

Looking at Data

Data Visualisation mini-course

Gavin L. Simpson

31/03/2019

Data Visualisation

Some data visualisations are better than others

Why?



Data Visualisation

Some data visualisations are better than others

- Taste
 - The eye is in the beholder
- Choice of data
- Human visual perception

What makes bad figures?

Aesthetic

Tacky, tasteless, ugly, hodgepodge, inconsistent design

Substantive

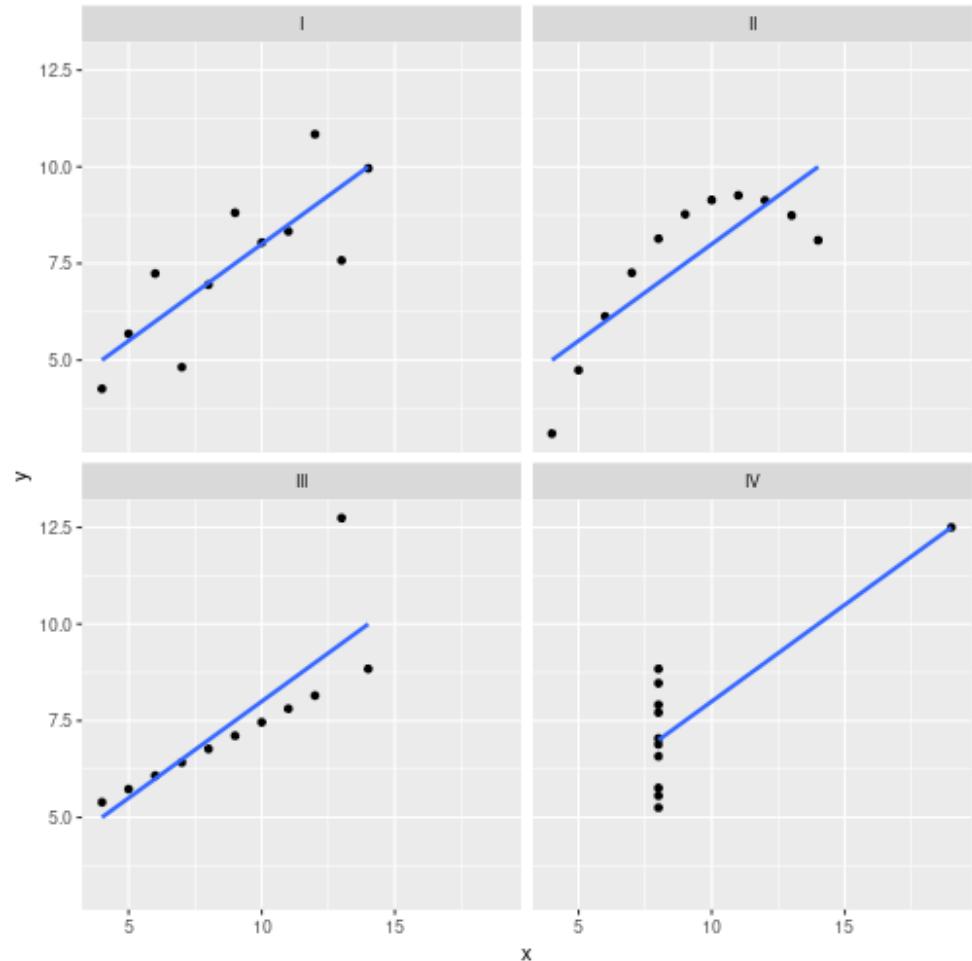
Graph has problems because of the data being presented

Perceptual

Inspite of good taste and good data, a graph may be *confusing* or *misleading* because of how people perceive and process what they are looking at

Always plot your data

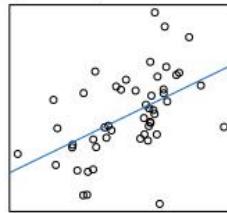
- Anscombe's quartet
- Scatterplot
 - 2 quantities mapped to x and y axis
- and in each set have the same
 - mean
 - variance
- Regression lines have same
- and have the same correlation



