Data Exploration & Visualization

Course Introduction

Dr. ZENG Wei

DSAA 5024

Hong Kong University of Science and Technology (Guangzhou)

Course Overview

- **Lecturer**: Dr. ZENG Wei
  - Assistant Professor @ CMA & DSA, HKUST (GZ)
  - Research: Data visualization, AR/VR, human-computer interaction
  - Web: zeng-wei.com
  - Email: weizeng@hkust-gz.edu.cn
  - Office: W1-6F-601
  - Office hours: Wed. 2 – 5 pm
- **TAs:**
  - Xingchen Zeng (xzeng159@connect.hkust-gz.edu.cn)
  - Liangwei Wang (lwang344@connect.hkust-gz.edu.cn)

Happy to talk with you about **anything!**

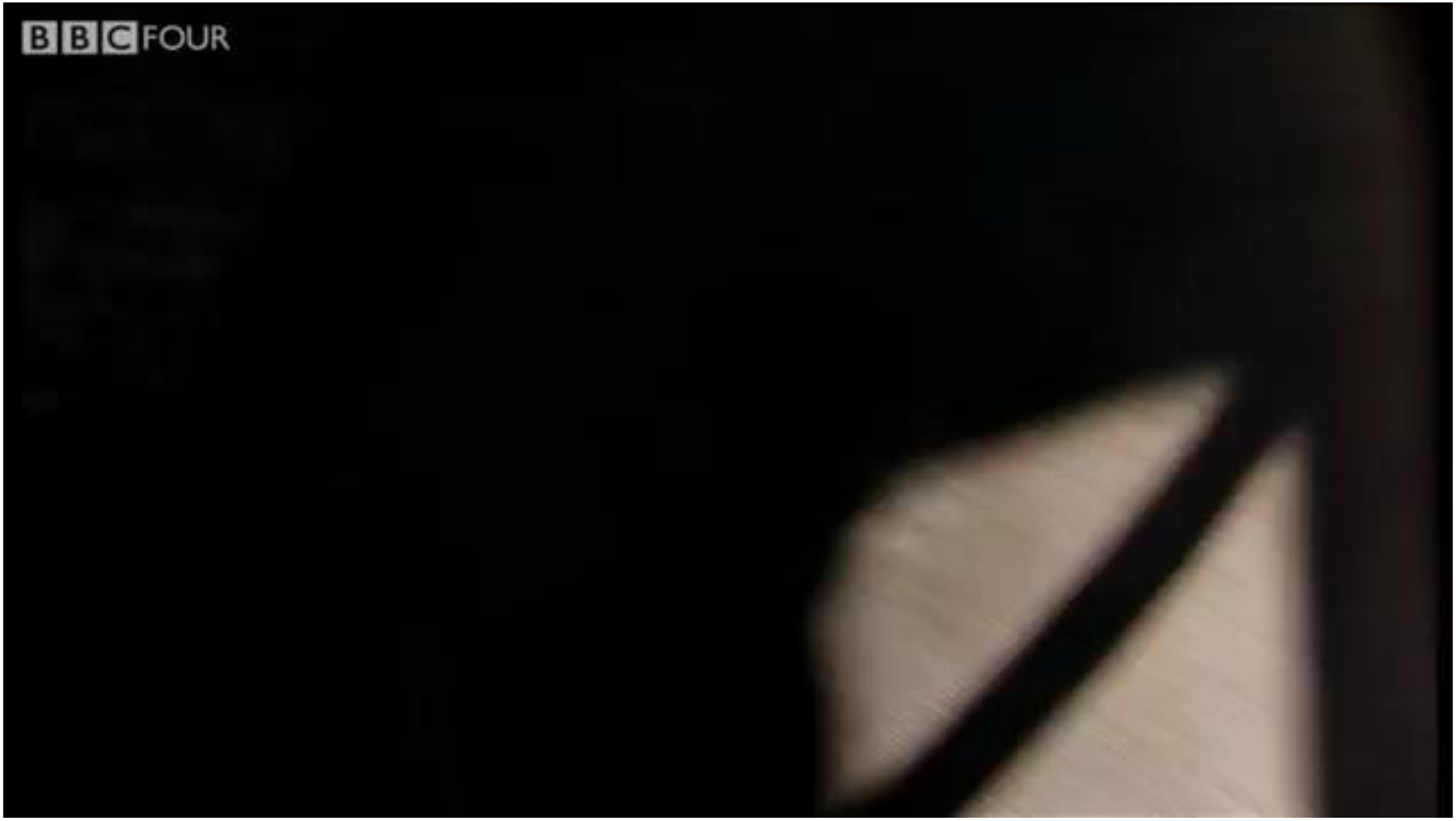
Course Overview

- **Course**
  - 13 weeks – Starting today (3 Sep. 2024)!
  - Tuesday 3:00 – 5:50 pm
  - Venue: Rm 102, E4
  - Assessment
    - Class participation 10%
    - Assignments 20%
    - Quiz x2 20%
    - Final Project 50%
      (report, presentation, demo)

What do you expect from the course?

Examples

- “Hans Rosling’s 200 Countries, 200 Years, 4 Minutes - The Joy of Stats”, BBC 2010.



Examples

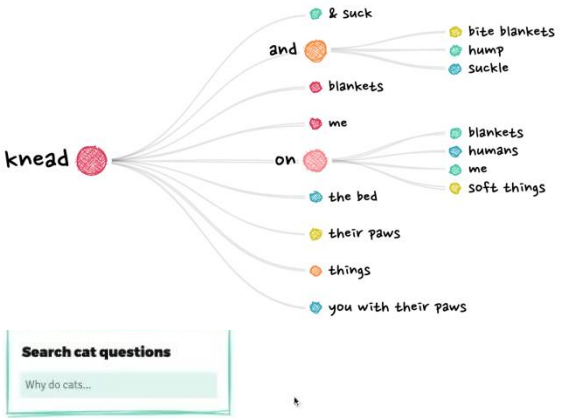
- “The genius of the London Tube Map”, Design legend Michael Bierut @ TED 2018

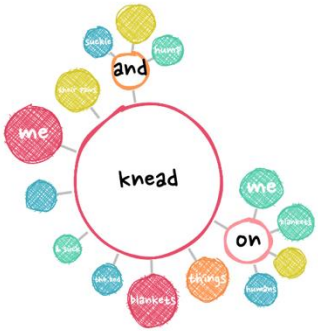
Examples

- A Network of Science: 150 Years of Nature Papers – Nature, 2019
  - Interactive version: https://www.nature.com/articles/d41586-019-03305-w

Examples

- Why do cats & dogs …?

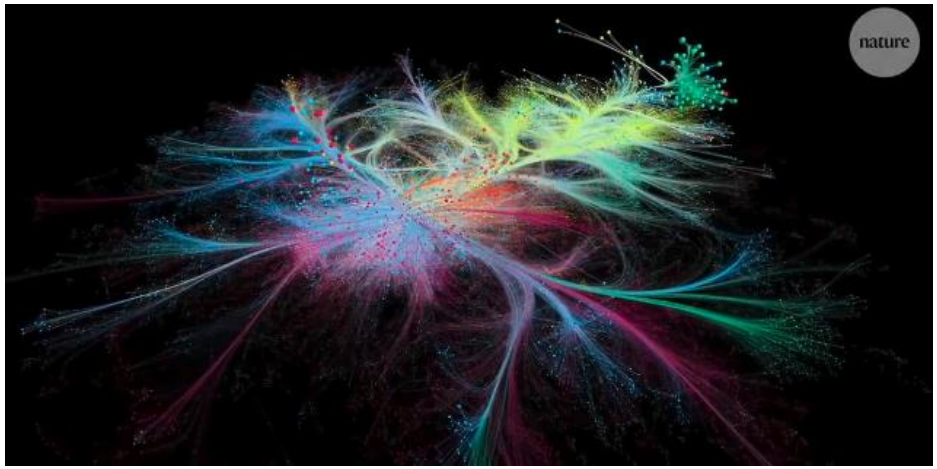


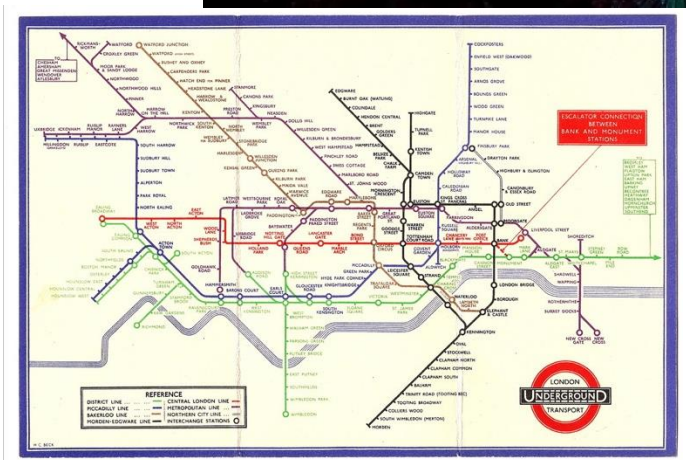


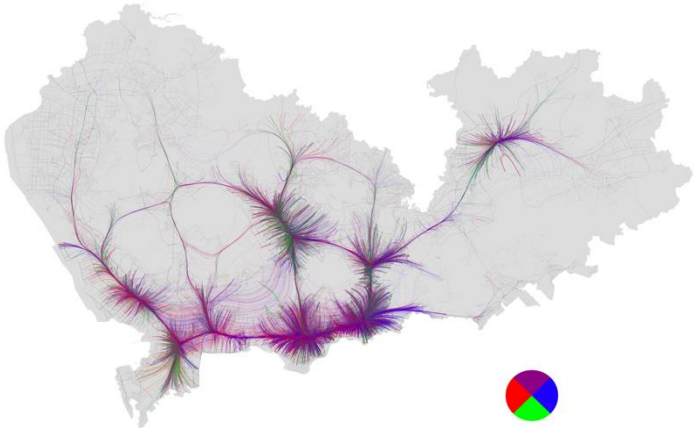
Credit: Nadieh Bremer https://whydocatsanddogs.com/

Examples

- What are the commons and differences among them?







Motivating example:
"How the Swedes Book their Summer Vacations"

- Tripadvisor (circa 2015)
  - "Tell me if there's anything interesting about the way the Swedes book their summer vacations"

Motivating example:
"How the Swedes Book their Summer Vacations"

- Context
  - Tripadvisor just merged with another company. So data resided in multiple (unconnected) databases
  - The sizes of the databases are massive. They don't fit in a typical hard drive of a desktop computer
  - The boss is very interested in the answer to this problem because it represents a business advantage
  - (The student working on this problem quit after 2 weeks of banging his head against this)

- Why is this hard?
- How would you get started?

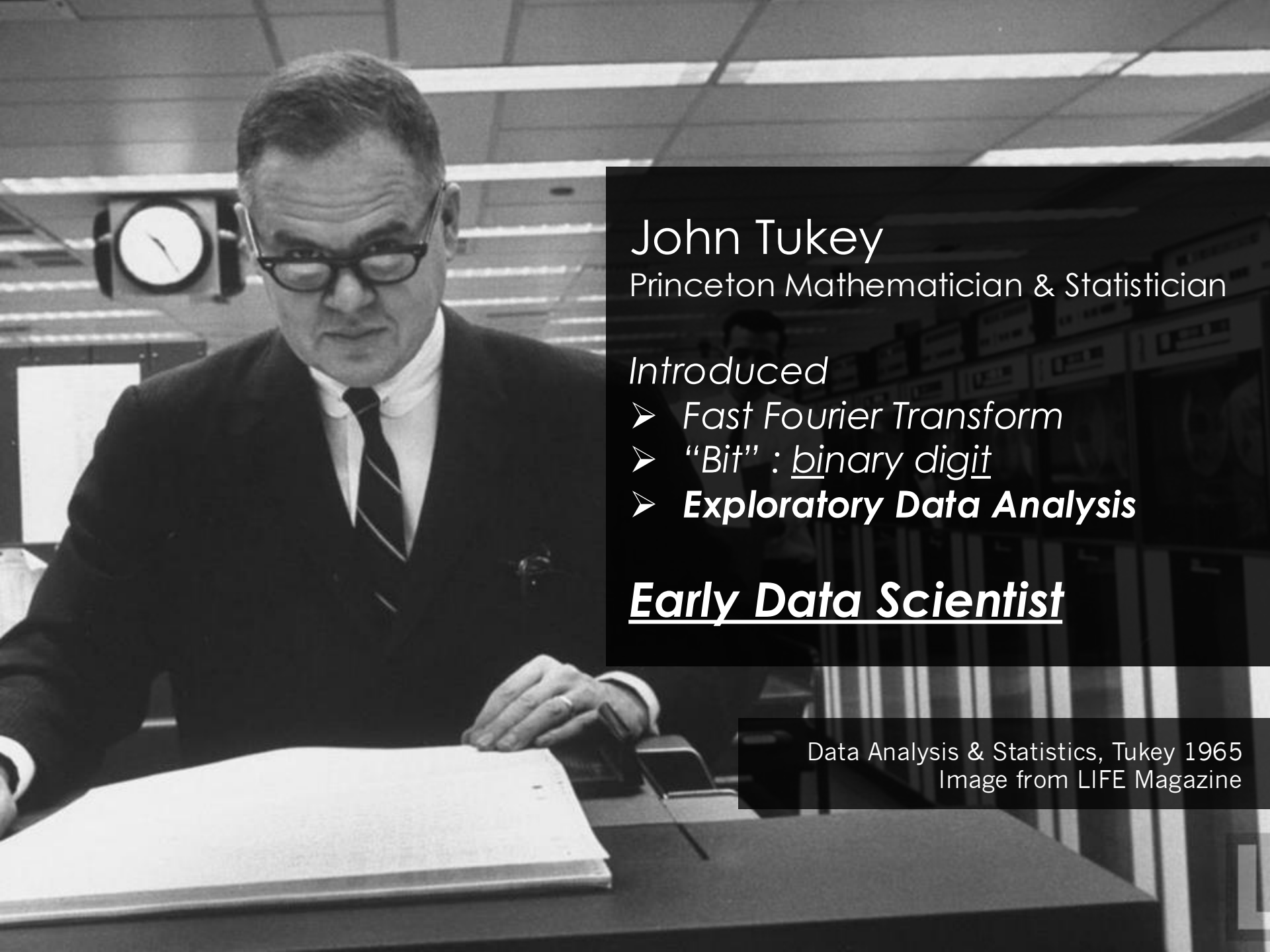
Motivating example:
"How the Swedes Book their Summer Vacations"

- Why is this hard?
  - Problem is ill-defined. What does "interesting" mean?
    - Is "interesting" how Swedes book summer versus winter vacations?
    - Or "interesting" when compared to other Europeans (or or others around the world) in their summer vacations?
  - Data is too disparate
    - How do I extract the right data from multiple databases when I don't know what the question is?
  - Data is too large
    - Can't fit the whole database into (Excel | R | memory | hard drive), so how do I find a representative sample?
  - No idea what's in the data or what it looks like
    - Unclear what kinds of statistics (or machine learning) is needed
    - Don't know how to visualize this data

Reflecting on the example

- "Wicked problem": problem that is impossible to solve because of incomplete, contradictory, and changing requirements that are often difficult to recognize. (source: wiki)

- Computer scientists hate these types of problems…
  - but in a way, this is how data science became its own field,
  - and our focus is on data visualization as an exploratory tool.



John Tukey

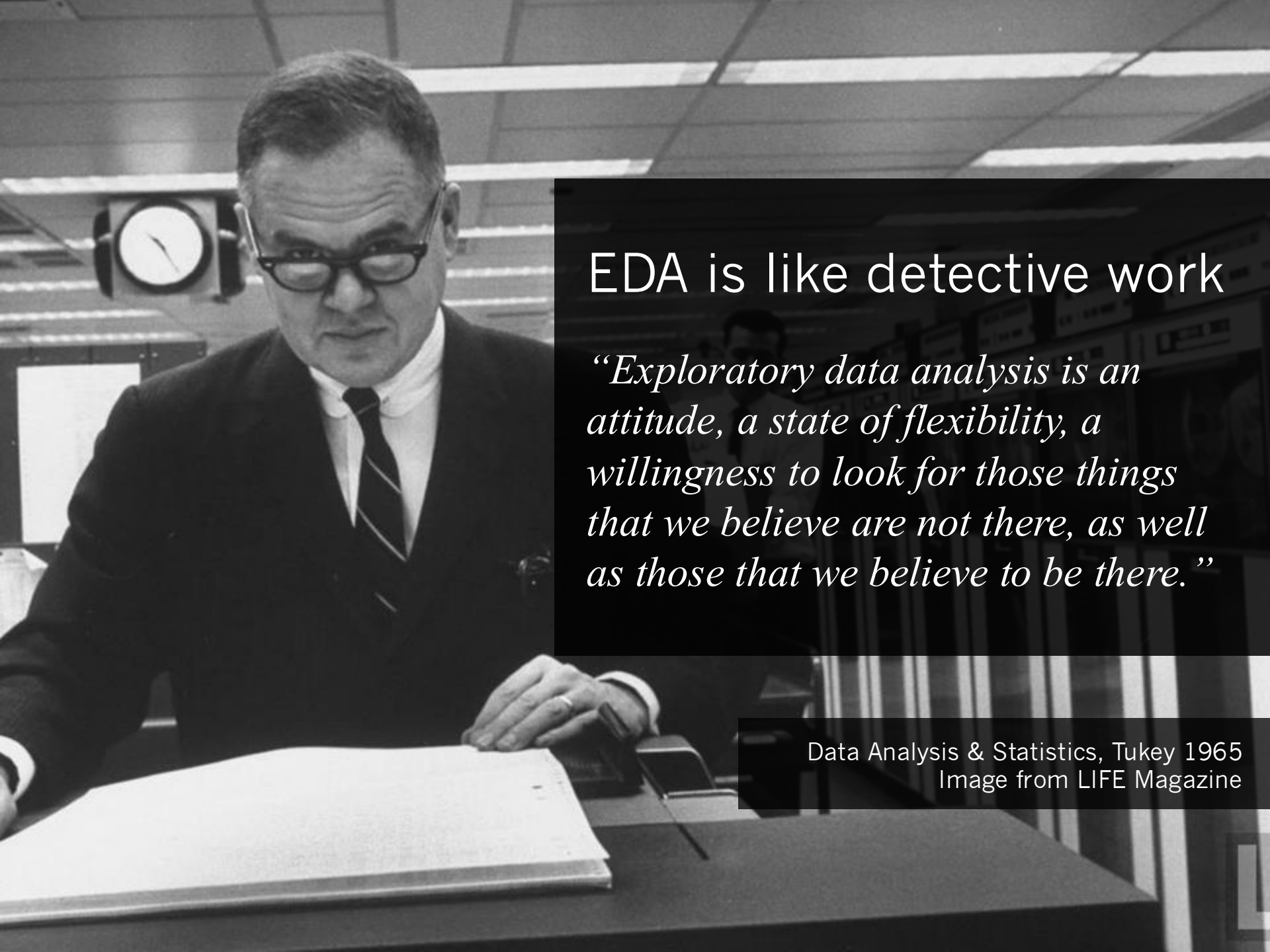
Princeton Mathematician & Statistician

Introduced

- *Fast Fourier Transform*
- *"Bit" : binary digit*
- ***Exploratory Data Analysis***

Early Data Scientist

Data Analysis & Statistics, Tukey 1965
Image from LIFE Magazine



EDA is like detective work

"Exploratory data analysis is an attitude, a state of flexibility, a willingness to look for those things that we believe are not there, as well as those that we believe to be there."

Data Analysis & Statistics, Tukey 1965
Image from LIFE Magazine

Course Goals

- **Principles:** Understand visualization!
  - Concepts
    - History, relation to other fields about data visualization
    - Perception & cognition
  - Design & evaluation
    - Visual channels: color, position, shape, etc.
    - Effectiveness and expressiveness
- **Techniques:** Create visualization!
  - Algorithms
    - Data transformation, visual mapping, etc.
  - Techniques
    - Visualization techniques, interaction techniques
  - ~~Mechanics~~
    - ~~Tableau, Javascript, D3, etc.~~

Learn by yourself!

Course Topics

- Introduction: this week

- Part I: Principles (week 2 – 5)
  - Data transformation, visual channels, color vision, effectiveness and expressiveness

- Part II: Techniques (week 6 – 9)
  - Charts, multiple views, interaction techniques

- Part IV: Topics (week 10 – 13)
  - Spatial, movements, time series, graphs

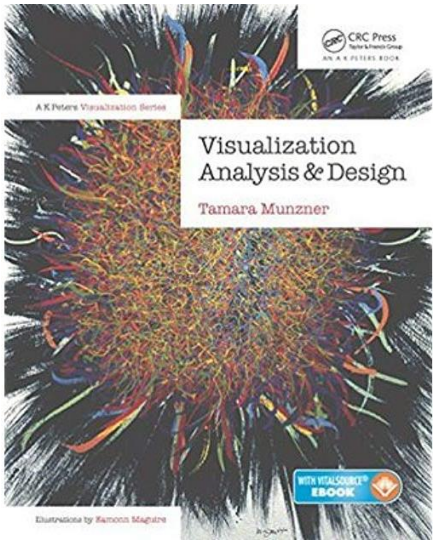
Textbook

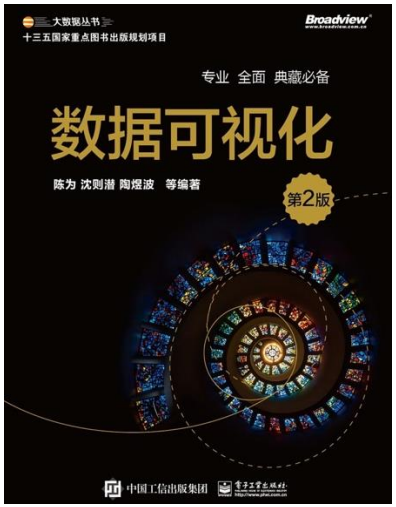
Recommendations

“Visualization Analysis and Design”

1st Edition (2014)

- By Tamara Munzner, A K Peters/CRC Press, ISBN 9781466508910

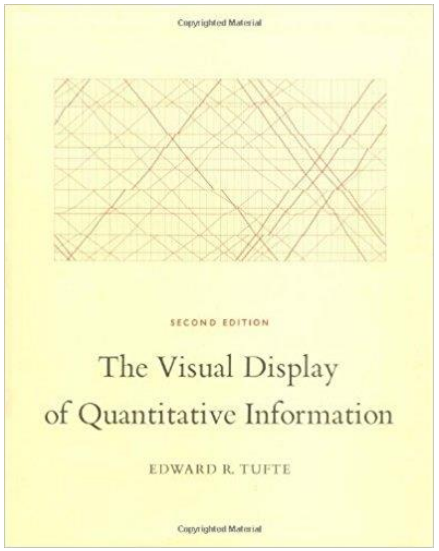


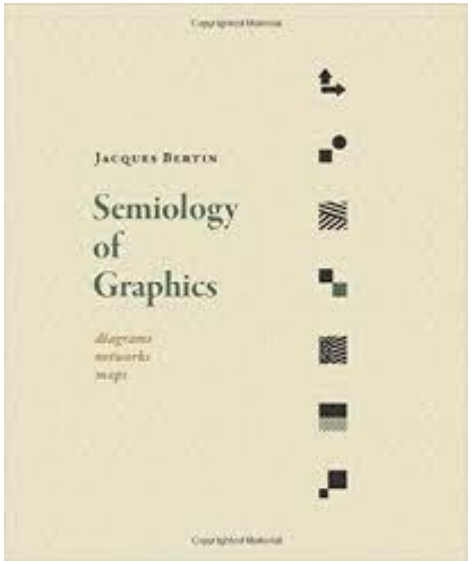


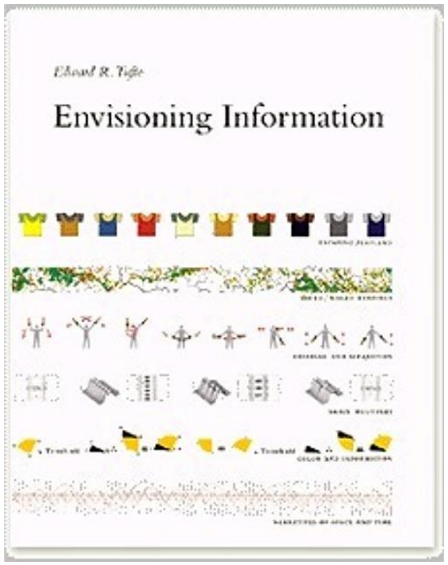
“数据可视化” 2nd Edition (2019)

