Data Visualization & Storytelling

Marysia Winkels

PyData

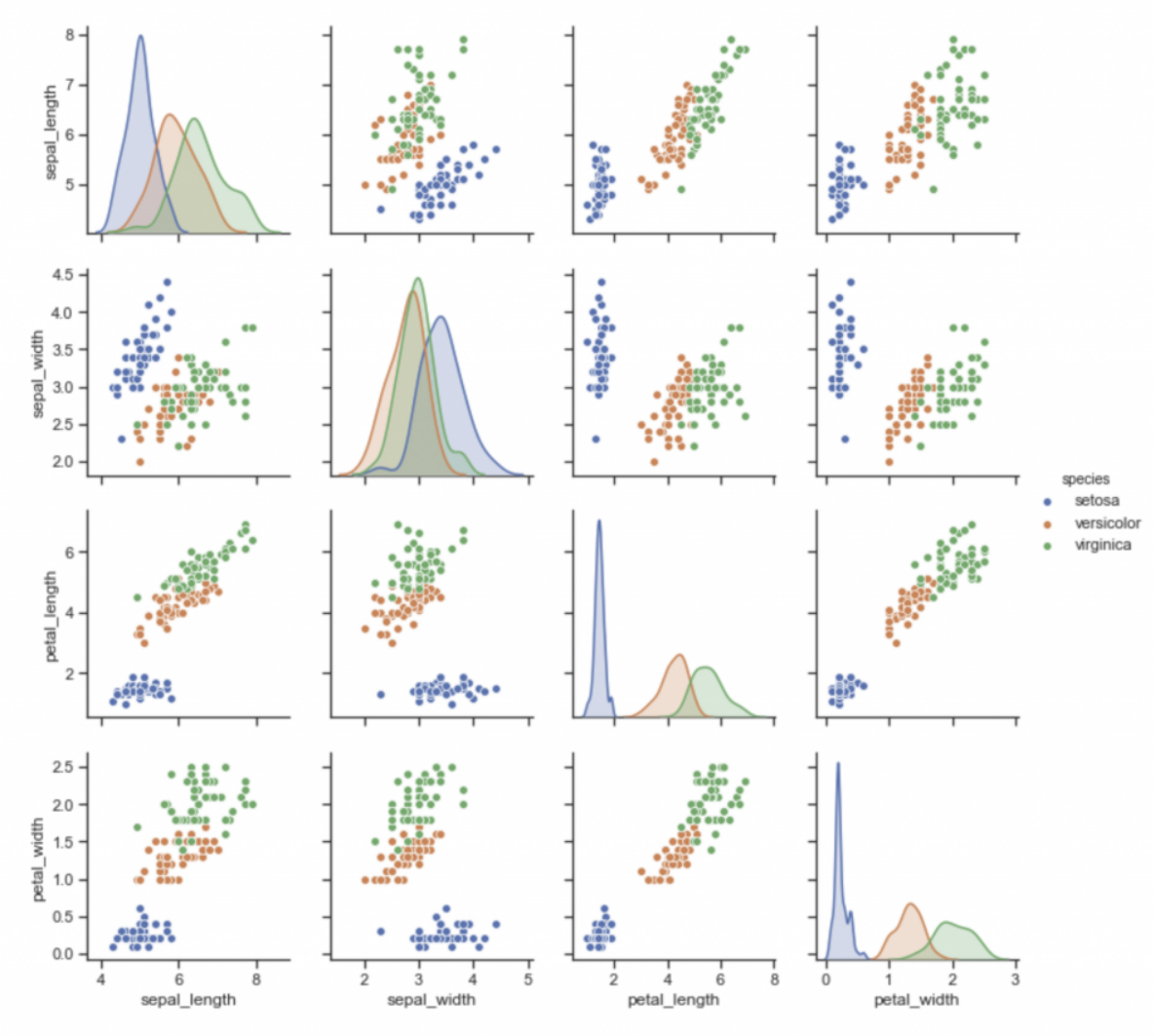
🤦‍♀️ About me

- Data Scientist
- PyData Amsterdam
- Storyteller

Data visualization can be used to tell a story

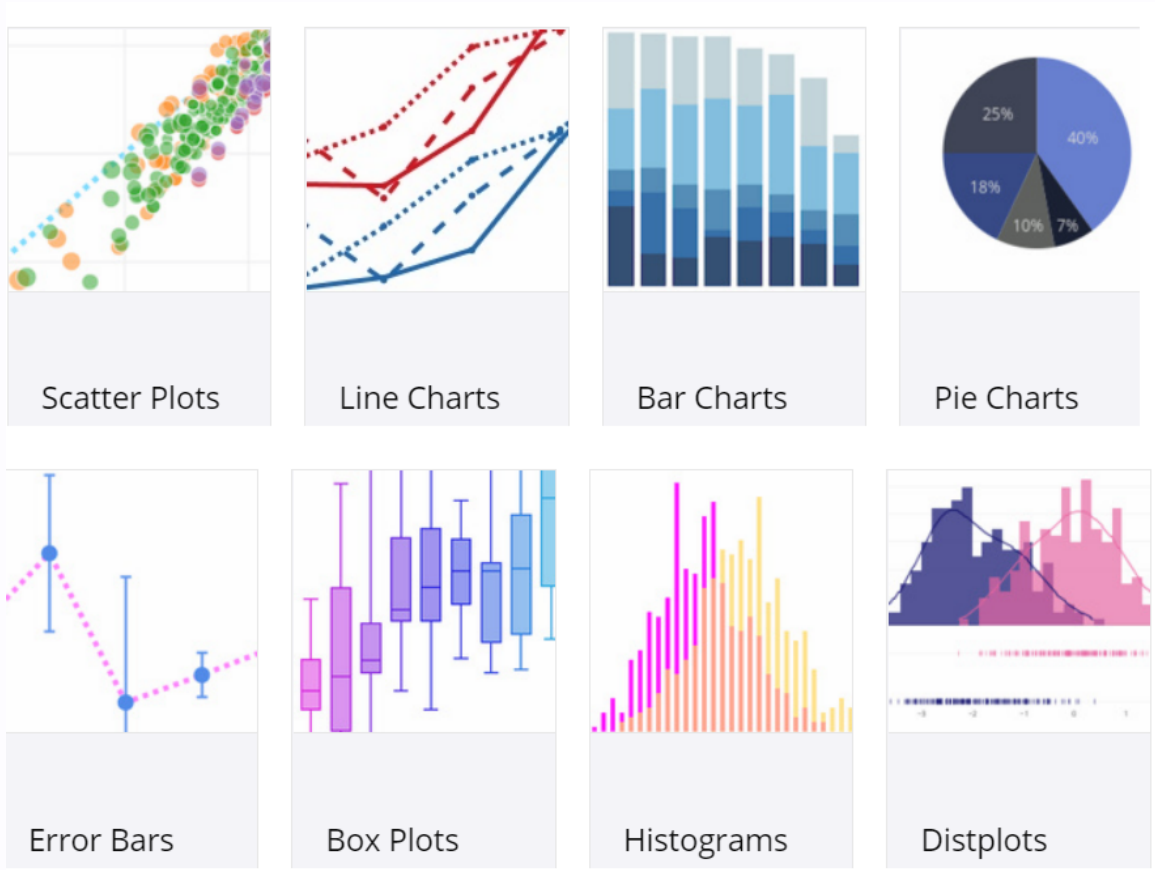
This talk will not be about:

- Data visualization for data analysis
- What the best plotting library is in Python



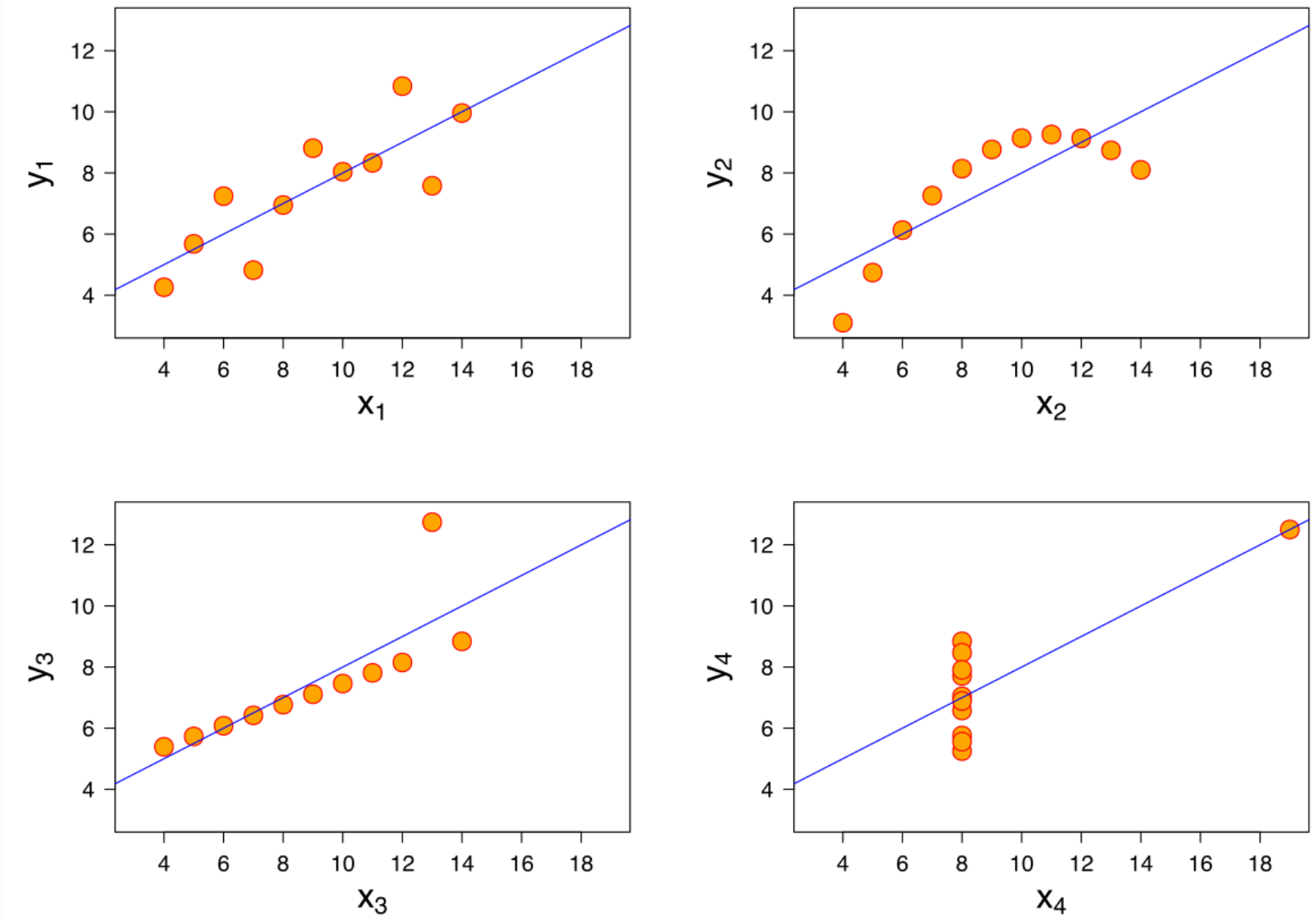
This talk **will** be about:

- *Why* you should tell a story with your data visualization
- *How* you can tell a story with your data visualization



The Importance of Data Visualization

Summary statistics are not enough to describe your data



	x	y
0	55.3846	97.1795
1	51.5385	96.0256
2	46.1538	94.4872
3	42.8205	91.4103
4	40.7692	88.3333
...	...	...
137	39.4872	25.3846
138	91.2821	41.5385
139	50.0000	95.7692
140	47.9487	95.0000
141	44.1026	92.6923

142 rows × 2 columns

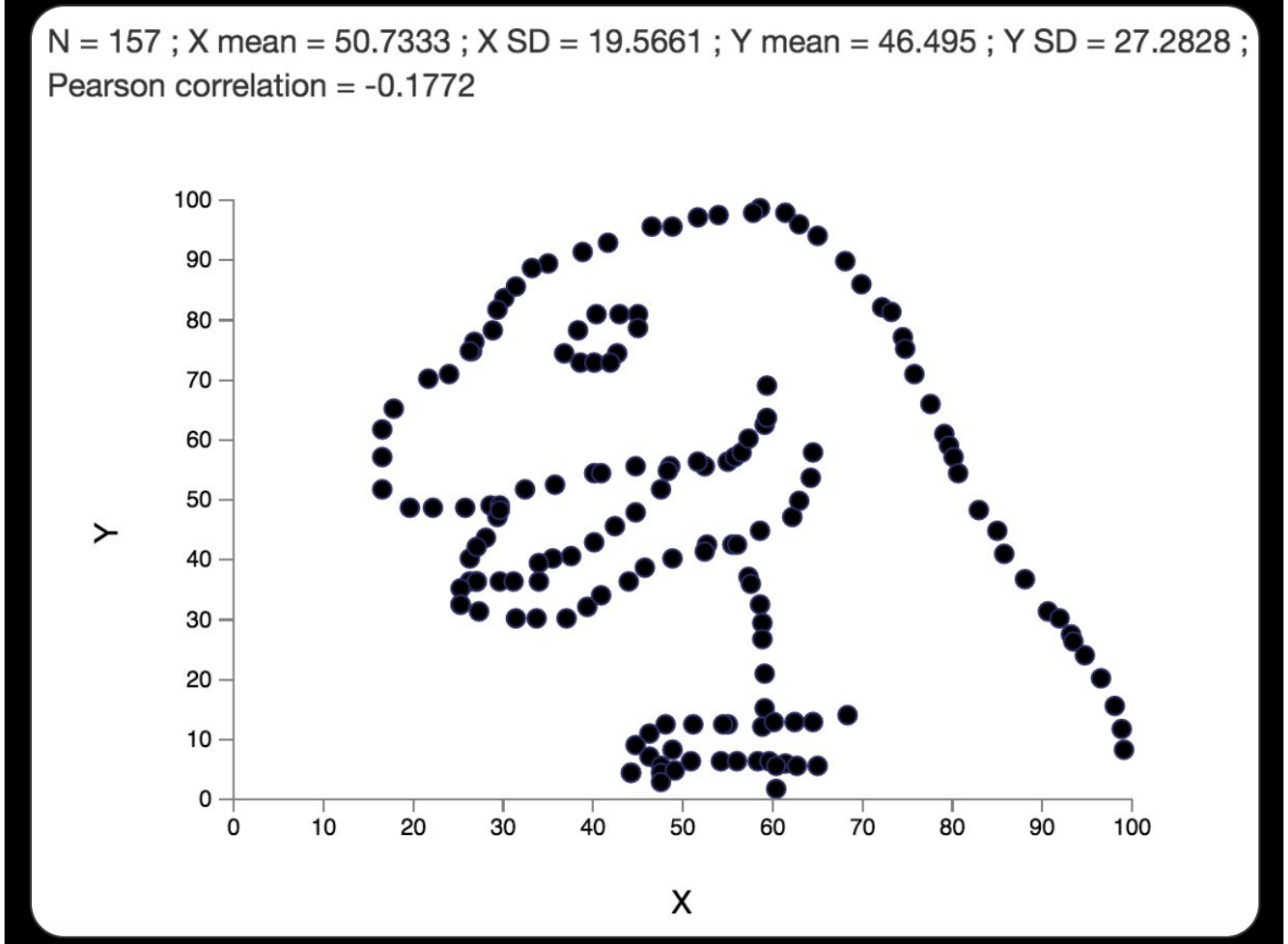
	x	y
0	55.3846	97.1795
1	51.5385	96.0256
2	46.1538	94.4872
3	42.8205	91.4103
4	40.7692	88.3333
...	...	...
137	39.4872	25.3846
138	91.2821	41.5385
139	50.0000	95.7692
140	47.9487	95.0000
141	44.1026	92.6923

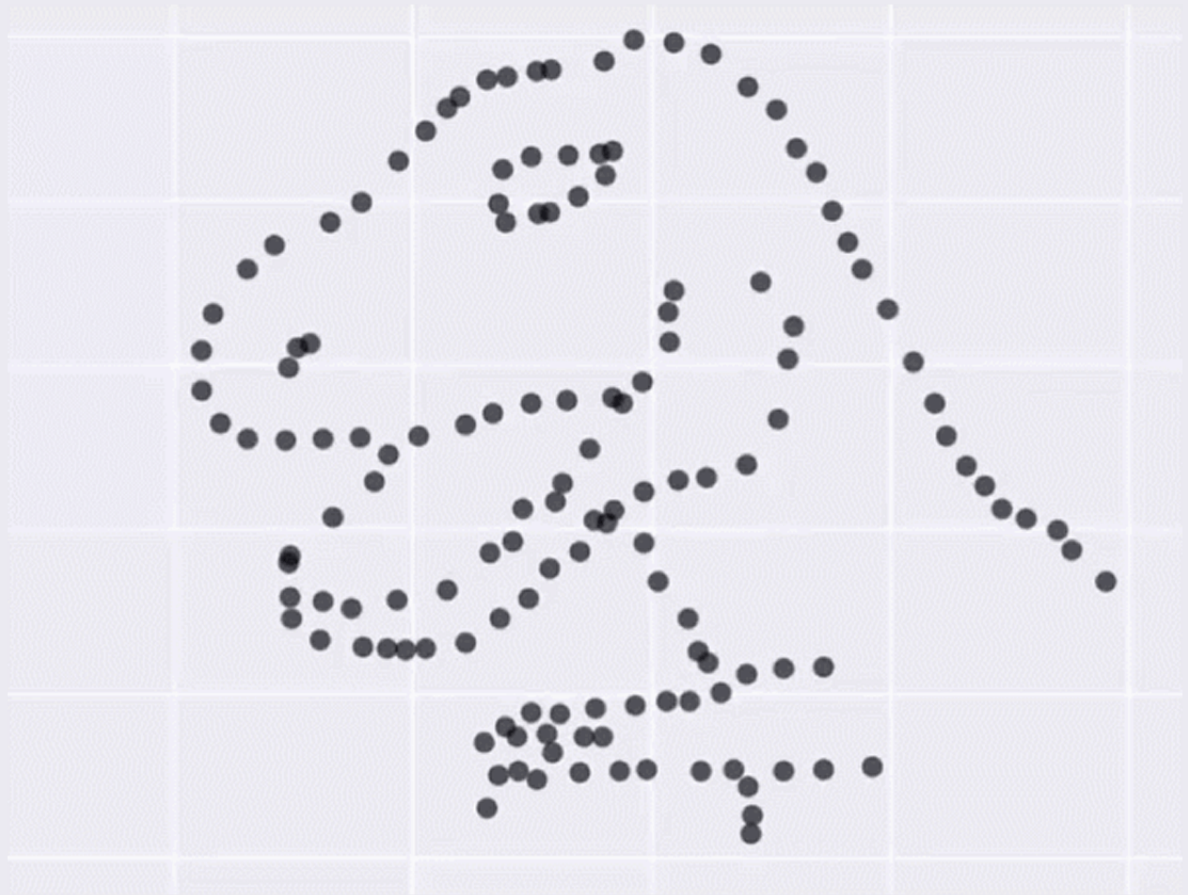
142 rows × 2 columns

```
N = 142
X mean = 54.26327323943662
X std = 16.765142039116 79
Y mean = 47.832252816901416
Y std = 26.935403486939116
Pearson correlation = -0.0645
```

