

Data Analysis & Visualisation

Week 2

Raoul Grouls, 16-9-2025

Recap leerdoelen les 1

- de gestalt principles & five guidelines toepassen op visualisaties
- een virtual environment activeren met uv
- nieuwe features extraheren met behulp van regular expressions
- Een script vanaf de terminal opstarten
- click gebruiken voor command line arguments bij een script
- begrijpt de opzet van een project (src folder, data/raw en data/processed, pyproject.toml, notebooks) en kan dit zelf opzetten
- kan een eigen git-repo maken
- Regular expressions toepassen:
 - start ^
 - end \$
 - or: [Bb]
 - ranges [a-zA-Z]
 - any char .
 - zero or more a*
 - one or more a+
 - not in range [^a-zA-Z]
 - shortcuts (\w, \s, \d)
 - lookbehind (?<=...)
 - lookahead (?=...)

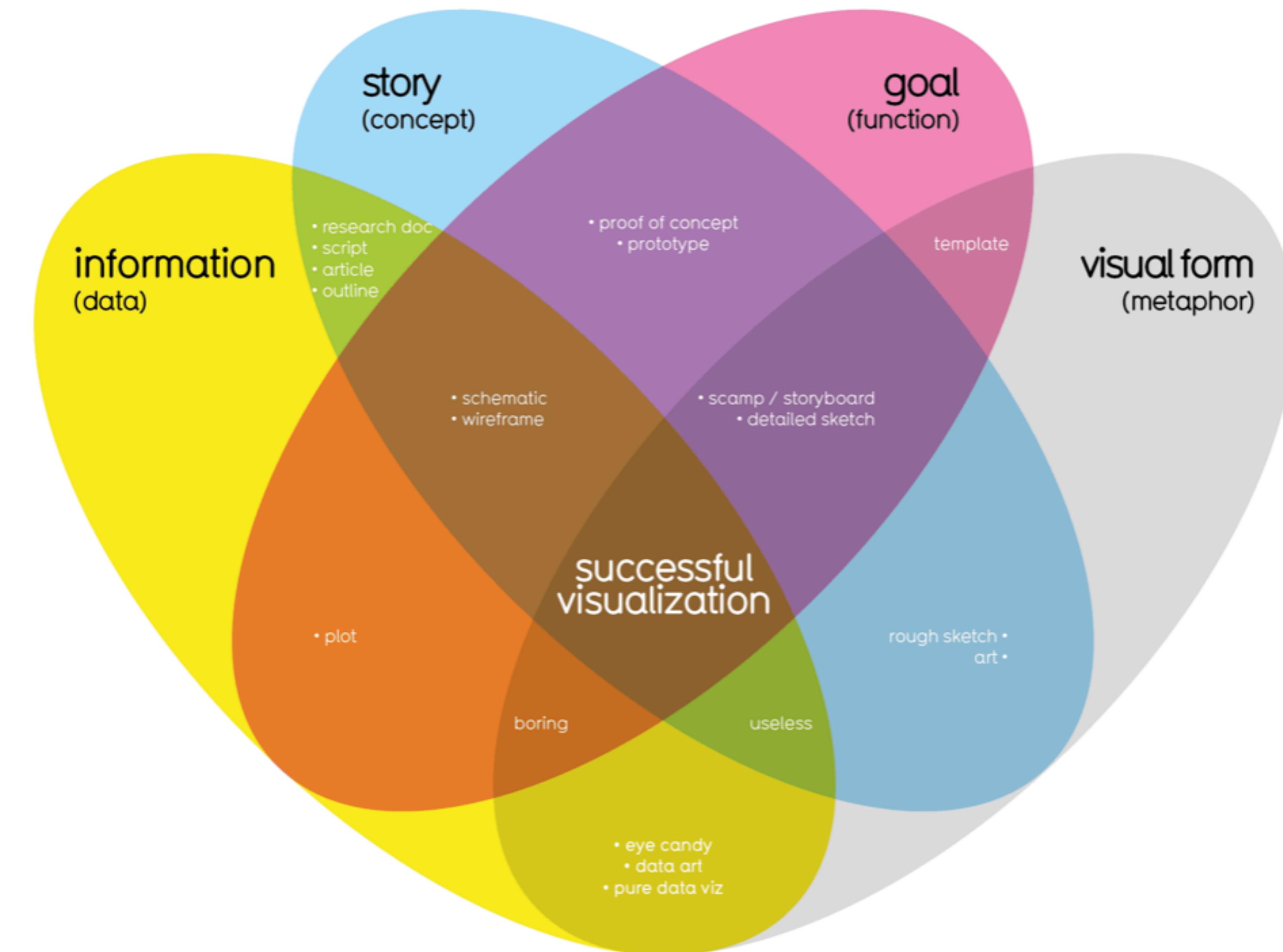
leerdoelen les 2

- Leren toepassen van visualisatie principes
- Omgaan met venv, uv, path, scripts en git
- Oefenen met nieuwe features extraheren met behulp van regular expressions
- Vergelijken van categorieën met behulp van data visualisaties:
 - Barplots
 - Barbell plot
 - heatmaps
- Werken met palettes (en list comprehensions)
- Pandas
 - Pandas groupby & aggregate
 - Pandas cut

