MASTER R WEEK 2 SUMMARY

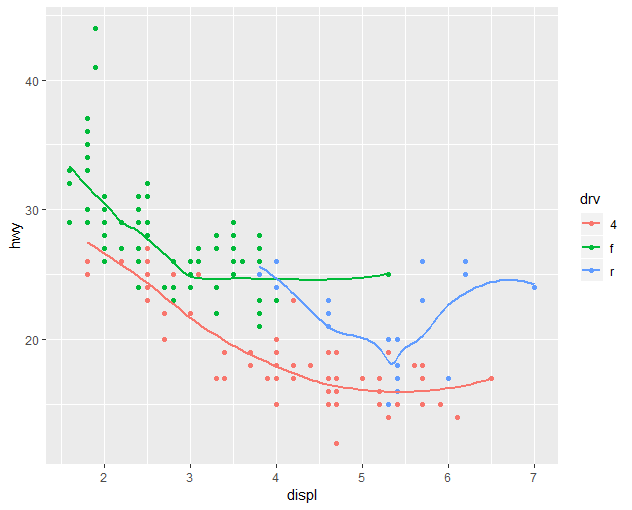
###3.6.1 Exercises

####question2

ggplot(data = mpg, mapping = aes(x = displ, y = hwy, color = drv)) +

geom\_point() +

geom\_smooth(se = FALSE)

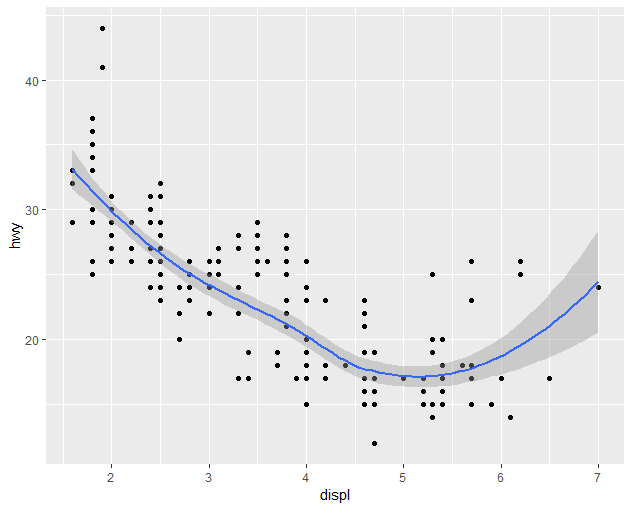


####question5

ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +

geom\_point() +

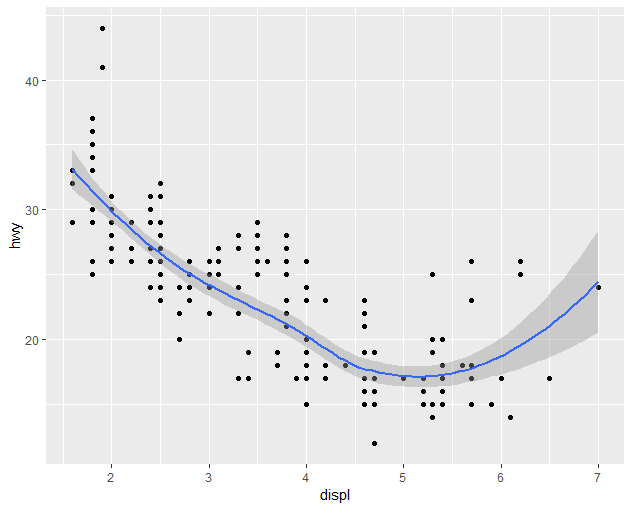
geom\_smooth()



ggplot() +

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

geom\_smooth(data = mpg, mapping = aes(x = displ, y = hwy))###no difference in this two graphs as the layers are the same



