Case Study 2

Mohamed Megahed

4/19/2020

# Class Practice - 5 (Univariate Analysis)

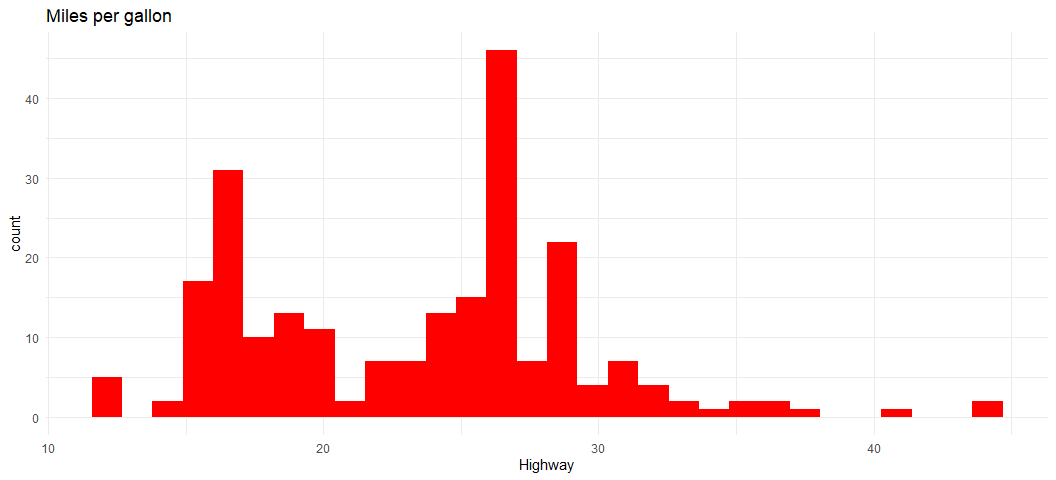
library(ggplot2)

**1. Plot the histogram using ggplot()**

*ggplot(cars, aes(hwy)) + geom\_histogram()*

ggplot(mpg, aes(hwy)) + geom\_histogram(fill = "red")+ labs(x = "Highway", title = "Miles per gallon") +  
 theme\_minimal()

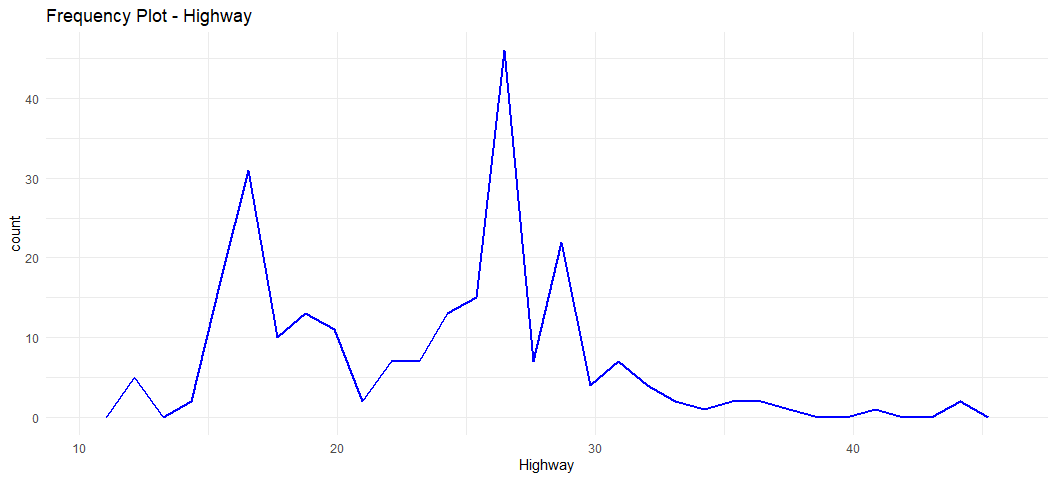
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



