

Data Visualization Using Spreadsheets

Professor Widom's Instructional Odyssey

www.professorwidom.org



Data Visualization

“A picture is worth a thousand words”

- Data visualizations can be enlightening and powerful
- Creating good visualizations is a science and an art

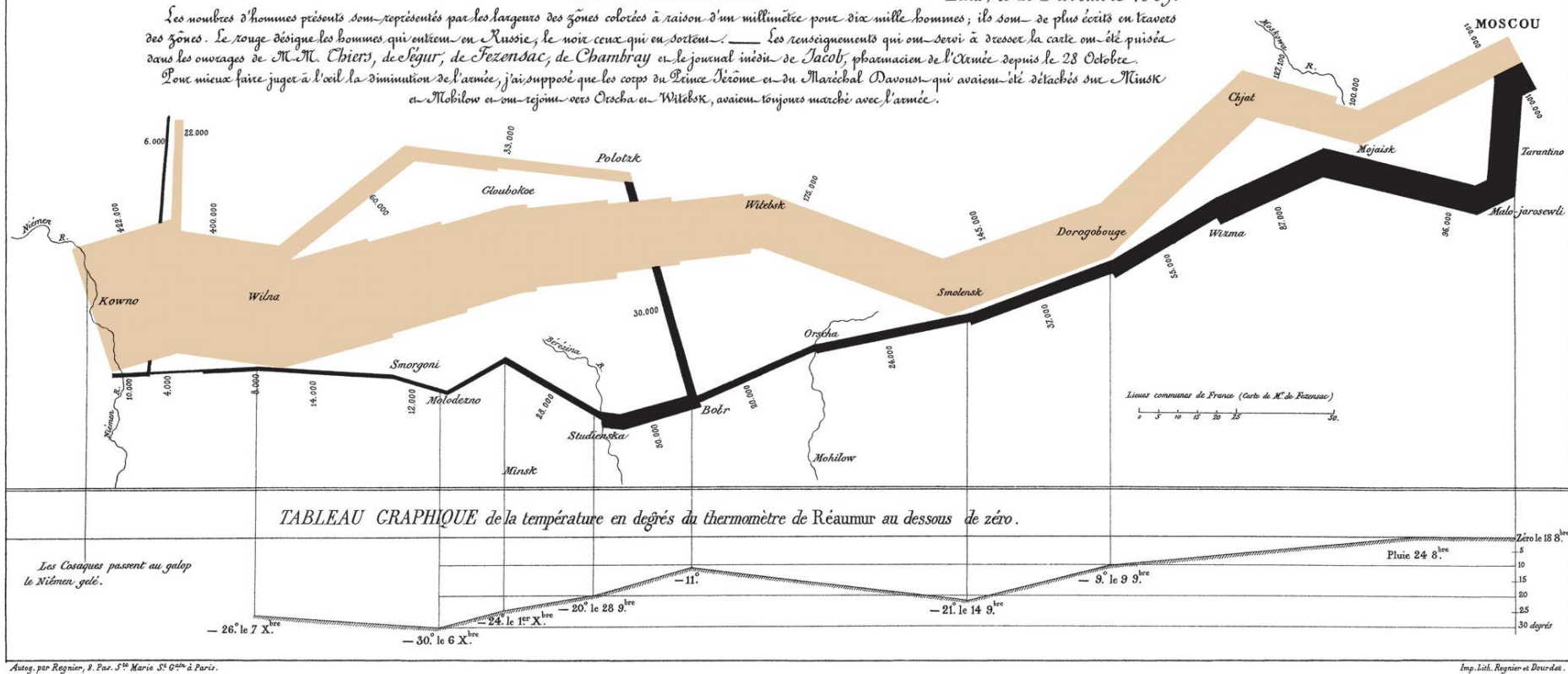
Early Data Visualization

Napoleon's Army

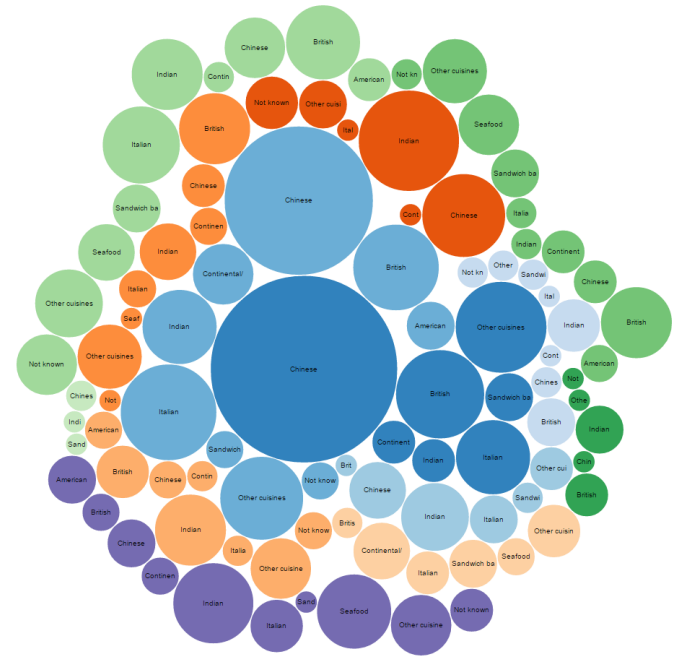
