Case Study 2

Mohamed Megahed

4/19/2020

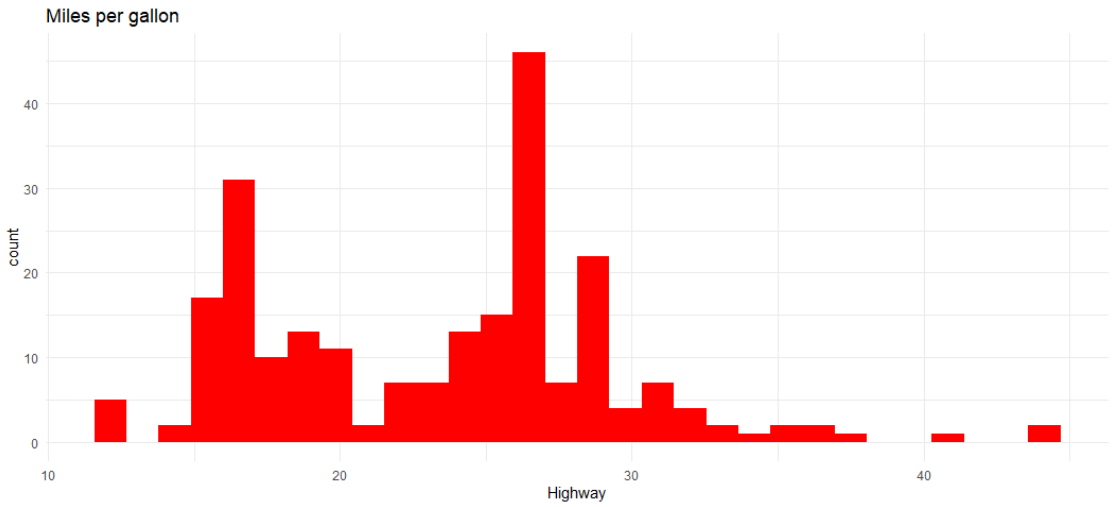
Class Practice - 5 (Univariate Analysis)

```
library(ggplot2)
```

1. Plot the histogram using ggplot()

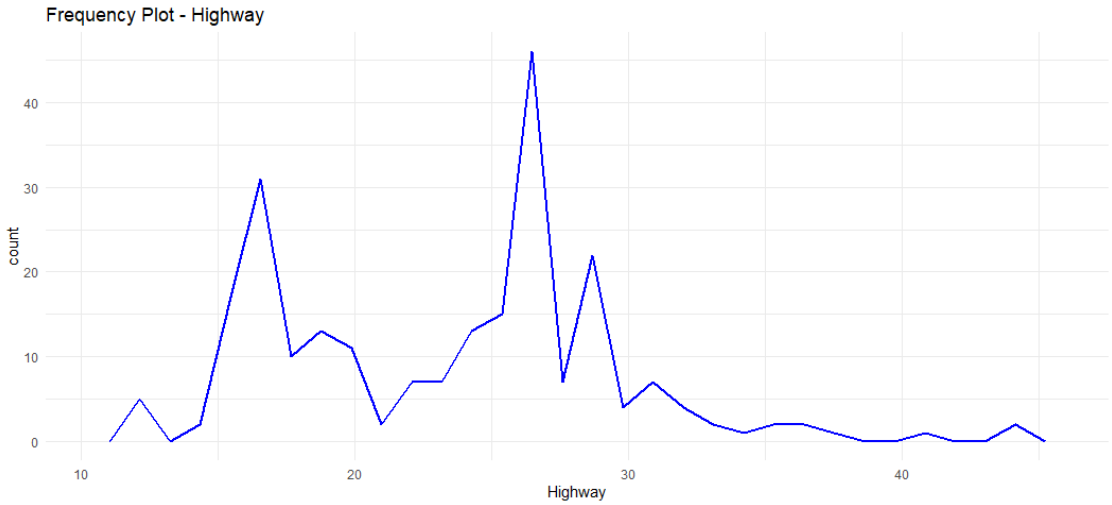
ggplot(cars, aes(hwy)) + geom_histogram()

```
ggplot(mpg, aes(hwy)) + geom_histogram(fill = "red")+ labs(x = "Highway", tit
le = "Miles per gallon") +
 theme_minimal()

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



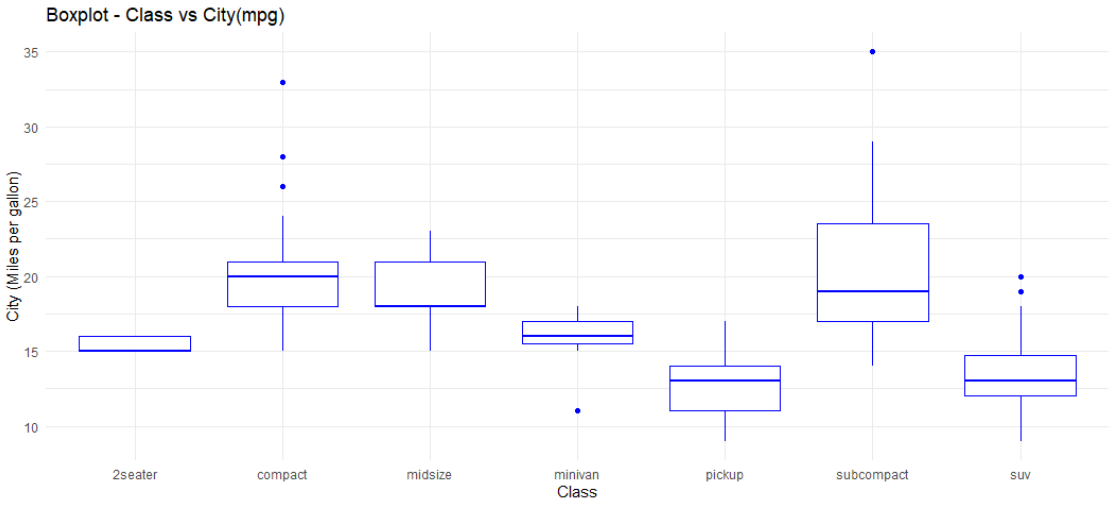
2. Plot the distribution of the variables using geom_freqpoly() *ggplot(cars, aes(hwy)) + geeom_freqpoly()*

