Introduction to data visualization

What, how, why, and more

Yan-holtz.com/teaching

What

Data Visualization (Dataviz)

=

Visual representation of data

Who sells more weapons ?

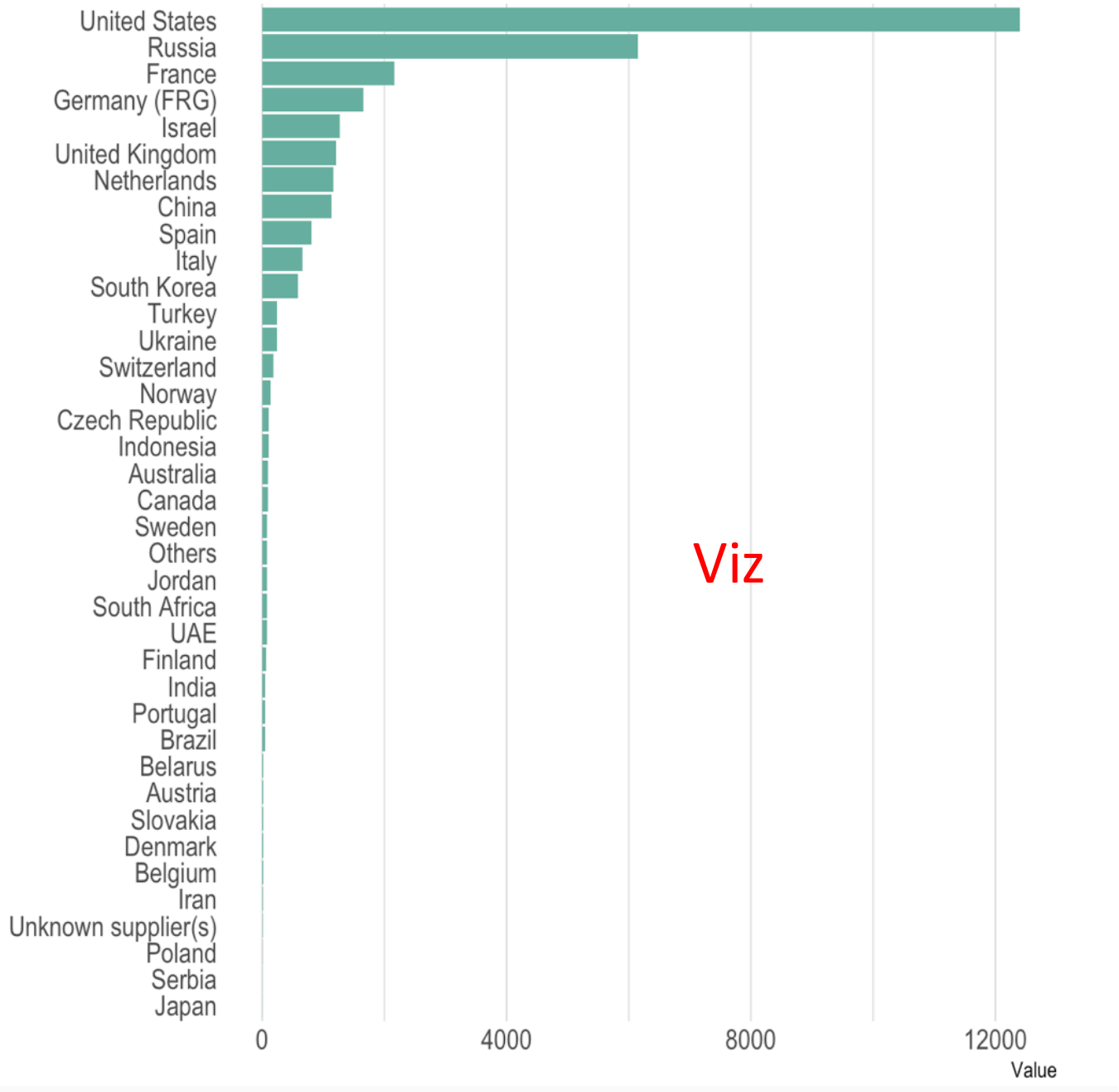
Country	Value
United States	12394
Russia	6148
Germany (FRG)	1653
France	2162
United Kingdom	1214
China	1131

Data

Who sells more weapons ?

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

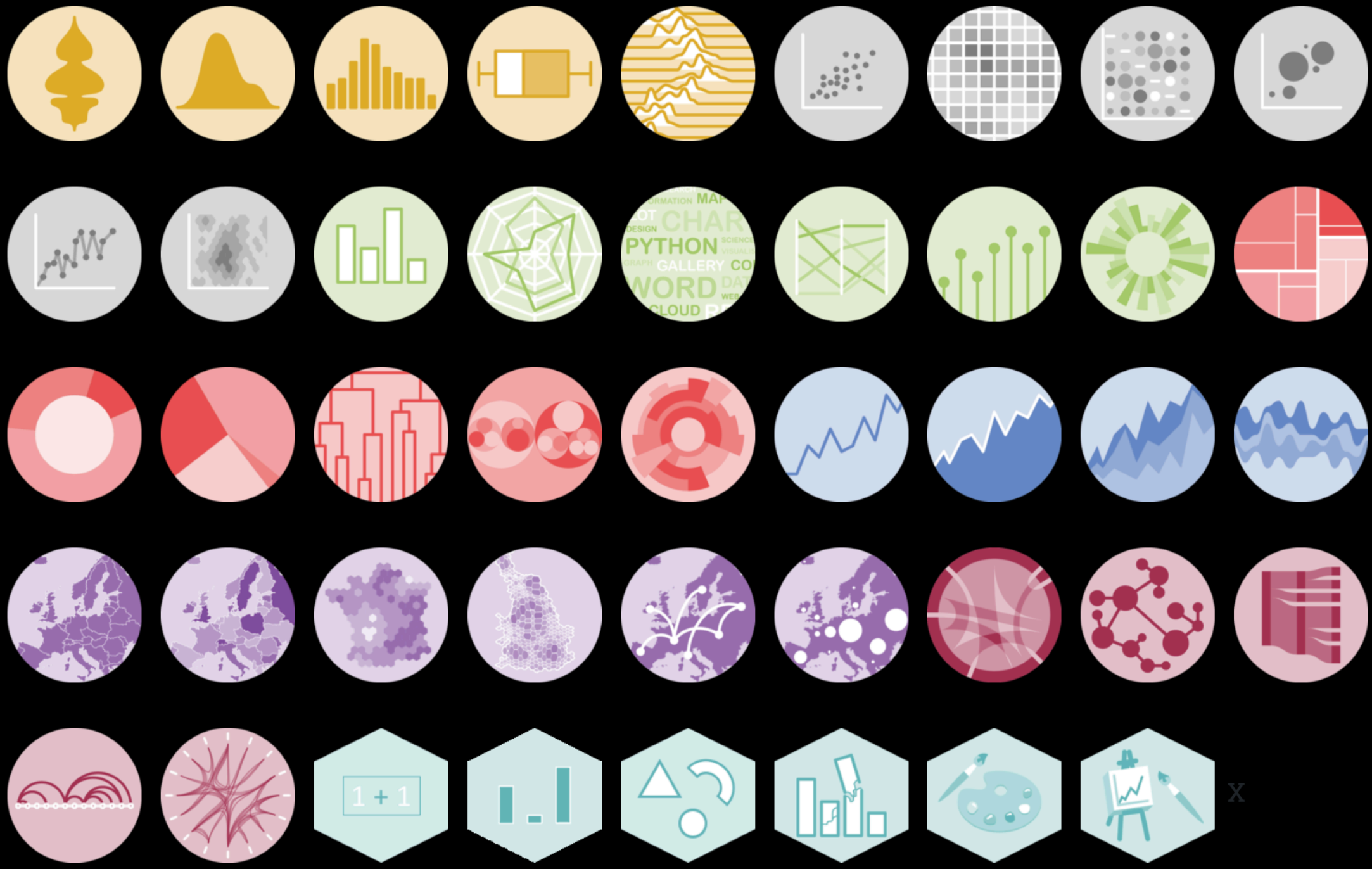
Data



Viz

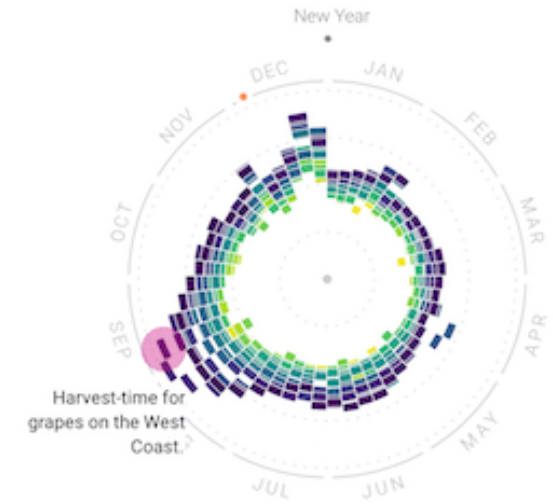
Dataviz = diversity

What kind of chart do you know?


PYTHON
GALLERY
WORD
CLOUD
1 + 1

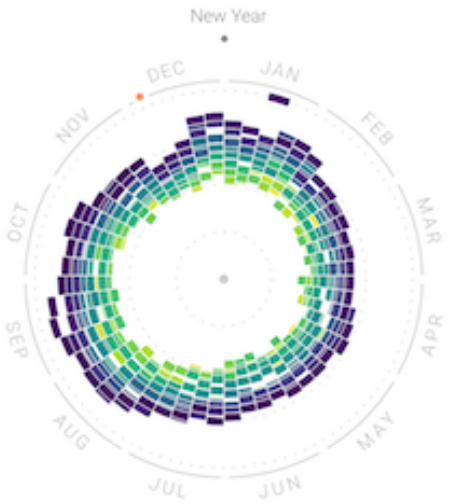
Dataviz > chart

Chart < Infographic < Interactivity < Dashboard < Story Telling

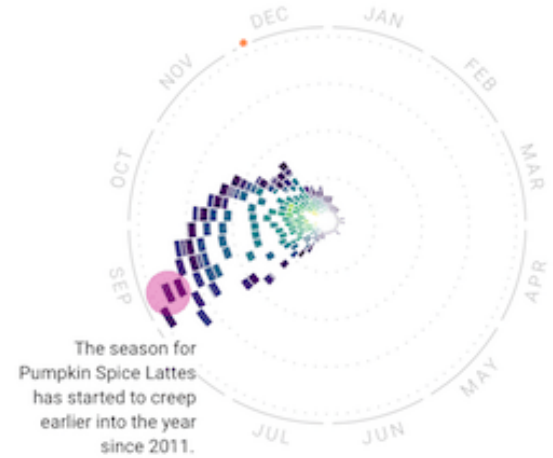


GRAPE

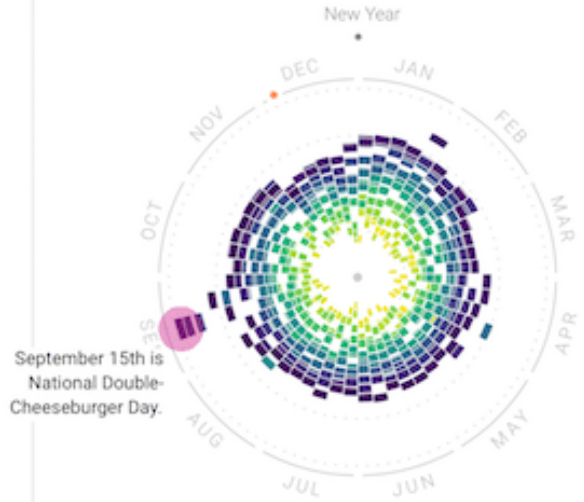
ELDERBERRY



CHILI PEPPER



PUMPKIN SPICE LATTE



CHEESEBURGER

CHANTERELLE

Flowing Data

What Could Disappear

Maps show coastal and low-lying areas that would be permanently flooded, without engineered protection, in three levels of higher seas. Percentages are the portion of dry, habitable land within the city limits of places listed that would be permanently submerged.

