# DATA VISUALIZATION

## Introduction

Tobias Isenberg
tobias.isenberg@inria.fr

# INSTRUCTORS

## Tobias ISENBERG

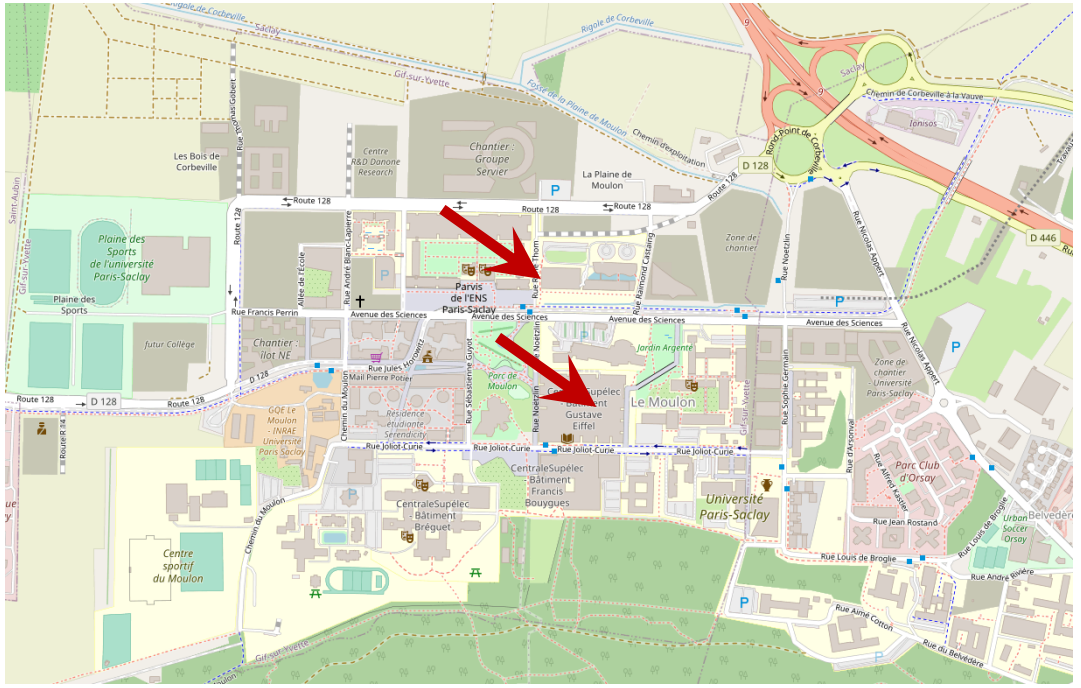
*tobias.isenberg@inria.fr* https://tobias.isenberg.cc/

**OFFICE**

Digiteo-Moulon (building 660),
room 1044 (01-69156433)

**OFFICE HOURS**: by appointment

**CONTACT PREFERENCE**: e-mail



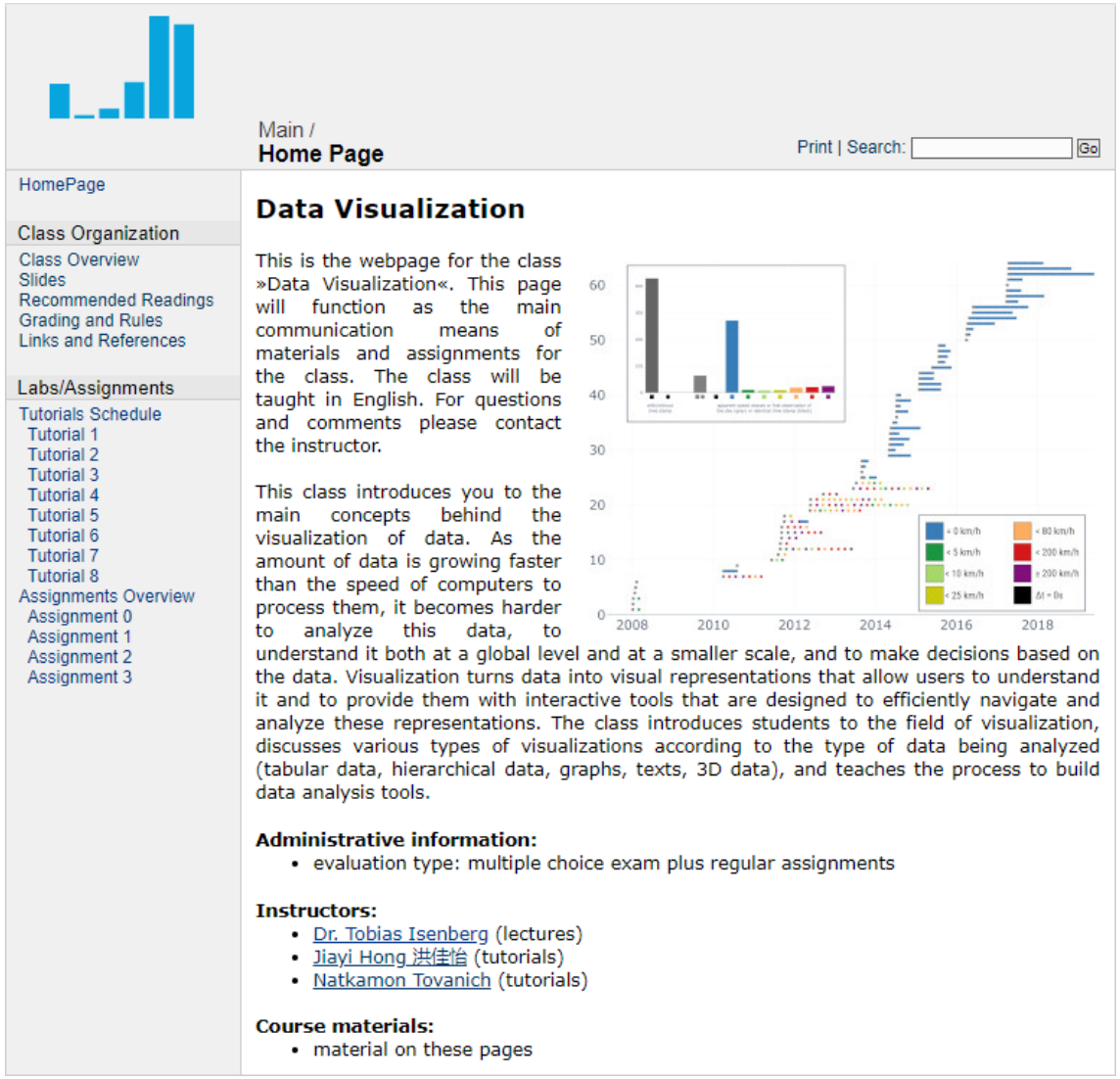
# INSTRUCTOR – Tutorials


## Natkamon TOVANICH

*natkamon.tovanich@inria.fr*


contact the TA for any questions about
the tutorials and assignments
**primary contact:** via e-mail

# COURSE INFO https://tobias.isenberg.cc/visualization/



Main /
**Home Page**

Print | Search: Go

HomePage

Class Organization
Class Overview
Slides
Recommended Readings
Grading and Rules
Links and References

Labs/Assignments
Tutorials Schedule
Tutorial 1
Tutorial 2
Tutorial 3
Tutorial 4
Tutorial 5
Tutorial 6
Tutorial 7
Tutorial 8
Assignments Overview
Assignment 0
Assignment 1
Assignment 2
Assignment 3

## Data Visualization

This is the webpage for the class »Data Visualization«. This page will function as the main communication means of materials and assignments for the class. The class will be taught in English. For questions and comments please contact the instructor.

This class introduces you to the main concepts behind the visualization of data. As the amount of data is growing faster than the speed of computers to process them, it becomes harder to analyze this data, to understand it both at a global level and at a smaller scale, and to make decisions based on the data. Visualization turns data into visual representations that allow users to understand it and to provide them with interactive tools that are designed to efficiently navigate and analyze these representations. The class introduces students to the field of visualization, discusses various types of visualizations according to the type of data being analyzed (tabular data, hierarchical data, graphs, texts, 3D data), and teaches the process to build data analysis tools.

**Administrative information:**
- evaluation type: multiple choice exam plus regular assignments

**Instructors:**
- Dr. Tobias Isenberg (lectures)
- Jiayi Hong 洪佳怡 (tutorials)
- Natkamon Tovanich (tutorials)

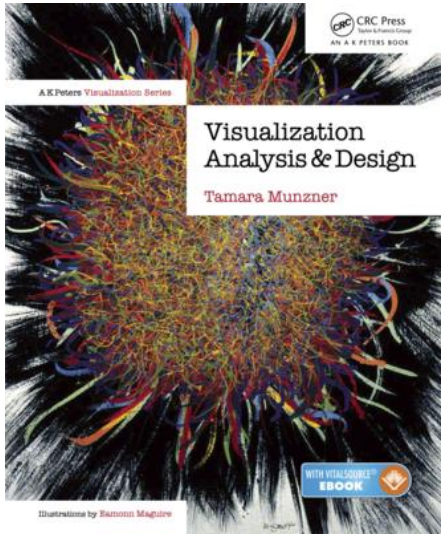
**Course materials:**
- material on these pages

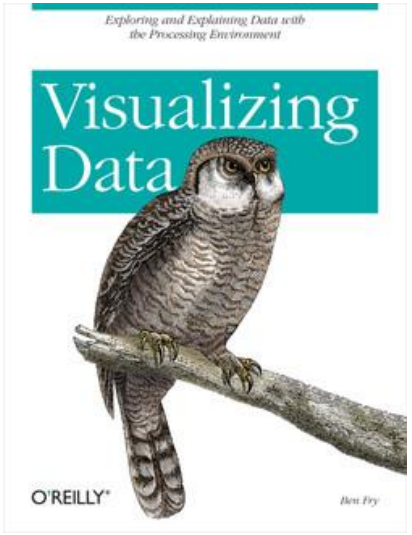
- class overview
- slides (after the lectures)
- tutorial information
- assignment details
- contact information

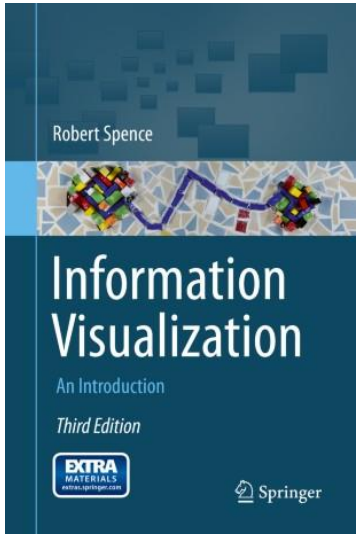
# COURSE INFO https://tobias.isenberg.cc/visualization/

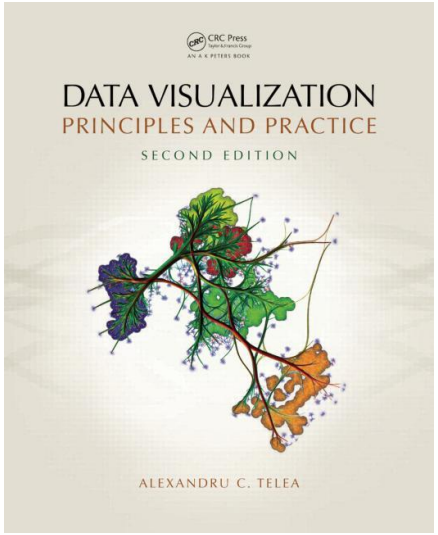
1. General class overview
   Introduction to information visualization
2. Perception and color
3. Multi-dimensional data visualization
4. Graphs and trees
5. Time-dependent data
6. Storytelling with data
7. Interaction

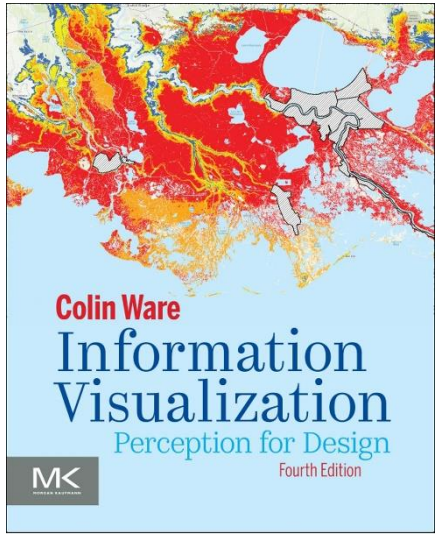
# RECOMMENDED READINGS



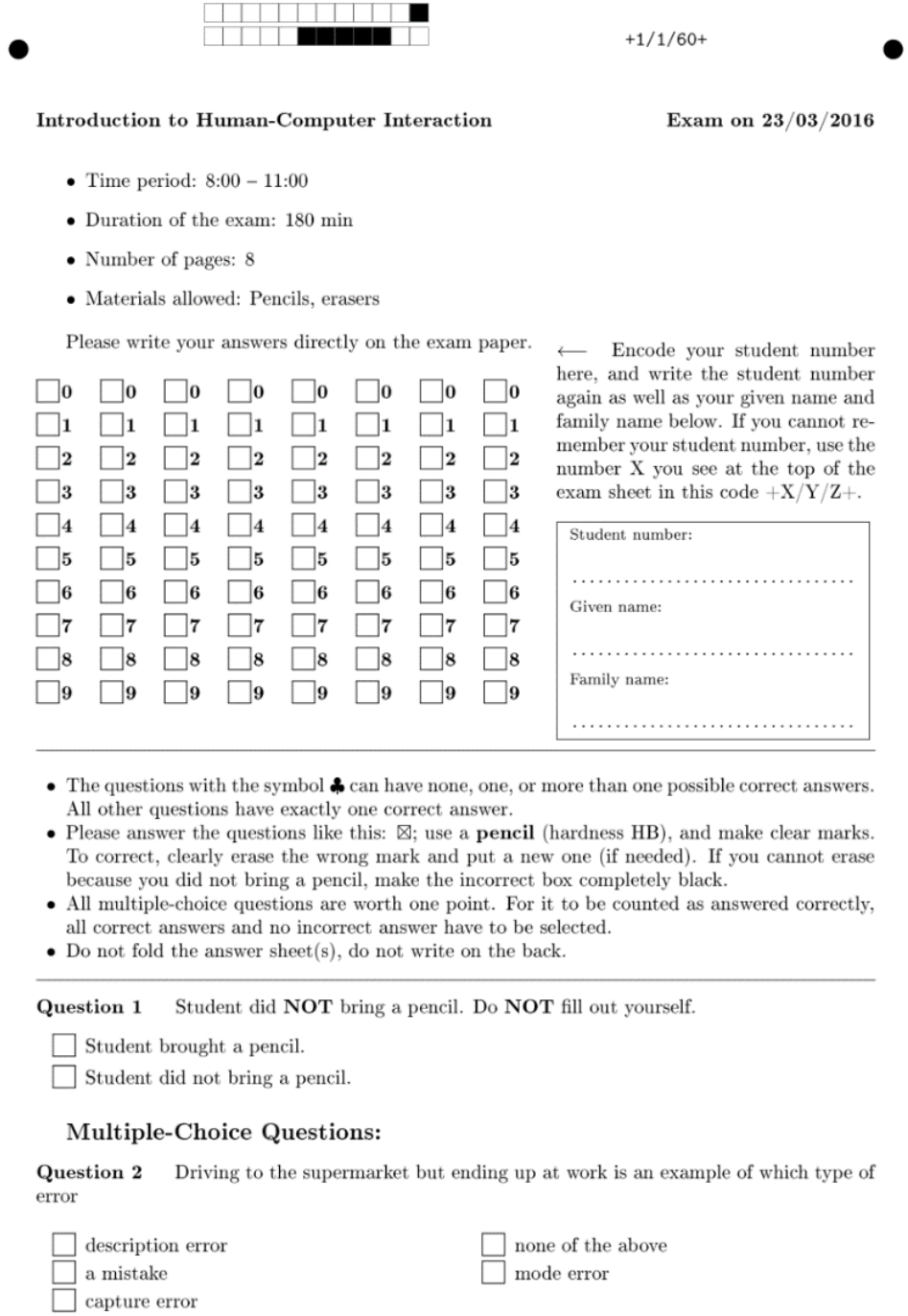




→ links to the books on the class website

# GRADING SCHEME

- Written multiple-choice exam: 50%
- Assignments/project: 50%
  - More information later, after the break



+1/1/60+

**Introduction to Human-Computer Interaction** **Exam on 23/03/2016**

- Time period: 8:00 – 11:00
- Duration of the exam: 180 min
- Number of pages: 8
- Materials allowed: Pencils, erasers

Please write your answers directly on the exam paper.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ☐0 | ☐0 | ☐0 | ☐0 | ☐0 | ☐0 | ☐0 | ☐0 |
| ☐1 | ☐1 | ☐1 | ☐1 | ☐1 | ☐1 | ☐1 | ☐1 |
| ☐2 | ☐2 | ☐2 | ☐2 | ☐2 | ☐2 | ☐2 | ☐2 |
| ☐3 | ☐3 | ☐3 | ☐3 | ☐3 | ☐3 | ☐3 | ☐3 |
| ☐4 | ☐4 | ☐4 | ☐4 | ☐4 | ☐4 | ☐4 | ☐4 |
| ☐5 | ☐5 | ☐5 | ☐5 | ☐5 | ☐5 | ☐5 | ☐5 |
| ☐6 | ☐6 | ☐6 | ☐6 | ☐6 | ☐6 | ☐6 | ☐6 |
| ☐7 | ☐7 | ☐7 | ☐7 | ☐7 | ☐7 | ☐7 | ☐7 |
| ☐8 | ☐8 | ☐8 | ☐8 | ☐8 | ☐8 | ☐8 | ☐8 |
| ☐9 | ☐9 | ☐9 | ☐9 | ☐9 | ☐9 | ☐9 | ☐9 |

⟵ Encode your student number here, and write the student number again as well as your given name and family name below. If you cannot remember your student number, use the number X you see at the top of the exam sheet in this code +X/Y/Z+.

Student number: .....

Given name: .....

Family name: .....

- The questions with the symbol ♣ can have none, one, or more than one possible correct answers. All other questions have exactly one correct answer.
- Please answer the questions like this: ☒; use a **pencil** (hardness HB), and make clear marks. To correct, clearly erase the wrong mark and put a new one (if needed). If you cannot erase because you did not bring a pencil, make the incorrect box completely black.
- All multiple-choice questions are worth one point. For it to be counted as answered correctly, all correct answers and no incorrect answer have to be selected.
- Do not fold the answer sheet(s), do not write on the back.

**Question 1** Student did **NOT** bring a pencil. Do **NOT** fill out yourself.

☐ Student brought a pencil.
☐ Student did not bring a pencil.

**Multiple-Choice Questions:**

**Question 2** Driving to the supermarket but ending up at work is an example of which type of error

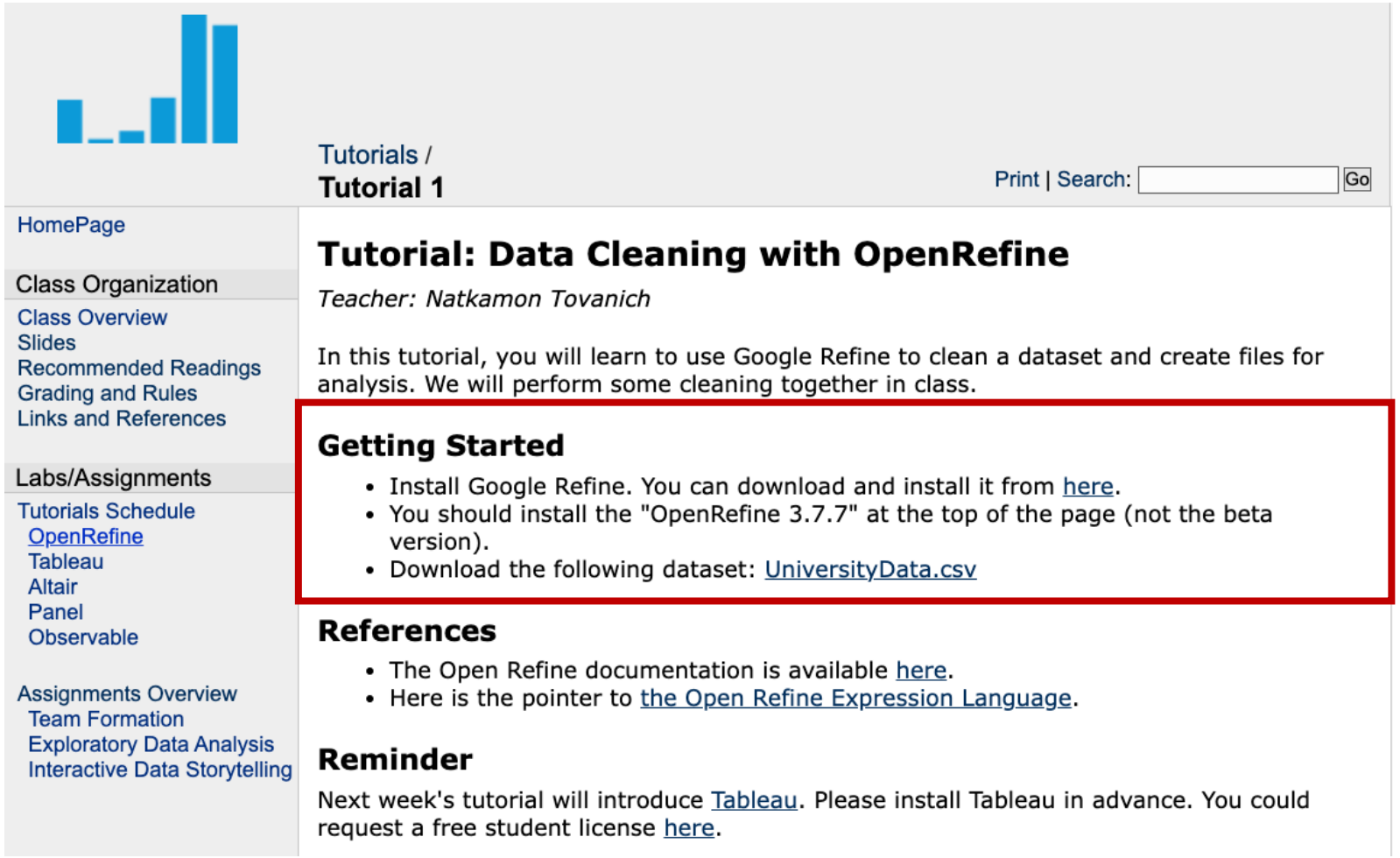
☐ description error
☐ a mistake
☐ capture error
☐ none of the above
☐ mode error

# Tutorials

https://tobias.isenberg.cc/visualization/

1. Data wrangling with OpenRefine
2. Working with Tableau
3. Data vis with Altair # 1: Basics
4. Data vis with Altair # 2: Exercise
5. Data vis with Altair # 3: Interaction & selection
6. Data apps with Streamlit
7. Streamlit continued, Data dashboards with Panel or Cartographic Visualization in Altair

# Prepare for Tutorial



Tutorials /
**Tutorial 1**

Print | Search: Go

HomePage

Class Organization
- Class Overview
- Slides
- Recommended Readings
- Grading and Rules
- Links and References

Labs/Assignments
- Tutorials Schedule
  - OpenRefine
  - Tableau
  - Altair
  - Panel
  - Observable
- Assignments Overview
  - Team Formation
  - Exploratory Data Analysis
  - Interactive Data Storytelling

## Tutorial: Data Cleaning with OpenRefine

*Teacher: Natkamon Tovanich*

In this tutorial, you will learn to use Google Refine to clean a dataset and create files for analysis. We will perform some cleaning together in class.

