Data Visualization

elena casiraghi

Ufficio 6006, Dipartimento di Informatica

Mail to: elena.casiraghi@unimi.it

casiraghi@di.unimi.it

UNIVERSITÀ DEGLI STUDI DI MILANO

Let’s start with some nice DataVis / SciVis video

Visual explanation of exponential growth and epidemics

The Big Data era

"All too much. Monstrous amounts of data."

The Economist. 25 Febbraio 2010

2008:

[…] According to a **2008** study by International Data Corp (IDC), a market-research firm,

around **1,200 exabytes** of digital **data will be generated this year (2010).** […]

What about the information that was actually consumed in 2008?

An example:

In 2008 American households **were bombarded with 3.6 zettabytes of information (about 34 gigabytes per person per day)**.

Only 5% of the information that is created is "structured", meaning it comes in a standard format of words or numbers that can be read by computers.

The rest are data flows **less easily retrievable, understandable and usable**.

In 2020, there will be around 40 trillion gigabytes (~= 40 zettabytes) of data

(From IDC report "The Digital Universe in 2020")

"The digital universe would double every two years until 2020"

2010 -> 1.2 zettabyte

2012 -> 2.4 zettabyte

2014 -> 4.8 zettabyte

2016 -> 9.6 zettabyte

2018 -> 19.2 zettabyte

2020 -> 38.4 **(~= 40 zettabytes)**

In 2012, only 0.5% of all data was analyzed and useful data was the 22%.

(Source: The Guardian)

In 2012 only 22% of all the data had the potential for analysis (Source: IDC).

By 2020, the percentage of useful data, i.e., the information that has the potential for analysis, would jump to 37% (Source: IDC).

2017: the 2. 90% of all data has been created in the last two years.

(Source: IBM)

In 2012 we had 2.5 billion internet users.

In 2014 we had 3 billion

In 2019 we have 4.1 billion people online AND

in 2019, internet users spent 1.2 billion years online (Source Digital 2019)

Data inflation

2

Unit	Size	What it means
Bit (b)	1 or 0	Short for "binary digit", after the binary code (1 or 0) computers use to store and process data
Byte (B)	8 bits	Enough information to create an English letter or number in computer code. It is the basic unit of computing
Kilobyte (KB)	1,000, or 2^{10}, bytes	From "thousand" in Greek. One page of typed text is 2KB
Megabyte (MB)	1,000KB; 2^{20} bytes	From "large" in Greek. The complete works of Shakespeare total 5MB. A typical pop song is about 4MB
Gigabyte (GB)	1,000MB; 2^{30} bytes	From "giant" in Greek. A two-hour film can be compressed into 1-2GB
Terabyte (TB)	1,000GB; 2^{40} bytes	From "monster" in Greek. All the catalogued books in America's Library of Congress total 15TB
Petabyte (PB)	1,000TB; 2^{50} bytes	All letters delivered by America's postal service this year will amount to around 5PB. Google processes around 1PB every hour
Exabyte (EB)	1,000PB; 2^{60} bytes	Equivalent to 10 billion copies of *The Economist*
Zettabyte (ZB)	1,000EB; 2^{70} bytes	The total amount of information in existence this year is forecast to be around 1.2ZB
Yottabyte (YB)	1,000ZB; 2^{80} bytes	Currently too big to imagine

Source: *The Economist*

The prefixes are set by an intergovernmental group, the International Bureau of Weights and Measures. Yotta and Zetta were added in 1991; terms for larger amounts have yet to be established.

5. Internet users generate about 2.5 quintillion bytes of data each day.
(Source: Data Never Sleeps 5.0)

In 2012 we had 2.5 billion internet users.

In 2014 we had 3 billion

In 2019 we have 4.1 billion people online.

TOOLS FOR BIG DATA EXPLORATION, ANALYSIS, AND PROCESSING ARE NEEDED

Using big data, Netflix saves $1 billion per year on customer retention.

(Source: Statista, Inside Big Data)

Netflix provides recommendations through big data analysis.

Which data? searches, ratings, re-watched programs, and so on.

In 2009 Netflix invested $1 million in enhancing their recommendation algorithm.

In 2015 Netflix invested $651 million in enhancing their technologies and developments

In 2018, the budget reached $1.3 billion.

FROM THE BEGINNINGS TO 2003 humans generated about 5 exabytes of data,

NOW in TWO DAYS we generate 5 exabytes.

5 exabytes: more than 200.000 years of DVD videos.

What kind of data?

- chat, e-mail, phone-calls, photos, …
- Unstructured data produced by processes and by communications between, e.g., computers, mobiles, manufacturing systems…

Erich Schmidt, ex- CEO di Google 2010

"It is a very sad thing that nowadays there is so little useless information,"

Oscar Wilde's quotation (1894)

ALL THIS DATA IS USEFUL!

Examples:

- to understand user preferences
- to follow the behaviour of machineries and predict damages

  (**A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications - DOI:** 10.1109/ACCESS.2019.2953499)
- to monitor persons' health

  (**Human Digital Twin for Fitness Management - DOI:** 10.1109/ACCESS.2020.2971576)

4V for big data

- **Volume** (storage of large amounts of data),
- **Variety** (the heterogenic nature of data),
- **Velocity** (the speed of acquisition, opposed to long processing time) of data,
- **Value** (the significance of the information carried by data).

If we “see” (understand) data we may:

- **reduce their volume**: discard redundant/unuseful infos;
- **speed computation;**
- **increase value** by extrapolating hidden infos;
- **exploit their eterogeneity** by applying ad hoc data-fusion techniques;

So how do we deal with so much data?

... the information anxiety

The 70s: **Richard Saul Wurman** (born March 26, 1935), American architect and graphic designer, predicts the data and information explosion and firstly mentions the Information architecture (IA) field.

Information architecture (IA): focuses on organizing, structuring, and labeling data and information in an effective and sustainable way.

The goals of IA:

- to help users **understand information** and complete tasks

  understand information = how pieces fit together, how items relate and interact with each other

- define principles to bring information to the digital landscape

  principles of information design

  principles of shared information environments

OTHER DEFINITIONS OF IA:

1 - A subset of data architecture where usable data (a.k.a. information) is constructed and designed or arranged in a fashion most useful or empirically holistic to the users of this data.

2 - The conceptual framework surrounding information, providing context, awareness of location and sustainable structure.

Therefore Information Architecture helps winning the *in*formation anxiety

Information Anxiety (Wurman):

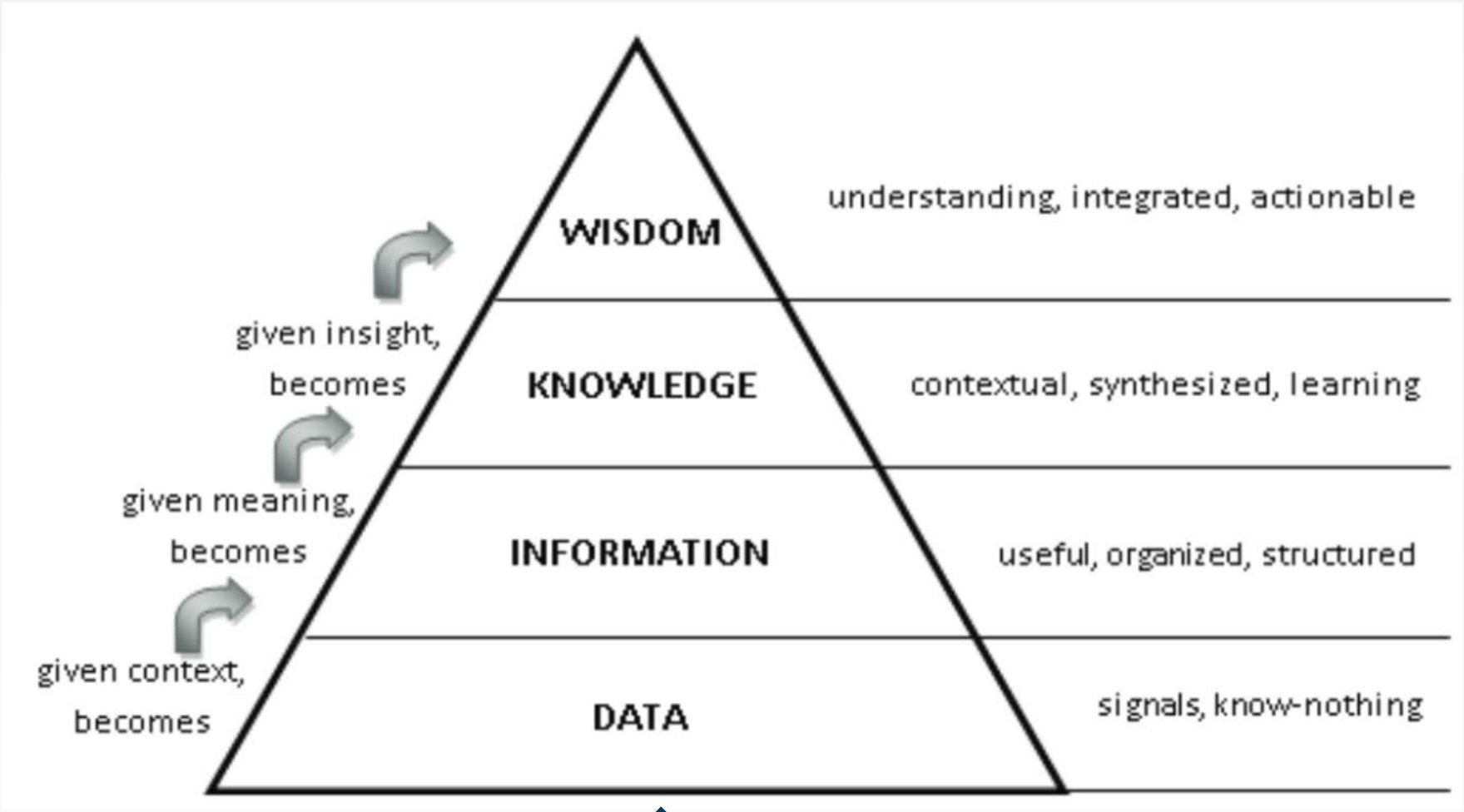
"caused by an overwhelming flood of data, much of it from computers and much of it unintelligible."

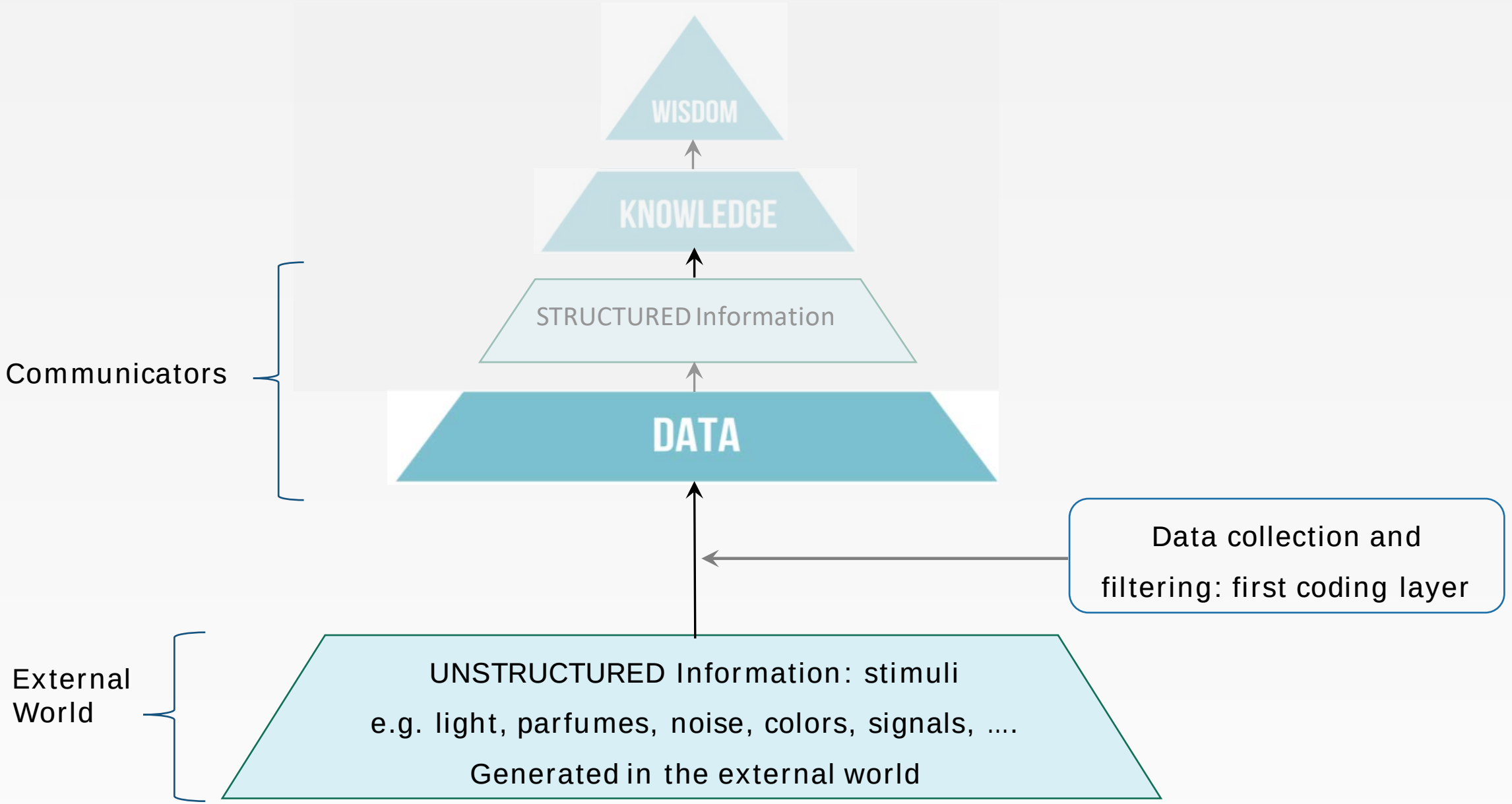
Source, Rowley, Jennifer.

