Lecture 6

Visualization

History of Data Science, Spring 2022 @ UC San Diego

Suraj Rampure

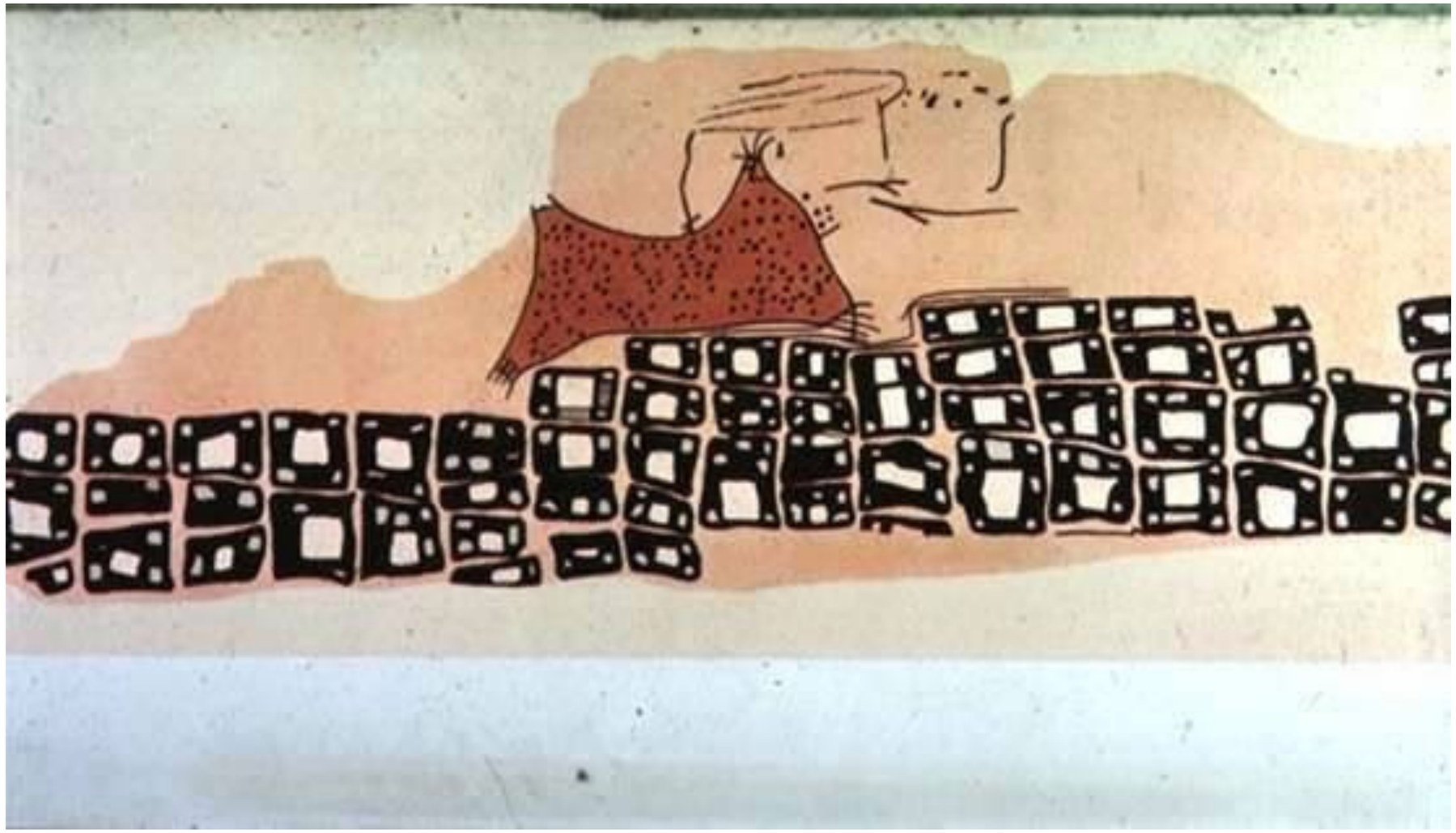
Announcements

- Homework 6 is released and is due on **Sunday, May 8th at 11:59PM**.
  - You'll get to make a website!
- Homework 4 is graded! Make sure to look at the solutions, posted on Slack and on the course website.
  - **Make sure to make an honest attempt on each question; you may not get credit for the homework if you do not.**
- Please try and attend in-person if you are able to!