**2. Plot the distribution of the variables using geom\_freqpoly()** *ggplot(cars, aes(hwy)) + geeom\_freqpoly()*

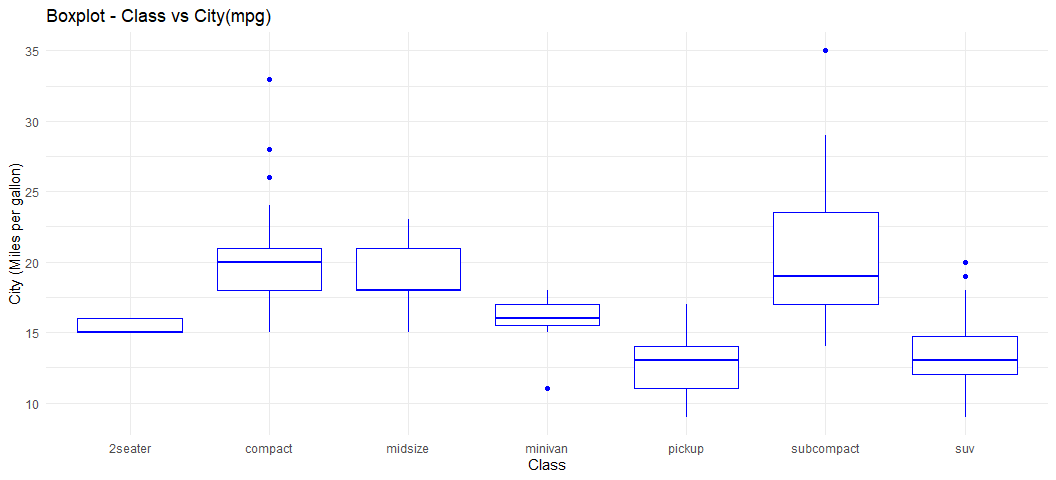
ggplot(mpg, aes(hwy)) + geom\_freqpoly(size = 0.9,colour = "blue") + labs(x = "Highway", title = "Frequency Plot - Highway") + theme\_minimal()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



**3. Plot box plot using geom\_boxplot()**

ggplot(mpg, aes(class,cty)) + geom\_boxplot(size = 0.5,colour = "blue") + labs(x = "Class",y= "City (Miles per gallon)", title = "Boxplot - Class vs City(mpg)") + theme\_minimal()



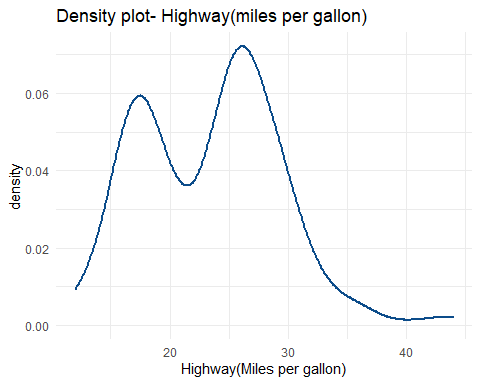
**4 What is the purpose of Histograms and Density plot?**

\_Histogram Plot: we use it to visualize the type of distribution usually for continous data. as it is a great way to get started exploring a single variable. A histogram divides the variable into bins, counts the data points in each bin, and shows the bins on the x-axis and the counts on the y-axis.

Density plots are usually a much more effective way to view the distribution of a variable

**5 Name another Univariate plot?** *geom\_density()*

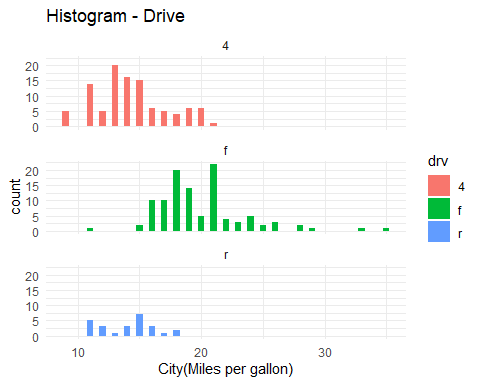
ggplot(mpg, aes(hwy)) + geom\_density(size = 0.8,colour = "#0c4c8a") + labs(x = "Highway(Miles per gallon)", title = "Density plot- Highway(miles per gallon)") + theme\_minimal()



