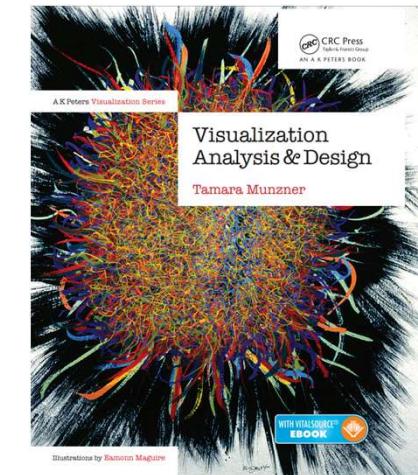
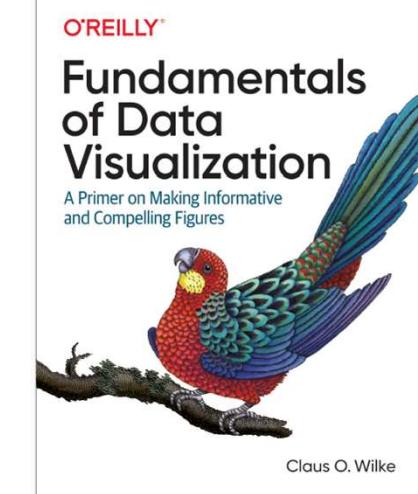


# Introduction to Data Visualization

Veera Muangsin

# References

- Book: [Fundamentals of Data Visualization](#), Claus O. Wilke, 2019.
- Book: [Visualization Analysis & Design](#), Tamara Munzner, CRC Press, 2014.  
<http://www.cs.ubc.ca/~tmm/talks.html>
- [A Tour Through the Visualization Zoo](#), Jeffrey Heer, Michael Bostock and Vadim Ogievetsky, Communications of the ACM, 2010.  
<https://dl.acm.org/citation.cfm?id=1743567>
- Book: [The Visual Display of Quantitative Information](#), Edward R. Tufte, 2001.



DOI:10.1145/1743546.1743567  
Article development led by [queue.acm.org](https://queue.acm.org)  
A survey of powerful visualization techniques, from the obvious to the obscure.  
BY JEFFREY HEER, MICHAEL BOSTOCK, AND VADIM OGIEVETSKY

## Examples

- <https://public.tableau.com/en-us/s/gallery>
- <https://community.powerbi.com/t5/Data-Stories-Gallery/bd-p/DataStoriesGallery>
- <https://www.gapminder.org/tools/>
- <https://flowingdata.com/>
- <https://truth-and-beauty.net/>
- <http://datavis.ca/milestones/>

## A Tour Through the Visualization Zoo

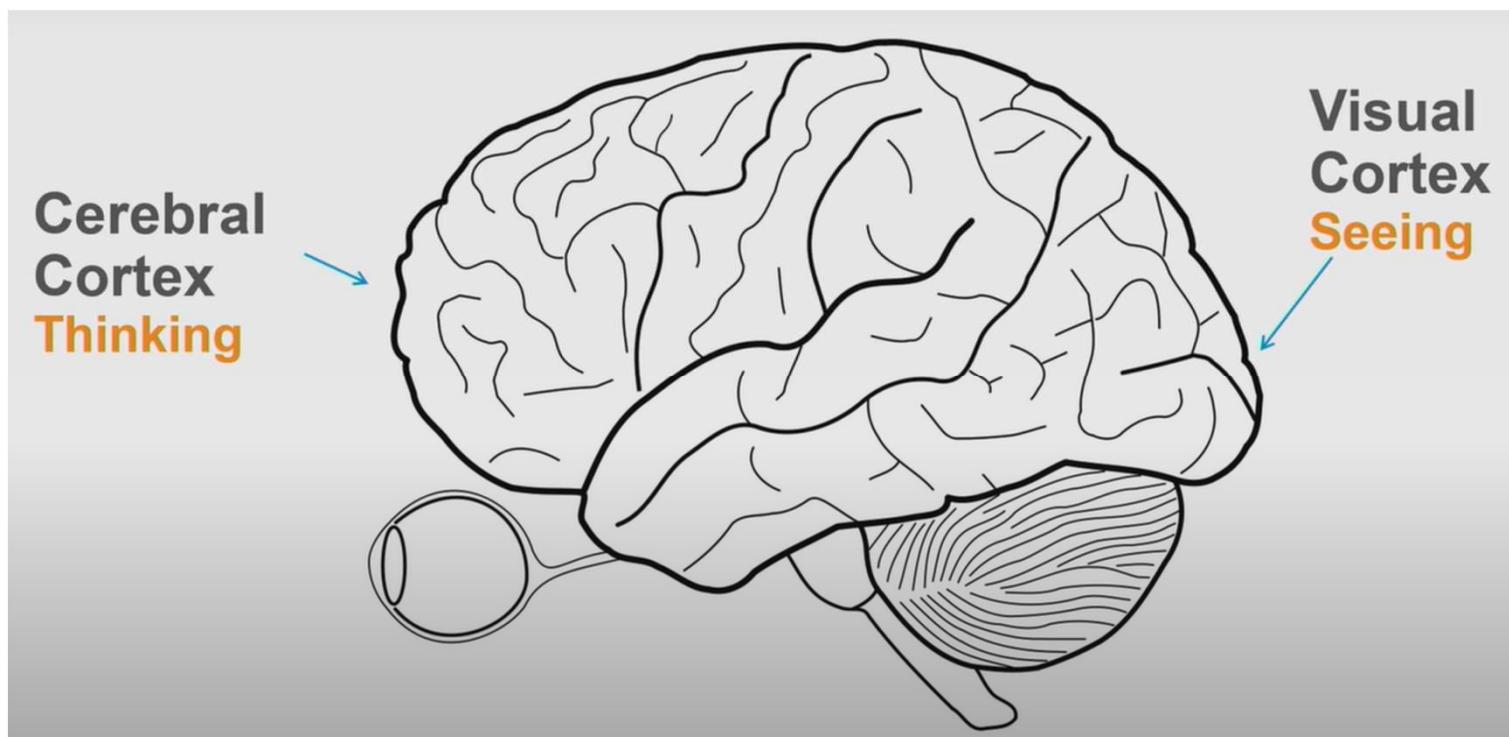
THANKS TO ADVANCES in sensing, networking, and data management, our society is producing digital information at an astonishing rate. According to one estimate, in 2010 alone we will generate 1,200 exabytes—60 million times the content of the Library of Congress. Within this deluge of data lies a wealth

# What is Data Visualization?

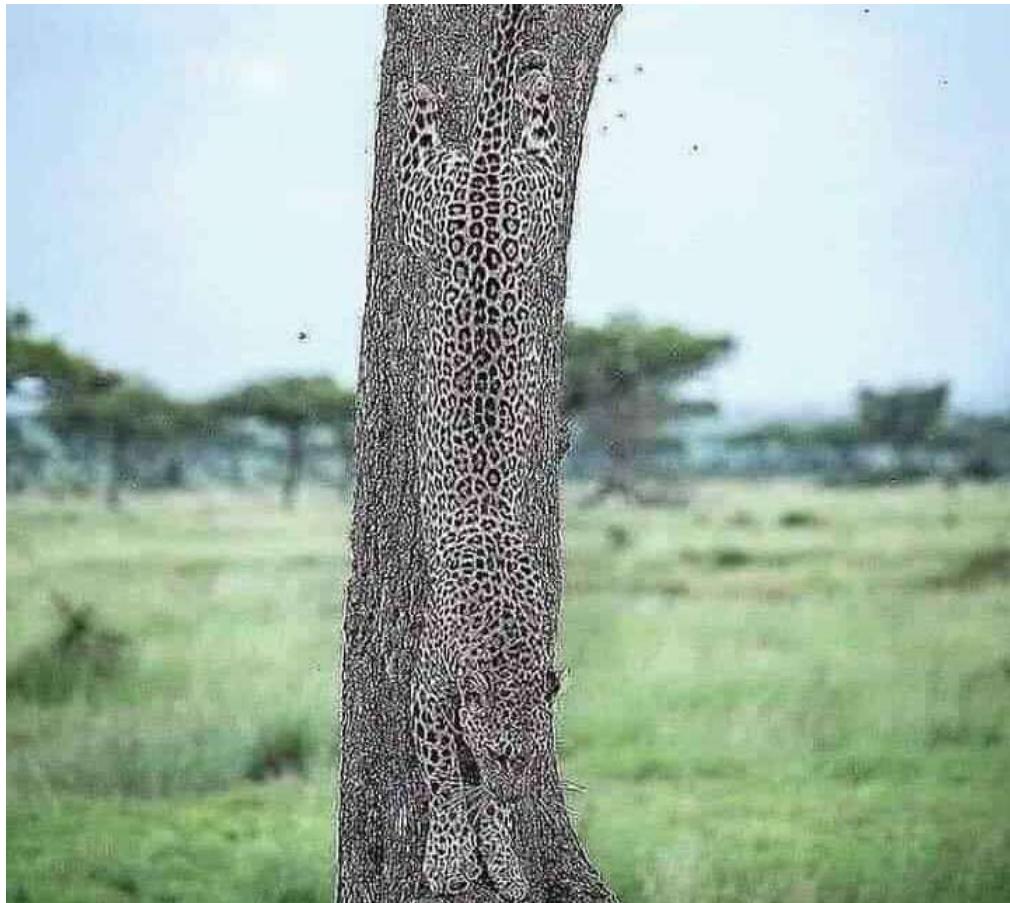
The process of converting raw data  
into easily understood visual representation  
that enables effective communication  
and decision making

# Why Visualization?

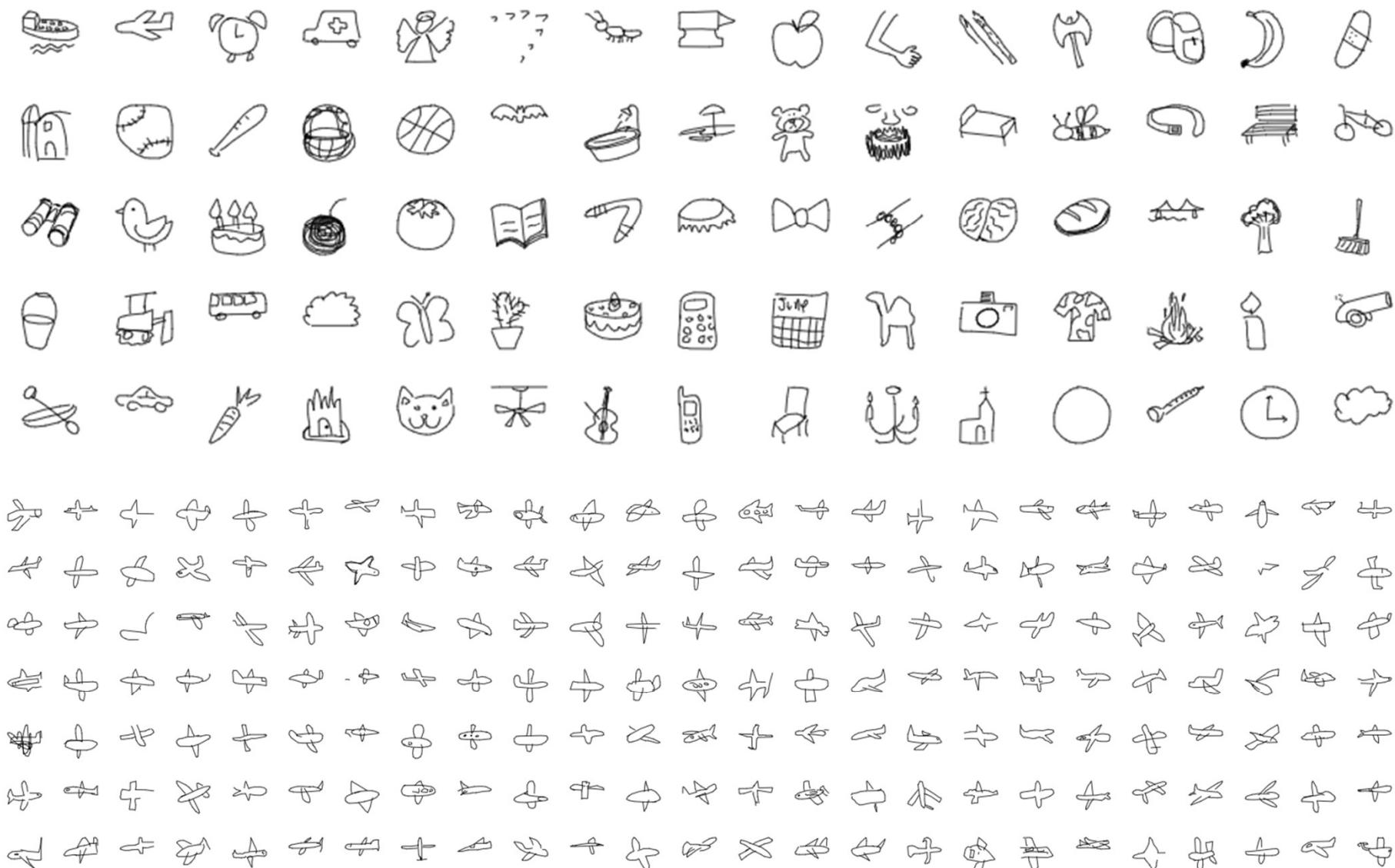
- Of all human senses, visual sense is the most powerful.
  - Smell for dog, sonar for dolphin, etc.
- 90% of all information transmitted to our brains is visual.
- Human is good at finding patterns.



# See or Die



# Abstraction



<https://quickdraw.withgoogle.com/data>

# Language: Can You Raed Tihs?

*We do not raed ervey lteter by itslef  
but the wrod as a wlohe.*

*We do not read every letter by itself  
but the word as a whole.*

# Visual Representation

Data visualization displays measured quantities using a combination of geometric primitives.

Marks are geometric primitives.

④ Points



④ Lines

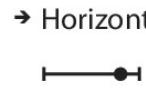


④ Areas



The appearance of marks corresponds to the data.

④ Position



④ Color



④ Shape

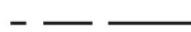


④ Tilt



④ Size

→ Length



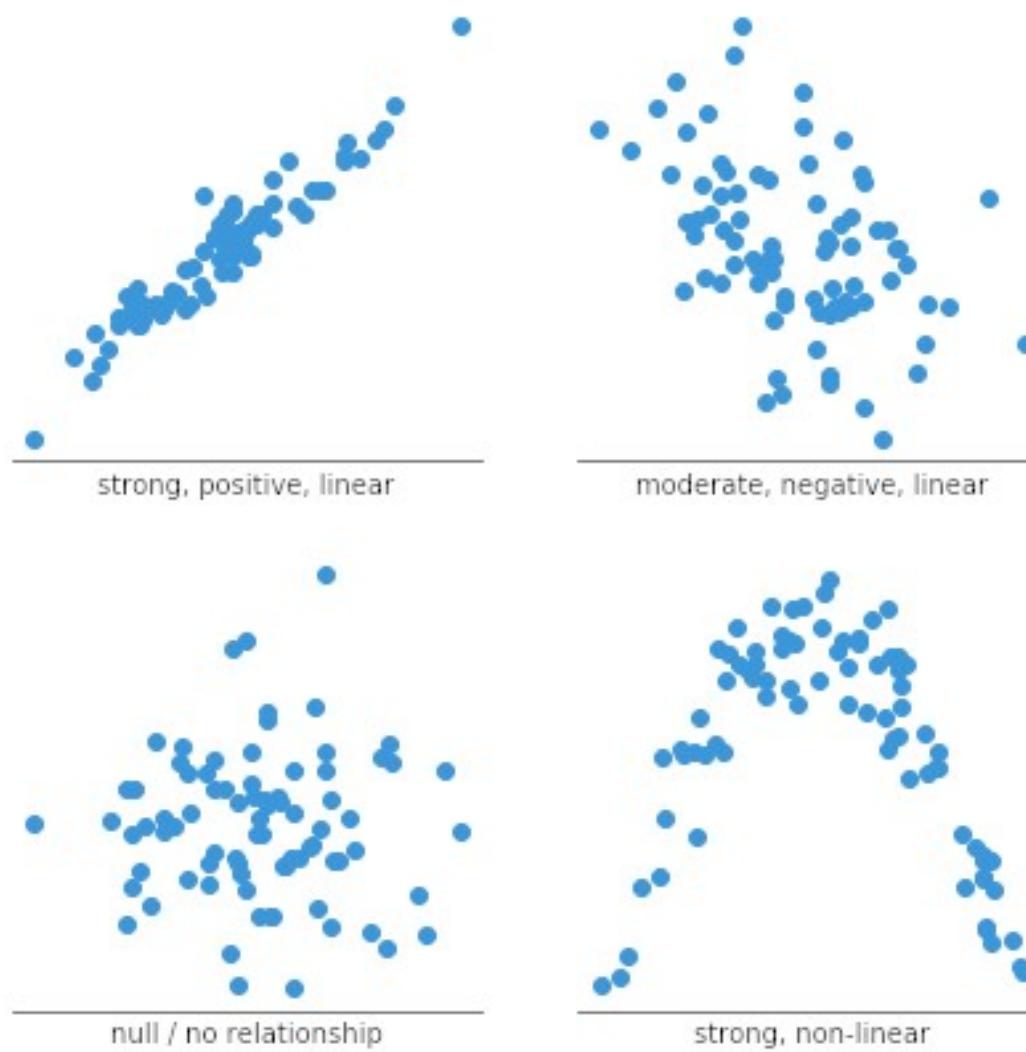
→ Area



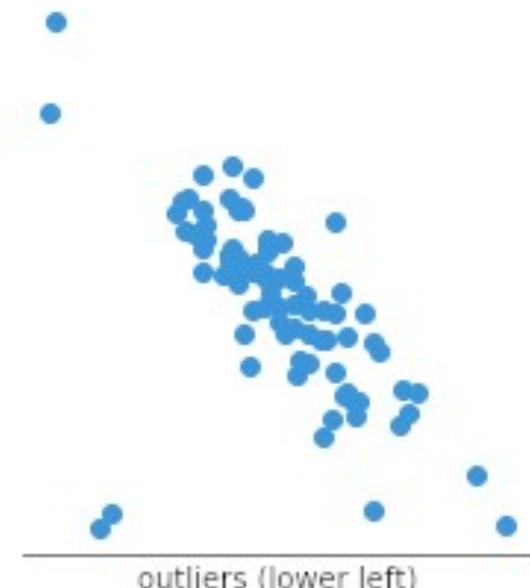
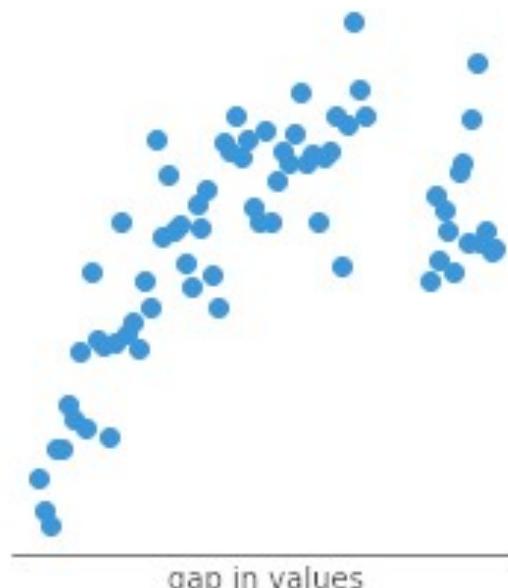
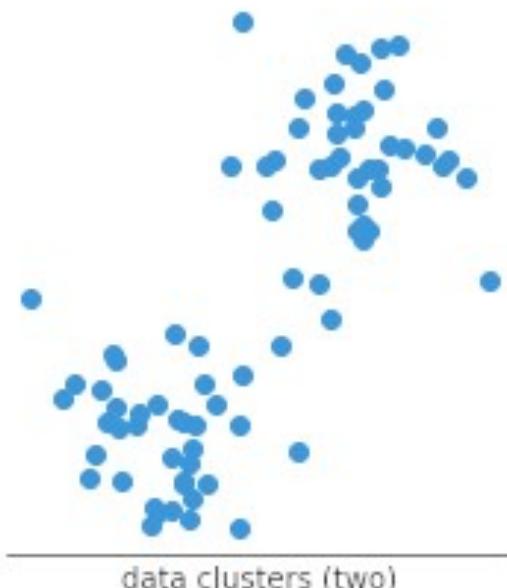
→ Volume