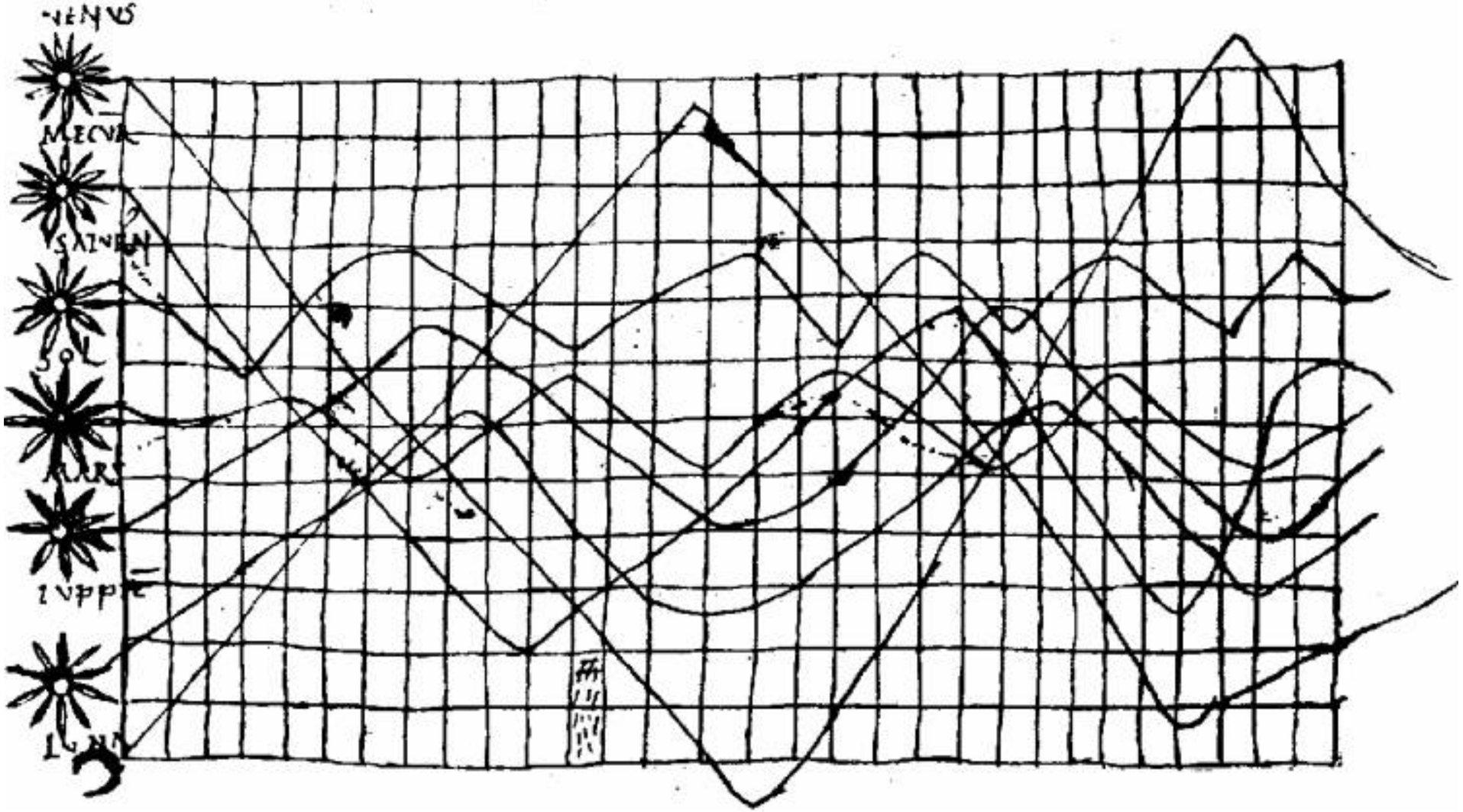
Agenda

- Today, we'll look at several examples of old visualizations.
- In addition, we will try to re-create some of these visualizations on our own in a Jupyter Notebook.
  - If the background of a slide is grey, it means that we'll re-create the visualization on that slide in the lecture notebook.
- Follow along!
- Last time, we didn't cover the derivation of the Gaussian distribution – we will revisit it later.

Early examples of visualizations



6200 BC: A map depicting the town of Konya, Turkey.



950 (AD): A line plot depicting the positions of the sun, moon, and planets throughout the year.

z

difformis vniformiter variatio reddit vniformiter difformem. ¶ Latum vniformiter difforme. Si vniformiter excessus graduum quibus quelibet excedit aliam quamlibet equidistantium sibi inuicem fuerit eadem proportio alterius ad alterius proportionem equalitatis. Tunc illud totum erit vniformiter difforme. Si vero excessus graduum quibus quelibet excedit aliam quamlibet equidistantium sibi inuicem fuerit eadem proportio quam habent quelibet distantiarum suarum proportionem equalitatis. Tunc illud totum erit latum vniformiter difformiter. Si autem vniformiter difformis ut patet ex diffinitionibus membrorum secunde diuisionis. Rursus si nulla proportio seruatur tunc nulla poterit attendi vniformitas in latitudine talis et sic non erit vniformiter difformis ¶ Latum vniformiter difformiter difformiter difforme. Illud quod inter excessus graduum eque distantium non seruat eandem proportionem sicut in secunda parte patebit. Notandum tamen est quod sicut in supradictis diffinitionibus vbi loquitur de excessu graduum inter se eque distantium debet accipi distancia secundum partes latitudinis extensiue et non intensiue vt ita ut loquuntur de excessibus difformitatum vbi distantia sumitur duarum summarum in quibus gradualis

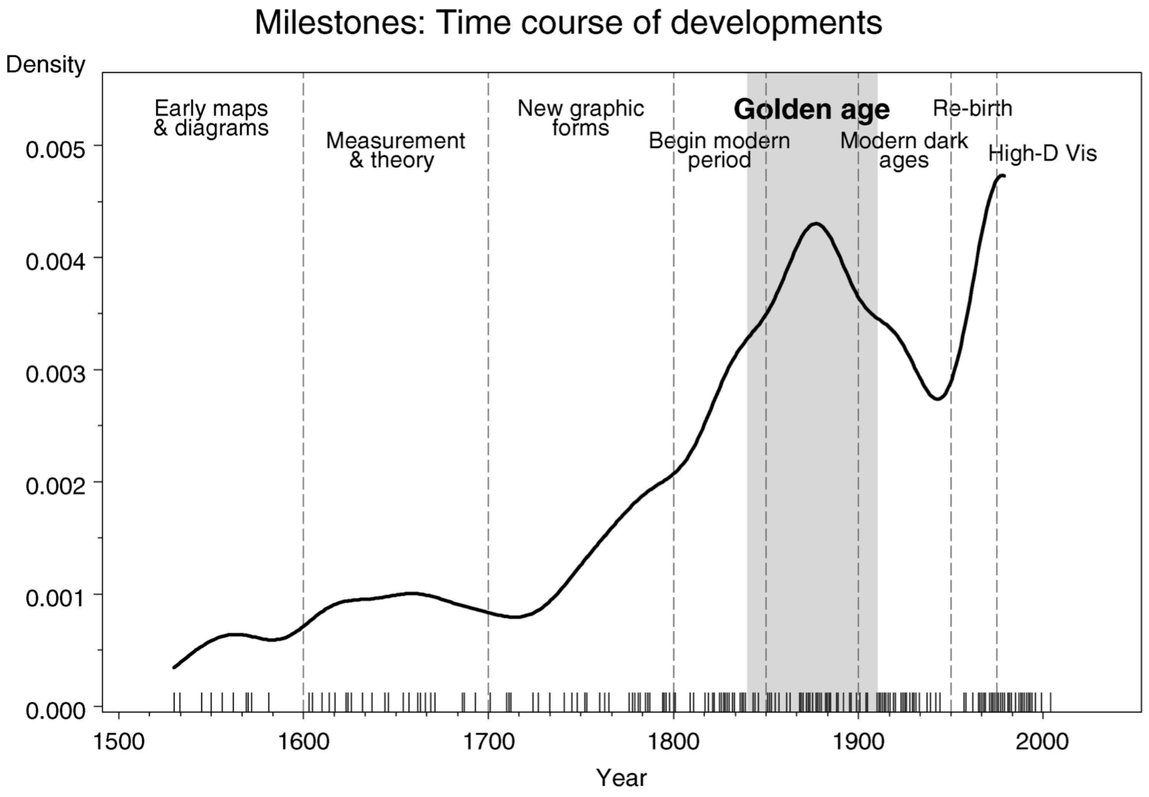
Equitis secunda pars in qua ut supradicta intelligantur ad sensum per figuras geometricas ostenduntur. Et ut omnem speciem latitudinis in presenti materia via occurrat apparentior latitudines ad figuras geometricas applicant. Hec pars diuidit per tria capitula quorum primum ponet diffinitiones. Secundum suppositiones

1350: Nicole Oresme plotted functions of time (e.g. velocity) as bar charts.

Today, we would use line charts or scatter plots to show the same information,.

