

THE YOUNG ZOOLOGIST GROUP PRESENTS

DATA-VIZ

A DATA VISUALIZATION WORKSHOP SERIES

YAN HOLTZ



CHOOSING THE RIGHT CHART FOR YOUR DATA

1.

29th of September - 10:00 AM - 12:00 AM

DATA VISUALIZATION IN R (BEGINNER)

2.

4th of October - 10:00 AM - 12:00 AM

DATA VISUALIZATION IN R (ADVANCED)

3.

8th of October - 10:00 AM - 12:00 AM

CEDRIC SCHERER



JOIN THE DATA-VIZ CHALLENGE



Apply the skills you learned to plot our dataset
and get a chance to win some amazing prizes!

REGISTER AT: RBZS.BE/WORKSHOPS2021

Supported by:



Hosted by:

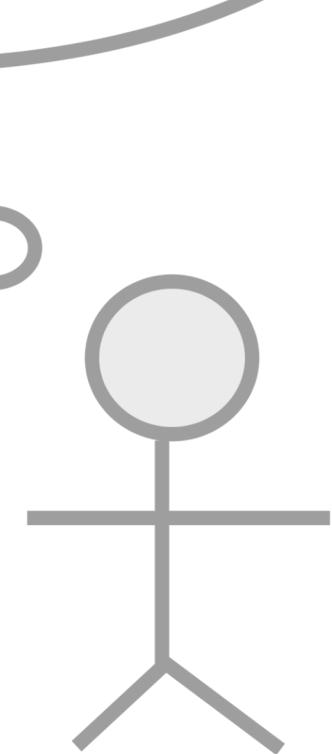


from Data to Viz

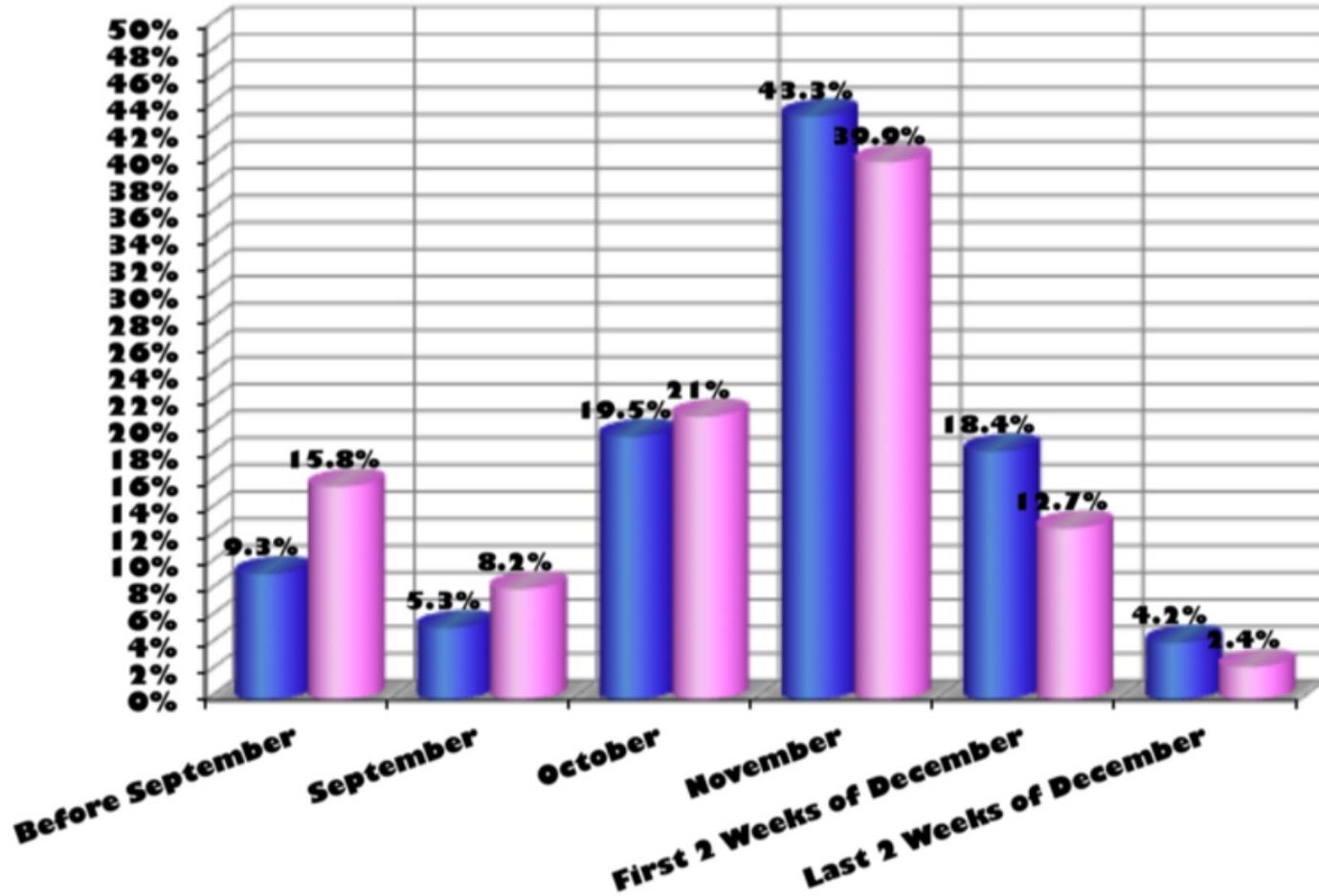
Choosing the right chart for your data

What should I do with my
data ??

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

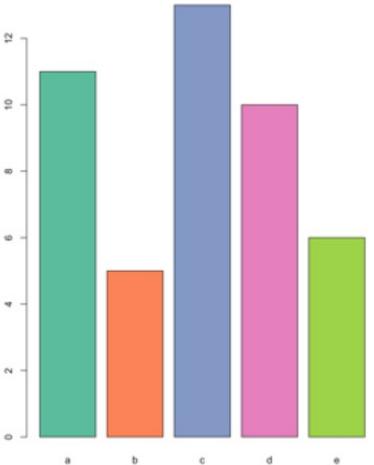


■ Men ■ Women



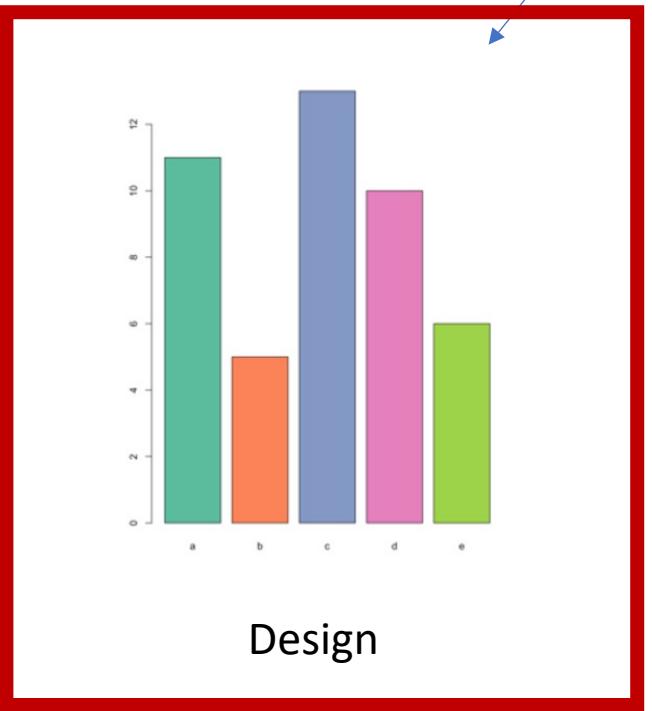
...

What about
that?



Design

Code



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



Scatterplot



Scatterplot



2d density chart

Who sells more weapons ?

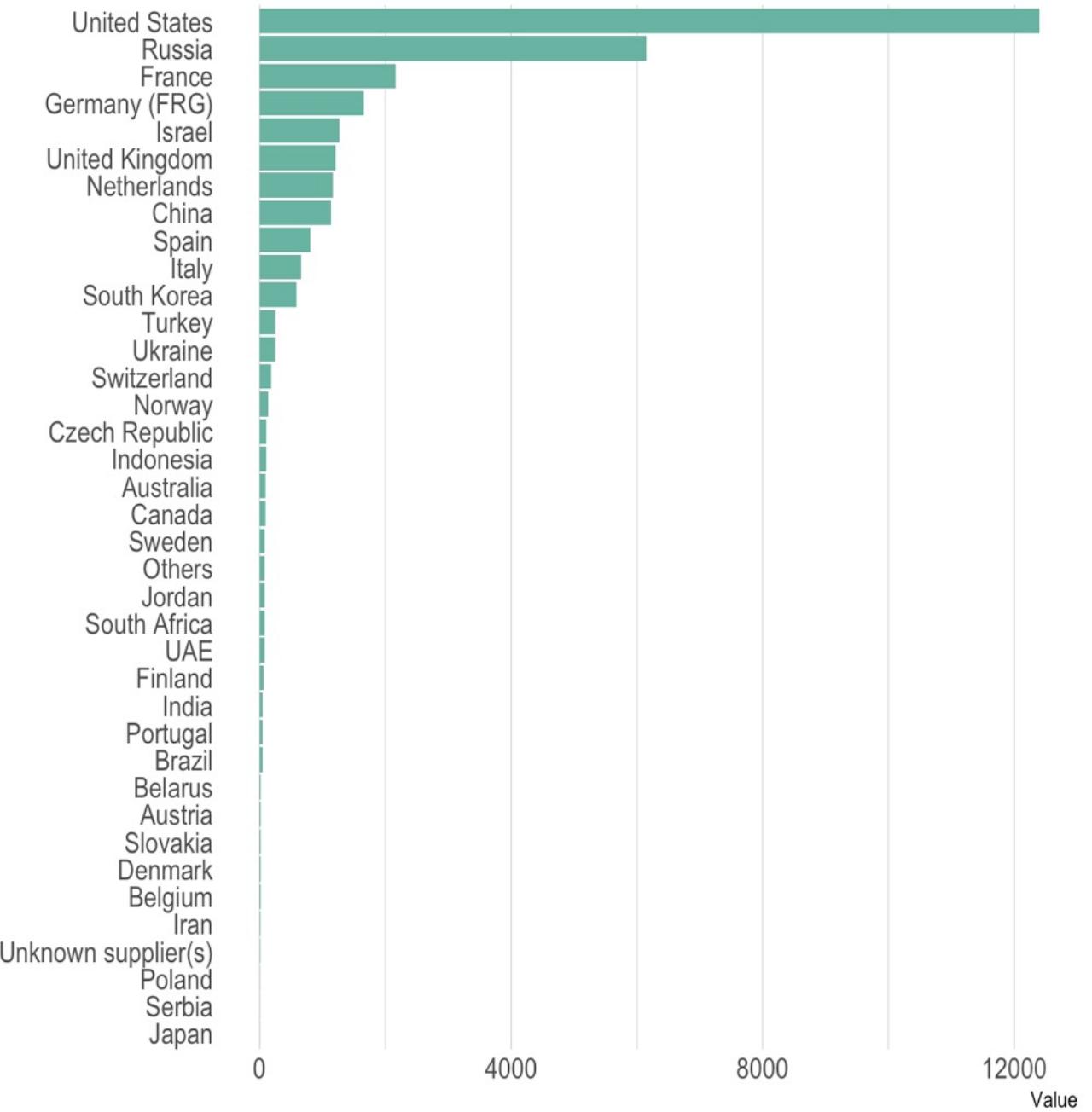
Country	Value
United States	12394
Russia	6148
Germany (FRG)	1653
France	2162
United Kingdom	1214
China	1131



?

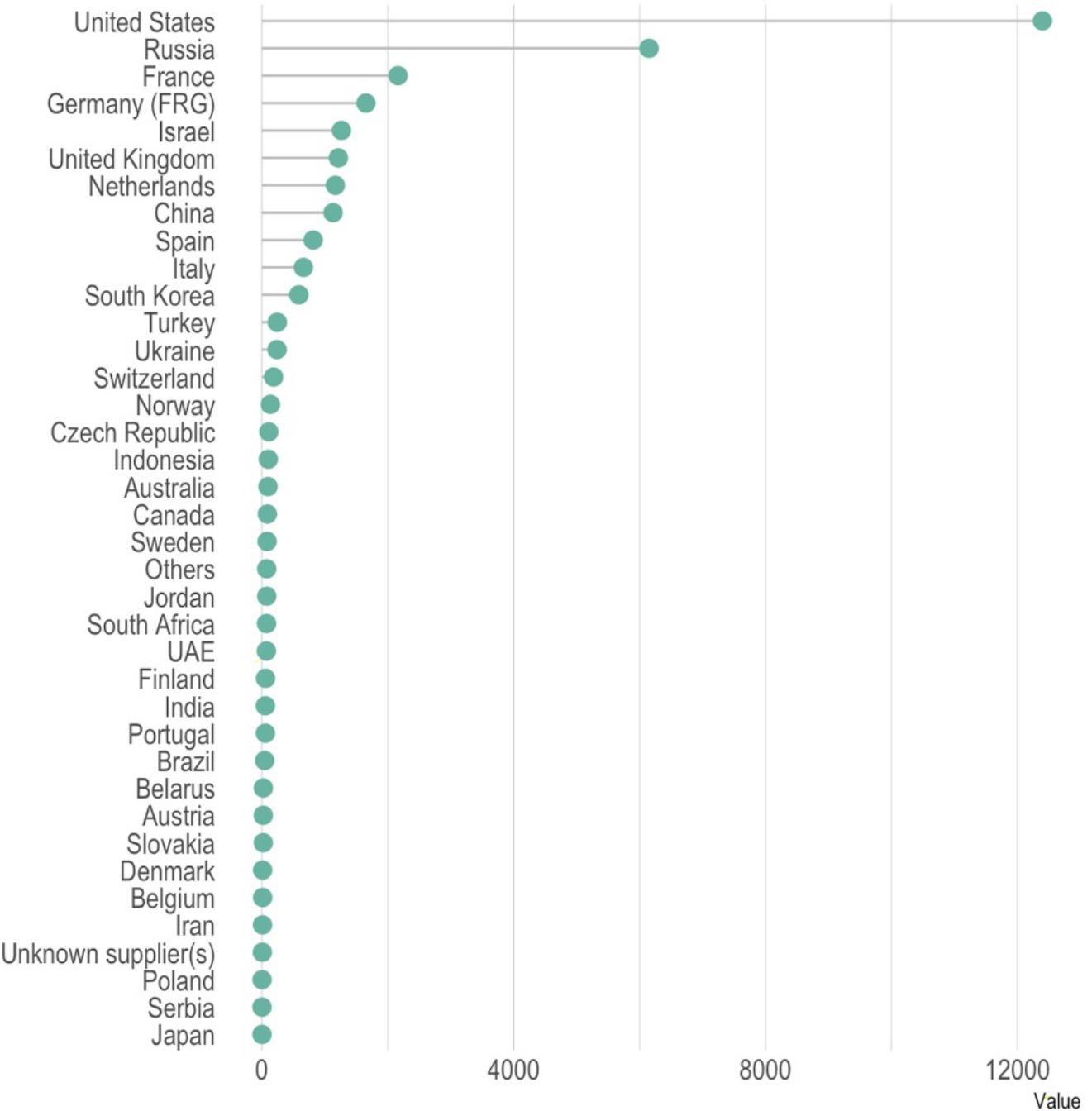
Who sells more weapons ?

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



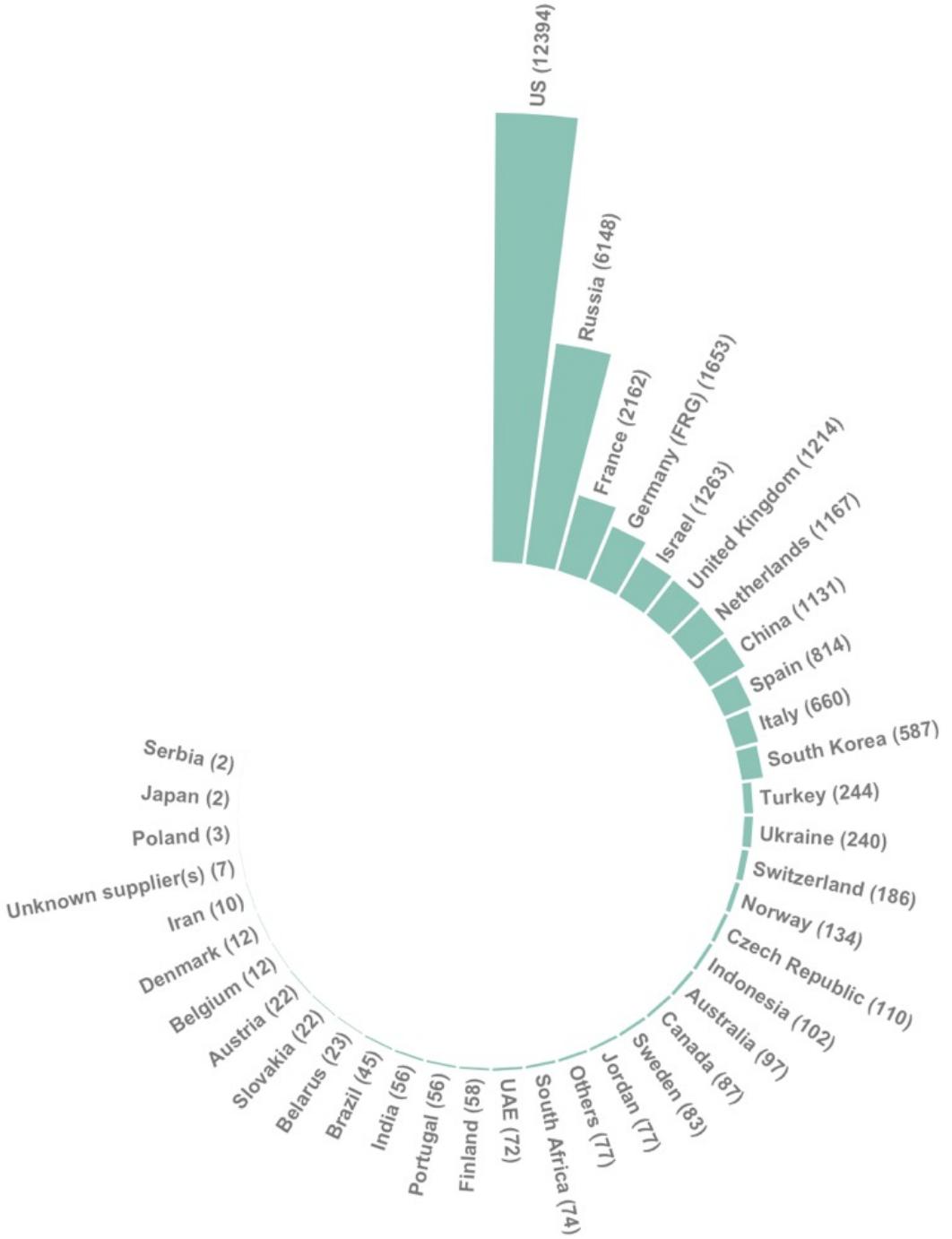
Who sells more weapons ?