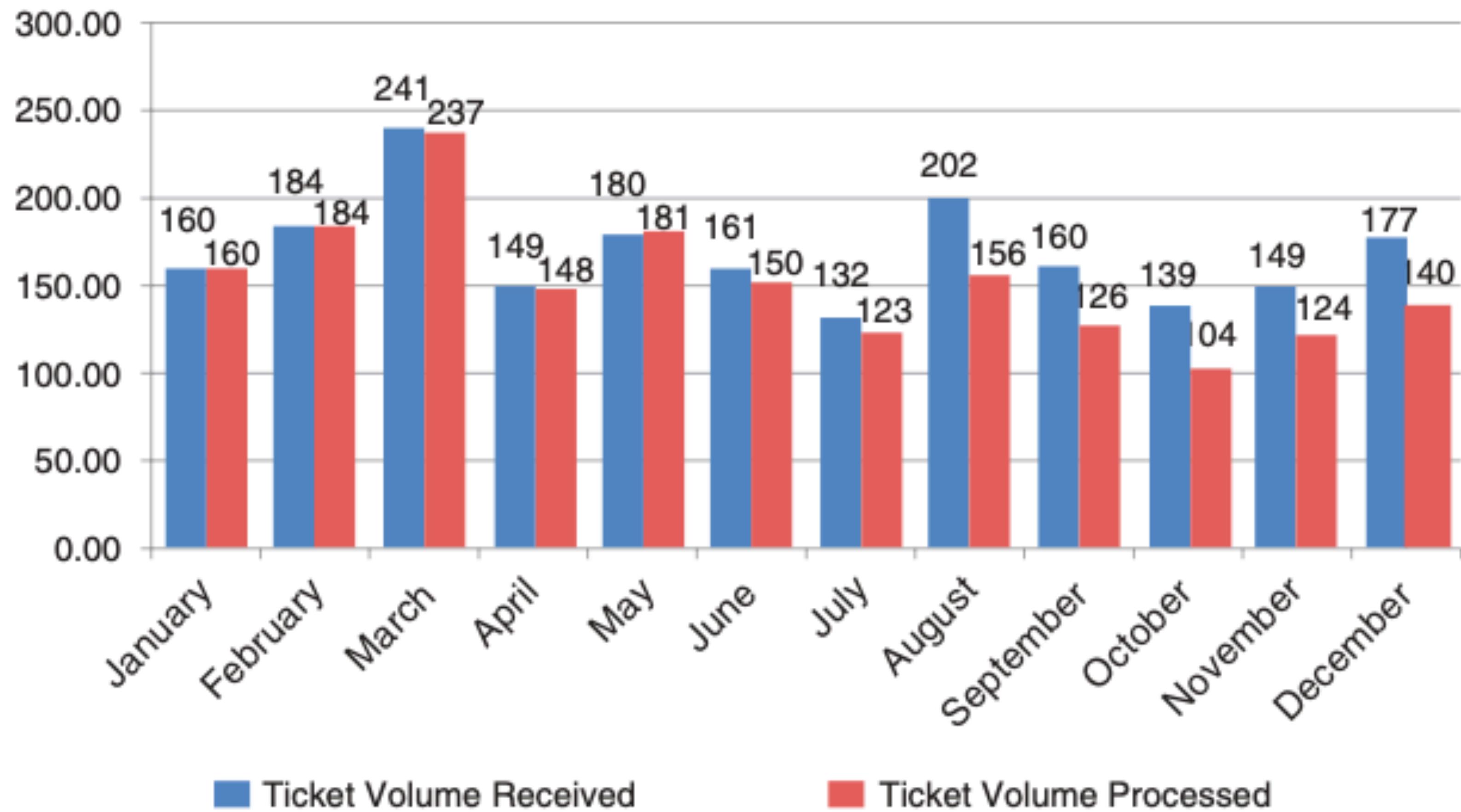
rollover for more detail

What Makes a Good Visualization?

explicit (implicit)