Today's waterways Land submerged by rising oceans

Select sea level rise over current level:

25 feet: Potential level in coming centuries, based on historical climate data.

12 feet: Potential level in about 2300 if nations make only moderate pollution cuts.

5 feet: Probable level in about 100 to 300 years.

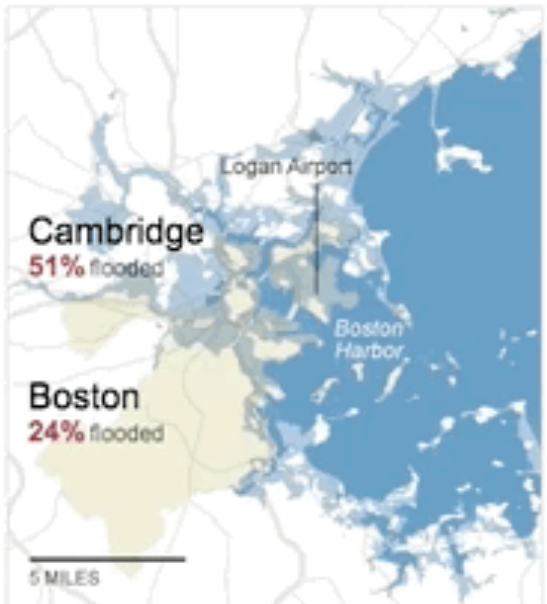
0 feet: Today's sea levels and land area.

Notes on sea level estimates

Baltimore 5% flooded

Fells Point and the city's ports are submerged.

Boston



Back Bay, the South End and the airport are permanently submerged. What's left of downtown is an island.

Charleston, S.C. 42% flooded

Much of the original city and its ports are flooded.

New York Times

Real-Time Active Users by Channel

Right now

10

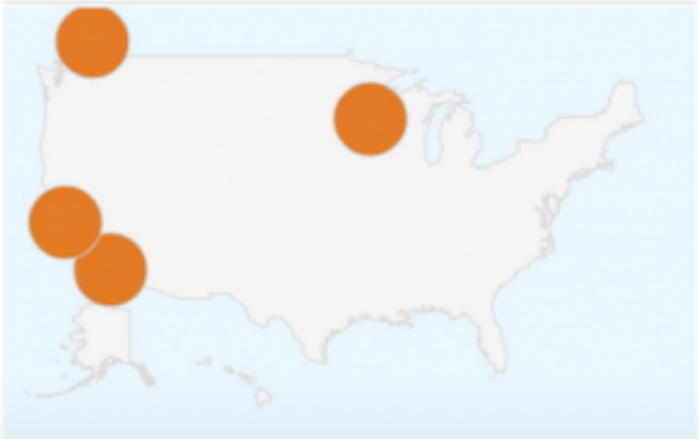
active users on site

ORGANIC DIRECT REFERRAL SOCIAL

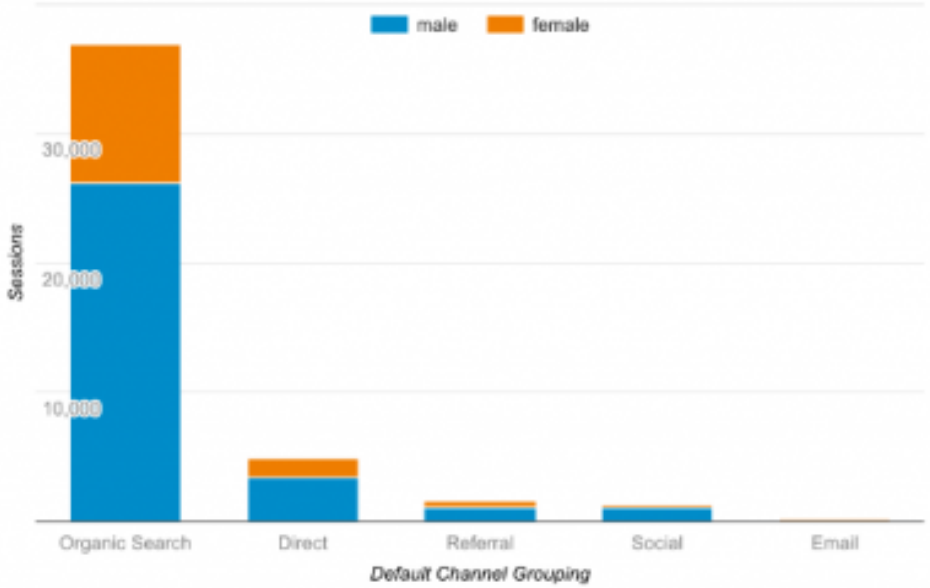
50% 30% 10% 10%

Real-Time Pageviews (Last 30 Min)

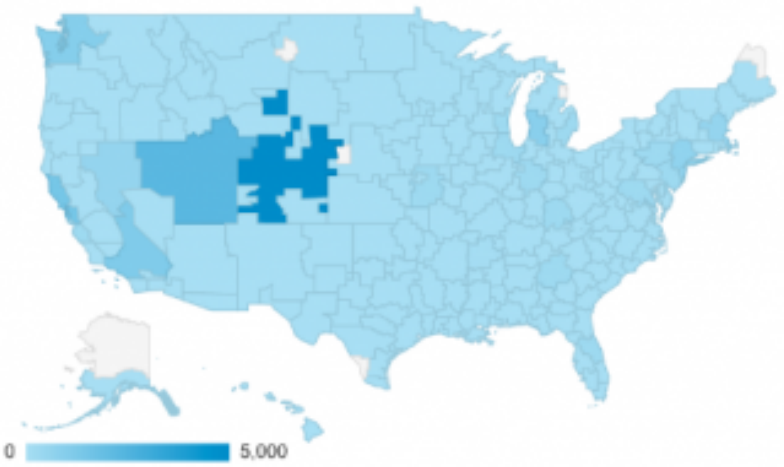
Real-Time Active Users (US Cities Only)



Sessions by Channel Pivoted by Gender (Stacked)



Revenue by Metro (US Only)



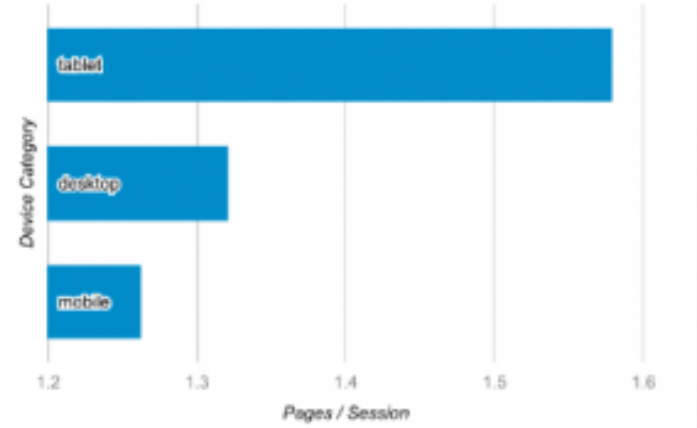
Revenue

$27,570.00

% of Total: 100.00% ($27,570.00)

Sessions by Social Network

Pages / Session by Device Category



Story Telling with data: The Pudding

https://pudding.cool/2018/08/wiki-death/

Why

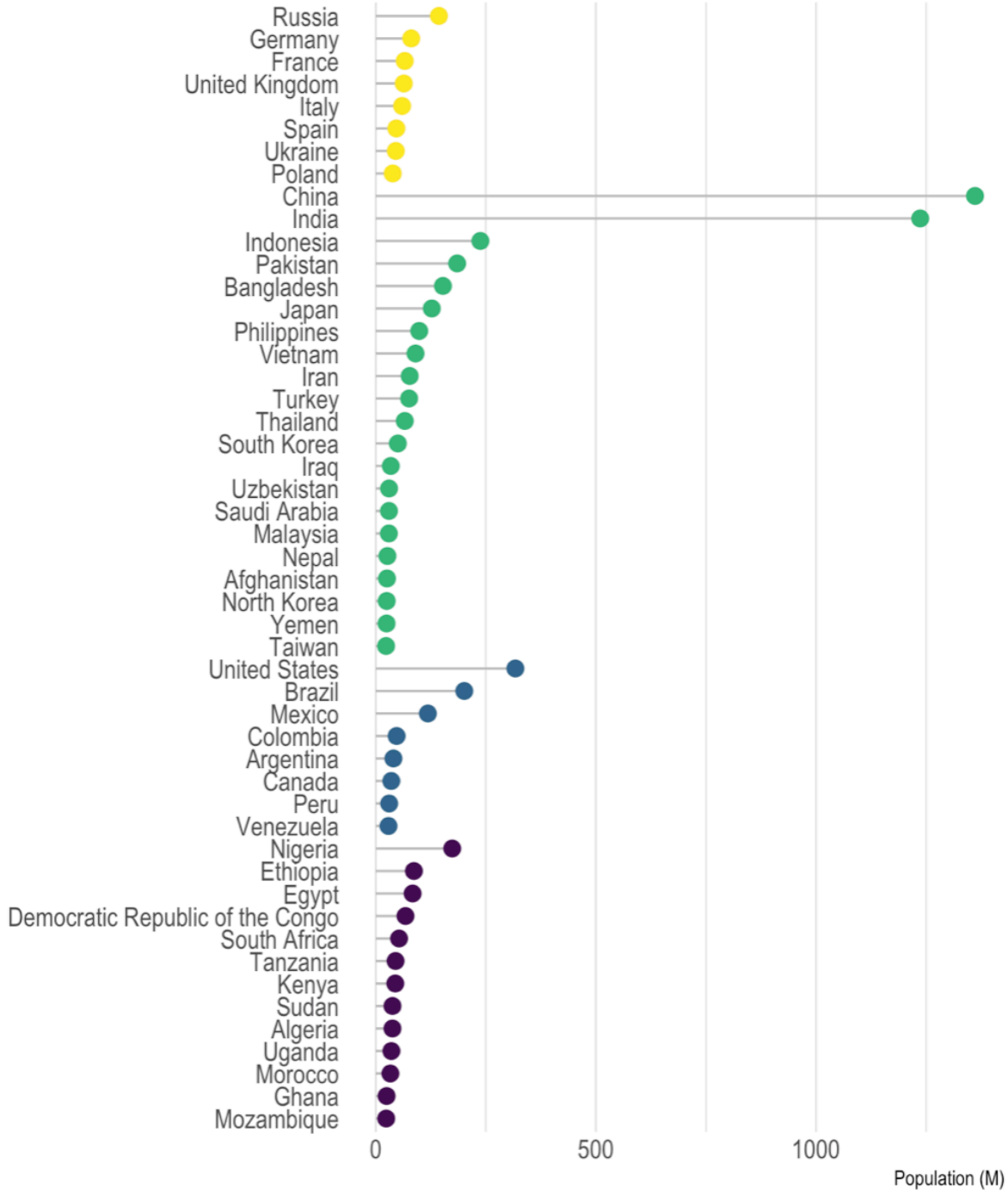
« A picture is worth a thousand words »

World population distribution

	A	B	C	D
1	region	subregion	key	value
2	Asia	Southern Asia	Afghanistan	25500100
3	Europe	Northern Europe	Ã…land Islands	28502
4	Europe	Southern Europe	Albania	2821977
5	Africa	Northern Africa	Algeria	37900000
6	Oceania	Polynesia	American Samoa	55519
7	Europe	Southern Europe	Andorra	76246
8	Africa	Middle Africa	Angola	20609294
9	Americas	Caribbean	Anguilla	13452
10			Antarctica	-1
11	Americas	Caribbean	Antigua and Barbuda	86295
12	Americas	South America	Argentina	40117096
13	Asia	Western Asia	Armenia	3024100
14	Americas	Caribbean	Aruba	101484
15	Oceania	Australia and New Zealand	Australia	23254142
16	Europe	Western Europe	Austria	8501502
17	Asia	Western Asia	Azerbaijan	9235100
18	Americas	Caribbean	Bahamas	-1
19	Asia	Western Asia	Bahrain	1234571
20	Asia	Southern Asia	Bangladesh	152518015
21	Americas	Caribbean	Barbados	274200
22	Europe	Eastern Europe	Belarus	9465500
23	Europe	Western Europe	Belgium	11175653
24	Americas	Central America	Belize	312971
25	Africa	Western Africa	Benin	10323000
26	Americas	Northern America	Bermuda	64237
27	Asia	Southern Asia	Bhutan	740990
28	Americas	South America	Bolivia	10027254
29	Americas	Caribbean	Bonaire	-1
30	Europe	Southern Europe	Bosnia and Herzegovina	3791622
31	Africa	Southern Africa	Botswana	2024904
32			Bouvet Island	-1
33	Americas	South America	Brazil	201032714
34	Africa	Eastern Africa	British Indian Ocean Territory	-1
35	Americas	Caribbean	British Virgin Islands	29537

- Biggest country?
- Rank?
- Distribution by continent?

