Analytics in Practice

#Libraries used

library(ggplot2)  
library(ggforce)  
library(gghighlight)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyverse)

## -- Attaching packages ---------------------------------------------------------------------------- tidyverse 1.3.0 --

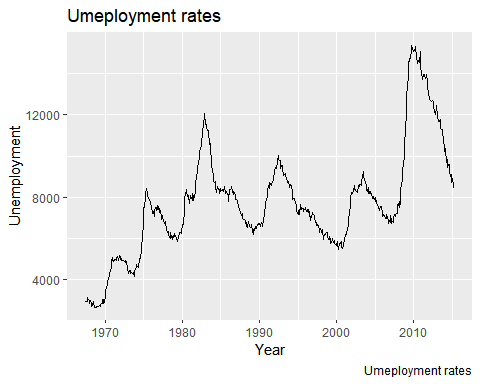
## v tibble 3.0.0 v purrr 0.3.3  
## v tidyr 1.0.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

## -- Conflicts ------------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

#Class Practice 8

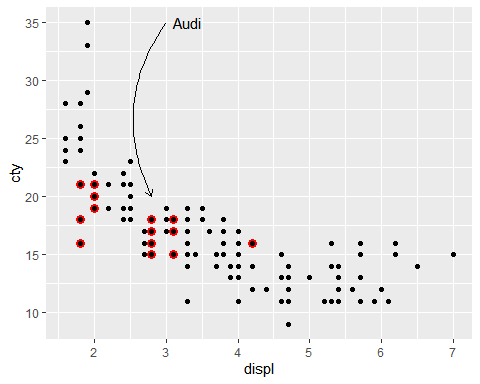
#1. Plot the following graph and label as shown:

ggplot(economics, aes(date, unemploy)) + geom\_line() +   
 labs(title="Umeployment rates", y="Unemployment", x="Year", caption="Umeployment rates")

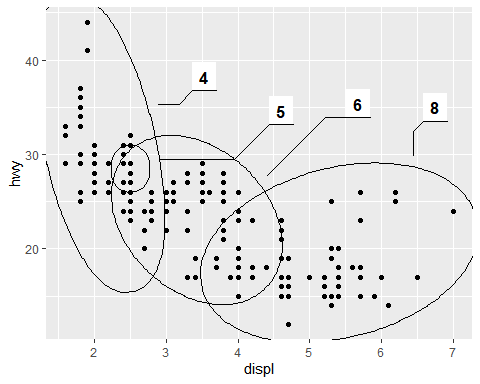


#2. Execute the following program to annotate the graph:

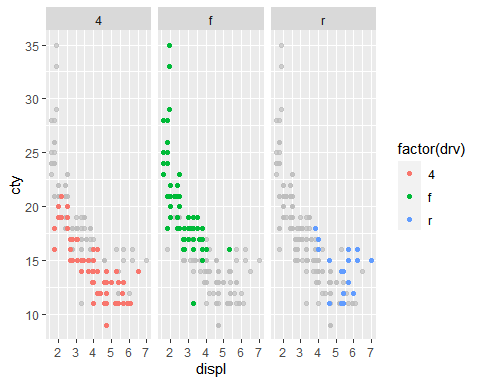
p <- ggplot(mpg, aes(displ, cty)) +  
 geom\_point(  
 data = filter(mpg, manufacturer == "audi"),   
 colour = "red",  
 size = 3  
 ) +  
 geom\_point()   
p +   
 annotate(  
 geom = "curve", x = 3, y = 35, xend = 2.8, yend = 20,   
 curvature = .3, arrow = arrow(length = unit(2, "mm"))  
 ) +  
 annotate(geom = "text", x = 3.1, y = 35, label = "Audi", hjust = "left")



ggplot(mpg, aes(displ, hwy)) +  
 geom\_point() +   
 ggforce::geom\_mark\_ellipse(aes(label = cyl, group = cyl))



