



**GHENT
UNIVERSITY**



SSHT! DATA VISUALIZATION — FEBRUARY 11, 2020

DATA VISUALIZATION

Prof. dr. Dries Benoit — dries.benoit@ugent.be

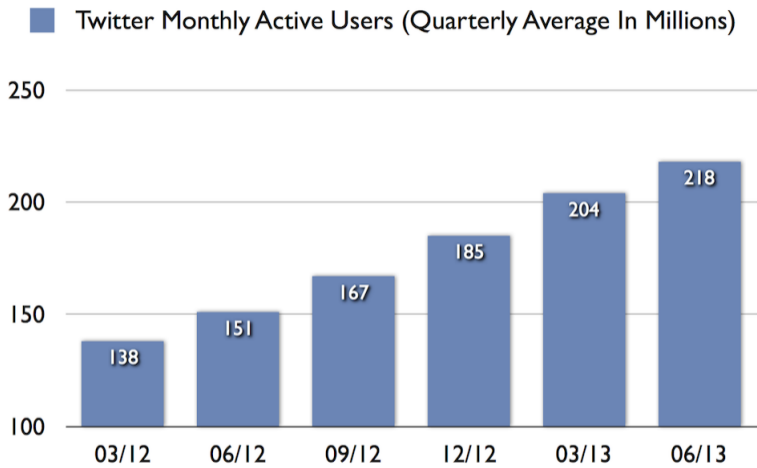


OVERVIEW

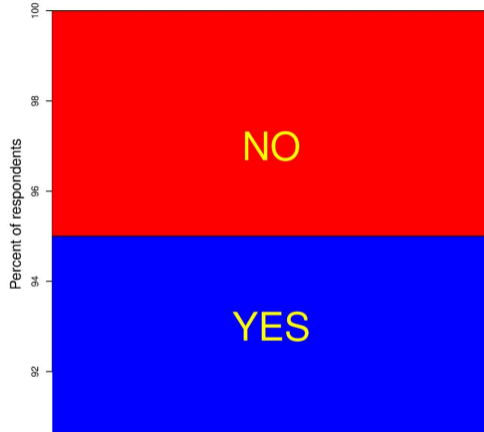
- 1 MOTIVATION
- 2 THEORY
- 3 DON'T'S
- 4 EXERCISE
- 5 TAKE-AWAYS

MOTIVATION



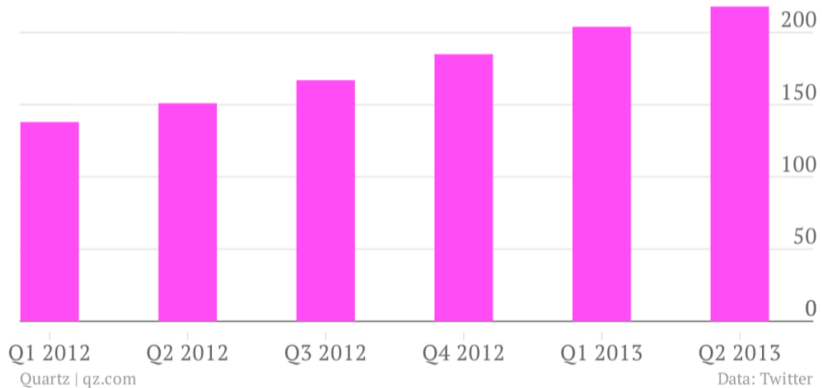


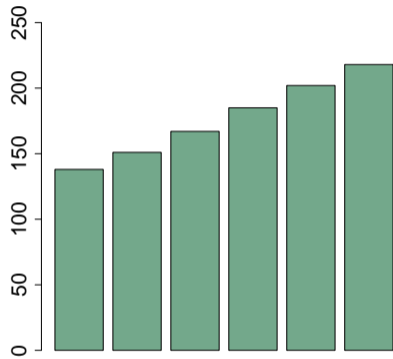
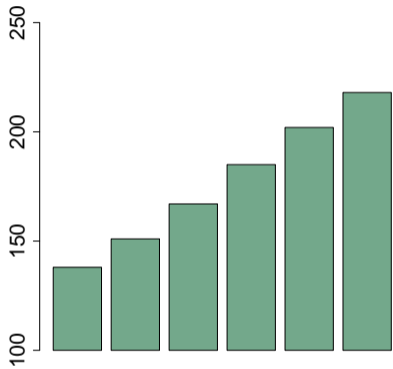
Is truncating the Y-axis dishonest?



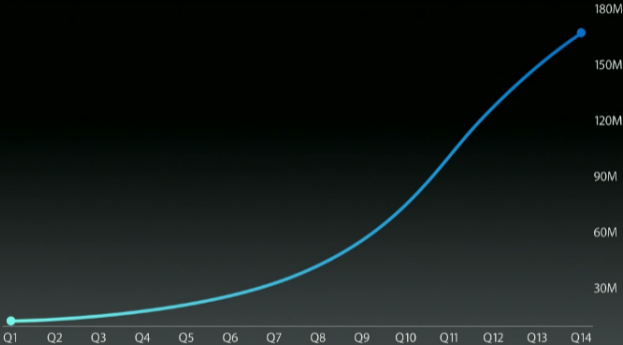
Twitter monthly active users

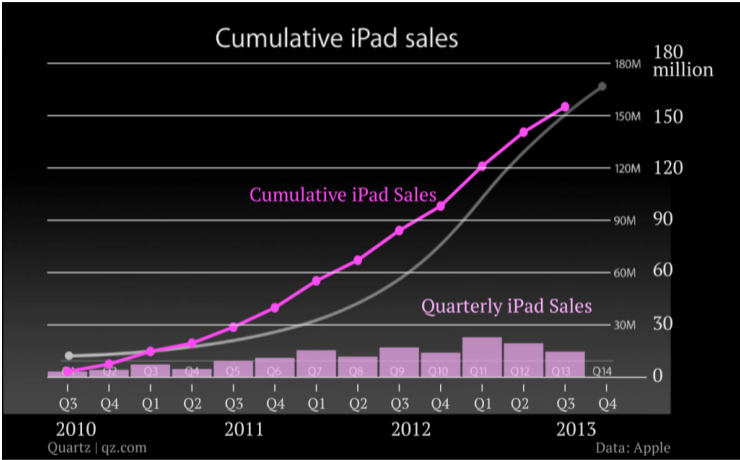
250 million users





Cumulative iPad sales





SOME DEFINITIONS

Definition

Visualization is the power or process of forming a mental picture or vision of something not actually present to the sight; a picture thus formed.

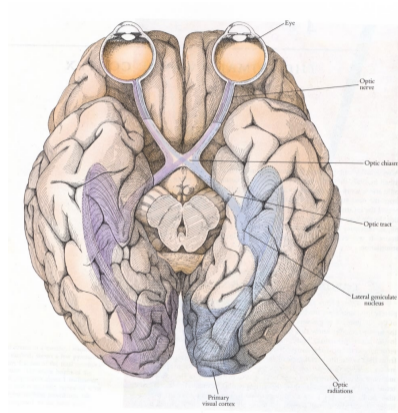
Definition

Visualization is the construction of images which represent important aspects of some situation or process.

- Plot
- Graph

- Diagram
- Picture

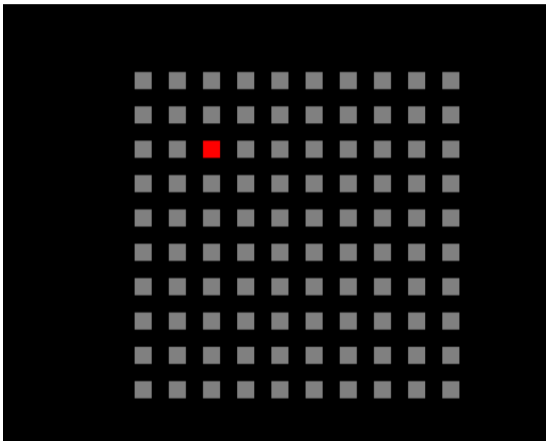
WHY VISUALIZE?



