

# Good and Bad Practices for Data Vis

Álvaro Figueira, PhD.

MSc in Data Science - Data Visualization



DEPARTAMENTO DE CIÊNCIA DE COMPUTADORES  
FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DO PORTO



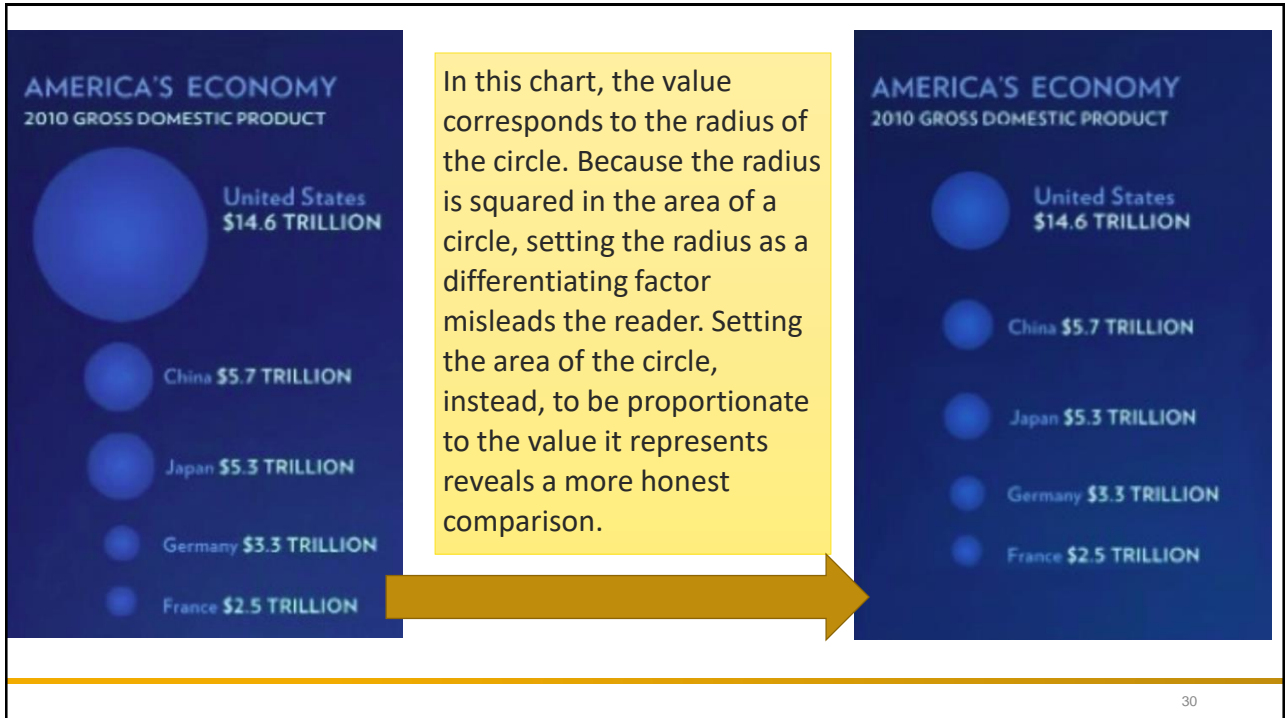
FACULDADE DE CIÊNCIAS  
UNIVERSIDADE DO PORTO



