Dr Isabella Gollini

Probability and Statistics

Solutions 1

1. (i)
```r
data(morley)

dim(morley)
```
```
## [1] 100   3
```
```r
head(morley)
```
```
##     Expt Run Speed
## 001    1   1   850
## 002    1   2   740
## 003    1   3   900
## 004    1   4  1070
## 005    1   5   930
## 006    1   6   850
```
```r
tail(morley)
```
```
##     Expt Run Speed
## 095    5  15   810
## 096    5  16   940
## 097    5  17   950
## 098    5  18   800
## 099    5  19   810
## 100    5  20   870
```

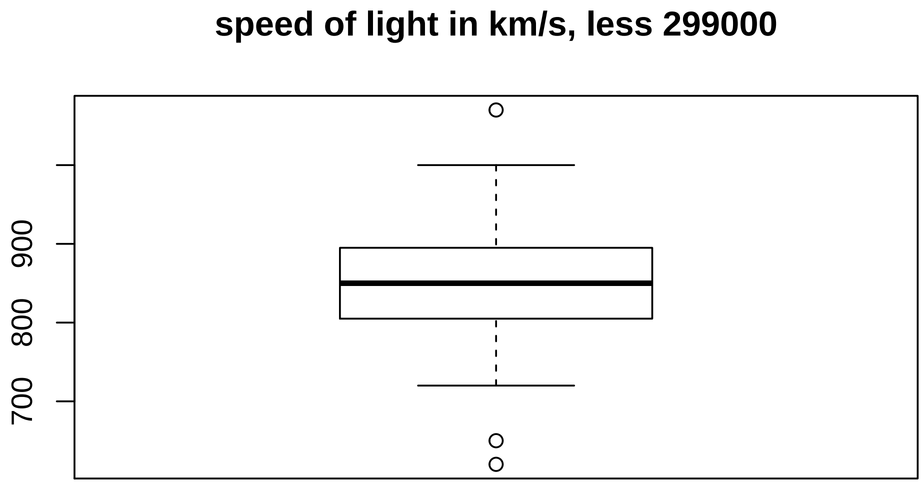
(ii)
```r
speed <- morley$Speed

summary(speed)
```
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   620.0   807.5   850.0   852.4   892.5  1070.0
```

The average recorded speed of light was $299000 + 852.40 = 299852$ to the nearest km/s.
The interquartile range was $Q_3 - Q_1 = 892.50 - 807.50 = 85$.

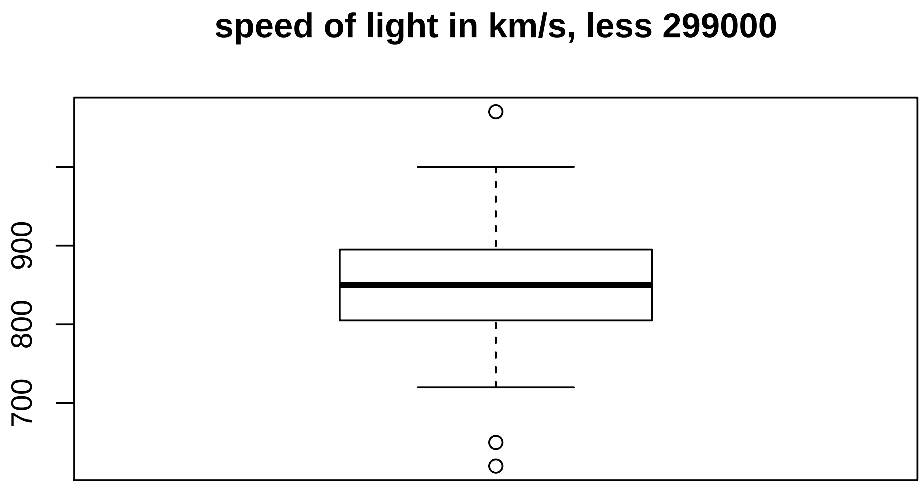
```r
IQR(speed)
```
```
## [1] 85
```

(iii)
```r
boxplot(speed,
        main = "speed of light in km/s, less 299000")
```



If you save the boxplot to an object you can find the outliers:

```r
bp_speed <- boxplot(speed,
                    main = "speed of light in km/s, less 299000")
```



In particular

```r
bp_speed$out
```

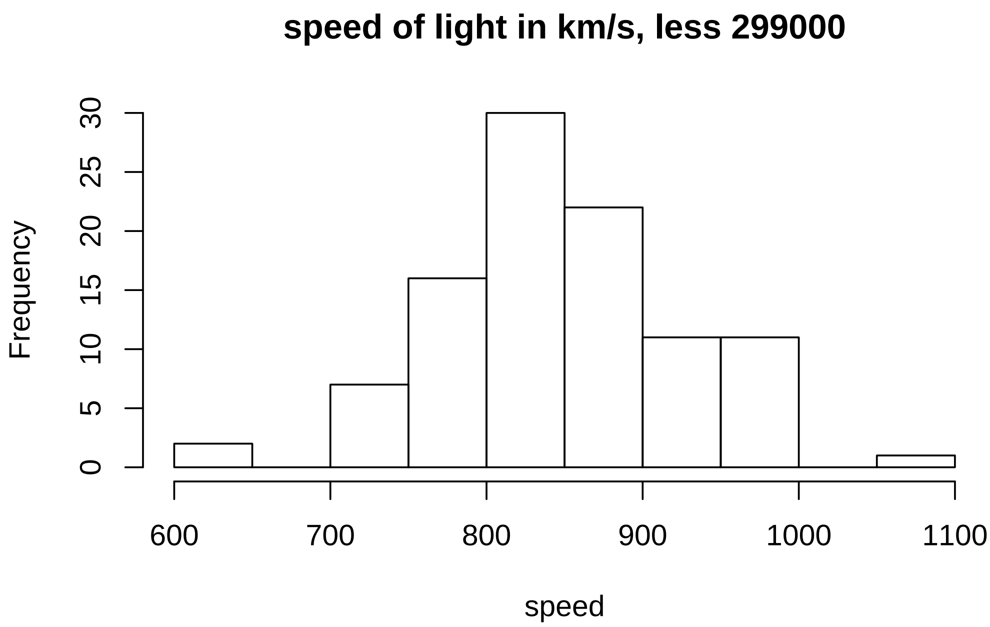
```
## [1] 1070  650  620
```

gives the outliers.

The values of the three “outliers” are 620, 650, 1070.

(iv)
```r
cutpoints <- seq(600, 1100, by = 50)

hist(speed,
     breaks = cutpoints,
     main = "speed of light in km/s, less 299000")
```



2. (i)
```r
cointossing <- read.table(file = "cointossing.dat")
colnames(cointossing) <- "heads"
```

(ii)
```r
summary(cointossing)
```

```
##      heads
##  Min.   :1.00
##  1st Qu.:4.00
##  Median :5.00
##  Mean   :4.97
##  3rd Qu.:6.00
##  Max.   :9.00
```

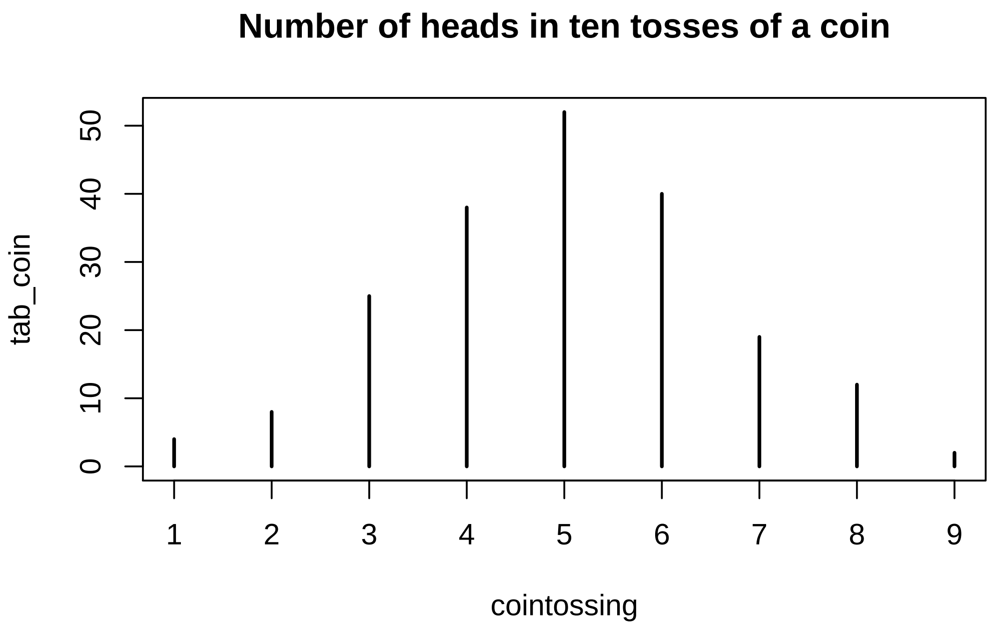
The average number of heads observed in ten tosses was 4.97.
The smallest number observed was 1 and the largest 9.

(iii) The session command and the frequency distribution are shown below.