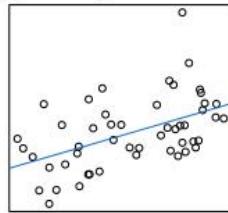
Visualising data

All correlations: $r(50) = 0.5$

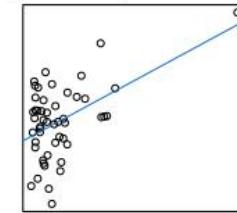
(1) Normal x, normal residuals



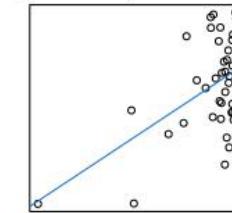
(2) Uniform x, normal residuals



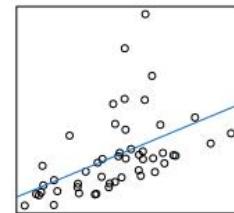
(3) +-skewed x, normal residuals



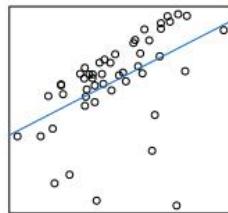
(4) --skewed x, normal residuals



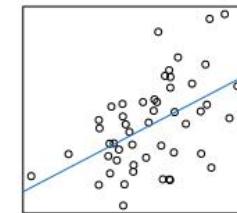
(5) Normal x, +-skewed residuals



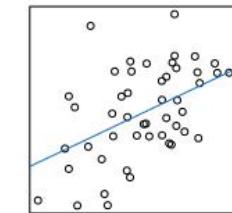
(6) Normal x, --skewed residuals



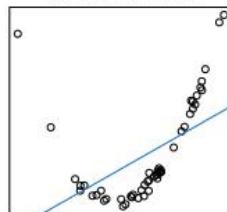
(7) Increasing spread



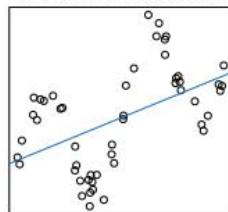
(8) Decreasing spread



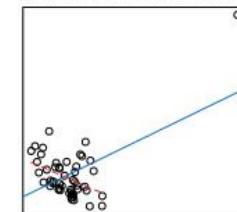
(9) Quadratic trend



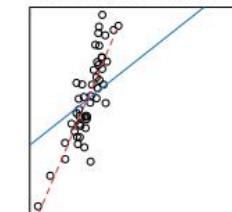
(10) Sinusoid relationship



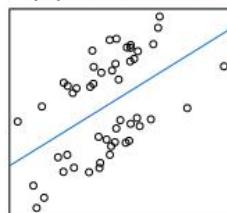
(11) A single positive outlier



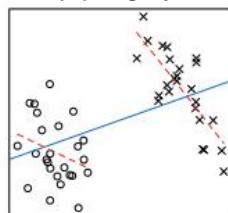
