Introduction to Data Visualization

The Graduate Center at CUNY

Michelle A. McSweeney

About Me

- Michelle McSweeney
  - mmcsweeney@gc.cuny.edu
  - Senior Data Science Curriculum Developer, Codecademy
  - Linguistics PhD
  - Chemistry & Math BA
  - Books
    - Data
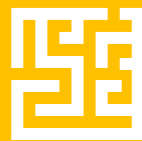   - OK
    - Pragmatics of Text Messages

Introduction

Name

Why you are taking this course/program

Data you are interested in OR things you like to learn about (that is NOT data analysis/viz)

Overview of Class Conceptually

DATA   VISUALIZATION   DESIGN   SEMINAR/STUDIO

Overview of Class Practically

- Projects
- Reflections
- Pin Ups
- Portfolio
- White Paper

What is Data Visualization?

Brief History of Data Visualization

A very Euro-centric retelling that needs to be interrogated

Data | Dedomenai

- 300 B.C. in Euclid's Dedomenai (Δεδομεναι)
- Dedomenai means "that which is given", or—more concisely: data.
- Ptolemaic Egypt, the library of Alexandria, focus on epistemology

1600-1699 | Measurement

- Early 1600's, tables of measurements appear in Western Europe. Arranging observations into a table is a technology unto itself and still the basic way we think about "data."

  - Today we might color code such a table and call it a heat map.

- 1670 Gabriel Mouton in France proposes a method for measuring length at the intersection of science and nature.

- His construct has divisible units where the basic unit is related to the size of the Earth.
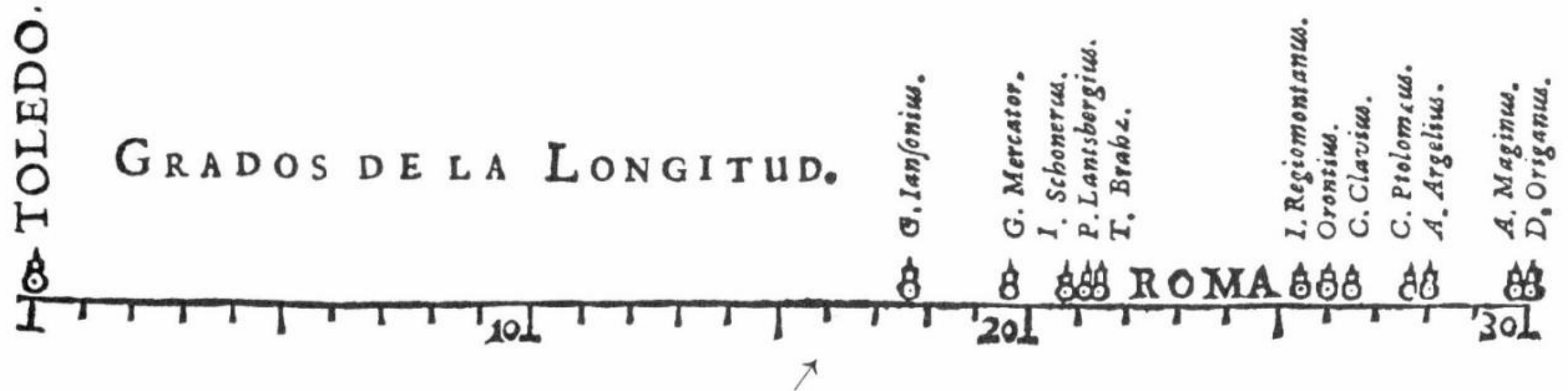
The Longitude Problem

- Latitude is easy: the equator is a nice center point, a starting point, or a zero.

- Longitude is harder.
    - There is no zero point to start from (we've since picked Greenwich, England)
    - It is impossible to use celestial bodies such as the sun without being able to tell time.

- As with many hard problems, it took a long time to solve for it, but the solving for it changed the world.
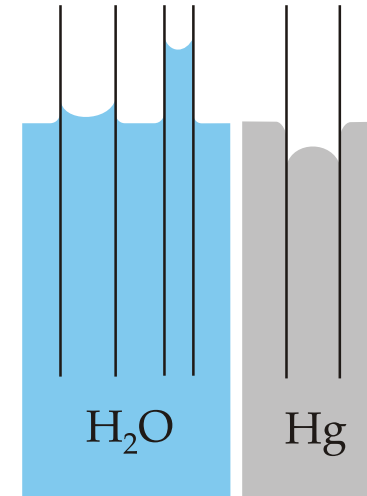
1644, Michael Florent van Langren
Difference in Longitude between Toledo and Rome

