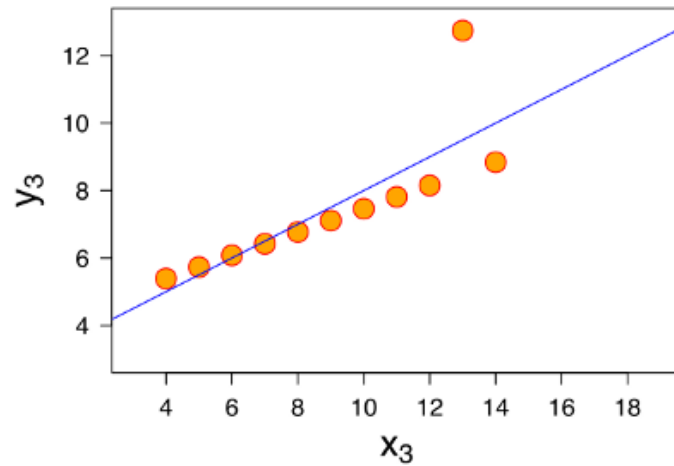
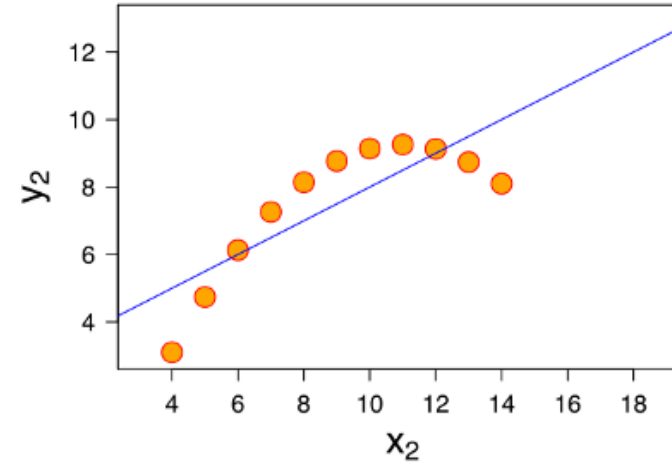
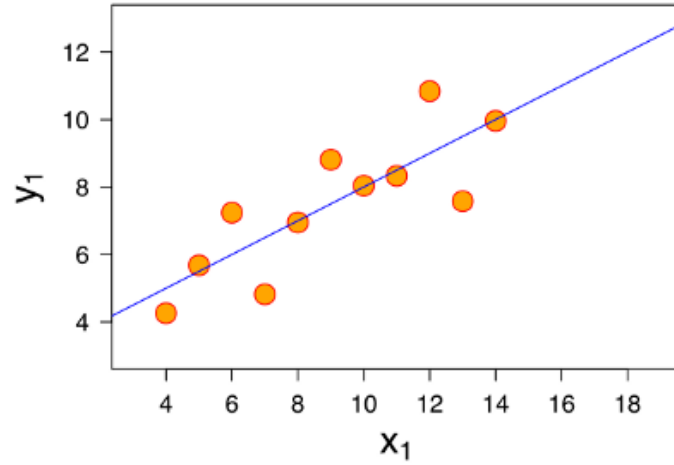
O Quarteto de Anscombe

- Em 1973 o estatístico Francis Anscombe realizou um experimento para demonstrar a importância da visualização no processo de análise de dados.
- Este experimento ficou conhecido por “O Quarteto de Anscombe”.

Dados do experimento

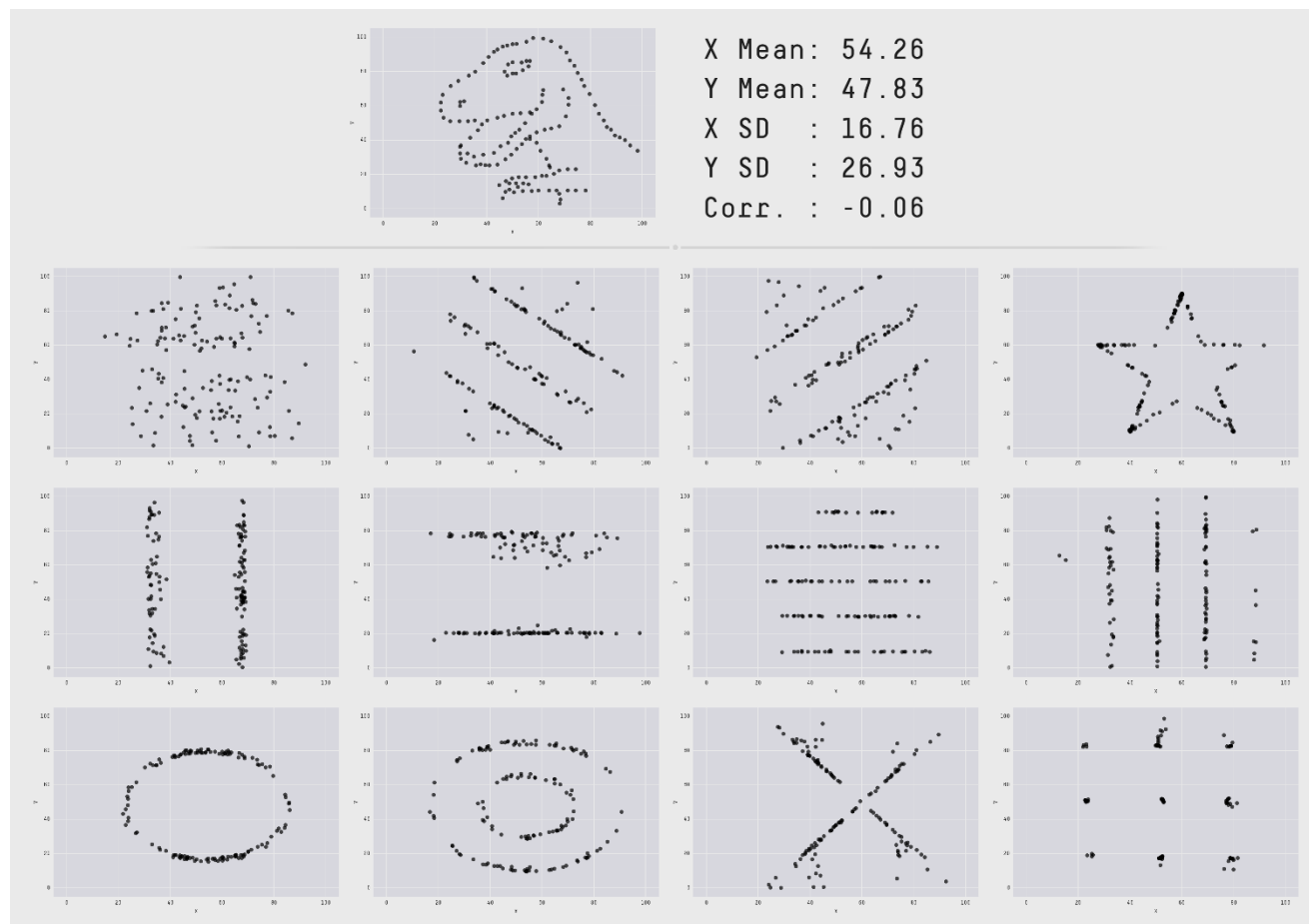
	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10	8,04	10	9,14	10	7,46	8	6,58
	8	6,95	8	8,14	8	6,77	8	5,76
	13	7,58	13	8,74	13	12,74	8	7,71
	9	8,81	9	8,77	9	7,11	8	8,84
	11	8,33	11	9,26	11	7,81	8	8,47
	14	9,96	14	8,1	14	8,84	8	7,04
	6	7,24	6	6,13	6	6,08	8	5,25
	4	4,26	4	3,1	4	5,39	19	12,5
	12	10,84	12	9,13	12	8,15	8	5,56
	7	4,82	7	7,26	7	6,42	8	7,91
	5	5,68	5	4,74	5	5,73	8	6,89
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

