

INFORMATION VISUALIZATION

Introduction

Petra Isenberg, Anastasia Bezerianos, Katerina Batziakoudi

petra.isenberg@inria.fr

anastasia.bezerianos@universite-paris-saclay.fr

aikaterini.batzikoudi@universite-paris-saclay.fr



COURSE INFO + EVALUATION

We'll use ecampus
also contact us is the ecampus forum

<https://ecampus.paris-saclay.fr/course/view.php?id=154627>
(title: 2025 [HCI/HCID] Interactive Information Visualization)



Go ahead and **self-register** !

GRADING SCHEME

Assignments: 50% of the grade

- Quizzes every week
- Assignments related to your project
- Peer feedback

Project: 50% of the grade

(of course calibrated by the difficulty and quality of the project / prototype)

GRADING

weekly

- Quizzes (easy!)
 - on the theoretical or practical topic of the previous week
 - will happen in the first 10min of the lecture (on e-campus, so have internet)
- Assignments, working towards a bigger project
 - check e-campus on how to submit them and when
 - in some cases you will need to provide Peer feedback to your colleagues
 - LATE PENALTIES: every hour of delay costs you marks!

READINGS

No specific readings necessary

Will announce readings on a per-lecture basis for those interested in learning more

PROJECT

We will work on a real problem with real data

Attempt to make an impact

AFTER TODAY YOU WILL...

have gained an overview of the research area

learned basic principles of data representation and interaction

understand how to ask research questions

Why

VISUALIZE DATA?

DATA EXPLORATION / INSIGHT

confirmatory analysis

- start with a hypothesis about the data
- confirm that it is true

focus of fully automated analysis methods

exploratory analysis

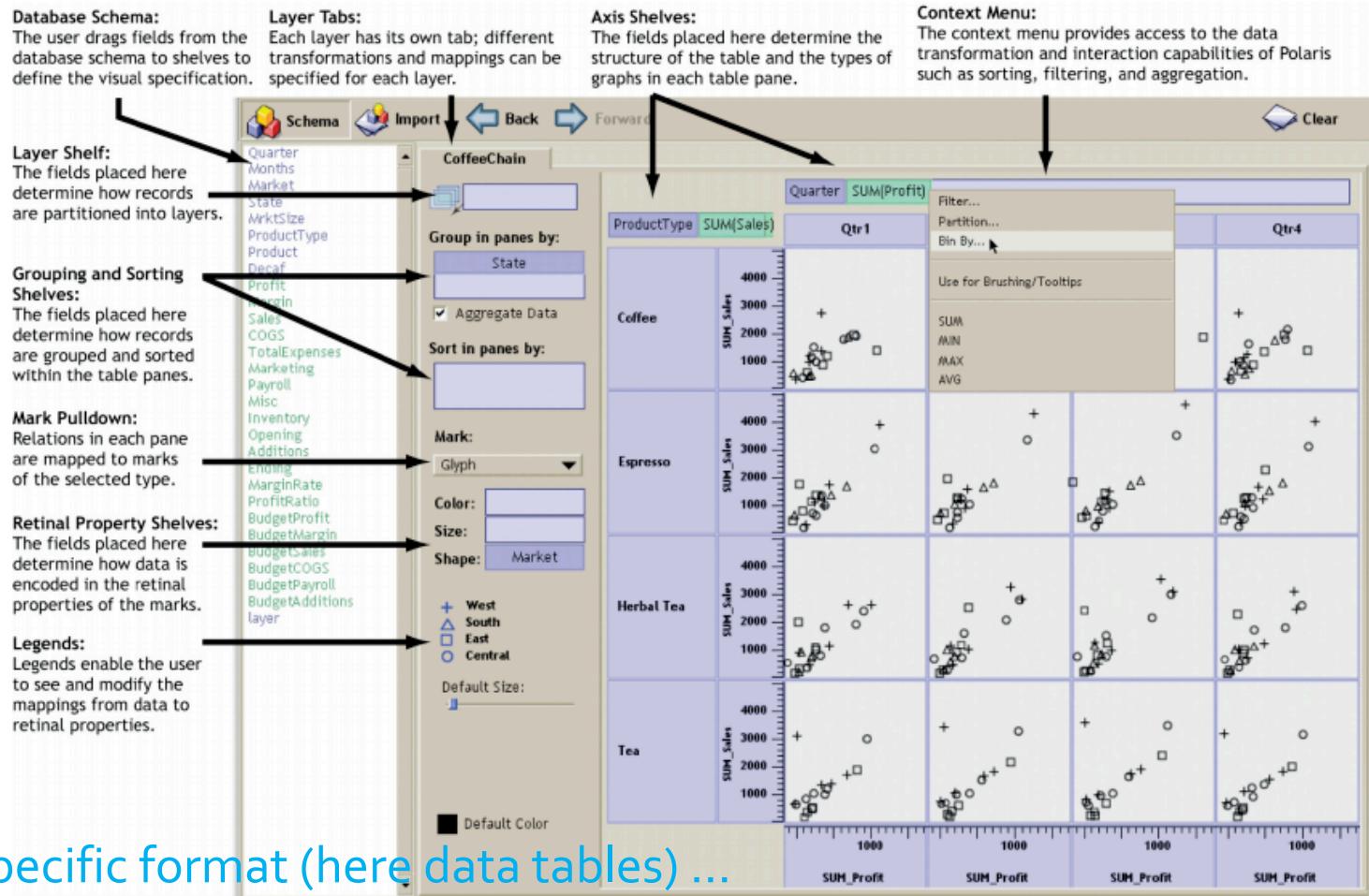
- likely no a-priori information about the data
- not sure about patterns and information present
- explore to create hypotheses & confirm later

focus of visualization

DATA EXPLORATION / INSIGHT

Often with expert / complex tools

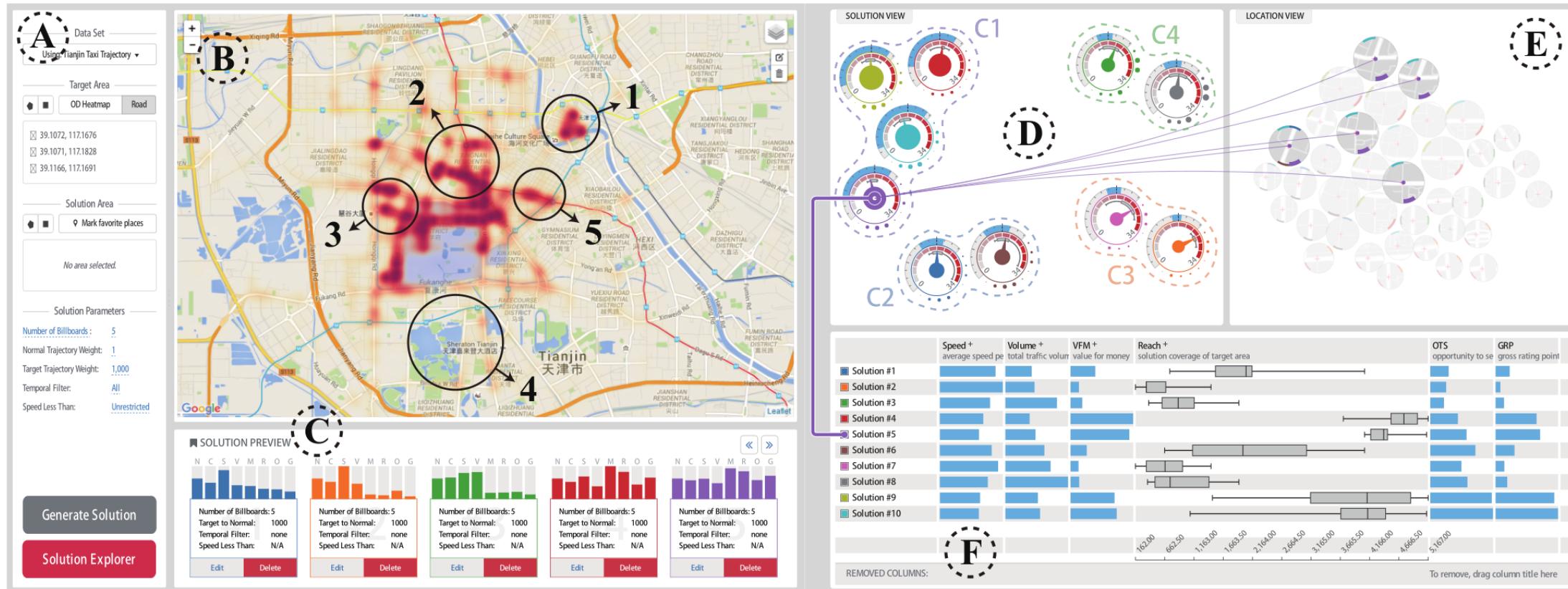
POLARIS



... Either for any data of a specific format (here data tables) ...

Taxi Trajectories for Selecting Billboard Locations

Dongyu Liu, Di Weng, Yuhong Li, Jie Bao, Yu Zheng, Huamin Qu, and Yingcai Wu



... Or for dedicated data / tasks

PERSUASION, EDUCATION, MOTIVATION



<https://www.visualcapitalist.com/map-press-freedom-around-the-world/>

"All the News
That's Fit to Print!"

The New York Times

VOL. CLXX... No. 58,976

© 2021 The New York Times Company

NEW YORK, SUNDAY, FEBRUARY 21, 2021

\$6.00

U.S. VIRUS DEATHS NEARING 500,000 IN JUST ONE YEAR

MORE THAN IN 3 WARS

Empty Spaces in Cities,
Towns, Restaurants,
Homes and Hearts

By JULIE BOSSMAN

CHICAGO — A nation mired by misery and loss is confronting a number that still has the power to shock: 500,000.

Roughly one year since the first known death by the coronavirus in the United States, an unfathomable toll is nearing — the loss of half a million people.

No other country has counted so many deaths in the pandemic. More Americans have perished from Covid-19 than on the battlefield of World War I, World War II and the Vietnam War combined.

The milestone comes at a hopeful moment: New virus cases are down sharply, deaths are slowing and vaccines are steadily being administered.

But there is concern about emerging variants of the virus, and it may be months before the pandemic is contained.

Each death has left untold numbers of mourners, a ripple effect of loss that has swept over towns and cities. Each death has left an empty space in communities

The Toll: America Approaches Half a Million Covid Deaths

Feb. 20, 2020: First report of a U.S. death, in Washington State

Each dot represents one death from Covid-19 in the U.S.



APRIL 24
51,380

MAY 27
100,422

Garland Faces Resurgent Peril Of Extremism

Oklahoma City Attack Shaped His Views

By MARK LEIBOVICH

WASHINGTON — Judge Merrick B. Garland always made a point of wearing a cap and tie when he surveyed the wreckage at the site of the 1995 Oklahoma City bombing, the worst domestic terrorist attack in American history.

He had been dispatched from Washington to oversee the case for the Justice Department, and he told colleagues that he viewed his daily uniform as a gesture of respect for a community left devastated after Timothy J. McVeigh placed a 7,000-pound bomb in a Ryder truck and blew up the Alfred P. Murrah Federal Building, killing 168 people, including 19 children.

"It really looked like a war zone," Judge Garland said in recalling the destroyed and still-smoldering building, part of an era history he participated in for the Oklahoma City National Memorial and Museum. "The site was lit up like a sun, like the middle of the day." The worst part, he said, was seeing the demolished day care center. "There was nothing there," he said. "It was just a big empty concrete." His own daughters were 4 and 2 at the time.

The Oklahoma City case, he lat-

— 96 days later, 50,000 deaths had been reported.

— It took 33 days to reach the next 50,000 deaths.

— 63 days to reach the next 50,000 deaths.

STORMS EXPOSING A NATION PRIMED FOR CATASTROPHE

CLIMATE CHANGE WRATH

Unprepared for Threats

Facing Power Grids,
Water and Roads

This article is by Christopher Flavelle, Reed Plawar and Hiroko Tabuchi.

Even as Texas struggled to restore electricity and water over the past week, signs of the risks posed by increasingly extreme weather to America's aging infrastructure were cropping up across the country.

The continent-spanning winter storms triggered blackouts in Texas, Oklahoma, Mississippi and several other states. One-third of oil production in the nation was halted. Drinking-water systems in Ohio were knocked offline. Road networks nationwide were paralyzed and vaccination efforts in 20 states were disrupted.

