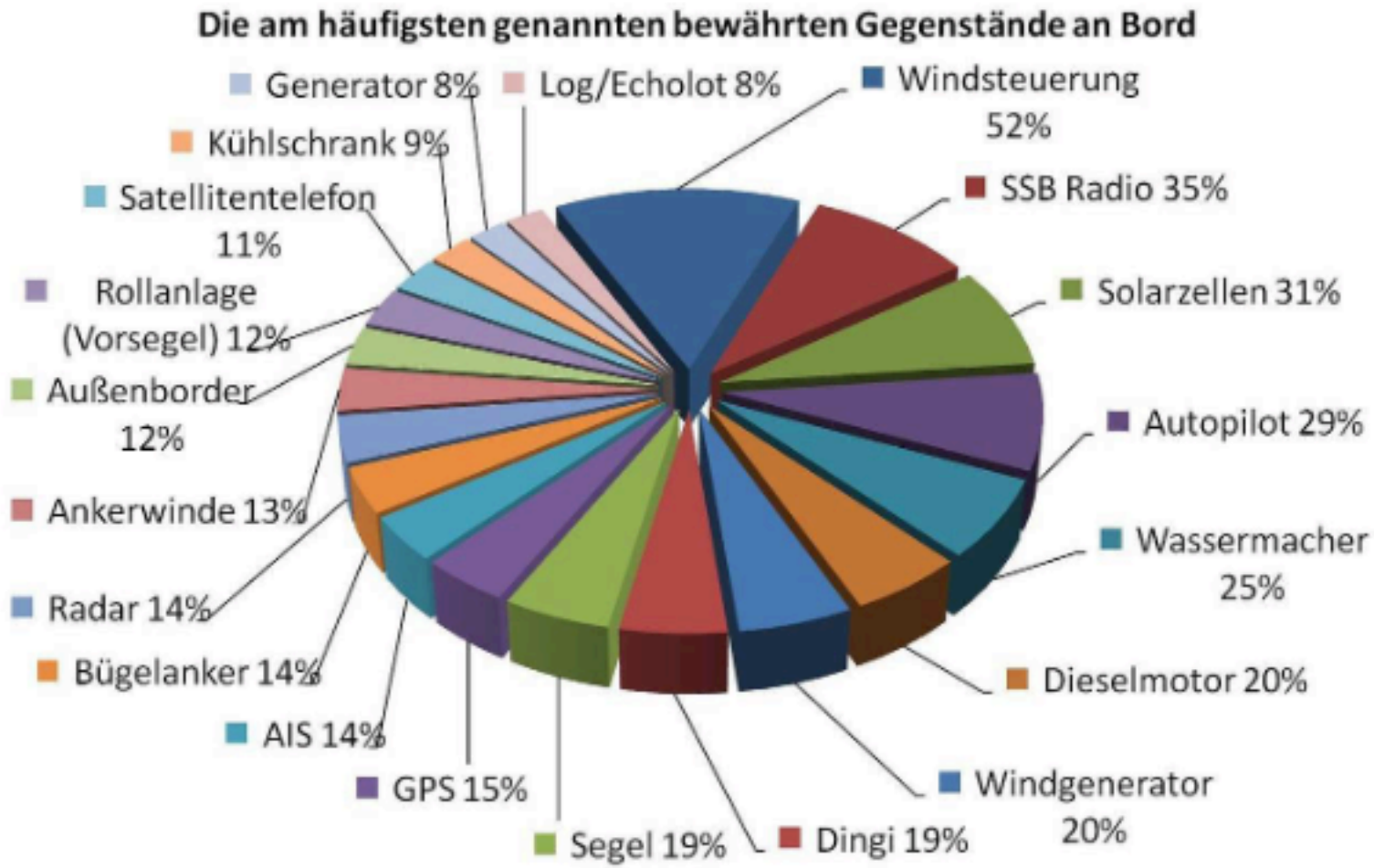
An exploration of dataviz bad practices

An overview of dataviz caveats

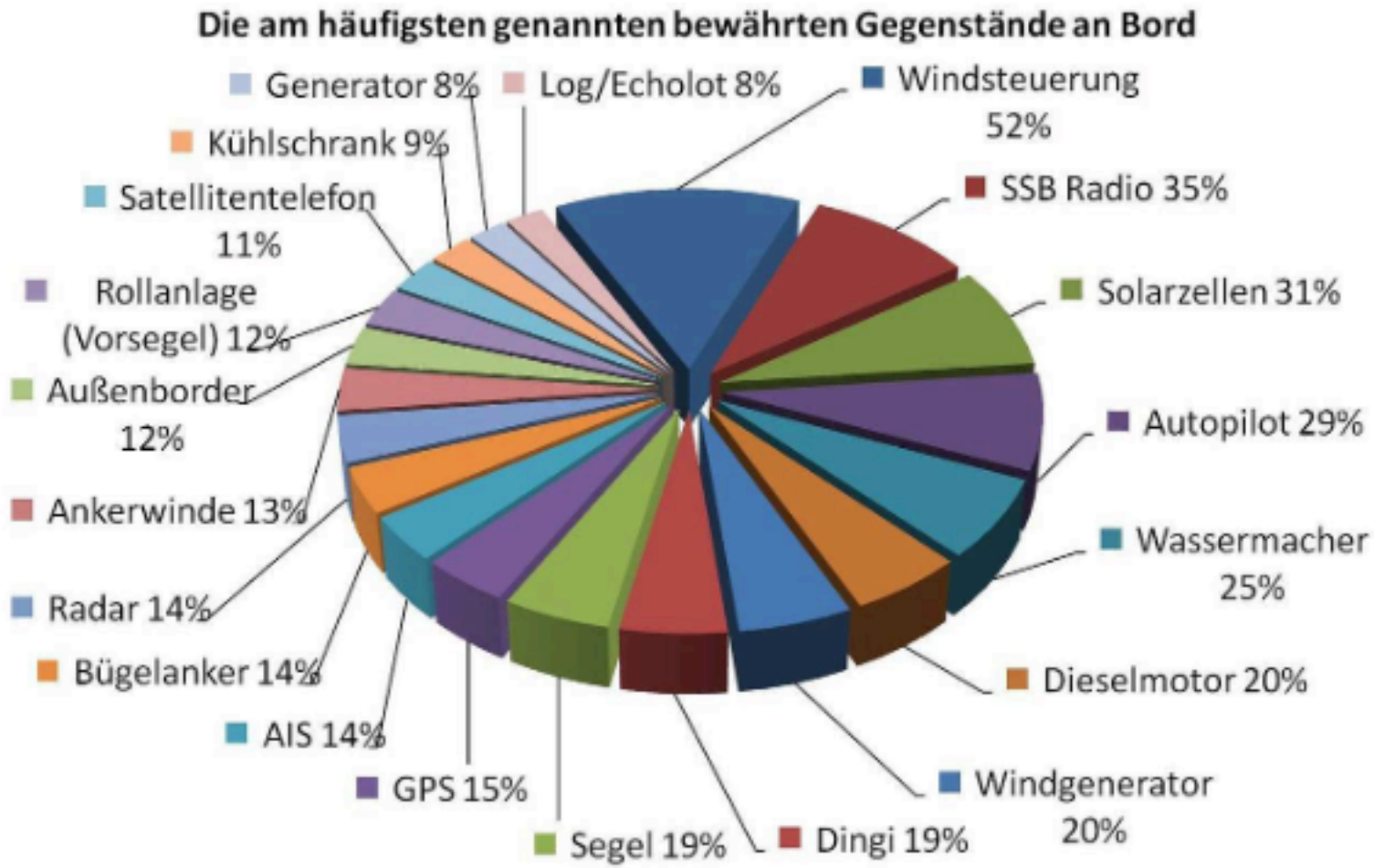
Yan-holtz.com/teaching

What's wrong with this chart?



Source: WTF Visualizations

Calculation
+
3D
+
Pie chart



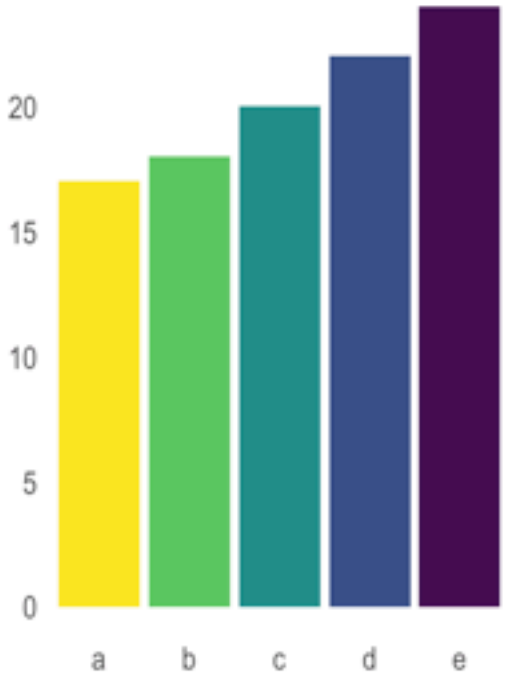
Source: WTF Visualizations

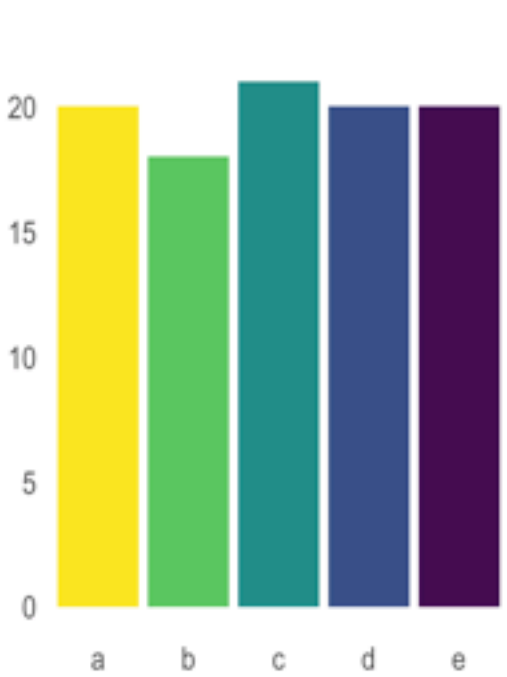
What's wrong with pie chart?

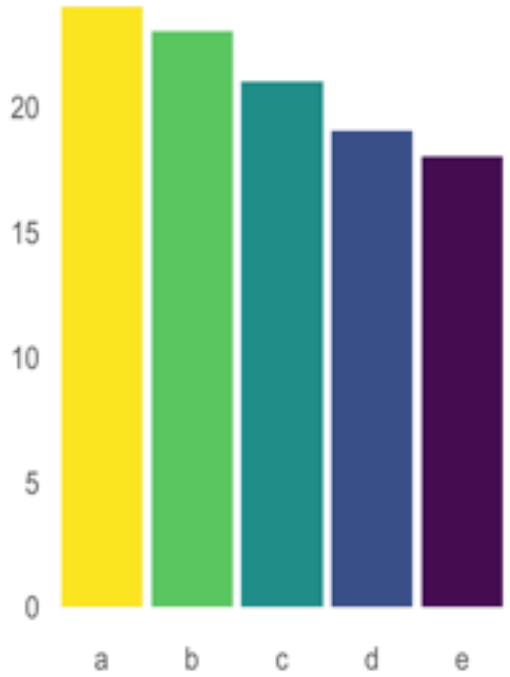
What can you see?

What's wrong with pie chart?