### Getting Started

- Install Google Refine. You can download and install it from here.
- You should install the "OpenRefine 3.7.7" at the top of the page (not the beta version).
- Download the following dataset: UniversityData.csv

### References

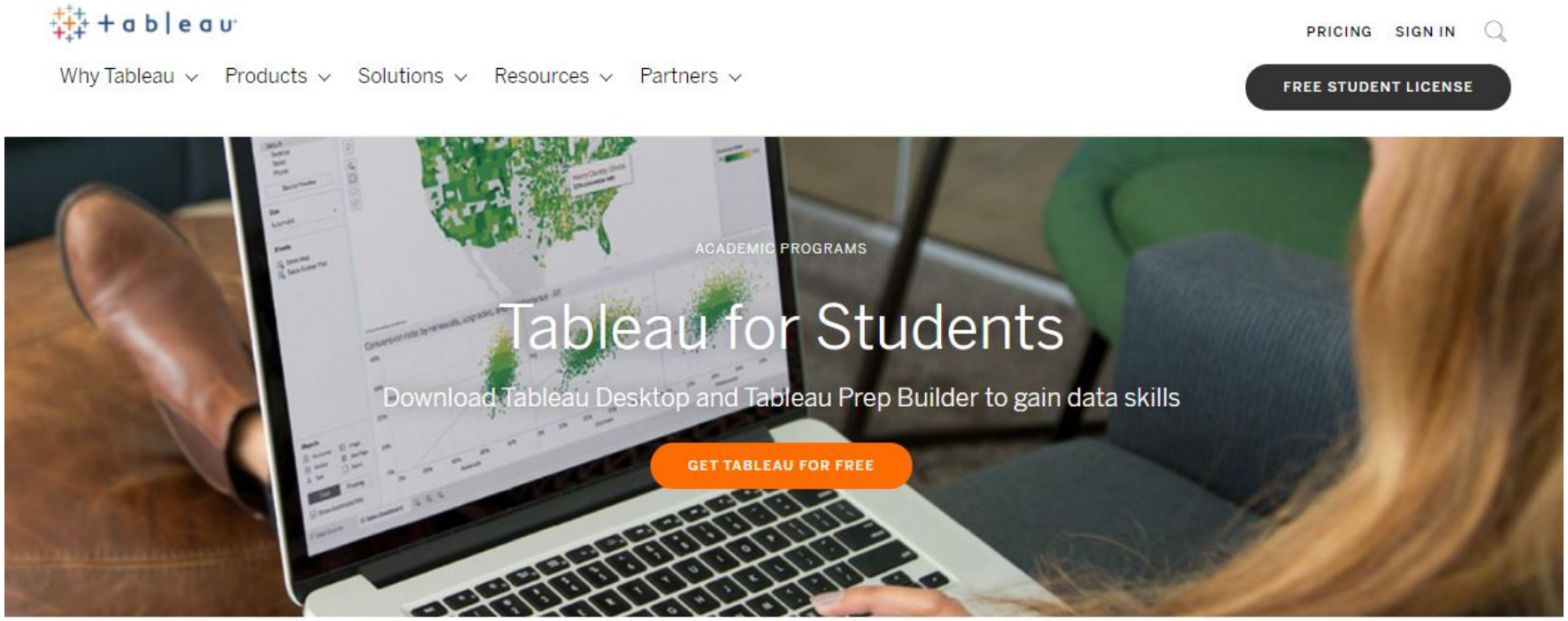
- The Open Refine documentation is available here.
- Here is the pointer to the Open Refine Expression Language.

### Reminder

Next week's tutorial will introduce Tableau. Please install Tableau in advance. You could request a free student license here.

# Prepare for EACH Tutorial

Install Tableau and register for the student license




We offer free one-year Tableau licenses to students at accredited academic institutions through our Tableau for Students program. Receive access to our entire eLearning suite once verified.

# Assignments

https://tobias.isenberg.cc/visualization/

- two major assignments for the class
- detailed instructions on the class website
- assignment results count for 50% of the final grade
- work in groups of 2 students (= Assignment 0)
- find a group and let us know about groups **before** the third class (deadline: 23:00 on day before)
- strict deadlines; point deduction if late

# Assignment 1

https://tobias.isenberg.cc/visualization/

**Data Proposals and Exploratory Data Analysis** (30 points)

Part 1 (group work)

- pick a topic area of interest
- find a dataset that can provide insights into that topic
- describe the dataset

Part 2 (individual work)

- each team member should find a research question about the dataset
- produce one exploratory visualization that effectively answers that research question
- each team member must pick different questions
- write up on findings you gain

# Assignment 2

https://tobias.isenberg.cc/visualization/

Create an interactive data storytelling (70 points)

- at least 3 different views/visualization components
- multiple views coordinated with linked highlighting
- at least 2 UI widgets allow users to filter the data or update certain views interactively
- interactive tooltips are shown when users hover over marks, at least in one view

You are free to use libraries introduced in the class or any code-based libraries.
However, you are not allowed to use dashboard software for the final project.

- online tool
- code base
- video presentation
- presentation slide

# Schedule

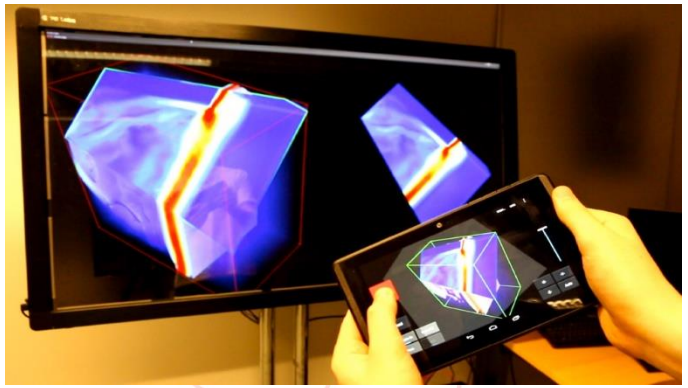
| week 1 | week 2 | week 3 | week 4 | week 5 | week 6 | week 7 | week 8 | week 9 | week 11 | final deadline |
|---|---|---|---|---|---|---|---|---|---|---|
| Jan. 23 | Jan. 30 | Feb. 6 | Feb. 13 | Feb. 20 | Feb. 27 | Mar. 6 | Mar. 13 | Mar. 20 | Mar. 27 | Mar. 31 |
| L1 | L3 | T1 | break | break | T3 | L5 | L6 | L7 | exam | – |
| L2 | L4 | T2 | break | break | T4 | T5 | T6 | T7 | exam | – |
| start A0 | | start A1 | | | start A2 | | | | | |
| | | A0 due Feb. 5, 23:00 CET | | | A1 due Feb. 26, 23:00 CET | | | | | A2 due 23:00 CET |

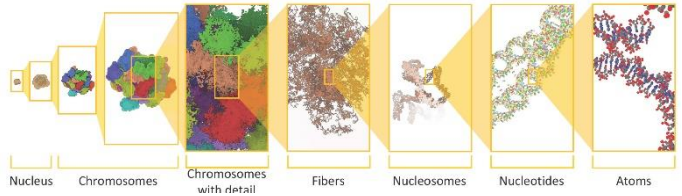
# ANY QUESTIONS?

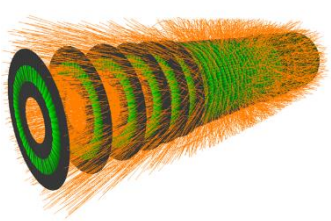
- please ask questions during the lecture
  - if I speak too fast
  - if I use terms or expressions that you do not understand
  - if you do not understand a concept
  - there are no dumb questions!
  - there is no impact on the grade!
- e-mail us if you have questions in-between lectures

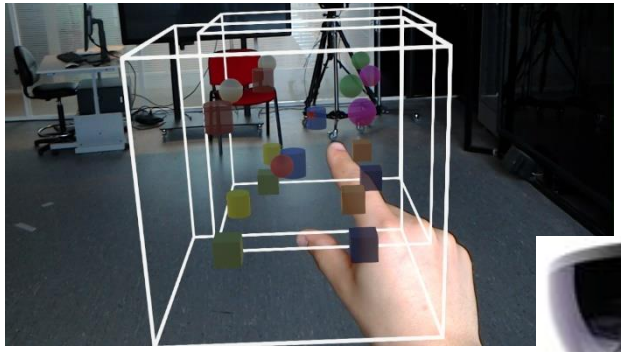
# FINAL NOTE BEFORE WE START

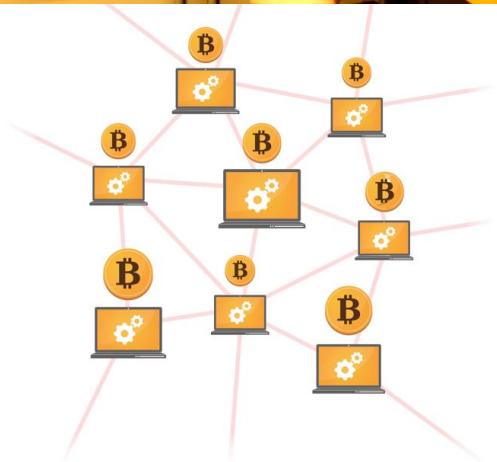
- we are always looking for good students with visualization background for internships/MSc theses and, later PhD theses – contact us if interested

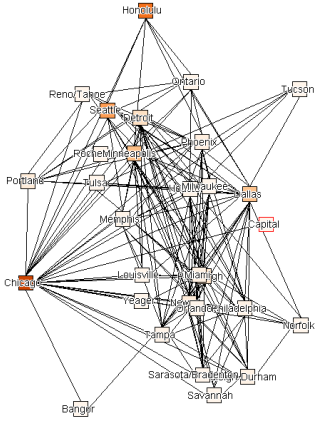


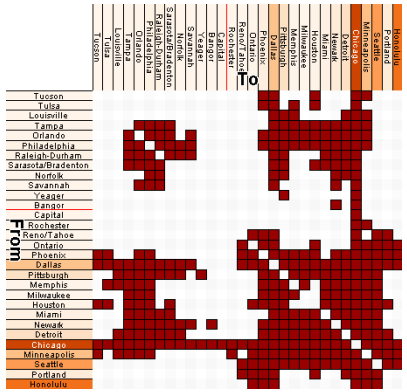


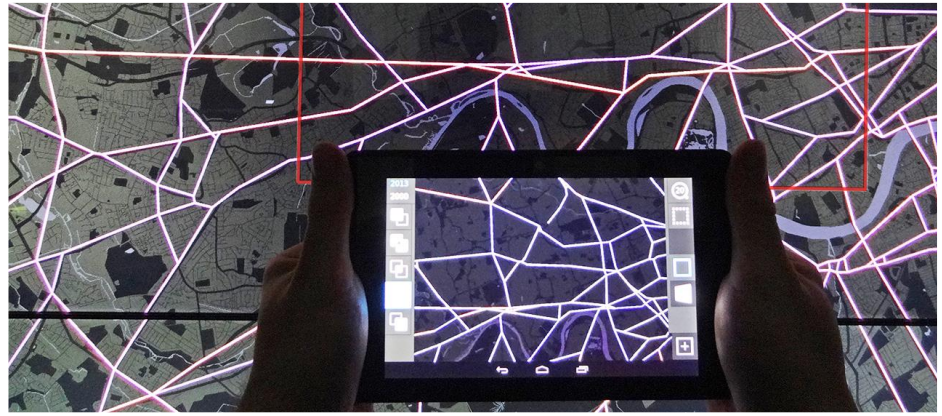











Why

# VISUALIZE DATA?

# AFTER TODAY YOU WILL…

have gained an overview of visualization basics

learned basic principles of data representation and interaction

understand **data types**, **marks**, **visual variables**

# 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

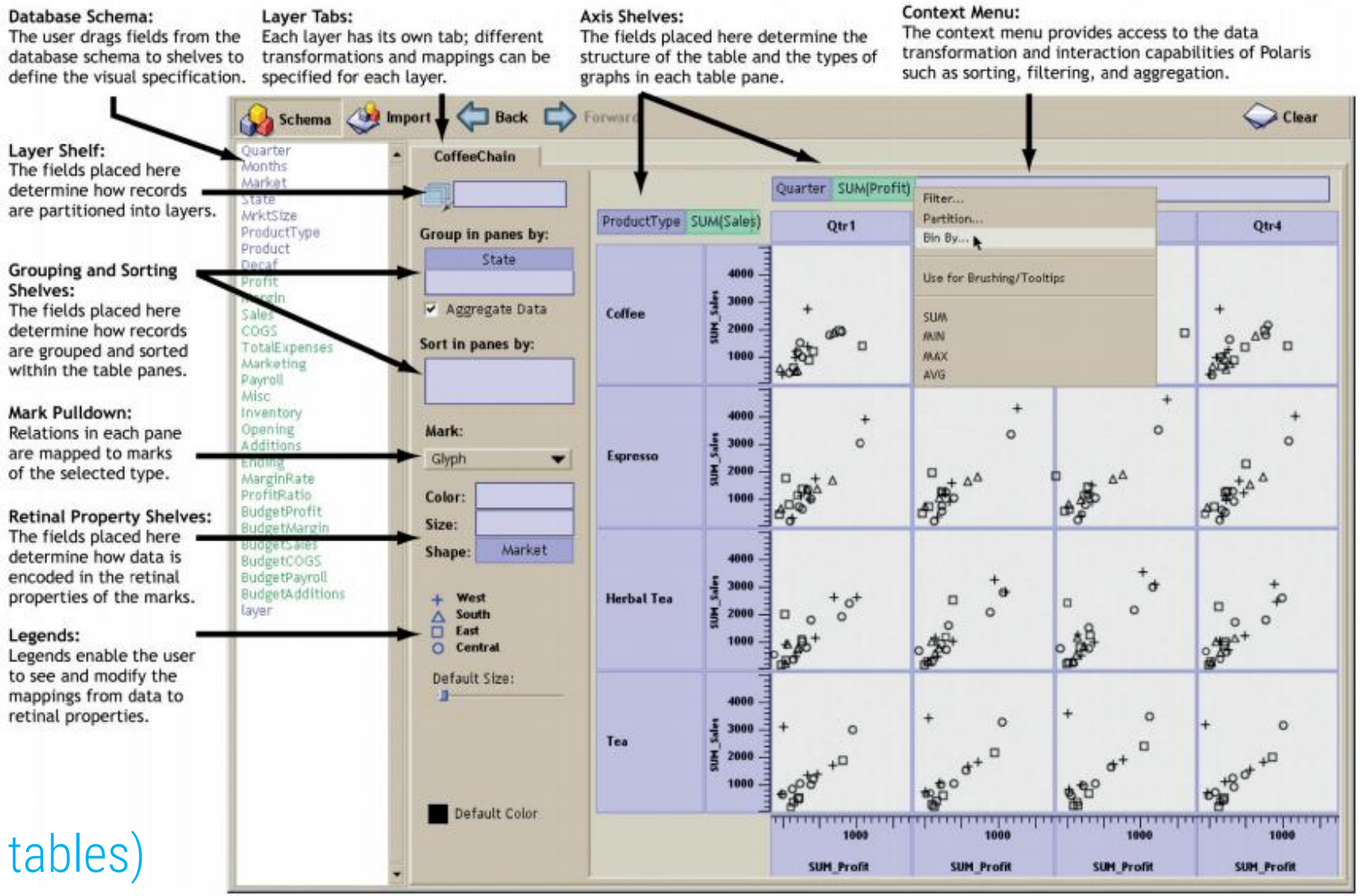
- likely no a-priori information about the data
- not sure about patterns and information present
- explore to create hypotheses & confirm later

**focus of visualization**

# DATA EXPLORATION / INSIGHT

Often with expert / complex tools

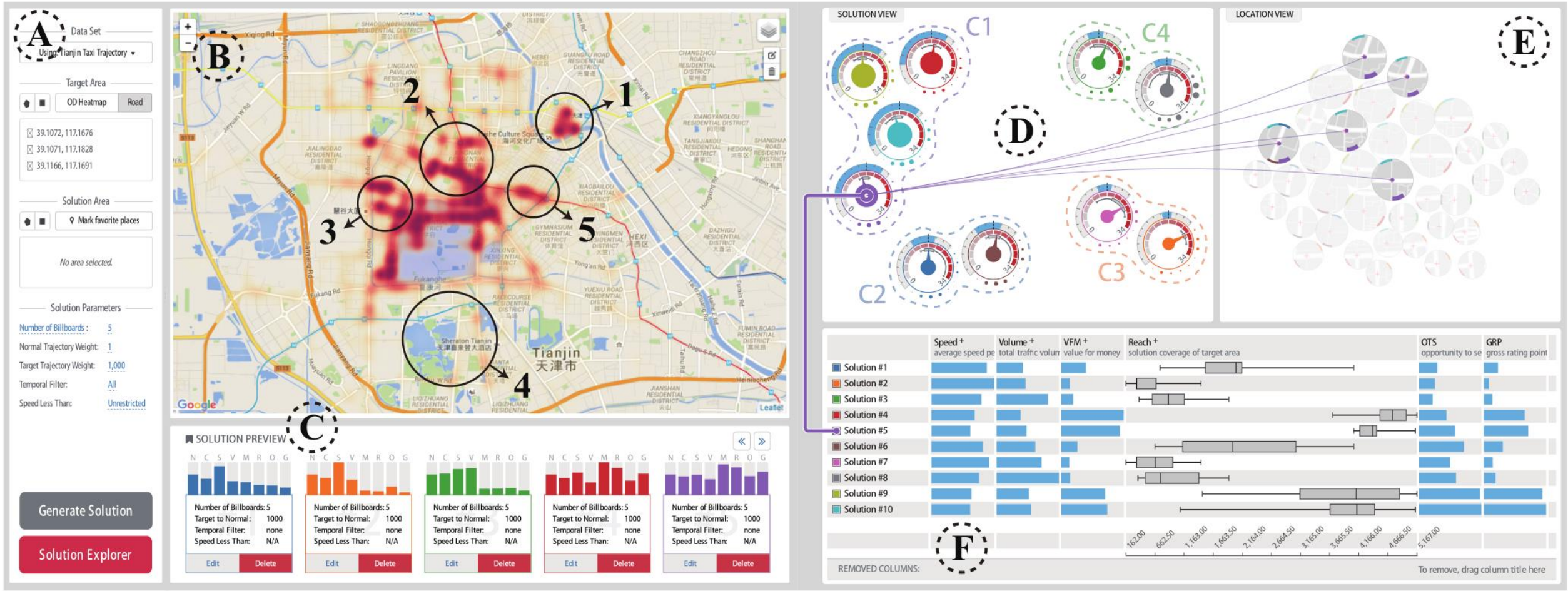
POLARIS




Either for any data (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

# PERSUATION, EDUCATION, MOTIVATION

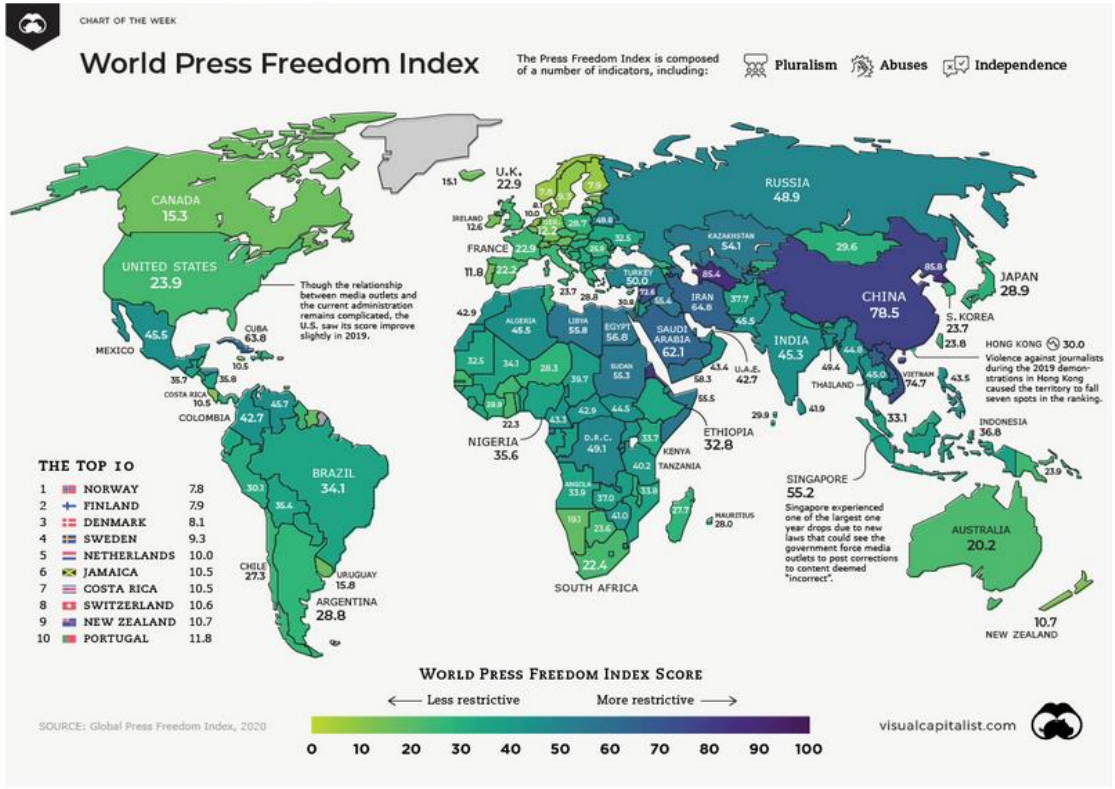


POLITICS

## Mapped: The State of Press Freedom Around the World

Published 10 months ago on May 1, 2020
By Nick Routley

*View a more detailed version of this map*




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

"All the News That's Fit to Print"

# The New York Times

Late Edition

Today, plenty of sunshine, chilly, high 33. Tonight, increasing cloudiness, low 23. Tomorrow, intervals of snow and rain, little to no accumulation, high 41. Weather map, Page 33.

VOL. CLXX ... No. 58,976 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 BOSMAN

CHICAGO — A nation numbed 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.

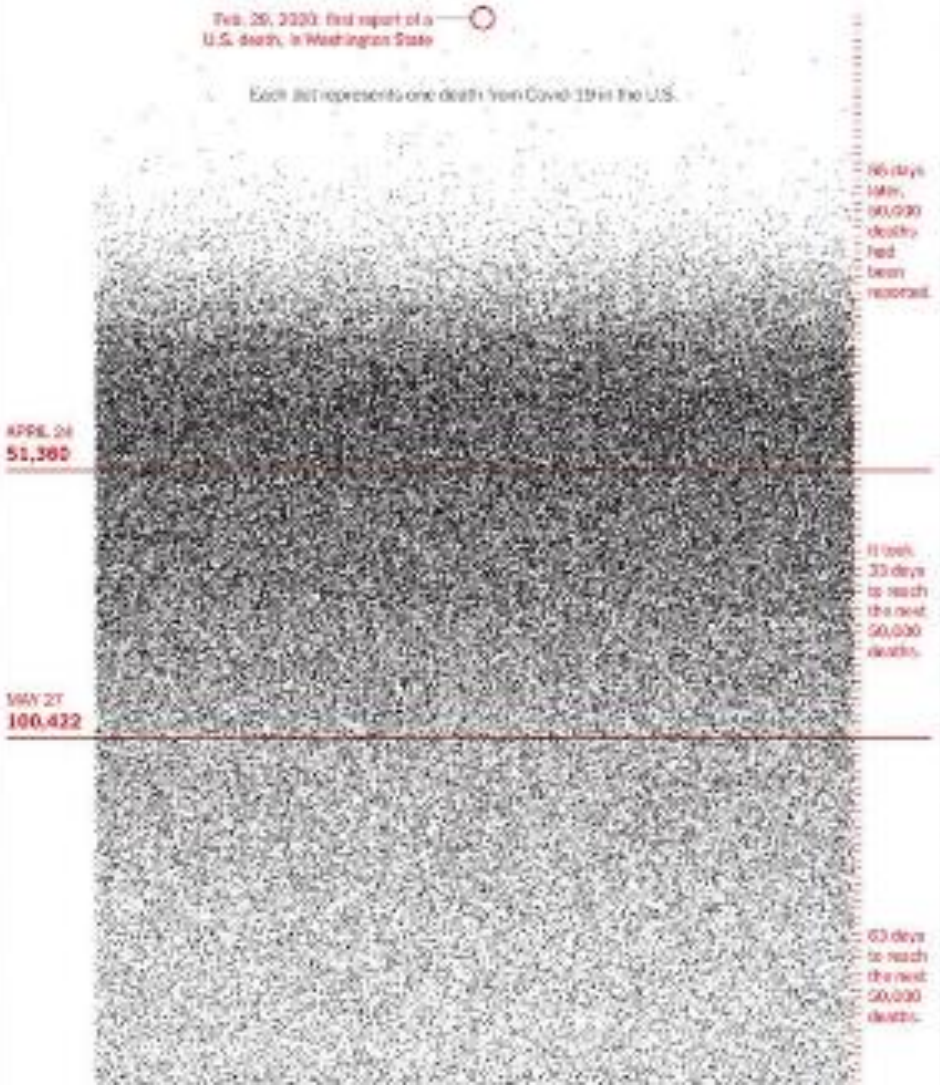
No other country has counted so many deaths in the pandemic. More Americans have perished from Covid-19 than on the battlefields of World War I, World War II and the Vietnam War combined.