28

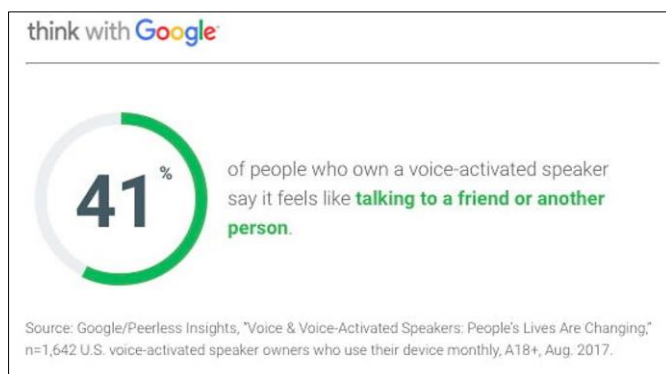


29



30

## Cheating with colors

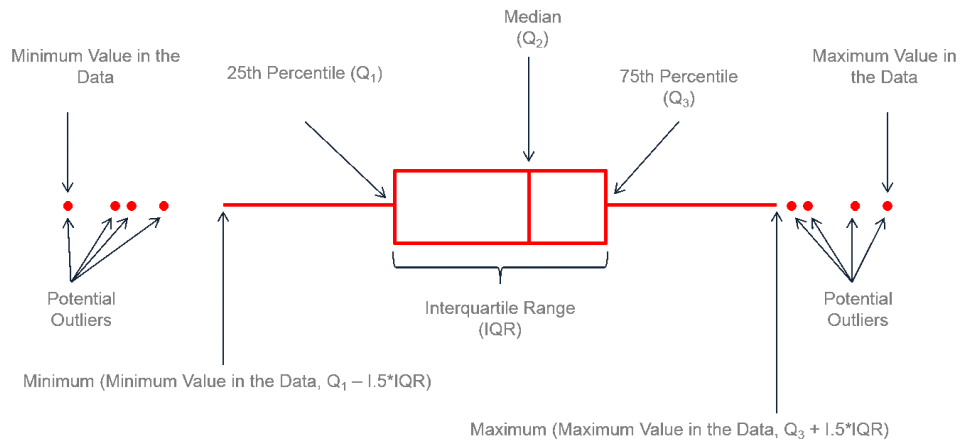


In this graph, the 41% (the minority) is shaded grey, whereas the other 59% is shaded with a vibrant, noticeable green.

31

31

## Boxplots

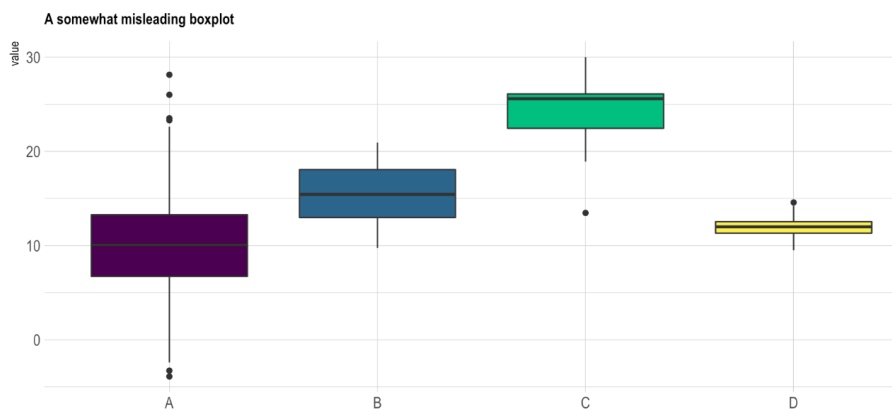


DATE, ADDITIONAL DETAILS (set this text using "Header & Footer")

32

32

## Cheating with boxplots



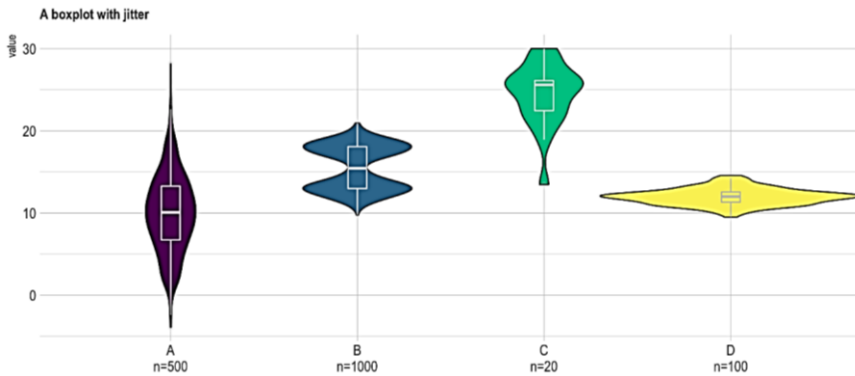
The boxplot assumes that a distribution has only one hump and is some variation of Bell curve with adjustable parameters.