The crisis carries a profound warning. As climate change brings more frequent and intense storms, floods, heat waves, wildfires and other extreme events, it is placing growing strain on the foundations of the country's economy: its network of roads and railways, drinking-water systems, power plants, electrical

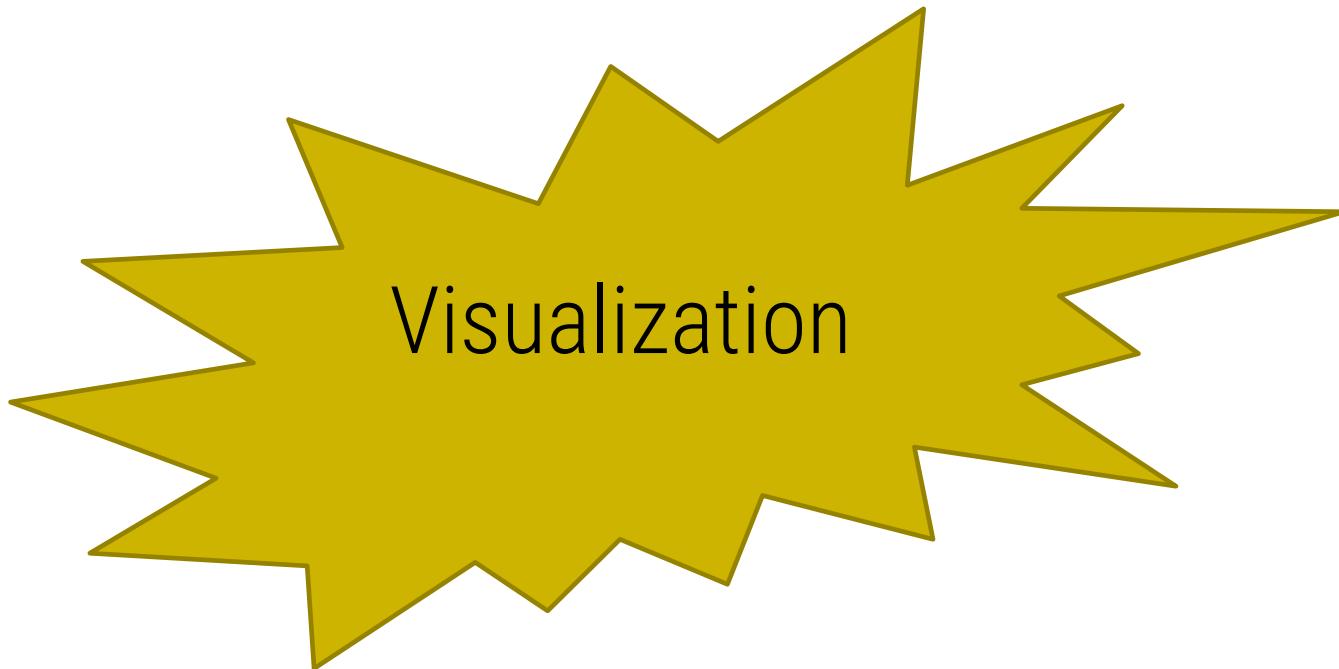


BIG RESEARCH QUESTION IN VISUALIZATION

how can humans effectively access data?

- understand its structure?
- make comparisons?
- make decisions?
- gain new knowledge?
- convince others?
- ...

OK, BUT WHY...



EXAMPLE

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

Raw Data from Anscombe's Quartet

[Source: Anscombe's quartet, Wikipedia]

STATISTICAL ANALYSIS

For all four columns, the statistics are identical

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

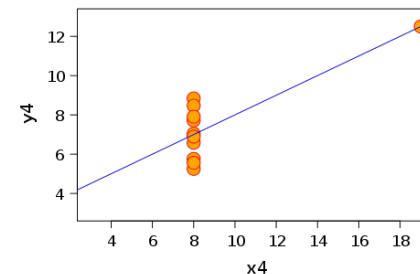
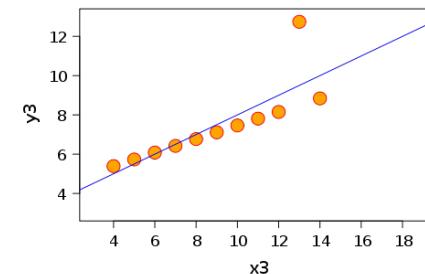
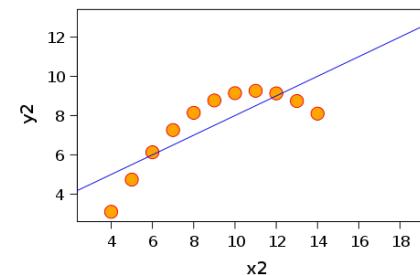
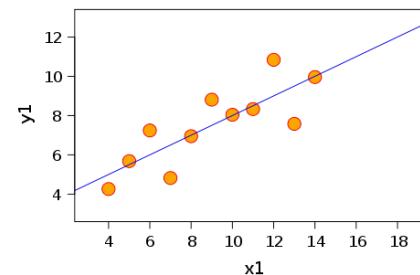
Mean of x	9.0
Variance of x	11.0
Mean of y	7.5
Variance of y	4.12
Correlation between x and y	0.816
Linear regression line	$y = 3 + 0.5x$

[Source: Anscombe's quartet, Wikipedia]

VISUAL REPRESENTATION OF THE DATA

Visual representation reveals a different story

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

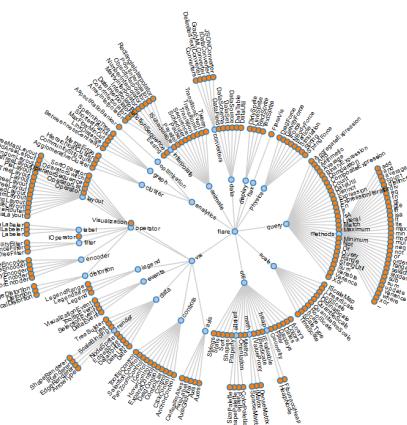


[Source: Anscombe's quartet, Wikipedia]

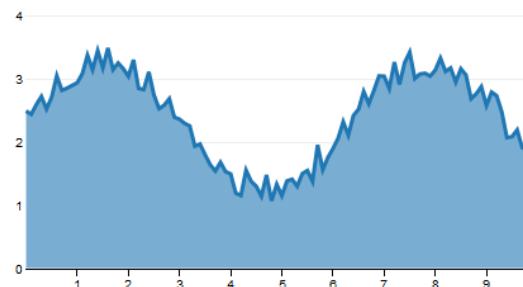
Why visual data representations?

- Vision is our most dominant sense
- We are very good at recognizing visual patterns
- We need to see and understand in order to explain, reason, and make decisions

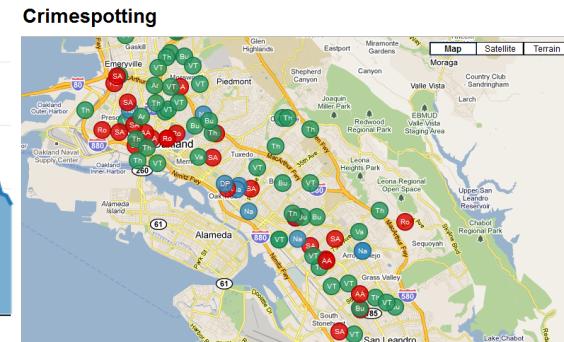
common examples:



graphs / hierarchies



charts



maps

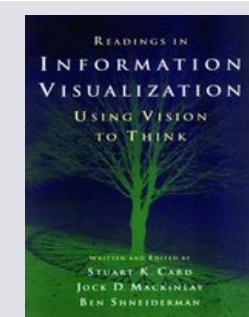
all examples from: <http://vis.stanford.edu/protovis/>

Visualization

- Create visual representation
- Includes interaction

Official Definition:

*The use of computer-supported, interactive,
visual representations of data
to amplify cognition.*
[Card et al., 1999]



Hint: but there is more to it

Functions of Visualizations

- Recording information
 - Tables, blueprints, satellite images
- Processing information
 - needs feedback and interaction
- Presenting information
 - share, collaborate, revise
 - for oneself, for one's peers, and to teach
- Seeing the unseen

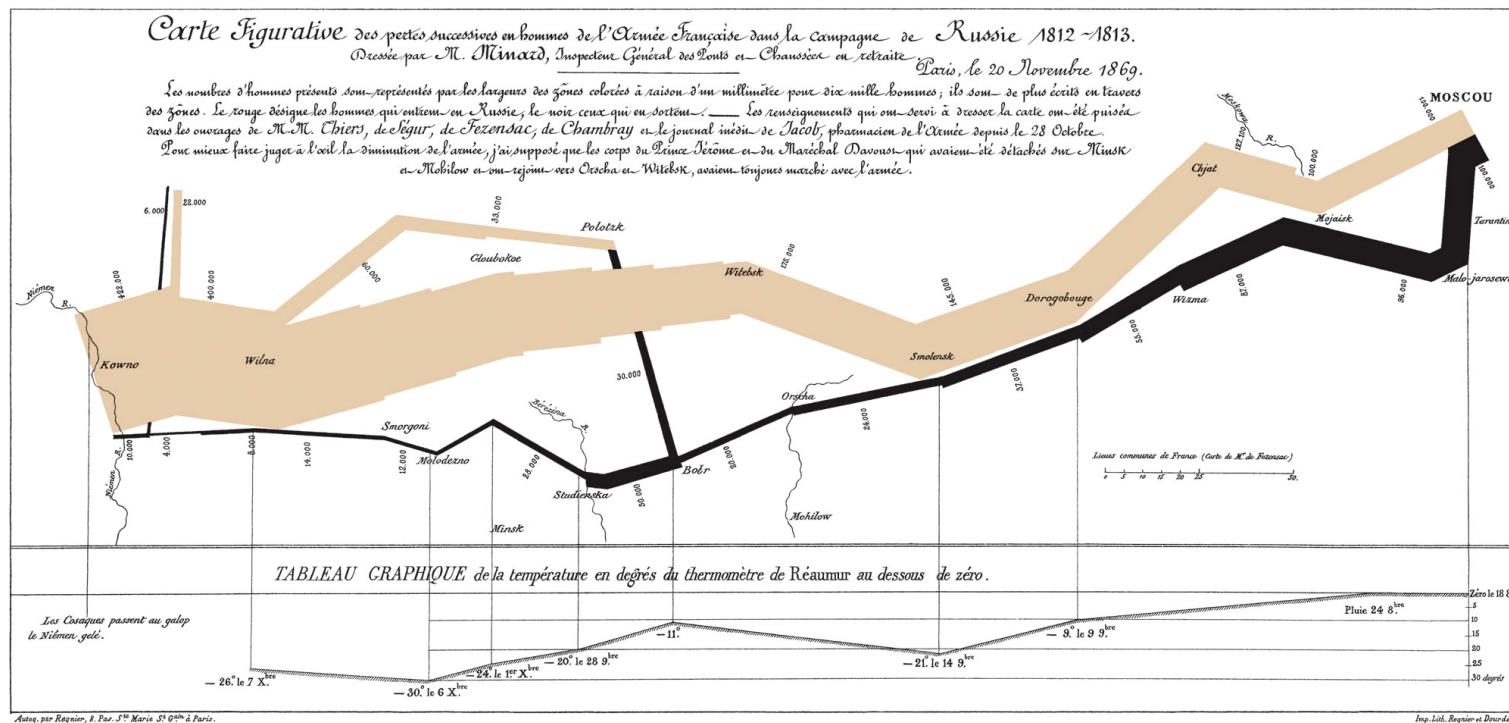
HISTORICAL EXAMPLES

Visualization of data has been practiced for hundreds of years...

NAPOLEON'S MARCH ON MOSCOW

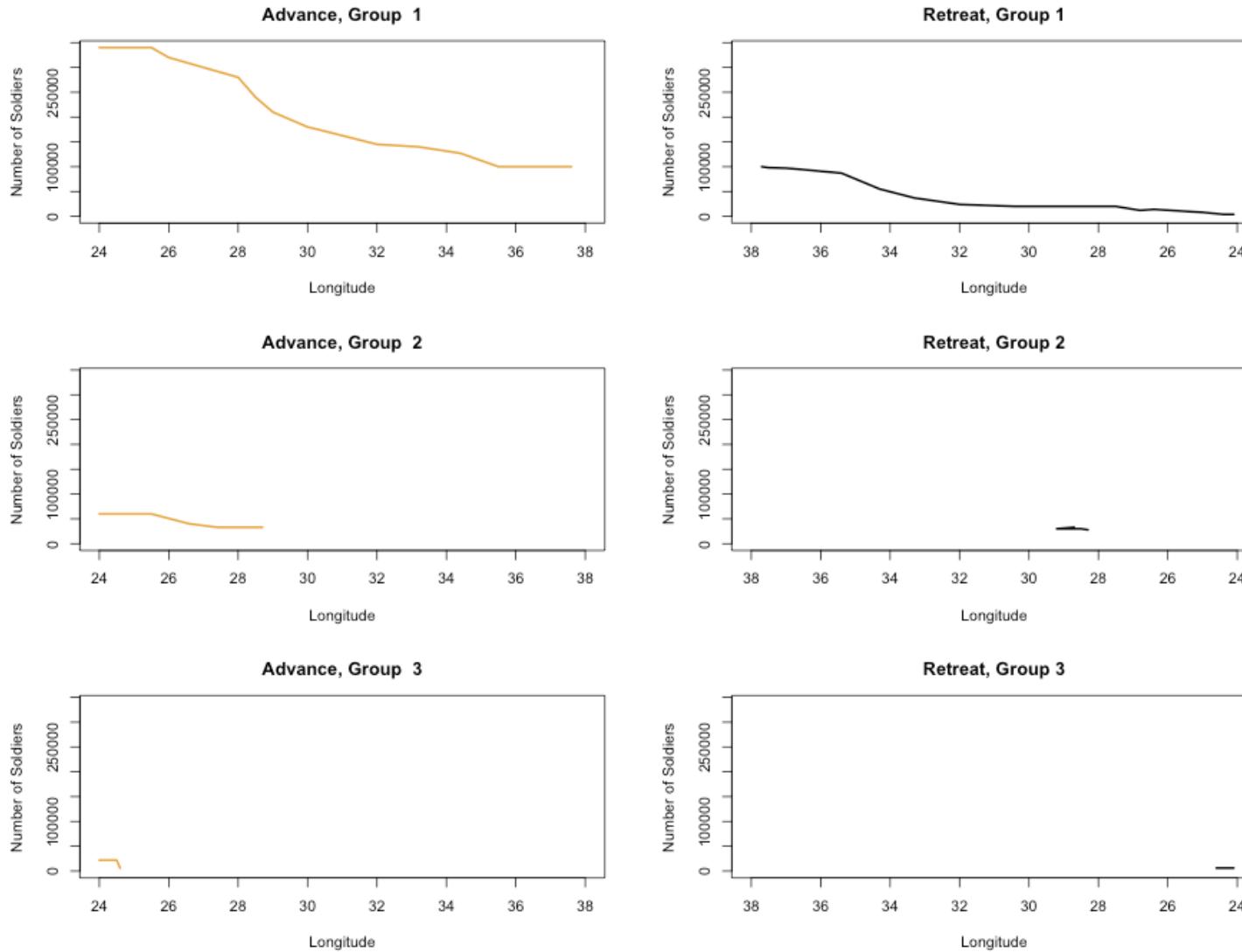
By C. Minard. Named the best statistical graphic ever drawn (by Edward Tufte)