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



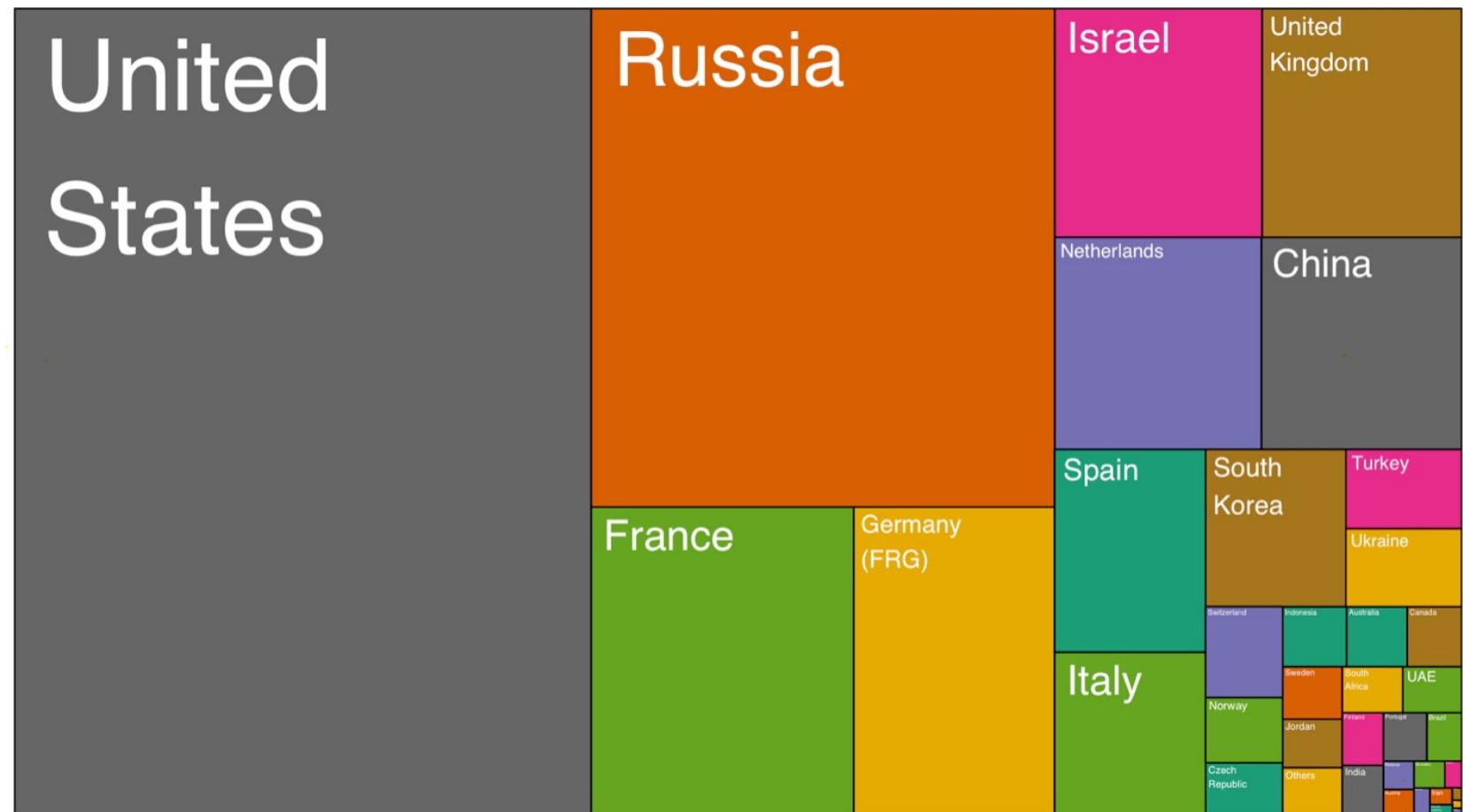
Who sells more weapons ?

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



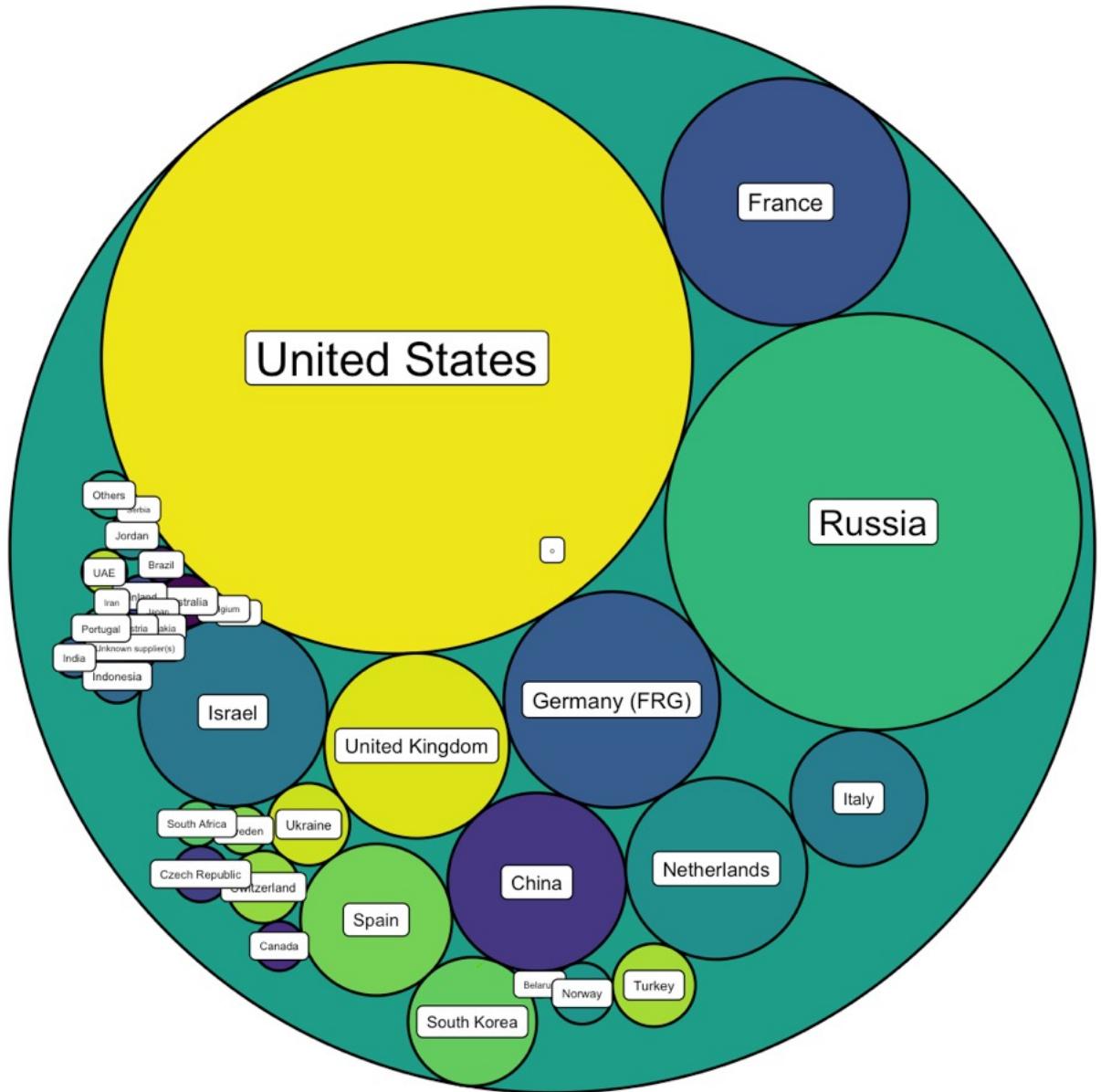
Who sells more weapons ?

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



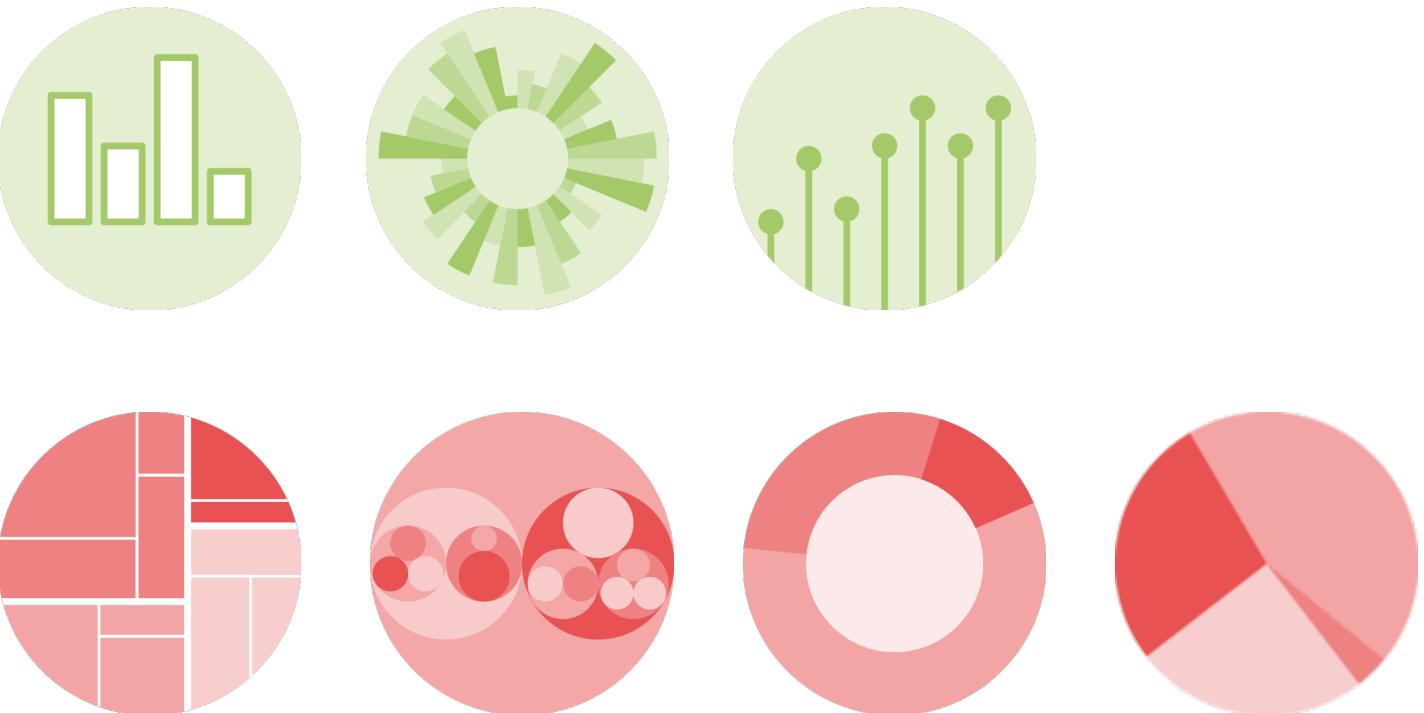
Who sells more weapons ?

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



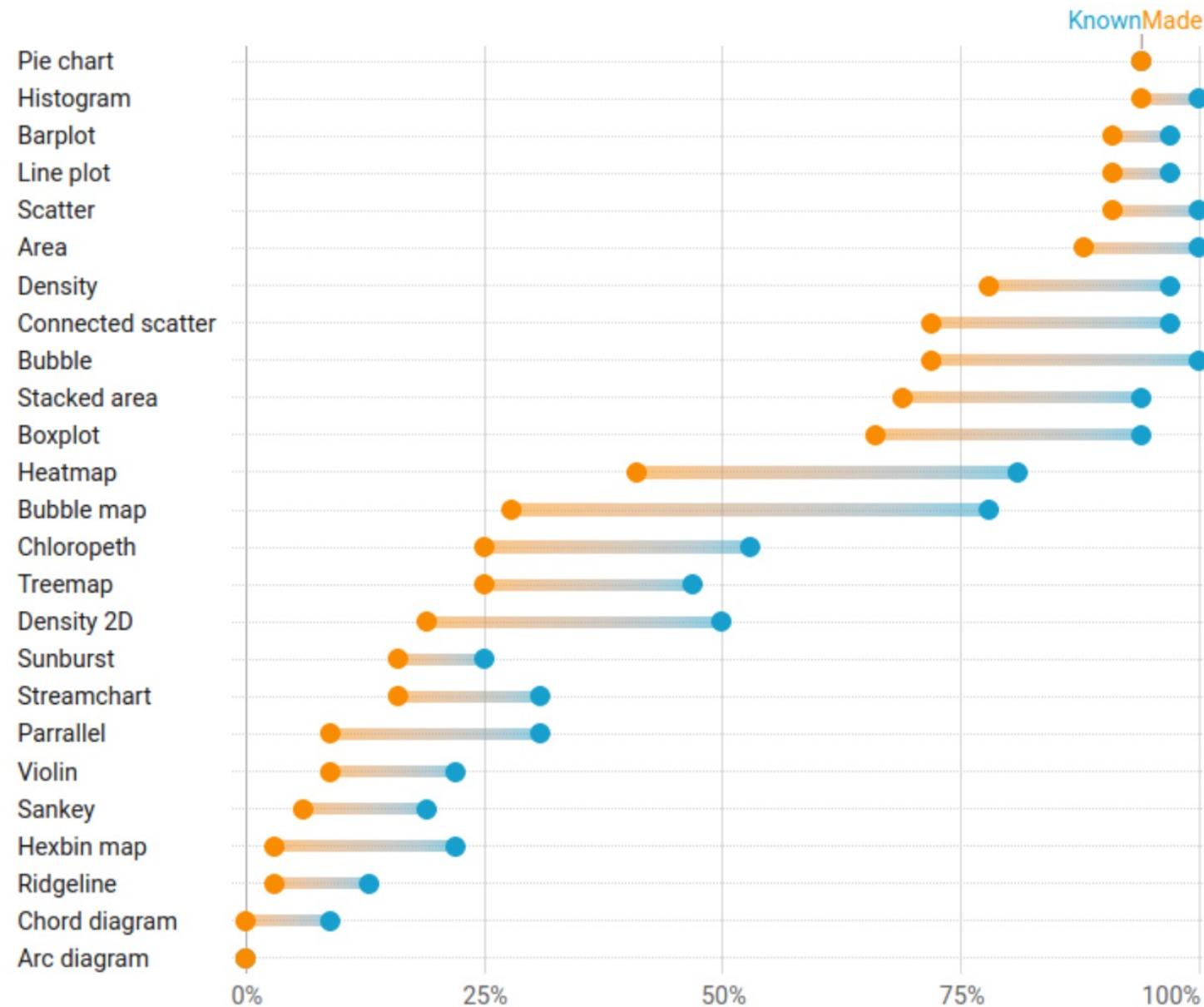
Who sells more weapons ?

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



Charts you made and charts you know 18/12

Dataviz training session 18/12

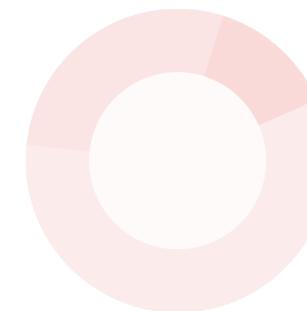
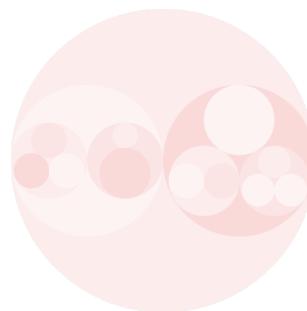
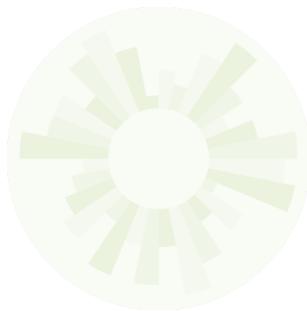


Who sells more weapons ?