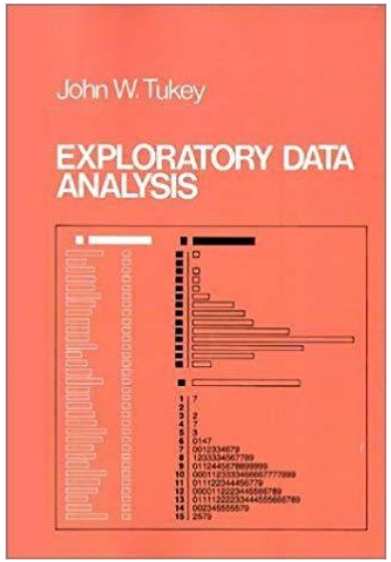
- By 陈为等, 电子工业出版社, ISBN 9787121357275

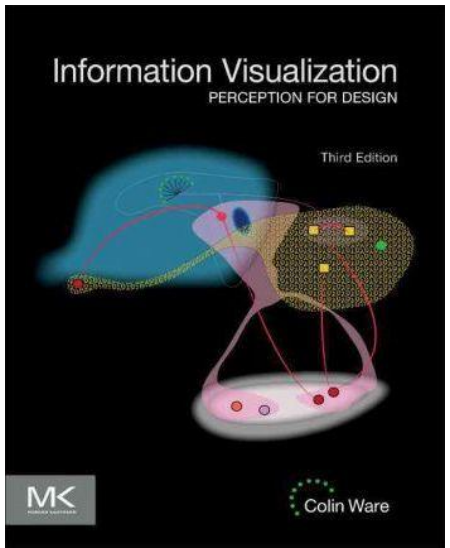
Other reading materials

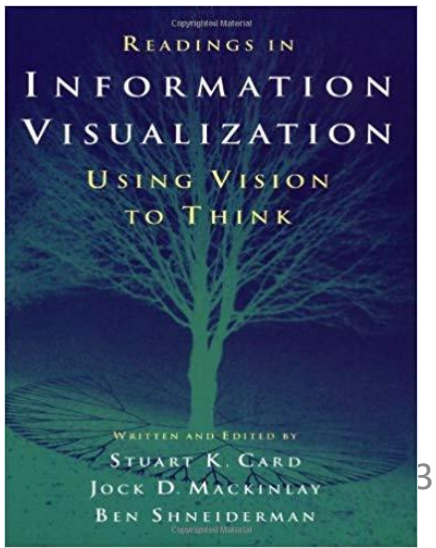












Resources

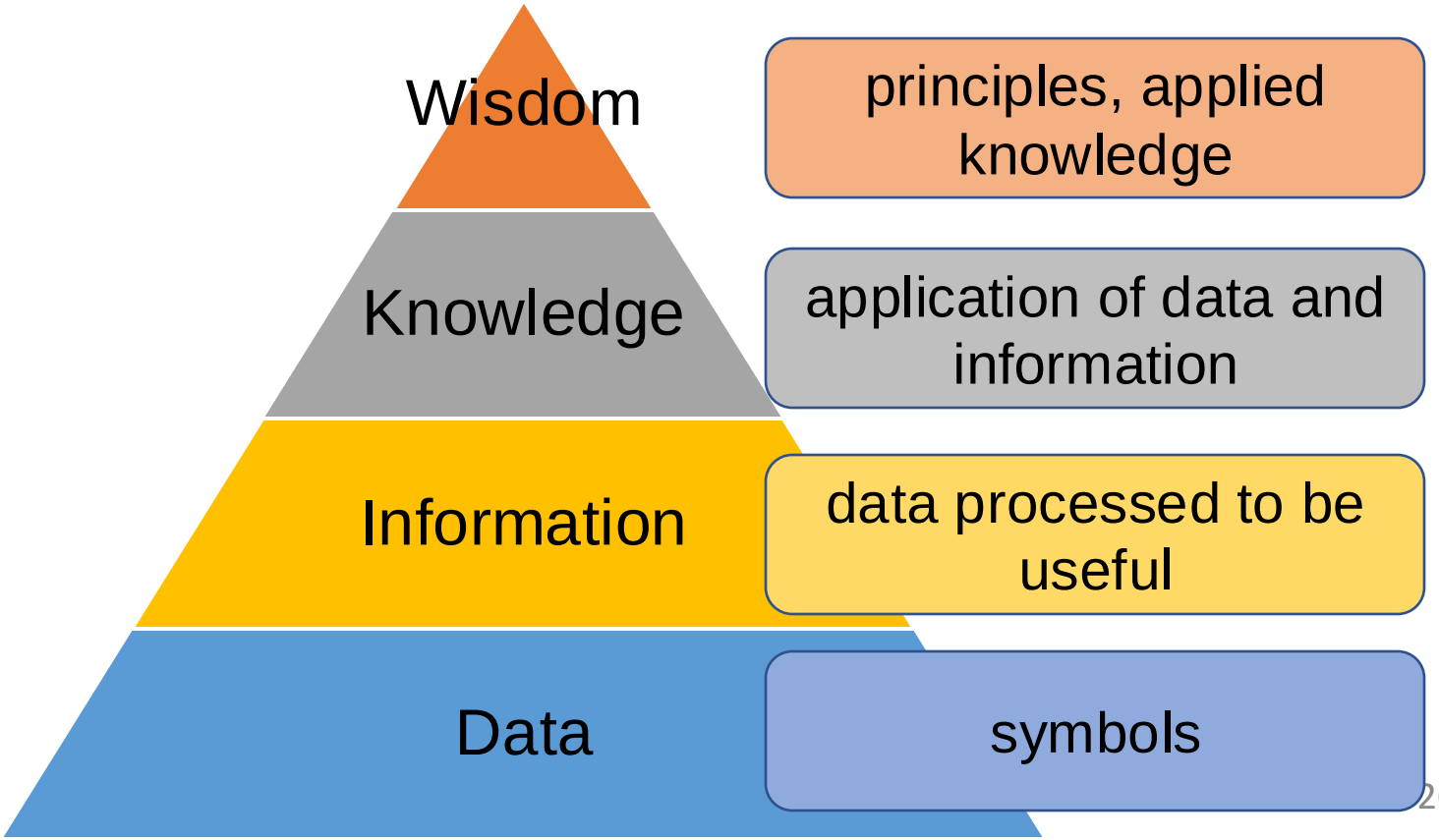
- *Journals & Conferences*
  - *IEEE VIS*, *EuroVis*, *PacificVis*, *ACM CHI*
- *Blogs & webs*
  - *Information is beautiful:* https://informationisbeautiful.net/
  - *Perceptual edge* by Stephen Few https://www.perceptualedge.com/
- *Toolkits:*
  - *Tableau* https://www.tableau.com/
  - *Power BI* https://powerbi.microsoft.com/en-us/
  - *D3* https://d3js.org/, https://bl.ocks.org/mbostock
  - *Echarts* https://www.echartsjs.com/en/index.html
  - *Processing* https://processing.org/

Outline

- **Introduction to data visualization**
  - What: concept and examples
  - Why: using vision to think
- **History of data visualization**
  - Evolution of techniques
  - Impact and utilization
- **Disciplines**
  - Subfields: SciVis, InfoVis, VAST
  - Relations with other fields

DIKW Model

- Data-information-knowledge-wisdom (DIKW) hierarchy (Ackoff, 1989)



Data Visualization

- There are many definitions
  - "The goal of visualization in computing is to gain insight by using visual machinery." - McCormick et al. 1987
  - "... to gain insight into an information space..." - Senay and Ignatius 1990
  - "... maximize human understanding and communication." - Owen 1999
  - "... graphics can be more precise and revealing than conventional statistical computations" - Tufte 2001
  - "... to assist humans in solving problems" - Purchase et al. 2008
  - "... unveiling of the underlying structure..." - Berkeley 2010
- All relate to data/information, purpose, means, and users.

Data Visualization

- The use of computer-supported, interactive, visual representations of data to amplify cognition.

  - Card, Mackinlay & Shneiderman

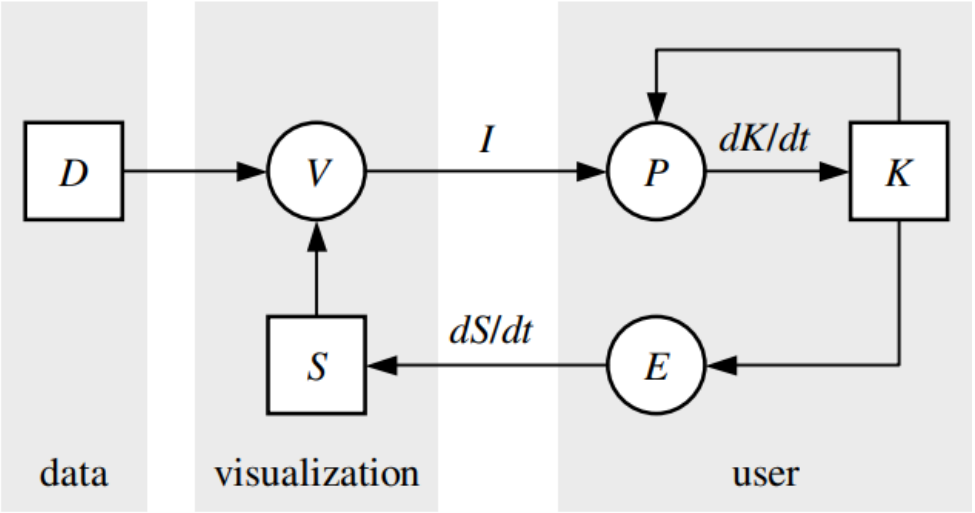
- Computer-based visualization systems provide visual representations of datasets intended to help people carry out some tasks more effectively.

  - Tamara Munzer

  - **Dataset:** data, data processing techniques
  - **Visual representations:** visual channels, color vision, design, data visualization techniques
  - **Interaction:** interaction techniques
  - **Task:** tasks and validation

Data Visualization

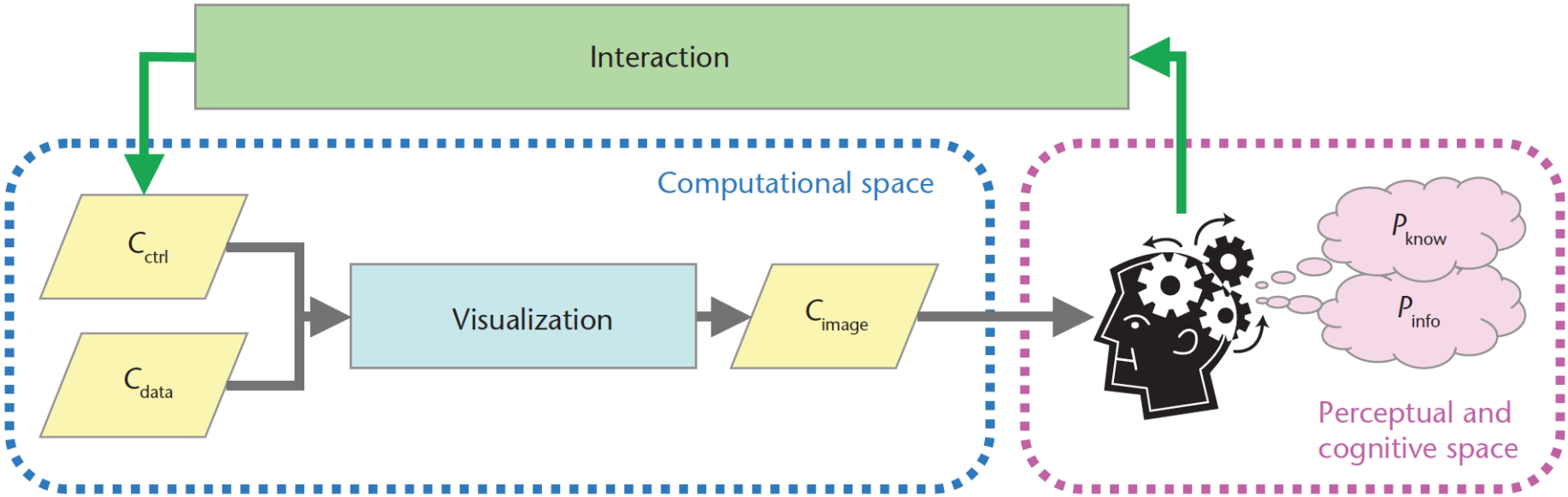
- *"Value of Visualization*," van Wijk, 2005
  - "A simple model of visualization"



- Squares are objects
  - D = data, S = specification, K = knowledge, I = image
- Circles are actions
  - V = visualize, P = perceive, E = explore

Data Visualization

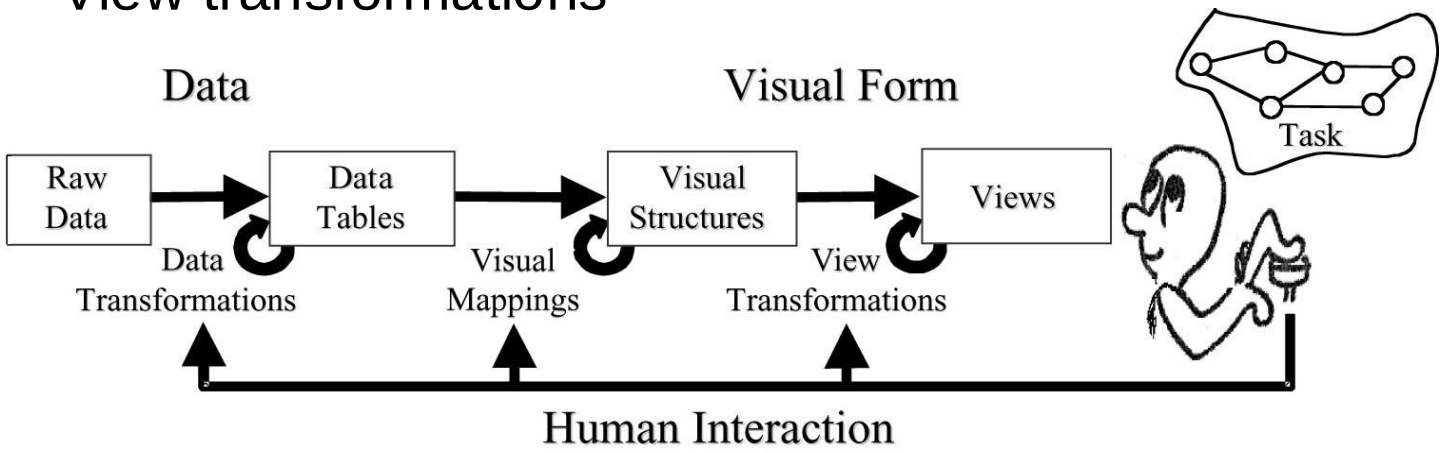
- A typical visualization process maps from *data* and *control parameters* to *images*.
  - Data: symbols
  - Control parameters: viewpoint, data selections, etc.
  - Images: visual representations



Chen et al., 2009

Data Visualization

- Information visualization reference model
  - Data transformations
  - Visual mappings
  - View transformations

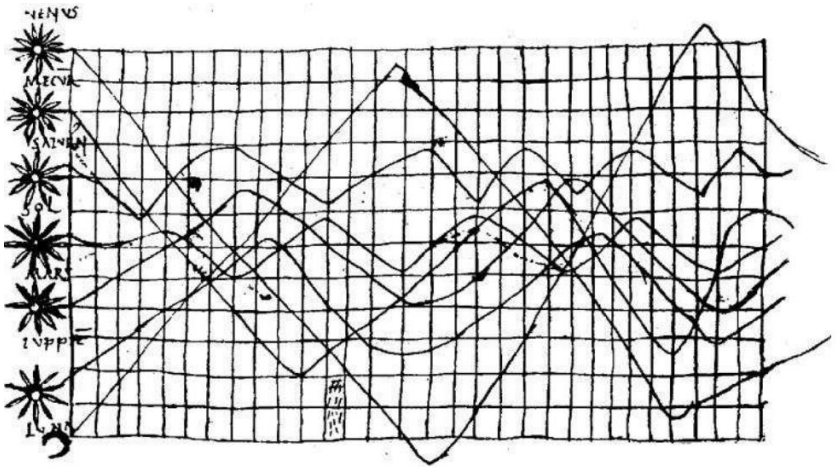


Raw Data: idiosyncratic formats
Data Tables: relations (cases by variables) + meta-data
Visual Structures: spatial substrates + marks + graphical properties
Views: graphical parameters (position, scaling, clipping, …)

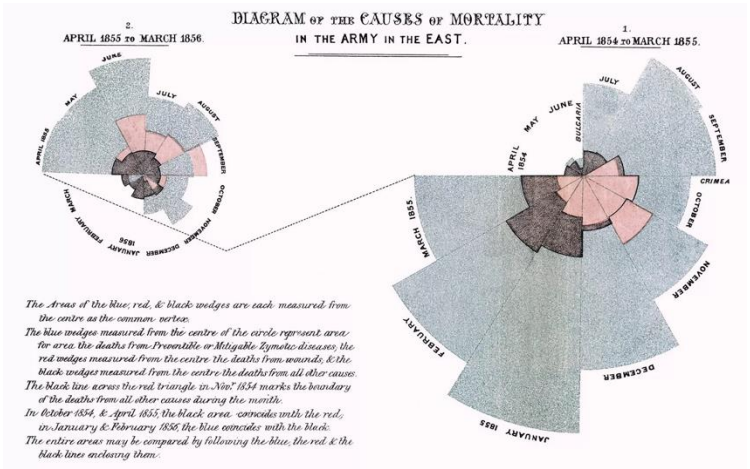
Credit: Stuart, Mackinlay, Shneiderman, 1999

Why Visualization?

- The motivations of visualization:
  - as a tool for ***recording***
    - For example, mapping planetary movements
  - as a tool for ***communication***
    - For example, Nightingale's rose chart
  - as a means for ***analysis***
    - To discover patterns in data



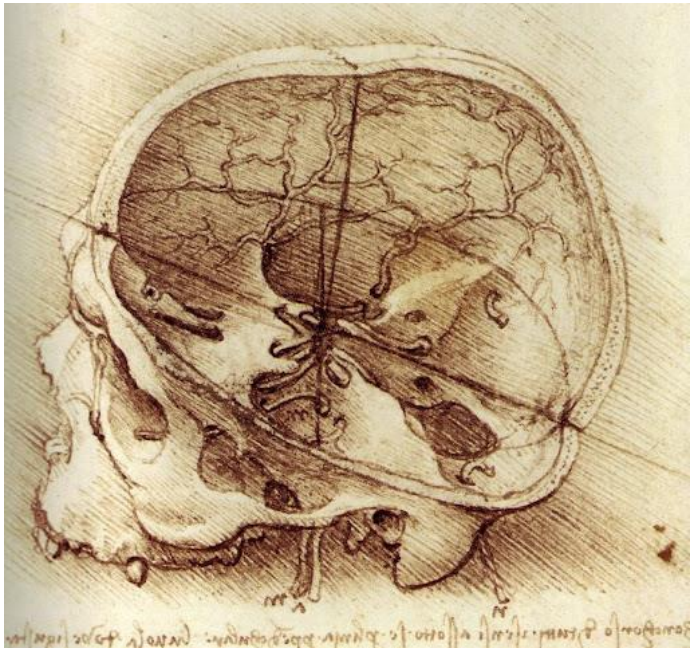
Planetary Movement, 10th century, unknown astronomer

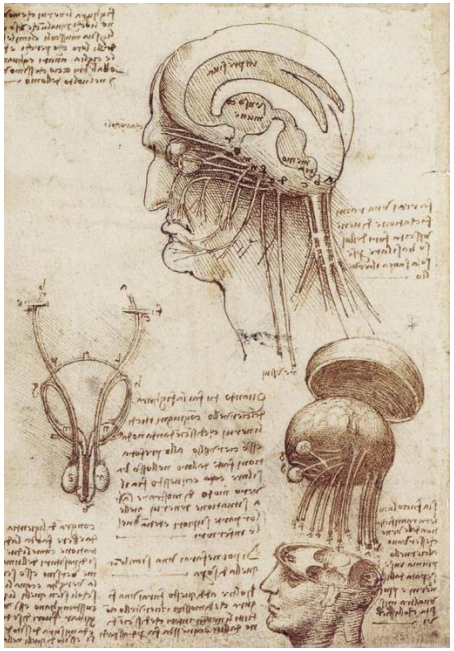


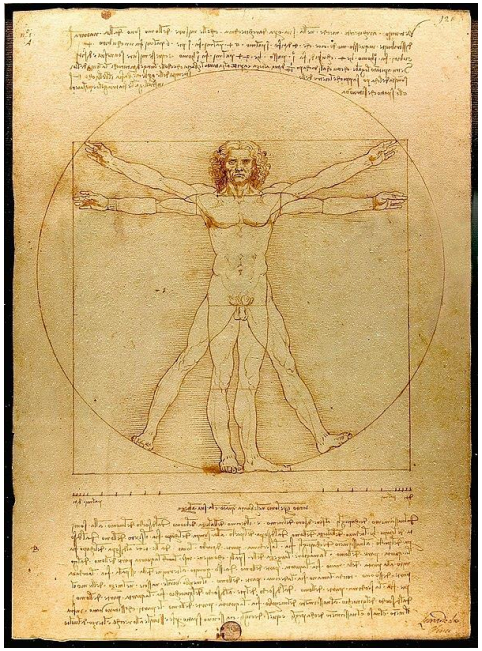
Nightingale's Rose: "Diagram of Causes of Mortality in the Army of the East" by Florence Nightingale, 1858

Use of Visualization to **Record**

- Leonardo da Vinci (1452 - 1519)
  - The human brain and skull
  - Vitruvian Man
  - Da Vinci called his visualization process "sapere vedere", translated as "knowing how to see”.