(12) A single negative outlier



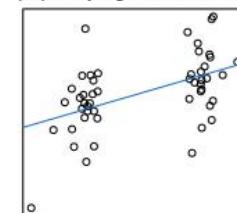
(13) Bimodal residuals



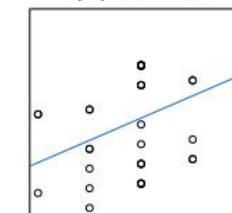
(14) Two groups



(15) Sampling at the extremes



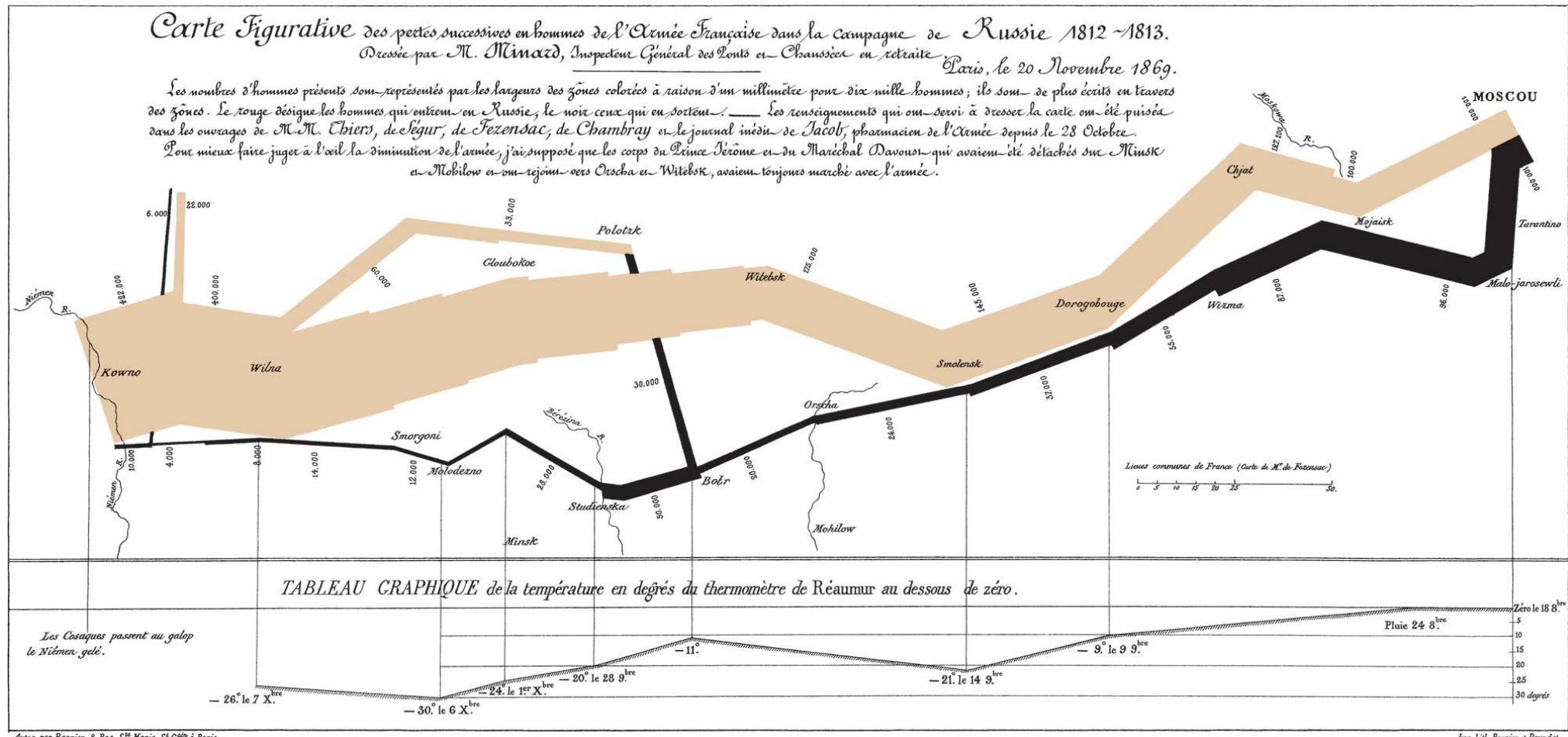
(16) Coarse data



Visualising data



Visualising data

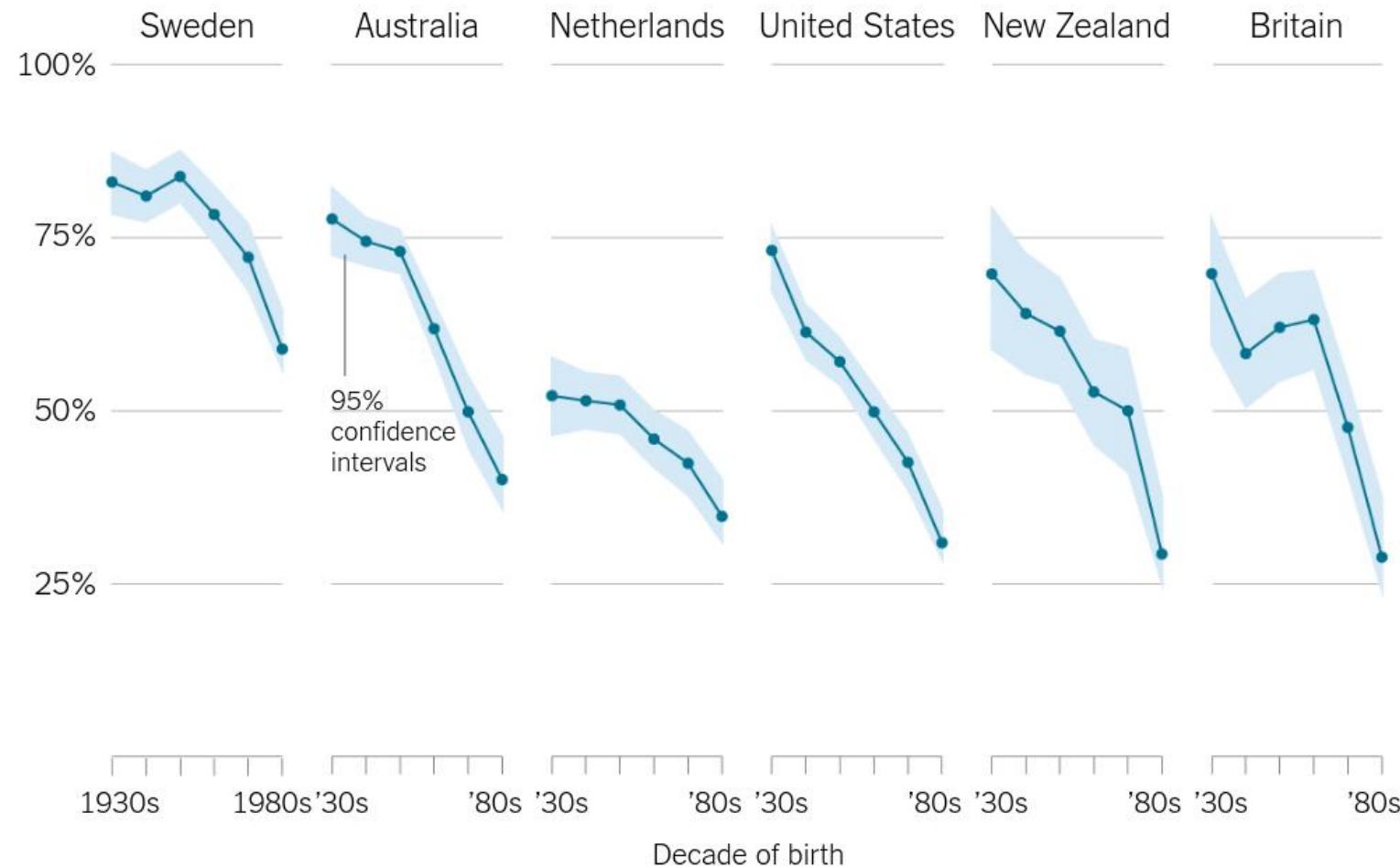




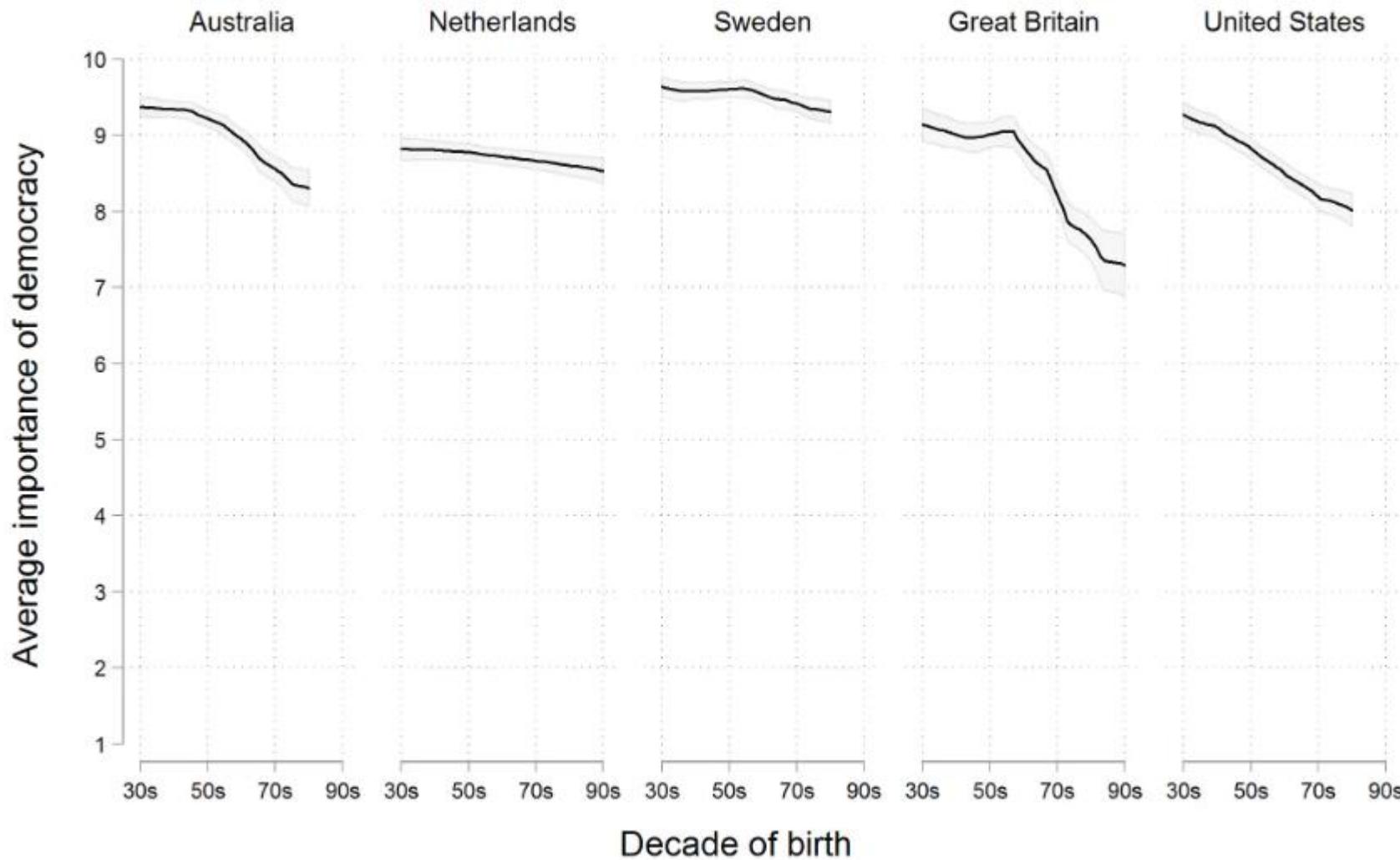
Data has a better idea

Bad data

Percentage of people who say it is “essential” to live in a democracy

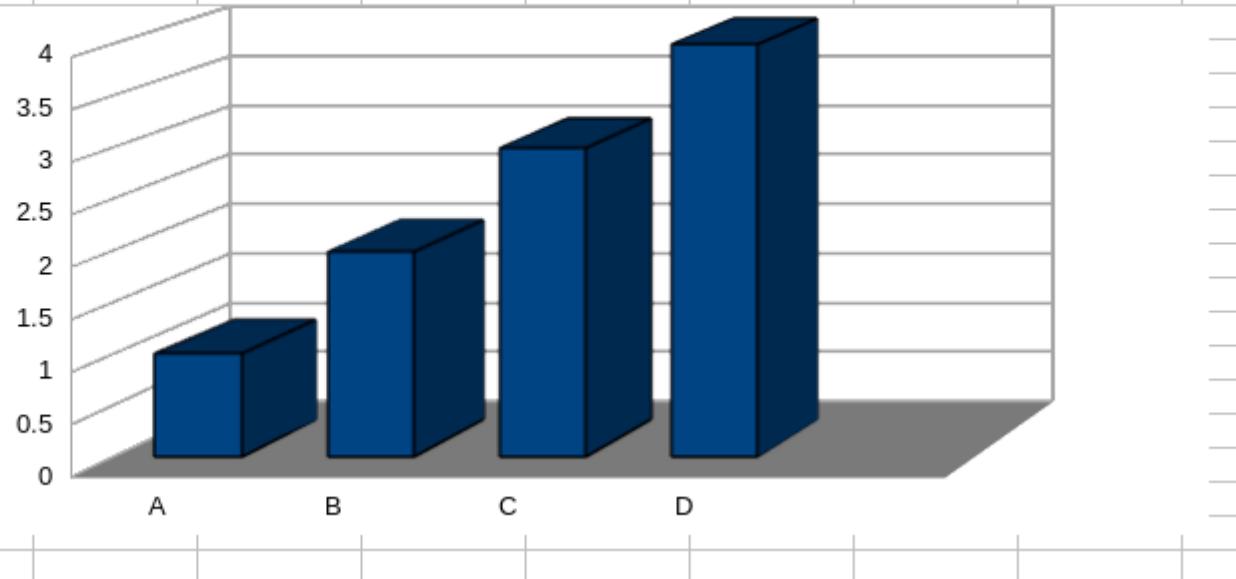


Bad data

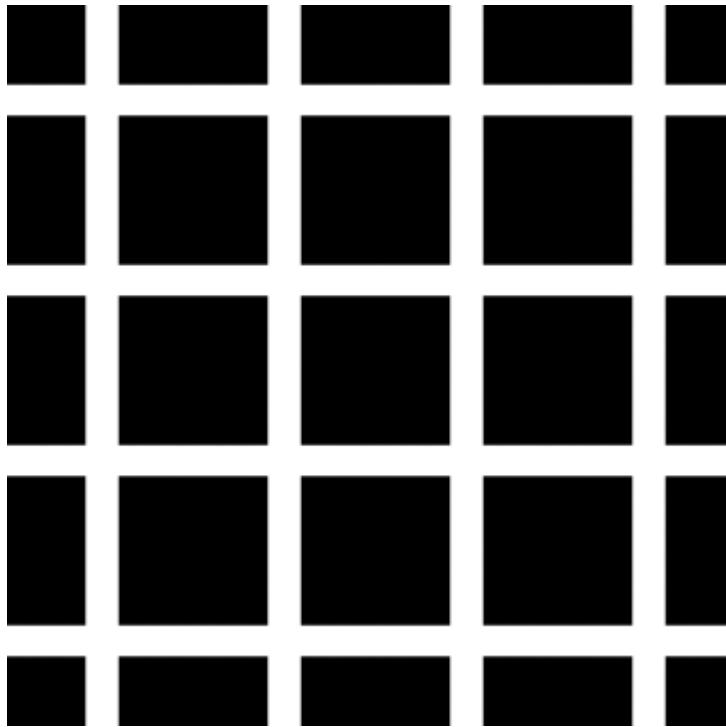


Bad Data

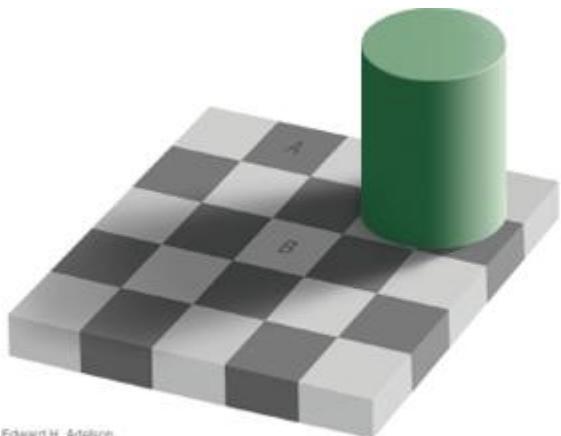
	A	B	C	D	E	F	G	H	I	J
1	A		1							
2	B		2							
3	C		3							
4	D		4							
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										



Perception

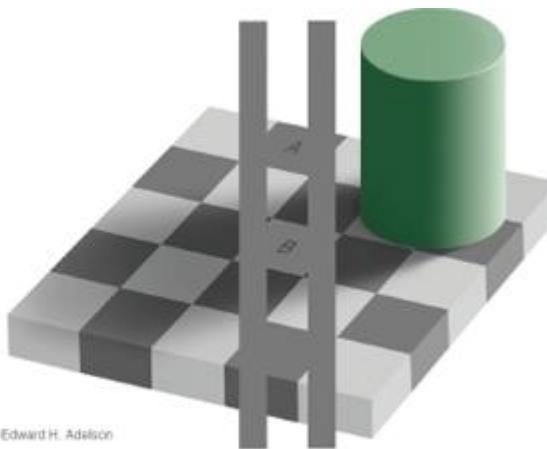


Perception



Edward H. Adelson

(a)

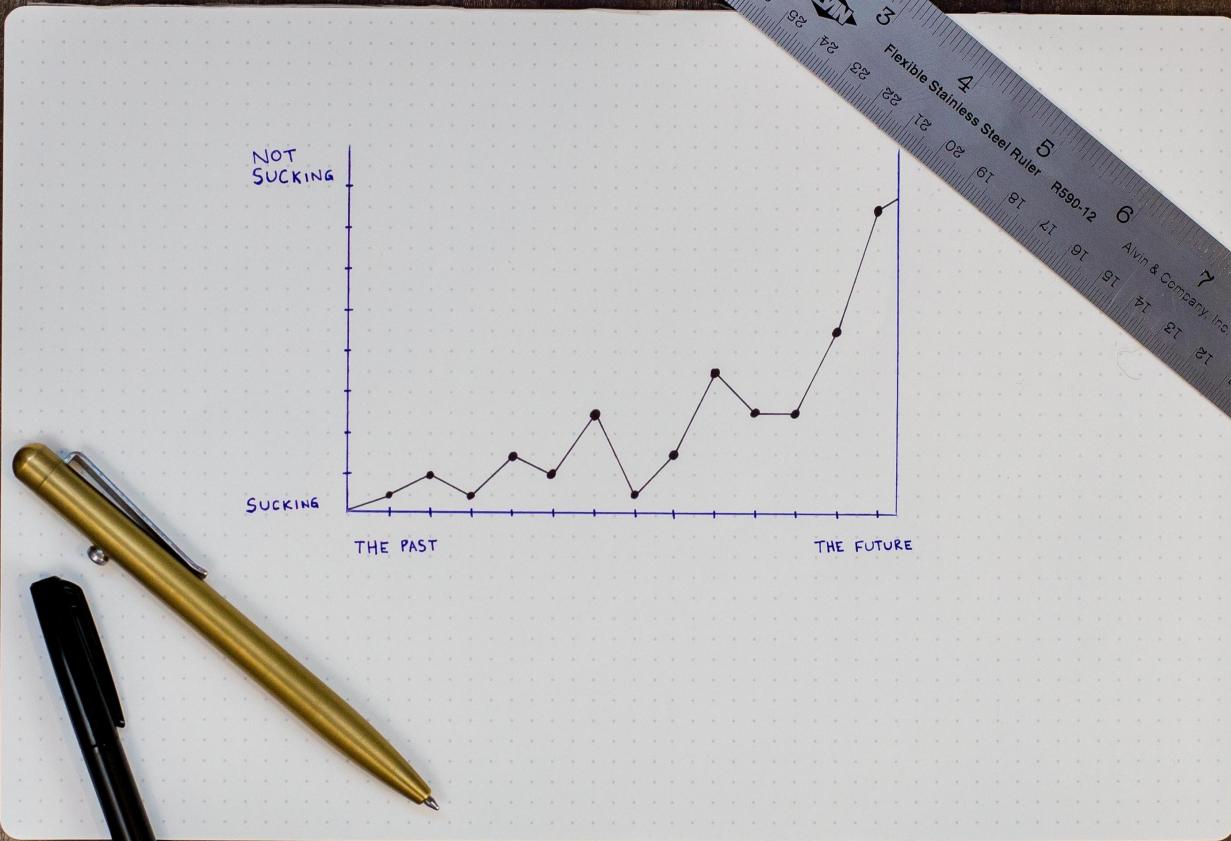


Edward H. Adelson

(b)