- Includes: spatial layout linked with stats on: army size, temperature, time
- Tells a story in one overview



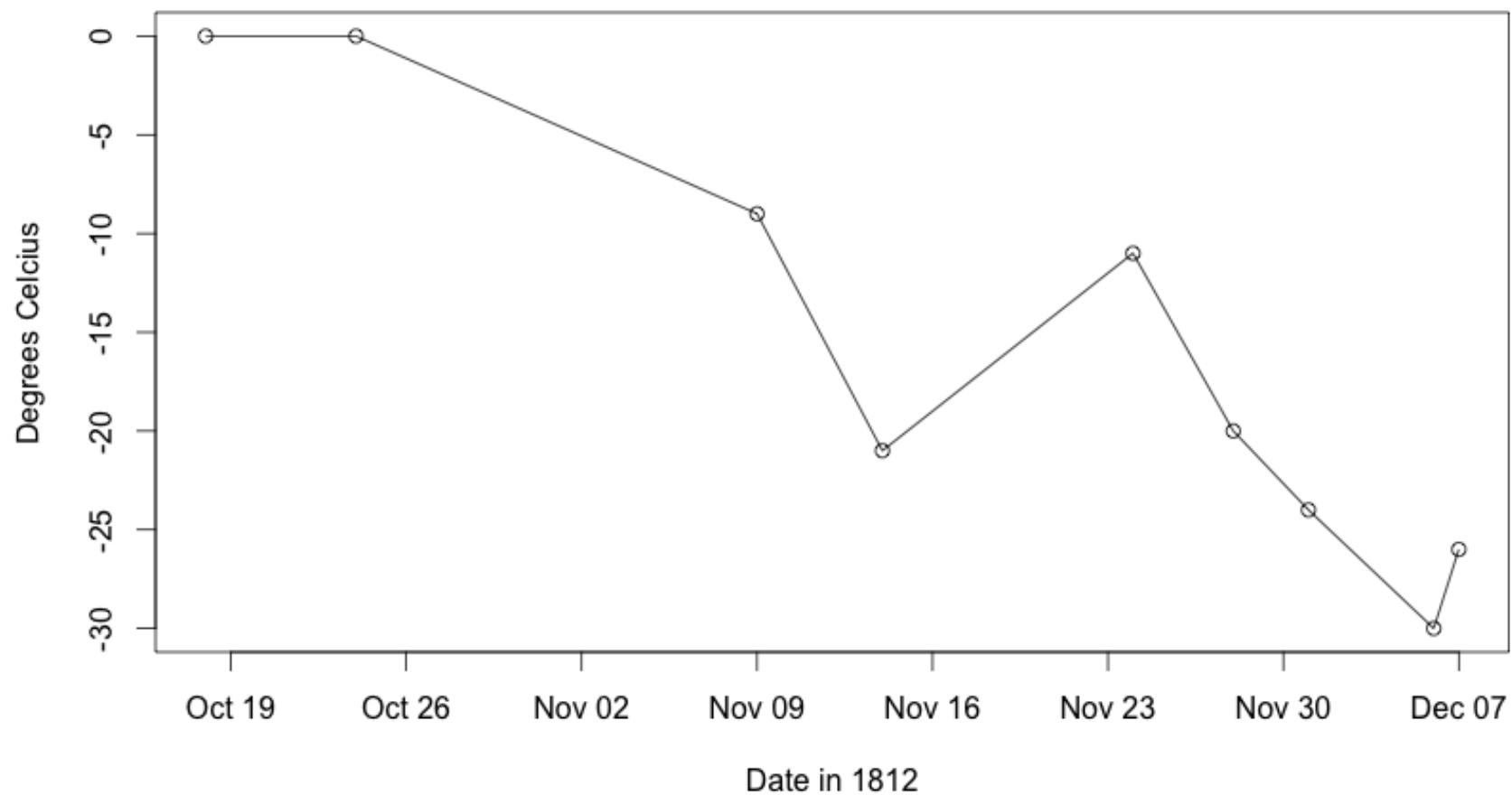
Charles Minard, 1869

More info: The Visual Display of Quantitative Information (Tufte)



<https://thoughtbot.com/blog/analyzing-minards-visualization-of-napoleons-1812-march>

Temperature During The Retreat



<https://thoughtbot.com/blog/analyzing-minards-visualization-of-napoleons-1812-march>



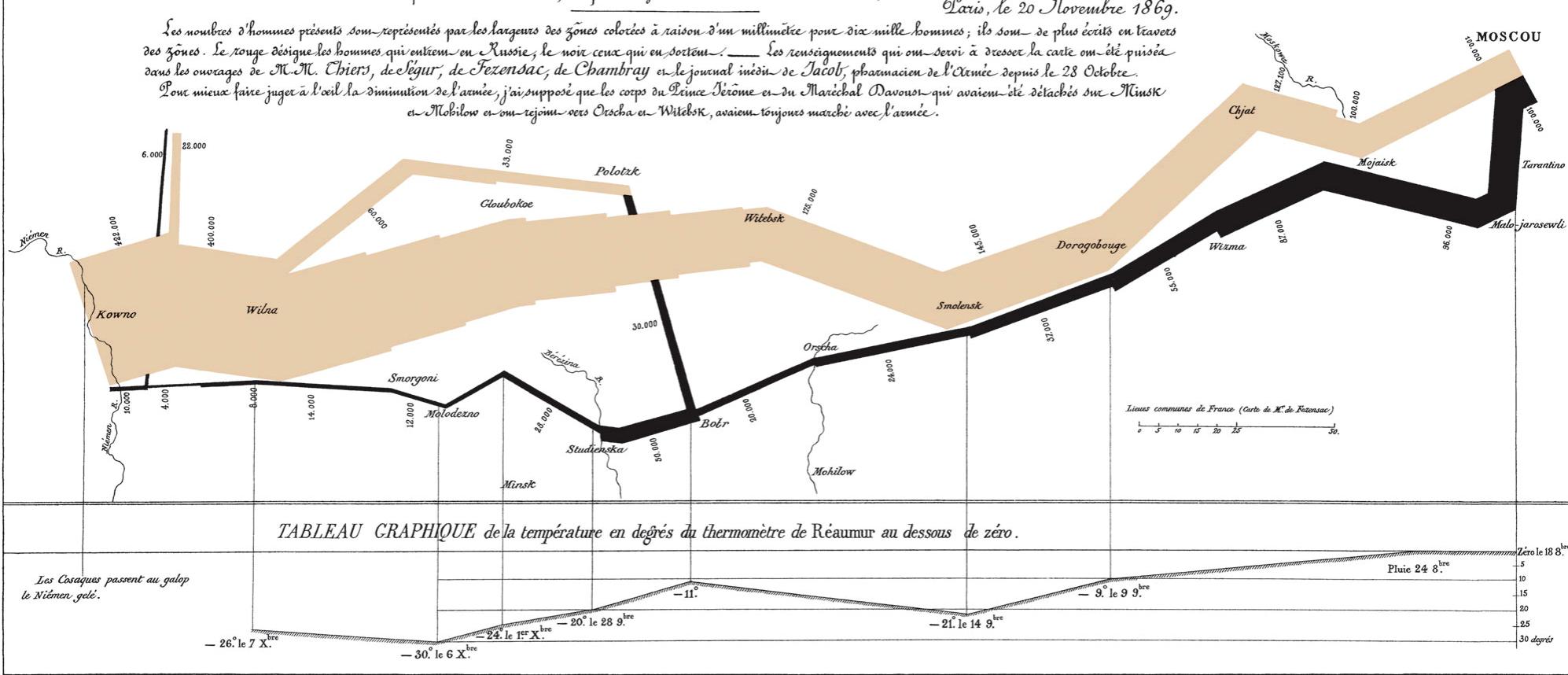
Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

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 zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. 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. Chiers, de Clément, 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 Davout qui avaient été détachés sur Minsk et Mohilow et qui rejoignirent l'armée vers Orsha et Wilek, avaient toujours marché avec l'armée.

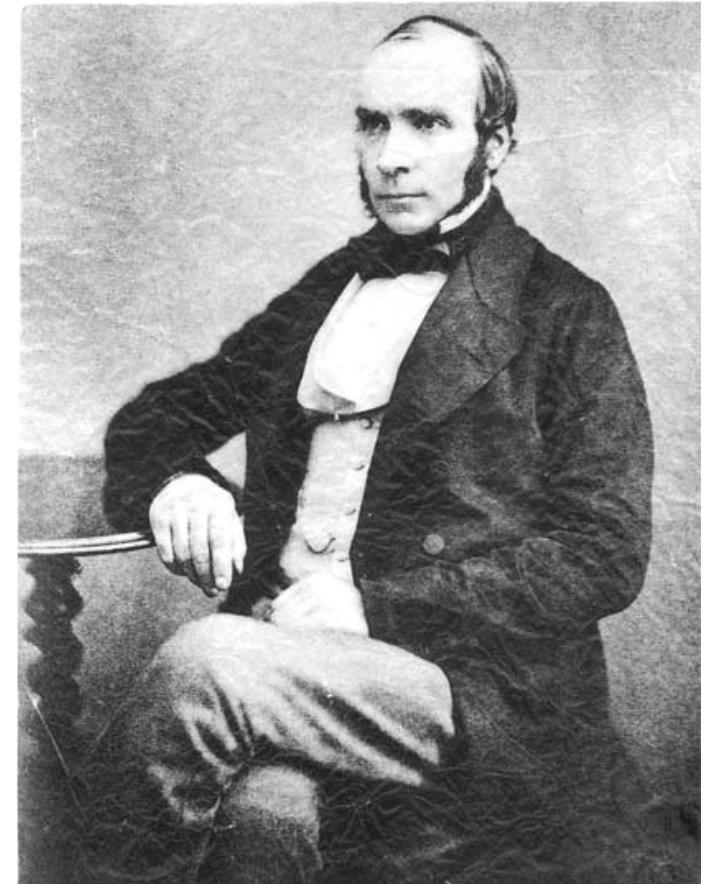


Autog. par Regnier, 8. Pas. S^{te} Marie S^{te} G^{me} à Paris.

Imp. Lith. Regnier et Dourdet.

THE BROADWAY STREET PUMP

- In 1854 cholera broke out in London
 - 127 people near Broad Street died within 3 days
 - 616 people died within 30 days
- “Miasma in the atmosphere”
- Dr. John Snow was the first to link contaminated water to the outbreak of cholera
- How did he do it?
 - he talked to local residents
 - identified a water pump as a likely source
 - used maps to illustrate his theory
 - convinced authorities to disable the pump



More info here: http://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak



30

JOHN SNOW, 1854

TODAY

Visualizations are done by many people...



KANTAR Information is Beautiful Awards 2019

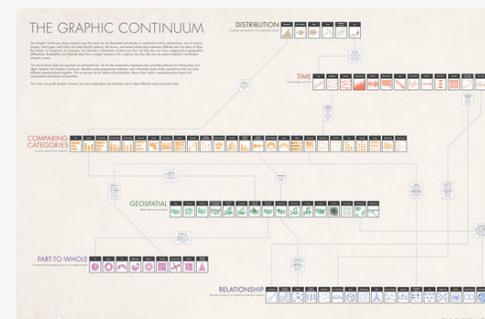
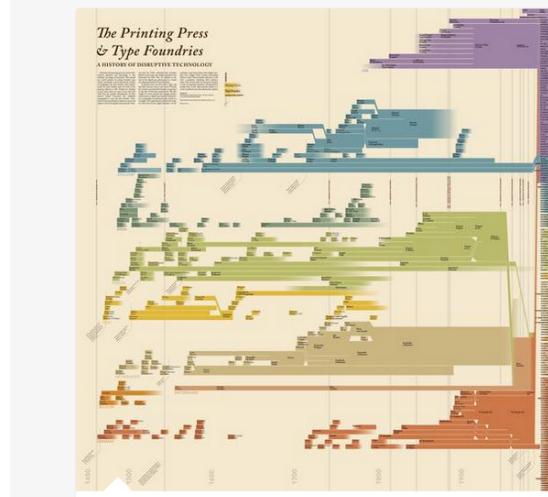
About News Awards Entry Showcase Sponsor Attend the Awards ceremony

Awards Challenges

2017 2016 2015 2014 2013 2012

All Shortlist winners Gold Silver Bronze Rising Star Outstanding Individual Best-Non-English-Language Outstanding Outfit Student Community Most Beautiful

All Arts, Entertainment & Culture Breaking News Humanitarian Leisure, Games & Sport Maps, Places & Spaces News & Current Affairs People, Language & Identity Politics & Global Visualization & Information Design Science & Technology Unusual



Project Ukko - Seasonal Wind Predictions for the Energy Sector

Weather forecasts predict future wind conditions only in the range of weeks. Climate predictions look at big changes over years and decades. However, for energy traders, wind farm managers and...

