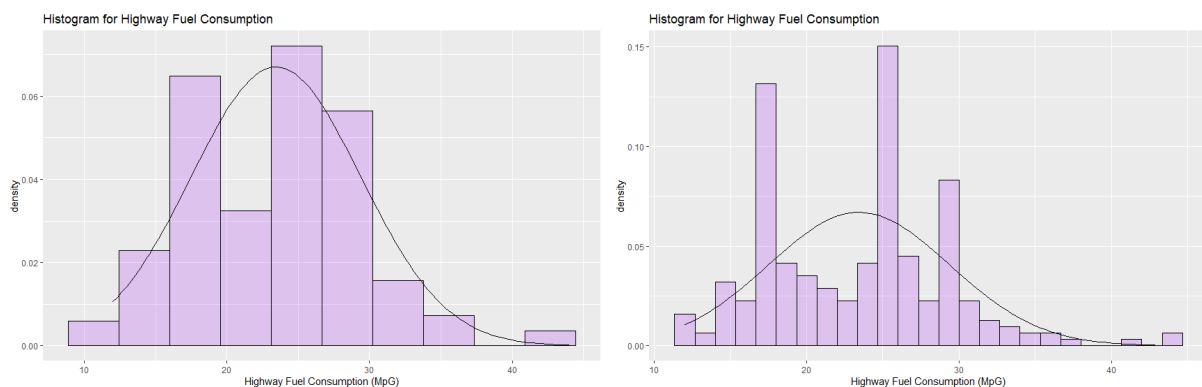


Assignment 8

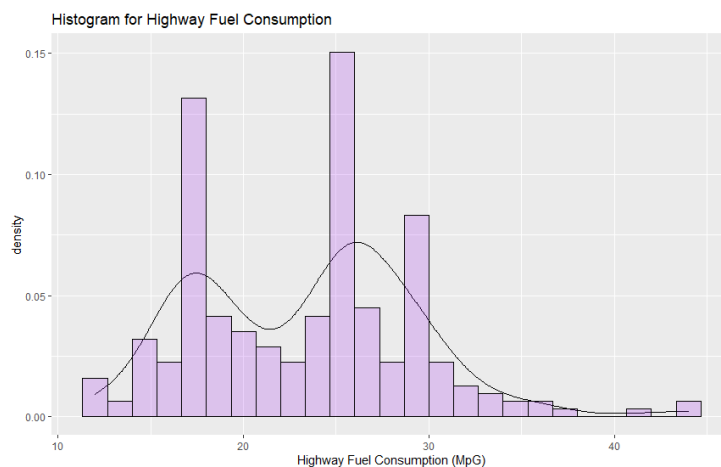
1. Draw a histogram of the *hwy* variable in the *mpg* dataset.

```
ggplot(mpg, aes(x=hwy)) +  
  geom_histogram(aes(y = ..density..),  
                 bins = 10, color = "black",  
                 fill = "purple", alpha = 0.2) +  
  stat_function(fun = dnorm, args = list(mean = mean(mpg$hwy), sd = sd(mpg$hwy))) +  
  labs(x= 'Highway Fuel Consumption (MpG)', title = 'Histogram for Highway Fuel Consumption')
```



2. Can it be argued that the *hwy* variable follows a normal distribution?

After watching the histogram with 10 bins, it may seem that *hwy* follows a normal distribution but the histogram with 25 bins clearly shows this is not the case. After performing a Shapiro-Wilk test the results confirm the claim that *hwy* does not follow a normal distribution. This is clear because the fuel consumption depends on the number of cylinders and the dataset has information for several number of cylinders. Therefore, it can be expected to have two normal distributions for 4 and 8 cylinders, like in the following histogram.



We can see that *hwy* follows a multi modal distribution.

"Written by Carlos Vintimilla"