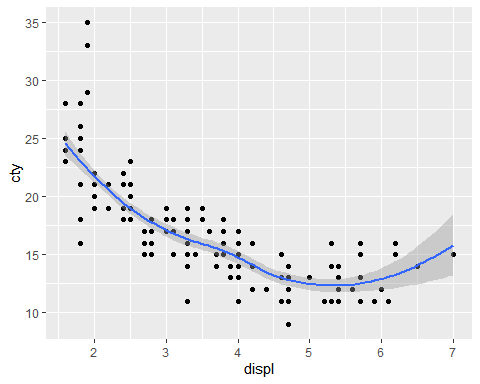
ggplot(mpg, aes(displ, cty,   
 colour = factor(drv))) +  
 geom\_point() +   
 gghighlight::gghighlight() +   
 facet\_wrap(vars(drv))



#3. Try following plots:

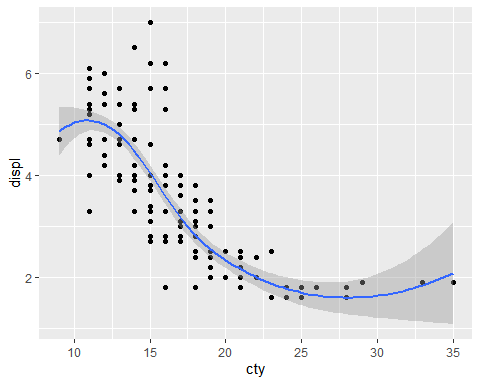
#Plot 1:  
ggplot(mpg, aes( displ,cty)) +   
 geom\_point() +   
 geom\_smooth()

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



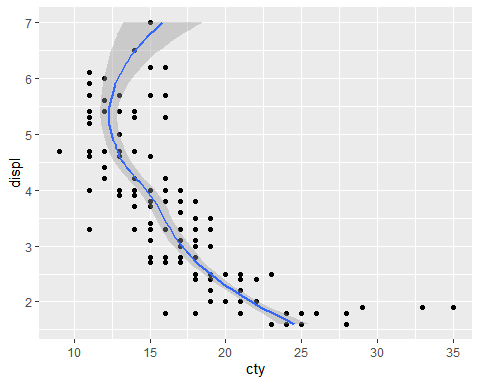
#Plot 2:  
ggplot(mpg, aes(cty,displ)) +   
 geom\_point() +   
 geom\_smooth()

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



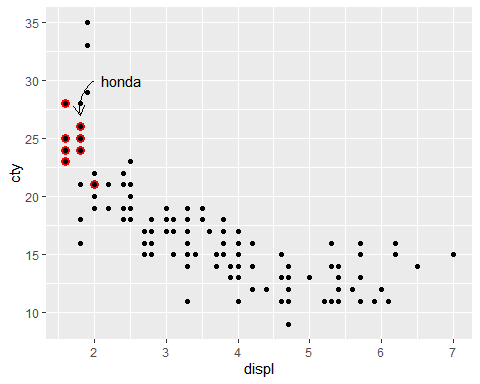
#Plot 3:  
ggplot(mpg, aes(displ, cty)) +   
 geom\_point() +   
 geom\_smooth() +   
 coord\_flip()

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



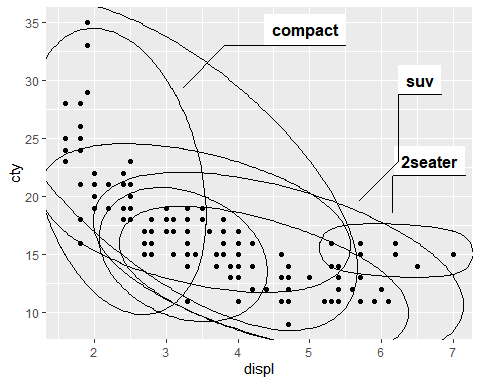
#4. Using “Annotation” discussed earlier, highlight car type “Honda”

b<-mpg  
a <- ggplot(mpg, aes(displ, cty)) +  
 geom\_point(  
 data = filter(mpg, manufacturer == "honda"),   
 colour = "red",  
 size = 3  
 ) +  
 geom\_point()   
a +   
 annotate(  
 geom = "curve", x = 2, y = 30, xend = 1.8, yend = 27,   
 curvature = .3, arrow = arrow(length = unit(3, "mm"))  
 ) +  
 annotate(geom = "text", x = 2.1, y = 30, label = "honda", hjust = "left")