```
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()
```



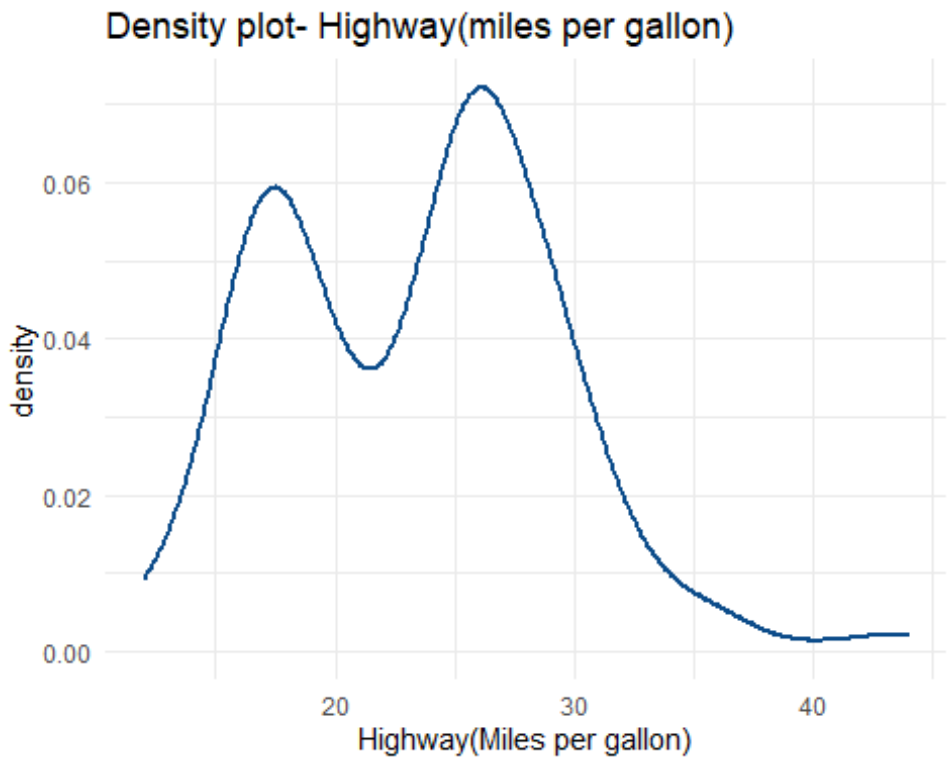
4 What is the purpose of Histograms and Density plot?

_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()*

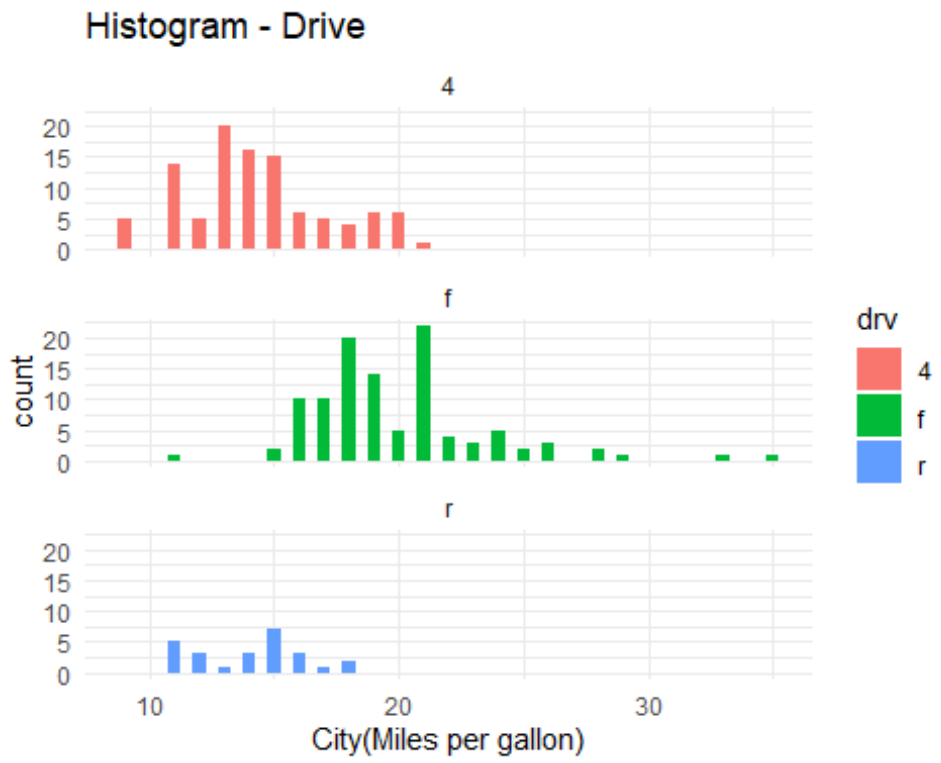
```
ggplot(mpg, aes(hwy)) + geom_density(size = 0.8,colour = "#0c4c8a") + labs(x = "Highway(Miles per gallon)", title = "Density plot- Highway(miles per gallon)") + theme_minimal()
```



6. Plot the following using ggplot() + facet_wrap()

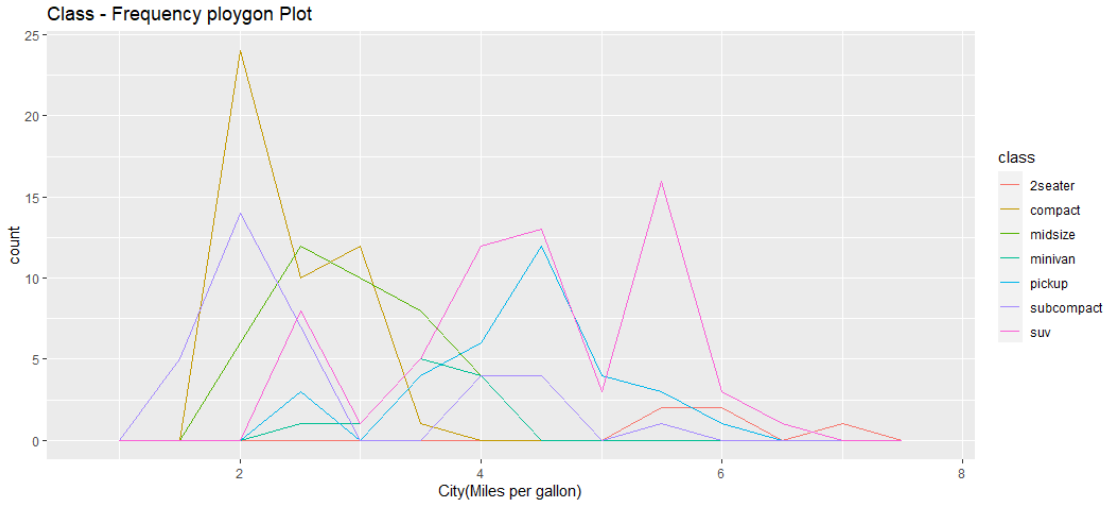
ggplot(mpg, aes(cty, fill = drv)) + geom_histogram(binwidth = 0.5) + facet_wrap(~drv, ncol = 1)

