

# Introduction to Data Storytelling

PyLadiesCon 2024

# About me

---

- LLM stuff @ Cohere
- PyData Amsterdam
- Storyteller



# What this workshop is about

- Why data storytelling is important
- How to choose a good **foundation** for presenting your data
- How to draw the **focus** of the reader to the right places
- How you can use a visualization to drive action **forward**

# How this workshop is structured

- Slides and presentation
- Exercises in the repo

# Foundation

# Why do we need data storytelling?

Account	Sales Volume	% Change vs prior	Avg # of UPCs	% ACV Selling	Price per Pound
A	\$15,753	3.60%	1.15	98	\$10.43
B	\$294,164	3.20%	1.75	83	\$15.76
C	\$21,856	-1.20%	1.00	84	\$12.74
D	\$547,265	5.60%	1.10	89	\$9.45
E	\$18,496	-4.70%	1.00	92	\$14.85
F	\$43,986	-2.40%	2.73	92	\$12.86
G	\$86,734	10.60%	1.00	100	\$17.32
H	\$11,645	37.90%	1.00	85	\$11.43
I	\$11,985	-0.70%	1.00	22	\$20.82
J	\$190,473	-8.70%	1.00	72	\$11.24

# Why don't we present the table as is?

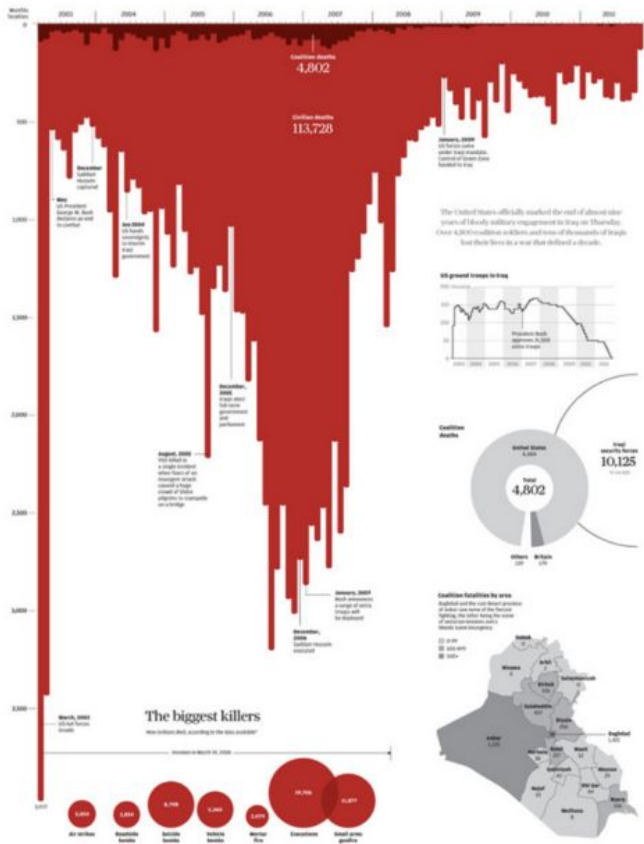
- Information overload
- Distraction
- **You** know is relevant

Account	Sales Volume	% Change vs prior	Avg # of UPCs	% ACV Selling	Price per Pound
A	\$15,753	3.60%	1.15	98	\$10.43
B	\$294,164	3.20%	1.75	83	\$15.76
C	\$21,856	-1.20%	1.00	84	\$12.74
D	\$547,265	5.60%	1.10	89	\$9.45
E	\$18,496	-4.70%	1.00	92	\$14.85
F	\$43,986	-2.40%	2.73	92	\$12.86
G	\$86,734	10.60%	1.00	100	\$17.32
H	\$11,645	37.90%	1.00	85	\$11.43
I	\$11,985	-0.70%	1.00	22	\$20.82
J	\$190,473	-8.70%	1.00	72	\$11.24

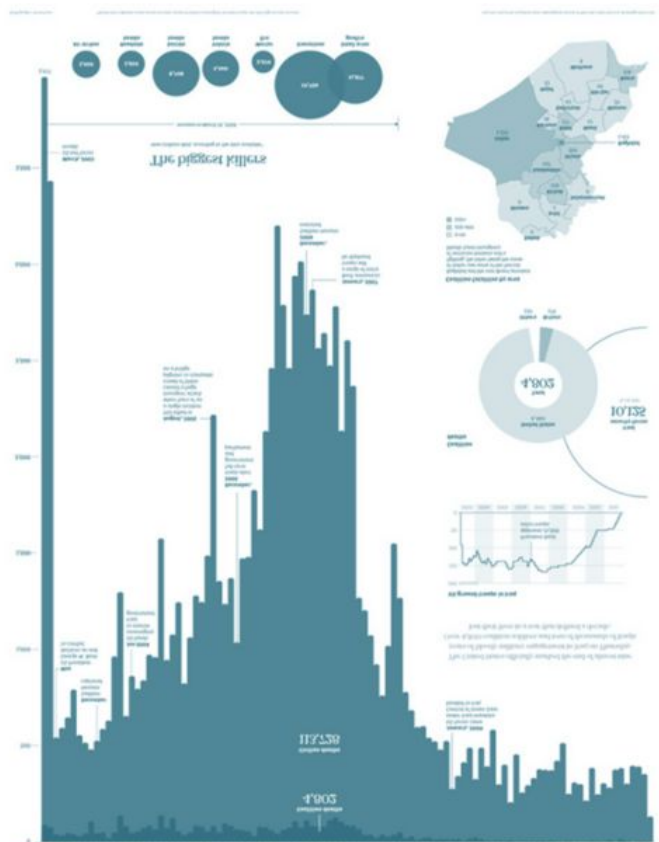




## Iraq's bloody toll



## Iraq: Deaths on the decline



# Two types of data visualization

EXPLORATORY



EXPLANATORY



The best **exploratory visualization** is  
not necessarily the best **explanatory  
visualization**

The way you present your **data** matters

# Example

## New client tier share

Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A	77	7.08%	\$4.68	25%
A+	19	1.75%	\$3.93	21%
B	338	31.07%	\$5.98	32%
C	425	39.06%	\$2.81	15%
D	24	2.21%	\$0.37	2%

## New client tier share

Tier	# of Accounts	% Accounts	Revenue (\$M)	% Revenue
A+	19	2%	\$3.9	21%
A	77	7%	\$4.7	25%
B	338	31%	\$6.0	32%
C	425	39%	\$2.8	15%
D	24	2%	\$0.4	2%
All other	205	19%	\$0.9	5%
TOTAL	1,088	100%	\$18.7	100%

## New client tier share

TIER	ACCOUNTS		REVENUE	
	#	% OF TOT	\$M	% OF TOT
A+	19	2%	\$3.9	21%
A	77	7%	\$4.7	25%
B	338	31%	\$6.0	32%
C	425	39%	\$2.8	15%
D	24	2%	\$0.4	2%
All other	205	19%	\$0.9	5%
TOTAL	1,088	100%	\$18.7	100%

# Exercise

(5 minutes)

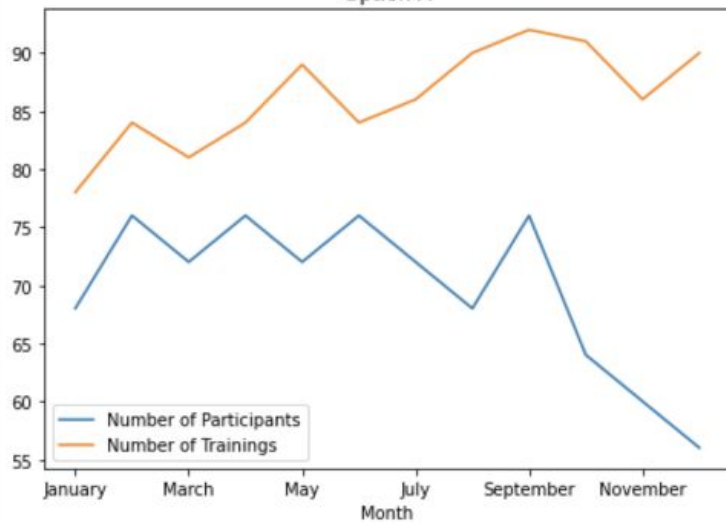
**`exercises/exercise_1.md`**

— — —

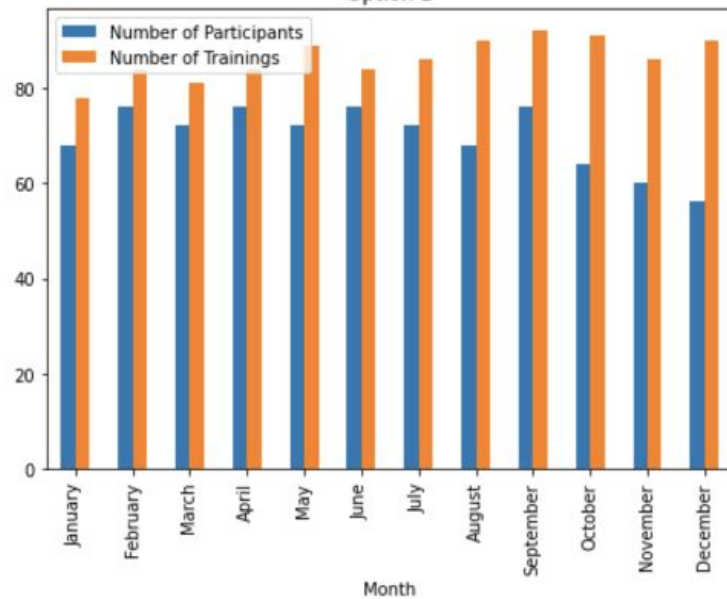


What if you *do* want to visualize the data?