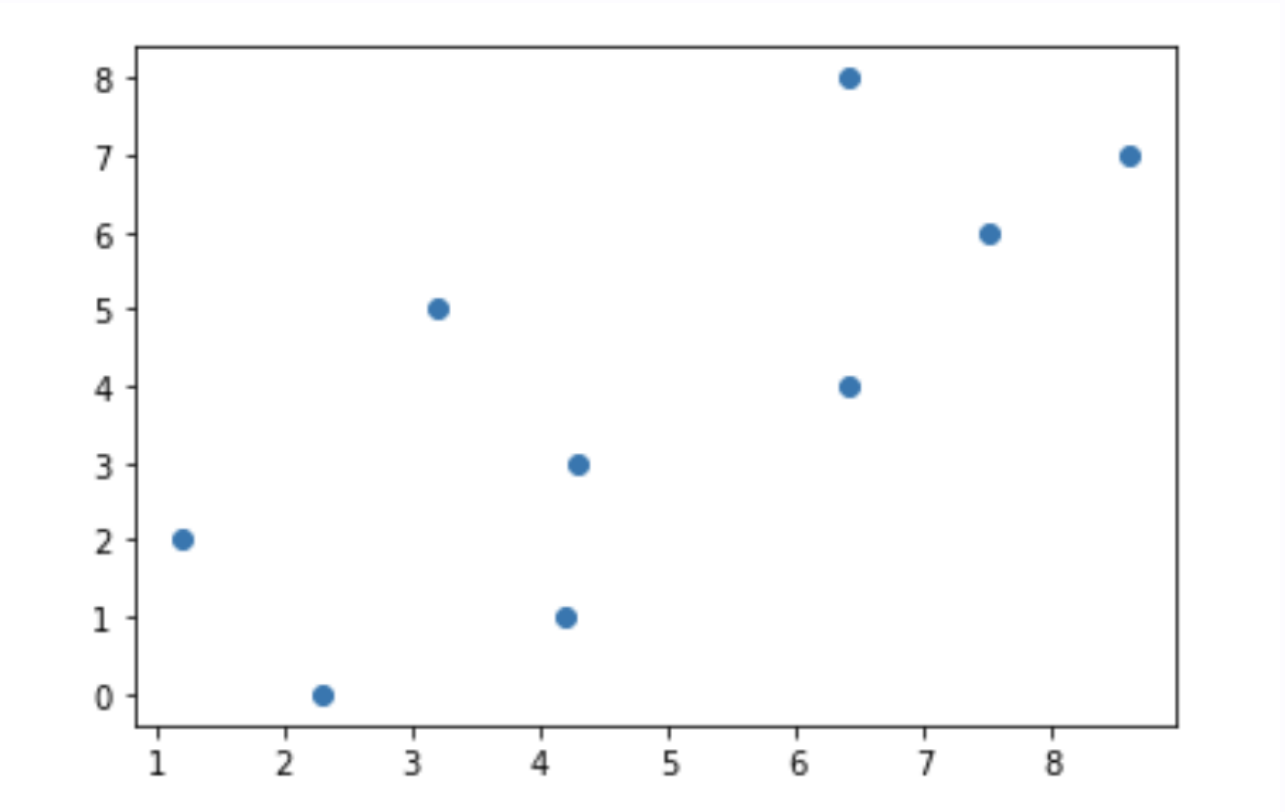
@albertocairo@mastodon.social
@AlbertoCairo

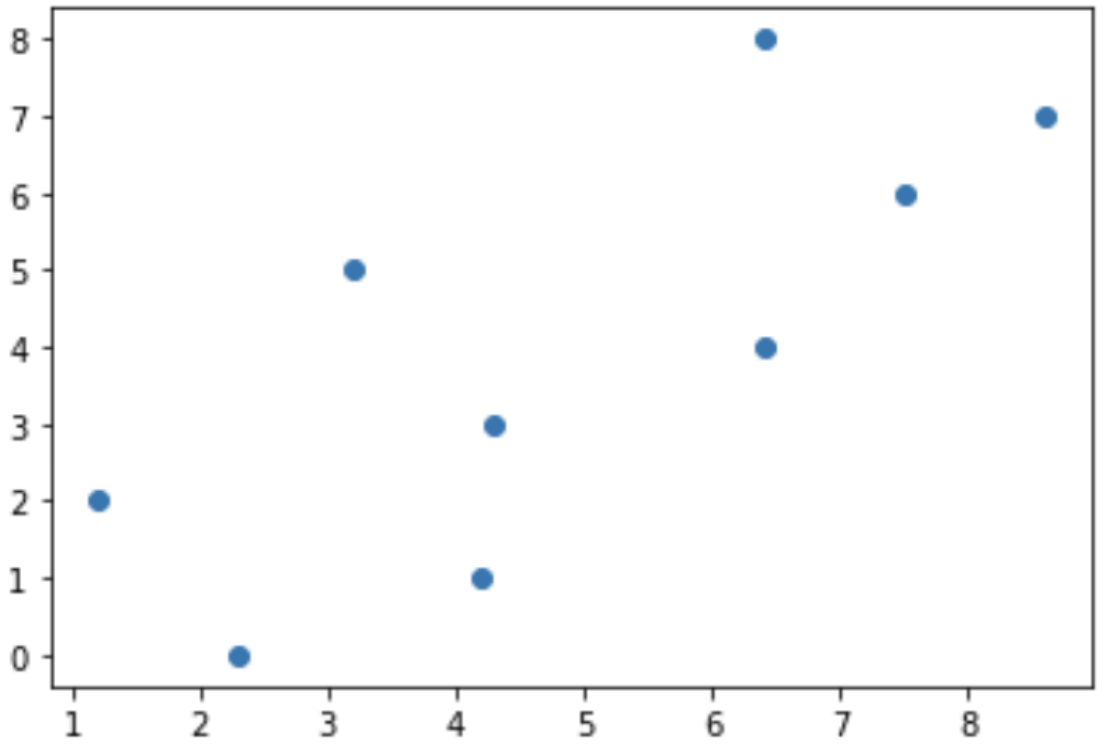
Don't trust summary statistics. Always visualize your data first robertgrantstats.co.uk/drawmydata.html

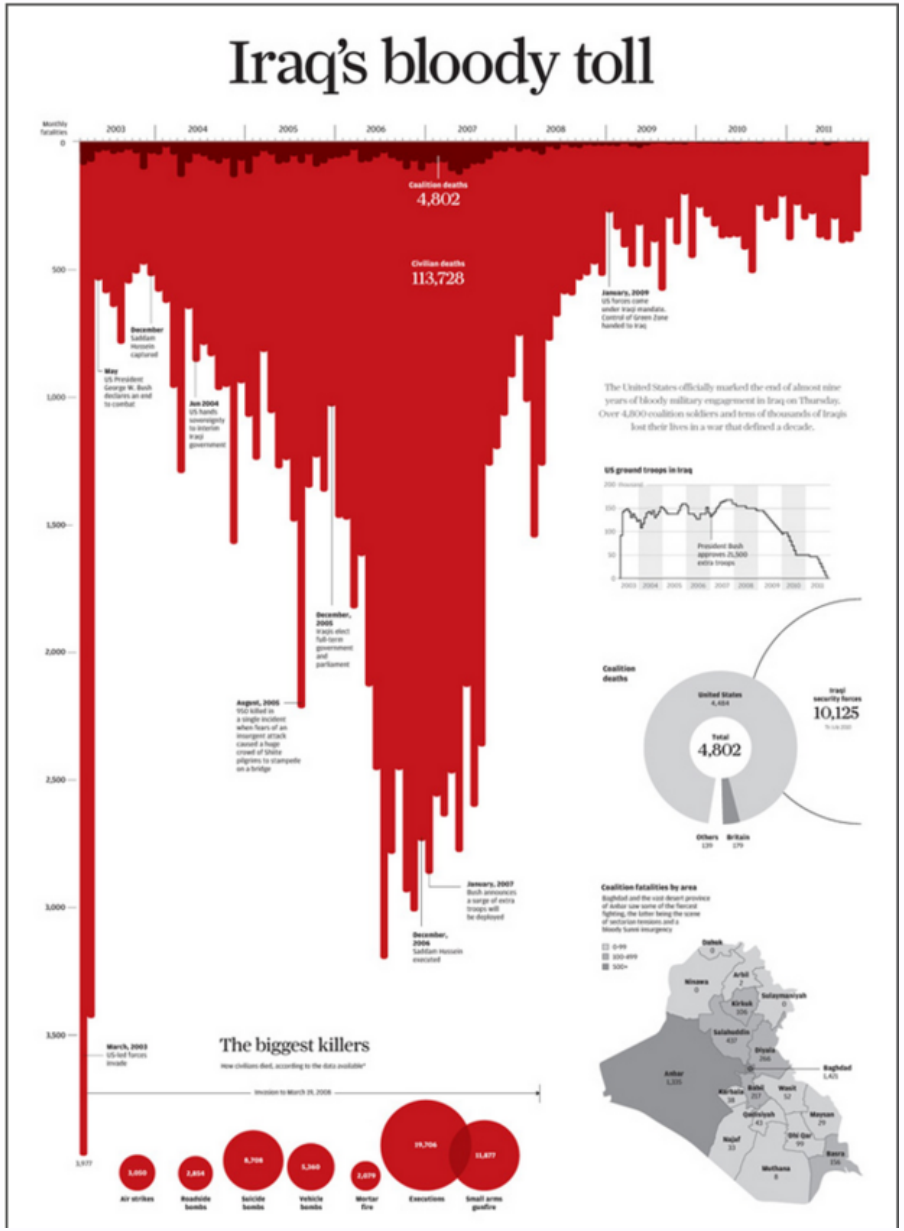


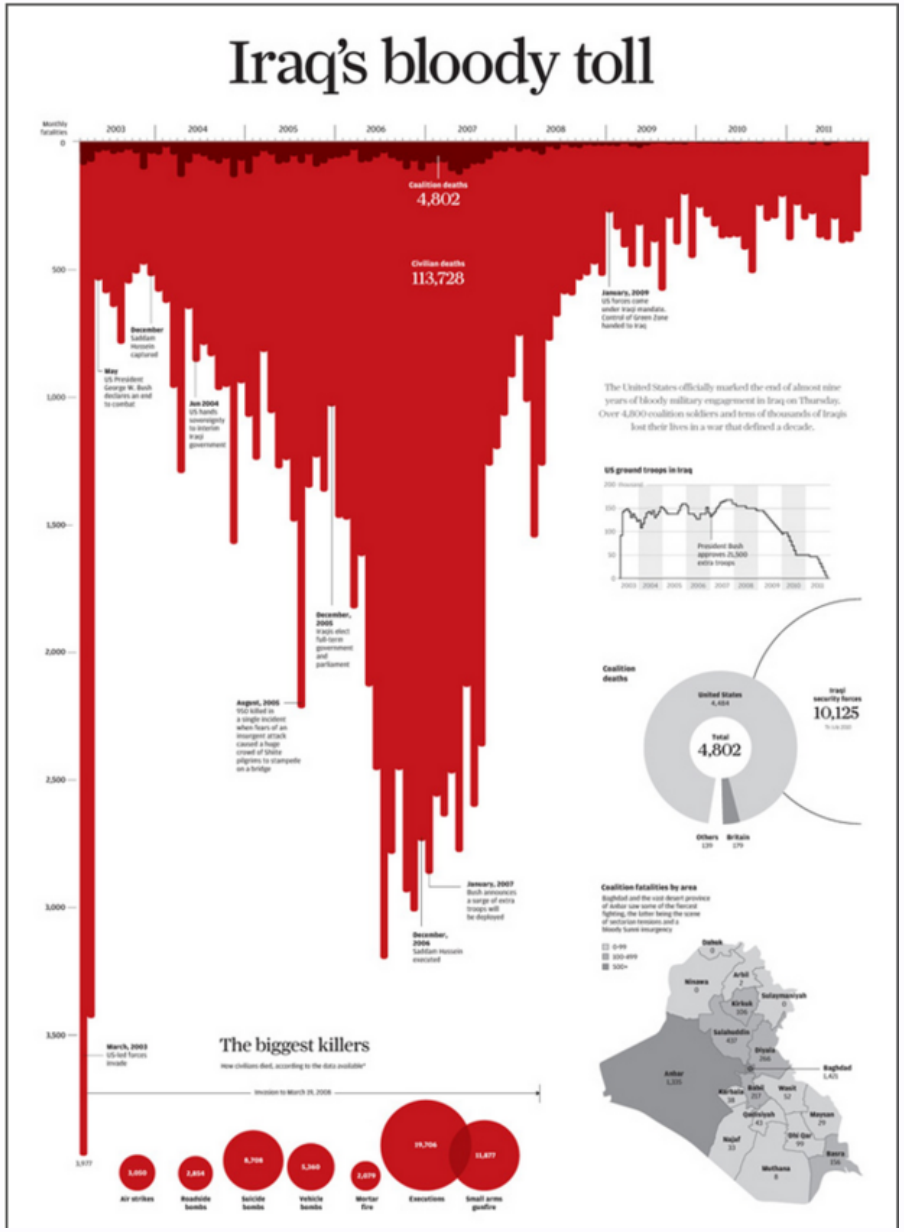


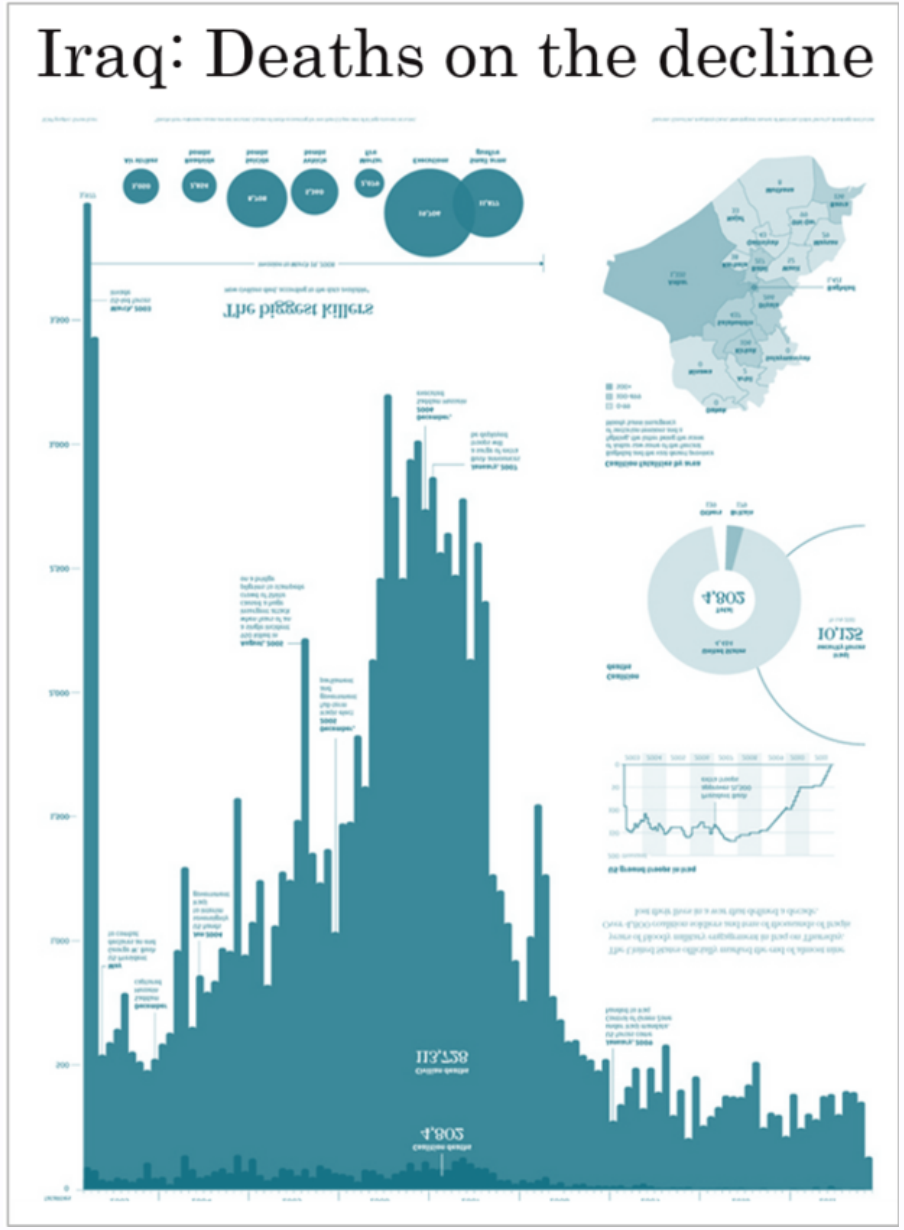
So what is data storytelling?