Dados do experimento

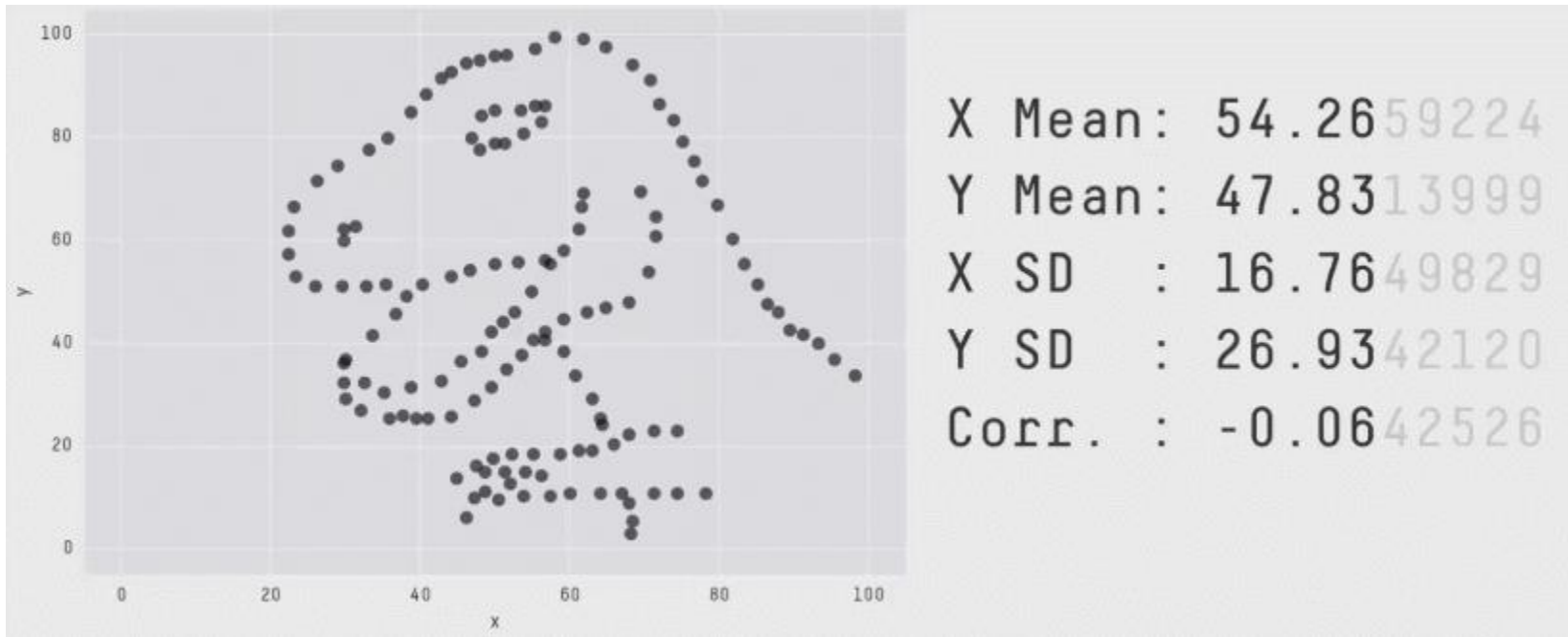


The Datasaurus Dozen

- Em 2016, Alberto Cairo criou um conjunto de dados que ele chamou de Datasaurus e usou-o para ilustrar a afirmação "never trust summary statistics alone; always visualize your data".



The Datasaurus Dozen



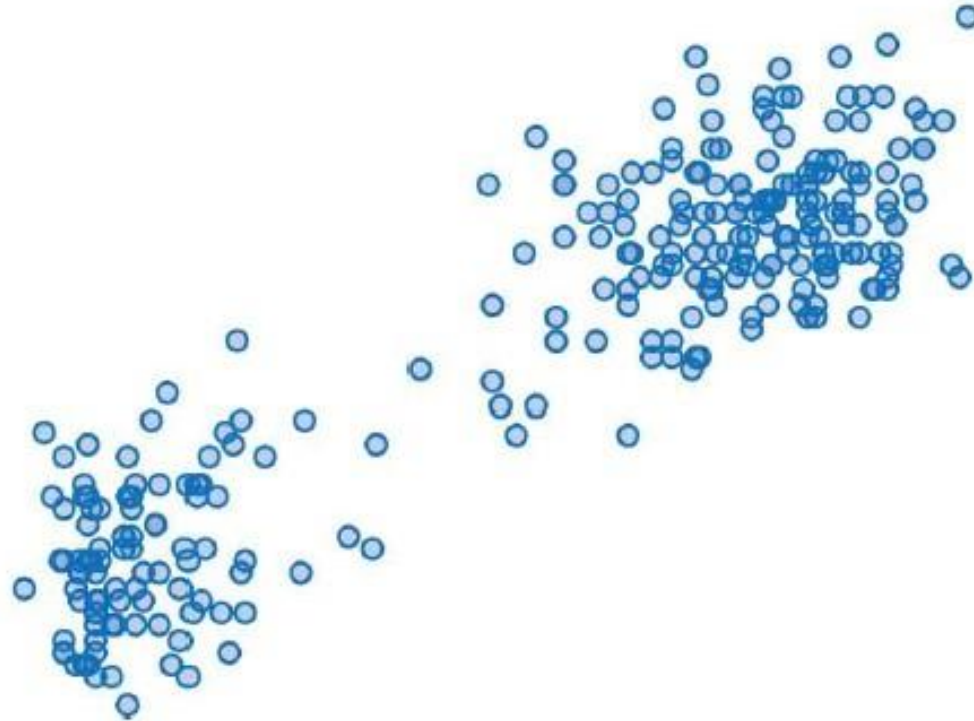
Princípios Gestalt da Percepção Visual

Os Princípios Gestalt

- Como percebemos a informação e como podemos utilizar esta informação para gerar visualizações mais efetivas?
- A Teoria Gestalt foi desenvolvida no início do século 20 por psicólogos alemães e nos mostra como organizamos elementos visuais em grupos.
- Utilizada em áreas como teoria da informação, ciência da visão e neurociência cognitiva.
- Este conhecimento nos ajuda a entender como as pessoas entendem as visualizações.

