Lab 5a

Aram

April 29, 2019

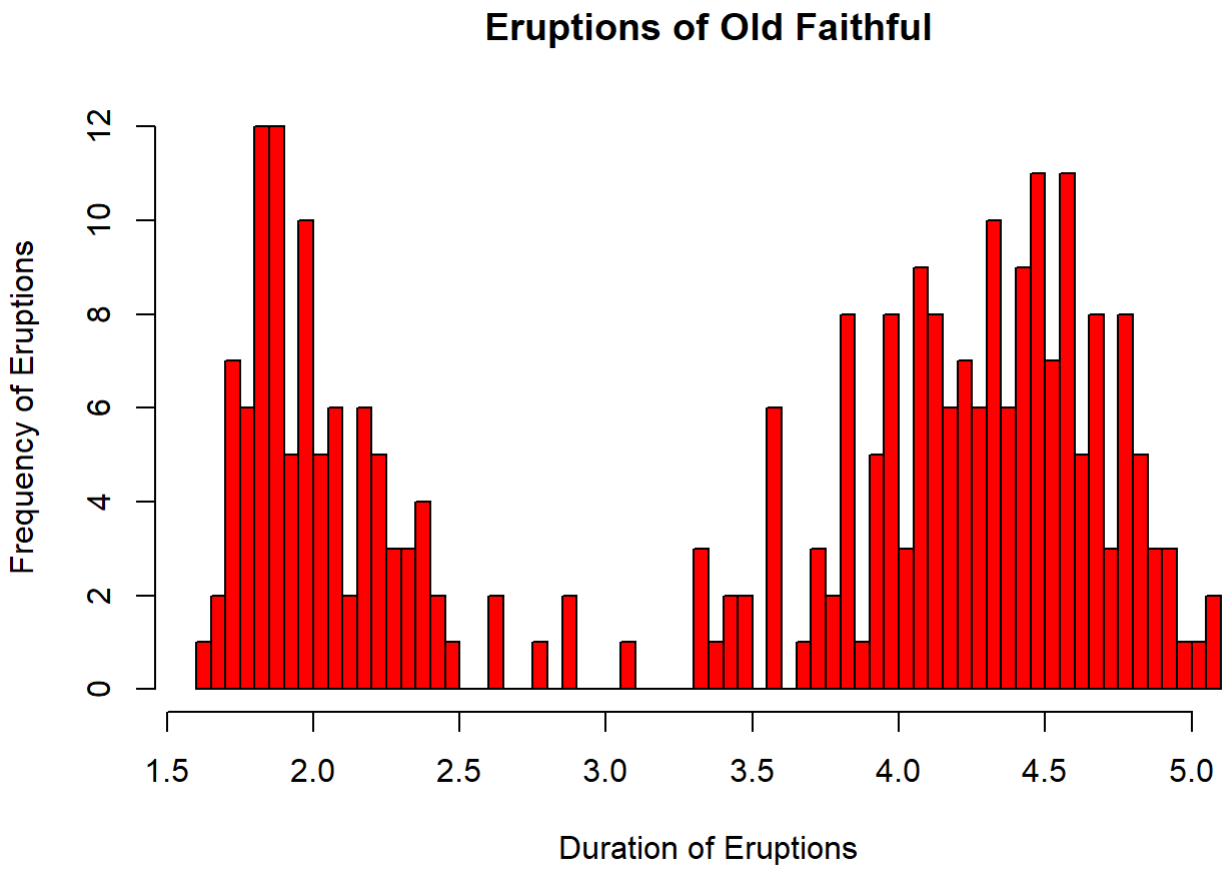
Introduction

This dataset consists of the waiting time between eruptions and how long the eruptions are for Old Faithful geyser in Yellowstone National Park, Wyoming, USA

```
data("faithful")
head(faithful, n = 3)
```

```
##   eruptions waiting
## 1     3.600      79
## 2     1.800      54
## 3     3.333      74
```

```
wd <- (faithful$eruptions)

hist(wd, breaks = 50, xlab = "Duration of Eruptions", ylab = "Frequency of Eruptions", col = "red", main = "Eruptions of Old Faithful")
```



