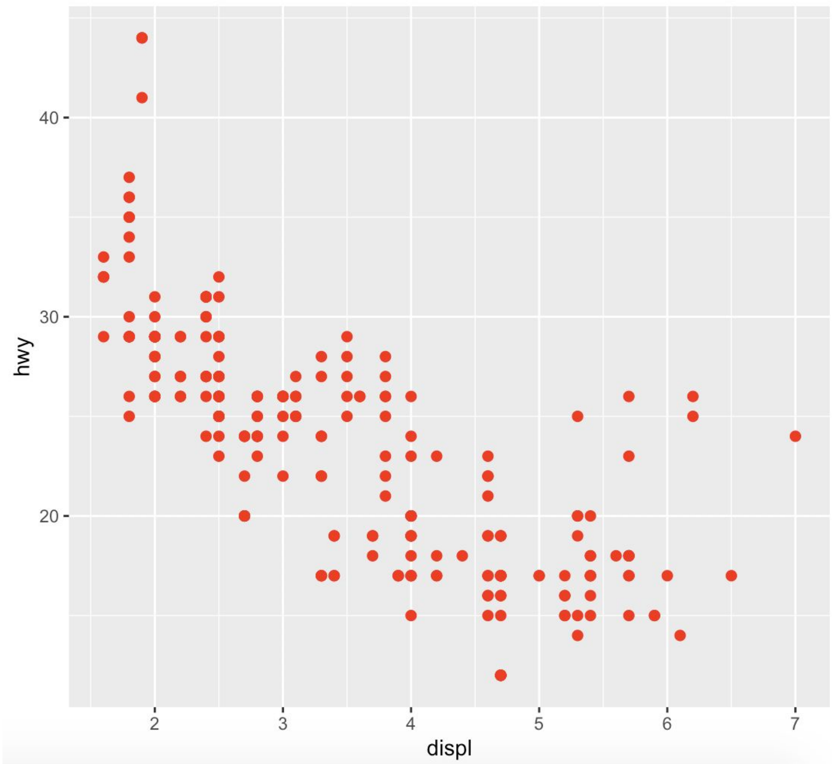
Lab 1

Alexandra Holland, Maura Glynn, Tyler Kramer, Jonathan Zaremba

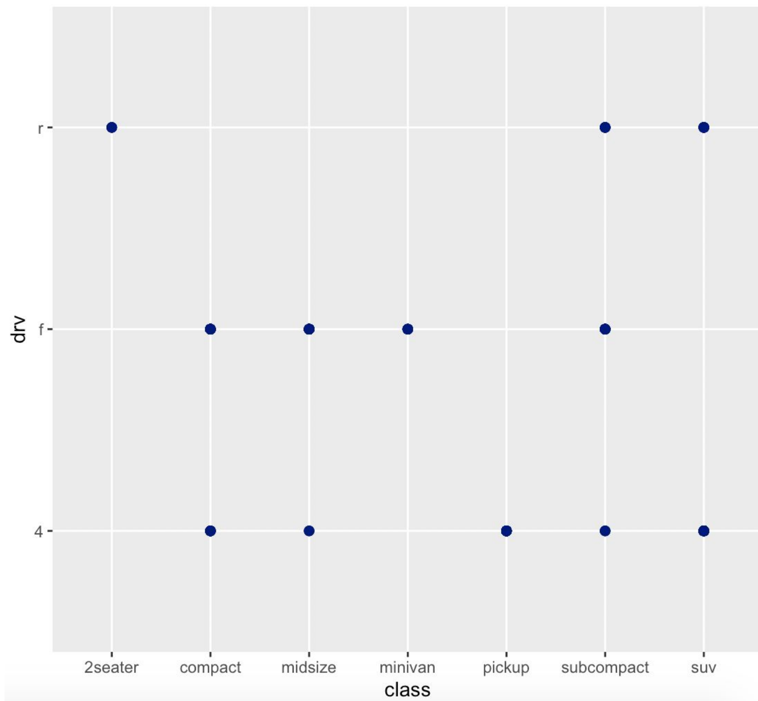
Exercise 1)

a. Engine Size (*displ*) vs. Highway Fuel Efficiency (*hwy*)



