

# Píndoles estadístiques UEB-VHIR

## *Visualització moderna de dades. Més enllà dels diagrames de barres*

 Vall d'Hebron  
Institut de Recerca  
**VHIR**

 **ei**  
ueb.vhir.org

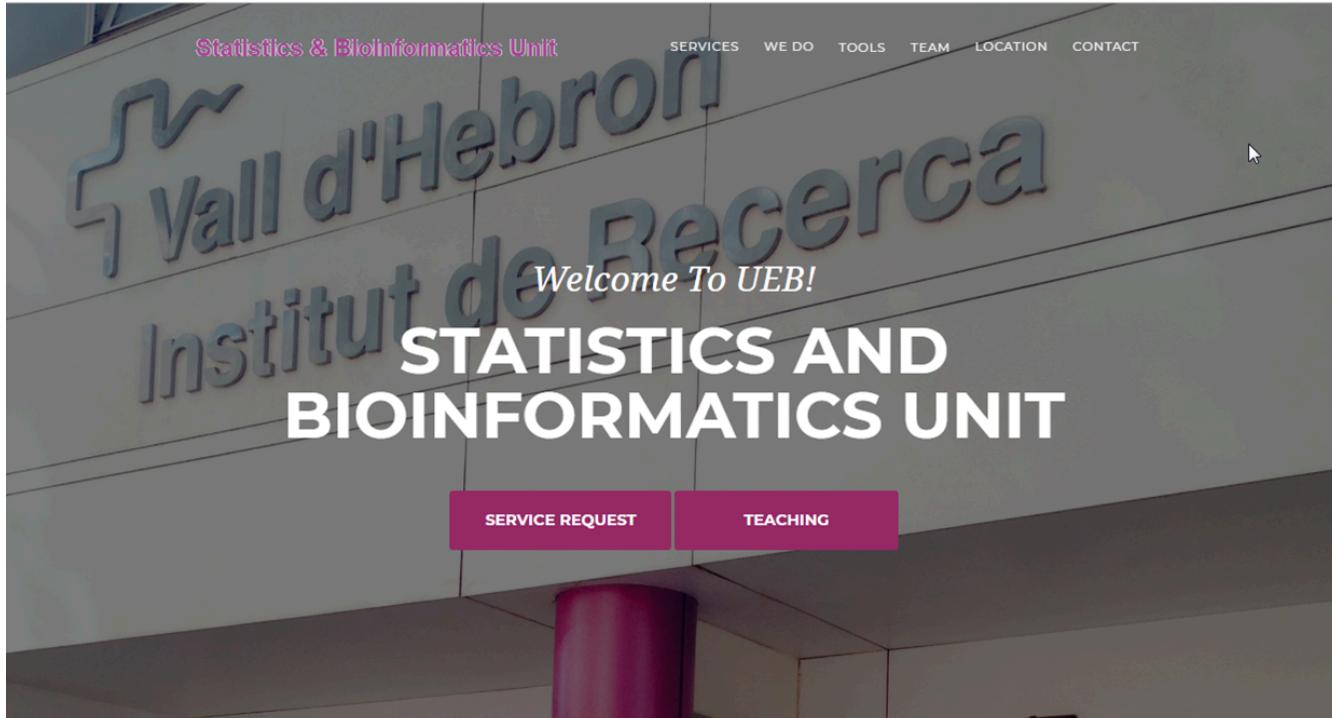
**Miriam Mota i Alex Sanchez-Pla**  
**Unitat d'Estadística i Bioinformàtica (VHIR)**  
**Dept. de Genètica, Micro i Estadística (UB)**

Divendres 31 de Gener de 12:30 a 13:30  
Sala d'Actes de Traumatologia i Rehabilitació

Les píndoles estadístiques són sessions divulgatives, organitzades per la Unitat d'Estadística i Bioinformàtica (UEB) del VHIR, on es presenten problemes i solucions estadístiques dirigides als professionals interessats del Campus Vall d'Hebron

# Statistics and Bioinformatics

Unit (UEB)



<http://ueb.vhir.org>

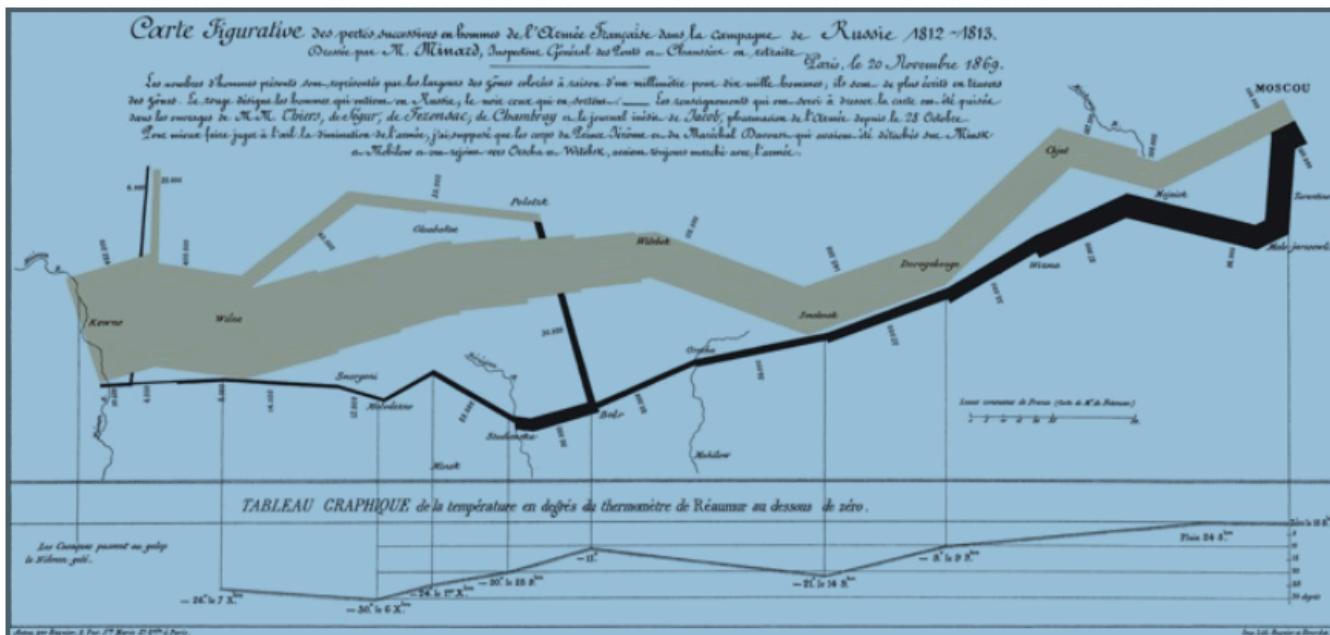
# Outline of the talk

- What is data visualization? ("WHAT")
- Why we need data visualization? ("WHY")
- Stages of data visualization ("HOW")
- Getting more concrete. ("WHICH", "WHEN")
- Tools for Data Visualization ("WITH WHAT")
- Summary (and recommendations)
- References and links

# Outline of the talk

- **What is data visualization? ("WHAT")**
- Why we need data visualization? ("WHY")
- Stages of data visualization ("HOW")
- Getting more concrete. ("WHICH", "WHEN")
- Tools for Data Visualization ("WITH WHAT")
- Summary (and recommendations)
- References and links

# What is data visualization?



"A good sketch is better than a long speech"  
(Napoleon)

# A famous animated visualization



Hans Rosling - *The best Stats you have ever seen*

# So what is data visualization?

- "*the process of acquiring, interpreting and comparing data in order to clearly communicate complex ideas, thereby facilitating the identification and analysis of meaningful patterns.*"
- "*A computer-based system providing visual representations of datasets intended to help people carry out some task more effectively*"  
*(Tamara Munzner)*"

perception vs cognition

human in the loop needs the details

computer-based visualization systems providing  
visual representations of datasets to help people  
carry out some task more effectively

intended task

measurable definitions of effectiveness

# Outline of the talk

- What is data visualization? ("WHAT")
- **Why we need data visualization? ("WHY")**
- Stages of data visualization ("HOW")
- Getting more concrete. ("WHICH", "WHEN")
- Tools for Data Visualization ("WITH WHAT")
- Summary (and recommendations)
- References and links

# Why do we visualize data?

(“WHY”?)

X Lisa Charlotte Rost: Why do we visualise data? @lisacrost

# Why do YOU visualise data?

---

Priorities?

- Attention & Beauty
- Understanding
- Implication



understanding

beauty

# Why do we visualize data?

Lisa Charlotte Rost: Why do we visualise data?

@lisacrost

# Why do YOU visualize data?

Priorities

Attention & Beauty

## Why Visualizing Data Matters

© 2014 Data Movement LLC

# Why do we visualize data?

Lisa Charlotte Rost: Why do we visualise data? @lisacrost

## Why do YOU visualize?

Priorities

Attention & Beauty

WHY VISUALIZE?

