



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

96801 - LANGUAGE LABORATORY: COMMUNICATION OF STATISTICS AND DATA BUSINESS ANALYTICS

LESSON 2 - 12/11/2025

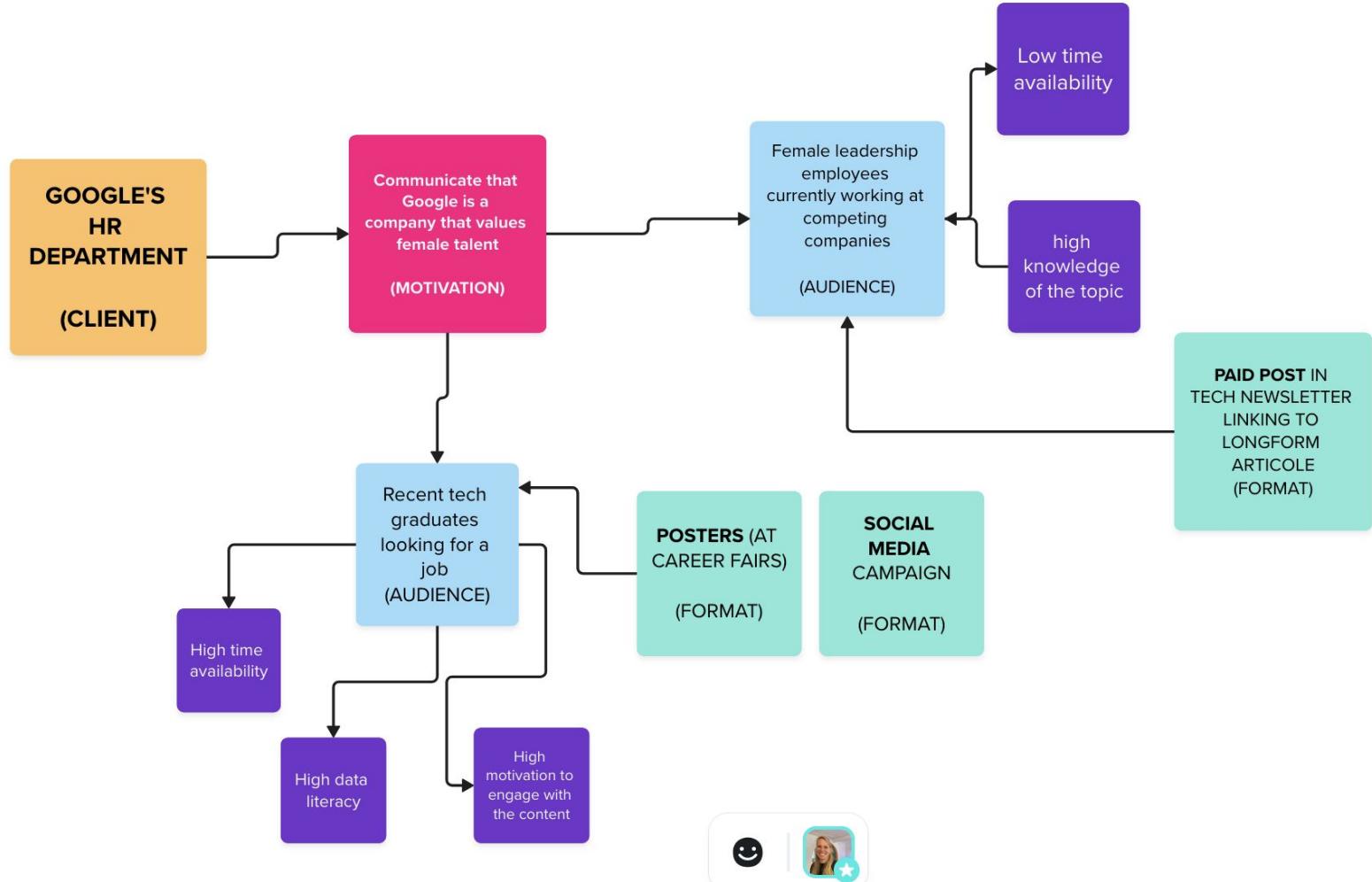
Alice Corona

Dipartimento di Scienze Statistiche "Paolo Fortunati"

EXERCISE → A. THINKING OF A DATA COMMUNICATION SCENARIO

Pick one of the following datasets. Think about potential data communication scenarios that could be associated with the data.

- List the **possible clients** that might be interested in communicating the data, along with their **motivations**
- List the **possible audience segments** to whom the clients might be interested in communicating the data, along with the **characteristics** of each segment
- Think about what **data storytelling formats** might be suitable for each audience segment/motivation



THE EXAM

THE DELIVERABLES

- **1 PROJECT PITCH**

In the last in-person lesson, you will present to the whole class a pitch with your idea of the project. The pitch will be of about 5 slides. A template will be provided, and there will be a lesson explaining how to craft a pitch.  Deadline: 11/12/2025

- **THE ACTUAL PROJECT**

A presentation, a PDF report, a series of Instagram graphics..whatever you choose as the appropriate format. You will receive a Google Form survey to submit the exam and the project pitch. Each team member has to answer the Google Form (with the same files). Deadline: chosen exam session

THE EXAM

WHAT TO INCLUDE:

- **Cover slide:** names of participants
- **SLIDE 1:** The scenario
- **SLIDE 2:** Data sources (where will you find the data)
- **SLIDE 3:** About the audience (characteristics & needs)
- **SLIDE 4:** Data questions (what questions about the data would be relevant?)
- **SLIDE 5:** The format and potential structure
- **OPTIONAL EXTRAS:** Interesting insights or viz ideas

THE EXAM

WHAT TO INCLUDE:

- **Cover slide:** names of participants
- **SLIDE 1:** The scenario
- **SLIDE 2:** Data sources (where will you find the data)
- **SLIDE 3:** About the audience (characteristics & needs)
- **SLIDE 4:** Data questions (what questions about the data would be relevant?)
- **SLIDE 5:** The format and potential structure
- **OPTIONAL EXTRAS:** Interesting insights or viz ideas

THE EXAM

SLIDE 1: The scenario

We work in the HR department of Google and are asked to prepare a presentation on Google's employee current diversity and on strategies to improve it. We will deliver the presentation to a committee of senior-level members of the department.

[INCLUDES: context/client/employer, audience, stakeholders, communication goal]

THE EXAM

WHAT TO INCLUDE:

- **Cover slide:** names of participants
- **SLIDE 1:** The scenario
- **SLIDE 2: Data sources (where will you find the data)**
- **SLIDE 3:** About the audience (characteristics & needs)
- **SLIDE 4:** Data questions (what questions about the data would be relevant?)
- **SLIDE 5:** The format and potential structure
- **OPTIONAL EXTRAS:** Interesting insights or viz ideas

THE EXAM

WHAT TO INCLUDE:

- **Cover slide:** names of participants
- **SLIDE 1:** The scenario
- **SLIDE 2:** Data sources (where will you find the data)
- **SLIDE 3: About the audience (characteristics & needs)**
- **SLIDE 4:** Data questions (what questions about the data would be relevant?)
- **SLIDE 5:** The format and potential structure
- **OPTIONAL EXTRAS:** Interesting insights or viz ideas

THE EXAM

SLIDE 4: The audience

AUDIENCE	AUDIENCE CHARACTERISTICS	AUDIENCE NEEDS	DESIRED OUTCOME	MAIN MESSAGE
Segment #1	<i>of the topic of the medium; of the format <u>formato</u>; data literacy <u>graphicacy</u></i>	<i>time availability; attention span; interest <u>for</u> the topic; interest <u>for</u> the angles; motivation; level of detail required;</i>	ACCEPT AGREE APPROVE BUY CONDAMN INVEST LEARN PAY FOR PERSUADE RAISE AWARENESS REMEMBER SHARE SUPPORT UNDERSTAND VOTE	

THE EXAM

WHAT TO INCLUDE:

- **Cover slide:** names of participants
- **SLIDE 1:** The scenario
- **SLIDE 2:** Data sources (where will you find the data)
- **SLIDE 3:** About the audience (characteristics & needs)
- **SLIDE 4:** Data questions (what questions about the data would be relevant?)
- **SLIDE 5:** The format and potential structure
- **OPTIONAL EXTRAS:** Interesting insights or viz ideas

THE EXAM

SLIDE 5: Data questions

- What is the rank position of Google, compared to other 10 tech companies, for share of employees working in tech?
- How are Google's male and female employees distributed along the seniority ladder (junior - mid level - senior) & is there a gender gap in this distribution?
- What is the ratio of female new hires, out of total new hires, in the last 5 years? Has it improved over time & is it enough to close the gender gap among tech employees?
- How many years will it take for Google to close the gender gap among tech employees, with the current rate of new female hires?
- From what universities are Google's competitors hiring female tech talent?

THE EXAM

WHAT TO INCLUDE:

- **Cover slide:** names of participants
- **SLIDE 1:** The scenario
- **SLIDE 2:** Data sources (where will you find the data)
- **SLIDE 3:** About the audience (characteristics & needs)
- **SLIDE 4:** Data questions (what questions about the data would be relevant?)
- **SLIDE 5: The format and potential structure**
- **OPTIONAL EXTRAS:** Interesting insights or viz ideas

THE EXAM

WHAT TO INCLUDE:

- **Cover slide:** names of participants
- **SLIDE 1:** The scenario
- **SLIDE 2:** Data sources (where will you find the data)
- **SLIDE 3:** About the audience (characteristics & needs)
- **SLIDE 4:** Data questions (what questions about the data would be relevant?)
- **SLIDE 5:** The format and potential structure
- **OPTIONAL EXTRAS:** Interesting insights or viz ideas

BEFORE DEC. 1st

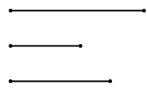
Fill in the survey about your group project.

VISUALIZING DATA

Introduction to data viz design, choosing a chart type, using text

**WHAT IS A
DATA
VISUALIZATION
?**

VISUAL ENCODINGS (MARKS & ATTRIBUTES)



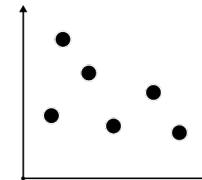
LENGTH



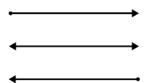
AREA



HUE



POSITION



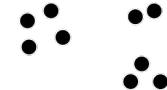
DIRECTION



VOLUME



LIGHT/SATURATION



PROXIMITY



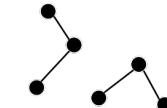
ANGLE/CURVE
& SLOPE



SHAPE



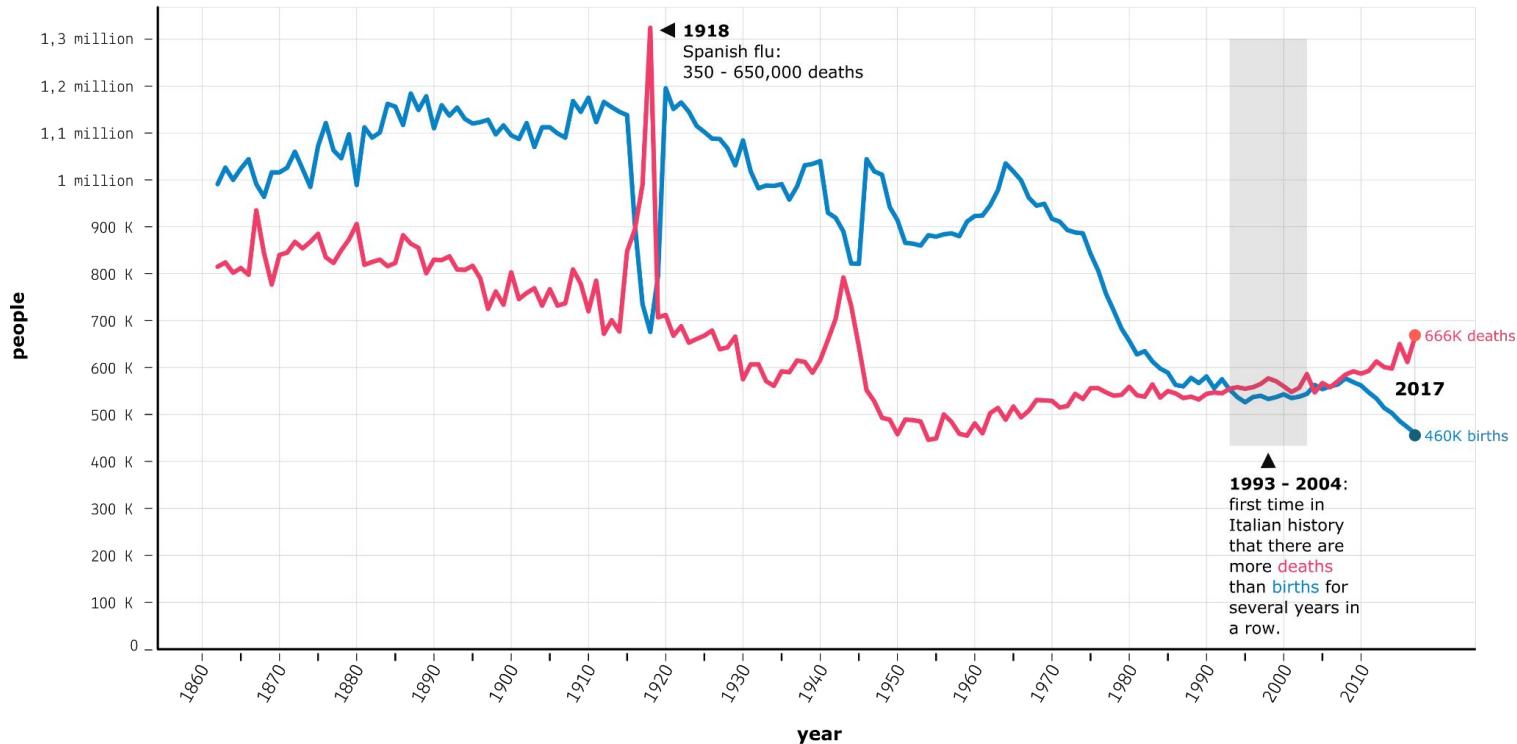
TEXTURE



CONNECTION

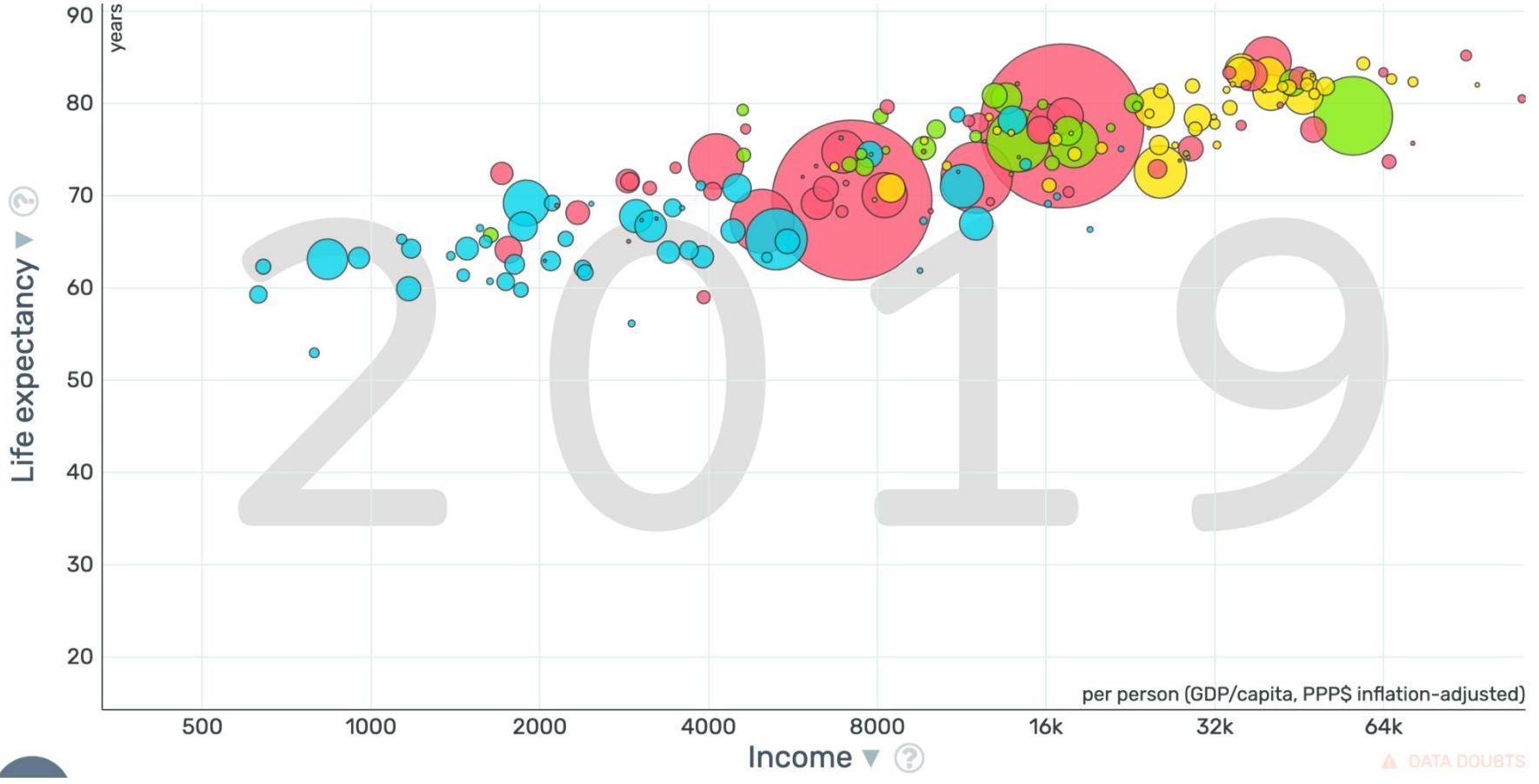
Births and Deaths in Italy

1861 - 2017



Fonte: Istat · Changes in Italian population from 1861 to today (*I cambiamenti della popolazione italiana dal 1861 a oggi*). Data updated on 18/12/2018



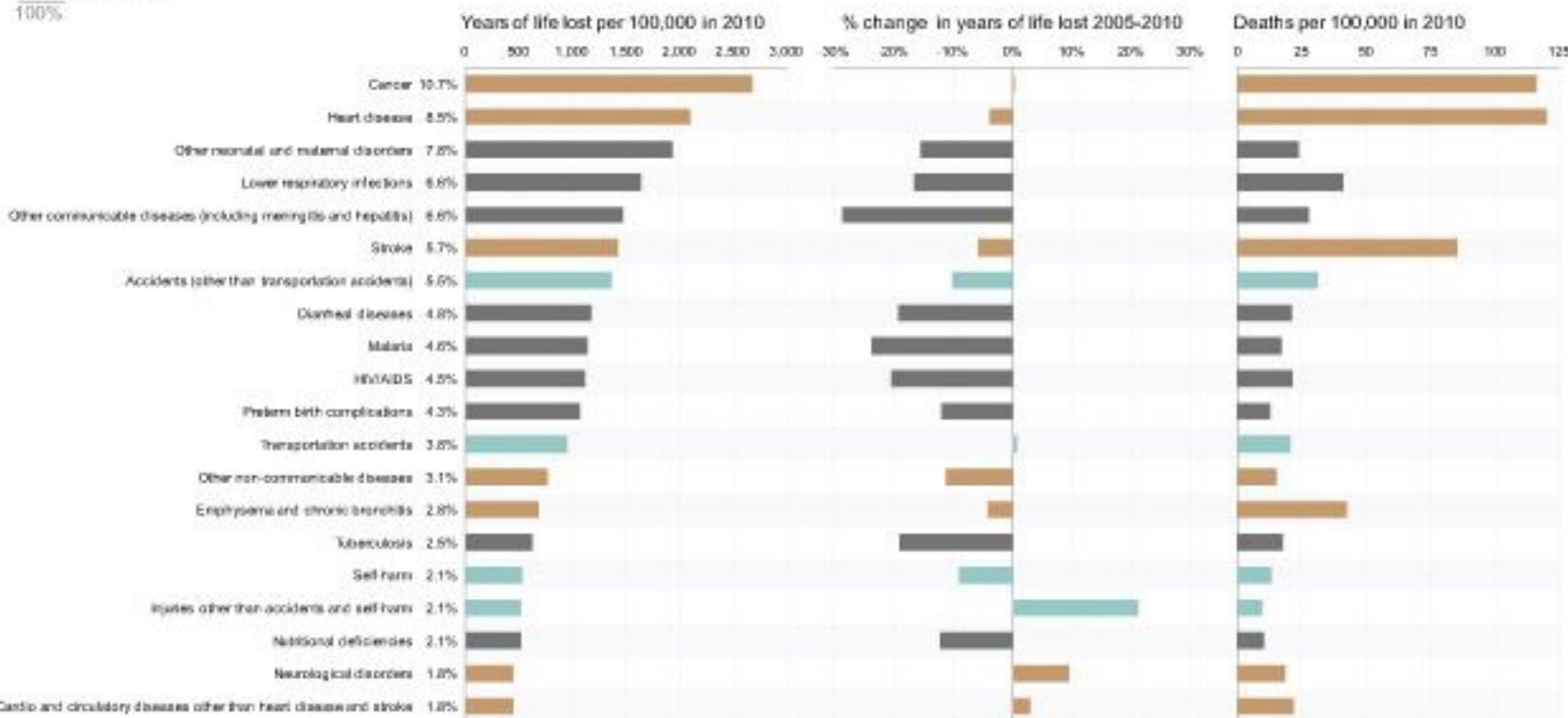


GAPMINDER

Global Causes of Lost Life

44% ■ Communicable, maternal, neonatal, and nutritional disorders
 43% ■ Non-communicable diseases
 13% ■ Injuries

Comparing the number of deaths alone, as shown in the rightmost graph below, doesn't tell the entire story. Some causes of death have a greater effect on the young, which can be seen when comparing years of life lost in the leftmost graph.



