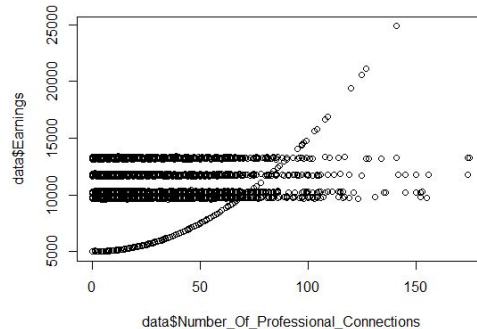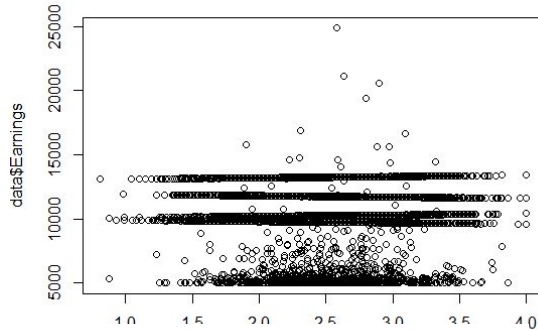# Prediction Challenge 3

Bennett Garcia

Analysis of The Data

- I initially plotted earnings against all the different attributes in side of the dataset.
- Both GPA and Professional Connections showed interesting behavior.
- The graph against Professional Connections also tipped me off that major might play a large role since there were 6 distinct lines.
- I decided to split the dataset by major and make individual models for each. This seemed necessary since some majors exhibited linear behavior while others showed exponential.

Deciding on a model

- Linear regression, random forest and neural networks showed low MSE for 4 of the 6 majors.
- Linear regression had the lowest average MSE. Neural networks also showed good results, but had slightly higher MSE. Random forest also showed promise, but ended up averaging higher then both.
- I decided to use neural networks because they seemed more interesting and I thought I could improve their MSE.

The Other Majors

- The business major and other major weren't showing good results when trained on a neural network.
- I dove deeper into the data and saw that the other major was the major showing the exponential behavior.
- Business also seemed to have odd behavior as well. The first 500 students followed one linear pattern while the other 500 followed a different one.
- I decided to abandon the neural network for these majors and use an exponential model and two distinct linear regression models for the other major and business major respectively.

Training the Neural Net

- I used the nnet package to create and train my network.
- I used one hidden layer with 32 nodes and trained for 100 epochs.
- I also normalized the earnings to avoid exploding gradients.
- I kept training the networks until there loss functions converged on a value of less than .0005

A simple neural network

input layer   hidden layer   output layer