“Picture superiority effect”



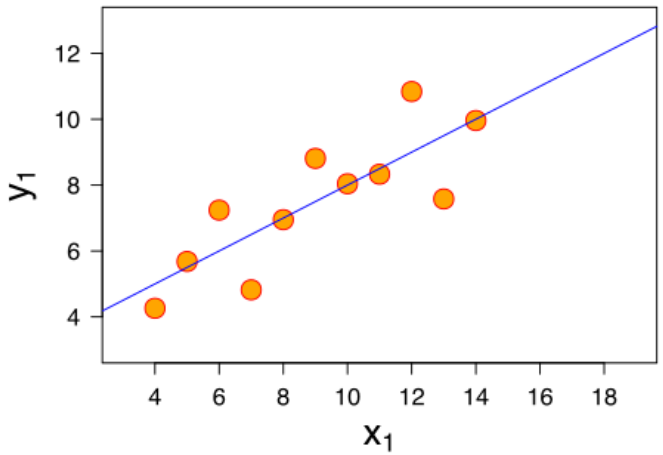
OK, but we have summary statistics!

Y = 3 + 0.5x
Cor = 0.8

Mean(x) = 9
Var(x) = 11

Mean(Y) = 7.5
Var(Y) = 4.1

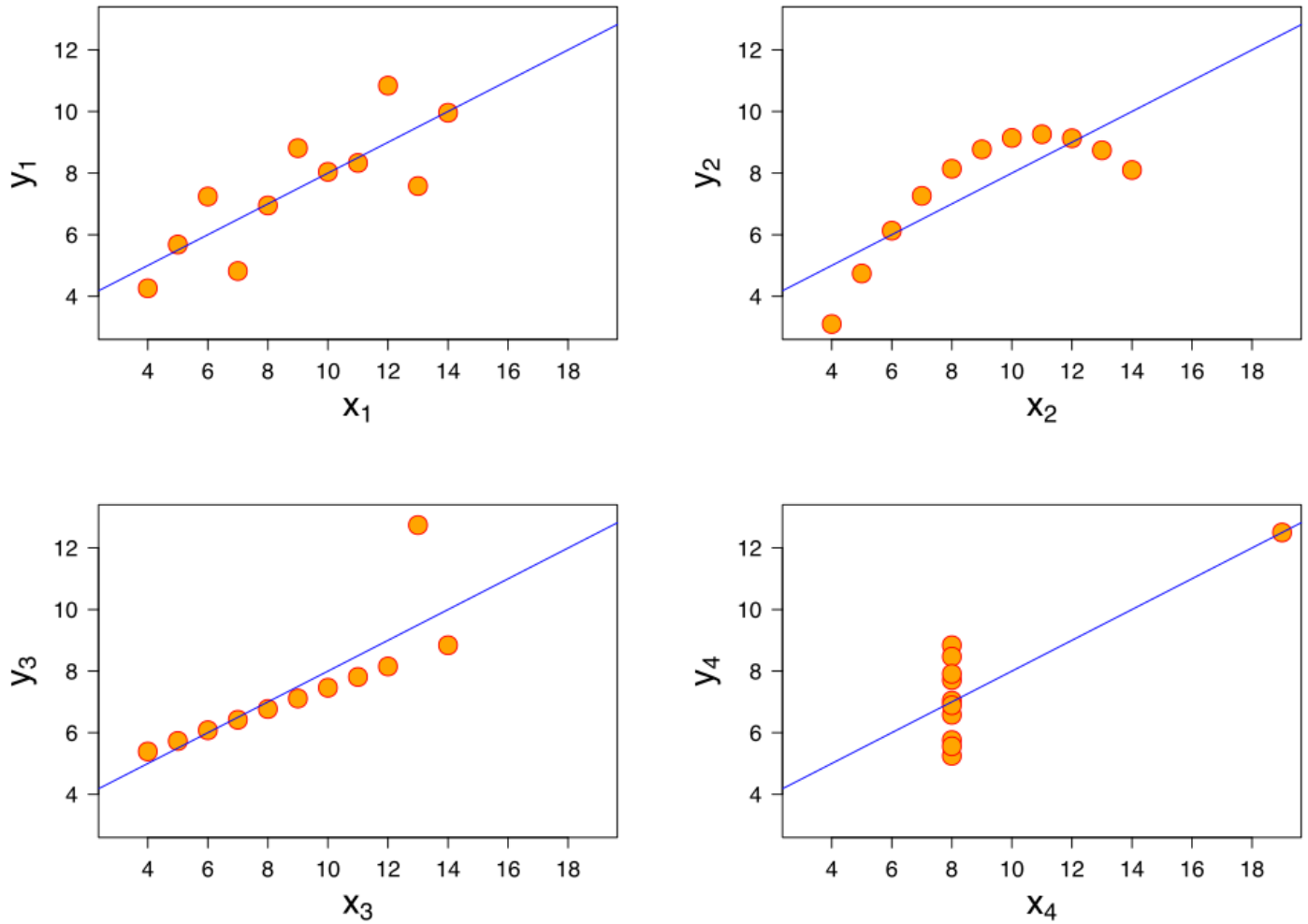
Anscombe et al. 1973



Y = 3 + 0.5x
Cor = 0.8

Mean(x) = 9
Var(x) = 11

Mean(Y) = 7.5
Var(Y) = 4.1



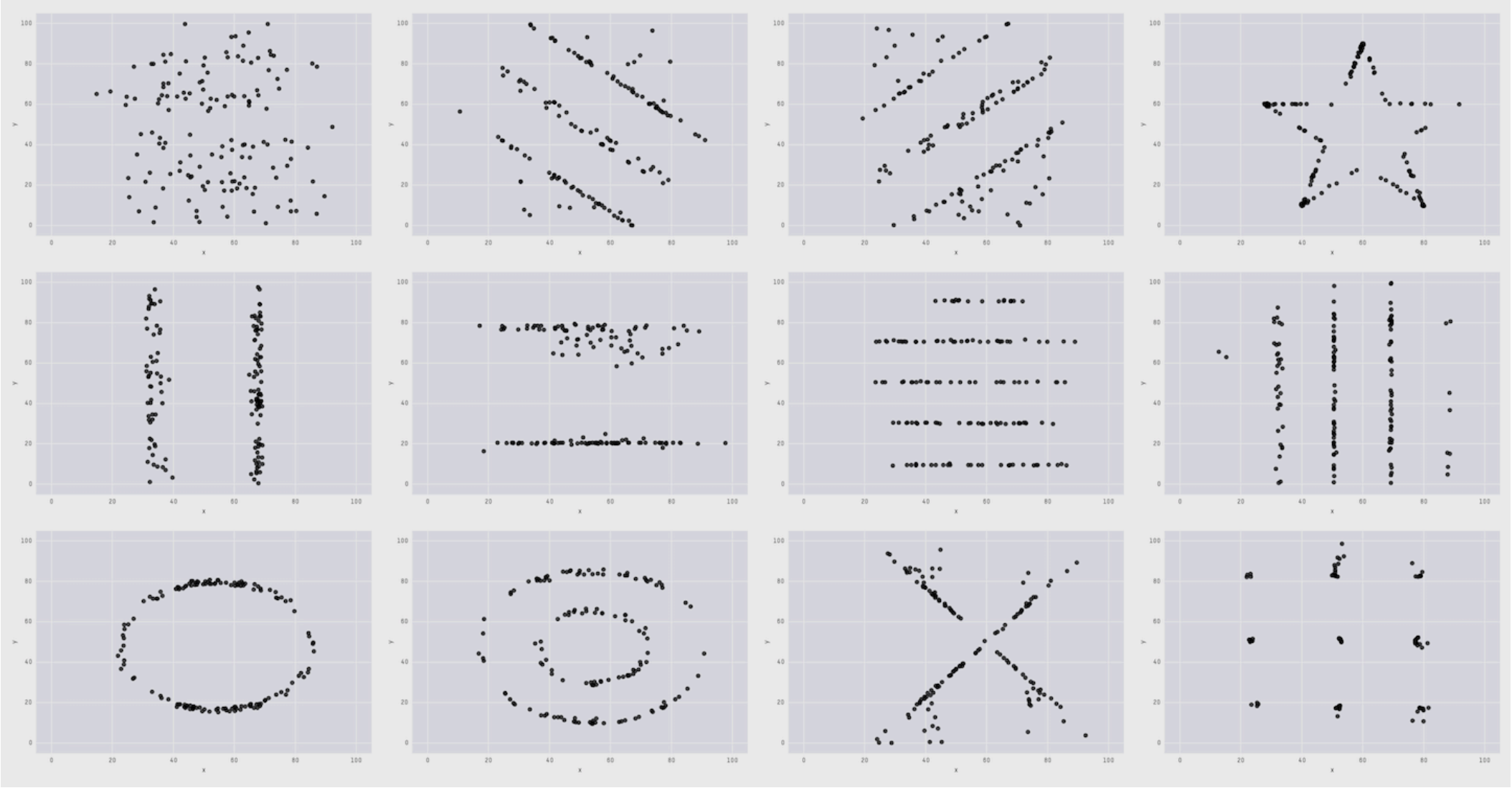
Y = 3 + 0.5x
Cor = 0.8

Mean(x) = 9
Var(x) = 11

Mean(Y) = 7.5
Var(Y) = 4.1

