

# *SCIENTIFIC AND LARGE DATA VISUALIZATION*

*October 18, 2020*

*Fundamentals of Information Visualization – Part II*

*Daniela Giorgi*

*Visual Computing Lab, CNR-ISTI*

*Recap from the previous lesson*

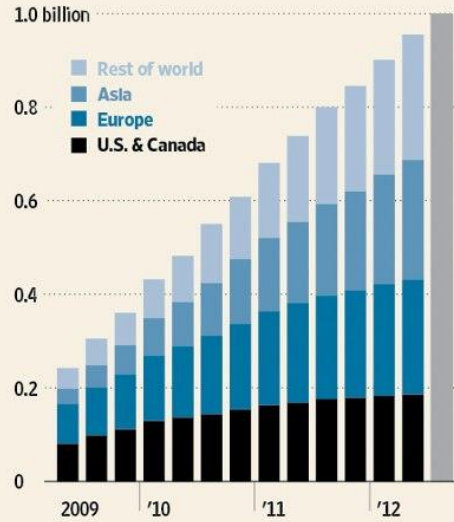


## Facebook Nation | The social network reaches 1 billion active users

If Facebook were a nation, it would have the third-largest population. But if you compared its revenue to national economies, it would rank 156th.

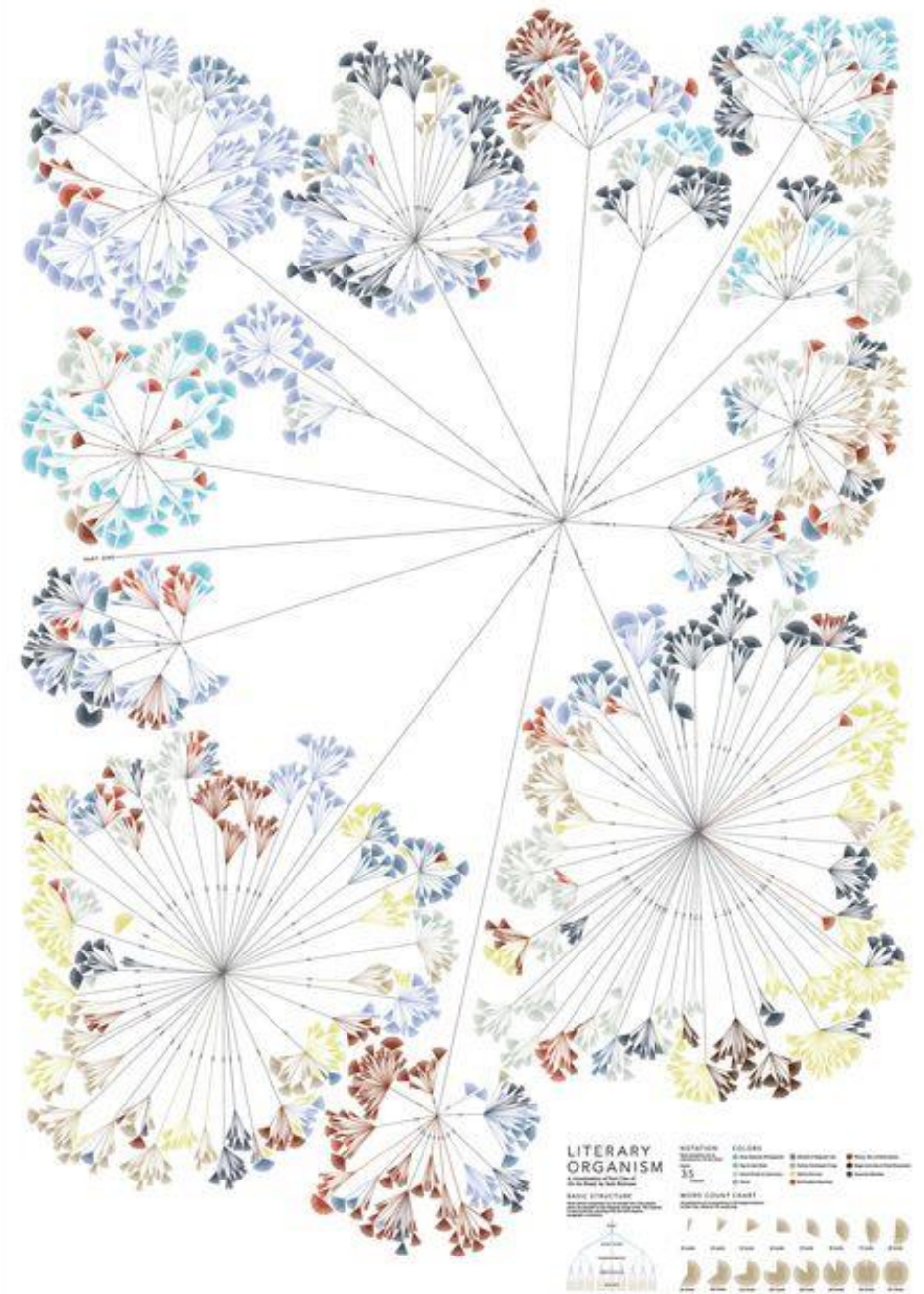
Total population, billions of people		Ranking		2011 gross domestic product, billions of dollars
1.34	China	#1	#154	Swaziland 3.98
1.24	India	#2	#155	Fiji 3.81
1 billion users*	Facebook	#3	#156	Facebook \$3.71 billion
0.31	U.S.	#4	#157	Barbados 3.69
0.24	Indonesia	#5	#158	Togo 3.59

Facebook's monthly active users by region, quarterly (current breakdown not available)

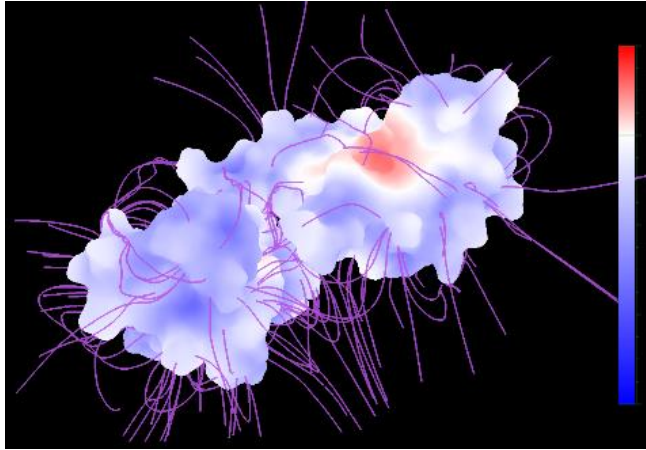


Note: For purposes of reporting MAUs and revenue by geographic region, Europe includes Russia and Turkey, Asia includes Australia and New Zealand, and rest of world includes Africa, Latin America, and the Middle East.

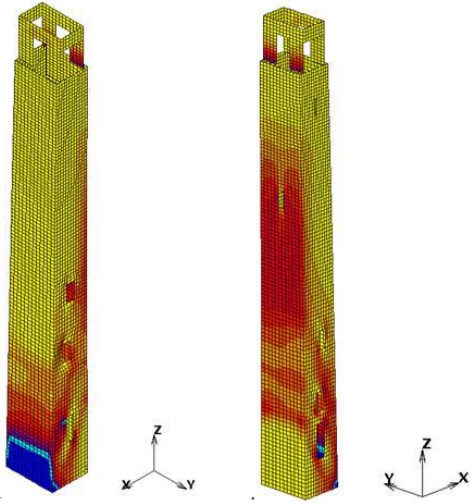
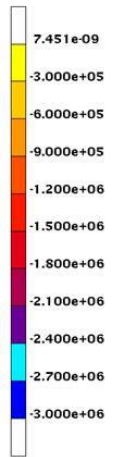
Sources: The World Bank (economic and population data); the company (regional data); FactSet (quarterly revenue) The Wall Street Journal



## 1. Visualization is in the mainstream

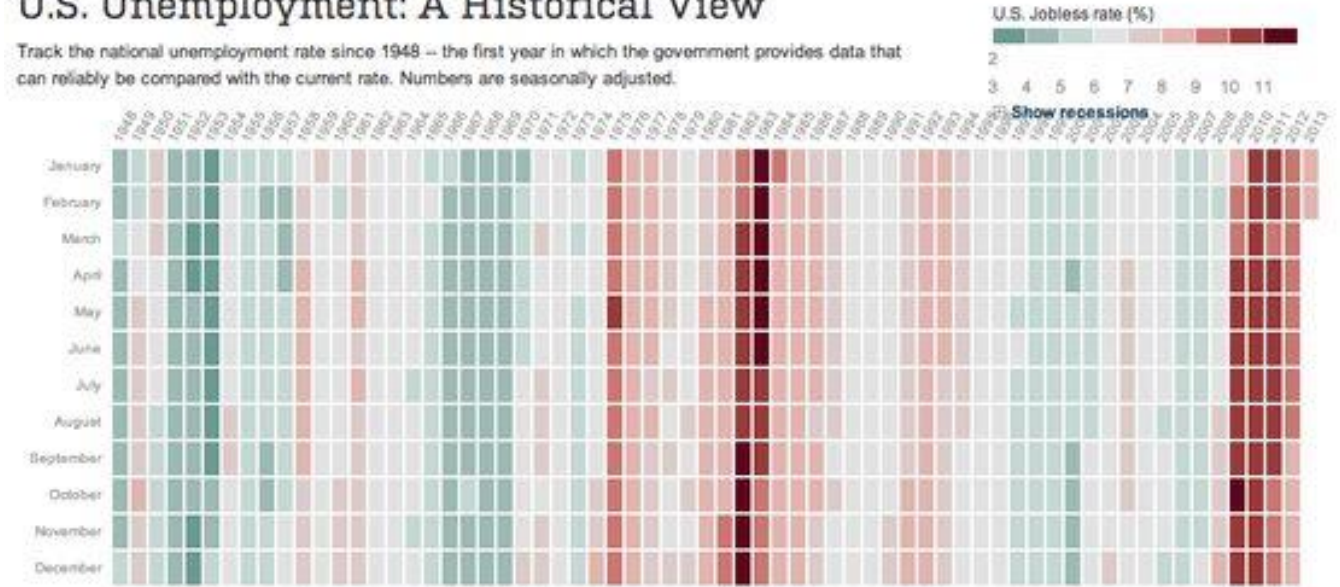


Inc: 343  
Time: 3.410e+00



## U.S. Unemployment: A Historical View

Track the national unemployment rate since 1948 -- the first year in which the government provides data that can reliably be compared with the current rate. Numbers are seasonally adjusted.

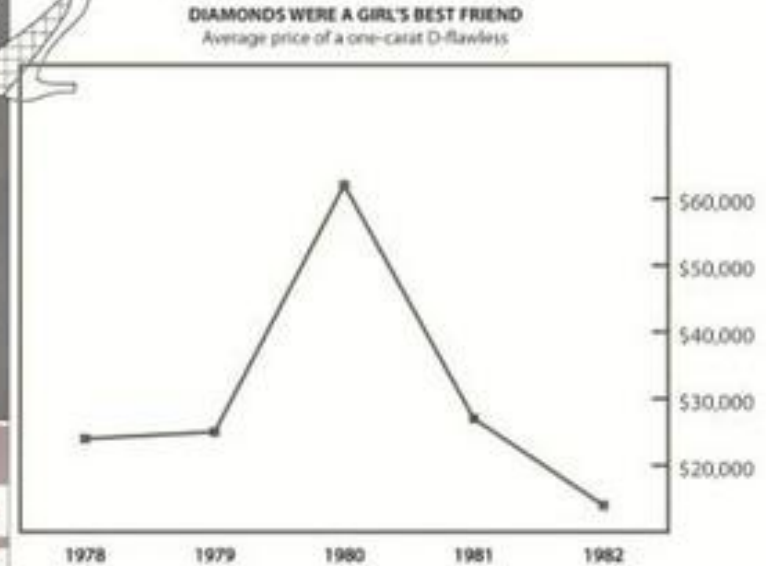
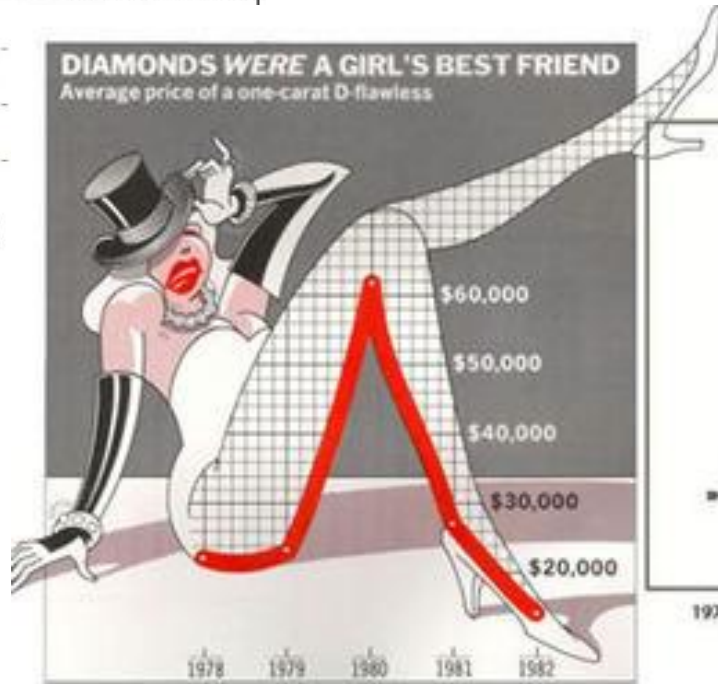
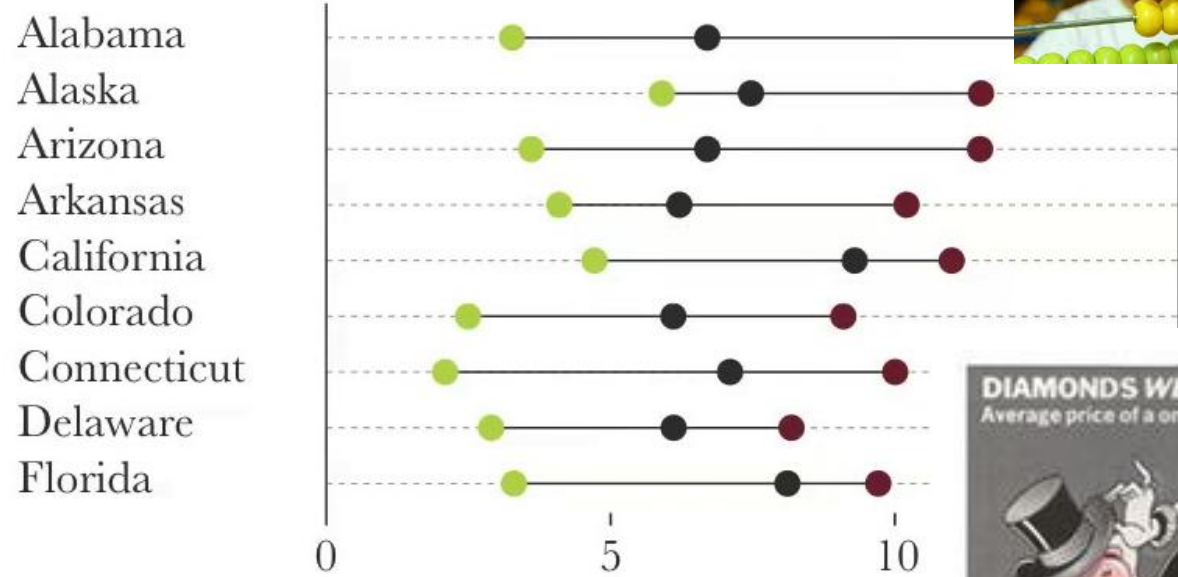


Sources: Bureau of Labor Statistics; Current Population Survey. Updated: March 8, 2013

## 2. Difference with scientific visualization

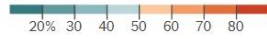
# Unemployment rate (%)

● Current    ● Historical maximum    ● Historical minimum

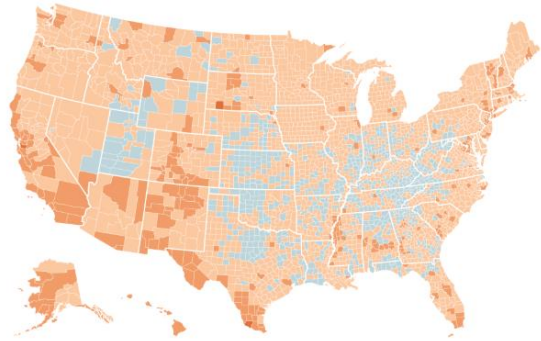


## 3. Visualization and cognition

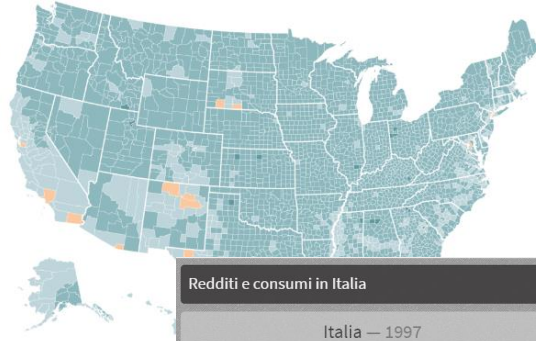
Percentage of adults per county who think ...