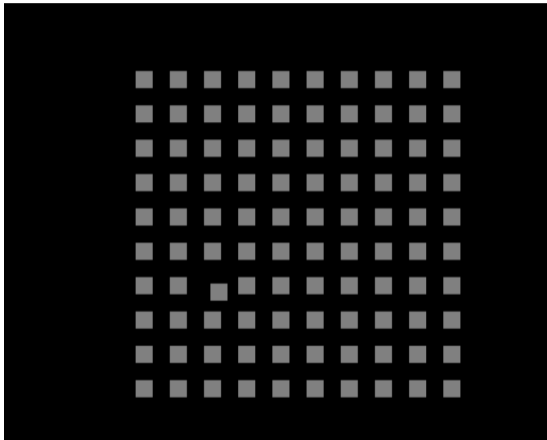
EXAMPLE

Partij	Verkiezingen 06/10	Peiling 9/12	Peiling 5/13	Peiling 10/13
N-VA	28,2	36,3	32,1	27,9
CD&V	17,6	18,5	17,4	19
SP.A	14,9	14,5	14,7	13,4
VB	12,6	9,5	10,6	10,6
VLD	14	10,7	10,1	13,8
Groen!	7,1	7,9	9,5	10
LDD	3,7	1,3	0,4	1,3
PVDA+	5	5	2,5	3,9
Andere	2	1,2	2,6	0,1

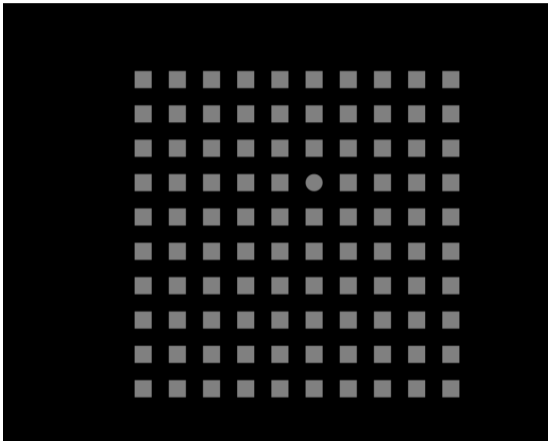
EXAMPLE



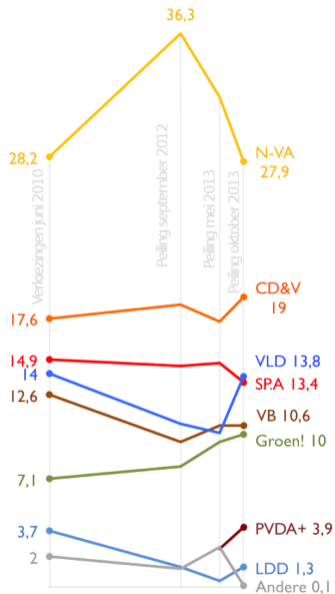
EXAMPLE



EXAMPLE



EXAMPLE



HOW CAN VISUALIZATION BE USEFUL?

- Communication

Visualisation provides a quick way to communicate a very rich message.

- Discovery

Visualisation provides a way of displaying a large amount of data so we can uncover new facts and relationships.

- Insight

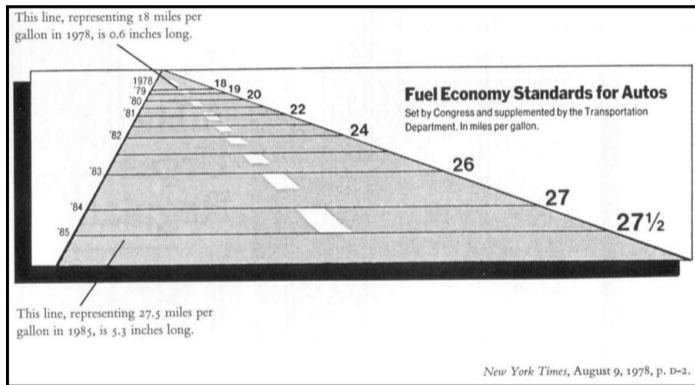
Visualisation provides a way to obtain better insight into things we already know.

THEORY



TUFTE'S PRINCIPLES

Data integrity



TUFTE'S PRINCIPLES

Lie factor:

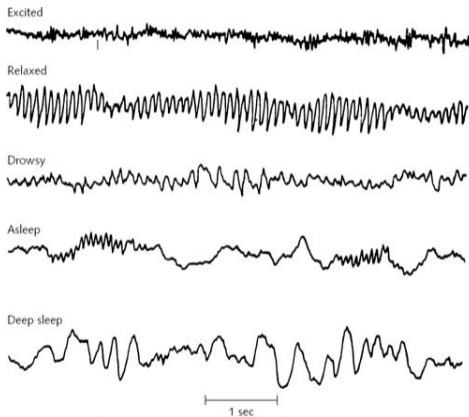
$$\text{lie factor} = \frac{\text{size of effect shown in graph}}{\text{size of effect in data}}$$

with

$$\text{size of effect} = \frac{|\text{second value} - \text{first value}|}{\text{first value}}$$