This is a visualization, it's also the start of modern statistics: quantifying uncertainty.
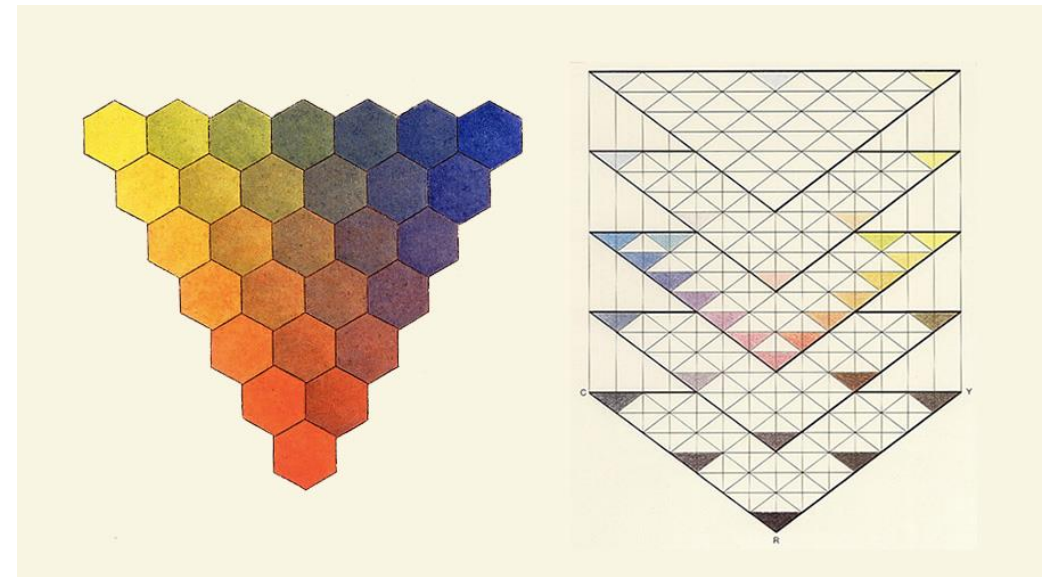
1700 - 1799 | Representation

- 1712 Literal line chart
  - Francis Huckabee
  - He was researching capillary action between plates and represented how high the water reached between different plates of glass.

- 1724 Abstract line chart
  - Nicolaus Samuel Cruquius
  - Barometric observations

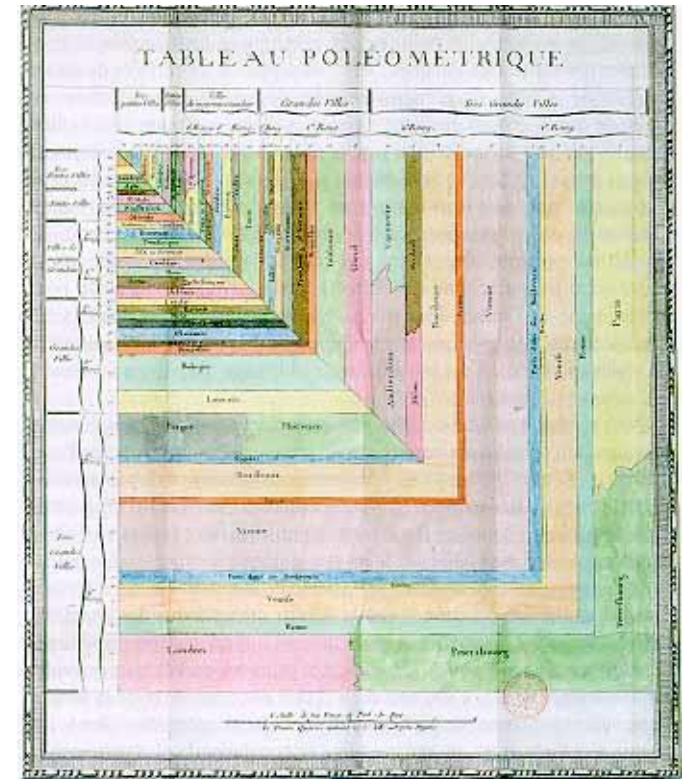Capillary action of H2O vs Hg between plates of glass.

H_2O Hg

1700-1799 | Representation

- Jacques Barbeu-Dubourg's timeline "Carte chronographique"
  - 54 foot long timeline of important world events & people (from the POV of an 18th century French man)
  - Invented synchronoptic view (multiple events at the same time)
- Lots of color systems (Mayer, Harris, Lambert). Quantifying color is abstract, and this gives us color scales, chloropleths, etc.