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

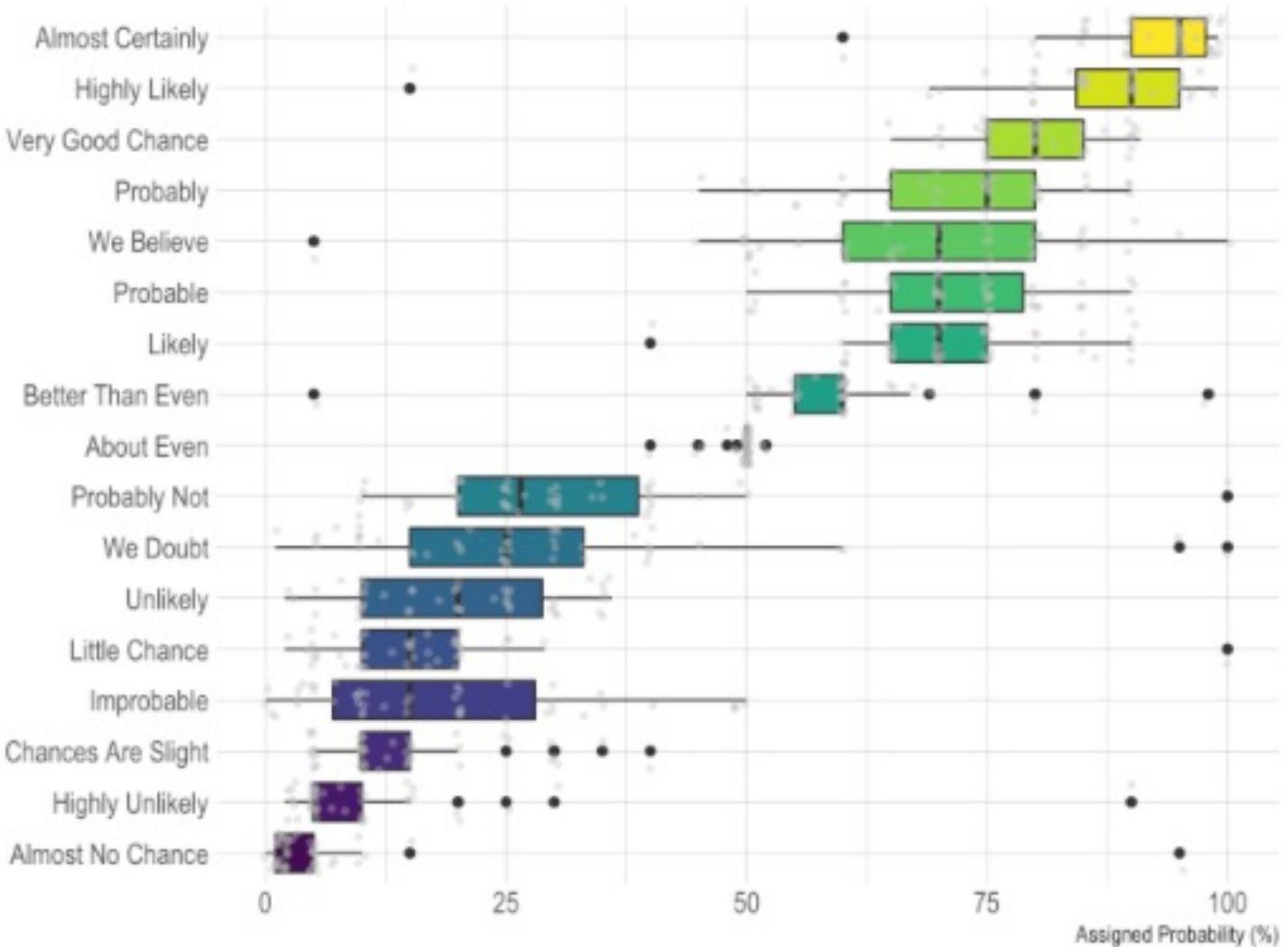
Several
observations
per group

1 categorical
variable

1 numeric
variable

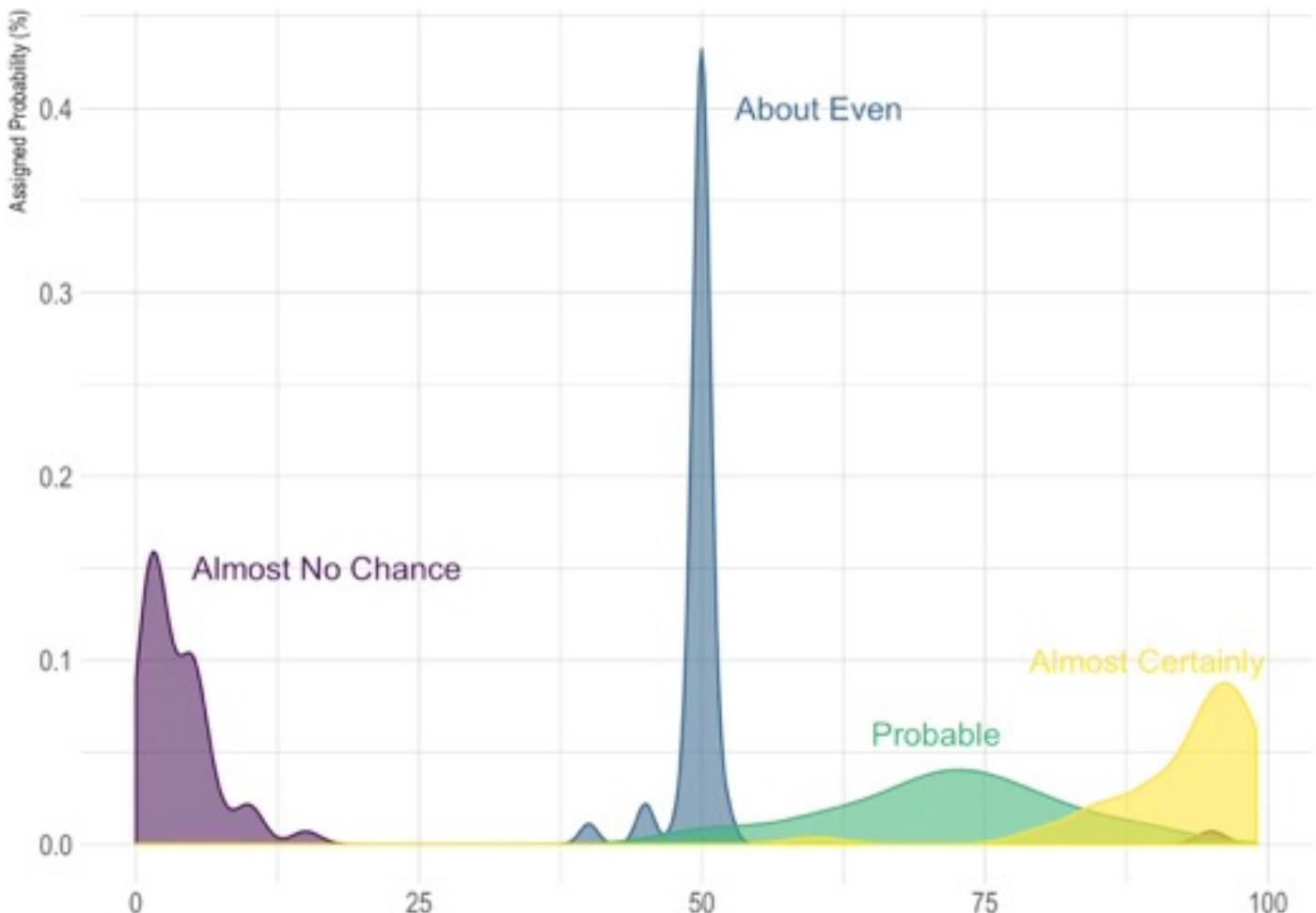
Perception of probability

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



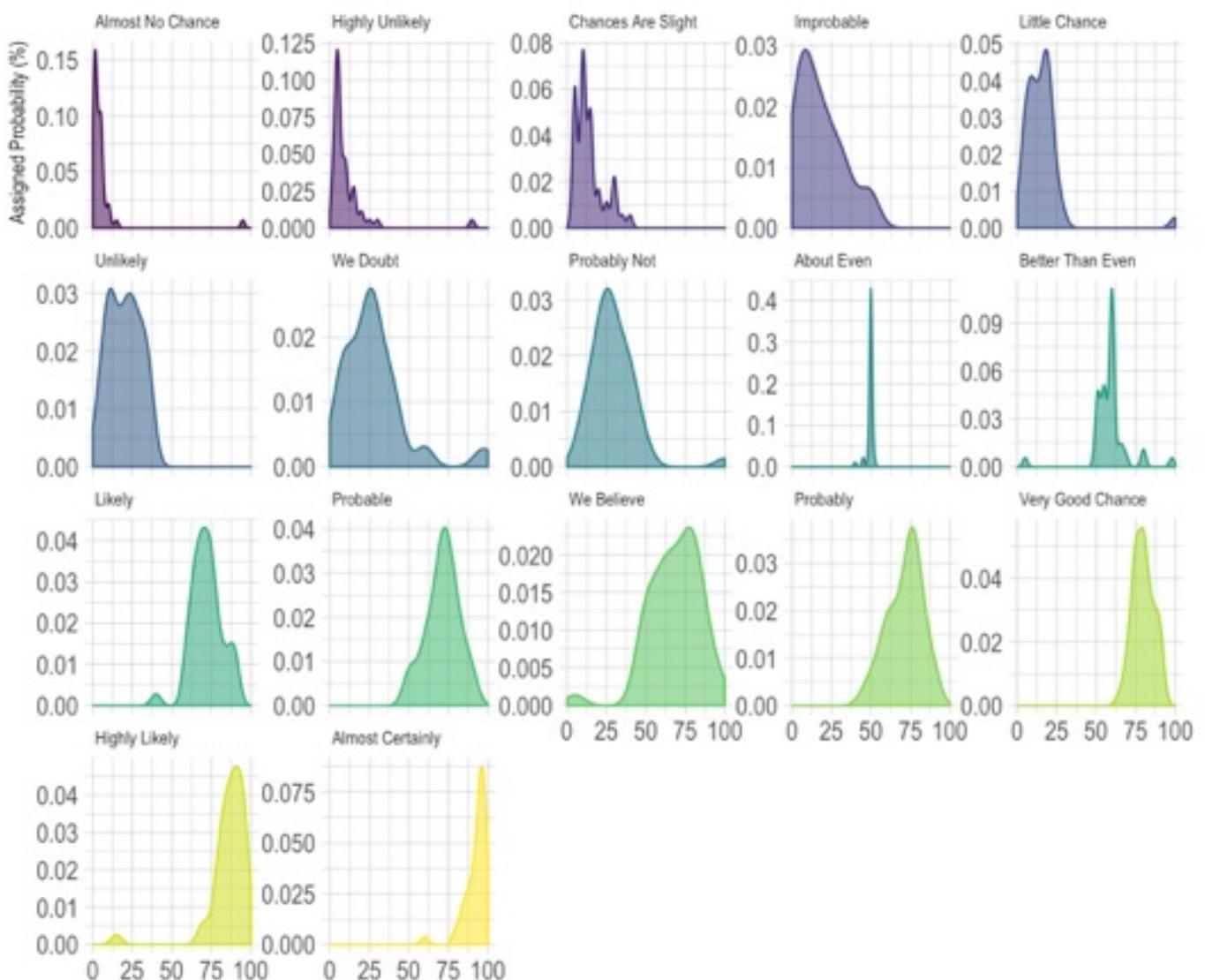
Perception of probability

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



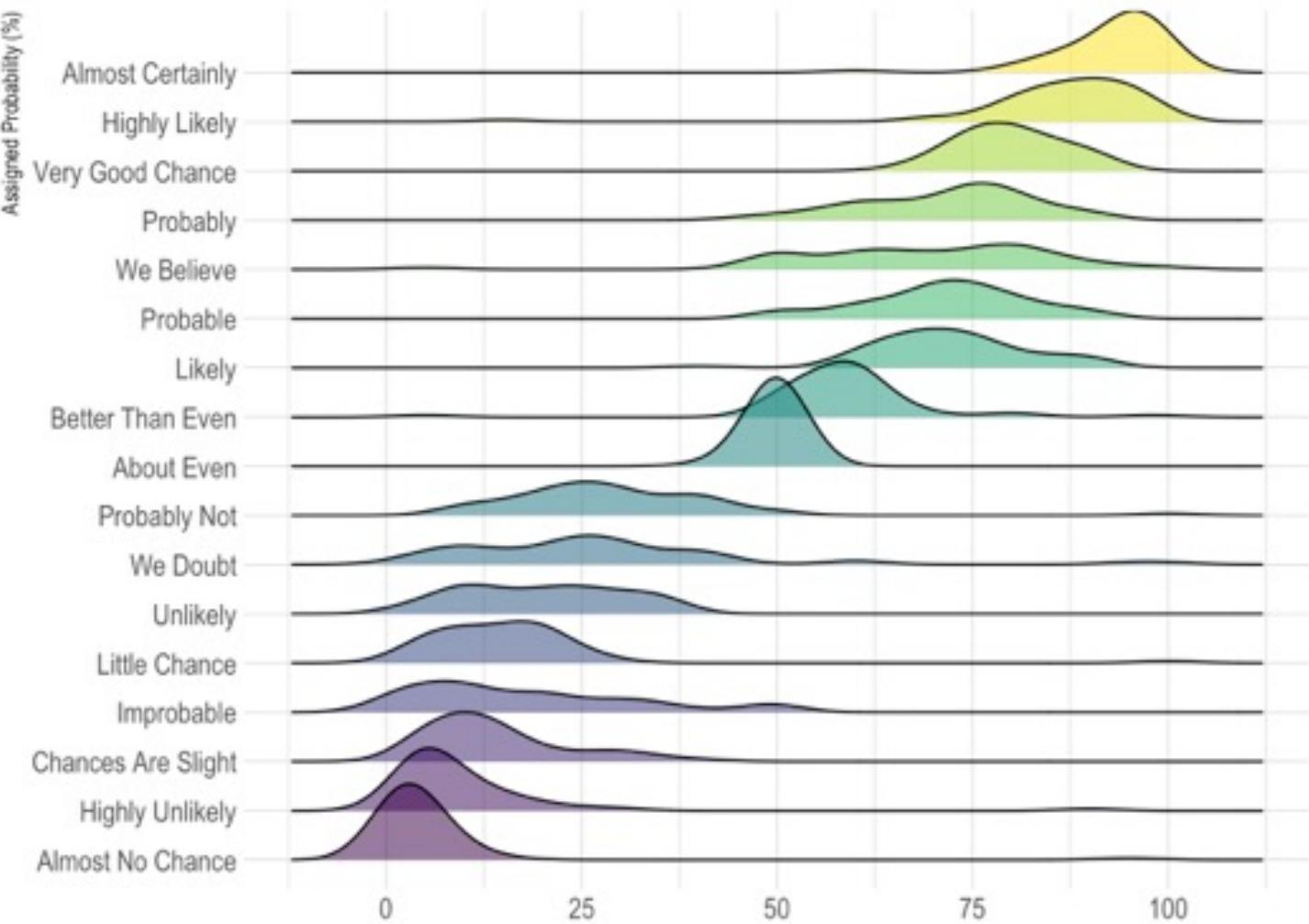
Perception of probability

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



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



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

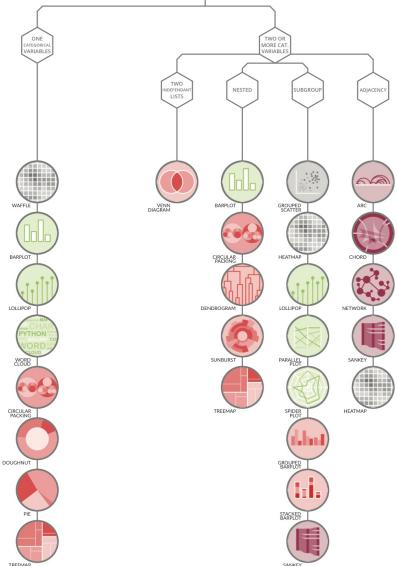
'From Data to Viz' is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps :

- 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.

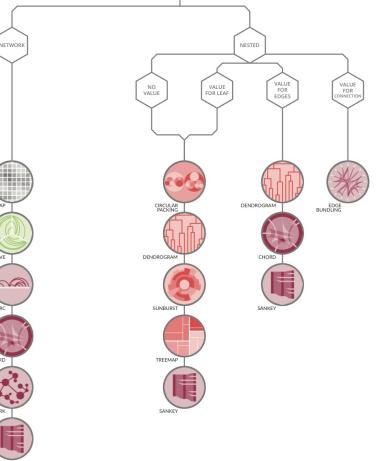
Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more, visit:

data-to-viz.com

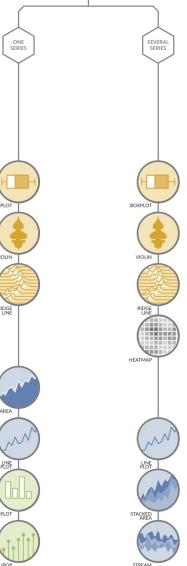
CATEGORIC



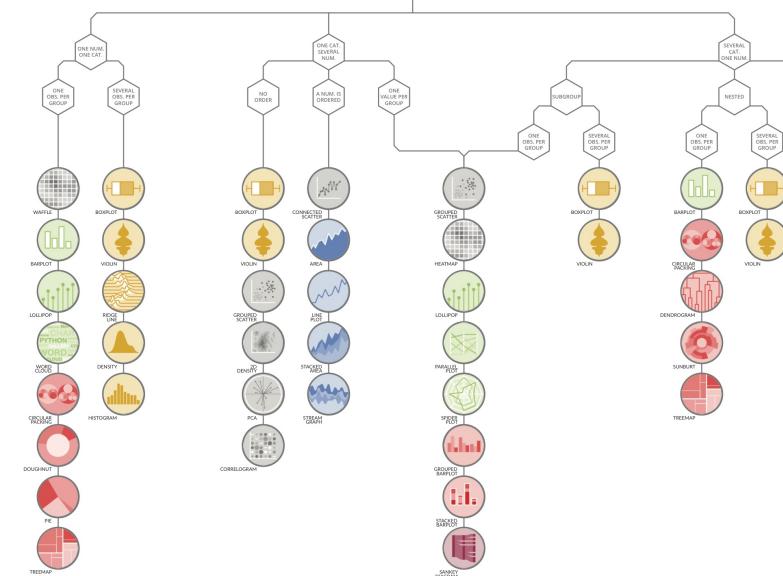
RELATIONAL



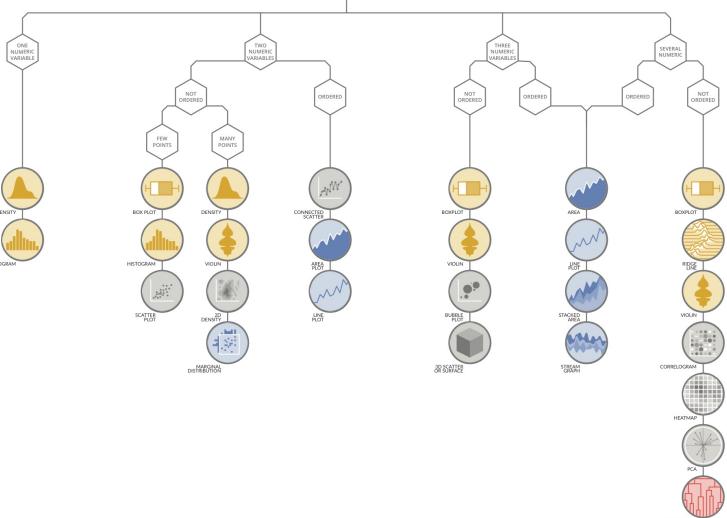
TIME SERIES



CATEGORIC AND NUMERIC



NUMERIC



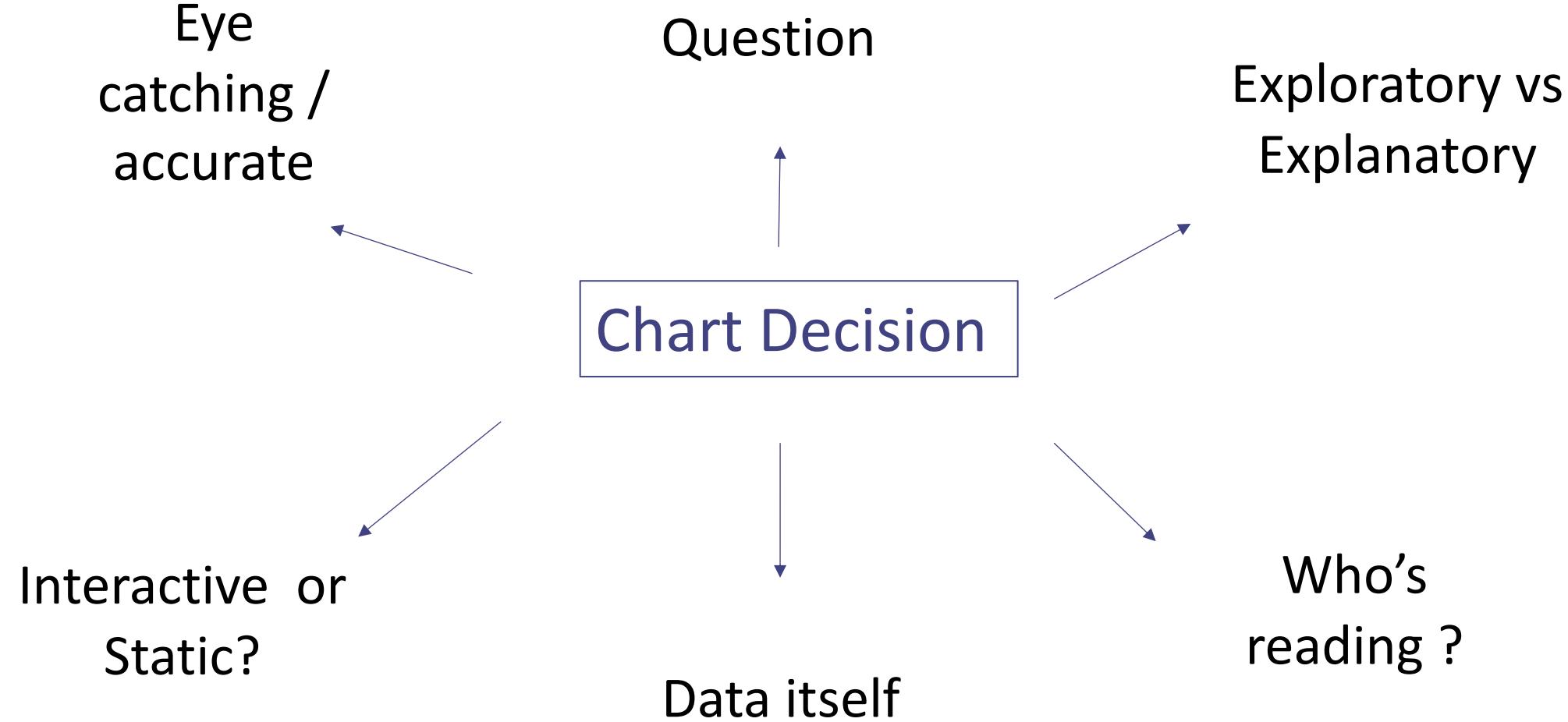
Data-to-viz.com

There are no limits in dataviz!

<https://xeno.graphics/>

WHAT YOU SHOULD DO

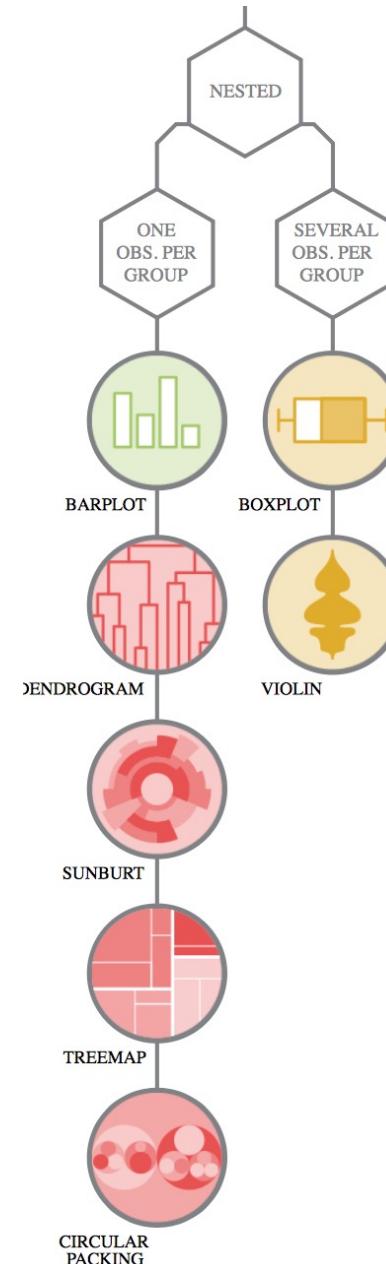
About 20 examples of storytelling with data



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

WHAT DO YOU WANT TO SHOW ?

