

SCIENTIFIC AND LARGE DATA VISUALIZATION

October 17, 2020

Fundamentals of Information Visualization – Part I

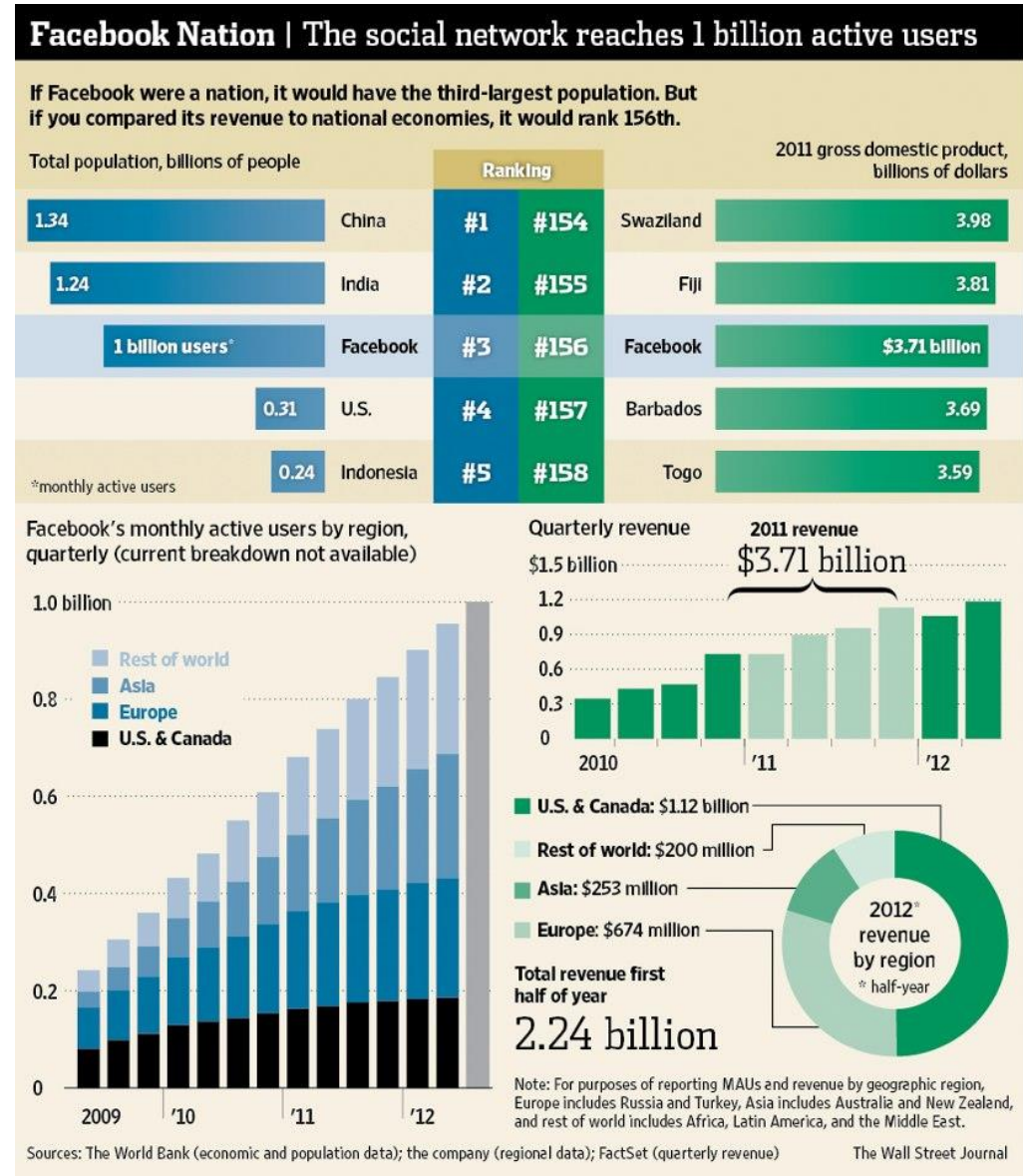
Daniela Giorgi

Visual Computing Lab, CNR-ISTI

An introduction to Information Visualization

Visualization and infographics

For data journalism



[Wall Street Journal]

Visualization and infographics

For data journalism

A LAS PUERTAS DEL CIELO

"La iglesia del Sagrado Corazón de Jesús es una metáfora que se refiere a la existencia humana y su relación con Dios. Esto queda en evidencia a simple vista: a medida que la obra alcanza altura, los personajes (representados por estatuas) están más cerca del cielo". Así la explica Carmelo D'Agostino, el más antiguo de los capuchinos de Córdoba.

En la parte más baja, al nivel del zócalo, se ven arañas, tortugas, escorpiones y sapos, que simbolizan el pecado. Sobre ellos se yerguen columnas con diferentes formas, símbolos de las diversas culturas anteriores a Cristo. En un nivel superior se encuentran los atlantes, que son los hombres que cargan con las torres de la iglesia, es decir con el peso del pecado. Inmediatamente arriba de los atlantes, el arquitecto Augusto Ferrari ubicó a los doce apóstoles, representantes de las virtudes que Dios espera alcancen los seres humanos. Luego están los santos, que lograron esas virtudes y viven en la gloria de Dios. A la misma altura de los santos, está la estatua del Sagrado Corazón. Por último, en lo más alto, en medio de las torres, Ferrari colocó a San Francisco, para que desde ahí custodie a toda la ciudad.

BOVEDA ESTRELLADA

En el cielo raso, están pintadas las constelaciones del firmamento de Córdoba a medianoche.



Fuente: párroco Sebastián Gasmann - Augusto C. Ferrari (1871-1970), de Ediciones Licipodio. Infografía: Juan Colombata y Enzo Oliva.

FICHA TÉCNICA



CONSTRUCCIÓN: 1927-1933 (*)
 ESTILO: Románico y gótico
 INGENIEROS: Giralt, Nores y Olmedo
 ESCULTOR: Ramacciotti
 DECORADOR: Alfaro

(*) Se inauguró en 1933, pero se terminó hasta el último detalle recién en la década de 1960



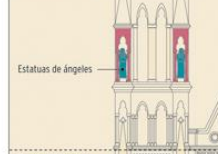
ORNAMENTOS DE LA FACHADA

Combina elementos góticos y románicos. Está construida de manera que, a medida que alcanza altura, los personajes se encuentran más cerca del cielo.



SÍMBOLOS HACIA EL INFINITO

ESPIRITUS CELESTES CRIADOS POR DIOS PARA SU MINISTERIO



LOS QUE LOGRAN LAS VIRTUDES Y VIVEN EN LA GLORIA DE DIOS

Corimisa de cabezas de animales
 Estatua de la caridad

REPRESENTANTES DE LAS VIRTUDES

Apóstoles
 San Pablo
 Mosaico veneciano

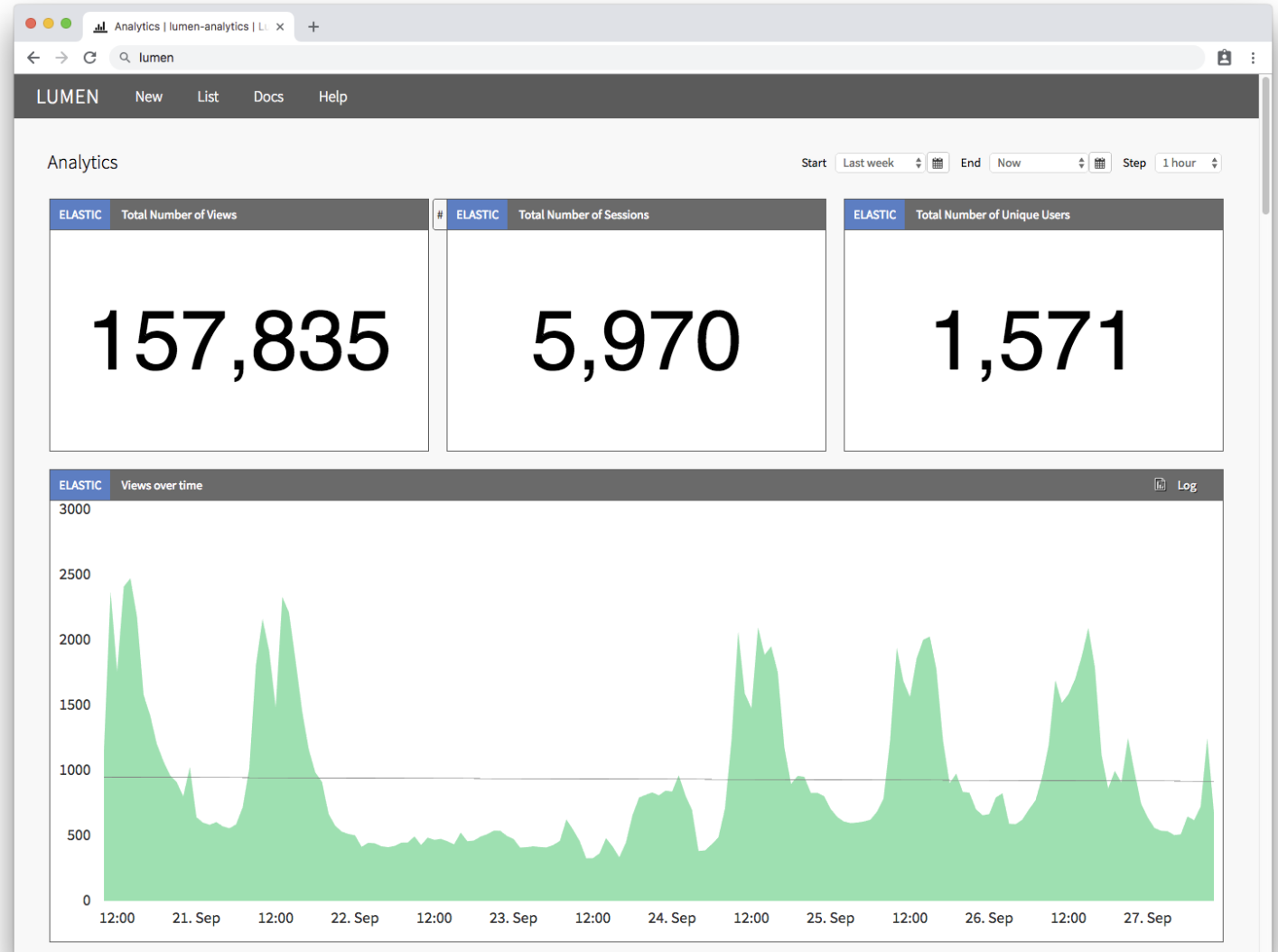
SÍMBOLOS DEL PECADO

Atlantes
 Arañas, tortugas, escorpiones y sapos al pie de las columnas.
 Columnas multiformes: A los costados de los portales, representan las razas de todos los pueblos.
 Columnas lisas: Son para dar la impresión de fuerza.

[Juan Colombata & Enzo Oliva – La Voz del Interior]

Visualization and infographics

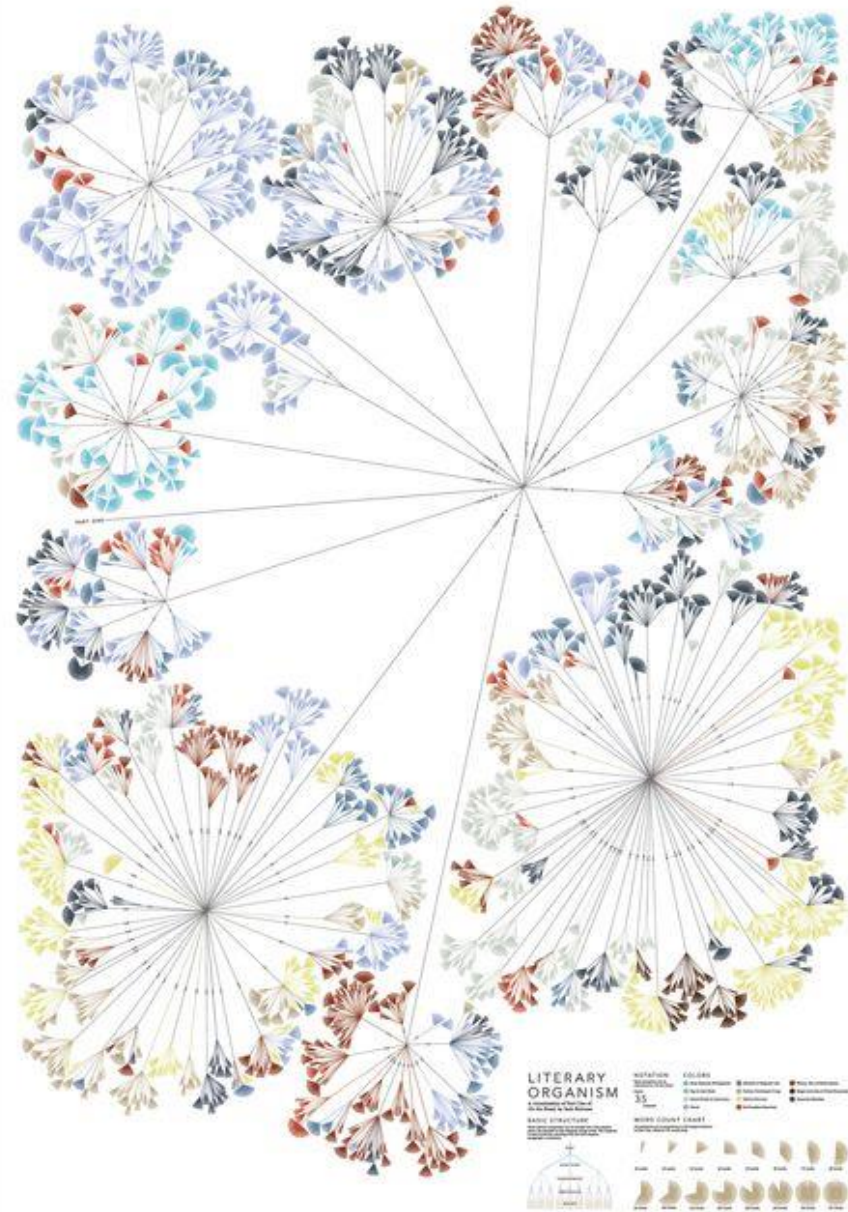
For data science
(and for companies)



[Lumen dashboard by Netflix]

Visualization and infographics

For data science
(and for research)



[Stephanie Posavec – Writing without words
<http://www.stefanieposavec.com/writing-without-words>]

Definitions

- Information visualization is the use of computer-supported, interactive, visual representations of abstract data to amplify cognition [S. T. Card, 1999]

Definitions

- Information visualization is the use of computer-supported, **interactive**, visual representations of **abstract data** to amplify **cognition** [S. T. Card, 1999]

Definitions

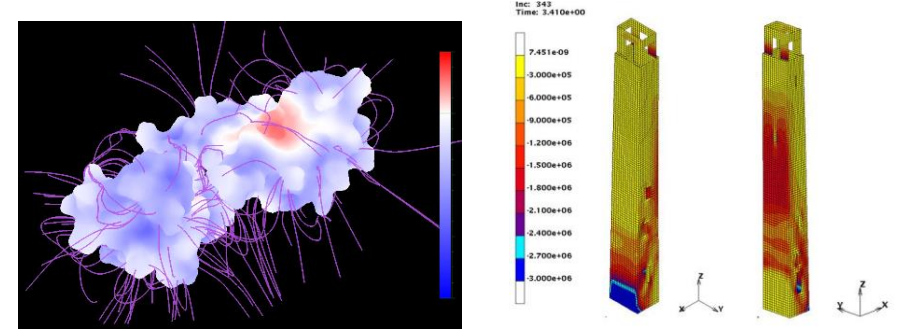
- Information visualization is the use of computer-supported, **interactive**, visual representations of **abstract data** to amplify **cognition** [S. T. Card, 1999]
- To visualize is to make certain phenomena and portions of reality visible and understandable; many of these phenomena are not naturally accessible to the bare eye, and many of them are **not even of visual nature** [J. Costa, 1998]

Related disciplines

Scientific visualization

(Learn more with guest lecturers M. Callieri and M. Zoppé)

- *Scientific visualization* concerns the visualization of scientific data and phenomena (usually having a spatial representation)

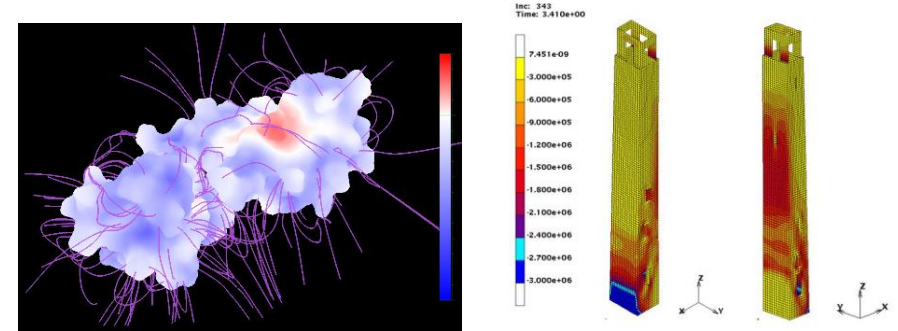


Related disciplines

Scientific visualization

(Learn more with guest lecturers M. Callieri and M. Zoppé)

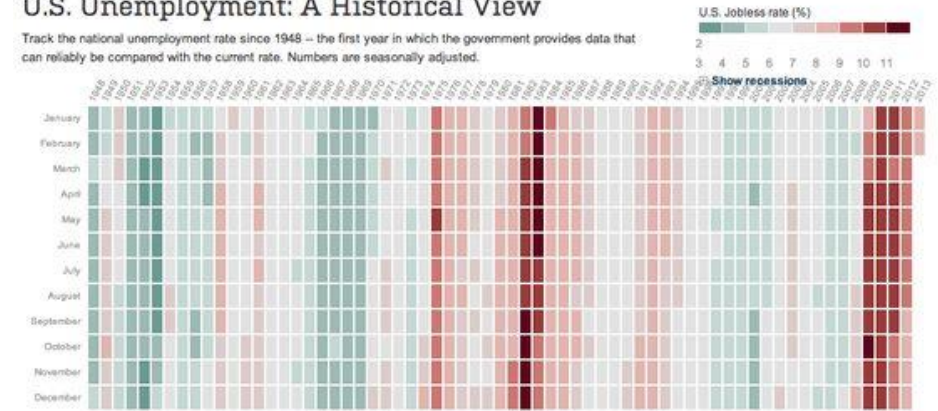
- *Scientific visualization* concerns the visualization of scientific data and phenomena (usually having a spatial representation)



- In scientific visualization there is a *natural* relationship between what is represented and its representation, whereas in information visualization the relationship is *conventional*

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

Related disciplines

Visual analytics

- *Visual analytics* is an outgrowth of the fields of information visualization and scientific visualization that focuses on analytical reasoning facilitated by interactive visual interfaces [Wikipedia]
- Visual analytics brings together several scientific and technical communities from computer science, information visualization, cognitive and perceptual sciences, interactive design, graphic design, and social sciences
- The accent is more on reasoning and inference

Why visualization

Numerical vs graphical
communication

Unemployment rate (%)