“Modern” data visualization

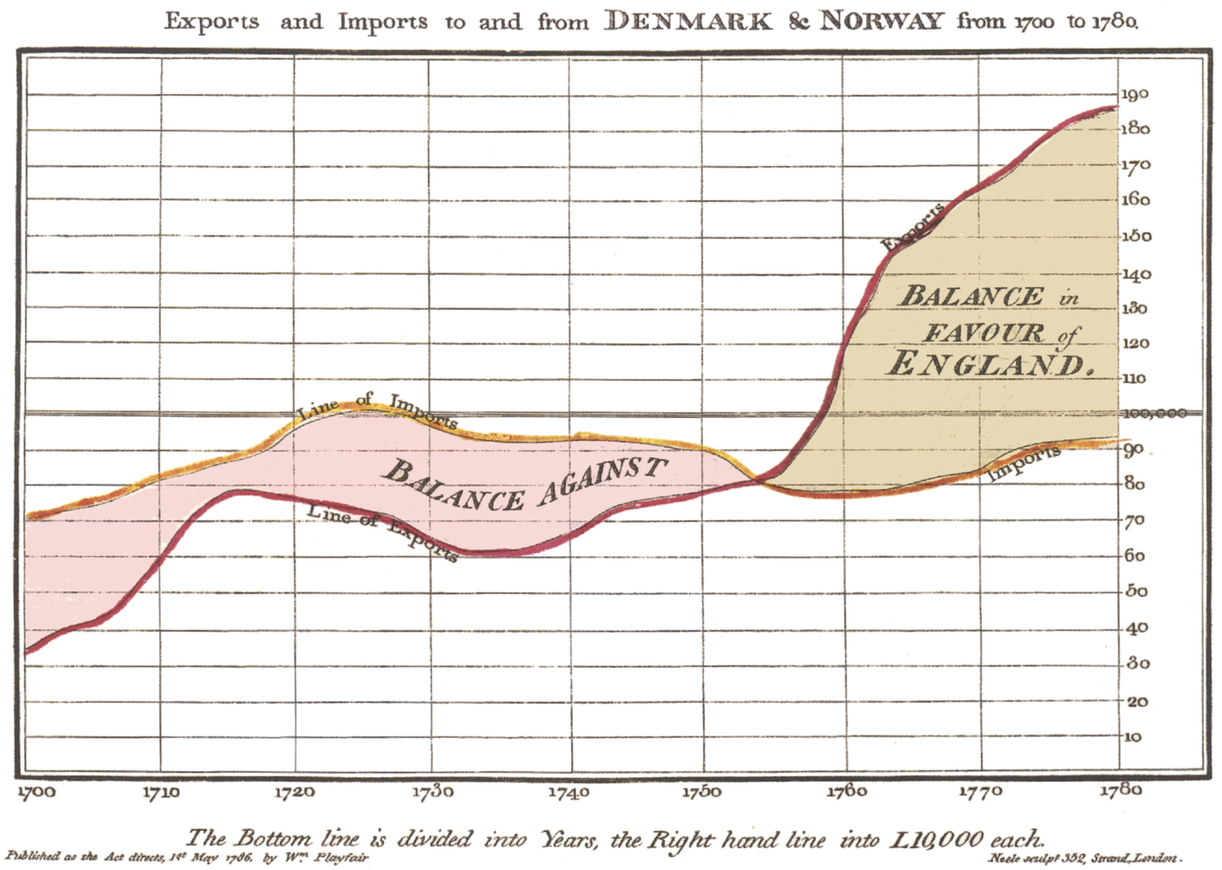
- In the late 1600s, civilizations started to gather large amounts of information about their citizens (e.g. births and deaths) and trade (e.g. imports and exports).
- The term *statistics* comes from the latin term *statisticum*, which means “of the state,” and was introduced around 1750.