**6. Plot the following using ggplot() + facet\_wrap()**

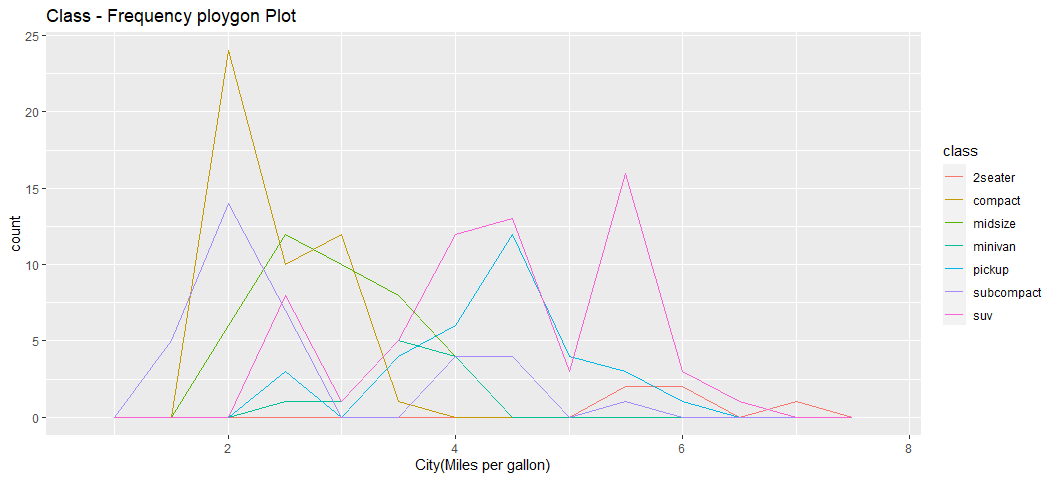
*ggplot(mpg, aes(cty, fill = drv)) + geom\_histogram(binwidth = 0.5) + facet\_wrap(~drv, ncol = 1)*

ggplot(mpg, aes(cty, fill = drv)) + geom\_histogram(binwidth = 0.5) + facet\_wrap(~drv, ncol = 1) + labs(x = "City(Miles per gallon)", title = "Histogram - Drive") + theme\_minimal()



**7. Plot the following graph:**

ggplot(mpg, aes(displ, colour = class)) + geom\_freqpoly(binwidth = 0.5) + labs(x = "City(Miles per gallon)", title = "Class - Frequency ploygon Plot")