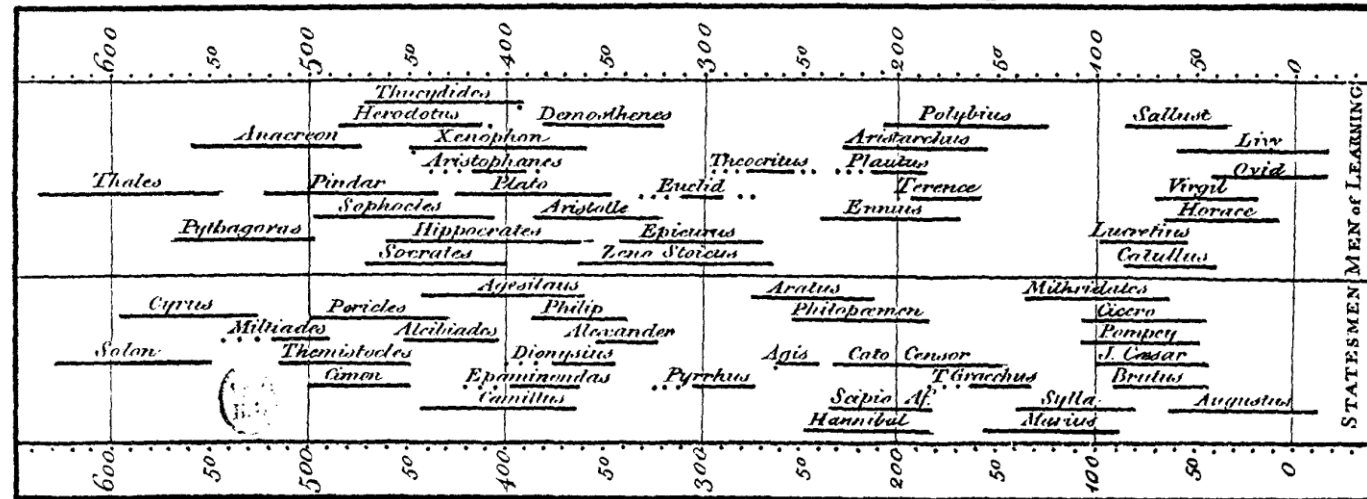# 1775-1850 | Visualization & Storytelling

- 1782 de Fourcroy's proportional squares (tableau graphique)

- 1798 Seaman – yellow fever maps of dots & circles – foundations for demographic maps

- 1819 Dupin chloropleth of literacy in France

- 1820's Faraday includes diagrams in his scientific publications

- 1829 Guerry polar area charts

- Visualization explosion (ideas generate ideas) – but 2 contributors stand out: Priestley & Playfair
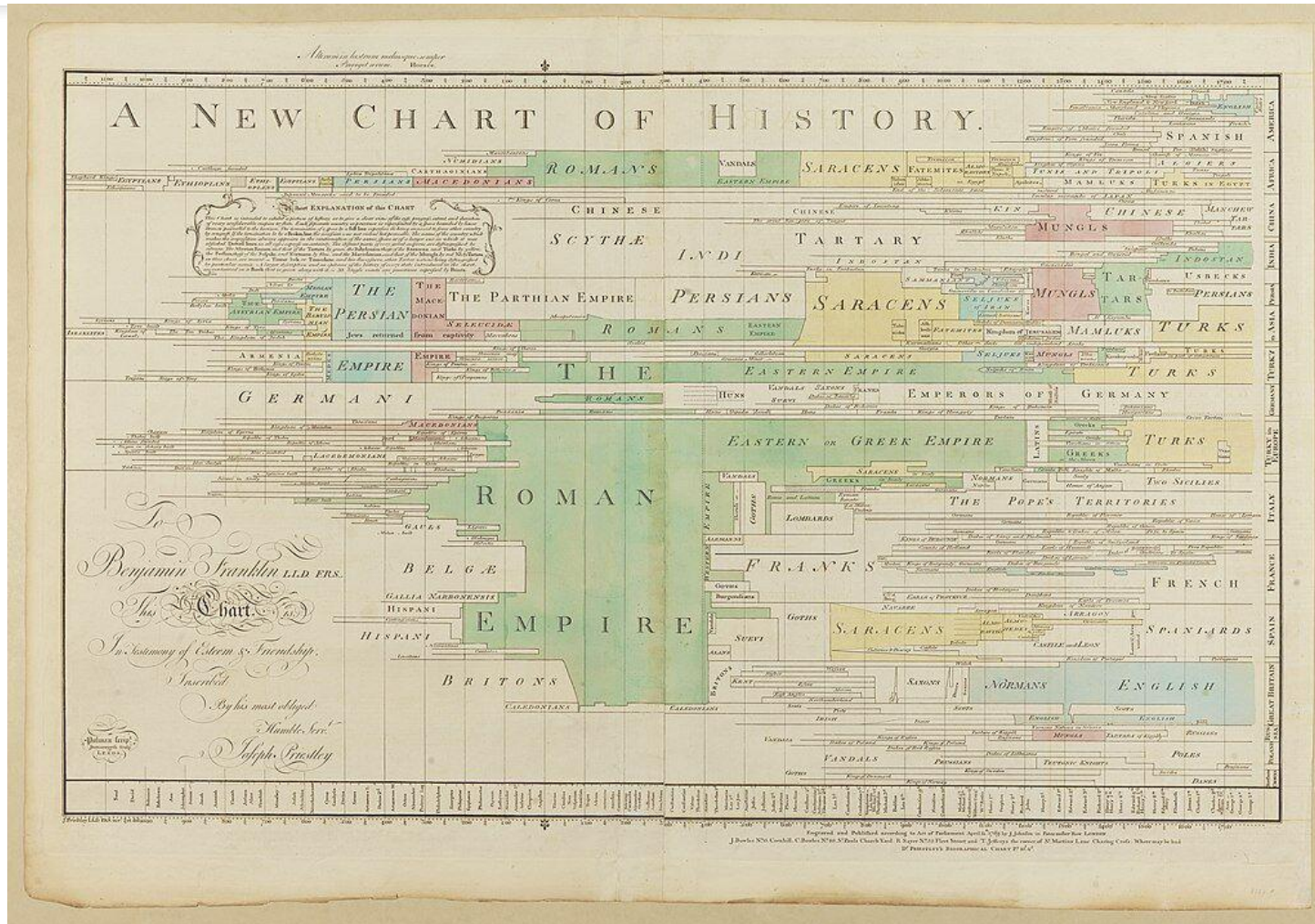
1760's | Joseph Priestley