# Identifying Correlational Relationships



# Identifying other patterns



# Raw data vs. Visualization

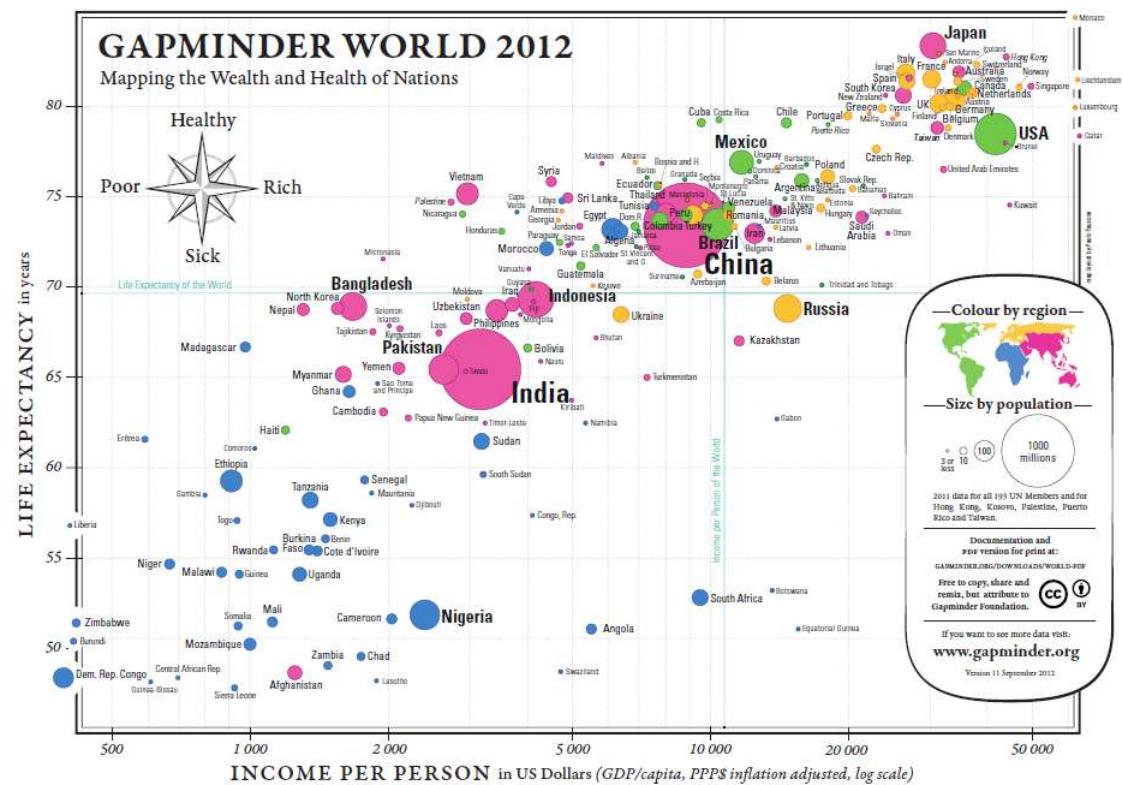
country	region	income per person	life expectancy	population
Afghanistan	asia_west	1840	57.2	30700000
Albania	europe_east	10400	77	2920000
Algeria	africa_north	13200	76.8	37600000
Andorra	europe_west	41900	82.6	82400
Angola	africa_sub_saharan	6000	61.7	25100000
Antigua and Barbuda	america_north	19100	77	96800
Argentina	america_south	19200	76.1	42100000
Armenia	europe_east	7510	74.3	2880000
Australia	east_asia_pacific	42600	82.3	22800000
Austria	europe_west	44400	80.9	8520000
Azerbaijan	europe_east	15900	70.2	9260000
Bahamas	america_north	23000	73.7	372000
Bahrain	asia_west	41500	76.3	1300000
Bangladesh	asia_west	2710	71.3	156000000
Barbados	america_north	15400	76.8	282000
Belarus	europe_east	17500	71.8	9470000
Belgium	europe_west	41000	80.3	11100000
Belize	america_north	7970	71.6	337000
Benin	africa_sub_saharan	1860	62.6	9730000
Bhutan	asia_west	7030	72.9	753000

Thailand	east_asia_pacific	14400	77.2	67800000
Timor-Leste	east_asia_pacific	2030	72	1160000
Togo	africa_sub_saharan	1260	60.4	6860000
Tonga	east_asia_pacific	5130	70	105000
Trinidad and Tobago	america_north	31300	72.8	1340000
Tunisia	africa_north	10400	77.1	10900000
Turkey	europe_east	20300	78.6	74600000
Turkmenistan	asia_west	12200	68.8	5270000
Uganda	africa_sub_saharan	1640	58.6	36300000
Ukraine	europe_east	8320	71.1	45300000
United Arab Emirates	asia_west	59800	76.4	8900000
United Kingdom	europe_west	36700	80.7	64300000
United States	america_north	50500	78.9	313000000
Uruguay	america_south	18500	76.5	3400000
Uzbekistan	asia_west	4770	69.3	29500000
Vanuatu	east_asia_pacific	2900	63.3	247000
Venezuela	america_south	17700	75.3	29900000
Vietnam	east_asia_pacific	4910	73.6	90500000
Yemen	asia_west	3790	67.9	24900000
Zambia	africa_sub_saharan	3510	54.5	14700000
Zimbabwe	africa_sub_saharan	1850	54.1	14700000

# Raw data vs. Visualization

country	region	income per person	life expectancy	population
Afghanistan	asia_west	1840	57.2	30700000
Albania	europe_east	10400	77	2920000
Algeria	africa_north	13200	76.8	37600000
Andorra	europe_west	41900	82.6	82400
Angola	africa_sub_saharan	6000	61.7	25100000
Antigua and Barbuda	america_north	19100	77	96800
Argentina	america_south	19200	76.1	42100000
Armenia	europe_east	7510	74.3	2880000
Australia	east_asia_pacific	42600	82.3	22800000
Austria	europe_west	44400	80.9	8520000
Azerbaijan	europe_east	15900	70.2	9260000
Bahamas	america_north	23000	73.7	372000
Bahrain	asia_west	41500	76.3	1300000
Bangladesh	asia_west	2710	71.3	156000000
Barbados	america_north	15400	76.8	282000
Belarus	europe_east	17500	71.8	9470000
Belgium	europe_west	41000	80.3	11100000
Belize	america_north	7970	71.6	337000
Benin	africa_sub_saharan	1860	62.6	9730000
Bhutan	asia_west	7030	72.9	753000

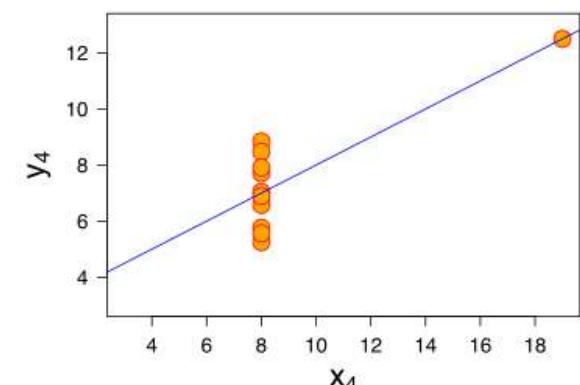
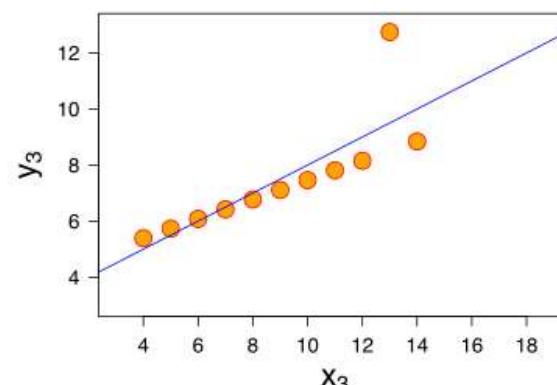
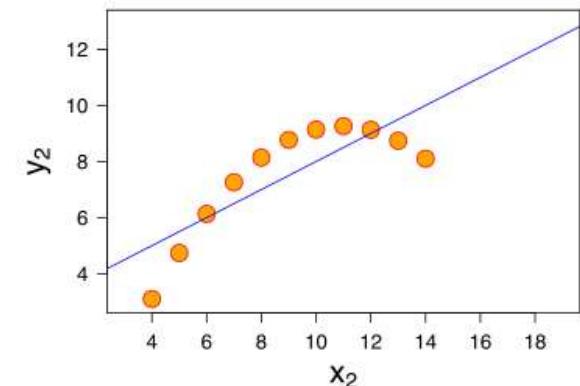
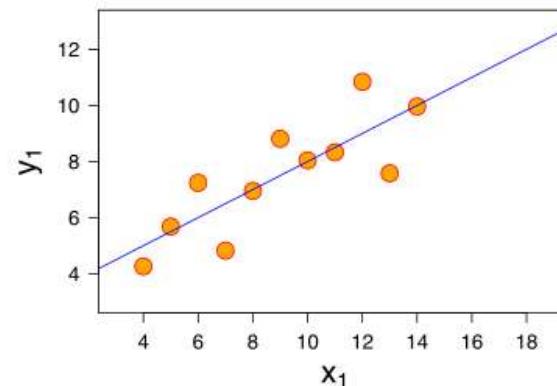
Thailand	east_asia_pacific	14400	77.2	67800000
Timor-Leste	east_asia_pacific	2030	72	1160000
Togo	africa_sub_saharan	1260	60.4	6860000
Tonga	east_asia_pacific	5130	70	105000
Trinidad and Tobago	america_north	31300	72.8	1340000
Tunisia	africa_north	10400	77.1	10900000
Turkey	europe_east	20300	78.6	74600000
Turkmenistan	asia_west	12200	68.8	5270000
Uganda	africa_sub_saharan	1640	58.6	36300000
Ukraine	europe_east	8320	71.1	45300000
United Arab Emirates	asia_west	59800	76.4	8900000
United Kingdom	europe_west	36700	80.7	64300000
United States	america_north	50500	78.9	313000000
Uruguay	america_south	18500	76.5	3400000
Uzbekistan	asia_west	4770	69.3	29500000
Vanuatu	east_asia_pacific	2900	63.3	247000
Venezuela	america_south	17700	75.3	29900000
Vietnam	east_asia_pacific	4910	73.6	90500000
Yemen	asia_west	3790	67.9	24900000
Zambia	africa_sub_saharan	3510	54.5	14700000
Zimbabwe	africa_sub_saharan	1850	54.1	14700000



# Anscombe's quartet

- Demonstrate the importance of visual representation of data
- All four data set have almost identical statistics but different graphs

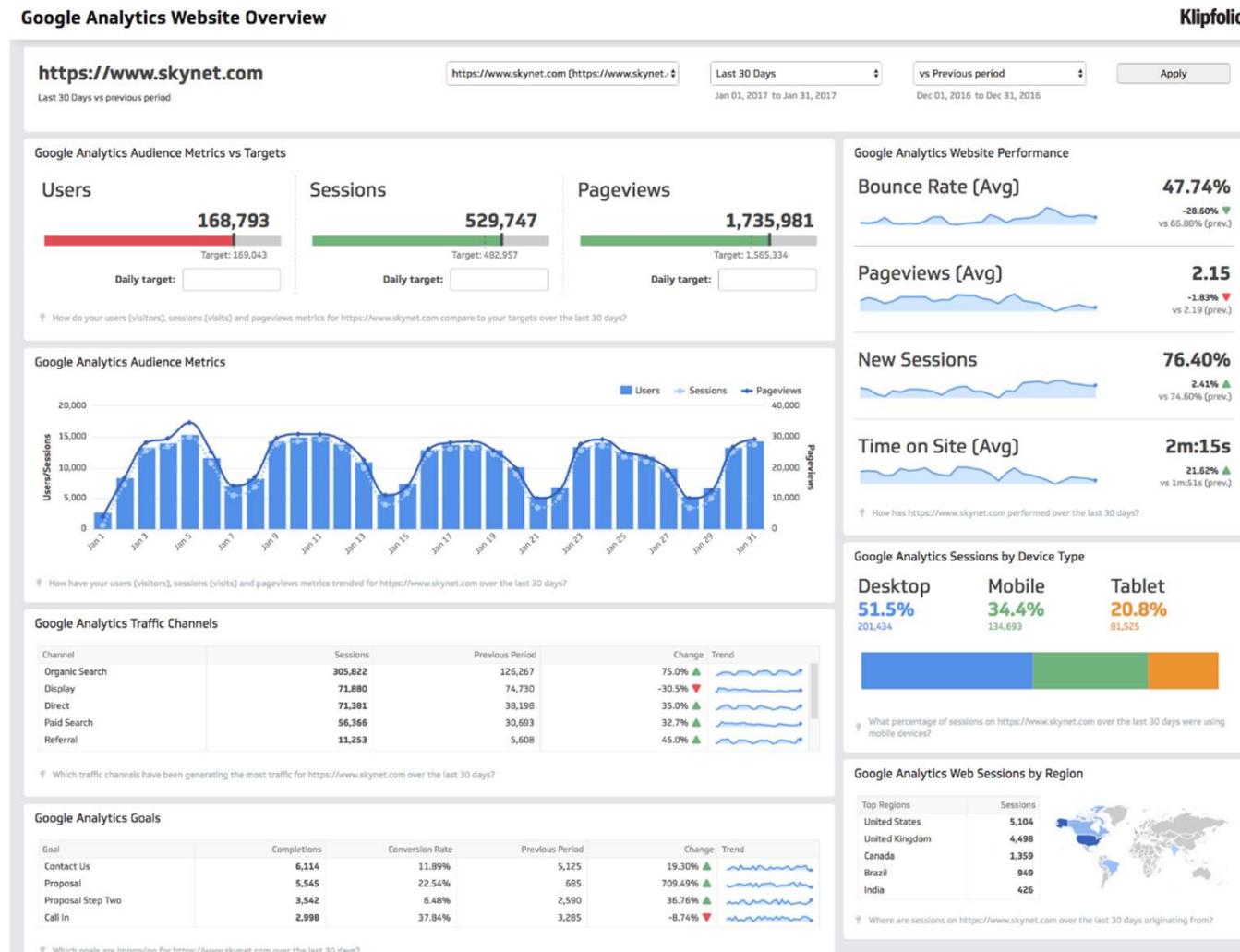
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	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



Property	Value
Mean of $x$	9
Sample variance of $x$	11
Mean of $y$	7.50
Sample variance of $y$	4.125
Correlation between $x$ and $y$	0.816
Linear regression line	$y = 3.00 + 0.500x$
Coefficient of determination of the linear regression	0.67

# Dashboard

Display many measures or key performance indicators (KPIs) on different charts.



# Stock Market Technical Chart

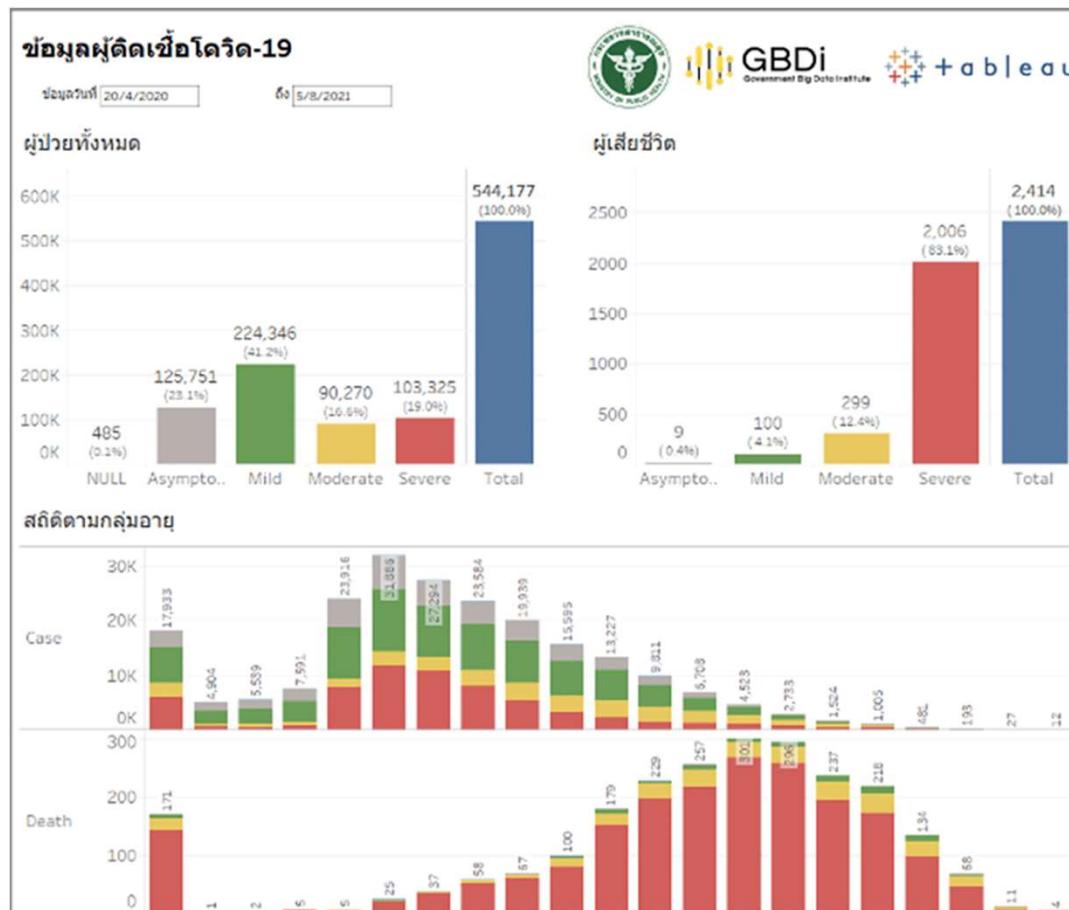


# JHU COVID-19 Dashboard (2019)

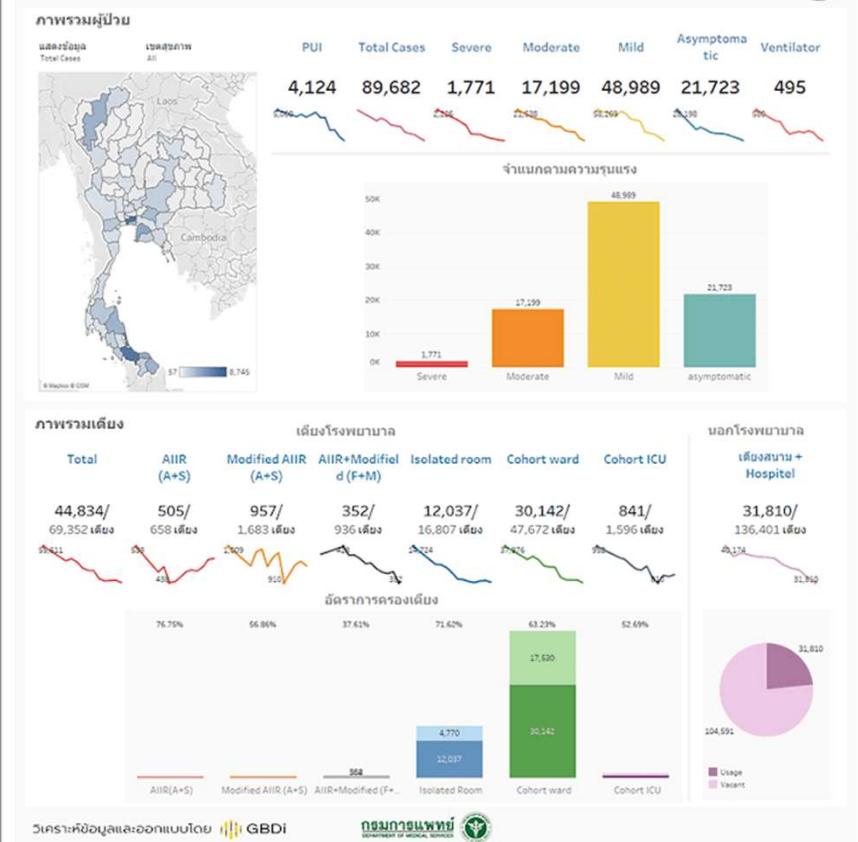
<https://coronavirus.jhu.edu/map.html>



# Thailand Covid-19 Dashboard



รายงานสถานการณ์ตีเสียงร้องรับผ้าไว้ COVID-19



<https://www.tableau.com/blog/how-tableau-community-volunteered-their-data-skills-help-government-thailand>