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**




## *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 coat 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 oral 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 concave." His own daughters were 4 and 2 at the time.

The Oklahoma City case, he lat-

## 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, Rod Plasser 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. Rail 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 stress on the foundations of the country's economy: its network of roads and railways, drinking-water systems, power plants, electrical

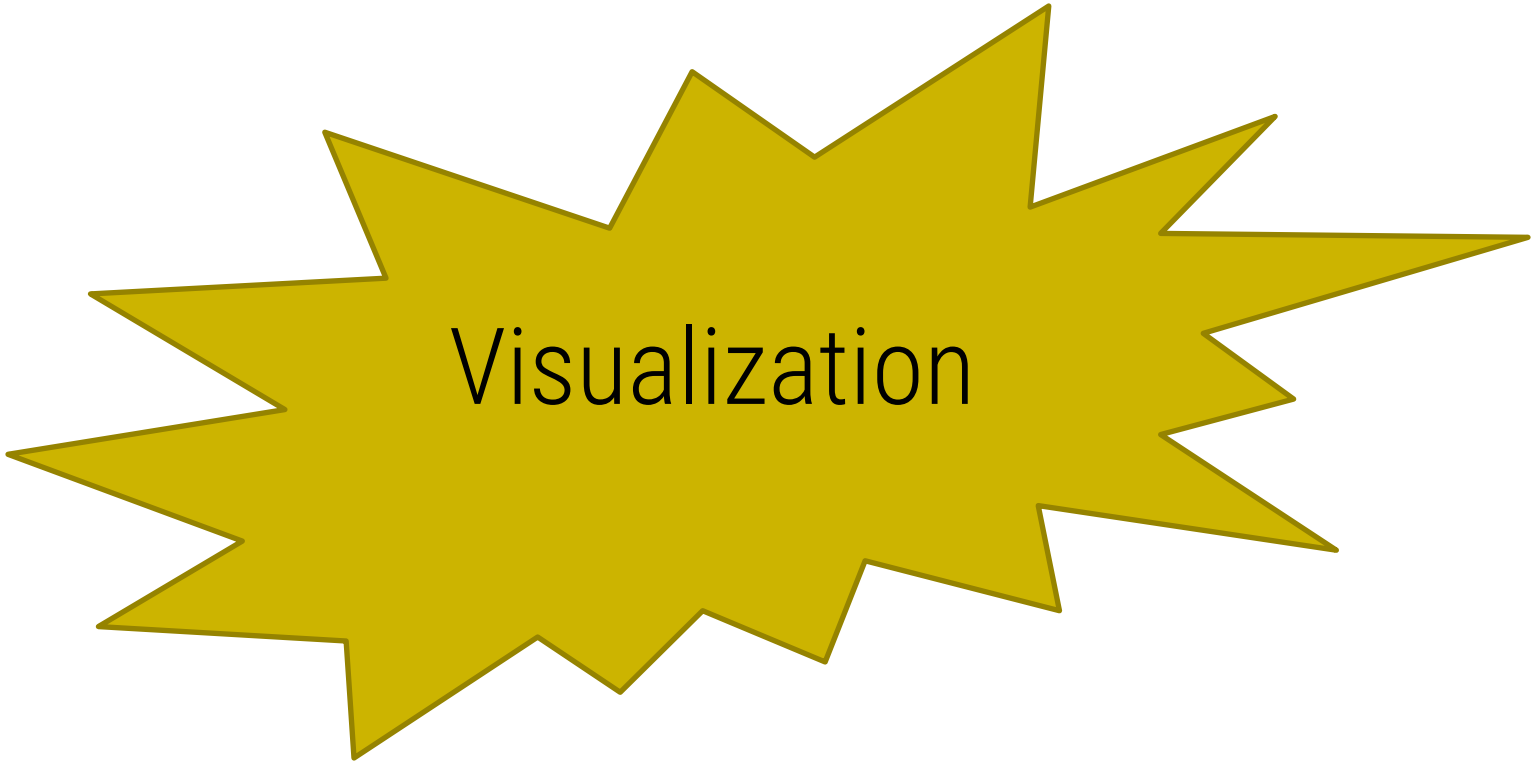

→ personal information and motivation

# BIG RESEARCH QUESTIONS 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 |

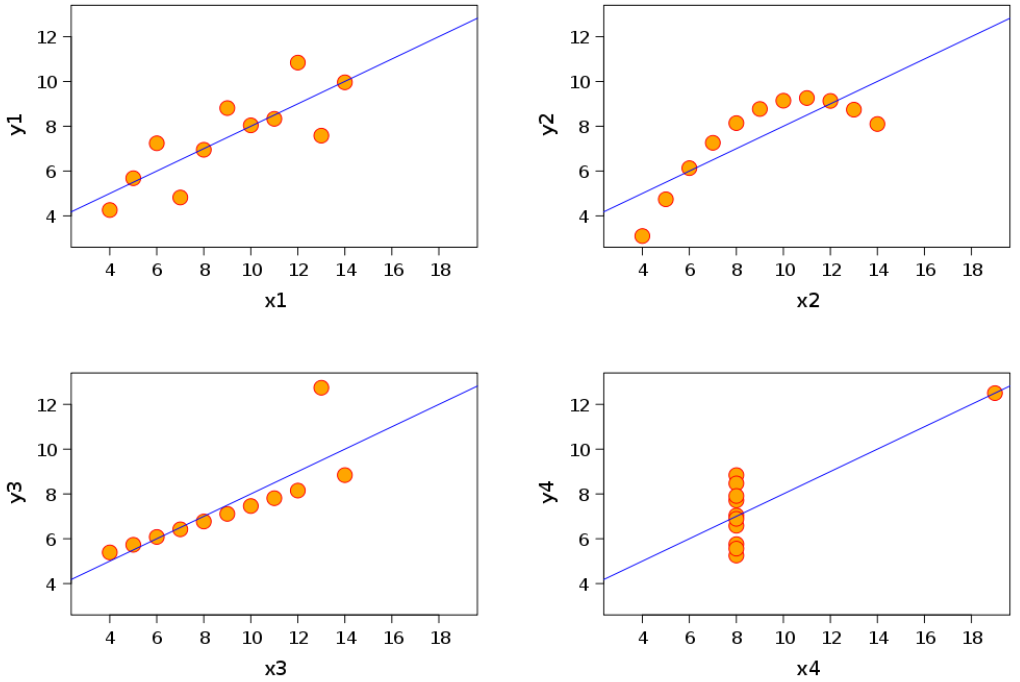
| | |
|---|---|
| Mean of *x* | 9.0 |
| Variance of *x* | 11.0 |
| Mean of *y* | 7.5 |
| Variance of *y* | 4.12 |
| Correlation between *x* and *y* | 0.816 |
| Linear regression line | $y = 3 + 0.5x$ |

[Source: Anscombe's quartet, Wikipedia]

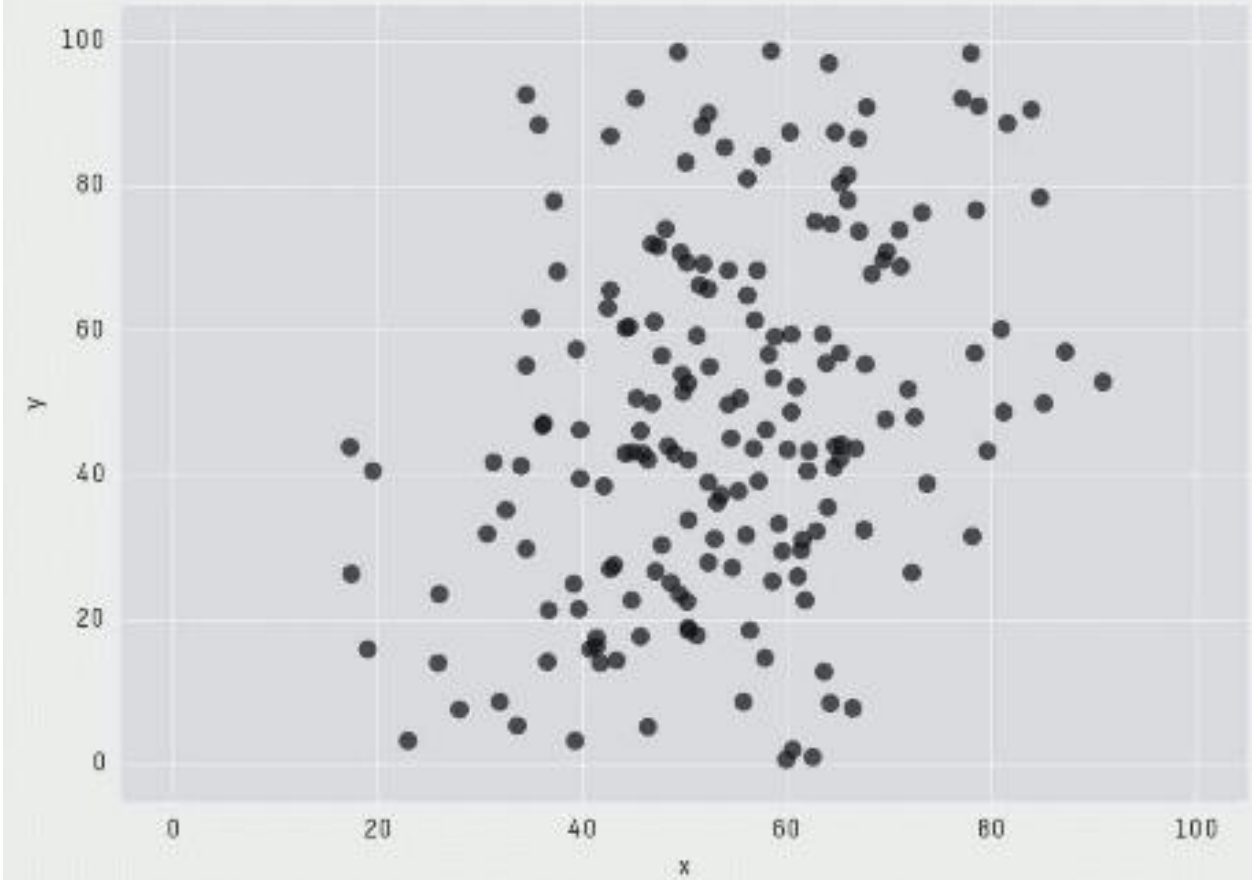
# VISUAL REPRESENTATION OF THE DATA

## Visual representation reveals a different story

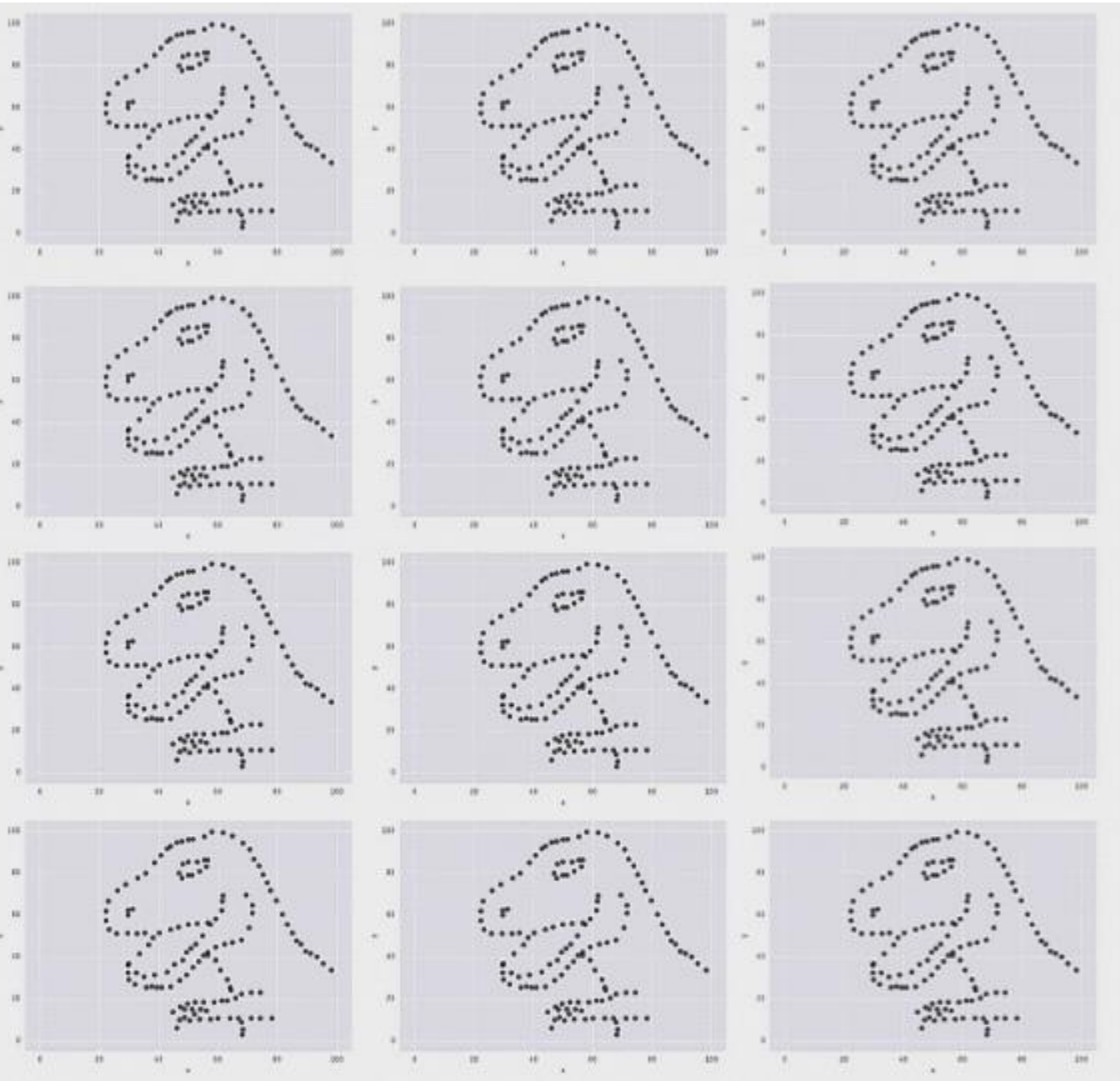
| I | | II | | III | | IV | |
|---|---|---|---|---|---|---|---|
| x | y | x | y | x | y | x | y |
| 10.0 | 8.04 | 10.0 | 9.14 | 10.0 | 7.46 | 8.0 | 6.58 |
| 8.0 | 6.95 | 8.0 | 8.14 | 8.0 | 6.77 | 8.0 | 5.76 |
| 13.0 | 7.58 | 13.0 | 8.74 | 13.0 | 12.74 | 8.0 | 7.71 |
| 9.0 | 8.81 | 9.0 | 8.77 | 9.0 | 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 |

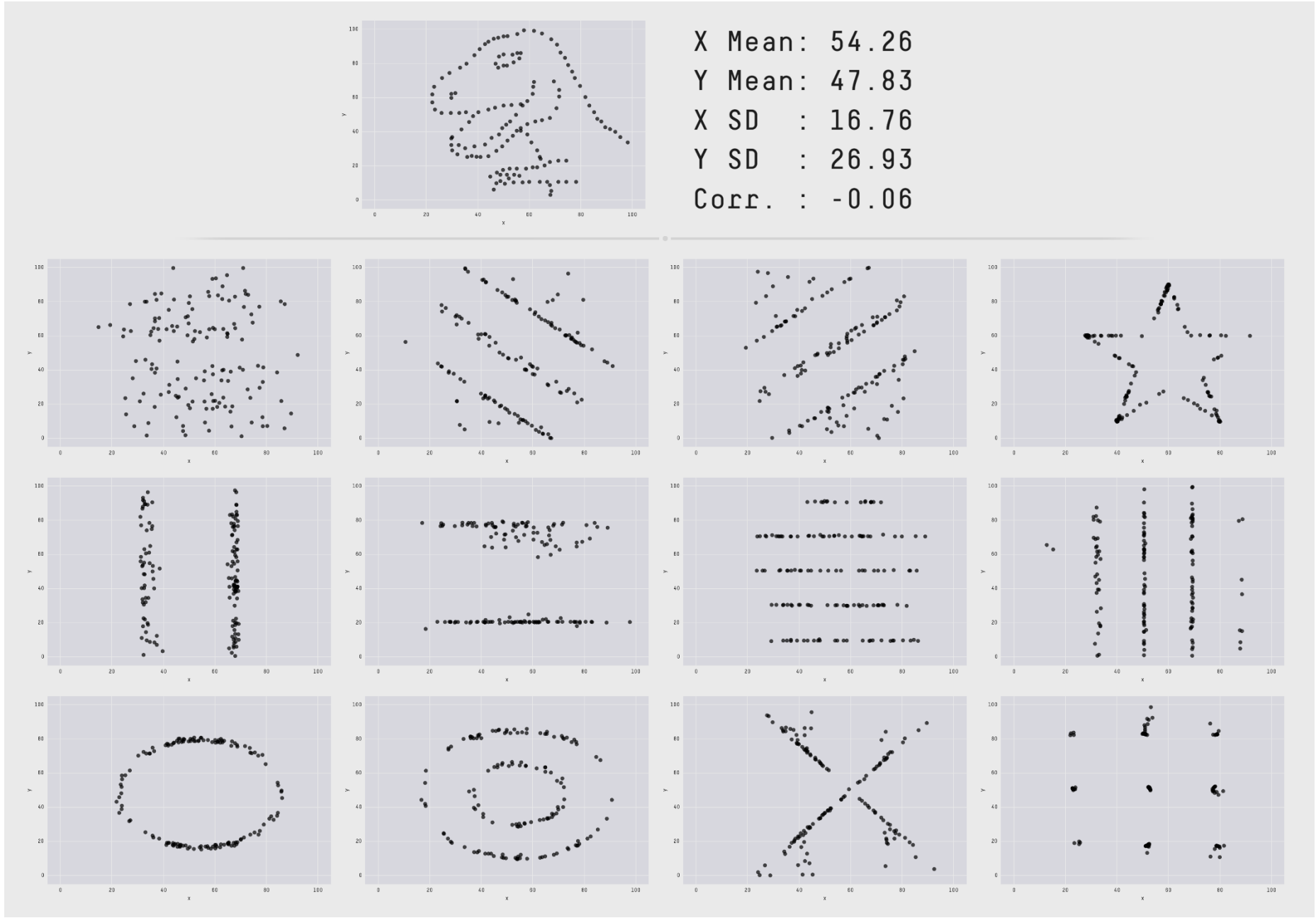



[Source: Anscombe's quartet, Wikipedia]




X Mean: 54.0236753
Y Mean: 48.0970794
X SD : 14.5298540
Y SD : 24.7943127
Corr. : +0.3280926



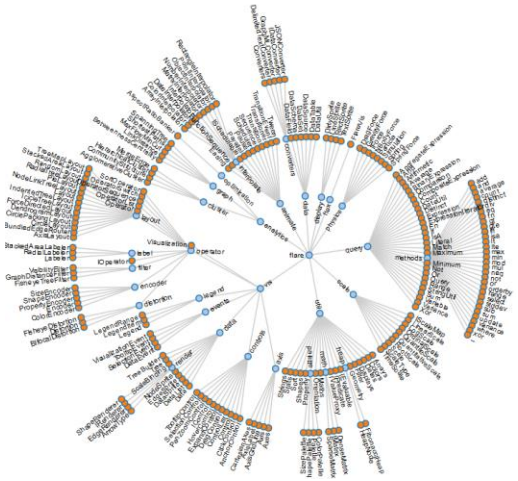



https://www.autodeskresearch.com/publications/samestats

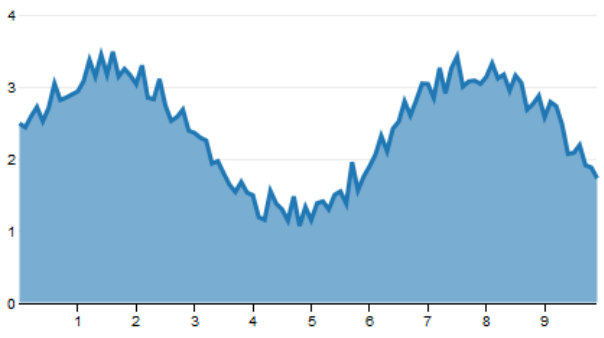
# Why visual data representations?