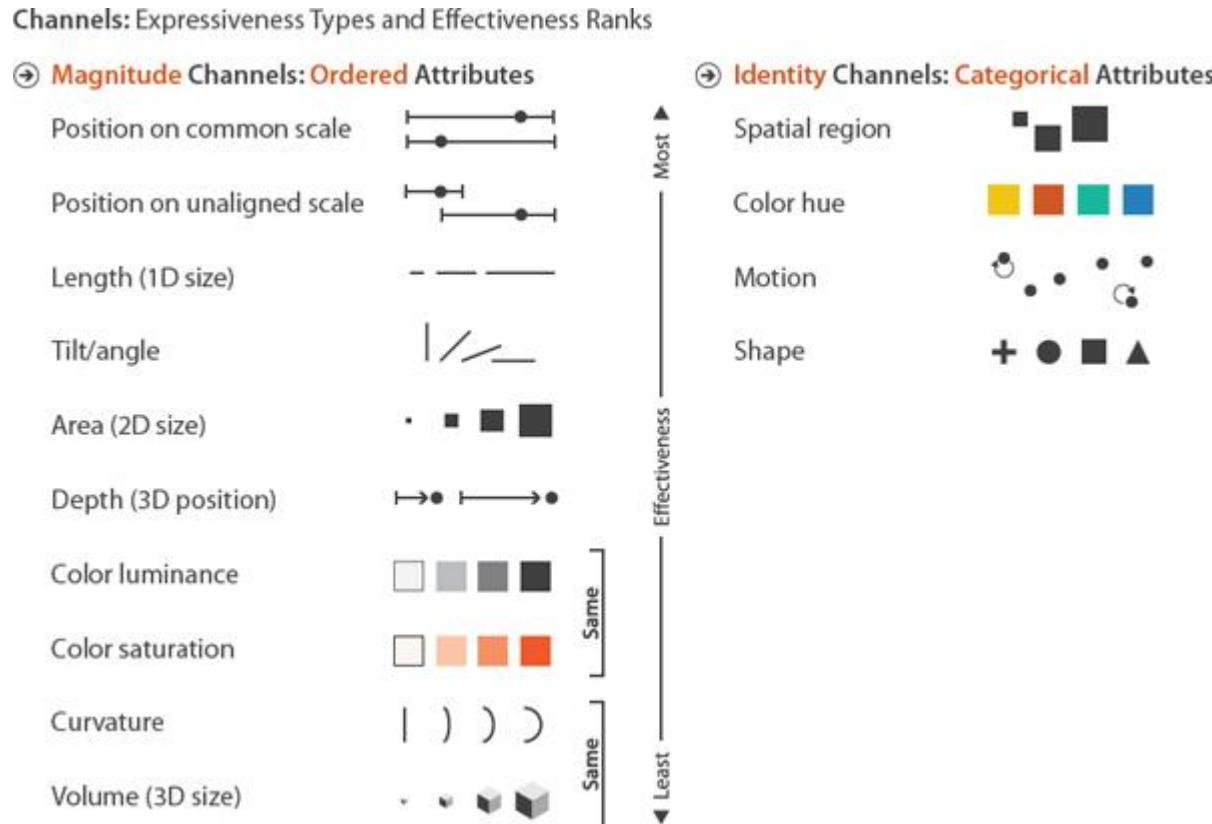
Mapping



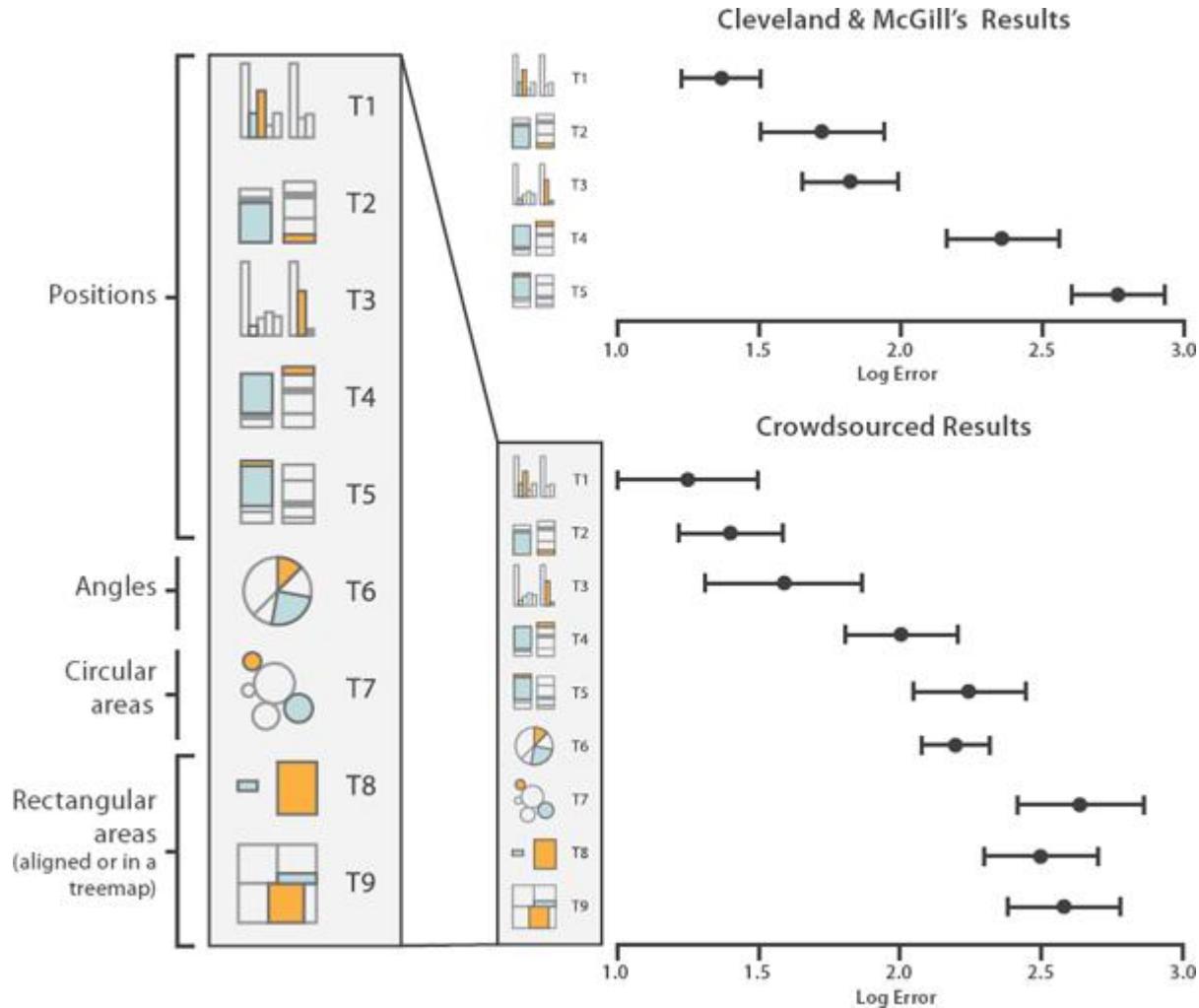
Mapping

Drawing a graph involves mapping data to visual attributes

Some mappings more effective than others



Accuracy of Mappings



Colour



Colour models

Colour models

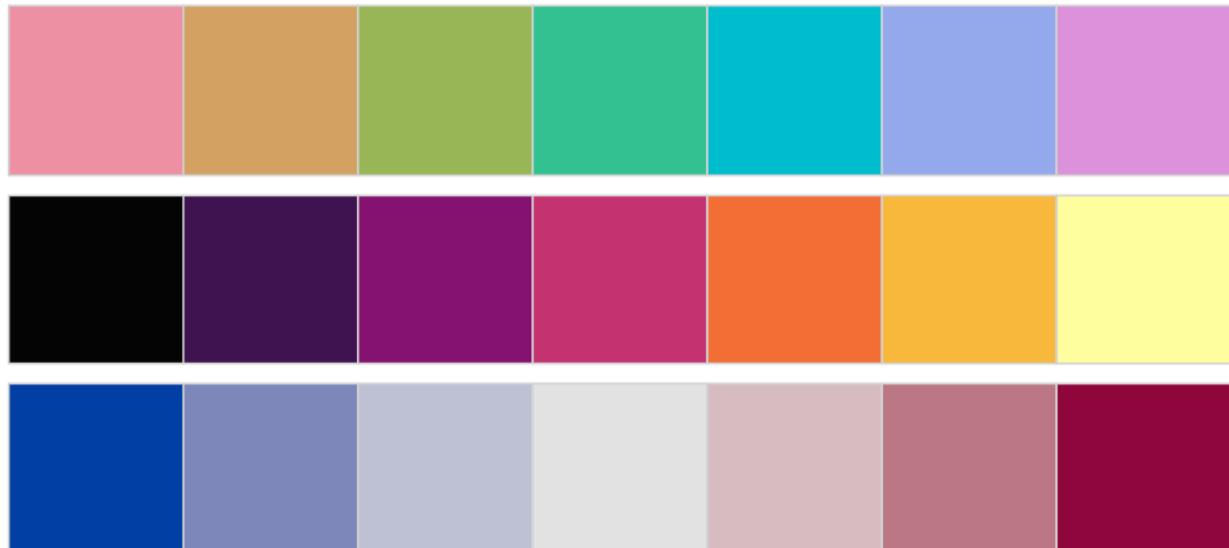
- Red Green Blue – RGB
 - Hexadecimal (base 16)
 - '0'–'9' with 'A'–'F' for 10-15
 - Encode 256 values of each colour
 - In R "#AA6633"
 - TVs, Digital cameras, etc
- Cyan Yellow Magenta Black – CYMK
 - Covers a much wider range of colours than RGB
 - Magazines, printing
- Hue Chroma Luminance – HCL
 - Hue – colour
 - Chroma – how much of the colour
 - Luminance – brightness
 - Designed to reflect human colour perception

Colour palettes

A colour palette is a the colour scheme or selection used to represent data or design on a graph

