

## Week 2 Assignment

laknath Gunathilake

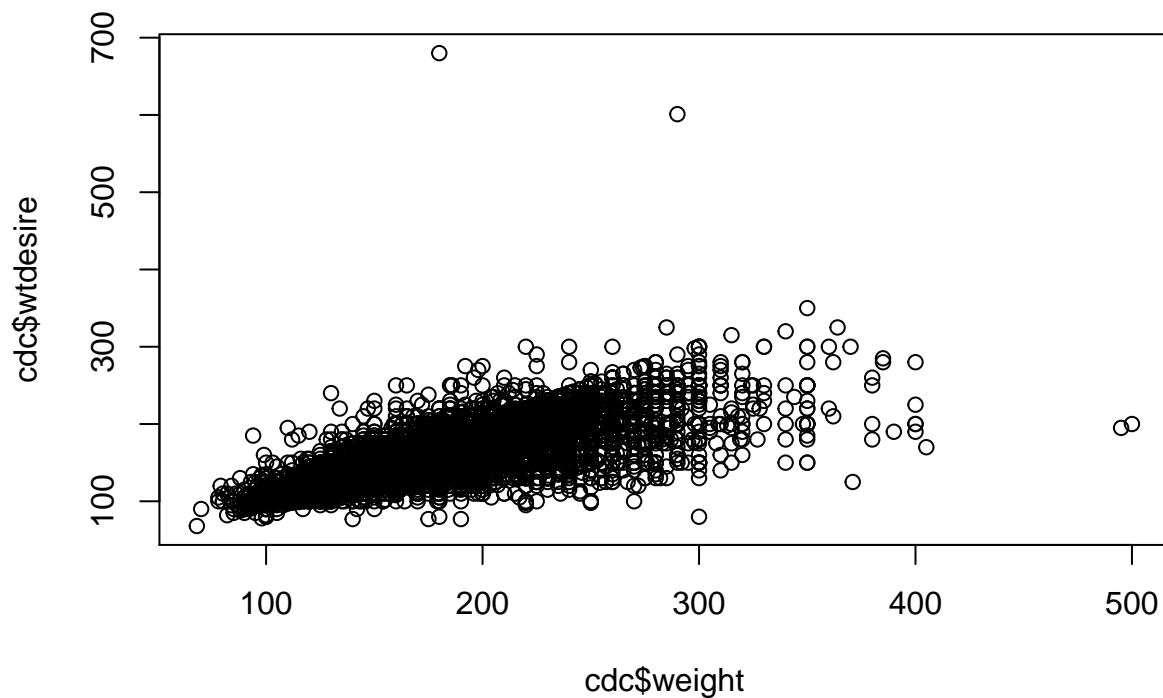
January 15, 2020

```
source("http://www.openintro.org/stat/data/cdc.R")
```

### NOW ON YOUR OWN

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

```
plot( cdc$weight , cdc$wtdesired )
```



The actual weight and desired weight seem to have a strong association between 75 and 275lb. However, this relationship is weaker for those who are above 300lb. This would make sense since, we can assume that those who are overweight/obese hope to lose weight, and thus have a desired weight that is lower than their current weight.

2. Let's consider a new variable: the difference between desired weight (wtdesired) 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$wtdesired - cdc$weight
```

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?

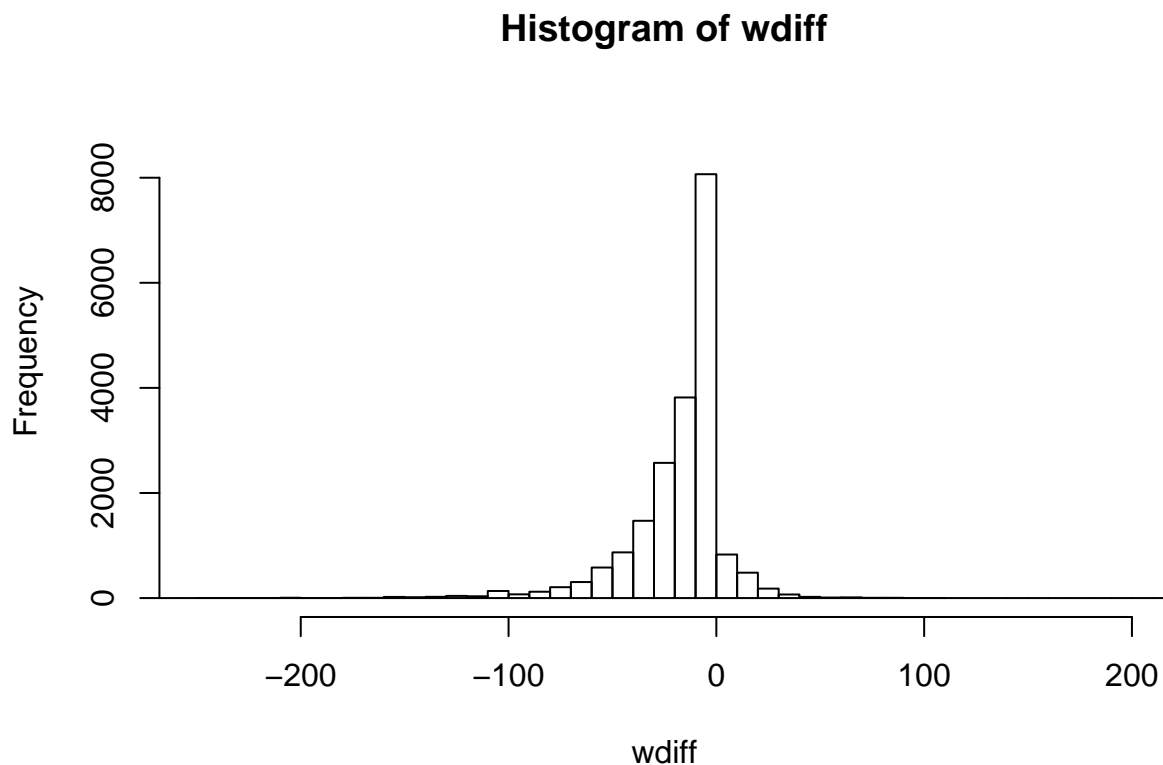
```
class(wdiff)
```

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

- The class of Data is an Integer
- the actual weight and the desired weight are the same: or a person weighs exactly how much they want to weigh.
- negative wdiff= This means that the actual weight is higher than desired weight (They would hope to lose weight)
- positive wdiff= This means that the actual weight is lower than the desired weight (they would hope to gain weight)

4. Describe the distribution of wdiff in terms of its center, shape, and spread, including any plots you use.