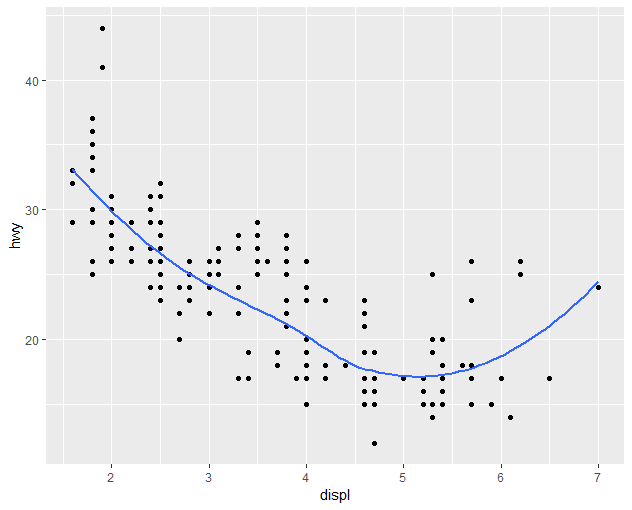
####question6

##graph1

ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +

geom\_point() +

geom\_smooth(se = FALSE)

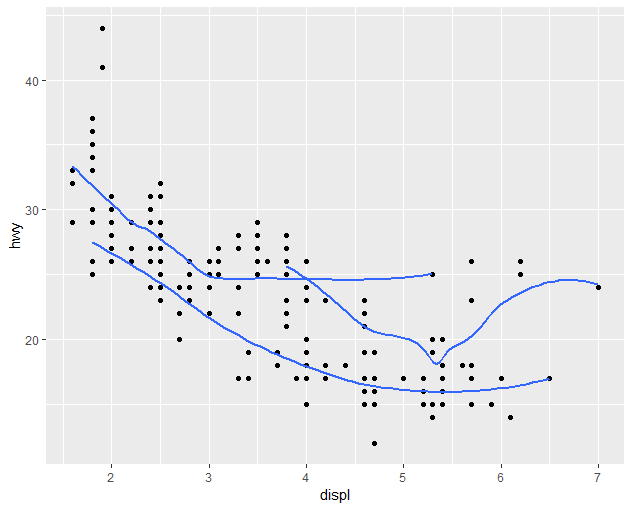


##graph2

ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +

geom\_point() +

geom\_smooth(mapping = aes(x = displ, y = hwy, group = drv), se = FALSE)

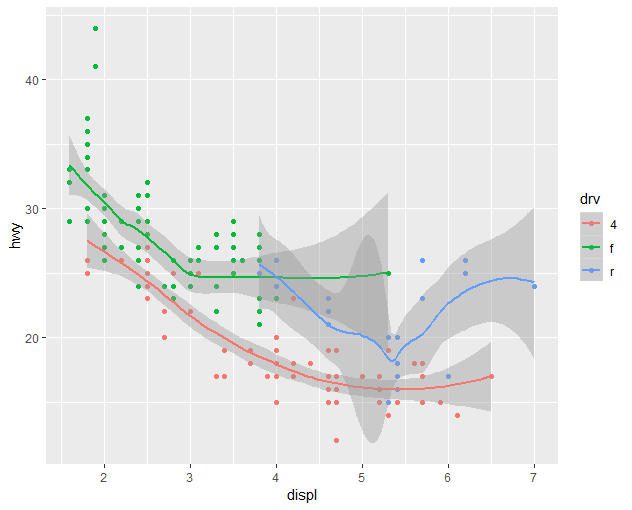


##graph3

ggplot(data = mpg, mapping = aes(x = displ, y = hwy, color = drv)) +

geom\_point() +

geom\_smooth()

