R Project 1

```
heart.data <- read.csv("~/Downloads/ClevelandHeartData.csv")
```

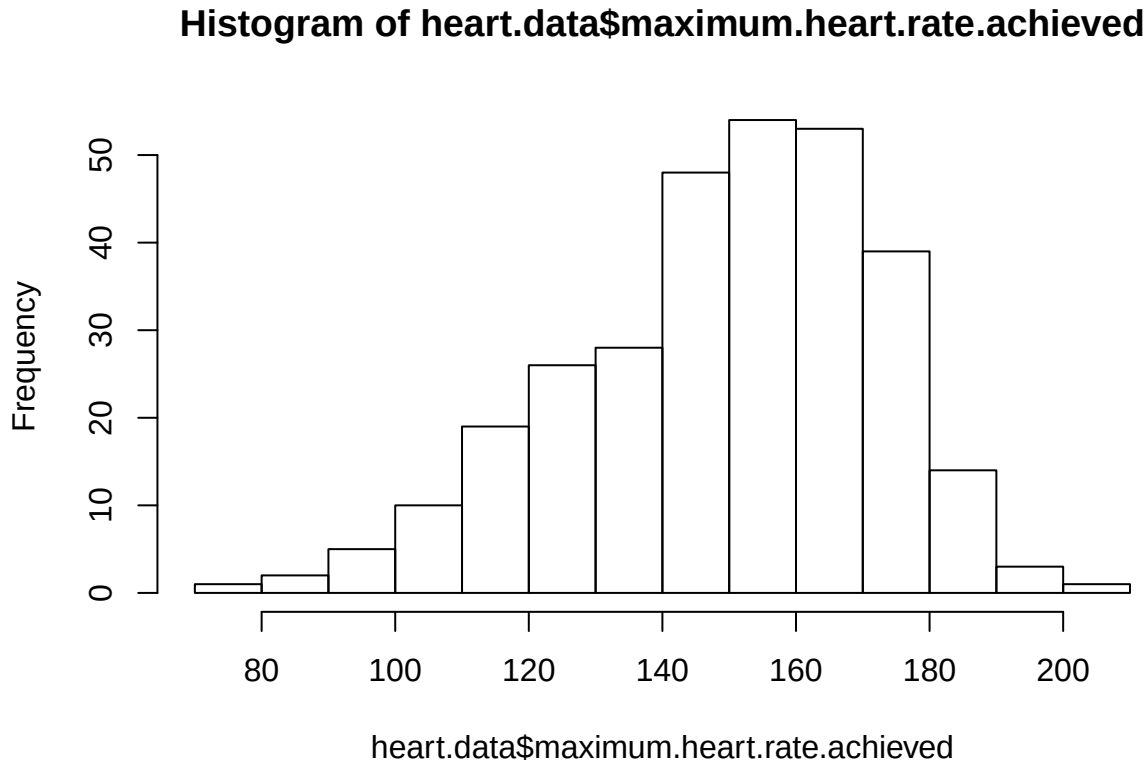
#1. How many of the 303 patients were diagnosed with heart disease?

```
sum(heart.data$presence.or.absence.of.heart.disease..0..50.diameter.narrowing.1..50.diameter.narrowing.
```

```
## [1] 139
```

2. Make a histogram of the data in the "maximum heart rate achieved" column.

```
hist(heart.data$maximum.heart.rate.achieved)
```



3. Report the sample mean, median, and sample standard deviation for the maximum heart rate.

```
mean(heart.data$maximum.heart.rate.achieved)
```

```
## [1] 149.6073
```

```
median(heart.data$maximum.heart.rate.achieved)
```