© 2018 Lisa Charlotte Rost

24 x 781

Rational Optimism

Information Graphics

HIDE & SHOW

DATA

visible, understandable, natural, accessible, to the "BARE" eye

Phenomena & Reality

GOAL

Visual Communicator

SYSTEMIC AND EXCITING INTELLECTUAL CHAOS

Information Architects

Defined organization, art + science discipline, community & practice

Design

BLACK HOLE

between DATA & KNOWLEDGE

Visualization as Technology

extensions of self

means to goal

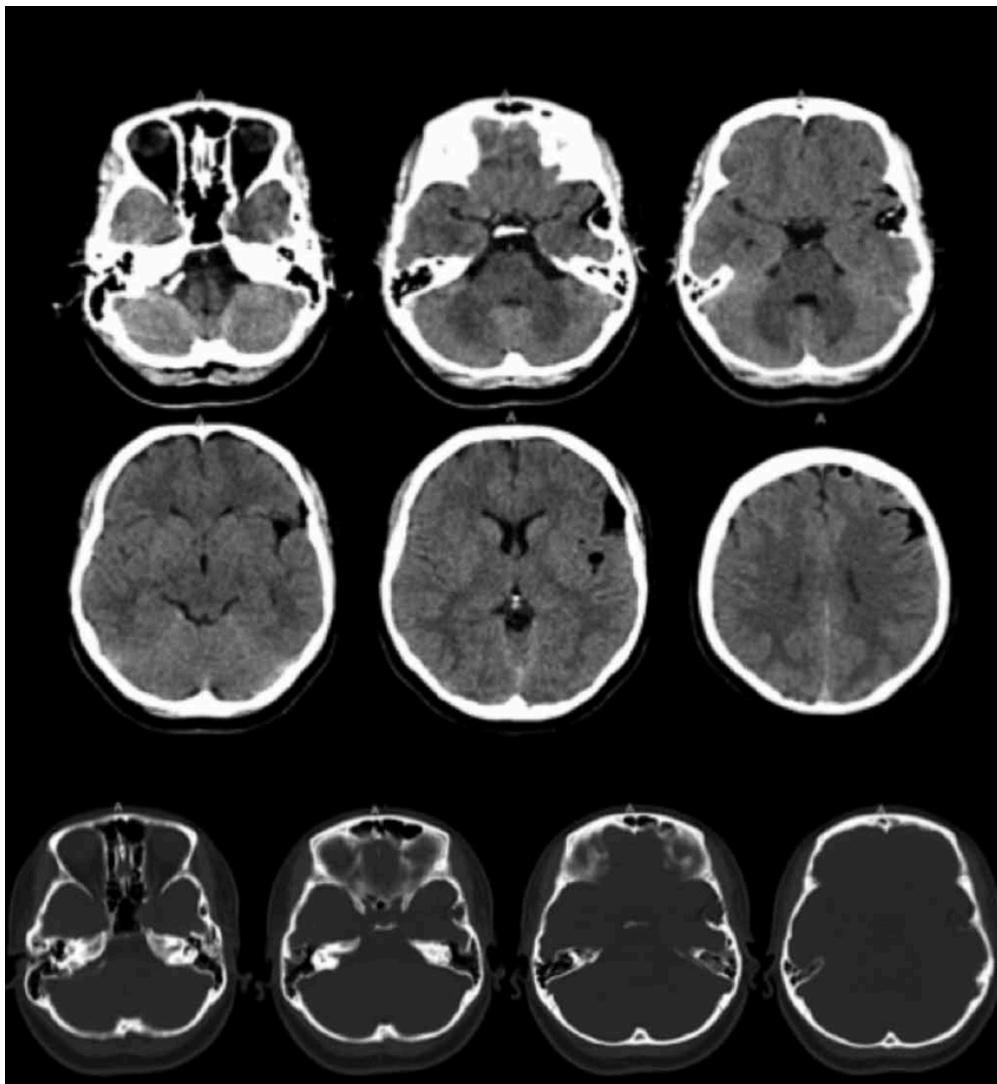
object, process, method

means to fulfill a human purpose

DIKW hierarchies

@hj\_dewaard  
@BY  
01.23.2018

# Dataviz to **record** information



# Dataviz to **explore** data (1a)

1		2		3		4	
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	6.80

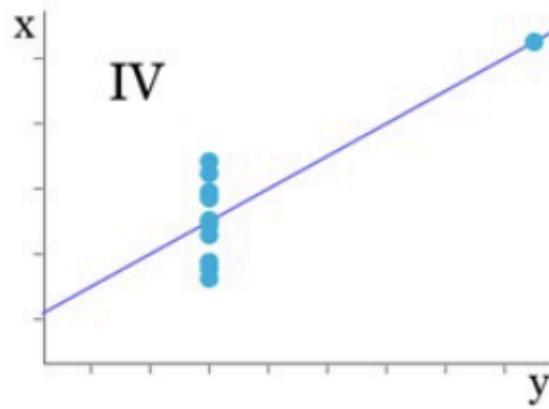
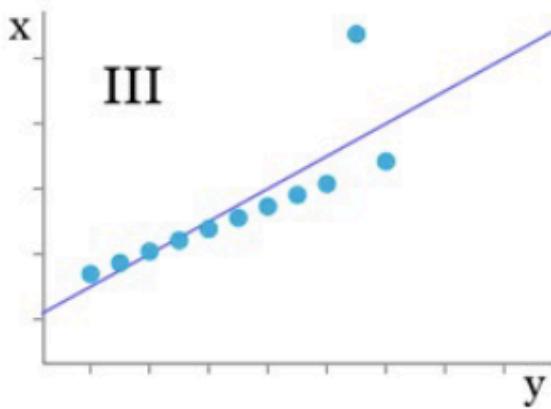
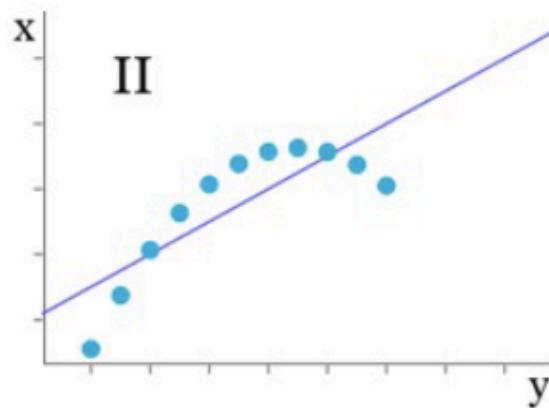
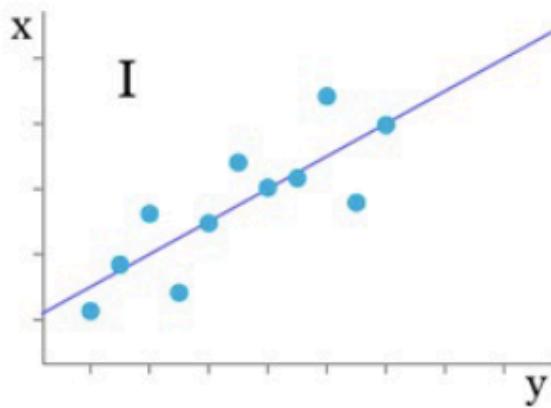
# Dataviz to **explore** data (1b)

For all four datasets:

$N = 11$

- mean of x values = 9
- variance of x values = 11
- mean of y values = 7.5
- variance of y values = 4.12
- correlation between x & y = 0.816
- regression line:  $y = 3 + 0.5x$