- 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:

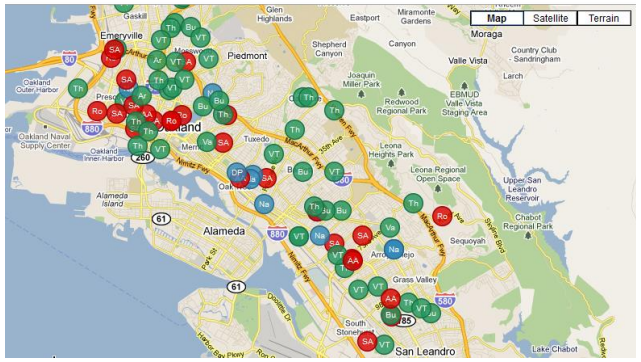


graphs / hierarchies




charts

Crimespotting

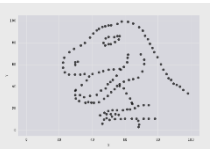



maps

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

# Other benefits of visualization

- expand human working memory
  - offload cognitive resources to the visual system
- reduce search
  - by representing a large amount of data in a small space
- enhance the recognition of patterns
  - by making them visually explicit
- aid monitoring of a large number of potential events
- provides a manipulable medium & allows exploration of a space of parameter values

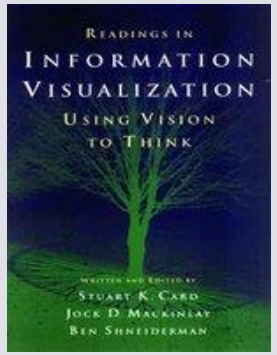


# Visualization

- Create visual representation
- Includes interaction

Official Definition:

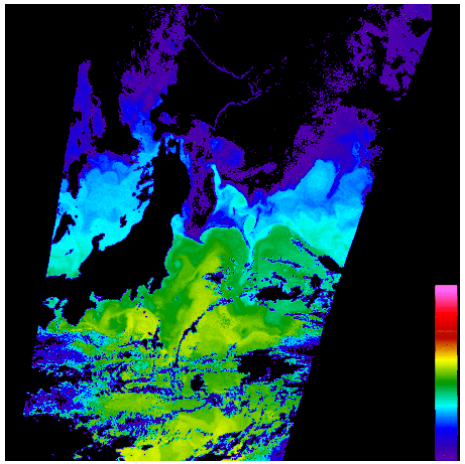
*The use of computer-supported, interactive, visual representations of data to amplify cognition.*
[Card et al., 1999]

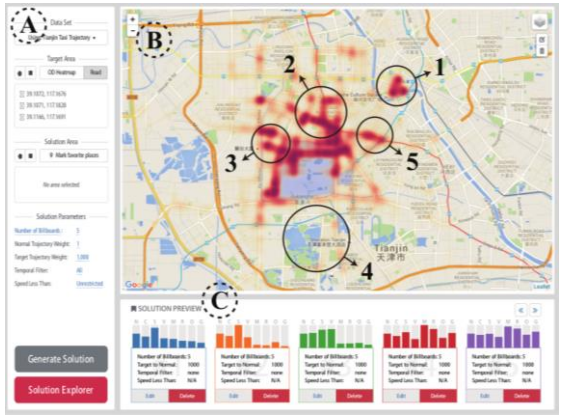



Hint: but there is more to it

# 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

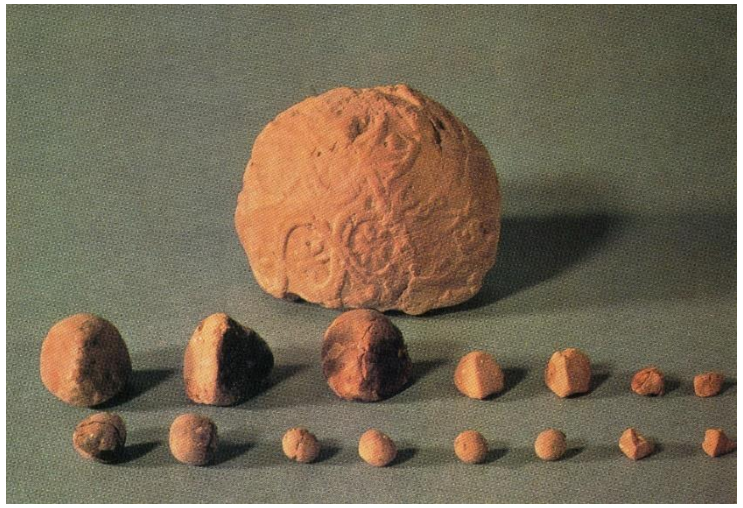





# HISTORICAL EXAMPLES

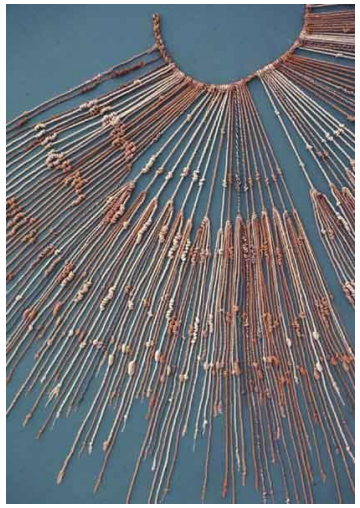
Visualization of data has been practiced for hundreds of years…

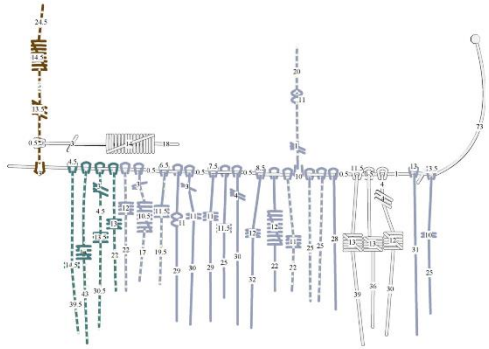
# VERY EARLY EXAMPLES



**5500 BC – Mesopotamian Clay Tokens**

tokens that externalize information, support visual thinking, and enhance cognition way before paper and writing were invented





**2600 BC – Quipus**

assemblies of knotted ropes that were used in South America as a data storage; color, relative position of knots, knot types, and rope length were used to encode categorical and quantitative variables

**500 BC – Pebble Voting**

the earliest participatory visualizations were probably voting systems

# 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




**JOHN SNOW, 1854**

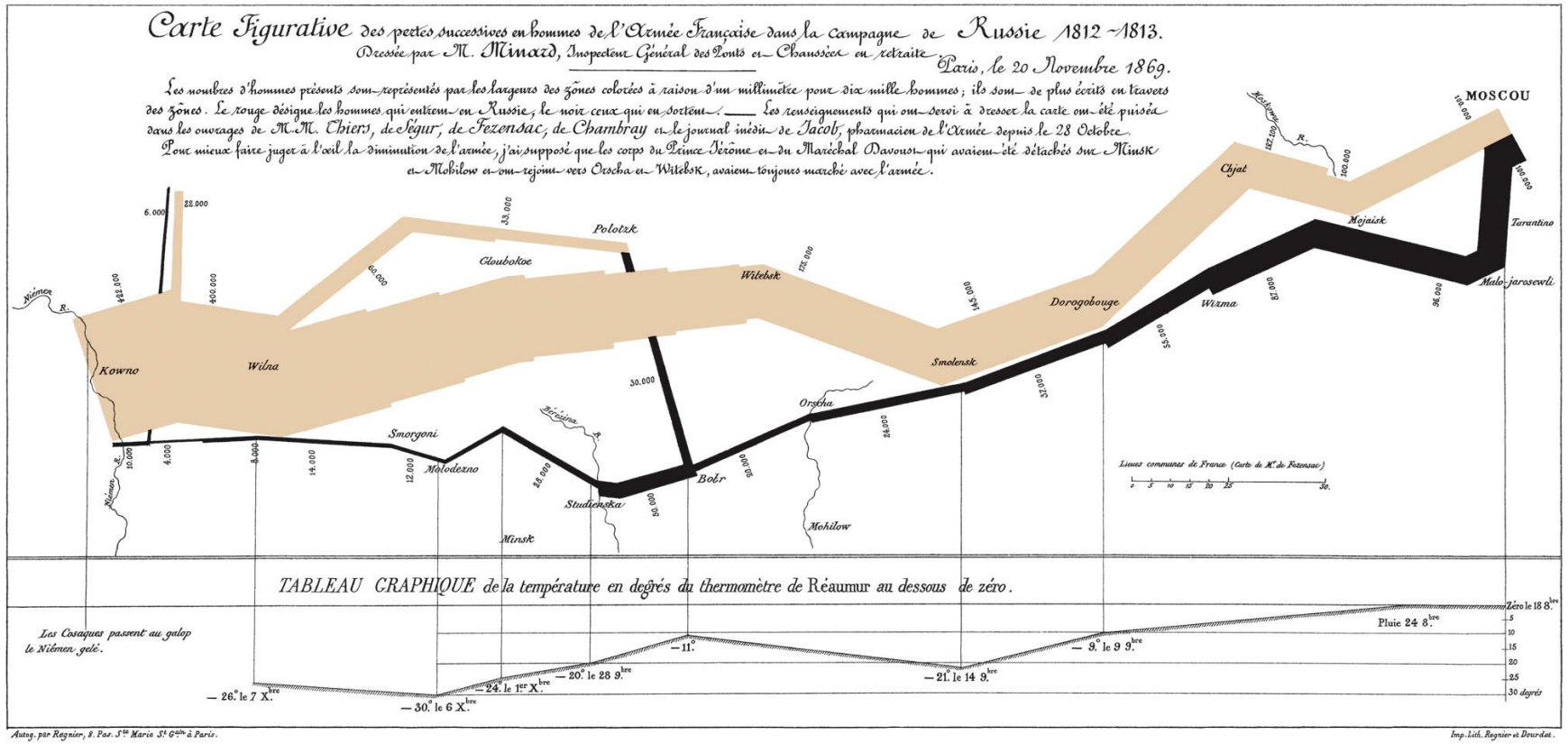
Replica memorial of the Broadway street pump

by Wikipedia author Jamzze

# NAPOLEON'S MARCH ON MOSCOW

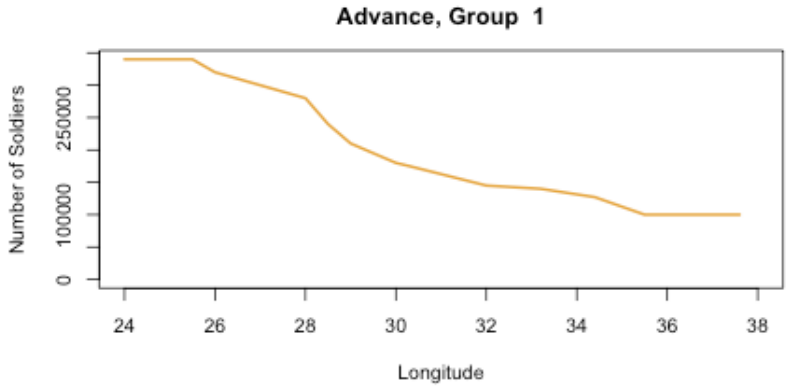
Named the best statistical graphic ever drawn (by Edward Tufte)

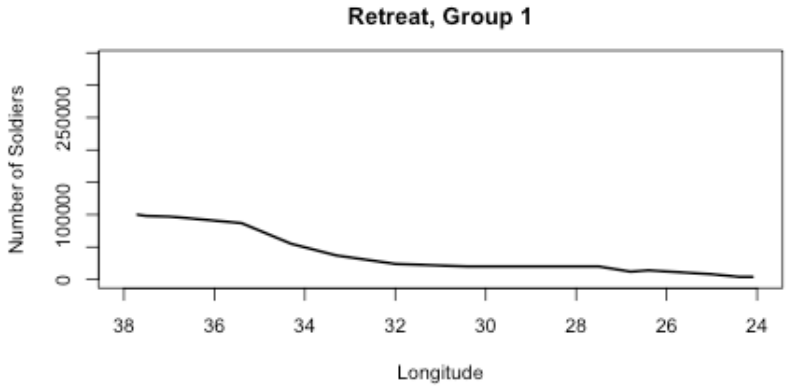
- Includes: spatial layout linked with stats on: army size, temperature, time
- Tells a story in one overview

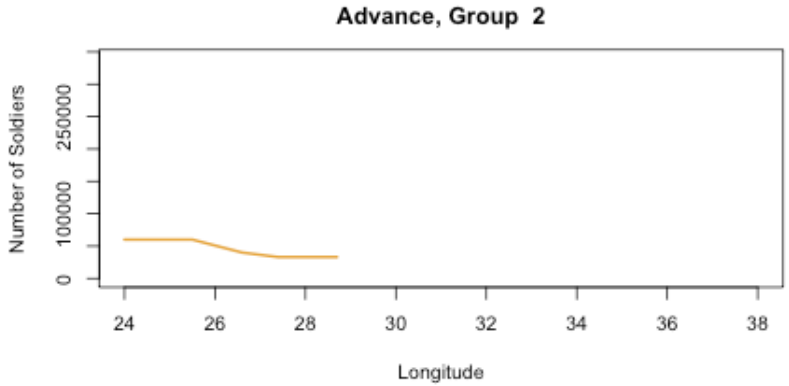


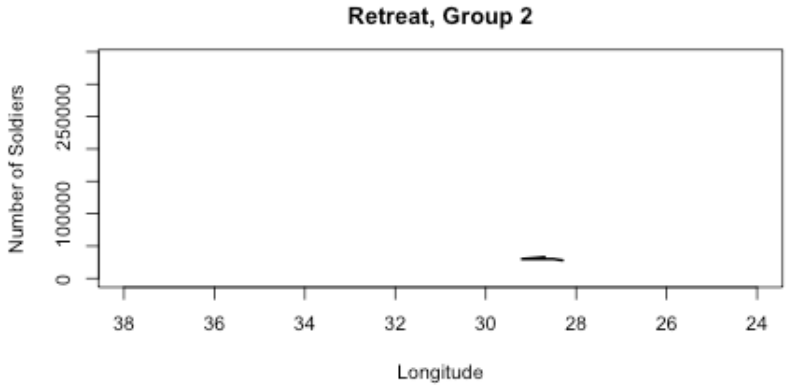

Charles Minard, 1869

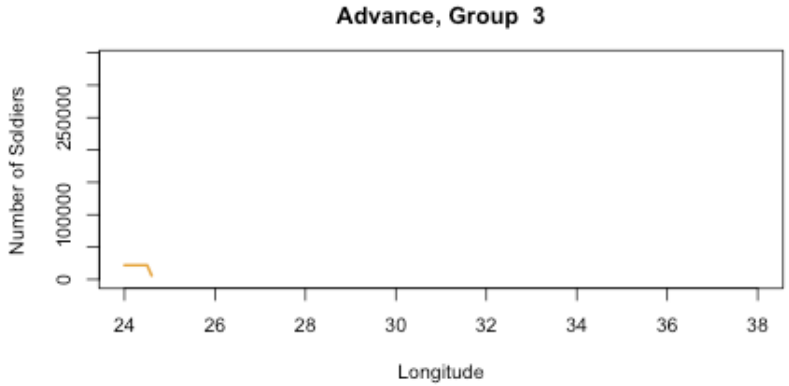
More info: The Visual Display of Quantitative Information (Tufte)

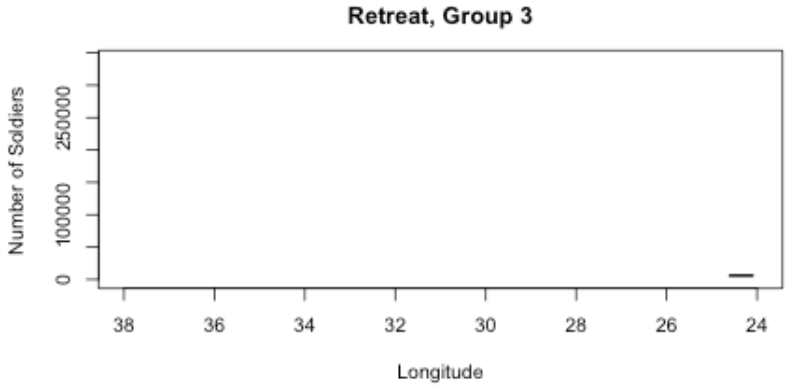


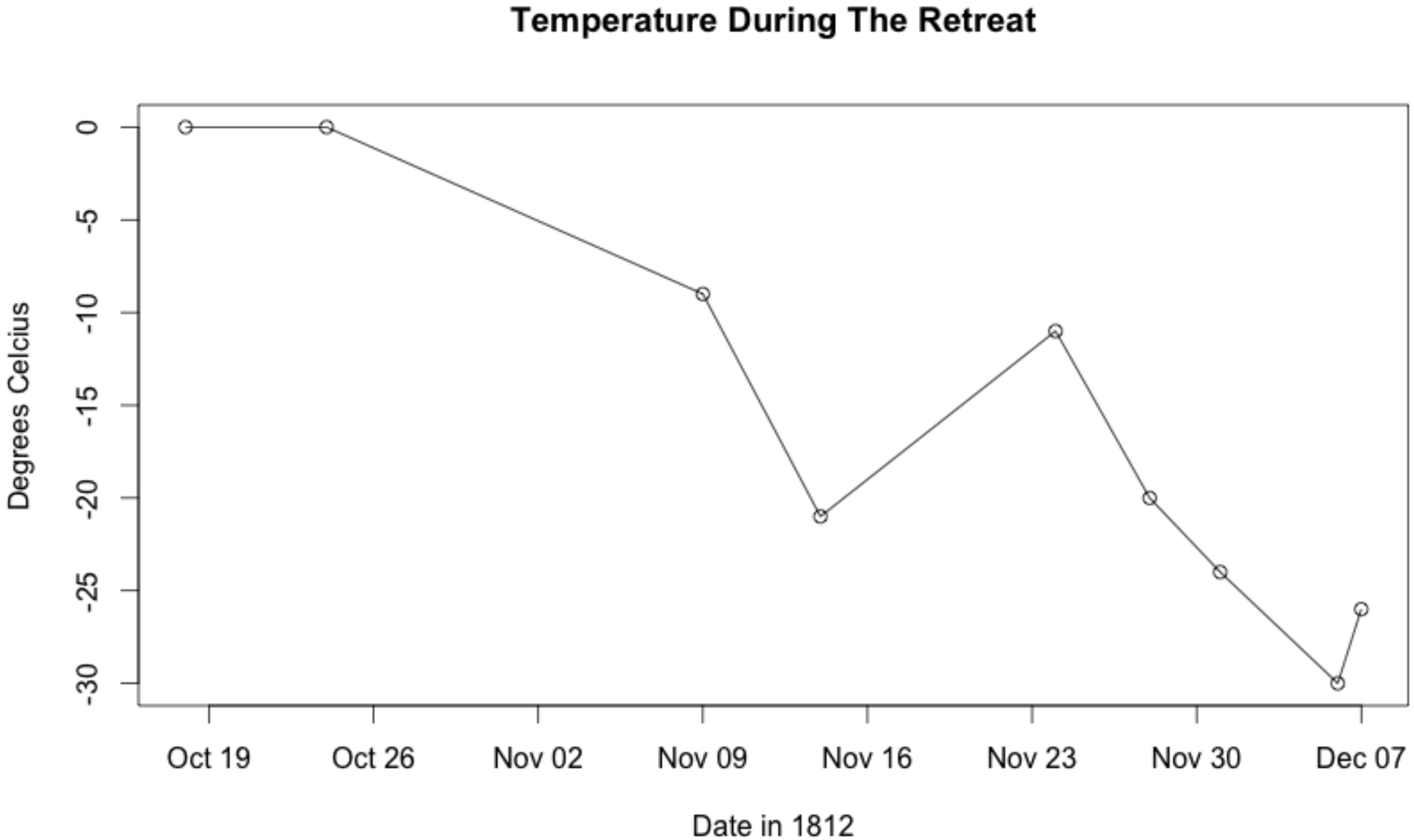



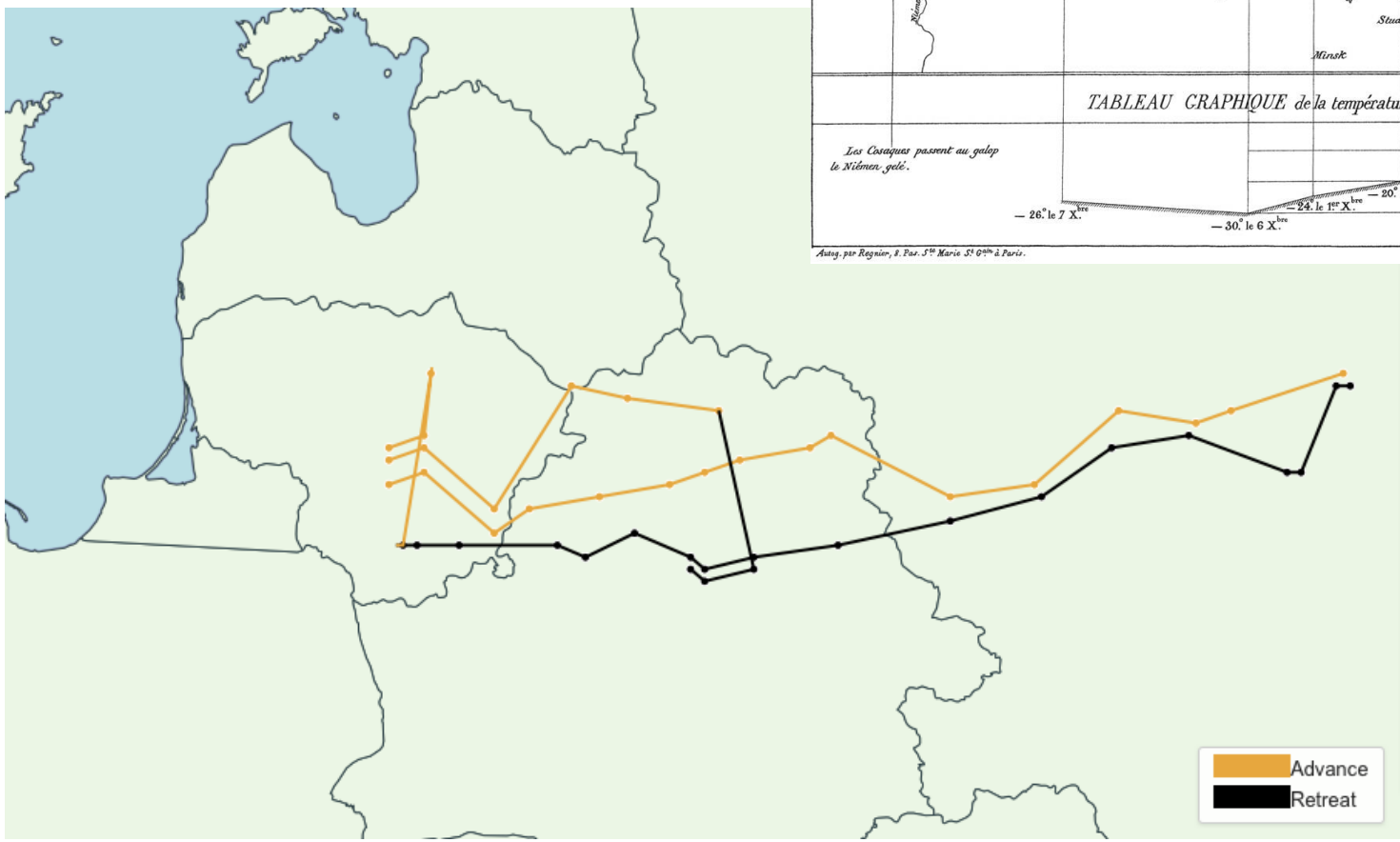




https://thoughtbot.com/blog/analyzing-minards-visualization-of-napoleons-1812-march




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


Advance
Retreat




# 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.



## 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 1.er 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

Autog. par Regnier, 8. Pas. S.te Marie S.t Gath. à Paris.

Imp. Lith. Regnier et Dourdet.