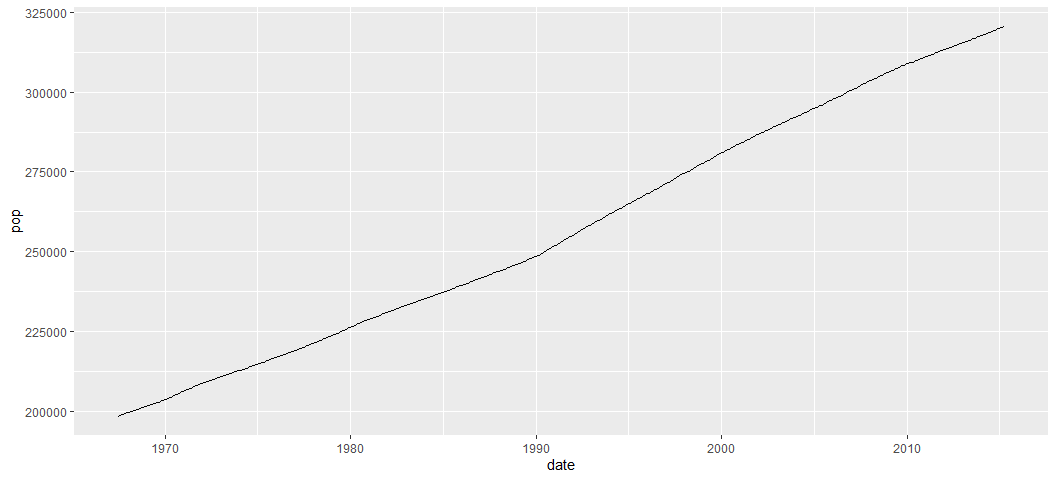


# Class Practice - 6

### Time Series:

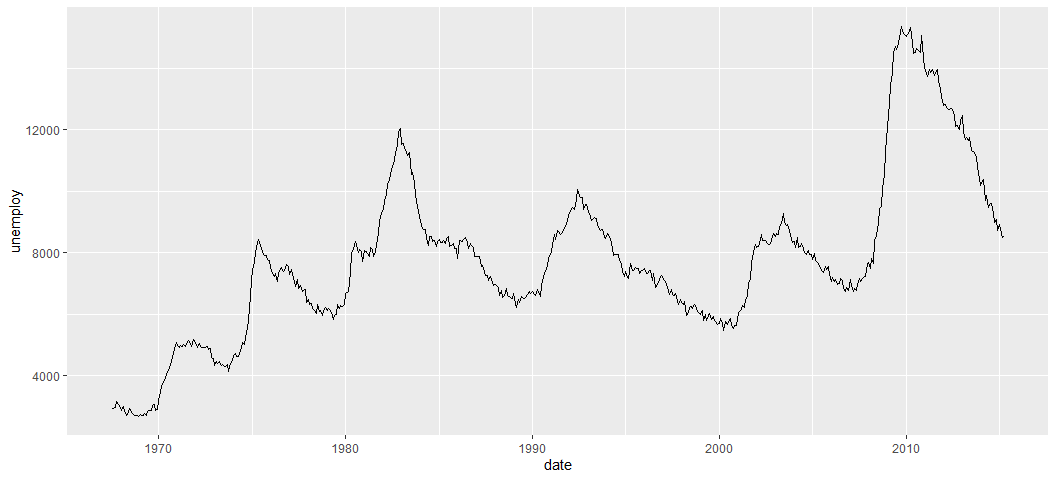
**1. Use the ‘economics’ dataset and Plot the following using geom\_line():** *ggplot(economics, aes(date, pop)) + geom\_line()*

ggplot(economics, aes(date, pop)) + geom\_line()



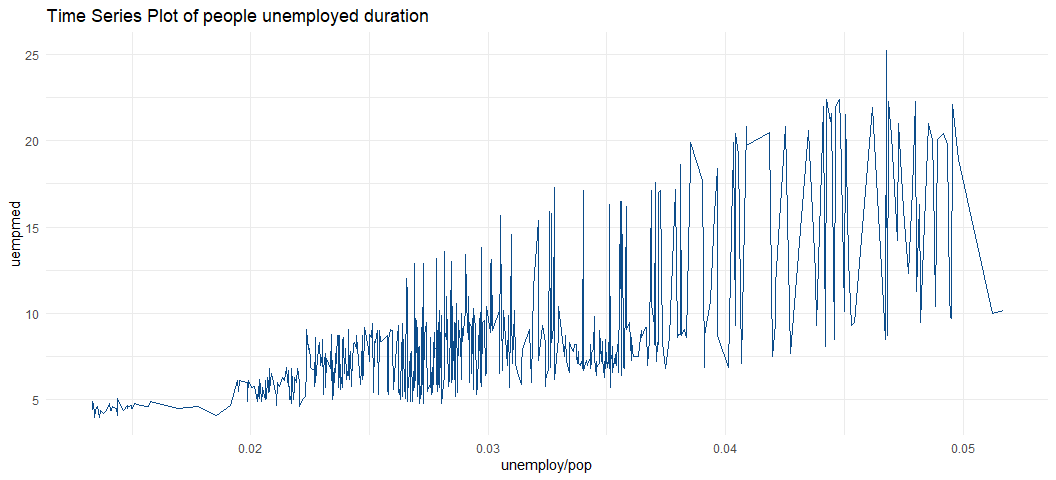
**2.Plot the unemployment growth rate over a period of time:** *ggplot(economics, aes(date, unemploy)) + geom\_line()*

ggplot(economics, aes(date, unemploy)) + geom\_line()

