

Introduction to data visualization

What, how, why, and more

Yan-holtz.com/teaching

What

A picture is worth a thousand words

Taking data,
finding the
good way to
show it

Using picture to
represent data

List of chart
type

John snow map

Don't have to
be boring

Building chart -> Building infographiz -> building interactivity -> Dashboard -
> story telling

Looking for a chart ?

What you
can do

What you
should do

Caveats to
avoid

How to
build it

Who sells more weapons ?

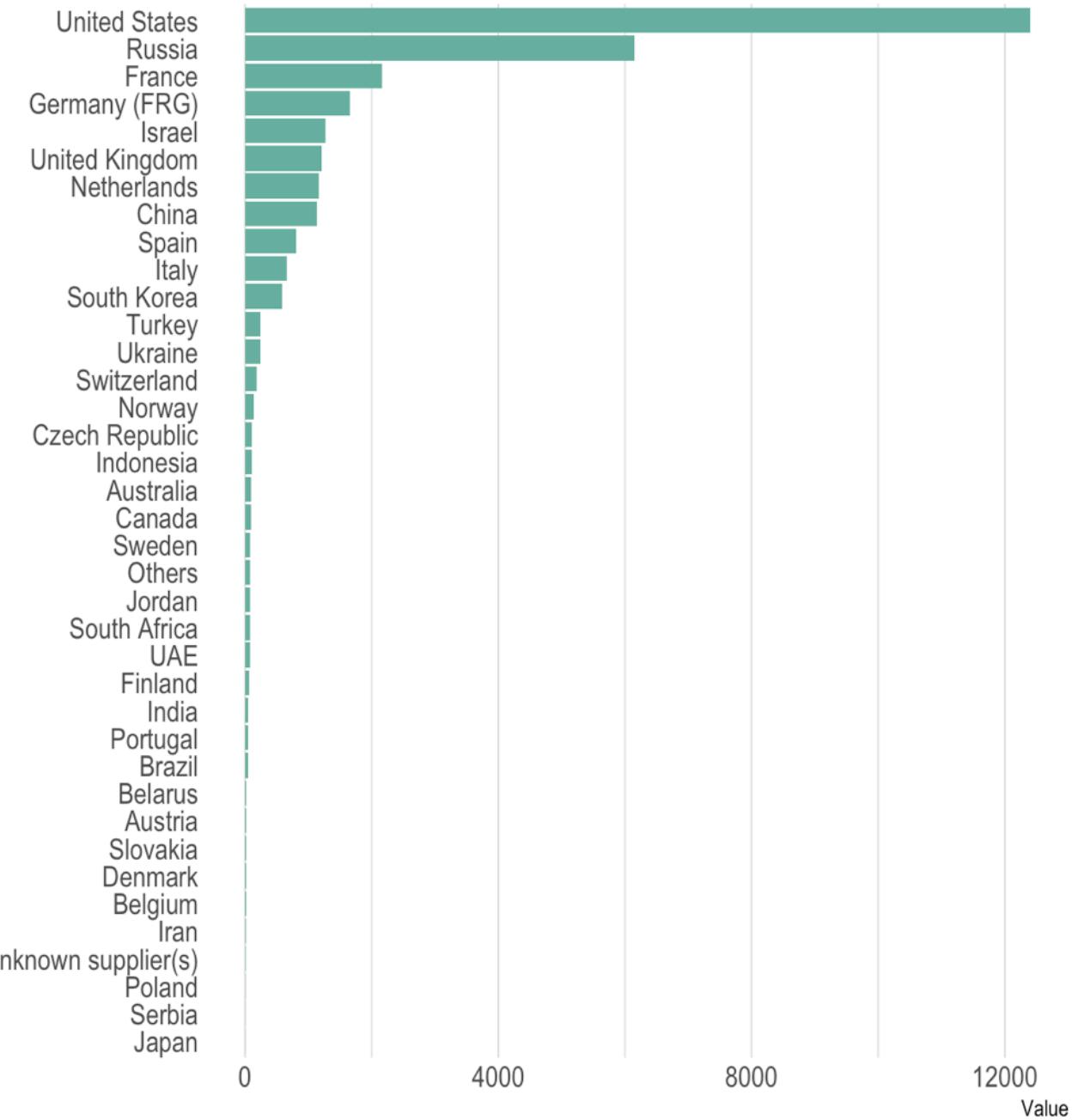
Country	Value
United States	12394
Russia	6148
Germany (FRG)	1653
France	2162
United Kingdom	1214
China	1131



?

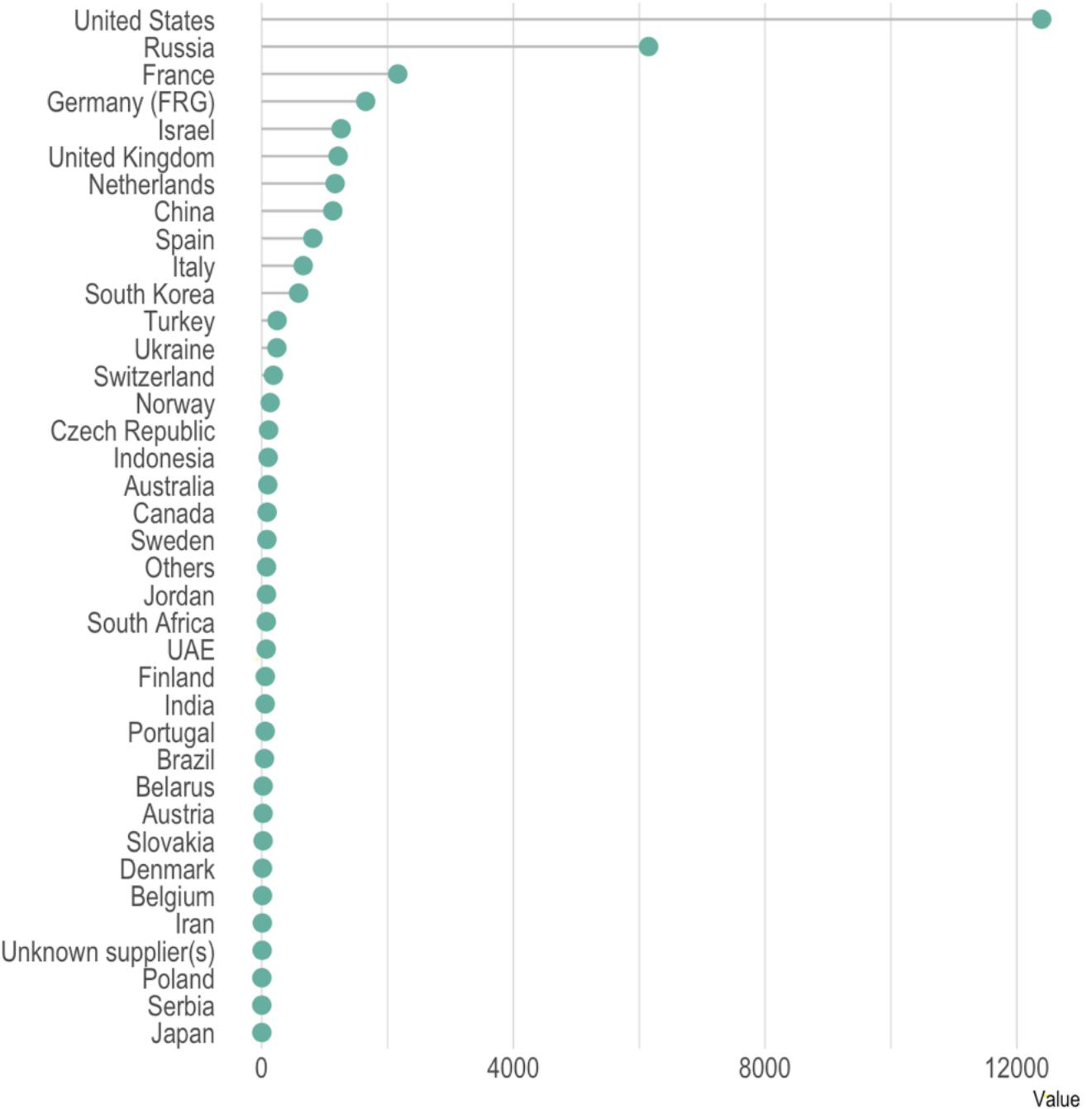
Who sells more weapons ?