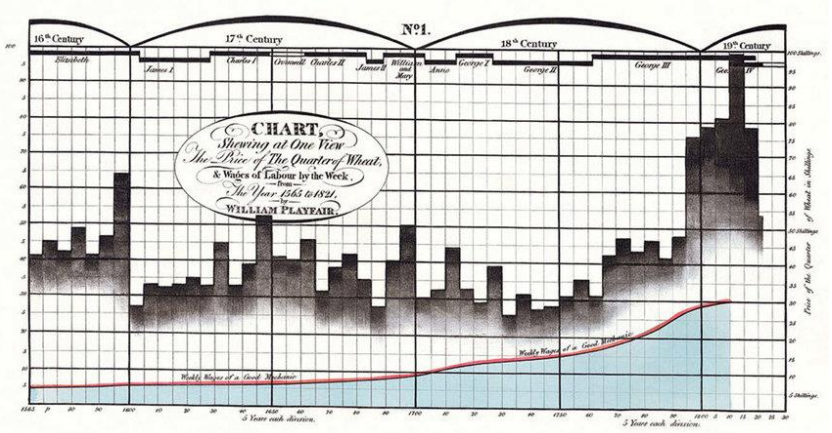




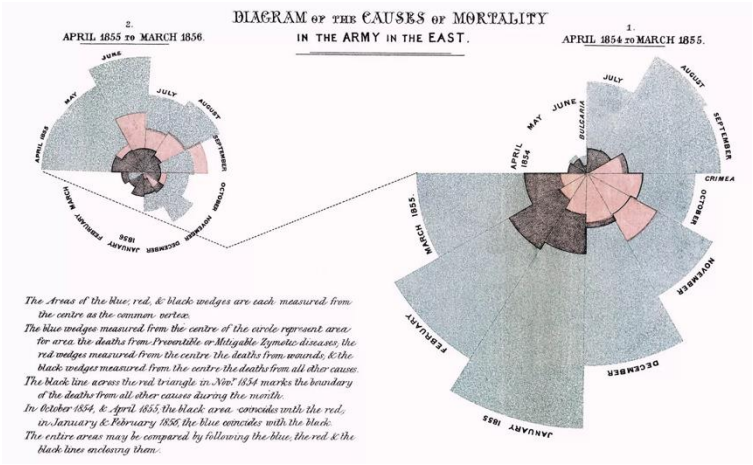


Use of Visualization to **Communicate**

- The example by Playfair (1821) was used to make the argument that workers life have improved over time



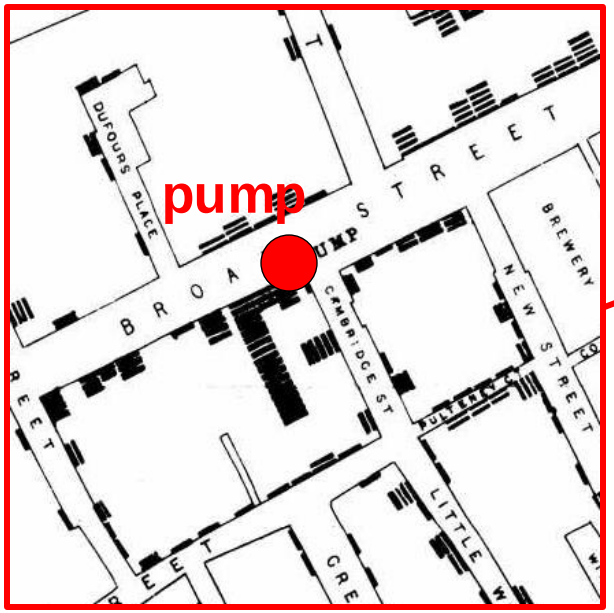
- The Nightingale's Rose (1858) was used to persuade the Secretary of State for War to improve hospital conditions

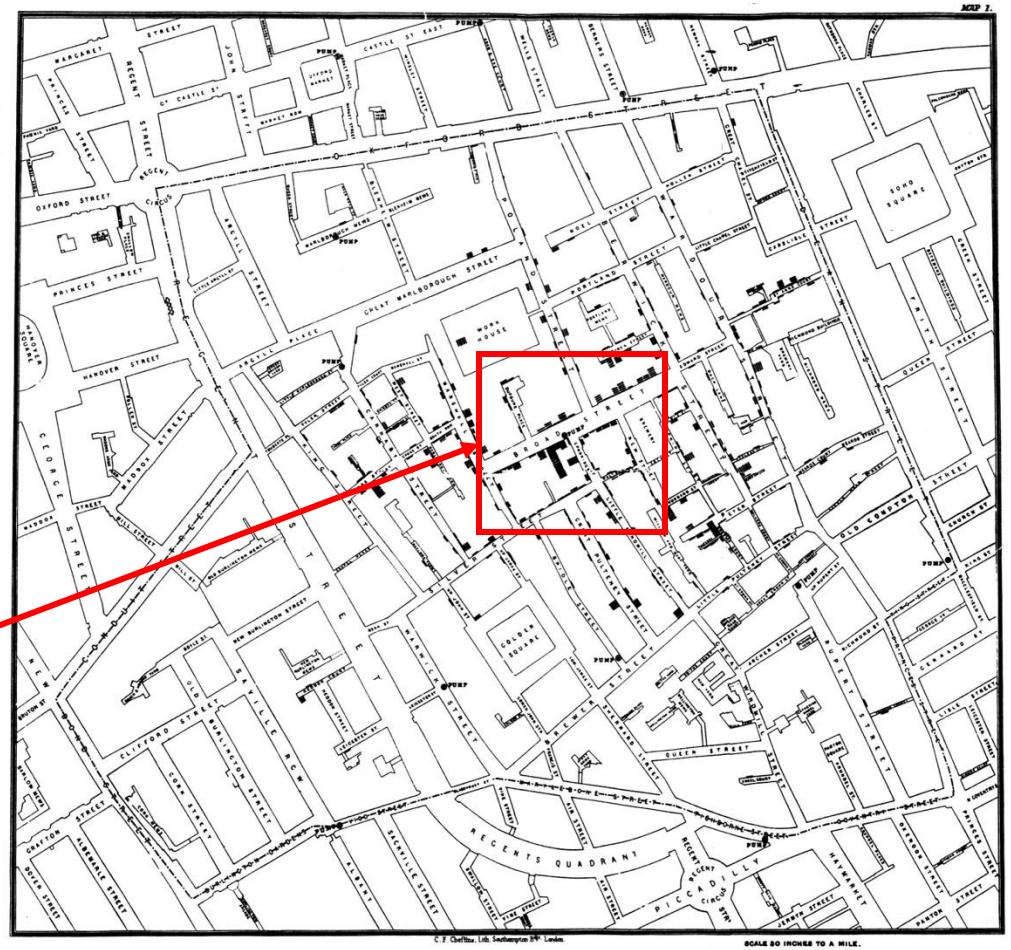


Use of Visualization to **Analysis**

generate hypothesis, evaluate hypothesis

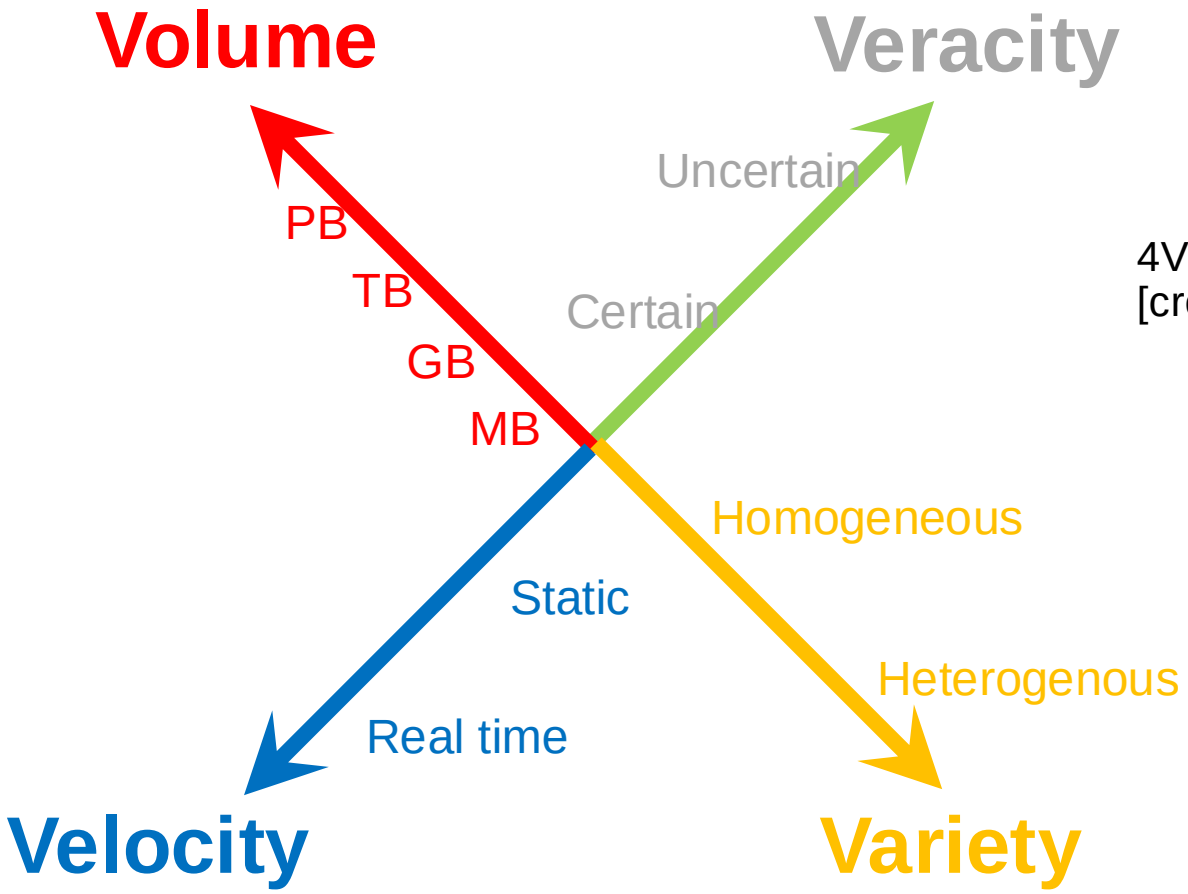
- Map of the 1854 London cholera outbreak, John Snow, 1854





The Age of Big Data

- Big data ⇒ smart data ⇒ (visual) data science



4Vs of Big Data
[credit: Gartner Group]

Data visualization is a key

NEWS

Analysts: Data visualization tools key to 'big data' analytics success

Mark Brunelli, Senior News Editor
Published: 30 Nov 2011

Demand for data visualization tools is rising sharply, partly as a result of more companies seeking to gain valuable business insights through "big data" analytics initiatives. But achieving success with data visualization often requires fresh thinking about how to present information to business users, especially in big-data environments, according to data management analysts.

Data

BY CHRIS TAYLOR, TIBCO 08.26.13 12:53 PM

Using vision to think

- Visualization is really about external cognition, that is, how resources outside the mind can be used to boost the cognitive capabilities of the mind.

  - Sturt Card

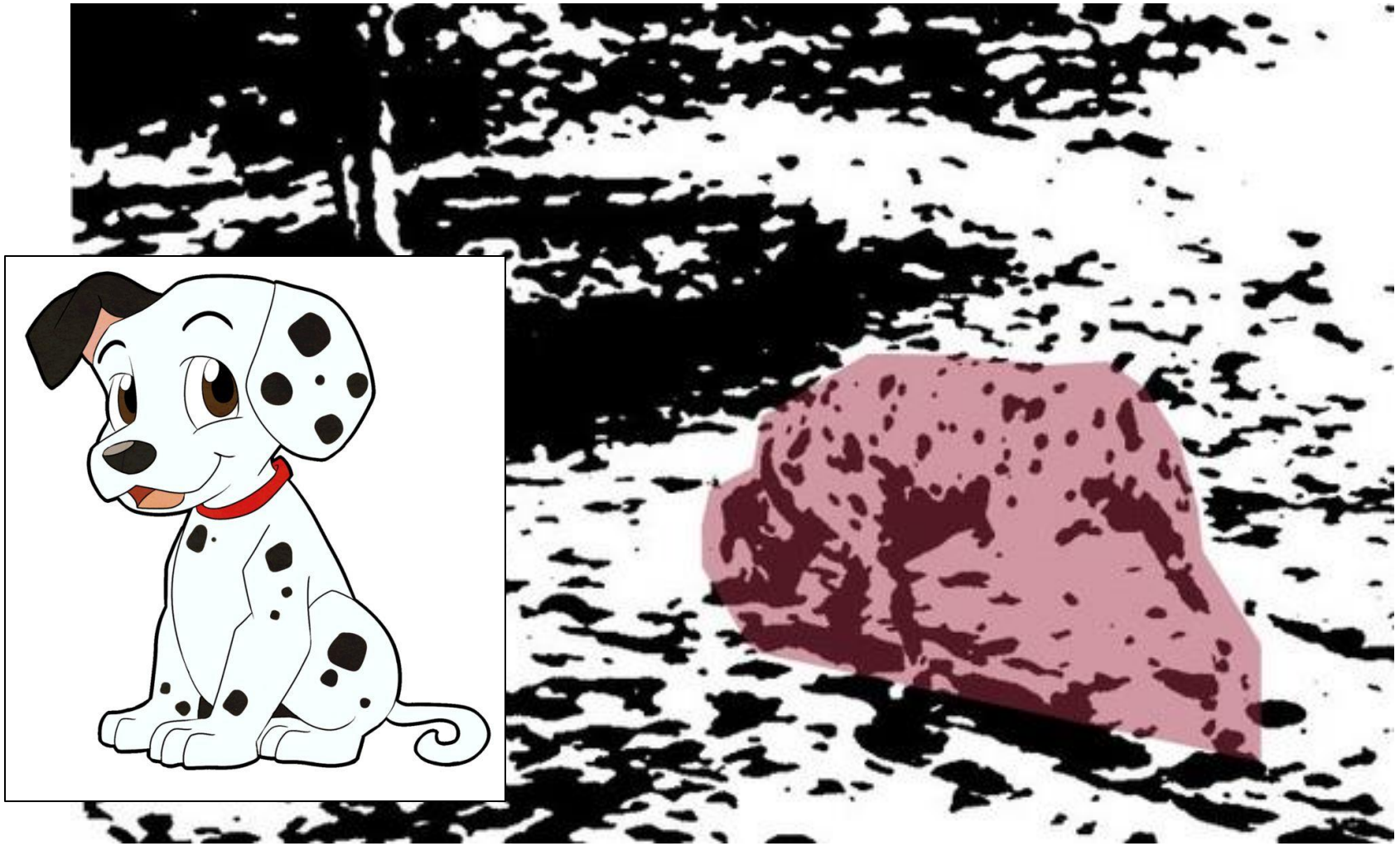
- What you see when you see a thing depends upon what the thing you see is.

  But what you see the thing as depends upon what you know about what you are seeing…

  - Visual Perception: Physiology, Psychology and Ecology

Dalmatian illusion

Dalmatian illusion



Anscombe's quartet

- Mean: $u_x = {(x_1 + x_2 + \cdots + x_n)}/{n}$
- Sample variance: $S_x{}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - u_x)^2$
- Correlation: $\boldsymbol{\rho}_{x,y} = \frac{\sum_{i=1}^{n} (x_i - u_x)(y_i - u_y)}{(n-1) S_x S_y}$

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

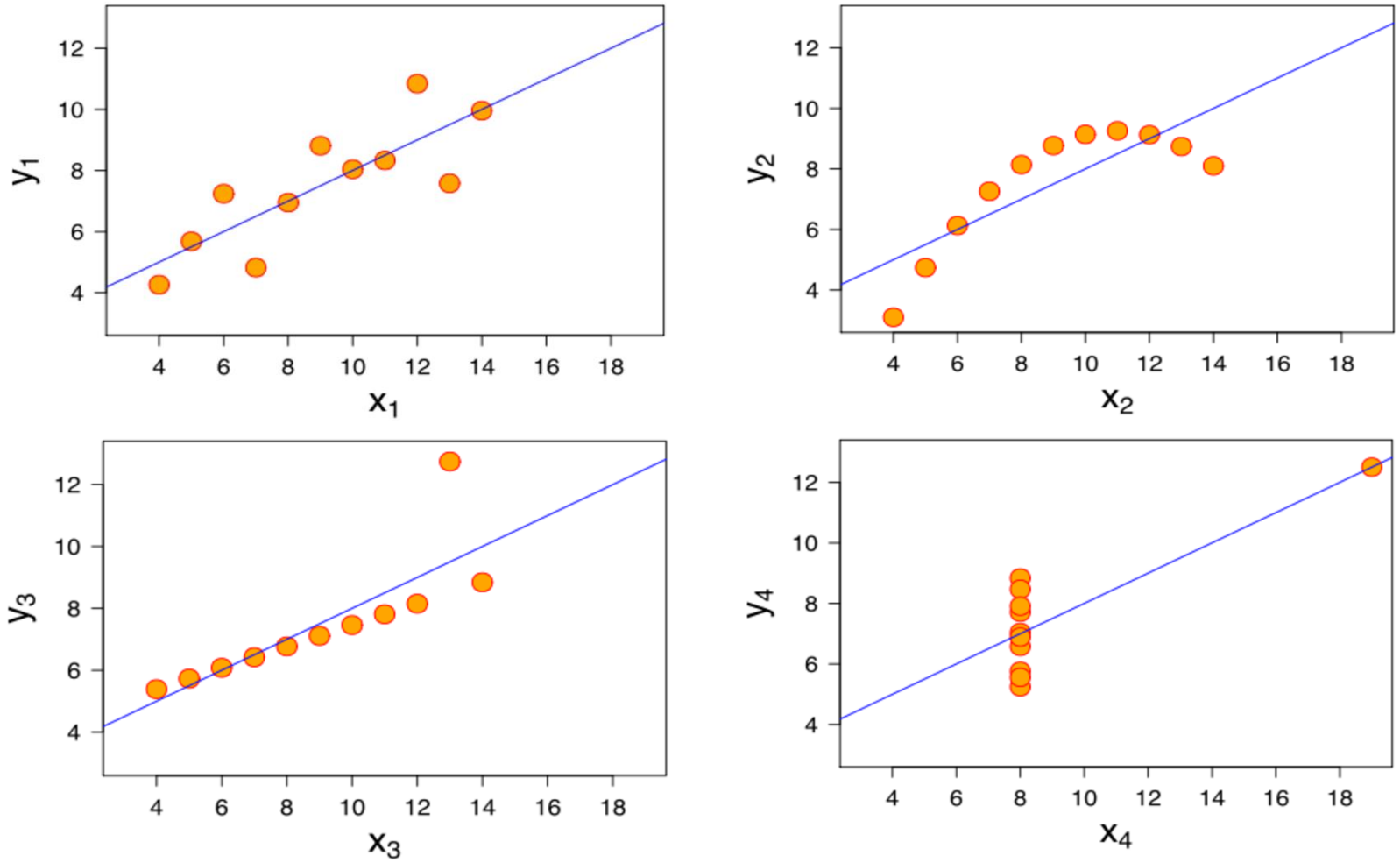
Anscombe's quartet

- Mean: $u_x = 9, u_y = 7.5$
- Sample variance: ${S_x}^2 = 11$, ${S_y}^2 = 4.125$
- Correlation: $\boldsymbol{\rho}_{x,y} = 0.816$

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

Anscombe's quartet

- The differences are significant when plotted



Traffic data

- Shenzhen taxi data

Update regularly every 20/30/60 seconds

```
粤BM43V0,114.099365,22.543400,2016-01-01 00:05:10,1511413,0,45,0,,,1,蓝色,
粤BD9P55,114.040932,22.714832,2016-01-01 00:05:12,1157858,1,90,0,,,0,蓝色,
粤B2HU45,113.874550,22.670233,2016-01-01 00:05:09,1376477,62,80,0,,,0,蓝色,
粤BN43F1,114.137535,22.551950,2016-01-01 00:05:11,1511529,44,180,0,,,0,蓝色,
粤BL3C37,113.997200,22.555967,2016-01-01 00:04:56,1228953,69,276,0,,,1,蓝色,
粤B152YS,114.040001,22.651699,2016-01-01 00:05:10,1454811,30,256,0,,,0,蓝色,
粤B263YS,114.037300,22.705099,2016-01-01 00:05:11,1454778,2,0,0,,,0,蓝色,
粤B6HR48,114.107880,22.611349,2016-01-01 00:05:11,1372113,0,126,0,,,1,蓝色,
粤BL6C35,114.145485,22.716700,2016-01-01 00:04:09,1228931,0,57,1,,,0,蓝色,
粤B5WR39,113.945747,22.523899,2016-01-01 00:04:26,1433675,15,90,0,,,1,蓝色,
粤B5WR39,113.947365,22.523733,2016-01-01 00:04:41,1433675,28,90,0,,,1,蓝色,
粤B4K1S2,113.928146,22.492018,2016-01-01 00:05:12,1608266,0,135,1,,,0,蓝色,
粤BQ74Q5,114.173218,22.603050,2016-01-01 00:05:12,1519417,11,90,0,,,0,蓝色,
粤SQZ583,113.734352,23.019917,2016-01-01 00:05:06,1198307,51,282,0,,,0,,
粤BF7644,113.814880,22.610283,2016-01-01 00:05:11,1467026,70,90,0,,,1,蓝色,
粤BR5127,113.887779,22.561642,2016-01-01 00:04:05,1344821,0,0,0,,,0,黄色,
粤B4V1Q2,114.074364,22.531767,2016-01-01 00:05:06,1571096,75,45,0,,,1,蓝色,
粤SQS507,113.934853,23.091600,2016-01-01 00:05:11,1197523,55,152,0,,,0,,
粤SQM455,113.819901,22.818916,2016-01-01 00:05:11,1197425,69,190,0,,,0,,
粤SYP417,114.163002,22.840567,2016-01-01 00:05:12,1294012,0,0,0,,,0,,
```

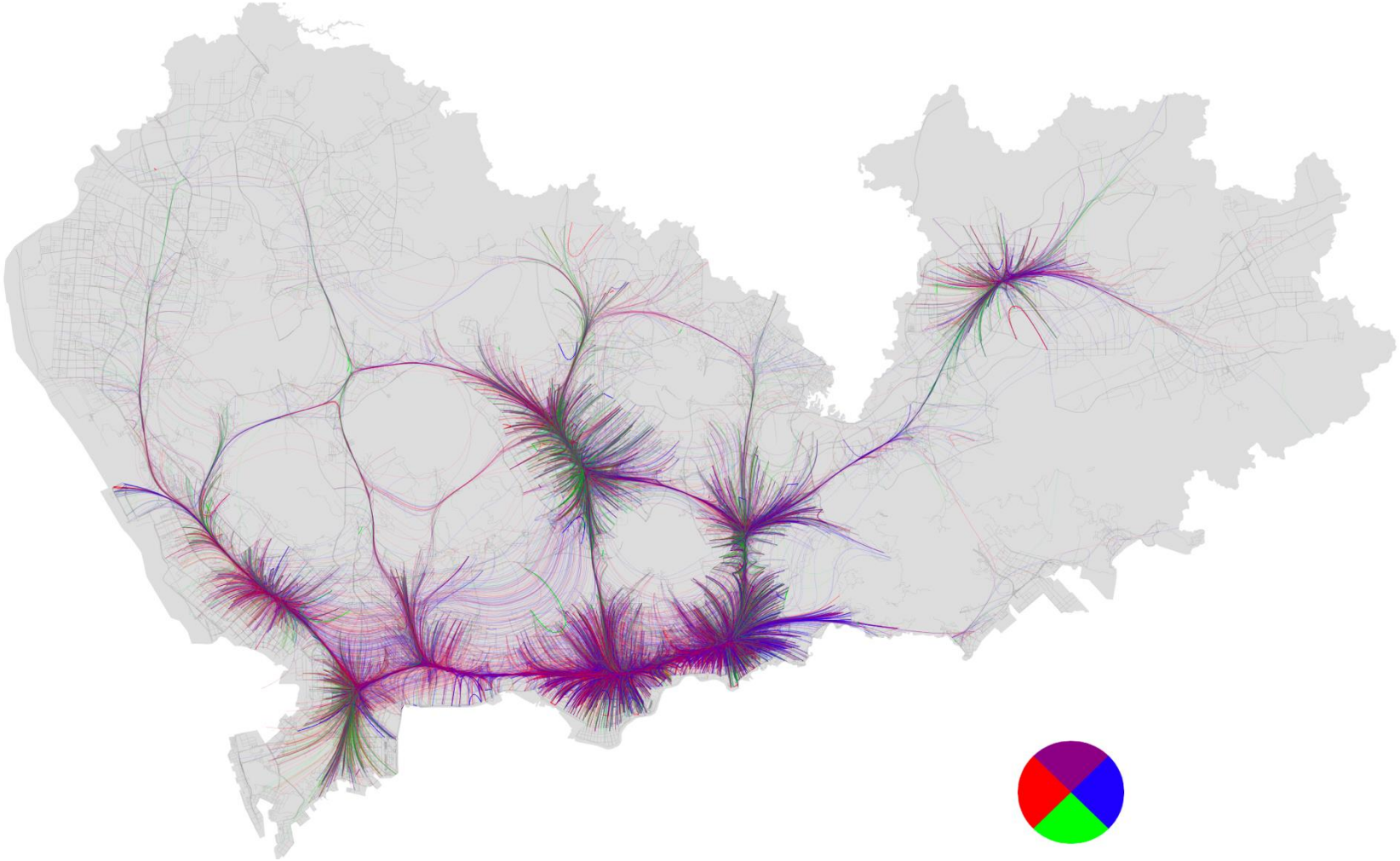
Taxi ID

Position (long, lat)