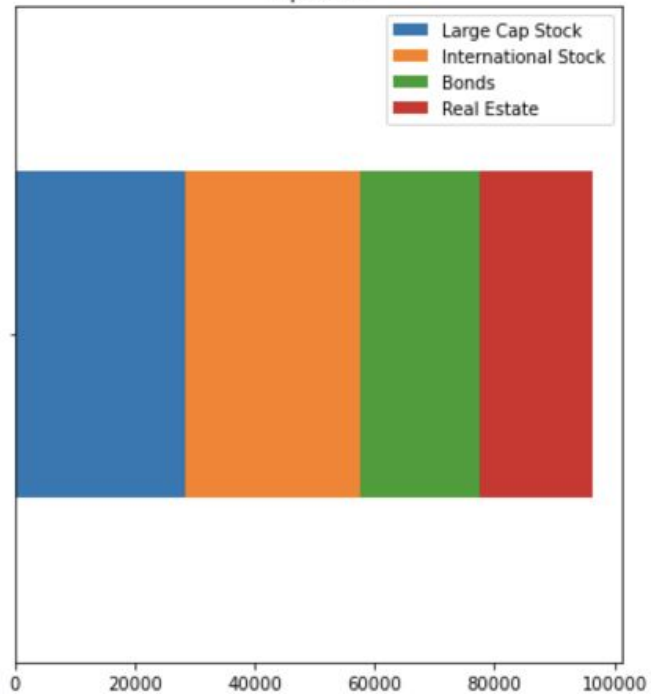
Option A



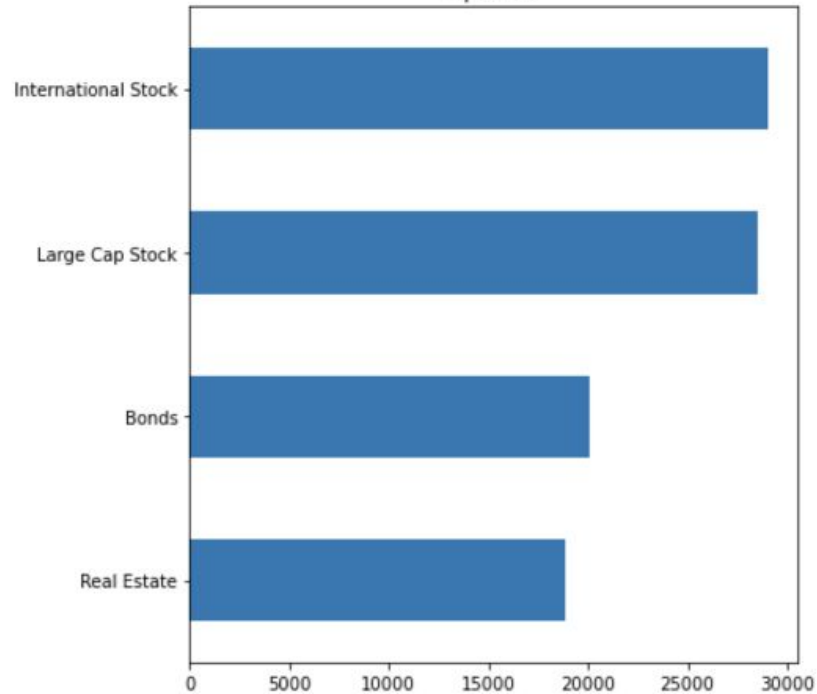
Option B



Option A



Option B

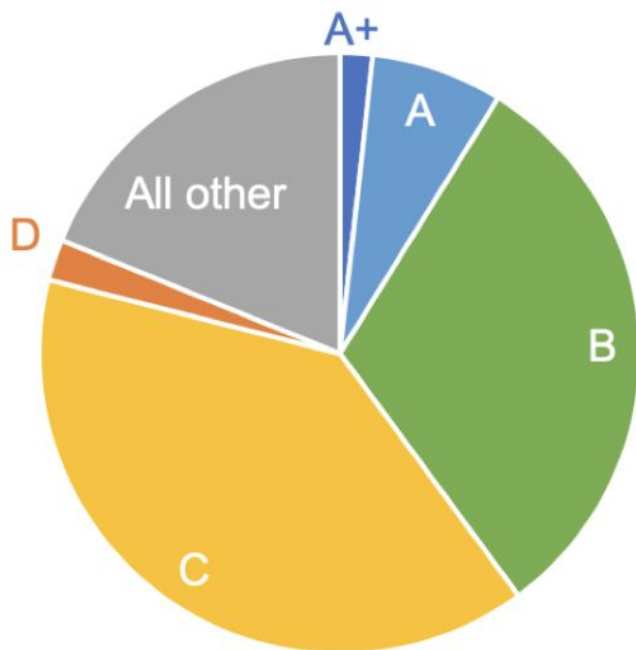


## New client tier share

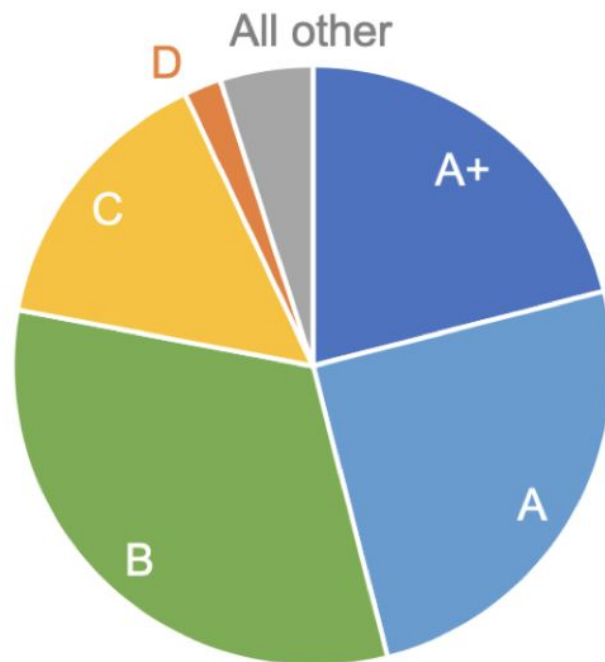
TIER	ACCOUNTS		REVENUE	
	#	% OF TOT	\$M	% OF TOT
A+	19	2%	\$3.9	21%
A	77	7%	\$4.7	25%
B	338	31%	\$6.0	32%
C	425	39%	\$2.8	15%
D	24	2%	\$0.4	2%
All other	205	19%	\$0.9	5%
TOTAL	1,088	100%	\$18.7	100%

# New client tier share

% of Total **Accounts**



% of Total **Revenue**



Before deciding on a specific chart type, **explore your options.**

# Exercise

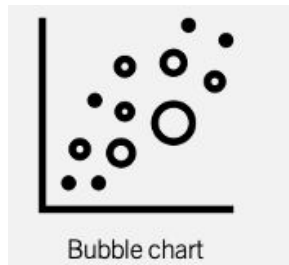
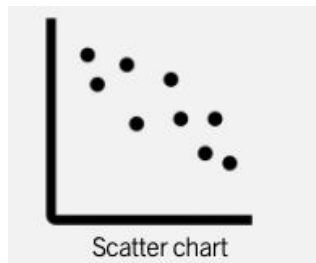
(10 minutes)

`exercises/exercise_2.md`

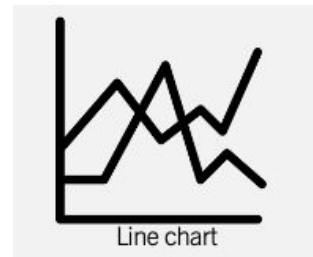
---

# How to choose the right graph

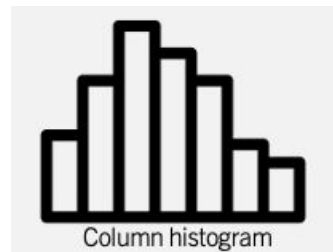
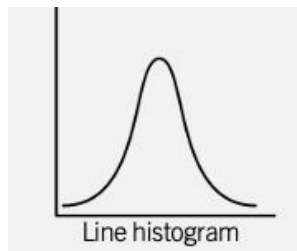
## Relationship



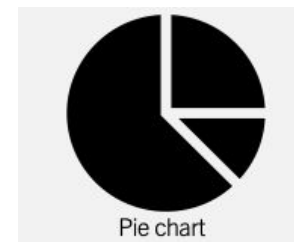
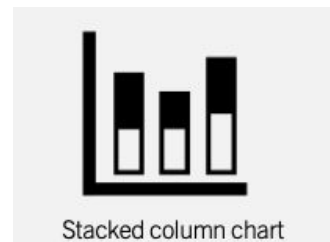
## Comparison