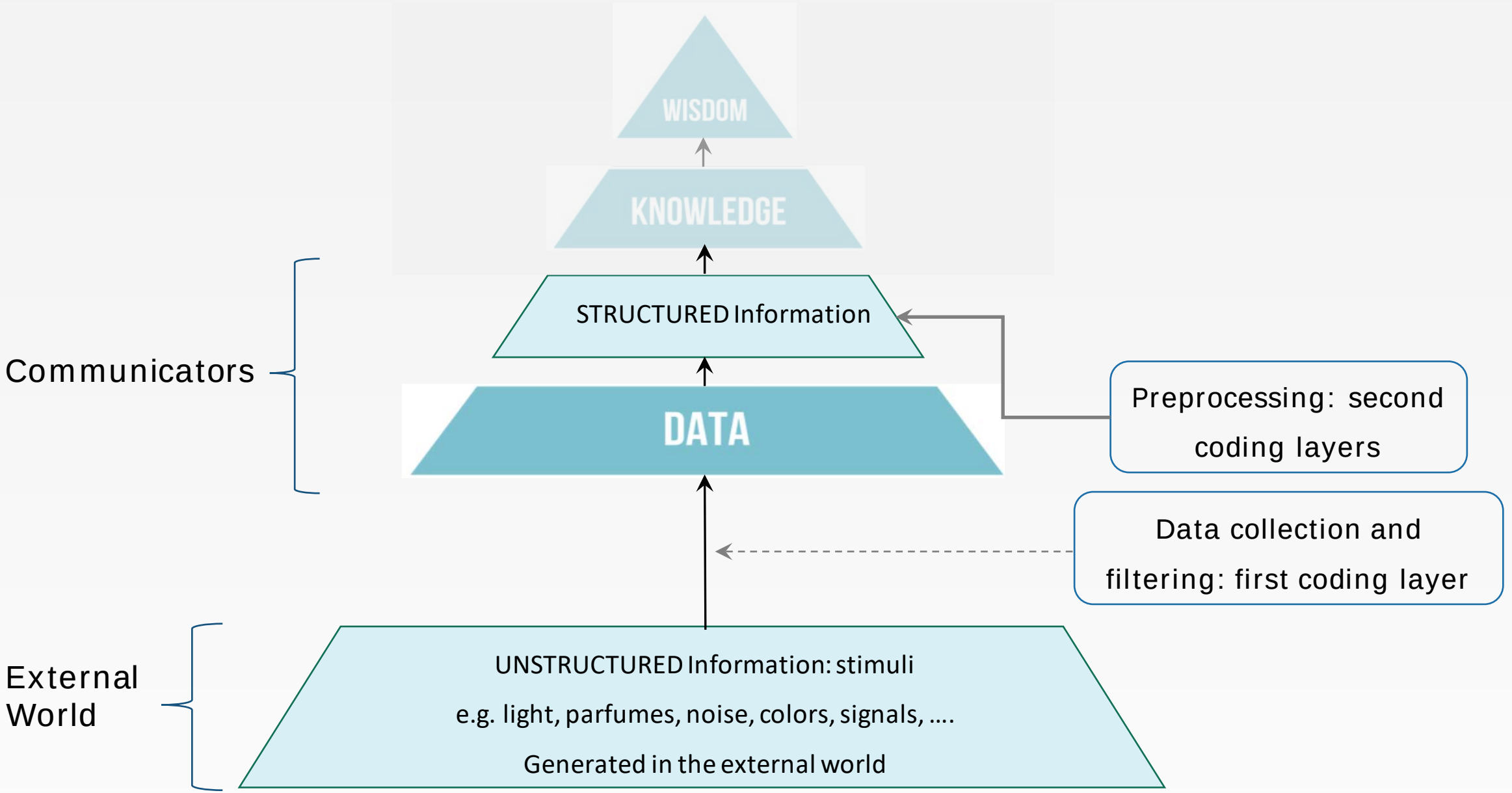
"The wisdom hierarchy: representations of the DIKW hierarchy".

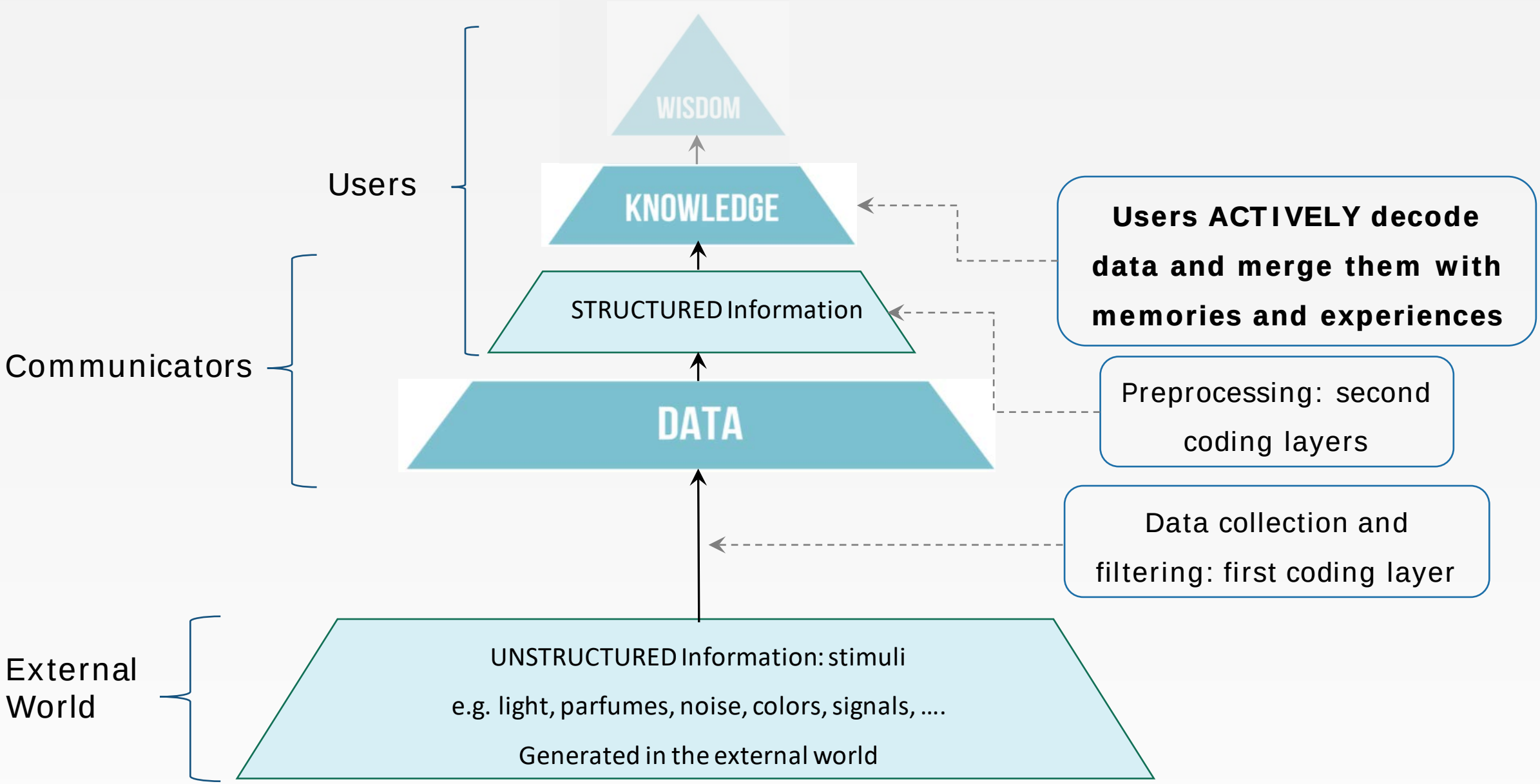
Journal of Information and Communication Science. 33 (2): 163–180, 2007

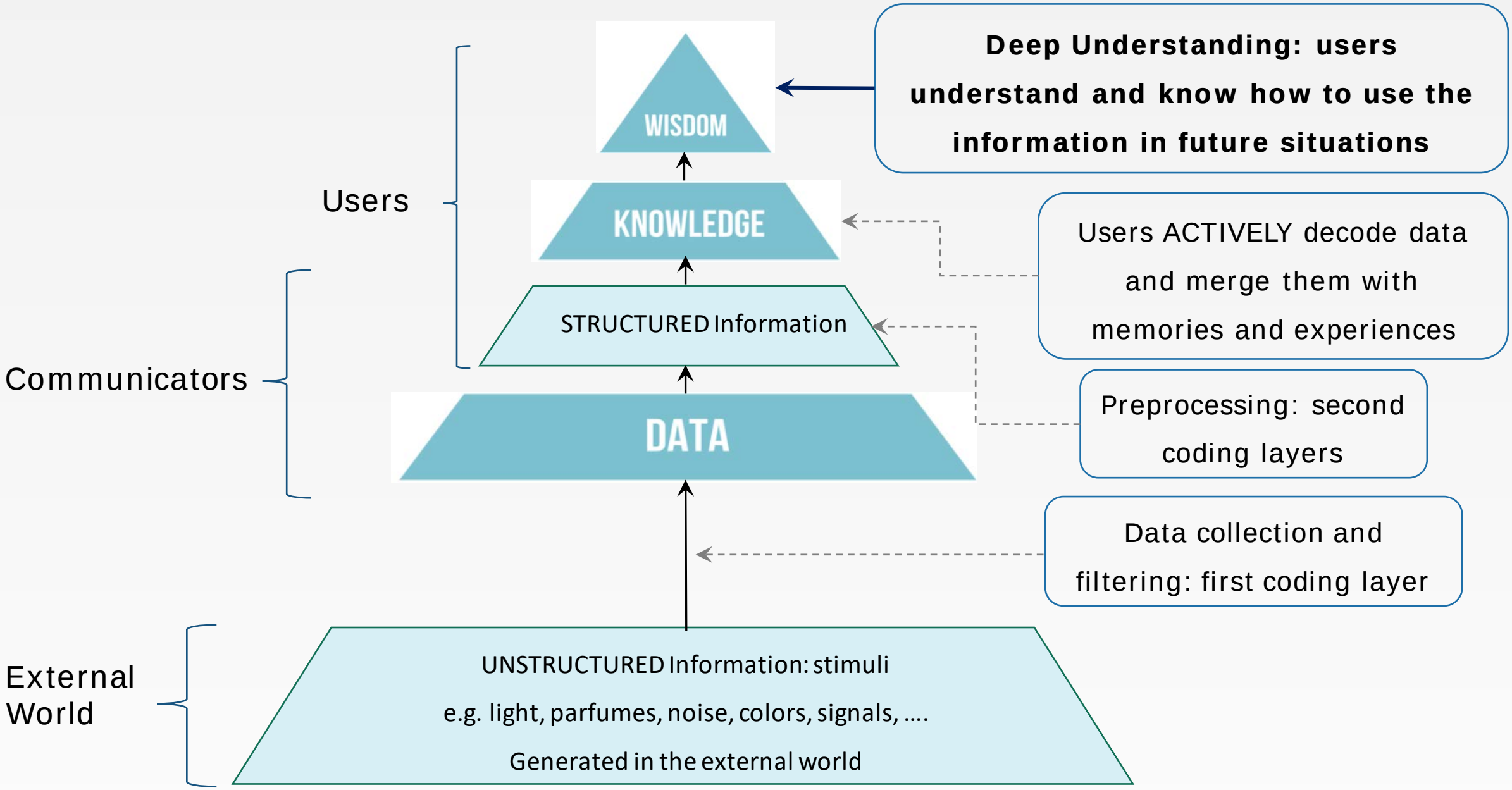
DIKW (Data, Information, Knowledge, Wisdom) Pyramid

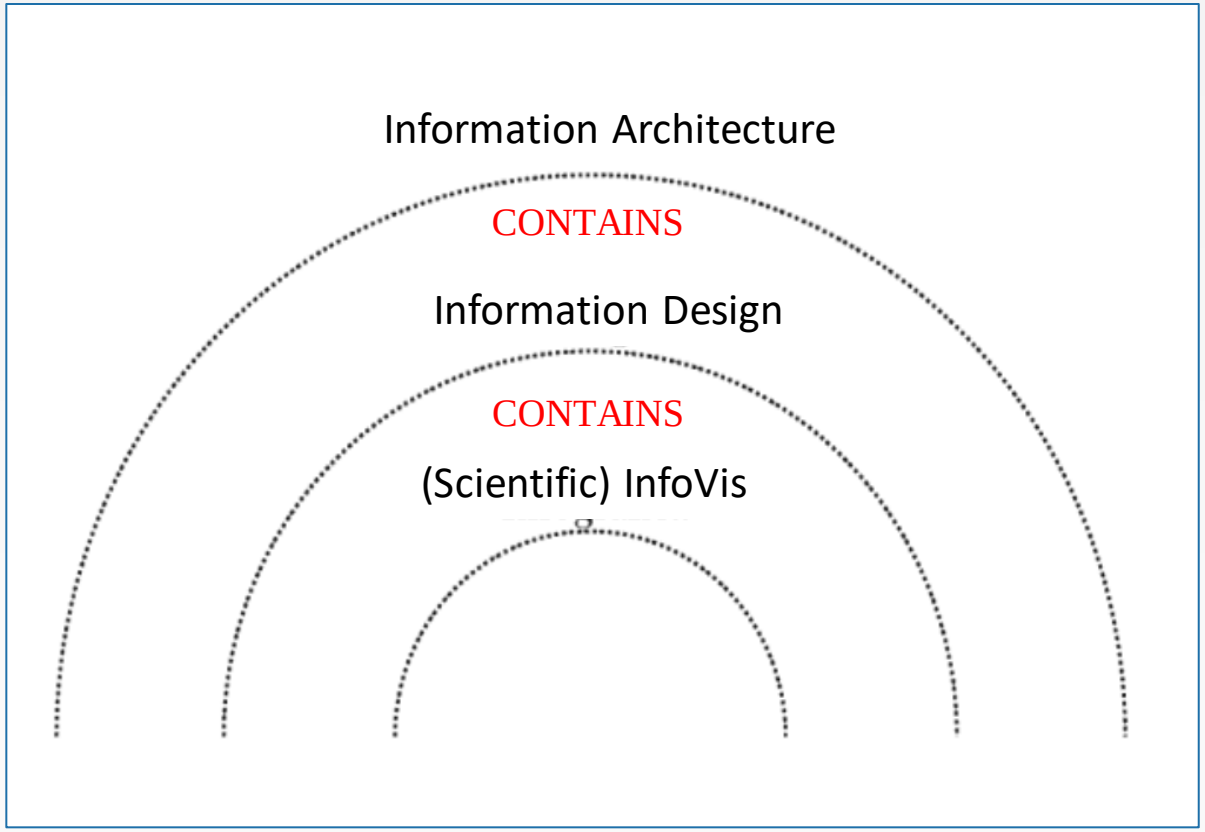



WISDOM
KNOWLEDGE
STRUCTURED Information
DATA
Communicators
Data collection and
filtering: first coding layer
External
World
UNSTRUCTURED Information: stimuli
e.g. light, parfumes, noise, colors, signals, ….
Generated in the external world


WISDOM
KNOWLEDGE
STRUCTURED Information
DATA
Communicators
Preprocessing: second coding layers
Data collection and filtering: first coding layer
UNSTRUCTURED Information: stimuli
e.g. light, parfumes, noise, colors, signals, ….
Generated in the external world
External World