# TODAY

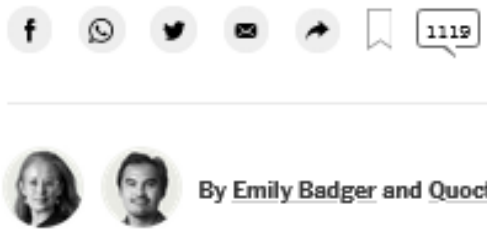
Visualizations are done by many people…


Professionals (journalists, analysts, …)

TheUpshot

# *Where Have All the Houses Gone?*

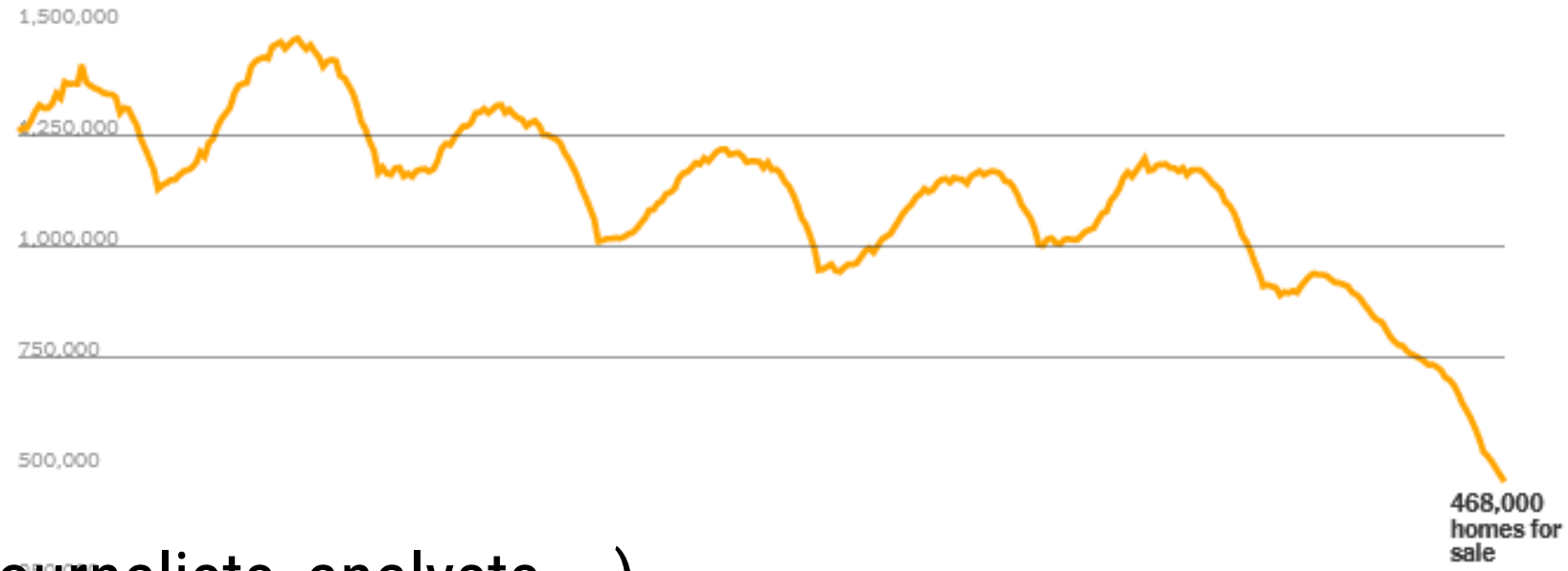
The inventory of homes for sale is startlingly low. The pandemic is part of the reason, but it's not the whole story.



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, …)

KANTAR
Information is Beautiful
Awards 2019

About News Awards Entry Showcase Sponsor Attend the Awards ceremony

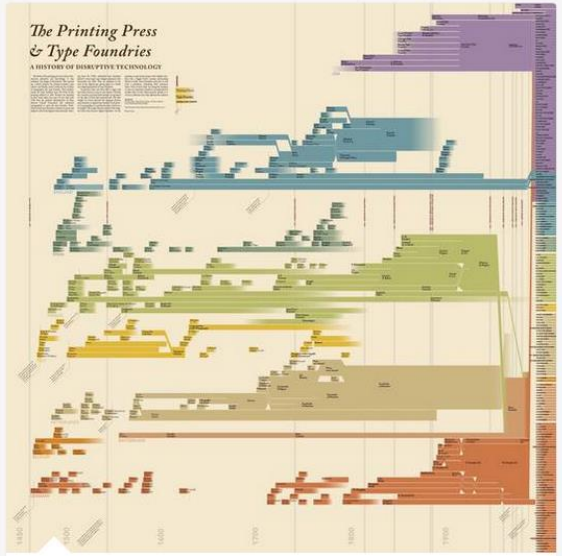
Awards Challenges

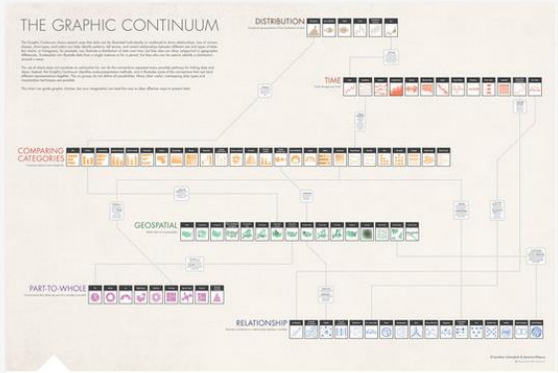
All 2019 2018 2017 2016 2015 2014 2013 2012

Gold Silver Bronze Rising Star Outstanding Individual Best-Non-English-Language Outstanding Outfit Student

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




**The Graphic Continuum**

The Graphic Continuum is our view of the many different types of visualizations available to us when we encode and present data. We've plotted nearly 90 different graphics across five main...


**Project Ukko - Seasonal Wind Predictions for the Energy Sector**

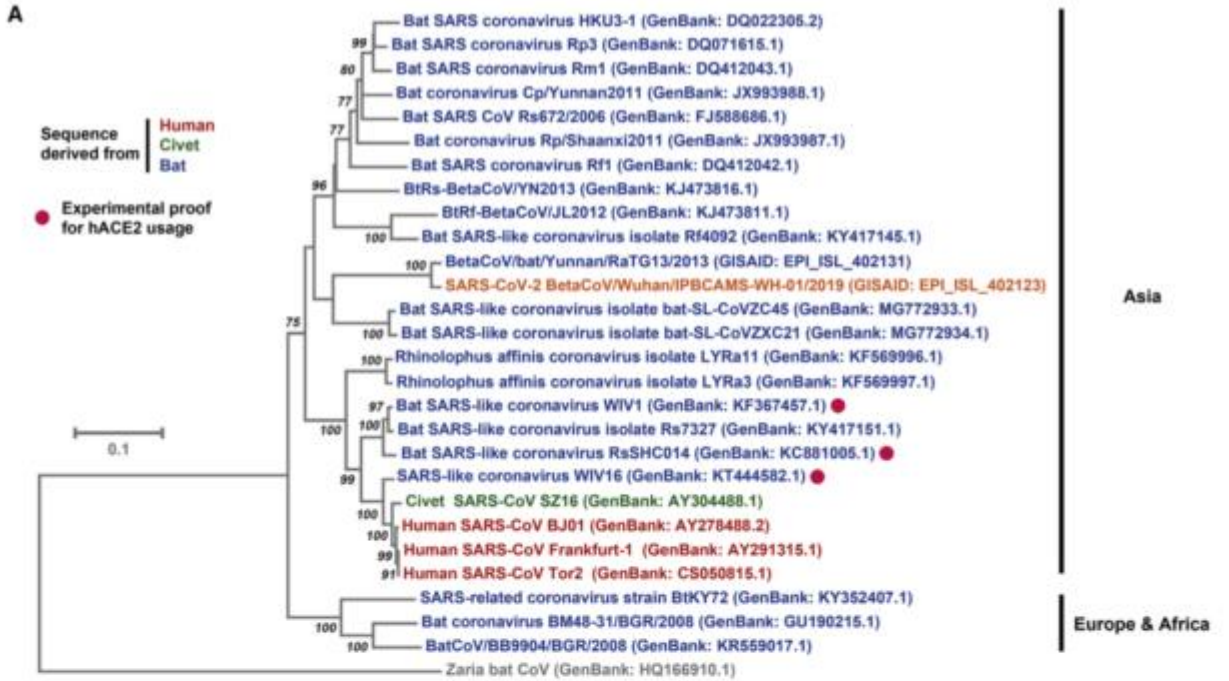
Weather forecasts predict future wind conditions only in the range of weeks. Climate predictions look at big changes over years and decades. However, for energy traders, wind farm managers and...

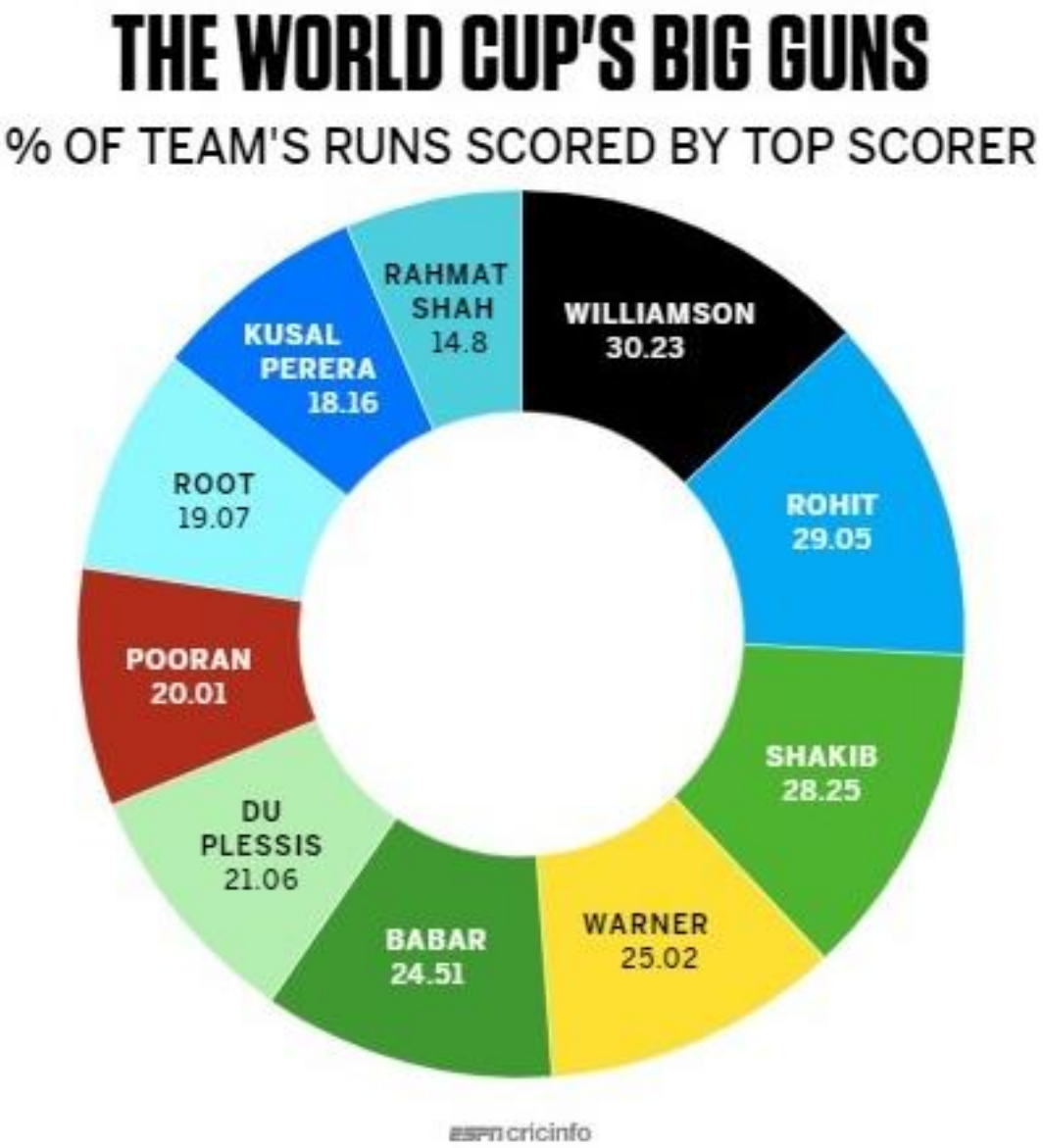
Google for beautiful examples

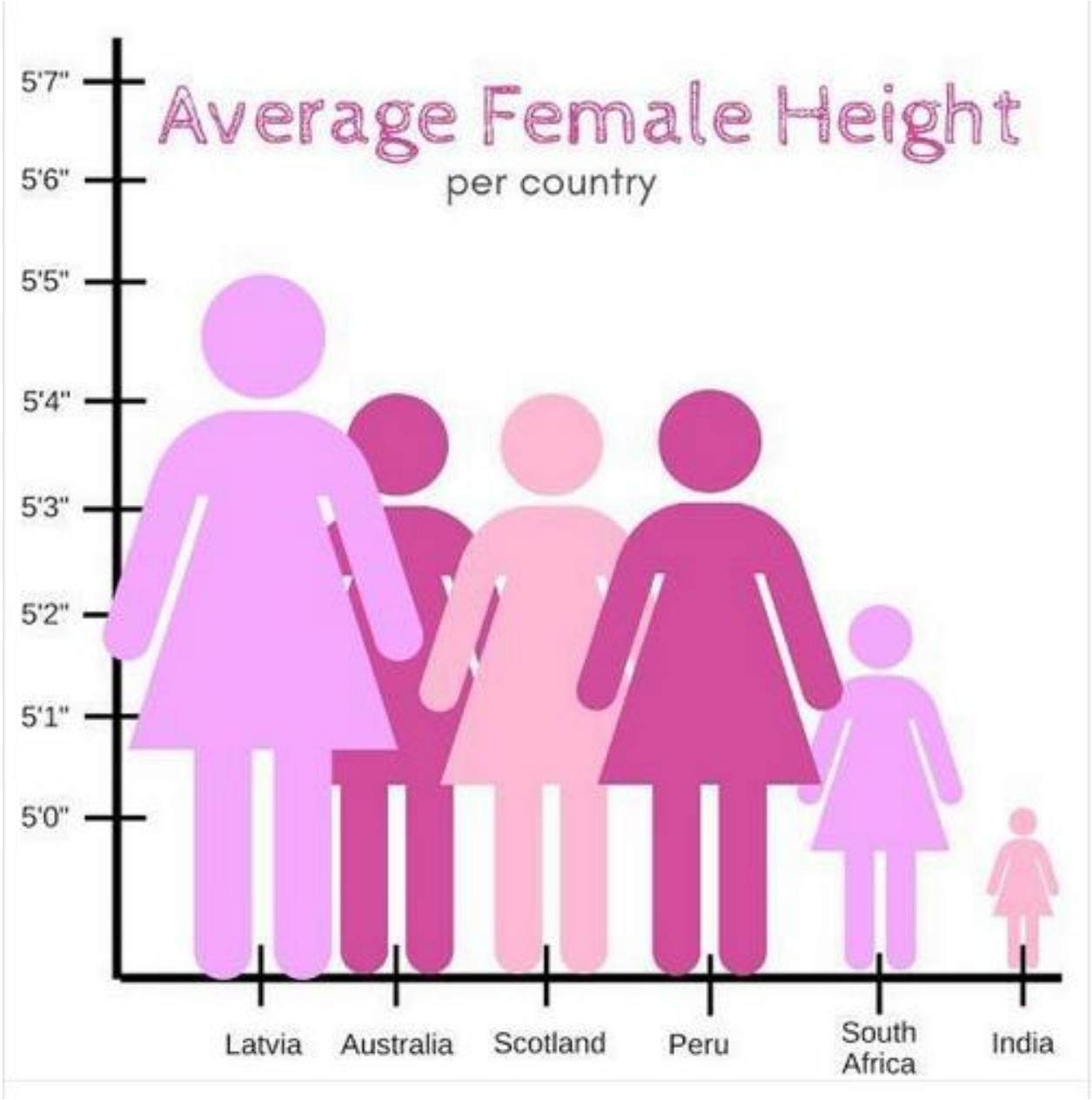
Enthusiasts, freelancer, …

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






People who shouldn't

# RESOURCES

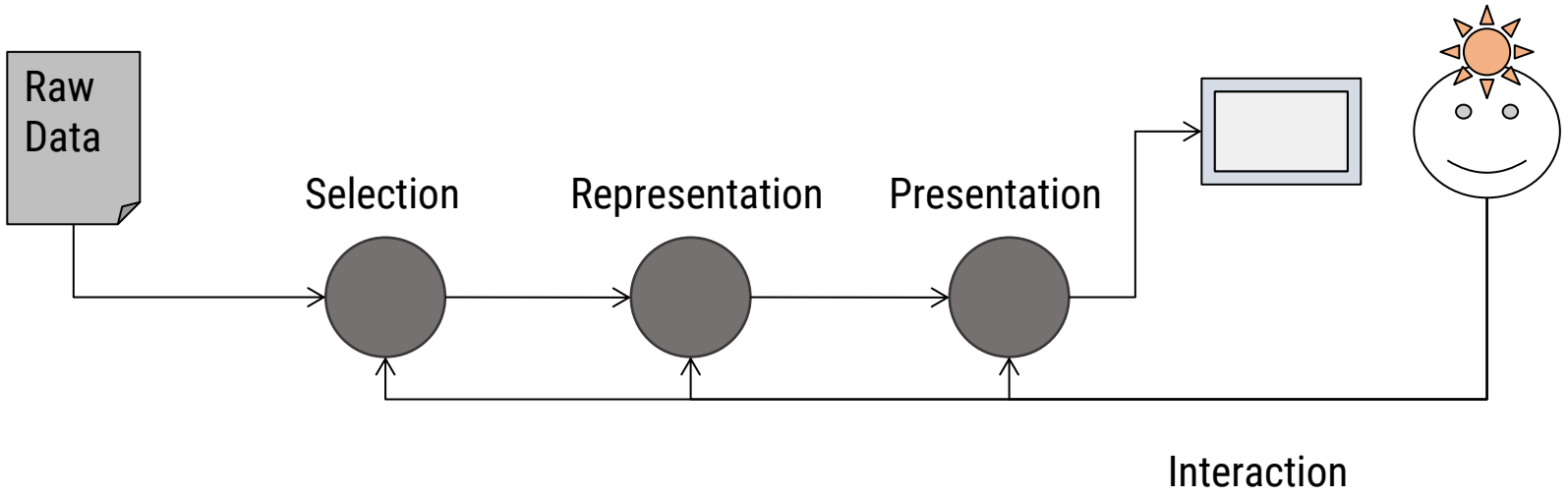
- Visualization conferences
  - ieeevis.org
- Blogs
  - http://eagereyes.org/
  - http://flowingdata.com/
  - http://www.informationisbeautiful.net/
  - https://www.visualisingdata.com/blog/
  - https://pudding.cool/
  - https://junkcharts.typepad.com/
  - https://badvisualisations.tumblr.com/
- Books
  - 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

# VISUALIZATIONS


Cem

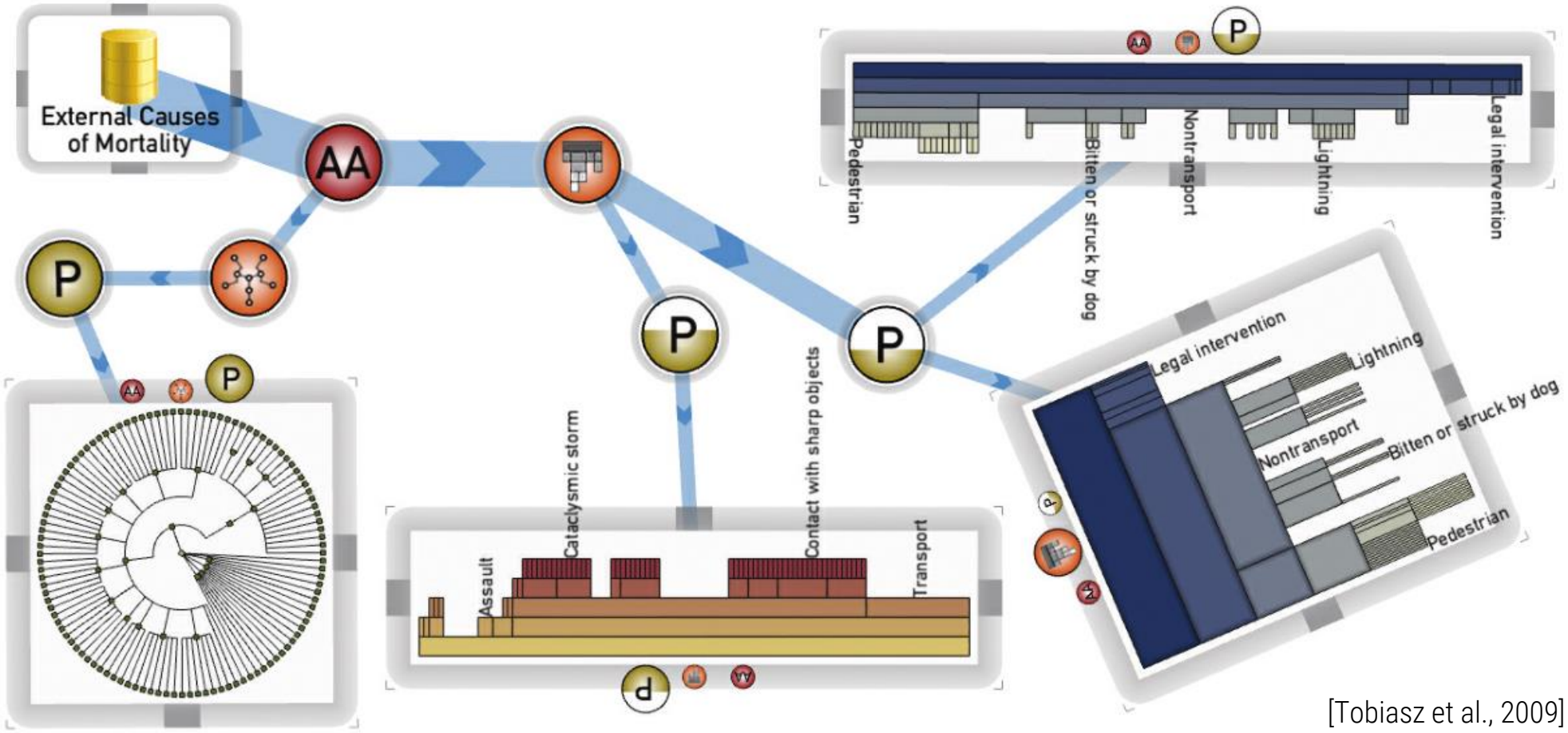
# HOW DO WE ARRIVE AT A VISUALIZATION?