## Distribution



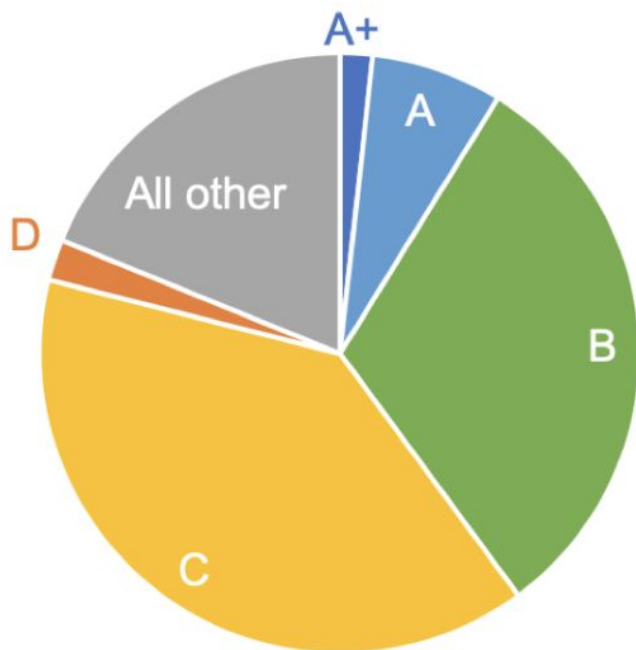
## Composition



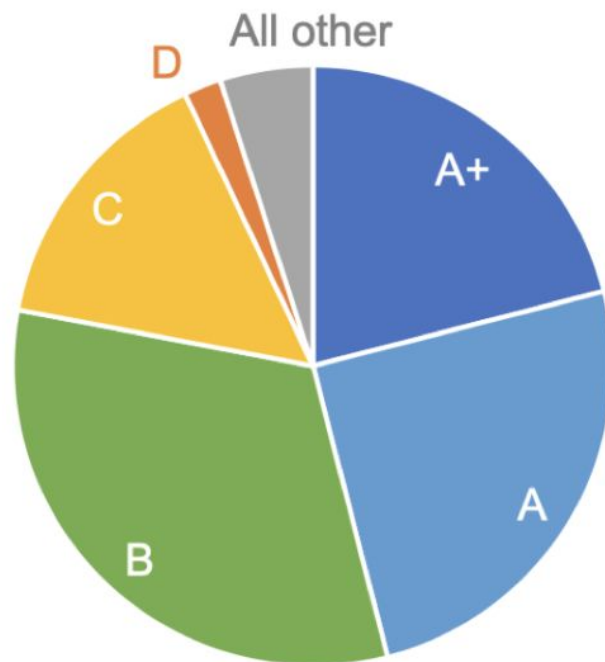


# New client tier share

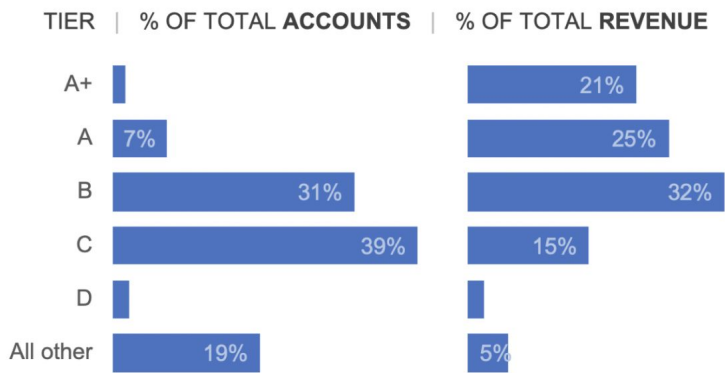
% of Total **Accounts**



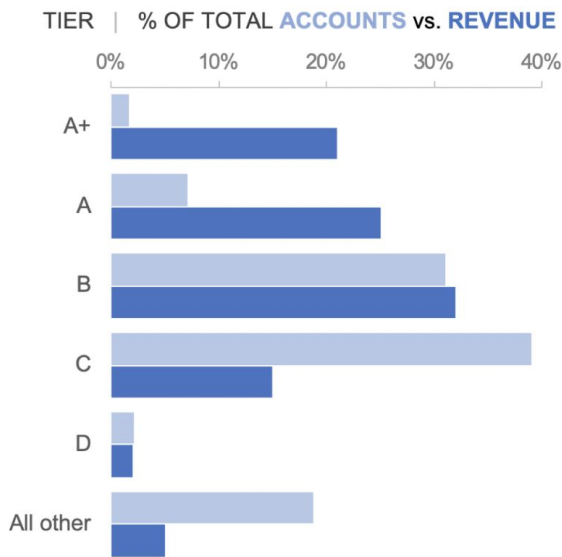
% of Total **Revenue**



## New client tier share

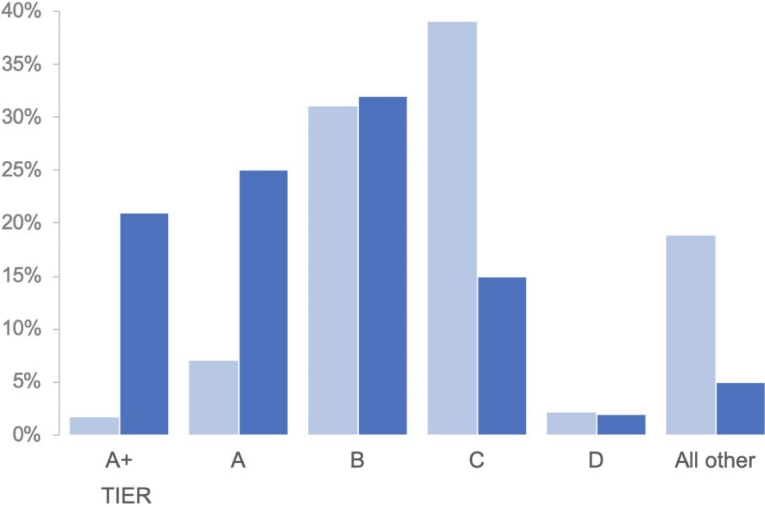


## New client tier share



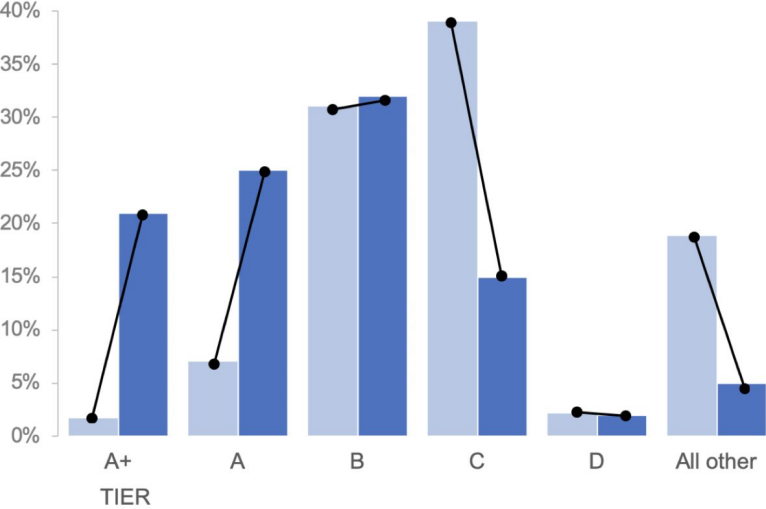
# New client tier share

% OF TOTAL ACCOUNTS vs. REVENUE



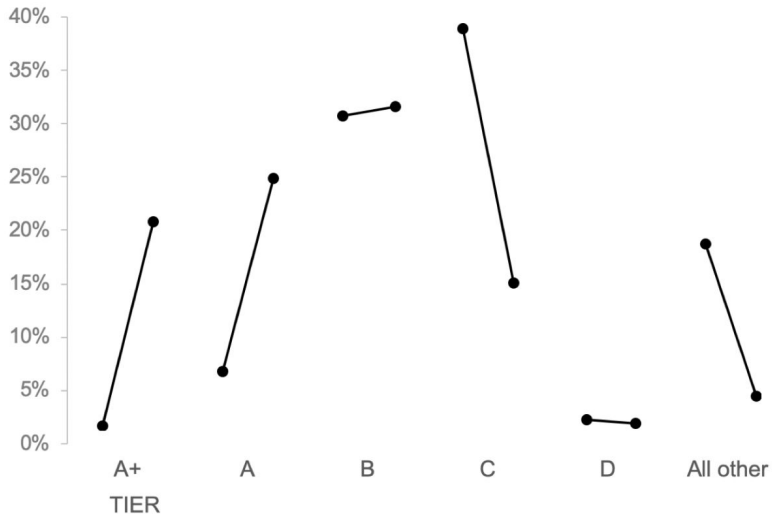
# New client tier share

% OF TOTAL ACCOUNTS vs. REVENUE

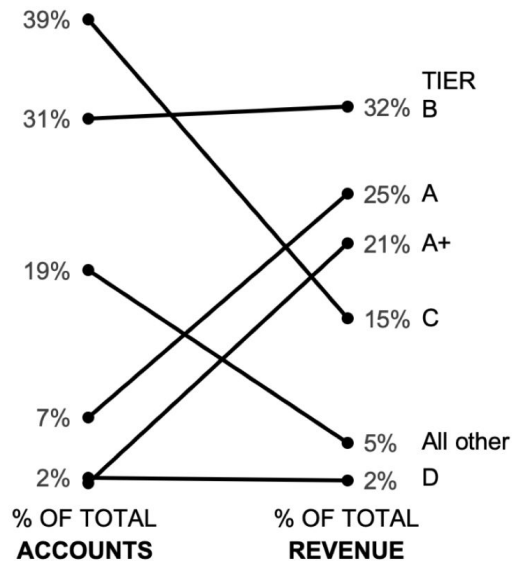


## New client tier share

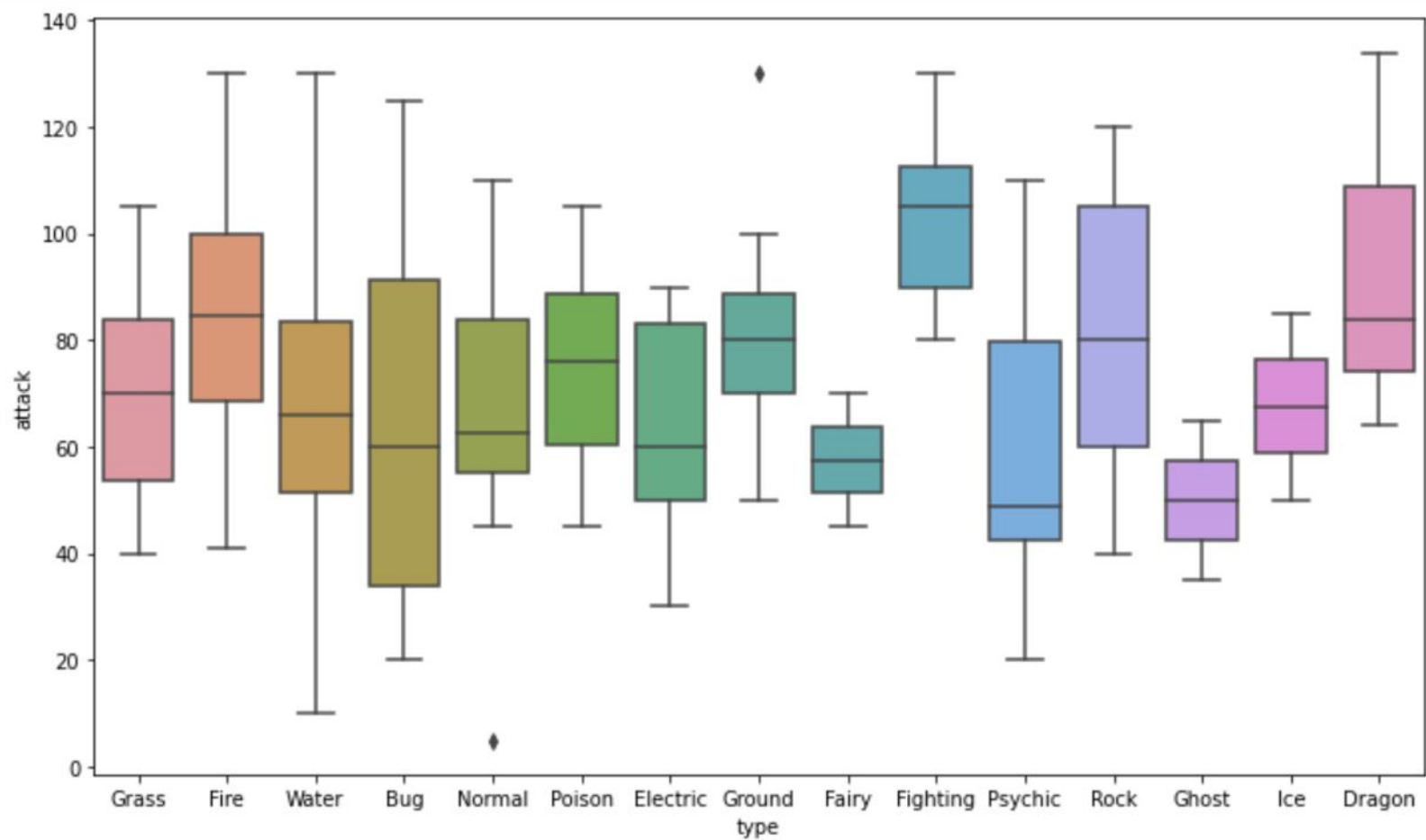
% OF TOTAL **ACCOUNTS** vs. **REVENUE**

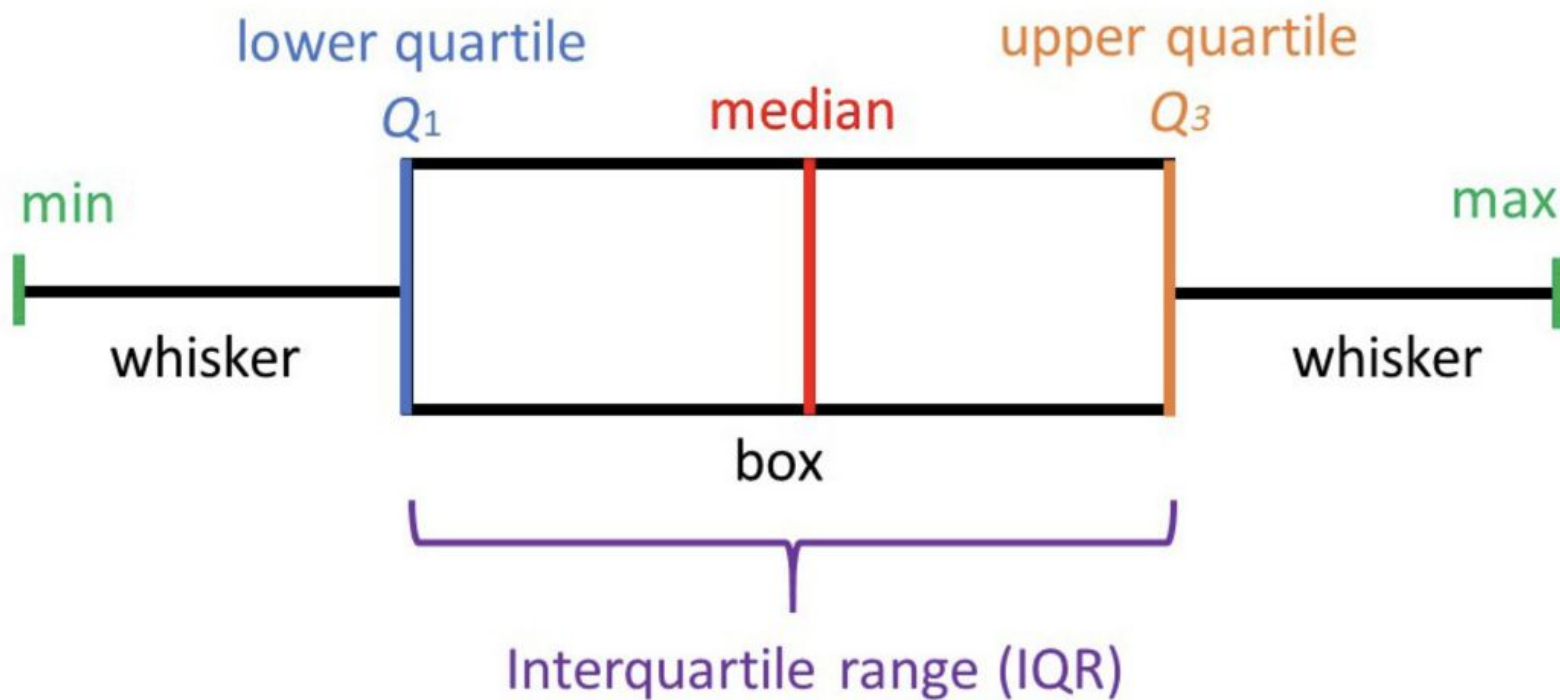


## New client tier share



Make sure the reader is **familiar** with  
your type of chart

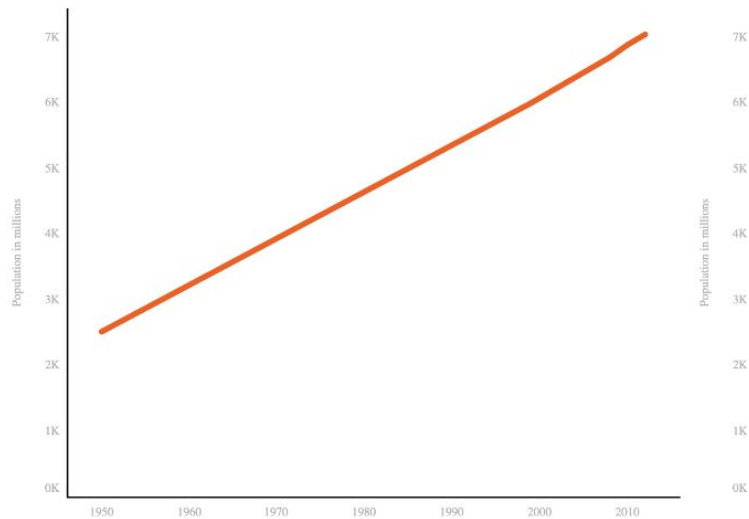




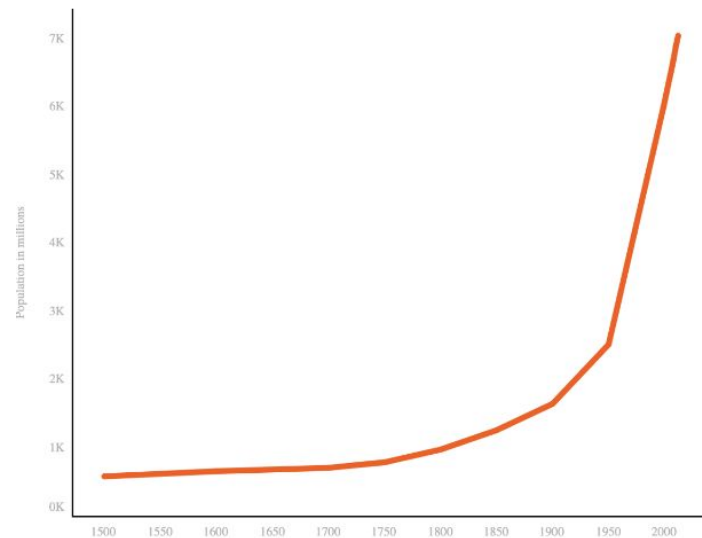
Choose the data you display  
carefully

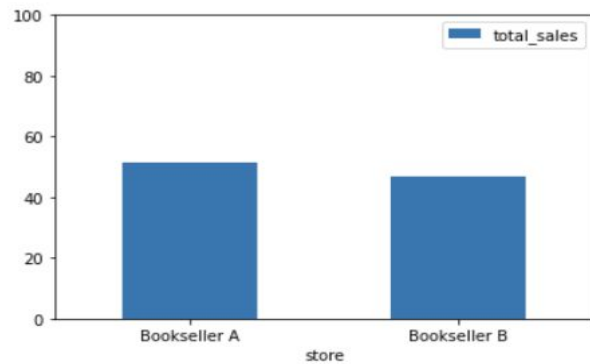
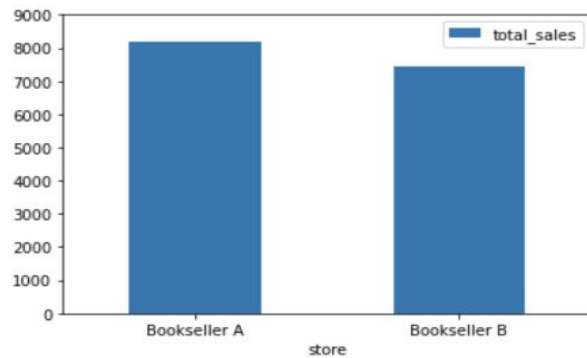
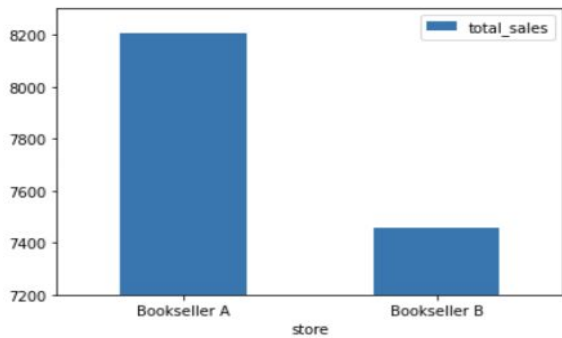


