

Data Analysis and Visualization

DR. MUHAMMAD FAISAL CHEEMA

FAISAL.CHEEMA@NU.EDU.PK

About me

Education:

BS(CS) – IIUI

MS(CS) – **FAST-NU**

PhD. – Leipzig University, Germany

Interests:

Data science

Visual Analytics

Information Visualization

Etc.



About me

Assistant Professor – FAST-NU, 2019

Office: 505-C (New CS Block)

Extension: 515

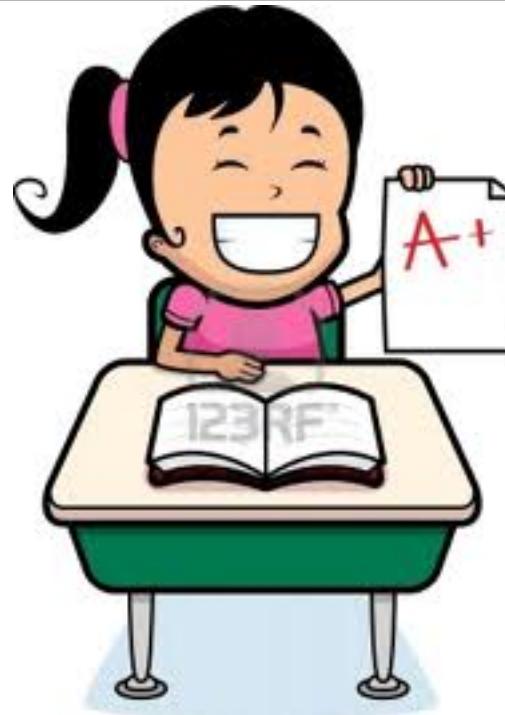
Office hours:

Will be displayed on office door

About you?

You are here because?

- There is no other option ☺
- What if you had an option?



Questions

“He who asks a question is a fool for five minutes; he who does not ask a question remains a fool forever”

Chinese Proverb

“The wise man doesn't give the right answers, he poses the right questions.”

Claude Levi-Strauss

“A wise man can learn more from a foolish question than a fool can learn from a wise answer.”

Bruce Lee

Expectations

Remember this is a core course.

- Quality, Extent and Depth of contents largely depends on class attitude.
- Take-up the lead yourself.

Advice on deliverables

- Start early
- Discuss progress regularly
- Complete and submit all deliverables
- Share ideas but do not share your code with anyone
- Be creative

Some Rules

- Raise your hand before asking any question and then WAIT for the permission

- Never ever miss a class
 - **No retakes** (except for the Mid/Final Exam*)
 - **No late submissions accepted**

- Never ever “sleep” in the class
 - You might miss a quiz 😞

- Never use mobile phone in the class

- **Above all, whatever you do, please do not disturb others**

* Conditional: as per university policy

Dishonesty, Plagiarism

All parties involved in any kind of cheating in any exam (Quizzes, Assignments & Projects) will get **0** in that whole set of assessments. E.g. cheating in 1 assignment will lead to 0 in all assignments.

Habitual cases will be forwarded to disciplinary committee.

Dishonesty, Plagiarism

You can fool some of the people **all of the time**, and **all of the people** some of the time, but you can not fool all of the people all of the time.

Abraham Lincoln,

16th president of US (1809 - 1865)

Tentative Division of Marks

Quizzes (5-10)	10
Assignments (3-5)	15
Mid Term Exam	25
Project	10
Final Exam	40
Total	<u>100</u>

Some course topics

Introduction to data visualization, value of visualization, examples of good and bad visualizations. Human perception and cognition, design criticism, graphical integrity principles.

Introduction to Information Visualization and Scientific Visualization. The visualization pipeline - conceptual and implementation perspective.

Data types and dataset types. Basic data transformation and simplification techniques including filtering & queries, pivoting, aggregation etc. Advance data and visual simplification techniques including visual aggregation, dimensionality reduction, and interaction techniques etc.

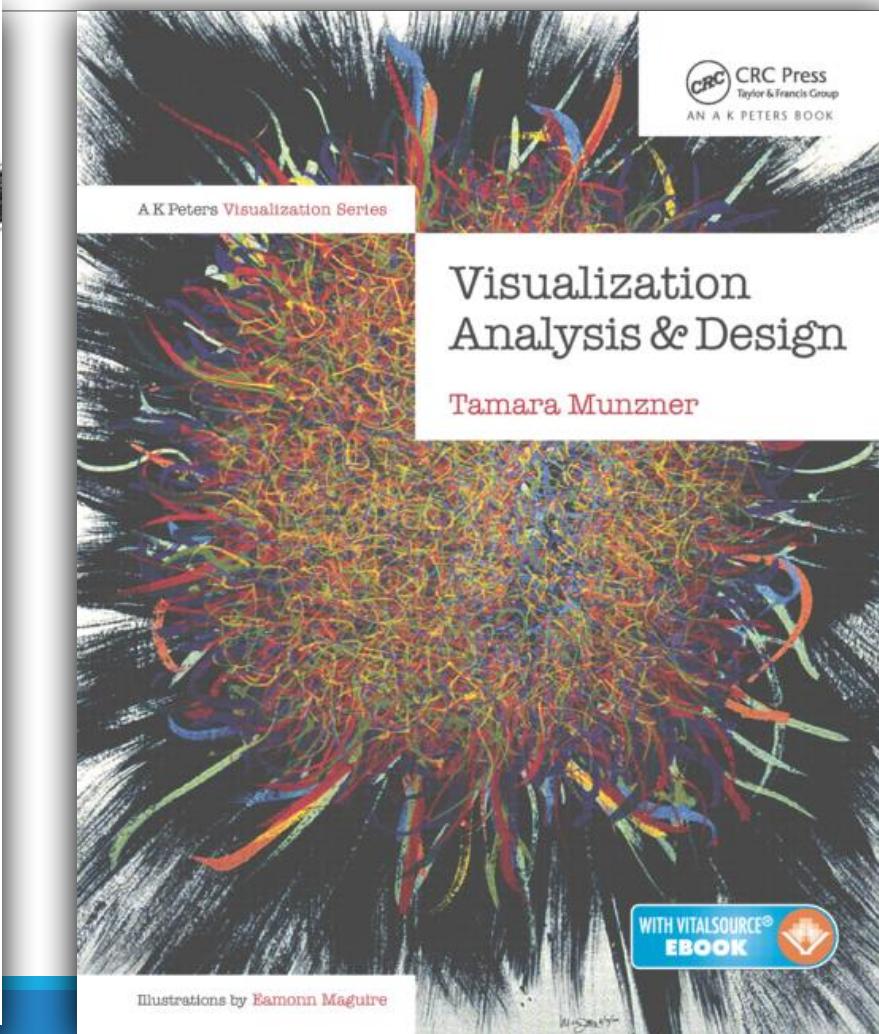
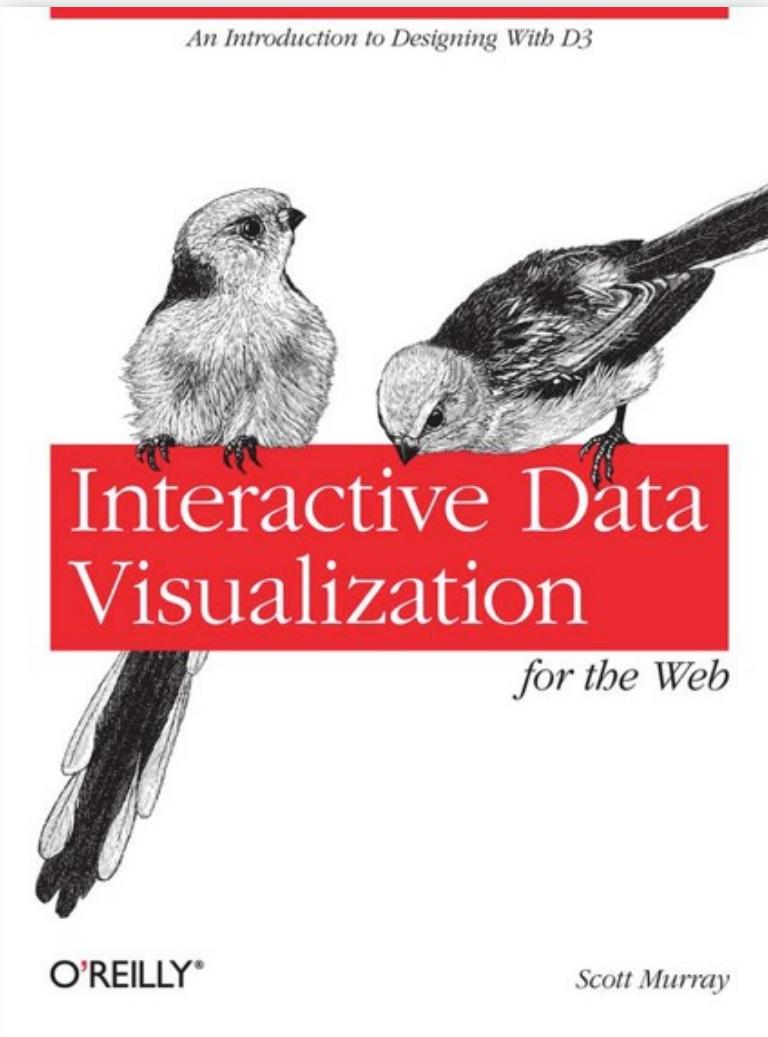
Visualization building blocks (marks, channels and their ranking), visual encoding, expressiveness and effectiveness of a visual encoding, color models and color spaces.