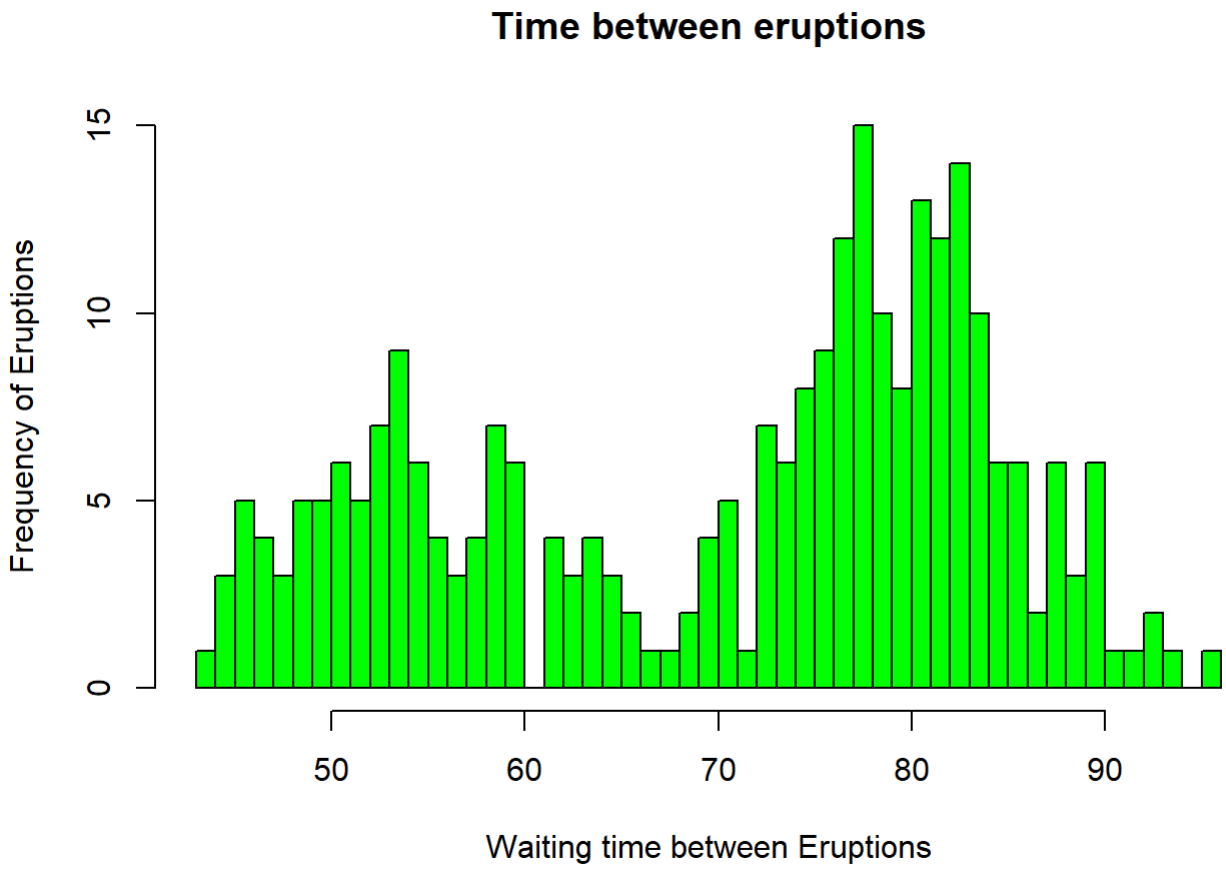

```
summary(faithful$eruptions)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.600   2.163   4.000   3.488   4.454   5.100
```

Discussion about Distribution of the duration of eruptions

The overall shape of the distribution is bimodal, meaning the data centers around two means. The two means are around 110 and 270 seconds respectively. The pattern seems to be that most of the eruptions of Old Faithful last around slightly under 2 minutes and around 4.5 minutes, there is little variation with eruptions that last around 2 minutes or the first mean, there is a lot of variation in the second mean where eruptions last around 4 to 5 minutes.

