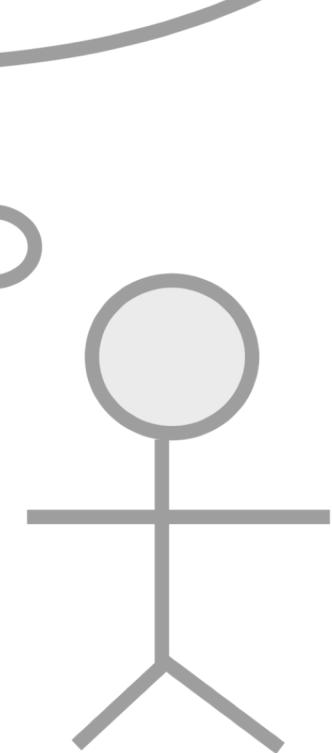


from Data to Viz

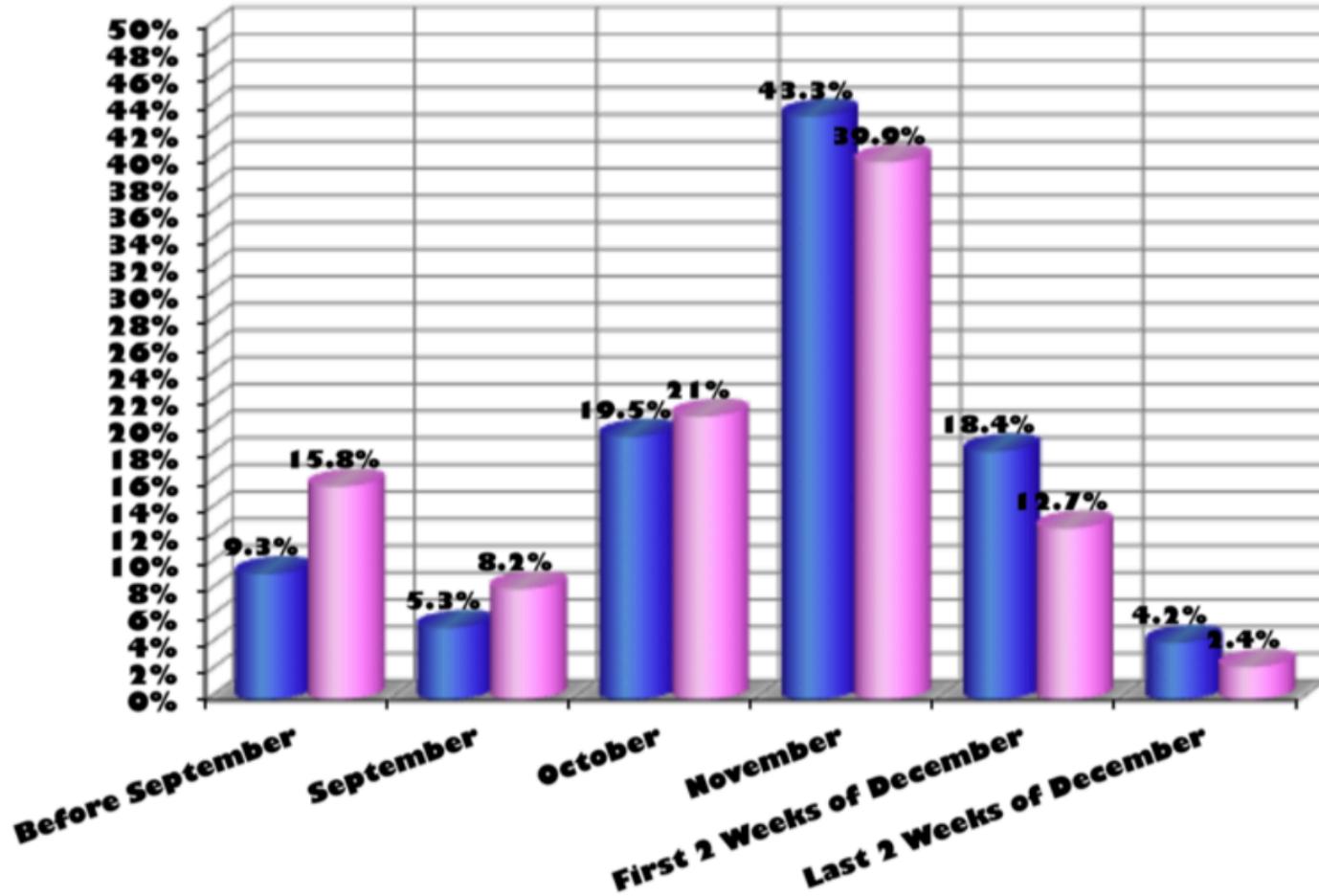
A classification of chart types based on input data format

| Id | feature 1 |
|-----|-----------|
| A | 10 |
| B | 12 |
| C | 15 |
| ... | ... |

What should I do with my
data ??



■ Men ■ Women



...

What about
that?

Looking for a chart ?

What you
can do

What you
should do

Caveats to
avoid

How to
build it

WHAT YOU CAN DO

A classification of chart types based on data input format

Who sells more weapons ?

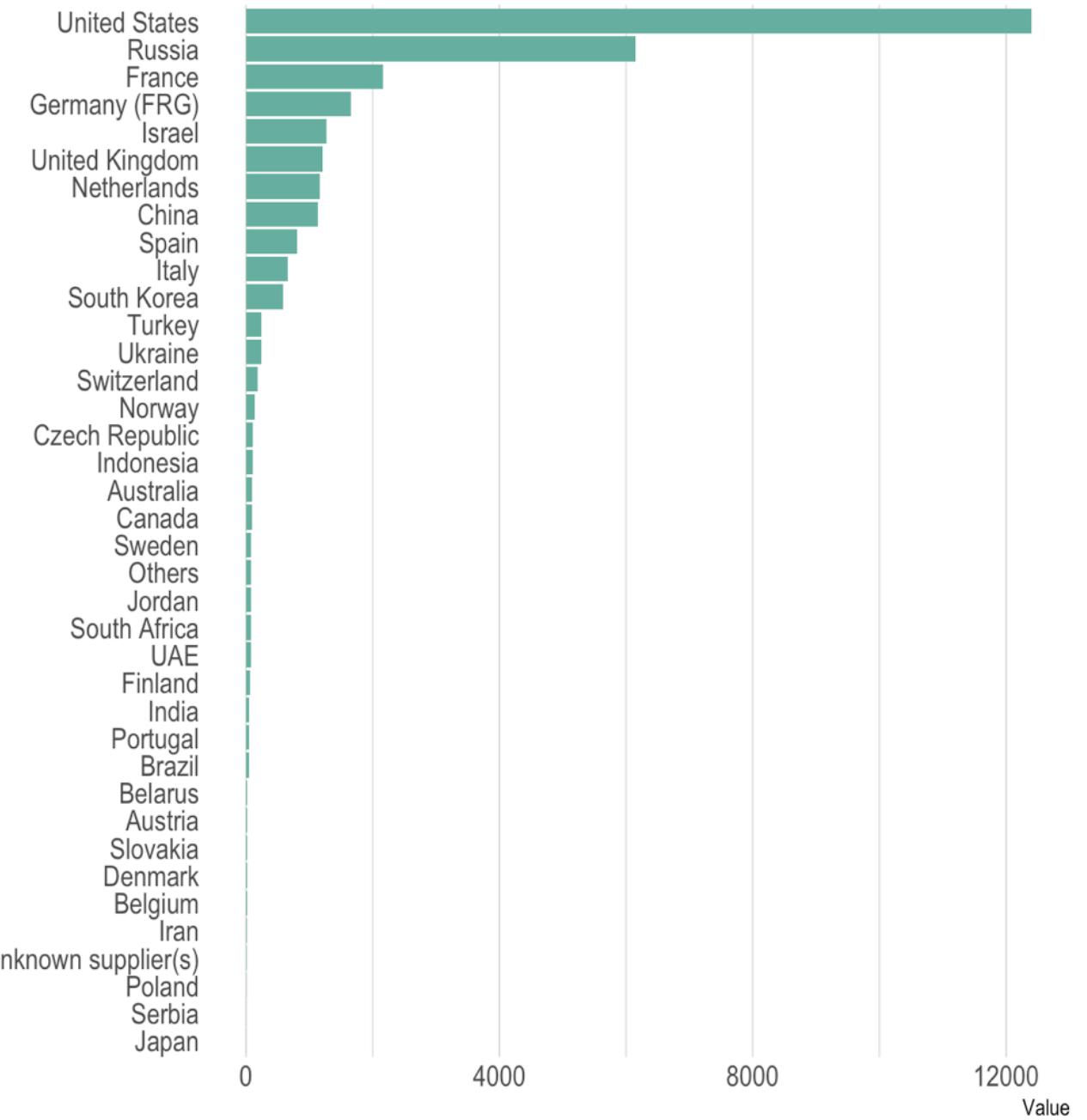
| Country | Value |
|----------------|-------|
| United States | 12394 |
| Russia | 6148 |
| Germany (FRG) | 1653 |
| France | 2162 |
| United Kingdom | 1214 |
| China | 1131 |



?

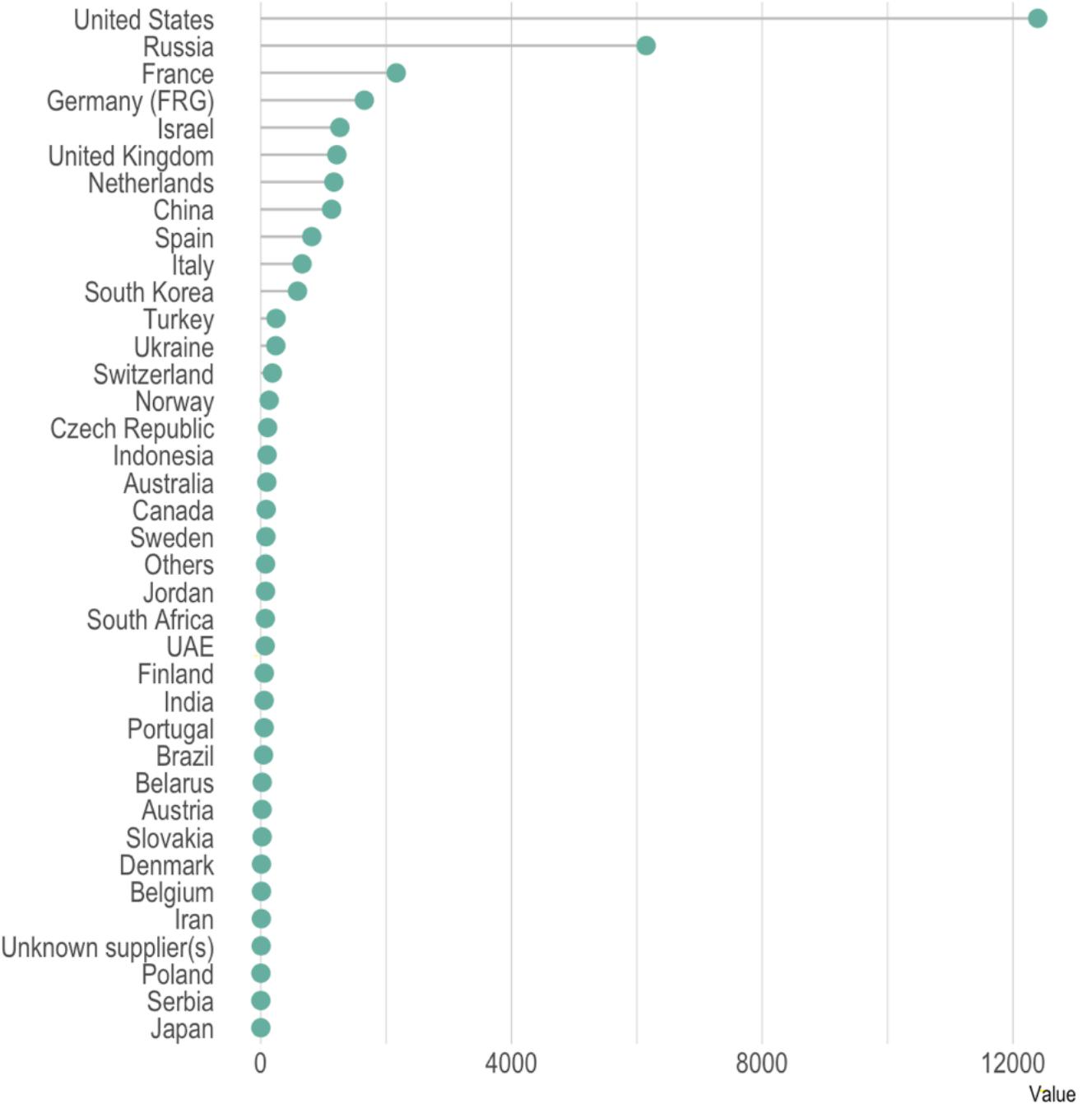
Who sells more weapons ?

| Country | Value |
|----------------|-------|
| United States | 12394 |
| Russia | 6148 |
| Germany (FRG) | 1653 |
| France | 2162 |
| United Kingdom | 1214 |
| China | 1131 |



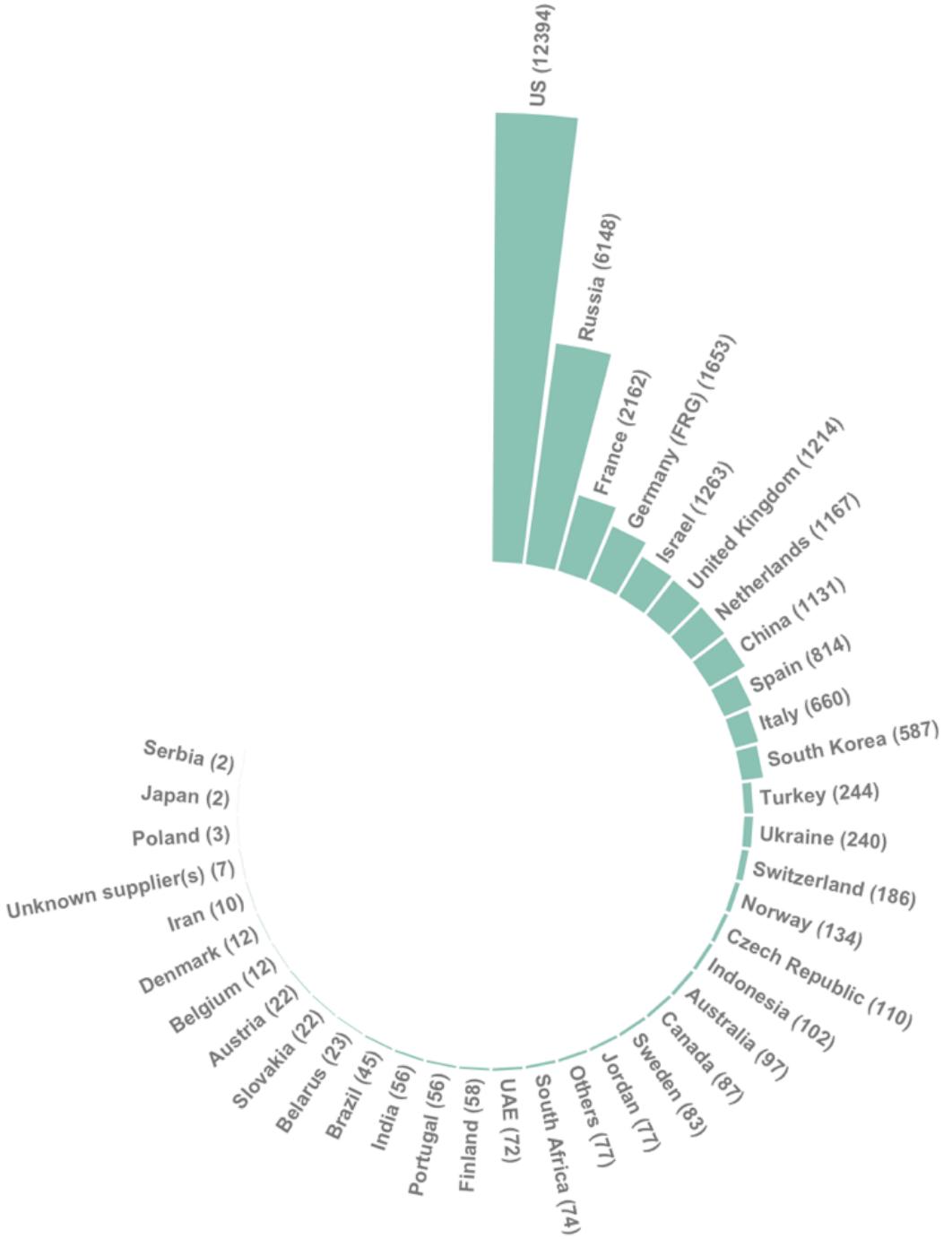
Who sells more weapons ?

| Country | Value |
|----------------|-------|
| United States | 12394 |
| Russia | 6148 |
| Germany (FRG) | 1653 |
| France | 2162 |
| United Kingdom | 1214 |
| China | 1131 |



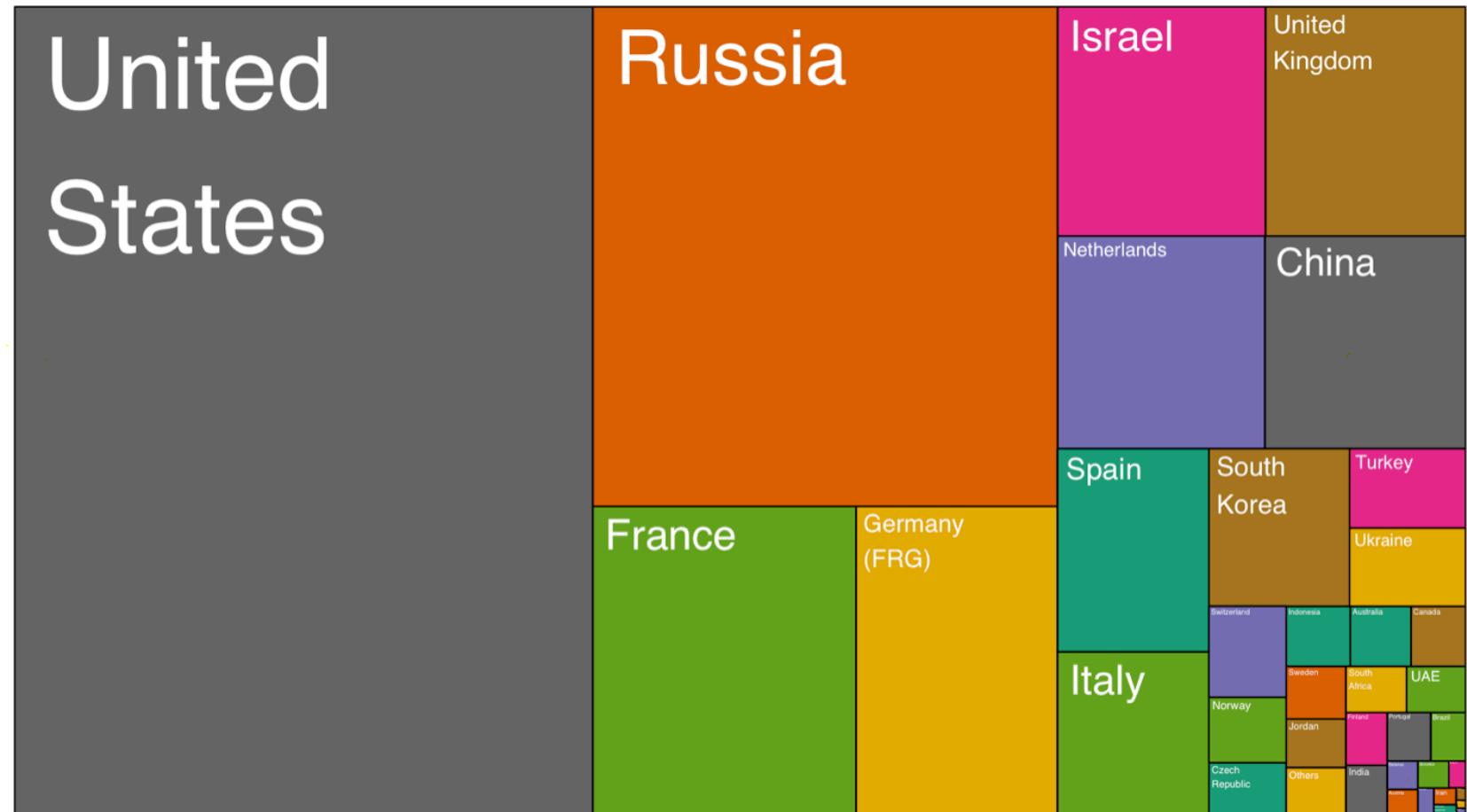
Who sells more weapons ?

| Country | Value |
|----------------|-------|
| United States | 12394 |
| Russia | 6148 |
| Germany (FRG) | 1653 |
| France | 2162 |
| United Kingdom | 1214 |
| China | 1131 |