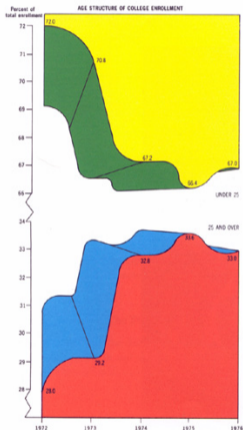
TUFTE'S PRINCIPLES

Data-Ink

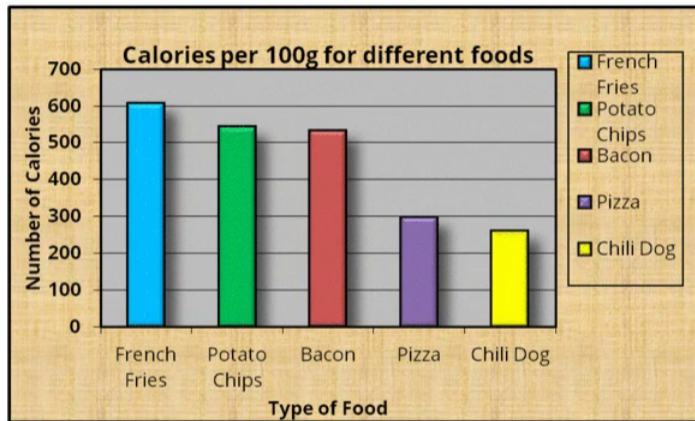


TUFTE'S PRINCIPLES

Chart-junk



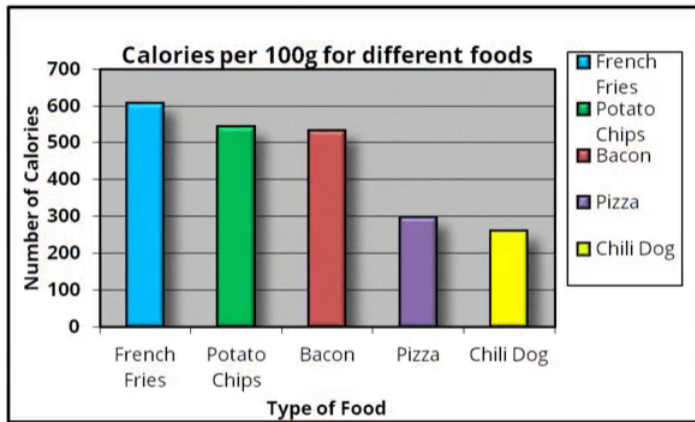
Remove backgrounds



Created by Darkhorse Analytics

www.darkhorseanalytics.com

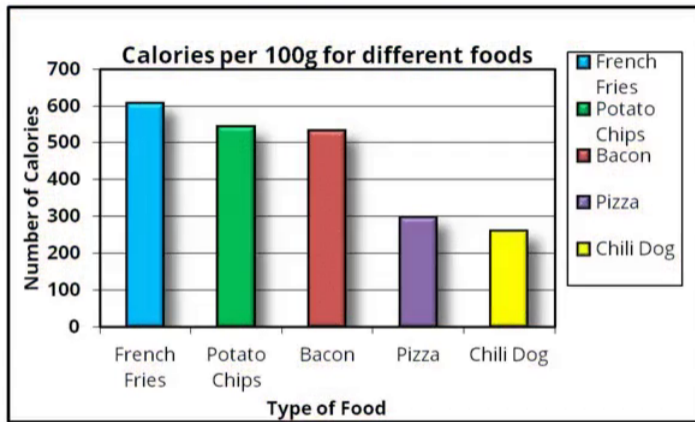
Remove backgrounds



Created by Darkhorse Analytics

www.darkhorseanalytics.com

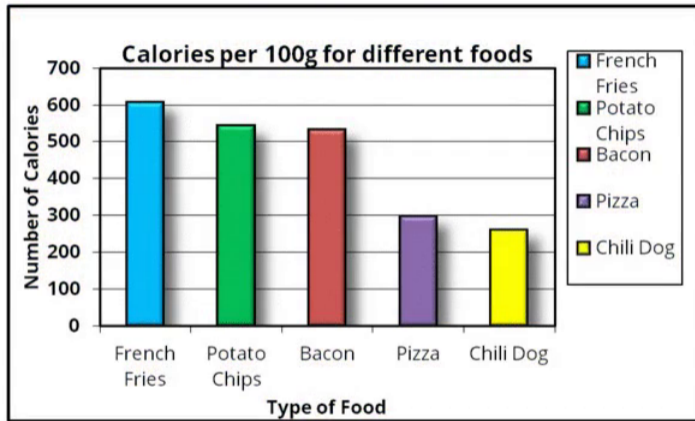
Remove backgrounds



Created by Darkhorse Analytics

www.darkhorseanalytics.com

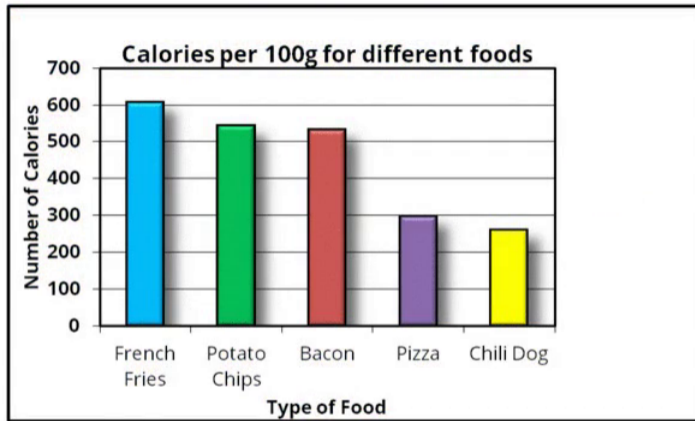
Remove redundant labels



Created by Darkhorse Analytics

www.darkhorseanalytics.com

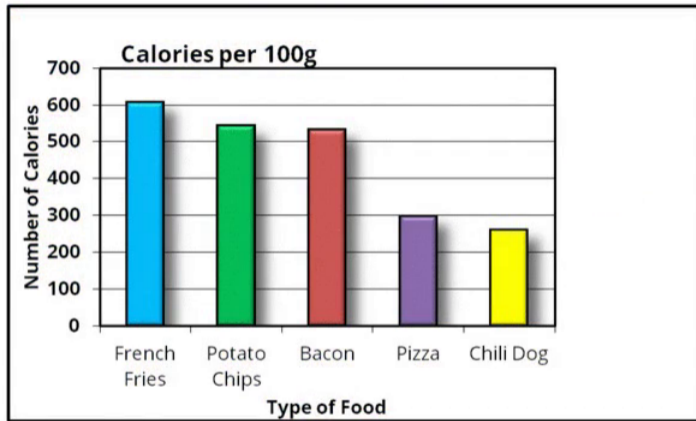
Remove redundant labels



Created by Darkhorse Analytics

www.darkhorseanalytics.com

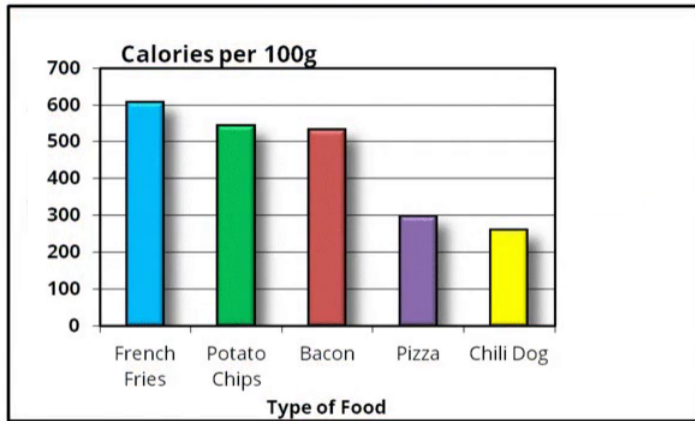
Remove redundant labels



Created by Darkhorse Analytics

www.darkhorseanalytics.com

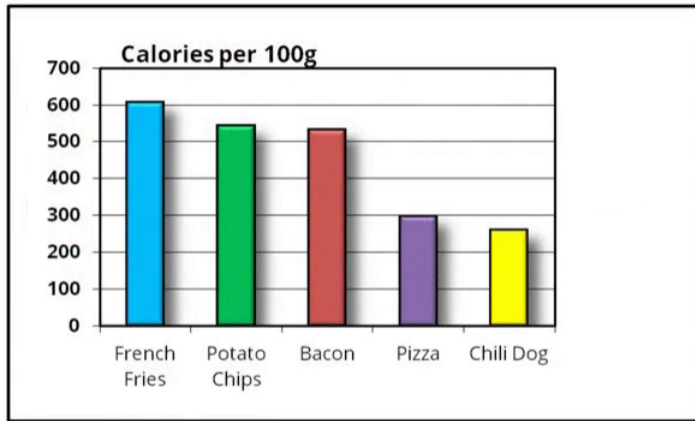
Remove redundant labels



Created by Darkhorse Analytics

www.darkhorseanalytics.com

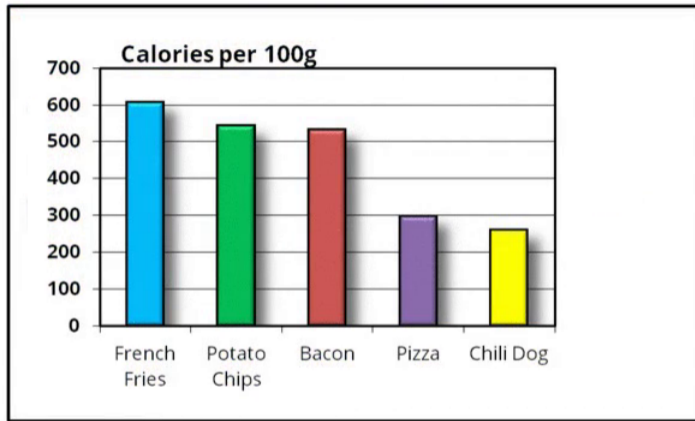
Remove redundant labels



Created by Darkhorse Analytics

www.darkhorseanalytics.com

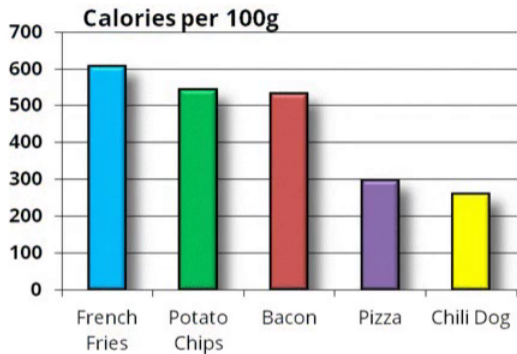
Remove borders



Created by Darkhorse Analytics

www.darkhorseanalytics.com

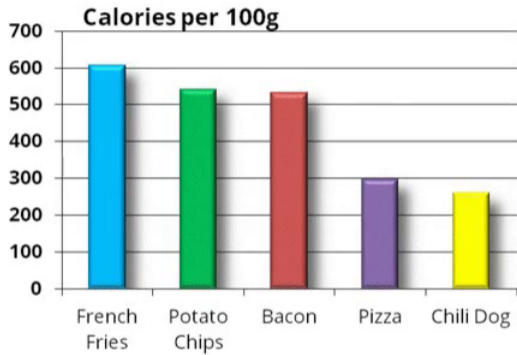
Remove borders



Created by Darkhorse Analytics

www.darkhorseanalytics.com