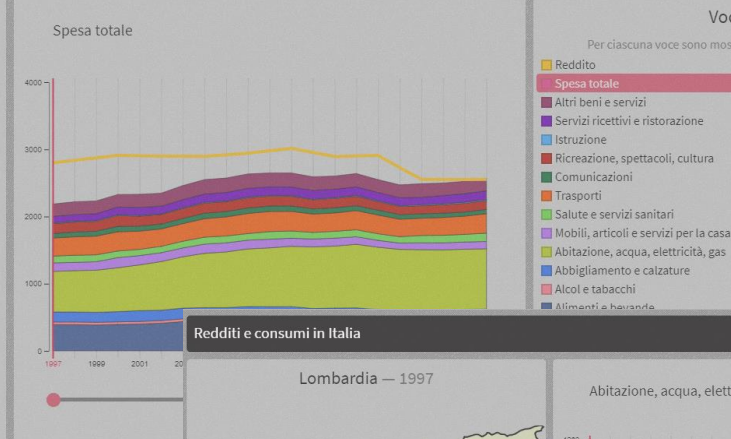
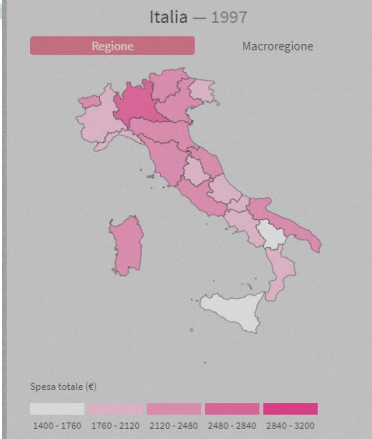
Global warming will harm people in the United States



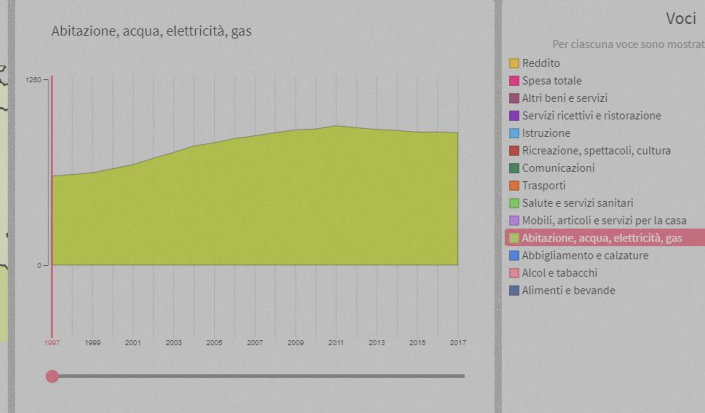
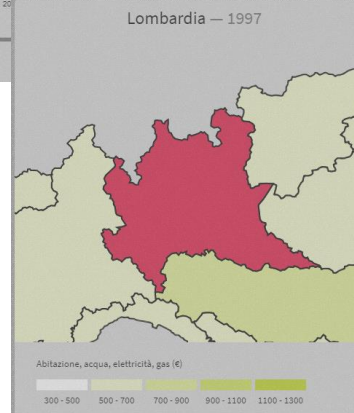
Global warming will harm me, personally



### Redditi e consumi in Italia



### Redditi e consumi in Italia

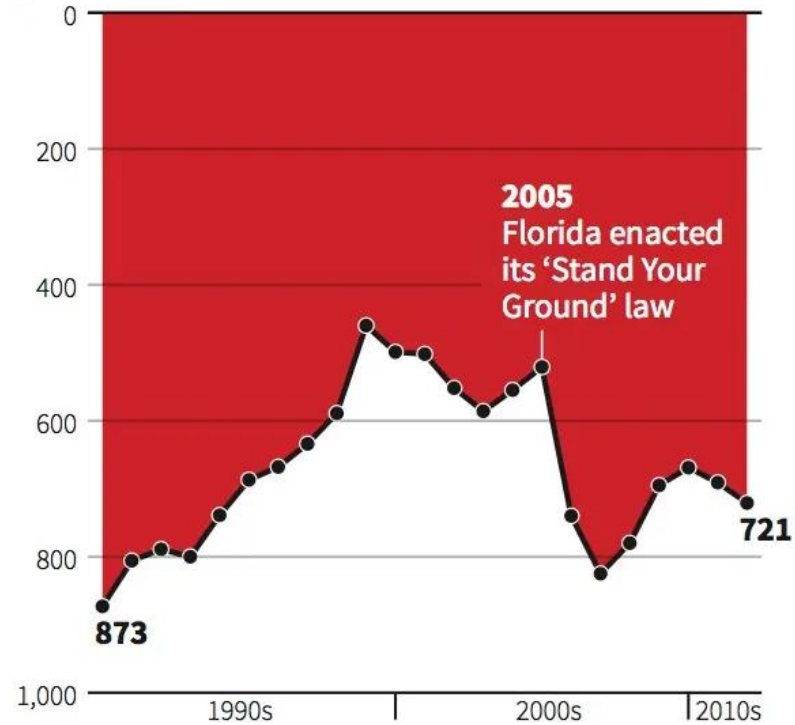


Fonte dati: Banca d'Italia (redditi) e Istat (spese).

## 4. Explanation, exploration, confirmation

# Gun deaths in Florida

Number of murders committed using firearms

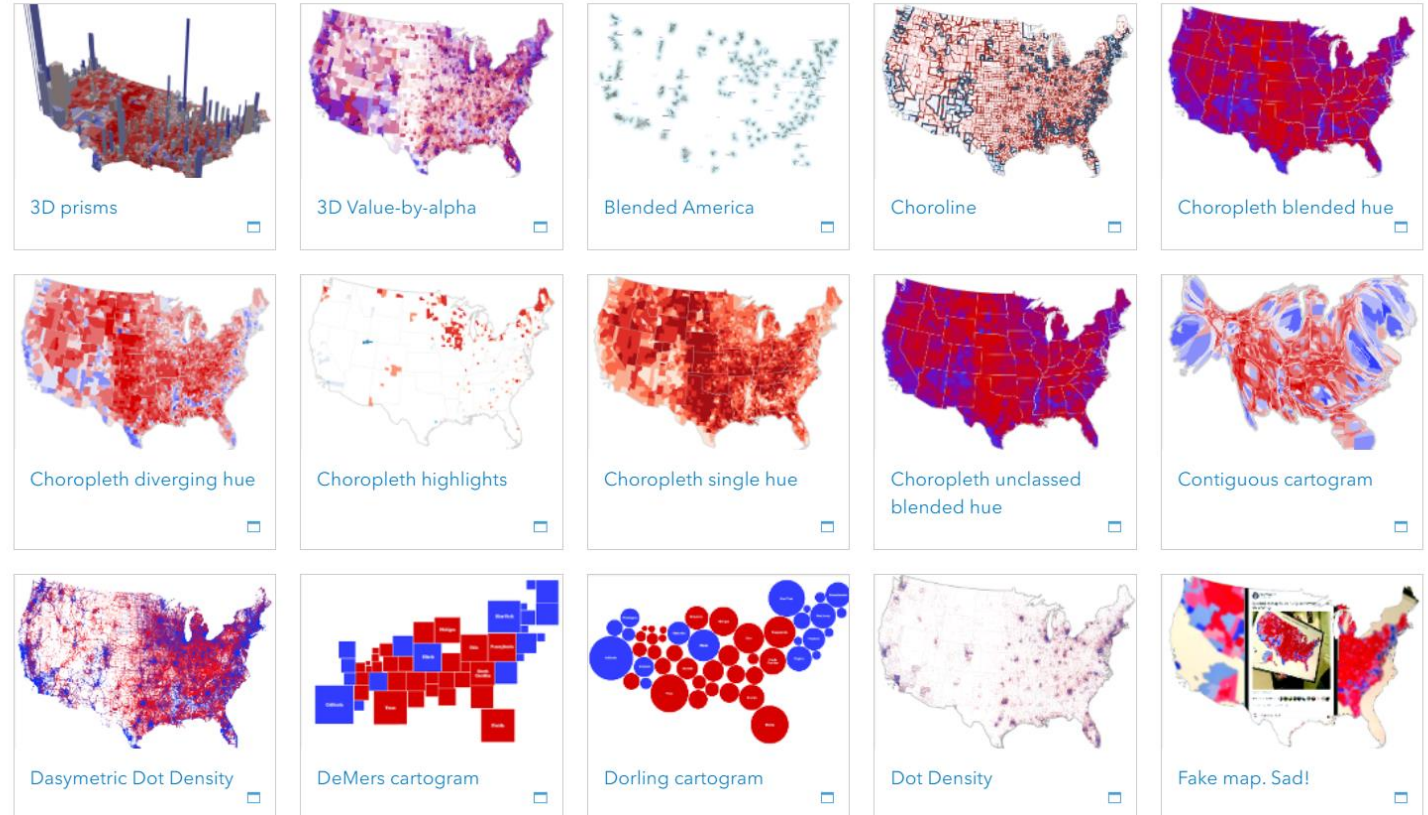


Source: Florida Department of Law Enforcement

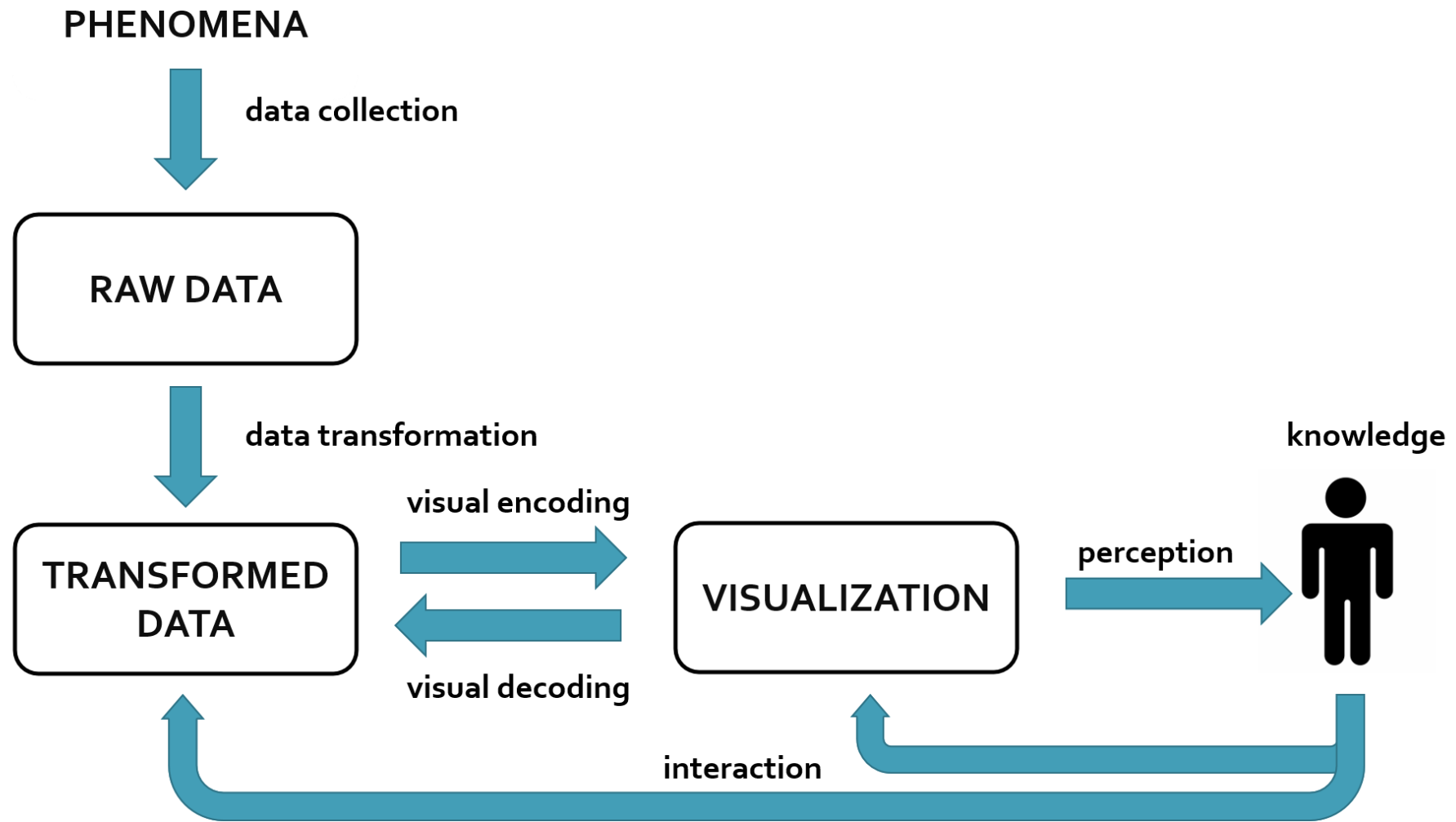
C.Chan 16/02/2014



Thematic maps of the 2016 Presidential election (lower 48 states)



5. Different visualizations, different narratives

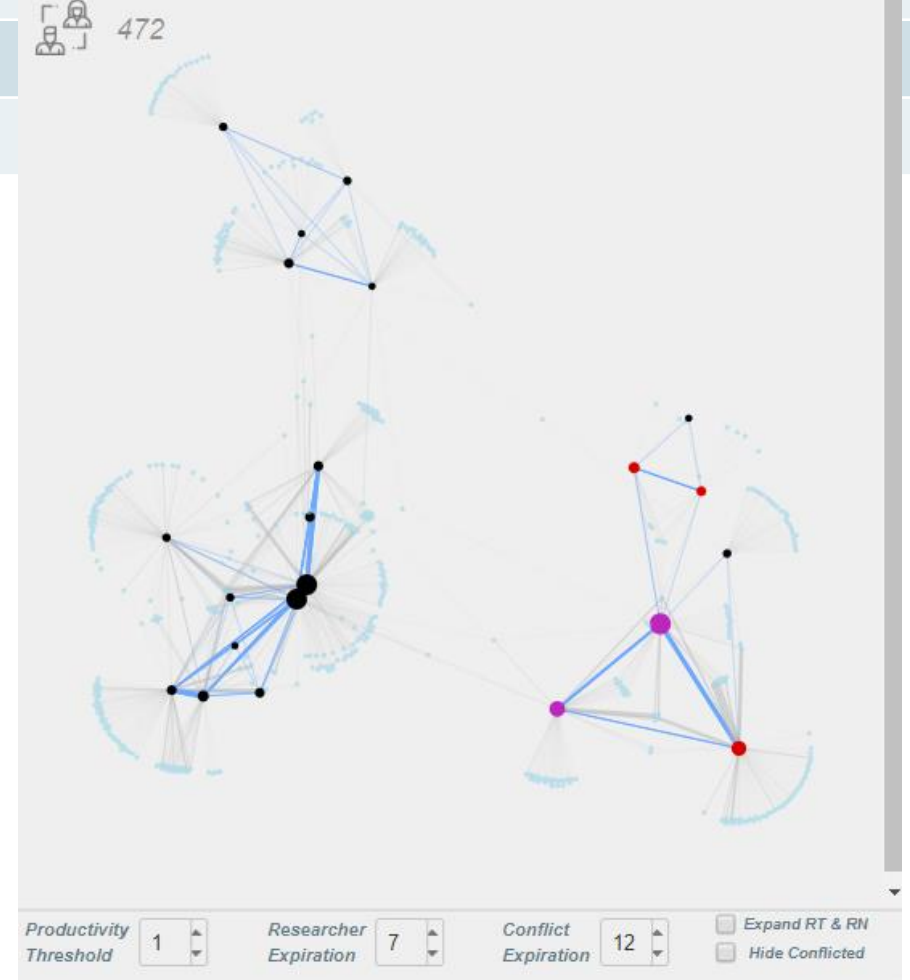


6. The visualization pipeline, from data to insight

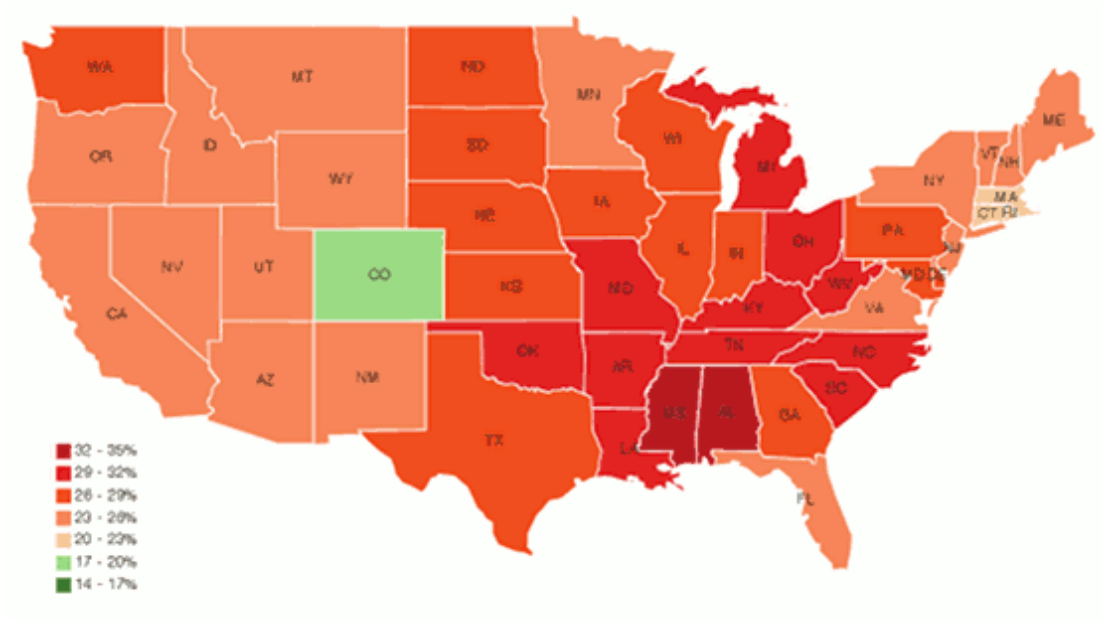
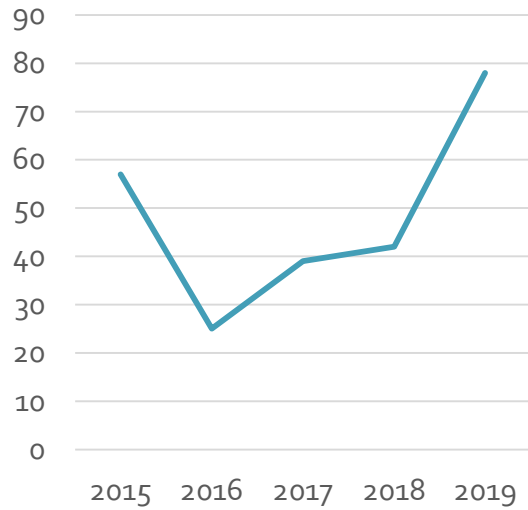
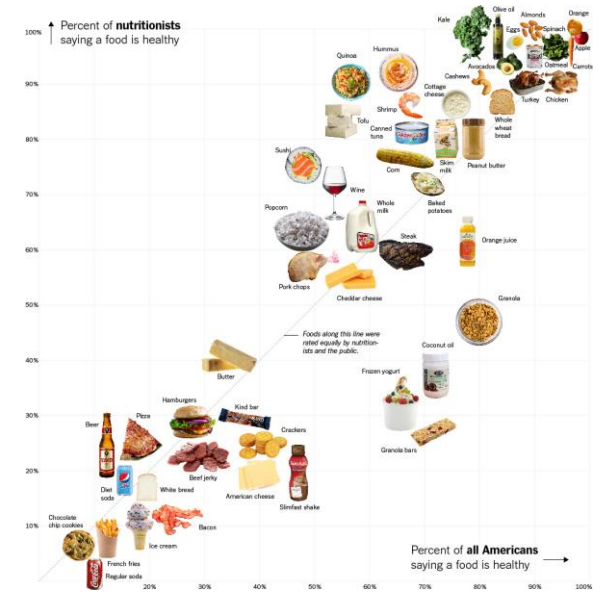
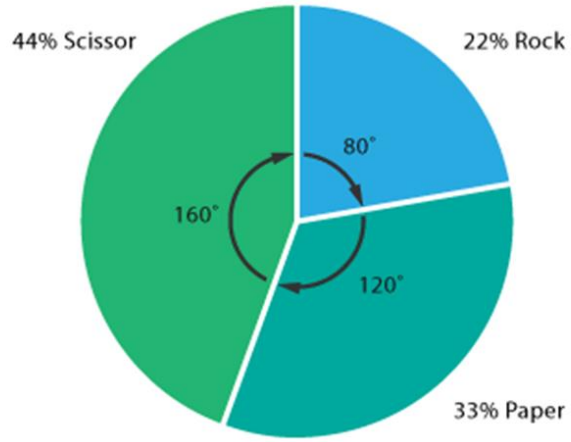
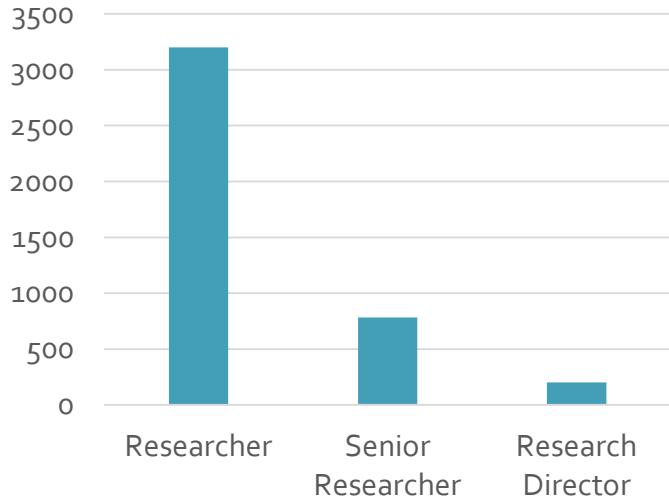