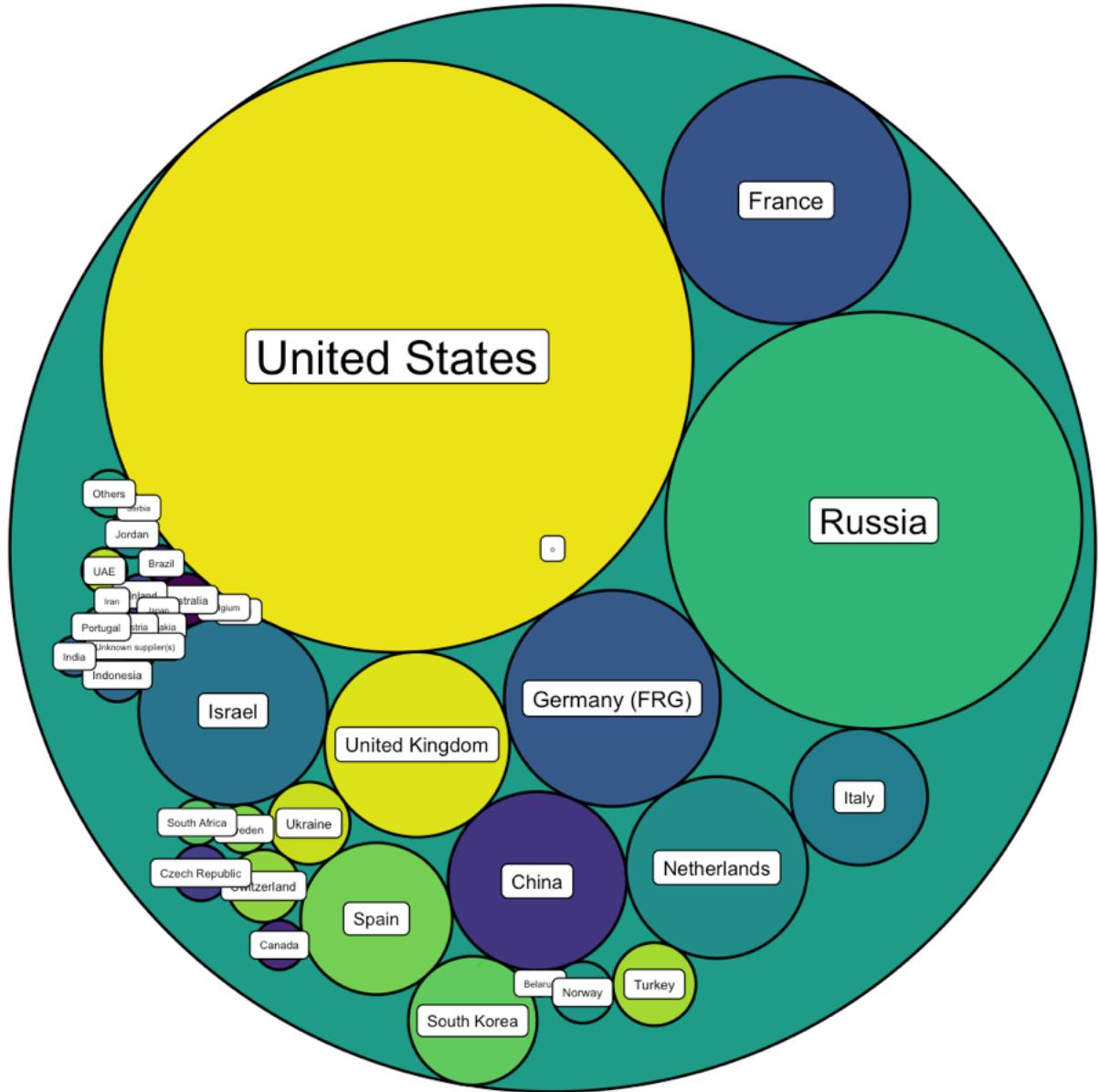
Who sells more weapons ?

| Country | Value |
|----------------|-------|
| United States | 12394 |
| Russia | 6148 |
| Germany (FRG) | 1653 |
| France | 2162 |
| United Kingdom | 1214 |
| China | 1131 |



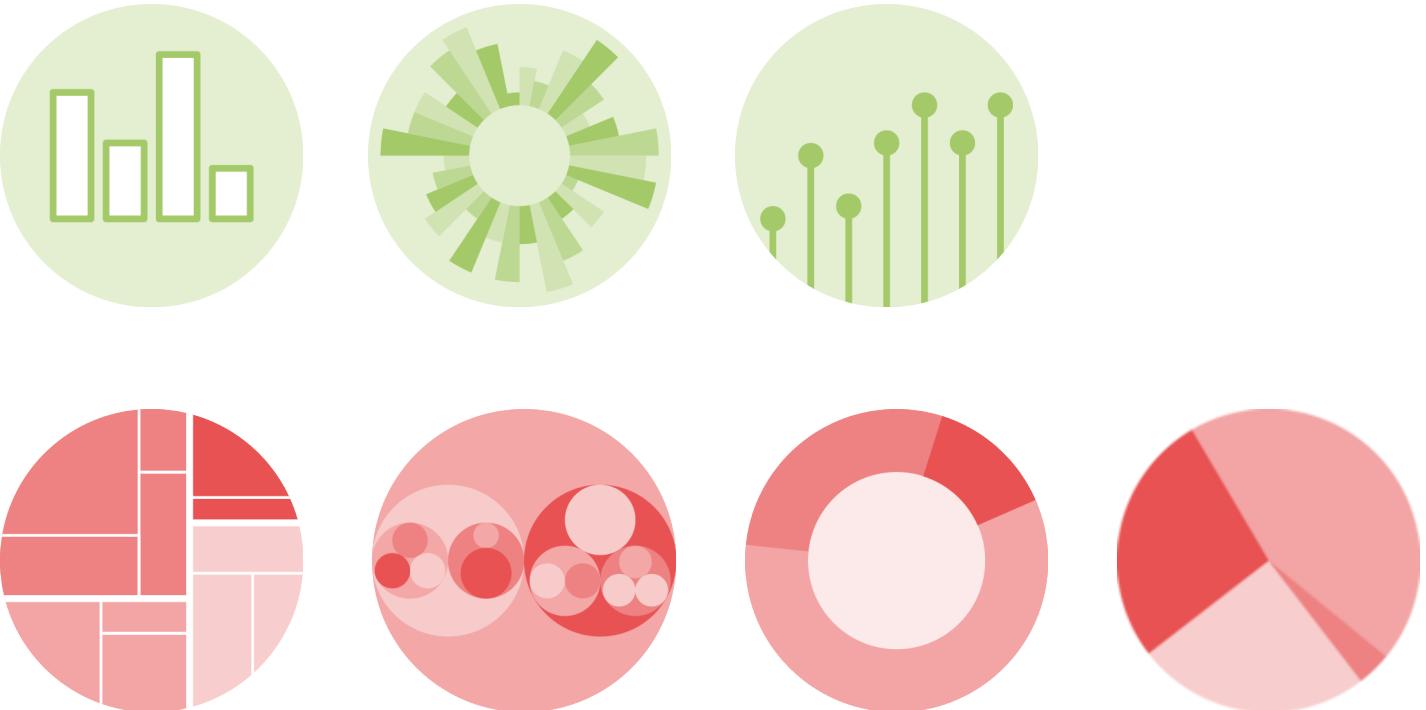
Who sells more weapons ?

| Country | Value |
|----------------|-------|
| United States | 12394 |
| Russia | 6148 |
| Germany (FRG) | 1653 |
| France | 2162 |
| United Kingdom | 1214 |
| China | 1131 |



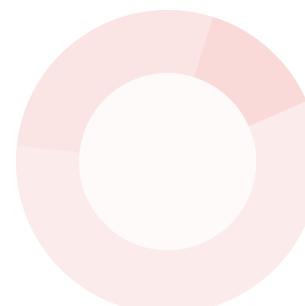
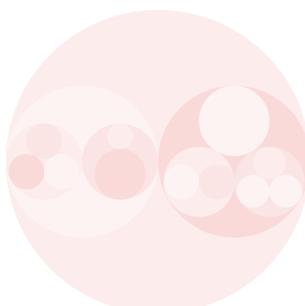
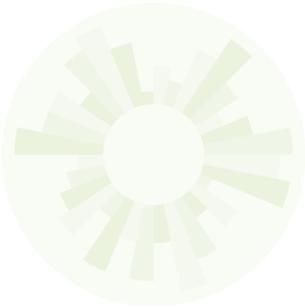
Who sells more weapons ?

| Country | Value |
|----------------|-------|
| United States | 12394 |
| Russia | 6148 |
| Germany (FRG) | 1653 |
| France | 2162 |
| United Kingdom | 1214 |
| China | 1131 |



Who sells more weapons ?

| Country | Value |
|----------------|-------|
| United States | 12394 |
| Russia | 6148 |
| Germany (FRG) | 1653 |
| France | 2162 |
| United Kingdom | 1214 |
| China | 1131 |



1 categorical variable

1 numerical variable

1 observation per group

Perception of probability

| text | value |
|------------------|-------|
| Improbable | 33 |
| Almost Certainly | 98 |
| Likely | 60 |
| Almost Certainly | 98 |
| Unlikely | 10 |
| Probably Not | 25 |
| About Even | 50 |
| Probably | 75 |

Perception of probability

| text | value |
|------------------|-------|
| Improbable | 33 |
| Almost Certainly | 98 |
| Likely | 60 |
| Almost Certainly | 98 |
| Unlikely | 10 |
| Probably Not | 25 |
| About Even | 50 |
| Probably | 75 |

Several
observations
per group

1 categorical
variable

1 numeric
variable

Perception of probability

text **value**

Improbable 33

Almost Certainly 98

Likely 60

Almost Certainly 98

Unlikely 10

Probably Not 25

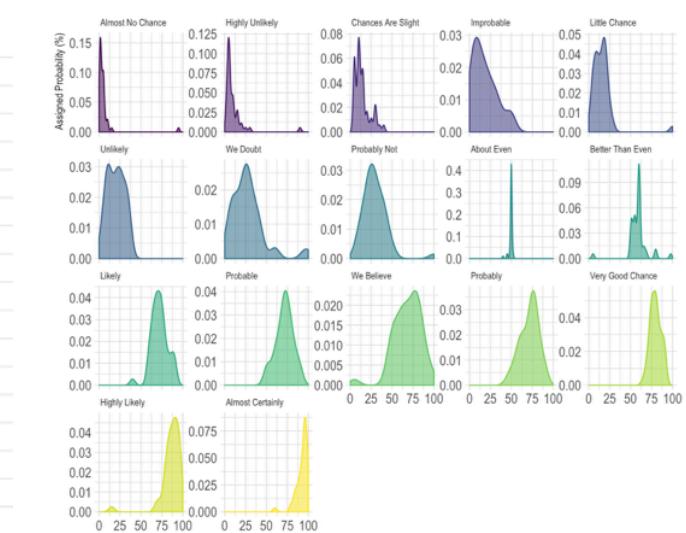
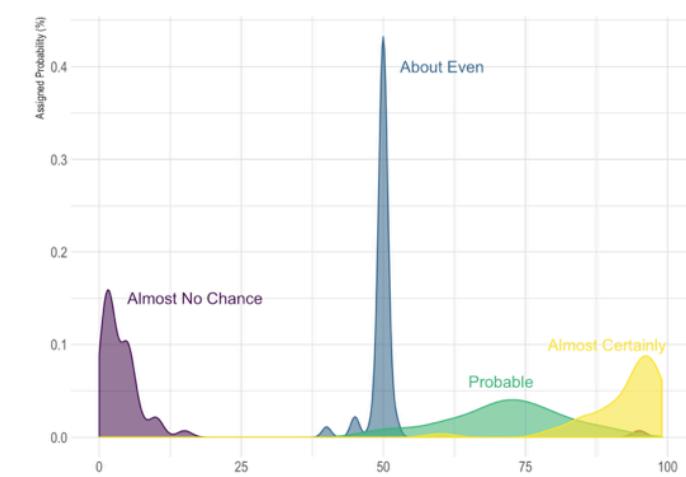
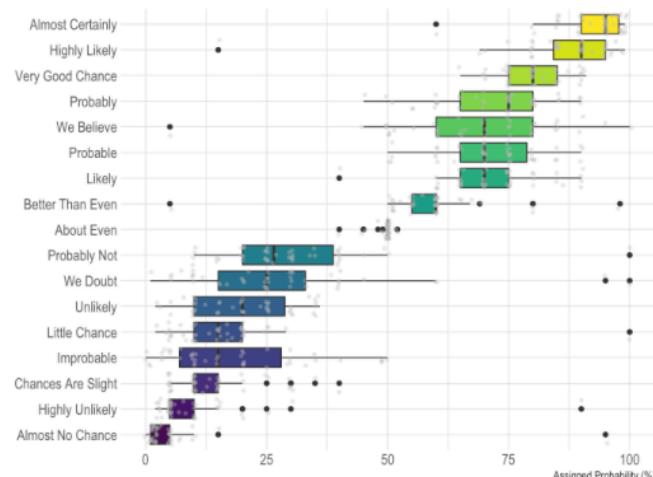
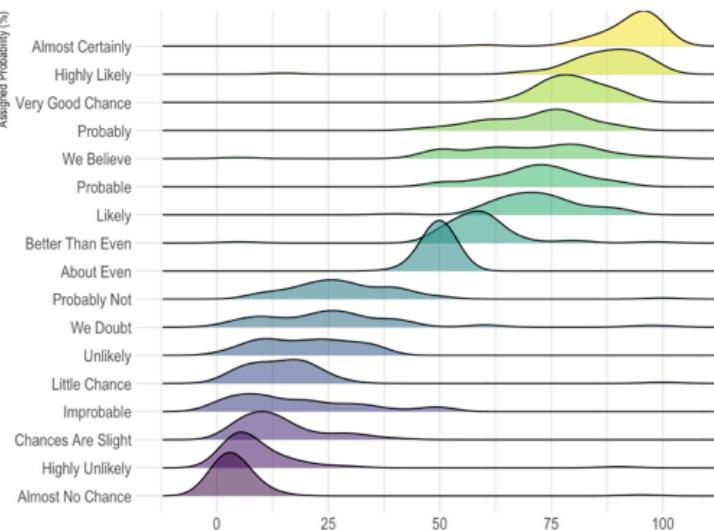
About Even 50

Probably 75

1 categorical
variable

1 numeric
variable

Several
observation
per group



So...

- Knowing the **possibilities** is the **first step** in chart choice
- Not easy to know **all** the **chart types**
- Hard to figure out options **from a dataset**



Let's build a decision tree



from Data to Viz

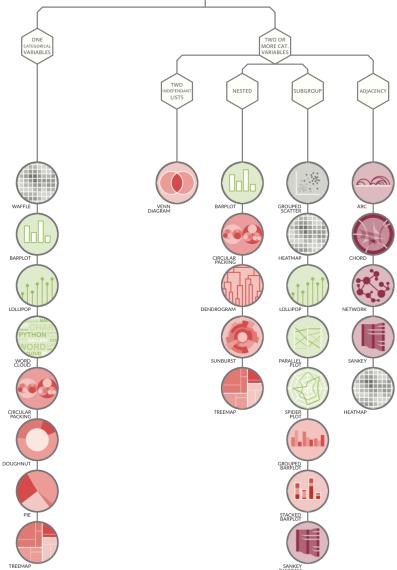
'From Data to Viz' is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps :

- 1 Identify what type of data you have.
 - 2 Go to the corresponding decision tree and follow it down to a set of possible charts.
 - 3 Choose the chart from the set that will suit your data and your needs best.

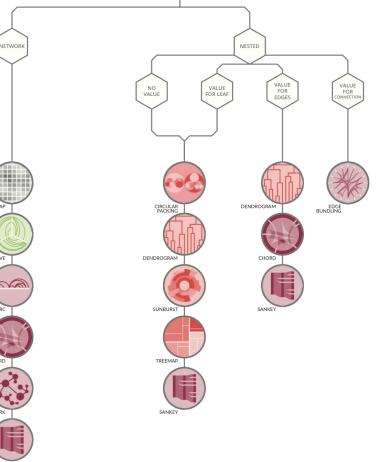
Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more, visit:

data-to-viz.com

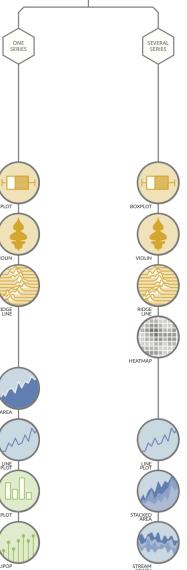
CATEGORIC



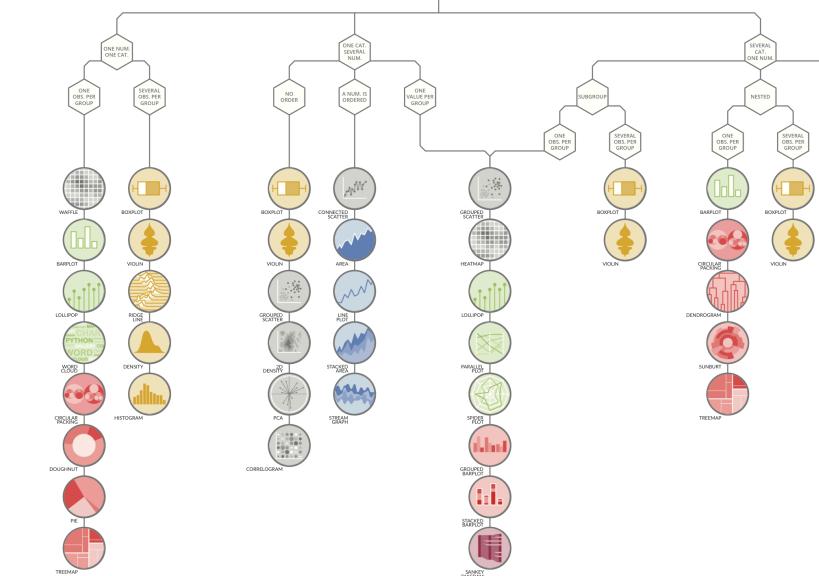
RELATIONAL



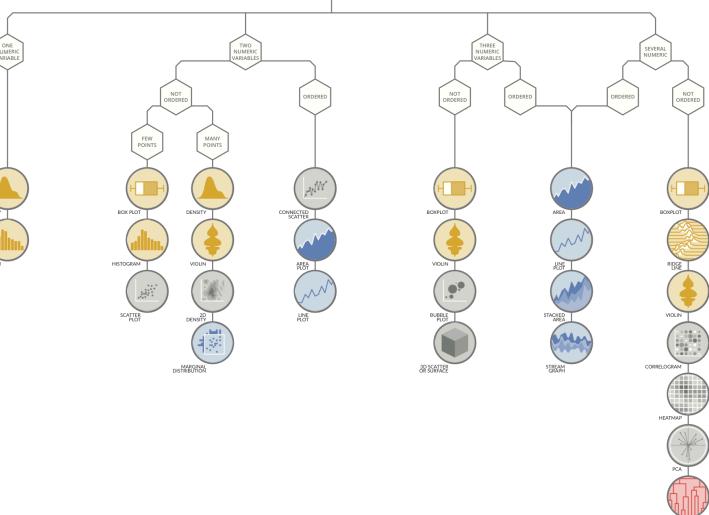
TIME SERIES



CATEGORIC AND NUMERIC

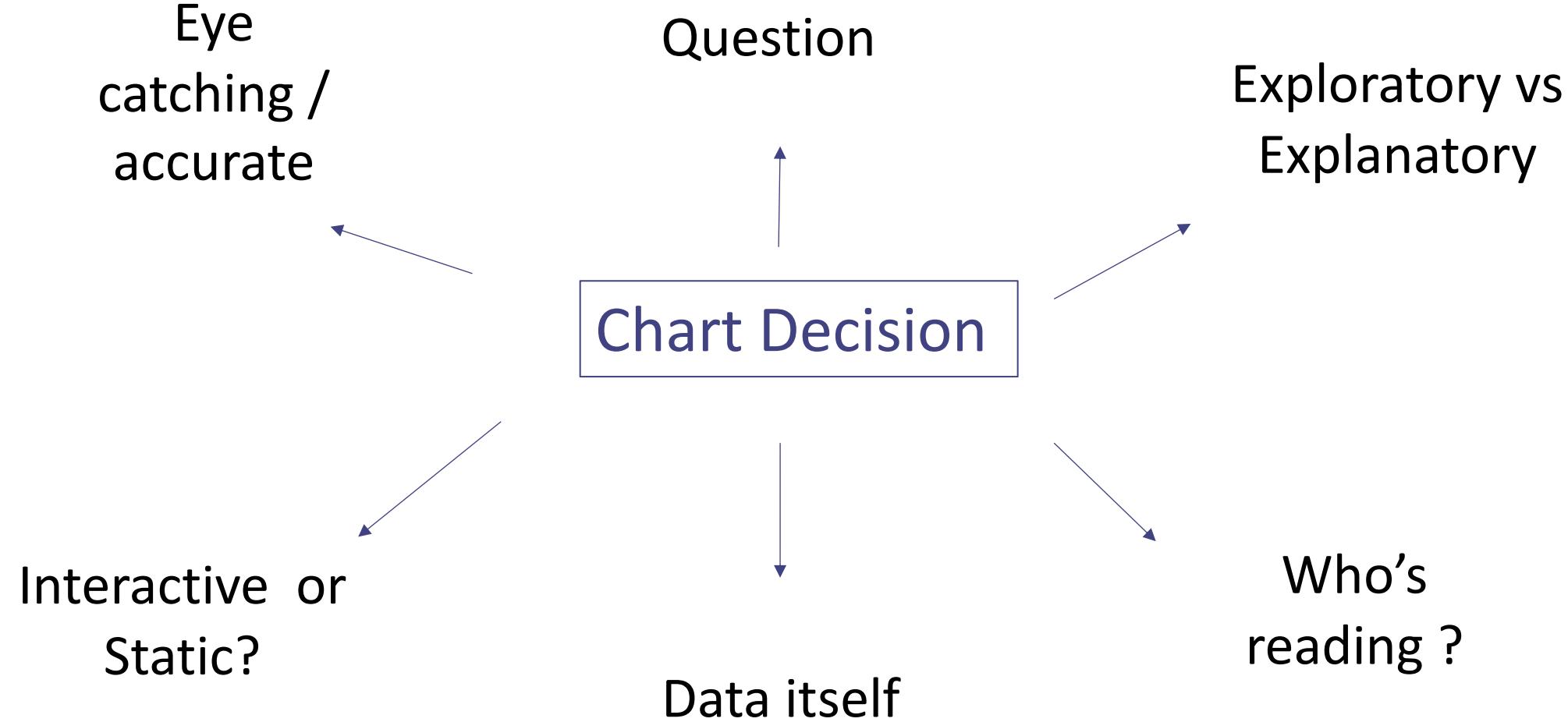


NUMERIC



WHAT YOU SHOULD DO

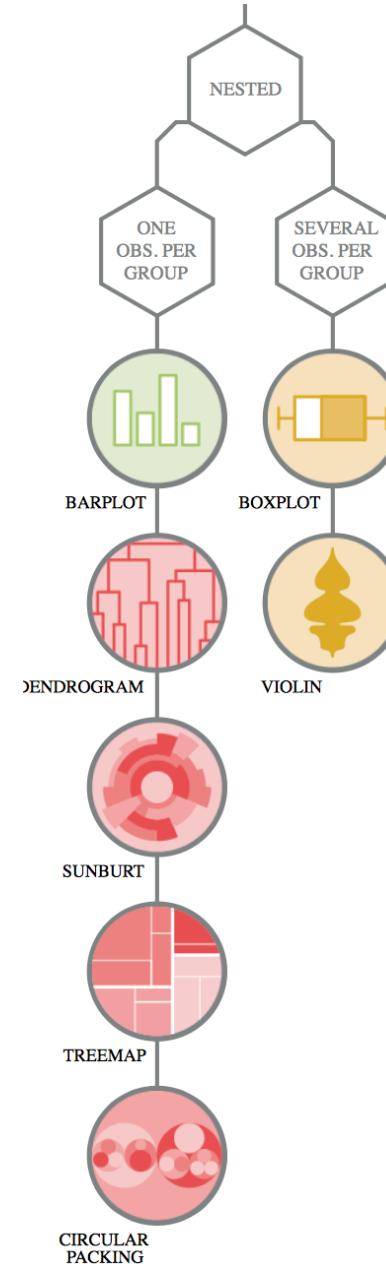
About 20 examples of storytelling with data



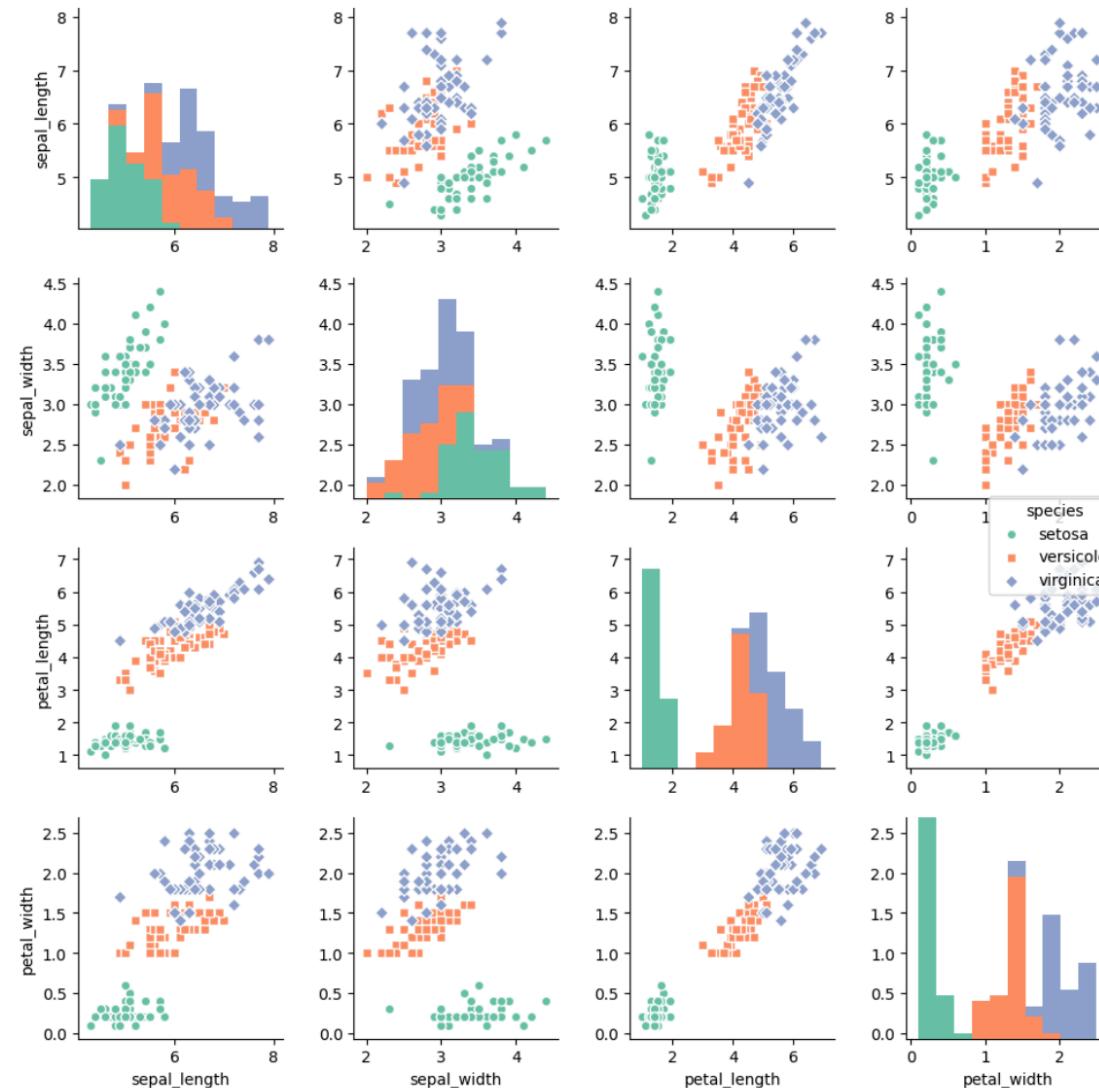
Question – Explo/Expla – Reader – Data – Interactivity – Eye catching

WHAT DO YOU WANT TO SHOW ?

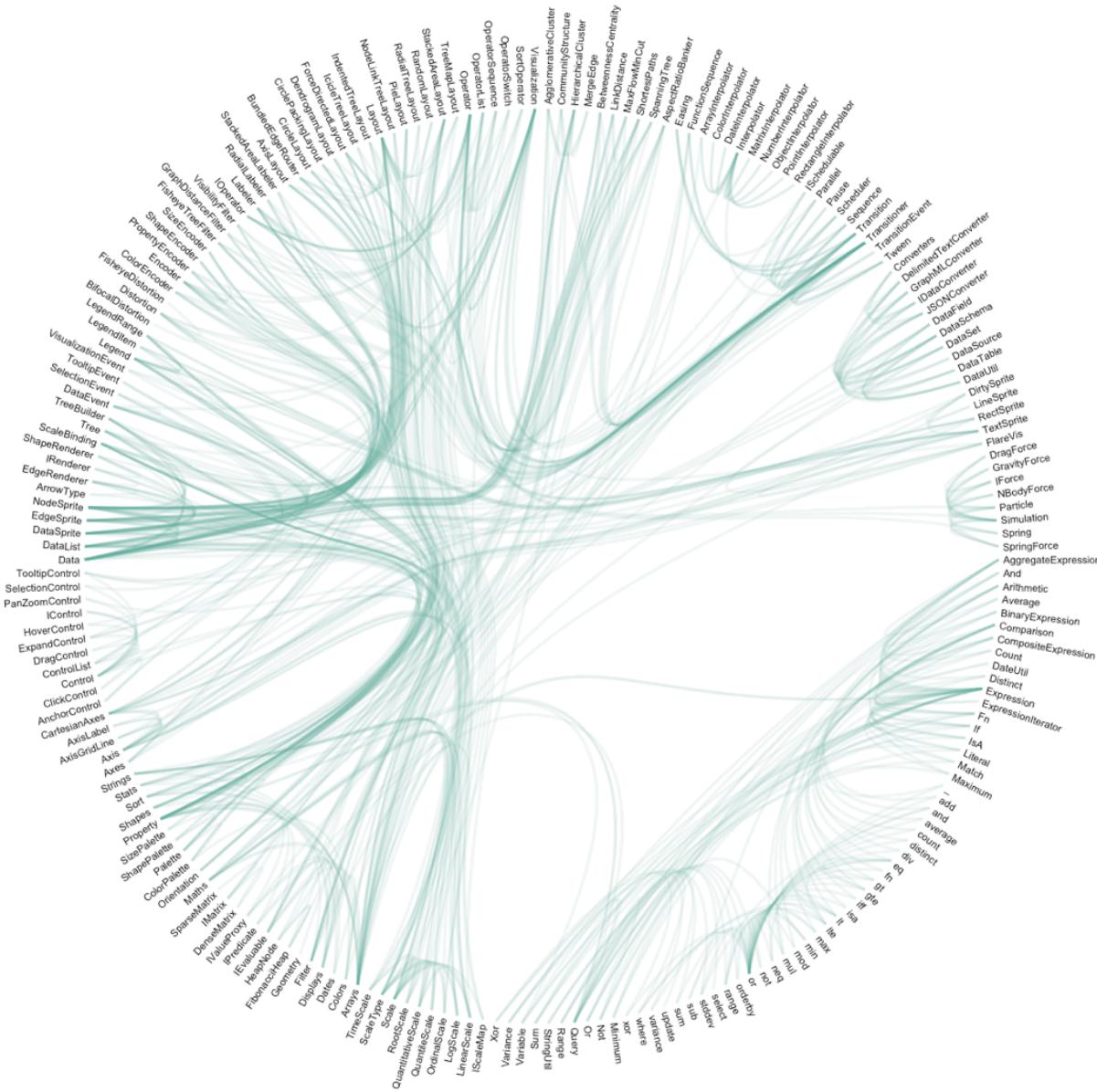
- Distribution
- Evolution
- Correlation
- Maps
- Ranking
- Flow
- Part of a whole



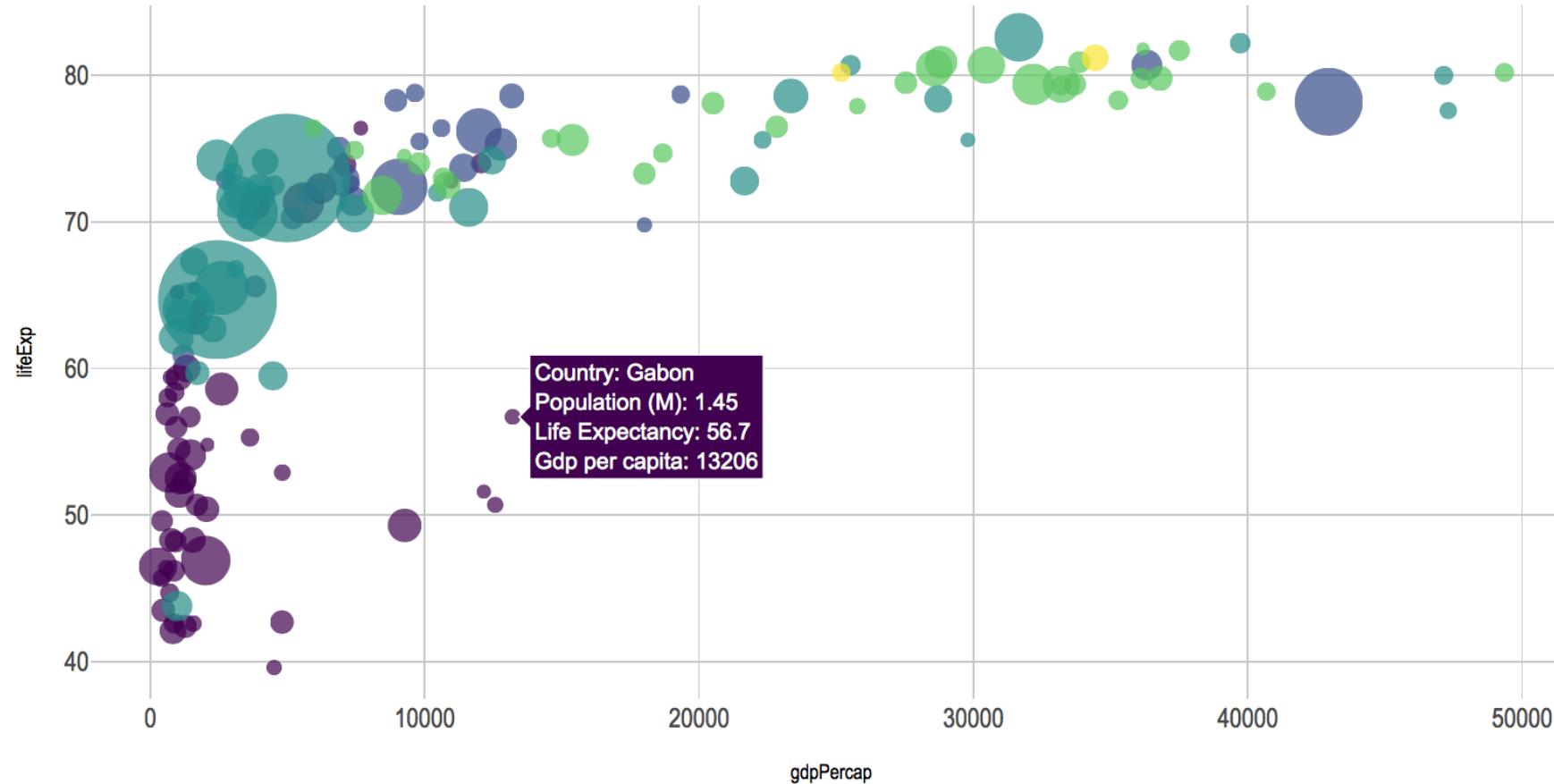
Question – **Exploratory Data Analysis** – Reader – Data – Interactivity – Eye catching



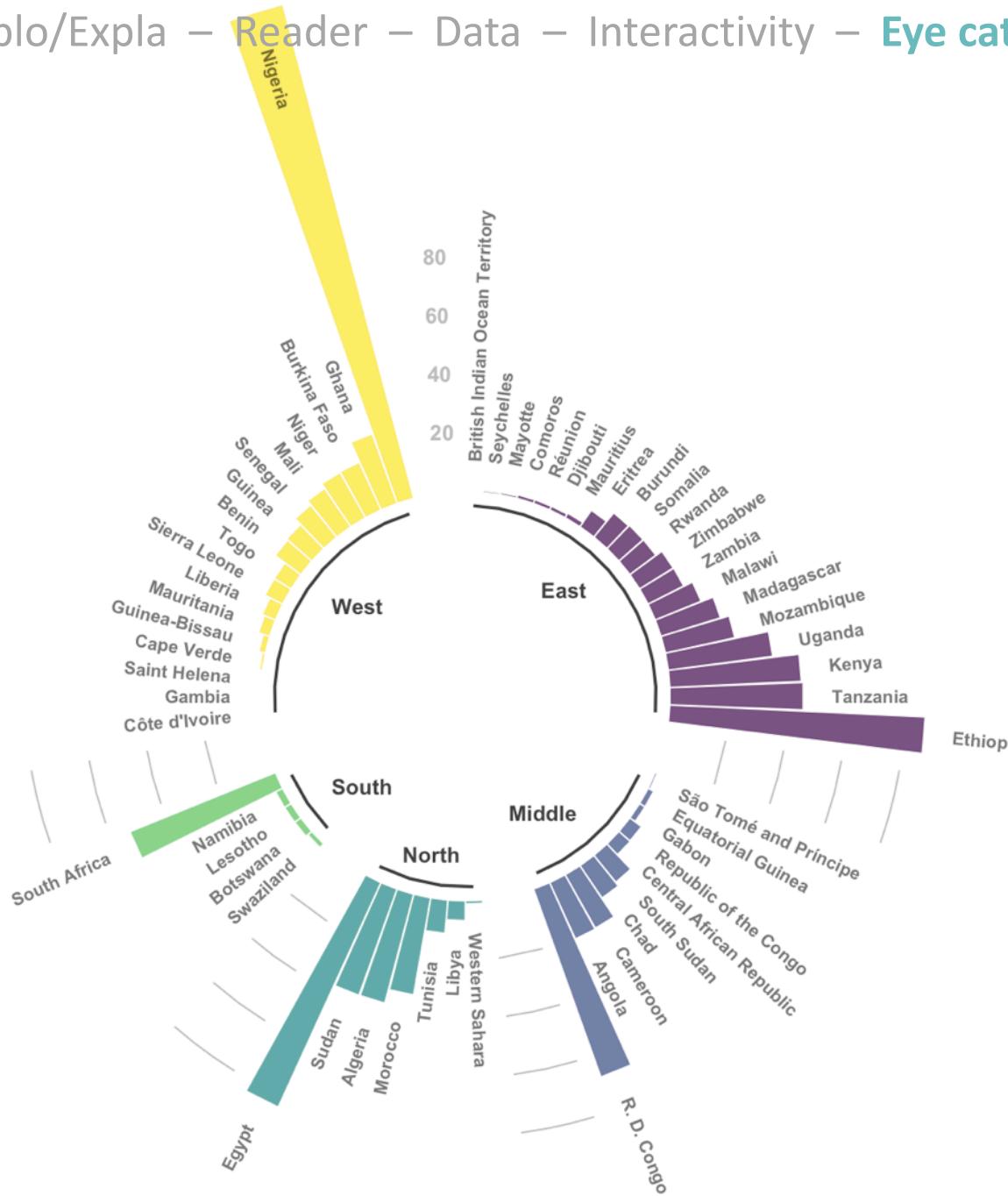
Question – Explo/Expla – **Reader** – Data – Interactivity – Eye catching

