Assignment 2

Andrew Hayes, id = 21321503

29 September, 2022

## Starting with R-Markdown

In the following R-Markdown document some data are created followed by calculation of some summary statistics and display of graphical summaries. All the results are embedded for you in the report when you knit the document into a report.

The following R chunk creates a dataset in a vector and stores it in R’s memory using the name x. You will have been given some directions in how to adapt this dataset on Blackboard.

x = c(10, 23, 14, 12, 34, 26, 28, 24)

The mean of this data is

mean(x)

## [1] 21.375

The summary statistics (minimum, maximum, , median, mean and ) obtained from the summary() function are:

# Insert your code here  
summary(x)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 10.00 13.50 23.50 21.38 26.50 34.00

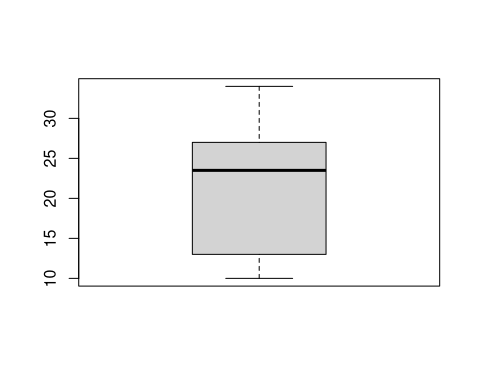
The five number summary which uses Tukey’s method to estimate the lower and upper quartiles ( and ) is given below. Notice the small differences in these quartiles.

# Insert your code here  
fivenum(x)

## [1] 10.0 13.0 23.5 27.0 34.0

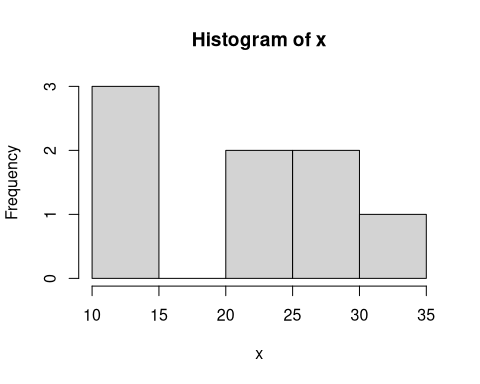
The boxplot of the data below also uses Tukey’s method. I would describe the shape of the distribution using the boxplot as right-skewed, as the tail on the right is significantly longer than the tail on the left. However, the median is offset to the right of the box, which would normally indicate a left-skew. One possible reason for this inconsistency is the small size of the dataset used for this boxplot, as boxplots are not very accurate for small data sets.

boxplot(x)



A histogram is given below. I would describe the shape of the distribution using the histogram as right-skewed, as it peaks on the left, and decreases as it goes to the right.

hist(x)



Use the help system in R to learn how to use the breaks argument in the hist function to include around 10 breakpoints. To use the help system type help(hist)

hist(x, breaks = 10)