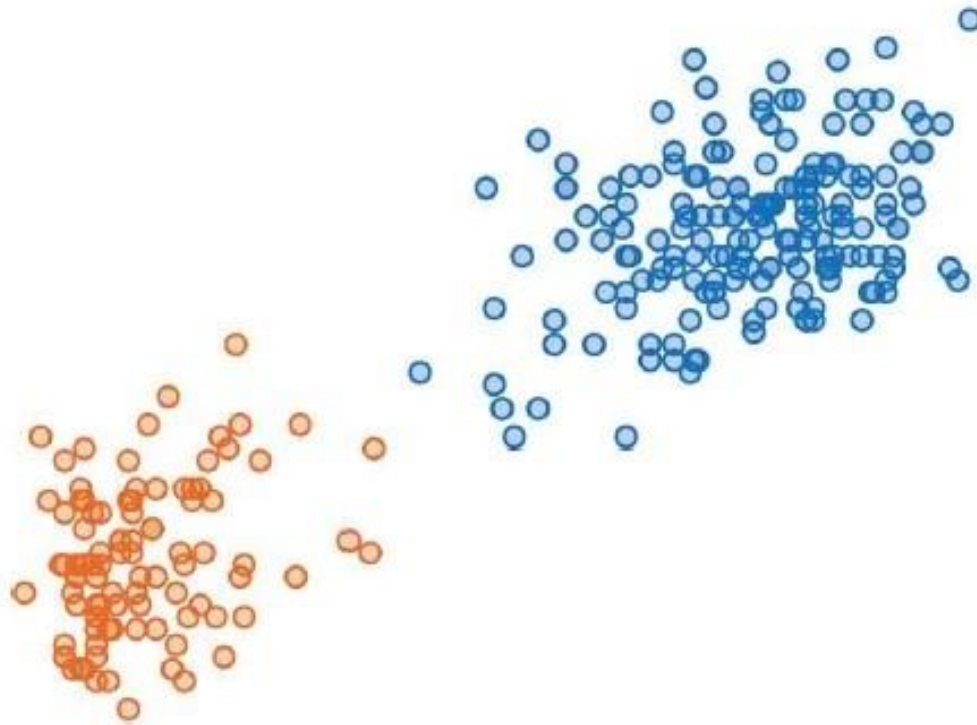
1. PROXIMIDADE

- Temos a percepção que elementos próximos formam um grupo.



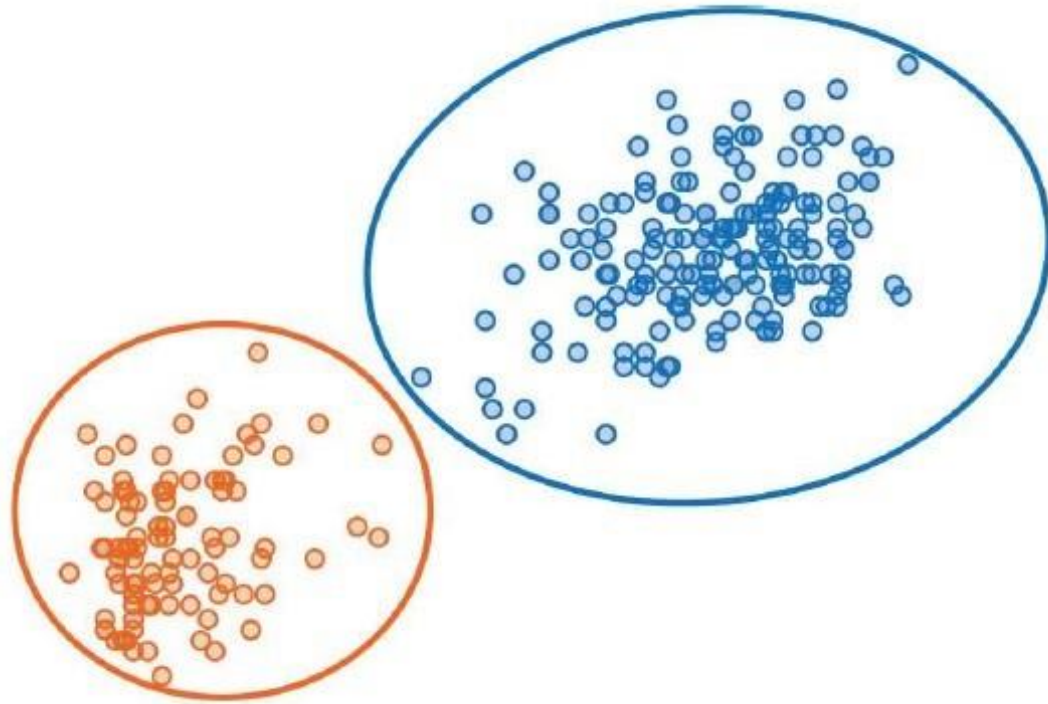
2. SIMILARIDADE

- Agrupamos objetos que possuam a mesma cor, formato ou direção.



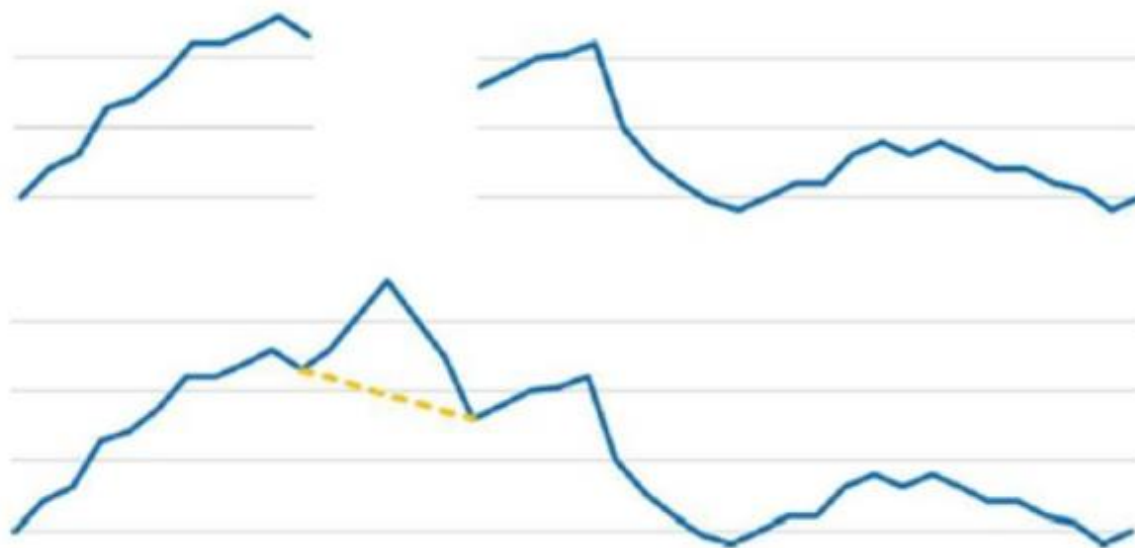
3. FRONTEIRAS

- Objetos cercados são percebidos como grupos.