Ticket Trend



Please approve the hire of 2 FTEs

to backfill those who quit in the past year

Ticket volume over time

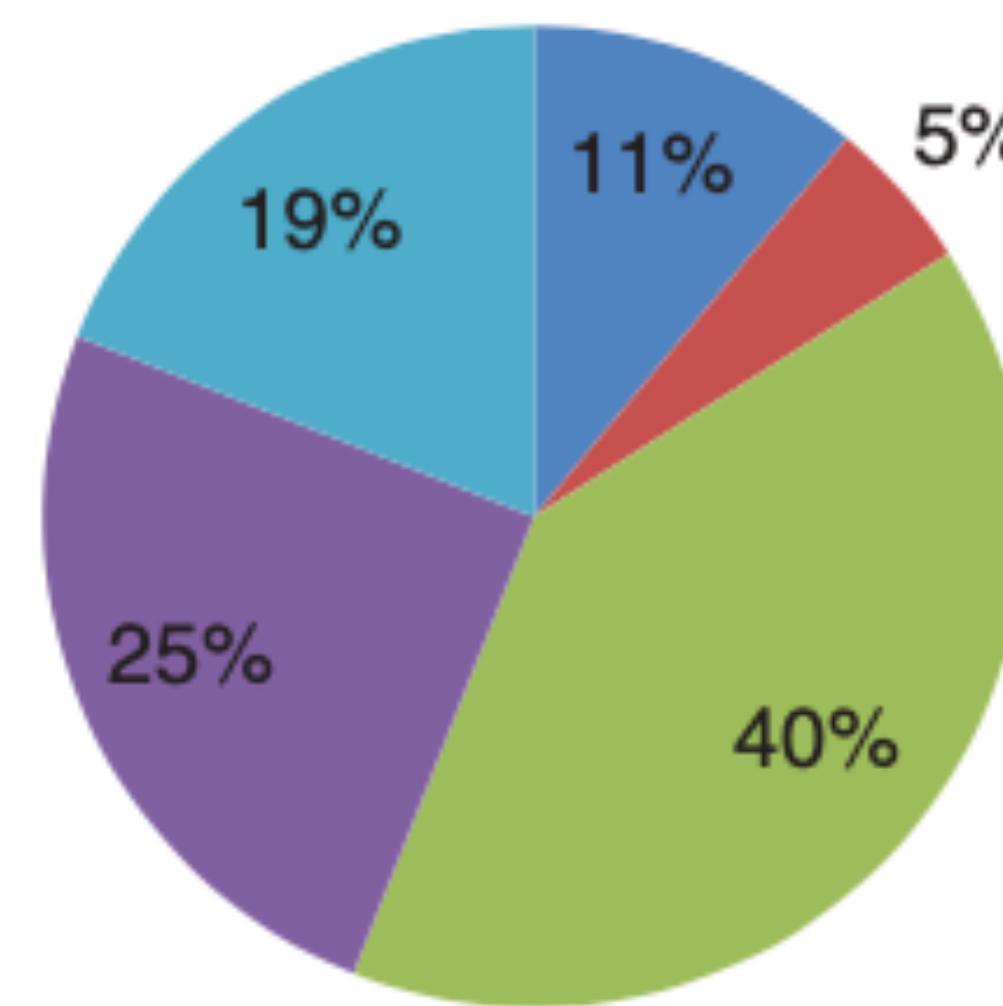


Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

Survey Results

PRE: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



POST: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited

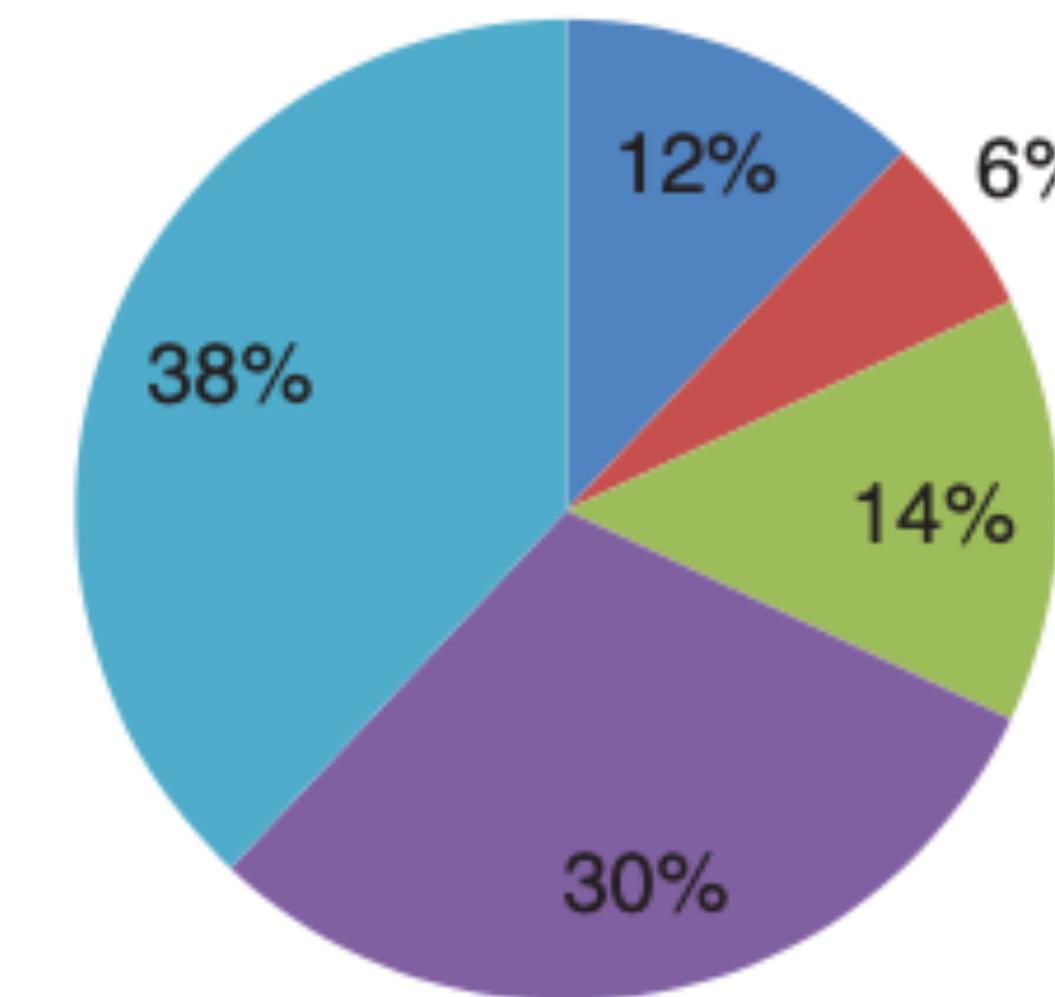


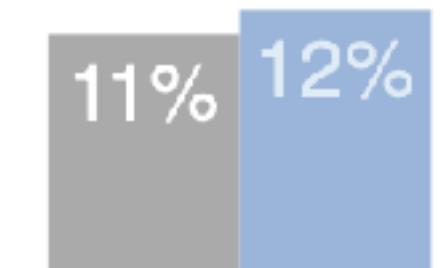
FIGURE 0.4 Example 2 (before): showing data

Pilot program was a success

How do you feel about science?

BEFORE program, the majority of children felt just OK about science.

40%



AFTER program, more children were *Kind of interested* & *Excited* about science.

38%

19%

30%

25%

14%

Bored

Not great

OK

Kind of
interested

Excited

Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

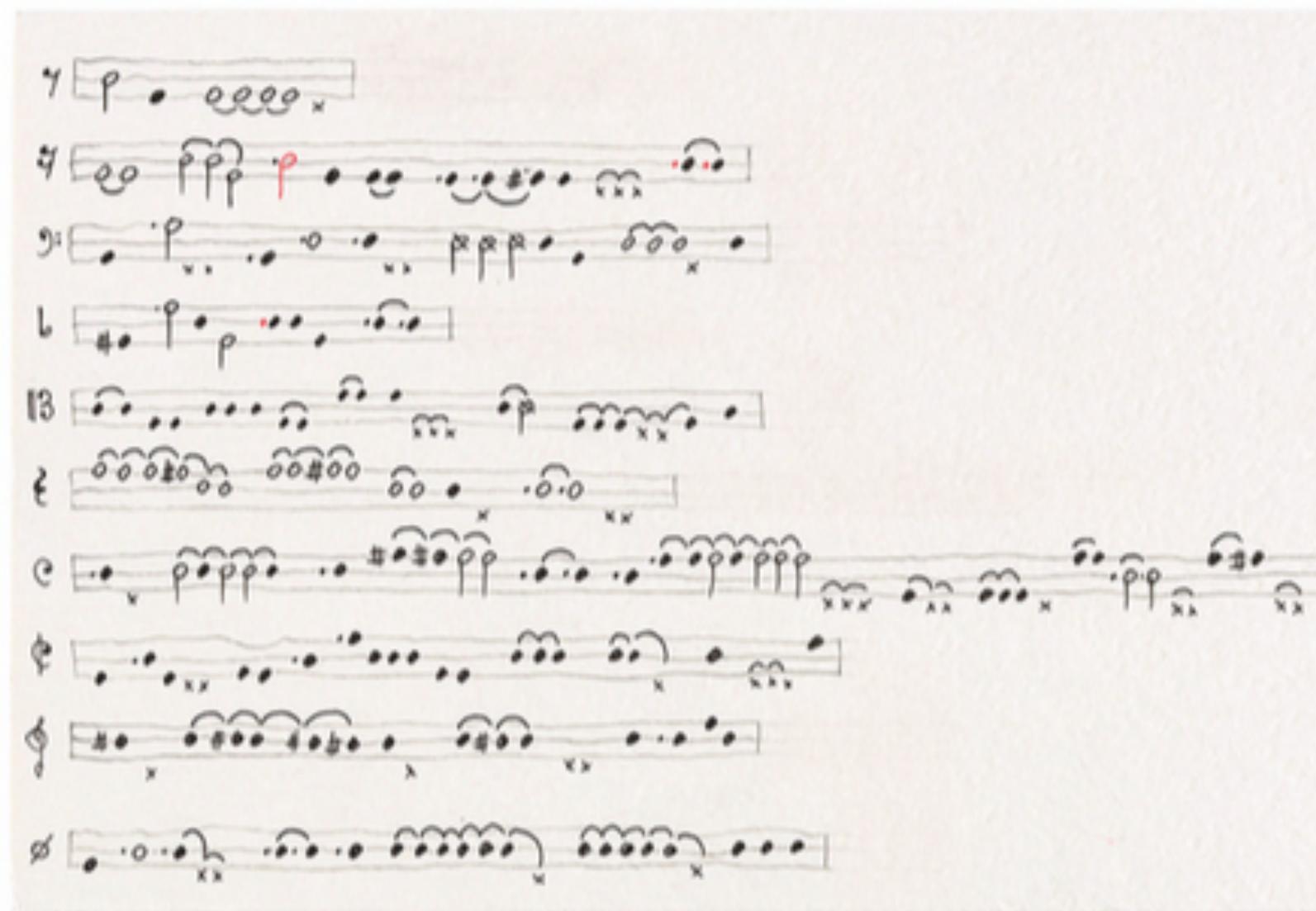
FIGURE 0.5 Example 2 (after): storytelling with data

Afweging 1: snelheid

- Meer story, minder information?
- preattentive processing vs bestuderen?

GIORGIA

week seven



DEAR DATA
WEEK 07: MUSICAL COMPLAINTS

HOW TO READ IT: Each "note" is a single complaint I said.
(i.e. every single time I expressed dissatisfaction or annoyance about a situation or particular thing)

Each "score" represents a typology of things I complained about, featuring complaints in chronological order.

SCORES:

- 1 - ME AS A PERSON (e.g. "I am so... angry / obsessive...")
- 2 - ME AT WORK (e.g. "I should've done ...")
- 3 - WORK (e.g. "this project isn't going well!")
- 4 - TECHNOLOGY (e.g. "The computer is not working!")
- 5 - SERVICE/FOOD (e.g. "OMG the waiter is so slow!")
- 6 - SOMEONE (e.g. "He's really a jerk...")
- 7 - COLD (e.g. "I am freezing! the A.C. is crazy!")
- 8 - HOW I FEEL (e.g. "so tired!!", "so bored!...")
- 9 - BOYFRIEND (e.g. "You're enabling! you haven't...")
- 10 - OTHER (e.g. "I spent 1 hour waiting for...")

POSITIONS OF NOTES:

- 1 → ACTUAL need to complain
- 2 → average " "
- 3 → MORE " "
- 4 → MISSING COMPLAINTS:
thought of complaining
but didn't do!

