Lab 1

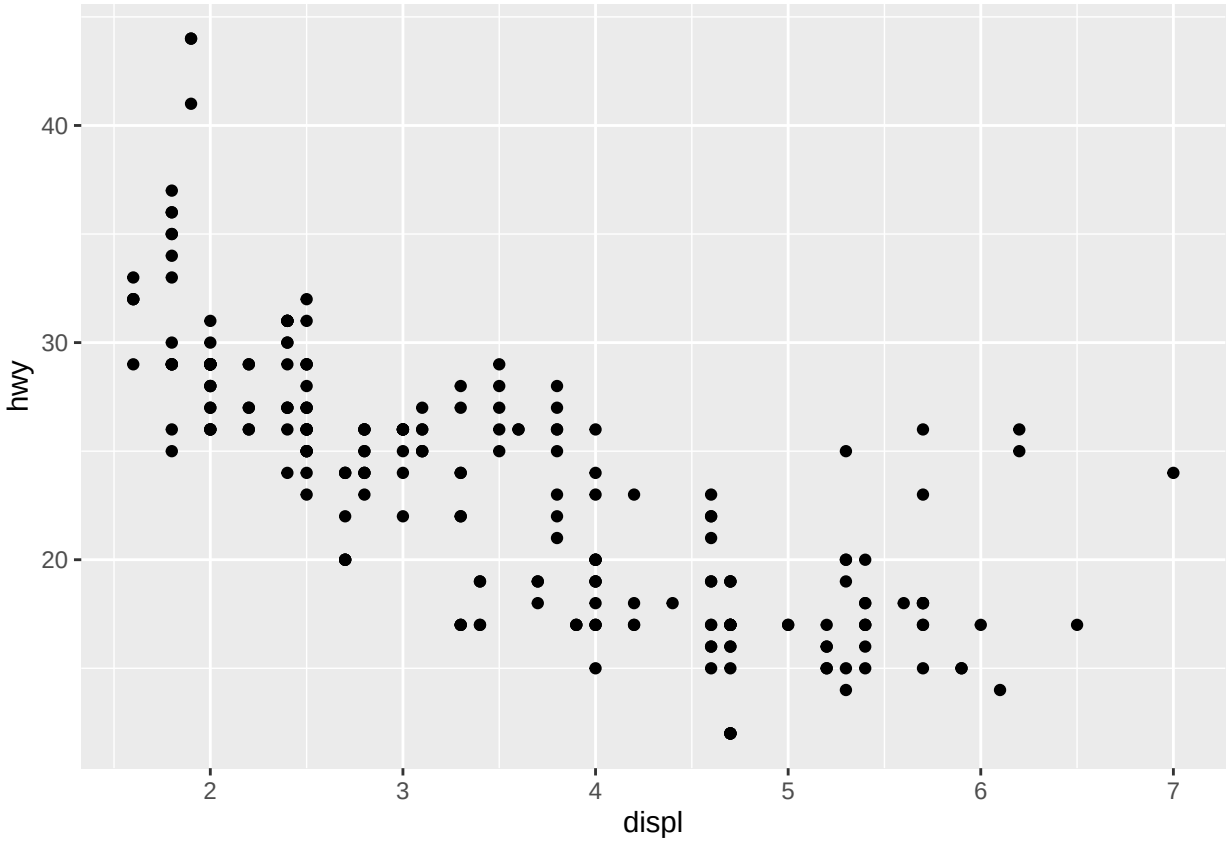
Bushong Boys (Group 17)