	CURRENT	Historical maximum	Historical minimum
Alabama	6.7	14.4	3.3
Alaska	7.5	11.5	5.9
Arizona	6.9	11.5	3.6
Arkansas	6.2	10.2	4.1
California	9.3	11.0	4.7
Colorado	6.1	9.1	2.5
Connecticut	7.1	10.0	2.1
Delaware	6.1	8.2	2.9
Florida	8.1	9.7	3.3

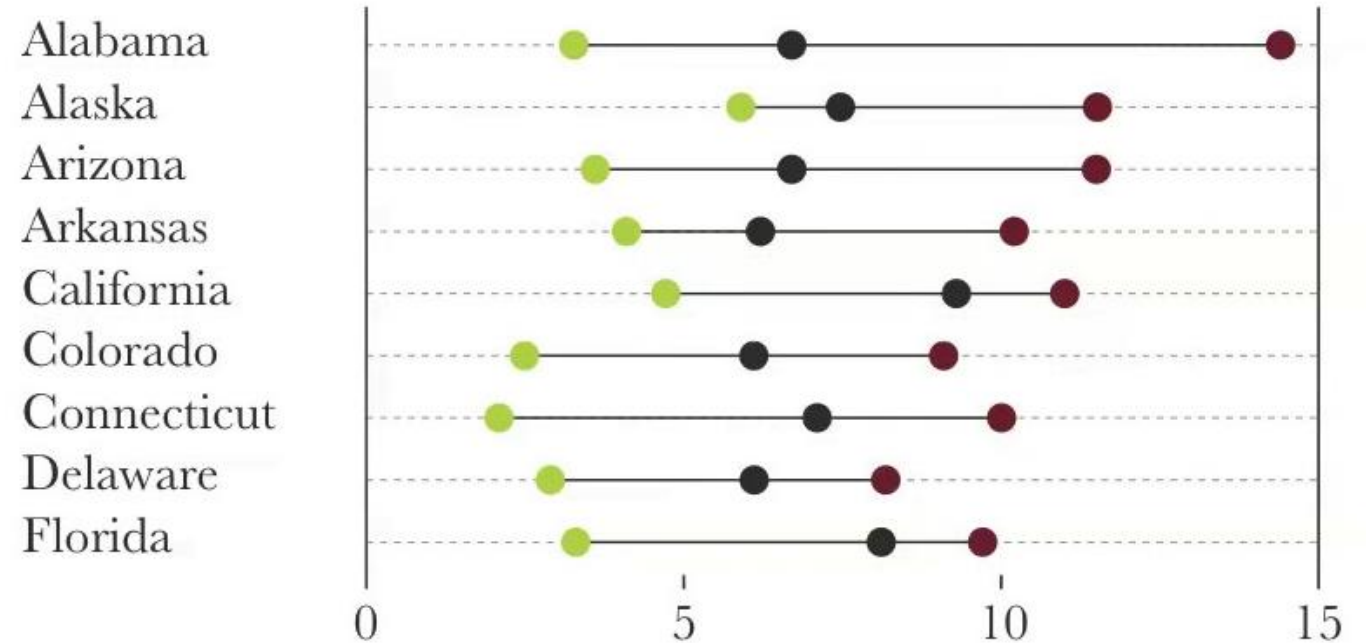
- Which country has the closest unemployment rate to its historical minimum?

Why visualization

Numerical vs graphical
communication

Unemployment rate (%)

● Current ● Historical maximum ● Historical minimum



- Which country has the closest unemployment rate to its historical minimum?

Why visualization

The processing of information in
the human brain

- We are a visual species. The human visual system is very good at identifying and analyzing patterns
- Visualizations are like *cognitive artifacts* (man-made things that aid or enhance our cognitive abilities [Norman, 1991])



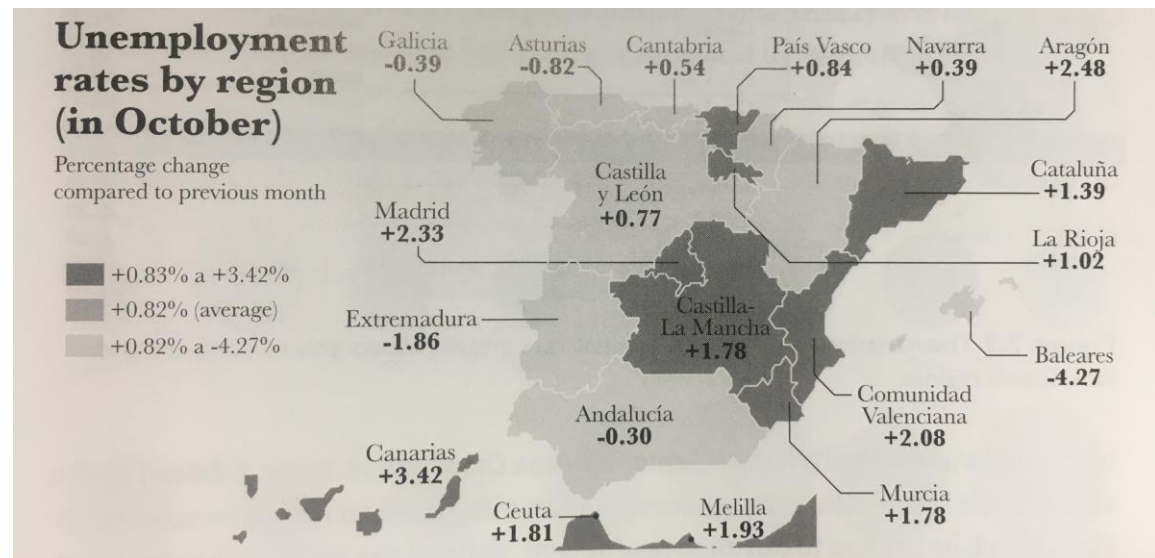
- Visualization has to do with *distributed cognition* (our cognitive system is not only made up of our brain, mind, and sensors, but also of the environment around us, which we use to store and manipulate information, without having to retain all information in our mind)
- Visualizations enable *parallel* processing of information, instead of *sequential*

Why visualization

- Infographics is a visual tool for communication, for the understanding and for the analysis [A. Cairo, *The functional art*, 2013]
- Visualization is a *technology* [A. Cairo, *ibid.*]
 - Function constraints form: as the design of a technological object must depend on the task it should help with, the graphics form should be constrained by the function of the presentation

Why visualization

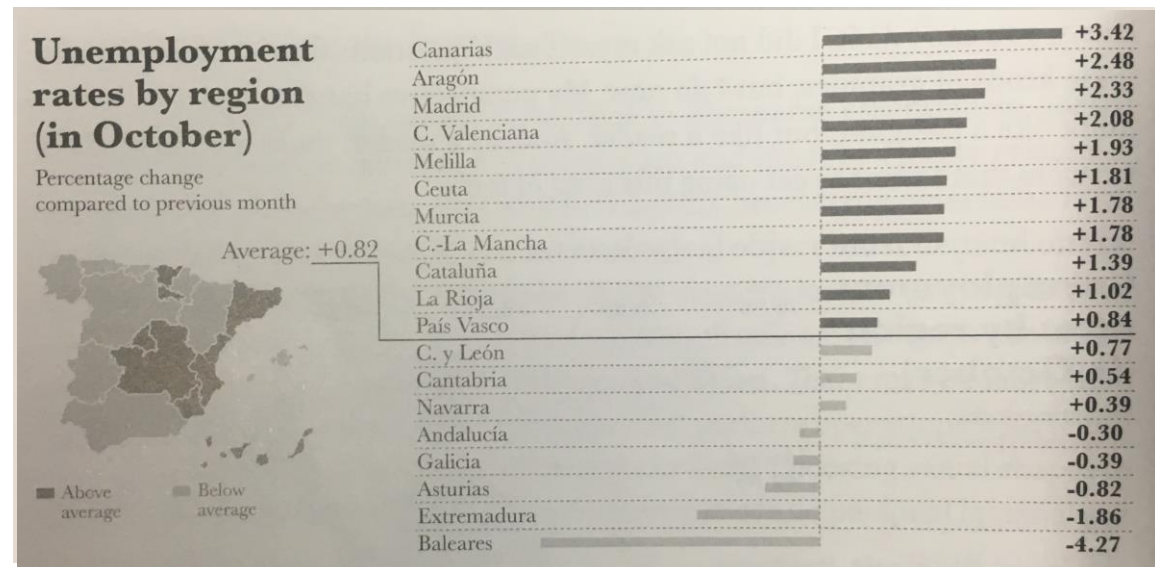
- Infographics is a visual tool for communication, for the understanding and for the analysis [A. Cairo, *The functional art*, 2013]
- Visualization is a *technology* [A. Cairo, *ibid.*]
 - Function constraints form: as the design of a technological object must depend on the task it should help with, the graphics form should be constrained by the function of the presentation



[A. Cairo, *The Functional Art*, pag. 37]

Why visualization

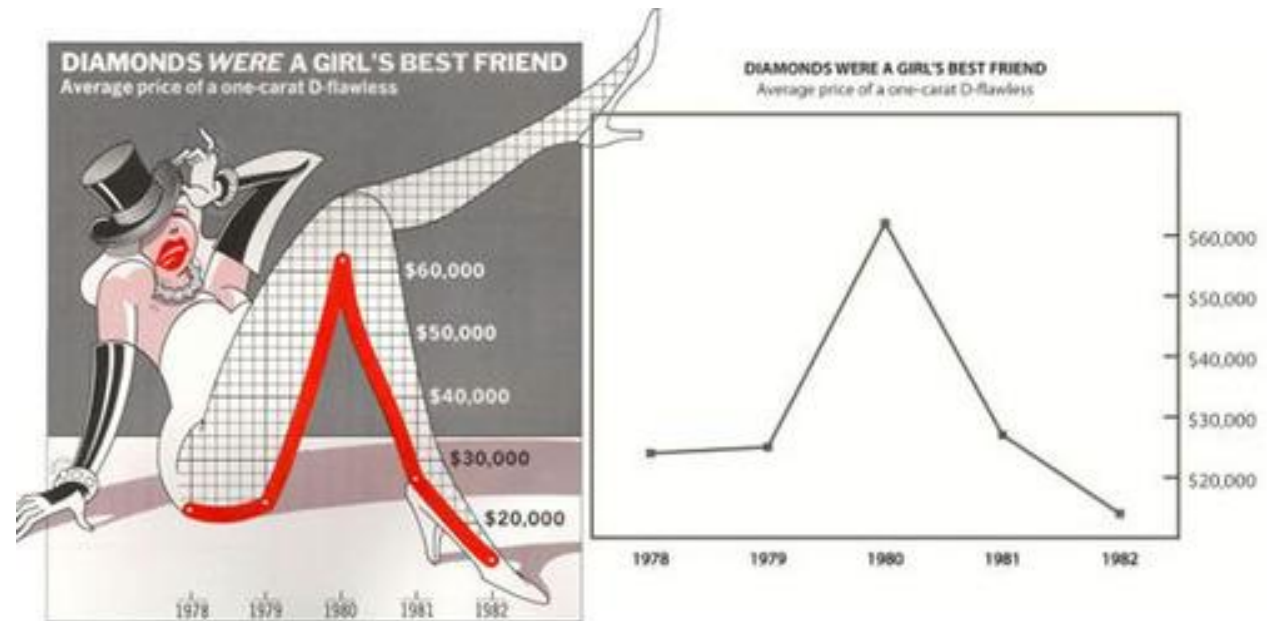
- Infographics is a visual tool for communication, for the understanding and for the analysis [A. Cairo, *The functional art*, 2013]
- Visualization is a *technology* [A. Cairo, *ibid.*]
 - Function constraints form: as the design of a technological object must depend on the task it should help with, the graphics form should be constrained by the function of the presentation



[A. Cairo, *The Functional Art*, pag. 38]

Why visualization

- Infographics is a visual tool for communication, for the understanding and for the analysis [A. Cairo, *The functional art*, 2013]
- Visualization is a *technology* [A. Cairo, *ibid.*]
 - The purpose of information visualization is insight, not pictures [B. Shneiderman]

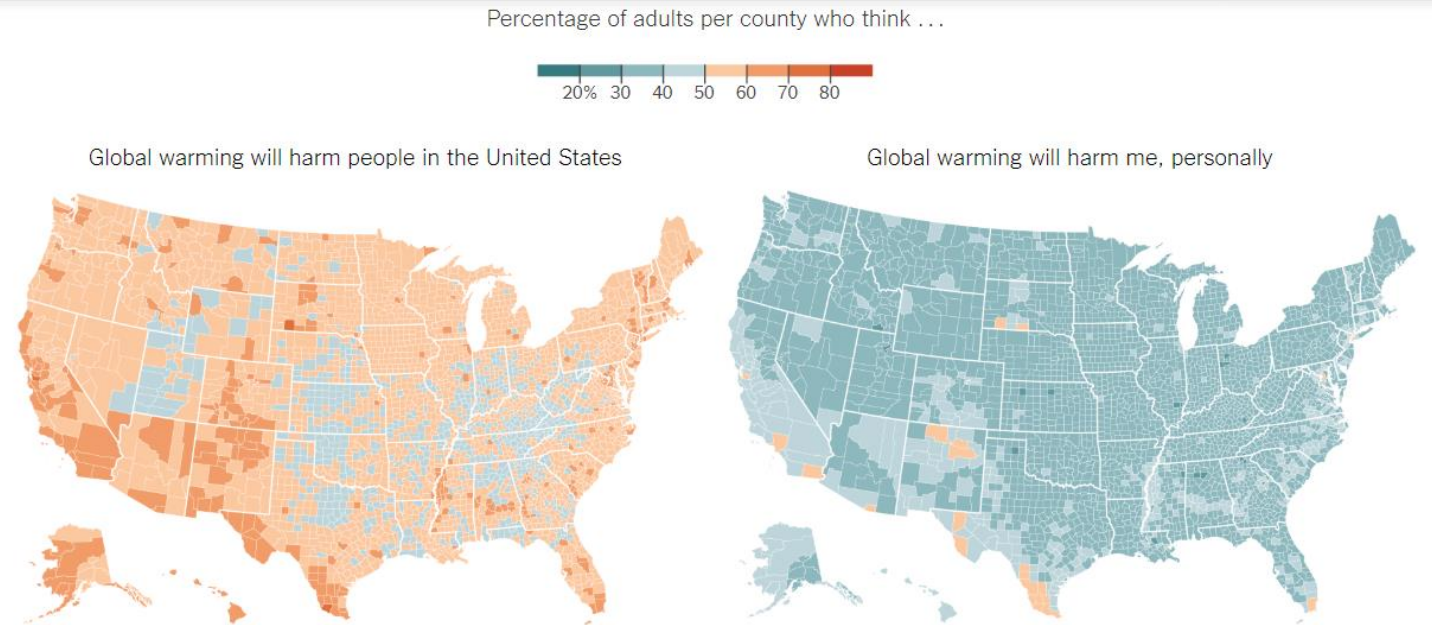


[A long-standing debate: Nigel Holmes' Diamond chart on TIME, and a minimalist version]

Why visualization

- *Explanatory visualizations* are tools to *present* information, *communicate* data and messages, *explain* something to somebody else

Three main purposes:
Explanation, exploration,
confirmation

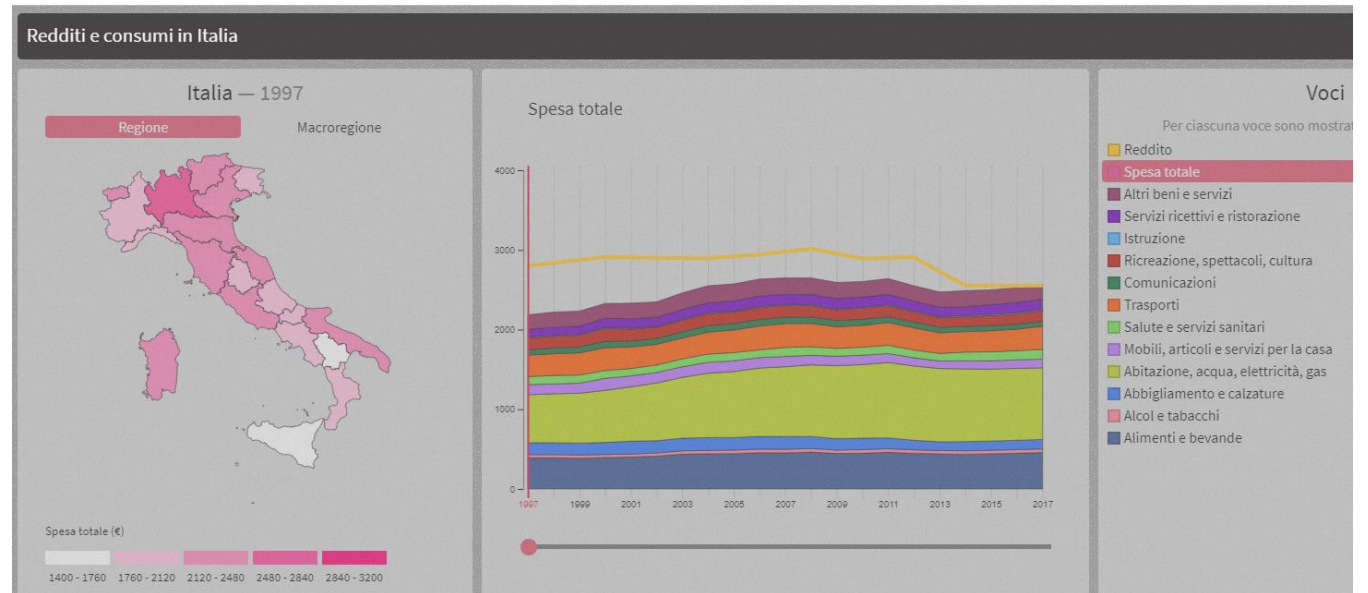


[*New York Times, How Americans think about climate change, in six maps*]

Why visualization

Three main purposes:
Explanation, exploration,
confirmation

- *Exploratory visualizations* are tools to *analyse* what is being presented. Very often, the output of exploratory analysis is not only the answer to the original questions, but the generation of new questions

