Name: Student ID:

Where applicable, answer the following questions using R code and write the code you used in the space below. We will be working with the `flights` data frame in the `stat20data` package.

1. Create a scatter plot of the relationship between distance traveled and air time (in minutes) of every flight that left the Bay in 2020. Copy your code below.

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 flight takes over 200 minutes per hour? Answer each question and explain your choices.

3. Use the `lm()` function to fit a linear model by least squares that predicts the air time 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. Remake your scatter plot but add a line to superimpose your model onto the visualization using the `geom_smooth(method = "lm")` layer. You do not have to copy the code down again. Sketch your final plot below.

6. Imagine that you wanted to predict the air time 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.

7. The `mse_lm` function below is meant to mirror the formula for the Sample Mean Squared Error (MSE) for a linear model, but it's currently incomplete! Complete the function in the gray area below using `b0`, `b1`, `x` and `y`. *Hint: In this function, b0 and b1 are numbers and x and y are vectors. How does the MSE formula translate to vector calculations in R?*

```
mse_lm <- function(b0,b1,x,y) {

  mean(                                                                    )

}
```

8. *Using the `flights` data, the coefficients from your model and your completed `mse_lm` function*, calculate the Sample MSE. Copy the code you used below. Remember that the sample MSE is a summary statistic!

9. To spot check your answer to the last question, *use your model object `m1` and the `fitted()`/`residuals()` functions* to write a data pipeline, beginning with `flights`, to calculate the sample MSE for the linear model. Copy the code you wrote below.