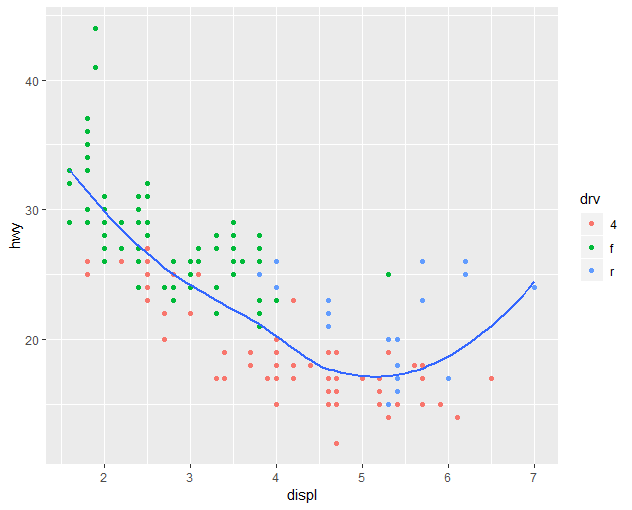


##graph4

ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +

geom\_point(mapping = aes(color = drv)) +

geom\_smooth(se = FALSE)

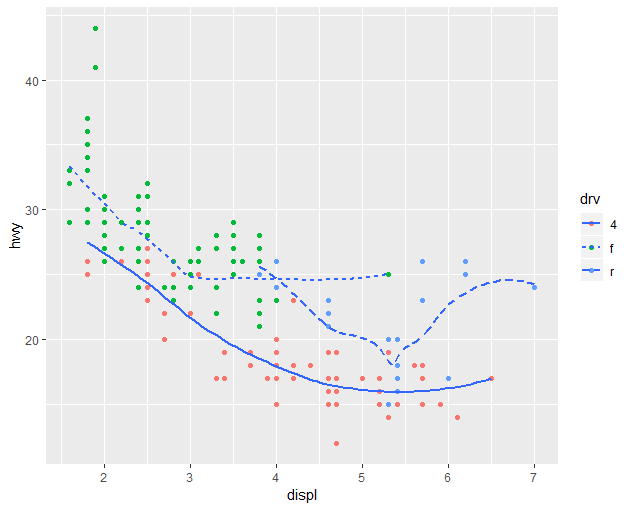


##graph5

ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +

geom\_point(mapping = aes(color = drv)) +

geom\_smooth(se = FALSE, mapping = aes(linetype = drv))

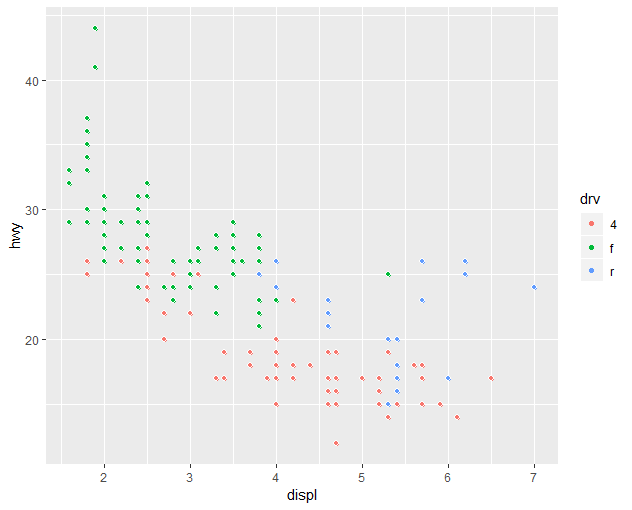


##graph6

ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +

geom\_point(mapping = aes(color = drv)) +

geom\_point(shape = 21, color = "white", stroke = 1)



###COMBINING ALL GRAPHS TOGETHER

library(cowplot)

library(gridExtra)

p1 <- ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +

geom\_point() +

geom\_smooth(se = FALSE)

p2 <- ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +

geom\_point() +

geom\_smooth(mapping = aes(group = drv), se = FALSE)

p3 <- ggplot(data = mpg, mapping = aes(x = displ, y = hwy, color=drv)) +

geom\_point() +

geom\_smooth(se = FALSE)

p4 <- ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +

geom\_point(mapping = aes(color=drv)) +

geom\_smooth(se = FALSE)

p5 <- ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +

geom\_point(mapping = aes(color=drv)) +

geom\_smooth(se = FALSE, mapping = aes(linetype = drv))