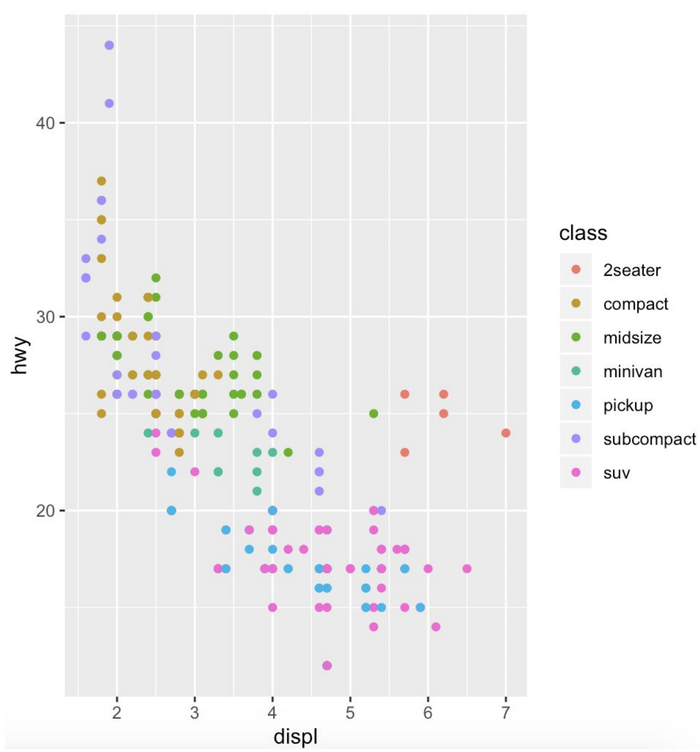
This graph does display the relationship between engine size and highway fuel efficiency that would be expected as larger engines almost always utilize more fuel and do not have high fuel efficiency. This is shown through the negative trend and clustering of the data points, implying a negative correlation between the two variables.

Class vs. Drive (*drv*)

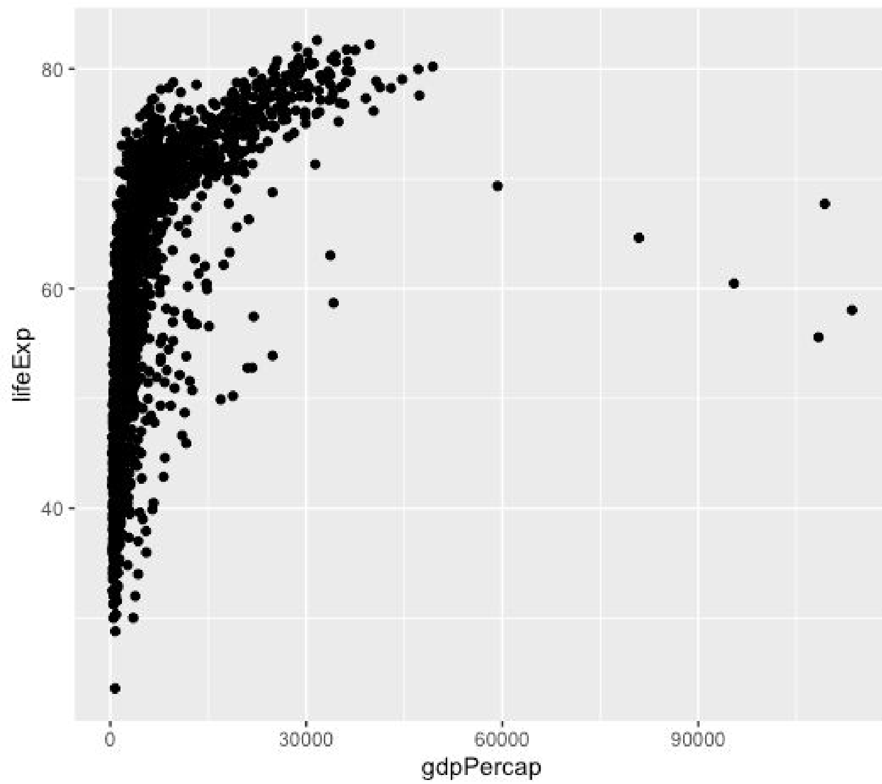


A scatter plot of class versus type of drive is essentially useless because both of these variables of the data set are qualitative, meaning they account for non-numeric values and values that cannot be measured. Graphically, comparing qualitative data via scatter plot does not give useful insight.