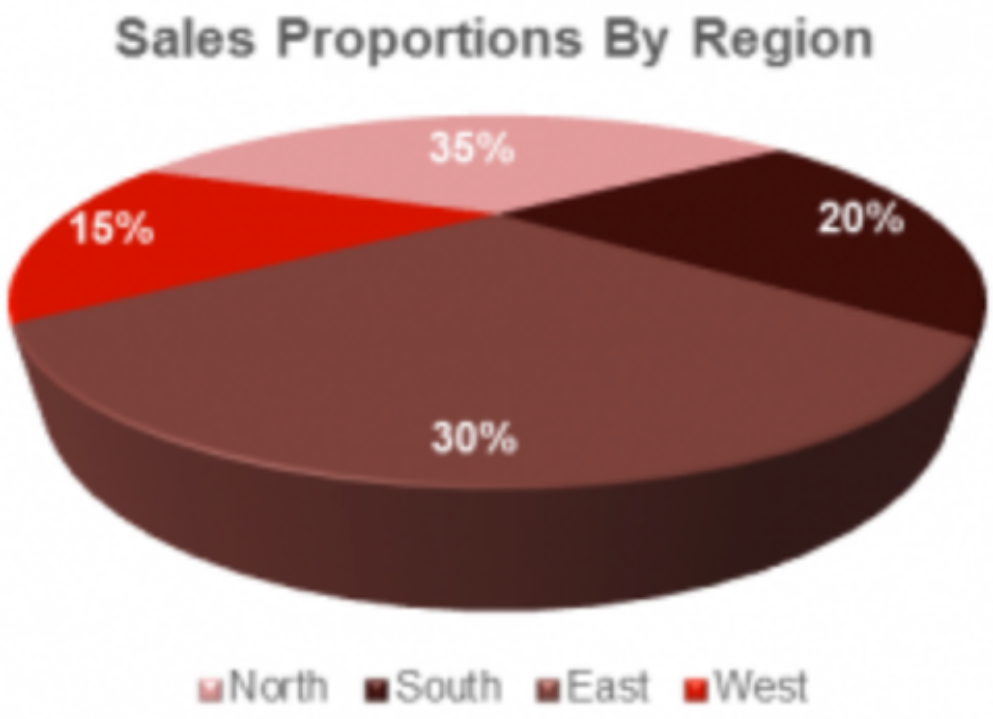
It is hard to distinguish angles



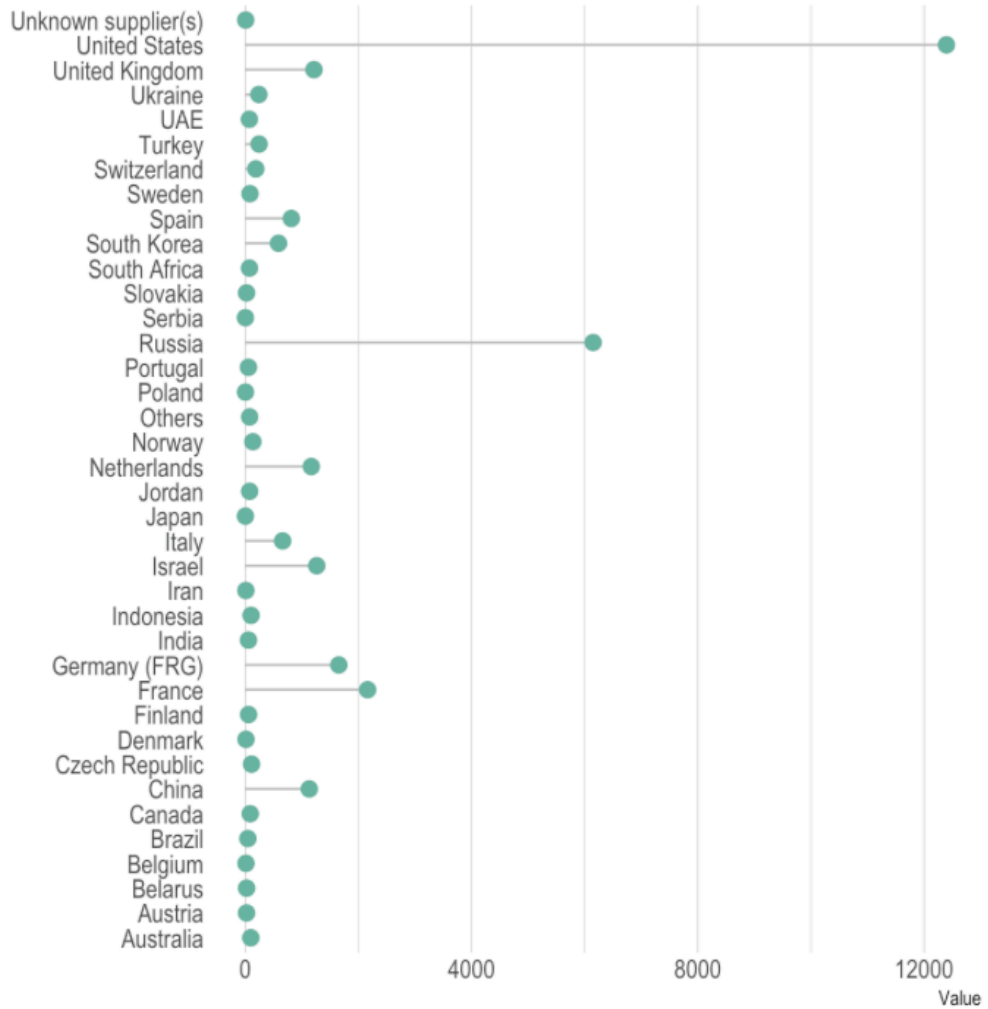




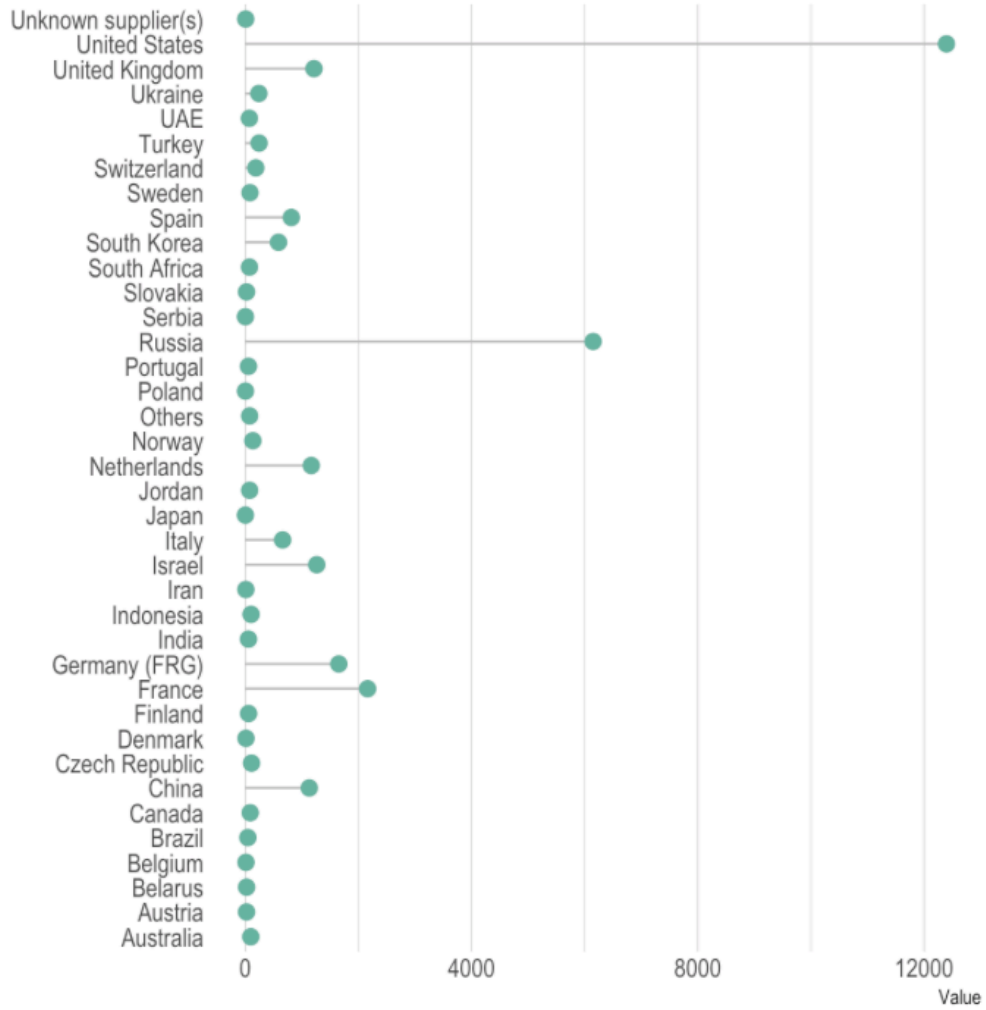
What's wrong with 3D

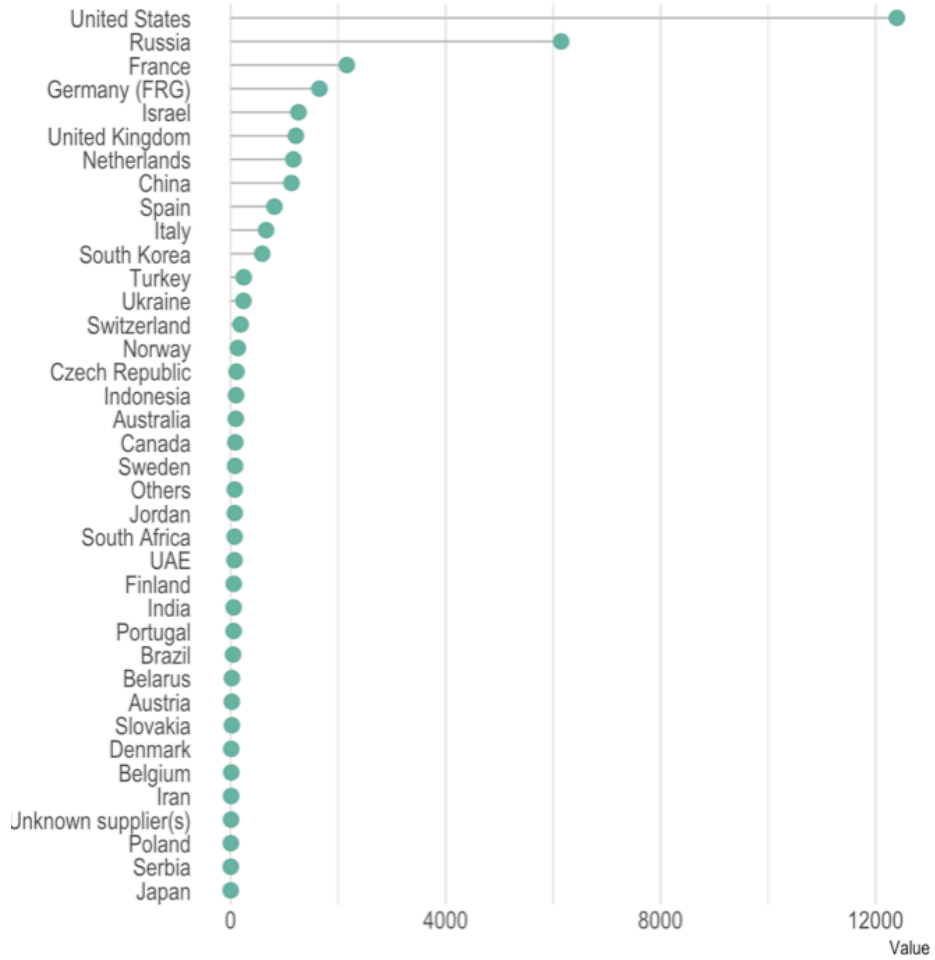


What could be better here?