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



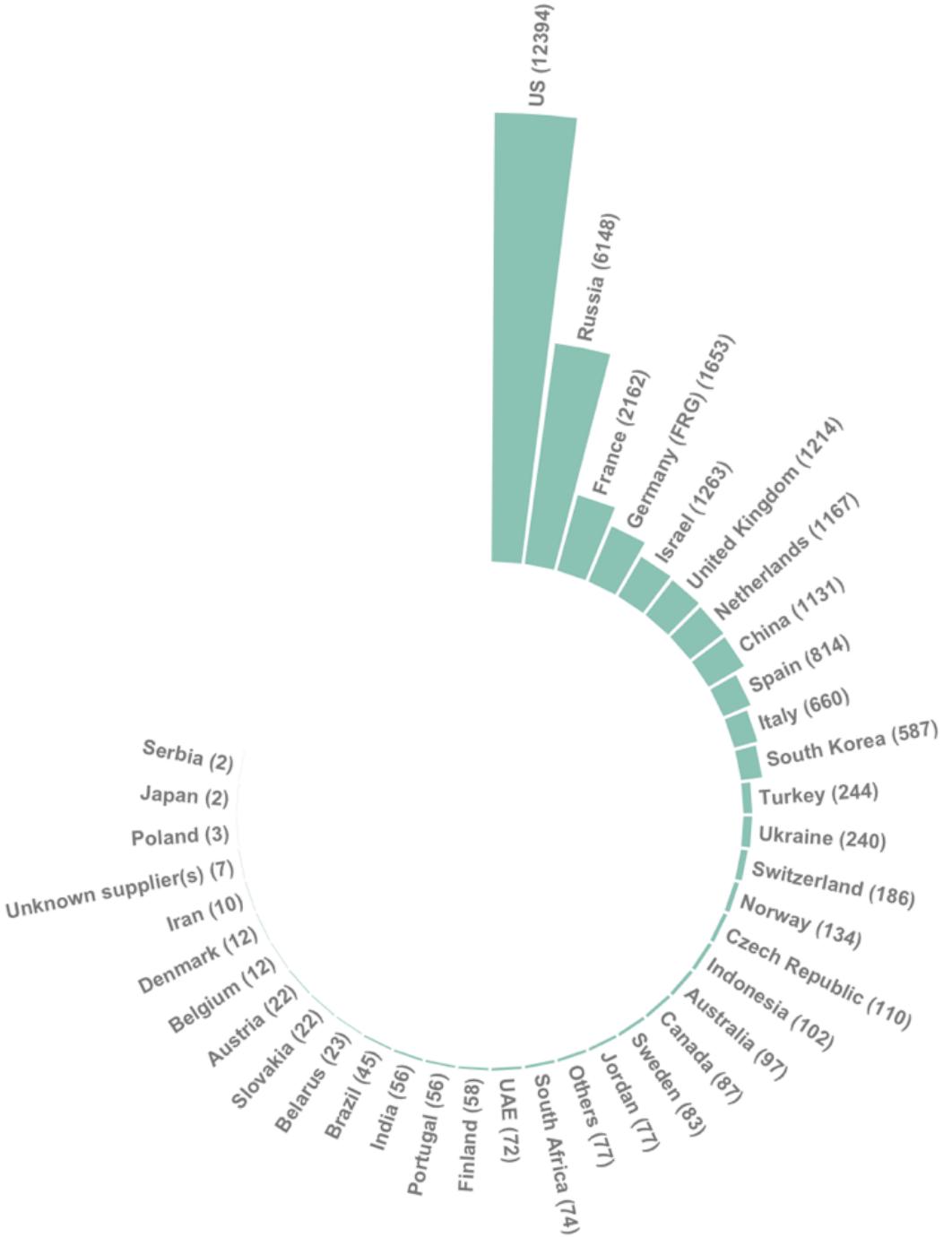
Who sells more weapons ?

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



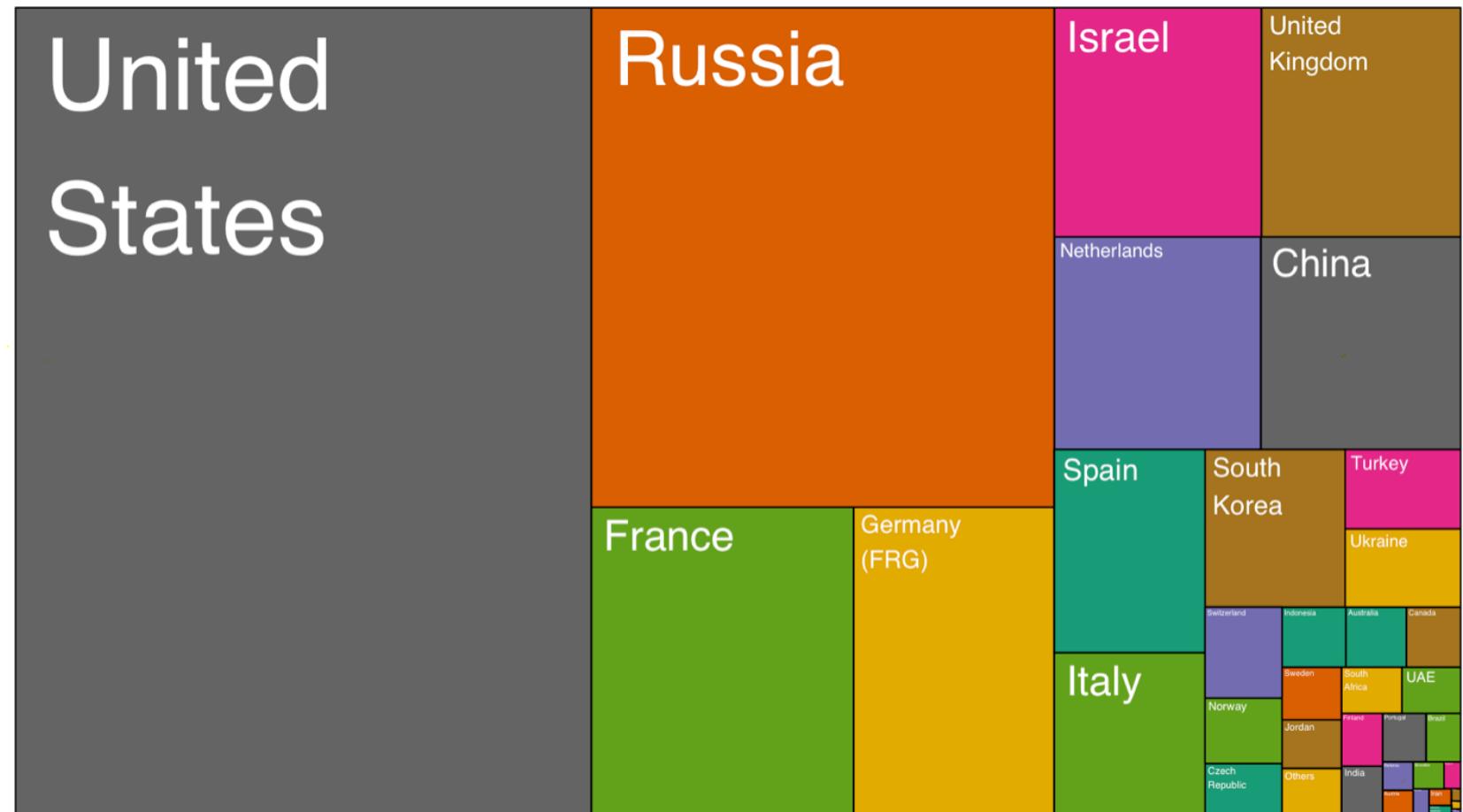
Who sells more weapons ?