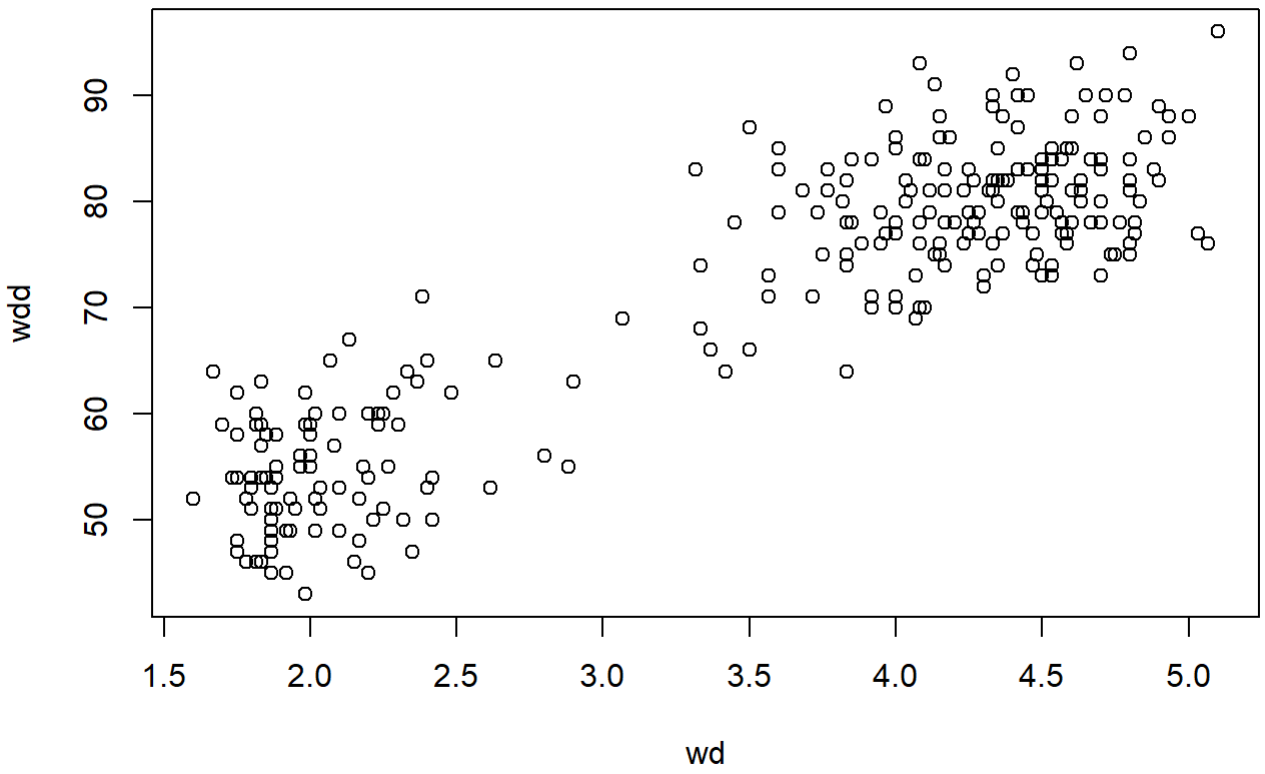
```
wdd <- (faithful$waiting)

hist(wdd, breaks = 50, xlab = "Waiting time between Eruptions", ylab = "Frequency of Eruptions", col = "green", main = "Time between eruptions")
```