- Priestley shaped how we think about data and visualization by abstracting data from its measurement.
- Chart of biography (notably biased) quantified lifespan using bars. This was a big deal. Time represented as discrete length was new.
- 1765 A Chart of Biography

A Specimen of a Chart of Biography.

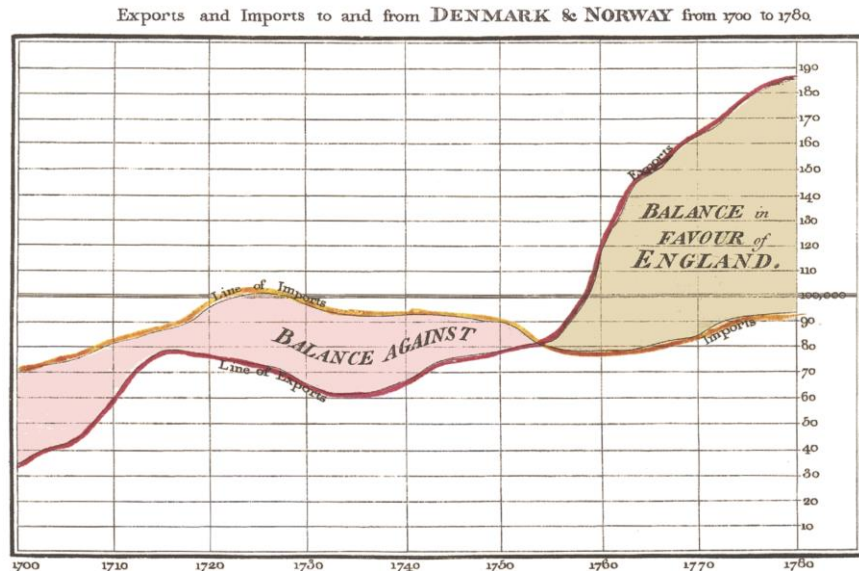# 1760's | Joseph Priestley

1769 A New Chart of History

For as biased as this looks today, it was revolutionary for the time. His argument is that to understand history, you have to understand interactions, and any boundary (country, time, place) is artificial)

Region is on the y axis, time on the x.
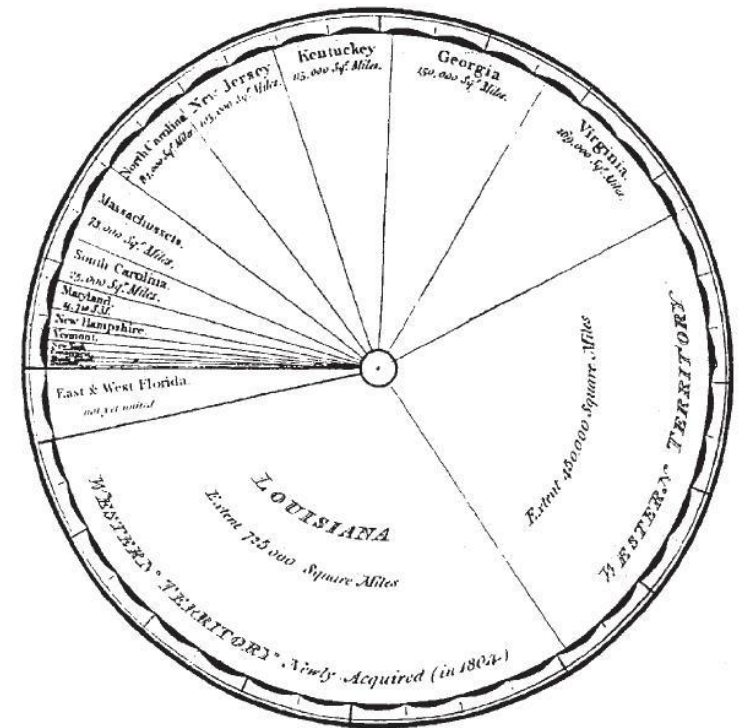
Golden Age of Data Visualization

