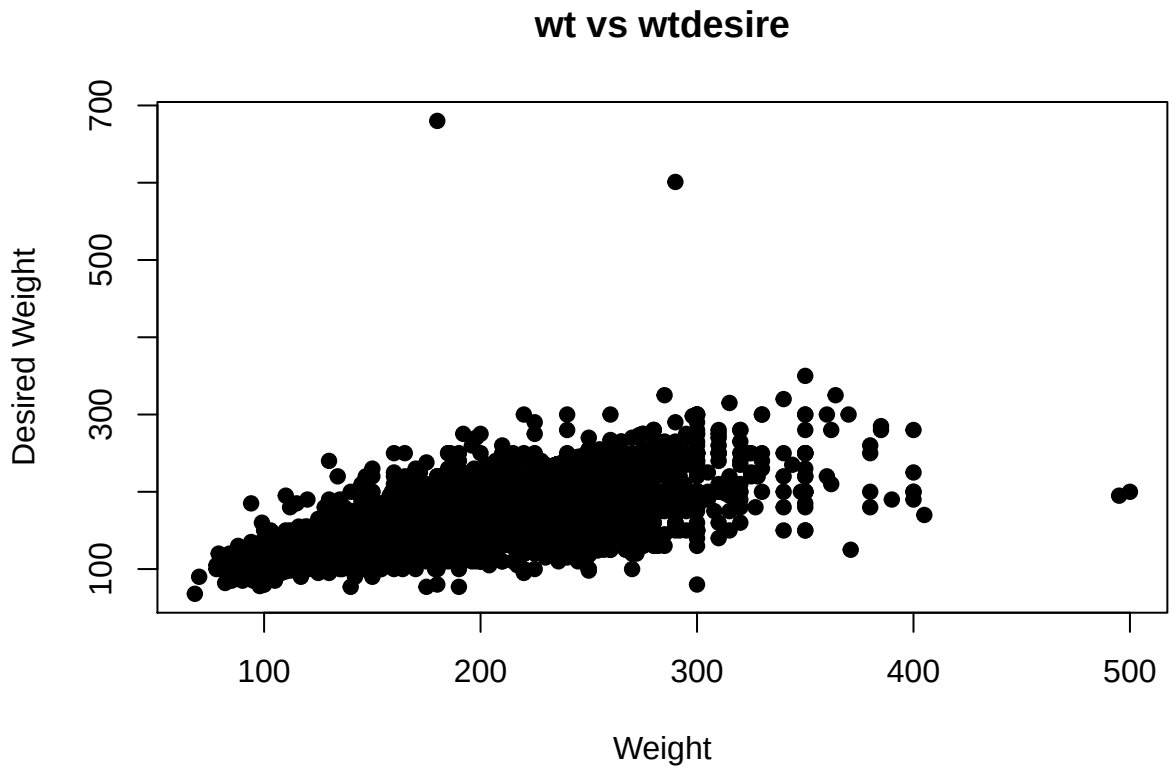
Kumar-Lab1

sanjivek

September 5, 2015

1) Make a scatterplot of weight versus desired weight. Describe the relationship between these two variables.

```
source("c:/Users/sanjivek/Documents/Lab1/more/cdc.R")
attach(cdc)
plot(weight,wtdesire, main="wt vs wtdesire",xlab="Weight", ylab="Desired Weight", pch=19)
```




```
cor(weight,wtdesire)
```

```
## [1] 0.8000521
```

Conclusion: There is a Positive strng correlation between weight and wtdesire as the correlation Coefficient is 0.80 which is near to 1 and scatterplot also shows that in the diagram.

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<-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?

