NAMES:

In this problem set you'll practice fitting and evaluating different predictive models to see if you can detect overfitting. The data is available in the folder "Datasets" on bcourses and you can download it from there. For the following questions, fill in the space below with the R code you used.

- 1. Visualize the relationship between x and y with a ggplot and comment on what you see, describing the strength, direction, and shape/form of the association.
- 2. Split your data into training and testing sets; seventy percent of the data should be allocated to the training set.
- 3. Fill out the table shown below. You will fit a model with a polynomial having the degree specified, and report the testing and training RMSE in each case.

Degree	Training RMSE	Testing RMSE
1		
2		
3		
4		
5		
10		
20		
25		

To help you, here is some code that will calculate training RMSE for you, provided you have fit a linear model called m1 and make a training set called train:

You will need to modify this code slightly to help you find the testing RMSE. Write the modified code in the space below given a linear model m1.

4. Describe the pattern in the results you see and explain it.