Victoria Haley

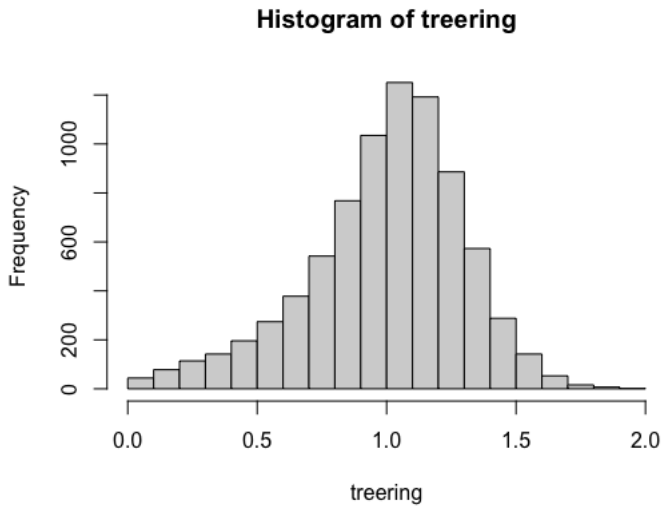
IST 772 Winter 2023 Week 1 homework

- Exercise 1
  - Mean: The sum of a group of values divided by the total number of values in that group. AKA the average.
  - Median: The value that is halfway between the start and end of a group of values.
  - Mode: The most common occurring value in a group/dataset.
  - Variance: The average "squared deviation" from the mean.
  - Standard Deviation: The square root of the Variance. SD is better for communicating because it is calibrated in the same units as your observation.
  - Histogram: A bar graph that shows the frequency distribution of values in a dataset.
  - Normal distribution: A bell-shaped distribution with symmetric tails and gradual curves toward a peak in the middle.
  - Poisson distribution: A distribution pattern with different shapes at different levels of lambda. Good for modelling arrival times

- Exercise 3
  - R-code:
    - data()
    - summary(Orange)
      - output: summary(Orange)
      - Tree age circumference
      - 3:7 Min. : 118.0 Min. : 30.0
      - 1:7 1st Qu.: 484.0 1st Qu.: 65.5
      - 5:7 Median :1004.0 Median :115.0
      - 2:7 Mean : 922.1 Mean :115.9
      - 4:7 3rd Qu.:1372.0 3rd Qu.:161.5
      - Max. :1582.0 Max. :214.0
  - For this exercise, I chose the 'Orange' data set.

  Of this dataset, only 'age' and 'circumference' are numerical variables.
  The Mean of the 'age' variable is 922.1 In technical terms, this means that the sum of all 35 age observations divided by 35 equals 922.1. In practical terms, this means that the average age of all orange trees in this data set is 922.1 years old.
  The Mean of the 'circumference' variable is 115.9. In technical terms, this means that the sum of all 35 circumference observations divided by 35 equals 115.9. In practical terms, this means that the average circumference of all orange trees in this data set is 115.9.
  The Median of the 'age' variable is 1004. This means that when looking at the observations of the age variable in order, 1004 is the halfway point.
  The Median of the 'circumference' is 115. This means that when looking at the observations of the circumference variable in order, 115 is the halfway point.

- Exercise 4
    - R-code:
        - data()
        - hist(treering)
            - output:



- For this exercise, I chose the 'treering' data set.
- The shape of this histogram can be described as "bell-shaped". The Normal Distribution shape best describes this histogram. I came to this conclusion based on the bell-shaped curve of this graph, with the roughly symmetric tails and gradual curve towards the peak in the middle.