# Dataviz to **explore** data (1c)



# Dataviz to communicate (1)

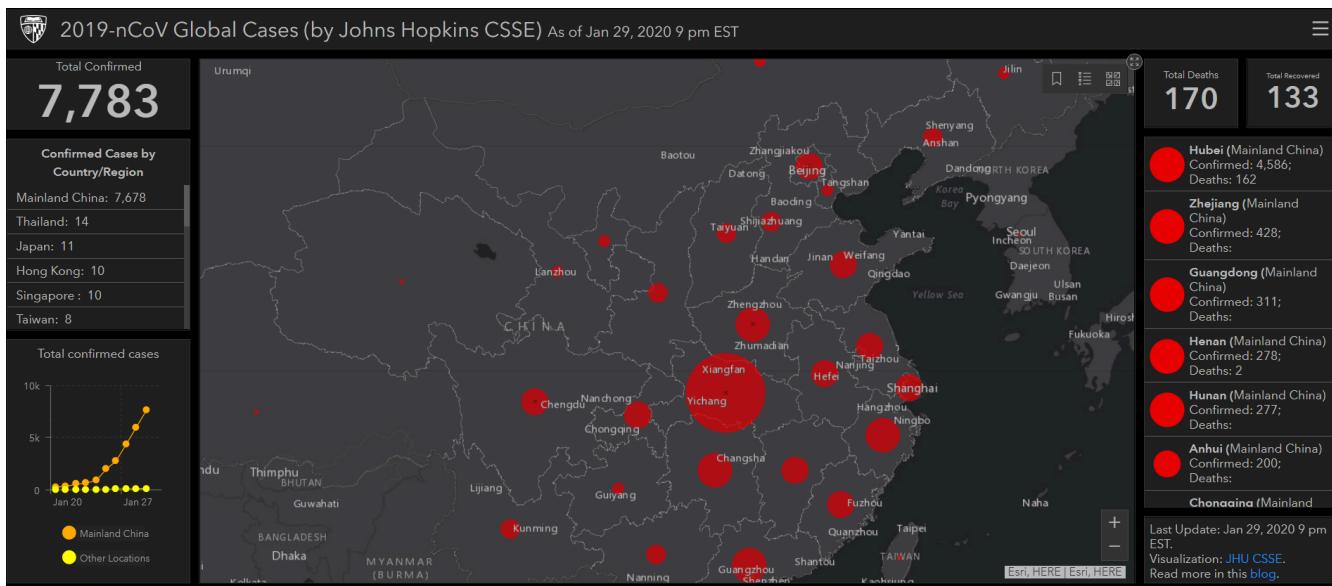
## INFOGRAPHICS FOOD HEALTHY NUTRITION

The infographic is divided into several sections:

- FRUITS**: Includes a watermelon icon and placeholder text.
- VEGETABLES**: Includes a broccoli icon and placeholder text.
- HEALTHY PROTEIN**: Includes a meat icon and placeholder text.
- WHOLE GRAINS**: Includes a bread icon and placeholder text.
- HEALTHY OILS**: Includes a oil bottle icon and placeholder text.
- WATER**: Includes a water bottle icon and placeholder text.
- JUNK FOOD HIGH CALORIE**: Includes a crossed-out fork and knife icon and placeholder text.
- Food Icons**: A central bowl containing various food icons like a pig, chicken, bread, vegetables, fruit, and a cow, surrounded by a flexing muscular arm icon.
- Bottom Row Icons**: A row of small food icons including a head of lettuce, a bulb, a radish, a zucchini, a pig's head, a corn cob, a cow's head, a chicken's head, a fish, and a leaf.

Shutterstock watermark is present throughout the image.

# Dataviz to tell a story (2)



John Hopkins **interactive** tracking dashboard

# Outline of the talk

- What is data visualization? ("WHAT")
- Why we need data visualization? ("WHY")
- **Stages of data visualization ("HOW")**
- Getting more concrete. ("WHICH", "WHEN")
- Tools for Data Visualization ("WITH WHAT")
- Summary (and recommendations)
- References and links

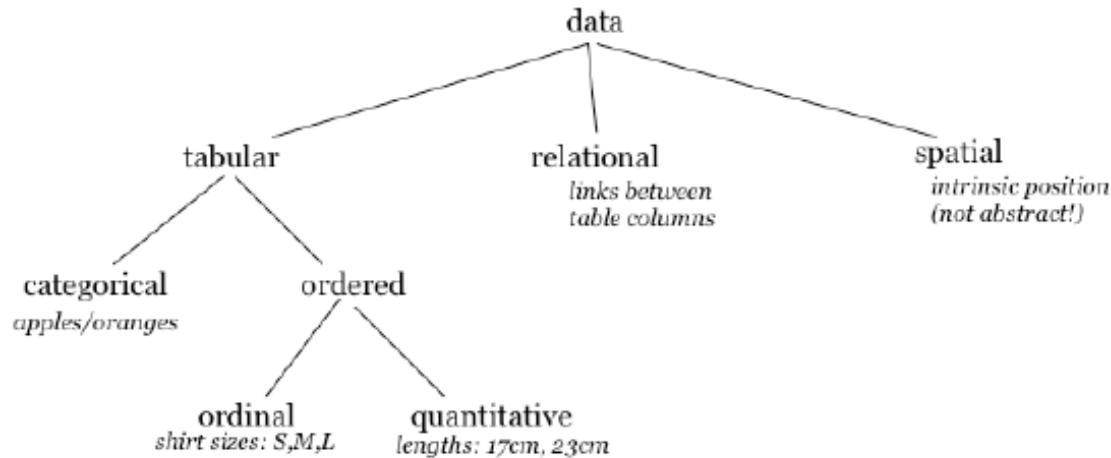
# Stages of data visualization

(“HOW?”)

To get from data to visualization need to understand:

- Properties of data
- Components of image
- Mapping data to image

# Properties of data



# Components of the image

## Visual variables

“marks” - geometric primitives

	Points	Lines	Areas	Best to show
<i>Shape</i>		possible, but too weird to show	cartogram	qualitative differences
<i>Size</i>			cartogram	quantitative differences
H				qualitative differences
V				quantitative differences
S				qualitative differences
<i>Texture</i>				qualitative & quantitative differences

“channels” - control appearance of marks

# Mapping data to images



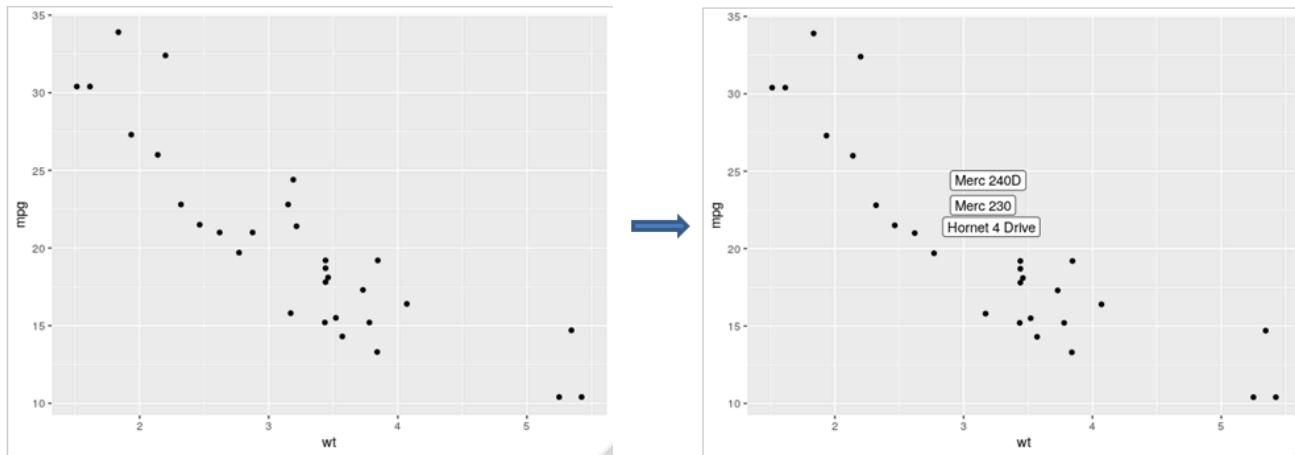
# Principles of data to visualization

- A series of rules to warrant better graphics when followed.
- Problem: Many lists of rules
- Example: **Core principles of Data Visualization**

<https://policyviz.com/product/core-principles-of-data-visualization-cheatsheet/>

# Core principles (1)

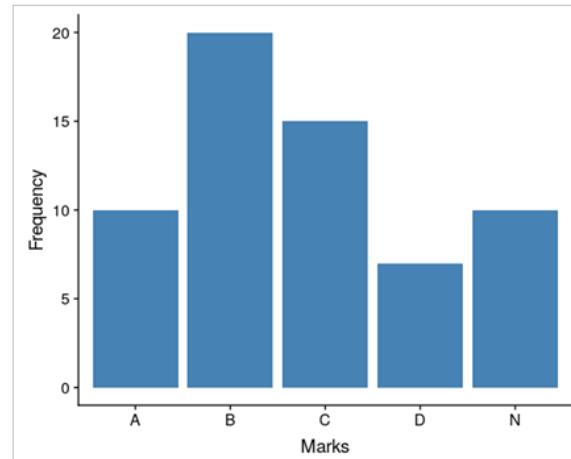
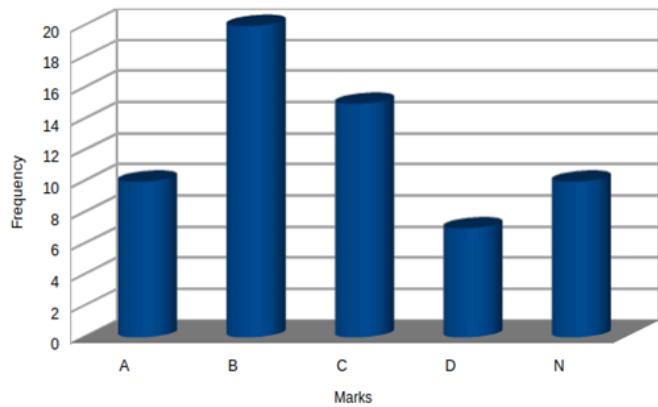
Show your data



Relevant data may be shown in the plot.