Visualization of high dimensional data, multi-dimensional data, network and hierarchical data, geo-spatial data (maps and cartography), text data and visual storytelling.

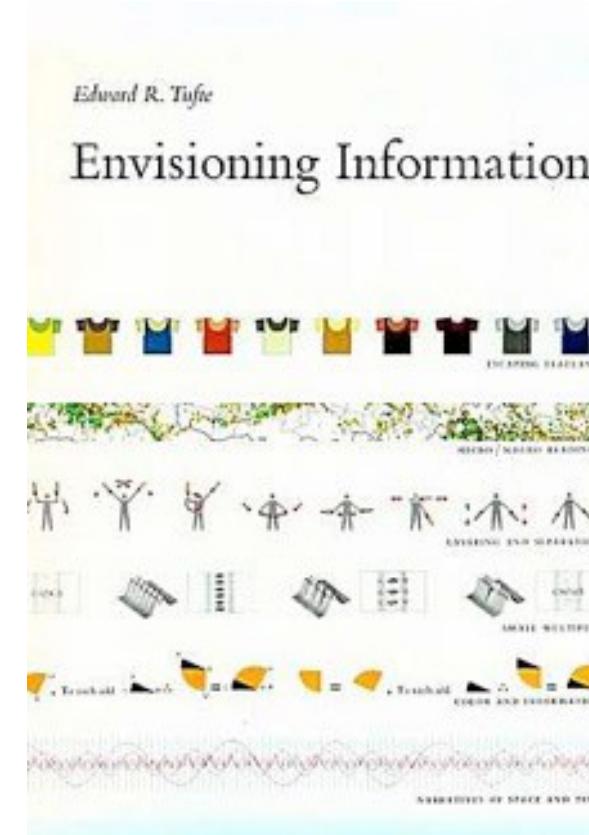
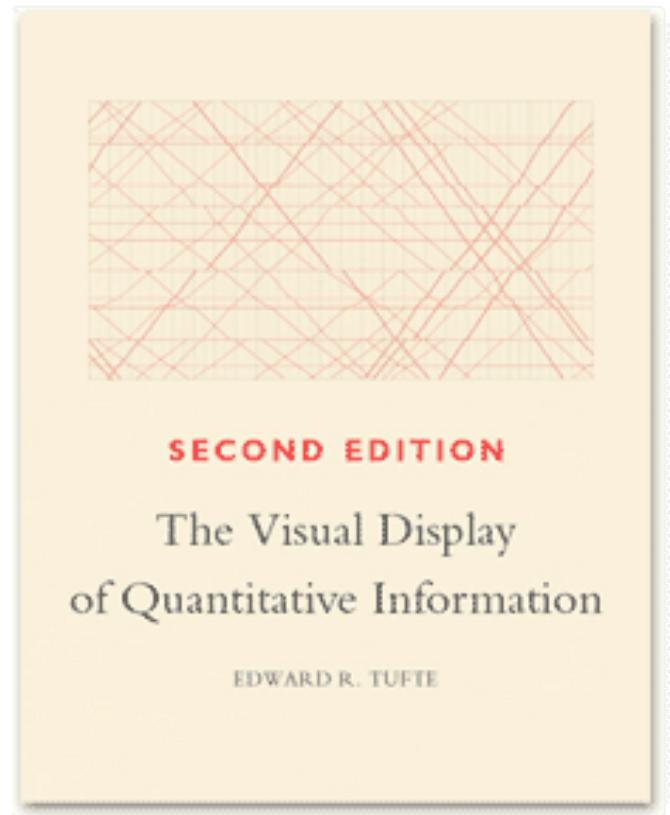
Interaction techniques: Designing multiple coordinated views with brushing and linking, navigation through large datasets (overview + context | overview, filter, details on demand).

Leading trends and advanced topics in Data Visualization

Books



More books on VIZ



www.edwardtufte.com

Prerequisites

- No general pre-req
- COMPUTER GRAPHICS Course (Preferred)
- Programming experience
- Javascript, python, html, css, etc.

Willingness to learn new software & tools

- This can be time consuming

Course Goals

- **Evaluate** and **critique** visualization designs
- **Implement** interactive data visualizations
- **Apply** fundamental principles & techniques
- **Develop** a substantial visualization project

How are these defined?

Computer Graphics

Computer Vision

Image Processing

Data Visualization

Visual Computing

Why create visualization

Answer questions (or discover them)

Make decisions

See data in context

Expand memory

Support graphical calculation

Find patterns

Present argument or tell a story

Inspire

- **Record** information

-
- Blueprints, photographs, seismographs,
...

- **Analyze** data to support reasoning

- Develop and assess hypotheses
(visual *exploration*)
- Find patterns and discover errors in
data

- **Communicate** information to others

- Share and persuade (visual
explanation)