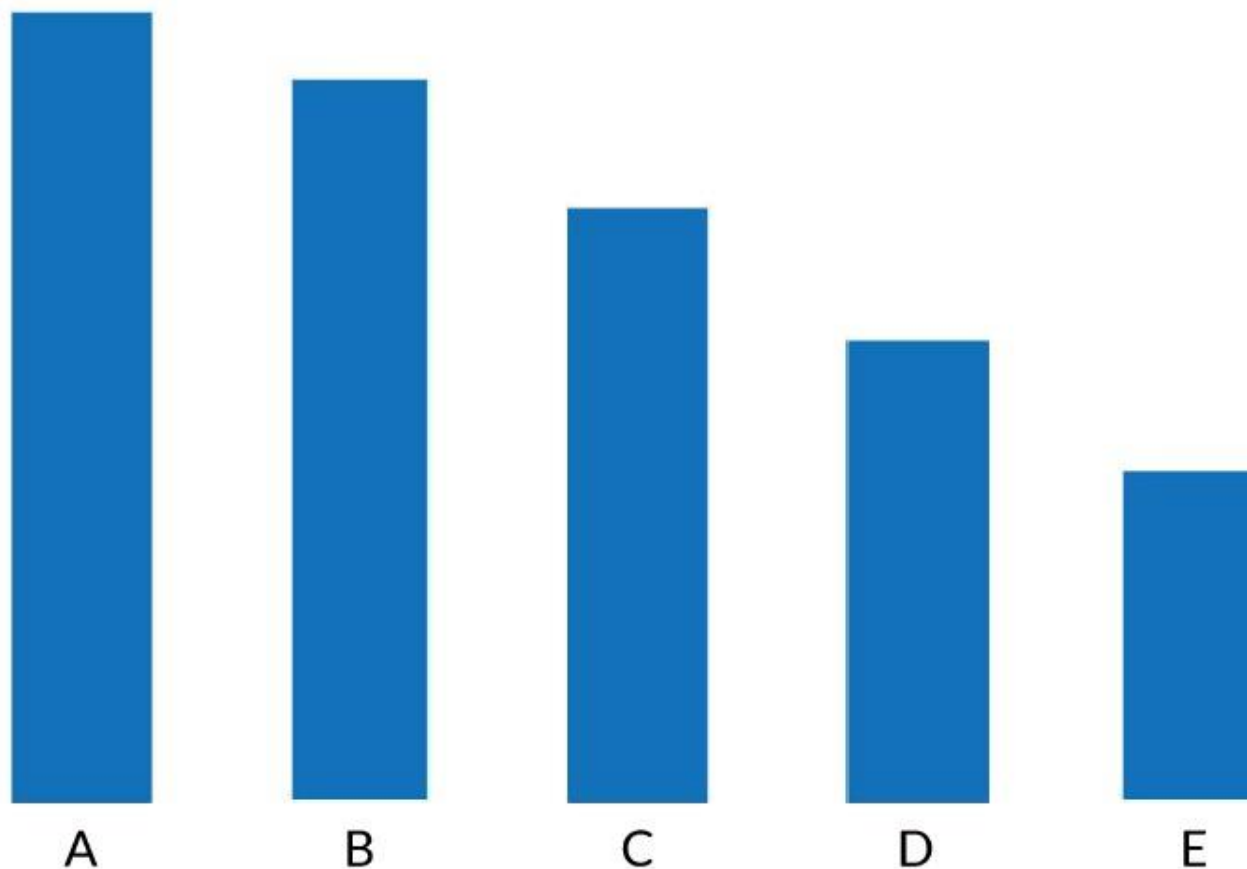
4. FECHAMENTO

- Em uma visualização com espaços sem dados, tendemos a completá-los da forma mais direta e simples possível.



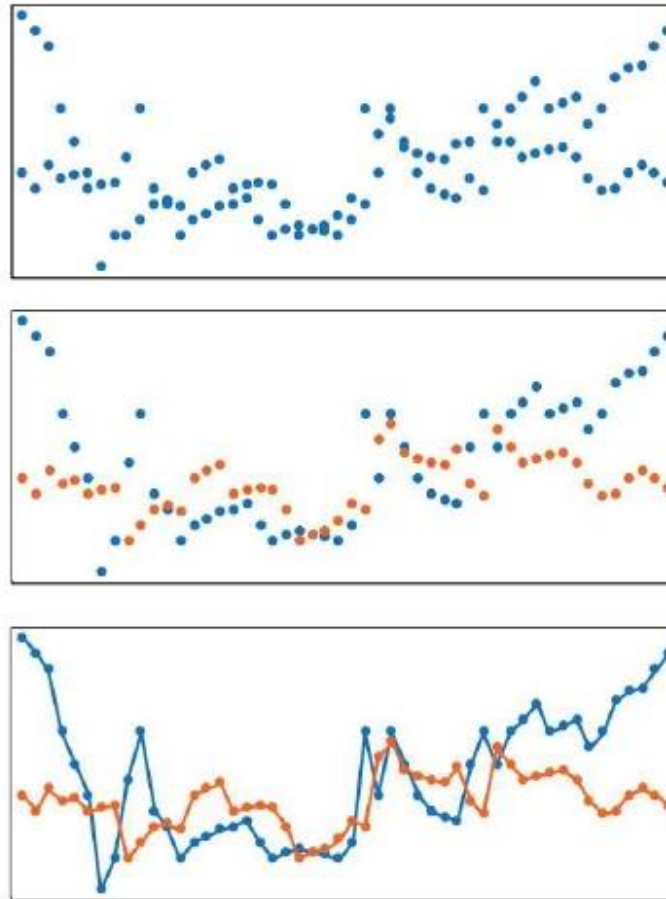
5. CONTINUIDADE

- Objetos com algum grau de continuidade ou alinhamento são percebidos como um grupo.



6. CONEXÃO

- Percebemos elementos conectados como parte de um mesmo grupo.



Processamento Pré-atentivo

Processamento Pré-atentivo

- Área derivada dos princípios Gestalt que busca formas de chamar atenção para informações visuais específicas.

Quantos números 3 existem nesta imagem?

756395068473

658663037576

860372658602

846589107830

E agora?

756 395068473

658663037576

860 372658602

846589107830

Utilizando em textos

Sem atributos pré-atentivos

O que estamos fazendo bem? Ótimos produtos. Esses produtos são claramente os melhores de sua classe. Peças substitutas são enviadas quando necessário. Vocês me enviam juntas sem que eu precise pedir. Os problemas são resolvidos prontamente. O escritório de cobranças resolveu rapidamente um problema que eu tinha. O serviço ao cliente supera as expectativas. O gerente de contas até fez contato fora do horário comercial. Vocês têm uma excelente empresa — continuem o bom trabalho!

E agora?

