EDP308: STATISTICAL LITERACY

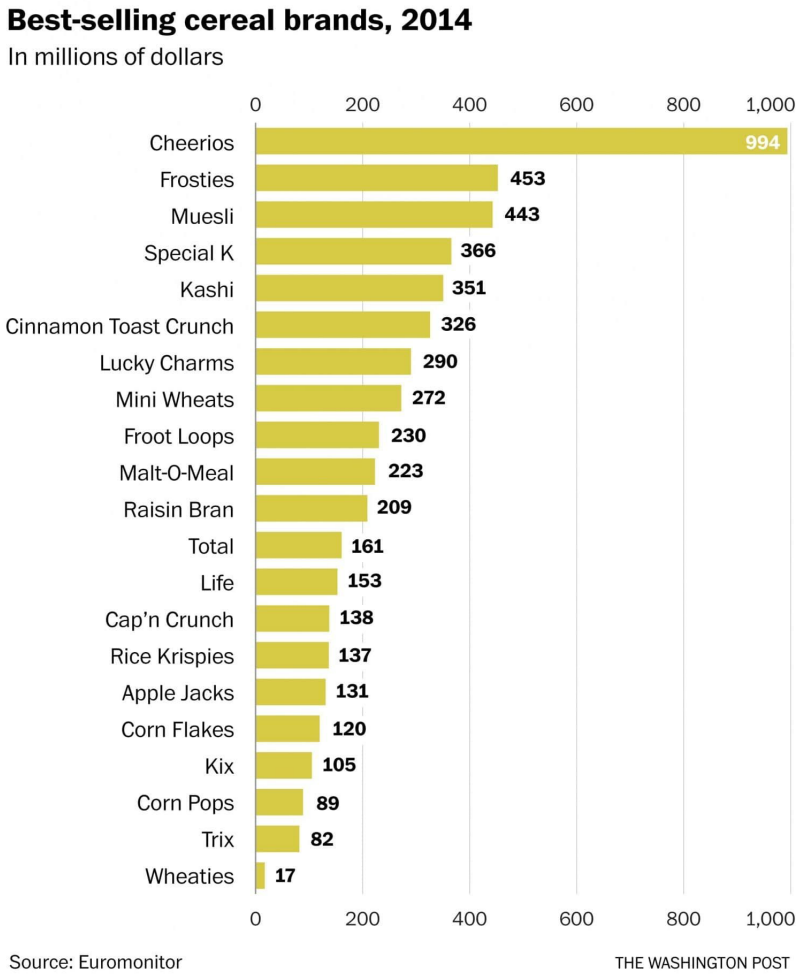
The University of Texas at Austin, Fall 2020
RAZ: Rebecca A. Zárate, MA

Overview

- Data Visualizations
  - Bar Graph
  - Histogram
  - Box Plot
  - Line Graph
  - Area Graph
  - Scatter Plot
- Historical Moment: Florence Nightingale
  - Polar Plot
- Visualizations with R

Data Visualizations

Visualizations: Bar Graphs

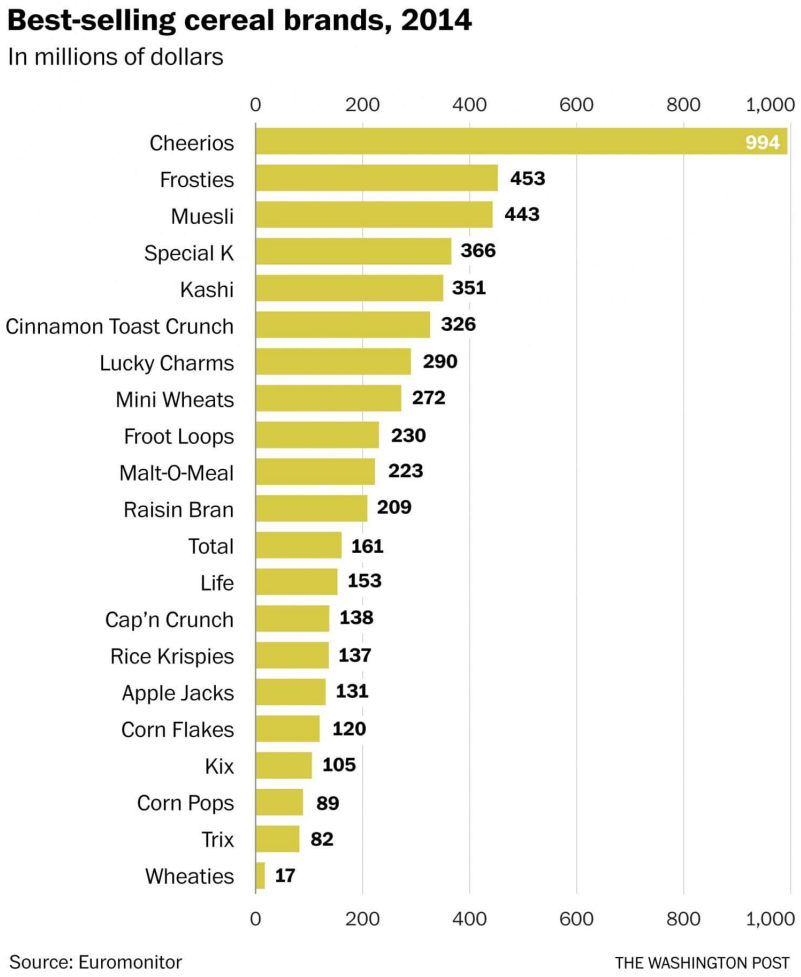


- Bar graphs are used for Categorical (nominal) data.
  - The bars to not touch because they are distinct groups.
  - Usually represent count (frequency) data and relative frequency.

What information do we have?

Which cereal is best-selling?