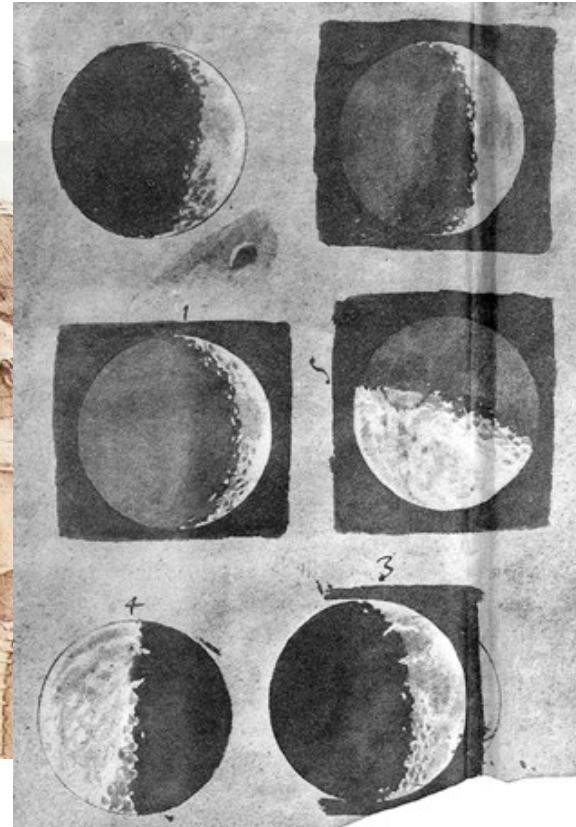
Use cases of visualization usage

Historical Examples

Record



Leonardo DaVinci, ca. 1500

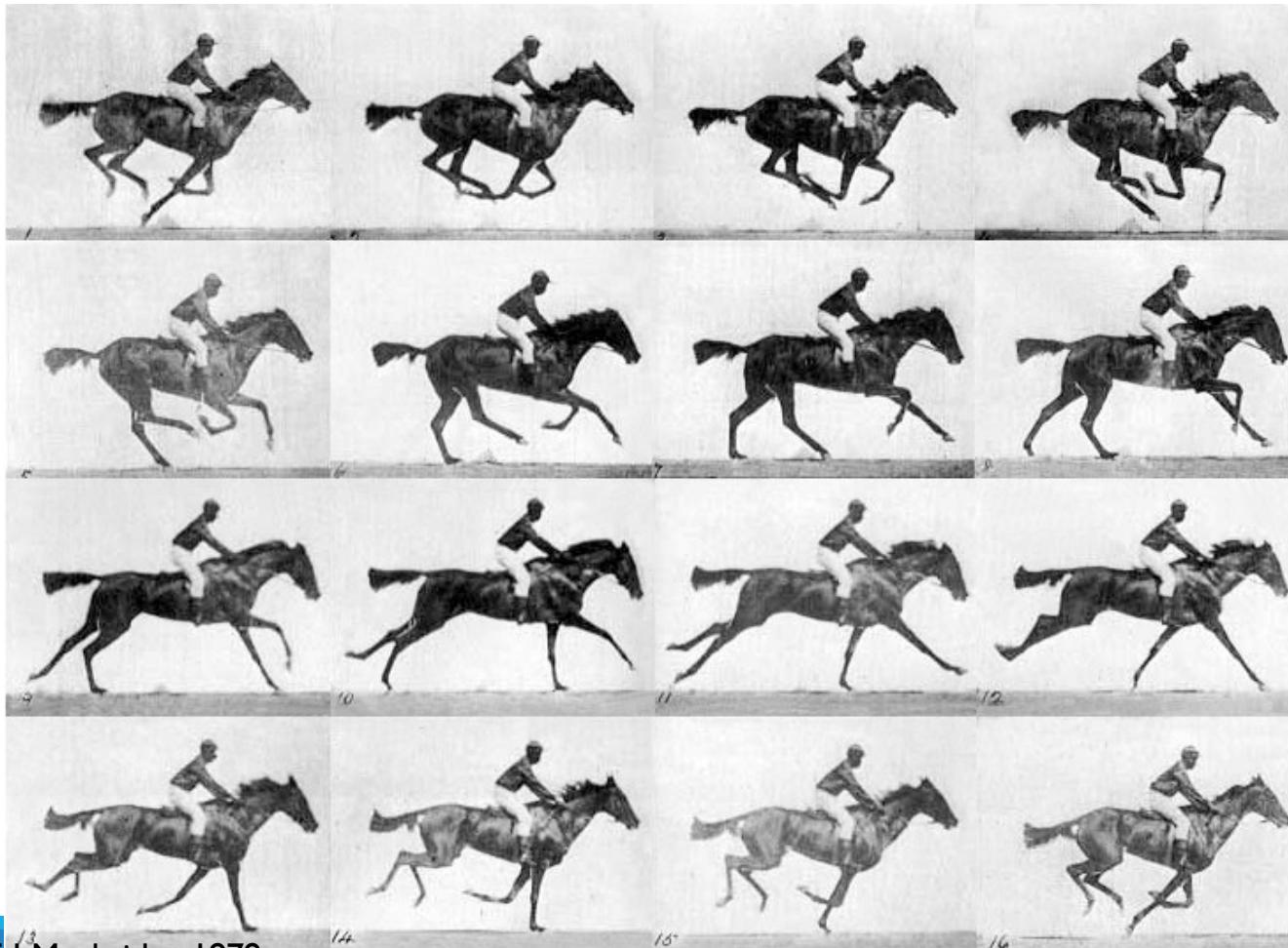


Galileo Galilei, 1616



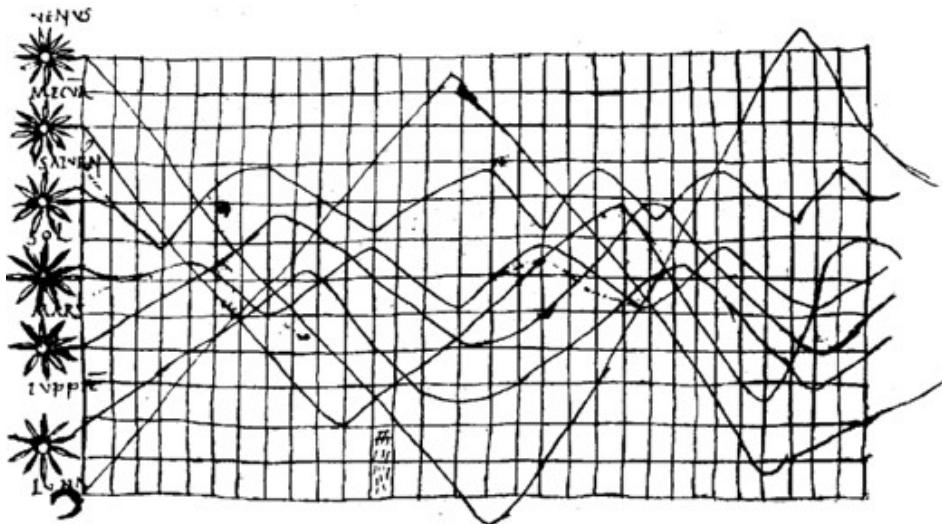
William Curtis (1746-1799)

