

```

#Getting started
source("http://www.openintro.org/stat/data/cdc.R")
#Exercise 1. How many cases are there in this data set?
#Exercise 1-1,1-2: 20000 cases and 9 variables
dim(cdc)
#Exercise 1-3: 4 discrete data, 5 categorical.
str(cdc)

```

```

#Summaries and tables
summary(cdc$weight)
smoke = table(cdc$smoke100);smoke
barplot(smoke)
mosaicplot(table(cdc$gender,cdc$smoke100))

```

```

#Exercise 2-1
height = cdc$height
# IQR of height
IQR(height)
# IQR of age
age = cdc$age
IQR(age)

```

```

#Exercise 3
#male seems to have higher smoking rate than female.

```

```

#Quantitative data
bmi = (cdc$weight/cdc$height^2)*703
boxplot(bmi~cdc$gender)
hist(bmi,breaks=50)

```

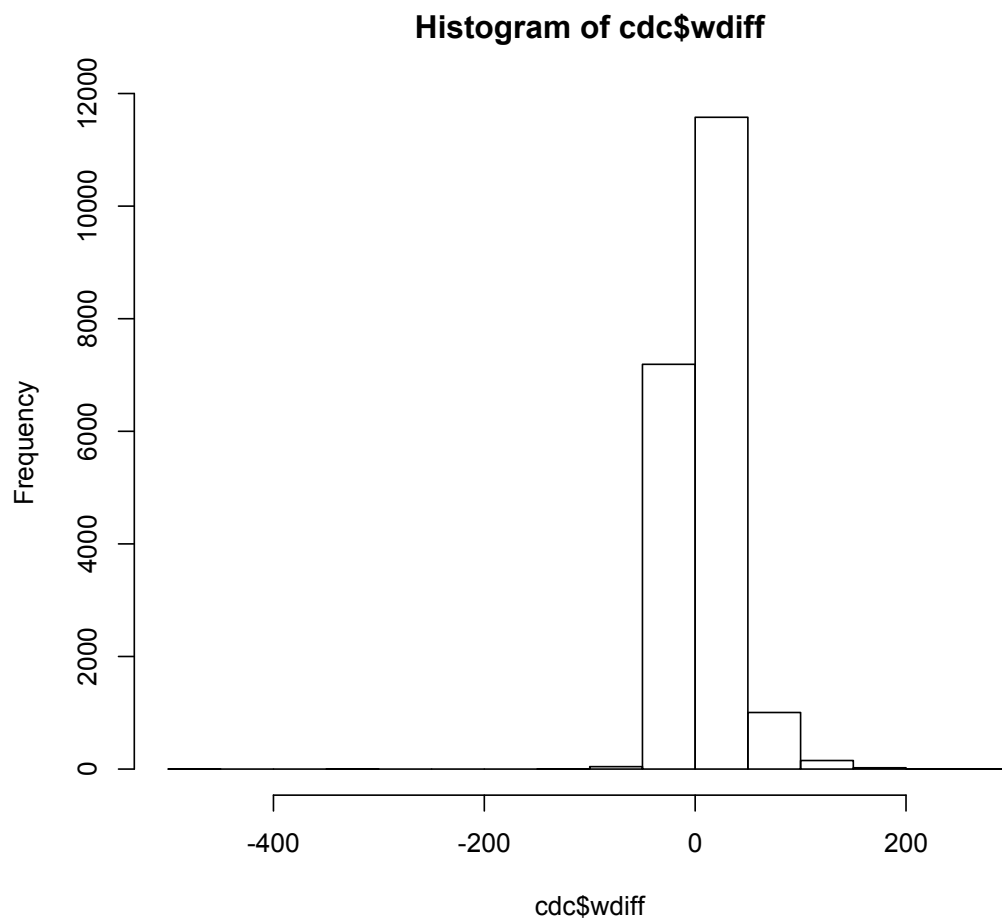
```

#####Assignment#####
#QBS Assignment 1
#Part A.
#Q1
cdc$wdiff = cdc$weight- cdc$wtdesired

```

#Q2

```
hist(cdc$wdiff)
```



# from the graphics we can see more people are weighed a bit higher than they wish to be.

```
prop.table(table(cdc$wdiff>0))
```

```
FALSE  TRUE
```

```
0.3618 0.6382
```

#Also, from the above calculation, we could conclude that the 64% people have higher weight than they desire to have.

#Q3

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

```
[1] 169.683
```

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

```
[1] 40.08097
```

```
length(which(cdc$weight>(mean(cdc$weight)-  
sd(cdc$weight))&cdc$weight<(mean(cdc$weight)+sd(cdc$weight))))/le  
ngth(cdc$weight)  
[1] 0.7076
```