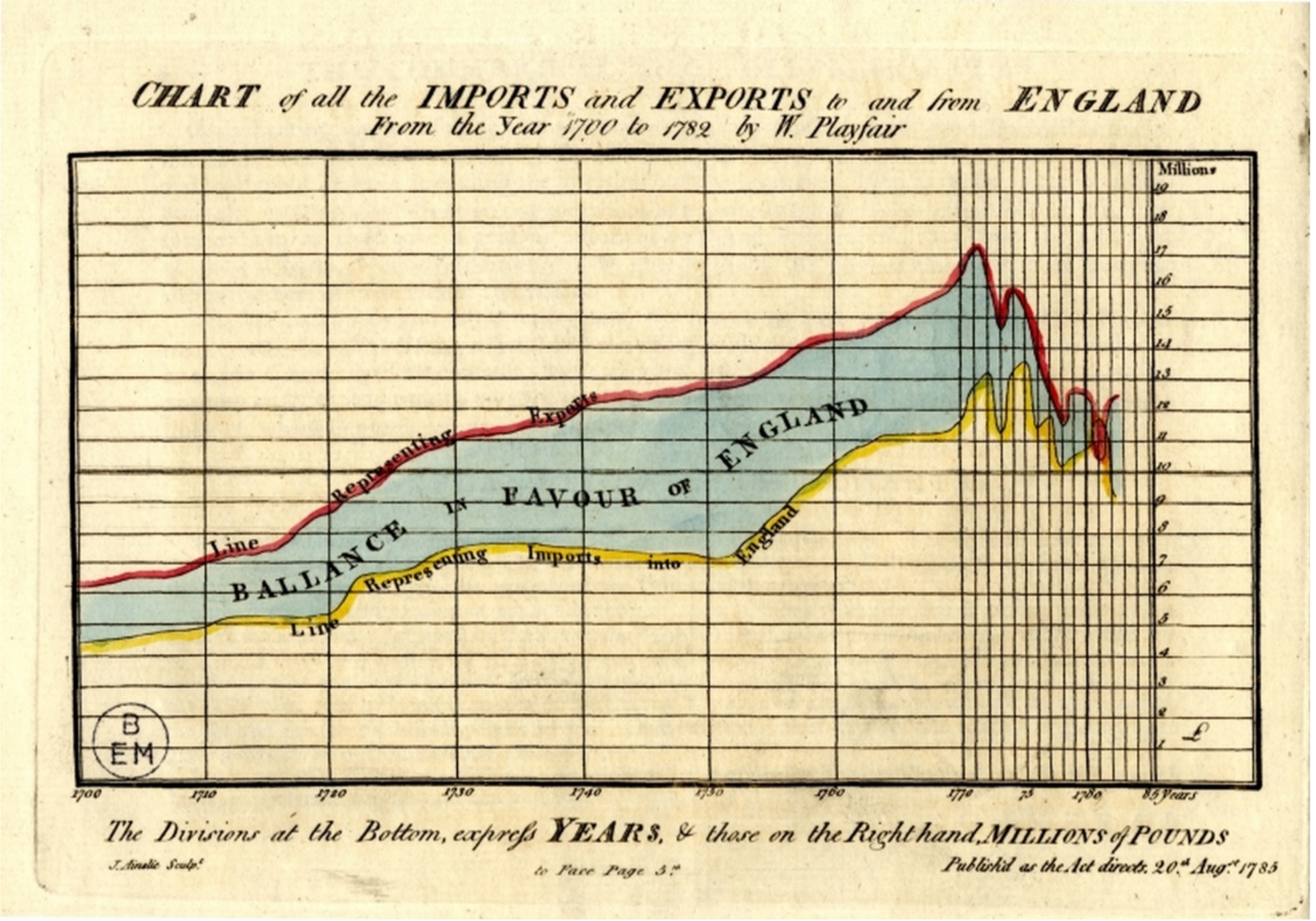


Source

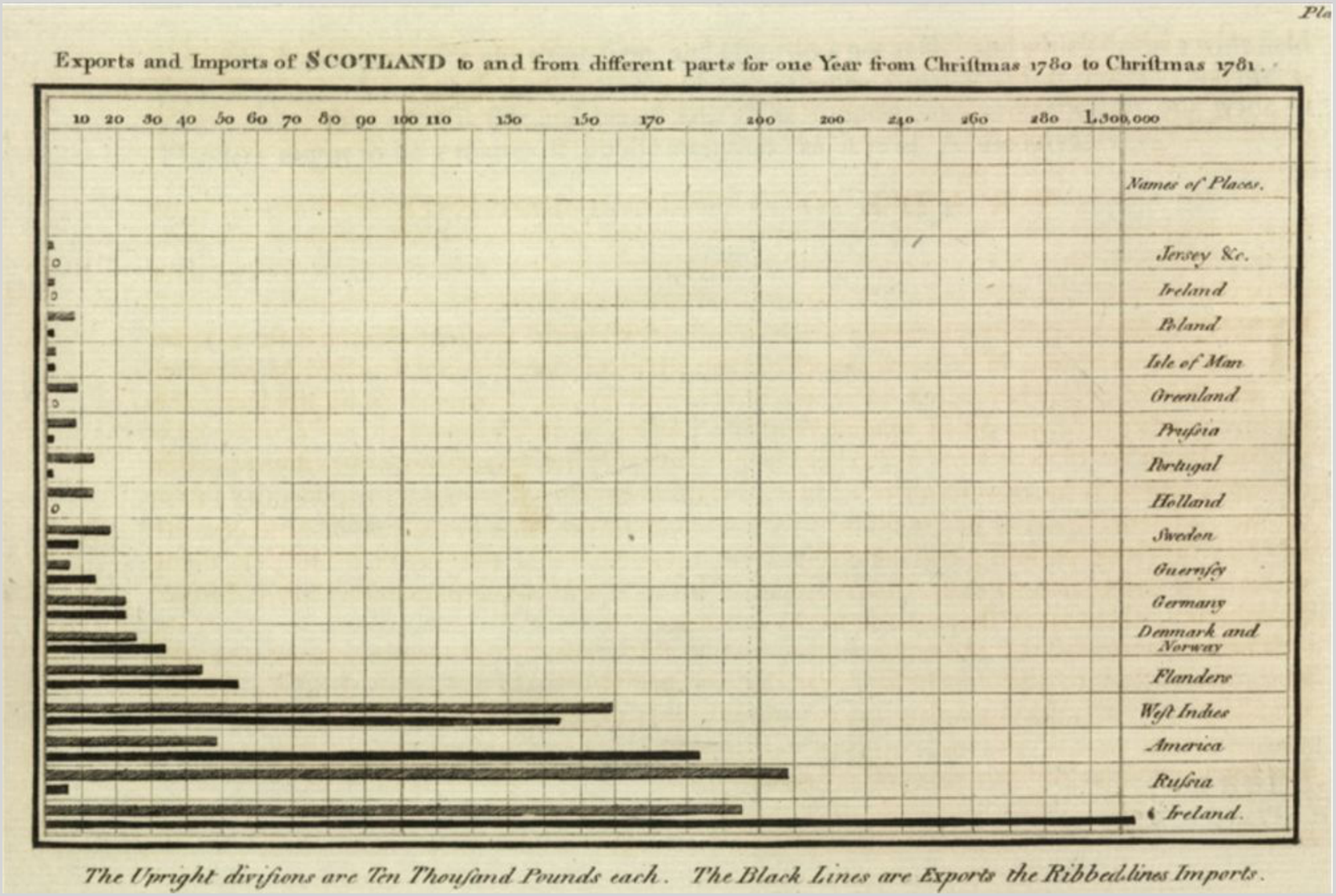
William Playfair

- William Playfair (1759-1823), of Scotland, is known as the “father of data visualization”. He is credited for developing:
  - line charts
  - bar charts
  - pie charts
- One of his most famous visualizations, shown to the right, depicts England’s imports and exports to Denmark and Norway (1786).
  - Interactive version here.

