Lab 1

Team name: Team 4

Author: Jonah, Miranda, Manavi, Charlotte, Lucy

Discussants (other team members present in lab):

Exercises

Load CDC data:

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

```
## [1] "genhlth"  "exerany"  "hlthplan" "smoke100" "height"   "weight"
## [7] "wtdesire" "age"      "gender"
```

Exercise 1:

head(cdc)
tail(cdc)
There are 20000 responses in the dataset. There are 9 variables in the dataset.
genhlth -> ordinal categorical
exerany -> unordered (regular) categirical
hlthplan -> regular categorical
smoke100 -> regular categorical
height -> continuous
weight -> continuous
wtdesire -> continuous
age -> continuous
gender -> regular categorical

summary(cdc*weight*)*table*(*cdc*smoke100)/20000 # percentage
barplot(table(cdc$smoke100))

Exercise 2:

```
summary(cdc$height)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   48.00   64.00   67.00   67.18   70.00   93.00
```

```
 # 1st -> 64.00 3rd -> 70.00

summary(cdc$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.00   31.00   43.00   45.07   57.00   99.00
```

```
# 1st -> 31.00 3rd -> 57.00


IQRHeight <- (70 - 64)
IQRAge <- (57.00 - 31.00)
```

```r
table(cdc$gender)
```

```
##
##     m     f
##  9569 10431
```

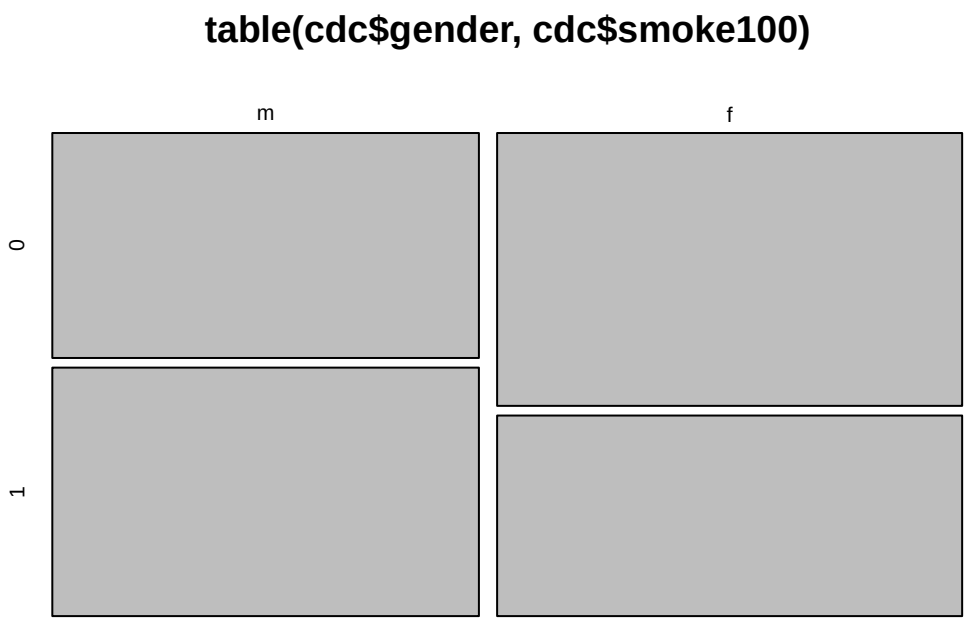
```r
#9569 males in the sample

table(cdc$genhlth)/20000
```

```
##
## excellent very good      good      fair      poor
##   0.23285   0.34860   0.28375   0.10095   0.03385
```

```r
# 23% are in excelent health
```