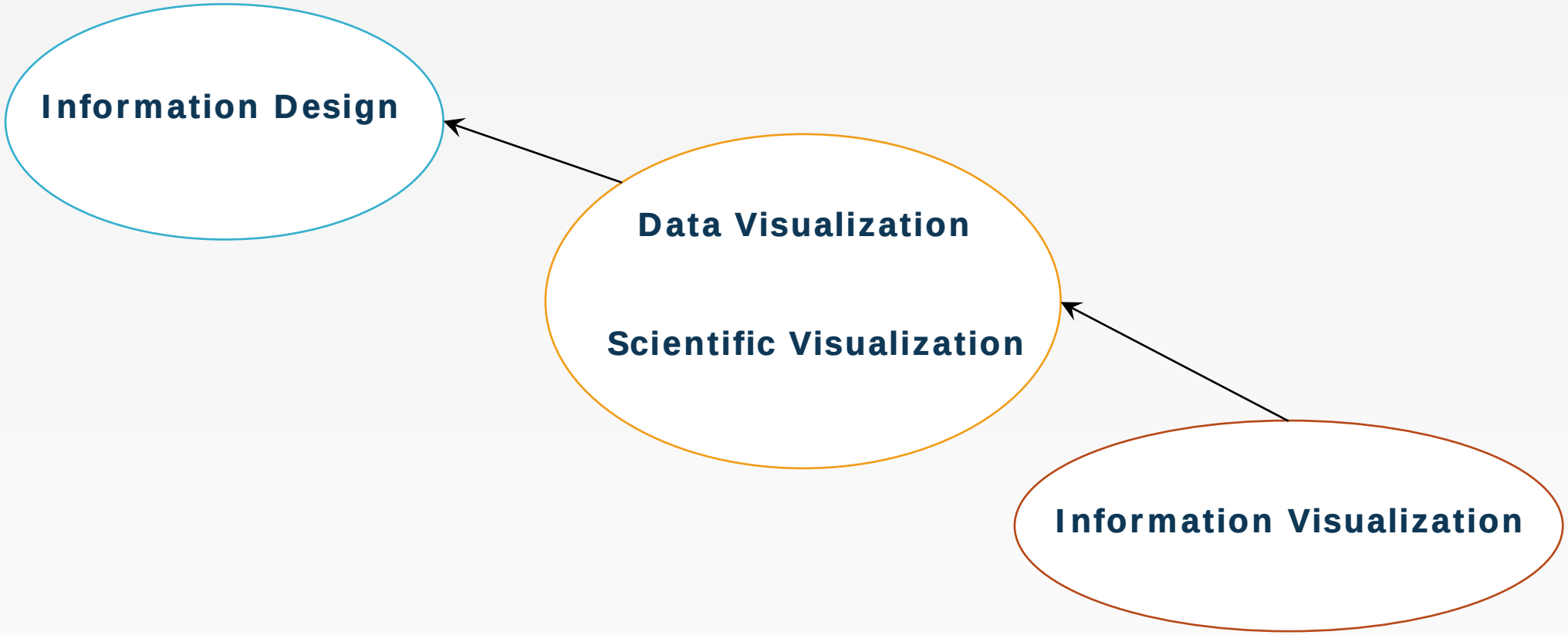

WISDOM
KNOWLEDGE
STRUCTURED Information
DATA
UNSTRUCTURED Information: stimuli
e.g. light, parfumes, noise, colors, signals, ….
Generated in the external world
Users
Communicators
External World
Users ACTIVELY decode data and merge them with memories and experiences
Preprocessing: second coding layers
Data collection and filtering: first coding layer


WISDOM
KNOWLEDGE
STRUCTURED Information
DATA
UNSTRUCTURED Information: stimuli
e.g. light, parfumes, noise, colors, signals, ….
Generated in the external world
Users
Communicators
External World
Deep Understanding: users understand and know how to use the information in future situations
Users ACTIVELY decode data and merge them with memories and experiences
Preprocessing: second coding layers
Data collection and filtering: first coding layer


Information Architecture
CONTAINS
Information Design
CONTAINS
(Scientific) InfoVis

WE HAVE

Information architecture organizes the information to make it useful.



Information Design (ID) "is the detailed planning of specific information that is to be provided to a particular audience to meet specific objectives".

Simplifies, Integrates, Filters and Selectively Emphasizes information in a way that is user-oriented.

Inherits by dataViz and InfoViz and SciViz the exploitation of psychology and physiology principles (e.g.: how users access, learn, and remember information; the impact of colors, shapes, and patterns, learning styles).

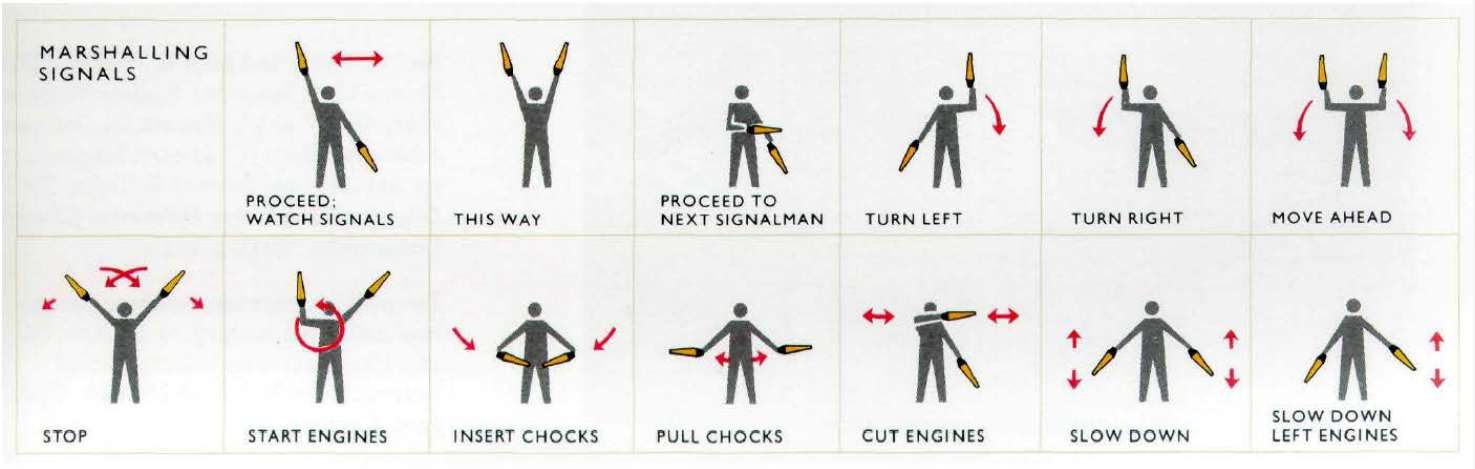
Encompasses many different responsibilities and tasks, including:

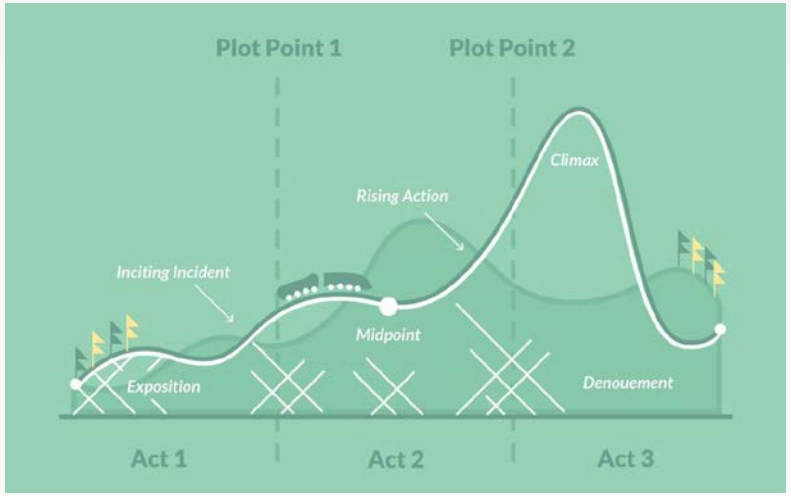
a) Analysis of users' needs and learning styles
b) Selection of the most effective layouts, colors, fonts, and graphics
c) Identification of most effective navigational infrastructures
d) Application of principles of simplification, synthesis and integration
e) Testing of readability, contrast and legibility in adverse situations and for handicapped users

(Luigi Canali De Rossi: What is Information Design)

INFORMATION VISUALIZATION

Information Visualization helps us finding the best way to visualize infos





Born in journalism, **Information graphics or infographics** are graphic visual representations of information, data, or knowledge intended to present information quickly and clearly.

(Smiciklas, Mark. The Power of Infographics: Using Pictures to Communicate and Connect with Your Audiences, 2012, ISBN 9780789749499.)

STRUCTURED infos are presented through mixtures of (many) images, (many) plots and graphs, (few) text.

Principles of perception and vision are used to transmit information and ideas.

Data visualization is the graphical display of abstract information for two purposes: sense-making (also called data analysis) and communication. Important stories live in our data and data visualization is a powerful means to discover and understand these stories, and then to present them to others. The information is abstract in that it describes things that are not physical.

Stephen Few, **Data Visualization for Human Perception**

World population growth with respect to preceding year

InfoVis & Infographic	**Unstructured data**	**Presentation**	**Exploration and Analisys**	**Structured, Scientific data**	**Data Visualization & Scientific Visualization**

Visualize = “make certain phenomena and portions of reality visible and understandable; many of these phenomena are not directly or naturally accessible to the naked eye and many are not even visual in nature” (Joan Costa)

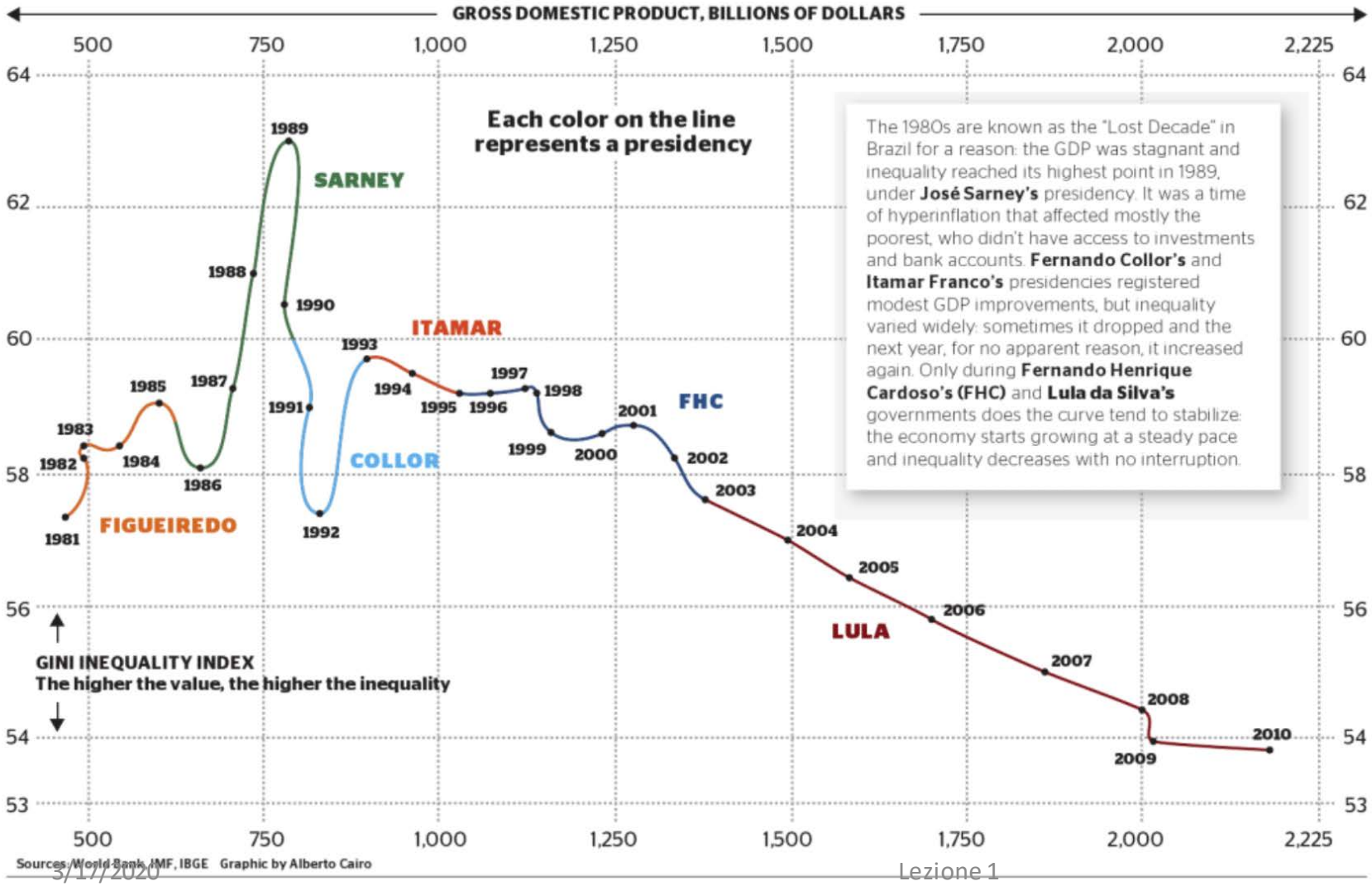
Visualize = “make certain phenomena and portions of reality visible and understandable; many of these phenomena are not directly or naturally accessible to the naked eye and many are not even visual in nature” (Joan Costa)

Visualizations may be:

Figurative Visualizations:	they display real data. e.g.: results of election per states / countries / cities, results of experiments for different algorithms
Not figurative or Abstract:	they represent informations and abstract data that can’t be measured. e.g.: user manuals for hoven, turistic maps (where turistic spots are highlighted, aerplane disaster, opinions about elections

When the Brazilian Economy Improves, Inequality Doesn't Drop

The graphic below shows the correlation between Brazilian GDP (horizontal axis) and inequality (vertical axis) between 1981 and 2010. The position of the points, each representing a year, depends on how high GDP and inequality were. You can notice, for instance, that the economy grew between 1986 and 1989 because the line tends to move to the right, but inequality also grew, as the point representing 1989 is much higher than the ones before. You can also see that, during Lula da Silva's government, the economy expanded almost as much as during the terms of the other presidents who preceded him combined.



Figurative Visualization: It represents real measurable data

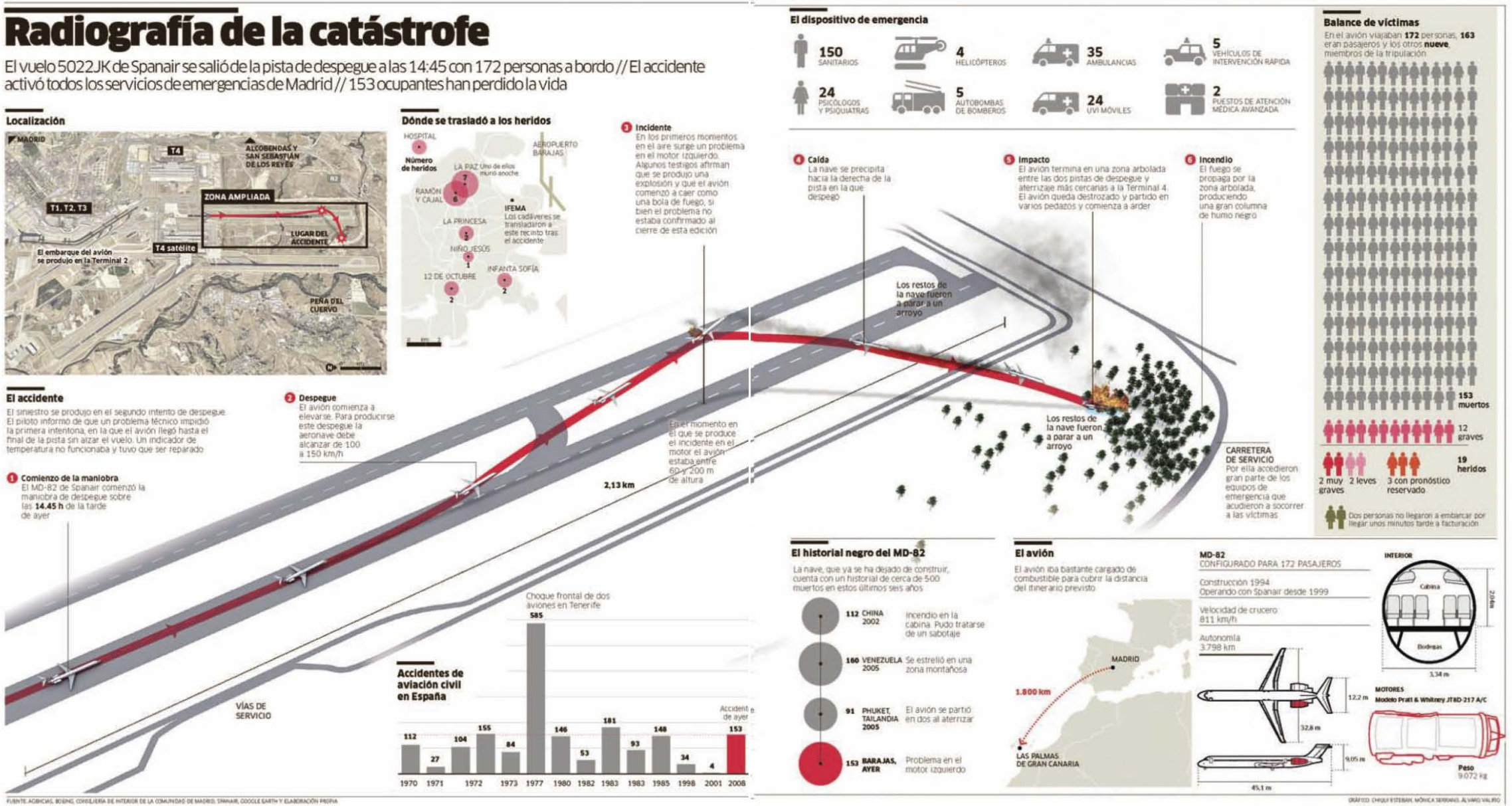


Figura 1.10 *Público* (Spagna). Incidente aereo all'aeroporto di Barajas, Madrid, 21 agosto 2008. Infografica di Chiqui Esteban, Mónica Serrano, Álvaro Valiño.

