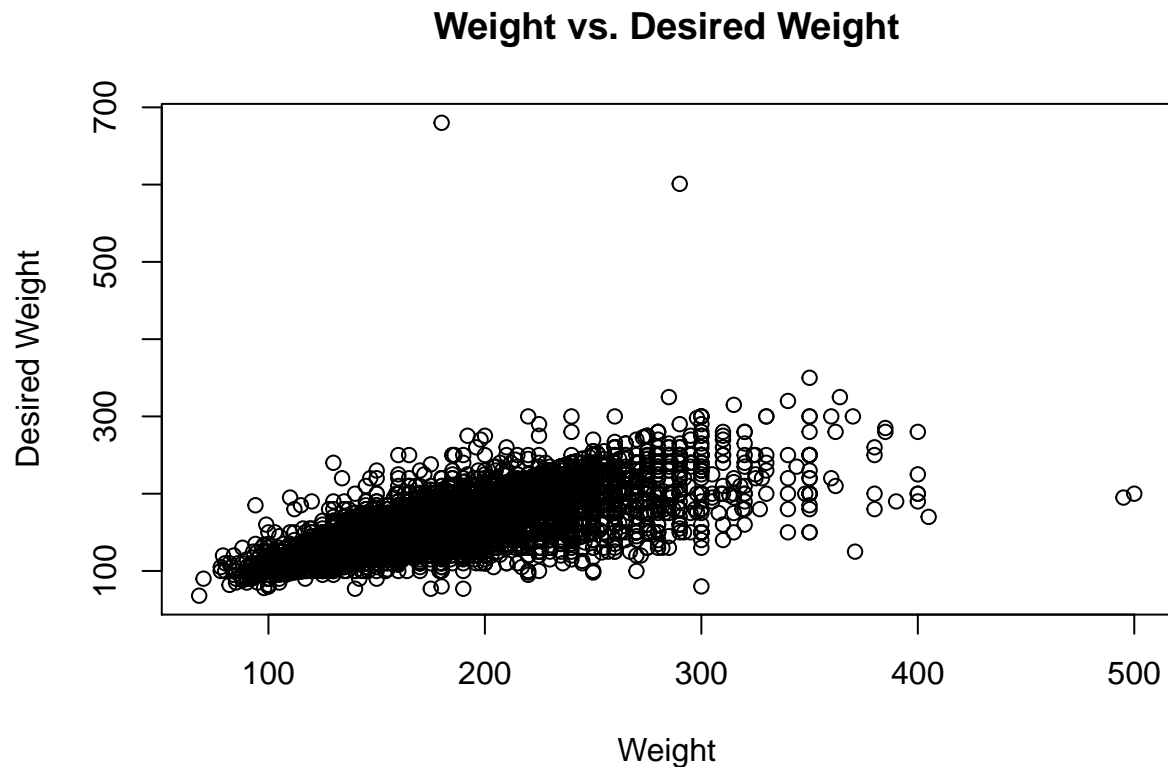# IS606 Lab 1

*Aadi Kalloo*

*Due February 7, 2016*

**Question 1:**

```
source("http://www.openintro.org/stat/data/cdc.R")
plot(x = cdc$weight, y = cdc$wtdesire, type = "p", xlab = "Weight", ylab = "Desired Weight")
title("Weight vs. Desired Weight")
```



It can be seen that as weight increases, desired weight also increases.