Last 60 years



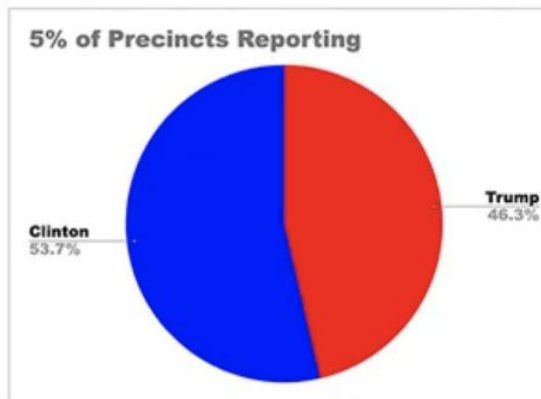
Last 500 years



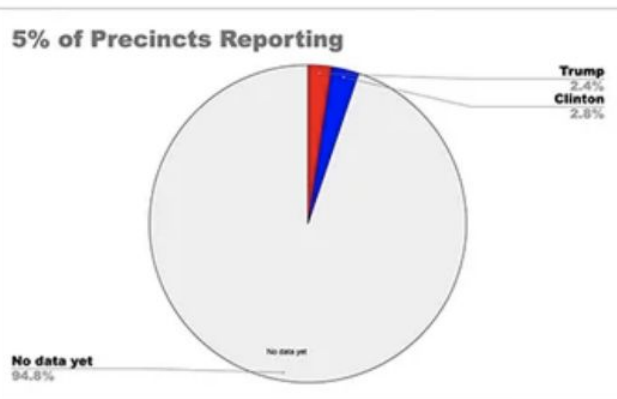


# Media, on election night:

**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**

# Exercise

(10 minutes)

Create a visualization

See `exercises/exercise_3.md`

---

**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

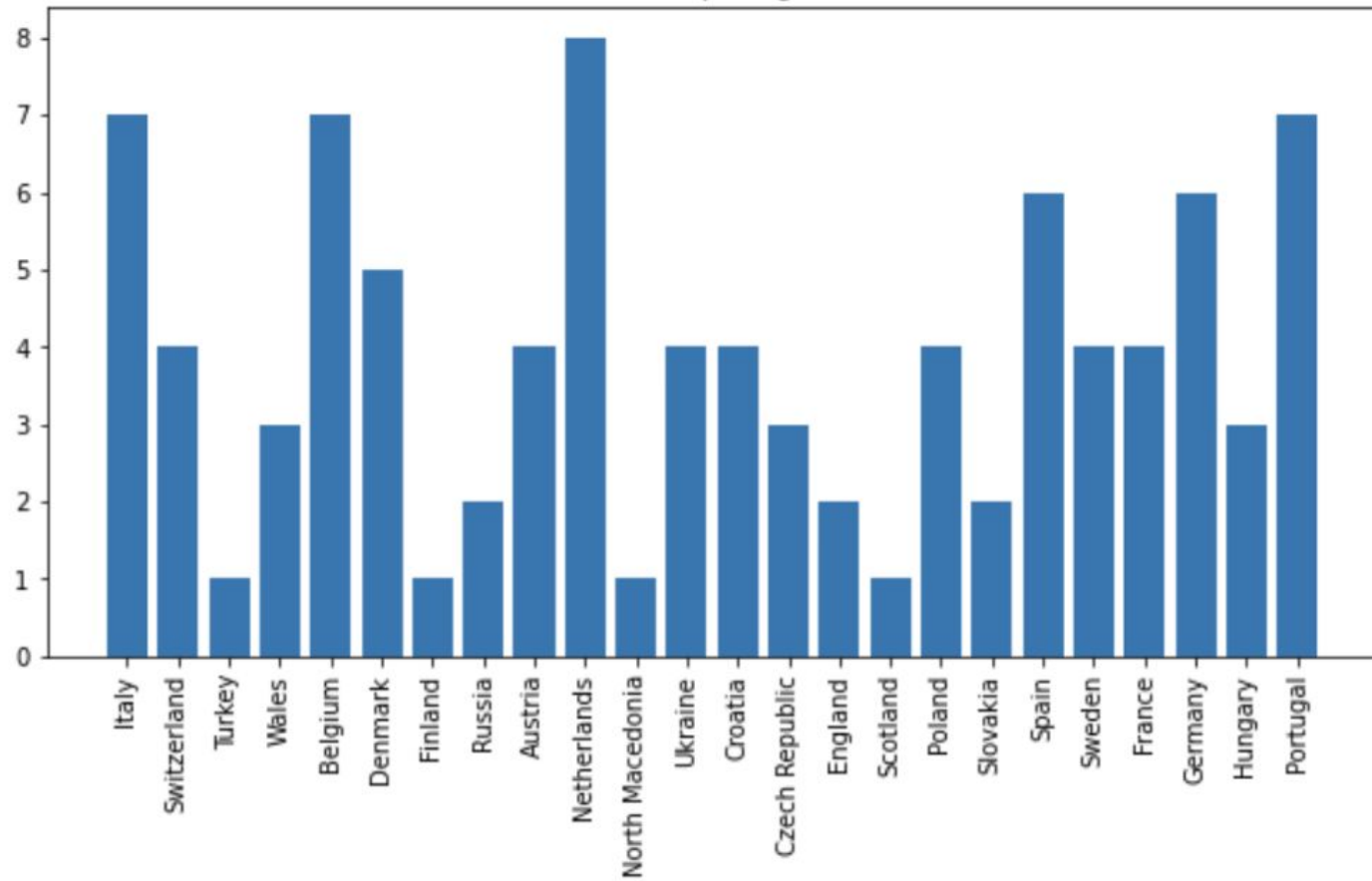
9 9 9 9 9 9 9 9 8 9 9 9 9  
9 9 9 9 8 9 9 9 9 9 9 8 9  
9 9 8 9 9 9 9 8 9 9 9 9 9

Find the 1s

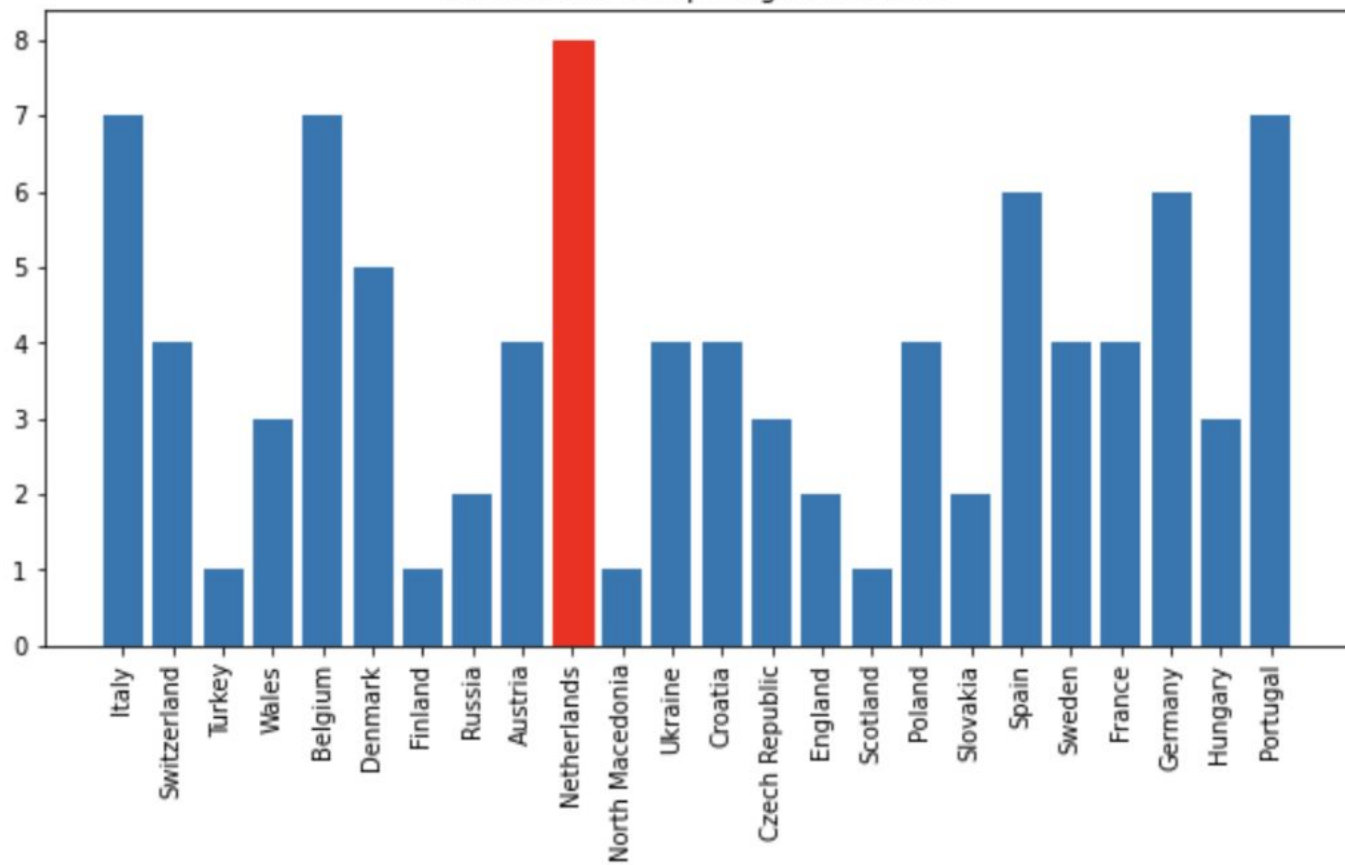
9 9 9 9 9 9 9 9 1 9 9 9 9  
9 9 9 9 1 9 9 9 9 9 9 1 9  
9 9 1 9 9 9 9 1 9 9 9 9 9