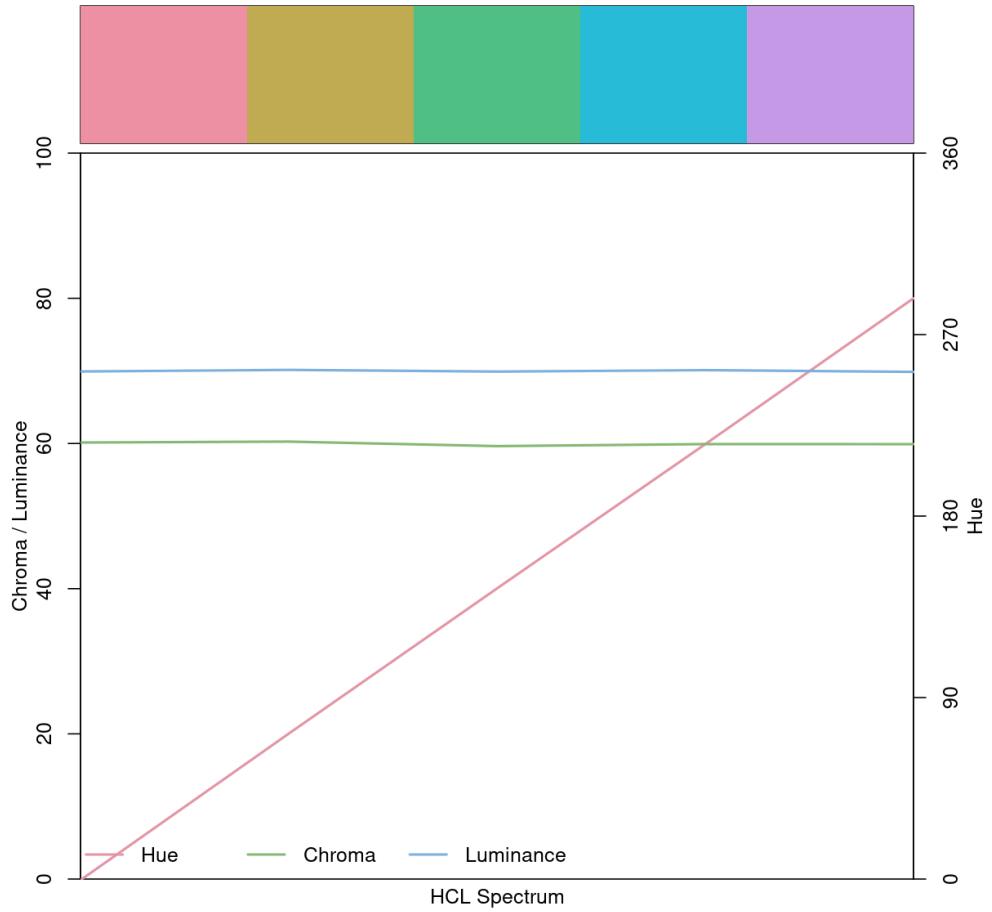
Want more than a numerical mapping – want perceptually uniform mappings

- discrete
- sequential
- diverging



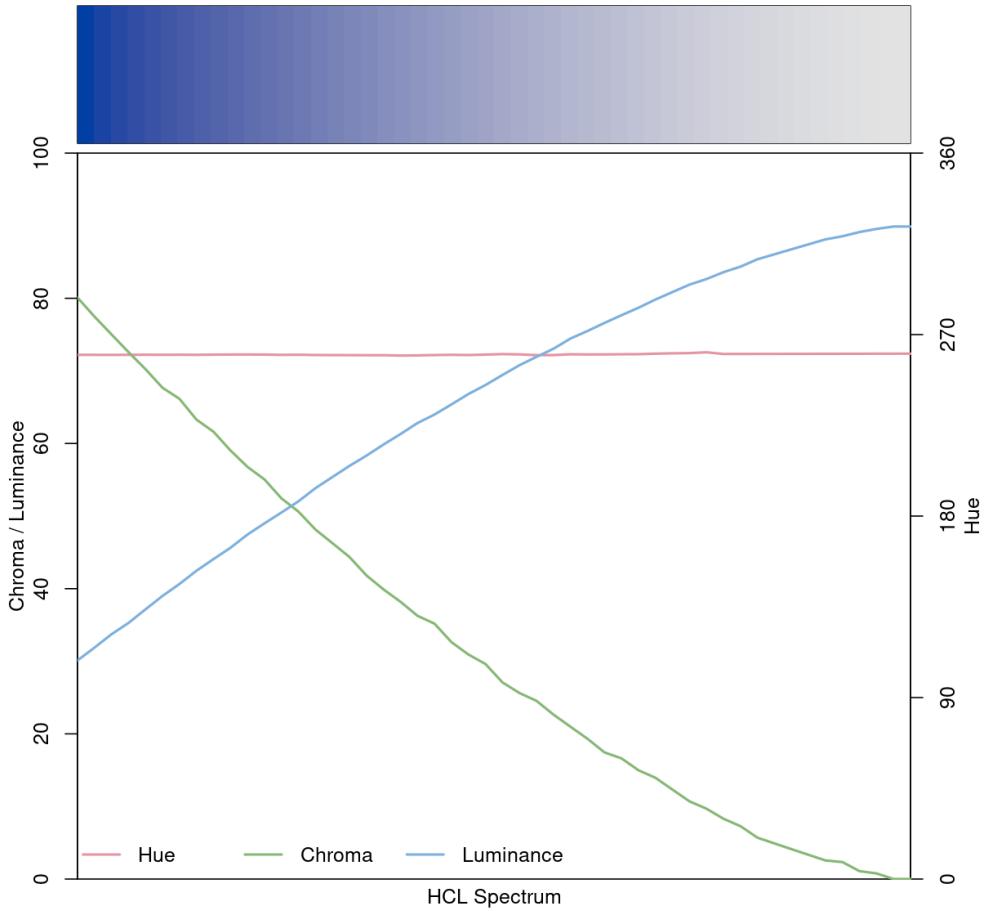
Discrete palettes

- Categorical data
- Easily distinguishable
- Favour no one colour
- Vary H, constant C & L



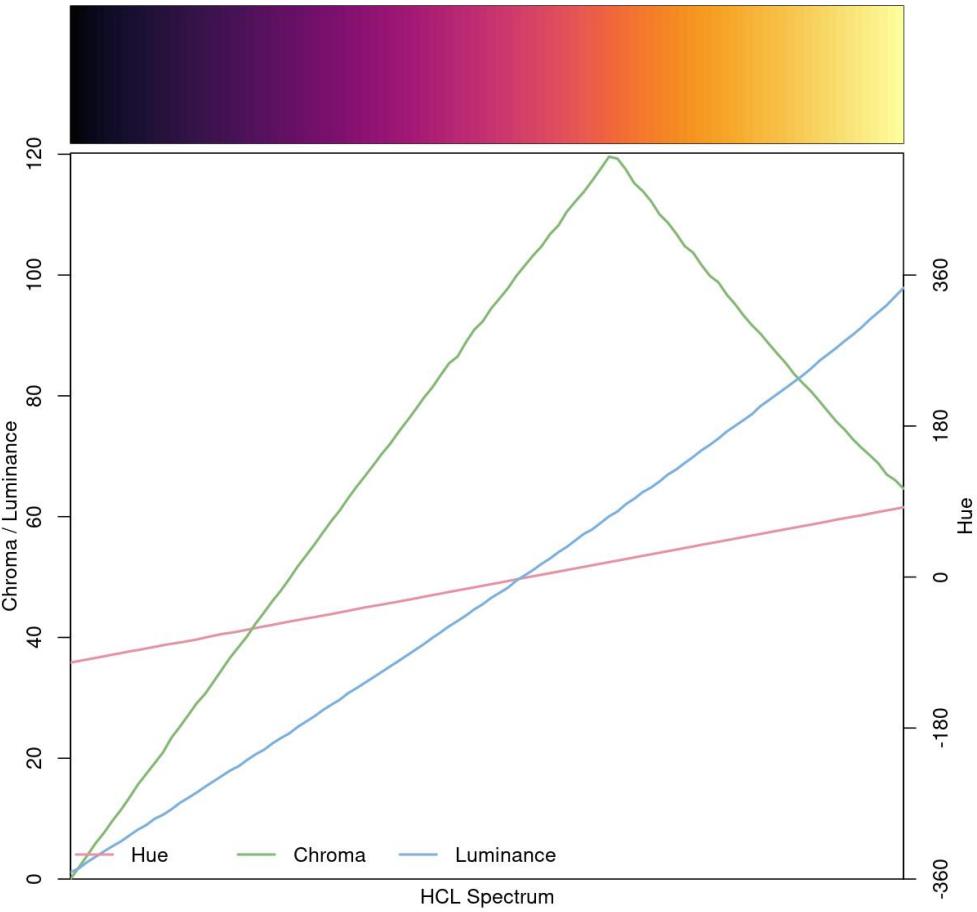
Sequential palettes

- Continuous data
- Brightness & intensity of colour vary
- Vary C & L, constant H