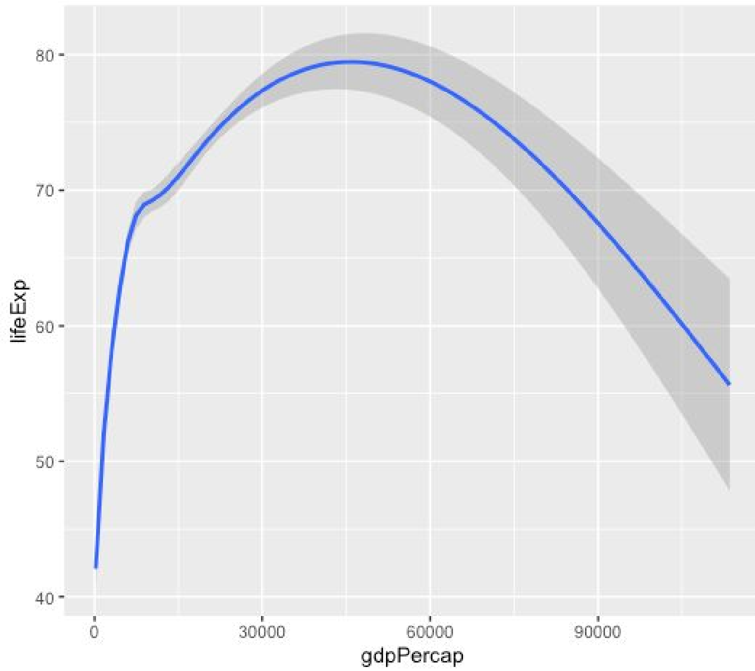
b. Engine Size (*displ*) vs. Highway Fuel Efficiency (*hwy*) - scaled by *class*



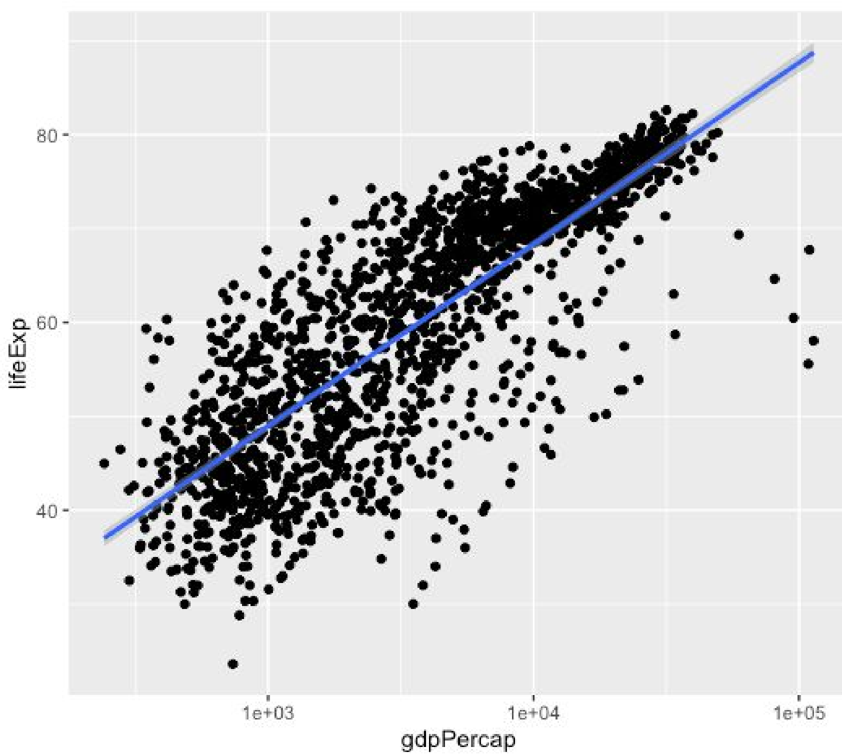
The outlying group of vehicles that seems to have a higher than expected highway mileage for having a large engine size appear to be 2seater cars, most likely sports cars which would have large engine but are designed for reaching high speeds so it makes sense that these points would not follow the trend as closely.



