Lab 2

Wesley Wessen

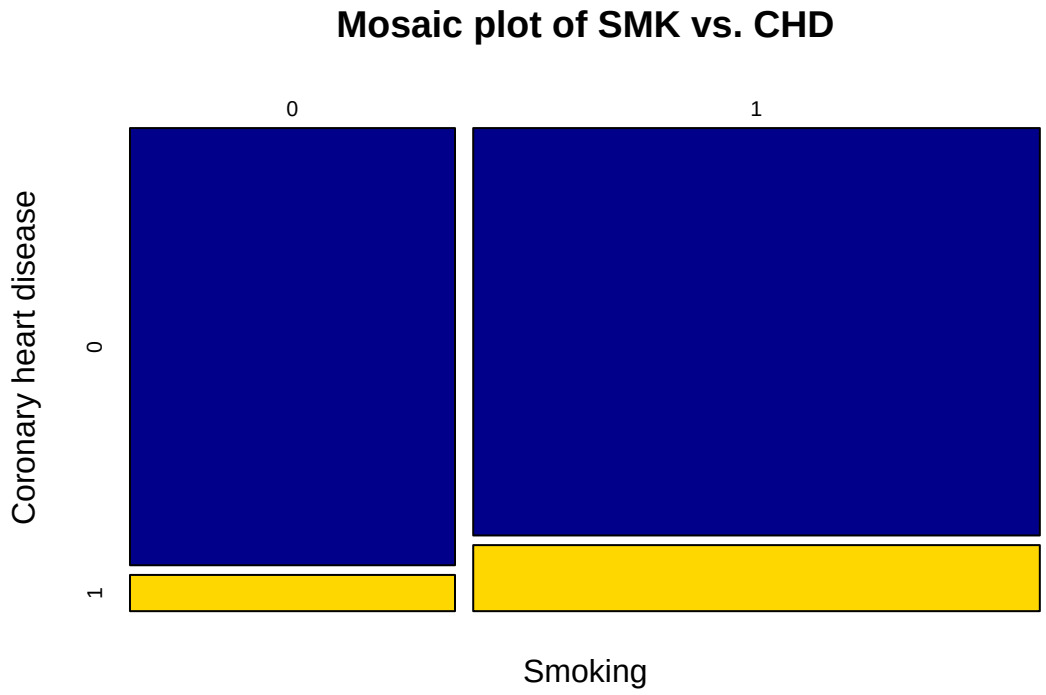
10/21/2020

Contingency table: coronary heart disease and smoking

```
##        Coronary heart disease
## Smoking   0   1
##       0 205  17
##       1 333  54
```

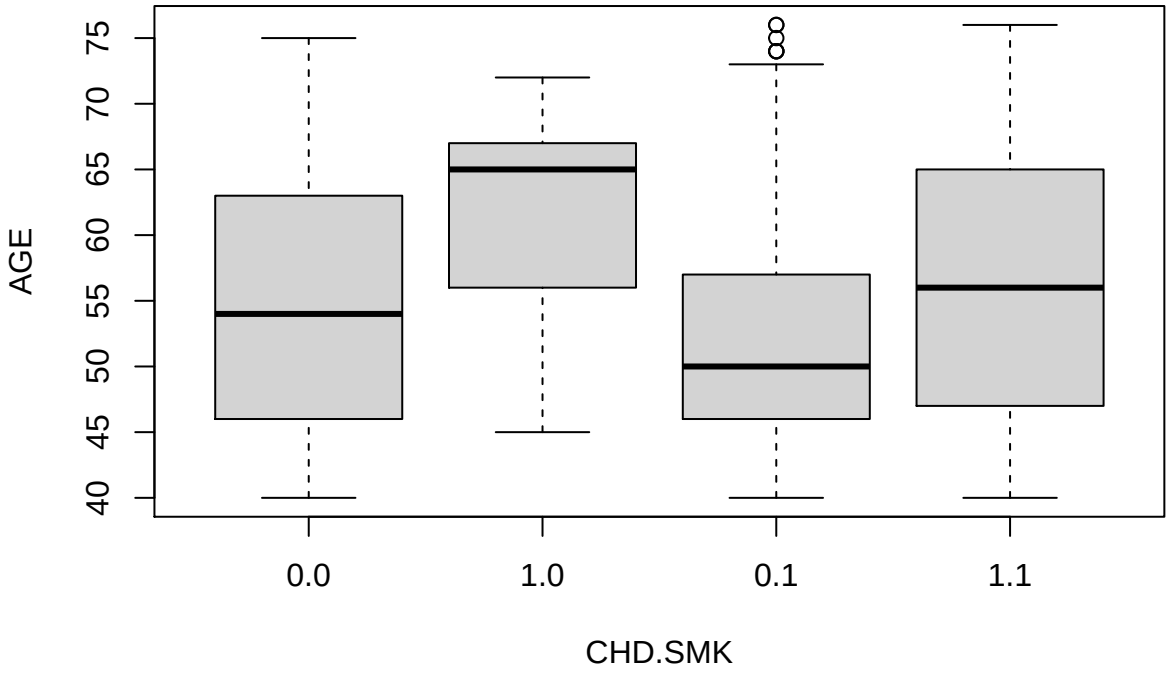
Comments: 0 indicates no smoking/no coronary heart disease (CHD), 1 indicates a positive for smoking/CHD.

