

Keith Allison

Worked on Alone

Q1: Plots

The scatterplot, Cleveland dot-plot, the QQ plot, and the coplot show every data point.

Q2: Plots

The histogram and the boxplot and the show aggregated or summarized data.

Q3: Conditioning Variables

A conditional variable is a variable that is used to split a plot of two variables by the value of a third variable. It can be used in the context of graphical data exploration to examine the interaction between three variables using a coplot.

Q4: Dispersion 1

Measures of spread

- Standard Deviation
- Interquartile range
- Coefficient of variation

Q5: Dispersion 2

Standard deviation is the root mean squared deviation from the mean. It describes the distribution of data around the mean. For normally distributed variables, + or - 1 standard deviation includes 68% of the values and + or - 2 standard deviations includes 95% of the values. Standard deviation can be used to understand the spread of the data because we know a majority of the values are within 1 standard deviation of the mean and almost all values are within 2 standard deviations of the mean.

The interquartile range is the range between the 25th and 75th quantiles of the data. Similarly to the standard deviation, it can tell us between what values a majority of the data are between.

Q6: Data Exploration

One data set I worked with included possible risk factors for cancer and cancer diagnoses. Data exploration was important to find the possible risk factors that actually had an effect on the rates of diagnosis. Box plots were invaluable for determining which factors had an effect because they allowed me to use continuous risk factor variables and binary cancer diagnosis variables and find the differences between those individuals who got cancer and those who did not.

Data exploration was important to find correlation between some of the predictor variables in the data set. Scatterplots were useful to find correlation between risk factors, which were usually continuous or discrete variables.