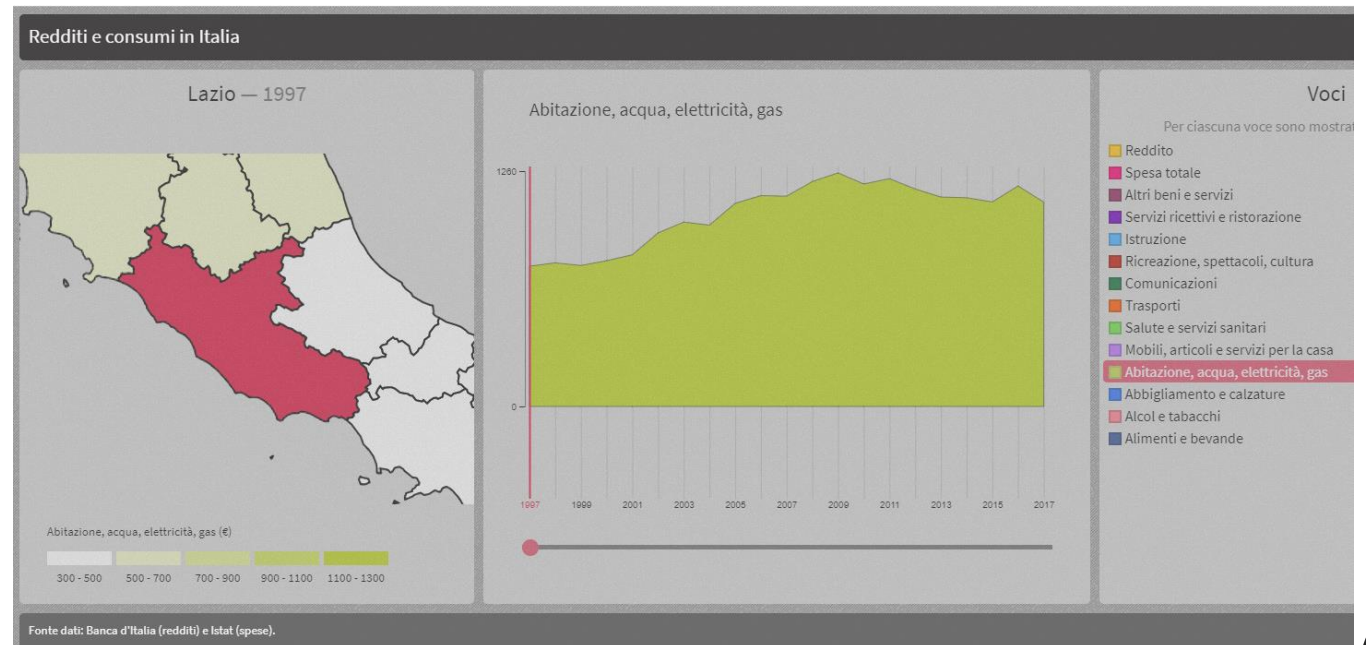


[Matteo Loporchio, AA 2018-2019, <https://mloporchio.github.io/D3IE-Italy/>]

Why visualization

Three main purposes:
Explanation, exploration,
confirmation

- *Exploratory visualizations* are tools to *analyse* what is being presented. Very often, the output of exploratory analysis is not only the answer to the original questions, but the generation of new questions

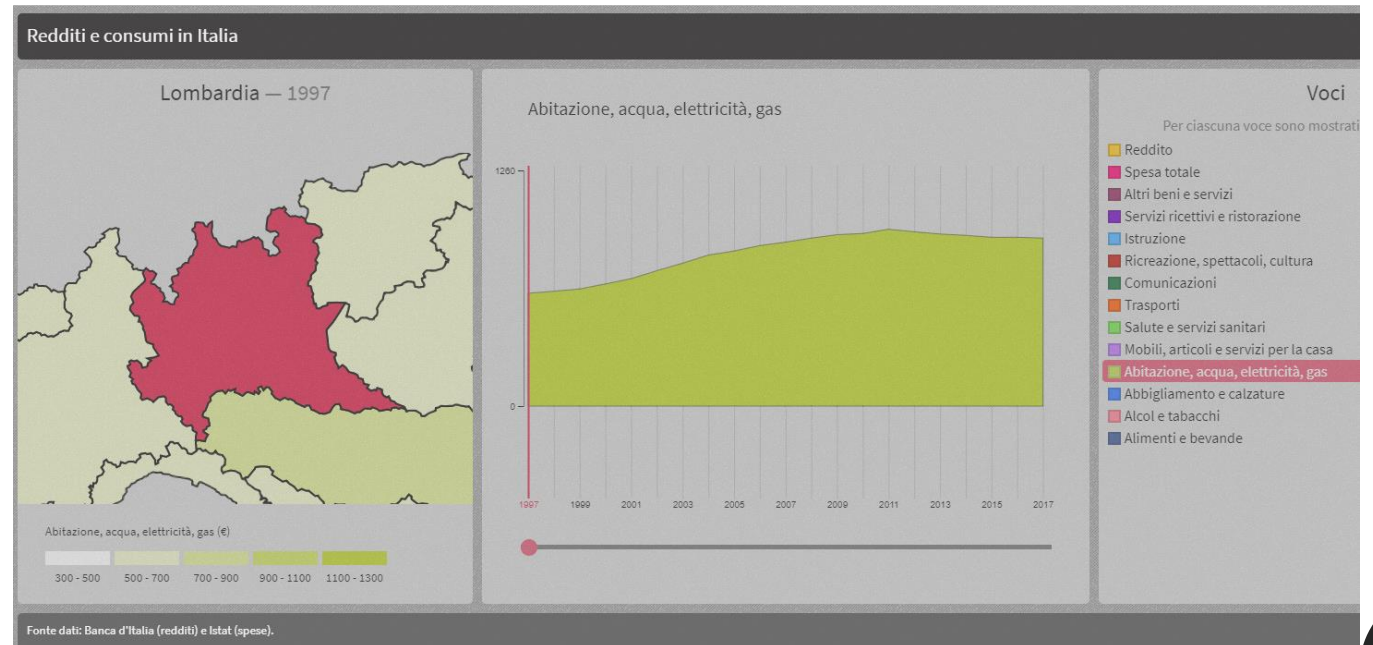


[Matteo Loporchio, AA 2018-2019, <https://mloporchio.github.io/D3IE-Italy/>]

Why visualization

Three main purposes:
Explanation, exploration,
confirmation

- In *confirmatory visualizations*, the reader has an hypothesis, and uses the visualization to see whether his/her hypothesis is true or not



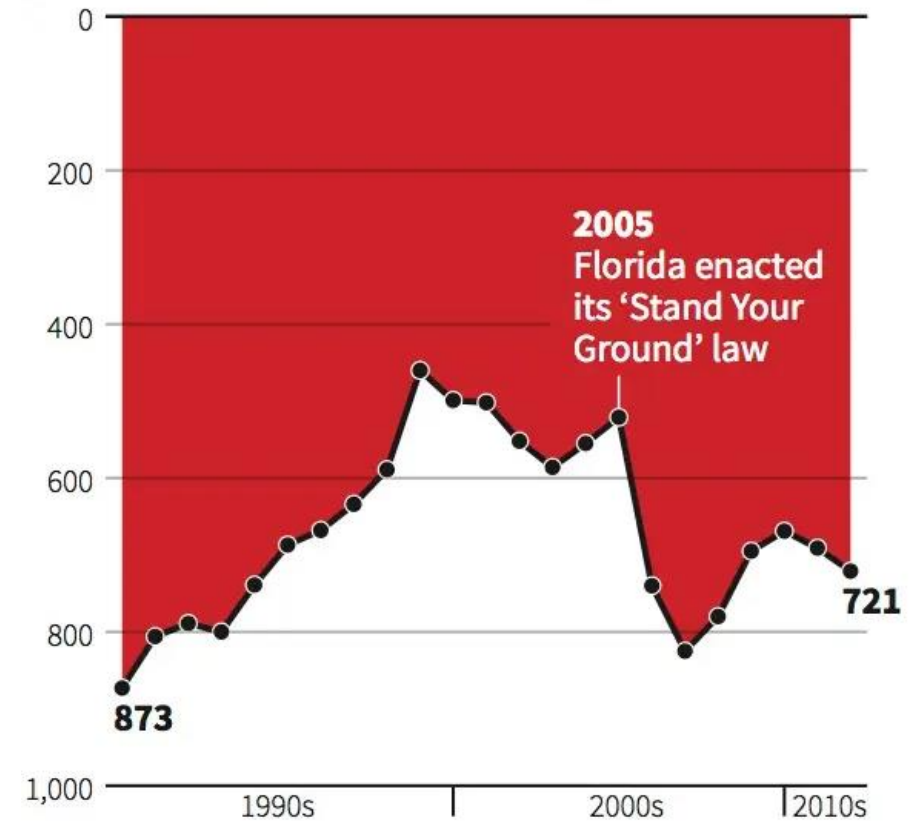
[Matteo Loporchio, AA 2018-2019, <https://mloporchio.github.io/D3IE-Italy/>]

Use and abuse of visualizations

Visualizations can be misleading

Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

[Business Insider, <https://www.businessinsider.com/gun-deaths-in-florida-increased-with-stand-your-ground-2014-2?IR=T>]

Use and abuse of visualizations

Visualizations can be misleading

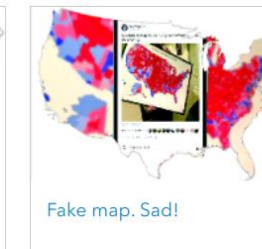
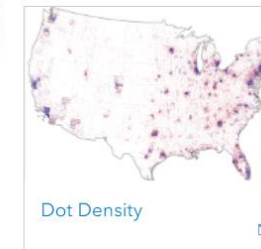
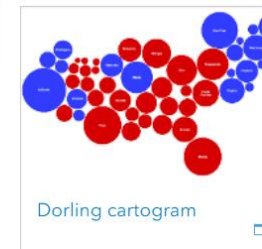
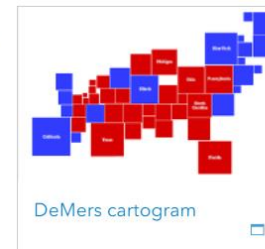
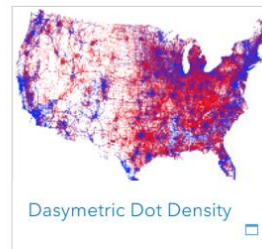
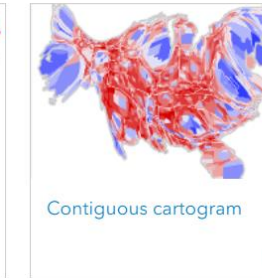
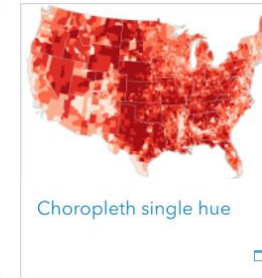
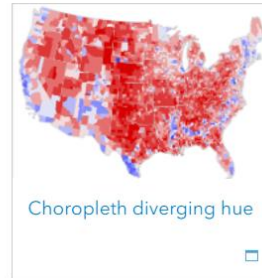
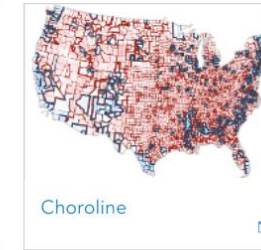
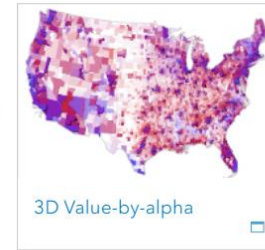
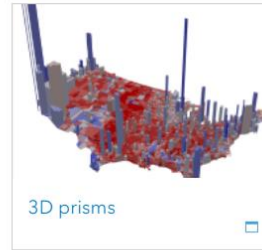


[CNN, <https://edition.cnn.com/2019/10/01/politics/trump-impeach-this-map-fact-check/index.html>]

Use and abuse of visualizations

Fact checking

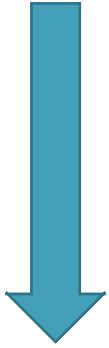
Thematic maps of the 2016 Presidential election (lower 48 states)



[<http://carto.maps.arcgis.com/apps/MinimalGallery/index.html?appid=b3d1fe0e8814480993ff5ad8doc62c32#>]

