Please write the needed R code and provide the corresponding graphics and answers to the questions posed.

Place all of your R functions/code in a single .R file. Please use comments to identify and separate what you do for problem 1 and 2. With respect to your graphics, please generate a PNG and copy them into a single MS Word document. You will submit:

1. an R file containing your code
2. an MS Word document that contains your graphics and answers to questions posed

**Problem 1**

For this problem, you will make several scatter plots to answer a few key questions about professor salaries. (*Hint: Consider using a mix of visualization and regression*)

First, read in the Salaries.csv file included into a data frame. (Bibliographic note: This is the “Salaries” dataset found at: <https://vincentarelbundock.github.io/Rdatasets/datasets.html>)

1. Do men and women make a discernable difference overall in their salaries as their time as a professor progresses?
2. Relative to salary, what are typical times in terms of years of service for each rank?
3. 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?