	ID	Gender	Age	Assumption date	Role	Seat
M. Rossi	01066	M	43	01/03/2009	Researcher	Genoa
L. Bianchi	01043	F	36	15/06/2015	Senior Researcher	Genoa
G. Neri	01432	M	59	01/09/1998	Research Director	Pisa
T. Smith	02312	M	32	01/01/2019		
D. Verdi	00043	F	55	01/05/2005		
W. James	01203	M	48	01/03/2004		

RESEARCHER NETWORK



## 7. Data types, items and attributes



## 8. Fundamental graphs

## *Fundamental graphs [Continued]*



# Stacked bar charts

Representation of additional attributes

As many bars as the number of categories for the first attribute, and as many segments within each bar as the values for the other categorical attribute

Good to communicate proportion

## Group Data By

- Starting Year
- Type of Grant
- Nation
- Subject Area (Scopus Only)
- Group Subjects by Macroareas

## General Information

Total People Analyzed 4537  
Scopus Profiles Found 4003  
Orcid Profiles Found 3339  
Orcid Profiles attached to Scopus 1233

Hover on the graph bars to view additional details.

## About the current Graph

This view shows the distribution of the (main) research areas of ERC winners divided by nation.

The results come from the Scopus website, but the areas are grouped according to the [ERC-defined Research Domains](#) (or Panels).

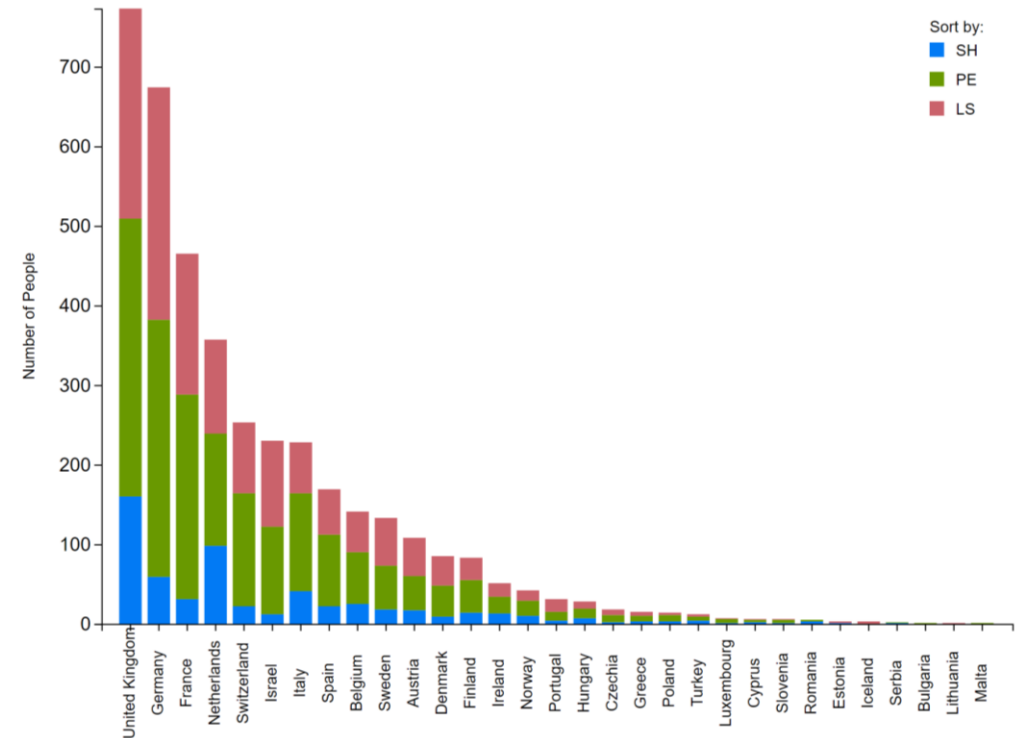
Hover on the chart legend to see the meaning of each group.

Click on the chart legend to sort the bars by a subject area. By default bars are sorted by total people in each country.

## ERC Academic Profiles Statistics

A tool that aims to show how academic profiles on the internet are sometimes harder than expected to find.

### Subject Area (Scopus Only)



[Davide Rucci, AA 2018-2019, <https://drdav.github.io/ERC-Academic-Profiles/>]

# Grouped bar charts

Representation of additional attributes

The same bar chart is repeated multiple times for the number of categories in the additional attribute

Good for comparison of individual values

## Group Data By

- Starting Year
- Type of Grant
- Nation
- Subject Area (Scopus Only)

## General Information

Total People Analyzed	4537
Scopus Profiles Found	4003
Orcid Profiles Found	3339
Orcid Profiles attached to Scopus	1233

Hover on the graph bars to view additional details.

## About the current Graph

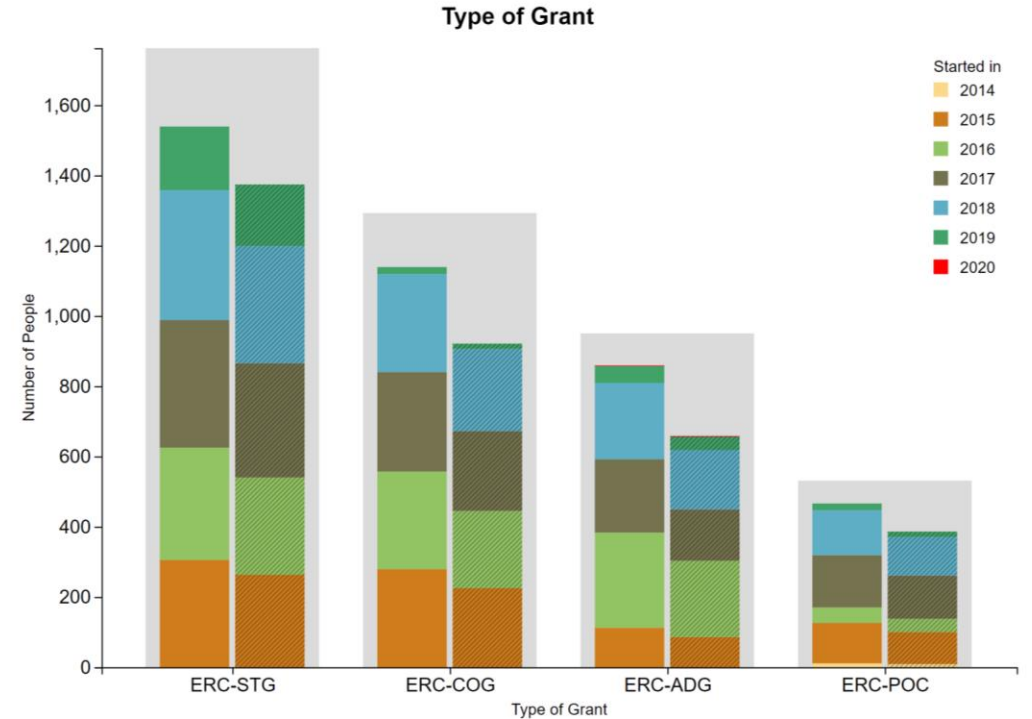
The bars show how many ERC-winning people have a (findable) Scopus and/or Orcid profile, grouped by the type of grant they have received from the EU.

Solid-colored columns represent Scopus profiles, while a diagonal pattern overlay is used for Orcid's. Grey bars show the total number of people that have received a grant of a given type.

The columns are subsequently divided by the year in which their project started.

## ERC Academic Profiles Statistics

A tool that aims to show how academic profiles on the internet are sometimes harder than expected to find.



[Davide Rucci, AA 2018-2019, <https://drdav.github.io/ERC-Academic-Profiles/>]

# Grouped bar charts

Representation of additional attributes

The same bar chart is repeated multiple times for the number of categories in the additional attribute

Good for comparison of individual values

## Group Data By

- Starting Year
- Type of Grant
- Nation
- Subject Area (Scopus Only)

## General Information

Total People Analyzed	4537
Scopus Profiles Found	4003
Orcid Profiles Found	3339
Orcid Profiles attached to Scopus	1233

Hover on the graph bars to view additional details.

## About the current Graph

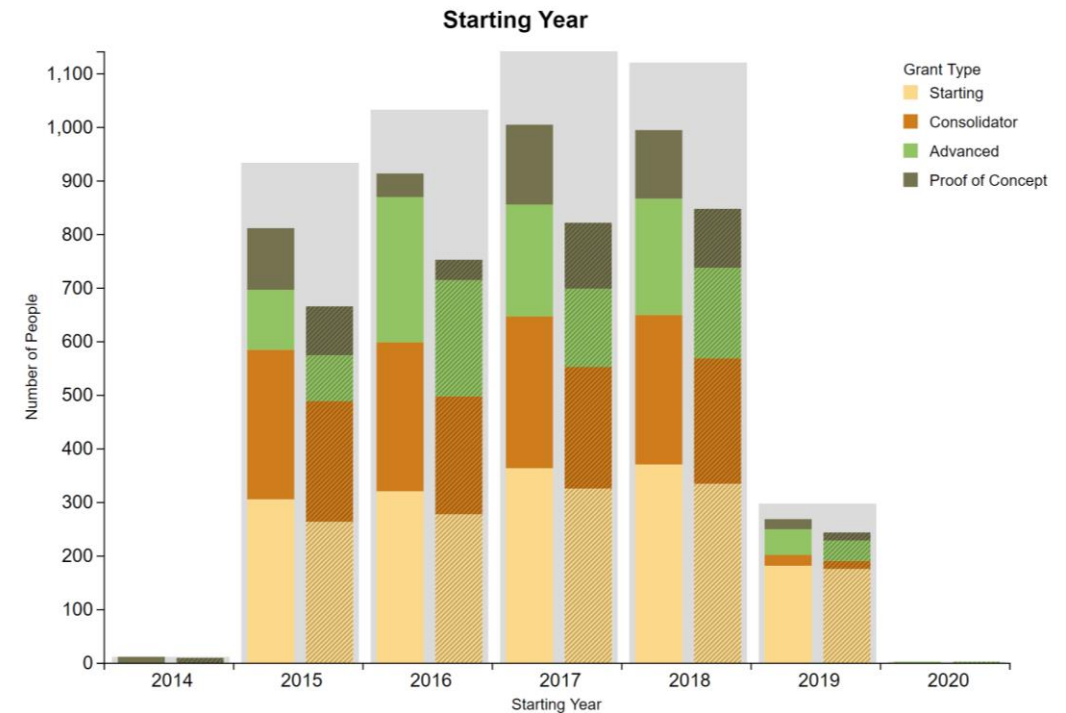
The bars show how many ERC-winning people have a (findable) Scopus and/or Orcid profile, grouped by the year in which they have started their EU project.

Solid-colored columns represent Scopus profiles, while a diagonal pattern overlay is used for Orcid's. Grey bars show the total number of people that started in a given year.

The columns are subsequently divided by type of project, namely Starting Grants, Consolidator Grants, Advanced Grants or Proofs of Concept.

## ERC Academic Profiles Statistics

A tool that aims to show how academic profiles on the internet are sometimes harder than expected to find.



[Davide Rucci, AA 2018-2019, <https://drdav.github.io/ERC-Academic-Profiles/>]

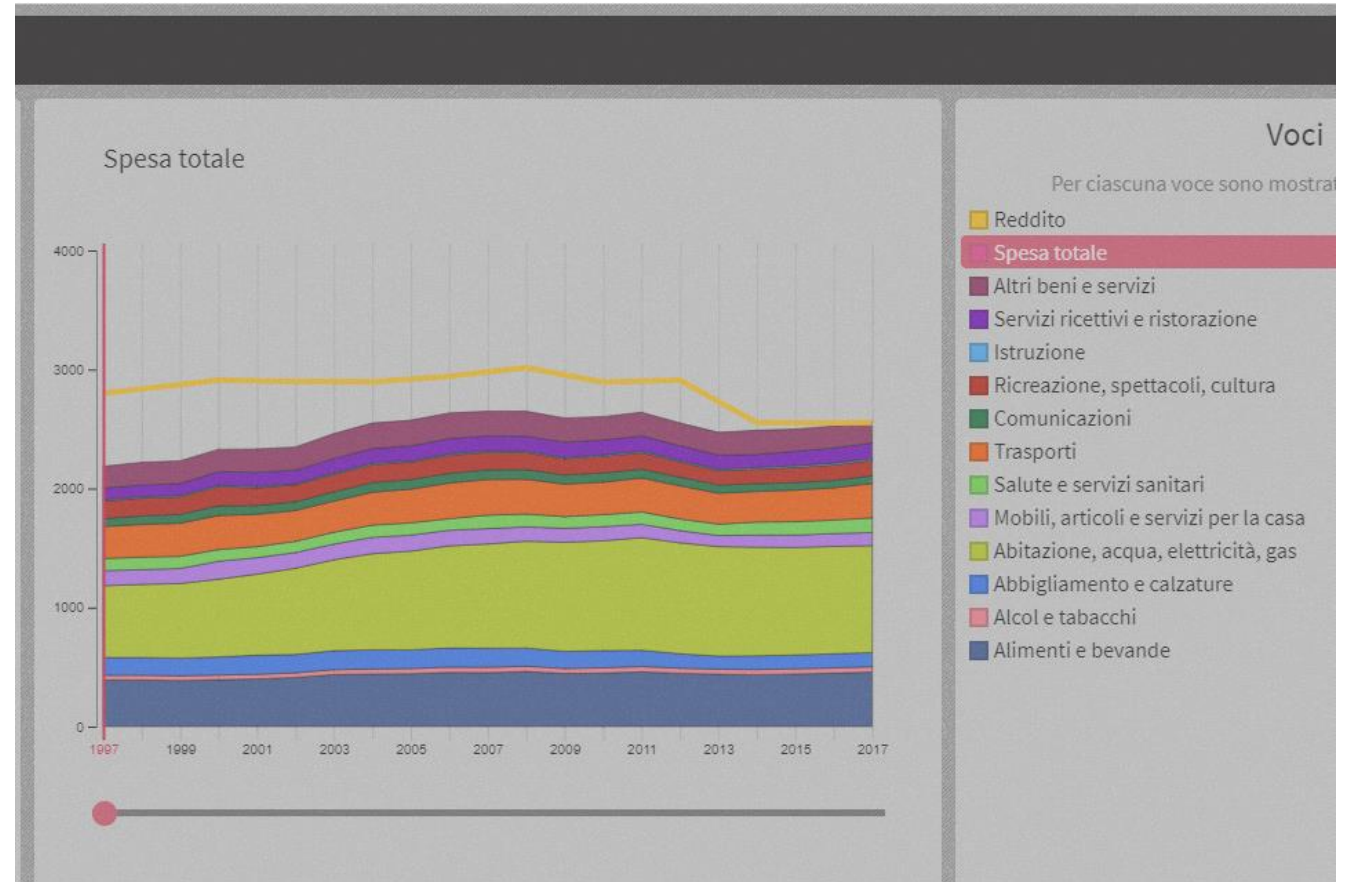
# Stacked line charts

Representation of additional attributes

Values across categories are stacked one on top of the other

The shape of lines is affected by the shape of the lines below

Not meaningful for data that should not be summed



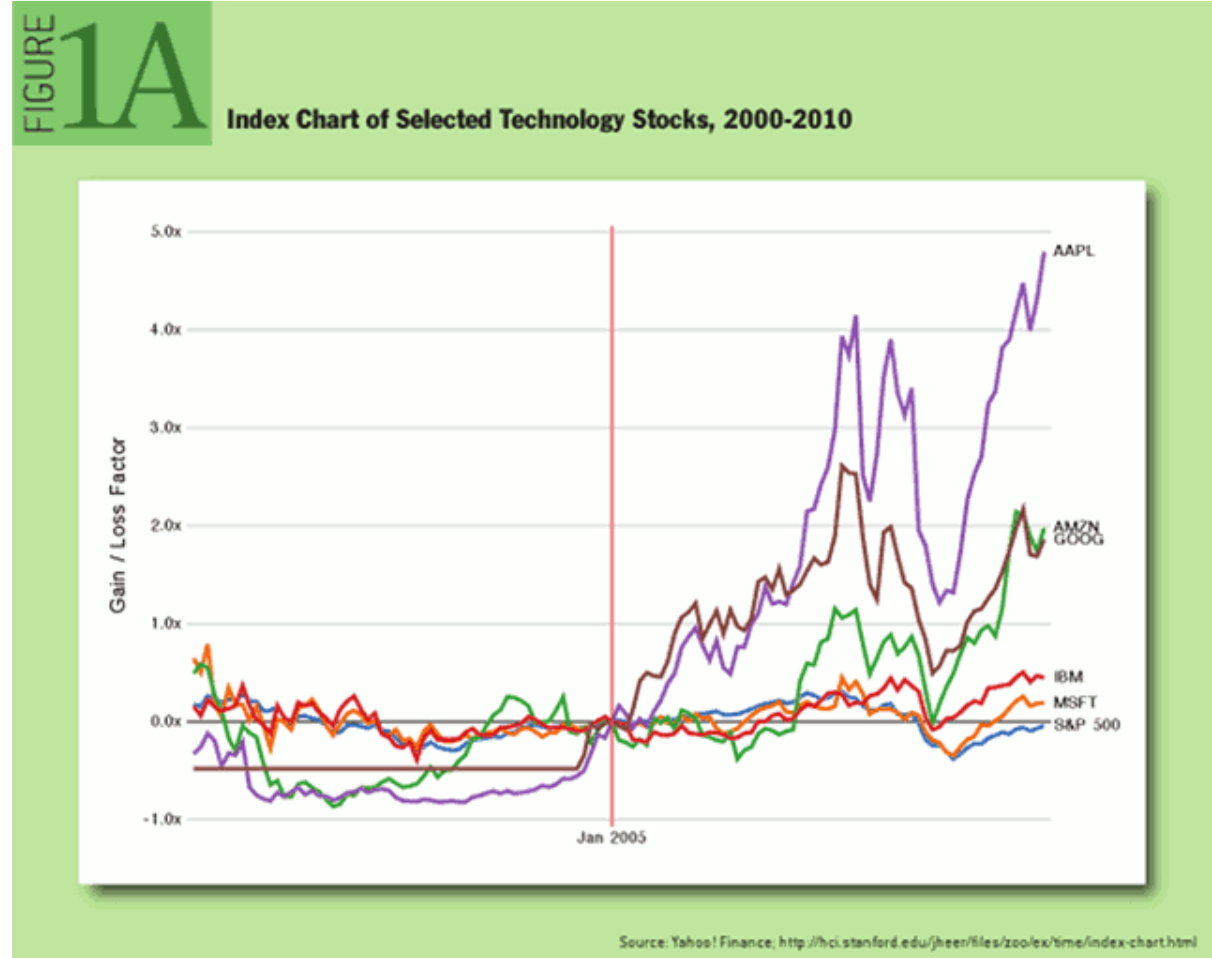
[Matteo Loporchio, AA 2018-2019, <https://mloporchio.github.io/D3IE-Italy/>]