# Core principles (2)

Reduce the clutter

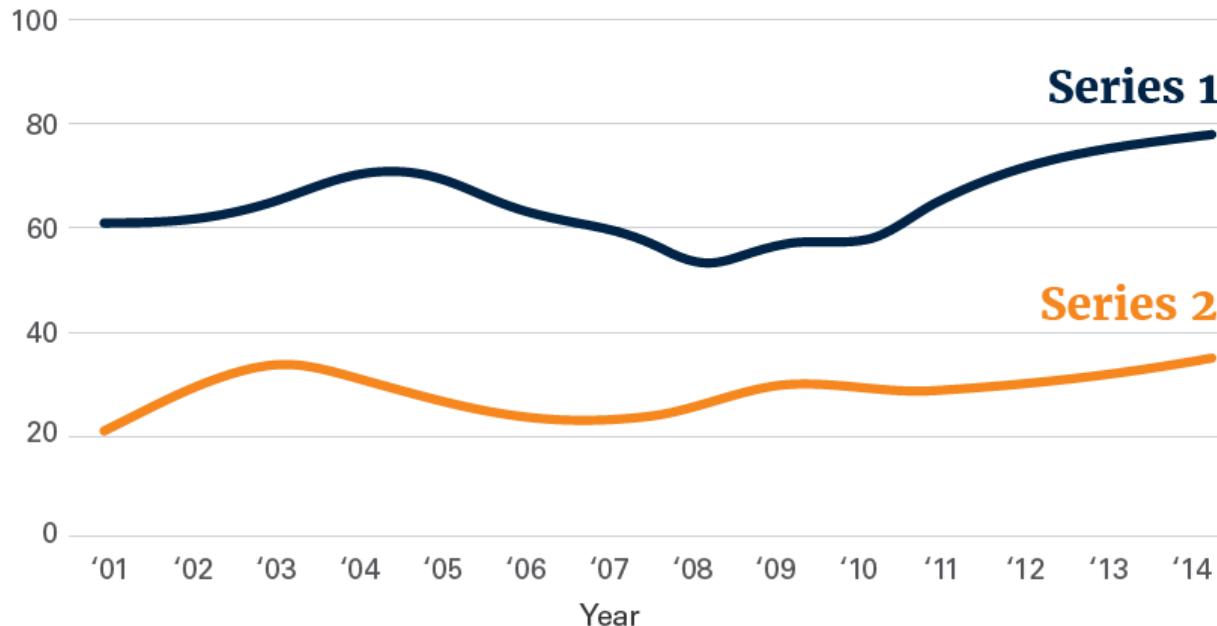


# Core principles (3)

Integrate text and graph

**Chart Title Here**

(Y axis label here)



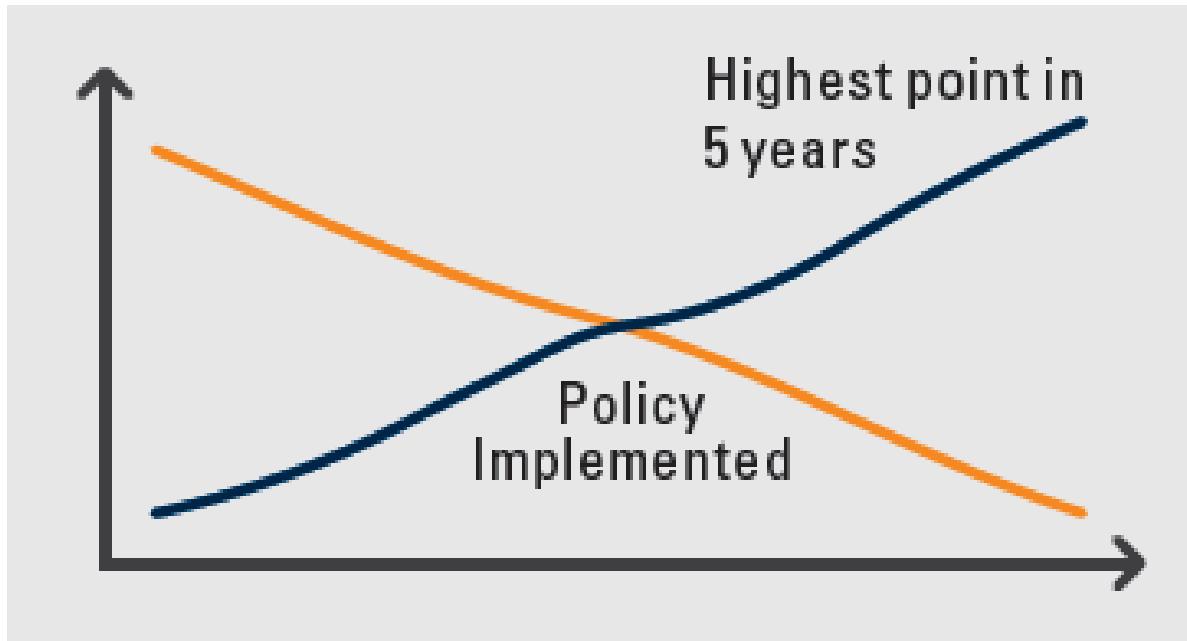
# Core principles

## Audience



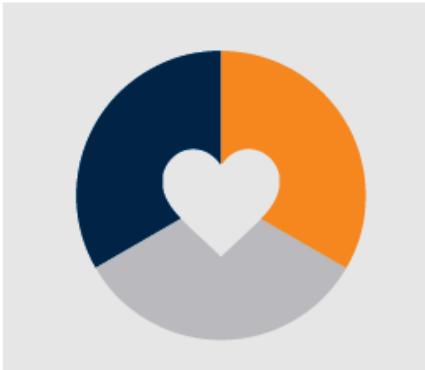
# Core principles

## Include annotation



# Core principles

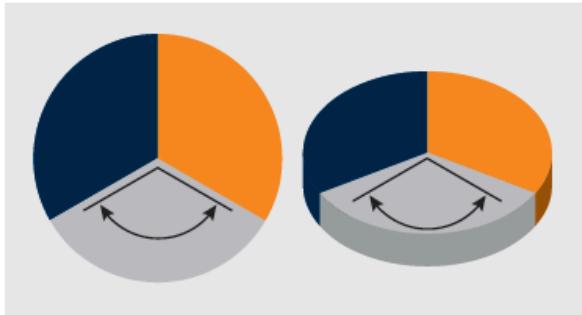
## Use pie charts with care



We are not very good at discerning quantities from the slices of the pie chart. Other chart types—for example, bars, stacked bars, treemaps, or slope charts—may be a better choice.

# Core principles

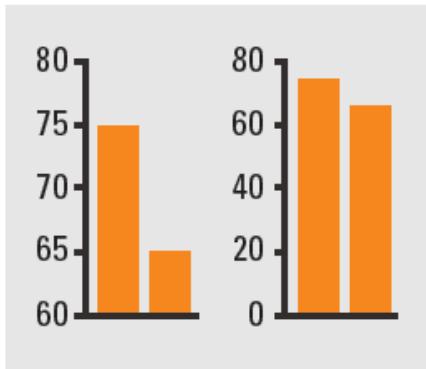
## Avoid 3D



Using 3D when you don't have a third variable will usually distort the perception of the data and should thus be avoided.

# Core principles

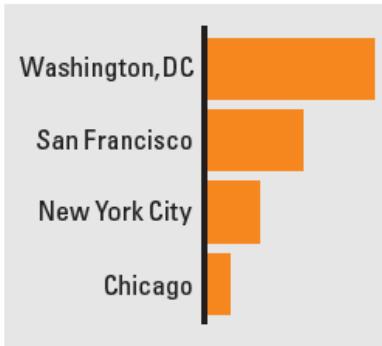
## Start bar and column charts at zero



Bar and column charts that do not start at zero overemphasize the differences between the values. For small changes in quantities, consider visualizing the difference or the change in the values.

# Core principles

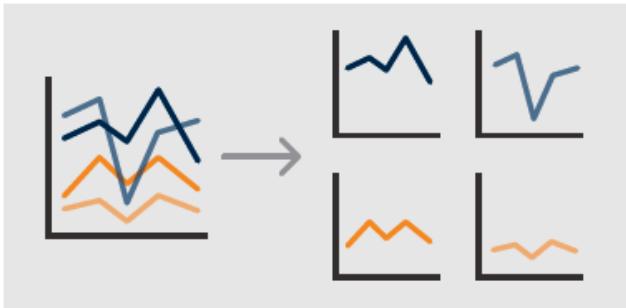
## Make labels easy to read



When applicable, rotate bar and column charts to make the labels horizontal. If possible, make vertical axis labels horizontal, possibly below the title. In general, make labels clear, concise, and easy for your reader to understand.

# Core principles

## Try small multiples



Breaking up a complicated chart into smaller chunks can be an effective way to visualize your data.