8
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8


8
7
6
5
4
3
2
1
0
1
2
3
4
5
6
7
8


Iraq's bloody toll
2003
2004
2005
2006
2007
2008
2009
2010
2011
Coalition deaths
4,802
Civilian deaths
113,728
December
Saddam Hussein captured
May
US President George W. Bush declares an end to combat
Jan 2004
US hands sovereignty to interim Iraqi government
January, 2009
US forces come under Iraqi mandate. Control of Green Zone handed to Iraq
The United States officially marked the end of almost nine years of bloody military engagement in Iraq on Thursday. Over 4,800 coalition soldiers and tens of thousands of Iraqis lost their lives in a war that defined a decade.
US ground troops in Iraq
President Bush approves 21,500 extra troops
December, 2005
Iraqis elect full-term government and parliament
August, 2005
Coalition deaths
United States
Iraqi security forces
10,125
Total
4,802
Others
Britain
January, 2007
Bush announces a surge of extra troops will be deployed
December, 2006
Saddam Hussein executed
March, 2003
US-led forces invade
The biggest killers
How civilians died, according to the data available*
Air strikes
Roadside bombs
Suicide bombs
Vehicle bombs
Mortar fire
Executions
Small arms gunfire
Coalition fatalities by area
Baghdad and the vast desert province of Anbar saw some of the fiercest fighting, the latter being the scene of sectarian tensions and a bloody Sunni insurgency
Dohuk
Nineveh
Arbil
Kirkuk
Salaheddin
Diyala
Baghdad
Anbar
Babil
Wasit
Karbala
Qadisiyah
Maysan
Najaf
Dhi Qar
Basra
Muthana


Iraq's bloody toll
Coalition deaths
4,802
Civilian deaths
113,728
The United States officially marked the end of almost nine years of bloody military engagement in Iraq on Thursday. Over 4,800 coalition soldiers and tens of thousands of Iraqis lost their lives in a war that defined a decade.
US ground troops in Iraq
Coalition deaths
Total
4,802
United States
4,484
Others
Britain
Iraqi security forces
10,125
Coalition fatalities by area
The biggest killers
How civilians died, according to the data available*
Air strikes
Roadside bombs
Suicide bombs
Vehicle bombs
Mortar fire
Executions
Small arms gunfire

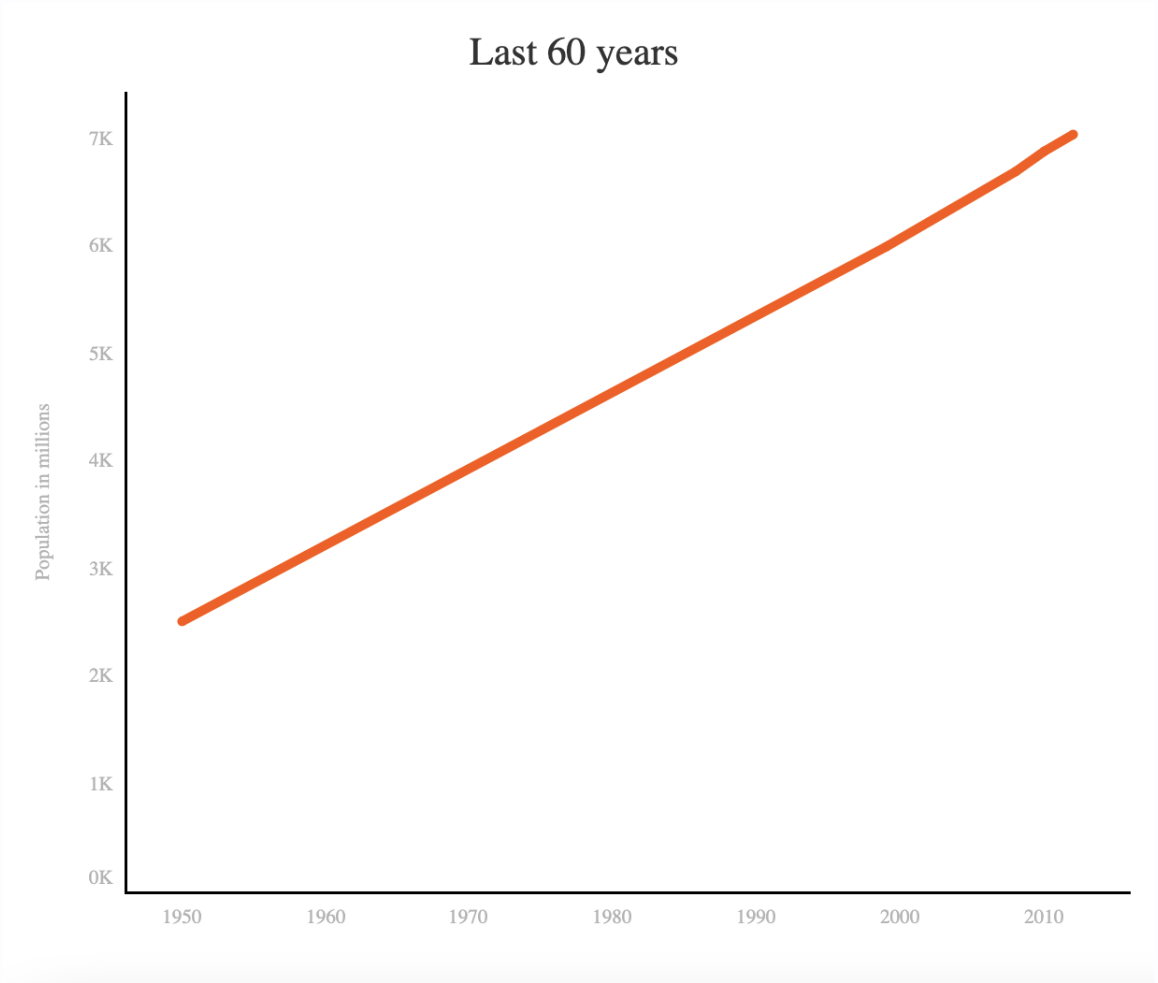

Iraq: Deaths on the decline

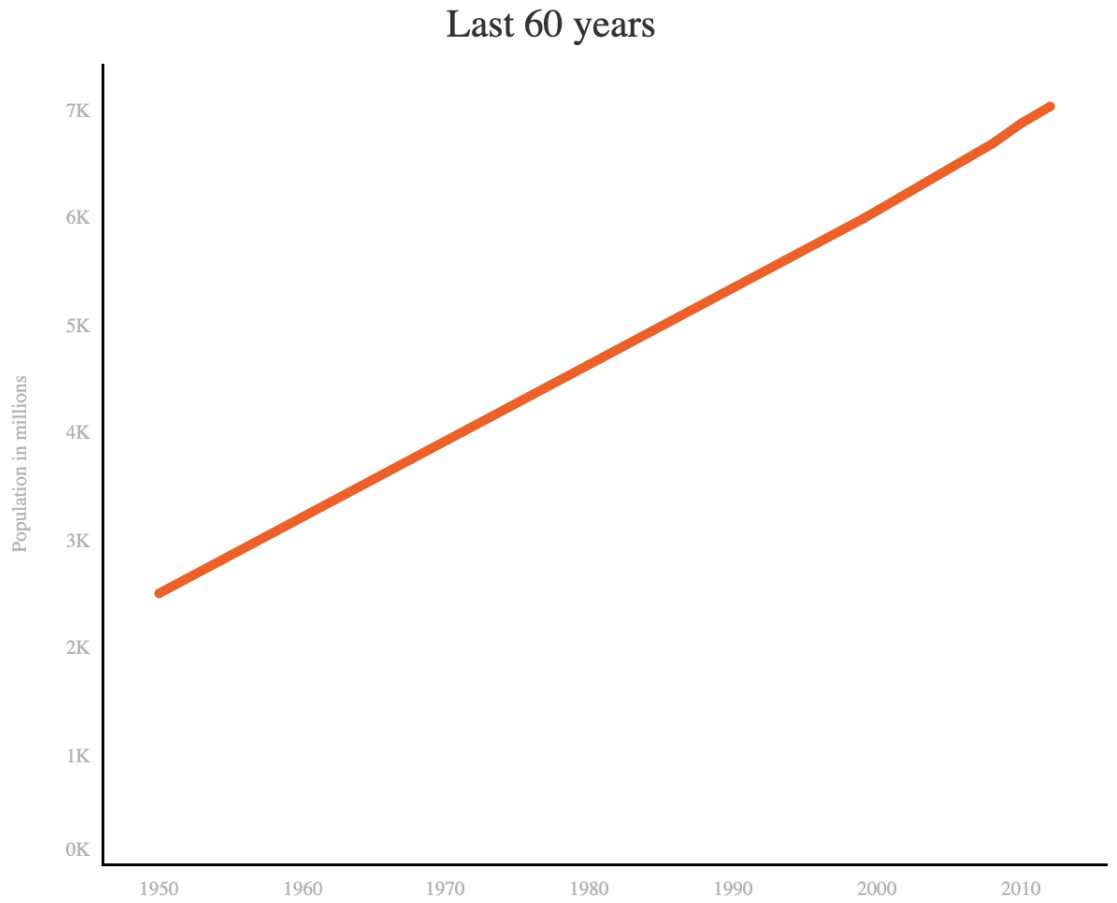
Types of data visualization

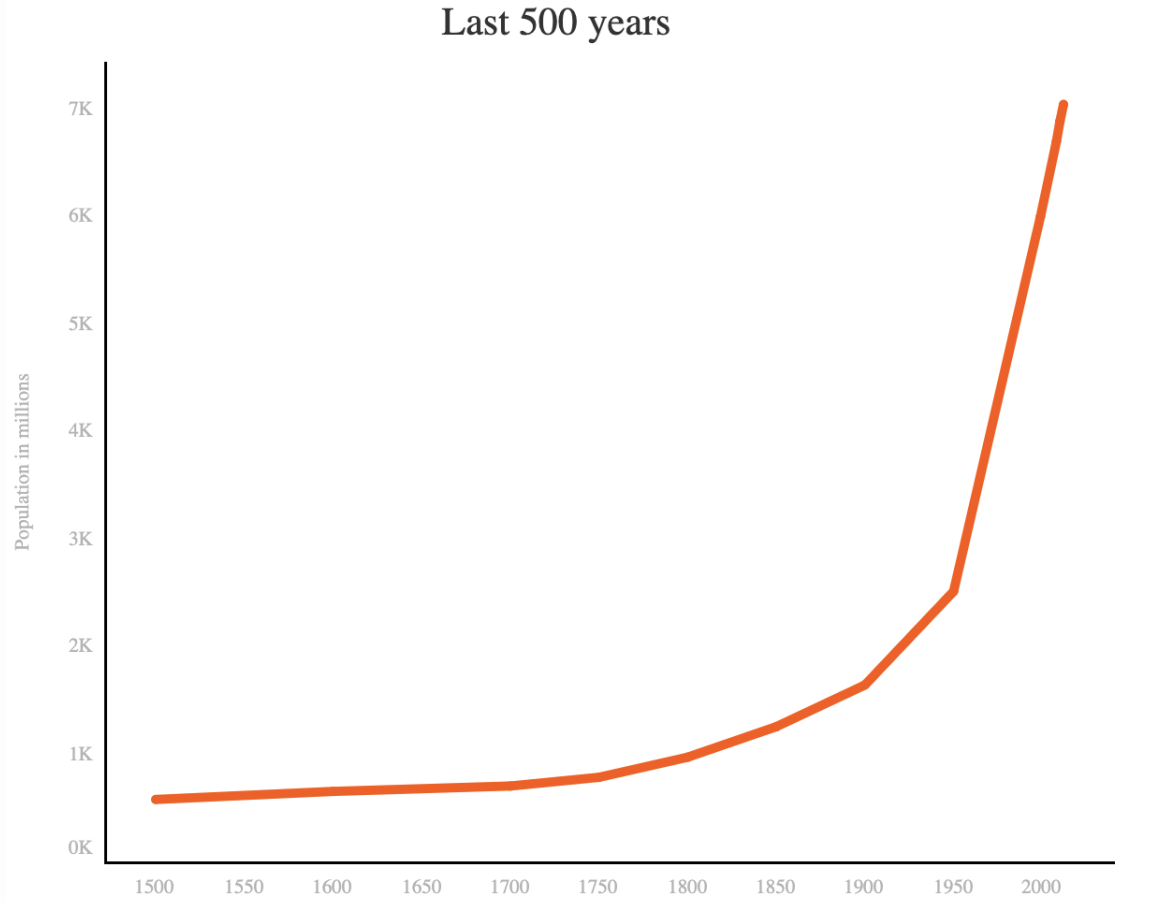
- Exploratory
- Explanatory

The best exploratory visualization is not necessarily the best explanatory visualization

Explanatory visualizations tell a story


Last 60 years
Population in millions
7K
6K
5K
4K
3K
2K
1K
0K
1950
1960
1970
1980
1990
2000
2010


Last 60 years
Population in millions
7K
6K
5K
4K
3K
2K
1K
0K
1950
1960
1970
1980
1990
2000
2010


Last 500 years
Population in millions
7K
6K
5K
4K
3K
2K
1K
0K
1500
1550
1600
1650
1700
1750
1800
1850
1900
1950
2000

Key points of a story:

- Place
- People
- Purpose

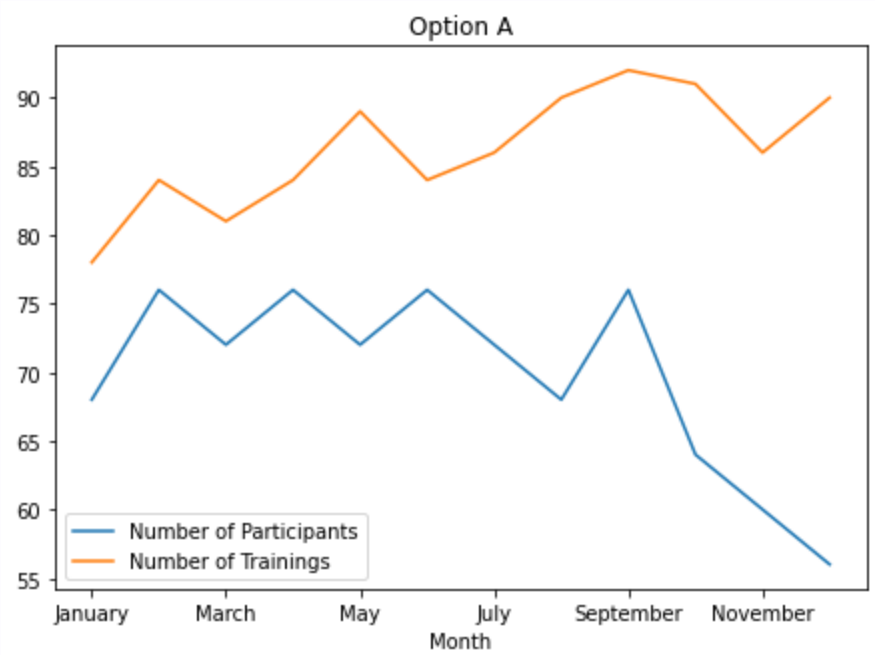
Key points of a **data** story:

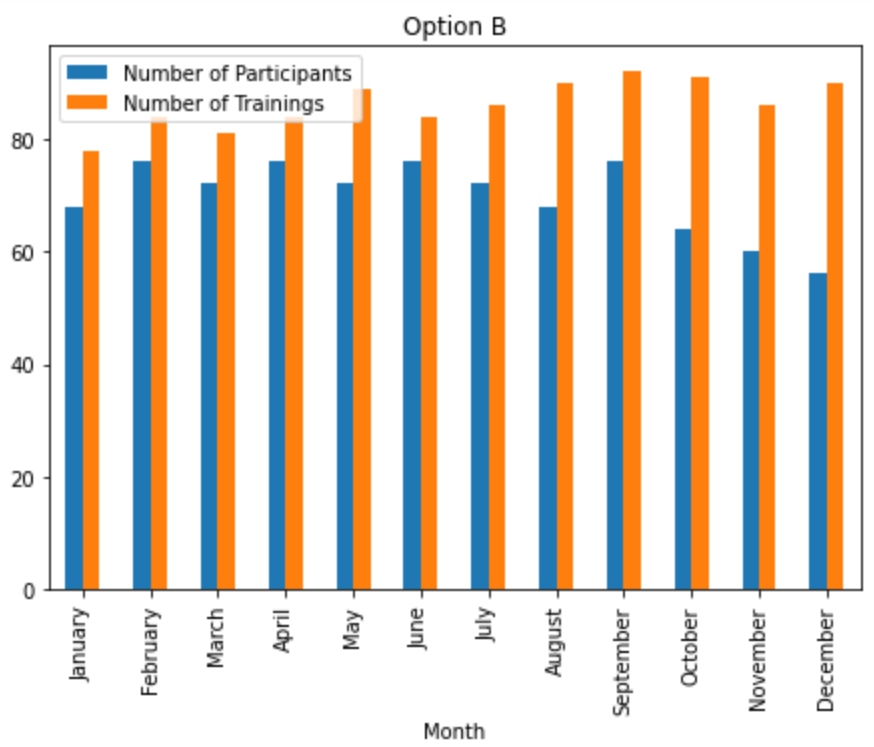
- Foundation
- Focus
- Forward

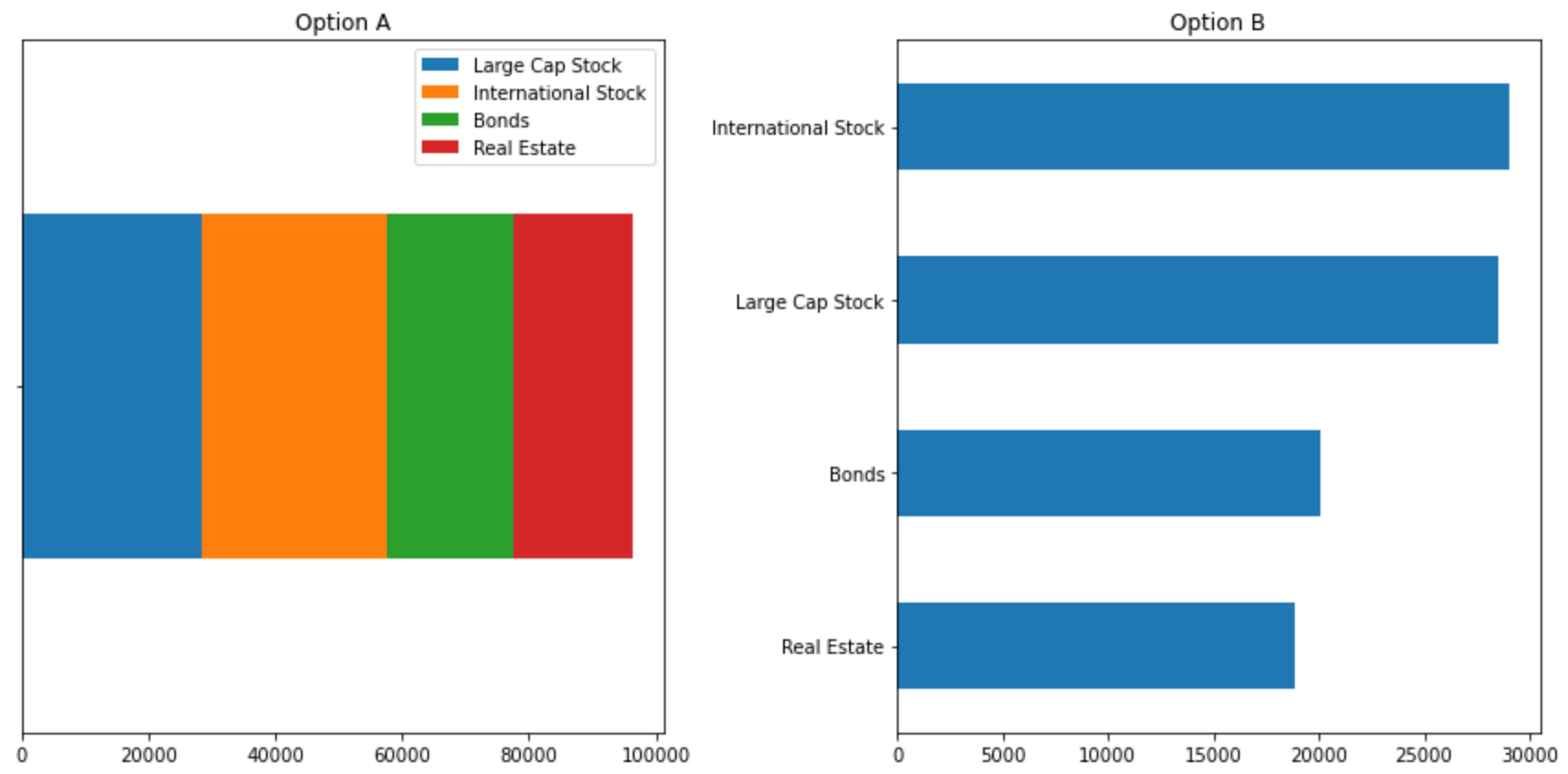
FOUNDATION

Types of charts:

- line plots
- scatter plots
- bar charts
- pie charts
- ... etc.


Option A
90
85
80
75
70
65
60
55
Number of Participants
Number of Trainings
January
March
May
July
September
November
Month


Option B
Number of Participants
Number of Trainings
80
60
40
20
0
January
February
March
April
May
June
July
August
September
October
November
December
Month


Option A
Large Cap Stock
International Stock
Bonds
Real Estate
0
20000
40000
60000
80000
100000
Option B
International Stock
Large Cap Stock
Bonds
Real Estate
0
5000
10000
15000
20000
25000
30000

How to determine the right chart for your use case?