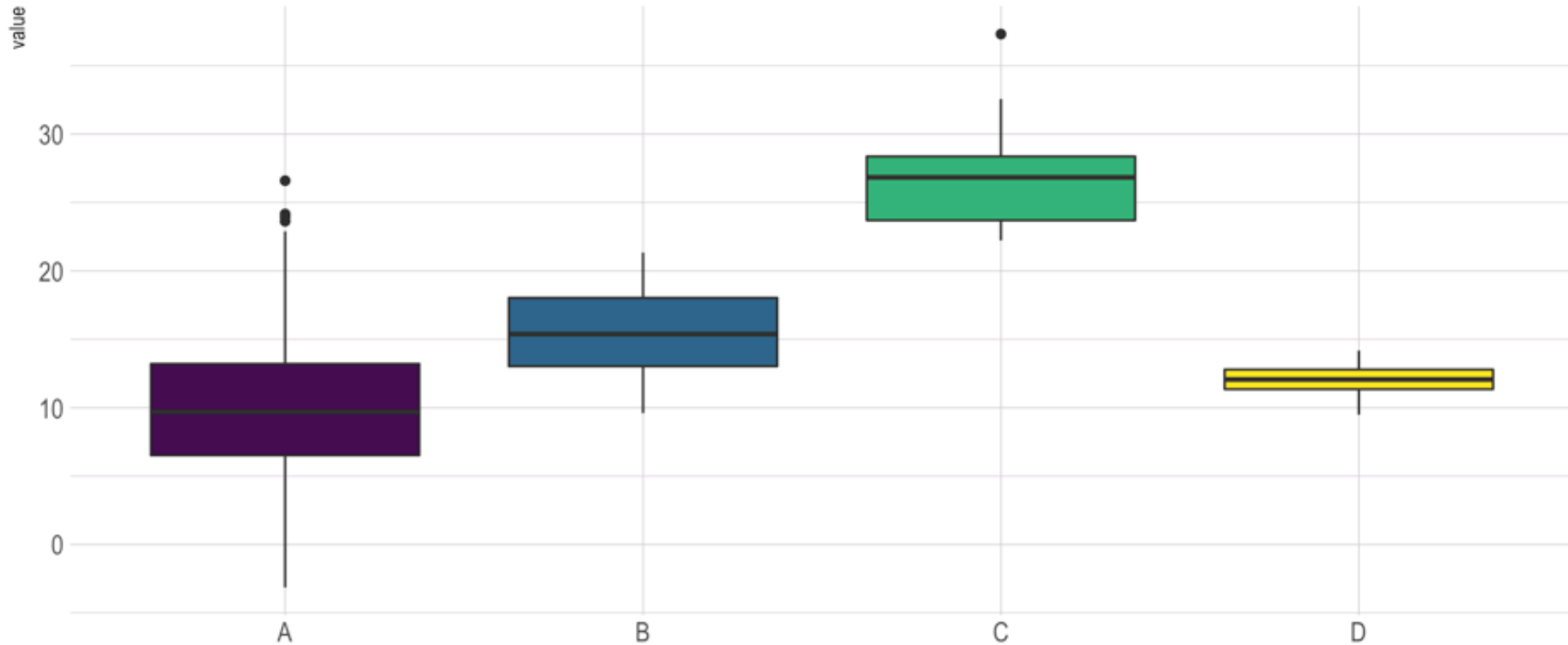


Order your data

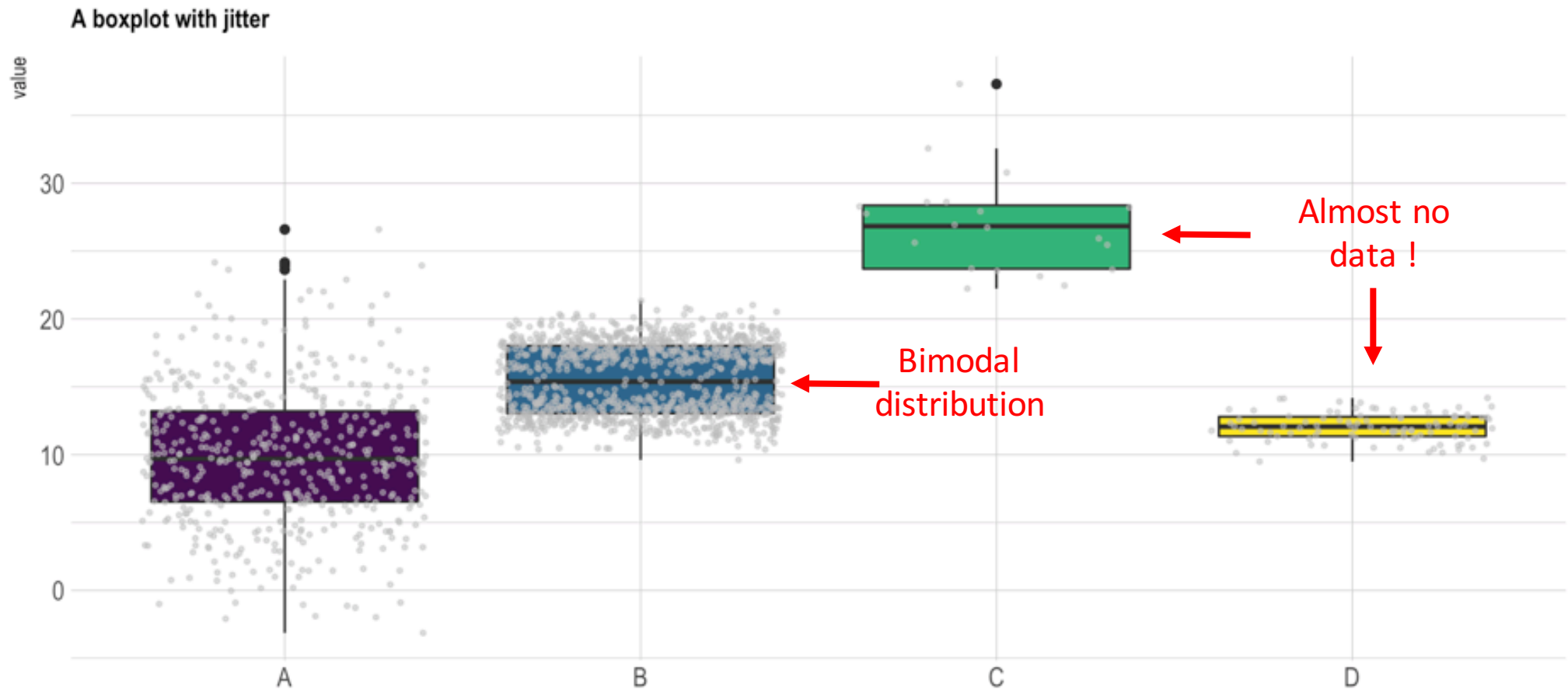




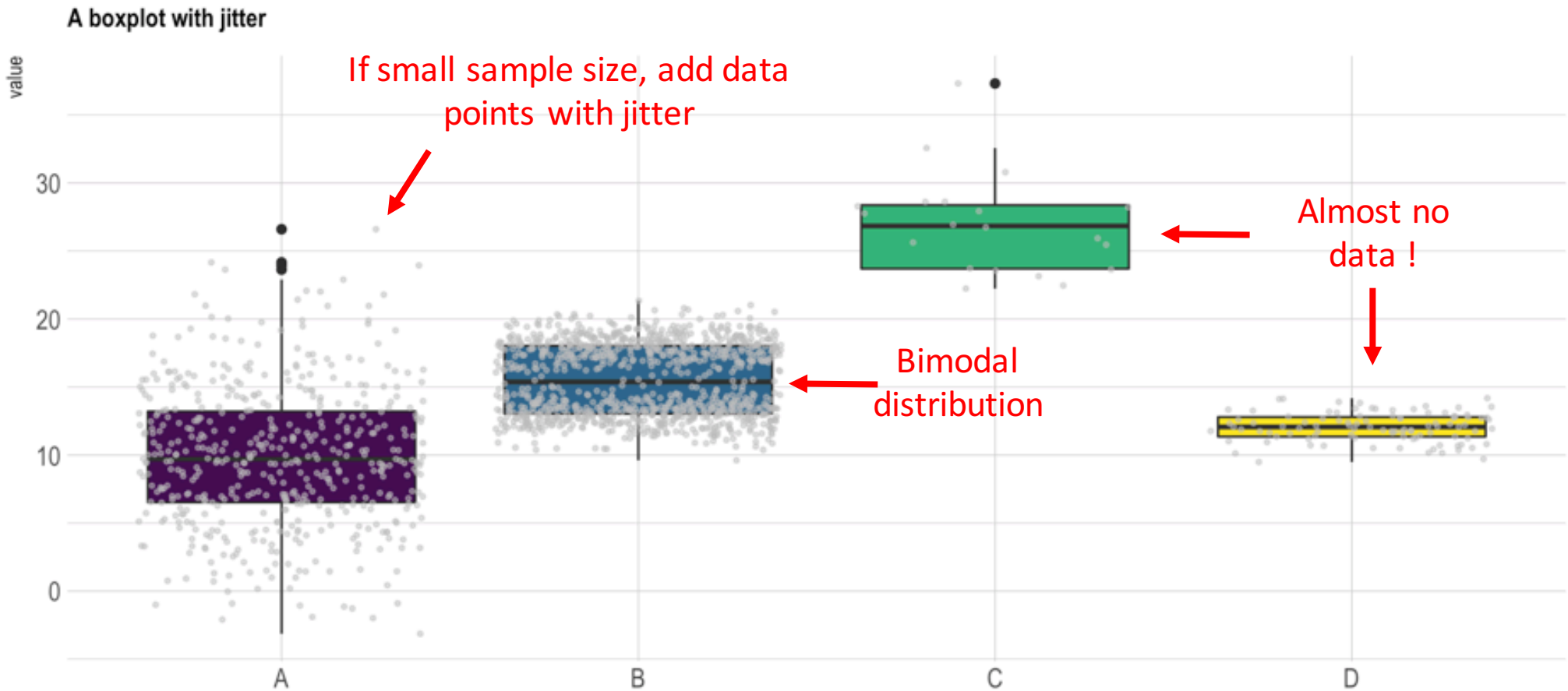
Anything wrong here?