VISUAL ENCODINGS (MARKS & ATTRIBUTES)



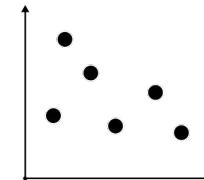
LENGTH



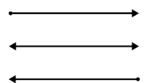
AREA



HUE



POSITION



DIRECTION



VOLUME



LIGHT/SATURATION



PROXIMITY



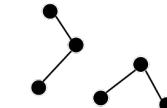
ANGLE/CURVE
& SLOPE



SHAPE



TEXTURE



CONNECTION

Vincent van Gogh moves to England to work at the London branch of Goupil & Cie

Legend

Categories of terms

(*) color = category

size = frequency
in the letter

- family, affections
- human figure
- work
- geography
- painting
- religion
- money
- time
- other artists
- other

5
20

Letters and words

(+) total number
of written letters

(^o) total number
of written words

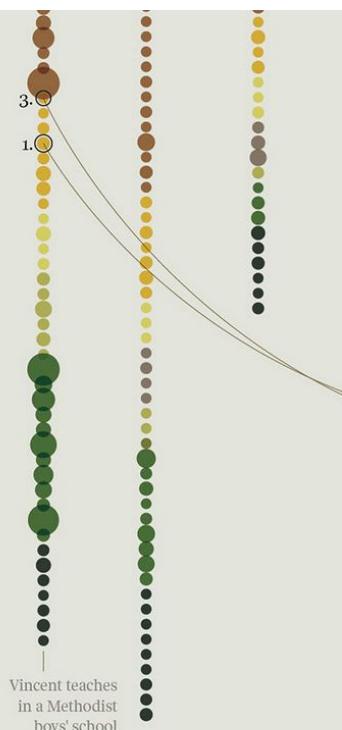
10
20
30

10.000
50.000
100.000

The 5 most mentioned names

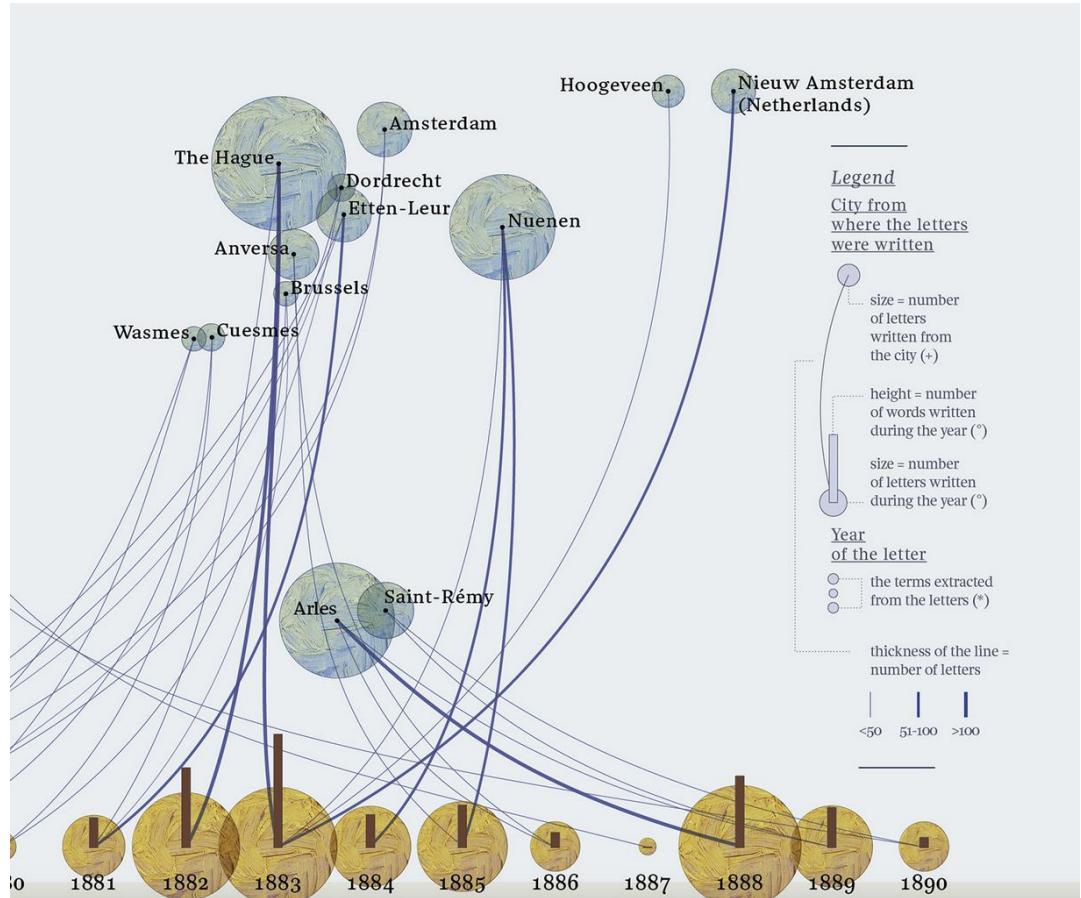
1. Theodorus Van Gogh
(Vincent Van Gogh's father)
2. Eugène Henri Paul Gauguin
(painter)
3. Anna Cornelia Carbentus
(Vincent Van Gogh's mother)
4. Jean-François Millet
(painter)
5. Clasina Maria Hoornik
(partner)

1. ○
first mention
1/. ○
last mention



Vincent teaches
in a Methodist
boys' school

The visualization shows the letters written
to his brother Theodorus «Theo».
For each year are indicated: total number of
letters written, total number of written words.



Federica Fragapane - From Vincent to Theo

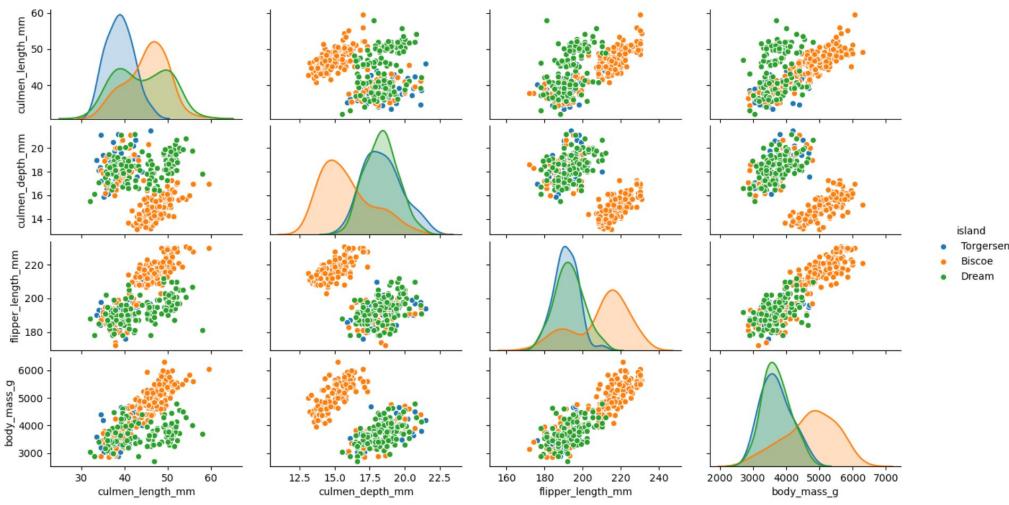
**WHY
VISUALIZE
DATA
?**

EXPLORATORY DATA VISUALIZATION

- Supports the finding of insights in the data.
- Interpretation and comprehension rely mostly on the viewer.
- Requires cognitive effort → useful if the intended audience is knowledgeable and/or motivated.
- Can overlook editorial details like: captions, clear axis titles, journalistic titles, colors that highlight or convey meaning

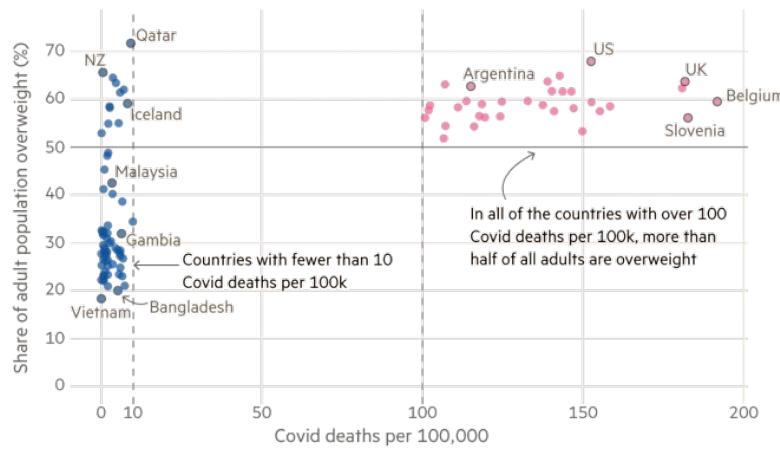
EXPLANATORY DATA VISUALIZATION

- Supports the communication of the meaning of the data.
- Visualizer takes responsibility to make clear key insights and to facilitate their interpretation.
- Editorial choices are always present, to various degrees: explicit and clear legends, journalistic titles, annotations and arrows, reference lines; smart use of color, highlights etc.



Covid's obesity connection

Prevalence of adult obesity vs Covid mortality



Sources: Johns Hopkins Coronavirus Resource Center; WHO Global Health Observatory
© FT



THINK OF DATA VIZ AS...

“The representation and presentation
of data to **facilitate** understanding”

~ Andy Kirk in *Data Visualization. A Handbook for Data Driven Design* (2016)



PERCEPTION

- Refers to the processes involved in forming information from sensory inputs.
- We are dealing with **data visualization** → we discuss specifically about the sense of sight specifically.
- Note, seeing is different from perceiving (which implies being consciously aware of what you see).



COGNITION

- Refers to the thinking processes through which you take in, make sense of and utilize information.
- Not only does it utilize information that came to the brain through perception, it also sends information (retrieved from memory and past experiences) to the areas of the brain doing the visual processing.



LONG-TERM MEMORY

You need to save information here - and be able to retrieve it - in order to understand data and use it to build knowledge, to make decisions or to craft persuading arguments, etc.

X

Theories of selective attention



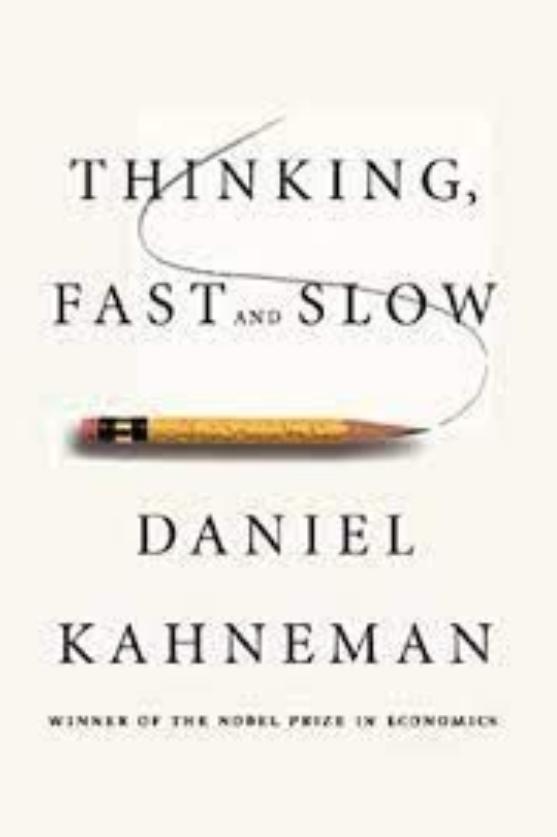
Copy link



0:03 / 4:59



Many more useful videos by Khan Academy in the unit Processing the environment, especially in the 3 modules on Attention, Memory and Cognition.



THINKING, FAST AND SLOW

DANIEL
KAHNEMAN

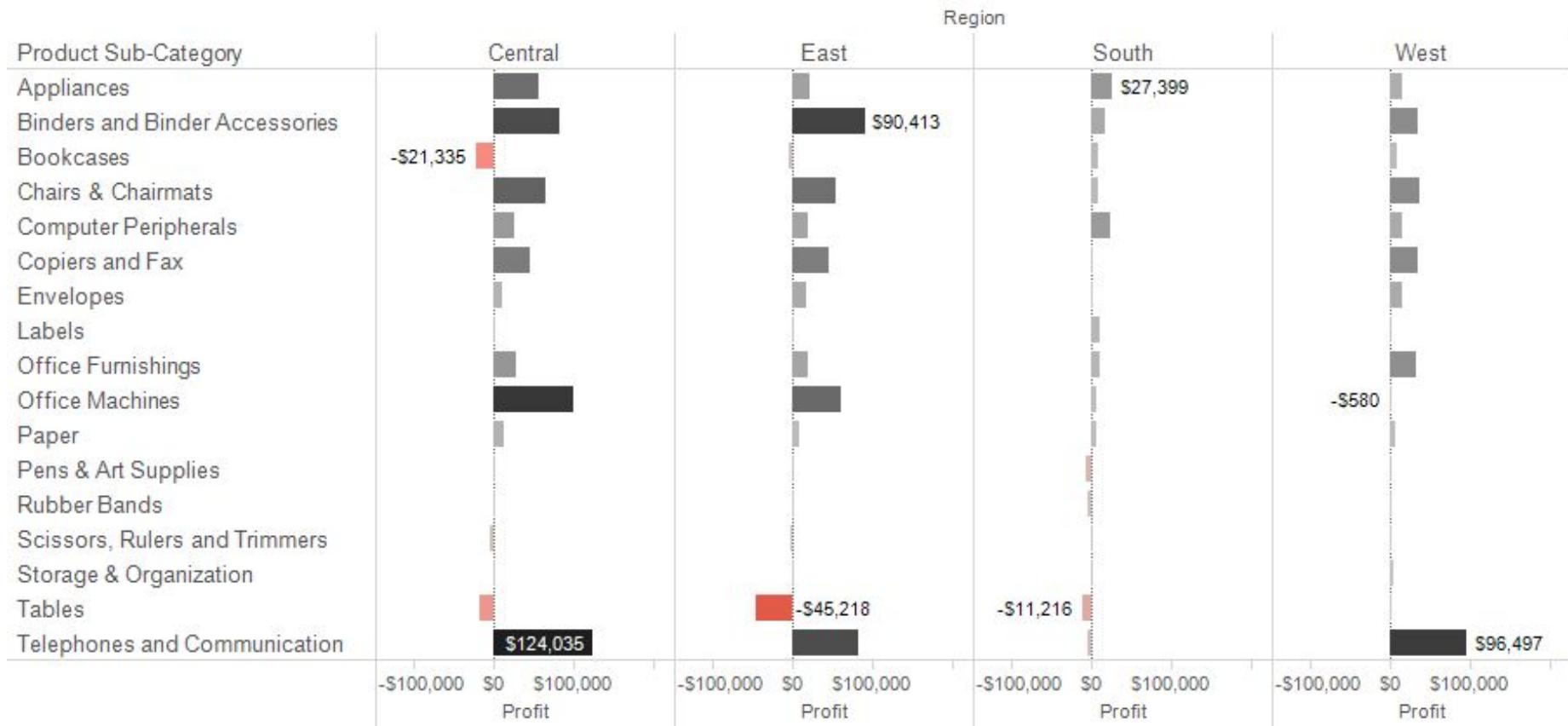
WINNER OF THE NOBEL PRIZE IN ECONOMICS

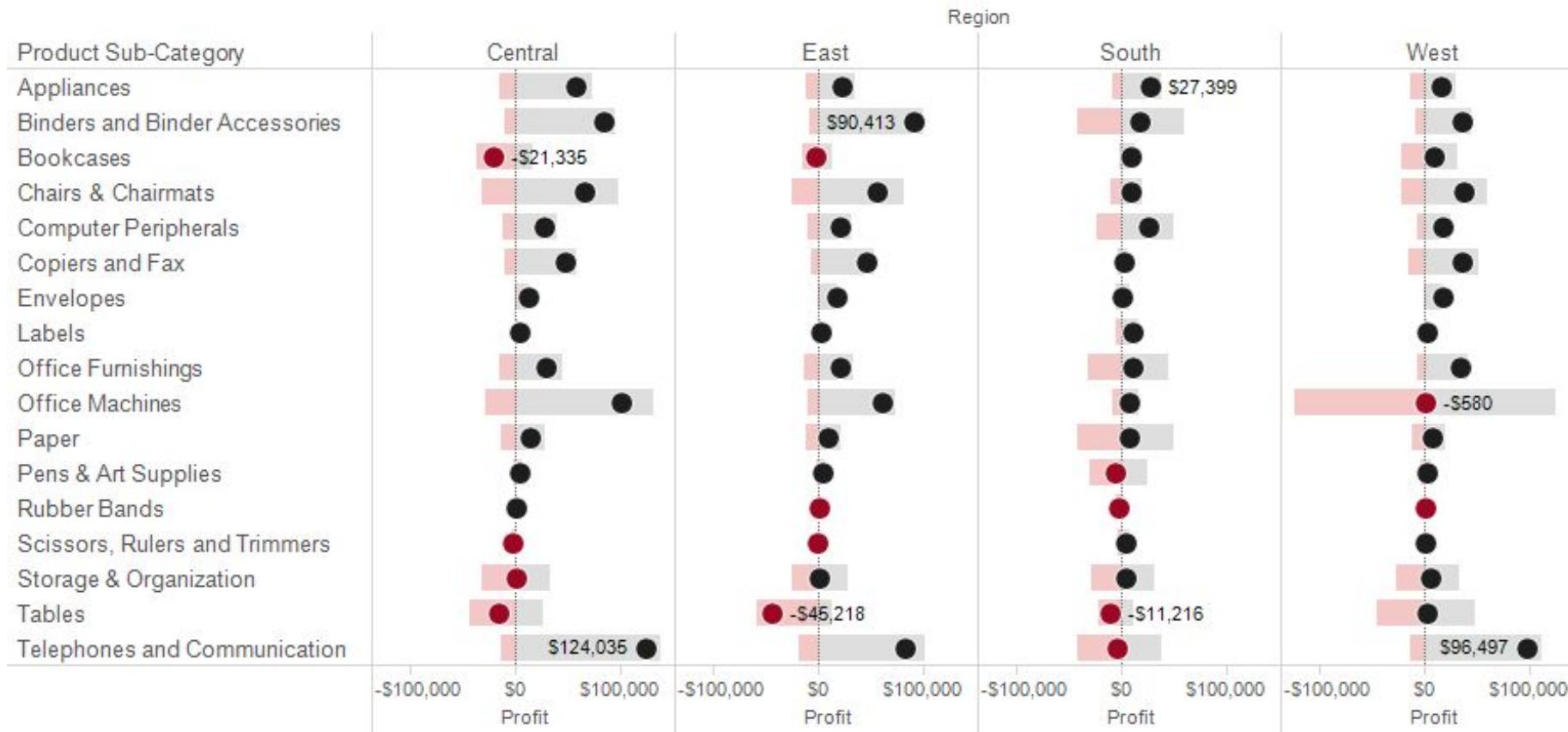
*In his mega bestseller, *Thinking, Fast and Slow*, Daniel Kahneman, the renowned psychologist and winner of the Nobel Prize in Economics, takes us on a groundbreaking tour of the mind and explains the two systems that drive the way we think.*

System 1 is fast, intuitive, and emotional; System 2 is slower, more deliberative, and more logical. The impact of overconfidence on corporate strategies, the difficulties of predicting what will make us happy in the future, the profound effect of cognitive biases on everything from playing the stock market to planning our next vacation—each of these can be understood only by knowing how the two systems shape our judgments and decisions.

Engaging the reader in a lively conversation about how we think, Kahneman reveals where we can and cannot trust our intuitions and how we can tap into the benefits of slow thinking. He offers practical and enlightening insights into how choices are made in both our business and our personal lives—and how we can use different techniques to guard against the mental glitches that often get us into trouble.

~ From the book description. *Thinking, Fast and Slow* by Daniel Kahneman





EVERY DESIGN CHOICE MUST BE

DELIBERATE.

IT IS INEVITABLY SUBJECTIVE,

BUT IT SHOULD NEVER BE

ARBITRARY.

Alberto Cairo

Visualizing data with impact (data.europa.academy)

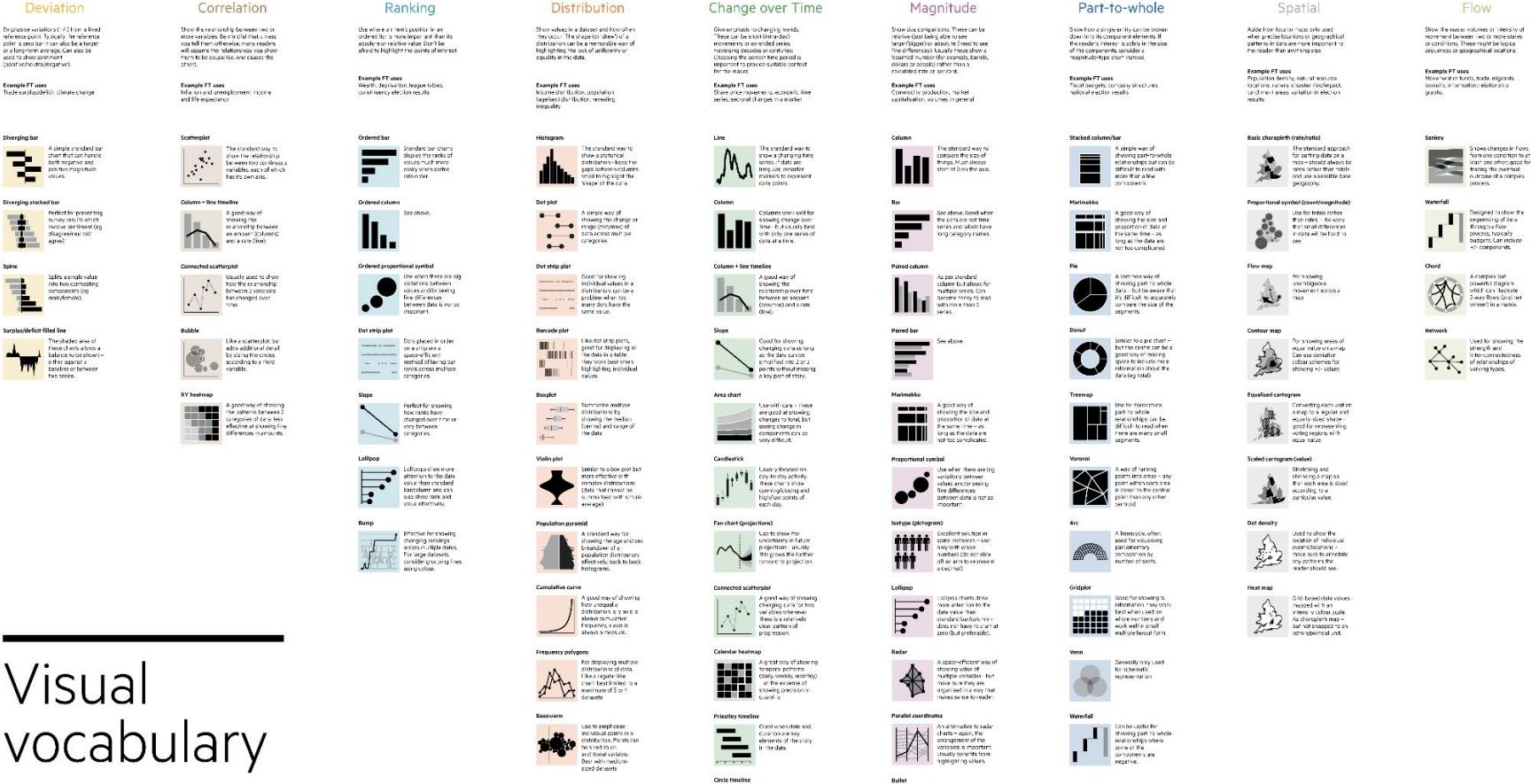
**WHAT ARE THE
CHOICES WE
HAVE TO MAKE
?**