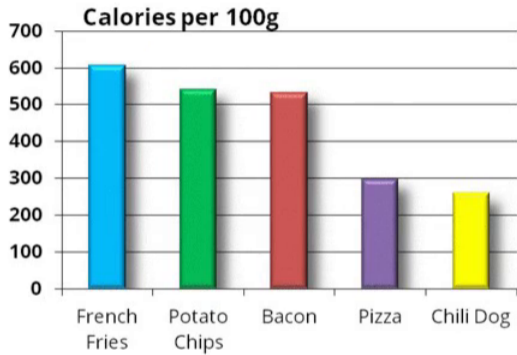
Remove borders



Created by Darkhorse Analytics

www.darkhorseanalytics.com

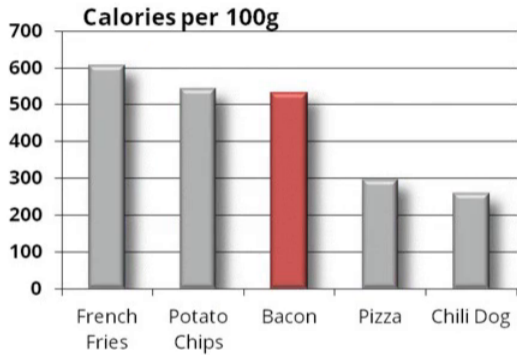
Reduce colors



Created by Darkhorse Analytics

www.darkhorseanalytics.com

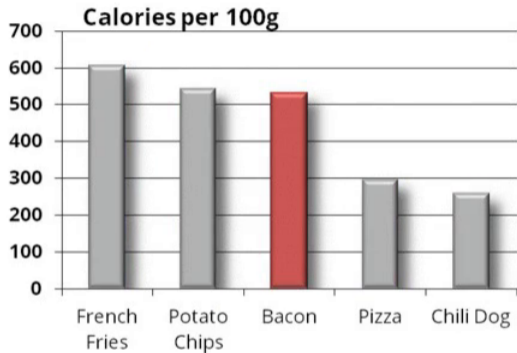
Reduce colors



Created by Darkhorse Analytics

www.darkhorseanalytics.com

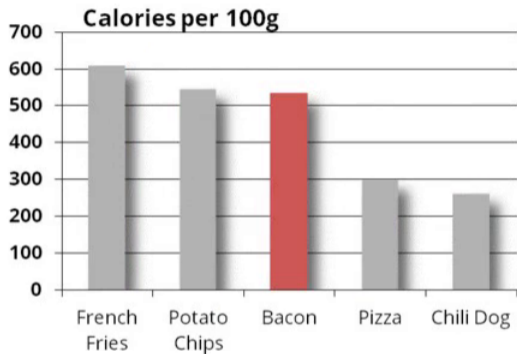
Remove special effects



Created by Darkhorse Analytics

www.darkhorseanalytics.com

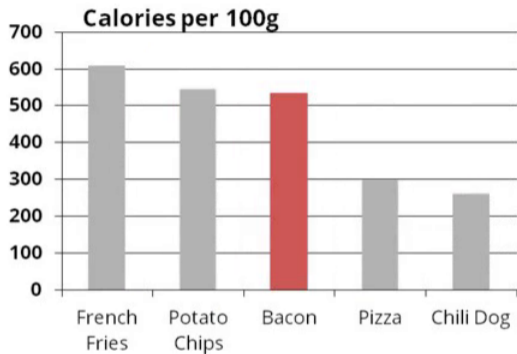
Remove special effects



Created by Darkhorse Analytics

www.darkhorseanalytics.com

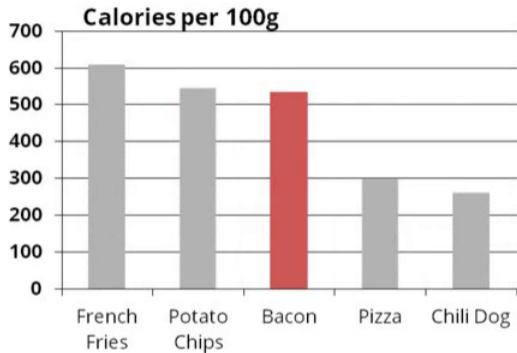
Remove special effects



Created by Darkhorse Analytics

www.darkhorseanalytics.com

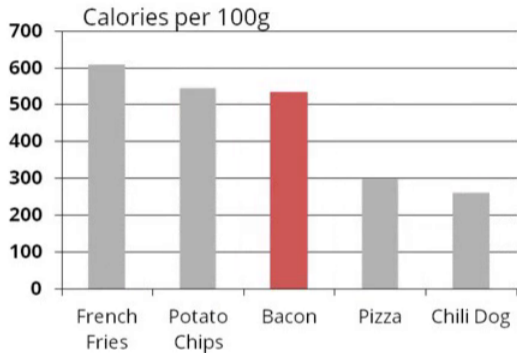
Remove bolding



Created by Darkhorse Analytics

www.darkhorseanalytics.com

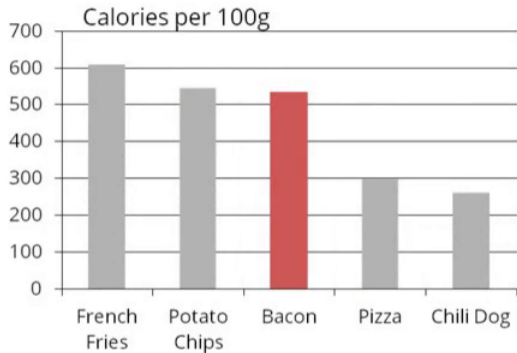
Remove bolding



Created by Darkhorse Analytics

www.darkhorseanalytics.com

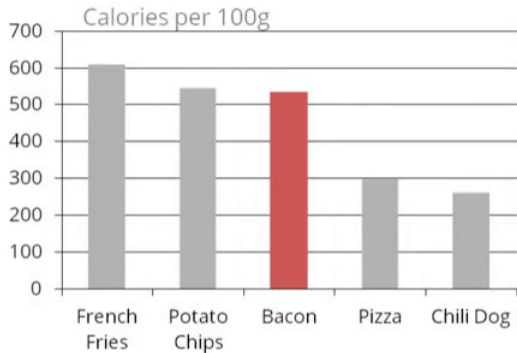
Lighten labels



Created by Darkhorse Analytics

www.darkhorseanalytics.com

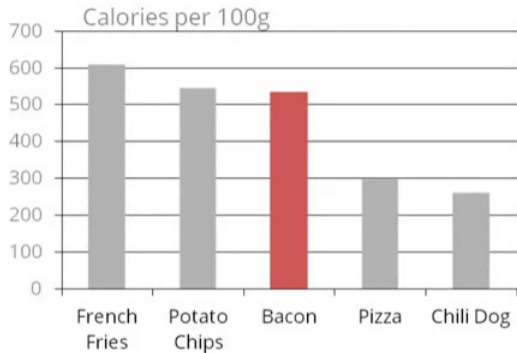
Lighten labels



Created by Darkhorse Analytics

www.darkhorseanalytics.com

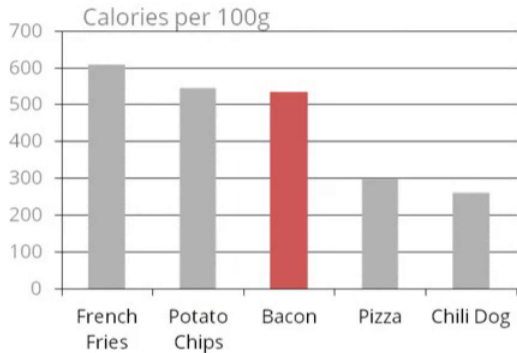
Lighten labels



Created by Darkhorse Analytics

www.darkhorseanalytics.com

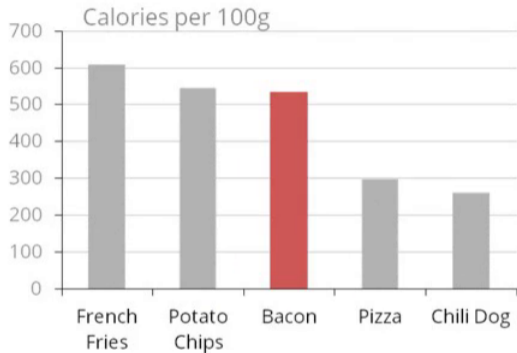