Abstract Visualization: it represents a situation

Visualizations use conventional methods that are not related to the represented data.

Art and perception are part of the design.

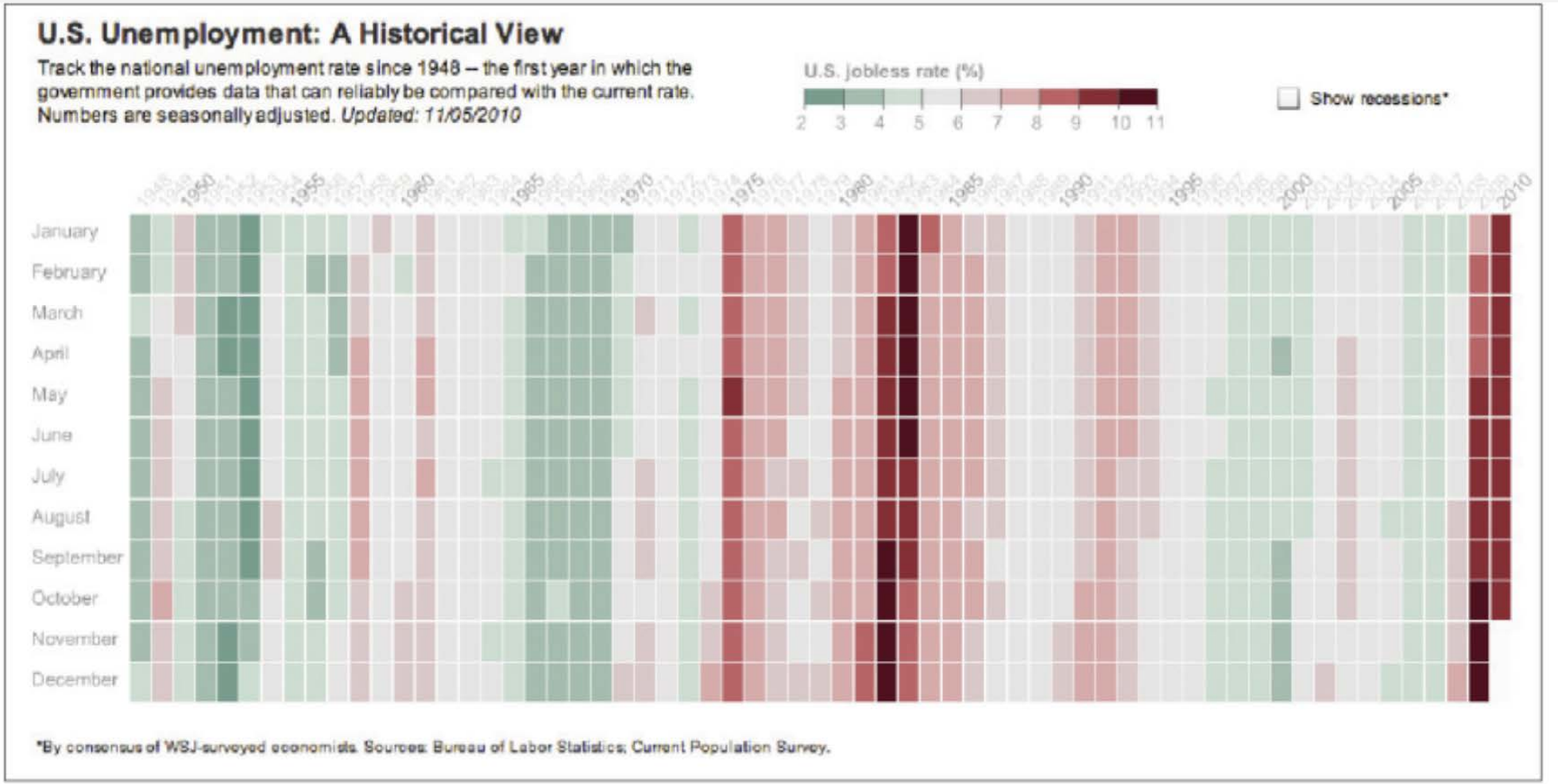


Figura 1.11 *The Wall Street Journal*. Grafico web "U.S. Unemployment: A Historical View". (Riprodotto per gentile concessione di *The Wall Street Journal*, Copyright © 2010 Dow Jones & Company, Inc. All Rights Reserved Worldwide)

Nathan Yau.

[Causes of death]

10 LEADING CAUSES OF DEATH BY AGE

Watch how they changed from 1999 through 2016.

Female | Male | 2001 | Play | Pause

	Age 1-4	5-9	10-14	15-19	20-24	25-34	35-44	45-54	55-64	65-74	75-84	85+
#1	Accidental injury	Transport accidents	Transport accidents	Transport accidents	Transport accidents	Malignant neoplasms	Malignant neoplasms	Malignant neoplasms	Malignant neoplasms	Malignant neoplasms	Malignant neoplasms	Ischaemic heart diseases
2	Transport accidents	Malignant neoplasms	Malignant neoplasms	Assault	Assault	Transport accidents	Transport accidents	Ischaemic heart diseases	Ischaemic heart diseases	Ischaemic heart diseases	Ischaemic heart diseases	Cerebrovascular diseases
3	Malignant neoplasms	Accidental injury	Accidental injury		Malignant neoplasms	Accidental injury	Accidental injury	Cerebrovascular diseases	Chronic lower respiratory diseases	Chronic lower respiratory diseases	Cerebrovascular diseases	Other forms of heart disease
4	Assault	Assault	Assault	Intentional self-harm	Intentional self-harm	Intentional self-harm	Ischaemic heart diseases	Diseases of liver	Cerebrovascular diseases	Cerebrovascular diseases	Other forms of heart disease	
5	Congenital malform. of circ. sys.	Cerebral palsy	Intentional self-harm	Accidental injury		Assault	HIV		Diabetes mellitus	Other forms of heart disease	Chronic lower respiratory diseases	Other degen. diseases of nerv. sys.
6	Ill-defined and unknown causes	Other forms of	Other forms of heart disease	Ill-defined and unknown causes	Ill-defined and unknown causes		Intentional self-harm	Diabetes mellitus	Other forms of heart disease	Diabetes mellitus	Diabetes mellitus	Influenza and pneumonia
7	Metabolic disorders	Congenital malform. of circ. sys.	Cerebral palsy			Other forms of heart disease	Diseases of liver	Transport accidents	Diseases of liver	Renal failure	Other degen. diseases of nerv. sys.	Symptomatic mental disorders
8	Other forms of	Congenital malform. of nervous sys.		Metabolic disorders		Ill-defined and unknown causes	Cerebrovascular diseases	Accidental injury	Hypertensive diseases	Hypertensive diseases	Influenza and pneumonia	Chronic lower respiratory diseases
9	Congenital malform. of nervous sys.	disorders	Metabolic	Cerebral palsy	Systemic connective tissue	Cerebrovascular diseases		diseases	Other bacterial diseases	Influenza and pneumonia	Hypertensive diseases	Hypertensive diseases
10	Other bacterial diseases	Neoplasms of uncertain behaviour Other disorders of the nervous	Ill-defined and unknown causes Chronic lower respiratory diseases	Other bacterial diseases	Cerebrovascular diseases	Diabetes mellitus	Ill-defined and unknown causes	Intentional self-harm	Renal failure		Symptomatic mental disorders	Diseases of arteries and capillaries

Per 100,000 people in age group

10 50 100 500 1000

At the wikipedia page we can see many different ways to see the U.S. unemployment during different periods and its correlation with other factors, e.g. carceration, changes in net employment, presidents, ….

Compare the seen page with the one describing the Italian unemployment rate

or the youth italian unemployment rate

WHAT DO YO NOTICE WHEN COMPARING THE PAGES?

WHY ARE DATAVIZ/SCIVIZ SO USEFUL

Let's watch a video:

https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen

INFORMATION VISUALIZATION
DATA and SCIENTIFIC VISUALIZATION (DATAVIZ - SCIVIZ)
definitions

The use of computer-supported, interactive, visual representations of

abstract data

or

physical data

to amplify cognition. Cognition is the acquisition or use of knowledge.

(Card, S. and Mackinlay, J. and Shneiderman, B., Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers, 1999)

DATA and SCIENTIFIC VISUALIZATION (DATAVIZ - SCIVIZ): ONE definition

The mean

The use of computer-supported, interactive, visual representations of

abstract data

or

physical data

The aim

to amplify cognition. Cognition is the acquisition or use of knowledge.

(Card, S. and Mackinlay, J. and Shneiderman, B., Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers, 1999)

DATA and SCIENTIFIC VISUALIZATION (DATAVIZ - SCIVIZ): ONE definition

The use of computer-supported, interactive, visual representations of

abstract data — InfoViZ

or

physical data — DATAVIZ/SCIVIZ

The aim to amplify cognition. Cognition is the acquisition or use of knowledge.

(Card, S. and Mackinlay, J. and Shneiderman, B., Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers, 1999)

“Excellence in statistical graphics consists of complex ideas communicated with clarity, precision and efficiency.”

Edward Tufte, “Envisioning Information”

Edward Tufte, “ The Visual Display of Quantitative Information”

Principles of Infographics (and of Data Visualization)

Infographics, (Scientific/Data Visualizations) should:

- show the data
- induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production, or something else
- avoid distorting what the data has to say
- present many numbers in a small space
- make large data sets coherent
- encourage the eye to compare different pieces of data
- reveal the data at several levels of detail, from a broad overview to the fine structure
- serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- be closely integrated with the statistical and verbal descriptions of a data set.

Graphics *reveal* data. Indeed graphics can be more precise and revealing than conventional statistical computations.

EXAMPLE

Matt Ridley, “The rational Optimist: how prosperity evolves” (2010)

Forecasts:

- the growing fertility of women in poor regions is the reason why the earth has to maintain 7 billion people with forecasts of 9 billion in twenty years and even more in the distant future.
- in developed” countries, fertility rates are lower than 2.1.
- 2.1 = “substitution level” : if the rate drops well below 2.1, the population will decrease over time, if it rises too much, the population will be very young with catastrophic consequences – e.g. high rates of violence and crime.

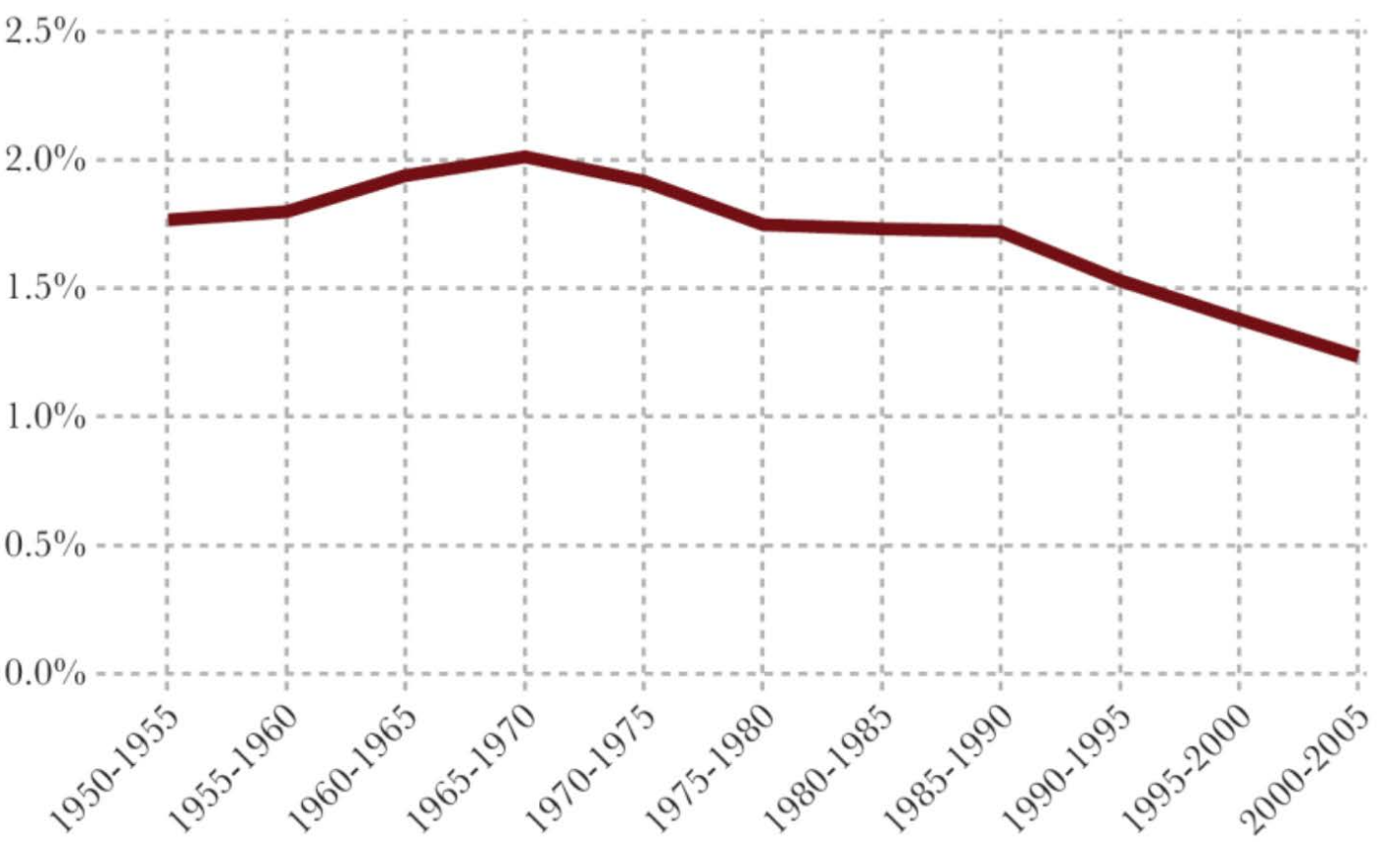
Ridley says that:

- On average, fertility rate in poor countries in the years prior to 2010 was decreasing
- On average fertility in rich countries is growing
- If the trend continues like this, global fertility will stand at 2.1 and the global population will be 9 billion, so DO NOT PANIC!

Is this true?