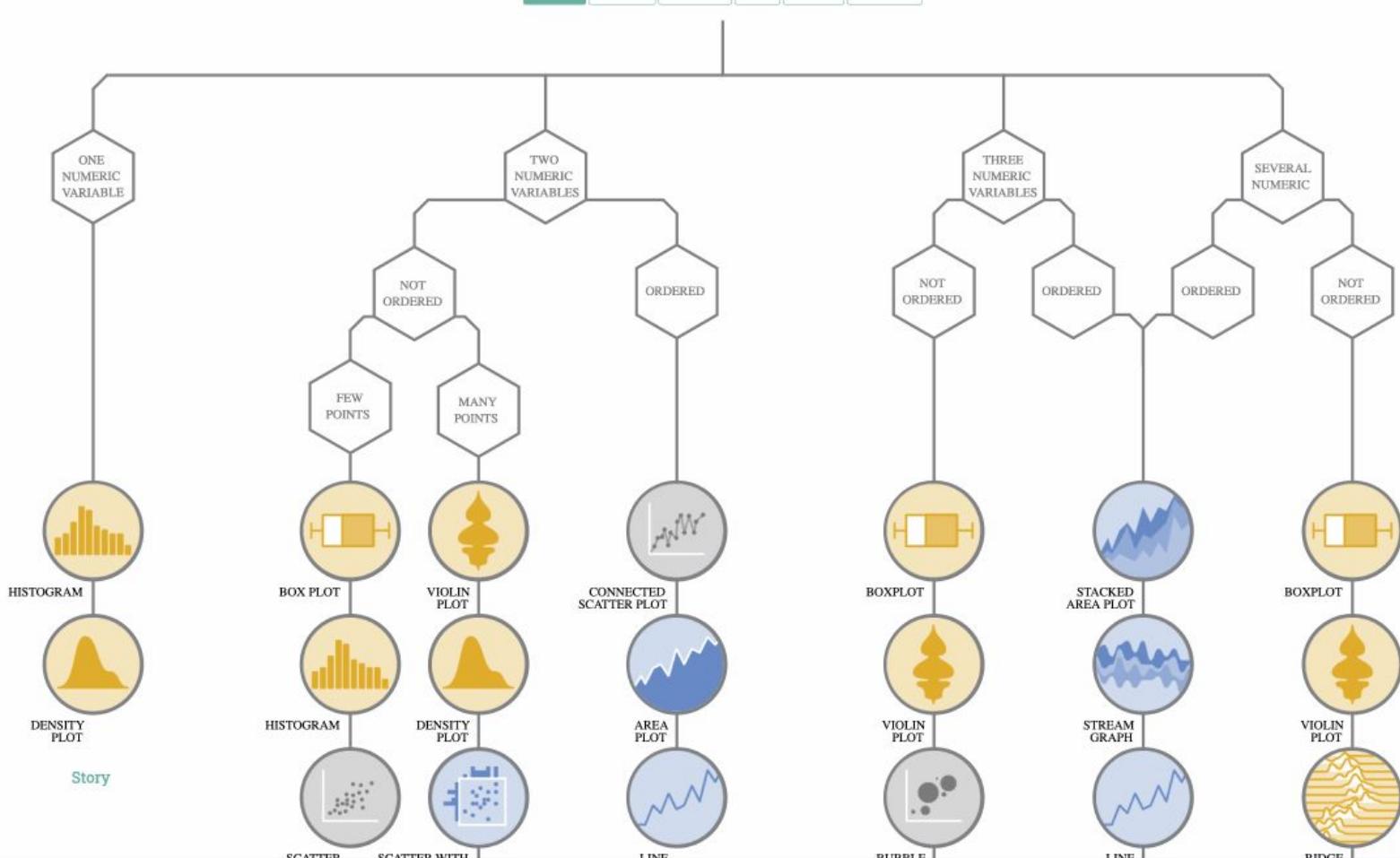
tips for choosing a

CHART TYPE

Visual vocabulary

Visual Vocabulary • Financial Times





Filter by chart name or AKA

Reference Type: ○ Example ● Solution

| Chart Families: ● Categorical ○ Hierarchical ● Relational ● Temporal ● Spatial

	Amazon QuickSight	ArcGIS	ChartJS	Charticulator	D3.js	Data Illustrator	Datawrapper	Flourish	FusionCharts	Gephi	Google Charts	Google Data Studio
Bar chart	●			●	●●●	○	●●●	○	○		●●	○
Clustered bar chart	●				●	○	●●●○	○	○		●●	
Bullet chart				●	●		●●			○		
Radar chart			○		●					○		
Polar chart			●	●	●							
Connected dot plot				●	●●	○	●●●○	●				
Pictogram					○							

[The Chartmaker Directory](#)

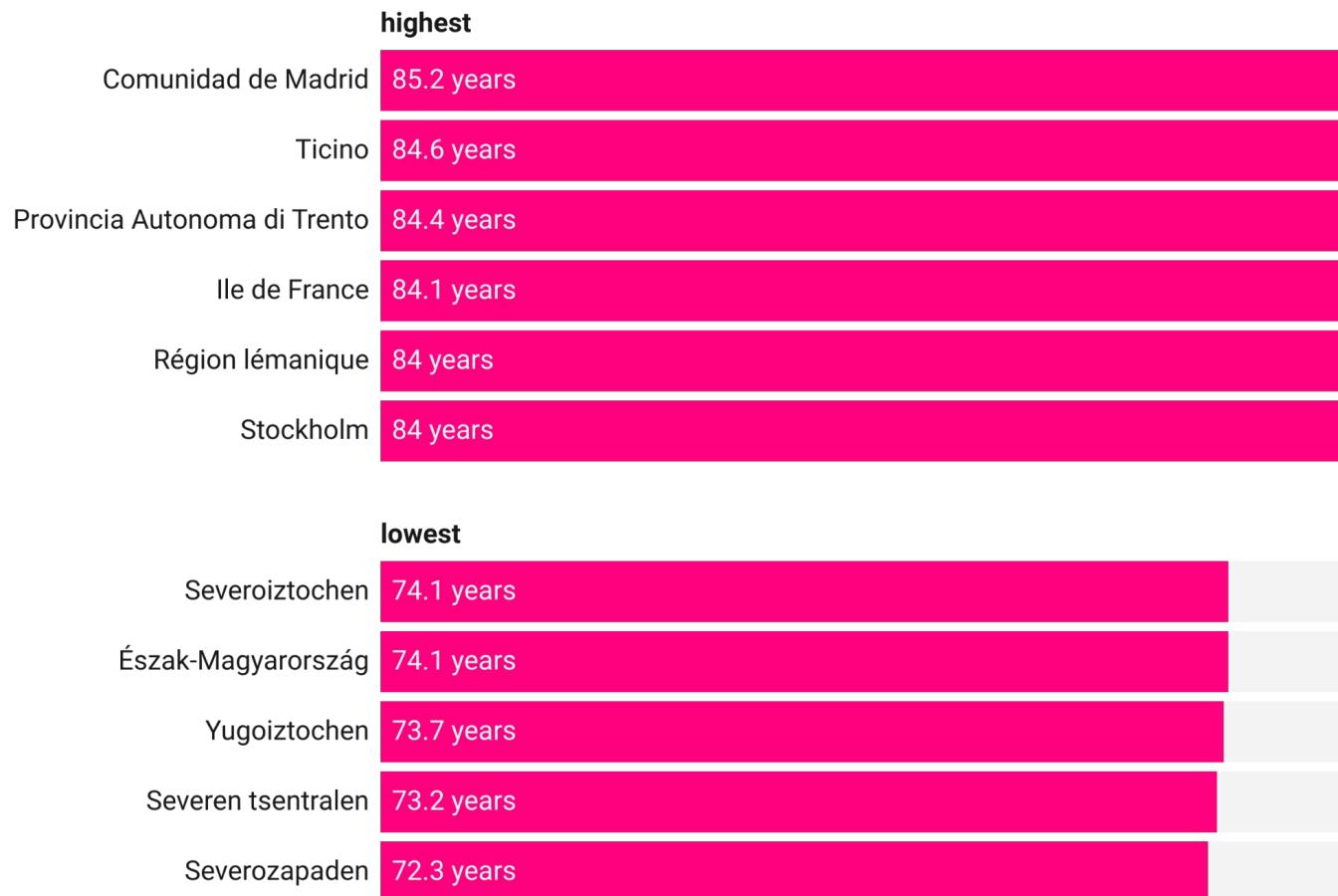
	A	B	C	D	E	F	G	H	I	J	
1	NUTS2 code	country	NUTS2 name	2011	2012	2013	2014	2015	2016	2017	2018
2	AL01	AL	Veri								
3	AL02	AL	Qender								
4	AL03	AL	Jug								
5	AT11	AT	Burgenland	78.1	77.9	78.4	79.3	78.5	79.3	79.1	
6	AT12	AT	Niederösterreich	78.0	78.2	78.3	78.8	78.7	79.0	79.4	
7	AT13	AT	Wien	77.1	77.3	77.4	77.9	77.8	78.5	78.6	
8	AT21	AT	Kärnten	78.6	78.4	79.0	79.1	78.6	78.9	79.4	
9	AT22	AT	Steiermark	78.4	79.0	79.1	79.5	78.7	79.6	79.3	
10	AT31	AT	Oberösterreich	78.6	78.7	78.8	79.4	79.0	79.6	79.6	
11	AT32	AT	Salzburg	79.1	79.4	79.6	80.0	79.9	80.5	80.1	
12	AT33	AT	Tirol	79.6	79.7	80.1	80.7	80.3	80.6	81.2	
13	AT34	AT	Vorarlberg	79.4	79.3	80.0	80.3	80.1	80.4	80.1	
14	BE10	BE	Région de Bruxelles-Capitale/Brussels H	78.5	77.5	77.8	78.2	78.2	78.4	78.5	
15	BE21	BE	Prov. Antwerpen	79.2	78.8	79.2	80.1	80.1	80.2	80.3	
16	BE22	BE	Prov. Limburg (BE)	79.1	79.3	79.9	80.5	80.6	80.6	81.0	
17	BE23	BE	Prov. Oost-Vlaanderen	78.2	78.1	78.5	79.3	79.1	79.5	79.8	
18	BE24	BE	Prov. Vlaams-Brabant	79.6	79.4	79.5	80.1	79.9	80.4	80.6	
19	BE25	BE	Prov. West-Vlaanderen	78.8	78.9	78.7	79.6	79.7	80.2	80.0	
20	BE31	BE	Prov. Brabant wallon	78.8	78.3	79.6	79.1	79.7	80.3	79.6	
21	BE32	BE	Prov. Hainaut	74.7	74.7	75.0	75.9	75.8	75.7	76.4	
22	BE33	BE	Prov. Liège	76.3	76.4	76.9	77.3	77.0	77.5	77.5	
23	REF1	REF	Prov. Luxembourg (REF)	76.0	76.7	76.9	77.2	77.0	76.9	77.1	

+ ≡ TOTAL_NUTS2_life expectancy at birth 2011.2022 ▾ FEMALE_NUTS2_life expectancy at birth 2011.2022 ▾ MALE_NUTS2_life expectancy at birth

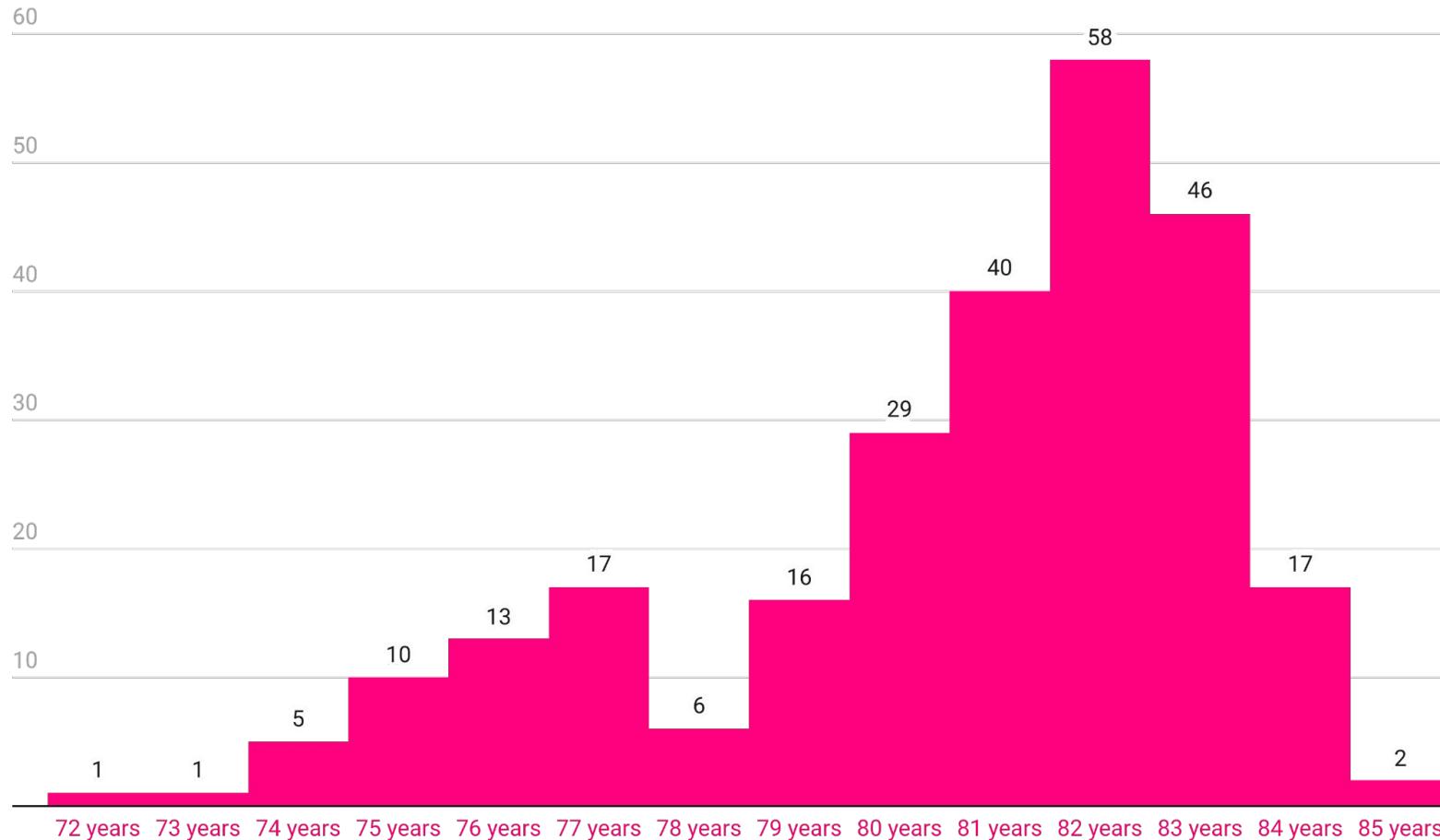
EXERCISE

- 👉 Open [this dataset](#) and make your own copy.
- 👉 Make some charts in Excel / Google Sheets to portray insights. You can do some basic aggregations of the data or calculations if needed.
- 👉 Upload PNG / Screenshots [in this folder](#)

Top regions with highest and lowest life expectancy at birth in 2022



Distribution of life expectancy at birth in European regions in 2022

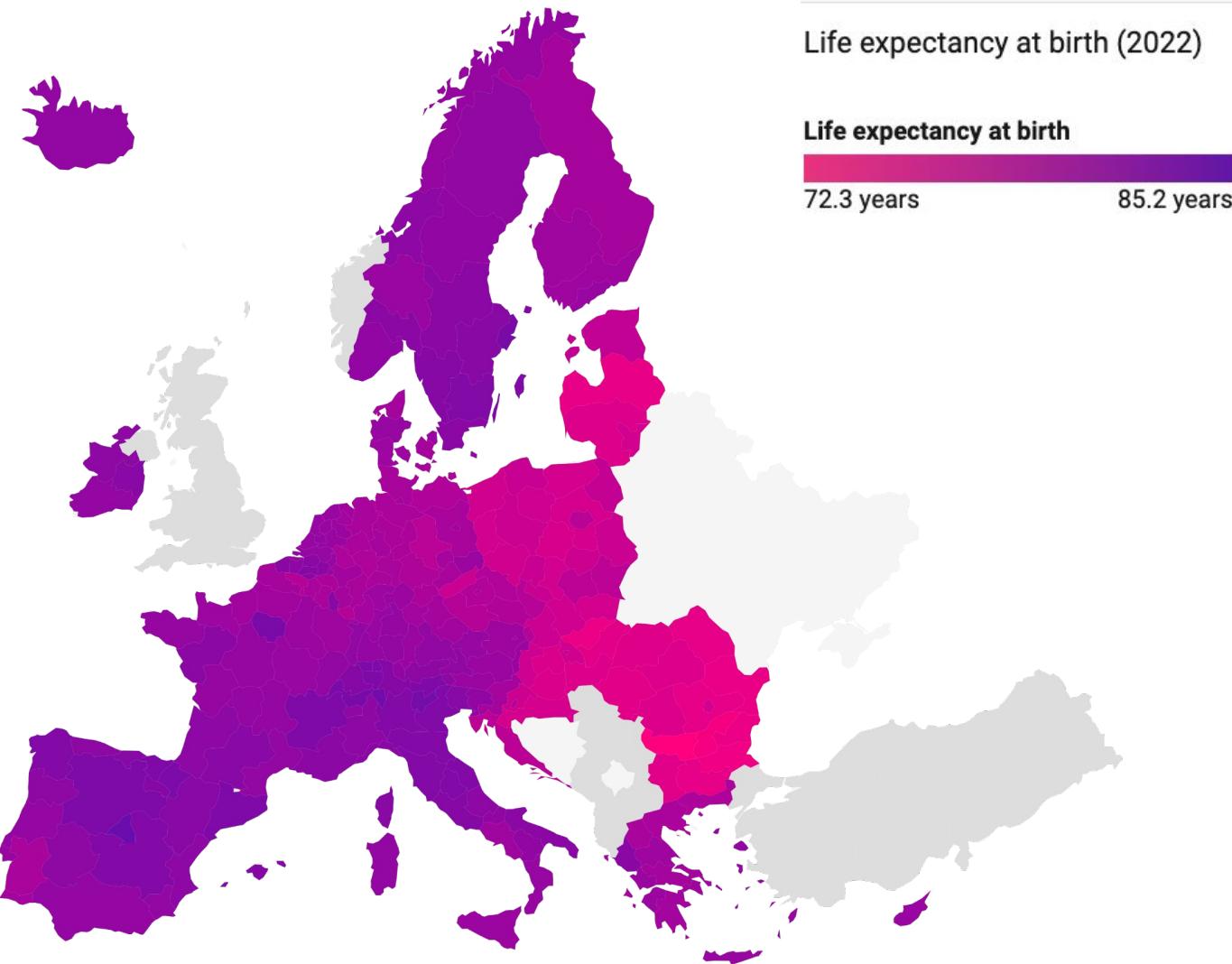


Life expectancy at birth (2022)

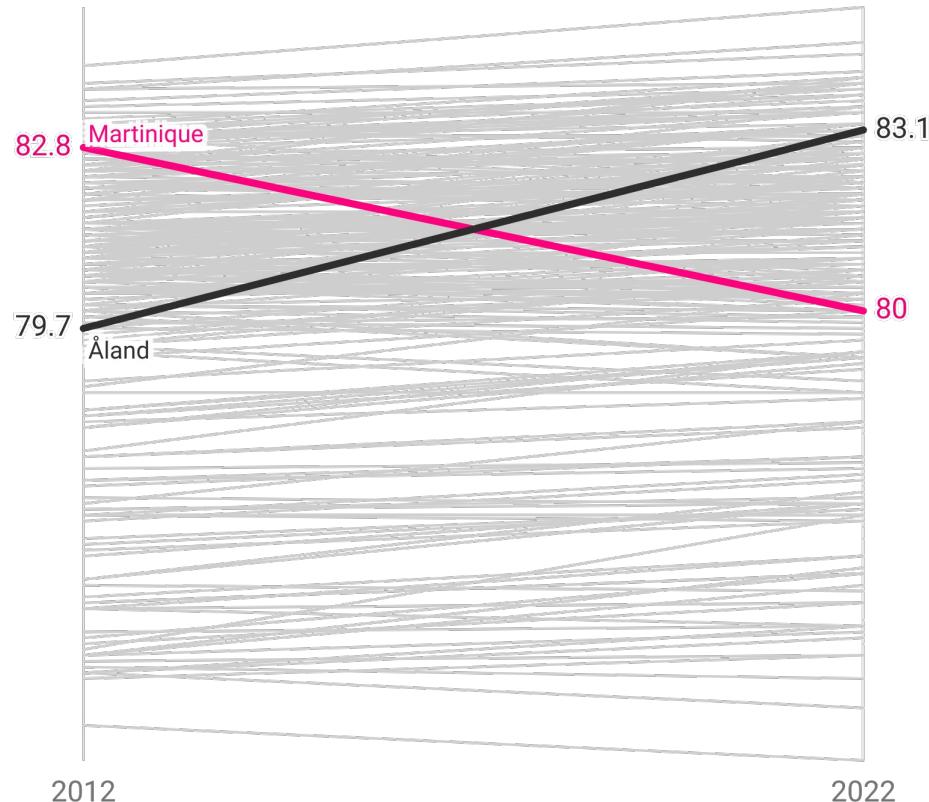
Life expectancy at birth

72.3 years

85.2 years



Life expectancy in 2012 and 2022 in different European regions

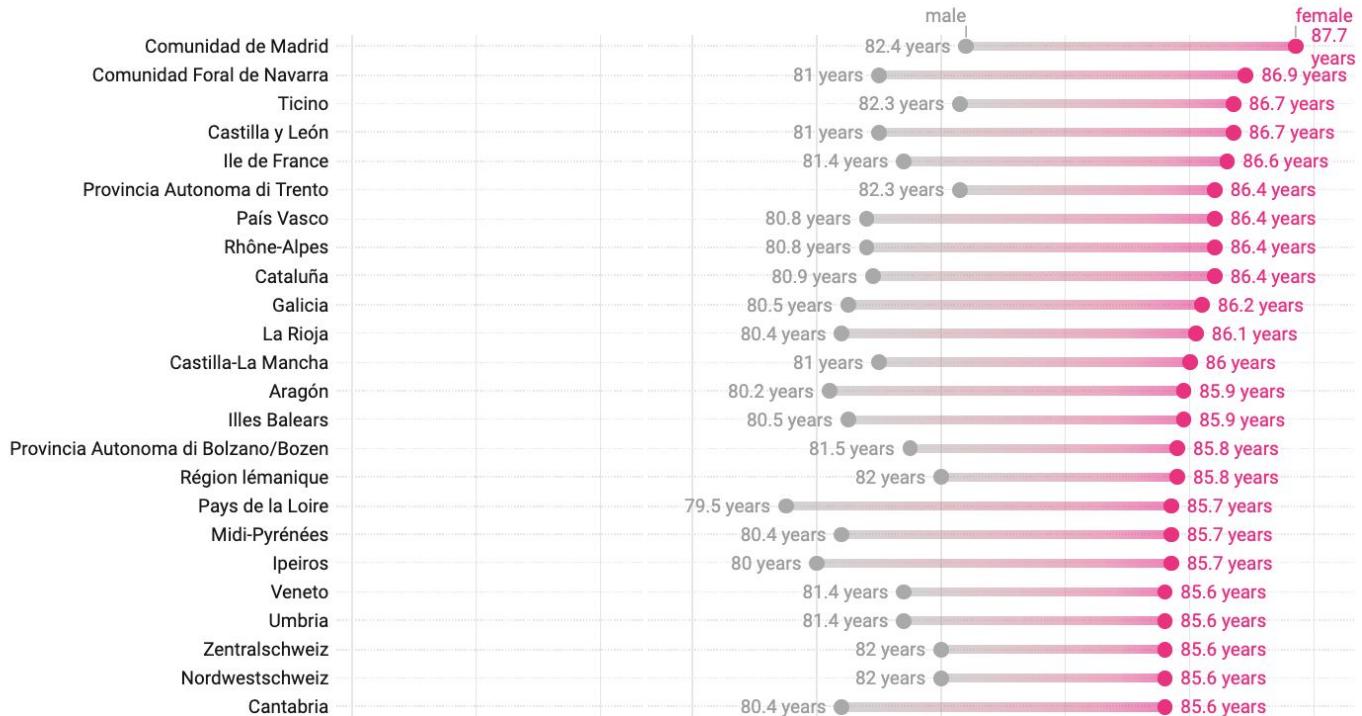


Life expectancy in EU regions in 2022 vs. 2012

Regione (NUTS2)	2022	2012	change ▾
Åland	83.1 years	79.7 years	+3.4
Sostinės regionas	76.5 years	74.2 years	+2.3
Prov. Brabant wallon	82.8 years	81.0 years	+1.8
Prov. Antwerpen	82.8 years	81.1 years	+1.7
Prov. Oost-Vlaanderen	82.6 years	80.9 years	+1.7
Região Autónoma da Madeira	79.3 years	77.6 years	+1.7
Stockholm	84.0 years	82.4 years	+1.6
Småland med öarna	83.5 years	81.9 years	+1.6
Norra Mellansverige	82.7 years	81.1 years	+1.6
Norte	82.5 years	80.9 years	+1.6

Additional 251 rows not shown.