The visualization pipeline

PHENOMENA

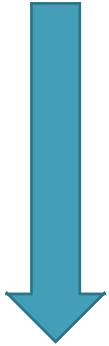


data collection



DATA

PHENOMENA

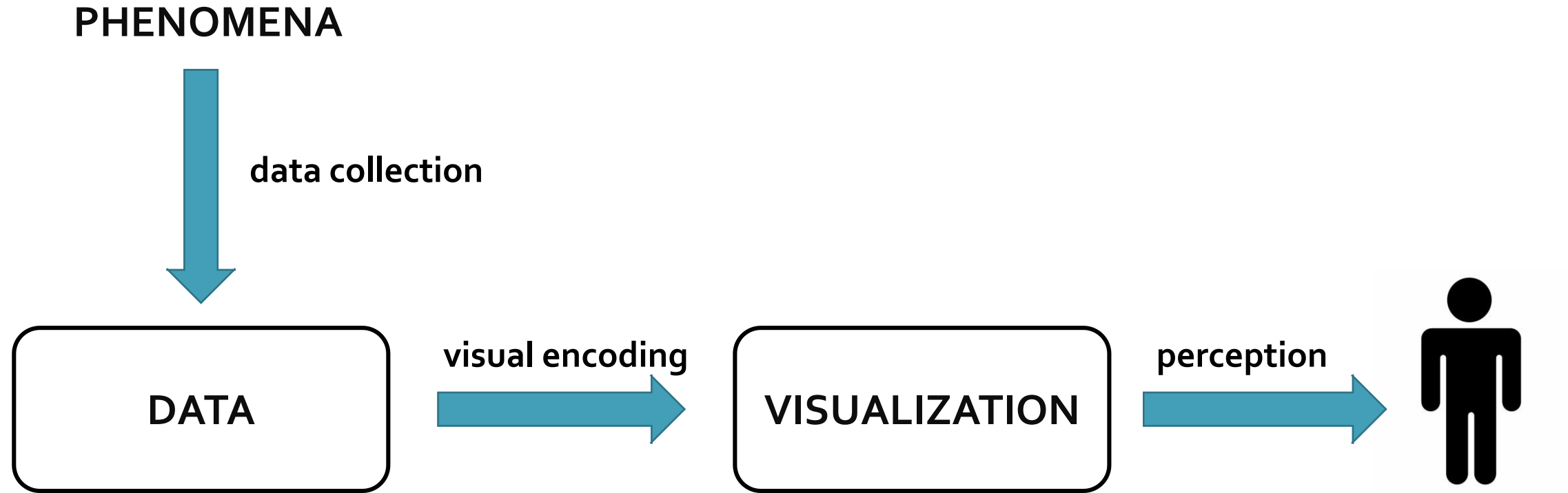


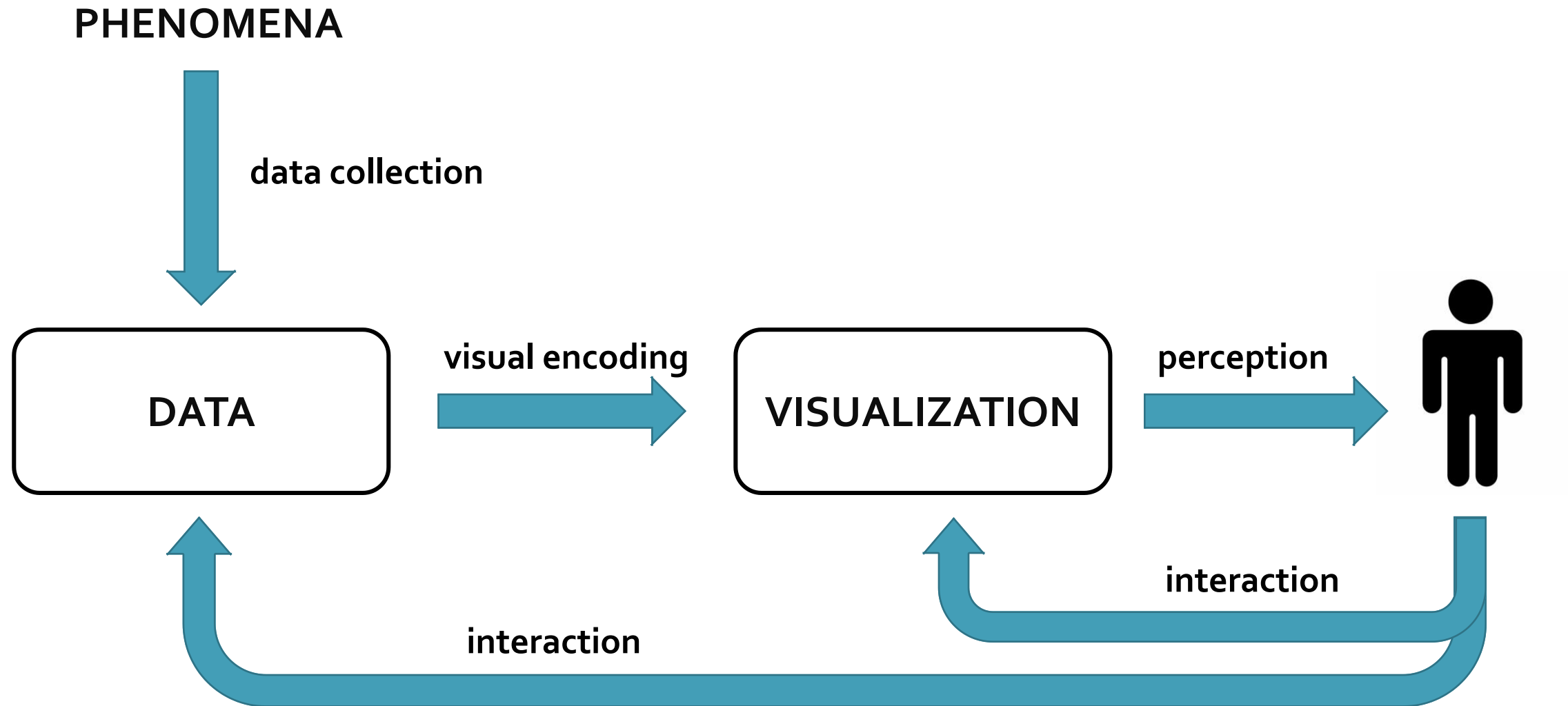
data collection

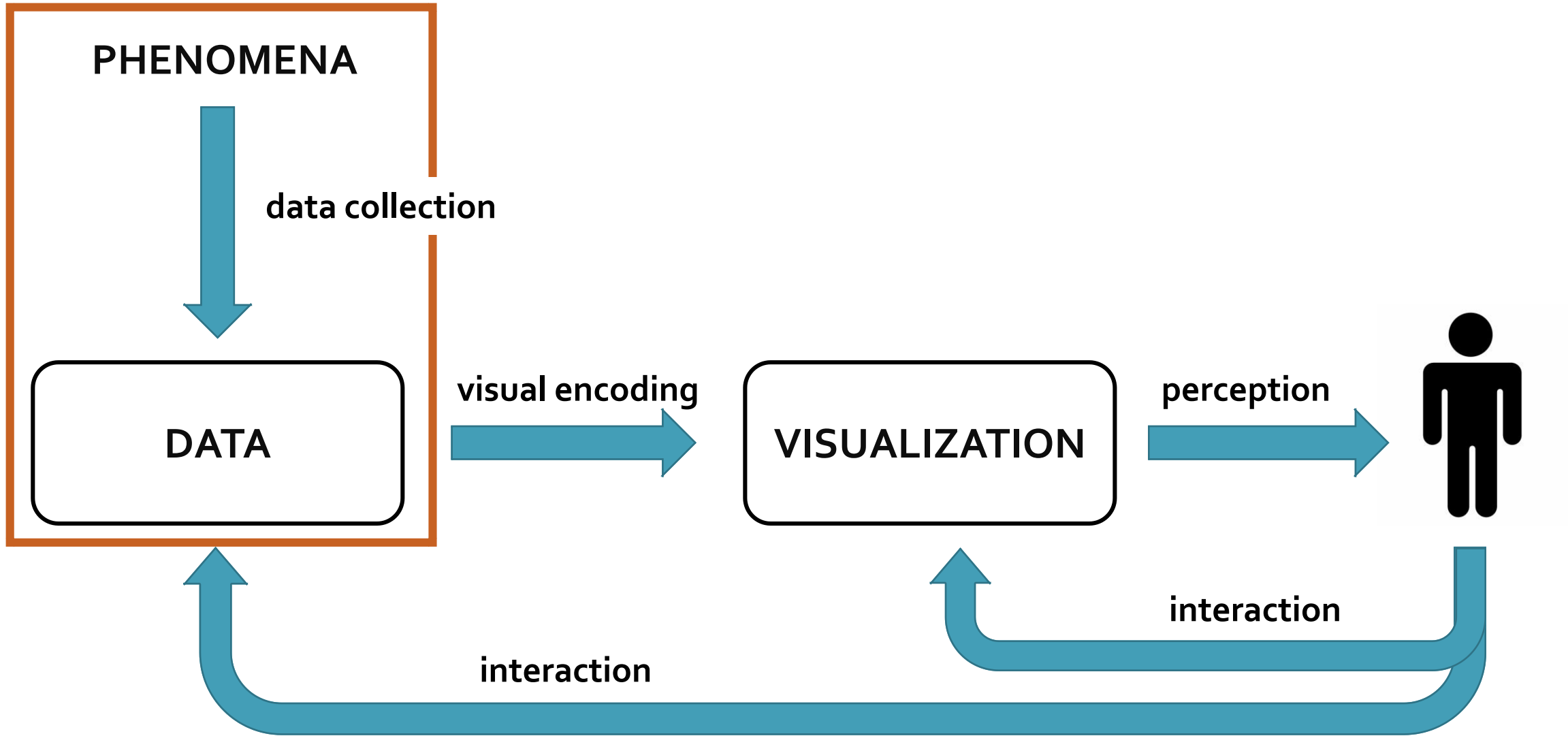


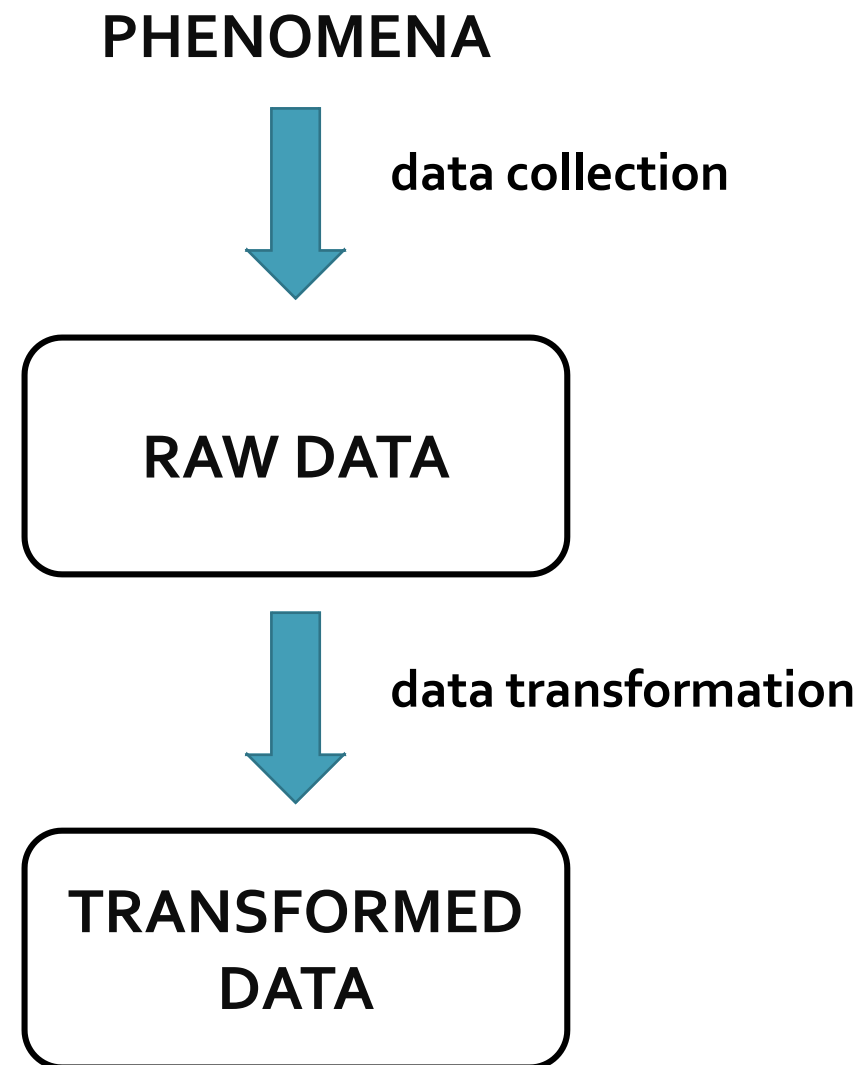
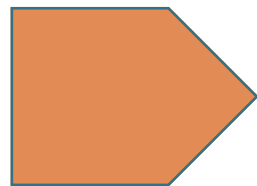
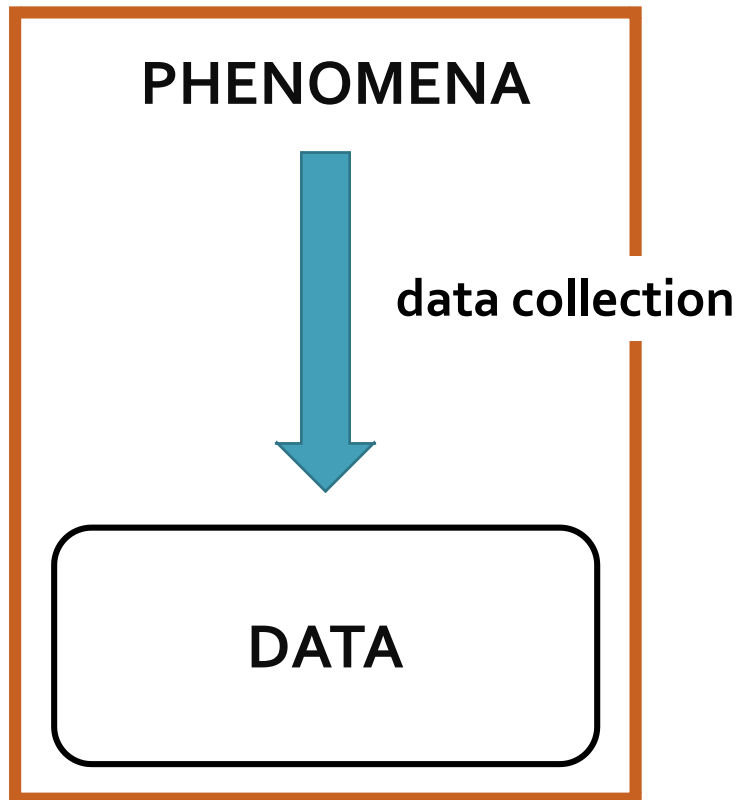
visual encoding











PHENOMENA



data collection



data transformation



visual encoding



perception



interaction



PHENOMENA



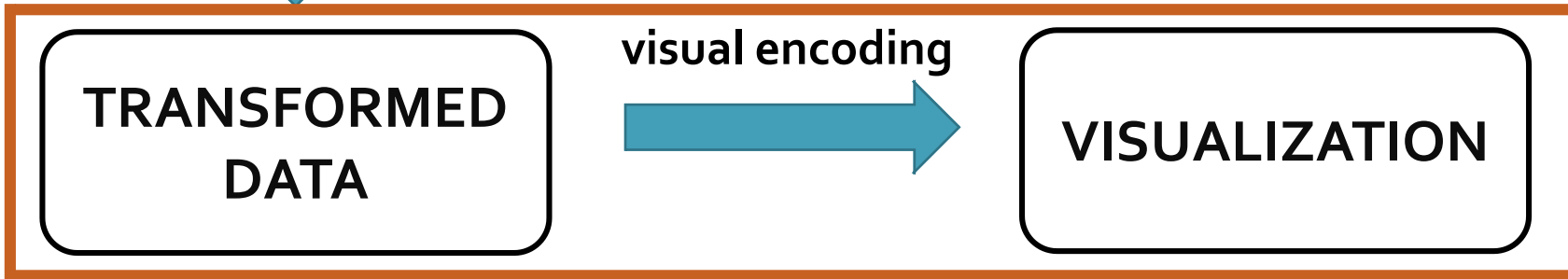
data collection



RAW DATA



data transformation



TRANSFORMED DATA

visual encoding



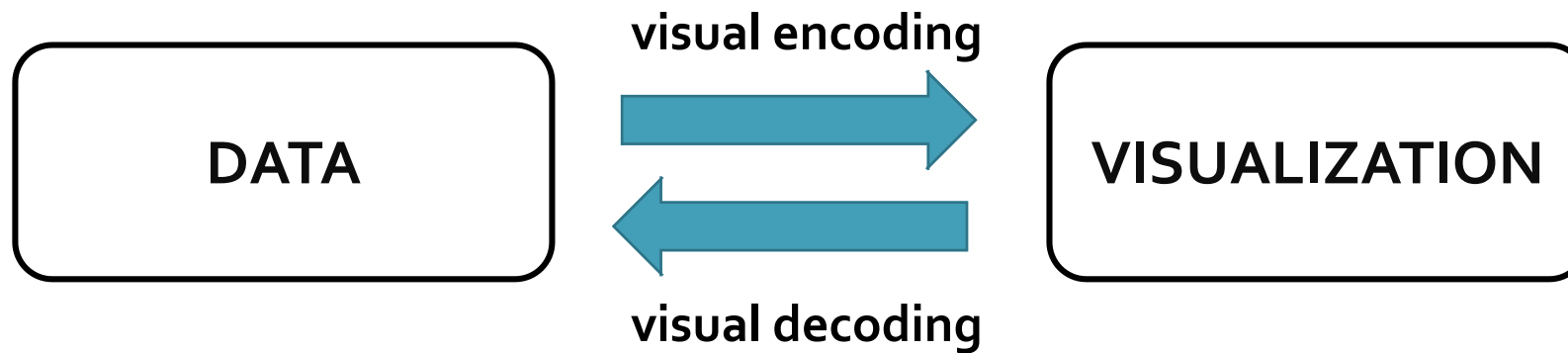
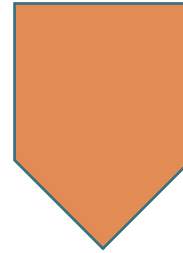
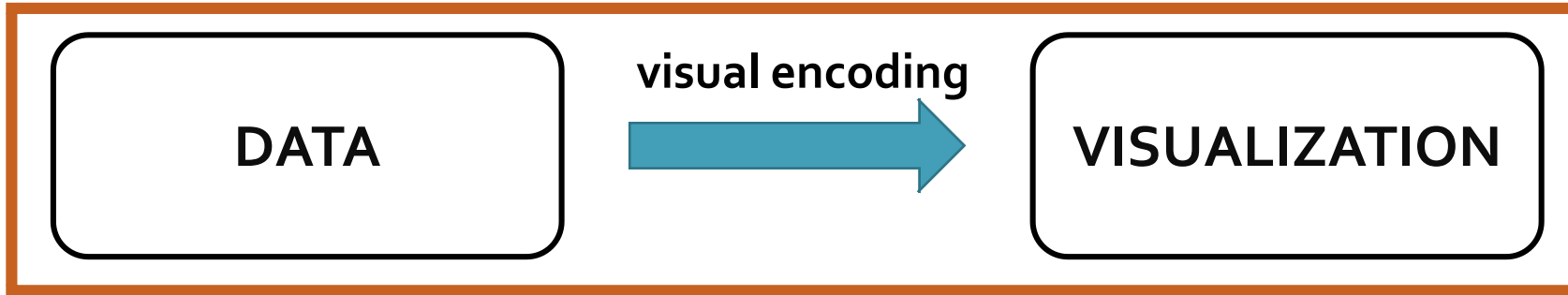
VISUALIZATION

perception



interaction





PHENOMENA



data collection



RAW DATA



data transformation



TRANSFORMED DATA

visual encoding



visual decoding



VISUALIZATION

perception



interaction



PHENOMENA



data collection



RAW DATA



data transformation



TRANSFORMED DATA

visual encoding



visual decoding



VISUALIZATION

perception



interaction



PHENOMENA



data collection



RAW DATA



data transformation



TRANSFORMED DATA

visual encoding



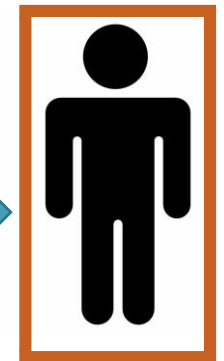
visual decoding



VISUALIZATION



perception



knowledge



interaction



PHENOMENA



data collection



RAW DATA



data transformation



TRANSFORMED DATA

visual encoding



visual decoding



VISUALIZATION

perception



knowledge

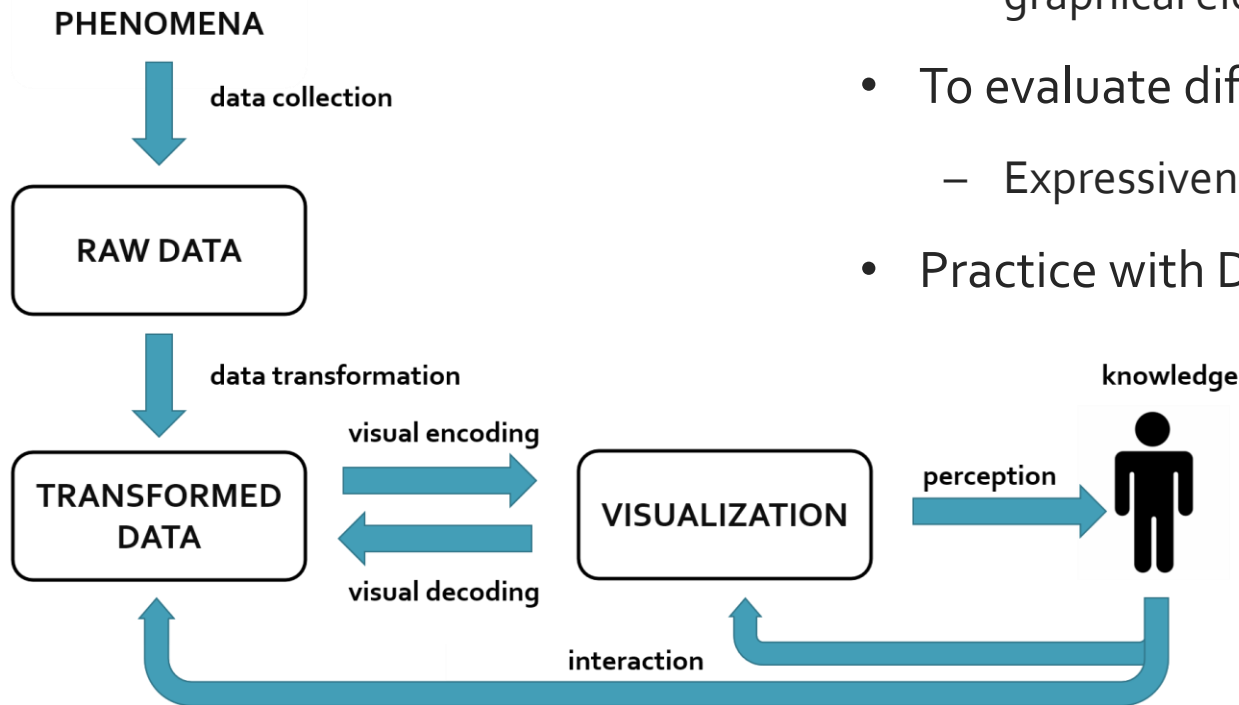


interaction



What we will learn

- To decide *what* to visualize
 - Data abstraction, selection, and transformation
- To decide *how* to visualize it (encoding)
 - Learning the design space and the vocabulary of graphical elements
- To evaluate different solutions (decoding, perception)
 - Expressiveness and effectiveness
- Practice with D3.js and Python



Data types, attributes, transformations

Data

- Data = Factual information such as measurements or statistics, used as a basis for reasoning, discussion or calculation [Mirriam Webster dictionary]
 - Data are generated by sensors, computers, phones, humans...
 - Between the beginning of time and 2003, humanity generated roughly 5 exabytes of data [E. Schmidt, former Google CEO, 2010]. We now produce the same volume of bits every two days [A. Cairo, 2013]
 - Structured vs unstructured data

Data items and attributes

- Data = collection of *items* and *attributes* of items
- *Items* are the objects/entities you want to analyse
- *Attributes* are the properties of these objects/entities

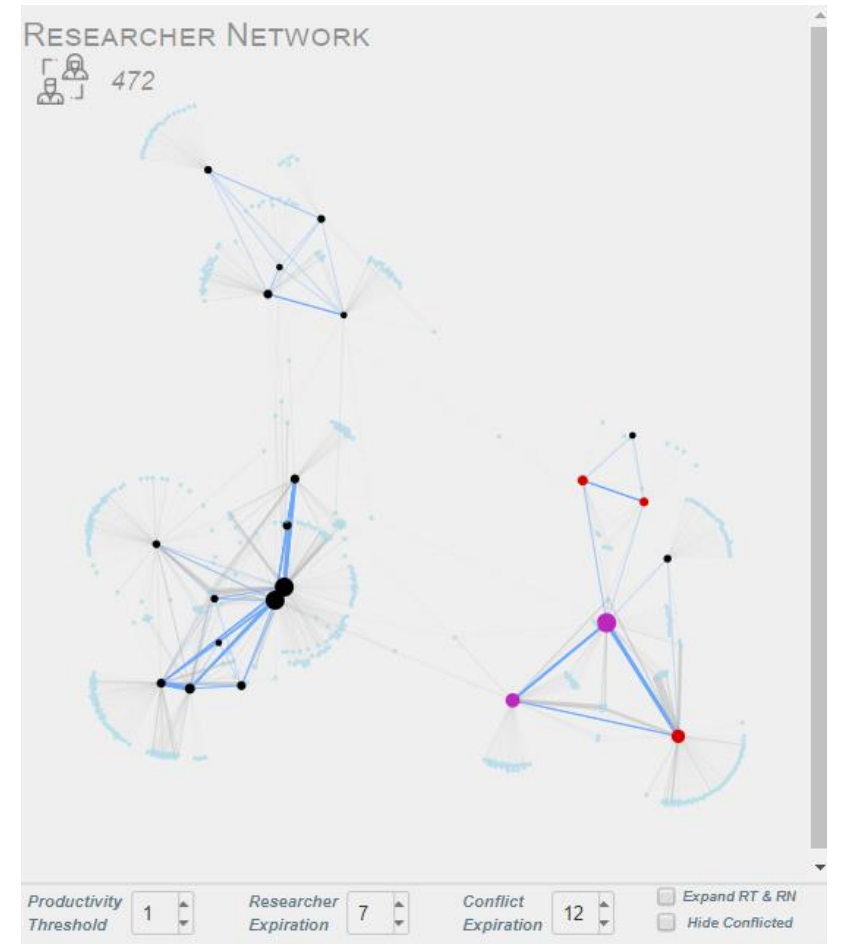
Items	Attribute 1	Attribute 2	Attribute 3	...		
	ID	Gender	Age	Assumption date	Role	Seat
M. Rossi	01066	M	43	01/03/2009	Researcher	Genoa
L. Bianchi	01043	F	36	15/06/2015	Senior Researcher	Genoa
G. Neri	01432	M	59	01/09/1998	Research Director	Pisa
T. Smith	02312	M	32	01/01/2019	Technologist	Pisa
D. Verdi	00043	F	55	01/05/2005	Senior Researcher	Pisa
W. James	01203	M	48	01/03/2004	Technician	Bologna

[CNR imaginary employess]

Data types

- A *table* is a grid of columns and rows, with rows representing items and columns representing attributes
- A *network* is a collection of nodes representing items, connected by links; both nodes and links can have attributes

Tables
Networks



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

Attribute types

- *Nominal* attributes are attributes whose values describe categories, which do not have an intrinsic ordering
- *Ordinal* are attributes whose values describe categories, which can be ordered (and order is meaningful)
- *Quantitative* attributes are attributes whose values represent measured quantities. Their values can be ordered, but also the distance between values can be computed and has a meaning.