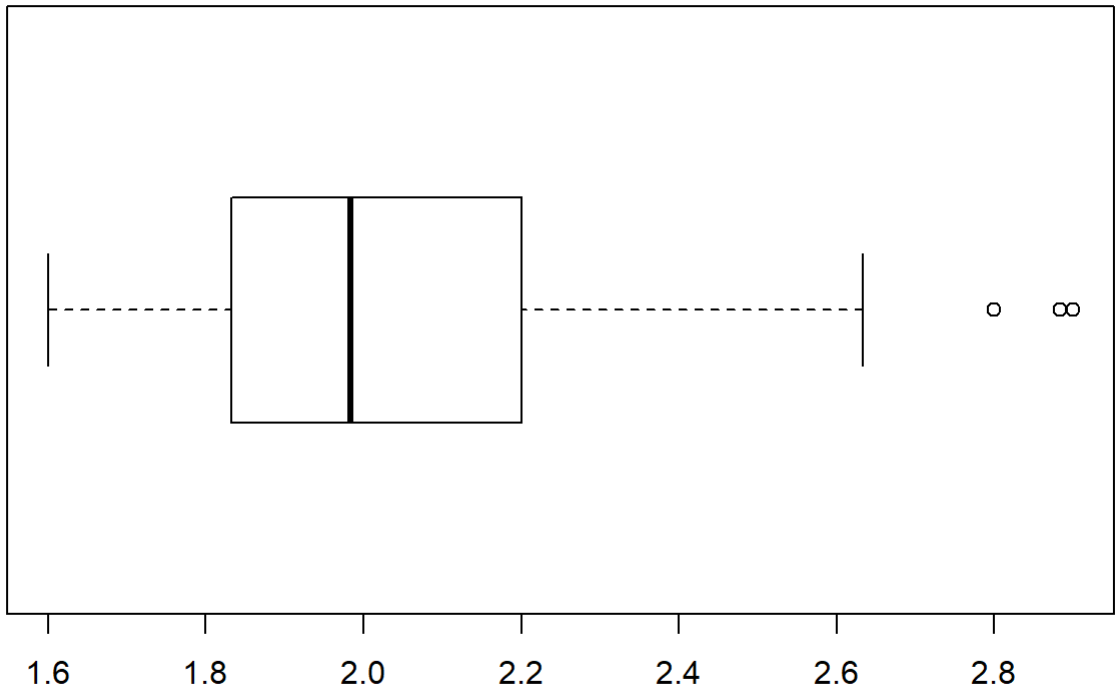
Analytical Discussion behind time

between Eruptions Plotting a histogram of the time between eruptions we can see some degree of bimodal distribution in the data for Old Faithful, given this we can estimate that the average time between eruptions ly around 50 minutes or 80 minutes. The data centers around these two means, biased toward the latter of 80 minutes. We can use the duration of the current eruption to predict the waiting time to the next eruption. It might be worth testing if the shorter eruptions correlated to a shorter waiting time between eruptions and the longer eruptions correlated to a longer waiting time between eruptions. In other words a positive correlation.

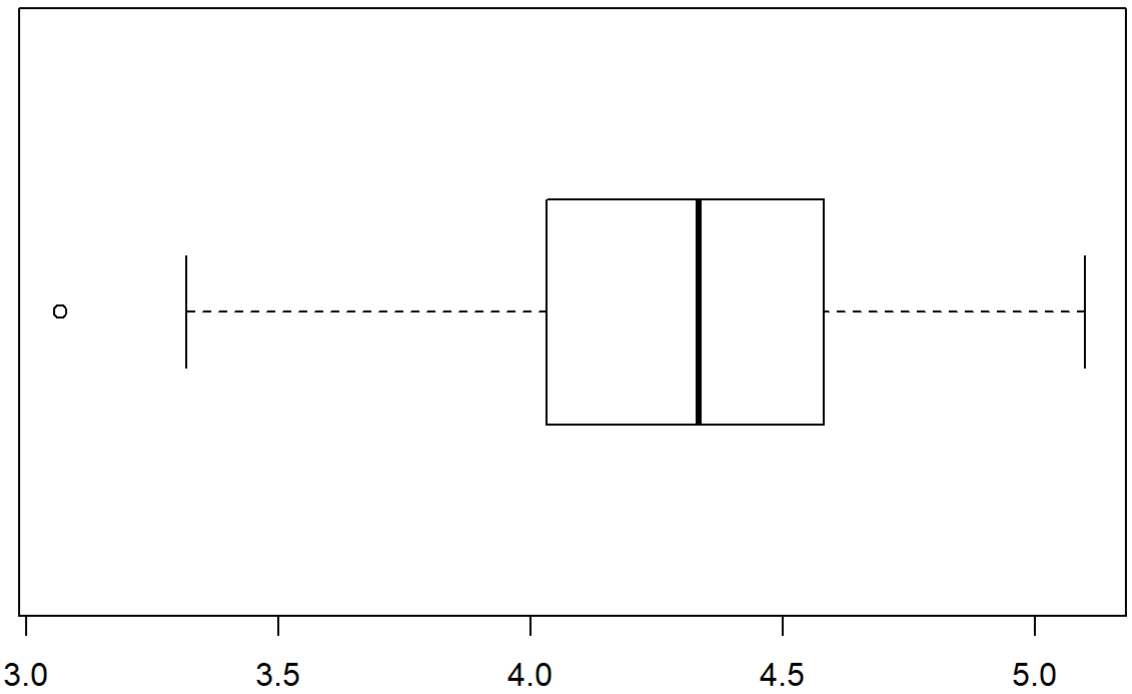
```
plot(wd,wdd)
```




```
wd2 <- subset(wd, wd < 3)
wd3 <- subset(wd, wd >= 3)

boxplot(wd2, horizontal = TRUE)
```




```
boxplot(wd3, horizontal = TRUE)
```




```
summary(wd2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.600   1.833   1.983   2.038   2.200   2.900
```

```
summary(wd3)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.067   4.033   4.333   4.291   4.583   5.100
```

Discussion behind numeric summaries of Eruption time subgroups

Taking a look at the 5 number summaries for each subgroup, we can observe that for the first subgroup where waiting time is less than 3 minutes, the average eruption time is 2 minutes and for the subgroup where waiting time is more than or equal to 3 minutes, the average eurption time is 4.2 minutes. This matches our estimate earlier at the beginning of the analysis that the distribution is bimodal centered around 2 means.

Question 8

If you arrive at Old Faithful after an eruption of less than 3 minutes you would expect to wait less to the next eruption than if you had arrived just after an eruption of greater than 3 minutes because the waiting time is less for the subgroups where eruptions lasted less than 3 minutes compared to the subgroup where eruptions lasted more than 3 minutes.