FROM:
GIORGIA LUPI
NY BROOKLYN
NY - USA



SEND TO:

STEFANIE POSAVEC

NEW YORK CITY

LONDON

- UK -

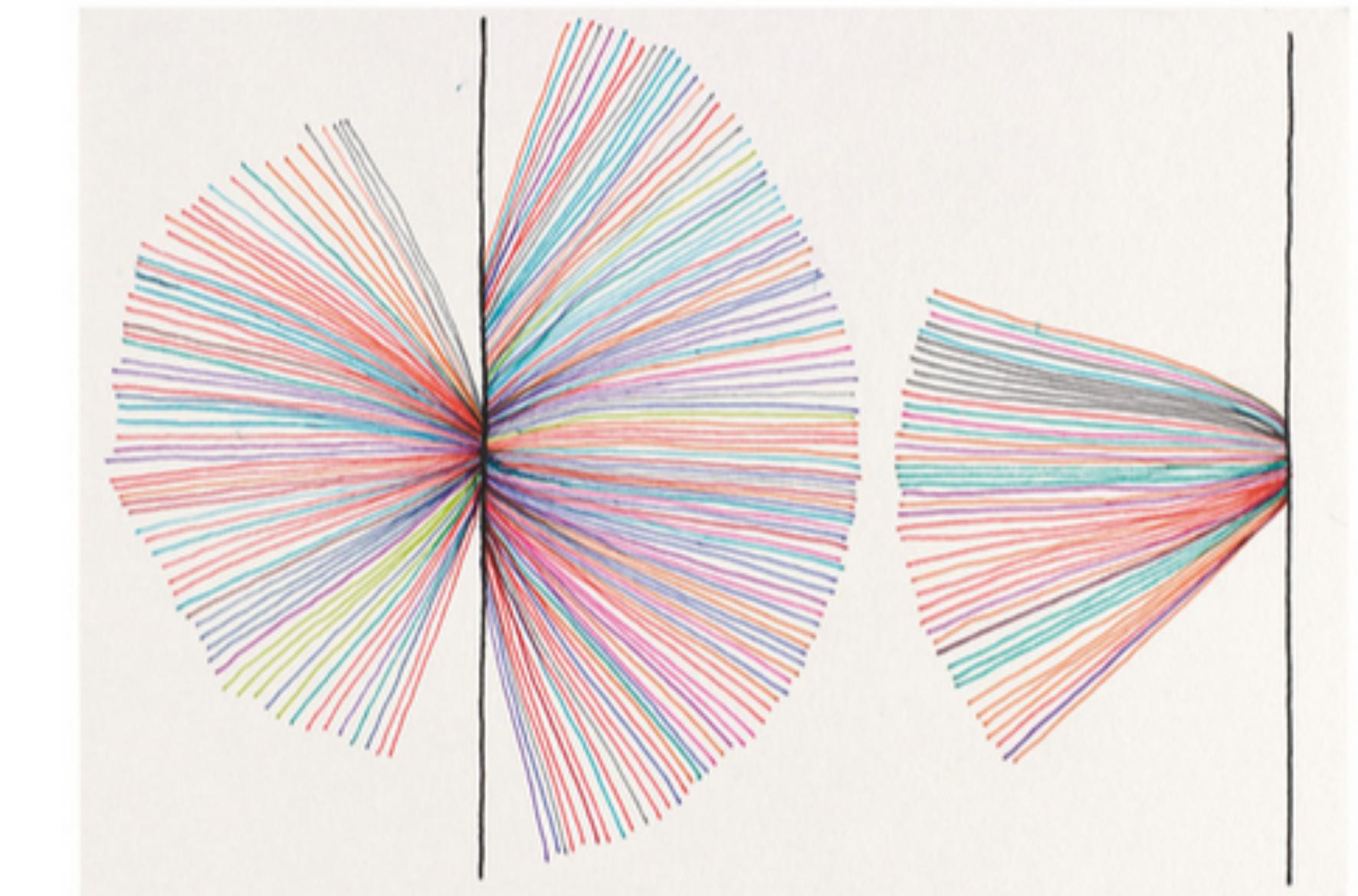
ENGLAND

**DELIVERED BY
HAND (SPECIAL NYC
DELIVERY!)**

What better visual reference than a musical score to show the repetitiveness of Giorgia's protests and the "level" of complaint: whether they are justified or totally out of place.

a week of complaints

Stefanie



DEAR DATA - WEEK 07

A WEEK OF COMPLAINTS*
AND GENERAL GRUMPINESS

HOW TO READ IT: (E ~~THREW AWAY~~ THREW DANNY MY PENS WHEN I FINISHED
(COMPLAINT #7) WHAT IS WRONG WHEN?)

ALL COMPLAINTS IN CHRONOLOGICAL ORDER,
EXCEPT "COMPLAINTS TO ME"
BECAUSE ((COMPLAINT #1))

MESSED UP THE ORDER +
BY THAT TIME I HAD
ALREADY SPENT 1.5 HOURS
ON THIS... ARGH! AND
(COMPLAINT #2) THIS PART
OF THE DRAWING IS NOT
SYMMETRICAL WHICH
BOTHERS ME :)

TYPE OF COMPLAINT:
PRIVATE OUTWARD COMPLAINTS
COMPLAINING COMPLAINTS TO ME

WEATHER HEALTH
HUSBAND HUNGER
ANIMALS MYSELF
FAMILY TECHNOLOGY/
MEDIA
SOCIETY MONEY
THE WORLD TODAY INHABITANT
ACQUAINTANCES OBJECTS
STRANGERS TRANSPORT
MY APPEARANCE
FRIENDS
WORK

MAIN STATS

PRIVATE COMPLAINTS: 67
OUTWARD COMPLAINTS: 100

COMPLAINTS TO ME: 43

NO OF PEOPLE WHO
COMPLAINED TO ME: 5

NO OF PEOPLE I
COMPLAINED TO: 10

COMPLAINTS MOST RECENT
TO ME: (SOMEBODY'S DAD)
YOU YELLING!

* SECOND MOST COMPLAINTS ABOUT
MY HUSBAND: (SOMEBODY'S
DAD)

* AND A WEEK OF COMPLAINTS AFTER WHEN I FRAMED UP THIS DRAWING! (COMPLAINT #5)

FROM:
S POSAVEC
LONDON
UK



TO:
GIORGIA LUPI

BROOKLYN, NY

USA

**DELIVERED BY
HAND (SPECIAL NYC
DELIVERY!)**

Note the hand-drawn stamps: these postcards were delivered in person in New York!

What is Consciousness?

Make up your own mind



A field that exists in its own parallel "realm" of existence outside reality so can't be seen.
(Substance Dualism)



A sensation that "grows" inevitably out of complicated brain states.
(Emergent Dualism)



A physical property of all matter, like electromagnetism, just not one the scientists know about.
(Property Dualism)



All matter has a psychic part. Consciousness is just the psychic part of our brain.
(Pan Psychism)



Simply, mental states are physical events that we can see in brain scans.
(Identity Theory)



Consciousness and its states (belief, desire, pain) are simply functions the brain performs.
(Functionalism)



Literally just behaviour. When we behave in a certain way, we appear conscious.
(Behaviourism)



An accidental side-effect of complex physical processes in the brain.
(Epiphenomenalism)

HOW LONG CAN YOU WORK ON MAKING A ROUTINE TASK MORE
EFFICIENT BEFORE YOU'RE SPENDING MORE TIME THAN YOU SAVE?
(ACROSS FIVE YEARS)

		HOW OFTEN YOU DO THE TASK					
		50/DAY	5/DAY	DAILY	WEEKLY	MONTHLY	YEARLY
HOW MUCH TIME YOU SHAVE OFF	1 SECOND	1 DAY	2 HOURS	30 MINUTES	4 MINUTES	1 MINUTE	5 SECONDS
	5 SECONDS	5 DAYS	12 HOURS	2 HOURS	21 MINUTES	5 MINUTES	25 SECONDS
	30 SECONDS	4 WEEKS	3 DAYS	12 HOURS	2 HOURS	30 MINUTES	2 MINUTES
	1 MINUTE	8 WEEKS	6 DAYS	1 DAY	4 HOURS	1 HOUR	5 MINUTES
	5 MINUTES	9 MONTHS	4 WEEKS	6 DAYS	21 HOURS	5 HOURS	25 MINUTES
	30 MINUTES	6 MONTHS	5 WEEKS	5 DAYS	1 DAY	2 HOURS	
	1 HOUR	10 MONTHS	2 MONTHS	10 DAYS	2 DAYS	5 HOURS	
	6 HOURS			2 MONTHS	2 WEEKS	1 DAY	
	1 DAY				8 WEEKS	5 DAYS	