The Visualization Pipeline

From [Spence, 2000]

# HOW DO WE ARRIVE AT A VISUALIZATION?




[Tobiasz et al., 2009]

# 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

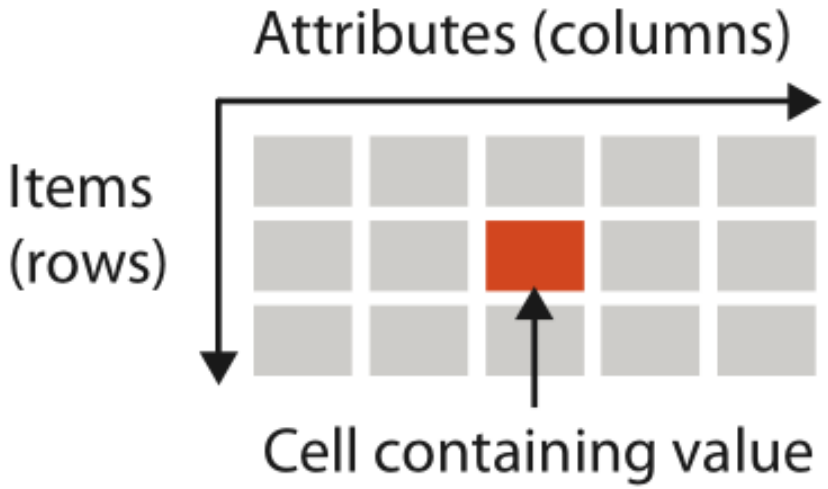
# DATA

- 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 AND ATTRIBUTE TYPES

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

# TABLES

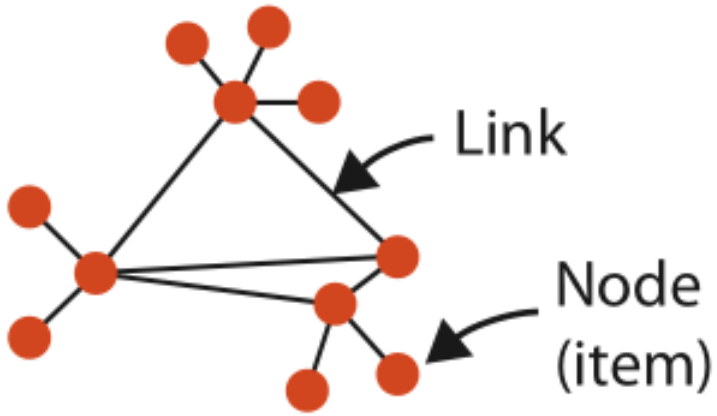



Example:

Items: drinks
Attributes: color, calories, name, …

[Munzner, Visualization Analysis and Design]

# NETWORKS




Example:

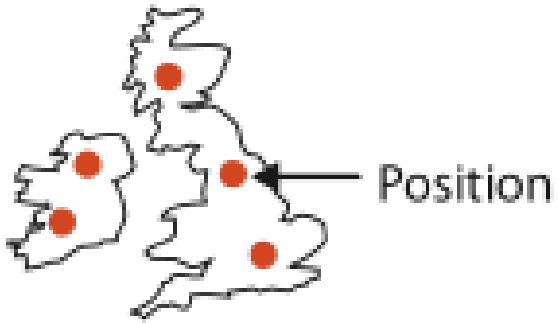
Item = nodes: people
Item = links: co-authorship

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

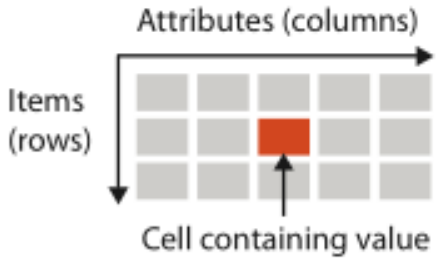



Item = countries
Positions = location on the planet

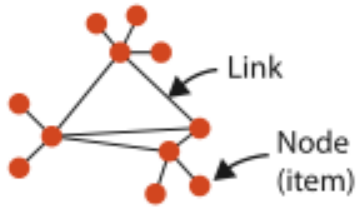
[Munzner, Visualization Analysis and Design]

## Dataset Types

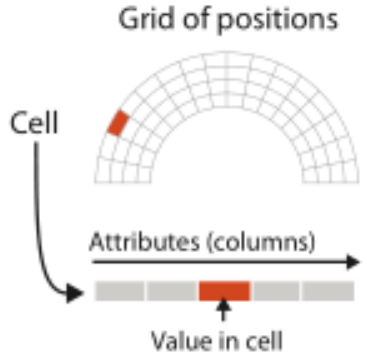
### → Tables




### → Networks

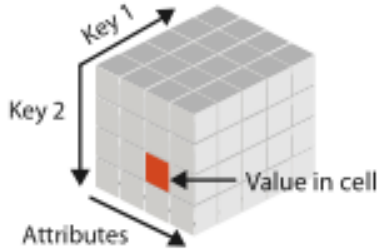



### → Fields (Continuous) (ordered)



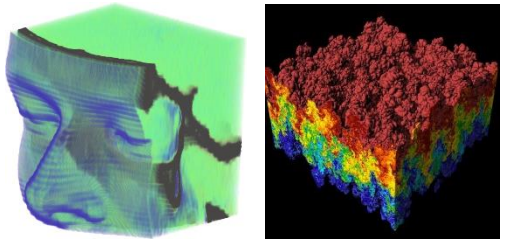

### → Geometry (Spatial)


### → *Multidimensional Table*




### → *Trees*

### → *3D Fields*



### → *3D Geometries*

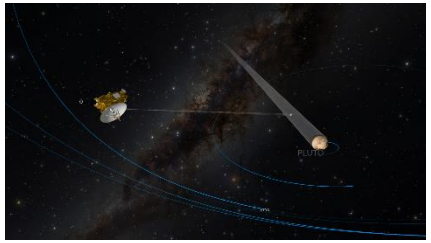
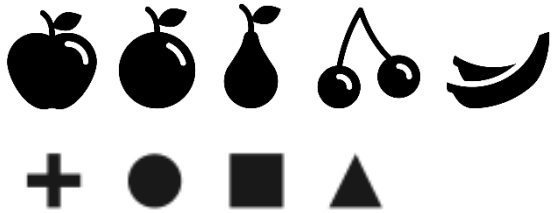


**Figure 2.4.** The detailed structure of the four basic dataset types.

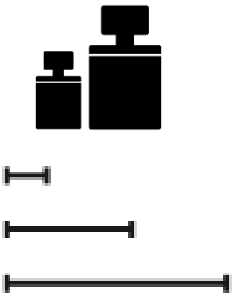
[Munzner, Visualization Analysis and Design]

# 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** calculated with: =, ≠, <, >
  - Quantitative
    - Counts and amounts, 5kg / 10kg
    - you can do =, ≠, <, >, - , +, x, ÷



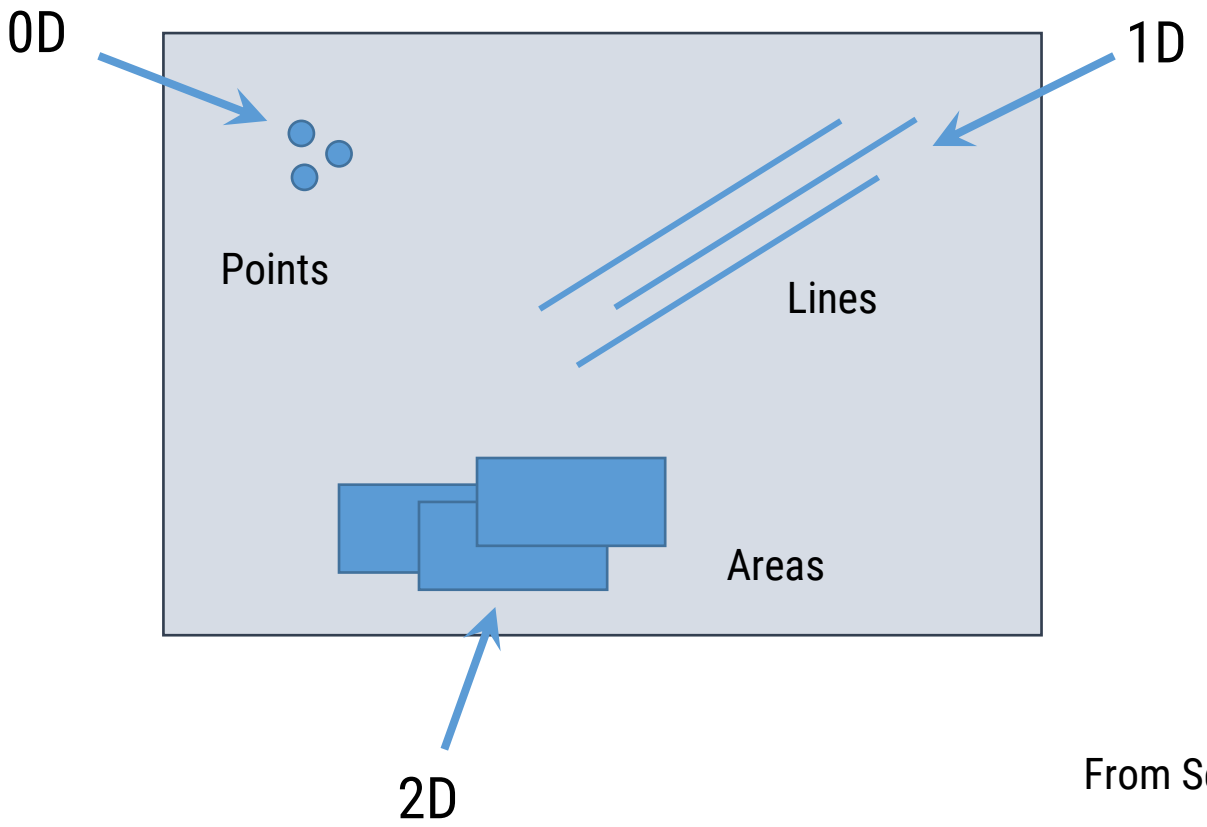
# WHY IS THIS IMPORTANT?

- 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

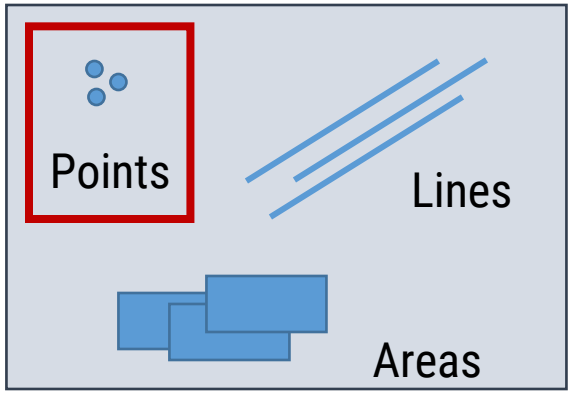
Basic geometric element data depict items or links




From Semiology of Graphics (Bertin)

# POINTS

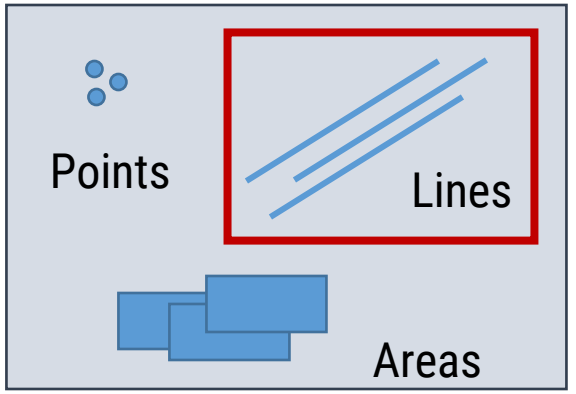
- “A point represents a location on the plane that has **no theoretical length or area**. This signification is independent of the size and character of the mark which renders it visible.”
- a location
- marks that indicate points can vary in all visual variables




From Semiology of Graphics (Bertin)

# LINES

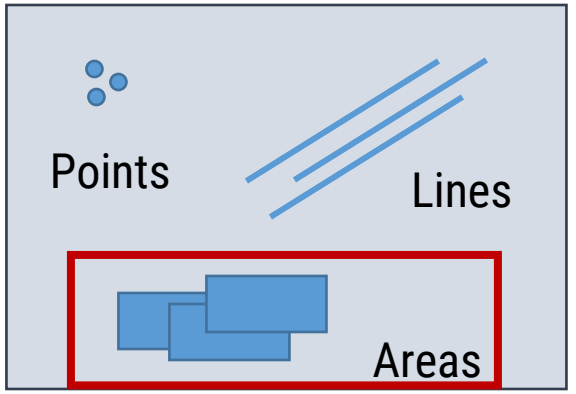
- “A line signifies a phenomenon on the plane which has **measurable length but no area**. This signification is independent of the width and characteristics of the mark which renders it visible.”
- a boundary, a route, a connection




From Semiology of Graphics (Bertin)

# AREAS

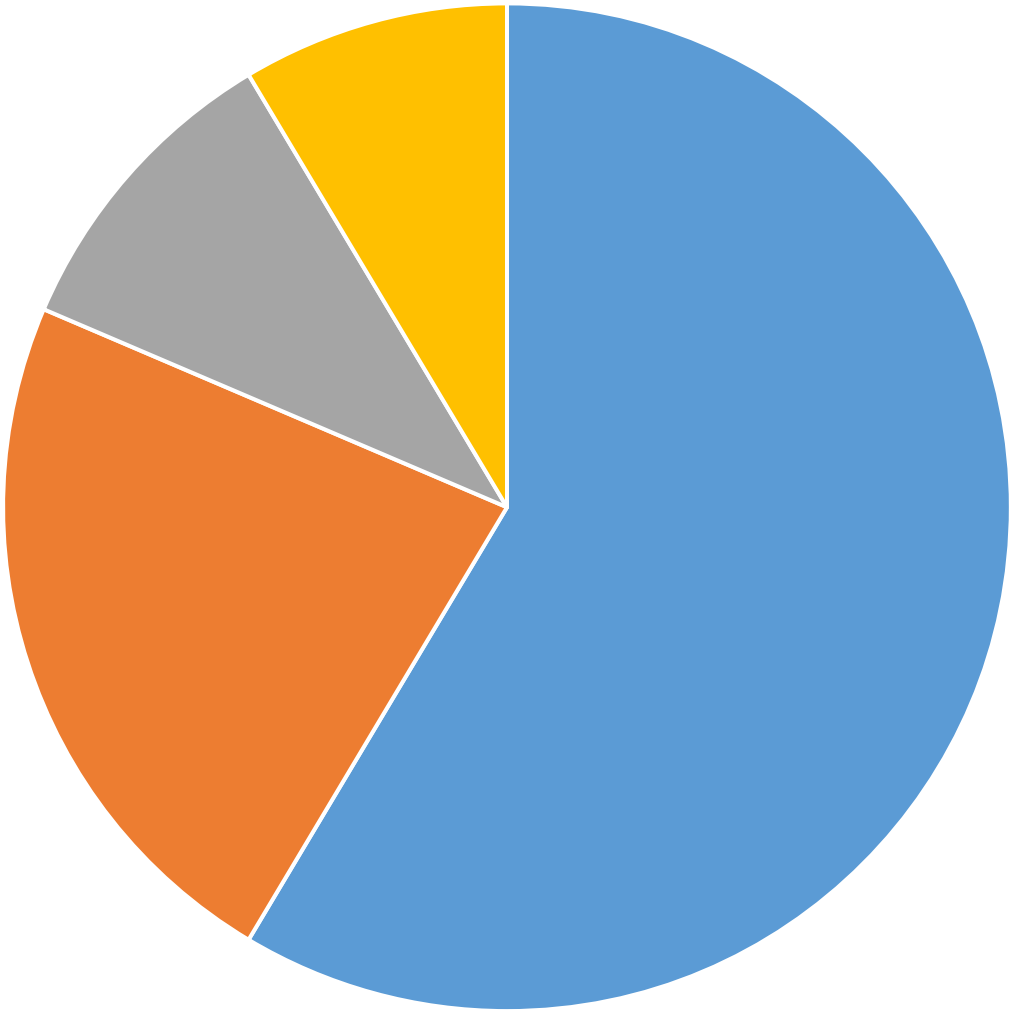
- “An area signifies something on the plane that **has measurable size**. This signification applies to the entire area covered by the visible mark.”
- an area can change in position but not in size, shape or orientation without making the area itself have a different meaning



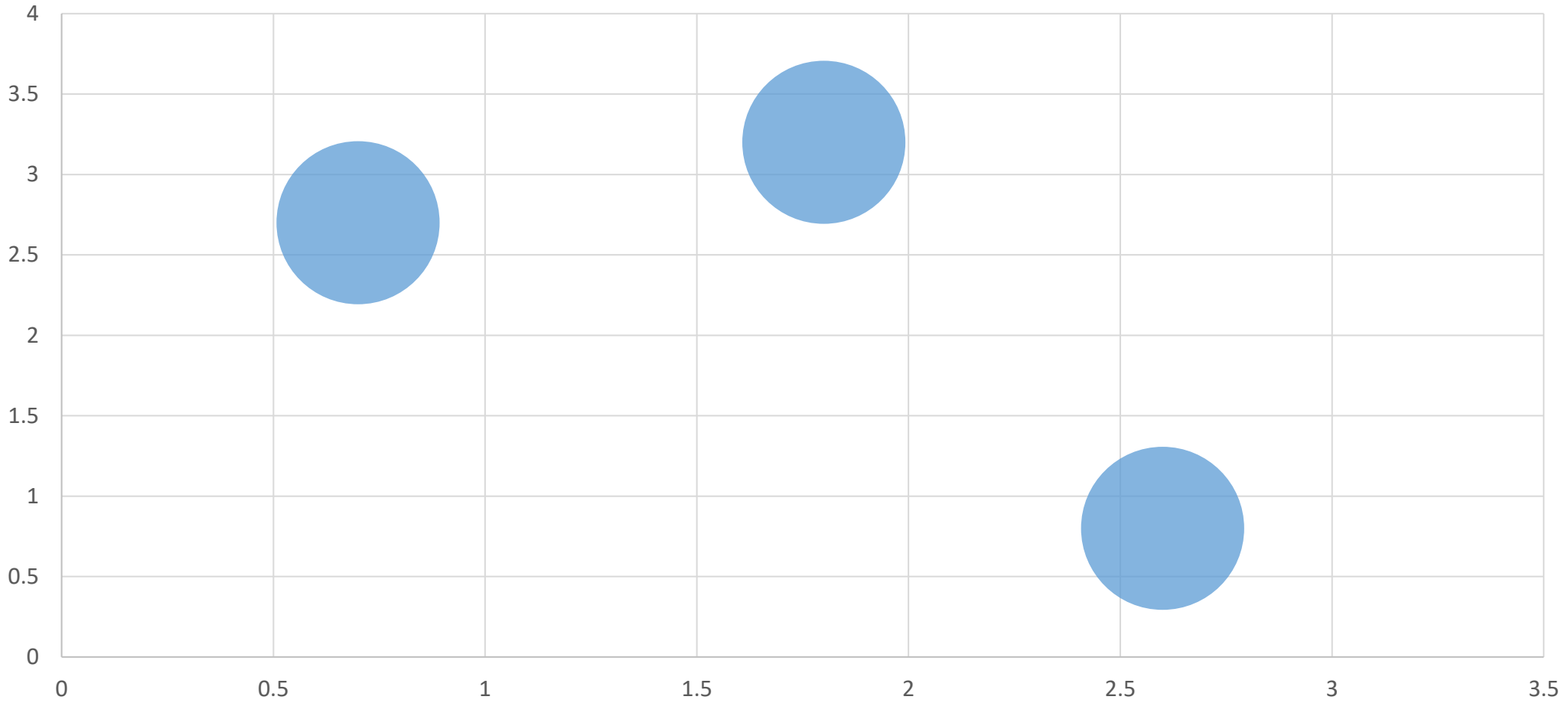

From Semiology of Graphics (Bertin)

# which marks do you see?

# which marks do you see?



# which marks do you see?



# VISUAL CHANNELS

Also often called visual variables

Modify marks
independent of the
dimensionality of the mark

**Position**

- Horizontal
- Vertical
- Both

**Color**

**Shape**

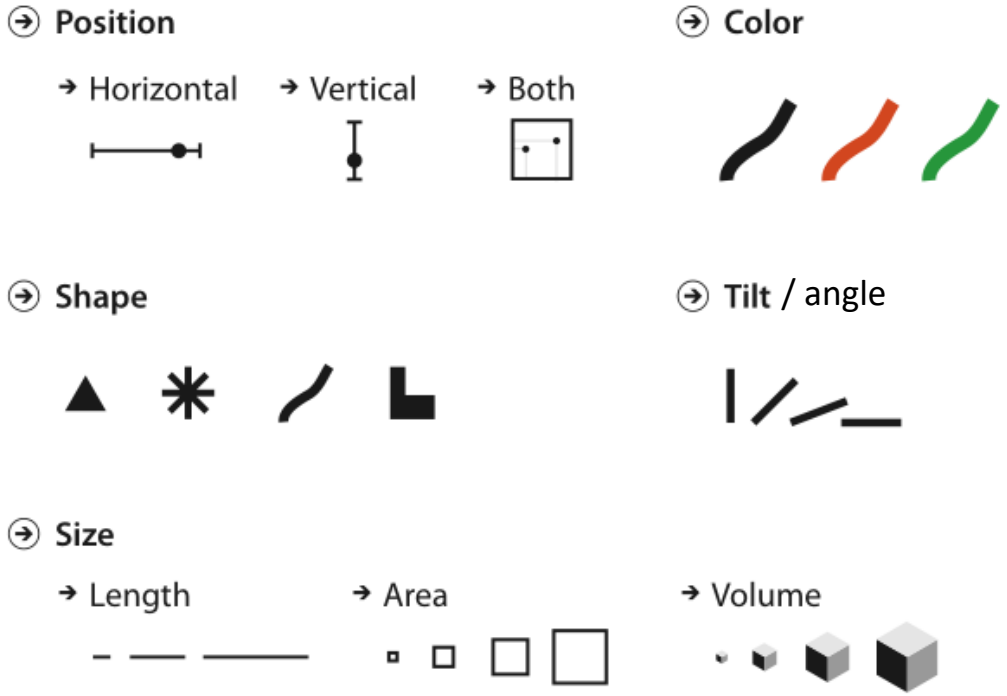
**Tilt / angle**

**Size**

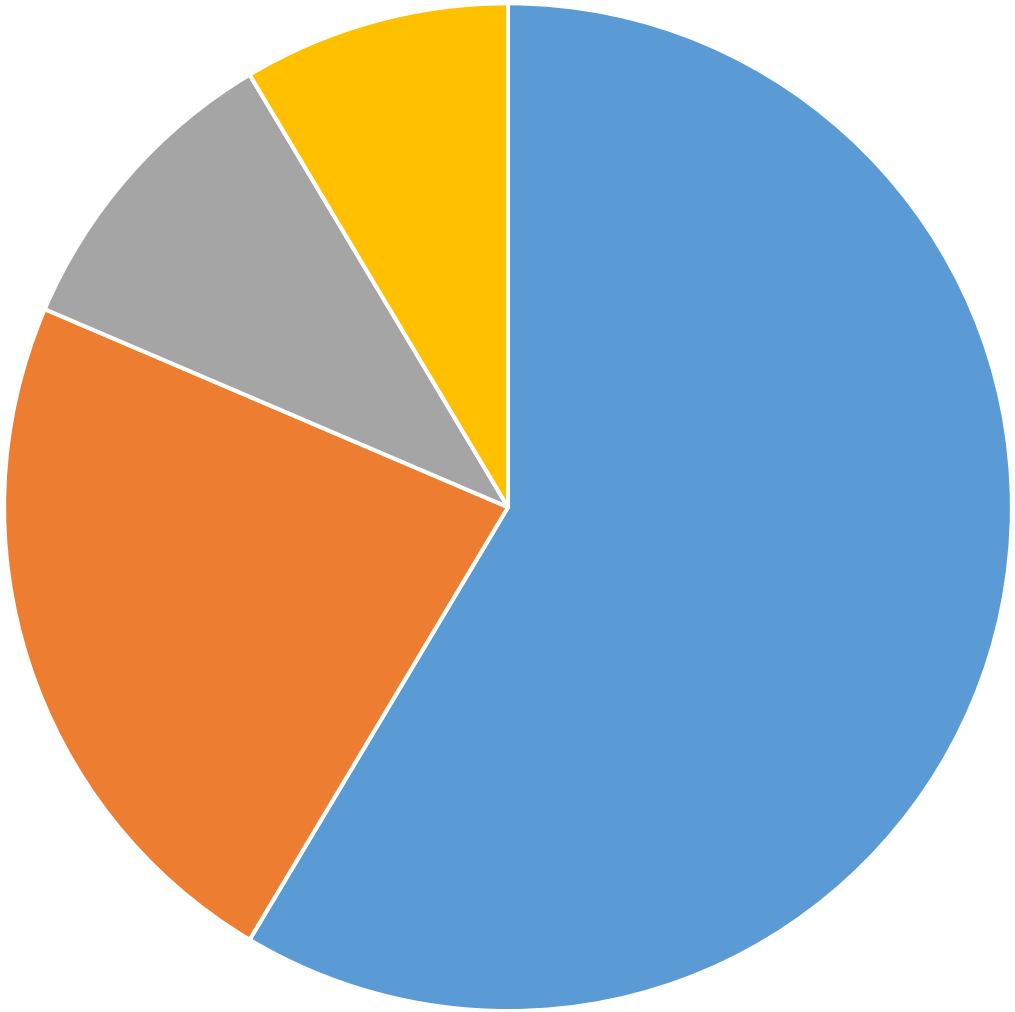
- Length
- Area
- Volume

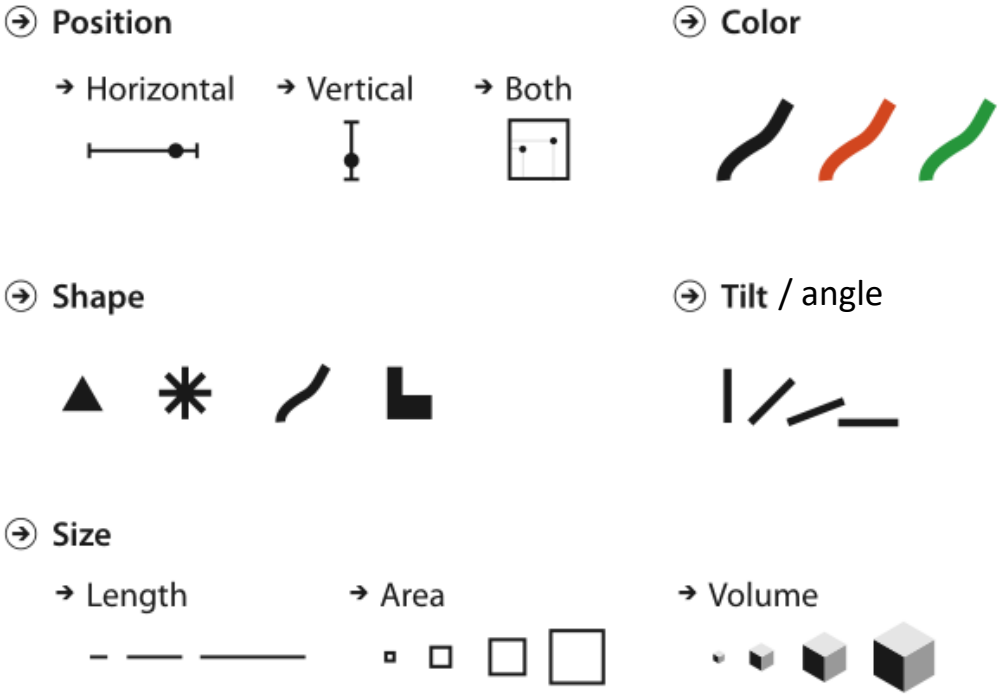
These are the most common but there are more