Life expectancy by gender in European regions



SEEMS EASY.

SEEMS EASY.

BUT...

Top regions with highest and lowest life expectancy at birth in 2022



Top regions with highest and lowest life expectancy at birth in 2022

2012

highest

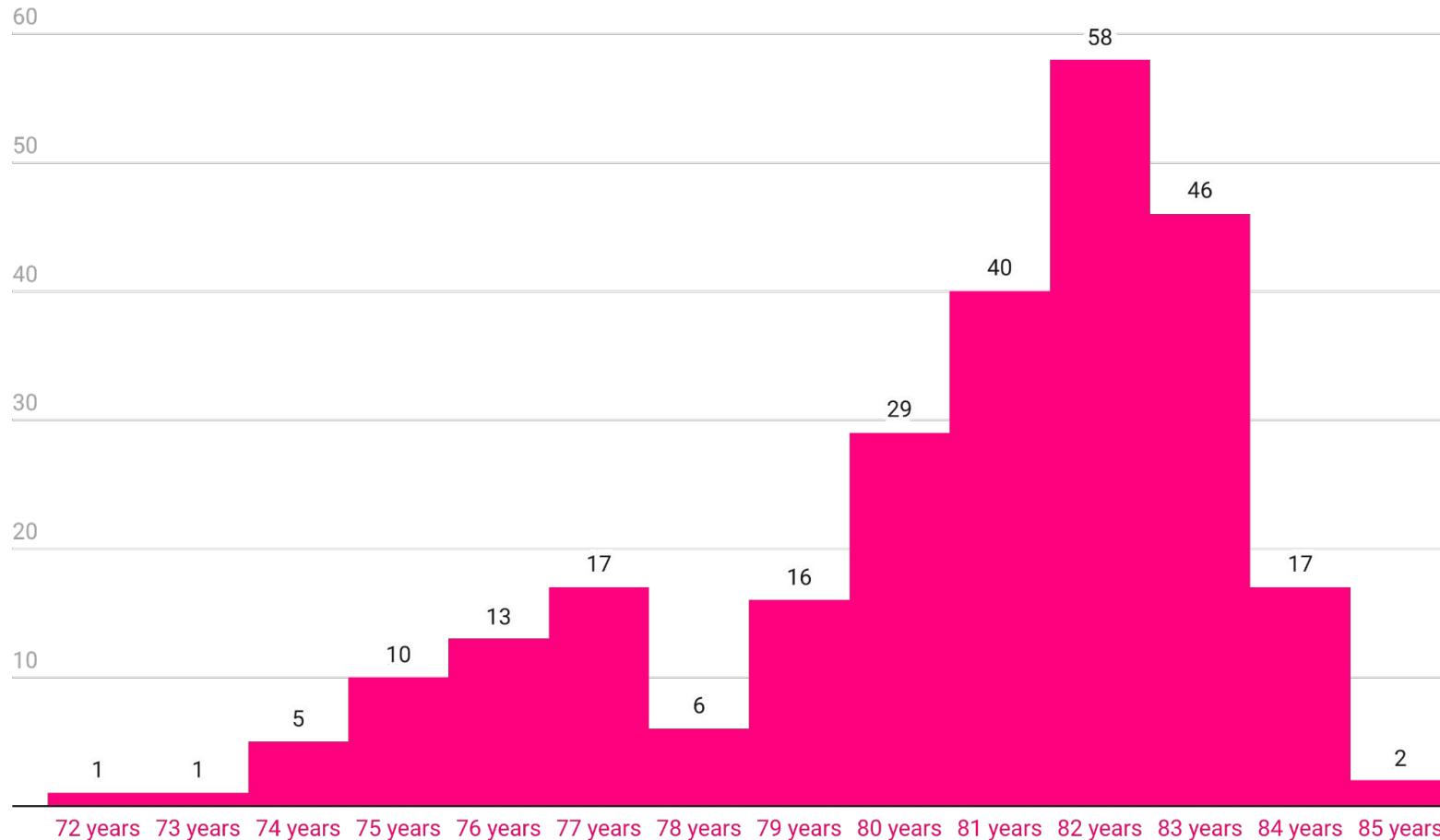
Comunidad de Madrid	85.2 years
Ticino	84.6 years
Provincia Autonoma di Trento	84.4 years
Ile de France	84.1 years
Région lémanique	84 years
Stockholm	84 years

REFERENCE

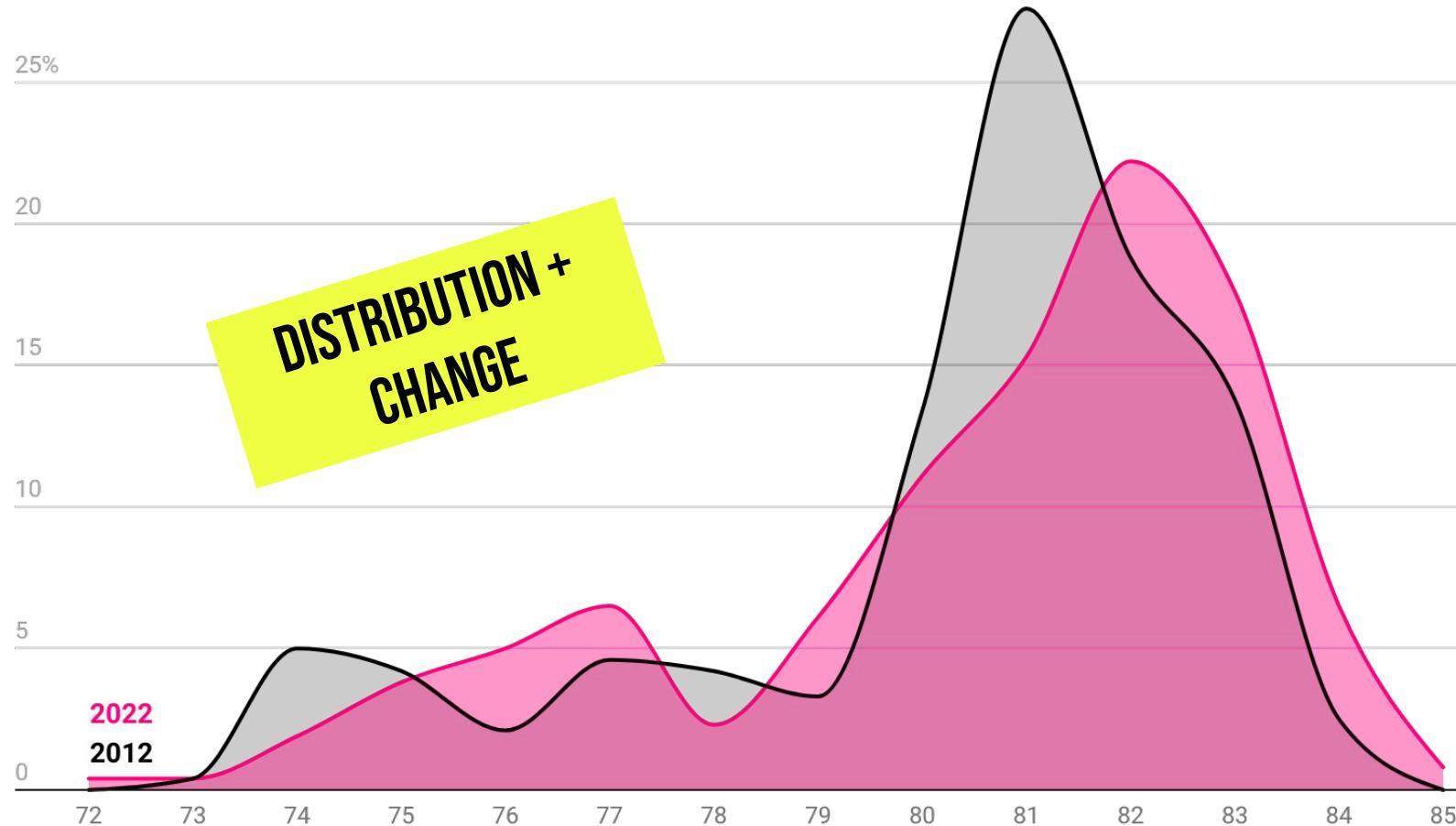
lowest

Severoiztochen	74.1 years
Észak-Magyarország	74.1 years
Yugoiztochen	73.7 years
Severen tsentralen	73.2 years
Severozapaden	72.3 years

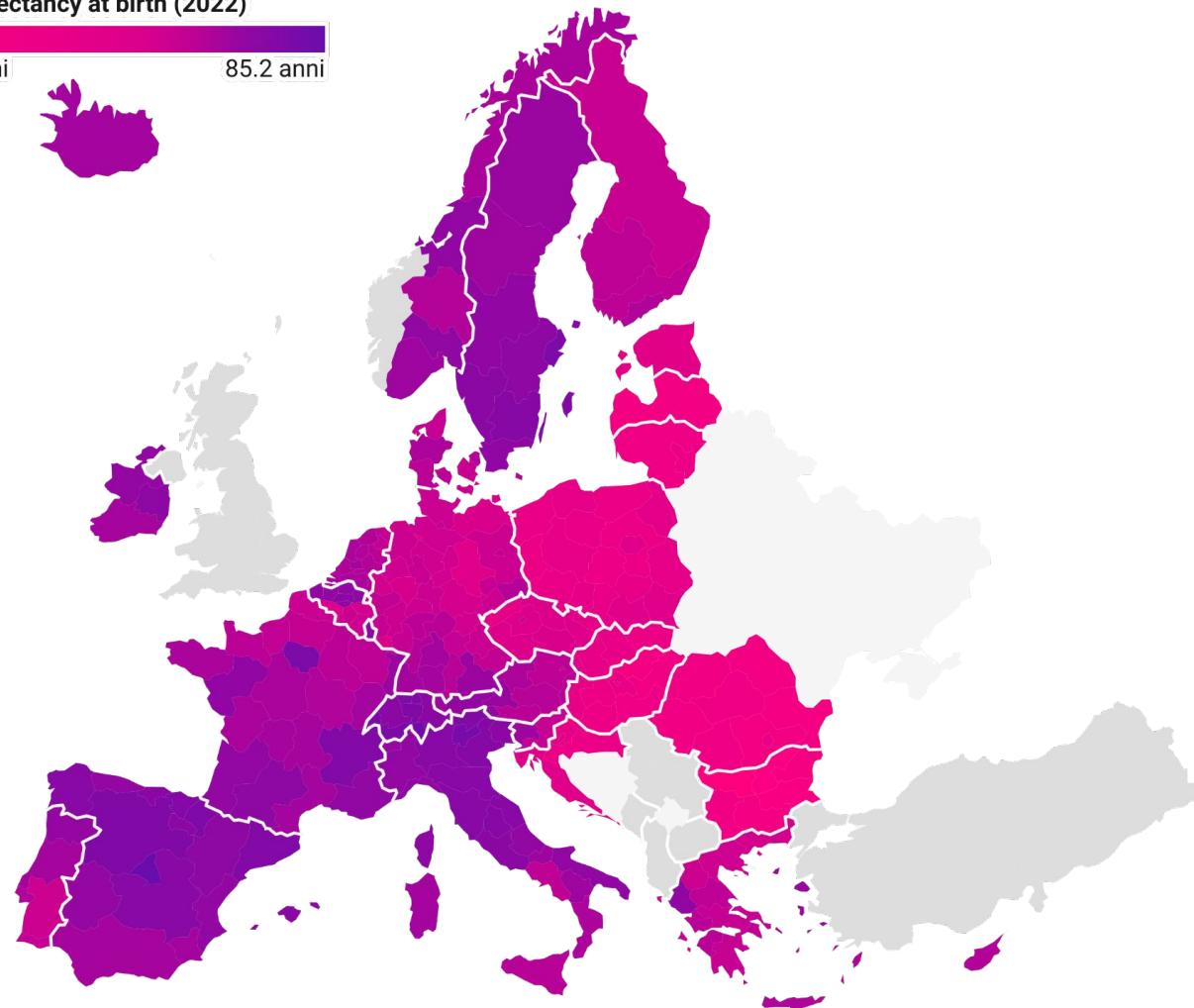
Distribution of life expectancy at birth in European regions in 2022



Distribution of life expectancy at birth in European regions in 2012 and 2022



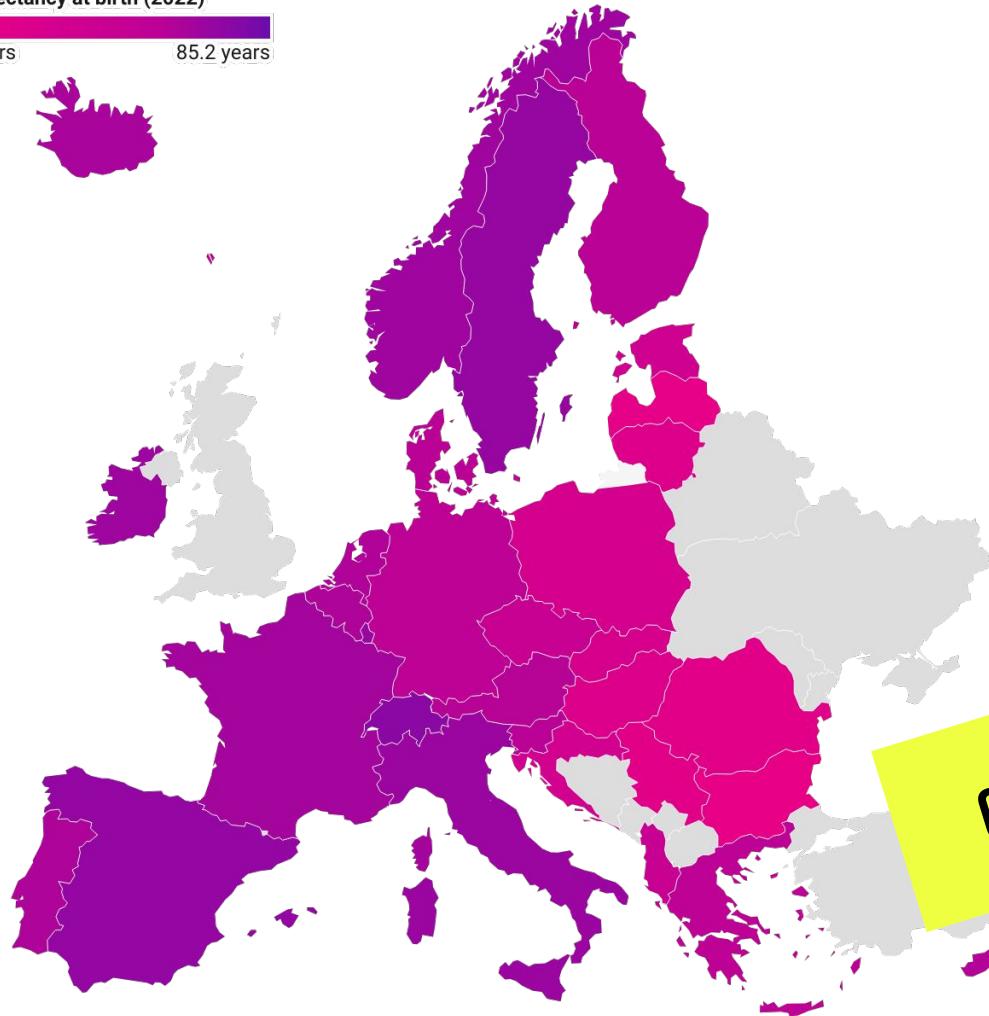
Life expectancy at birth (2022)



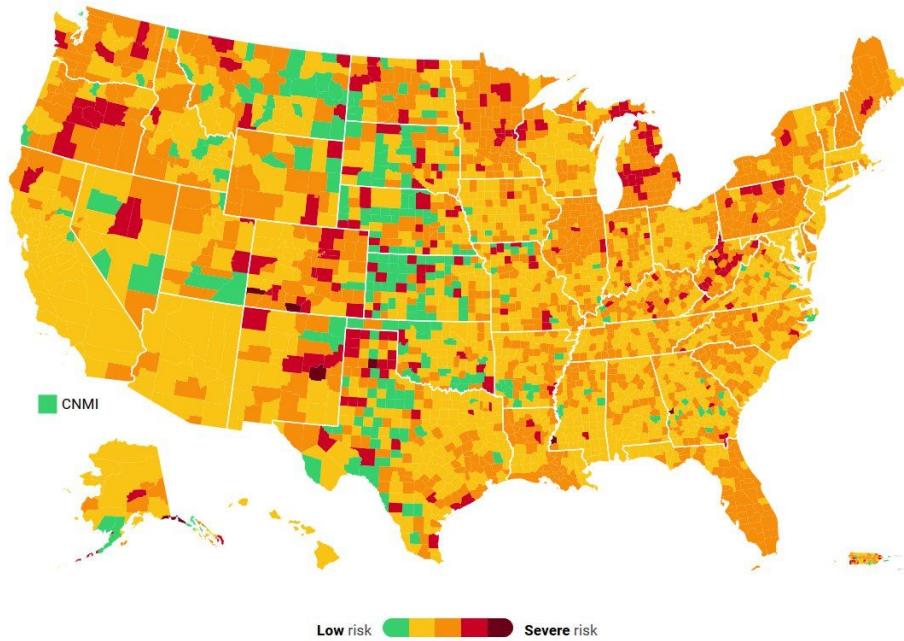
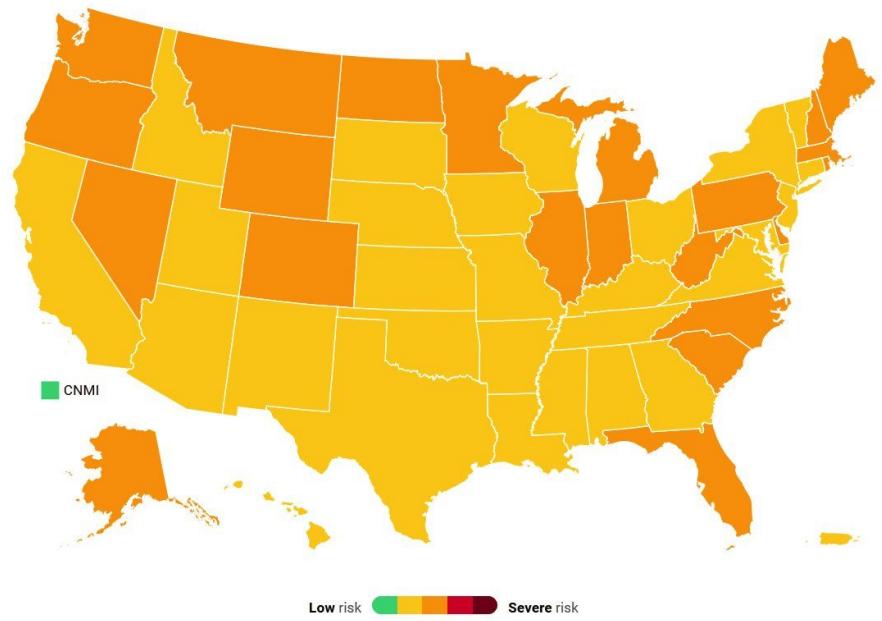
Life expectancy at birth (2022)