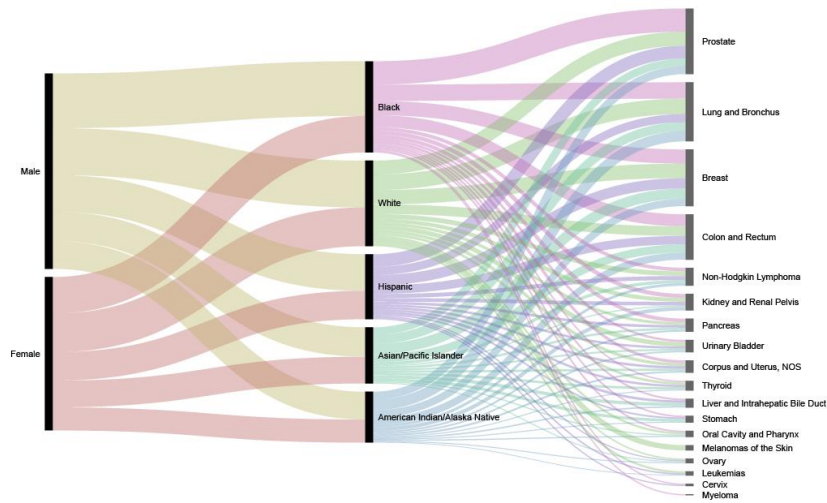
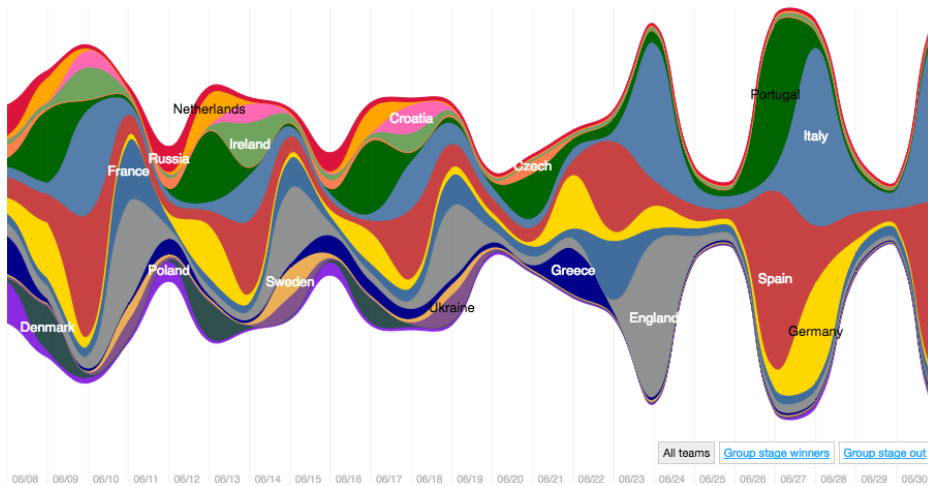
Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite, Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Thiers, de Leguér, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre. Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mohilow et qui rejoignent vers Orscha et Witebsk, avaient toujours marché avec l'armée.