		Nominal ↓	Quantitative ↓		Ordinal ↓	
	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	Researcher	Pisa
D. Verdi	00043	F	55	01/05/2005	Senior Researcher	Pisa
W. James	01203	M	48	01/03/2004	Researcher	Bologna

Attribute types

- *Nominal* attributes are attributes whose values describe categories, which do not have an intrinsic ordering
- *Ordinal* are attributes whose values describe categories, which can be ordered (and order is meaningful)
- *Quantitative* attributes are attributes whose values represent measured quantities. Their values can be ordered, but also the distance between values can be computed and has a meaning.



Nominal



Quantitative



Ordinal

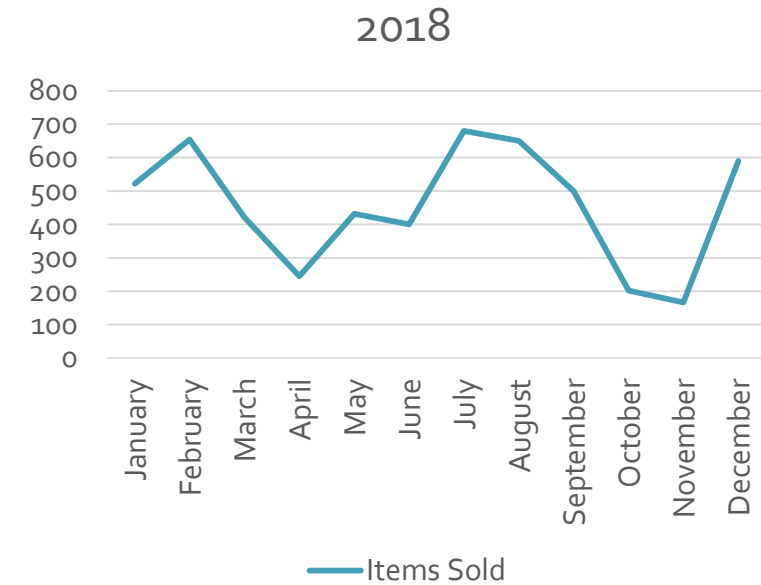


	ID	Gender	Age	Assumption date	Role	Seat
M. Rossi	01066	M	43	01/03/2009	Researcher	Genoa
L. Bianchi	01043	F	36	15/06/2015	Senior Researcher	Genoa
G. Neri	01432	M	59	01/09/1998	Research Director	Pisa
T. Smith	02312	M	32	01/01/2019	Researcher	Pisa
D. Verdi	00043	F	55	01/05/2005	Senior Researcher	Pisa
W.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

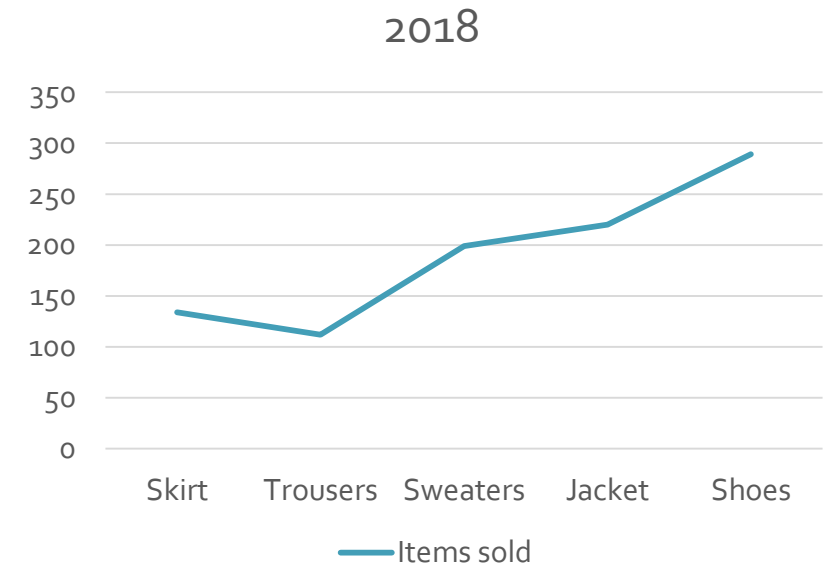
Attribute types

How to select the appropriate visual representation according to the attribute type

- Good idea



- Not a good idea



Attribute semantics

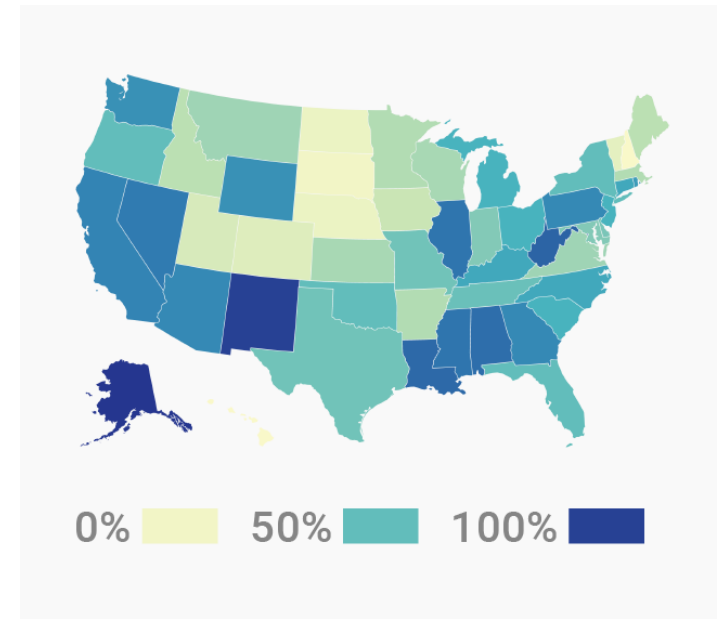
i.e., the meaning of attributes

- Spatial and temporal attributes
 - e.g., location (city) is nominal and spatial; latitude and longitude are quantitative and spatial
 - e.g., date of assumption is temporal
- Sequential, diverging, cyclic attributes
 - e.g., temperature is diverging (it has a zero value, and positive/negative values); month of the year is cyclic
- Hierarchical attributes
 - e.g., type of product (clothes) with subcategories (shirt, skirt, trouser, sweater)

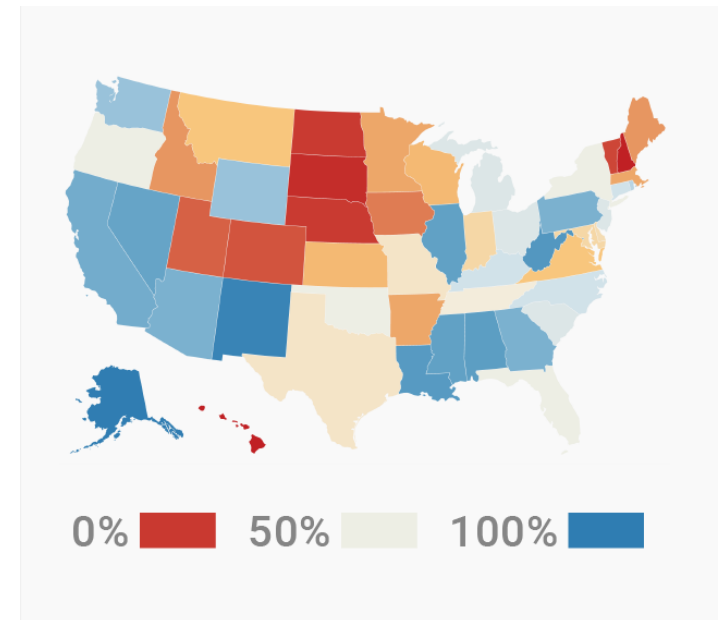
Attribute semantics

How to select the appropriate visual representation according to the attribute semantics

- Good idea



- Not a good idea



Attribute transformation

(Back on this, later on in the course)

- Selection of the attributes to visualize
- Aggregation of attributes
 - mean, max, average, sum, median, stddev...
 - spatial aggregation (e.g., by province, region, country)
 - aggregation by time resolution (e.g., by week, month, year)
- Geographical coding and decoding
 - e.g., from city names to latitude and longitude coordinates
- Binning
 - from quantitative to ordinal
- Re-expression
 - normalization, percentage...

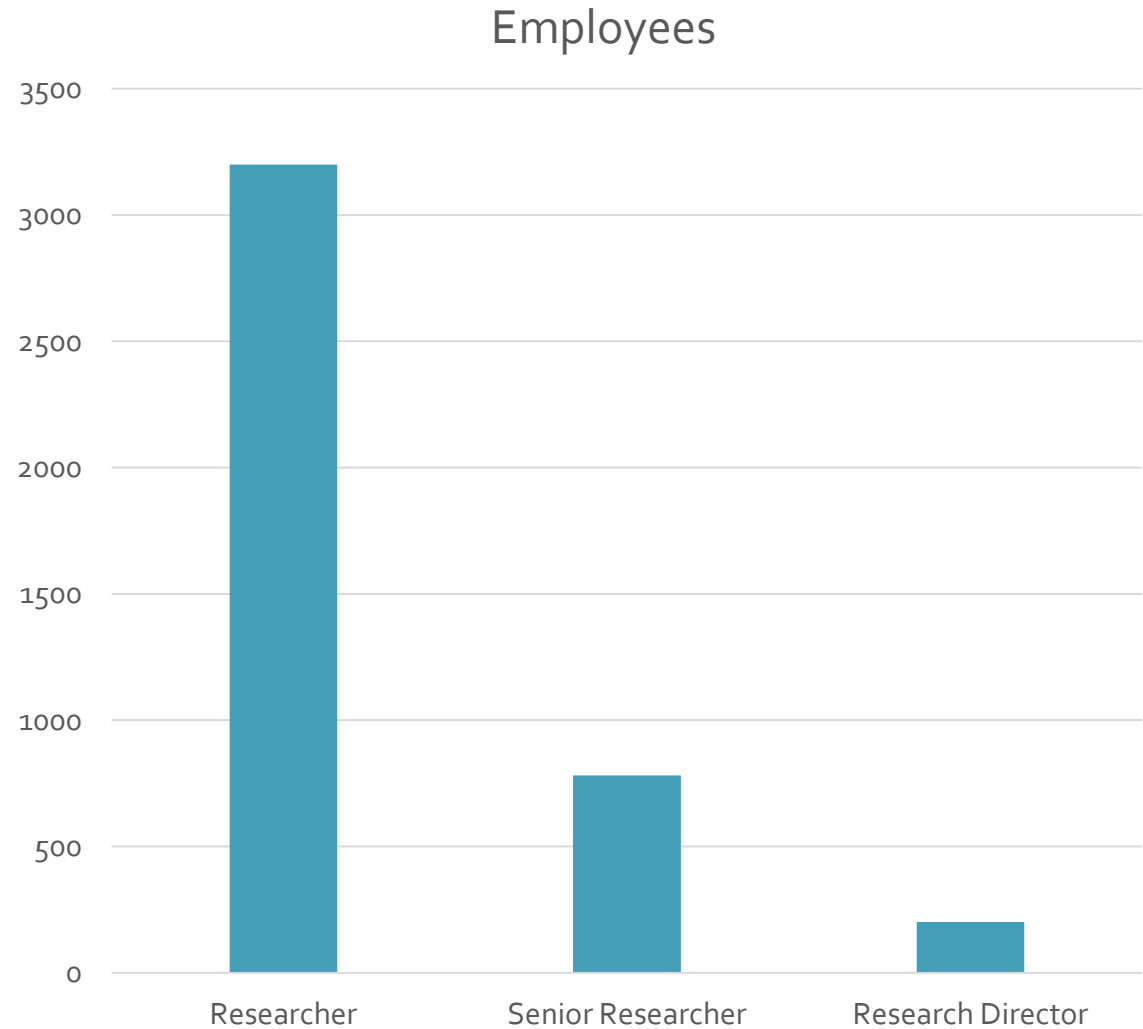
Fundamental graphs – Part I

Bar charts

Relationship between a set of categories and measured quantities

Bivariate data: nominal/ordinal (typically independent) and quantitative (typically dependent)

Do not confuse with histograms!



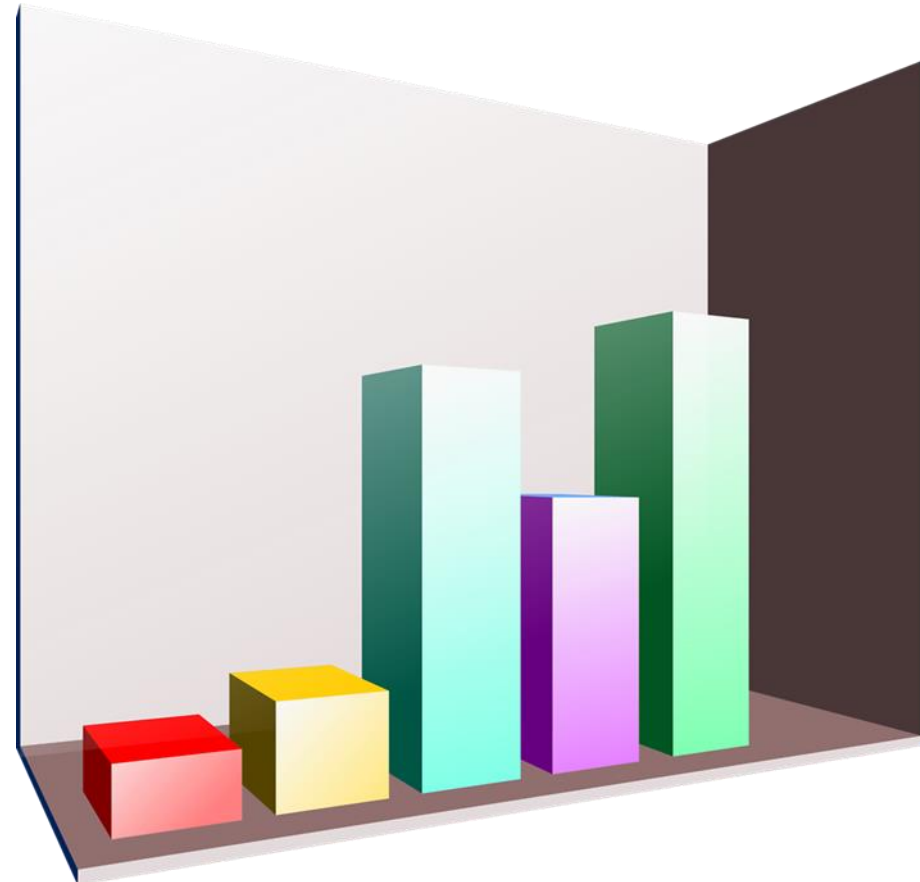
Bar charts

- Uhm...

Relationship between a set of
categories and measured
quantities

Bivariate data: nominal/ordinal
(typically independent) and
quantitative (typically dependent)

Do not confuse with histograms!

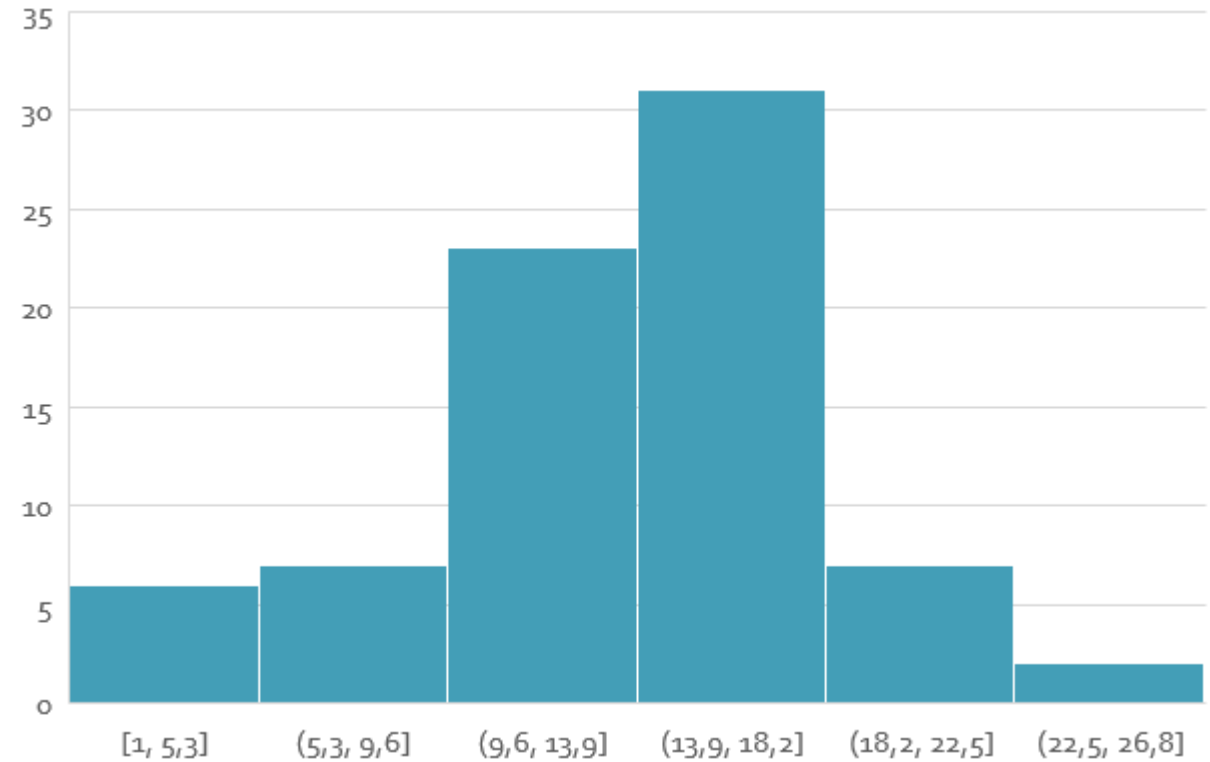


Histograms

Frequency of items

Bivariate data: an independent variable quantized in intervals (bins) and a dependent variable

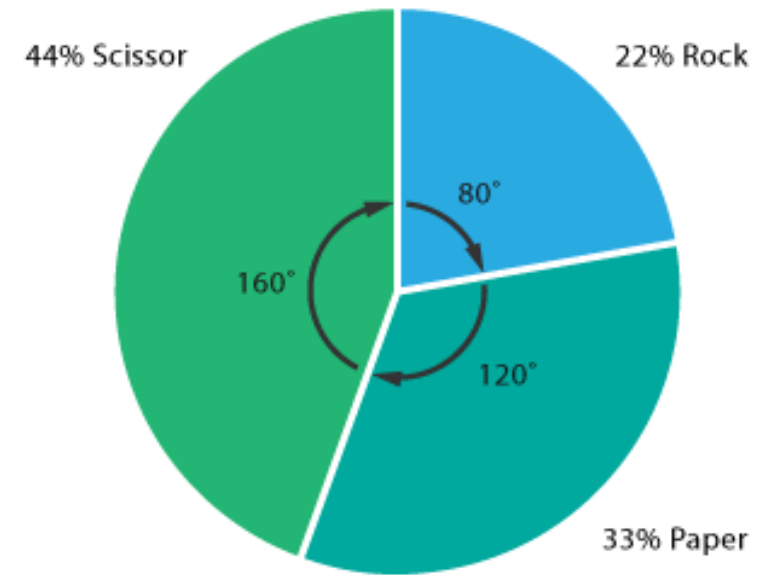
Do not confuse with bar charts!