72.3 years

85.2 years

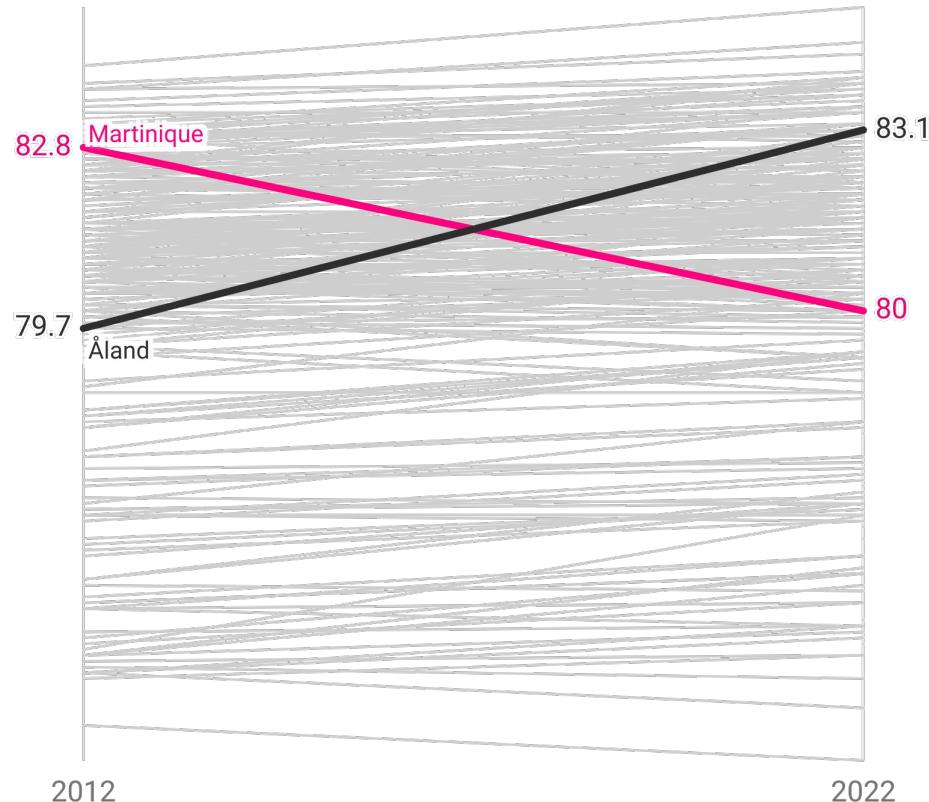


GRANULARITY

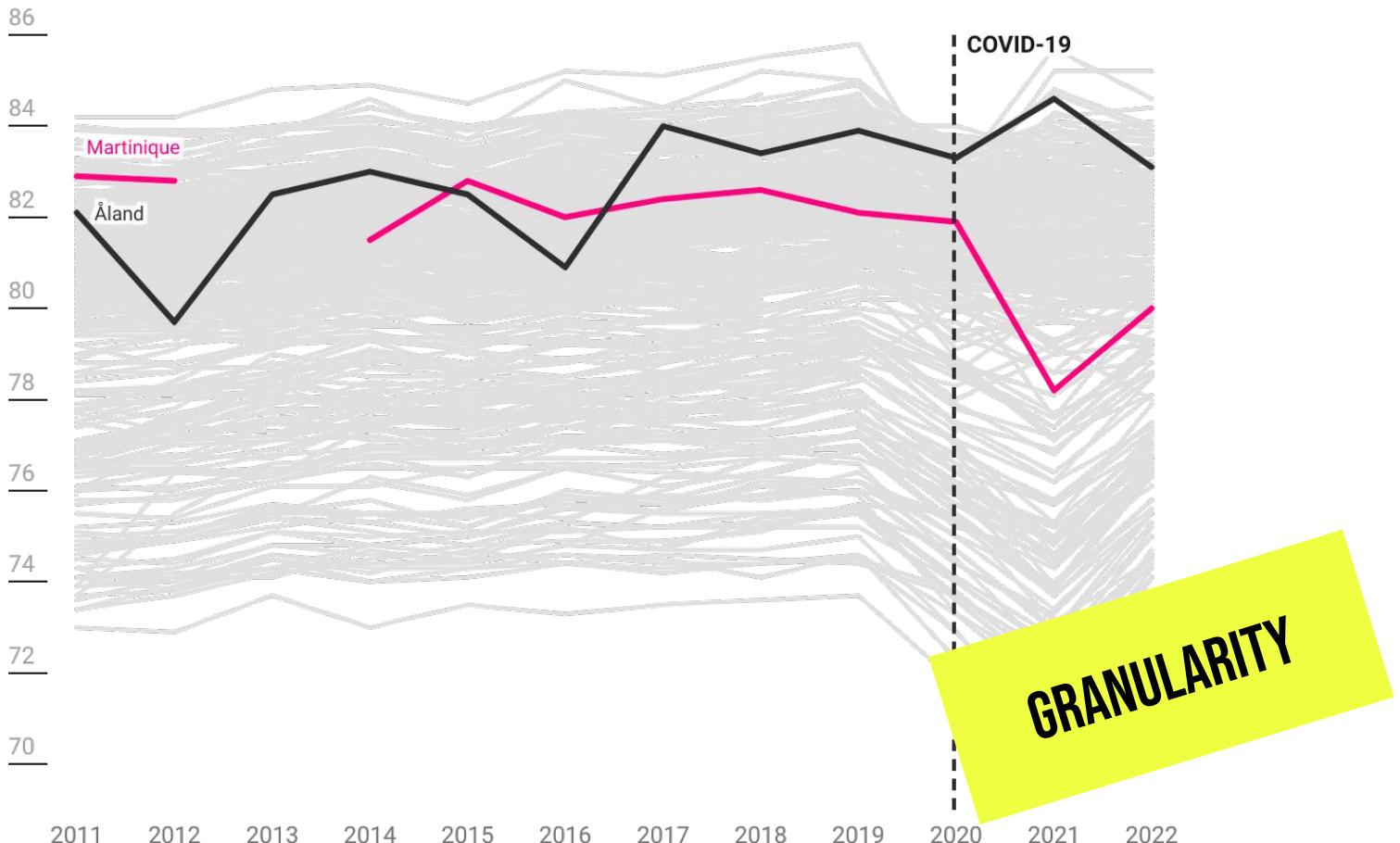


[Twitter thread](#) • Dr. Neurofourier (Map by COVID ACT NOW)

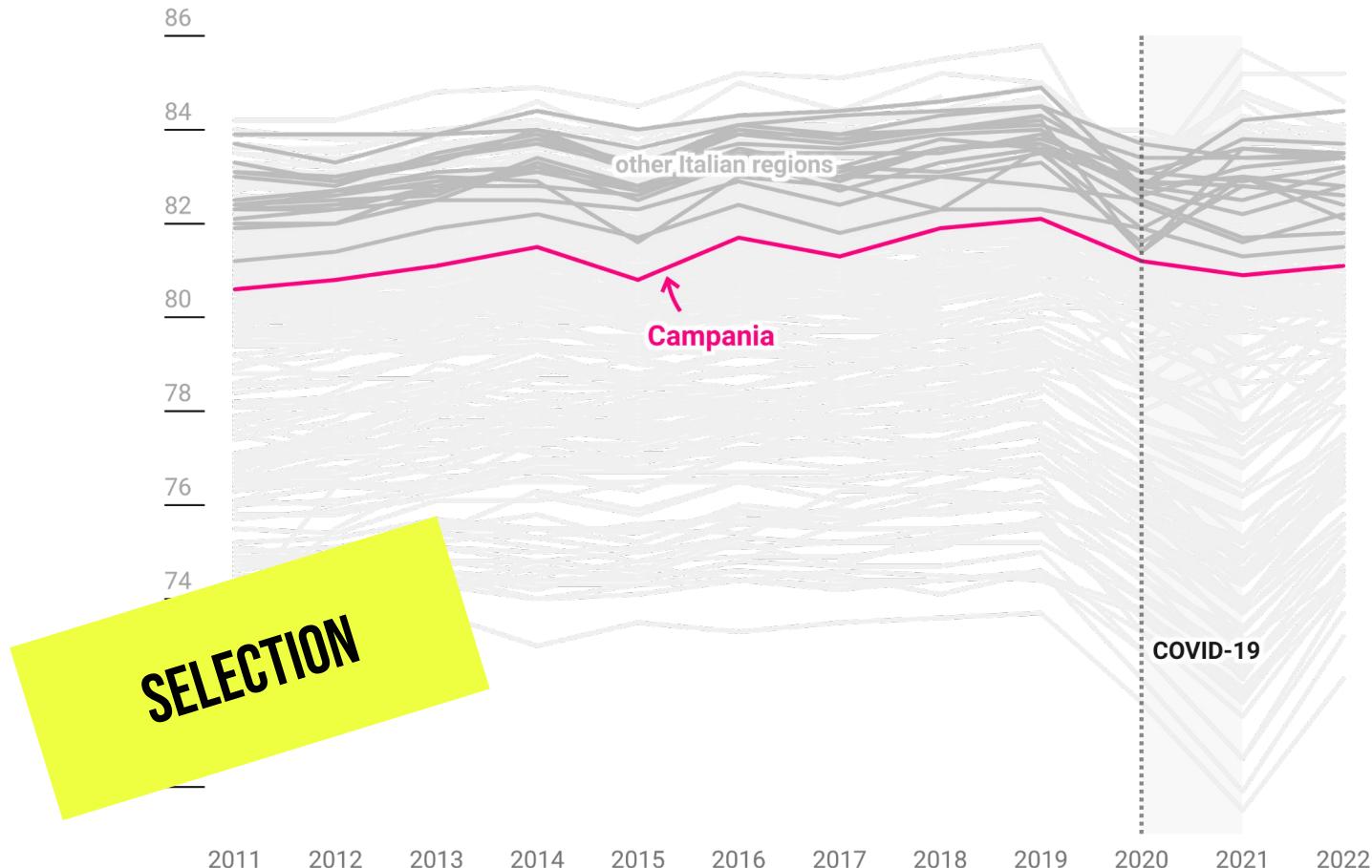
Life expectancy in 2012 and 2022 in different European regions



Life expectancy in 2012-2022 in different European regions



Life expectancy in 2012-2022 in different European regions



Life expectancy in EU regions in 2022 vs. 2012

Regione (NUTS2)	2022	2012	change ▾
Åland	83.1 years	79.7 years	+3.4
Sostinės regionas	76.5 years	74.2 years	+2.3
Prov. Brabant wallon	82.8 years	81.0 years	+1.8
Prov. Antwerpen	82.8 years	81.1 years	+1.7
Prov. Oost-Vlaanderen	82.6 years	80.9 years	+1.7
Região Autónoma da Madeira	79.3 years	77.6 years	+1.7
Stockholm	84.0 years	82.4 years	+1.6
Småland med öarna	83.5 years	81.9 years	+1.6
Norra Mellansverige	82.7 years	81.1 years	+1.6
Norte	82.5 years	80.9 years	+1.6

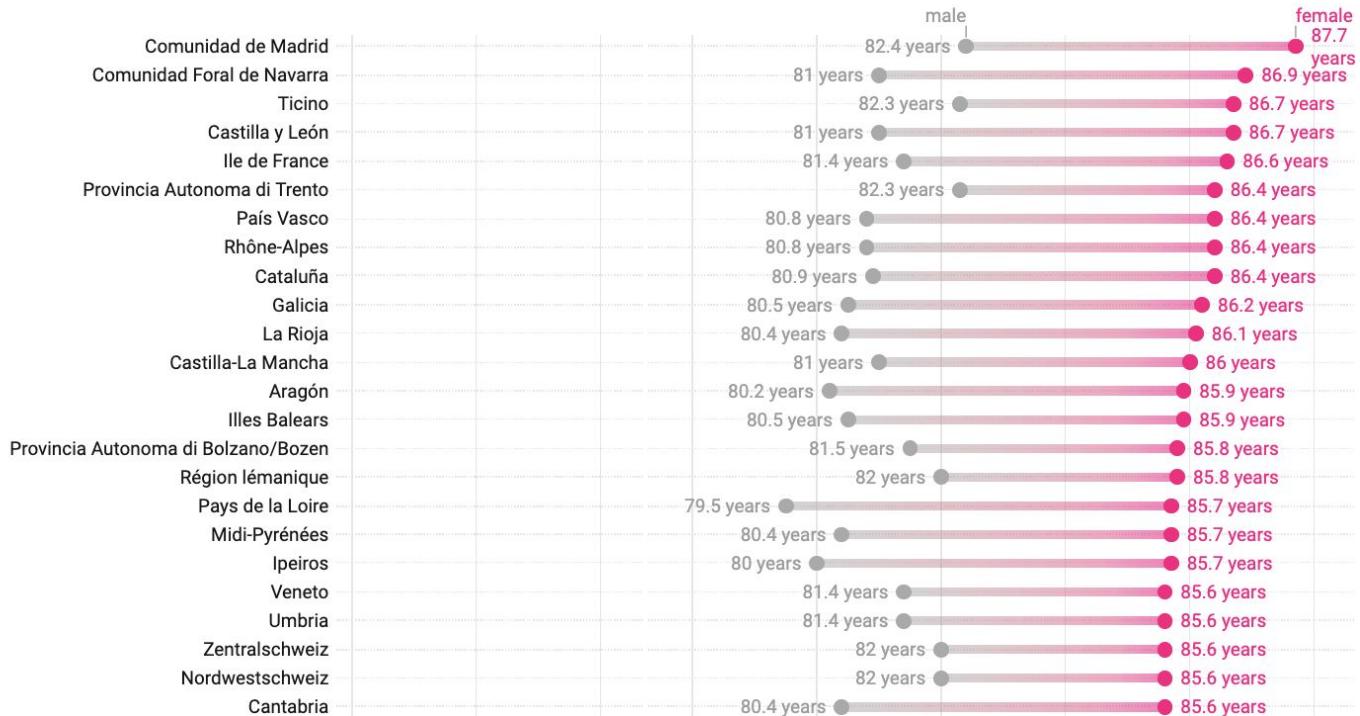
Additional 251 rows not shown.

Life expectancy in EU regions in 2022 vs. 2012

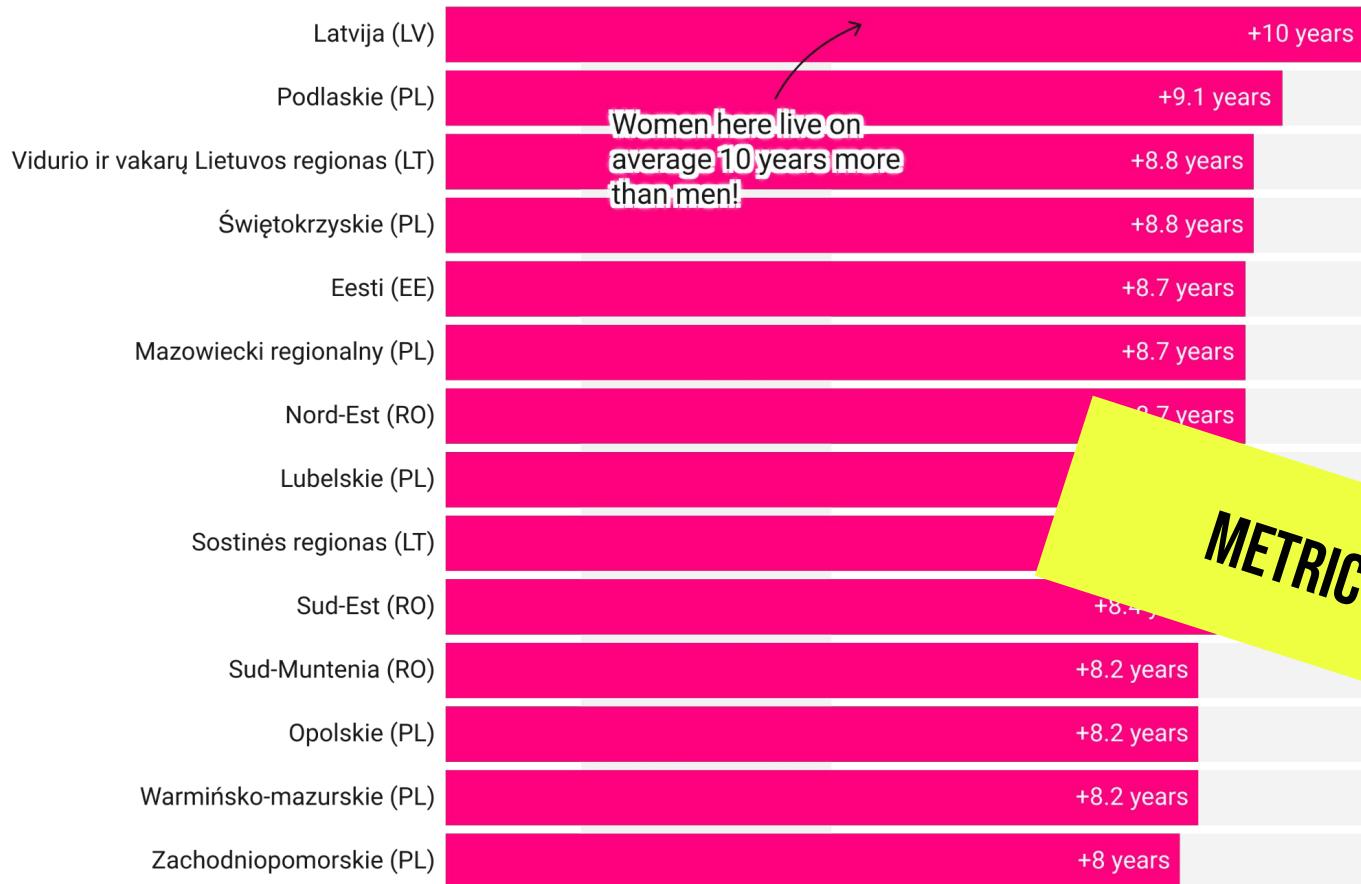
Region (NUTS2)	2022	2012	change ▲
Region Vojvodine	74.5 years		
Mayotte	74.3 years		
Martinique	80.0 years	82.8 years	-2.8
Ísland	82.1 years	83.0 years	-0.9
Ionia Nisia	80.5 years	81.4 years	-0.9
Saarland	79.2 years	80.0 years	-0.8
Trier	80.6 years	81.3 years	-0.7
Bremen	79.2 years	79.9 years	-0.7
Guyane	78.6 years	79.3 years	-0.7
Sachsen-Anha	78.8 years	79.4 years	-0.6

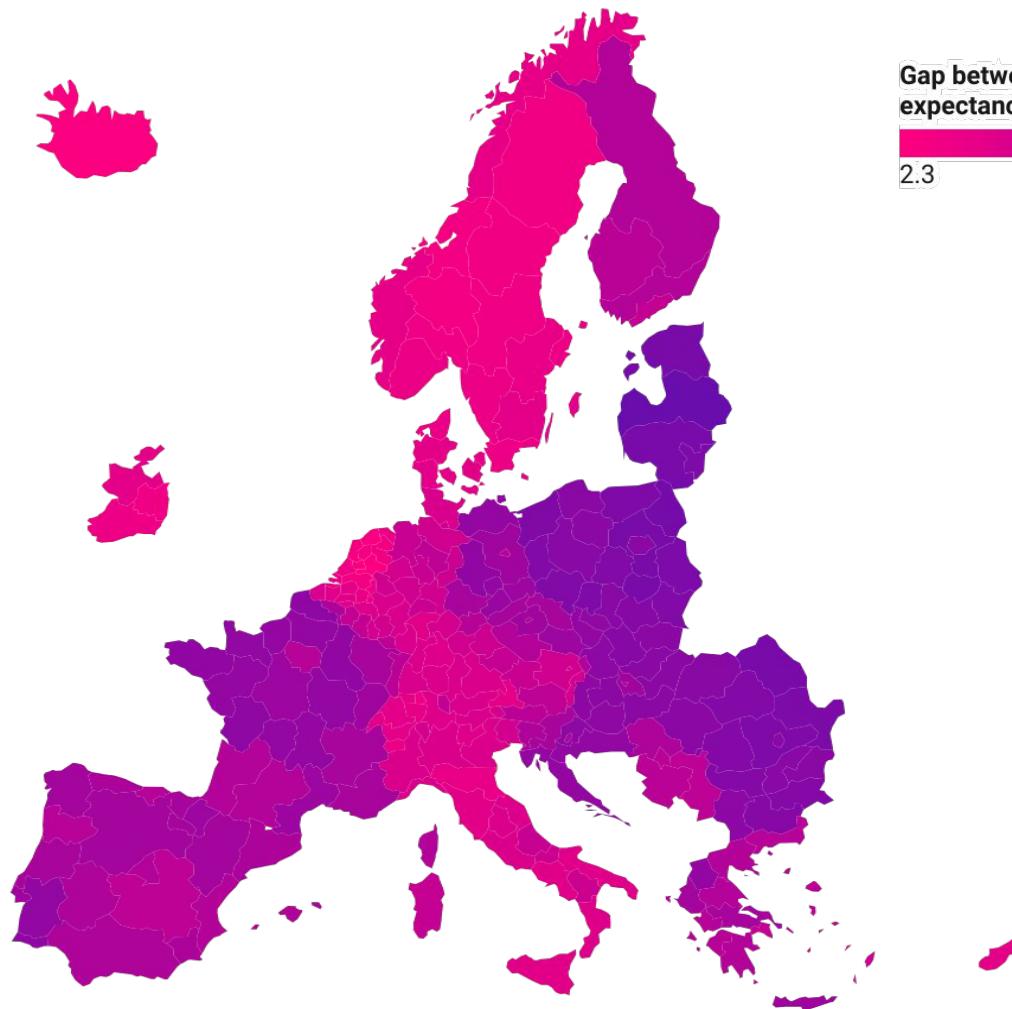
SORTING

Life expectancy by gender in European regions



European regions with the highest difference in female life expectancy compared to male life expectancy in 2022





Gap between female and male life expectancy

2.3

10.0

METRIC + ENCODING



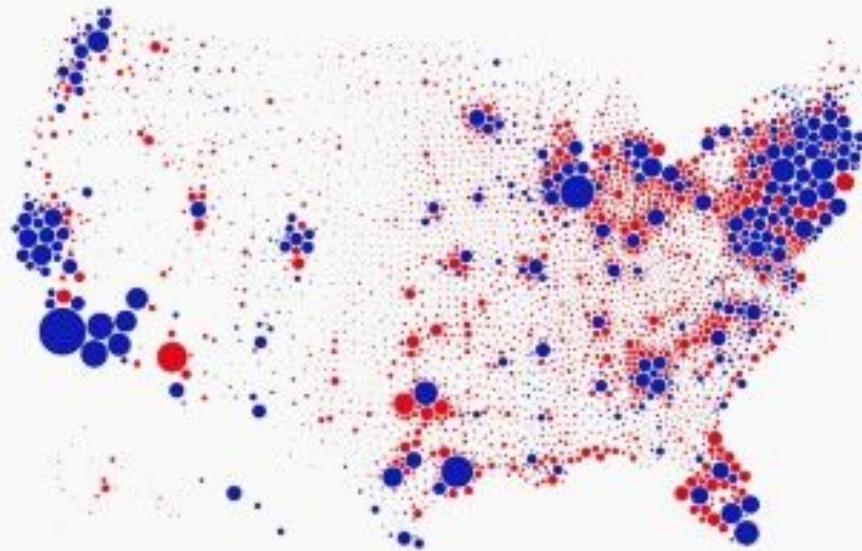
Donald J. Trump 
@realDonaldTrump



4:05 AM · 01 Oct 19 · Twitter for iPhone

53.1K Retweets 199K Likes



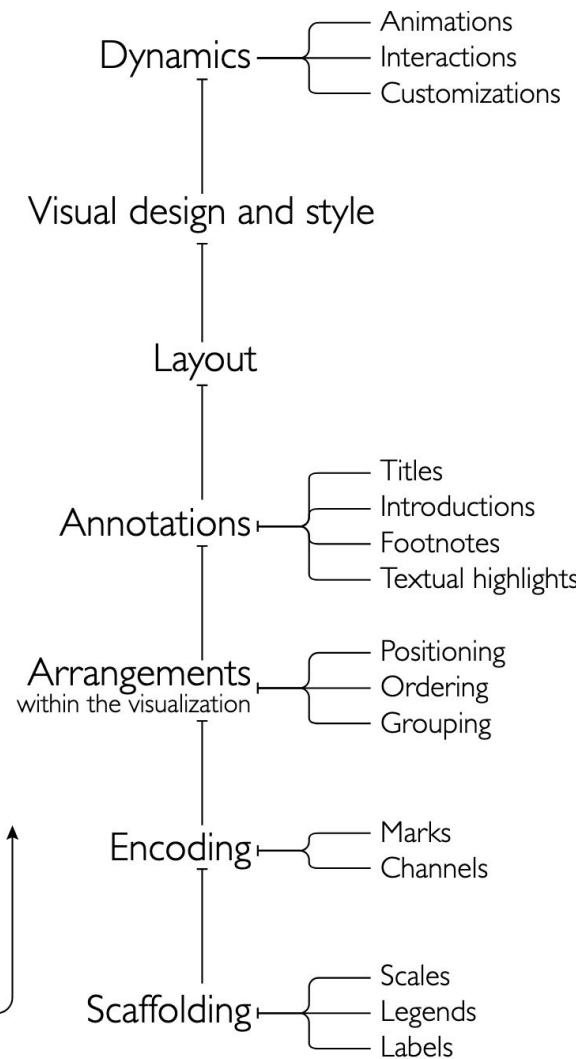


METRIC + ENCODING

Visualization: Layers and elements to think about

The nature of the data to
be visualized influences
everything else

**Read from
the bottom-up**



Alberto Cairo
Visualizing data with
impact(data.europa
academy)

