21001911 – Assignment 1

Question 1

```
set.seed(123)
```

```
> employee_data <- data.frame(Employee_ID = 1:100,Name=paste("Employee",1:100),Age=sample(22:60,100,r
eplace=TRUE),Salary=round(runif(100,min=50000,max=100000),2),YearsOfExperince = sample(1:20,100,repla
ce=TRUE),Perfromance_Rating=sample(1:5,100,replace=TRUE))
> write.csv(employee_data, "employee_data.csv",row.names=FALSE)
```

a.

```
> employee_data <- read.csv("employee_data.csv")
> str(employee_data)
'data.frame':   100 obs. of  6 variables:
 $ Employee_ID       : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Name              : chr  "Employee 1" "Employee 2" "Employee 3" "Employee 4" ...
 $ Age               : int  24 39 53 38 59 53 47 34 41 59 ...
 $ Salary            : num  58699 90071 57314 91136 66550 ...
 $ YearsOfExperince  : int  8 3 16 1 12 18 14 8 5 17 ...
 $ Perfromance_Rating: int  5 1 1 1 5 5 4 2 2 4 ...
```

b.

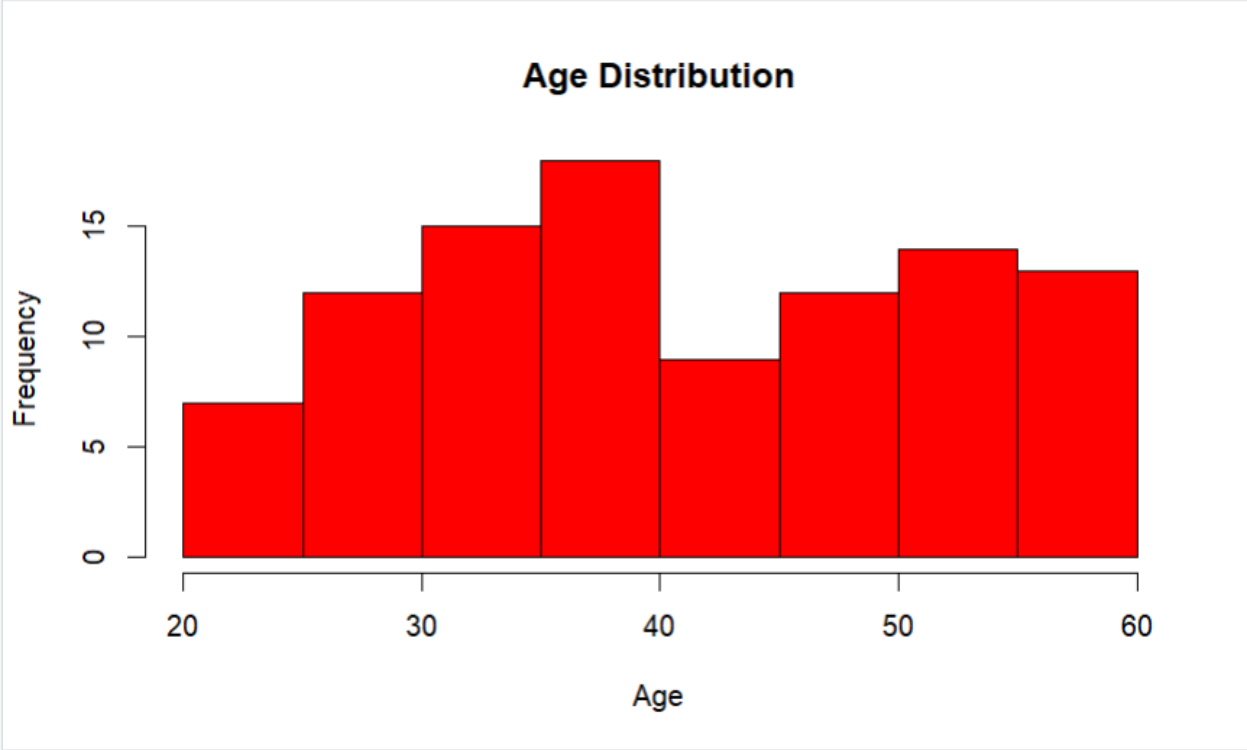
```
> avg_age <- mean(employee_data$Age)
> avg_age
[1] 41.3
```

c.

```
> salary_range <- range(employee_data$Salary)
> salary_range
[1] 51003.70 99652.24
```

d.

```
> hist(employee_data$Age,main="Age Distribution", xlab="Age",col="red",border="black")
```



e.

```
> plot(employee_data$YearsOfExperince,employee_data$Salary, main="Salary over yeaers of experienc
e",xlab="Years Of Experince", ylab="Salary",pch=16,col="pink")
```