Modern Data Visualization



Dynamic Data Visualization

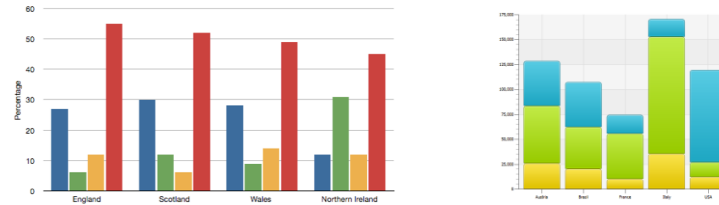


<https://www.youtube.com/watch?v=jbkSRLYSojo>

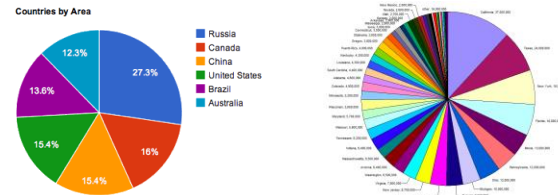
Basic Data Visualization

Don't underestimate the power of basic visualizations

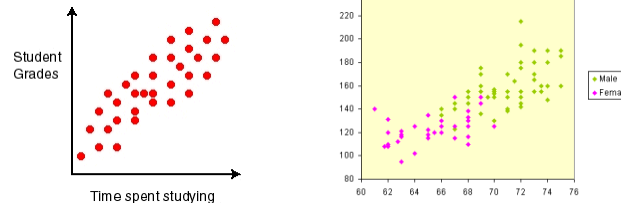
- Bar charts



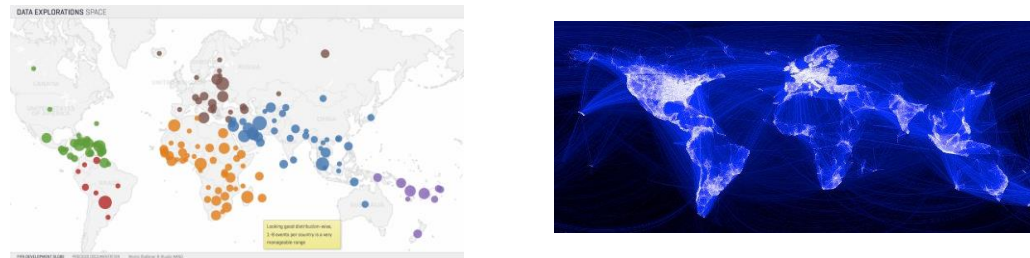
- Pie charts



- Scatterplots



- Maps



Bar Charts

When one axis is categories
and the other is numeric

Your Turn

Ten countries with the highest population,
bar chart showing populations

Pie Charts

To compare relative sizes of categories



Your Turn

Pie chart showing relative number of cities with negative longitude and positive longitude

Label the two slices “west” for west of the Prime Meridian (negative longitude), and “east” for east of the Prime Meridian (positive longitude)

Scatterplots

When both axes are numeric



Your Turn

Switch to CitiesExt.csv

Use a scatterplot to explore whether there is any relationship between the latitude of cities in a country (x-axis) and the population of that country (y-axis)

Why are there lines of dots?

Maps

Values by geographic region



Your Turn

Map of European countries colored by whether or not the country is in the European Union

Data Visualization with Spreadsheets

Convenient and useful

- Suggested charts are often good ones
- Google Sheets has fewer features than Microsoft Excel (for now)
Though some people prefer the simplicity

For help while working with charts

- Web search, sometimes
- Or just keep fiddling 😊

World Cup Data Visualization

1. Create a **bar chart** showing the average number of minutes played by players in each of the four positions.
2. Create a **stacked bar chart** for teams that played more than 4 games, showing their number of wins, draws, and losses (not percentages).
3. Create a **pie chart** showing the relative percentage of teams with 0, 1, and 2 red cards. *Note: the pie should have three slices.*
4. Create a **scatterplot** of players showing passes (y-axis) versus minutes (x-axis). (Why are there some lines of dots?)
5. Create a **map** of countries colored light to dark blue based on how many goals their team made (“goalsFor”).
6. Create a **pie chart** showing the relative percentage of players making ≤ 0.25 passes per minute, ≥ 0.5 passes per minute, and between 0.25 and 0.5. Hint: `=countif()` and `=countifs()` ★

Titanic Data Visualization (Extra)

1. Create a **bar chart** showing the average fare paid by passengers in each class, with the three bars labeled "first", "second", "third".
2. Create a **stacked bar chart** showing the number of passengers in each class, divided into male and female (three bars). Then reverse roles and show the number of passengers of each gender, divided into class (two bars).
3. Create a **pie chart** showing the relative number of male survivors, male non-survivors, female survivors, and female non-survivors (four slices).
4. For passengers in 3rd class, create a **scatterplot** of fare (y-axis) versus age (x-axis). Don't worry about missing ages.