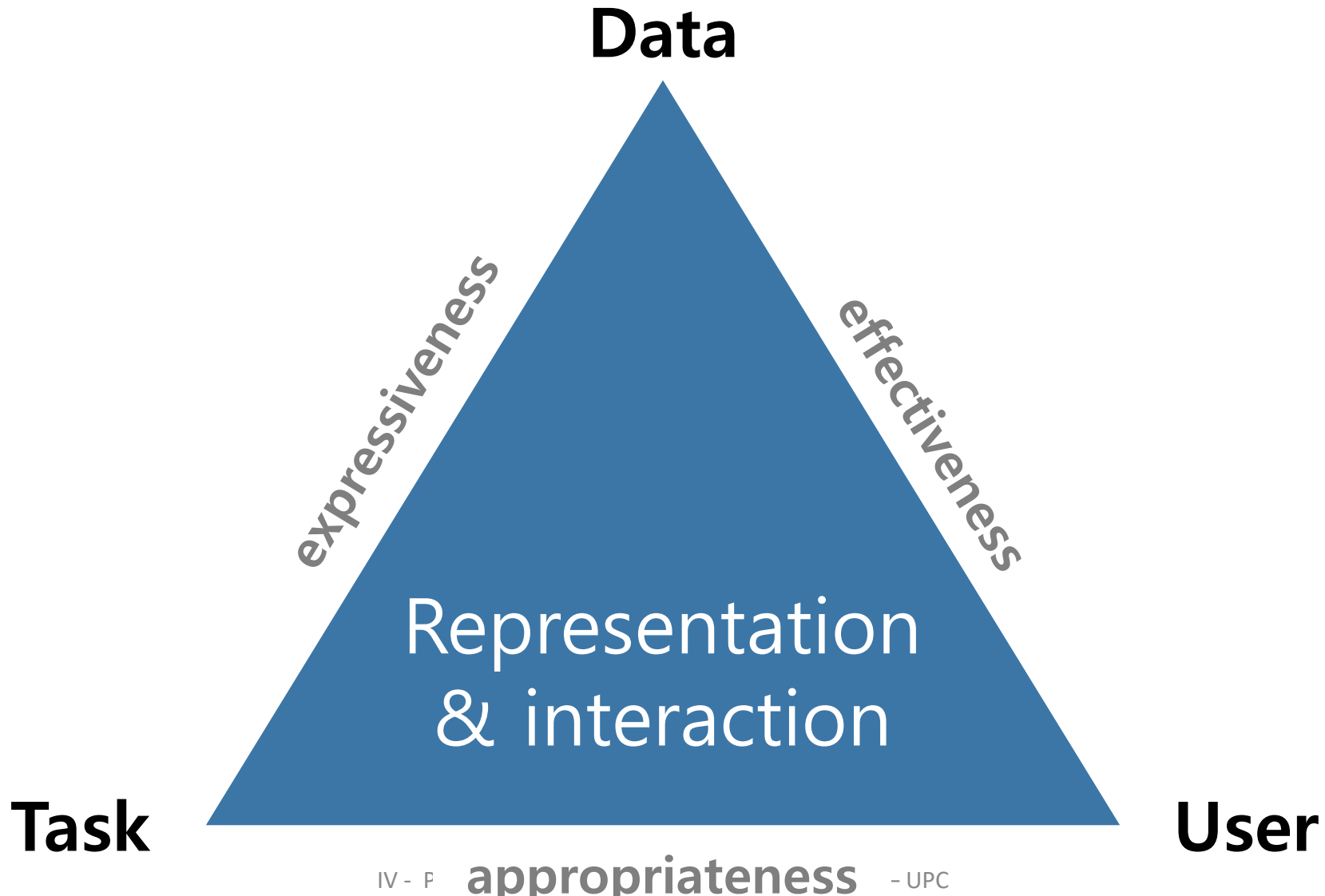
INTRODUCTION TO VISUALIZATION

PERE-PAU VÁZQUEZ – VIRVIG GROUP – UPC

"Computer-based visualization systems provide **visual representations** of **datasets** designed to **help people** carry out **tasks** more effectively."

Tamara Munzner

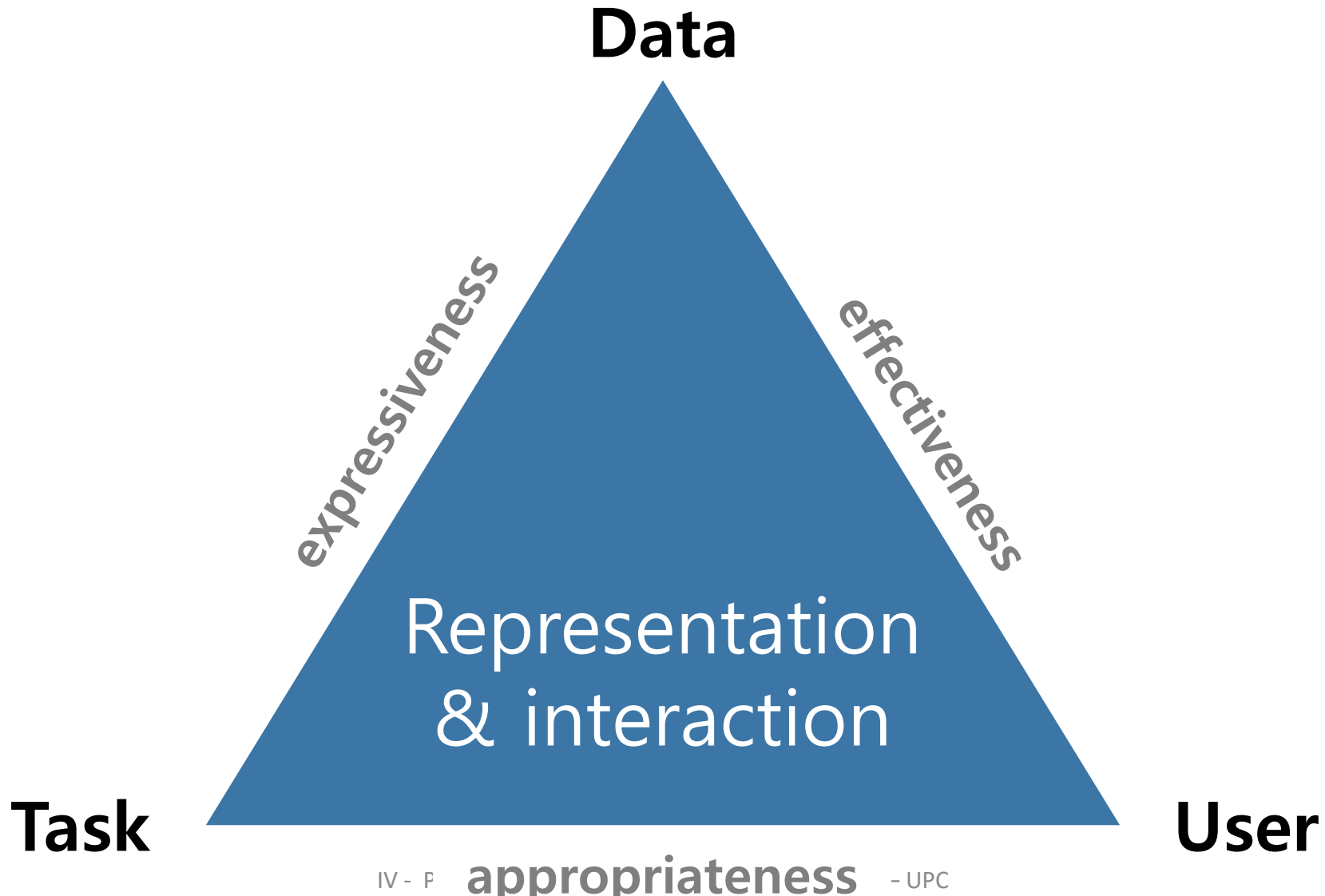
VISUALIZATION



VISUALIZATION

- Who are the users?
  - Their needs, their work environment, their background, physical condition, cultural specifics, hardware specifics…
- What is the data?
  - Scale (quantitative vs qualitative), type (1-dimensional, 2D, …, temporal…), number of variables…
- What are the tasks?
  - Low level (e.g., search for an element) or high level (e.g., find the most successful runner in different categories)

VISUALIZATION



VISUALIZATION

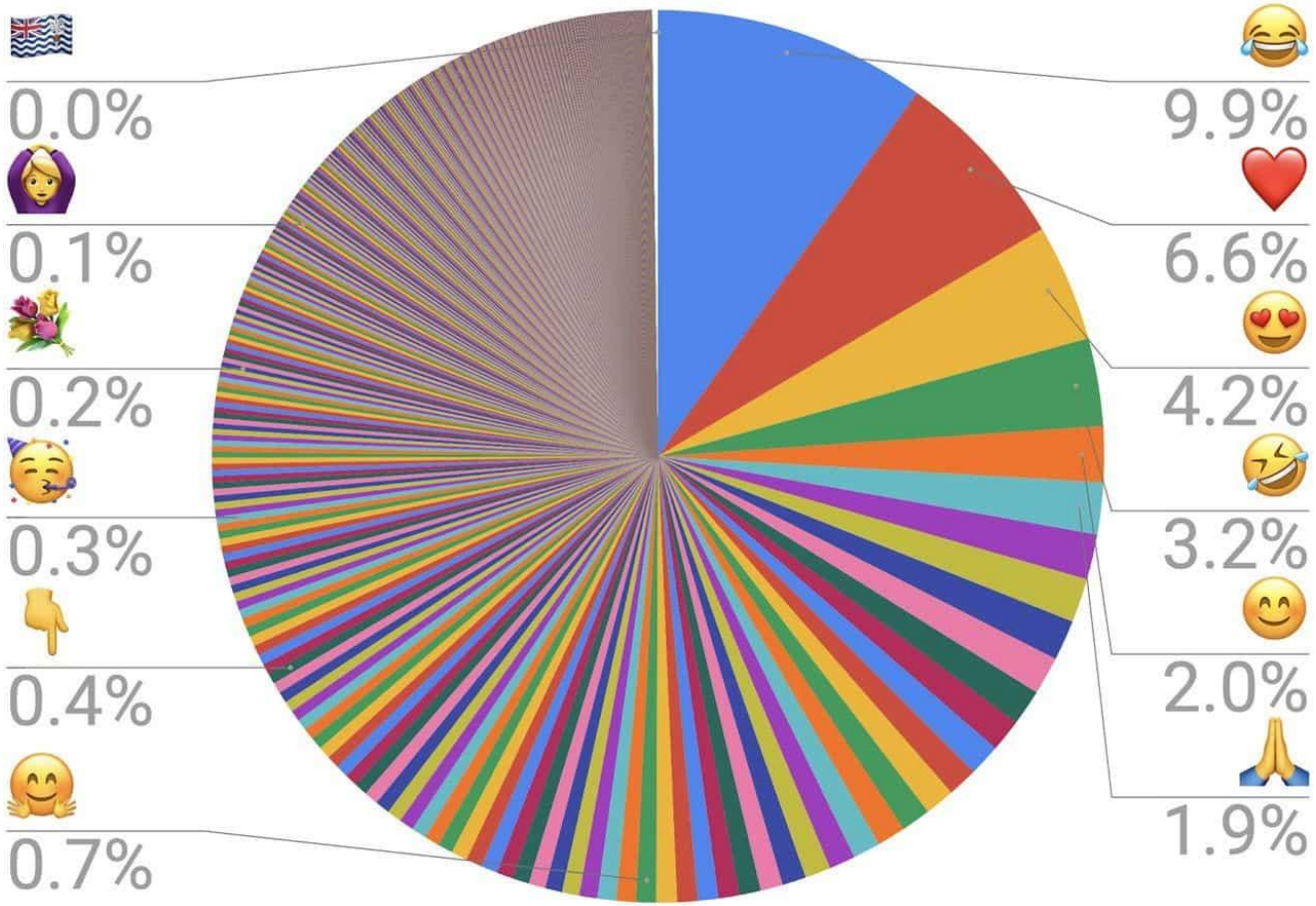
- *Expressiveness:* Show exactly **the information in the data**
  - Nothing more and nothing less
- *Effectiveness:* Take into account the **cognitive capabilities** of the human visual system, and
  - the task, application background, and other context-related information…
- *Appropriateness:* Cost-value ratio that assesses the benefit of the visualization process with respect to achieving the task
  - Mainly **time** (computation) and **space** (screen-space) efficiency

VISUALIZATION

- Carry out tasks more **effectively**
  - Effectiveness requires match between data/task and representation
    - Set of representations is huge
    - Many are ineffective mismatch for specific data/task combo
    - Increases chance of finding good solutions if you understand full space of possibilities
  - What counts as effective?
    - Novel: enable entirely new kinds of analysis
    - Faster: speed up existing workflows
  - How to validate effectiveness
    - Many methods, must pick appropriate one for your context

NOT VISUALIZATION

We got a data set and we throw it to any chart type



https://www.oldstreetsolutions.com/good-and-bad-data-visualization

NOT VISUALIZATION

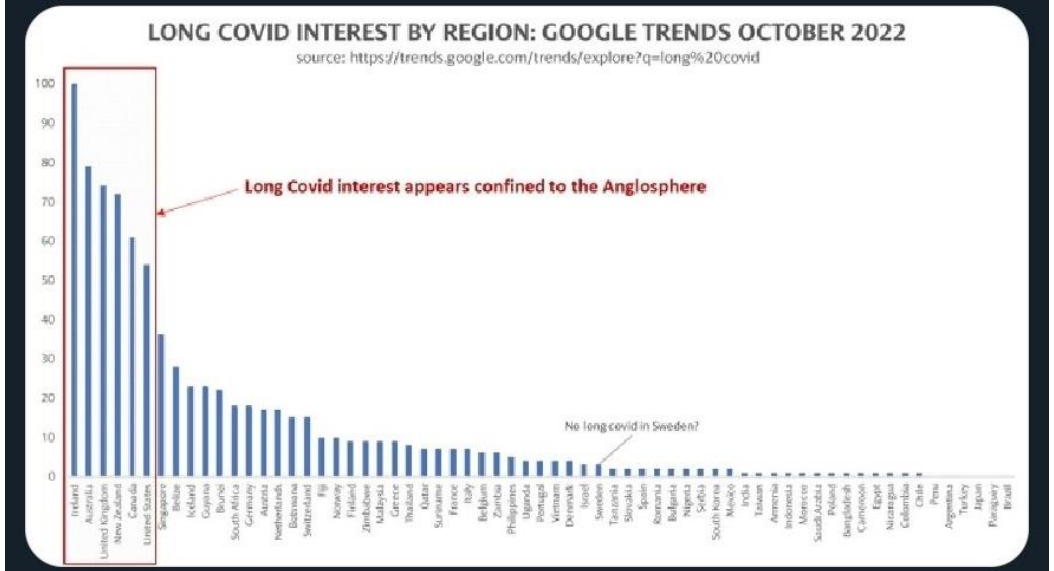
NOT VISUALIZATION

Get some [random] data and create a “visualization” from it

Based upon google searches, it appears that "long covid" only exists in the anglosphere nations: Ireland, Australia, UK, Canada, New Zealand, and the USA.

Does covid somehow know if you speak English?

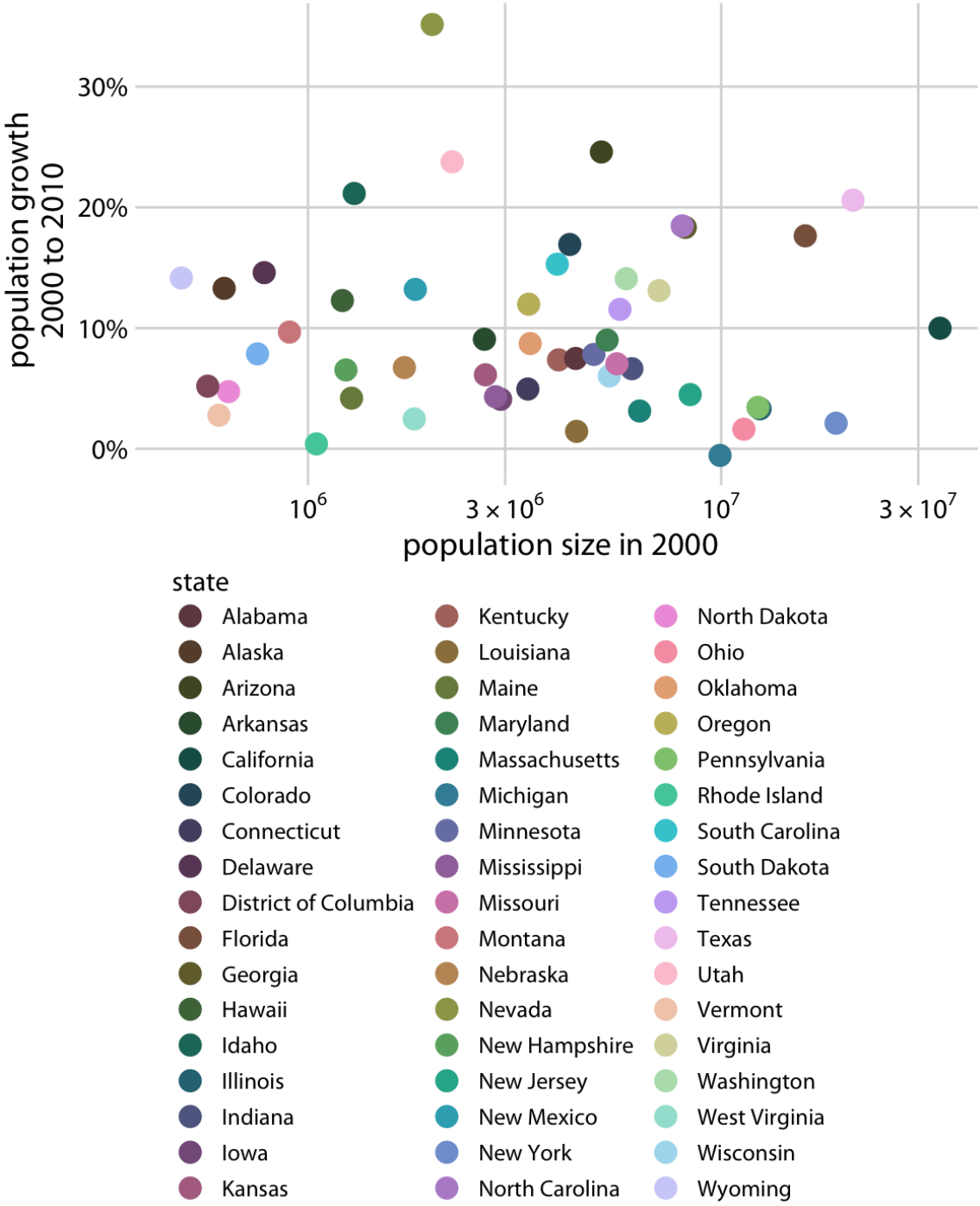
Why no interest in long covid in Japan or Sweden or Italy or Brazil?



373 702 1,694

NOT VISUALIZATION

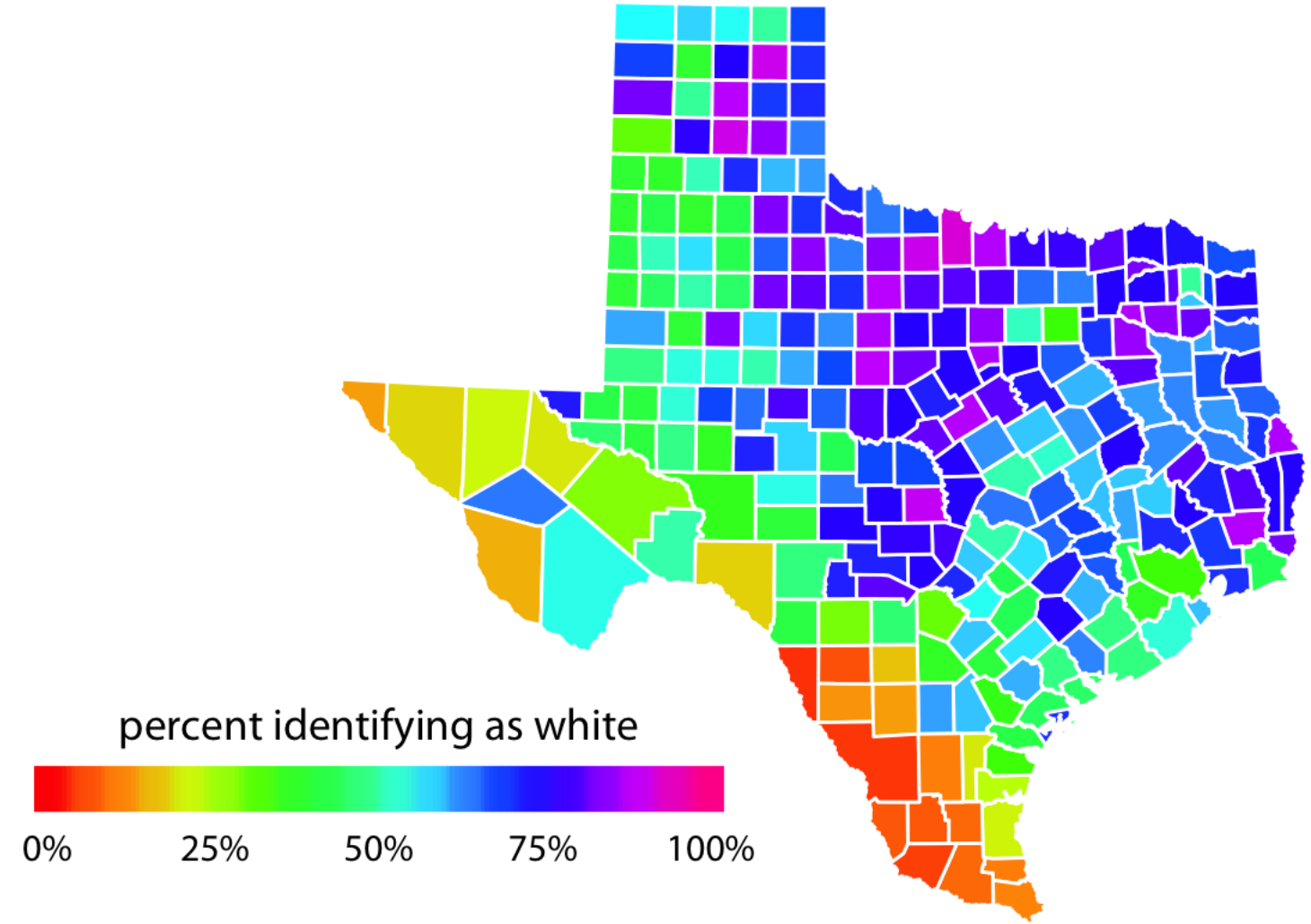
Encoding too much or irrelevant information



https://clauswilke.com/dataviz/index.html

NOT VISUALIZATION

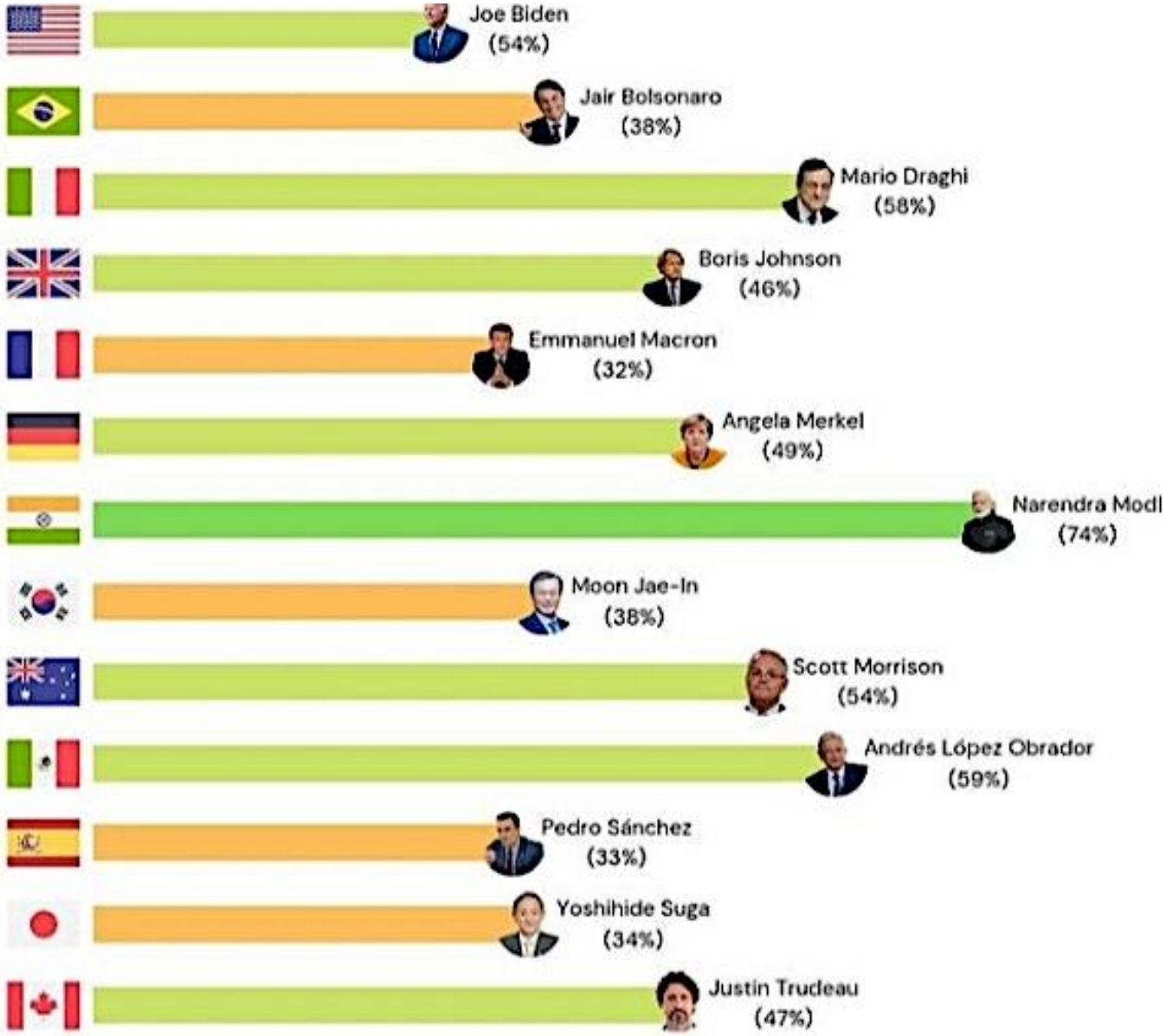
Using unsuitable palettes



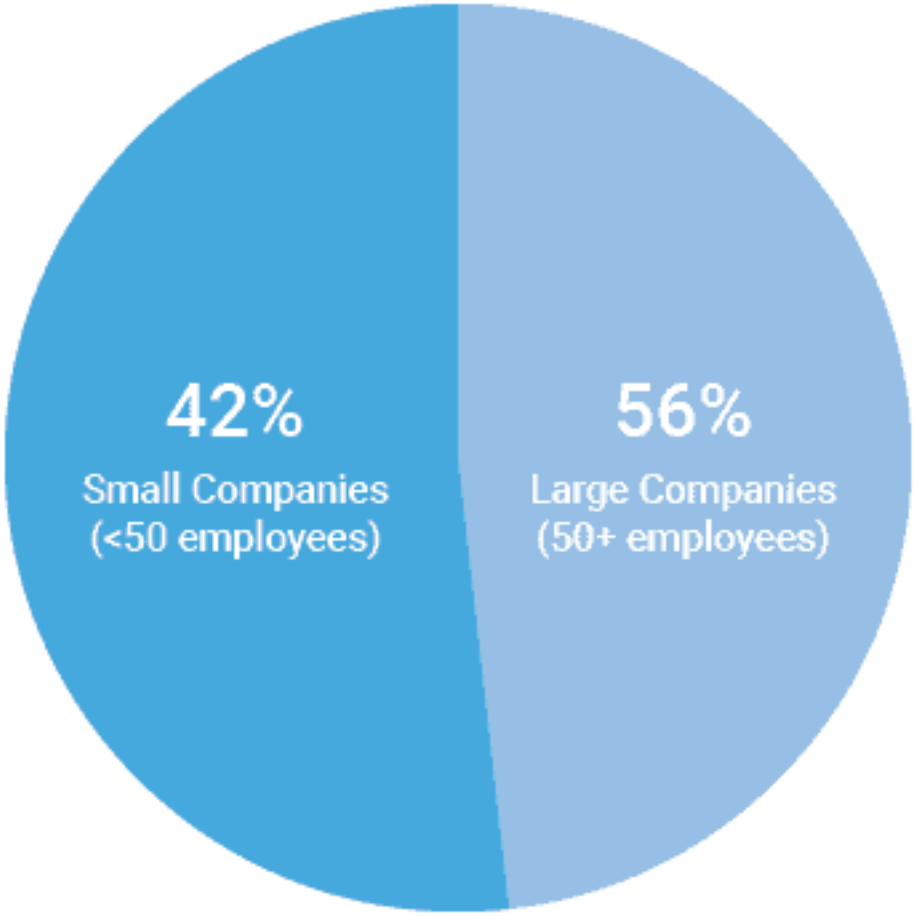
https://clauswilke.com/dataviz/index.html

EXAMPLES

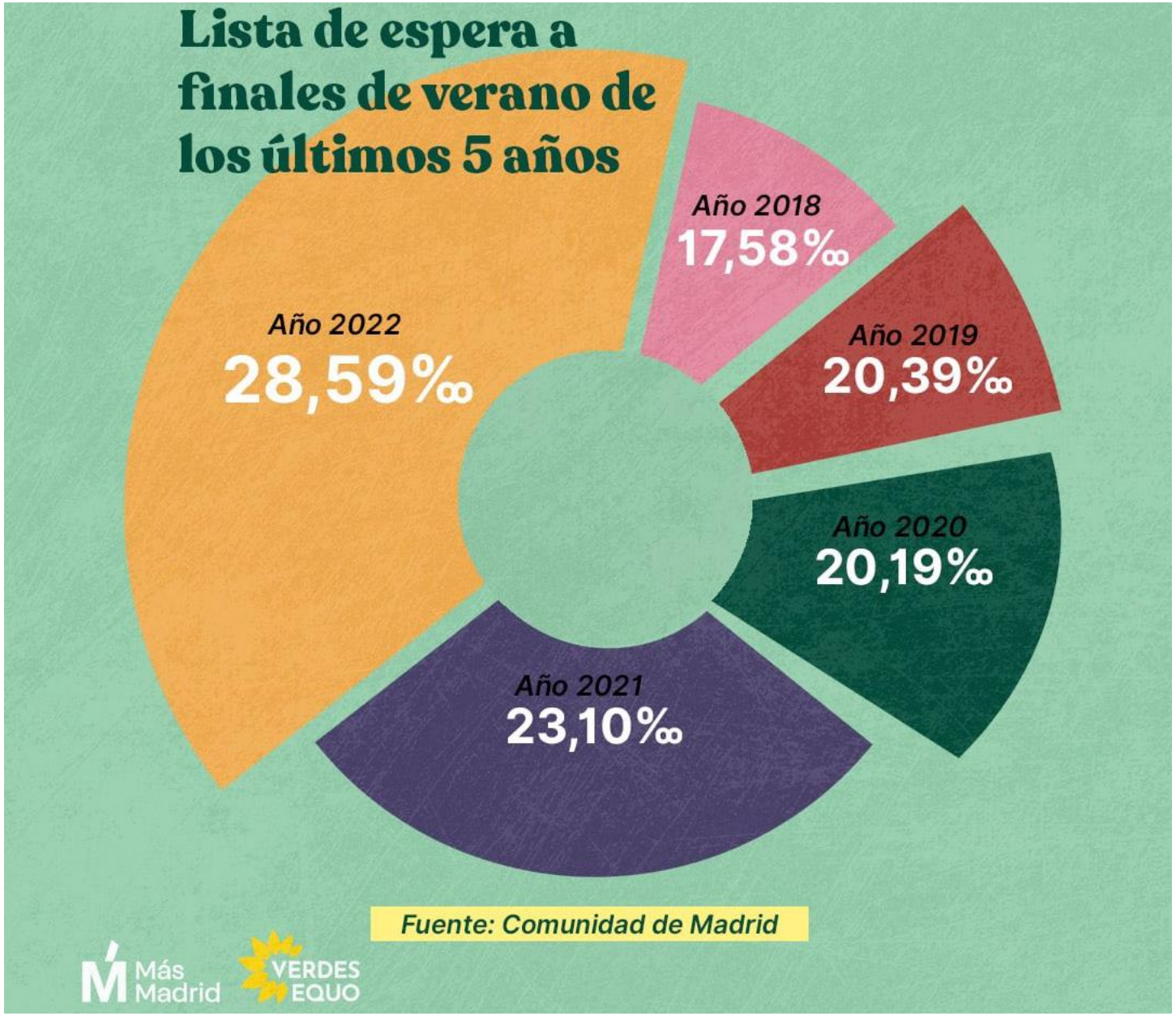
Approval rate of some world leaders



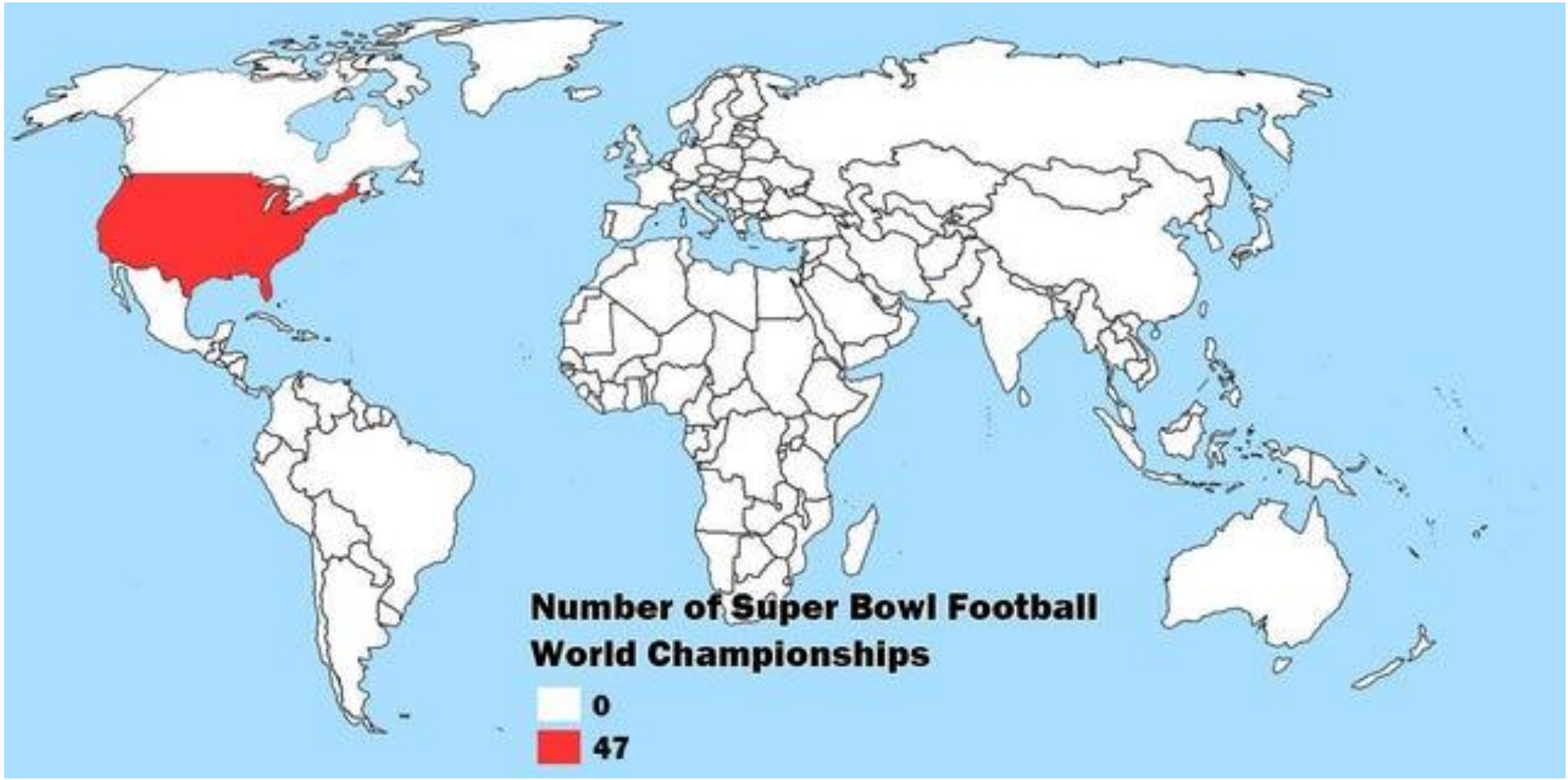
EXAMPLES


42%
Small Companies
(<50 employees)
56%
Large Companies
(50+ employees)