Ridley's book cites data taken from United Nations (UN) and World Bank databases, but in the book the only trend seen is that of world population growth compared to the previous year.

No other graph or picture.



The picture in Ridley's book

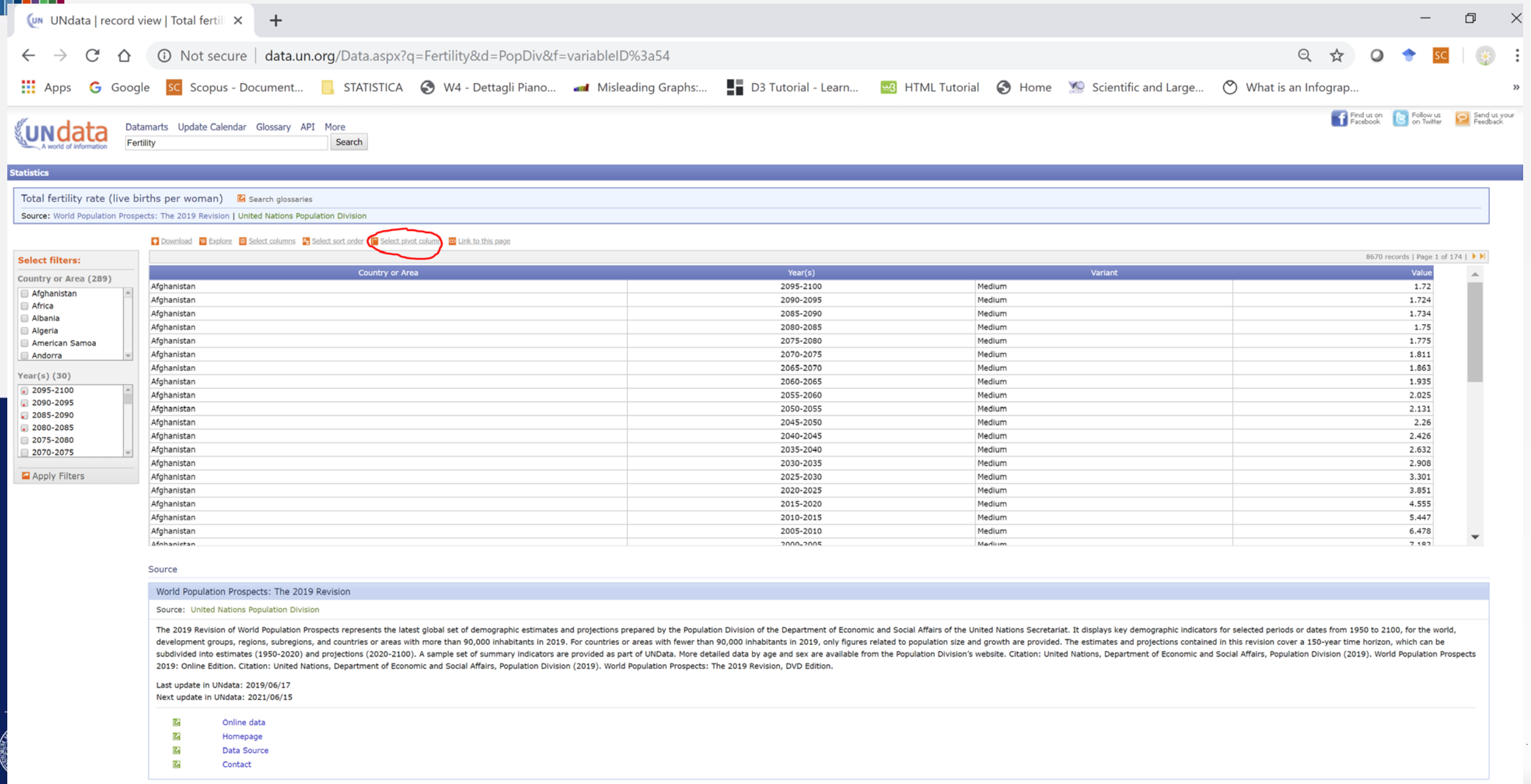
Mean world fertility rate

(children per woman)

http://data.un.org/ (search for UN data: found it while searching for UN website).

It's a web database with any sort of data.

Let's download the nations' fertility reates


UNdata | record view | Total fertil
Not secure | data.un.org/Data.aspx?q=Fertility&d=PopDiv&f=variableID%3a54
Apps
Google
Scopus - Document...
STATISTICA
W4 - Dettagli Piano...
Misleading Graphs:...
D3 Tutorial - Learn...
HTML Tutorial
Home
Scientific and Large...
What is an Infograp...
UNdata
A world of information
Datamarts
Update Calendar
Glossary
API
More
Fertility
Search
Find us on Facebook
Follow us on Twitter
Send us your Feedback
Statistics
Total fertility rate (live births per woman)
Search glossaries
Source: World Population Prospects: The 2019 Revision | United Nations Population Division
Download
Explore
Select columns
Select sort order
Select pivot column
Link to this page
8670 records | Page 1 of 174
Select filters:
Country or Area (289)
Afghanistan
Africa
Albania
Algeria
American Samoa
Andorra
Year(s) (30)
2095-2100
2090-2095
2085-2090
2080-2085
2075-2080
2070-2075
Apply Filters
Country or Area | Year(s) | Variant | Value
Afghanistan | 2095-2100 | Medium | 1.72
Afghanistan | 2090-2095 | Medium | 1.724
Afghanistan | 2085-2090 | Medium | 1.734
Afghanistan | 2080-2085 | Medium | 1.75
Afghanistan | 2075-2080 | Medium | 1.775
Afghanistan | 2070-2075 | Medium | 1.811
Afghanistan | 2065-2070 | Medium | 1.863
Afghanistan | 2060-2065 | Medium | 1.935
Afghanistan | 2055-2060 | Medium | 2.025
Afghanistan | 2050-2055 | Medium | 2.131
Afghanistan | 2045-2050 | Medium | 2.26
Afghanistan | 2040-2045 | Medium | 2.426
Afghanistan | 2035-2040 | Medium | 2.632
Afghanistan | 2030-2035 | Medium | 2.908
Afghanistan | 2025-2030 | Medium | 3.301
Afghanistan | 2020-2025 | Medium | 3.851
Afghanistan | 2015-2020 | Medium | 4.555
Afghanistan | 2010-2015 | Medium | 5.447
Afghanistan | 2005-2010 | Medium | 6.478
Source
World Population Prospects: The 2019 Revision
Source: United Nations Population Division
The 2019 Revision of World Population Prospects represents the latest global set of demographic estimates and projections prepared by the Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat. It displays key demographic indicators for selected periods or dates from 1950 to 2100, for the world, development groups, regions, subregions, and countries or areas with more than 90,000 inhabitants in 2019. For countries or areas with fewer than 90,000 inhabitants in 2019, only figures related to population size and growth are provided. The estimates and projections contained in this revision cover a 150-year time horizon, which can be subdivided into estimates (1950-2020) and projections (2020-2100). A sample set of summary indicators are provided as part of UNdata. More detailed data by age and sex are available from the Population Division's website. Citation: United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019: Online Edition. Citation: United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects: The 2019 Revision, DVD Edition.
Last update in UNdata: 2019/06/17
Next update in UNdata: 2021/06/15
Online data
Homepage
Data Source
Contact

UNdata | record view | Total fertil

Not secure | data.un.org/Data.aspx?q=Fertility&d=PopDiv&f=variableID%3a54

Apps | Google | Scopus - Document... | STATISTICA | W4 - Dettagli Piano... | Misleading Graphs:... | D3 Tutorial - Learn... | HTML Tutorial | Home | Scientific and Large... | What is an Infograp...

UNdata
A world of information

Datamarts | Update Calendar | Glossary | API | More

Fertility | Search

Find us on Facebook | Follow us on Twitter | Send Feedb

Statistics

Total fertility rate (live births per woman) | Search glossaries

Source: World Population Prospects: The 2019 Revision | United Nations Population Division

Download | Select pivot column | Link to this page

Select download format:

Structured

XML

Value separated

Comma | Semicolon | Pipe

289 records | Page 1 of 6

Select filters:

Country or Area (289)

- Afghanistan
- Africa
- Albania
- Algeria
- American Samoa
- Andorra

Year(s) (30)

- 2035-2040
- 2030-2035
- 2025-2030
- 2020-2025
- 2015-2020

Apply Filters

Coun		1960-1965	1965-1970	1970-1975	1975-1980	1980-1985	1985-1990	1990-1995	1995-2000	2000-2005	2005-2010	2010-2015	2015-2020	2020-2025	2025-2030	2030-2035	2035-2040	2040-2045	2045-2050	2050-2055	2055-2060	2060-2065	2065-2070	2070-2075	2075-2080	2080-2085	2085-2090	2090-2095	2095-2100
Afghanistan		7.45	7.45	7.45	7.45	7.45	7.469	7.482	7.654	7.182	6.478	5.447	4.555	3.851	3.301	2.908	2.632	2.426	2.26	2.131	2.025	1.935	1.863	1.811	1.775	1.75	1.734	1.724	1.72
Africa		6.699	6.706	6.703	6.64	6.501	6.187	5.724	5.351	5.077	4.9	4.73	4.438	4.155	3.893	3.651	3.434	3.242	3.071	2.922	2.79	2.675	2.57	2.478	2.397	2.322	2.255	2.195	2.141
Albania		6.23	5.259	4.6	3.9	3.409	3.15	2.786	2.384	1.946	1.64	1.714	1.62	1.545	1.489	1.461	1.465	1.486	1.508	1.534	1.558	1.582	1.603	1.619	1.636	1.655	1.665	1.678	1.688
Algeria		7.648	7.648	7.572	7.175	6.315	5.302	4.12	2.885	2.384	2.724	2.96	3.05	2.789	2.59	2.433	2.312	2.213	2.131	2.064	2.013	1.976	1.942	1.916	1.898	1.881	1.865	1.852	1.843
American S																													
Andorra																													
Angola	Medium	6.5	6.9	7.3	7.5	7.456	7.456	7.4	7.1	6.75	6.55	6.35	6	5.55	5.213	4.904	4.616	4.342	4.106	3.882	3.684	3.508	3.351	3.204	3.077	2.958	2.846	2.756	2.667	2.577
Anguilla	Medium																												
Antigua and Barbuda	Medium	4.5	4.3	4	3.26	2.24	2.14	2.07	2.09	2.2	2.16	2	2	2	1.95	1.912	1.874	1.847	1.82	1.805	1.79	1.779	1.775	1.771	1.765	1.764	1.76	1.758	1.759	1.761
Argentina	Medium	3.154	3.127	3.09	3.05	3.15	3.4	3.15	3.053	2.914	2.63	2.48	2.37	2.33	2.268	2.199	2.138	2.076	2.025	1.974	1.933	1.899	1.87	1.849	1.831	1.818	1.806	1.798	1.792	1.783	1.778
Armenia	Medium	4.494	4.9	4.453	3.447	3.037	2.6	2.5	2.6	2.38	1.75	1.65	1.72	1.72	1.756	1.758	1.76	1.764	1.765	1.77	1.77	1.772	1.774	1.774	1.776	1.775	1.776	1.777	1.777	1.778	1.779	1.778
Aruba	Medium	5.65	5.15	4.399	3.301	2.651	2.45	2.358	2.3	2.174	1.953	1.816	1.76	1.8	1.9	1.884	1.868	1.859	1.849	1.844	1.839	1.836	1.83	1.827	1.825	1.823	1.821	1.821	1.818	1.818	1.817
Asia	Medium	5.831	5.591	5.797	5.745	5.056	4.097	3.689	3.497	2.896	2.607	2.447	2.328	2.21	2.152	2.093	2.045	1.997	1.945	1.905	1.876	1.856	1.834	1.815	1.796	1.782	1.775	1.768	1.764	1.76	1.758
Australia	Medium	3.18	3.406	3.274	2.871	2.535	1.989	1.91	1.859	1.863	1.787	1.774	1.952	1.885	1.832	1.784	1.759	1.746	1.731	1.723	1.719	1.72	1.723	1.722	1.724	1.727	1.73	1.735	1.735	1.739	1.741
Australia/New Zealand	Medium	3.274	3.529	3.381	2.957	2.59	2.023	1.922	1.889	1.898	1.815	1.804	1.985	1.91	1.844	1.796	1.768	1.751	1.735	1.726	1.722	1.722	1.724	1.724	1.724	1.727	1.73	1.734	1.735	1.739	1.741
Austria	Medium	2.105	2.565	2.778	2.569	2.037	1.651	1.596	1.448	1.484	1.388	1.38	1.399	1.449	1.529	1.571	1.603	1.631	1.652	1.67	1.684	1.695	1.707	1.716	1.721	1.728	1.735	1.737	1.74	1.741	1.744
Azerbaijan	Medium	5.2	5.6	6	5.4	4.6	3.8	3.3	3.2	2.9	2.25	1.9	1.83	2.09	2.083	1.993	1.918	1.864	1.821	1.79	1.768	1.754	1.743	1.735	1.734	1.736	1.733	1.734	1.736	1.737	1.741
Bahamas	Medium	4.05	4.31	4.5	3.58	3.54	2.95	3.05	2.65	2.64	2.328	1.87	1.91	1.81	1.76	1.709	1.68	1.663	1.658	1.656	1.66	1.663	1.67	1.673	1.68	1.689	1.693	1.701	1.705	1.711	1.718
Bahrain	Medium	6.97	7.17	6.97	5.95	5.23	4.63	4.08	3.4	2.95	2.65	2.25	2.12	1.998	1.891	1.799	1.723	1.67	1.637	1.619	1.612	1.616	1.621	1.629	1.64	1.651	1.662	1.672	1.682	1.694

Source

World Population Prospects: The 2019 Revision

Source: United Nations Population Division

The 2019 Revision of World Population Prospects represents the latest global set of demographic estimates and projections prepared by the Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat. It displays key demographic indicators for selected periods or dates from 1950 to 2100, for the world, development groups, regions, subregions, and countries or areas with more than 90,000 inhabitants in 2019. For countries or areas with fewer than 90,000 inhabitants in 2019, only figures related to population size and growth are provided. The estimates and projections contained in this revision cover a 150-year time horizon, which can be subdivided into estimates (1950-2020) and projections (2020-2100). A sample set of summary indicators are provided as part of UNData. More detailed data by age and sex are available from the Population Division's website. Citation: United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019: Online Edition. Citation: United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects: The 2019 Revision, DVD Edition.

Last update in UNdata: 2019/06/17
Next update in UNdata: 2021/06/15

- Online data
- Homepage
- Data Source
- Contact

I downloaded this file (.csv)

I cleaned it to have just nations with all numbers (there was some nations with no numbers….)