# *Line charts series*

Representation of additional attributes

Collection of lines in the same chart

Easy to compare values, but overlapping curves can make the graph hard to read



[A tour through the visualization zoo, <https://queue.acm.org/detail.cfm?id=1805128>]



# Bubble charts

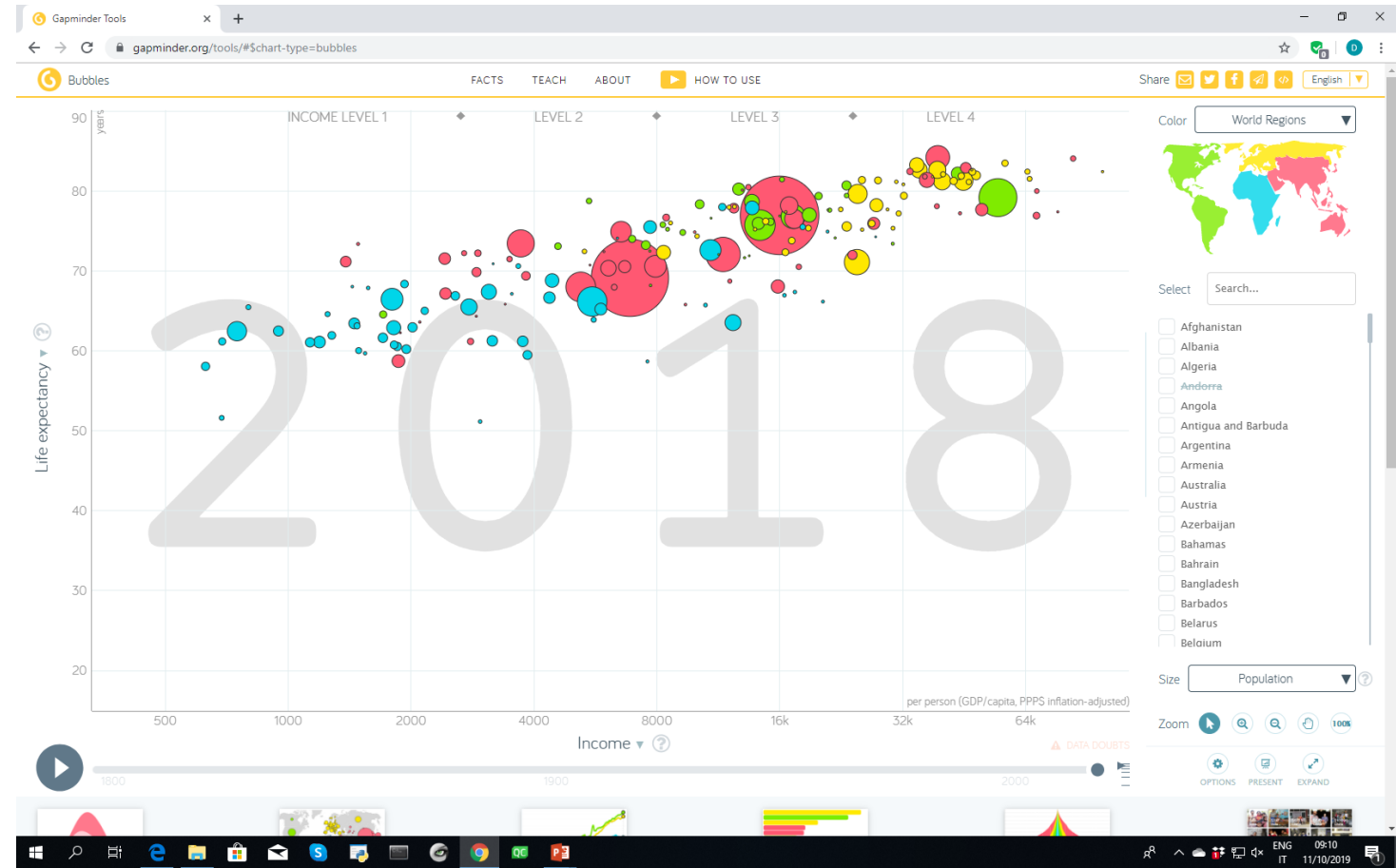
Representation of additional attributes

Variant on scatterplots

Circle areas represent an additional attribute

Colors can distinguish categories or represent additional attributes

Limited data size capacity



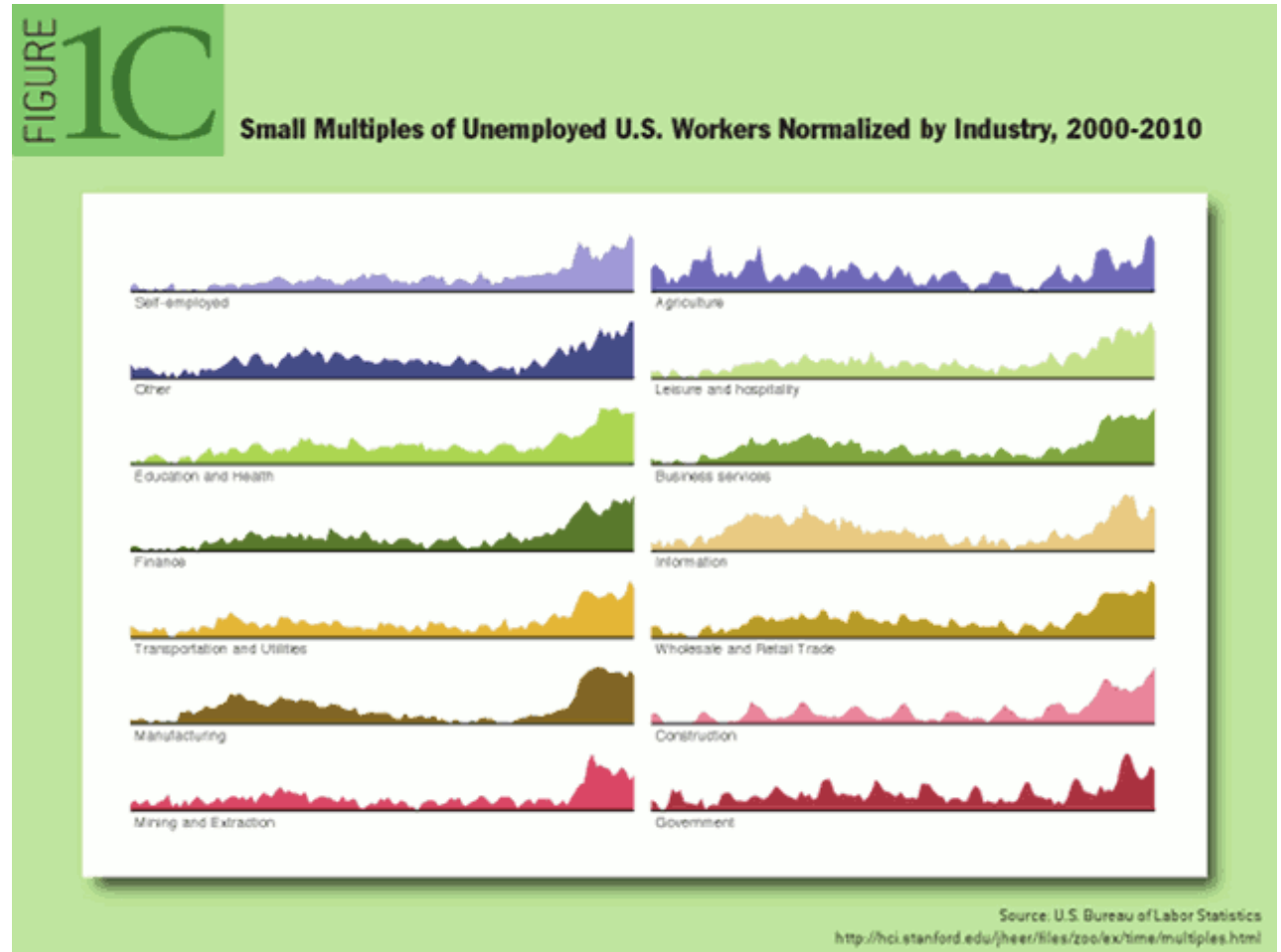
[Gapminder <https://gapminder.org/>; Hans Rosling Ted Talk [https://www.ted.com/talks/hans\\_rosling\\_shows\\_the\\_best\\_stats\\_you\\_ve\\_ever\\_seen?language=it#t-283414](https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen?language=it#t-283414)]

# *Small multiples*

Representation of additional attributes

Each graph in its own chart

It can be done with any graphs

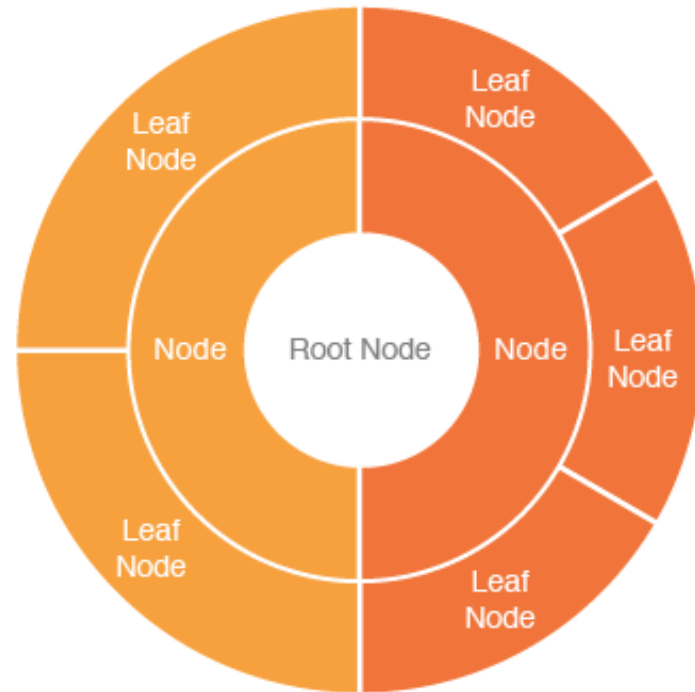
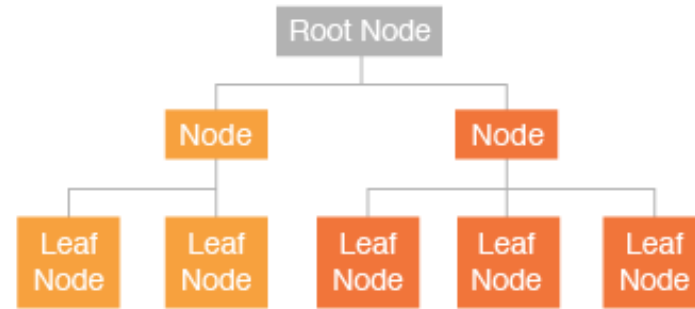


[A tour through the visualization zoo, <https://queue.acm.org/detail.cfm?id=1805128>]

# *Sunburst charts*

To show hierarchies

Each ring represents a level in the hierarchy, moving outwards

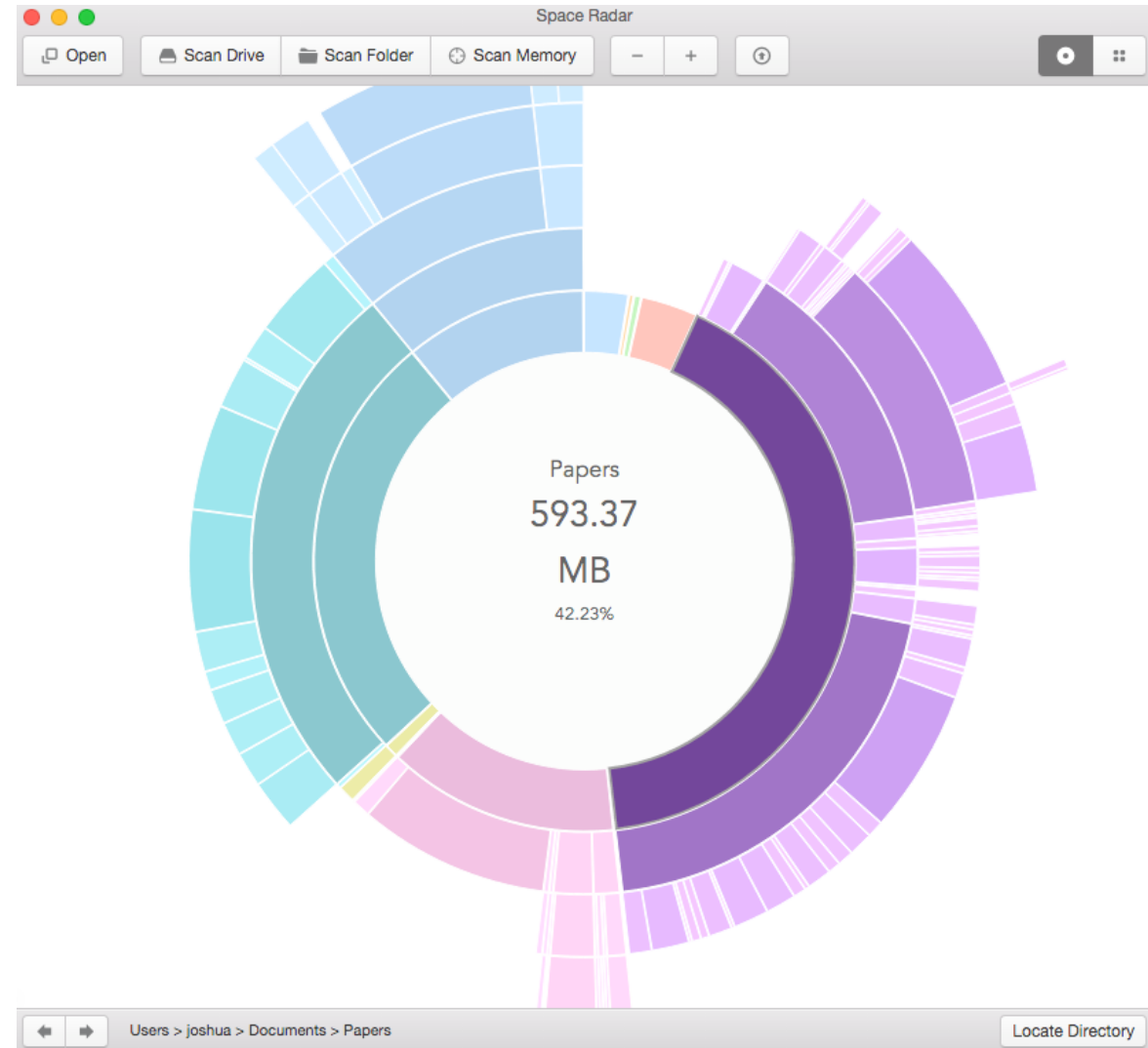


[Data Visualization Catalogue, <https://datavizcatalogue.com/>]

# *Sunburst charts*

To show hierarchies

Each ring represents a level in the  
hierarchy, moving outwards



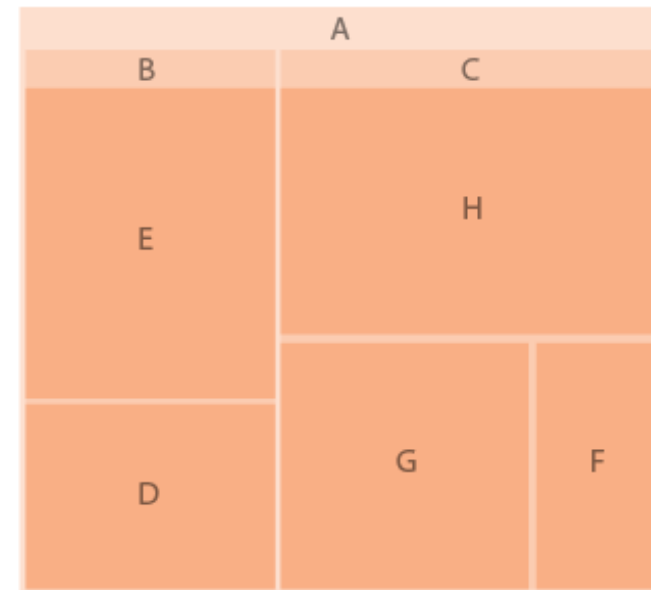
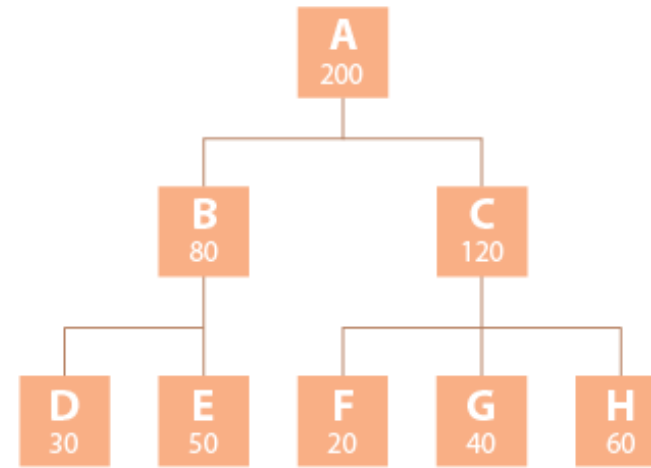
[Space Radar <https://github.com/zz85/space-radar>]