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

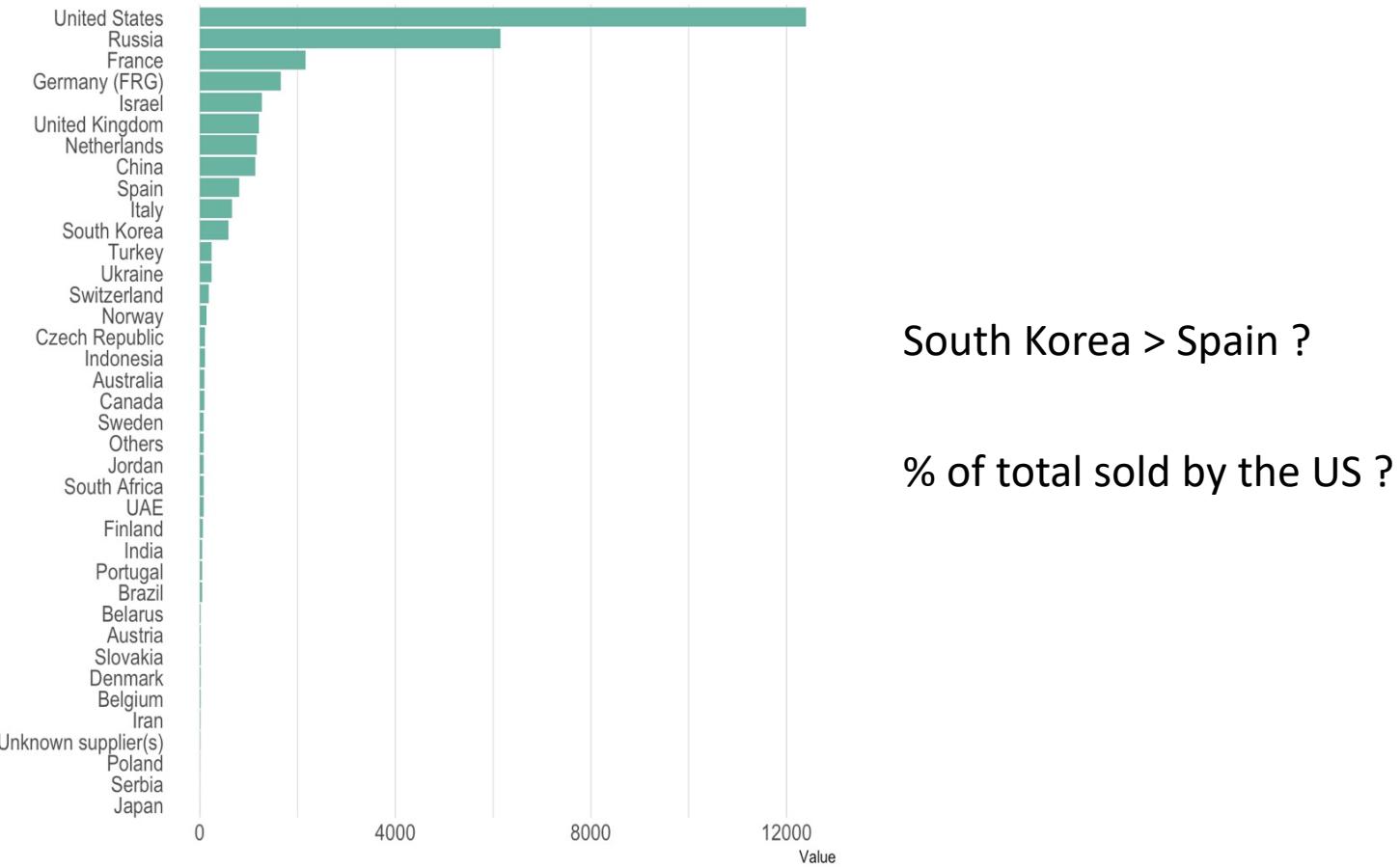


Who sells more weapons ?

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

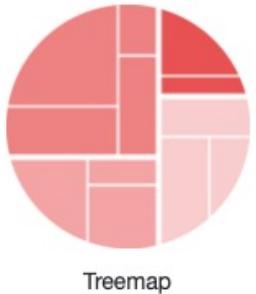


Barplot



Who sells more weapons ?

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



United States

Russia

Israel

United Kingdom

Netherlands

China

Spain

South Korea

Turkey

Ukraine

France

Germany (FRG)

Italy

Norway

Switzerland

Indonesia

Australia

Canada

UAE

Sweden

South Africa

Jordan

Iceland

Portugal

Brazil

Czech Republic

Others

India

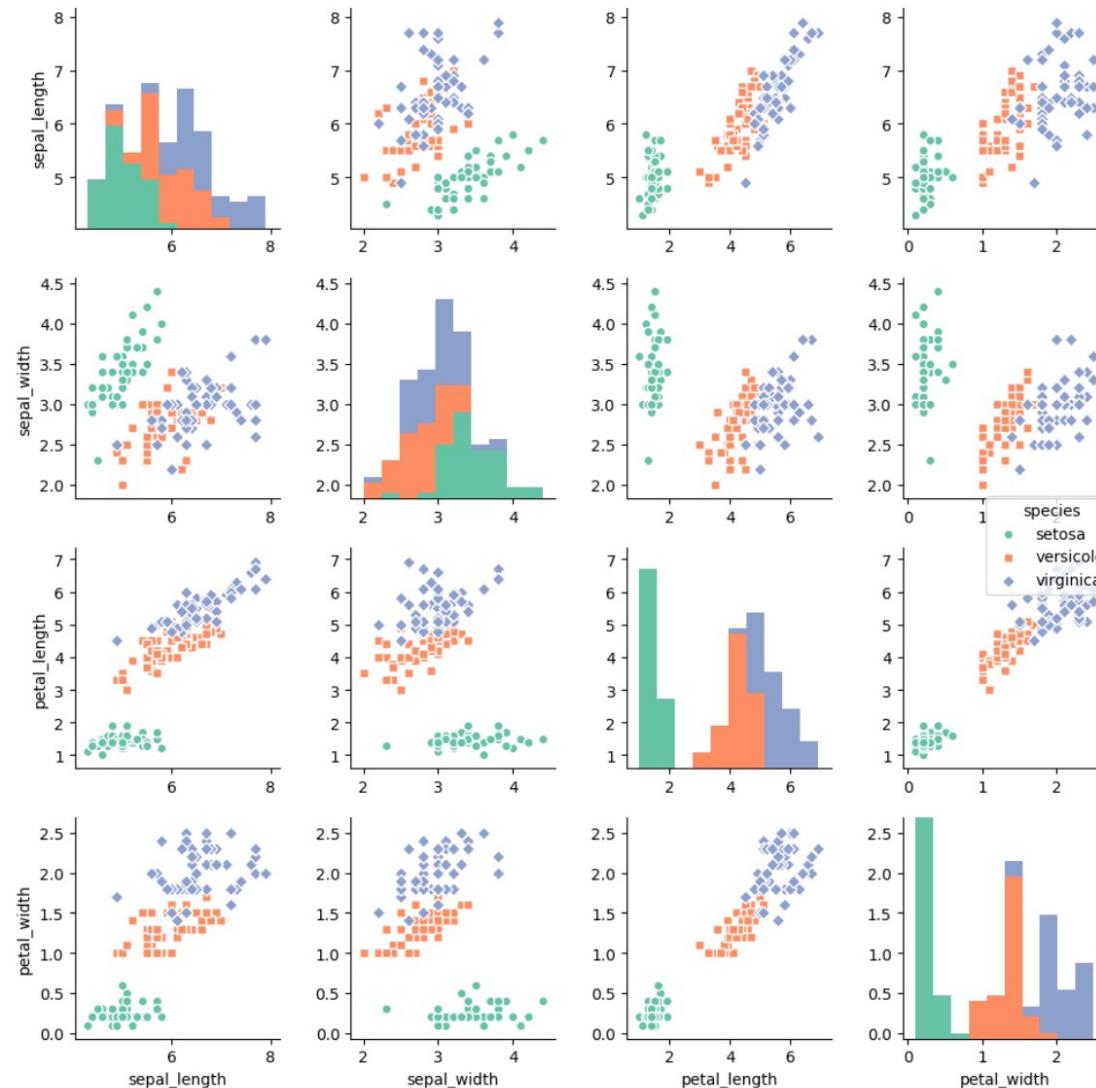
Malta

Other

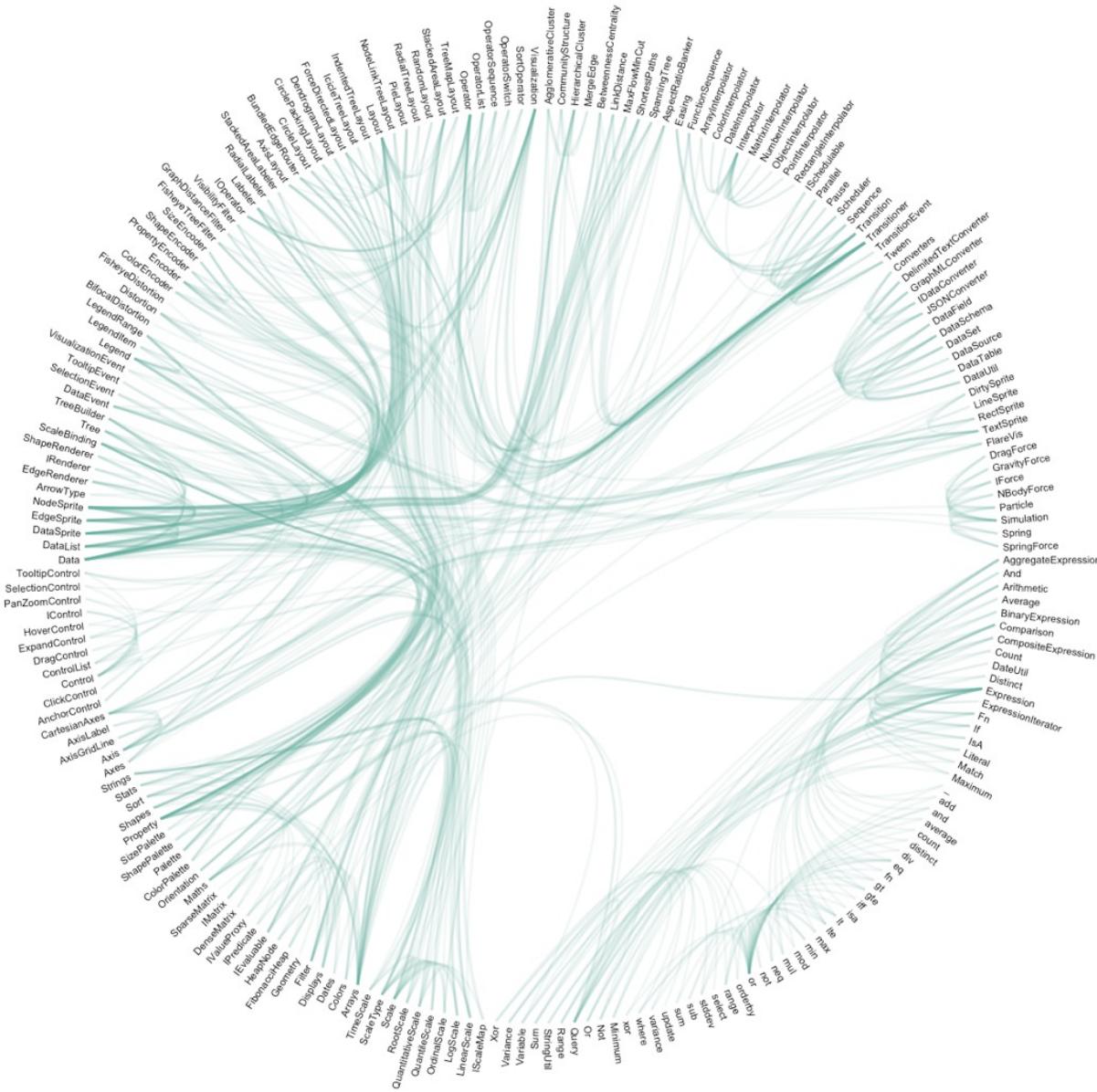
South Korea > Spain ?

% of total sold by the US ?

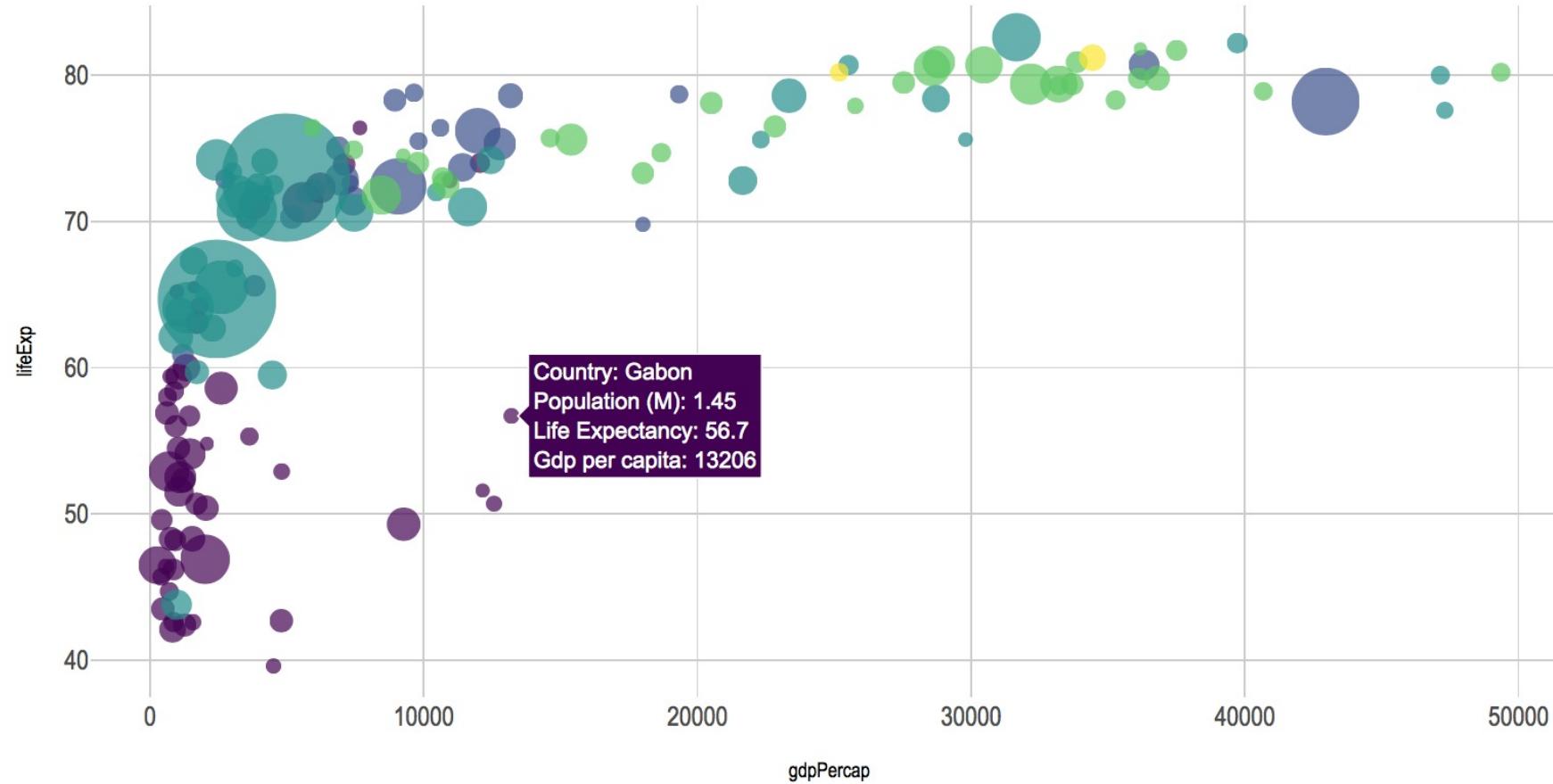
Question – **Exploratory Data Analysis** – Reader – Data – Interactivity – Eye catching



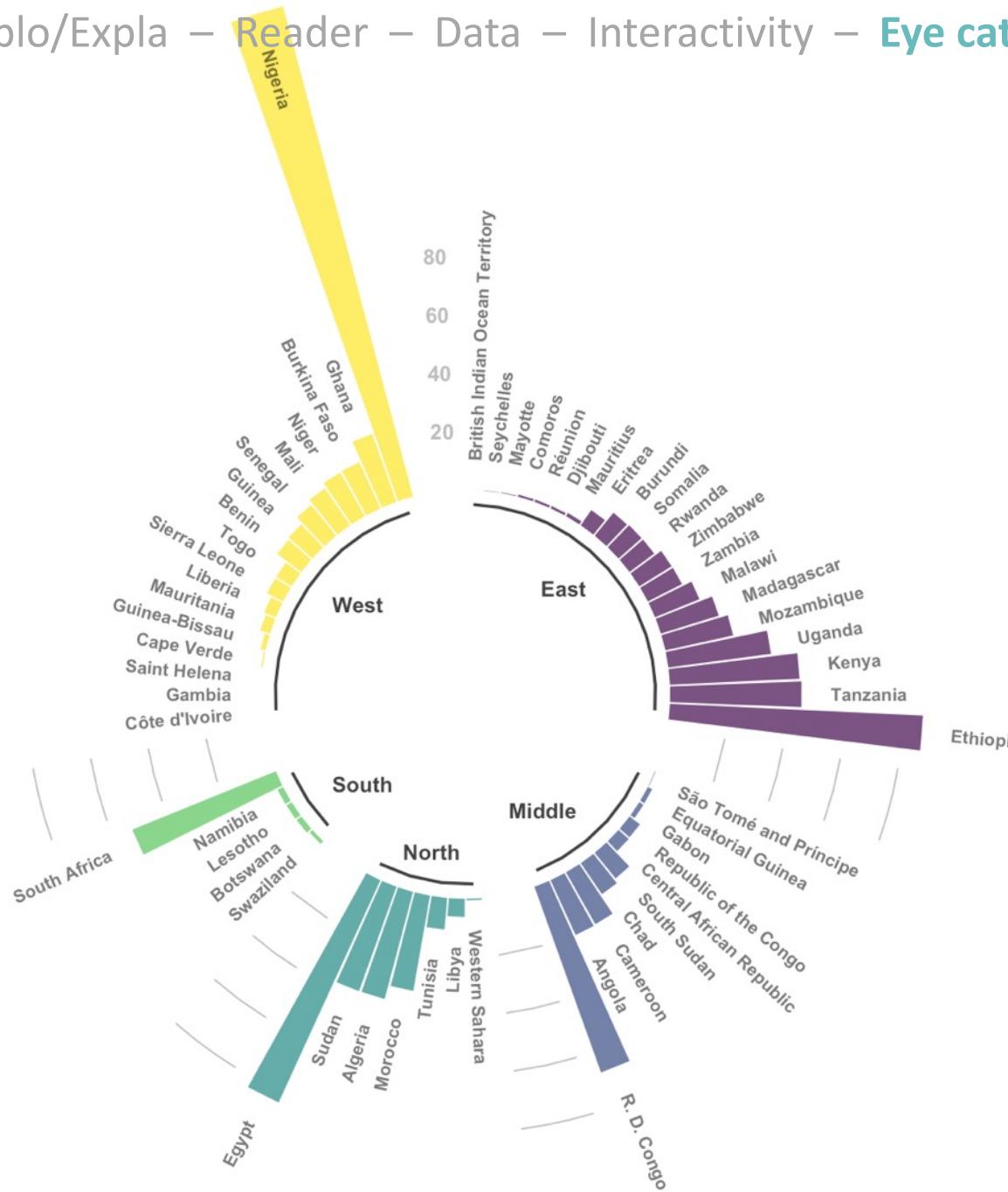
Question – Explo/Expla – **Reader** – Data – Interactivity – Eye catching