```
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")
```

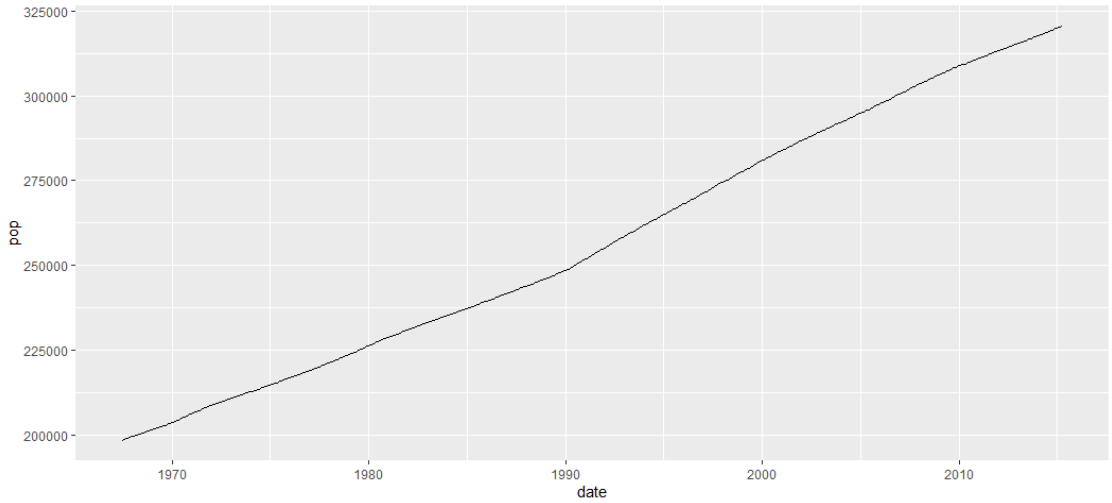


Class Practice - 6

Time Series:

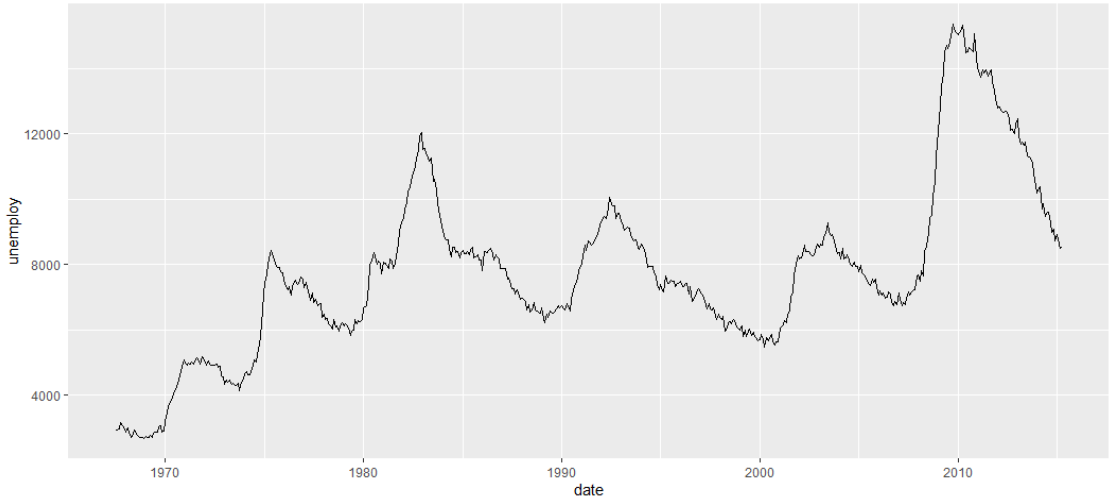
1. Use the 'economics' dataset and Plot the following using geom_line():
ggplot(economics, aes(date, pop)) + geom_line()