# Treemaps

Alternative way to visualize hierarchies

Each rectangle represents a category, with nested subcategory rectangles

When a quantity is assigned to a category, its area size is displayed in proportion to that quantity and to the other quantities within the same parent category in a part-to-whole relationship



[Data Visualization Catalogue, <https://datavizcatalogue.com/>]



### Occupations of memorable people born in present day Norway

3501 BC - 2015

**Visualizations**

MAKE A

- TREE MAP
- STACKED
- LINE CHART
- MAP

GROUP PEOPLE BY

- PLACES
- OCCUPATIONS**

---

FILTERED BY

**ALL** | FEMALES | MALES

**BORN** | DIED

3501 BC AND 2015

**BORN IN** | DIED IN

Norway

All Cities

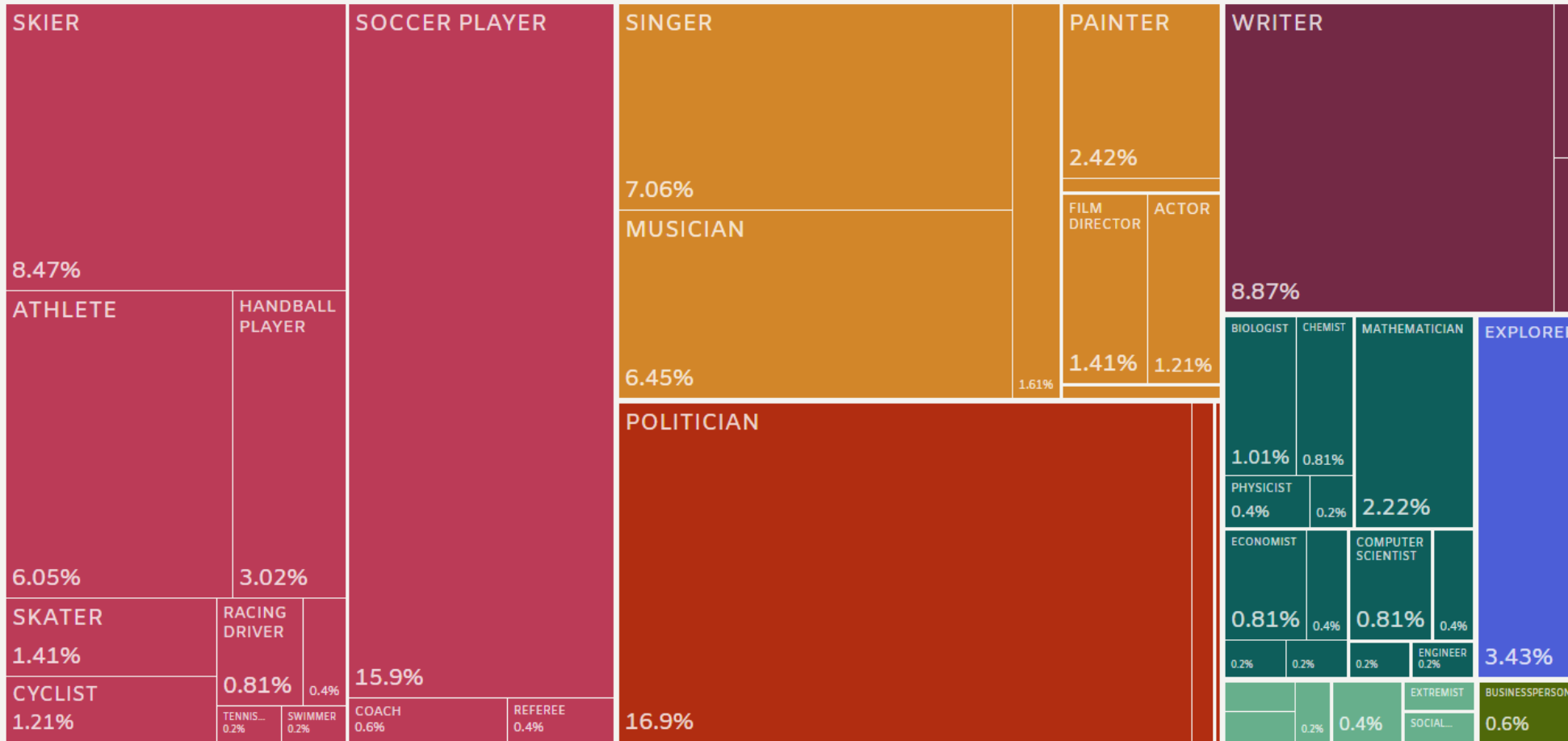
WORKING IN

All Occupations

---

ADVANCED OPTIONS

HPI | 4

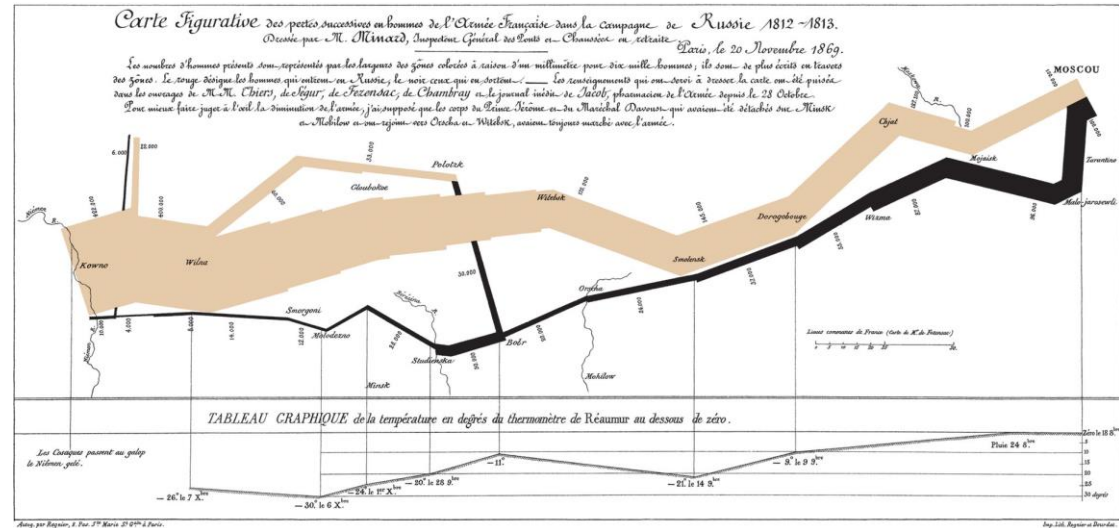


■ EXPLORER
 ■ SCIENCE & TECHNOLOGY
 ■ SPORTS
 ■ ARTS
 ■ PUBLIC FIGURE
 ■ HUMANITIES
 ■ INSTITUTIONS
 ■ BUSINESSPERSON

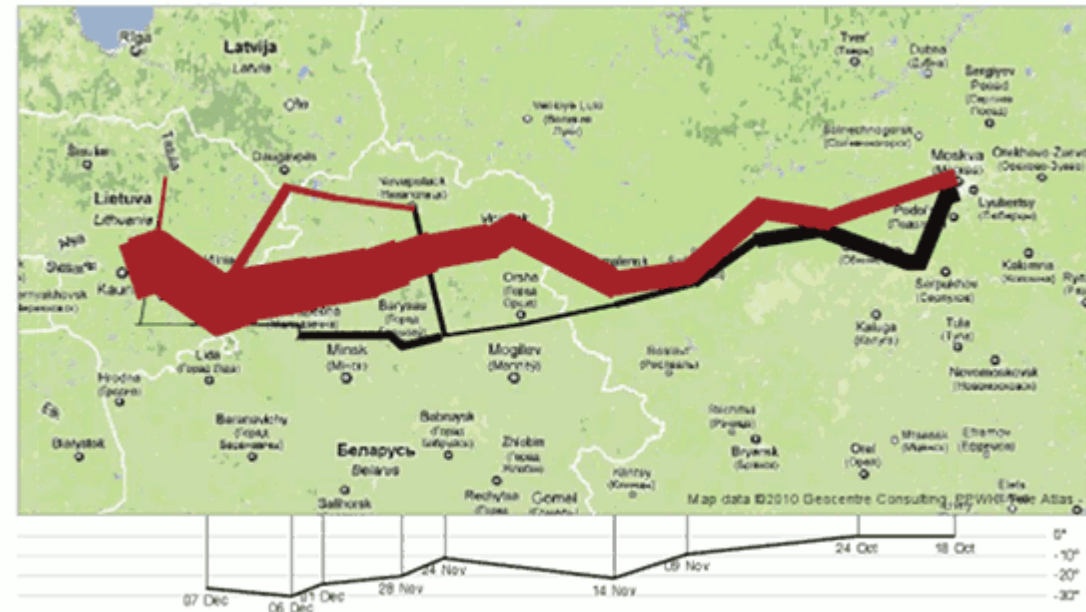
[Pantheon Project, MIT, <https://pantheon.world/explore/viz>]

# Flow maps

Depict the movement of entities  
by placing stroked lines on top of  
maps (space and time)  
Thickness, color, etc. can encode  
additional information

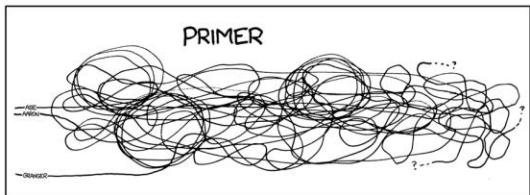
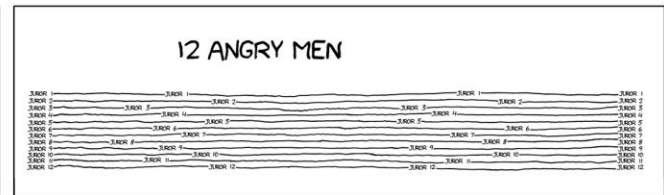
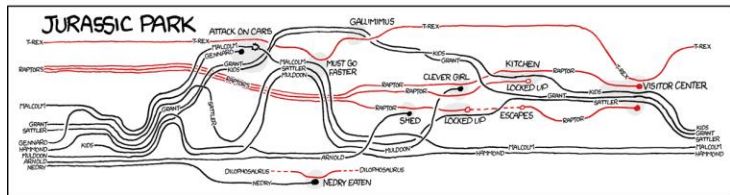
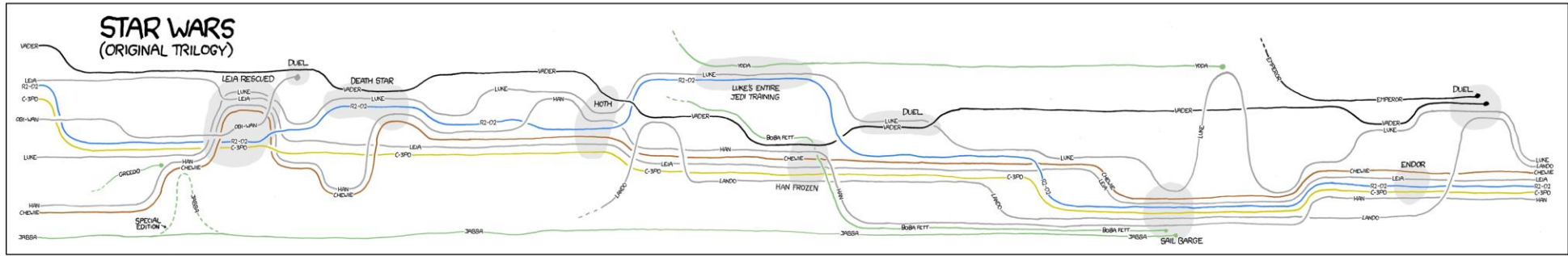
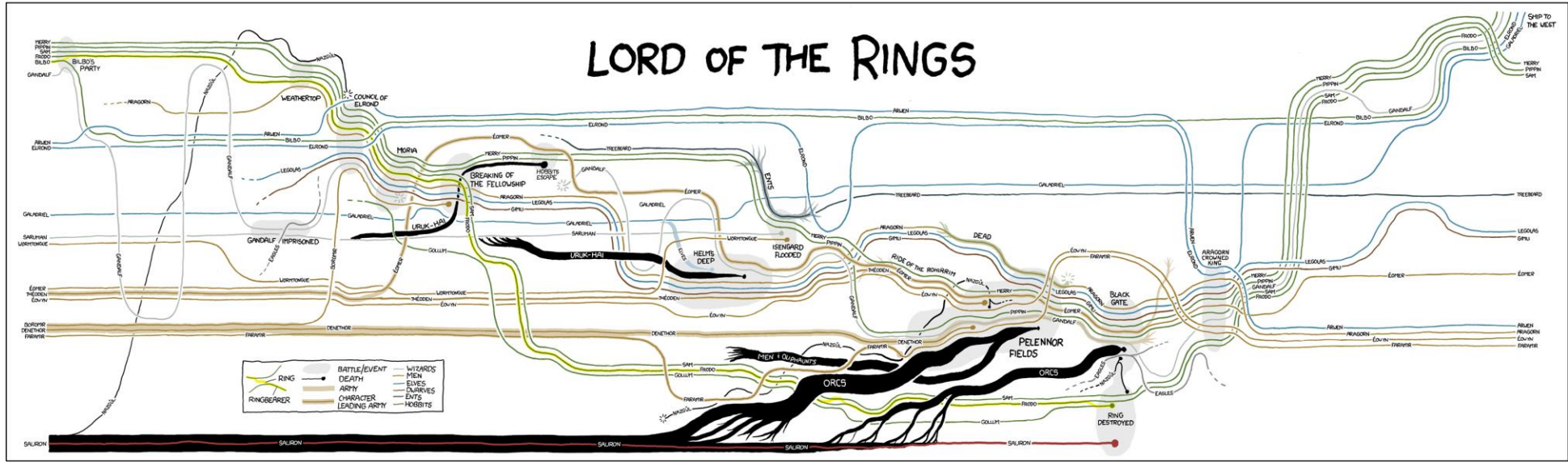


[Minard's map of Napoleon's Russian campaign in 1812]



[A tour through the visualization zoo, <https://queue.acm.org/detail.cfm?id=1805128>]

THESE CHARTS SHOW MOVIE CHARACTER INTERACTIONS.  
 THE HORIZONTAL AXIS IS TIME. THE VERTICAL GROUPING OF THE  
 LINES INDICATES WHICH CHARACTERS ARE TOGETHER AT A GIVEN TIME.





*but also...*



# Wordclouds

Representation of how frequently words appear in a given body of text through the size of the word

Variations in arrangement and color

Mainly used for aesthetic reasons



[Worldcloud generator, <https://www.jasondavies.com/wordcloud/>]

# Wordclouds

Representation of how frequently words appear in a given body of text through the size of the word

Variations in arrangement and color

Mainly used for aesthetic reasons

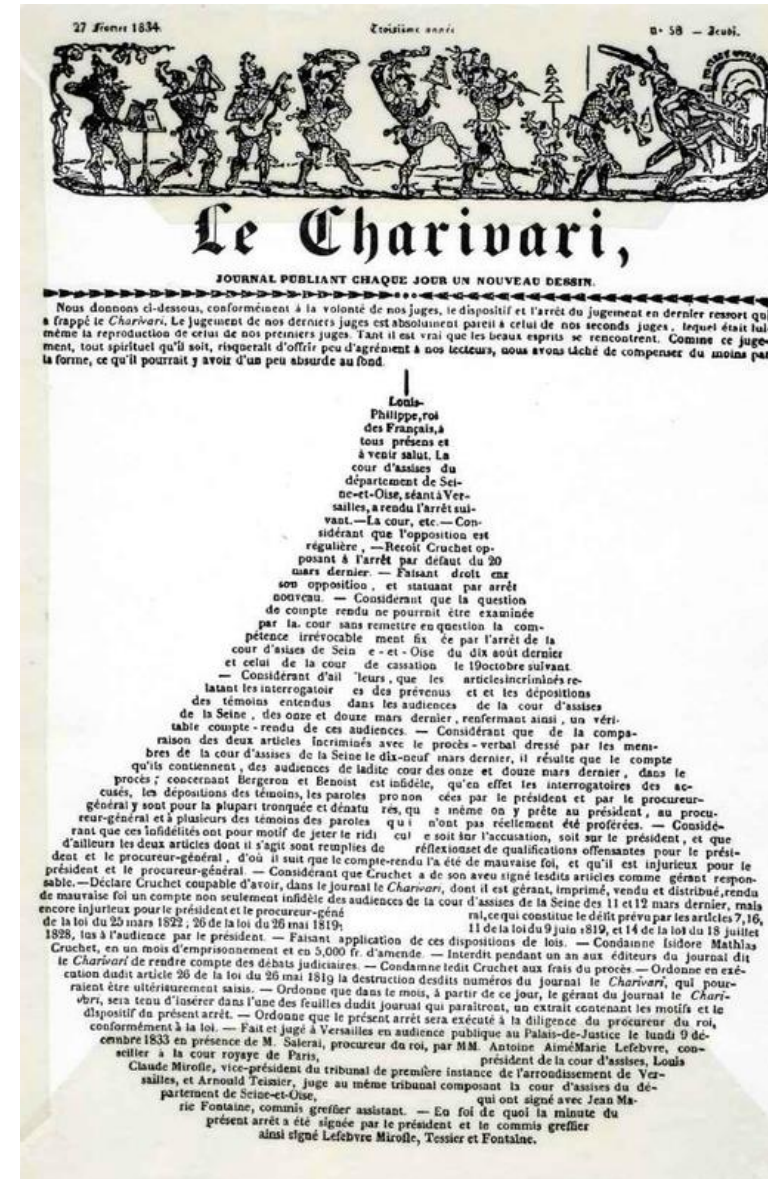


[ [www.wordclouds.com](http://www.wordclouds.com) ]

# Calligrams

Texts arranged so that they form a thematically related image

The image created by the words illustrates the text by expressing visually something associated to (or contradicting) what the text says