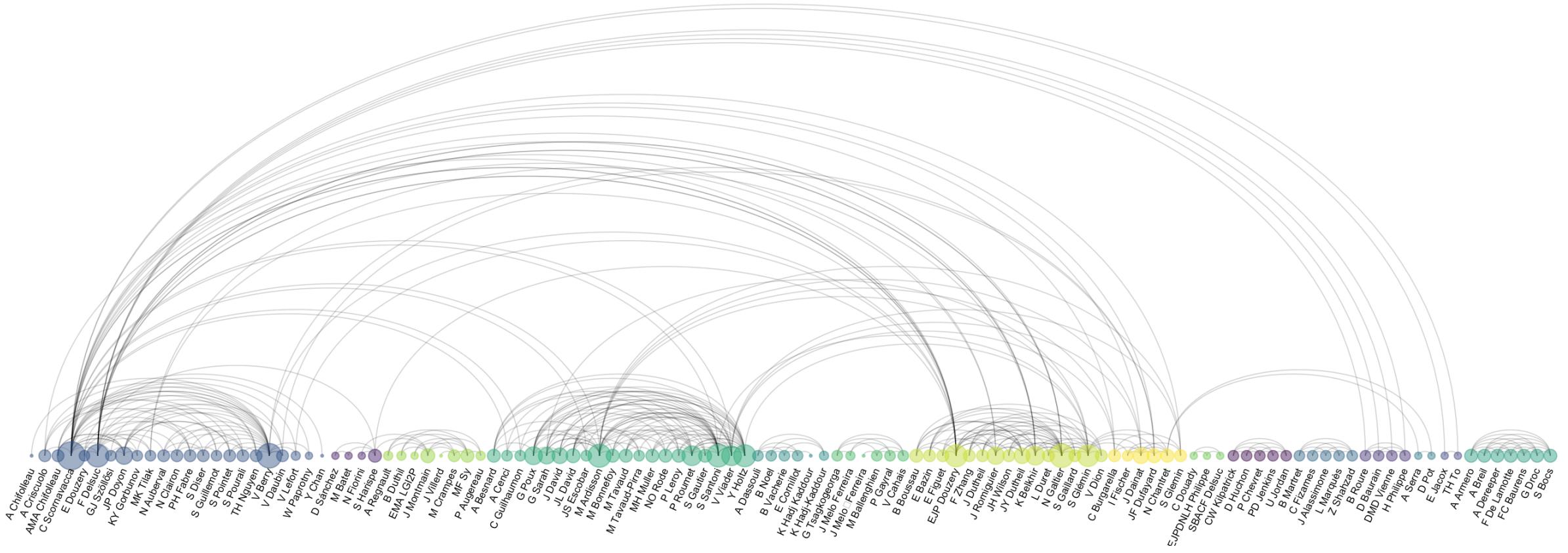


Question – Explo/Expla – Reader – Data – **Interactivity** – Eye catching



Question – Explo/Expla – Reader – Data – Interactivity – Eye catching





Co-authorship network of a researcher



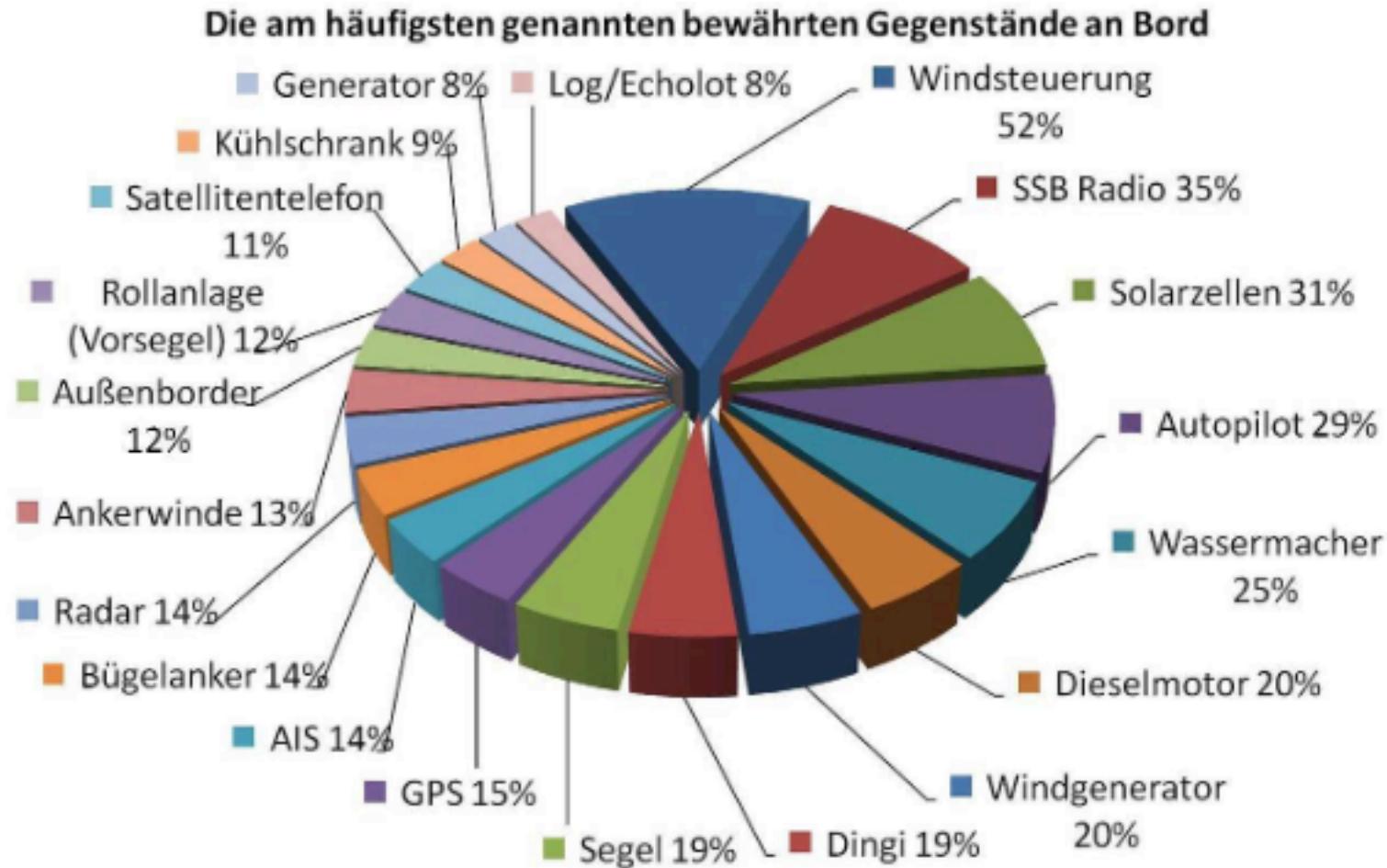
Where surfers travel.

data-to-viz.com | NASA.gov | 10,000 #surf tweets recovered

WHAT YOU SHOULD NOT DO

A gallery of most common caveats

What's wrong with
this chart?



Source: [WTF Visualizations](#)

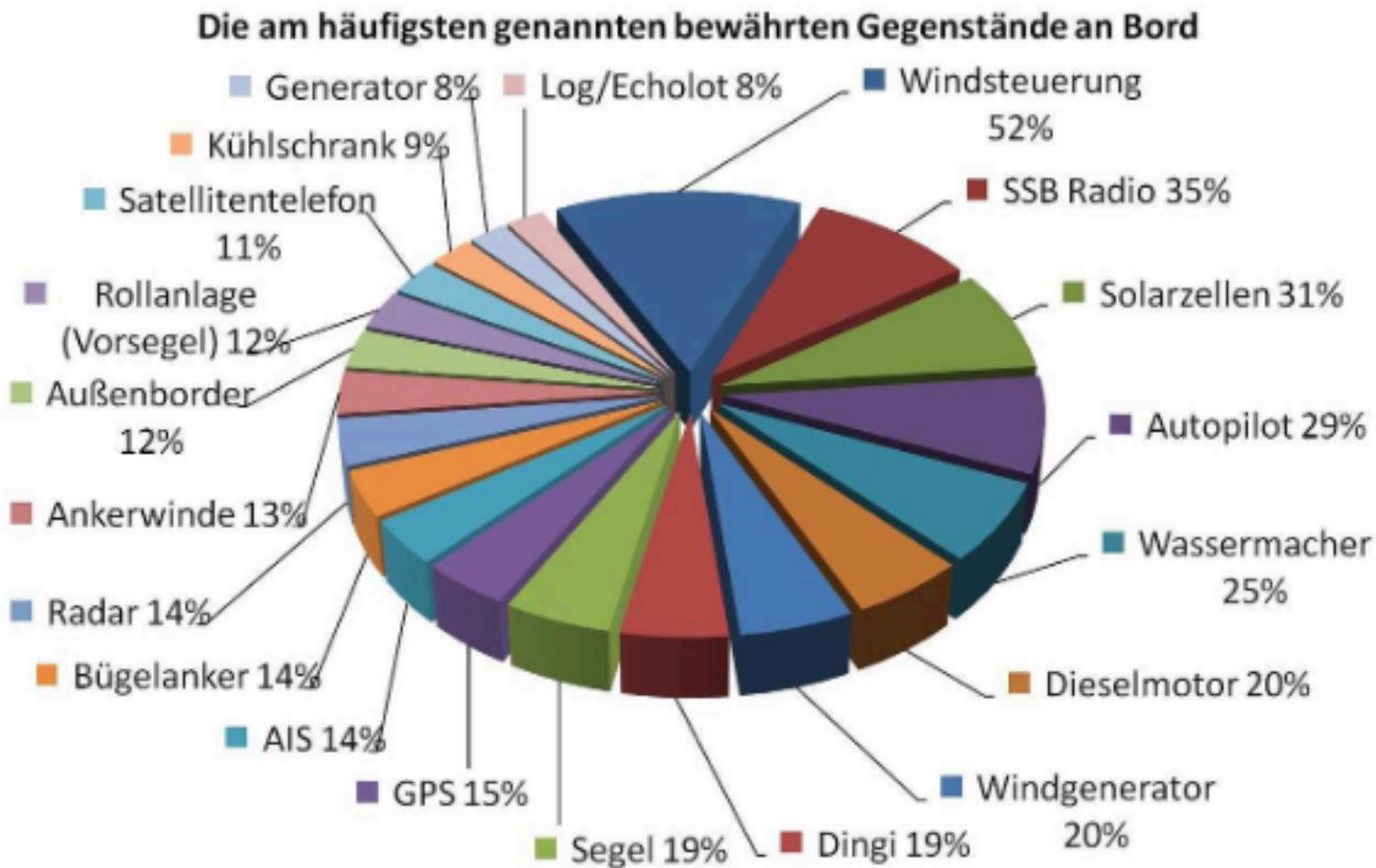
Calculation

+

3D

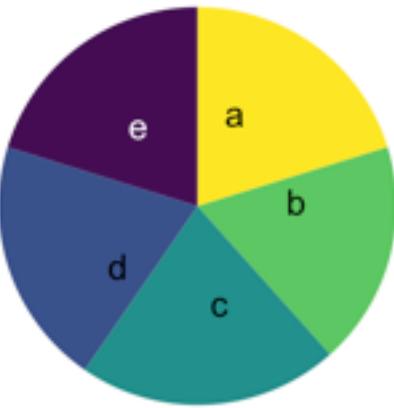
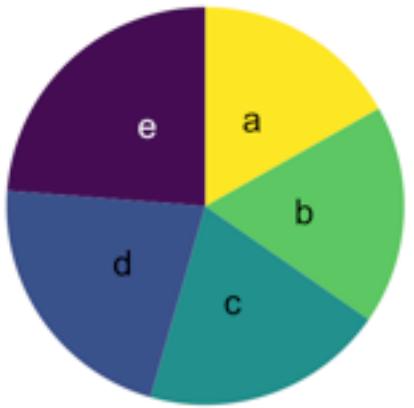
+

Pie chart



Source: [WTF Visualizations](#)

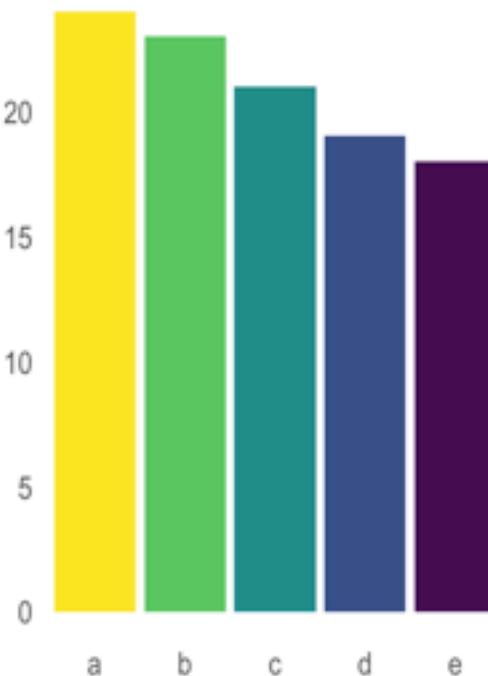
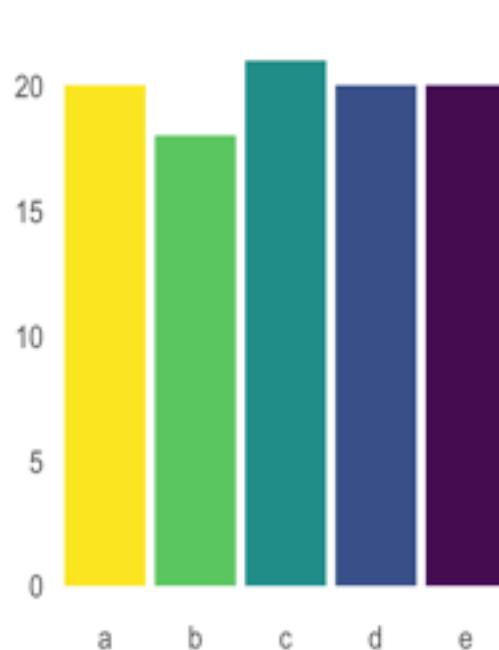
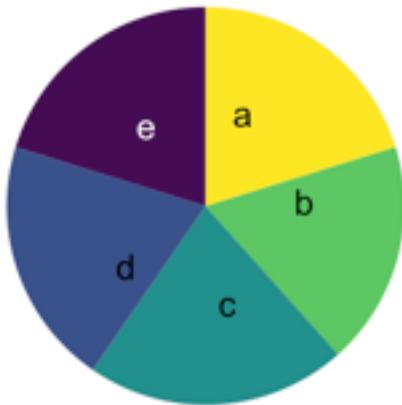
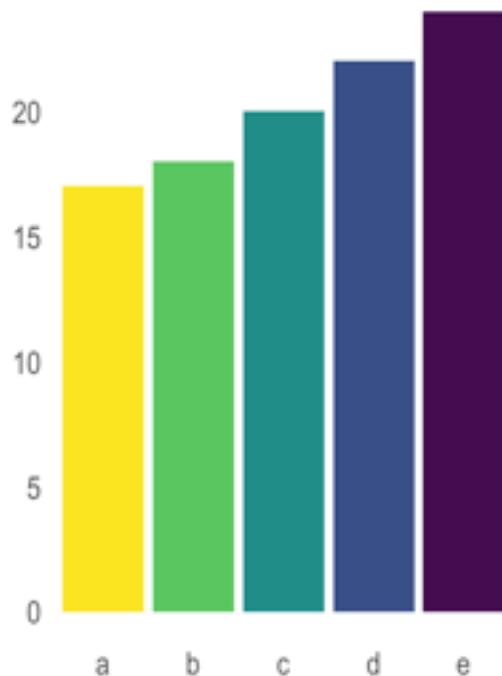
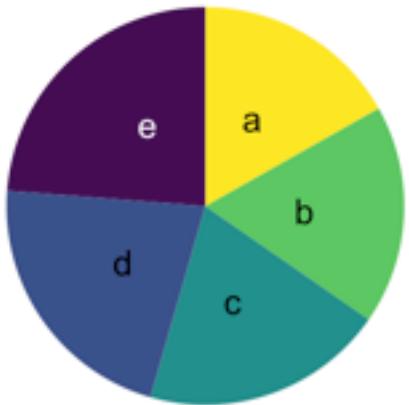
What's wrong with pie chart?



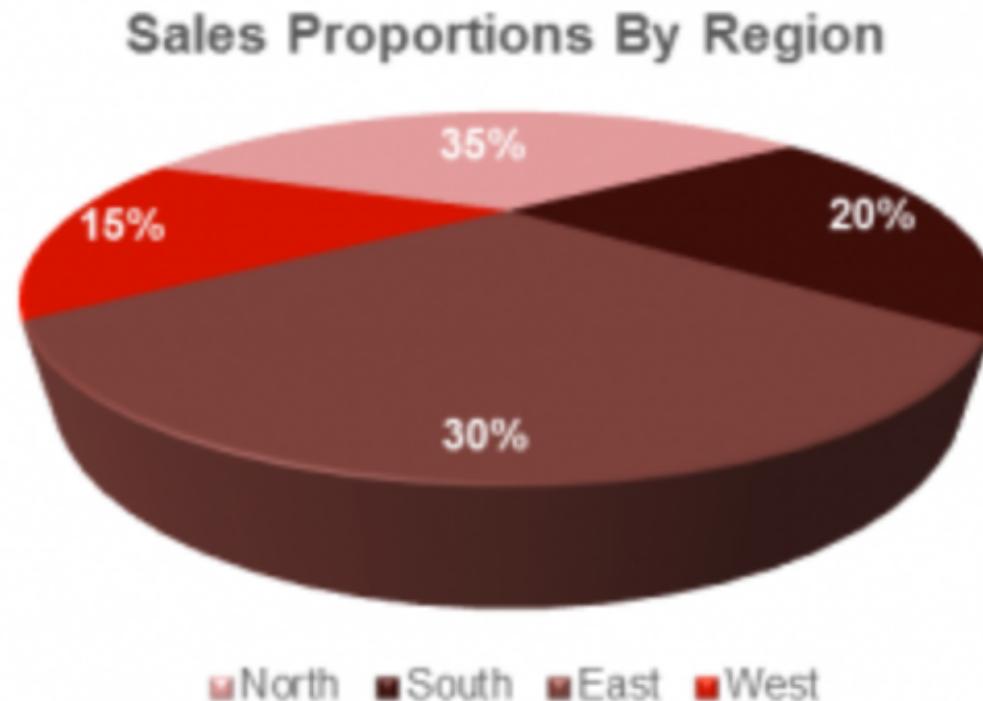
What can you see?

What's wrong with
pie chart?