Record

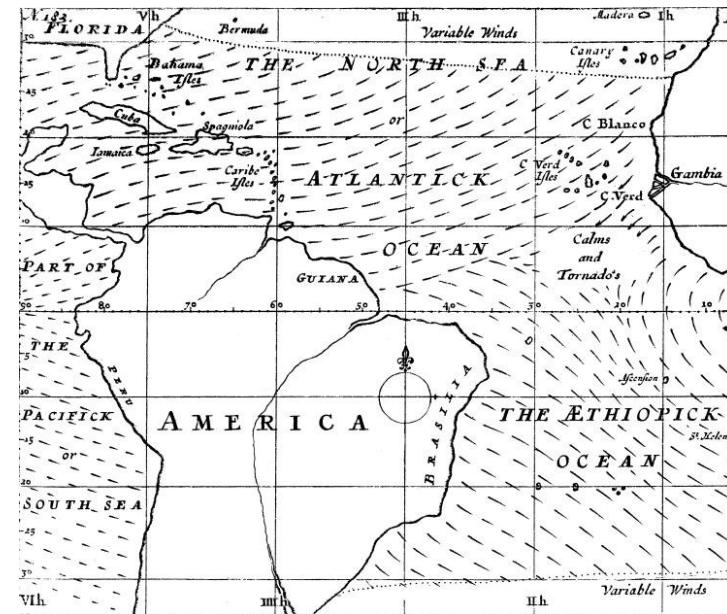


E.J. Muybridge, 1878

Analyze

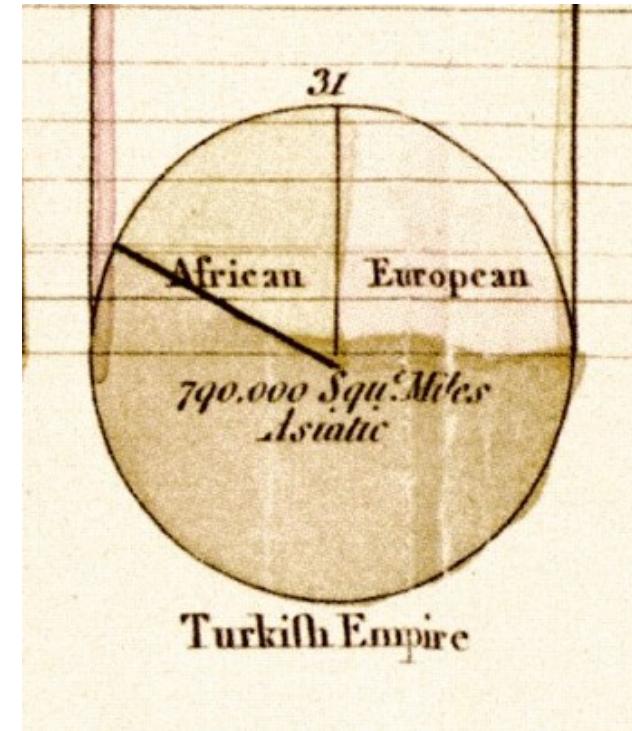
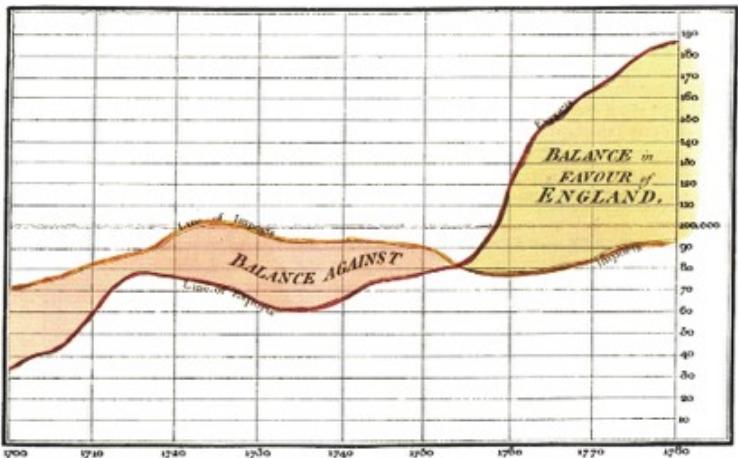
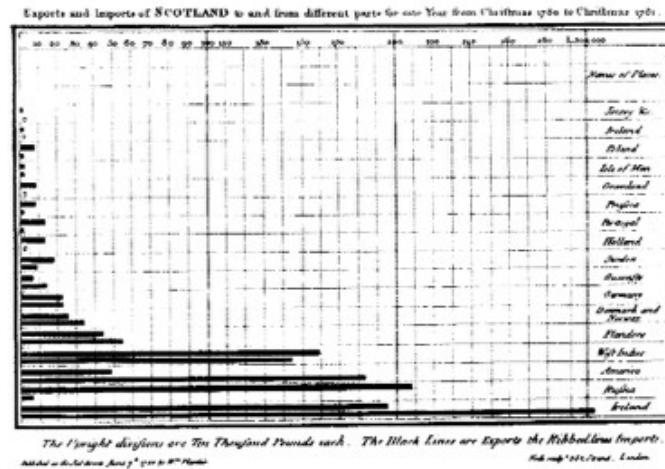


Planetary Movement Diagram, c. 950



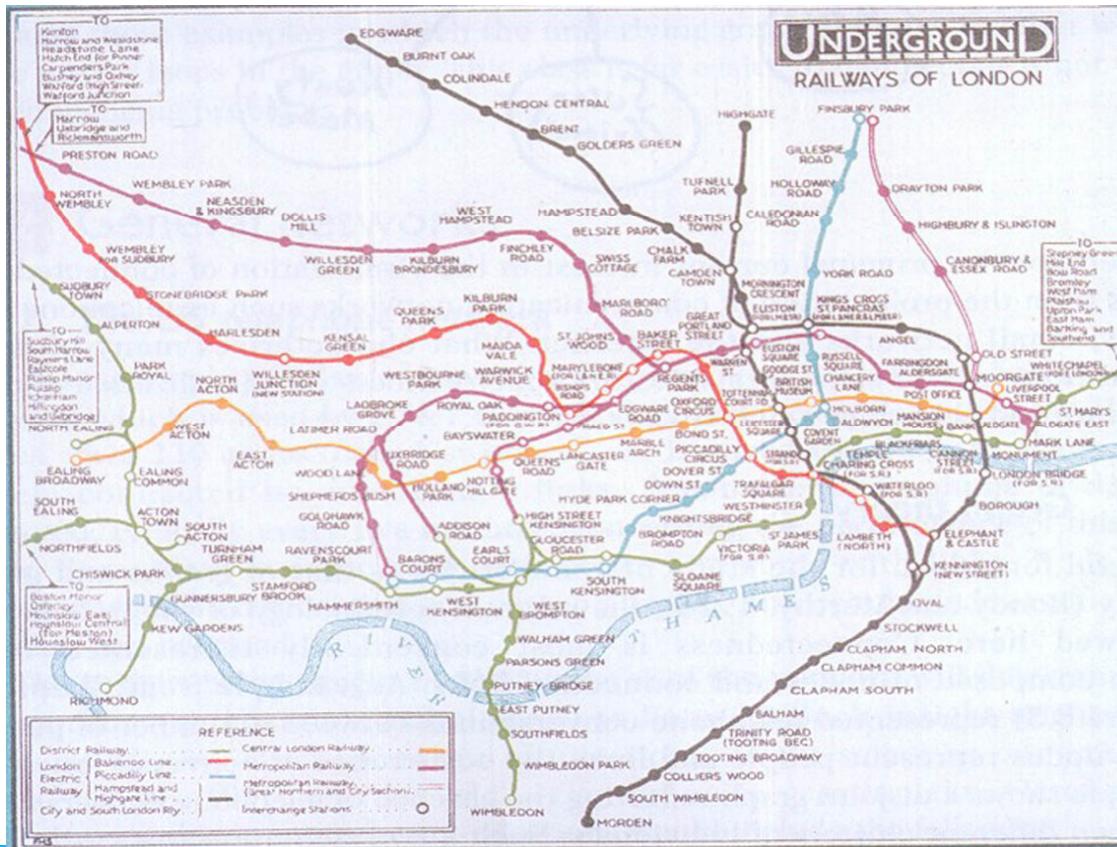
Halley's Wind Map, 1686

Analyze



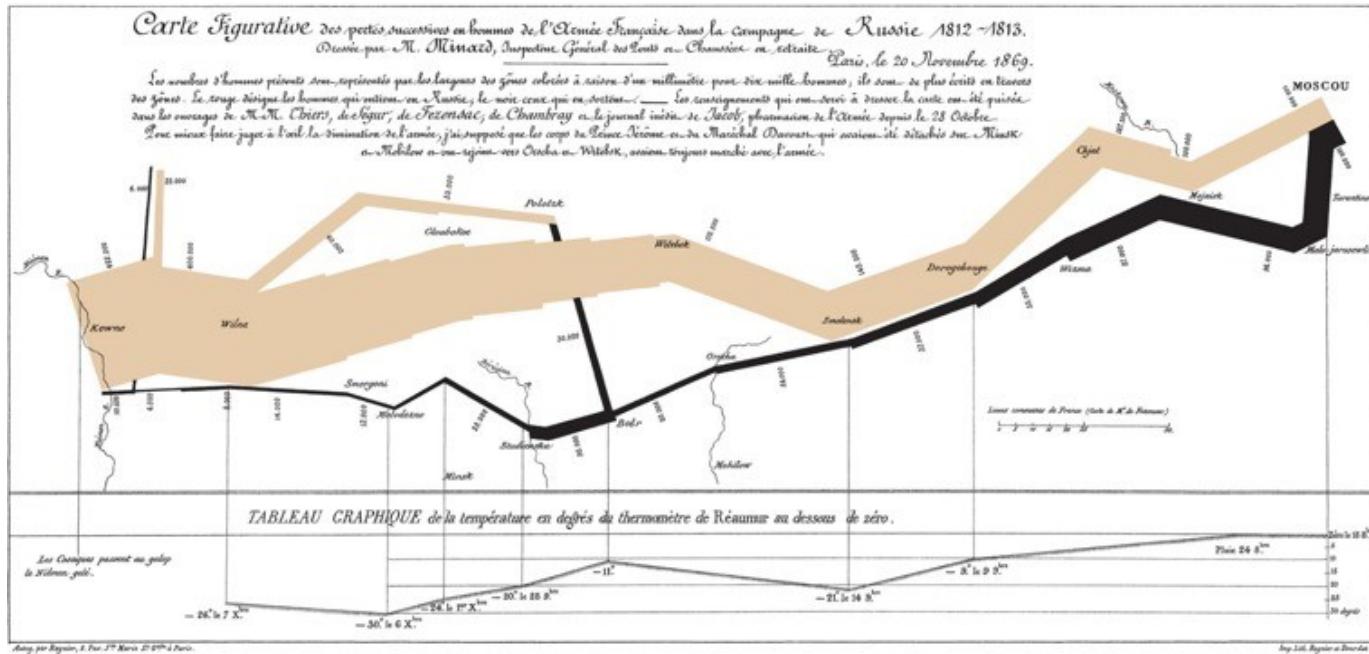
W. Playfair, 1801

Communicate



London Subway Map, 1927

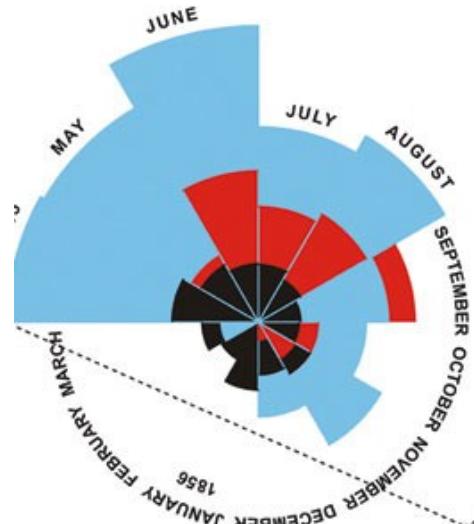
“A good sketch is better than a long speech” (Napoleon)