Enthusiasts, freelancers, ...

Where Have All the Houses Gone?

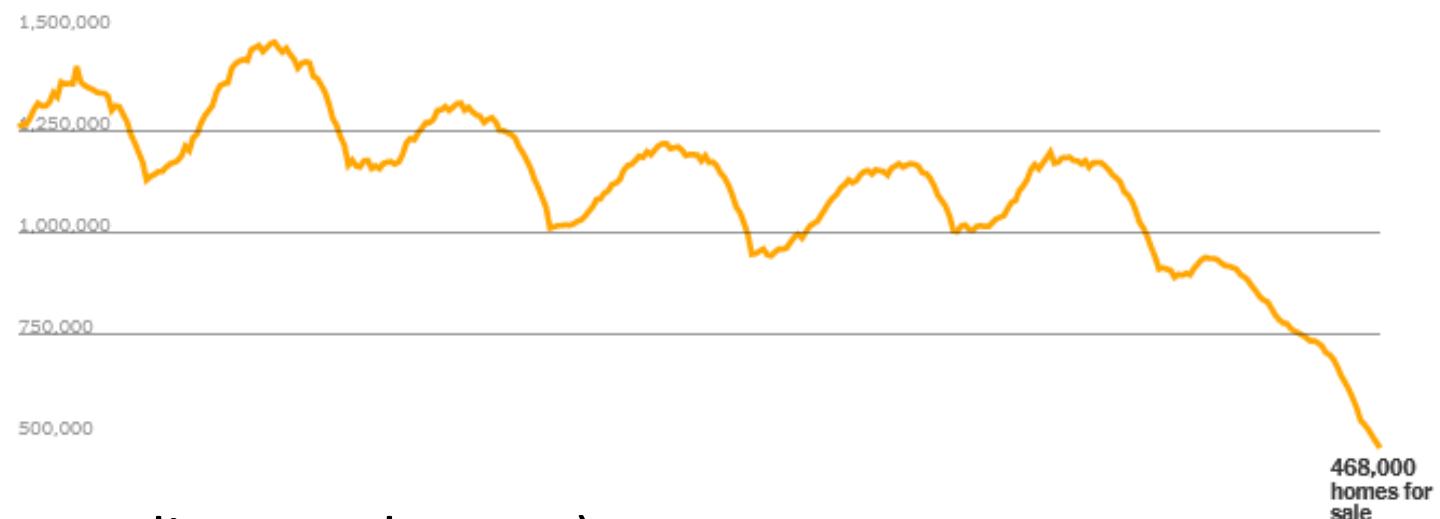
The inventory of homes for sale is startlingly low. The pandemic is part of the reason, but it's not the whole story.



By [Emily Badger](#) and [Quoctrung Bui](#)

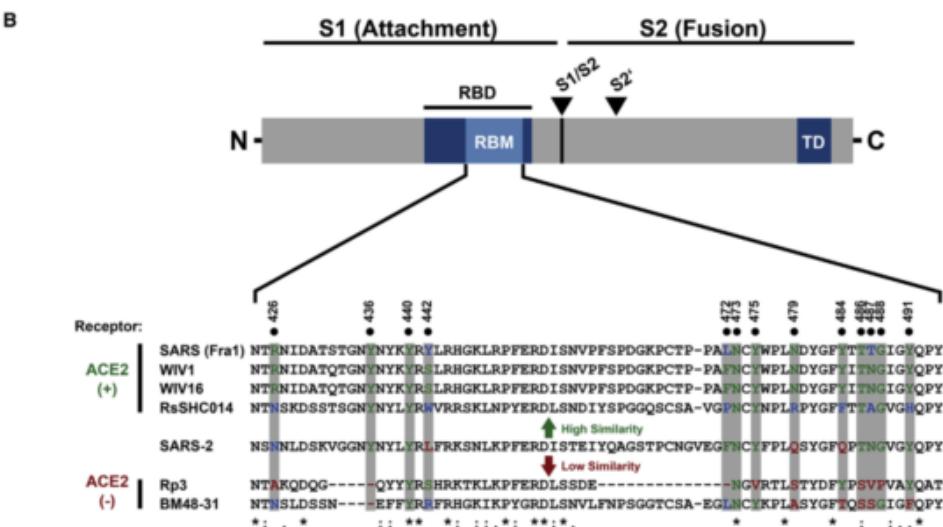
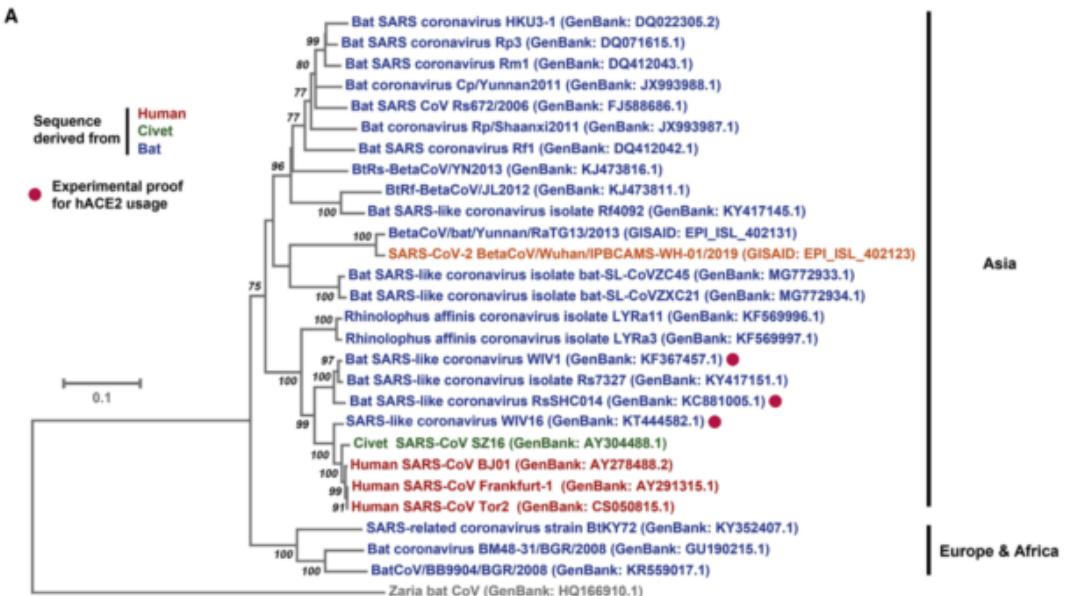
Published Feb. 26, 2021 Updated March 2, 2021

The number of homes for sale nationally has plummeted



Professionals (journalists, analysts, ...)

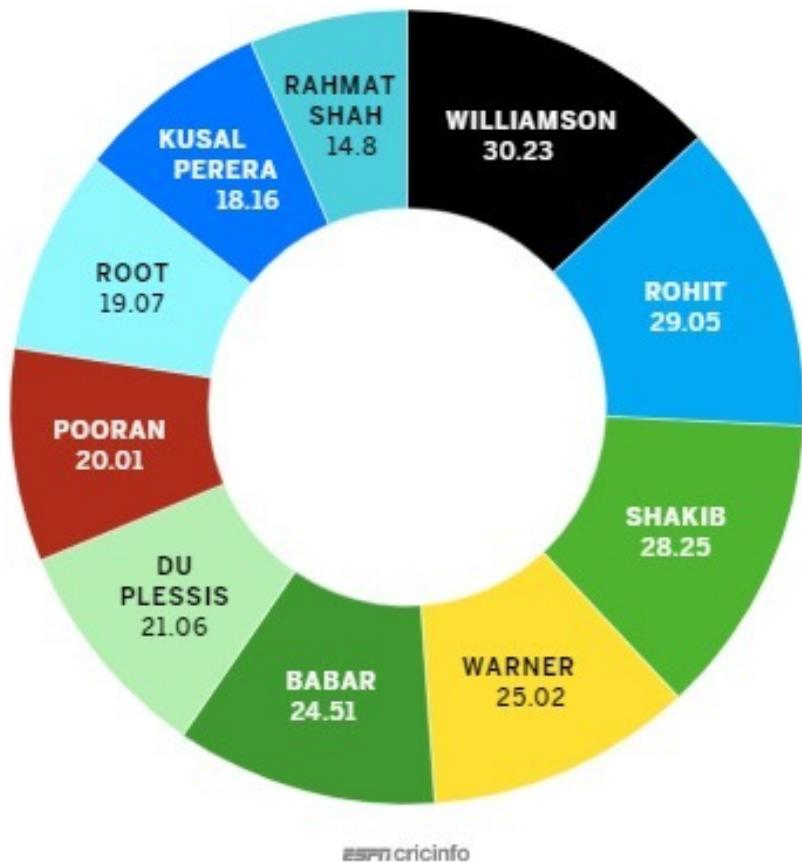
Markus Hoffmann, Hannah Kleine-Weber, Simon Schroeder, Nadine Krüger, Tanja Herrler, Sandra Erichsen, Tobias S. Schiergens, Georg Herrler, Nai-Huei Wu, Andreas Nitsche, Marcel A. Müller, Christian Drosten, Stefan Pöhlmann,
SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor, Cell,
Volume 181, Issue 2,
2020,



