Analytics in Practice

#Libraries used

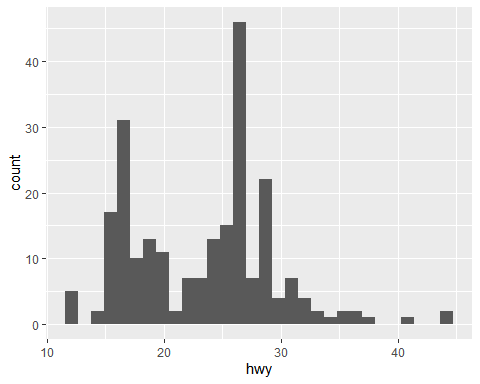
library(ggplot2)

#Class Practice 5

#1. Plot the histogram using ggplot()

ggplot(mpg, aes(hwy)) + geom\_histogram()

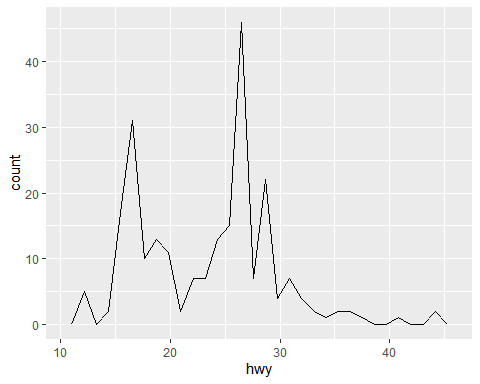
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#2. Plot the distribution of the variables using geom\_freqpoly()

ggplot(mpg, aes(hwy)) + geom\_freqpoly()

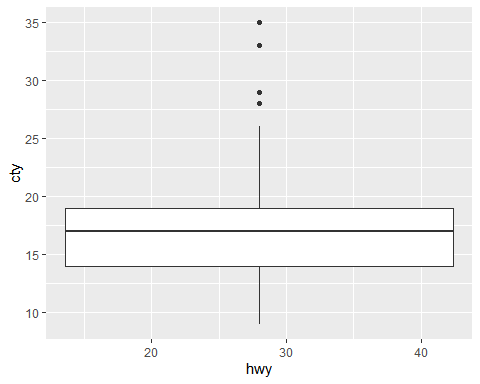
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



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

ggplot(mpg, aes(hwy,cty)) + geom\_boxplot()

## Warning: Continuous x aesthetic -- did you forget aes(group=...)?



#4. What is the purpose of Histograms and Density plot?

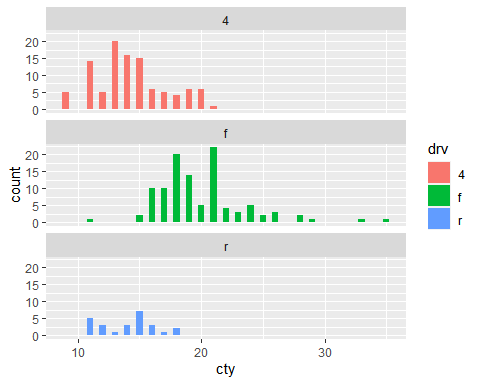
Histogram: A histogram divides the variable into bins, counts the data points in each bin, and shows the bins on the x-axis and the counts on the y-axis.

Density Plot: A Density Plot visualises the distribution of data over a continuous interval or time period. An advantage Density Plots have over Histograms is that they’re better at determining the distribution shape because they’re not affected by the number of bins used (each bar used in a typical histogram).

#5. Name another Univariate plot? Boxplot, Scatterplot

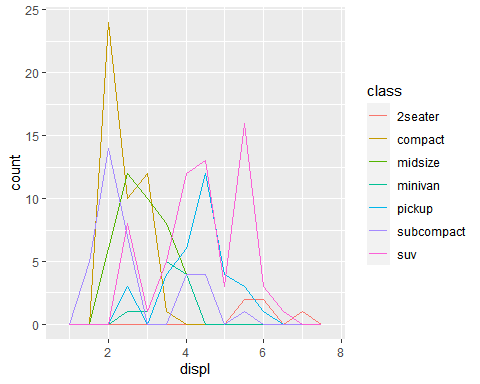
#6. Plot the following using ggplot() + facet\_wrap()

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



#7. Plot the following graph:

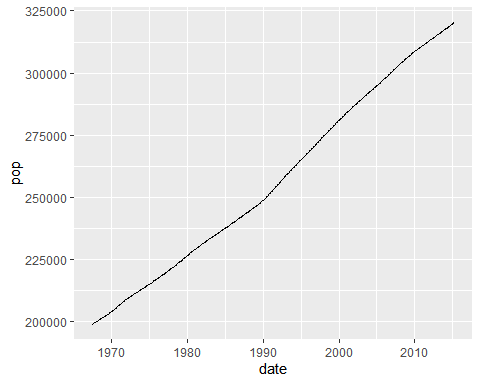
ggplot(mpg, aes(displ, colour = class)) + geom\_freqpoly(binwidth = 0.5)



#Class Practice 6

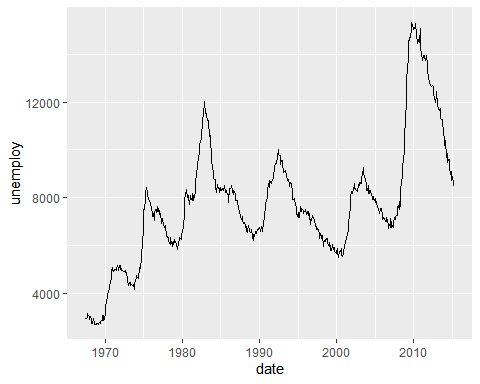
#1. Use the “economics” dataset and Plot the following using geom\_line():