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

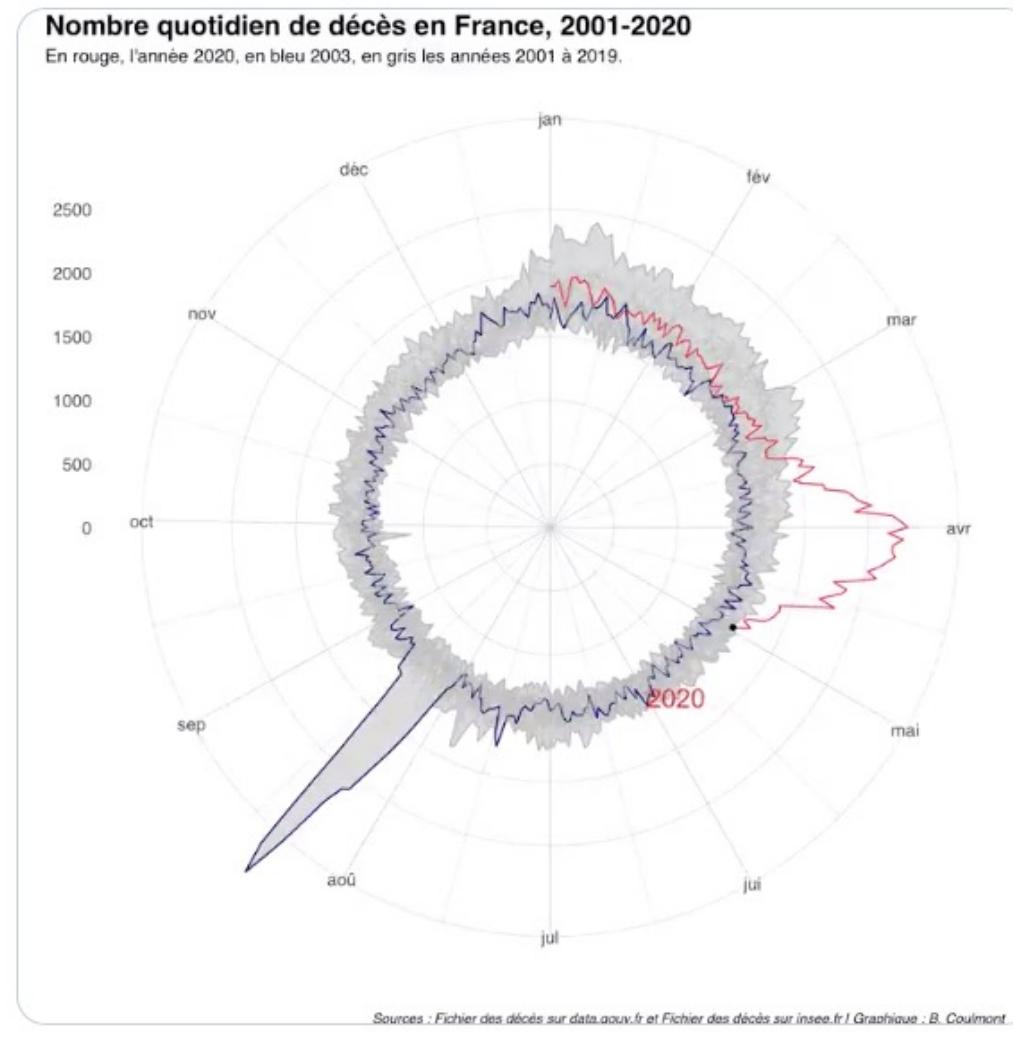


Question – Explo/Expla – Reader – Data – Interactivity – Eye catching





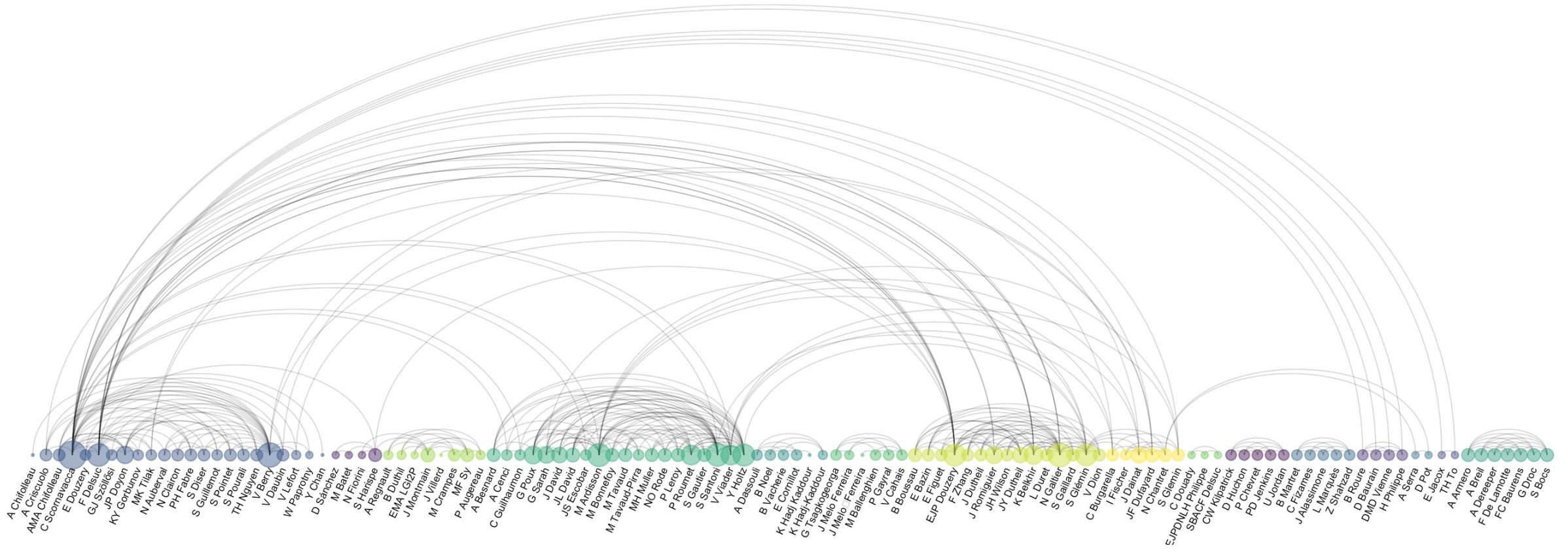
Nombre quotidien de décès en France, 2001-2020, en version animée, coordonnées polaires.



[Link](#)

12:43 PM · 2 déc. 2020 · Twitter Web App

6,5 k Retweets **1 k** Tweets cités **16,1 k** J'aime



Co-authorship network of a researcher



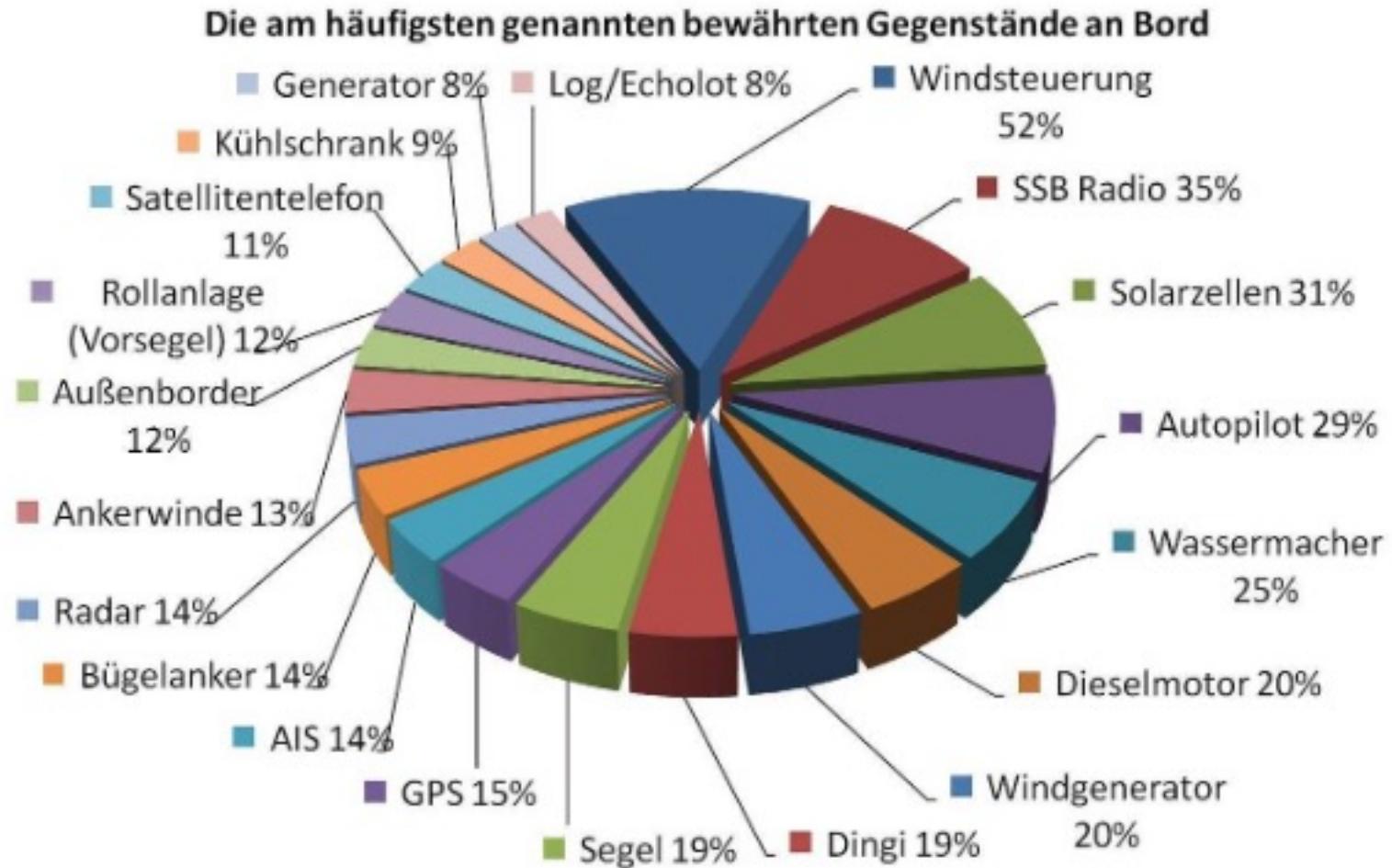
Where surfers travel.

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

WHAT YOU SHOULD **NOT** DO

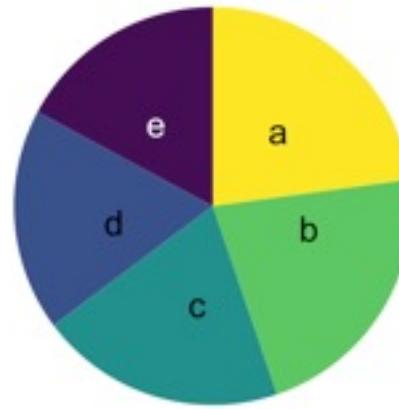
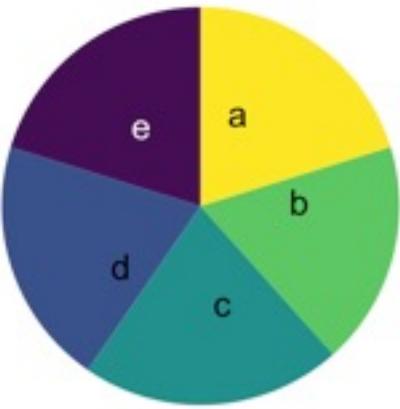
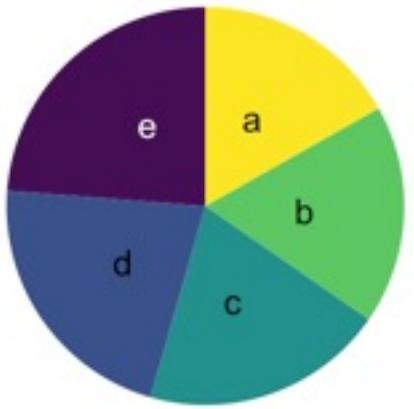
A gallery of most common caveats

What's wrong with
this chart?



Source: [WTF Visualizations](#)

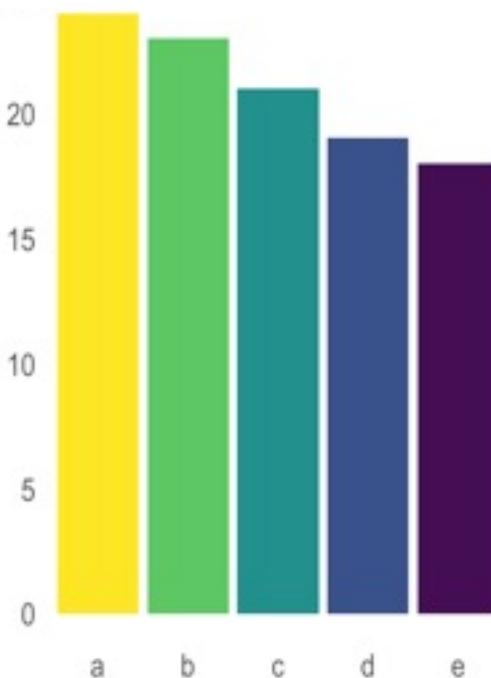
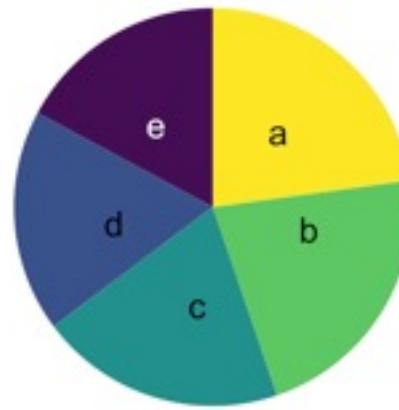
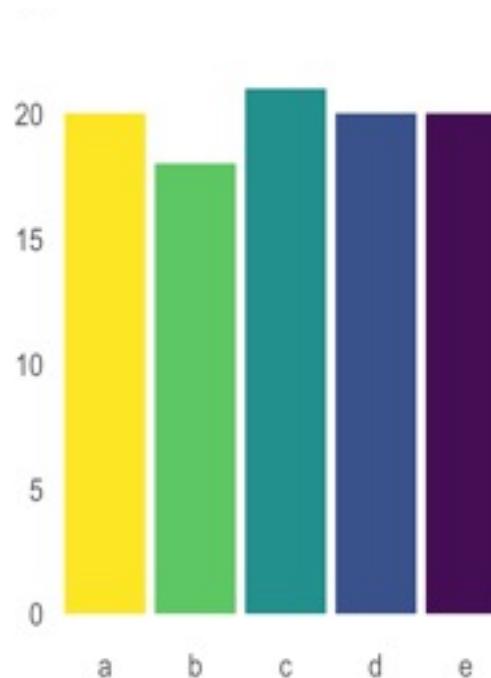
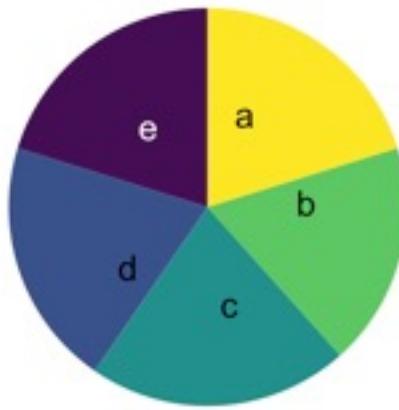
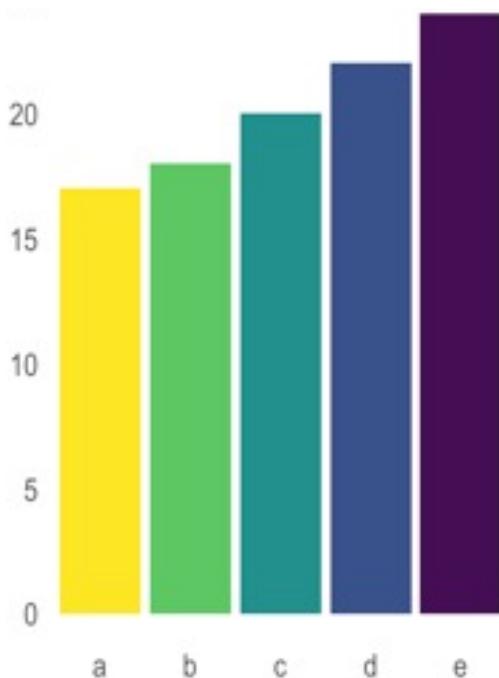
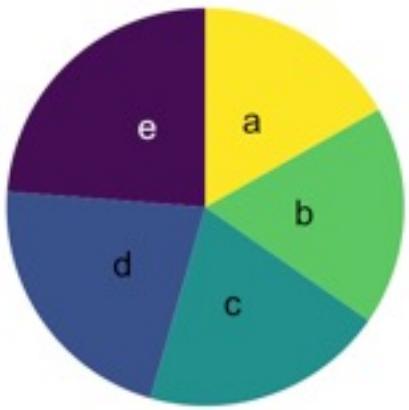
What's wrong
with pie chart?



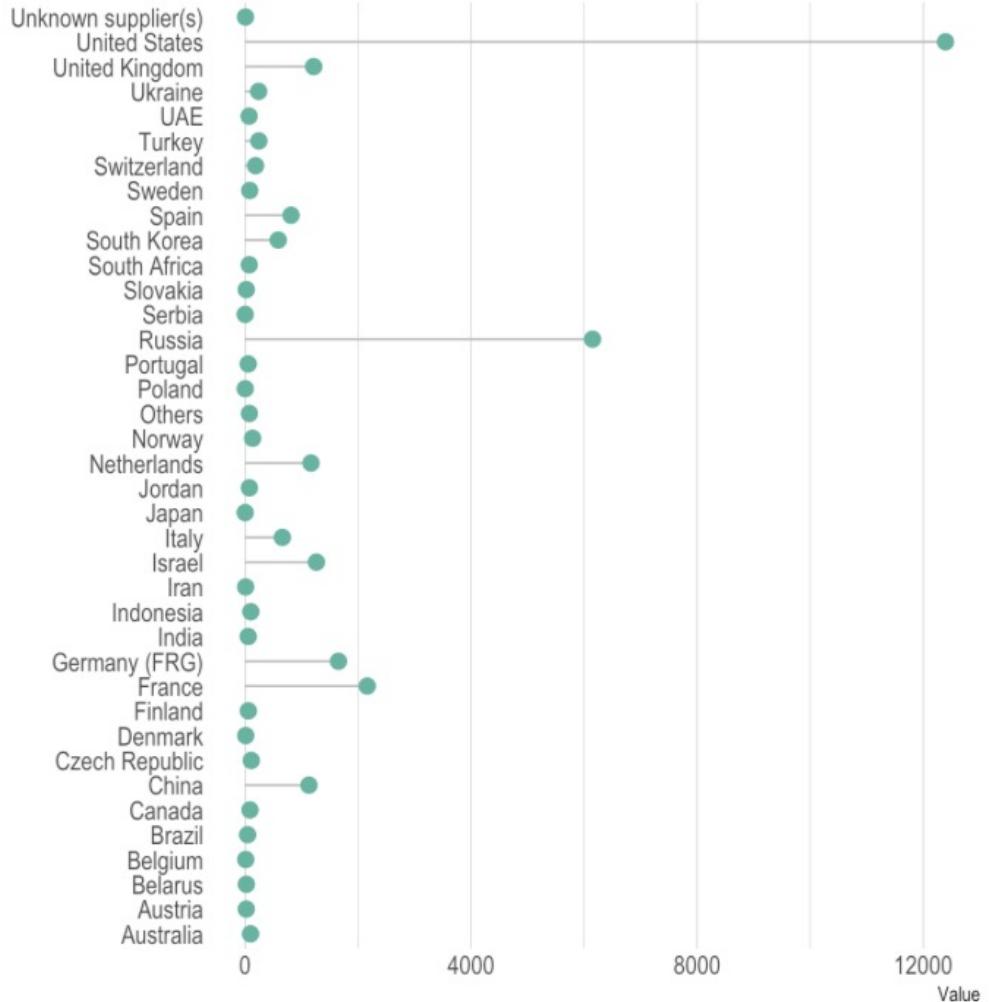
What can you see?