**Question 2:**

```
cdc$wdiff <- cdc$wtdesire - cdc$weight
```

**Question 3:**

```
str(cdc$wdiff)
```

```
##  int [1:20000] 0 -10 0 -8 -20 0 -9 -10 -20 -10 ...
```
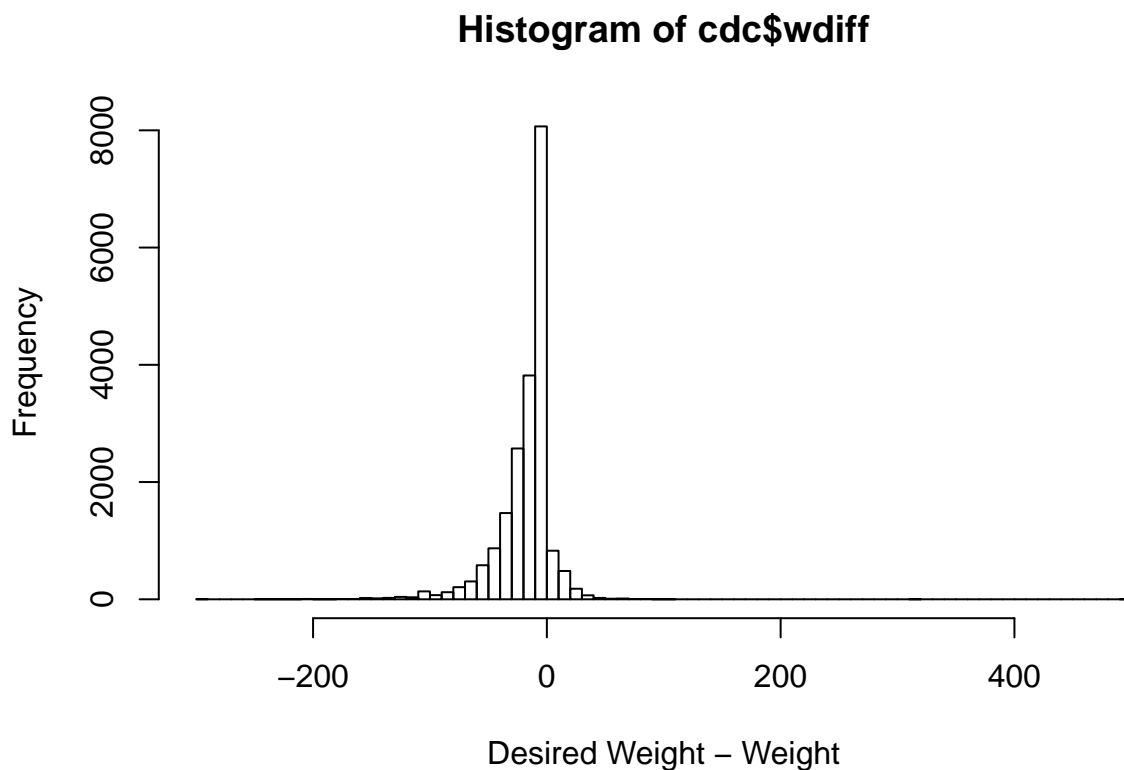
```
summary(cdc$wdiff)
```

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

The *wdiff* variable contains numeric (integer) data that is continuous. It has contains a minimum of -300 and a maximum of 500. If a value is negative, it means that a participant's weight is greater than their desired weight and, therefore, they want to lose weight. Conversely, if *wdiff* contains a positive value, it means the participant wants to gain weight.
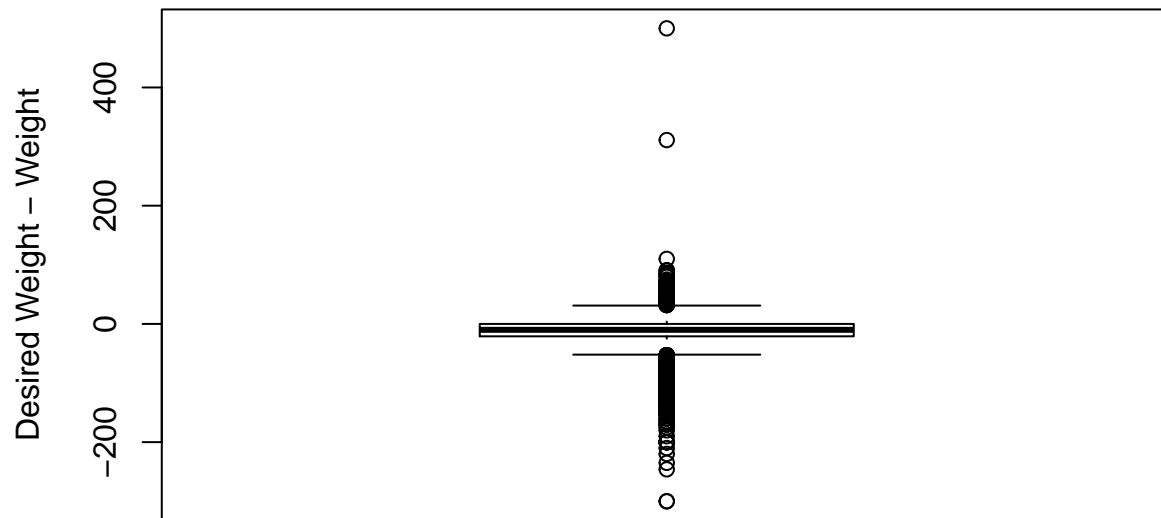
**Question 4:**

```
hist(cdc$wdiff, breaks=100, xlab = "Desired Weight - Weight")
```

