

Curs 01 – Why

Ştiinţa datelor vs. analiza statistică

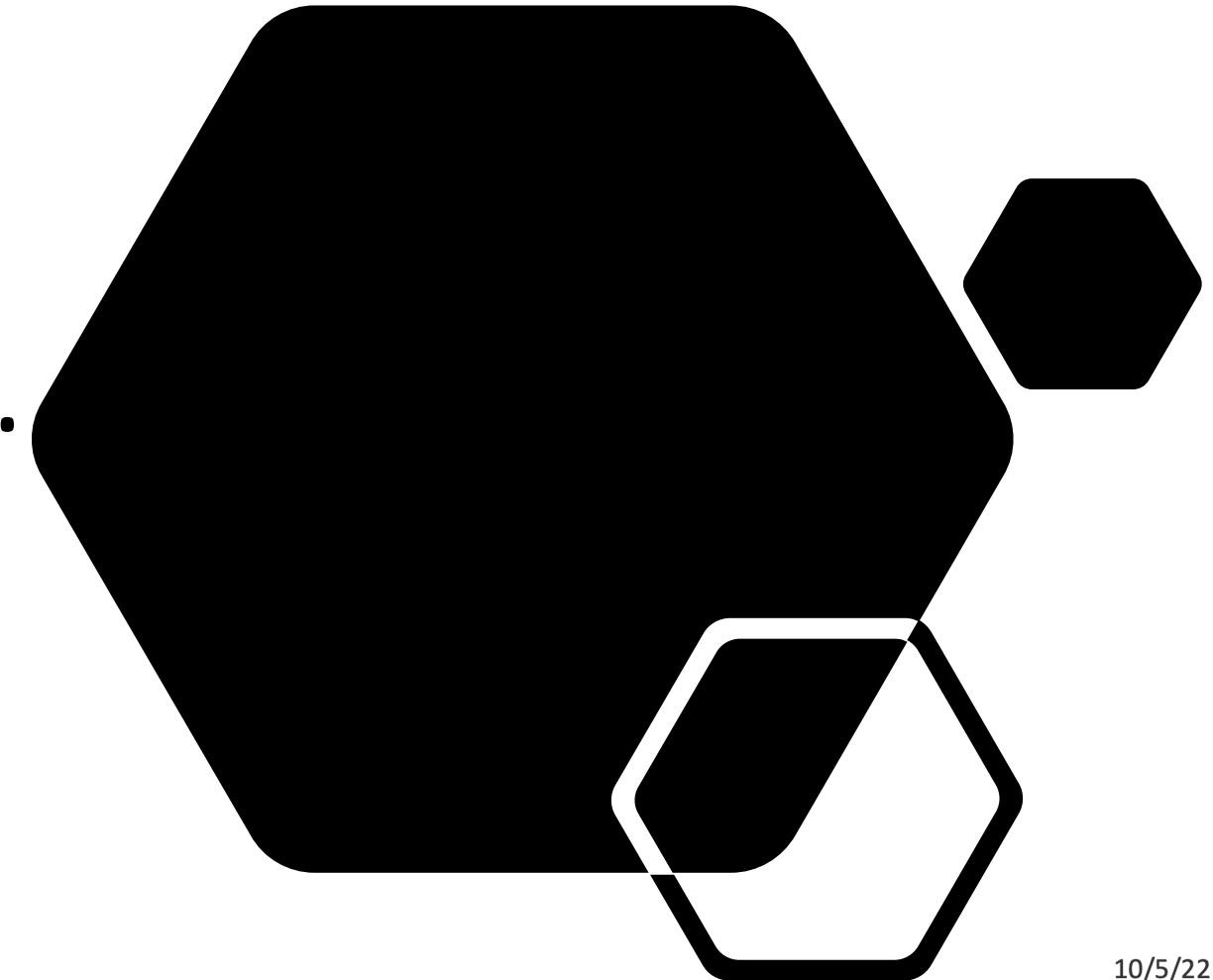
Why?

Obiective ale analizei statistice

Structura cursului

1. Why?
 - Știința datelor vs. analiza statistică
 - Why?
 - Obiective ale analizei statistice
2. Cauzalitate
3. Măsurare
4. Modelare și eșantionare
5. Tehnici de analiză
6. Predicție
7. Programare și ML
8. ML și Deep Learning
9. Producția
10. Why Privacy?
11. Privacy Preserving Algorithms
12. Privacy Architectures and Federated Learning

Ştiinţa datelor vs. analiza statistică



10/5/22

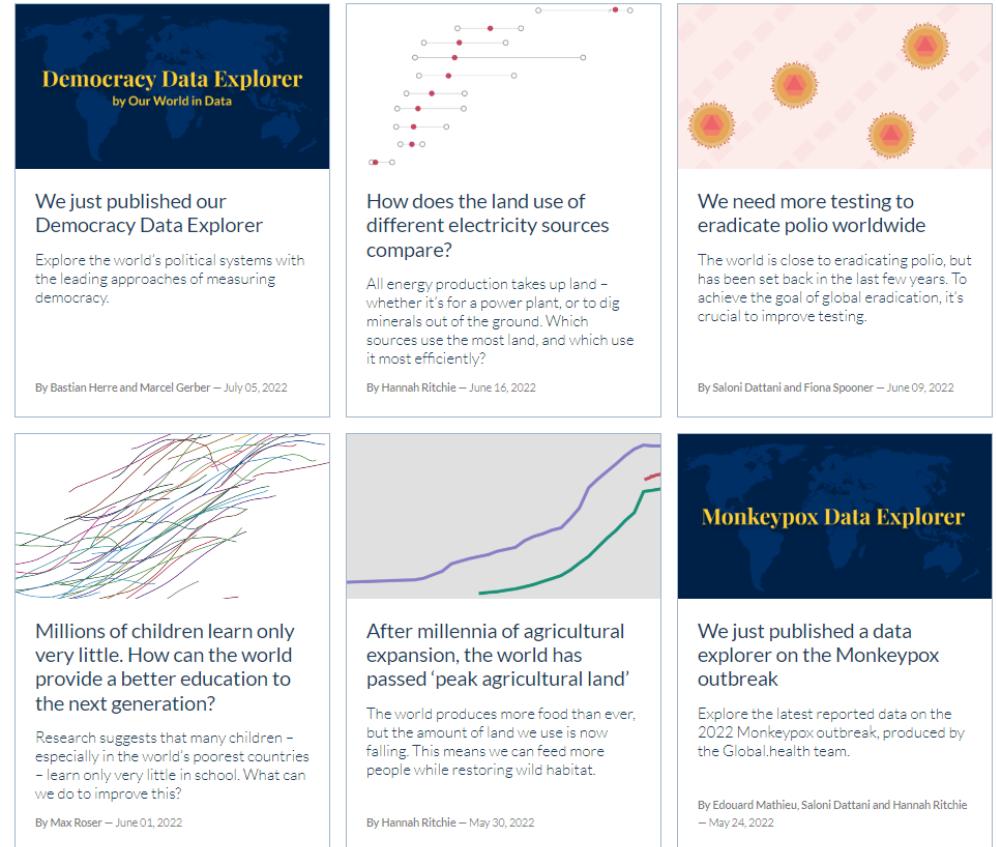


Ce este o știință?

- Studiul prin metode falsificabile al unei forme de ordine a lumii
 - Științe ale **naturii**: chimie, fizică, biologie, geografie, meteorologie, ecologie
 - Științe ale **minții umane**: psihologie
 - Științe ale **creațiilor oamenilor**: istorie, istoria artei, lingvistică, economie, sociologie, științe politice, știința calculatoarelor...
 - **Și știința datelor?**
 - Din acest pdv. nu sunt științe: matematica, filosofia, teologia (?), arta

Care este obiectul științei datelor?

- Analiza statistică a datelor este folosită în practic toate științele naturii, omului și societății – de ex.
 - Încălzirea globală
 - „personalitatea”, „IQ”
 - Inegalitatea socială, mobilitatea socială
- Știința datelor studiază formele de ordine din date
 - „patternuri”



Analiza statistică vs. știința datelor /ML

Analiza statistică a datelor

- Date structurate
 - Unități și variabile coerente
- Modele explicative vin din disciplinele științifice
 - Nu din date
- Obiectul analizei: fenomenul
- Scop: înțelegere și predicție

Ştiința datelor / ML

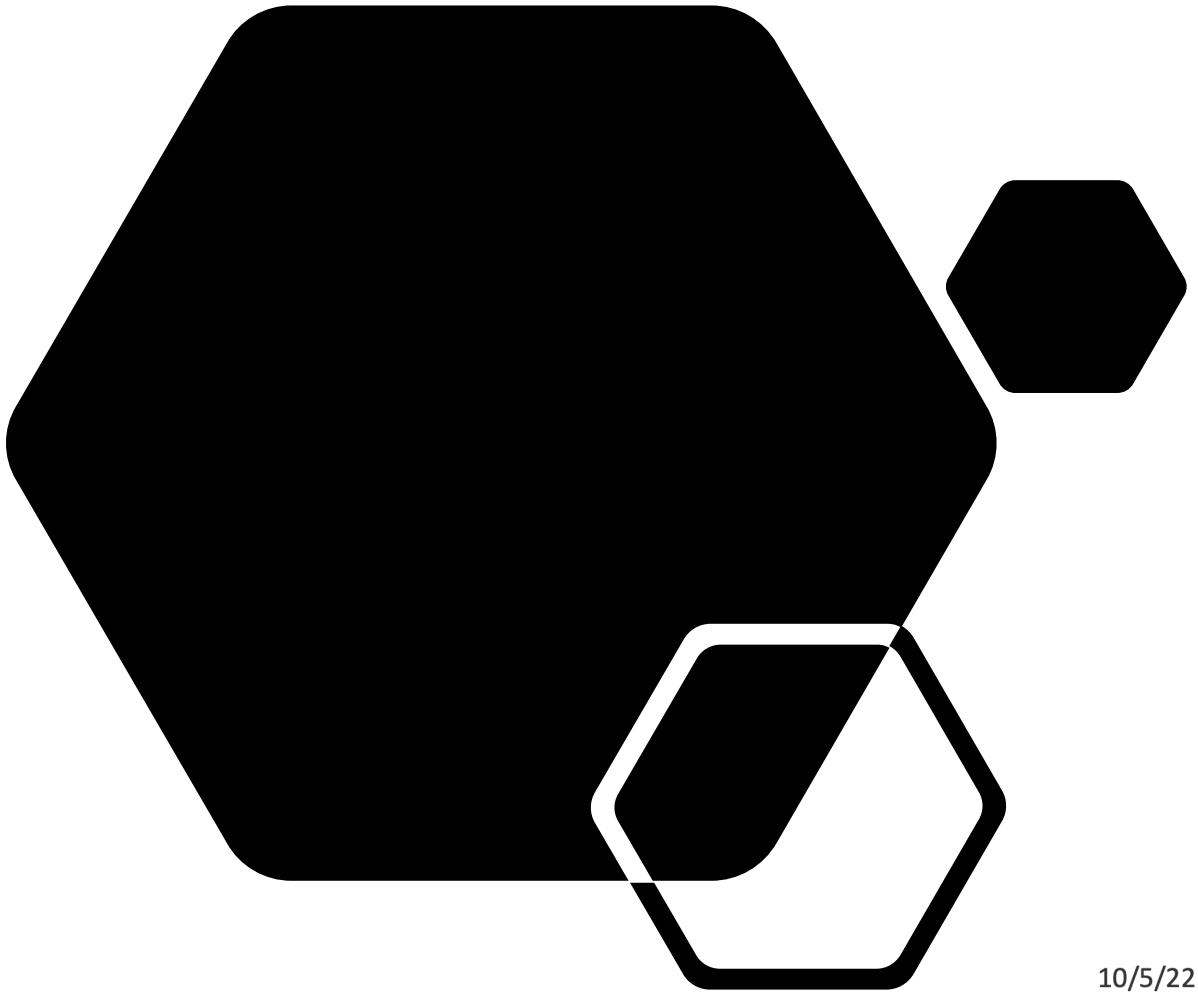
- Date vaste / nestructurate
 - Structuri eterogene
 - Unități și variabile diverse
- Modelele sunt generate automat, ad-hoc din date
- Obiectul analizei: patternuri
- Scop: clasificare și predicție

Importanța analizei statistice pentru DS

- Analiza statistică oferă modelul conceptual pentru DS
- Ce sunt pattern-urile?
 - Distribuții
 - Tendințe
 - Asocieri și corelații
 - Dimensiuni
 - Inegalități
 - Estimări cauzale
 - Categorii
- Interpretări, limitări, mod de folosire



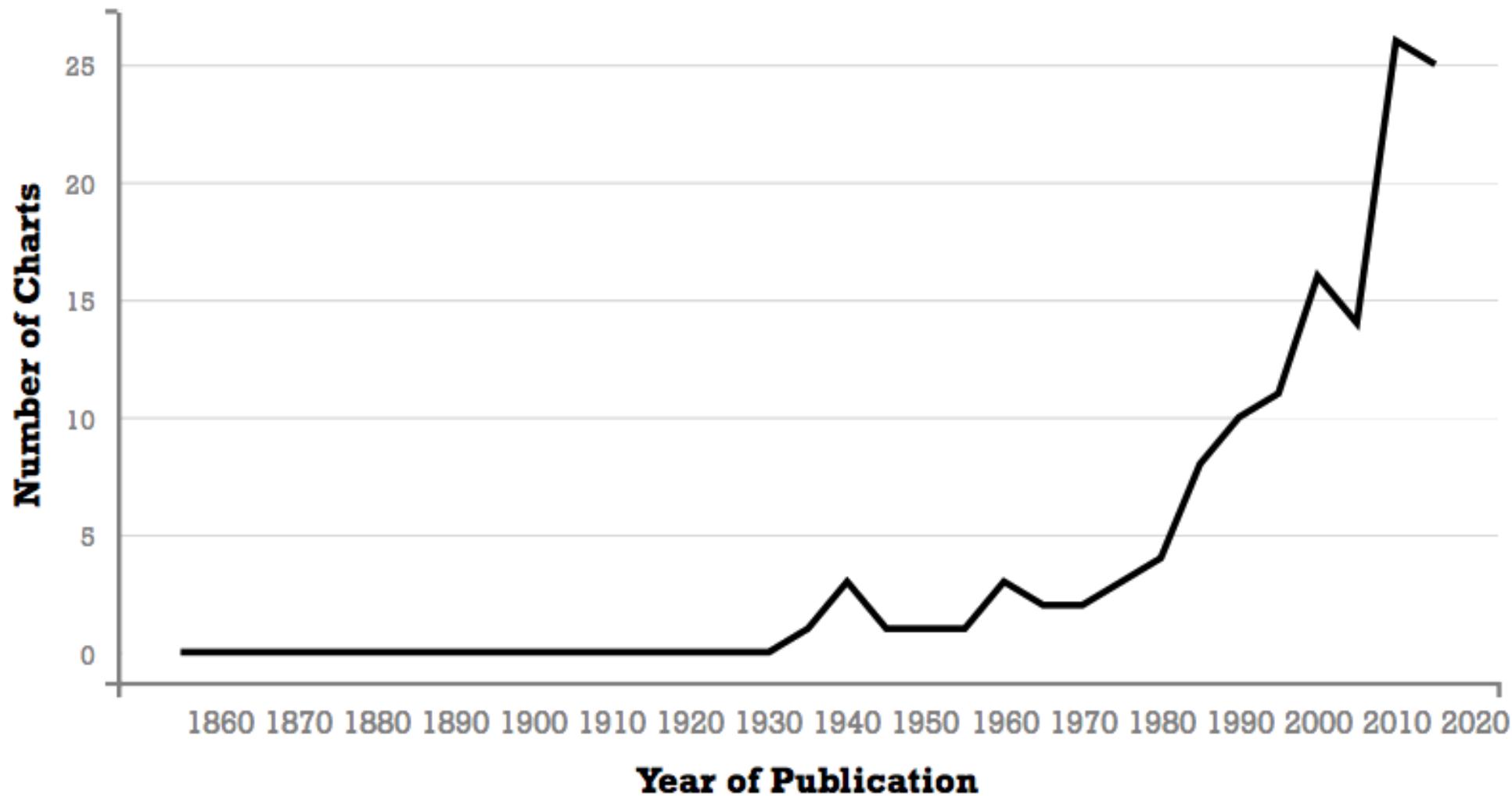
Why?



10/5/22

The Rise of the Chart in the Newspaper

The # of Charts in a September Issue of the New York Times (Checked Every Five Years)



[Sursa](#)

● Big Data
Search term

● Data Science
Search term

● Data Analytics
Search term

+ Add comparison

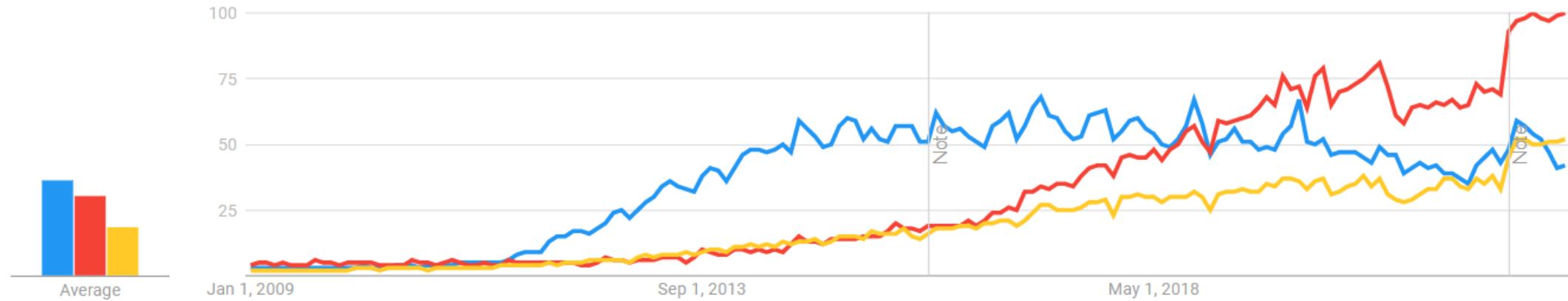
Worldwide ▾

1/1/09 - 8/10/22 ▾

All categories ▾

Web Search ▾

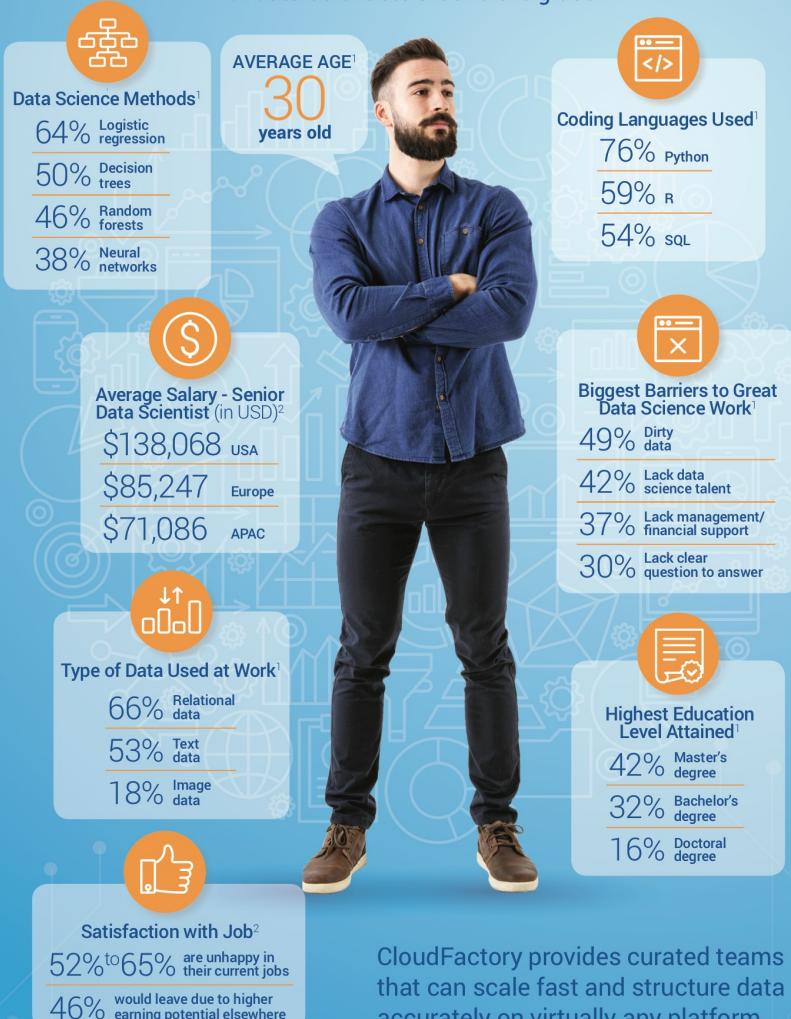
Interest over time ?



Sursa

The Life of a DATA SCIENTIST

Data scientists extract knowledge, insights, or solutions from big data. Here's a look at the life of a data scientist, based on two surveys of data scientists around the globe.



Sursa

10/5/22

Sources:
1 - 2017 State of Data Science & Machine Learning, Kaggle
2 - 2018 Data Science Salary Report, Big Cloud

cloudfactory.com



ANATOMY OF A

DATA SCIENTIST

SALARY

Average salary of data scientists is **\$120,000/year**

BENEFITS

- Harvard Business Review called data science the **"Sexiest Job of the 21st Century"**
- One of the fastest growing careers in the United States
- 94%** of data science graduates have found jobs since 2011



EDUCATION

- 88%** of all data scientists have at least a Master's degree
- 46%** of data scientists have a PhD

SKILLS

- Programming languages (R, Python, SQL, Hive, etc.)
- Statistics
- Multivariable calculus and linear algebra
- Machine learning
- Software engineering
- Wrangle, visualize, and communicate data to management

RESPONSIBILITIES

- Conduct research
- Extract, clean, and analyze data from varied sources
- Solve problems
- Build automation tools
- Communicate findings to management

CAREER POSSIBILITIES

- The majority of data scientists work in the **technology industry**.
- Other options include marketing, consulting, healthcare and pharmaceuticals, finance, government, gaming, and many more.



THE COMPUTER MERCHANT, LTD.
THE IT STAFFING COMPANY

Sursa

11

De la analiza
datelor la
știința
datelor

The six Vs of big data

Big data is a collection of data from various sources, often characterized by what's become known as the 3Vs: *volume, variety and velocity*. Over time, other Vs have been added to descriptions of big data:

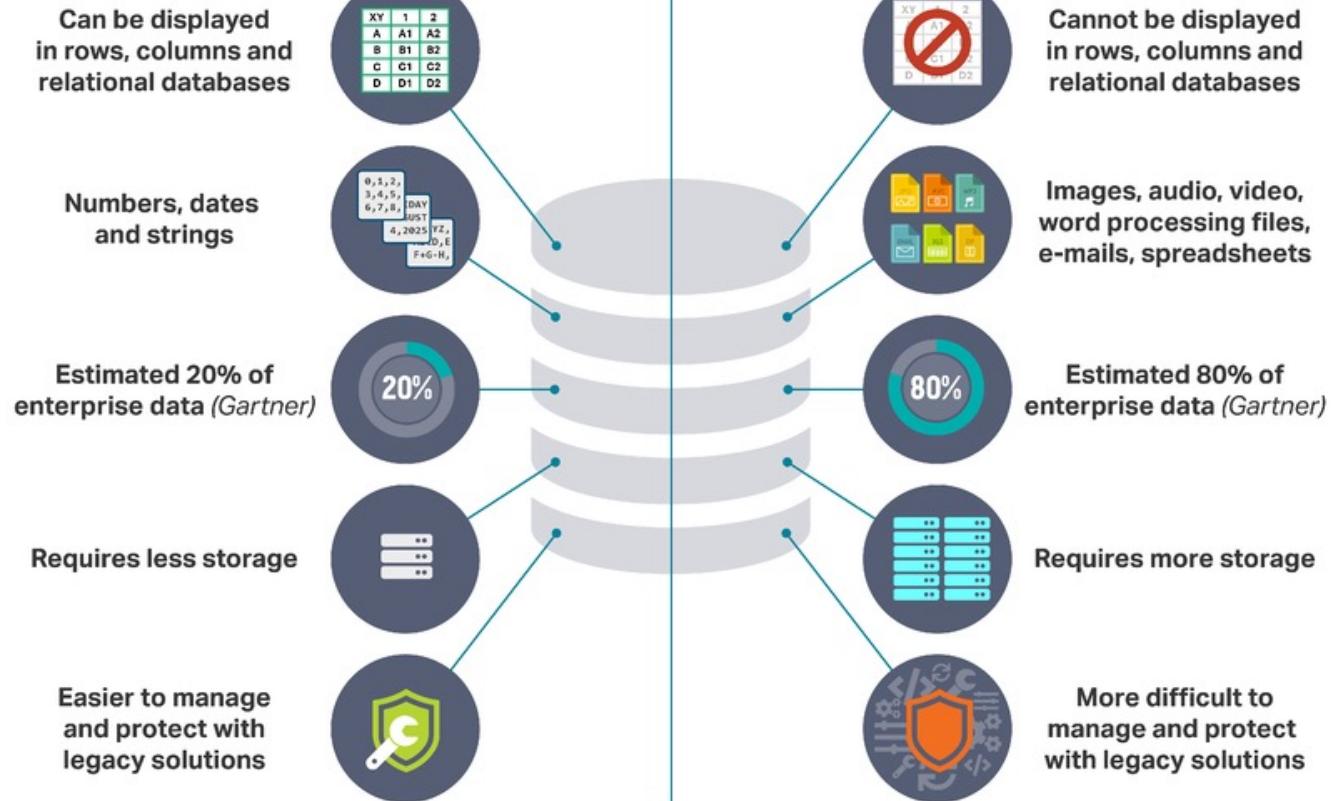
VOLUME	VARIETY	VELOCITY	VERACITY	VALUE	VARIABILITY
The amount of data from myriad sources. 	The types of data: structured, semi-structured, unstructured. 	The speed at which big data is generated. 	The degree to which big data can be trusted. 	The business value of the data collected. 	The ways in which the big data can be used and formatted. 