1) Choose your chart

What would you like to show?

categories | time | part to whole | distribution | geospatial | relationship | exact value

barchart
one measure

grouped bar
two variables

diverging stacked bar
opposing variables

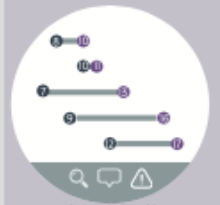
deviation bar
delta between

floating bar
delta between

stacked bar
one category + total

panel bar
multiple categories

lollipop
like bar but thinner



dumbbell
two groups

proportional
sizing

x/y coordinate plot
measure combination

vertical waterfall
visual calculation

bullet graph
bad/ok/good

parallel coordinate
multi variate data

pictograph
using icons

wordcloud
not recommended

line chart
continuous time

dot-line chart
aggregated in time

area chart
one measure

column chart
one measure

stacked column
like column + total

stacked area
like area + total

deviation column
above or below target

deviation line
versus cumulative target

waterfall
change in time

timeline
order of events

sparklines
mini trend

slopegraph
two time stamps

100% stacked bar
one categorie + 100%

100% stacked column
one timestamp + 100%

100% stacked area
continuous time + 100%

100% bar chart
adding up to 100%

100% stacked bar
category in time

100% stacked column
multiple categories

Sankey diagram
flow

waffle chart
100 blocks filled

tree map
nested part to whole

100% waterfall
breakdown

parallel set
part to multiple whole

Marimekko chart
plus extra variable

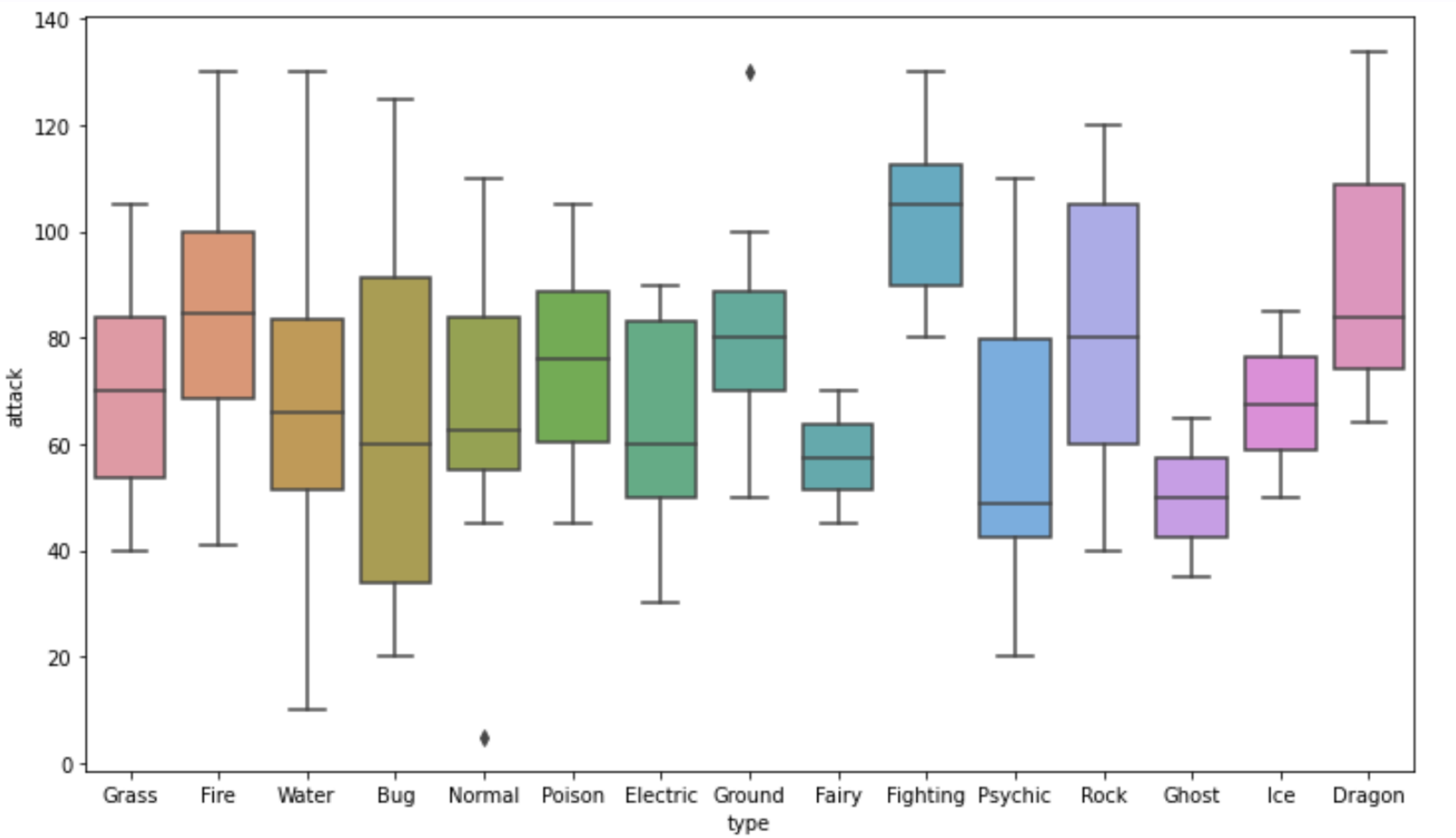
nested area
parts inside other parts

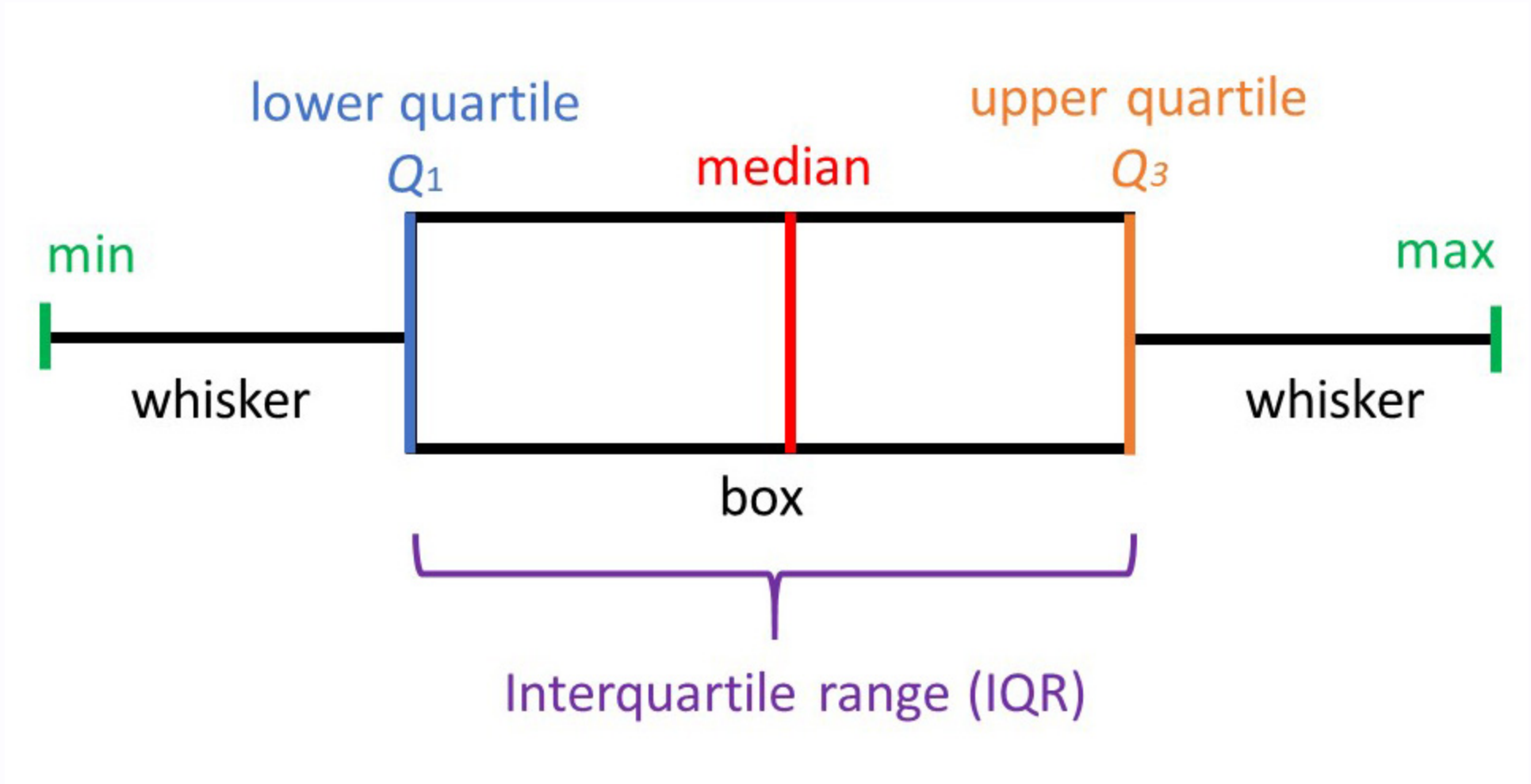
Pareto chart
80 / 20 analysis

pie chart
not recommended

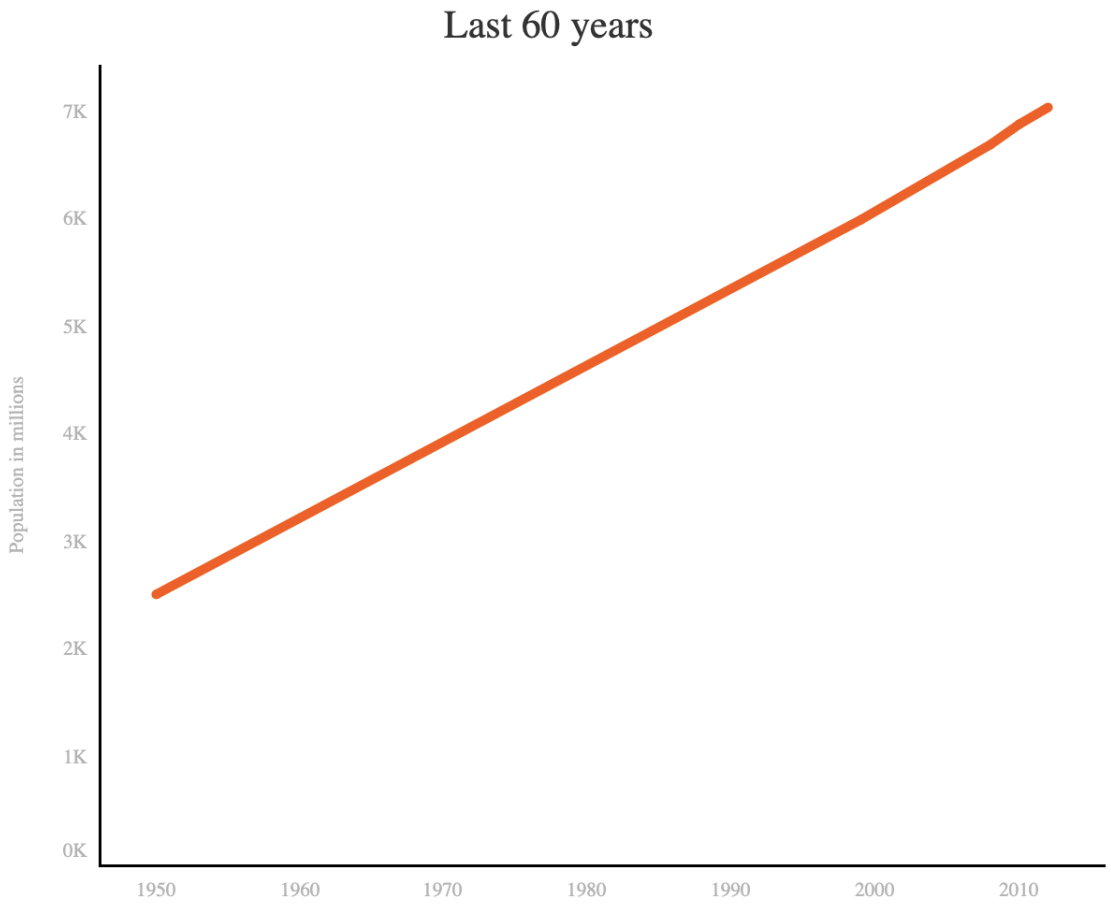
donut chart
not recommended

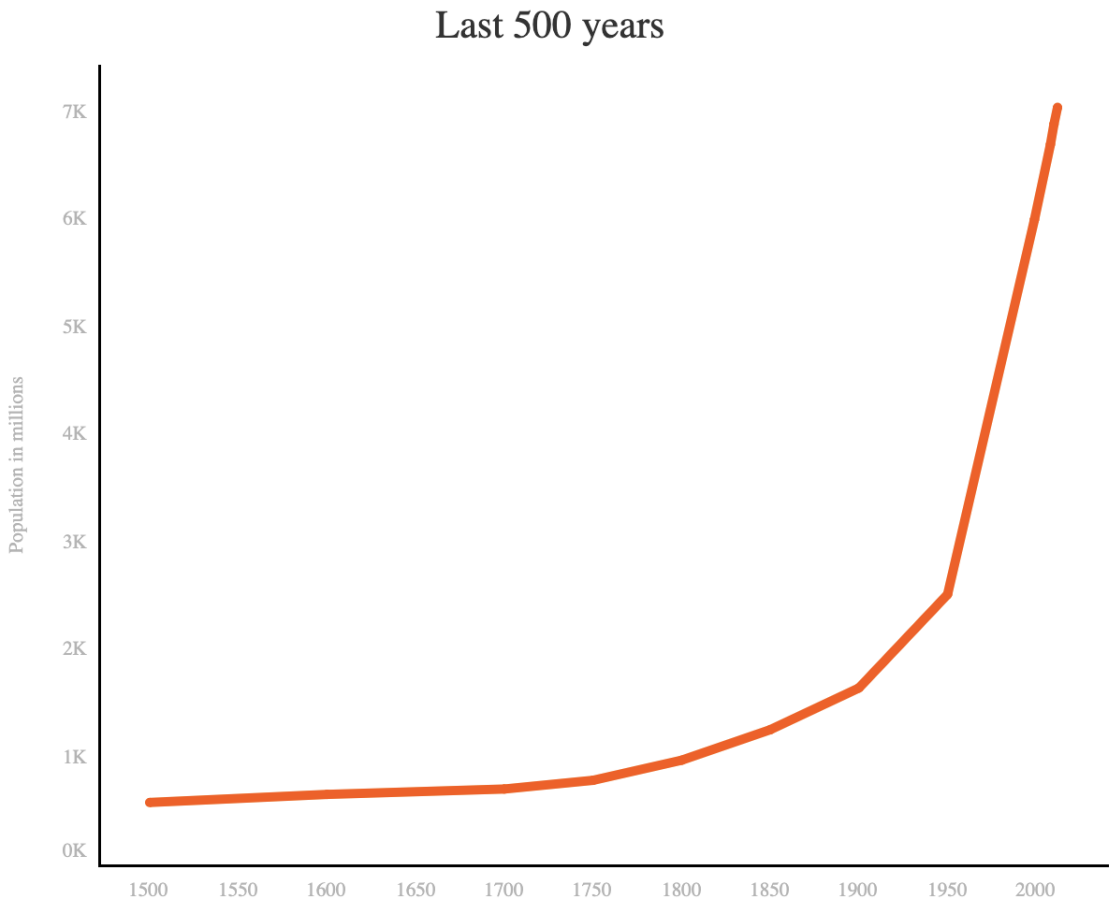
Make sure the reader is familiar with your type of chart

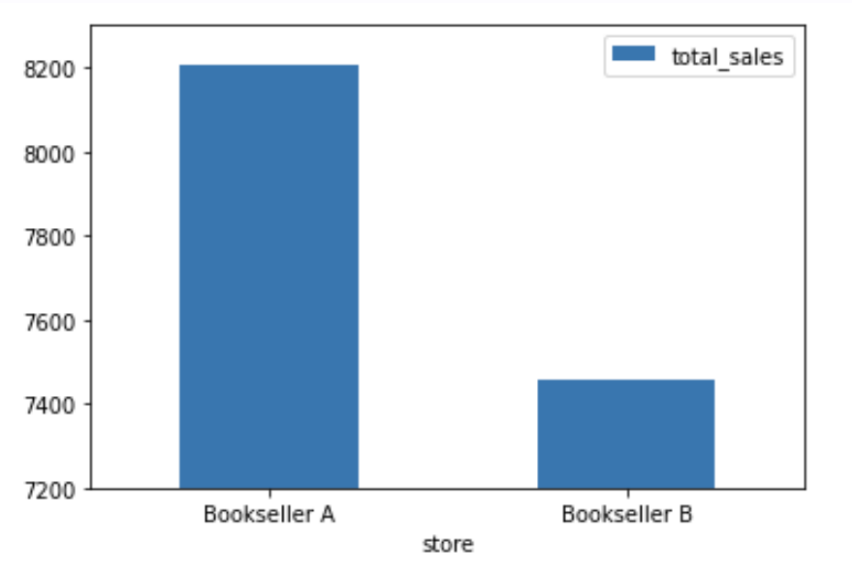

attack
type
0
20
40
60
80
100
120
140
Grass
Fire
Water
Bug
Normal
Poison
Electric
Ground
Fairy
Fighting
Psychic
Rock
Ghost
Ice
Dragon

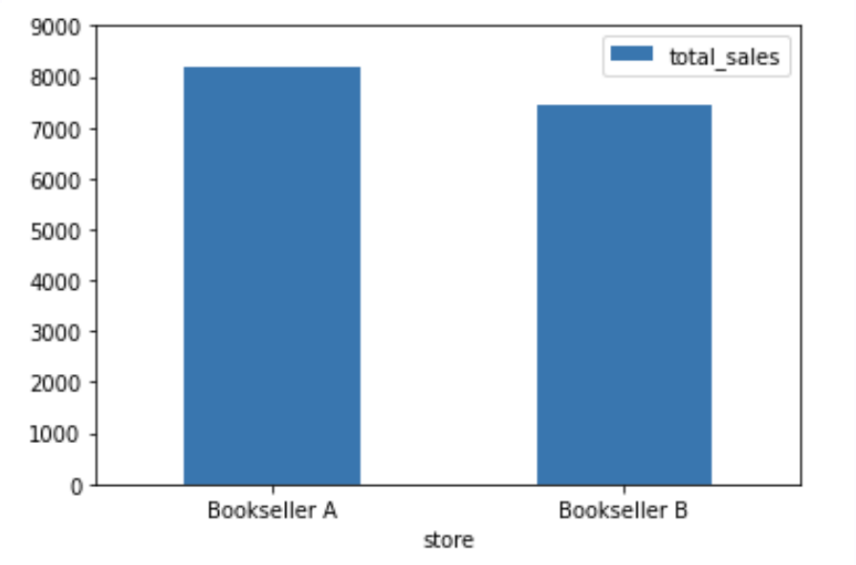

lower quartile
Q_1
median
upper quartile
Q_3
min
max
whisker
whisker
box
Interquartile range (IQR)