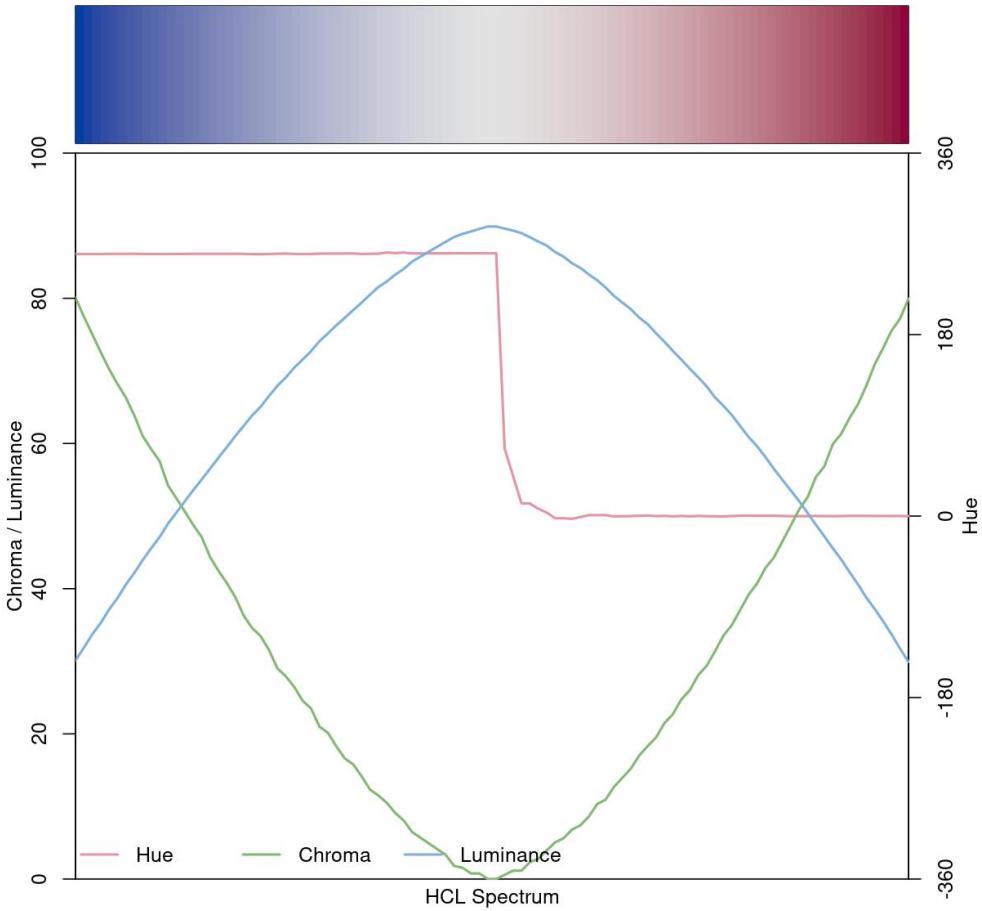
Sequential multi-hue palettes

- Continuous data
- Can vary everything if careful
- Vary H, C & L

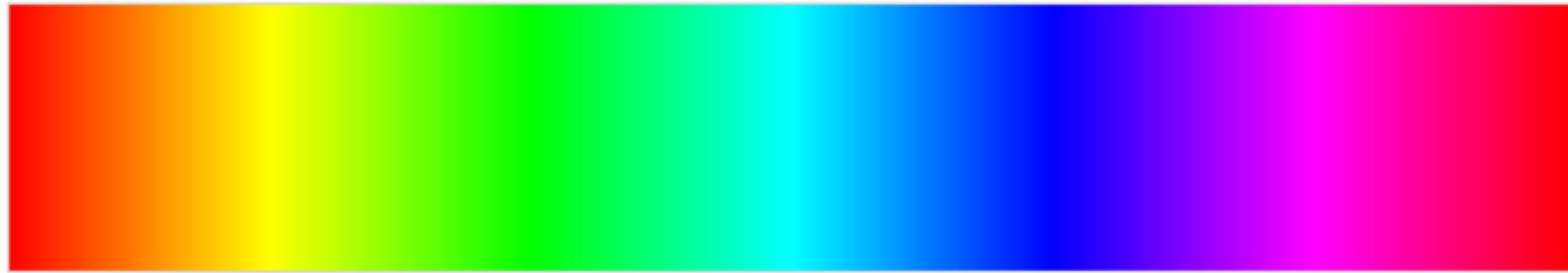


Diverging palettes

- Continuous data where mid-point means something (0)
- Single hue in each arm
- C & L are balanced in each arm
- C goes to 0 at mid-point

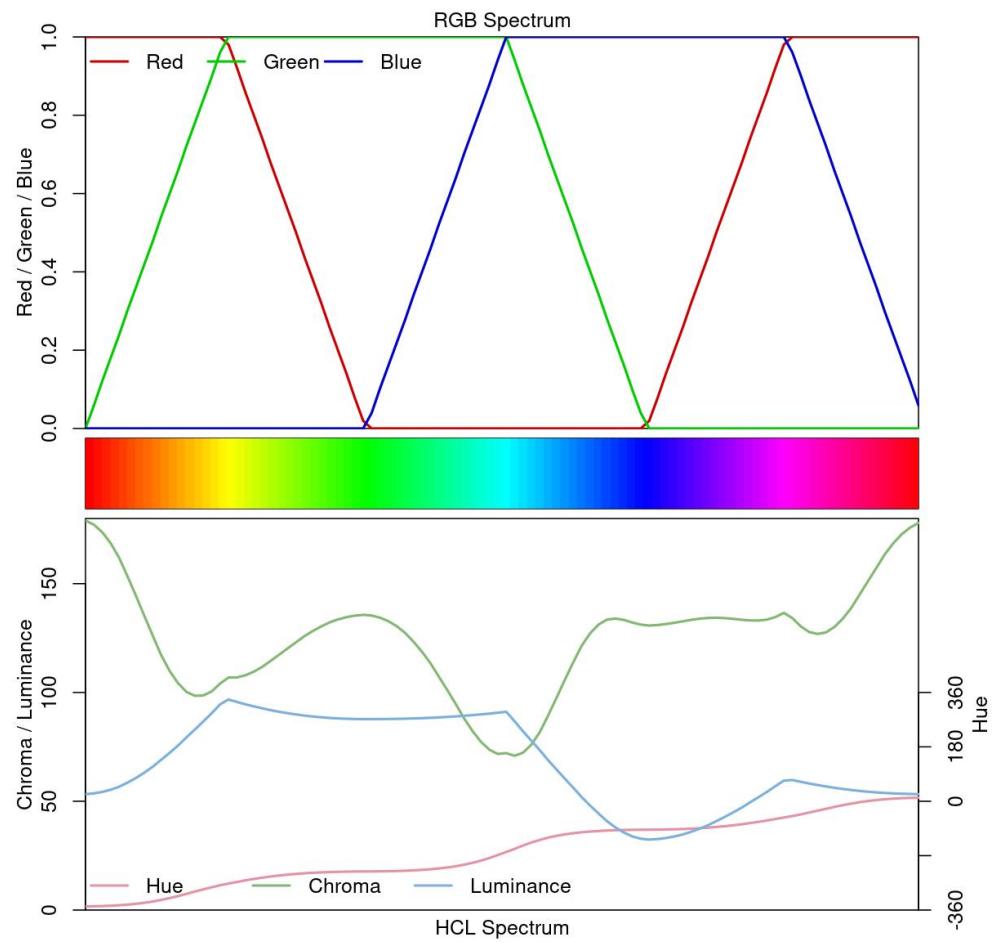


Rainbow



#endrainbow

- Luminance is not linear or even monotonic
- Colour vision deficiency



#endrainbow



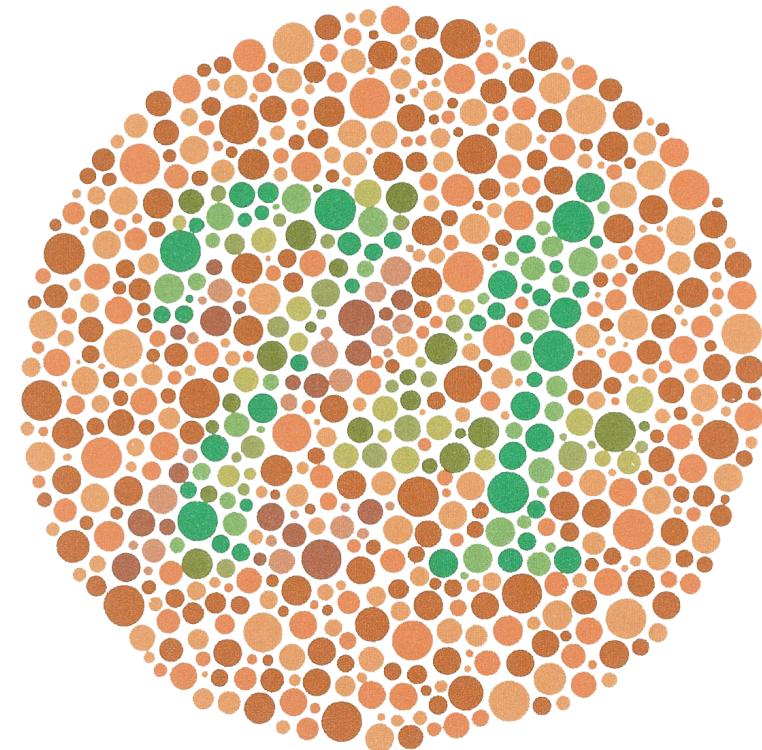
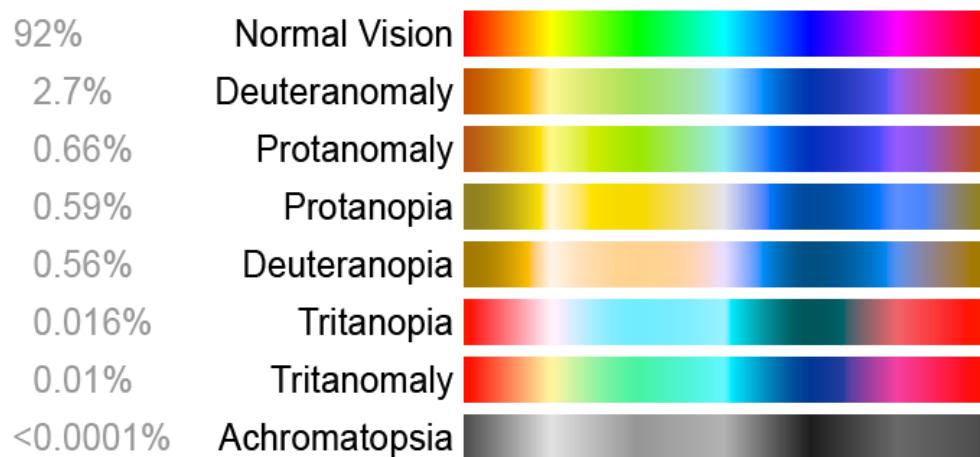
#endrainbow