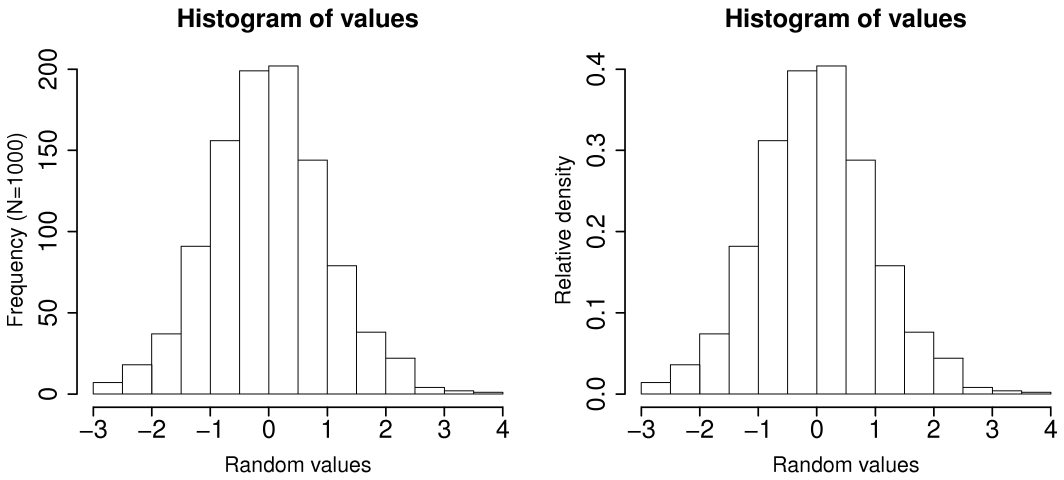
Visualizations: Bar Graphs



- Cereal = Categorical, nominal data
- Millions of dollars = Continuous, ratio
- Cheerios is best selling brand
  - About twice as much as the next best-selling brand Frosties

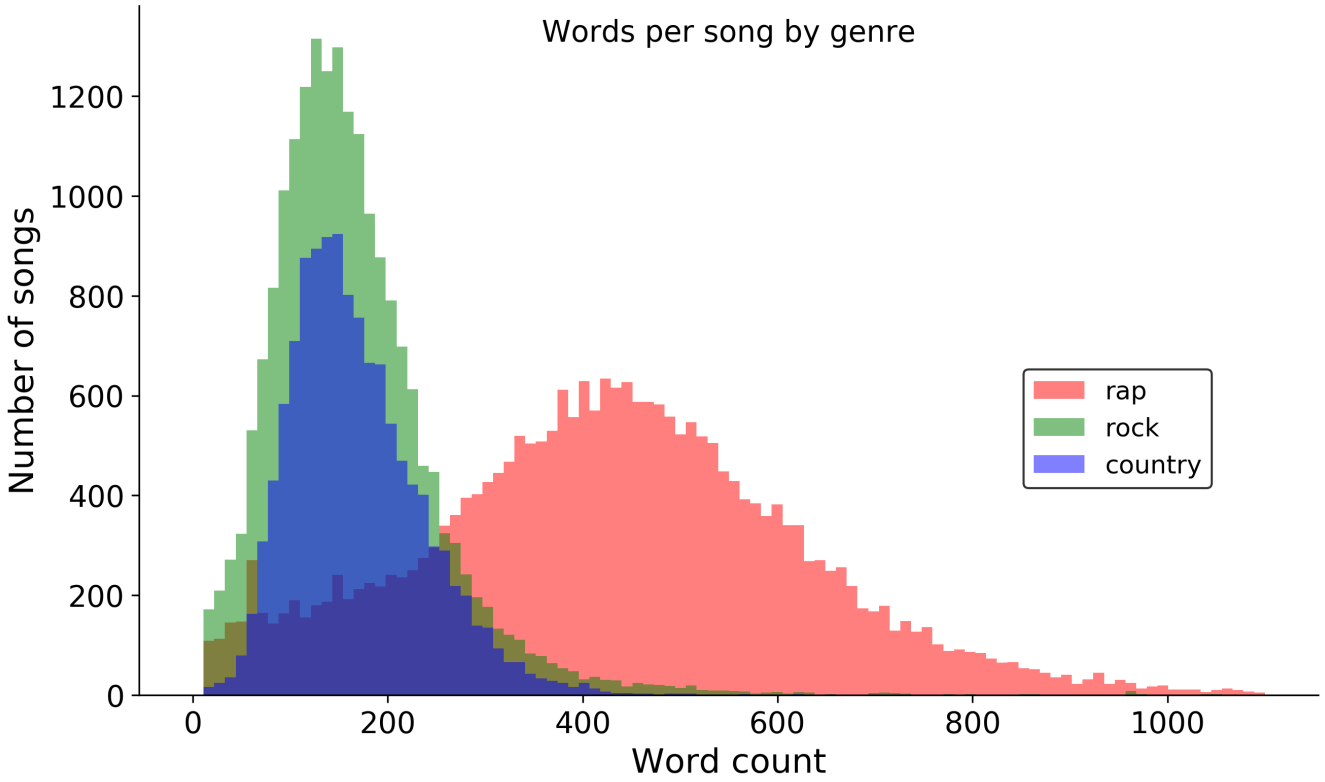
Visualizations: Histograms

- Look very similar to bar graphs, but in a histogram the bars touch each other. Why?
  - Because Histograms are used for CONTINOUS data, bar graphs are for DISCRETE CATEGORICAL data
    - Sometimes this general rule is violated… Stay vigilant!