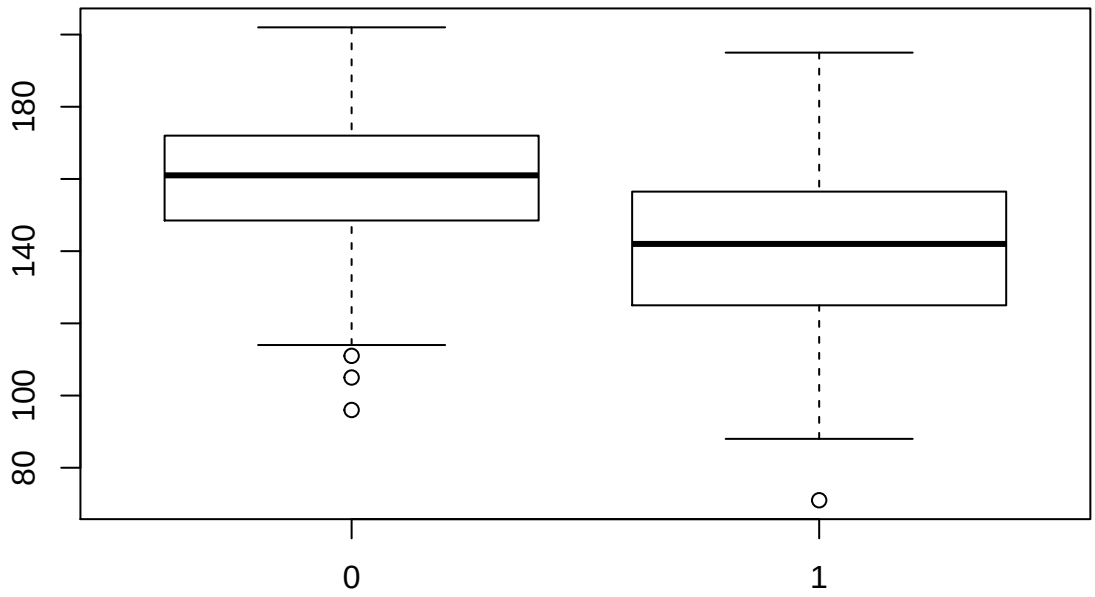
```
## [1] 153
```

```
sd(heart.data$maximum.heart.rate.achieved)
```

```
## [1] 22.875
```

4. Make a side-by-side boxplot of the maximum heart rate for those patients diagnosed with heart disease vs. those patients not diagnosed with heart disease.

```
boxplot(maximum.heart.rate.achieved~presence.or.absence.of.heart.disease..0..50.diameter.narrowing.1..5
```



5. Do a t-test to test whether the maximum heart rate is significantly different for those patients diagnosed with heart disease vs. those patients not diagnosed with heart disease. Report both the p-value and the 95% confidence interval for the difference of the population means. Clearly state your conclusion.

```
heart.data.withDisease <- heart.data[heart.data$presence.or.absence.of.heart.disease..0..50.diameter.na
heart.data.withoutDisease <- heart.data[heart.data$presence.or.absence.of.heart.disease..0..50.diameter
t.test(heart.data.withDisease$maximum.heart.rate.achieved, heart.data.withoutDisease$maximum.heart.rate
```

```
##
##  Welch Two Sample t-test
##
## data:  heart.data.withDisease$maximum.heart.rate.achieved and heart.data.withoutDisease$maximum.hear
## t = -7.8579, df = 272.27, p-value = 9.106e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -23.90912 -14.32900
## sample estimates:
## mean of x mean of y
##   139.259   158.378
```