1759 - 1823| William Playfair

- Bar chart, line chart, and area chart of economic data
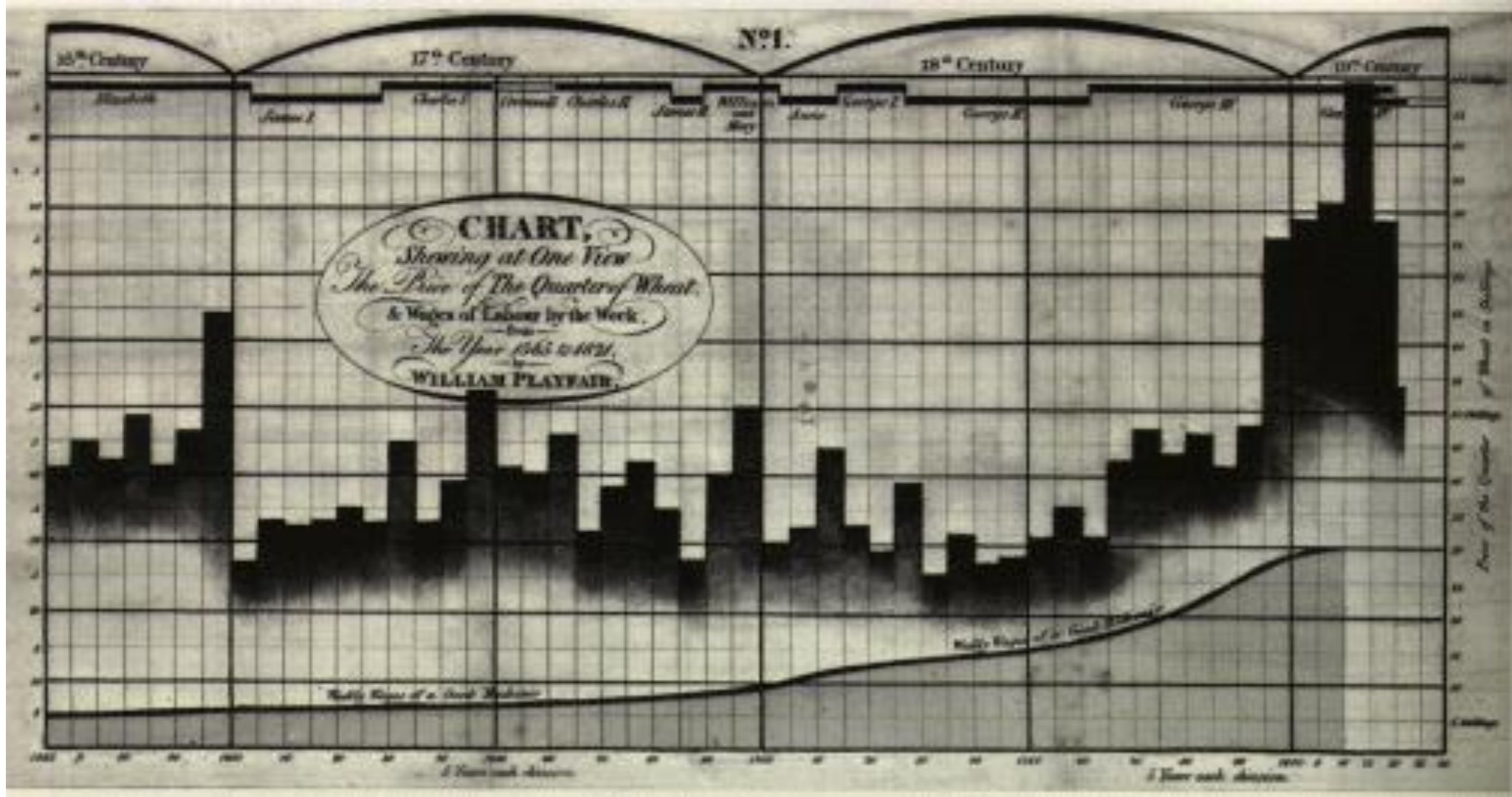- Pie chart & circle graph to show part-to-whole relationships