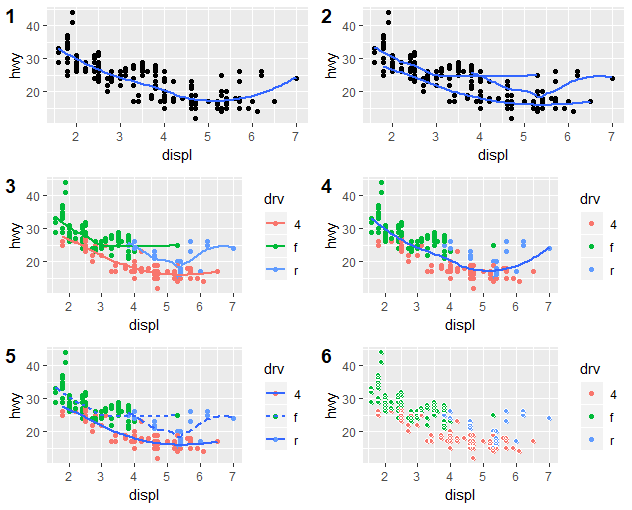
p6 <- ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +

geom\_point(mapping = aes(color=drv)) +

geom\_point(shape = 21, color = "white", stroke = 1)

theme\_set(theme\_gray())

plot\_grid(p1, p2, p3, p4, p5, p6, labels=c("1","2","3", "4","5","6"), ncol=2, nrow = 3)



#3.7 STATISTICAL TRANSFORMATIONS

####3.7.1 Exercises

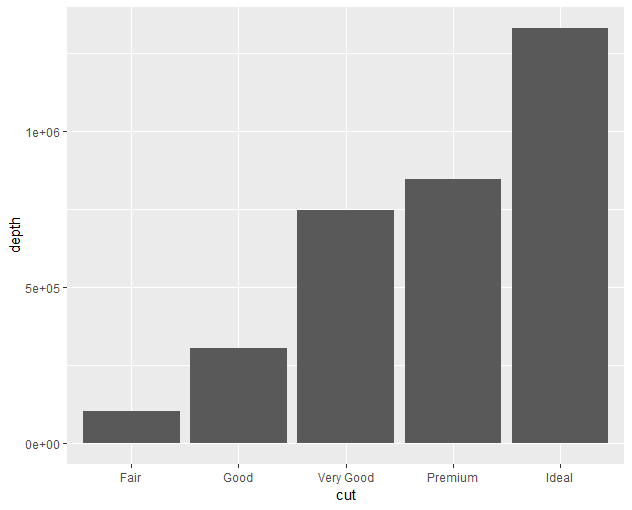
####Question1:default geom for stat\_summary==geom\_pointrange

####Question2:geom\_col==it makes the heights of the bars representative of the values in the data, thus leaving the data as is

###geom\_bar bins together the values and makes the heights of the bars a proportion of the total cases

ggplot(data = diamonds) +

geom\_col(mapping = aes(x = cut, y = depth))

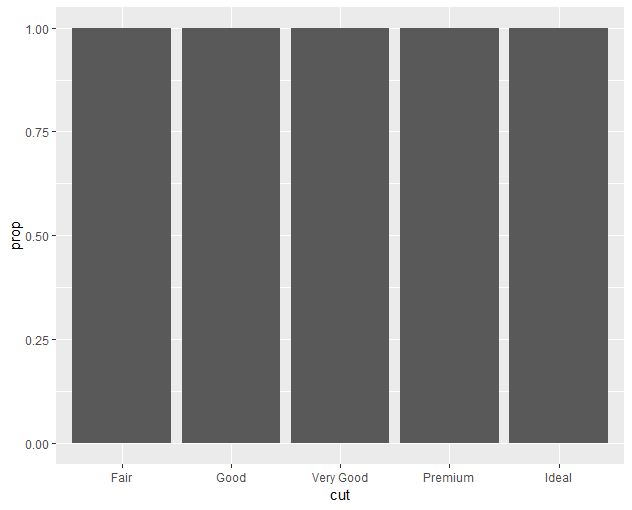


####Qusetion4:stat\_smooth computes y:predicted value, ymin:lower C.I, ymax: upper C.I, se:standard error