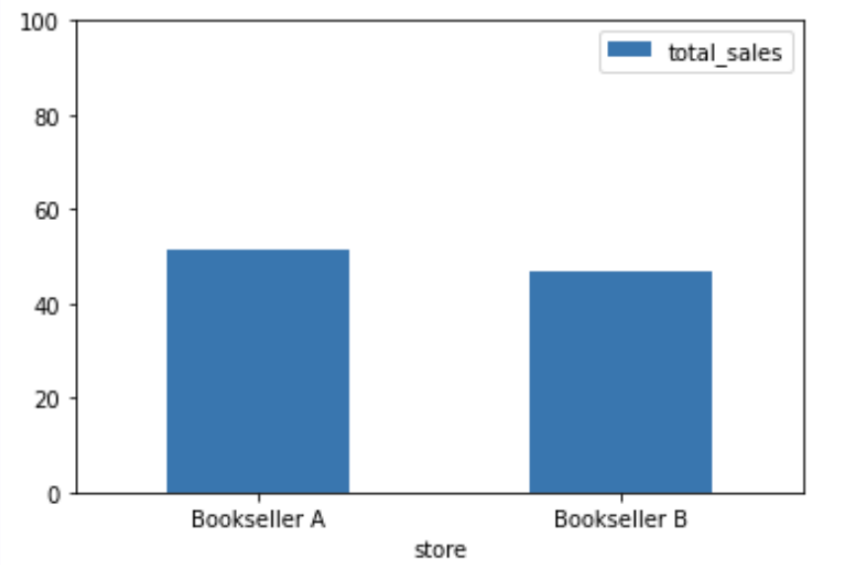
Choose the data to display carefully


Last 60 years
Population in millions
7K
6K
5K
4K
3K
2K
1K
0K
1950
1960
1970
1980
1990
2000
2010


Last 500 years
Population in millions
7K
6K
5K
4K
3K
2K
1K
0K
1500
1550
1600
1650
1700
1750
1800
1850
1900
1950
2000

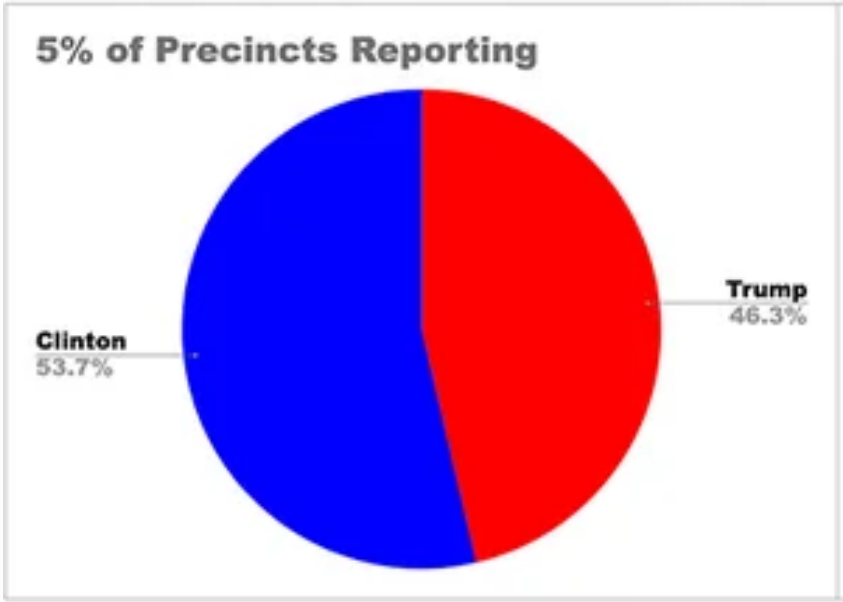

total_sales
8200
8000
7800
7600
7400
7200
Bookseller A
Bookseller B
store


total_sales
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Bookseller A
Bookseller B
store

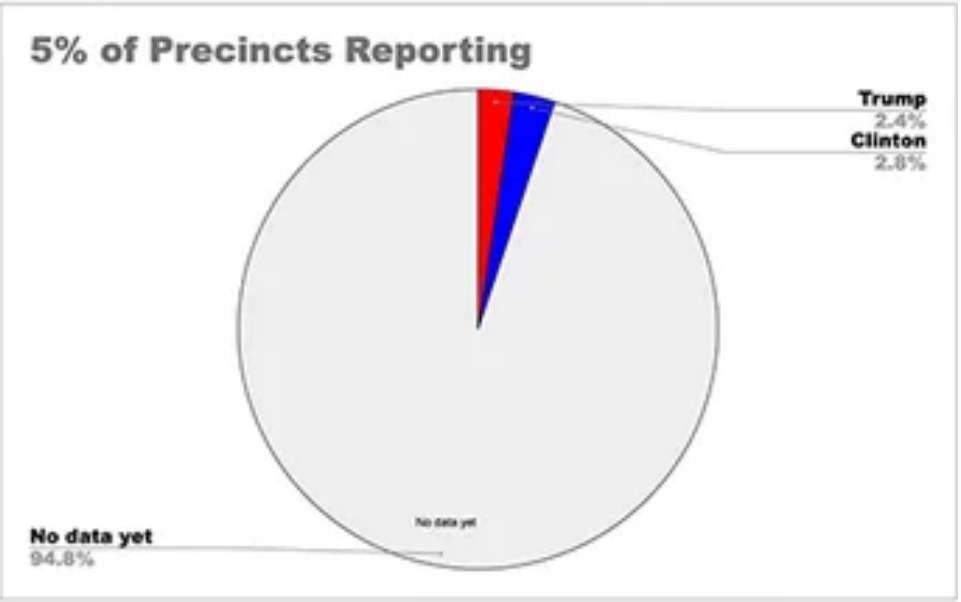

total_sales
100
80
60
40
20
0
Bookseller A
Bookseller B
store

Media, on election night:

Don't use charts like this:



Use charts like THIS:



And include ALL projected mail-in votes in the totals, so if there are lots absentee/mail-in ballots that haven't been counted, they SHOW UP as not counted yet!

*Data from CNN 2016 Presidential election results by county

FOCUS

8 5 3 4 3 9 3 5 3 7 2 9 1 2

7 2 8 3 5 4 6 7 3 4 9 6 5 4

3 4 9 8 2 9 6 8 5 1 1 9 2 1

2 3 5 8 2 4 7 8 9 3 4 4 6 9

3 9 2 5 4 6 7 2 6 8 9 8 7 3

8 5 3 4 3 9 3 5 3 7 2 9 1 2

7 2 8 3 5 4 6 7 3 4 9 6 5 4

3 4 9 8 2 9 6 8 5 1 1 9 2 1

2 3 5 8 2 4 7 8 9 3 4 4 6 9

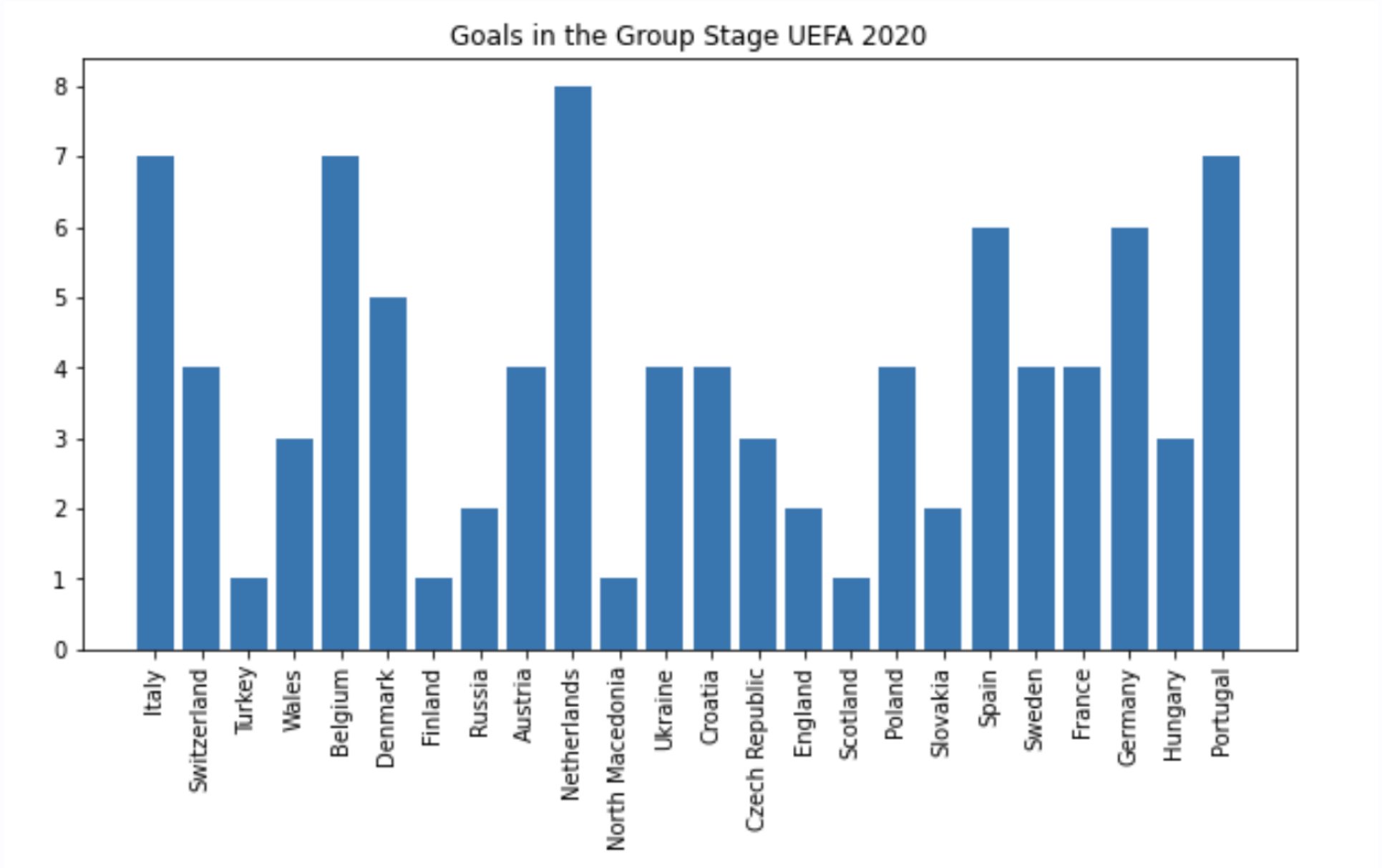
3 9 2 5 4 6 7 2 6 8 9 8 7 3

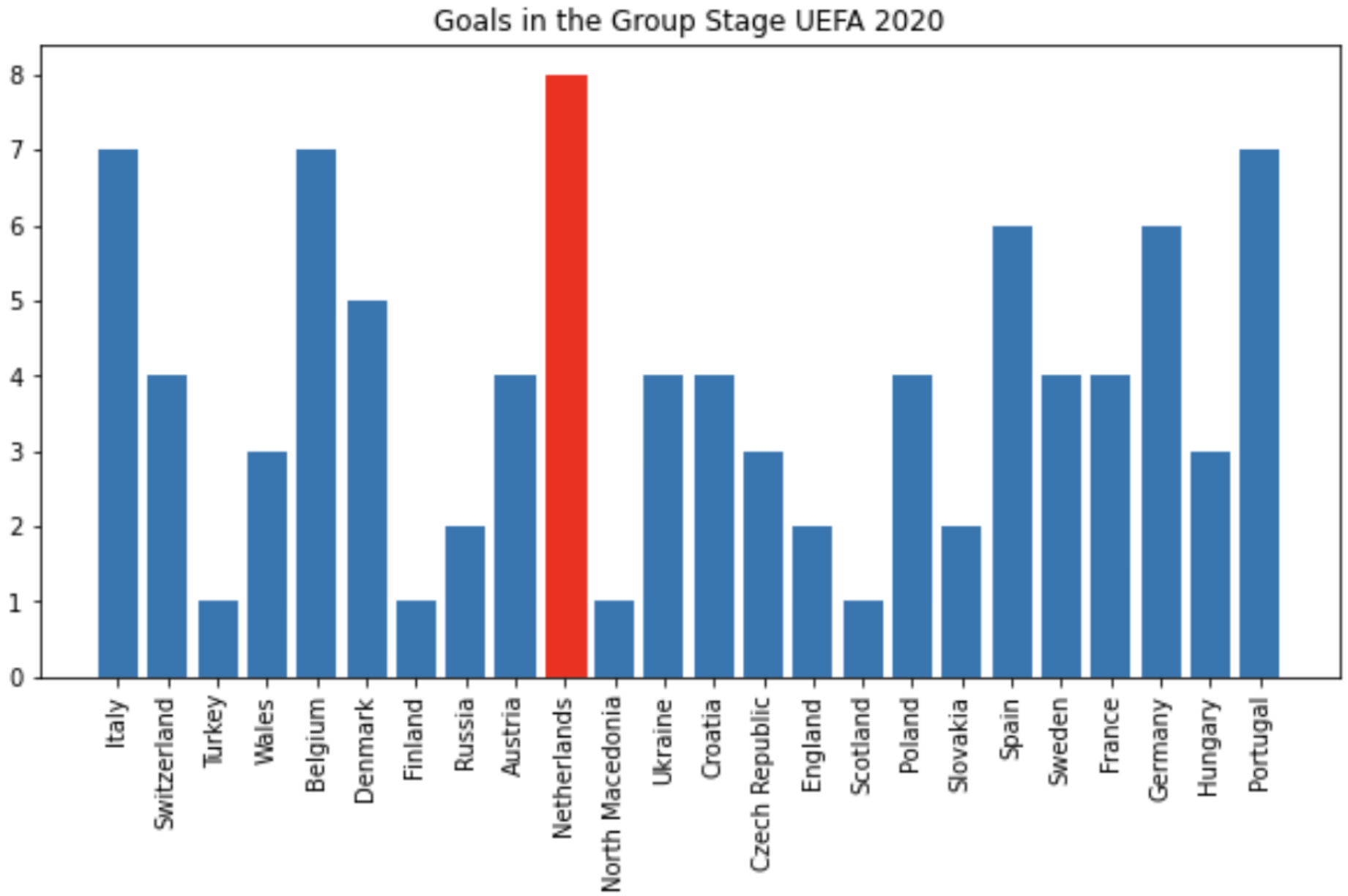
Find the 8s

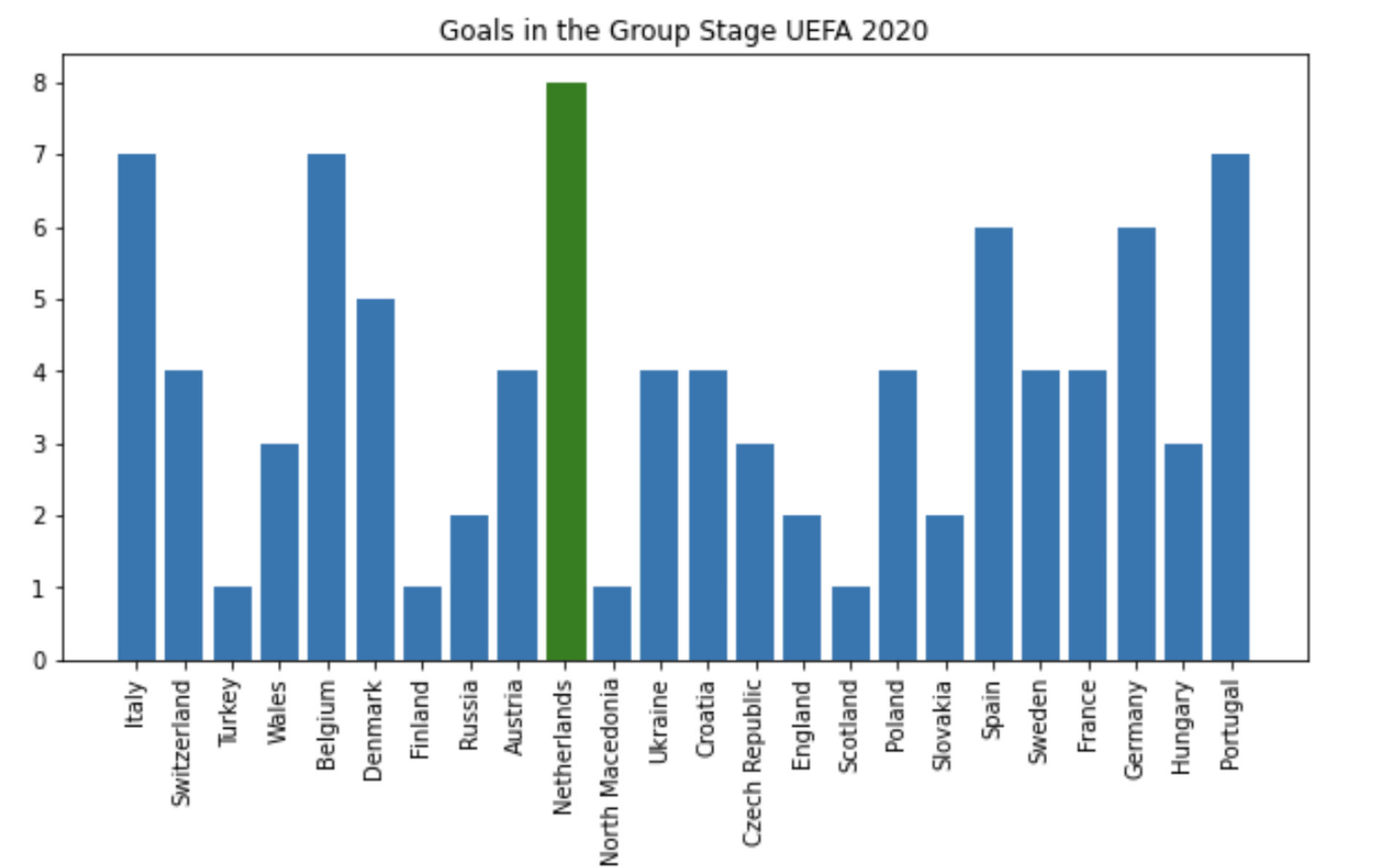
9 9 9 9 9 9 9 9 8 9 9 9 9
9 9 9 9 8 9 9 9 9 9 9 8 9
9 9 8 9 9 9 9 8 9 9 9 9 9