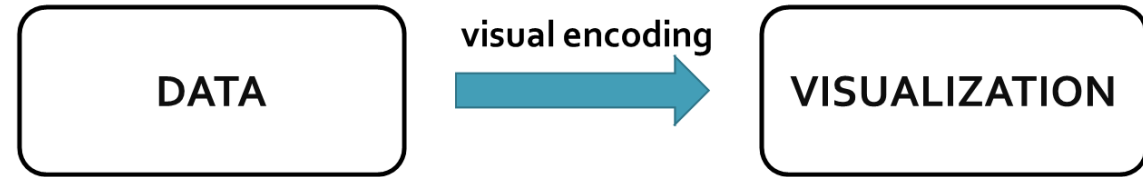
[Le Charivari, 1834, <https://en.wikipedia.org/wiki/Calligram>]

# *Visual encoding and decoding*

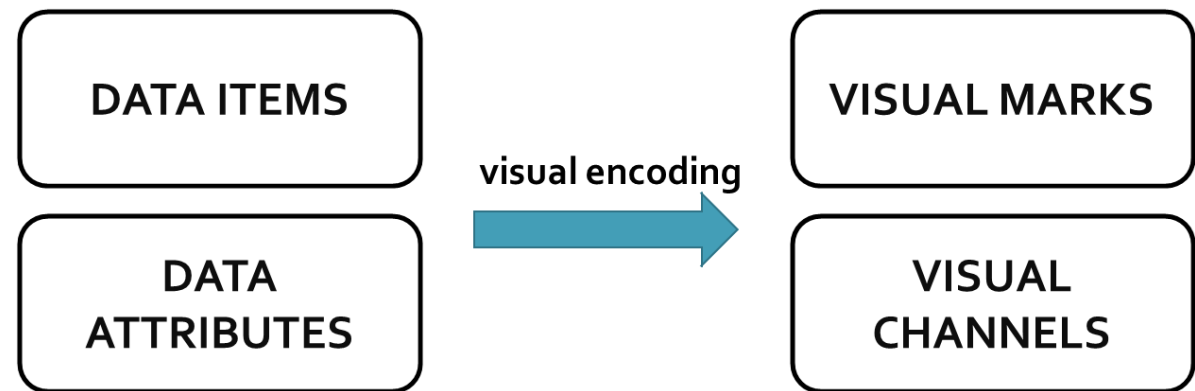


# *Visual encoding*

- Visual encoding means going from data to visual representations



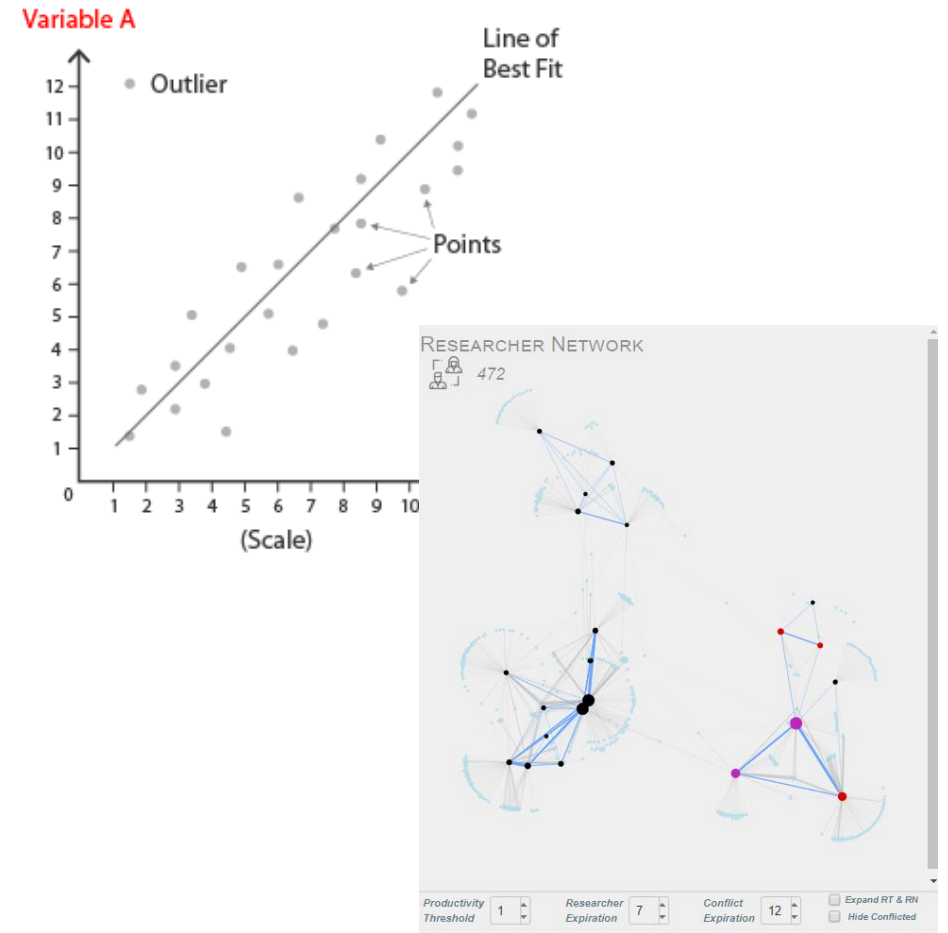
- Encoding requires not only to choose a graph that is appropriate for the data at hand, but also to select the individual graphical elements



# Graphical elements

Marks  
Channels  
Contextual components

- *Marks* are visual objects representing data items
  - points
  - bars
  - lines
  - areas
  - ...

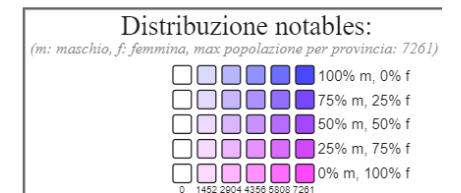
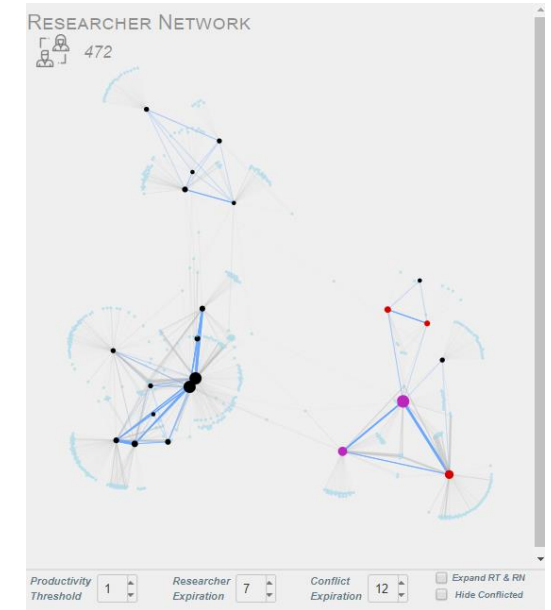
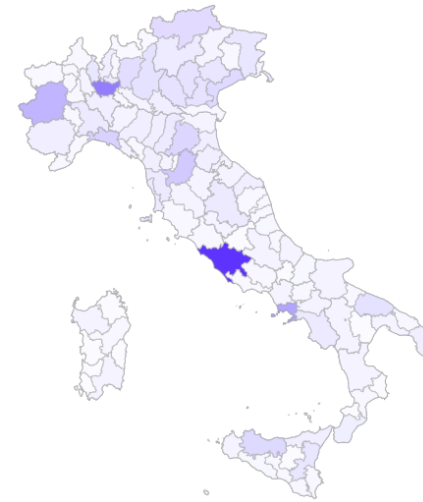


[Mario Salinas, AA 2017-2018, ReviewerNet: <https://reviewernet.org/>]

# Graphical elements

Marks  
Channels  
Contextual components

- *Channels* are the visual properties that encode data attributes
  - position
  - size (length, width, area)
  - angle and slope
  - color (multiple channels)
  - shape and texture
  - ...





# *Graphical elements*

- Contextual components
  - Legends, labels, annotations
  - Grids, reference lines
  - ...

Marks

Channels

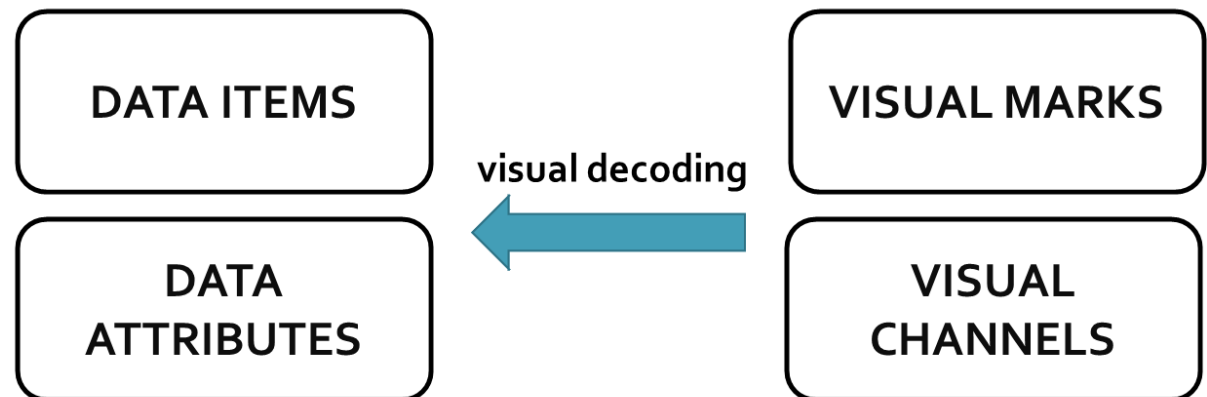
Contextual components

---



# *Visual decoding*

- Visual decoding means deconstructing a visual representation it into its major units, by identifying:
  - the graphical elements
    - what are the visual marks? What are the visual channels?
  - the mapping rules (i.e., the information that the graphical elements represent)
    - what data items do the mark represent? What attributes do the channel represent?
- It is useful for evaluating and redesigning visualizations



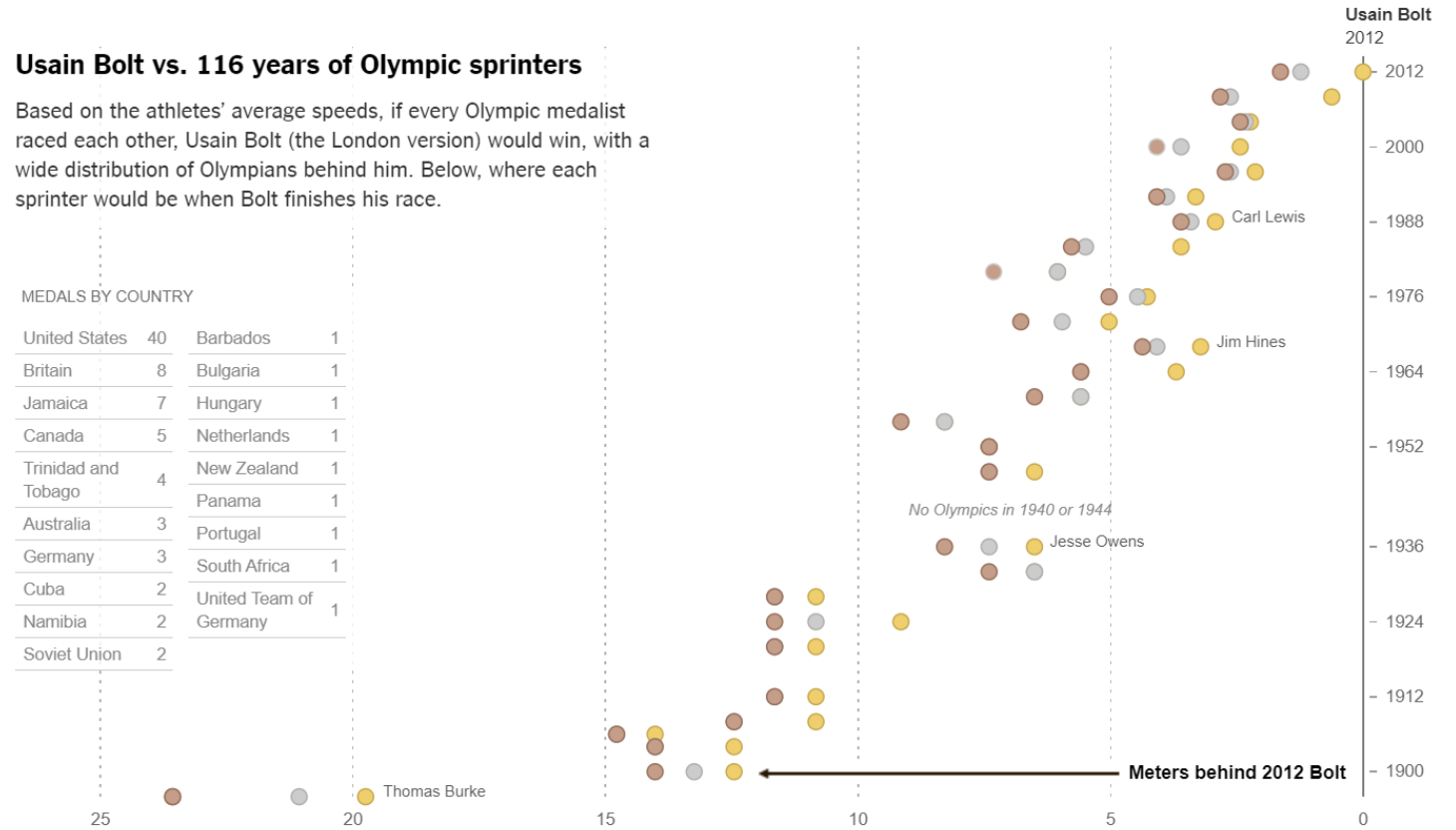
# Visual decoding

## Usain Bolt vs. 116 years of Olympic sprinters

Based on the athletes' average speeds, if every Olympic medalist raced each other, Usain Bolt (the London version) would win, with a wide distribution of Olympians behind him. Below, where each sprinter would be when Bolt finishes his race.

### MEDALS BY COUNTRY

United States	40	Barbados	1
Britain	8	Bulgaria	1
Jamaica	7	Hungary	1
Canada	5	Netherlands	1
Trinidad and Tobago	4	New Zealand	1
Australia	3	Panama	1
Germany	3	Portugal	1
Cuba	2	South Africa	1
Namibia	2	United Team of Germany	1
Soviet Union	2		



This chart includes medals for the United States and Australia in the "Intermediary" Games of 1906, which the I.O.C. does not formally recognize.

[New York Times: One Race, Every Medalist Ever,

[http://archive.nytimes.com/www.nytimes.com/interactive/2012/08/05/sports/olympics/the-100-meter-dash-one-race-every-medalist-ever.html?\\_r=1](http://archive.nytimes.com/www.nytimes.com/interactive/2012/08/05/sports/olympics/the-100-meter-dash-one-race-every-medalist-ever.html?_r=1)]

# Quality evaluation

Expressiveness

Effectiveness

---

- *Expressiveness principle*: the visual representation should represent all of, and only, the relationships that exist in the data
  - Consider the lie factor: a value to describe the relation between the size of effect shown in a graphic and the size of effect shown in the data [Edward Tufte]
  - Do not present information that is not in the data

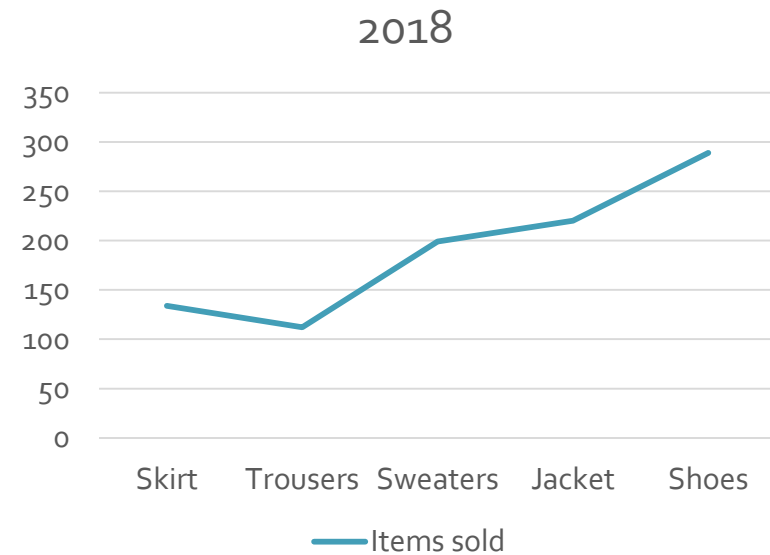


# Quality evaluation

Expressiveness

Effectiveness

- *Expressiveness principle*: the visual representation should represent all of, and only, the relationships that exist in the data
  - Consider the lie factor: a value to describe the relation between the size of effect shown in a graphic and the size of effect shown in the data [Edward Tufte]
  - Do not present information that is not in the data

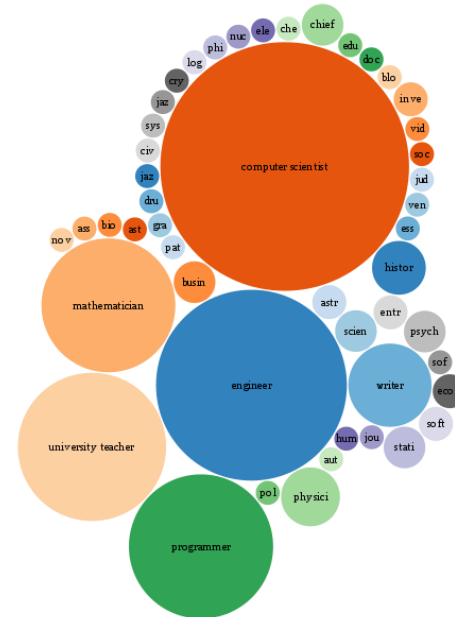


# Quality evaluation

Expressiveness

Effectiveness

- *Expressiveness principle*: the visual representation should represent all of, and only, the relationships that exist in the data
  - Consider the lie factor: a value to describe the relation between the size of effect shown in a graphic and the size of effect shown in the data [Edward Tufte]
  - Do not present information that is not in the data



# *Quality evaluation*

Expressiveness

Effectiveness

---

- *Effectiveness principle*: the relevance of information displayed should match the effectiveness of the channel
  - relevant information should be prioritized then encoded with the most effective/accurate channels



# Quality evaluation

- *Effectiveness principle*: the relevance of information displayed should match the effectiveness of the channel
  - relevant information should be prioritized then encoded with the most effective/accurate channels