1/20/2020

Exercise 1

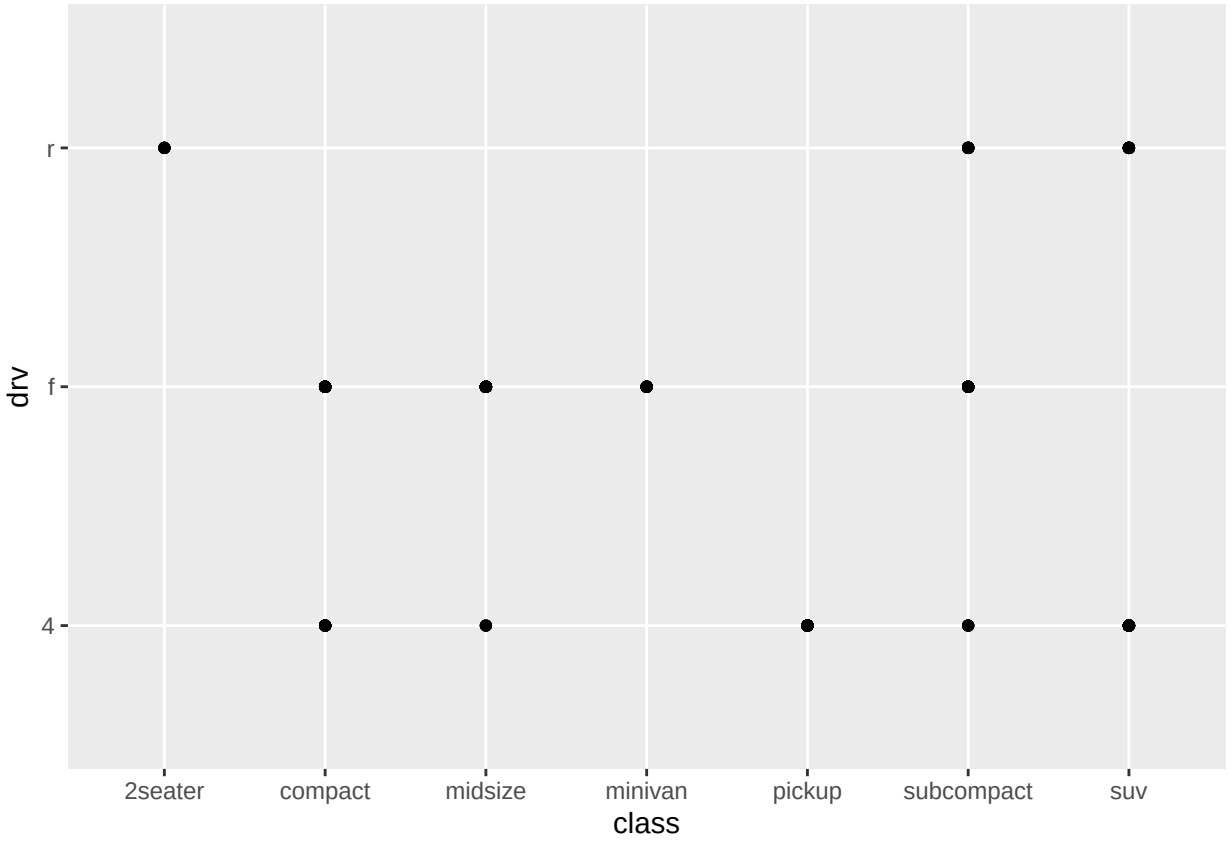
Comparing disp to hwy:

In the scatterplot below, 38 models of cars engine sizes are compared to fuel efficiency on the highway. As you can see, there is an inverse relationship between these two variables. This makes logical sense since as the engine size of a car goes up, so does it actual size. Generally the larger a car is, the worse its fuel efficiency.