Exercise 3:

```r
mosaicplot(table(cdc$gender, cdc$smoke100))
```



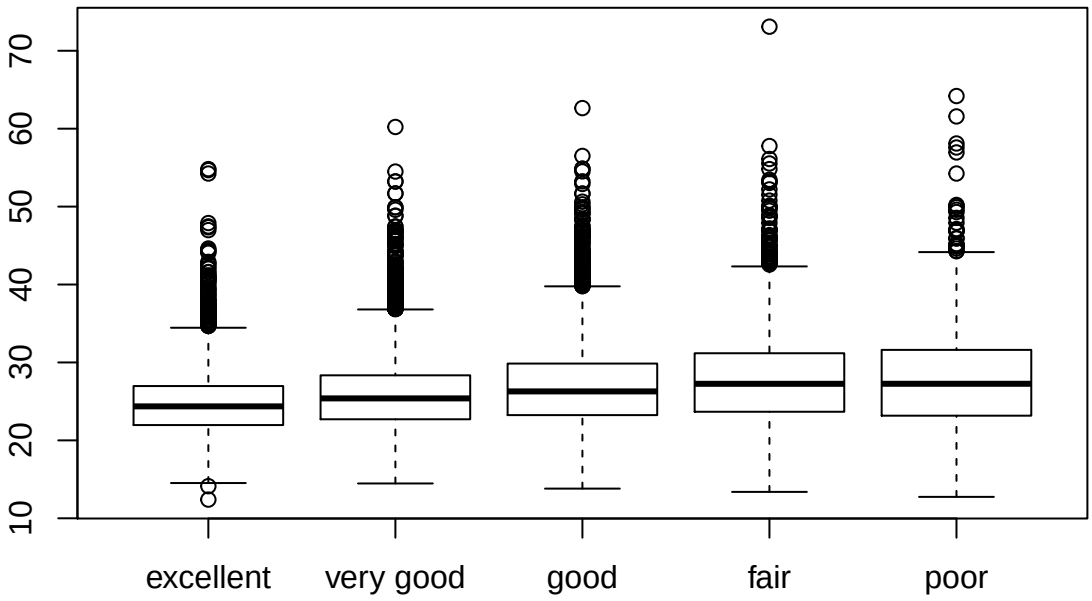

```r
# This revealed that slightly more males smoke than females.
# This is shown because the portion of men with a 1 response (1 is equivalent to True in binary)
# is higher than that of females with a 1 response.
```

Exercise 4:

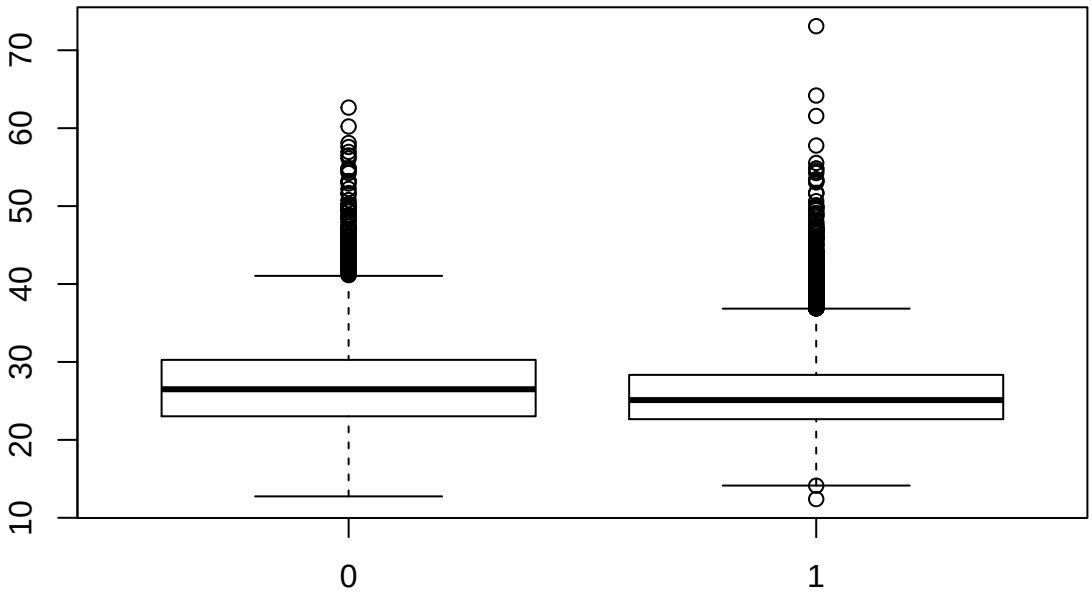
```r
under23_and_smoke <- subset(cdc, cdc$age < 23 & cdc$smoke100 == 1)
```