# Core principles

## Use maps carefully

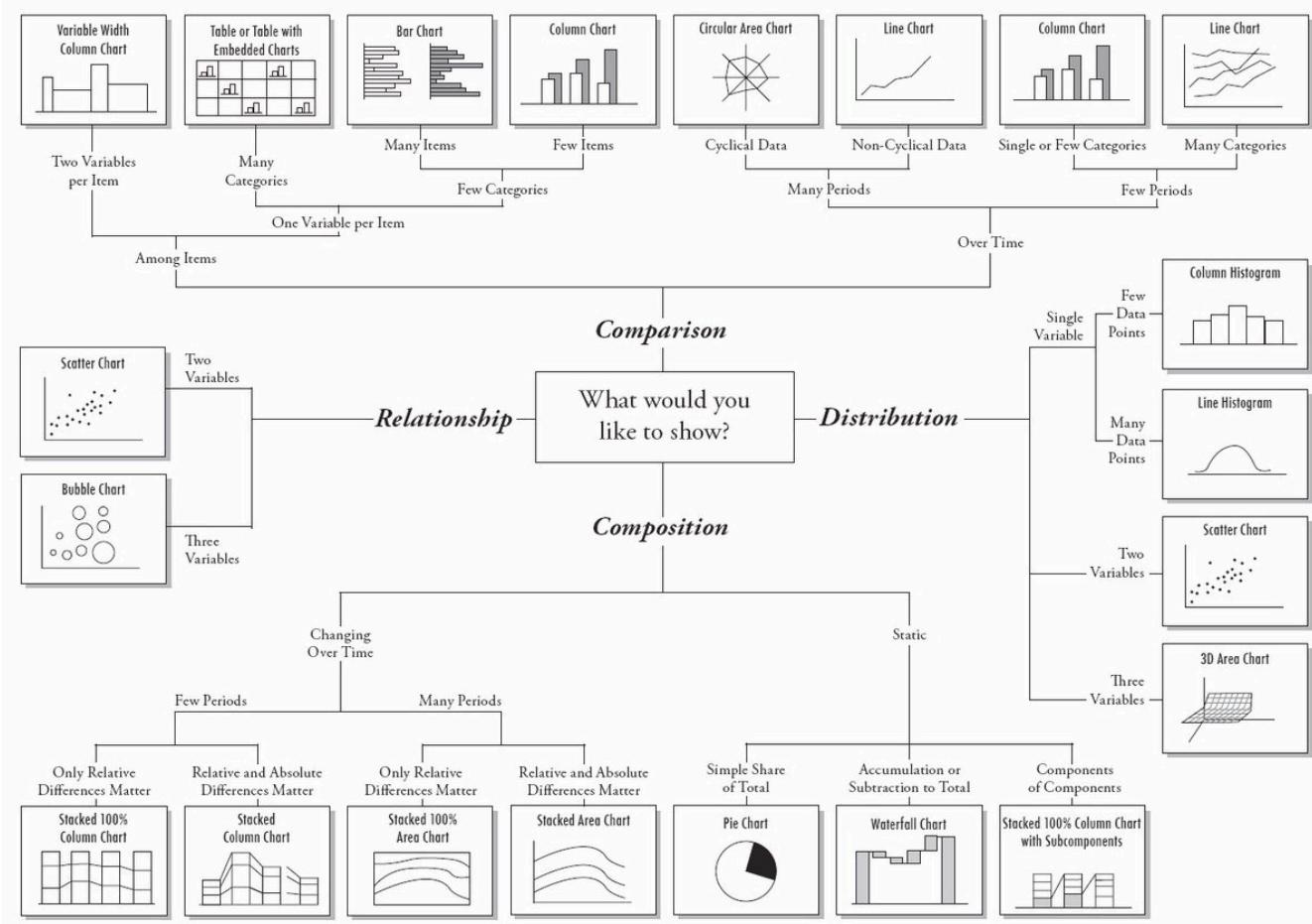


Use maps carefully, always being sure it is the geographic point you are trying to make. Column and bar charts, for example, are often better at enabling comparisons between geographic units.

# Outline of the talk

- What is data visualization? ("WHAT")
- Why we need data visualization? ("WHY")
- Stages of data visualization ("HOW")
- **Getting more concrete. ("WHICH", "WHEN")**
- Tools for Data Visualization ("WITH WHAT")
- Summary (and recommendations)
- References and links

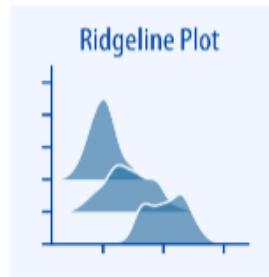
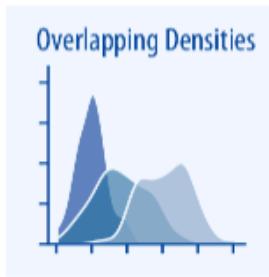
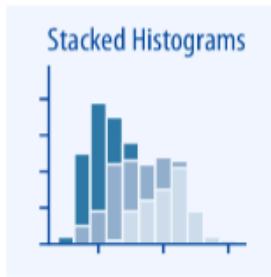
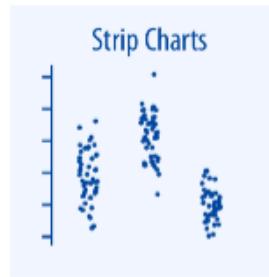
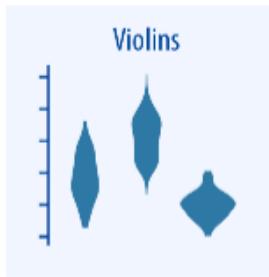
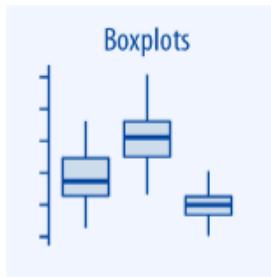
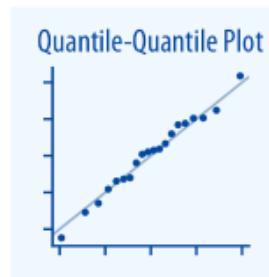
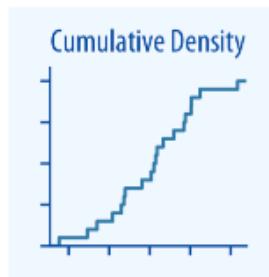
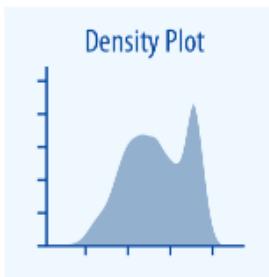
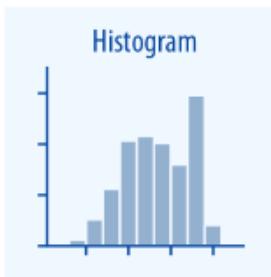
# What would you like to show?



You have to know your point!

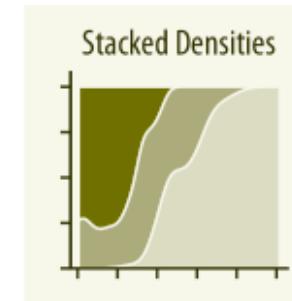
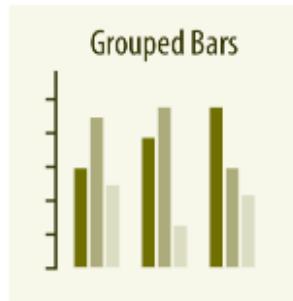
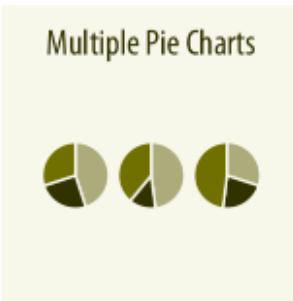
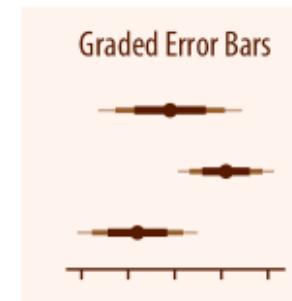
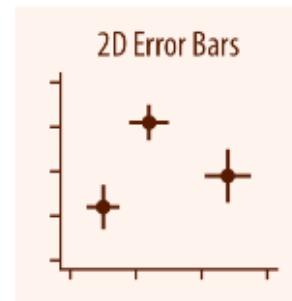
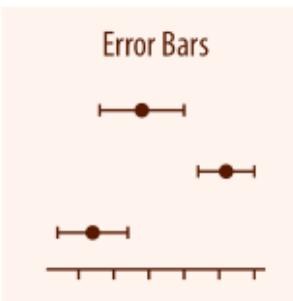
# Distribution

Where do your values fail? Are there outliers?