####Question5: we set the group = 1 because if not, geom\_bar assumes that all groups have equal x values

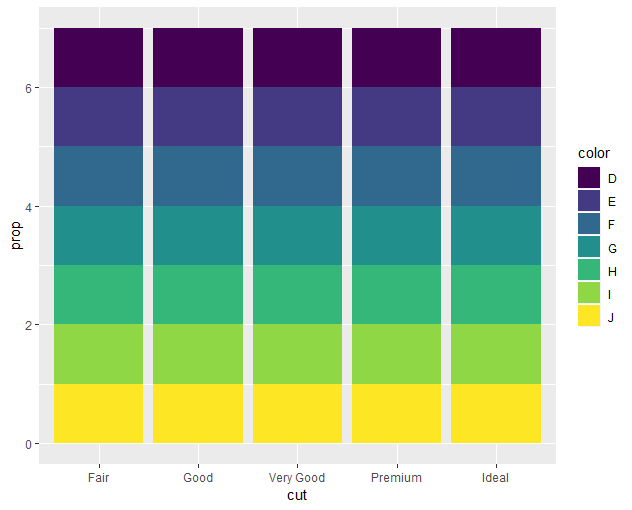
ggplot(data = diamonds) +

geom\_bar(mapping = aes(x = cut, y = ..prop..))



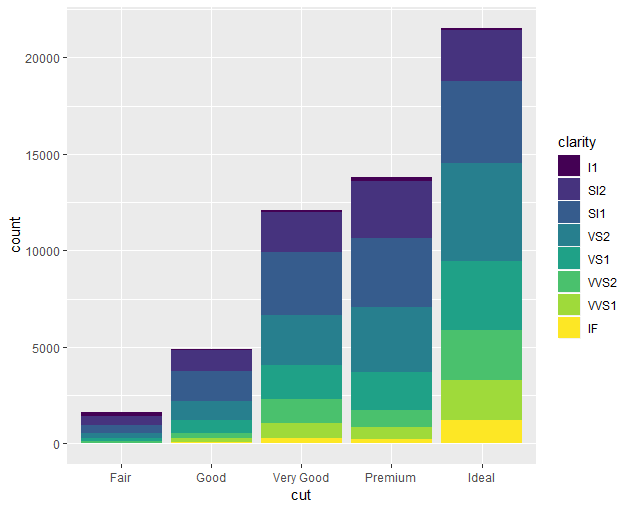
ggplot(data = diamonds) +

geom\_bar(mapping = aes(x = cut, fill = color, y = ..prop..))



ggplot(data = diamonds) +

geom\_bar(mapping = aes(x = cut, fill = clarity))



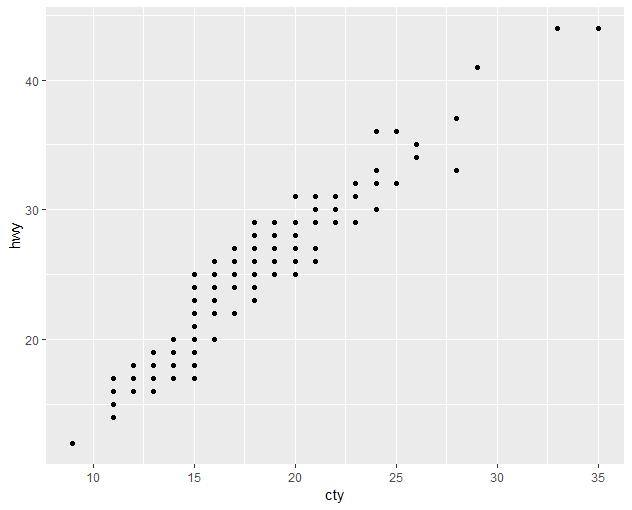
#3.8 POSITION ADJUSTMENTS

##3.8.1 Exercises

###Question 1 What is the problem with this plot? How could you improve it?