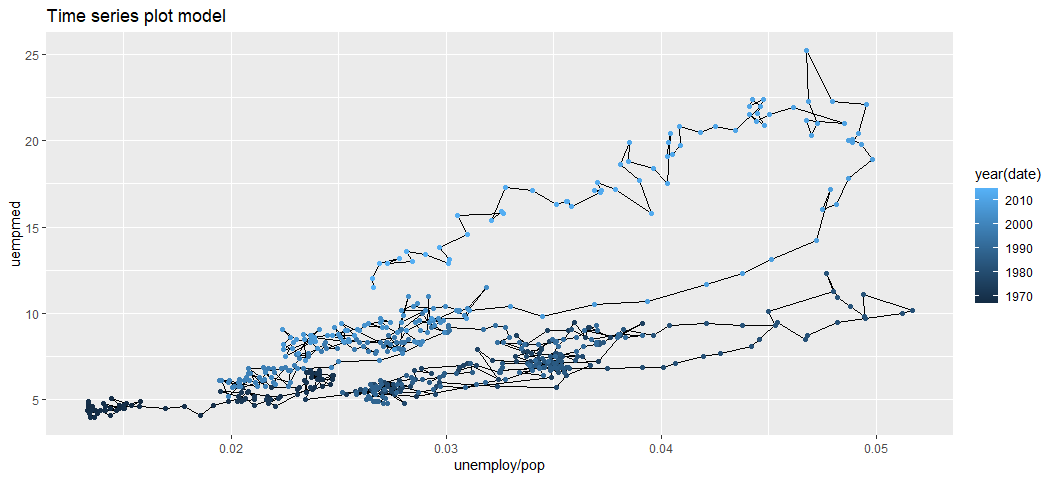


**3 Plot the graph to show how long people were unemployed?**

ggplot(economics, aes(unemploy / pop, uempmed)) + geom\_line(colour = "#0c4c8a") + ggtitle("Time Series Plot of people unemployed duration") + theme\_minimal()

 **4 Plot the below graph:** *(HINT: Use POSIXlt() ; ggplot() + geom\_path() + geom\_point()*

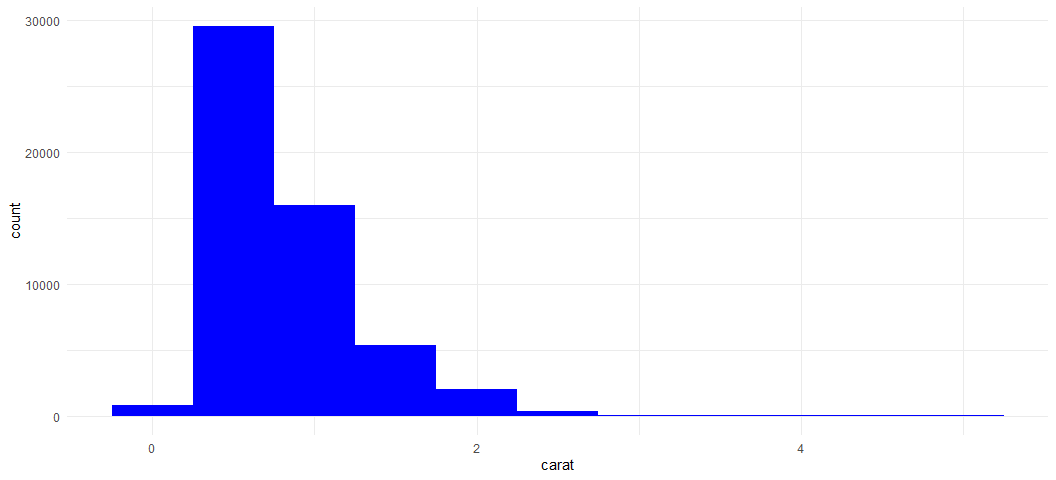
year <- function(x) as.POSIXlt(x)$year + 1900  
ggplot(economics, aes(unemploy / pop, uempmed)) + geom\_path() + geom\_point(aes(colour = year(date)))+ ggtitle("Time series plot model")