# which visual channels encode data?

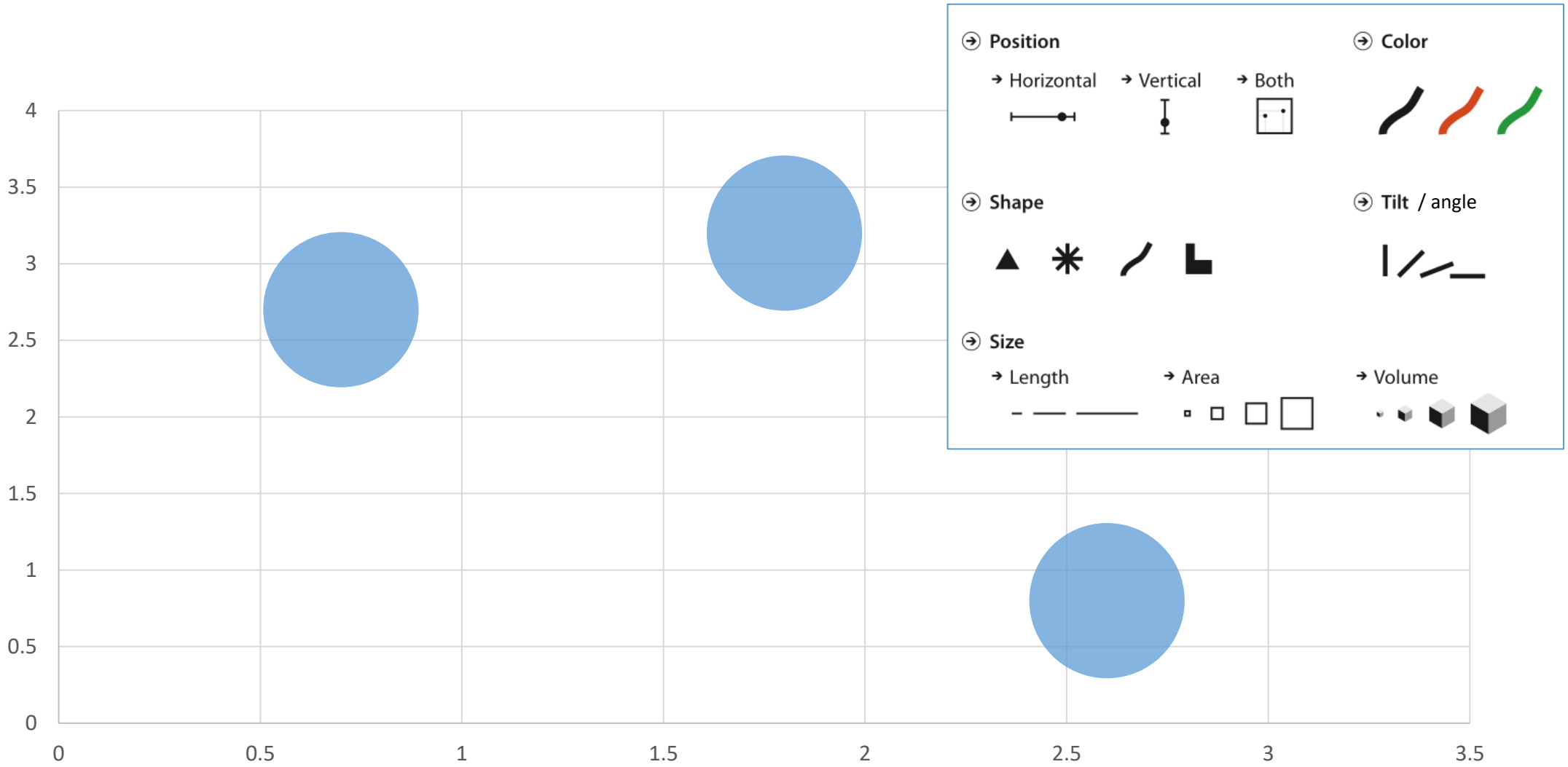


# which visual channels encode data?



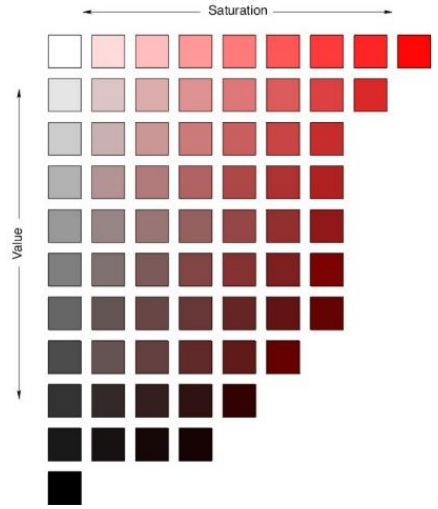


# which visual channels encode data?



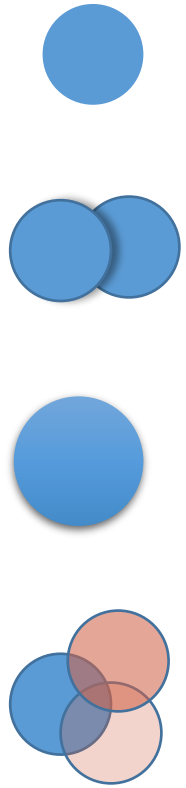
# ADDITIONAL CHANNELS

- motion
  - direction, acceleration, speed, frequency, onset, ‘personality’
- saturation
  - color as Bertin uses it largely refers to hue, saturation != value



# ADDITIONAL CHANNELS

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



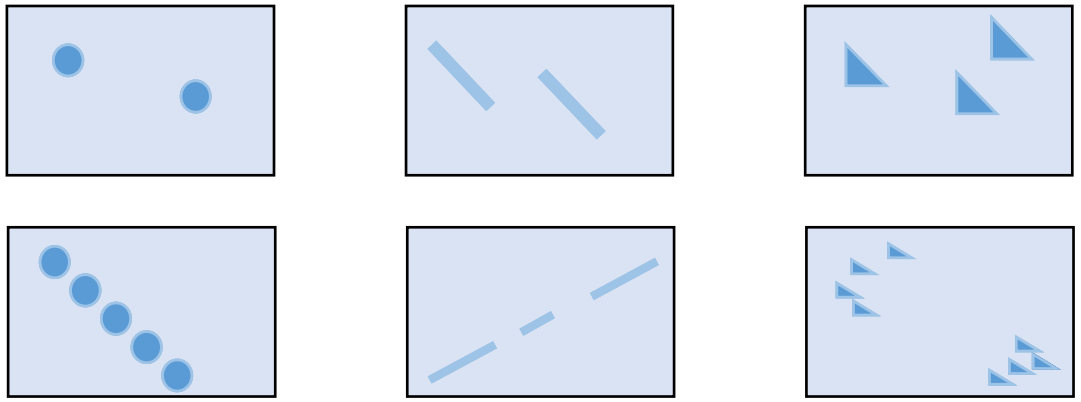
# CHARACTERISTICS OF VISUAL VARIABLES

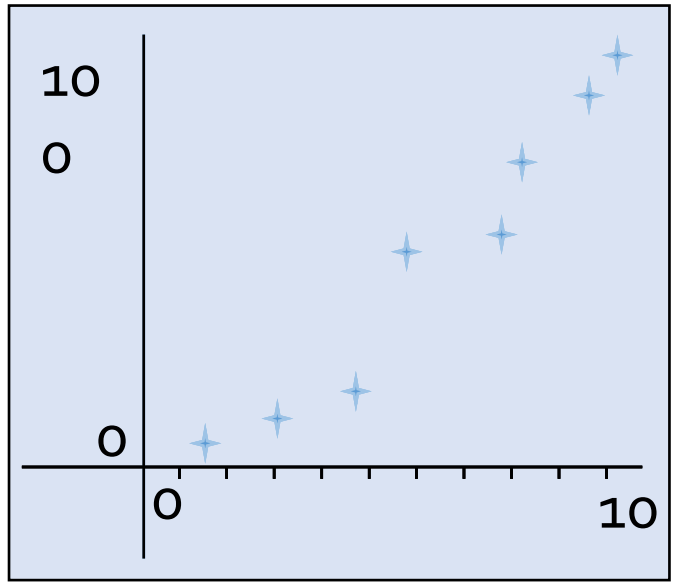
- selective:
  Can this variable allow us to spontaneously differentiate/isolate items from groups?
- associative:
  Can this variable allow us to spontaneously group items in a group?
- ordered:
  Can this variable allow us to spontaneously perceive an order?
- quantitative:
  Is there a numerical reading obtainable from changes in this variable?
- length (resolution):
  Across how many changes in this variable are distinctions possible?

From Semiology of Graphics (Bertin)

# VISUAL VARIABLE: POSITION

- ✓ selective
- ✓ associative
- ✓ order
- ✓ quantitative
- ✓ length (resolution)

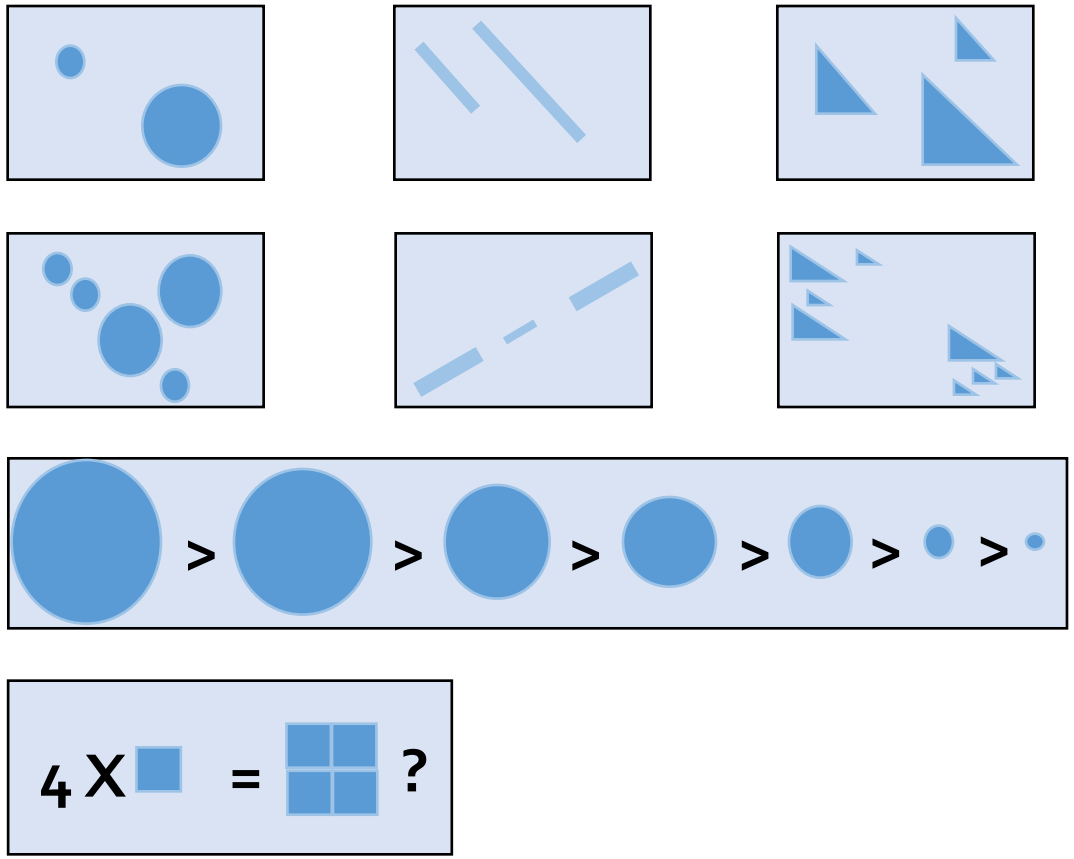





From Semiology of Graphics (Bertin)

# VISUAL VARIABLE: SIZE

- ✓ selective
- ✓ associative
- ✓ order
- ≈ quantitative
- ✓ length (resolution)



# SIZE

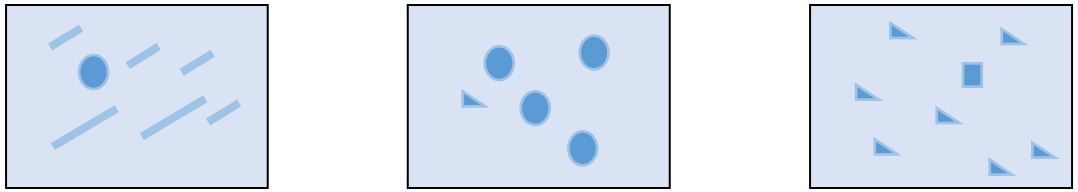
points

lines

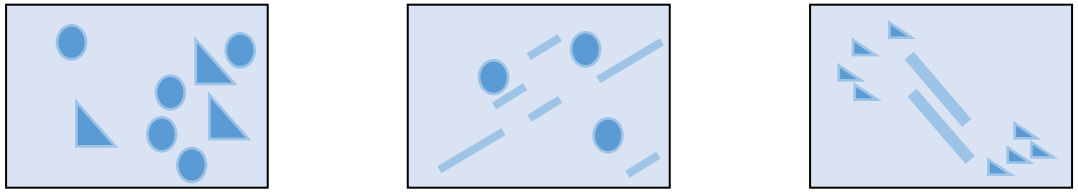
areas

# VISUAL VARIABLE: SHAPE

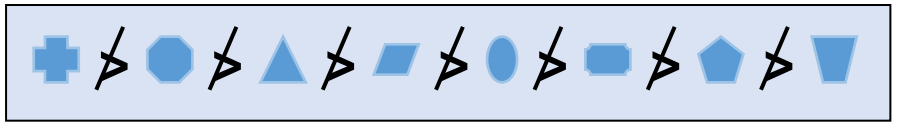
≃ • selective



≃ • associative



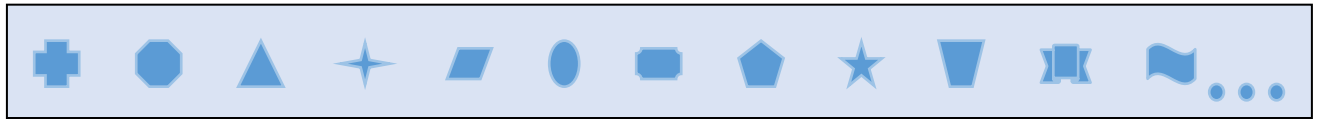
≠ • ordered



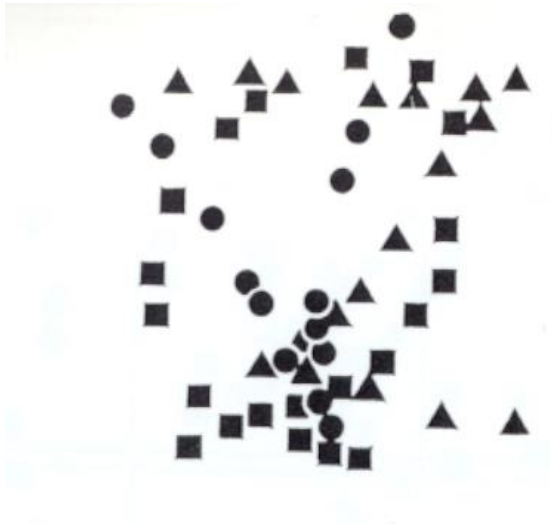
≠ • quantitative

✓ • length (resolution)

– infinite

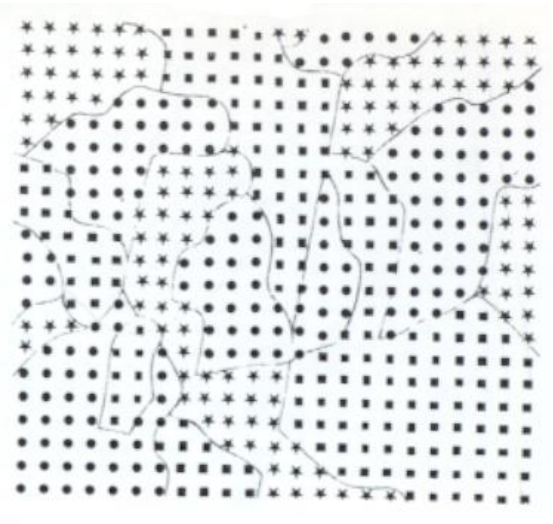


# SHAPE



points

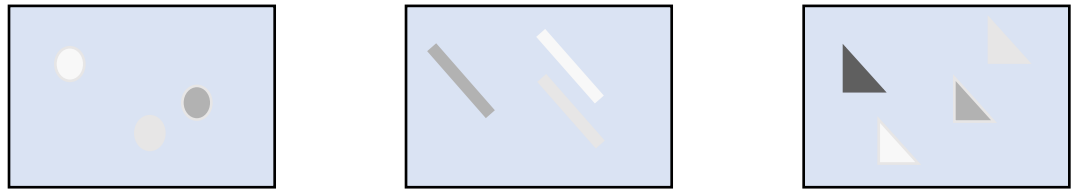
lines

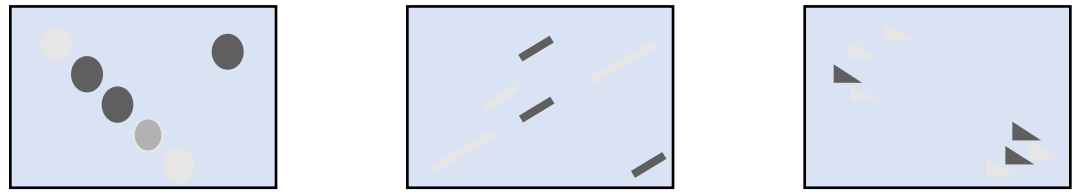


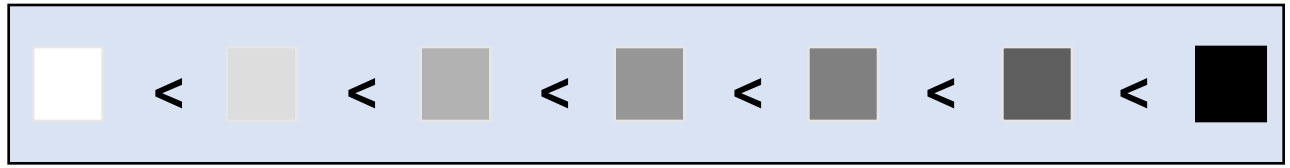
areas

# VISUAL VARIABLE: VALUE

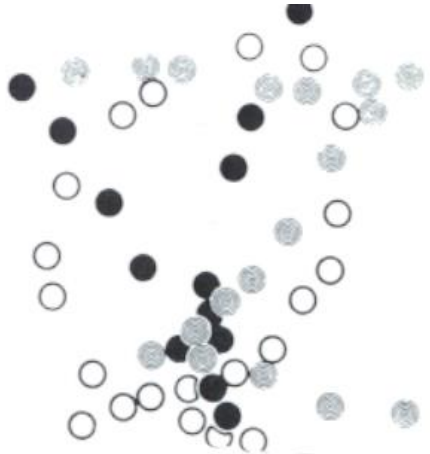
- ✓ selective
- ✓ associative
- ✓ ordered
- ≠ quantitative
- ✓ length (resolution)
  - theoretically infinite but practically limited
  - association and selection ~ < 7 and distinction ~ 10