ICONS: ALEXANDRZ/ADBE STOCK

©2018 TECHTARGET. ALL RIGHTS RESERVED


Date structure vs. nestructurate

Structured Data vs Unstructured Data



Business intelligence vs. știința datelor

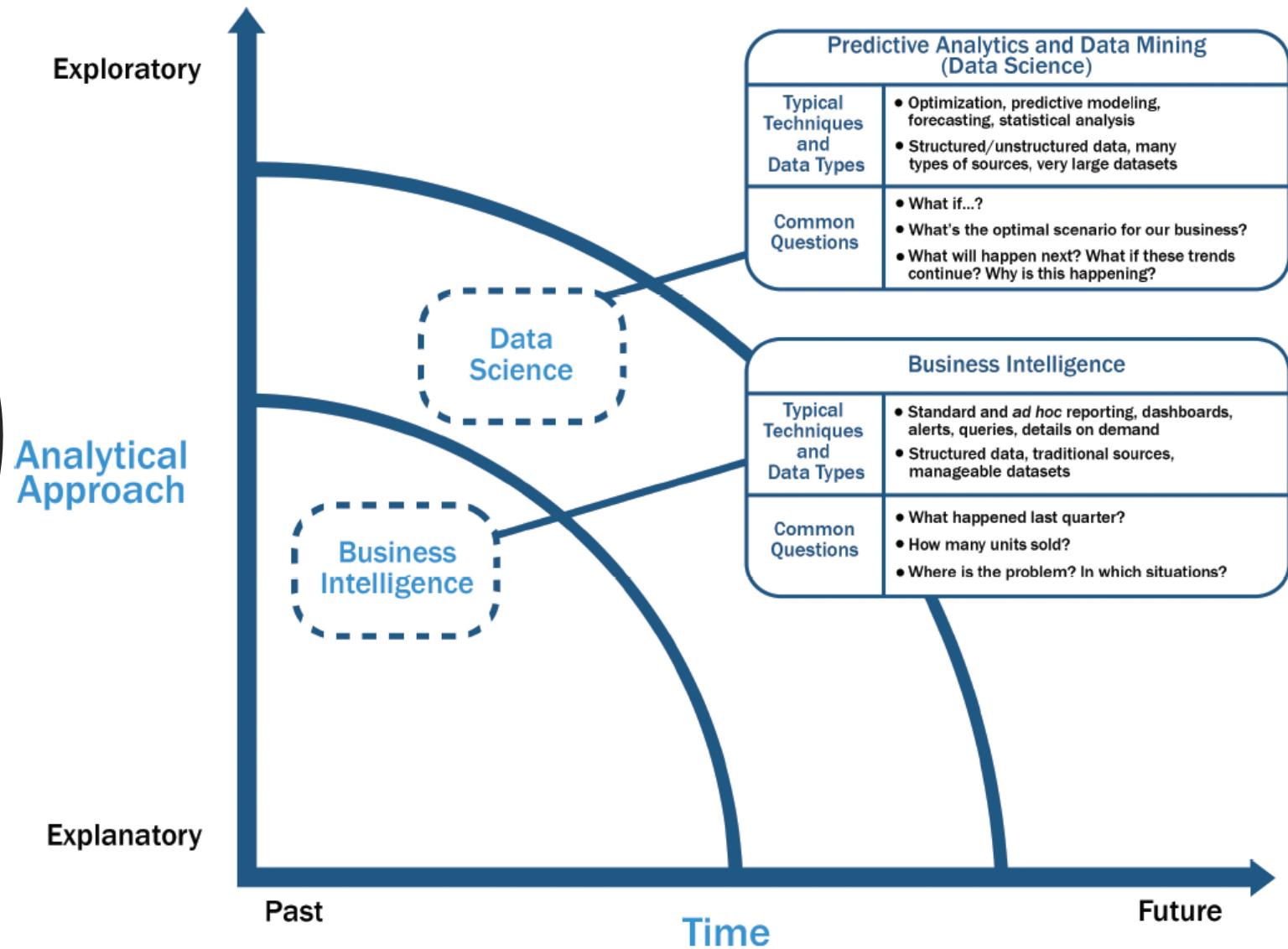


FIGURE 1-8 Comparing BI with Data Science

Utilitatea în business a științei datelor

TABLE 1-2 *Business Drivers for Advanced Analytics*

Business Driver	Examples
Optimize business operations	Sales, pricing, profitability, efficiency
Identify business risk	Customer churn, fraud, default
Predict new business opportunities	Upsell, cross-sell, best new customer prospects
Comply with laws or regulatory requirements	Anti-Money Laundering, Fair Lending, Basel II-III, Sarbanes-Oxley (SOX)

Ecosistemul științei datelor

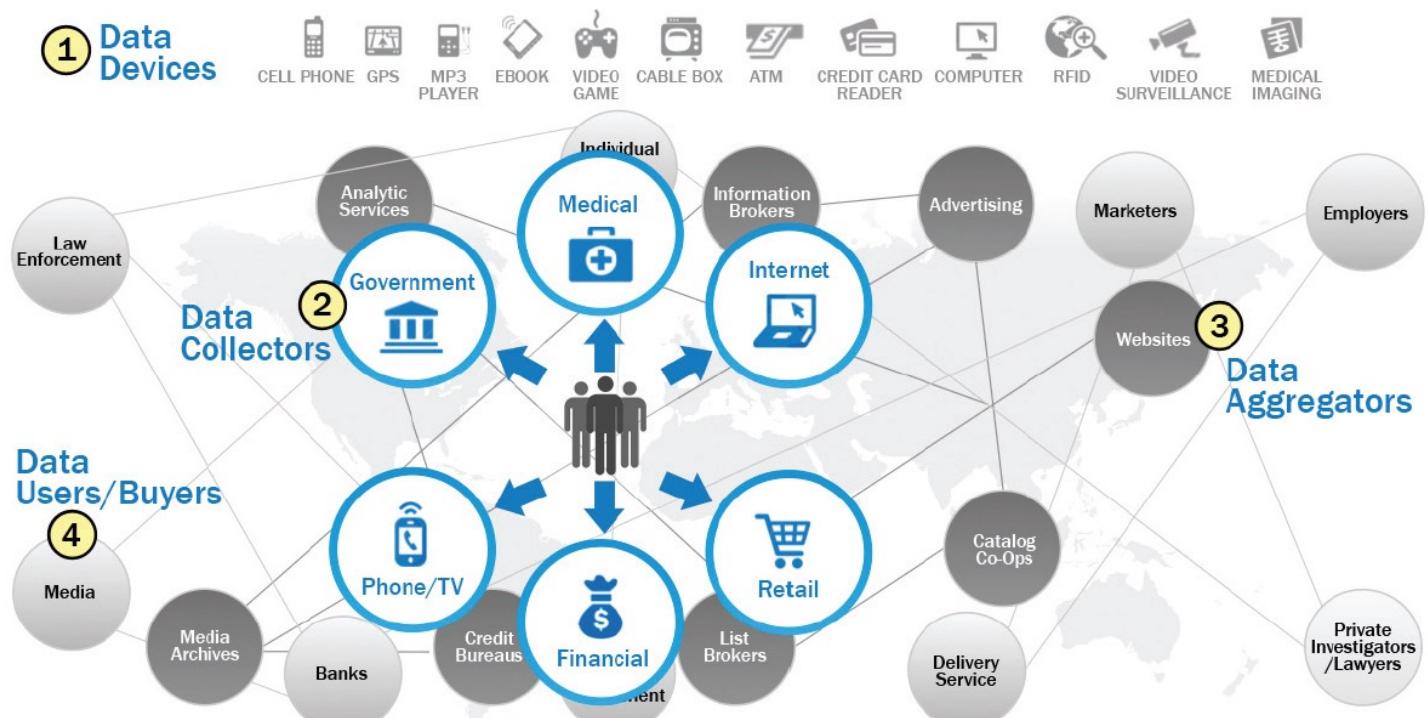
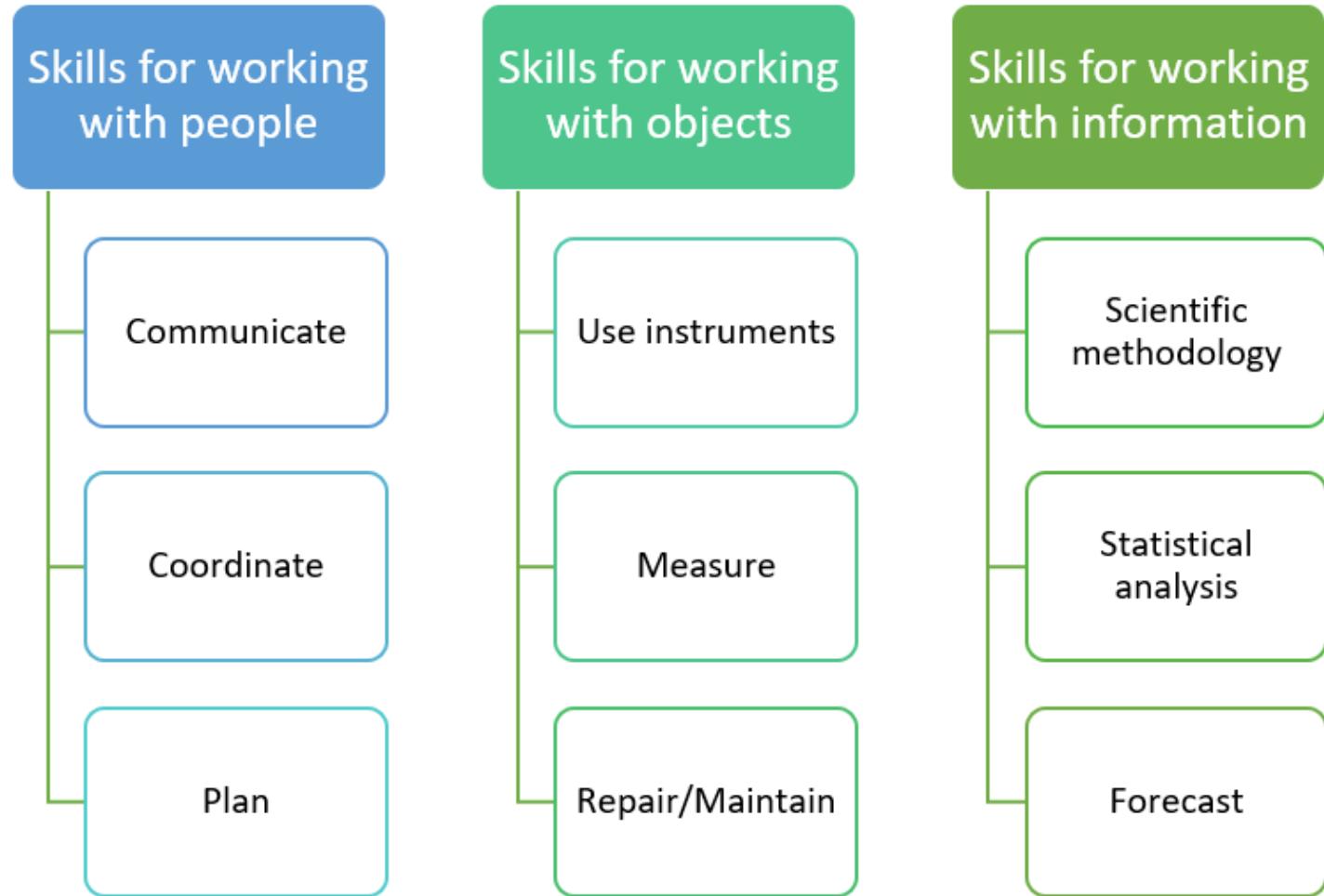
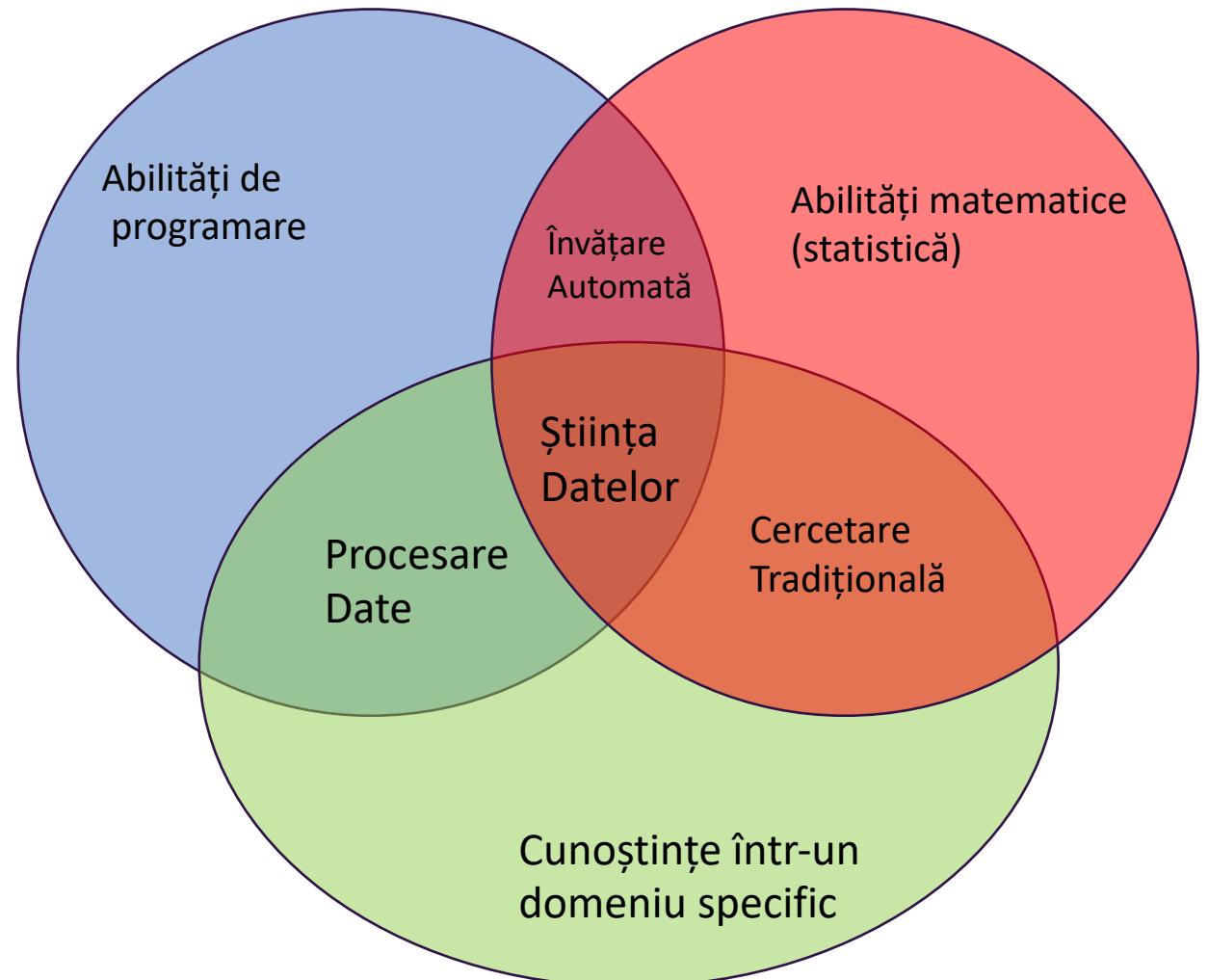


FIGURE 1-11 Emerging Big Data ecosystem



Abilități specifice DS

- Design-ul cercetării
 - Formulare a întrebărilor
 - Colectarea & analiza a datelor
 - Definirea modelului
- Automatizarea analizei
 - Programare
 - Procesarea datelor
- Comunicare
 - Formularea diagnozei, deciziei, predicției etc
 - Comunicarea cu beneficiarii



Importanța Științei Datelor – personalități

Daryl Morey [[Wiki](#), [Interview](#)]

- MIT Sloan Sports Analytics.
- Directorul general al Houston Racket (baschet)

Mircea Lucescu

- Antrenor român de fotbal [[Wiki](#)]

Nate Silver [[Wiki](#), [FiveThirtyEight](#)]

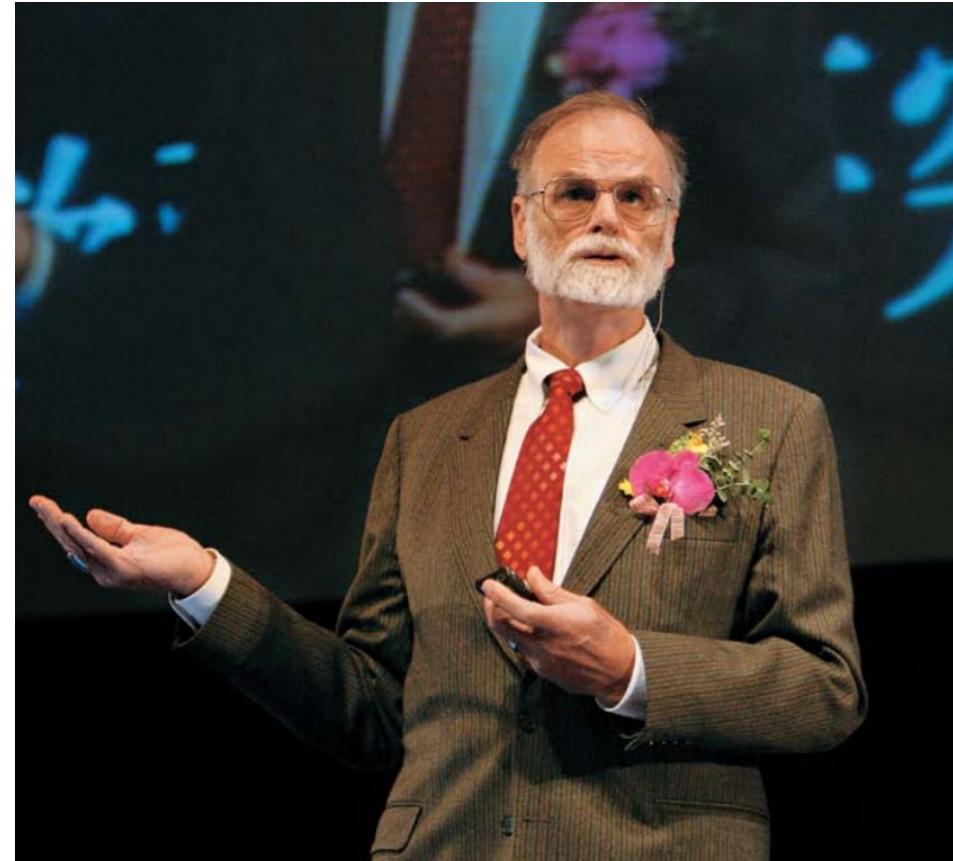
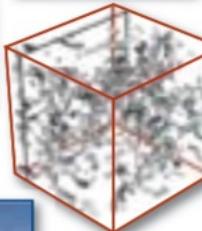
- Predictii cu ajutorul datelor publice
- Cine o să câștige alegerile ?

Science Paradigms

- Thousand years ago:
science was **empirical**
describing natural phenomena
- Last few hundred years:
theoretical branch
using models, generalizations
- Last few decades:
a computational branch
simulating complex phenomena
- Today: **data exploration** (eScience)
unify theory, experiment, and simulation
 - Data captured by instruments or generated by simulator
 - Processed by software
 - Information/knowledge stored in computer
 - Scientist analyzes database/files using data management and statistics

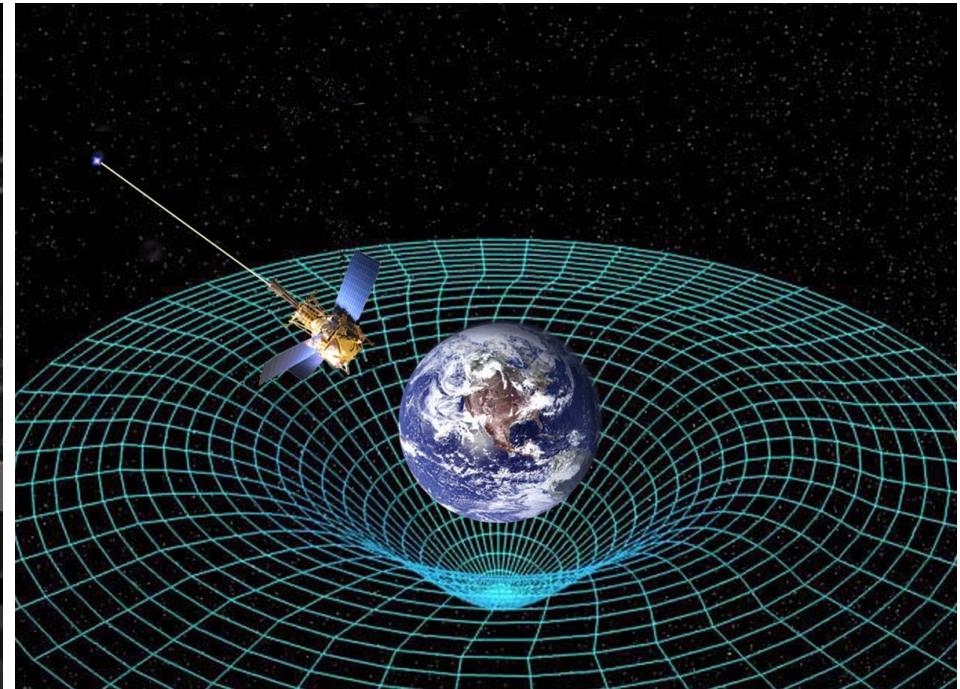
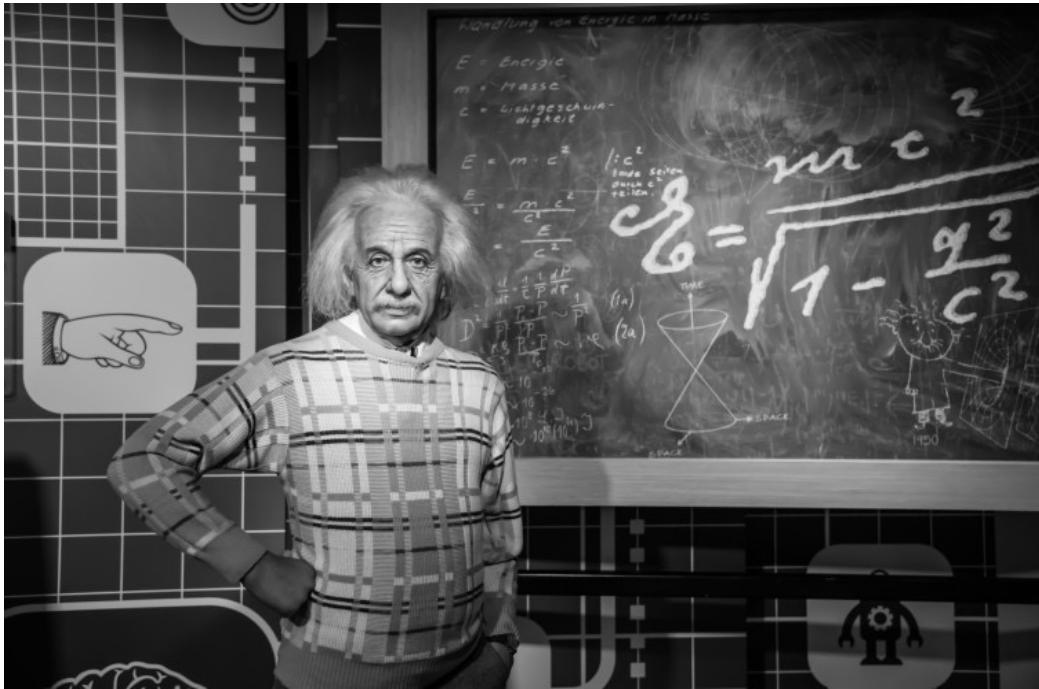


$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G p}{3} - K \frac{c^2}{a^2}$$



Jim Gray's Fourth paradigm of science
[[Hey et al.](#)]

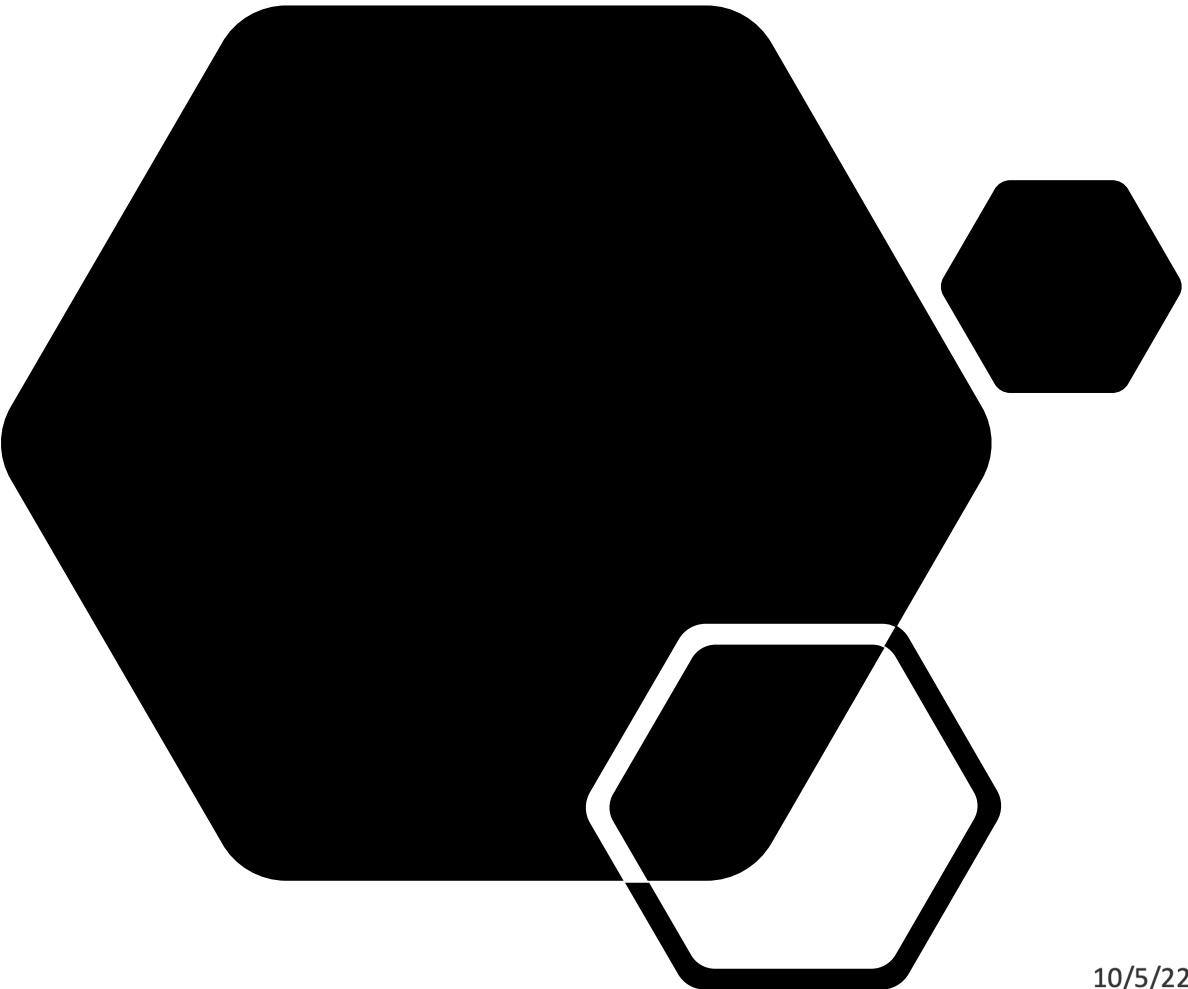
O nouă Teorie a Relativității?



Vezi: [Einstein's Relativity and Everyday Life, de Clifford M. Will](#)

Obiective ale analizei datelor

Utilizarea datelor... de la străbunicii Big
Data până azi



10/5/22

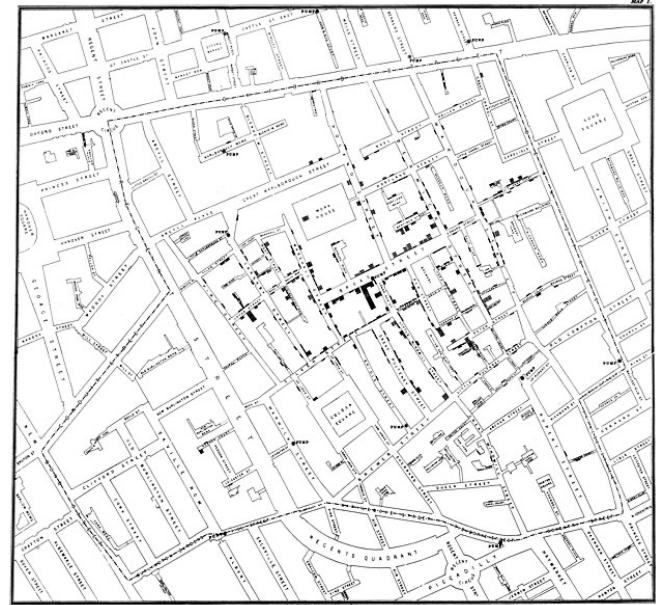
Focus pe esențial

Similitudini remarcabile în ultimii 250 de ani

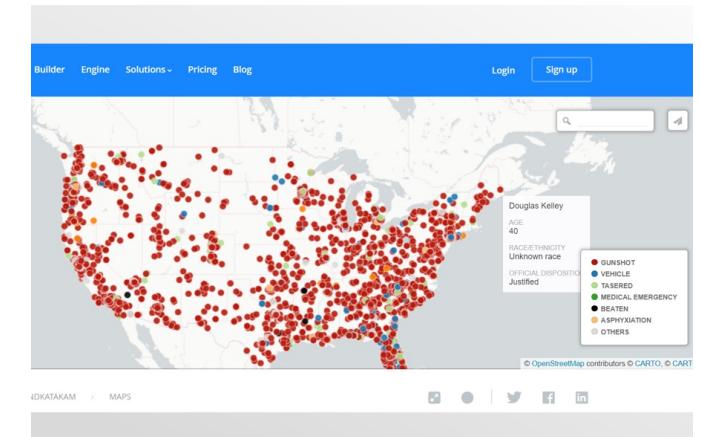
- Reducerea complexității – găsirea esențialului
- Evidențierea mesajului

Transformări datorate tehnologiei

- Sinteză unui volum mult mai mare de date
- Vizualizări dinamice
 - Includerea timpului ca dimensiune
- Vizualizări interactive
 - Stratificarea variabilelor prezentate
- „Eye candy”



John Snow, 1854



Fatal encounters, 2015

Datele permit...

1. Priviri de ansamblu („bird's eye view”)

- Sinteza situației în câteva dimensiuni („variabile” și „unități”)
- Problematica măsurării și a reducerii dimensionalității

2. Evidențierea deficitului și excesului

- Distribuții: ce e prea mult? Ce e prea puțin?

