Analytics in Practice

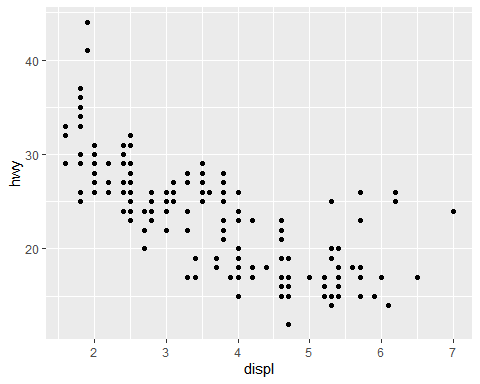
#Libraries used

library(ggplot2)

#Class Practice 1

#1. Plot the following using ggplot():

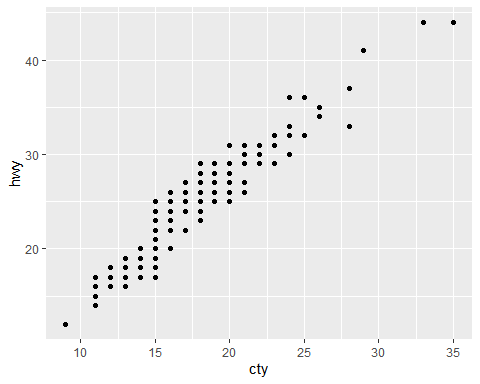
ggplot(mpg,aes(x=displ,y=hwy))+geom\_point()



#2. How would you describe the relationship between cty and hwy (plot graph using ggplot)? Do you have any concerns about drawing conclusions from that plot?

Cty and hwy are positively correlated. The strongest relationship can be seen between 20 -30 hwy after that there are very few points to support the fact at extreme right.

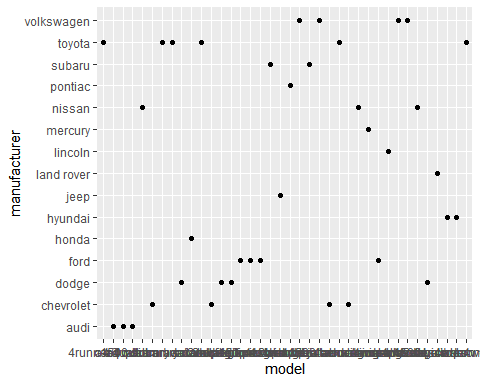
ggplot(mpg,aes(x=cty,y=hwy))+geom\_point()



#3. Plot the following graph and explain the output:

The visualization between model and manufacture is not clear as there are several data points for model and a few for manufacture. Thus, the name of all model is overlapping.

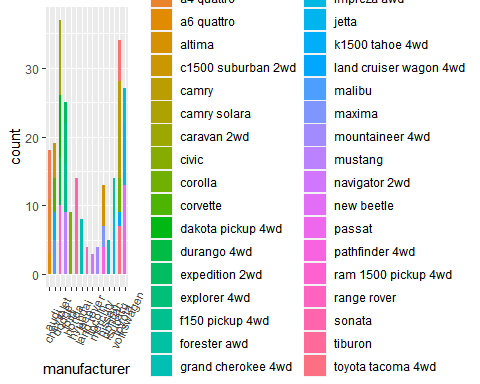
ggplot(mpg, aes(model, manufacturer)) + geom\_point()



#4. Is the above graph useful? How could you modify the data to make it more informative?

No, the above graph is not useful as we can’t make out anything due to poor data visualization. We can modify the data by using geom\_bar with color coding the models.

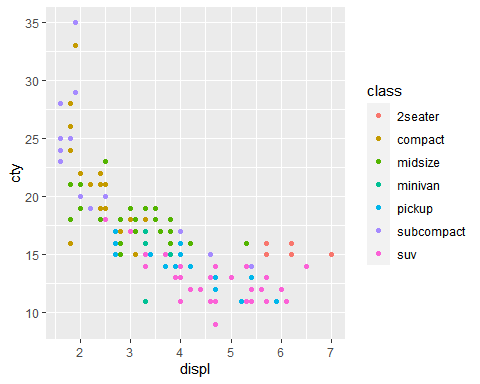
ggplot(mpg, aes(manufacturer))+ geom\_bar(aes(fill=model), width = 0.5) + theme(axis.text.x = element\_text(angle=65, vjust=0.6))