Expressiveness

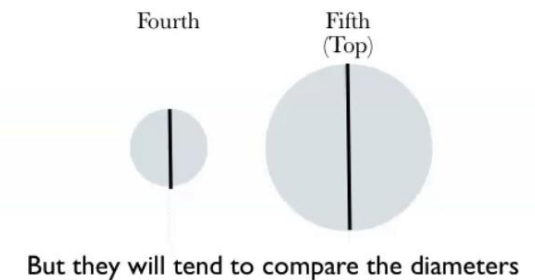
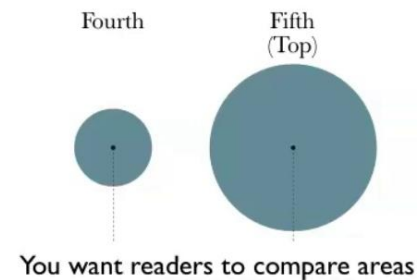
Effectiveness

*How much you would save if federal income taxes were reduced 20%*



SOURCE: The New York Times

GRAPHIC: ACME

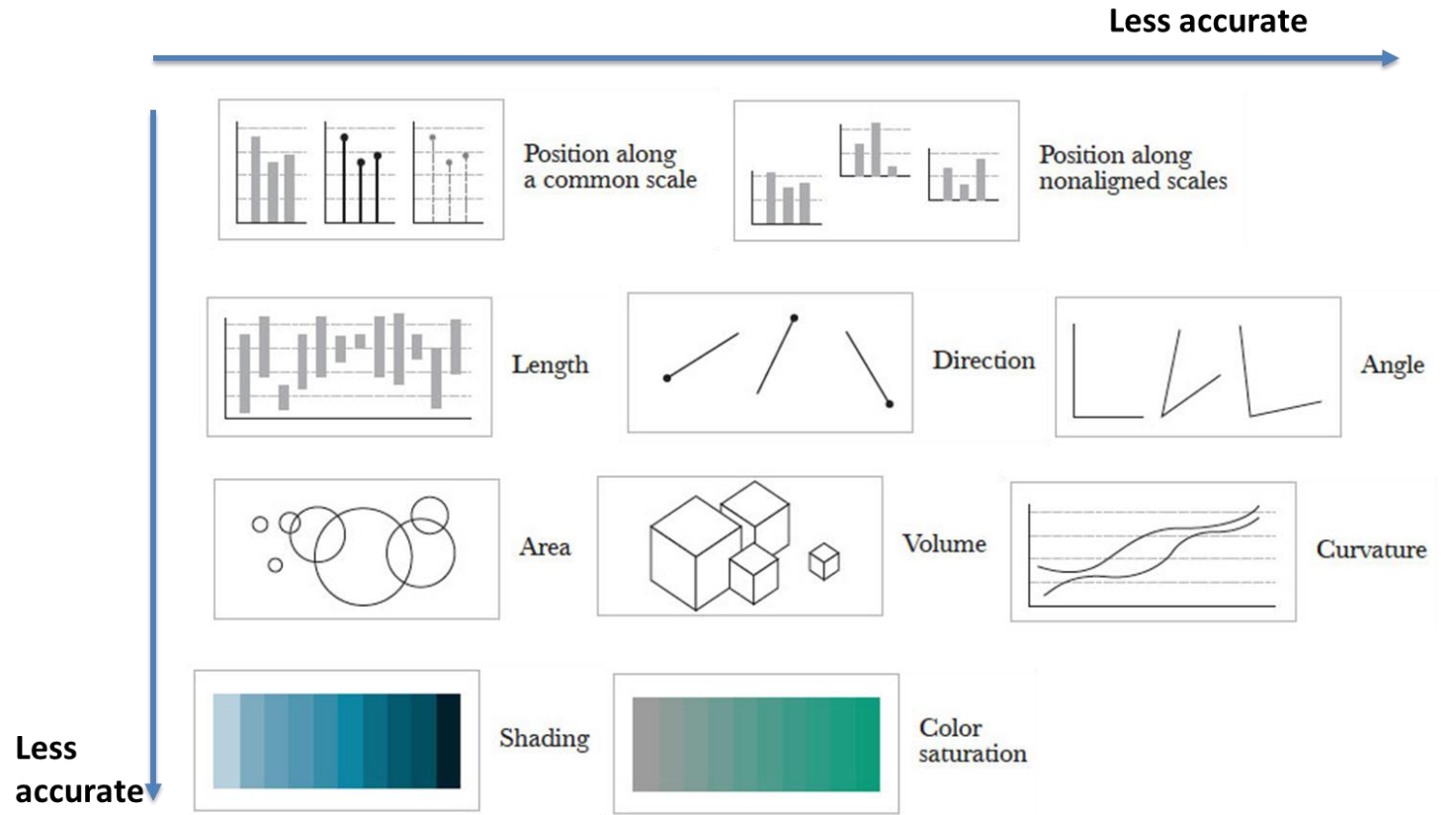


[A. Cairo, *The functional art*]




# Perceptual accuracy

Quantitative attributes



[Cleveland and McGill, 1984 – Adapted from A. Cairo, *The functional art*]

# *Perceptual accuracy*



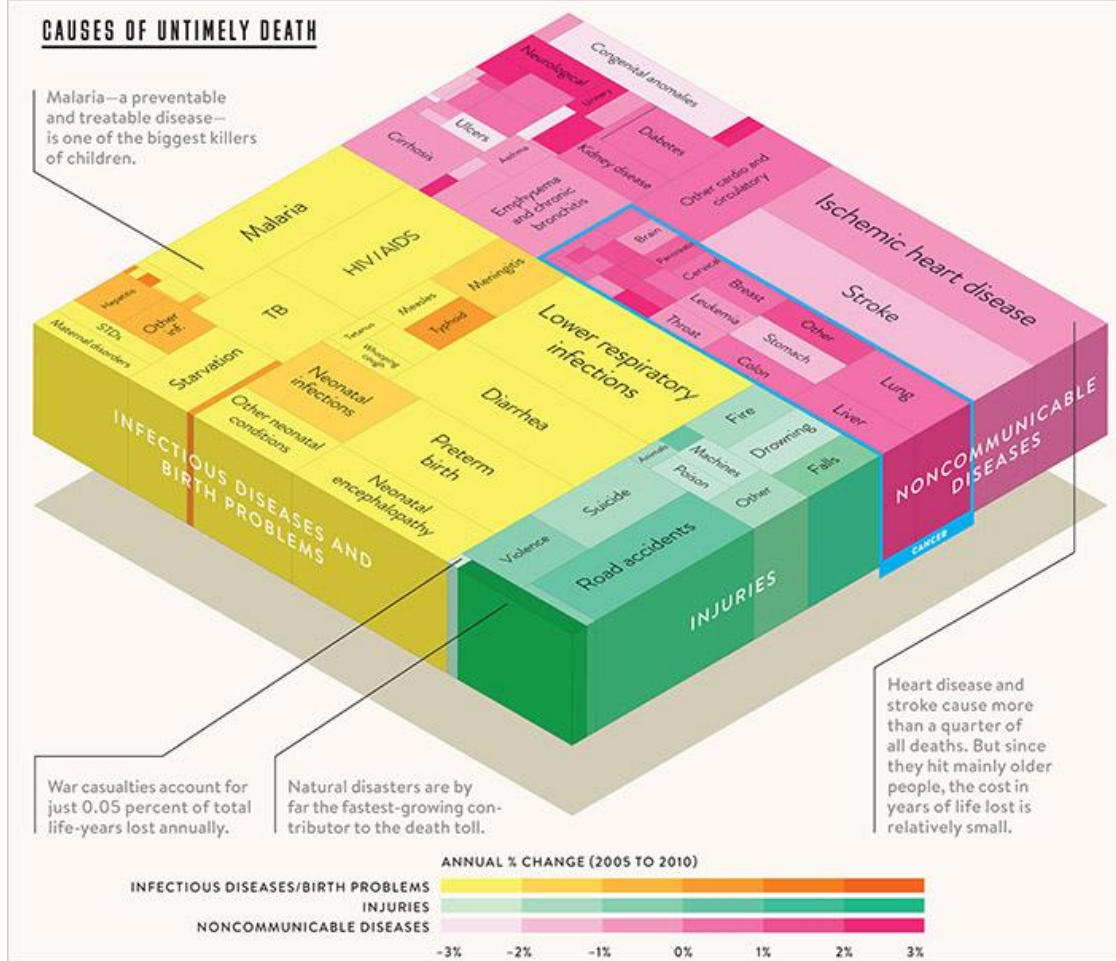
<b>QUANTITATIVE</b>	<b>ORDINAL</b>	<b>CATEGORICAL</b>
Position	Position	Position
Length	Density	Hue
Angle	Saturation	Texture
Slope	Hue	Connection
Area	Texture	Containment
Volume	Connection	Density
Density	Containment	Saturation
Saturation	Length	Shape
Hue	Angle	Length
	Slope	Angle
	Area	Slope
	Volume	Area
		Volume

less accurate

[Data Visualization Course by John C. Hart, for Coursera, 2015]

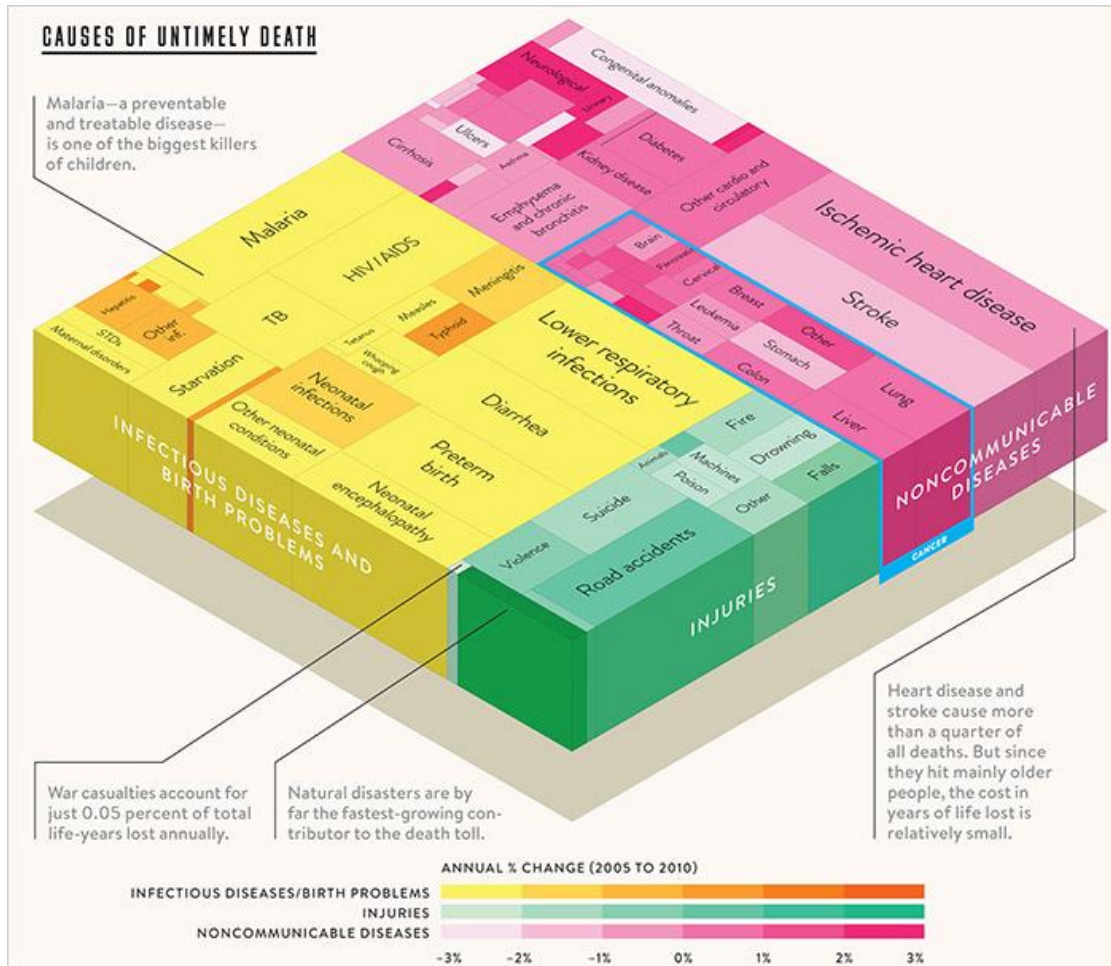
# *Evaluation examples*





"this graph [...] shows that while the number of people dying from communicable diseases is still far too high, those numbers continue to come down. In fact, fewer kids are dying, more kids are going to school and more diseases are on their way to being eliminated. But there remains much to do to cut down the deaths in that yellow block even more dramatically." [Bill Gates, Wired Magazine]

[Example taken from <http://www.perceptualedge.com/>]



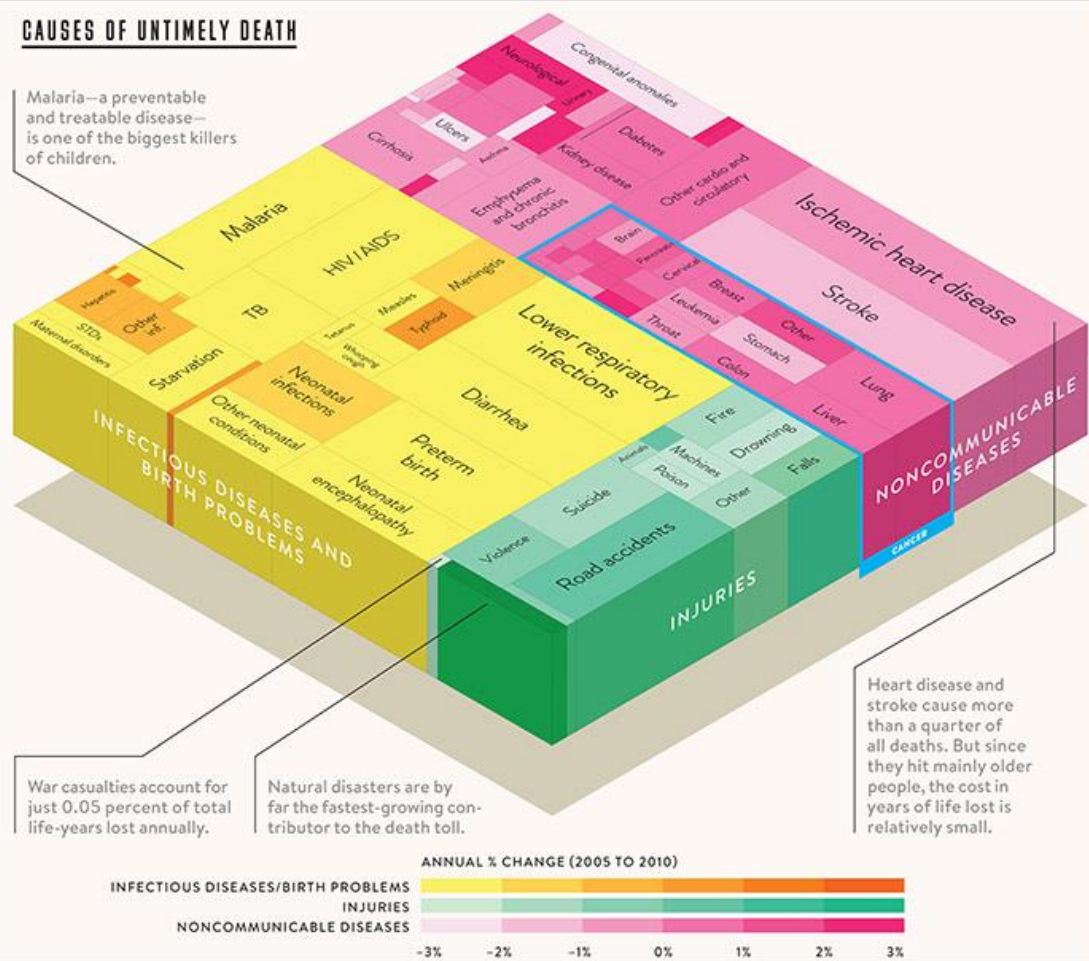
"this graph [...] shows that while the number of people dying from communicable diseases is still far too high, those numbers continue to come down. In fact, fewer kids are dying, more kids are going to school and more diseases are on their way to being eliminated. But there remains much to do to cut down the deaths in that yellow block even more dramatically." [Bill Gates, Wired Magazine]

- The size of rectangles represent years of life lost
- The color hue represents groups of causes
- The color intensity represents the change from one year to another



## CAUSES OF UNTIMELY DEATH

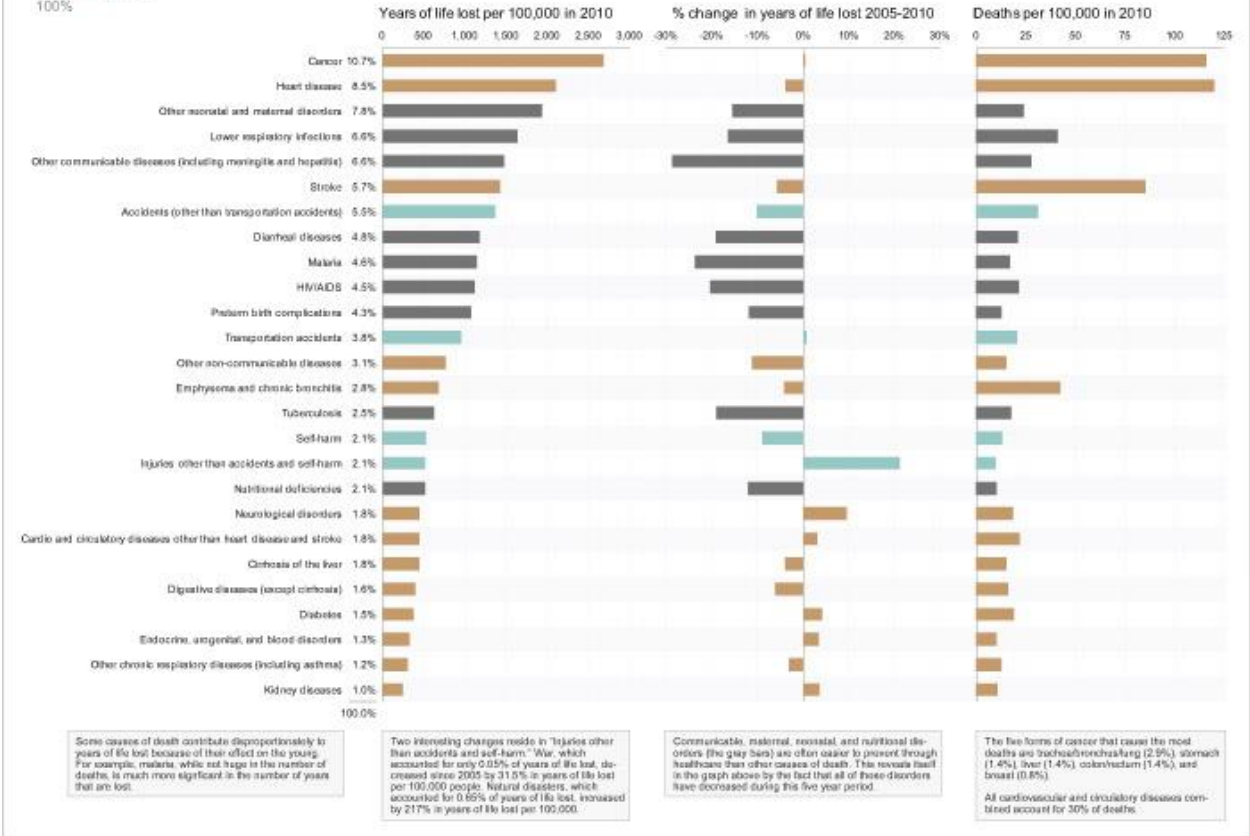
Malaria—a preventable and treatable disease—is one of the biggest killers of children.