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



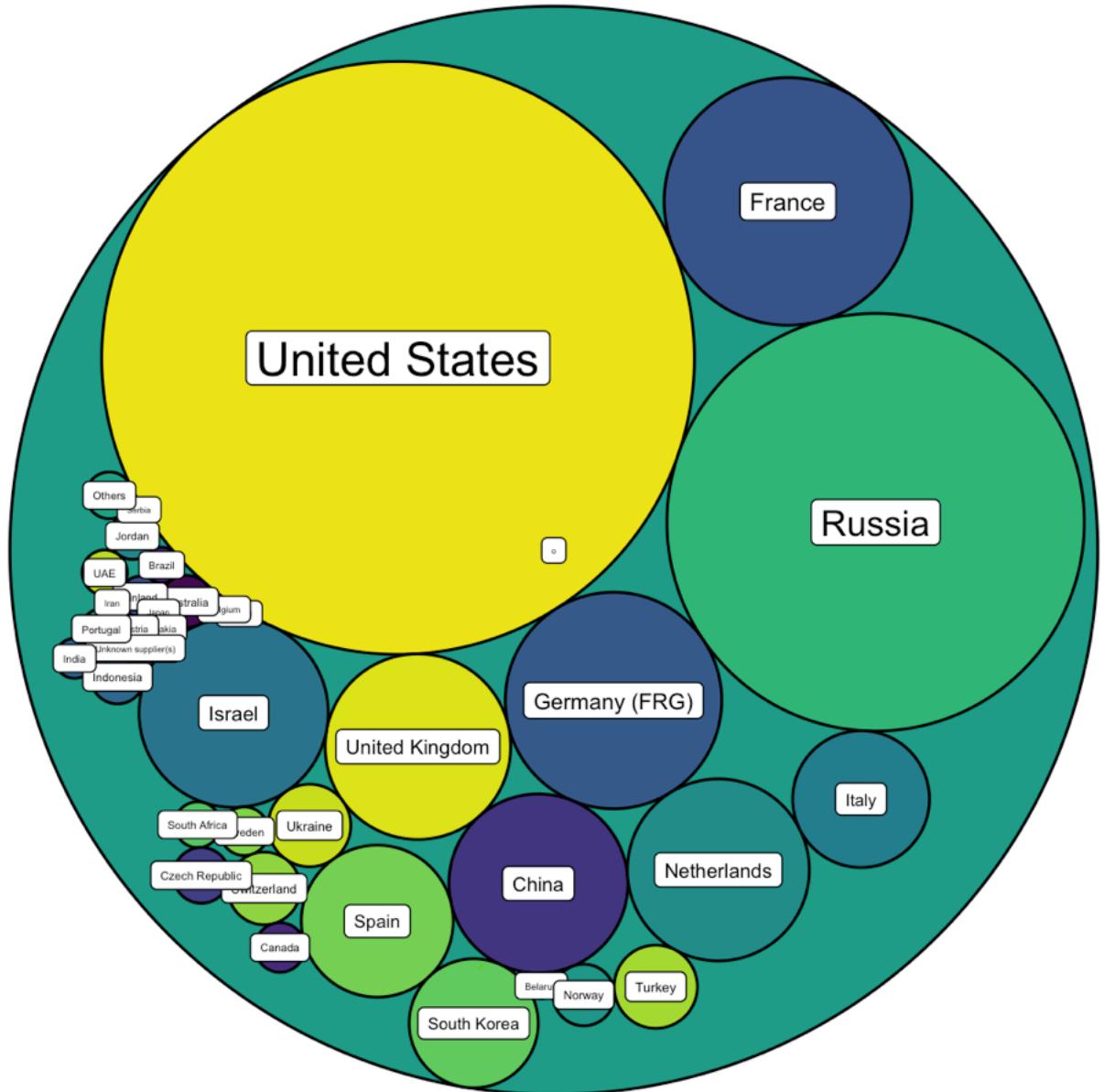
Who sells more weapons ?

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



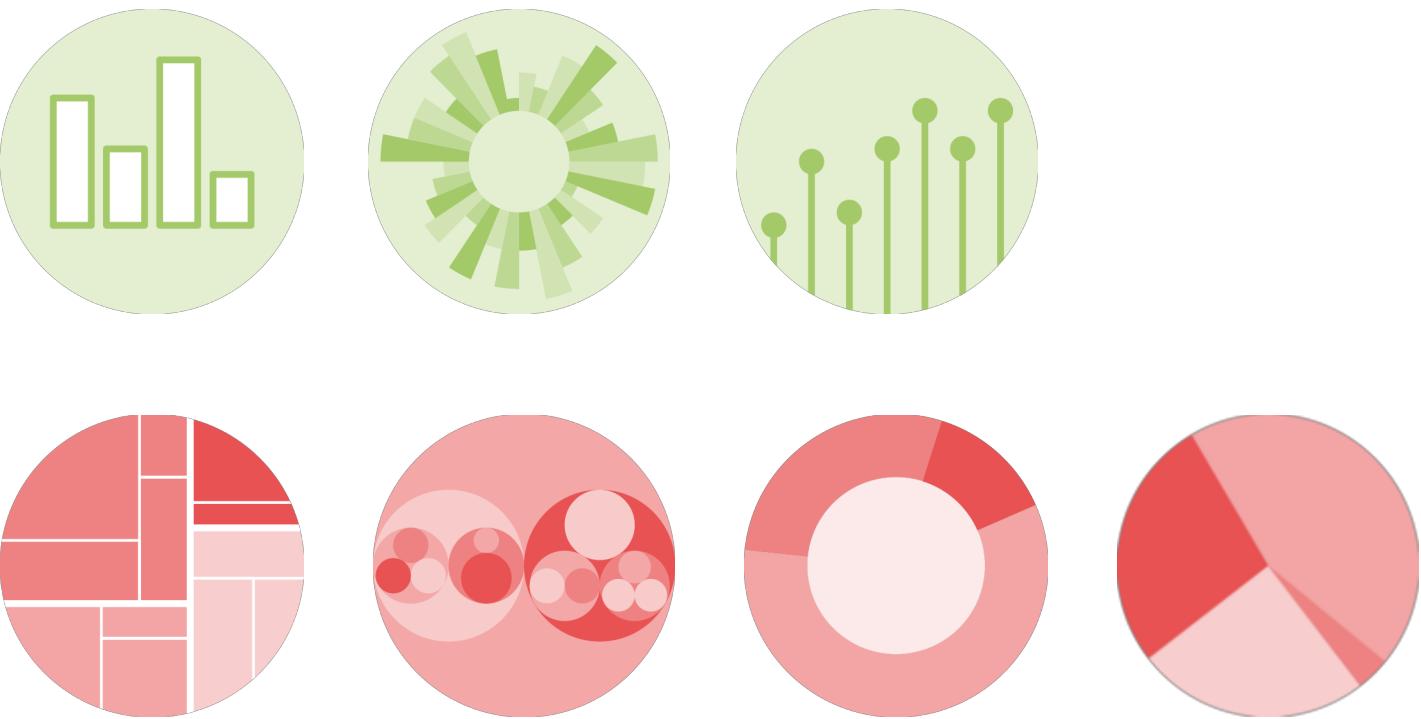
Who sells more weapons ?