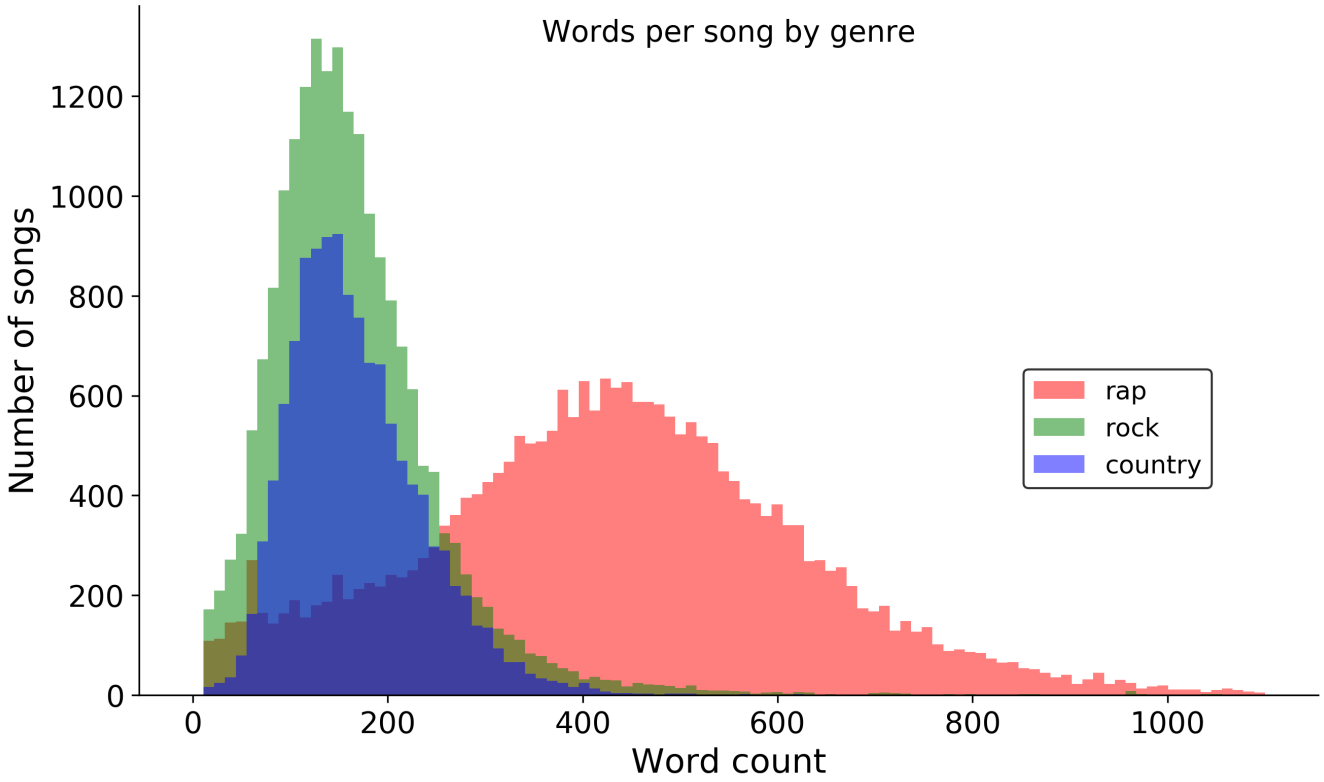
*Note the different Y axes

Another Visual: Histogram



What information do we have? What are the variables?
What does the color represent, what kind of data?
Which has more variability?

Another Visual: Histogram



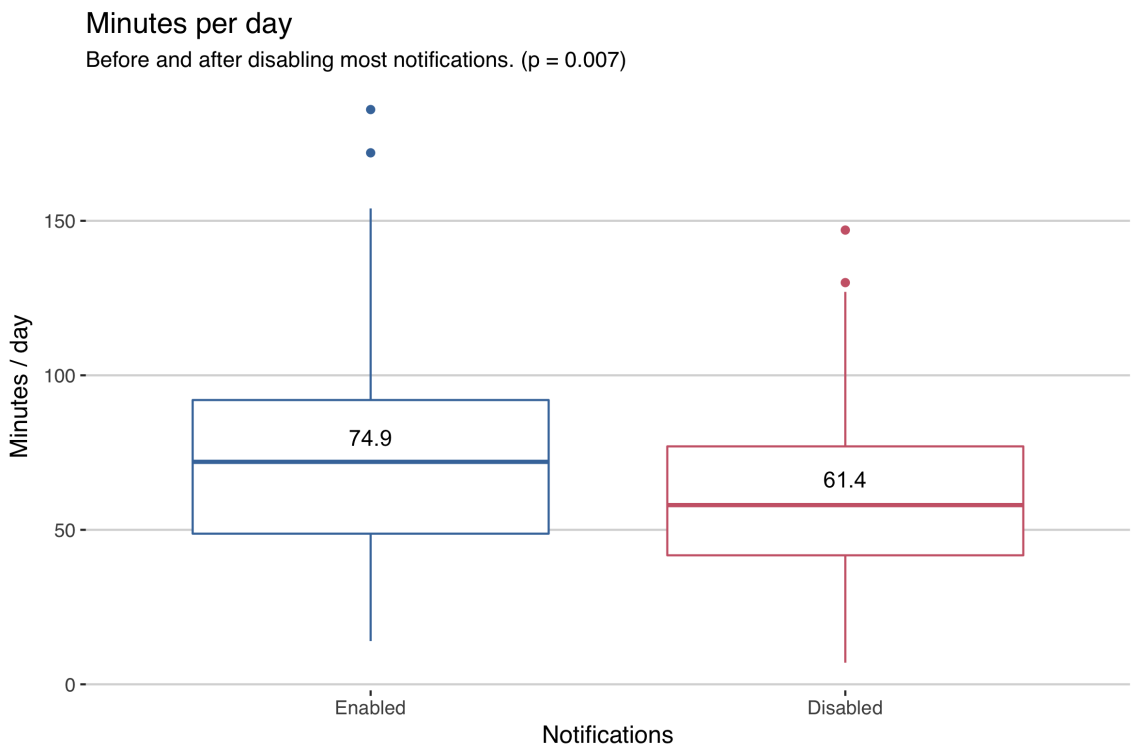
1) Frequency (number of songs), 2) Word Count