Lighten lines



Created by Darkhorse Analytics

www.darkhorseanalytics.com

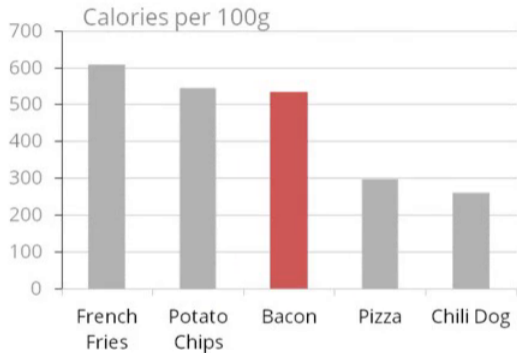
Lighten lines



Created by Darkhorse Analytics

www.darkhorseanalytics.com

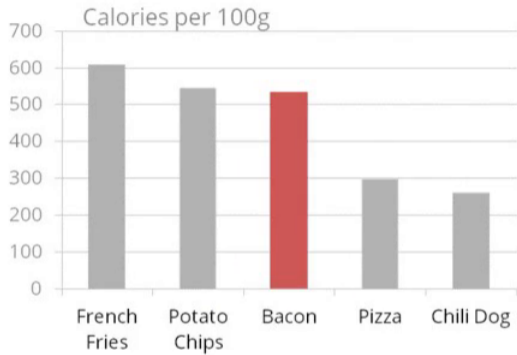
Lighten lines



Created by Darkhorse Analytics

www.darkhorseanalytics.com

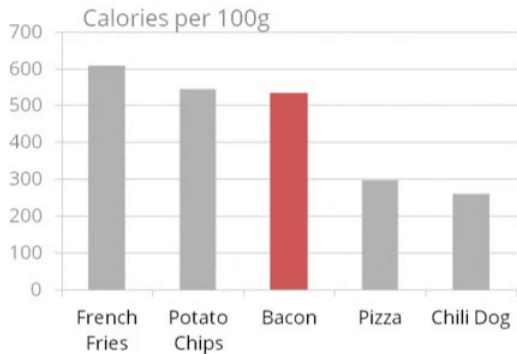
Lighten lines



Created by Darkhorse Analytics

www.darkhorseanalytics.com

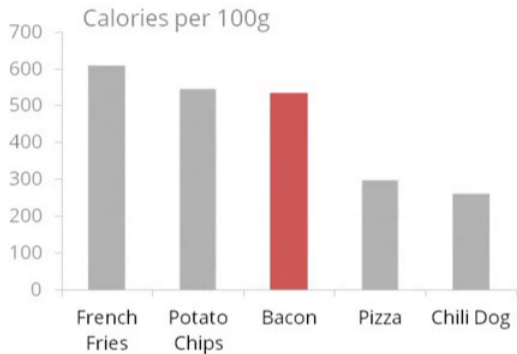
Or remove lines



Created by Darkhorse Analytics

www.darkhorseanalytics.com

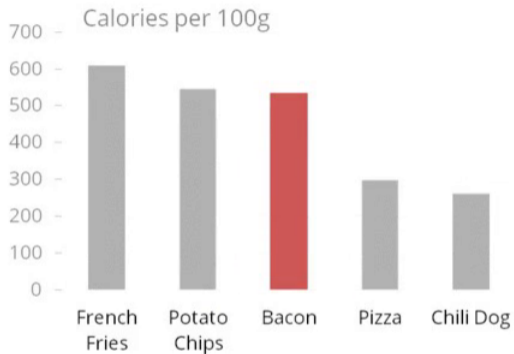
Or remove lines



Created by Darkhorse Analytics

www.darkhorseanalytics.com

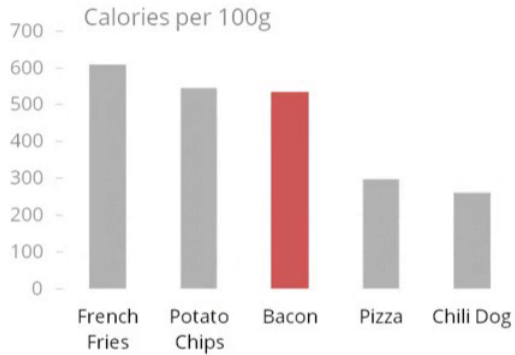
Or remove lines



Created by Darkhorse Analytics

www.darkhorseanalytics.com

Direct label

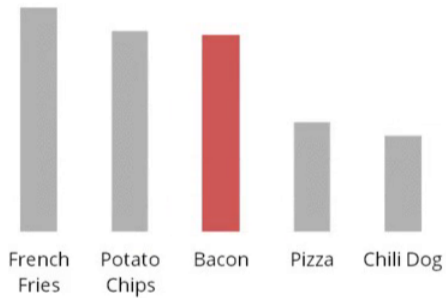


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Direct label

Calories per 100g

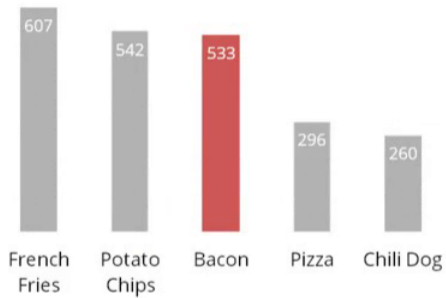


Created by Darkhorse Analytics

www.darkhorseanalytics.com

Direct label

Calories per 100g



Created by Darkhorse Analytics

www.darkhorseanalytics.com

Less
is **more**
(effective)

Created by **Darkhorse Analytics**

www.darkhorseanalytics.com

Less
is **more**
(attractive)

Created by **Darkhorse Analytics**

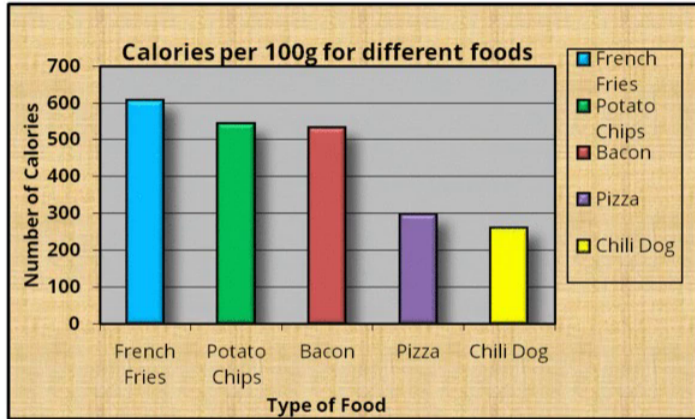
www.darkhorseanalytics.com

Less
is **more**
(impactive)