Afweging 2: confidence

- Absolute vs relative numbers
- Confidence comes with higher numbers

Errorbars

- Standard error
- Confidence intervals
- Standard deviation
- Range

Errorbars

Suppose statistically independent n samples x_1, x_2, \dots, x_n from a population

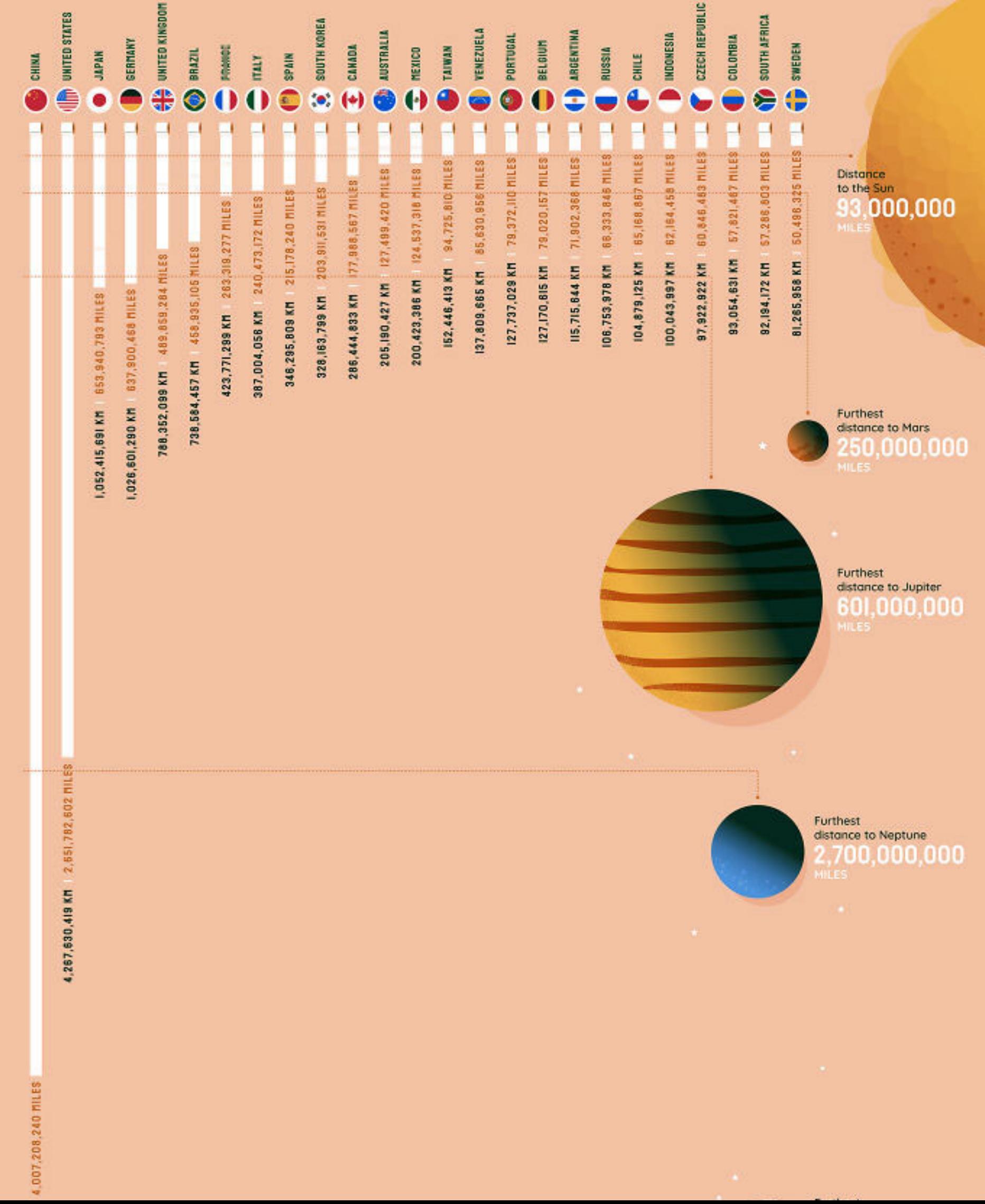
Suppose the population has a standard deviation of σ

The standard error is calculated with $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Because we rarely know the σ of the population, we often estimate it with the sample standard deviation.

THE COUNTRIES THAT USE THE MOST TOILET PAPER

► TOILET PAPER USAGE PER YEAR PER COUNTRY



DEATH TOLL

[HIGHEST TO LOWEST]

the disease is new to medicine,
and data is still coming in.

200M

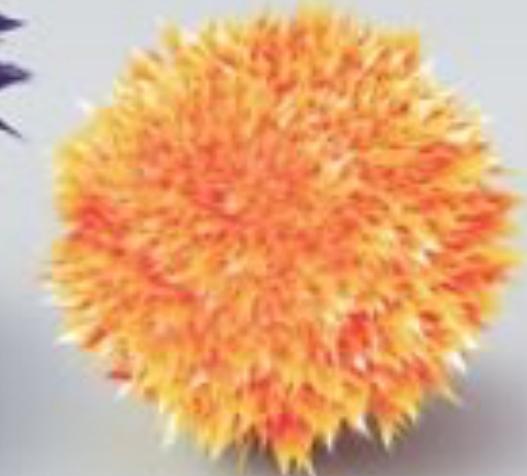
Black Death (Bubonic Plague)
1347-1351



The plague originated
in rats and spread to
humans via infected fleas.

56M

Small Pox
1520



The outbreak wiped
out 30-50% of Europe's
population. It took more than
200 years for the continent's
population to recover.

40-50M

Spanish Flu
1918-1919



30-50M

Plague of Justinian
541-542



Smallpox killed an estimated 90% of
Native Americans. In Europe during the
1800s, an estimated 400,000 people
were being killed by smallpox annually.
The first ever vaccine was created to
ward off smallpox.

The death toll of this plague
is still under debate as new
evidence is uncovered, but
many think it may have
helped hasten the fall of
the Roman Empire.



25-35M
HIV/AIDS
1981-PRESENT



12M
The Third Plague
1855



5M
Antonine
Plague
165-180



3M
17th Century
Great Plagues
1665



1.1M
Asian Flu
1957-1958



1M
Russian Flu
1889-1890



1M
Hong Kong Flu
1968-1970



1M
Cholera 6
outbreak
1817-1923

A series of Cholera outbreaks spread
around the world in the 1800s killing
millions of people. There is no solid
consensus on death tolls. ▾