1786 time series line chart
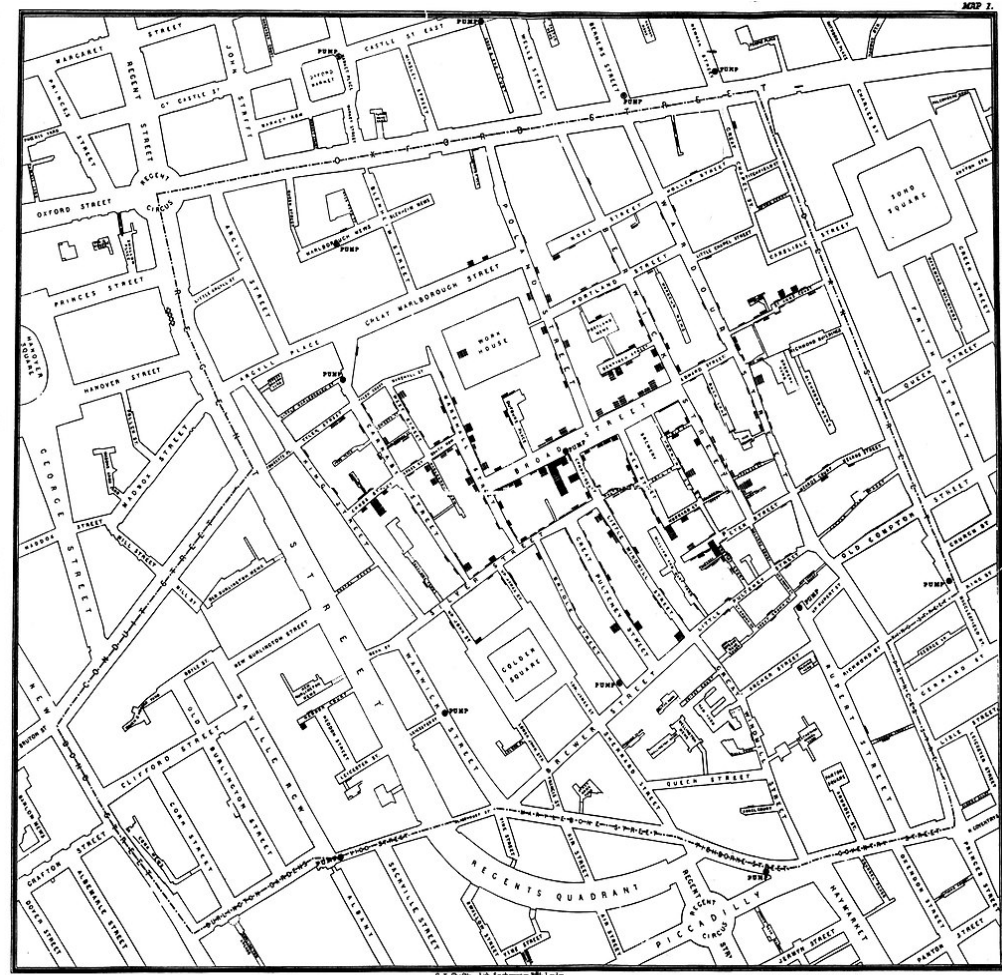
1801 area representation by state

1821, Playfair Price of wheat, weekly wages, and reigning monarch from 1565 to 1820

1800-1899 | New takes on standard charts

- Evolution of statistics into modern statistics

- Focus on morality & mortality of people and society.

- Industrialization is making print more accessible and widespread urbanization in W. Europe, U.S. & Canada accelerates technology & communication.

- More print leads to greater transfer of ideas and greater collection of data and quick evolution of ideas