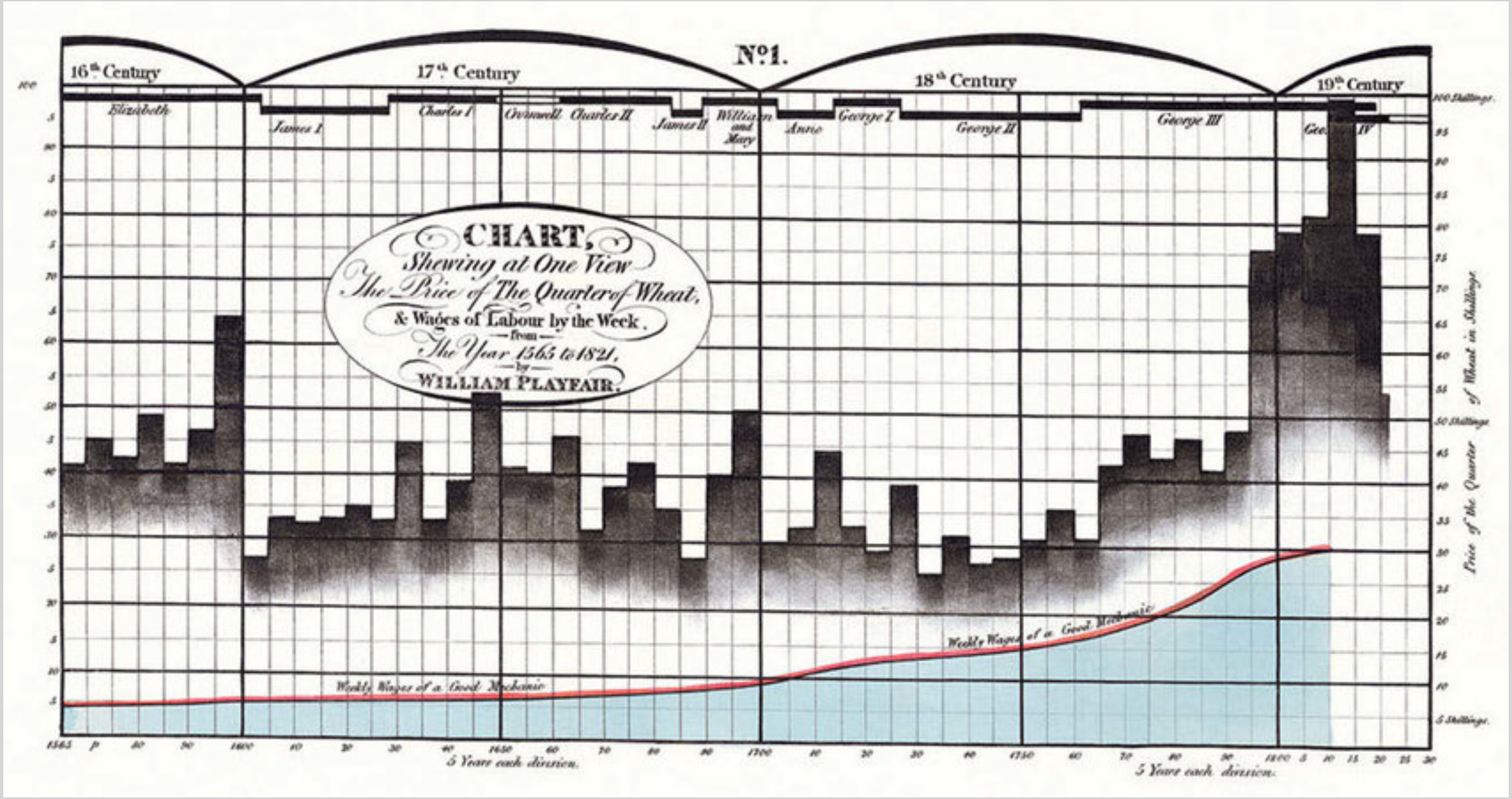




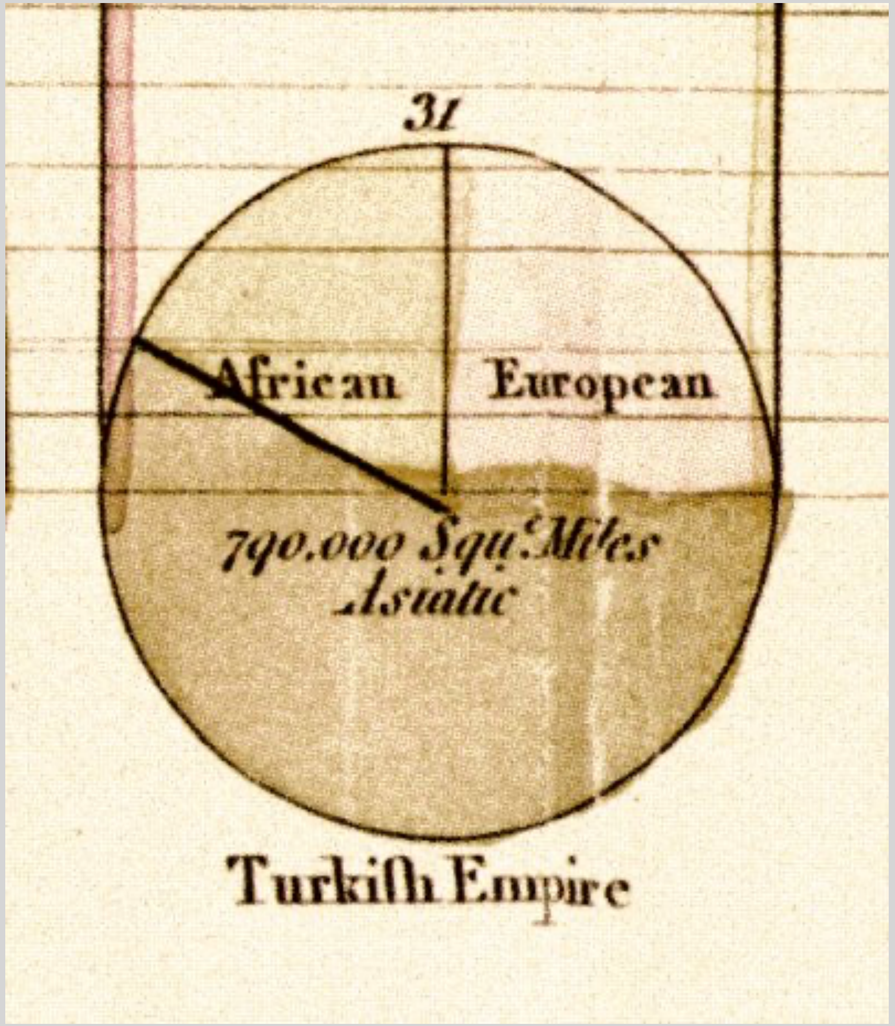
1785: Another line chart by Playfair, depicting the total imports and exports to England over a period of 85 years.



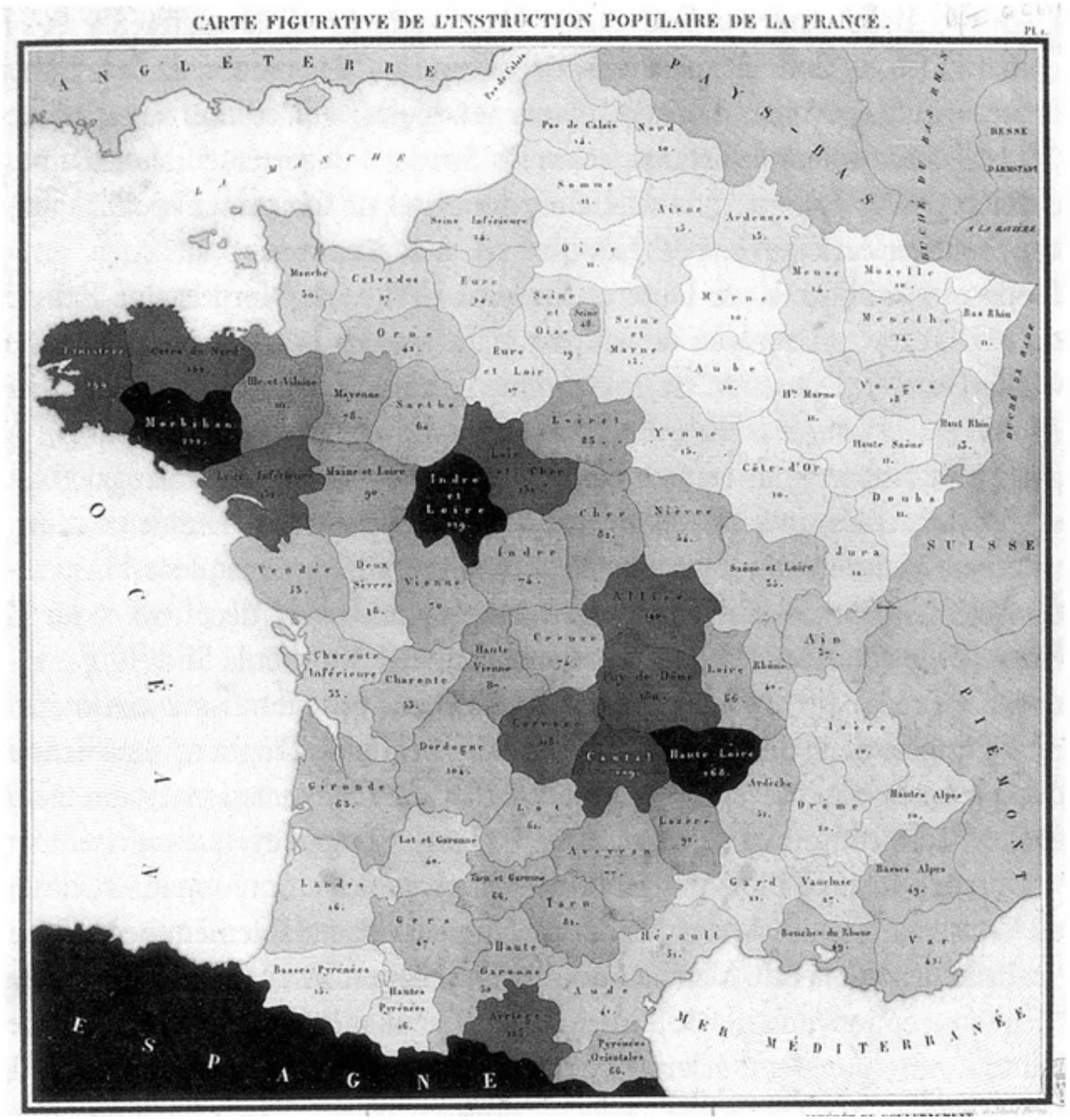
1796: The first known example of a bar chart, also by Playfair, depicting the imports and exports of Scotland to various countries in 1781.



