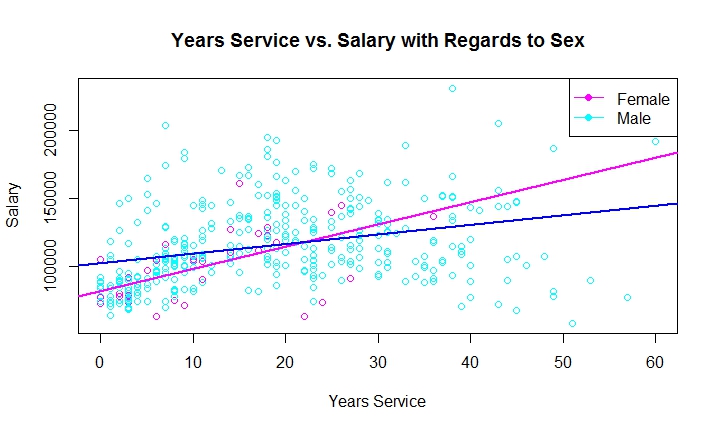
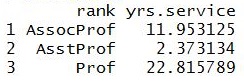
**Problem 1**

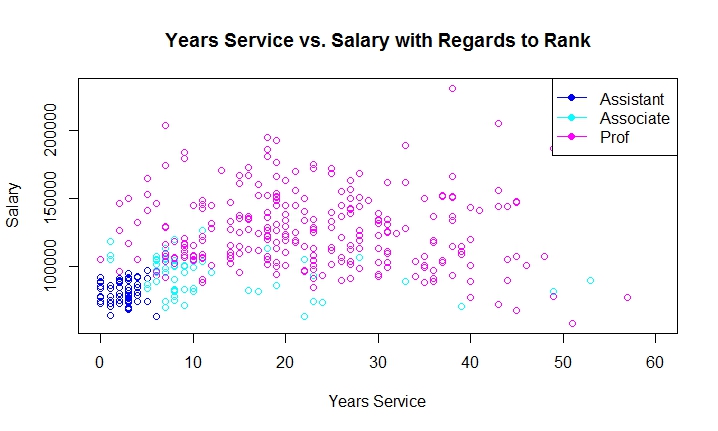
1. Do men and women make a discernable difference overall in their salaries as their time as a professor progresses?

Yes, they do. The first thing you will notice about this graph is that there are many more male professors than female. At the beginning of their career, men start out making more money than women, but as time goes on, women end up making more money more quickly than men. Men and women end up making the same amount of money around 21 to 22 years into their career.



1. Relative to salary, what are typical times in terms of years of service for each rank?

To the right, I have a table with the average years of service for each rank, and below is a graph with years of service vs. salary in terms of rank. In summary, assistant professors do not make very much money and only stay assistant professors for a maximum of about 5 years; associate professors make a little more money than assistant professors and stay associate professors for 11 years on average; and full professors make much more money that assistant and associate professors and stay full professors for much longer.



1. What general relationship exists between discipline and salary as time of service progresses?

**Problem 2**

For this problem, you will explore and model house prices in Boston. In essence, you are going to build the core of your own Zillow app. To begin, first install the MASS package. There is a dataset called “Boston” which includes housing prices from the 1970s. To access this data, do the following:

1. library(MASS)
2. > data(Boston)
3. ?Boston # This will take you to the Boston documentation

The data will then be contained in the Boston data frame.

1. Explore the data using visualizations.
   1. What are some relationships that stand out as important to you and why?
   2. What hypotheses come from your data exploration?
   3. Please include at least 3 plots of interest and explain them.
2. Use regression modeling to construct a predictive model.
   1. What modeling assumptions (e.g. transformation of certain variables, inclusion/exclusion of certain variables, etc.) did you make and why?
   2. What is the final model you settled on? (i.e. the coefficients)
   3. How good is your fit?
   4. Predict prices for a random 10% of the houses contained in the data and compare to actual values. What is the mean and standard deviation of the errors?