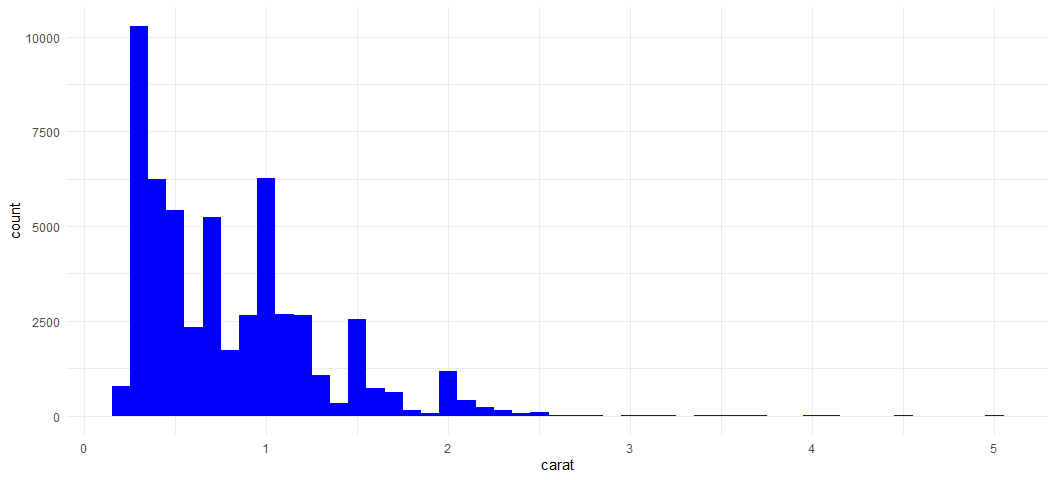
# Class Practice - 7

**1.Explore the distribution of the ‘carat’ variable in the diamonds dataset. What binwidth reveals the most interesting patterns?**

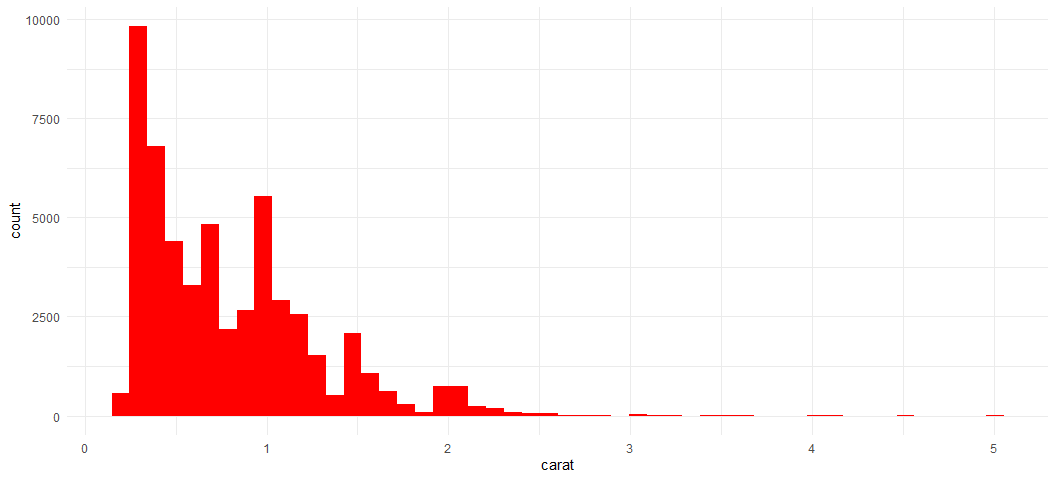
ggplot(diamonds,aes(carat))+geom\_histogram(binwidth=0.5,fill = "blue")+theme\_minimal()



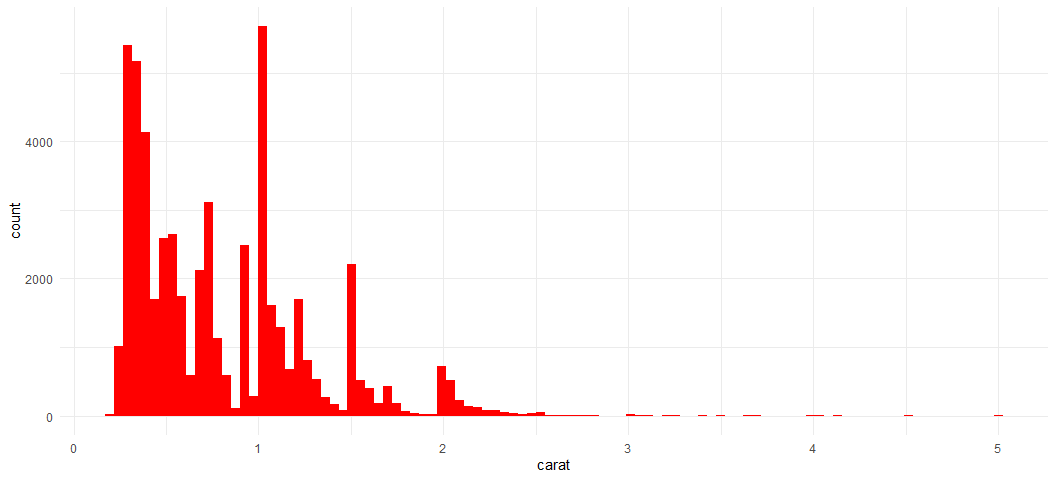
ggplot(diamonds,aes(carat))+geom\_histogram(binwidth=0.1,fill = "blue")+theme\_minimal()