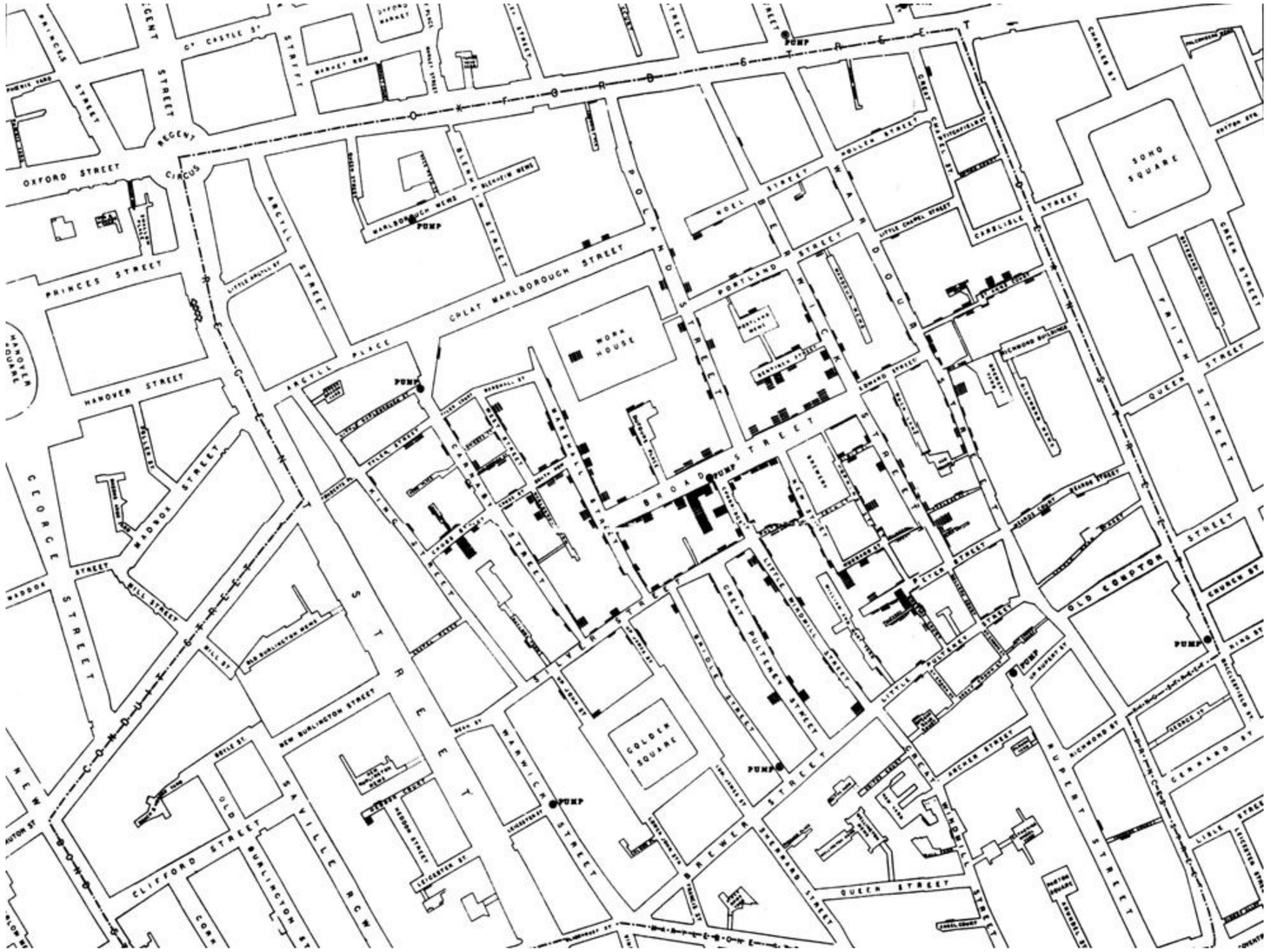
1821: Another Playfair visualization, showing the relationship between weekly labor wages and the cost of a "quarter" of wheat, along with a timeline of English monarchs, from 1565 to 1821.



1801: Playfair's pie chart depicting the distribution of the Turkish Empire.