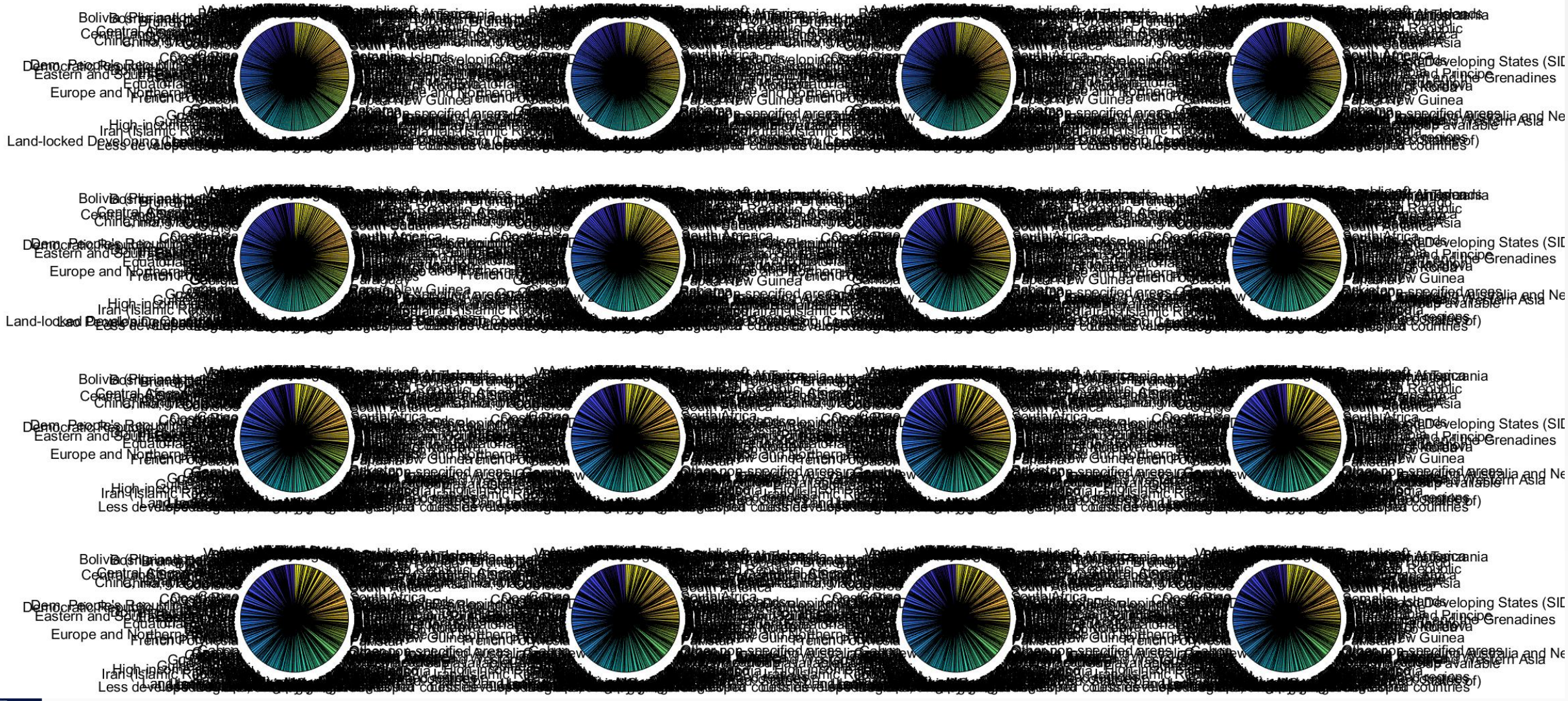
This is the cleaned file

Let’s try to make some plot….

Which one???

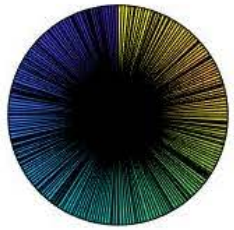
5 minutes to think….

What about pie chart?

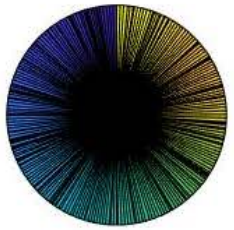


mmm… let’s remove legends

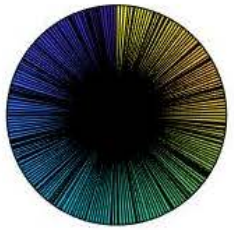
1950-1955



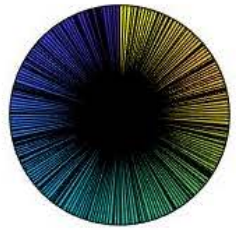
1955-1960



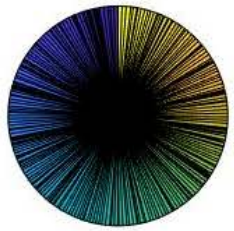
1960-1965



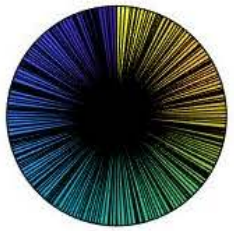
1965-1970



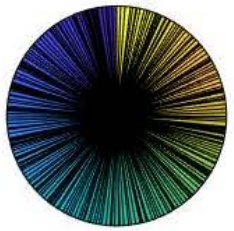
1970-1975



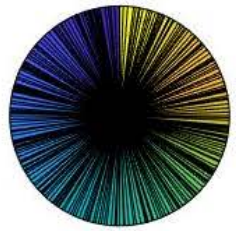
1975-1980



1980-1985

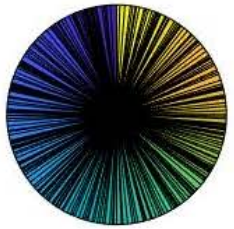


1985-1990

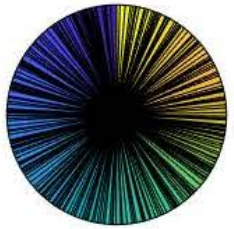


1990-1995

1995-2000

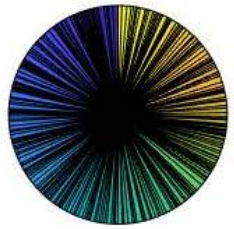


2000-2005

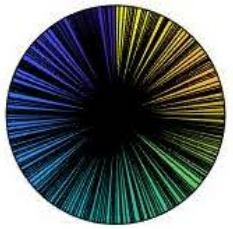


2005-2010

2010-2015

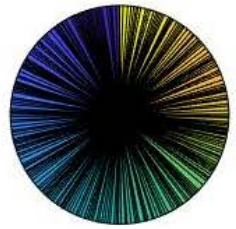


2015-2020



2020-2025

2025-2030

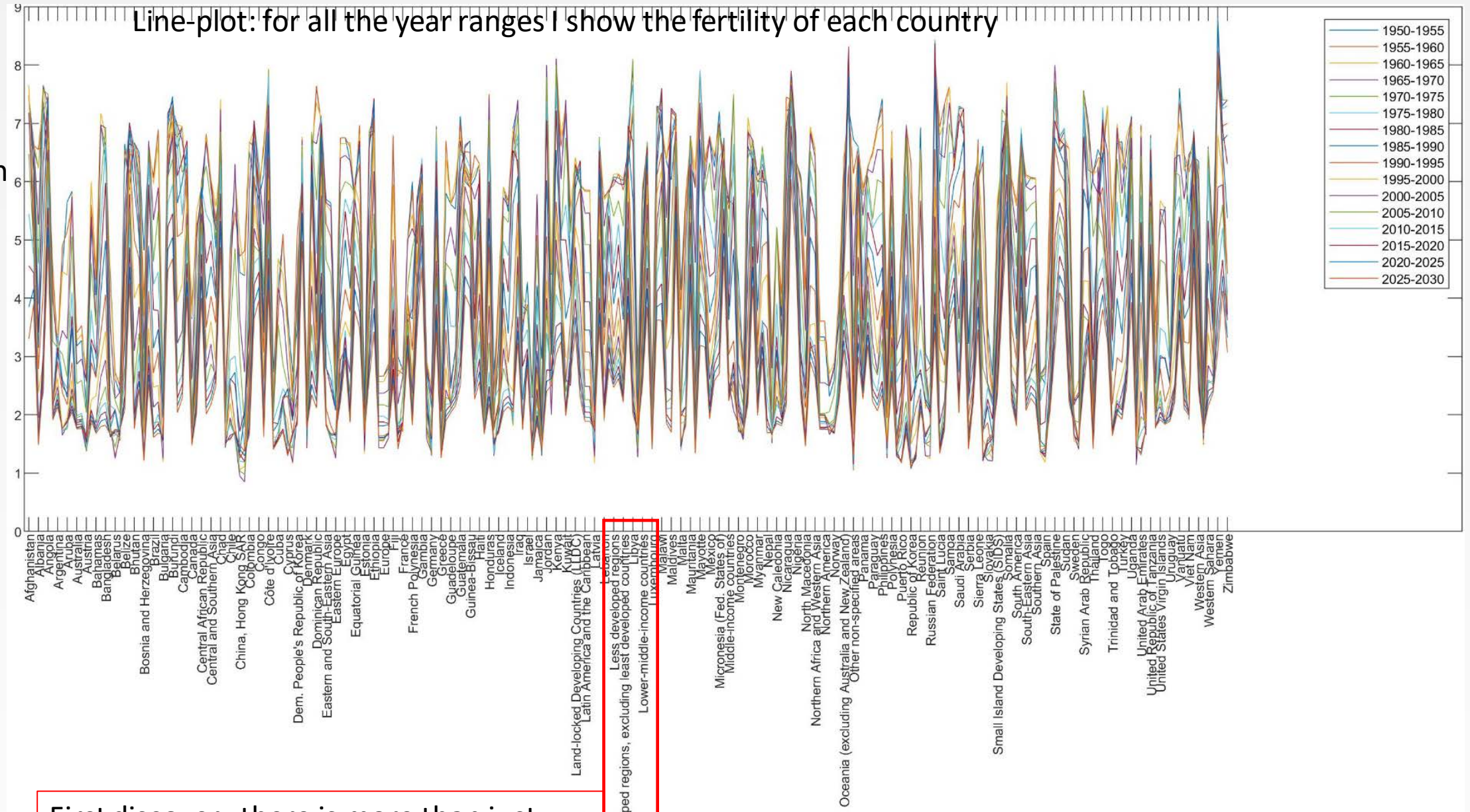


Let’s try another plot ?

……

Line-plot: for all the year ranges I show the fertility of each country

Children per woman

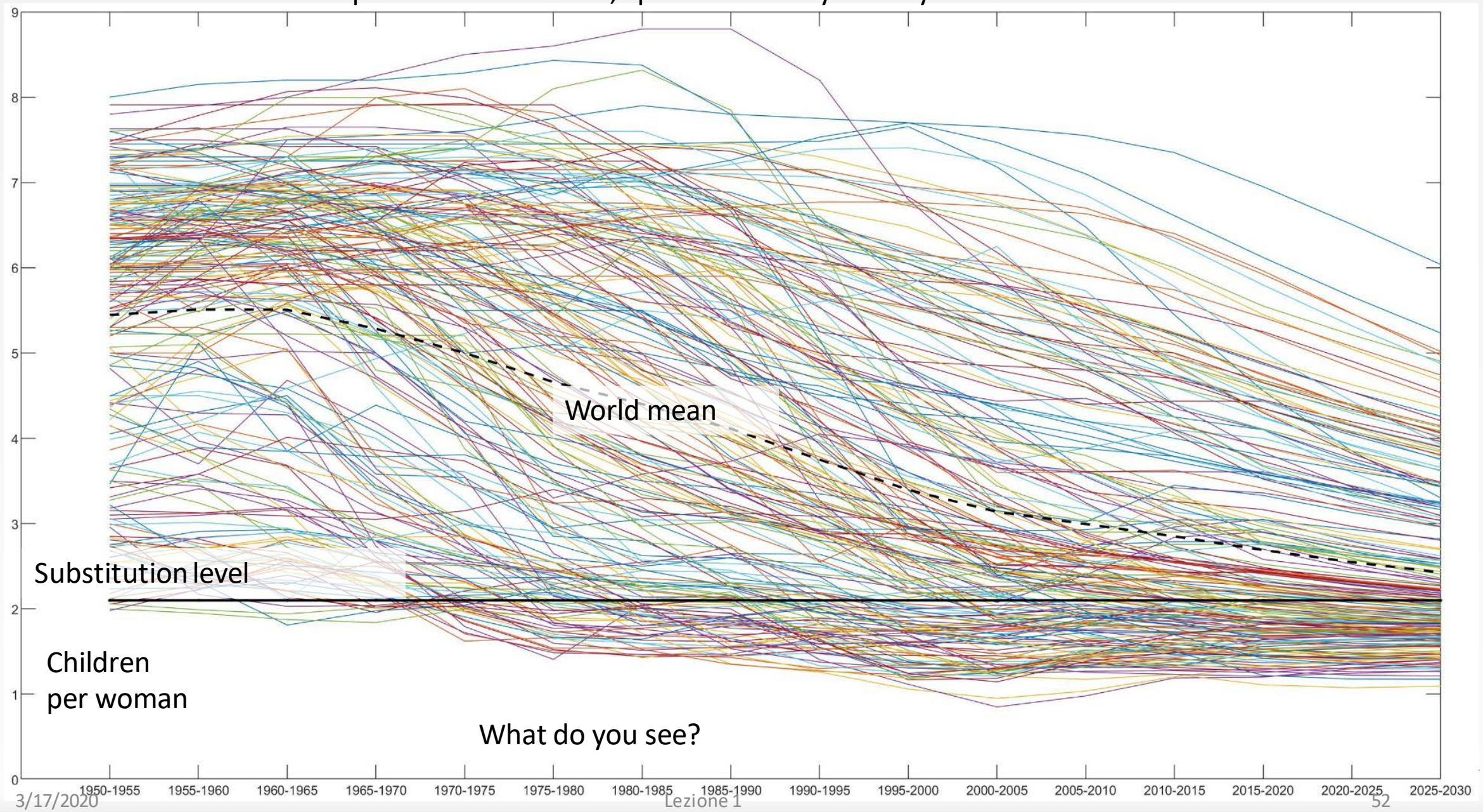


First discovery there is more than just nations

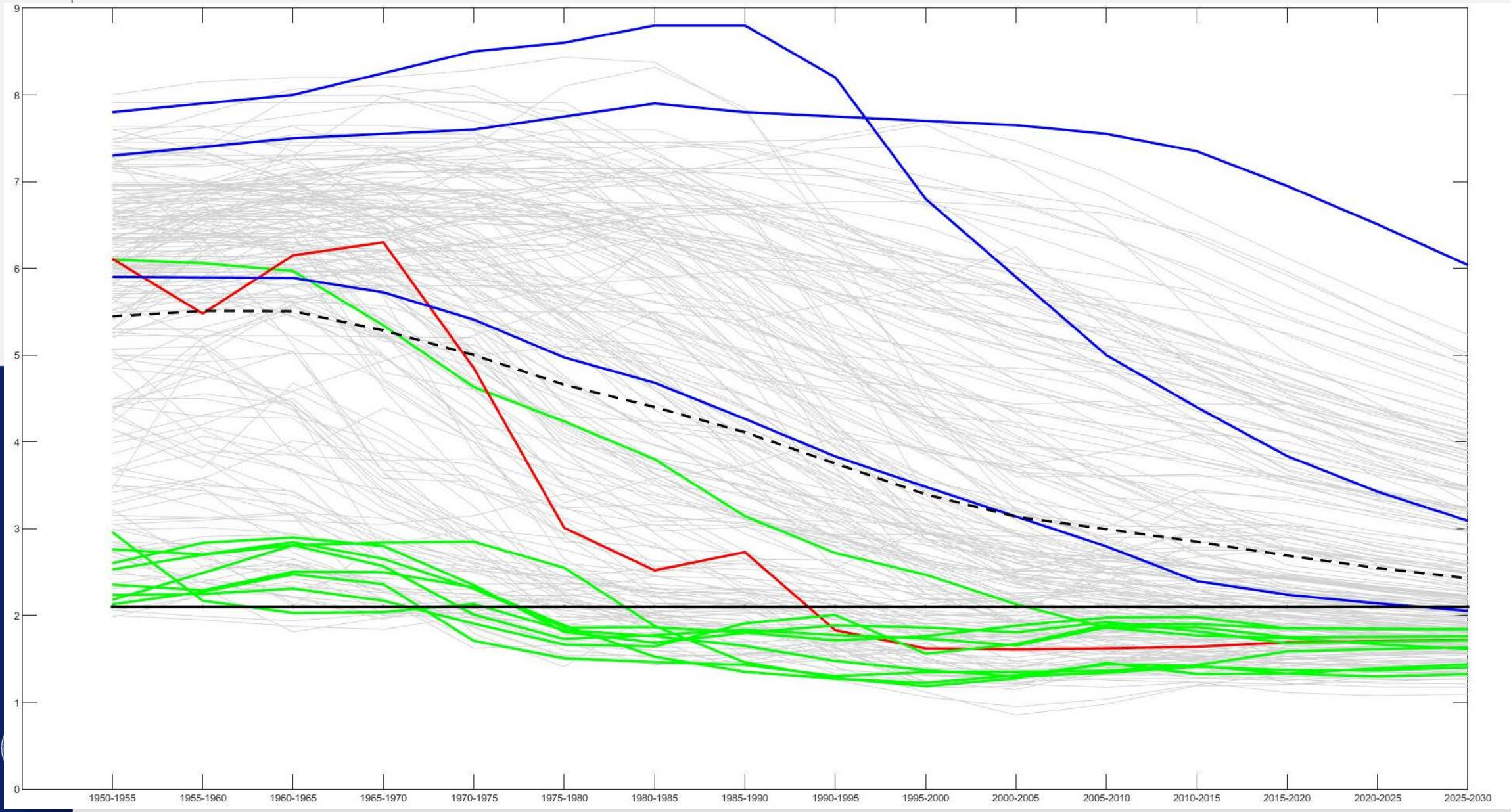
Let's clean again to keep only the nations. Here

NOW?? WHICH PLOT IS BEST??