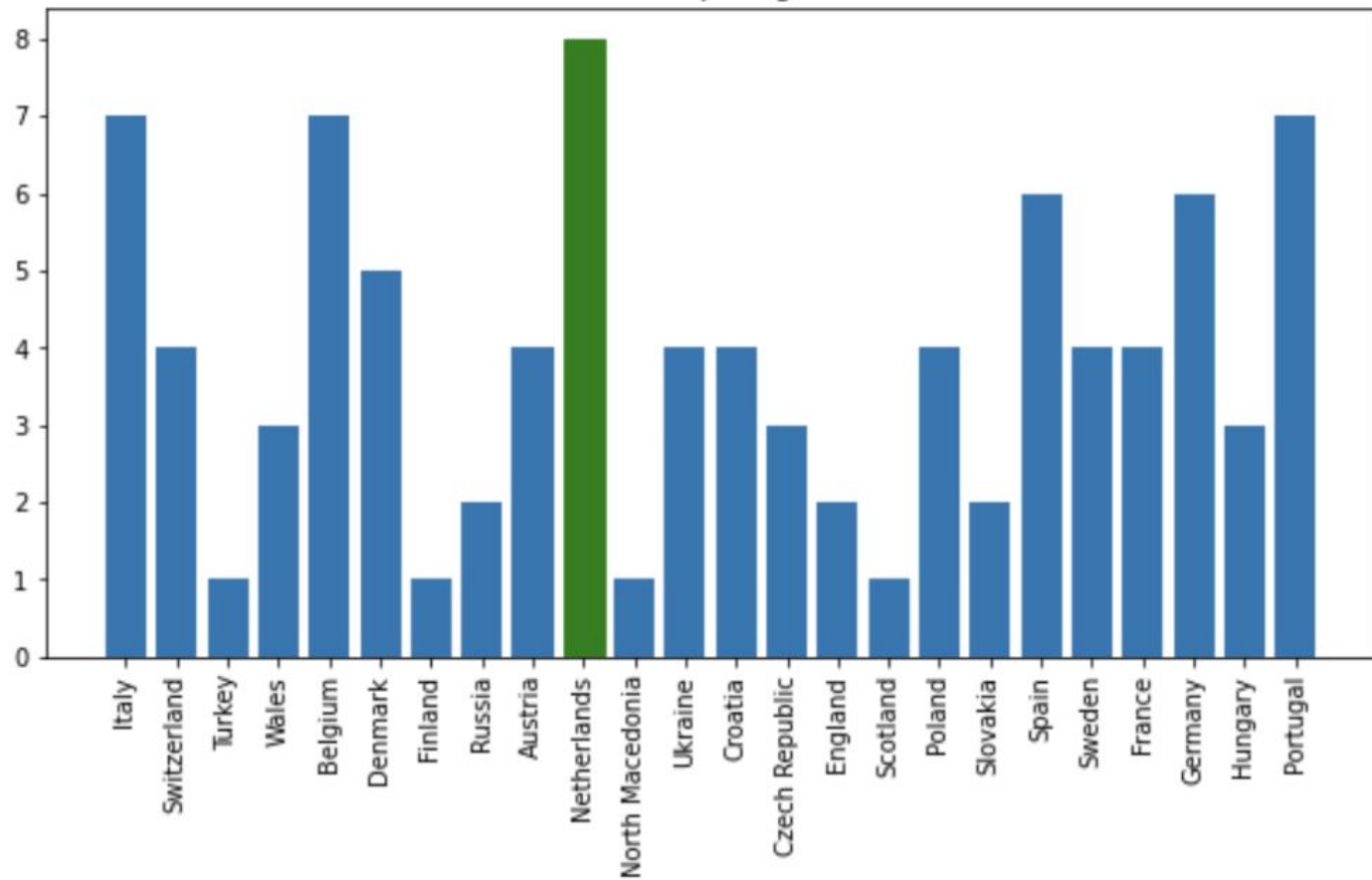
Goals in the Group Stage UEFA 2020



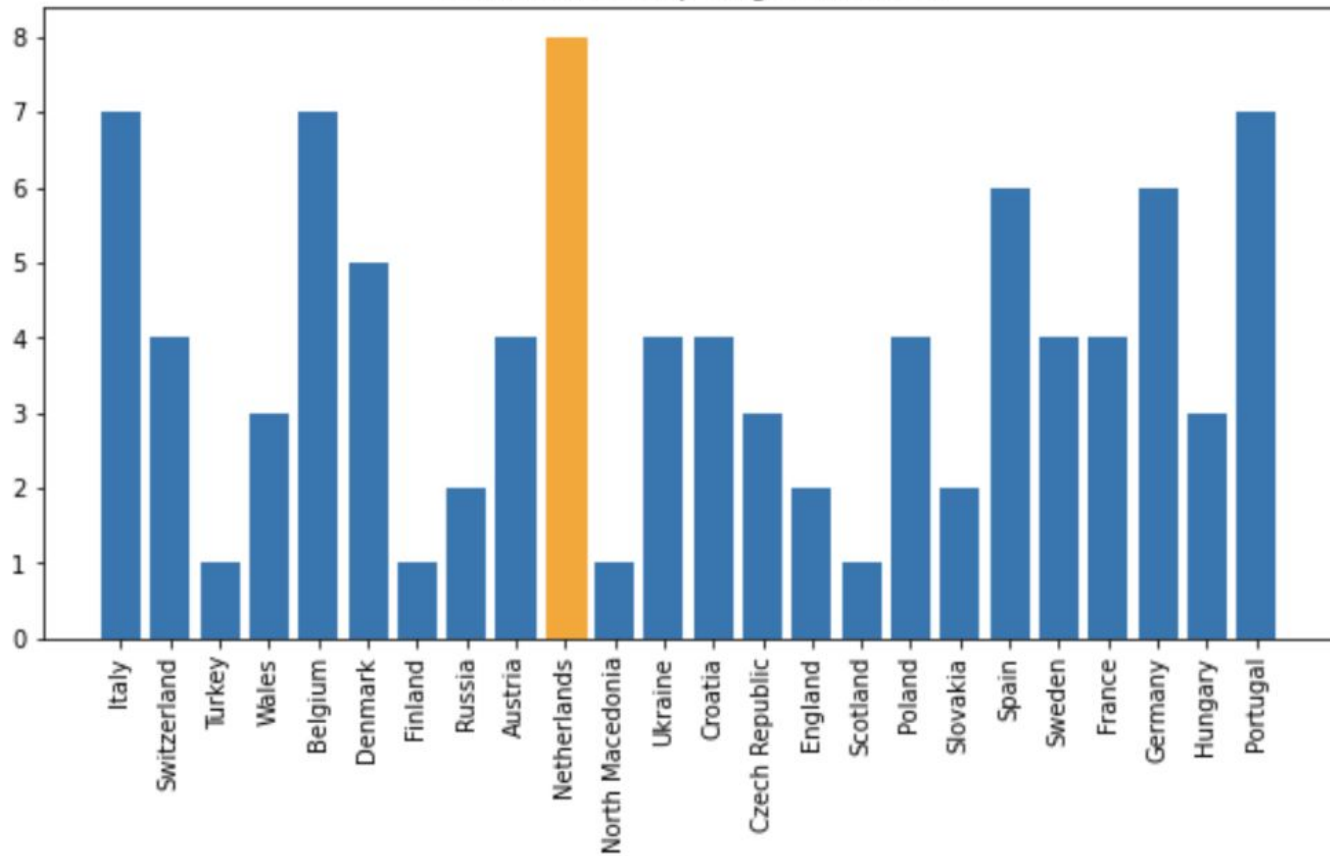
Goals in the Group Stage UEFA 2020



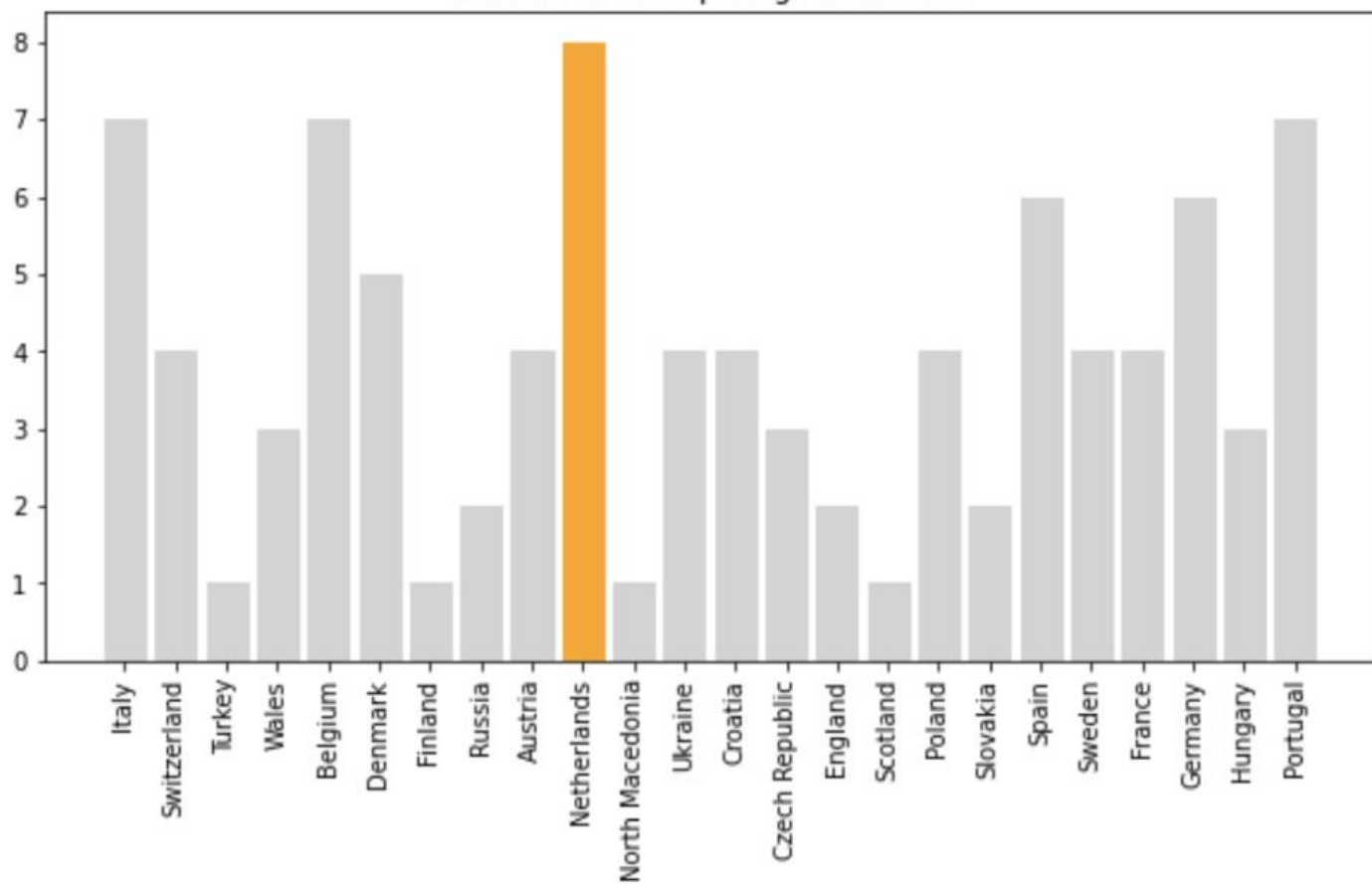
Goals in the Group Stage UEFA 2020



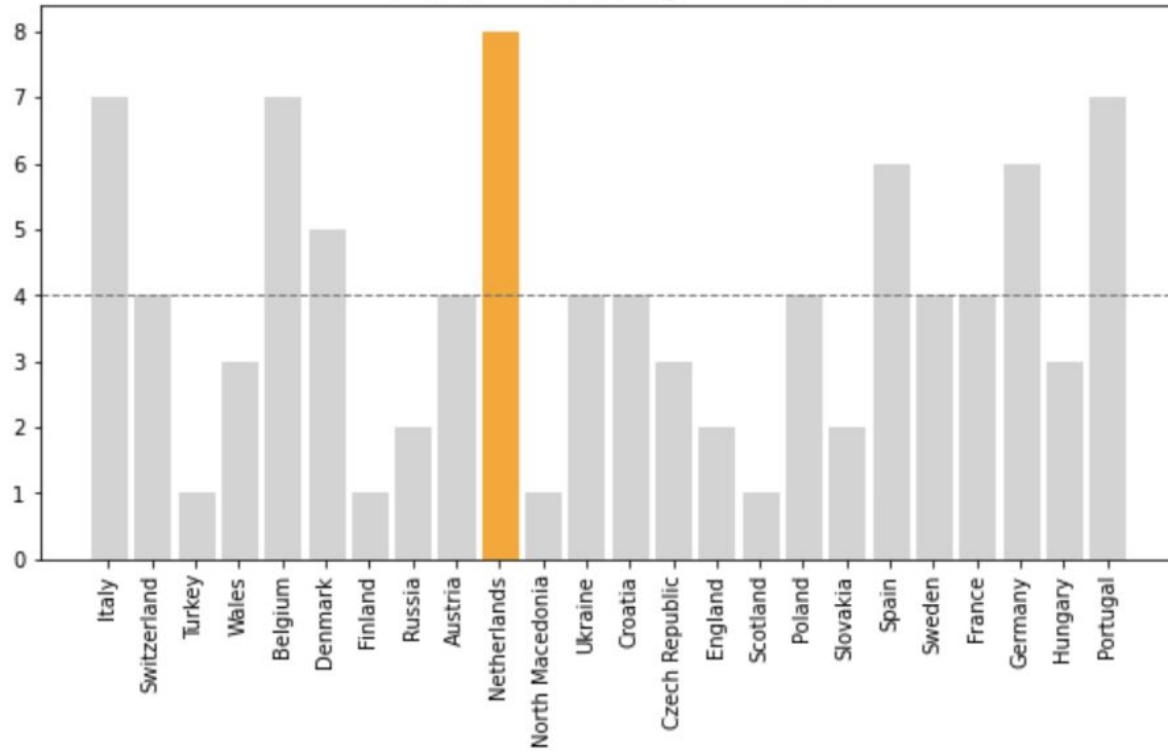
Goals in the Group Stage UEFA 2020



Goals in the Group Stage UEFA 2020

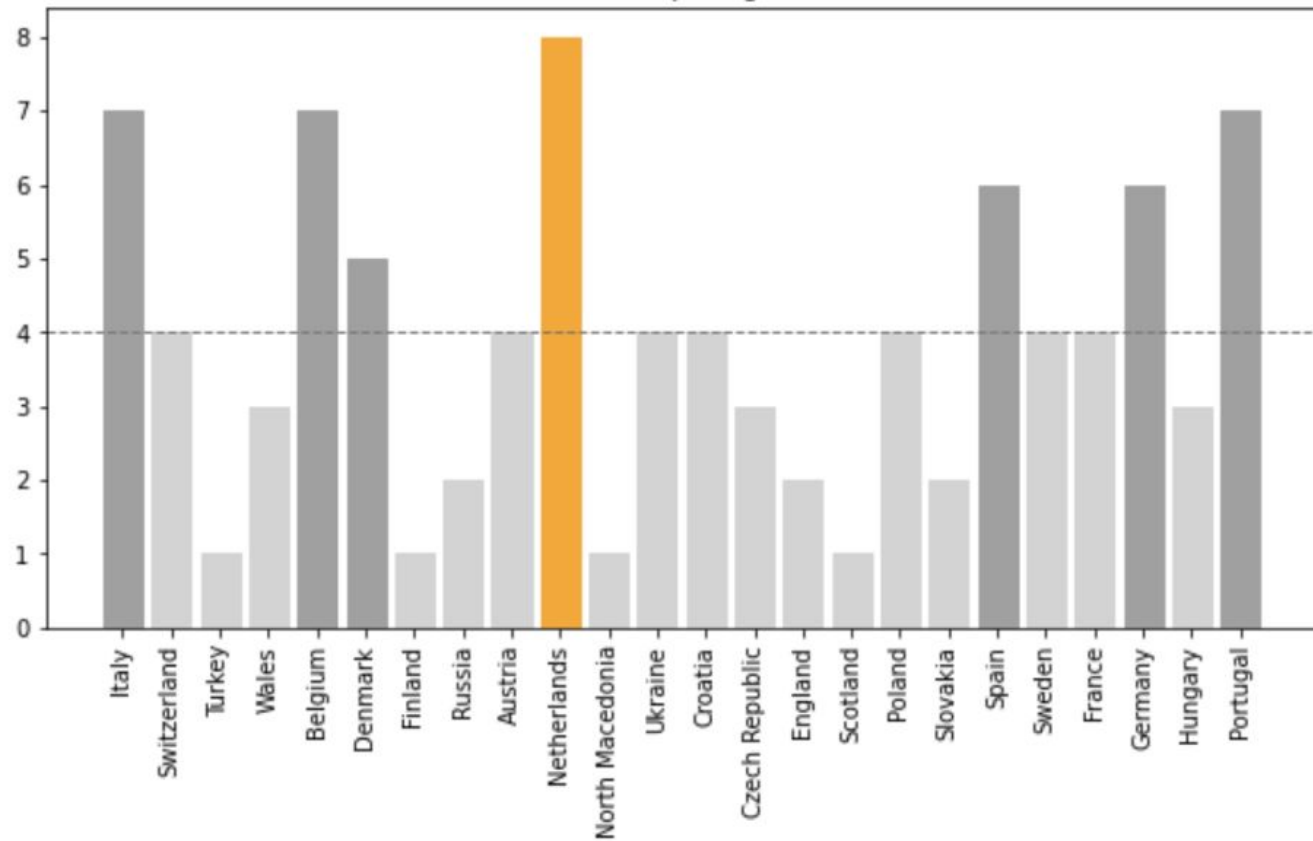


Goals in the Group Stage UEFA 2020



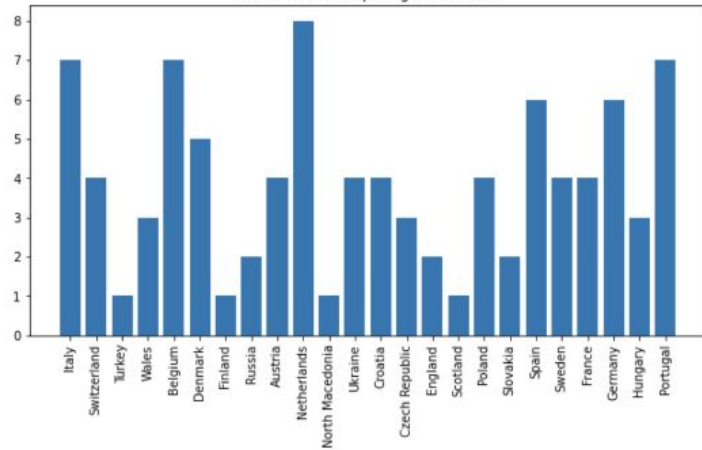
Most goals scored in the group phase in 2016