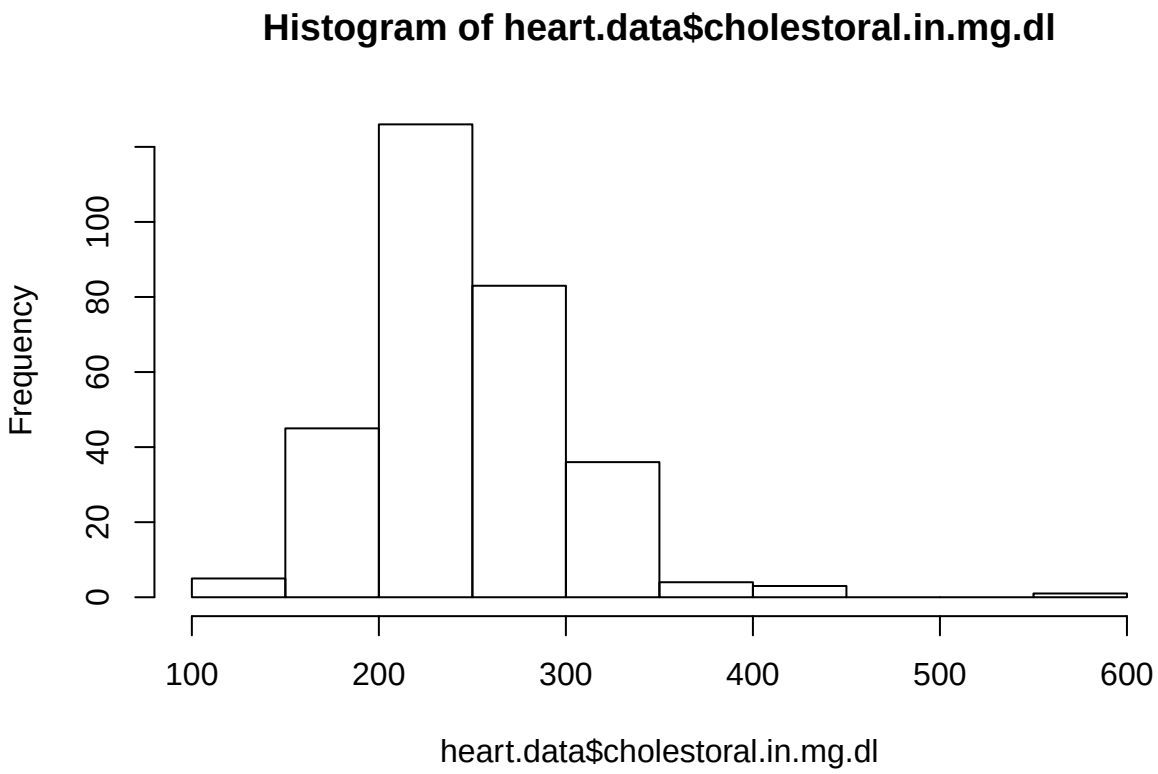
6. Do a Wilcoxon-Mann-Whitney test to test whether the maximum heart rate is significantly different for those patients diagnosed with heart disease vs. those patients not diagnosed with heart disease. Report the p-value and clearly state your conclusion.

```
wilcox.test(heart.data.withDisease$maximum.heart.rate.achieved, heart.data.withoutDisease$maximum.heart
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  heart.data.withDisease$maximum.heart.rate.achieved and heart.data.withoutDisease$maximum.hear
## W = 5806.5, p-value = 1.861e-13
## alternative hypothesis: true location shift is not equal to 0
```

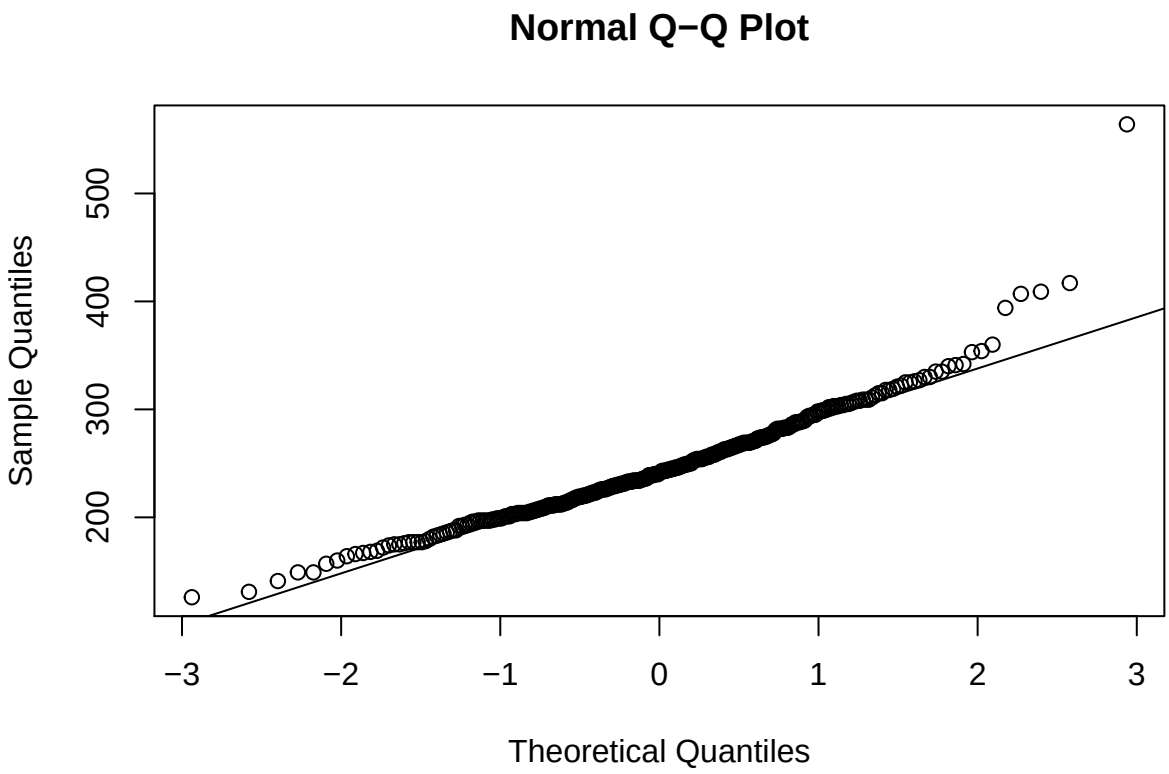
7. Make a histogram of the data in the cholesterol column.

```
hist(heart.data$cholestoral.in.mg.dl)
```