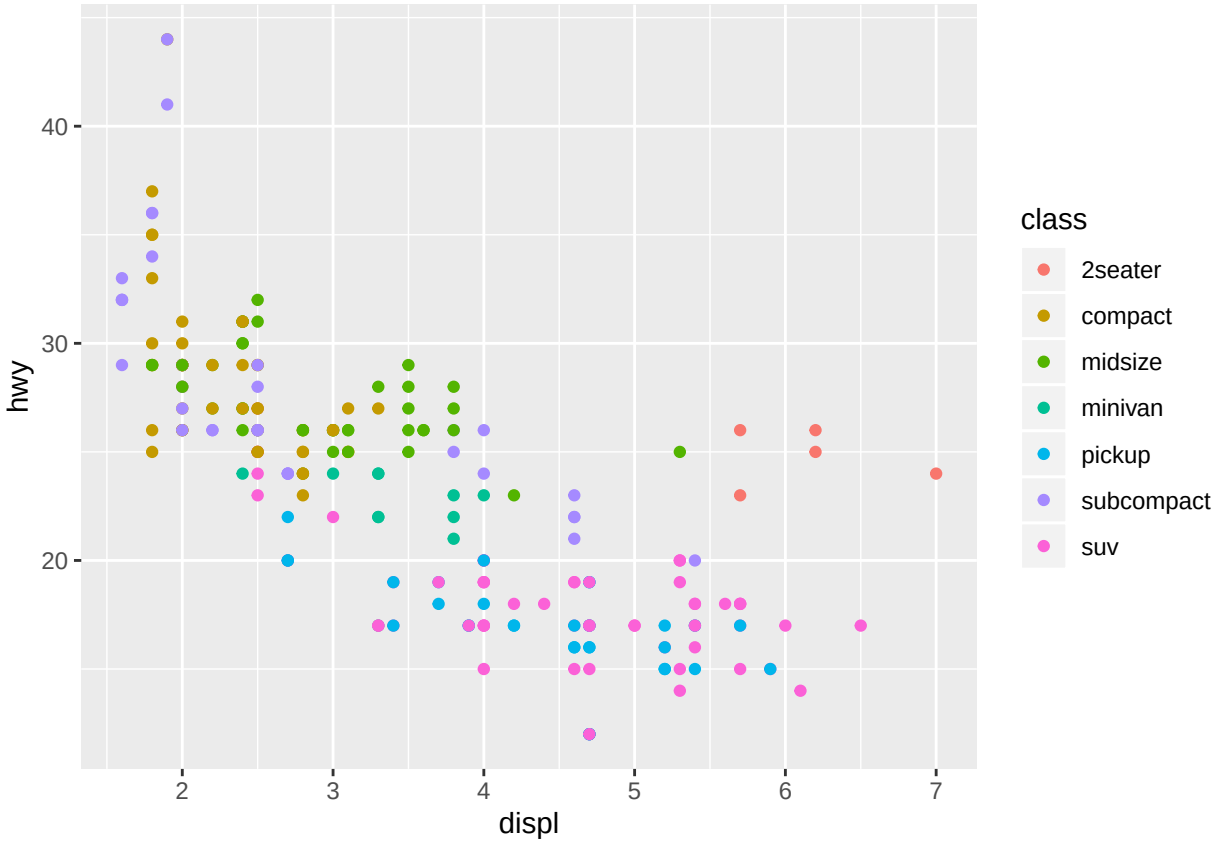
Comparing class to drv:

The plot below compares a car's class to its type of drive. As seen by the graph, there is no inherent relationship between these two things since they are features of a car. There are plenty of SUVs with both front, rear, and four-wheel drive. Therefore, this a very un-useful visualization.



Back to disp and hwy:

You may notice from our previous scatterplot comparing engine size and fuel efficiency that there is a group of cars with large engine sizes but have a higher fuel efficiency than a linear model might have predicted. One of our hypotheses was that the type of car was influencing this characteristic. Once adding the class of car as another level in the form of color, it was revealed that all of the outlier points were two seaters cars. After putting thought into this fact, it makes sense since they are smaller cars, but have larger engines for sporting capabilities.