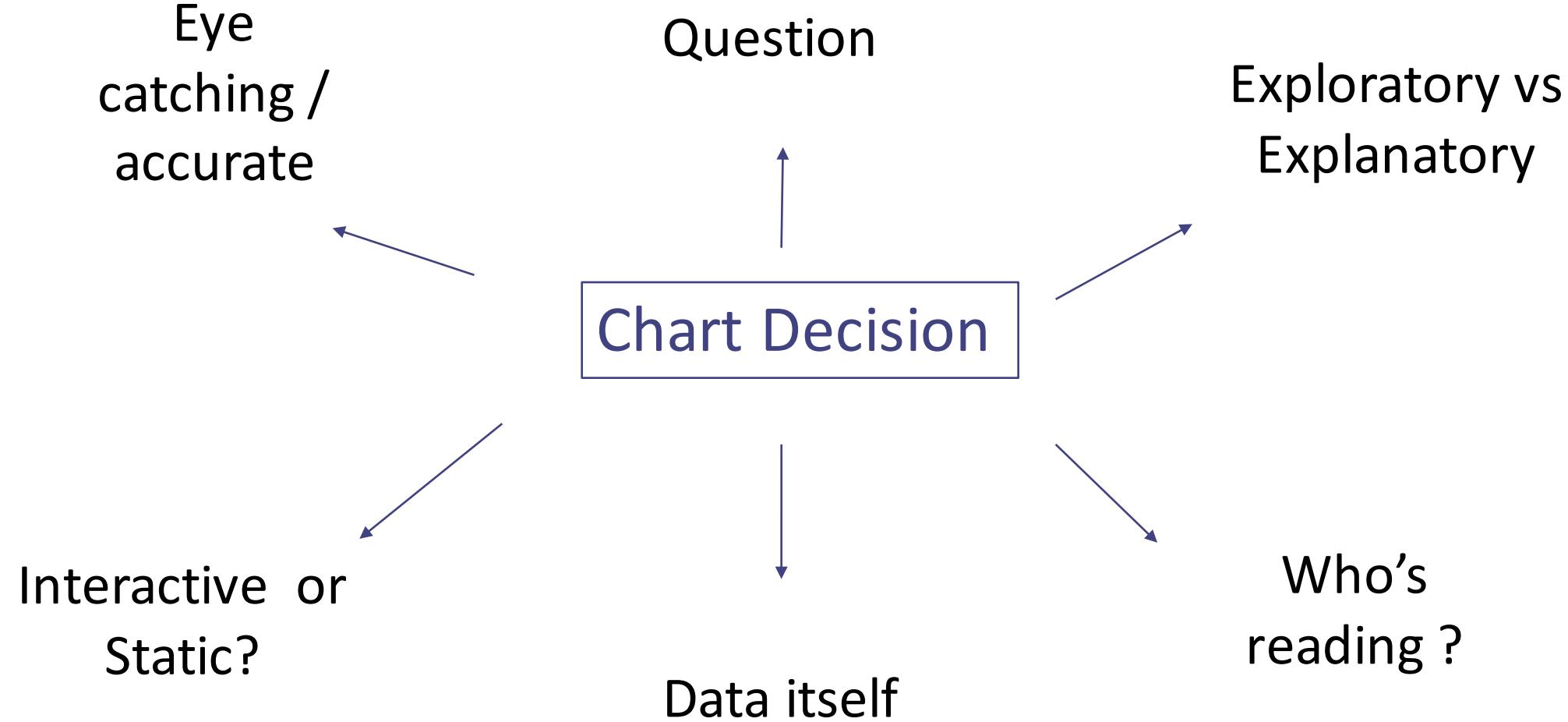
Country	Value
United States	12394
Russia	6148
Germany (FRG)	1653
France	2162
United Kingdom	1214
China	1131



Who sells more weapons ?