3. Tendențe (evoluții, involuții): încotro merge lumea?

- Serii de timp (Predictii, extrapolări)

4. Inegalități

- Asocieri și corelații

5. Clasificări

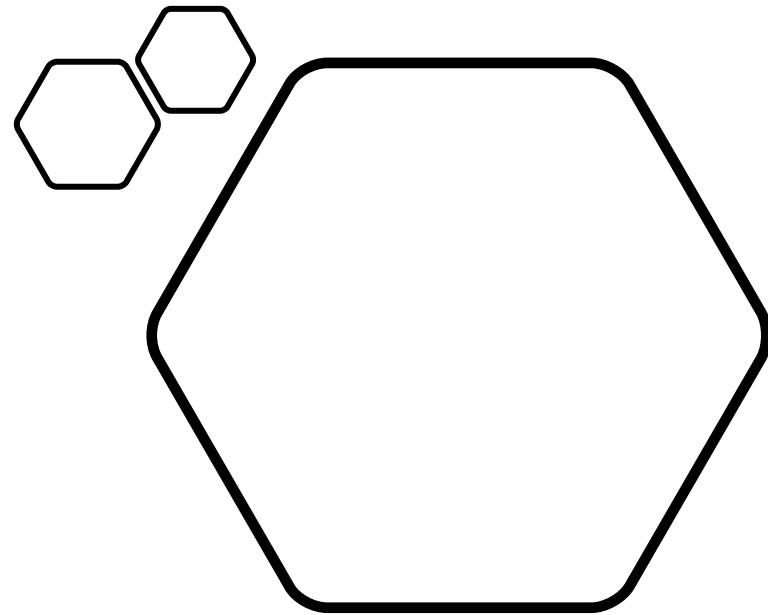
- Analiza cluster

6. Cauzalitate

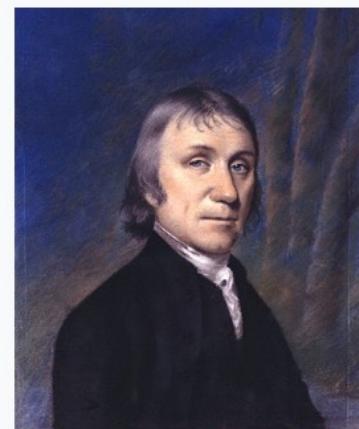
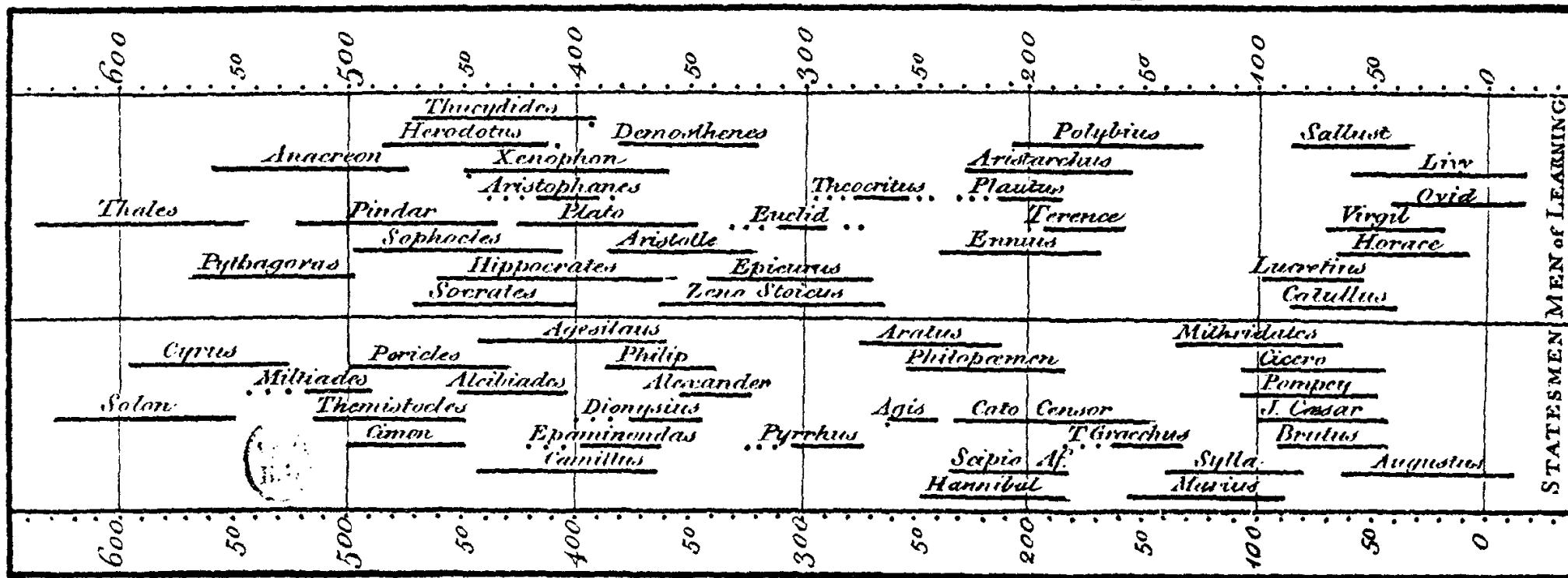
- Analiza de regresie, path

1. Priviri de ansamblu

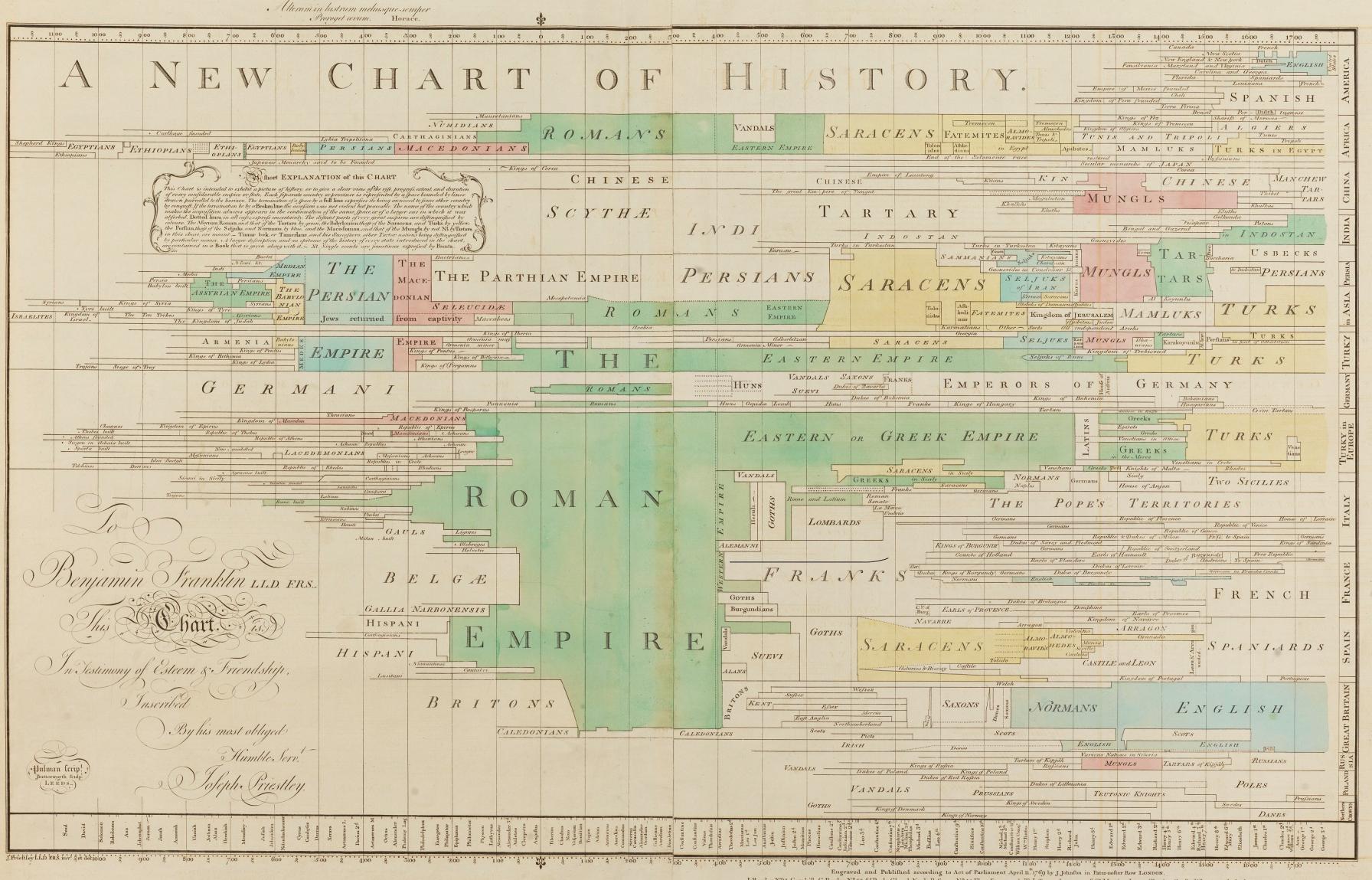
Reducerea dimensionalității



A Specimen of a Chart of Biography.

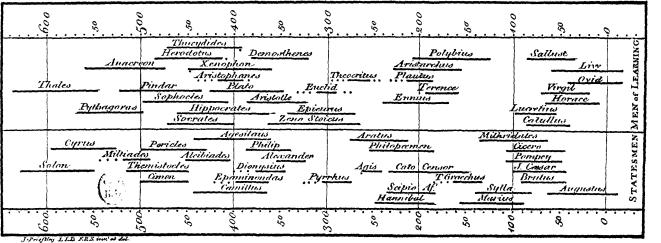
Priestley by Ellen Sharples (1794)^[1]

Born	24 March [O.S. 13 March] 1733 Birstall, Yorkshire, England
Died	6 February 1804 (aged 70) Northumberland, Pennsylvania, U.S.
Occupation	Chemist · natural philosopher · theologian · grammarian · teacher
Known for	<ul style="list-style-type: none"> Discovery of oxygen (O_2) and nine other gases (including carbon monoxide (CO), nitric oxide (NO), nitrous oxide (N_2O), ammonia (NH_3), sulphur dioxide (SO_2), and nitrogen peroxide (N_2O_4)) Discovery of the carbon cycle
Awards	<ul style="list-style-type: none"> Fellow of the Royal Society (1766)^[2]



Sursă

A Specimen of a Chart of Biography.



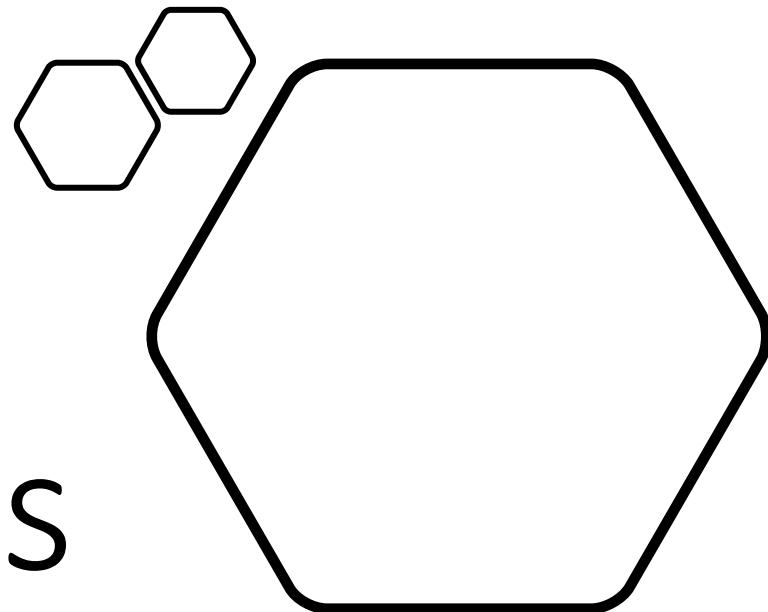
Sursă: [Notable People](#)



2. Deficit și exces

Distribuții:

- Ce este prea mult? Ce este prea puțin?
- Cazuri limită, anomalii



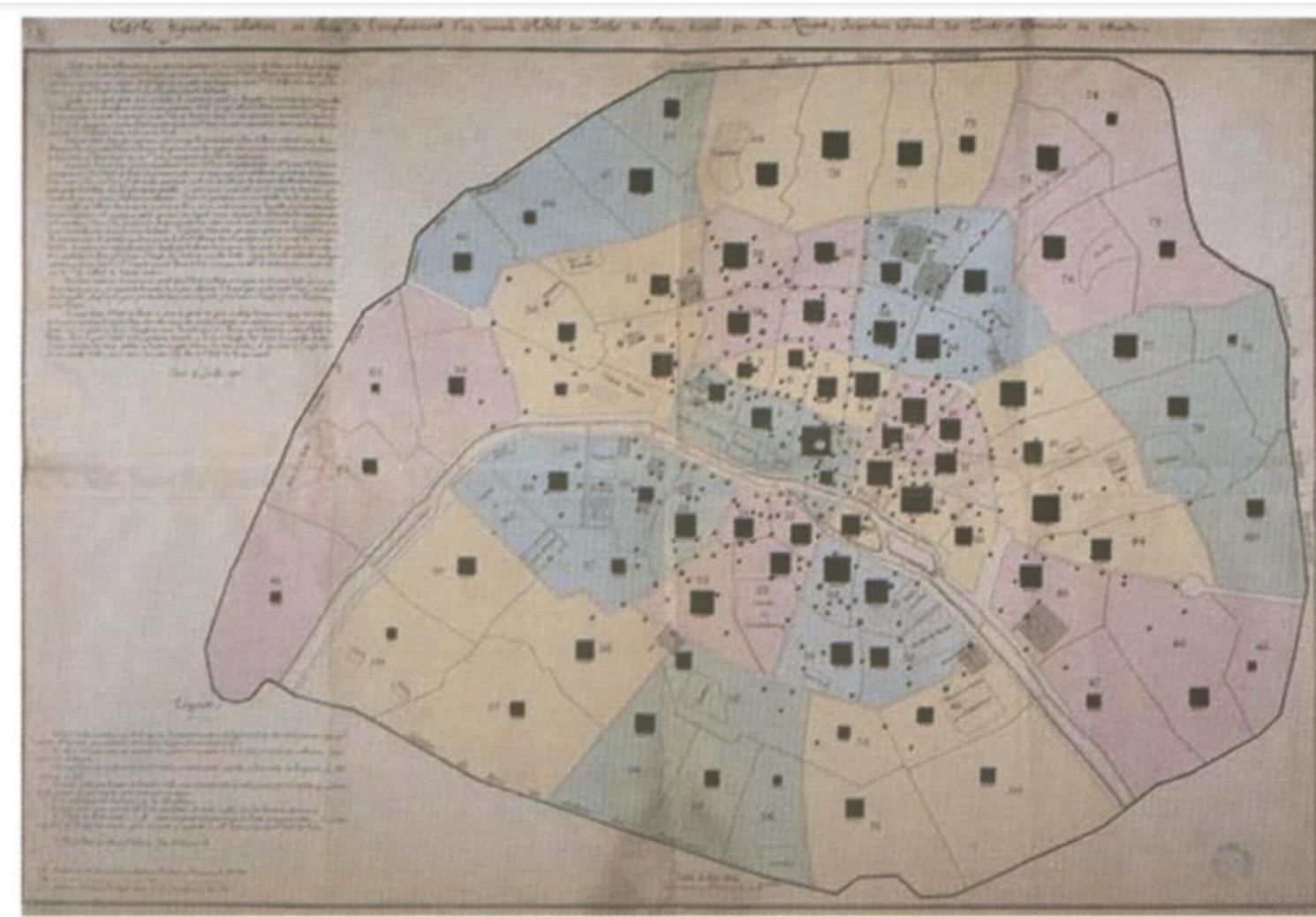


FIGURE 3 *Minard's center-of-gravity map of Paris.*

Note. The visual center of gravity of population density was used to choose location for a new post office. (Col. et cliché ENPC, 10970/C589; used by permission.)

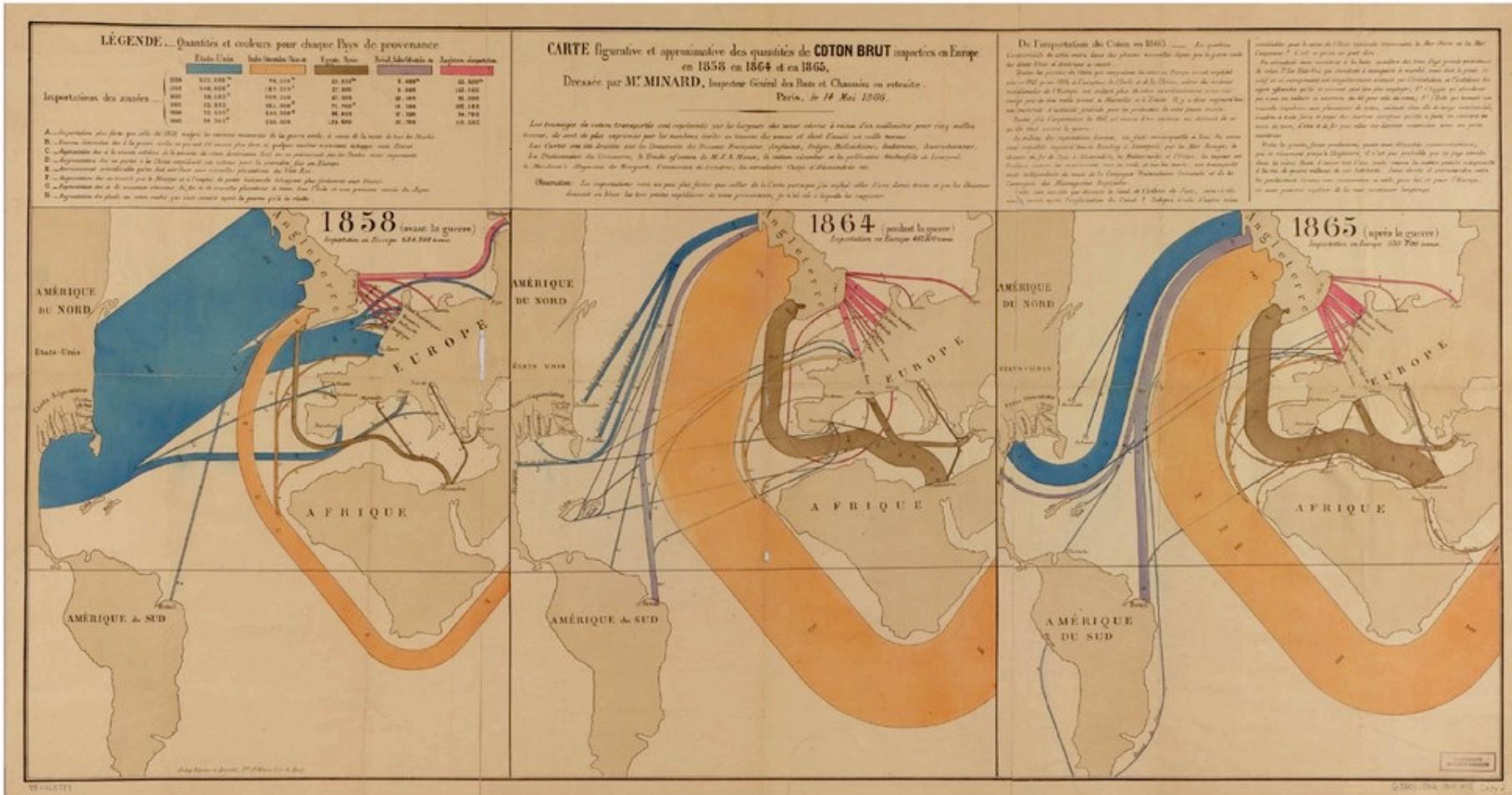


© Science Museum Pictorial / Science & Society Picture Library
10970/C589

Charles Joseph Minard

Born	27 March 1781 Dijon, France
Died	24 October 1870 (aged 89) Bordeaux, France
Nationality	French
Alma mater	École Polytechnique
Known for	<i>Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813</i>
Scientific career	
Fields	Civil engineering and information graphics
	Signature

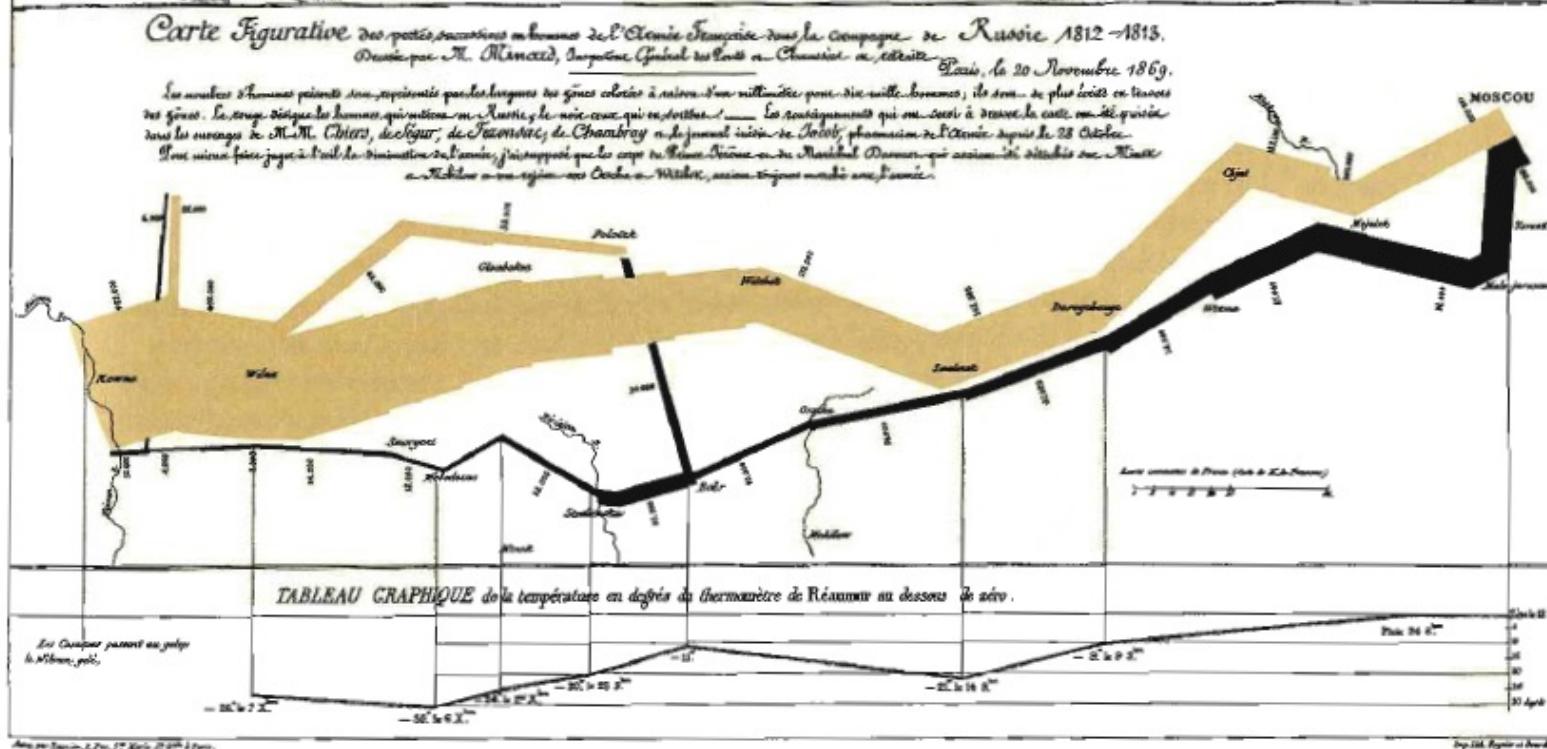
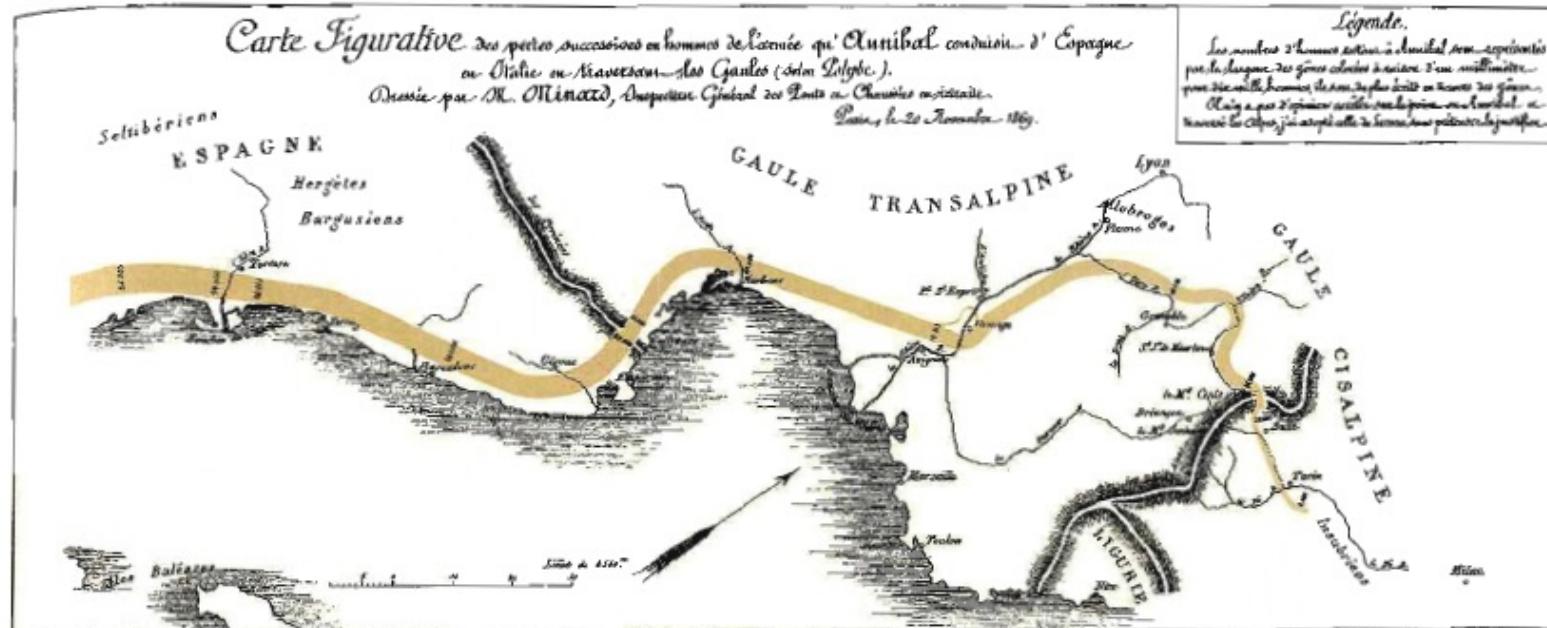
Ch. J. Minard



1869



© Science Museum Pictorial / Science & Society Picture Library

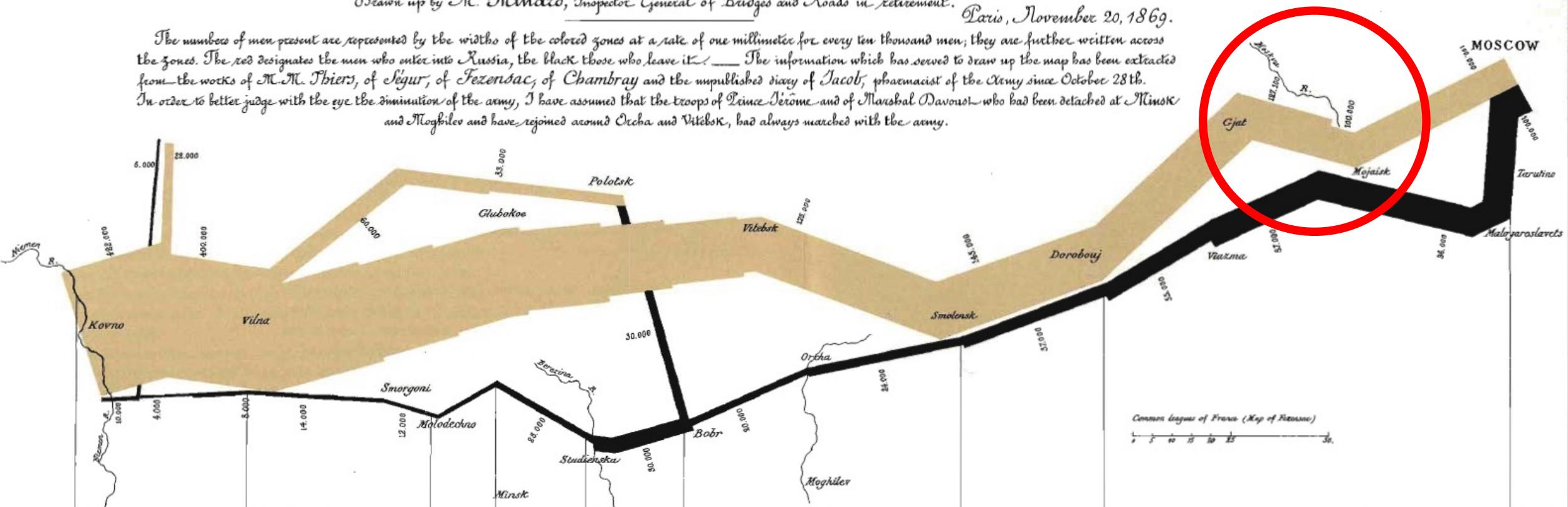


Charles Joseph Minard	
Born	27 March 1781 Dijon, France
Died	24 October 1870 (aged 89) Bordeaux, France
Nationality	French
Alma mater	École Polytechnique
Known for	<i>Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813</i>
Fields	Civil engineering and information graphics
Signature	

Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812 ~1813.

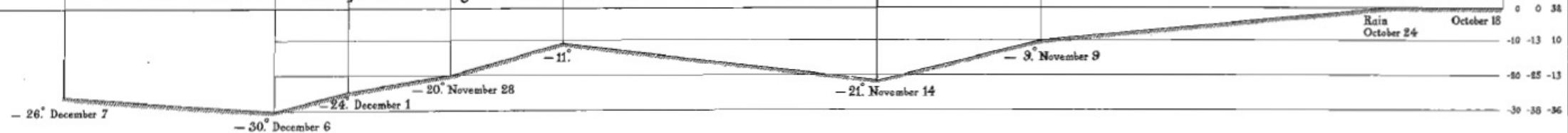
Drawn up by M. Minard, Inspector General of Bridges and Roads in retirement. Paris, November 20, 1869.

The numbers of men present are represented by the widths of the colored zones at a rate of one millimeter for every ten thousand men; they are further written across the zones. The red designates the men who enter into Russia, the black those who leave it. — The information which has served to draw up the map has been extracted from the works of M.M. Thiers, of Cléger, of Fezensac, of Chambray and the unpublished diary of Jacob, pharmacist of the Army since October 28th. In order to better judge with the eye the diminution of the army, I have assumed that the troops of Prince Jérôme and of Marshal Davout, who had been detached at Minsk and Moghilev and have rejoined around Orsha and Vitebsk, had always marched with the army.



GRAPHIC TABLE of the temperature in degrees of the Réaumur thermometer below zero.

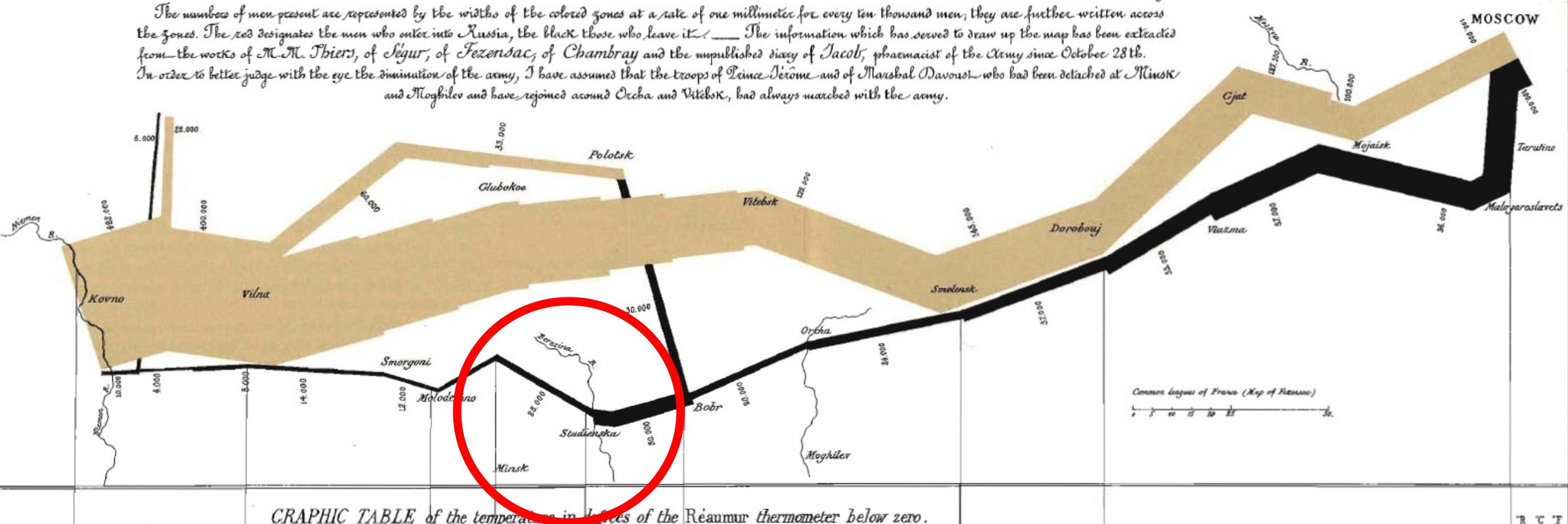
The Cossacks pass the frozen Niemen at a gallop.



Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812 ~ 1813.

Drawn up by M. Minard, Inspector General of Bridges and Roads in retirement. Paris, November 20, 1869.

The numbers of men present are represented by the widths of the colored zones at a rate of one millimeter for every ten thousand men; they are further written across the zones. The red designates the men who enter into Russia, the black those who leave it. — The information which has served to draw up the map has been extracted from the works of M.M. Thiers, of Cléger, of Fezensac, of Chambray and the unpublished diary of Jacob, pharmacist of the Army since October 28th. In order to better judge with the eye the diminution of the army, I have assumed that the troops of Prince Jérôme and of Marshal Davout, who had been detached at Minsk and Moghilev and have rejoined around Orsha and Vitebsk, had always marched with the army.



GRAPHIC TABLE of the temperatures in degrees of the Réaumur thermometer below zero.

The Cossacks pass the frozen Niemen at a gallop.

— 26° December 7 — 24. December 1 — 20° November 28 — 11°

— 30° December 6

— 21° November 14

— 9° November 9

	R	C	F
October 18	0	0	32
Rain October 24	-10	-13	10
	-20	-25	-13
	-30	-38	-36

Imp. Lith. Rognier et Bourdet.

Belligerents	Strength
 French Empire  Confederation of Poland  Kingdom of Italy  Kingdom of Naples  Kingdom of Saxony  Kingdom of Bavaria  Kingdom of Westphalia  Kingdom of Württemberg  Grand Duchy of Hesse  Grand Duchy of Berg  Grand Duchy of Baden  Swiss Confederation  Austrian Empire  Kingdom of Prussia  Kingdom of Spain ^[1]  Kingdom of Denmark	 Russian Empire Supported by:  United Kingdom
612,000 ^[2] - 685,000 ^[3] total:	488,000 - 623,000 total: ^[2]
<ul style="list-style-type: none"> • 449,000 First wave^[4] • 134,000 Borodino^[5] • 108,000 Maloyaroslavets^[6] • 33,000 Berezina^[7] 	<ul style="list-style-type: none"> • 98,000 First wave^[4] • 118,000 Borodino^[8] • 129,000 Maloyaroslavets^[9] • 126,000 Berezina^[7]
Casualties and losses	Total military and civilian deaths:
400,000–484,000	350,000–410,000
<ul style="list-style-type: none"> • 200,000^{[10][11]}–284,000 dead^{[12][13]} <ul style="list-style-type: none"> • 100,000 killed in combat^{[11][10]} • 50,000 died in captivity^{[14][12][2]} • 50,000 wounded^[13] • 50,000 deserted^{[13][12]} • 50,000 prisoners^{[2][12]} 	<ul style="list-style-type: none"> • 210,000 dead^[15] • 150,000 wounded^[16] • 50,000 deserted^[16]
c. 1,000,000^[17]	

[Sursa](#)

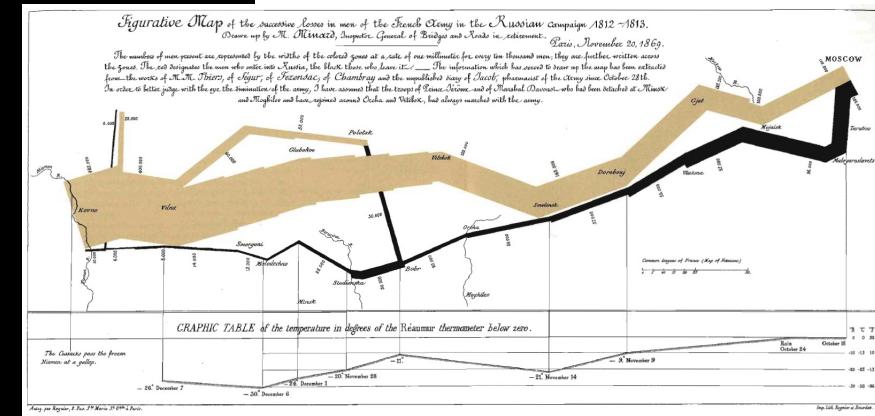
Sursă: [The Fallen of WW2](#)
(video)

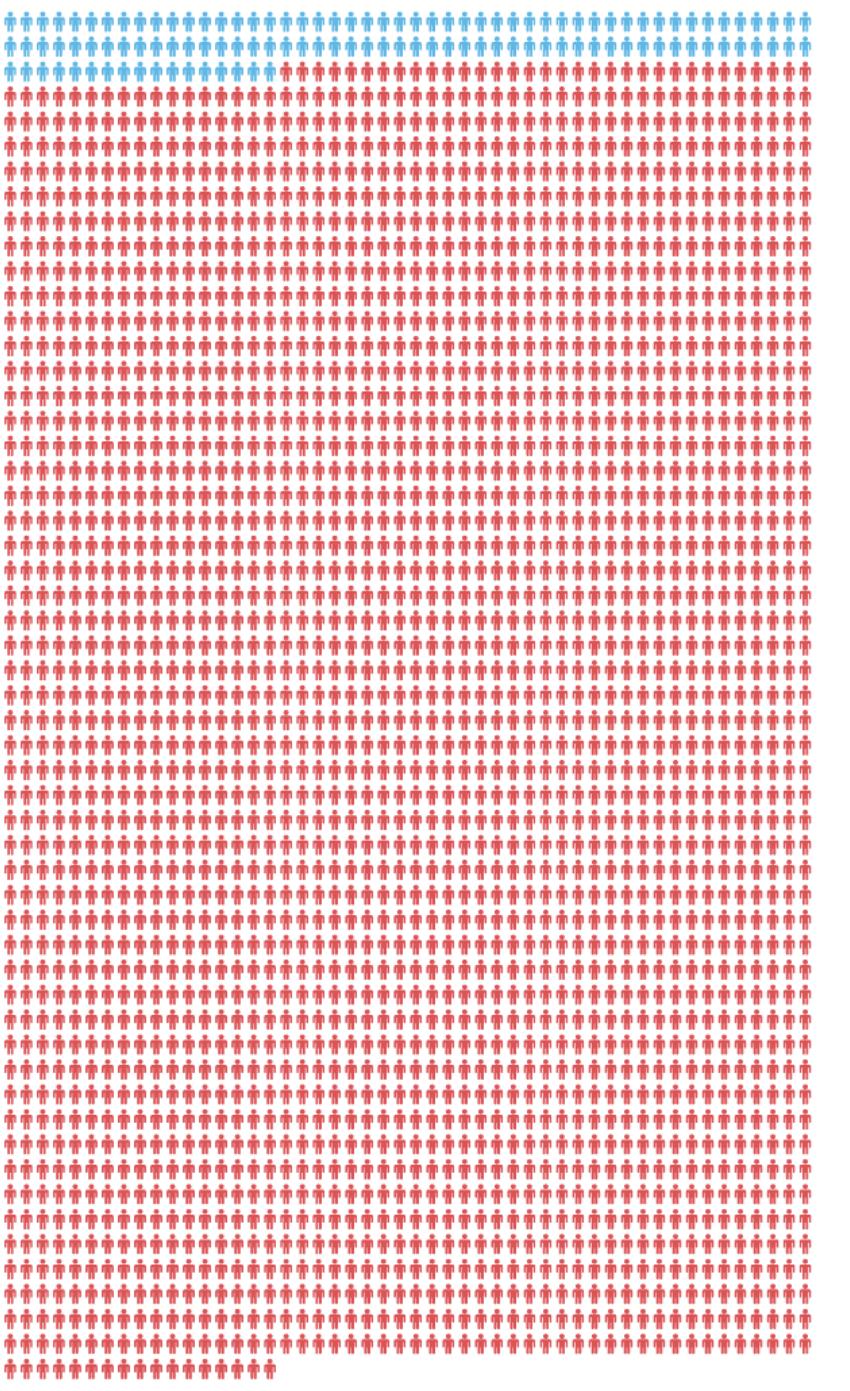
⚠ Not secure | fallen.io/ww2/



The Fallen of World War II

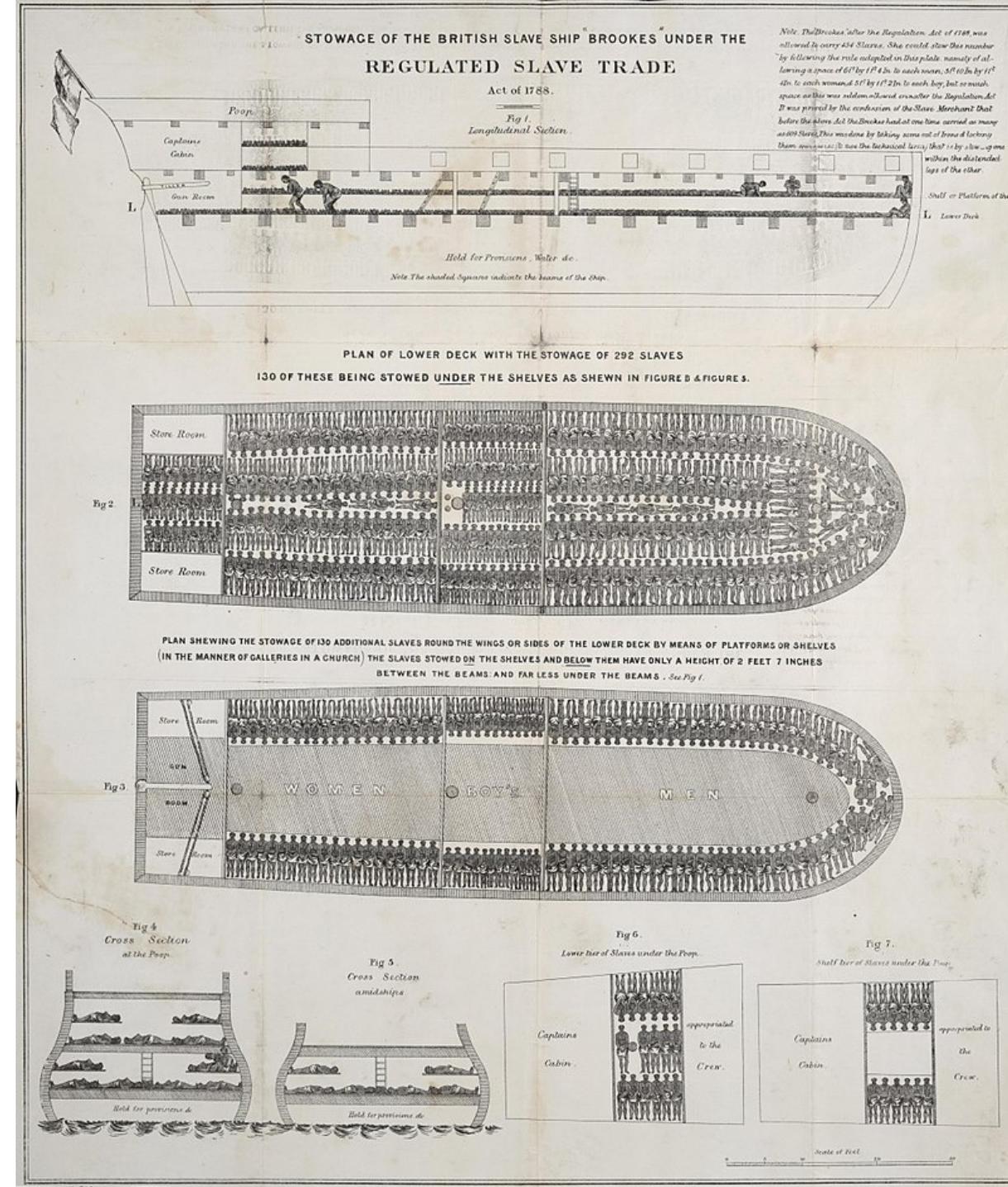
Neil Halloran





People killed in terrorist attacks in 2017: Europe (Blue) vs Africa (Red)

[Sursa](#)





SOURCE: Fatal Encounters



- GUNSHOT
- VEHICLE
- TASERED
- MEDICAL EMERGENCY
- BEATEN
- ASPHYXIATION
- OTHERS

© OpenStreetMap contributors © CARTO, © CARTO



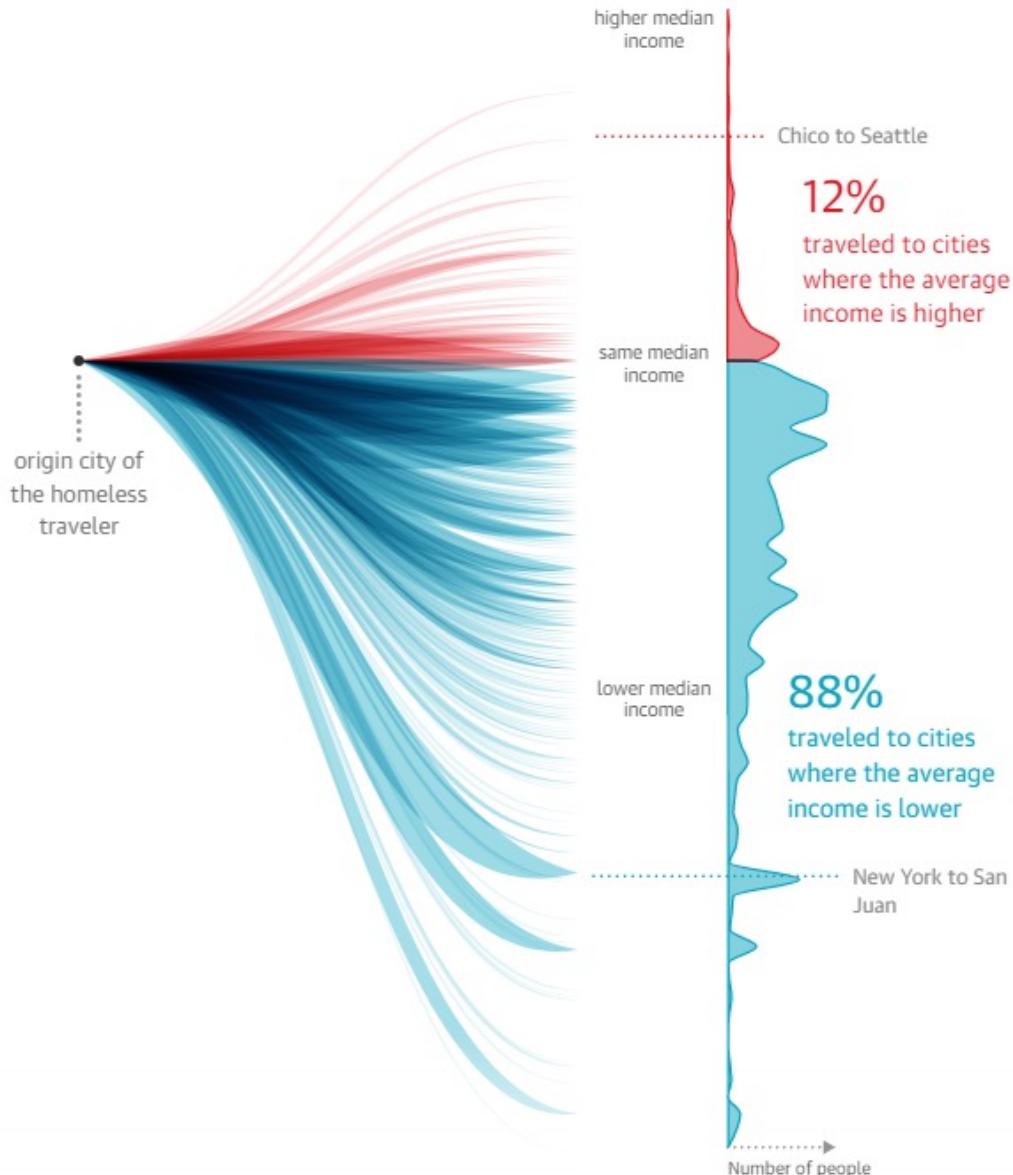
ANANDKATAKAM



MAPS



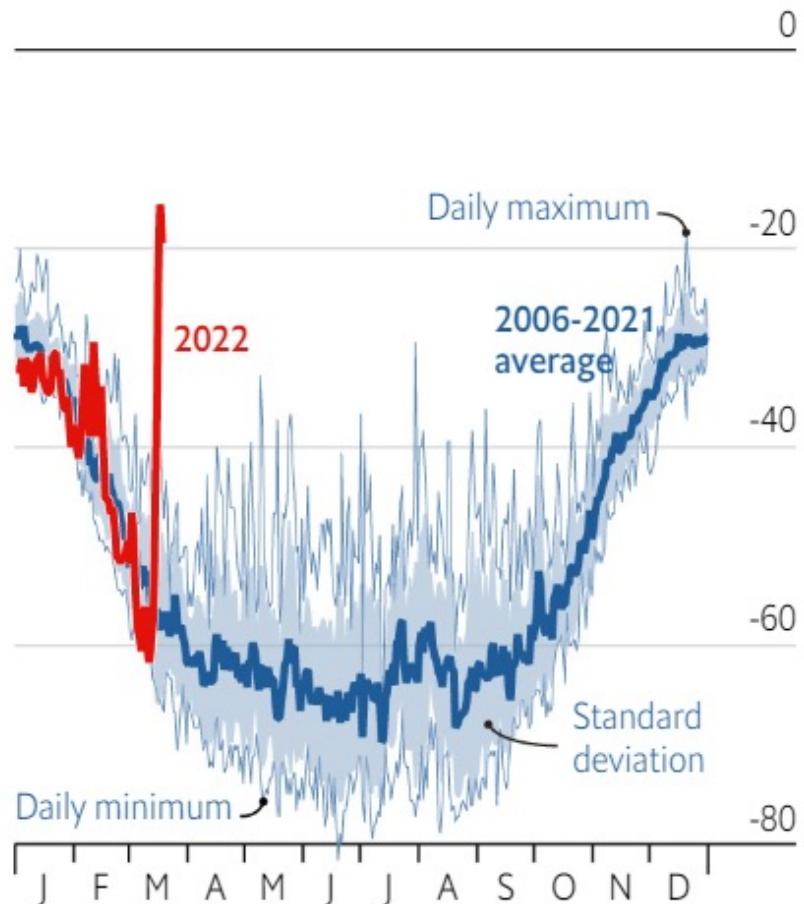
Most ticket recipients are relocated to places with a lower median income



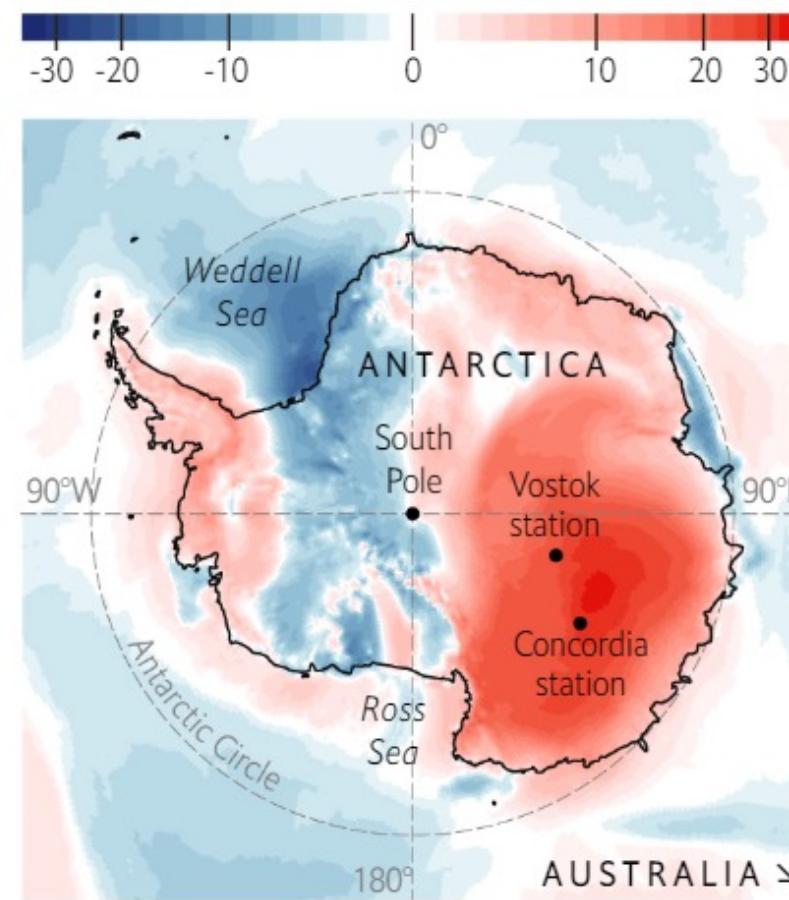
Sursa

Baby, it's (a lot less) cold outside

Antarctica, daily average temperature,
Concordia station, °C



Temperature anomalies, March 19th 2022
compared to 1979-2000 baseline, °C



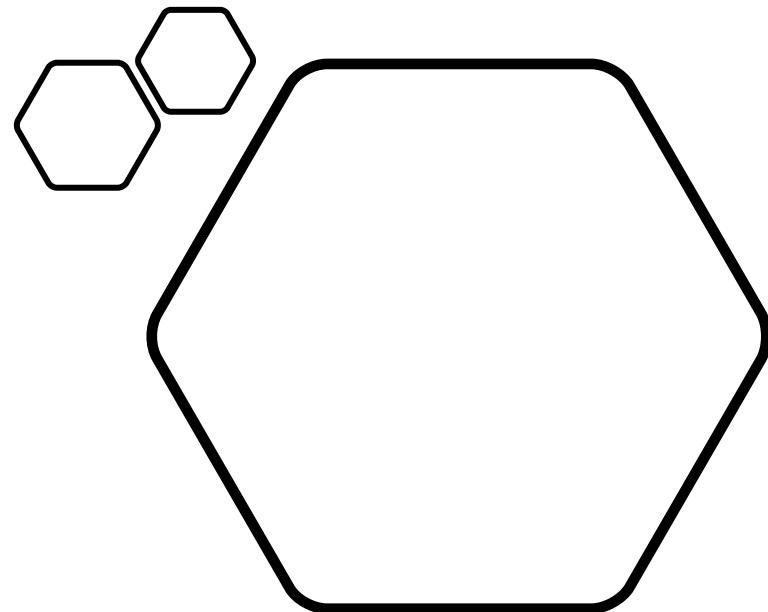
Sources: ClimateReanalyzer.org; National Centres for Environmental Information;
National Oceanic and Atmospheric Administration

Source:
The Economist

3. Tendințe

Serii de timp:

- Banii și viața
- Evoluția civilizațiilor
- Evoluția tehnologiei



Like an ice age, in reverse; CO₂ levels are far higher than previous interglacial periods, and have risen remarkably fast

Atmospheric CO₂ levels, parts per million

Source — Vostok ice core — Law Dome ice core — Mauna Loa Observatory

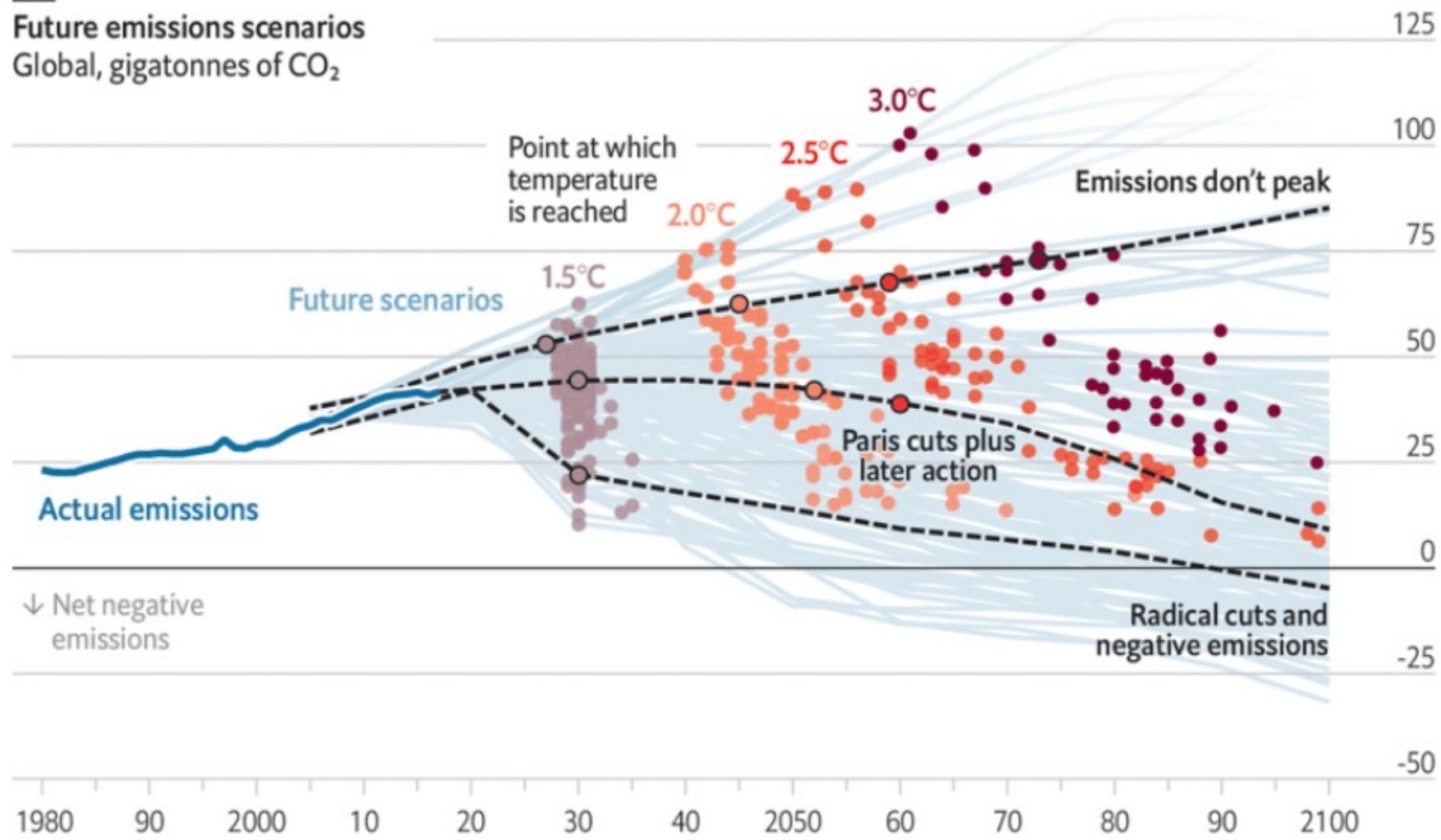


Source:
The Economist

Scenarios for future CO₂ emissions, with three representative pathways picked out

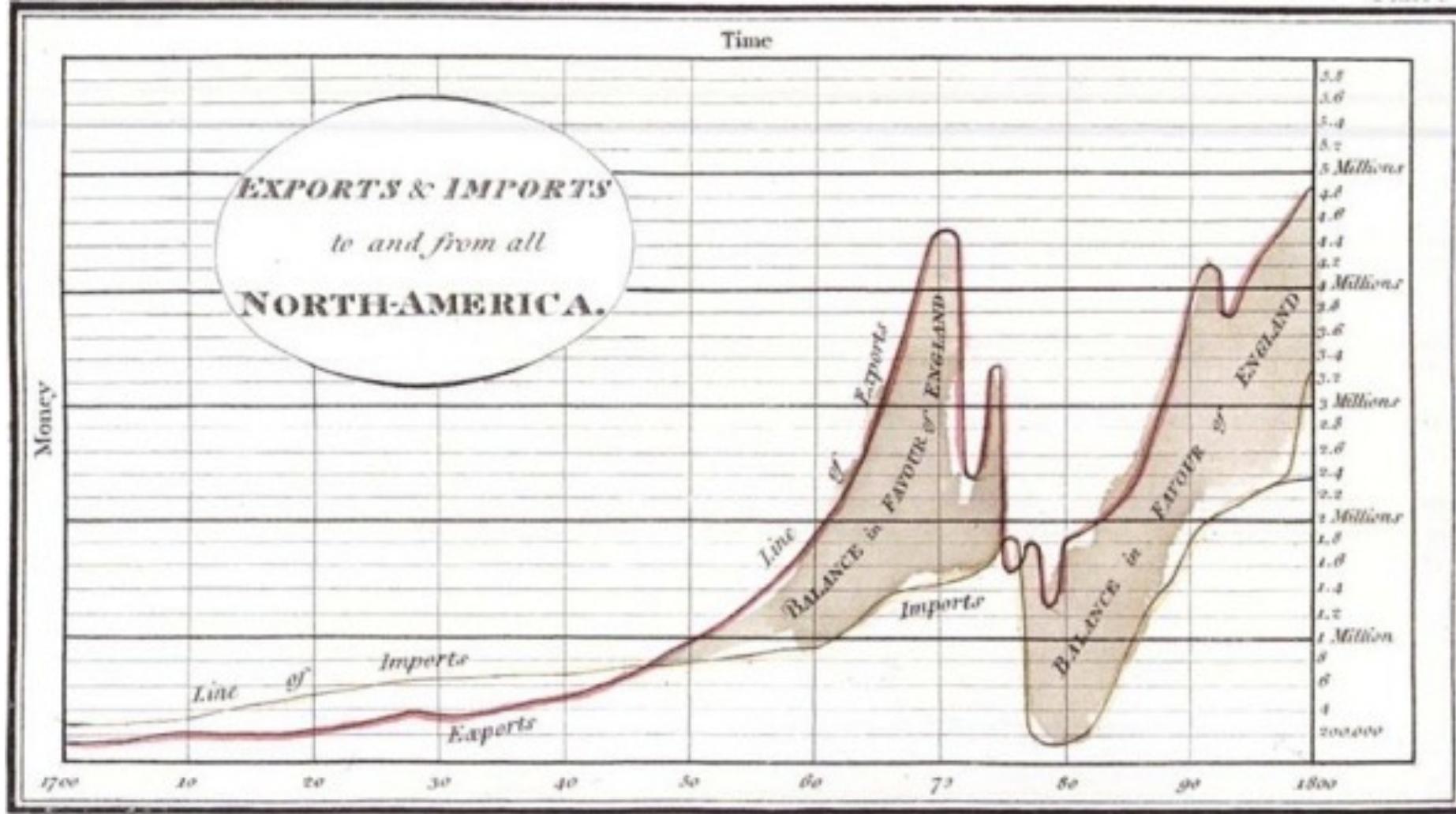
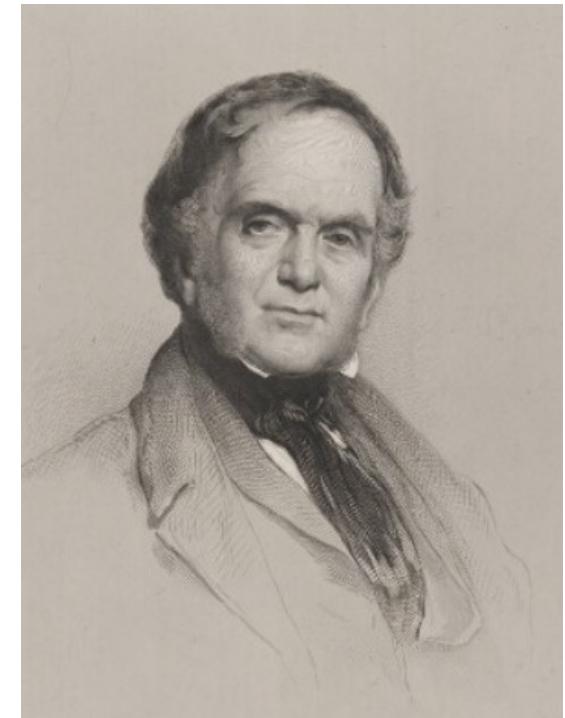
Future emissions scenarios

Global, gigatonnes of CO₂



The Economist

Source:
The Economist



William Playfair

Born	September 22, 1759 Benvie, Forfarshire, Scotland
Died	11 February 1823 (aged 63) London, England
Nationality	Scottish
Known for	inventor of statistical graphs, writer on political economy, and secret agent for Great Britain
Family	John Playfair (brother) James Playfair (brother) William Henry Playfair (nephew)

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



The Bottom line is divided into Years, the Right hand line into £10,000 each.

