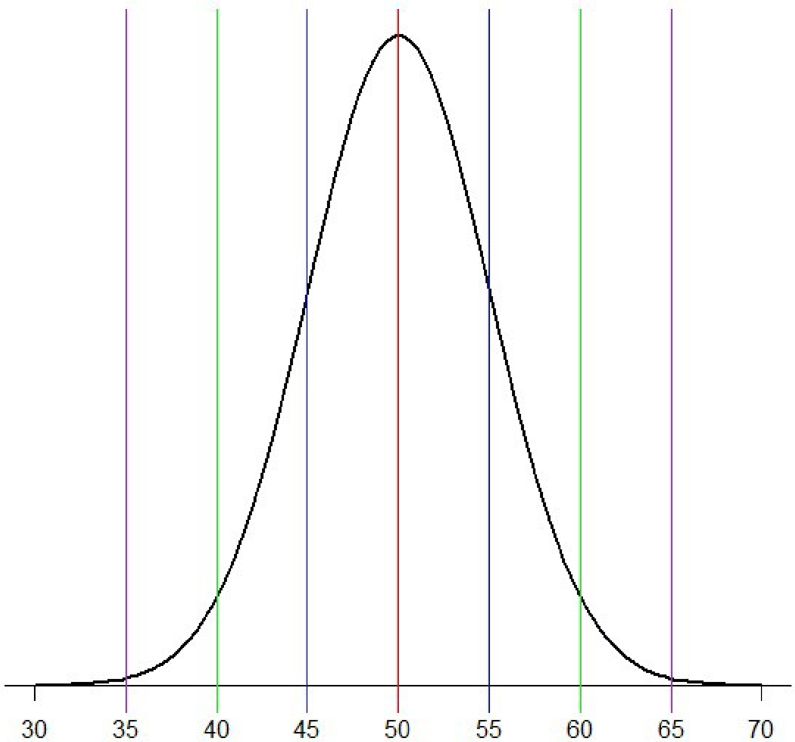
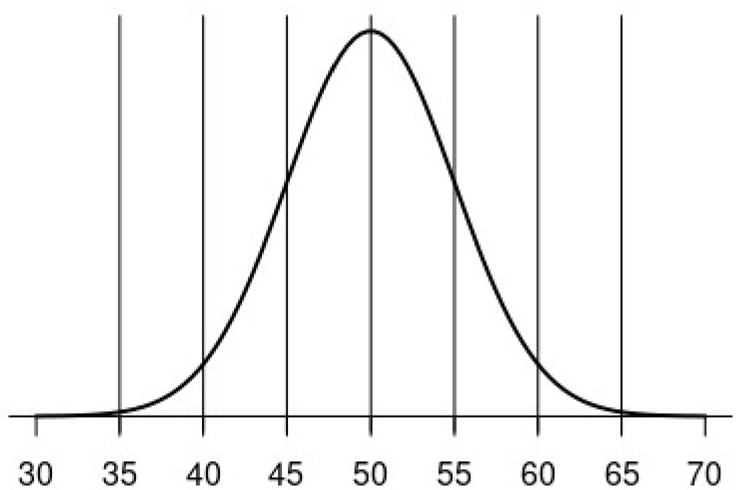
Assignment 1 (Unit 4)

i. Get this graph in R/R Studio using the codes provided in the class slide of session 18:




```
#Define mean and SD
 pop_mean<-50
 pop_sd<-5
 #Define lower and upper limits
 LL <-pop_mean-pop_sd
 UL <-pop_mean+ pop_sd
 #Create a sequence of 100 x values based on pop mean and sd
 x <-seq(-4,4, length=100)*pop_sd+pop_mean
 y <-dnorm(x, pop_mean, pop_sd)
 plot(x,y, type="l", lwd=2, axes=F, xlab="", ylab="")
 sd_axis_bounds= 5
 axis_bounds<-seq(-sd_axis_bounds*pop_sd + pop_mean, sd_axis_bounds*pop_sd+ pop_mean,
by=pop_sd)
 axis(side=1, at=axis_bounds, pos=0)
 abline(v=50)
 abline(v=50+pop_sd)
```

```
abline(v=50-pop_sd)
abline(v=50+2*pop_sd)
abline(v=50-2*pop_sd)
abline(v=50+3*pop_sd)
abline(v=50-3*pop_sd)
```



ii. Add annotation in this graph as follows:

45-55: mean ± 1SD = 67% data with double edge arrow

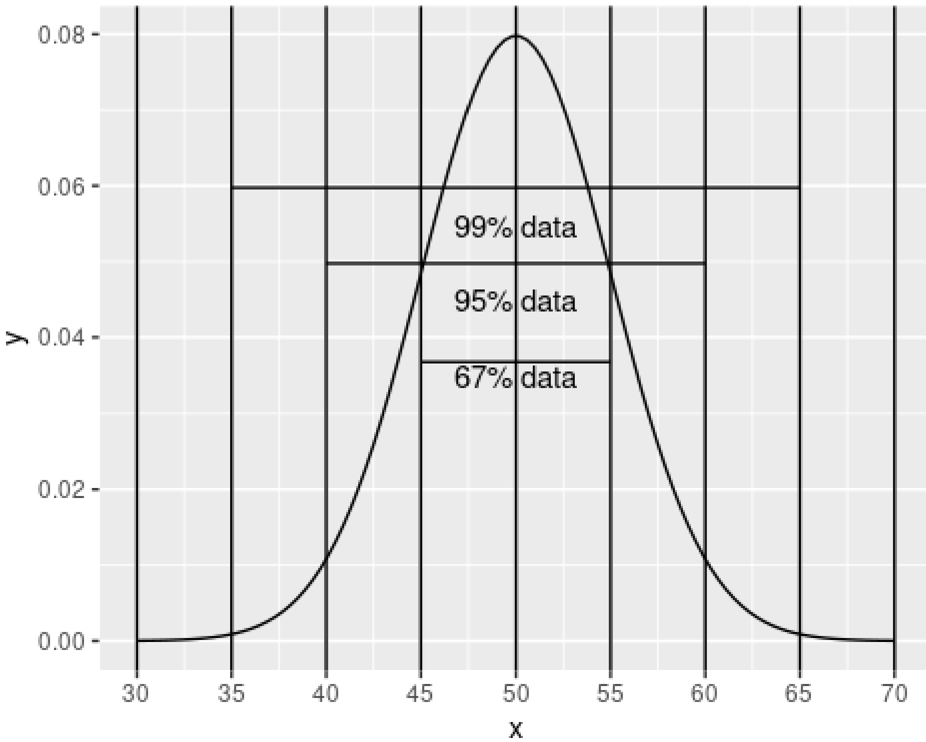
40-60: mean ± 2SD = 95% data with double edge arrow

35-65: mean ± 3SD = 99% data with double edge arrow

Note: You can use the "ggplot2" package if you want!

```
library(ggplot2)
 ggplot(data=as.data.frame(x,y),aes(x=x,y=y))+geom_line()+scale_x_continuous(breaks =
axis_bounds)+geom_vline(xintercept = seq(30,70,by=pop_sd))+annotate(geom = "text",x =
50,y=mean(y)+0.01, label = "67% data")+annotate(geom = "segment",x = 45,y=mean(y)
```