8. Make a qqplot and do the Shapiro-Wilk test to test whether the cholesterol values are distributed normally.

```
qqnorm(heart.data$cholestoral.in.mg.dl)
qqline(heart.data$cholestoral.in.mg.dl)
```

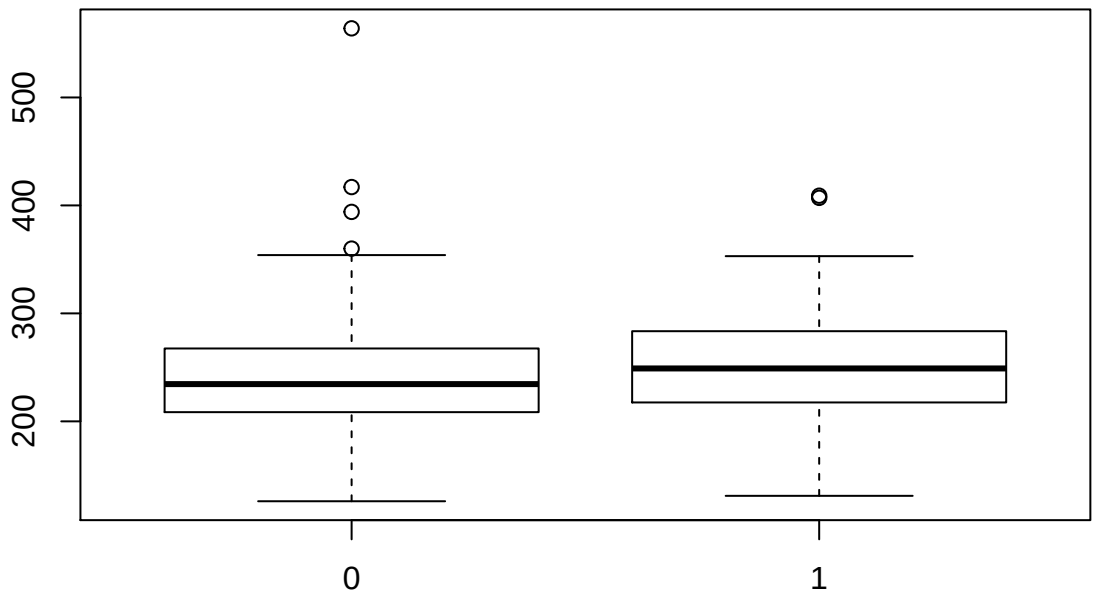



```
shapiro.test(heart.data$cholestoral.in.mg.dl)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  heart.data$cholestoral.in.mg.dl
## W = 0.94725, p-value = 5.912e-09
```

9. Make a side-by-side boxplot of the cholesterol values for those patients diagnosed with heart disease vs. those patients not diagnosed with heart disease.

```
boxplot(cholestoral.in.mg.dl~presence.or.absence.of.heart.disease..0..50.diameter.narrowing.1..50.diame
```



10. Do a t-test to test whether cholesterol is significantly different for those patients diagnosed with heart disease vs. those patients not diagnosed with heart disease. Report both the p-value and the 95% confidence interval for the difference of the population means. Clearly state your conclusion.

```
t.test(heart.data.withDisease$cholestoral.in.mg.dl, heart.data.withoutDisease$cholestoral.in.mg.dl)
```

```
##
##  Welch Two Sample t-test
##
## data:  heart.data.withDisease$cholestoral.in.mg.dl and heart.data.withoutDisease$cholestoral.in.mg.dl
## t = 1.4924, df = 298.64, p-value = 0.1366
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.815018 20.484170
## sample estimates:
## mean of x mean of y
##  251.4748  242.6402
```

11. Propose your own question to answer with this data set (for example, is the maximum heart rate significantly different for men and women in this data set?). Clearly state the question you are trying to answer, the test you performed, the results of the test, and your conclusion. If there are any relevant plots include them as well.