Colour vision deficiency

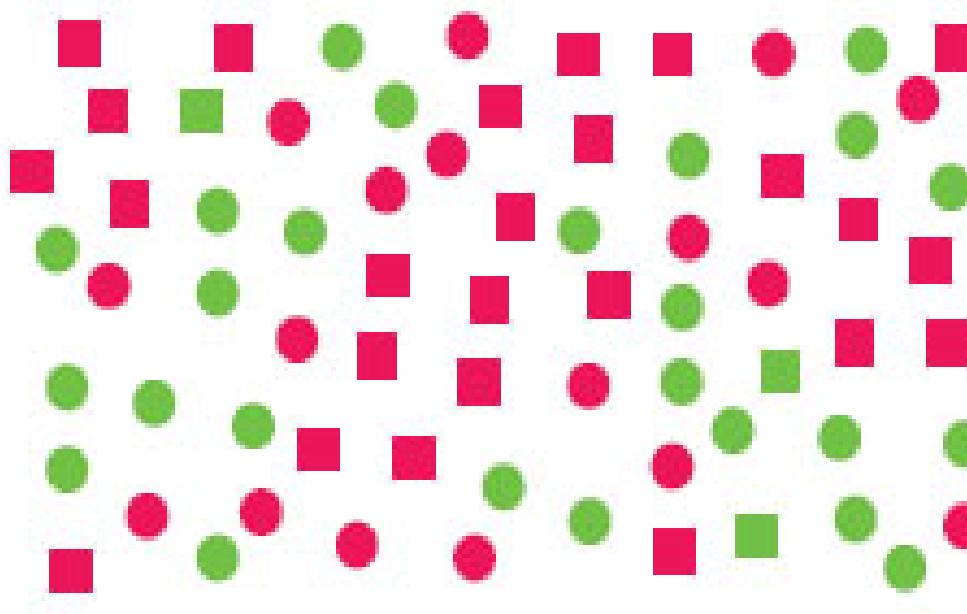
Decreased ability to see colour or differences in colour

- Red-Green CVD is sex-linked
- Gene carried on X chromosome
- Blue-Yellow CVD is not; chromosome 7
- -anomaly vs -anopia



Pop

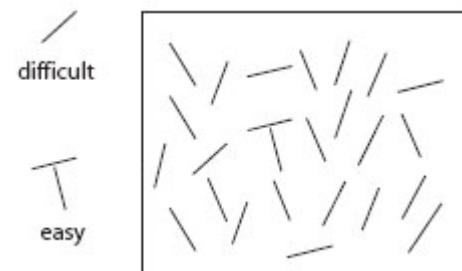
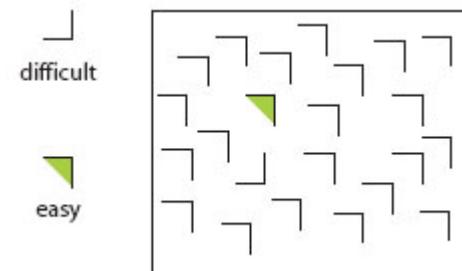
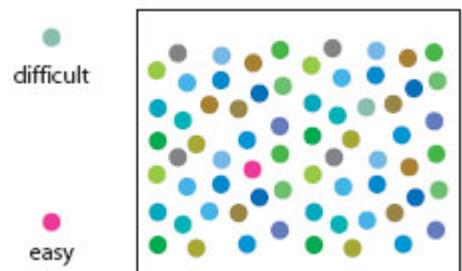
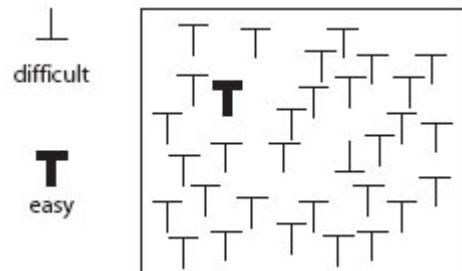
Can you see the green squares?