```
tab_coin <- table(cointossing)
tab_coin

## cointossing
##  1  2  3  4  5  6  7  8  9
##  4  8 25 38 52 40 19 12  2
```

(iv)
```
plot(tab_coin, main = "Number of heads in ten tosses of a coin")
```

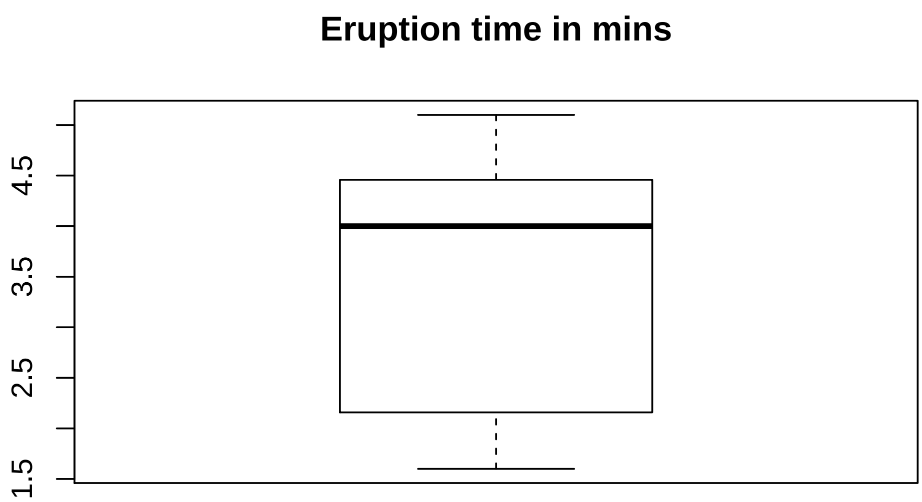


3. (i)
```
data(faithful)
```

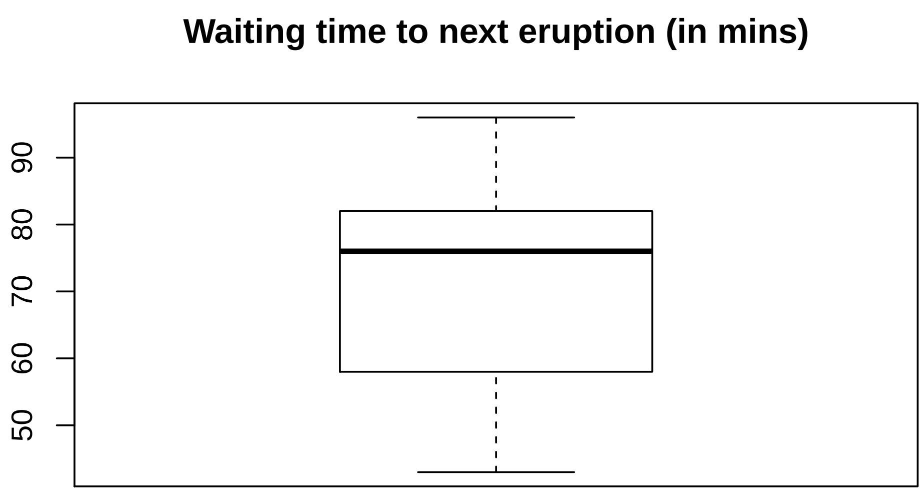
(ii)
```
summary(faithful)