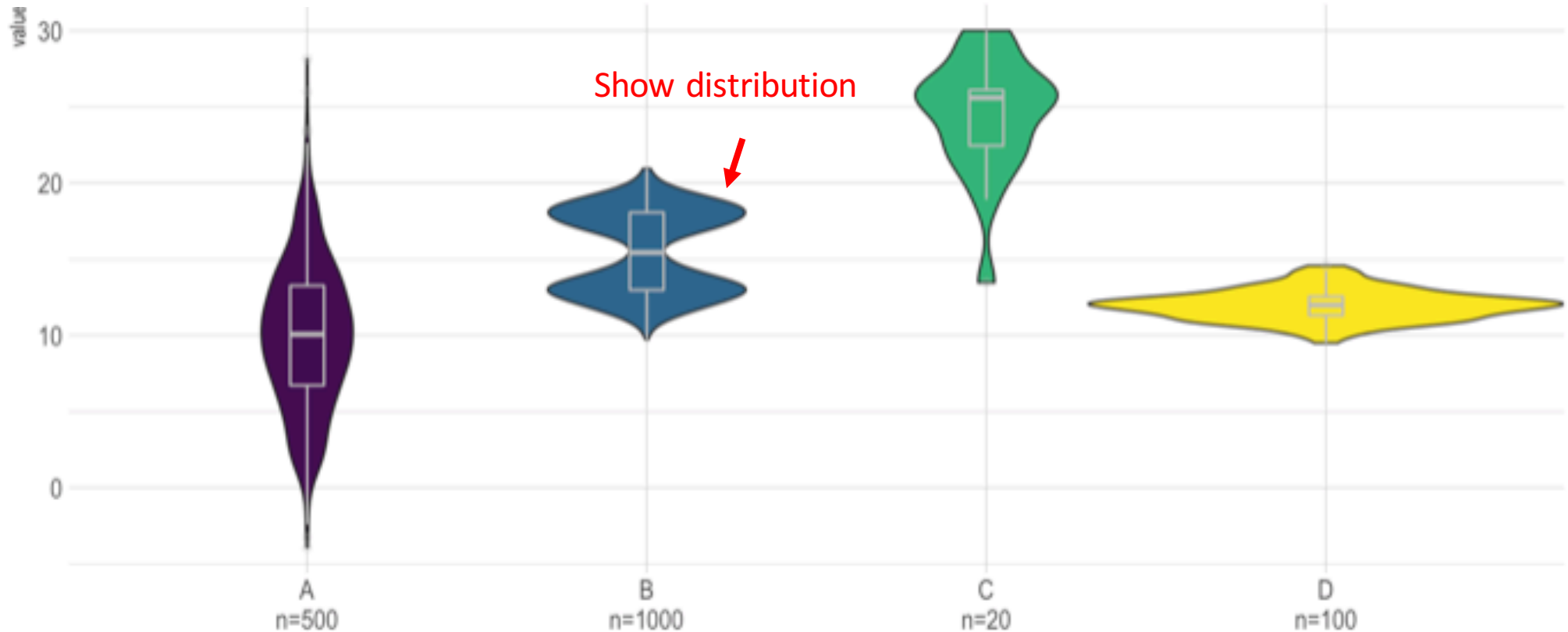
Boxplot = hide information



Boxplot = hide information



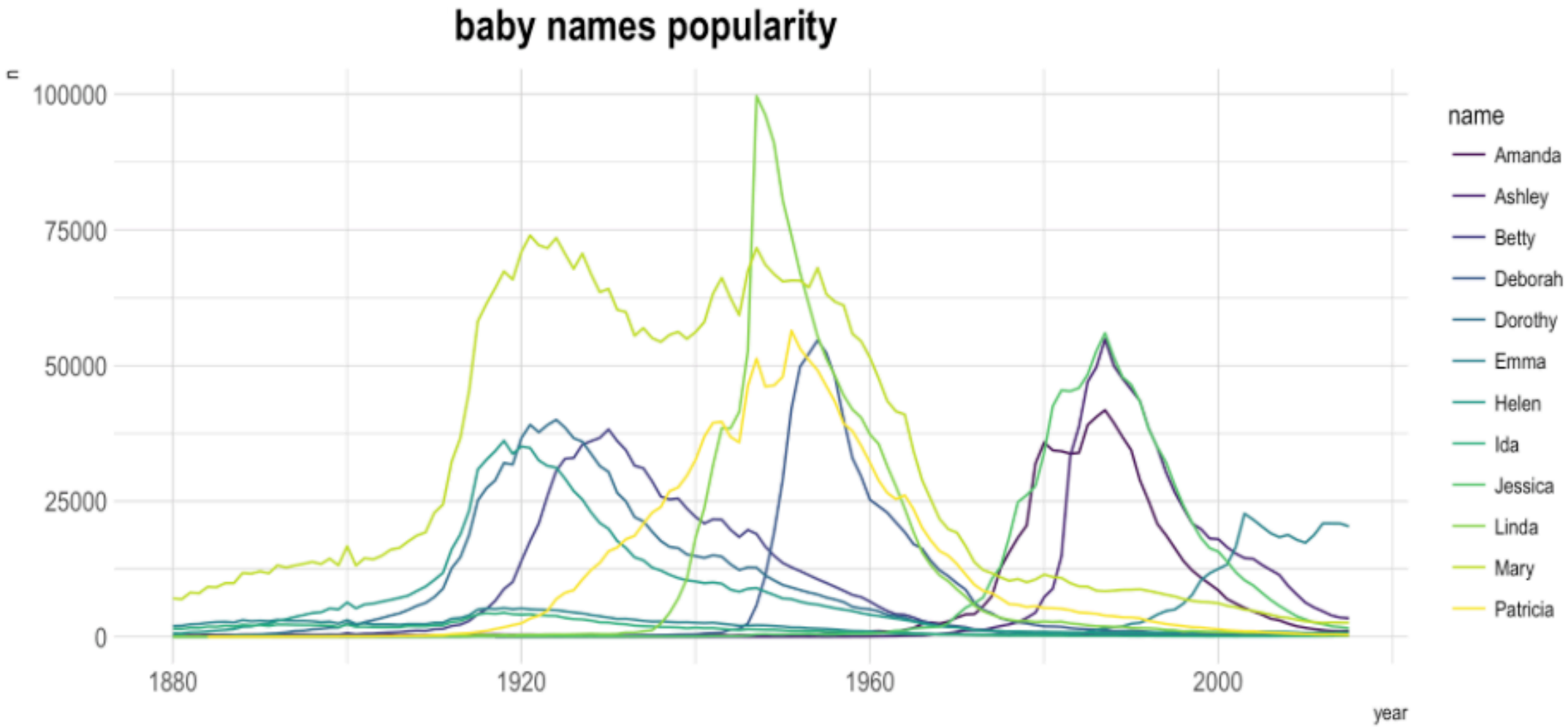
If big sample size, use violin plot



Show distribution

Show sample size

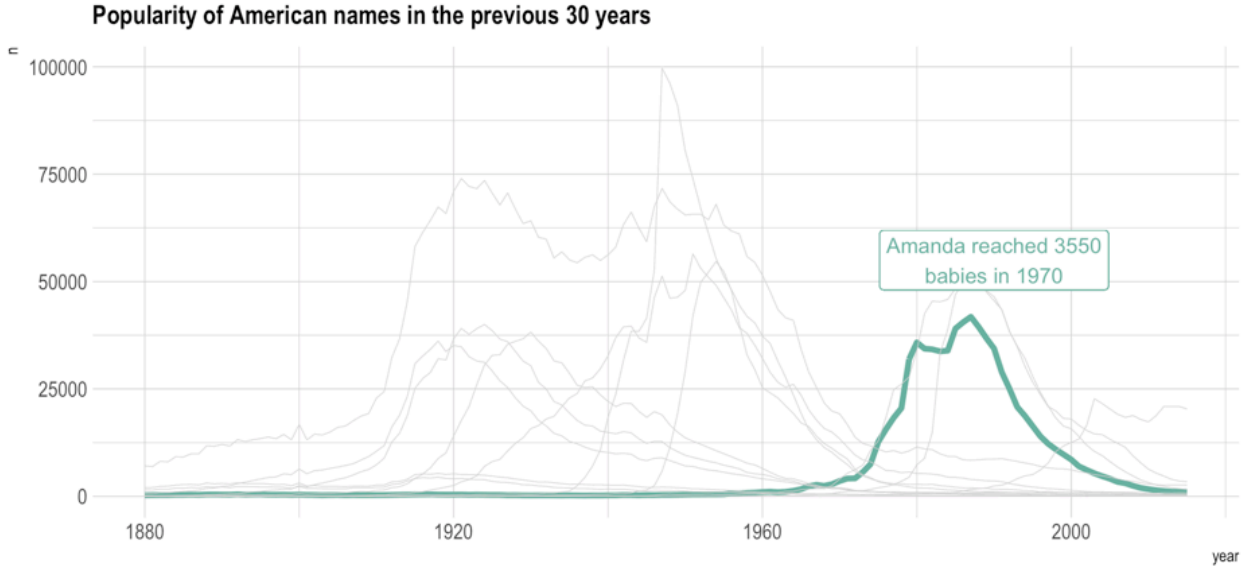
Look at this chart

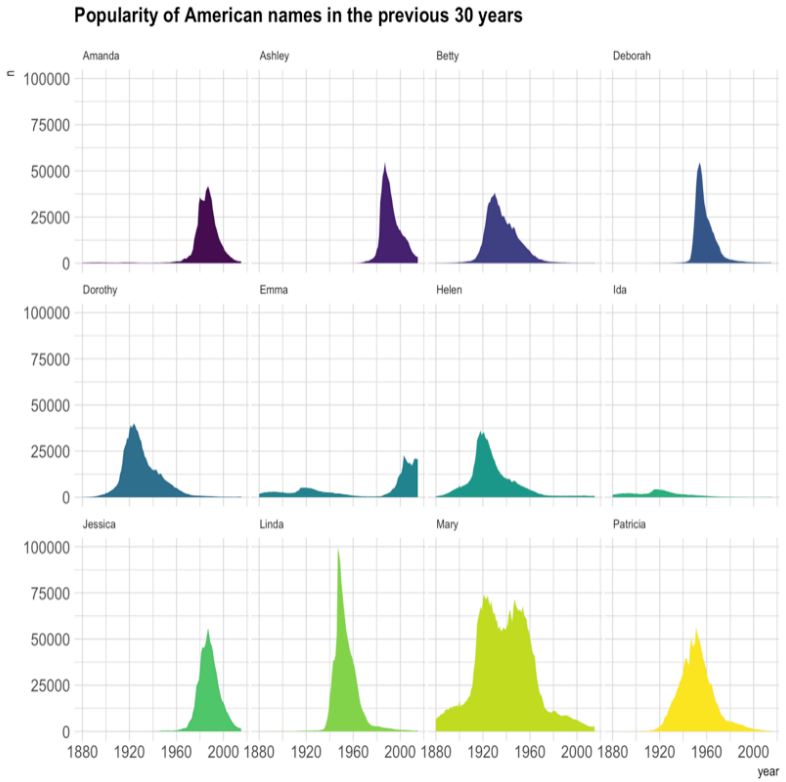


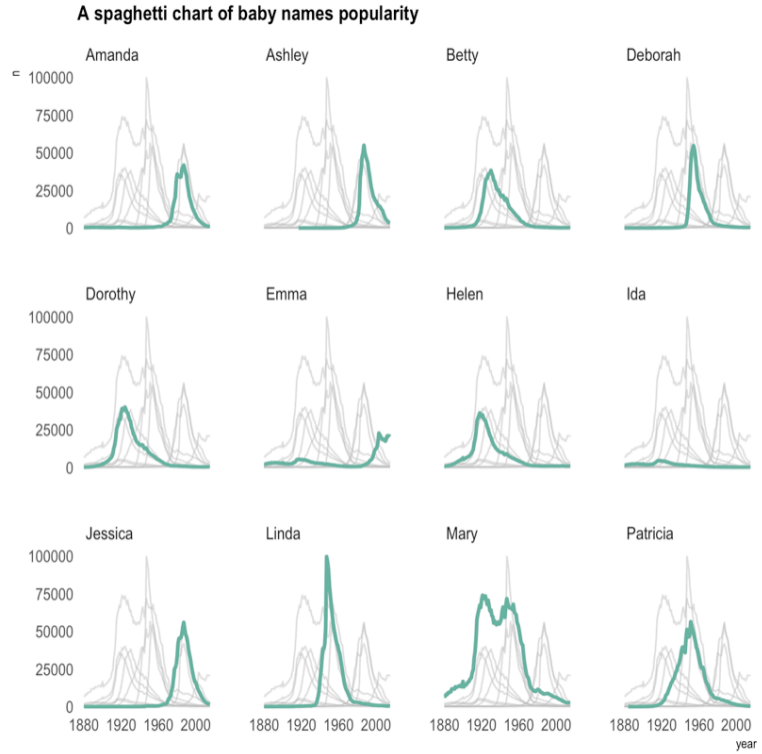
What do you remember?

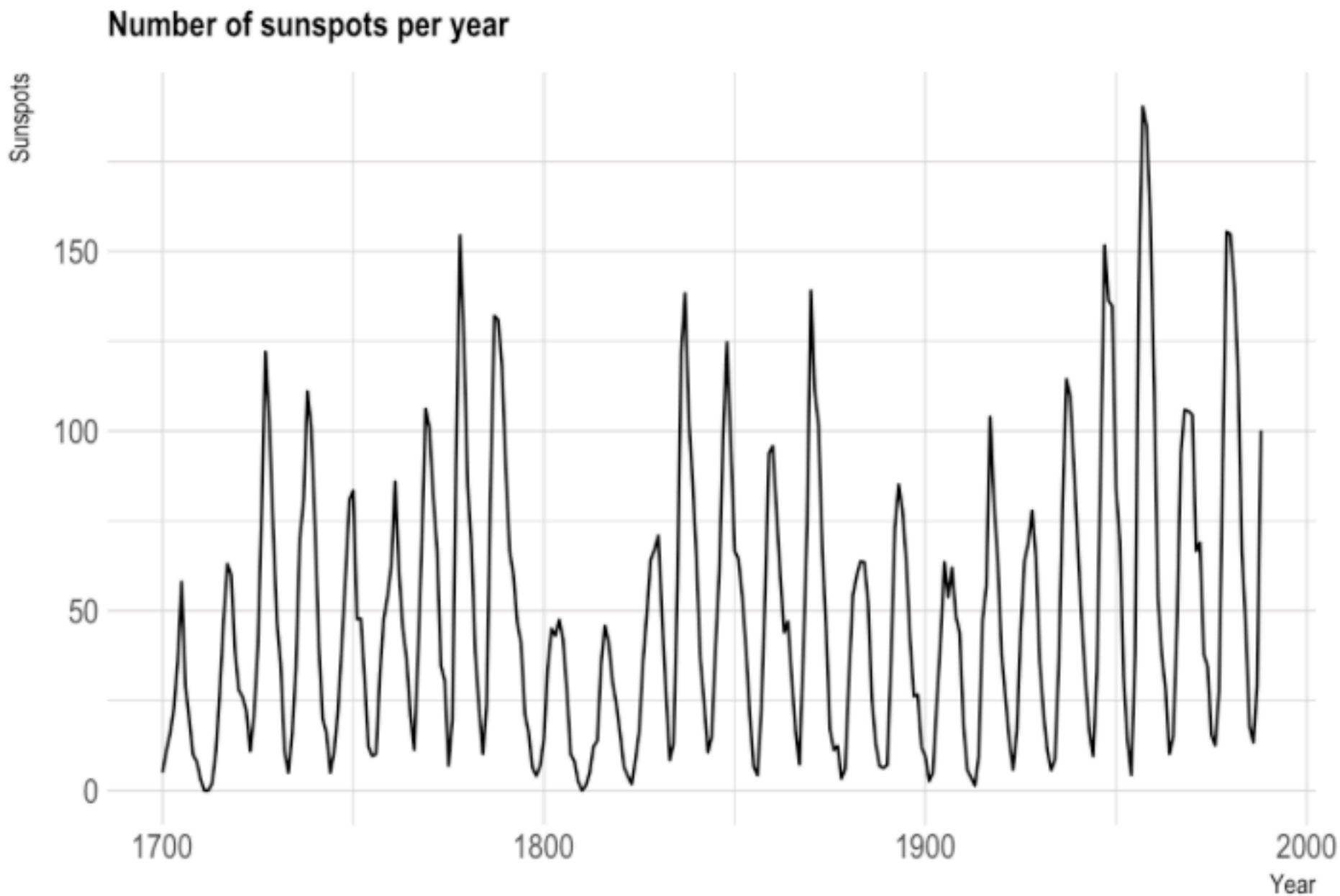
The Spaghetti chart

[Read more]