Instead we can use violin plot.

33

33

## Cheating with boxplots



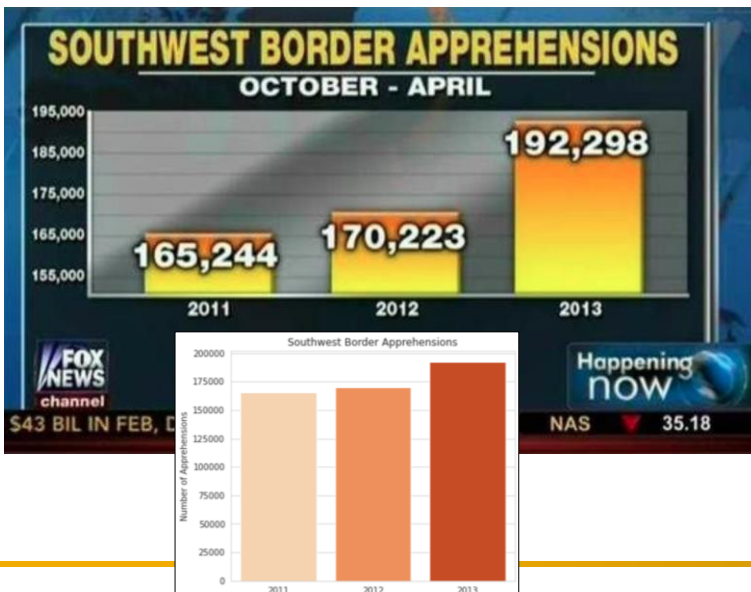
**BoxplotA** is nowhere representative of the large range of its distribution.

**BoxplotB** has two peaks, but the median assumes that a valley is really the peak of a singular-peak distribution.

34

34

## Chopping off the Y-Axis



This chart dramatizes an increase in southwest border apprehensions by not showing the full scale.

The increase is large (a 20,000 increase from 2012 to 2013) but compared to the complete scale, it is visually enormous.

A zero baseline makes any human draw numerical conclusions, such as **two times the number of border apprehensions** in 2013.

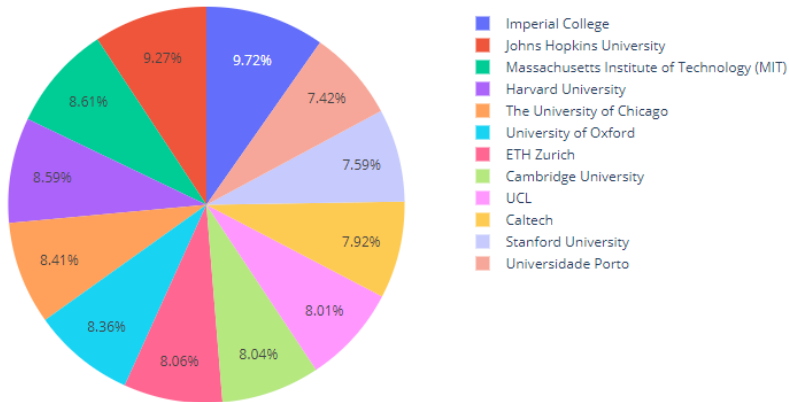
Even though numerically it is made obvious this **is not true**, the bar lengths are suggestive of the opposite.

35

35

## Problems with Pie Charts

Volume de tweets



Pie charts make a correspondence between an intuitive 10 scale into an unintuitive 360 degrees scale.