Goals in the Group Stage UEFA 2020

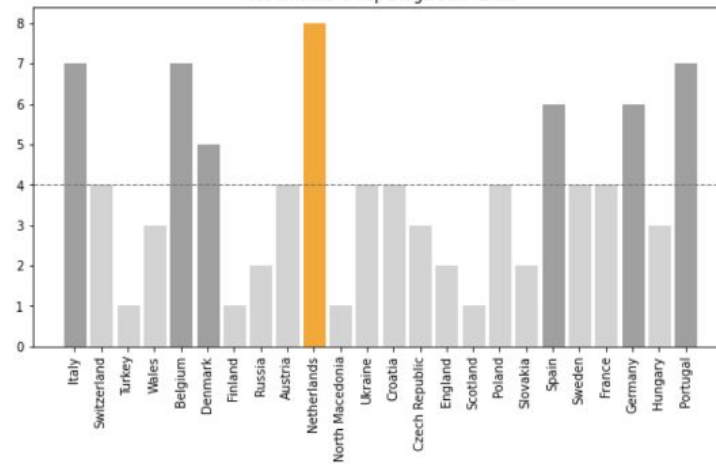


Most goals scored in the group phase in 2016

Goals in the Group Stage UEFA 2020



Goals in the Group Stage UEFA 2020



Most goals scored in the group phase in 2016

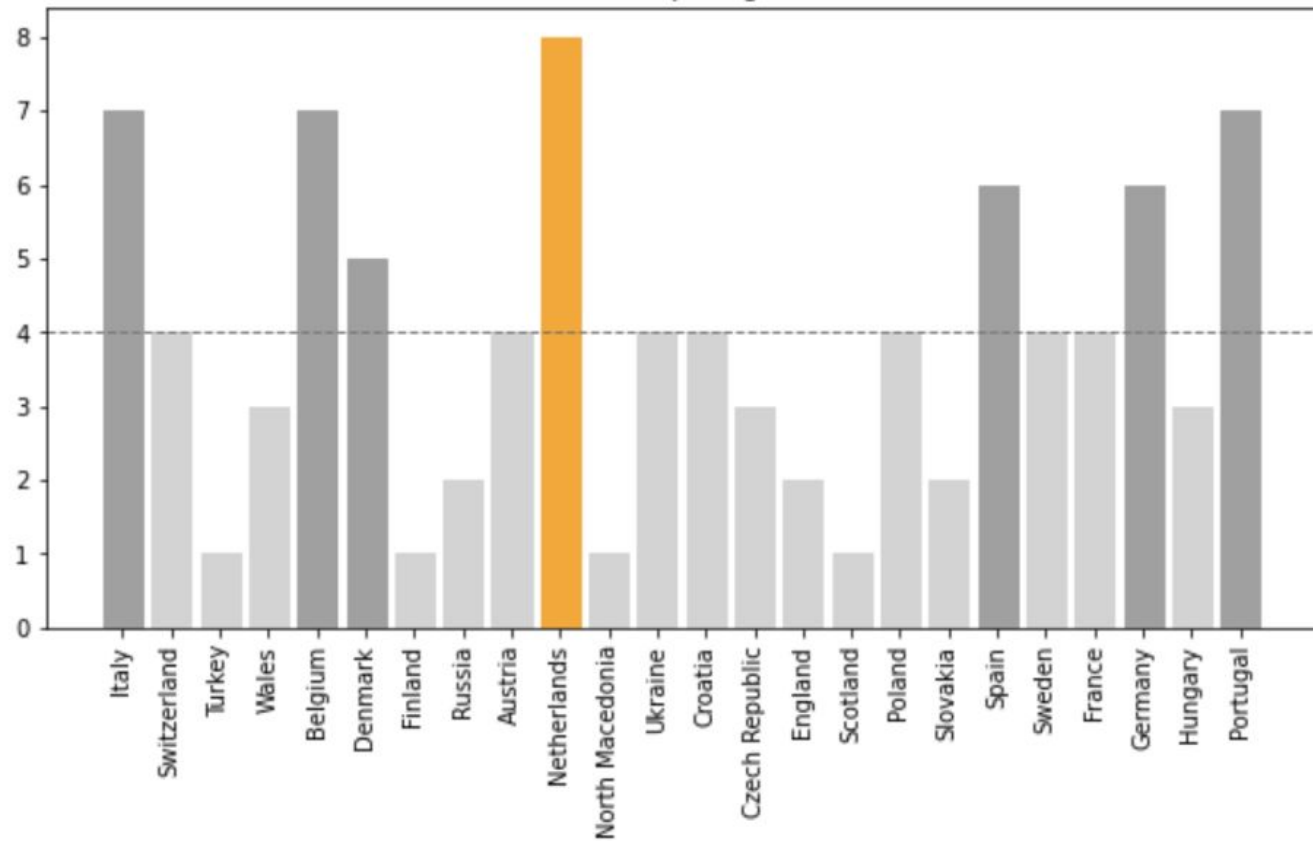


# General rules

- Use **color** to draw focus to or away from certain elements
- Use the **title** or additional **annotation elements** to guide attention
- Use text to add **additional context**
- **Group your data** logically
- Keep your graph **clean!**

Visualization **do not** have to be pretty,  
but it **does help** to keep readers  
interested and therefore focussed

Goals in the Group Stage UEFA 2020



Most goals scored in the group phase in 2016

# Exercise

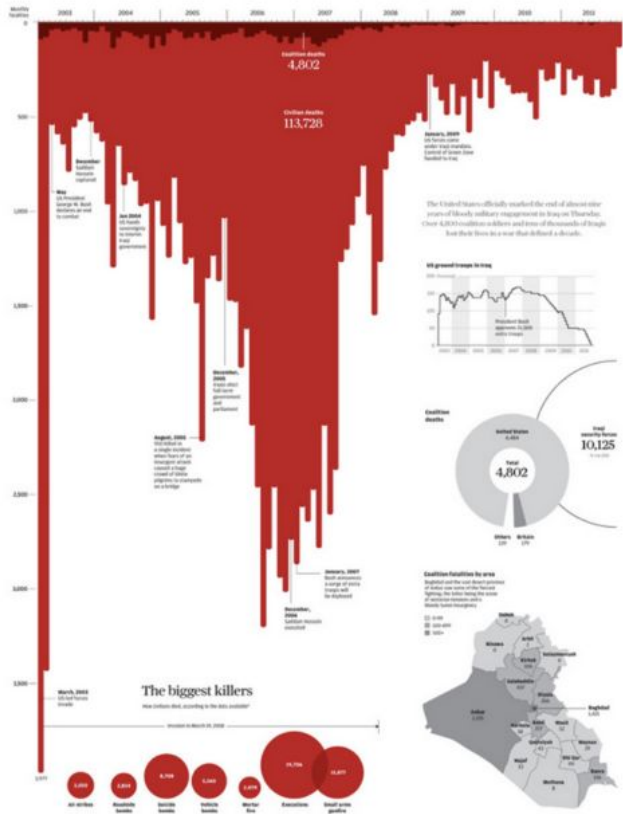
(10 minutes)

## Guide the focus

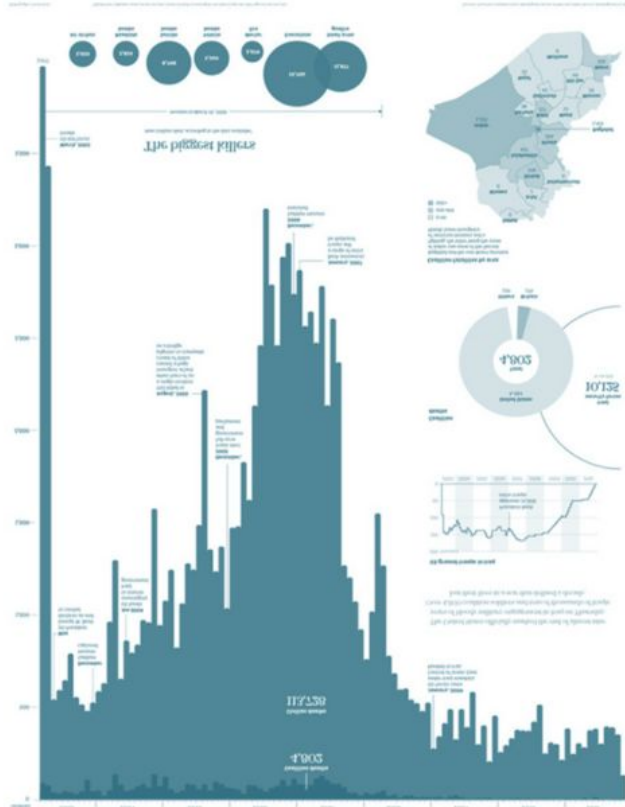
Redesign your previous graph. See instructions in `exercises/exercise_4.md`