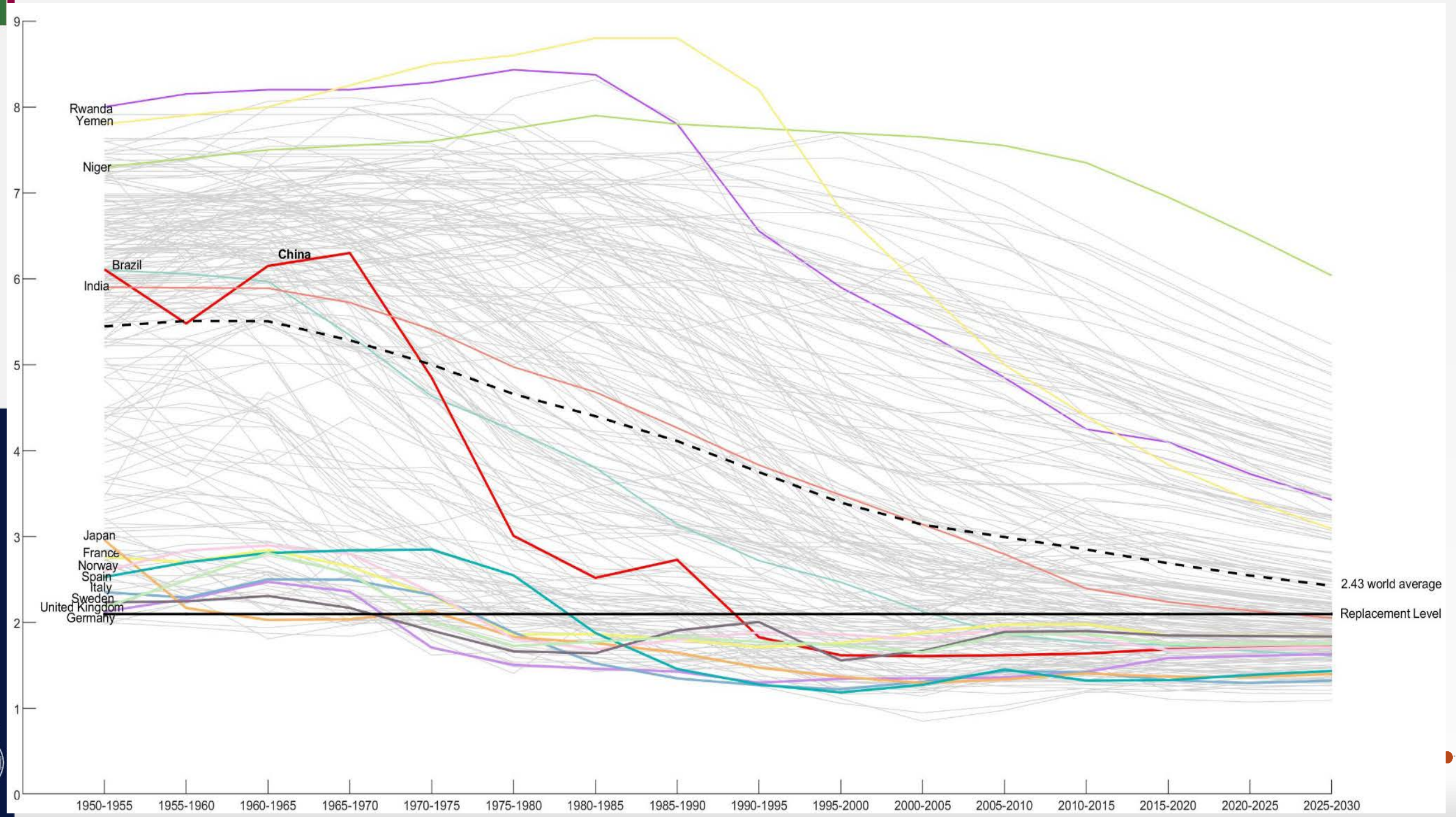
Let’s reverse the plot. For each nation, I plot the fertility in the years



What if I only consider European countries and the poorest/biggest countries?

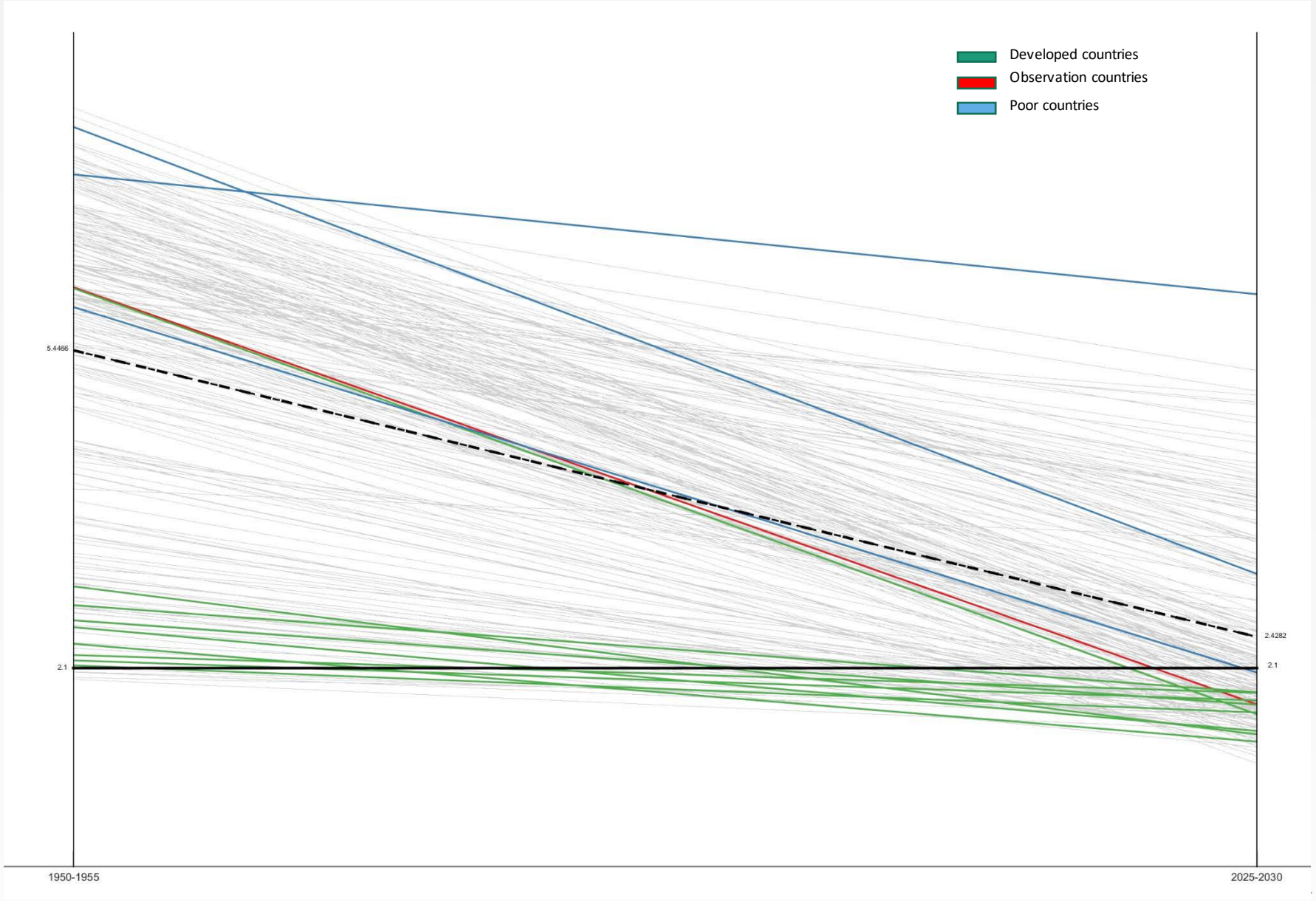


Per Country Fertility Rate (Children per woman) trend from 1950 to 2030



Now, what if we want to enforsize important details?

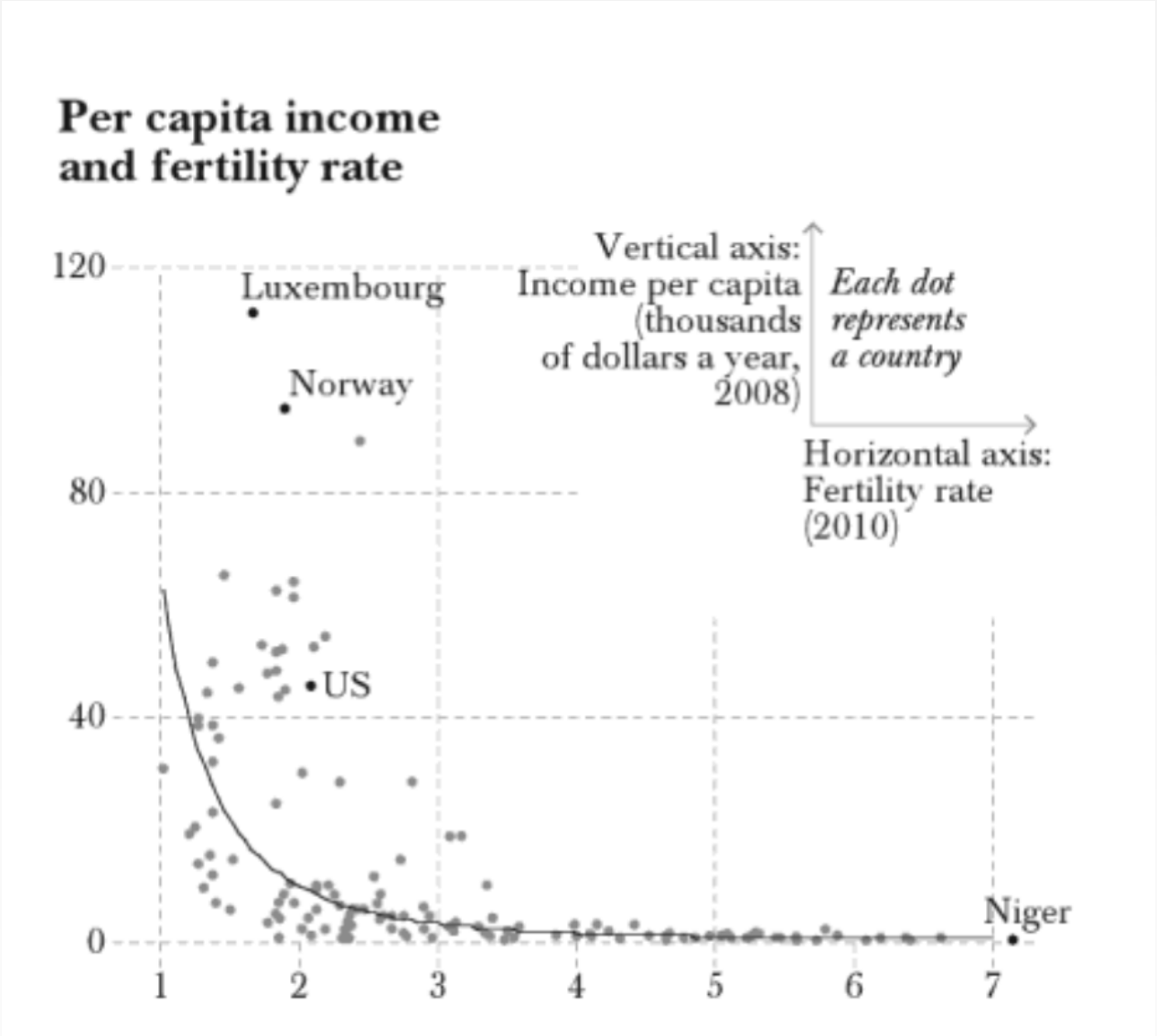
we are only interested in what will happen in the future (2030)



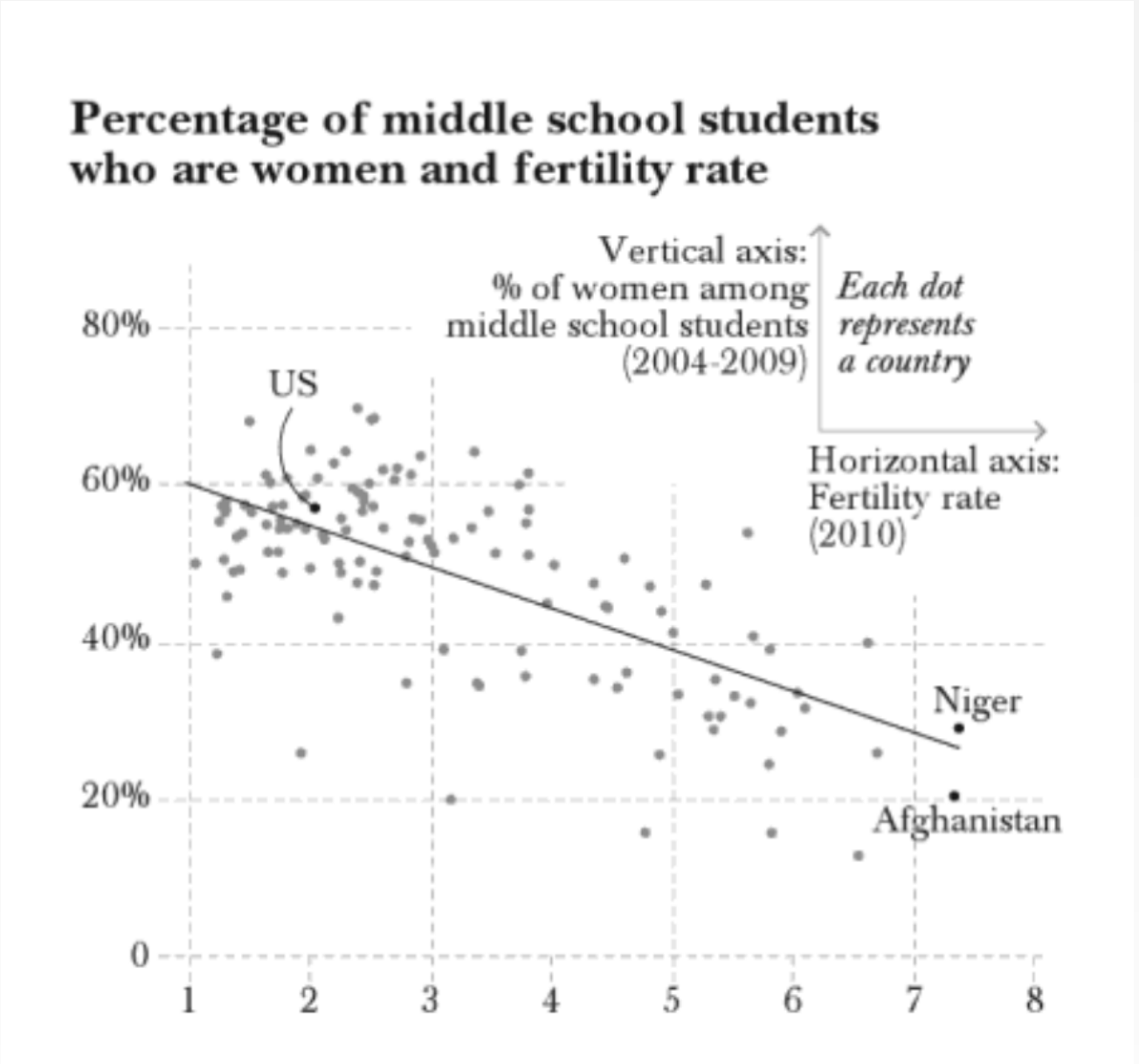
Simpler but effective.