1M
Japanese
Smallpox Epidemic
735-737

600K
18th Century
Great Plagues
1817-1923

200K
Swine Flu
2009-2010

100-150K
Yellow Fever
LATE 1800s

11.3K
Ebola
2014-2016

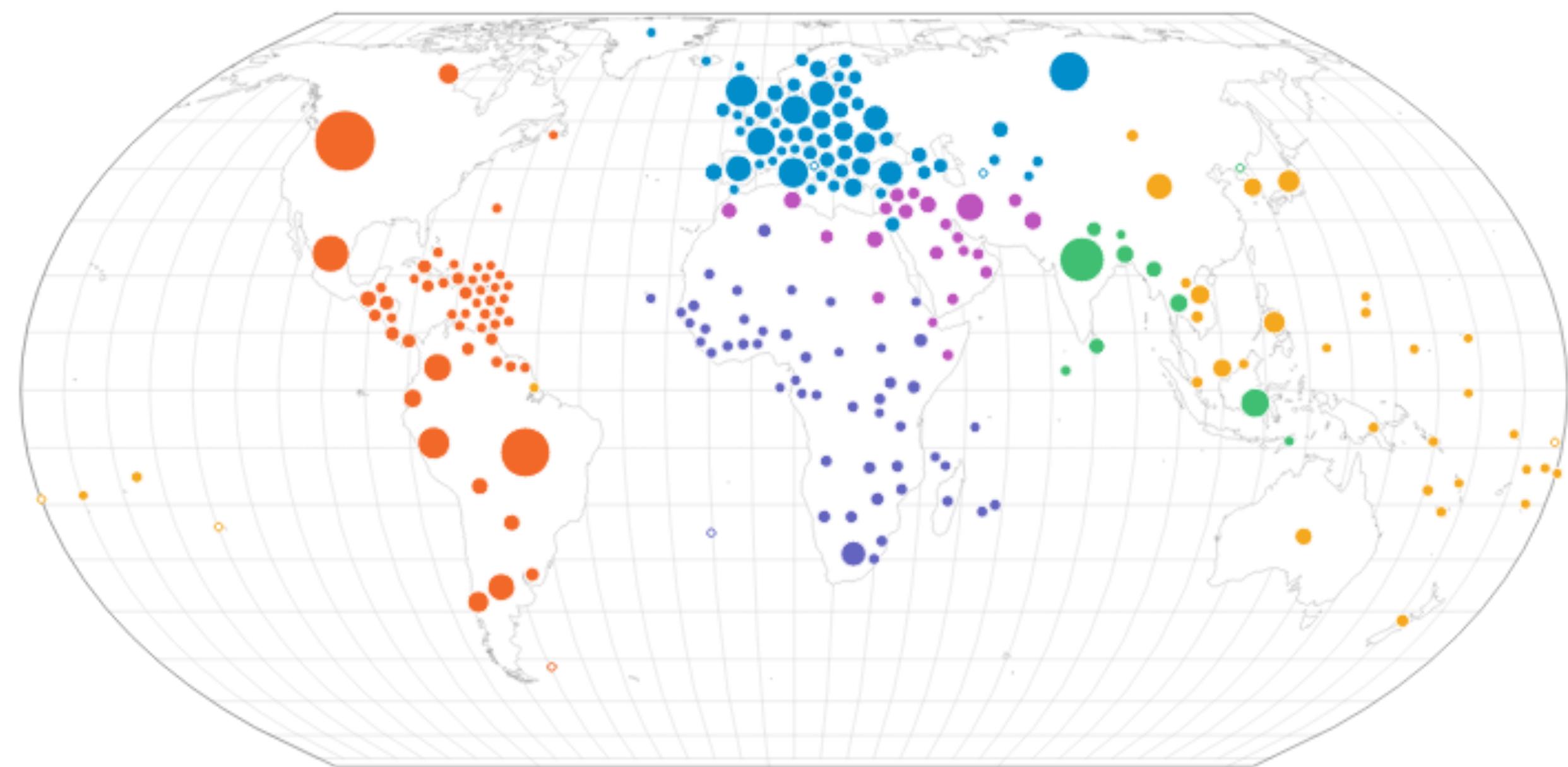
850
MERS
2015-PRESENT

770
SARS
2002-2003

4.7K*
Novel Coronavirus (COVID-19)
2019-PRESENT

Number of COVID-19 deaths reported to WHO (cumulative total)

World



7,031,216
Reported COVID-19 deaths

11 February 2024

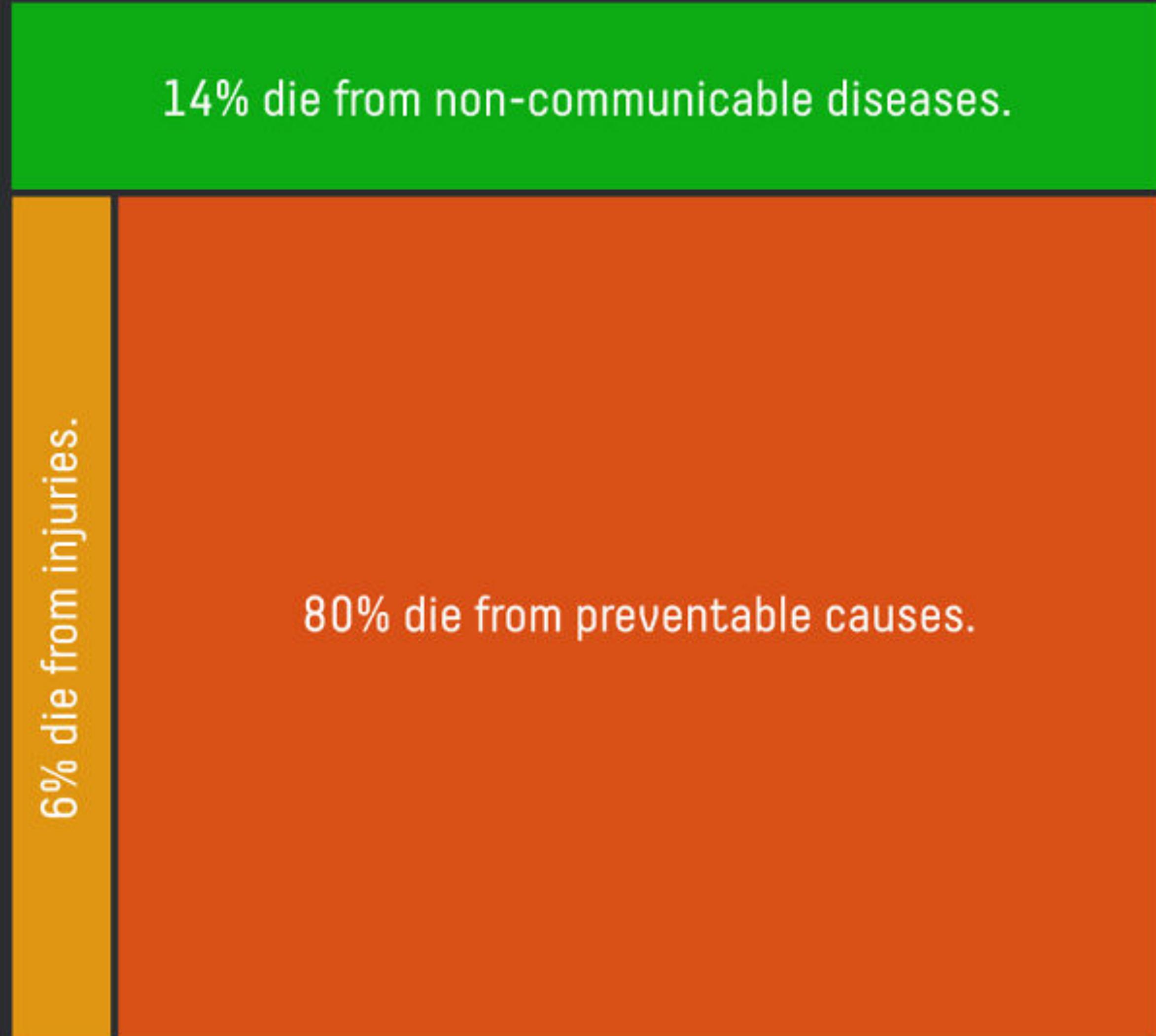
Number of COVID-19 deaths reported to WHO (cumulative total)

