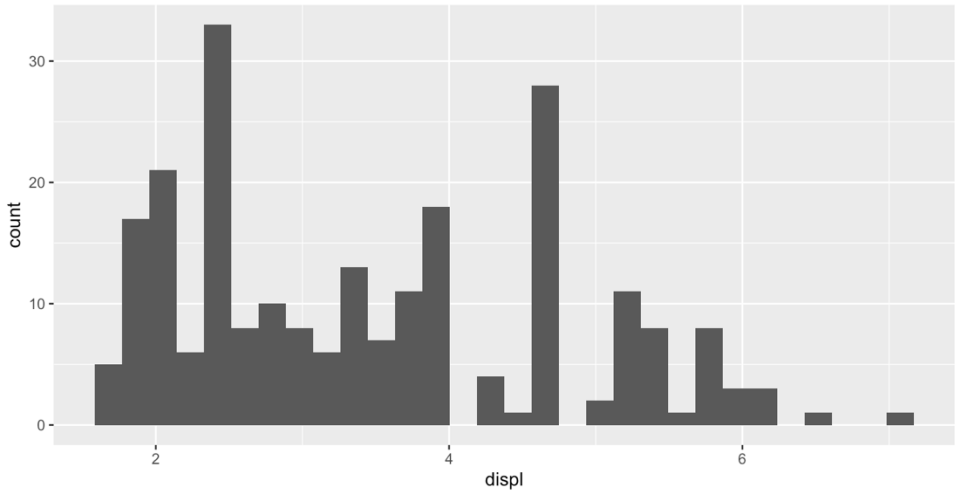
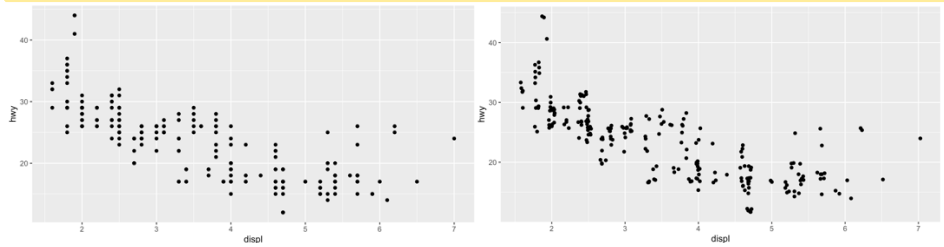
HW 1

1. Which of these variables in mpg are categorical?
   a. displ
   b. cty
   c. class
2. How many rows (cars) in mpg have year 1999?
   a. 116
   b. 117
   c. 118
   d. 119
3. Which code produced the figure below?
   a. ggplot(data=mpg)+geom_histogram(aes(x=displ))
   b. ggplot(data=mpg)+geom_bar(aes(x=displ))
   c. ggplot(data=mpg)+geom_density(aes(x=displ))
   d. ggplot(data=mpg)+geom_boxplot(aes(x=displ))

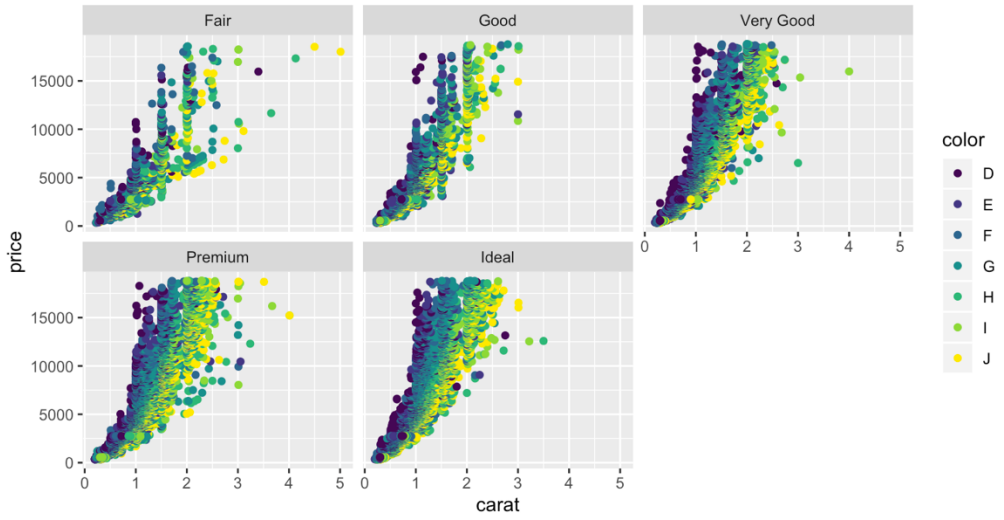


4. Each dot in the plot below corresponds to multiple observations that overlap. How can we visualize all the points as in the second plot?
   a. ggplot(data = mpg) + geom_point(mapping = aes(x = displ, y = hwy),position = "avoid")
   b. ggplot(data = mpg) + geom_point(mapping = aes(x = displ, y = hwy),position = "scatter")
   c. ggplot(data = mpg) + geom_point(mapping = aes(x = displ, y = hwy),position = "jitter")



5. How do you produce the following plot?
   a. diamonds %>% ggplot()+geom_point(aes(x=carat,y=price,color=color))+facet_wrap(~cut)
   b. ggplot(data=diamonds)+geom_point(aes(x=carat,y=price),color=color)+facet_wrap(~cut,nrow=2)

c. diamonds %>% ggplot()+geom_scatter(aes(x=carat,y=price,color=color))+facet_wrap(~cut)



6. Using the flights data from the nycflights13 package, find the mean departure delay (dep_delay) for flights that departed in month 1 (and where the dep_delay was not na). Hint: use filter and summarise.
   a. 7.0
   b. 8.0
   c. 9.0
   d. 10.0
7. You can count the number of flights per year, month and day by first grouping by those variables and then using n() in summarise to count the number of rows for each grouping:

   flights %>% group_by(year, month, day) %>% summarise(flights = n())

   When you do this how many flights were there on 2/4/13?
   a. 930
   b. 931
   c. 932
   d. 933

8. Which plane (`tailnum`) has the worst on-time record (arr_delay)? -> N384HA

9. What hour of day should you fly if you want to avoid delays (dep_delay) as much as possible?

10. Delays are typically temporally correlated: even once the problem that caused the initial delay has been resolved, later flights are delayed to allow earlier flights to leave. Using `lag()`, how is the delay of a flight related to the delay of the immediately preceding flight. Compute the correlation coef of these two variables using cor().