Lab1

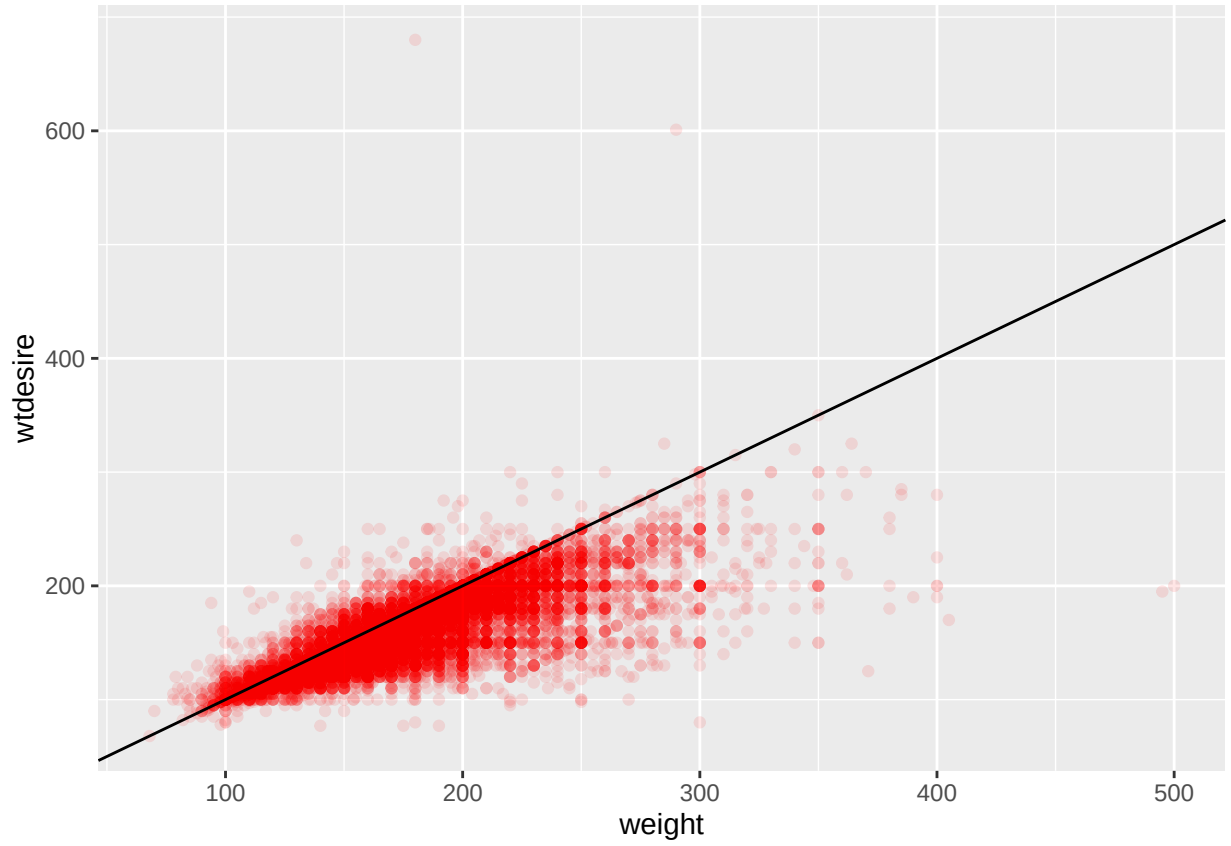
Ahsanul Choudhury

September 4, 2016

```
source("more/cdc.R")
```

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

```
ggplot(cdc, aes(x=weight, y=wtdesire)) + geom_point(alpha=1/10, colour="red") +   geom_abline(intercept=
```




```
##There is a positive realtionship between the variables
```

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`.