```
ggplot(economics, aes(date, pop)) + geom_line()
```



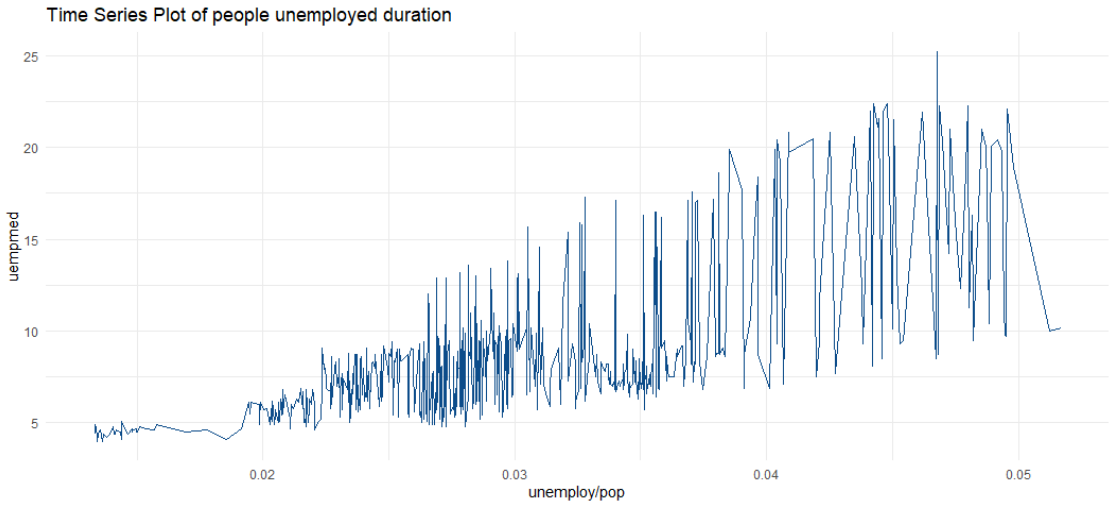
2.Plot the unemployment growth rate over a period of time: *ggplot(economics, aes(date, unemploy)) + geom_line()*