# Traffic Congestion Dashboard



<https://public.tableau.com/app/profile/veera.muangsin/viz/CongestionbasedonTravelTimeIndex/TravelTimeIndexTTIDashboard>

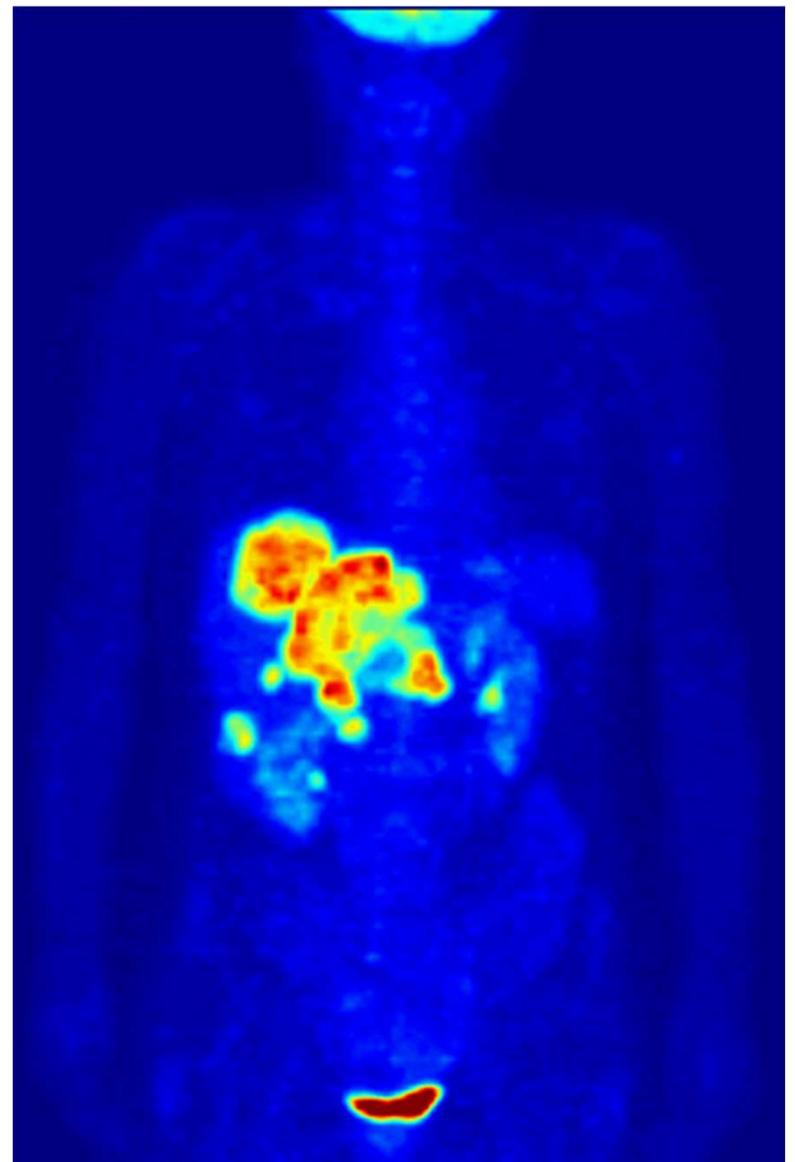
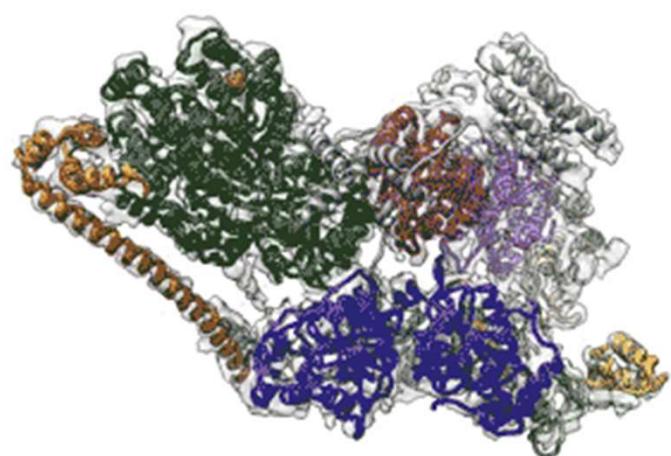
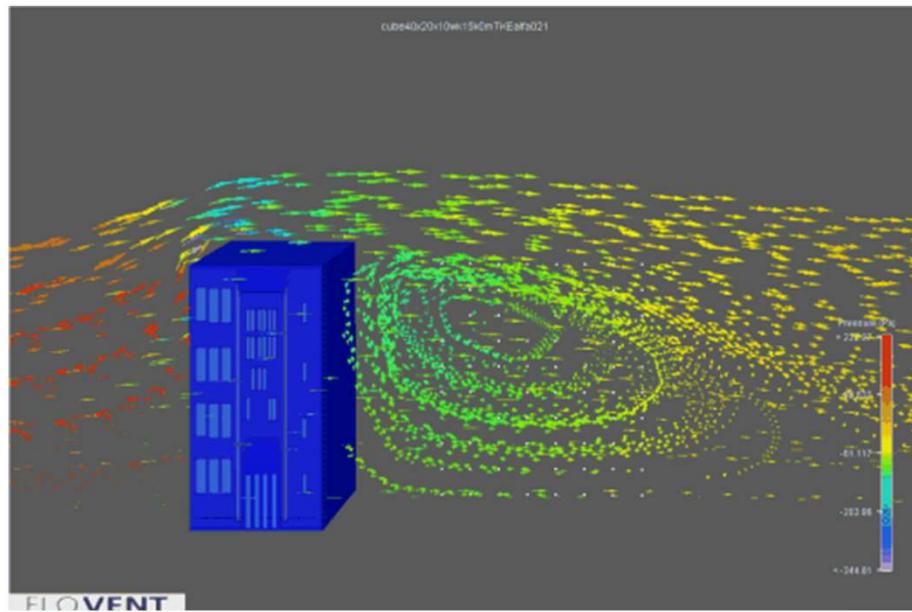
# Infographics

# Storytelling via graphical data representation.

Aesthetic is more important than accuracy.

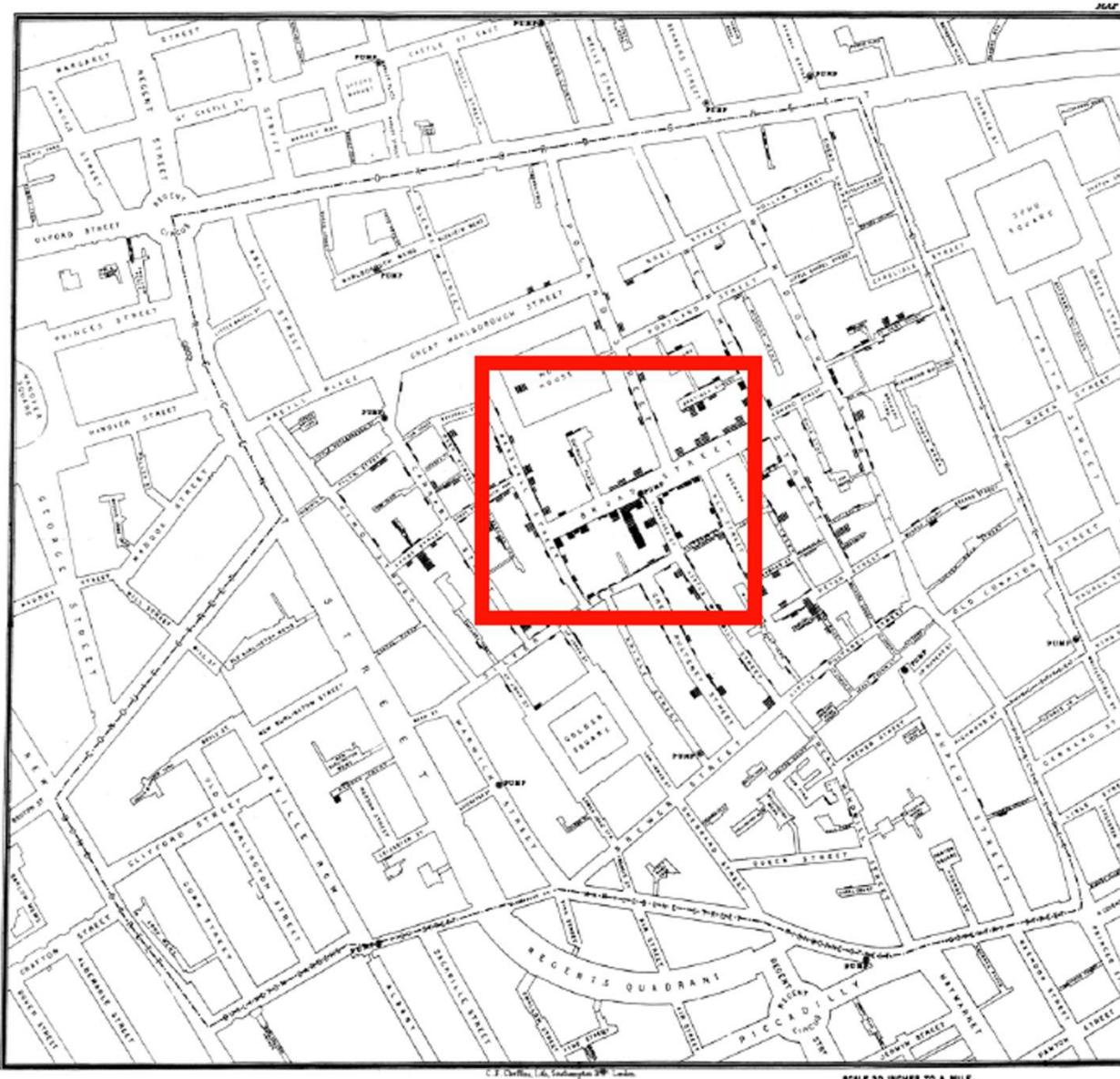
# Scientific Visualization

Visualize data from scientific instruments and simulation.



# Examples

# Classic Example: Dr.John Snow's Cholera Map (1855)



To stop the outbreak of cholera in London in 1854, **Dr. John Snow marked the cholera deaths on a map**. This map visualization indicated that the water from a pump on **Broad Street** was to blame as a large number of deaths were marked close to that pump. Snow's visualization is one of the most important early examples of epidemiology, that **clearly linked cholera's spread to water and not air**.

Snow, 1855 in  
*On the Mode of  
Communication of Cholera*

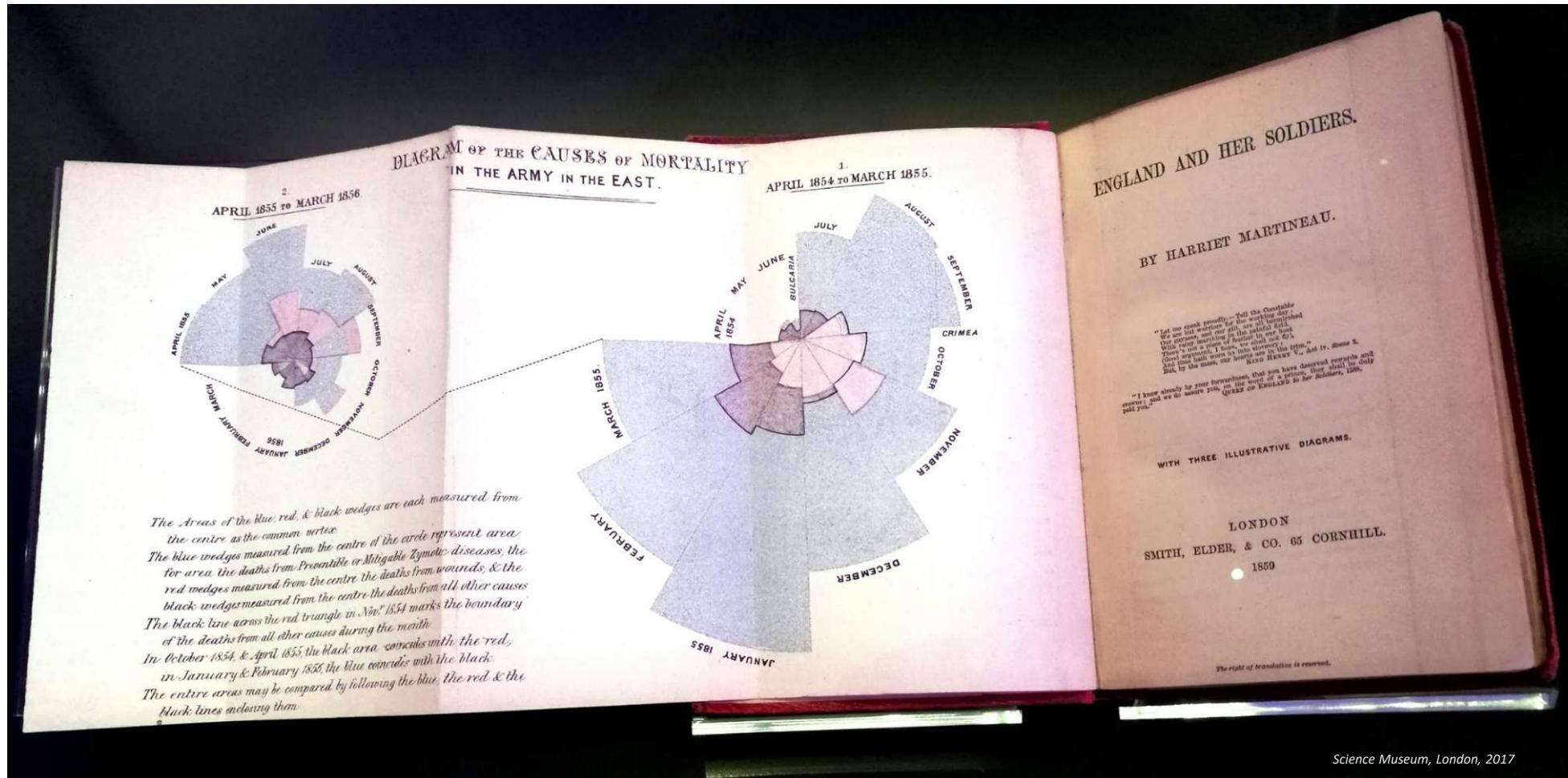
# Classic Example: Dr.John Snow's Cholera Map (1855)



To stop the outbreak of cholera in London in 1854, **Dr. John Snow marked the cholera deaths on a map**. This map visualization indicated that the water from a pump on **Broad Street** was to blame as a large number of deaths were marked close to that pump. Snow's visualization is one of the most important early examples of epidemiology, that **clearly linked cholera's spread to water and not air**.

Snow, 1855 in  
*On the Mode of  
Communication of Cholera*

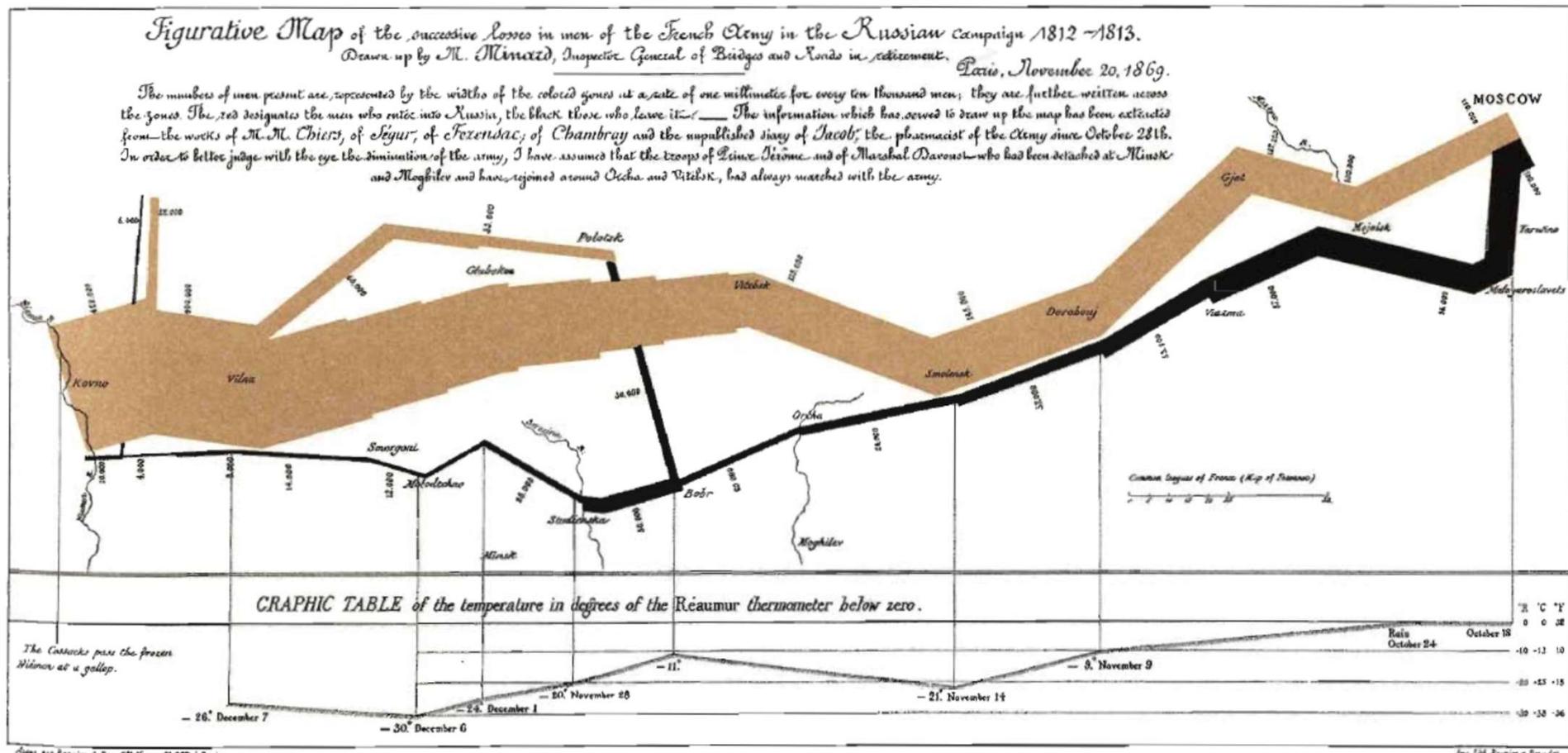
# Classic Example: Florence Nightingale's Diagram (1859)



Science Museum, London, 2017

- Polar area diagrams (or coxcomb diagram) showed the causes of soldiers' deaths in each month.
- The right diagram showed the data of the first year when she arrived the hospital.
- Blue area showed deaths from preventable diseases due to bad conditions in hospital. Red area showed deaths from battlefield wounds. Black represented other causes.
- The left diagram showed the data of the second year after she implemented hygiene practices.