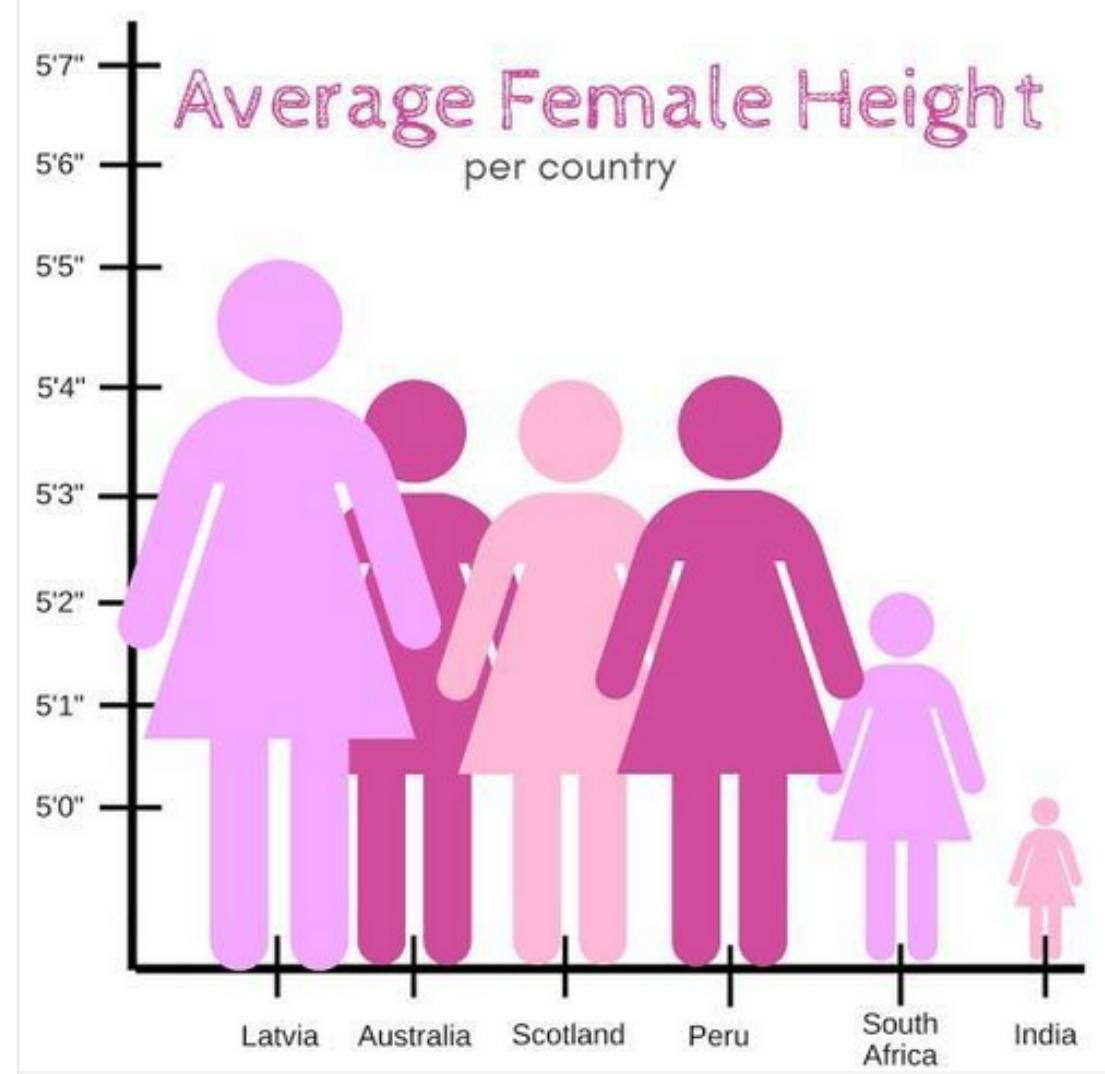
Researchers

THE WORLD CUP'S BIG GUNS

% OF TEAM'S RUNS SCORED BY TOP SCORER



People who shouldn't



RESOURCES

- Visualization conferences
 - ieevis.org
- Blogs
 - <http://eagereyes.org/>
 - <http://flowingdata.com/>
 - <http://www.informationisbeautiful.net/>
 - <https://www.visualisingdata.com/blog/>
 - <https://pudding.cool/>
 - <https://junkcharts.typepad.com/>
 - <https://badvisualisations.tumblr.com/>
- Catalogues
 - <https://datavizcatalogue.com/>
 - <https://www.treevis.net/>
 - <https://textvis.lnu.se/>
 - <https://browser.timeviz.net/>
 - <https://trustmlvis.lnu.se/>
 - <http://dataphys.org/list/> ... and many more
- Books
 - Textbooks
 - Readings in Information Visualization: Using Vision to Think (a bit old now but good intro)
 - Information Visualization (Robert Spence – a light intro, I recommend as a start)
 - Information Visualization Perception for Design (Colin Ware, focused on perception and cognition)
 - Interactive Data Visualization: Foundations, Techniques, and Applications (Ward et al.)
 - Visualization Analysis and Design (Tamara Munzner, most recent book)
 - Examples
 - Beautiful Data (McCandless)
 - Now You See it (Few)
 - Tufte Books: Visual Display of Quantitative Information (and others)
 - ... (many more, ask me for details)

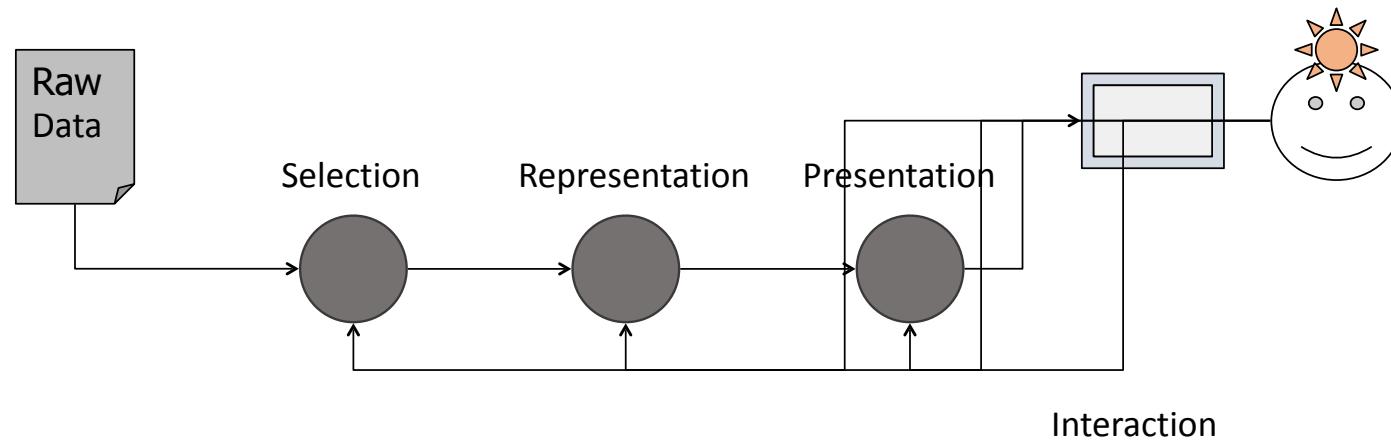
It is difficult to create

good
↑

VISUALIZATIONS



HOW DO WE ARRIVE AT A VISUALIZATION?



The Visualization Pipeline

From [Spence, 2000]

(REPRESENTATION VS. PRESENTATION)

12

12

12

12

12

XII

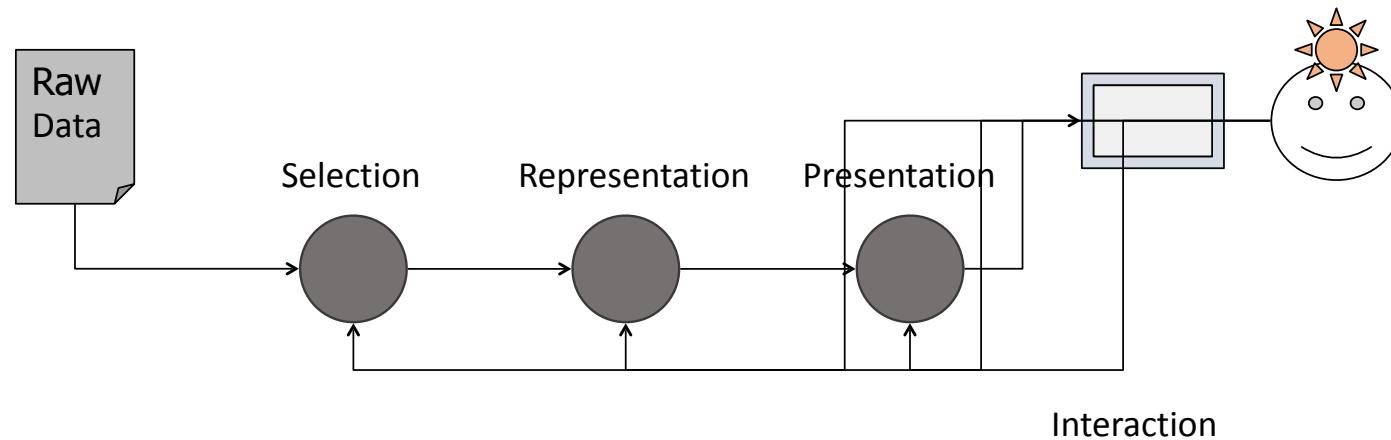
x99

XII

XII

XII

HOW DO WE ARRIVE AT A VISUALIZATION?



The Visualization Pipeline

From [Spence, 2000]

PITFALLS

- Selecting the wrong data
- Selecting the wrong data structure
- Filtering out important data
- Failed understanding of the types of things that need to be shown
- Choosing the wrong representation
- Choosing the wrong presentation format
- Inappropriate interactions provided to explore the data

RECAP

- So far you
 - learned what visualization is
 - learned about the advantages of visualization
 - saw a number of examples (historical and new)
- Next
 - you will get to know your data
 - you will learn about the basic components of visualization

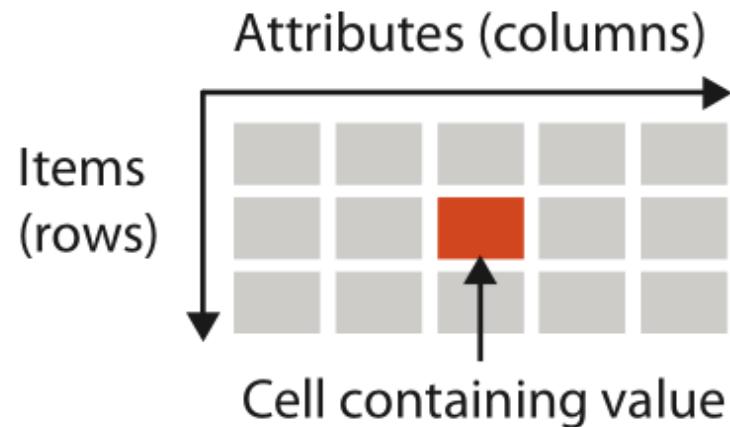
DATA

- Data is the foundation of any visualization
- The visualization designer needs to understand
 - the data properties
 - know what meta-data is available
 - know what people want from the data

DATA SET TYPES

(some of them, more later & even more in the cited literature)

TABLES



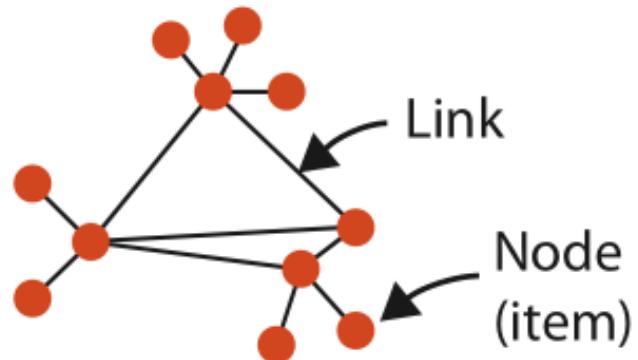
Example:

Items: drinks

Attributes: color, calories, name, ...

[Munzner, Visualization Analysis and Design]

NETWORKS



Example:

Item = nodes: people

Item = links: co-authorship

Node attributes: name, experience, ...

Link attributes: #of papers

[Munzner, Visualization Analysis and Design]

GEOMETRY (SPATIAL)

Specifies information about the shape of items with explicit spatial position

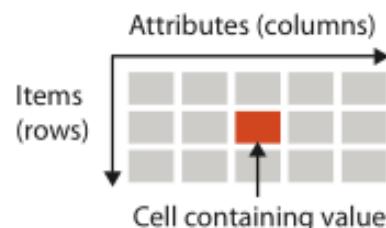


Item = countries
Positions = location on the planet

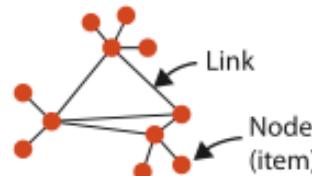
[Munzner, Visualization Analysis and Design]

→ Dataset Types

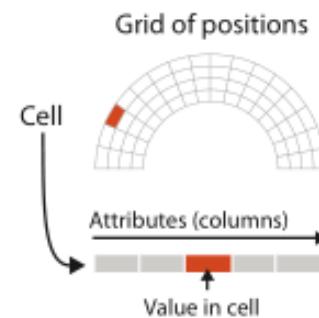
→ Tables



→ Networks



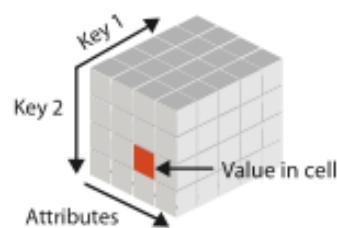
→ Fields (Continuous)



→ Geometry (Spatial)



→ Multidimensional Table



→ Trees



Figure 2.4. The detailed structure of the four basic dataset types.

[Munzner, Visualization Analysis and Design]

ATTRIBUTE TYPES

- Nominal (sometimes called categorical)

- Fruits: apples, oranges
 - Can be compared =, ≠



- Ordered

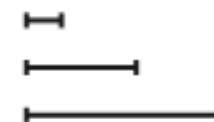
- Ordinal

- T-shirt sizes: S, M, L, XL
 - Can be compared & ordered, but not measured: =, ≠, <, >



- Quantitative

- Counts and amounts, 5kg / 10kg
 - you can do =, ≠, <, >, - , +, ×, ÷



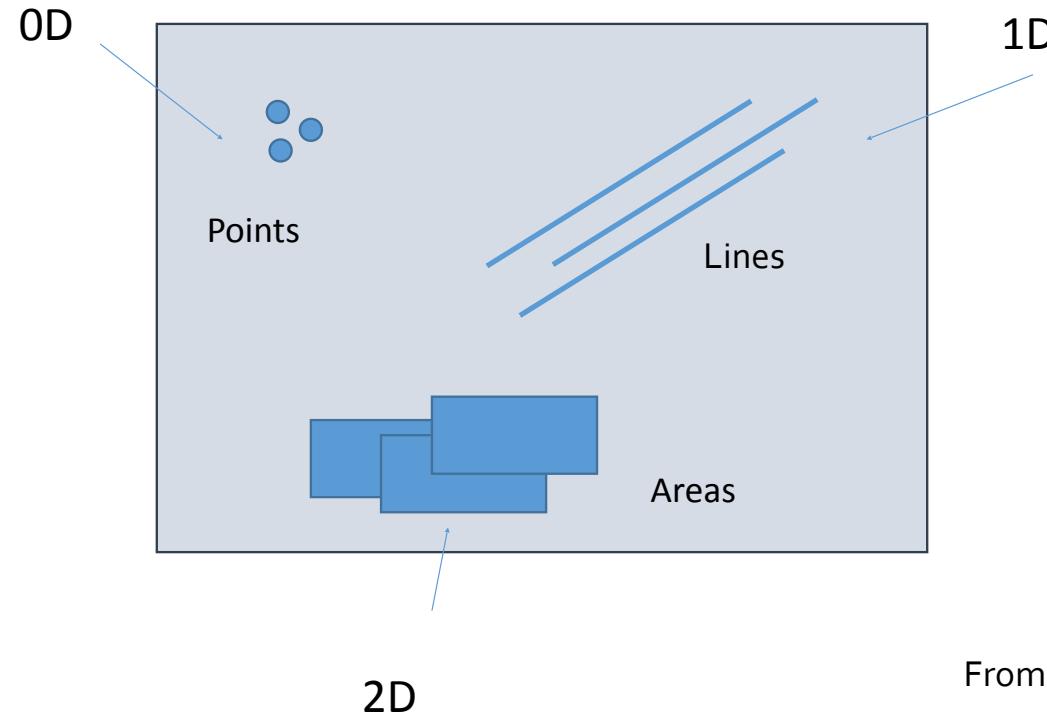
WHY IS THIS IMPORTANT?