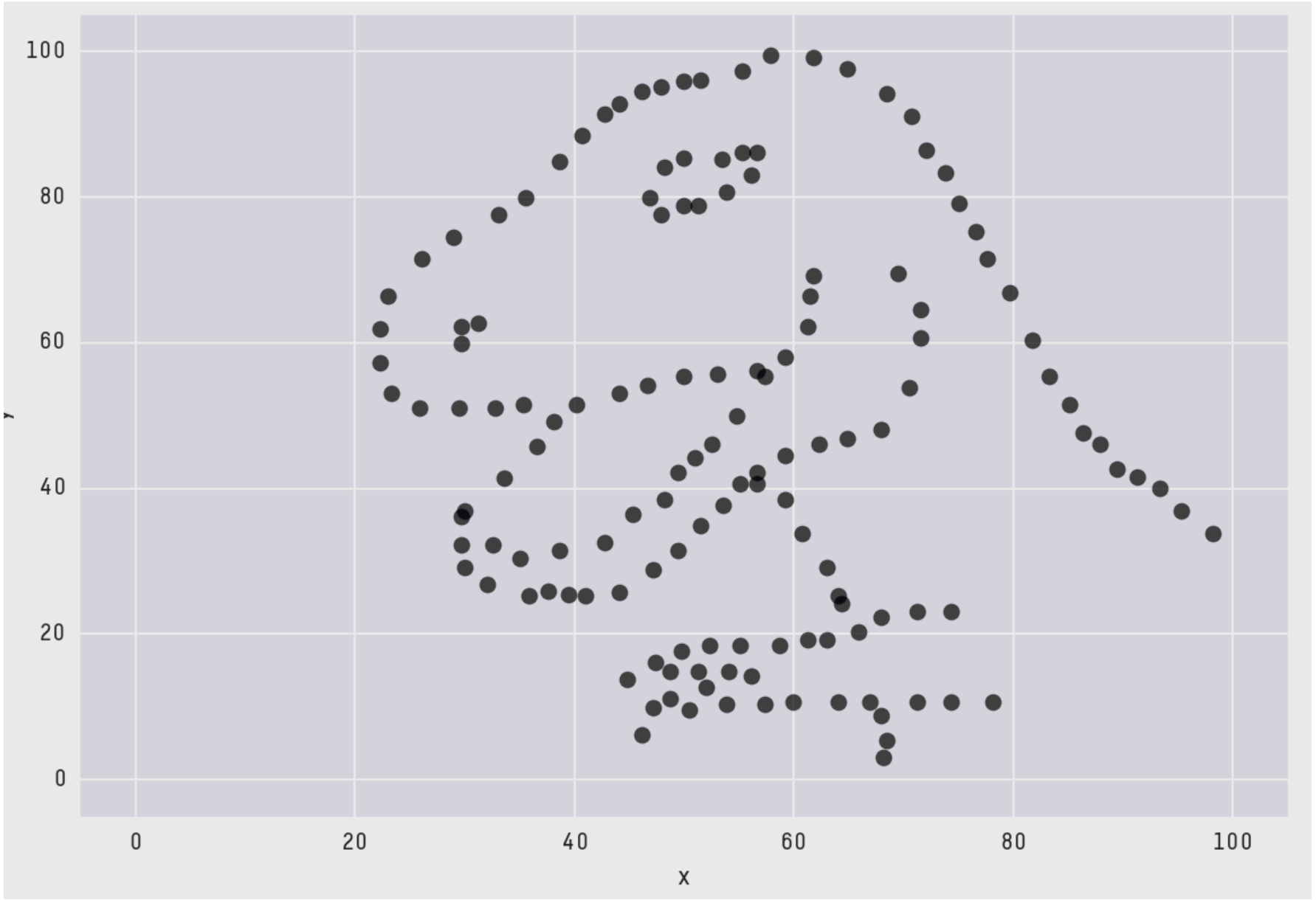
Anscombe's quartet

Anscombe et al. 1973



The Datasaurus Dozen

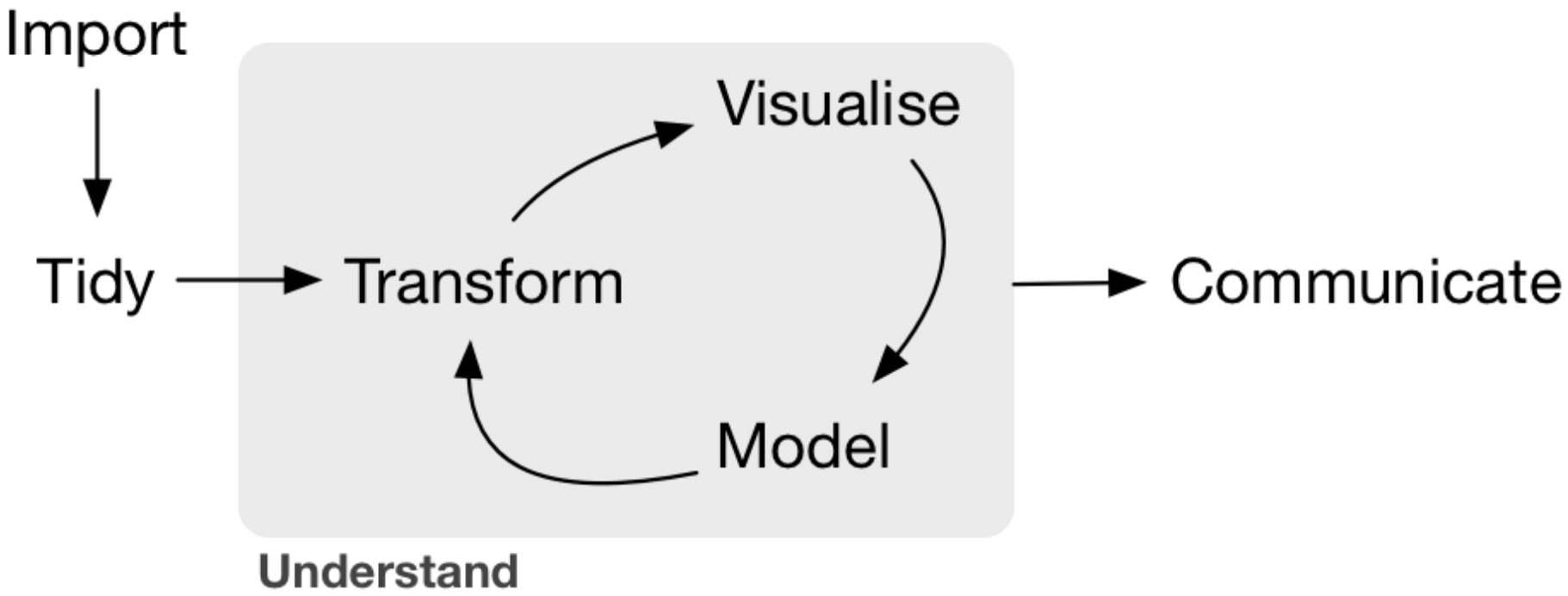
Cairo, 2017
Matejka & Fitzmaurice, 2017



The Datasaurus Dozen

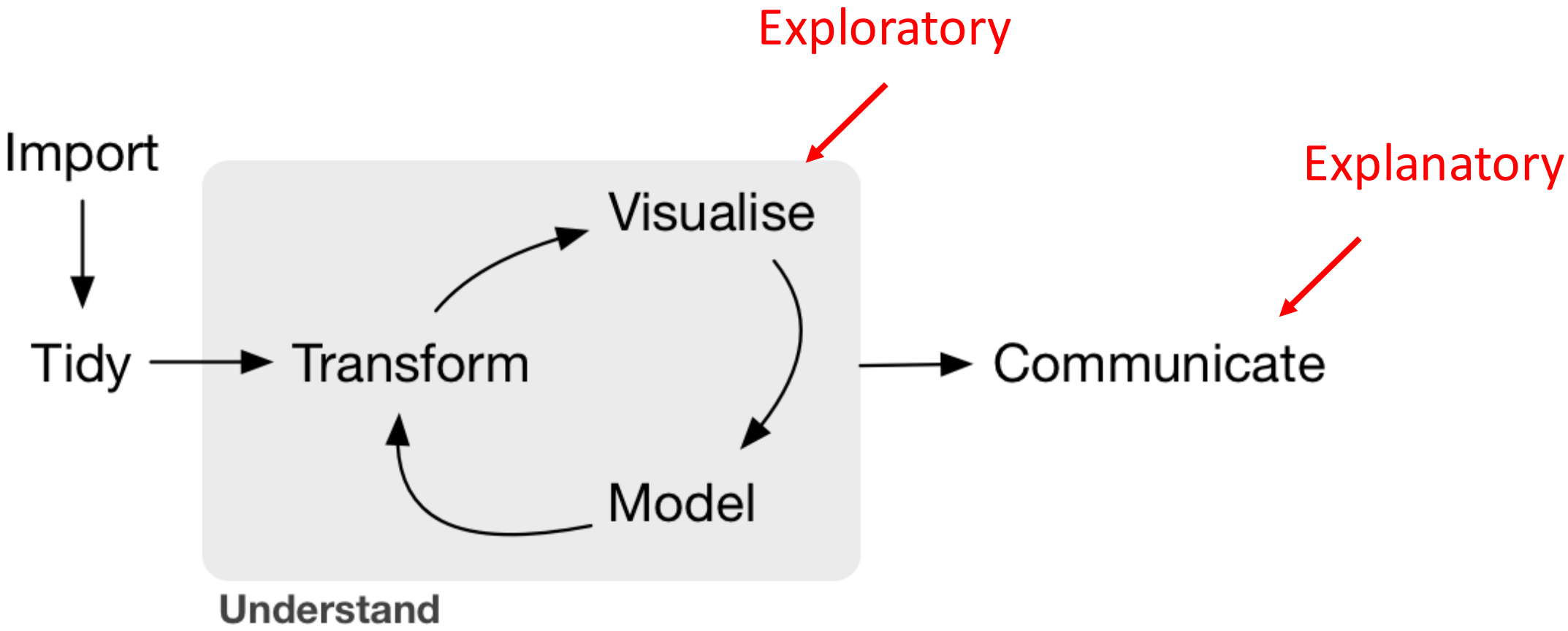
Cairo, 2017
Matejka & Fitzmaurice, 2017

The Data Science Pipeline



R for Data Science, Whickam et al.

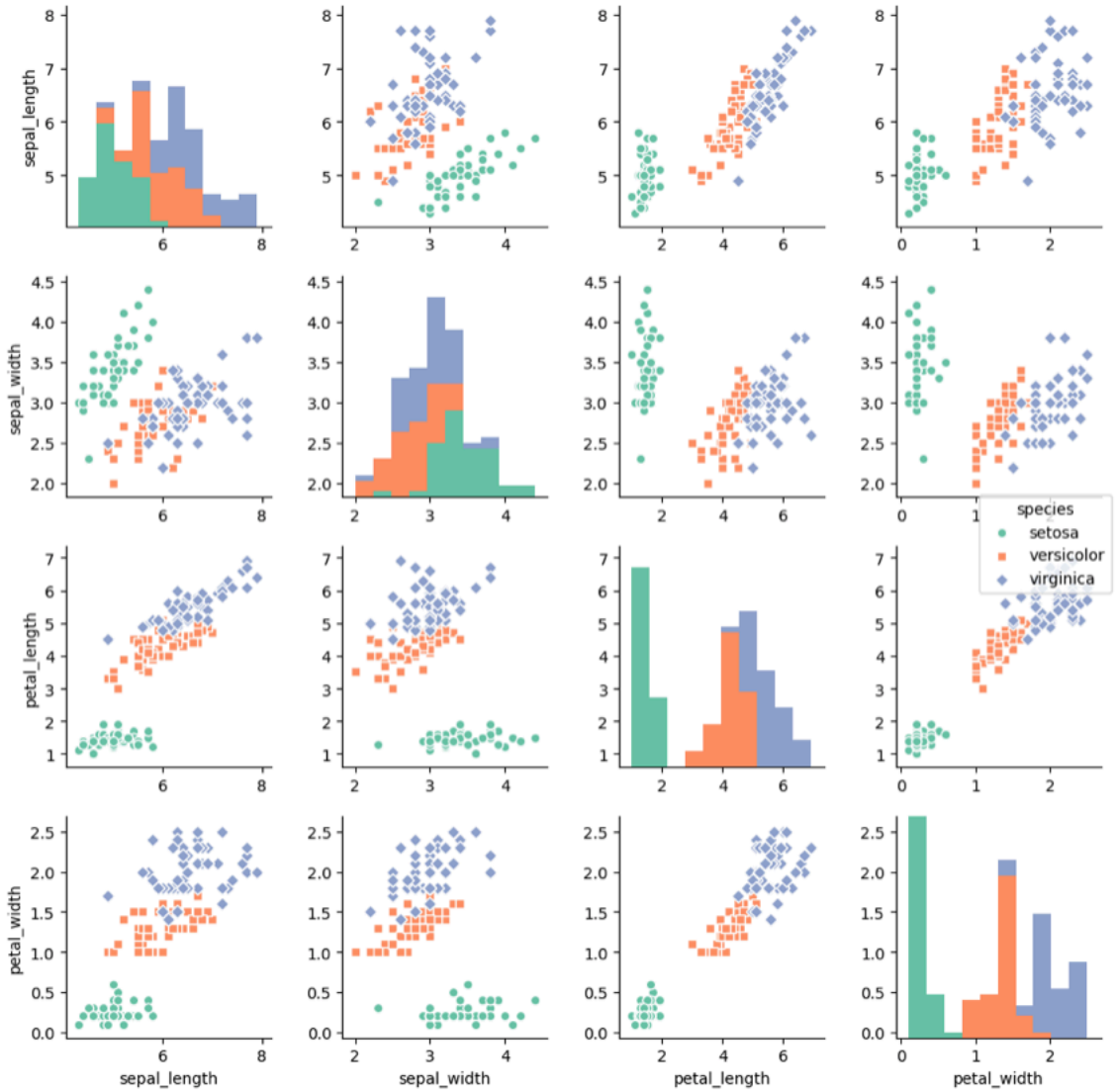
2 main types of dataviz:



R for Data Science, Whickam et al.