#Class Practice 2

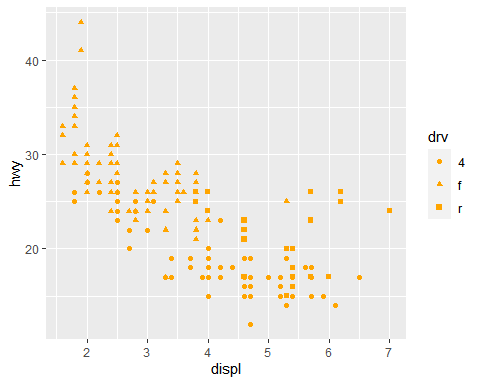
#1. Plot the following using ggplot()

ggplot(mpg, aes(x=displ, y = cty, colour = class)) + geom\_point()



#2. Plot the following (adding one more dimension):

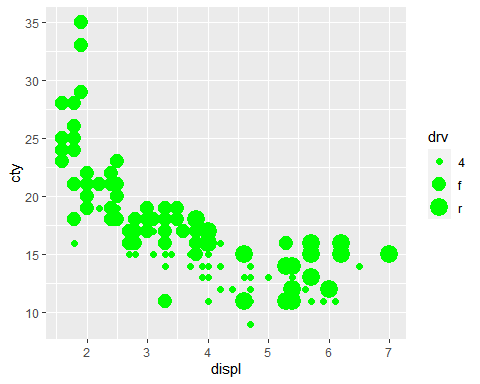
ggplot(mpg, aes(x=displ, y = hwy, shape = drv)) +   
 geom\_point(colour = "orange")



#3. Plot the following graph (adding more dimensions)

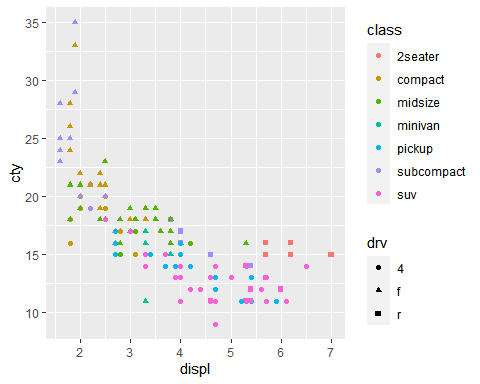
ggplot(mpg, aes(x=displ, y = cty, size = drv)) +   
 geom\_point(colour="green")

## Warning: Using size for a discrete variable is not advised.



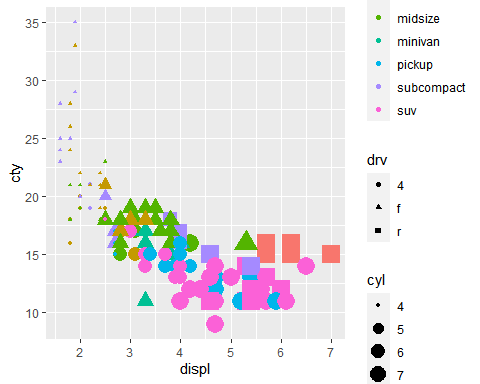
#4. Plot the following graph:

ggplot(mpg, aes(x=displ, y = cty, shape = drv,colour=class)) + geom\_point()



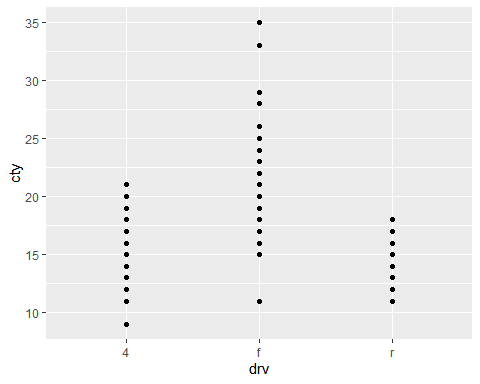
#5. Plot the following graph:

ggplot(mpg, aes(x=displ, y = cty, shape = drv,colour=class,size=cyl)) + geom\_point()



#6. How is drive train related to fuel economy? Plot and show

ggplot(mpg,aes(drv,cty)) + geom\_point()