**Histogram of cdc$wdiff**



```
boxplot(cdc$wdiff, ylab = "Desired Weight - Weight")
title("Boxplot Distribution of wdiff")
```

# Boxplot Distribution of wdiff



```r
summary(cdc$wdiff)
```

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

```r
IQR(cdc$wdiff)
```

```
## [1] 21
```

This distribution is negatively skewed with a center at -10.0. It's interquartile range is 21, with the 3rd quartile at 0.0 and 1st quartile at -21. A significant number of values lie below the 3rd quartile, resulting in its skewed character. We can see from this information that most people want to lose weight.

**Question 5:**

```r
boxplot(cdc$wdiff ~ cdc$gender)
title("Gender-stratified views on weight loss")
```

## Gender–stratified views on weight loss
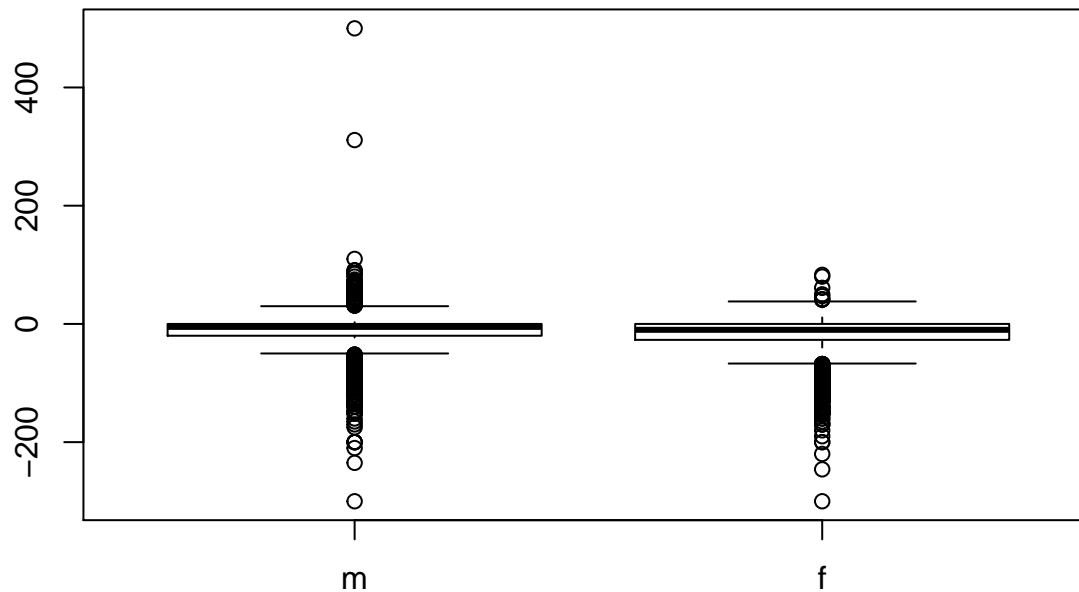


```
summary(cdc$wdiff[cdc$gender=="m"])
```

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

```
summary(cdc$wdiff[cdc$gender=="f"])
```

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

```
IQR(cdc$wdiff[cdc$gender=="m"])
```

```
## [1] 20
```

```
IQR(cdc$wdiff[cdc$gender=="f"])
```

```
## [1] 27
```

It seems as though men do tend to view their weight differently than women. The male distribution is centered around -5.0 while the female distribution is centered around -10.0. This indicates that the typical female desires to lose more weight than the typical male. Females also have a greater spread than males, which could indicate varying perceptions amongst females regarding weight loss. Furthermore, males have a greater number who want to gain weight when compared to females. This can most immediately be attributed to muscle-centered weight gain, which is quite common among males.

**Question 6:**

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

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

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

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

```
cdc$wtZscore <- (cdc$weight - mean(cdc$weight)) / (sd(cdc$weight))
wtproportion <- length(cdc$wtZscore[cdc$wtZscore <= 1]) / length(cdc$weight)
text <- "The proportion of weights within 1 standard deviation of the mean is "
print(paste0(text, wtproportion))
```

```
## [1] "The proportion of weights within 1 standard deviation of the mean is 0.84675"
```