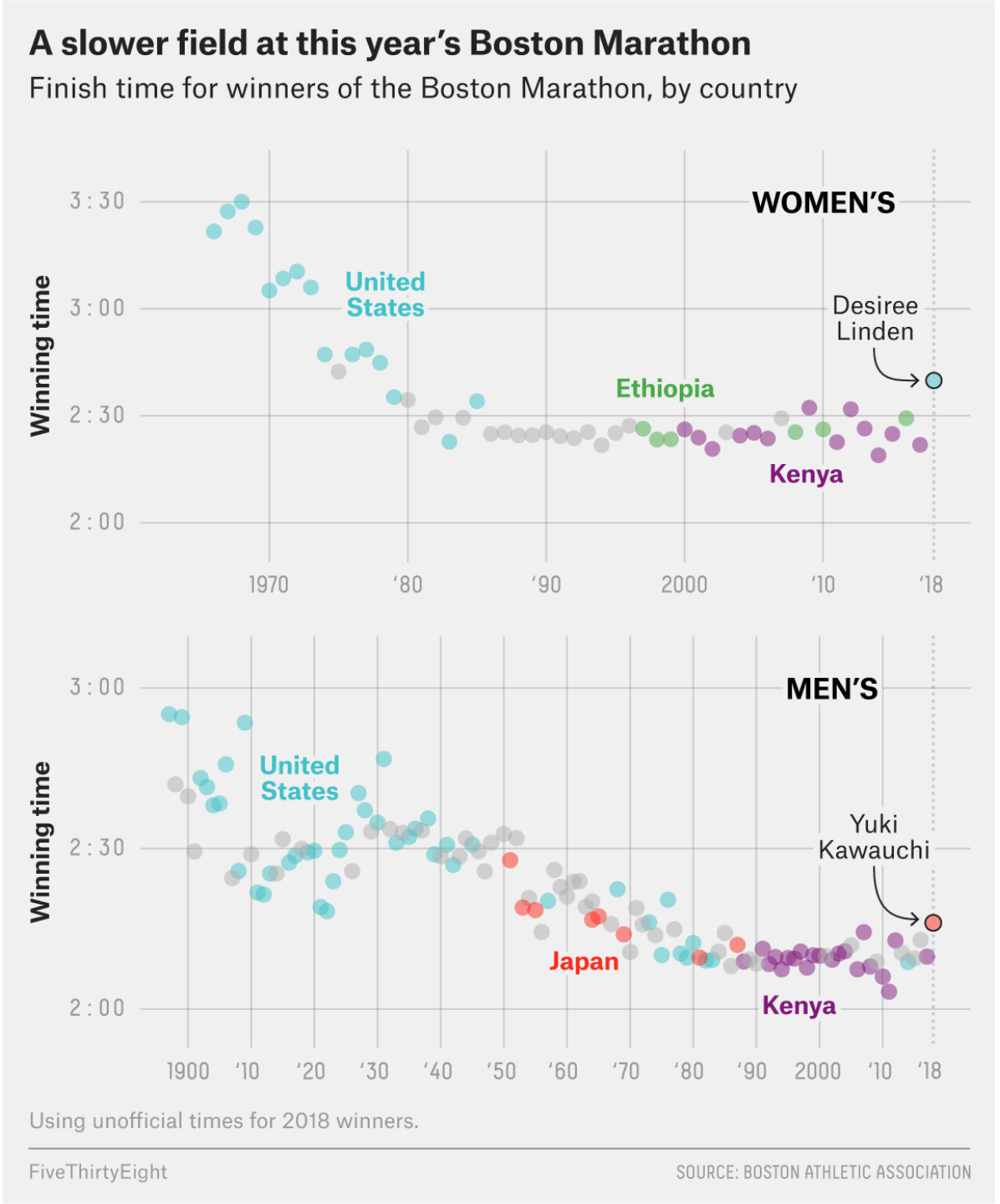
Exploratory

- Understand data
- Test hypothesis
- Find results



Explanatory

- Communicate result
- Show something specific
- Tells a story



Five thirty height

Building the right Viz

Looking for a chart ?

What you **can** do

What you **should** do

Caveats to avoid

How to build it

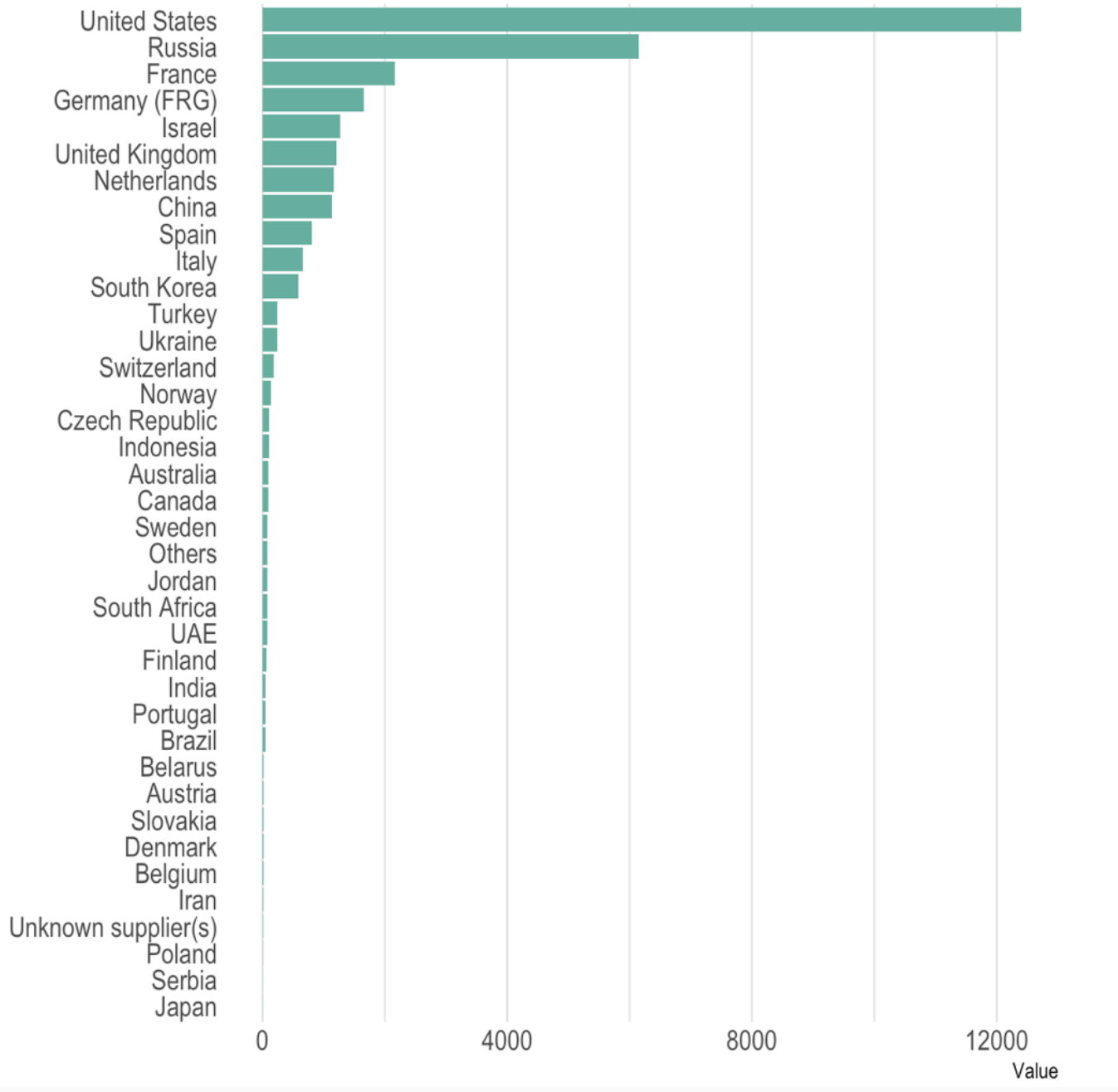
Who sells more weapons ?

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

?

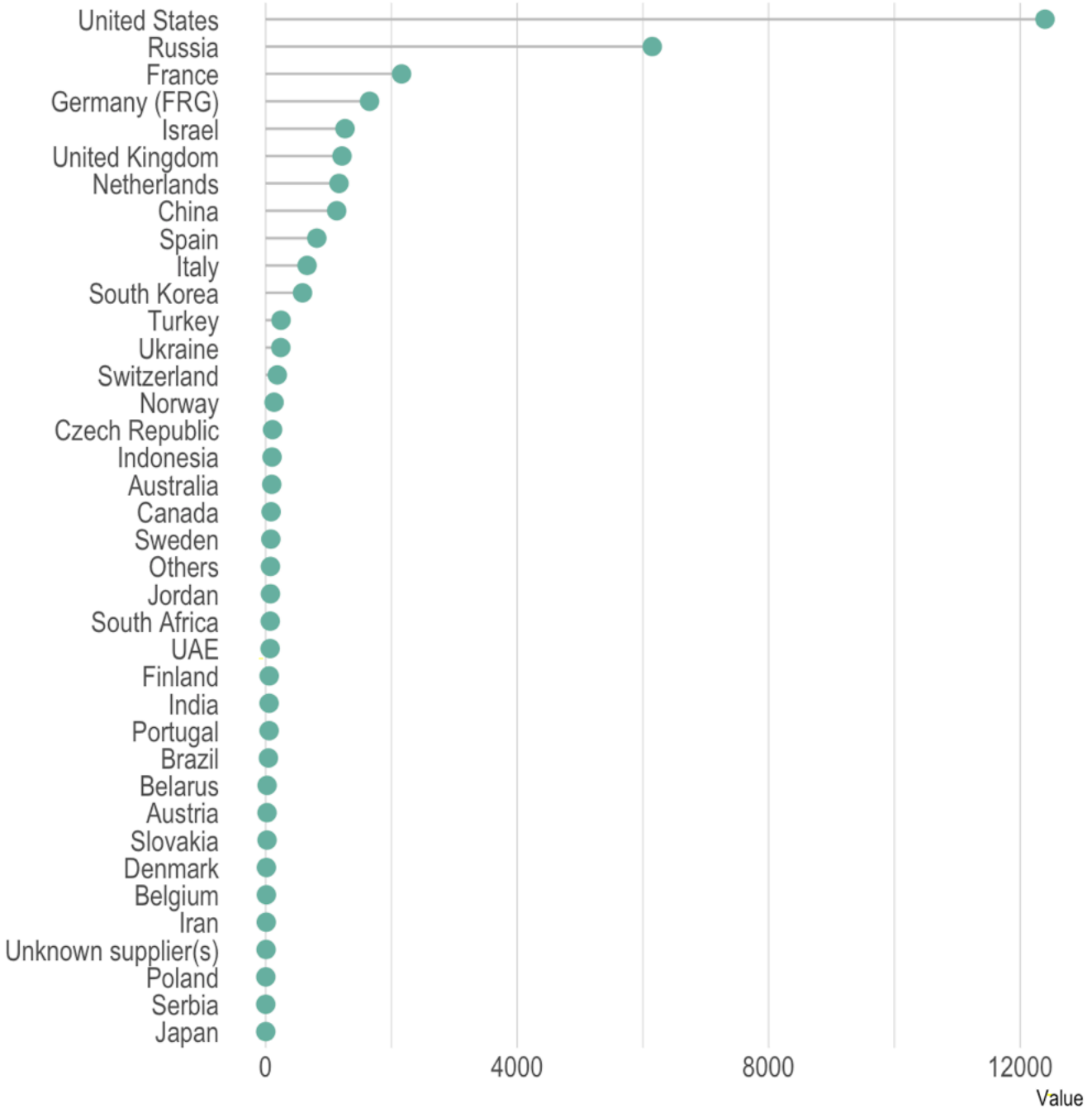
Who sells more weapons ?

