**Assignment 7.1**

**Question No.1:**

# To plot all the column values as histogram in a single command using the libraries of

> library(purrr)

> library(tidyr)

> library(ggplot2)

# and then using the following set of commands, where gather() converts the entire data frame into just two columns, where the first one would be the column name and the second one would be the value. So mtcars from a n x 11 column structure would convert into n x 2 column structure.

> mtcars %>% gather() %>% ggplot(aes(value)) + facet\_wrap(~ key, scales = "free") + geom\_histogram()

Else one can also use a loop programme in this fashion

> par(mfrow = c(2,3))

# To split the view of the plot area to accommodate the plots in 2 rows and 3 columns

> histcars <- for(col in 1:ncol(mtcars))

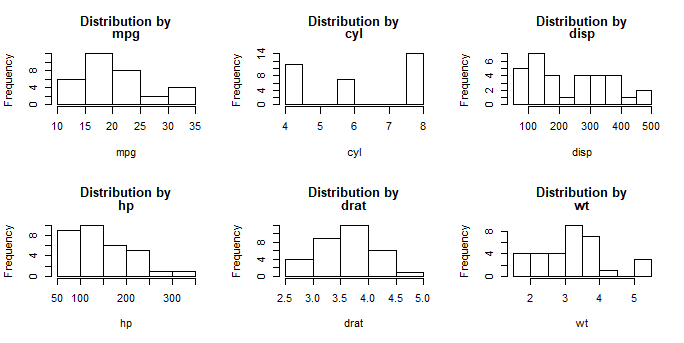
{

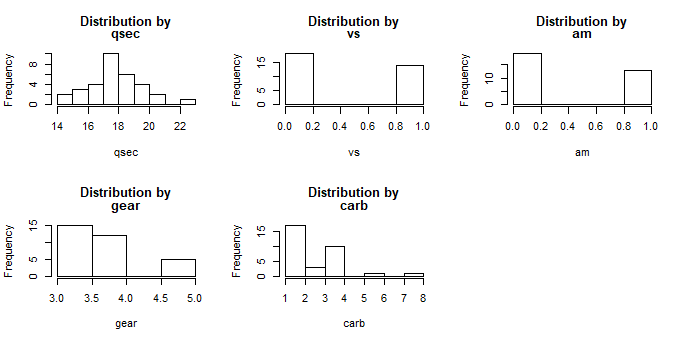
hist(mtcars[,col],

main = c("Distribution by", names(mtcars[col])),

xlab = names(mtcars[col]))

}





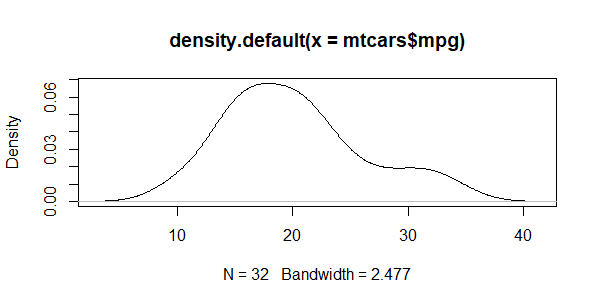
**Question NO: 2**

By observation of the histograms and the data, we can make an attempt to explain the possible probability distribution of the underlying data. Regarding the mtcars, the following distributions can be concluded

|  |  |
| --- | --- |
| **Variable** | **Distrbution** |
| mpg | Normal / Continuous variable – SW Test -88% |
| cyl | Multinomial / Discrete variable |
| disp | Multinomial / Discrete variable |
| hp | Multinomial / Discrete variable |
| drat | Normal / Continuous variable – SW Test – 89% |
| wt | Normal / Continuous variable – SW Test – 91% |
| qsec | Poisson /Normal / Continuous variable – SW Test – 41% |
| vs | Binomial / Discrete variable |
| am | Binomial / Discrete variable |
| gear | Multinomial / Discrete variable |
| carb | Multinomial / Discrete variable |

By observation the variables in mtcars have been classified into continuous and discrete variables. Later the continuous variables are tested for normality. The variables “mpg”, “drat”,”wt”, “qsec”, were tested for normality, using the Shapiro Wilk’s test, density plot, and qqnorm plot.

> plot(density(mtcars$mpg))



The above plot shows a skewed distribution towards the right, cannot comment on normality yet.

> shapiro.test(mtcars$mpg)

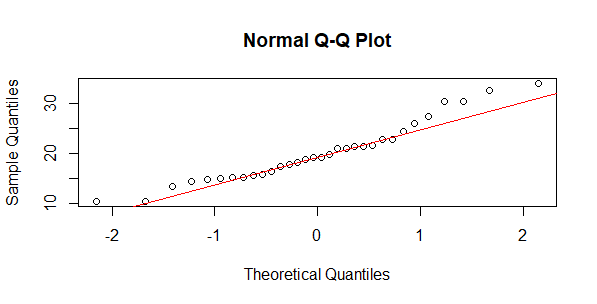
Shapiro-Wilk normality test

data: mtcars$mpg

W = 0.94756, p-value = 0.1229

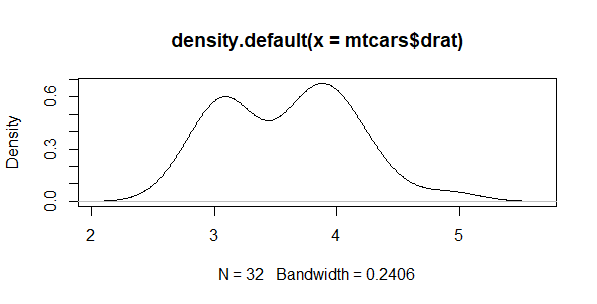
The above test has a p value greater than 0.05, so we cannot reject the null hypothesis “The variable is not normally distributed”. So in the absence of statistical significance we are forced to accept that the variable is normally distributed.