World

Country	Deaths
United States of America	1.2m
Brazil	702.1k
India	533.5k
Russian Federation	402.1k
Mexico	335k
United Kingdom of Great Britain And	232.1k

Show less

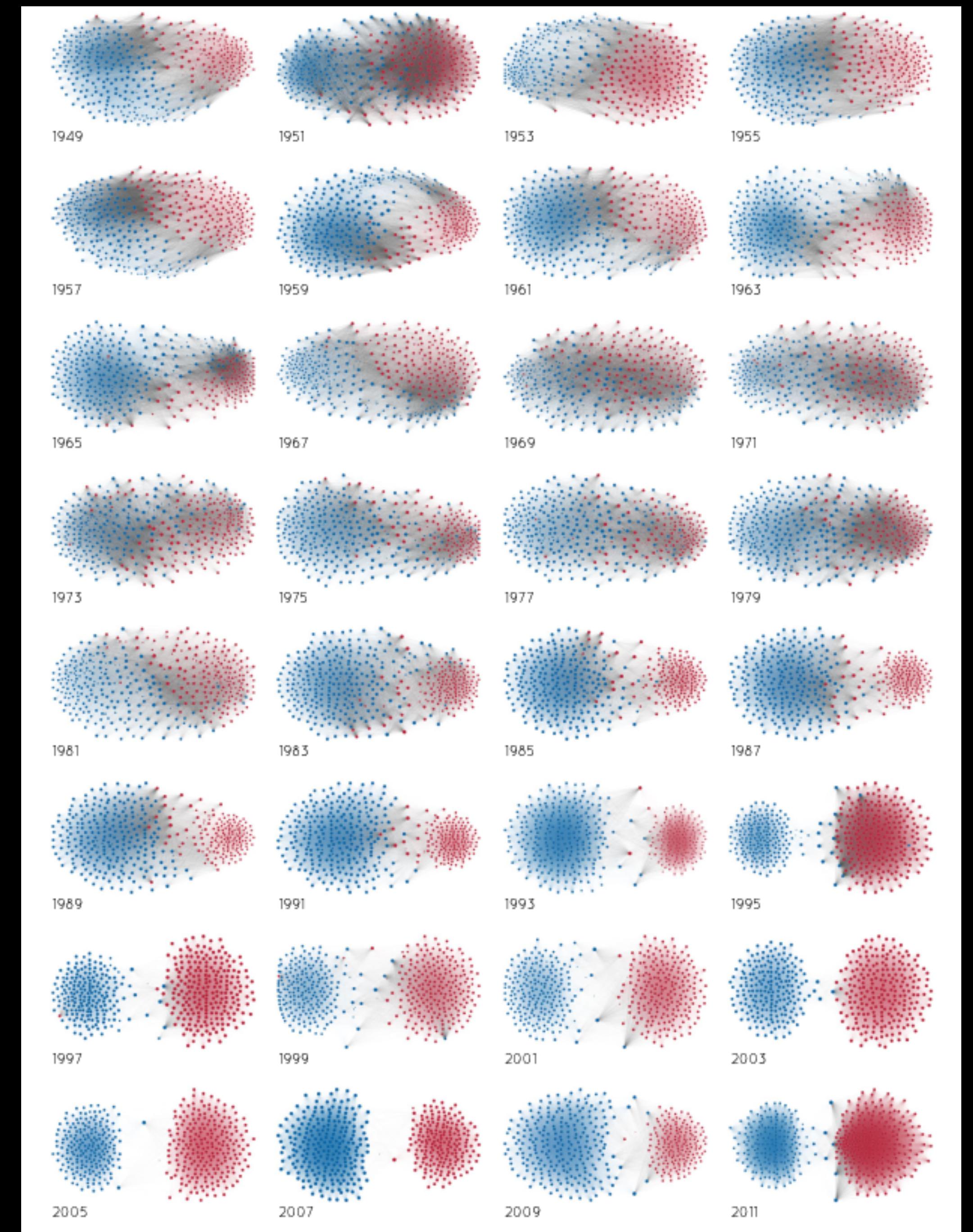
Causes of death in children under 5 (2013)



Source: IHME

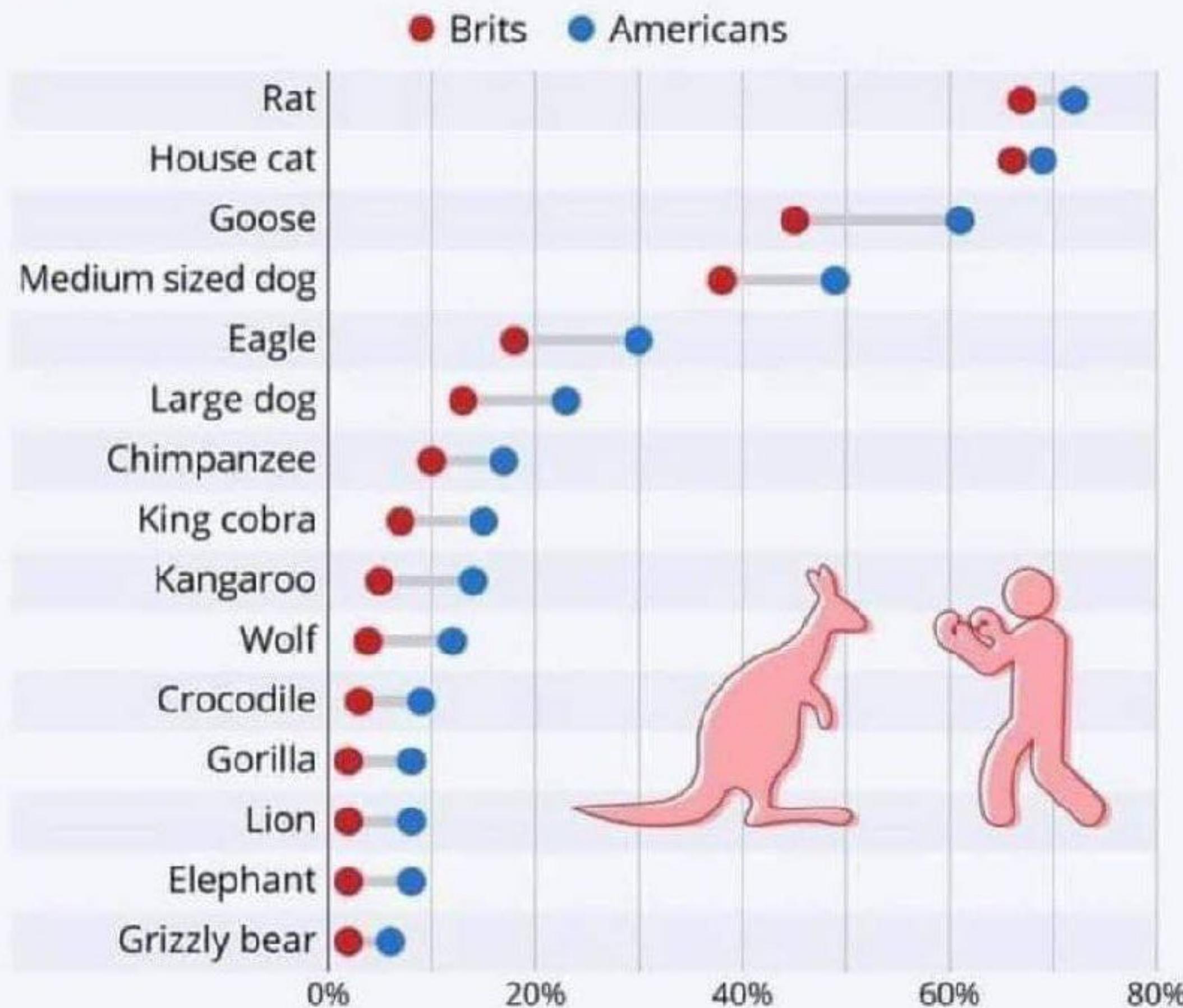
gatesnotes

The Rise of Partisanship and Super-Cooperators in the U.S. House of Representatives



Which Animals Could You Beat in a Fight?