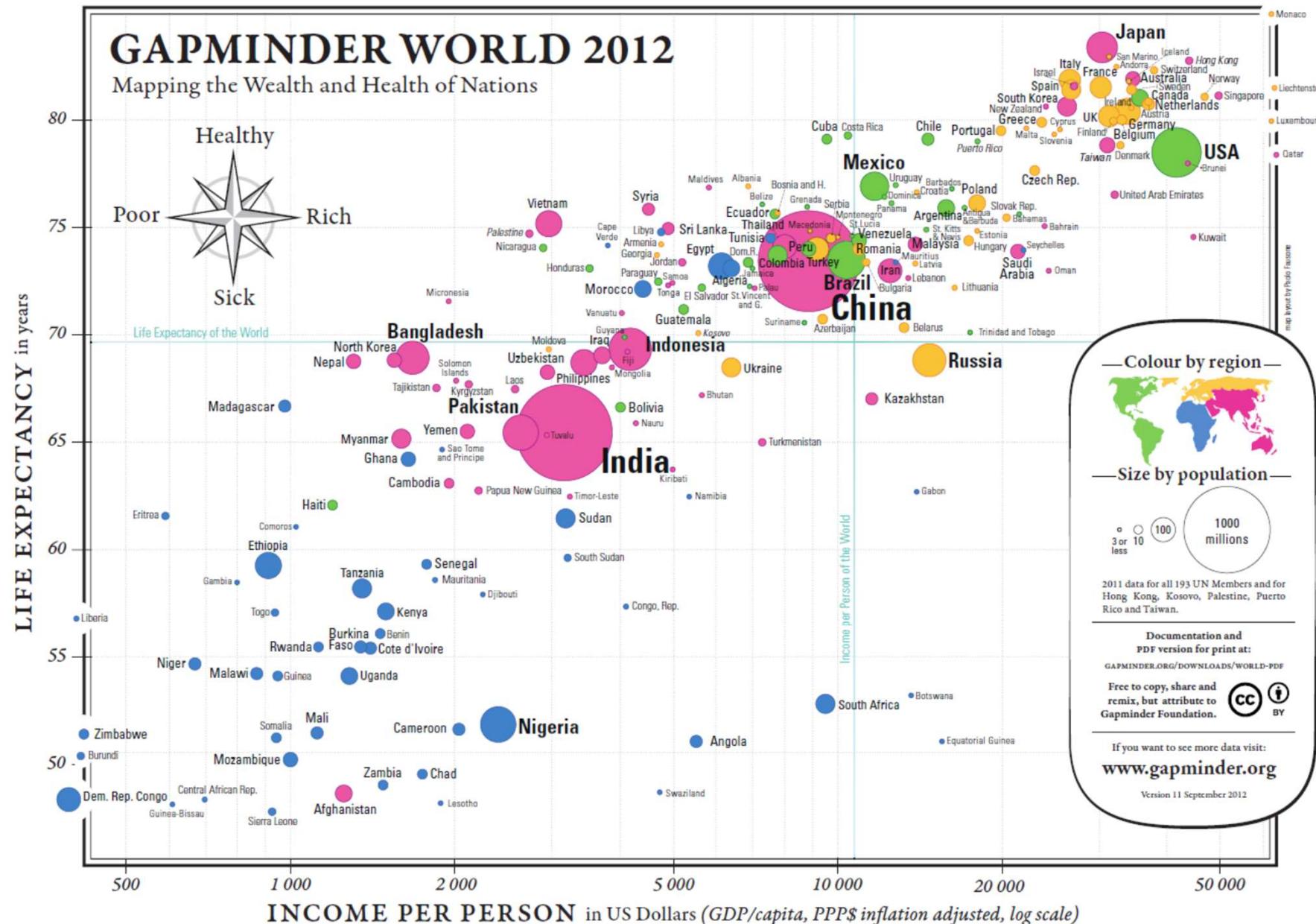
# Classic Example: Napoleon's army in Russia (1869)



- Charles Joseph Minard (1869)
- In 1812-1813, Napoleon led the army of 422,000 men from Poland border to Moscow (north-east direction) and then retreated. Only 10,000 men came back.
- Combined data map and time-series
- A flow line represents the route on the map, labeled with place names. Line thickness represents size of army at each place. The different colors represent the directions.
- The time-series chart show temperature and date.
- 6 variables are plotted: location of army on two-dimensional map, army's size, direction of army's movement, temperature, and date.

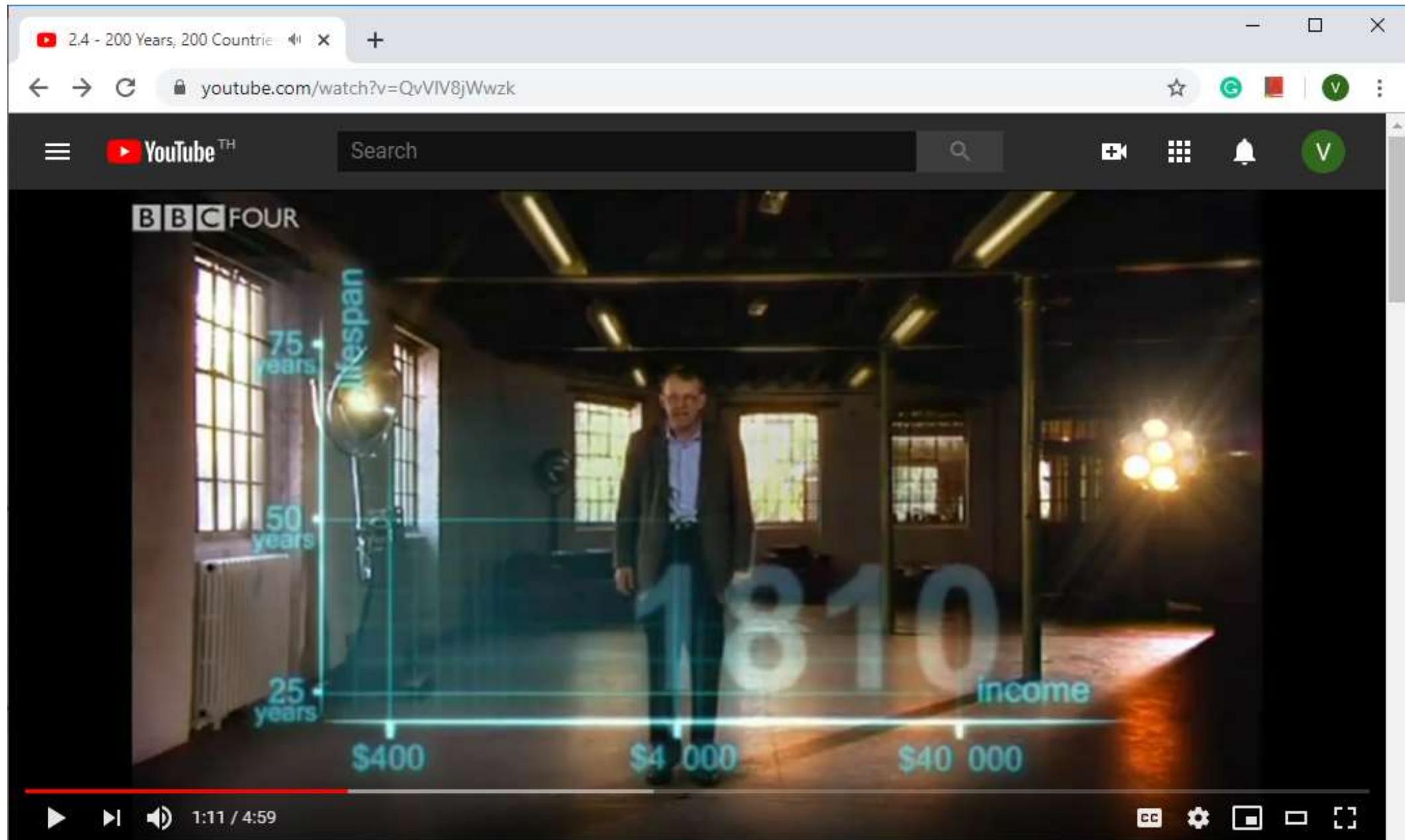
# Gap Minder

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



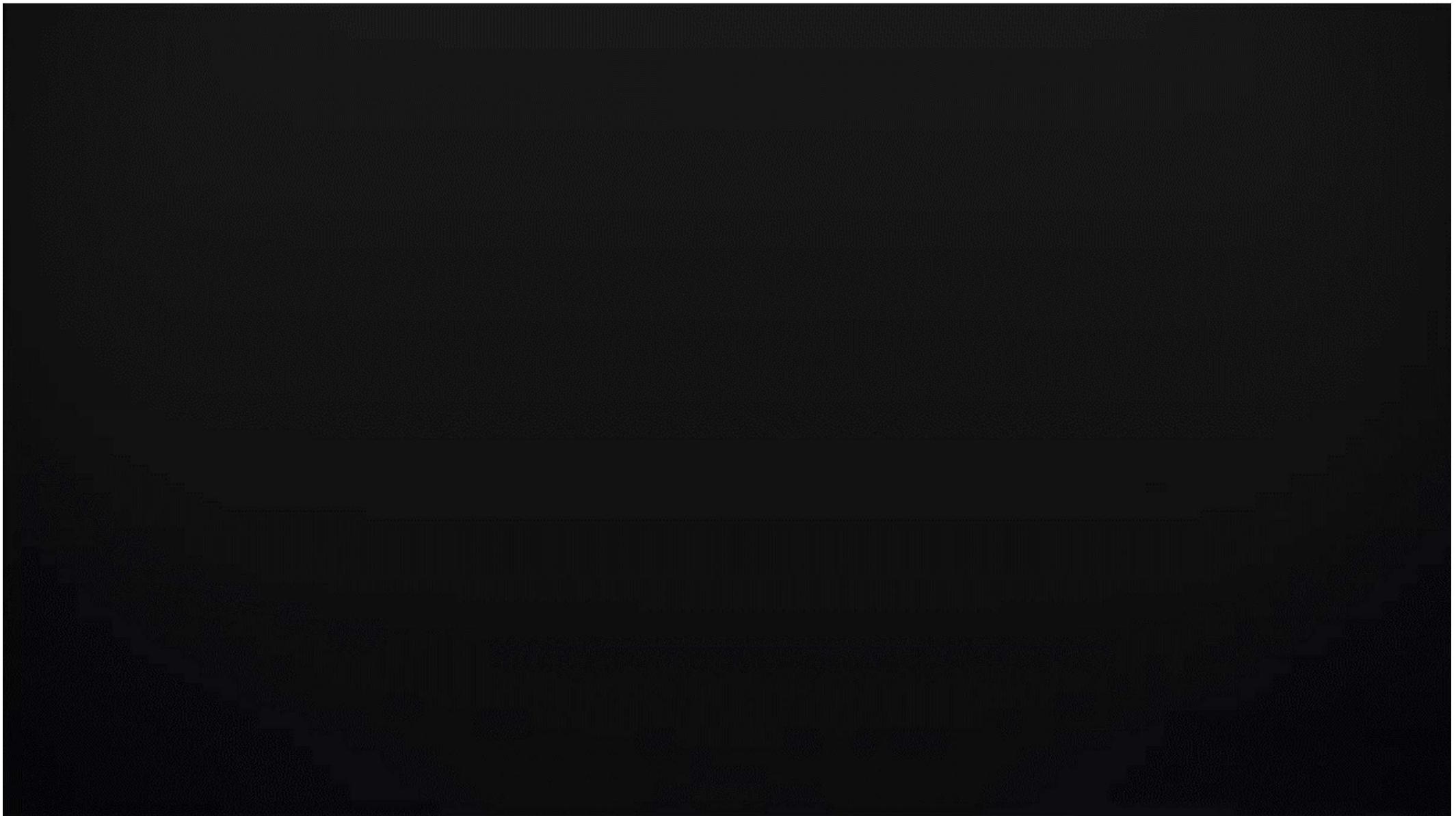
# Hans Rosling's Visualization Demo

<https://www.youtube.com/watch?v=QvVlV8jWwzk>



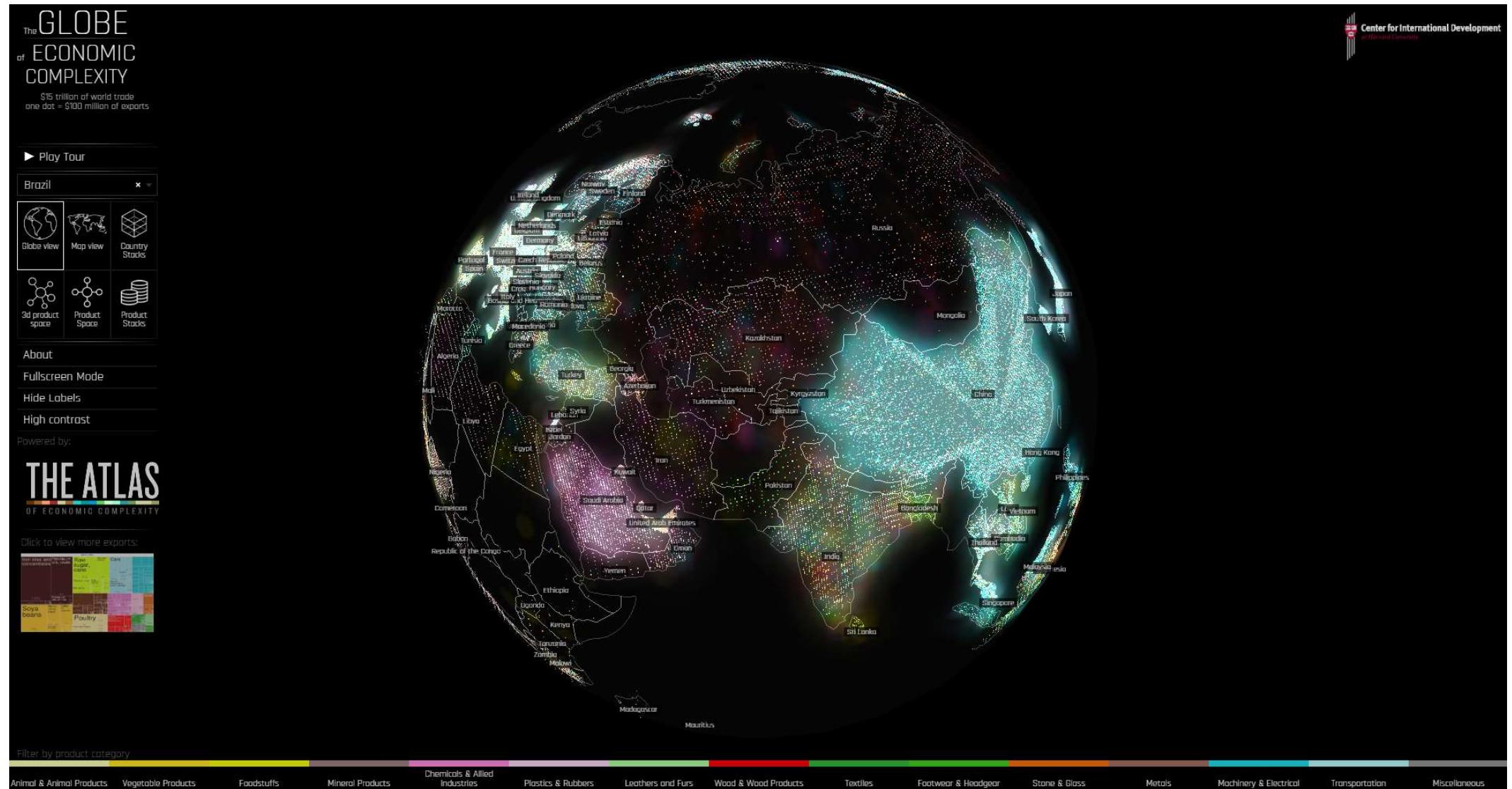
# Hans Rosling's Visualization Demo

<https://www.youtube.com/watch?v=QvVlV8jWwzk>



# The Globe of Economic Complexity

<http://globe.cid.harvard.edu/?mode=gridSphere&id=BA>



# The Globe of Economic Complexity

[https://youtu.be/Obuq\\_L2U4VU](https://youtu.be/Obuq_L2U4VU)

Imagine world economies as  
a cloud of confetti

# Principles of Graphical Excellence

What makes a good visualization?

What is a good data visualization?

Pretty charts and graphics?

# What is a good data visualization?

- Help viewer **understand the structure** of the data
- Help viewer **gain insight** into the data
- Help viewer **do tasks** more effectively
  - (e.g. presentation, story telling, question answering, analysis, exploration, decision making)

# Is Dr.Snow's visualization good?



- Help viewer **understand the structure** of the data
- Help viewer **gain insight** into the data
- Help viewer **do tasks** more effectively

# Principles of Graphical Excellence (Edward Tufte)

- Graphical excellence is well-designed presentation of interesting data,
- consists of complex ideas communicated with clarity, precision, and efficiency,
- gives the greatest number of ideas in the shortest time with the least ink in the smallest space,
- and tells the truth about the data.