3) Color is Categorical, music genre

Rap has greater variability in number of words

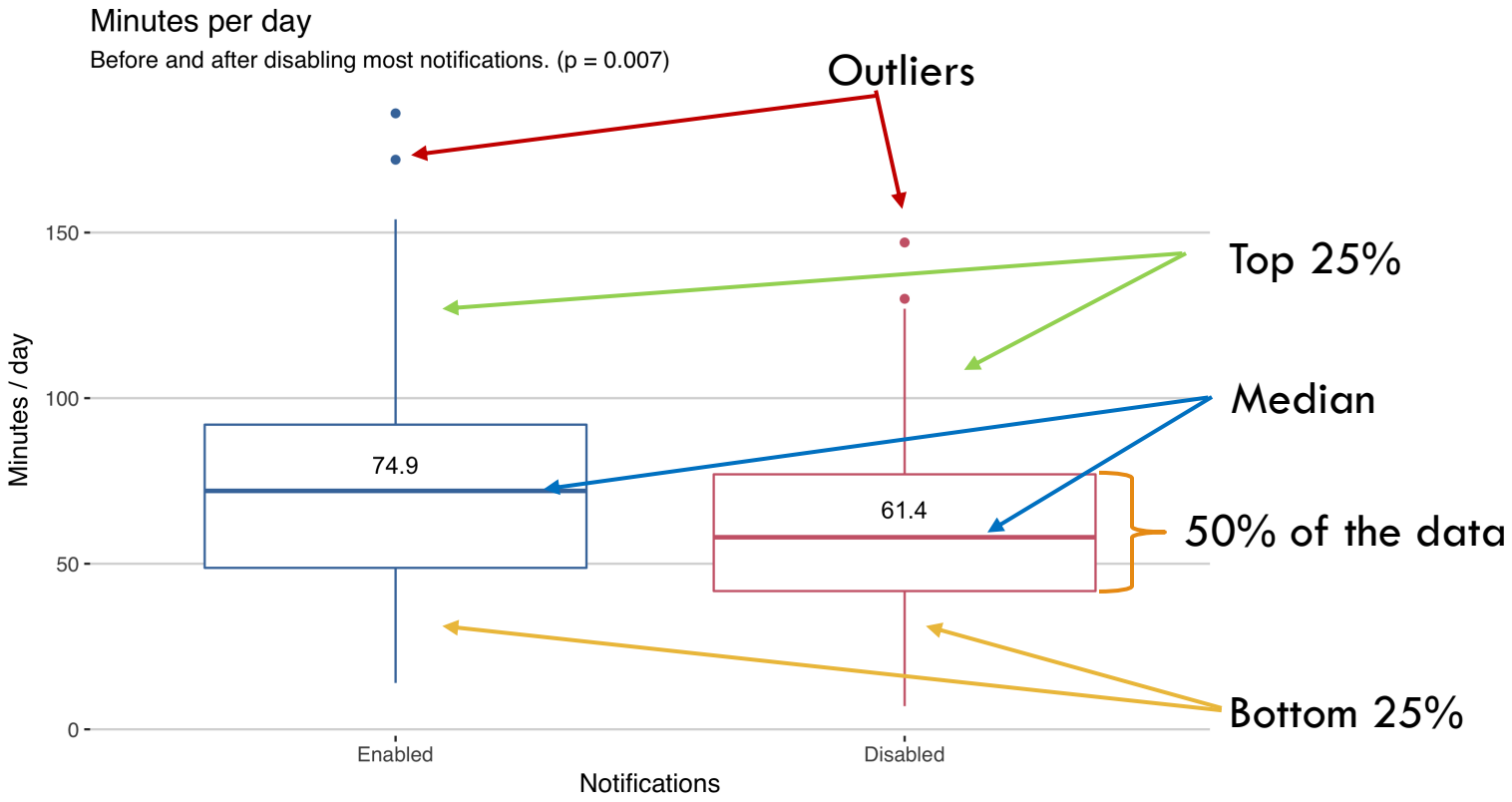
Visualization: Boxplots

- Boxplots show the median, the spread of the data (interquartile range), and outliers.
  - Also called "box-and-whisker" plots…
- Need ratio or interval data
- Categories are optional