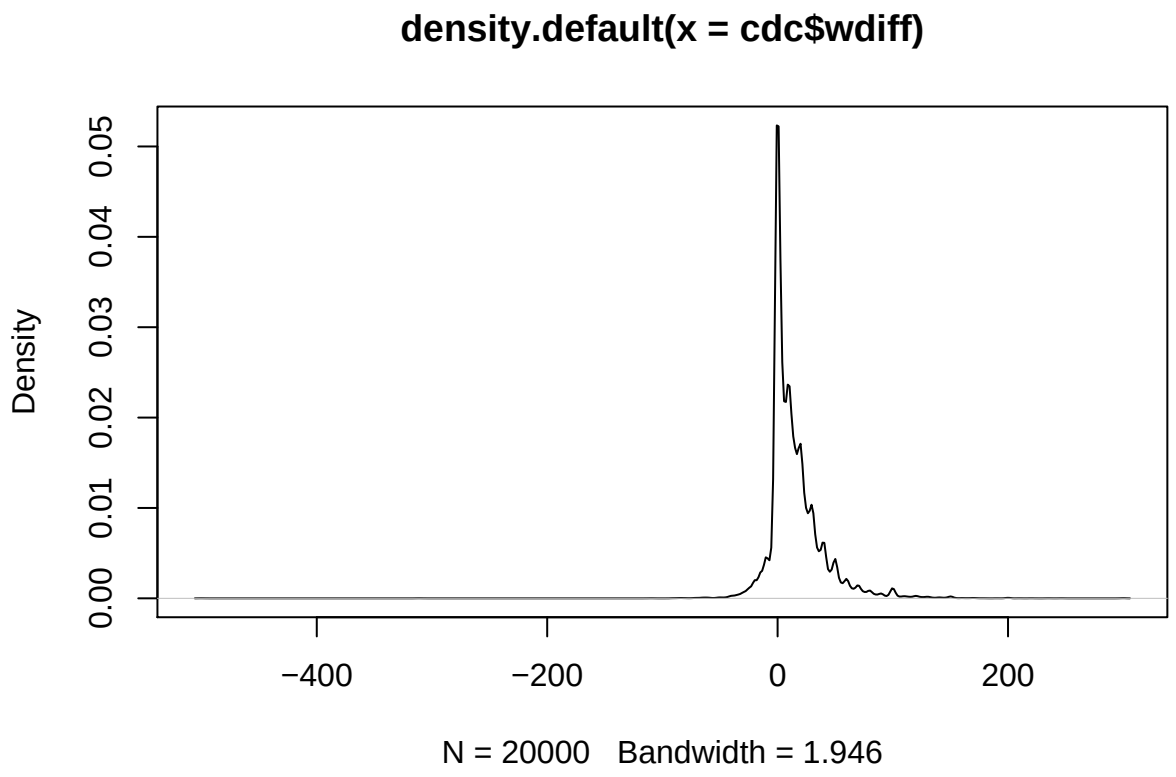
```
cdc$wdiff <- cdc$weight - cdc$wtdesire
```

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 numarical data. If `wdiff` is that means the person's current weight is their desired weight. if `wdiff` is positive that means the person wants to loose weight and if negative the person wants to gain weight.

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?

```
den <- density(cdc$wdiff)
plot(den)
```




```
##The density plot has a one clear peak so its a unimodal distribution and it is centered around 0.
```

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

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

```
##From the density plot and summary we can see majority or atleast 1/3 of the participent claims to be s
```

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

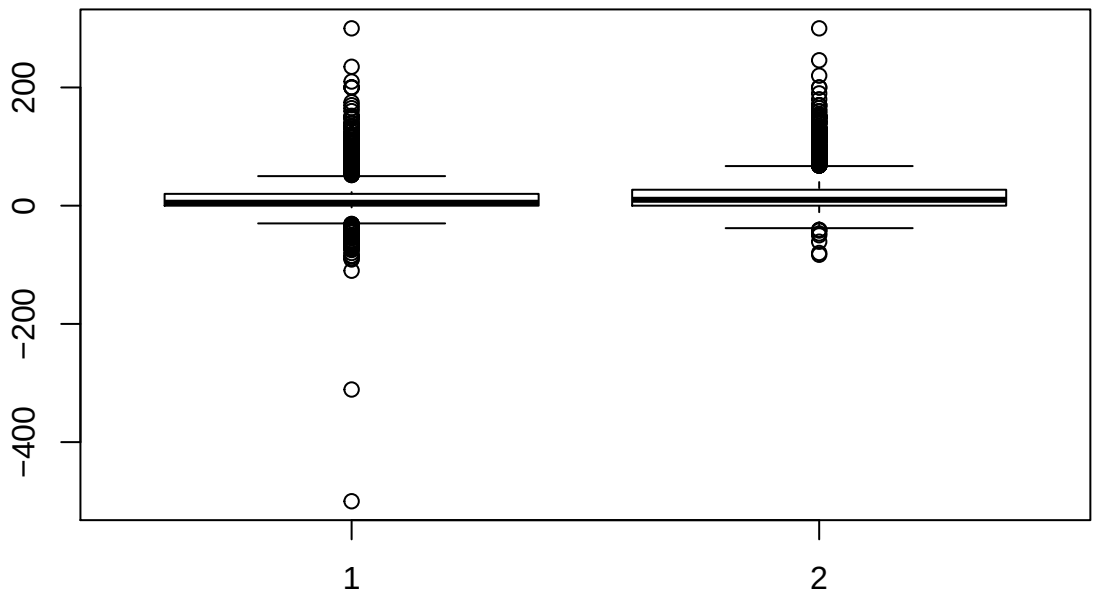
```
mwdiff <- cdc$gender=="m"
M <- cdc[mwdiff,]
summary (M$wdiff)
```

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

```
fwdiff <- cdc$gender=="f"
F <- cdc[fwdiff,]
summary (F$wdiff)
```

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

```
boxplot(M$wdiff,F$wdiff)
```




```
##Male participent has mean weight loss terget of 10.71 which is lower than the female mean target of 18
```

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.

```
mwt <- round(mean(cdc$weight), 2)
mwt
```

```
## [1] 169.68
```

```
#mean weight is 169.68
```

```
stdv <- sd(cdc$weight)
stdv
```

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

```
#standard deviation 40.08097
```

```r
onestdv <- subset(cdc, weight < (mwt + stdv) & weight > (mwt - stdv))
dim(onestdv)[1]/dim(cdc)[1]
```

```
## [1] 0.7076
```

```r
#proportion of weight within one standard diviation of mean are 0.7076
```