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

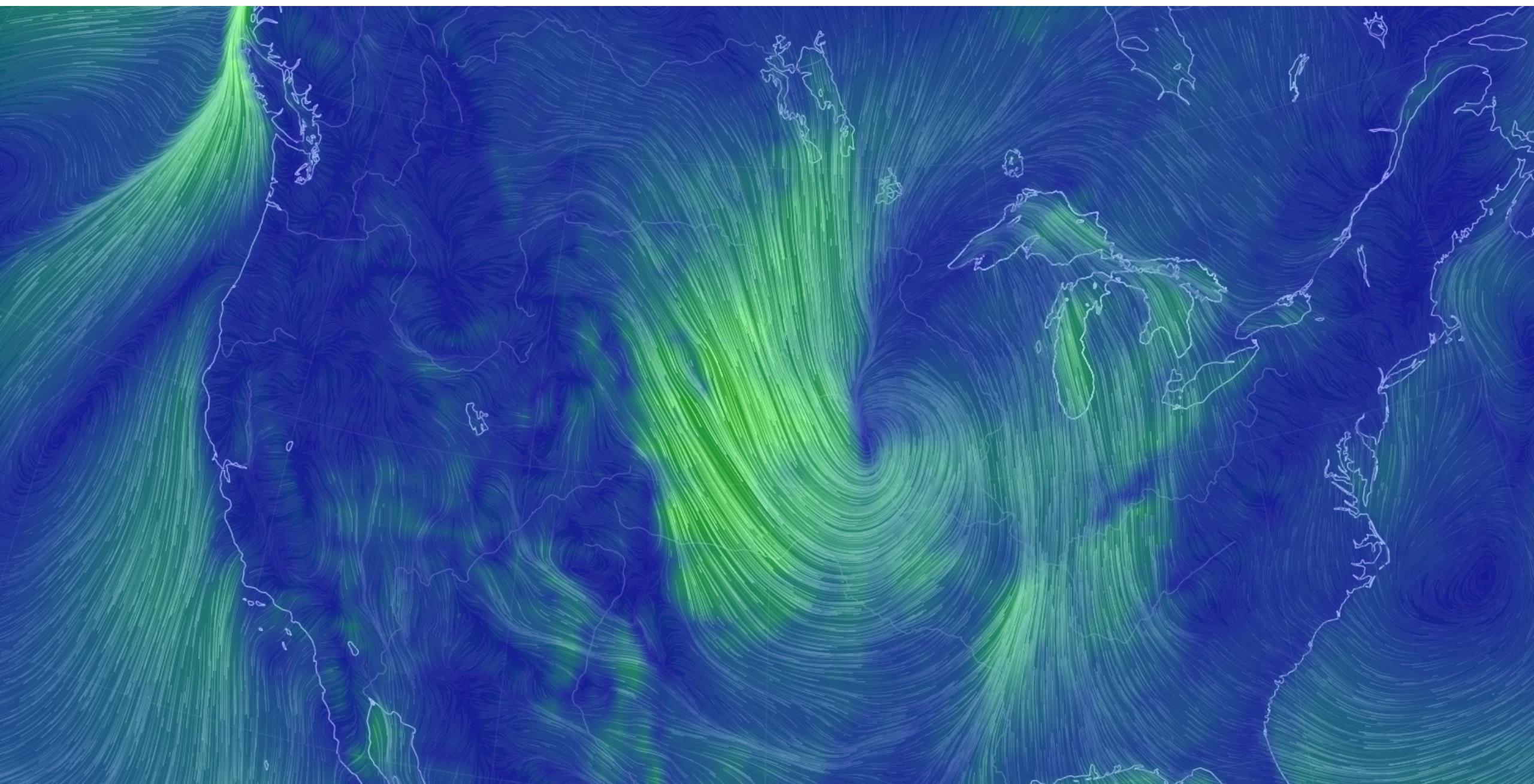


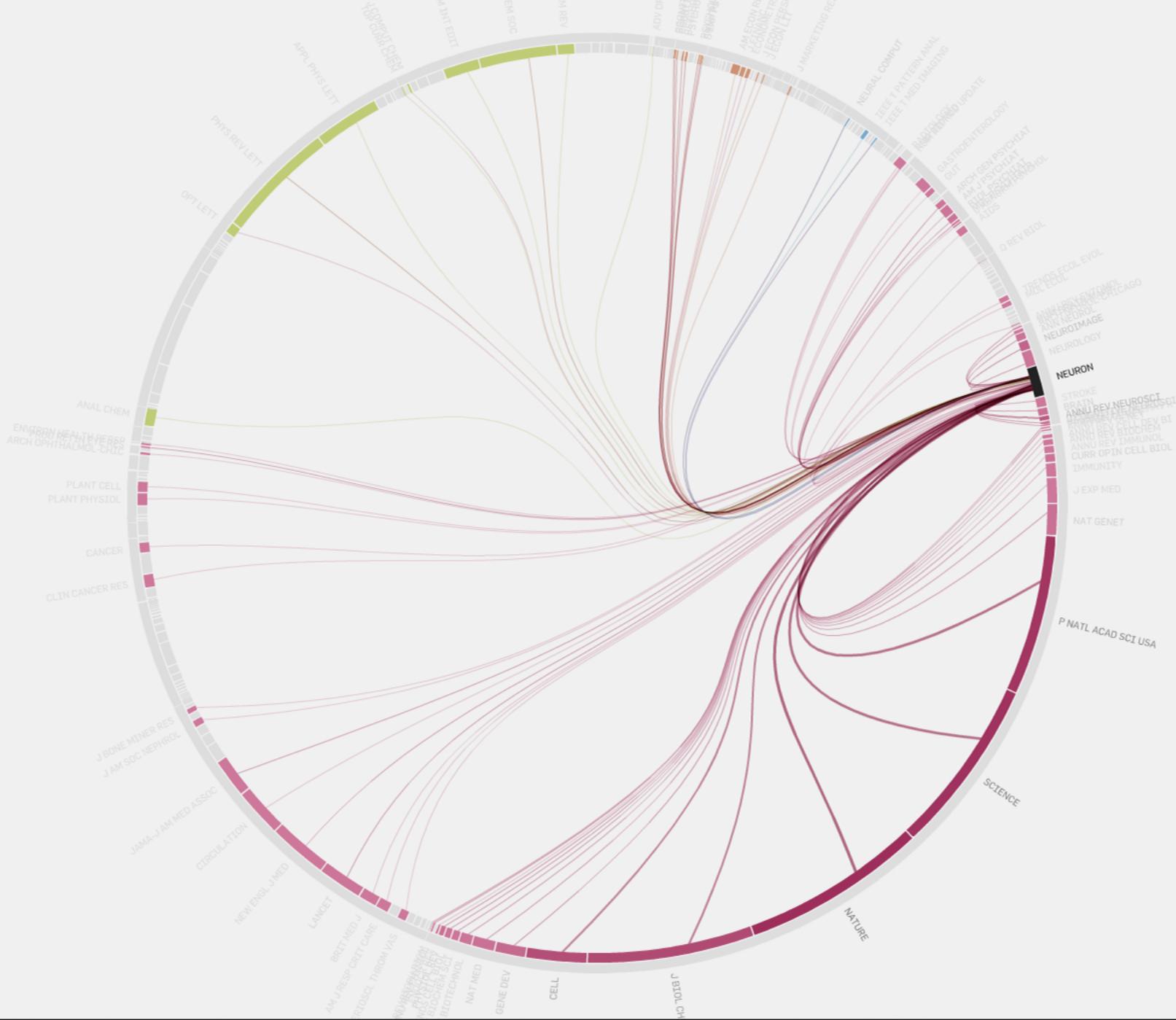




Where surfers travel.

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





Why

2/ Do you like Excel spreadsheet?

Can you understand them?

No, so you need to visualize it

Called the “picture superiority effect”

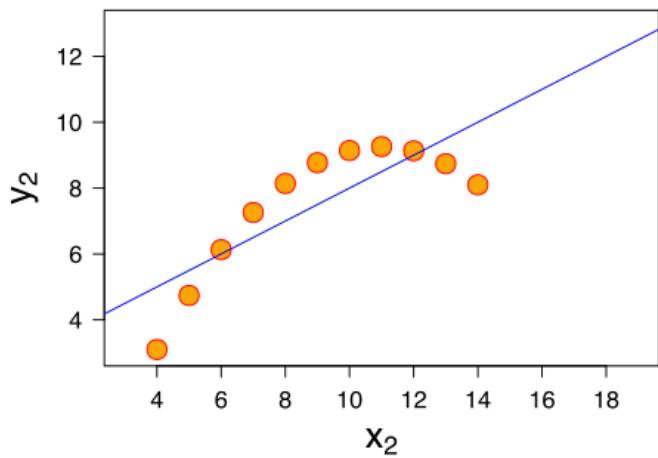
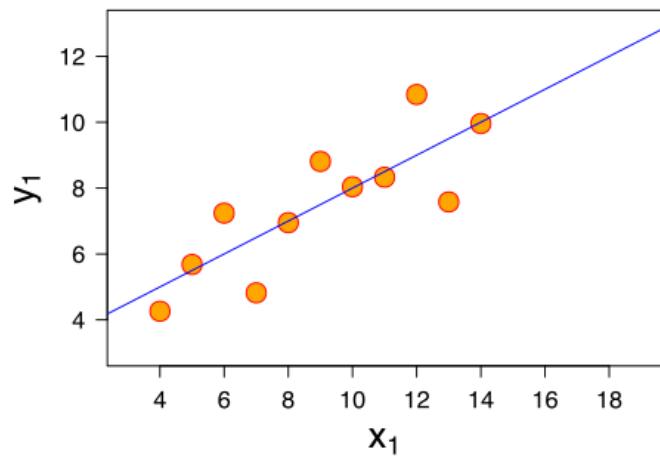
OK, but you have summary statistics!