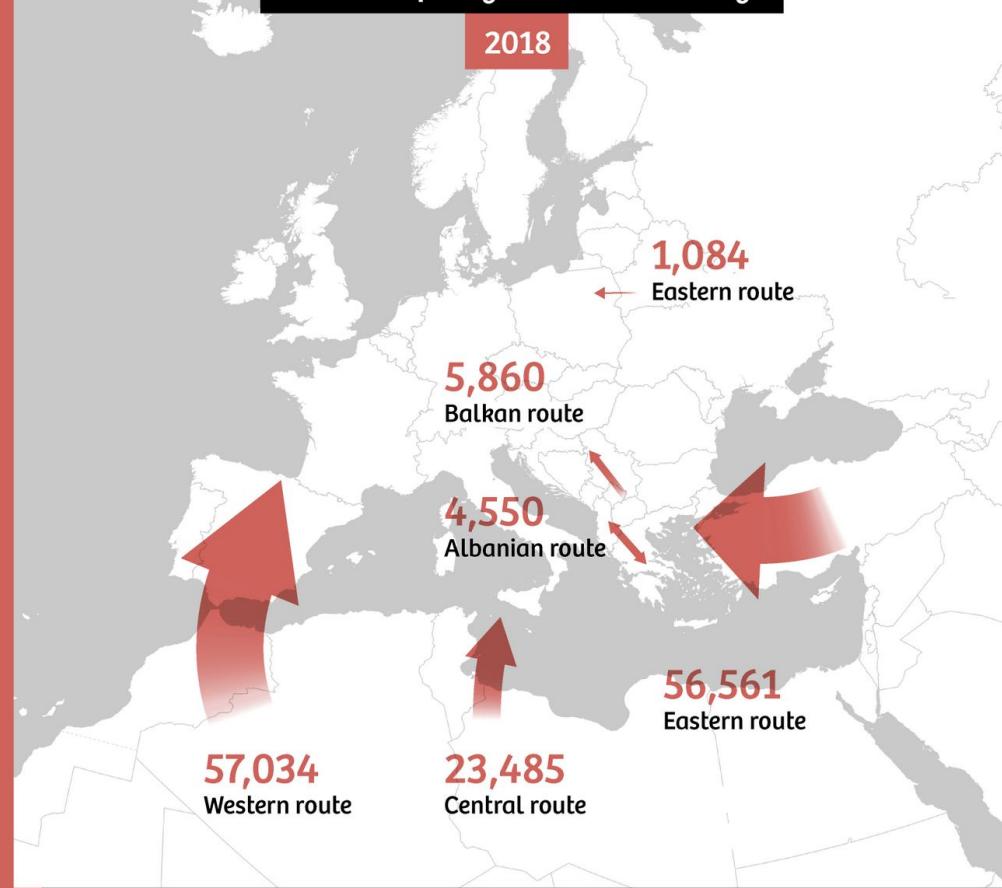
SEEMS EASY.

BUT...

METAPHORS

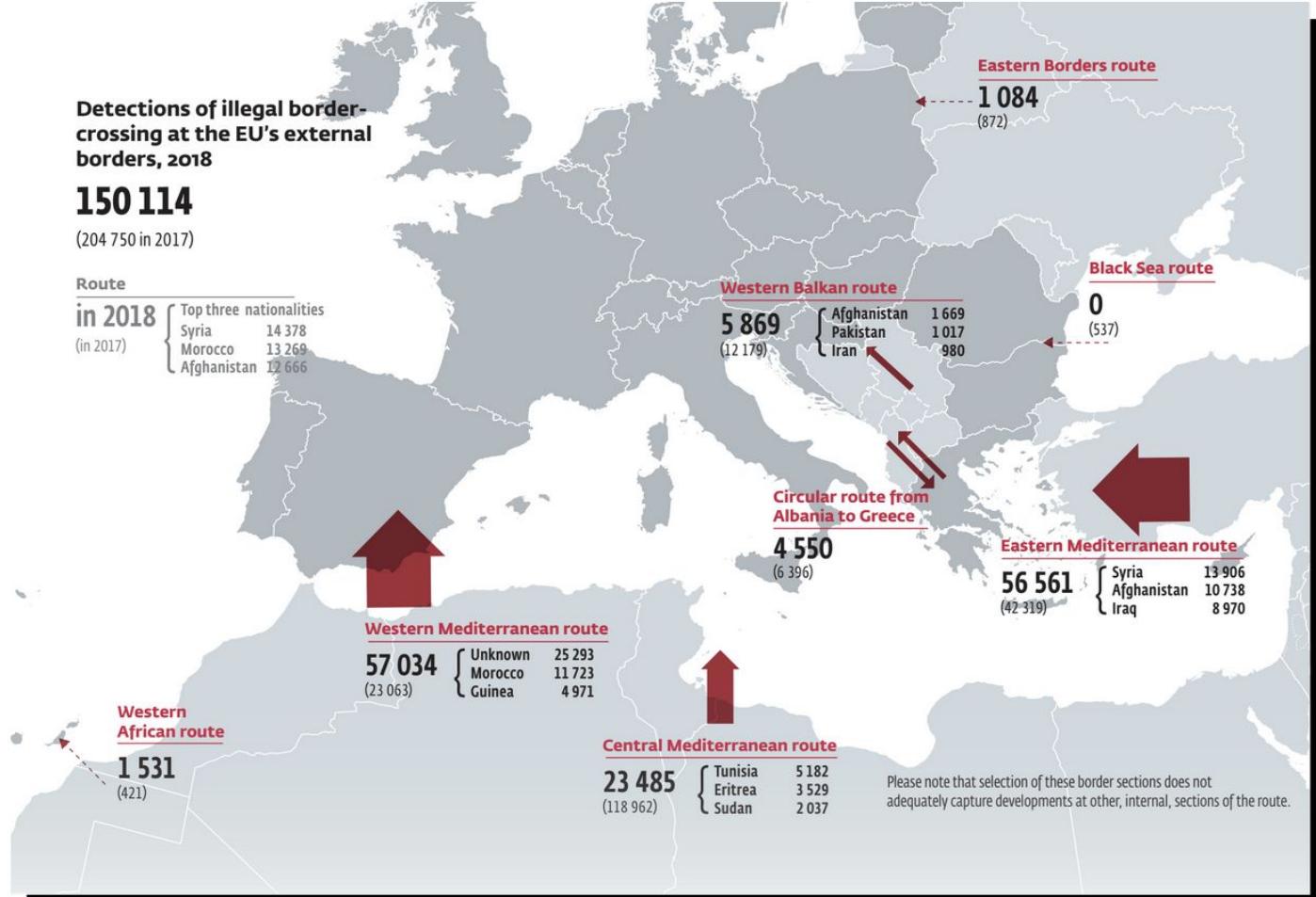
Number of illegal border crossings

2018



Source: Frontex

How maps in the media make us more negative about migrants



How maps in the media make us more negative about migrants

Mourir aux portes de l'Europe

MAPS BY DI PHILIPPE REKACEWICZ

Pré-frontière de l'Europe

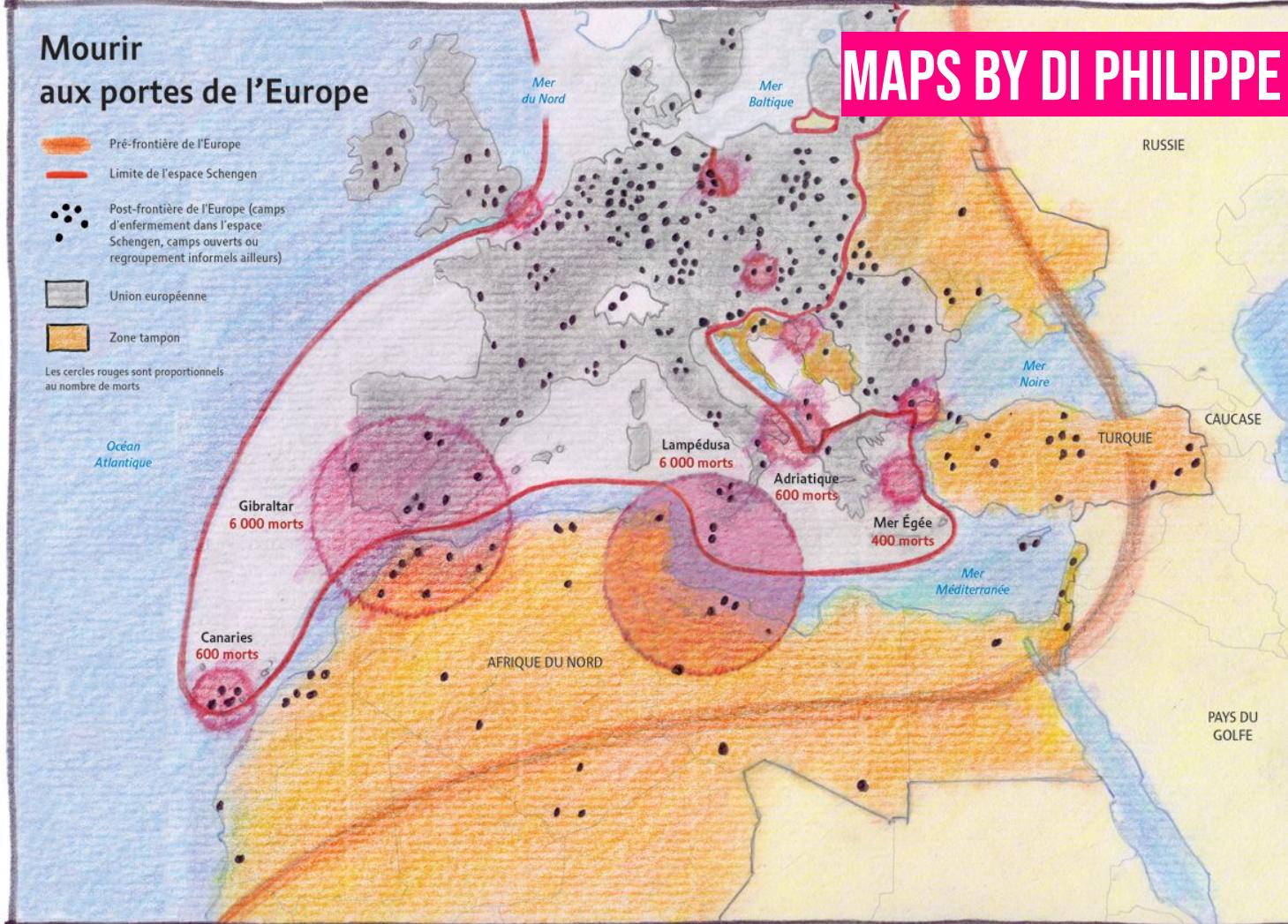
Limite de l'espace Schengen

Post-frontière de l'Europe (camps d'enfermement dans l'espace Schengen, camps ouverts ou regroupement informels ailleurs)

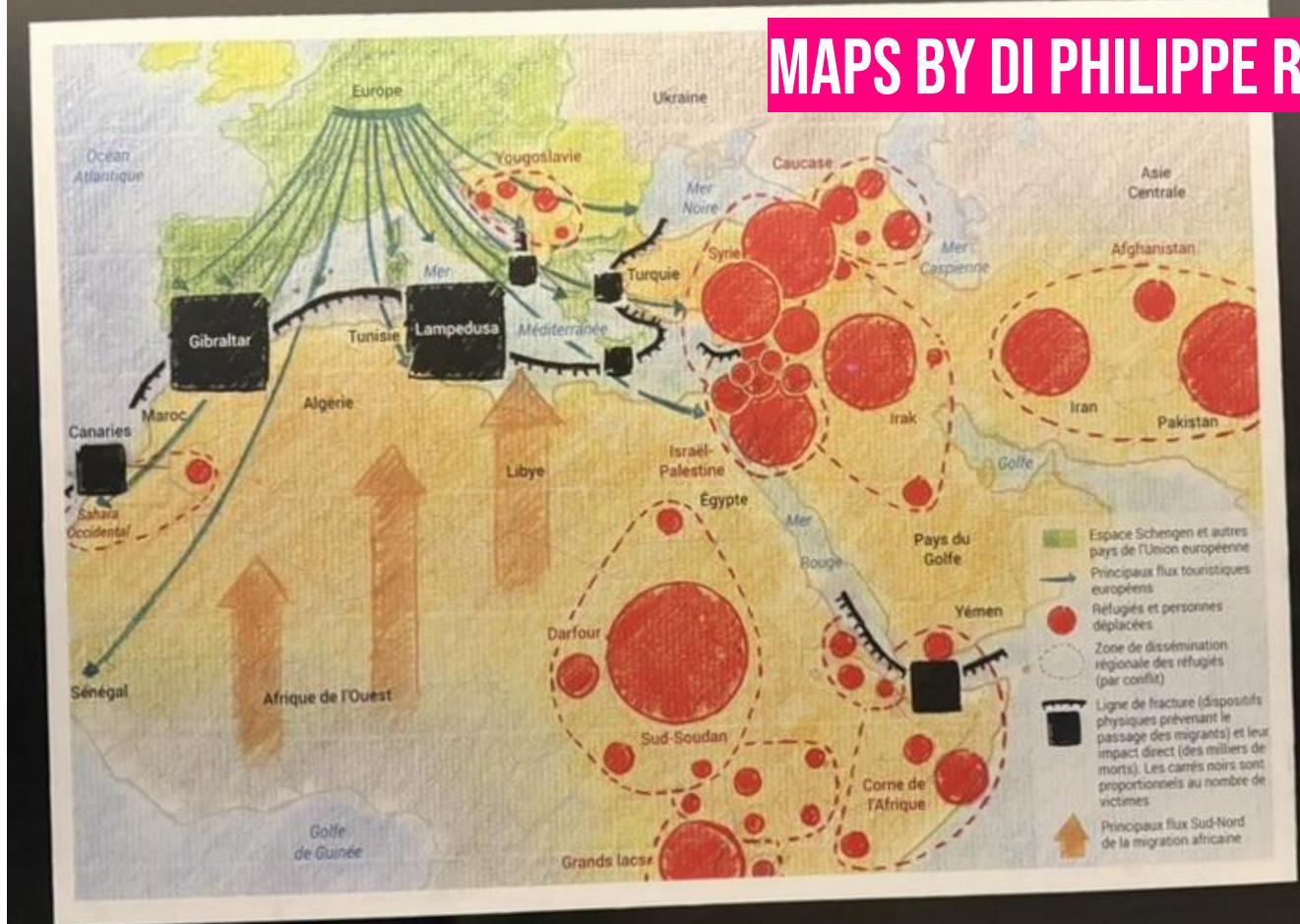
Union européenne

Zone tampon

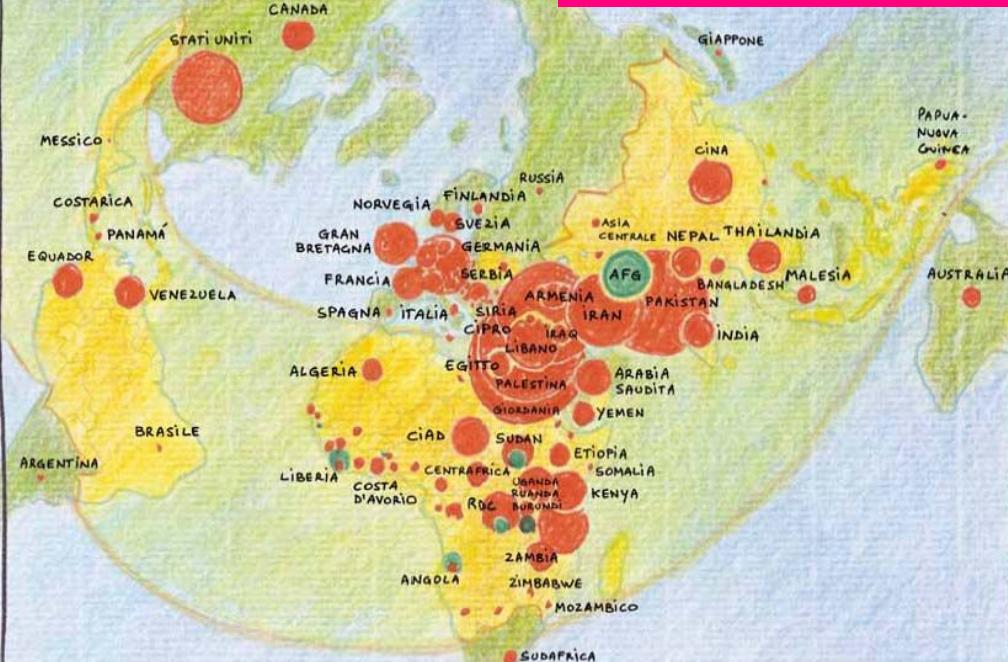
Les cercles rouges sont proportionnels au nombre de morts.



MAPS BY DI PHILIPPE REKACEWICZ



MAPS BY DI PHILIPPE REKACEWICZ

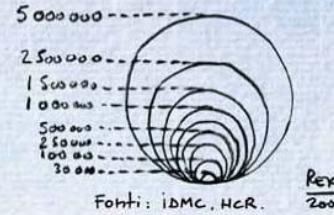


Reddito nazionale lordo (RNL) a parità di potere d'acquisto

- Superiore a 10 000 dollari
 - Inferiore a 10 000 dollari

Popolazione rifugiata: persone che hanno varcato un confine e alle quali è stato accordato uno status umanitario o una protezione temporanea

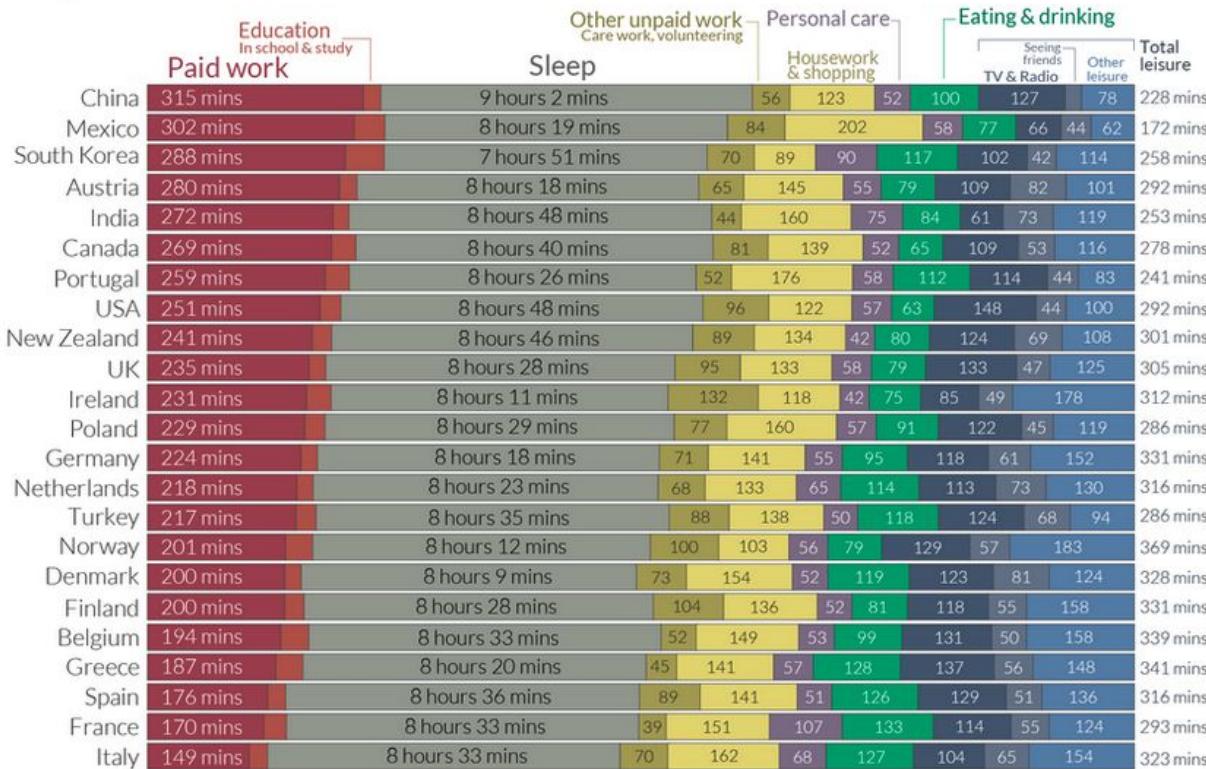
• Popolazione rimpatriata: rifugiati tornati al loro luogo d'origine. Sono affidati alla protezione dell'UNHCR che li assiste per diversi anni.



PERCEPTION

How do people spend their time?

Averages of minutes per day from time-use diaries for people between 15 and 64.



Data source: OECD Time Use Database, Gender Data Portal. For most countries surveys were conducted between 2009 and 2016, but surveys for some countries are older.
OurWorldInData.org – Research and data to make progress against the world's largest problems.