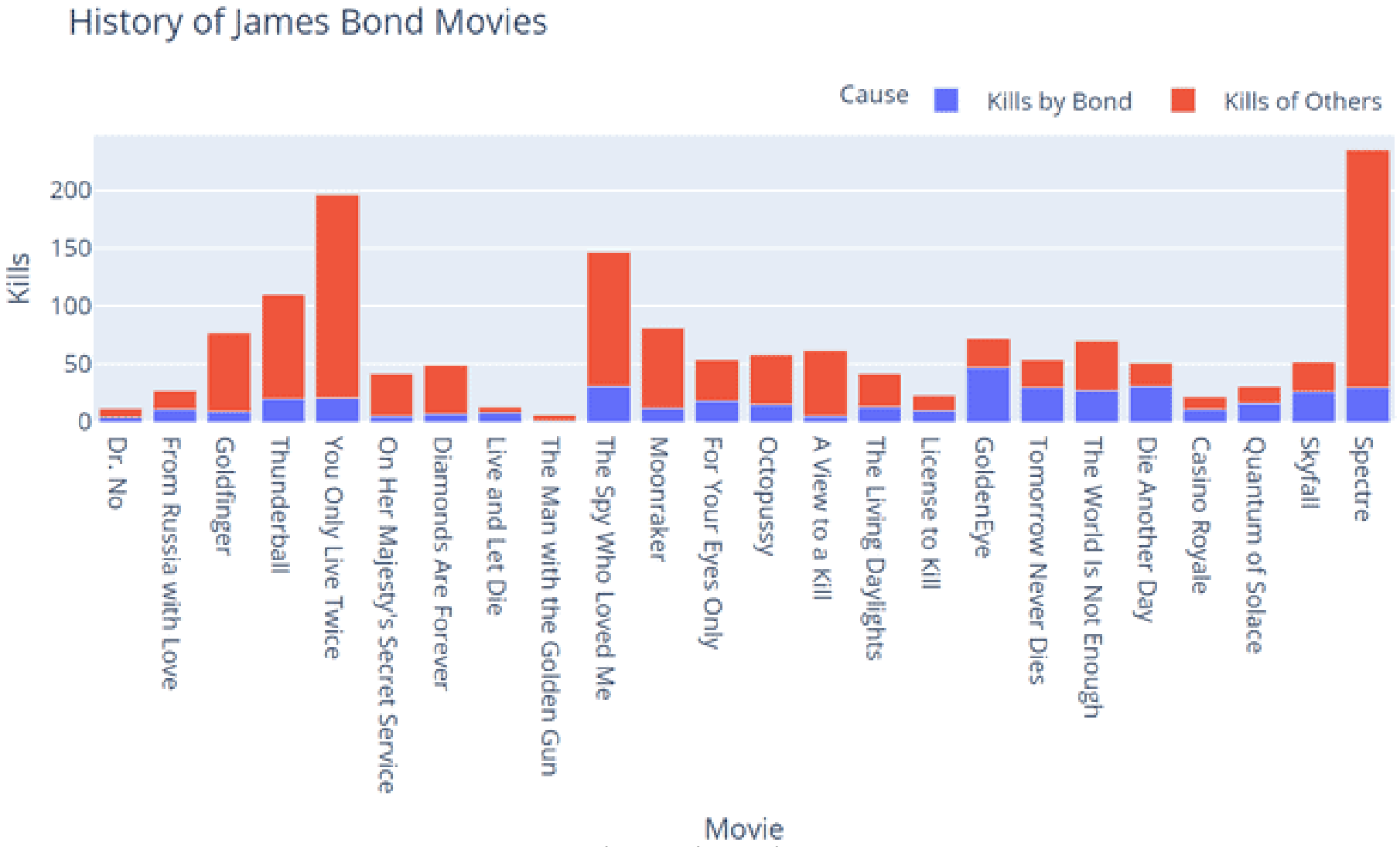
EXAMPLES


Lista de espera a finales de verano de los últimos 5 años
Año 2018
17,58‰
Año 2019
20,39‰
Año 2020
20,19‰
Año 2021
23,10‰
Año 2022
28,59‰
Fuente: Comunidad de Madrid
Más Madrid
VERDES EQUO

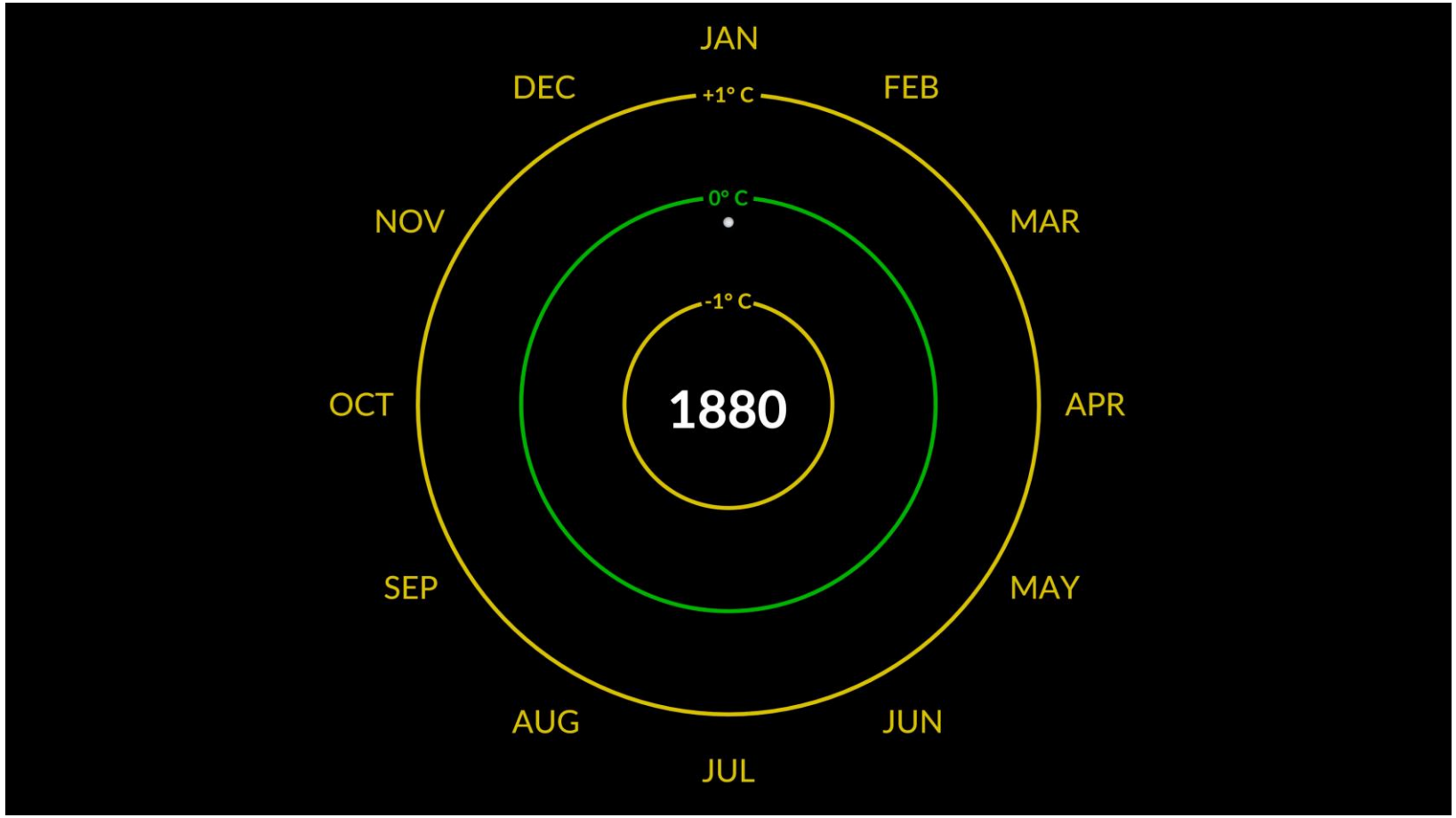
EXAMPLES



EXAMPLES

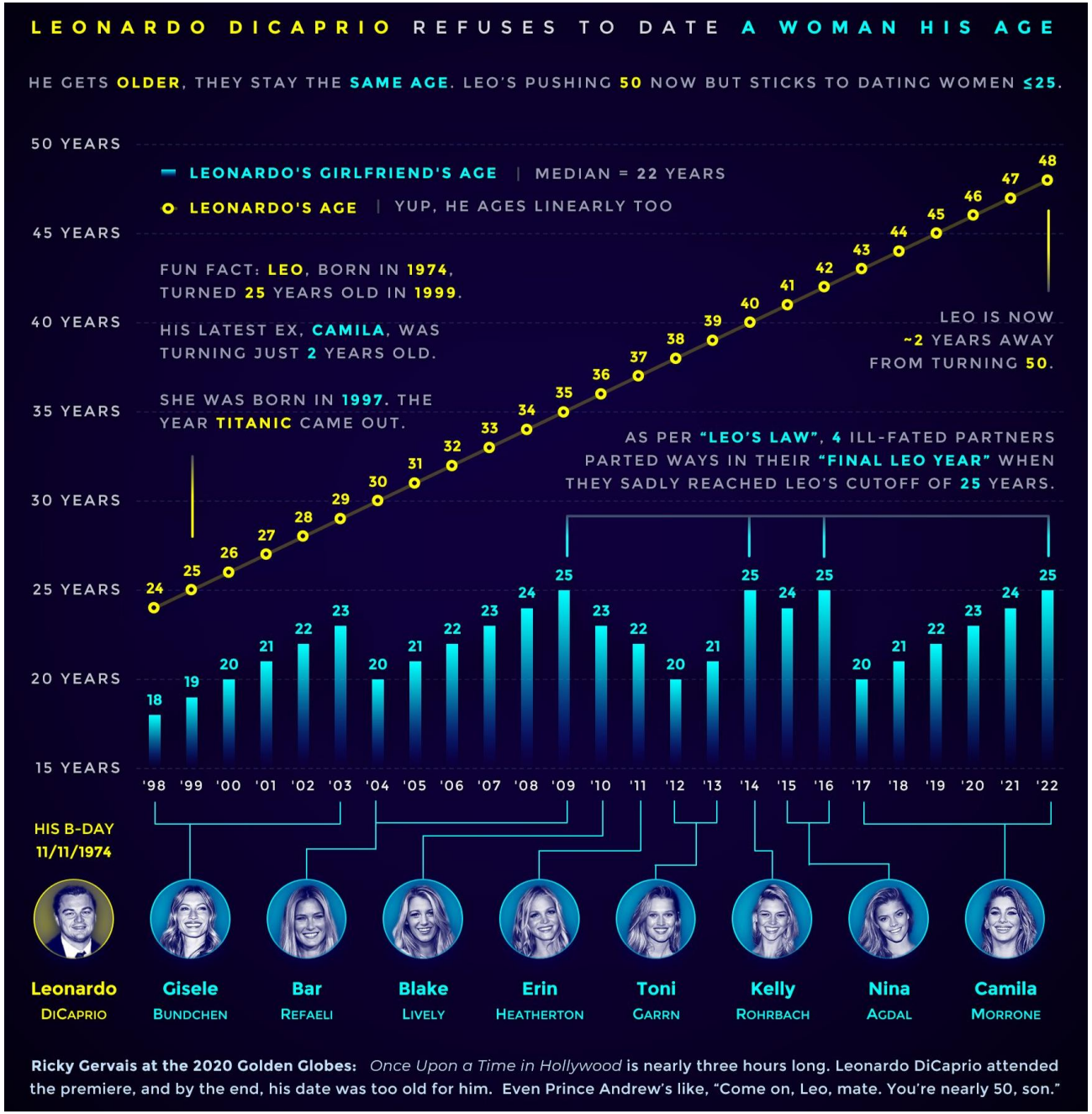

History of James Bond Movies
Cause
Kills by Bond
Kills of Others
Kills
200
150
100
50
0
Dr. No
From Russia with Love
Goldfinger
Thunderball
You Only Live Twice
On Her Majesty's Secret Service
Diamonds Are Forever
Live and Let Die
The Man with the Golden Gun
The Spy Who Loved Me
Moonraker
For Your Eyes Only
Octopussy
A View to a Kill
The Living Daylights
License to Kill
GoldenEye
Tomorrow Never Dies
The World Is Not Enough
Die Another Day
Casino Royale
Quantum of Solace
Skyfall
Spectre
Movie

EXAMPLES



https://climate.nasa.gov/climate_resources/300/video-climate-spiral/

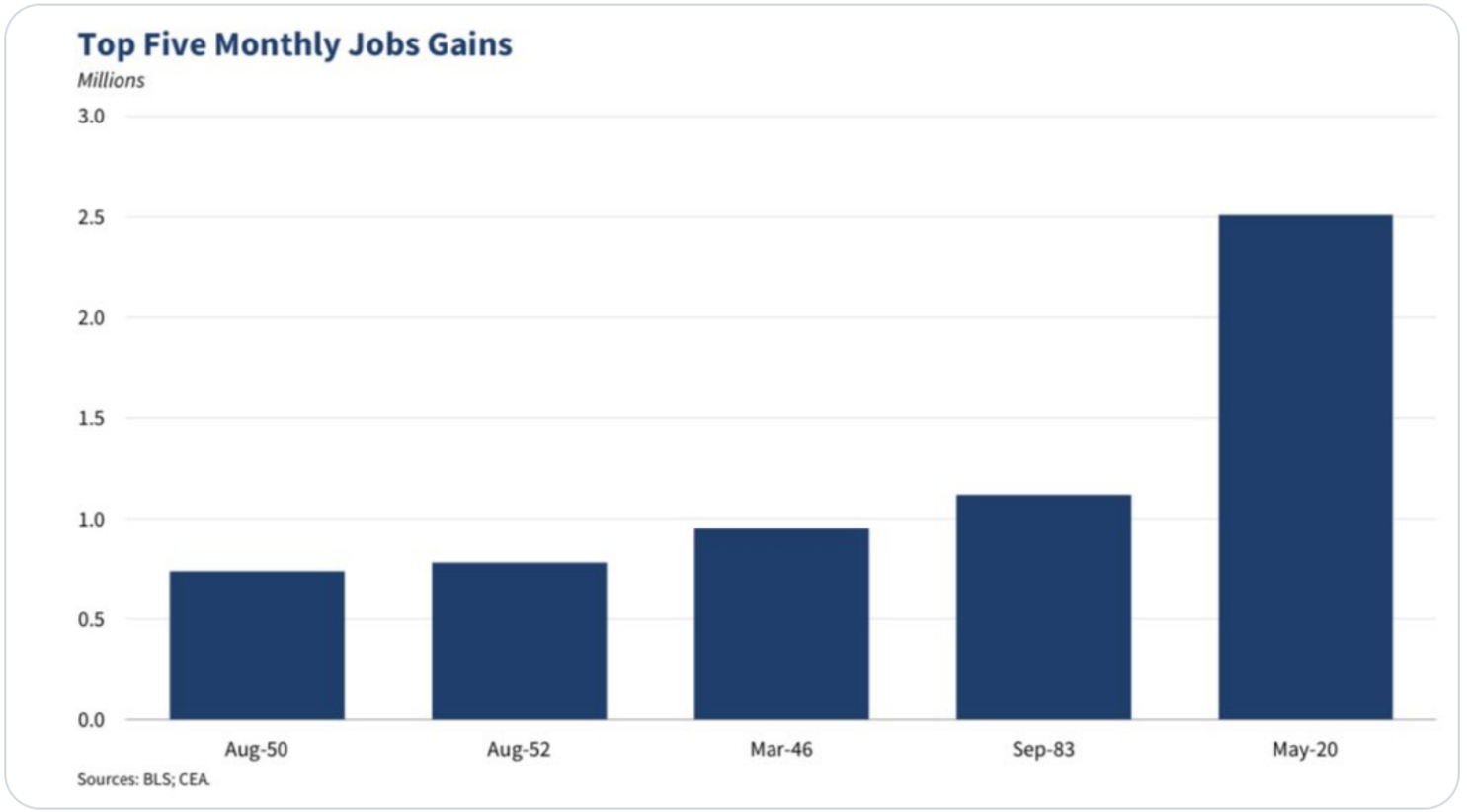
EXAMPLES



https://digg.com/data-viz/link/leonardo-di-caprio-dating-history-25-year-old-girlfriends-visualized-YjMeIPkdLJ

Donald J. Trump
@realDonaldTrump

Greatest Top Five Monthly Jobs Gains in HISTORY. We are #1!



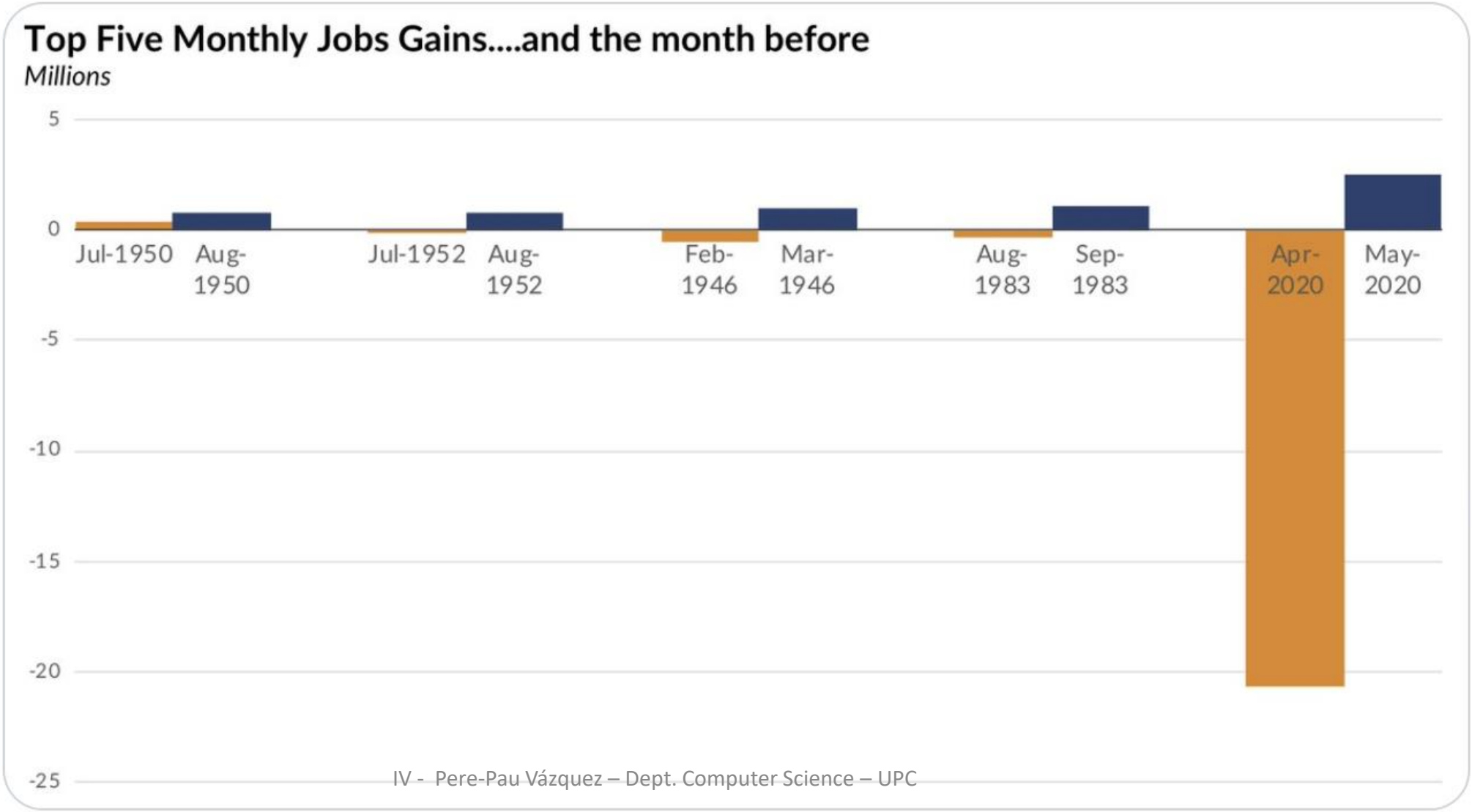
12:31 PM · Jun 5, 2020 · Twitter for iPhone

21.8K Retweets **72K** Likes

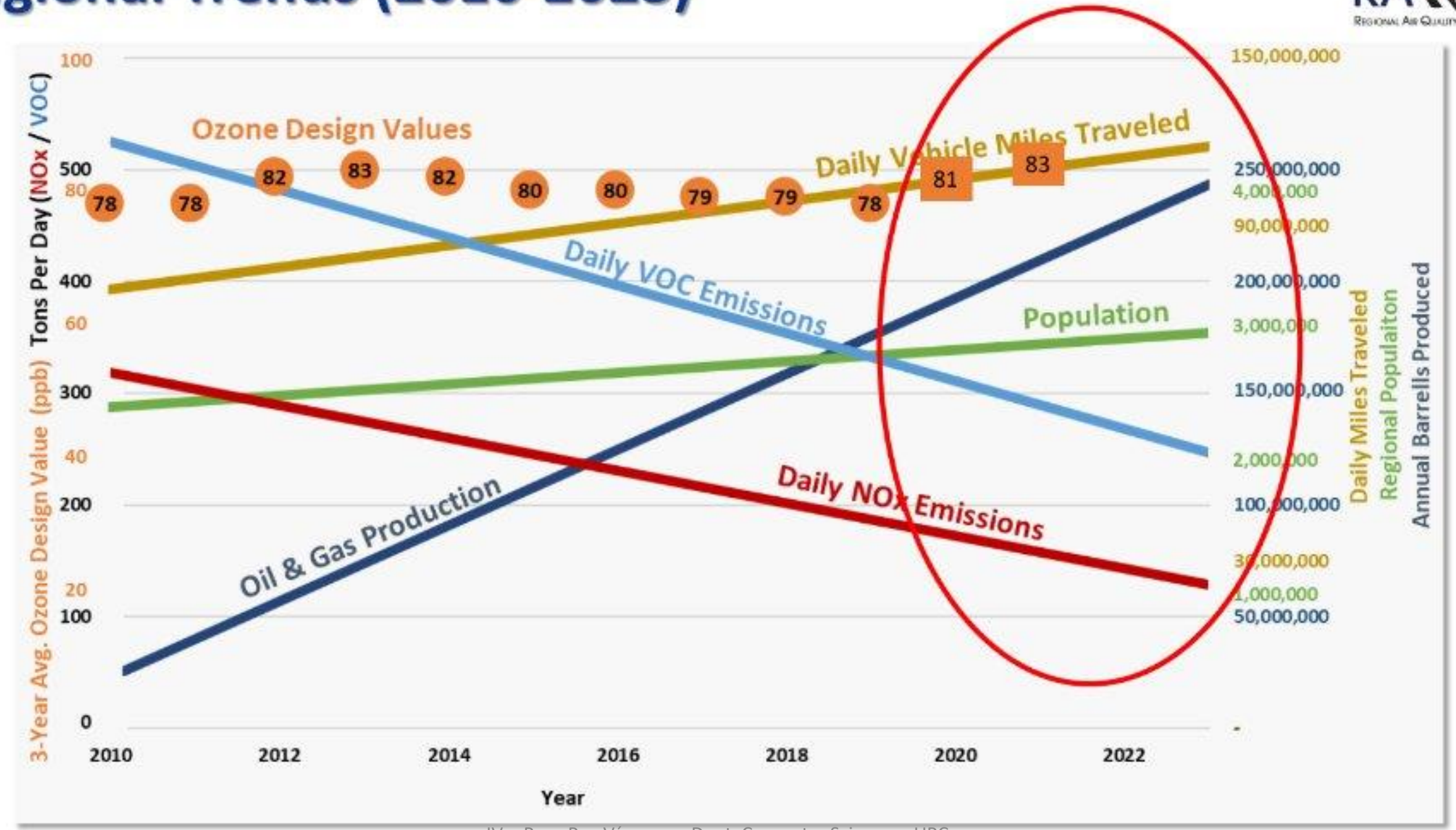
EXAMPLES

Jon Schwabish @jschwabish · 4h

I fixed this graph for you Mr. President. So, you know, it's not a complete misrepresentation of the facts..



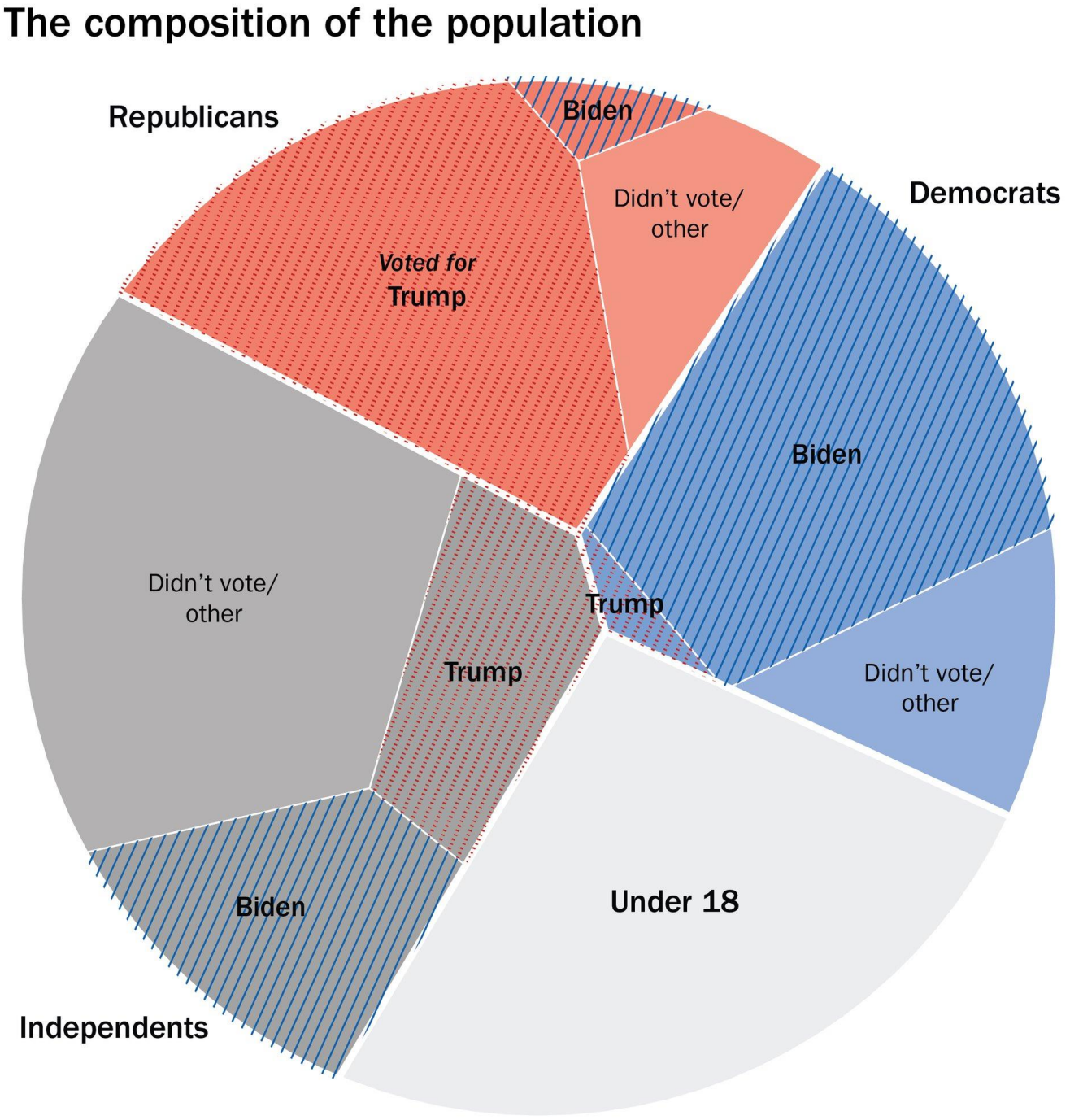
Regional Trends (2010-2023)



EXAMPLES

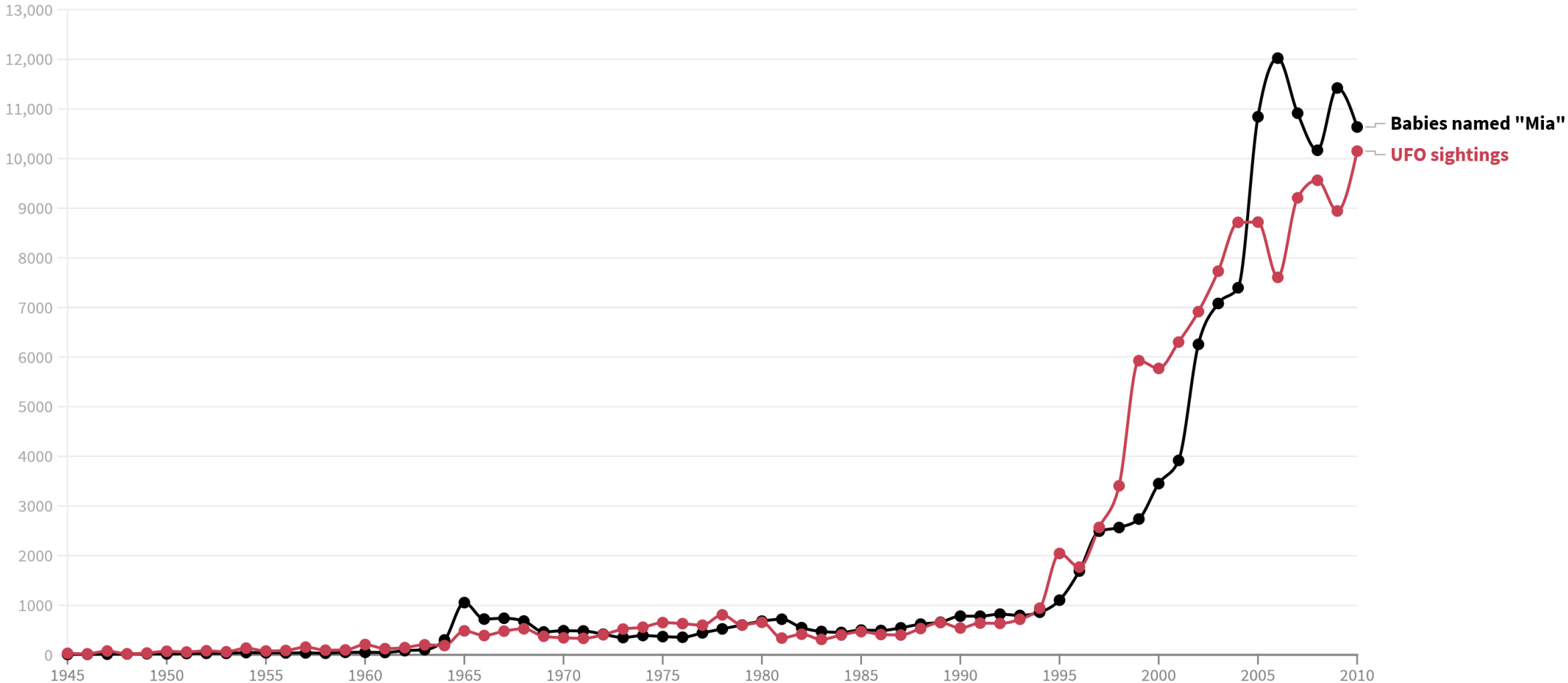

VOTO 2013
El pueblo decide
ELECCIÓN PRESIDENCIAL 2013
PORCENTAJES
50,66
% 49,07
Nicolás Maduro Moros
Henrique Capriles Rad