Created by **Darkhorse Analytics**

www.darkhorseanalytics.com

Before

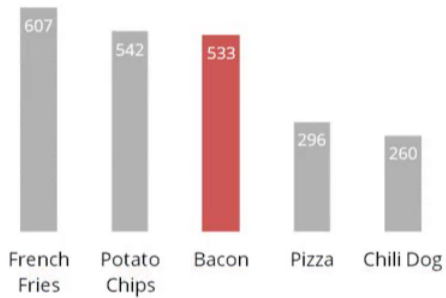


Created by Darkhorse Analytics

www.darkhorseanalytics.com

After

Calories per 100g

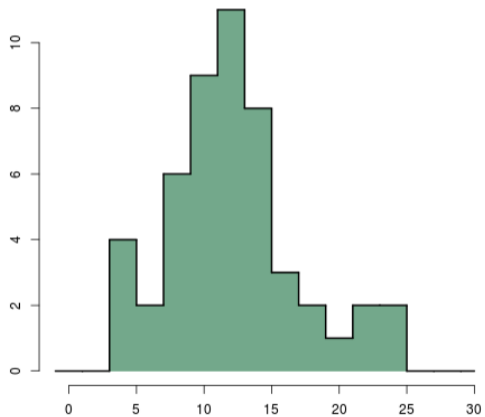


Created by Darkhorse Analytics

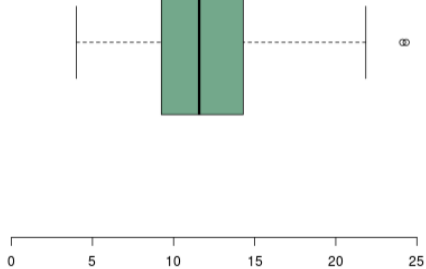
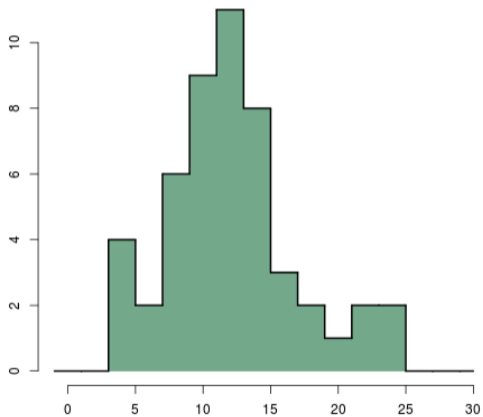
www.darkhorseanalytics.com

VISUAL SUMMARY OF ONE VARIABLE

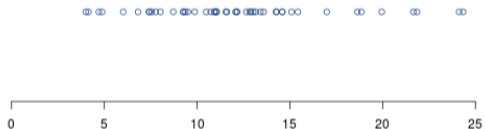
VISUAL SUMMARY OF ONE VARIABLE



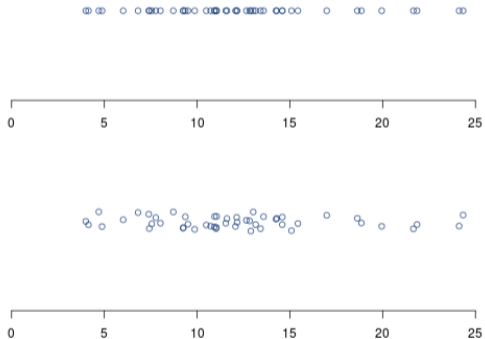
VISUAL SUMMARY OF ONE VARIABLE



VISUAL SUMMARY OF ONE VARIABLE

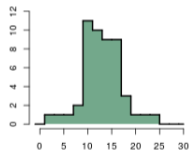
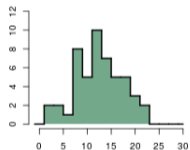
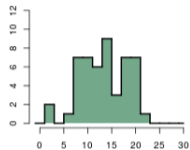
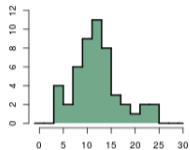


VISUAL SUMMARY OF ONE VARIABLE

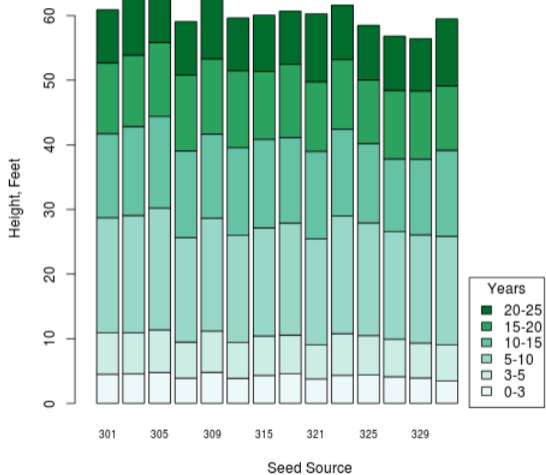


COMPARE DISTRIBUTIONS

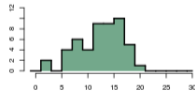
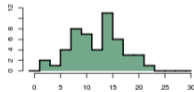
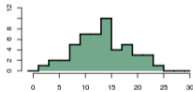
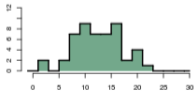
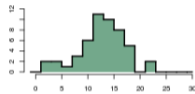
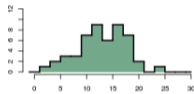
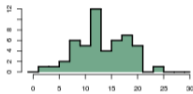
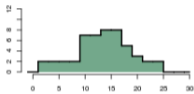
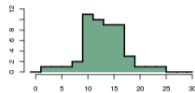
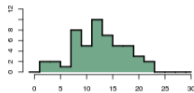
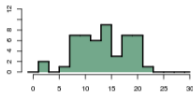
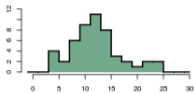
COMPARE DISTRIBUTIONS



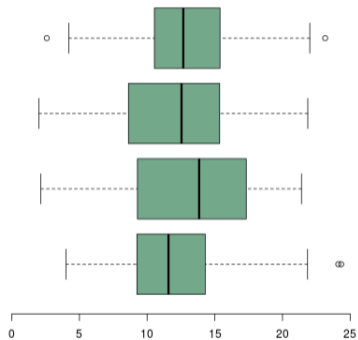
COMPARE DISTRIBUTIONS



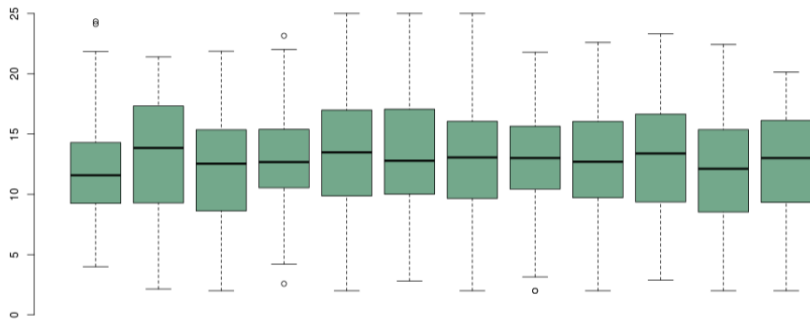
COMPARE DISTRIBUTIONS



COMPARE DISTRIBUTIONS

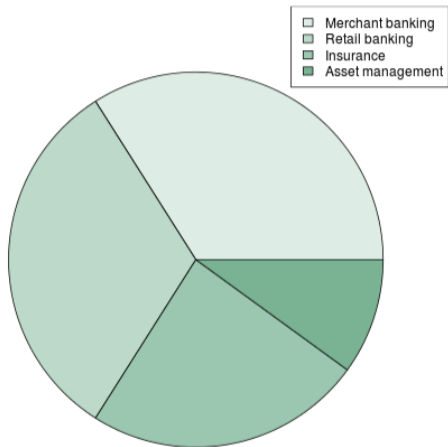


COMPARE DISTRIBUTIONS

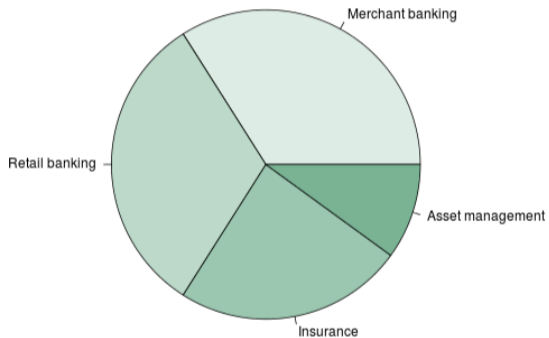
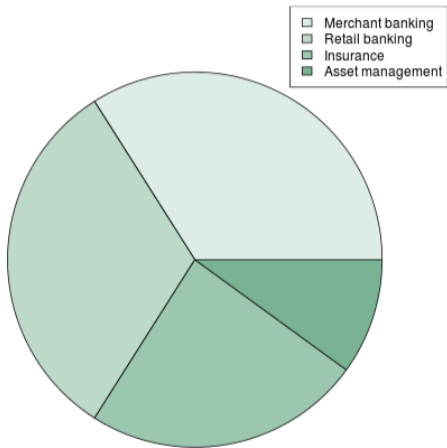