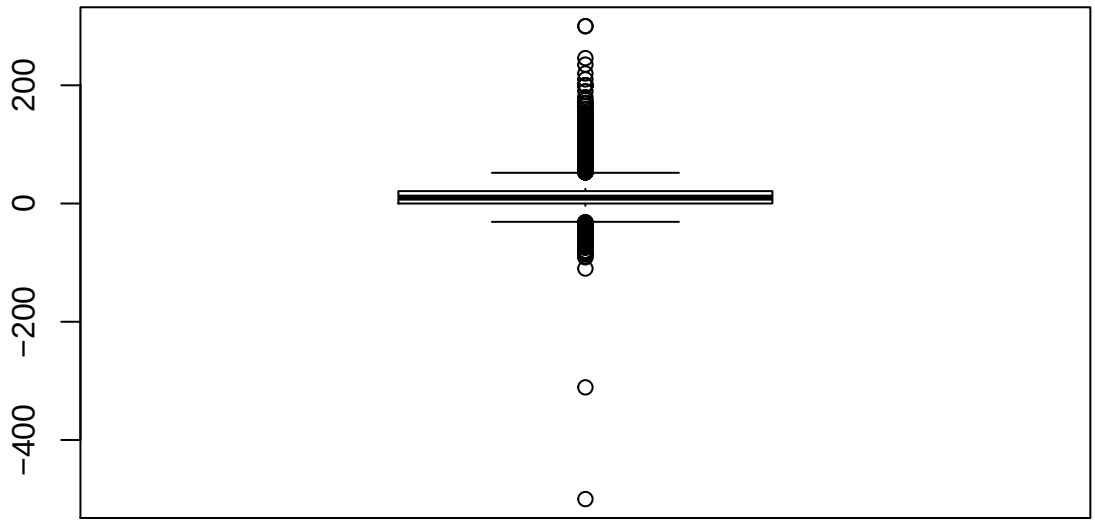
```
typeof(wdiff)
```

```
## [1] "integer"
```

Conclusion: Wdiff is an integer, If Wdiff is 0 then Weight and desirde weight are same, If Postive then Overweight and Negative then underweight.

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?

```
boxplot(wdiff)
```



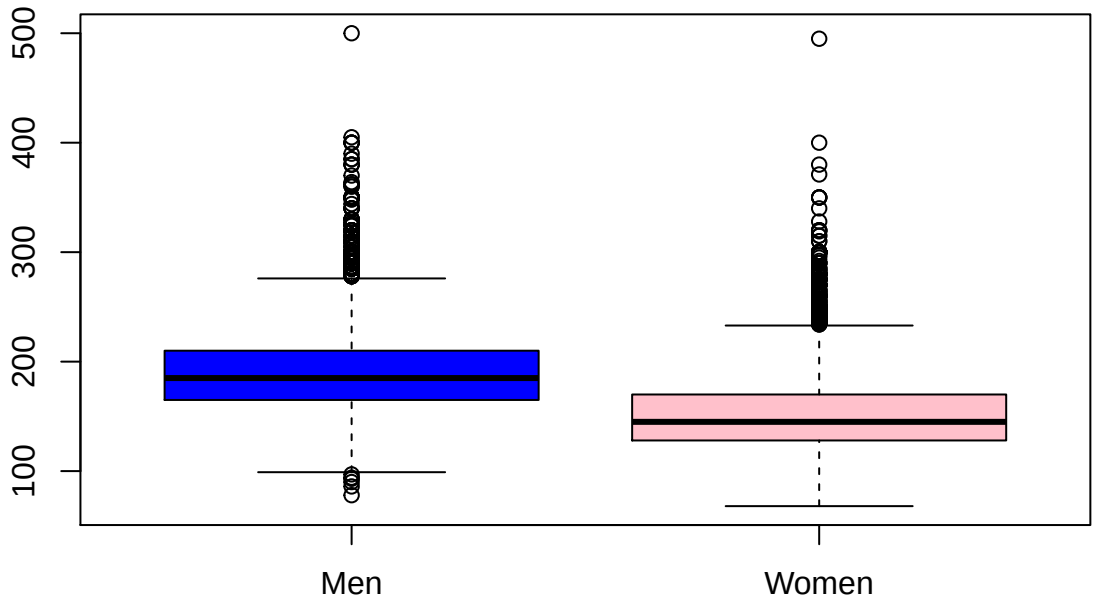

```
summary(wdiff)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -500.00    0.00   10.00   14.59   21.00  300.00
```

Conclusion: With density plot and Box Plot it shows that 50% of the people are overweight between 1 to 21 lbs and 25% people are overweight upto 52.5lbs with few outliers in overweight. 25% people are underweight upto 31.5 lbs with 3 outliers as shown below.