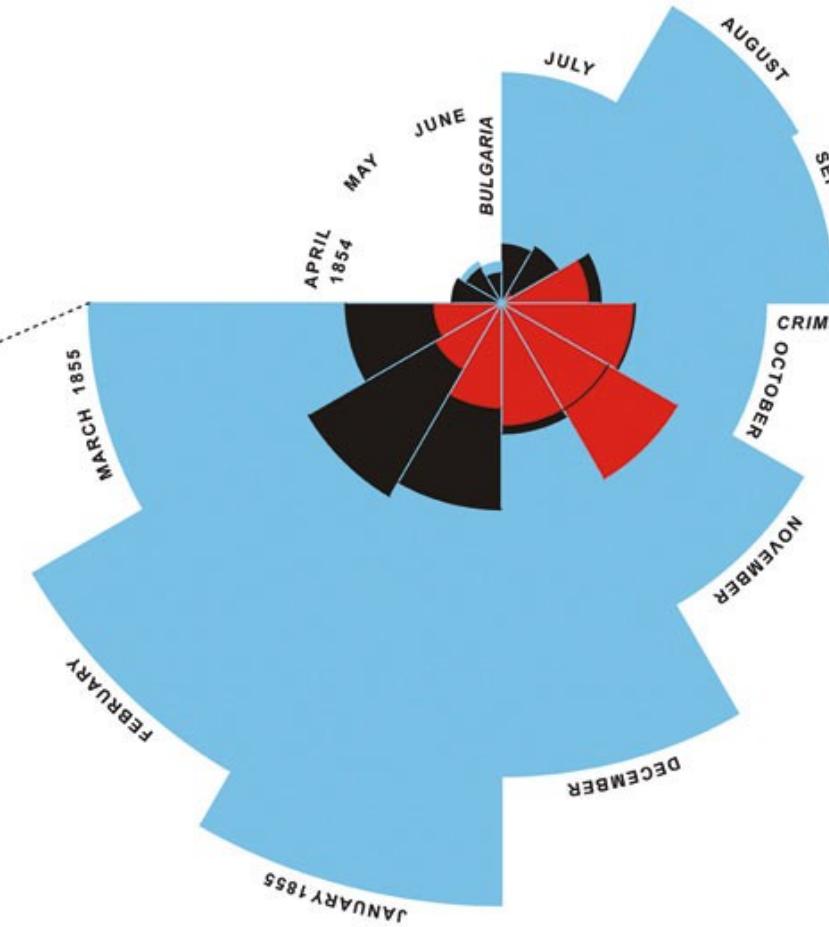
shows: size of the army, geographical coordinates, direction that the army was traveling, location of the army with respect to certain dates, temperature along the path of the retreat

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

2.
APRIL 1855 TO MARCH 1856.



1.
APRIL 1854 TO MARCH 1855



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

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

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

In October 1854, & April 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black

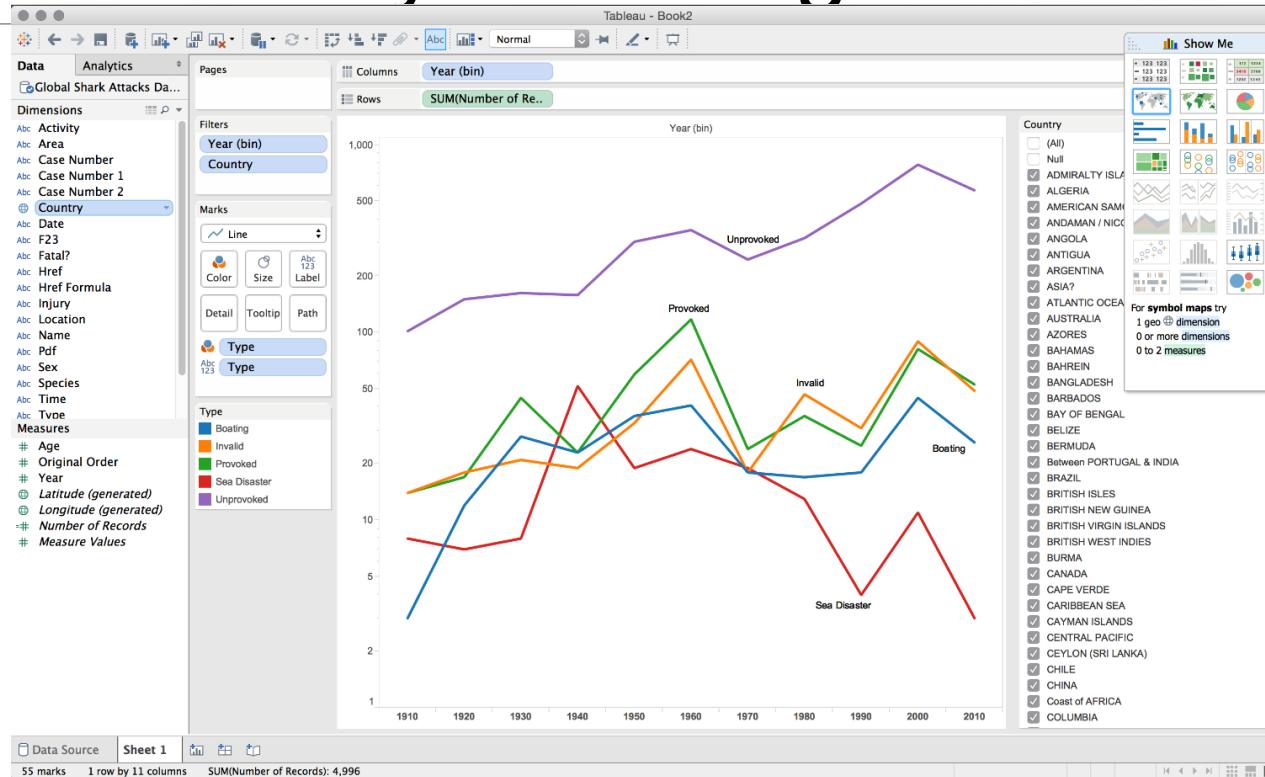
The entire areas may be compared by following the blue, the red & the black lines enclosing them. ©hugh-small.co.uk

Created by Florence Nightingale, 1858. Image from <http://www.economist.com/node/10278643> online.