- Greater collection of data leads to move visualization techniques, and there are thousands of tiny inventions in analysis & visual display of data.
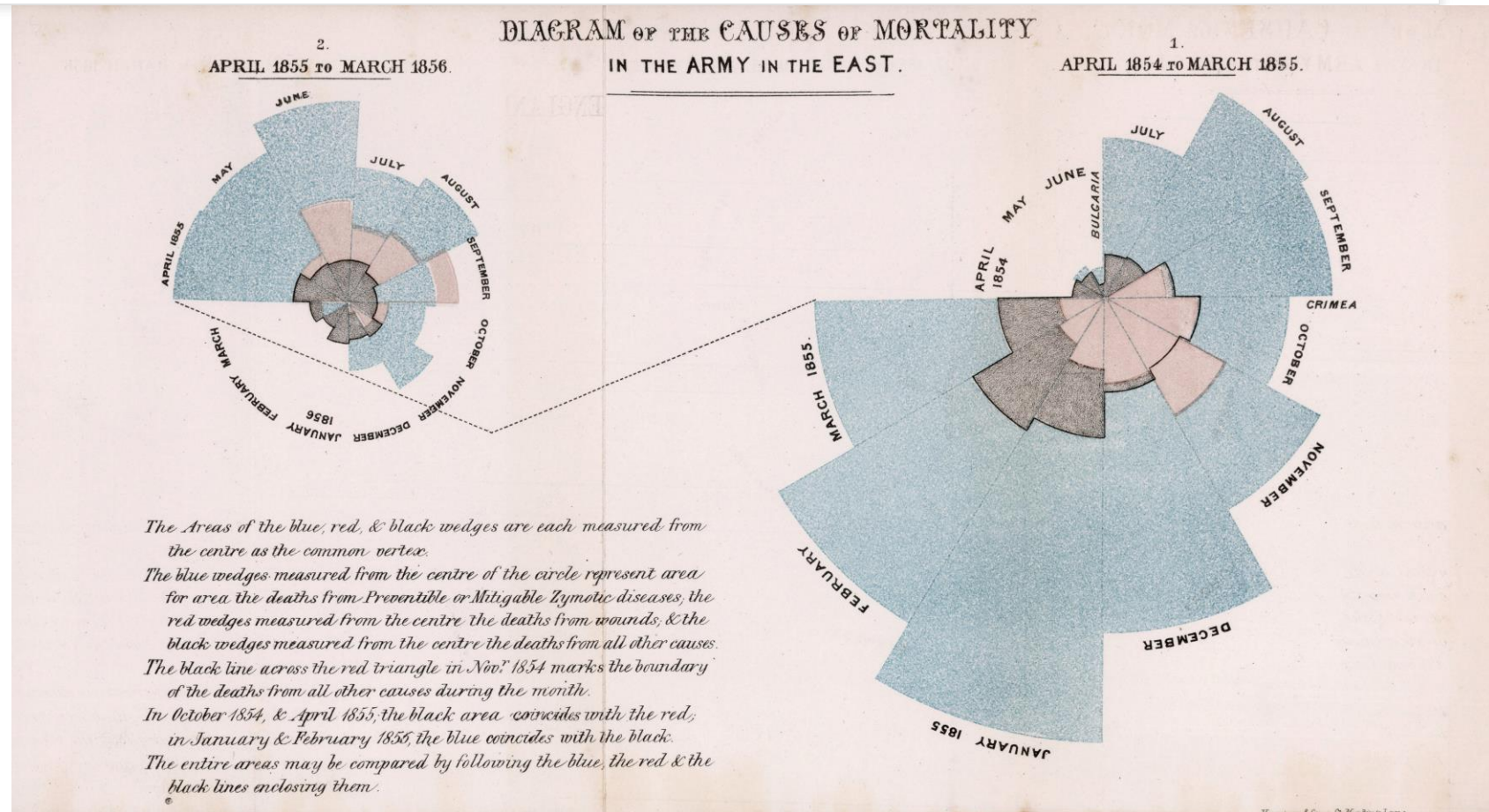
1854 | John Snow's Cholera Maps

- Major cholera outbreak in London

- John Snow mapped the cases using bar charts on a street map. Found the source of the outbreak (a well). Removed the handle and the outbreak ended.

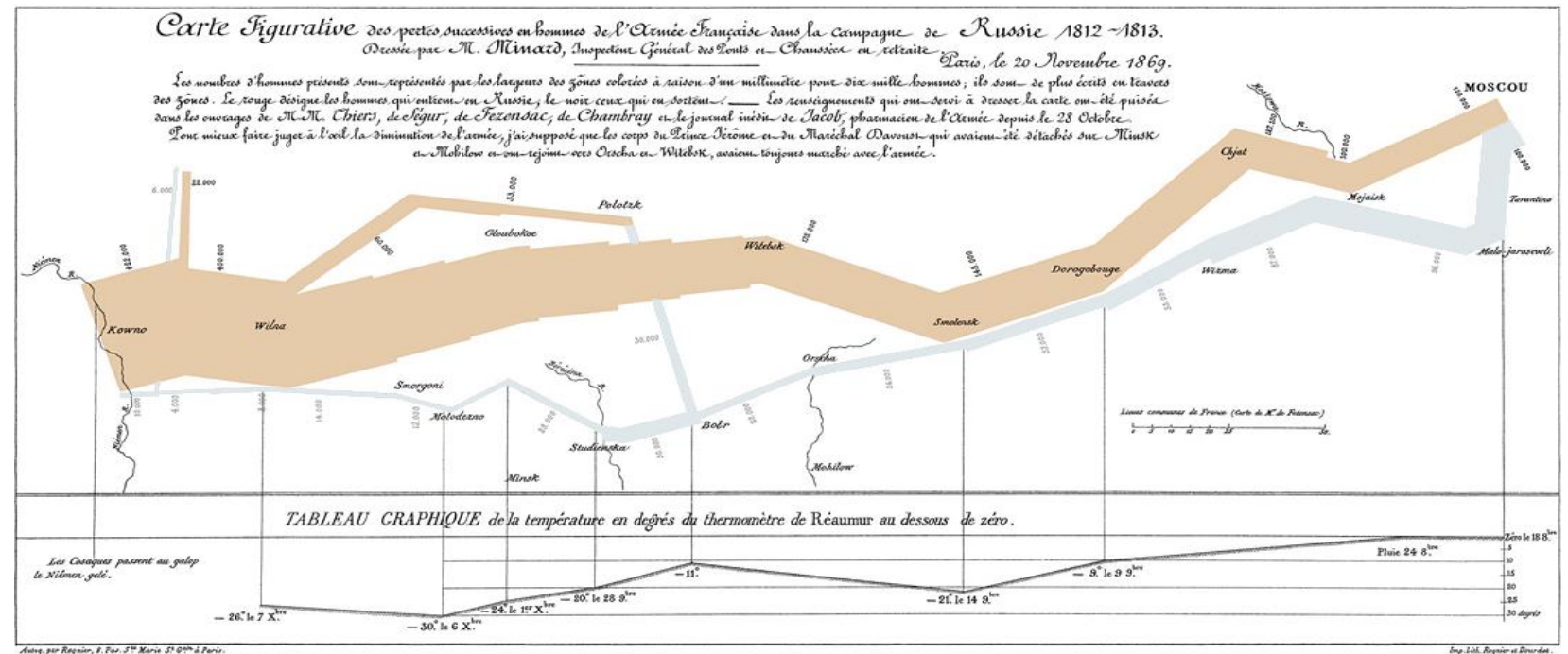- First epidemiological map – combination bar chart and map.