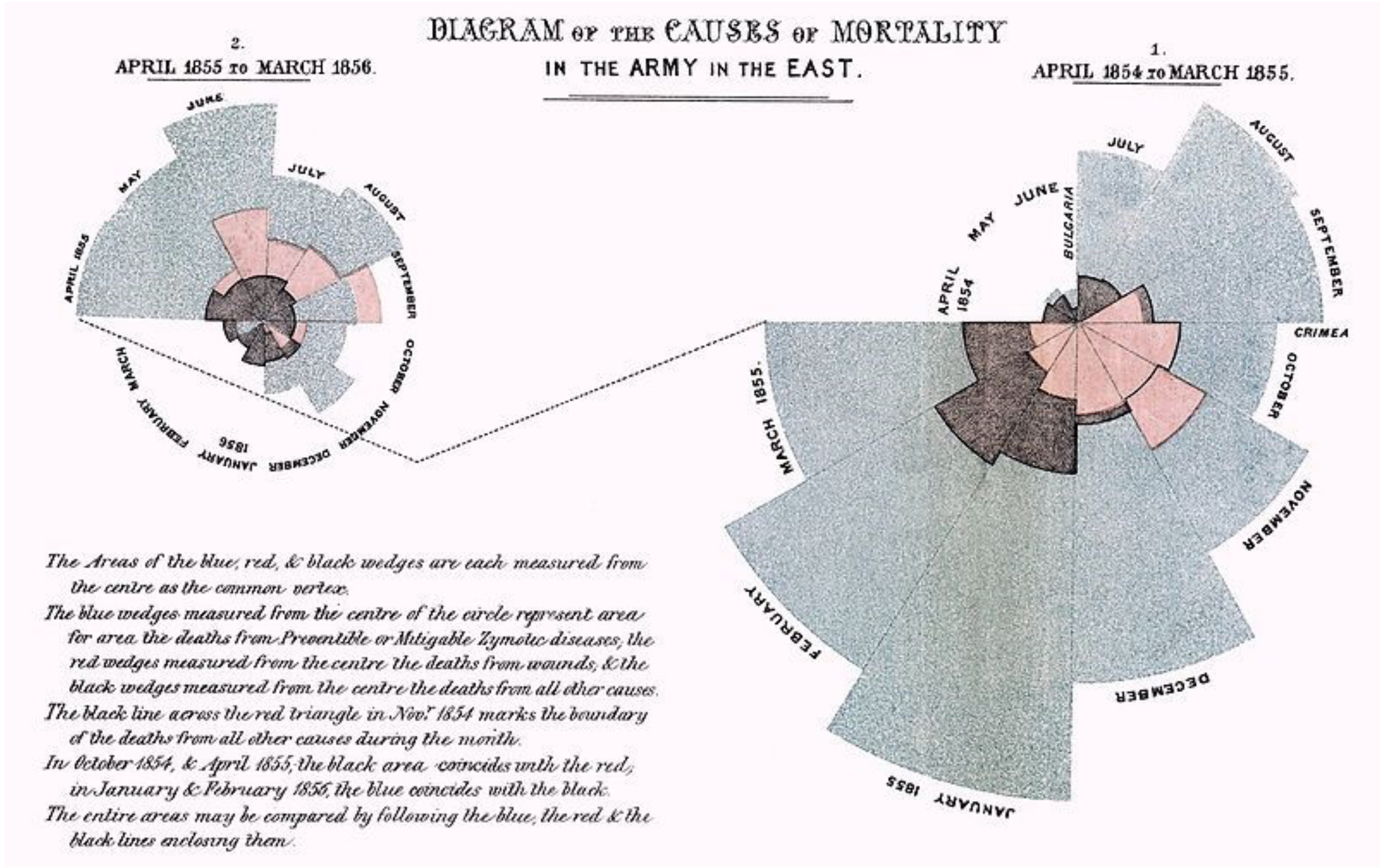


1826: Charles Dupin creates a choropleth, which describes the distribution of some quantity for each of several physical regions. His choropleth depicted rates of literacy in different parts of France.

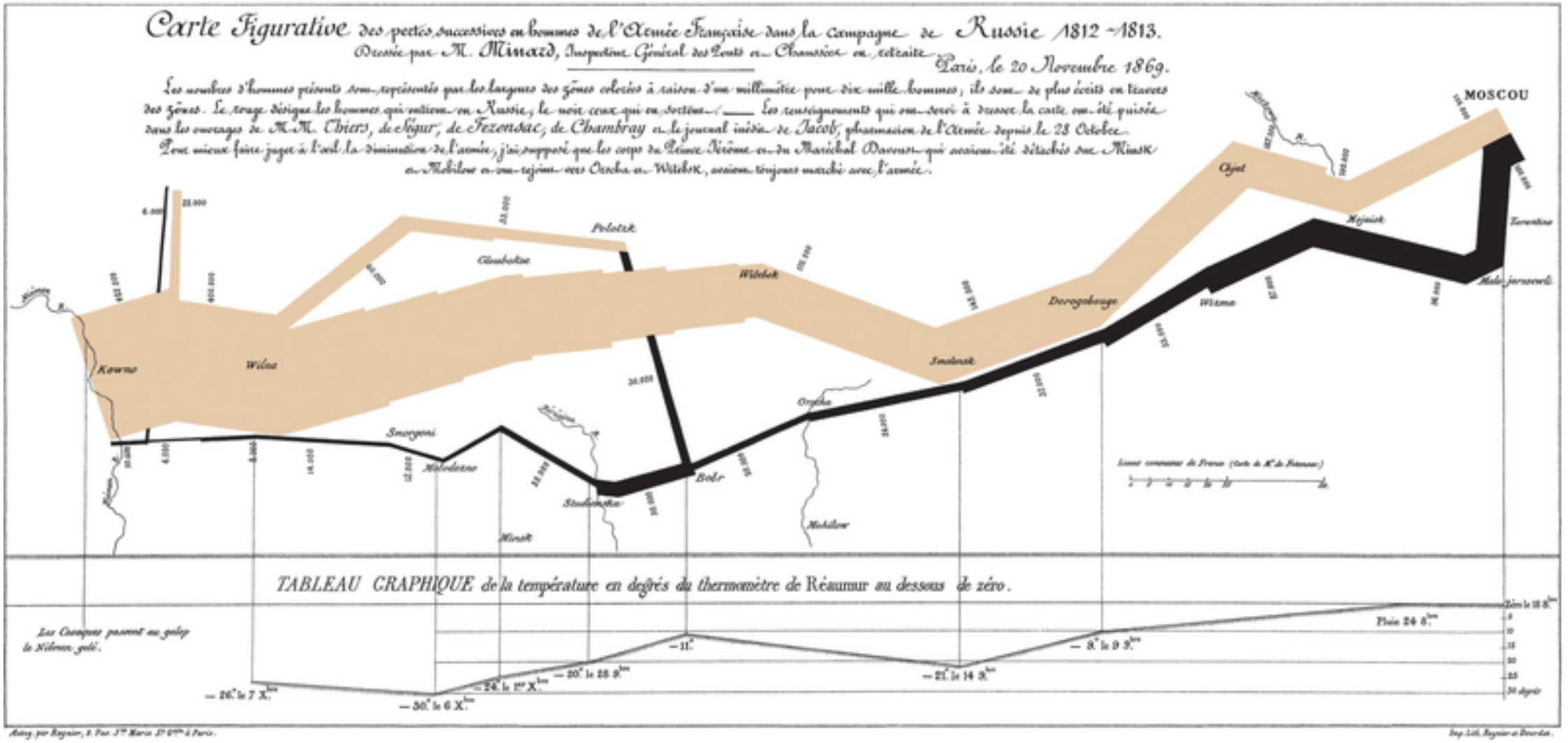


1854: John Snow mapped cholera deaths in SoHo, London. He noticed that many deaths were clustered around the Broad Street pump.