Use cases of visualization usage

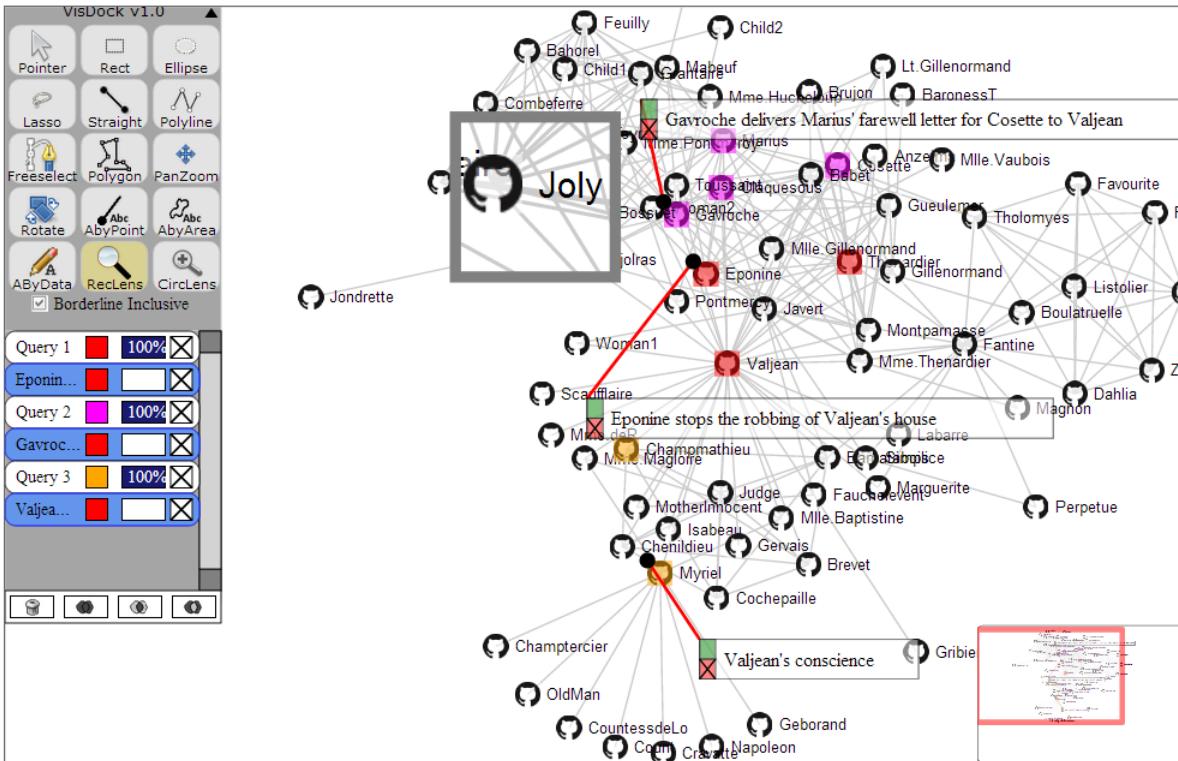
Recent Examples

Visualize data by choosing visual encodings



- Visualization palette (upper right) in the Tableau Desktop application for a dataset of shark attacks
- The “show me” feature in the tool (Mackinlay, 2007) will automatically highlight the suitable charts that can be used for the selected data

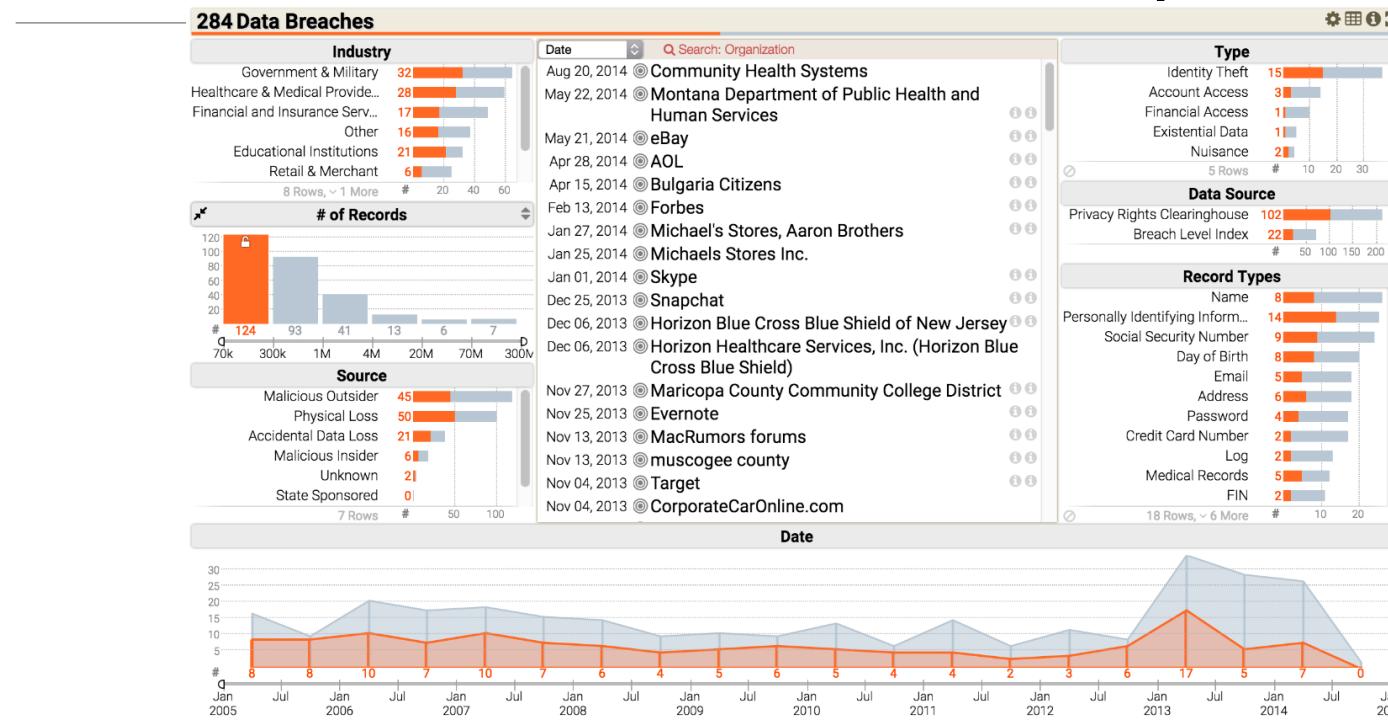
Select items to highlight, filter, or manipulate



Selection tools and data-aware annotations in an interactive node-link diagram representation of the social network for all of the characters in Victor Hugo's *Les Misérables*

- Characters are linked together if they appear in the same chapter in the book
- The textual annotations are connected to nodes using red lines and stay connected as the graph layout changes
- The toolbar on the upper left is part of the VisDock toolkit, and provides tools for annotation, navigation, and selection (Choi, 2015)

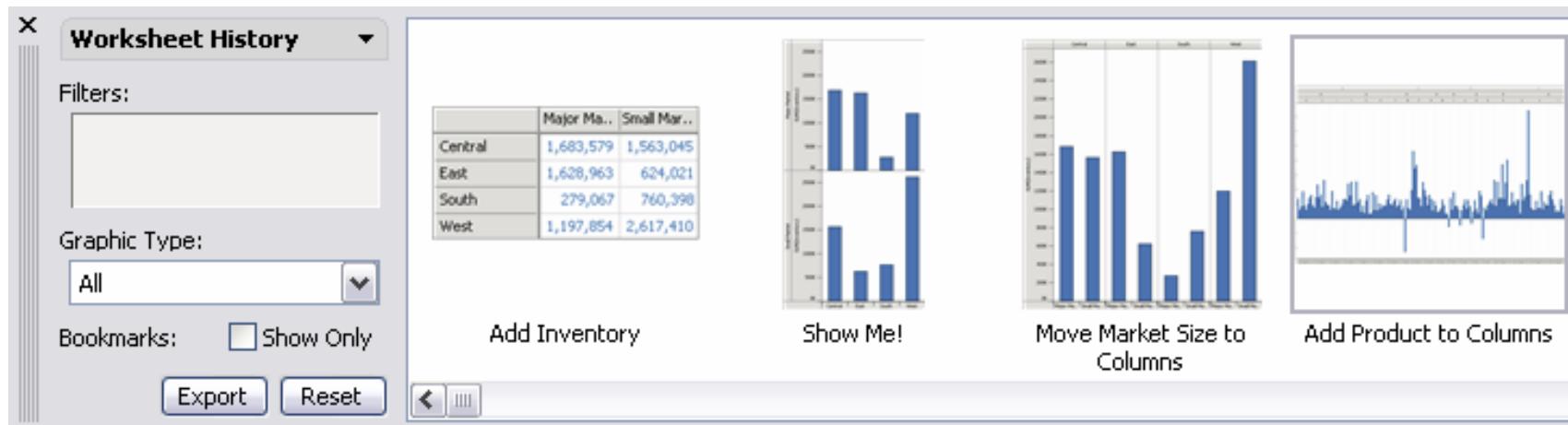
Coordinate views for linked exploration



Exploring 284 data breaches in the United States using the Keshif tool (<http://keshif.me/>), a multi-view visualization tool that shows different aspects of the data in separate views (Yalcin, 2016)