No, sumstat are dangerous

$$Y = 3 + 0.5x$$
$$\text{Cor} = 0.8$$

$$\text{Mean}(x) = 9$$
$$\text{Var}(x) = 11$$

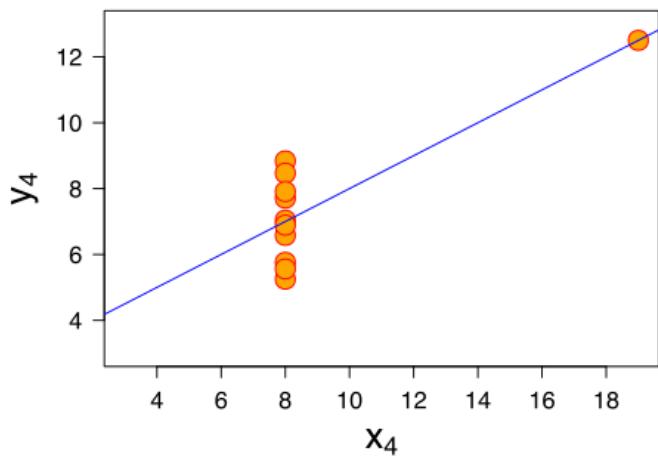
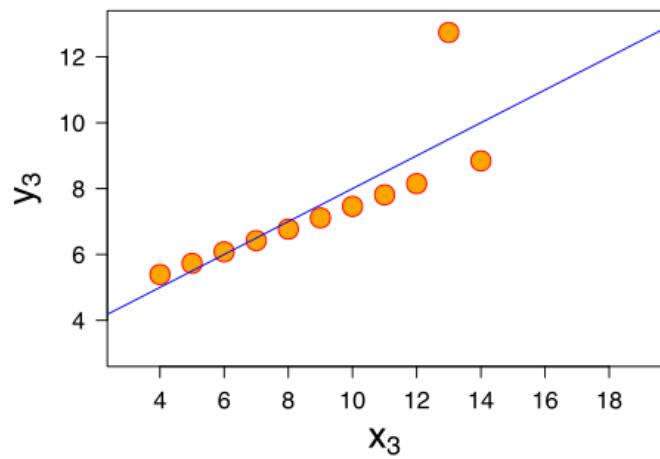
$$\text{Mean}(Y) = 7.5$$
$$\text{Var}(Y) = 4.1$$



$$Y = 3 + 0.5x$$
$$\text{Cor} = 0.8$$

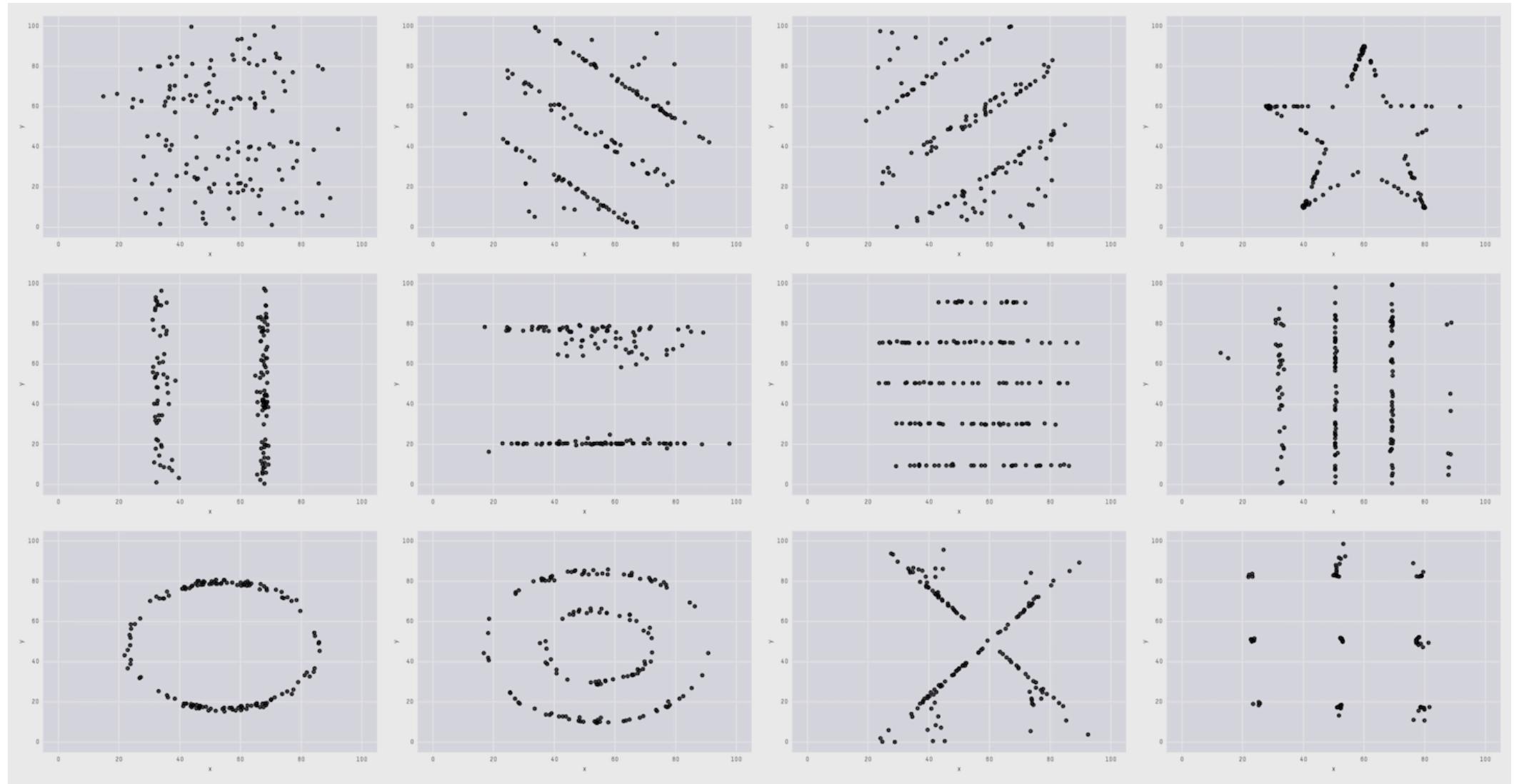
$$\text{Mean}(x) = 9$$
$$\text{Var}(x) = 11$$