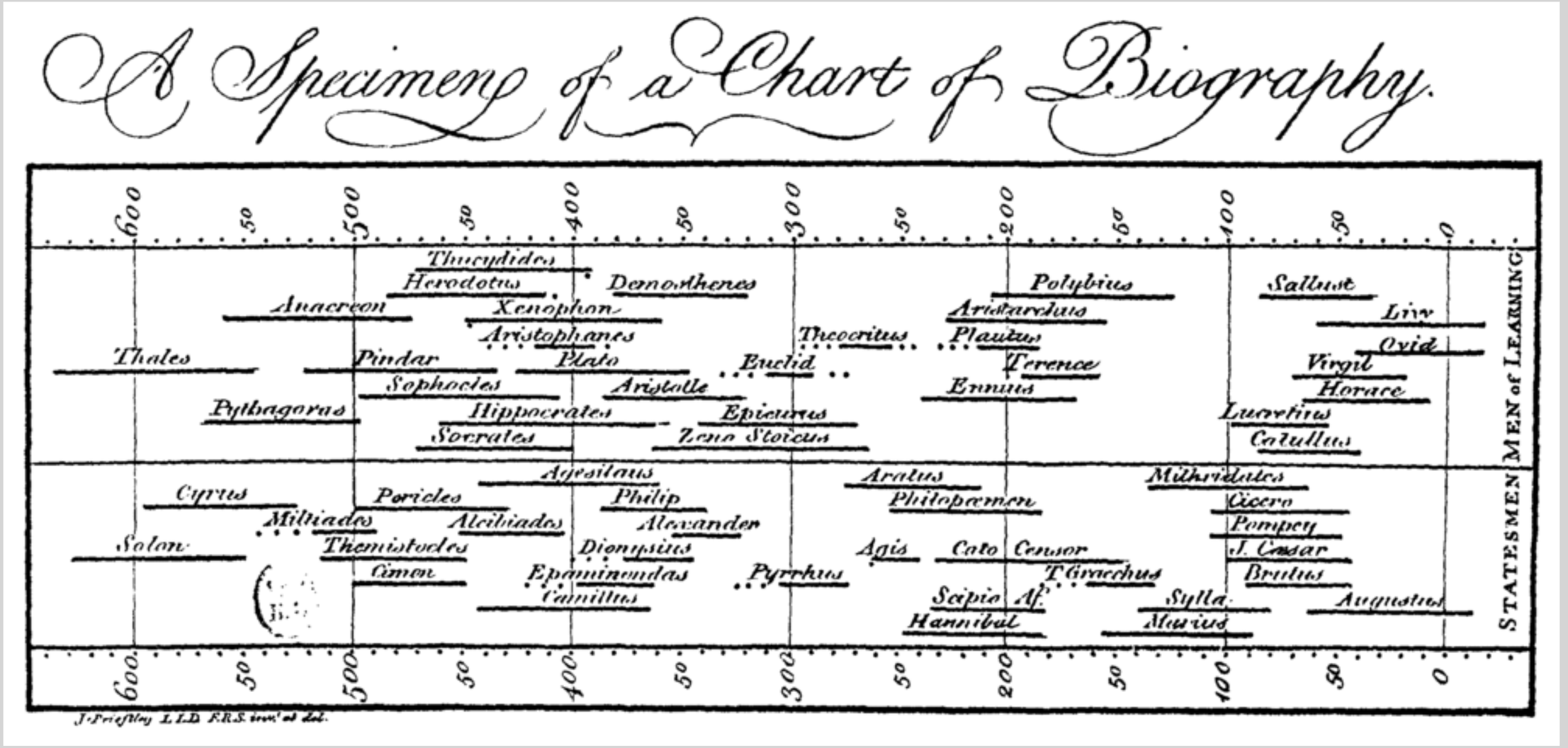
2020: The site of the Broad Street pump.



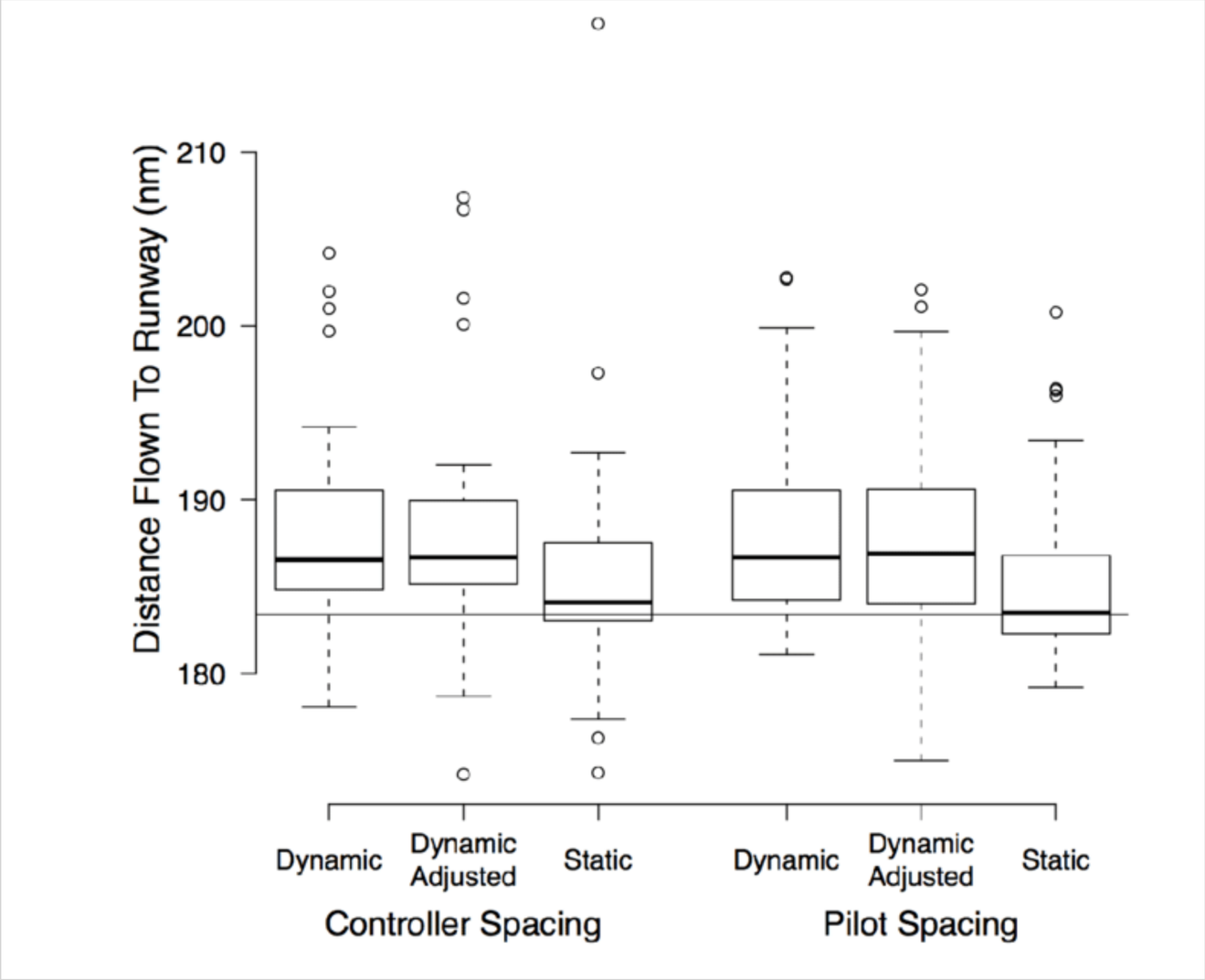
1855: Florence Nightingale's depiction of the deaths of British soldiers in the Crimean war. Florence Nightingale is known as the founder of modern nursing.



1869: Charles Joseph Minard's visualization of the French invasion of Russia (led by Napoleon).



1765: Joseph Priestley creates the "Chart of Biography", a timeline of the lifespans of several prominent figures in BC. This type of visualization is now occasionally called a "Gantt chart."



1973: John Tukey, who defined the term "Exploratory Data Analysis", created the box plot, which describes a numerical distribution using a 5 number summary.

That’s all!