Mosiac plot: coronary heart disease and smoking



Comments: The graph shows that the majority in the sample do not have CHD, regardless of whether they smoke or not. However, the amount of cases of CHD among those that smoke is higher than those that do not smoke.

Boxplots: distribution of age given CHD status and smoking status



Comments: The y-axis is the age of those in the sample. The x-axis indicates which category. The first number indicates CHD and the second number indicates SMK, 0 = no, 1 = yes. From the boxplot we can see those who smoke are getting CHD (1.1) at a younger age than those that do not (0.1). Those who are getting CHD without smoking (1.0) are getting it older as shown by the high center.The two middle groups show a smaller spread compared to the two outer ones, showing that there are less extreme values within those boxplots.

Qualitative characteristics: mean and sd of age given CHD status and smoking status

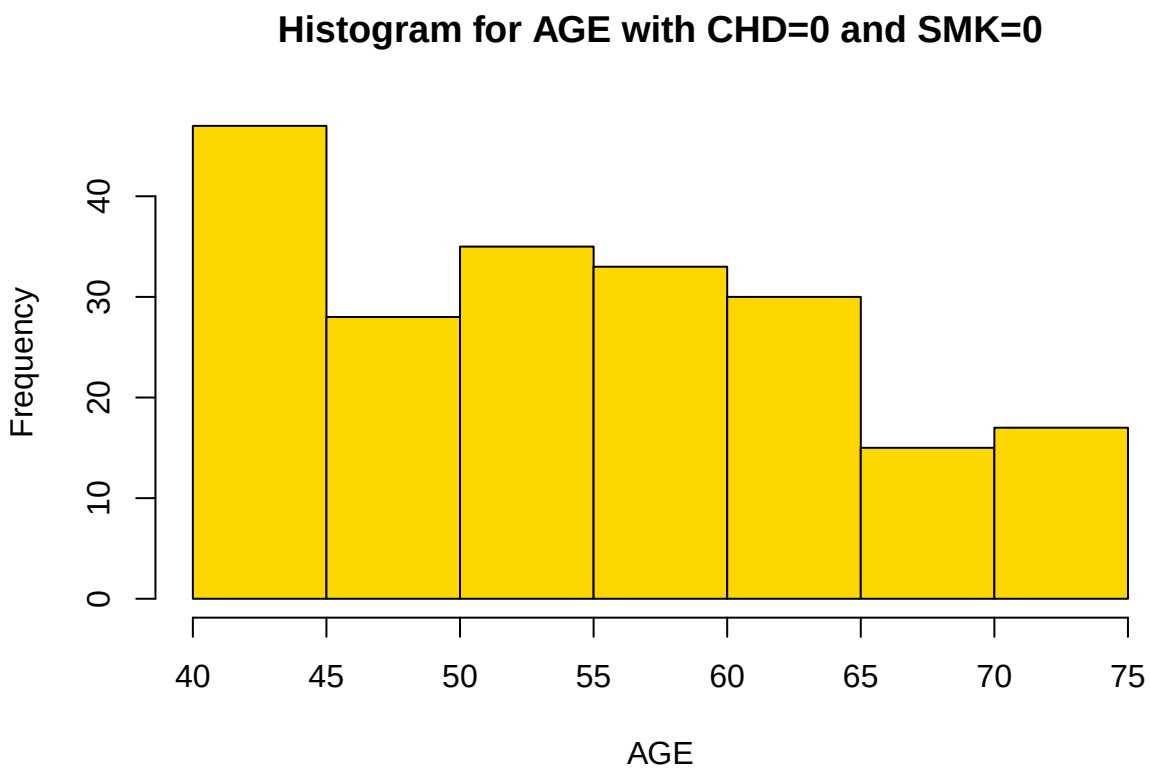
```
##   CHD SMK      AGE
## 1   0   0 54.88780
## 2   1   0 61.64706
## 3   0   1 52.22222
## 4   1   1 55.87037
```