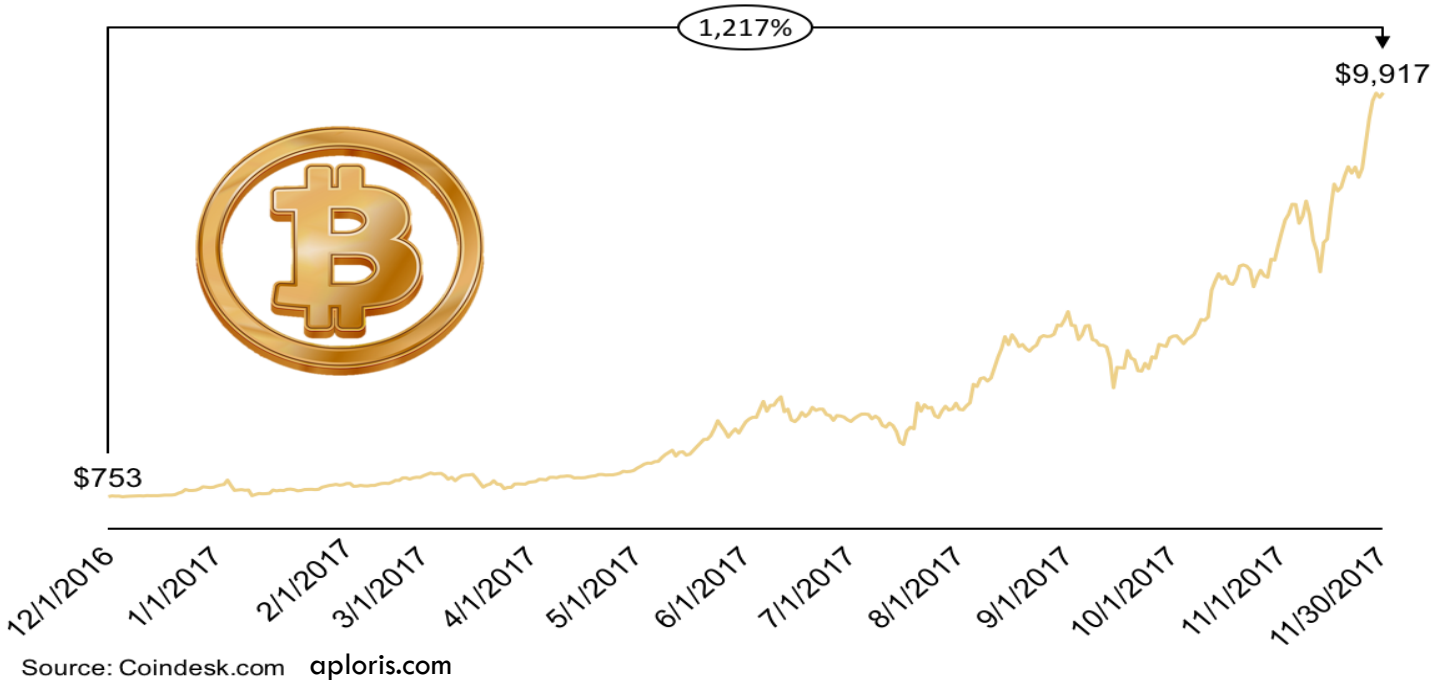
Visualization: Boxplots

- Boxplots show the median, the spread of the data (interquartile range), and outliers.



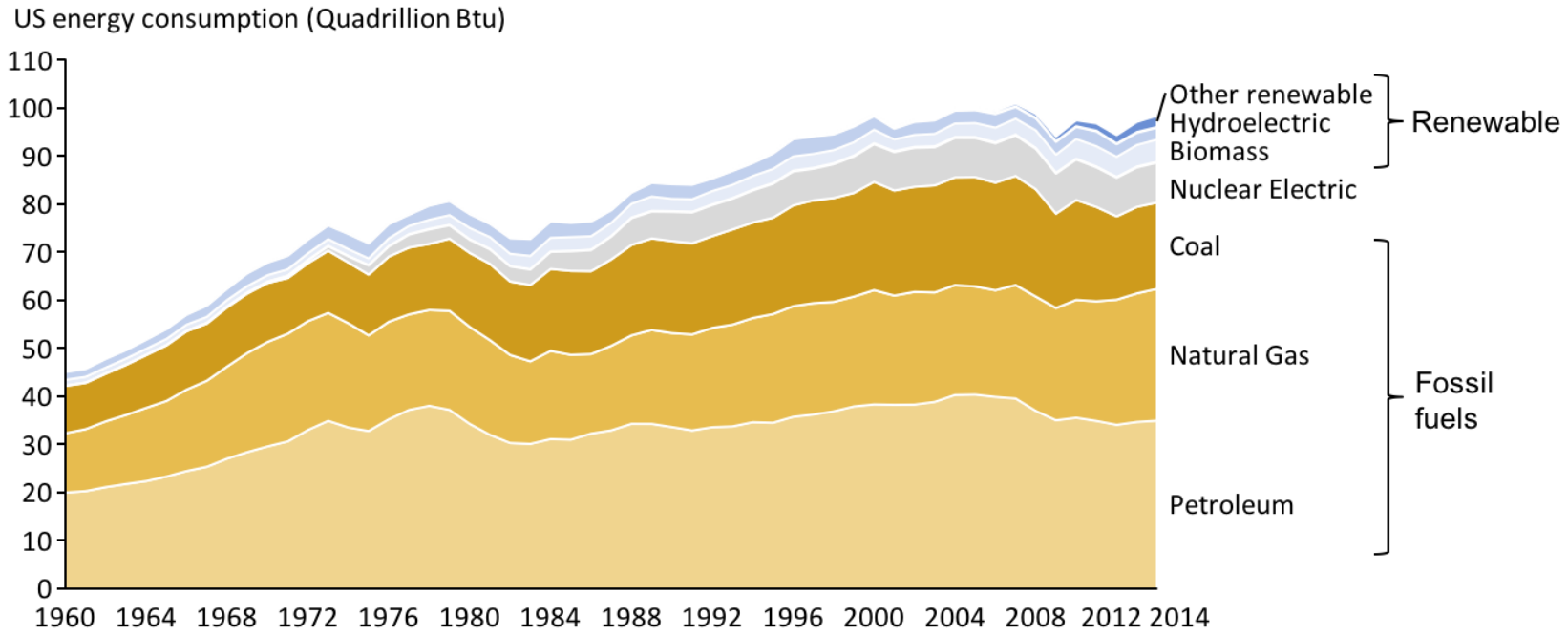
Visualization: Line Graphs

- Line graphs are typically chosen to show changes of time and trends
  - Lines are connected because time is treated as continuous