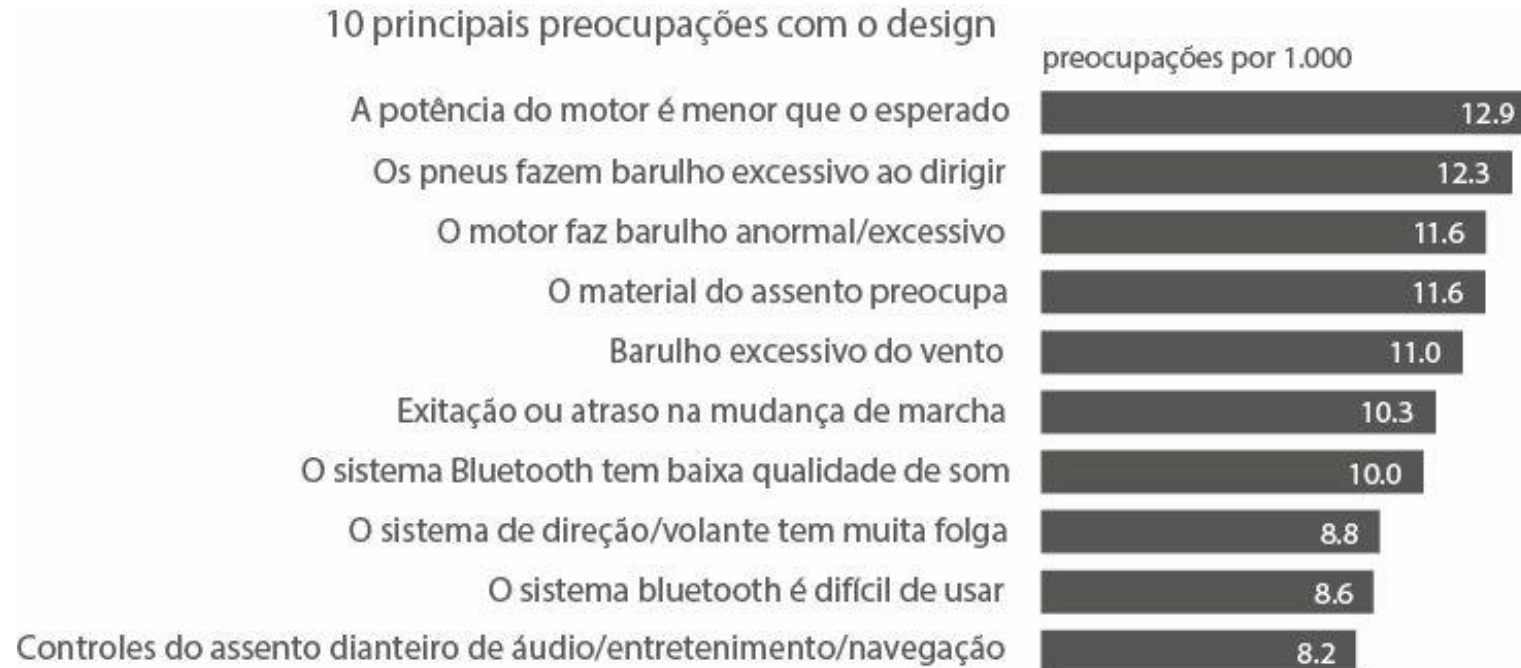
Negrito

O que estamos fazendo bem? Ótimos produtos. Esses produtos são claramente os melhores de sua classe. Peças substitutas são enviadas quando necessário. Vocês me enviam juntas sem que eu precise pedir. Os problemas são resolvidos prontamente. O escritório de cobranças resolveu rapidamente um problema que eu tinha. O serviço ao cliente supera as expectativas. O gerente de contas até fez contato fora do horário comercial. Vocês têm uma excelente empresa — continuem o bom trabalho!

Cor

O que estamos fazendo bem? Ótimos produtos. Esses produtos são claramente os melhores de sua classe. Peças substitutas são enviadas quando necessário. Vocês me enviam peças sem que eu precise pedir. Os problemas são resolvidos prontamente. O escritório de cobranças resolveu rapidamente um problema que eu tinha. O serviço ao cliente supera as expectativas. O gerente de contas até fez contato fora do horário comercial. Vocês têm uma excelente empresa — continuem o bom trabalho!

Utilizando em gráficos

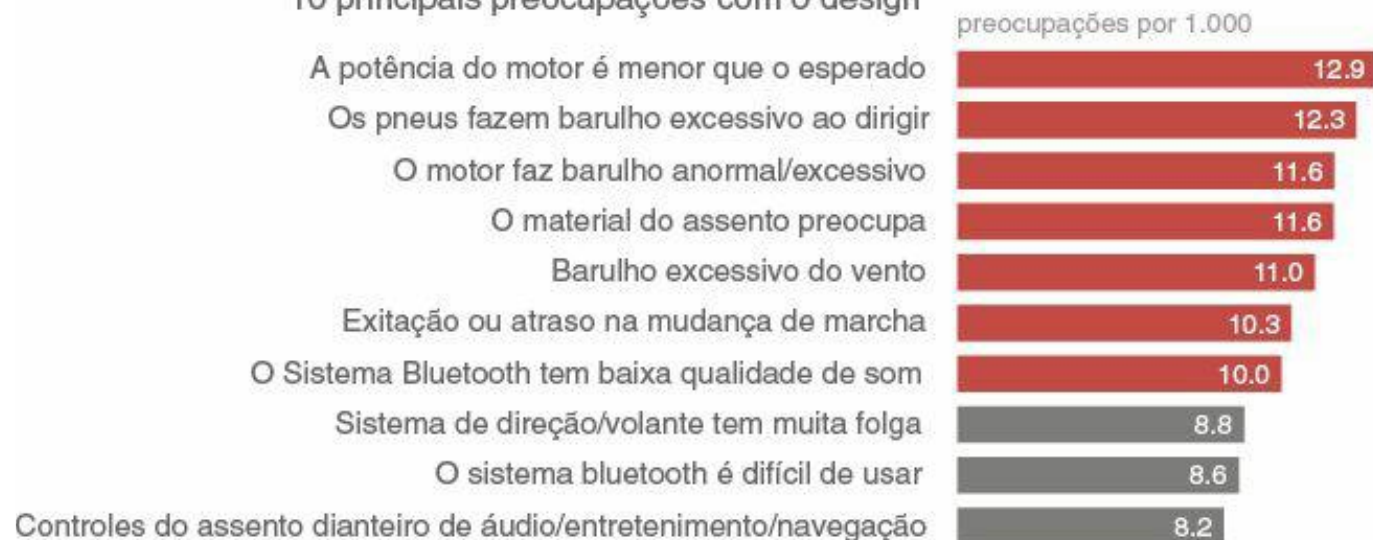


E agora?

7 das 10 principais preocupações com o design têm 10 ou mais preocupações por 1.000

Discussão: Essa é uma taxa padrão aceitável?

10 principais preocupações com o design



Mais um exemplo

Table 1. Our sales grew to \$600 million this year

	Q1	Q2	Q3	Q4
Bob	26	35	72	84
Ellie	22	15	61	35
Gerrie	19	20	71	55
Jack	22	95	13	64
Jon	83	62	46	48
Karen	30	65	98	82
Ken	38	28	45	71
Lauren	98	81	41	63
Steve	16	50	23	41
Valerie	46	24	30	57
Total	\$400	\$475	\$500	\$600

E agora?

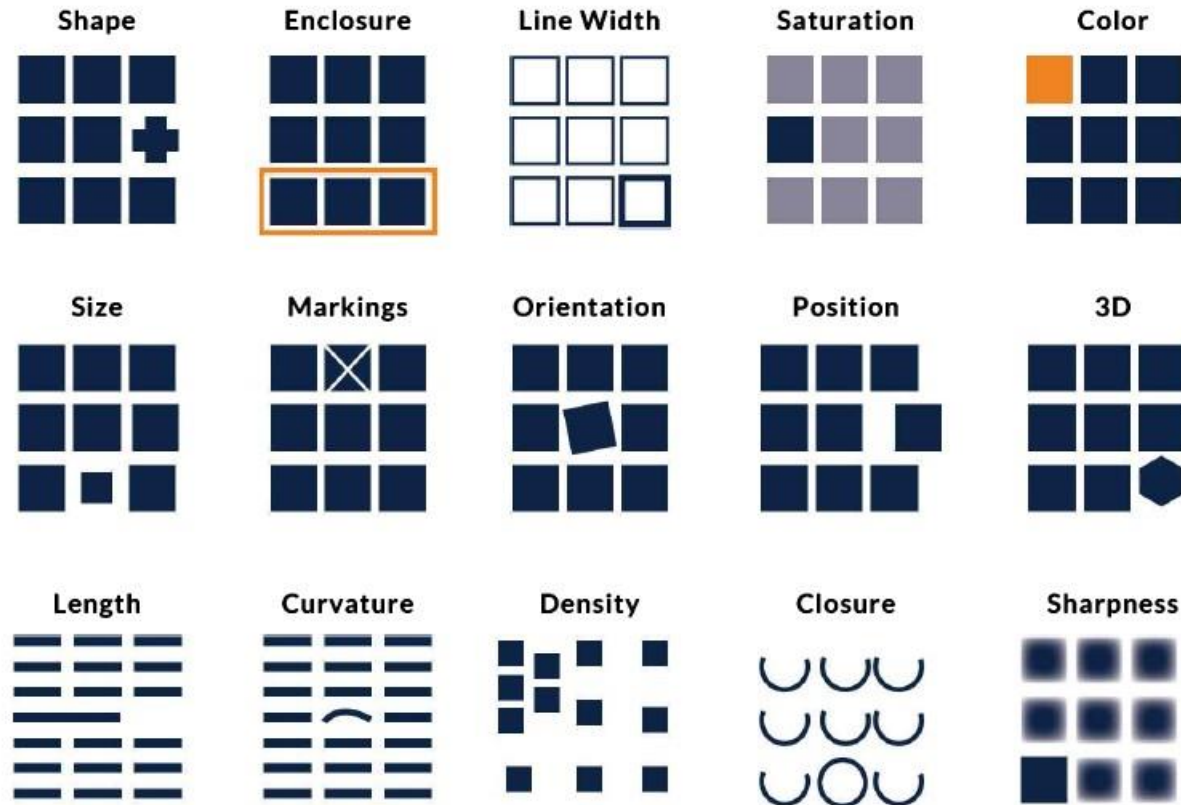
Table 1. Our sales grew to \$600 million this year