What's wrong with
pie chart?

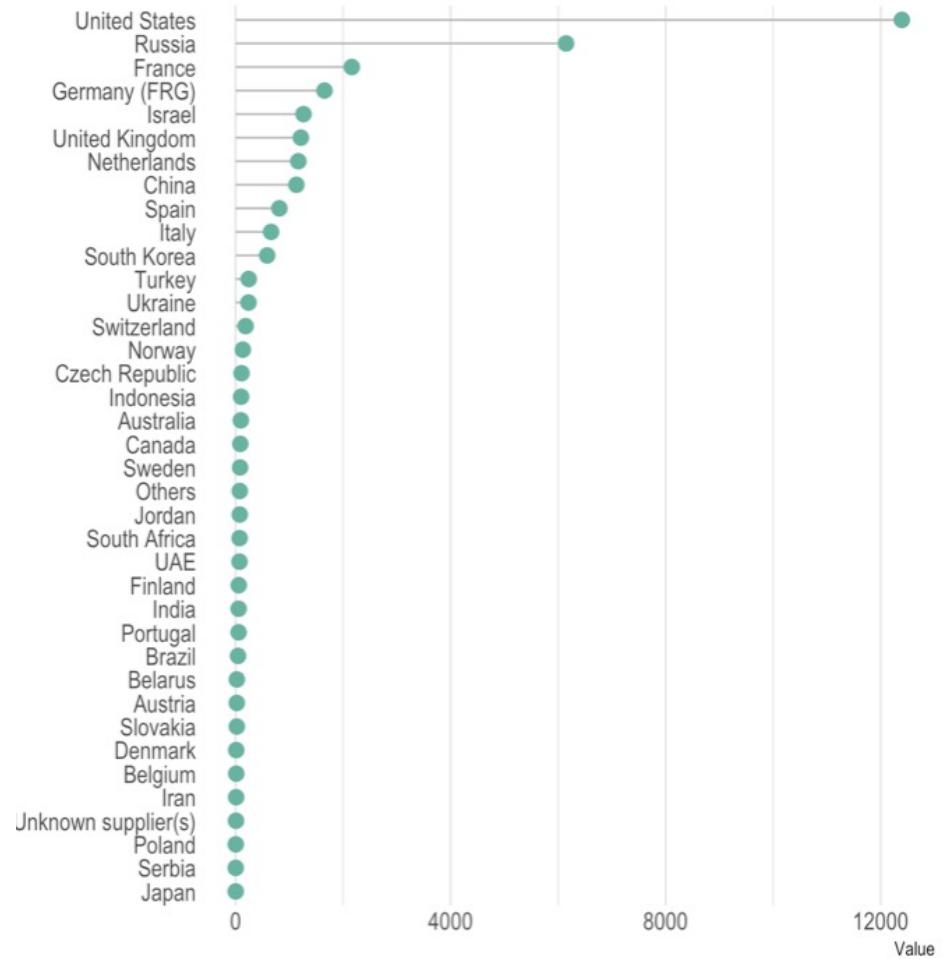
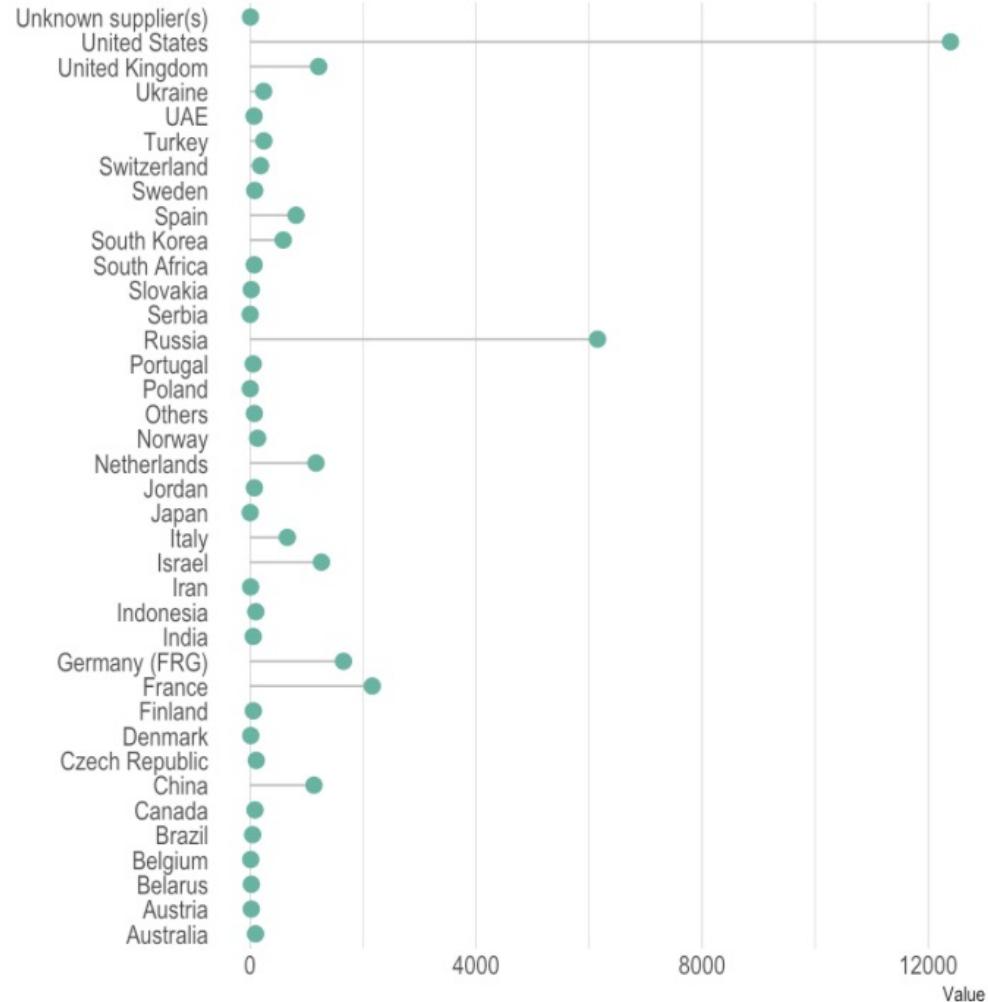
It is hard to
distinguish angles



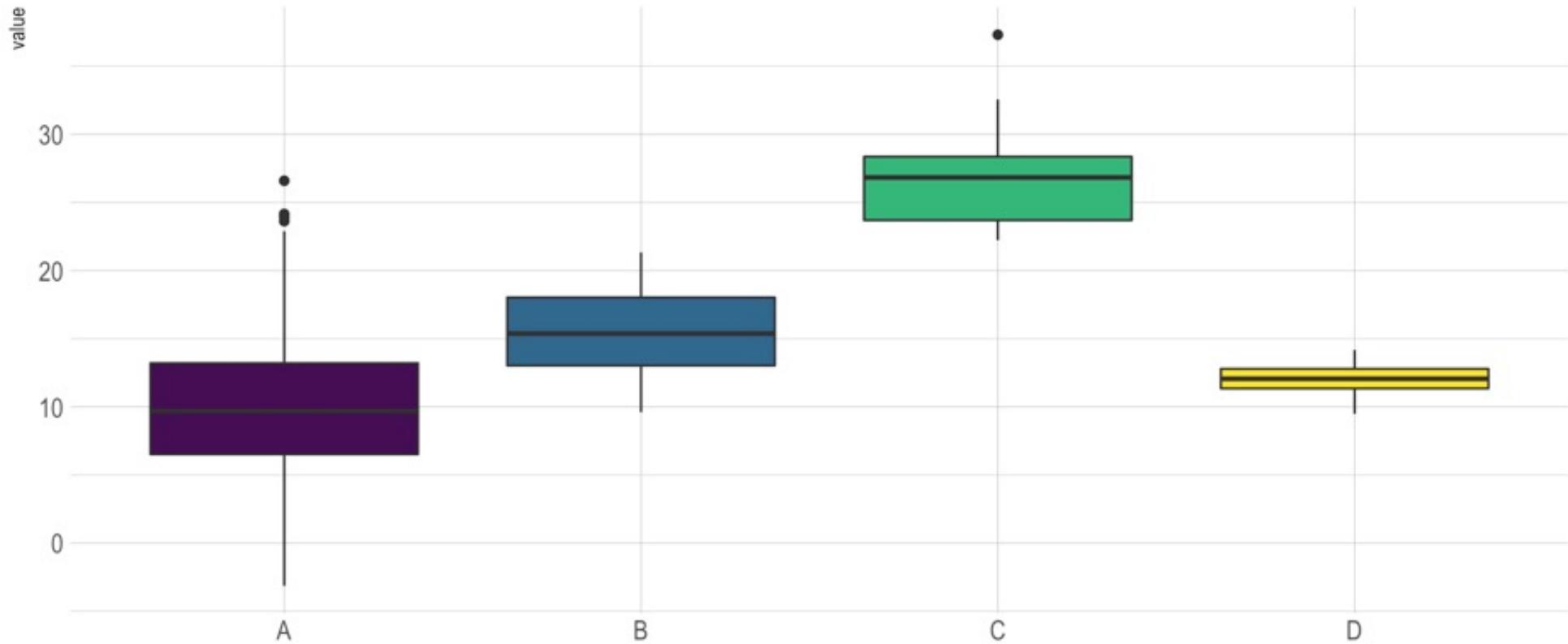
What could be better
here?



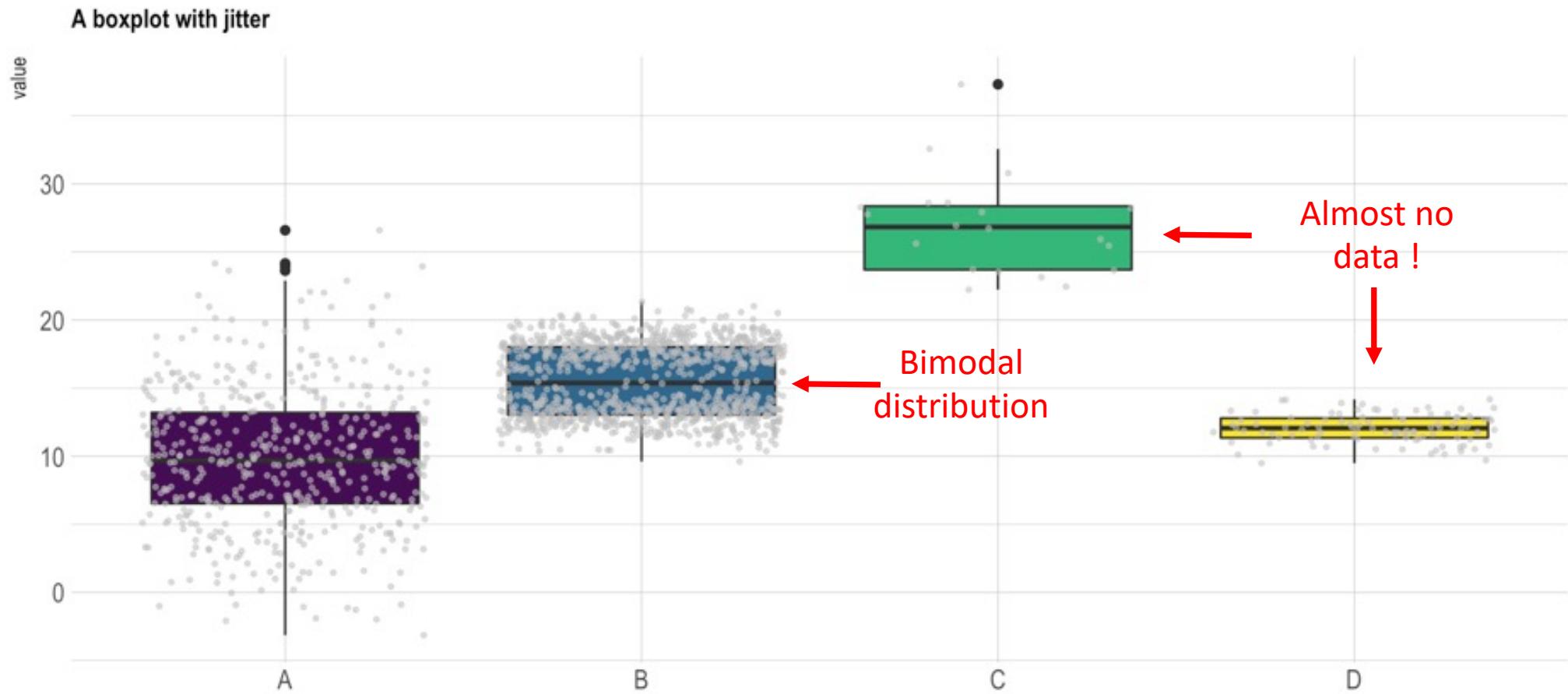
Order your data



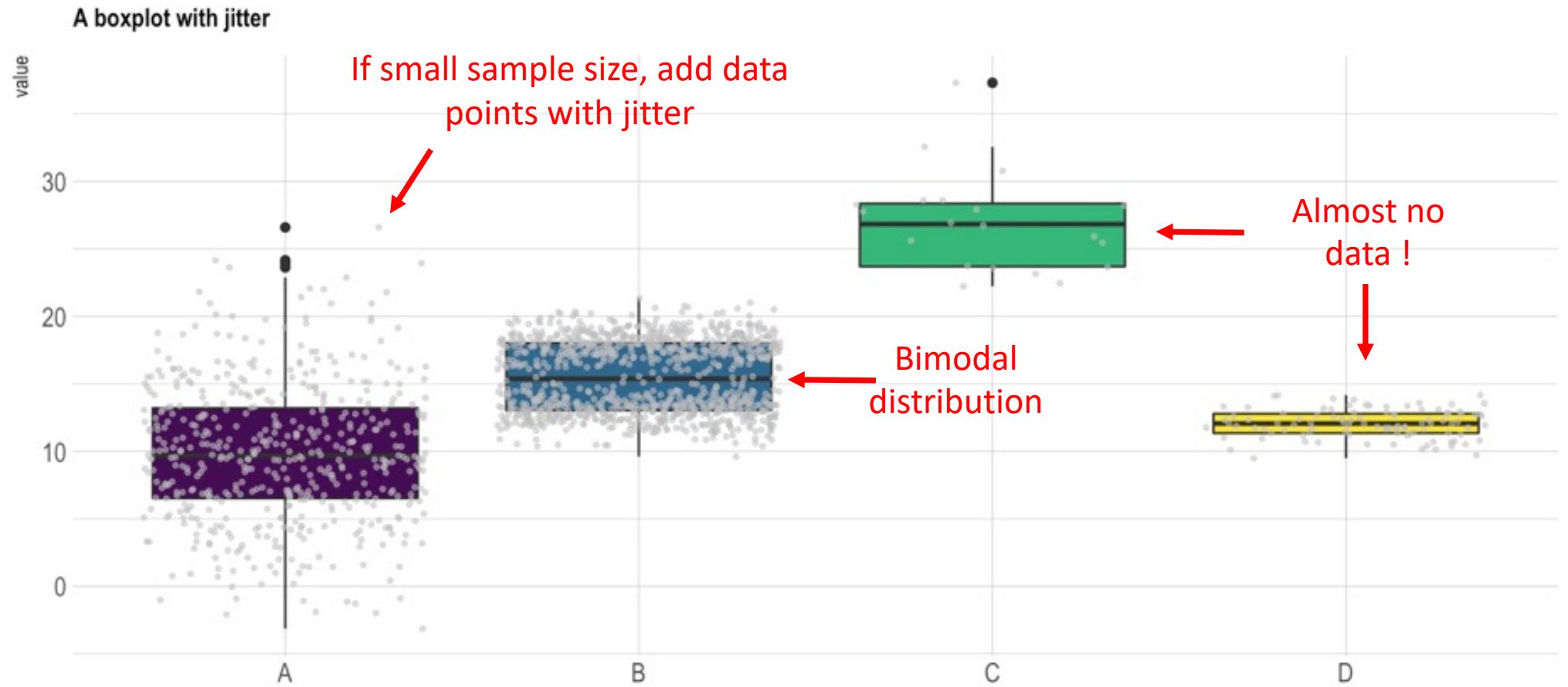
Anything wrong here?