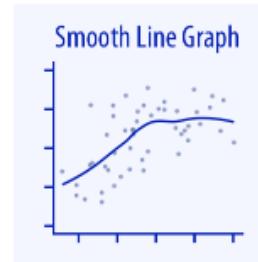
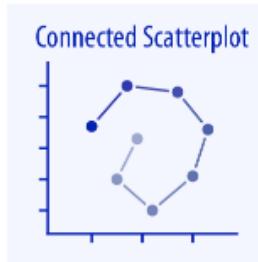
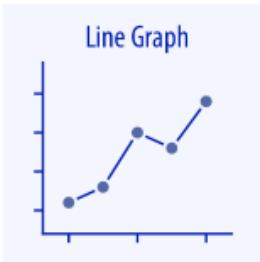
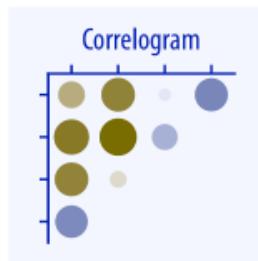
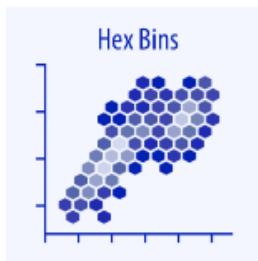
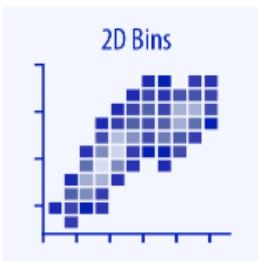
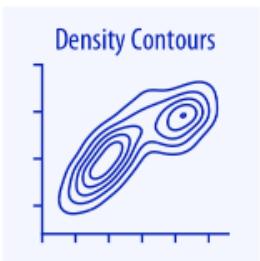
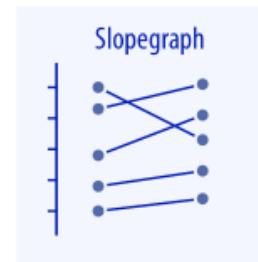
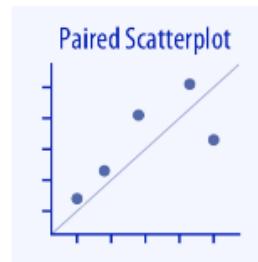
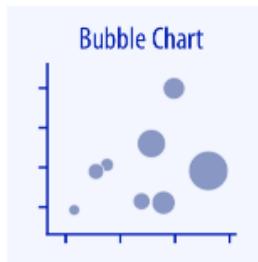
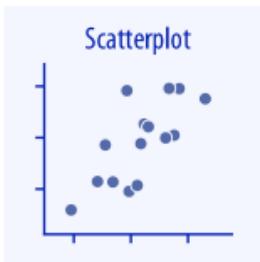
# Comparison

- How are these values similar or different?



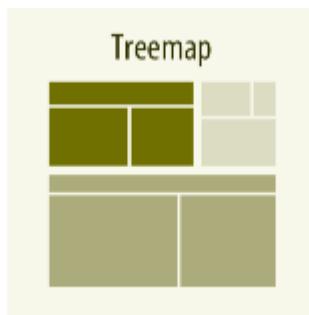
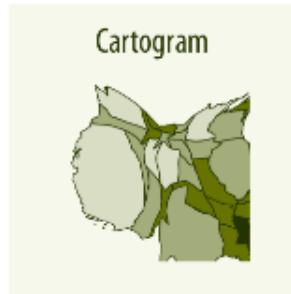
# Relationship

- How do these variables relate to one another?



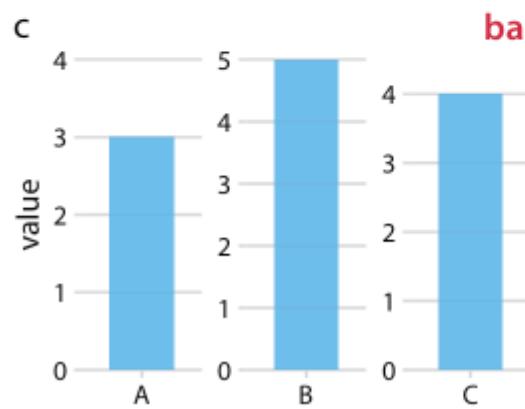
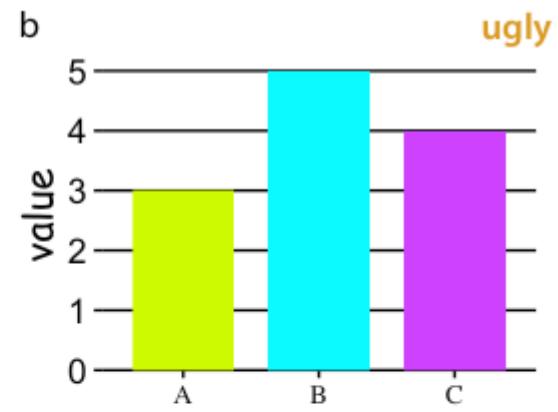
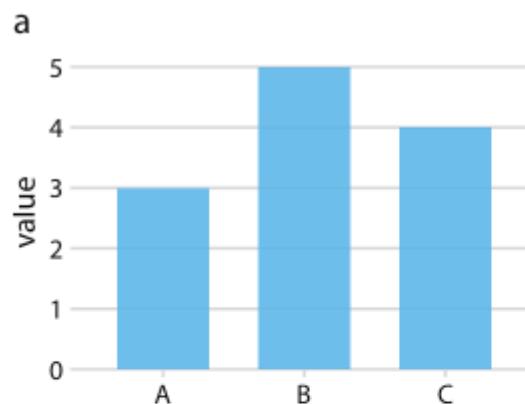
# Composition.

What's parts make up the whole?

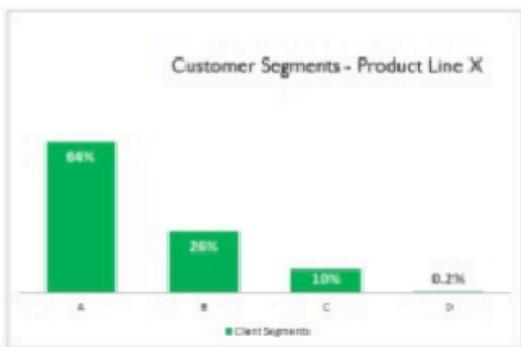
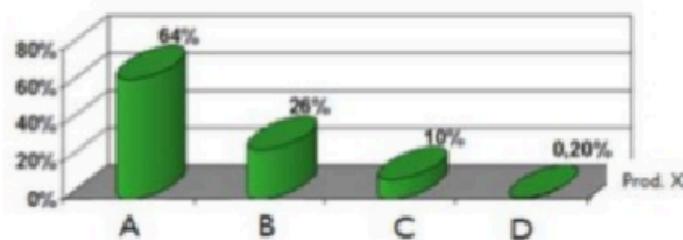
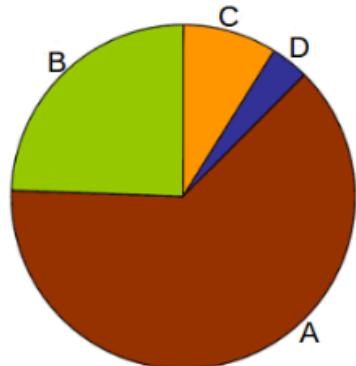