$$\text{Mean}(Y) = 7.5$$
$$\text{Var}(Y) = 4.1$$



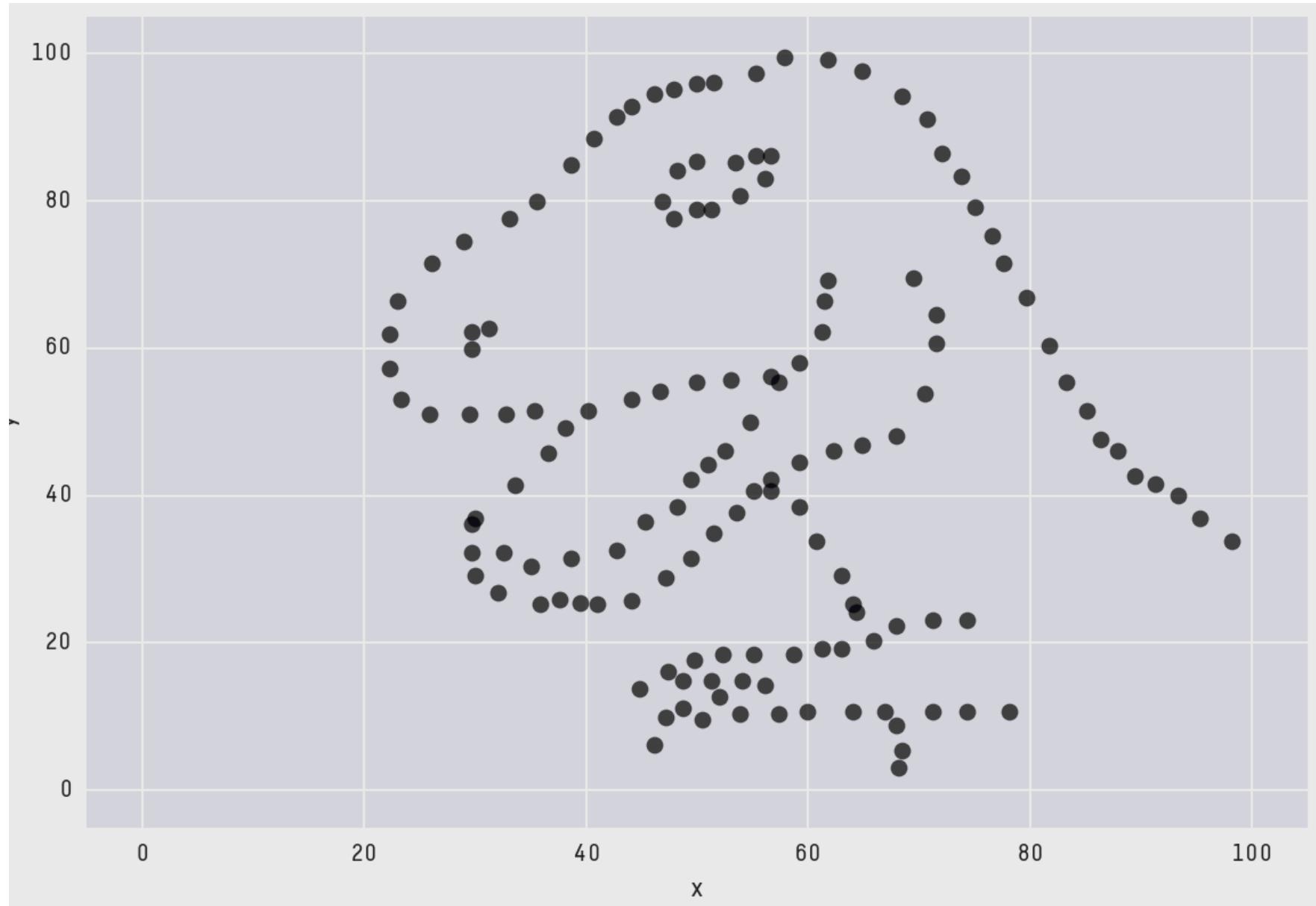
Plot your data

Anscombe's quartet



The Datasaurus Dozen

Cairo, 2017
Matejka & Fitzmaurice, 2017



The Datasaurus Dozen

Cairo, 2017

Matejka & Fitzmaurice, 2017

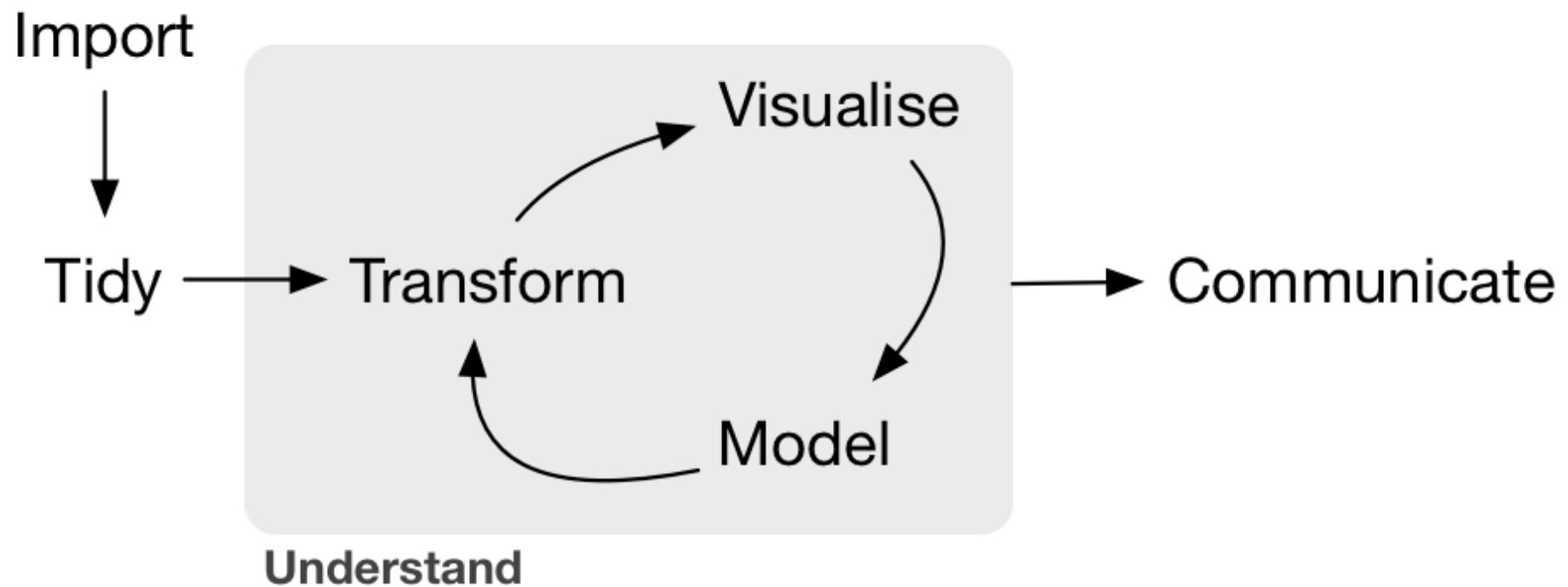
So dataviz = Understand data, make decision = EXPLO

But not only, also nice way to communicate!