```
ggplot(economics, aes(date, unemploy)) +  geom_line()
```



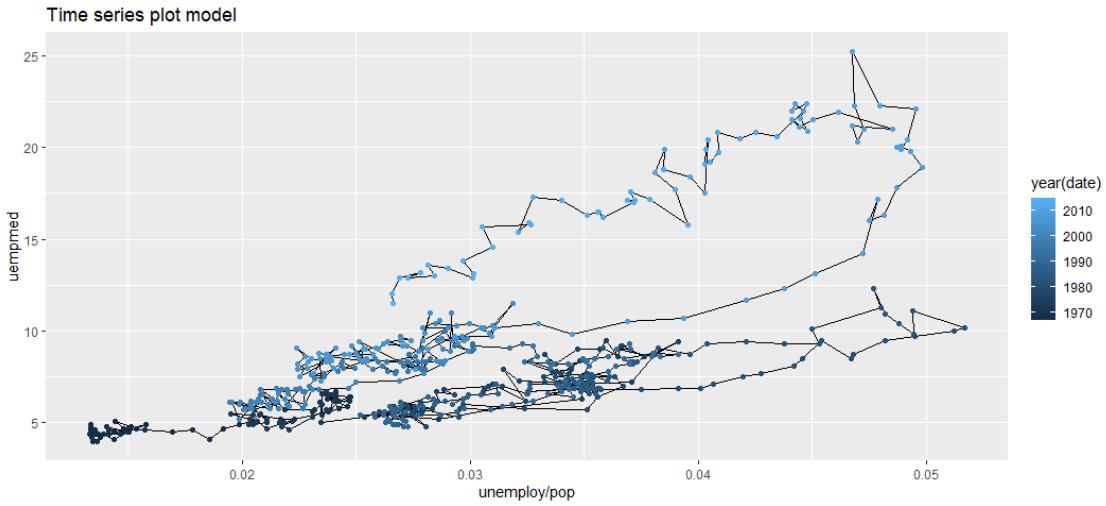
3 Plot the graph to show how long people were unemployed?