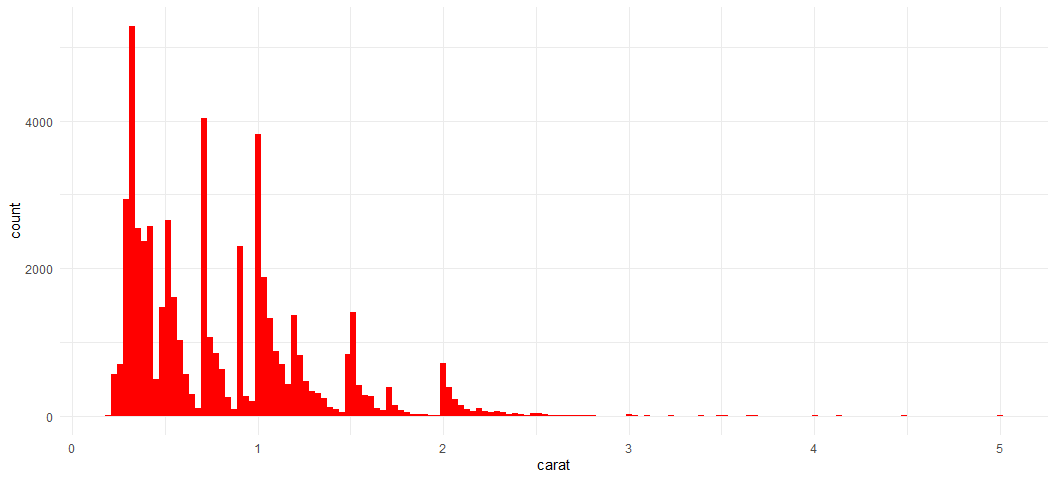
# Using different values of bins  
ggplot(diamonds,aes(carat))+geom\_histogram(bins=50,fill = "red")+theme\_minimal()



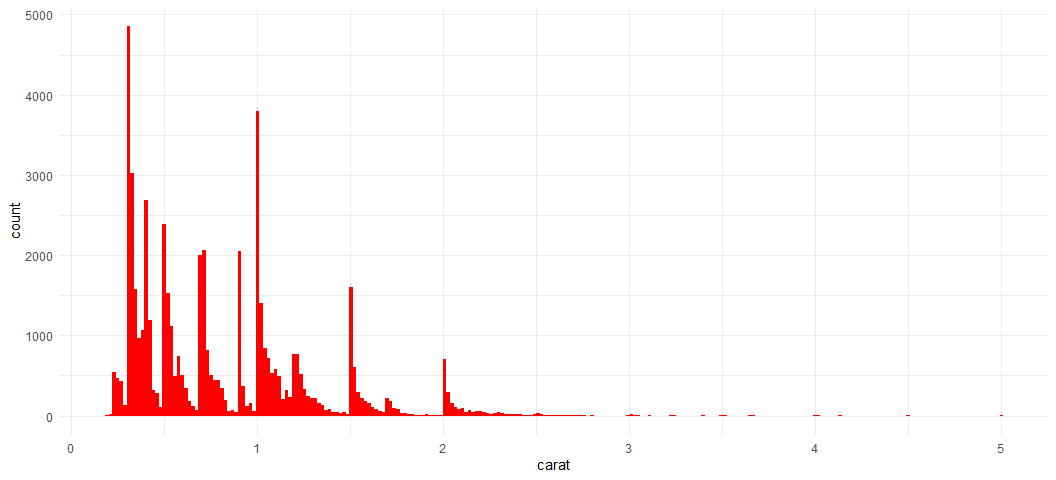
ggplot(diamonds,aes(carat))+geom\_histogram(bins=100,fill = "red")+theme\_minimal()



ggplot(diamonds,aes(carat))+geom\_histogram(bins=150,fill = "red")+theme\_minimal()



ggplot(diamonds,aes(carat))+geom\_histogram(bins=250,fill = "red")+theme\_minimal()



**2. Explore the distribution of the ‘price’ variable in the diamonds data. How does the distribution vary by cut?**

ggplot(diamonds,aes(price,fill=cut))+ geom\_density(alpha=0.3) + ggtitle("Density Distribution - Price Vs Cut")+theme\_minimal()

