Where applicable, answer the following questions using R code and write the code you used in the space below.

1. Create a scatter plot of the relationship between distance traveled and average speed in miles per hour of every flight that left the bay area in 2020 (recall the data set is called `flights`).

2. Is the problem of predicting average speed by distance traveled a regression problem or a classification problem? How about using distance to predict whether a plane travels over 400 miles per hour?

3. Use the `lm()` function to fit a linear model by least squares that predicts the average speed of the flight based on the distance and save it to `m1`.

4. Write out the equation of the linear model that you have fit.

5. Imagine that you wanted to predict the average speed for three flights: one in which a plane travels 10 miles, one in which a plane travels 500 miles and another in which the plane travels 5000 miles. Which of these three predictions, if any, do you expect to be accurate, and why? Comment on each of the three potential predictions.

6. Remake your scatter plot but add a line representing your linear model by adding a layer with `geom_smooth(method = "lm")`.

7. How could you modify your model to improve the predictions that it makes?