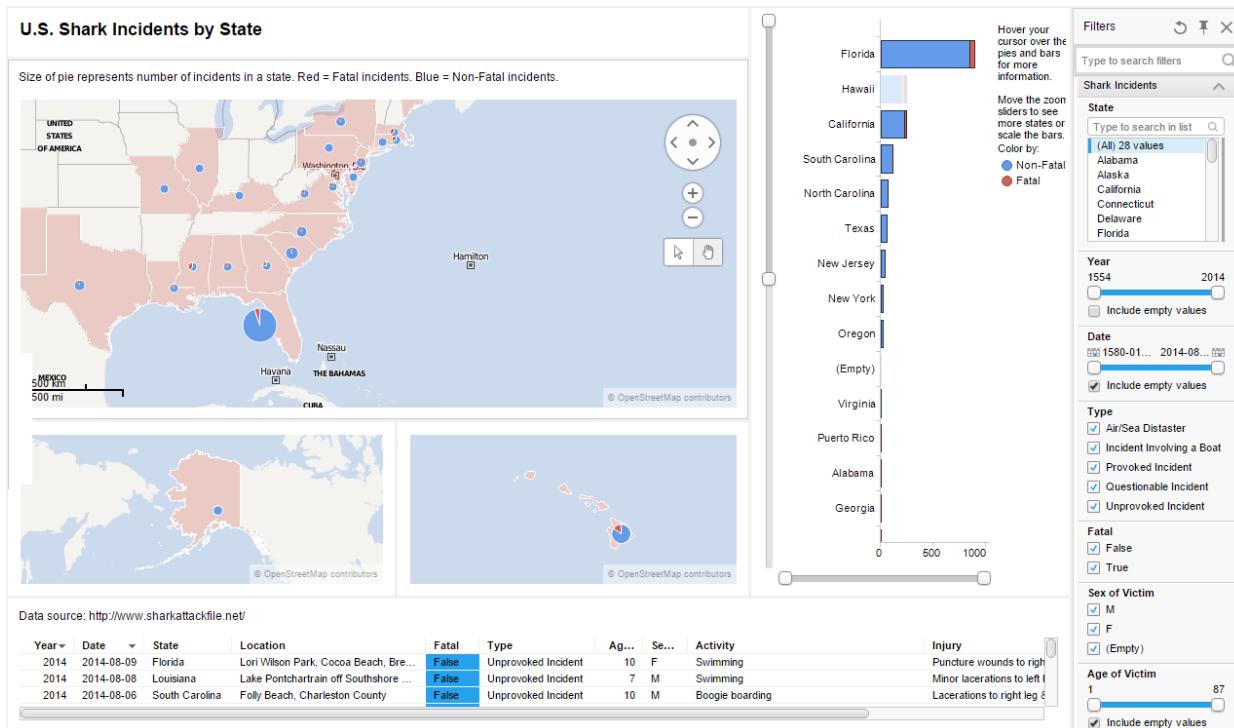
Selecting items in one view highlights them in others; for example, the user is currently hovering over the bar for “70k-300k” in the view titled “# of Records”, which causes those breaches (i.e. 124 breaches that exposed in the range of 70,000 to 300,000 records) to also be highlighted in orange in other views, including in the timeline at the bottom

Record analysis histories for re-visitation, review, and sharing



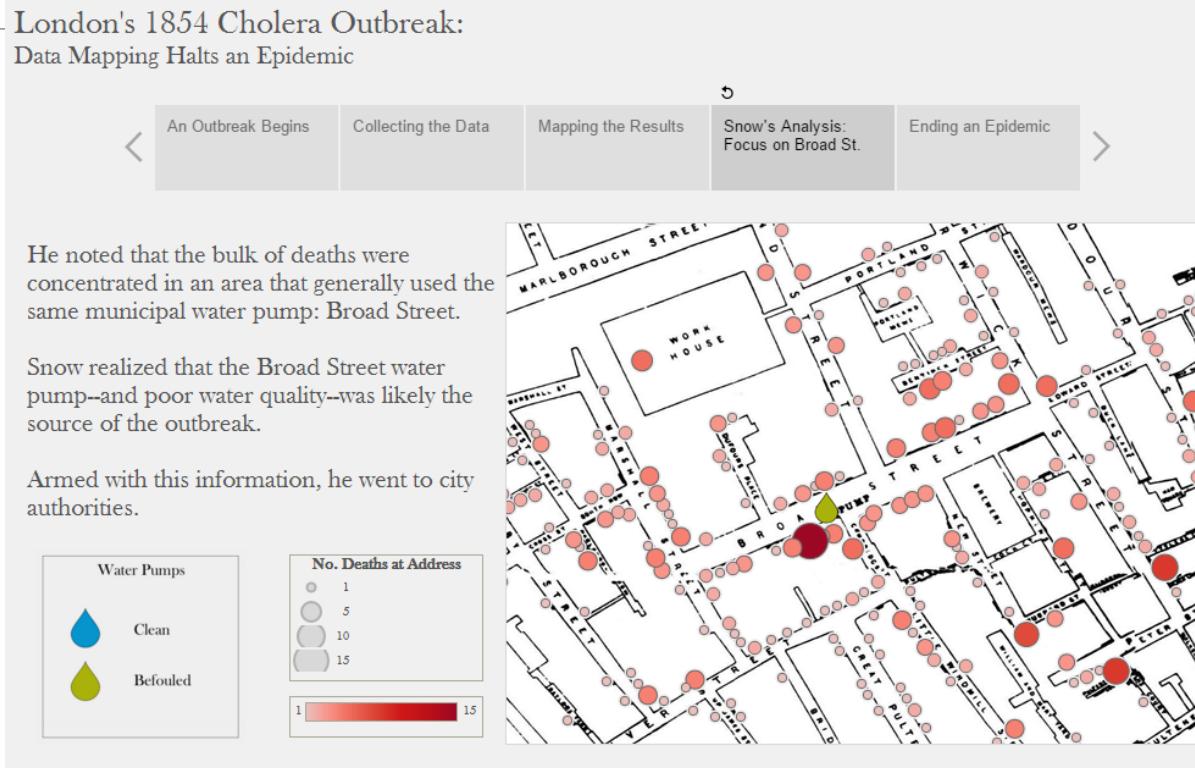
- Graphical history interface using thumbnails of previous visualization states organized in a comic-strip layout (Heer, 2008)
 - The labels describe the actions performed

Share views and annotations to enable collaboration



- Spotfire visualization dashboard of shark attacks published on the web
 - The dashboard can be interacted, causing views to update dynamically
 - The tool also allows for application bookmarking (storing the state for specific insights) as well as sharing the analysis on social media platforms such as Facebook, Twitter, and LinkedIn

Guide users through analysis tasks or stories



- Web-based visualization of London's 1854 cholera outbreak showing physician Jon Snow's use of visualization to find its source
 - This visualization was created in Tableau using its Story Points feature that allows users to build a narrative from data
 - The horizontal list of five boxes at the top of the display are the main points in the story, and viewers can be automatically guided through the story by moving to each point from left to right