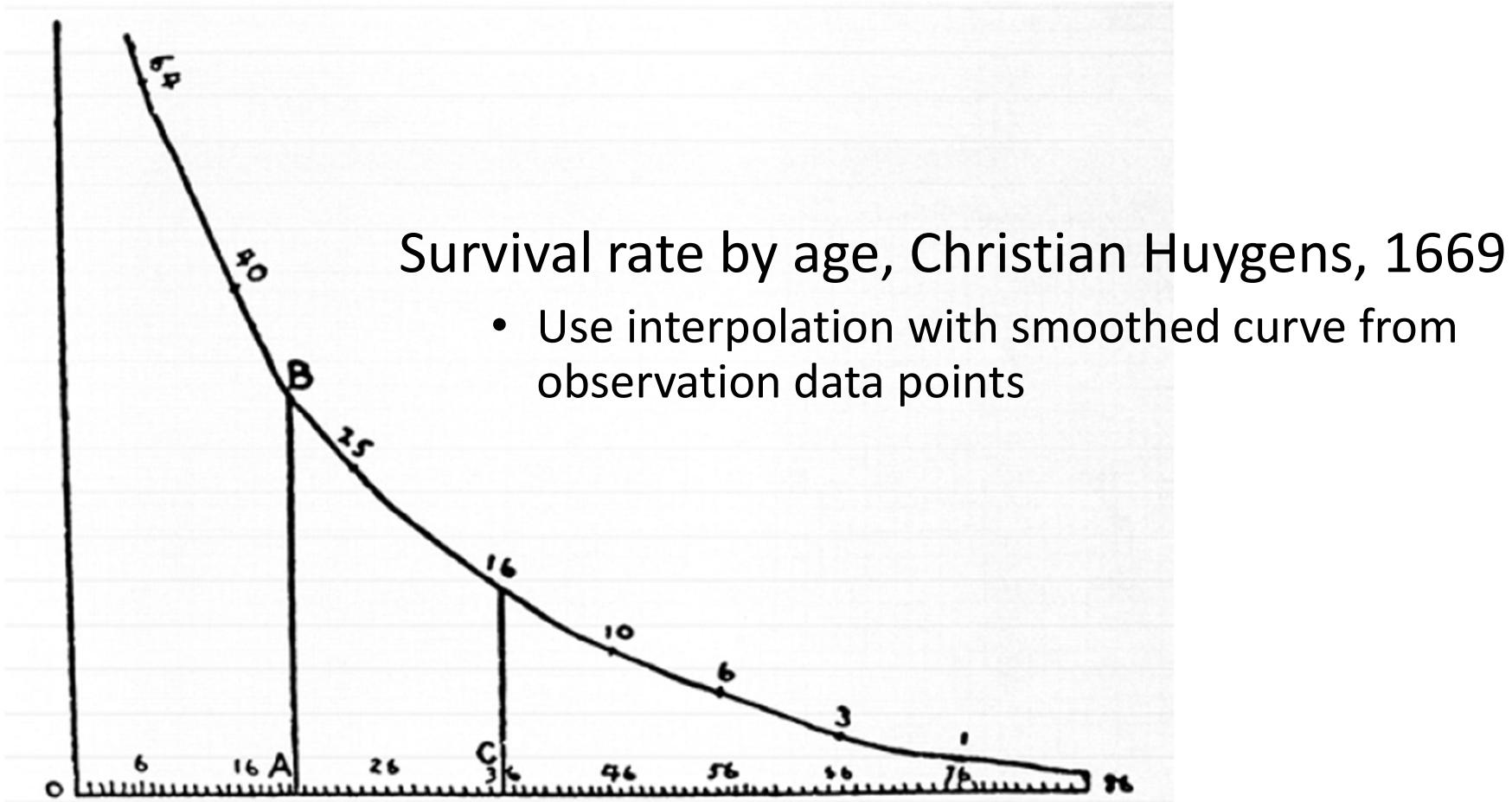
# Types of Visualization

# Main types of data visualization

- Relational graphics
- Time-series
- Spatial data map
- Network graph

# Relational Graphics

Show quantitative relationships between variables



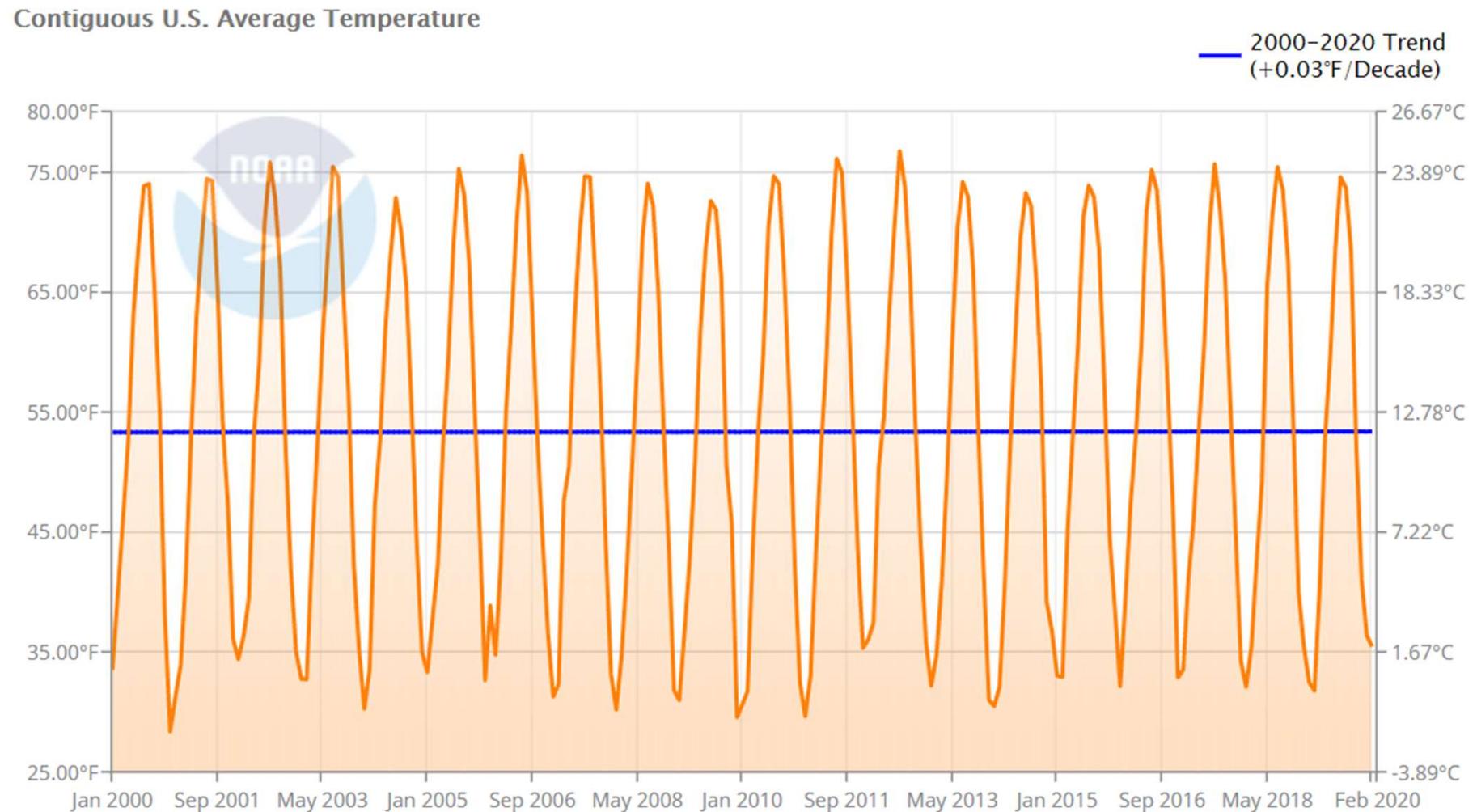
# Time series

Show change over time and repeating patterns

Daily barometric pressure in Oxford, Robert Plot, 1685



# Time series

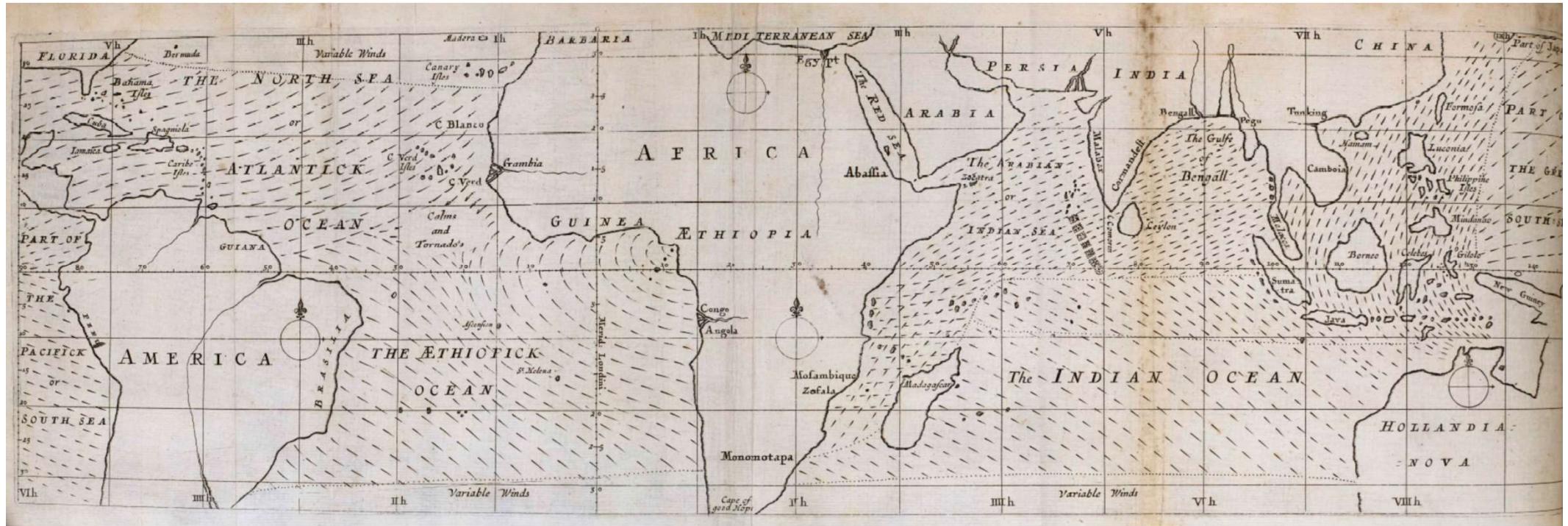


# Spatial data map

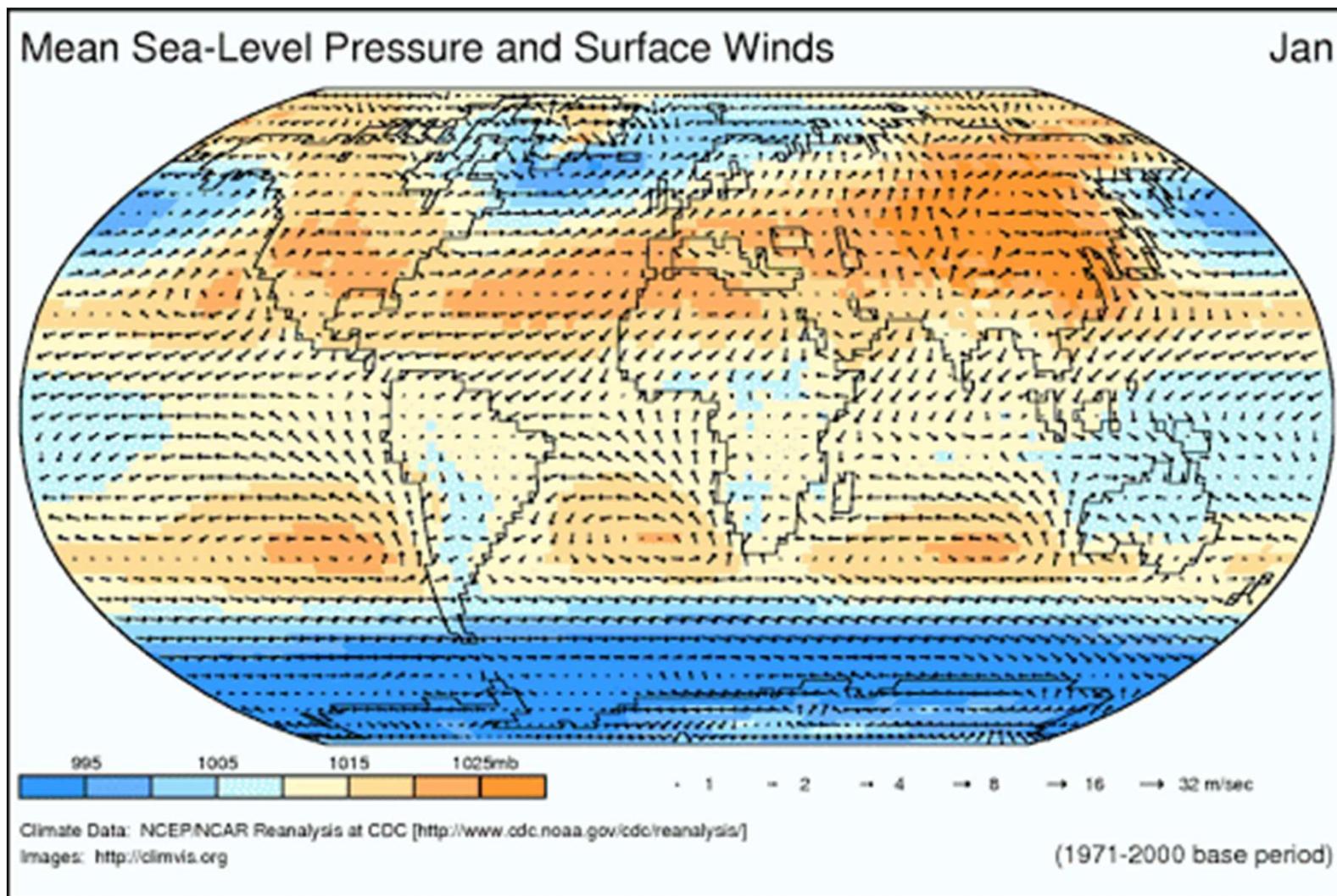
Show relationships between data and geolocation on map

Map of Trade Winds and Monsoons, Edmond Halley, 1686

- Use line symbols to represent direction



# Spatial data map



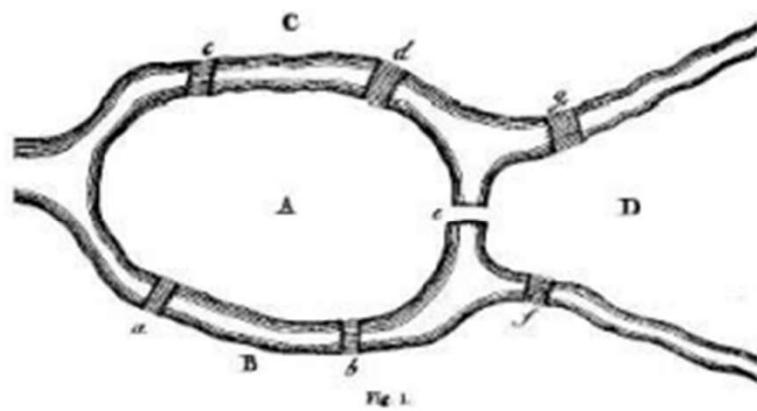
# Network graph

Show relationships between members of a set

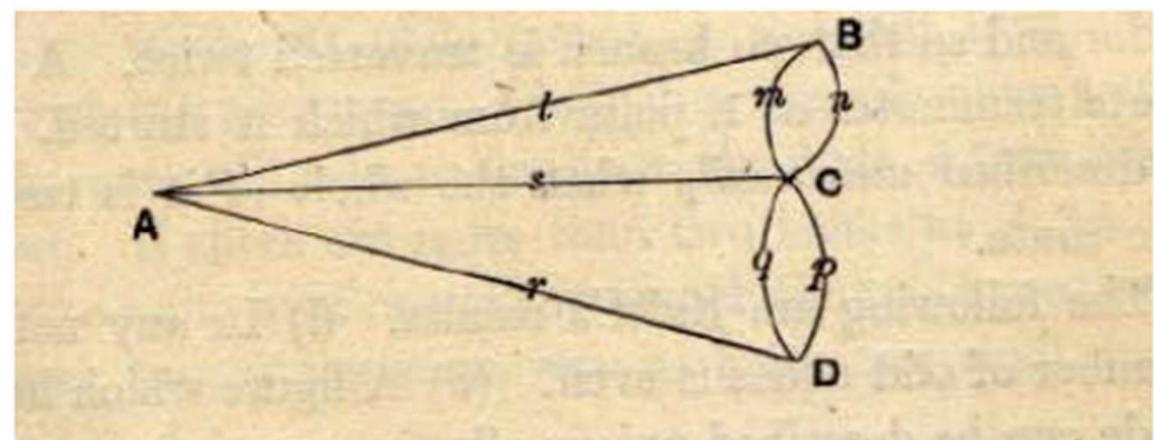
Node=entity Edge=relationship

The seven bridges of Königsberg problem, Euler (1736), Rouse Ball (1892)

*Is it possible to take a walk through the town in such a way as to cross over every bridge once, and only once?*

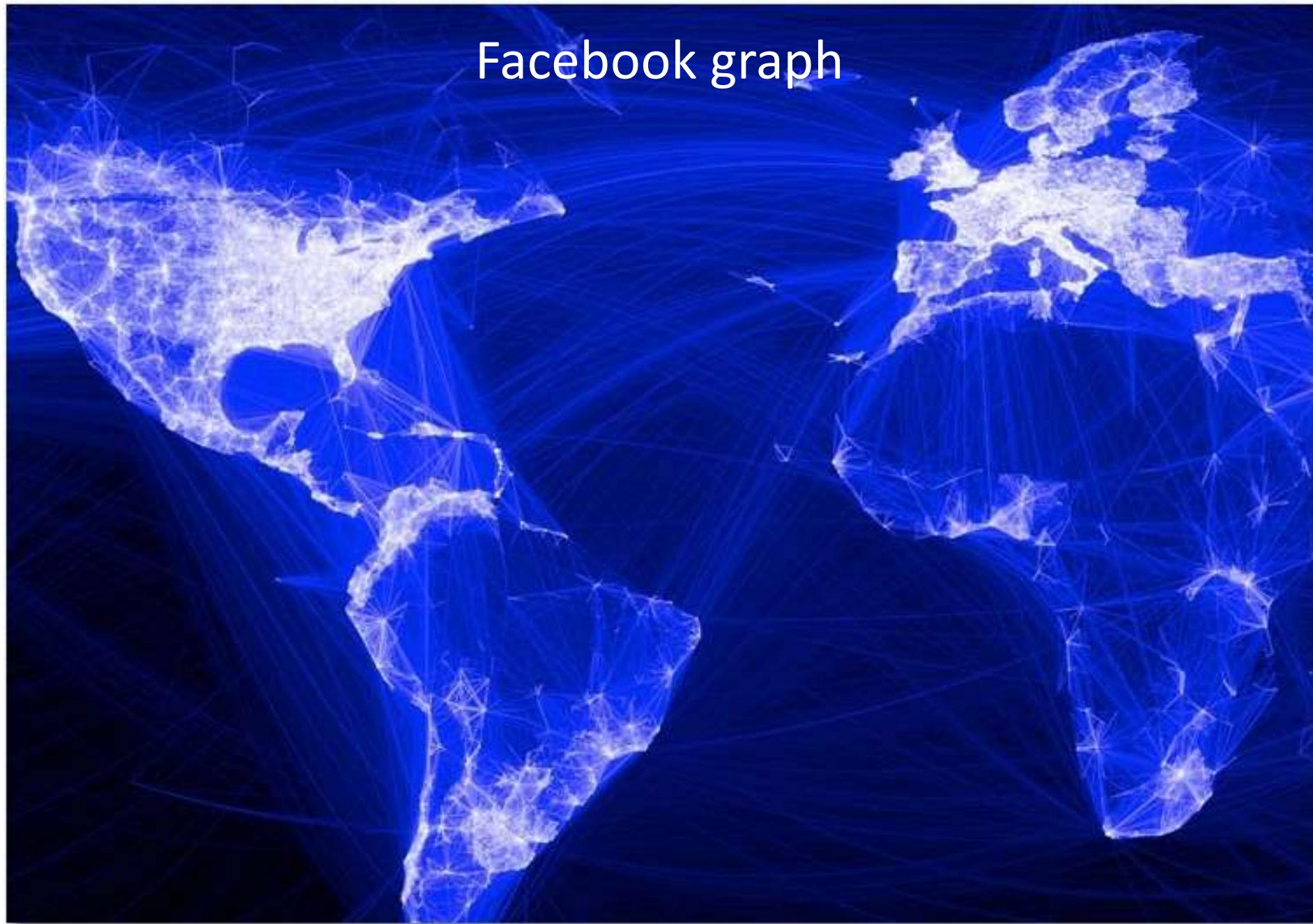


(a)