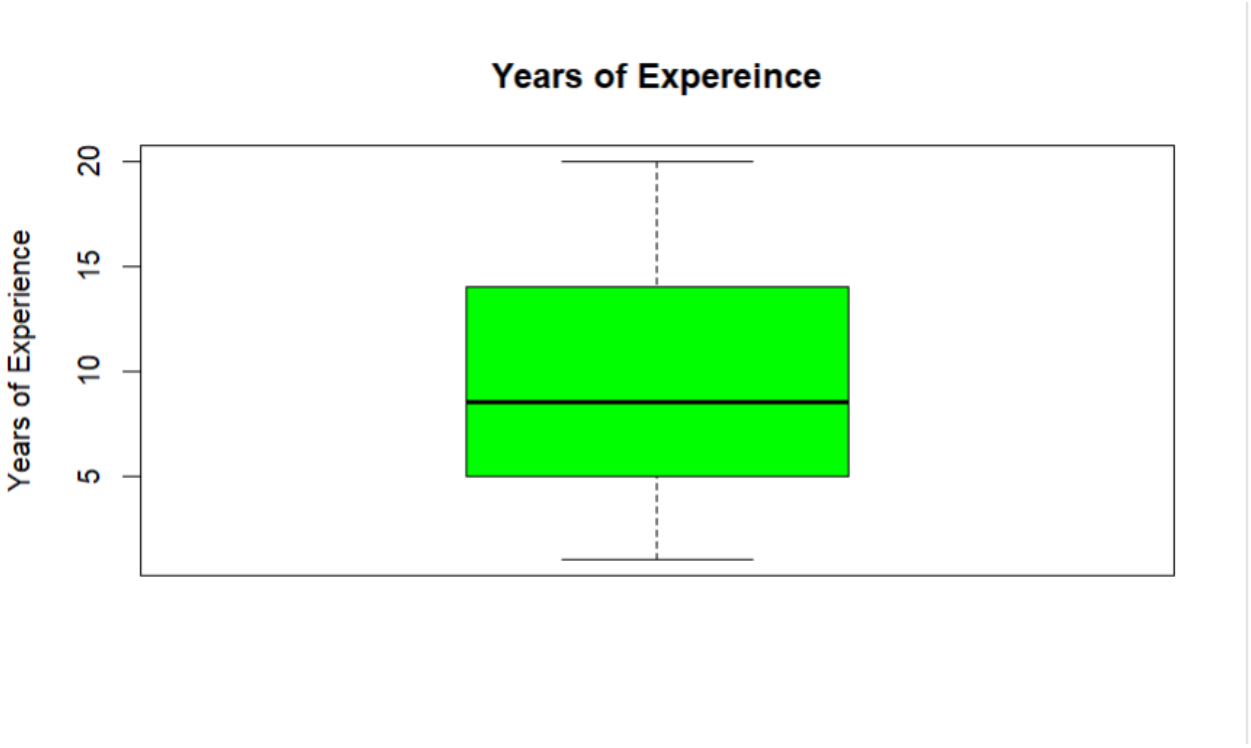
f.

```
> salary_mdeian <- median(employee_data$Salary)
> salary_avg <- mean(employee_data$Salary)
> salary_mdeian
[1] 76091.24
> salary_avg
[1] 75375.75
```

yes mean and median values are very close in values.

g.

```
> summary(employee_data$YearsOfExperince)
   Min. 1st Qu.  Median    Mean 3rd Qu.
   1.00    5.00    8.50    9.48   14.00
   Max.
  20.00
>
> boxplot(employee_data$YearsOfExperince,main="Years of Expereince",ylab="Years of Experience",col
="green")
```



2.

```
> install.packages("carData")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install
the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Chathuni Ranasinghe/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/carData_3.0-5.zip'
Content type 'application/zip' length 1821591 bytes (1.7 MB)
downloaded 1.7 MB

package 'carData' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\Chathuni Ranasinghe\AppData\Local\Temp\RtmpOWO2mI\downloaded_packages
> library(carData)
> data("Davis")
> summary(Davis)
 sex         weight          height          repwt
 F:112   Min.   : 39.0   Min.   : 57.0   Min.   : 41.00
 M: 88   1st Qu.: 55.0   1st Qu.:164.0   1st Qu.: 55.00
         Median : 63.0   Median :169.5   Median : 63.00
         Mean   : 65.8   Mean   :170.0   Mean   : 65.62
         3rd Qu.: 74.0   3rd Qu.:177.2   3rd Qu.: 73.50
         Max.   :166.0   Max.   :197.0   Max.   :124.00
                                         NA's   :17
     repht
 Min.   :148.0
 1st Qu.:160.5
 Median :168.0
 Mean   :168.5
 3rd Qu.:175.0
 Max.   :200.0
 NA's   :17
> male_proportion = sum(Davis$sex == "Male") / nrow(Davis)
> library(stats)
> conf_interval <- prop.test(sum(Davis$sex == "Male"), nrow(Davis), conf.level = 0.99)$conf.int
> males = subset(Davis, sex == "Male")
> females = subset(Davis, sex == "Female")
```