## Global Causes of Lost Life

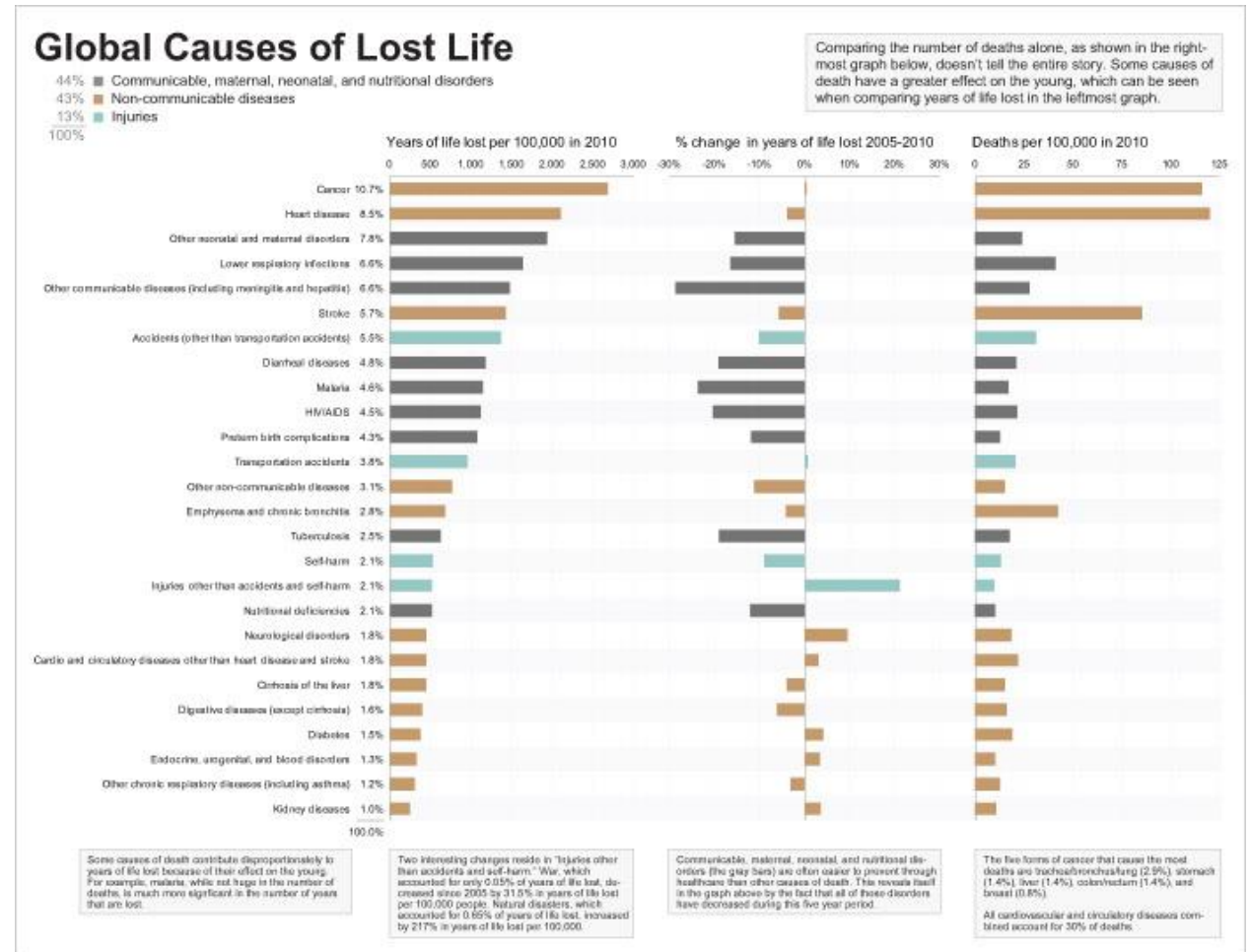
44% Communicable, maternal, neonatal, and nutritional disorders  
 43% Non-communicable diseases  
 13% Injuries  
 100%

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



"this graph [...] shows that while the number of people dying from communicable diseases is still far too high, those numbers continue to come down. In fact, fewer kids are dying, more kids are going to school and more diseases are on their way to being eliminated. But there remains much to do to cut down the deaths in that yellow block even more dramatically. " [Bill Gates, Wired Magazine]

- Three aligned bar charts (years of life lost for every single cause, amount of change, absolute values of deaths)
- Better representation of both quantities and changes, and proper space for labels





**CONSERVATIVE ASSET ALLOCATION MODEL.** Conservative investors tend to be more interested in safety of principal, liquidity and income, rather than in long-term growth or capital appreciation. These investors are willing to accept lower returns for the potential to reduce volatility.

6% International  
2% Small Cap  
2% Mid Cap  
8% Large Cap  
2% Real Estate Securities



5% High Yield Bonds  
60% Bonds  
15% Cash/Cash Equivalents

**MODERATELY CONSERVATIVE ASSET ALLOCATION MODEL.** Moderately conservative investors are interested in safety of principal, liquidity, and income, but also seek modest growth in the value of their investments. These investors are willing to take on a little more risk to achieve that growth, with the understanding that it may increase volatility.

12% International  
3% Small Cap  
5% Mid Cap  
17% Large Cap  
3% Real Estate Securities

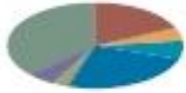


4% High Yield Bonds  
56% Bonds

*An alternative would be to use a multi-asset choice, in all or in part, to achieve a similar risk profile.*

**BALANCED ASSET ALLOCATION MODEL.** Balanced investors are equally interested in safety of principal and long-term growth. These investors generally want steady and sustained growth without the volatility that high-risk investments can bring.

18% International  
5% Small Cap  
7% Mid Cap  
25% Large Cap  
5% Real Estate Securities



3% High Yield Bonds  
37% Bonds

*An alternative would be to use a multi-asset choice, in all or in part, to achieve a similar risk profile.*

**AGGRESSIVE ASSET ALLOCATION MODEL.** Aggressive investors are primarily interested in long-term growth and are willing to take reasonable risks to achieve it. These investors are comfortable with the volatility that accompanies higher risk investments.

24% International  
6% Small Cap  
10% Mid Cap  
34% Large Cap  
6% Real Estate Securities



1% High Yield Bonds  
19% Bonds

**CONSERVATIVE ASSET ALLOCATION MODEL** Conservative investors tend to be more interested in safety of principal, liquidity and income, rather than in long-term growth or capital appreciation. These investors are willing to accept lower returns for the potential to reduce volatility.

6% International  
2% Small Cap  
2% Mid Cap  
8% Large Cap  
2% Real Estate Securities



5% High Yield Bonds  
60% Bonds  
15% Cash/Cash Equivalents

**MODERATELY CONSERVATIVE ASSET ALLOCATION MODEL** Moderately conservative investors are interested in safety of principal, liquidity, and income, but also seek modest growth in the value of their investments. These investors are willing to take on a little more risk to achieve that growth, with the understanding that it may increase volatility.

12% International  
3% Small Cap  
5% Mid Cap  
17% Large Cap  
3% Real Estate Securities

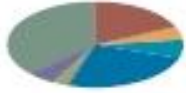


4% High Yield Bonds  
56% Bonds

*An alternative would be to use a multi-asset choice, in all or in part, to achieve a similar risk profile.*

**BALANCED ASSET ALLOCATION MODEL** Balanced investors are equally interested in safety of principal and long-term growth. These investors generally want steady and sustained growth without the volatility that high-risk investments can bring.

18% International  
5% Small Cap  
7% Mid Cap  
25% Large Cap  
5% Real Estate Securities



3% High Yield Bonds  
37% Bonds

*An alternative would be to use a multi-asset choice, in all or in part, to achieve a similar risk profile.*

**AGGRESSIVE ASSET ALLOCATION MODEL** Aggressive investors are primarily interested in long-term growth and are willing to take reasonable risks to achieve it. These investors are comfortable with the volatility that accompanies higher risk investments.

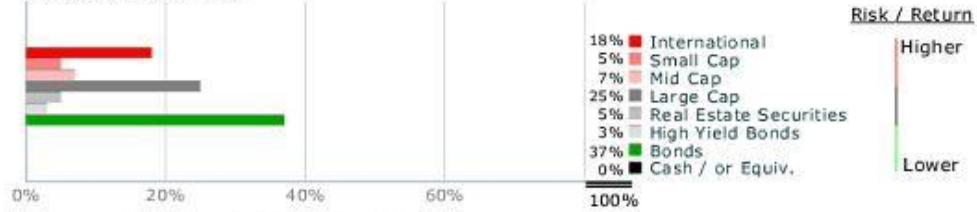
24% International  
6% Small Cap  
10% Mid Cap  
34% Large Cap  
6% Real Estate Securities



1% High Yield Bonds  
19% Bonds

### Balanced Asset Allocation Model

Balanced investors are equally interested in safety of principal and long-term growth. These investors generally want steady and sustained growth without the volatility that high-risk investments can bring.



### Aggressive Asset Allocation Model

Aggressive investors are primarily interested in long-term growth and are willing to take reasonable risks to achieve it. These investors are comfortable with the volatility that accompanies higher risk returns.



**CONSERVATIVE ASSET ALLOCATION MODEL** Conservative investors tend to be more interested in safety of principal, liquidity and income, rather than in long-term growth or capital appreciation. These investors are willing to accept lower returns for the potential to reduce volatility.

6% International  
2% Small Cap  
2% Mid Cap  
8% Large Cap  
2% Real Estate Securities



5% High Yield Bonds  
60% Bonds  
15% Cash/Cash Equivalents

**MODERATELY CONSERVATIVE ASSET ALLOCATION MODEL** Moderately conservative investors are interested in safety of principal, liquidity, and income, but also seek modest growth in the value of their investments. These investors are willing to take on a little more risk to achieve that growth, with the understanding that it may increase volatility.

12% International  
3% Small Cap  
5% Mid Cap  
17% Large Cap  
3% Real Estate Securities

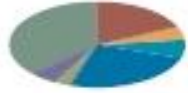


4% High Yield Bonds  
56% Bonds

An alternative would be to use a multi-asset choice, in all or in part, to achieve a similar risk profile.

**BALANCED ASSET ALLOCATION MODEL** Balanced investors are equally interested in safety of principal and long-term growth. These investors generally want steady and sustained growth without the volatility that high-risk investments can bring.

18% International  
5% Small Cap  
7% Mid Cap  
25% Large Cap  
5% Real Estate Securities



3% High Yield Bonds  
37% Bonds

An alternative would be to use a multi-asset choice, in all or in part, to achieve a similar risk profile.

**AGGRESSIVE ASSET ALLOCATION MODEL** Aggressive investors are primarily interested in long-term growth and are willing to take reasonable risks to achieve it. These investors are comfortable with the volatility that accompanies higher risk investments.

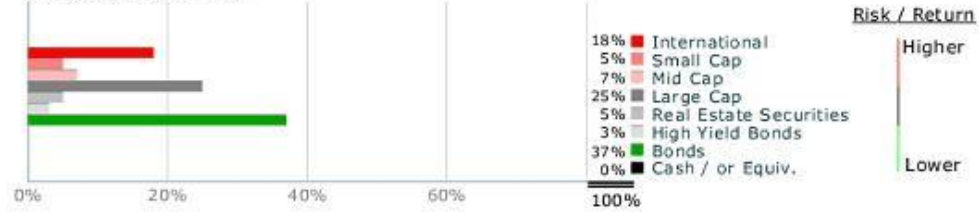
24% International  
6% Small Cap  
10% Mid Cap  
34% Large Cap  
6% Real Estate Securities



1% High Yield Bonds  
19% Bonds

### Balanced Asset Allocation Model

Balanced investors are equally interested in safety of principal and long-term growth. These investors generally want steady and sustained growth without the volatility that high-risk investments can bring.

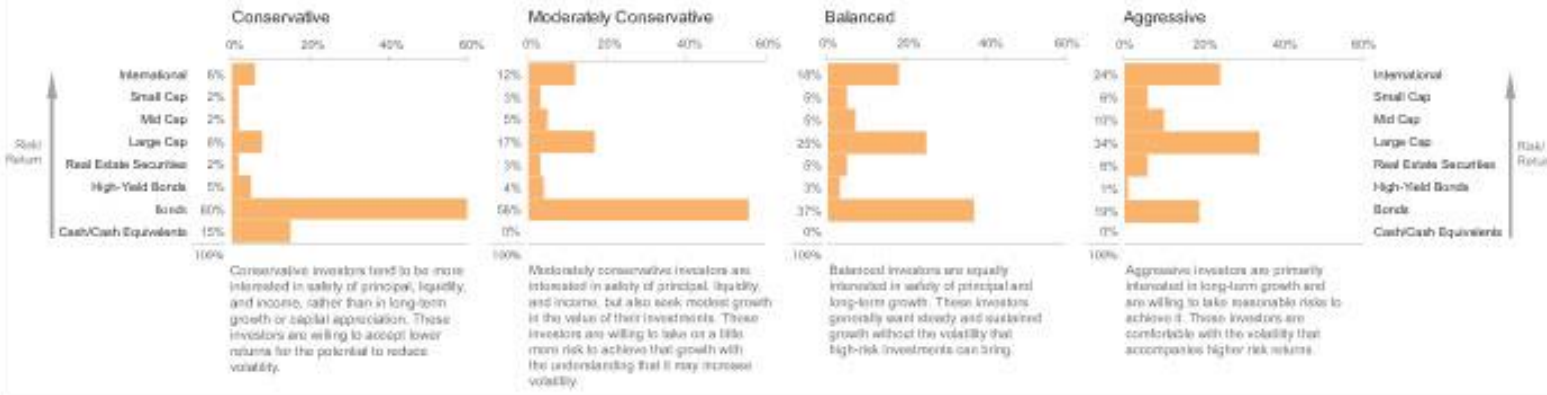


### Aggressive Asset Allocation Model

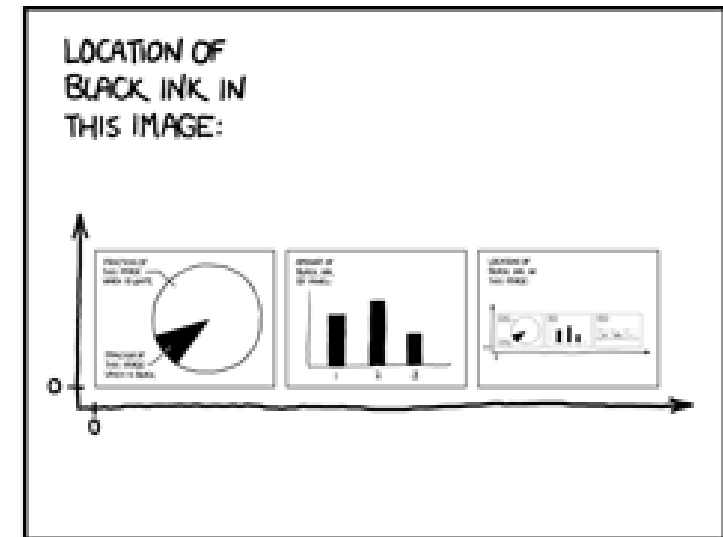
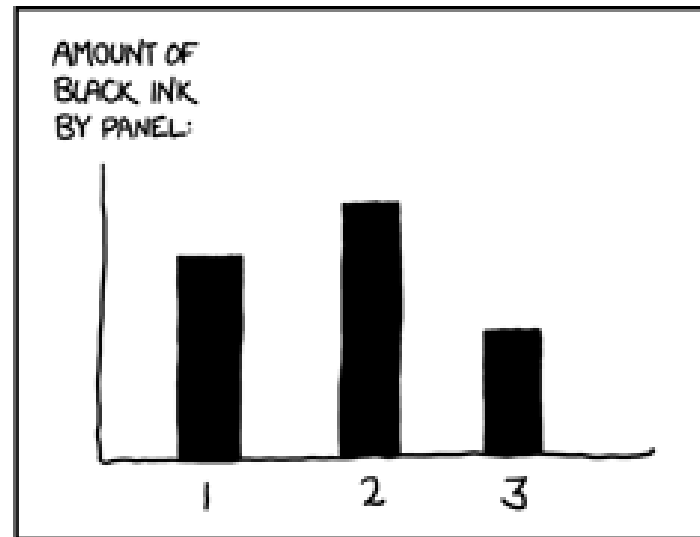
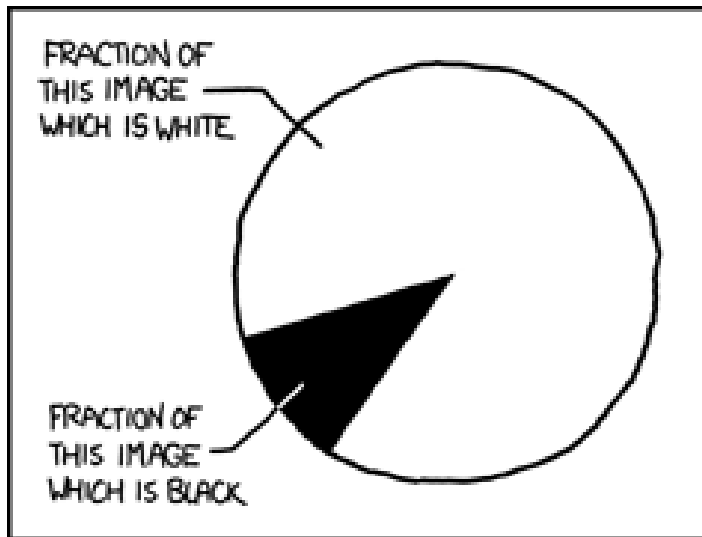
Aggressive investors are primarily interested in long-term growth and are willing to take reasonable risks to achieve it. These investors are comfortable with the volatility that accompanies higher risk returns.



### Asset Allocation Models



# Questions?



<https://xkcd.com/688/>