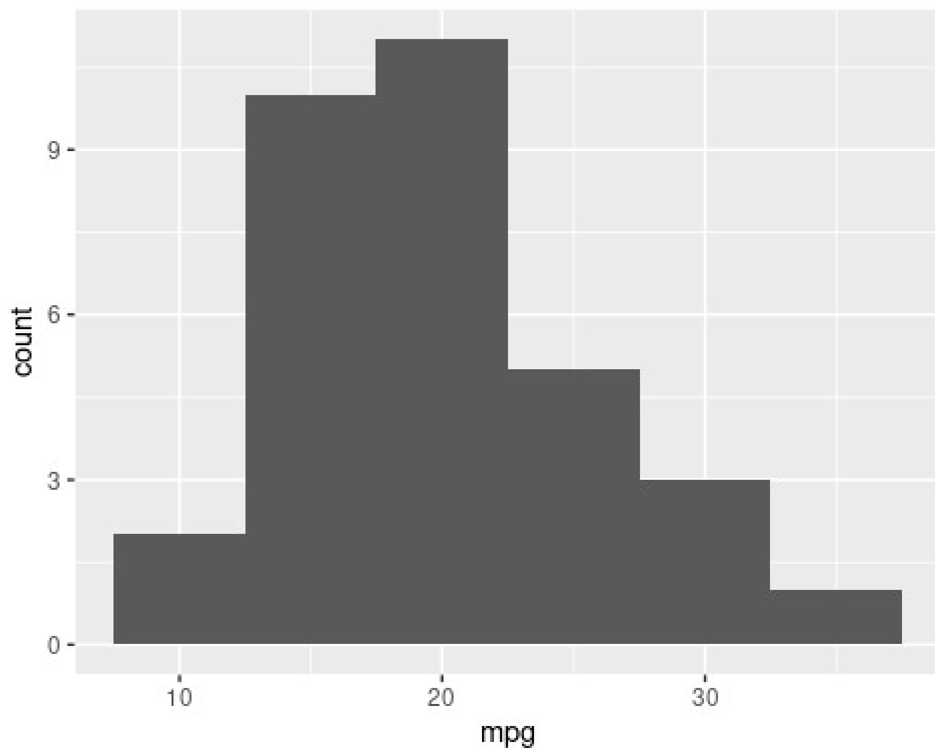
```
+0.012,xend = 55,yend = mean(y)+0.012)+annotate(geom = "text",x = 50,y=mean(y)+0.02, label
= "95% data")+annotate(geom = "segment",x = 40,y=mean(y)+0.025,xend = 60,yend = mean(y)
+0.025)+annotate(geom = "text",x = 50,y=mean(y)+0.03, label = "99% data")+annotate(geom =
"segment",x = 35,y=mean(y)+0.035,xend = 65,yend = mean(y)+0.035)
```



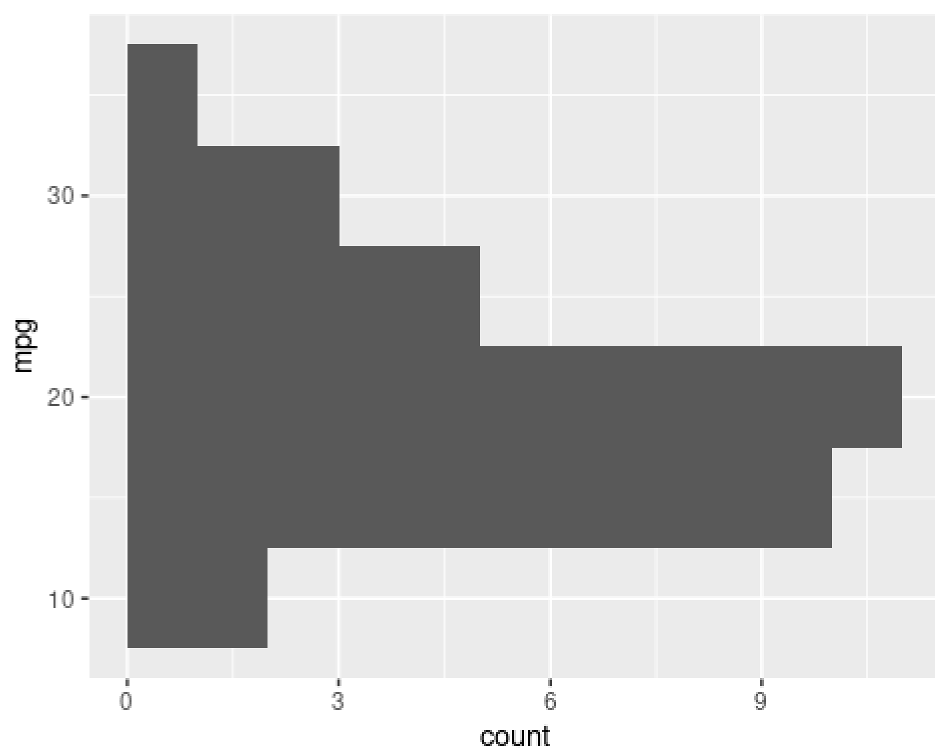
Assignment 2 (Unit 4):