```
ggplot(economics, aes(unemploy / pop, uempmed)) + geom_line(colour = "#0c4c8a
") + ggtitle("Time Series Plot of people unemployed duration") + theme_minima
l()
```



4 Plot the below graph: *(HINT: Use POSIXlt() ; ggplot() + geom_path() + geom_point()*

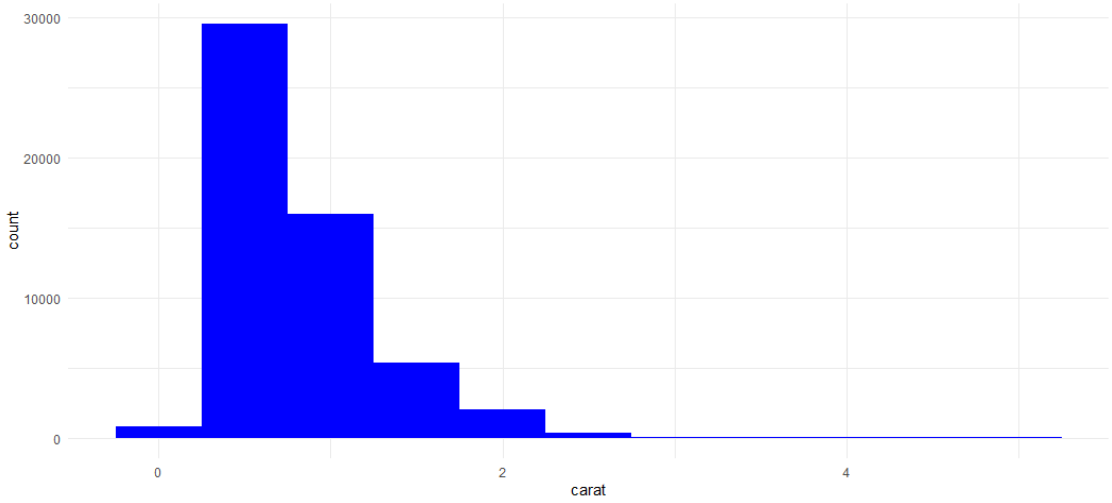
```
year <- function(x) as.POSIXlt(x)$year + 1900
ggplot(economics, aes(unemploy / pop, uempmed)) + geom_path() + geom_point(ae
s(colour = year(date)))+ ggtitle("Time series plot model")
```



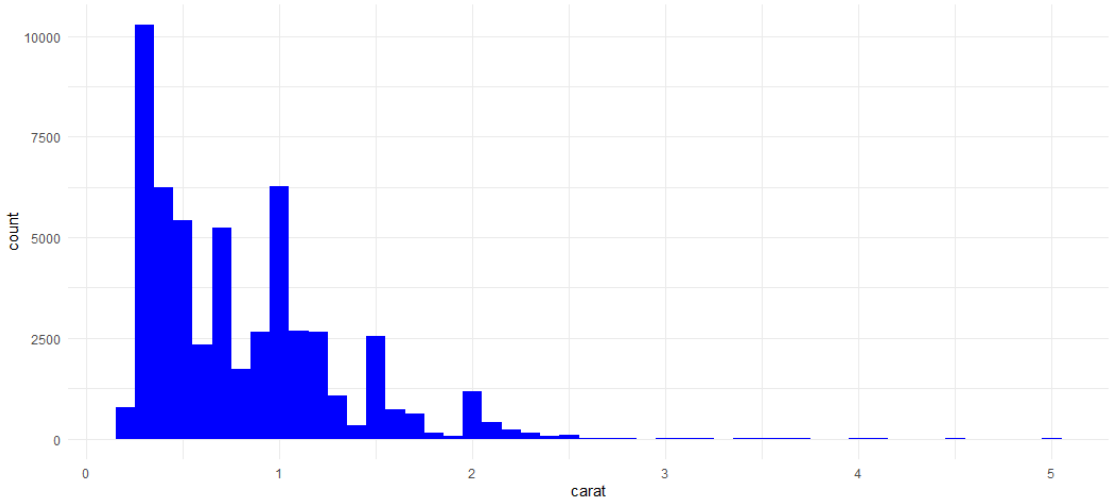
Class Practice - 7

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

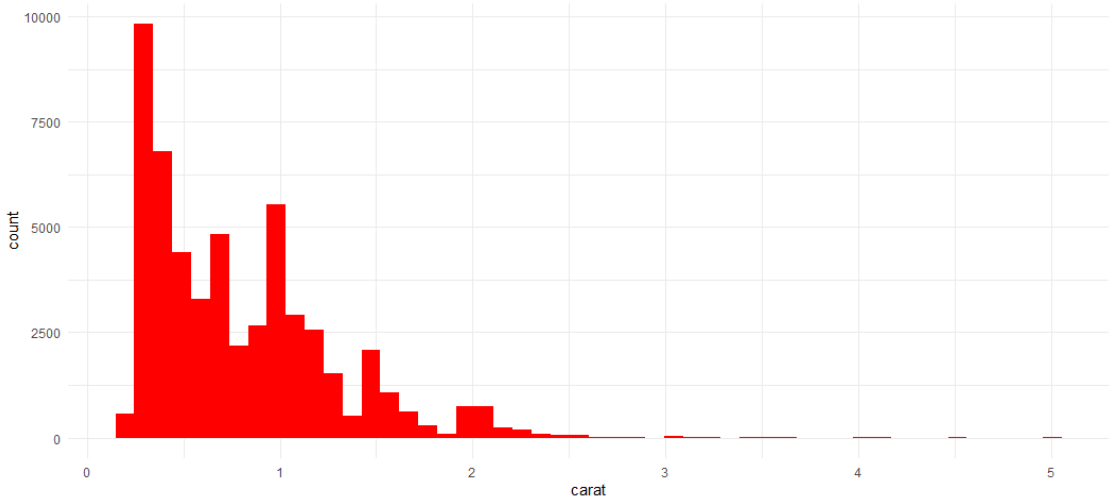
```
ggplot(diamonds,aes(carat))+geom_histogram(binwidth=0.5,fill = "blue")+theme_
minimal()
```



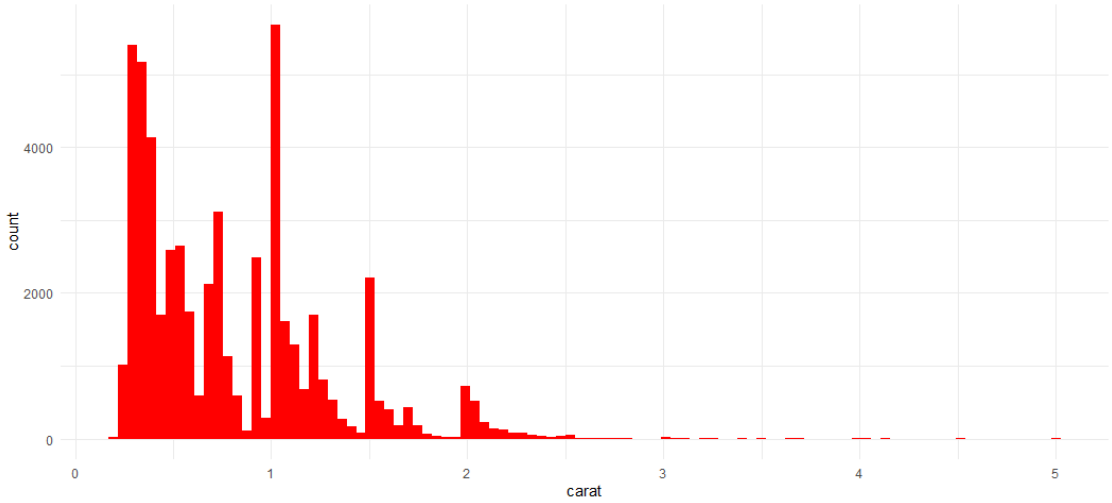