COMPARE SINGLE VALUES

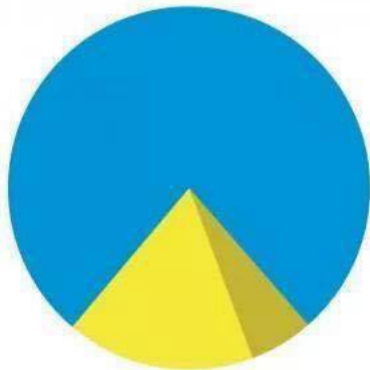
COMPARE SINGLE VALUES






COMPARE SINGLE VALUES



COMPARE SINGLE VALUES

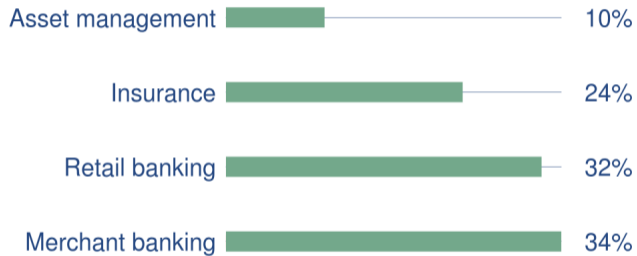


-  Sky
-  Sunny side of pyramid
-  Shady side of pyramid

COMPARE SINGLE VALUES



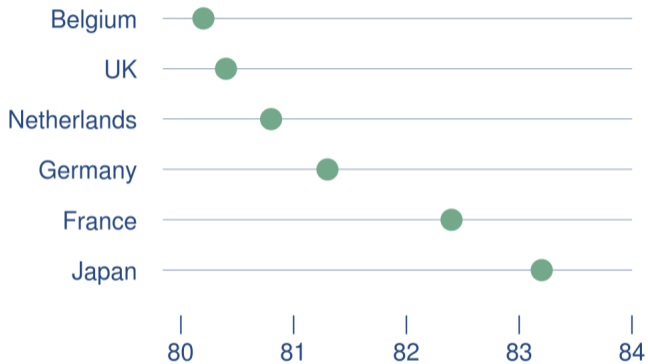
COMPARE SINGLE VALUES



COMPARE SINGLE VALUES

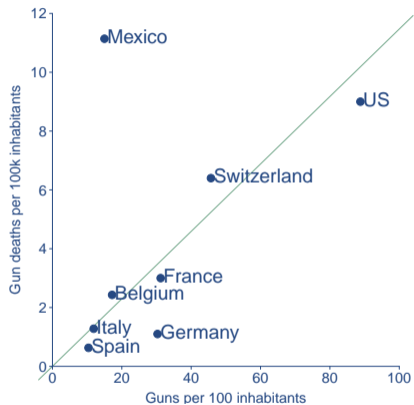


COMPARE SINGLE VALUES



COMPARE 2 CONTINUOUS VARIABLES

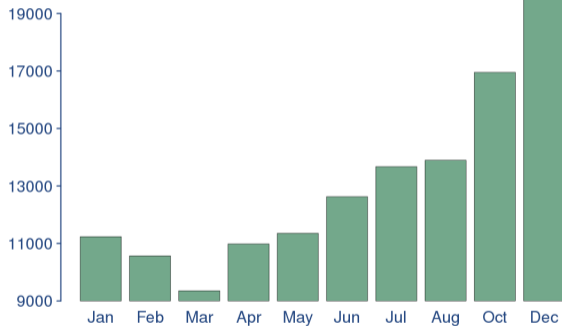
COMPARE 2 CONTINUOUS VARIABLES



SHOWING EVOLUTION

SHOWING EVOLUTION

Revenue 2013

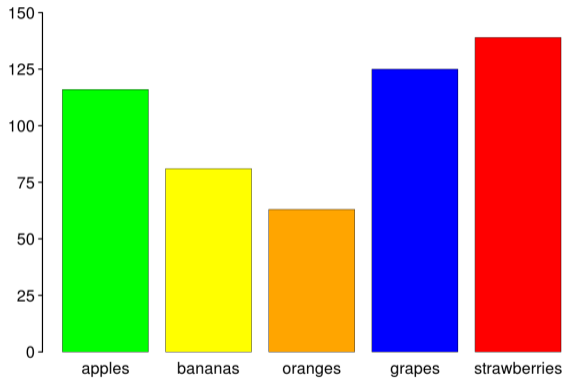


SHOWING EVOLUTION

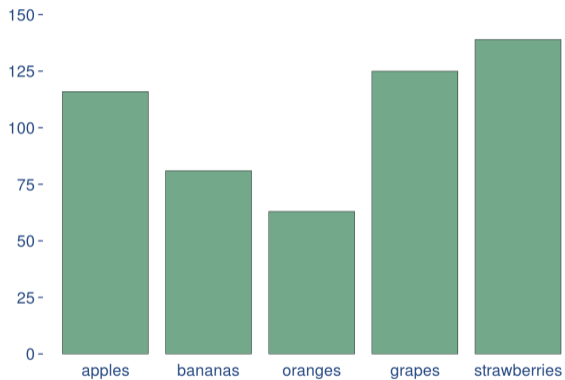


DON'T'S

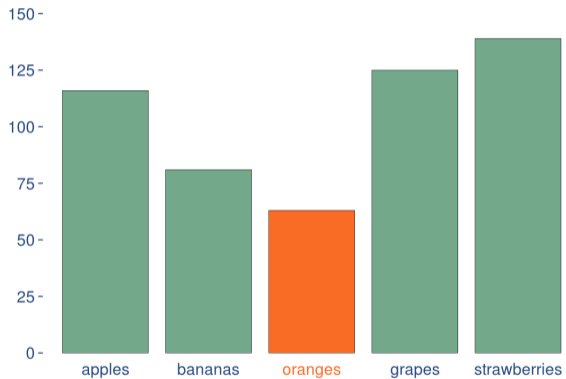
COLORS



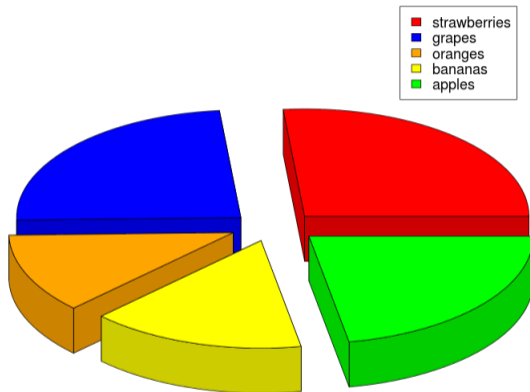
COLORS



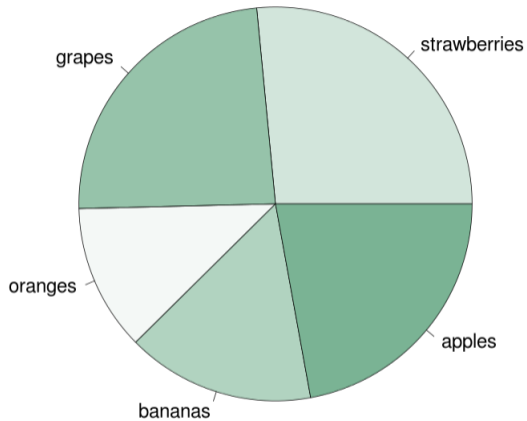
COLORS



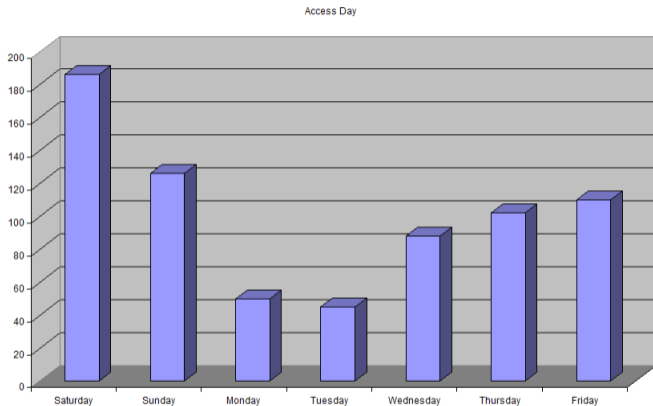
3D



3D

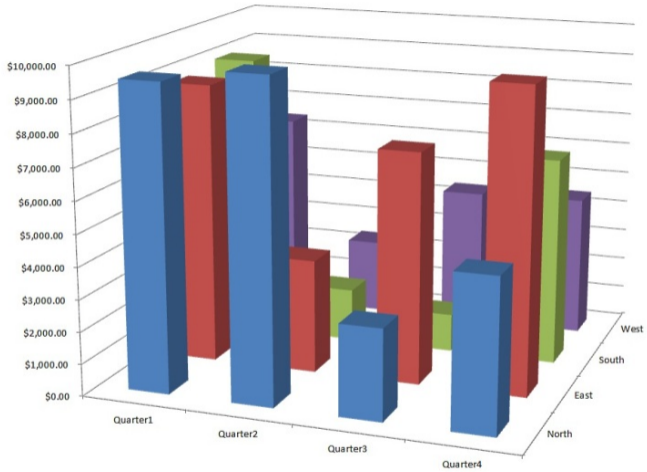


3D

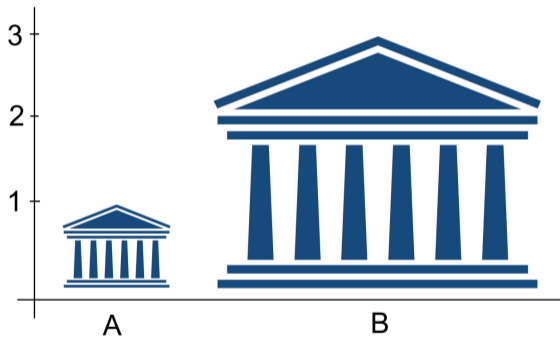


3D

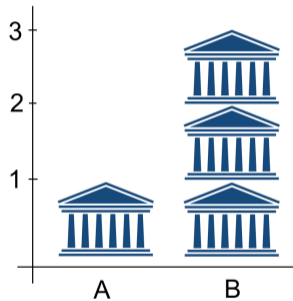
Quarterly Report by Region



SCALED PICTURES



SCALED PICTURES



SCALED AXES



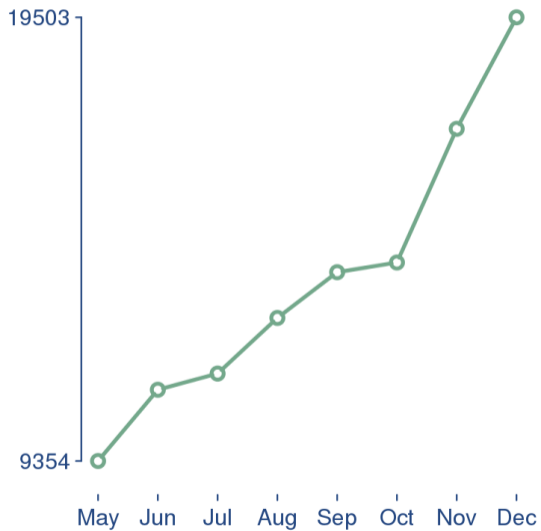
SCALED AXES



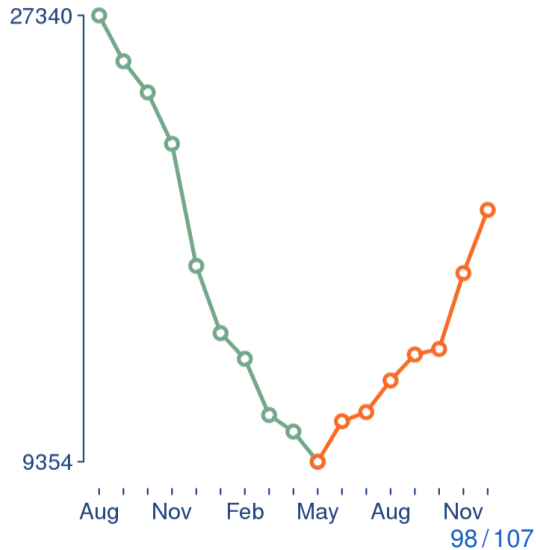
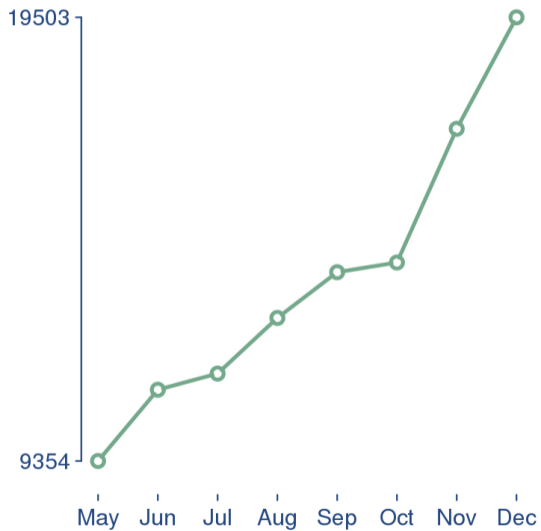
SCALED AXES