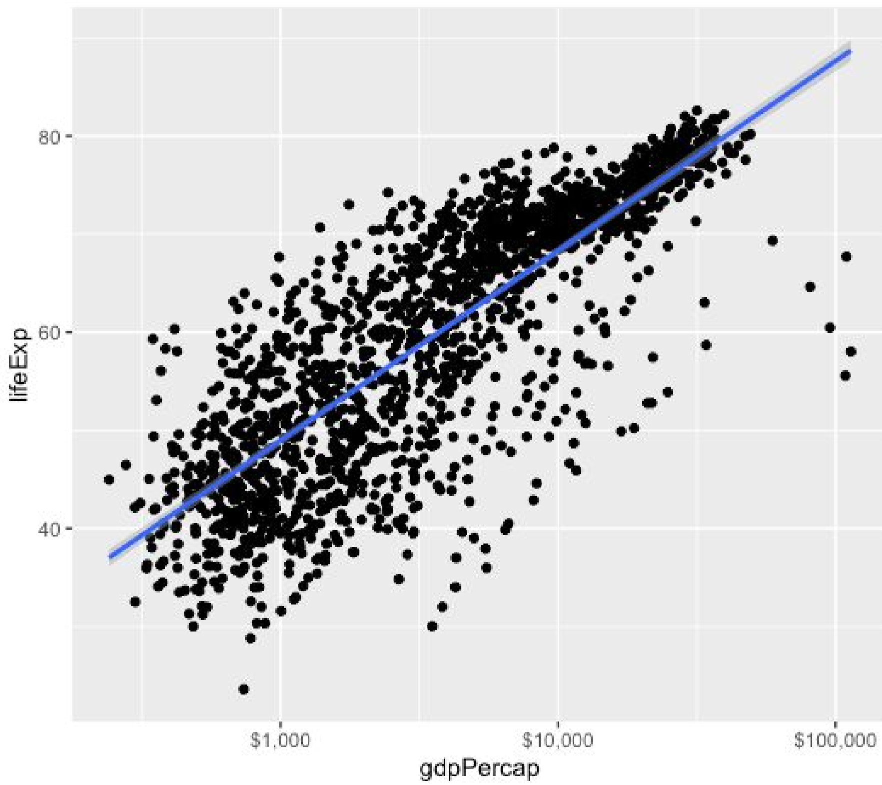
This is the initial, basic graph of Life Expectancy vs GDP per capita. It is very bunched toward the left, and the number of data points makes it difficult to see any real trends. There is a lot of variability in the low GDP per capita area, and very little data in the high GDP per capita area.



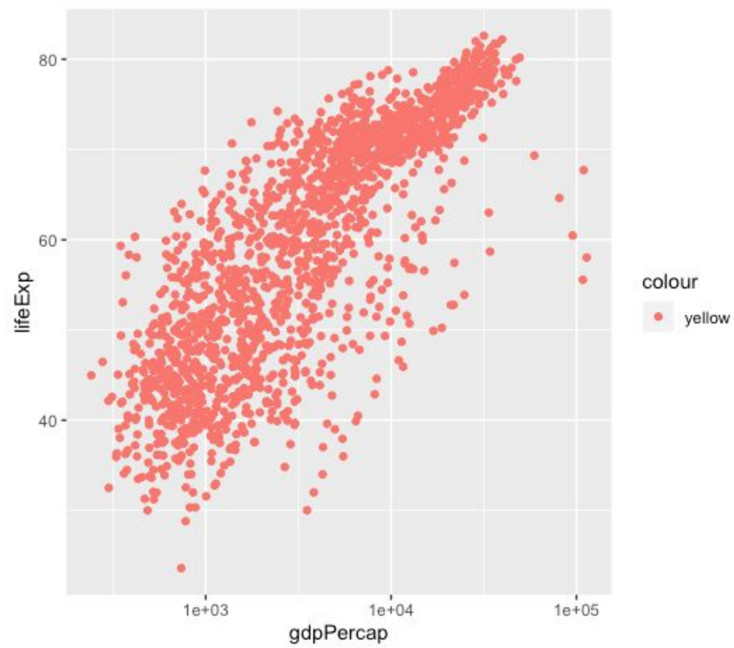
The smoothed version of the initial Life expectancy vs GDP per capita graph makes it seem as if there is a low life expectancy in the highest GDP per capita countries, however, if we adjust the scale, this trend might look different to account for the variability and lack of high GDP countries.