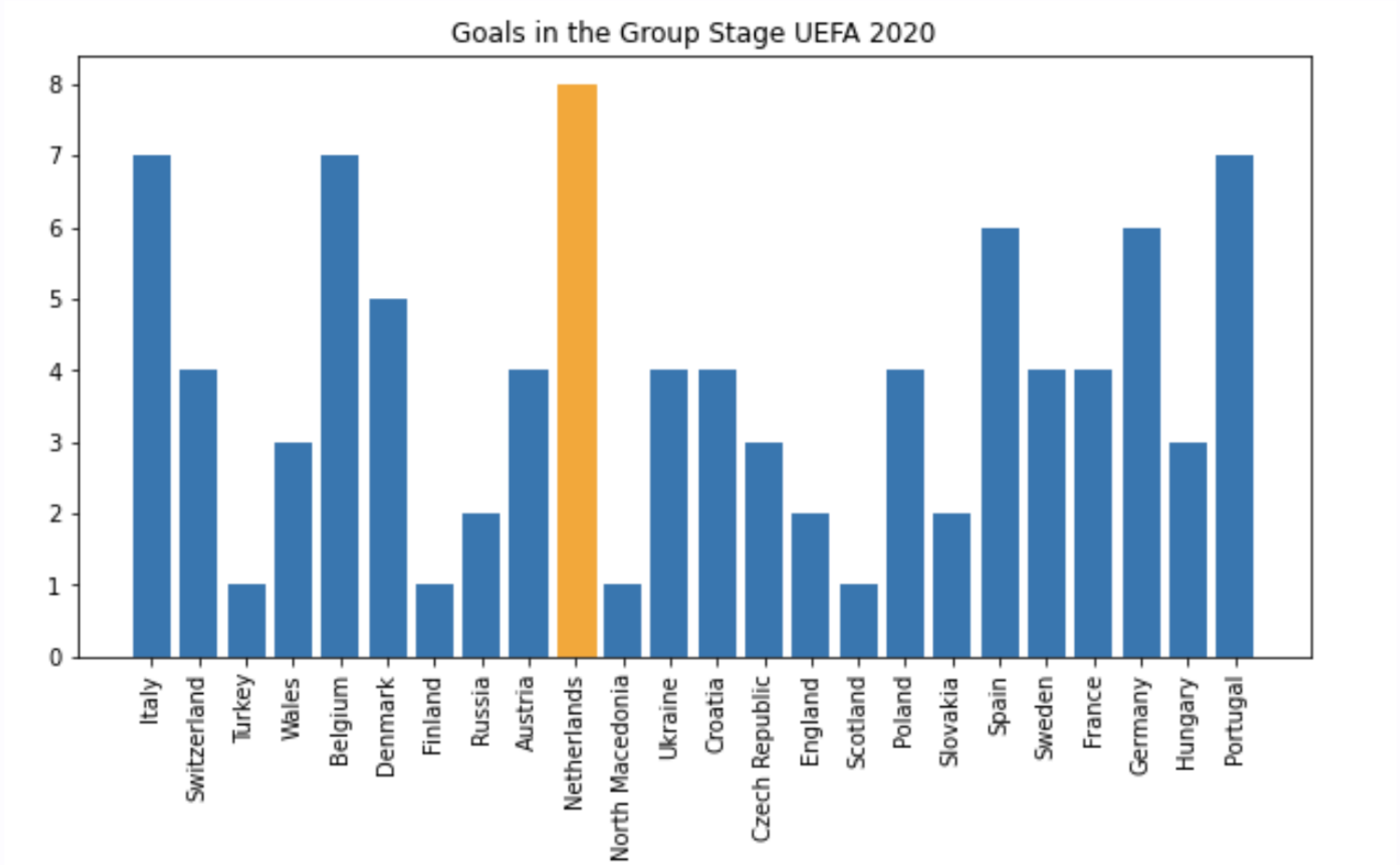
Find the 1s

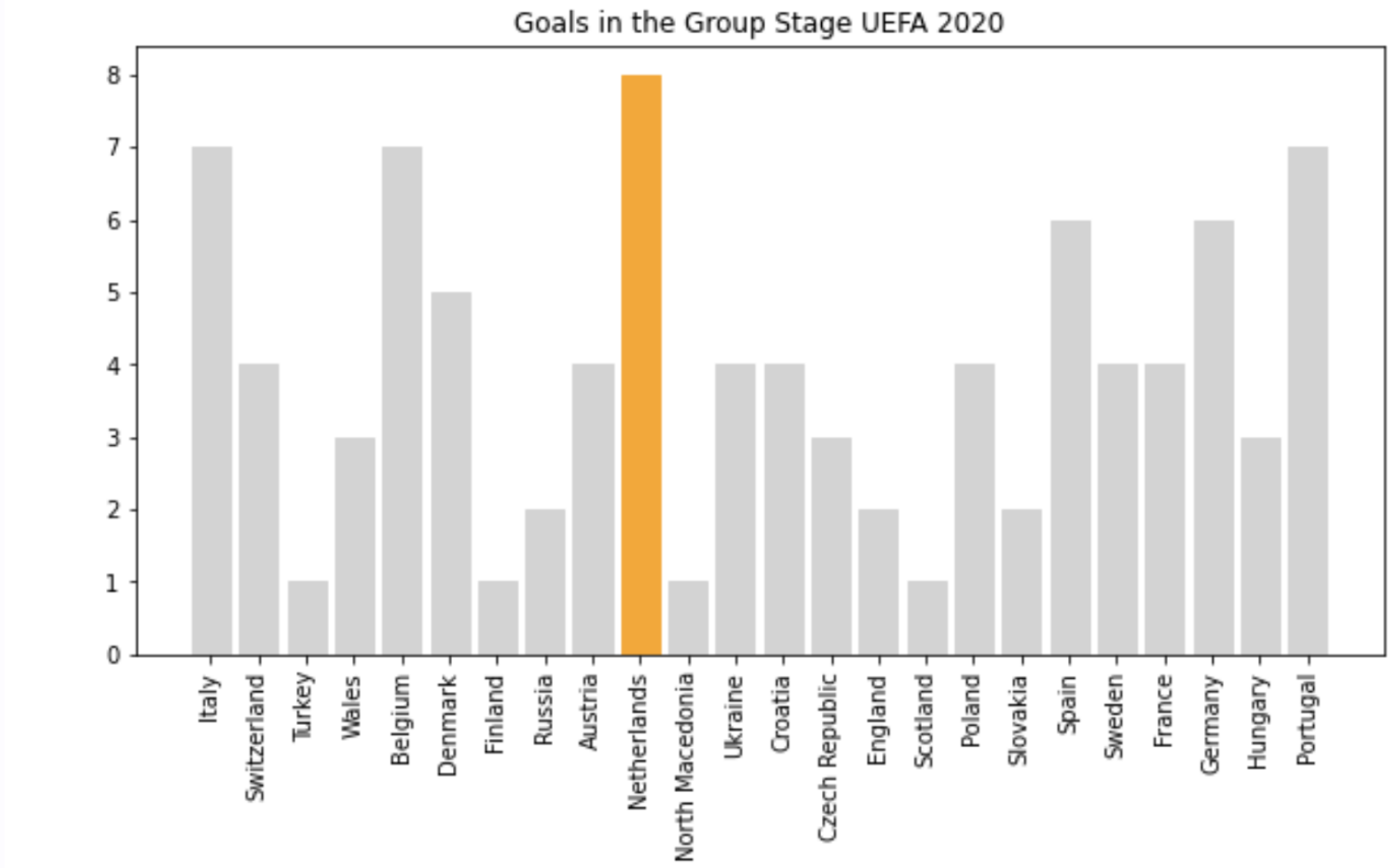
9 9 9 9 9 9 9 9 1 9 9 9 9
9 9 9 9 1 9 9 9 9 9 9 1 9
9 9 1 9 9 9 9 1 9 9 9 9 9

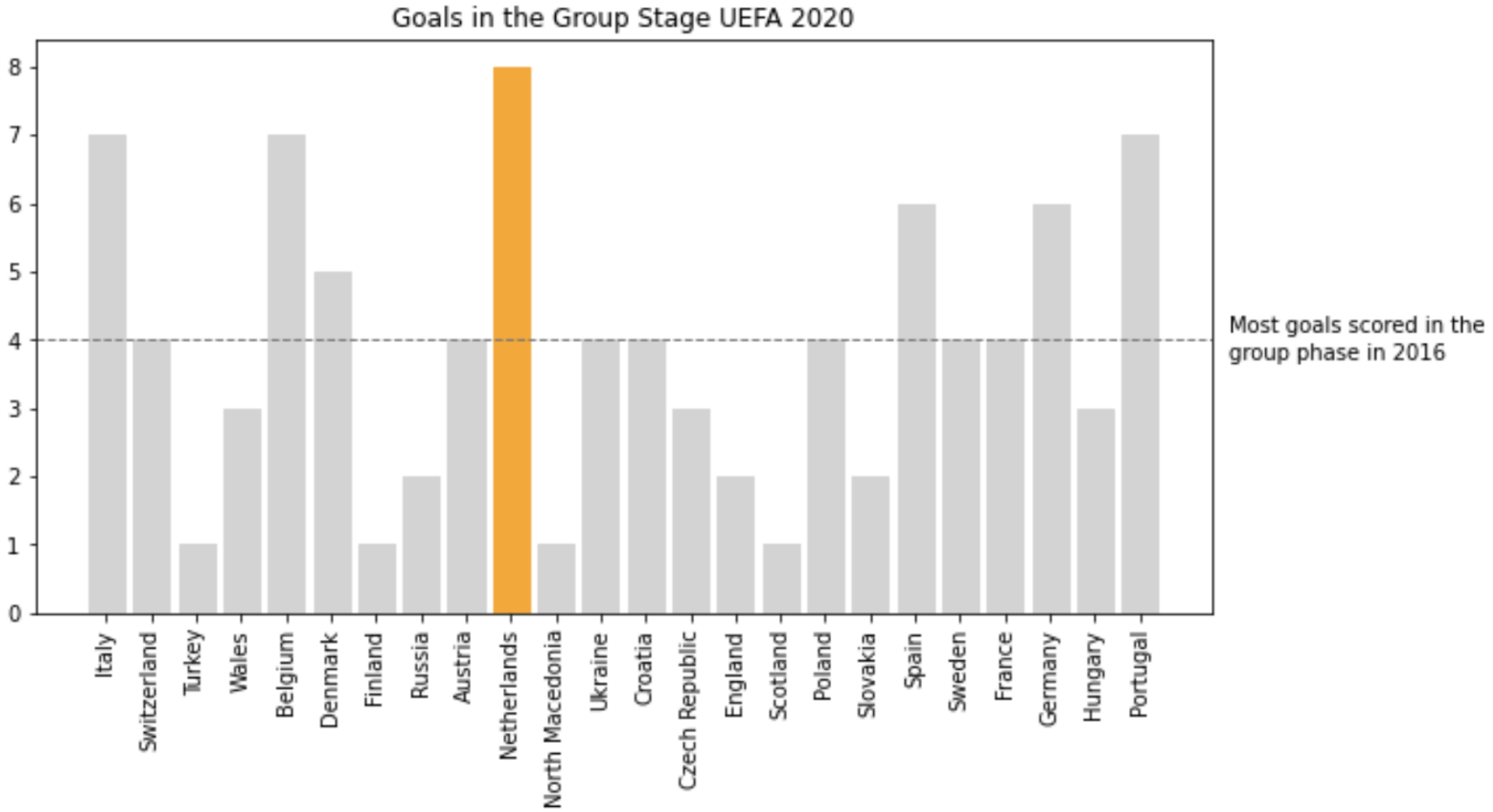

Goals in the Group Stage UEFA 2020
0
1
2
3
4
5
6
7
8
Italy
Switzerland
Turkey
Wales
Belgium
Denmark
Finland
Russia
Austria
Netherlands
North Macedonia
Ukraine
Croatia
Czech Republic
England
Scotland
Poland
Slovakia
Spain
Sweden
France
Germany
Hungary
Portugal

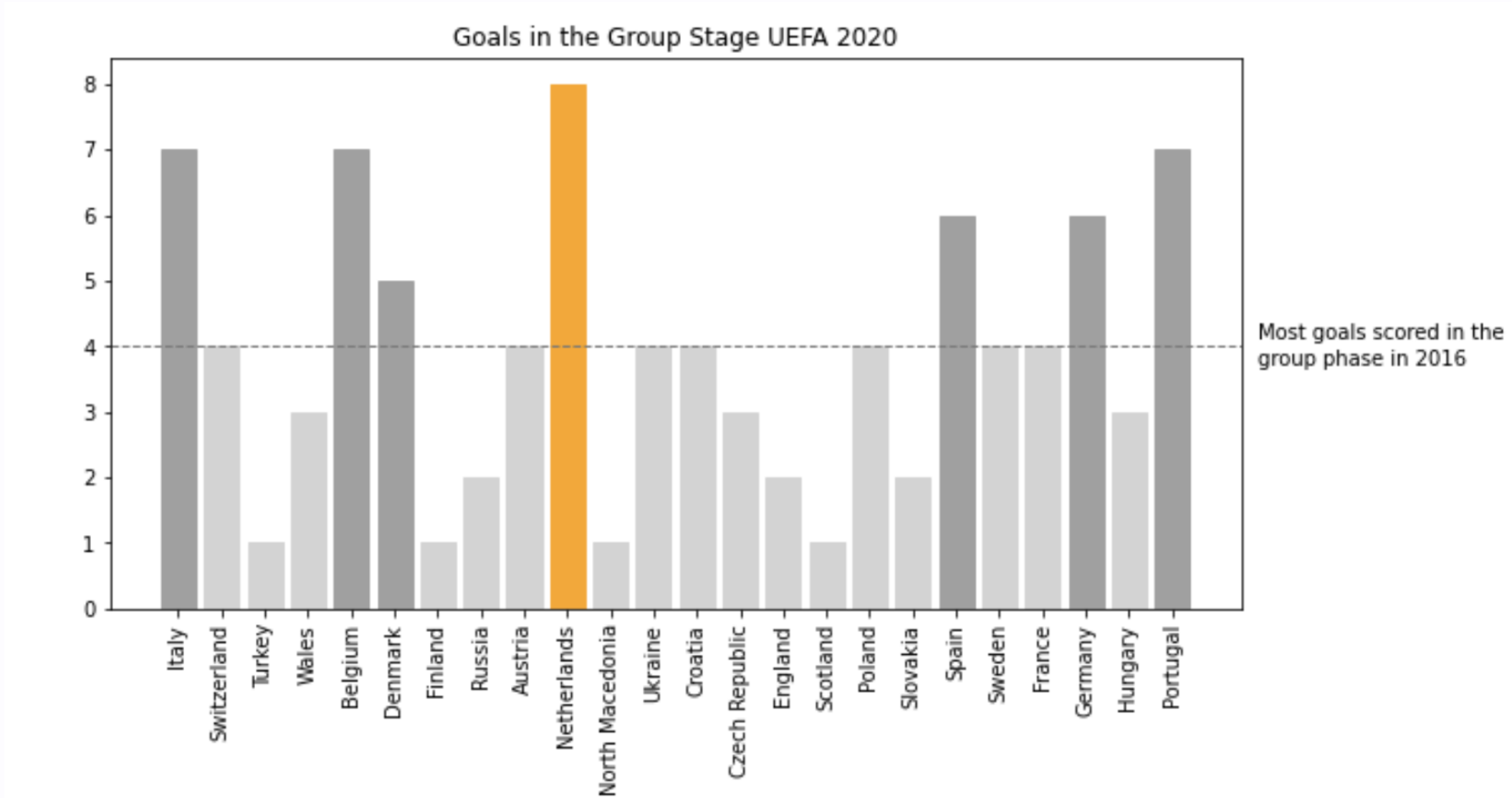

Goals in the Group Stage UEFA 2020
0
1
2
3
4
5
6
7
8
Italy
Switzerland
Turkey
Wales
Belgium
Denmark
Finland
Russia
Austria
Netherlands
North Macedonia
Ukraine
Croatia
Czech Republic
England
Scotland
Poland
Slovakia
Spain
Sweden
France
Germany
Hungary
Portugal

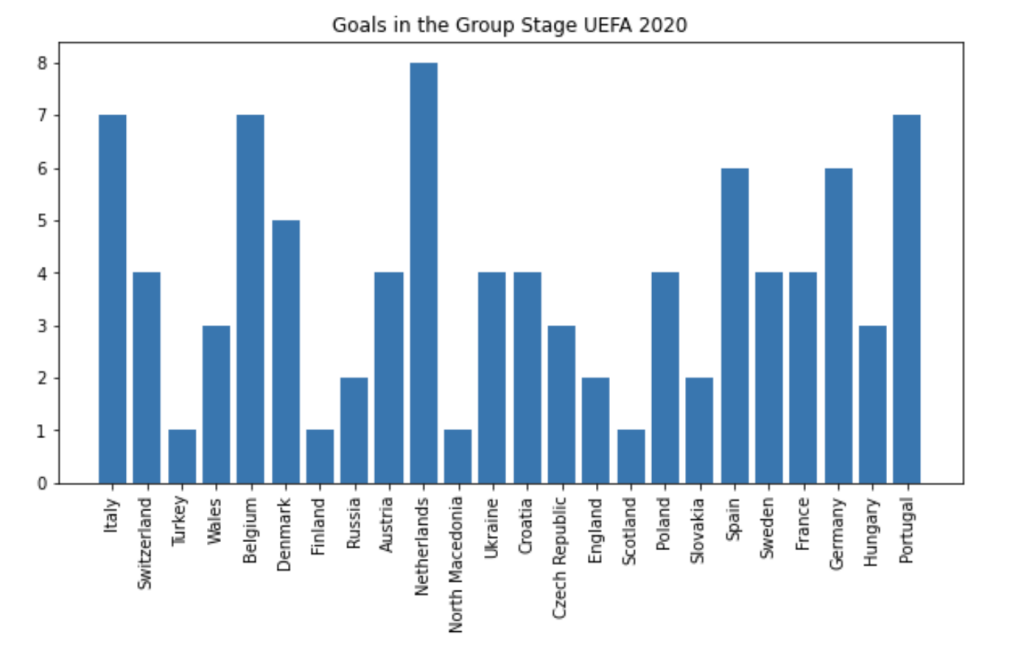

Goals in the Group Stage UEFA 2020
0
1
2
3
4
5
6
7
8
Italy
Switzerland
Turkey
Wales
Belgium
Denmark
Finland
Russia
Austria
Netherlands
North Macedonia
Ukraine
Croatia
Czech Republic
England
Scotland
Poland
Slovakia
Spain
Sweden
France
Germany
Hungary
Portugal

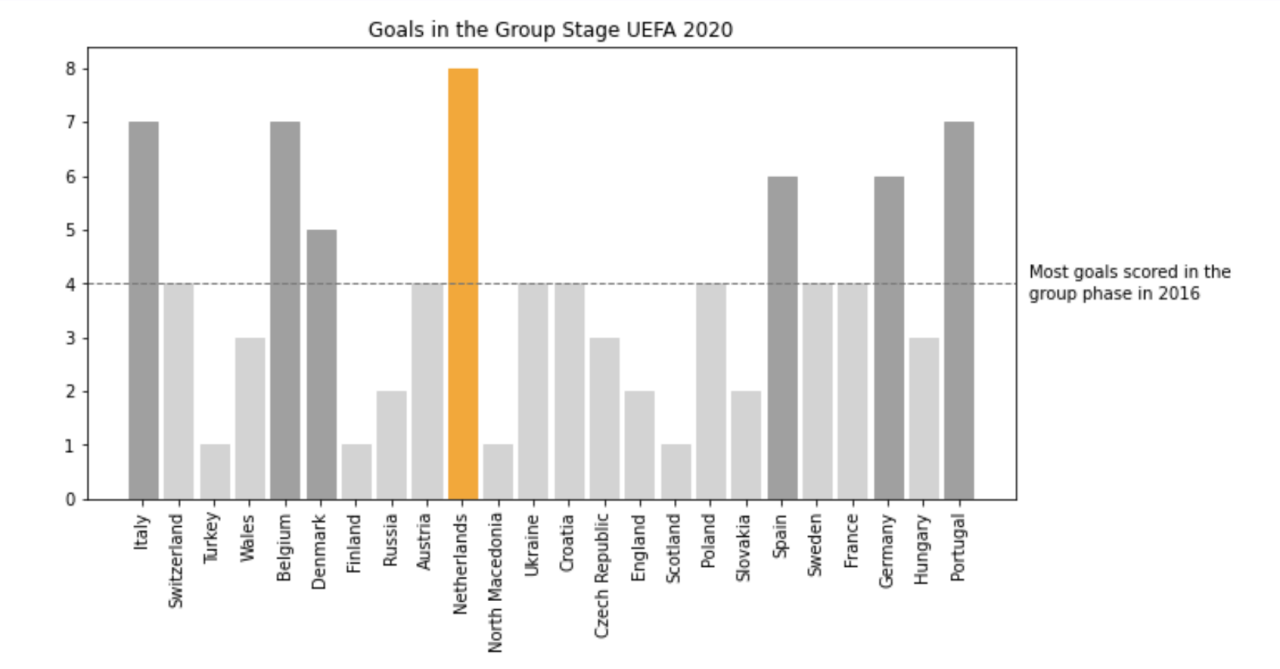

Goals in the Group Stage UEFA 2020
0
1
2
3
4
5
6
7
8
Italy
Switzerland
Turkey
Wales
Belgium
Denmark
Finland
Russia
Austria
Netherlands
North Macedonia
Ukraine
Croatia
Czech Republic
England
Scotland
Poland
Slovakia
Spain
Sweden
France
Germany
Hungary
Portugal