Visualization: Area Graphs

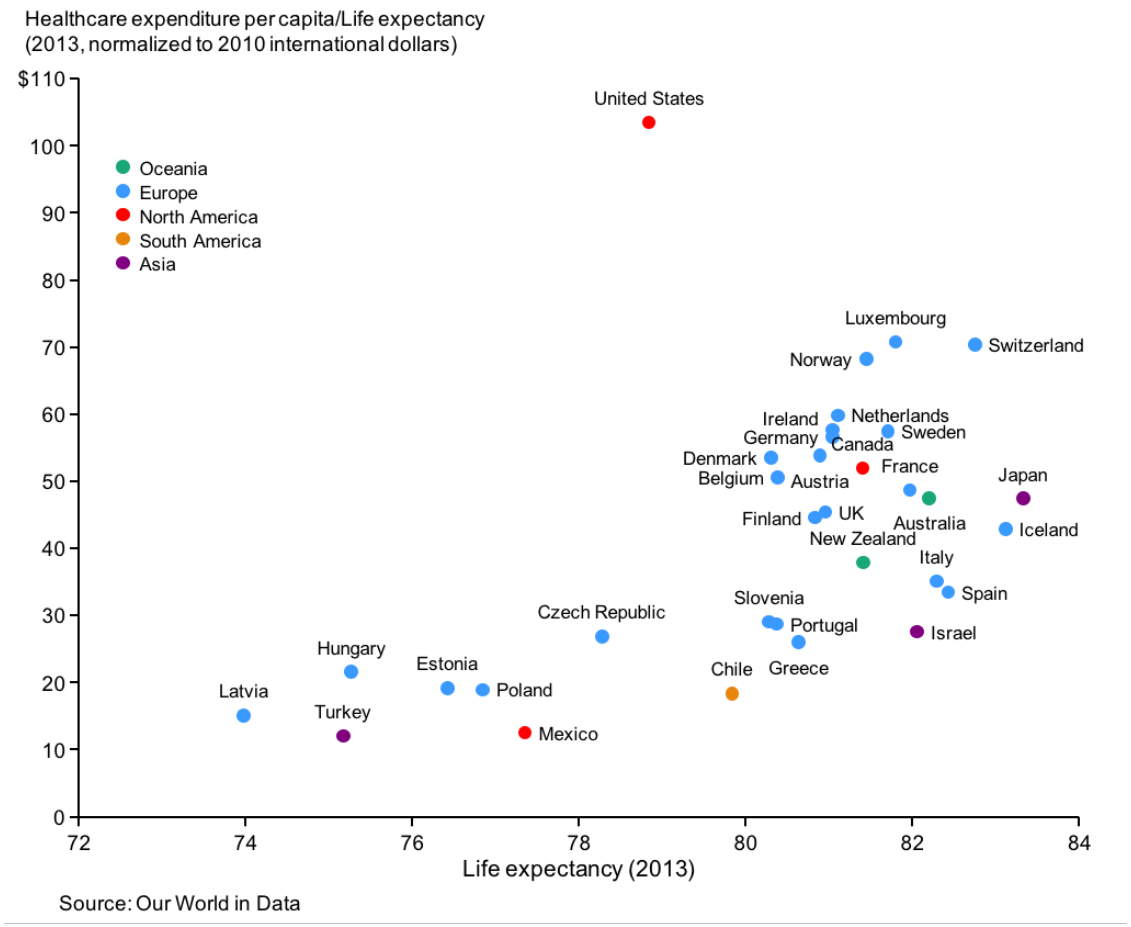
- Area graphs are based on line graphs to show change over time, but they also compare two or more quantities of different groups.



Source: US Energy Information Administration

Note: Other renewable includes geothermal, solar/PV, and wind

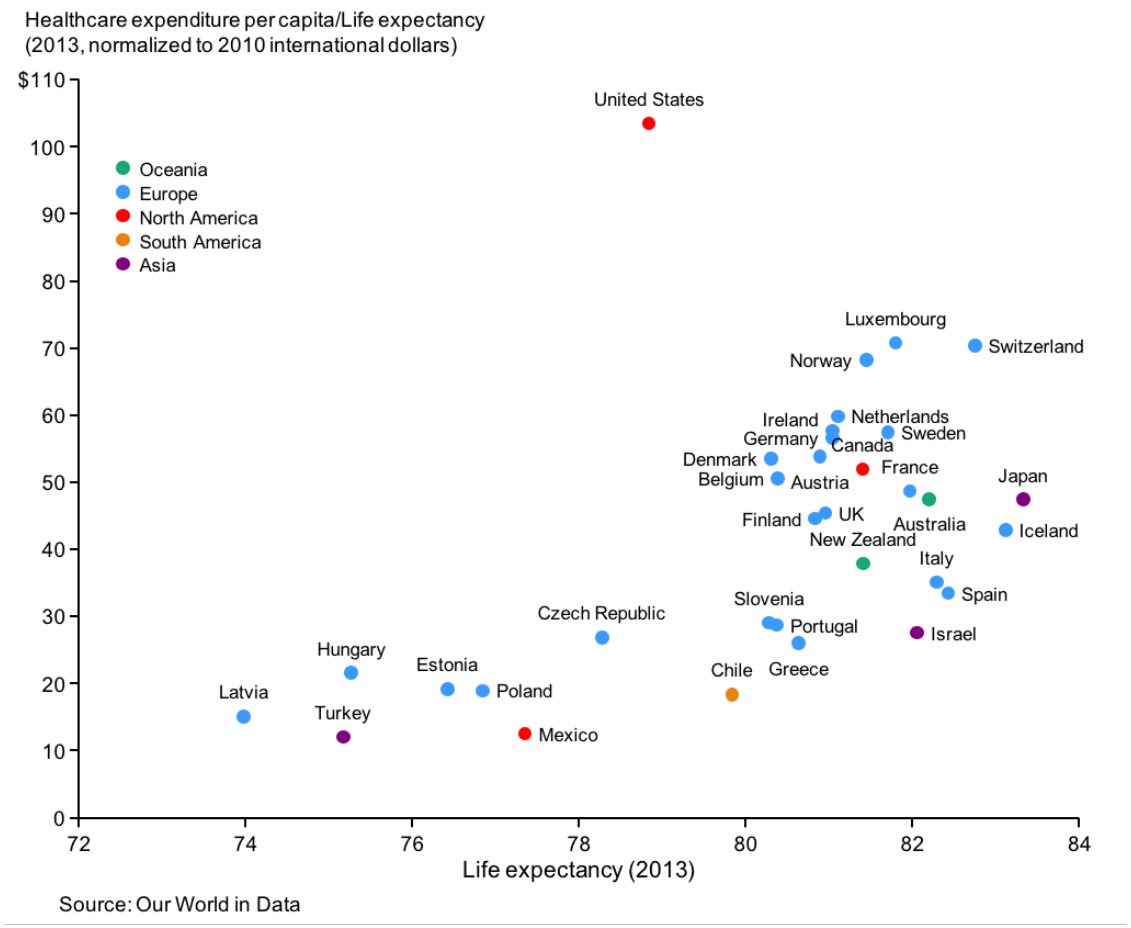
Visualization: Scatter Plots



- Scatter plots show the relationship between two continuous variables.

What kind of information do we have here? How many variables?