ggplot(data = mpg, mapping = aes(x = cty, y = hwy)) +

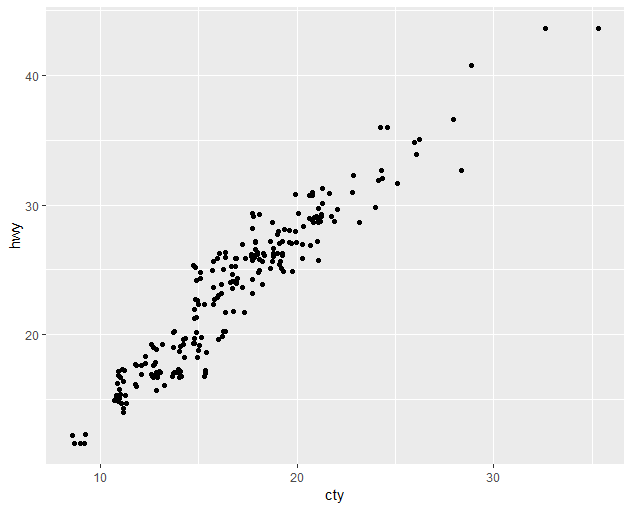
geom\_point()



####Ans: Many of the points are not shown in the plot, as they overlap. This can be improved using geom\_jitter

ggplot(data = mpg, mapping = aes(x = cty, y = hwy)) +

geom\_jitter()



####Question 2 What parameters to geom\_jitter() control the amount of jittering?

###Ans: The amount of jitter is controlled by increasing the distance between each of the points

####Qusetion 3 Compare and contrast geom\_jitter() with geom\_count().

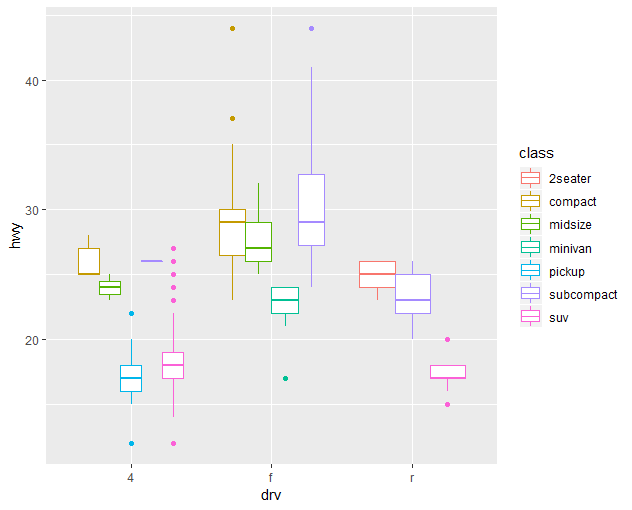
###Ans:geom\_count increases the size of overlapped points while geom\_jitter allows one to see all points while keeping them the same size

####Question 4 What’s the default position adjustment for geom\_boxplot()? Create a visualisation of the mpg dataset that demonstrates it.

