In this problem set you'll practice fitting and evaluating different predictive models to see if you can detect the presence of overfitting. The data for today's worksheet will be located on Ed!

1. Visualize the relationship between the variables x and y using `ggplot2`. Write your code below, and sketch the visualization you see. Comment on what you see, describing the *strength*, direction (positive/negative), and shape/form (pattern) of the association.

2. Split your data into training and testing sets so that seventy percent of the data is allocated to the training set. Write the code you used to do so below.

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 have a training set called `train`:

```
train |>
  mutate(yhat = predict(object = m1, newdata = _____),
         resid = _______) |>
  summarise(MSE = mean(resid^2))
```

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

4. Describe the pattern in the results you see in two to three sentences. Your description should explain if there is evidence of overfitting, and if so, where you see it.