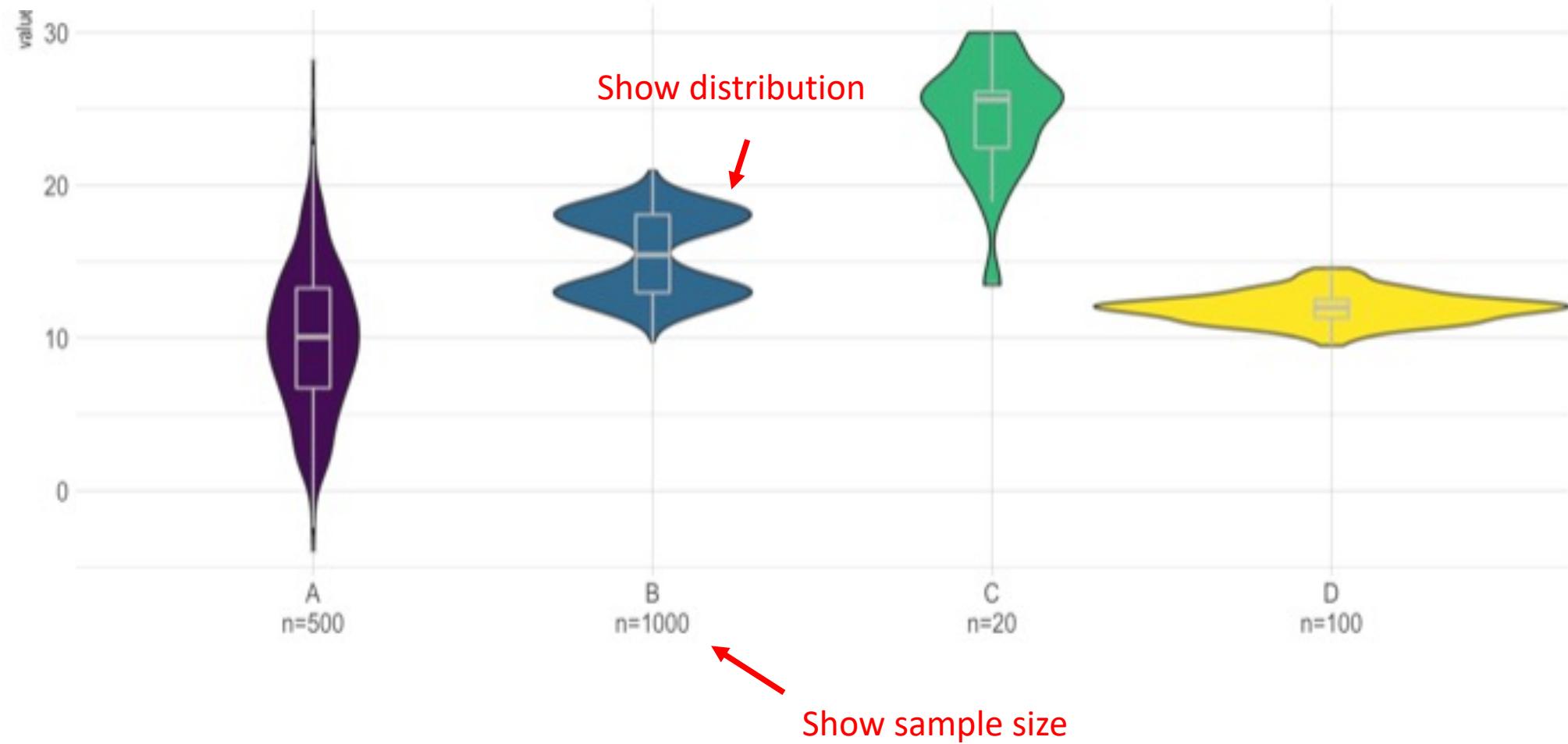
Boxplot = hide information



Boxplot = hide information



If big sample size, use violin plot





AMERICA'S ECONOMY

2010 GROSS DOMESTIC PRODUCT



United States
\$14.6 TRILLION



China \$5.7 TRILLION



Japan \$5.3 TRILLION

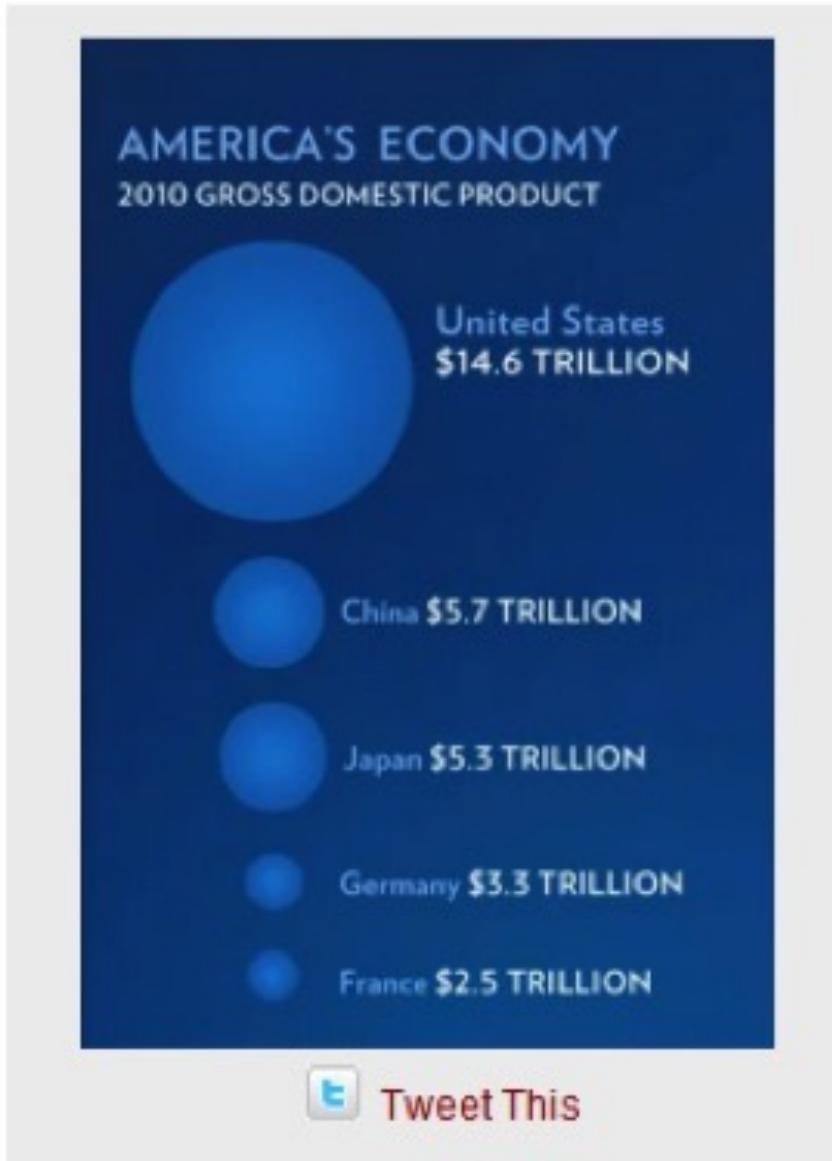


Germany \$3.3 TRILLION



France \$2.5 TRILLION

Size = radius

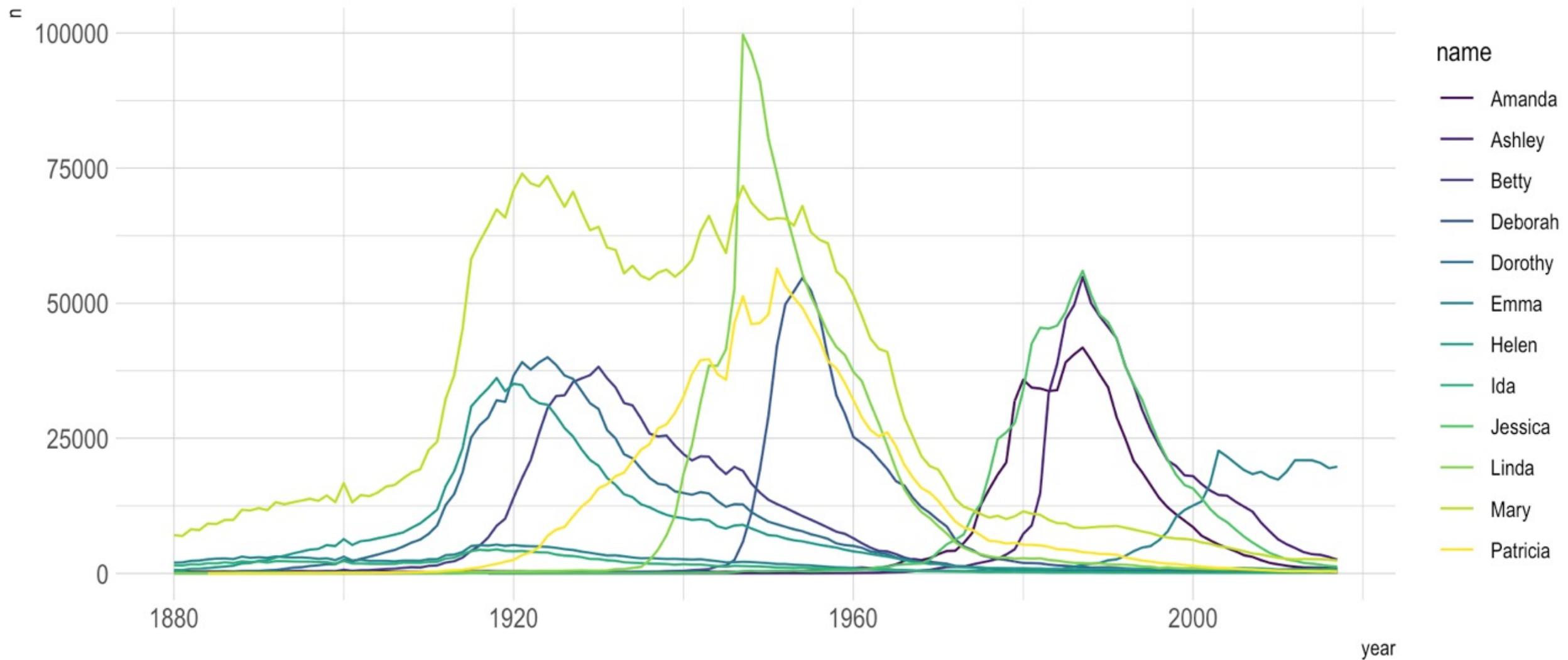


Size = area

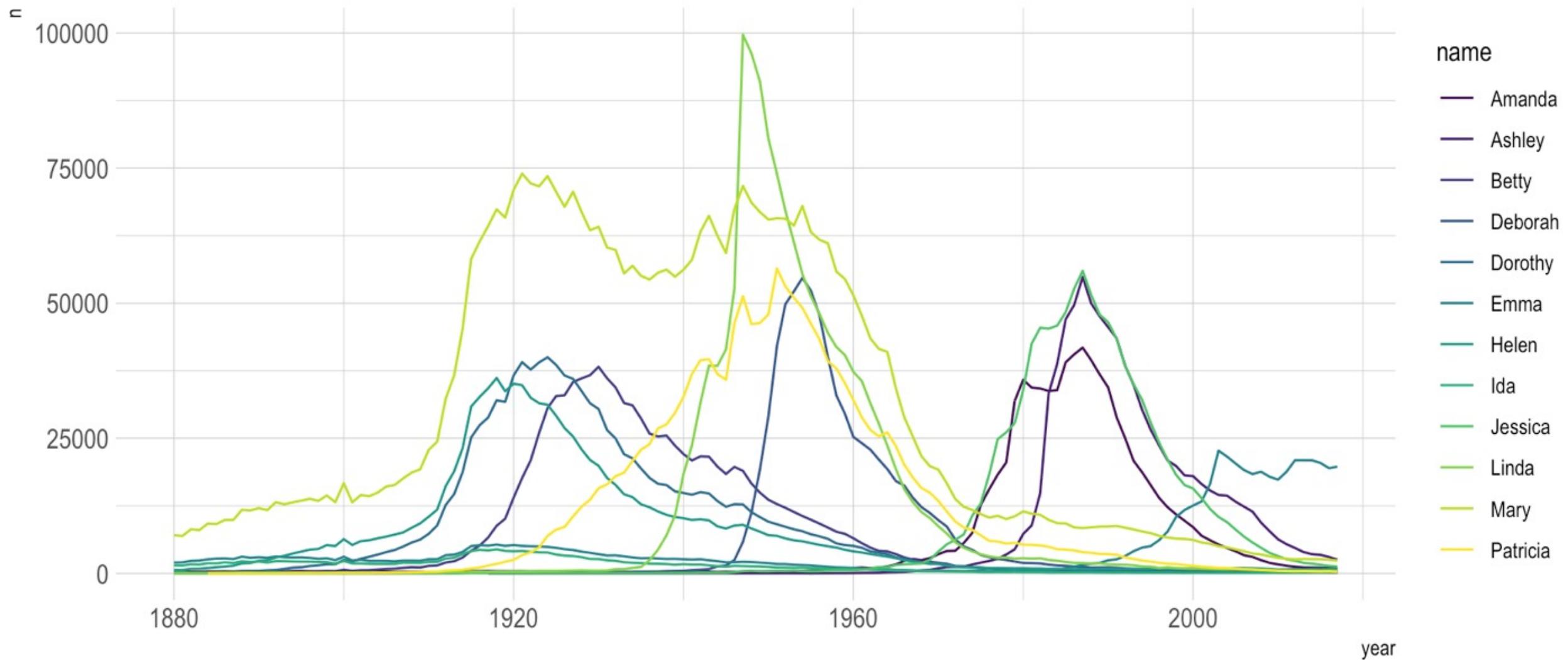


Source: [Fast Fedora blog](#)

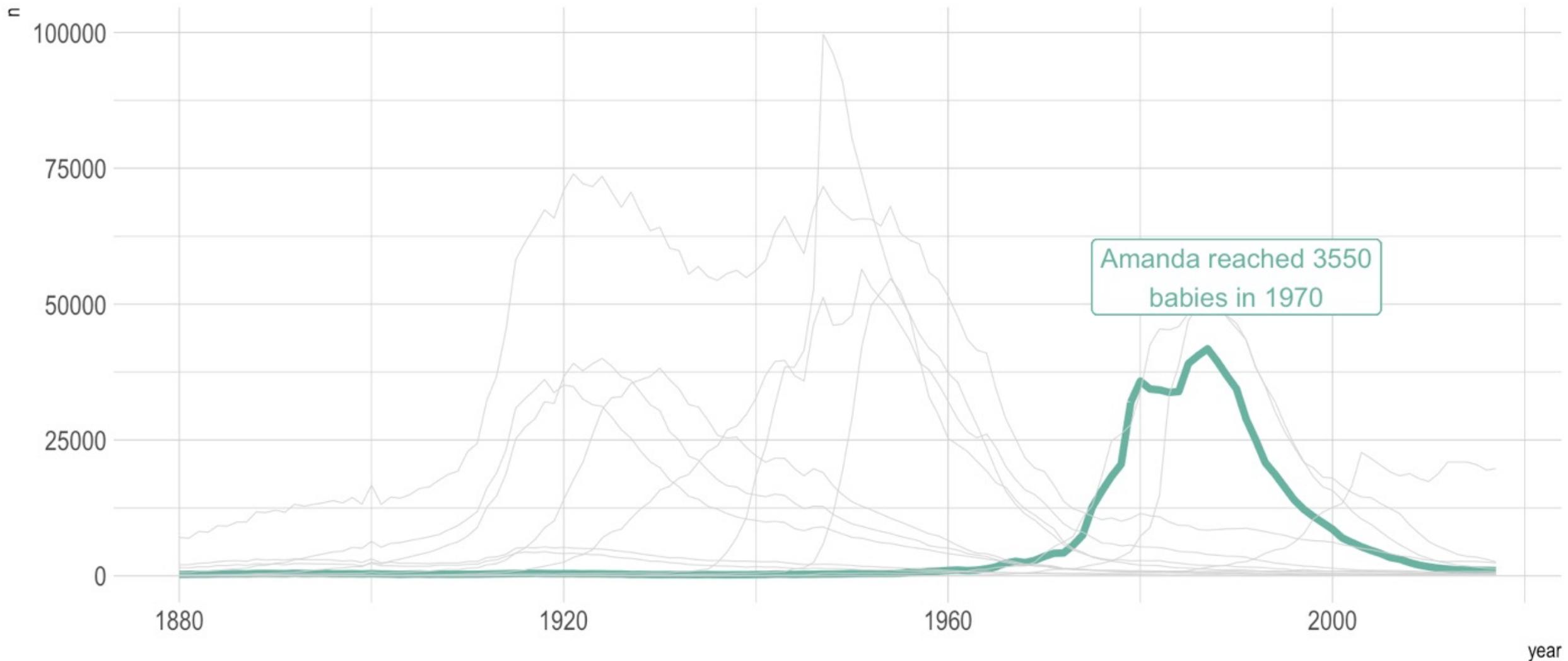
A spaghetti chart of baby names popularity



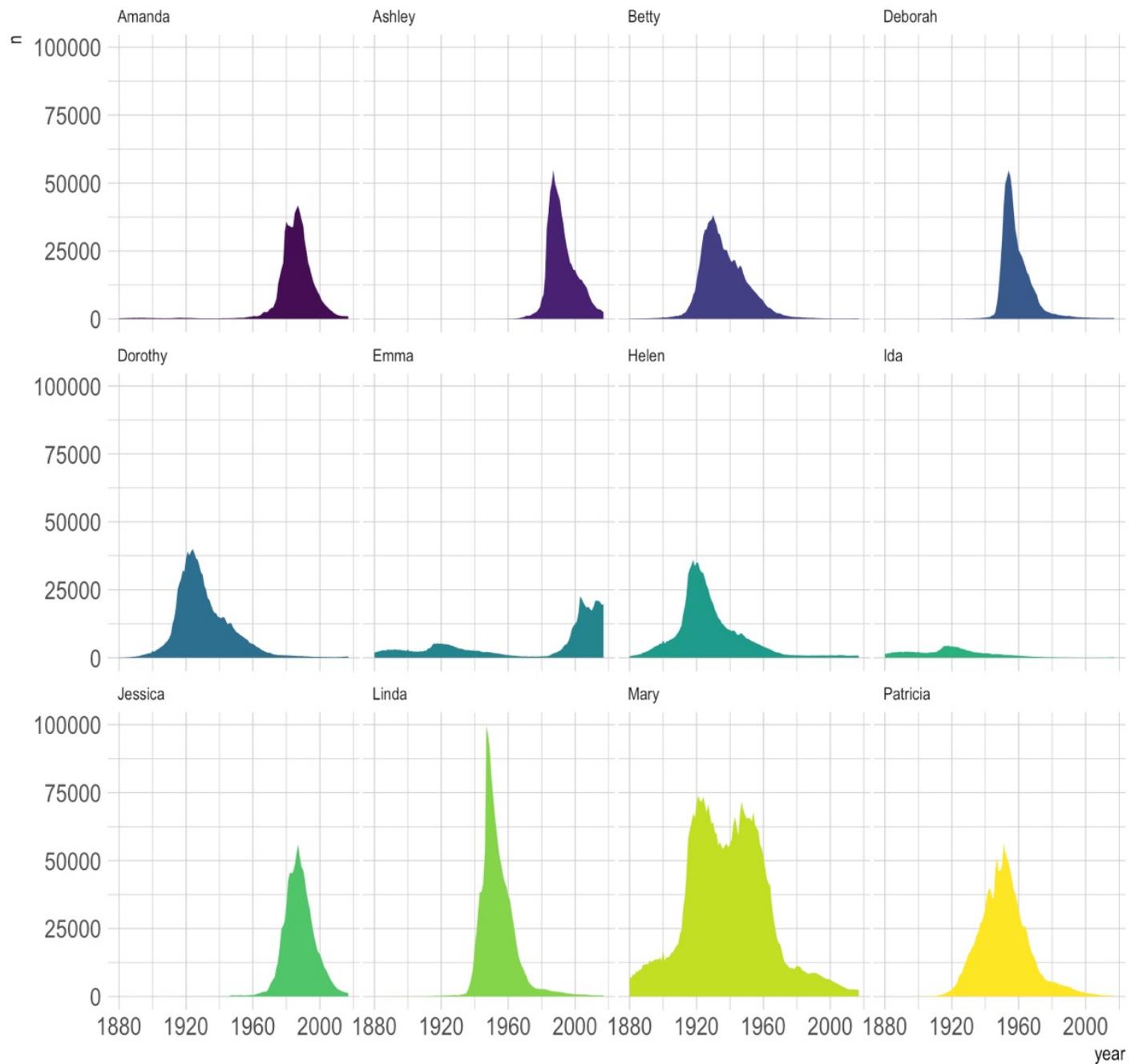
A spaghetti chart of baby names popularity