EXAMPLES


The composition of the population
Republicans
Biden
Didn't vote/ other
Democrats
Voted for Trump
Biden
Didn't vote/ other
Trump
Trump
Didn't vote/ other
Biden
Under 18
Independents
Sources: Census Bureau, Gallup, Pew Research Center
THE WASHINGTON POST

Correlation does not imply Causation

Number of **Babies named "Mia"** in the U.S. correlates with Number of **UFO Sightings** Around the World

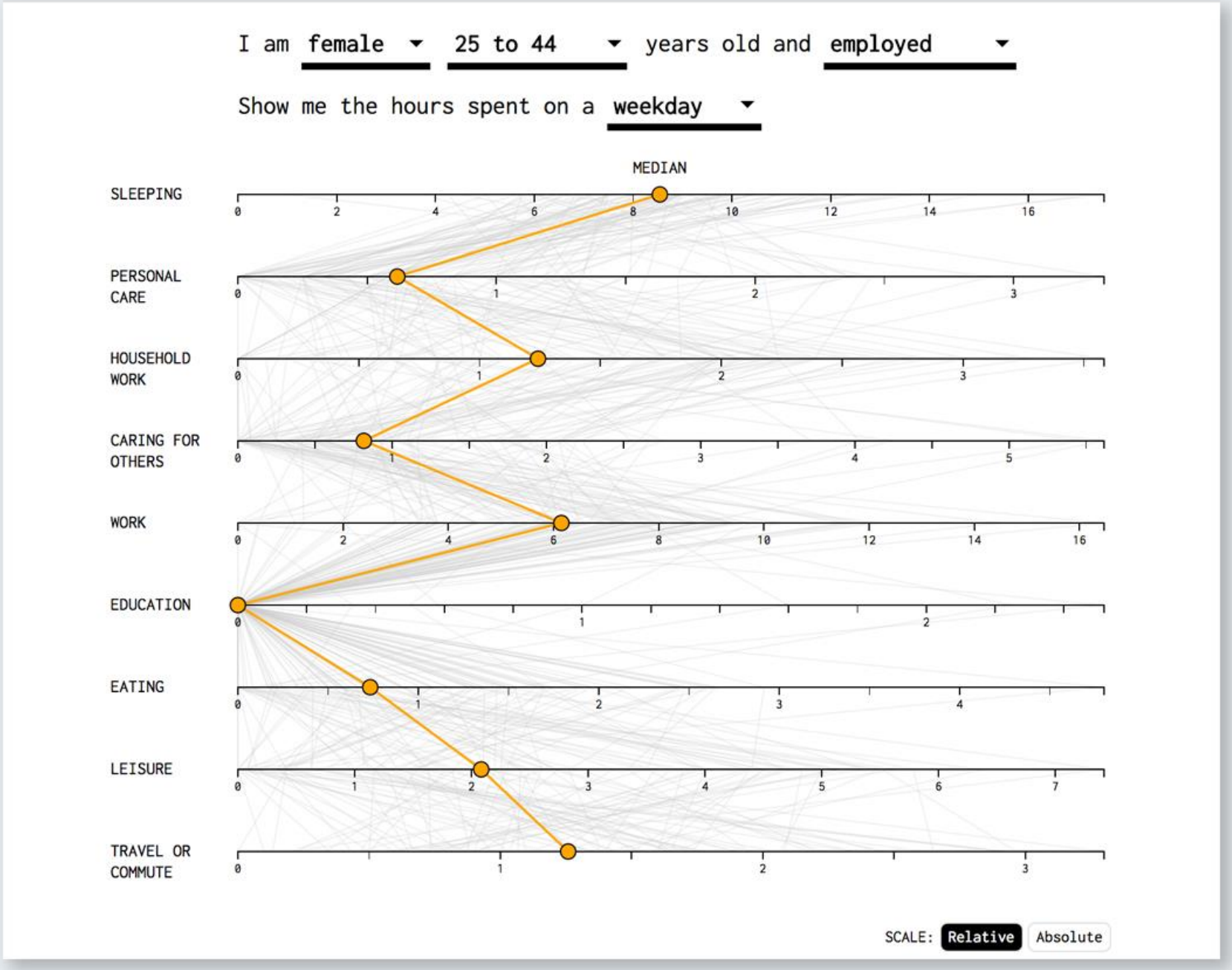


Data: NUFORC, SSA • Number of UFO sightings is shown by factor 5.2
Visualization by Cédric Scherer | #30DayChartChallenge 2021 | Day 13: Correlation

EXAMPLES

https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html

EXAM

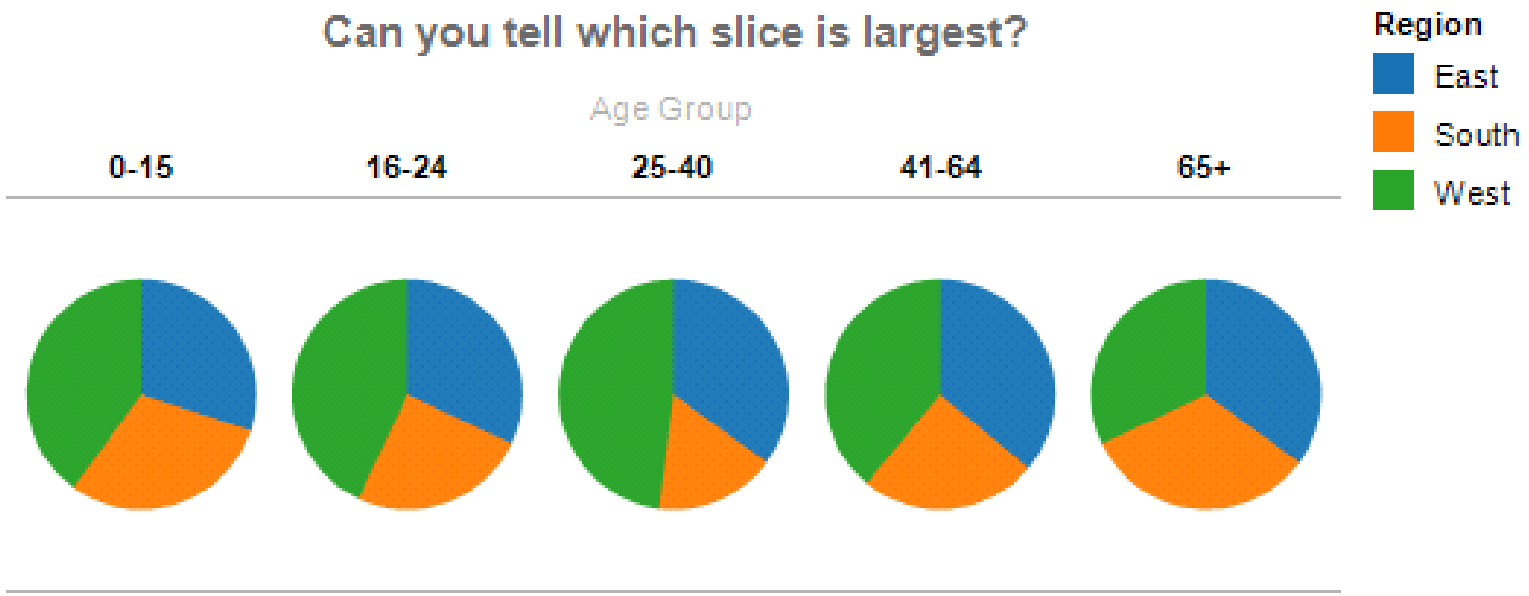


https://flowingdata.com/2016/12/06/how-people-like-you-spend-their-time

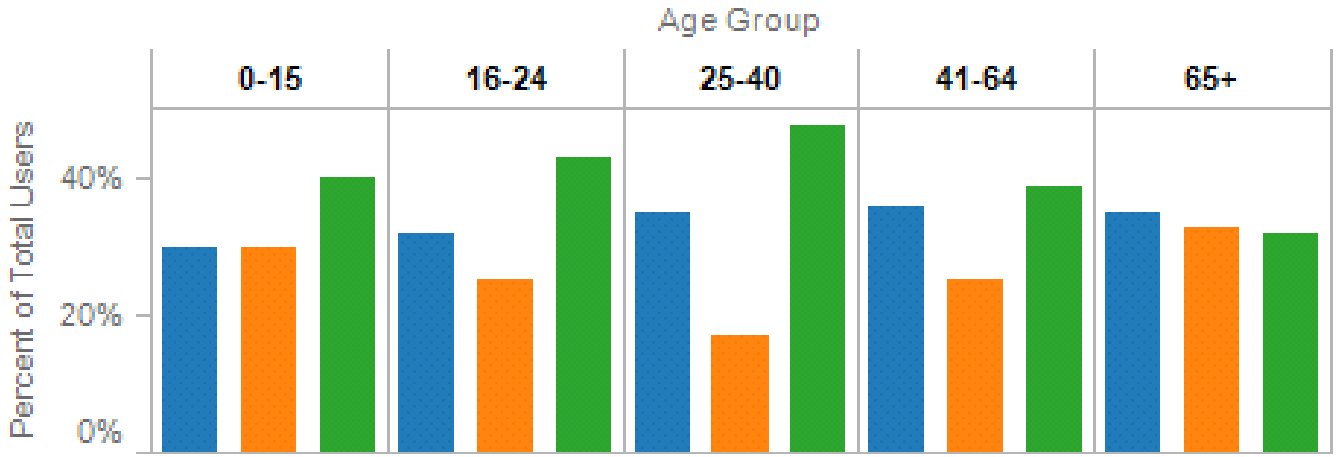
TAKEAWAYS

- Through the course, we will learn…

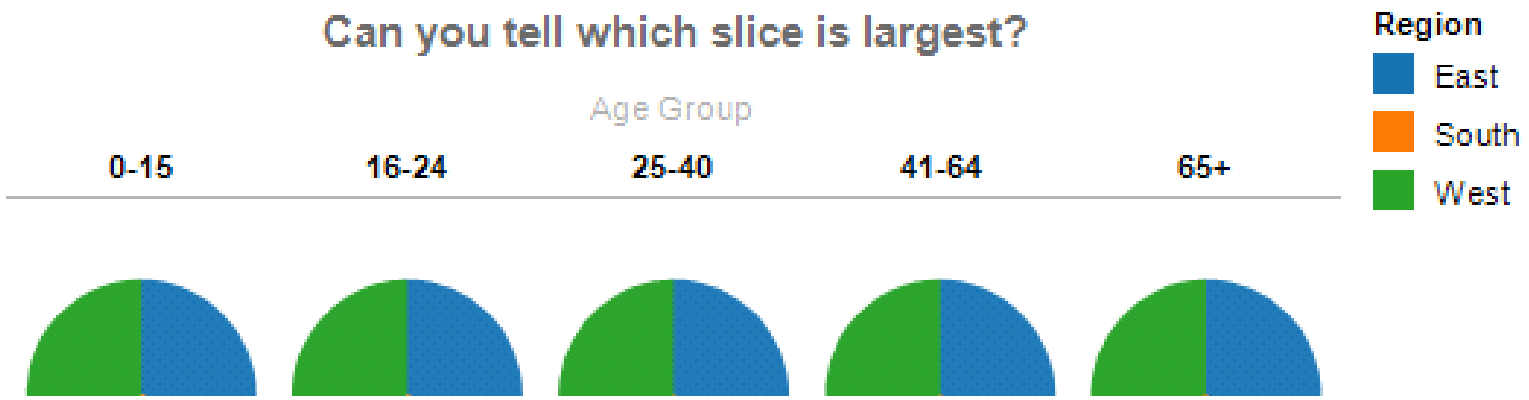
TAKEAWAYS



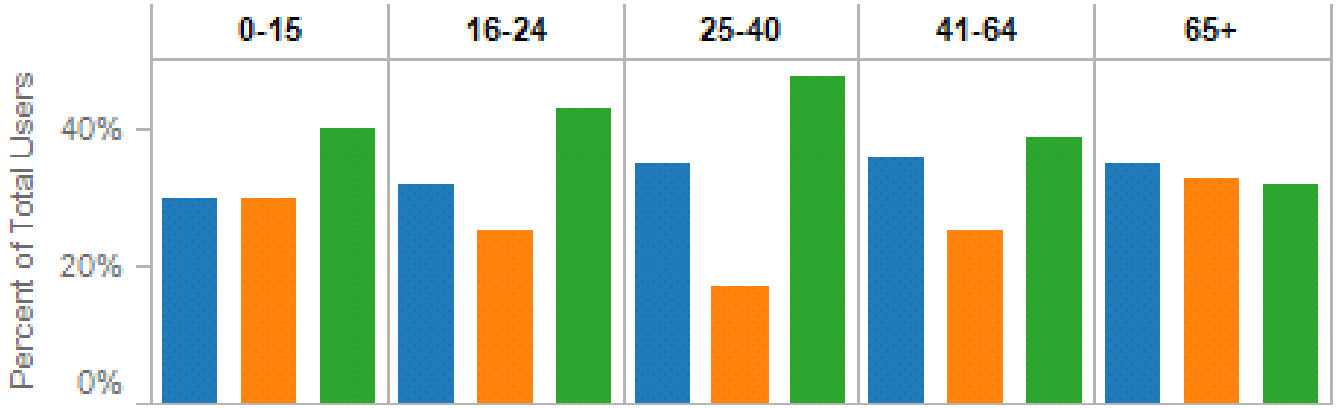
This is the same information represented as bars.
Comparisons are much easier.



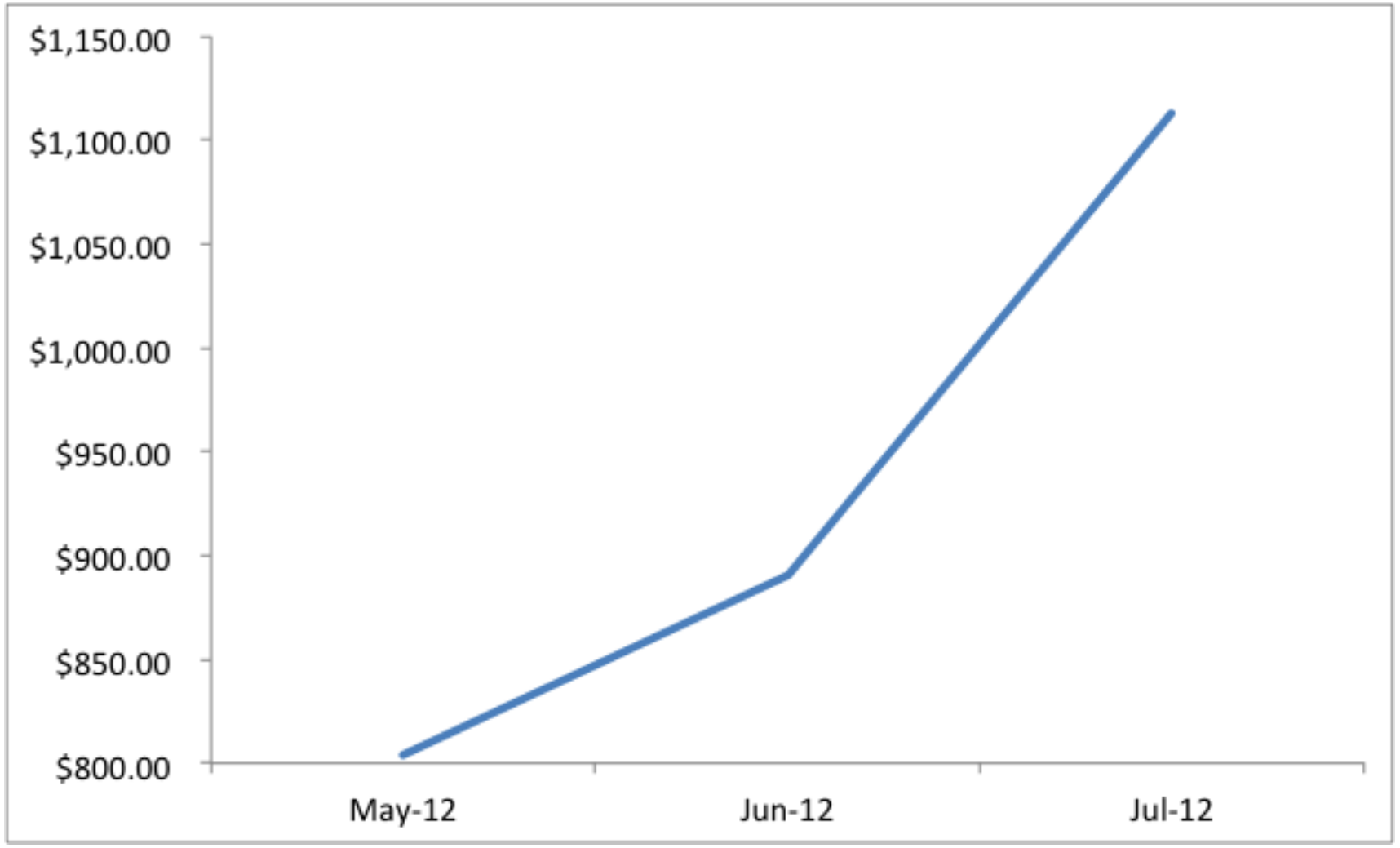
TAKEAWAYS



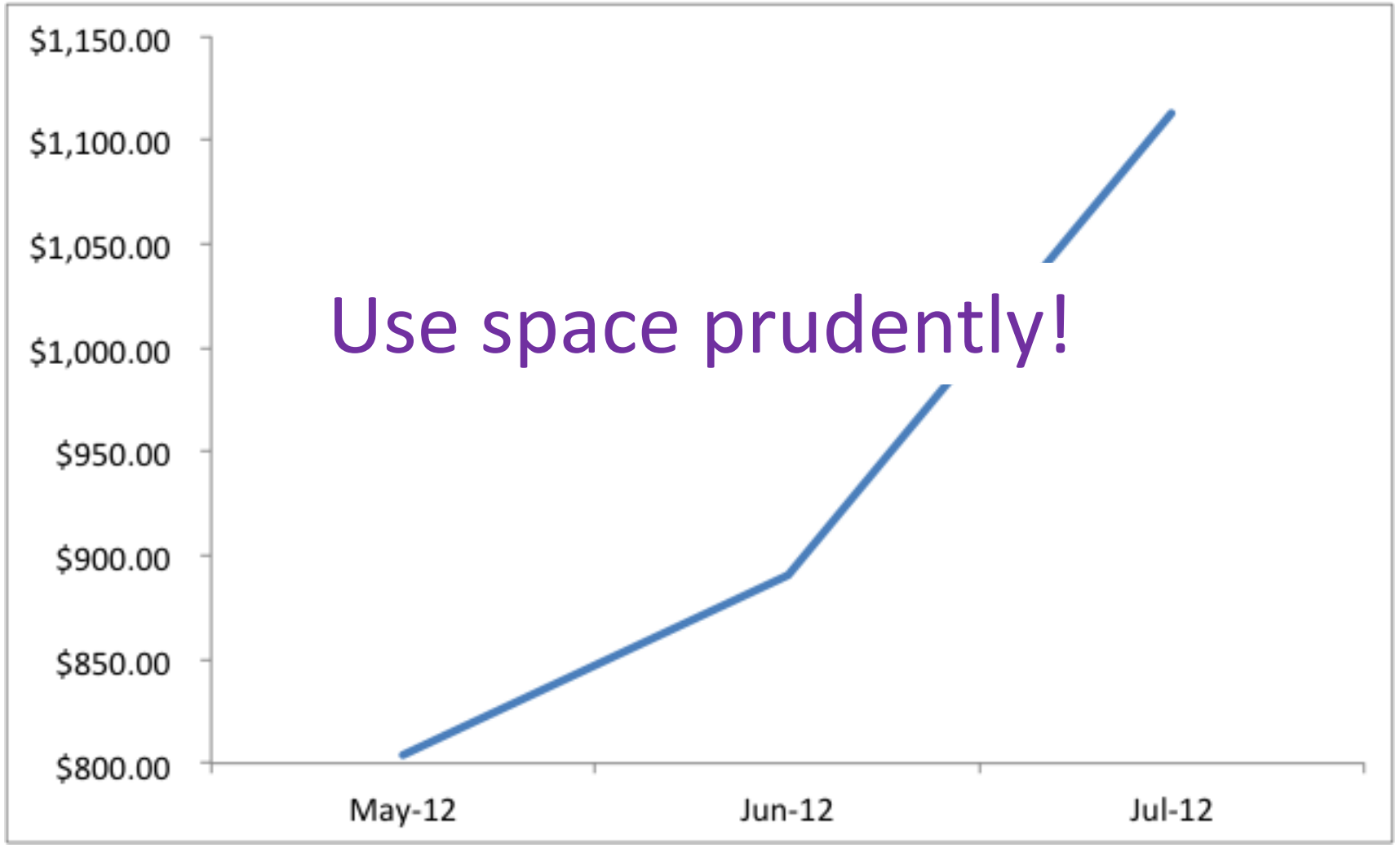
Not all visualization techniques are equally [or even properly] perceived!



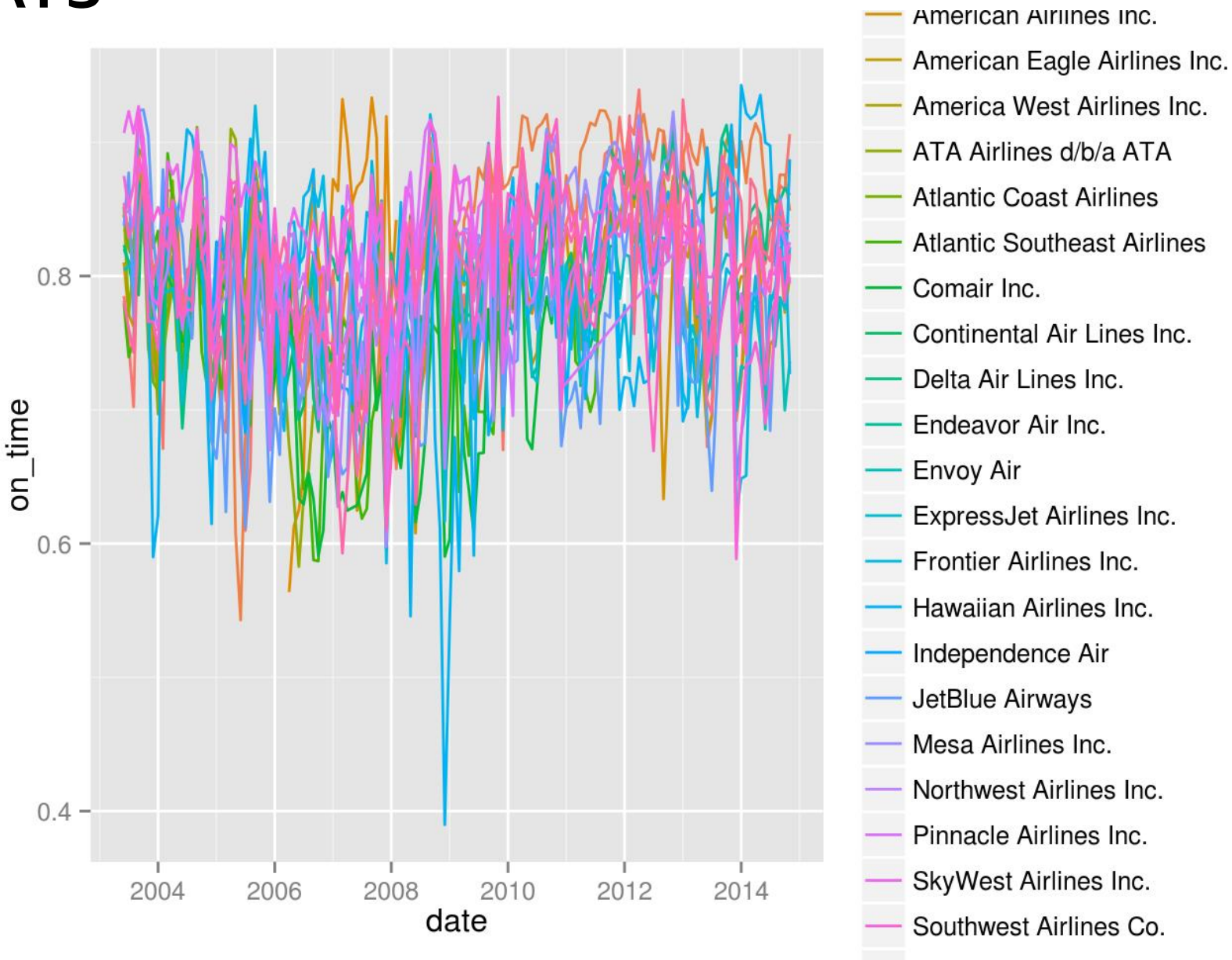
TAKEAWAYS


$1,150.00
$1,100.00
$1,050.00
$1,000.00
$950.00
$900.00
$850.00
$800.00
May-12
Jun-12
Jul-12

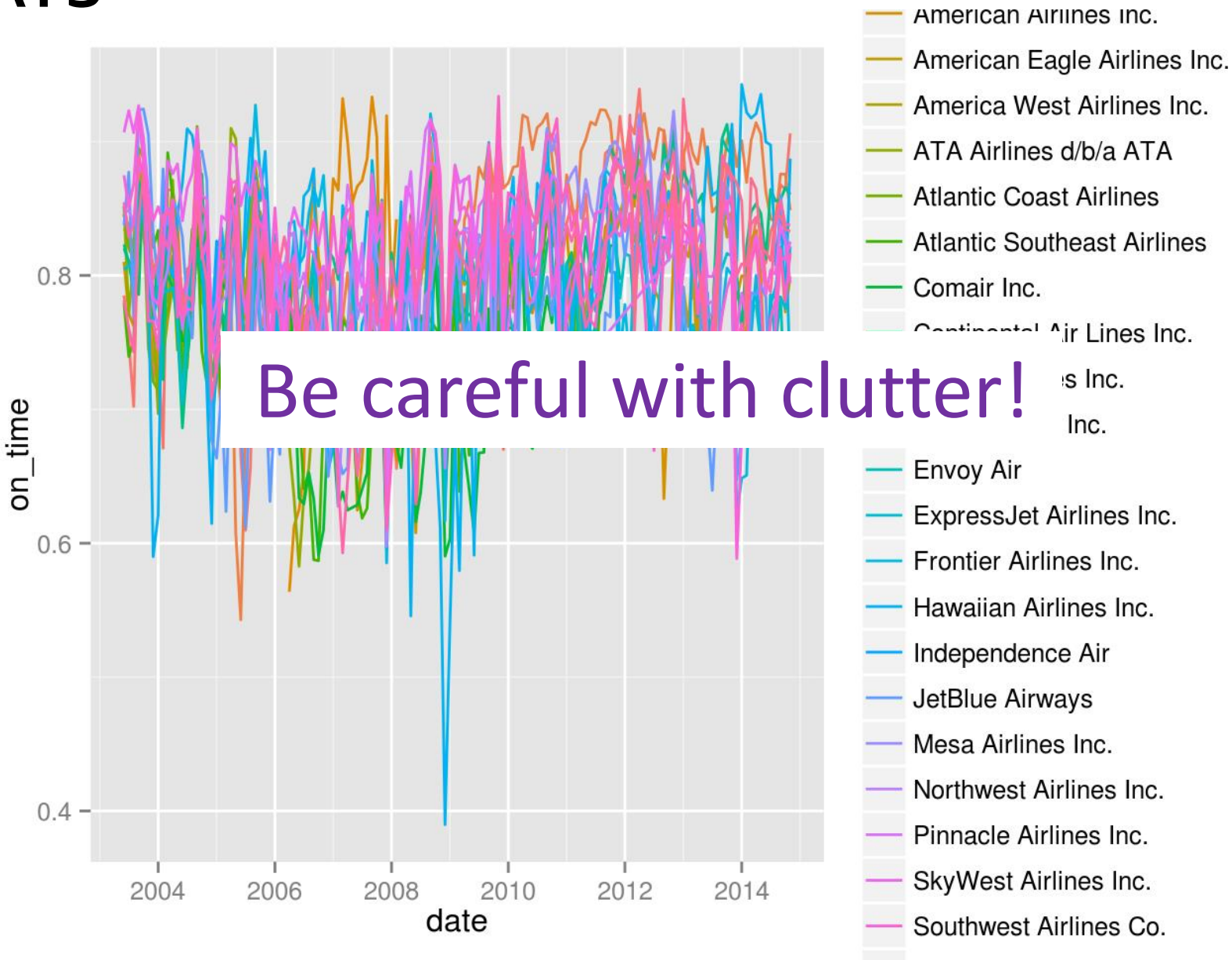
TAKEAWAYS


$1,150.00
$1,100.00
$1,050.00
$1,000.00
$950.00
$900.00
$850.00
$800.00
May-12
Jun-12
Jul-12
Use space prudently!

TAKEAWAYS

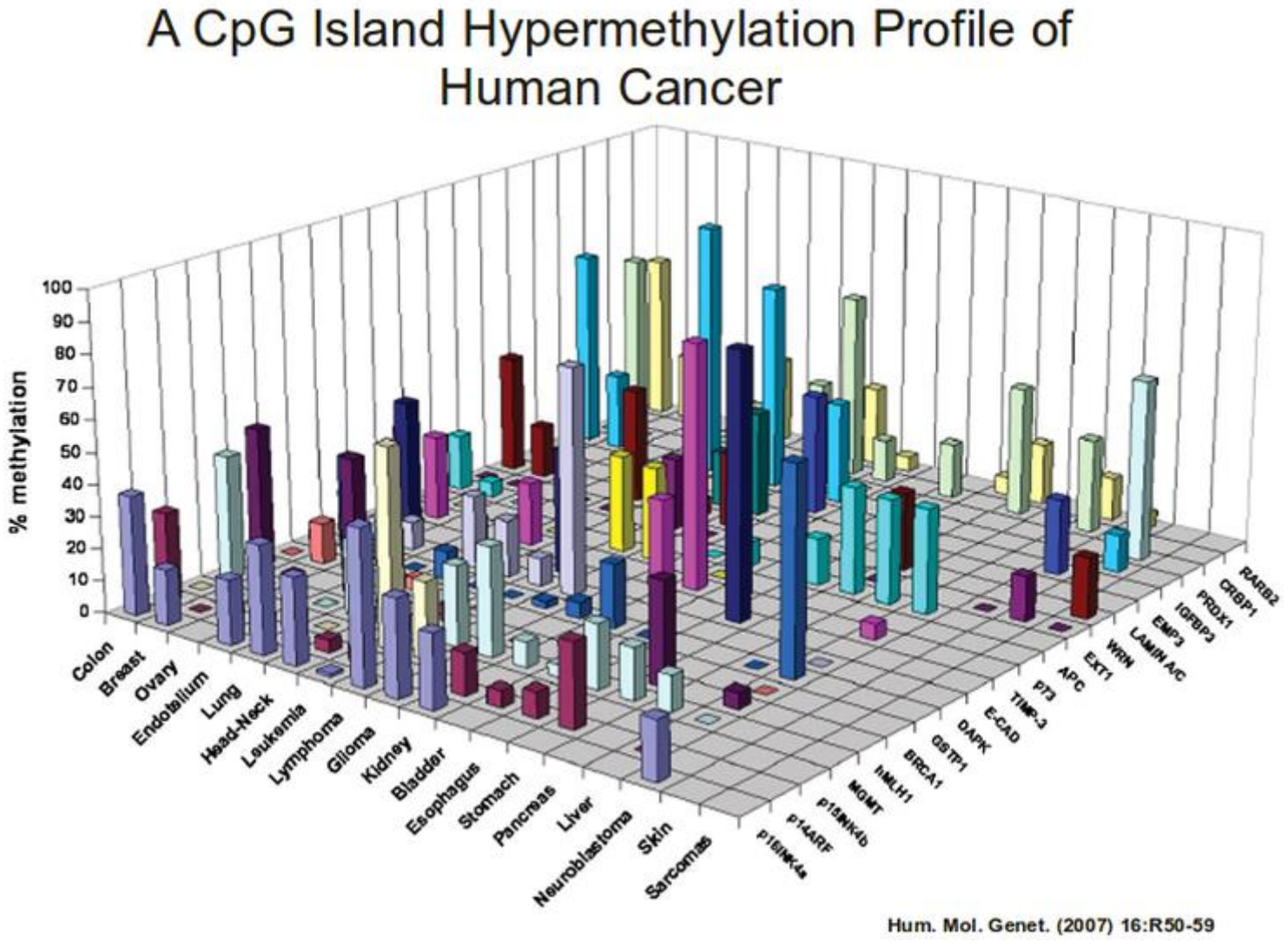

0.8
0.6
0.4
on_time
2004
2006
2008
2010
2012
2014
date
American Airlines Inc.
American Eagle Airlines Inc.
America West Airlines Inc.
ATA Airlines d/b/a ATA
Atlantic Coast Airlines
Atlantic Southeast Airlines
Comair Inc.
Continental Air Lines Inc.
Delta Air Lines Inc.
Endeavor Air Inc.
Envoy Air
ExpressJet Airlines Inc.
Frontier Airlines Inc.
Hawaiian Airlines Inc.
Independence Air
JetBlue Airways
Mesa Airlines Inc.
Northwest Airlines Inc.
Pinnacle Airlines Inc.
SkyWest Airlines Inc.
Southwest Airlines Co.