Pie charts

Show proportions and percentages between categories

Bivariate data: independent and dependent



Data			
Rock	Paper	Scissor	TOTAL
2	3	4	9
To calculate percentages			
$2/9=22\%$	$3/9=33\%$	$4/9=44\%$	100%
Degrees for each "pie slice"			
$(2/9) \times 360 = 80^\circ$	$(3/9) \times 360 = 120^\circ$	$(4/9) \times 360 = 160^\circ$	360°

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

Pie charts

Show proportions and percentages between categories

Bivariate data: independent and dependent

Primeiro Plano

Diagrama

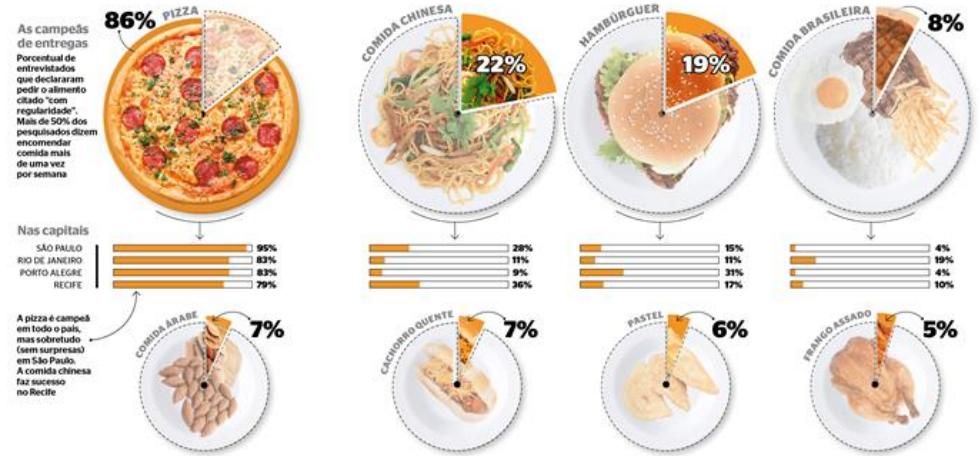
A NOTÍCIA EM PERSPECTIVA

Quando o brasileiro come fora

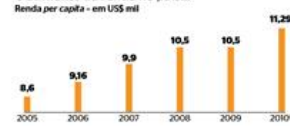
O crescimento da economia muda os hábitos alimentares e estimula o mercado de comida pronta

Luiz Salomão e Alberto Cairo

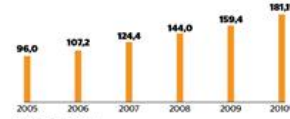
O **BRASILERO** vai cada vez menos à cozinha - e isso já se aproxima dos países desenvolvidos. Segundo o estudo Alimentação fora do lar na visão do consumidor brasileiro, da GS&M, consultoria de consumo especializada em varejo, **mais de 30% do gasto dos brasileiros com alimentação é feito com serviços de entrega ou refeições fora de casa.** So Estados Unidos, Portugal, Reino Unido e Espanha estão à frente. Esse percentual era de 24% em 2002. Oitent e quatro por cento dos 1.224 entrevistados em grandes capitais disseram que costumam comprar alimentos prontos para consumo, seja em supermercados, padarias ou restaurantes. A pesquisa atribui essa mudança, em primeiro lugar, ao crescimento econômico: o Brasil cresceu três vezes mais que a média das economias mais desenvolvidas nos últimos cinco anos, e a classe C já representa quase metade da população. Em segundo lugar, as mulheres já são 43% do total de trabalhadores. Por fim, os "domicílios unipessoais" passaram de 3,4 milhões em 1996 para 6,3 milhões em 2006.



O aumento da renda no país...



...acompanha o crescimento das empresas de comida pronta e restaurantes



34 | **EPiCA** | 1.º de dezembro de 2010

Porcentual do orçamento familiar para alimentação gasto em refeições fora de casa



Quem gasta mais fora de casa



Em dias úteis, almoço fora: no fim de semana, jantar fora



O brasileiro valoriza o sabor, a higiene e a aparência



Fatores que determinam o gasto fora de casa



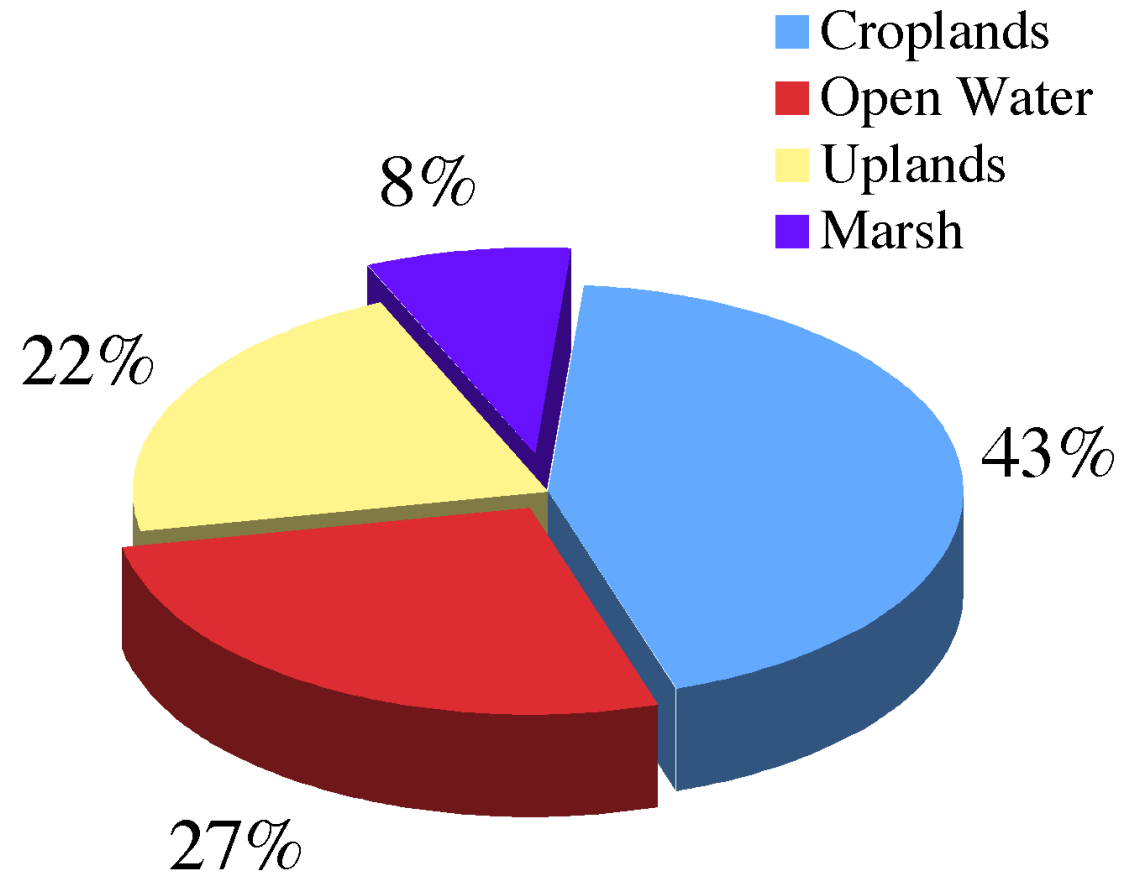
[Luiz Salomao]

Pie charts

- Uhm...

Show proportions and percentages between categories

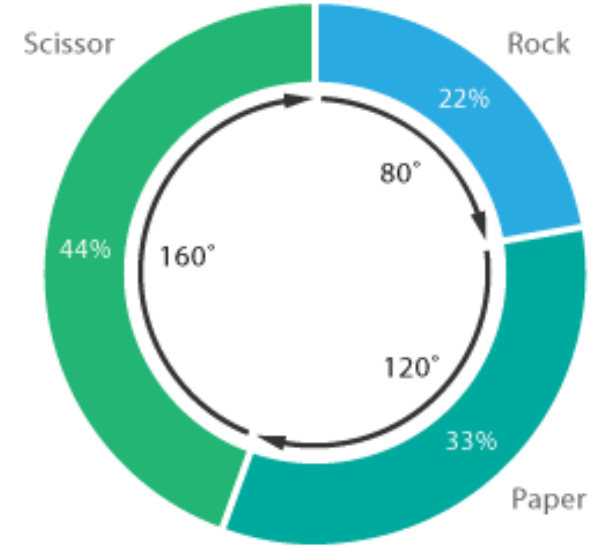
Bivariate data: independent and dependent



[<https://commons.wikimedia.org/wiki/File:TuleLakeHabitats.png>]

Donut charts

Pie charts with center area cut out
More focus on arc length than on proportion between slices

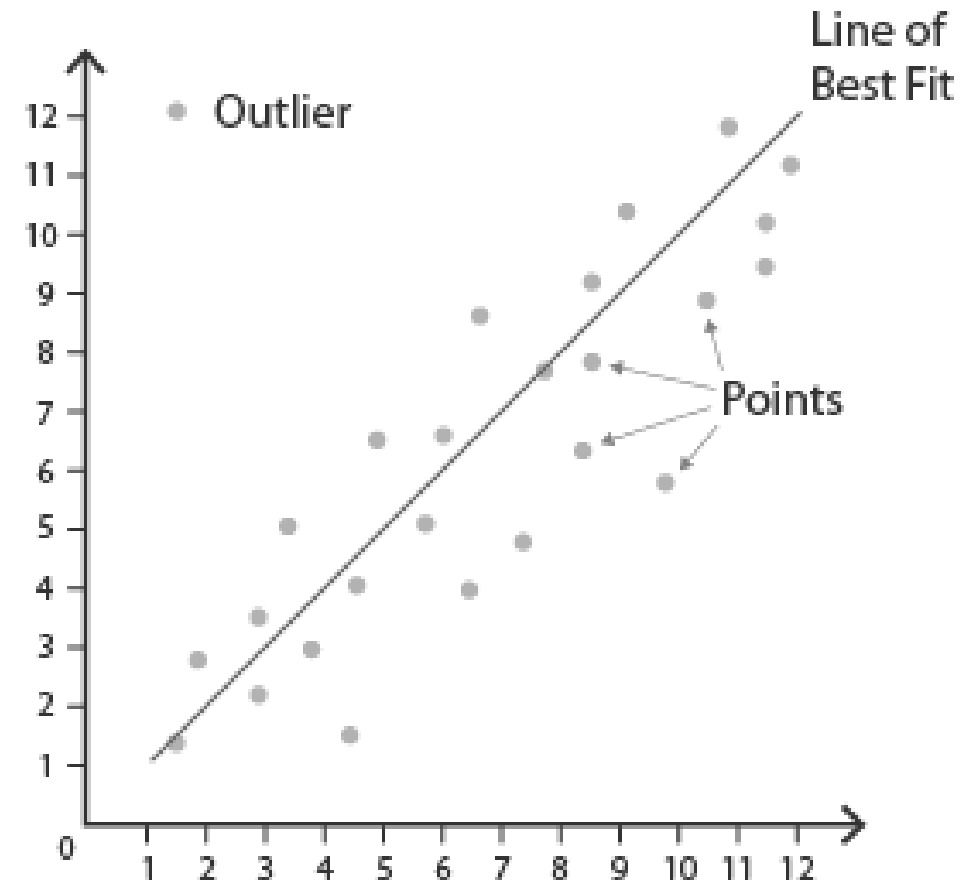


Data			
Rock	Paper	Scissor	TOTAL
2	3	4	9
To calculate percentages			
$2/9 = 22\%$	$3/9 = 33\%$	$4/9 = 44\%$	100%
Degrees for each "donut slice"			
$(2/9) \times 360 = 80^\circ$	$(3/9) \times 360 = 120^\circ$	$(4/9) \times 360 = 160^\circ$	360°

Scatter plots

Relationship between attributes
(and analysis of clusters, outliers,
etc.)

Bivariate data: two independent
quantitative attributes

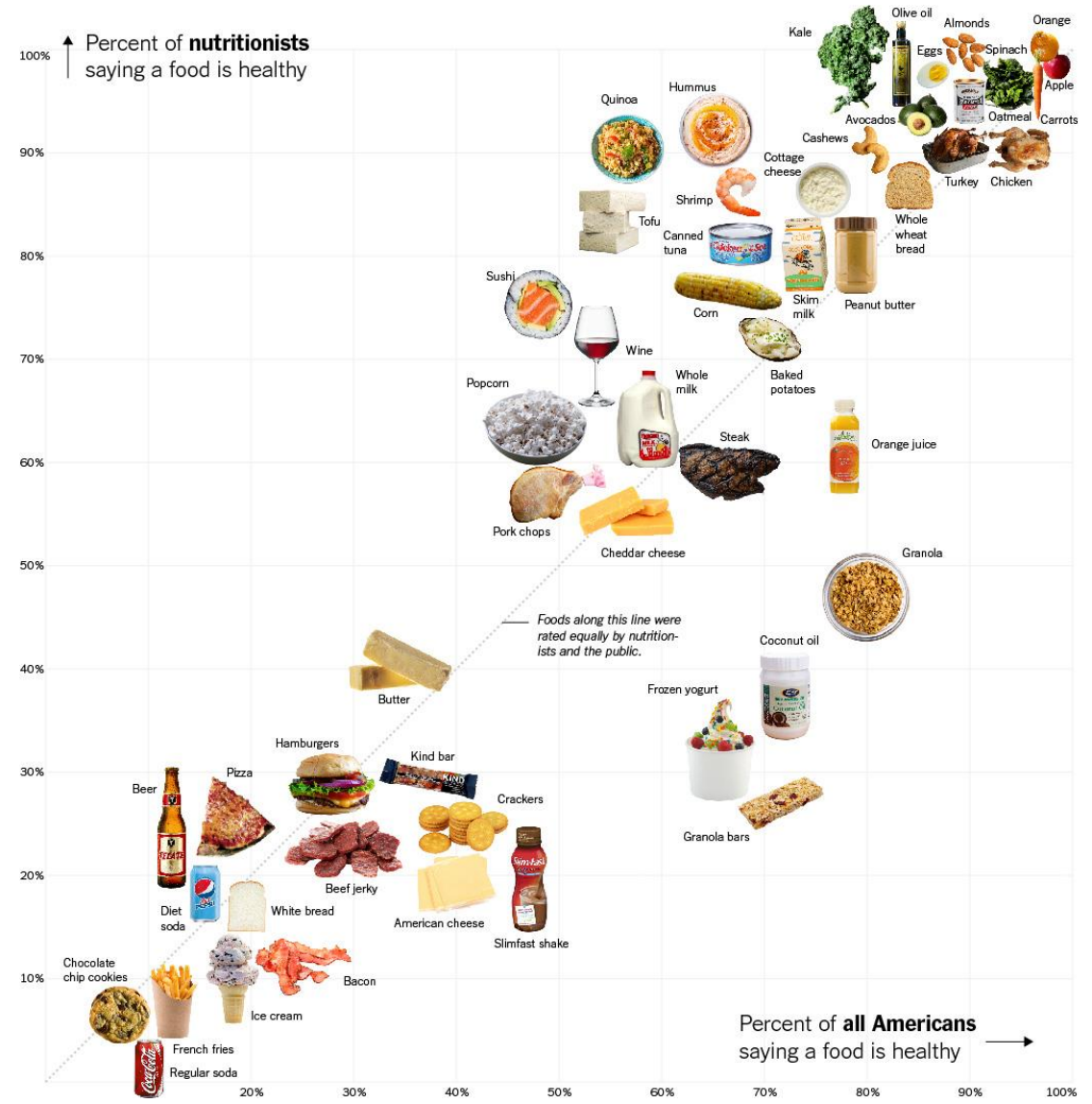


[Data Visualization Catalogue <https://datavizcatalogue.com/>]

Scatter plots

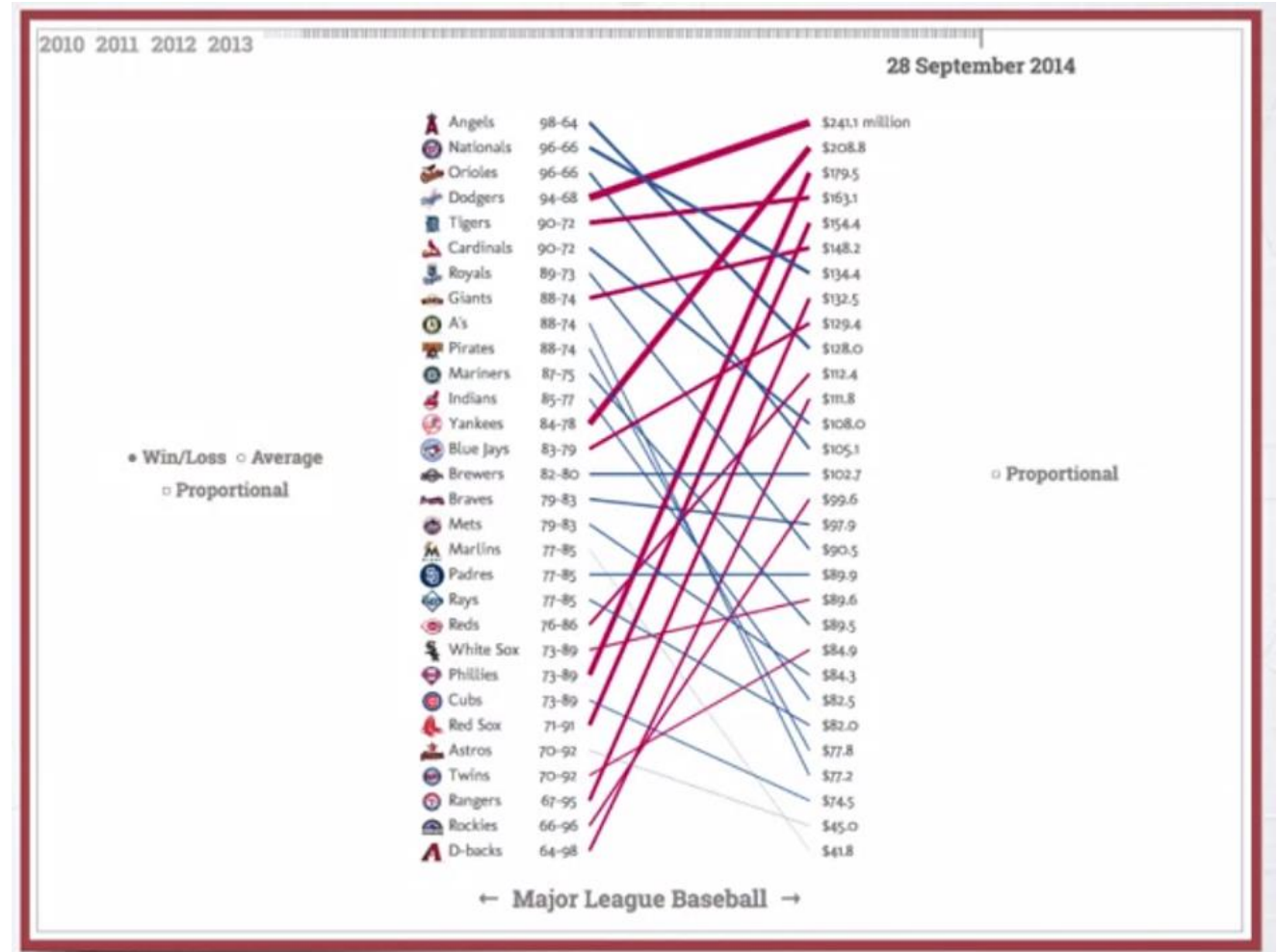
Relationship between attributes
(and analysis of clusters, outliers,
etc.)