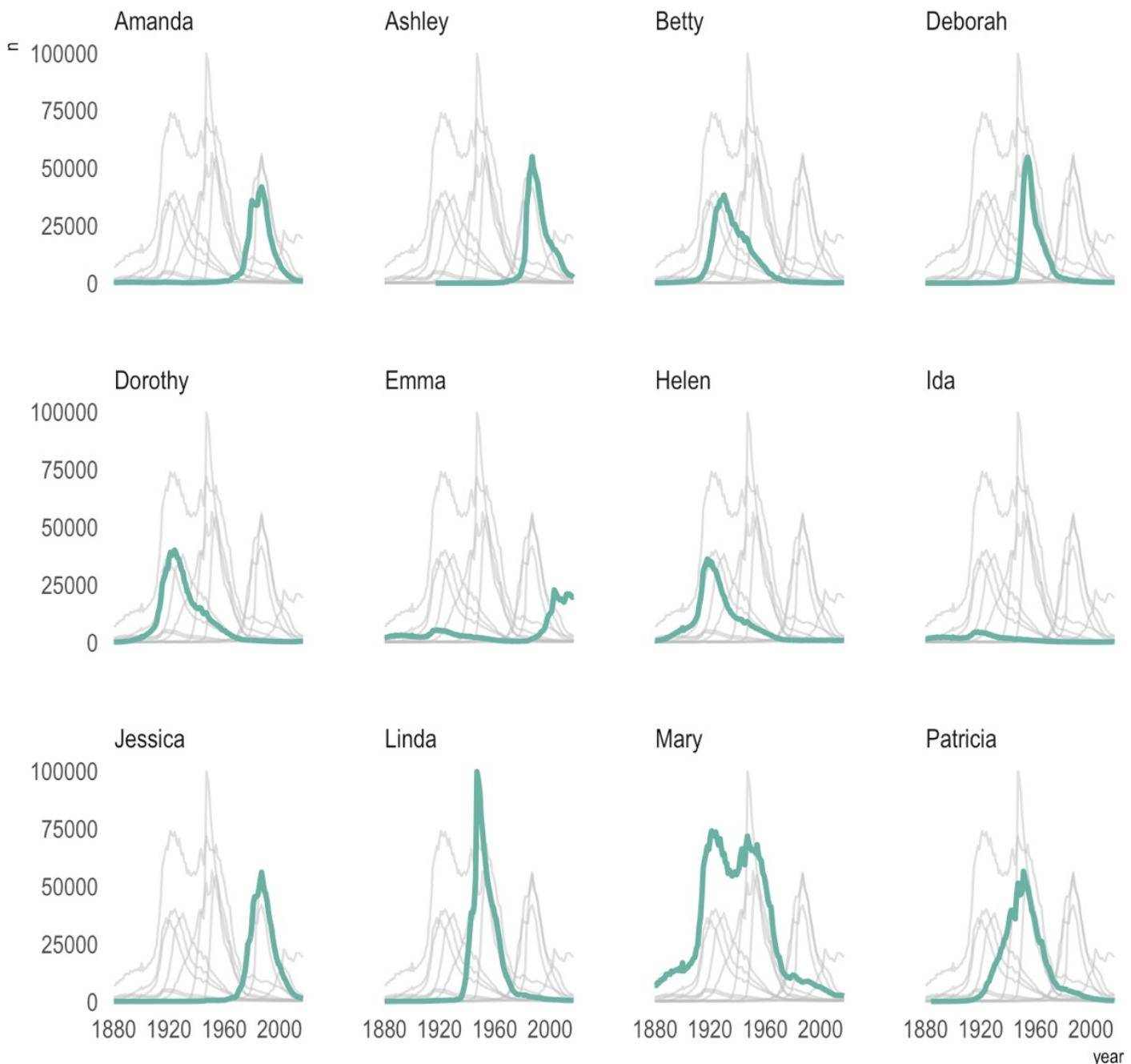
Popularity of American names in the previous 30 years

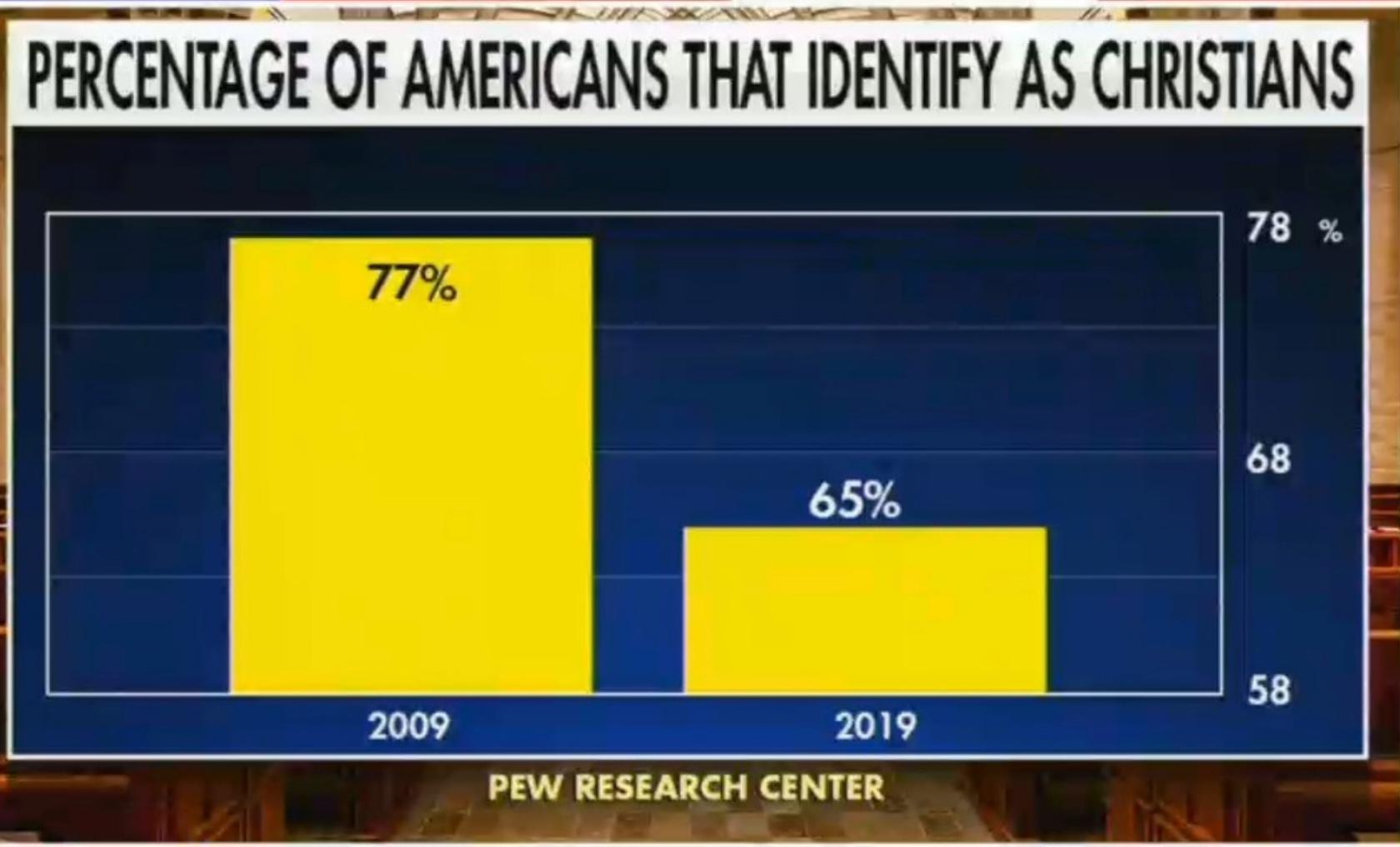


Popularity of American names in the previous 30 years



A spaghetti chart of baby names popularity





WE USED TO BE AN ENTHUSIASTICALLY CHRISTIAN NATION
TUCKER CARLSON • TONIGHT •

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.



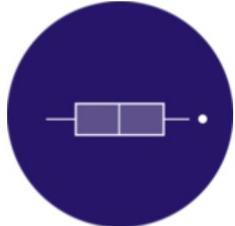
Pie chart

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

Design the idea



20 %

Build it



80 %



Easy

Hard

Limited

Flexible

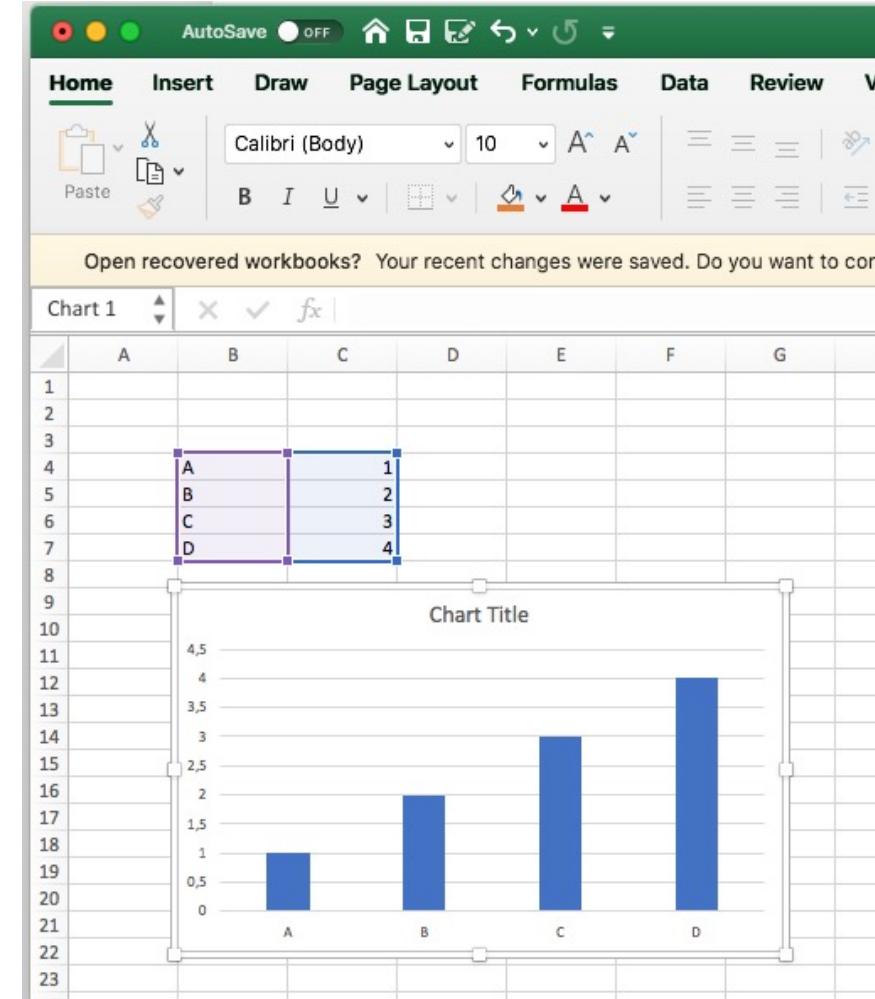




Easy

Excel

Limited

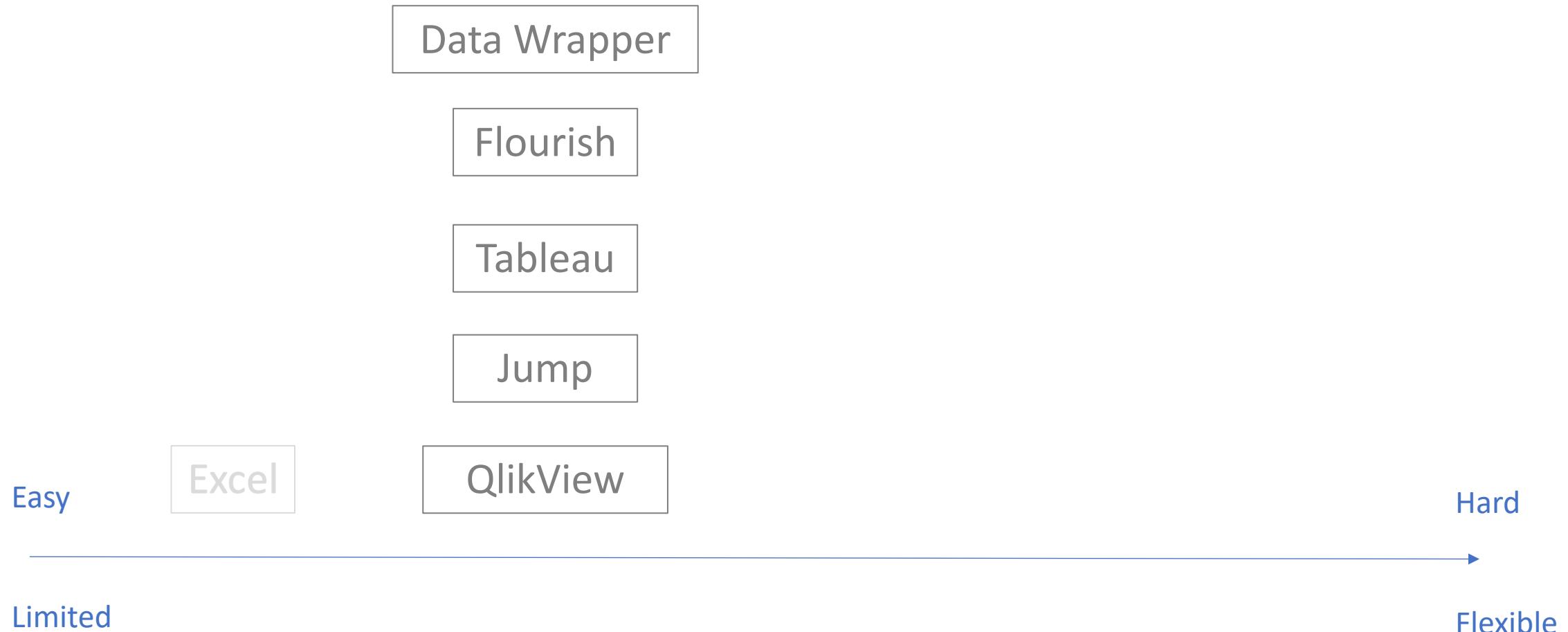


Hard

Flexible

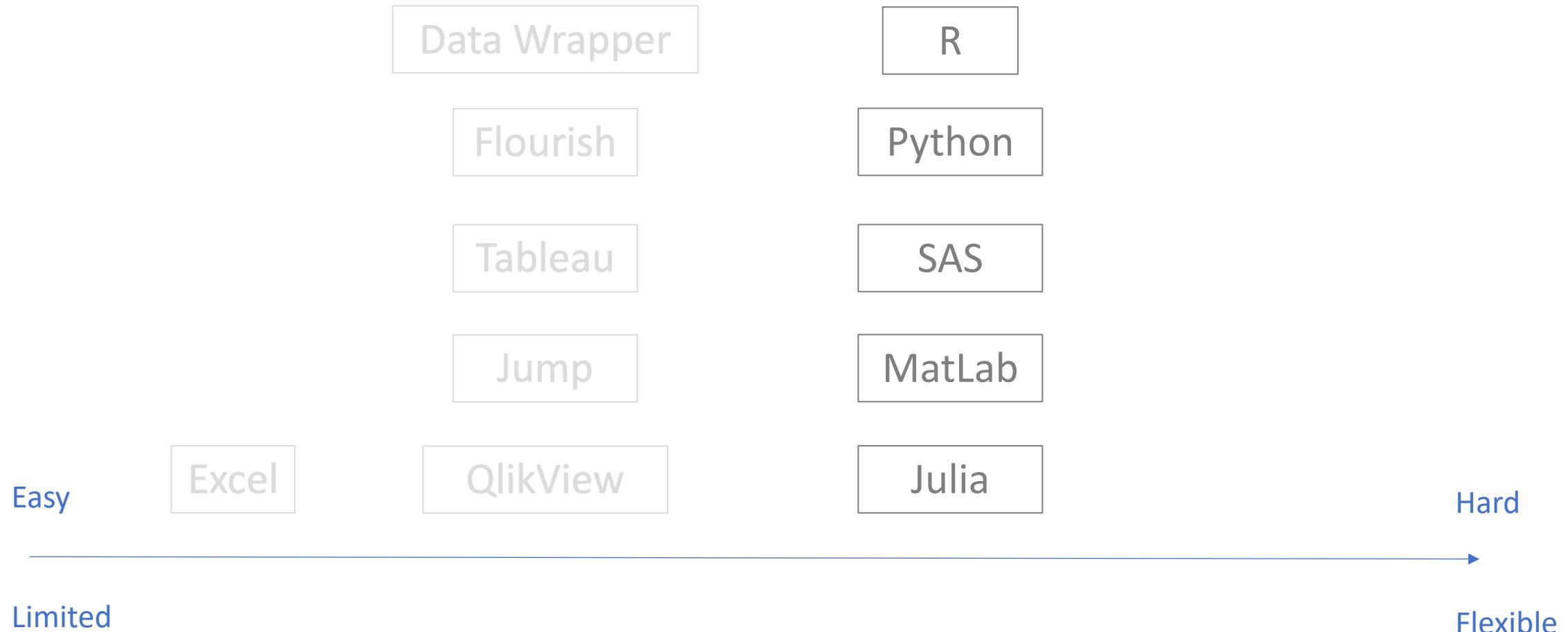


[Data Wrapper demo](#)





[R graph gallery demo](#)



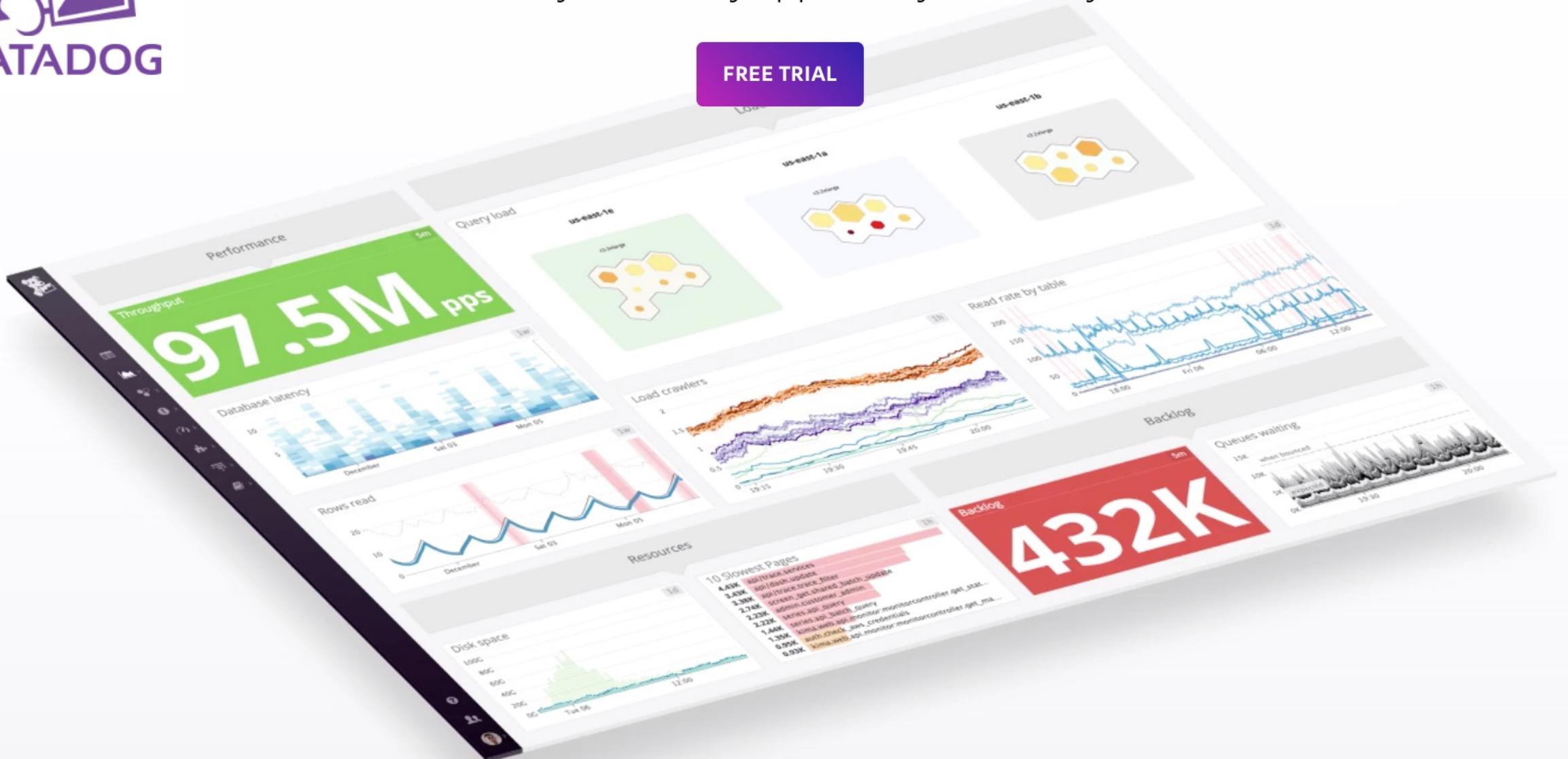




Modern monitoring & analytics

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

FREE TRIAL





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