Exercise 2

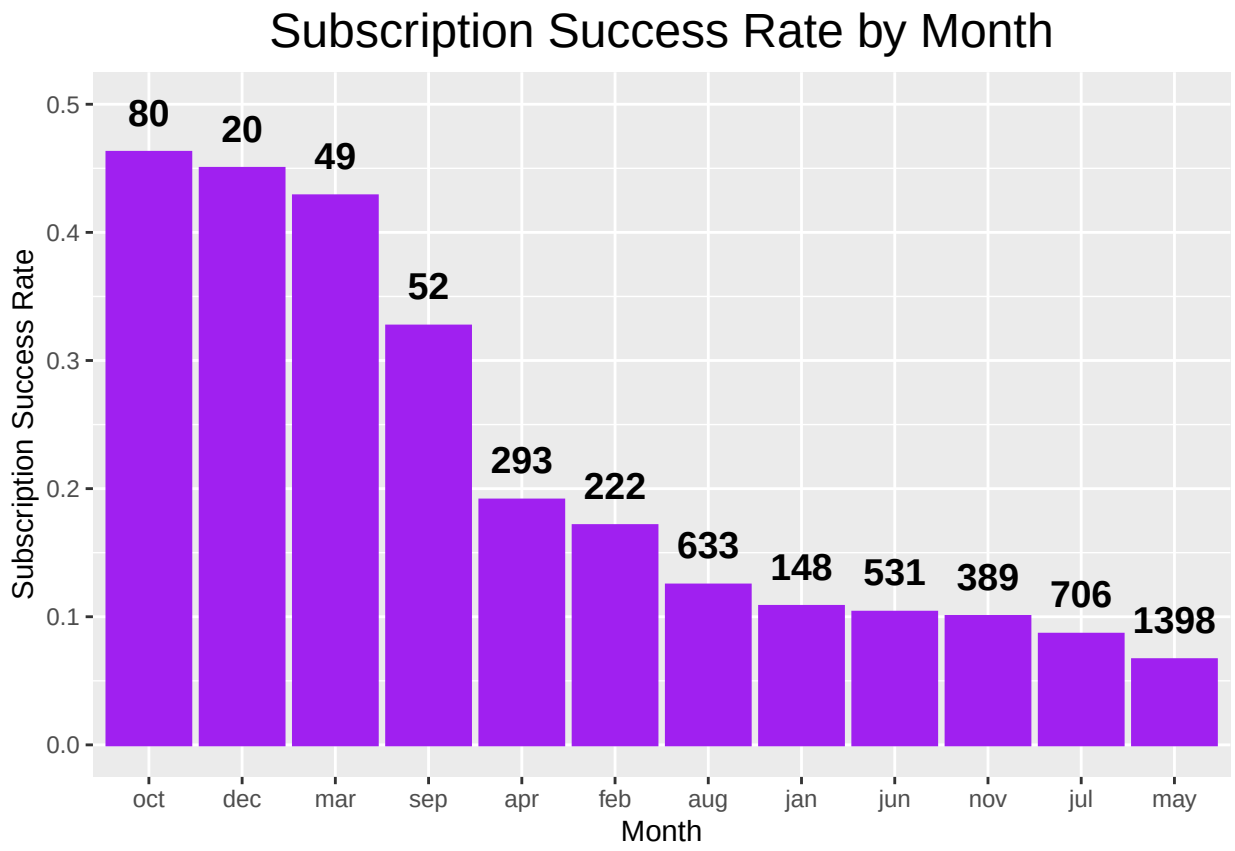
Recommendation 1 - Timing the Campaign

Description of the Graph:

Below, you will see a bar chart illustrating the month clients were called for an advertising campaign and the success rate of subscription from it. In addition, the black number above each bar is the number of clients reached during that month. For example, in October, 80 people were called to advertise the banks long-term savings account.

Advice:

Based on the results of this visualization, there are a few things we should observe. It would be naive to instantly conclude that the marketing campaign should take place during October, December, March, and September because this is when the success rate is the highest. The reasoning is these months have relatively small sample sizes and the actual mean might deviate far from the sample success rate. With that in mind, we recommend looking at a period of 3-4 months and there are two options based tolerance of risk. The first would be to choose the months October - December. This choice has the highest subscription success rate; however, three of these months have sample sizes under 80 which could yield in results that do not replicate this data. The more conservative approach would be to run the campaign February-April. This option still has high subscription success rates, but the greater sample sizes in February and April ensure a greater chance of success rates similar to these numbers.


Subscription Success Rate by Month
Subscription Success Rate
0.5
0.4
0.3
0.2
0.1
0.0
80
20
49
52
293
222
633
148
531
389
706
1398
oct
dec
mar
sep
apr
feb
aug
jan
jun
nov
jul
may
Month