	Q1	Q2	Q3	Q4
Bob	26	35	72	84
Ellie	22	15	61	35
Gerrie	19	20	71	55
Jack	22	95	13	64
Jon	83	62	46	48
Karen	30	65	98	82
Ken	38	28	45	71
Lauren	98	81	41	63
Steve	16	50	23	41
Valerie	46	24	30	57
Total	\$400	\$475	\$500	\$600

Table 1. Our sales grew to \$600 million this year

	Q1	Q2	Q3	Q4
Bob	26	35	72	84
Ellie	22	15	61	35
Gerrie	19	20	71	55
Jack	22	95	13	64
Jon	83	62	46	48
Karen	30	65	98	82
Ken	38	28	45	71
Lauren	98	81	41	63
Steve	16	50	23	41
Valerie	46	24	30	57
Total	\$400	\$475	\$500	\$600

Referência



Dicas para criação de visualizações melhores

1. APRESENTAR OS DADOS DE FORMA COERENTE

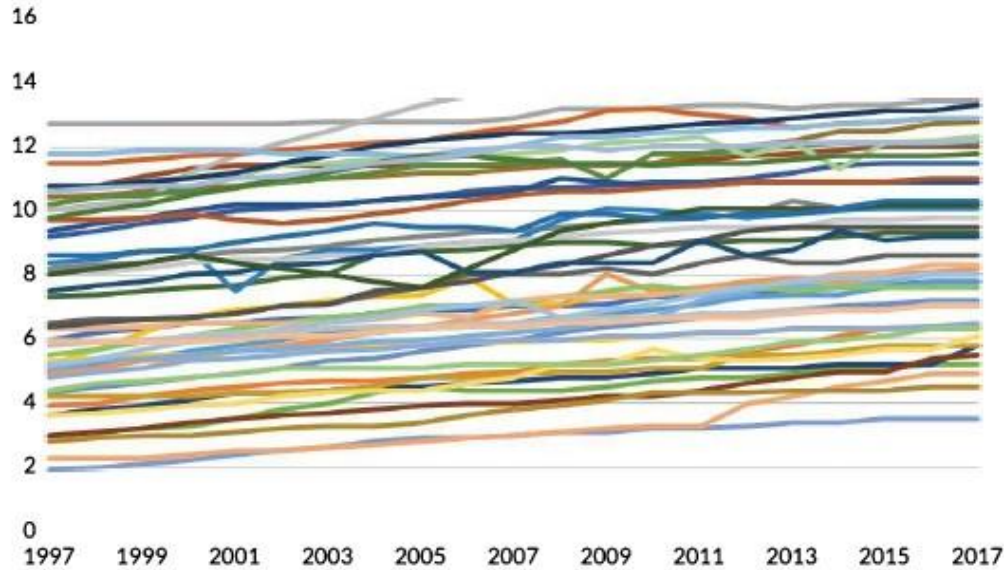
- Devemos apresentar TODOS os dados? Algumas vezes, sim...
- Dados demográficos dos E.U.A. segundo o censo de 2010



1. APRESENTAR OS DADOS DE FORMA COERENTE

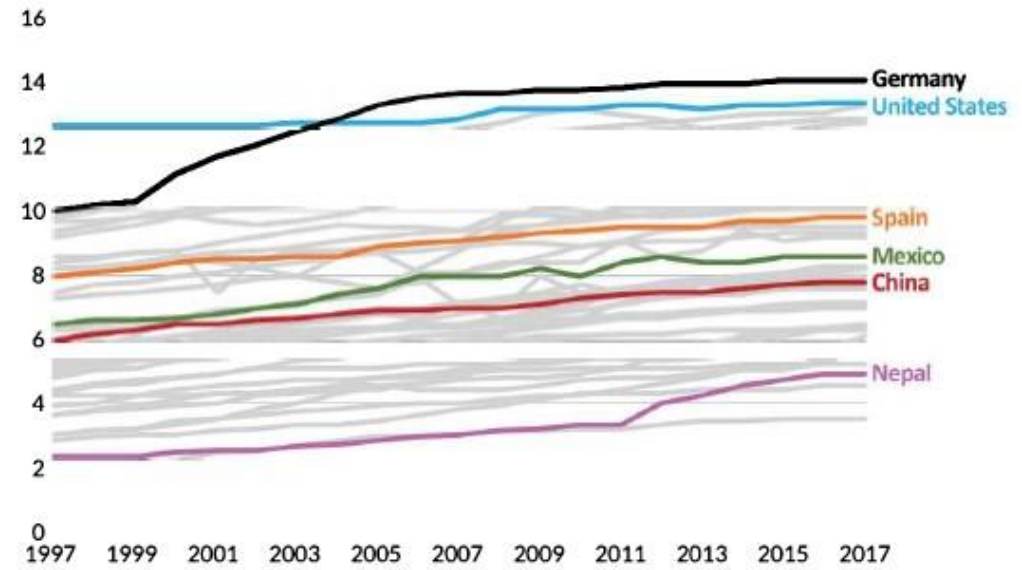
- Outras vezes, não, pelo menos não diretamente, ou devidamente destacando o que queremos dar maior importância.

Average years of schooling has increased around the world
(Number of years)



Source: Our World in Data

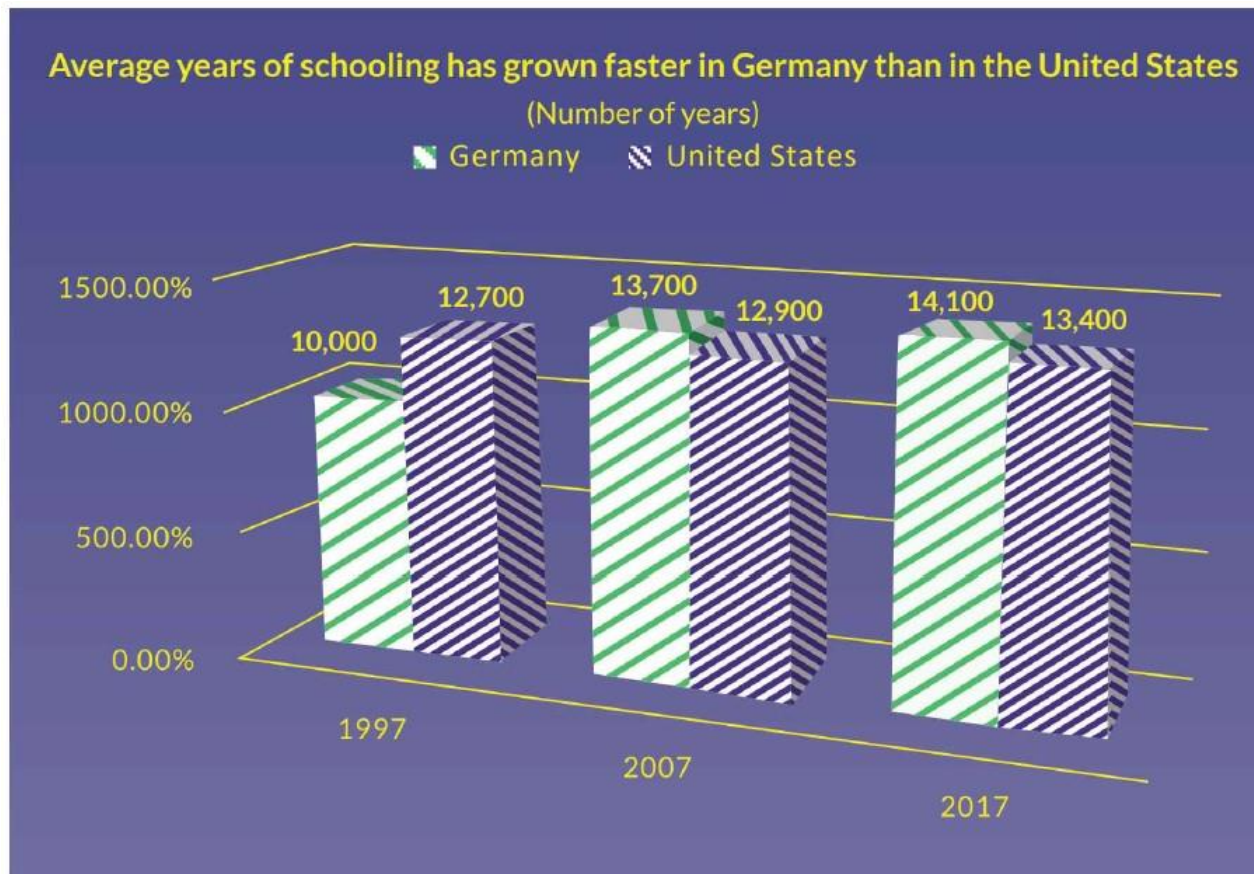
Average years of schooling has increased around the world
(Number of years)