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



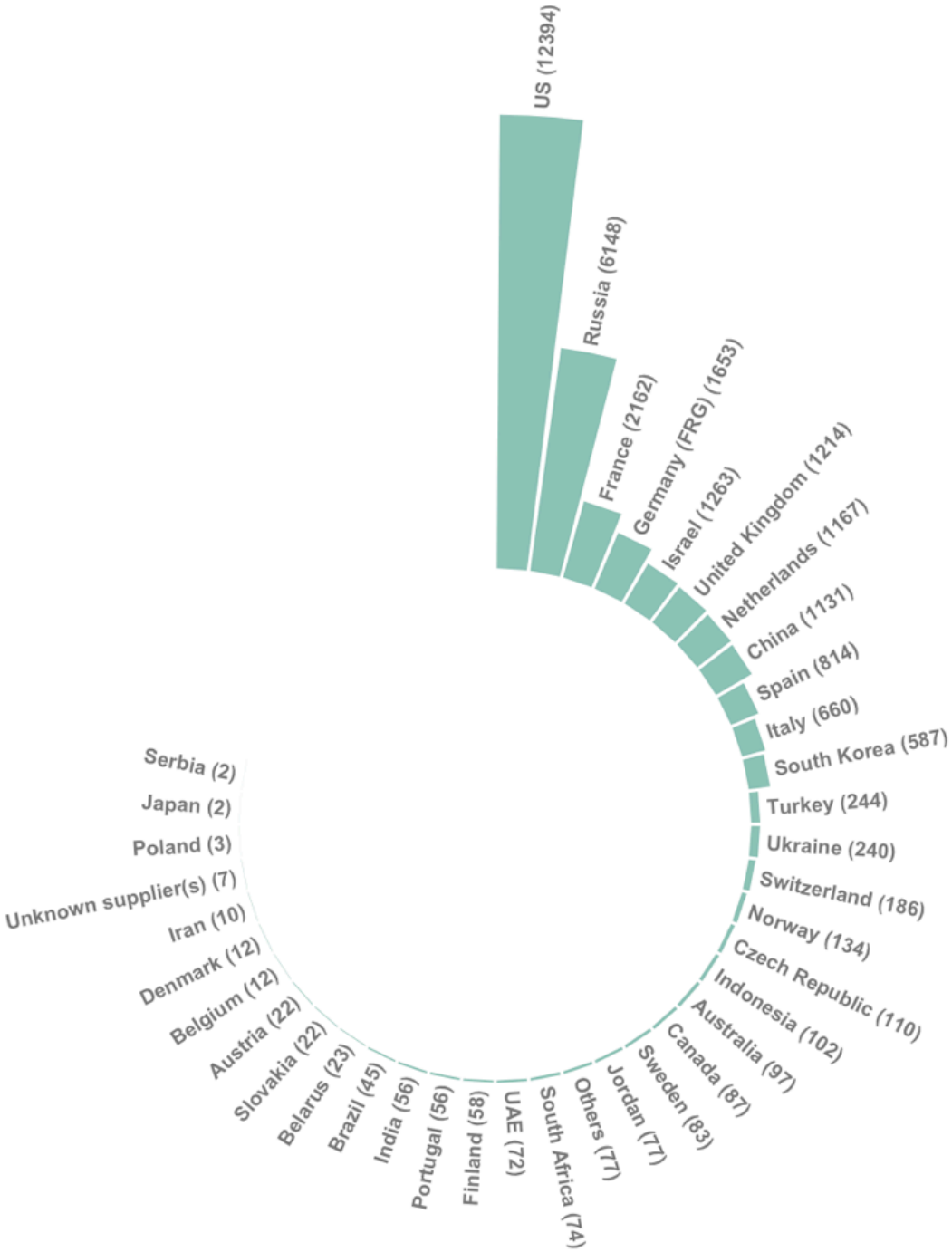
Who sells more weapons ?

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



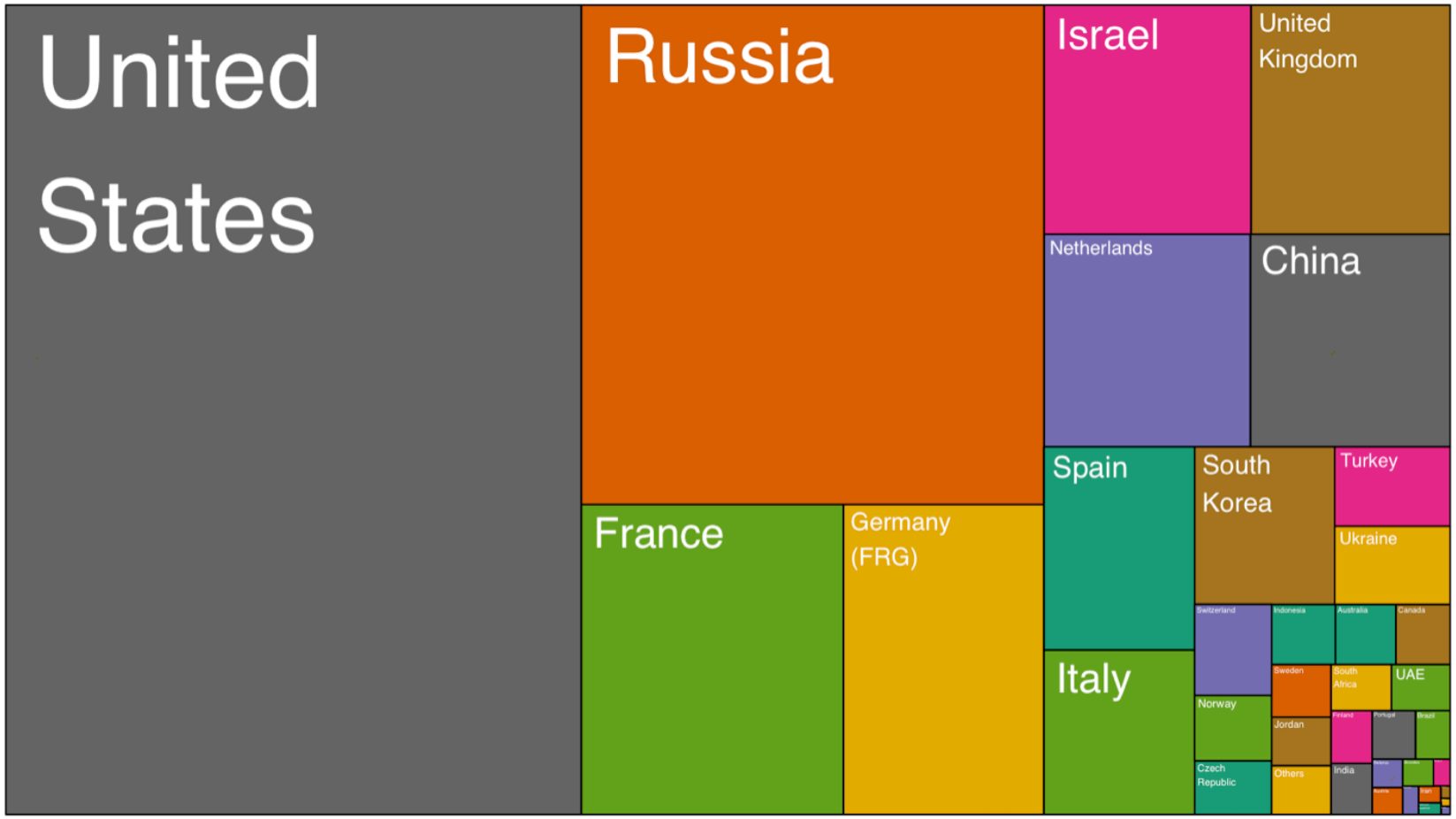
Who sells more weapons ?

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



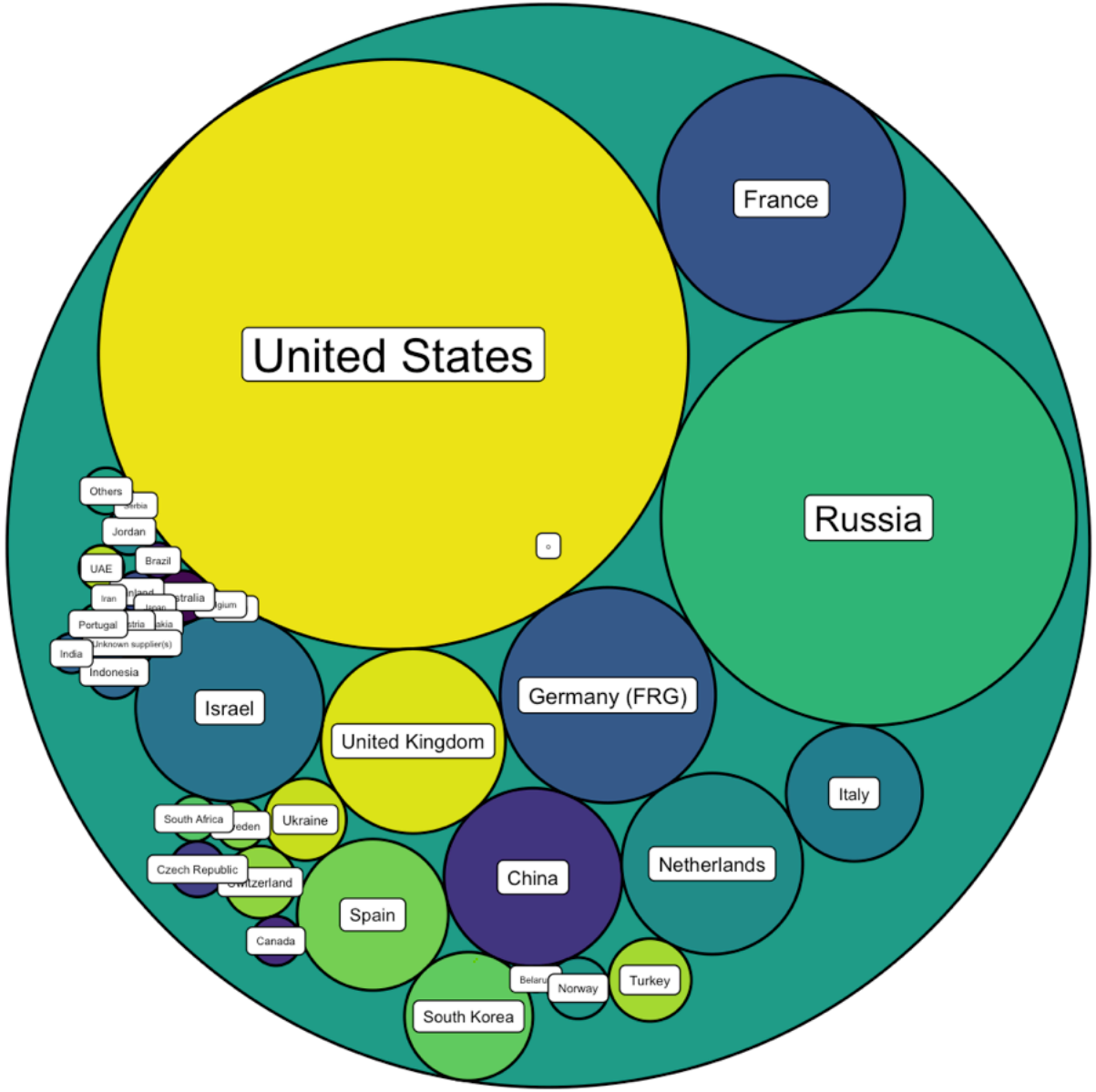
Who sells more weapons ?