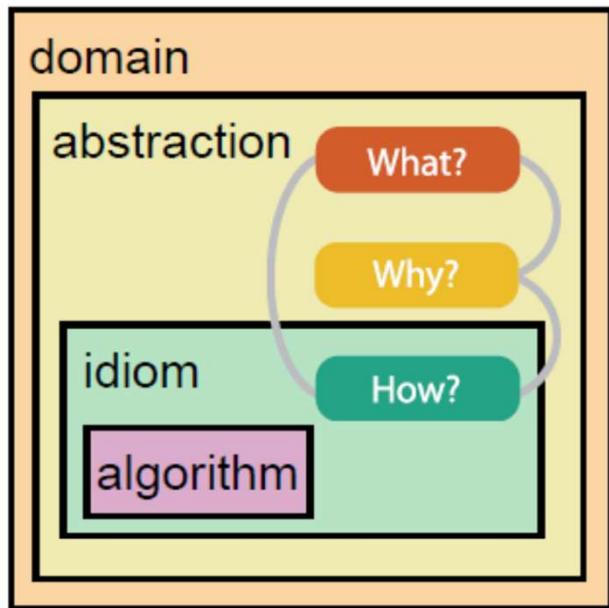
(b)

# Network Graph

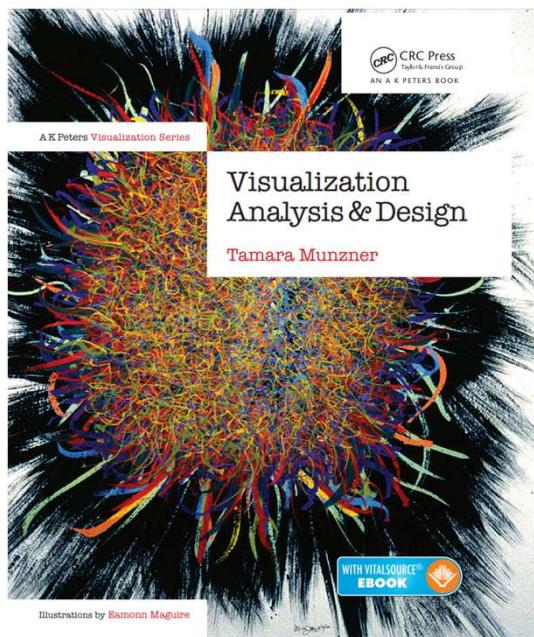


# Visualization Design

# Munzner's Visualization Design Methodology



- **What** is shown?
  - Data abstraction
- **Why** is the user looking at it?
  - Task abstraction
- **How** is it shown?
  - Visual encoding



What?

Why?

How?

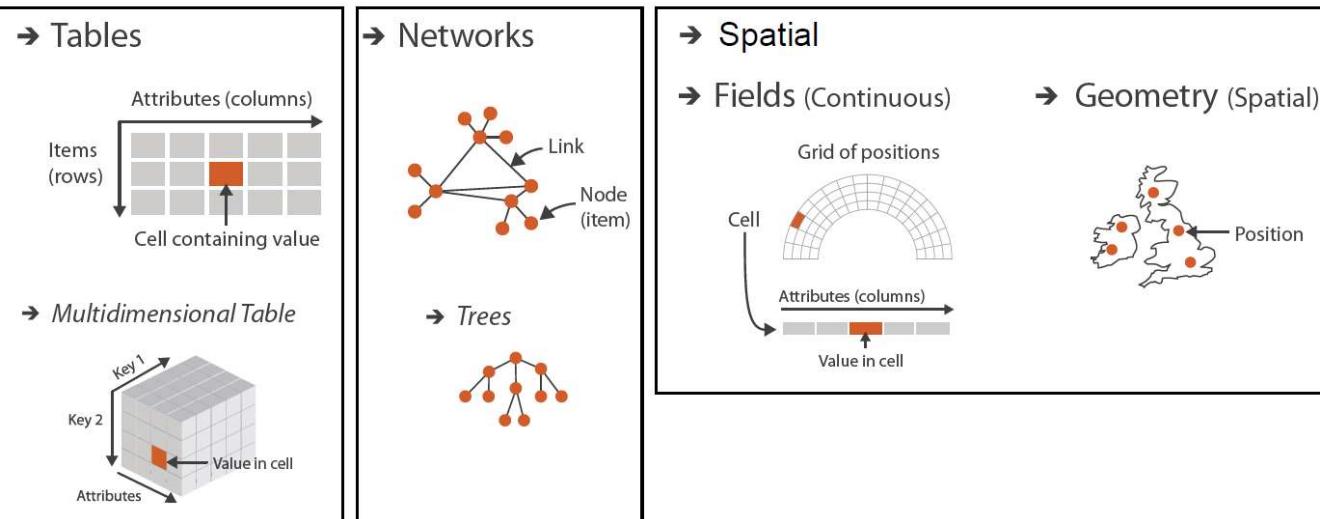
What is shown?

Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box		7/17/07
32	7/16/07	2-High	Medium Box		7/18/07
32	7/16/07	2-High	Medium Box	0.63	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	5	4-Not Specified	Small Pack	0.44	6/6/05
69	5	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05

## ➔ Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry
Items	Items (nodes)	Grids	Items
Attributes	Links	Positions	Positions

## ➔ Dataset Types



# Attribute Types

## ④ Attribute Types

→ Categorical

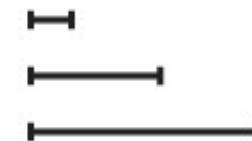


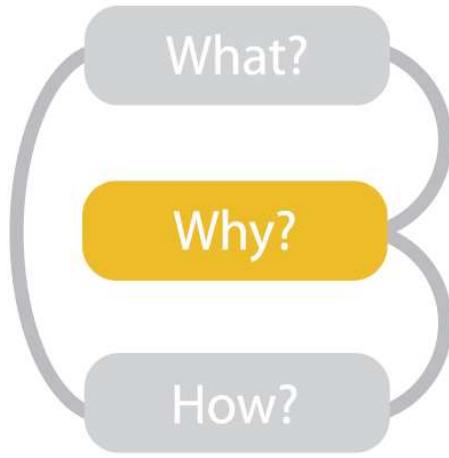
→ Ordered

→ *Ordinal*

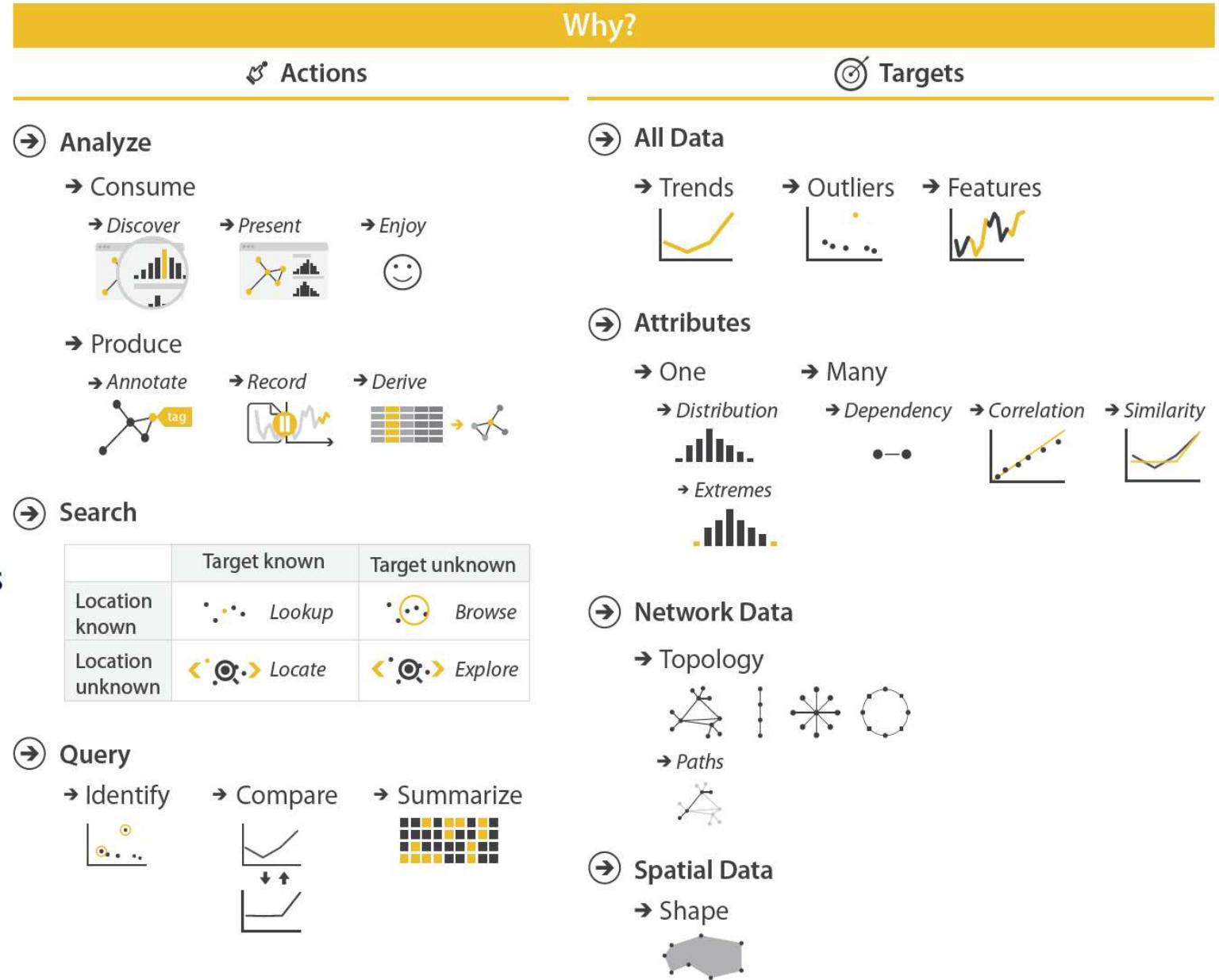


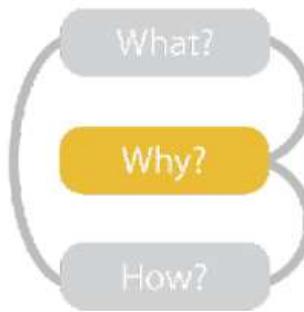
→ *Quantitative*



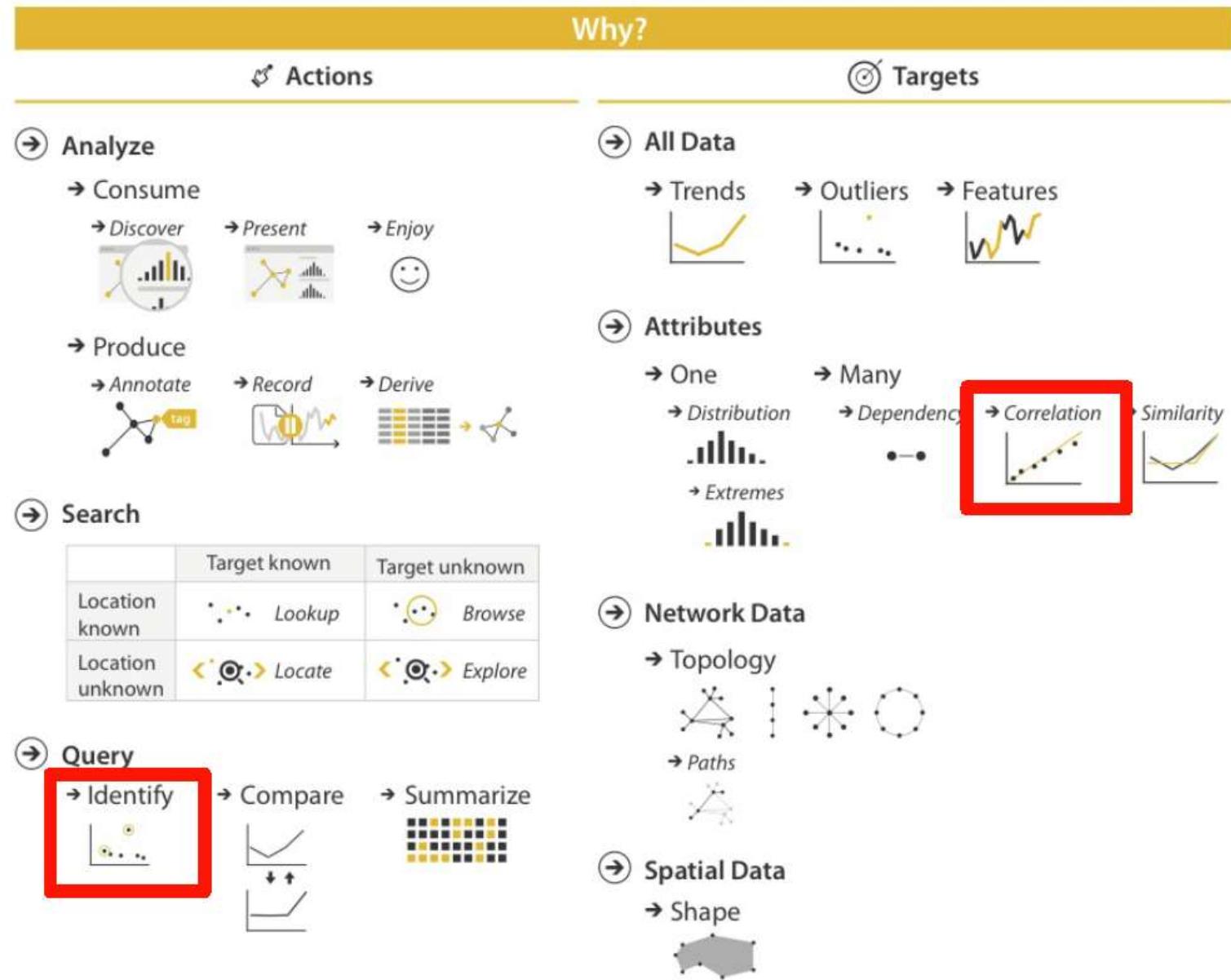
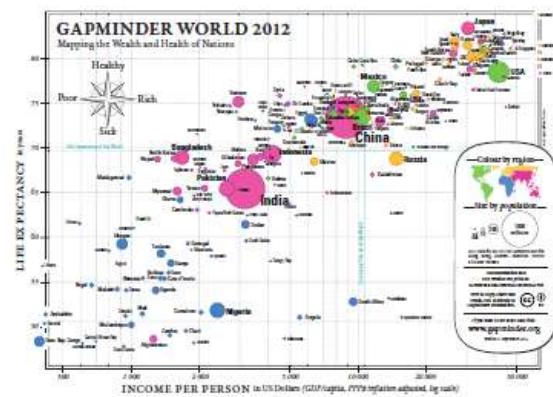


- {action, target} pairs
  - discover distribution
  - compare trends
  - locate outliers
  - browse topology





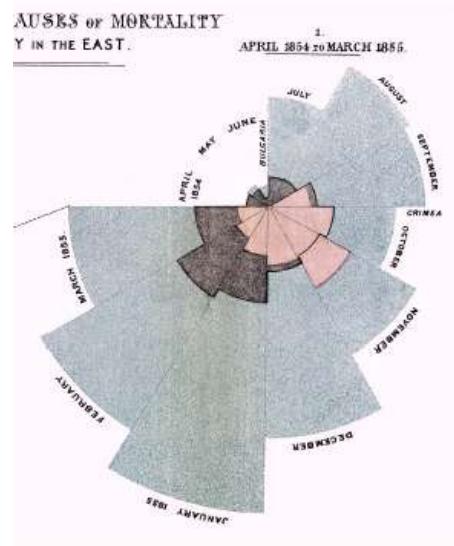
**Why** look at it?



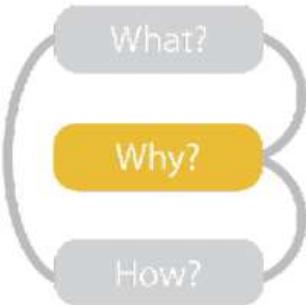
[Munzner, 2014]



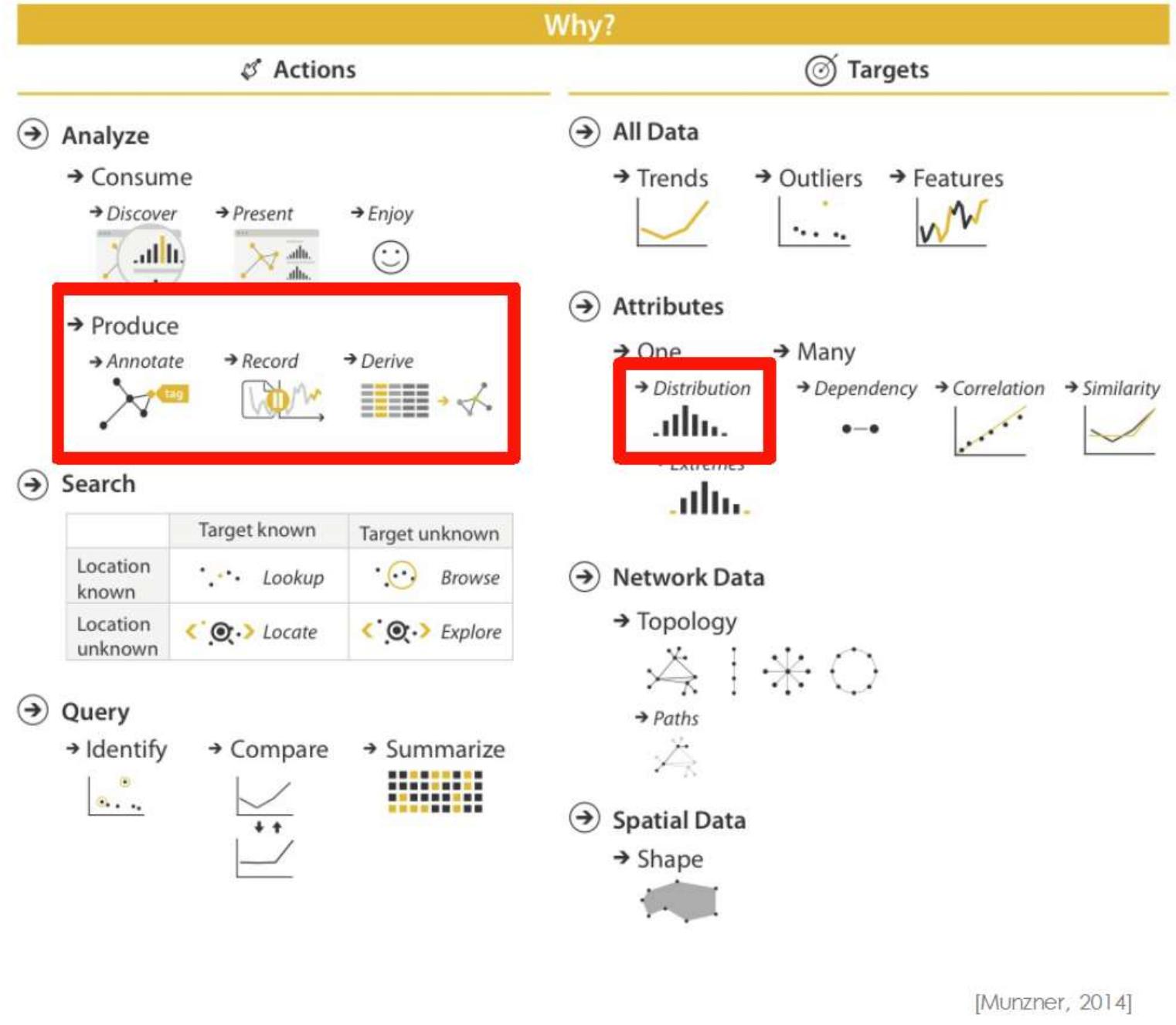
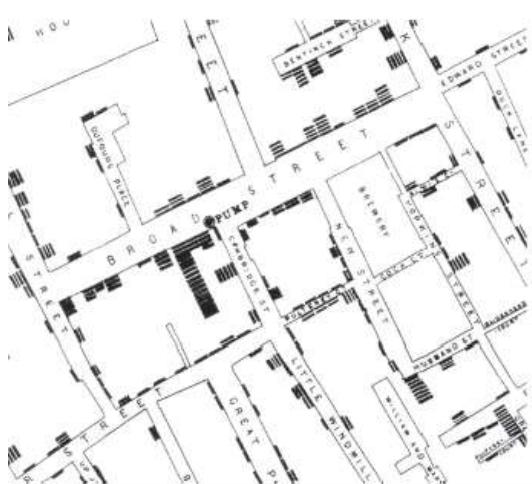
**Why** look at it?