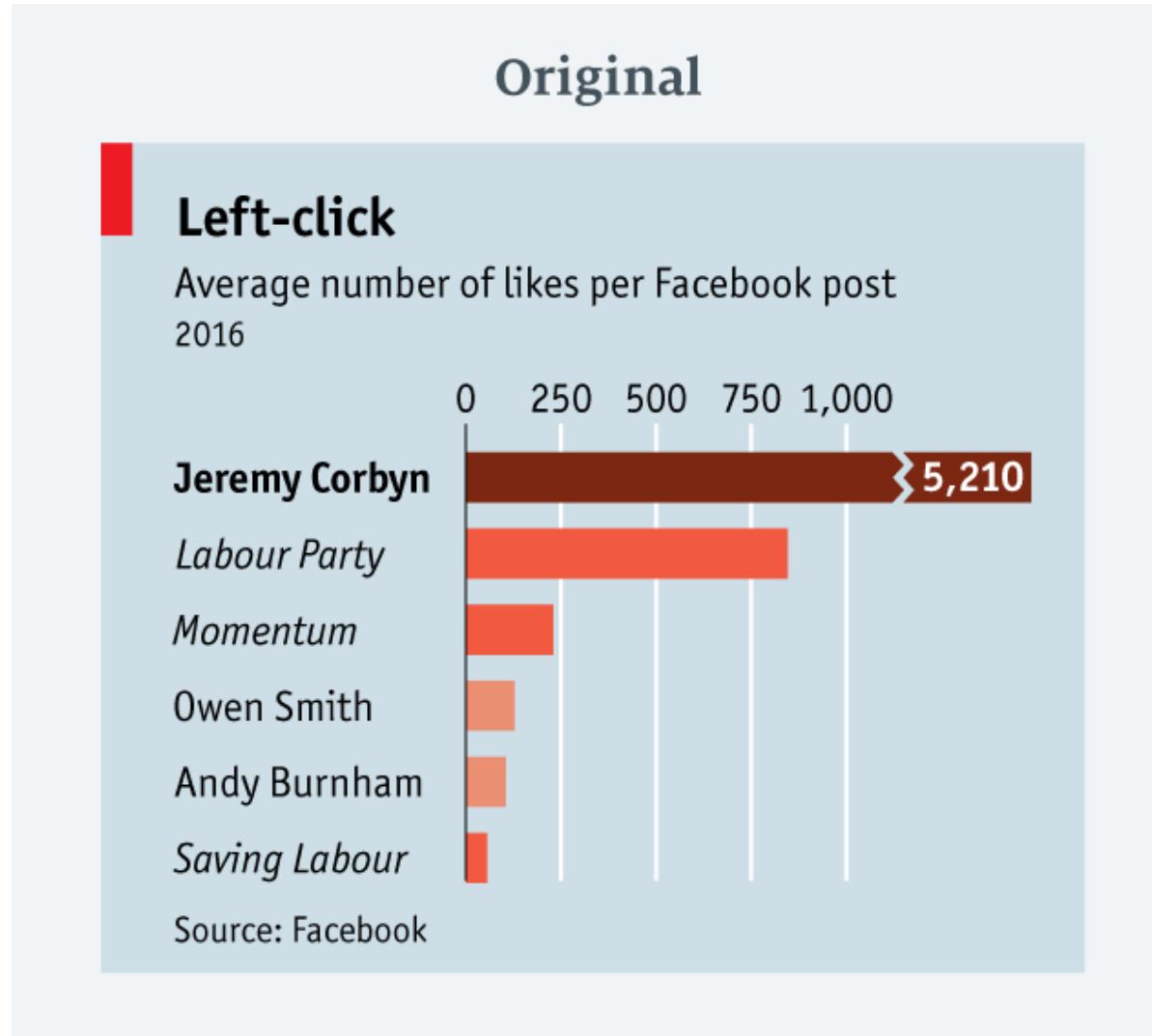
Preattentive pop-out

Some shapes, colours, angles more easy to spot

Can happen before (or almost before) before consciously looking at something



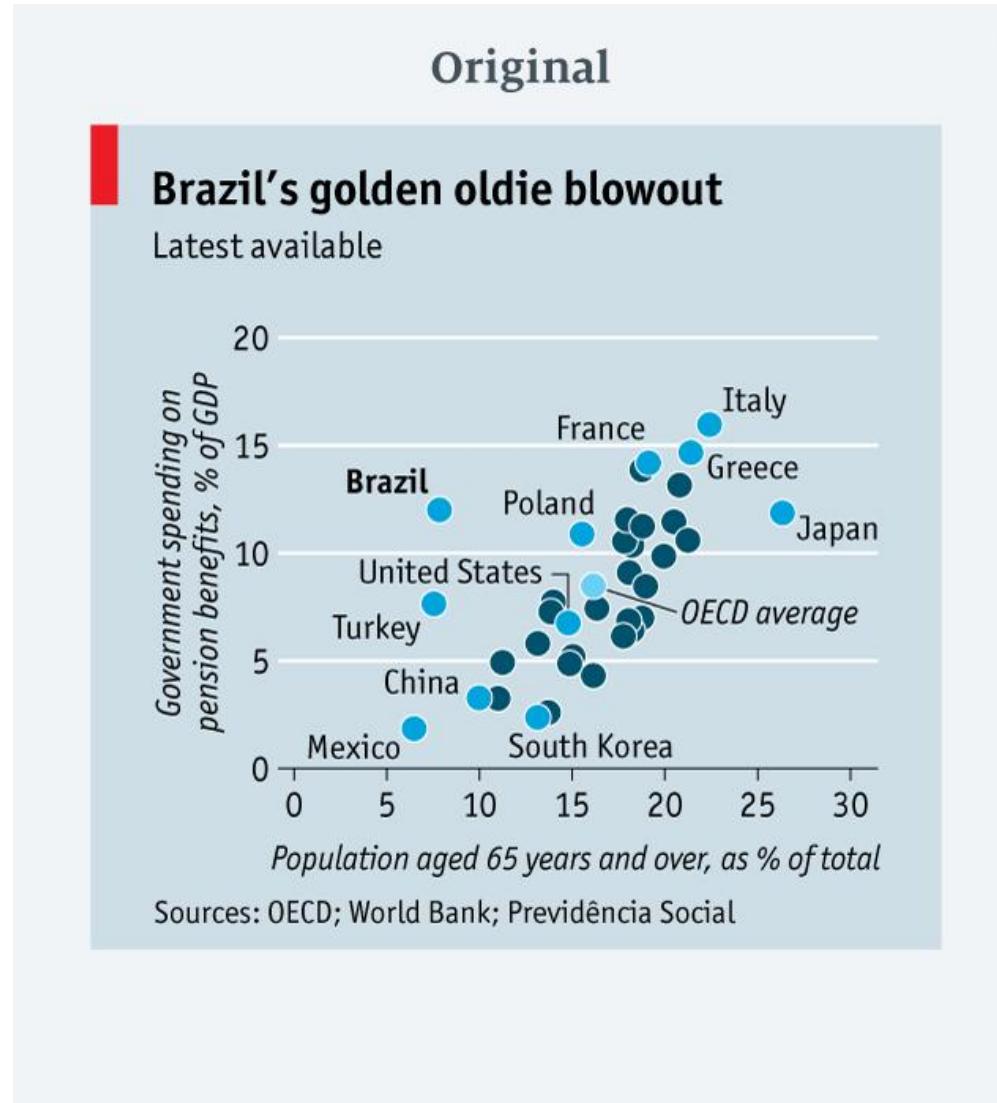
Bad graphs



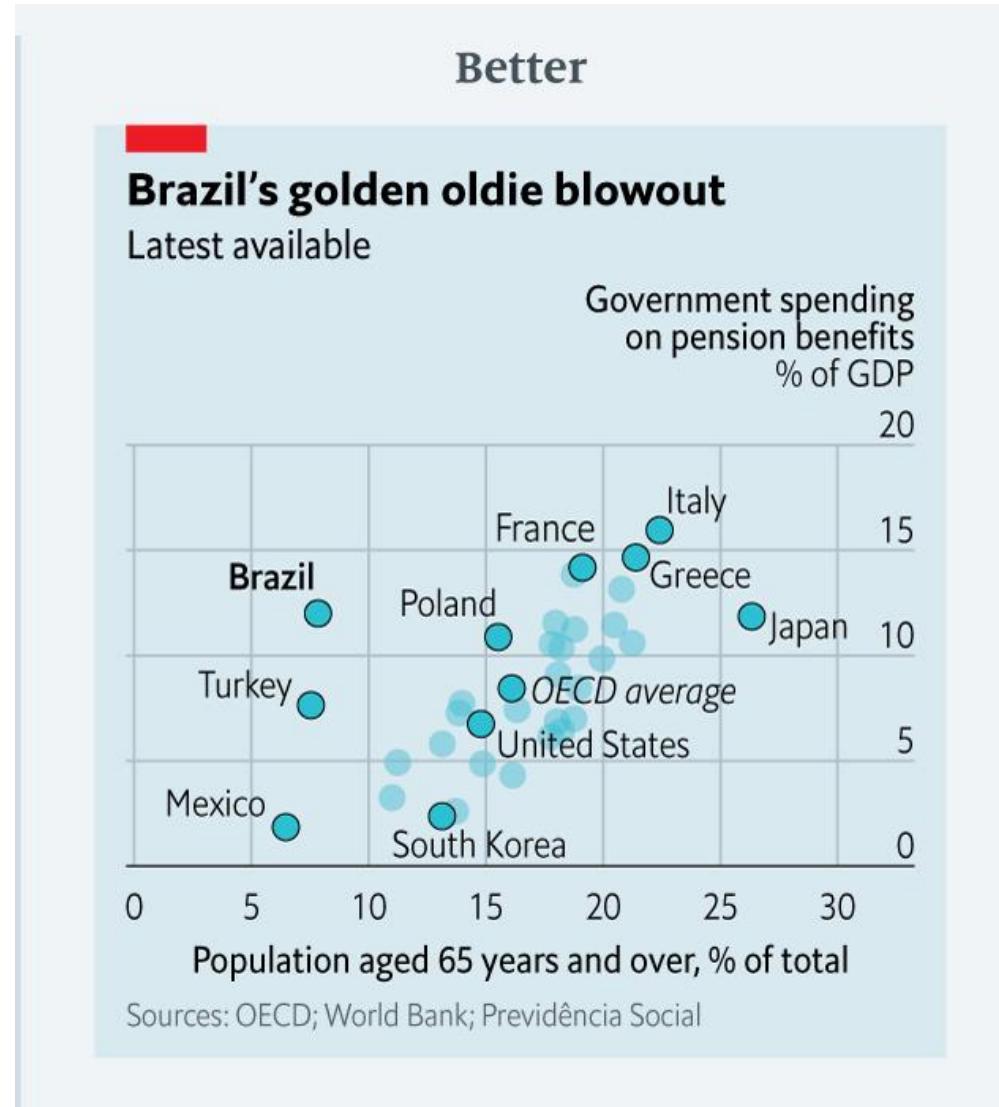
Bad graphs



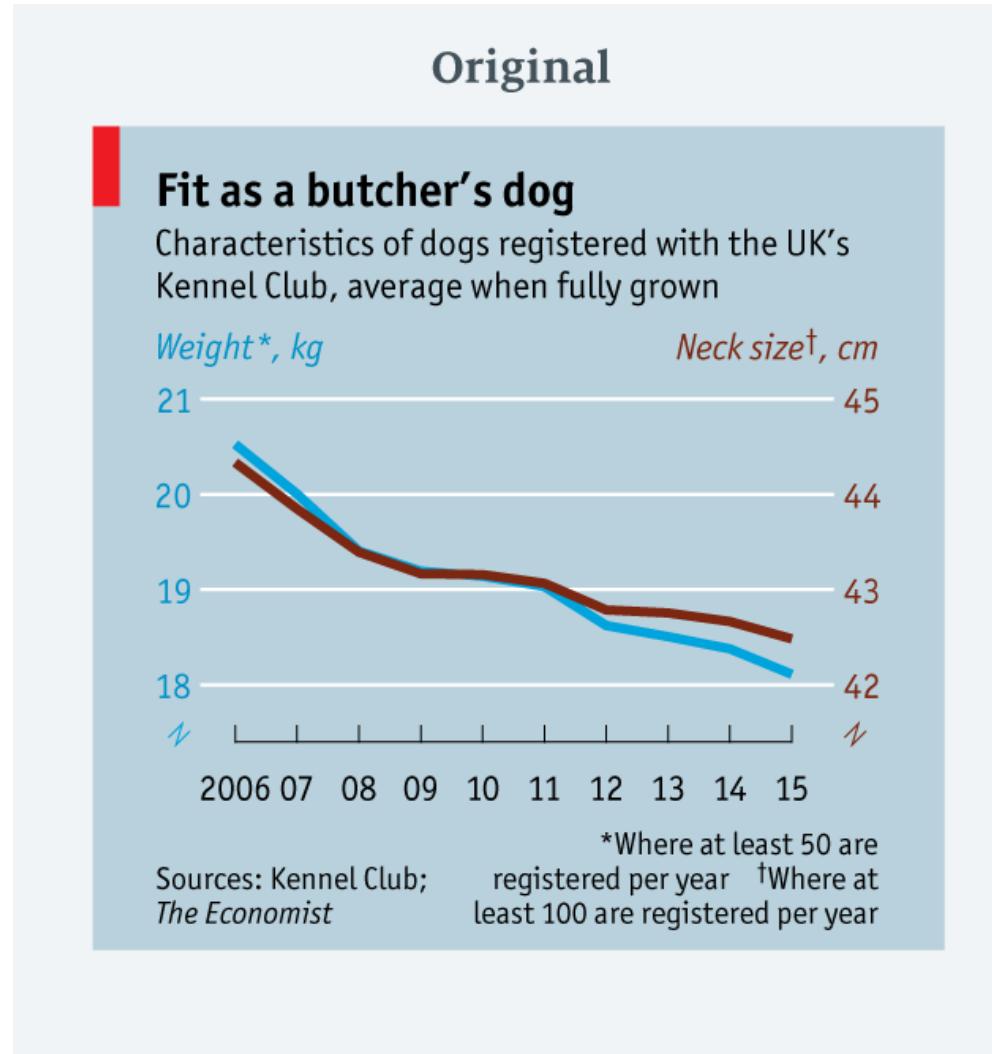
Bad graphs



Bad graphs



Bad graphs



Bad graphs