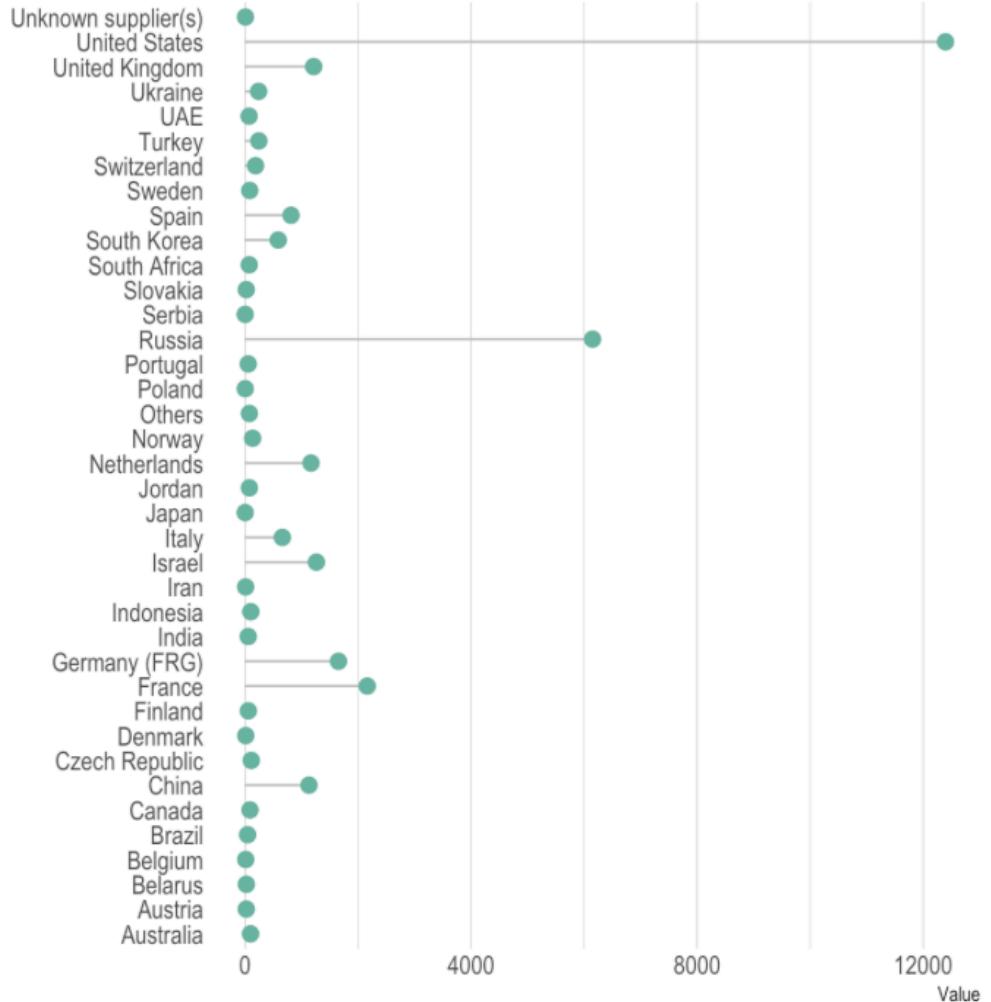
It is hard to
distinguish angles



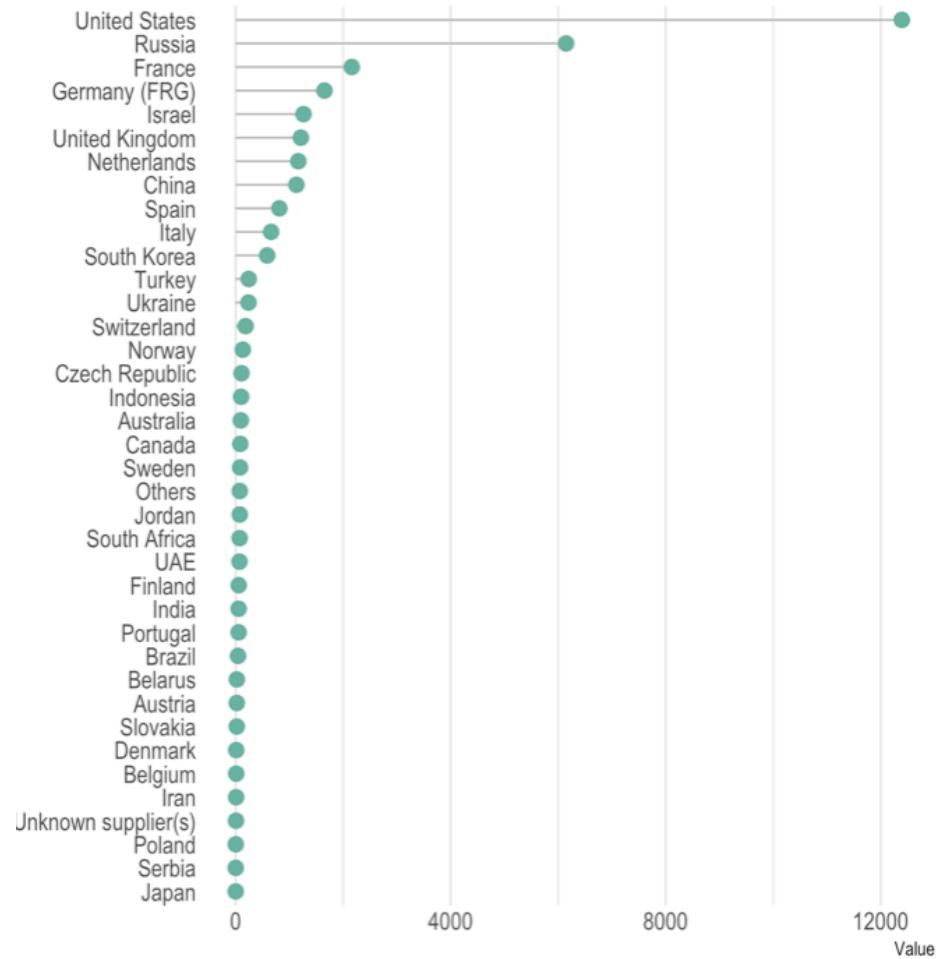
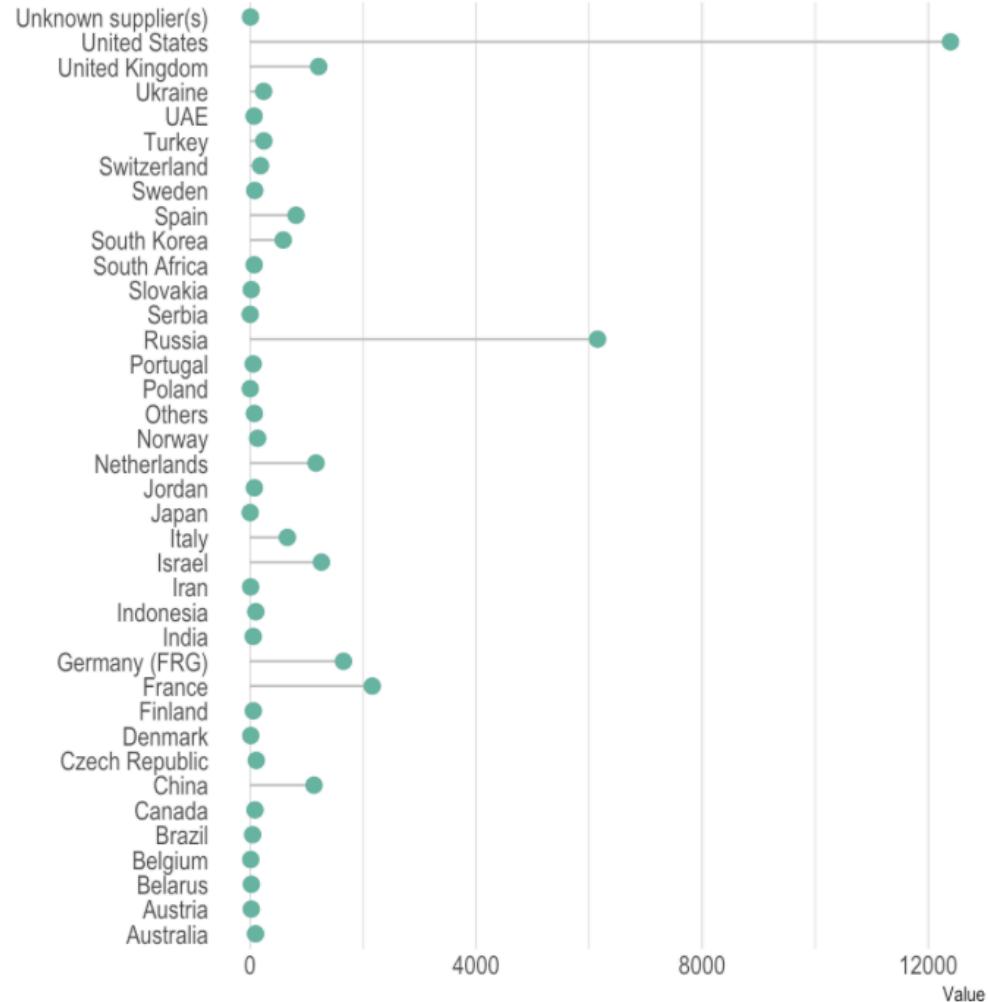
What's wrong with 3D



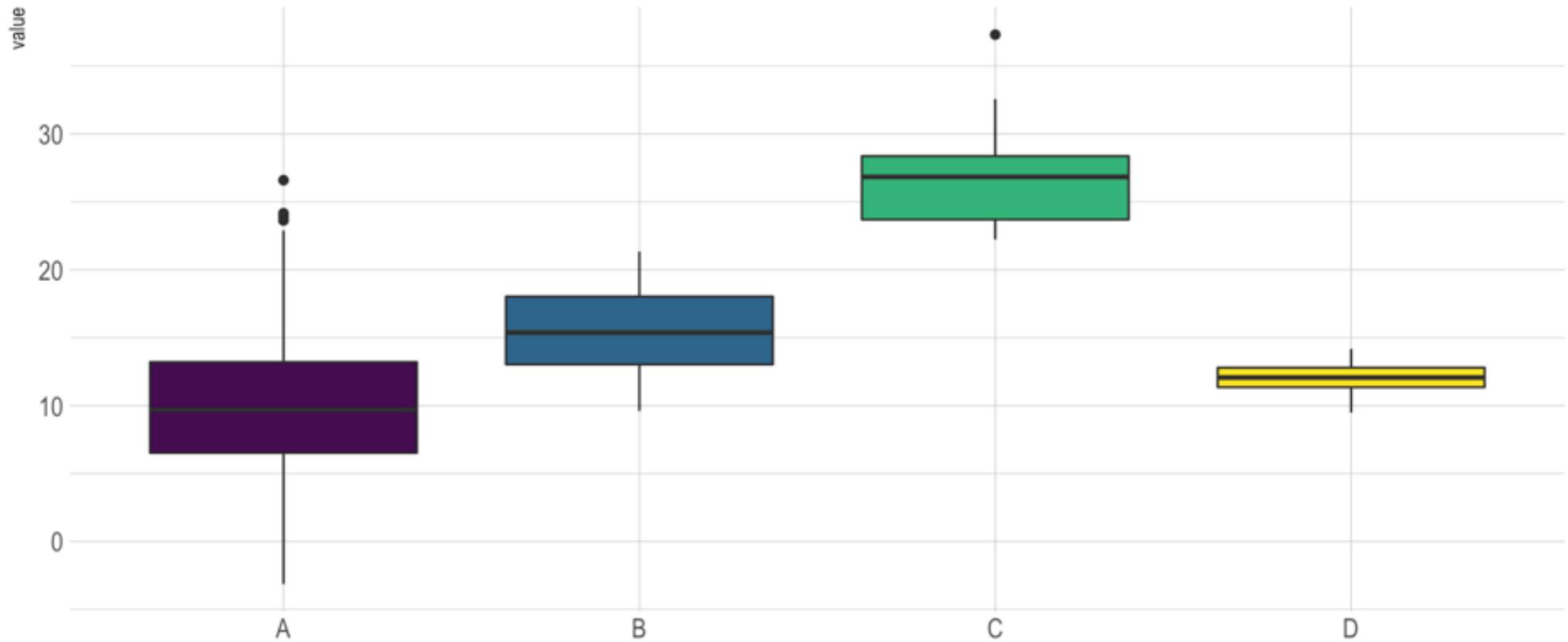
What could be better
here?



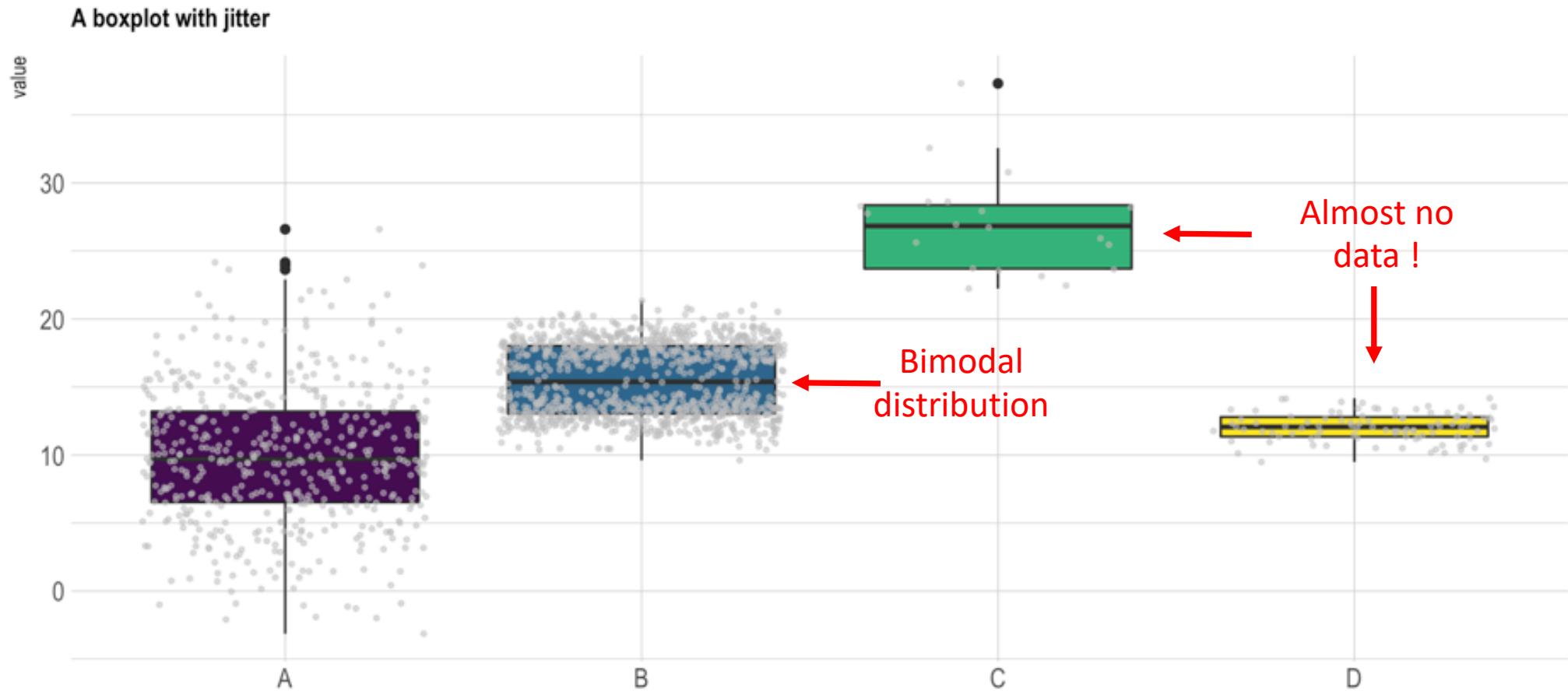
Order your data



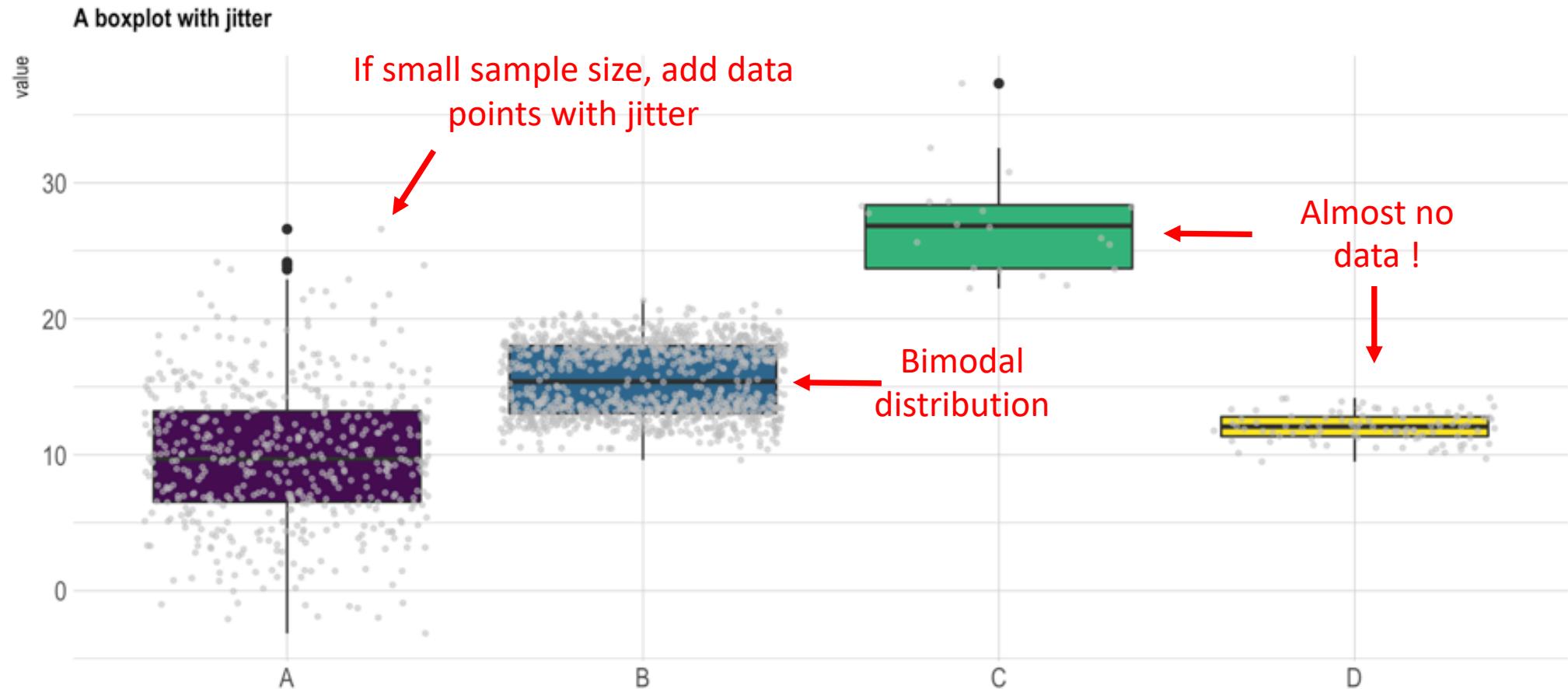
Anything wrong here?



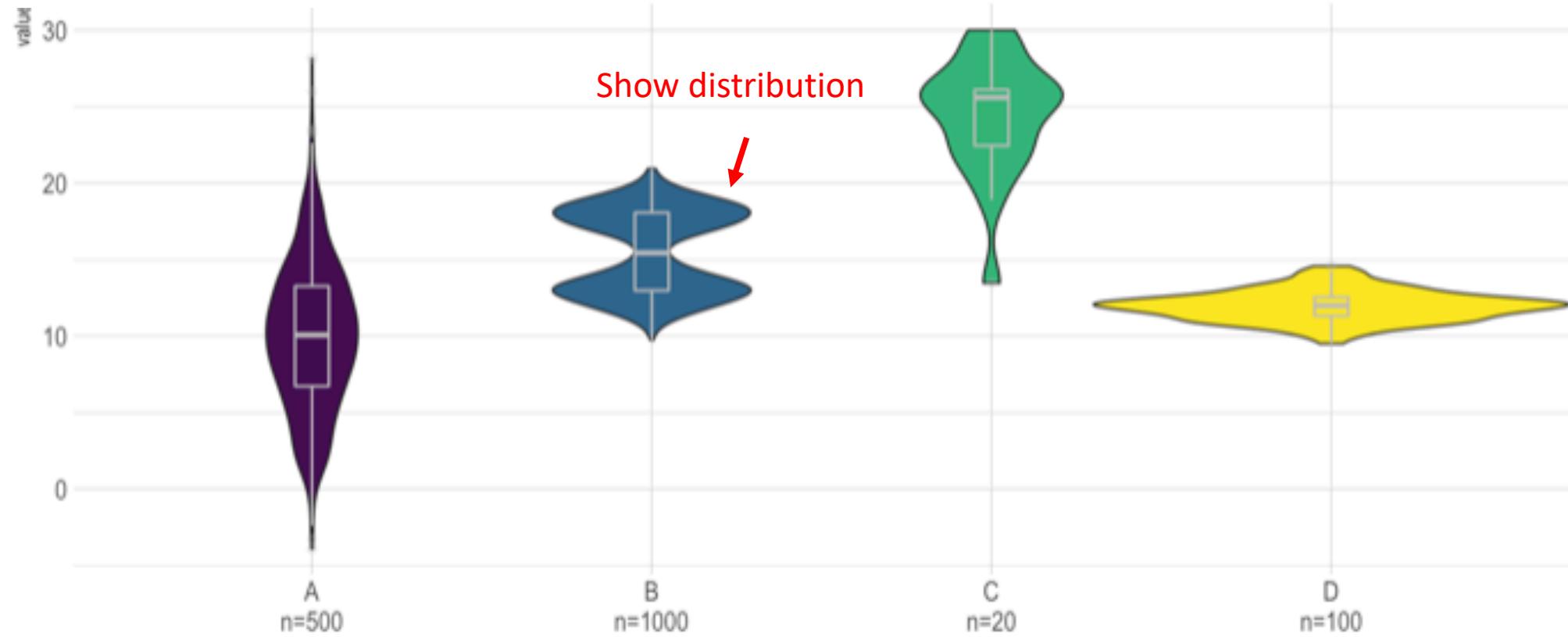
Boxplot = hide information



Boxplot = hide information

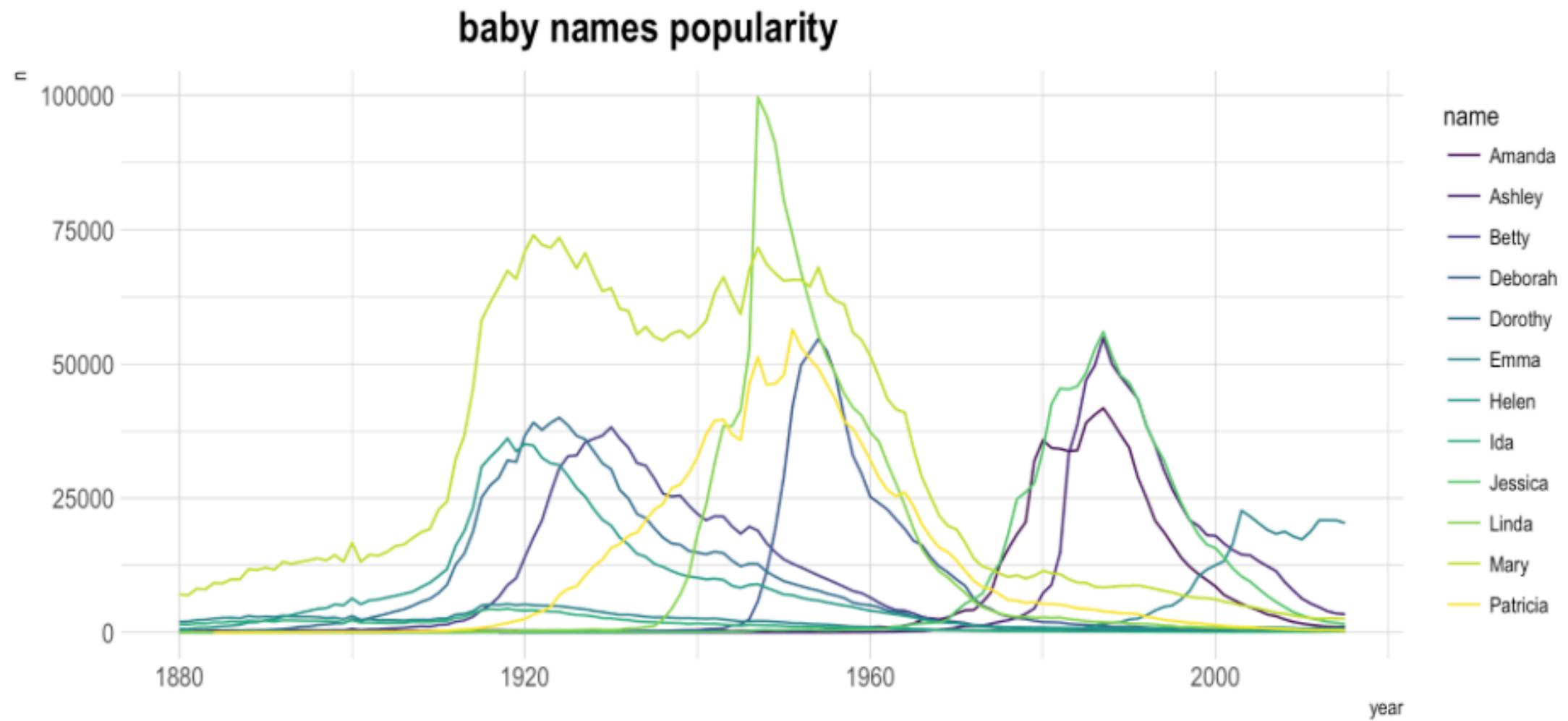


If big sample size, use violin plot



Show sample size

Look at this chart

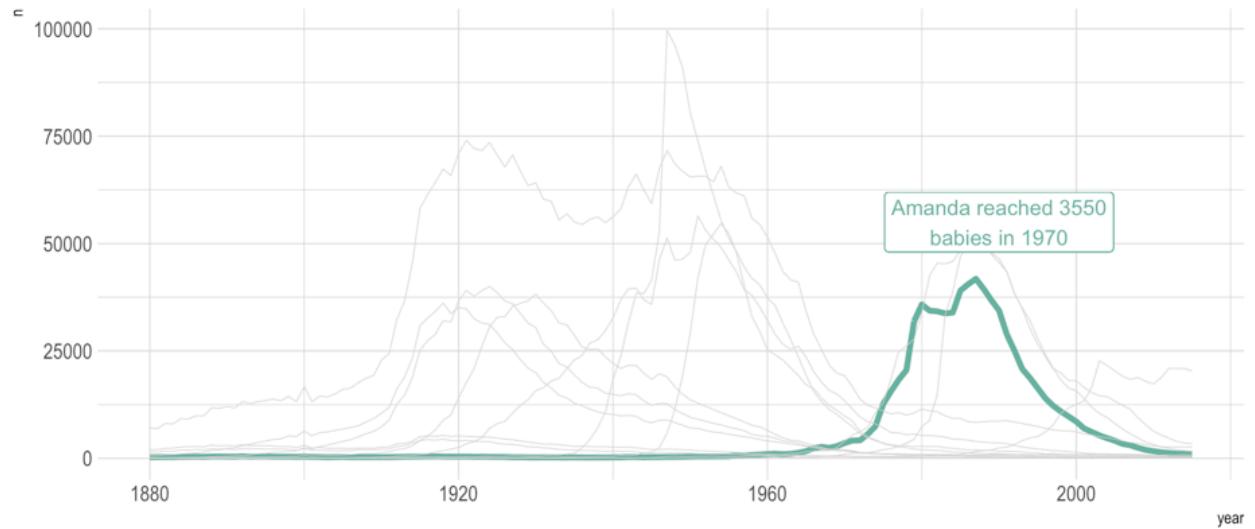


What do you remember?