— — —

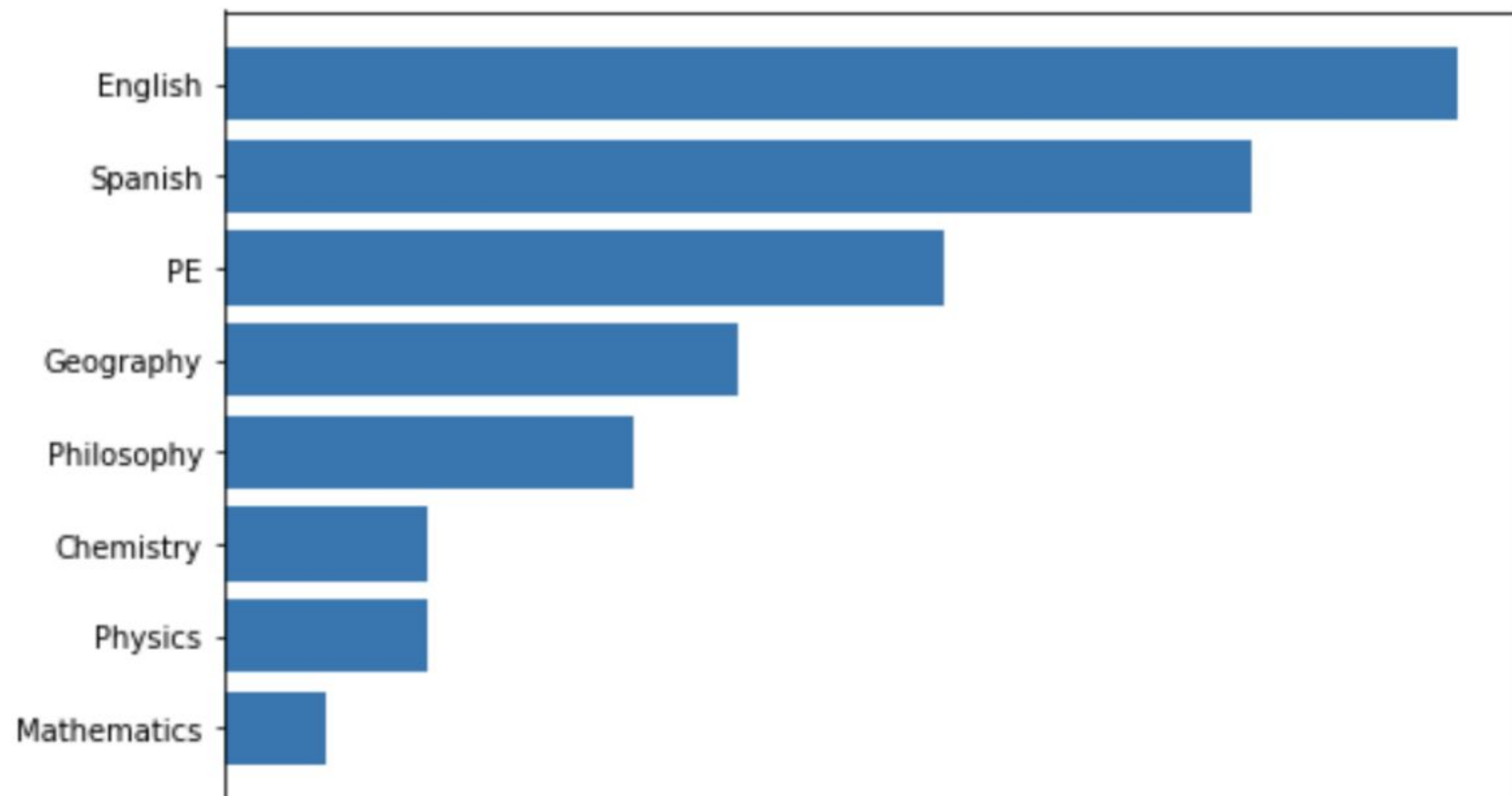
## Iraq's bloody toll



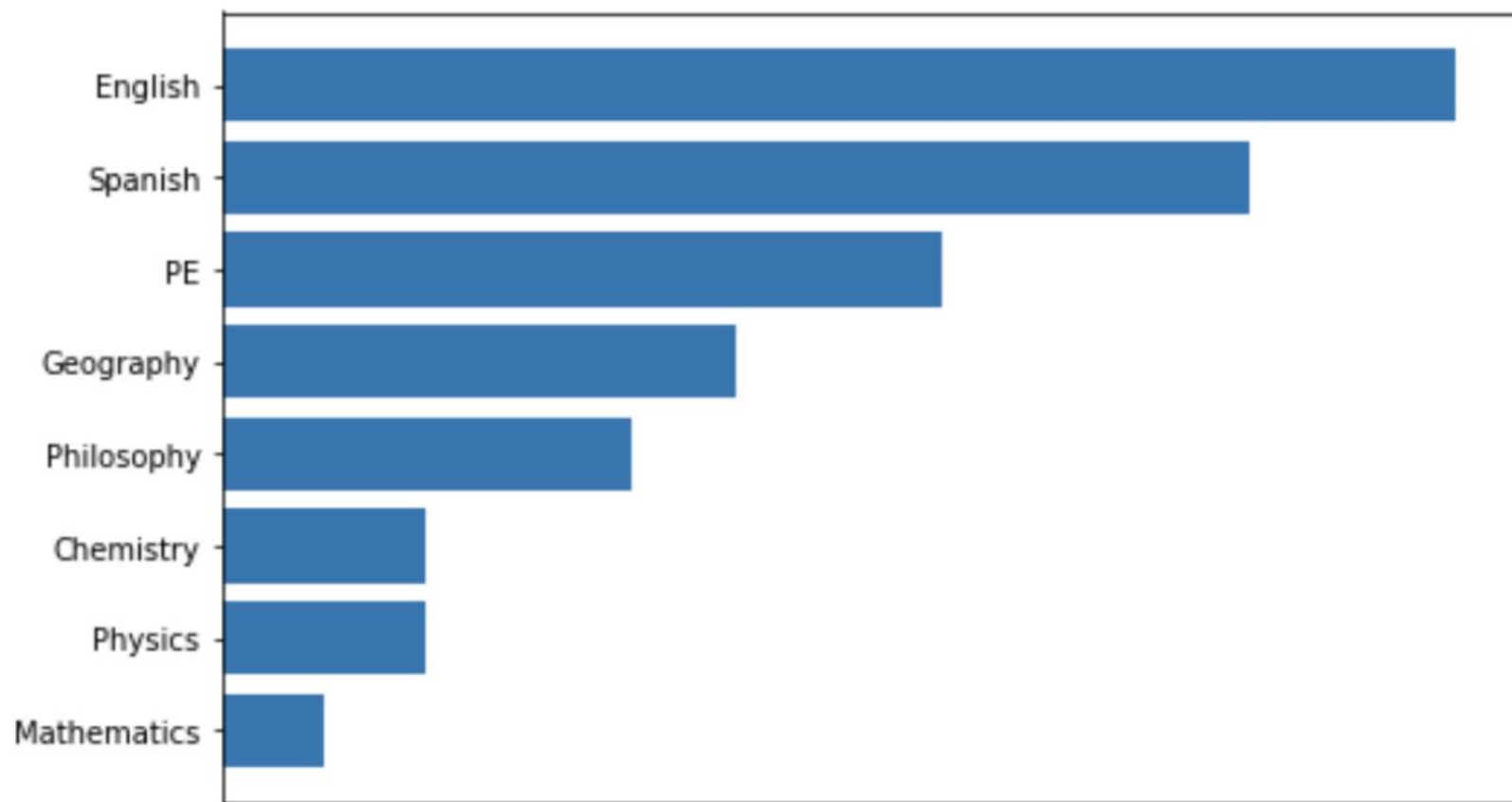
## Iraq: Deaths on the decline



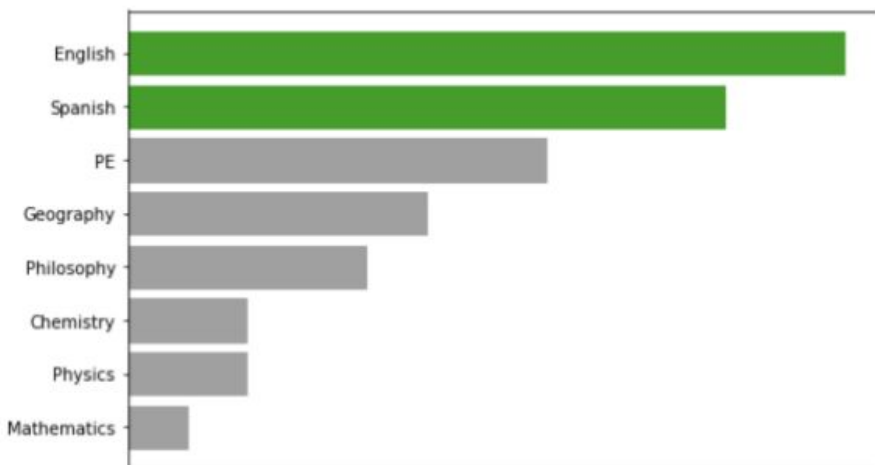
## Languages most popular among class 3A



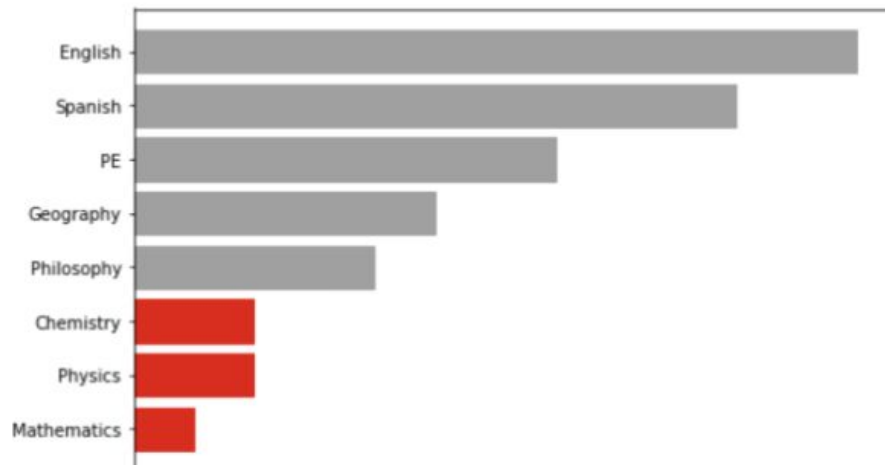
## STEM not popular among class 3A



Languages popular among class 3A



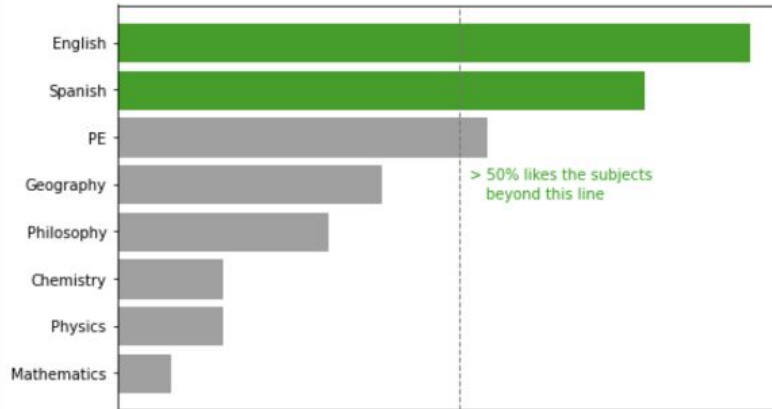
STEM not popular among class 3A





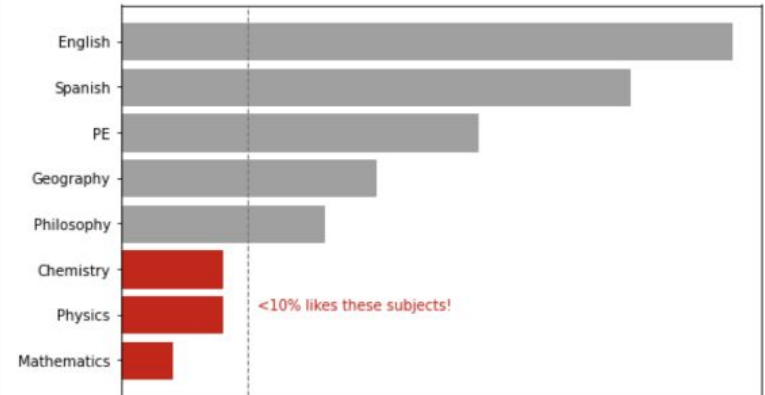
## Languages popular among class 3A

% of class 3A that likes the subject



## STEM subjects very unpopular among class 3A

% of class 3A that likes the subject



# Exercise

(10 minutes)

## Shift the message

Use the graph from the previous exercise. Create **two** very different messages.

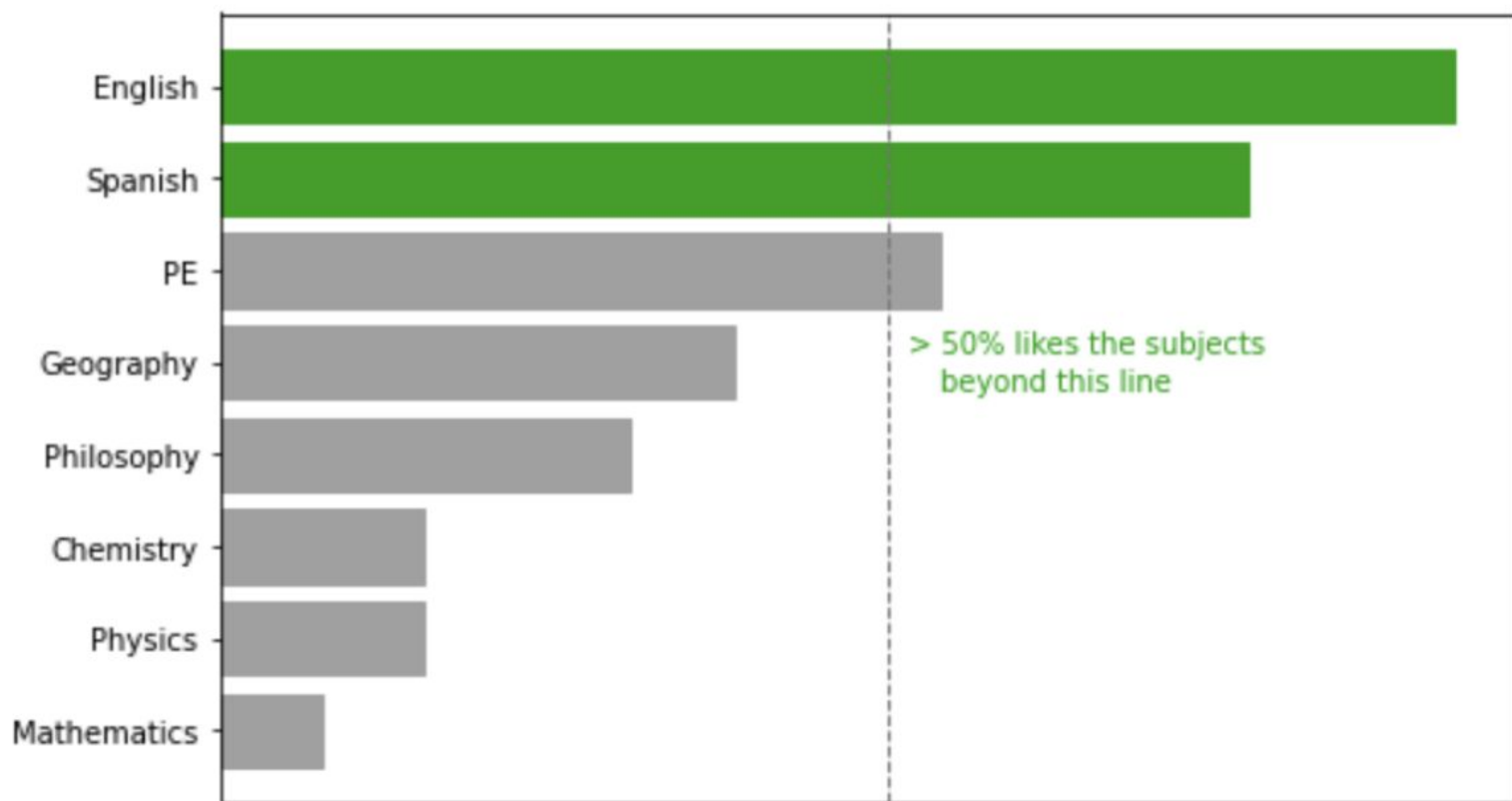
— — —

**Forward**

What is the **action** that should  
follow your visualization?