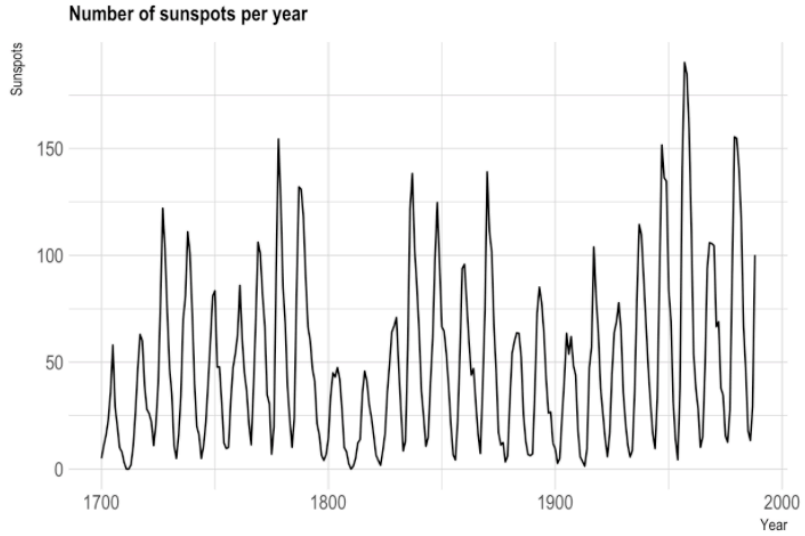


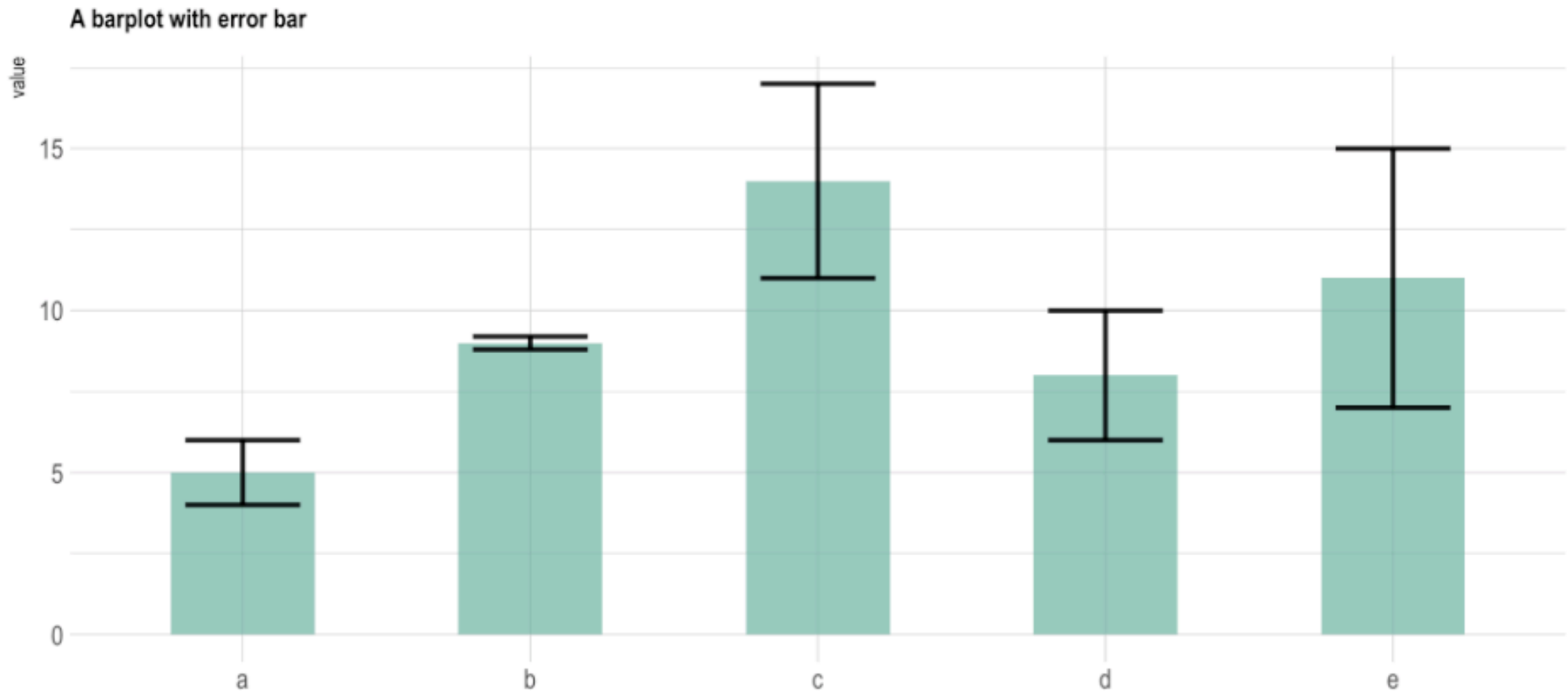



Number of sunspots per year
Sunspots
150
100
50
0
1700
1800
1900
2000
Year

Mind the aspect ratio

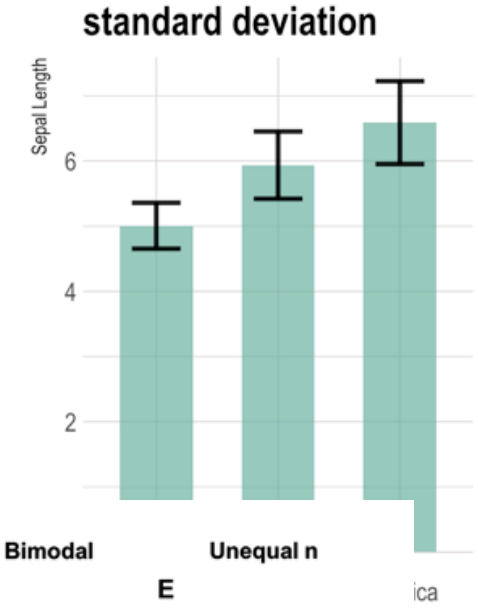
[Read more]

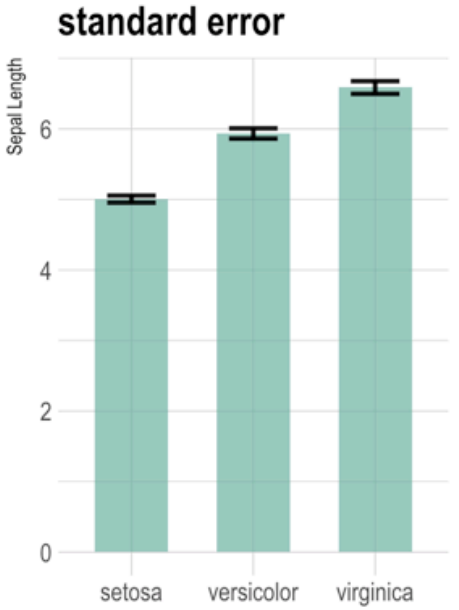


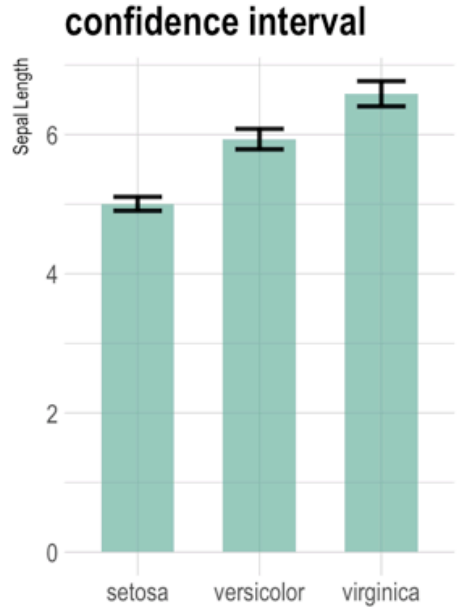

A barplot with error bar
value
15
10
5
0
a
b
c
d
e

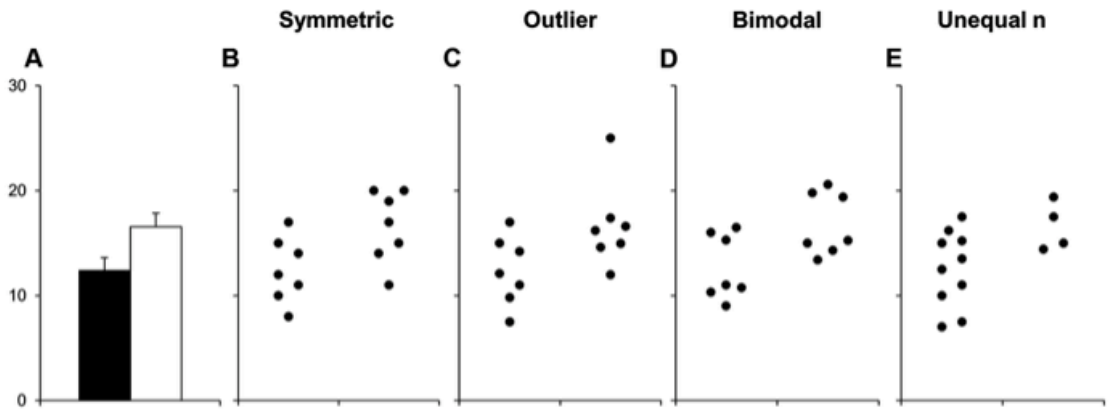
The issue with error bars

[Read more]







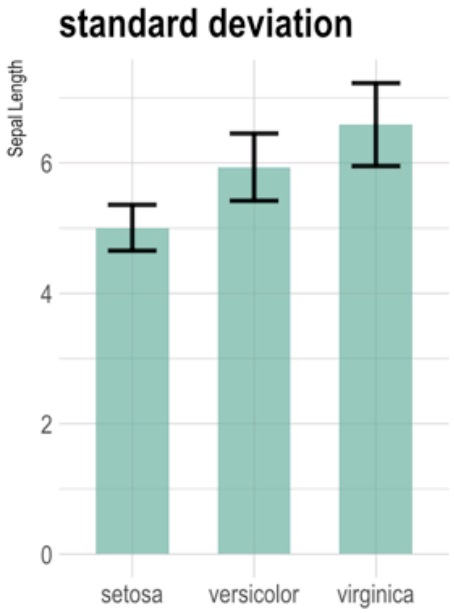


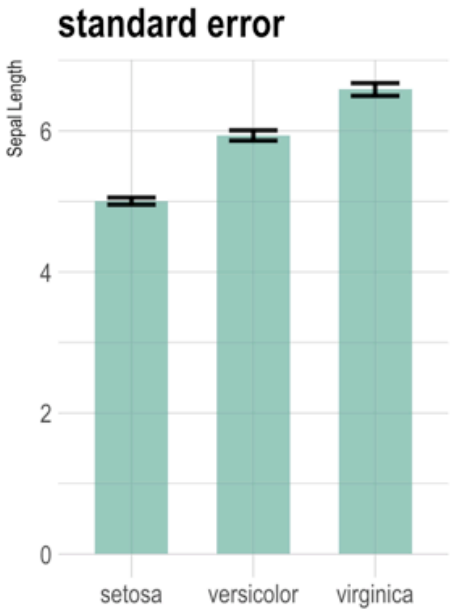
Test	p value			
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

