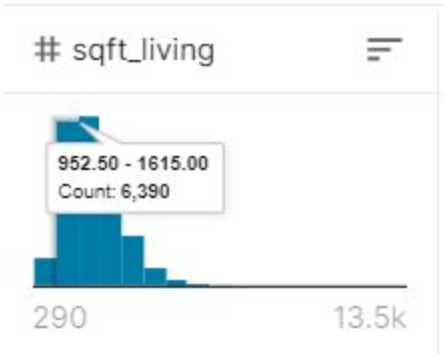
James Levy

1. Data Set details:
    a. What is the dataset
        i. The dataset involves many features that contribute to different houses. Although the dataset includes many features, we are only using two. In our case, we are using price as the value we are trying to predict (y). We are using the amount of living space in square footage as our x value. Here are some graphs for the distributions of our x and y variables.
        1. 

        2. 

        ii. The reason I am using living space as our variable for prediction is that I believe there is a correlation between space and size. It is my understanding that the larger the house, the more expensive it is. Of course, other factors like location play a big role, but in a void, i still expect there to be enough of a correlation to create an interesting regression.
    b. How did I split it
        i. I assume that the data inside of the file is random. Therefore, I split the test set and data set evenly. The first 1000 elements belong to the test set and the second 1000 elements belong to the test set.

*Result notes

I implemented the linear regression via gradient descent and via another simple version. The simple version seems to work fine and will give decent results when running the code. The gradient descent version overflows when the number of iterations/epochs is larger, this causes weird results. I think the logic is correct but seem unable to fix the overflow issue. Please check it anyways. Thank you. **this portion will run automatically after you close the pop up graph for the simple version.