Time

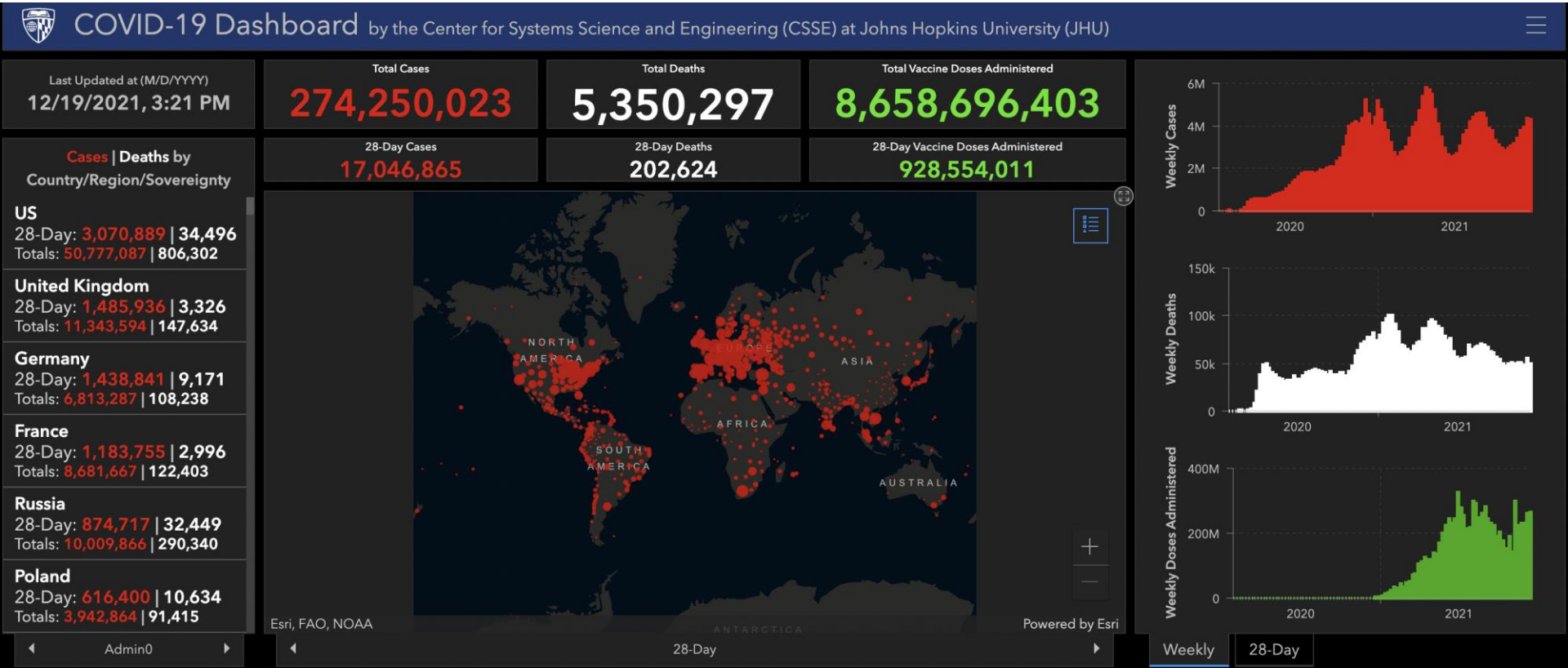
Traffic data

- Shenzhen taxi data



COVID-19 data

```
FIPS,Admin2,Province_State,Country_Region,Last_Update,Lat,Long_,Confirmed,Deaths,Recovered,Active,Combined_Key,Incident_Rate,Case_Fatality_Ratio
,,,Afghanistan,2021-01-01 05:23:07,33.93911,67.709953,52330,2189,41727,8414,Afghanistan,134.42648377951684,4.183068985285687
,,,Albania,2021-01-01 05:23:07,41.1533,20.1683,58316,1181,33634,23501,Albania,2026.409062478282,2.025173194320598
,,,Algeria,2021-01-01 05:23:07,28.0339,1.6596,99610,2756,67127,29727,Algeria,227.1553723362977,2.766790482883245
,,,Andorra,2021-01-01 05:23:07,42.5063,1.5218,8049,84,7432,533,Andorra,10417.394680644535,1.0436079016026836
,,,Angola,2021-01-01 05:23:07,-11.2027,17.8739,17553,405,11044,6104,Angola,53.407341533270525,2.3072978977952485
,,,Antigua and Barbuda,2021-01-01 05:23:07,17.0608,-61.7964,159,5,148,6,Antigua and Barbuda,162.36418593252185,3.1446540880503147
,,,Argentina,2021-01-01 05:23:07,-38.4161,-63.6167,1625514,43245,1426676,155593,Argentina,3596.605939532802,2.660389267640882
,,,Armenia,2021-01-01 05:23:07,40.0691,45.0382,159409,2823,142801,13785,Armenia,5379.561654597645,1.770916322164997
,,Australian Capital Territory,Australia,2021-01-01 05:23:07,-35.4735,149.0124,118,3,114,1,"Australian Capital Territory, Australia",27.563653352020555,2.542
```

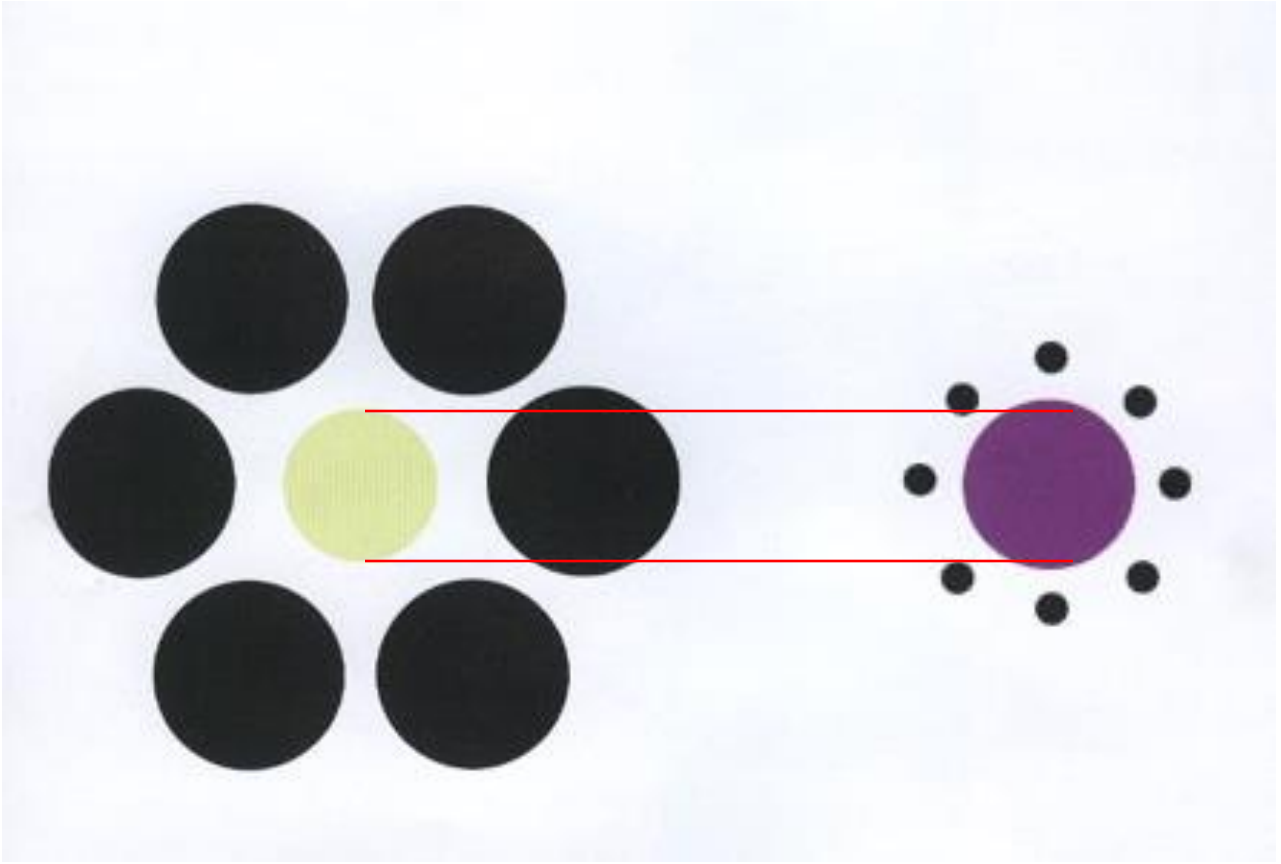


Human perception and cognition are powerful.

Can we solely rely on our vision?

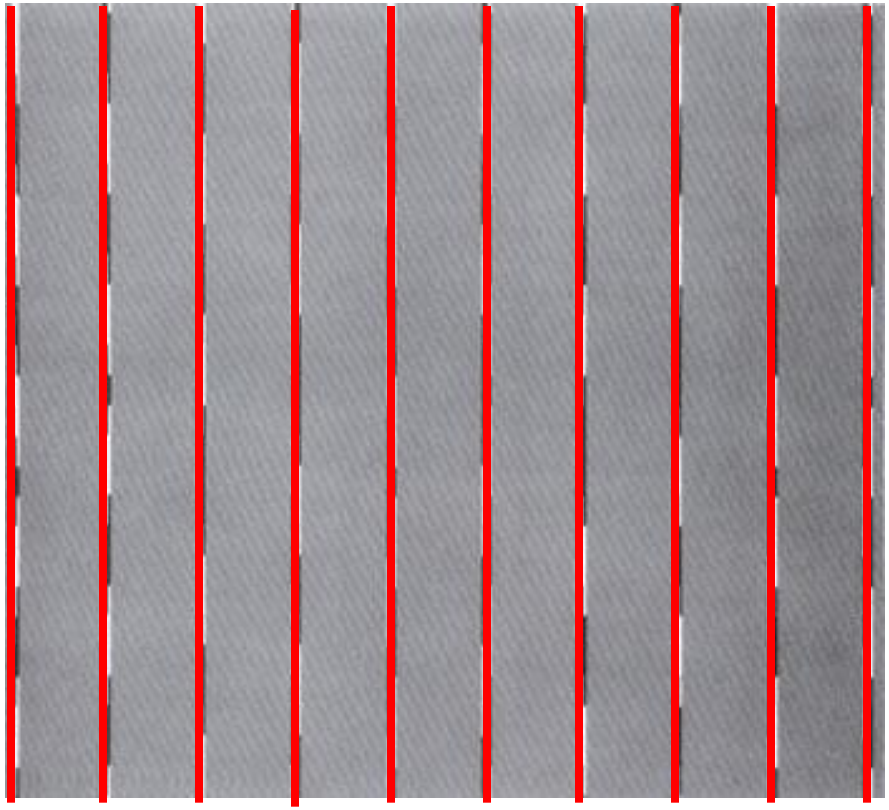
Optical illusion

- Use color, light and patterns to create images that can be deceptive or misleading to our brains.
  - Which circle is bigger?



Optical illusion

- Use color, light and patterns to create images that can be deceptive or misleading to our brains.
  - Do you see parallel lines?



Optical illusion

- Use color, light and patterns to create images that can be deceptive or misleading to our brains.
  - What do you see? A young or old lady?

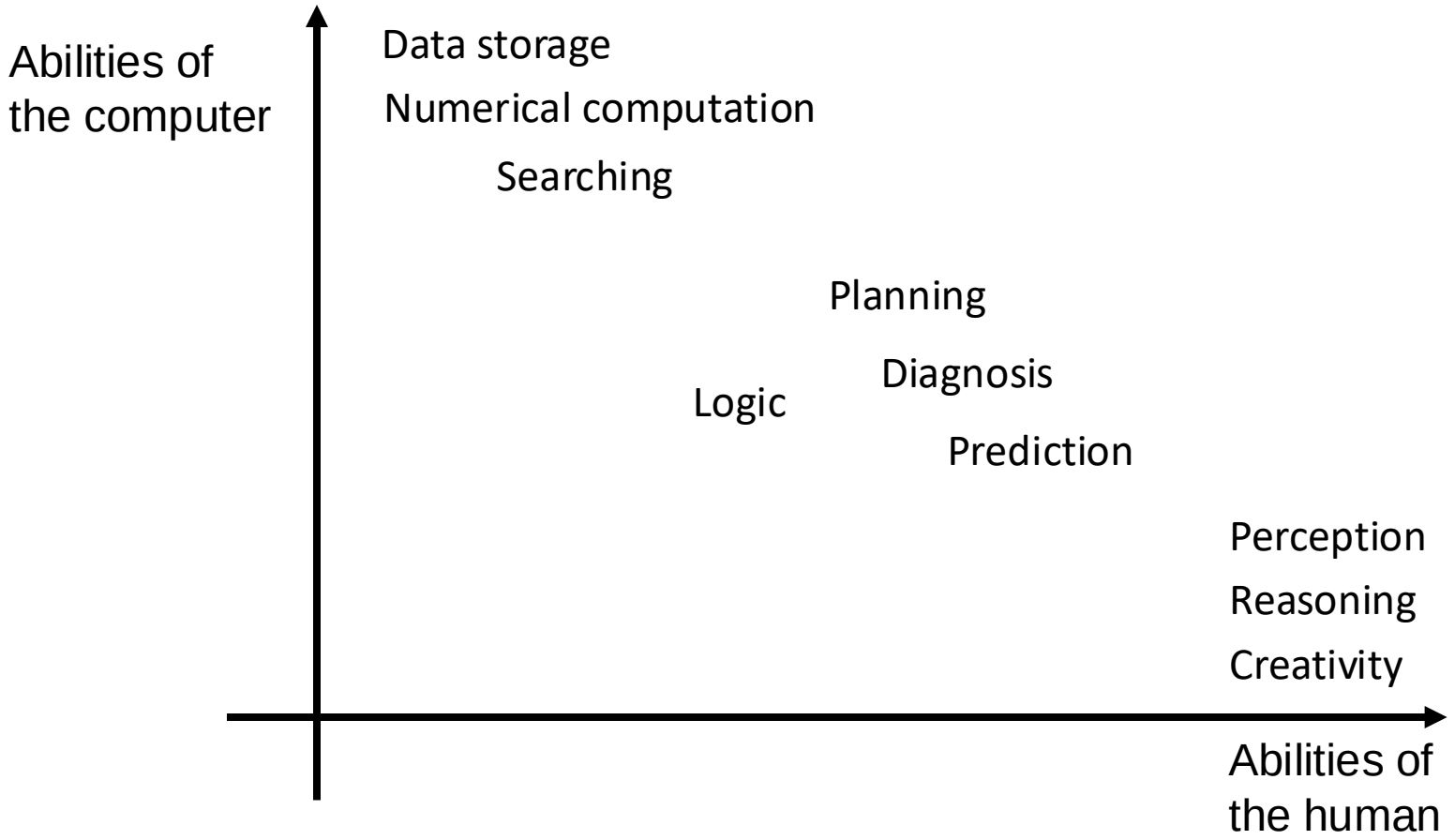
After you know what is there, it is hard to remember how "not" to see it!

https://michaelbach.de/ot/

http://www.ritsumei.ac.jp/~akitaoka/index-e.html

Keep human in the loop

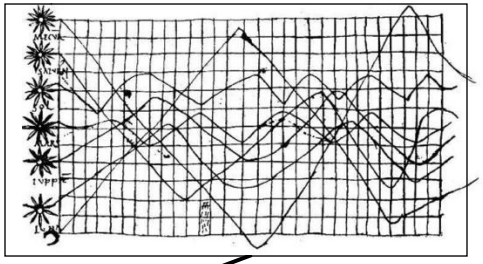
- Let computers do what computers are good at
- Let humans do what humans are good at

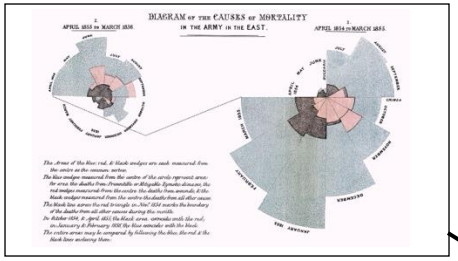


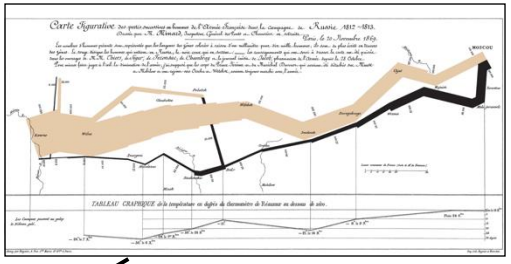
Outline

- **Introduction to data visualization**
  - What: concept and examples
  - Why: using vision to think
- **History of data visualization**
  - Evolution of techniques
  - Impact and utilization
- **Disciplines**
  - Subfields: SciVis, InfoVis, VAST
  - Relations with other fields

Forward

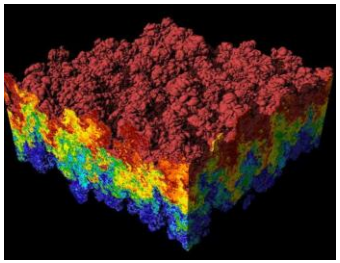




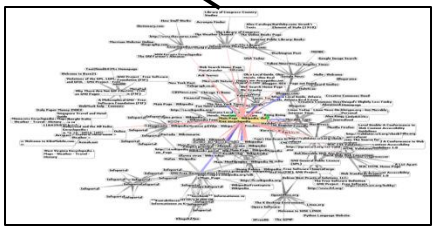


1000 1200 1400 1600 1800 2000

1990 2000 2010 2023



IEEE SciVis

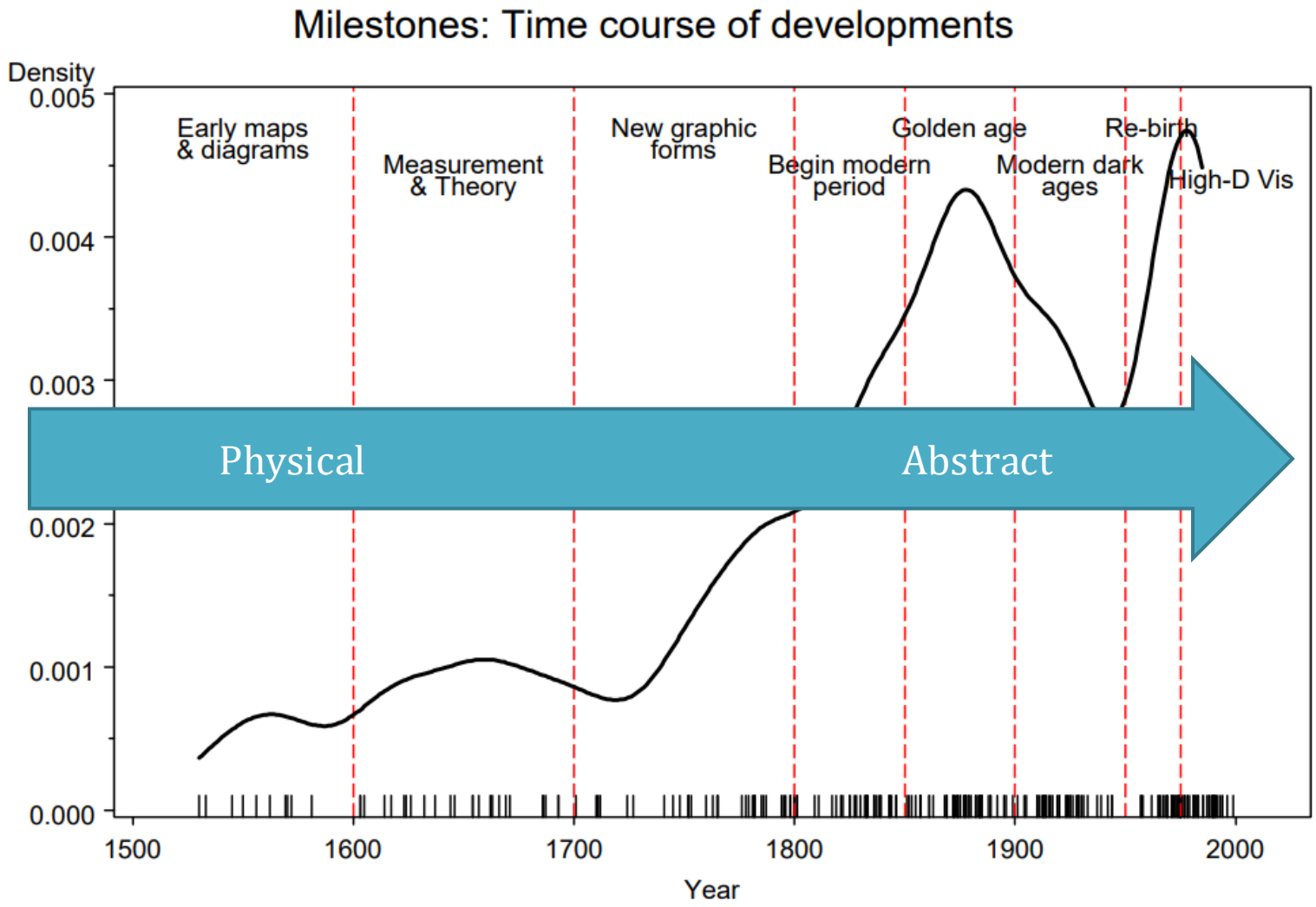


IEEE InfoVis

IEEE VAST

Data Exploration & Visualization

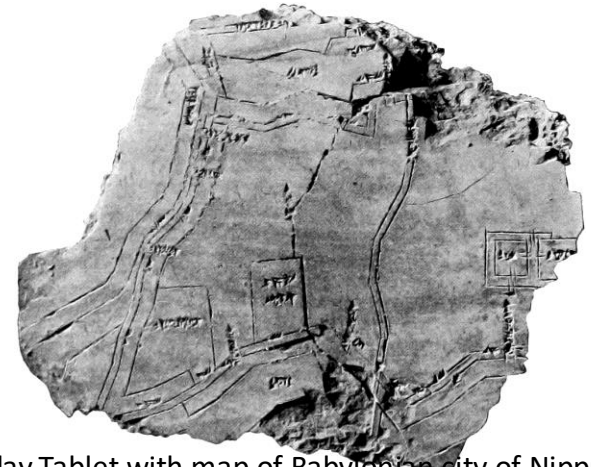
Forward



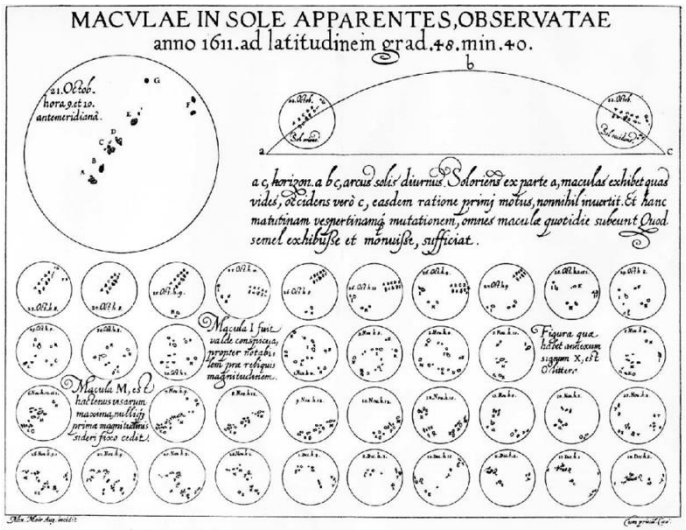
Each tick mark represents a milestone in the history of data visualization (referred to as a rug plot). Friendly, 2006

Forward

- Many of the techniques for “how” to make visualizations remain the same
  - Maps are ***old***
  - Small multiples were used in the 1600s
  - Playfair invented most modern techniques (bars, pies, lines) in the 1700s
  - Scatterplot was introduced in the 1800s
  - Jacques Bertin formalized the theory of data visualization in 1967.
  - High-dimensional visualization techniques were introduced by Tukey in the 1970s



Clay Tablet with map of Babylonian city of Nippur, ca. 1400BC

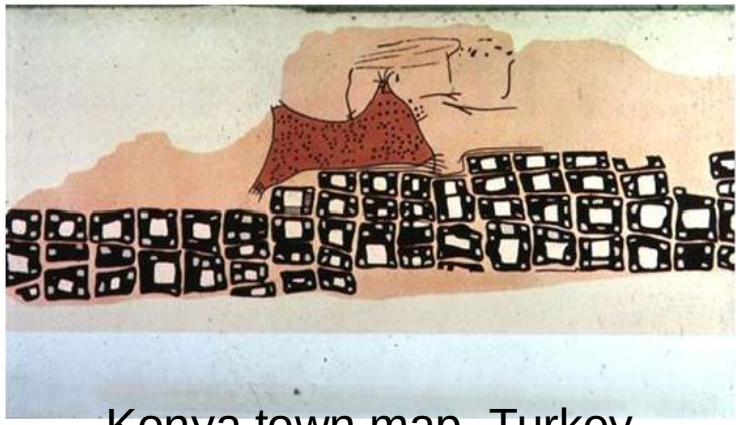


Christopher Scheiner’s visualization of changes in sunspots over time, 1626.