TAKEAWAYS

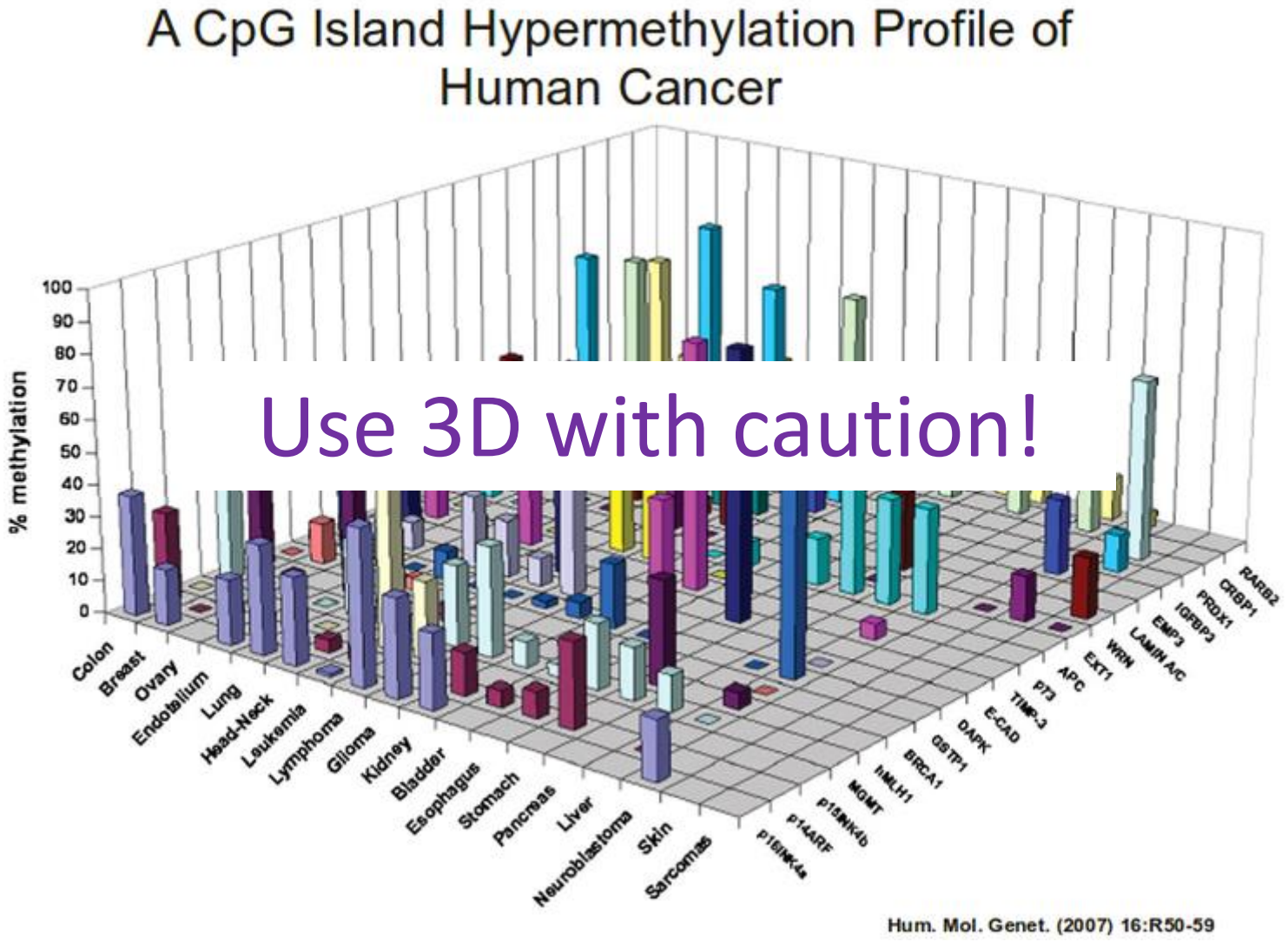


Be careful with clutter!

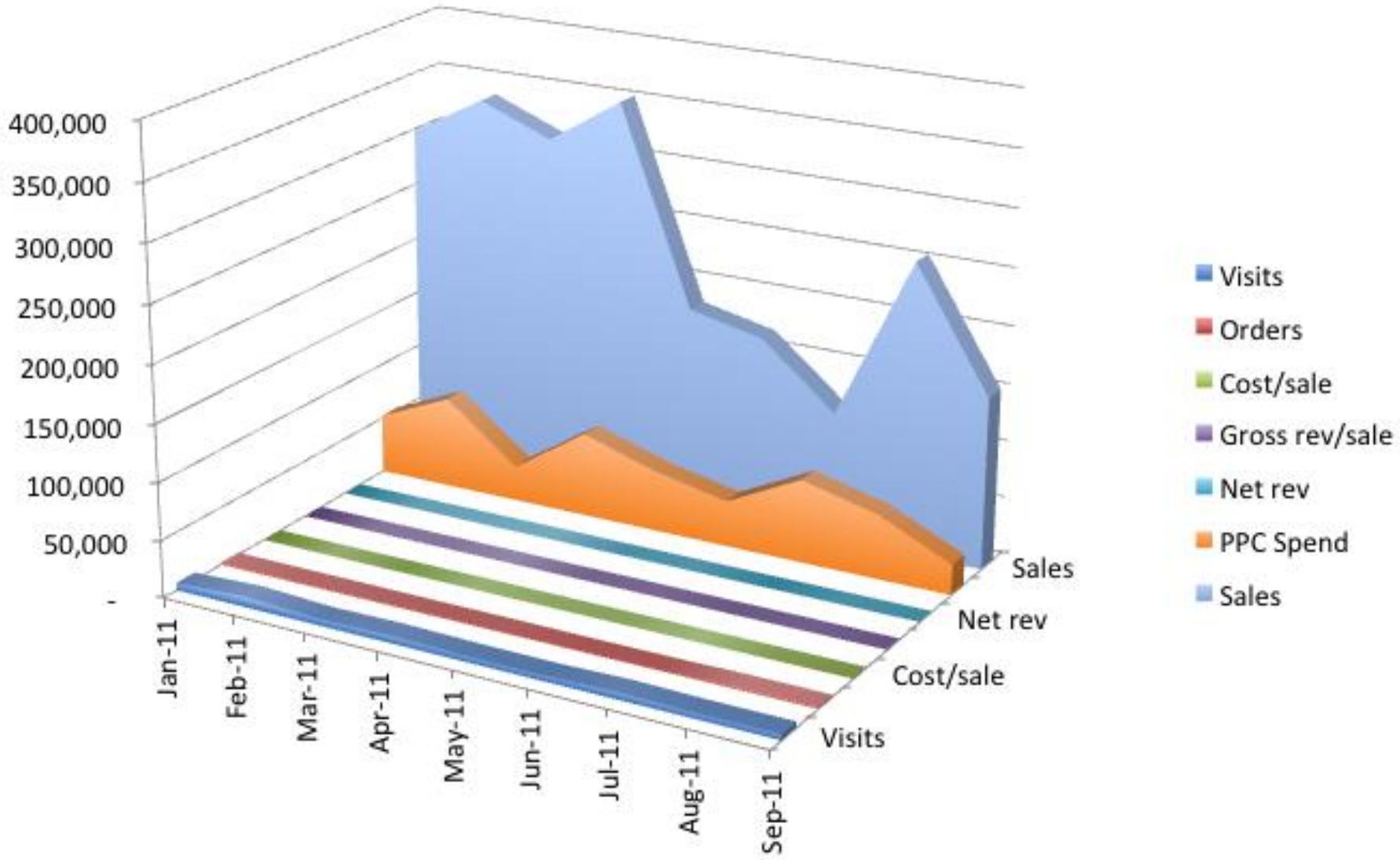
TAKEAWAYS


A CpG Island Hypermethylation Profile of Human Cancer
% methylation
100
90
80
70
60
50
40
30
20
10
0
Colon
Breast
Ovary
Endothelium
Lung
Head-Neck
Leukemia
Lymphoma
Glioma
Kidney
Bladder
Esophagus
Stomach
Pancreas
Liver
Neuroblastoma
Skin
Sarcomas
p16INK4a
p14ARF
p15INK4b
MGMT
hMLH1
BRCA1
GSTP1
DAPK
E-CAD
TIMP-3
PT3
APC
EXT1
WRN
LAMIN A/C
IGFBP3
PRDX1
CRBP1
RARB2
Hum. Mol. Genet. (2007) 16:R50-59

TAKEAWAYS


A CpG Island Hypermethylation Profile of Human Cancer
% methylation
100
90
80
70
60
50
40
30
20
10
0
Colon
Breast
Ovary
Endothelium
Lung
Head-Neck
Leukemia
Lymphoma
Glioma
Kidney
Bladder
Esophagus
Stomach
Pancreas
Liver
Neuroblastoma
Skin
Sarcomas
p16INK4a
p14ARF
p15INK4b
MGMT
hMLH1
BRCA1
GSTP1
DAPK
E-CAD
TIMP-3
PT3
APC
EXT1
WRN
LAMIN A/C
EMP3
IGFBP3
PRDX1
CRBP1
RARB2
Use 3D with caution!
Hum. Mol. Genet. (2007) 16:R50-59

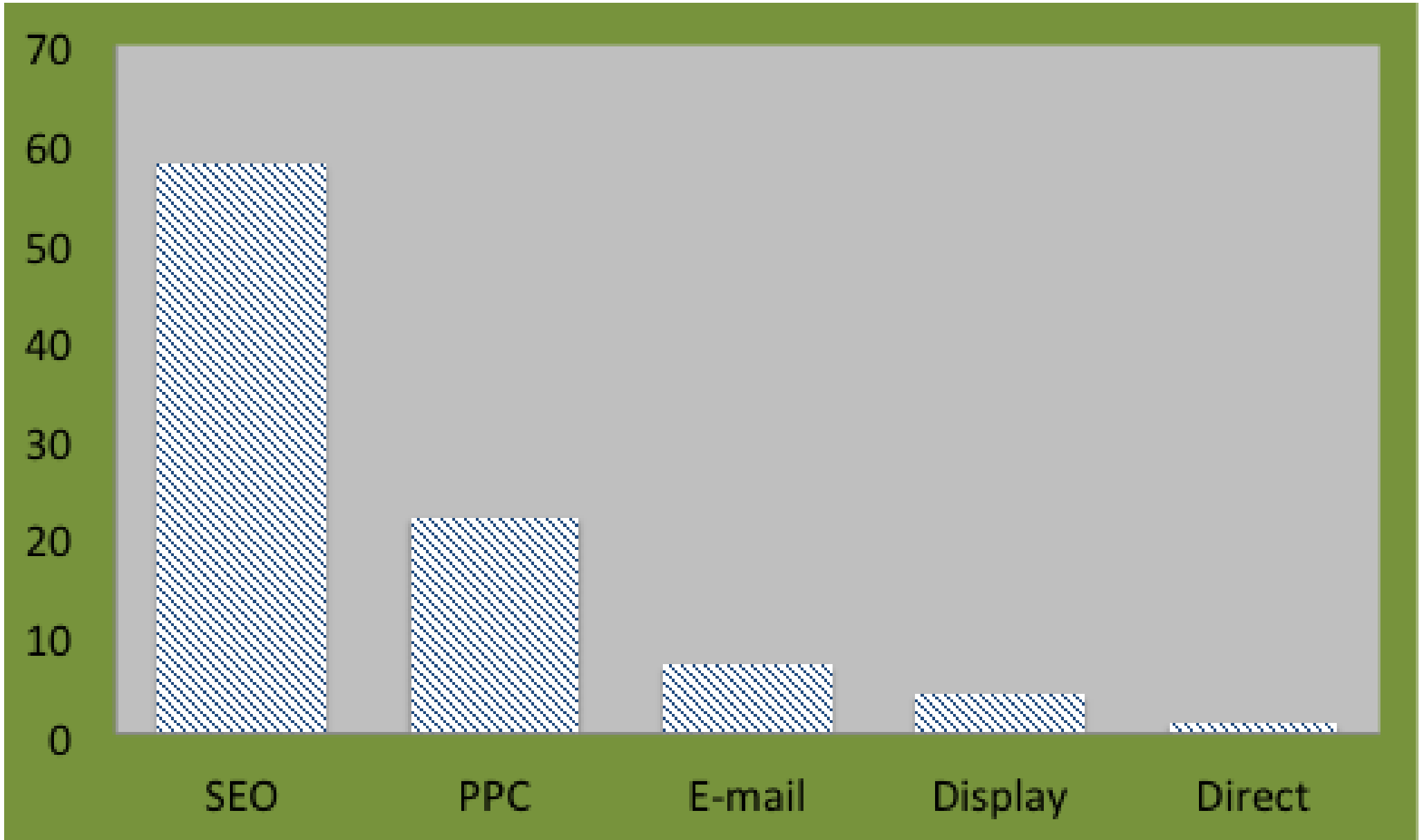
TAKEAWAYS


400,000
350,000
300,000
250,000
200,000
150,000
100,000
50,000
-
Jan-11
Feb-11
Mar-11
Apr-11
May-11
Jun-11
Jul-11
Aug-11
Sep-11
Sales
Net rev
Cost/sale
Visits
Visits
Orders
Cost/sale
Gross rev/sale
Net rev
PPC Spend
Sales

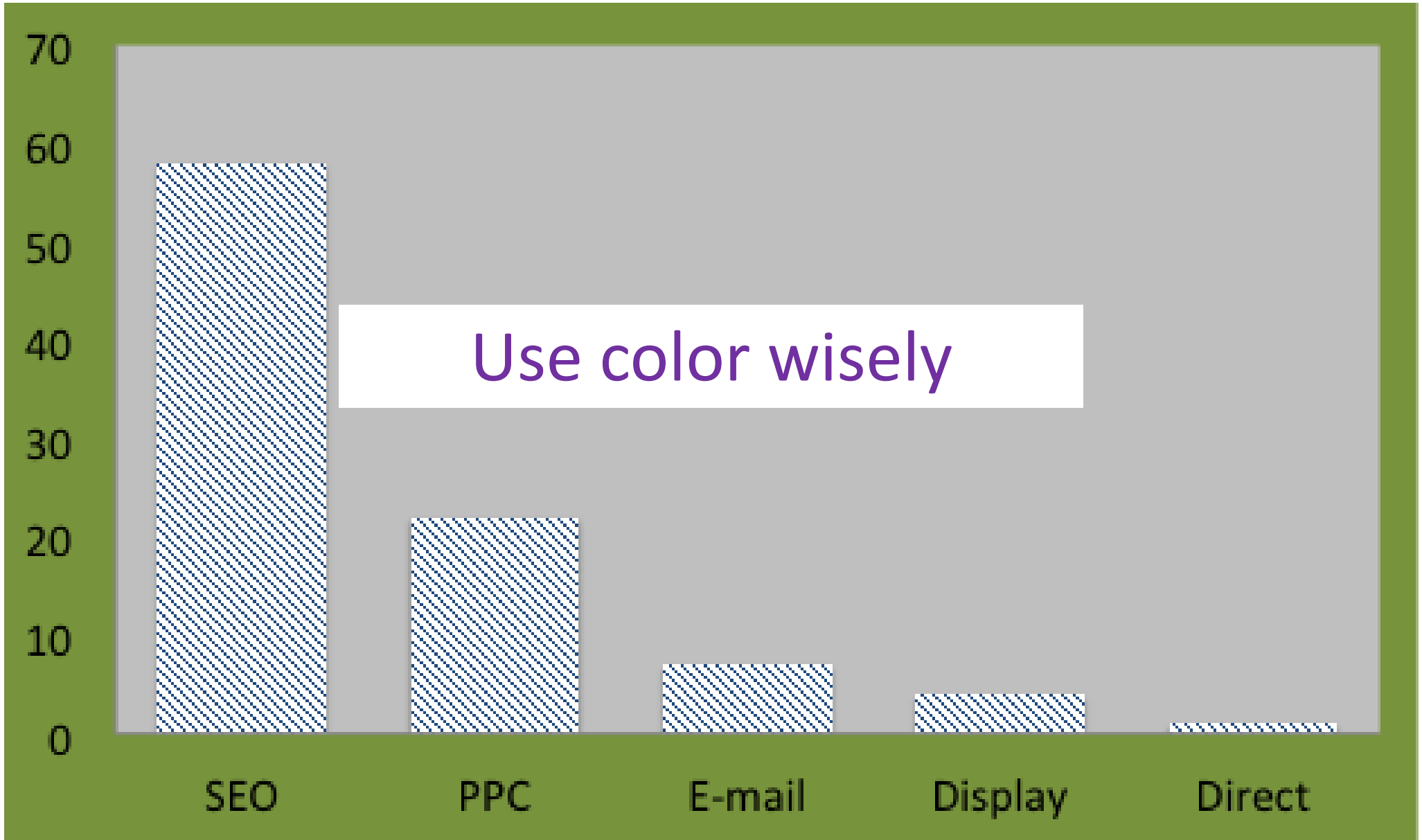
TAKEAWAYS

Be careful with scales!

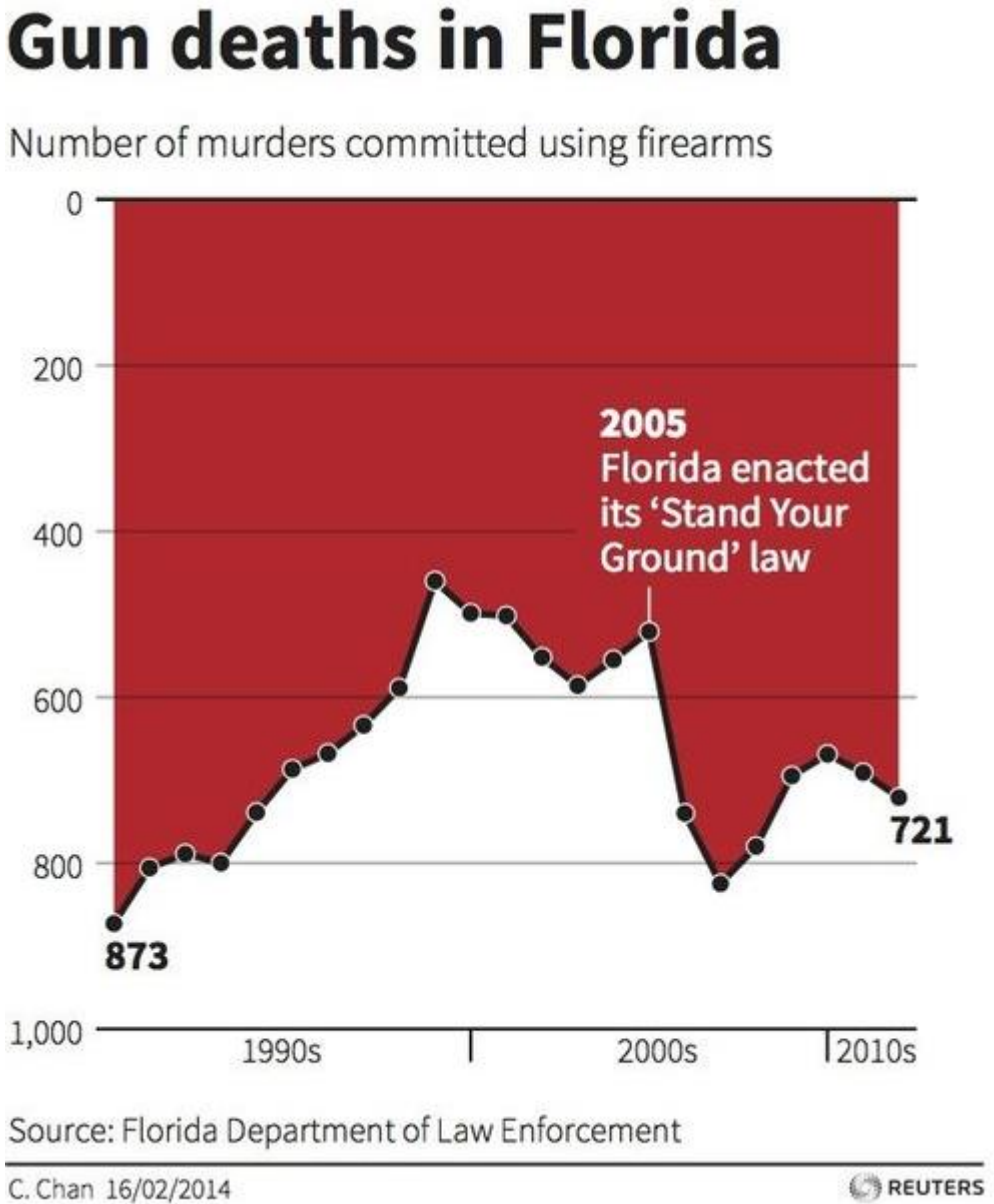
TAKEAWAYS



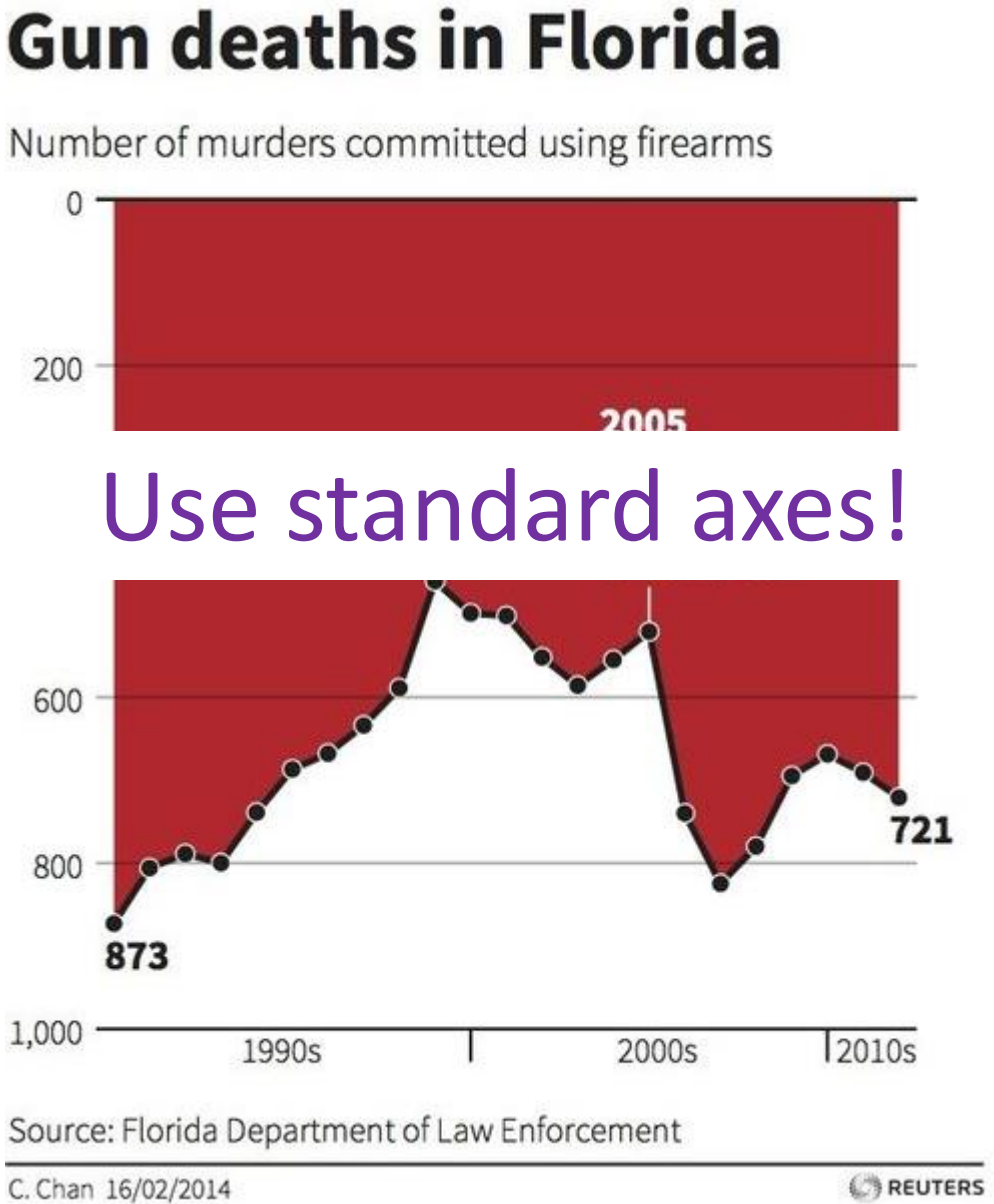
TAKEAWAYS


70
60
50
40
30
20
10
0
Use color wisely
SEO
PPC
E-mail
Display
Direct

TAKEAWAYS


Gun deaths in Florida
Number of murders committed using firearms
0
200
400
600
800
1,000
1990s
2000s
2010s
2005
Florida enacted its 'Stand Your Ground' law
873
721
Source: Florida Department of Law Enforcement
C. Chan 16/02/2014
REUTERS

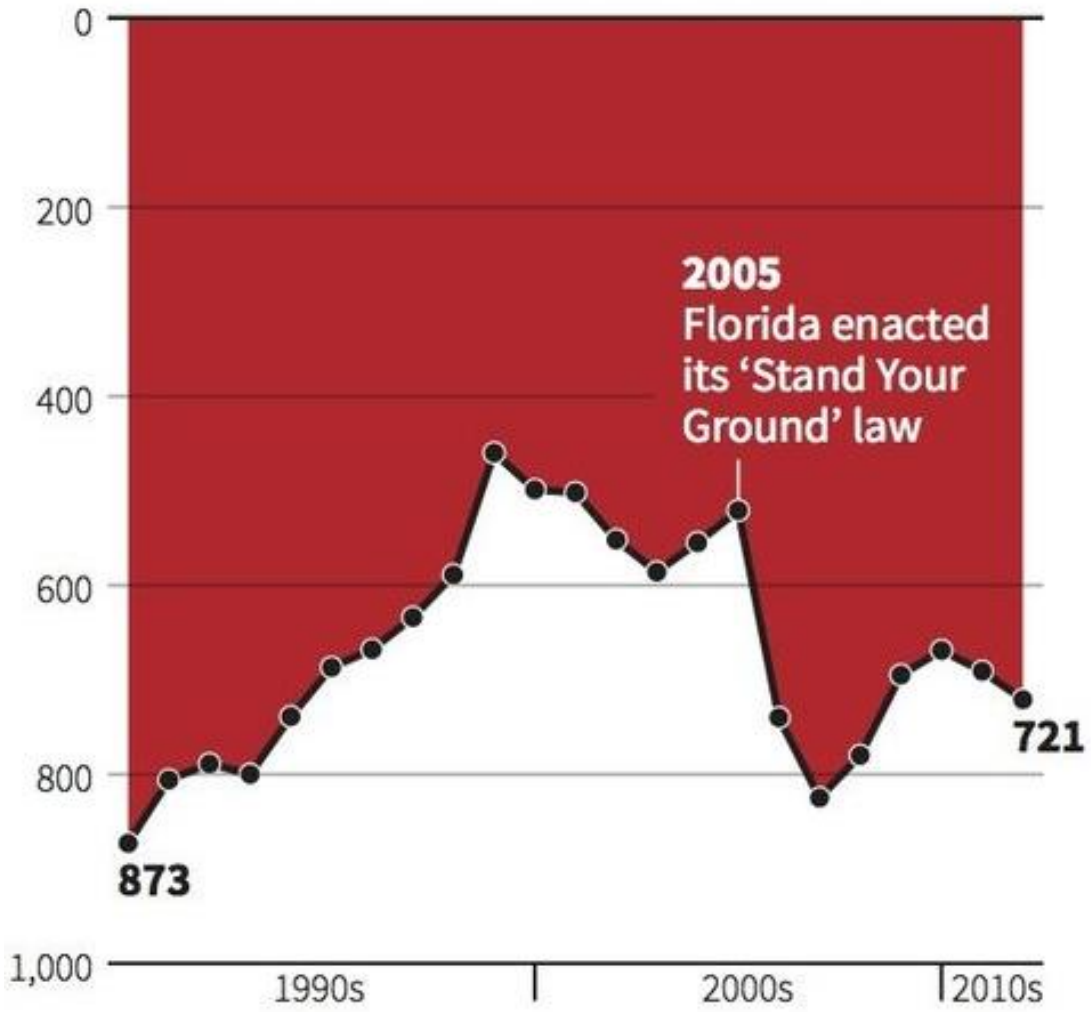
TAKEAWAYS



Use standard axes!