> qqnorm(mtcars$mpg); qqline(mtcars$mpg, col=2)



The above plot also does not show all the small balls falling on the straight red line to conclude that the date is perfectly normally distributed.

> plot(density(mtcars$drat))



The plot clearly shows that the distribution has two modes, so it is like a bimodal distribution and not like a perfect bell shaped curve. Also shows skewness to right and kurtosis also.

> shapiro.test(mtcars$drat)

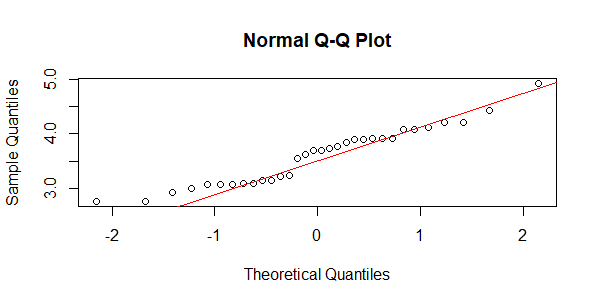
Shapiro-Wilk normality test

data: mtcars$drat

W = 0.94588, p-value = 0.1101

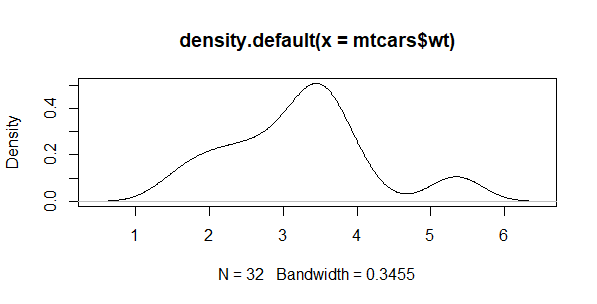
The p-value of the test is greater than 0.05, but it is 0.11, which means that with 89% confidence we can conclude that the data is normally distributed, but never with 95% confidence can we say that the “drat” data is normally distributed.

> qqnorm(mtcars$drat); qqline(mtcars$drat, col=2)



The qq plot shows that the data is not perfectly normally distributed. So the variable cannot be normally distributed and very few black balls are on the red line.

> plot(density(mtcars$wt))

****

The density plot shows that it is a skewed distribution with more observations on the left side of the mean.

> shapiro.test(mtcars$wt)

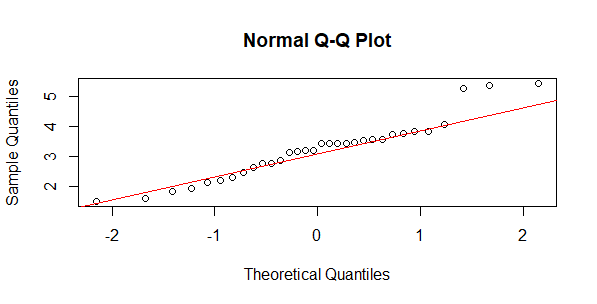
Shapiro-Wilk normality test

data: mtcars$wt

W = 0.94326, p-value = 0.09265

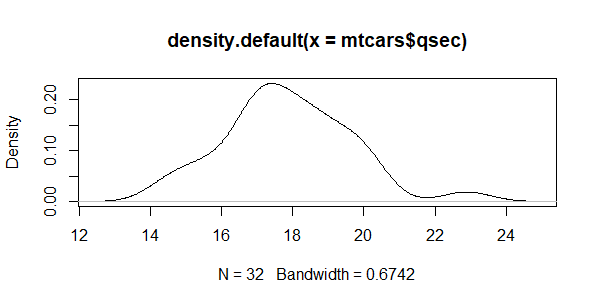
The Shapiro test rejects the normality assumption at 9.26% significance level so with 90.74% confidence we can say that the data is not normally distribution, but not at 95% confidence interval.

> qqnorm(mtcars$wt); qqline(mtcars$wt, col=2)



The qq plot also does not show that the data of “wt” is normally distributed. Only at 90% confidence it can be a non-normal distribution. Since we cannot reject the null hypothesis we will accept that the variable “wt” is normally distributed.

> plot(density(mtcars$qsec))

****

The density plot shows that there are high chances that the data might be normally distributed. But the nature and characteristic of the data gives the possibility of it being a poisson distribution since the data is expressed as a rate per a unit of time.

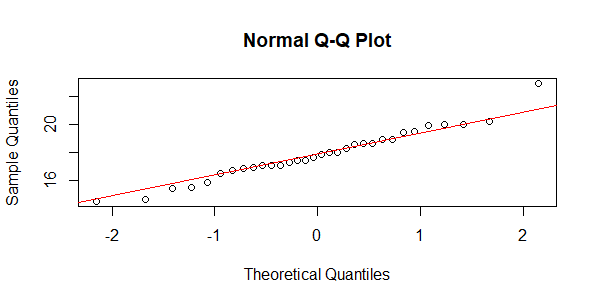
> shapiro.test(mtcars$qsec)

Shapiro-Wilk normality test

data: mtcars$qsec

W = 0.97325, p-value = 0.5935

The Shapiro’s test also confirms that there is no statistically significant evidence to believe that the data is not-normal.

> qqnorm(mtcars$qsec); qqline(mtcars$qsec, col=2)

**Question NO.3:**

> par(mfrow = c(1,3))

> boxcars <- for(col in 1:ncol(mtcars))

{

boxplot(mtcars[,col],main = c("distribution of",names(mtcars[col])),

xlab = names(mtcars[col]), ylab = "Value")

}