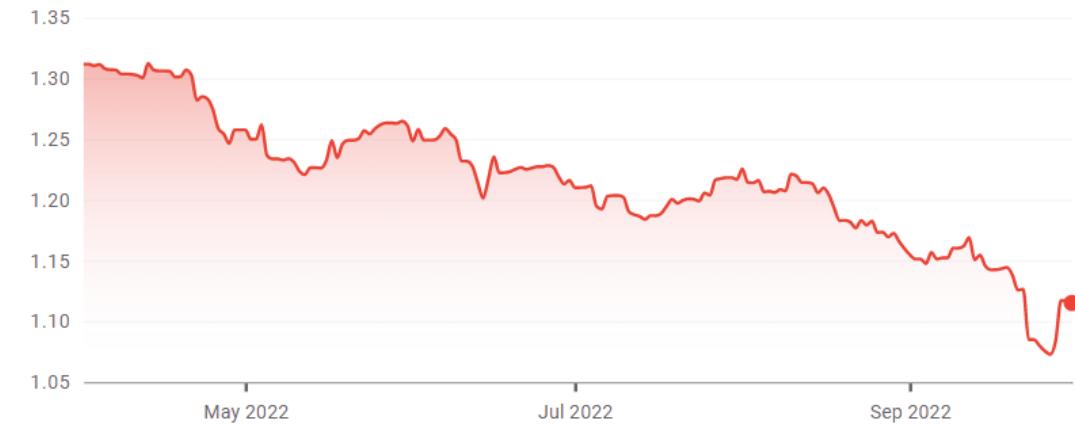
Published as the Act directs, 1st May 1786. by W^m Playfair
Neale sculpt^r 352, Strand, London.

Pound sterling to United States Dollar

1.1150 ↓14.98% -0.1965 6M

Oct 1, 4:46:25 PM UTC · Disclaimer

1D 5D 1M **6M** YTD 1Y 5Y MAX

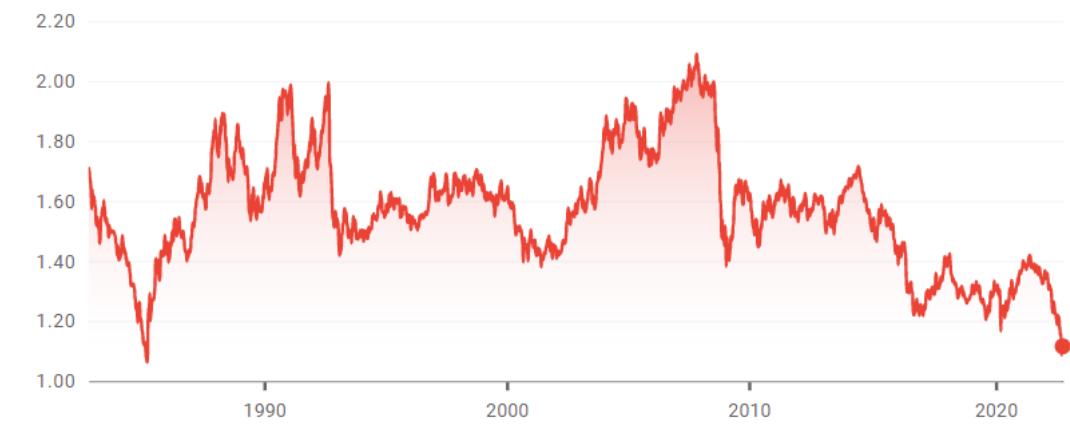


Pound sterling to United States Dollar

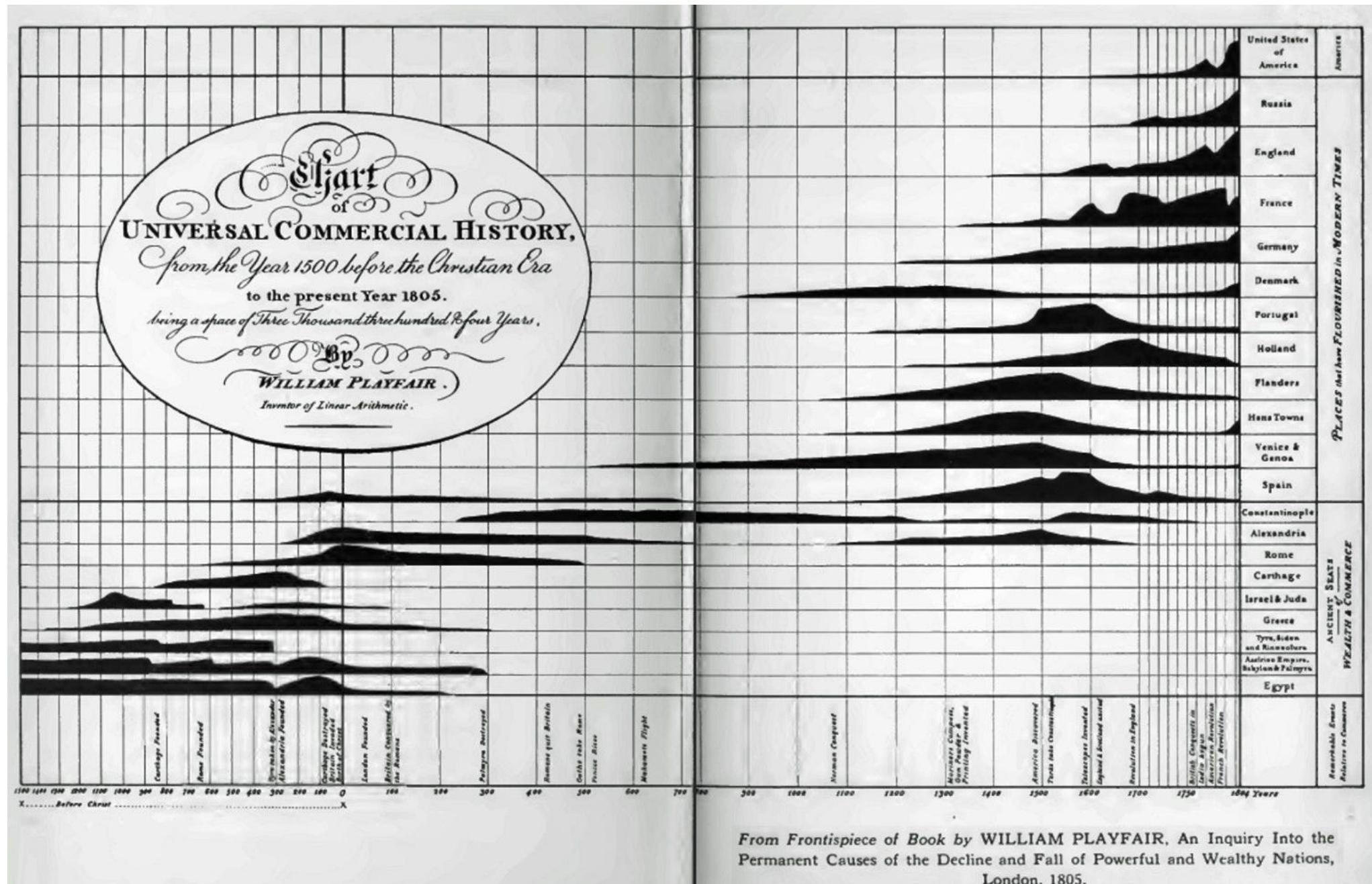
1.1150 ↓34.89% -0.5975 MAX

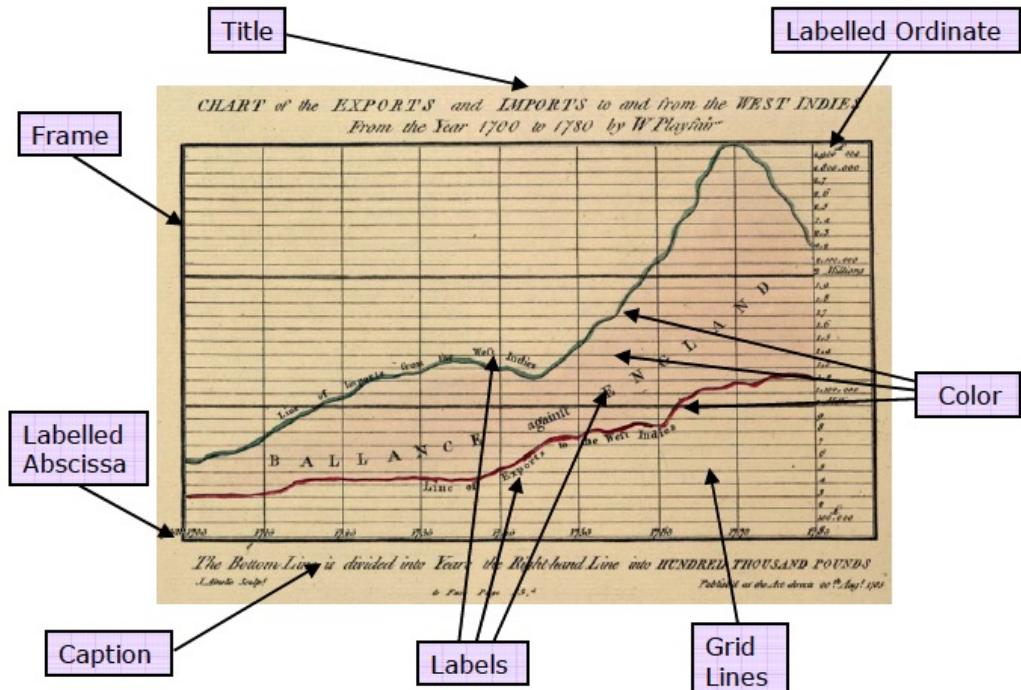
Oct 1, 4:47:25 PM UTC · Disclaimer

1D 5D 1M 6M **YTD** 1Y 5Y **MAX**

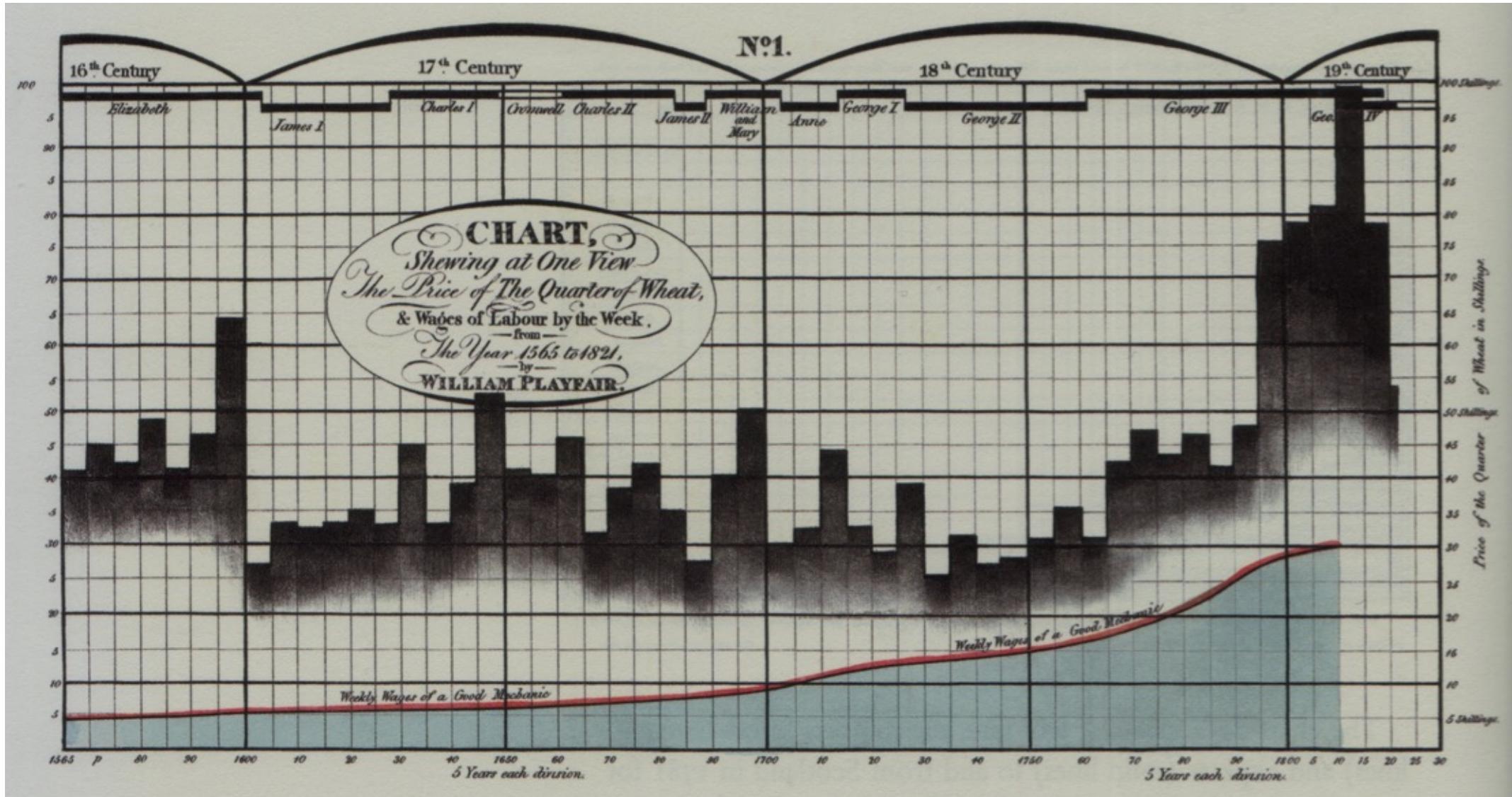


[Sursă](#)

[Sursă](#)

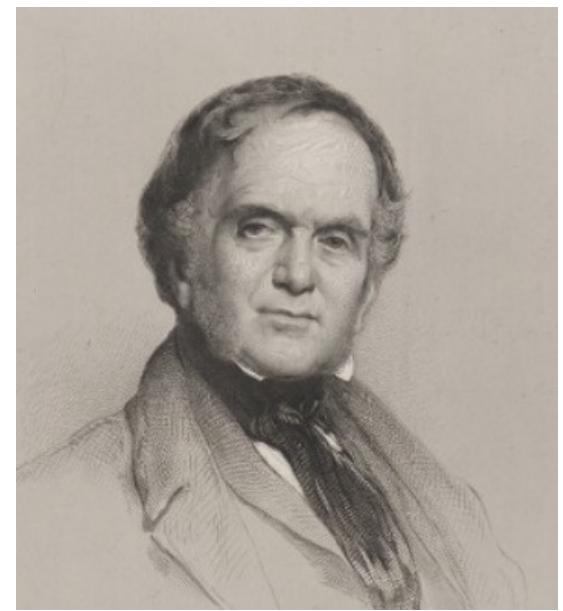


Playfair's charts are almost always constructed so that comparisons in different domains (lines, colors, labels, etc) do not exceed attentional and working memory capacity. Time series in the line graphs never exceed three or four in number; no more than three or four colors are used in a chart; labels are positioned next to the lines themselves rather than in legend boxes distant from the time series; similarly areas are labeled directly rather than being referenced in legends in some other region. Fig. 13 highlights the typical basic elements of a Playfair chart.



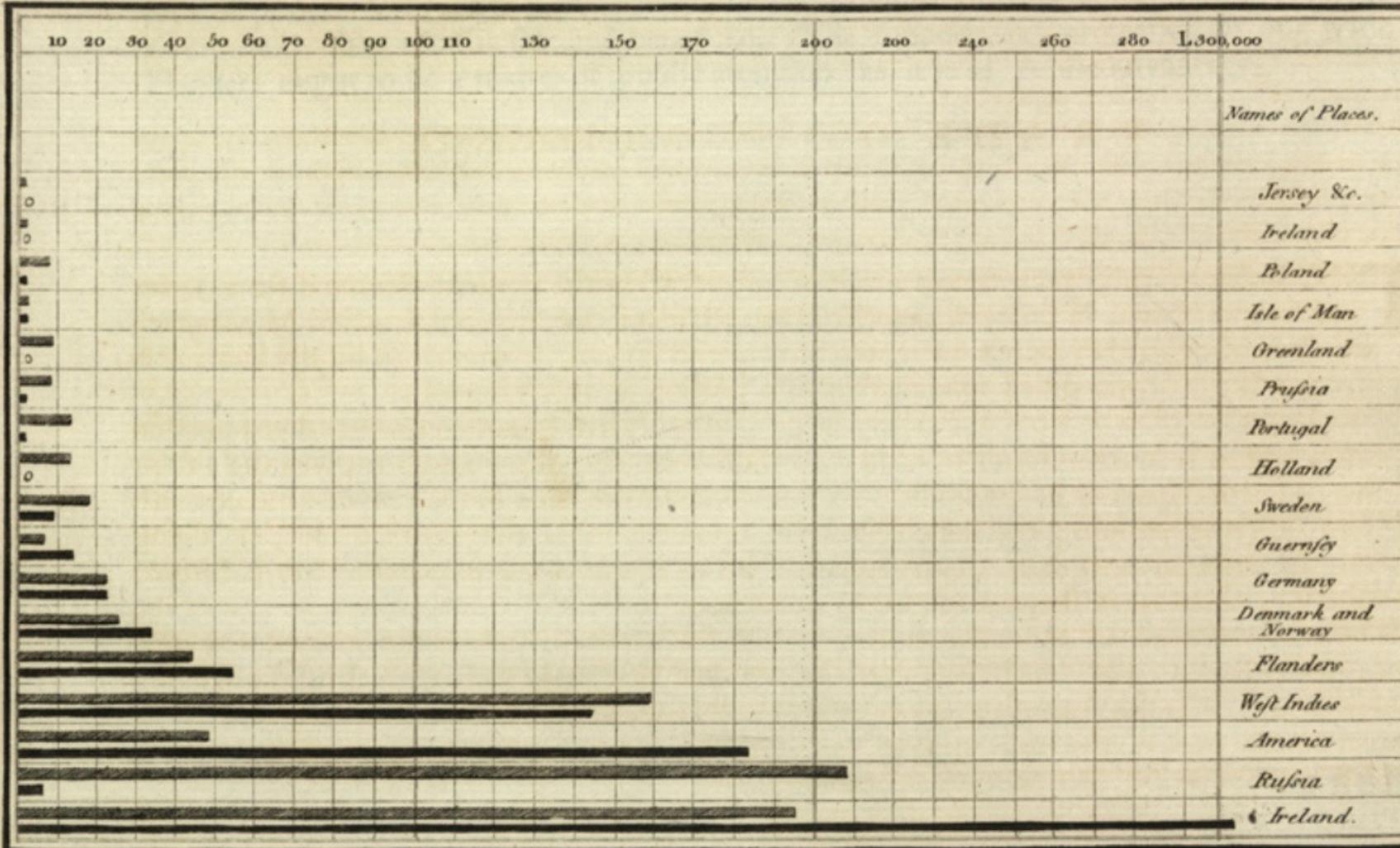
Discuție în
[The Economist](#)

Sursă



Pla

Exports and Imports of SCOTLAND to and from different parts for one Year from Christmas 1780 to Christmas 1781.

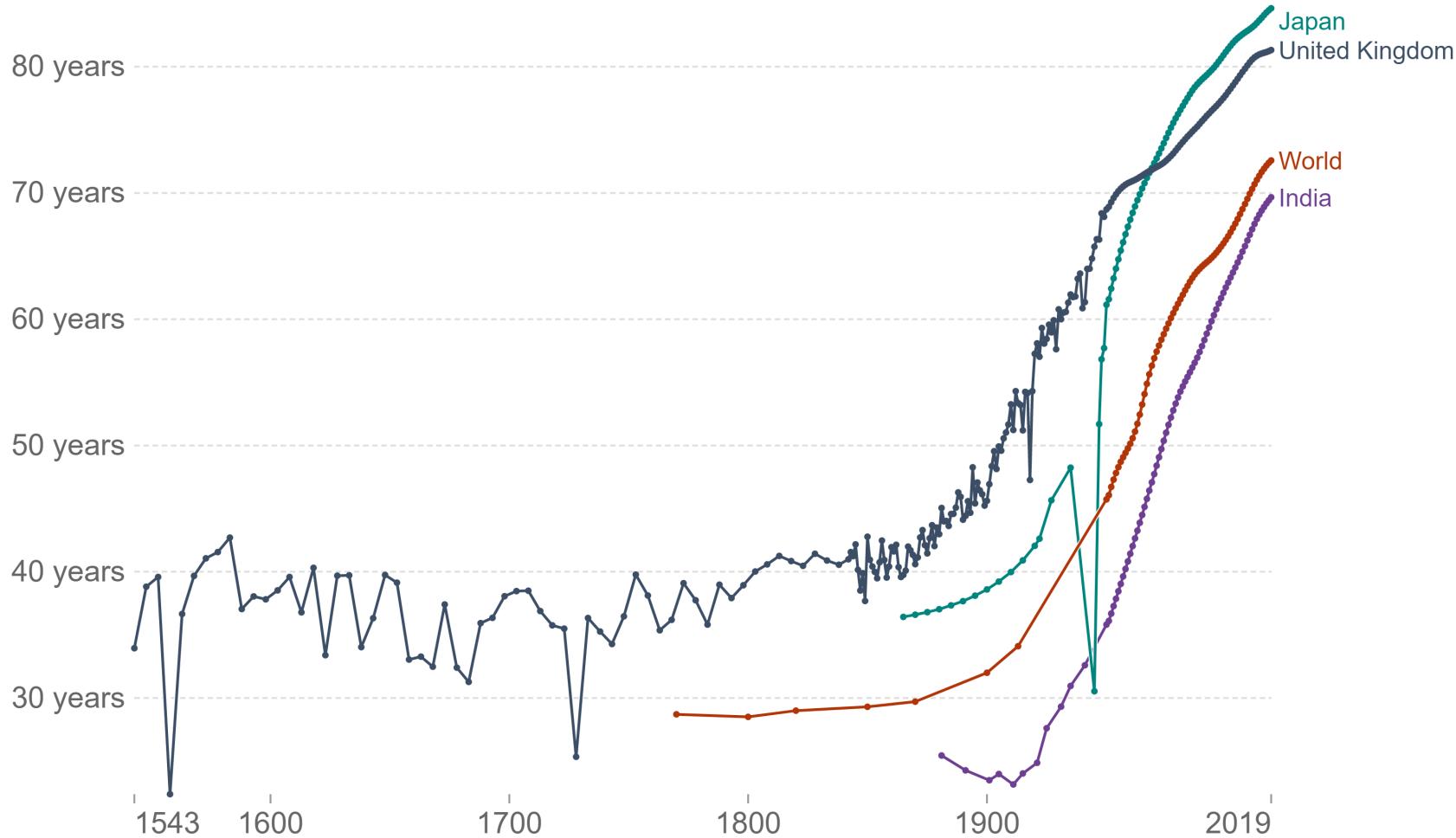


The Upright divisions are Ten Thousand Pounds each. The Black Lines are Exports the Ribbed lines Imports.