Gun deaths in Florida

Number of murders committed using firearms

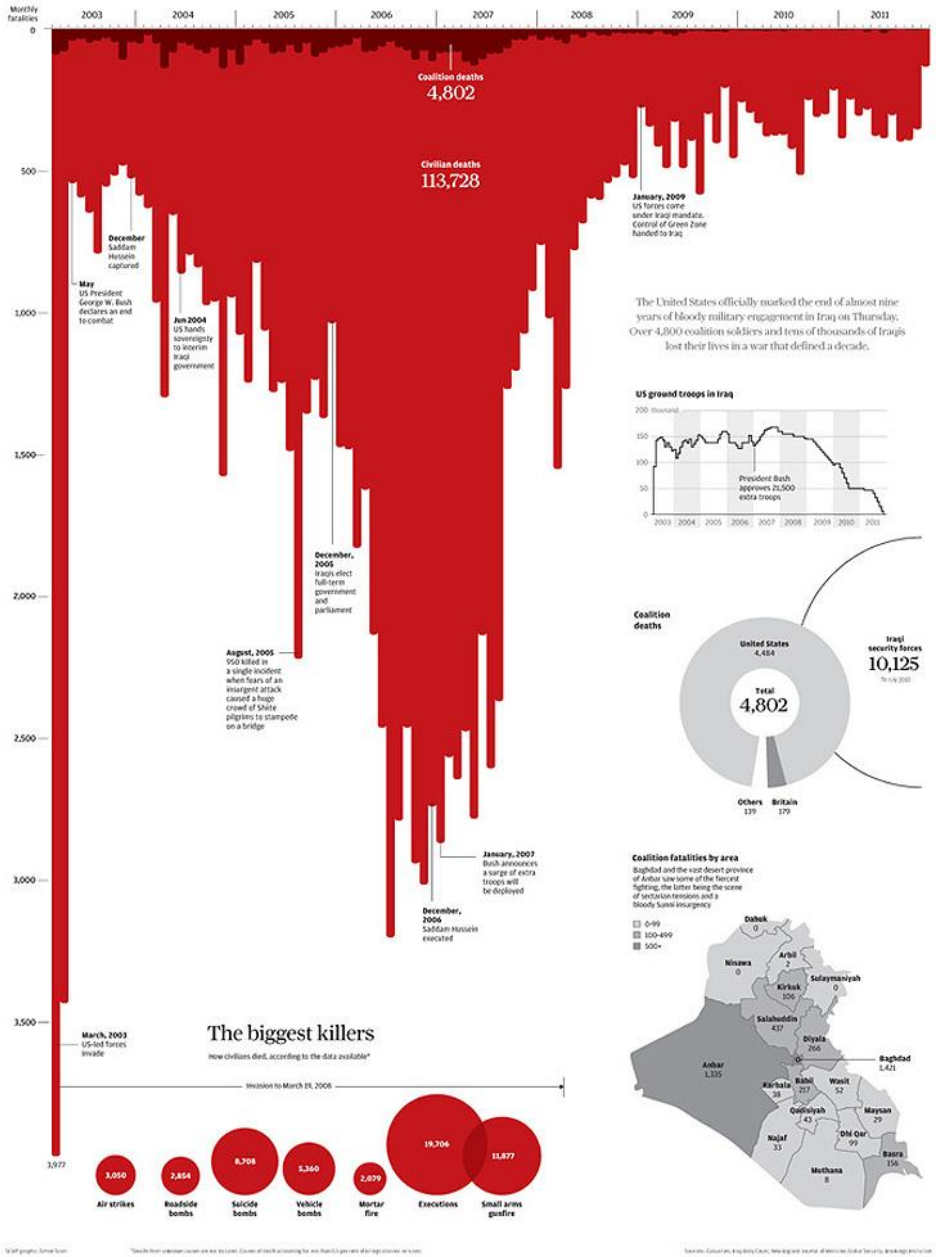


Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

Iraq's bloody toll




South China Morning Post
News
Comment
Lifestyle
Knowledge
More +
SUBSCRIBE
Infographics / Infographics
Iraq's bloody toll
Iraq's bloody toll
Monthly fatalities
0
500
2003
2004
2005
2006
2007
2008
2009
2010
2011
Coalition deaths
4,802
Civilian deaths
113,728
December
Saddam Hussein captured
May
US President George W. Bush
January, 2009
US forces come under Iraqi mandate. Control of Green Zone handed to Iraq
The United States officially marked the end of almost nine

OUTLINE

- Visualization. The basics
- History
- General Rules
- Data, Tasks, Users
- Visualization as a Design Process
- The Visualization Mantra

OUTLINE

- **Visualization. The basics**
- History
- General Rules
- Data, Tasks, Users
- Visualization as a Design Process
- The Visualization Mantra


BRACE YOURSELVES
DATA VISUALIZATION IS COMING
memegenerator.net

VISUALIZATION. THE BASICS

- T. Munzner's definition: "Computer-based visualization systems provide **visual** representations **of datasets** designed to **help people** carry out tasks more **effectively**"
  - Visual representations → human visual system
  - Datasets
  - People
  - Tasks
  - Effective
- Augment the capabilities of the human rather than simply replace them by some decision making method that is computational

VISUALIZATION. THE BASICS

- Visualization is related to understanding the underlying data
  - Cognitive process
  - Helping the user (human) to understand data using their excellent perception capabilities
    - Augmenting human capabilities (instead of replacing)
  - Helping the user to carry out tasks more effectively

- Putting **human in the loop** is fundamental

VISUALIZATION. THE BASICS

- Many applications
  - Exploratory data analysis
  - Presentation of results
  - Sometimes helping the developers build an automatic model
  - Understand the situation
  - Refine an algorithm
  - Build trust among users

VISUALIZATION. THE BASICS

- If the result is a calculation, then you should probably not be using visualization at all
  - If you know what you are looking for, you probably do not need the visualization

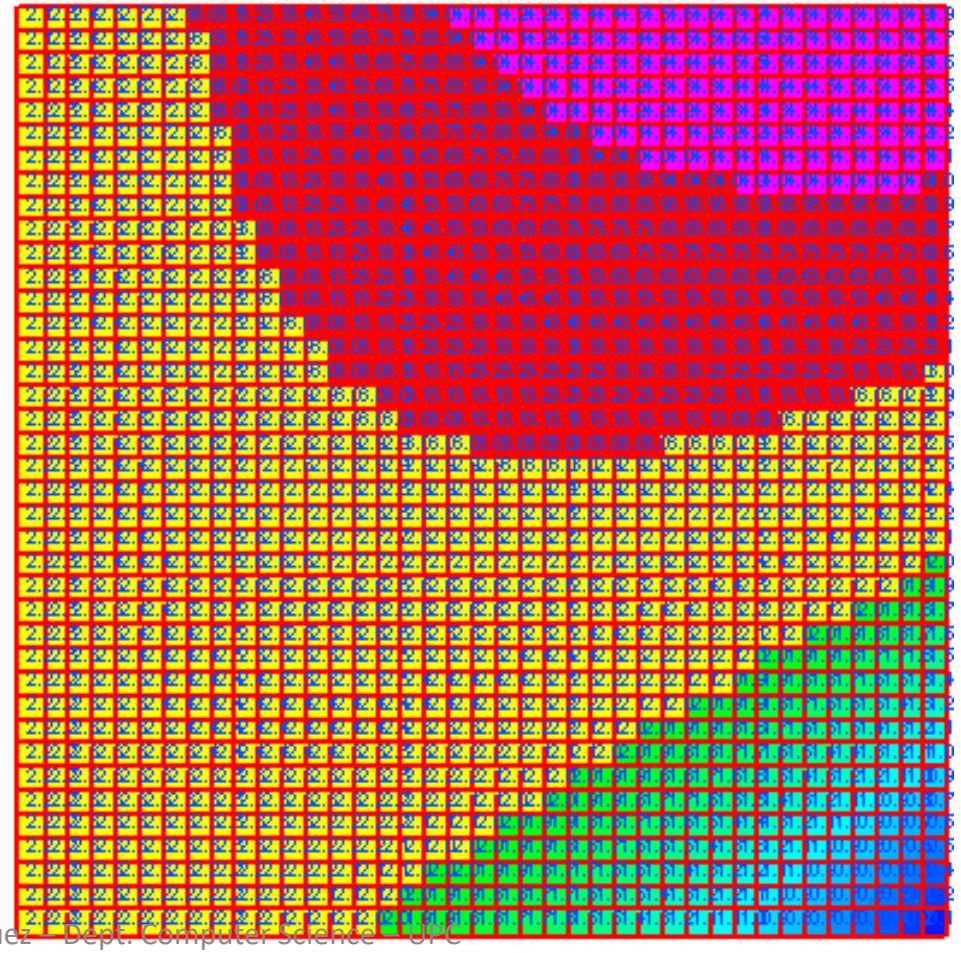
VISUALIZATION. THE BASICS

- **Cognitive process:** We want to understand this:

1.3	1.8	2.2	2.6	3.0	3.3	3.6	3.9	4.1	4.3	4.4	4.5
1.3	1.8	2.2	2.5	2.9	3.2	3.4	3.6	3.8	4.0	4.1	4.1
1.4	1.8	2.1	2.5	2.7	3.0	3.2	3.4	3.5	3.7	3.7	3.8
1.4	1.7	2.1	2.4	2.6	2.8	3.0	3.2	3.3	3.3	3.4	3.4
1.4	1.7	2.0	2.3	2.5	2.7	2.8	2.9	3.0	3.0	3.0	3.0
1.4	1.7	2.0	2.2	2.4	2.5	2.6	2.7	2.7	2.7	2.7	2.6
1.5	1.7	1.9	2.1	2.3	2.4	2.4	2.5	2.5	2.4	2.3	2.2
1.5	1.7	1.9	2.0	2.1	2.2	2.2	2.2	2.2	2.1	2.0	1.9
1.5	1.7	1.8	1.9	2.0	2.1	2.0	2.0	1.9	1.8	1.7	1.5
1.5	1.7	1.8	1.9	1.9	1.9	1.9	1.8	1.7	1.5	1.3	1.1
1.6	1.7	1.7	1.8	1.8	1.7	1.7	1.5	1.4	1.2	1.0	0.7
1.6	1.7	1.7	1.7	1.7	1.6	1.5	1.3	1.1	0.9	0.6	0.3

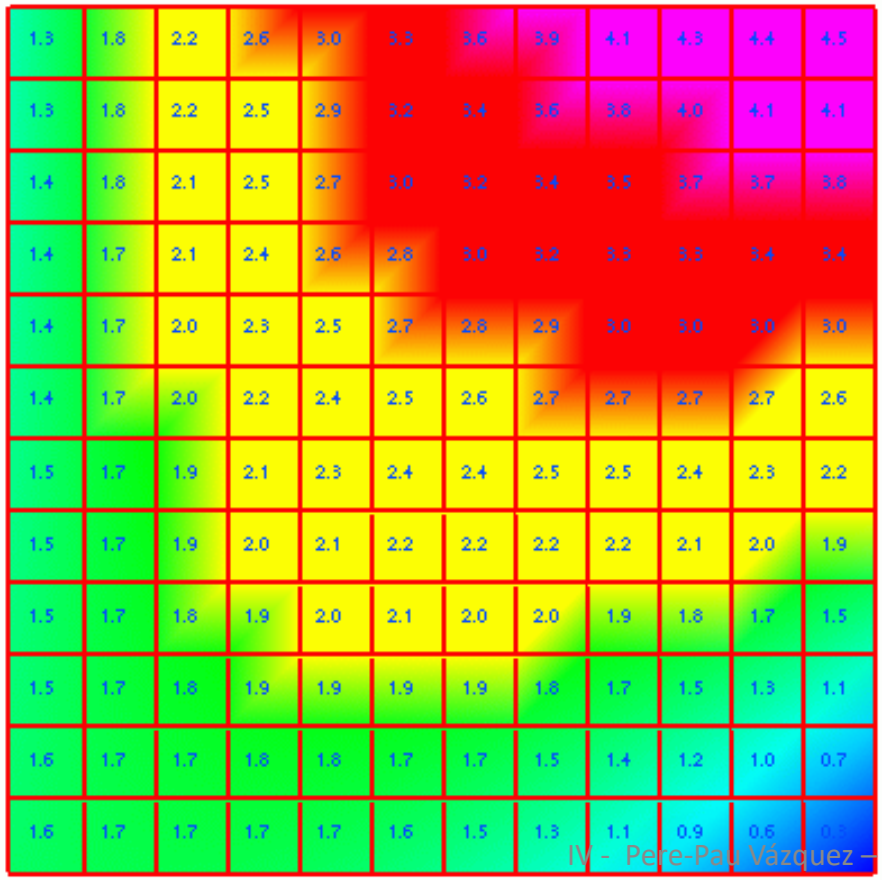
VISUALIZATION. THE BASICS

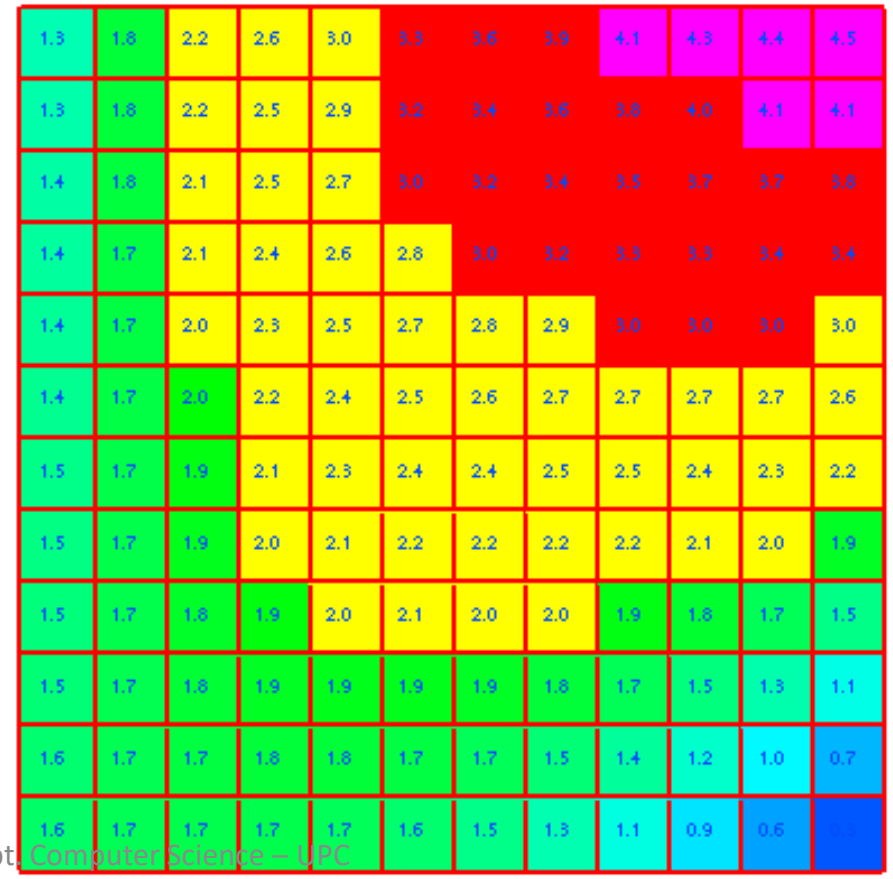
- What if we use **our perceptual system**?



VISUALIZATION. THE BASICS

- And even simplify?





VISUALIZATION. THE BASICS

- Representations of **datasets**
  - Summaries lose information, details matter
    - Confirm expected and find unexpected patterns
    - Assess validity of statistical model
  - Famous example, the Anscombe's quartet

VISUALIZATION. THE BASICS

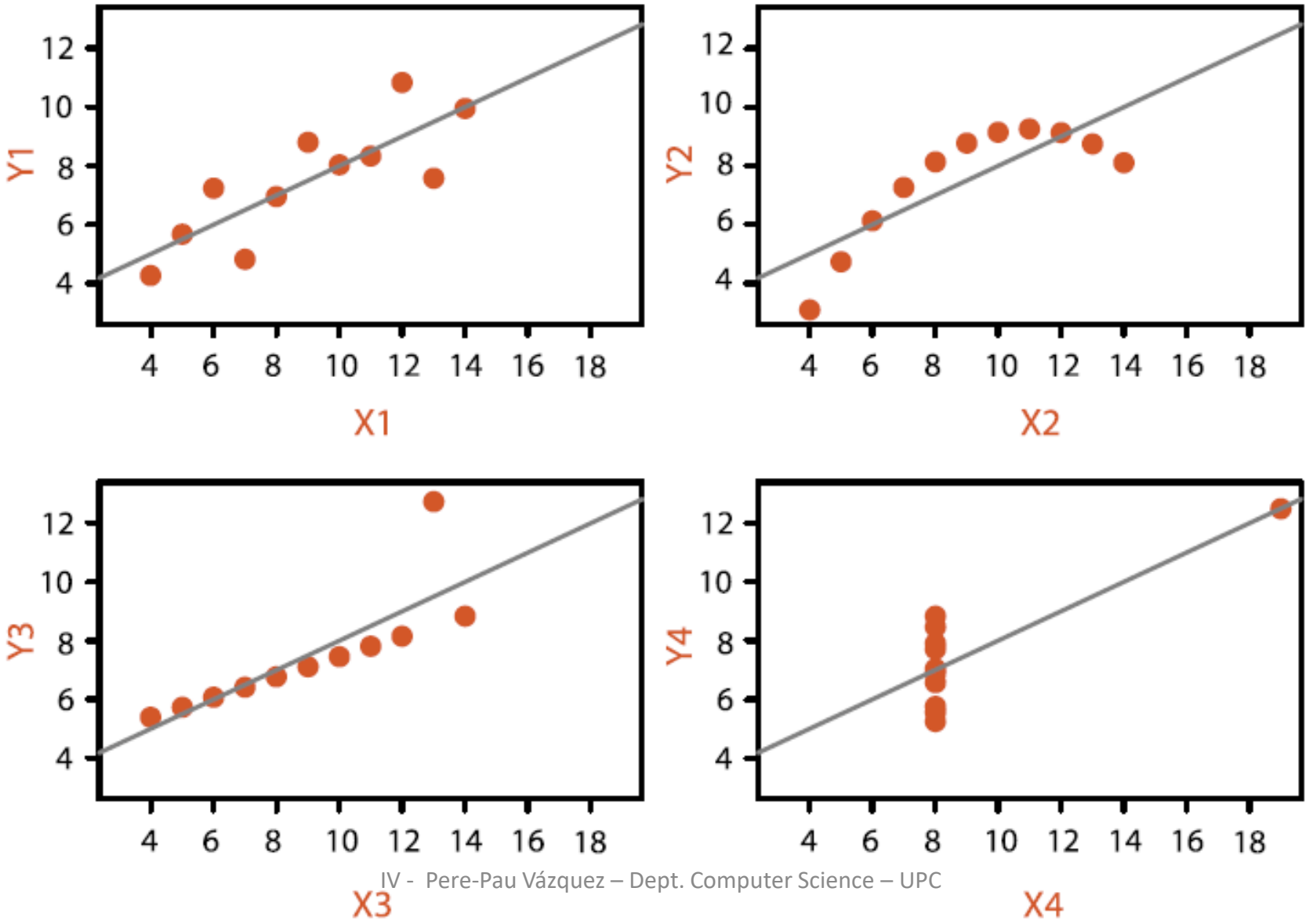
Anscombe's Quartet: Raw Data

	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	6.89
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	

VISUALIZATION. THE BASICS

- Same statistical properties:
  - Number of observations (n): 11
  - Mean x: 9.0
  - Mean y: 7.5
  - Equation of regression line: $y = 3 + 0.5x$
  - Sums of squares of $x - mean_x$: 110.0
  - Regression sums of squares: 27.50 (1 d.f.)
  - Residual sums of squares of y: 13.75 (9 d.f.)
  - Multiple R^2 : 0.667

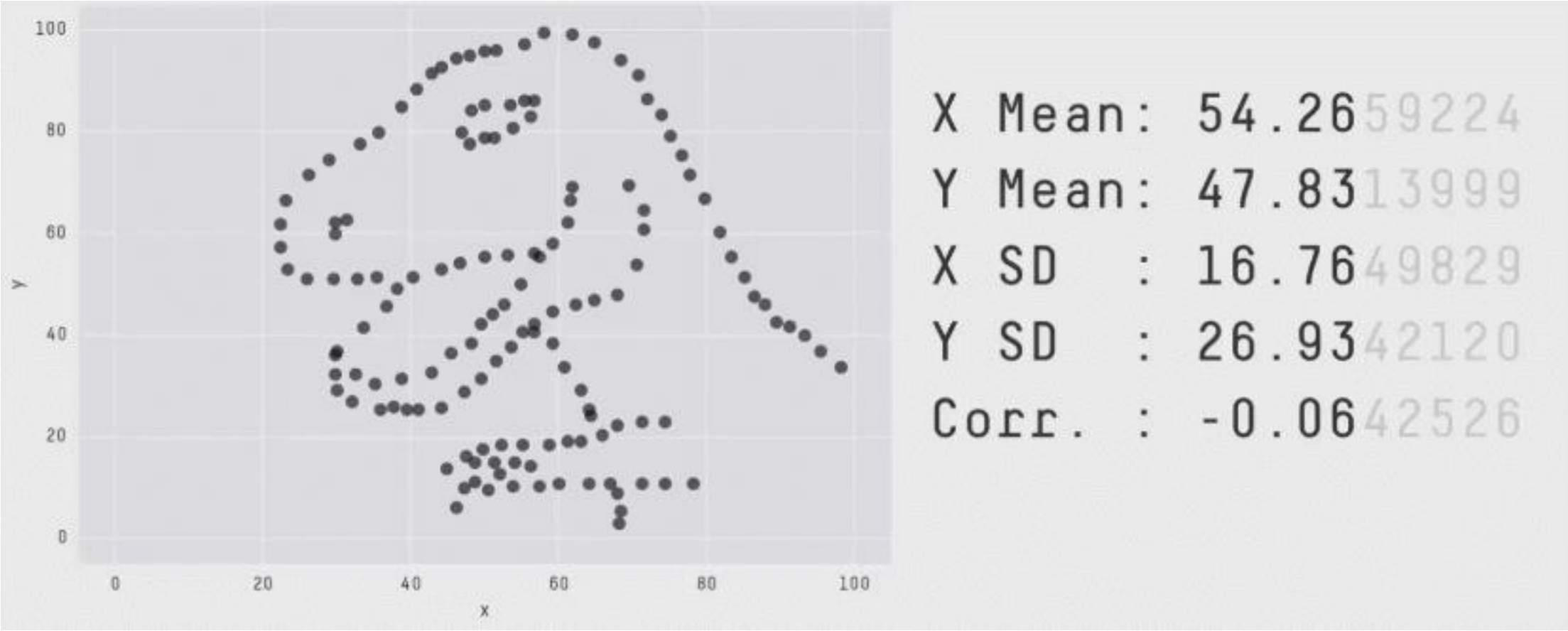
VISUALIZATION. THE BASICS



“...make both calculations and graphs.
Both sorts of output should be studied;
each will contribute to understanding.”

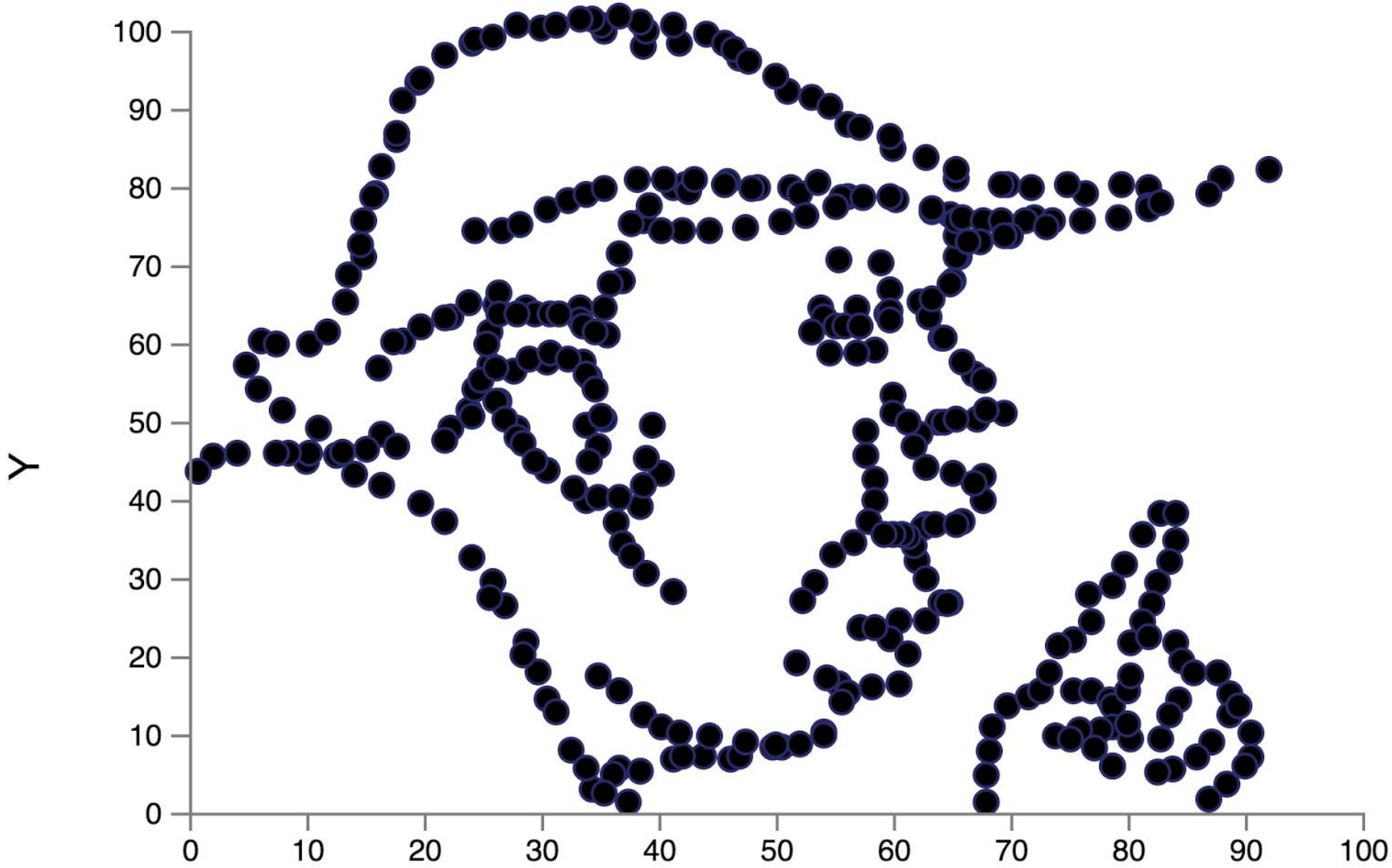
F. J. Anscombe, 1973

VISUALIZATION. THE BASICS

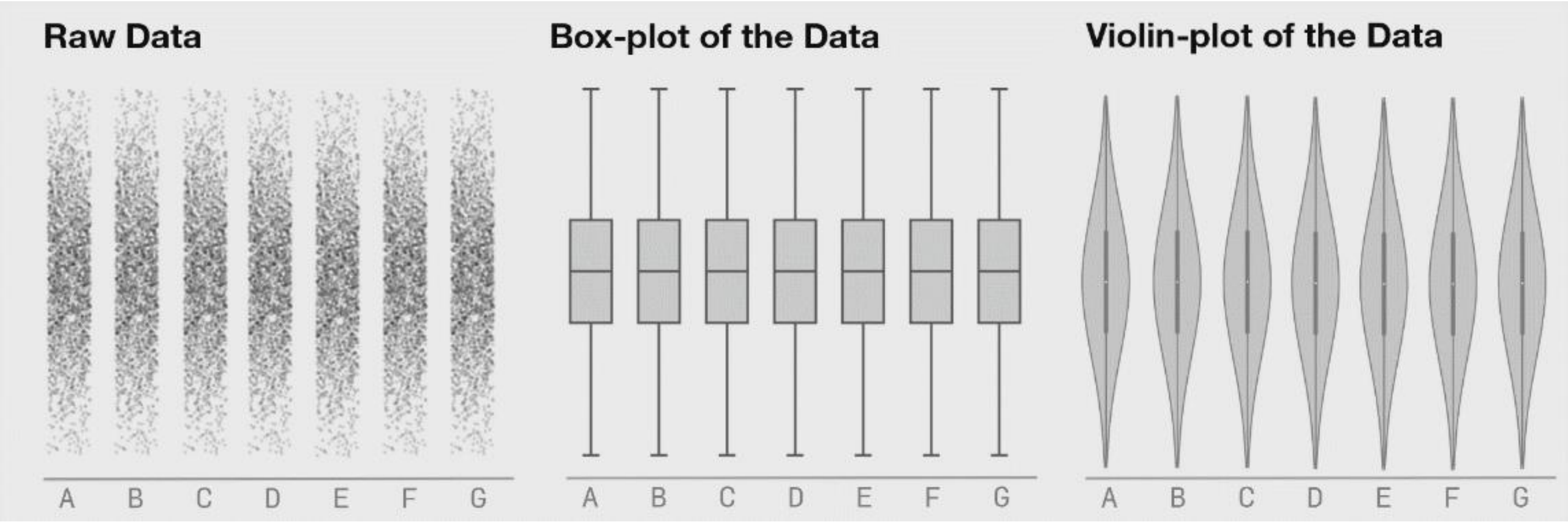


N = 383 ; X mean = 49.4905 ; X SD = 22.0677 ; Y mean = 50.4382 ; Y SD = 27.7443 ; Pearson correlation = -0.284

VISUA

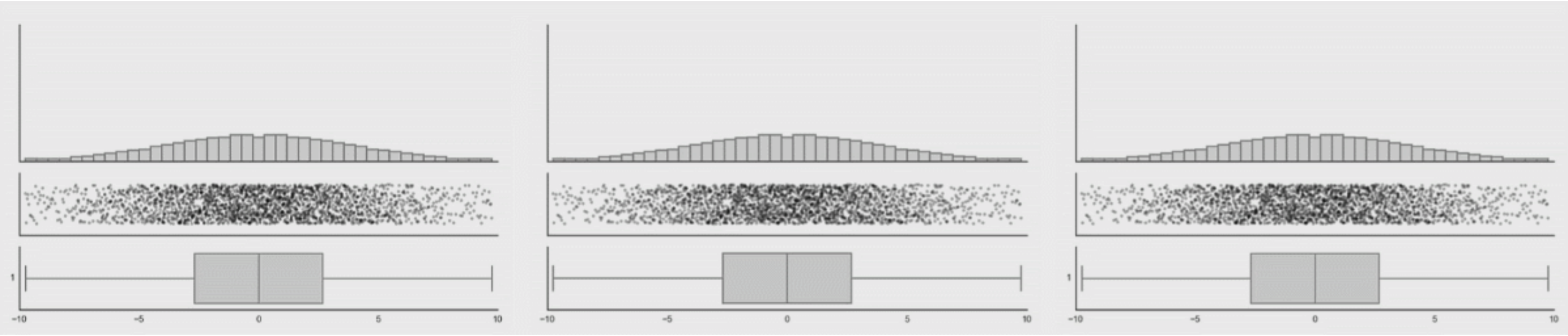


VISUALIZATION. THE BASICS



From Autodesk: https://www.autodesk.com/research/publications/same-stats-different-graphs

VISUALIZATION. THE BASICS



From Autodesk: https://www.autodesk.com/research/publications/same-stats-different-graphs

VISUALIZATION. THE BASICS

- Visualization requires a purpose
  - E.g., understanding of mortality rate, life expectancy, growth… (How the world has changed!) using publicly available data

    https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen, Hans Rosling, TED Talk 2006

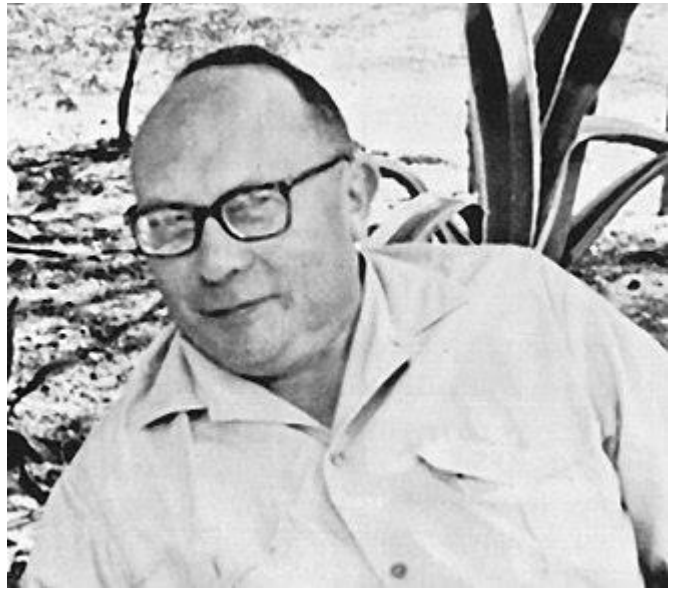
  - Other talks, just Google “Hans Rosling TED”
    - Also by Ola Rosling: https://youtu.be/0o0vn9PS-KY

VISUALIZATION. HOMEWORK

- Answer those questions:
  - How many variables (and which) are displayed in the following charts:
    - The one shown at 2:26
    - The one shown at 13:07
  - Find an example of visual representation that does not effectively communicate the message (minute and second, and reason why)
  - What happens in terms of variables when he “splits South Africa”?
  - Bonus: Who is the “ghost”?

"A graphic is not 'drawn' once and for all; ***it is 'constructed' and reconstructed until it reveals all the relationships*** *constituted by the interplay of the data. The best graphic* ***operations*** *are those* ***carried out by the decision-maker*** *themself."*

Jacques Bertin



VISUALIZATION. THE BASICS

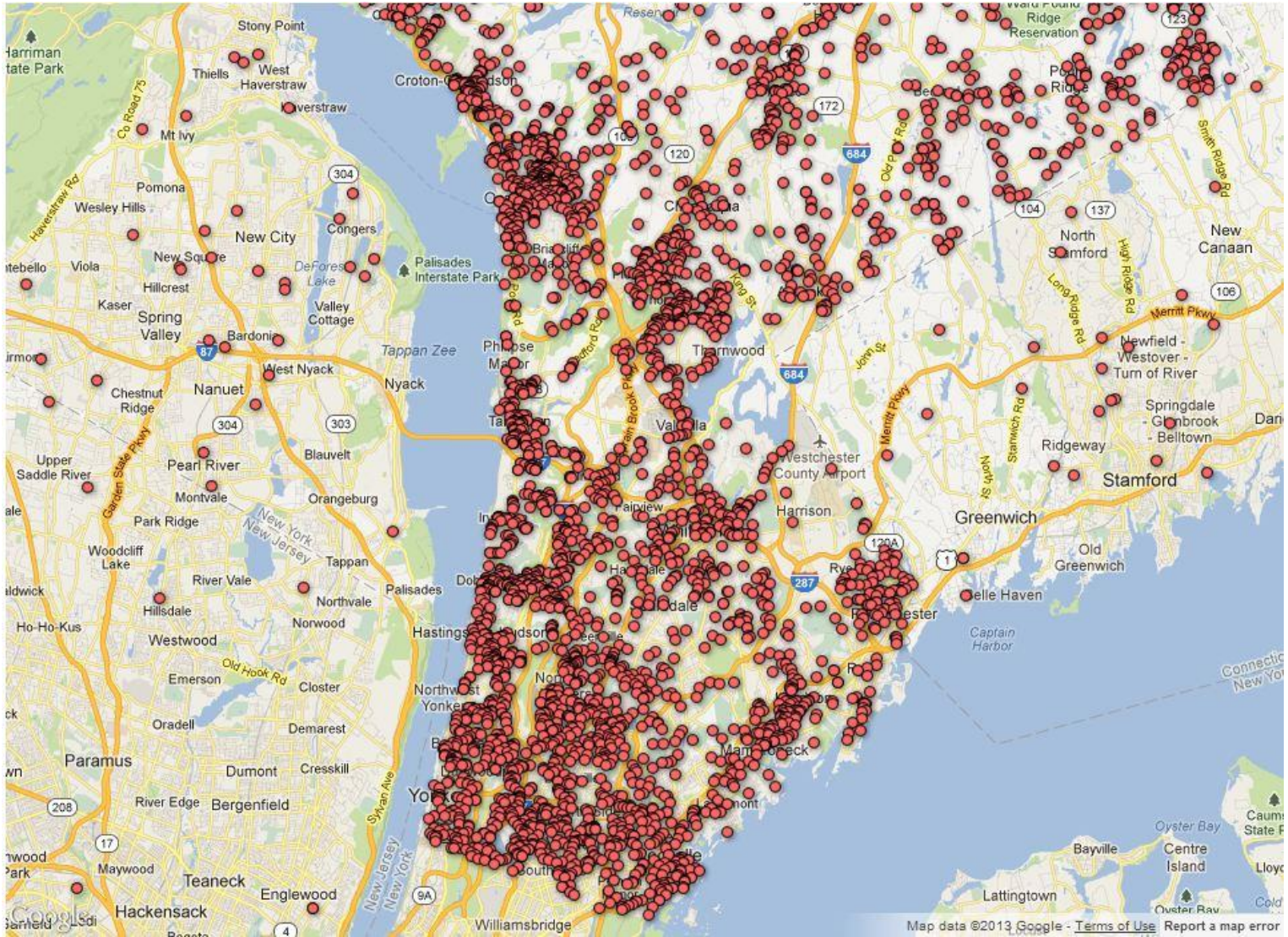
- Main applications of visualization
  - Explanatory: Present results
  - Analysis: Analyse hypothesis
  - Exploratory: Inspect data

VISUALIZATION. THE BASICS

- Presentation
  - Communicate data and ideas
  - Explain and inform
  - Provide evidence and support
  - Influence and persuade
    - See also Rosling's 2009 TED talk "Let my dataset change your mindset"

- Commonly only showing a few variables of the data
  - Yet its impact may be HUGE!!!