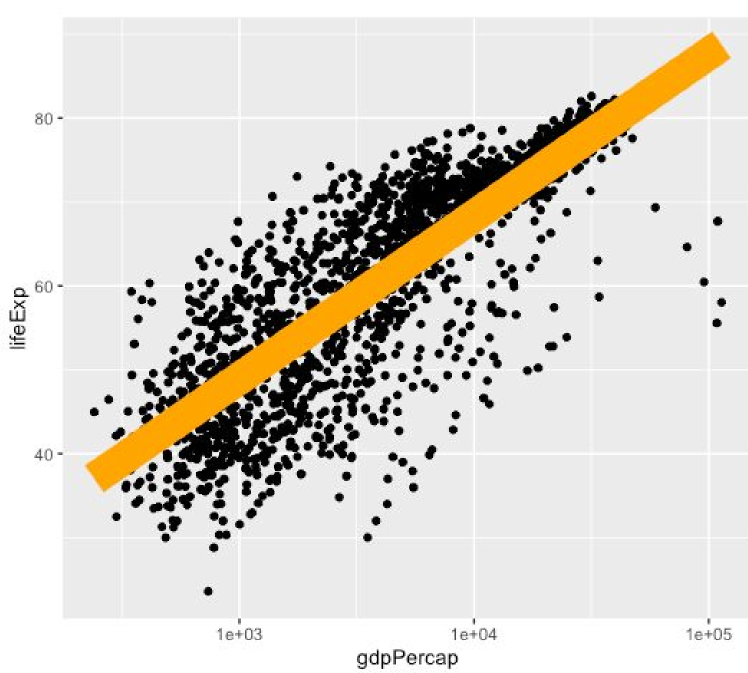
This is the smoothed and pointed version of the Life expectancy vs GDP per capita graph, adjusted to a logarithmic scale. The scale_x_log(10) function does a logarithmic transformation of the x scale by log base 10, and as a result, the values are more spread out on the lower limit side and more dense with the higher numbers.



The dollar() call changes the x-axis to be in terms of dollar amounts, rather than in scientific notation, so the graph labeling makes more sense and is easier to read. Other label options include NULL for no labels, or using waiver() computes default labels based on the data. Labels can be looked up in the ggplot documentation.