1858 | Florence Nightingale

- Argued for sanitary reform based on data.

- During Crimean war, she demonstrated that poor sanitation caused more loss of life than wounds and injuries.

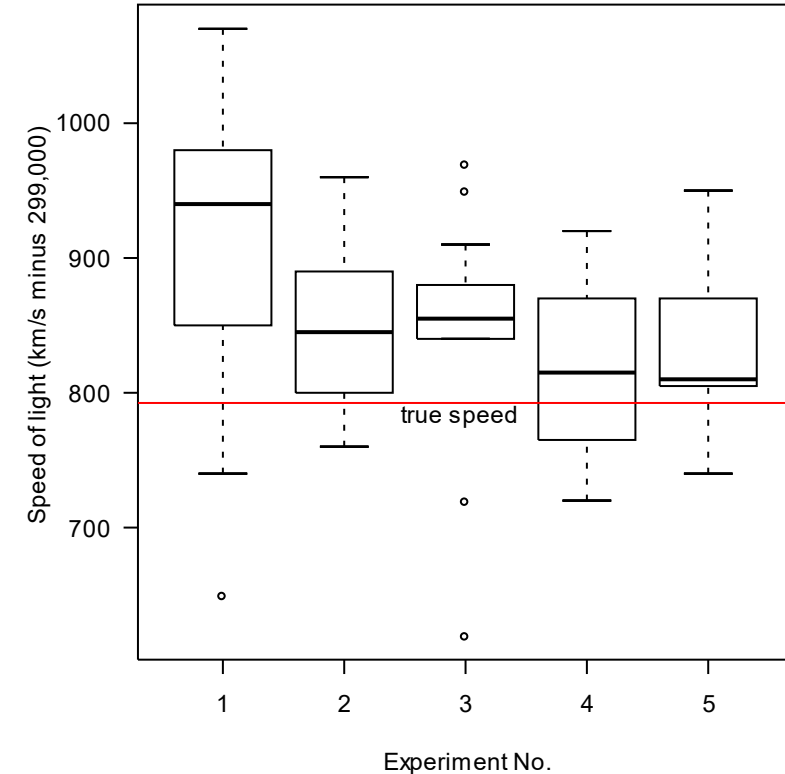- She was not wrong, but her visuals are not clear.

1869 | Charles Minard's Flow Chart

- First flow map
- SIX data types in one plane:
  - the number of Napoleon's troops;
  - the distance traveled;
  - temperature;
  - latitude and longitude;
  - direction of travel;
  - location relative to specific dates

1977 | Robert Tukey

- Visual representation of a variety of statistical information.
- In his book, *Exploratory Data Analysis*
- Box plot
  - Quantitative summary of a variable
    - Distribution
    - Mean
    - Quartiles
    - Outliers

Rebirth of Data Visualization

Rebirth of Data Visualization

- Edward Tufte
- Alberto Cairo
- Mike Bostock (D3)
- Aaron Koblin
- Amanda Cox
- David McCandless
- Ola Rosling
- Nathan Yau

- Eric Rodenbeck
- Fernanda Viégas
- George G. Robertson
- Hannah Fairfield
- Stephen Few
- koci hernandez
- Giorgia Lupi
- And so many more!

Creating a Dataset

- Your name

- Restaurant Recommendation

- Quick description

- Where it's located

Data Collection

- Google Form Survey