A story from past

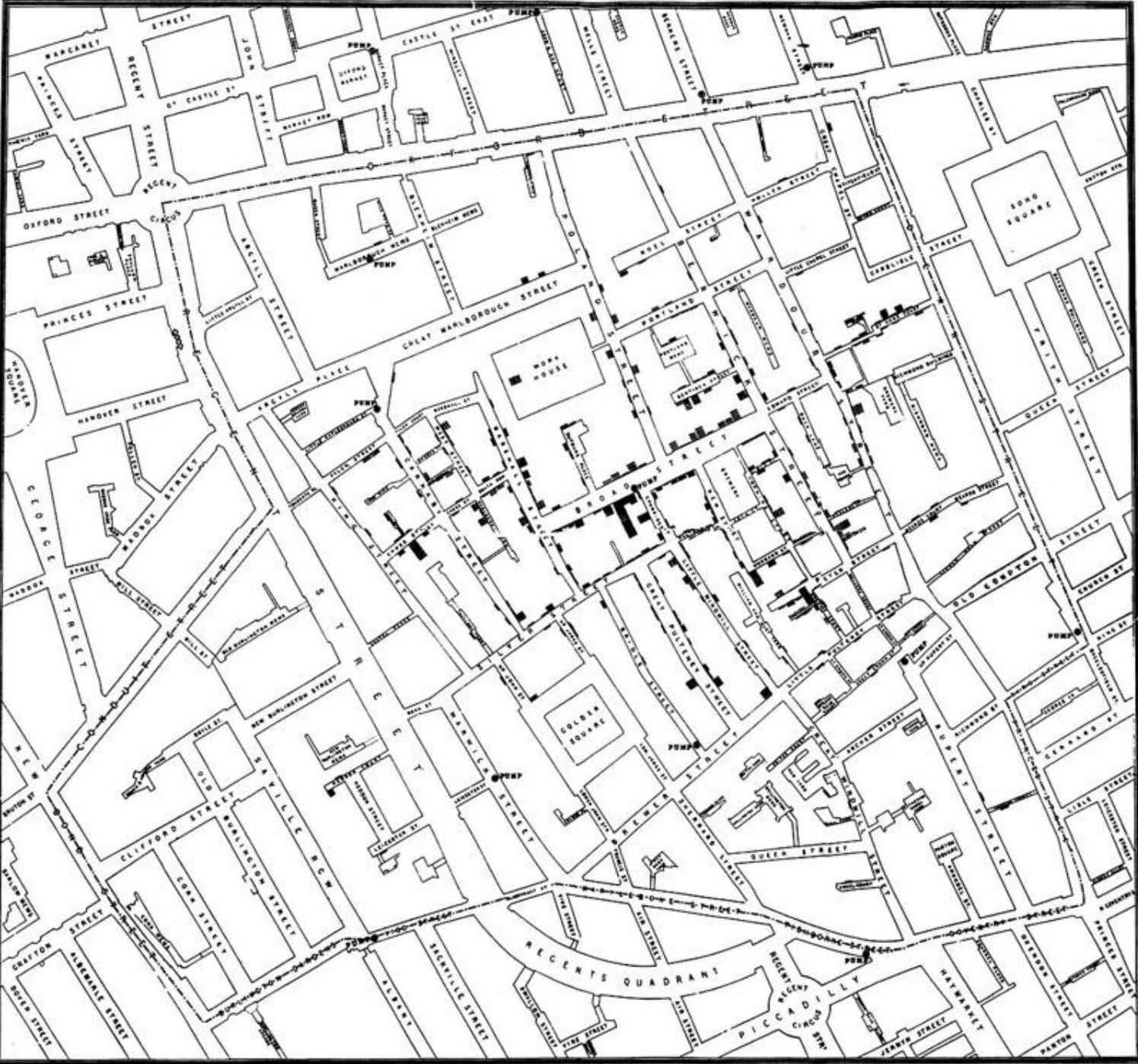
Emphasising use of *visualization in analysis*

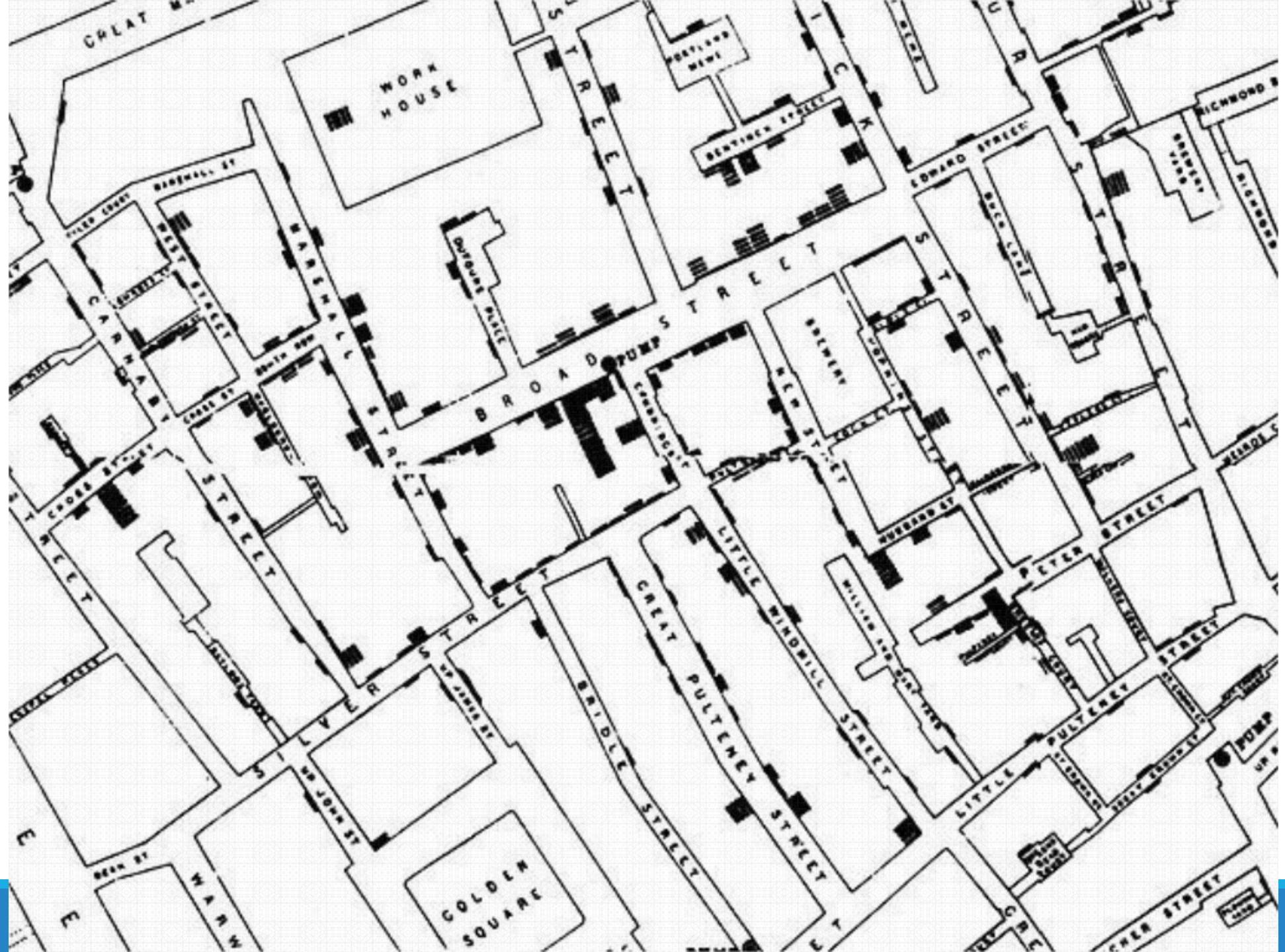
Cholera Epidemic in London, 1854

- >500 fatal attacks of cholera in 10 days
 - Concentrated in Broad Street area of London
 - Many died in a few hours
- Dominant theory of disease: caused by noxious odors
- Afflicted streets deserted by >75% inhabitants

John Snow

- Set out to investigate **cause**
- Suspected it might be due to water from community **pump**
- Tested water —> no obvious impurities
- What more evidence could there be?
 - Could list of 83 deaths, plotted on map





Investigation and aftermath

- Based on **visualization**, did case by case investigation
- Found that **61 / 83** positive identified as using well water from Broad Street pump
- Board ordered pump-handle to be removed from well
- Epidemic soon **ended**
- Solved centuries old question of how cholera spread

Methods used by Snow

- Placed data in appropriate **context** for assessing cause & effect

- Plotted on map, included well location
- Reveals proximity as cause
- Made quantitative **comparisons**
 - Fewer deaths closer to brewery, could investigate cause
- Considered **alternative** explanations & contrary cases
 - Investigated cases not close to pump, often found connection to pump
- Assessment of possible **errors** in numbers

Questions ?

Questions?

“He who asks a question is a fool for five minutes; he who does not ask a question remains a fool forever”

Chinese Proverb

“The wise man doesn't give the right answers, he poses the right questions.”

Claude Levi-Strauss

“A wise man can learn more from a foolish question than a fool can learn from a wise answer.”

Bruce Lee