Bivariate data: two independent
quantitative attributes



Slope charts

Alternative to scatter plots: each item is a line connecting two quantities

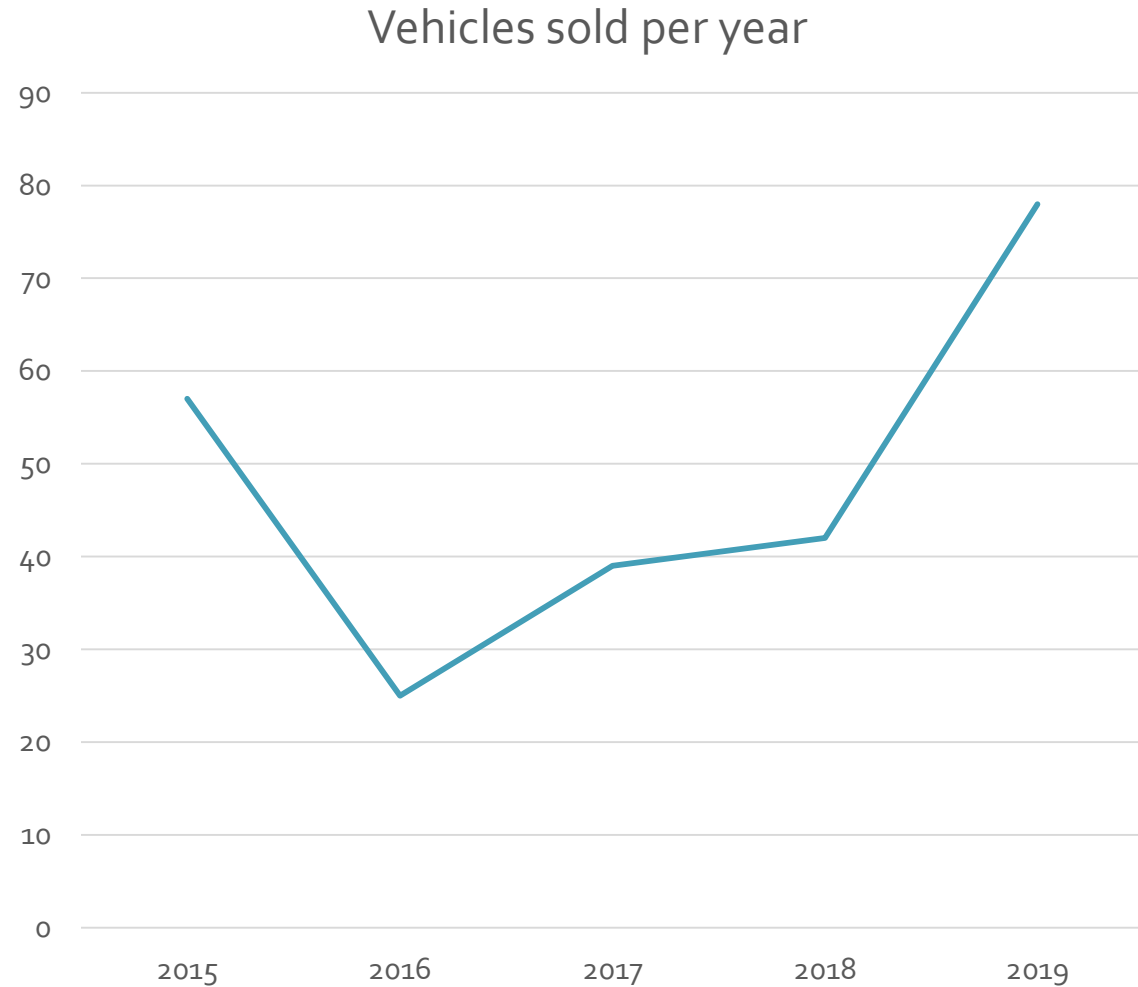


[Coursera, Information Visualization <https://www.coursera.org/learn/information-visualization-fundamentals/home/>]

Line charts

How a quantity changes in relation
to another quantity

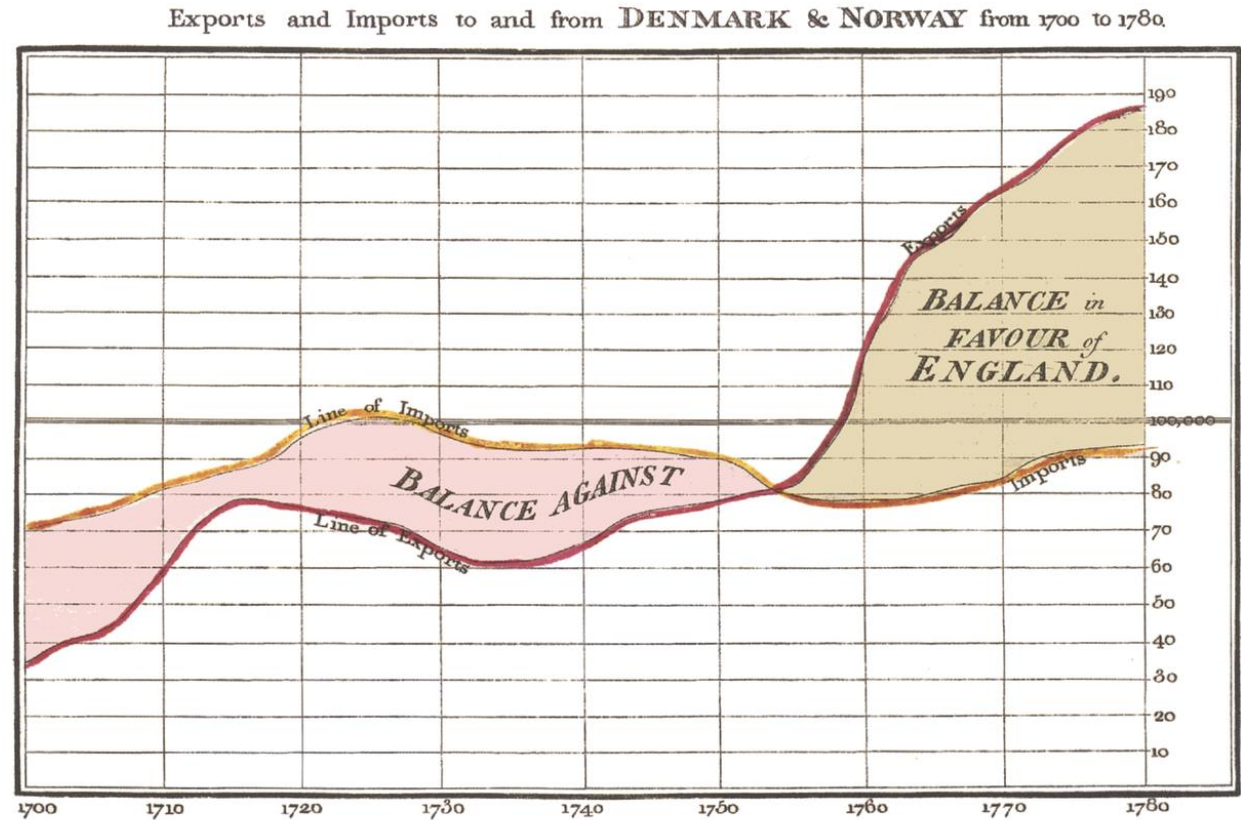
Bivariate data: one is quantitative,
one is usually time



Line charts

How a quantity changes in relation to another quantity

Bivariate data: one is quantitative, one is usually time

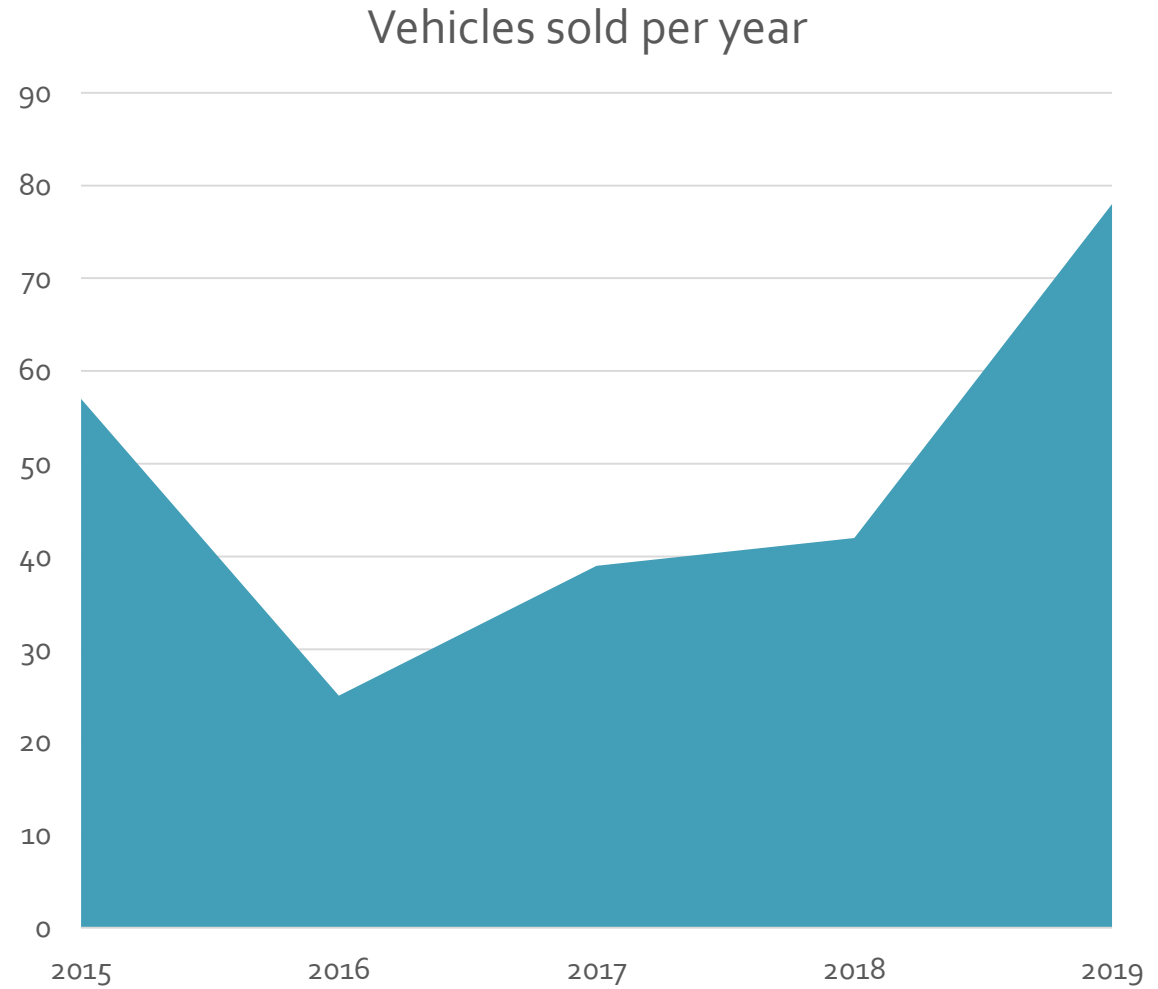


The Bottom line is divided into Years, the Right hand line into £10,000 each.
Published as the Act directs, 14th May 1786, by W^m Playfair. Neale sculpt 352, Strand, London.

[William Playfair, 1759-1823]

Area charts

Line charts with the area below the line filled up

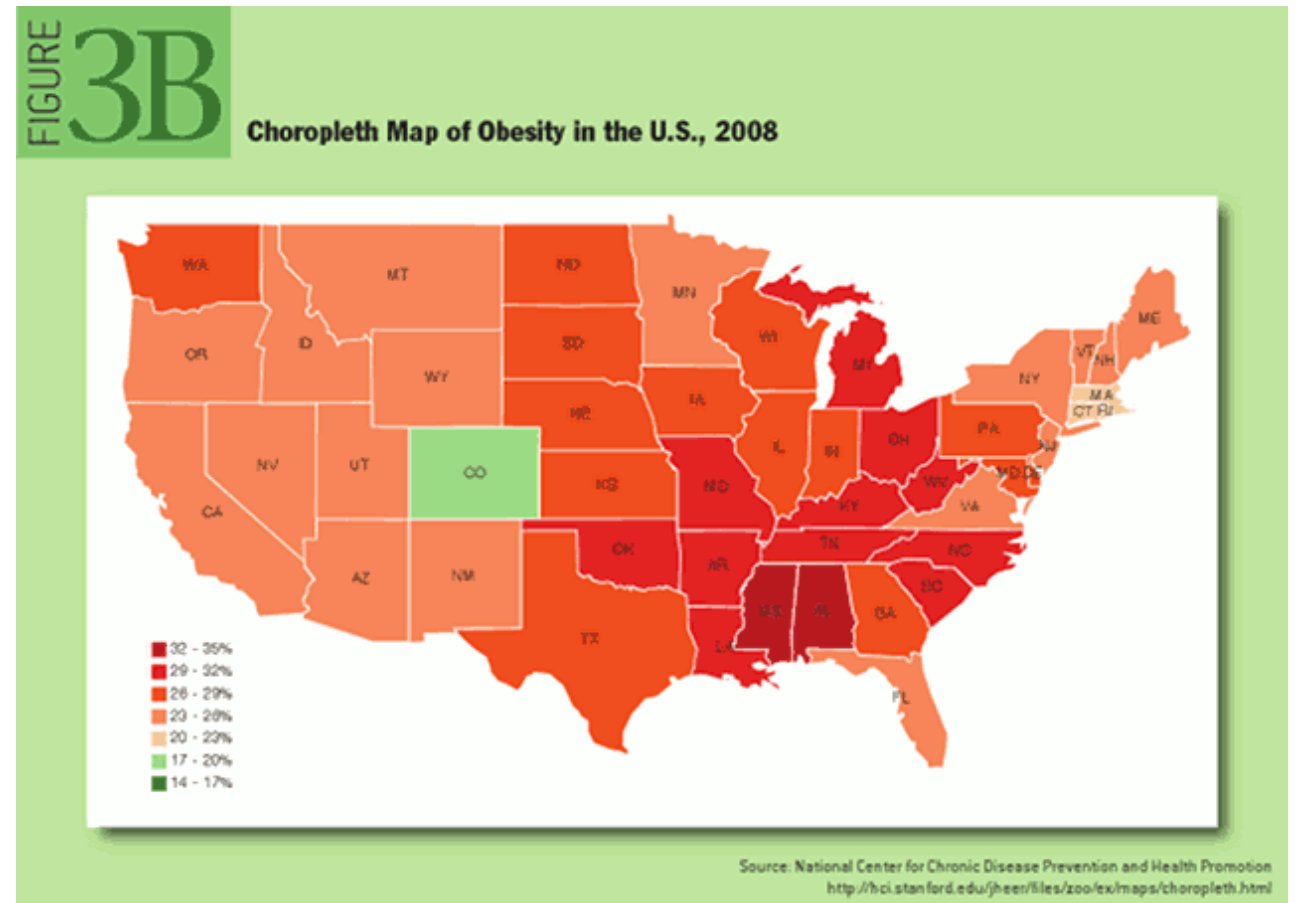


Choropleth maps

How a quantity distributes across geographical areas/regions

Colors, shades, patterns are used to render the quantity associated with areas/regions

Good overview (not accurate comparison) and beware of normalization

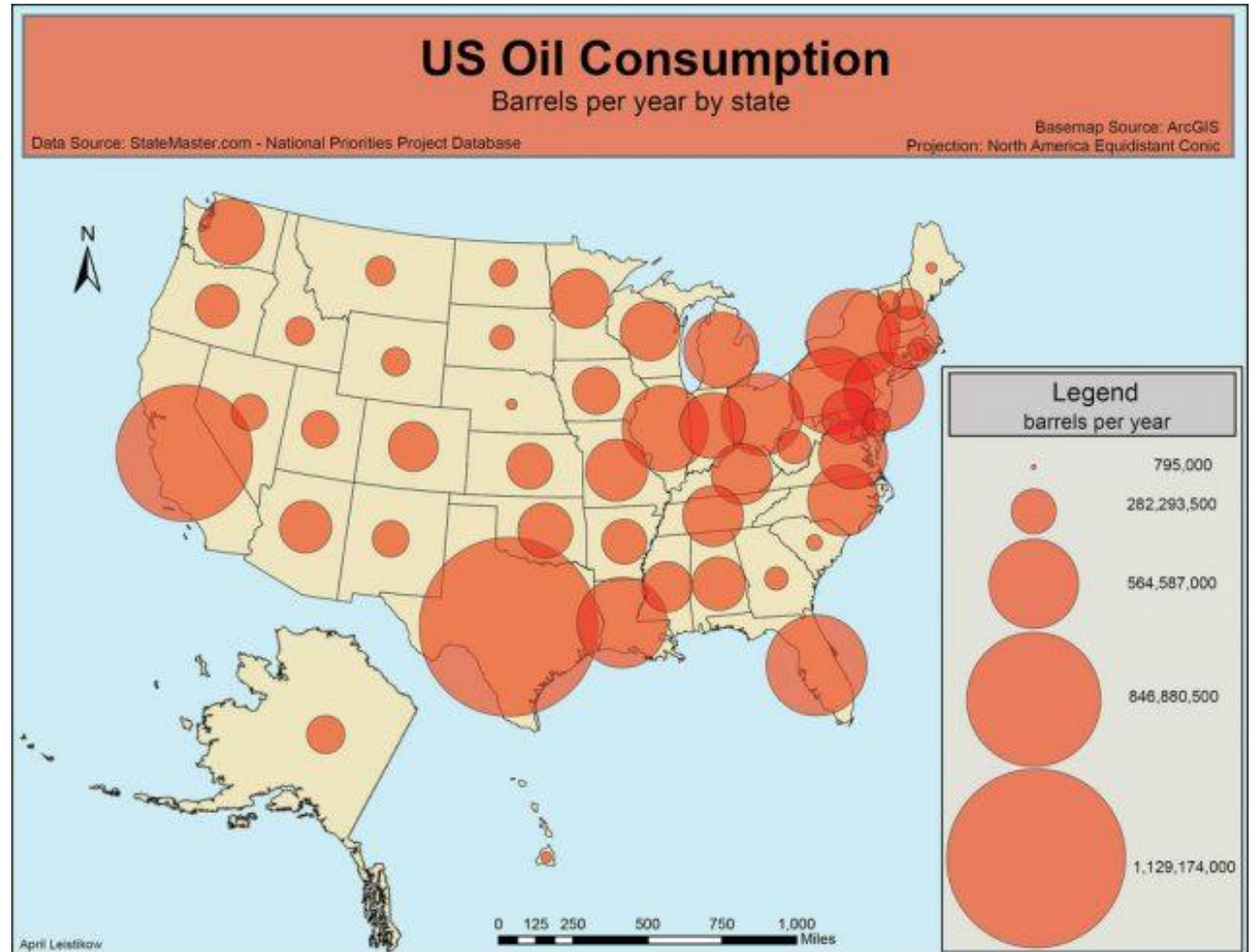


[A tour through the visualization zoo, <https://queue.acm.org/detail.cfm?id=1805128>]