##    eruptions        waiting
##  Min.   :1.600   Min.   :43.0
##  1st Qu.:2.163   1st Qu.:58.0
##  Median :4.000   Median :76.0
##  Mean   :3.488   Mean   :70.9
##  3rd Qu.:4.454   3rd Qu.:82.0
##  Max.   :5.100   Max.   :96.0
```

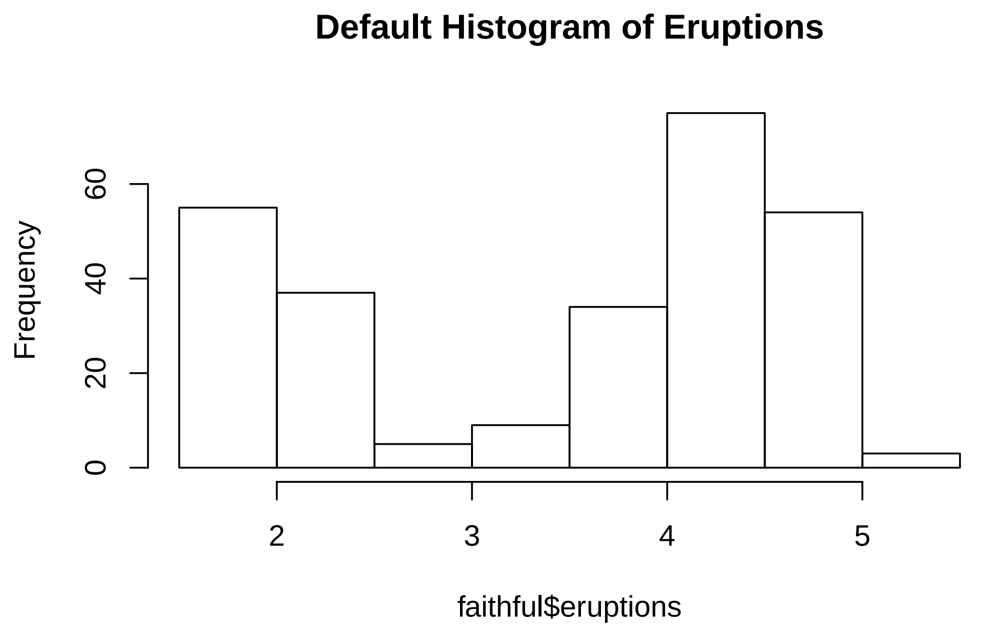
(iii)
```r
boxplot(faithful$eruptions,
        main = "Eruption time in mins")
```



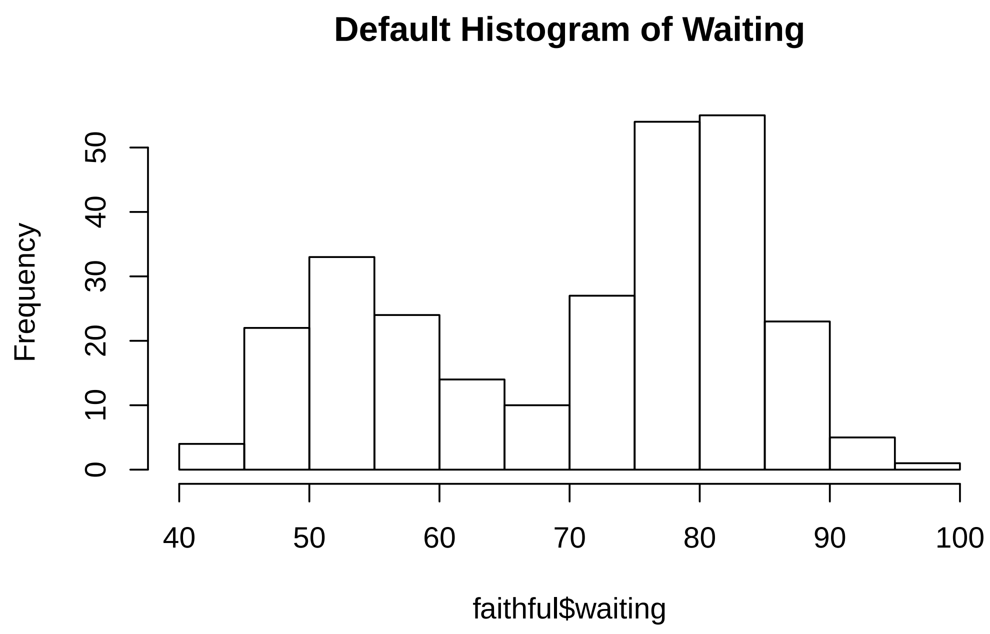

