Illapani_Lab1

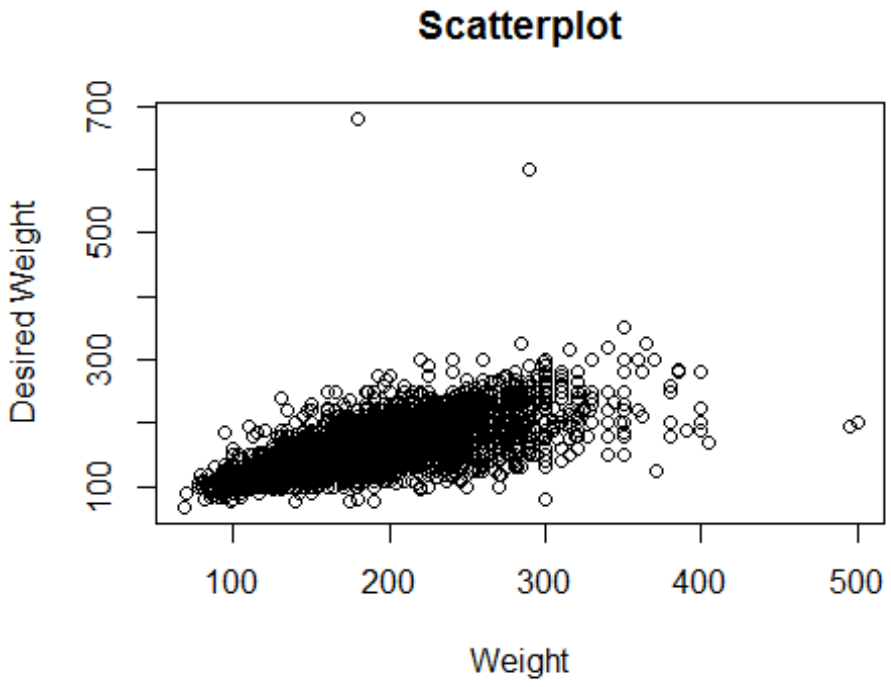
Srini Illapani

September 6, 2015

Lab1

1. Make a scatterplot of weight versus desired weight. Describe the relationship between these two variables.
   - The plot reveals a corelation between the variables weight and desired weight. The higher the weight the lower the desired weight the population wishes.

```
source("more/cdc.R")
plot(cdc$weight, cdc$wtdesire, main="Scatterplot",
     xlab="Weight", ylab="Desired Weight")
```



2. Let's consider a new variable: the difference between desired weight (`wtdesire`) and current weight (`weight`). Create this new variable by subtracting the two columns in the data frame and assigning them to a new object called `wdiff`.
   - Wdiff calulation using R

```
wdiff <- cdc$weight - cdc$wtdesire
head(wdiff)
```

```
## [1]  0 10  0  8 20  0
```

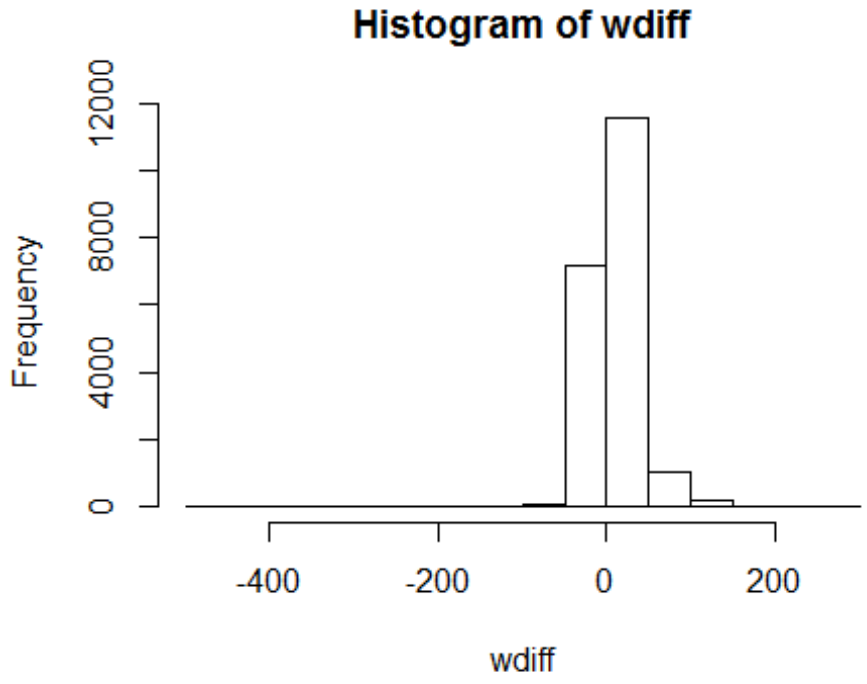
3. What type of data is `wdiff`? If an observation `wdiff` is 0, what does this mean about the person's weight and desired weight. What if `wdiff` is positive or negative?

   - 'wdiff' is of type Numerical and discrete. If wdiff is 0, the person does not desire to either reduce or increase the current weight. If wdiff is positive the person desires to reduce the weight, if wdiff is negative, the person desires to increase their weight.