# DO'S & DONT'S

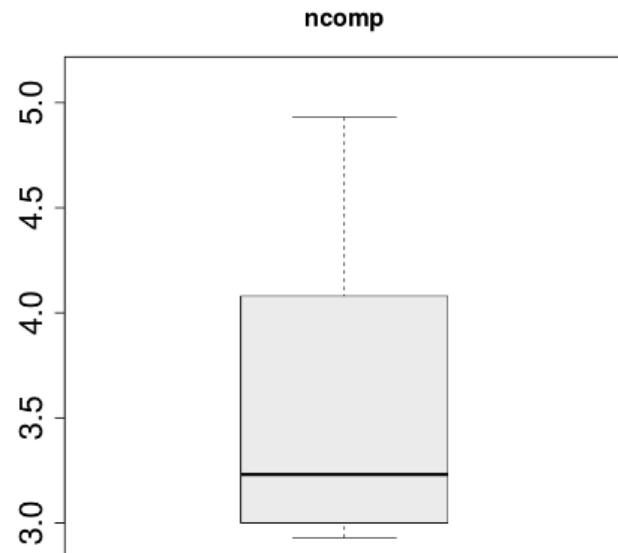
- There is room for improvement



# Distribution. Do's & Dont's



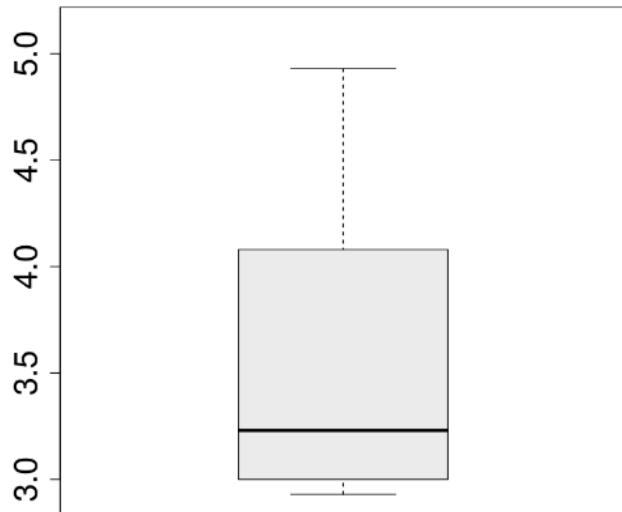
# Distribution. Do's & Dont's



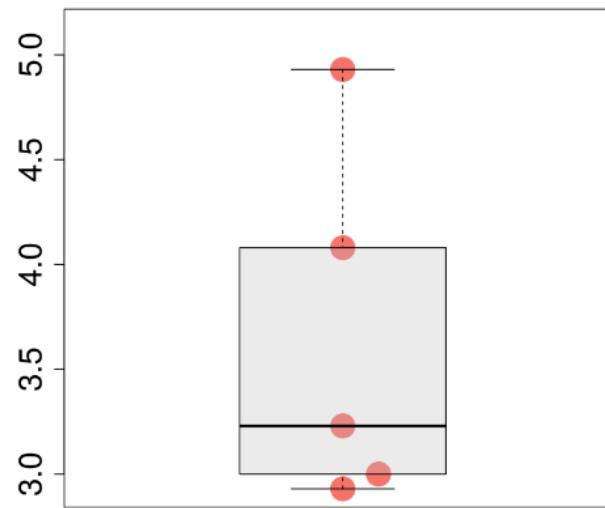
# Distribution. Do's & Dont's



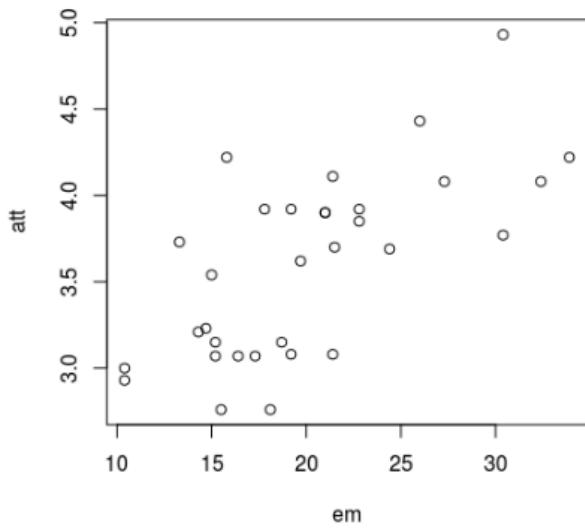
ncomp



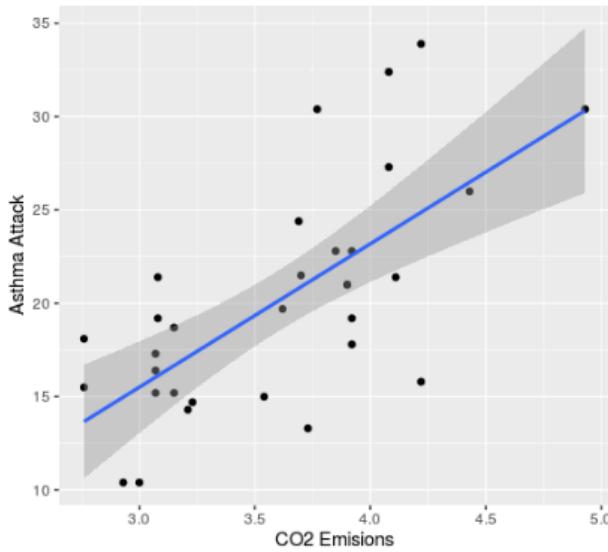
ncomp



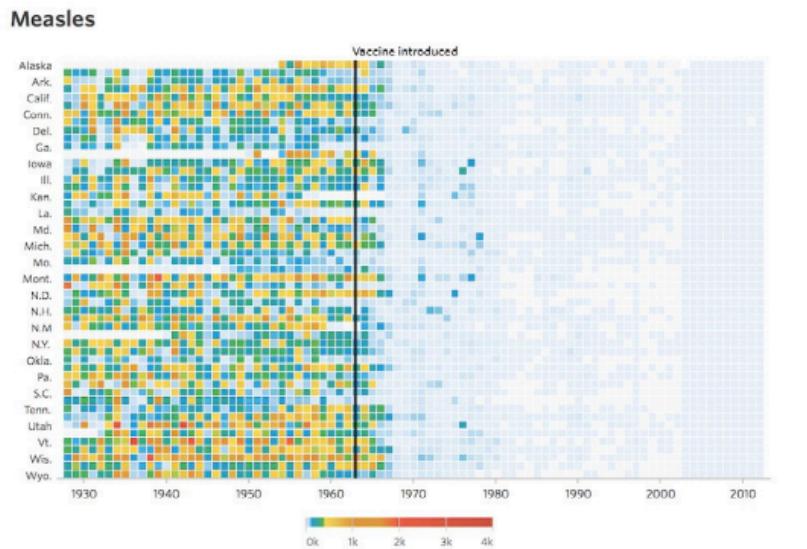
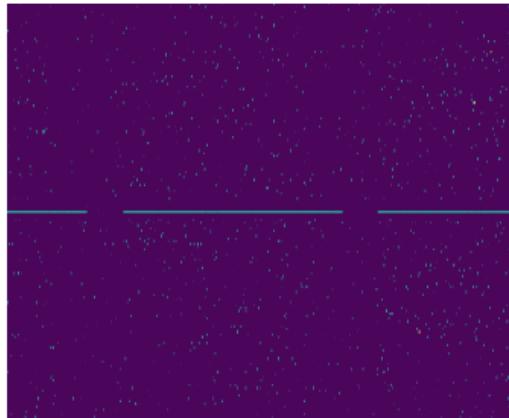
# Relationship. Do's & Dont's



Correlation between Asthma Attack  
and CO2 emissions



# Composition. Do's & Dont's.



# Composition. Do's & Dont's.



# Composition. Do's & Dont's.



## Casos confirmados en Asia

Actualizado a 29 de enero



# Outline of the talk

- What is data visualization? ("WHAT")
- Why we need data visualization? ("WHY")
- Stages of data visualization ("HOW")
- Getting more concrete. ("WHICH", "WHEN")
- **Tools for Data Visualization ("WITH WHAT")**
- Summary (and recommendations)
- References and links