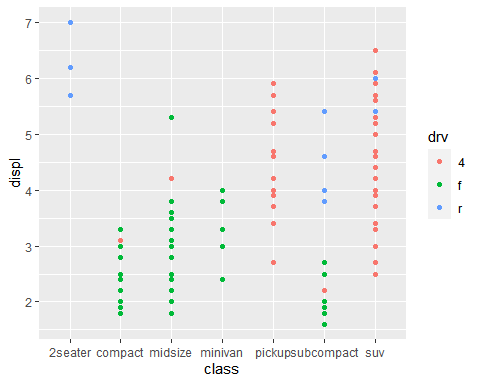


head(mpg)

## # A tibble: 6 x 11  
## manufacturer model displ year cyl trans drv cty hwy fl class   
## <chr> <chr> <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>   
## 1 audi a4 1.8 1999 4 auto(l5) f 18 29 p compa~  
## 2 audi a4 1.8 1999 4 manual(m5) f 21 29 p compa~  
## 3 audi a4 2 2008 4 manual(m6) f 20 31 p compa~  
## 4 audi a4 2 2008 4 auto(av) f 21 30 p compa~  
## 5 audi a4 2.8 1999 6 auto(l5) f 16 26 p compa~  
## 6 audi a4 2.8 1999 6 manual(m5) f 18 26 p compa~

#7. How is drive train related to engine size and class? Plot and show

ggplot(mpg,aes(class,displ,colour=drv)) + geom\_point()



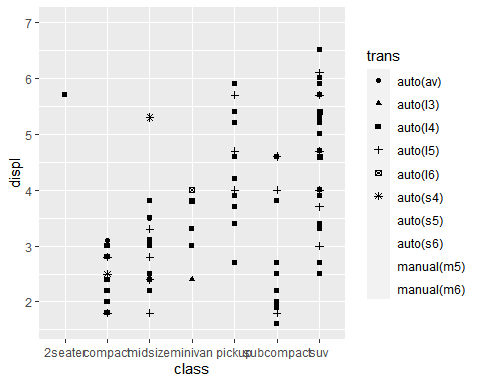
#8. What happens if you map trans variable to shape? Why?

The shape palette can deal with a maximum of 6 discrete values because more than 6 becomes difficult to discriminate; we have 10.

ggplot(mpg,aes(class,displ,shape=trans)) + geom\_point()

## Warning: The shape palette can deal with a maximum of 6 discrete values because  
## more than 6 becomes difficult to discriminate; you have 10. Consider  
## specifying shapes manually if you must have them.

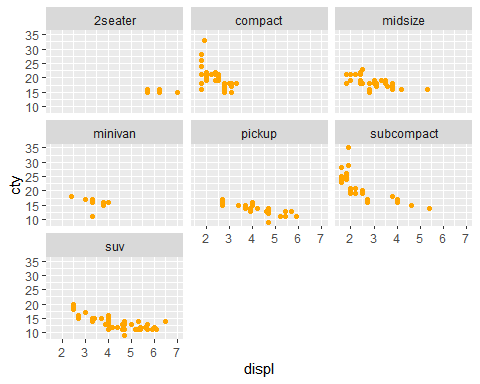
## Warning: Removed 96 rows containing missing values (geom\_point).



#Class Practice 3

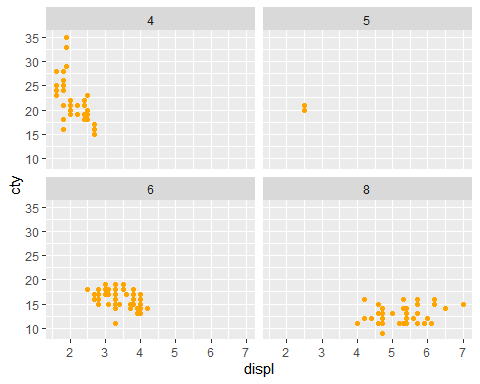
#1. Plot this graph using ggplot()

ggplot(mpg, aes(displ,cty)) +  
geom\_point(colour="orange") + facet\_wrap(~class)



#2. Use facetting to explore the 3-way relationship between fuel economy, engine size, and number of cylinders.

ggplot(mpg, aes(displ,cty)) +  
geom\_point(colour="orange") + facet\_wrap(~cyl)



#3. How does facetting by number of cylinders change your assessment of the relationship between engine size and fuel economy?

