Homework 1
Syracuse University
IST 772
Summer 2021

Question 1

Mean
The mean is a measure of central tendancy that represents where most of the data is or where the data typically is. It is calculated by taking the sum of all of the data points divided by the number of data points.

Median
The median is a measure of central tendancy that represents where the middle of the data is. It is calculated by ordering the data points in order of least to greatest and then taking the data point in the middle. If there are an even number of data points, the mean of the two middle data points are used. If there are an odd number of data points, there will be one data point in the middle.

Mode
The mode is a measure of central tendancy that represents the most frequently occuring value out of a set of data points. In the case of a tie, there can be more than one mode.

Variance
The variance is a measure of dispersion that represents the average squared deviance of a set of data points. It is calculated by taking the sum of squares of a set of data points and dividing by the number of data points.

Standard Deviation
The standard deviation is a measure of dispersion that represents the square root of the variance. It is calculated by taking the square root of the variance.

Histogram
A histogram is a chart that represents the frequency distribution of a set of data points. The data points are bucketed into a certain number of bins, and then the count of the number of data points within a certain bin falls on the y axis, and the bins fall on the x axis.

Normal Distribution
The normal distribution describes the shape of a distribution. When the distribution follows the shape of a bell curve, it is said to be a normal distribution.

Poisson Distribution
The poisson distribution describes the shape of a distribution. This distribution is commonly associated with time occurring between two events.

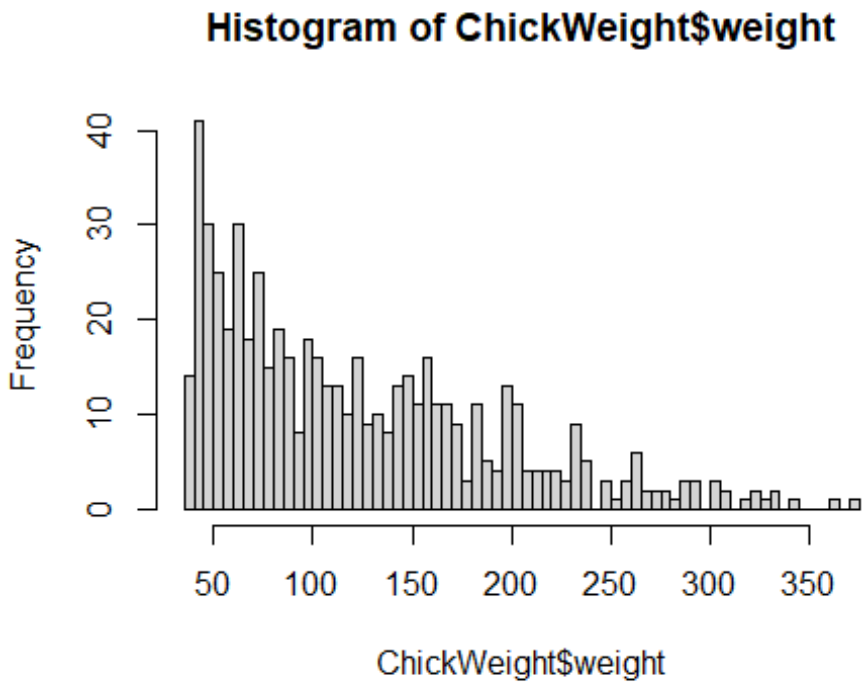
Question 3

```
# choose a data set that contains at least one numeric variable
mtcars <- mtcars

# summarize the mpg variable from the mtcars dataset
summary(mtcars$mpg)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   15.43   19.20   20.09   22.80   33.90

# The mean of the mpg variable in the mtcars dataset is 20.09. The mean is th
e
# sum of the data points divided by the number of data points. Here is the
# calculation of the mean the long way.

# here is a vector shwoing the data points in the mpg variable
mtcars$mpg

##  [1] 21.0 21.0 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2
10.4
## [16] 10.4 14.7 32.4 30.4 33.9 21.5 15.5 15.2 13.3 19.2 27.3 26.0 30.4 15.8
19.7
## [31] 15.0 21.4

# here is the sum of the data points in the mpg variable
sum(mtcars$mpg)

## [1] 642.9

# here is the number of data points in the mpg variable
length(mtcars$mpg)

## [1] 32

# here is the sum of the data points divided by the number of data points
sum(mtcars$mpg) / length(mtcars$mpg)

## [1] 20.09062

# The median of the mpg variable in the mtcars dataset is 19.20. The median
# is the middle point of the data points.Since there are an even number of da
ta
# points, 32 data points, as we saw from earlier, the median will take the
# mean of the two middle data points.

# here are the ordered data points from least to greatest
sort(mtcars$mpg)

##  [1] 10.4 10.4 13.3 14.3 14.7 15.0 15.2 15.2 15.5 15.8 16.4 17.3 17.8 18.1
18.7
## [16] 19.2 19.2 19.7 21.0 21.0 21.4 21.4 21.5 22.8 22.8 24.4 26.0 27.3 30.4
```

```
30.4
## [31] 32.4 33.9

# the median is the mean of numbers at positions 16 and 17
mean(sort(mtcars$mpg) [16], sort(mtcars$mpg) [17])

## [1] 19.2
```

Question 4

```
# choose a data set that contains at least one numeric variable
chickweight <- ChickWeight

# create a histogram of the weight variable from the ChickWeight dataset
hist(ChickWeight$weight, breaks = 50)
```



The chick weights seem to follow the shape of a poisson distribution. It fits this distrubtion because it is right skewed and tails off to the right and does not have a left tail.