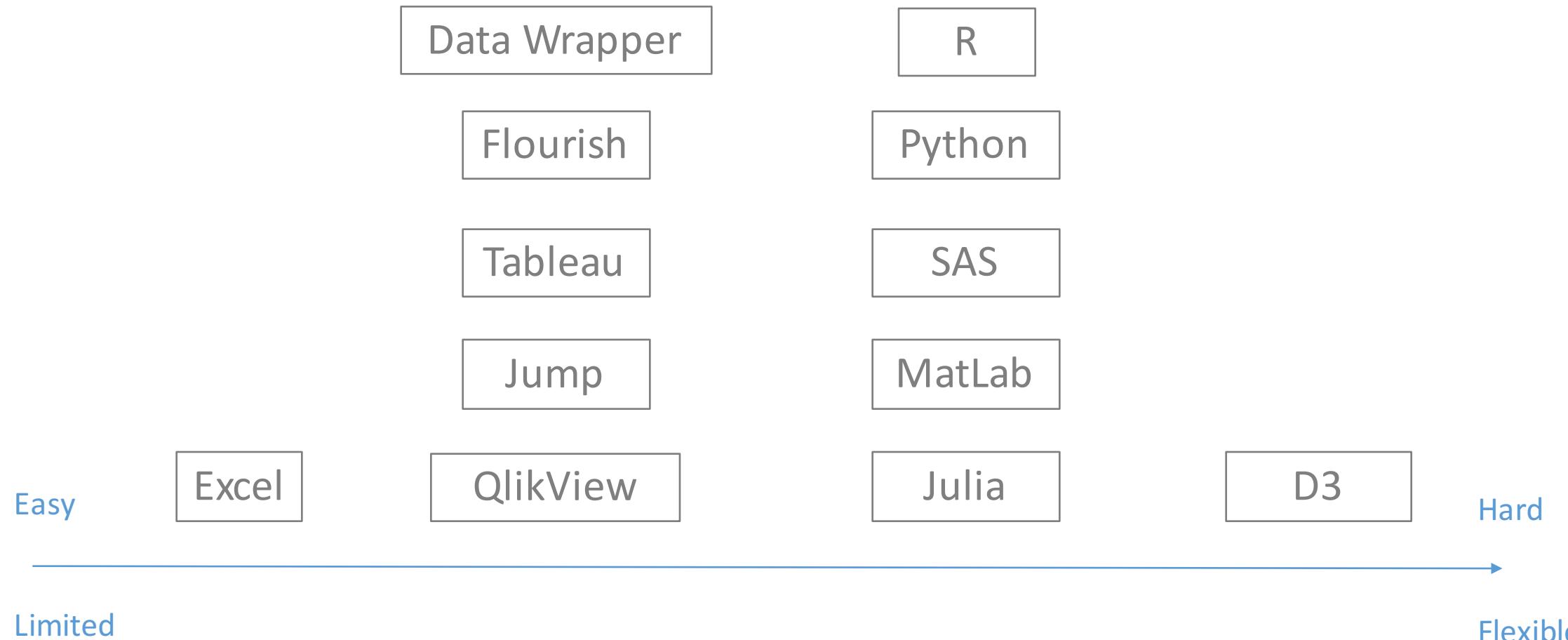
Communicate result = EXPLA

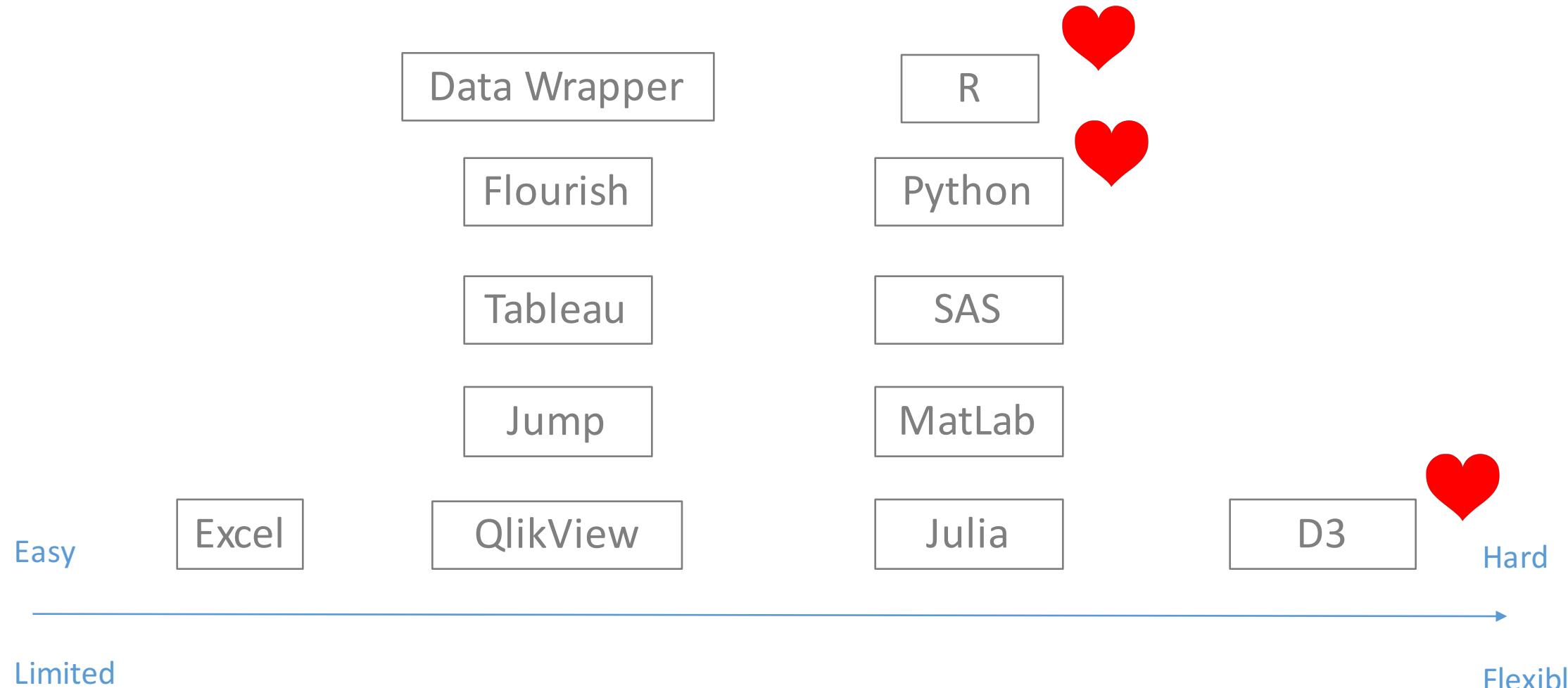
2/ Do you like Excel spreadsheet?

Can you understand them



How





Example:

Take a dataset

Explore it -> show explo graph

Make a visual one result is known

-> This is what you're gonna do during the rest of the course



from Data to Viz



R-graph-gallery.com

Python-graph-gallery.com

D3-graph-gallery.com



Thanks

Slides:

bit.ly/2SxJFIV

Cheat Sheet:

bit.ly/2AZySd9

Pimp my rmd:

bit.ly/2QDX7Hz



Yan.holtz.data@gmail.com



www.yan-holtz.com



@R_Graph_Gallery



github.com/holtzy