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



#2. Plot the unemployment growth rate over a period of time:

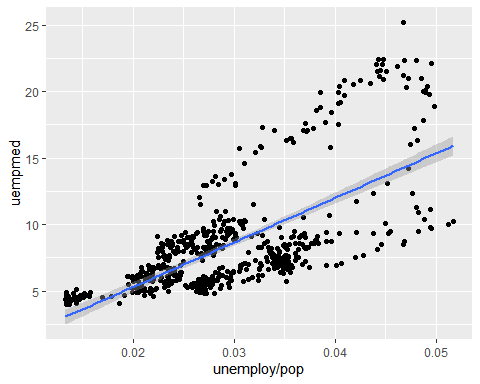
ggplot(economics, aes(date, unemploy)) + geom\_line()



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

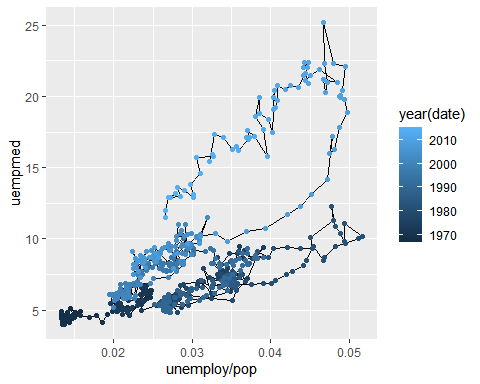
ggplot(economics, aes(unemploy / pop, uempmed)) + geom\_point() + geom\_smooth(method = "lm")

## `geom\_smooth()` using formula 'y ~ x'



#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)))



#Class Practice 6

#1. Explore the distribution of the “carat” variable in the diamonds dataset. What banwidth reveals the most interesting patterns?

Lower the binwidth, more interesting the patterns becomes.

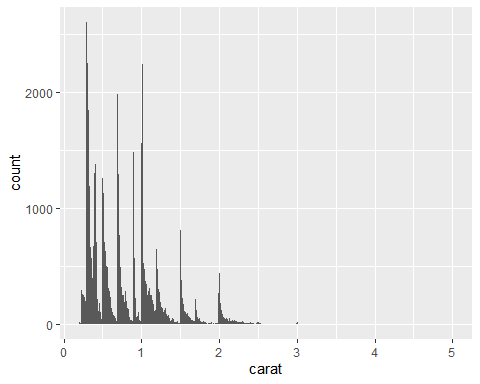
head(diamonds)

## # A tibble: 6 x 10  
## carat cut color clarity depth table price x y z  
## <dbl> <ord> <ord> <ord> <dbl> <dbl> <int> <dbl> <dbl> <dbl>  
## 1 0.23 Ideal E SI2 61.5 55 326 3.95 3.98 2.43  
## 2 0.21 Premium E SI1 59.8 61 326 3.89 3.84 2.31  
## 3 0.23 Good E VS1 56.9 65 327 4.05 4.07 2.31  
## 4 0.290 Premium I VS2 62.4 58 334 4.2 4.23 2.63  
## 5 0.31 Good J SI2 63.3 58 335 4.34 4.35 2.75  
## 6 0.24 Very Good J VVS2 62.8 57 336 3.94 3.96 2.48

summary(diamonds)

## carat cut color clarity depth   
## Min. :0.2000 Fair : 1610 D: 6775 SI1 :13065 Min. :43.00   
## 1st Qu.:0.4000 Good : 4906 E: 9797 VS2 :12258 1st Qu.:61.00   
## Median :0.7000 Very Good:12082 F: 9542 SI2 : 9194 Median :61.80   
## Mean :0.7979 Premium :13791 G:11292 VS1 : 8171 Mean :61.75   
## 3rd Qu.:1.0400 Ideal :21551 H: 8304 VVS2 : 5066 3rd Qu.:62.50   
## Max. :5.0100 I: 5422 VVS1 : 3655 Max. :79.00   
## J: 2808 (Other): 2531   
## table price x y   
## Min. :43.00 Min. : 326 Min. : 0.000 Min. : 0.000   
## 1st Qu.:56.00 1st Qu.: 950 1st Qu.: 4.710 1st Qu.: 4.720   
## Median :57.00 Median : 2401 Median : 5.700 Median : 5.710   
## Mean :57.46 Mean : 3933 Mean : 5.731 Mean : 5.735   
## 3rd Qu.:59.00 3rd Qu.: 5324 3rd Qu.: 6.540 3rd Qu.: 6.540   
## Max. :95.00 Max. :18823 Max. :10.740 Max. :58.900   
##   
## z   
## Min. : 0.000   
## 1st Qu.: 2.910   
## Median : 3.530   
## Mean : 3.539   
## 3rd Qu.: 4.040   
## Max. :31.800   
##

ggplot(diamonds,aes(carat)) + geom\_histogram(binwidth=0.01)



#2. Explore the distribution of the “price” variable in the diamonds data. How does the distribution vary by cut?

ggplot(diamonds, aes(x = cut, y = price)) + geom\_boxplot()