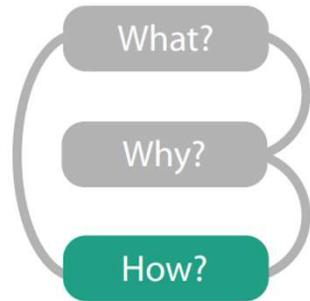


[Munzner, 2014]



**Why** look at it?





How shown?

## Marks

Geometric primitives used to construct any visualization

### → Points



### → Lines



### → Areas



## Visual Variables

The properties or appearance of marks that can be used to represent the data

### → Position

→ Horizontal



→ Vertical



→ Both



### → Color



### → Shape



### → Tilt



### → Size

→ Length

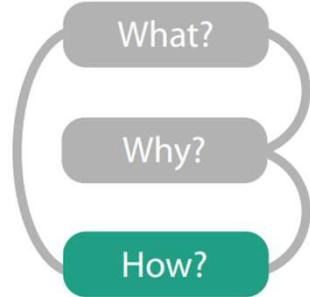


→ Area



→ Volume





How shown?

## Visual Encoding

Representing data with visual variables

## Encode

---

### ④ Arrange

→ Express



→ Separate



→ Order



→ Align



→ Use



### ④ Map

from categorical and ordered attributes

→ Color

→ Hue → Saturation → Luminance

→ Size, Angle, Curvature, ...



→ Shape



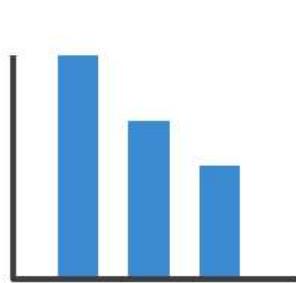
→ Motion

*Direction, Rate, Frequency, ...*



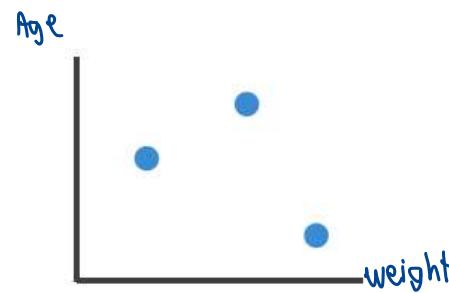
# Visual Encoding

- Combination of marks and visual variables (channels)



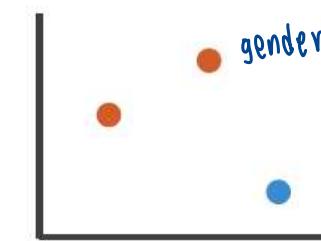
1:  
vertical position

mark: line



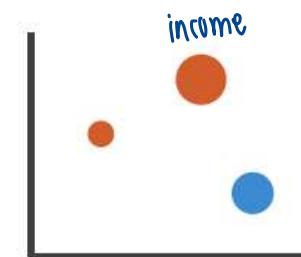
2:  
vertical position  
horizontal position

mark: point



3:  
vertical position  
horizontal position  
color hue

mark: point



4:  
vertical position  
horizontal position  
color hue  
size (area)

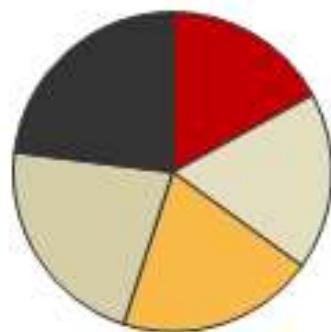
mark: point

# Effectiveness of Visual Variables

# Pie Charts vs. Bar Charts

Which one is better? Why?

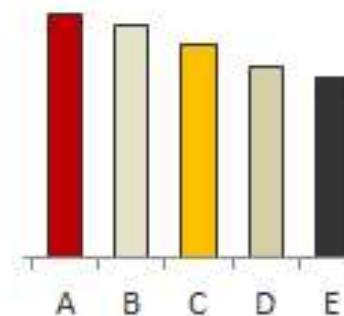
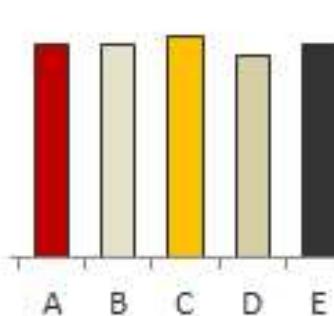
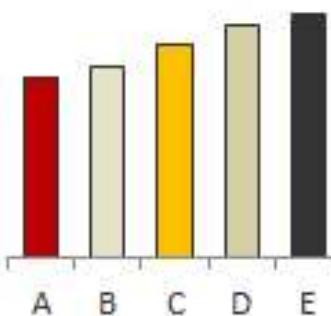
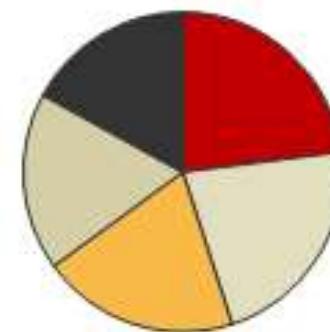
Question 1



Question 2



Question 3

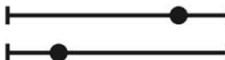


# Effectiveness of a visual variable depends on type of data

## Channels: Expressiveness Types and Effectiveness Ranks

### ④ Magnitude Channels: Ordered Attributes

Position on common scale



Position on unaligned scale



Length (1D size)



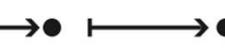
Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



Color saturation



Curvature



Volume (3D size)



### ④ Identity Channels: Categorical Attributes

Spatial region



Color hue



Motion



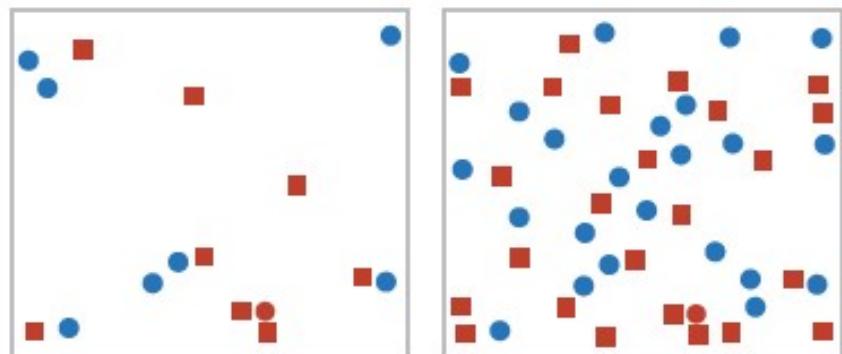
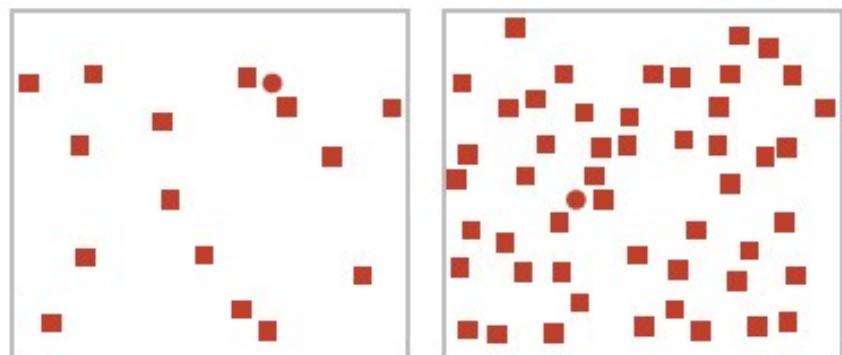
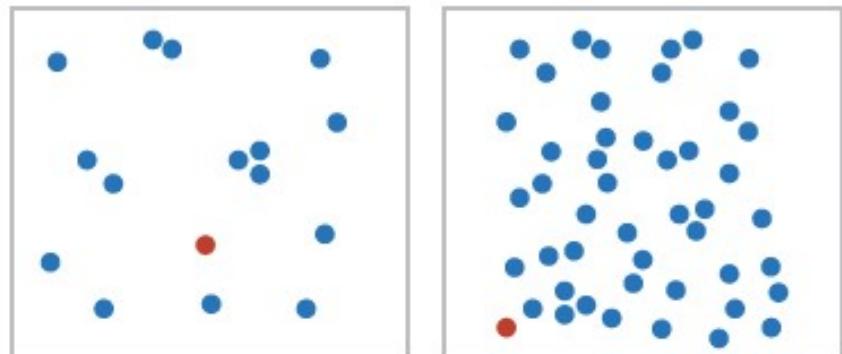
Shape



▲ Most  
Effectiveness  
Same  
Least ▼

# Popout Effect

- Find the **red circular dot**
- Popout: a distinct item stands out from many others immediately
- Speed is independent to number of distractors.
- Different visual variables have different level of popout effect.
- E.g. color has stronger popout effect than shape.
- Combination of visual variables weakens popout effect.
- Serial search is needed.
- Speed depends on number of distractor.



# Color and Order

Implicitly ordered

Luminance



Saturation

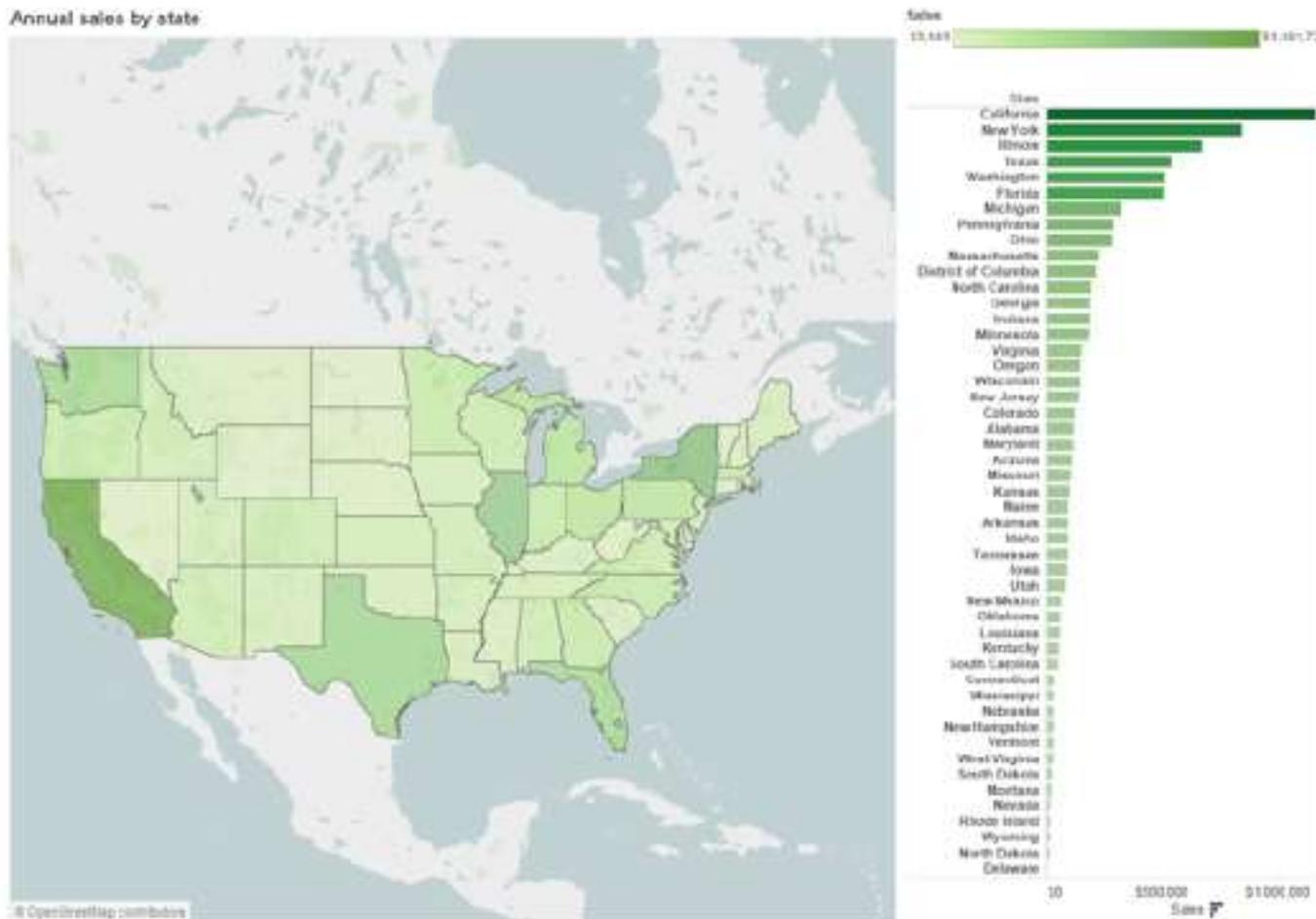


Unordered

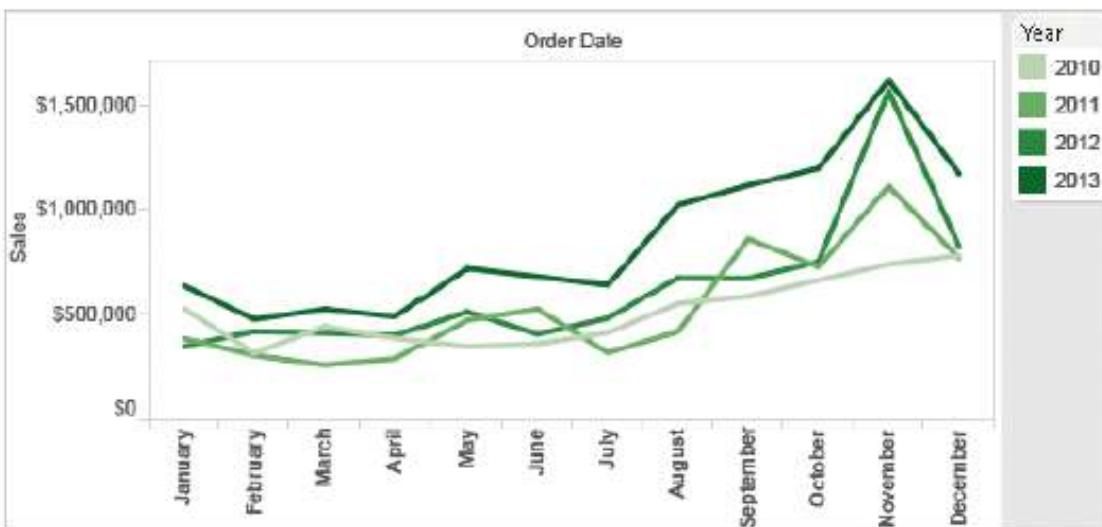
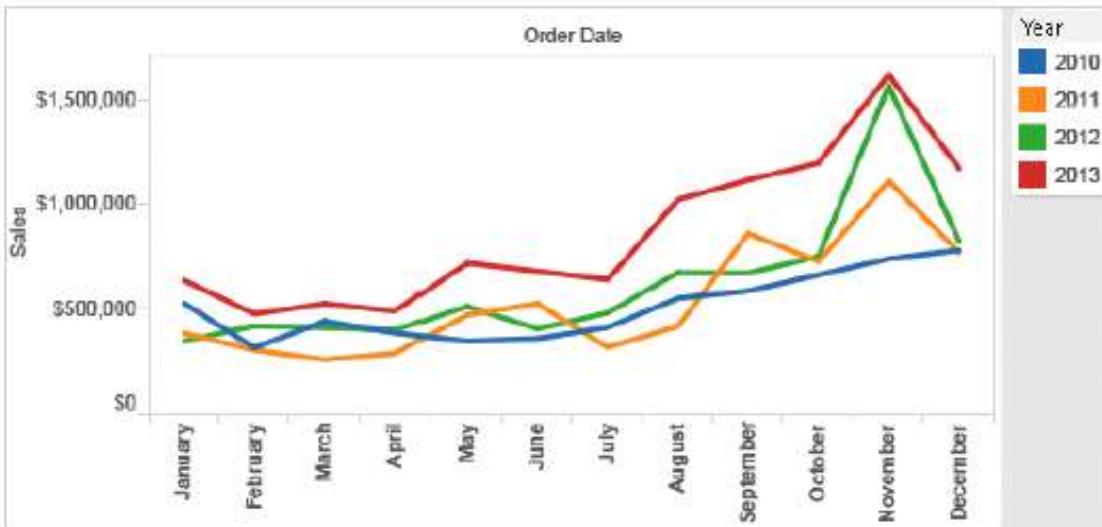
Hue



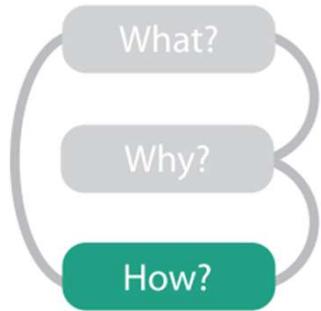
# Ordered by Saturation



# Categorical vs ordered color



# How to Select Chart Type



# How to select chart type

Based on

- Data [ **What** is shown? ]
- Task [ **Why** look at it? ]
- Effectiveness of visual variables
- Other factors (e.g. scalability)

What?

Why?

How?