```
hist(wdiff, breaks = 75, xlim = c(-250, 200), ylim = c(0, 9000))
```



```
mean(wdiff)
```

```
## [1] -14.5891
```

```
median(wdiff)
```

```
## [1] -10
```

```
min(wdiff)
```

```
## [1] -300
```

```
max(wdiff)
```

```
## [1] 500
```

```
library(moments)
```

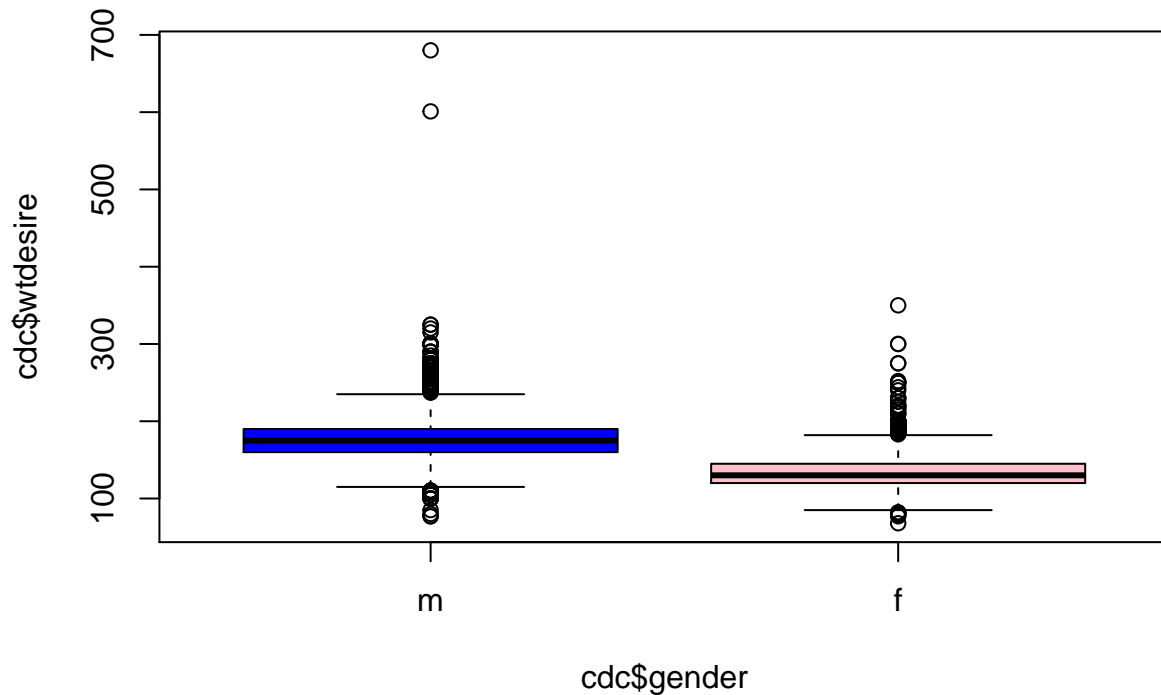
```
skewness(wdiff)
```

```
## [1] -1.447632
```

- the wdiff slightly left skewed
- wdiff has mean desired weight of -14.5 and a median desired weight of -10. Based on this we can conclude that most people are unhappy about their weight (They hope to have a weight lower than their current weight)

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

```
boxplot(cdc$wtdesired~cdc$gender,col=(c("blue","pink")))
```



```
tapply(cdc$wt desire, cdc$gender, mean, na.rm=TRUE)
```

```
##           m           f  
## 178.6166 133.5150
```

```
tapply(cdc$wt desire, cdc$gender, median, na.rm=TRUE)
```

```
##    m    f  
## 175 130
```

- It is evident that men tend to have a higher desired weight on average than women. On average, men seem to want to be around 178lb while women hope to be around the 133lb mark.

**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(cdc$weight)
```

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

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

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

```
within_one_negative<-mean(cdc$weight)-sd(cdc$weight)
```

```
within_one_positive<-mean(cdc$weight)+sd(cdc$weight)
```

```
within_one_sd<-cdc$weight[ cdc$weight >= within_one_negative & cdc$weight <= within_one_positive]  
length(within_one_sd)/length(cdc$weight)*100
```

```
## [1] 70.76
```

- 70 percents of the weights are within one standard deviation of the mean weight which is 169.68295