Source: Our World in Data

2. DIMINUIR A POLUIÇÃO VISUAL

- Devemos manter apenas os elementos visuais principais e que nos ajudem a passar a ideia correta sobre os dados.
- Antes...

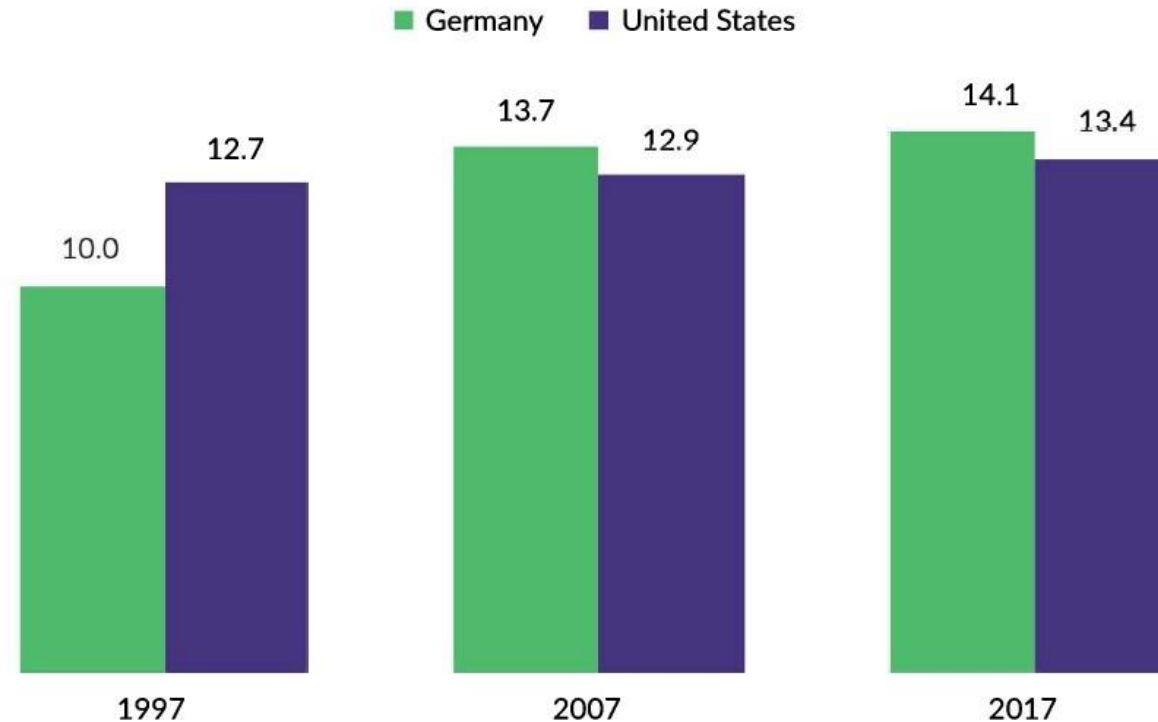


2. DIMINUIR A POLUIÇÃO VISUAL

- Depois...
- O que é melhor?

Average years of schooling has grown faster in Germany than in the United States

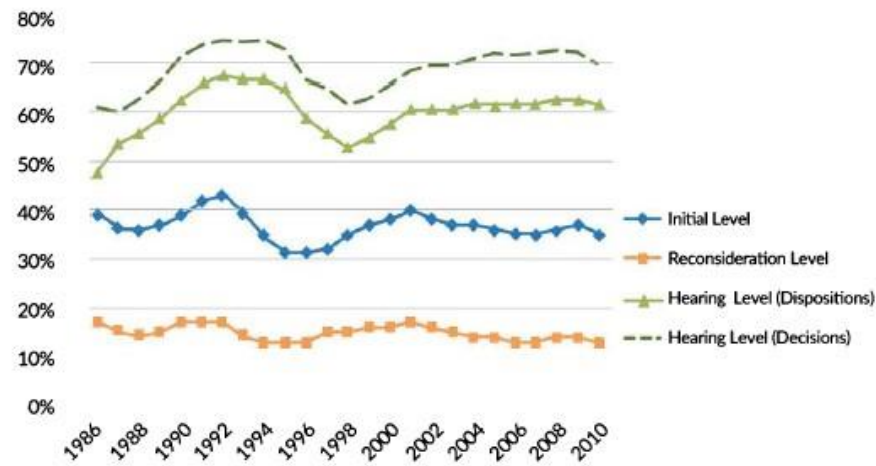
(Number of years)



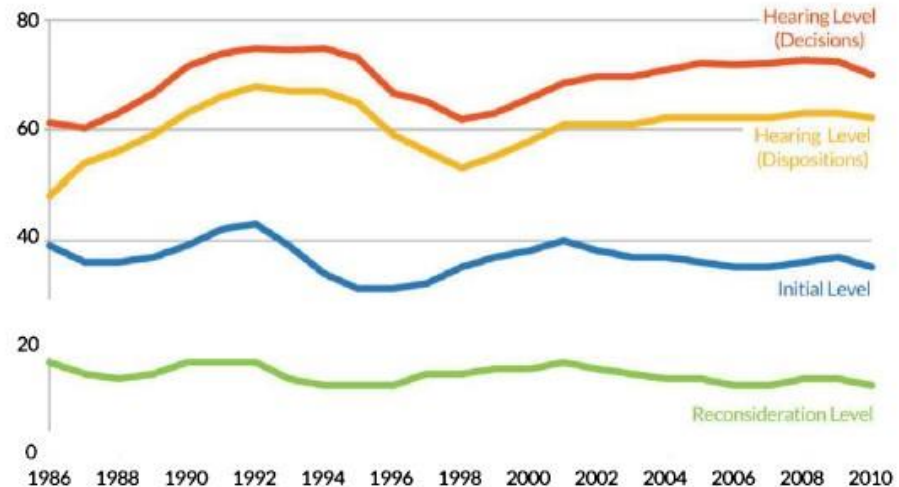
3. INTEGRAR GRÁFICOS E TEXTOS, COMO APOIO

- Apesar das visualizações estarem fortemente ligadas a elementos gráficos, podemos e devemos utilizar elementos textuais sempre que possível para diminuir a poluição visual e apresentar melhor os objetivos.