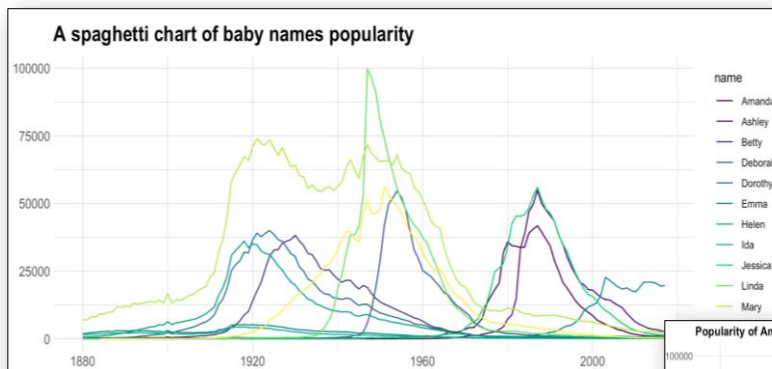
More than 3-4 categories make them confusing to read.

Combine a fundamentally flawed idea with special three-dimensional effects, legends, several pie slices, and exploded slices, and pie charts become often ineffective

36

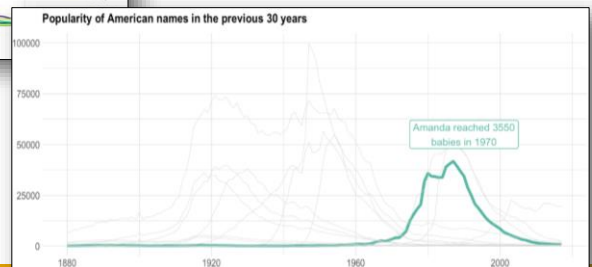
36

## Spaghetti Plots



Spaghetti plots are plots generally with five or more lines. Trying to follow a specific line is difficult, especially when lines of similar colors reach the same point and diverge.

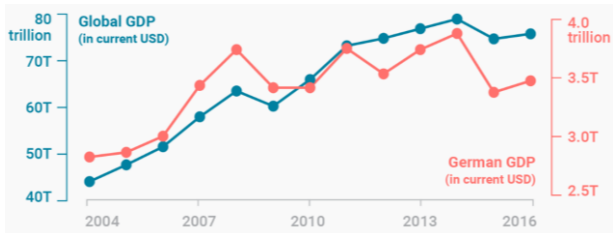
Instead, highlight at most three particular name of interests' evolution throughout time. Not only can they be compared to a general trend in the background, it tells a more clear story.



37

37

## Dual Axis Charts

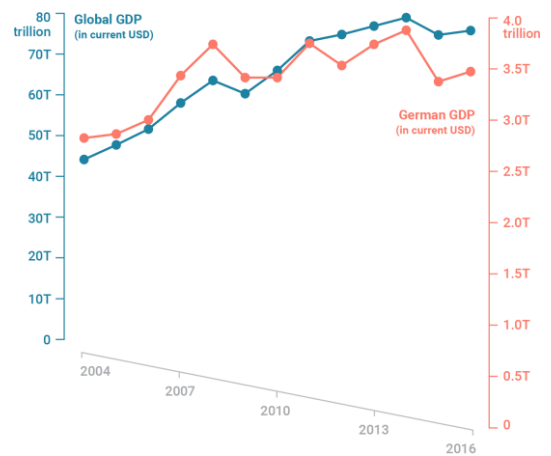


There are several issues with using dual axis plots. One issue with dual axis plots is the ability to manipulate where the y-axis baseline is.

Continuing to the baseline yields a different result.

While the chart looks like the German GDP and the global GDP go up at roughly the same rate (at least until 2014), actually they don't.

The global GDP increased by 80% until 2014; the GDP of Germany by 40%.