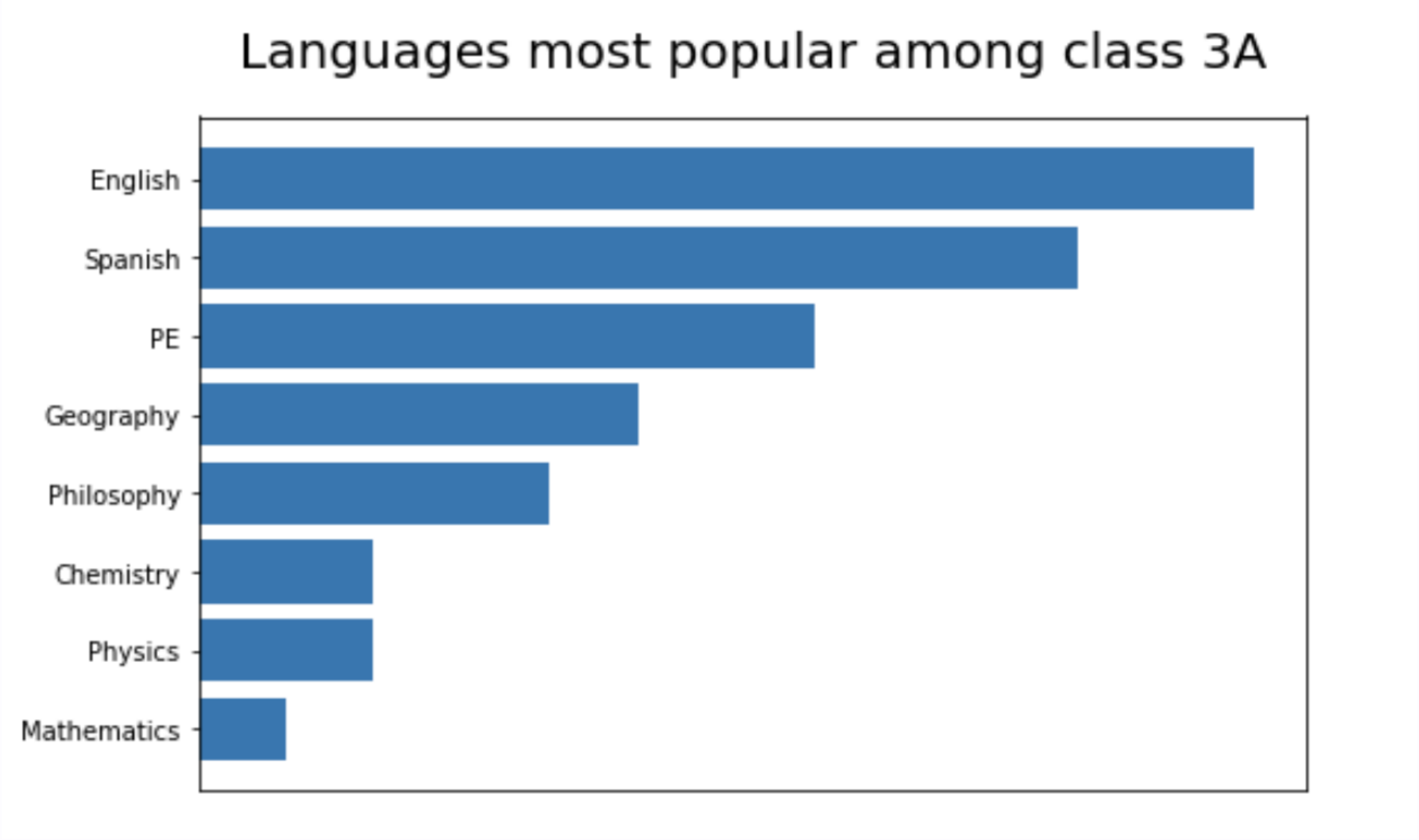

Goals in the Group Stage UEFA 2020
0
1
2
3
4
5
6
7
8
Italy
Switzerland
Turkey
Wales
Belgium
Denmark
Finland
Russia
Austria
Netherlands
North Macedonia
Ukraine
Croatia
Czech Republic
England
Scotland
Poland
Slovakia
Spain
Sweden
France
Germany
Hungary
Portugal

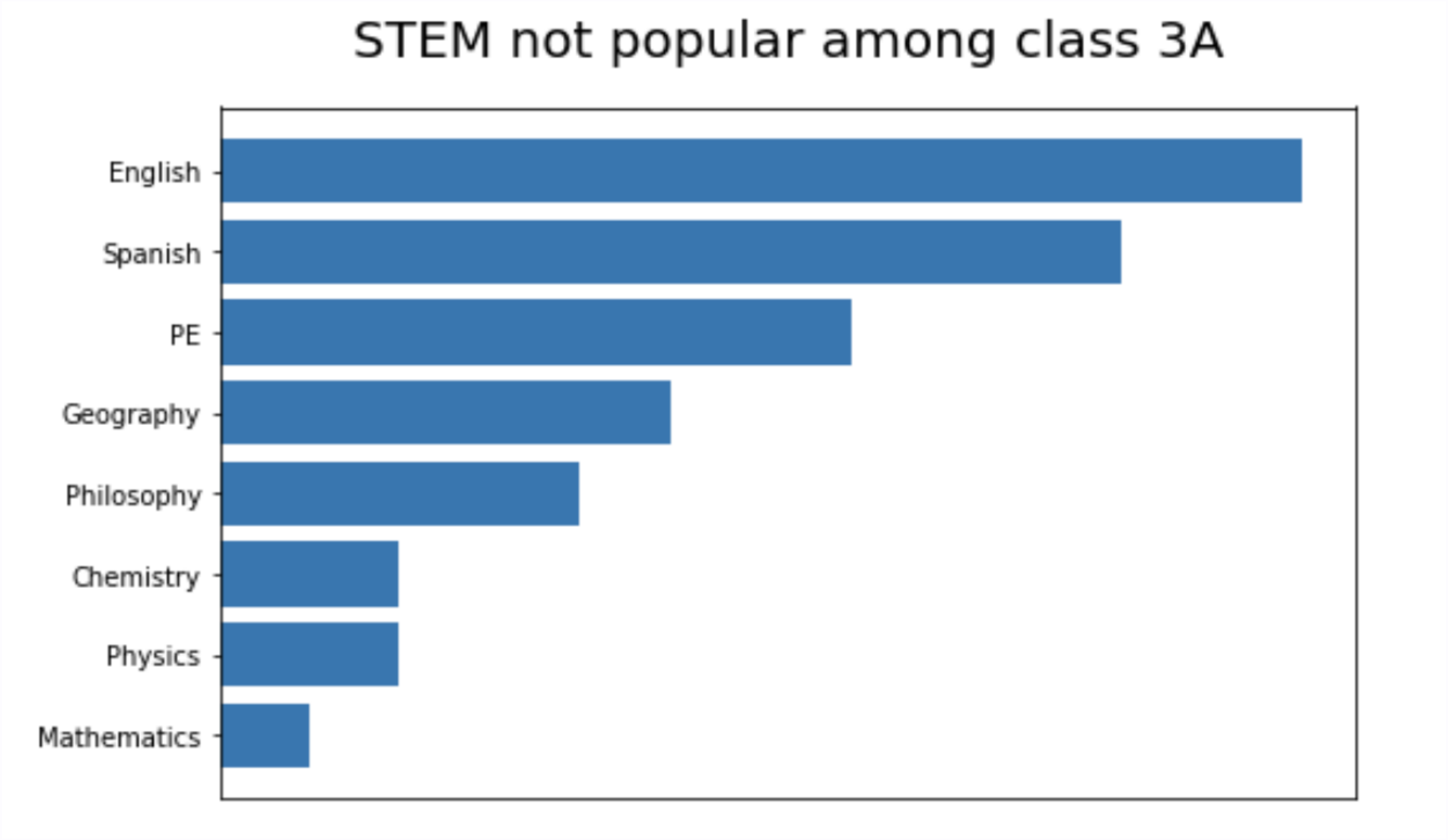

Goals in the Group Stage UEFA 2020
0
1
2
3
4
5
6
7
8
Italy
Switzerland
Turkey
Wales
Belgium
Denmark
Finland
Russia
Austria
Netherlands
North Macedonia
Ukraine
Croatia
Czech Republic
England
Scotland
Poland
Slovakia
Spain
Sweden
France
Germany
Hungary
Portugal
Most goals scored in the group phase in 2016

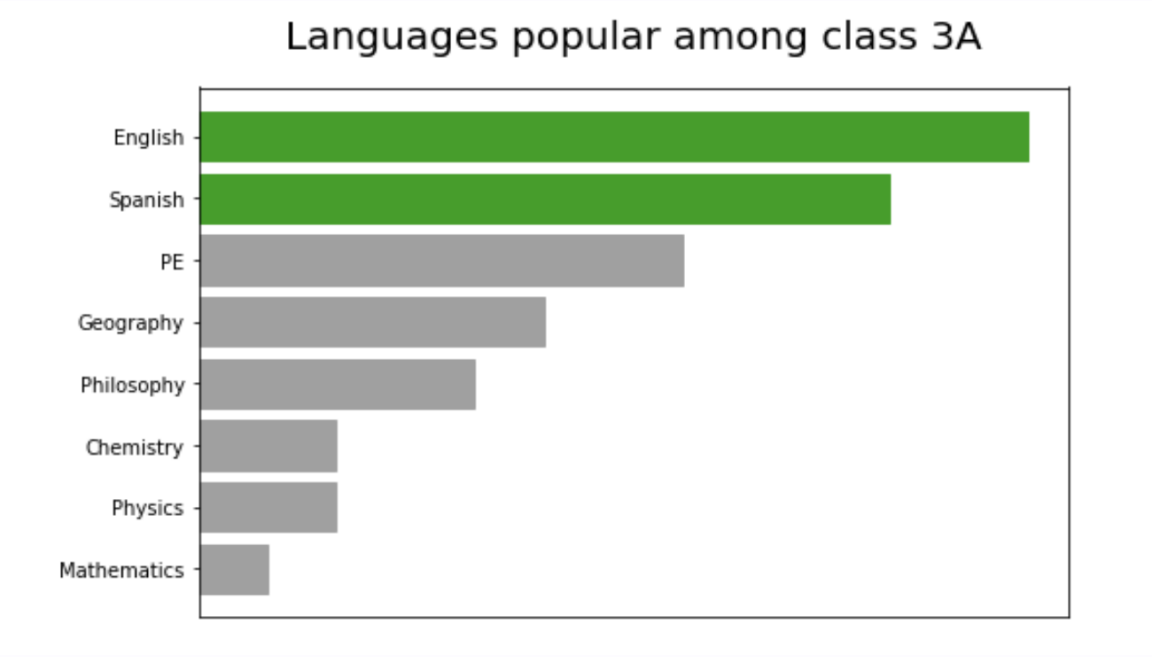

Goals in the Group Stage UEFA 2020
0
1
2
3
4
5
6
7
8
Italy
Switzerland
Turkey
Wales
Belgium
Denmark
Finland
Russia
Austria
Netherlands
North Macedonia
Ukraine
Croatia
Czech Republic
England
Scotland
Poland
Slovakia
Spain
Sweden
France
Germany
Hungary
Portugal
Most goals scored in the group phase in 2016

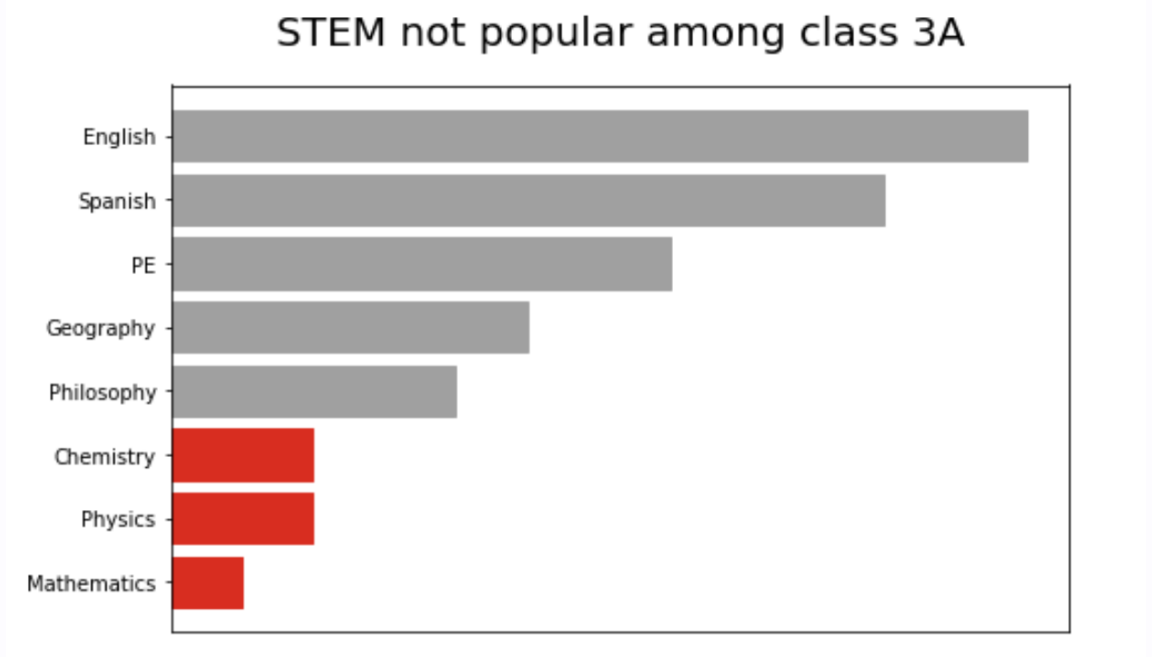

Goals in the Group Stage UEFA 2020
0
1
2
3
4
5
6
7
8
Italy
Switzerland
Turkey
Wales
Belgium
Denmark
Finland
Russia
Austria
Netherlands
North Macedonia
Ukraine
Croatia
Czech Republic
England
Scotland
Poland
Slovakia
Spain
Sweden
France
Germany
Hungary
Portugal

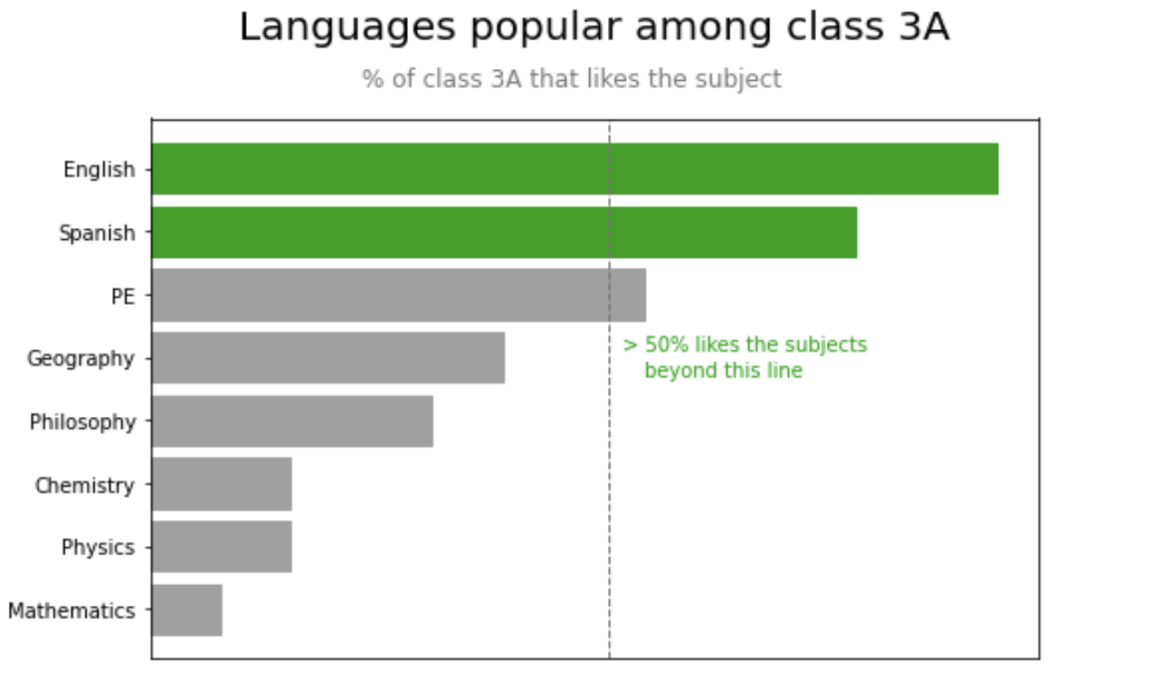

Goals in the Group Stage UEFA 2020
0
1
2
3
4
5
6
7
8
Italy
Switzerland
Turkey
Wales
Belgium
Denmark
Finland
Russia
Austria
Netherlands
North Macedonia
Ukraine
Croatia
Czech Republic
England
Scotland
Poland
Slovakia
Spain
Sweden
France
Germany
Hungary
Portugal
Most goals scored in the group phase in 2016

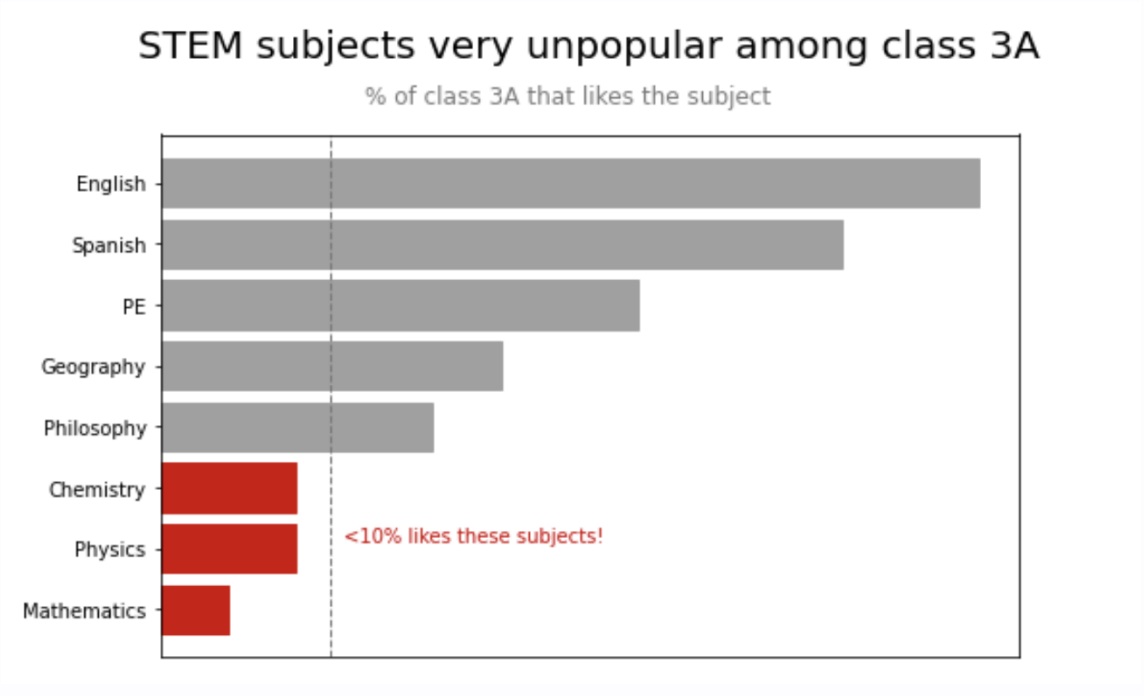

Languages most popular among class 3A
English
Spanish
PE
Geography
Philosophy
Chemistry
Physics
Mathematics