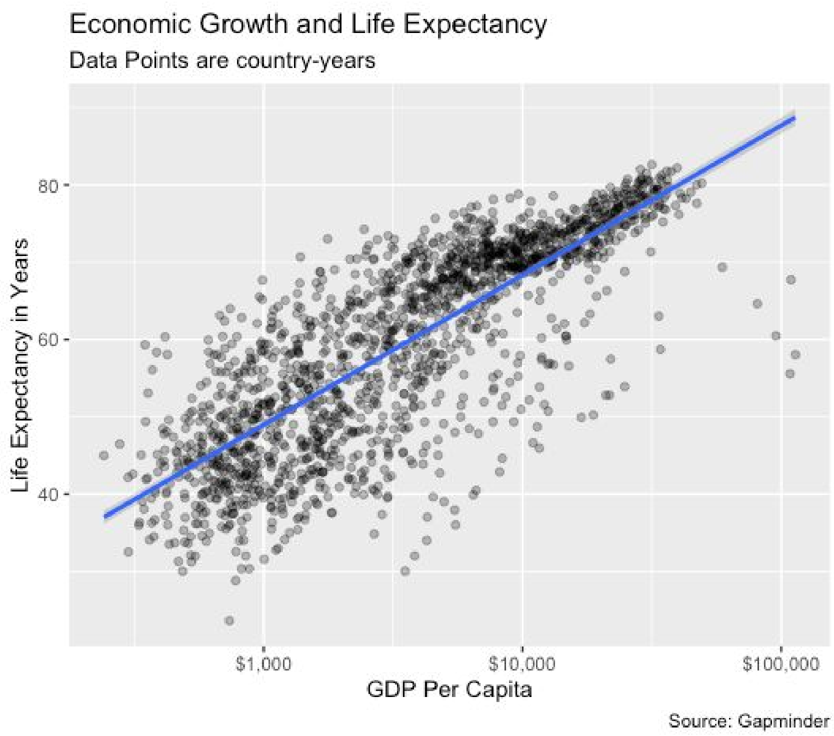


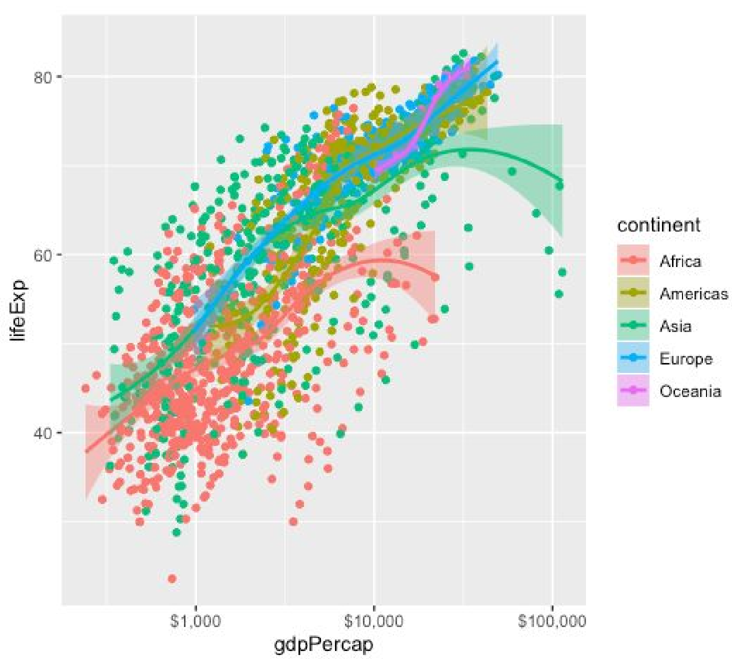
The points are not yellow because color is read as a variable and thus expects a dataset and is not considered to be an aesthetic piece. By placing it in the mapping portion of the ggplot, the graph is not read as utilizing yellow as an aesthetic but rather as a data set.



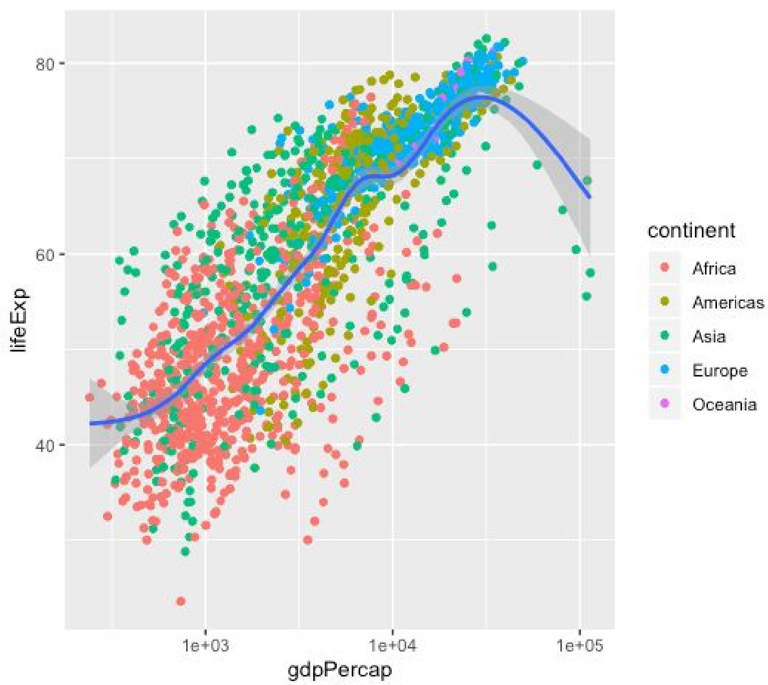

```
geom_smooth(color = "orange", se = FALSE, size = 8, method = "lm")
```

The arguments in this do various things. The ‘color’ changes the color of the trendline. The ‘se’ displays the confidence interval if set to true, or left as default. The ‘size’ changes the thickness of the line. The ‘method’ determines what function to smooth the data.