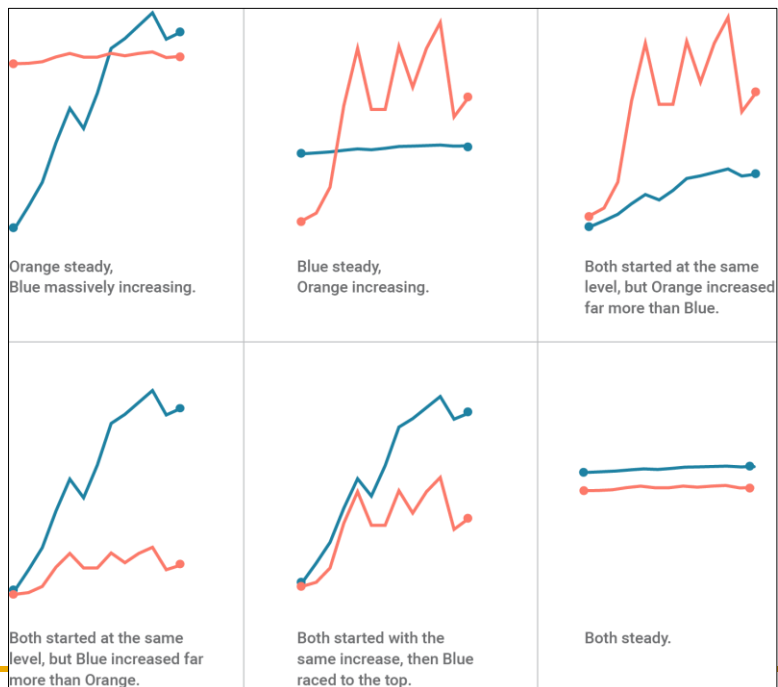
38

38

Tweaking the y-axis zero baseline and the scale is the equivalent of changing the data itself.

Especially with no coordinate lines, y-axis scales mean little and can be used to make any statement about the data you want it to.

This and many other plots are wrong not just because they violate the data's integrity, but because they make the relationship between them a main point, obfuscating the reader's understanding.

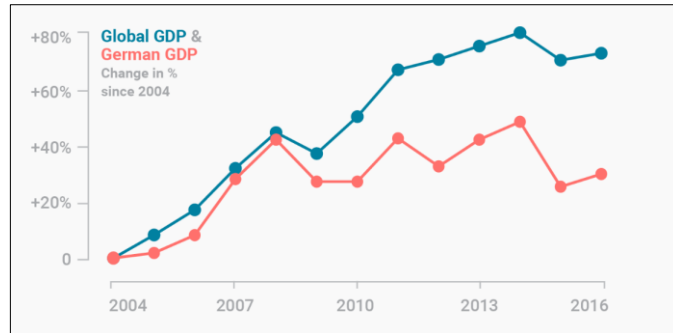
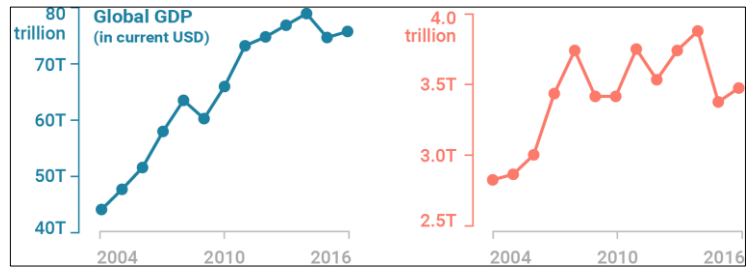


39

## Alternatives...

Two separate charts.

...or charts that show an increase metric on only one scale, such as % increase.

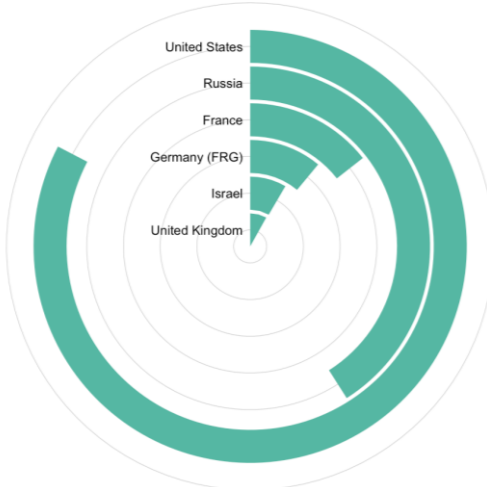


40

40

## Radial Bar Plots

The radial bar plot below shows the quantity of weapons exported by the top six largest exporters in 2017



Radial bar plots are often used because they are eye-catching, unique, and can be animated in a sleek way.

Radial bar plots are not based by length of bar but by, like the pie chart, the angle. The plots are based on a circular degree system, meaning that bars on the outside will always inevitably seem longer than ones closer to the center.

41

41