We only want to show the trend

Other plots taken by working on UN data



Number of children decrease when women go to school



Example: Anscombe's quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

APRIAMO MATLAB ED ANALIZZIAMO I DATI:

Command Window

New to MATLAB? See resources for Getting Started.

```
data =

  11×8 table

    Ax      Ay      Bx      By      Cx      Cy      Dx      Dy
    __    _____    __    _____    __    _____    __    _____

    10     8.04    10     9.14    10     7.46     8     6.58
     8     6.95     8     8.14     8     6.77     8     5.76
    13     7.58    13     8.74    13    12.74     8     7.71
     9     8.81     9     8.77     9     7.11     8     8.84
    11     8.33    11     9.26    11     7.81     8     8.47
    14     9.96    14      8.1    14     8.84     8     7.04
     6     7.24     6     6.13     6     6.08     8     5.25
     4     4.26     4      3.1     4     5.39    19     12.5
    12    10.84    12     9.13    12     8.15     8     5.56
     7     4.82     7     7.26     7     6.42     8     7.91
     5     5.68     5     4.74     5     5.73     8     6.89

    'Ax'    'Ay'    'Bx'    'By'    'Cx'    'Cy'    'Dx'    'Dy'

means: 9      7.5009           9      7.5009           9        7.5           9      7.5009
stds: 3.3166      2.0316      3.3166      2.0317      3.3166      2.0304      3.3166      2.0306
corr Ax-Ay = 0.82
corr Bx-By = 0.82
corr Cx-Cy = 0.82
corr Dx-Dy = 0.82

fx >>
```

Medie e deviazioni standard uguali

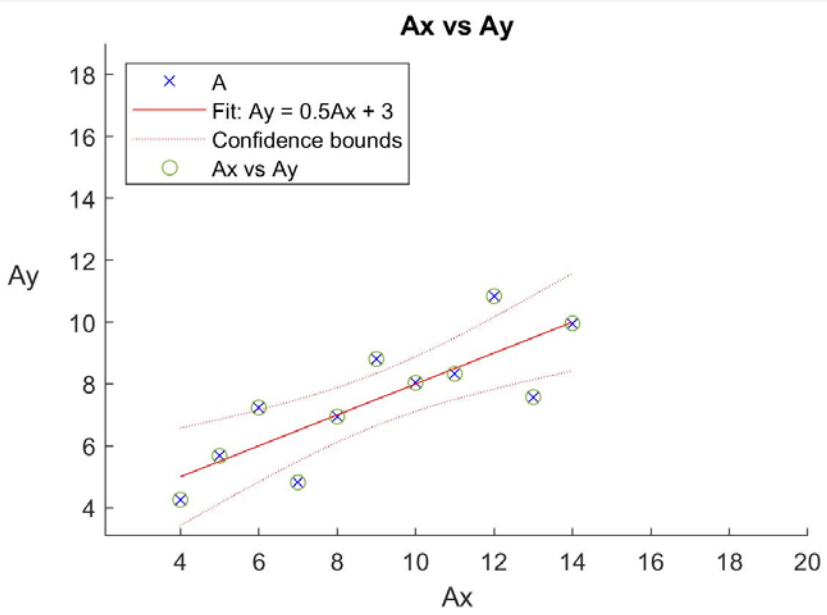
Editor - AnalizeAnscombe.m | Variables - cefs

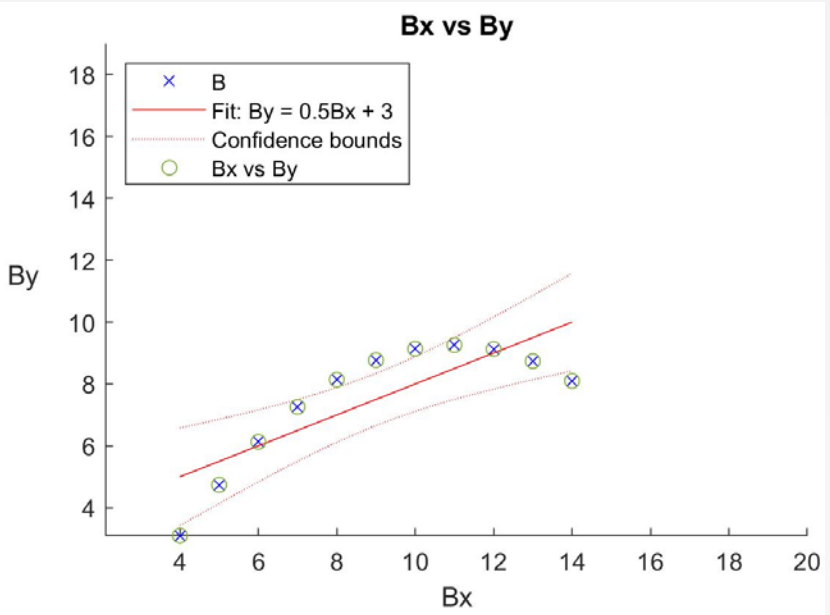
Command Window

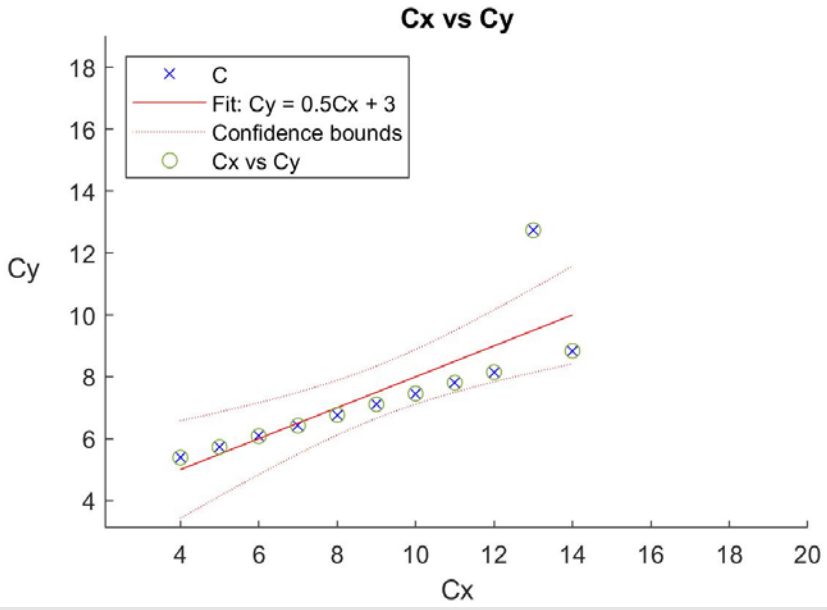
New to MATLAB? See resources for Getting Started.

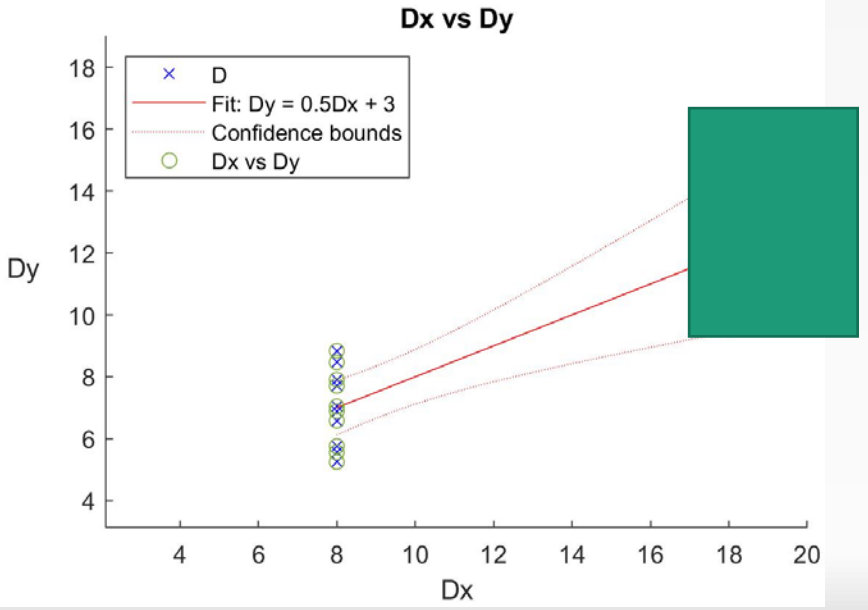
```
Linear Fits
Ax vs Ay
                    Estimate      SE        tStat       pValue
                    ________    ________    ______    _________

    (Intercept)       3.0001      1.1247    2.6673     0.025734
    x1               0.50009     0.11791    4.2415    0.0021696

Bx vs By
                    Estimate      SE        tStat       pValue
                    ________    ________    ______    _________

    (Intercept)       3.0009      1.1253    2.6668     0.025759
    x1                   0.5     0.11796    4.2386    0.0021788

Cx vs Cy
                    Estimate      SE        tStat       pValue
                    ________    ________    ______    _________

    (Intercept)       3.0025      1.1245    2.6701     0.025619
    x1               0.49973     0.11788    4.2394    0.0021763

Dx vs Dy
                    Estimate      SE        tStat       pValue
                    ________    ________    ______    _________

    (Intercept)       3.0017      1.1239    2.6708      0.02559
    x1               0.49991     0.11782     4.243    0.0021646
```

Same trend: linear fitting outputs the same result



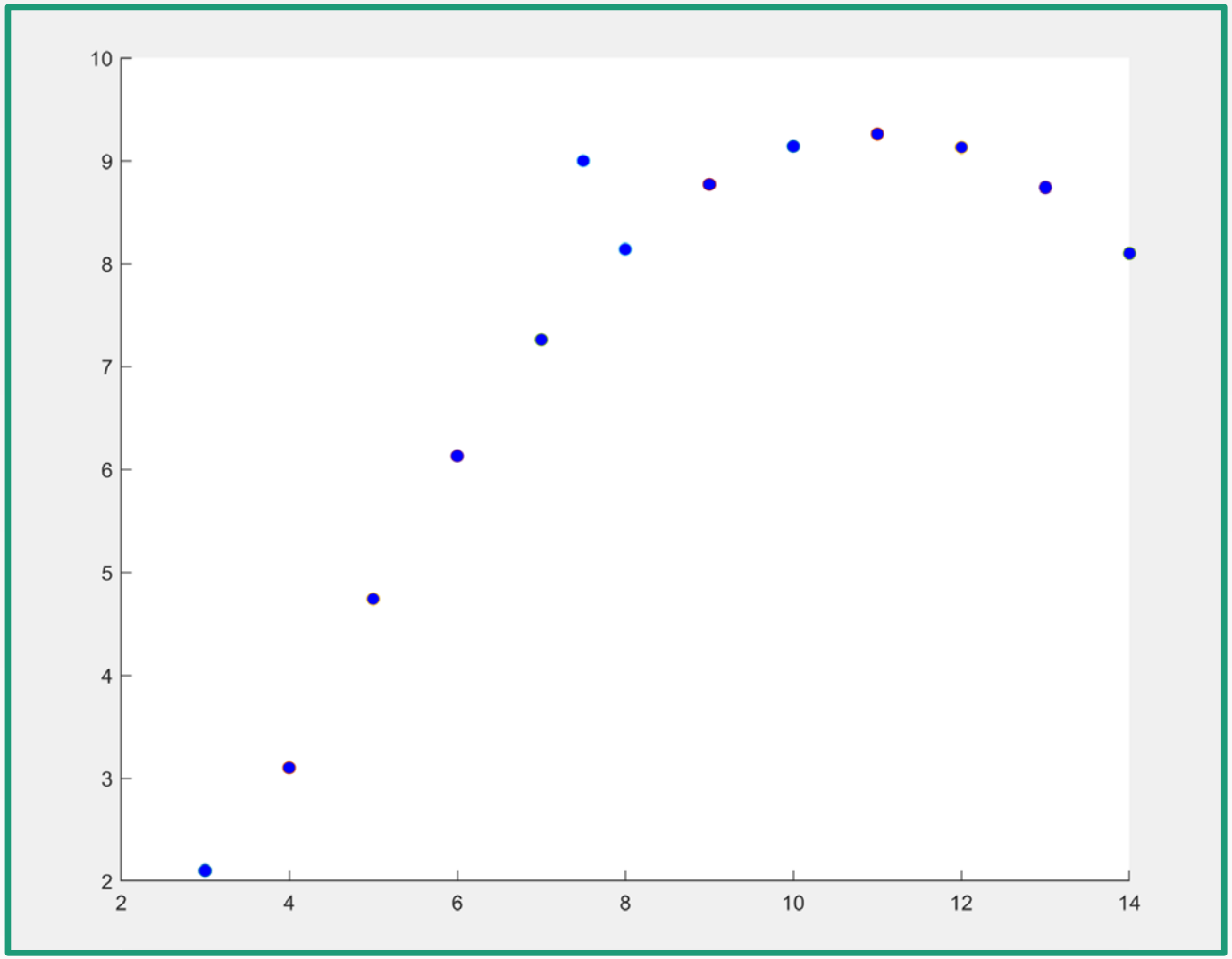






If we plot them?

II	
x	y
10.0	9.14
8.0	8.14
13.0	8.74
3.0	2.1
9.0	8.77
11.0	9.26
14.0	8.10
6.0	6.13
4.0	3.10
12.0	9.13
7.5	9.0
7.0	7.26
5.0	4.74



An outlier can be esily detected…

Let’s try to use MATLAB to redo the Plots

And, if you like,

HERE you find modifiable charts about COVID-19

Stephen Few:

Each of the examples that appear below illustrates quantitative information that is poorly designed for communication.

Look at Stephen’s solutions

Information Visualization

Scientific Visualization versus Information Visualization

Narrative Visualization: Telling Stories with Data