```r
boxplot(faithful$waiting,
        main = "Waiting time to next eruption (in mins)")
```



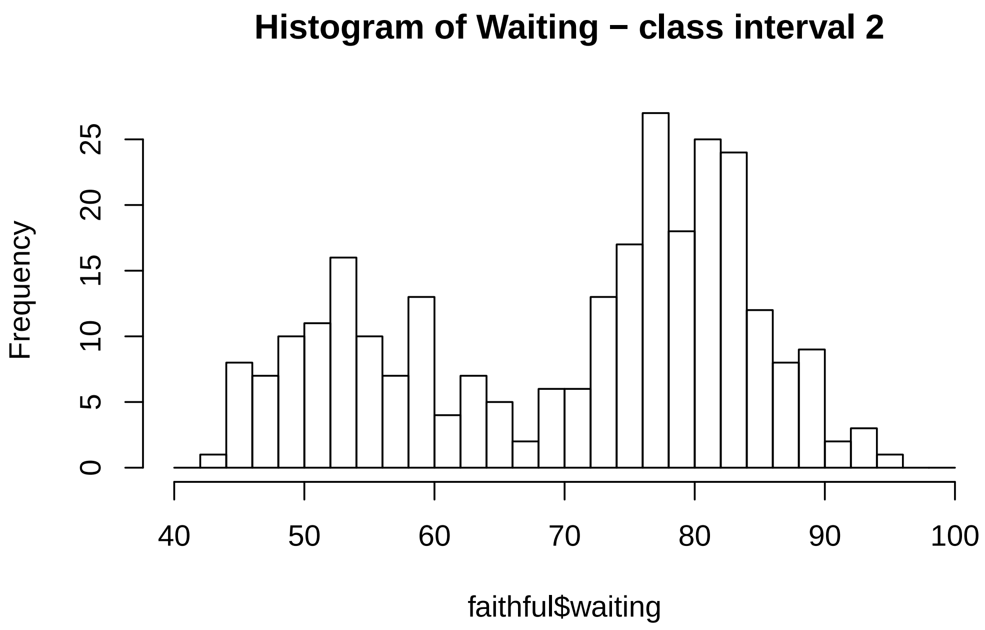
(iv)
```r
hist(faithful$eruptions,
     main = "Default Histogram of Eruptions")
```



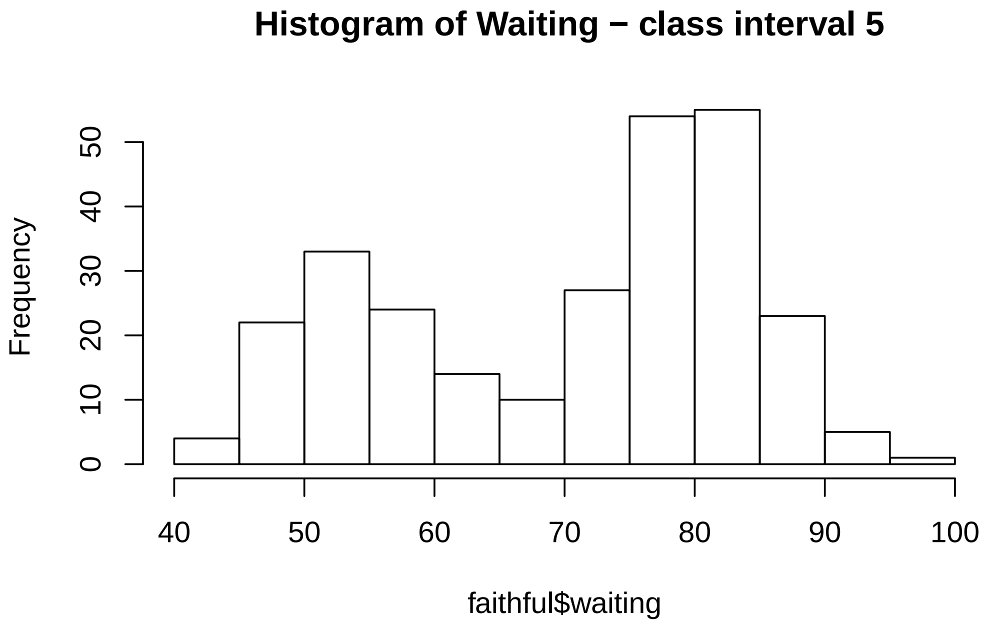