STEM not popular among class 3A
English
Spanish
PE
Geography
Philosophy
Chemistry
Physics
Mathematics


Languages popular among class 3A
English
Spanish
PE
Geography
Philosophy
Chemistry
Physics
Mathematics

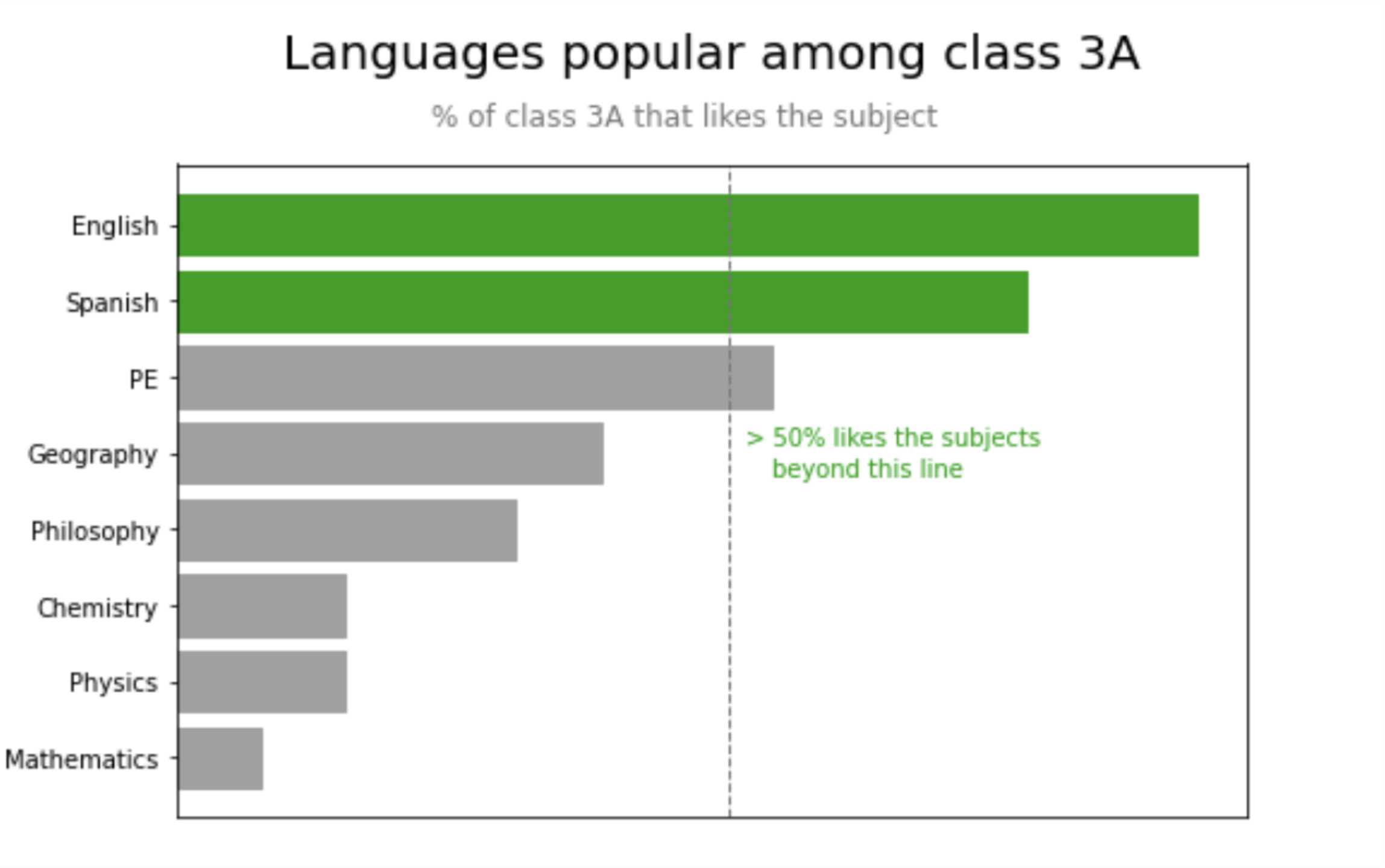

STEM not popular among class 3A
English
Spanish
PE
Geography
Philosophy
Chemistry
Physics
Mathematics


Languages popular among class 3A
% of class 3A that likes the subject
English
Spanish
PE
Geography
Philosophy
Chemistry
Physics
Mathematics
> 50% likes the subjects
beyond this line

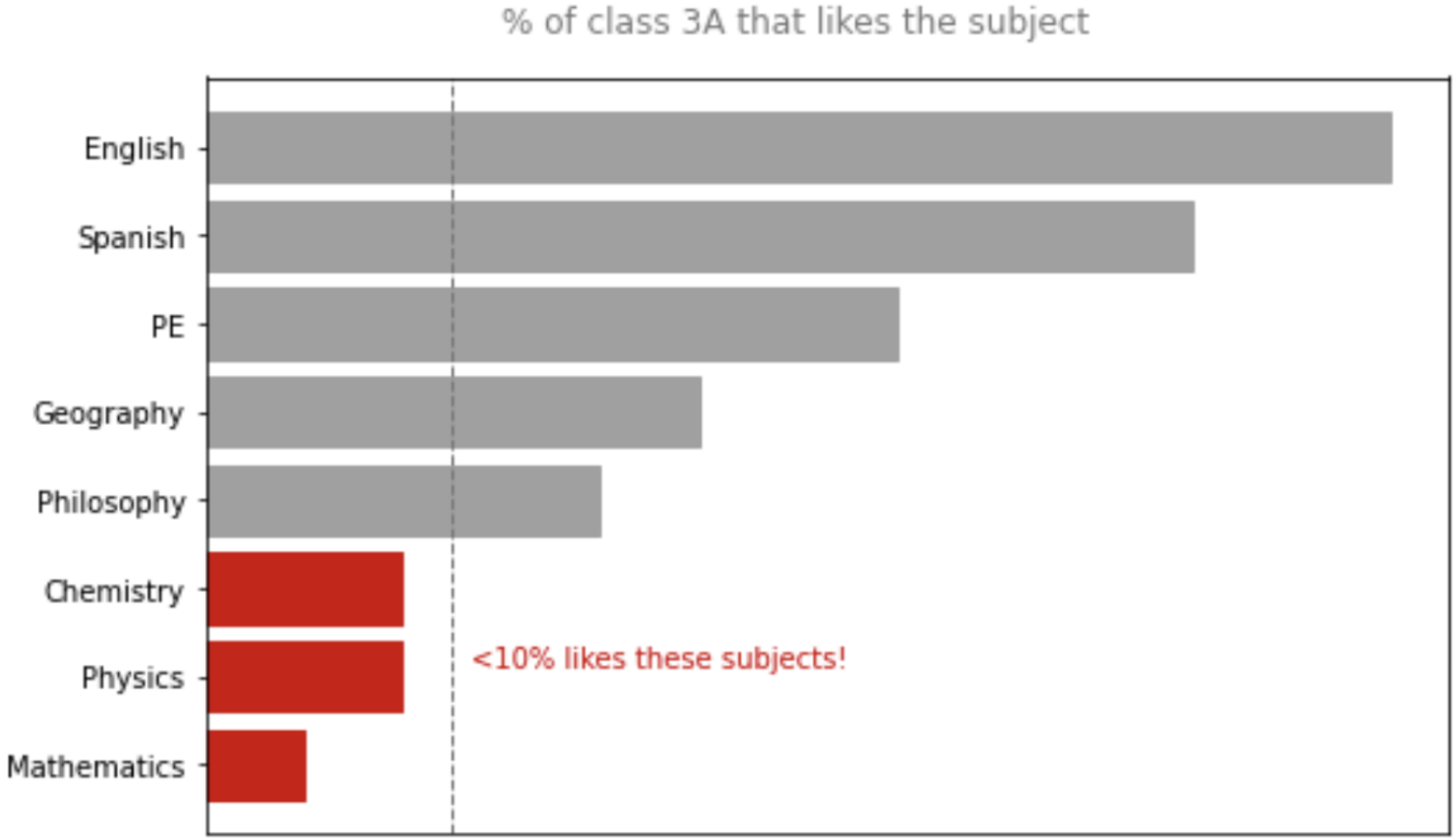

STEM subjects very unpopular among class 3A
% of class 3A that likes the subject
English
Spanish
PE
Geography
Philosophy
Chemistry
Physics
Mathematics
<10% likes these subjects!

FORWARD

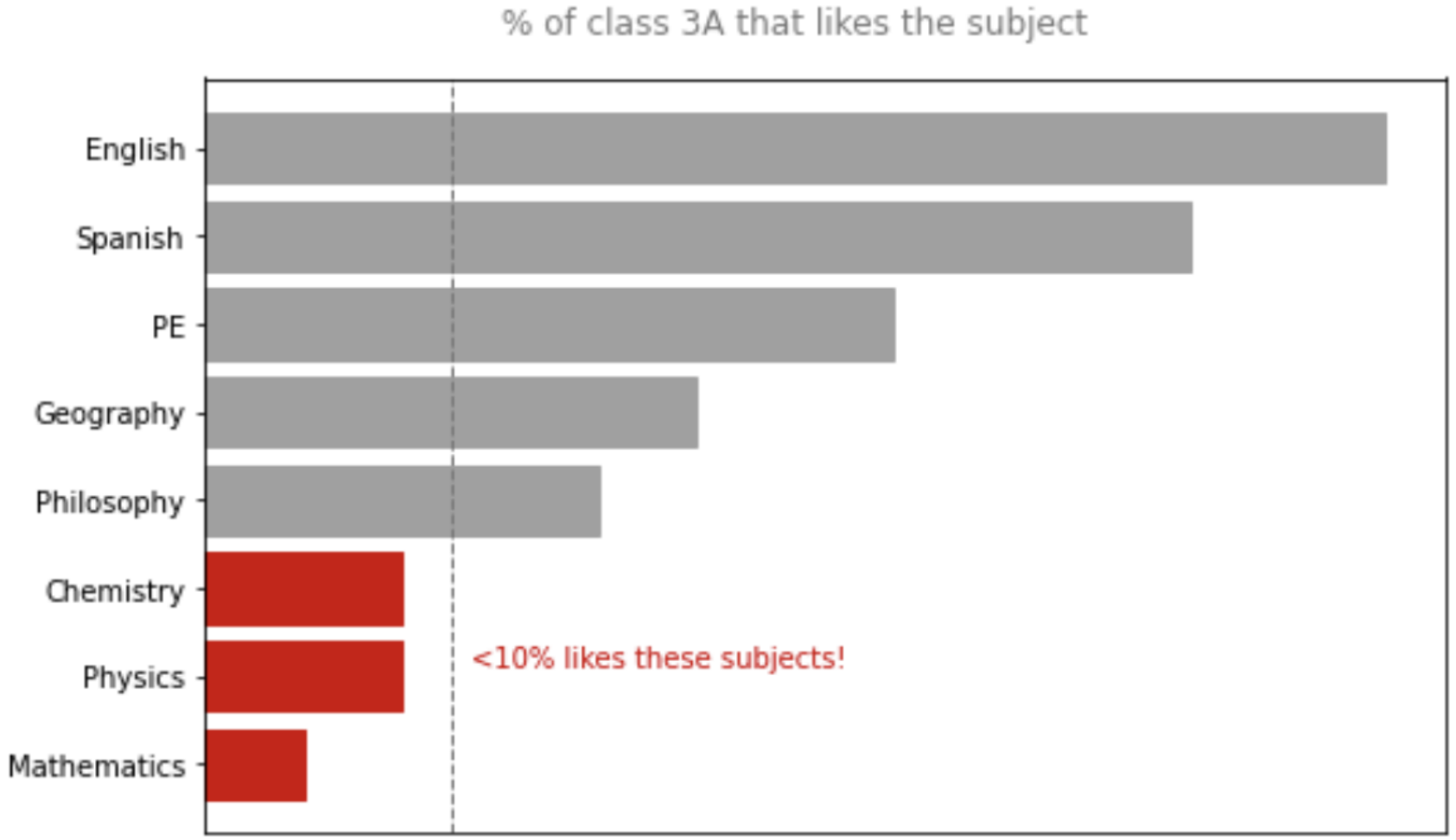
What is the take away from your visualization?


Languages popular among class 3A
% of class 3A that likes the subject
English
Spanish
PE
Geography
Philosophy
Chemistry
Physics
Mathematics
> 50% likes the subjects
beyond this line

STEM subjects very unpopular among class 3A

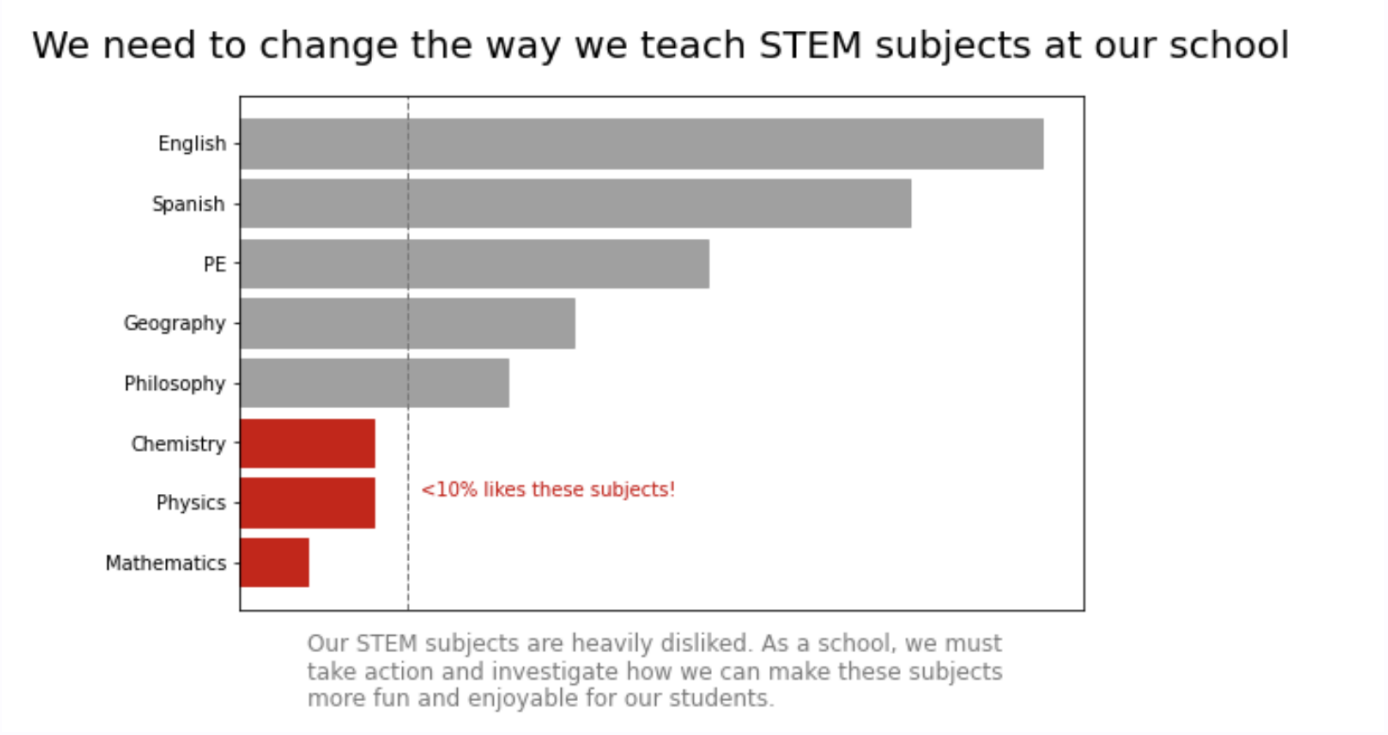


We need to change the way we teach STEM subjects at our school



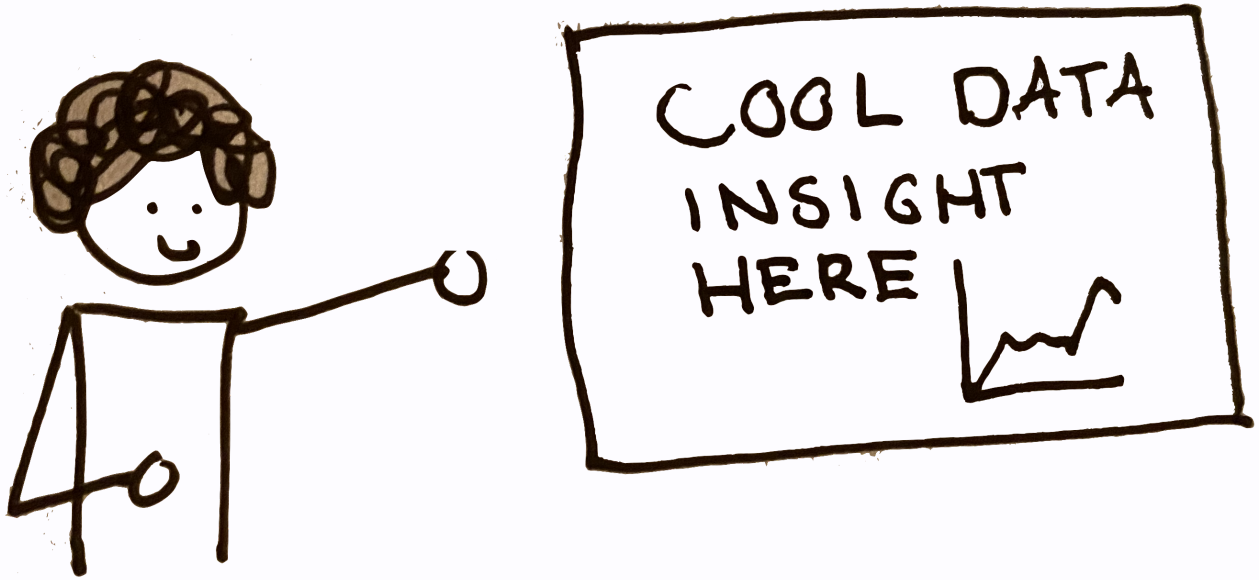
Forward

- Put the message first
- Support the conclusion
- Emphasise the action



Data Storytelling

- Foundation
- Focus
- Forward



Feel free to reach out!

- 🙋‍♀️ Marysia Winkels
- 🌐 marysia.nl
- 📧 hello@marysia.nl