# VALUE



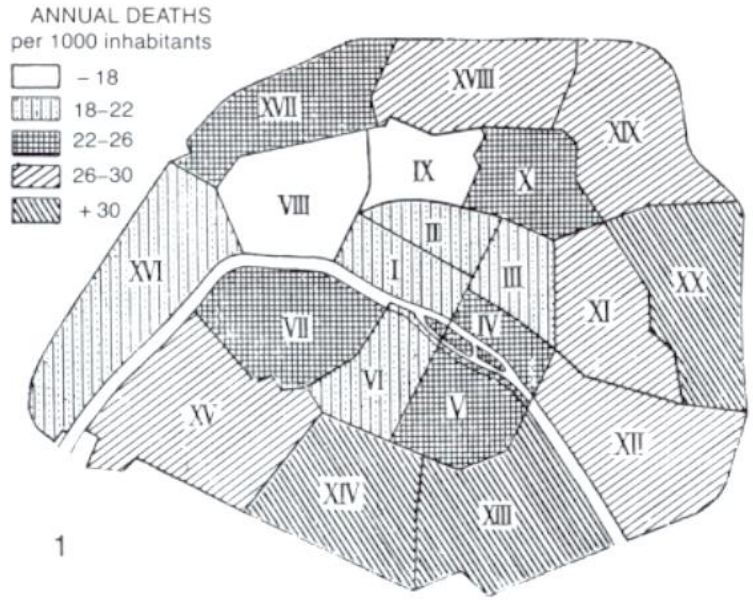
points

lines

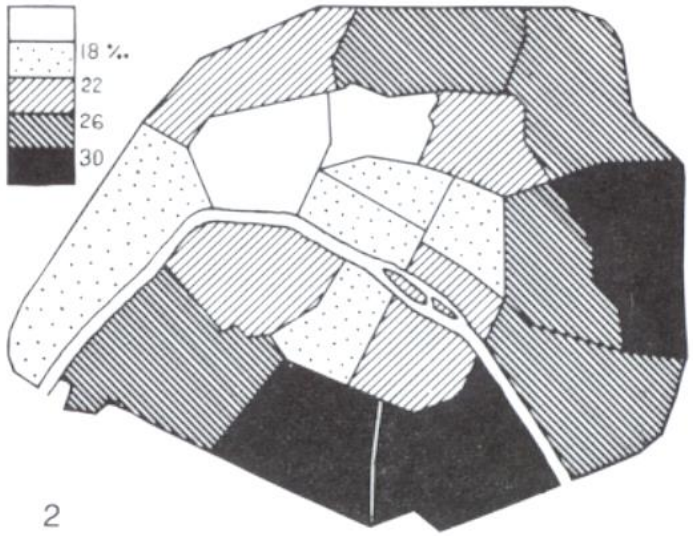
areas

# VALUE

- Ordered, cannot be reordered




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

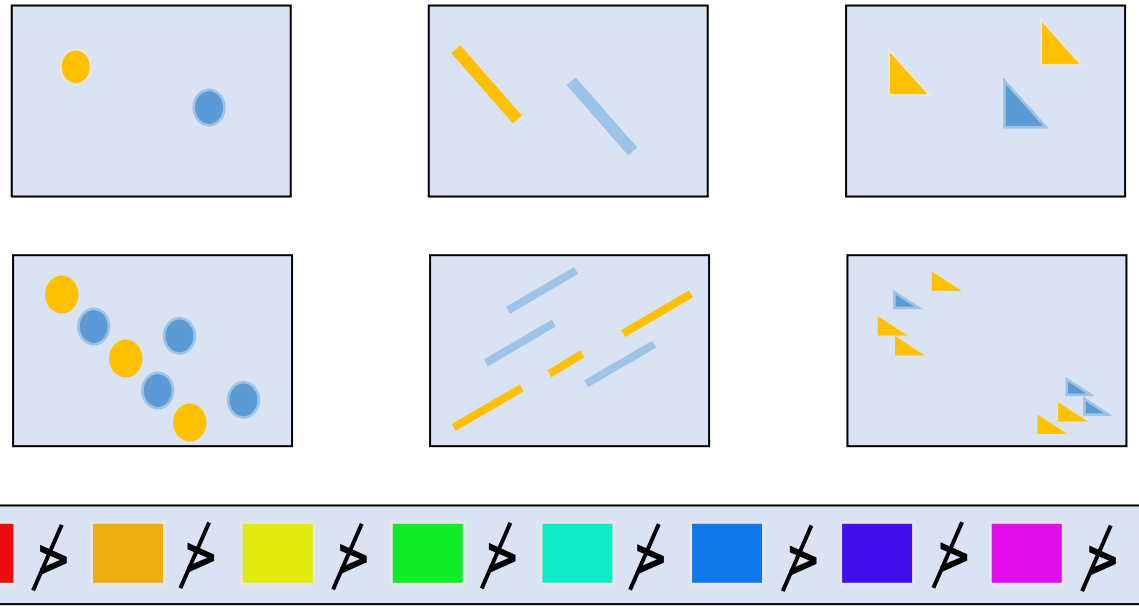



Values ordered correctly
Image much more useful

annual deaths per 1000 inhabitants, Paris

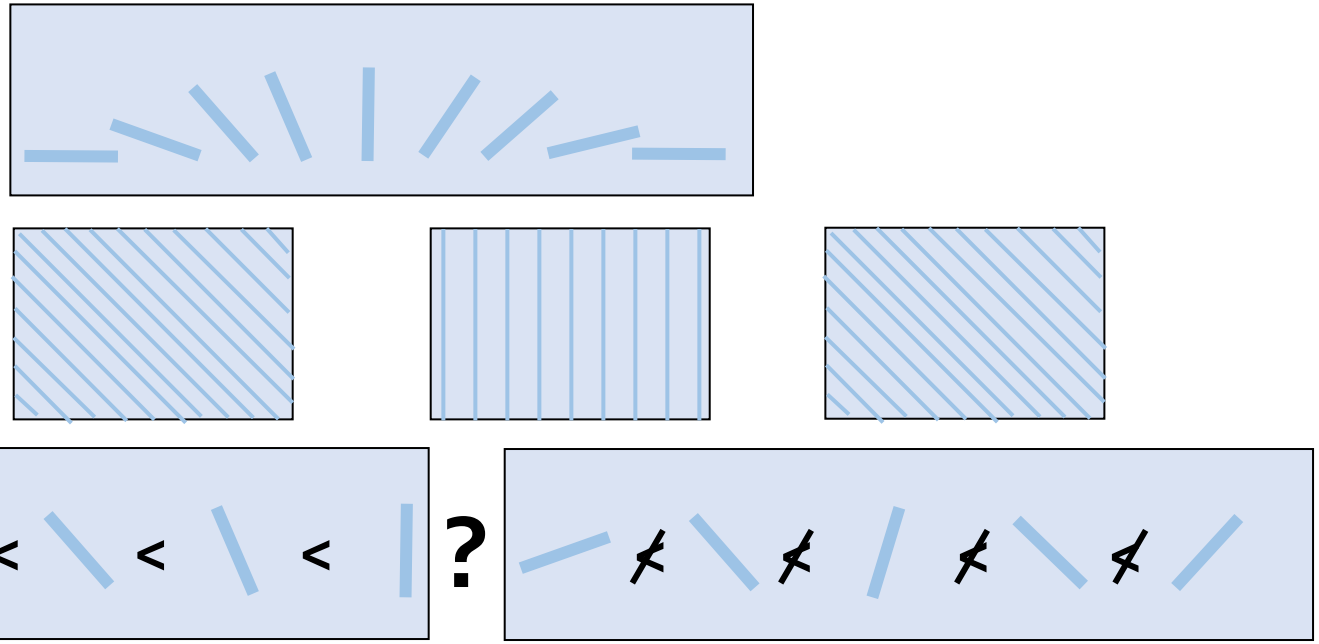
# VISUAL VARIABLE: COLOR

- ✓ selective
- ✓ associative
- ≠ ordered
- ≠ quantitative
- ✓ length (resolution)
  - theoretically infinite but practically limited
  - association and selection ~ < 7 and distinction ~ 10



# VISUAL VARIABLE: ORIENTATION

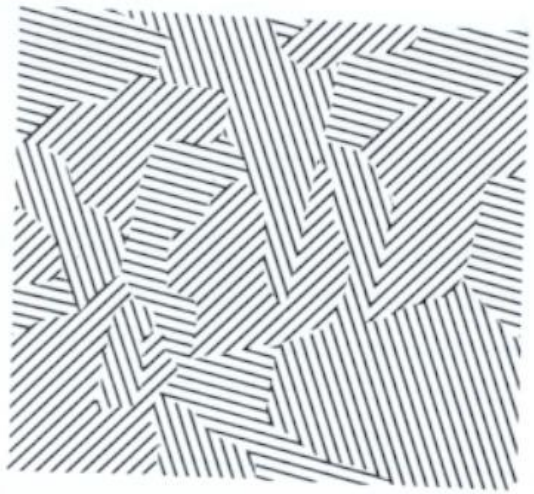
- ✓ selective
- ✓ associative
- ≠ ordered
- ≠ quantitative
- ✓ length (resolution)
  - ~5 in 2D; ? in 3D



# ORIENTATION

points

lines



areas

# VISUAL VARIABLE: TEXTURE

- ✓ selective

- ✓ associative

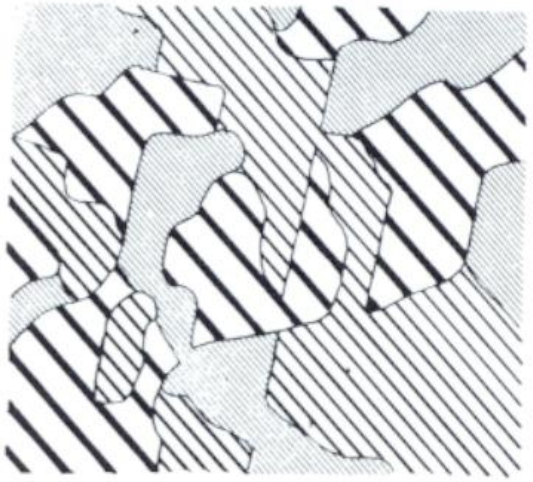
- ≠ ordered
- ≠ quantitative

- ✓ length (resolution)
  - theoretically infinite

# TEXTURE

points

lines



areas

# VISUAL VARIABLE: MOTION

- ✓ selective
  - motion is one of our most powerful attention grabbers
- ✓ associative
  - moving in unison groups objects effectively
- ≠ order
- ≠ quantitative
  - subjective perception
- ? length (resolution)
  - distinguishable types of motion?

# MOTION

# VISUAL VARIABLES

| Visual Variable | Selective | Associative | Quantitative | Order | Length |
|---|---|---|---|---|---|
| Position | Yes | Yes | Yes | Yes | Dependant on resolution |
| Size | Yes | Yes | Approximate | Yes | Association: 5; Distinction: 20 |
| Shape | With Effort | With Effort | No | No | Infinite |
| Value | Yes | Yes | No | Yes | Association: 7; Distinction: 10 |
| Hue | Yes | Yes | No | No | Association: 7; Distinction: 10 |
| Orientation | Yes | Yes | No | No | 4 |
| Grain | Yes | Yes | No | No | 5 |
| Texture | Yes | Yes | No | No | Infinite |
| Motion | Yes | Yes | No | Yes | Unknown |

# HOW TO CHOOSE CHANNELS?

EXPRESSIVENESS

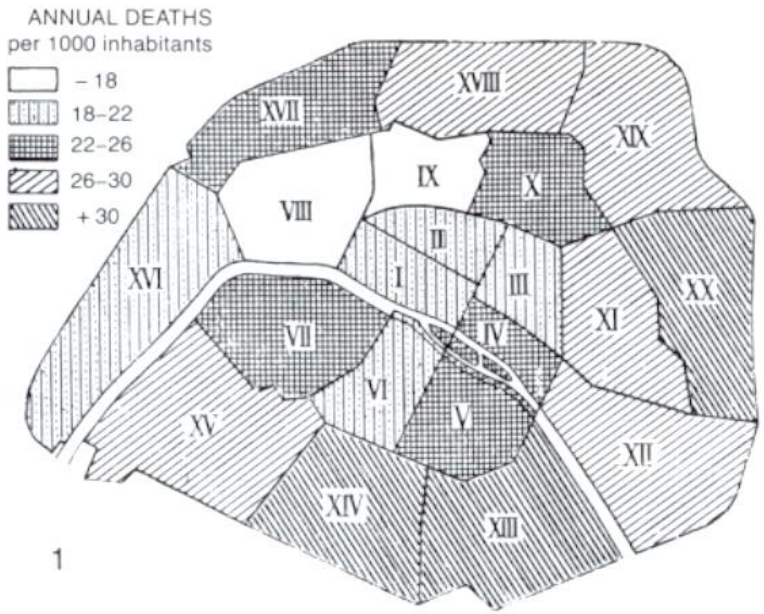
show **all of, and only,** the information in the attributes
(nothing more, nothing less)

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

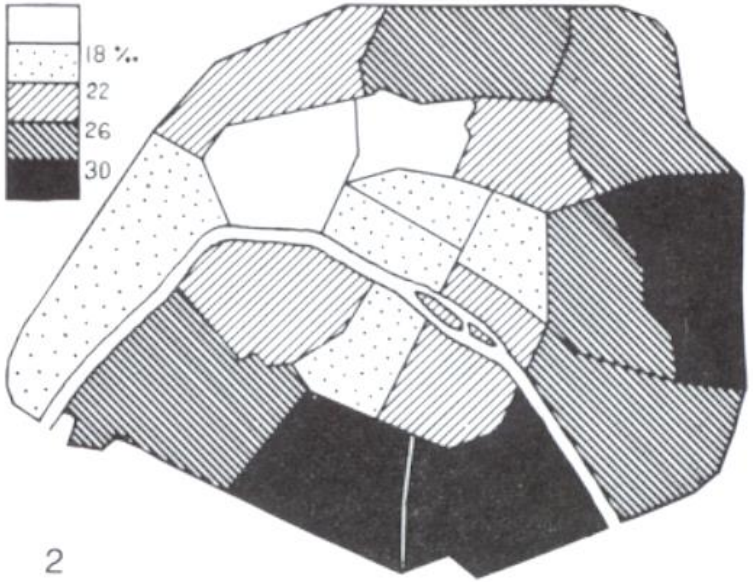
This mismatch is a common beginner's mistake

# EXPRESSIVENESS

luminance is ordered, cannot be reordered




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

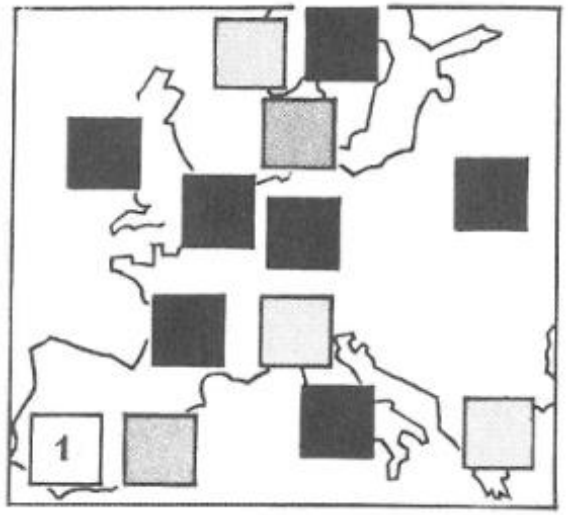



Values ordered correctly
Image much more useful

annual deaths per 1000 inhabitants, Paris

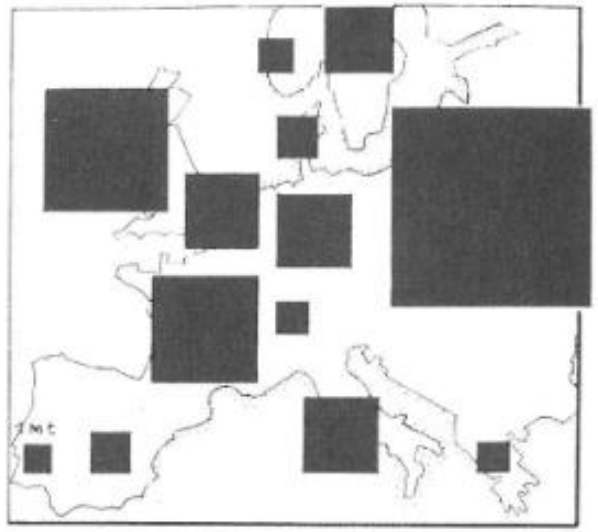
# EXPRESSIVENESS

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

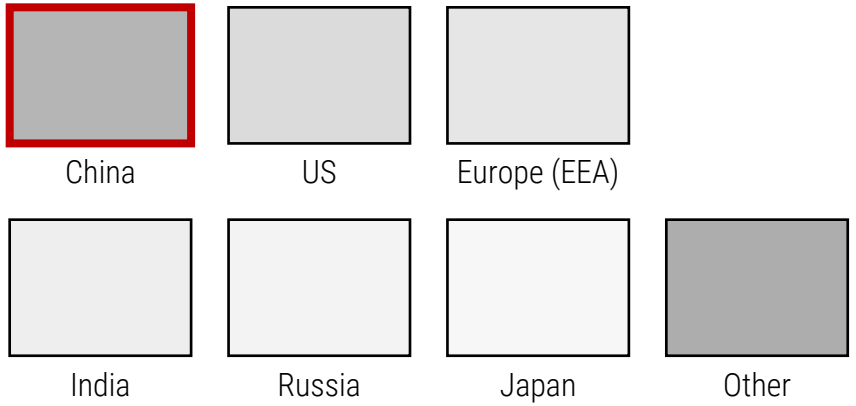
# HOW TO CHOOSE CHANNELS?

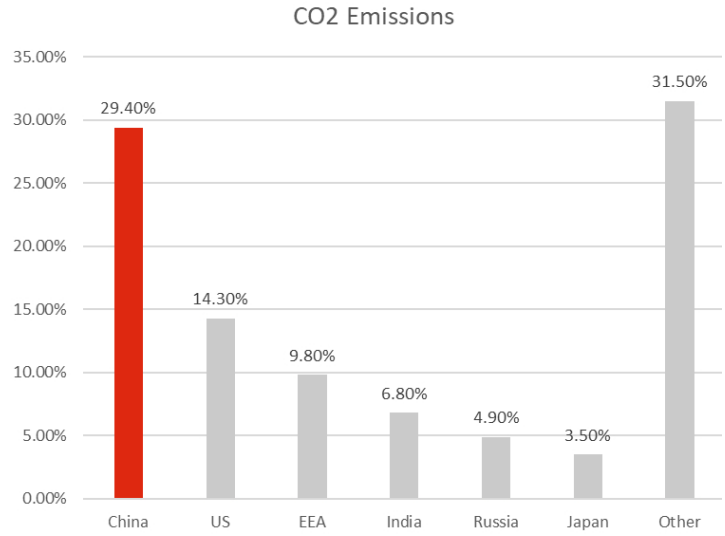
## EFFECTIVENESS

the importance of the attribute matches
the saliency of the channel:

the **most important attributes** should be encoded with the **most effective channels**

**annual $CO_2$ emissions (% of global emissions)**



# EFFECTIVENESS

**Accuracy**: How accurately values can be estimated?

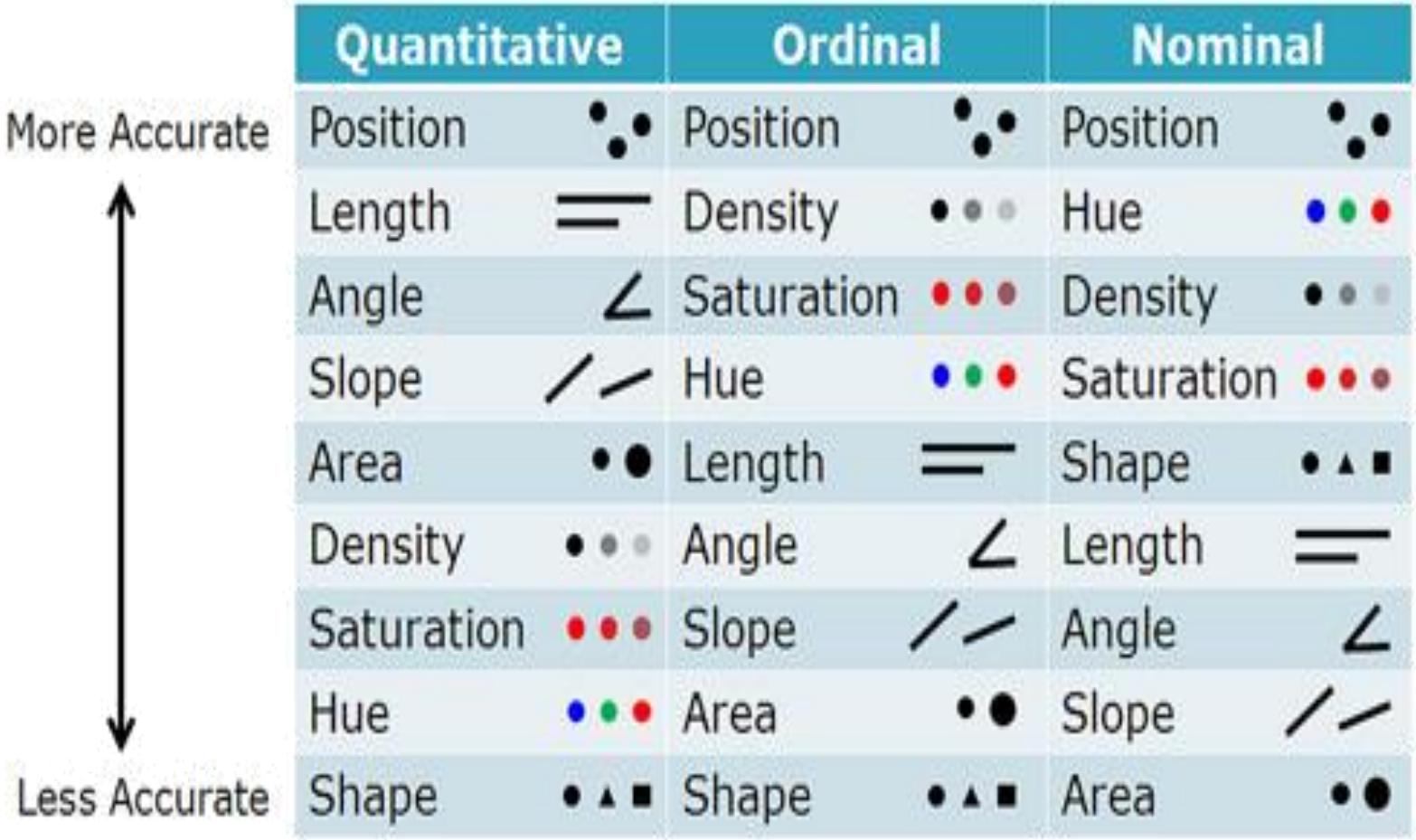
**Discriminability**: How many different values can be perceived?

**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?

# SUMMARY




Jacques Bertin, refined by Cleveland&McGill, then by Card&Mackinlay

## Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

Same

Same

Most

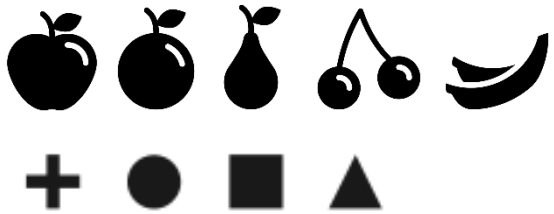
Effectiveness

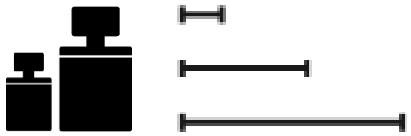
Least

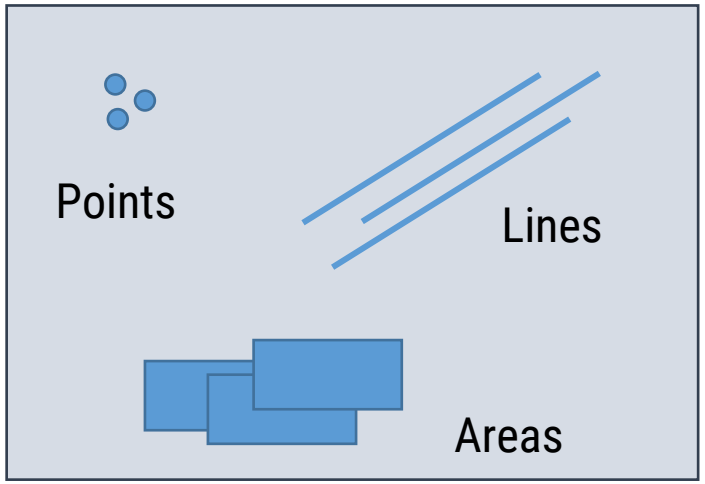
## Identity Channels: Categorical Attributes

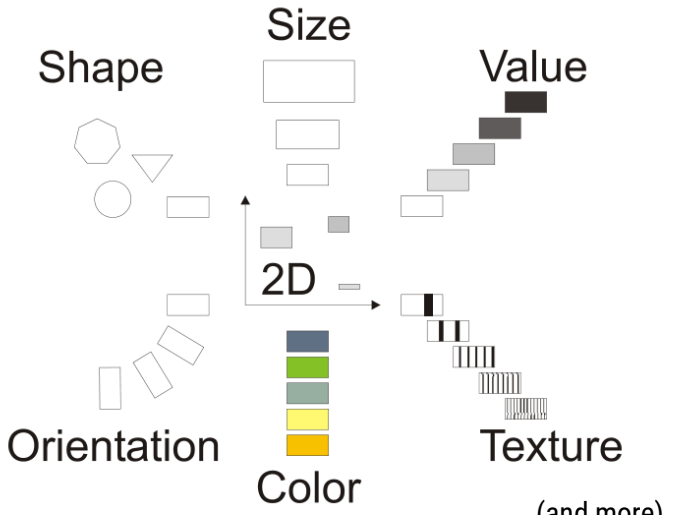
- Spatial region
- Color hue
- Motion
- Shape

# WHAT WE UNDERSTAND SO FAR








(and more)

| | Quantitative | Ordinal | Nominal |
|---|---|---|---|
| More Accurate | Position | Position | Position |
| | Length | Density | Hue |
| | Angle | Saturation | Density |
| | Slope | Hue | Saturation |
| | Area | Length | Shape |
| | Density | Angle | Length |
| | Saturation | Slope | Angle |
| | Hue | Area | Slope |
| Less Accurate | Shape | Shape | Area |