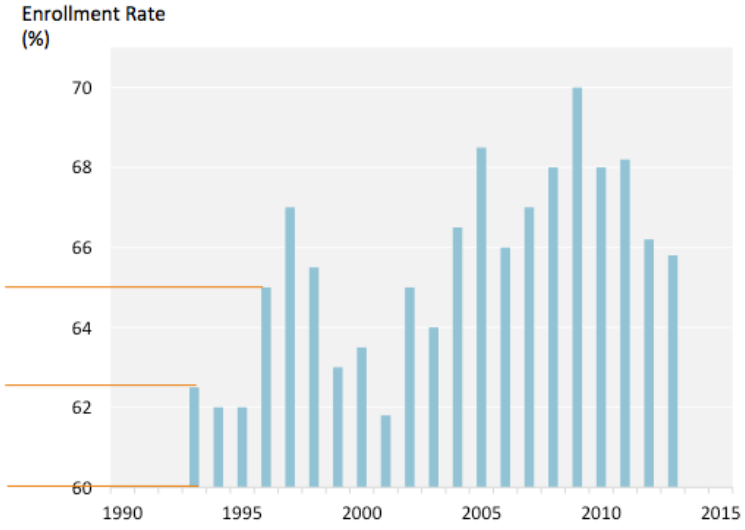
Visualization: Scatter Plots

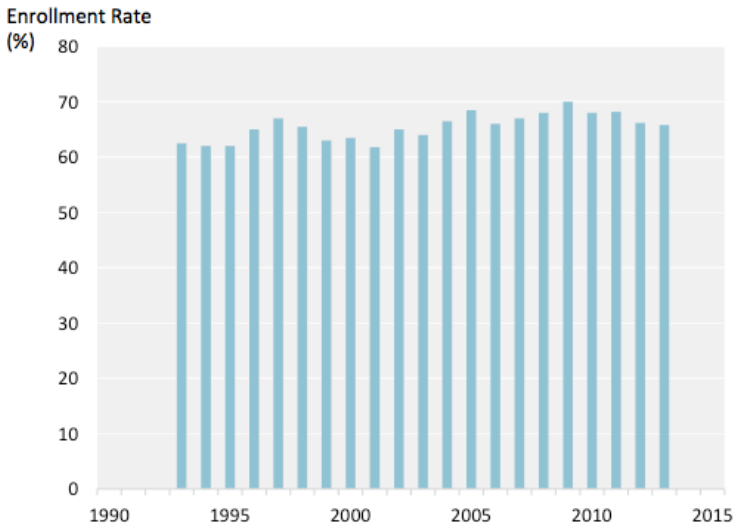


- Quantitative and Continuous:
  - Health care cost
  - Life Expectancy
- Continents (color)
  - Categorical

Quick Note on Zero

- Pay attention to axes… Sometimes they do not start at zero which can make things a bit misleading…





This is the same data but different start points for the Y-axis.

Historical Moment: Florence Nightingale

- **Florence Nightingale** (May 12, 1820 - August 13 1910)
  - "The Lady with the Lamp"
- Known for being the founder of modern nursing, serving as a nurse and training others in the Crimean War.
- In addition to saving lives directly as a nurse, Nightingale saved even more lives with her statistics.

Cause of Death for Soldiers in Crimean War

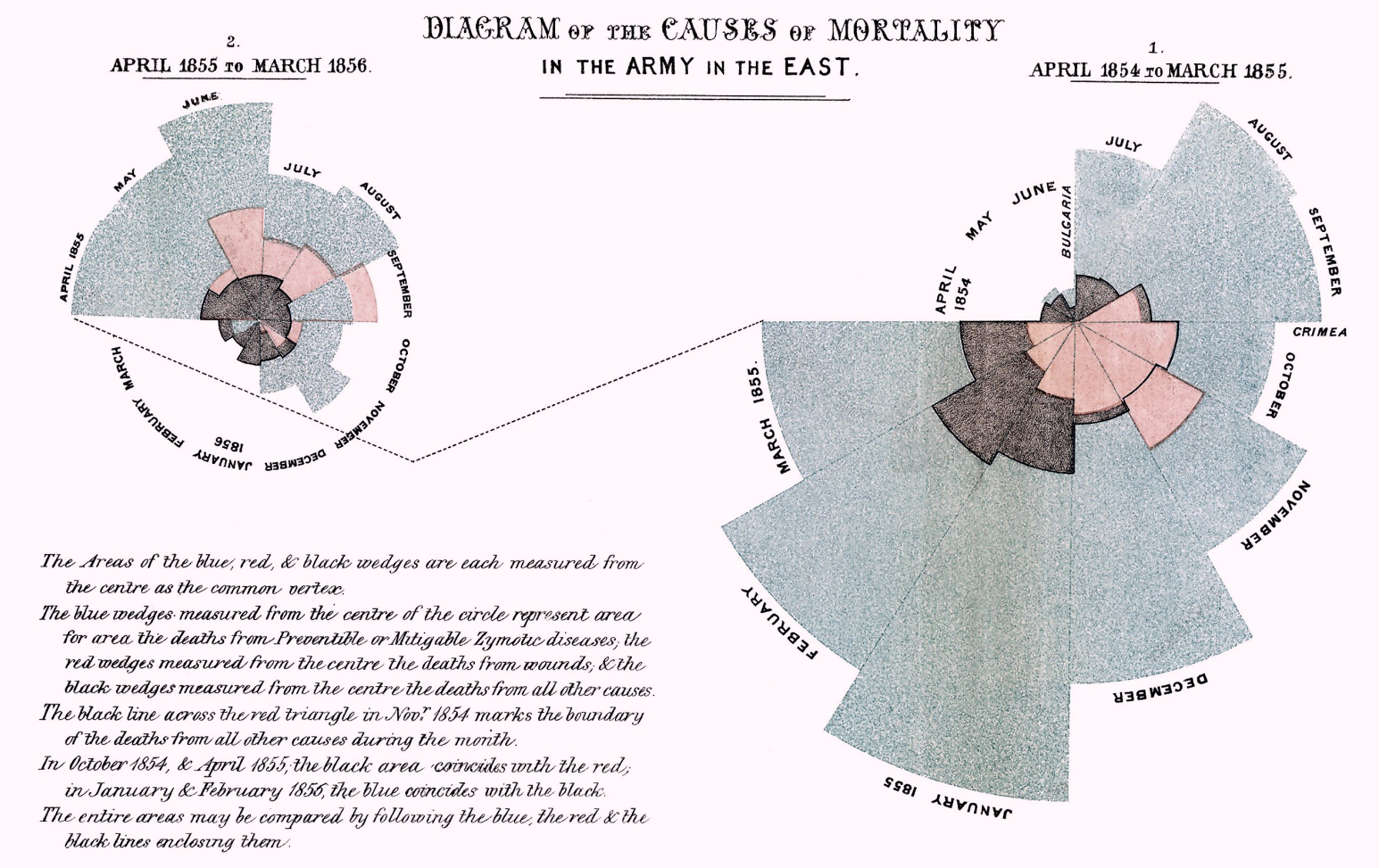
- At the time, most statistical work was presented in long tables, like the one below. All the information is there, but it is not readily digestible.