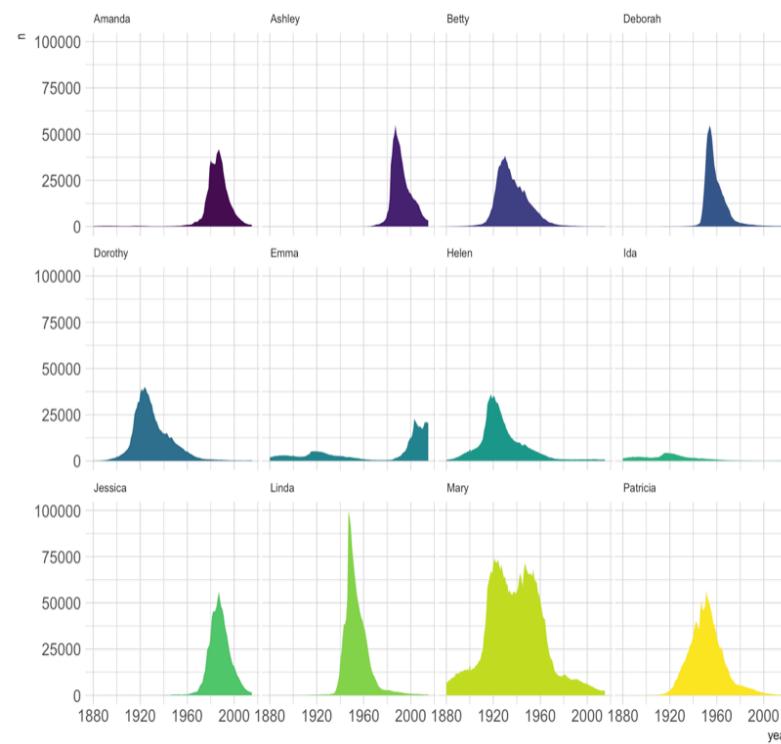
The Spaghetti chart

[Read more](#)

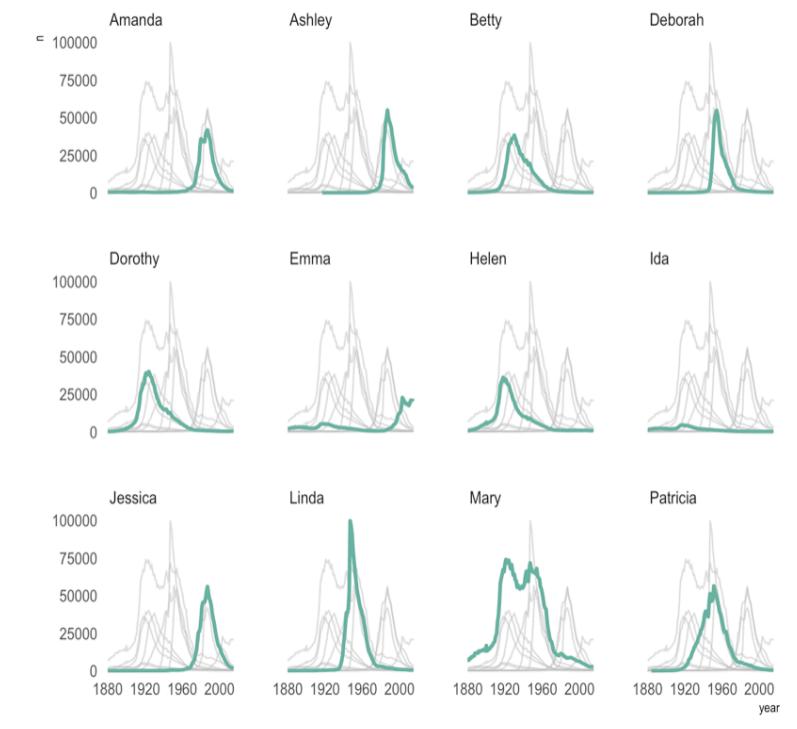
Popularity of American names in the previous 30 years



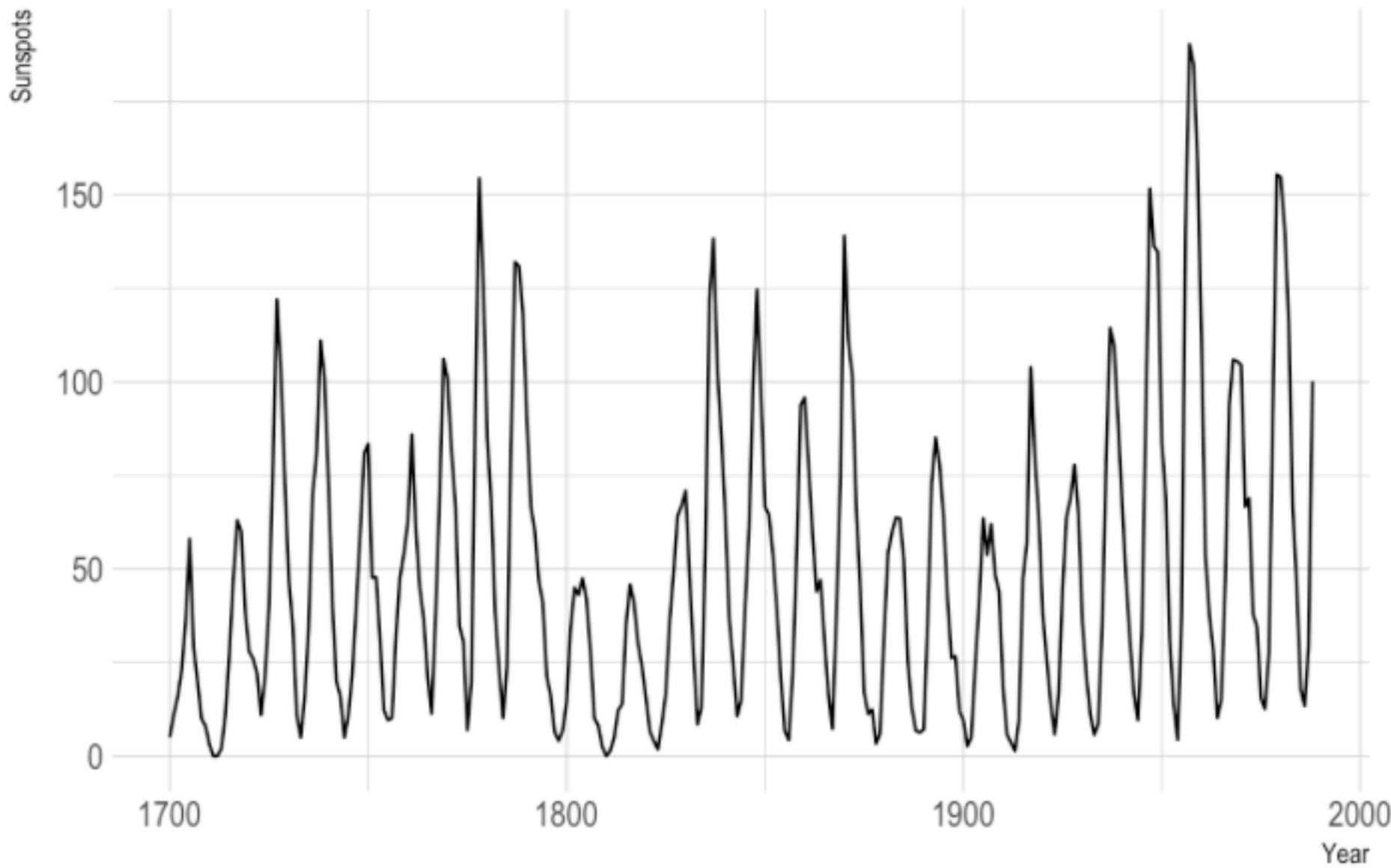
Popularity of American names in the previous 30 years



A spaghetti chart of baby names popularity

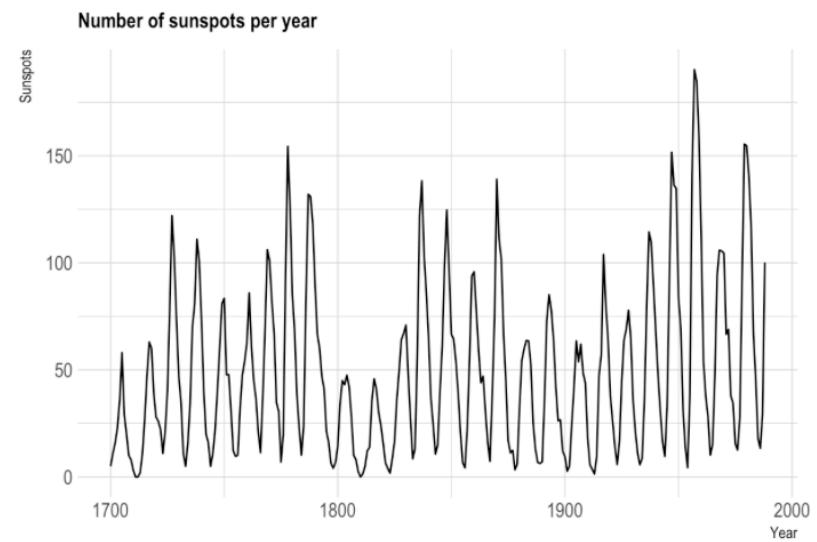


Number of sunspots per year



Mind the aspect ratio

[Read more](#)

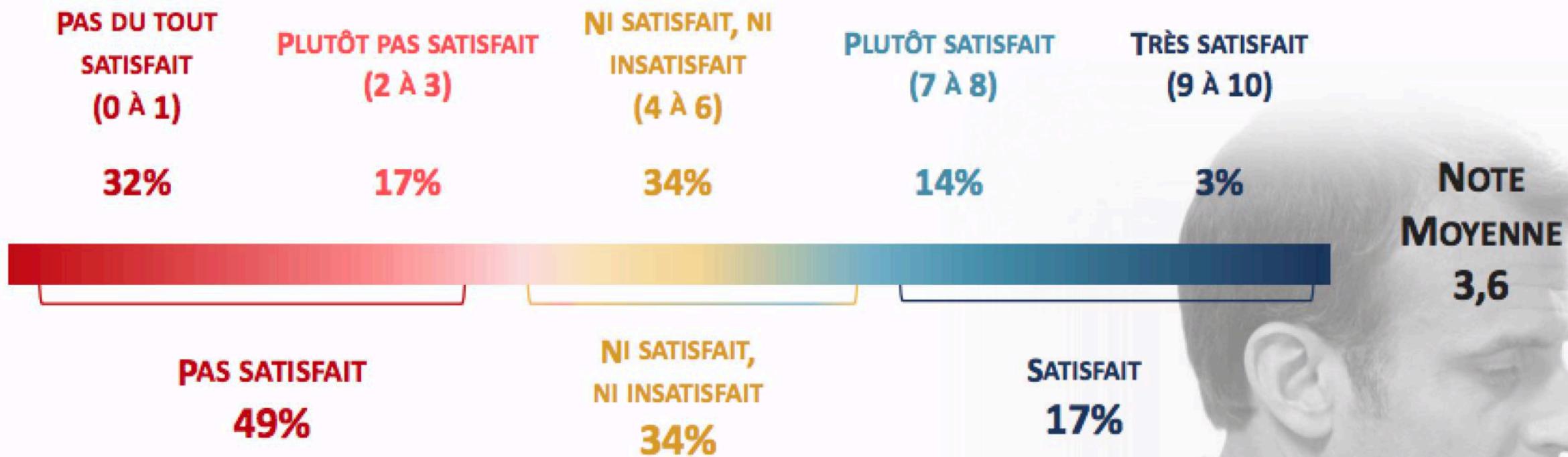


Number of sunspots per year



La satisfaction à l'égard de l'action d'Emmanuel Macron (1/3)

D'une manière générale, êtes-vous satisfait ou pas de l'action du président de la République Emmanuel Macron ?



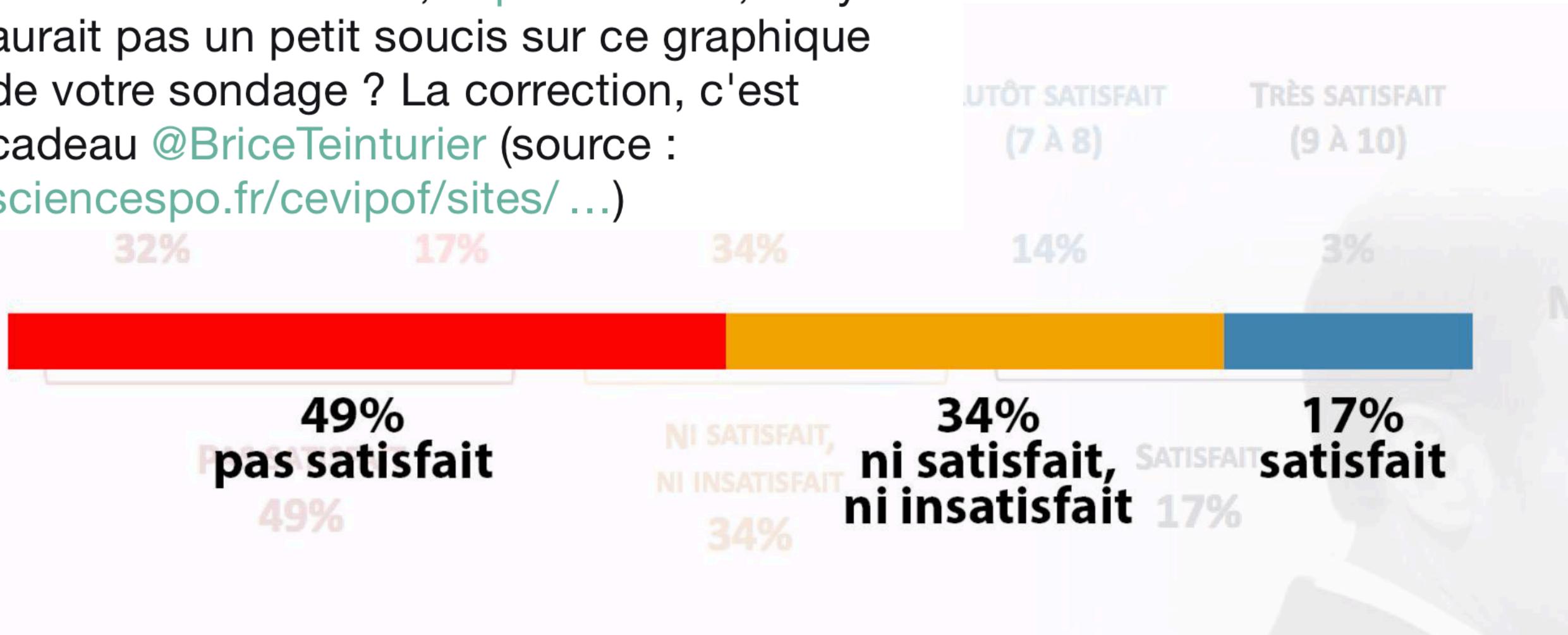


Nicolas Boeuf
@Nicolas_Boeuf

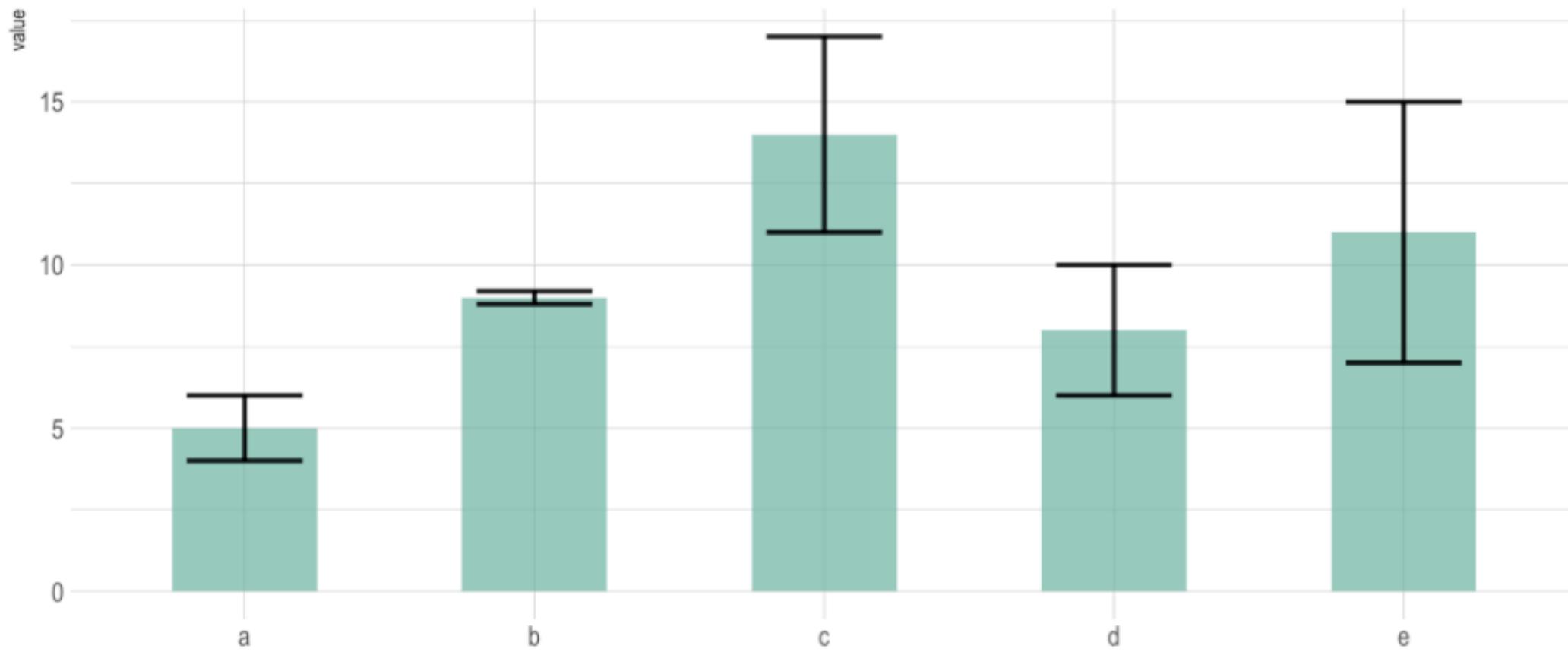
Abonné



17% de Français satisfaits de l'action de Macron... Dites donc, [@IpsosFrance](#), il n'y aurait pas un petit soucis sur ce graphique de votre sondage ? La correction, c'est cadeau [@BriceTeinturier](#) (source : sciencespo.fr/cevipof/sites/ ...)