i. Get stem-leaf plot, histogram and normal q-q plot of mpg variable of the built-in "mtcars" data of R.

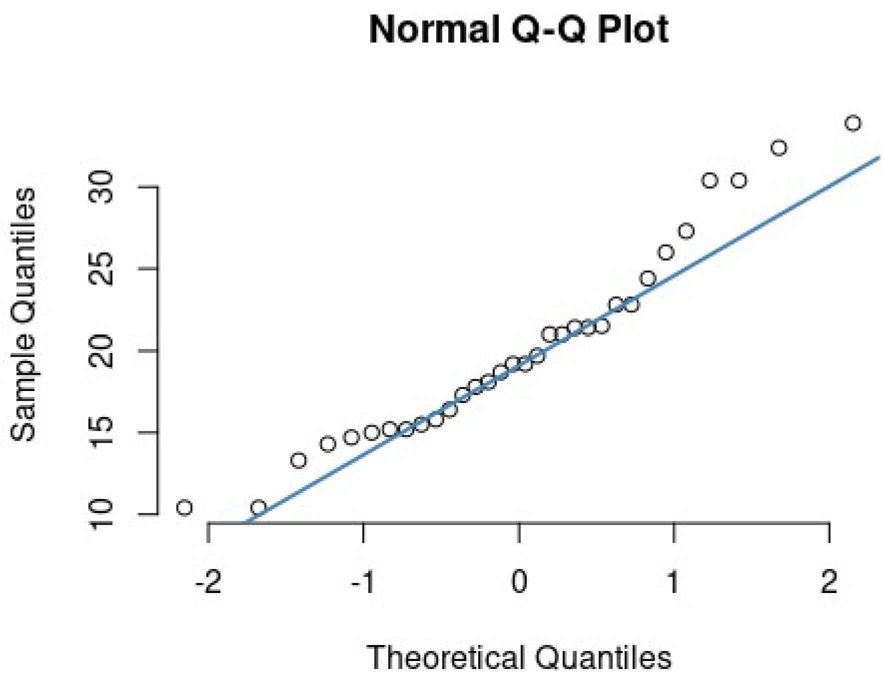
```
cars_data<-mtcars
 ggplot(data=cars_data,aes(x=mpg))+geom_histogram(binwidth = 5)
```




```
ggplot(data = cars_data,aes(mpg))+geom_histogram(binwidth = 5)+coord_flip()
```



```
cars_data<-mtcars
 qqnorm(cars_data$mpg, pch = 1, frame = FALSE)
 qqline(cars_data$mpg, col = "steelblue", lwd = 2)
```



ii. Test the normality of mpg variable of mtcars data using Shapiro-Wilk test of normality and take decision based on the p-value of this test for null and alternative hypothesis you wrote for this test. Explain why this test of normality must be used here!

```
shapiro.test(cars_data$mpg)

##
## Shapiro-Wilk normality test
 ##
## data: cars_data$mpg
 ## W = 0.94756, p-value = 0.1229
```

Null Hypothesis(H0): The data follows normal distribution Alternative Hypothesis(H1): The data does not follow normal distribution Since the p-value obtained from the test is not less than 0.05 we can say that the data comes from normal distribution i.e Null Hypothesis is accepted.

Since the number of samples (32)<50 we cannot use Kolmogorov-Smirnov. Thus, we have to use Shapiro-Wilk normality test.