## Idiom: Bar Chart

**marks:** lines

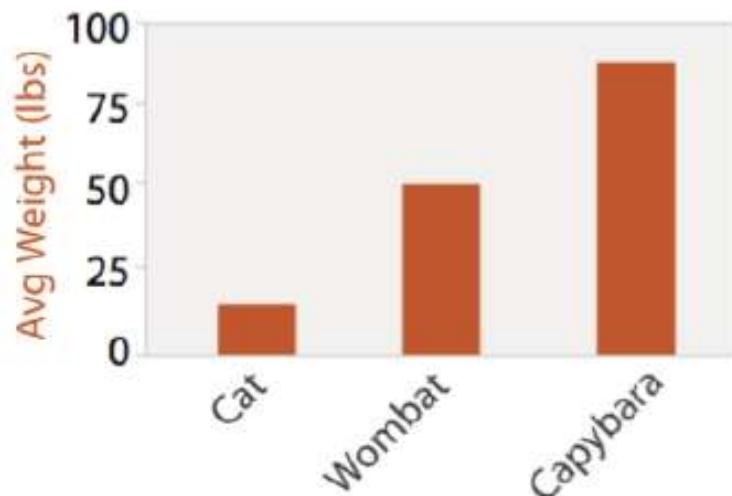
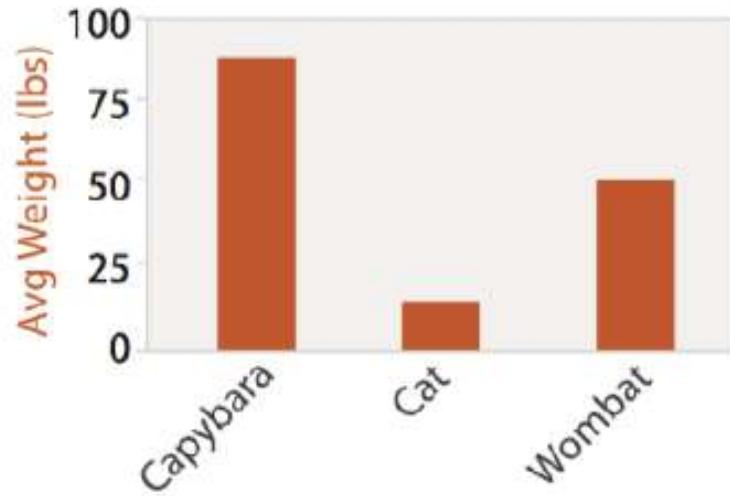
**visual variables:**

length for quantitative value,  
each mark separated horizontally, aligned  
vertically and ordered by label or length

**data:** table with 1 category attribute (key  
attribute) and 1 quantitative attribute

**tasks:** compare, lookup values

**scalability:** dozens to hundreds of levels for key attribute



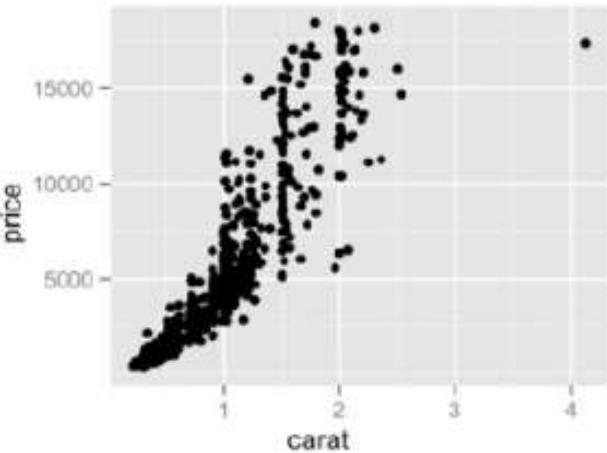
What?

Why?

How?

## Idiom: Scatterplot

**marks:** points

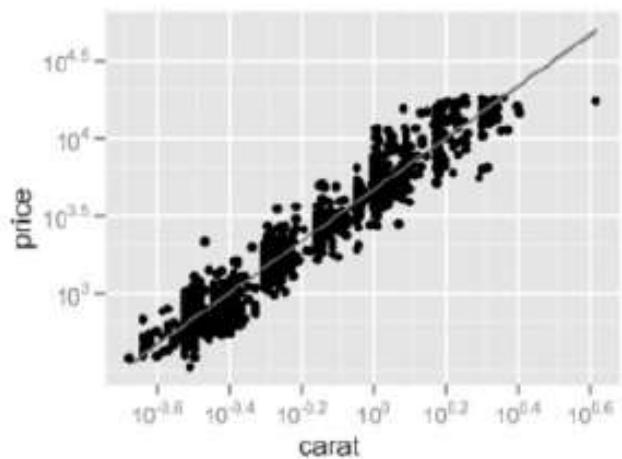


**visual variables:** position (horizontal + vertical)

**data:** table with only 2 quantitative attributes and no key (only values)

**tasks:** finding trends, outlier, distribution, correlation, clusters

**scalability:** hundreds of items



What?

Why?

How?

## Idiom: Line Chart

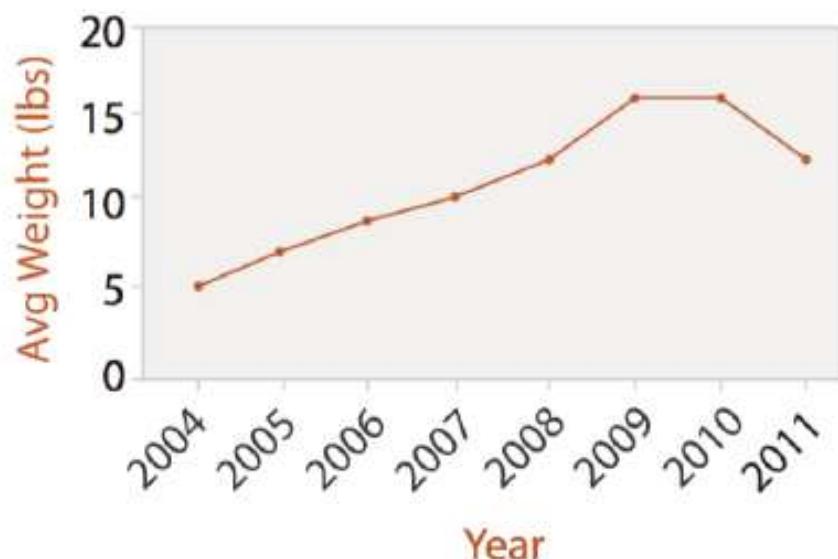
**marks:** points (a line connects the marks)

**visual variables:**

aligned length for quantitative value,  
horizontally separated and ordered by  
key attribute

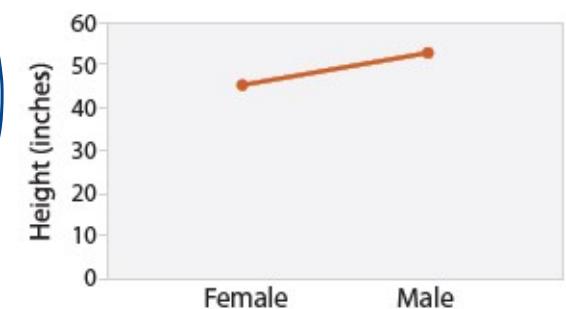
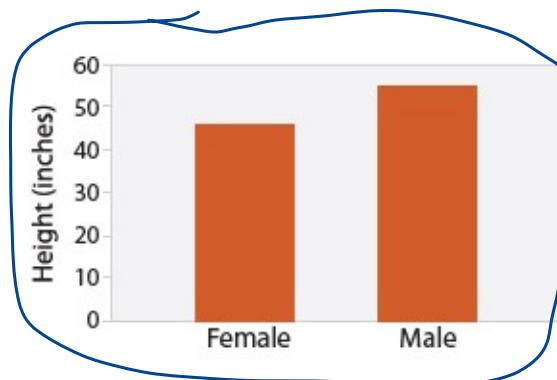
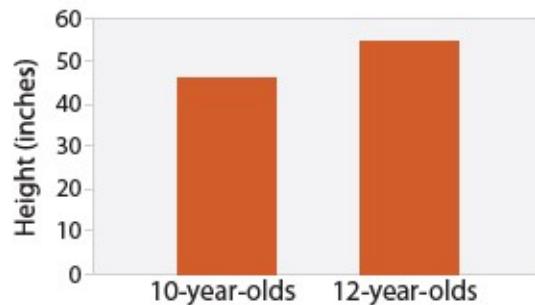
**data:** table with 1 ordered attribute (key  
attribute) and 1 quantitative attribute

**tasks:** find trend (line connecting the marks emphasizes the ordering  
of the items along key axis)



## Choosing bar vs line charts

- depends on type of key attrib
  - bar charts if categorical
  - line charts if ordered
- do not use line charts for categorical key attrs
  - violates expressiveness principle
    - implication of trend so strong that it overrides semantics!
    - “The more male a person is, the taller he/she is”

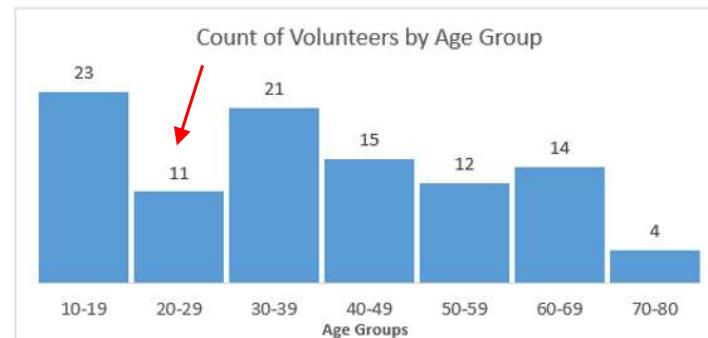
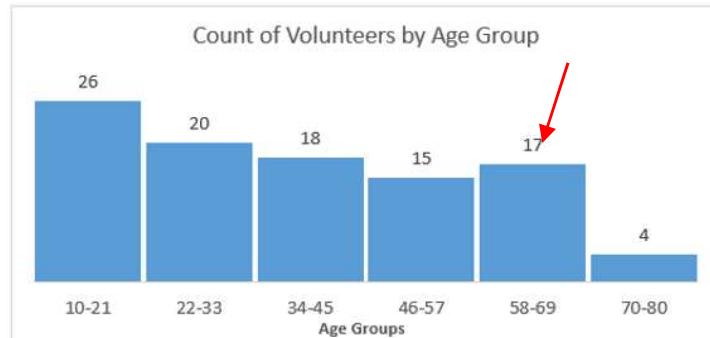
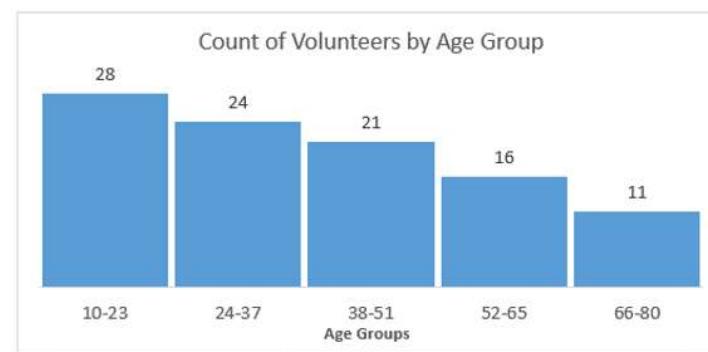
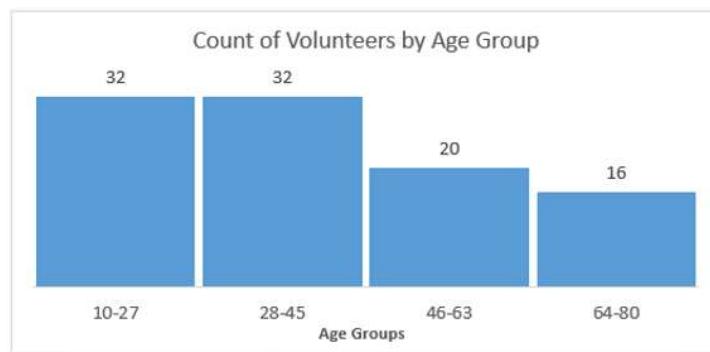
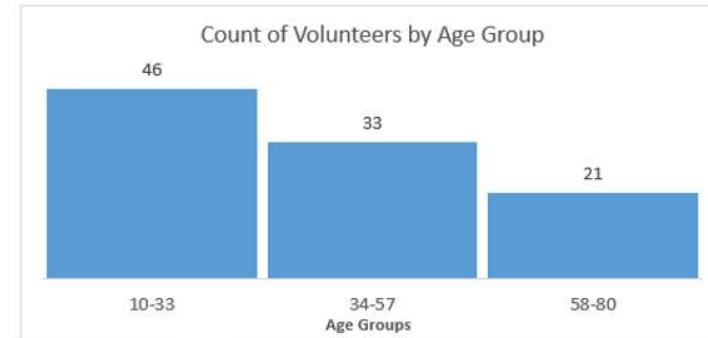
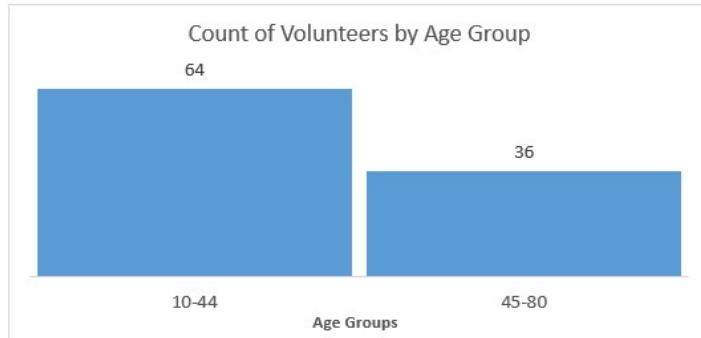


## Idiom: **histogram**

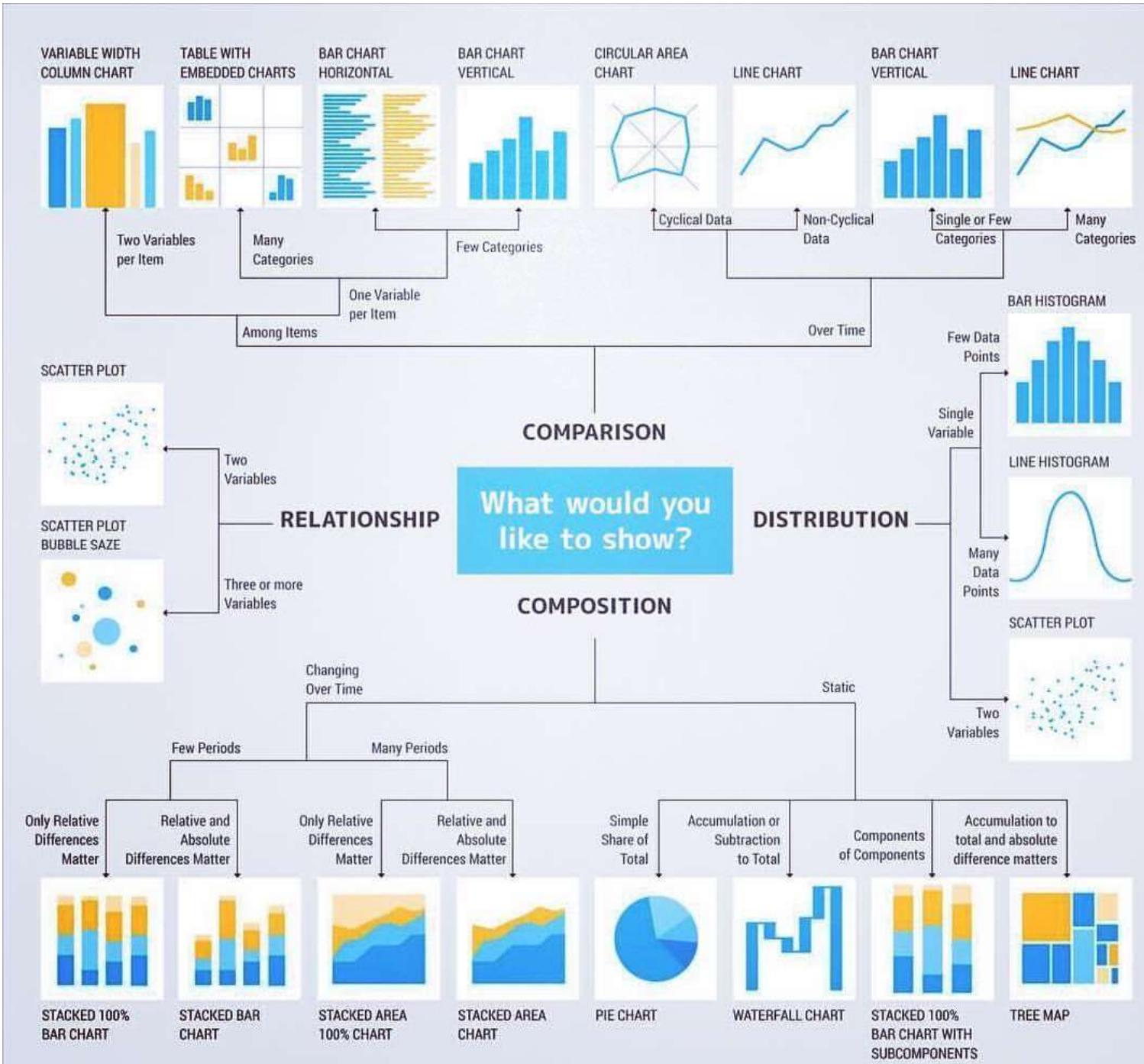
- static item aggregation
- task: find distribution
- data: table
- derived data
  - new table: keys are bins, values are counts
- bin size crucial
  - pattern can change dramatically depending on discretization
  - opportunity for interaction: control bin size on the fly



# Histogram: different bin sizes tell different stories



# How to choose your visualization



# Less is More

**Remove**  
to improve  
(the **data-ink** ratio)

Created by Darkhorse Analytics

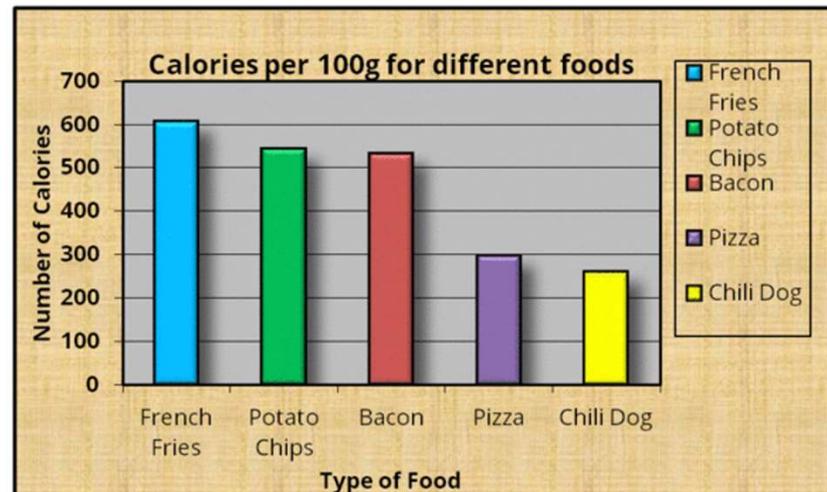
[www.darkhorseanalytics.com](http://www.darkhorseanalytics.com)

Note – In the above visual the author wanted to highlight the data for bacon so the bar for bacon was intentionally left red.

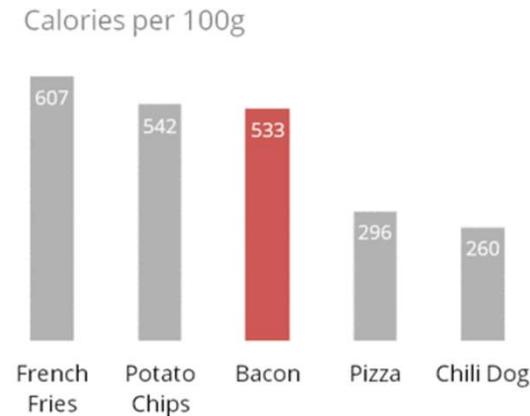
<https://www.e-nor.com/blog/data-visualization/makes-good-visualization>

# Less is More

Before



After



Note – In the above visual the author wanted to highlight the data for bacon so the bar for bacon was intentionally left red.