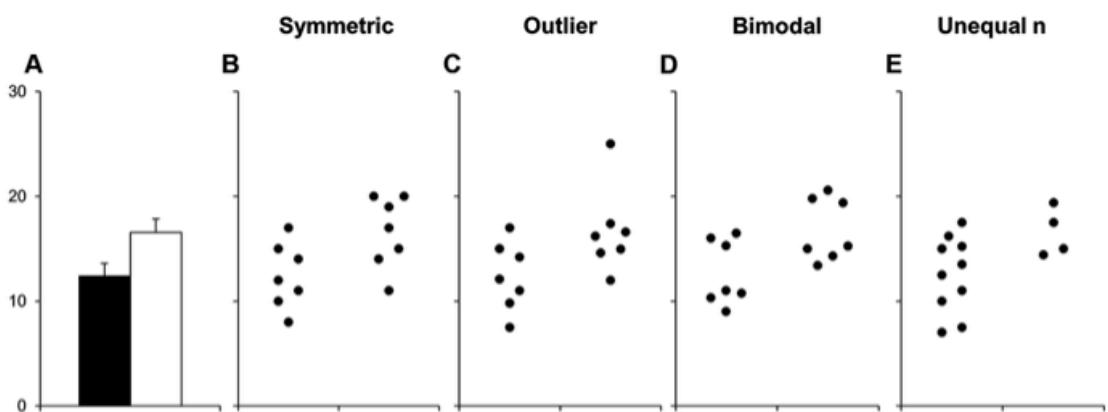


A barplot with error bar

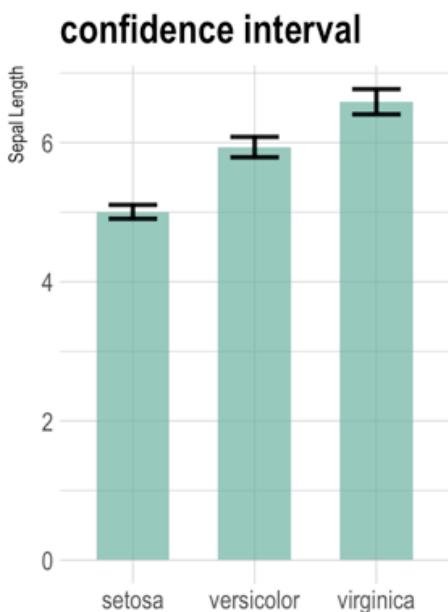
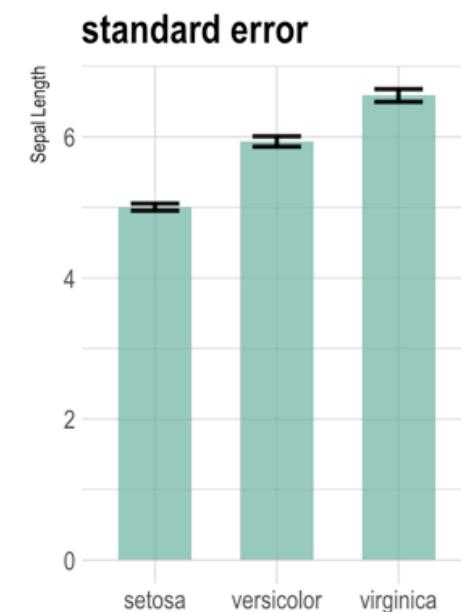
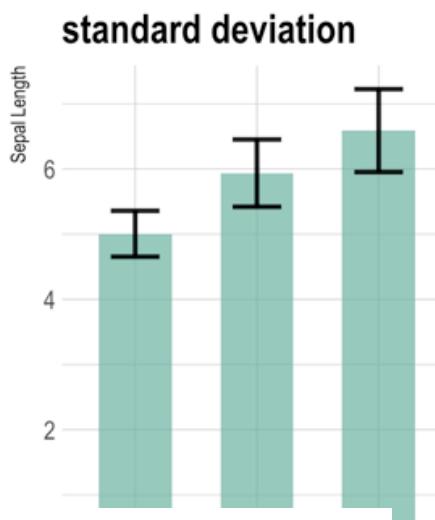


The issue with error bars

[Read more](#)

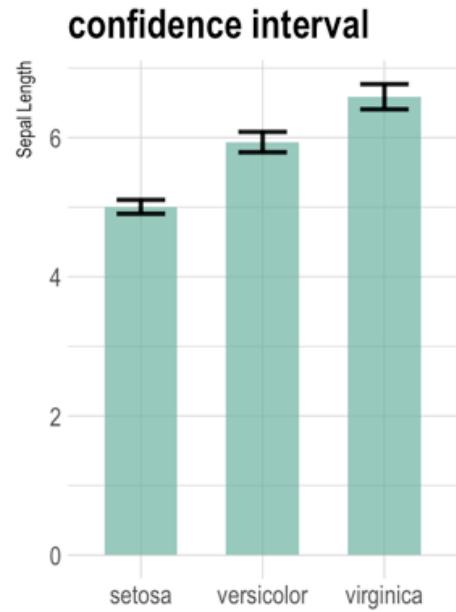
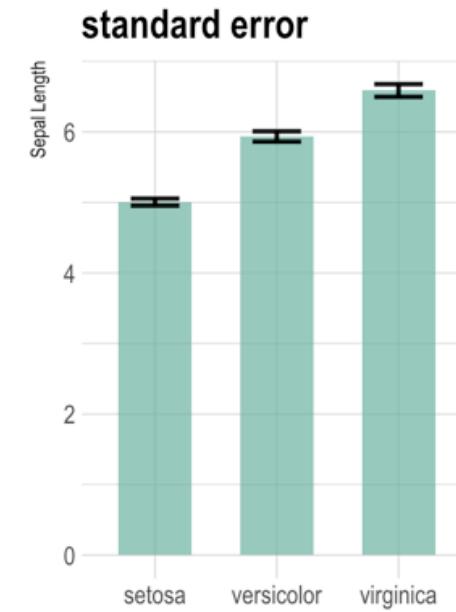
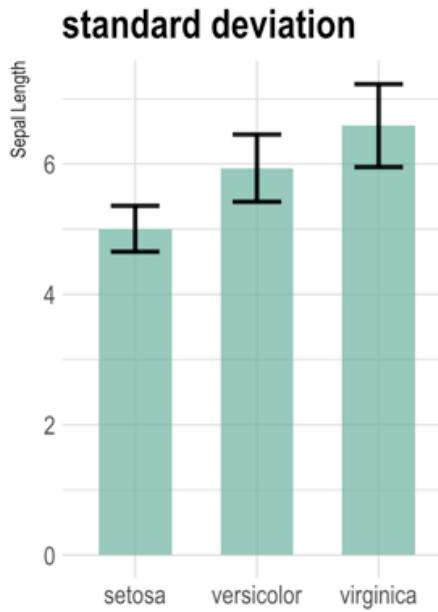


| Test | | p value | | |
|----------------------|--|---------|-------|-------|
| T-test: Equal var. | | 0.035 | 0.050 | 0.026 |
| T-test: Unequal var. | | 0.035 | 0.050 | 0.026 |
| Wilcoxon | | 0.054 | 0.073 | 0.128 |



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**

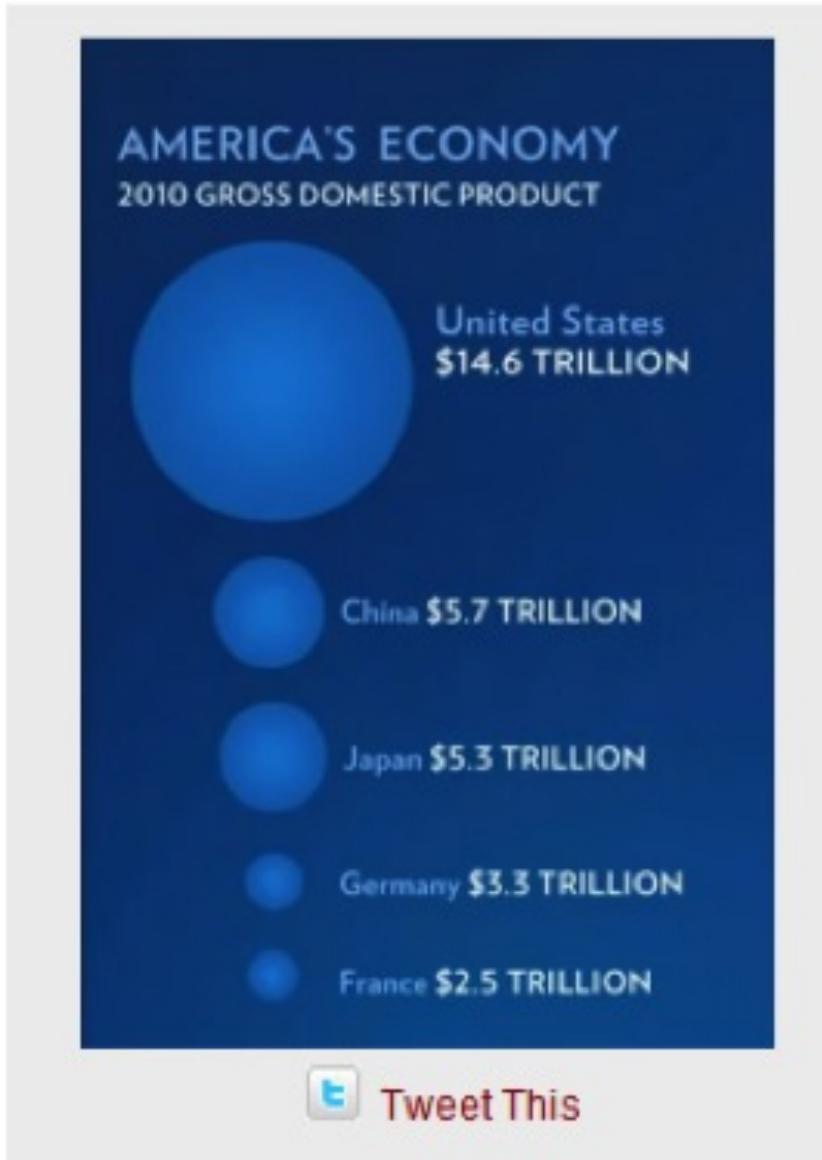


Germany **\$3.3 TRILLION**

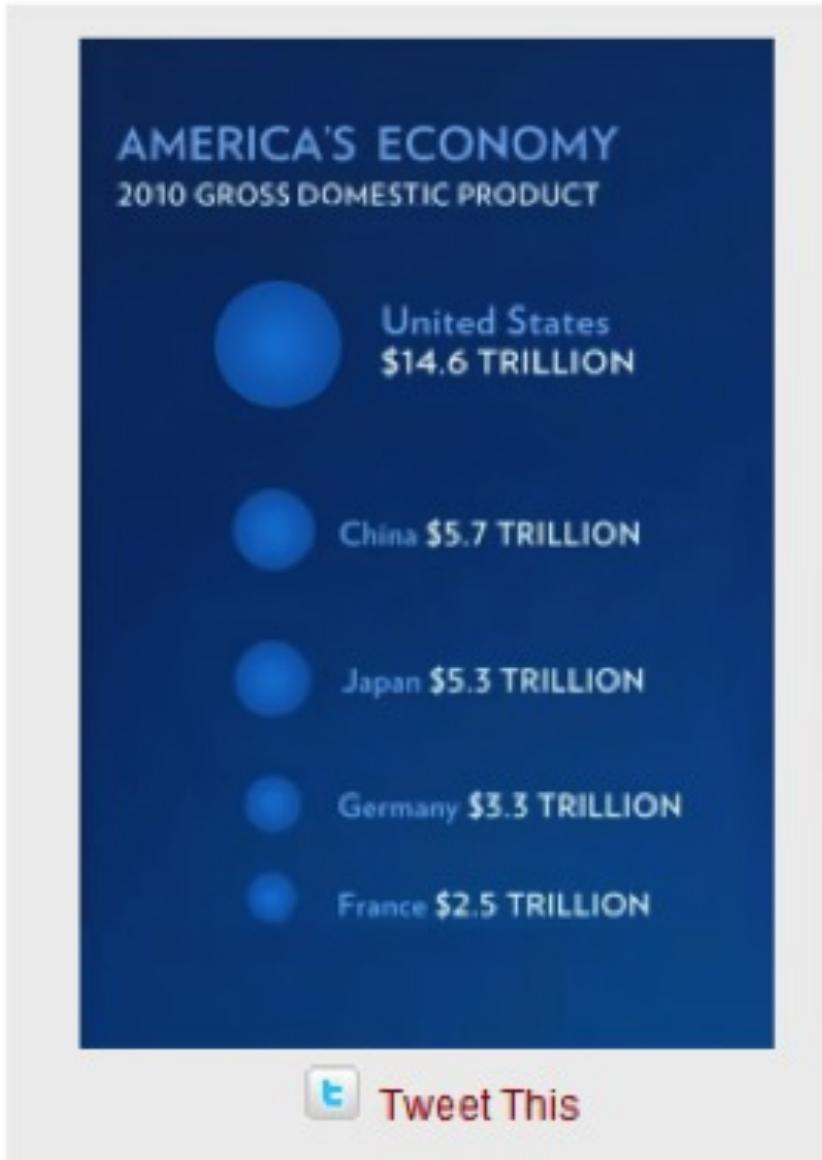


France **\$2.5 TRILLION**

Size = radius



Size = area

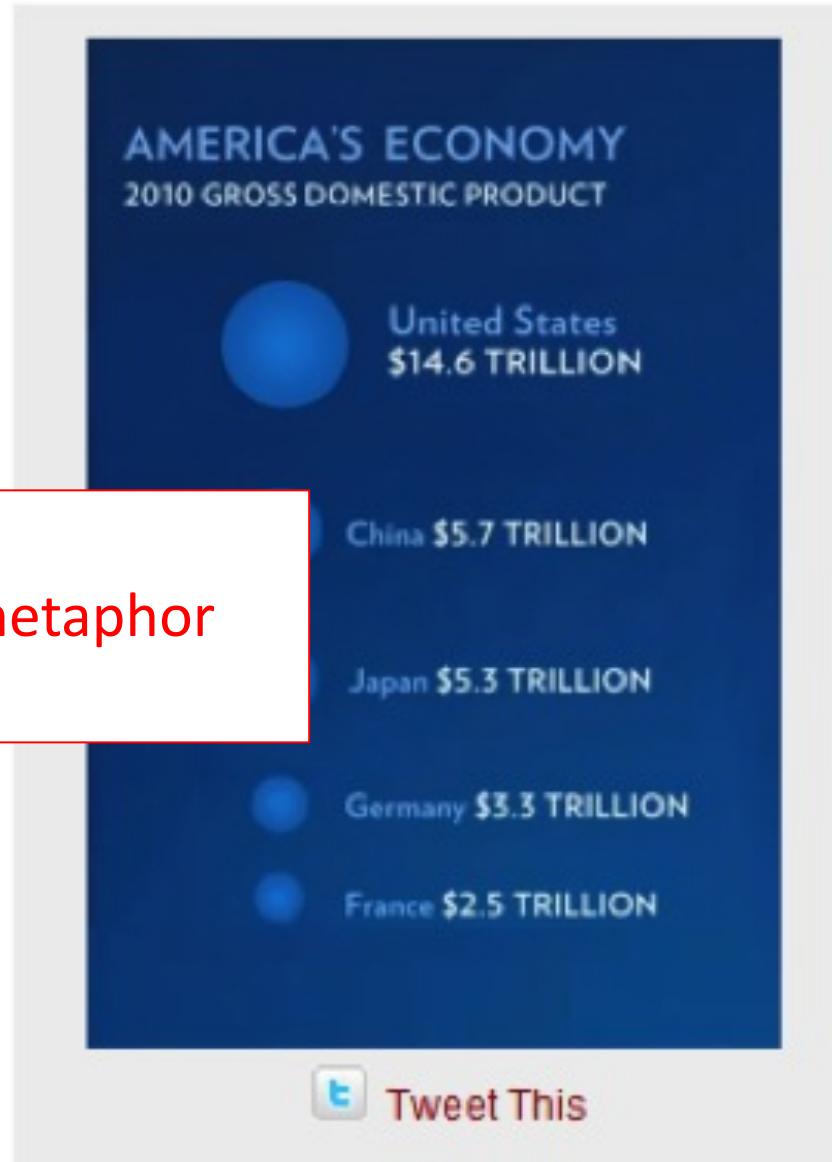


Source: [Fast Fedora blog](#)

Size = radius



Size = area



Source: [Fast Fedora blog](#)

SOUTHWEST BORDER APPREHENSIONS

OCTOBER - APRIL



Source: U.S. Border Patrol

Happening
NOW

\$43 BIL IN FEB, DOWN 3.4% FROM JAN'S REVISED

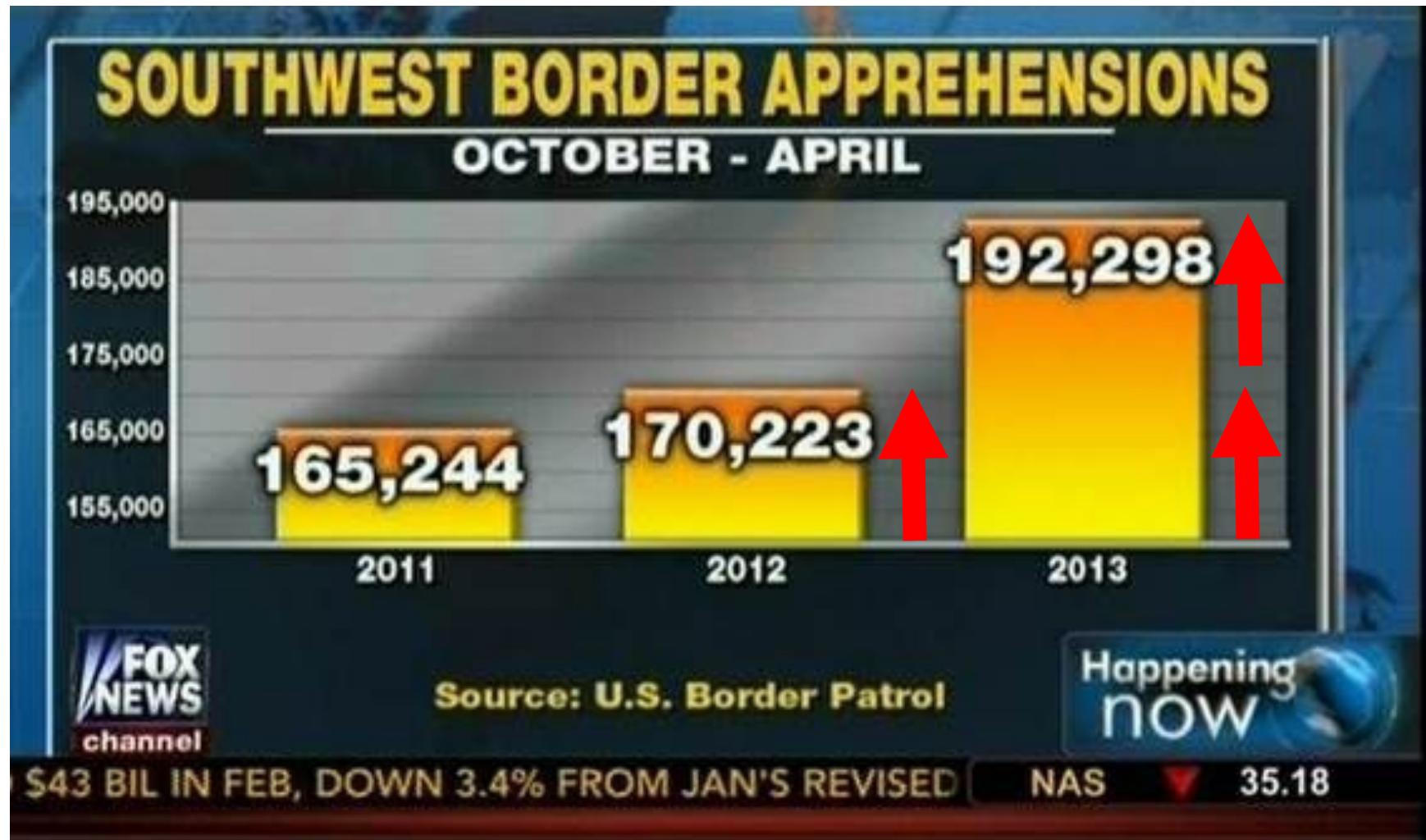
NAS

35.18

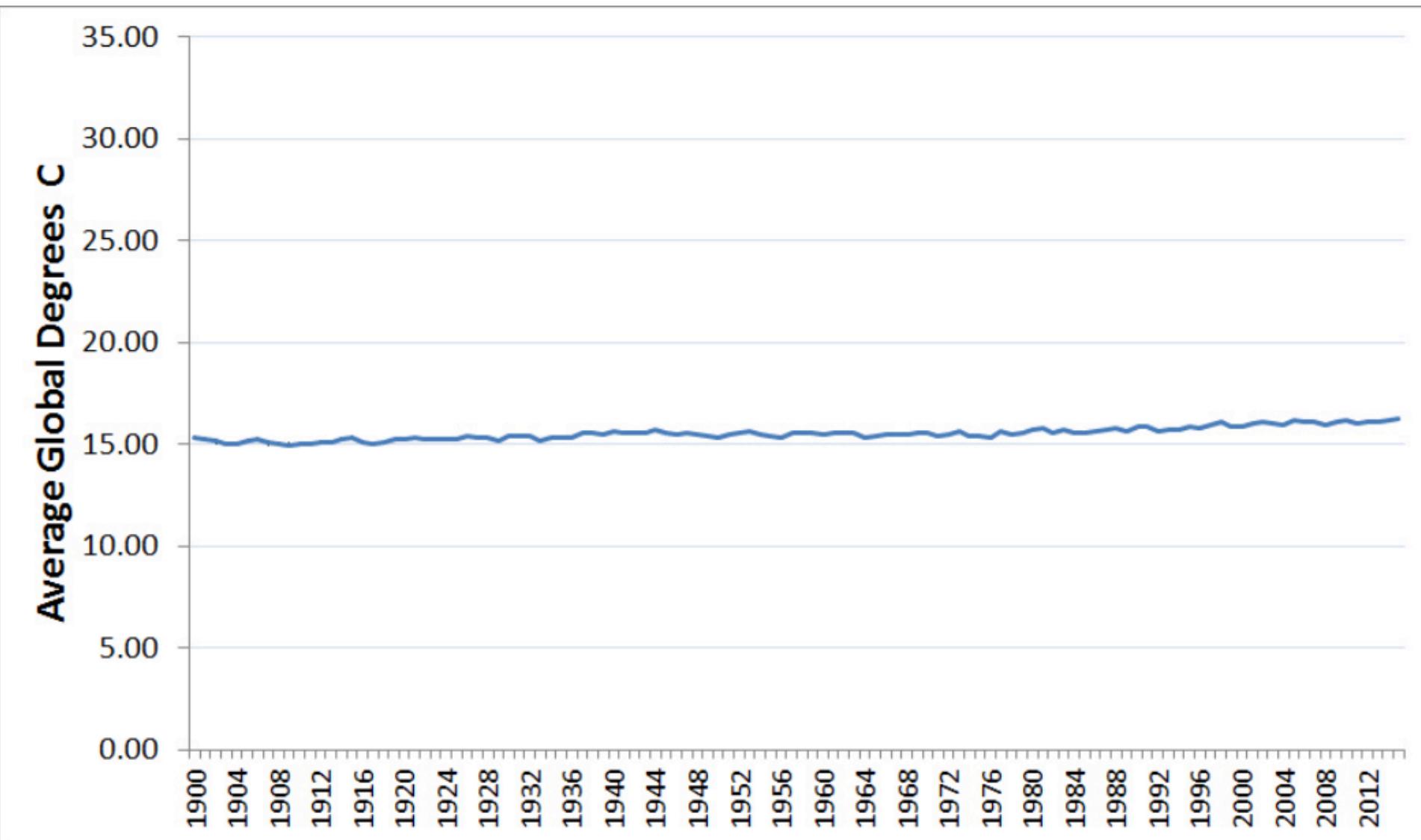
Source: Fox News, via [Media Matters for America](#)