Sursă

Life expectancy, 1543 to 2019

Our World
in Data

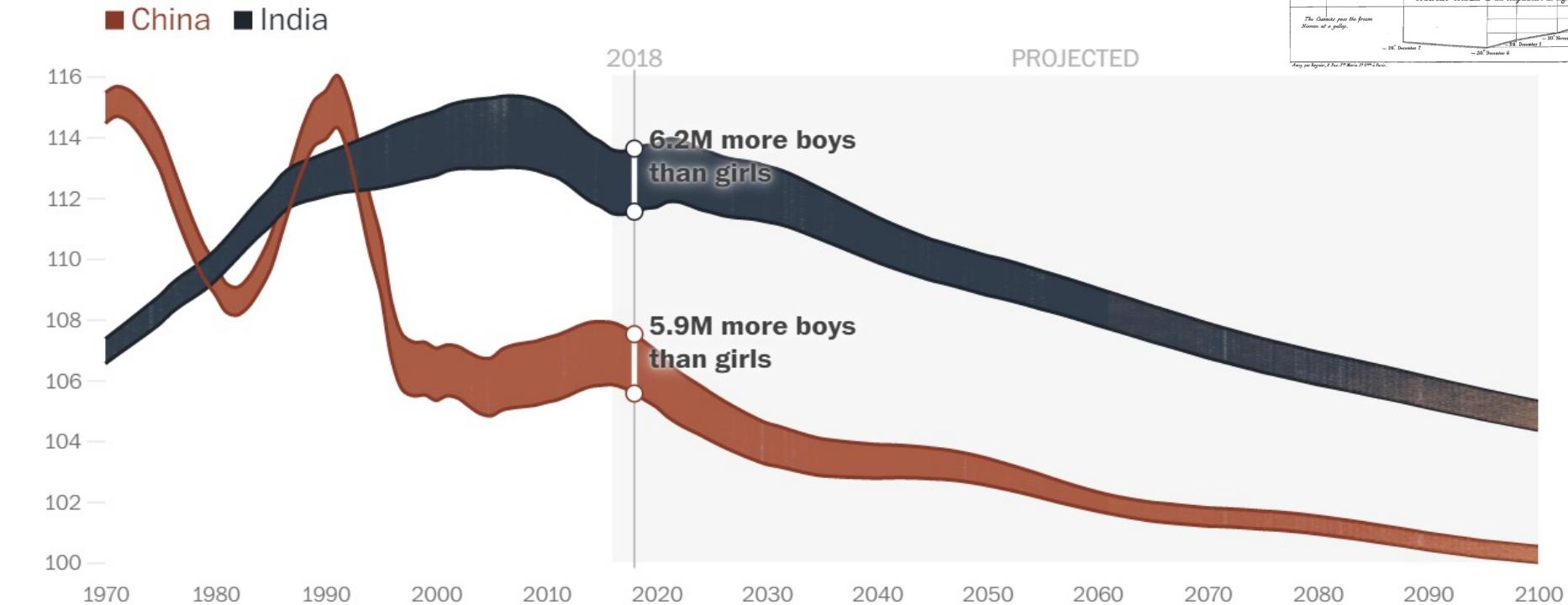
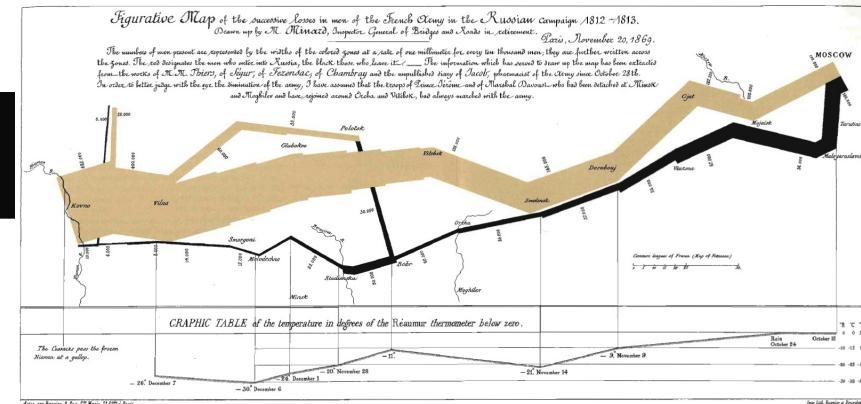


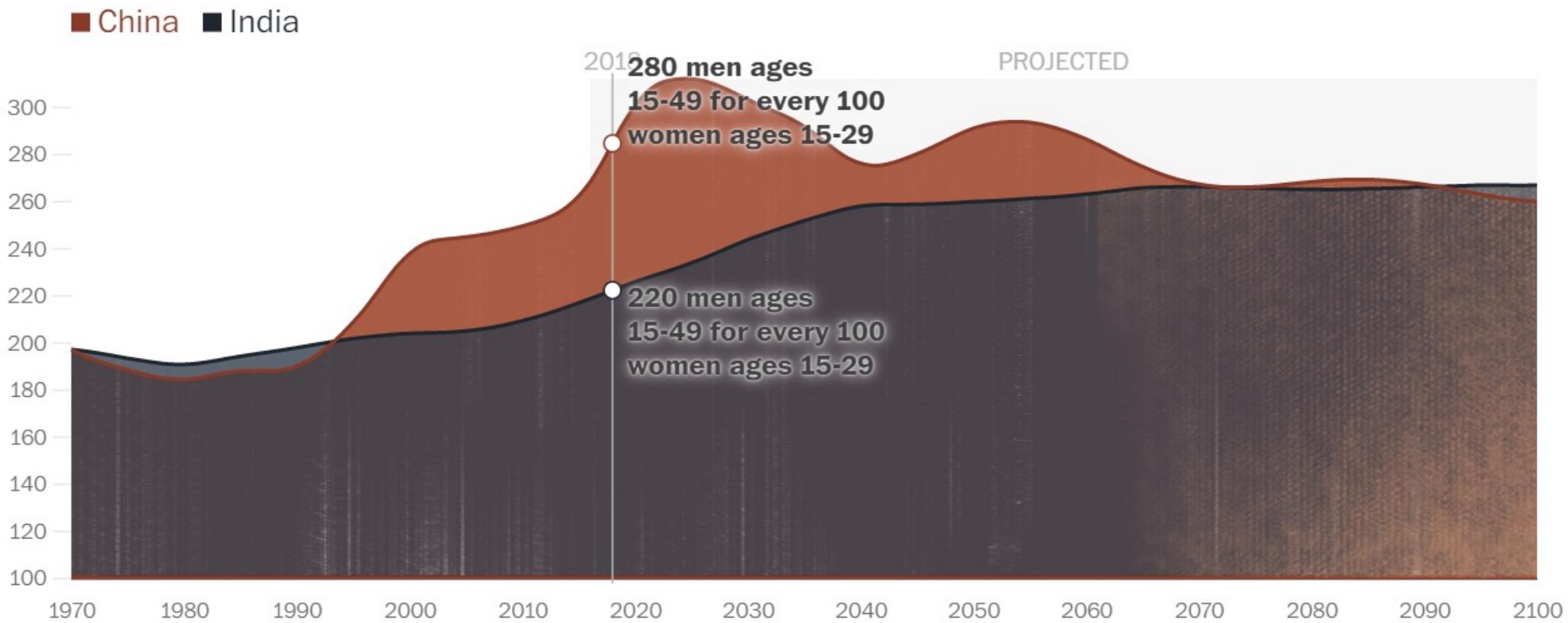
[Sursă](#)

Source: Riley (2005), Clio Infra (2015), and UN Population Division (2019)

OurWorldInData.org/life-expectancy • CC BY

Note: Shown is period life expectancy at birth, the average number of years a newborn would live if the pattern of mortality in the given year were to stay the same throughout its life.

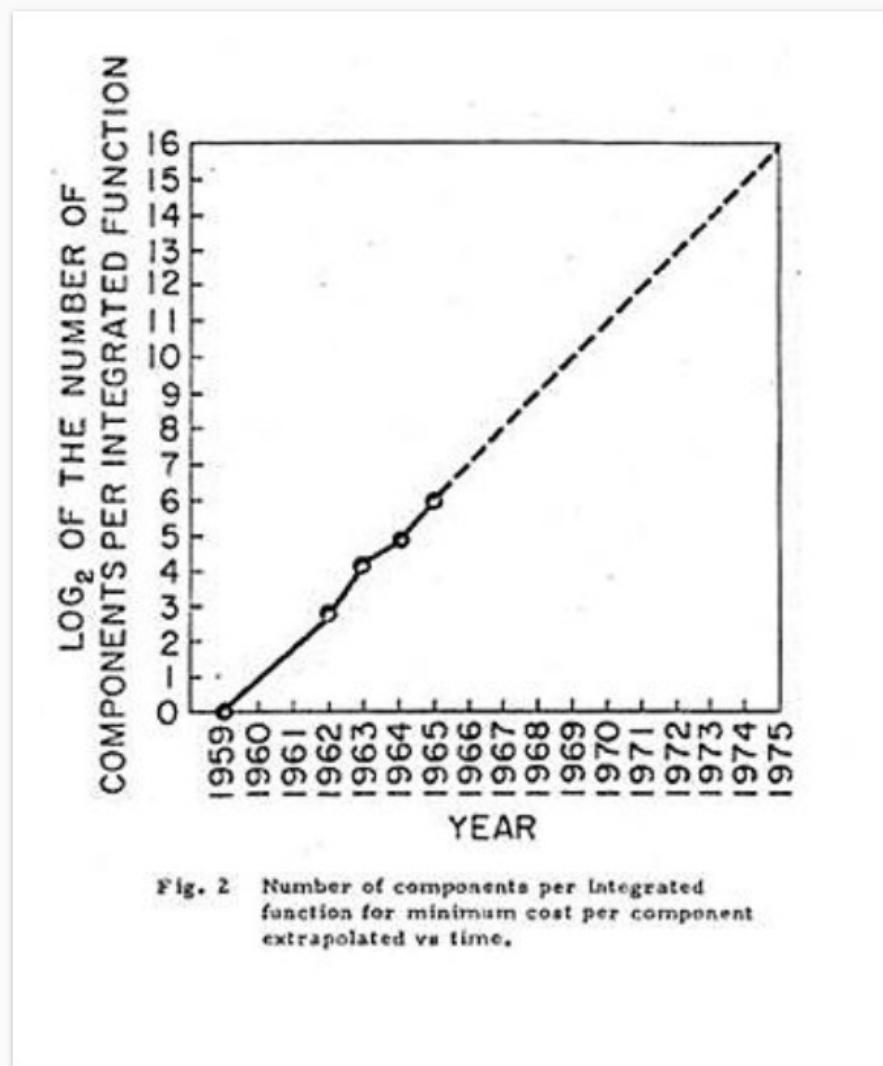




Source: United Nations World Population Prospects

[Sursa](#)

Moore's original graph from 1965: 'The Number of Components per Integrated Function'³

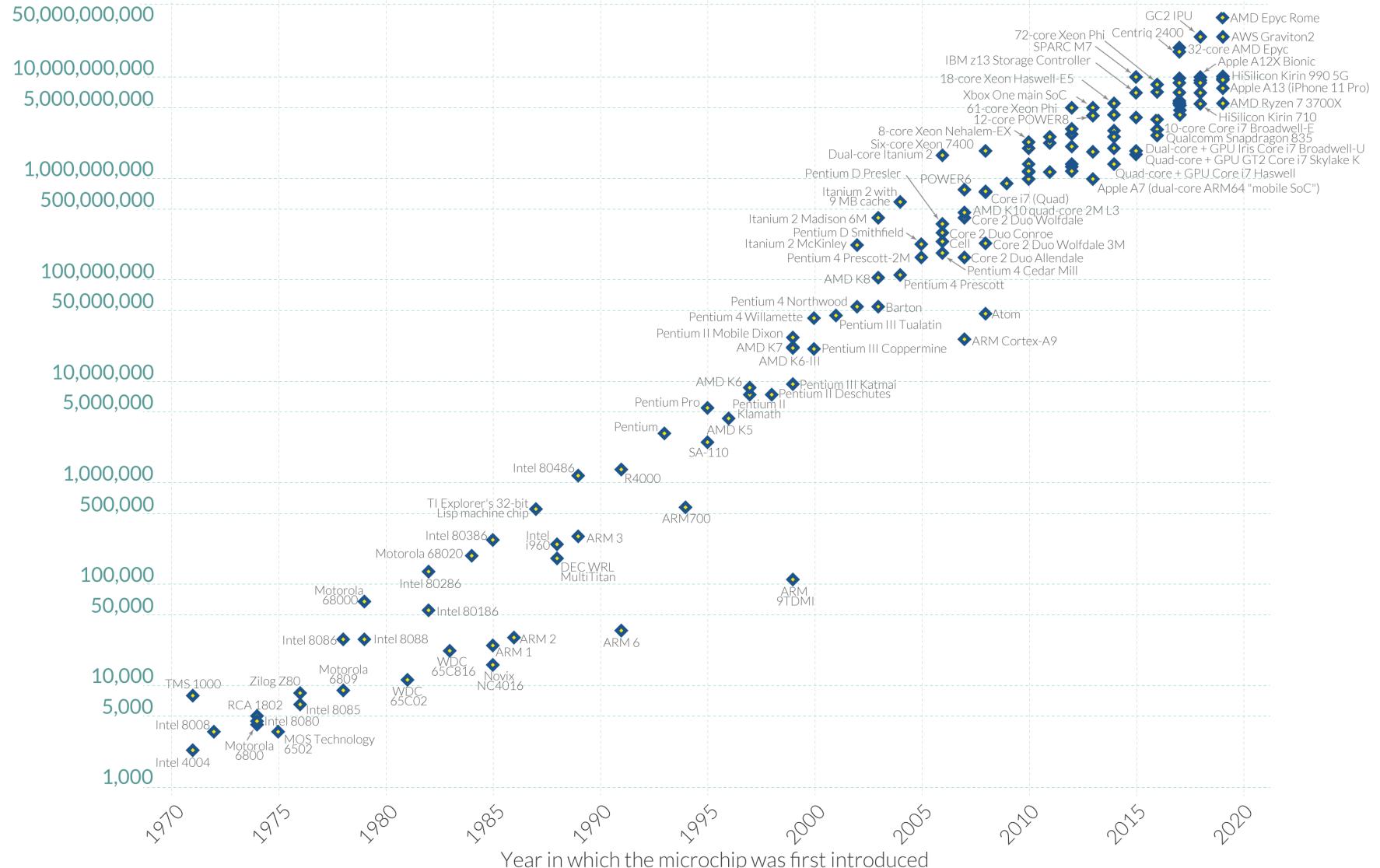


Sursa

Moore's Law: The number of transistors on microchips has doubled every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

Transistor count



Sursa

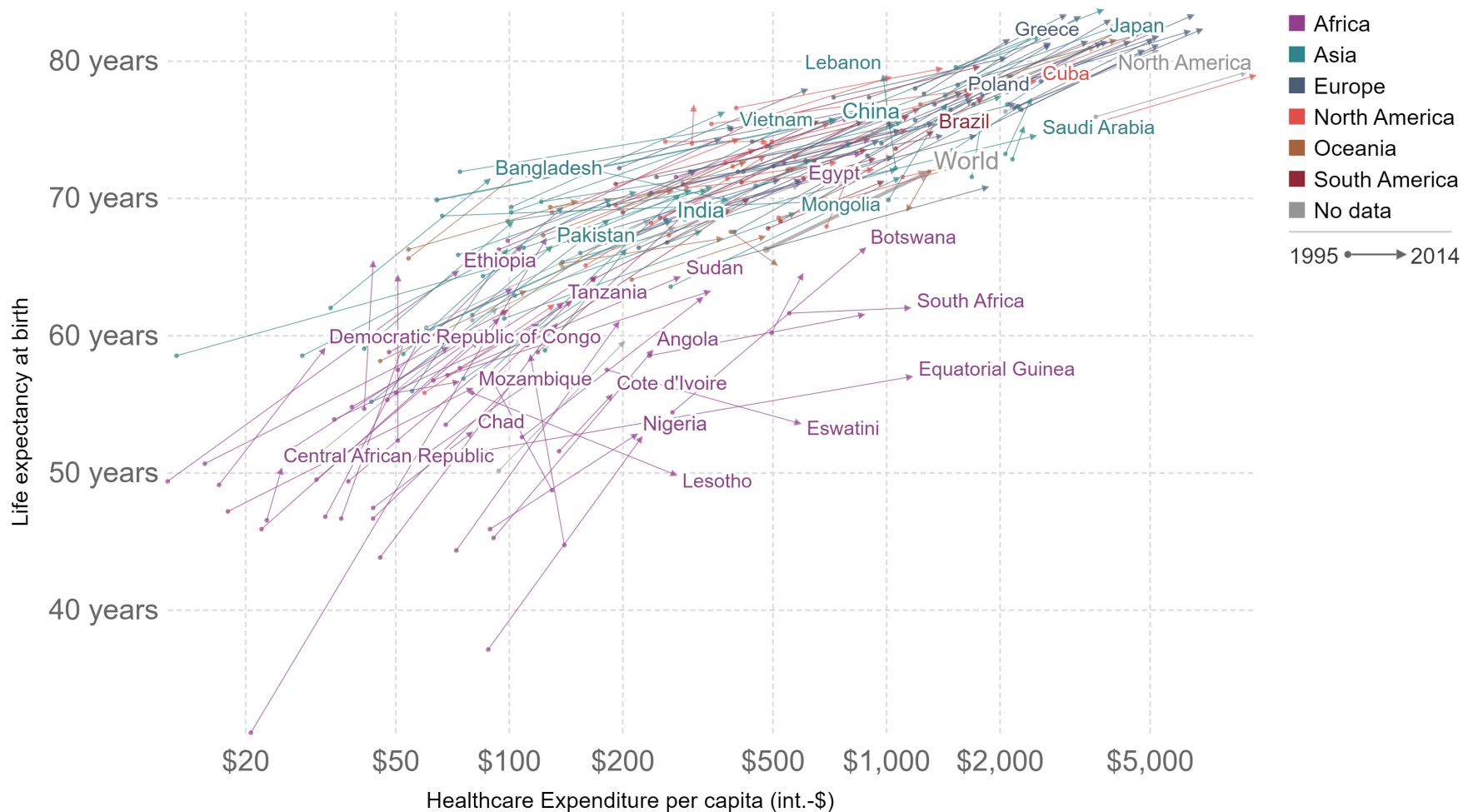
Data source: Wikipedia ([wikipedia.org/w/index.php?title=Transistor_count&oldid=921155120](https://en.wikipedia.org/w/index.php?title=Transistor_count&oldid=921155120))

OurWorldinData.org – Research and data to make progress against the world's largest problems.

Licensed under CC-BY by the authors Hannah Ritchie and Max Roser.

Life expectancy vs. healthcare expenditure, 1995 to 2014

Total healthcare expenditure per capita is adjusted for price differences between countries and for inflation and measured in international-\$.



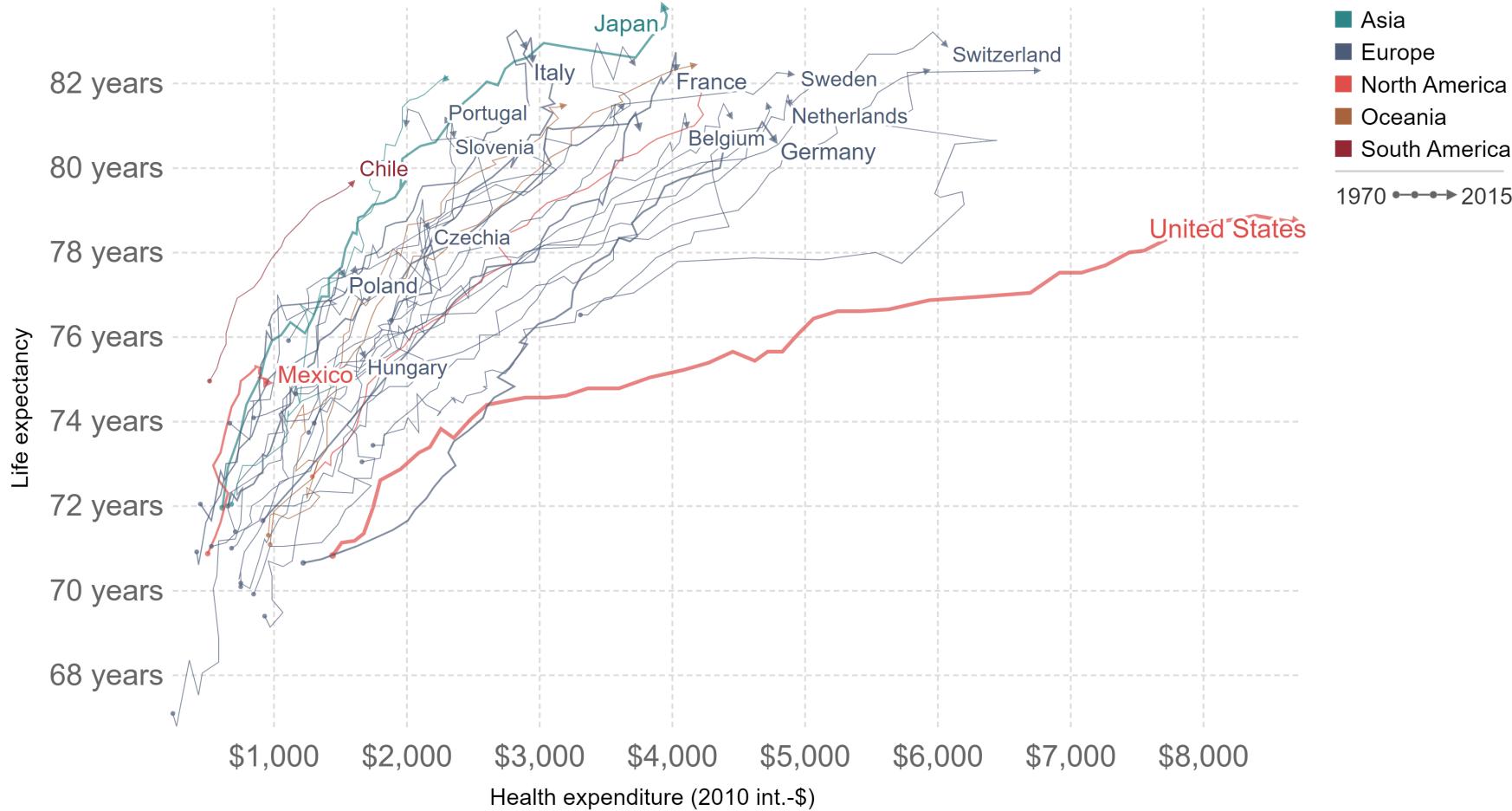
[Sursă](#)

Source: World Bank

CC BY

Life expectancy vs. health expenditure, 1970 to 2015

Health financing is reported as the annual per capita health expenditure and is adjusted for inflation and price level differences between countries (measured in 2010 international dollars).

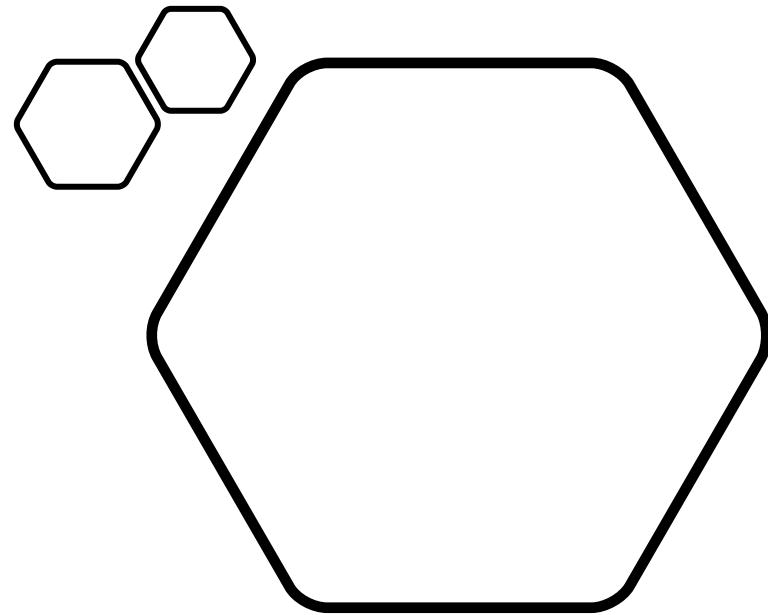


[Sursă](#)

Source: Data compiled from multiple sources by World Bank, Health Expenditure and Financing - OECDstat (2017)
OurWorldInData.org/the-link-between-life-expectancy-and-health-spending-us-focus • CC BY

4. Inegalități

Asocieri, corelații



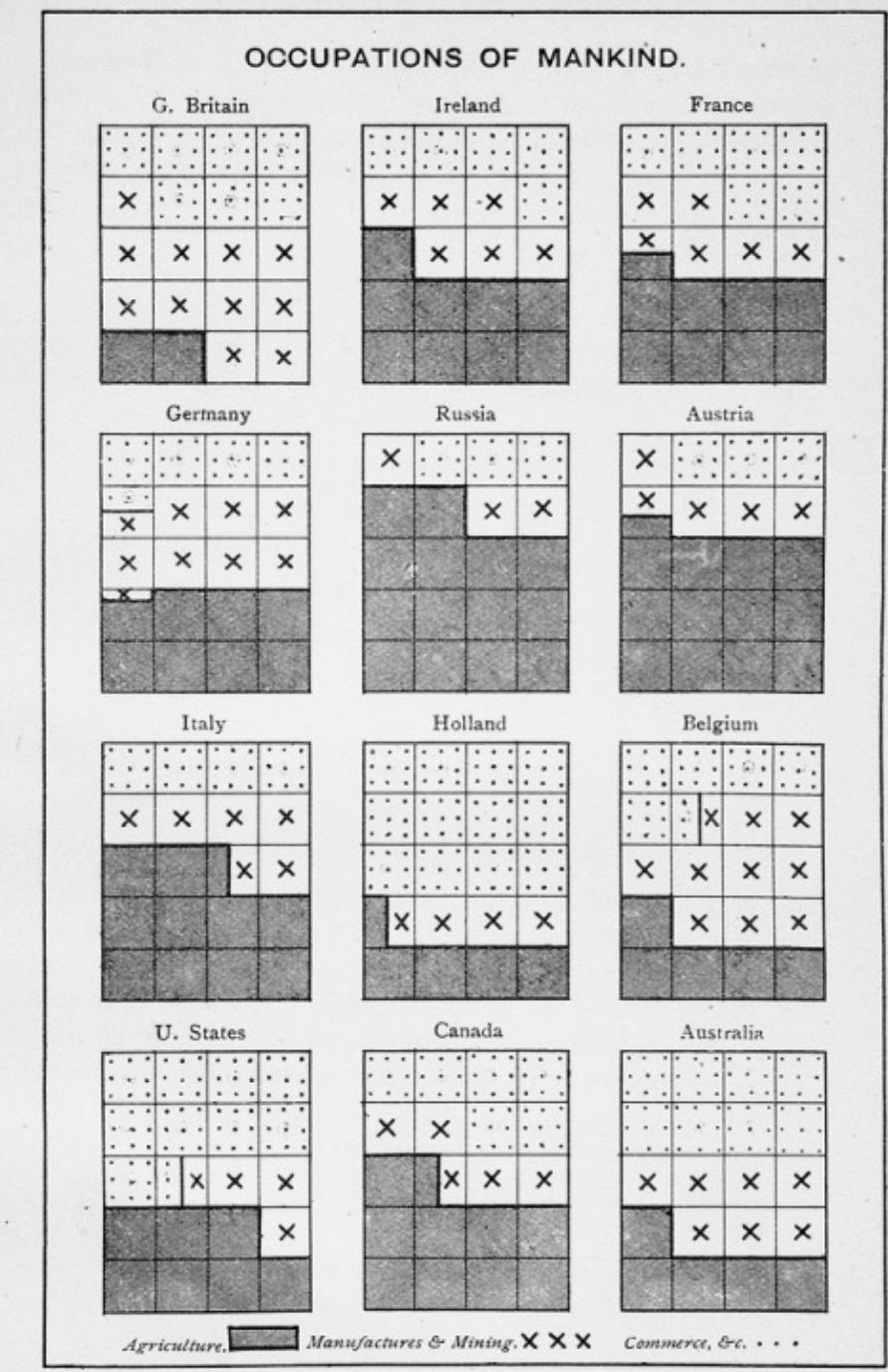


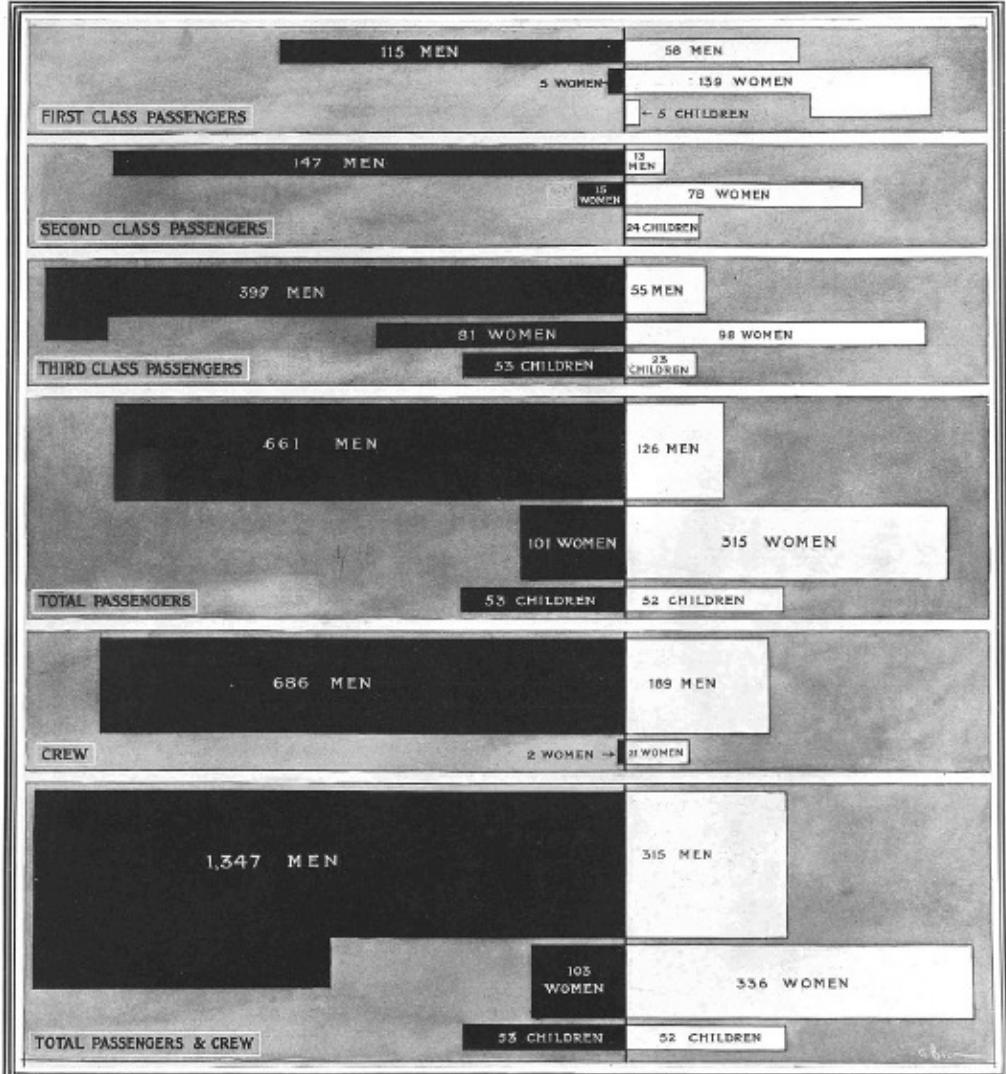
Figure 8.2 Square area diagrams showing occupations in several countries
Source: Mulhall, *Industries*, Plate III.

1912

THE LOSS of the "TITANIC."

