STA 2210 Homework 1 (Due on Monday 5/18 by 11:59pm)

The data set BM, a .csv file, contains data on percent body fat and other various measurements of body size, for a sample of 252 men. Write your R codes, in addition to your answer, to the following problems. (Don’t forget to refer to the R reference card to find helpful commands.)

BM <- read.csv(file.choose(), header = TRUE)

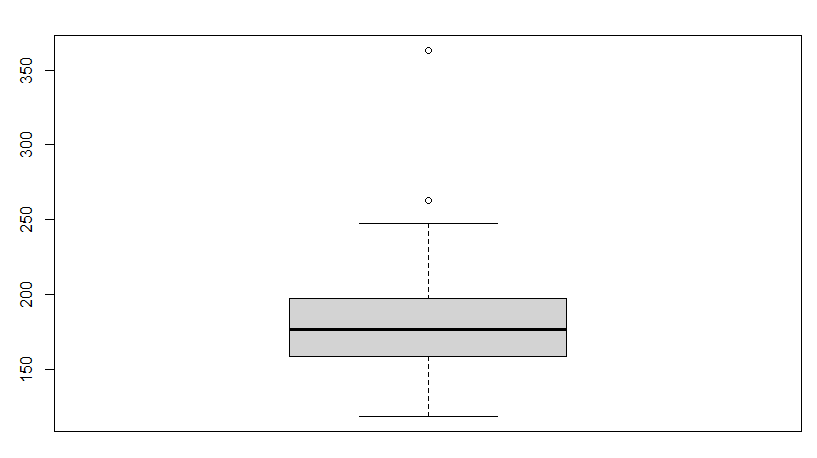
1. **Create a numerical summary for WEIGHT, and compute the interquartile range. Are any of the individuals outliers?**

Summary(BM$WEIGHT)

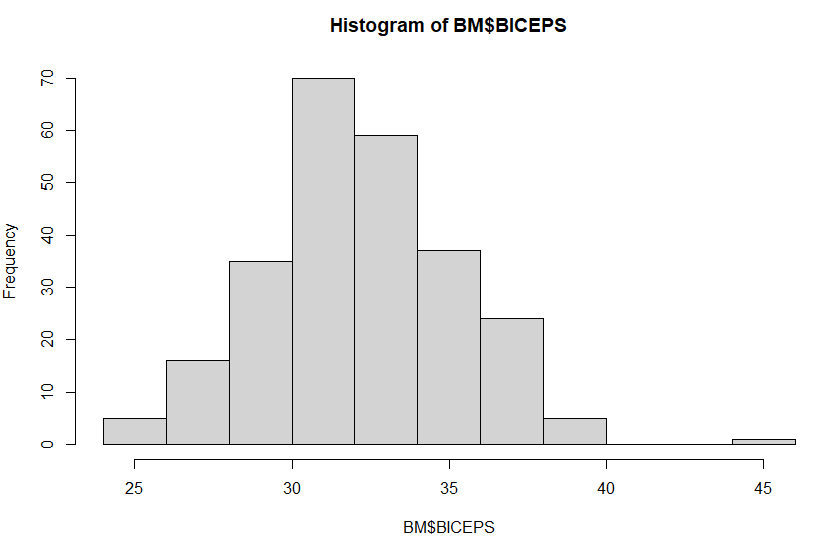
Min. 1st Qu. Median Mean 3rd Qu. Max.

118.5 159.0 176.5 178.9 197.0 363.1

IQR = Q3-Q1 = 197.0-159.0 = 38.0

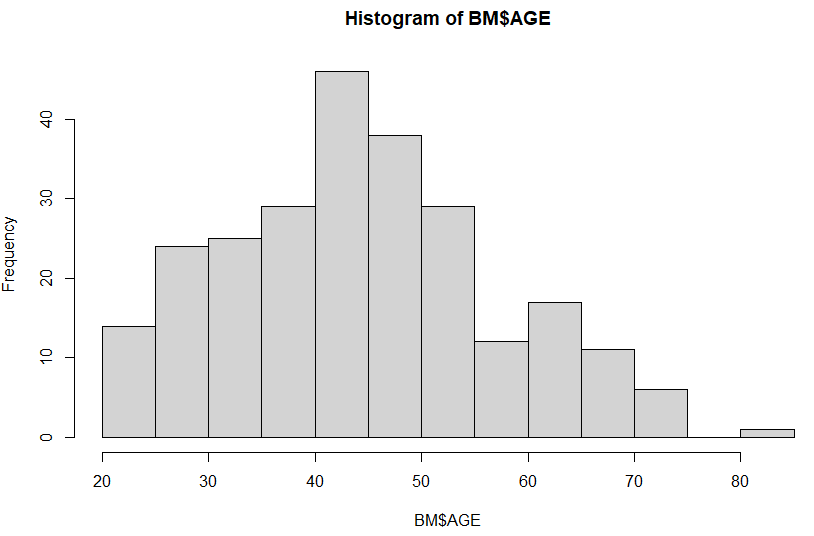
There are two outliers outside the maximum by looking at boxplot(BM$WEIGHT).

1. **Create a histogram for the following variables and describe the shape of the distribution.**
2. BICEPS = hist(BM$BICEPS)



The distribution has a unimodal shape with one peak near 30-32 cm and mostly symmetric skewness as the frequencies are relatively even going away from both sides of the peak. There is a slight pull of more observations to the right of the peak.

1. AGE = hist(BM$AGE)



The distribution is mostly unimodal in shape with a slight upward trend to the right of the peak, and has mostly symmetric skewness, with slight right skew due to higher frequency of age on the left near 20 than 80 on the right.

1. **What proportion of the men are older than 40 years?**

Prop40 <- subset(BM, AGE > 40) = 160 obs. / 252 total obs. = .635 = 63.5% of men are older than 40.

1. **What proportion of the men are taller than 68 in and weigh more than 200 lbs?**

Heightheavy <- subset(BM, HEIGHT > 68 & WEIGHT > 200) = 53 obs. / 252 total obs. = .210 = 21.0% of men are taller than 68 in. and weight more than 200 lbs.

1. **What proportion of the men have a wrist measurement that is less than 18 cm or a forearm measurement that is less than 28 cm?**

Wristforearm <- subset(BM, WRIST < 18 | FOREARM < 28) = 123 obs. / 252 total obs. = .488 = 48.8% of men have a wrist of less than 18 cm and forearm less than 28 cm.

1. **Find the mean and standard deviation of BODYFAT and determine the proportion of the individuals that are more than two standard deviations from the mean.**

mean(BM$BODYFAT) = 18.93849

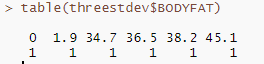
sd(BM$BODYFAT) = 7.750856

Bodyfat greater than 2 standard deviations = 18.93849 + (2 \* 7.750856) = 34.440202

Bodyfat less than 2 standard deviations = 7.750856 + (2 \* 7.750856) = 3.436778

threestdev <- subset(BM, BODYFAT > (34.440202) | BODYFAT < 3.436778)

= 6 observations of individuals more than two standard deviations from the mean.



Proportion: 6/252 = 0.0238 = 2.38% of men more than 2 standard deviations from the mean.