##Ans: the default position is dodge

ggplot(data = mpg, mapping = aes(x = drv, y = hwy, color = class)) +

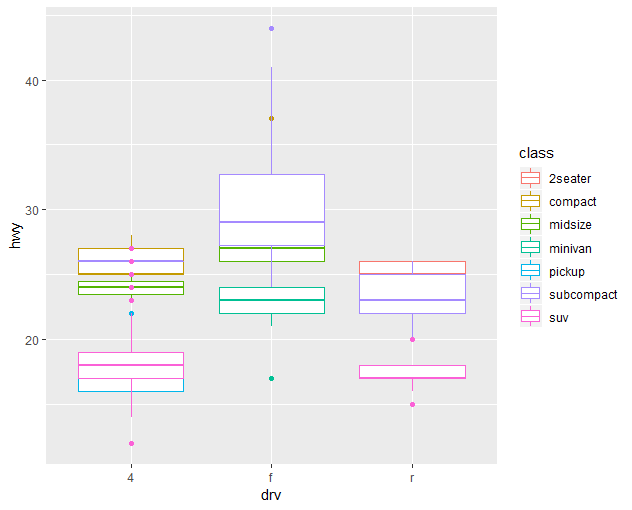
geom\_boxplot(position="dodge")



###we can also use identity

ggplot(data = mpg, mapping = aes(x = drv, y = hwy, color = class)) +

geom\_boxplot(position="identity")



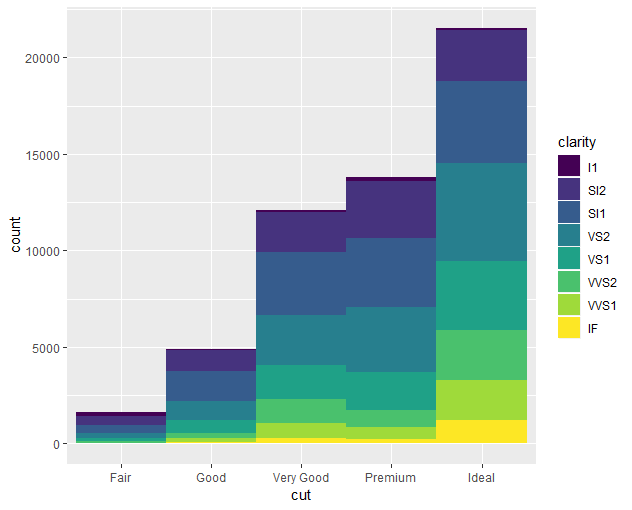
#3.9 COORDINATE SYSTEMS

####3.9.1 Exercises

###Question 1

ggplot(data = diamonds) +

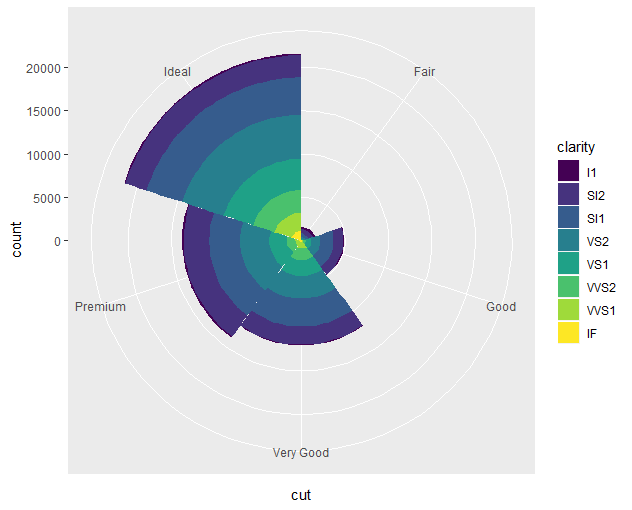
geom\_bar(mapping = aes(x = cut, fill = clarity), width = 1)



ggplot(data = diamonds) +

geom\_bar(mapping = aes(x = cut, fill = clarity), width = 1) +

coord\_polar()



###Question 2 What does labs() do?

##Ans: It labels the coordinates by modifying axis, legends and plit tables

###Question 3What’s the difference between coord\_quickmap() and coord\_map()?

####Question 4 What does the plot below tell you about the relationship between city and highway mpg? Why is coord\_fixed() important? What does geom\_abline() do?

ggplot(data = mpg, mapping = aes(x = cty, y = hwy)) +

geom\_point() +

geom\_abline() +

coord\_fixed()