Symbol maps

How a quantity distributes across two spatial coordinates

A symbol (usually a disk or a square) is placed in a point and scaled such that its area is proportional to the quantity associated to the point

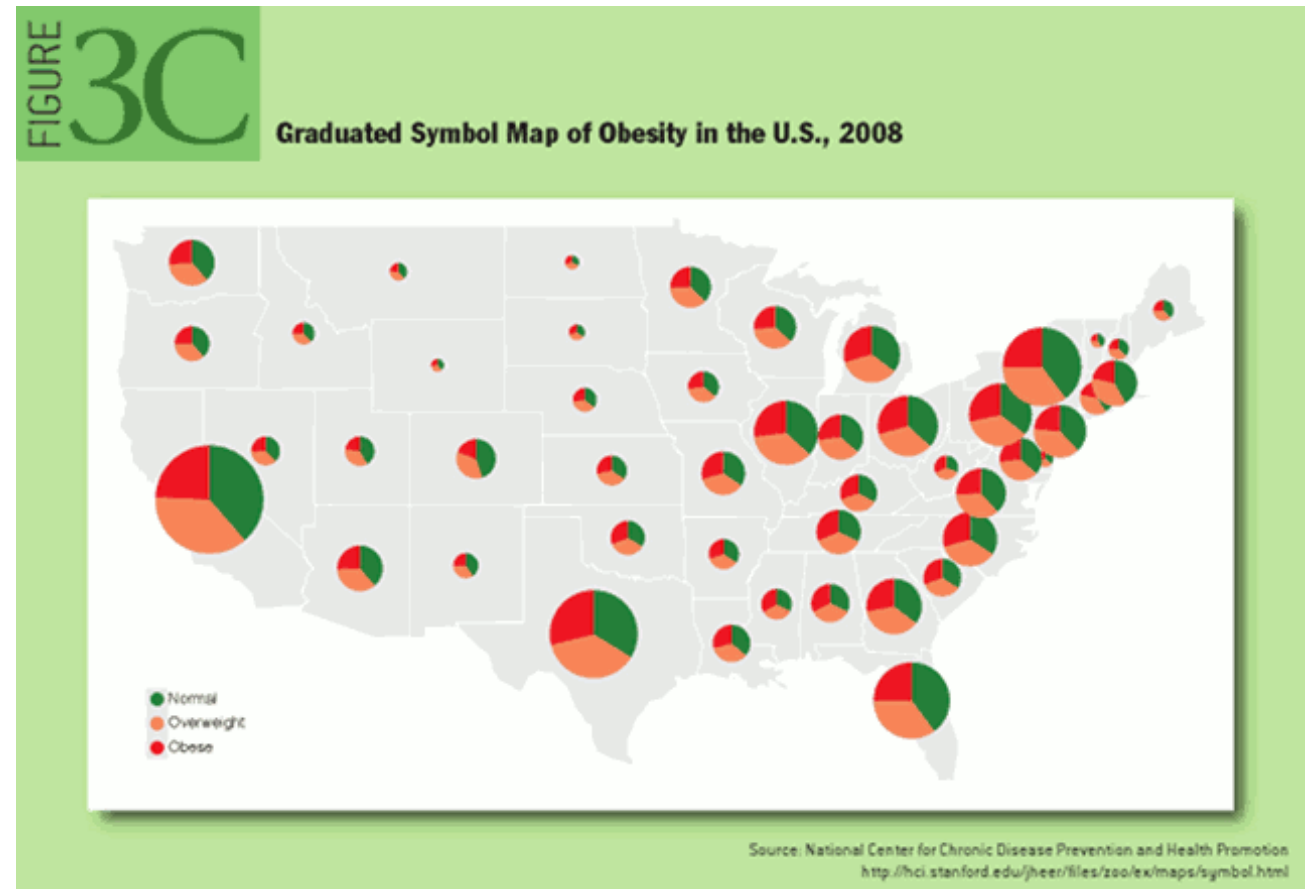


[<https://mappingignorance.org/2013/12/16/the-complexity-of-drawing-good-proportional-symbol-maps/>]

Symbol maps

How a quantity distributes across
two spatial coordinates

Different glyphs can be used as
symbols

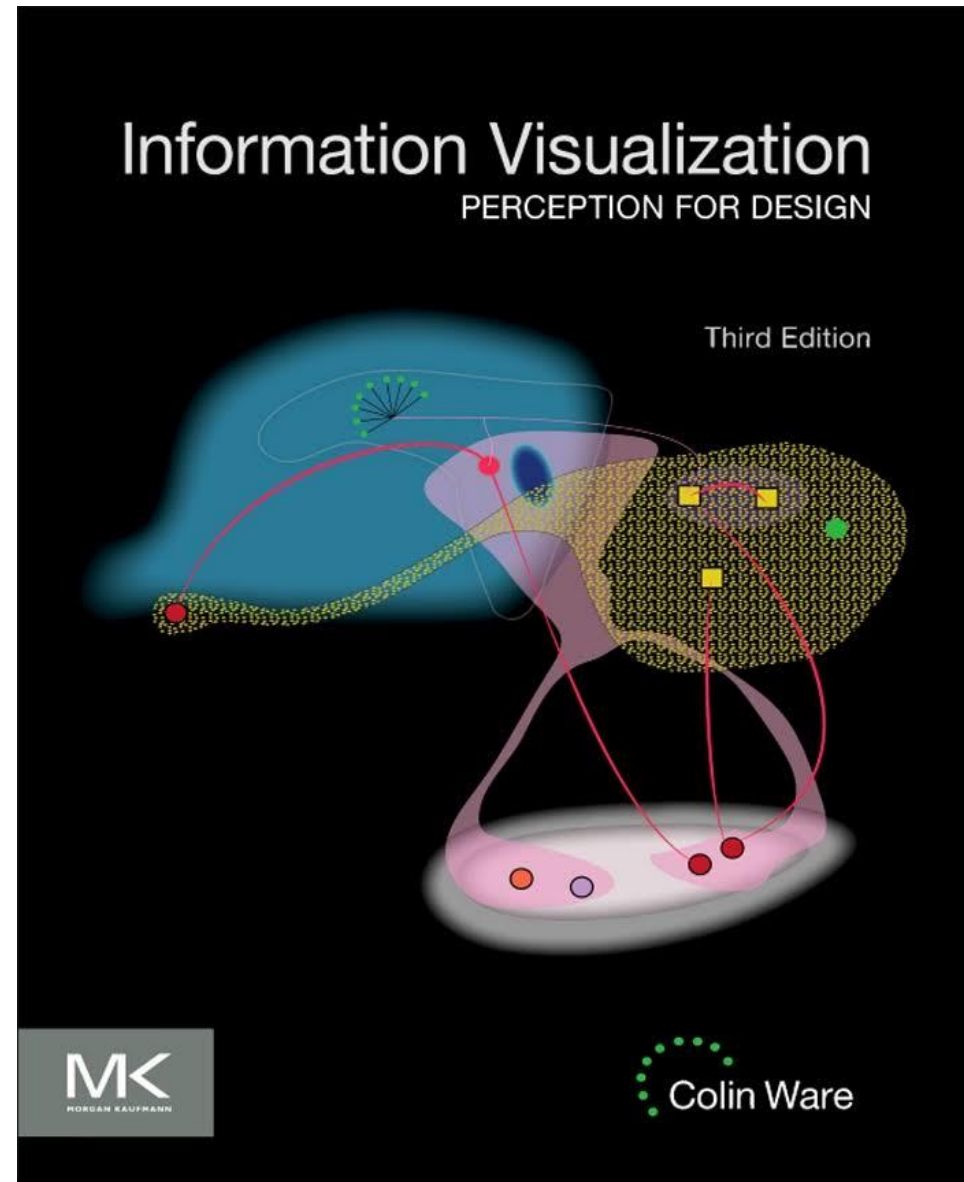


[A tour through the visualization zoo, <https://queue.acm.org/detail.cfm?id=1805128>]

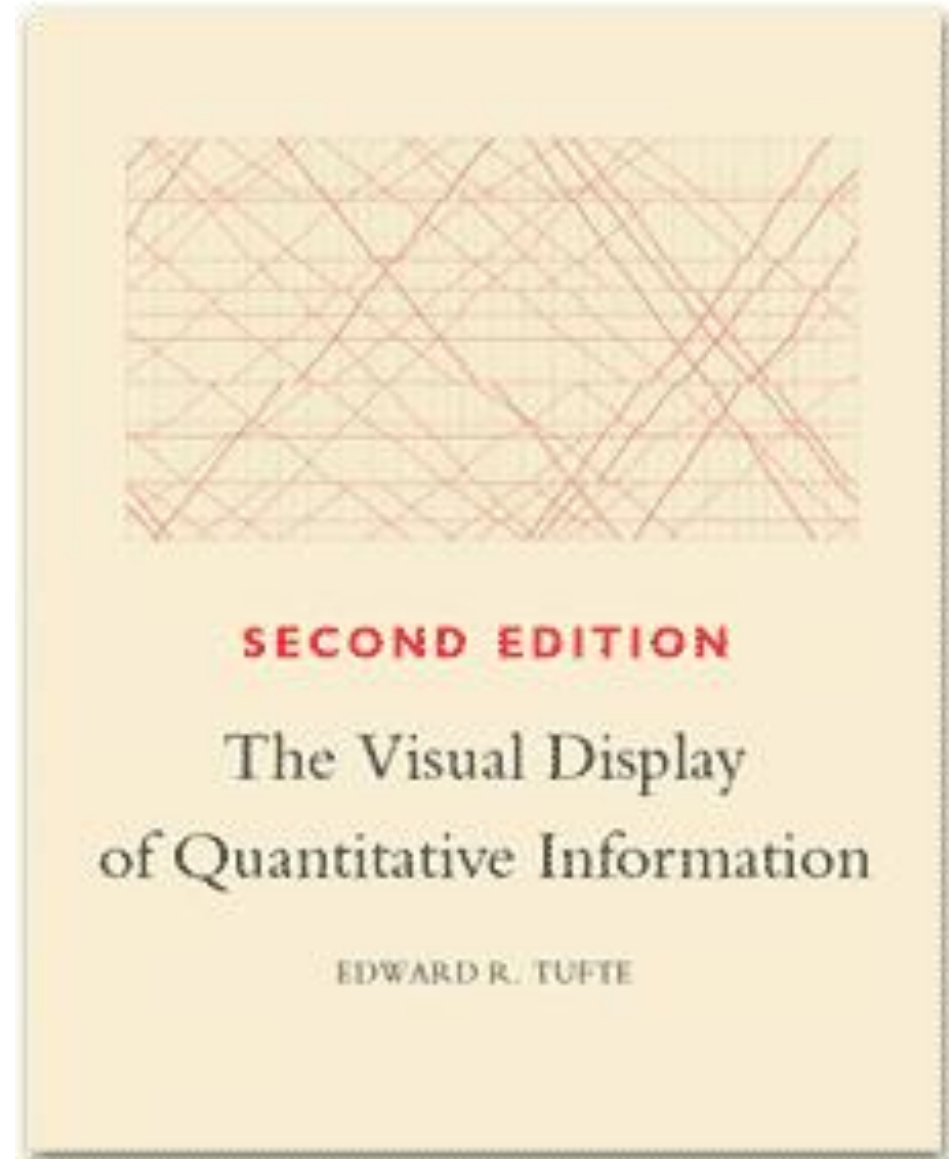
Fundamental graphs – To be continued

Additional references

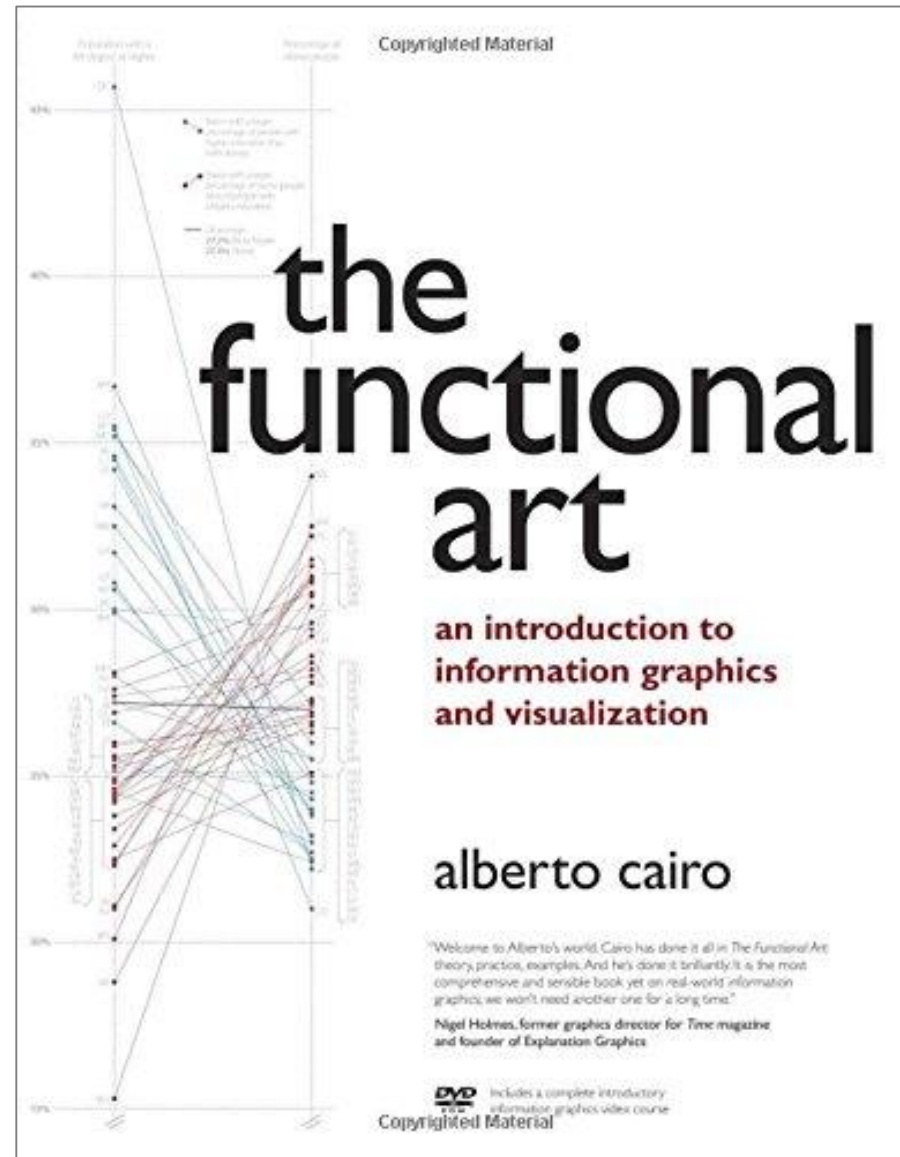
Books



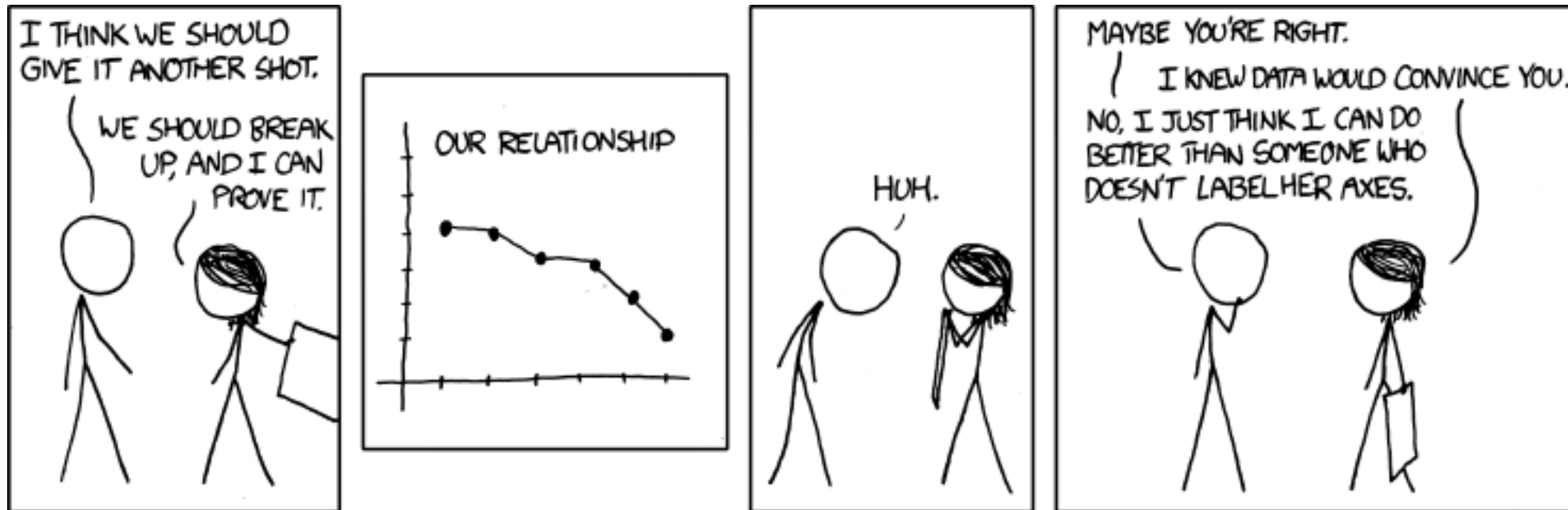
Books



Books



Questions?



[<https://xkcd.com/833/>]