- Nominal, ordinal, and quantitative data are best expressed in different ways visually
- Data types often have inherent tasks
 - geometry (understand spatial relationships)
 - trees (understand parent-child relationships)
 - ...
- But:
 - any data type (1D, 2D,...) can be expressed in a multitude of ways!

VISUALIZATION BUILDING BLOCKS

MARKS

Basic geometric element data depict items or links

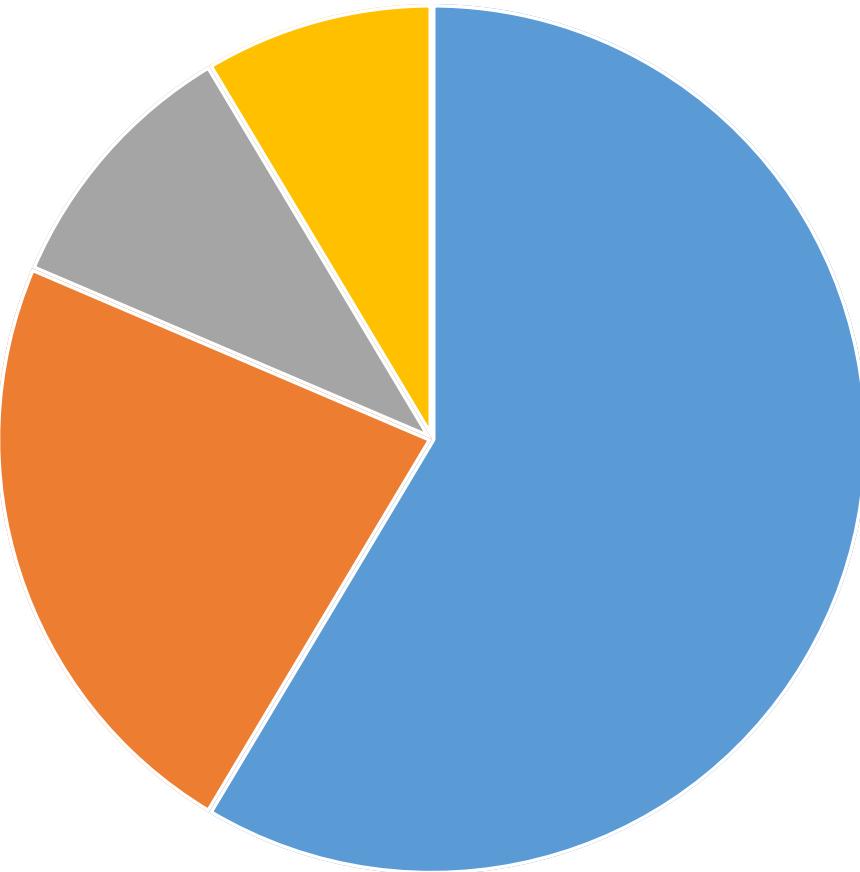


From Semiology of Graphics (Bertin) 53

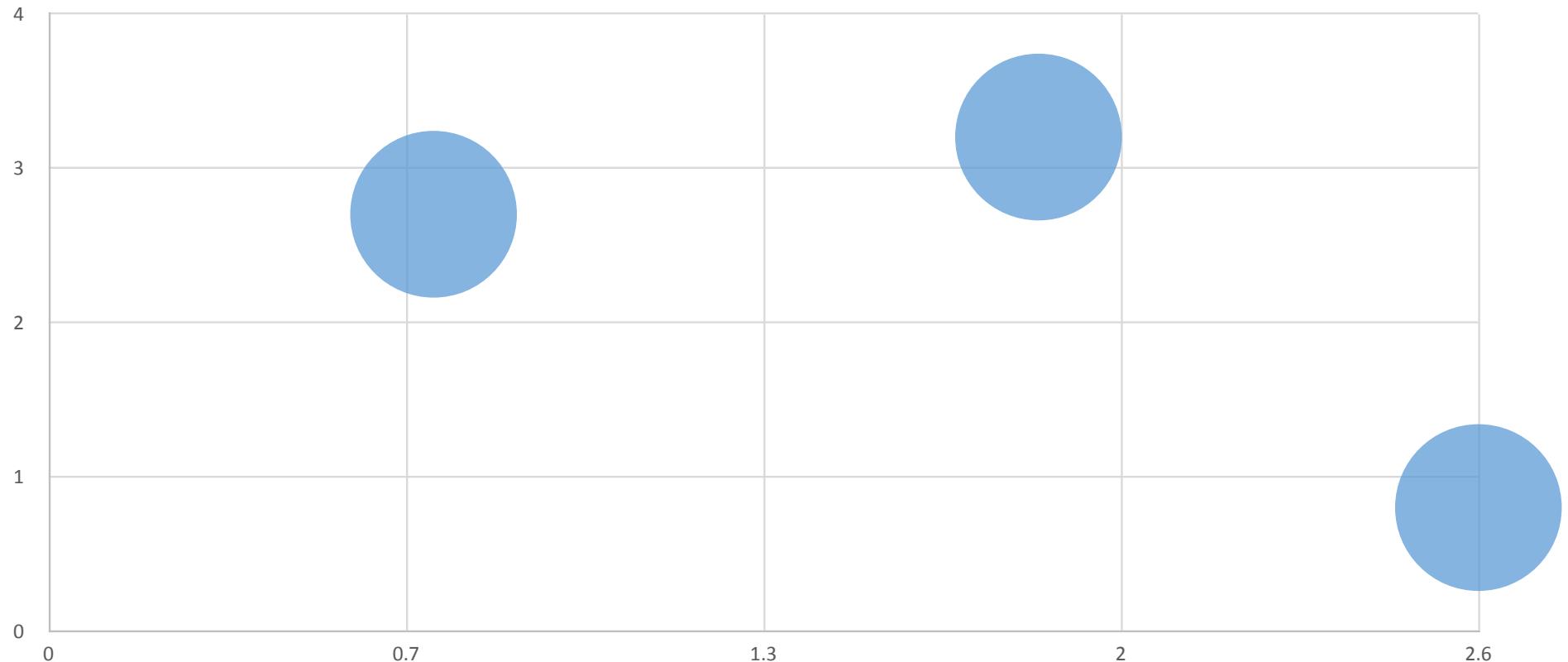
which marks do you see?



which marks do you see?



which marks do you see?



VISUAL CHANNELS

Also often called visual variables

④ Position

→ Horizontal



→ Vertical



→ Both



④ Color



Modify marks
independent of the
dimensionality of the
mark

④ Shape



④ Tilt / angle



④ Size

→ Length



→ Area

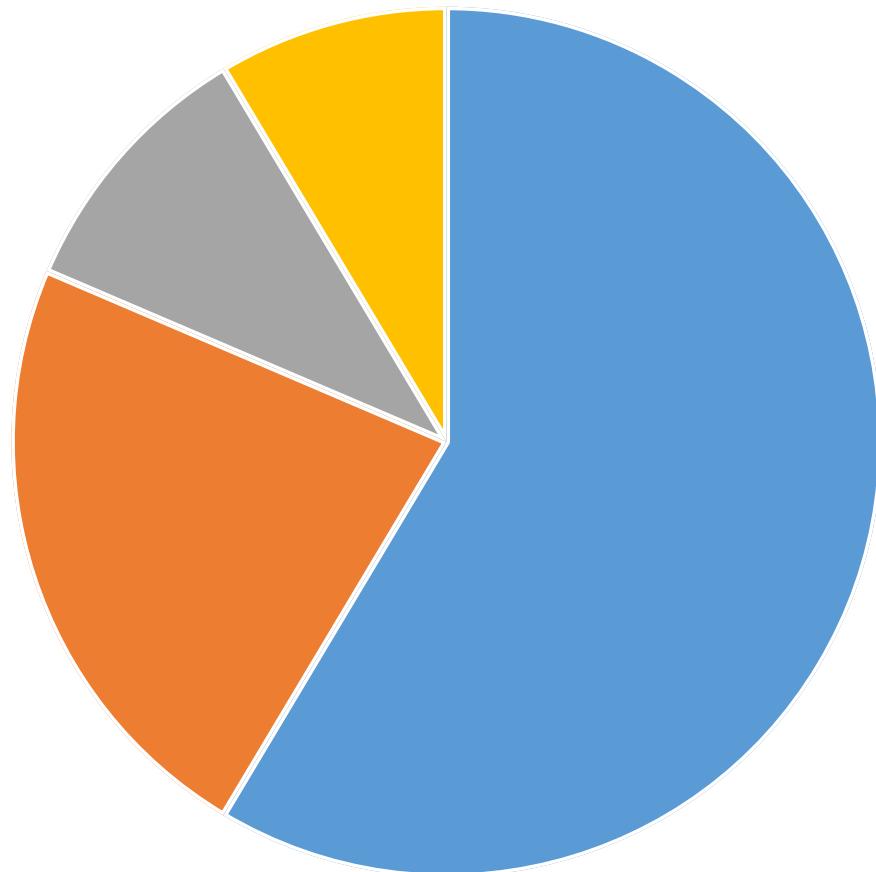


→ Volume



These are the most common but there are more

which visual channels encode data?



↪ Position

→ Horizontal → Vertical → Both



↪ Color



↪ Shape



↪ Tilt / angle

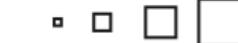


↪ Size

→ Length



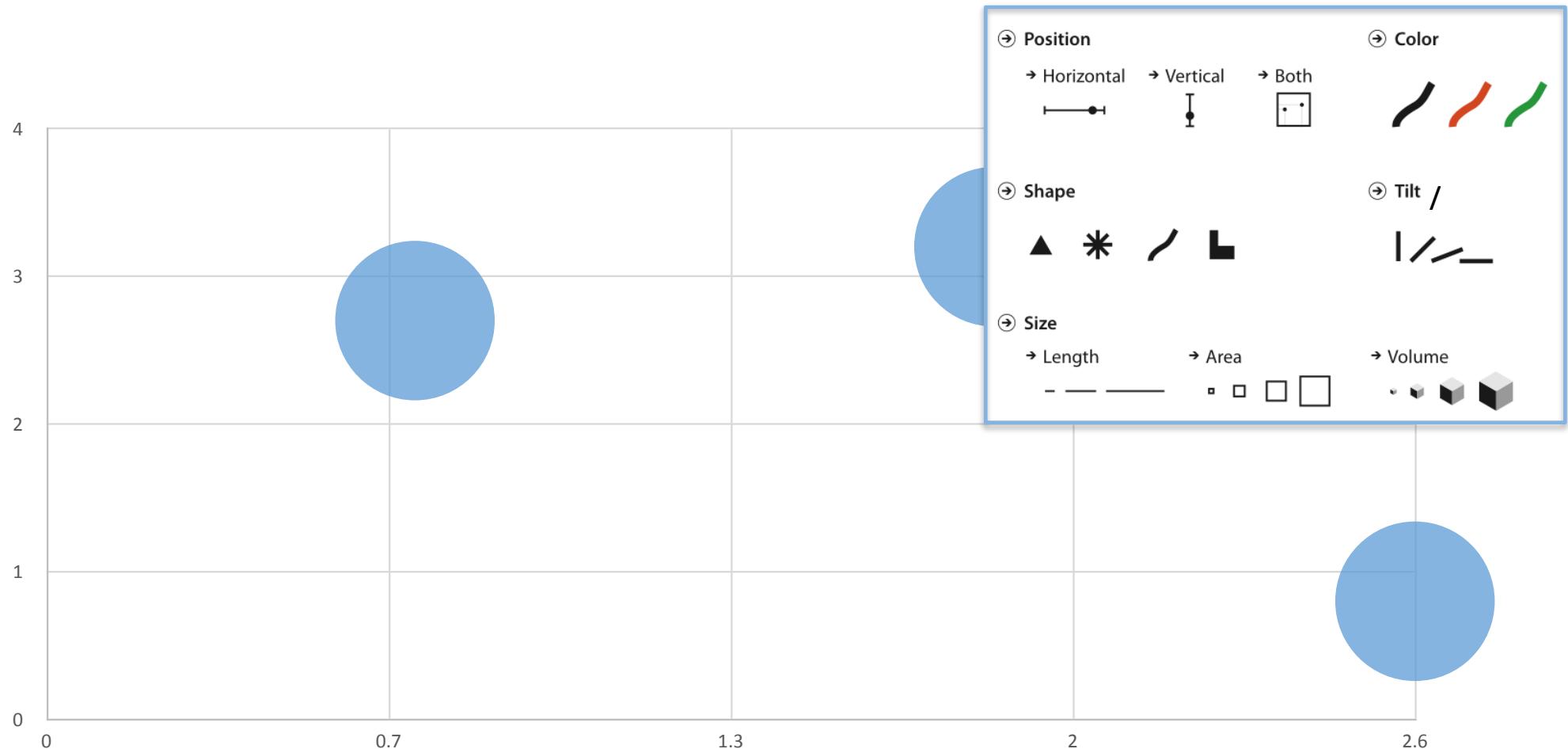
→ Area



→ Volume

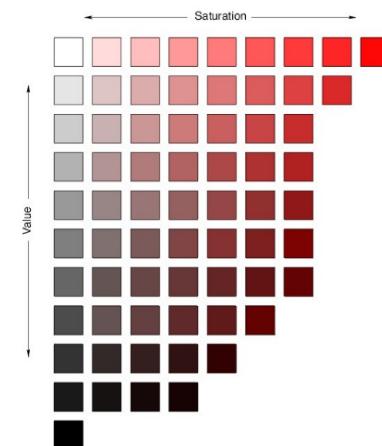


which visual channels encode data?