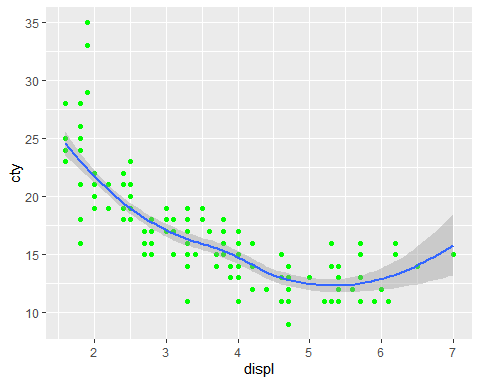
8 cylinder: Has a positive relation 6 cylinder: Has a negative relation 5 cylinder: Has a negative realation 4 cylinder: Not related

#Class Practice 4

#1. Plot the following graph (adding smooth()):

ggplot(mpg, aes(x=displ, y = cty)) + geom\_point(colour="green") + geom\_smooth()

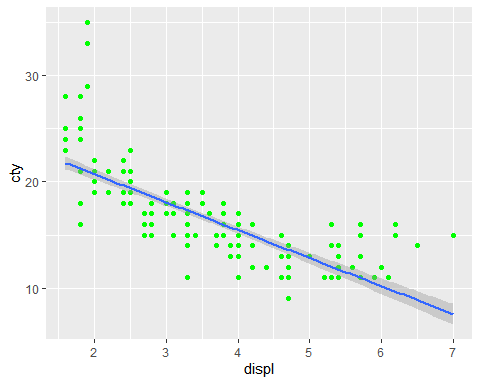
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



#2. Change the method to linear (method=”lm”)

ggplot(mpg, aes(x=displ, y = cty)) + geom\_point(colour="green") +  
 geom\_smooth(method = "lm")

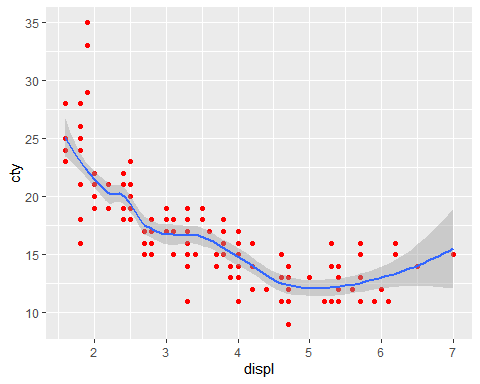
## `geom\_smooth()` using formula 'y ~ x'



#3. Changing the span parameter to control wiggliness of the line:

ggplot(mpg, aes(x=displ, y = cty)) + geom\_point(colour="red") + geom\_smooth(span=0.4)

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



#4. How is this smooth() function useful in interpreting results?

Points alone can be hard to interpret which we have already seen in above practice questions. Smooth() helps represent a line which is easier to understand.

#5. Can you use this to explain this to a senior management executive?

Yes, we can as we can clearly see that fuel economy is negatively related to engine size. As the engine size gets bigger, the fuel economy decreases. We also have a confidence interval for it.

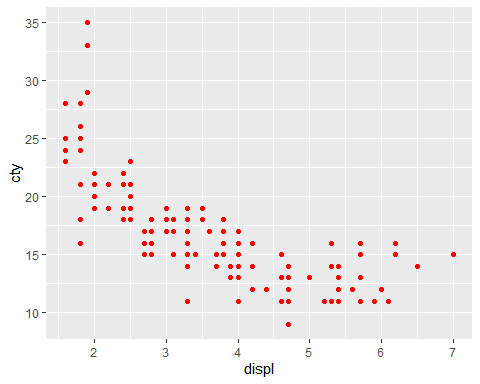
#6. Try changing span parameter, Values range from 0 to 1:

Smoothness increases with the span value.

ggplot(mpg, aes(x=displ, y = cty)) + geom\_point(colour="red") + geom\_smooth(span=0)

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Computation failed in `stat\_smooth()`:  
## span is too small



ggplot(mpg, aes(x=displ, y = cty)) + geom\_point(colour="red") + geom\_smooth(span=1)

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