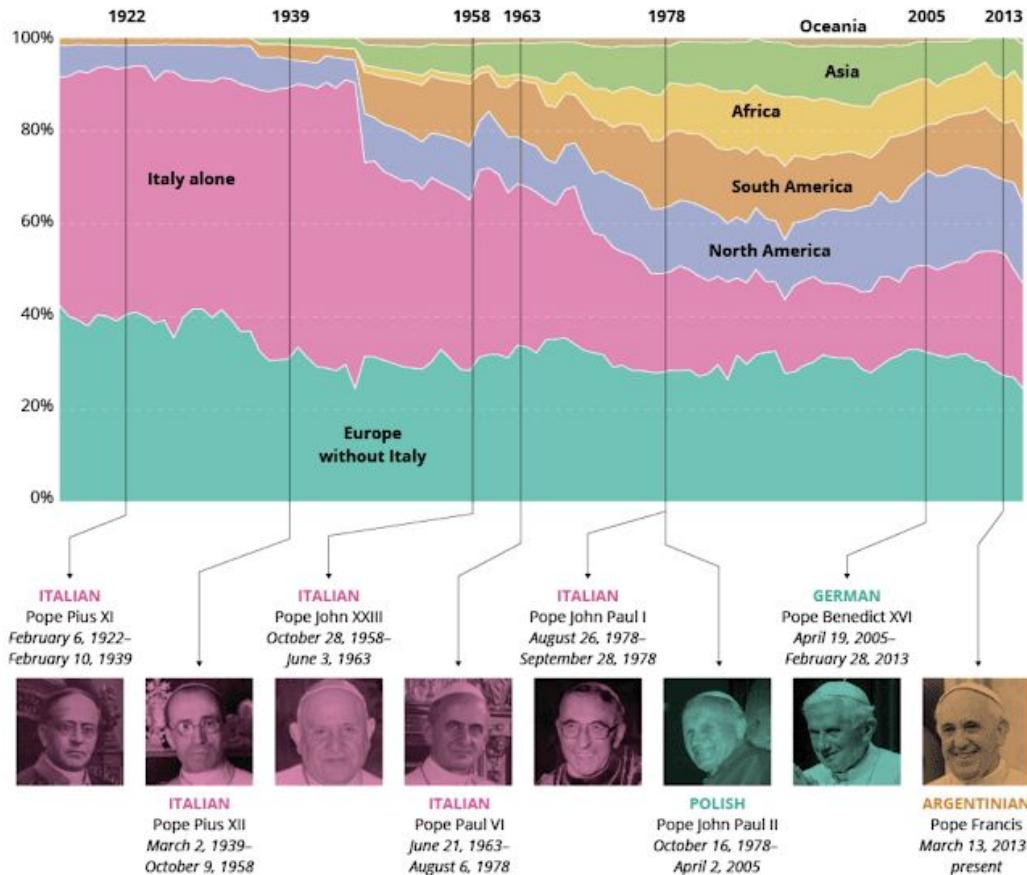
Licensed under CC-BY by the author Esteban Ortiz-Ospina.

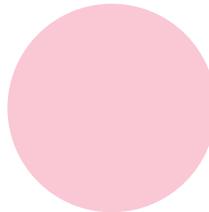
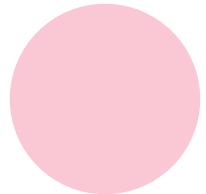
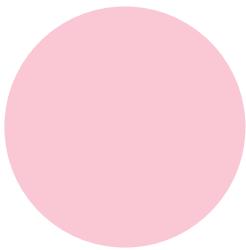
More Diverse Cardinals = More Diverse Popes

Once the College of Cardinals shifted from a Italian majority, they began electing non-Italian popes.

Voting-age cardinals, 1915-2015

— Year of papal conclave





4

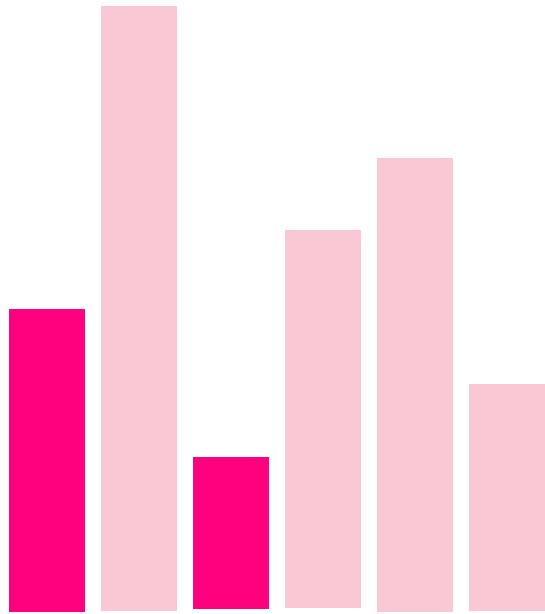
8

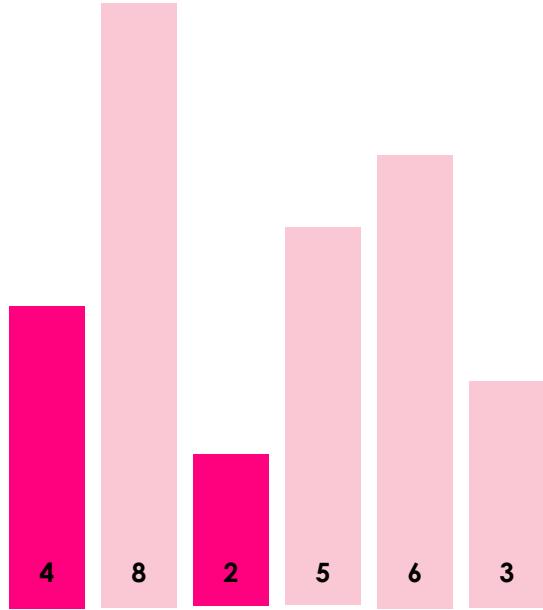
2

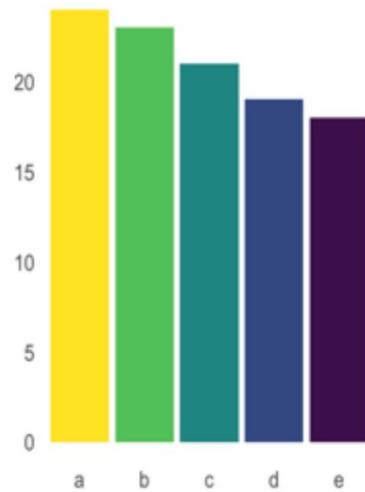
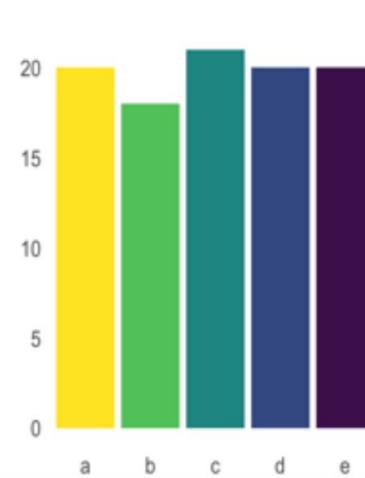
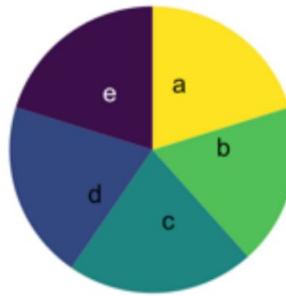
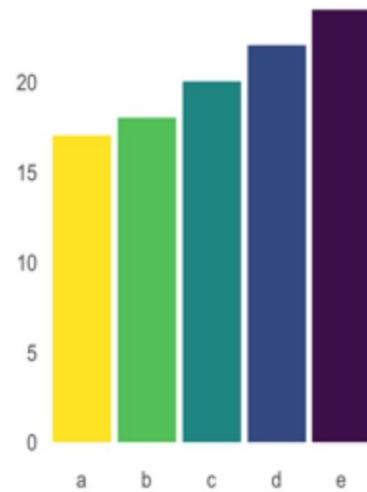
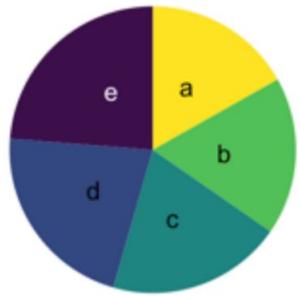
5

6

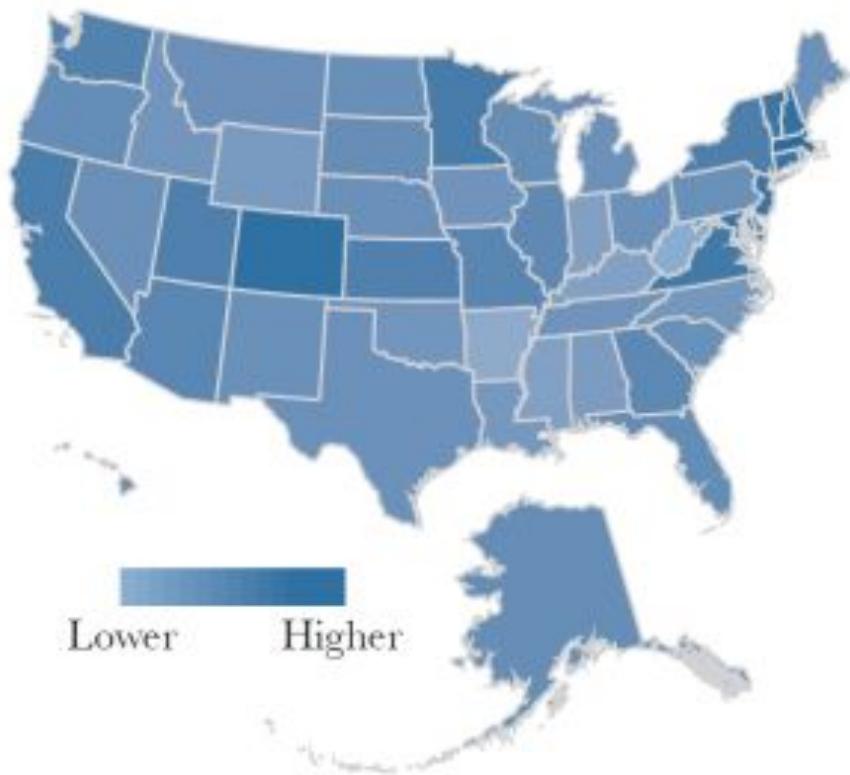
3



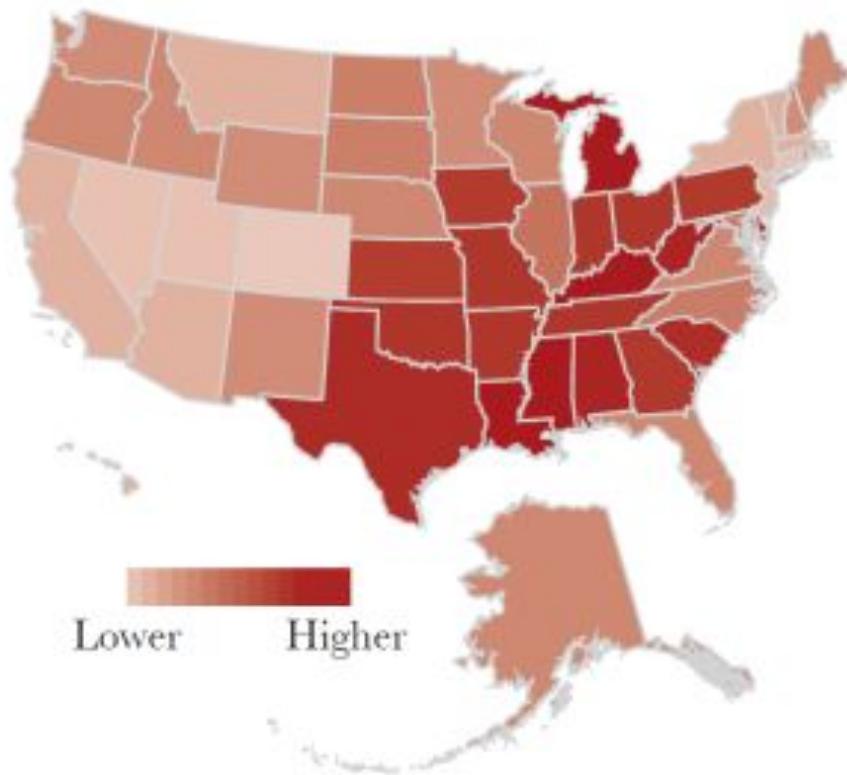


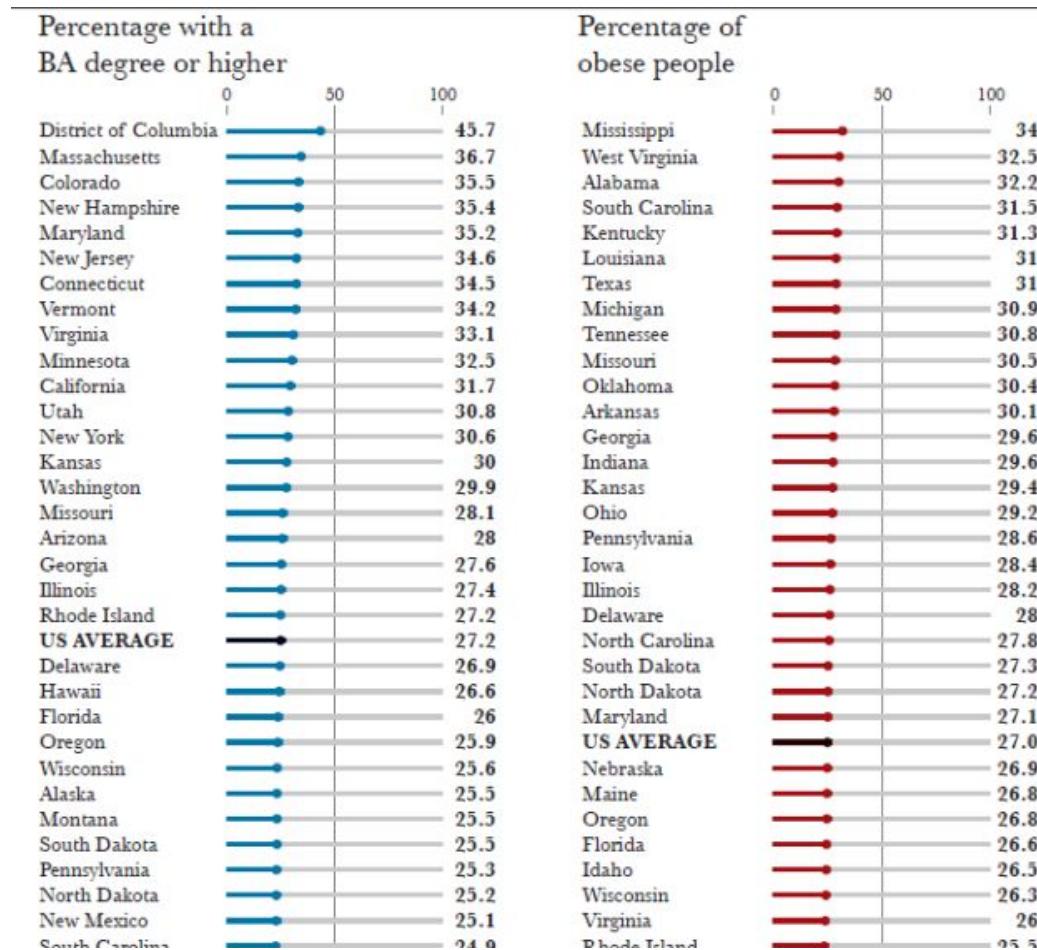


Percentage of people with a
BA degree or higher



Percentage of
obese people





BIAS

From Wikipedia, the free encyclopedia

For common errors in logic, see List of fallacies.



Cognitive biases are systematic patterns of deviation from norm and/or rationality in judgment.^{[1][2]} They are often studied in psychology, sociology and behavioral economics.^[1]

Although the reality of most of these biases is confirmed by reproducible research,^{[3][4]} there are often controversies about how to classify these biases or how to explain them.^[5] Several theoretical causes are known for some cognitive biases, which provides a classification of biases by their common generative mechanism (such as noisy information-processing^[6]). Gerd Gigerenzer has criticized the framing of cognitive biases as errors in judgment, and favors interpreting them as arising from rational deviations from logical thought.^[7]

Explanations include information-processing rules (i.e., mental shortcuts), called heuristics, that the brain uses to

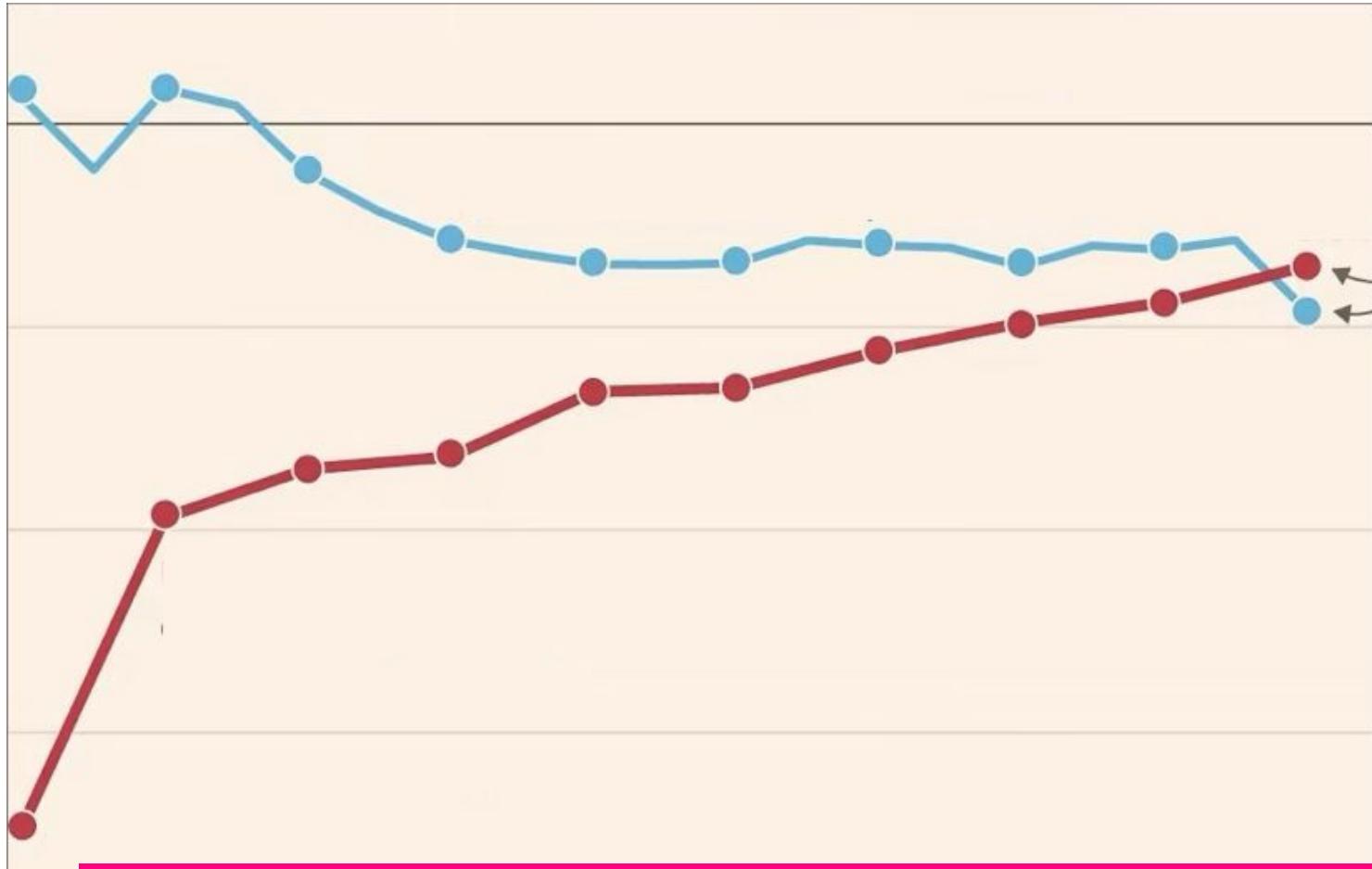


Chart by Financial Times • Example conceived by Raffaele Mastrolonardo of SkyNews24

Without more support, the **energy bills squeeze** will hit UK households far harder than the **global financial crisis and austerity**, especially lower-earners

Change in net disposable income after housing costs (%), by household income level



Sources: Tony Blair Institute for Global Change; IFS; FT analysis

FT graphic: John Burn-Murdoch / @jburnmurdoch

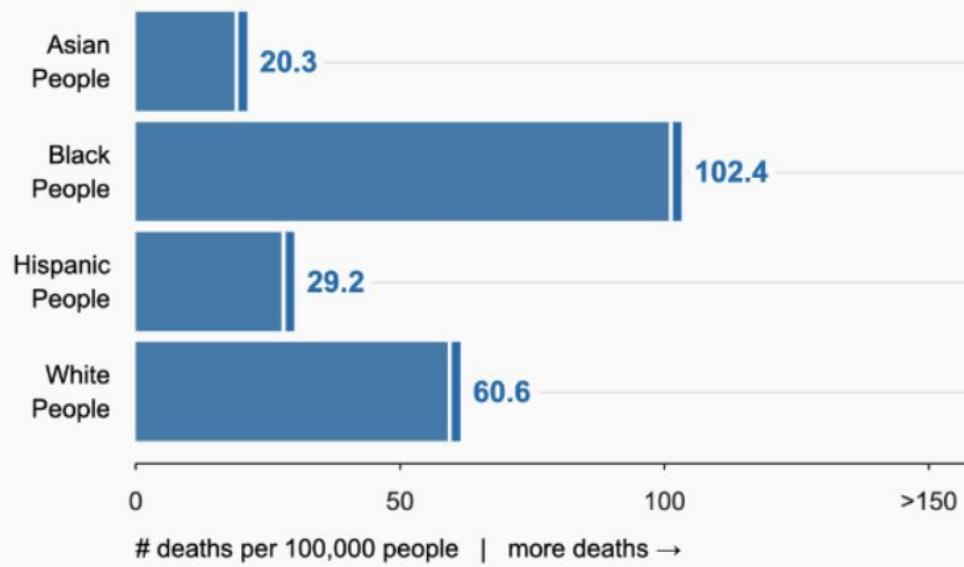
© FT

Chart by Financial Times • Example conceived by Raffaele Mastrolonardo of SkyNews24

Early deaths from heart disease

How many younger adults died each year from heart disease? Crude mortality rates for U.S. adults, ages 15-64

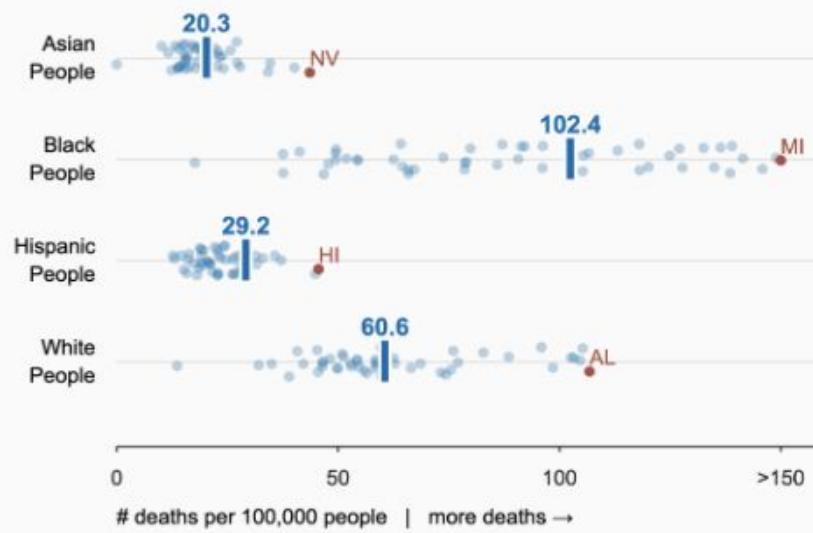
= National Average



Health outcomes vary widely by geography.
Michigan and **Alabama** have the most
deaths from heart disease for younger
Black and White adults.