Before 1600: Early Maps and Diagrams
1600-1699: Measurement and Theory

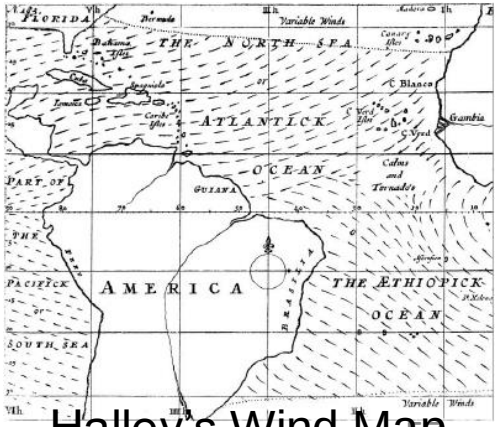
Before 1600 and 1600-1700

- Data visualization was closely tied to cartography – measuring time, distance, and space.
  - "How Maps Work", Alan MacEachren, 2004
- This is always interesting because maps are:
  - A abstraction of a "real" space
  - A visualization of multi-dimensional data
    - latitude, longitude, types (water, land, etc.), boundaries, etc.



Konya town map, Turkey, c.6200 BC

Anaximander of Miletus, c.550 BC



Halley's Wind Map, 1686

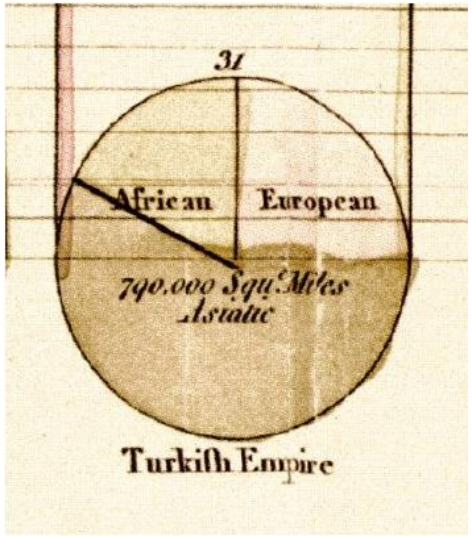
Before 1600 and 1600-1700

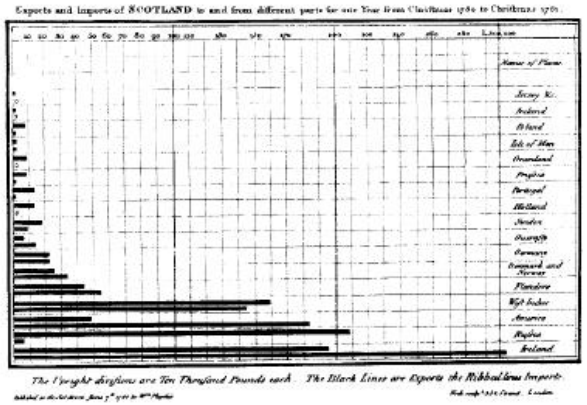
- It helps with one of the most common, and yet cognitively demanding tasks -- seeing patterns in space-time
  - Spatiotemporal hotspots (e.g. disease spread)
  - Trajectory analysis (e.g. car movements)
  - Routing (e.g. planning for bus routes)
- Despite two thousand years of effort, research in maps (and GIS) remains a “hot” topic
- This is a really good example of the utility of visualization
  - Visualization serves as an aid to human cognition
  - For example, humans are bad at remembering animated sequences
    - Maps can help “flatten” these sequences

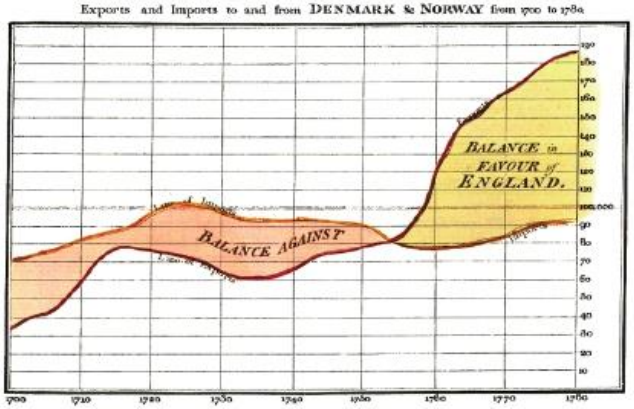
1700-1799: New Graphic Forms

William Playfair (1759-1823)

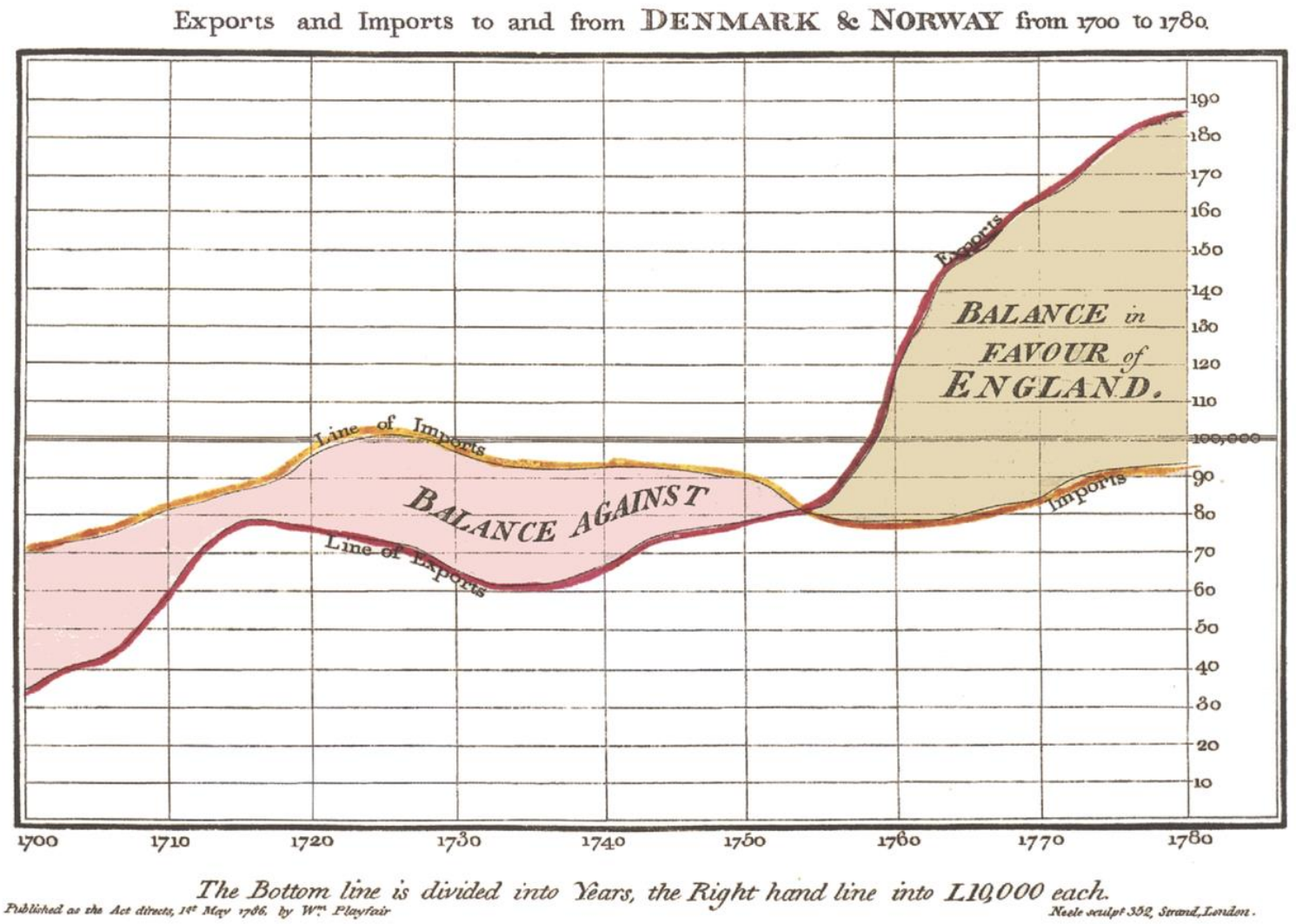
- Considered the inventor of most graphical forms used today:
  - Line graph and Bar chart (1786)
  - Pie chart and Circle Graph (1801)
  - Composite Visualizations (1821)





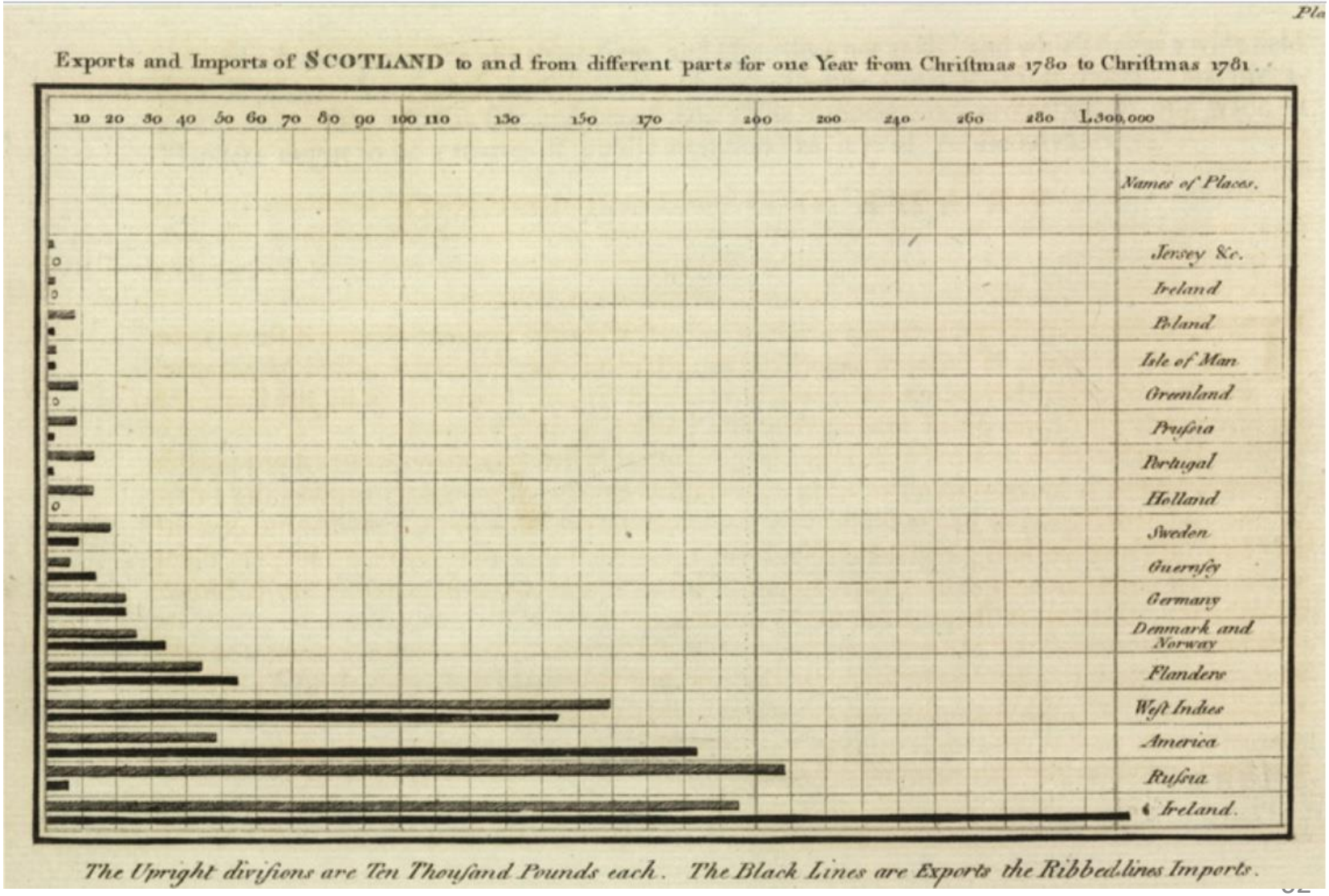


William Playfair (1759-1823)



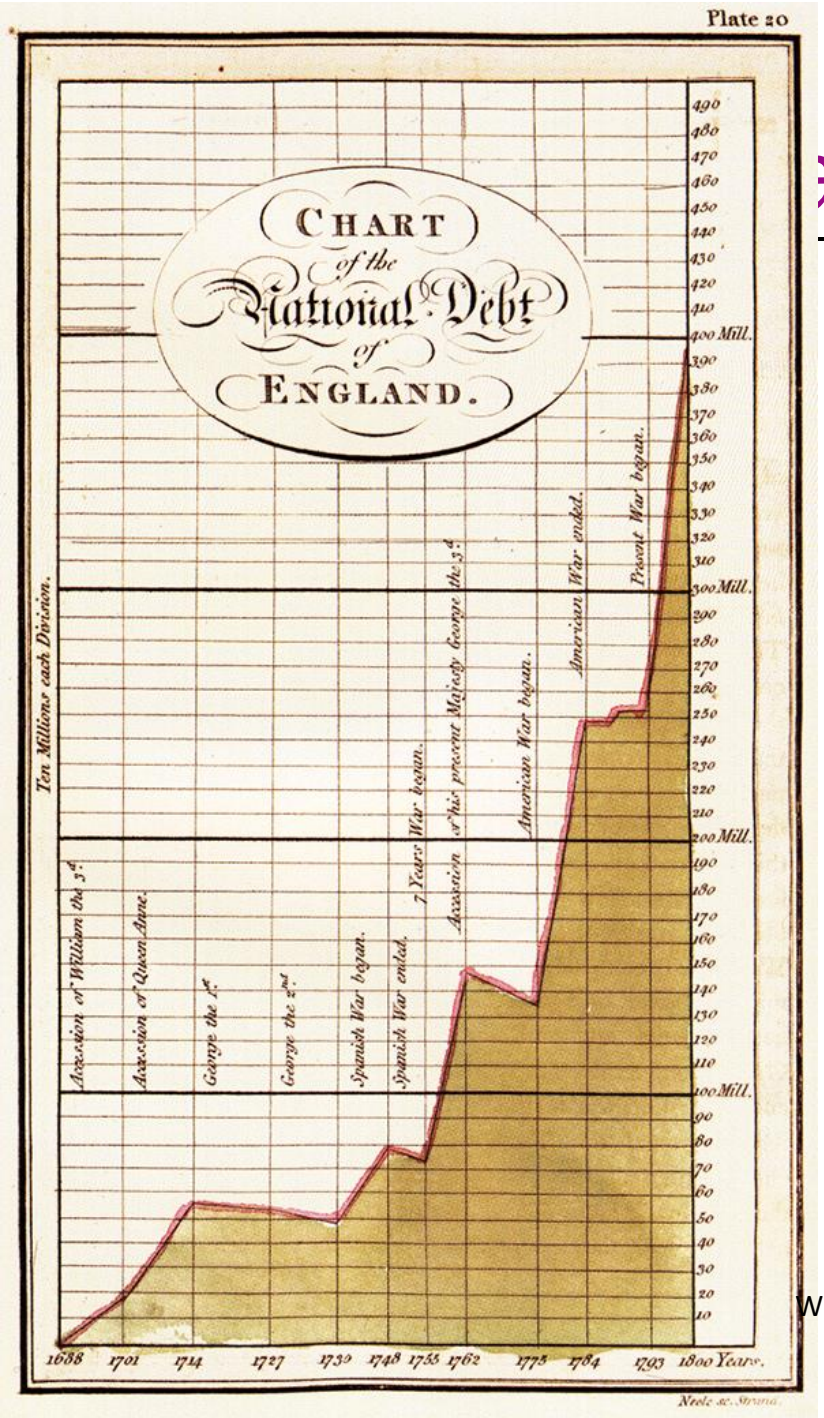
William Playfair: trade-balance time-series chart, 1786

William Playfair (1759-1823)



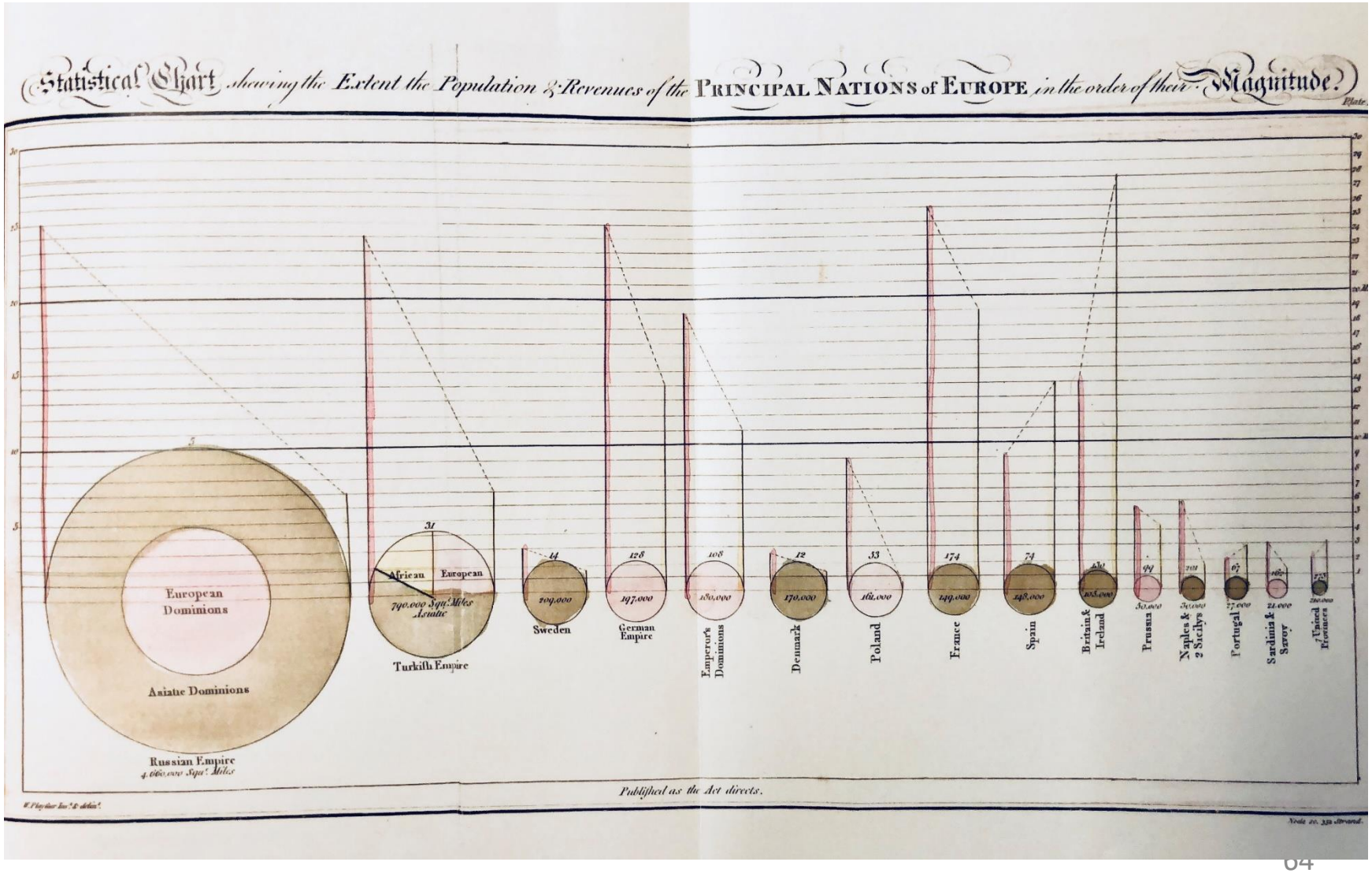
William Playfair: Exports and Imports of Scotland, 1786

air (1759-1823)



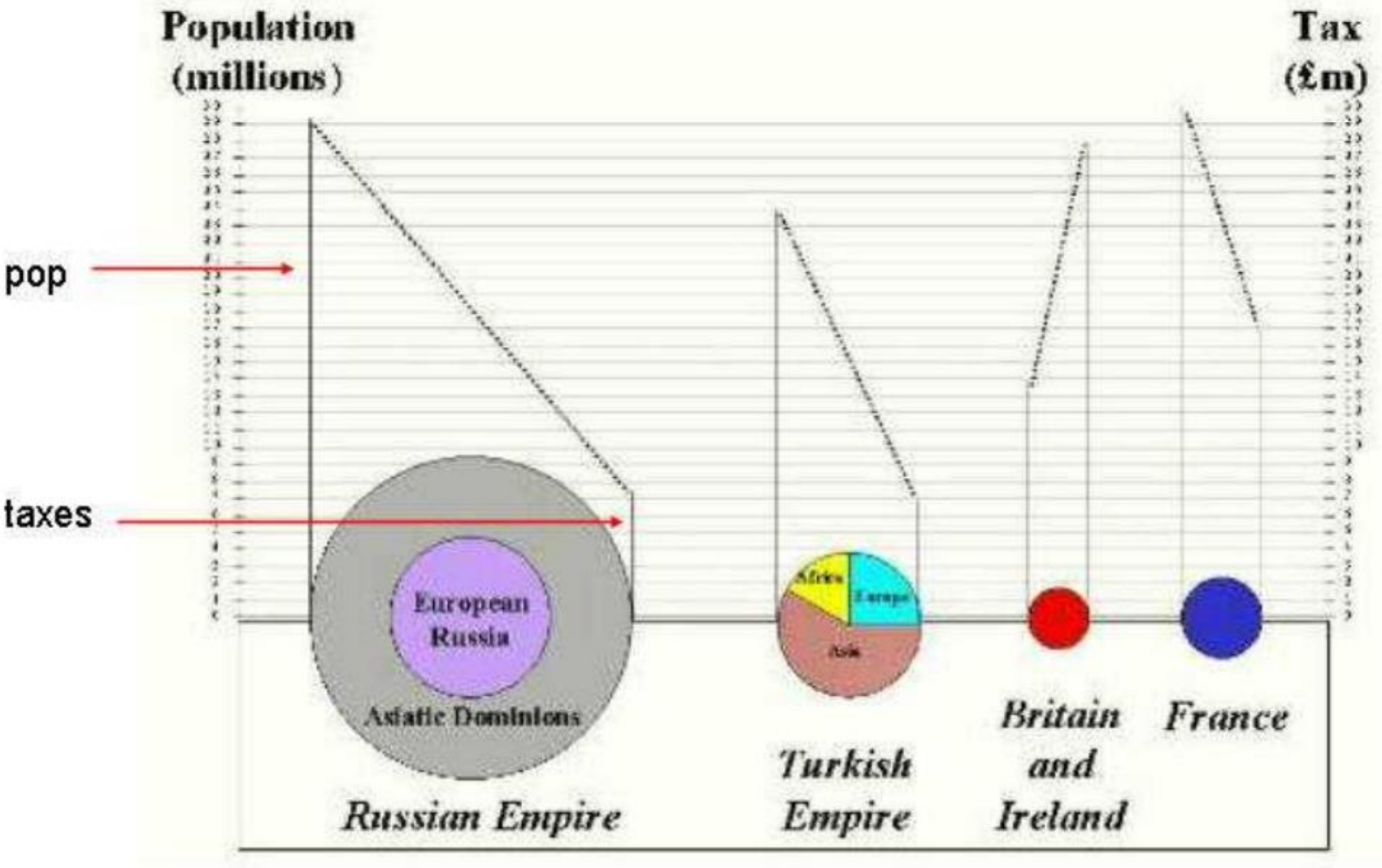
William Playfair, National debt of England, 1801

William Playfair (1759-1823)