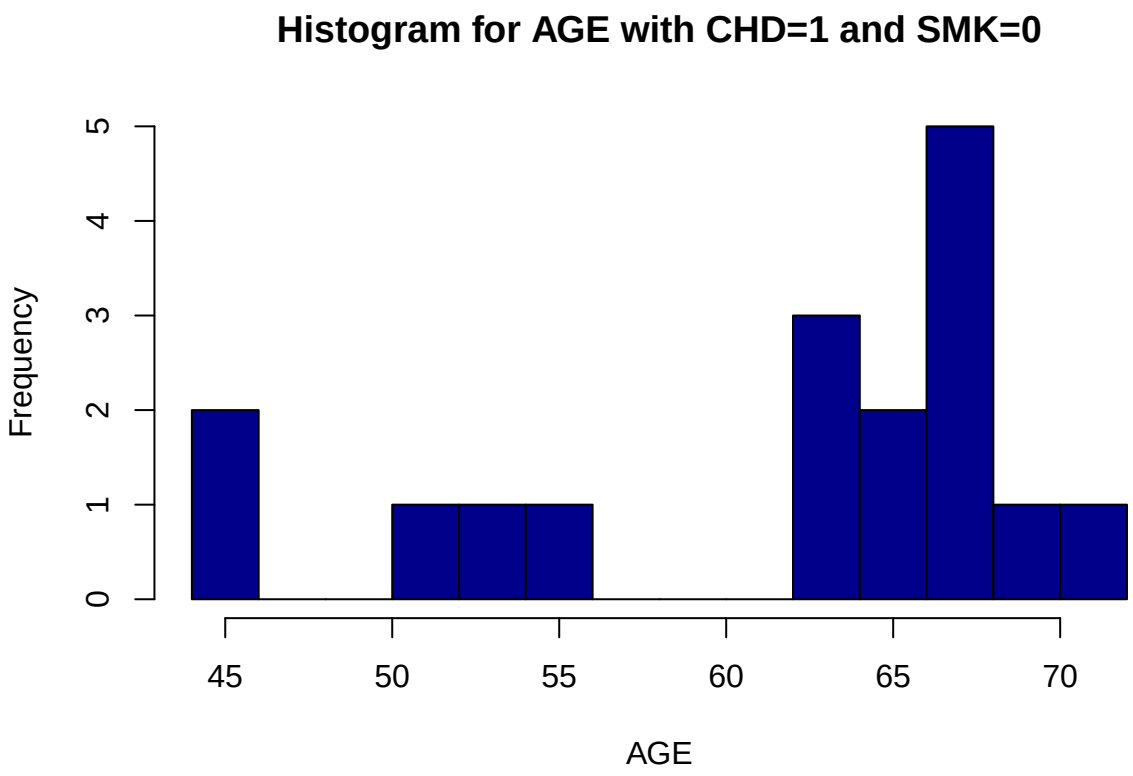
```
##   CHD SMK       AGE
```

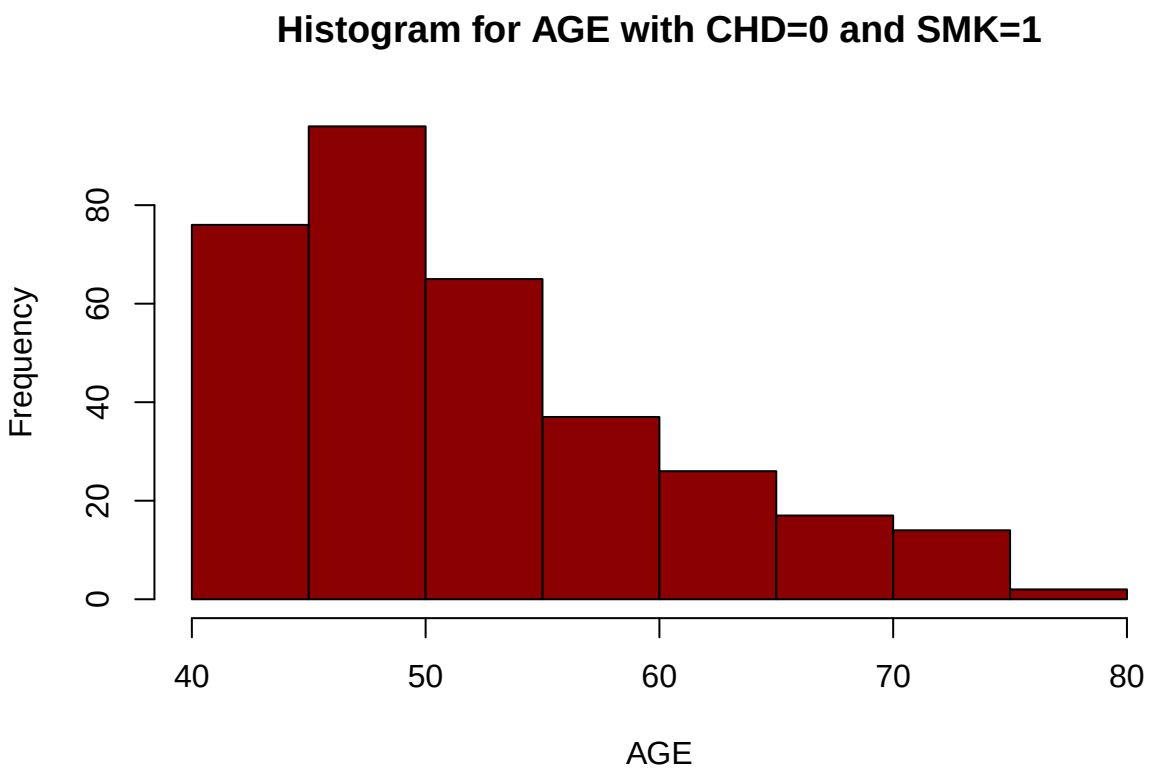
```
## 1   0   0  9.672181
## 2   1   0  8.146634
## 3   0   1  8.491944
## 4   1   1 10.356014
```

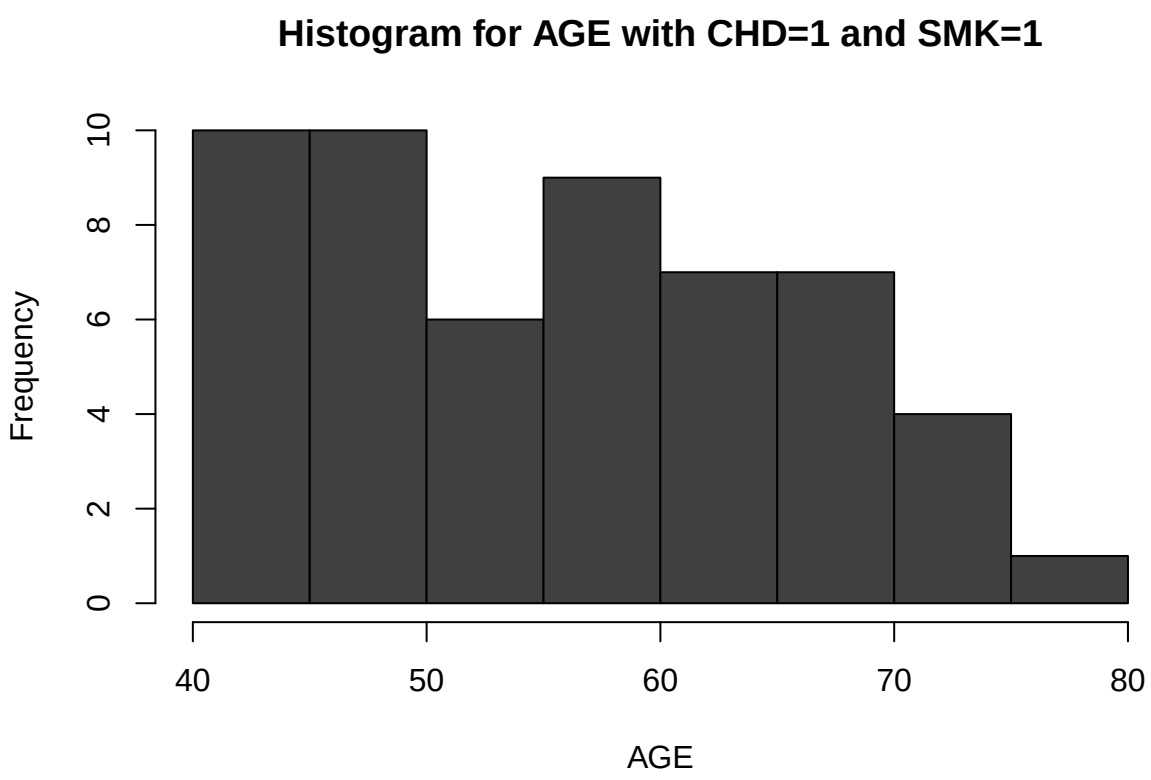
Comments: The differences of means of the ages of those who get CHD shows a nearly 6 year age gap between those that smoke (55.87) and those that don't (61.64). However, the standard deviation of 1.1 is 10.35 and 0.0 is 9.67, which is large compared to the others, since standard deviations are attracted to extreme values, there may be some extreme values in those groups.