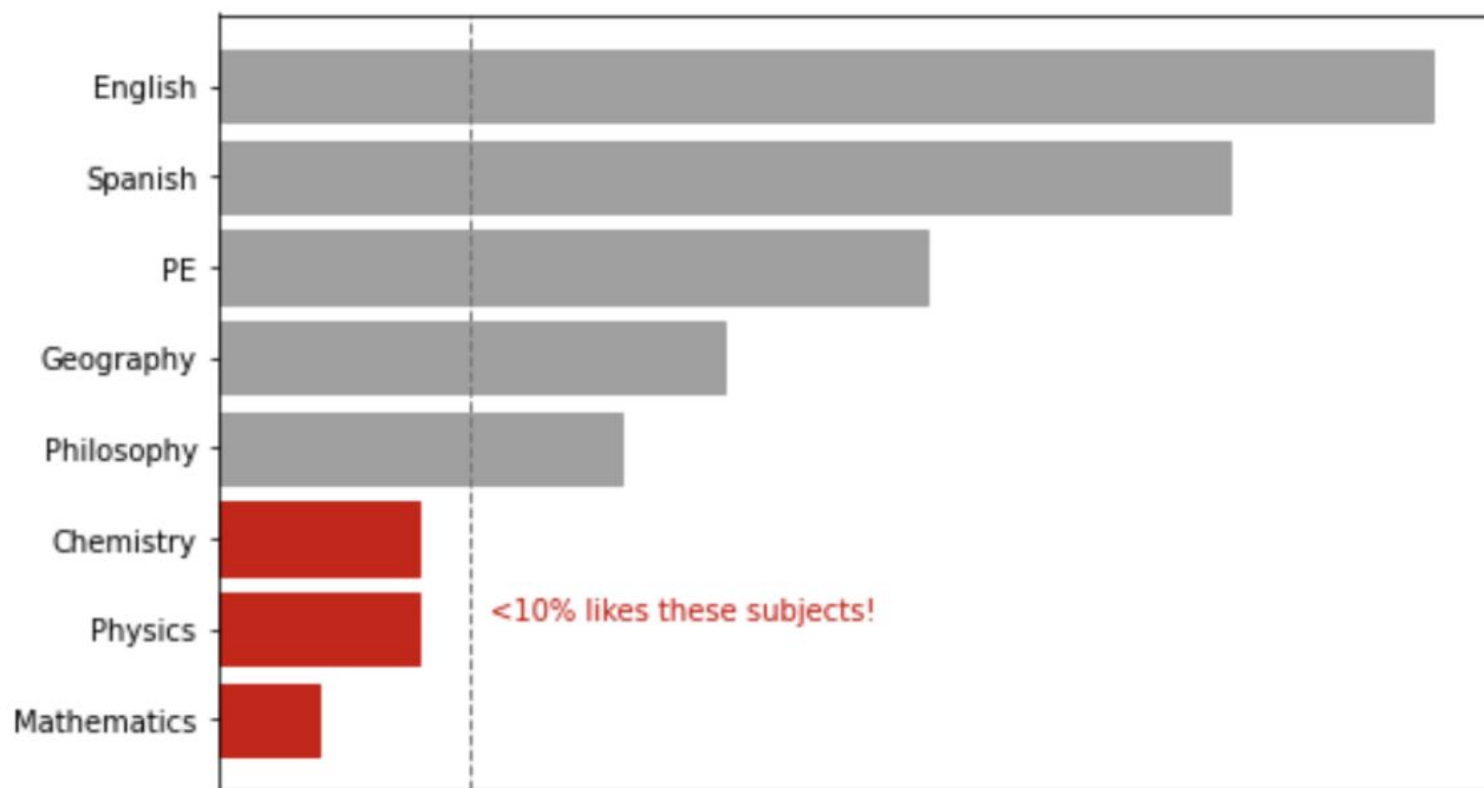
# Languages popular among class 3A

% of class 3A that likes the subject



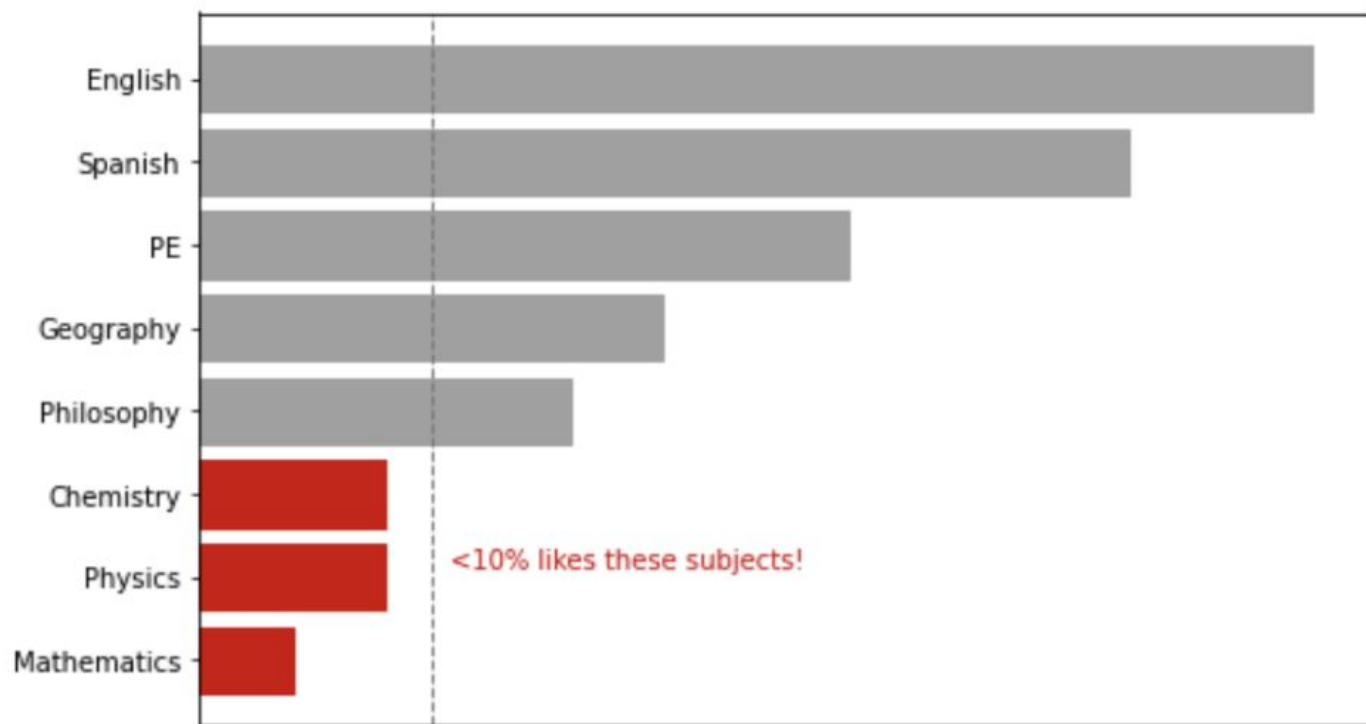
# STEM subjects very unpopular among class 3A

% of class 3A that likes the subject

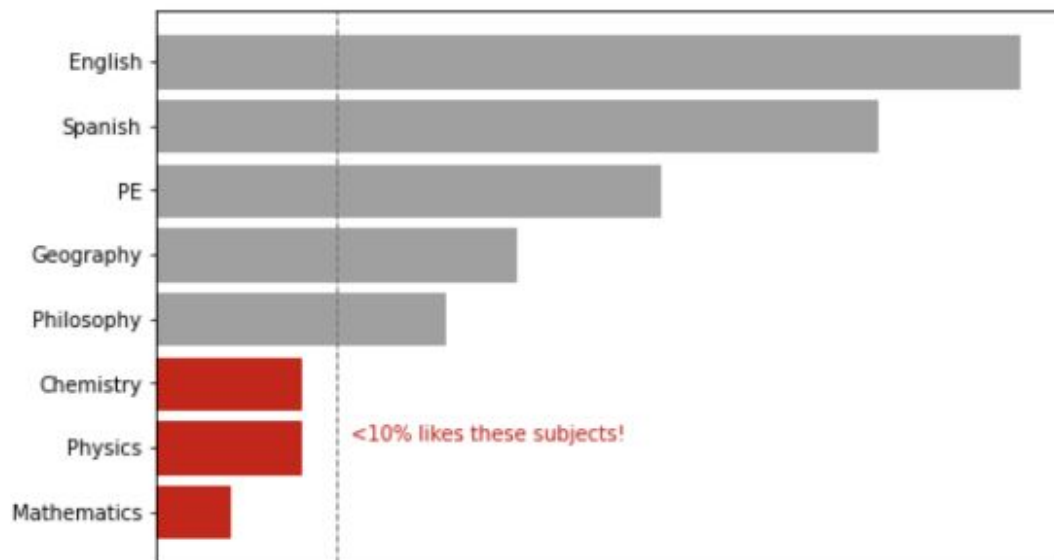


# We need to change the way we teach STEM subjects at our school

% of class 3A that likes the subject



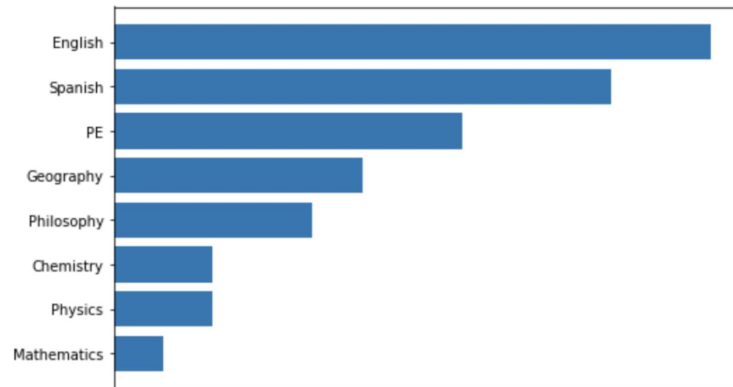
## We need to change the way we teach STEM subjects at our school



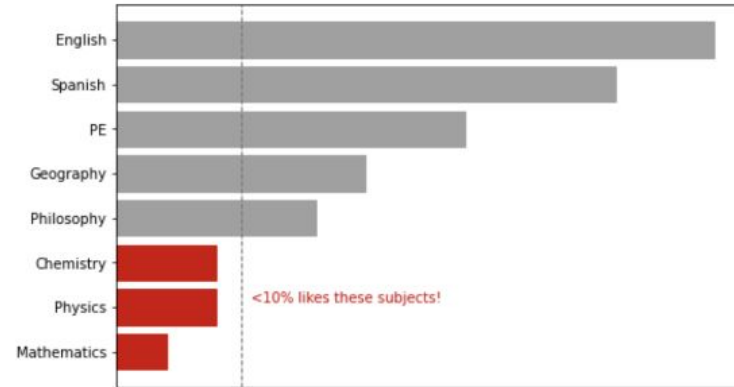
Our STEM subjects are heavily disliked. As a school, we must take action and investigate how we can make these subjects more fun and enjoyable for our students.



### STEM not popular among class 3A



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

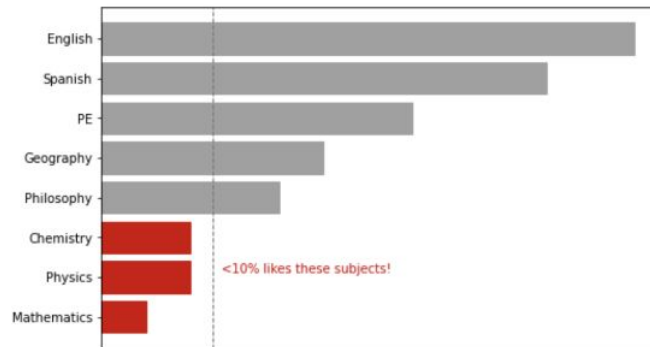


Our STEM subjects are heavily disliked. As a school, we must take action and investigate how we can make these subjects more fun and enjoyable for our students.

# Drive action forward

- Put the message **first**
- **Support** the conclusion
- Emphasize the **action**

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



Our STEM subjects are heavily disliked. As a school, we must take action and investigate how we can make these subjects more fun and enjoyable for our students.

# Exercise

## Drive the action

1. Create an appealing slide with the visualization, that effectively emphasizes the recommended action to take.
2. Create a narrative around this recommendation, and create multiple slides to make a convincing story!

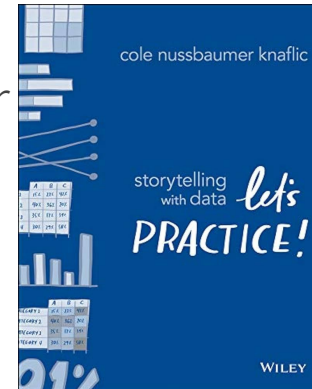
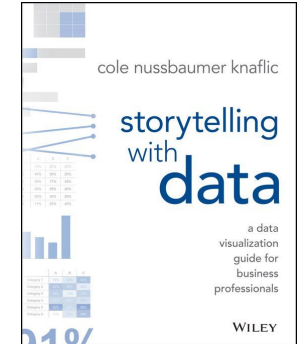
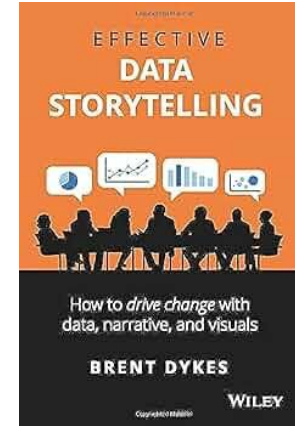
— — —

# Summary

- Why data storytelling is important
- How to choose a good **foundation** for presenting your data
- How to draw the **focus** of the reader to the right places
- How you can use a visualization to drive action **forward**

# Recommended resources

- *Effective Data Storytelling: How to Drive Change with Data, Narrative and Visuals*
  - Brent Dykes
- *Storytelling with Data: A Data Visualization Guide for Business Professionals*
  - Cole Nussbaumer-Knaflic
- *The Storytelling Edge: How to Transform Your Business, Stop Screaming into the Void, and make People Love You*
  - Shane Snow, Joe Luzauskas



# Feel free to reach out!

---



Marysia Winkels



[marysia.nl](https://marysia.nl)



[hello@marysia.nl](mailto:hello@marysia.nl)