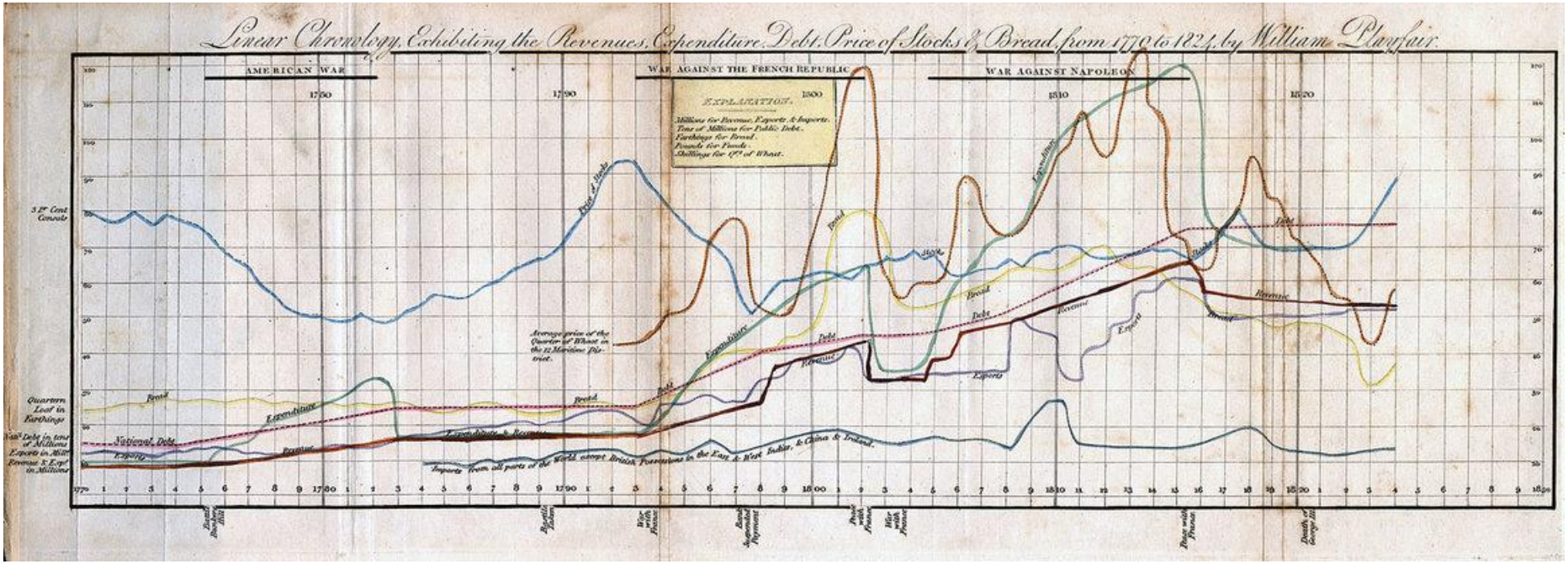
William Playfair: Comparison of population and taxes of nations, 1801

William Playfair (1759-1823)



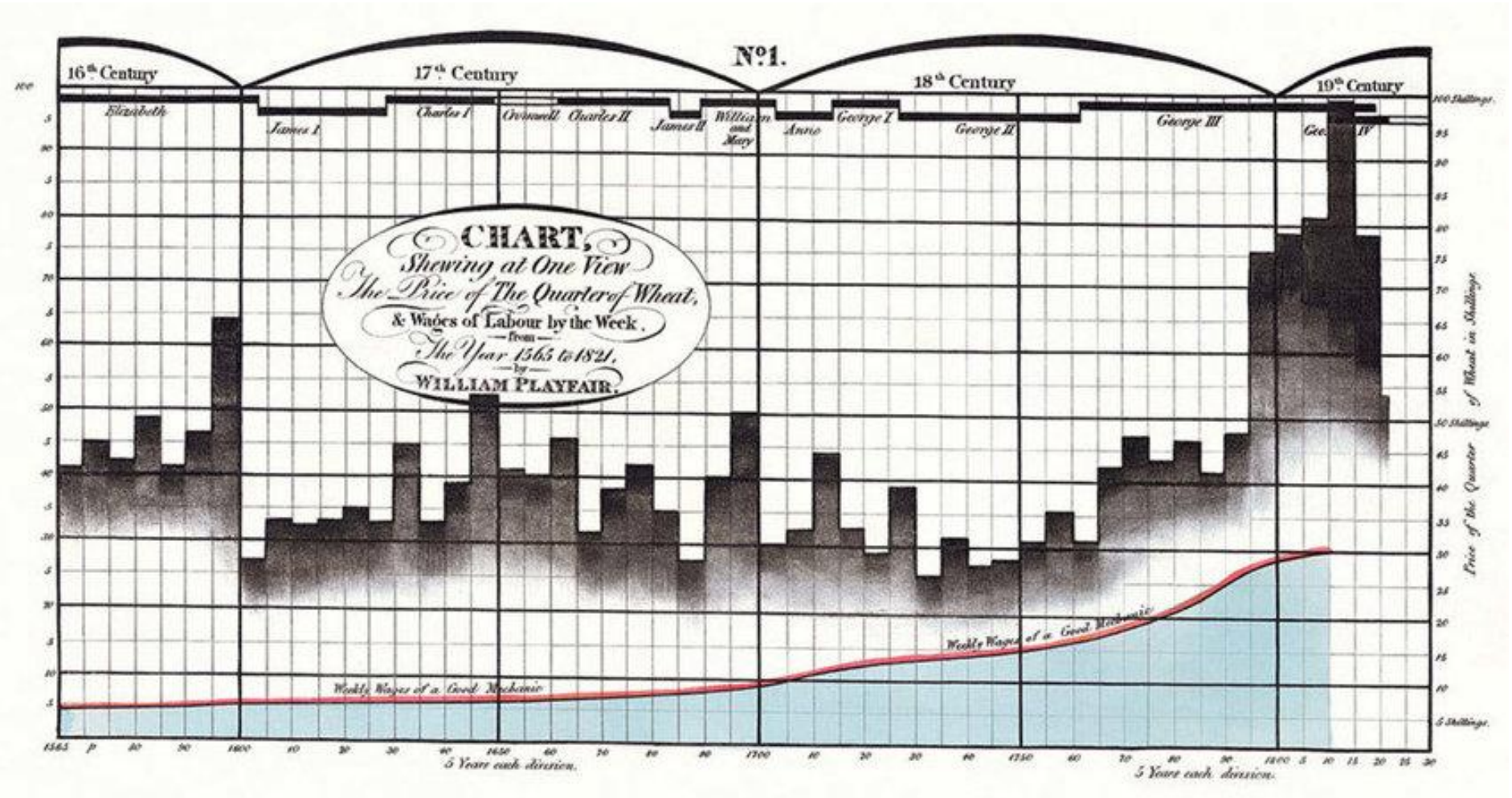
Recreation of the Willaim Playfair plot from previous slide. Friendly, 2006

William Playfair (1759-1823)



William Playfair: Bread and stock prices affected by war, 1824

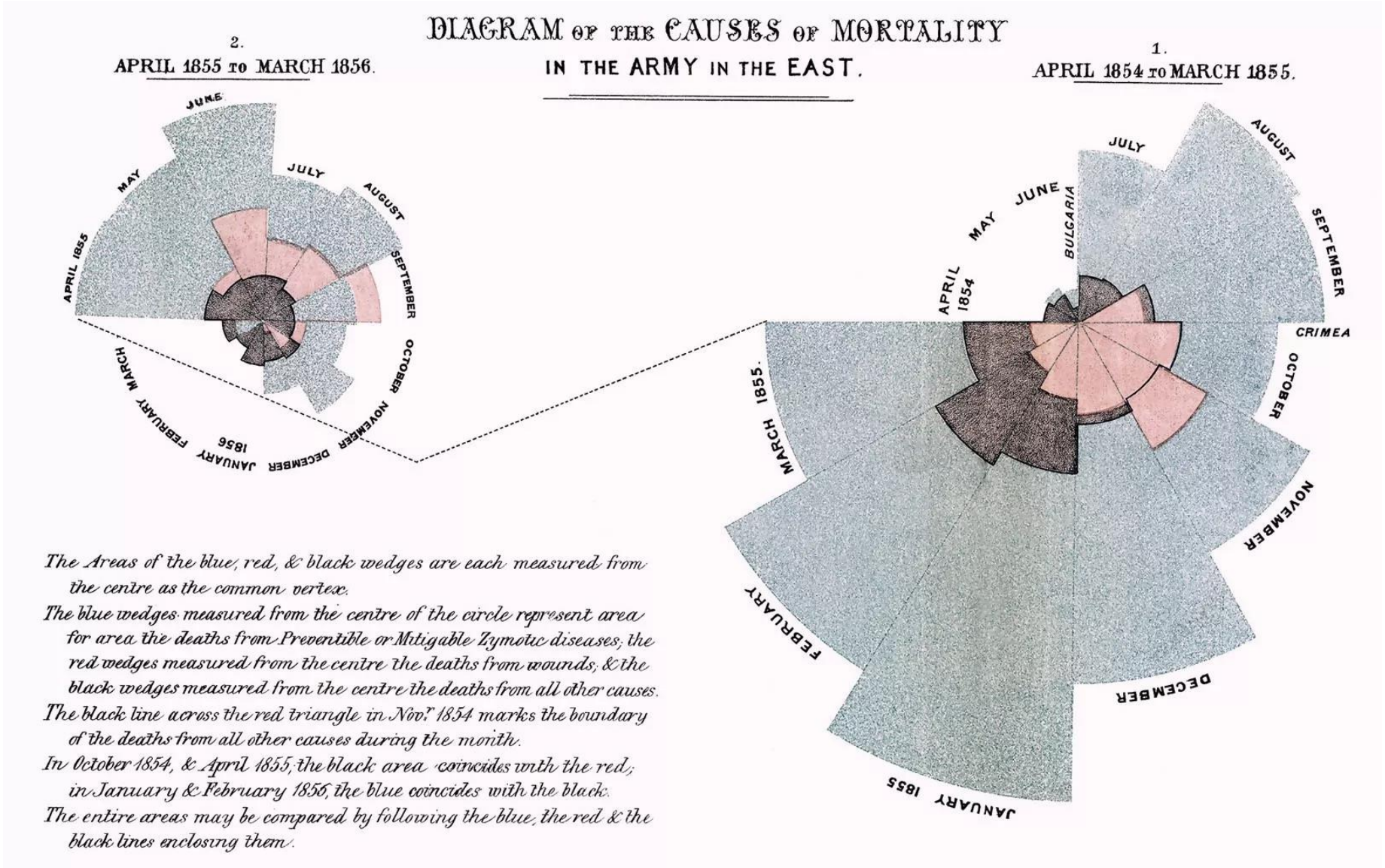
William Playfair (1759-1823)



William Playfair: Price of wheat and weekly wages, 1821
Note the use of 3 time-series datasets: the reigning monarchs (top), price of wheat (barchart), weekly wages of "a good mechanic" (line chart)

1800-1850: Beginnings of Modern Graphics

Nightingale's Rose Chart

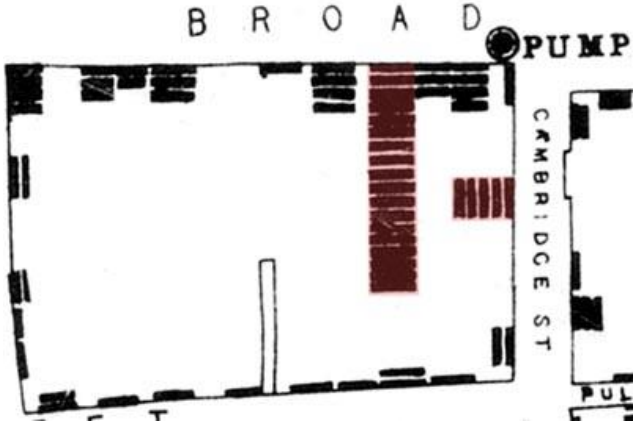


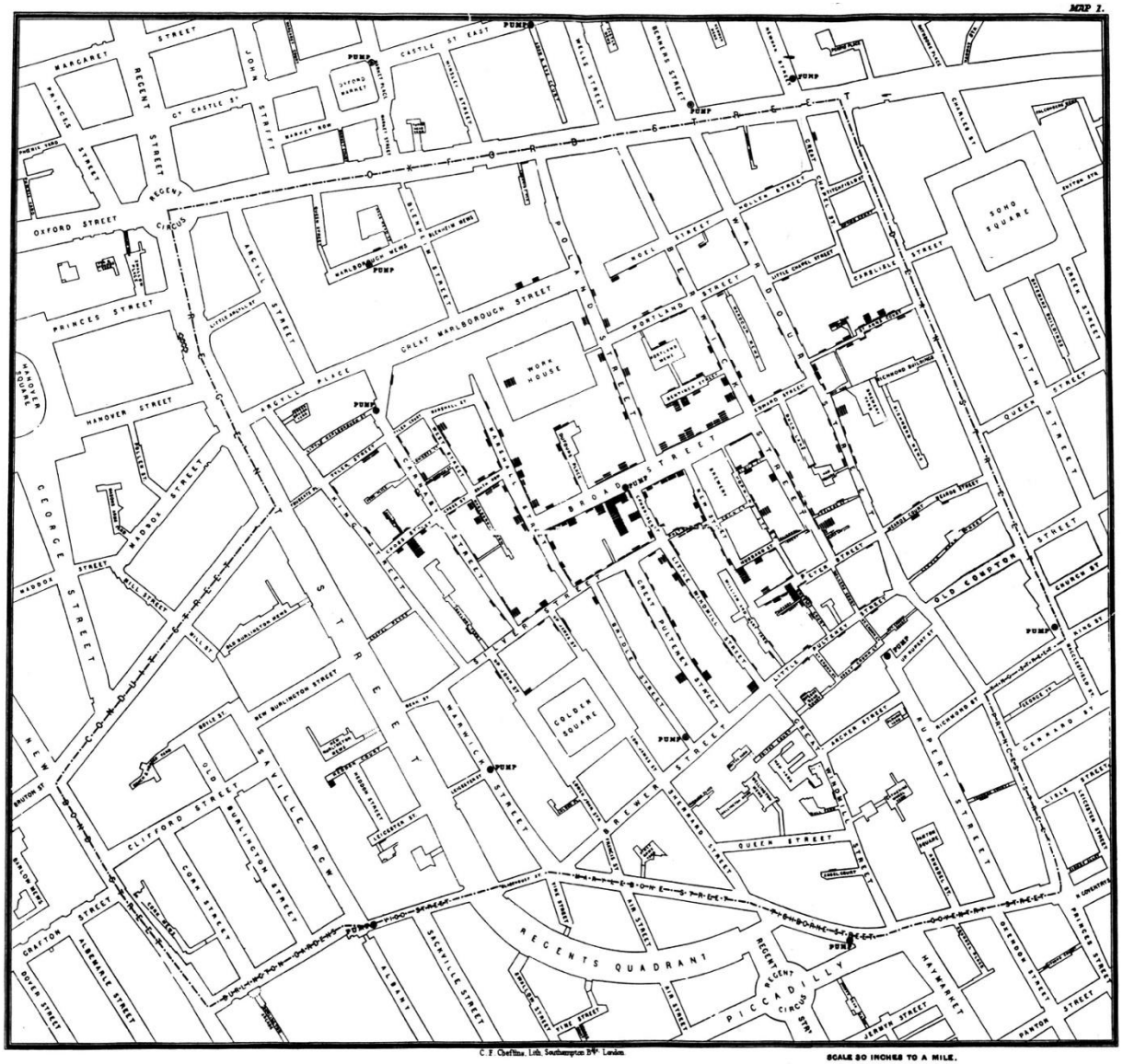
Guerry's Choropleth Map

- Andre-Michel Guerry, in 1832, published his essay on "Moral Statistics of France," used choropleth maps to argue for more social laws (note: "instruction" above is a short for "school instructions"; i.e. the amount of education)

Snow's Map of the 1854 London Cholera Outbreak

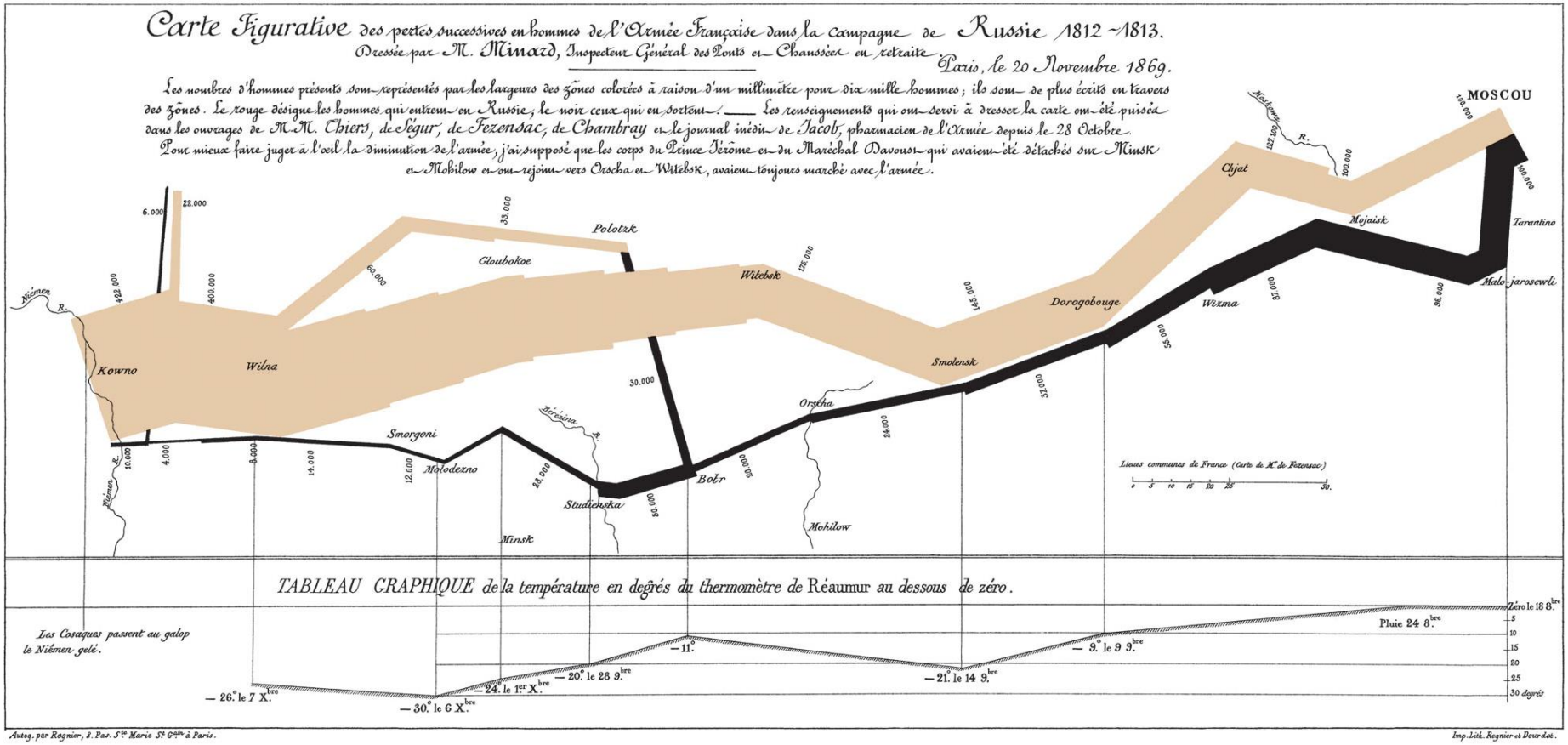
- A successful combination of statistic graphic (bar chart) with geographic map





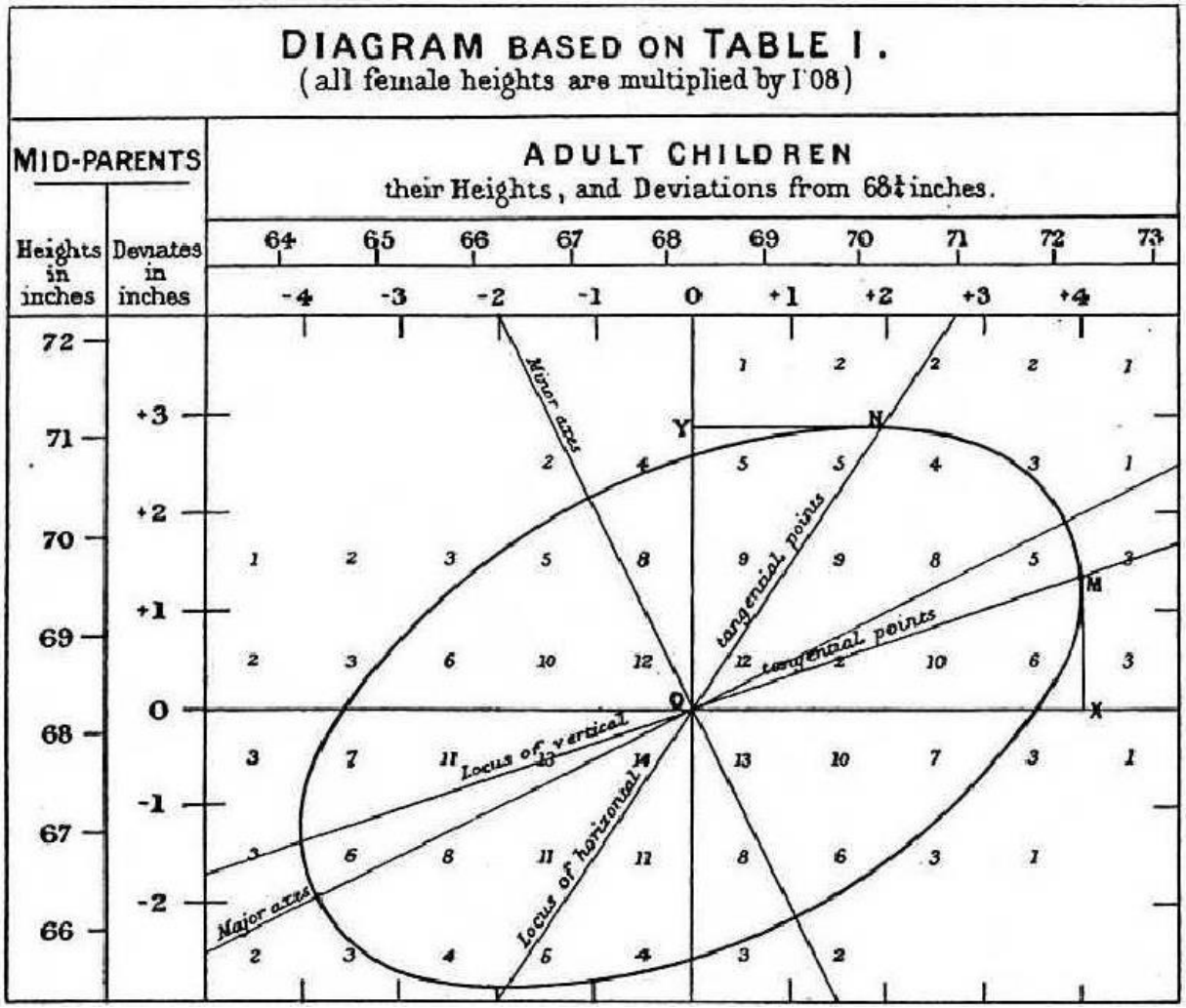
Map of Napoleon‘s Russian campaign of 1812

- Presenting much information in one single map
  - Number of Napoleon's troops; distance; temperature; latitude and longitude; direction of travel; location



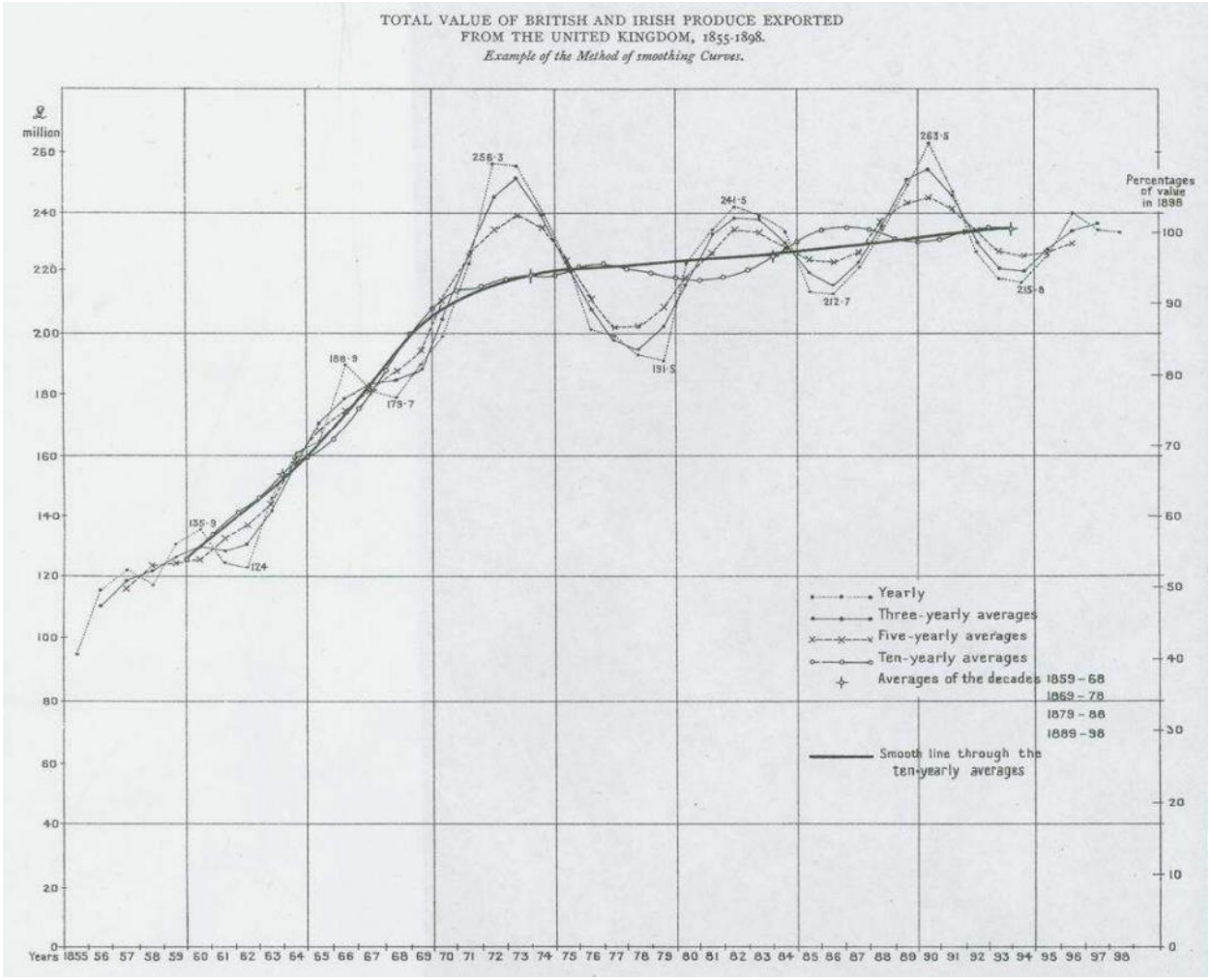
1850-1900: The Golden Age of Statistical Graphics