So Nightingale made a special graph…

	Zymotic Diseases.	Wounds and Injuries.	All other Causes.	Total
APRIL 1854	1.4	..	7.0	8.4
MAY	6.2	..	4.6	10.8
JUNE	4.7	..	2.5	7.2
JULY	150.0	..	9.6	159.6
AUGUST	328.5	.4	11.9	342.8
SEPTEMBER	312.2	32.1	27.7	372
OCTOBER	197.0	51.7	50.1	298.8
NOVEMBER	340.6	115.8	42.8	499.2
DECEMBER	631.5	41.7	48.0	721.2
JANUARY 1855	1022.8	30.7	120.0	1173.5
FEBRUARY	822.8	16.3	140.1	979.2
MARCH	480.3	12.8	68.6	561.7

Zymotic = related to infectious or contagious disease

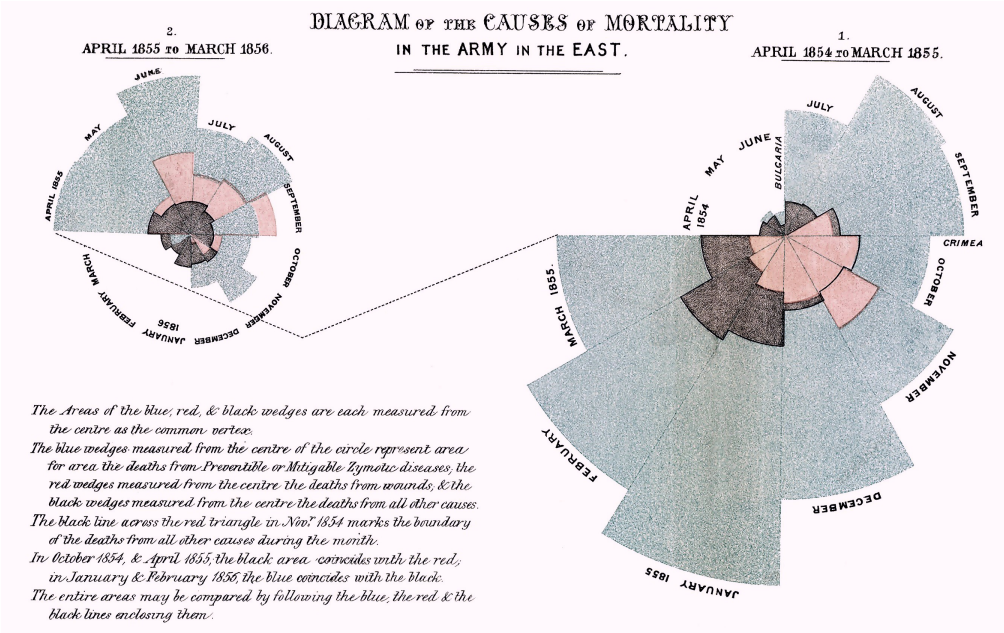
Polar Area Graph



Here is a link to an animated version of this visualization.

Polar Area Graph

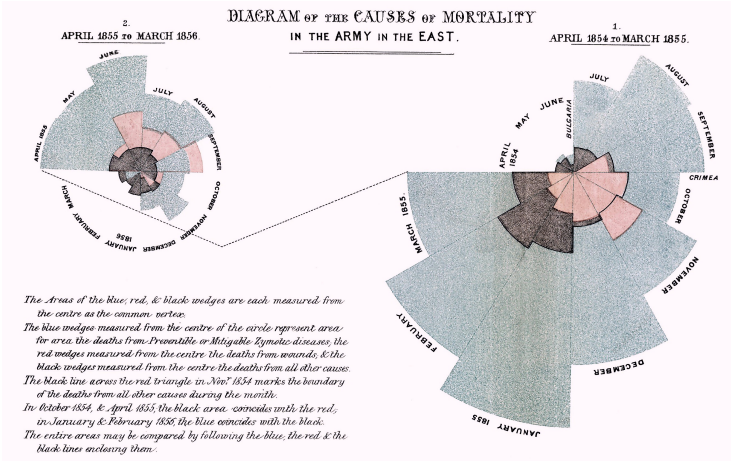
- The time during the war is split into months with wedges.
  - The red areas were deaths from wounds (like gunshots).
  - The black areas were deaths from accidents and "other" causes.
  - And the blue areas represent the deaths due to *preventable diseases*…



- There were a disproportionate number of disease caused deaths relative to the number of deaths from war injuries.

Saving Lives with Statistics

- This isn't just a pretty picture… Her irrefutable data and her ability to communicate those data through visualization revolutionized the hygiene and health care in hospitals. She saved more lives as a statistician than as a nurse by educating other health care professionals.



Historical Moment: Florence Nightingale

- "Statistics is the most important science in the whole world: for upon it depends the practical application of every other science and of every art: the one science essential to all political and social administration, all education, all organization based on experience, for it only gives results of our experience."-

  Florence Nightingale

Visualizations in R

So many options…!

- The data visualizations I showed here are some of the most common but are just a few of the *many* different data visualization options.
- R has many different tools to create visualizations. One of the most powerful is a package called "ggplot2"
  - This is beyond this scope of this class, but I *highly* encourage you to look into it!
  - For now, here is a link to a great R Graph Gallery