```
ggplot(diamonds,aes(carat))+geom_histogram(binwidth=0.1,fill = "blue")+theme_
minimal()
```



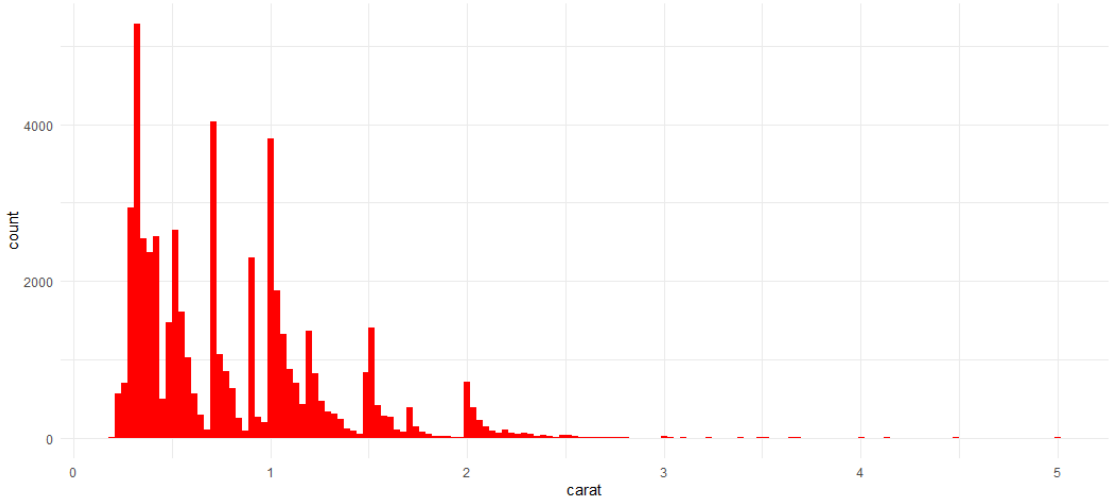