How many younger adults died each year from heart
disease? Crude mortality rates for U.S. adults, ages 15-64

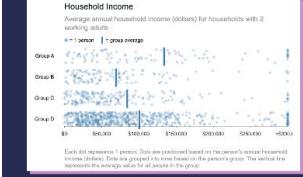
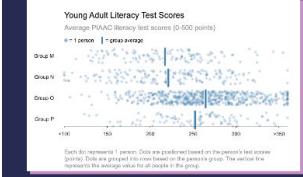
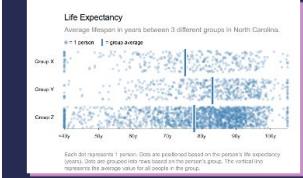
■ = National Average



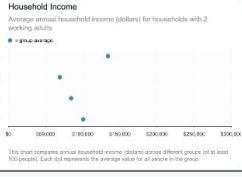
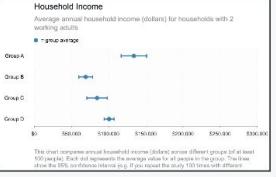
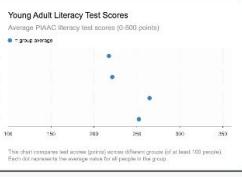
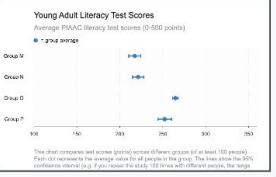
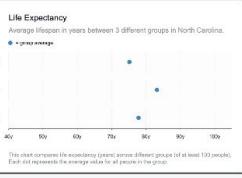
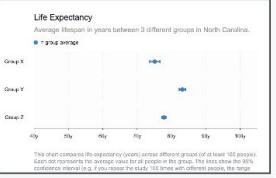
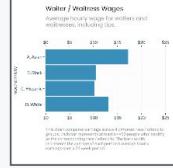
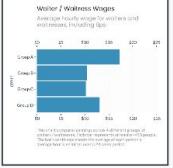
"Must Be a Tuesday": Affect, Attribution, and Geographic Variability in Equity-Oriented Visualizations of Population Health Disparities

• Holder, E., & Padilla, L.

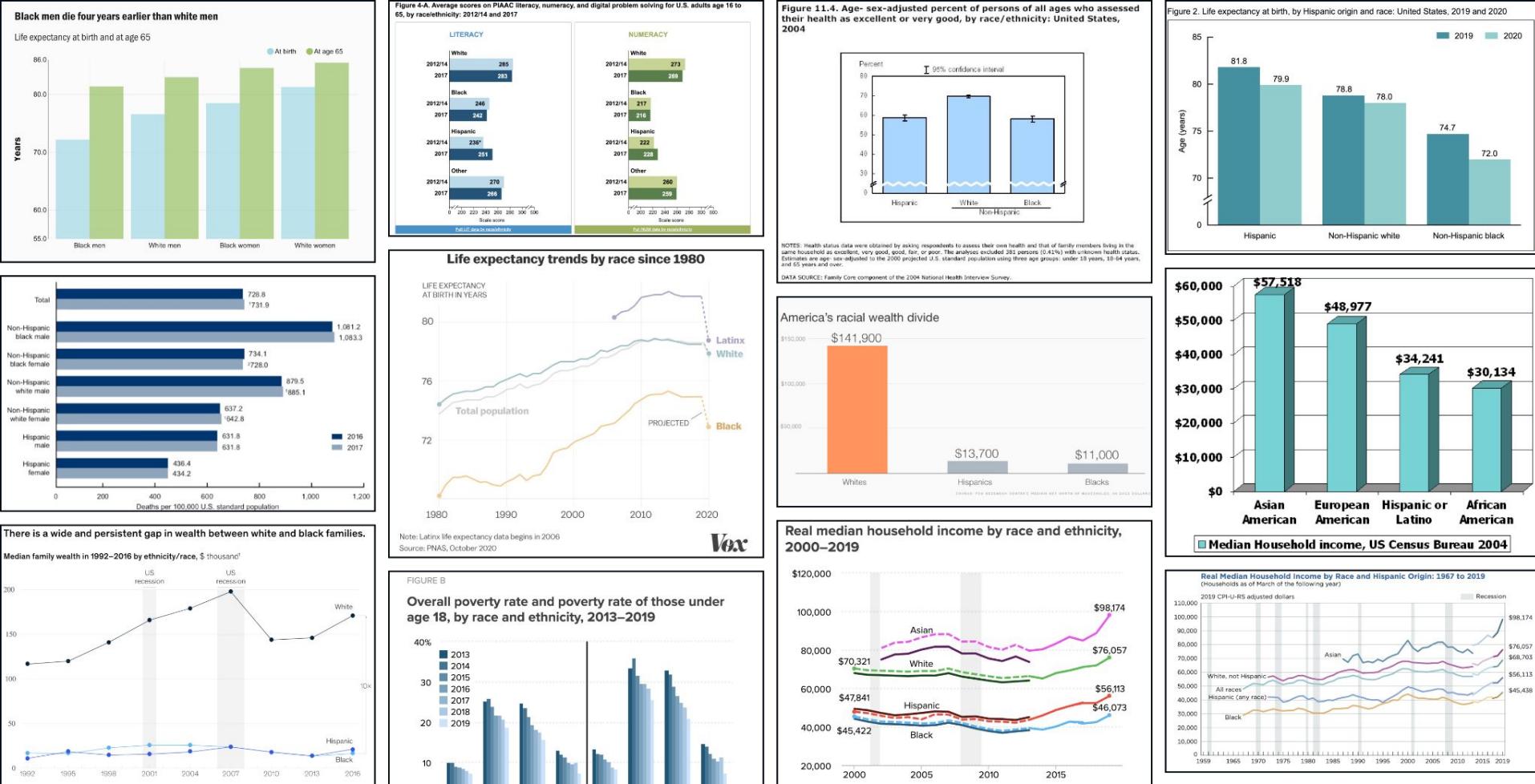
conditions showing variability



conditions hiding variability



Unfair Comparisons: How Visualizing Social Inequality Can Make It Worse • Holder, E., & Padilla, L.



COMMUNICATING DATA

Finding data sources



A GOOD DATA SOURCING PLAN IS A CONSEQUENCE OF THE DATA QUESTIONS YOU WANT TO ANSWER

WHAT **VARIABLES** DO I NEED TO ANSWER MY QUESTIONS?

WHAT **TIME** PERIOD DO I NEED TO COVER?

WHAT **GEOGRAPHICAL AREA** AND **SPATIAL GRANULARITY** DO I NEED?

WHAT ARE POSSIBLE **PRIMARY SOURCES**?

HOW MUCH **TIME** AND WHICH **SKILLS** DO I HAVE?

PORTALS: PUBLIC OPEN DATA

NATIONAL

National Statistical Offices

ISTAT (Italy), Bureau of the Census (USA), ONS (UK). See list here.

It is the place to find dataset related to a country demography, to the census, to socio-economic issues, jobs, and a lot of other topics. It usually contains geographical data files as well.

Public Administrations and Agencies

dati.gov.it (Italy), Greater London Authority (UK), Police Departments (USA)

Look for data portals dedicated to public entities, which can be like municipalities, civil protection, public universities, regional authorities, State Property Office, etc.

Government

data.gov (USA), Environmental Protection Agency (USA), NASA (USA)

In Italy (and probably elsewhere too) each Ministry has its own open data portal, where you can find data related to its activities. Some interesting ones are MIUR (Education) or Ministero dell'Interno (Interior) which contains an historical archive of election data. See here a list of US federal agencies and their data portals.

PORTALS: PUBLIC OPEN DATA

GLOBAL

EUROPE: [data.europa.eu](#)

Contains links to the different European / EU open data portals for EU countries, EU institutions and EU Stats Office

AFRICA: [Open Africa](#) or [Data Africa](#)

Volunteer-driven and independent open data portals with dataset from the continent

UNITED NATIONS: [UNdata](#)

Portal aggregating in a single place all the datasets produced by the UN departments and agencies, like UNICEF, UNHCR, WHO, ITU. You will also find that each agency also has its own Data page.

WORLD BANK: [World Bank Open Data](#)

Global development data and data on other topics related to the World Bank activities.

OECD: [OECD Data](#)

OECD datasets, including the famous [Better Life Index](#).

WORLD HEALTH ORGANIZATION: [Global Health Observatory data repository](#)

The Global Health Observatory, WHO's gateway to health-related statistics for more than 1000 indicators for its 194 Member States

PORTALS: PUBLIC OPEN DATA

AGGREGATORS

DataHub

Portal aggregating open data from various source. Offers a “data conciergerie” that will find you the data you need.

DataPortals

A data portal of data portals! (so does not contain data)

Google Public Data Explorer & Google Dataset Search

The first offers direct access to public datasets and lets you explore & visualize them directly within the interface. The second is just like Google Search, but for data.

Google Dataset Search

Just like Google Search, but for data.

Data.world

Often labeled “the Github for data”. Anybody can publish their curated datasets and follow others for updates. Offers nice integration options. Also the actual GitHub is a trove of datasets (see [FiveThirtyEight](#))

Humanitarian Data Exchange

Best place for all datasets that could be useful to humanitarian action - this includes also datasets that are very relevant in general (like population; schools; COVID-19, etc.)

PORTALS: PUBLIC OPEN DATA

CURATED LISTS

Data Is Plural (and archive)

r/datasets

Awesome Public Datasets

Storytelling with Data - Public Data Sources

Our World in Data

...

COMPANY DATA PAGES

Leadership intersectional representation

U.S.

● 2020 REPORT

● 2021 REPORT

21.1%
MEN
20.7%
MEN

8.5%
WOMEN
8.7%
WOMEN



ASIAN+

1.5%
MEN
1.8%
MEN

1.1%
WOMEN
1.3%
WOMEN



BLACK+

2.2%
MEN
2.5%
MEN

1.5%
WOMEN
1.4%
WOMEN



LATINX+

0.2%
MEN
N/A
MEN

0.3%
WOMEN
N/A
WOMEN



NATIVE
AMERICAN+

49.1%
MEN
47.7%
MEN

16.8%
WOMEN
17.8%
WOMEN



WHITE+

[Google Workforce Diversity annual report](#)

Companies will usually have a transparency section where they disclose data (willingly or because of legal obligations).

If they don't: check on a country's business registry, some of which provide also open/free data downloads

COMPANY DATA PAGES

Quarterly Earnings

2Q 2021	1Q 2021	4Q 2020	3Q 2020
Webcast Earnings Release Earnings Slides Balance Sheet Income Statement Earnings Call Transcript Follow Up Call Transcript Form 10-Q	Webcast Earnings Release Earnings Slides Earnings Call Transcript Follow Up Call Transcript Form 10-Q	Webcast Earnings Release Earnings Slides Earnings Call Transcript Follow Up Call Transcript Form 10-K	Webcast Earnings Release Earnings Slides Earnings Call Transcript Follow Up Call Transcript Form 10-Q

[Facebook Investor's Relations page](#)

Companies will usually have a transparency section where they disclose data (willingly or because of legal obligations).

If they don't: check on a country's business registry, some of which provide also open/free data downloads

COMPANY DATA PAGES

GOVERNMENT & LAW ENFORCEMENT REPORTS

Companies will usually have a transparency section where they disclose data about government requests for either content takedown or users' data.

([Google](#), [Facebook](#), [Microsoft..](#) and [here](#) a complete list)

FINANCIAL TRANSPARENCY DISCLOSURES

Many companies will publish financial transparency data on their website ([Shell](#), [Facebook](#)). If you don't find it, you can always look in public business filings and registries (for example: [SEC filings database](#) for the US, [Registro Imprese](#) for Italy, [Company House data for UK](#))

WORKFORCE DIVERSITY & INCLUSION REPORTS

Companies will often publish data on their workforce, to show hiring efforts towards a more diverse workforce. See [Google](#), [LinkedIn](#), [Pinterest](#), [Apple](#).

USER-GENERATED DATA

Some websites will publish interfaces to explore how users interact with their service. See [Google Trends](#) or [Stack Overflow Trends](#).

CROWDSOURCING



Obiezione Respinta
Mappa gli obiettori!

Info

Manifesto

FAQ

Segnala

Sos Aborto

Eventi & Stampa

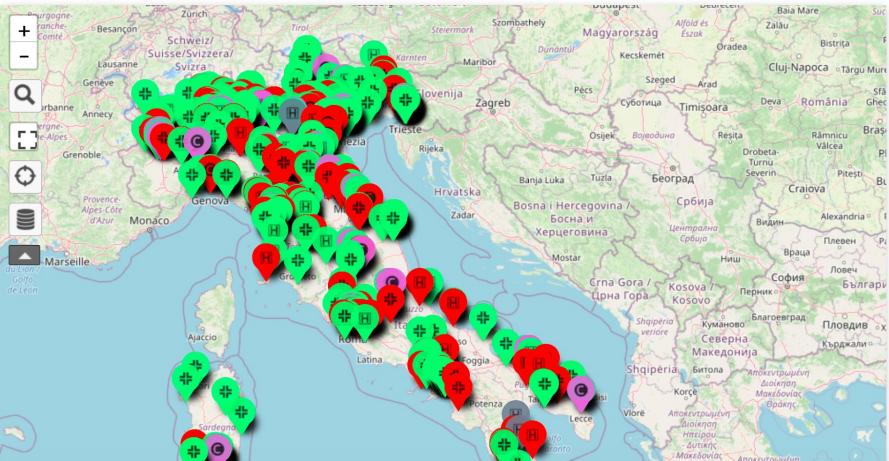
Telefono

Grafica

[Twitter icon](#) [Facebook icon](#) [Instagram icon](#) [Email icon](#)

Obiezione Respinta

Questa è la mappa di Obiezione Respinta. Puoi esplorarla usando la funzione "Cerca" per trovare la città che ti interessa. Potrai trovare dei segnalini con simboli diversi che indicano il tipo di servizio. (Farmacie, consultori, Ospedali, Queersutorie...ecc.). Il colore indica invece il tipo di segnalazione: quelli verdi sono stati considerati adeguati da altri utenti, quelli rossi no. Se cliccherai sul segnalino troverai ulteriori informazioni (in quella farmacia si rifiutano di vendere la pillola, in questo consultorio sono stata visitata correttamente...ecc) Puoi contribuire anche tu a questa inchiesta! Se vuoi fare una segnalazione guarda la pagina [Come segnalare](#).



The Italian project [Obiezione Respinta](#) uses a crowdsourced map to collect data about women's actual access to abortion and reproductive health rights.

Crowdsourcing leverages the wisdom of the crowd to build a dataset.

The core ideas is that data can be collected and refined by a more or less organized team of people, each carrying out one smaller tasks and contributing to the bigger dataset.

[See Wikipedia, Open Street Maps... many Google products also work thanks to crowdsourcing, perhaps without users' knowledge]

CROWDSOURCING

TOOLS:

FORMS AND SURVEYS

- [Google Forms](#)
- [TypeForm / FormBricks](#)

DEDICATED PLATFORMS

- [Ushahidi and Crowdmap](#)
- [Open Data Kit](#)

In all cases, beware of privacy, consent and framing!

[Here's a primer](#) on the main **ethical concerns**, and a [project/blog](#) with many articles on how to carry out **inclusive surveys**.

WEBSITE DATA

All Nobel Prizes

Share this



Between 1901 and 2021, the Nobel Prizes and the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel were awarded 609 times to 975 people and organisations. With some receiving the Nobel Prize more than once, this makes a total of 943 individuals and 25 organisations. Below, you can view the full list of Nobel Prizes and Nobel Prize laureates.

2021

The Nobel Prize in Physics 2021

"for groundbreaking contributions to our under-

Syukuro Manabe and Klaus Hasselmann "for th-
quantifying variability and reliably predicting gl-

Giorgio Parisi "for the discovery of the interpla-
from atomic to planetary scales"

The Nobel Prize in Chemistry 2021

Benjamin List and David MacMillan "for the de-

The Nobel Prize in Physics 2021

Syukuro Manabe
Klaus Hasselmann
Giorgio Parisi

Share this



Giorgio Parisi Facts



III. Niklas Elmehed © Nobel Prize Outreach

Giorgio Parisi
The Nobel Prize in Physics 2021

Born: 4 August 1948, Rome, Italy

Affiliation at the time of the award: Sapienza University of Rome, Rome, Italy

Prize motivation: "for the discovery of the interplay of disorder and fluctuations in physical systems from atomic to planetary scales."

Prize share: 1/2

You can automate the conversion of structured web content in a spreadsheet. This is useful, for example, to work with social media or user-generated data: from Twitter to Wikipedia, from LinkedIn to Imdb.

Mostly needs to be done by programming - working with **scrapers, Linked Open Data, or APIs.**

The Nobel Prize official website, [page listing all Nobel Prizes](#) and [page for Giorgio Parisi's 2021 Nobel Prize in Physics](#)

WEBSITE DATA: SCRAPING

WHAT IS WEB SCRAPING

In its essence, it means to access unstructured data in a web page and save it as a structured dataset.

A scraping program starts by parsing the HTML structure to retrieve the desired contents by navigating the tree.

The retrieved content is then saved in structured machine-readable format (ex. a Python dictionary).

easy

Scraping inside Google Sheets with dedicated formulas; or use [Chrome Scraper Extension](#).

intermediate

Use a **program** like [Import.io](#) or [Outwit Hub](#) (all are usually quite expensive)

advanced

Write your own script in a programming language of choice. For example, in **Python**, you could use libraries like `requests`, `urllib`, `Beautiful Soup` e `lxml` or a framework like `Scrapy`).

WEBSITE DATA: SCRAPING

```
▶ def get_root(url):
    r = requests.get(url)
    root = html.fromstring(r.text)
    if r.status_code == 200:
        return root
    else:
        return None
    print("HTTPS request failed with status code %s" % (r.status_code))

def process_winners(url):
    row = {}
    row['prize_id'] = abs(hash(url))
    row['source'] = url
    try:
        row['prize_year'] = int(url.split('/')[-4])
        row['prize_category'] = url.split('/')[-5]
    except ValueError:
        row['prize_year'] = int(url.split('/')[-2])
        row['prize_category'] = url.split('/')[-4]

    response_w = requests.get(url)
    root_w = html.fromstring(response_w.text)
    row['name'] = root_w.xpath('//header/h1')[0].text_content().strip()
    row['image_url'] = root_w.xpath('//div[@class="image"]/picture/source')[-1].get('data-srcset')
    row['prize_share'] = float(Fraction(root_w.xpath('//text()[contains(., "Prize share:")]')[0].split(':')[1].strip()))
```

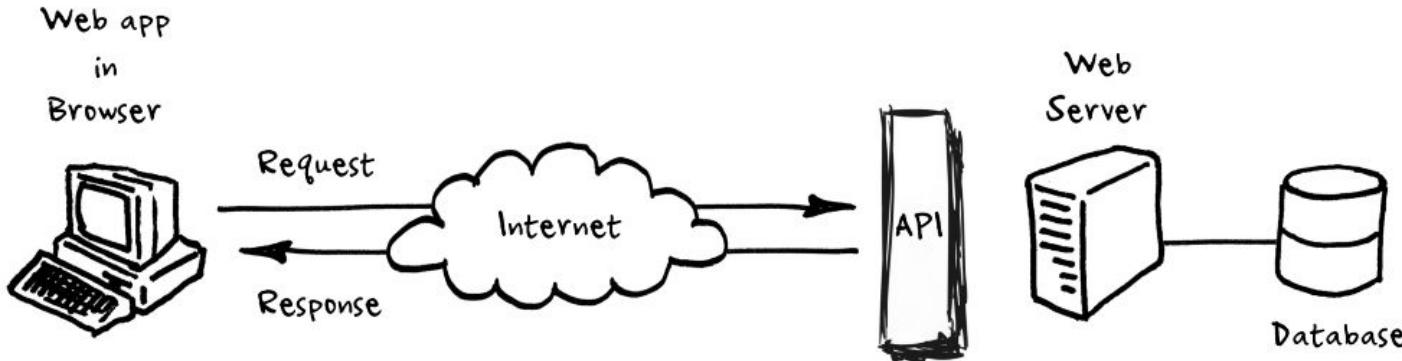
WEBSITE DATA: SCRAPING

prize_year	prize_category	name	image_url	prize_share	death_date	alive_status	birth_year	entity	acceptance
2020	physics	Roger Penrose	https://www.nobelprize.org/images/penrose-11175...	0.50		alive	1931	person	accepted
2020	physics	Reinhard Genzel	https://www.nobelprize.org/images/genzel-11175...	0.25		alive	1952	person	accepted
2020	physics	Andrea Ghez	https://www.nobelprize.org/images/ghez-111760-...	0.25		alive	1965	person	accepted
2020	chemistry	Emmanuelle Charpentier	https://www.nobelprize.org/images/charpentier-...	0.50		alive	1968	person	accepted
2020	chemistry	Jennifer A. Doudna	https://www.nobelprize.org/images/doudna-11176...	0.50		alive	1964	person	accepted

WEBSITE DATA: API

WHAT IS AN API

Application Programming Interface. It is a set of commands that allow web apps to access and exchange data.



What exactly IS an API?, Perry Eising



reddit

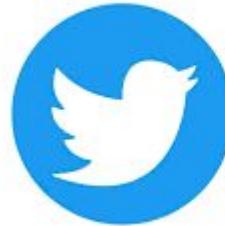
Quora



YouTube



Blogger



WIKIPEDIA
The Free Encyclopedia



EXERCISE → B. CHOOSE A DATA SOURCE

With your group, begin exploring what data sources are available to you already and what data you could obtain (through scraping, crowdsourcing etc.)

FOR NEXT WEEK

Create an account on Datawrapper

Fill in this form with the email you
have used to create the account