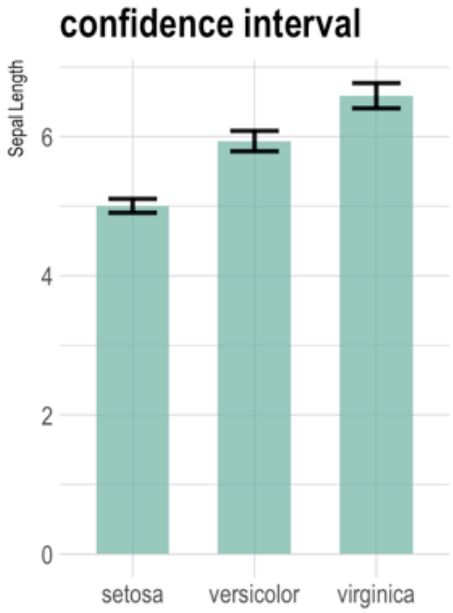
Weissgerber et al. 2015

The issue with error bars

Read more

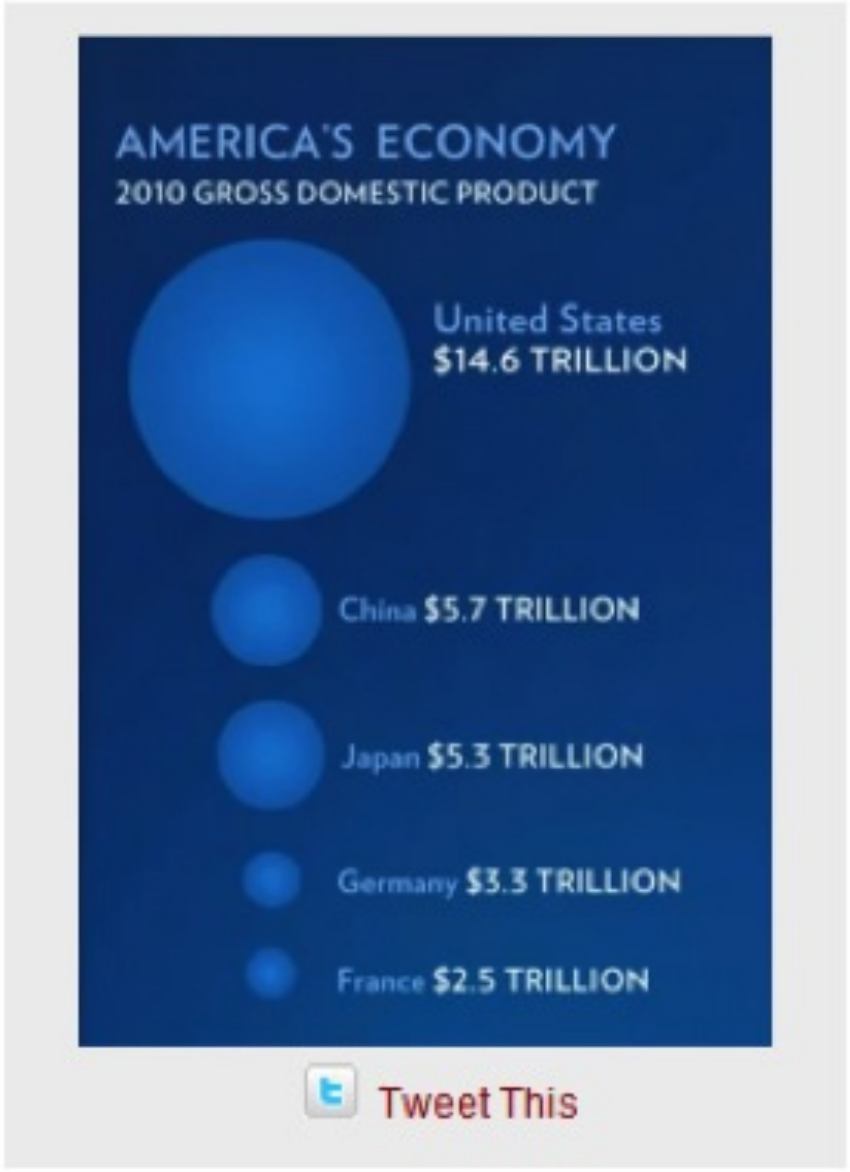




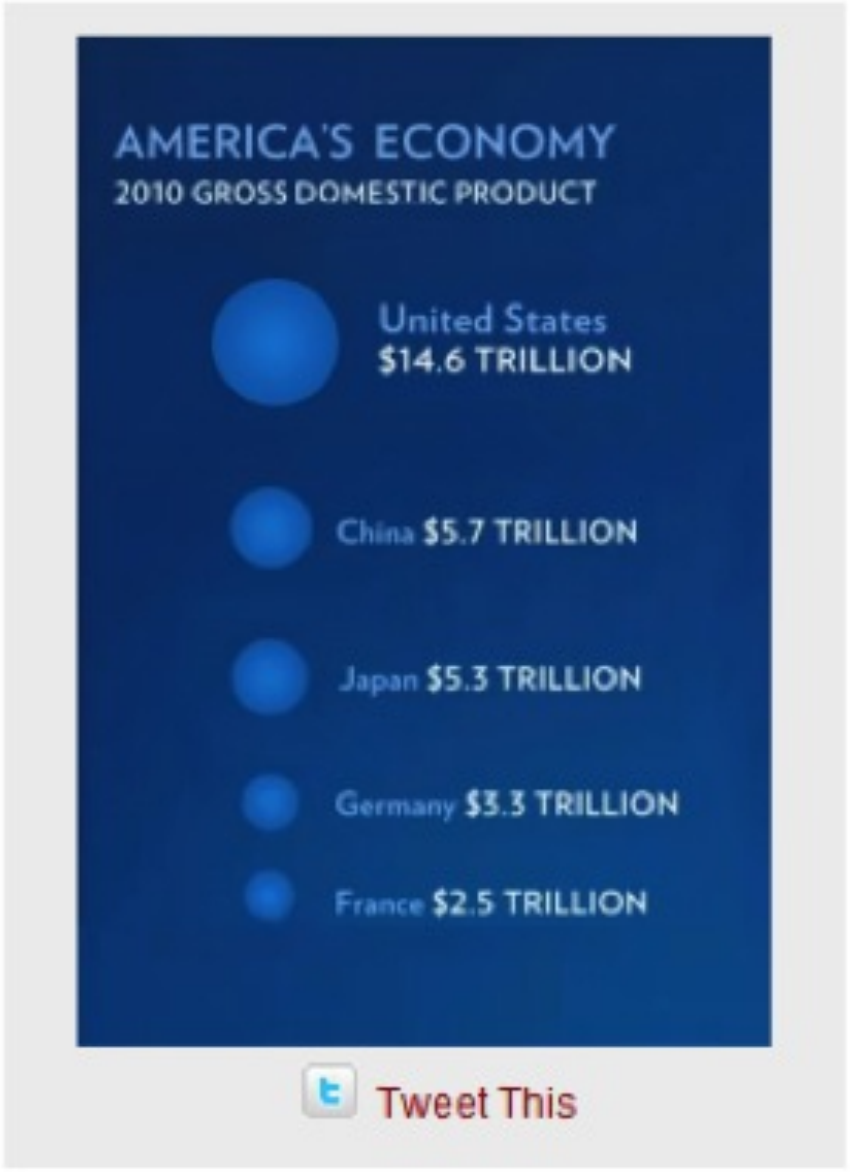


AMERICA'S ECONOMY
2010 GROSS DOMESTIC PRODUCT
United States
$14.6 TRILLION
China $5.7 TRILLION
Japan $5.3 TRILLION
Germany $3.3 TRILLION
France $2.5 TRILLION

Size = radius



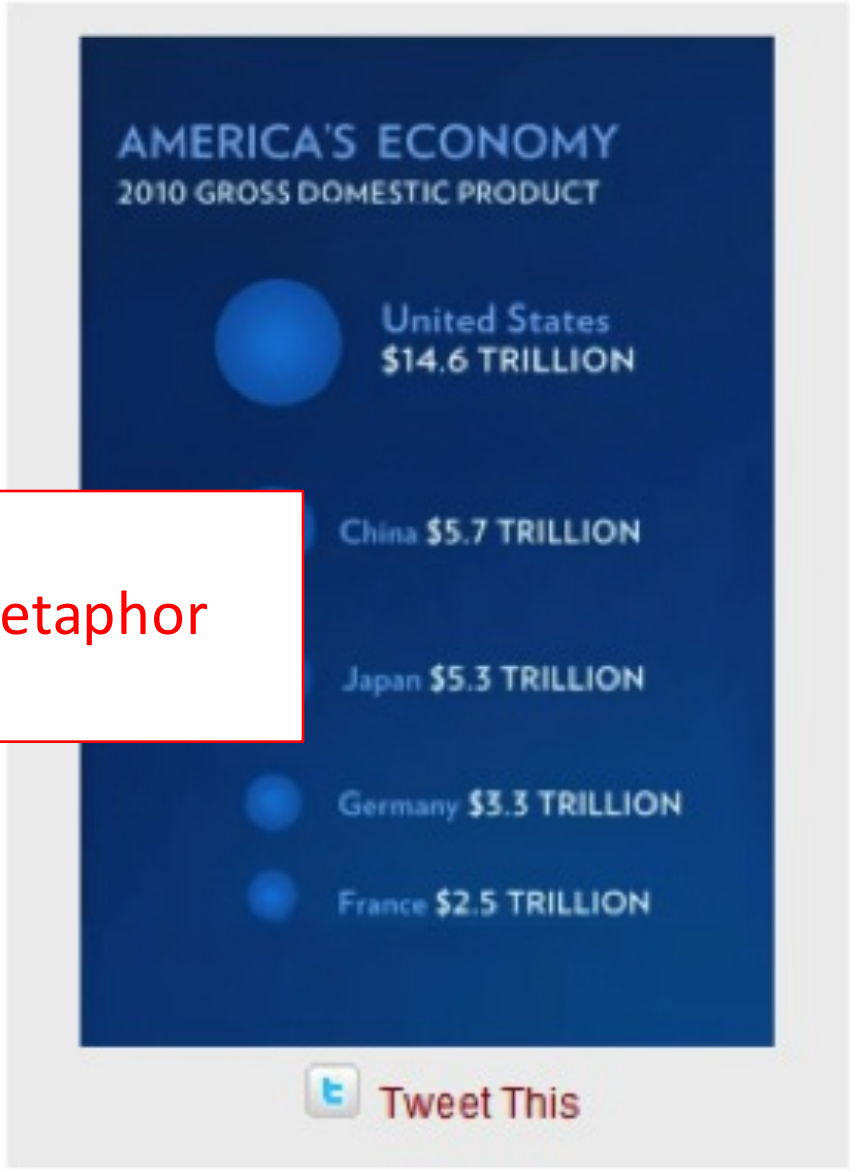
Size = area



Source: [Fast Fedora blog]