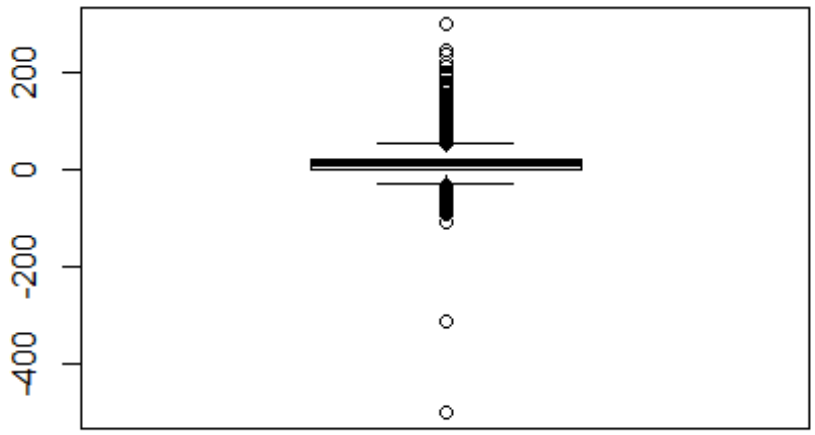
4. Describe the distribution of `wdiff` in terms of its center, shape, and spread, including any plots you use. What does this tell us about how people feel about their current weight?

   - The distribution of 'wdiff' is symmetric. The Histogram shows most of the population looking to reduce their weight on an average by 14.5 lbs. There are very few outliers. Most of the people desire to reduce their weight than those who desire to gain weight or keep same weight.

```
wdiff <- cdc$weight - cdc$wtdesire
hist(wdiff)
```

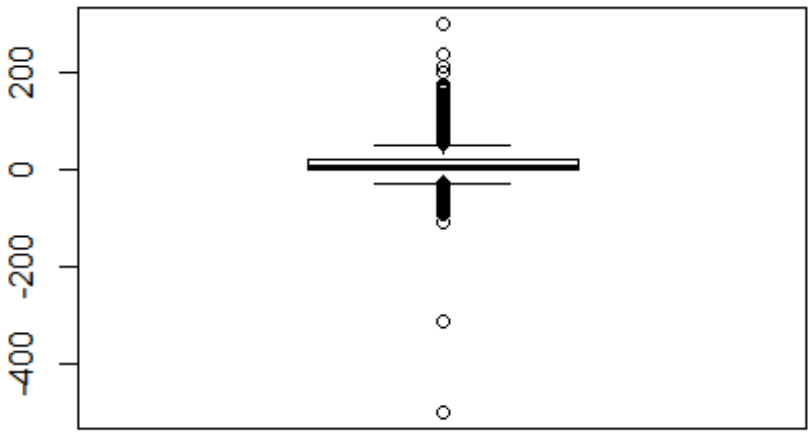



```
boxplot(wdiff)
```

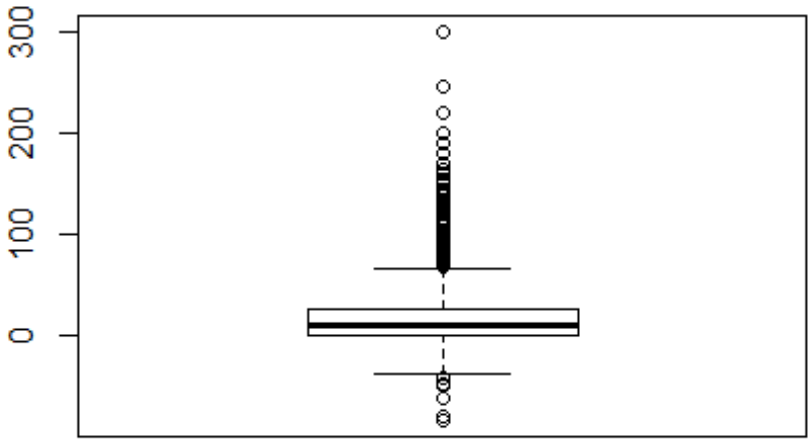


5. Using numerical summaries and a side-by-side box plot, determine if men tend to view their weight differently than women.
   - The box plots for men and women for their weight difference (weight - desired weight) looks similar except for some outliers on the lower end for the mens graph. We see men who would like to gain 300 and 500 lbs, this could be data error.
   - The summary measures reveal than women on an average would like to shed almost 8 lbs more compared to men.

```
mdata <- subset(cdc, cdc$gender == "m")
m_wdiff <- mdata$weight - mdata$wtdesire
boxplot(m_wdiff)
```




```
wdata <- subset(cdc, cdc$gender == "f")
w_wdiff <- wdata$weight - wdata$wtdesire
boxplot(w_wdiff)
```



```
summary(m_wdiff)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -500.00    0.00    5.00   10.71   20.00  300.00
```

```
summary(w_wdiff)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -83.00    0.00   10.00   18.15   27.00  300.00
```

6. Now it's time to get creative. Find the mean and standard deviation of `weight` and determine what proportion of the weights are within one standard deviation of the mean.
   - Mean = 169.7, SD = 40.08
   - We see that all the weights are within 4 standard deviations of the mean. So the proportion of the weights that is within one standard deviation of the mean is around 25%.

```
mean(cdc$weight)
```

```
## [1] 169.683
```

```
sd(cdc$weight)
```

```
## [1] 40.08097
```

```
hist(cdc$weight, breaks = 40)
```