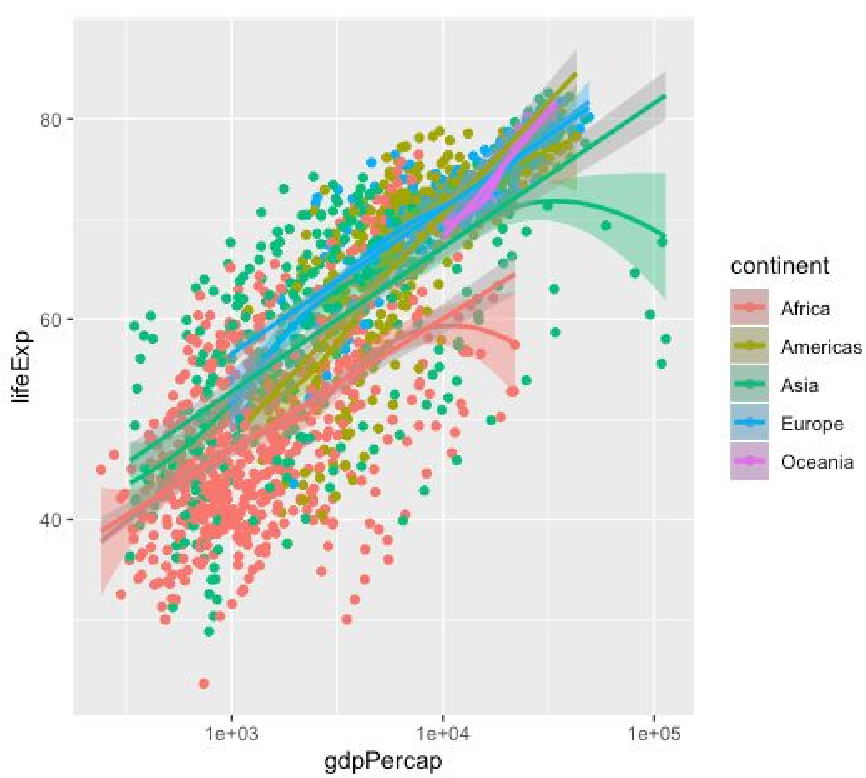




Fill = continent matches the color of each line to the continent that the country is on, and the error lines change as the data varies at each level for each country.



In this one, the code mapping the continent to the colors is only a part of the geom_point() command, whereas it was part of the ggplot initialization in the last code. Thus, it does not apply to every command, just ones where it is explicitly written.



In this one, the code repeating itself each time lends room for a lot of error and also makes the code less efficient and harder to read if others try to replicate it.

Exercise 2

Date:	January 21, 2020
To:	Portugeuse Bank Executives
From:	Big Data Energy Consultants
Subject:	Strategic Marketing Data Analysis

Background:
The Portuguese Bank (“the Company”) recently hired Big Data Energy Consultants (“BDE”) to analyze recent marketing data to offer insights to improve customer acquisition tactics.

Details:
Exhibit 1 (see attachments) examines the relationship between *Job Type* and *Success Rate* in the dataset provided the Company. This dataset includes a number of helpful characteristics of potential (and current) customers that were contacted in recent phone-marketing campaigns. BDE identified correlations between a potential customer’s job and their likelihood of subscribing to the bank term deposit product. BDE found that retirees, students, and managers were most likely, on average, to subscribe to the product based on the current marketing tactics. Blue collar workers, entrepreneurs, services workers were found to be least likely, on average, to be converted to a customer based on the current marketing tactics. Perhaps this finding can be utilized to shape caller lists to specifically target those high-success rate employment types (retirees, students, and managers) while spending less resources to recruit and convert the low-success rate employment types (blue-collar, entrepreneur, services, etc.). Another market tactic could be to consider shaping advertisements and the marketing team’s call scripts to target certain jobs. Retirees may prioritize the security of their wealth and personal interactions, whereas students may prefer a technological interface. Likewise, the current call scripts may not be appropriately suited to appeal to blue-collar workers and entrepreneurs.

Additionally, BDE found the highest rates of successful customer acquisition when calls were placed in the middle of the month, *Exhibit 2*. It is possible that customers are busy reconciling their personal finances during month-end and -beginning and are more likely to sign up for an account in the middle of the month. A marketing campaign that directs its resources in the middle of the month may enjoy greater success.

Attachments:

Exhibit 1: Job vs Success

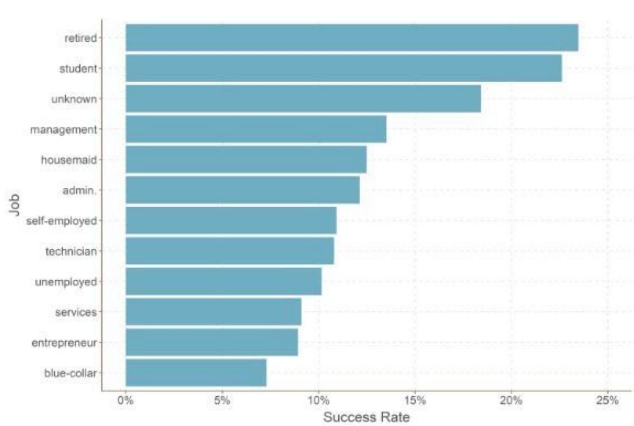


Exhibit 2: Success by Day