IMPROPER EXTRACTION



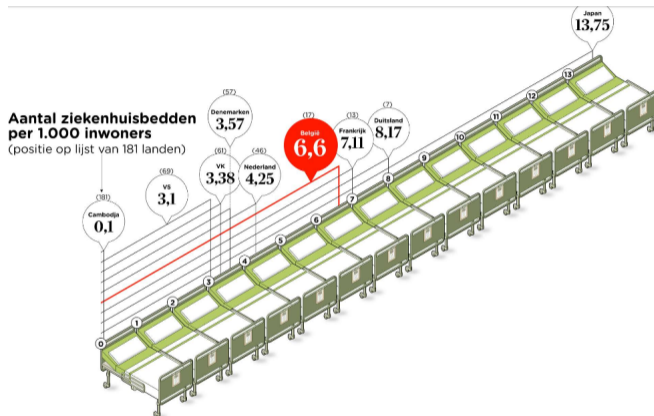
IMPROPER EXTRACTION



EXERCISE



EXERCISE



EXERCISE



EXERCISE



TAKE-AWAYS



TAKE-AWAYS

Do's

- Summarize single variable → histogram or boxplot
- Compare distributions → histogram or boxplot
- Compare single value → barchart or dotchart
- Compare 2 continuous variables → scatter plot
- Compare > 2 variables → small multiples
- Show evolution → linechart
- Scale your axes “fair”

TAKE-AWAYS

Don't's (or avoid's)

- Avoid using a legend → explicit labeling
- Avoid using many colors → color important datapoints
- Avoid using piecharts → use barcharts or dotcharts
- Avoid using 3D → no more dimensions than variables
- Avoid using surface or volume to compare values → use length
- Avoid “unfair” extraction of data

REFERENCES

- Book: Kabacoff, Data Visualization with R
<https://rkabacoff.github.io/datavis/>
- Gallery: The R Graph Gallery
<https://www.r-graph-gallery.com/>
- Other software: Python (matplotlib), D3.js, Tableau, PowerBI, ...

REFERENCES

- Book: Jean-Luc Doumont (2009) Trees, maps and theorems.
- Course: Ross Ihaka - Information Visualization
<https://www.stat.auckland.ac.nz/~ihaka/courses/120/>
- Website: Maarten Lambrechts - Data Journalist
<http://www.maartenlambrechts.be/>