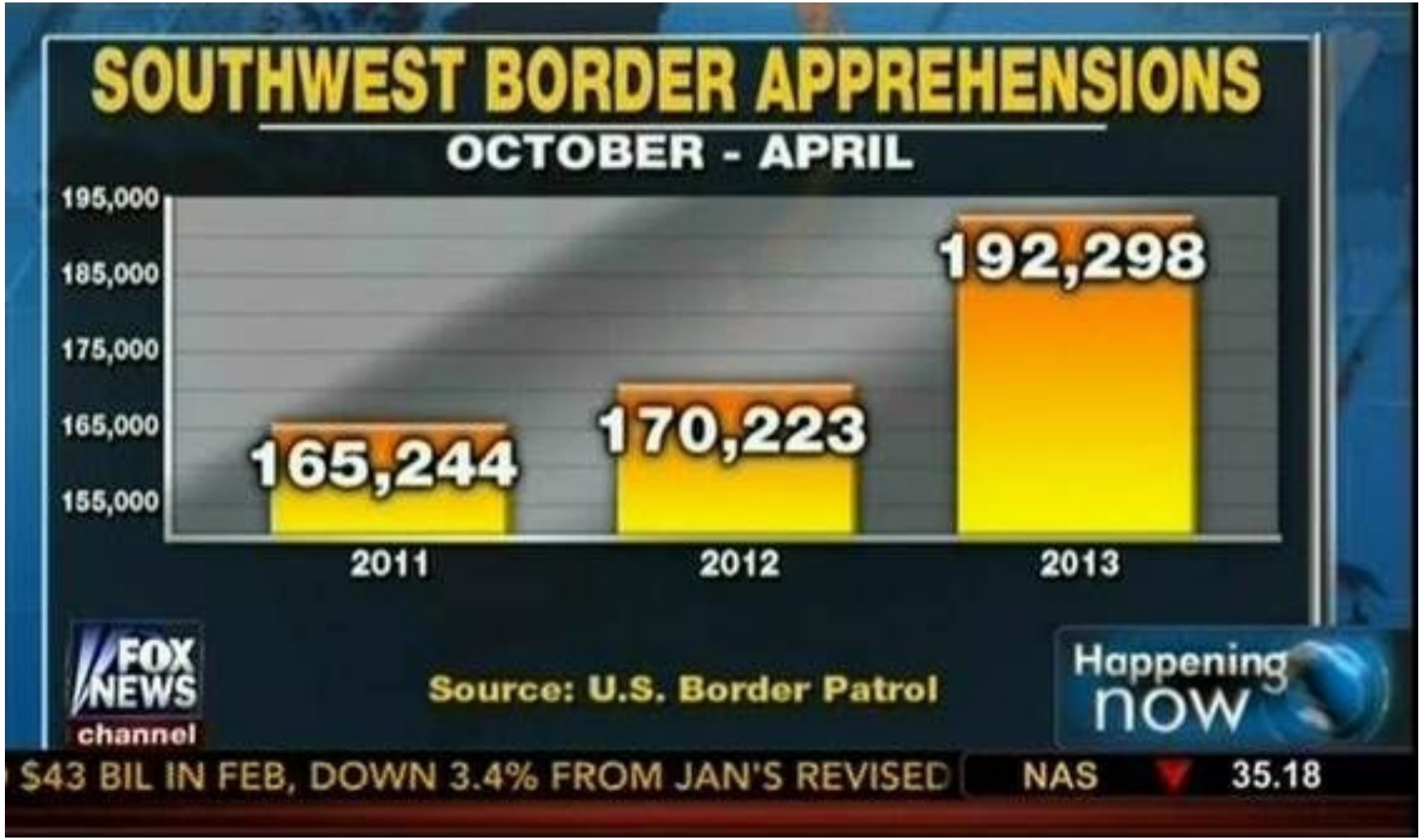
Size = radius

Size = area



Area = poor metaphor

Source: [Fast Fedora blog]

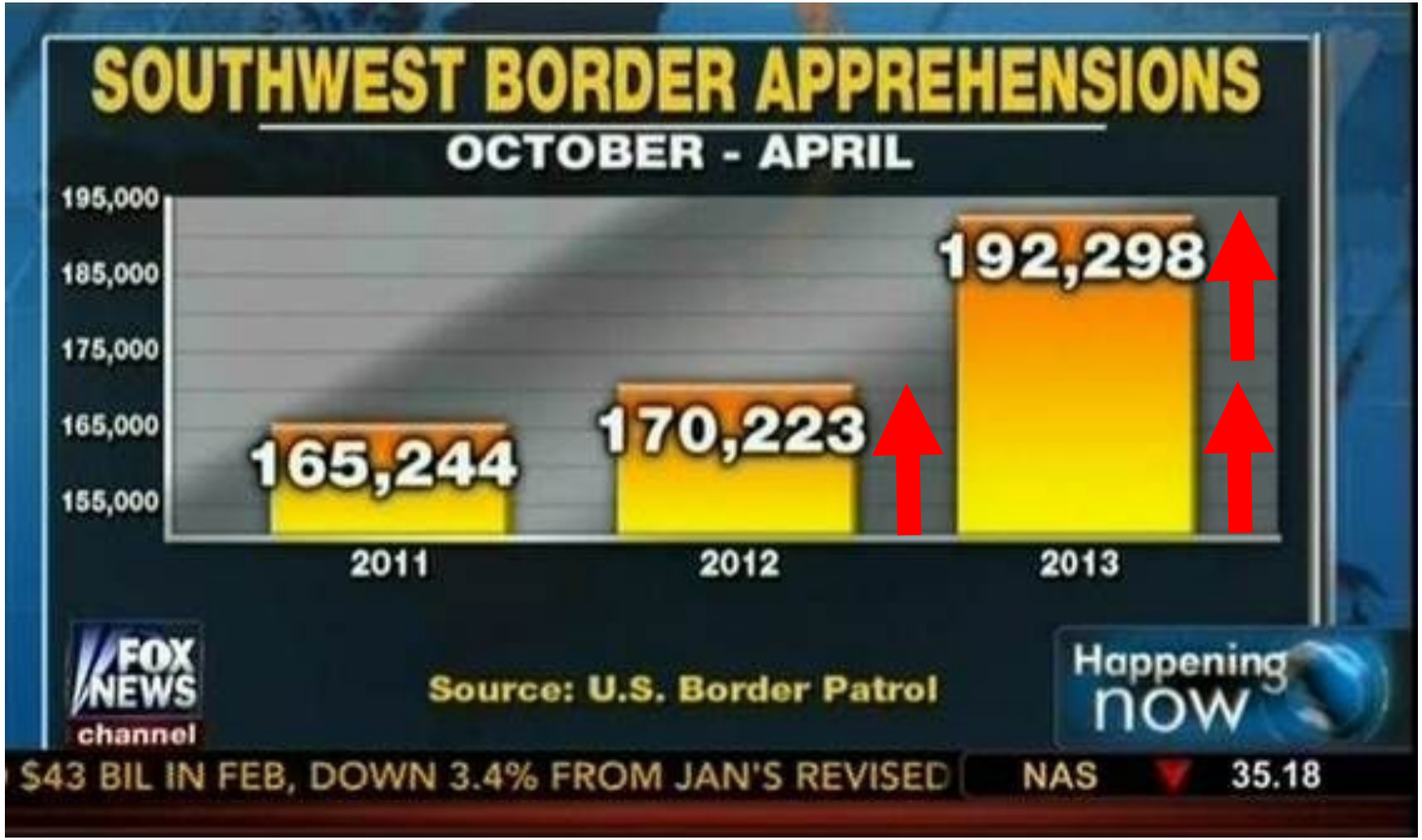


Source: Fox News, via [Media Matters for America]