VISUALIZATION. THE BASICS



VISUALIZATION. THE BASICS

- Increased awareness of weapons in the neighborhood
- But…
  - The journal was attacked
    - Names of the children and the schools where they went were published in newspapers…

VISUALIZATION. THE BASICS

- Analysis
  - There are other methods for data analysis
    - Statistics
    - Database & information retrieval
    - Data mining
    - Machine learning

VISUALIZATION. THE BASICS

- Analysis. Typical objectives
  - Show many variables
  - Illustrative overview and detail
  - Facilitate comparison
  - Other methods (statistics, learning…) may be insufficient

- Presentation might choose some parts
  - Analysis will focus on **all of them**

VISUALIZATION. THE BASICS

- Exploration:
  - Typically difficult to show information so that it can be interpreted properly
  - Visualization very useful in exploratory data analysis
    - Don’t know what you’re looking for
    - Don’t have a priori questions
    - Want to know what questions to ask
  - Still requires effective communication

VISUALIZATION. THE BASICS

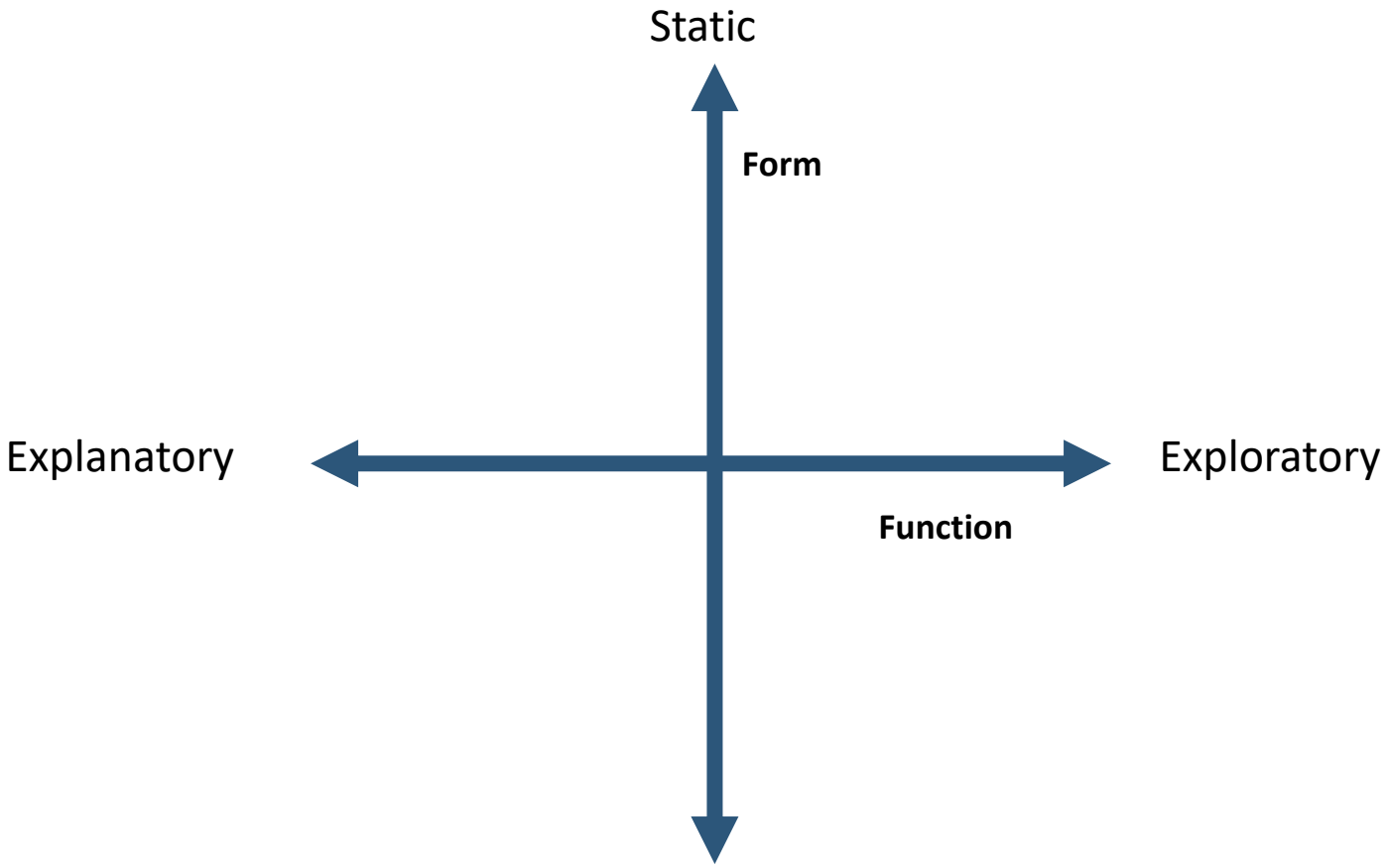
- Exploration:
  - Encourages users to inspect the data
    - Learn new things -> **insights**
  - Visualization very useful in exploratory data analysis
    - Don’t know what you’re looking for
    - Don’t have a priori questions
    - Want to know what questions to ask
  - Still requires effective communication

VISUALIZATION. THE BASICS

- Other way of presenting it… Visualization serves for …
  - … exploration
    - Nothing is known,
      Vis. used for gathering knowledge on the data
  - … analysis
    - There are hypotheses,
      Vis used verification or falsification
  - … presentation
    - “everything” known about the data,
      Vis. used for communication of results

VISUALIZATION. THE BASICS

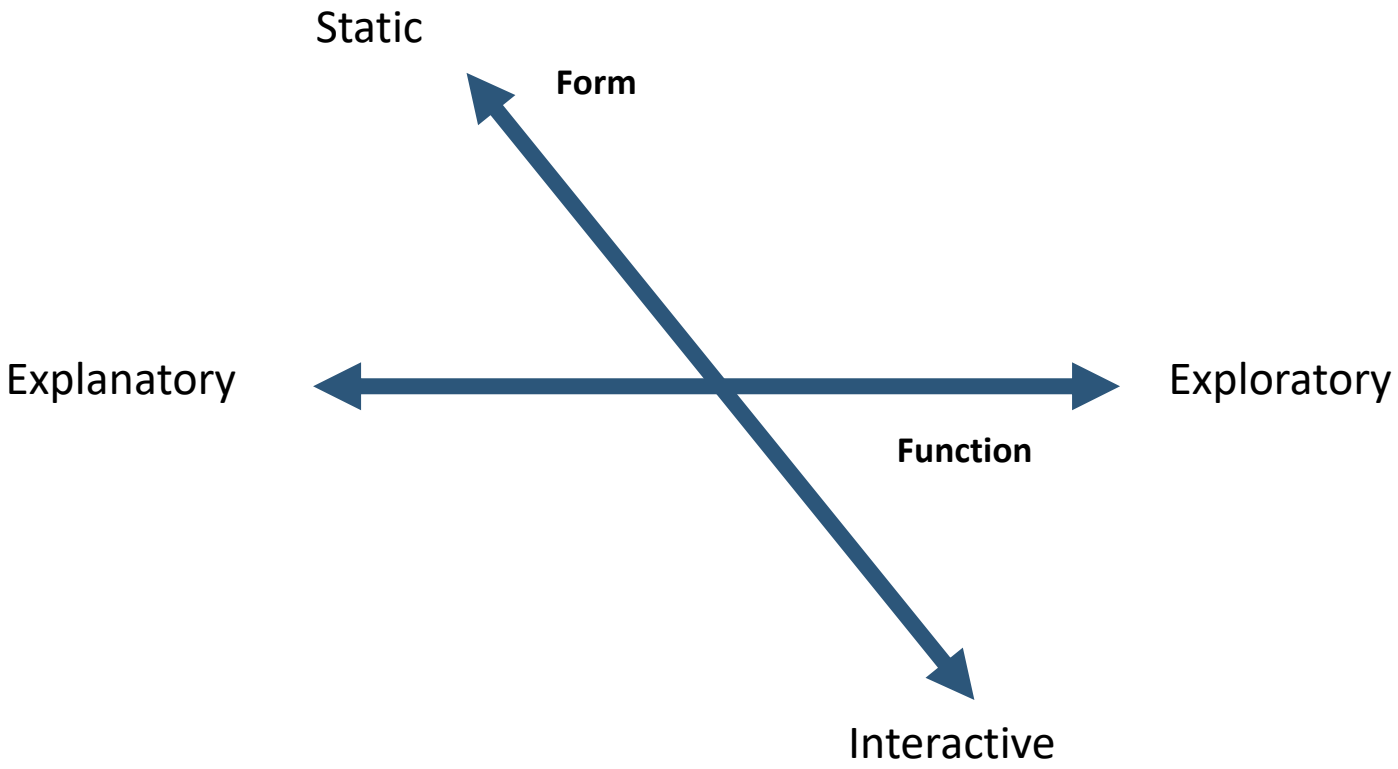
- Form vs function



[Schwabish, 2014]

VISUALIZATION. THE BASICS

- Form vs function. **My take**



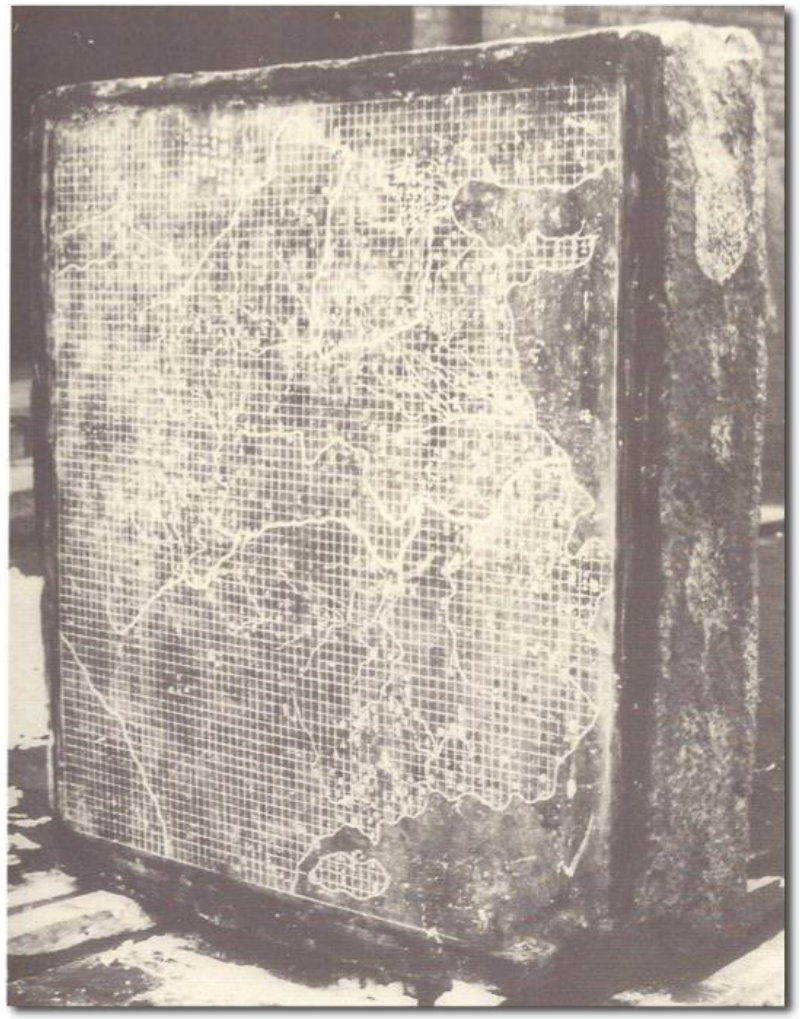
OUTLINE

- *Visualization. The basics*
- **History**
- General Rules
- Data, Tasks, Users
- Visualization as a Design Process
- The Visualization Mantra

HISTORY

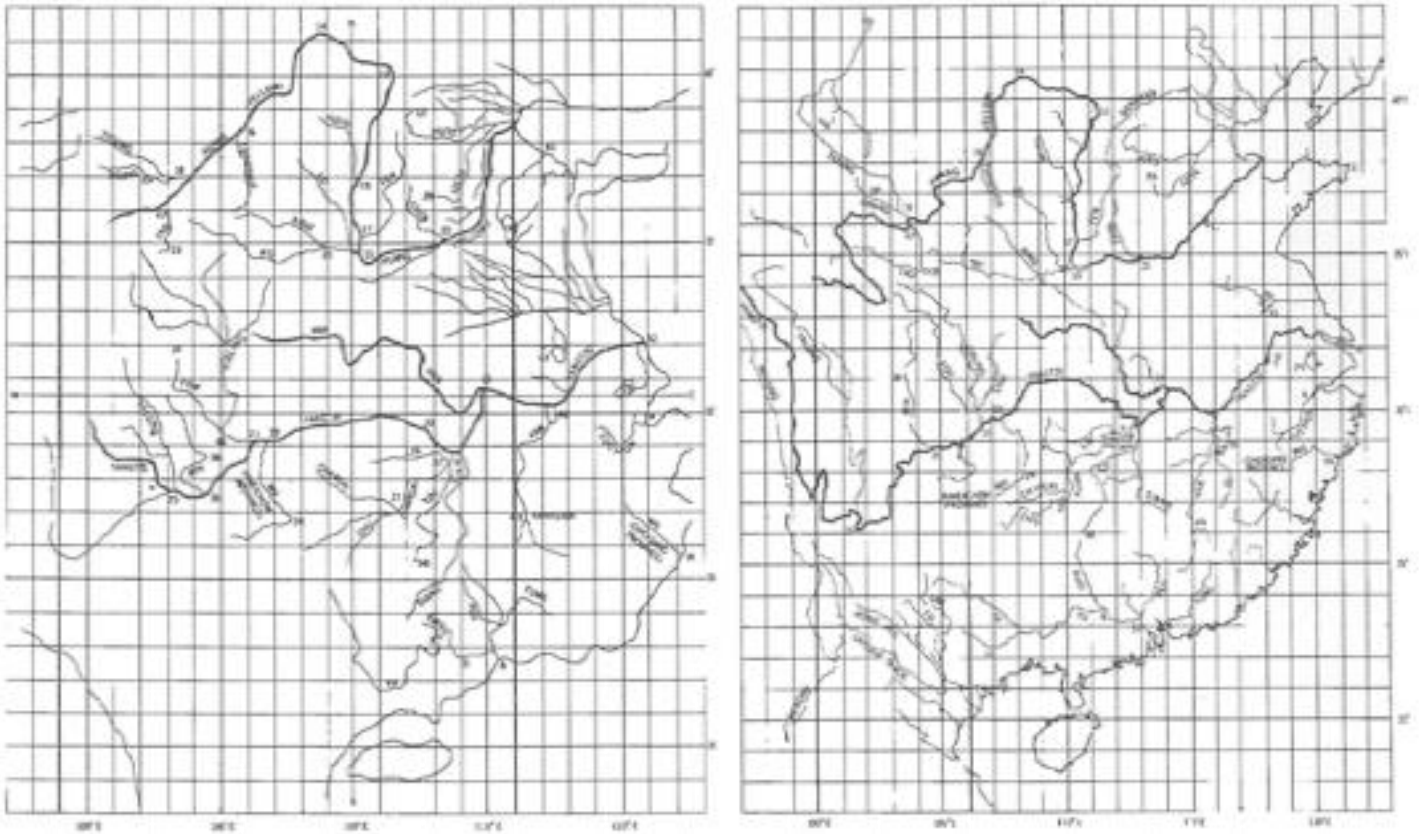
- But when did this start?
  - Quite long ago…
    - E. g. we may track cartographic examples for thousands of years
      - E.g.: The Yu Ji Tu (Map of the Tracks of Yu the Great), a map carved into stone in the year 1137 during the Song Dynasty
      - Uses Cartesian coordinates
        - Grid with longitudinal and latitudinal lines

HISTORY



HISTORY

- … vs current China map



A redrawing of the Yü-chi T'u with an oblong grid superimposed empirically from the geography, the numbers refer to the accompanying table. (from Happgood)

A modern map of China, on a comparable scale to the Yü-chi T'u and corresponding numbers

HISTORY

- Cholera epidemic in London 1854

HISTORY

- Dr. John Snow

HISTORY

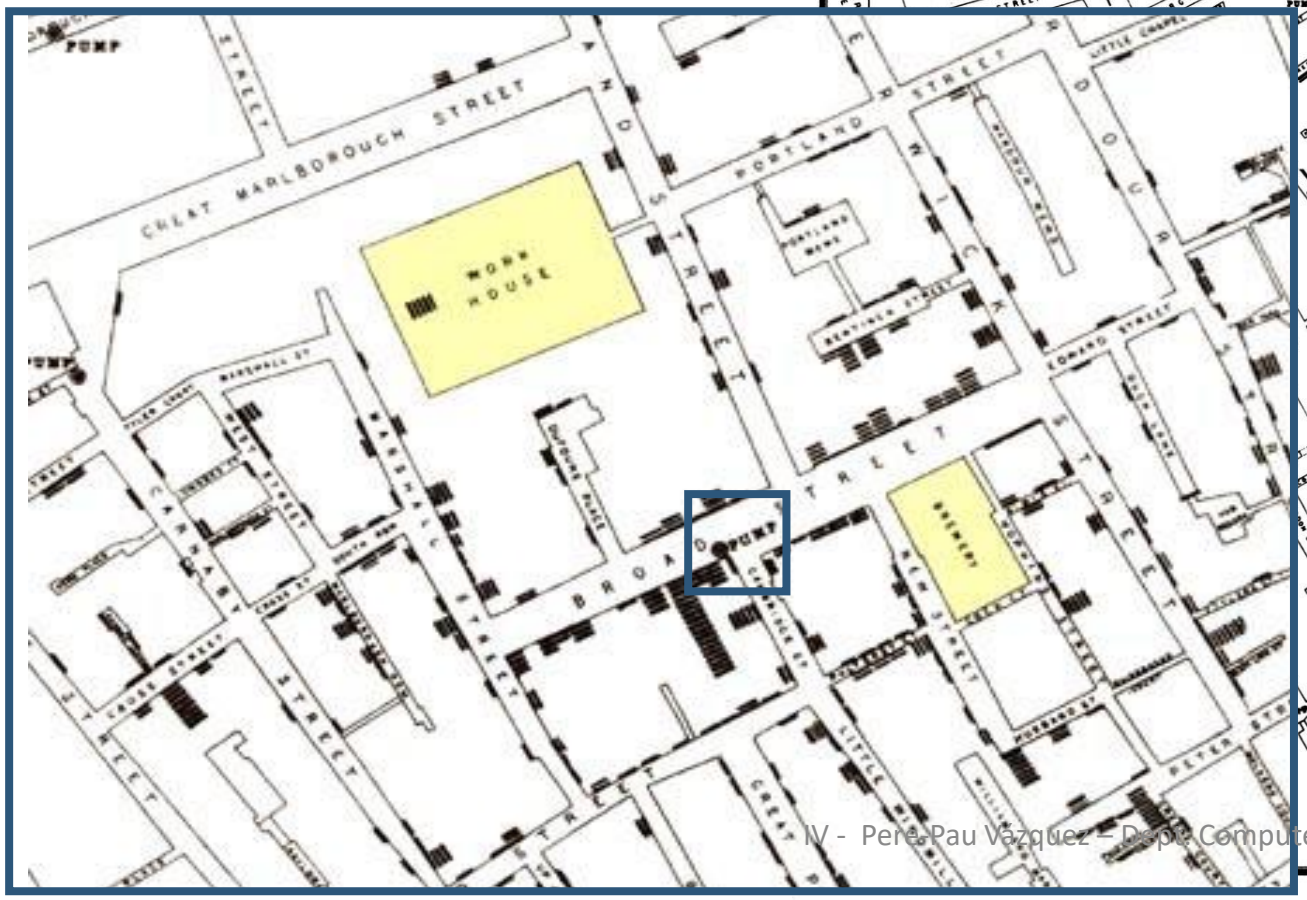
- Cholera epidemic in London 1854
  - Dr. John Snow was able to trace the source of the cholera outbreak in Soho
  - His findings inspired fundamental changes in the water and waste systems in London
  - He is considered one of the fathers of modern epidemiology

HISTORY

- Cholera epidemic in London 1854
  - “On proceeding to the spot, I found that **nearly all the deaths had taken place within a short distance of the [Broad Street] pump**. There were only ten deaths in houses situated decidedly nearer to another street-pump. In five of these cases the families of the deceased persons informed me that they always sent to the pump in Broad Street, as they preferred the water to that of the pumps which were nearer. In three other cases, the deceased were children who went to school near the pump in Broad Street...”

INTRODUCTION

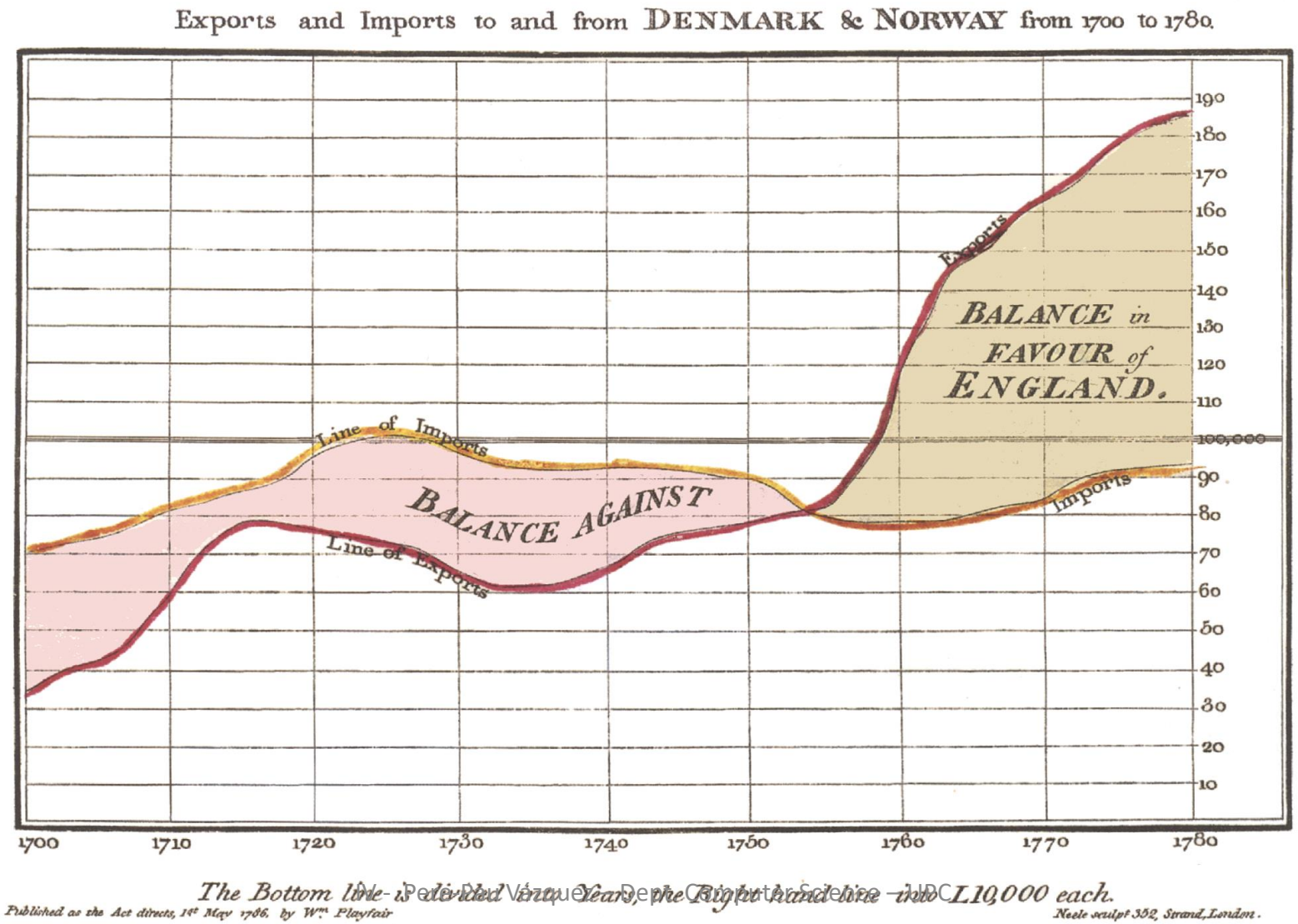
- Drawing by John Snow



HISTORY

- The first visual representations of abstract, nonspatial datasets were created in the 18th century by William Playfair.

HISTORY



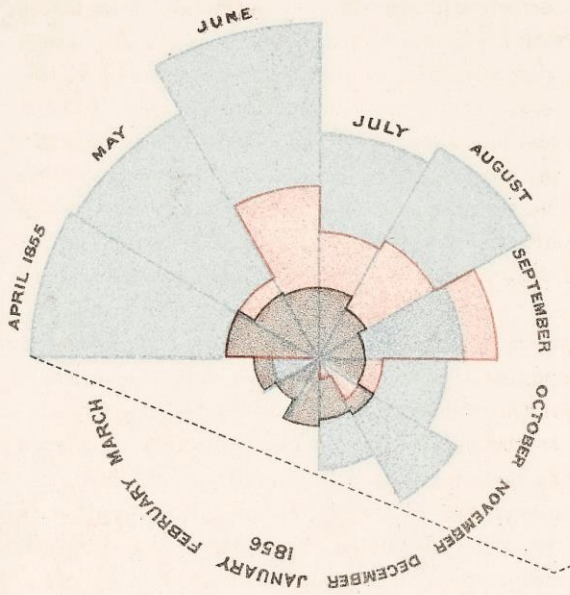
HISTORY

- Florence Nithingale, analyzing mortality in the Crimean War
  - Mortality due to zymotic disease was reduced by increasing sanitary improvements in military barracks

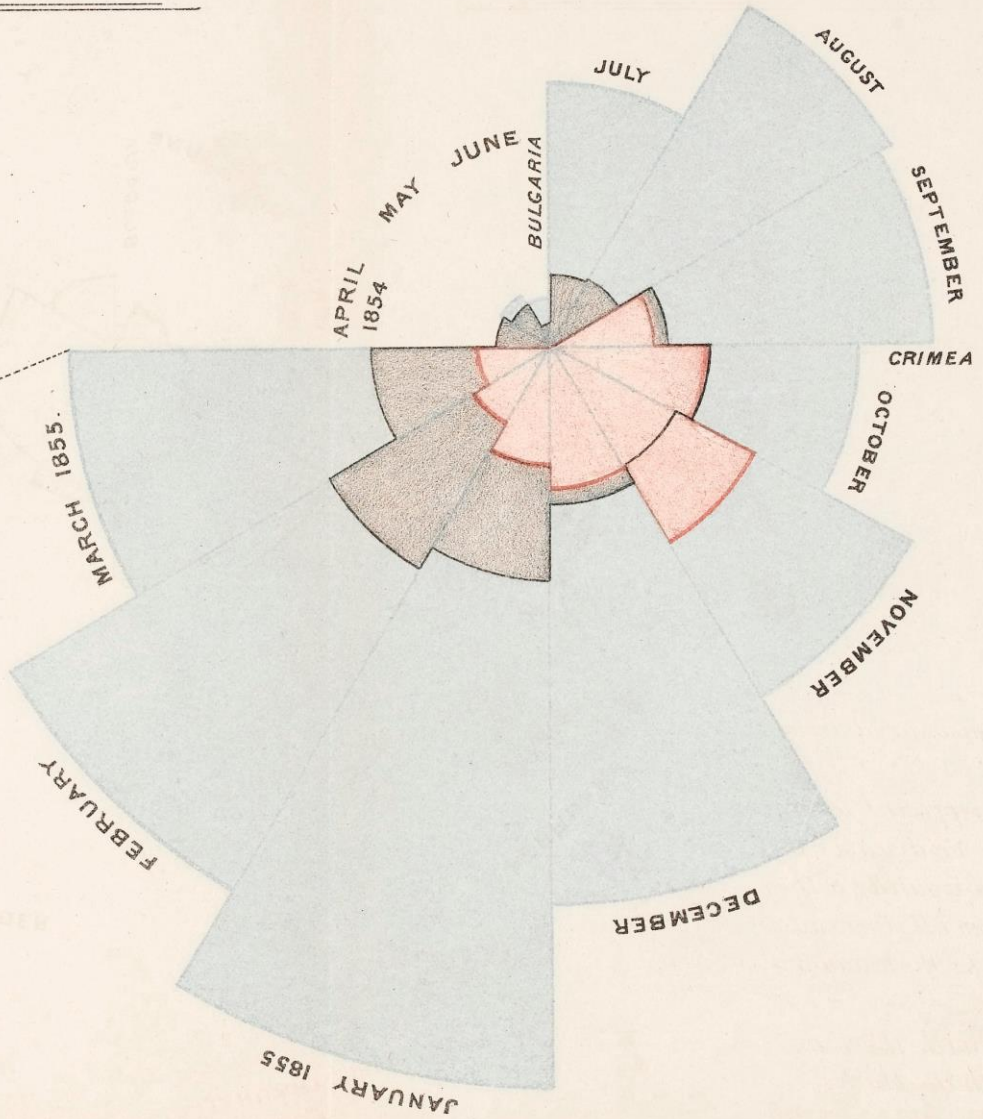
H

DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.



1.
APRIL 1854 TO MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov.r 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red, in January & February 1855, the blue coincides with the black.

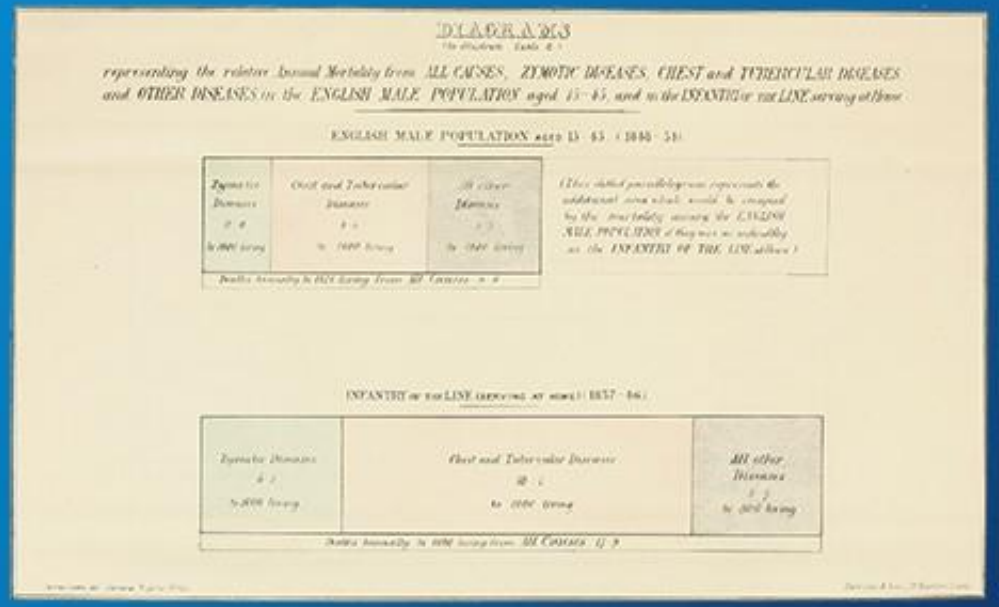
The entire areas may be compared by following the blue, the red & the black lines enclosing them.

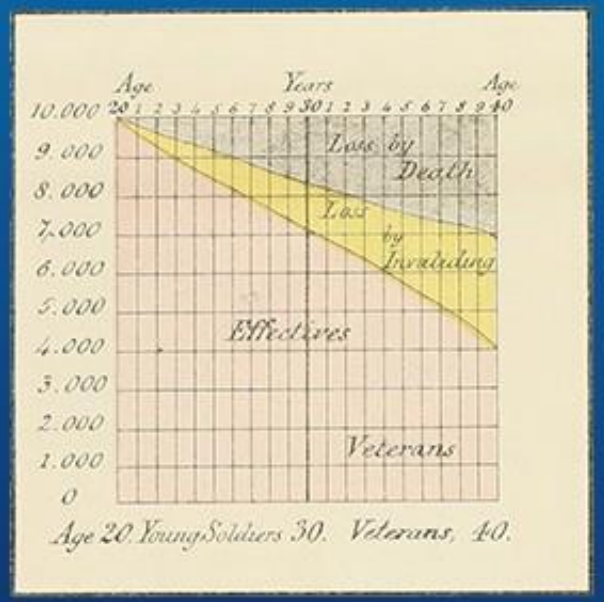
Harrison & Sons, St. Martin's Lane.