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

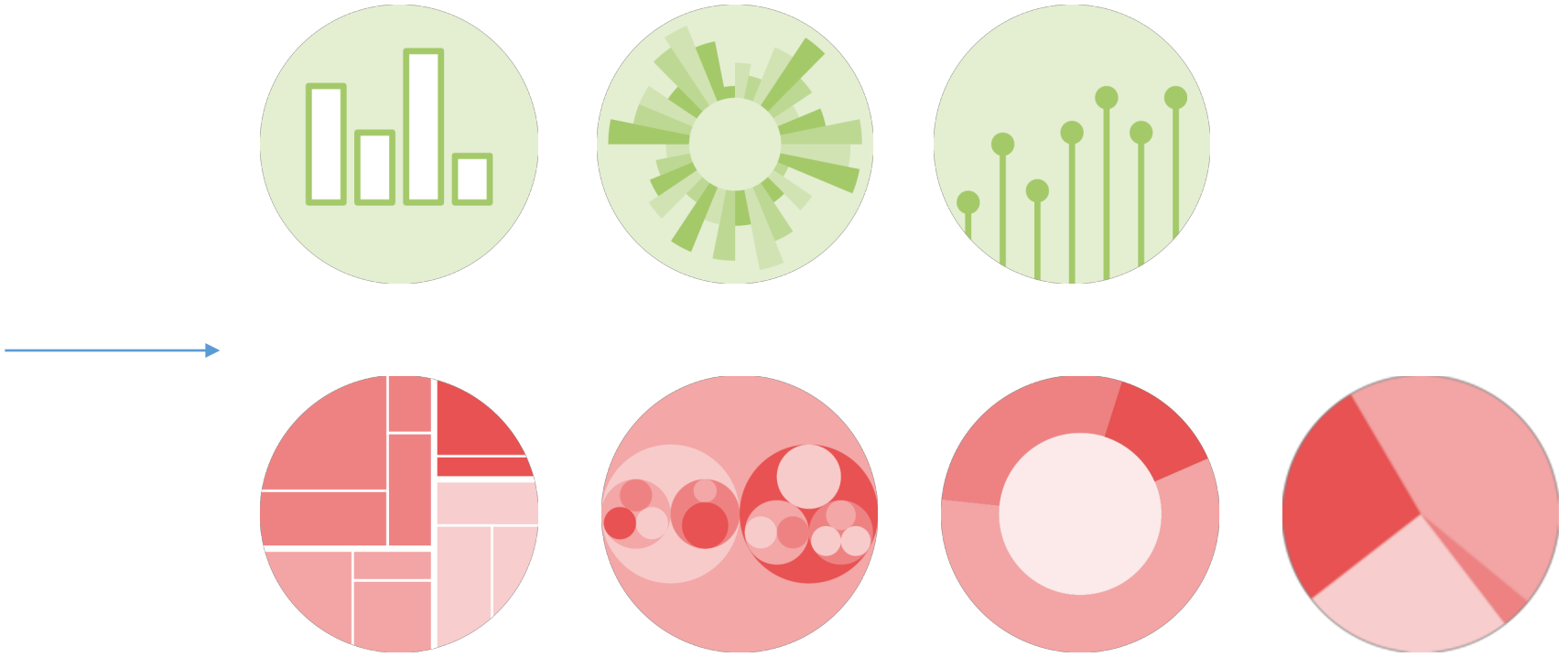
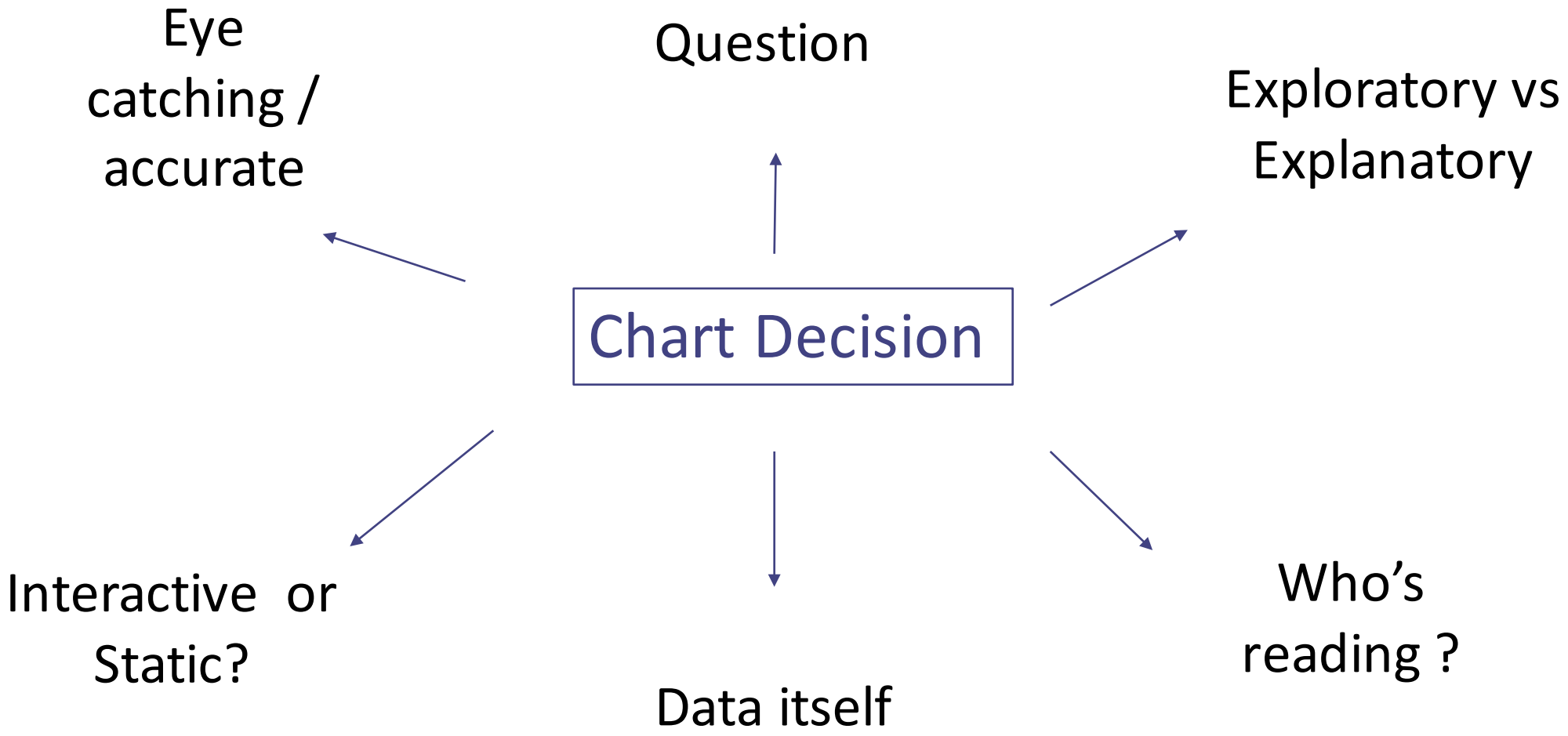
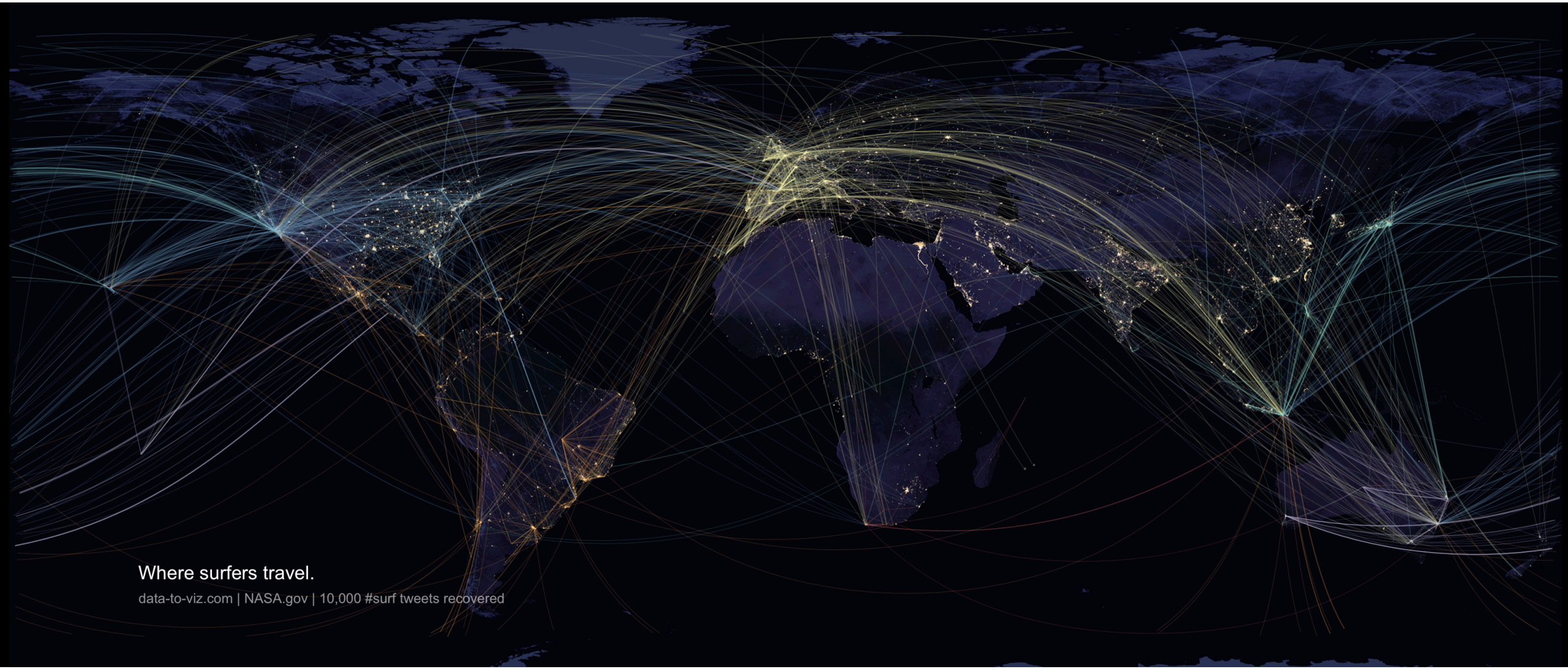