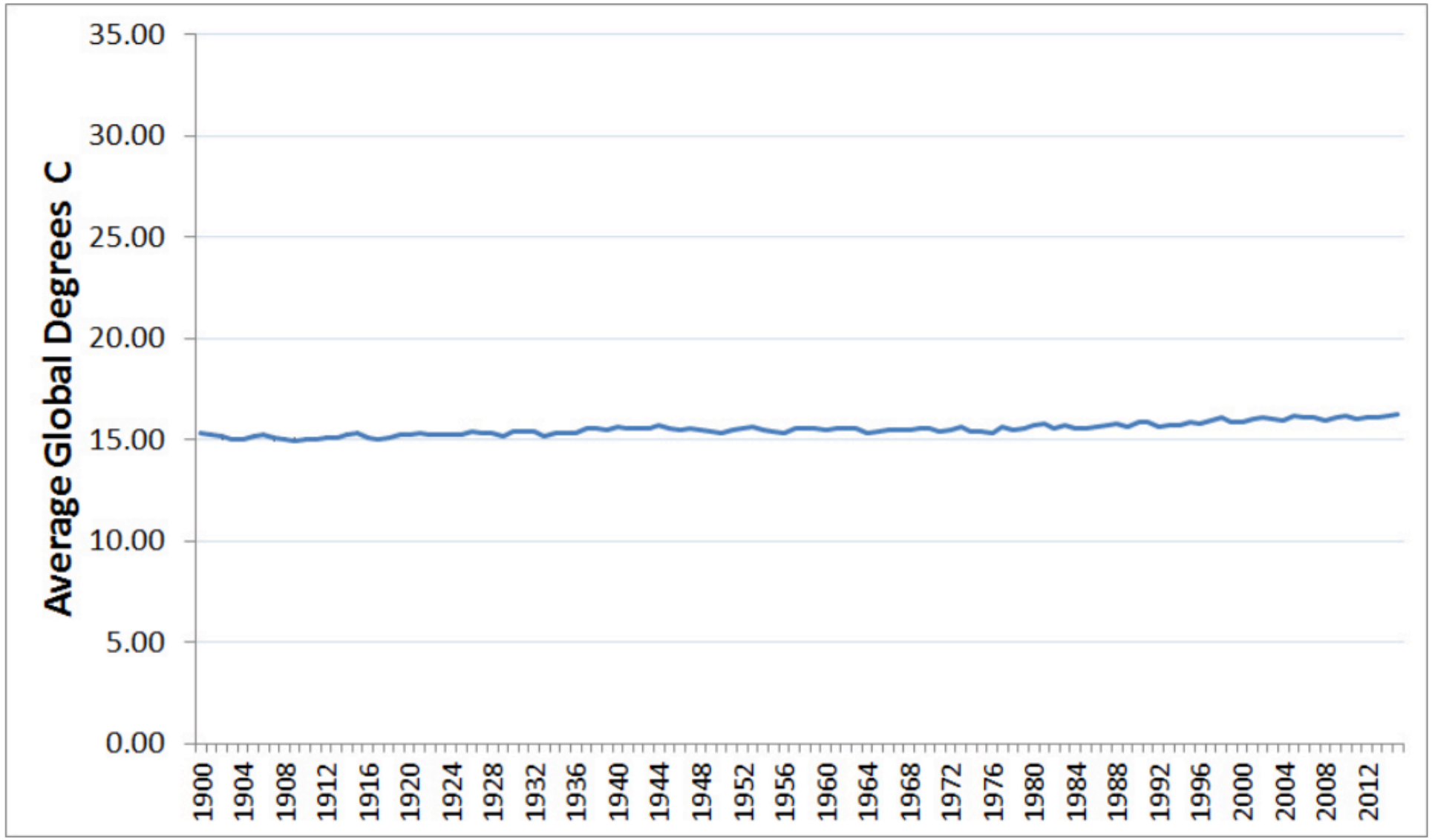
Obama adminitration

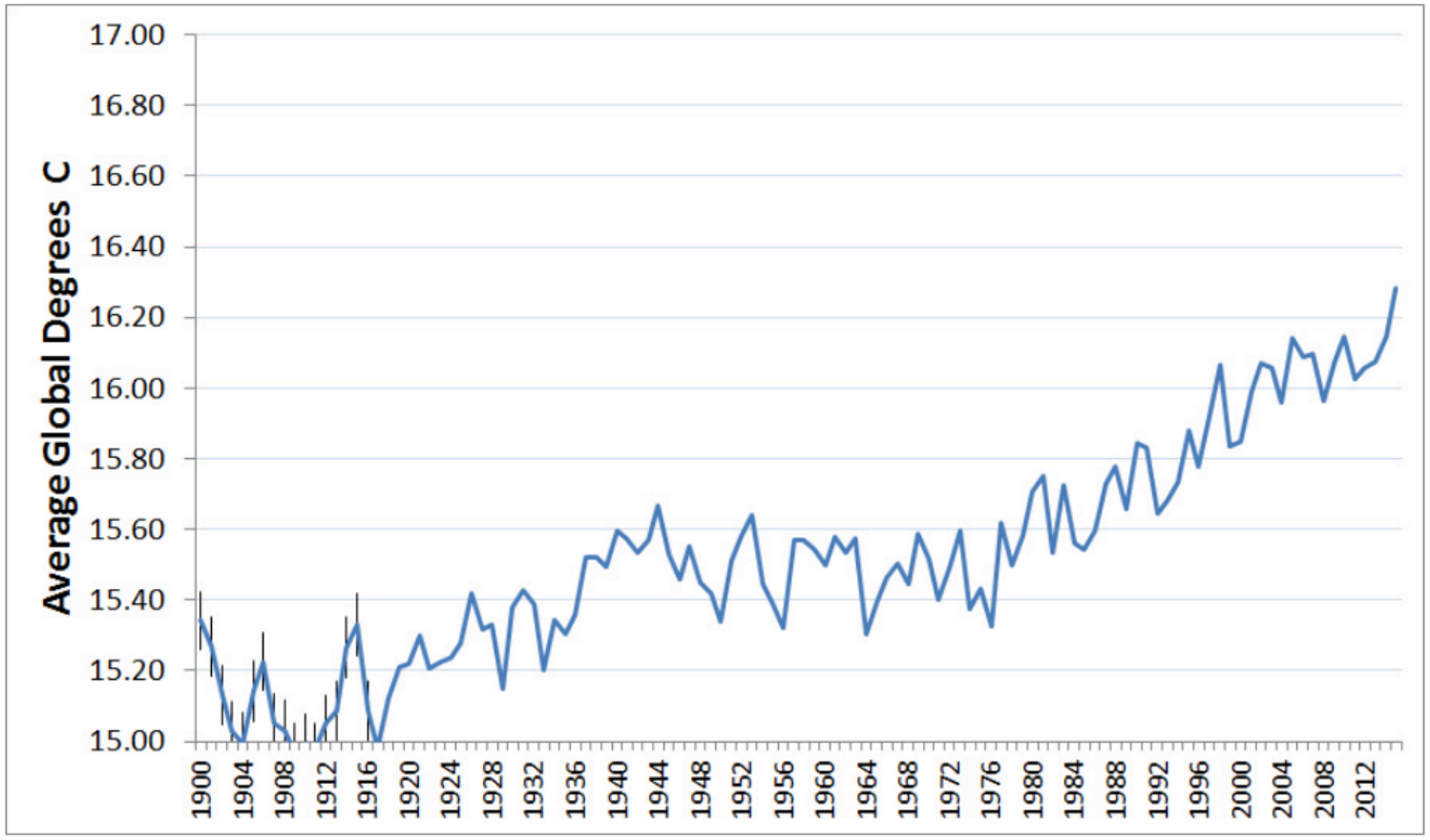
=

2x more apprehensions?

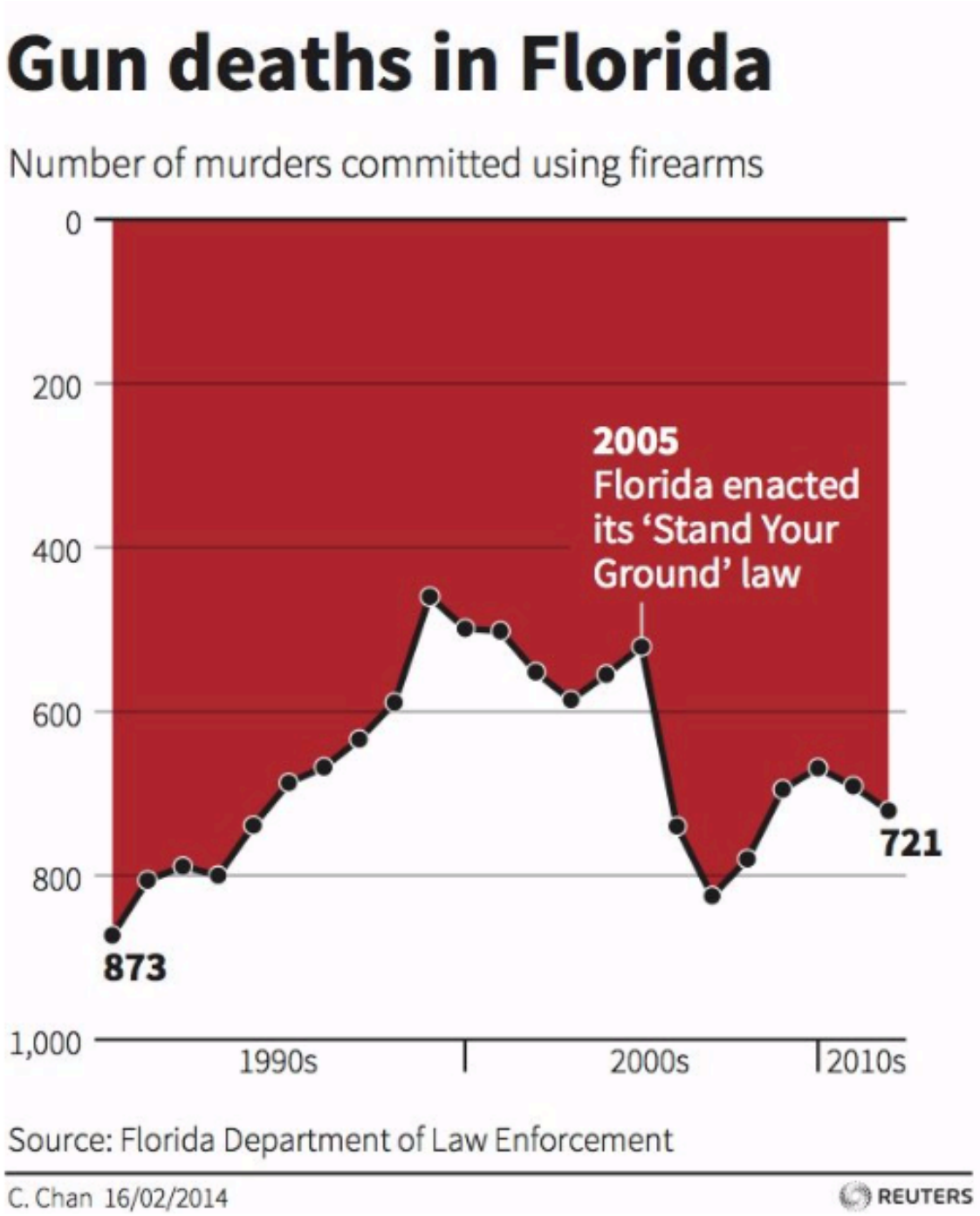


Source: Fox News, via [Media Matters for America]


Average Global Degrees C
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
1900
1904
1908
1912
1916
1920
1924
1928
1932
1936
1940
1944
1948
1952
1956
1960
1964
1968
1972
1976
1980
1984
1988
1992
1996
2000
2004
2008
2012



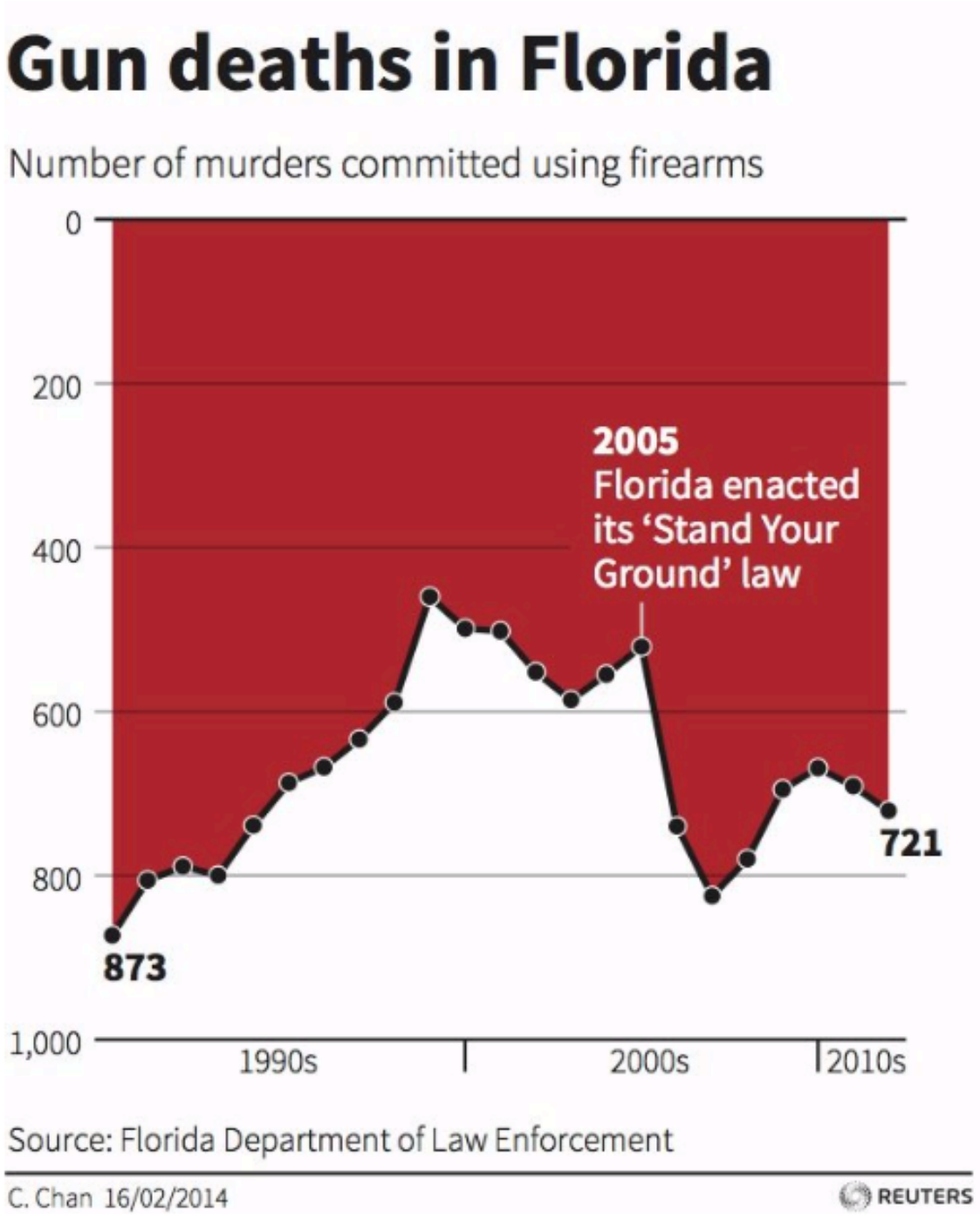
"In general, in a time-series, use a baseline that shows the data not the zero point" - Edward Tufte



Source: KD Nuggets

Do not be counter intuitive

Reverse Y Axis →



Source: KD Nuggets

Data to Viz

A collection of dataviz caveats

Data-to-viz.com/caveats

Order your data

When displaying the value of several entities, ordering them makes the graph much more insightful.

To cut or not to cut?

Cutting the Y-axis is one of the most controversial practice in data viz. See why.

The spaghetti chart

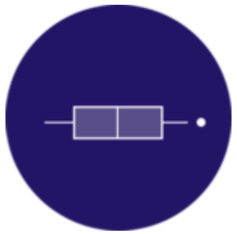
A line graph with too many lines becomes unreadable: it is called a spaghetti graph.

Pie chart

The human eye is bad at reading angles. See how to replace the most criticized chart ever.

Play with histogram bin size

Always try different bin sizes when you build a histogram, it can lead to different insights.



Do boxplots hide information?

Boxplots are a great way to summarize a distribution but hide the sample size and their distribution.

The problem with error bars

Barplots with error bars must be used with great care. See why and how to replace them.

Too many distributions.

If you need to compare the distributions of many variables, don't clutter your graphic.

Thanks

Slides: Yan-holtz/teaching

Data to viz: Data-to-viz.com

Yan.holtz.data@gmail.com

www.yan-holtz.com

@R_Graph_Gallery

github.com/holtzy