```r
hist(faithful$waiting,
     main = "Default Histogram of Waiting")
```



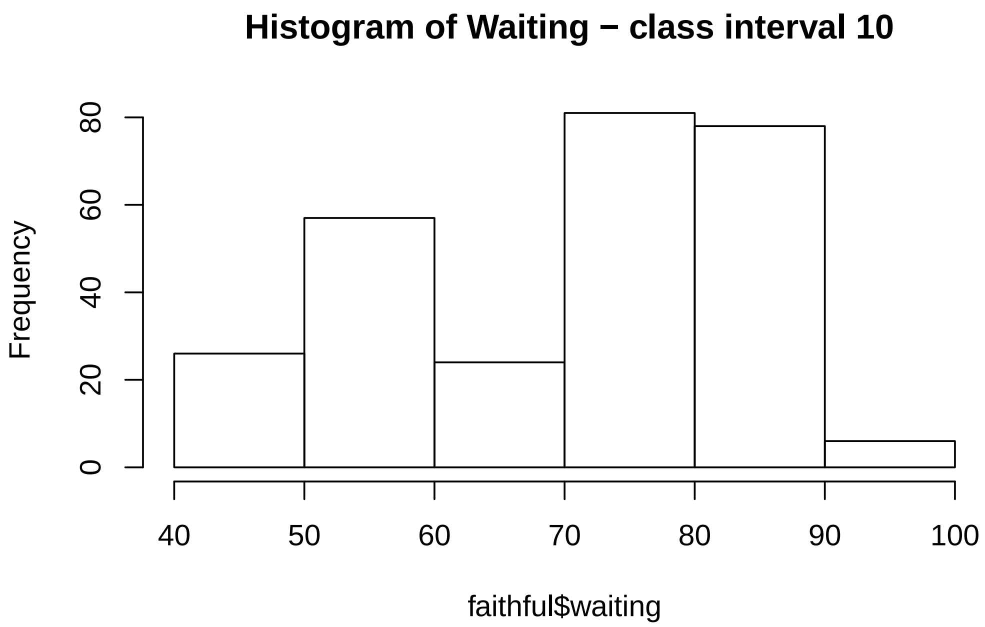

```r
hist(faithful$waiting,
     breaks = seq(40, 100, by = 2),
     main = " Histogram of Waiting - class interval 2")
```




```r
hist(faithful$waiting,
     breaks = seq(40, 100, by = 5),
     main = " Histogram of Waiting - class interval 5")
```




```r
hist(faithful$waiting,
     breaks = seq(40, 100, by = 10),
     main = " Histogram of Waiting - class interval 10")
```



(v) The histogram with class interval 5 may perhaps be judged to give the better illustration of the data. The smaller class interval 2 gives a picture that is too jagged, but the longer class interval 10 smoothes out too much of the detail of the distribution. However, such judgements are debatable.

A feature of the data that the histograms bring out is that for both variables the frequency distribution is what is known as *bimodal*, i.e., it has two peaks. This is not at all apparent from inspection of the descriptive statistics and the boxplots.