Histograms: distribution of age given CHD status and smoking status




Histogram for AGE with CHD=1 and SMK=0
Frequency
0
1
2
3
4
5
45
50
55
60
65
70
AGE


Histogram for AGE with CHD=0 and SMK=1
Frequency
0
20
40
60
80
40
50
60
70
80
AGE



Comments: The yellow graph shows us a pretty flat response for those that do not smoke and don't have CHD. Blue shows that when poeple don't smoke, they get CHD at an old age, above 60 years old. Red graph is skewed to the right, showing that those who were sampled were younger in age. Gray shows many smokers who had CHD were younger.

Final Comments

The above graphs show that the age that reported CHD is overall lower for smokers than non smokers. Those that don't smoke and didn't contract CHD had a wide distribution and standard deviation, and a centrally located mean, providing a control to what graphs should look like for a healthy population.The histogram and boxplot distribution for smokers without CHD shows that smokers as a group were yougner repondandts of the study. The wide distribution on the box plot, along with the consistent high frequencies for age groups shows that smokers are more at risk of getting CHD, when compared to non smokers that got CHD got it at an older age, as evident of the narrower distribution, high mean on the box plot, and the relatively low standard deviation of 8.1 years.Due to this, those who smoke can be said to be at higher risk of getting CHD.

Appendix

```
knitr::opts_chunk$set(echo = FALSE, fig.align = "center")
vars <- c("ID","CHD","CAT","AGE","CHL","SMK","ECG","DBP","SBP","HPT","CH","CC")
evans <- read.table("http://www.stat.ucdavis.edu/~affarris/evans.dat",
                    header=FALSE,
                    col.names=vars)
SmkChdTable <- table("Smoking"=evans$SMK,
                     "Coronary heart disease"=evans$CHD)

SmkChdTable
mosaicplot(SmkChdTable, main="Mosaic plot of SMK vs. CHD",
           color = c("blue4", "gold"))
boxplot(formula = AGE~CHD+SMK, data = evans, xlab="CHD.SMK")
aggregate(formula = AGE~CHD+SMK, data = evans, FUN = mean)

aggregate(formula = AGE~CHD+SMK, data = evans, FUN = sd)
hist(evans$AGE[evans$CHD==0 & evans$SMK==0], xlab = "AGE", ylab = "Frequency",
     main = "Histogram for AGE with CHD=0 and SMK=0", col = c("gold"), breaks = 10)

hist(evans$AGE[evans$CHD==1 & evans$SMK==0], xlab = "AGE", ylab = "Frequency",
     main = "Histogram for AGE with CHD=1 and SMK=0", col = c("blue4"), breaks = 10)

hist(evans$AGE[evans$CHD==0 & evans$SMK==1], xlab = "AGE", ylab = "Frequency",
     main = "Histogram for AGE with CHD=0 and SMK=1", col = c("red4"), breaks = 10)

hist(evans$AGE[evans$CHD==1 & evans$SMK==1], xlab = "AGE", ylab = "Frequency",
     main = "Histogram for AGE with CHD=1 and SMK=1", col = c("gray25"), breaks = 10)
```