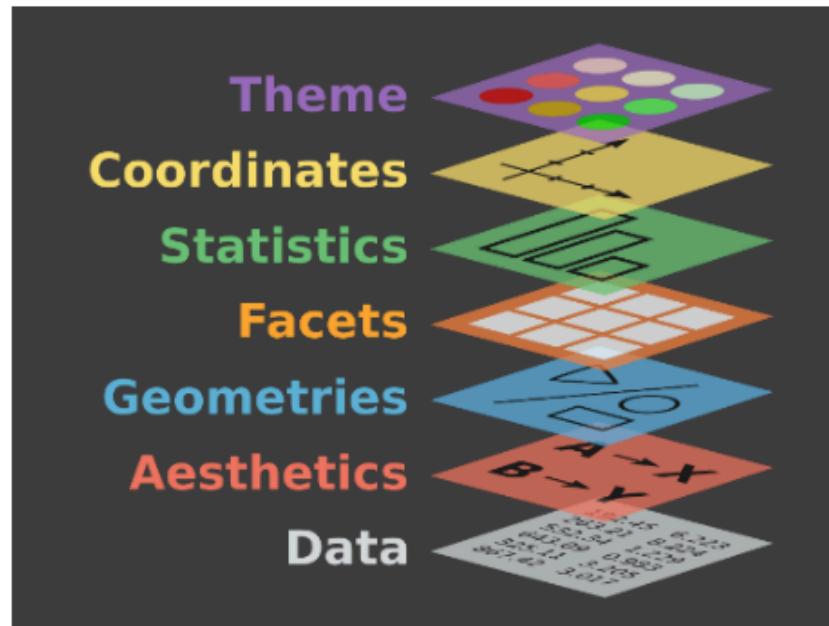
# TOOLS

# THE GLAMOUR OF GRAPHICS



# R & Cia

- ggplot2



<https://www.data-to-viz.com/>

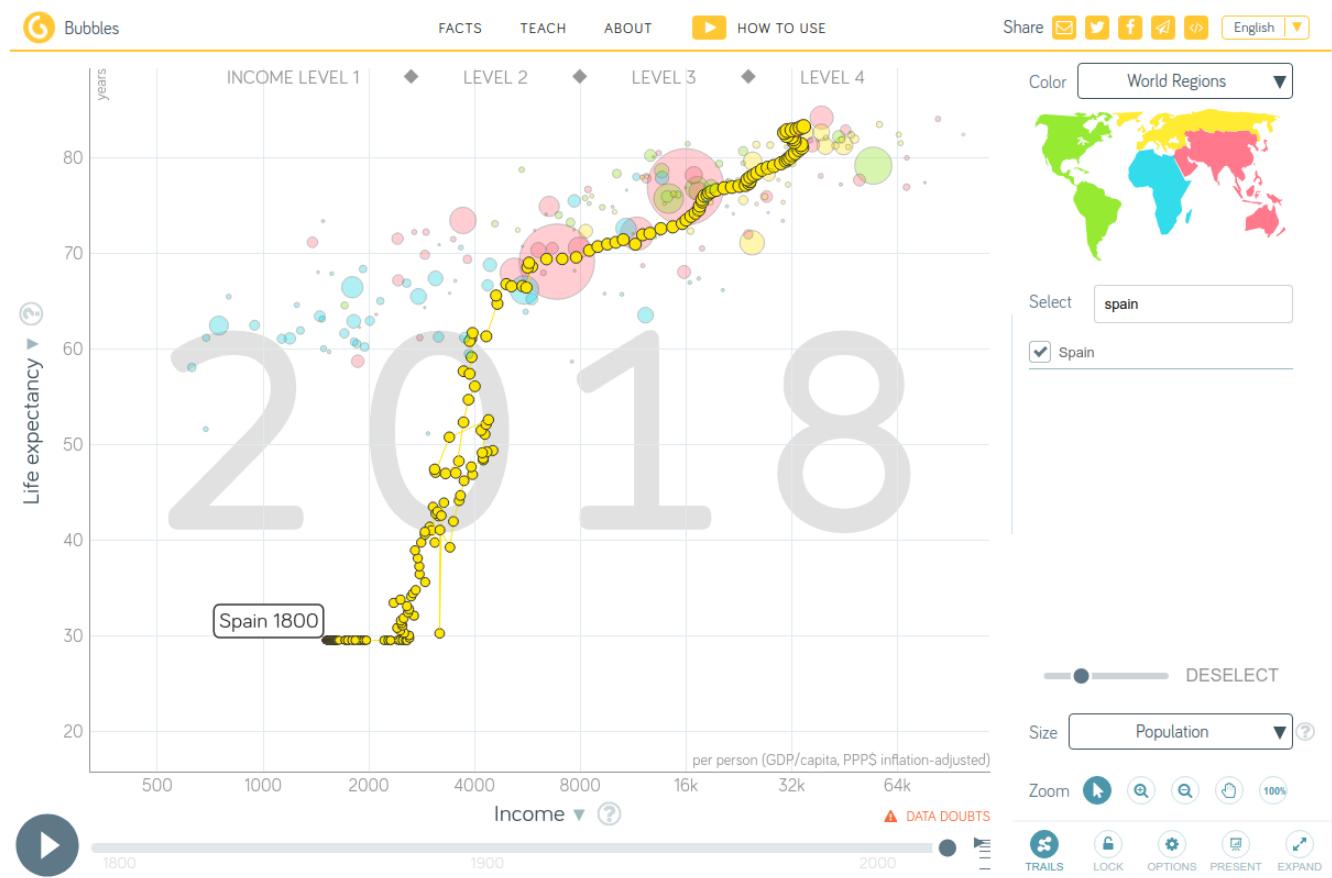
# R Friendly



<https://www.rcommander.com/>

<https://www.jamovi.org/>

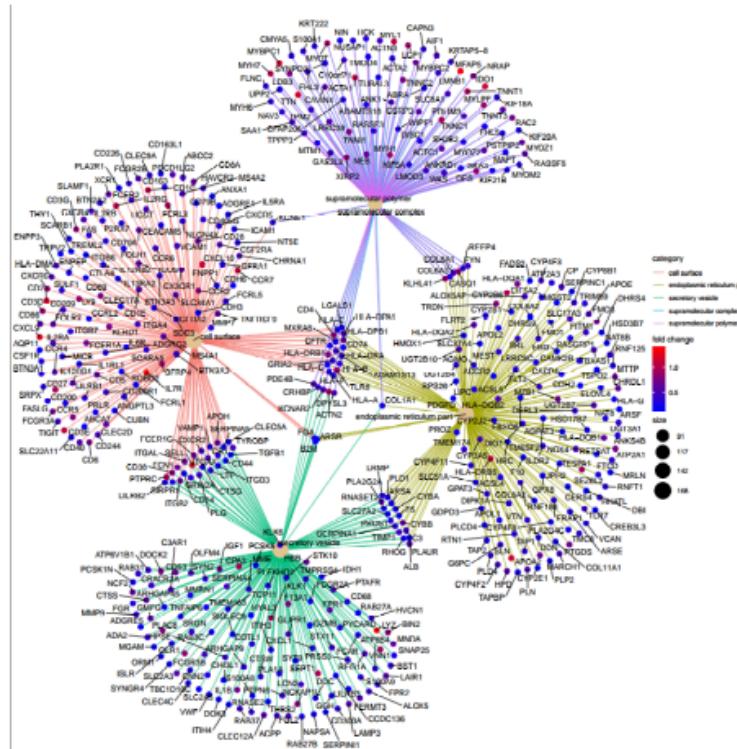
# Gapminder



<https://www.gapminder.org/>

# Omics data visualization

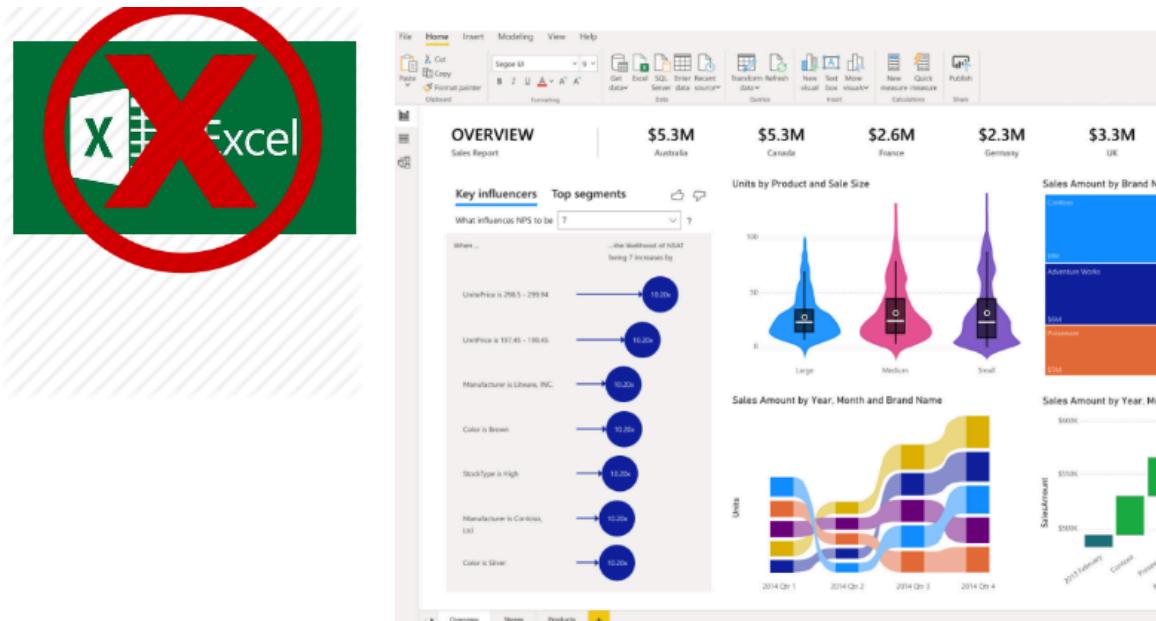
- Cytoscape
- GSEA
- EnrichmentMap
- PathVisio
- g: profiler



<https://cytoscape.org/>

# Power BI

- Dashboard tool



<https://powerbi.microsoft.com/es-es/>

# Outline of the talk

- What is data visualization? ("WHAT")
- Why we need data visualization? ("WHY")
- Stages of data visualization ("HOW")
- Getting more concrete. ("WHICH", "WHEN")
- Tools for Data Visualization ("WITH WHAT")
- **Summary (and recommendations)**
- References and links

# Summary and recommendations

- Data visualization is complex
  - Easy to do well but "sub-optimally"
  - Easy to make mistakes
- To improve
  - Adhere to graphical design principles
  - Follow recommendations based on "how to do data"
- Do the right visualization, the right way, with the right tool.

# QUESTIONS?

# Next 'Píndola'



Busqueu la fama, i aquí és on aneu a començar a pagar, amb suor

Estratègies per a la construcció de models i biomarcadors



28 de Febrer . 12:30 h