The Results Analysed and Shown in a Special
"Sphere" Diagram

Drawn from the Official Figures Given in the House of Commons



[Sursa](#)

Sursa

American exceptionalism

United States and its individual races and ethnicities, compared with other countries

Rank out of 146 countries, 2019 or latest available figure

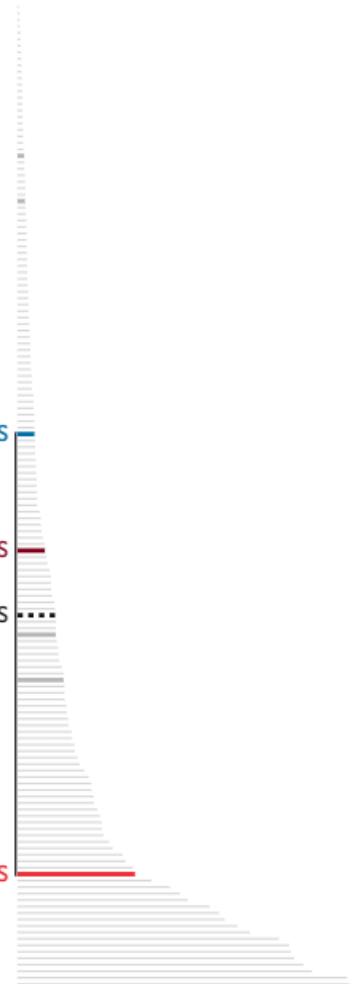
Life expectancy
Years, at birth



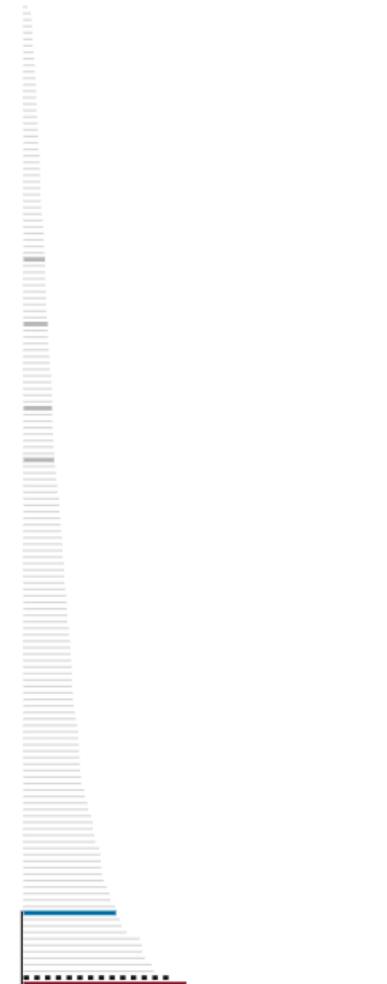
Infant mortality
Deaths per 1,000 births



Murders
Per 100,000 people

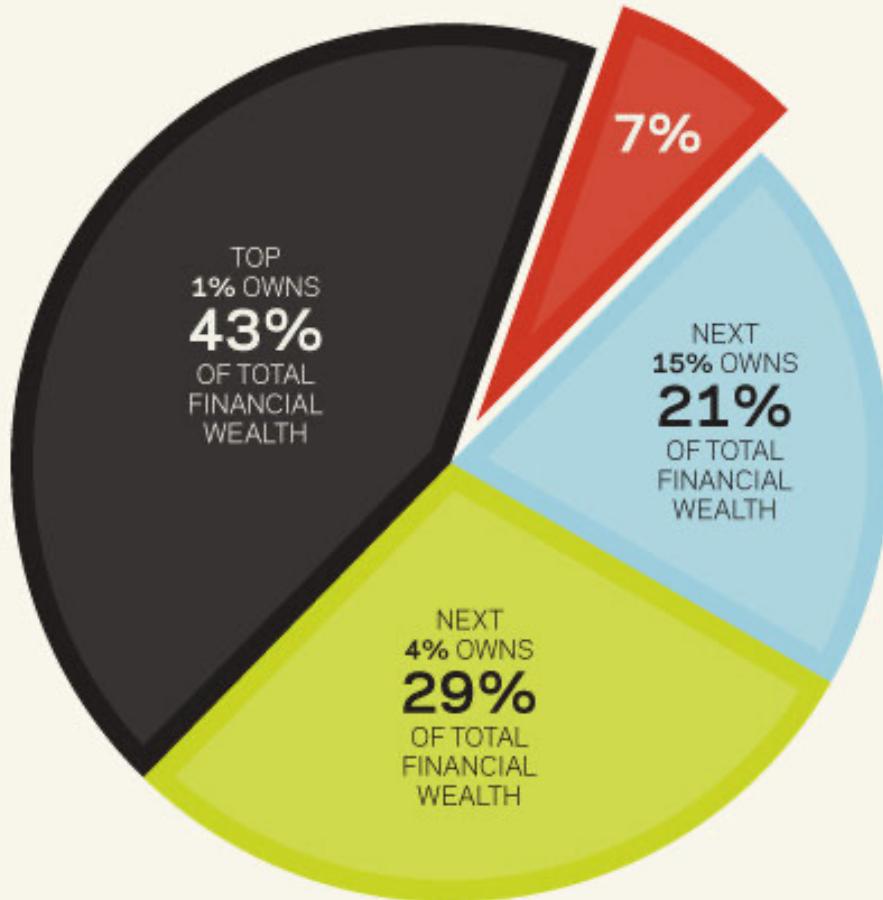


Prison population
Per 100,000 people



Sursă: [The Economist](#)

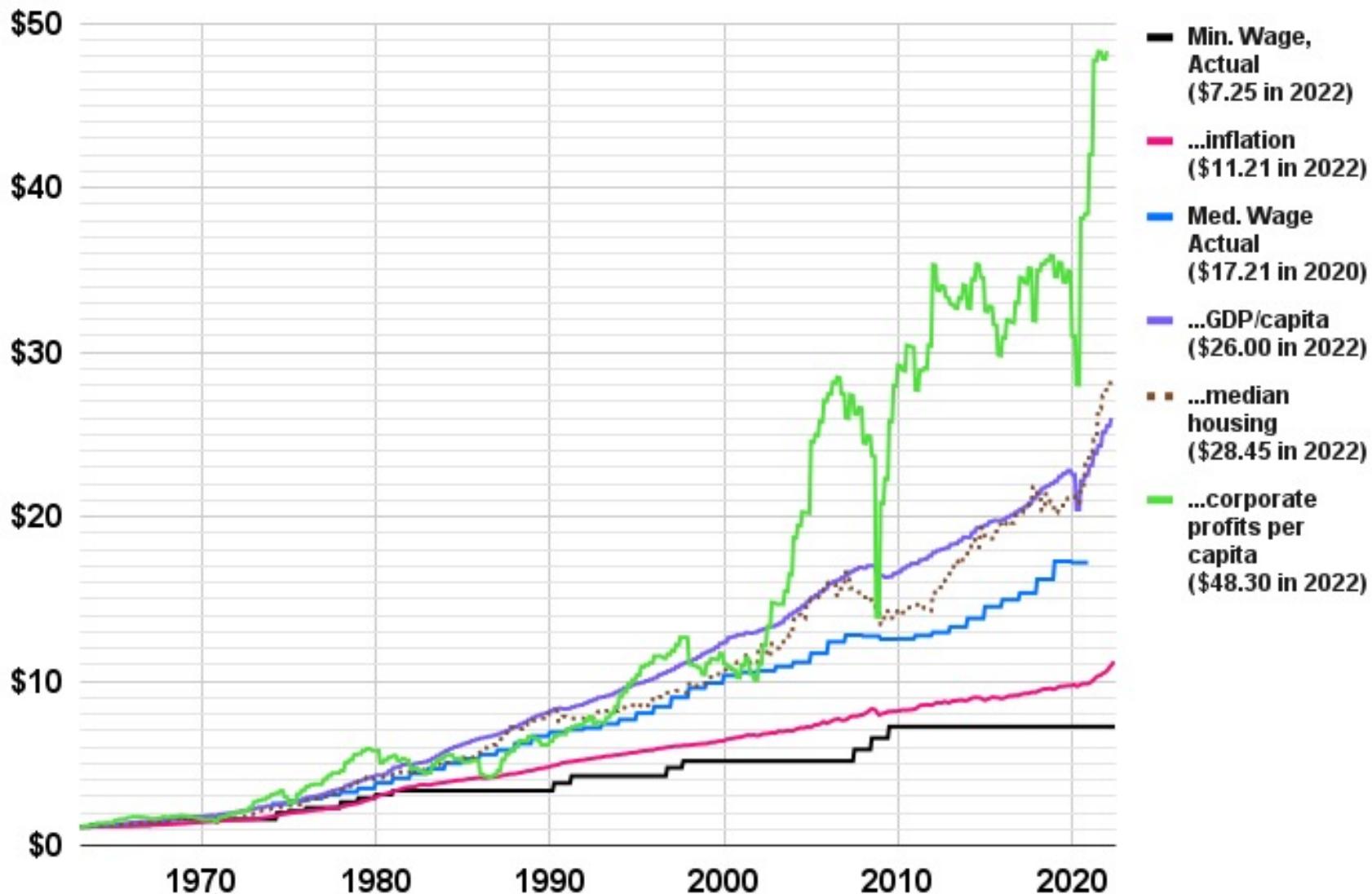
**80% OF AMERICANS SHARE ONLY 7% OF
ALL THE MONEY IN AMERICA**



THIS IS NOT WHAT DEMOCRACY LOOKS LIKE
DISTRIBUTION OF FINANCIAL WEALTH IN THE UNITED STATES

Sursa

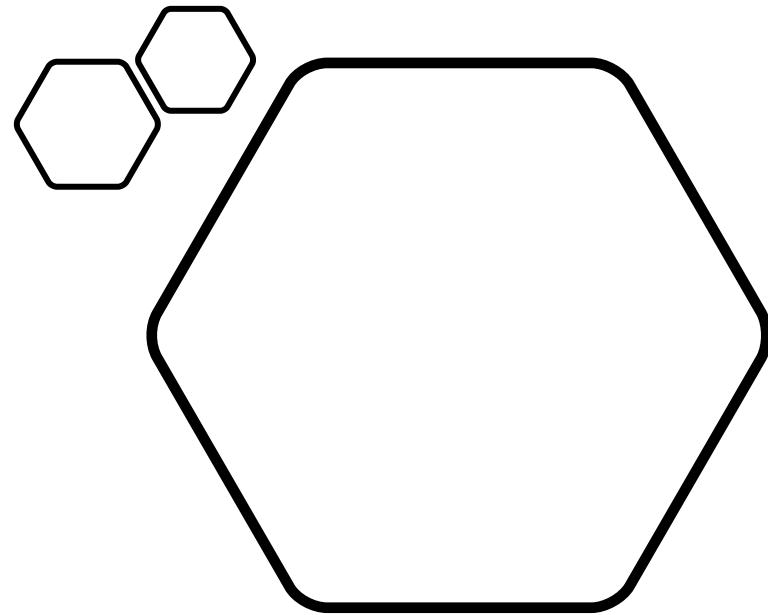
What would minimum wage be if it kept up with...?

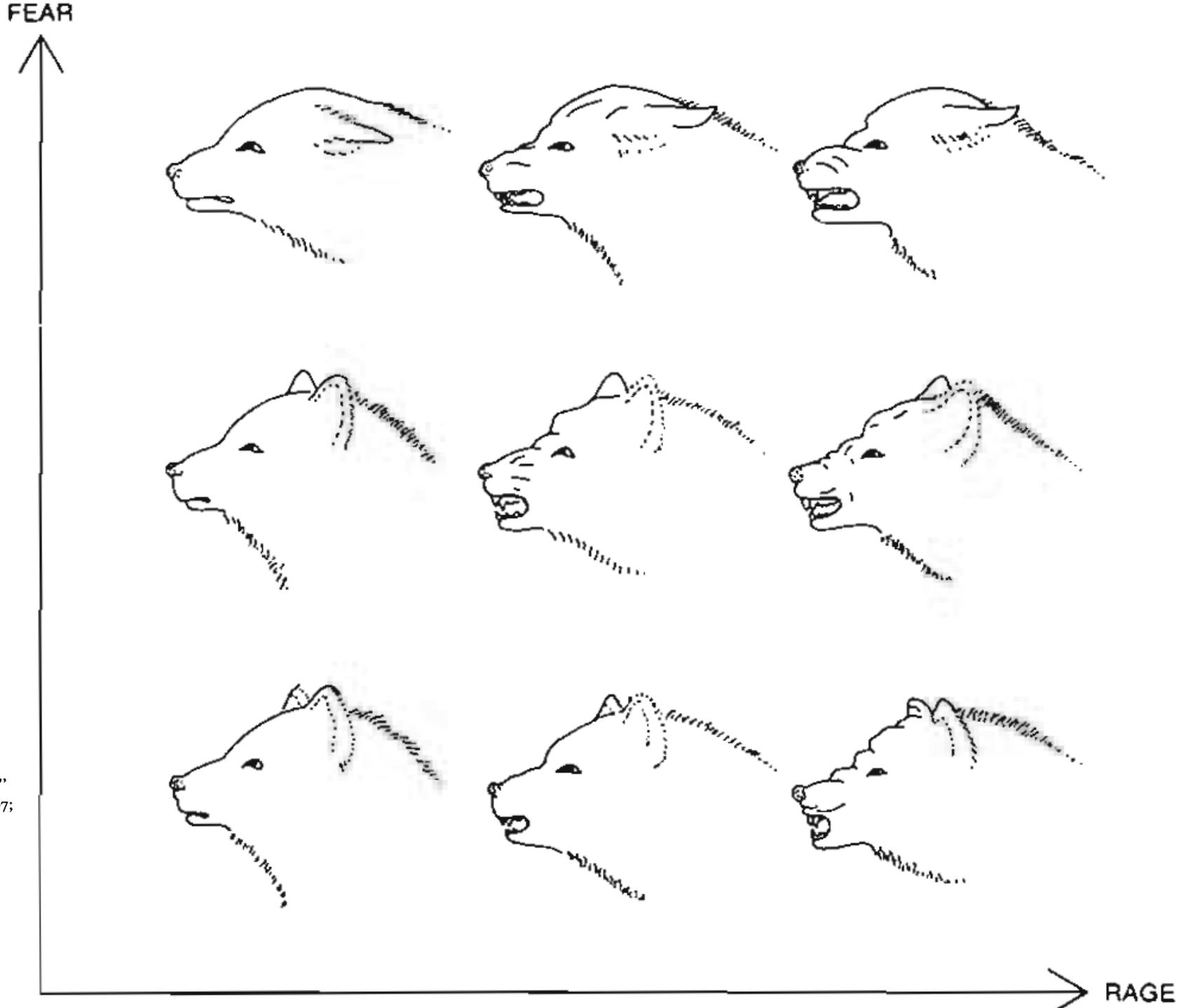


[Sursa](#)

5. Clasificări

Clasificări emergente și profilări





E. C. Zeeman, "Catastrophe Theory,"
Scientific American, 234 (April 1976), 67;
based on Konrad Z. Lorenz, *King
Solomon's Ring* (New York, 1952).

Sursă:
Tufte, 2001

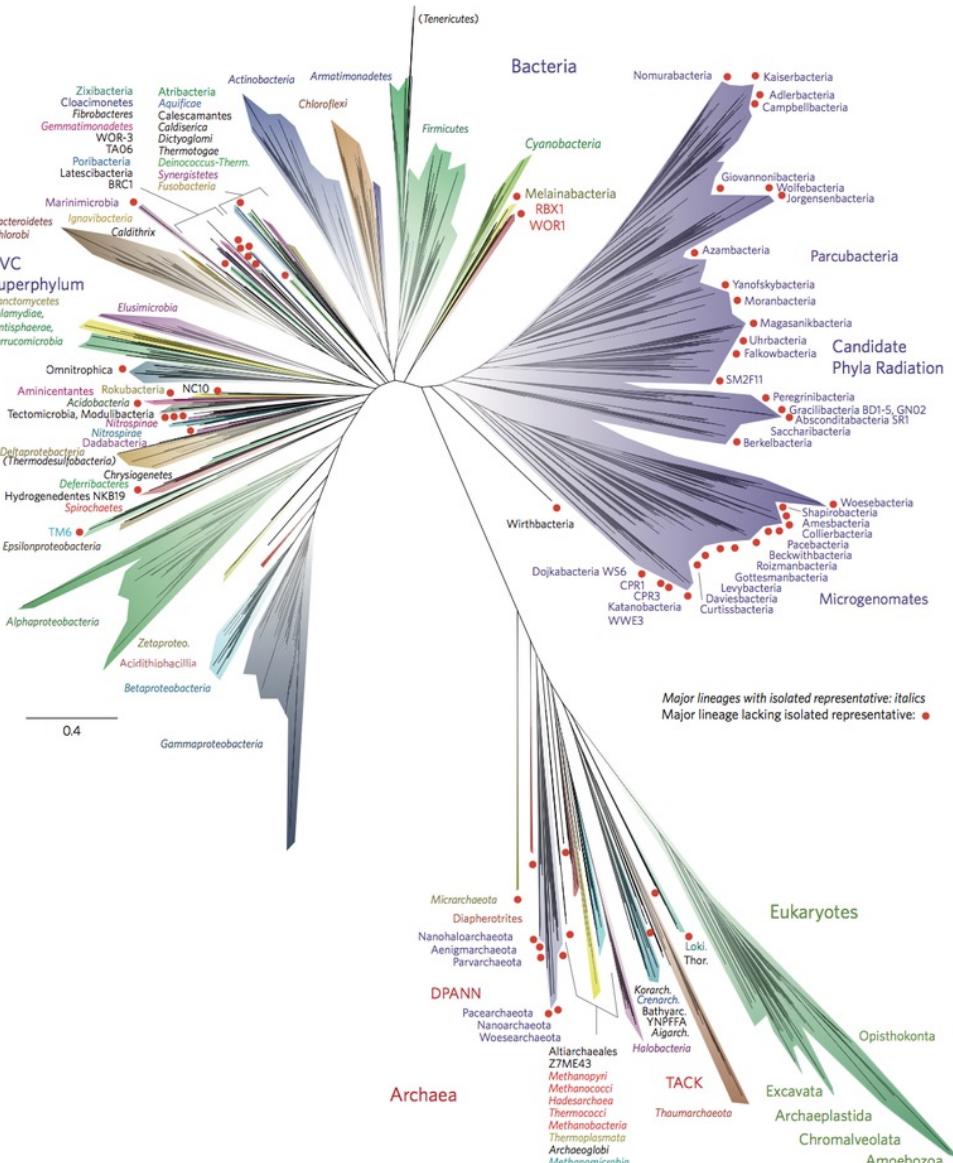
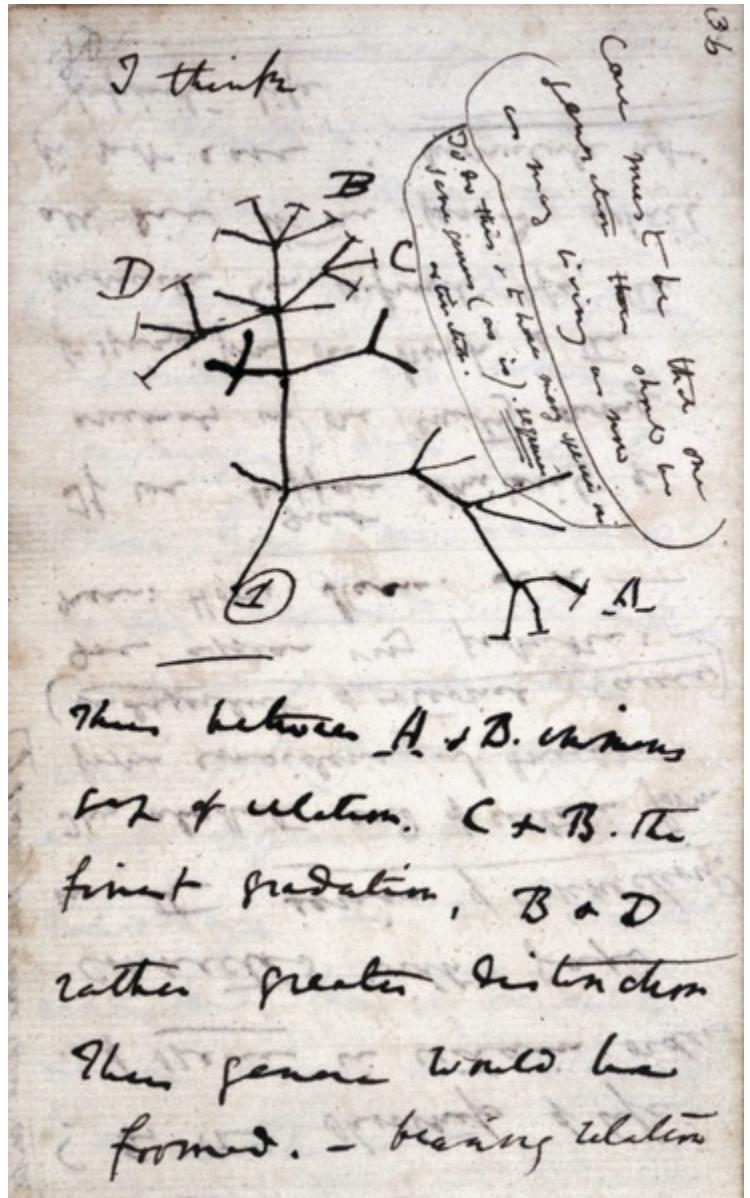
Clasificări

Clasificări emergente

- Identificarea tipurilor pe baza unor criterii
- Analiză cluster
- Exemple
 - Specii
 - Personalități
 - Genuri muzicale noi

Profilări

- Detalierea profilului unor tipuri pre-existente
- Tabele de contingență
- Exemple
 - Profilul culinar al orașelor
 - Profilul ocupațional al femeilor și bărbaților
 - Profilul utilizatorilor Android vs. iOS



How to Read the Circle of Life

Primordial life begins at the center and branches out in all directions, leading to the groups of species that exist today (colored rings)

Outer ring: Estimated proportion of all species*

Inner ring: Proportion of the groups named to date

Each black line represents at least 500 descendant species

Dark lines: Many species have been genetically sequenced

Light lines: Few species have been genetically sequenced

Nematodes (roundworms)

Lophotrochozoa (mollusks, segmented worms, brachiopods)

Deuterostomia (vertebrates, sea stars and urchins, certain worms)

Early diverging metazoa (cnidaria, comb jellies, sponges)

Many deuterostomia (gold) and plants (dark green) are already genetically sequenced (dark lines) because they are culturally or economically important (such as humans!)

Fungi

Arthropods (insects, arachnids, crustaceans)

Scientists have identified about one million arthropods (tan); millions more remain undescribed

Experts expect that most new species to be discovered will be bacteria (orange) and archaea (magenta)

Archaea (single-celled micro-organisms that tolerate extreme conditions)

Bacteria

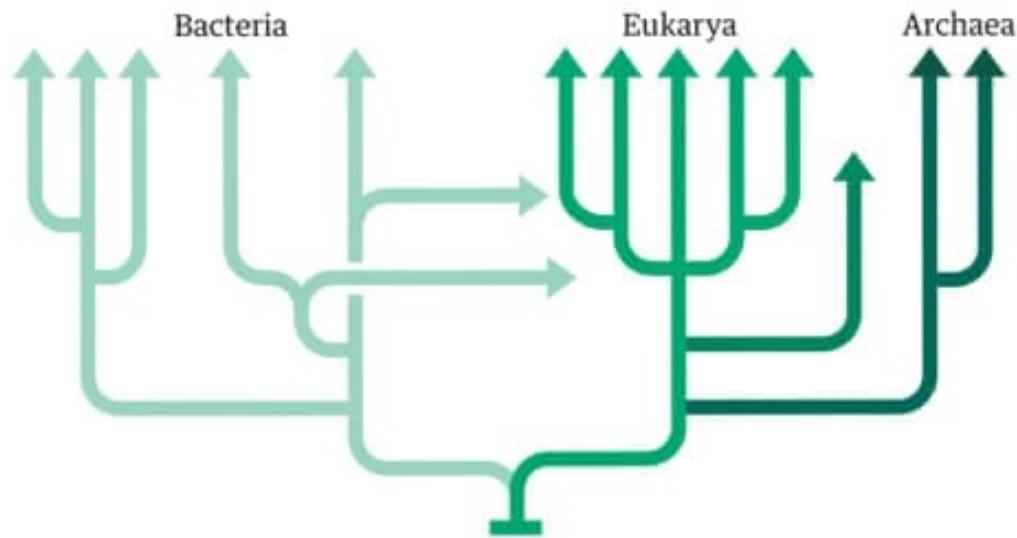
SARs[†] (diatoms, amoeboids, brown algae)

Early diverging archaeplastida (green algae, red algae)

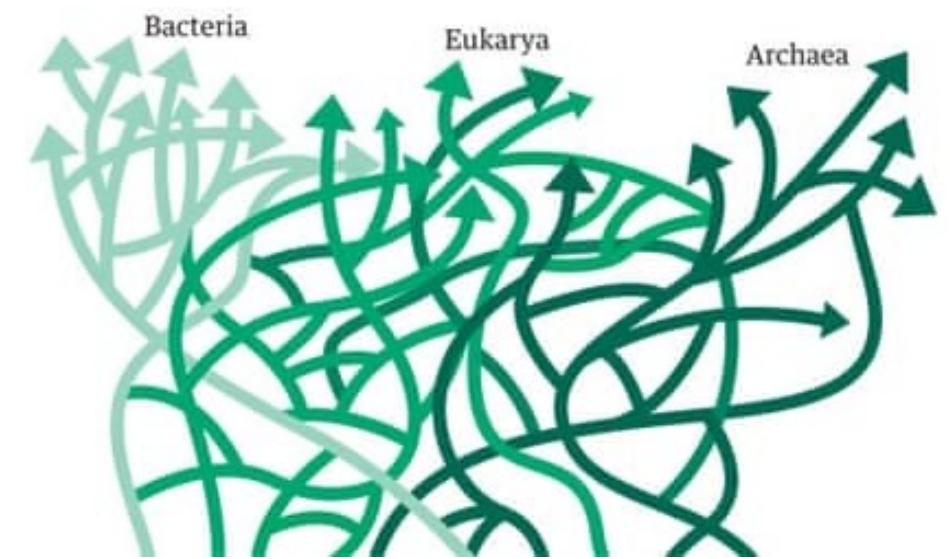
*Estimates vary widely; values shown are averages from multiple sources

[†]Stramenopiles, alveolates, Rhizarians

[Sursă](#)

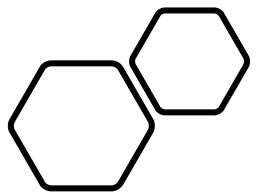


The Tree of Life (Ch. Darwin)

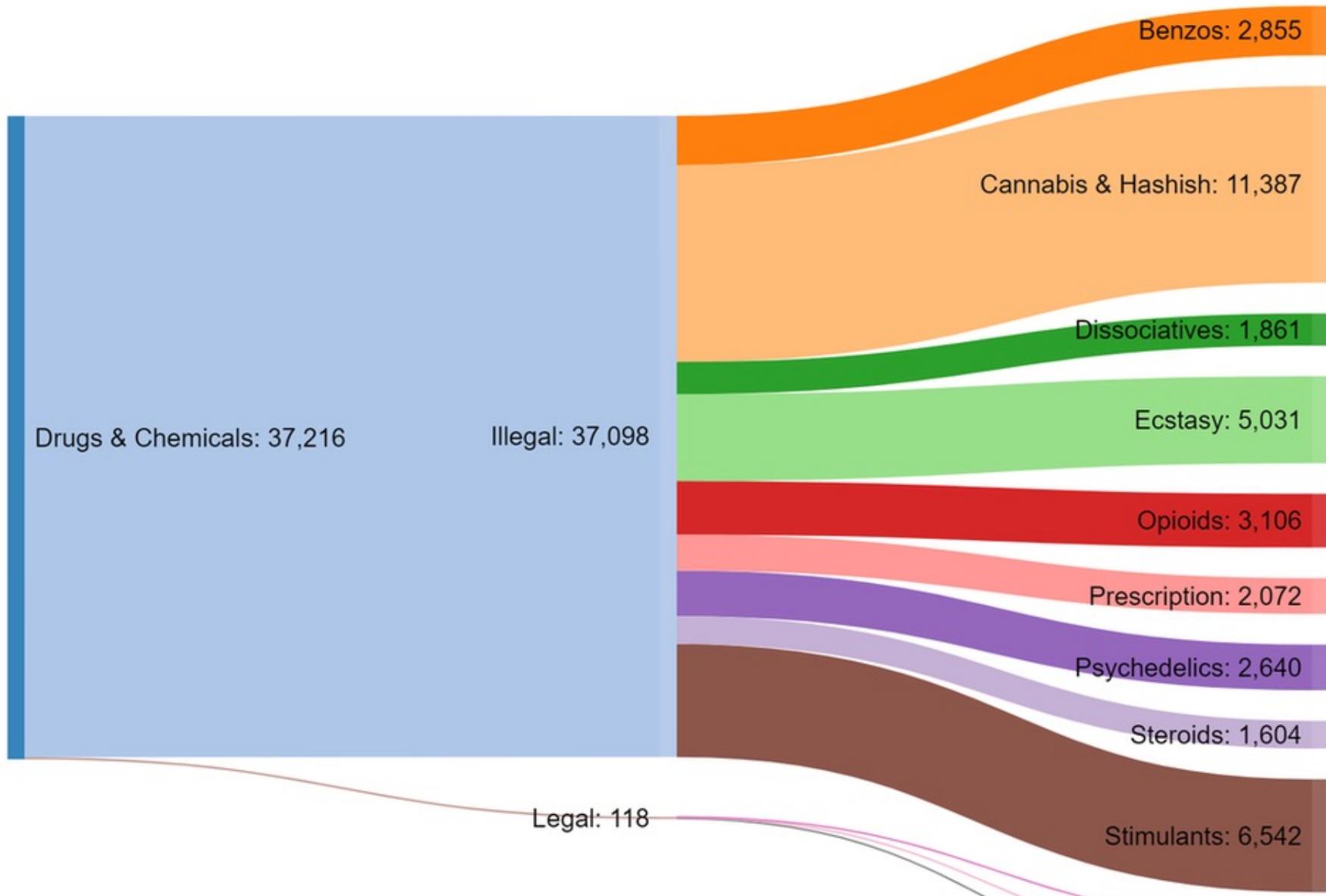


The Impenetrable Thicket of Life
(W Ford Doolittle)

Sursă



Profilări



[Sursa](#)

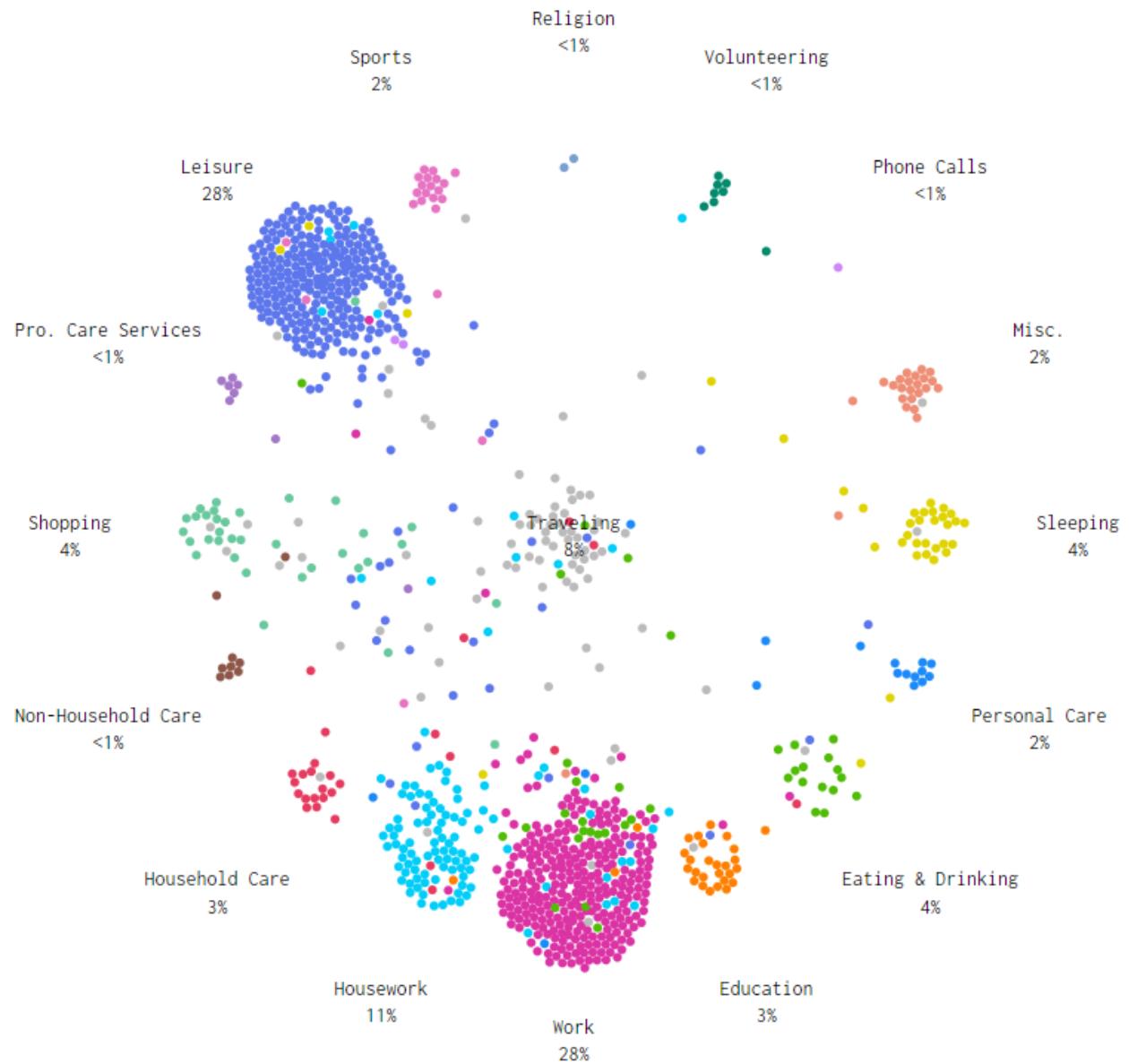
Number of listings per drugs category on
a popular darkweb market

3:22pm

SLOW MEDIUM **FAST**

Coffee break? Again, at the top of the hour, you see a shift in activity.