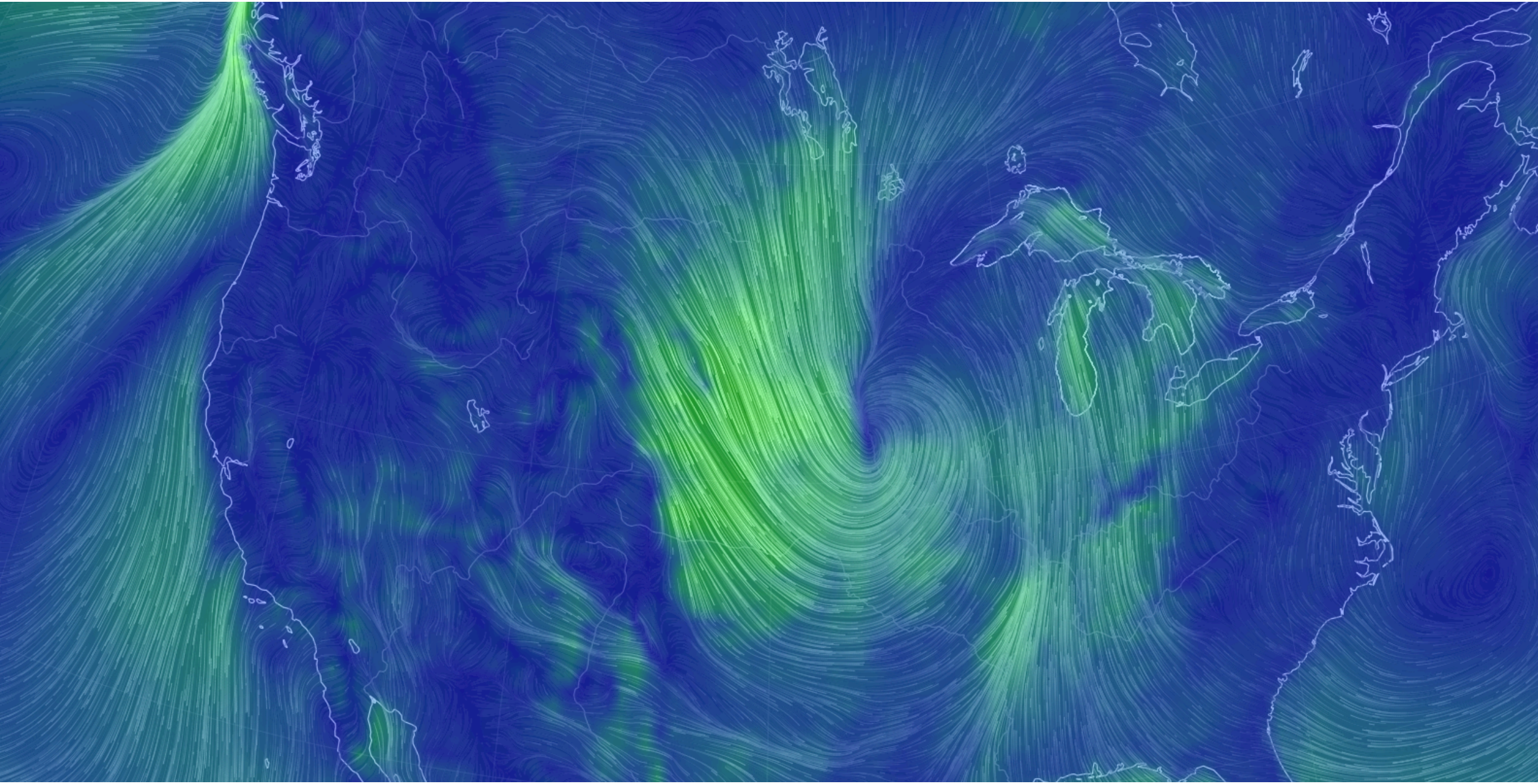
Chart Decision

?

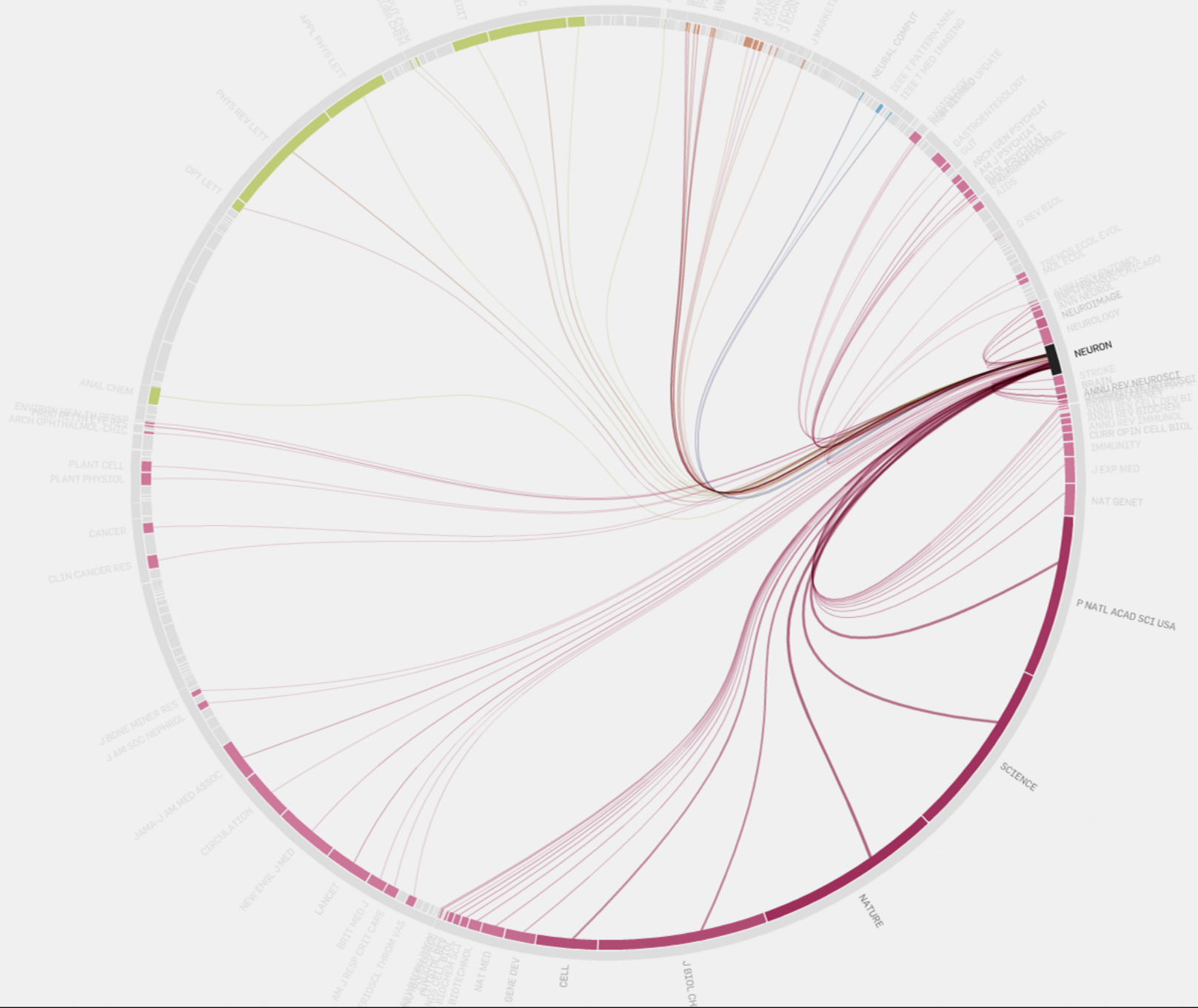

Eye catching / accurate
Question
Exploratory vs Explanatory
Chart Decision
Interactive or Static?
Data itself
Who's reading ?

Examples


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



NEURON

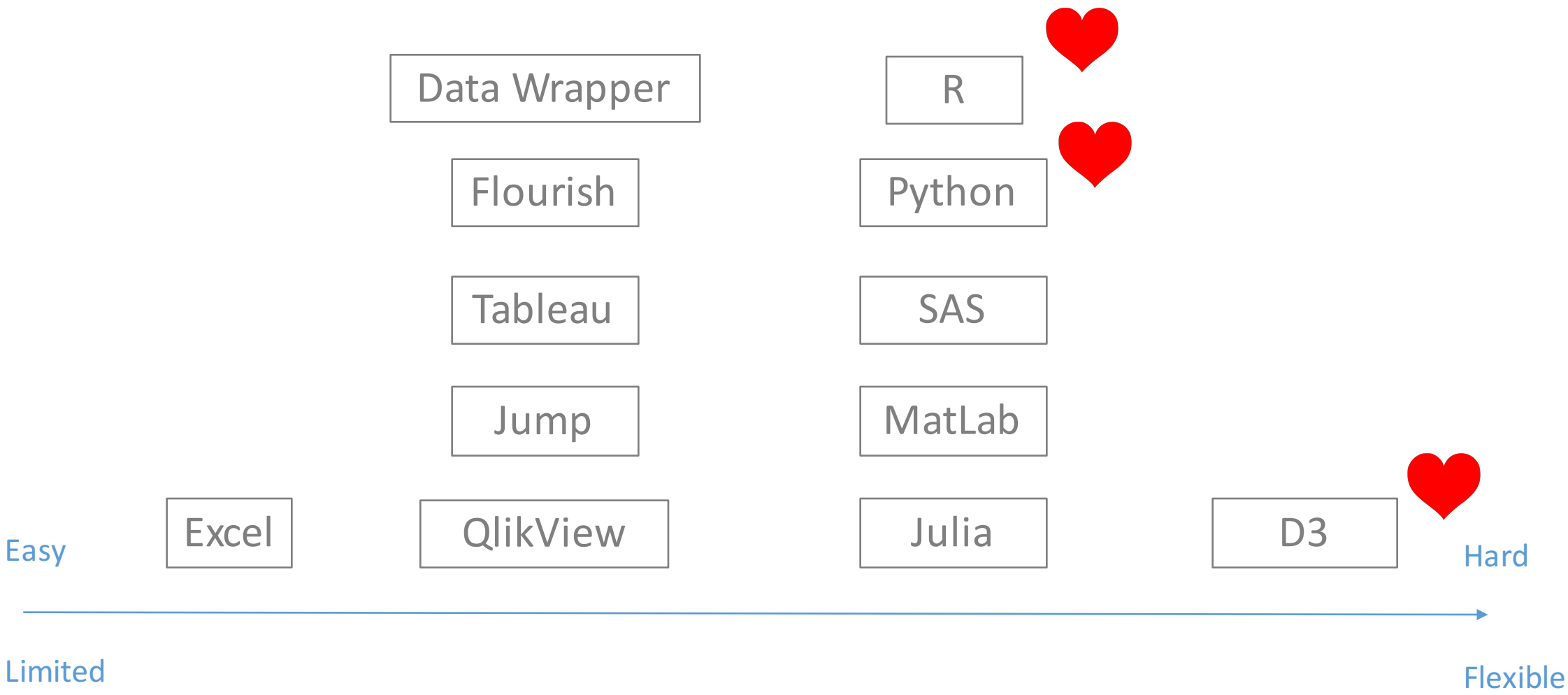


How

See the [chart maker directory](#) by visualisingdata.com

** Far from being exhaustive*

See the chart maker directory by visualisingdata.com



** Far from being exhaustive*

R-graph-gallery.com

Python-graph-gallery.com

D3-graph-gallery.com

Thanks

Slides: yan-holtz.com/teaching

 Yan.holtz.data@gmail.com

 www.yan-holtz.com

 @R_Graph_Gallery

 github.com/holtzy