Statistical Graphics



- Francis Galton (1822-1911)
- Study of correlation and regression
- Created bivariate visualizations to illustrate statistical patterns
- Left, Galton's smoothed correlation diagram for the data on heights of parents and children, 1886

Statistical Graphics

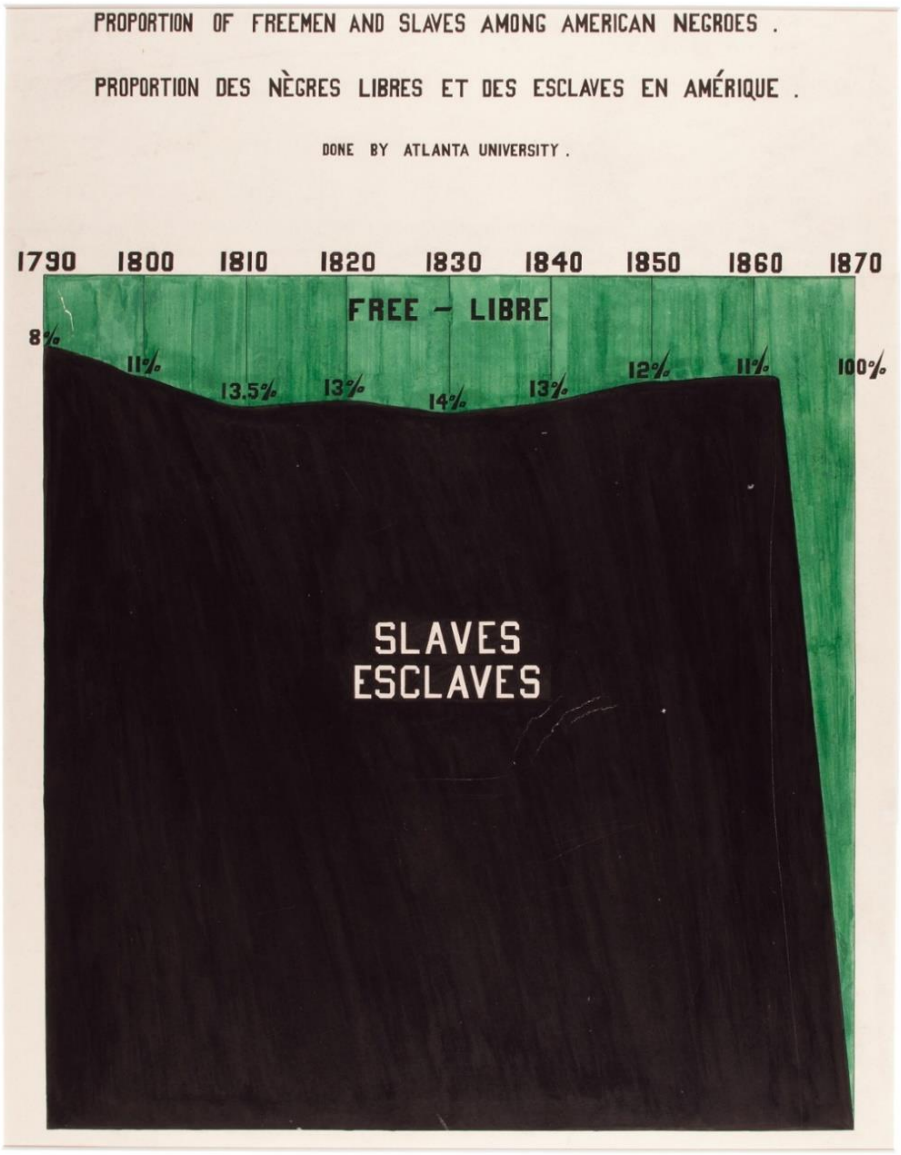


- Arthur Bowley's demonstration of smoothing a time-series graph, 1901

The Start of Infographics

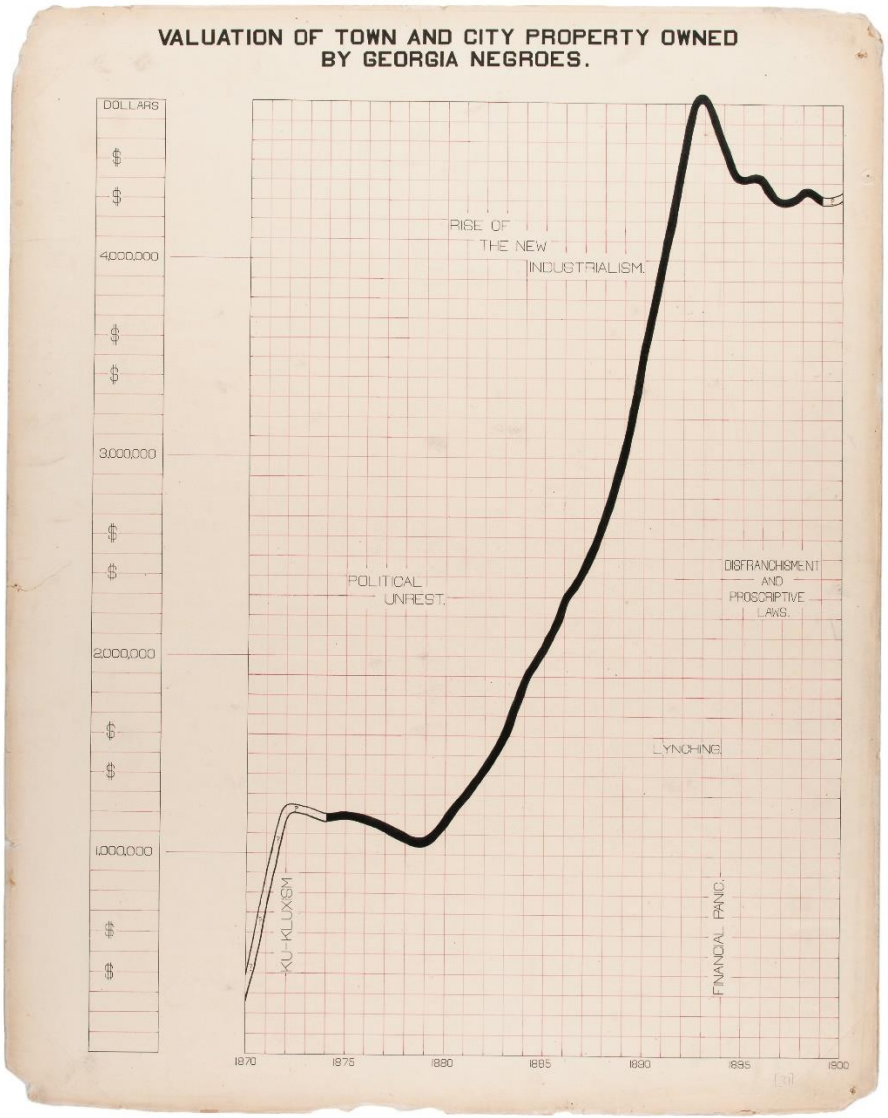
- W. E. B. Du Bois (1868-1963)
  - Prominent leader of civil rights movement
  - Created a number of visualizations for the World's Fair in 1900 in Paris ("*L'Exposition de Paris 1900")*

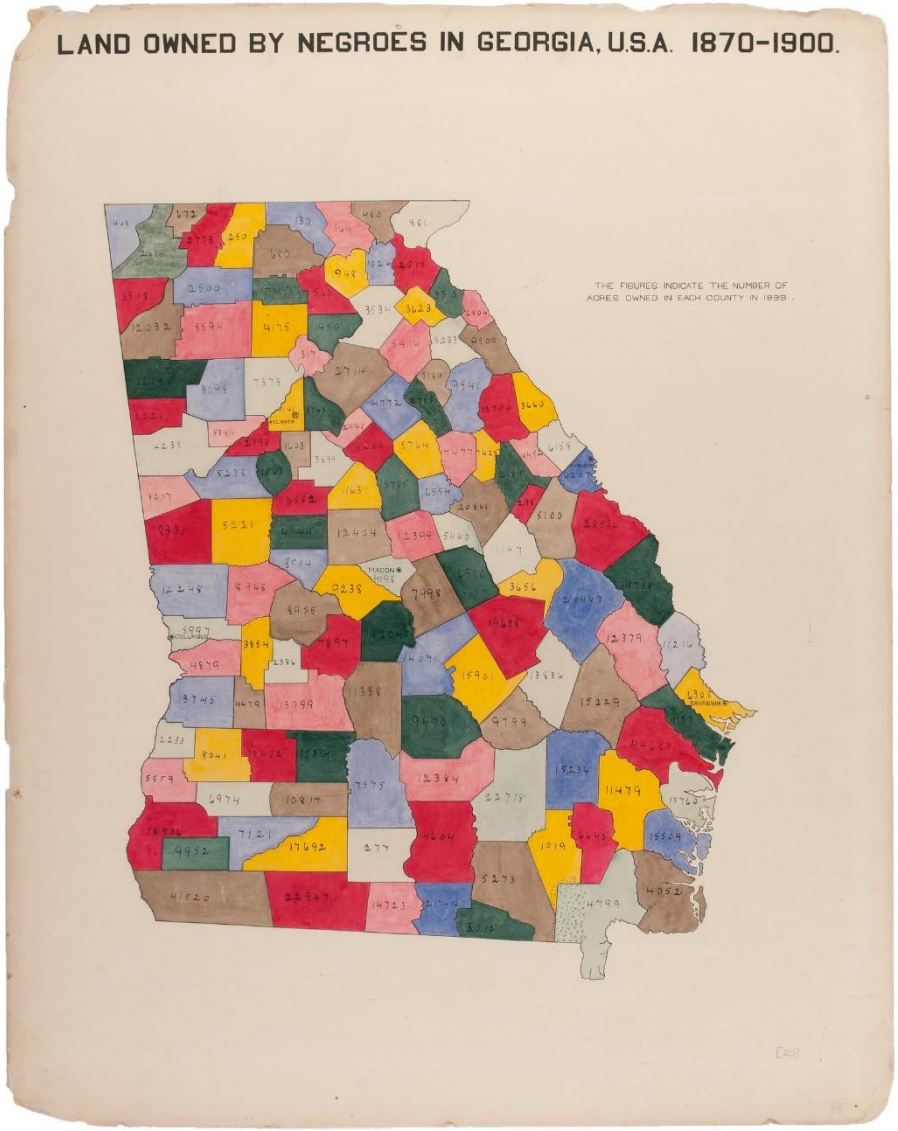
W. E. B. Du Bois


PROPORTION OF FREEMEN AND SLAVES AMONG AMERICAN NEGROES .
PROPORTION DES NÈGRES LIBRES ET DES ESCLAVES EN AMÉRIQUE .
DONE BY ATLANTA UNIVERSITY .
1790 1800 1810 1820 1830 1840 1850 1860 1870
FREE — LIBRE
8%
11%
13.5%
13%
14%
13%
12%
11%
100%
SLAVES
ESCLAVES


CITY AND RURAL POPULATION.
1890.
78,189 NEGROES IN CITIES OF OVER 10,000 INHABITANTS
8,025 NEGROES IN CITIES FROM 5,000 TO 10,000
37,689 NEGROES IN CITIES FROM 2,500 TO 5,000
734,952
NEGROES LIVING IN THE COUNTRY AND VILLAGES

W. E. B. Du Bois


VALUATION OF TOWN AND CITY PROPERTY OWNED BY GEORGIA NEGROES.
DOLLARS
4,000,000
3,000,000
2,000,000
1,000,000
RISE OF THE NEW INDUSTRIALISM.
POLITICAL UNREST.
DISFRANCHISEMENT AND PROSCRIPTIVE LAWS.
LYNCHING
KU-KLUXISM
FINANCIAL PANIC.
1870
1875
1880
1885
1890
1895
1900


LAND OWNED BY NEGROES IN GEORGIA, U.S.A. 1870-1900.
THE FIGURES INDICATE THE NUMBER OF ACRES OWNED IN EACH COUNTY IN 1899.

~~**1900-1950**: The Modern Dark Ages~~

1950-1975: Rebirth of Data Visualization

1975-1990: High-D, Interactive, and Dynamic Data Visualization

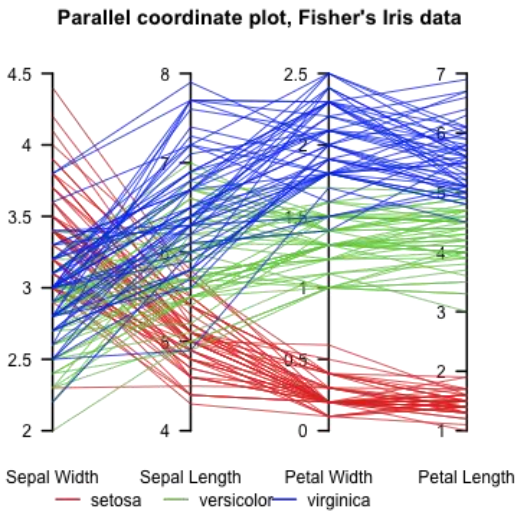
Key Events and People

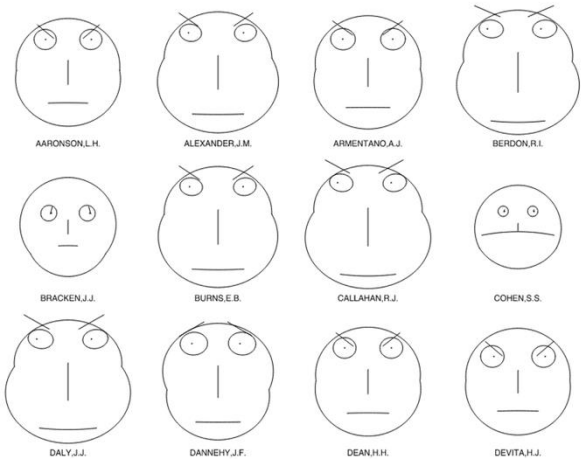
- John Tukey (1915-2000):
  - Coined the term “exploratory data analysis” (EDA), 1962
  - Advocated for the use of visualizations for EDA
  - Popularized the use of boxplots, stem-leaf plots, etc. in statistics research

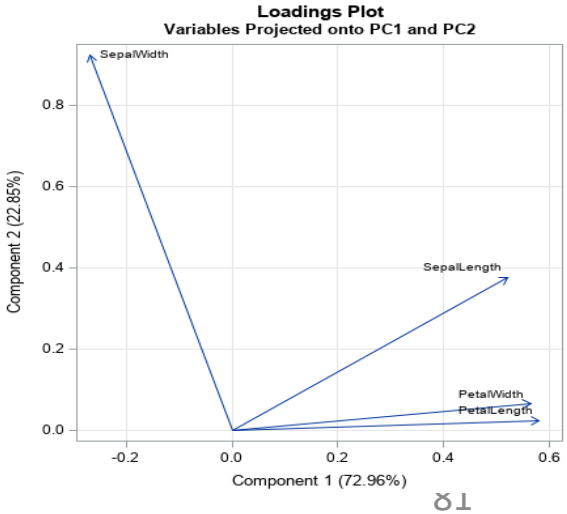
- Jacques Bertin (1918-2010):
  - Wrote the seminal book “*Sémiologie Graphique”*, or “Semiology of Graphics” (1967)

Inventions

- Statistical software with visualization components
  - FORTRAN, C, APL, S (S-plus)
- High-dimensional data visualization
  - Grand Tour (Asimov, 1985)
  - Projection Pursuit (Tukey, 1974)
  - Parallel Coordinates (Inselberg, 1985)
- High-dimensional data visualization
  - Scatterplot Matrix (Tukey, 1981)
  - Chernoff Faces (1973)
  - Biplots (1971)
- Perceptual and cognitive studies
  - Bill Cleveland's Bar vs. Pie study (1984)

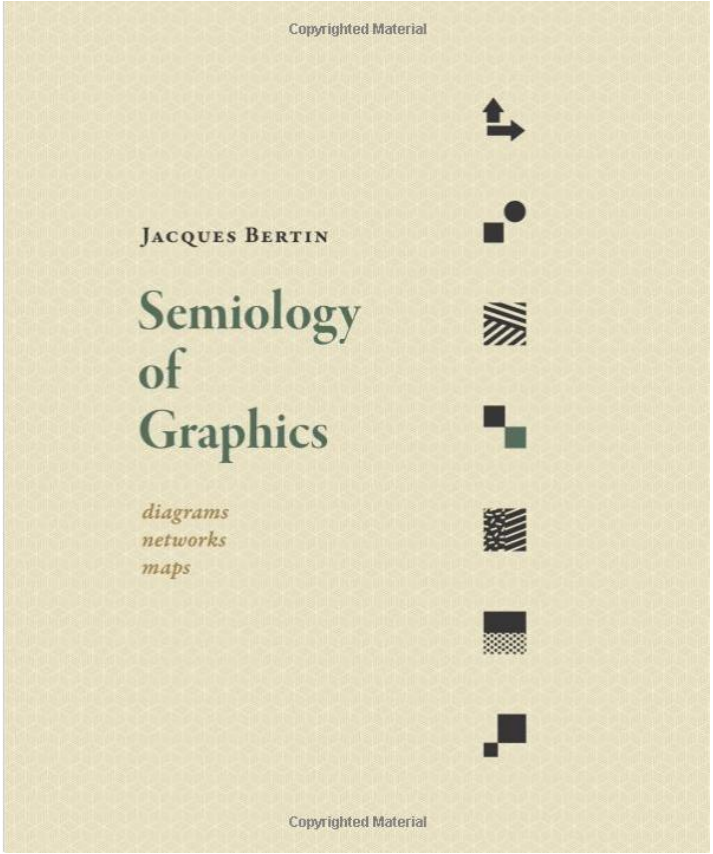






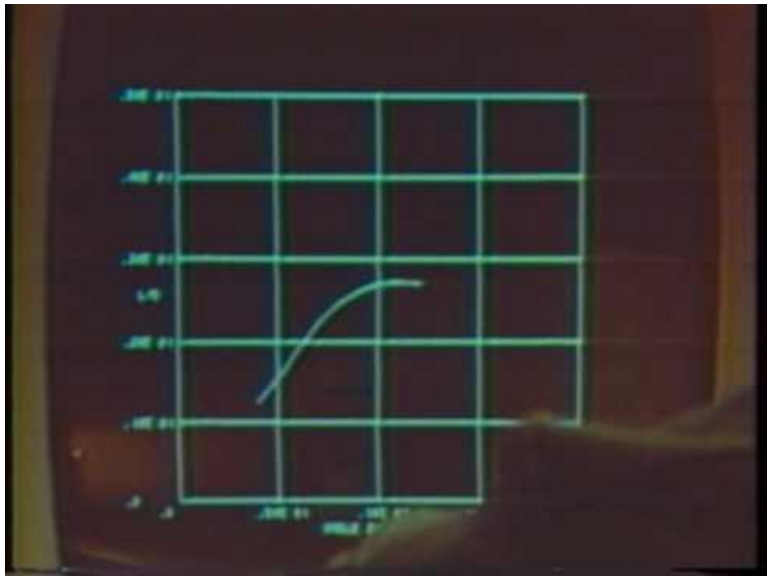
Semiology of Graphics

- Jacques Bertin's effort in formalizing visualization:
  - Considered the equivalent of *periodic table* to Chemistry, but for data visualization
  - This can be considered as the start of "InfoVis" as a scientific discipline

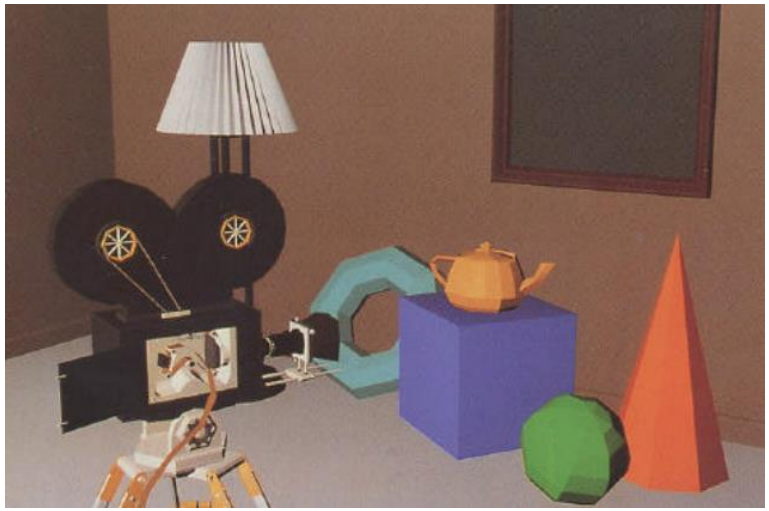


1990-Now: Graphics, HCI, Scientific Visualization, Information Visualization, and Visual Analytics

Computer Graphics (1969)

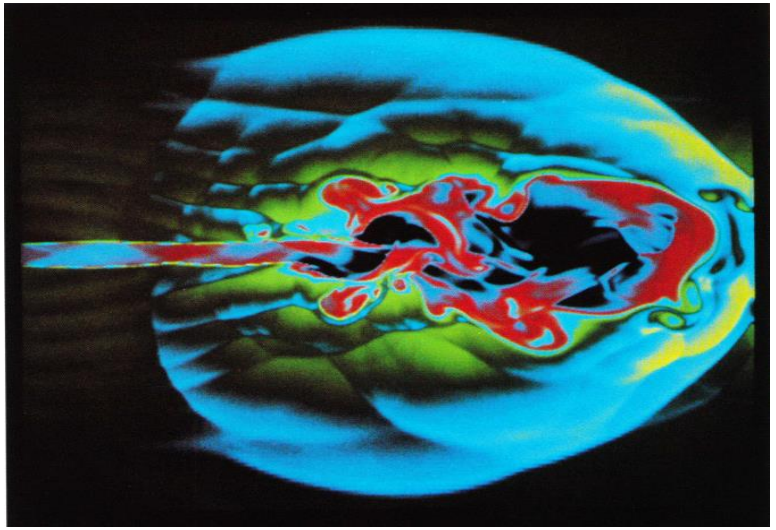


SketchPad, Sutherland, 1963



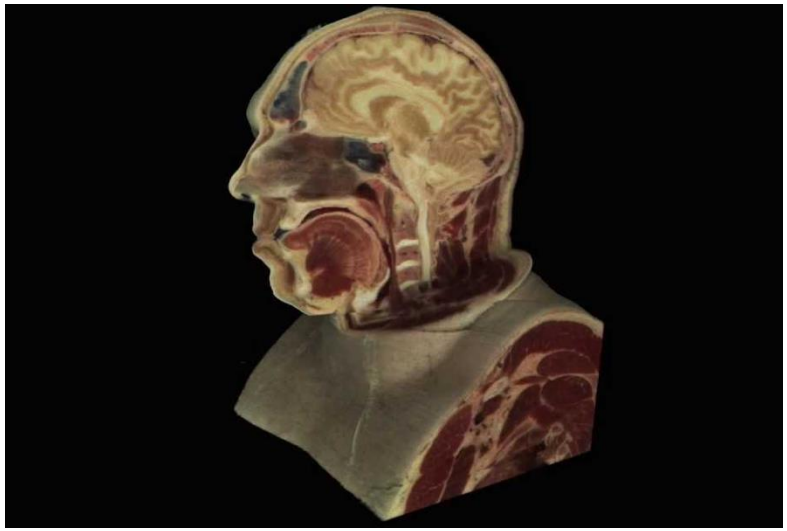
Ray Casting / Ray Tracing, 1967-80s

Computer Graphics (1969)