DI and SSI allowance rates have generally moved in tandem over the past 25 years
(Percent)



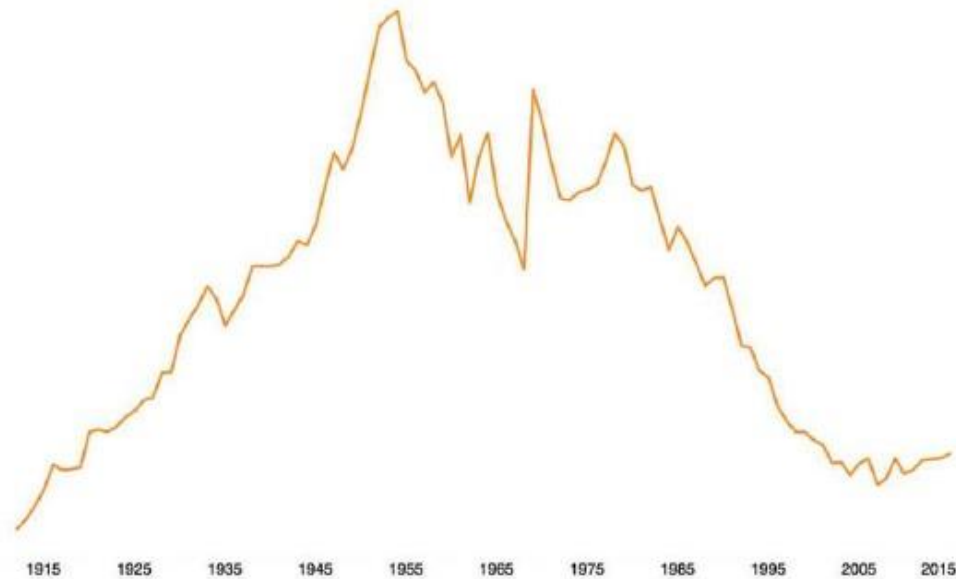
DI and SSI allowance rates have generally moved in tandem over the past 25 years
(Percent)



3. INTEGRAR GRÁFICOS E TEXTOS, COMO APOIO

Rise and Fall of the name **Neil** in the USA
Births 1912-2015

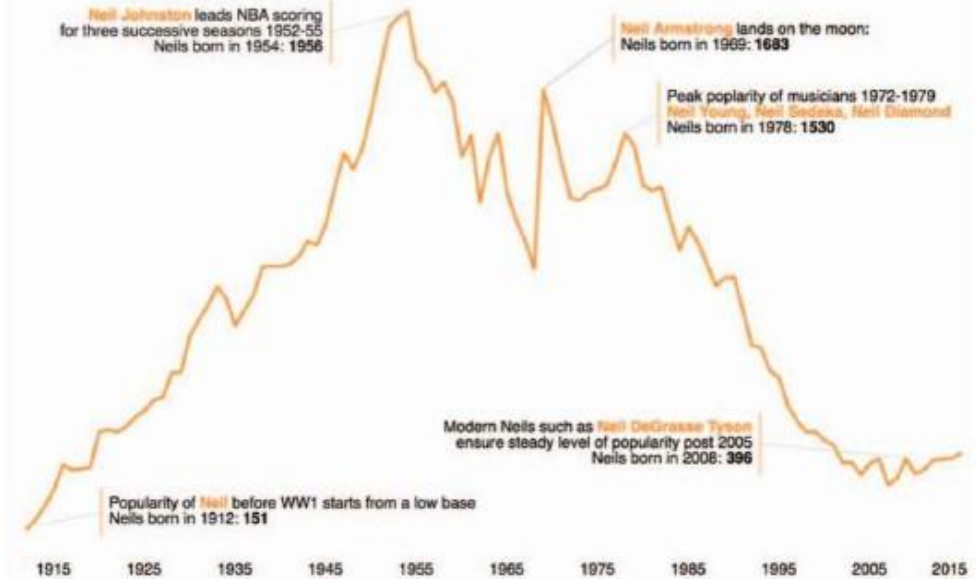
Source: data.gov



Visualisation: @theneilrichards

Rise and Fall of the name **Neil** in the USA
Births 1912-2015

Source: data.gov



Visualisation: @theneilrichards

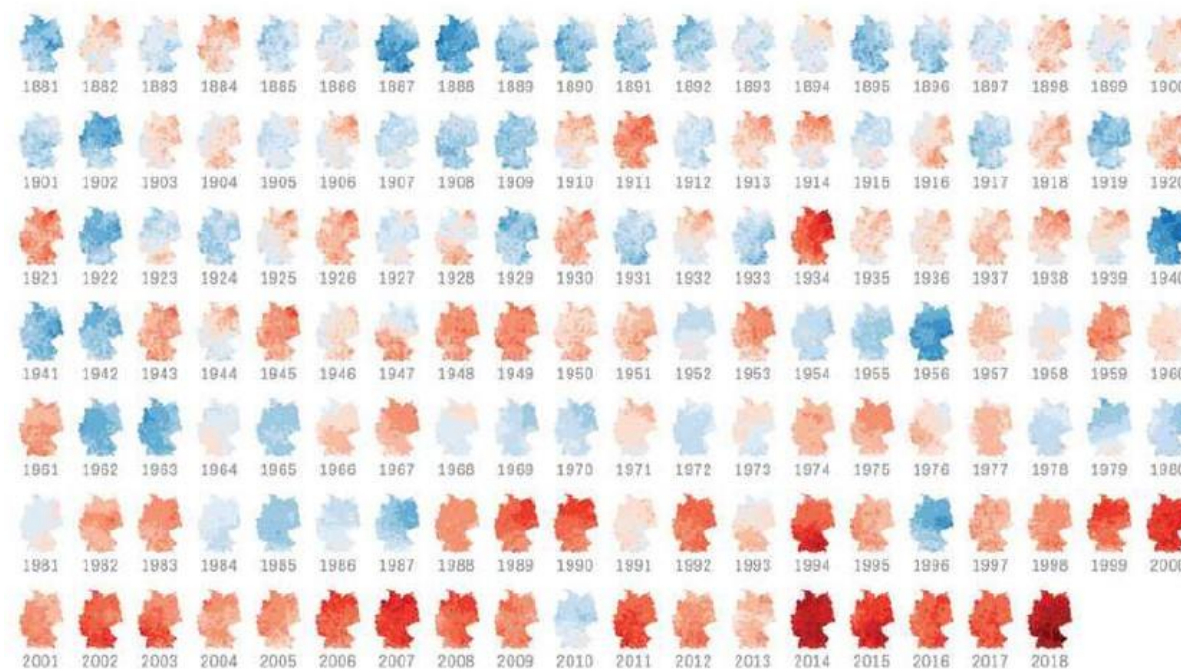
#SWDChallenge

4. EVITAR EXCESSO DE INFORMAÇÕES

- Podemos usar recursos visuais para facilitar a interpretação de informações muito densas.

Nine of the ten hottest years: all after the turn of the millennium

2018 was the hottest year since records started, with an average annual temperature of 10.5 ° C. 1940 the coldest with 6.6 ° C.



Source: German Weather Service (DWD), own calculations

Como escolher uma
visualização ideal para os
dados?

Como escolher o seu gráfico? Algumas sugestões