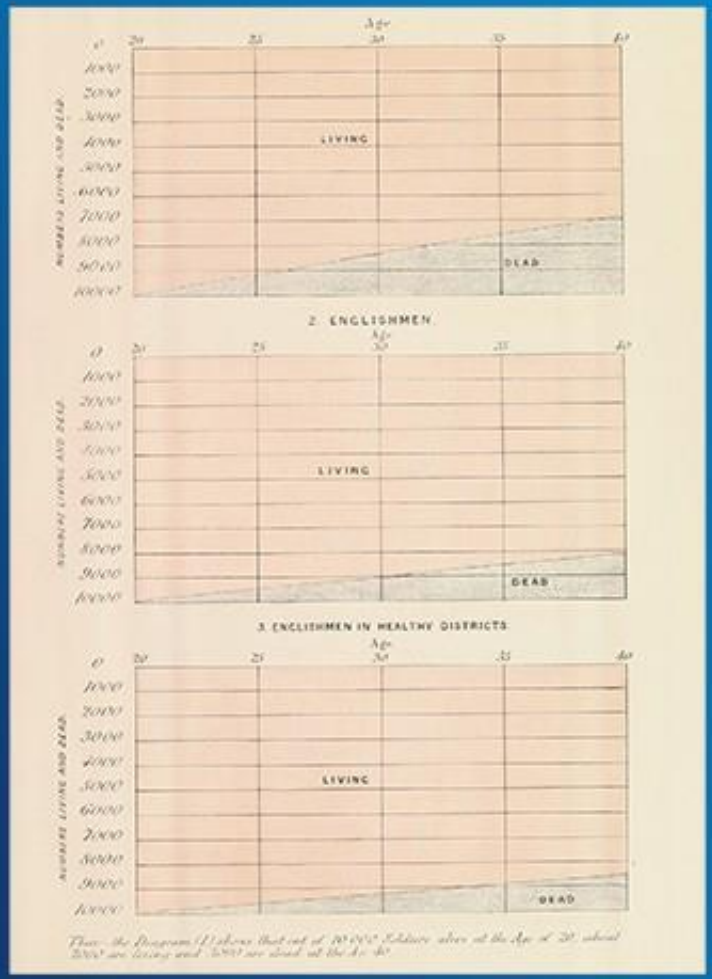
Diagram of the causes of mortality in the army. Credit: wellcomecollection.org

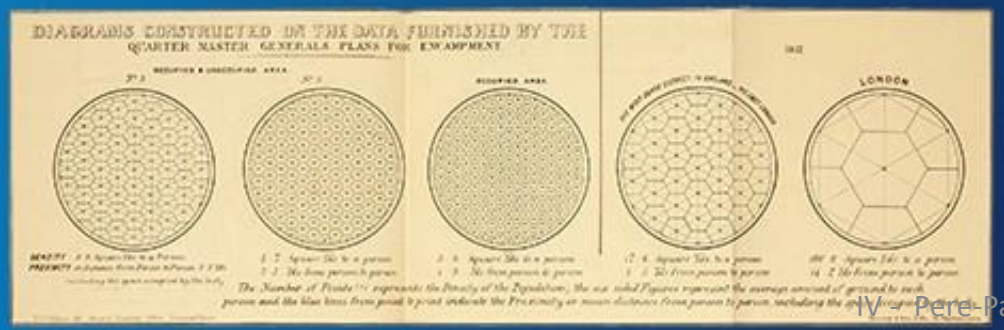
HISTORY

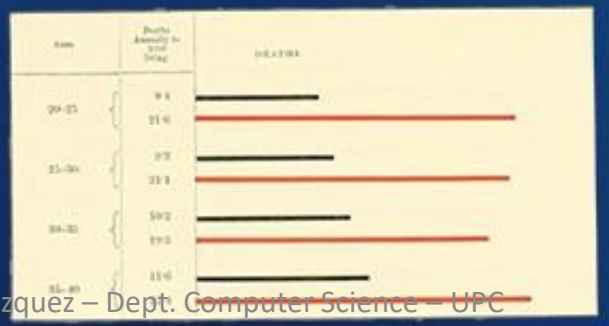
Diagrams from “Mortality of the British Army: at home and abroad” 1858 (https://archive.org/details/mortalityofbriti00lond/page/n53/mode/2up).







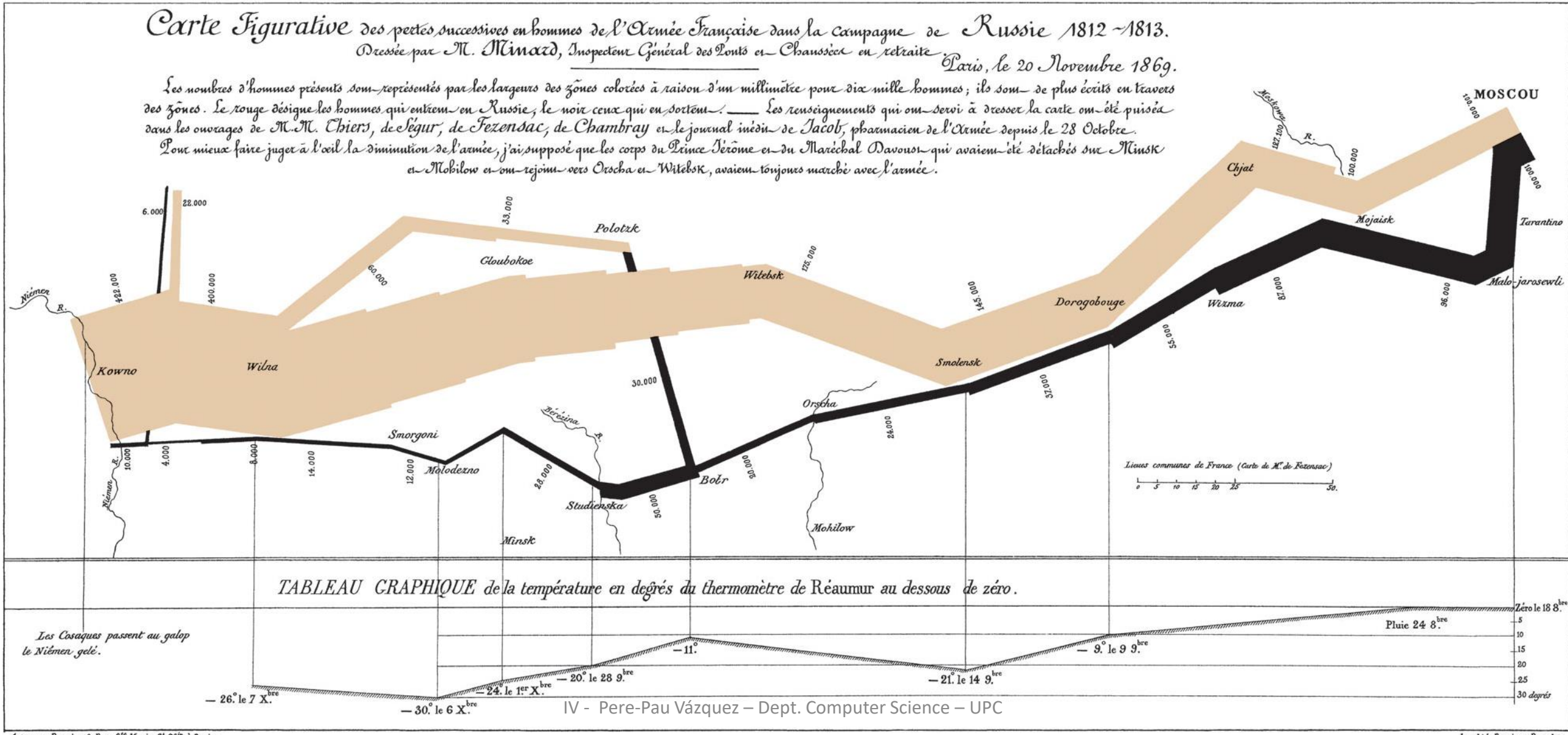




HISTORY

- Other examples… the Russian campaign by Napoleon

HISTORY


Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.
Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.
Paris, le 20 Novembre 1869.
Les nombres d'hommes présents sont représentés par les largeurs des zônes colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zônes. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. ___ Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M.M. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.
Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mobilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.
Niémen R.
Kowno
Wilna
Smorgoni
Niémen R.
Molodezno
Minsk
Berézina R.
Studienska
Bobr
Polotzk
Gloubokoe
Witebsk
Orscha
Mohilow
Smolensk
Dorogobouge
Chjat
Mojaisk
Wizma
Moskowa R.
MOSCOU
Tarantino
Malo-jarosewli
6.000
22.000
422.000
400.000
33.000
60.000
175.000
145.000
30.000
24.000
50.000
37.000
55.000
127.100
100.000
100.000
96.000
10.000
4.000
8.000
14.000
12.000
28.000
Lieues communes de France (Carte de M. de Fezensac)
0 5 10 15 20 25 50.
TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.
Les Cosaques passent au galop le Niémen gelé.
— 26.° le 7 X.bre
— 30.° le 6 X.bre
— 24. le 1er X.bre
— 20.° le 28 9.bre
— 11.°
— 21.° le 14 9.bre
— 9.° le 9 9.bre
Pluie 24 8.bre
Zéro le 18 8.bre
5
10
15
20
25
30 degrés
Autog. par Regnier, 8. Pas. Ste Marie St Gein à Paris.
Imp. Lith. Regnier et Dourdet.

HISTORY

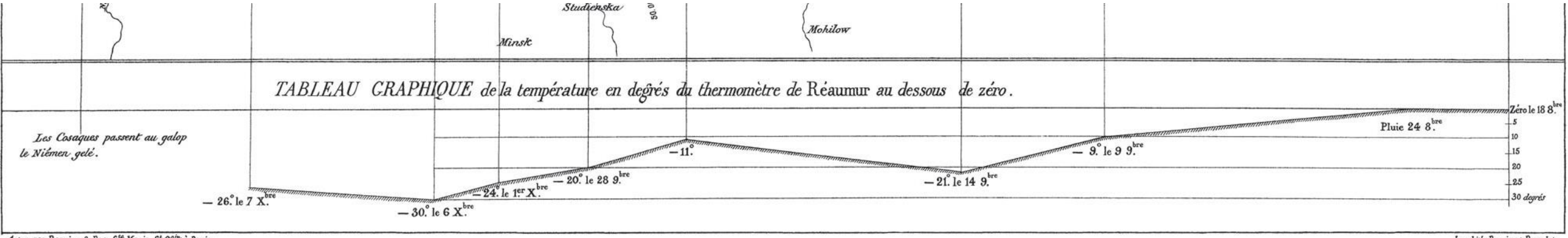
Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zônes colorées à raison d'un millimètre pour dix mille hommes ; ils sont de plus écrits en travers des zônes. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. ___ Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M.M. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mobilow en ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.

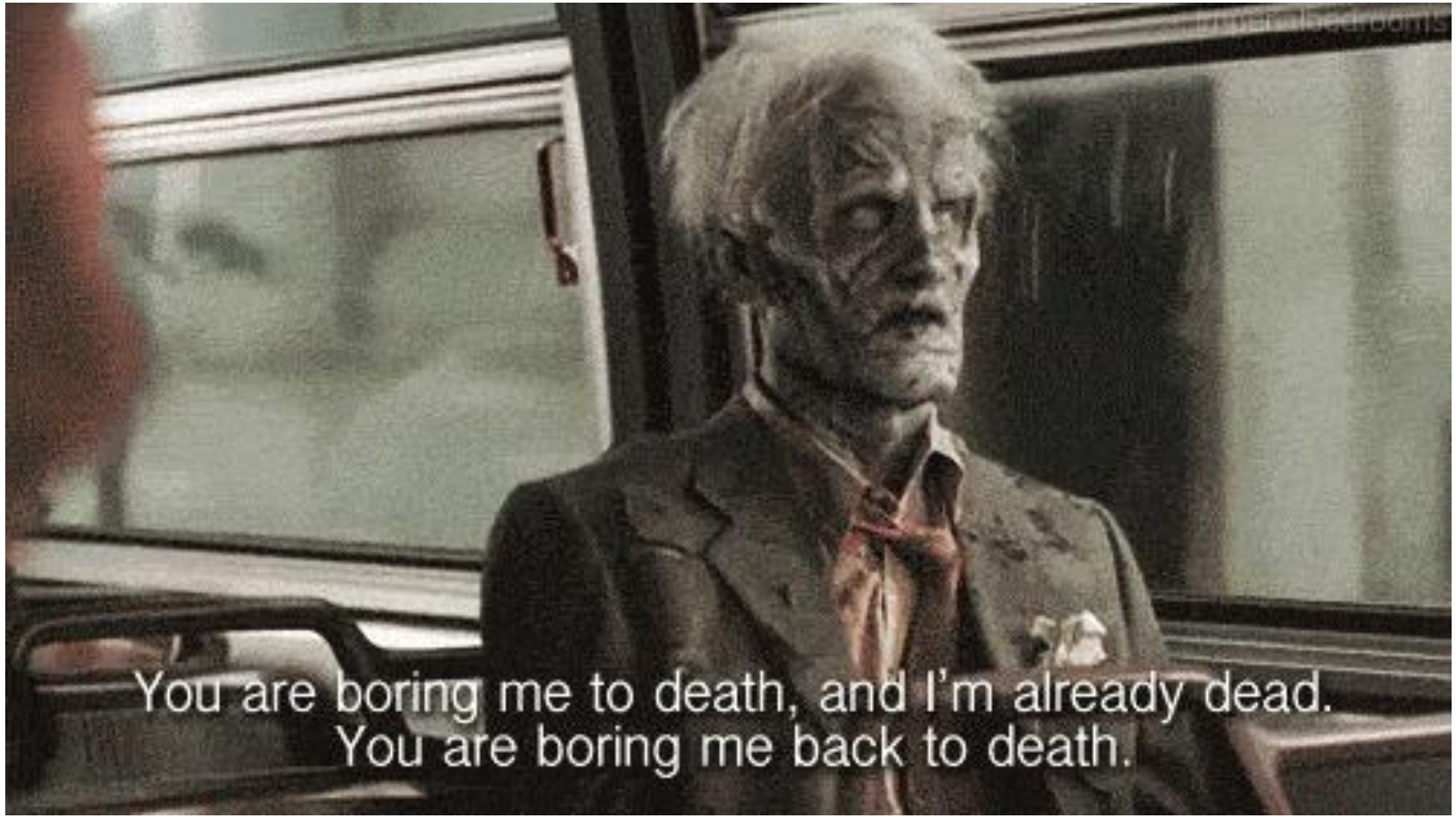


Autog. par Regnier, 8. Pas. Ste Marie St Gain à Paris.

Imp. Lith. Regnier et Dourdet.

INTRODUCTION TO VISUALIZATION

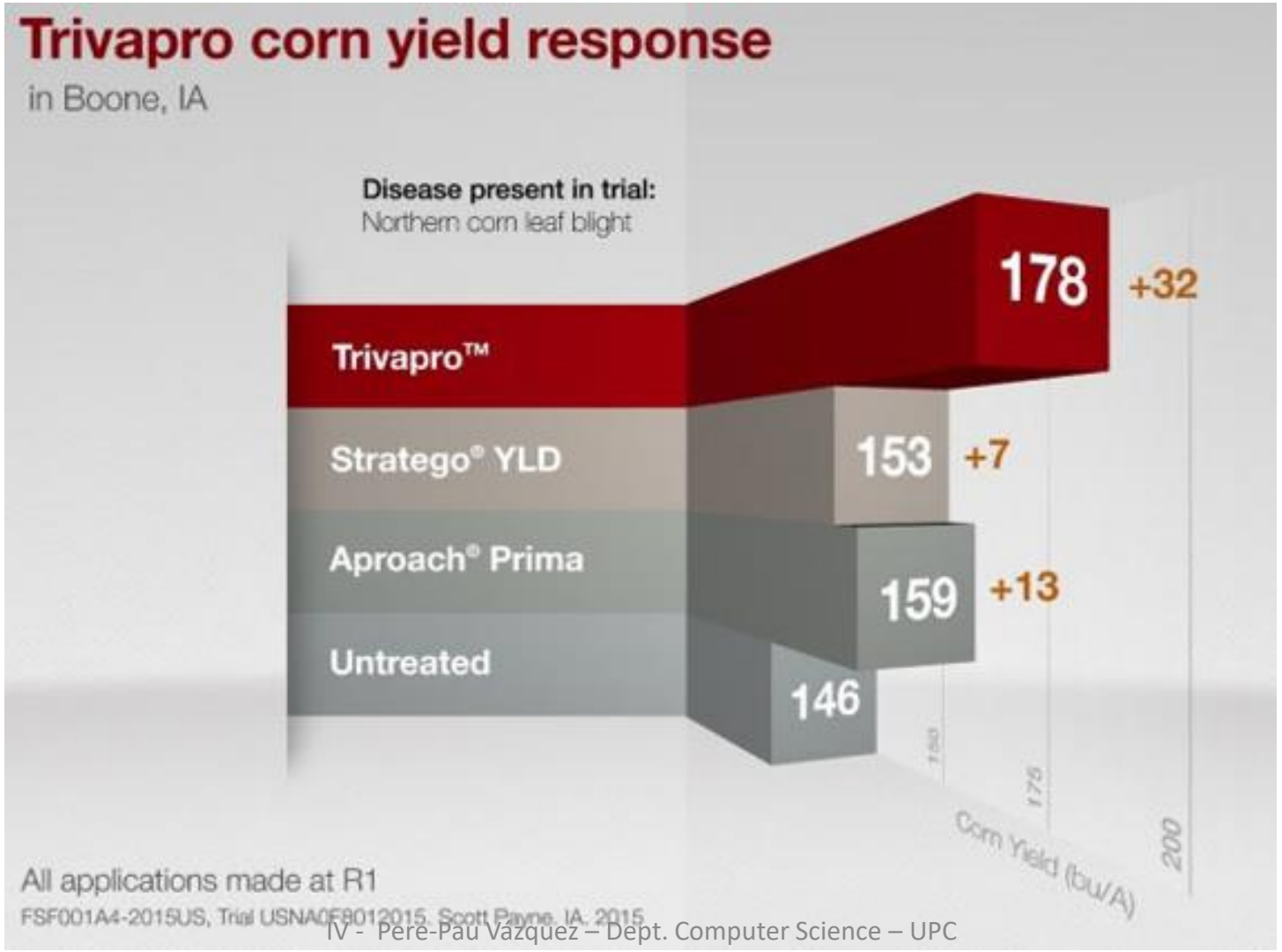
PERE-PAU VÁZQUEZ – VIRVIG GROUP – UPC


You are boring me to death, and I'm already dead.
You are boring me back to death.

EXERCISES

PERE-PAU VÁZQUEZ – VIRVIG GROUP – UPC

ANALYZE THE FOLLOWING CHART



ANALYZE THE FOLLOWING CHART

- Main issues:
  - Use of 3D
  - Bars do not start at zero
- Other issues
  - Redundancy in the data
  - Perspective projection
  - Lack of order in bars

ANALYZE THE FOLLOWING CHART

- **Use of 3D:** Using 3D for the bars and also at such a crooked angle makes it difficult to visually compare the data. Also, the perspective projection distorts the geometry even more. In particular, the nearest bar, the red one, looks much larger than the data it represents.
- **Bars do not start at zero:** if we only count the 3D part, the sizes of the bars are not proportional to the data, since they do not start at zero. If we count the base, maybe yes or maybe not (we can't measure it precisely enough), in any case, if the goal of the designers was to combine the two pieces, this makes it very difficult to appreciate the total length.

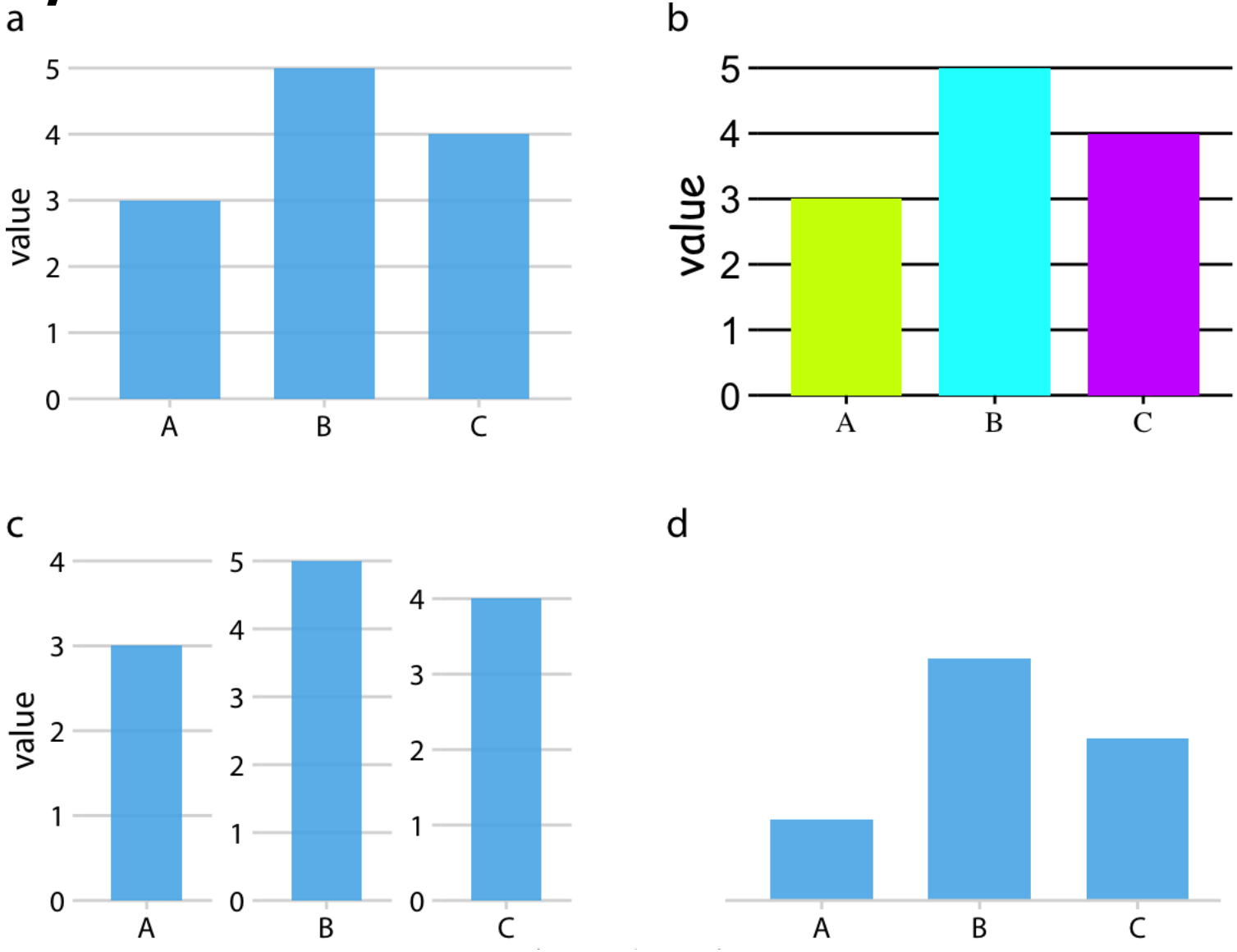
ANALYZE THE FOLLOWING CHART

- **Redundancy in the data:** since the numbers need to be put in because the data is not sufficiently understood, this makes the axes redundant. Also, since increments are also put in, which can be deduced from the numerical data, we have a second source of redundancy.
- **Perspective projection:** The 3D itself may be more or less harmful, a better projection would have facilitated the visual comparison
- **Lack of order in bars:** Categories are not ordered alphabetically or by quantity. There is no other apparent order

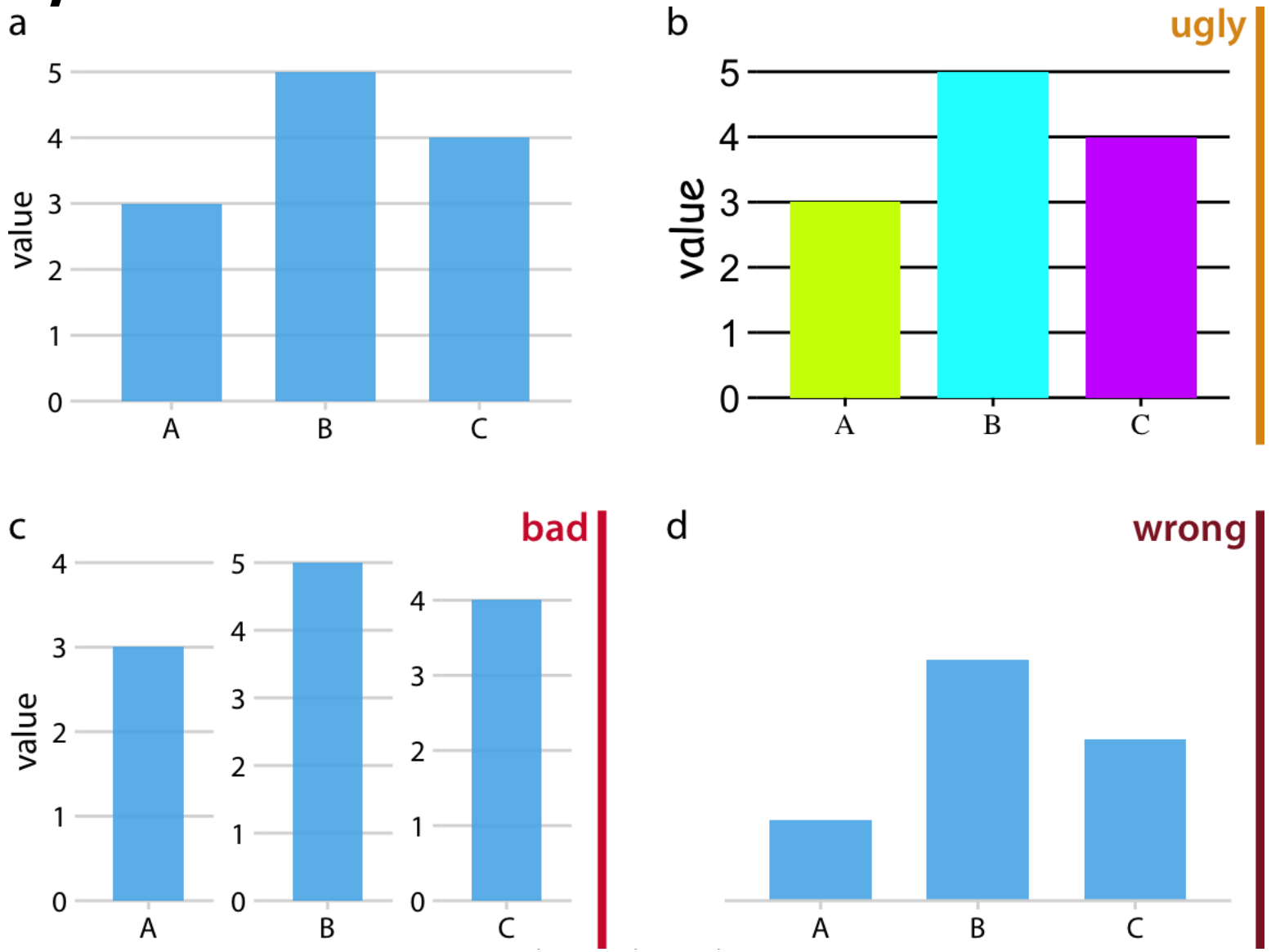
ANALYZE THE FOLLOWING CHART

- Extra points
  - The title is not entirely descriptive, since it not only shows the effect of Trivapro, but compares it with other products or with the fact of not using any treatment
  - The color also does not seem justified, although one would say that it is used to highlight the product that interests the most
  - It also uses a lot of space (a very poor ink ratio) to end up encoding only 4 values

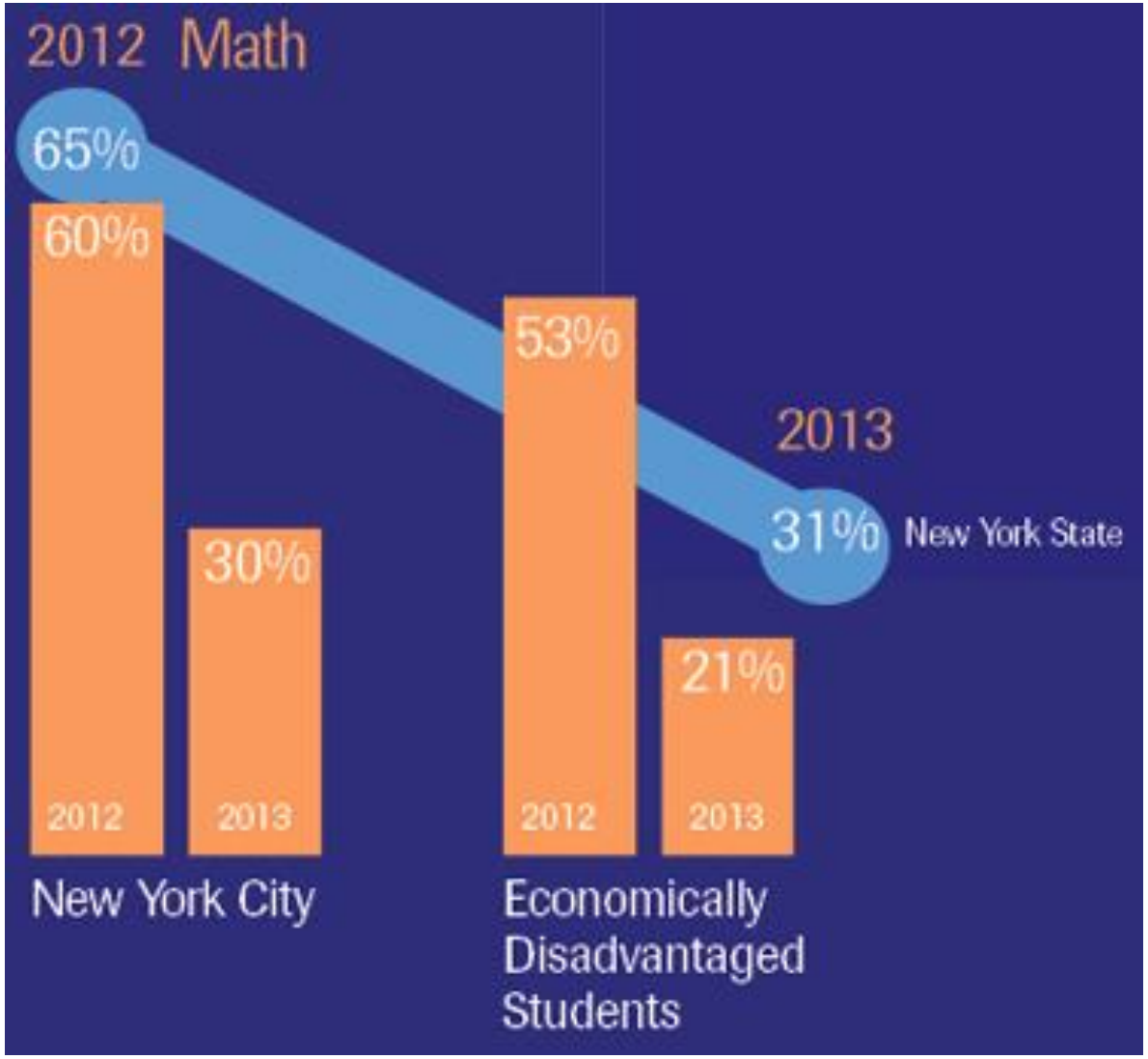
SAME DATA, DIFFERENT CHARTS...



SAME DATA, DIFFERENT CHARTS...



EXERCISES


2012 Math
65%
60%
30%
53%
21%
2013
31%
New York State
2012
2013
2012
2013
New York City
Economically Disadvantaged Students

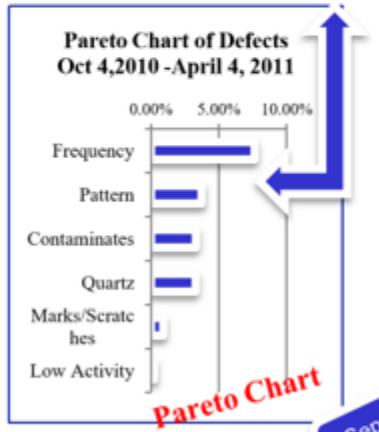
Finding the Skinny on Thin Film Sensor Reject Rates

DEFINE – 5/15/11

1) Problem Statement:
Production reject rate of thin film sensors increases after process change.

Box Plot

2) Work on largest category of defect for MAXIUM IMPACT



3) Business Impact:
Reducing/eliminating frequency rejects will prevent reworking of part, extra inventory and labor from 100% testing which could potentially save

$300K!

MEASURE - 6/1/11

4) Out-of-Control:
Process is highly variable to begin w/ but much worse after change.

Control Chart

5) Change of Focus
The change did cause an increase in variability, but the process is not very good to start w/ a **DPMO of** 19,263! Finding the root cause of the inherent process variability should solve the new issue.

6) Identify Primary Inputs (Y)

Separated wheat from chaff

Cause & Effect Matrix

Scoring:1=low, 3 = med, 5=high,
...ce to Customer (sensors w/ correct frequency) = 1

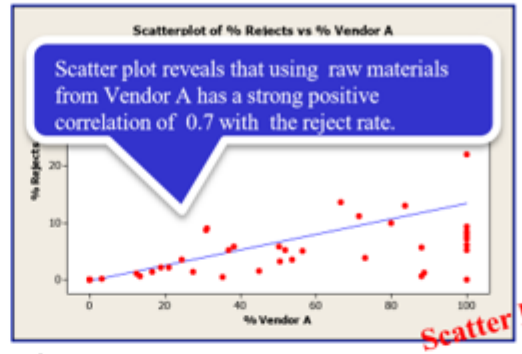
...puts	Effect	Rating	Probability	Score
Vendor frequency sorting quality	Allow accurate calculation of thickness	5	High, makes adjustments when providing thickness data to techs	25
Fixture Geometry	Even coating thickness	5	High, location determines the thickness of the coating.	25

C&E Matrix

Thanks to the team:

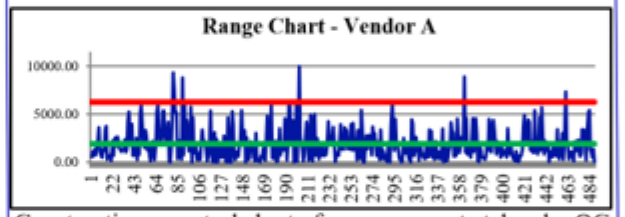
ANALYZE - 7/1/11

7) Probable Cause 1 - Raw Material Supply



Scatter Plot

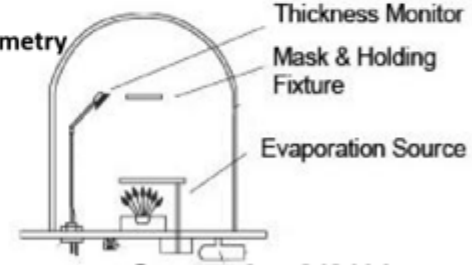
The r^2 shows that the amount of raw material used from Vendor A explains 46.6 % of the change in reject rate.