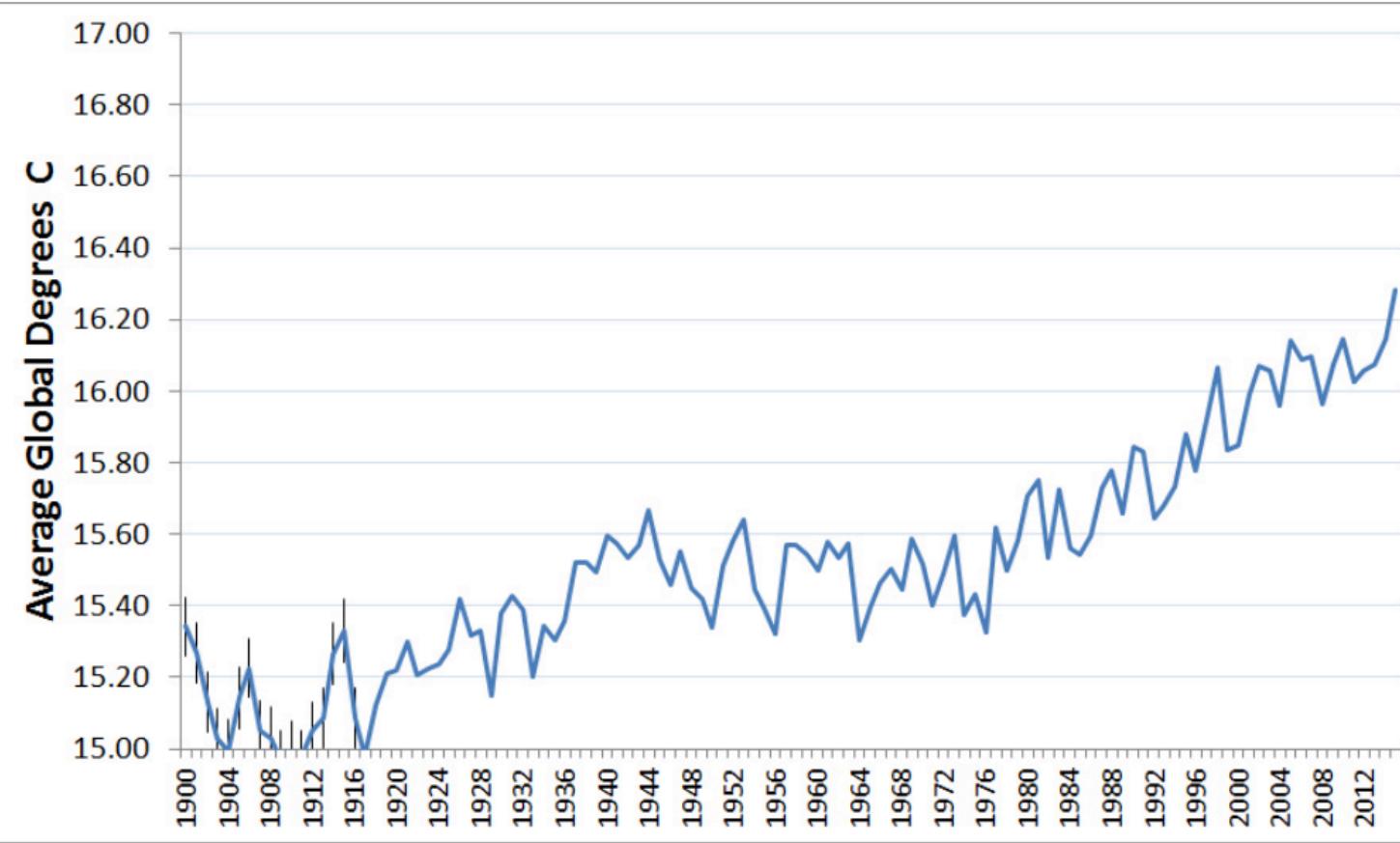
Obama
adminitration

=
2x more
apprehensions?



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

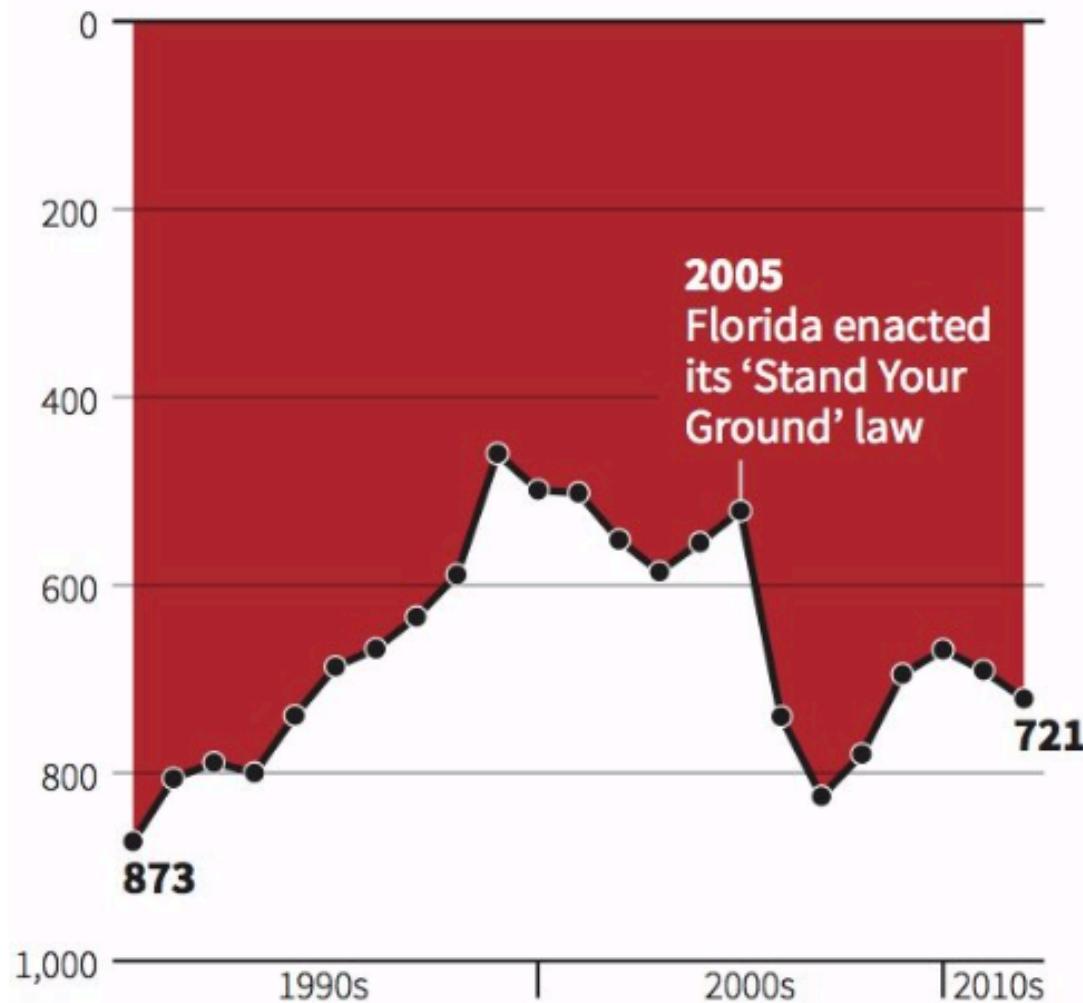




*“In general, in a time-series, use a baseline
that shows the data not the zero point” -*
Edward Tufte

Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

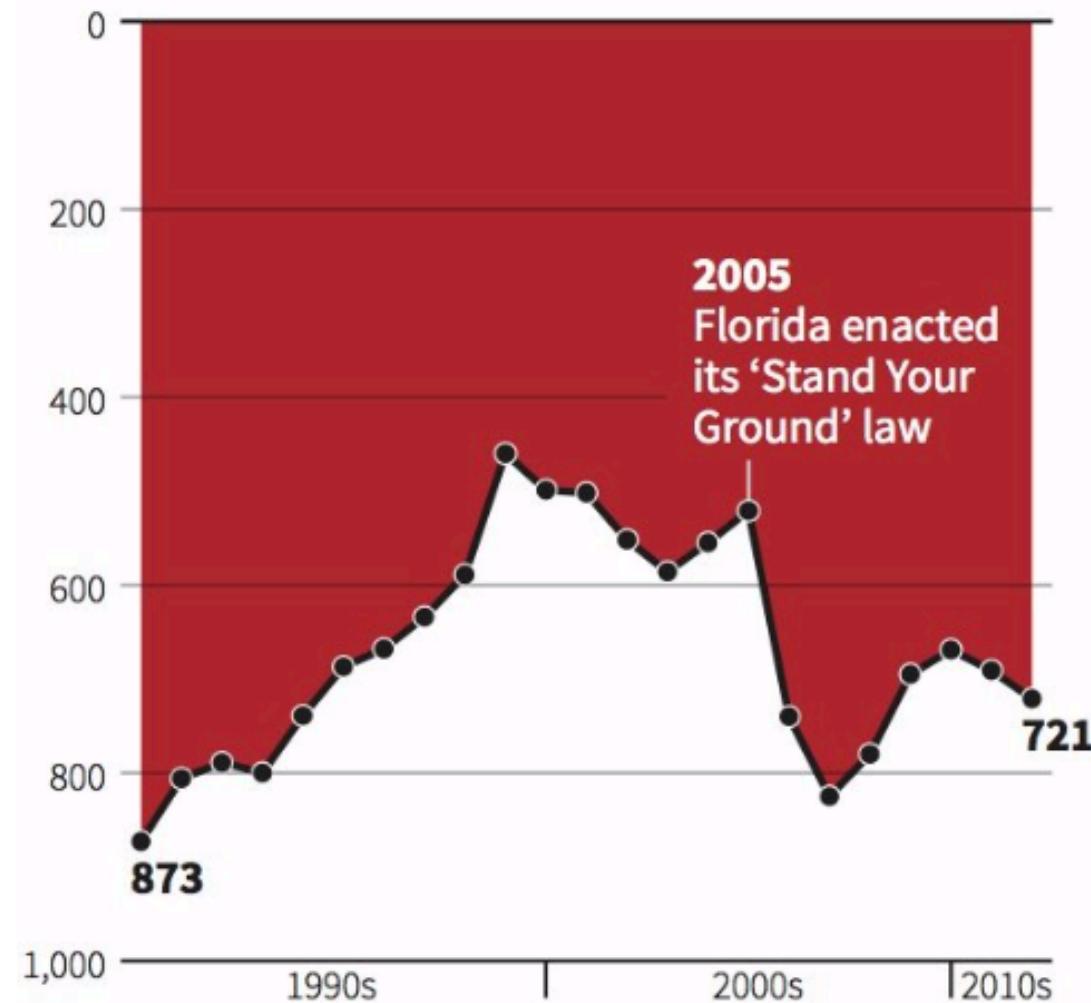
[Source: KD Nuggets](#)

Do not be
counter intuitive

Reverse Y Axis →

Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

Source: KD Nuggets

Data to Viz

A collection of
dataviz caveats

[Data-to-viz.com/caveats](https://www.dat-to-viz.com/caveats)



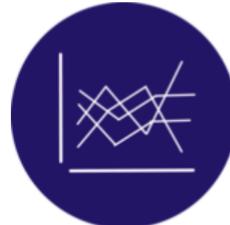
Order your data

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.



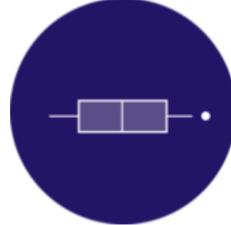
Pie chart

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



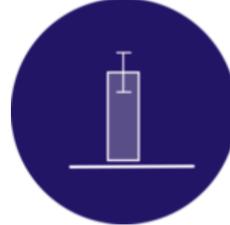
Play with histogram bin size

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.

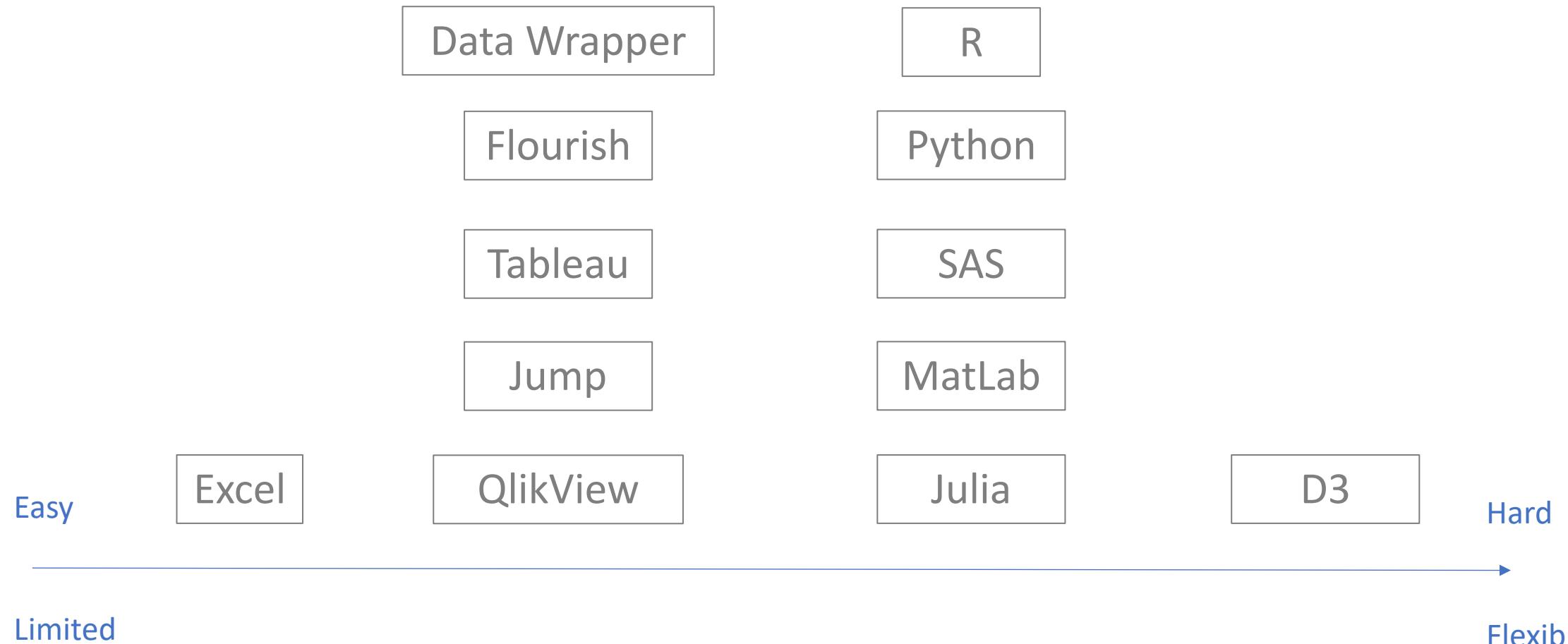


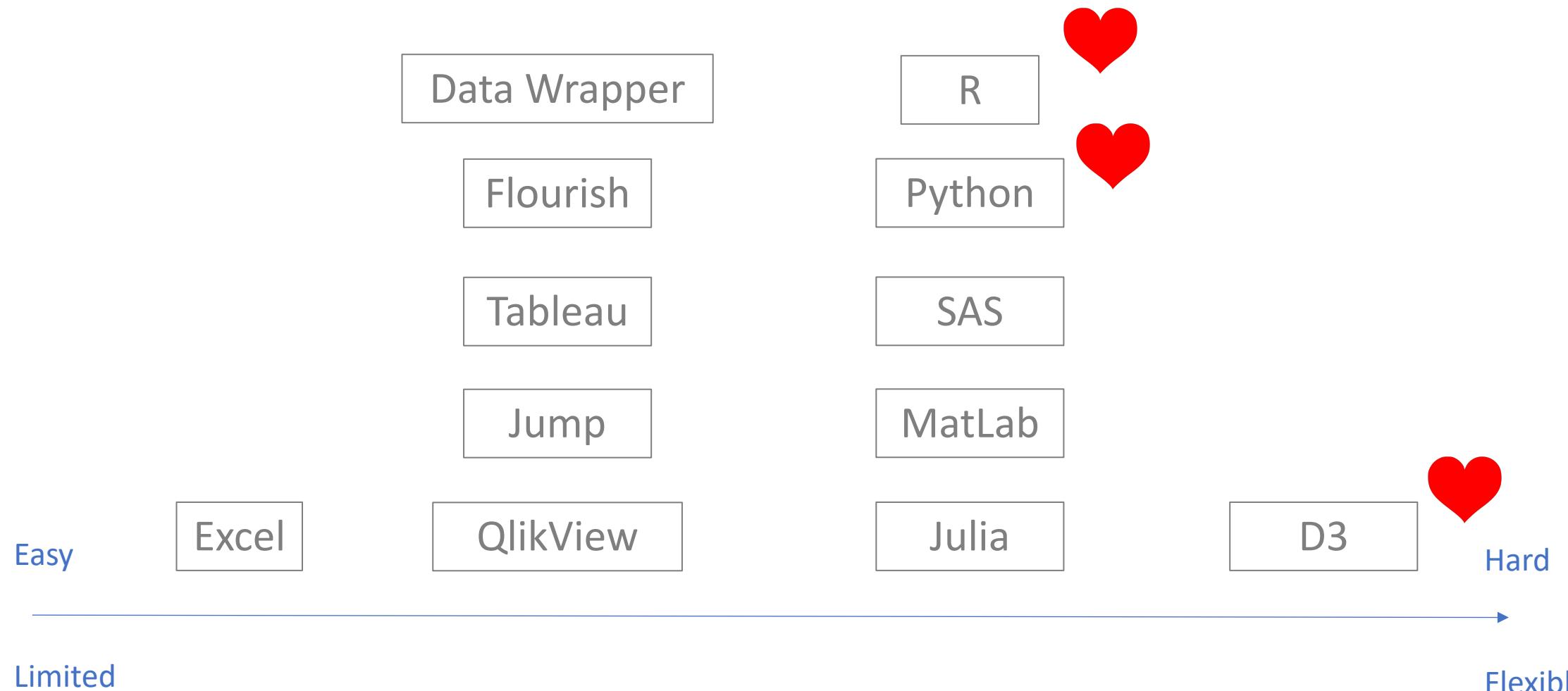
Too many distributions.

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

HOW TO DO IT

The R and Python graph galleries







from Data to *Viz*



[**R-graph-gallery.com**](#)

[**Python-graph-gallery.com**](#)

[**D3-graph-gallery.com**](#)



KANTAR
Information is Beautiful
Awards

Data-to-viz.com



@R_Graph_Gallery



github.com/holtzy/Talk



Yan.holtz.data@gmail.com



www.yan-holtz.com



Modern monitoring & analytics

See inside any stack, any app, at any scale, anywhere.

FREE TRIAL