Exercise 5:

```r
bmi <- (cdc$weight/cdc$height^2) * 703
boxplot(bmi ~ cdc$genhlth)
```



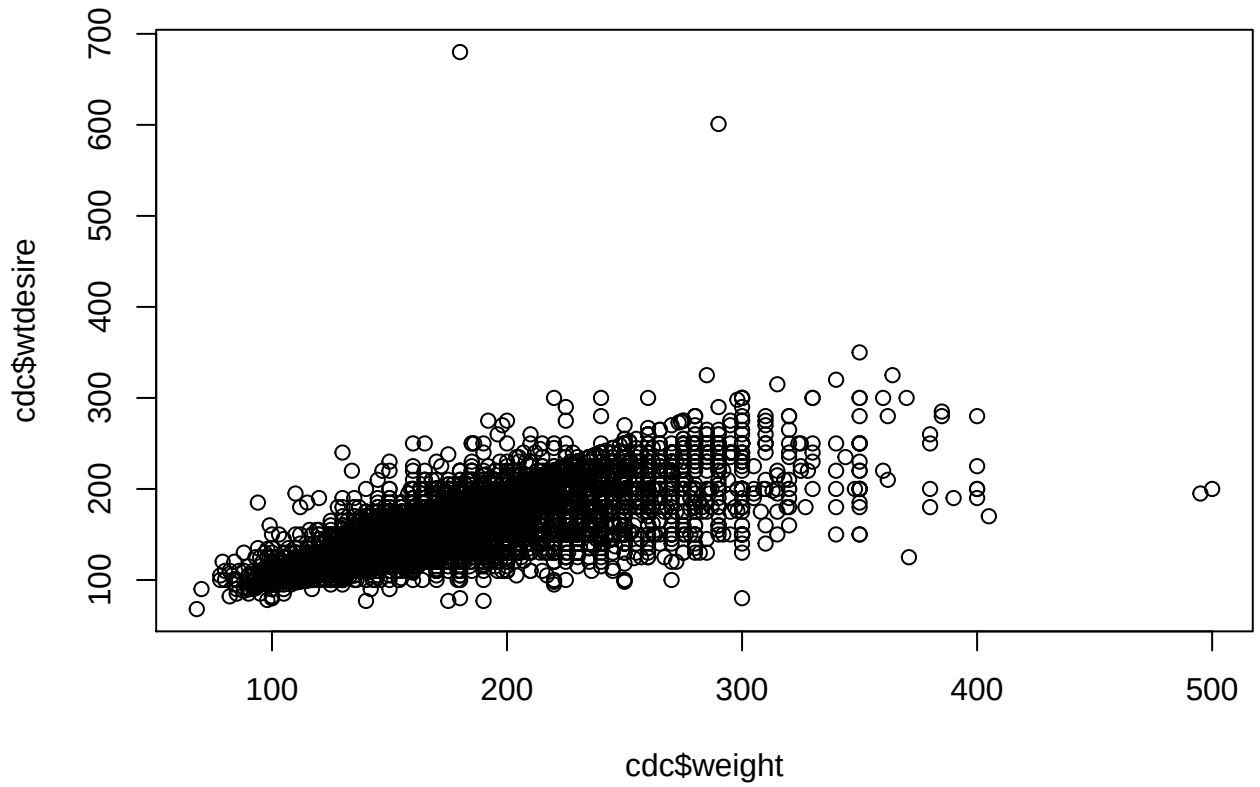

```
# This plot shows the bmi in accordance to thier general health.
# This is an obvious correlation that shows how bmi rises as general
# health becomes worse.

boxplot(bmi ~ cdc$exerany)
```




```
# This plot shows that bmi goes down as exercise increases.
# These have a causation relationship because as you exercise more, your body mass index will go down.
```

Exercise 6:

```r
plot(cdc$weight, cdc$wtdesire, type = 'p') # Plot of points
```




```r
# This plot shows a regression line that seems to be ~ (approximately) y = x
# This shows a correlation of people giving desired weights that are
# relatively close to their current weight.
# This is why the (estimated) regression lines has a slope of ~1.
# There are a few outliers which indicate people either making jokes about their weight,
# or about very serious dreams of weight loss.
```