```
# Using different values of bins
ggplot(diamonds,aes(carat))+geom_histogram(bins=50,fill = "red")+theme_minima
l()
```




```
ggplot(diamonds,aes(carat))+geom_histogram(bins=100,fill = "red")+theme_minim
al()
```



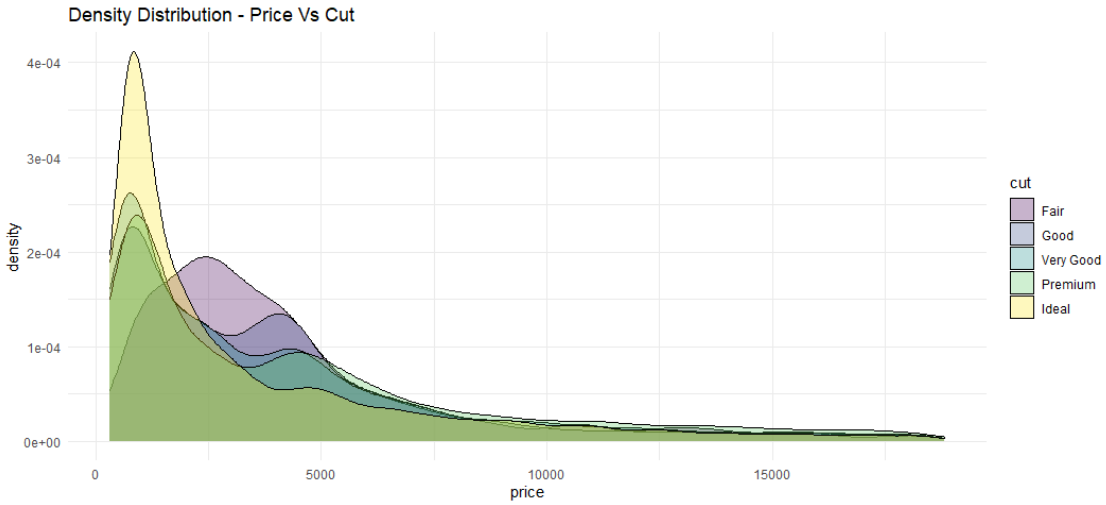

```
ggplot(diamonds,aes(carat))+geom_histogram(bins=150,fill = "red")+theme_minim
al()
```




```
ggplot(diamonds,aes(carat))+geom_histogram(bins=250,fill = "red")+theme_minim
al()
```

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("Dens
ity Distribution - Price Vs Cut")+theme_minimal()
```


Density Distribution - Price Vs Cut
4e-04
3e-04
2e-04
1e-04
0e+00
density
0
5000
10000
15000
price
cut
Fair
Good
Very Good
Premium
Ideal