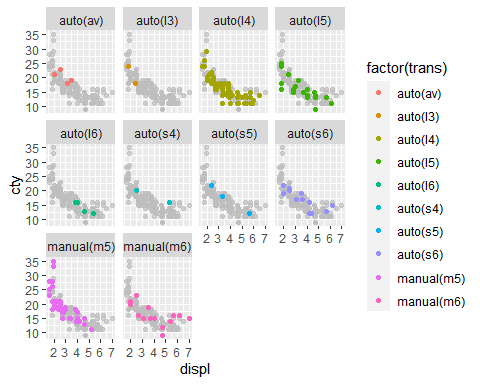
#5. Using ggforce() to highlight the “Class” of cars on “City” mileage based on engine (displ)

ggplot(mpg, aes(displ, cty)) + geom\_point() + ggforce::geom\_mark\_ellipse(aes(label = class, group = class))



#6. Using gghighlight() cluster “transmission” type

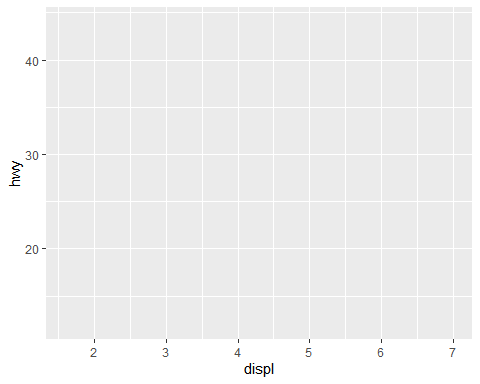
ggplot(mpg, aes(displ, cty, colour =factor(trans))) + geom\_point() + gghighlight::gghighlight() + facet\_wrap(vars(trans))



#Class Practice 9

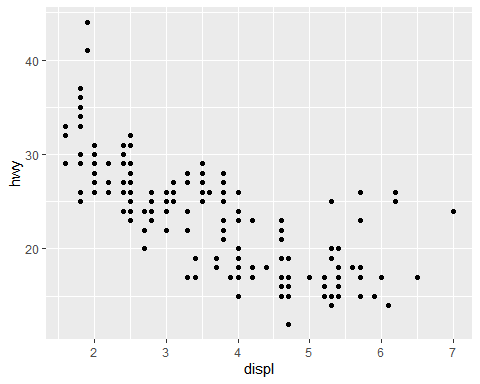
#1. Execute the following commands to build the first layer:

p<-ggplot(mpg, aes(displ, hwy))  
p



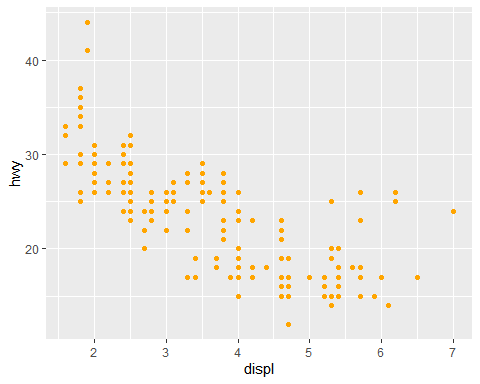
#Add aesthetics to the plot:

p + layer(  
 mapping = NULL,   
 data = NULL,  
 geom = "point",   
 stat = "identity",  
 position = "identity"  
)



#3. Can you change “colour” of points?

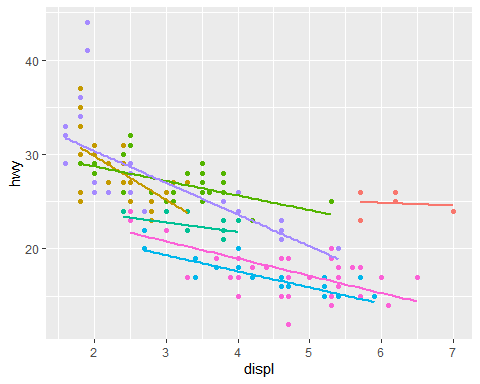
p + layer(  
 mapping = NULL,   
 data = NULL,  
 geom = "point",   
 stat = "identity",  
 position = "identity"  
)+ geom\_point(color = "orange")



#4. Fit the line to the data points by executing following ggplot() program:

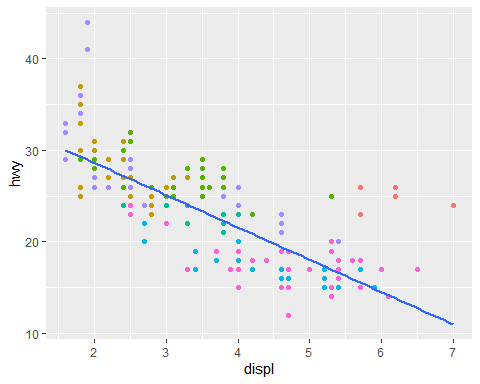
ggplot(mpg, aes(displ, hwy, colour = class)) +   
 geom\_point() +   
 geom\_smooth(method = "lm", se = FALSE) +  
 theme(legend.position = "none")

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



ggplot(mpg, aes(displ, hwy)) +   
 geom\_point(aes(colour = class)) +   
 geom\_smooth(method = "lm", se = FALSE) +   
 theme(legend.position = "none")

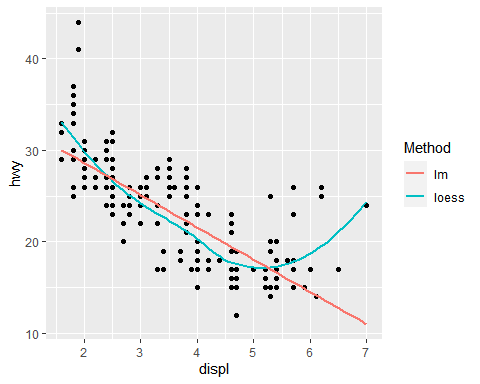
## `geom\_smooth()` using formula 'y ~ x'



#5. Add two different type of line “methods” by executing following program:

ggplot(mpg, aes(displ, hwy)) +   
 geom\_point() +  
 geom\_smooth(aes(colour = "loess"), method = "loess", se = FALSE) +   
 geom\_smooth(aes(colour = "lm"), method = "lm", se = FALSE) +  
 labs(colour = "Method")

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



#Class Practice 10

#Problem 31.1 Jenny averaged 70 on her quizzes during the first part of the quarter and 80 on her quizzes during the second part of the quarter. When she found out that her final average for the quarter was not 75, she went to argue with her teacher. Give a possible explanation for Jenny’s misunderstanding.

Ans. To find the average Jenny needs to know the sum of grades for all the quizzes to get the average by dividing it by total number.

#Problem 31.2 Suppose the following circle graphs are used to illustrate the fact that the number of elementary teaching majors at teachers’ colleges has doubled between 1993 and 2003, while the percent of male elementary teaching majors has stayed the same. What is misleading about the way the graphs are constructed?

Ans. The chart is misleading as the area is depicted larger for 2003 but the percentage remains same. A number would have been a better way of representation.

#Problem 31.3 What is wrong with the following line graph?

Ans. X and Y axis are not labelled.

#Problem 31.4 Doug’s Dog Food Company wanted to impress the public with the magnitude of the company’s growth. Sales of Doug’s Dog Food had doubled from 2002 to 2003, so the company displayed the following graph, in which the radius of the base and the height of the 2003 can are double those of the 2002 can. What does the graph really show with respect to the growth of the company?

Ans.This kind of visual representation is misleading as one can’t really quantify it.

#Problem 31.5 What’s wrong with the following graph?

Ans.Y-axis not labelled.

#Problem 31.6 Refer to the following pictograph: Ms McNulty claims that on the basis of this information, we can conclude that men are worse drivers than women. Discuss whether you can reach that conclusion from the pictograph or you need more information. If more information is needed, what would you like to know?

Ans. This graph is misleading as to come to the conclusion we need to know a lot of things, the core thing would be the total number of male and female drivers for whome this was recorded. Also, there is no numerical representation.

#Problem 31.7 Larry and Marc took the same courses last quarter. Each bet that he would receive the better grades.

Ans. Larry is correct since when we calculate their GPA, he has higher.

#Problem 31.8 Oil prices went up 20% one year and 30% the next. Is it true that over the two years, prices went up 50%?

Ans.False

#Problem 31.9 True or false? My rent went down 10% last year and then rose 20% this year. Over the two years my rent went up by 10%.

Ans. False

#Problem 31.10 Which graph could be used to indicate a greater decrease in the price of gasoline? Explain.

Ans.Graph A since the bars are relative and show the decrease much more intutively than the size of the car in Graph B. It takes away attention of the user as well as dosen’t convery the message clearly.