4) Using numerical summaries and a side-by-side box plot, determine if men tend to view their weight differently than women.

```
men <- subset(cdc$weight, gender == "m")
women <- subset(cdc$weight, gender == "f")
boxplot(men,women,names=c("Men","Women"),
col=c("Blue","Pink"))
```




```
summary(men)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    78.0   165.0   185.0   189.3   210.0   500.0
```

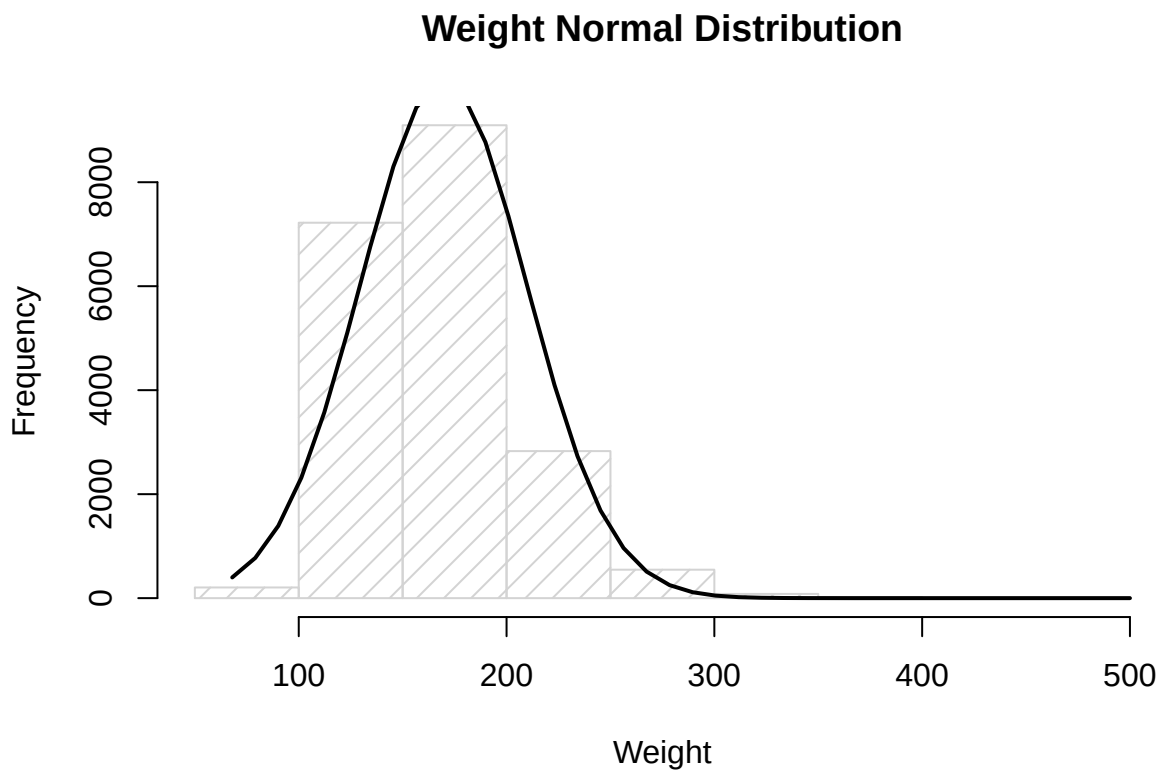
```
summary(women)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    68.0   128.0   145.0   151.7   170.0   495.0
```

Conclusion: Women tend to be 40 lbs less heavier than men and overall median weight is 145 lbs compared to men at 185 lbs.

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.

```
g<-cdc$weight
h<-hist(g, breaks=10, density=10, col="lightgray", xlab="Weight", main="Weight Normal Distribution")
  xfit<-seq(min(g),max(g),length=40)
  yfit<-dnorm(xfit,mean=mean(g),sd=sd(g))
  yfit <- yfit*diff(h$mids[1:2])*length(g)
  lines(xfit, yfit, col="black", lwd=2)
```




```
mn<- mean(g)
stnd<-sd(g)

# One Standard Deviation is mean +/-sd
LowProp<-mn-stnd
highProp<-mn+stnd
# Weight proportions for one standard deviations are
LowProp
```

```
## [1] 129.602
```

```
highProp
```

```
## [1] 209.7639
```