"Which of the following animals, if any, do you think you could beat in a fight if you were unarmed?"



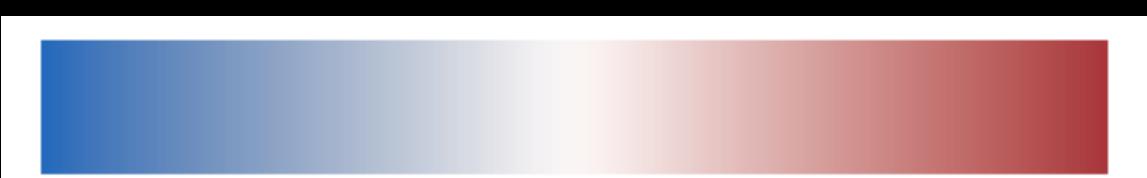
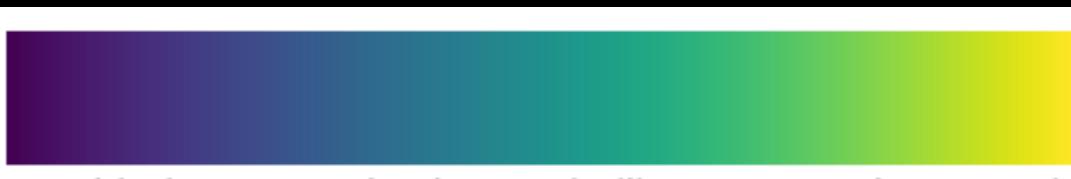
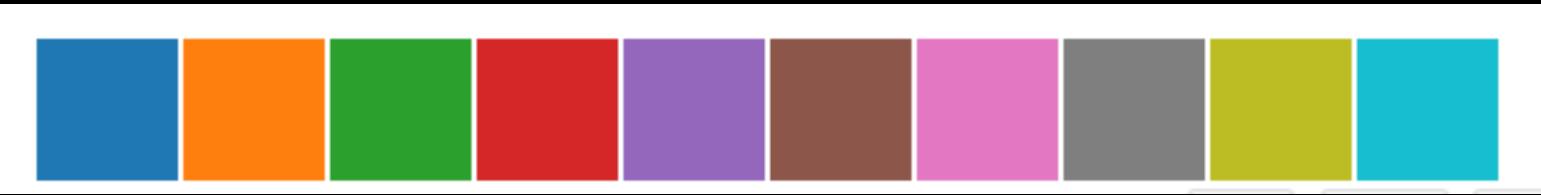
Survey of 2,082 GB adults (conducted 18-19 May 2021) & 1,224 U.S. adults (conducted 12-13 April 2021).

Source: YouGov

Color palettes

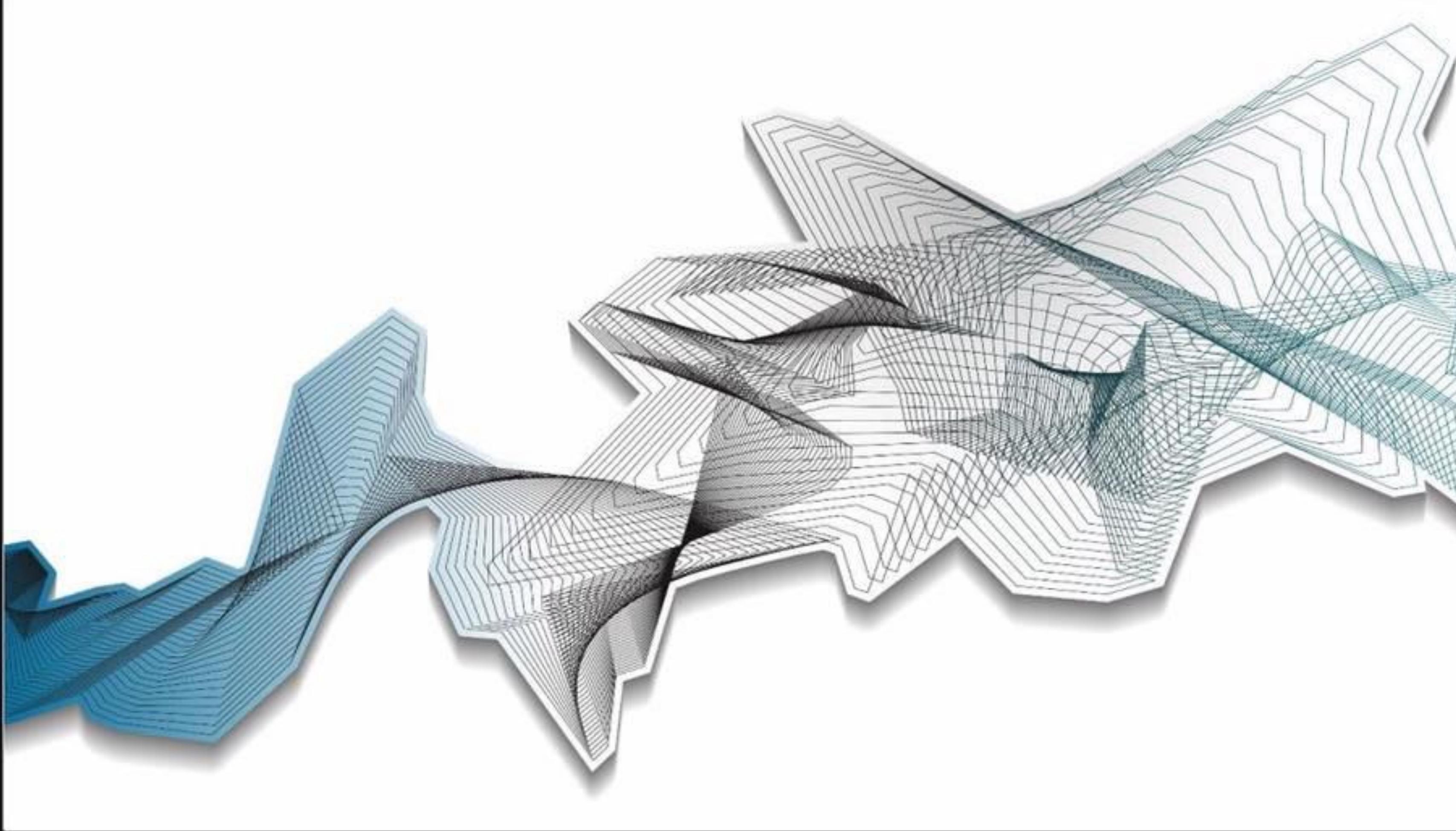
https://seaborn.pydata.org/tutorial/color_palettes.html

- **Qualitative:** representing categorical data
- **Sequential:** perceptually uniform
- **Diverging:** both large low and high values are interesting and span a midpoint value



Git

- The only sane way to work together on code
- <https://learngitbranching.js.org/>



Version Control

Git gets easier once you get the basic idea that branches are homeomorphic endofunctors mapping submanifolds of a Hilbert space.