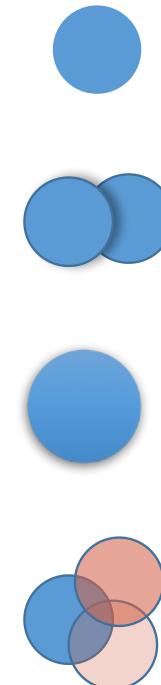
ADDITIONAL CHANNELS

- motion
 - direction, acceleration, speed, frequency, onset, 'personality'
- saturation
 - colour as Bertin uses largely refers to hue, saturation != value



ADDITIONAL CHANNELS

- flicker
 - frequency, rhythm, appearance
- depth? ‘quasi’ 3D
 - depth, occlusion, aerial perspective, binocular disparity
- illumination
- transparency



HOW TO CHOOSE CHANNELS?

EXPRESSIVENESS

show all of, and only, the information in the attributes
(nothing more, nothing less)

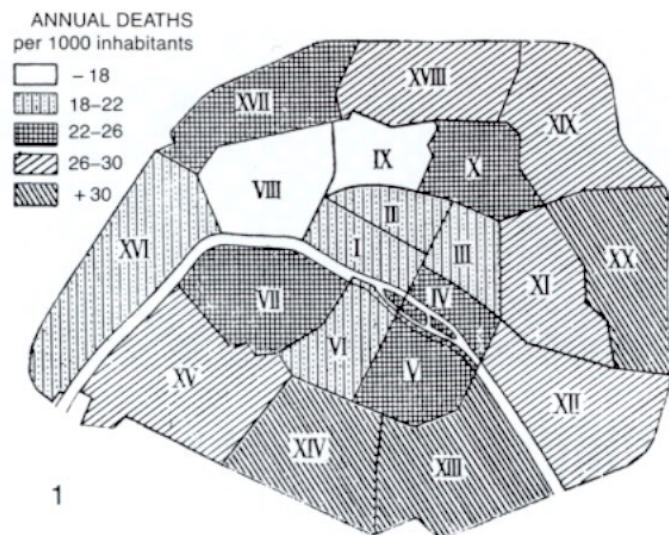
*Example:
an ordered attribute needs to look ordered,
an unordered attribute should not*



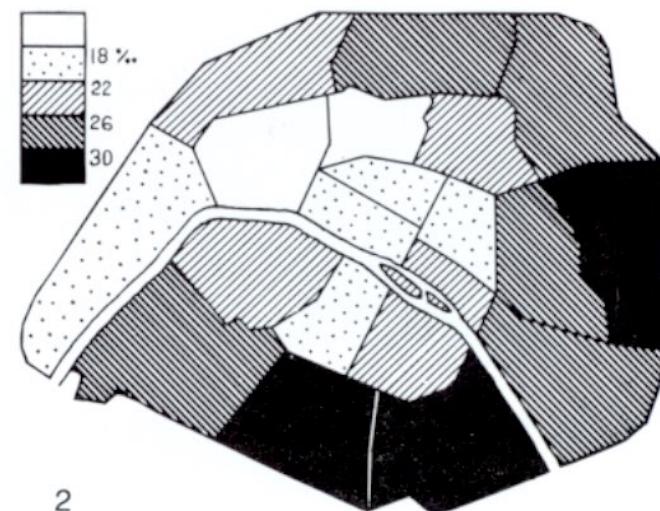
This mismatch is a common beginner's mistake

EXPRESSIVENESS

luminance is ordered, cannot be reordered



Values not ordered correctly according to scale
Information has to be read point by point

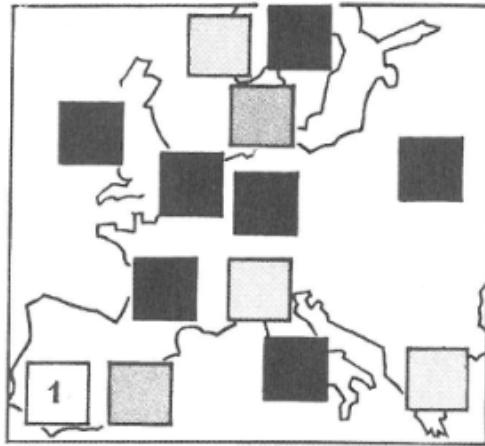


Values ordered correctly
Image much more useful

annual deaths per 1000 inhabitants, Paris

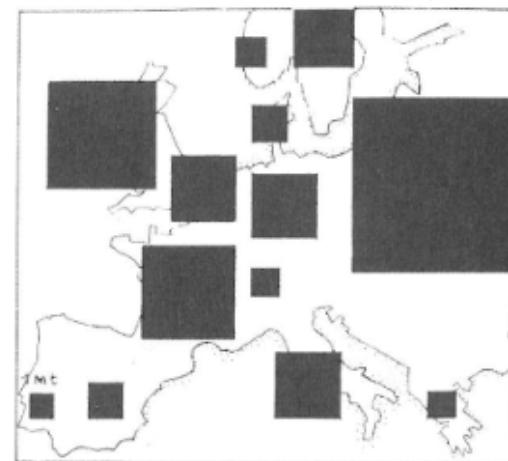
EXPRESSIVENESS

luminance is not quantitative



if Portugal is 1, what is France?

you need a legend!



if Portugal is 1, what is France?

still hard, but doable

HOW TO CHOOSE CHANNELS?

EFFECTIVENESS

the importance of the attribute matches the salience of the channel

(the most important attributes should be encoded with the most effective channels)

EFFECTIVENESS

Accuracy: How accurately values can be estimated.

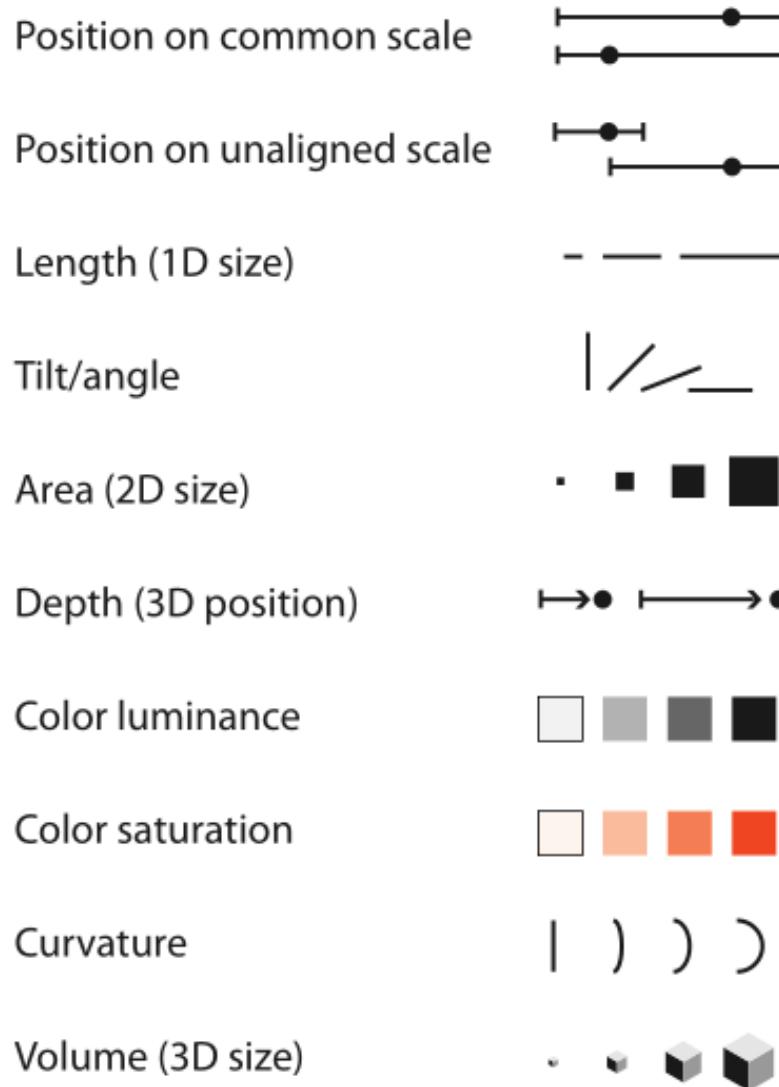
Discriminability: How many different values can be perceived.

Separability: How much interaction there is with multiple encodings.

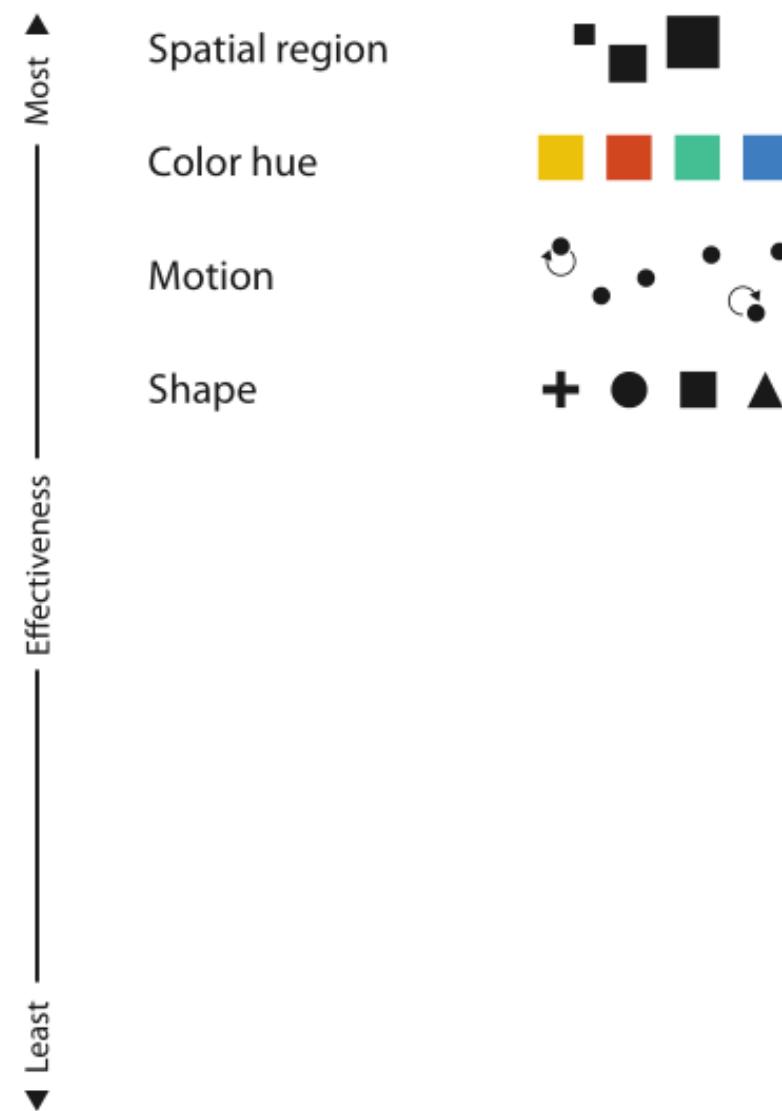
Popout: How easy it is to spot some values from the rest.

Grouping: How good a channel is in conveying groups.