Scientific Visualization (1990)

Air flow visualization from a super sonic jet, Haber and McNabb, 1990

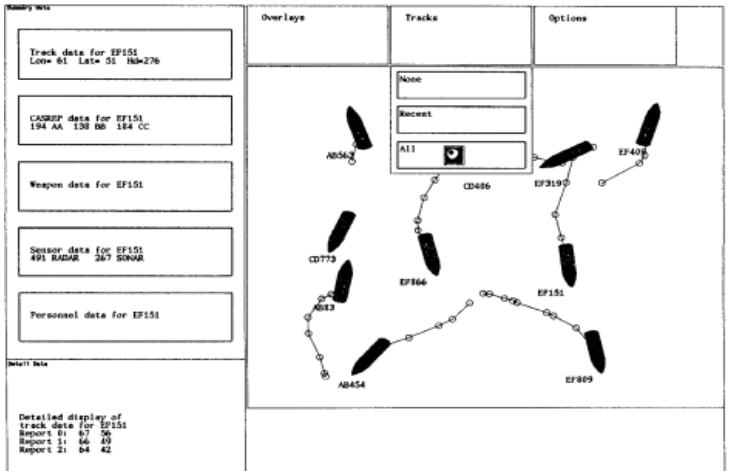


The “Visible Human Project”, 1994

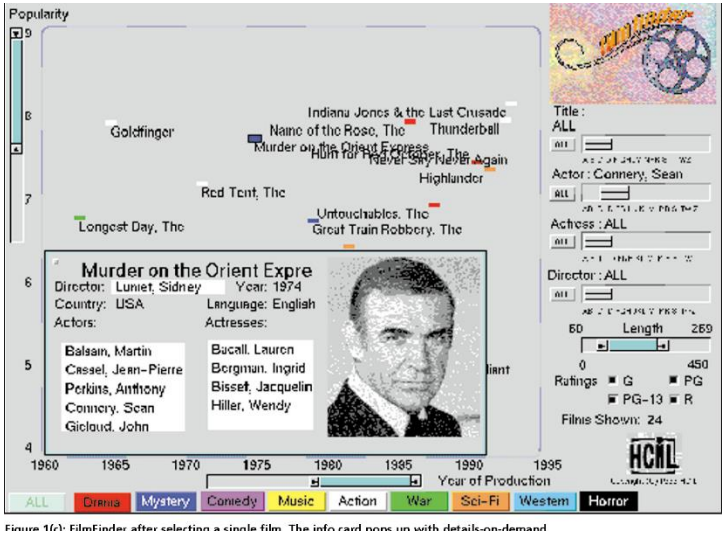
Computer Graphics (1969)

Scientific Visualization (1990)

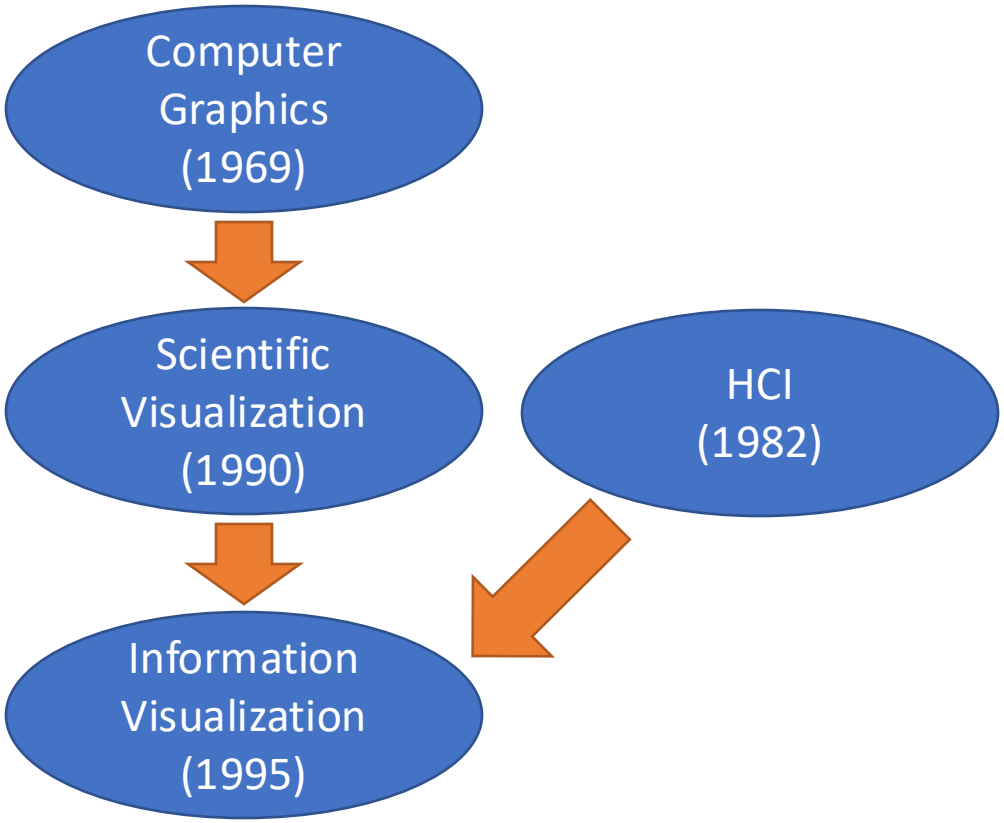
HCI (1982)

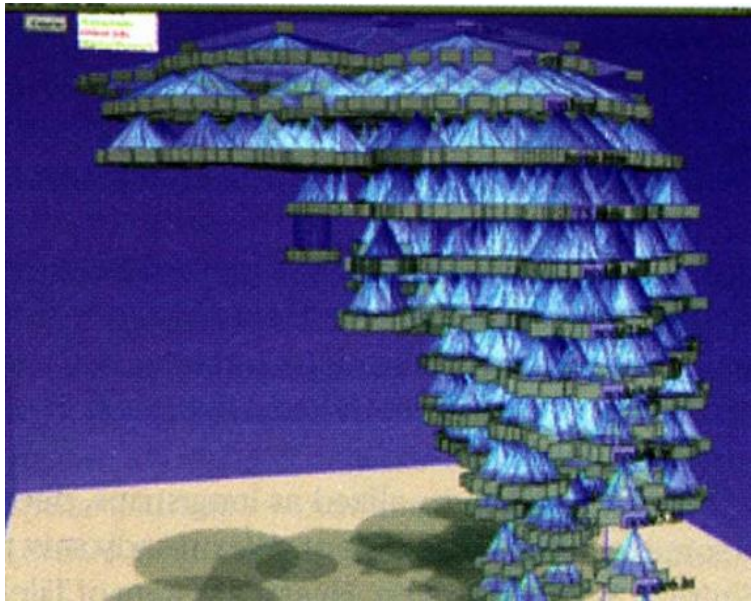


Eye tracking study, Rob Jacob, 1991

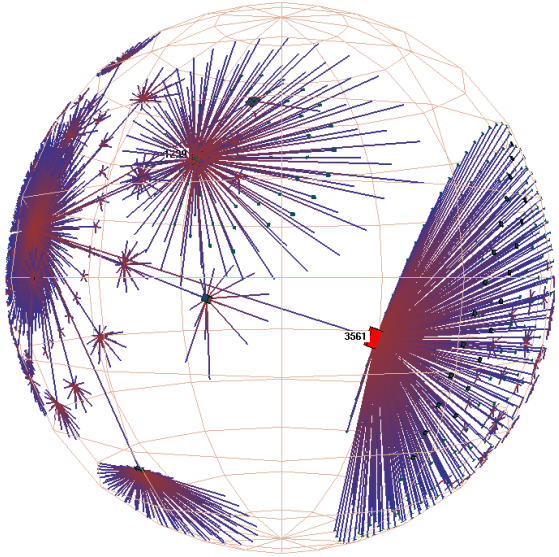


FilmFinder, Ahlberg and Shneiderman, 1994





Cone Tree, Robertson, 1991



Hyperbolic Tree, Munzner, 1998

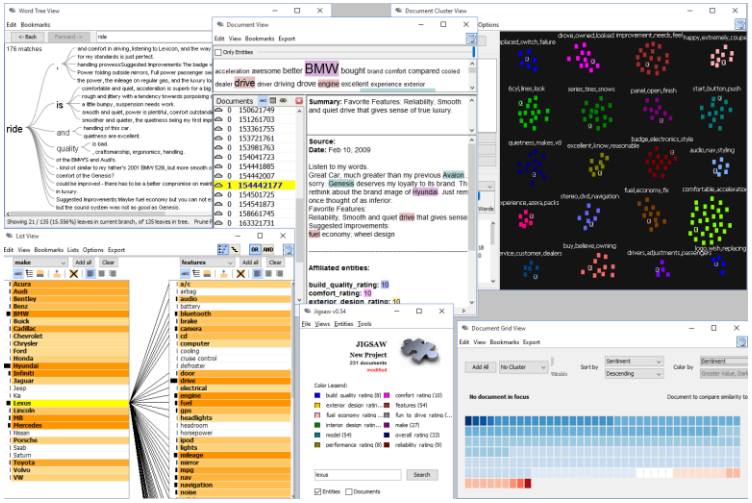
Computer Graphics (1969)

Scientific Visualization (1990)

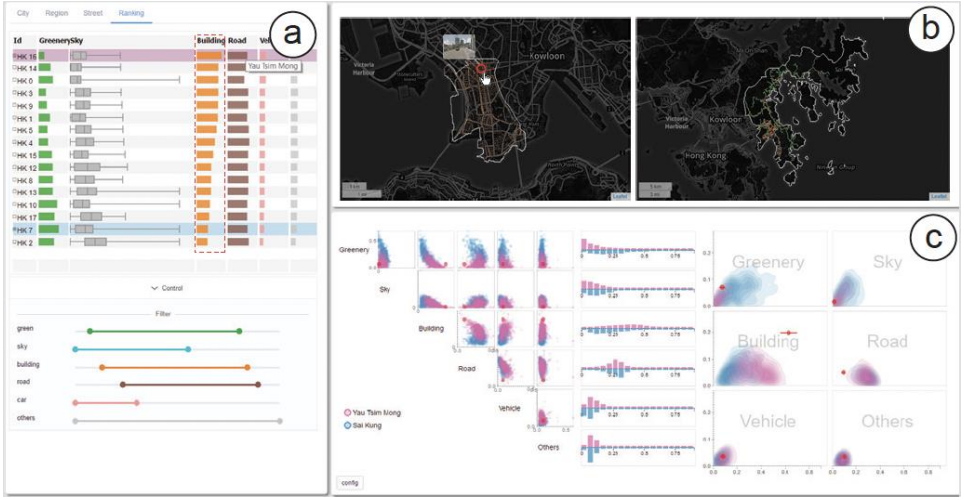
HCI (1982)

Information Visualization (1995)

Visual Analytics (2006)



Jigsaw, Stasko, 2007



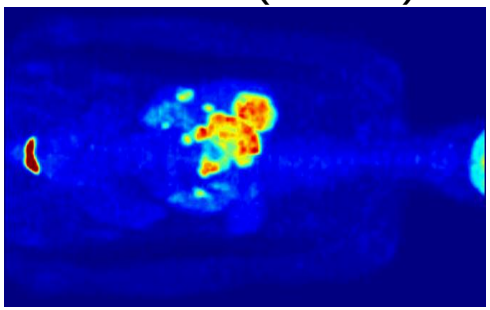
StreetVizor, 2017

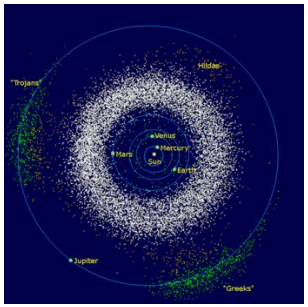
Outline

- **Introduction to data visualization**
  - What: concept and examples
  - Why: using vision to think
- **History of data visualization**
  - Evolution of techniques
  - Impact and utilization
- **Disciplines**
  - Subfields: SciVis, InfoVis, VAST
  - Relations with other fields

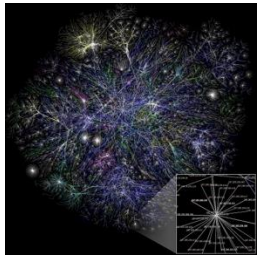
Subfields

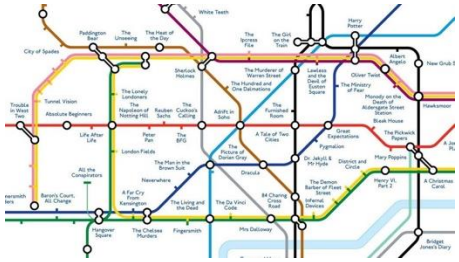
- Scientific Visualization (SciVis) – Spatial data



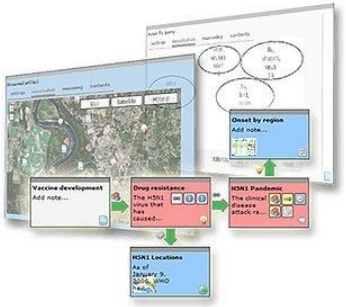


- Information Visualization (InfoVis) – Abstract data



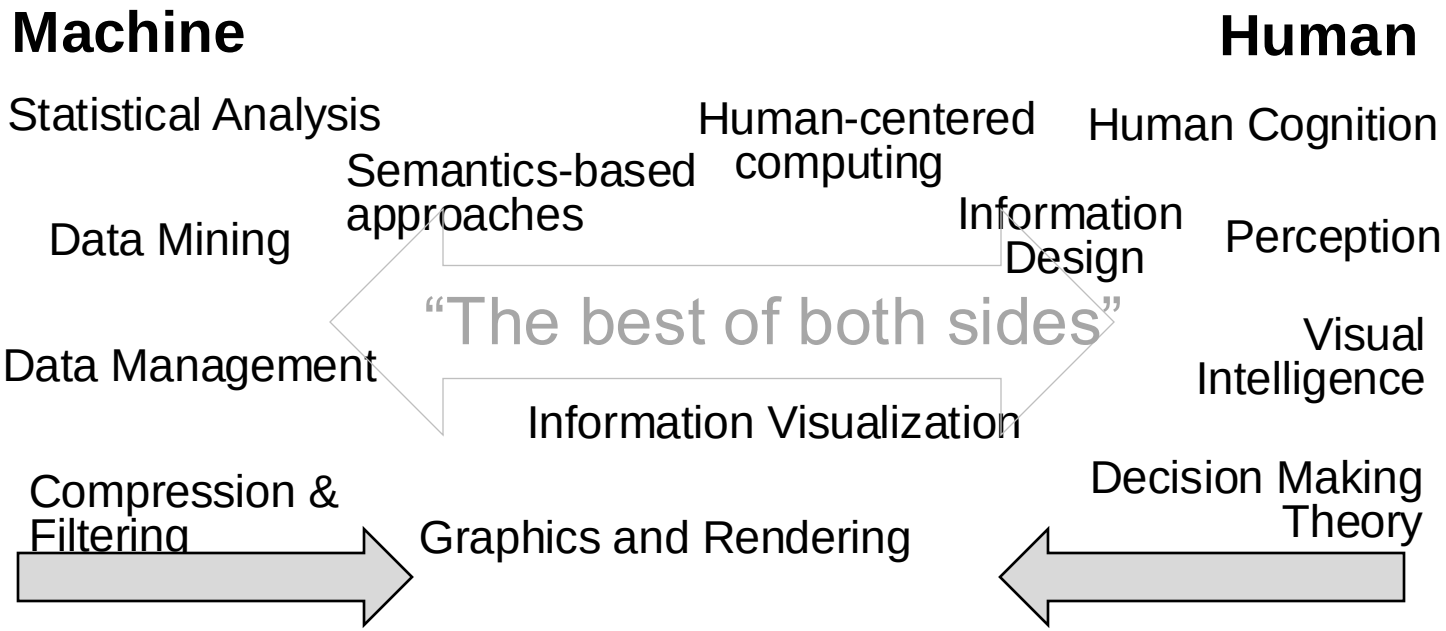


- Visual Analytics (VAST) – Analytical reasoning



Visual Analytics

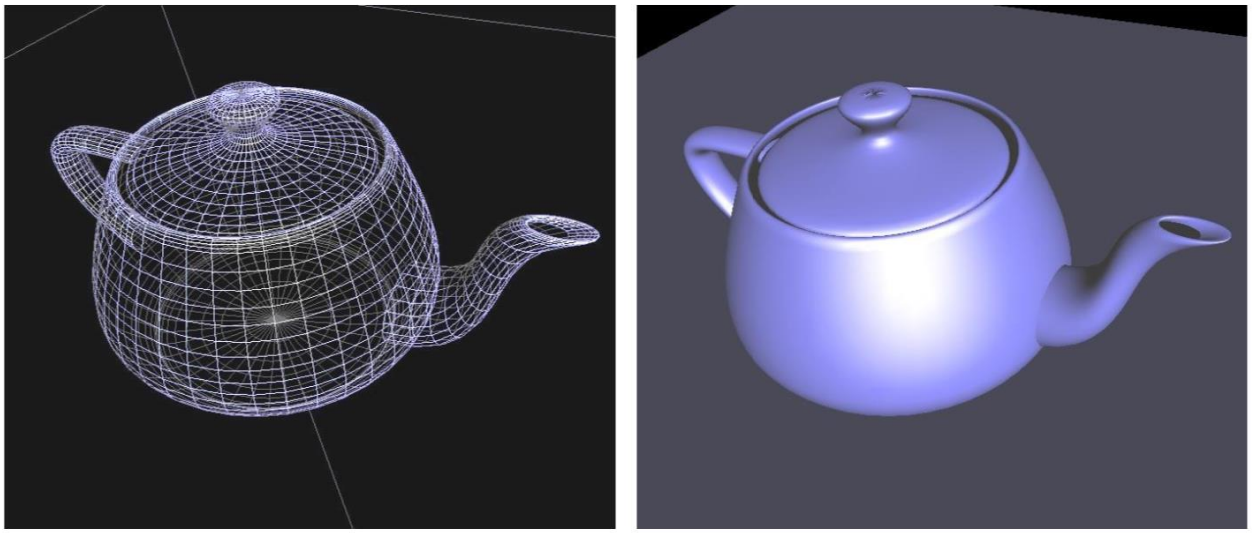
- The science of *analytical reasoning* facilitated by *interactive visual interfaces* [Thomas and Cook, 2004]
- Combines *automated analysis* with *interactive visualizations* [Keim et al., 2010].



credit: Keim et al., 2008

Vis. vs. Computer Graphics

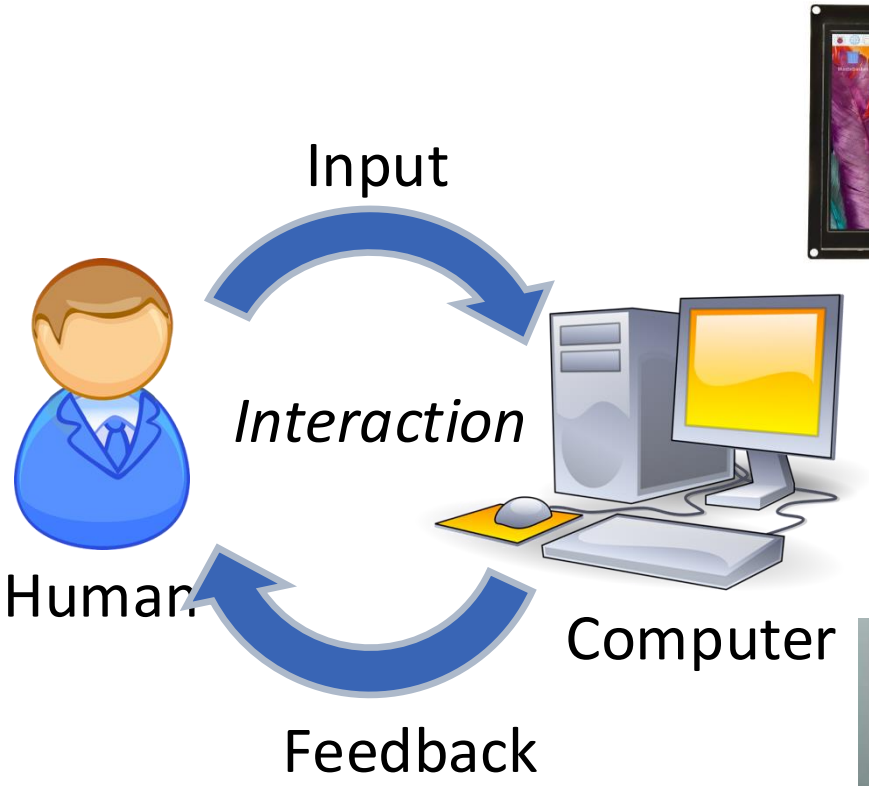
- **Computer graphics:** The process of generating an image with *a virtual camera*, *objects*, and *light sources*
  - Visualization represents data graphically and interacts with these representations in order to gain insight into the data
  - Computer graphics provides a powerful mechanism for creating, manipulating, and interacting with these representations



Real-time rendering [Akenine-Moeller]

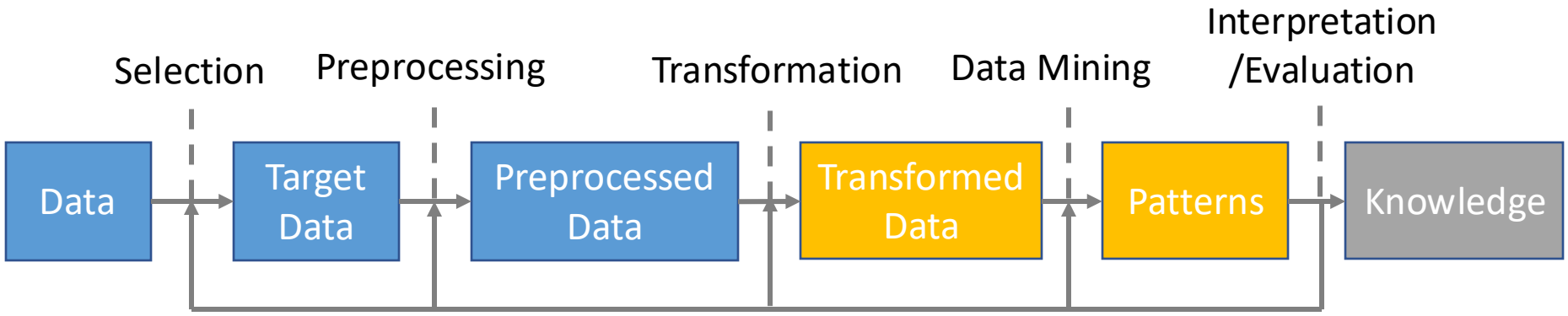
Vis. vs. HCI

- **Human-Computer Interaction**
  - Visualization focuses on creating visual images for dataset



Vis. vs. KDD

- **Data Mining:** Exploration & analysis, by **automatic** or **semi-automatic** means, of data in order to discover meaningful patterns.
  - **Supervised:** regression, classification, anomaly detection, etc.
  - **Unsupervised:** clustering, association, etc.
- Data exploration and visualization keeps human in the analysis loop
  - Exploratory data analysis (EDA), visual data mining.



The KDD process [Fayyad et al., 1996]

Summary

- **What is data visualization?**
  - The use of computer-supported, interactive, visual representations of data to amplify cognition.
- **Why data visualization?**
  - Data is huge, complex, and heterogeneous
  - The power of visual cognition
- **History of data visualization**
  - “A Brief History of Data Visualization”, Friendly, Handbook of Computational Statistics: Data Visualization
- **Discipline**
  - A multi-disciplinary research field embracing computer graphics, data mining, HCI, etc.