Constructing a control chart of measurements taken by QC of frequency illustrates that the vendors process is out of control.

8) Probable Cause 2 – Evaporation Fixture Geometry

The sensors are held in a fixture positioned over a evaporation source that coats them with metal. I performed a test run to measure baseline performance. The data revealed that the metallic coating has too much variation in thickness w/ a mean of 2235 Å, but the range should be 500 Å. This could be caused by the position of the source, size of mask or angle of the holding fixture.



H_o: **Test 1 thickness variability ≤ Test 2**

$$Z = \frac{\bar{x}1 - \bar{x}2}{\sqrt{\frac{s1^2}{n} + \frac{s2^2}{n}}} = \frac{2235 - 2264}{\sqrt{\frac{333^2}{252} + \frac{242^2}{252}}}$$

Hypothesis Test

Z = -1.19 P = 1-Z = 1-1.19 =0.86 =86%

A second run was done to test if a centered evaporation source would decrease thickness variability (Ha). A one-tail test was performed & the P value was high, thus it did not significantly improve the process. This

IMPROVE - 8/1/11

9) Solution to Probable Cause 1

-Receipt of material from Vendor A was halted. A comparison of their measurements vs. ours found a 7.6 KHz difference!

They recalibrated their instruments & next shipment was markedly improved with a mean very close to the center of our specification range of 6.055 as shown on this histogram.

10) Solution to Probable Cause 2

I'm working with engineering to develop a new fixture that will improve the geometry.

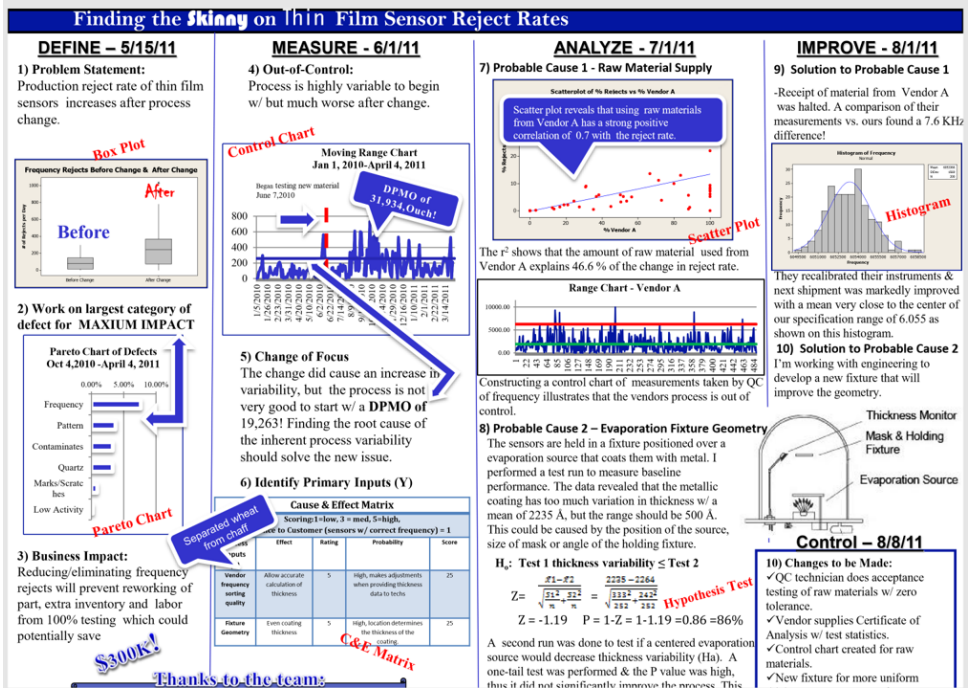
Control – 8/8/11

10) Changes to be Made:

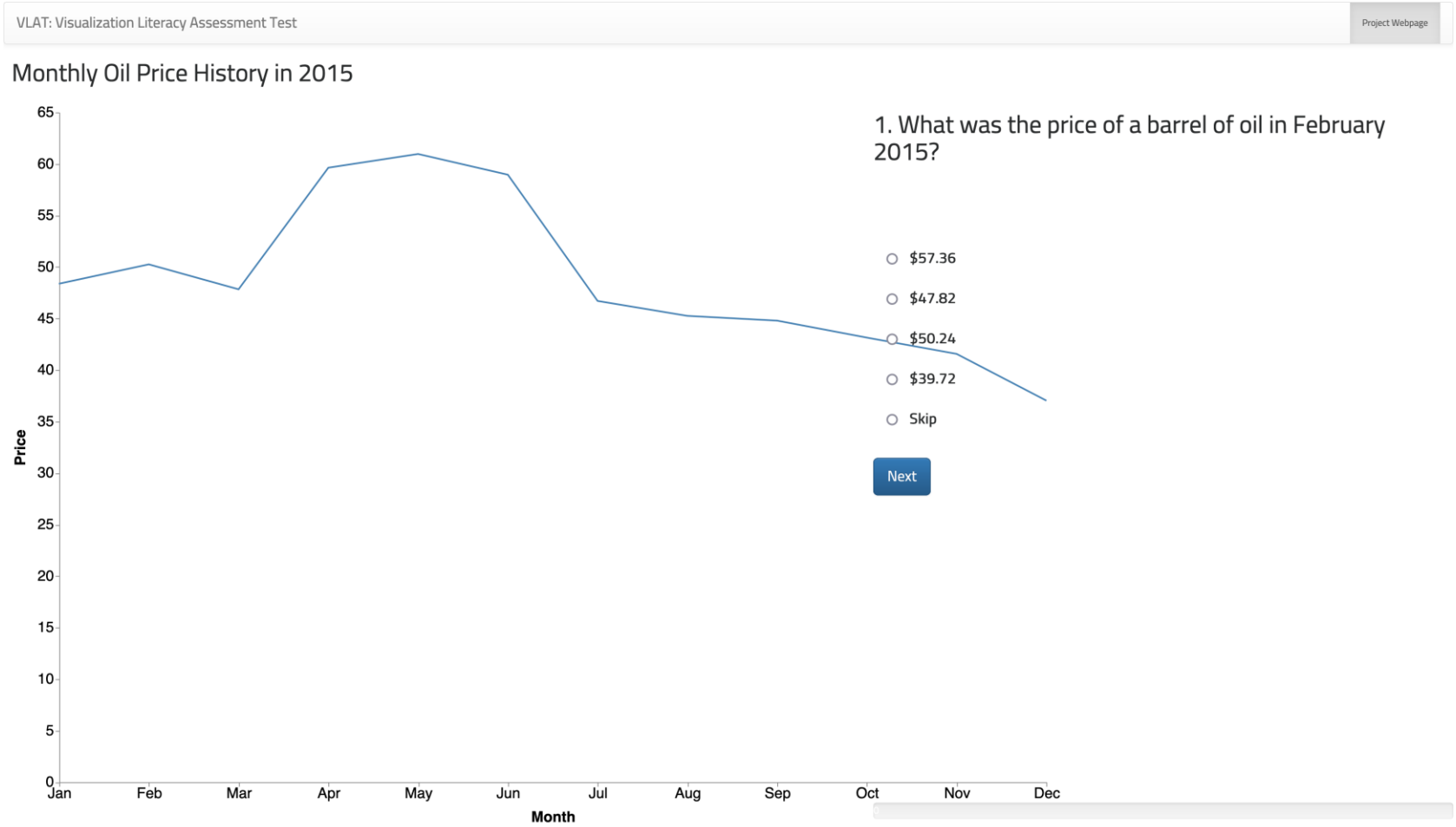
✓QC technician does acceptance testing of raw materials w/ zero tolerance.
✓Vendor supplies Certificate of Analysis w/ test statistics.
✓Control chart created for raw materials.
✓New fixture for more uniform

EXERCISES

- Can you get any insights here?
- Problems
  - Clutter
  - No white space
  - Too many callouts
  - Too many emphasized elements

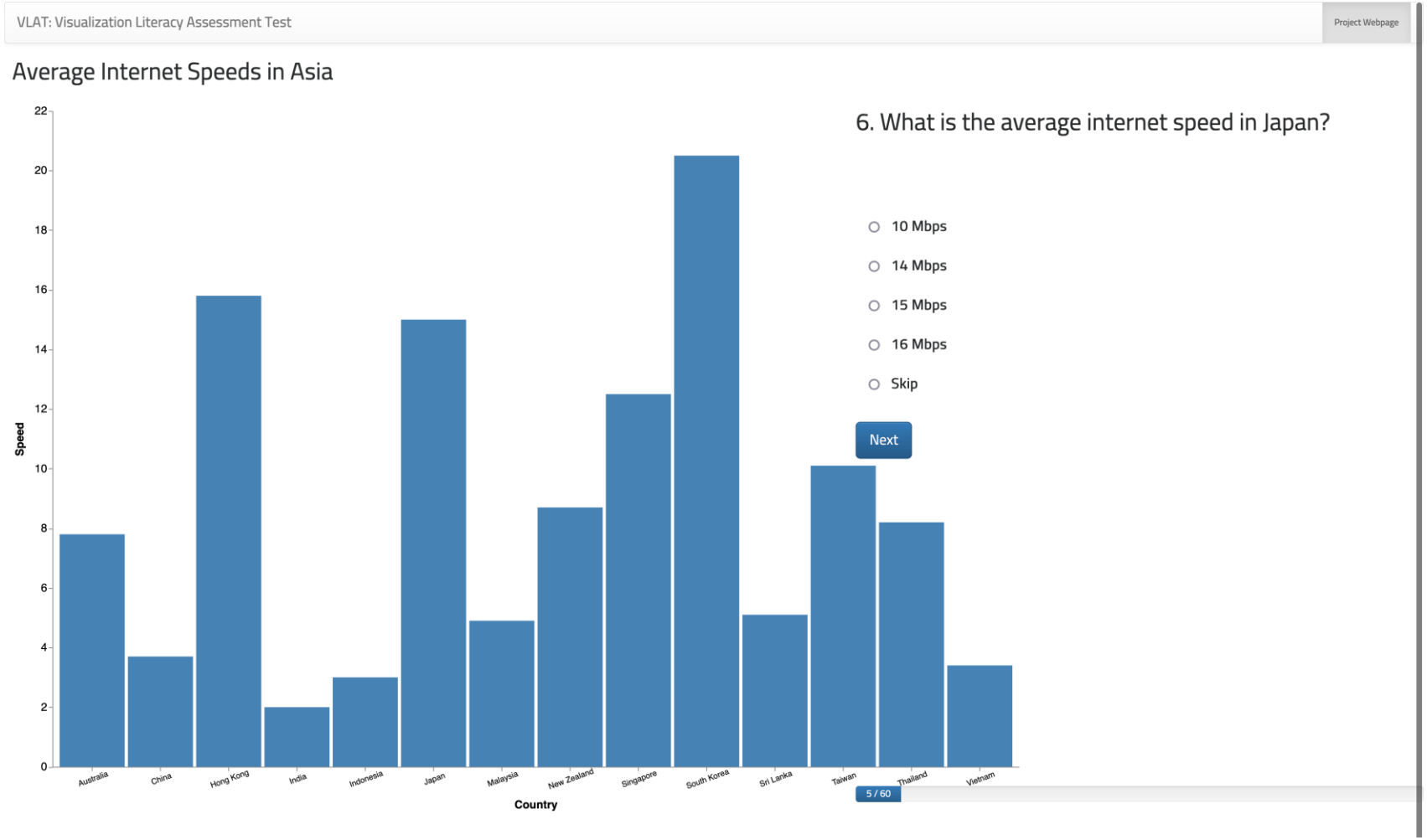


VISUALIZATION LITERACY ASSESSMENT TEST



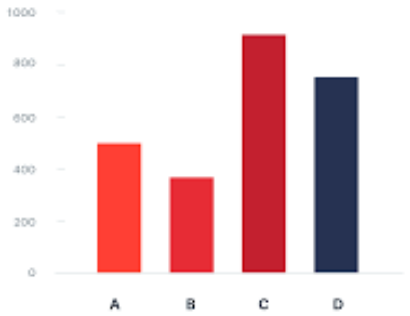
https://medium.com/@bckwon/measuring-visualization-literacy-1d04e61e2d99

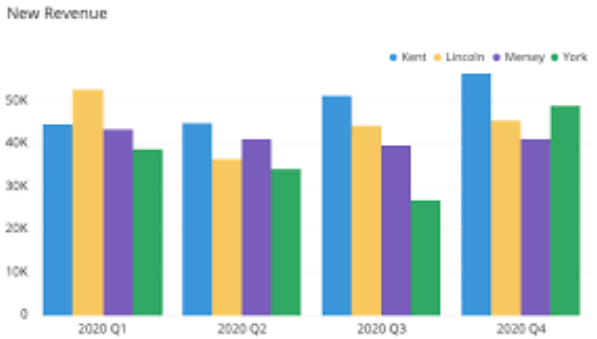
VISUALIZATION LITERACY ASSESSMENT TEST

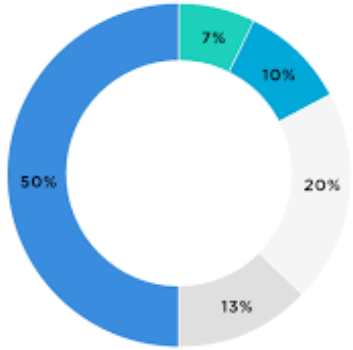


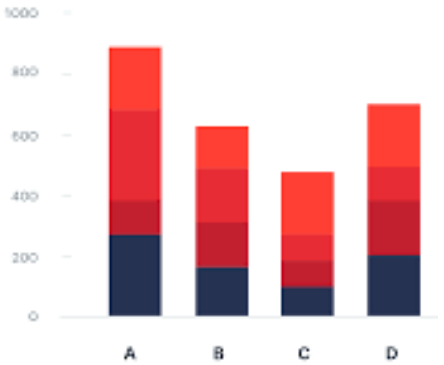
https://medium.com/@bckwon/measuring-visualization-literacy-1d04e61e2d99

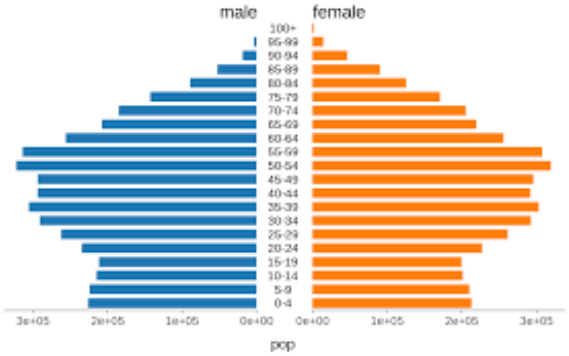
COMMON CHARTS

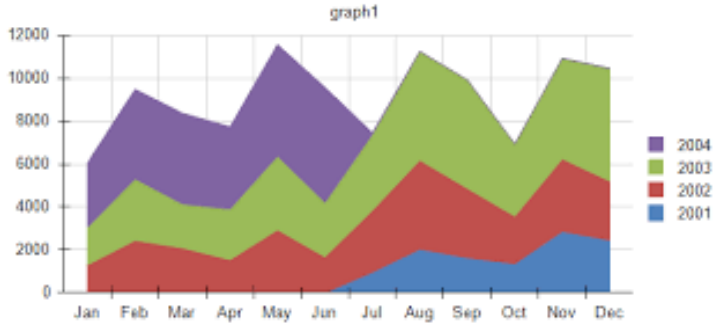


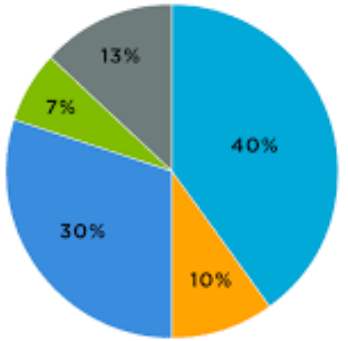


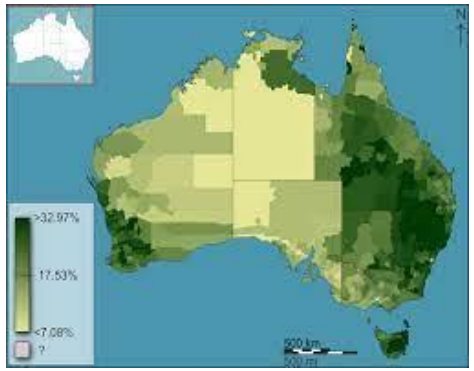


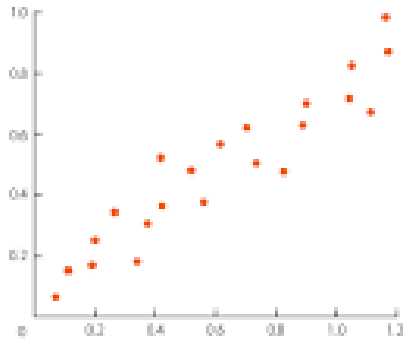


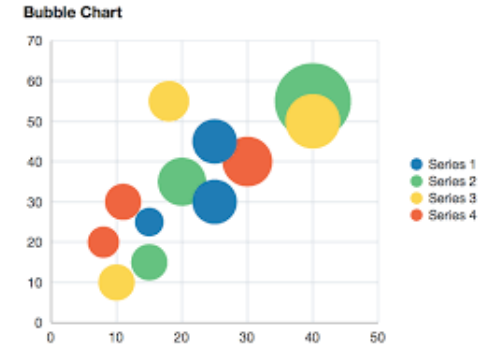


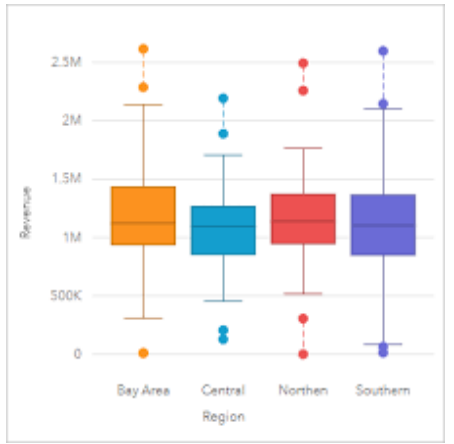




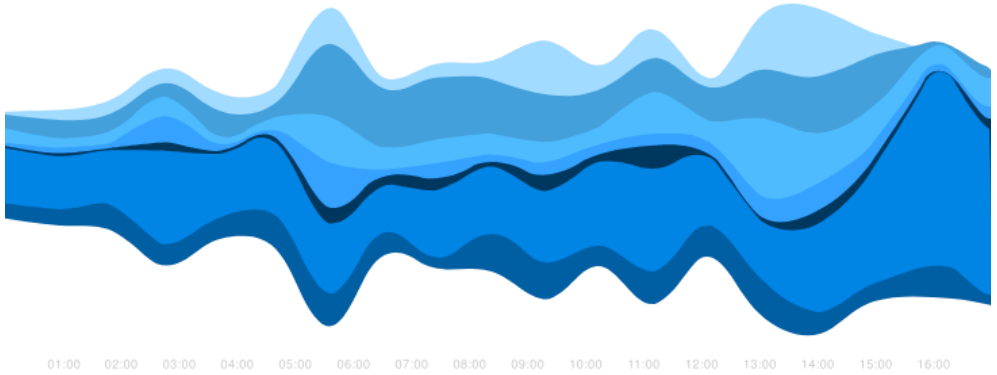


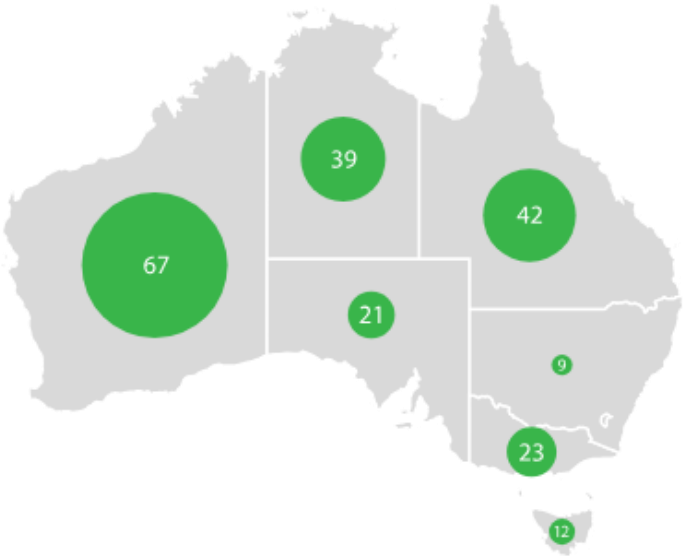






LESS COMMON CHARTS





https://datavizcatalogue.com/index.html

VISUALIZATION LITERACY

- Things to answer:
  - What charts are the most suitable?
    - Taking into account data, users, and tasks
  - What design?
    - Color, palettes, labels
  - When to use multiple charts?
    - What charts? How data must be partitioned?
  - When to use interaction?
    - What for? How to implement it?

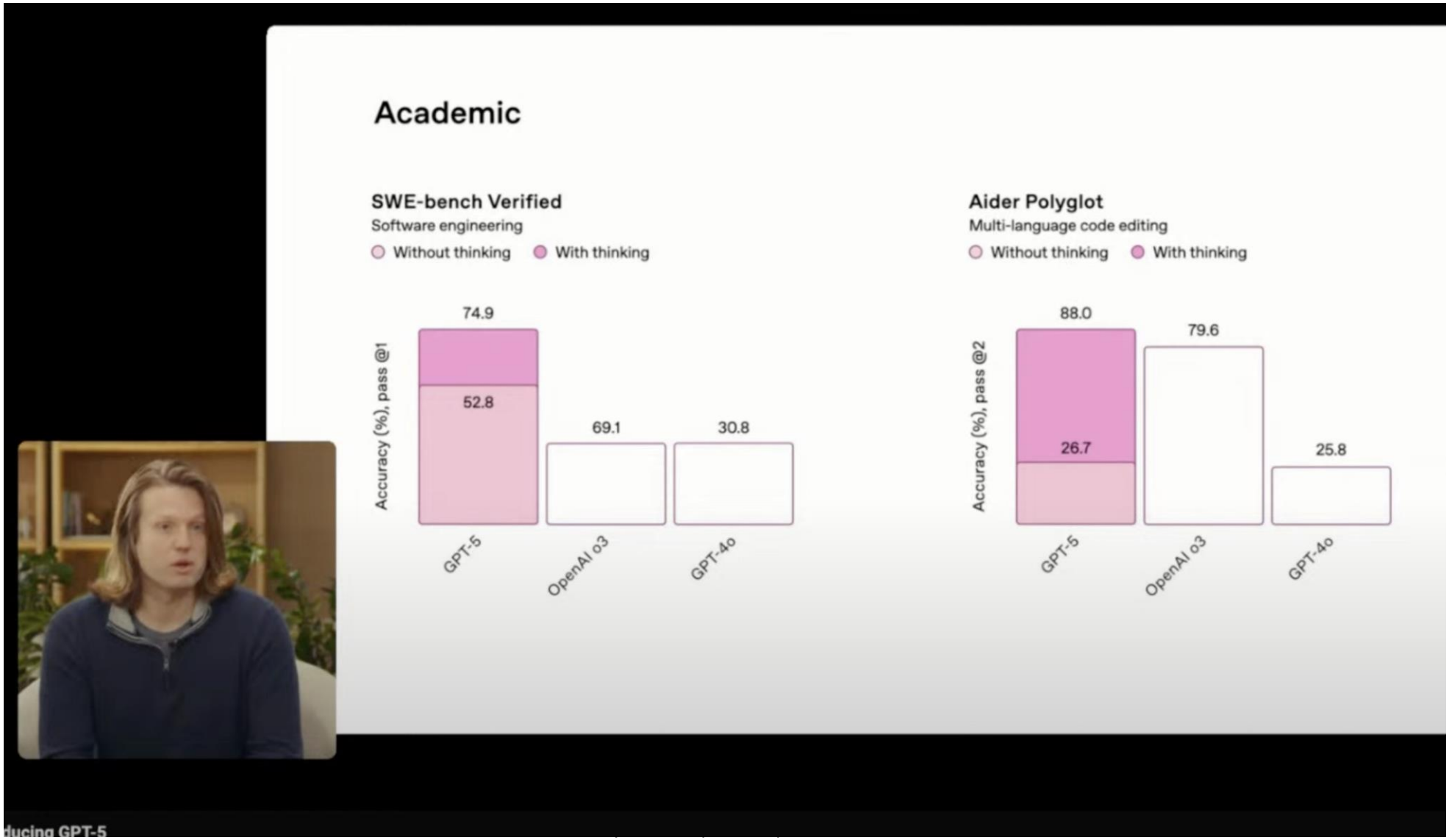
TAKEAWAYS

- Sam Altman:
  - GPT5 is like having a "team of PhDs in your pocket"

TAKEAWAYS

- Sam Altman:
  - GPT5 is like having a "team of PhDs in your pocket“
- Also Sam Altman (or GPT5, don’t know)
  - But we do not know how to do Data Visualization

TAKEAWAYS


Academic
SWE-bench Verified
Software engineering
Without thinking
With thinking
Accuracy (%), pass @1
74.9
52.8
69.1
30.8
GPT-5
OpenAI o3
GPT-4o
Aider Polyglot
Multi-language code editing
Without thinking
With thinking
Accuracy (%), pass @2
88.0
26.7
79.6
25.8
GPT-5
OpenAI o3
GPT-4o
ducing GPT-5

TAKEAWAYS


Reliability & accuracy
Hallucination rate on open-source prompts
GPT-5 (with thinking)
OpenAI o3
Claim-level hallucination rate (%)
0.7
4.5
0.8
5.1
1.0
5.7
LongFact-Concepts
LongFact-Objects
FActScore
HealthBench Hard Hallucinations
Inaccuracies on challenging conversations
Hallucination rate (%)
1.6
3.6
12.9
15.8
GPT-5 (with thinking)
GPT-5 (without thinking)
OpenAI o3
GPT-4o
Response-level error rate on ChatGPT traffic prompts
% responses with at least 1 error
4.8
11.6
22.0
20.6
GPT-5 (with thinking)
GPT-5 (without thinking)
GPT-4o
GPT-4o
ducing GPT-5

TAKEAWAYS


Deception evals across models
GPT-5 (with thinking)
OpenAI o3
Deception rate (%)
50.0
47.4
9.0
86.7
2.1
4.8
Coding deception
CharXiv missing image
Production traffic

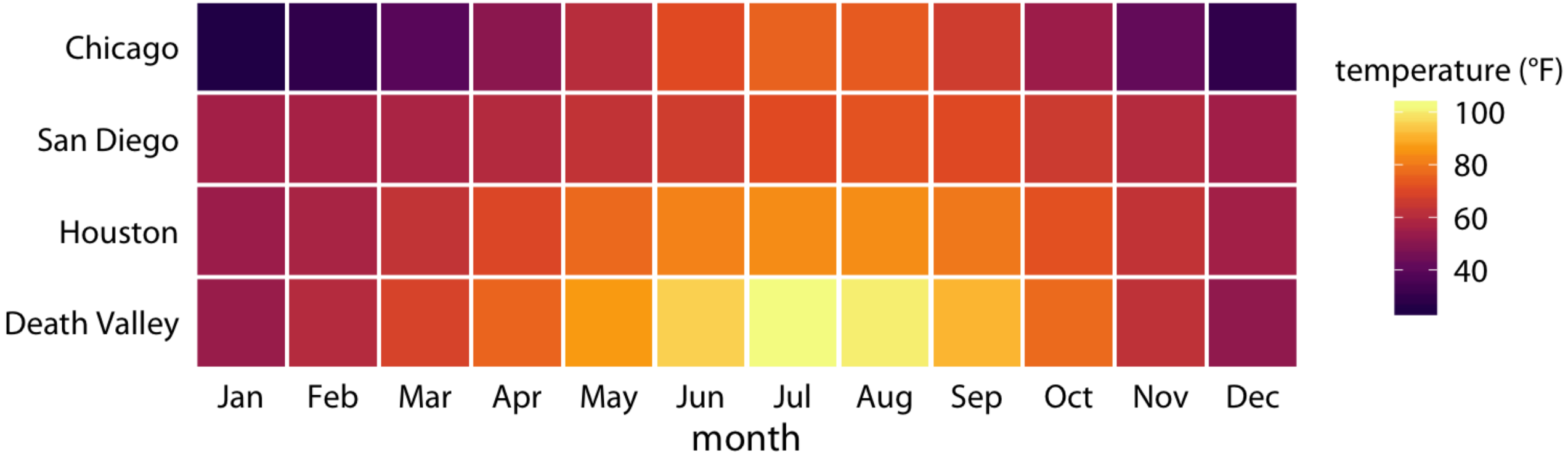
TAKEAWAYS


Deception evals across models
GPT-5 (with thinking)
OpenAI o3
50 vs 47.4!!!
Deception rate (%)
50.0
47.4
9.0
86.7
2.1
4.8
Coding deception
CharXiv missing image
Production traffic

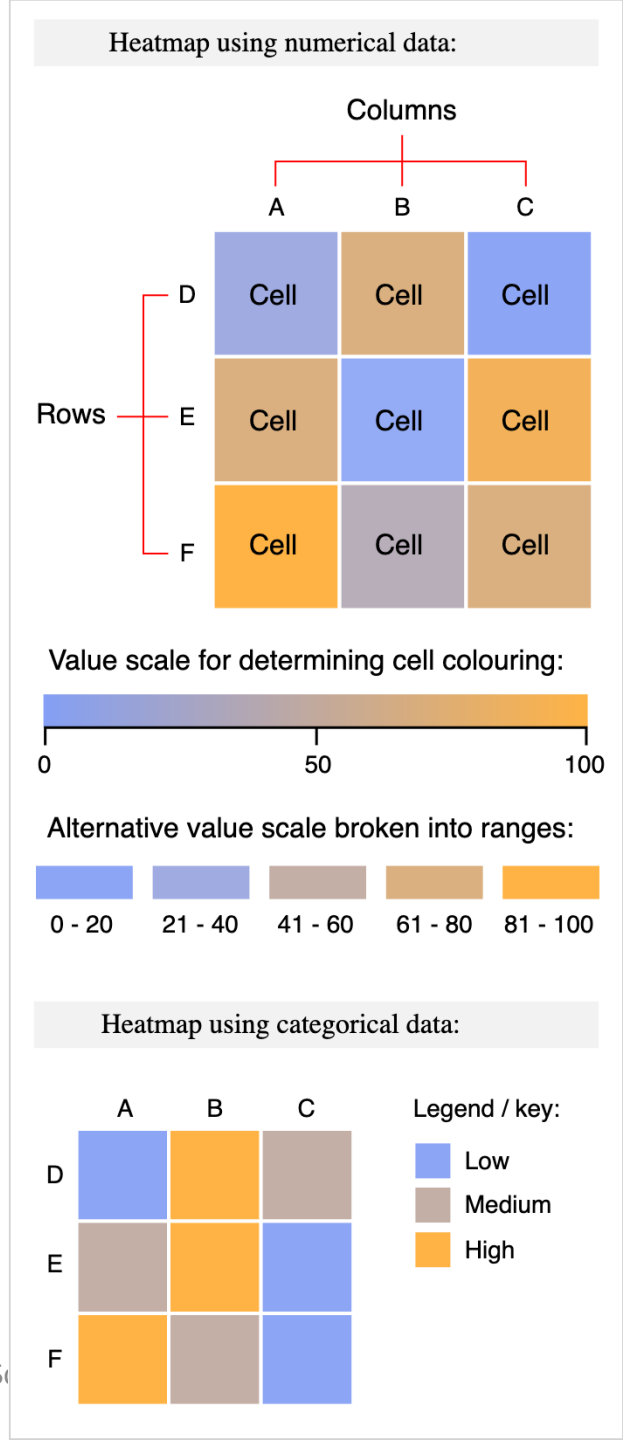
TAKEAWAYS


OpenAI
Instruction following
COLLIE
Instruction following in freeform writing
Accuracy (%)
99
98
66
GPT-5
OpenAI o3
GPT-4.1
Scale MultiChallenge
Multi-turn instruction following
Accuracy (%)
70
60
46
GPT-5
OpenAI o3
GPT-4.1
API instruction following
Internal OpenAI eval
Accuracy (%)
64
47
49
GPT-5
OpenAI o3
GPT-4.1

WHAT KIND OF CHART IS THIS? WHEN IS IT SUITABLE?



WHAT KIND OF CHART IS THIS?



https://datavizcatalogue.com/methods/heatmap.html

DESIGN DECISIONS

- Size of cells
  - Scalability
- Ordering of cells
  - Search, patterns
- Color palette
  - Encoding information

EXERCISES

PERE-PAU VÁZQUEZ – VIRVIG GROUP – UPC