④ Magnitude Channels: Ordered Attributes



⑤ Identity Channels: Categorical Attributes



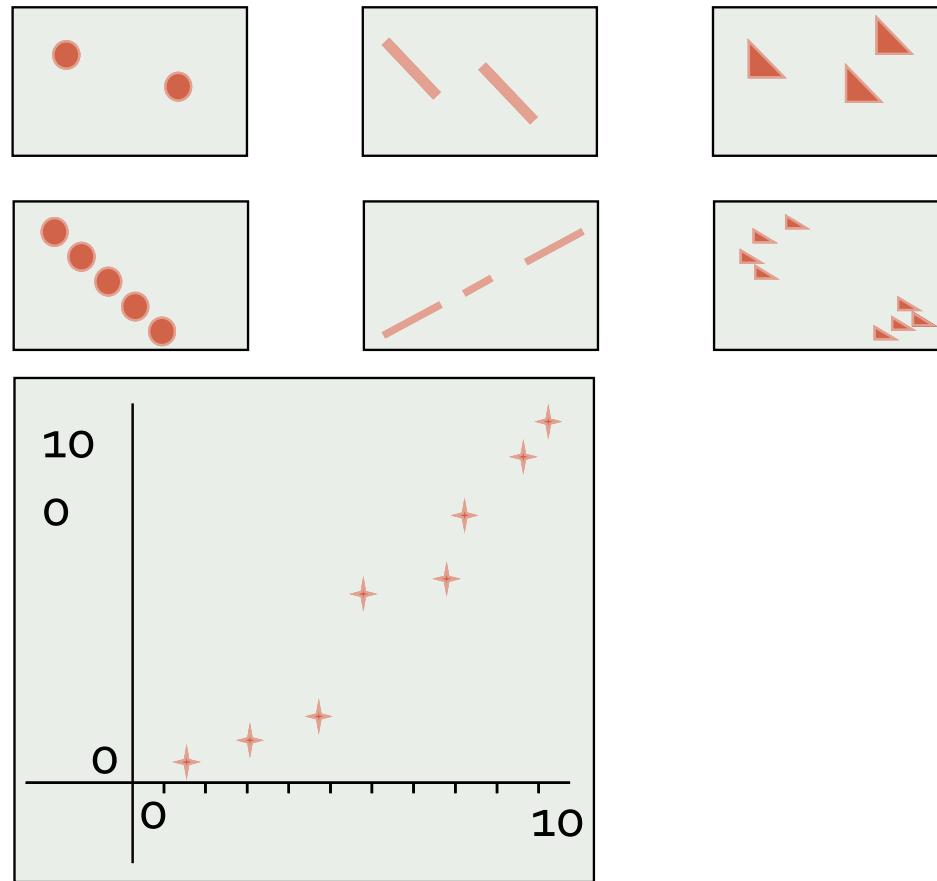
We will look at effectiveness a lot more throughout the lectures

We will look at effectiveness a lot more throughout the lectures,

(... but briefly for the most common ...)

Channel: Position

- ✓ high accuracy
- ✓ high discriminability
- ✓ high separability
- ✓ popout possible
- ✓ grouping possible
- ✓ ordering possible

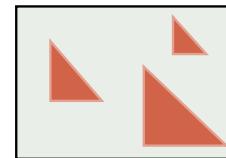
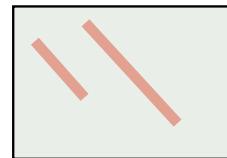
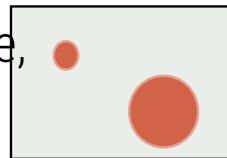


From Semiology of Graphics (Bertin)

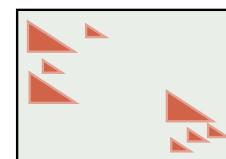
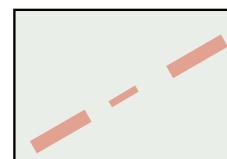
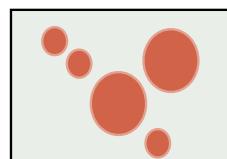
Channel: Size



- medium accuracy (medium in 2D size,
but great for 1D size = length)



- high discriminability

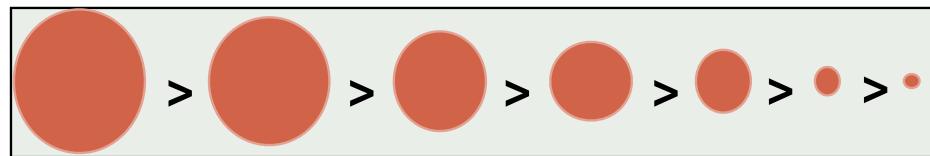


- high separability

$$4 \times \blacksquare = \begin{array}{|c|c|}\hline \blacksquare & \blacksquare \\ \hline \blacksquare & \blacksquare \\ \hline \end{array} ?$$



- popout possible



- grouping possible



- ordering possible

Size



points



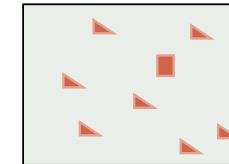
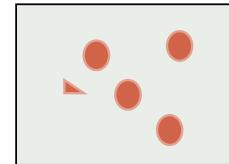
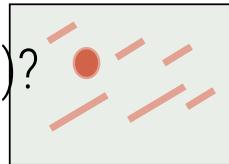
lines



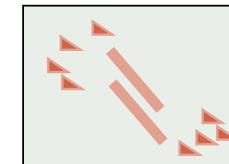
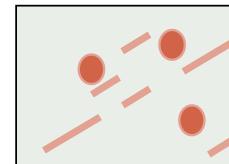
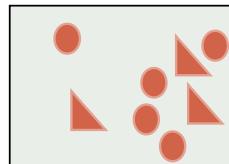
areas

Visual Variable: Shape

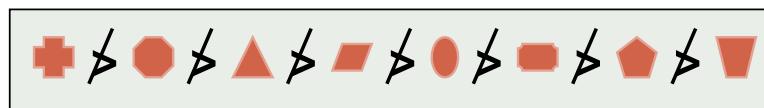
✗ low accuracy (what does it even mean)?



✓ high discriminability



✗ medium separability



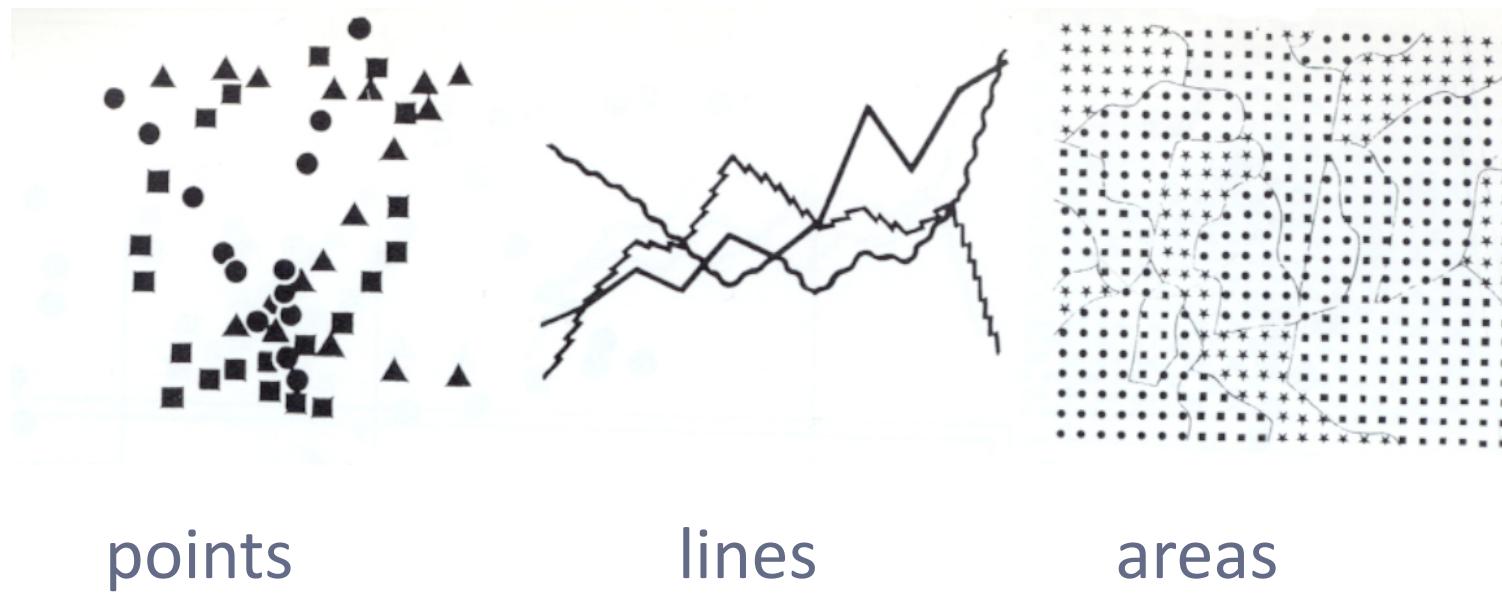
✓ popout possible

✓ grouping possible



✗ not ordered

Shape



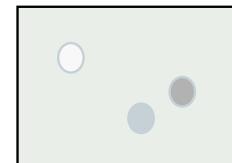
points

lines

areas

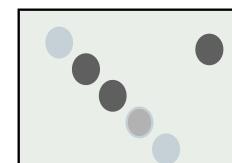
Channel: Value (color luminance)

✗ low accuracy

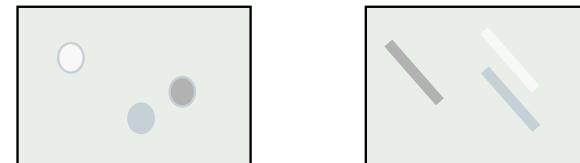


✓ high discriminability

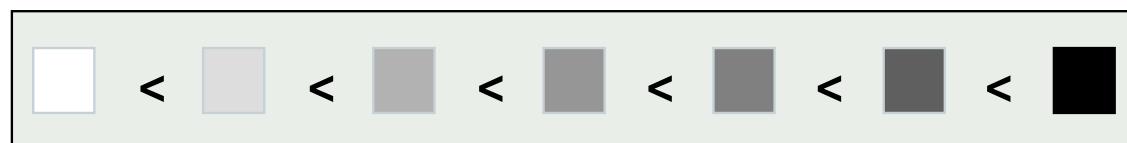
theoretically infinite but practically limited
association and selection ~ < 7 and distinction ~ 10



✓ medium separability



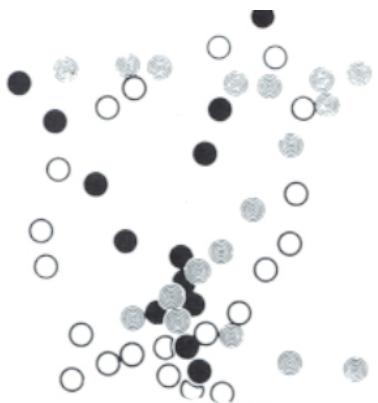
✓ popout possible



✓ grouping possible

✓ ordered

Value (Color luminance)



points



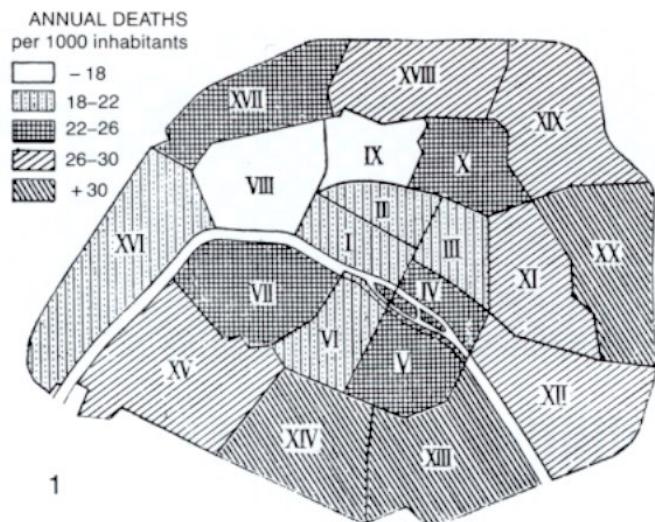
lines



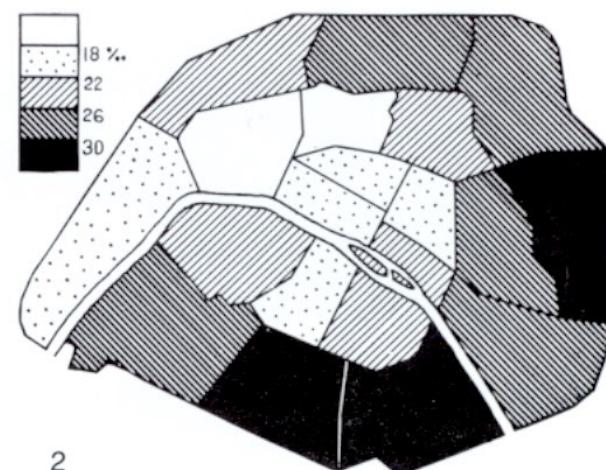
areas

Value (Color luminance)

ordered, cannot be reordered



Values not ordered correctly according to scale
Information has to be read point by point



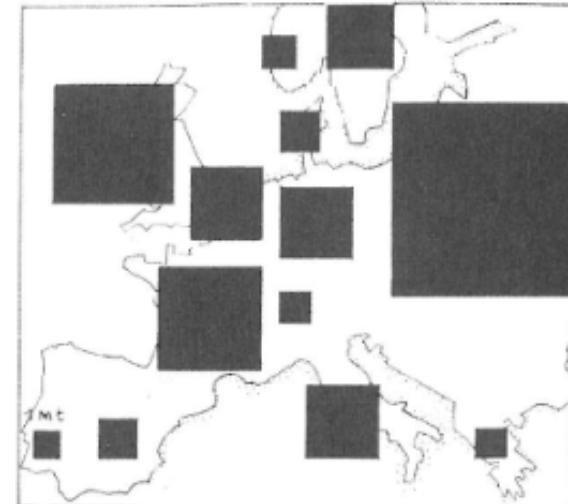
Values ordered correctly
Image much more useful

Value (Color luminance)

is not quantitative



if Portugal is 1,
what is France?
you need a legend!



if Portugal is 1,
what is France?
still hard, but doable

Channel: Colour

- ✗ low accuracy (what does it even mean?)
- ✓ high discriminability
theoretically infinite but practically limited
association and selection $\sim < 7$ and distinction ~ 10
- ✗ medium separability
- ✓ popout possible
- ✓ grouping possible
- ✗ not ordered