This is a simulation of 1,000 people's average day. It's based on 2014 data from the [American Time Use Survey](#), made way more accessible by the [ATUS Extract Builder](#).



The **SELFIEEXPLORATORY** is part of **SELFIECITY**

Help

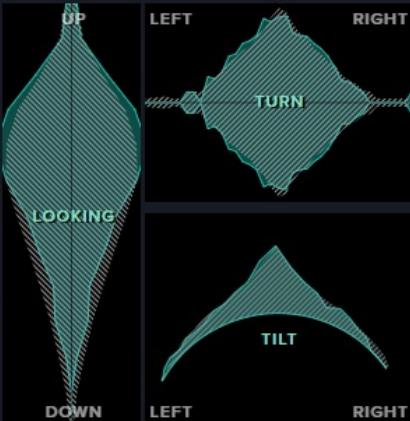
Share

Reset filters

DEMOGRAPHICS



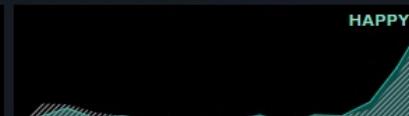
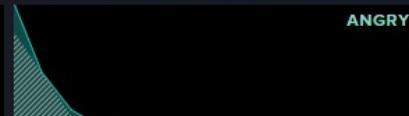
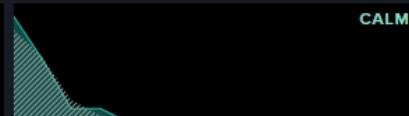
POSE



FEATURES



MOOD



640 of 3840 selfies.



Sursa
(interactiv)

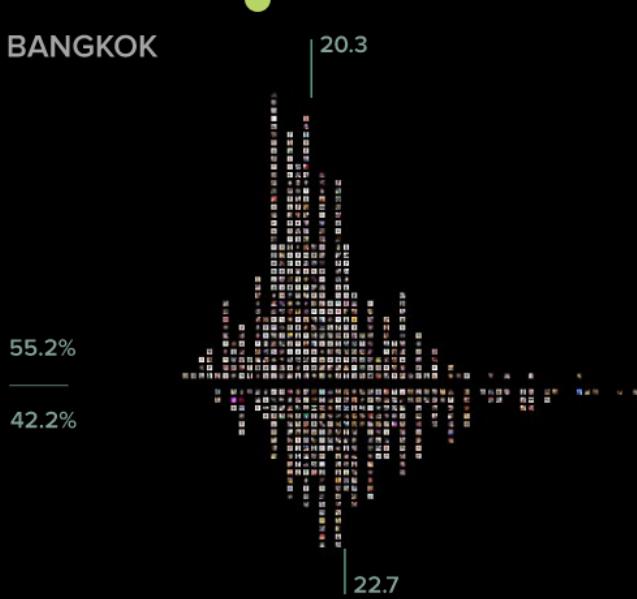
GENDER AND AGE PROFILES PER CITY

Case by case inspection of photos can reveal a lot of detail, but it is difficult to quantify the patterns observed.

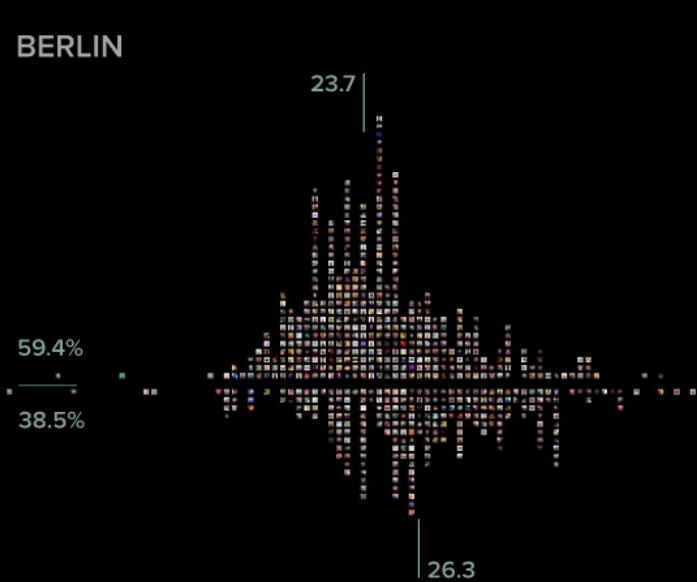
"Is it just me, or do Sao Paulo women actually tilt their heads more? Do New Yorkers or Berliners look older?"

In order to answer these questions, and supplement our rudimentary automatic face analysis with human judgment, we had thousands of photos inspected by Mechanical Turk workers, who estimated age and gender of the people on the photos. Here are the results:

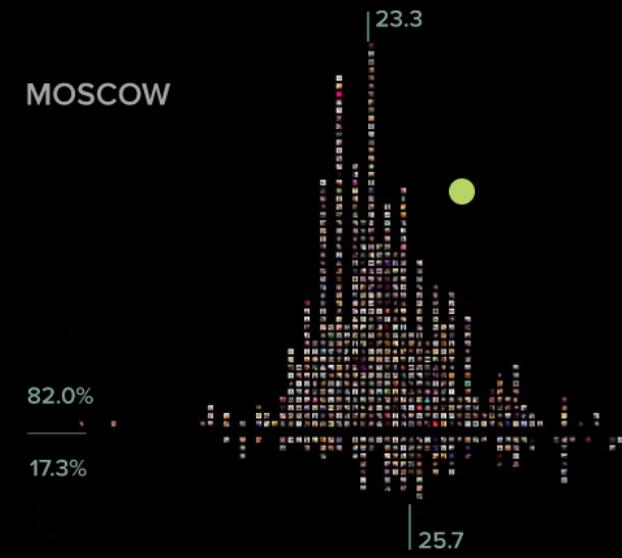
BANGKOK



BERLIN

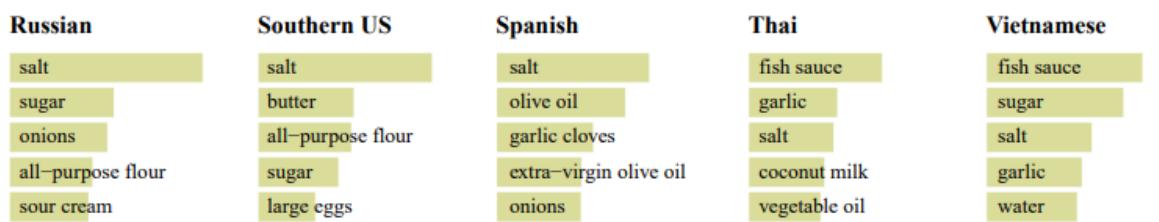
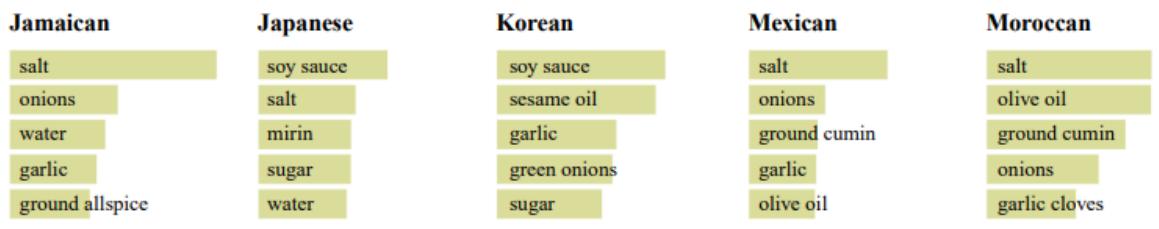
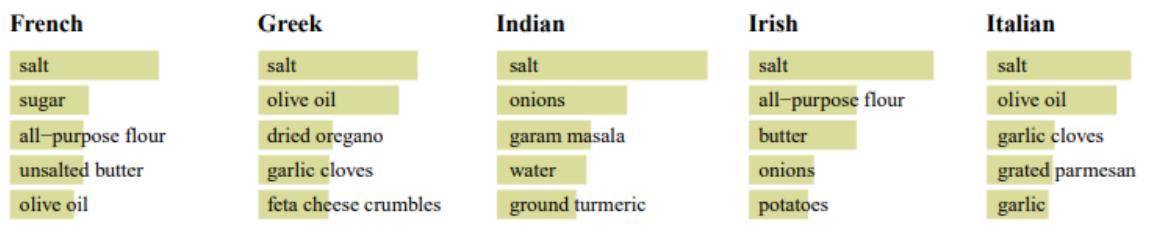
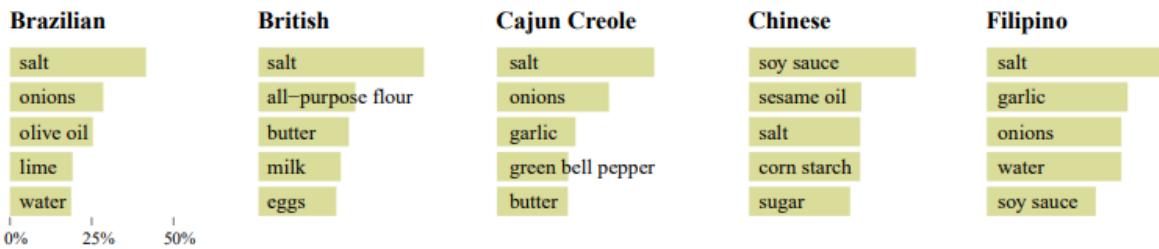


MOSCOW

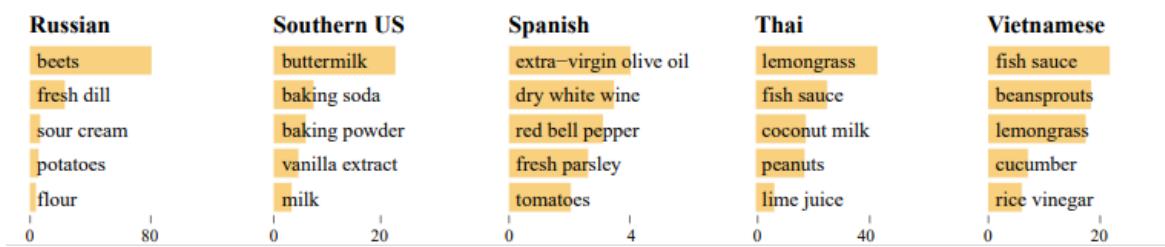
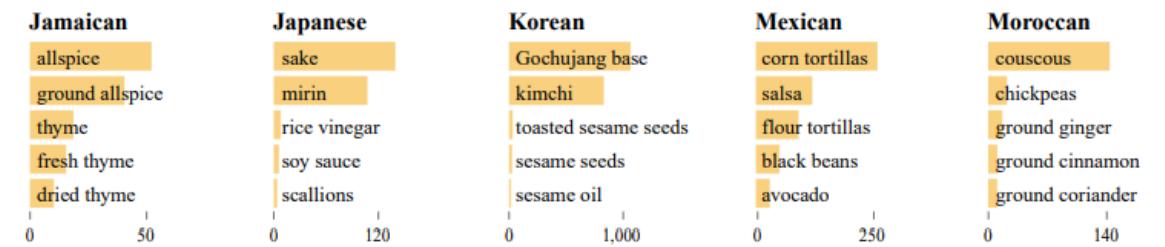
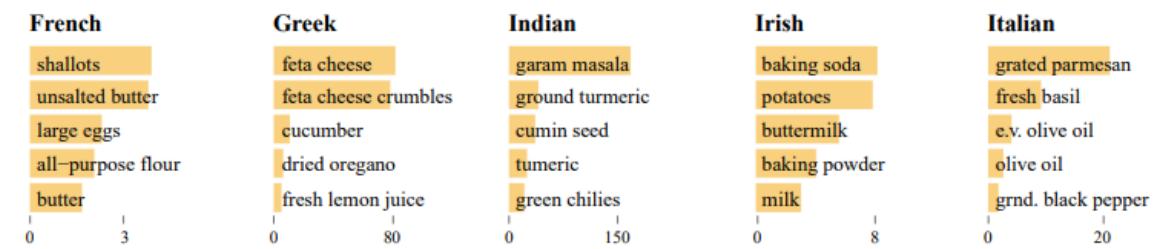
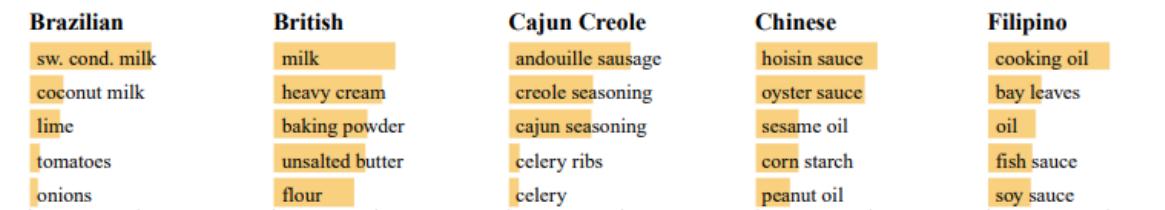


[Sursa](#)

Most Used Ingredients



Most Cuisine-Specific Ingredients



Books in my Living Room



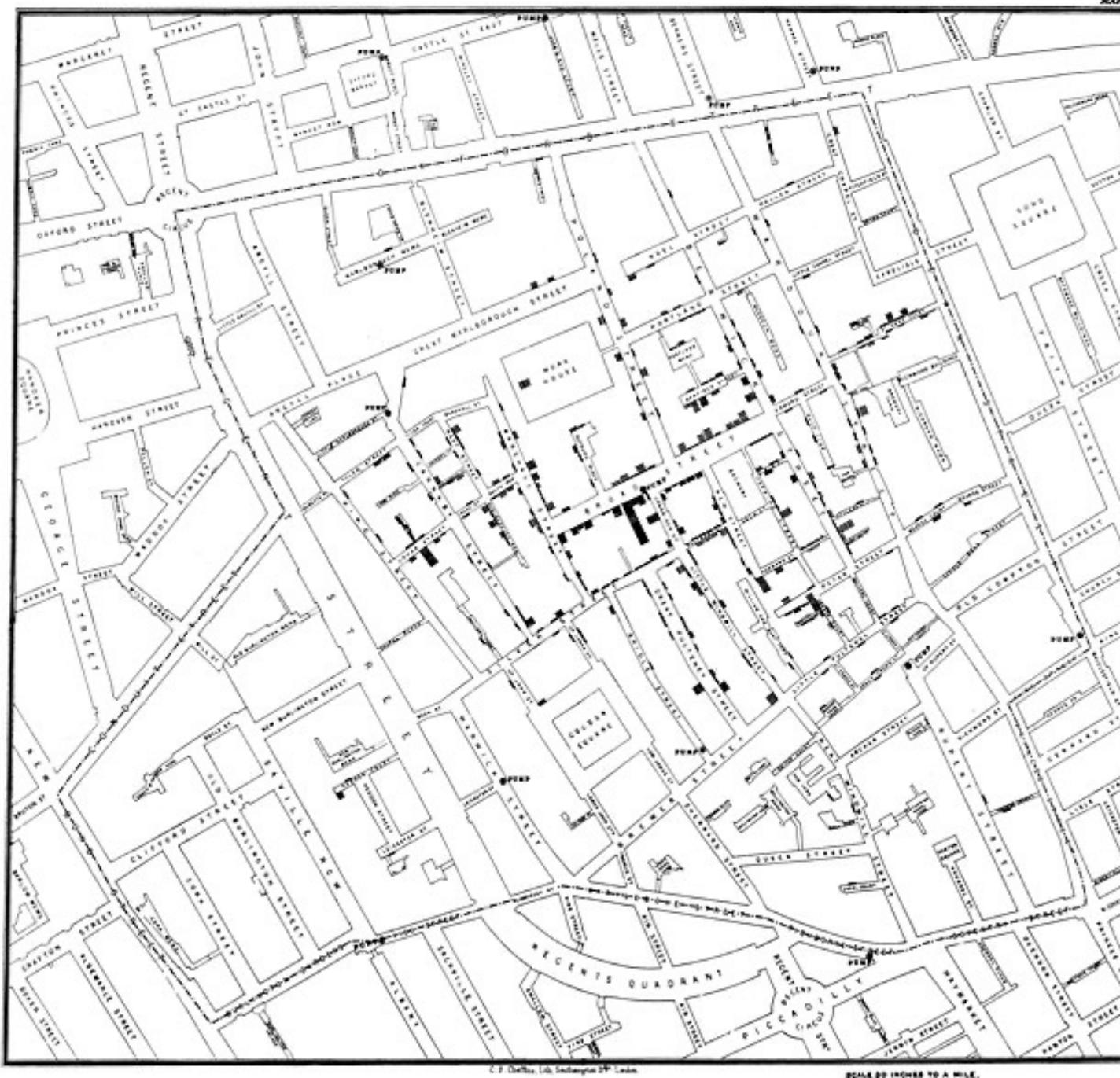
6. Cauzalitate

Corelația nu înseamnă (musai) cauzalitate!

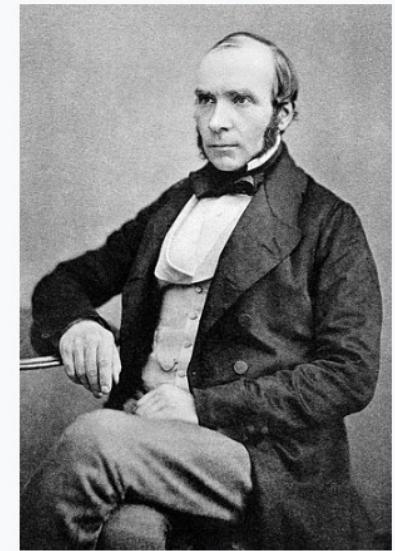
Analiză de regresie

Analiză path





John Snow



Born 15 March 1813
York, England

Died 16 June 1858 (aged 45)
London, England

Alma mater University of London

Known for Anaesthesia
Locating source of a cholera outbreak (thus establishing the disease as water-borne)

Scientific career

Fields Anaesthesia
Epidemiology

Signature

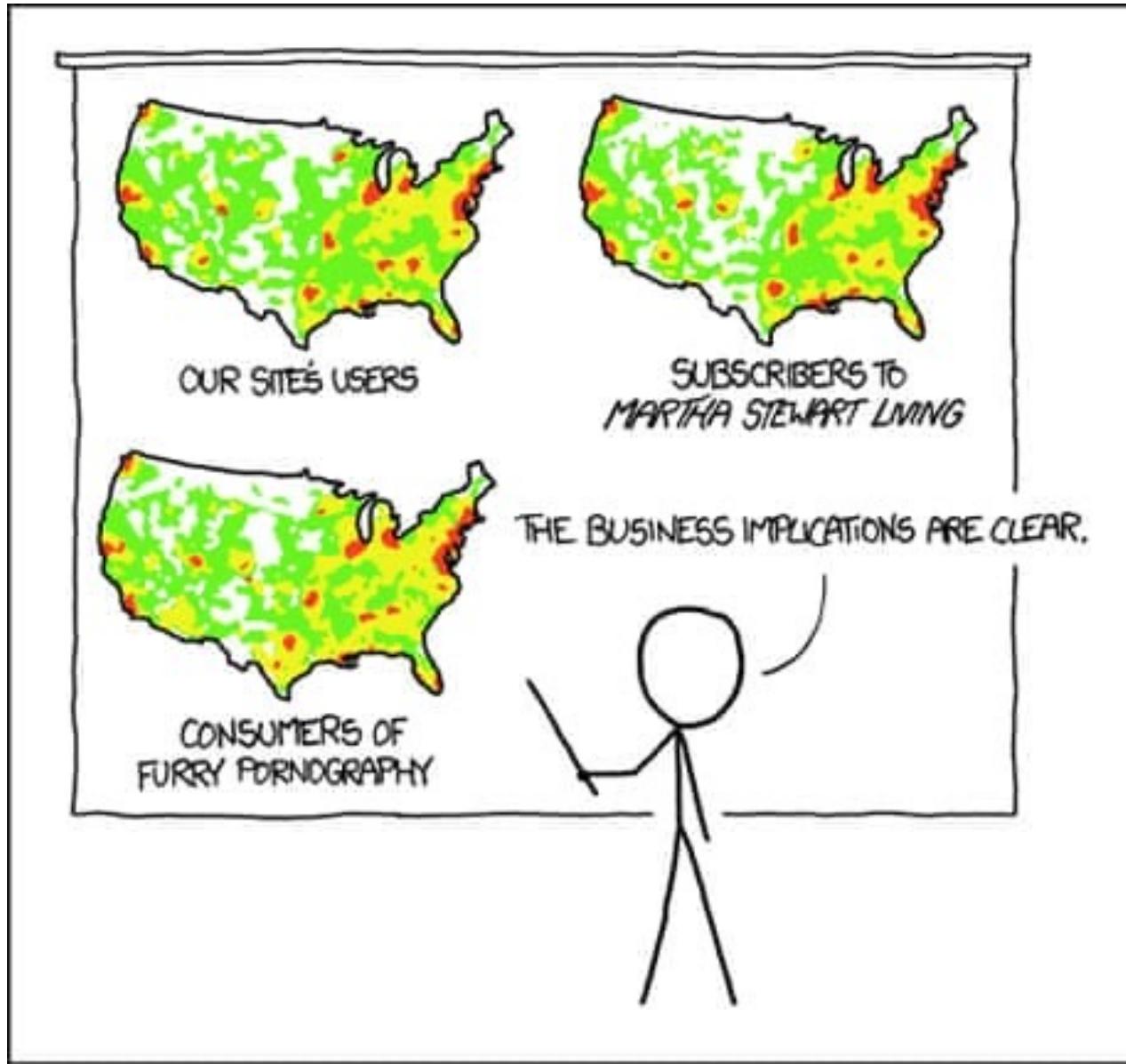
John Snow



„At a local brewery, the workers were allowed all the beer they could drink - it was believed they didn't drink water at all. But it had its own water supply too and there were consequently fewer cases.”

Sursă

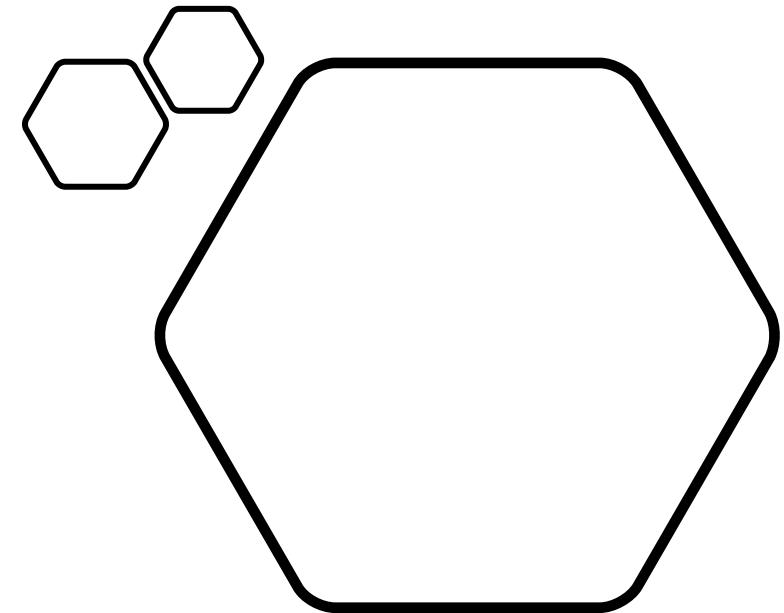
Discuție in
The Guardian



PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE
BASICALLY JUST POPULATION MAPS

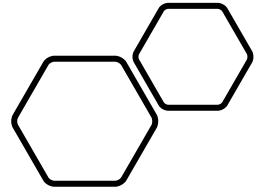
Concluzii

- Știința datelor (învățare automată) vs. analiză statistică
 - Transformare dinspre explicație spre predicție
 - Utilizarea datelor vaste și nestructurate
- Analiza statistică a datelor oferă vocabularul conceptual pentru știința datelor
- Pattern-urile au surse, utilizări și riscuri multiple



Referințe

1. Cao L. [Data science: a comprehensive overview](#). ACM Computing Surveys (CSUR). 2017 Jun 29;50(3):1-42.
2. The Economist. [Worth a Thousand Words](#). 2003 [Site]
3. Edward R. Tufte, [The graphical display of quantitative information](#). Graphics Press. 2001
4. Edward R. Tufte, [Beautiful evidence](#). Graphics Press, 2006.
5. Zach Gemignani. [20 Best examples of charts and graphics](#).
6. Tableau. [Data is beautiful: 10 of the best data visualisation examples from history to today](#)
7. [Gapminder](#)
8. Walt Hickey. [The 27 worst charts of all time](#). Business Insider, 2013.



Administrativ

Calendar

05 Oct	Why?	06 Oct	Cauzalitate
		13 Oct	Măsurare
18 Oct	Analiza datelor și COVID	20 Oct	Modelare și eșantionare
		27 Oct	Tehnici de analiză
02 Noi	Test lab 1-3	03 Noi	Predictia
		10 Noi	Programare și ML
16 Noi	Grilă curs 1-6	17 Noi	ML și Deep Learning
		24 Noi	Producția în arhitecturi ML
07 Dec	Q&A	08 Dec	Why privacy?
		17 Dec	Privacy Preserving Algorithms
11 Ian	Test Lab 4-9	12 Ian	Privacy Architectures and Federated Learning
		19 Ian	Studiu de caz

Echipa



Răzvan Rughiniş
✓ Fun Learning
✓ #makeitreal
✓ Biking



Dan Ciocîrlan
✓ Number representations
✓ Dota/fifa/table/şah
✓ Computer architecture

Echipa



Dan-Gabriel Badea

- ✓ Software Development
- ✓ Personal Development
- + Psychology
- ✓ Parties + Dancing



Damian Monea

- ✓ Data Scientist
- ✓ Professional memer
- ✓ When life gives you lemons, ask for salt and tequila

Echipa



Teodor Neacșu

- ✓ AI
- ✓ ML
- ✓ NLP



Andrei Ouatu

- ✓ Number Representation System
- ✓ Data Privac6
- ✓ Computer Architectures



Paul Sumedrea

- ✓ Computer Vision
- ✓ Cybersecurity
- ✓ Graph Neural Networks

Notare

- Curs 5p
 - 2.5p – Grilă curs 1-6 [16.11.2022]
 - 2.5p – Grilă curs 7-12 [sesiune]
- Lab 5p
 - 2p – participarea la laborator
 - 1p – test lab 1-3 [02.11.2022]
 - 2p – test lab 4-9 [11.01.2023]

PR Pins Awards

- 2 primele două scoruri la grila de midterm (departajare prin punctaj și apoi timp)
- 8 pentru implicare în laboratoare (negociate de asistenți)
- 2 pentru implicare în cursuri
- 2 primele două scoruri la grila din sesiune
- 2 contribuții în comunitate (sugestii pe conținut laboratoare, prezență pe forumuri)

