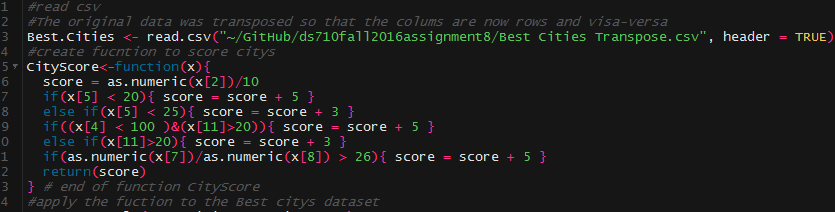
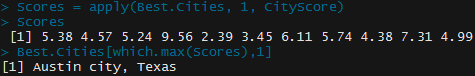
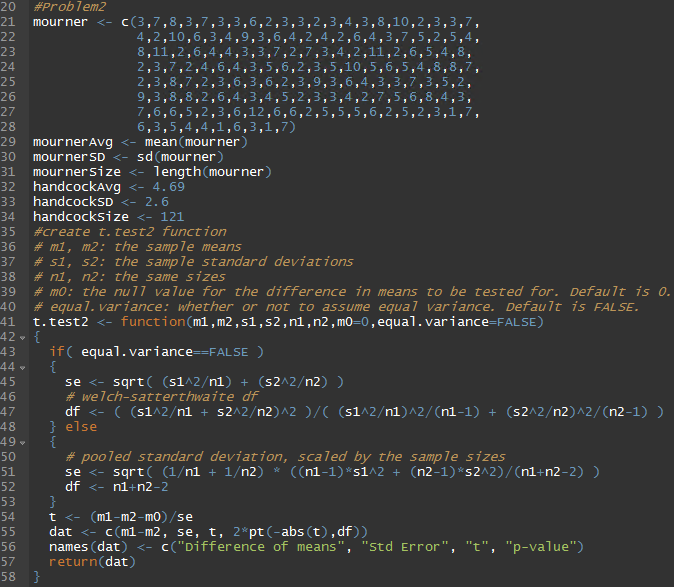
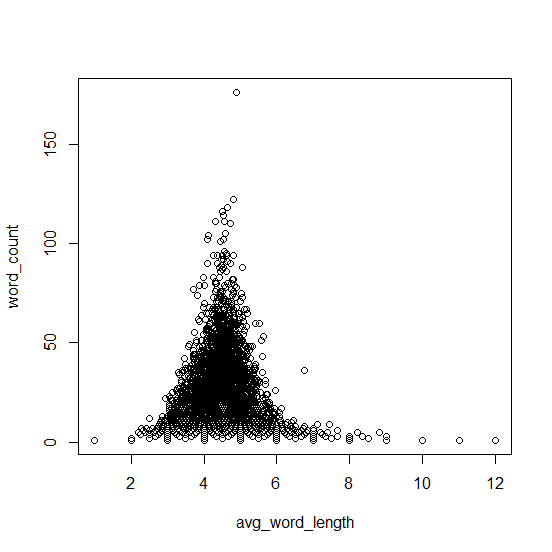
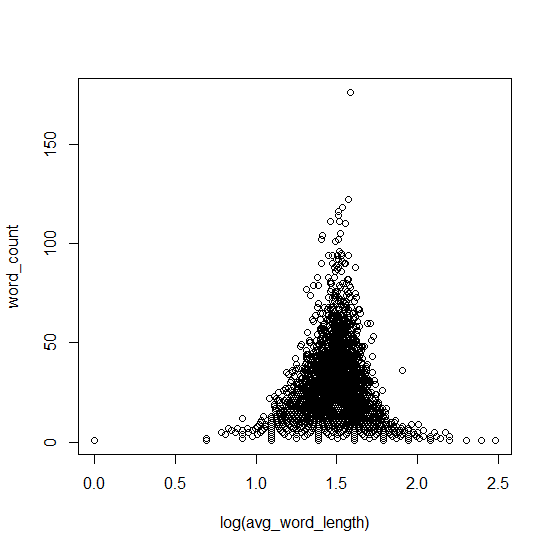
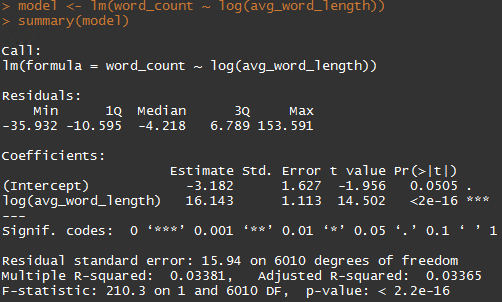
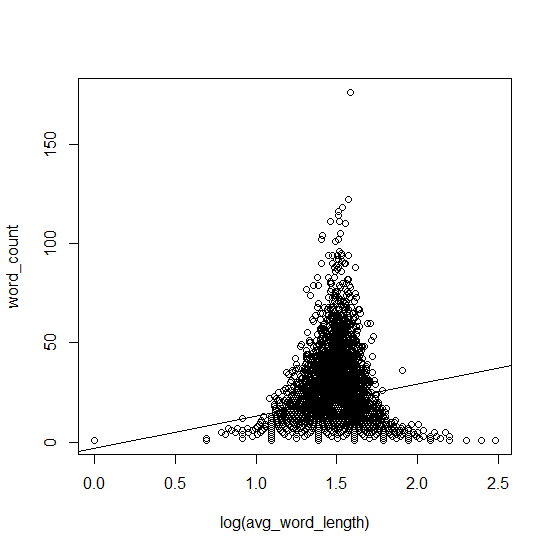
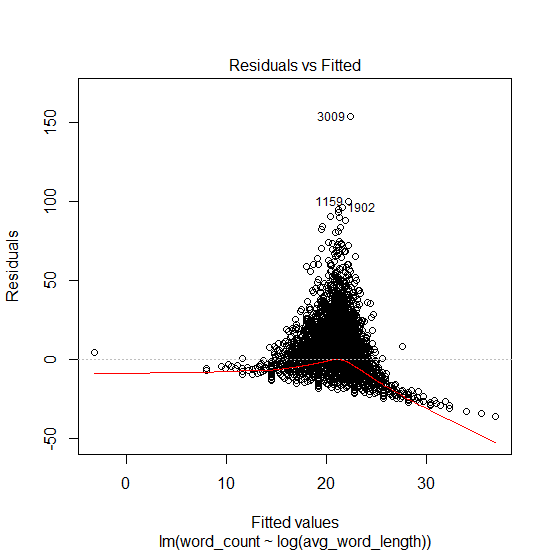
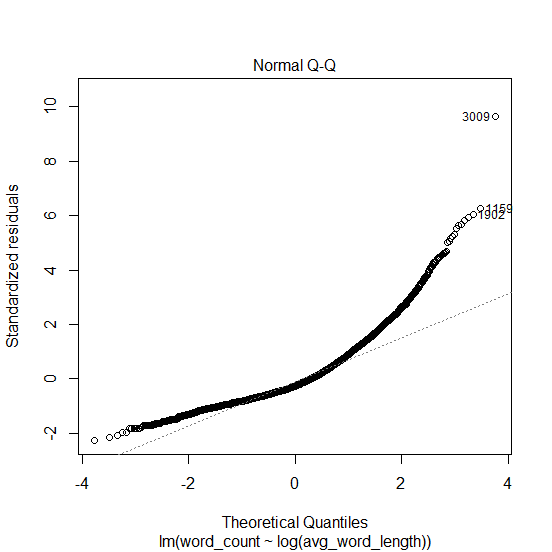
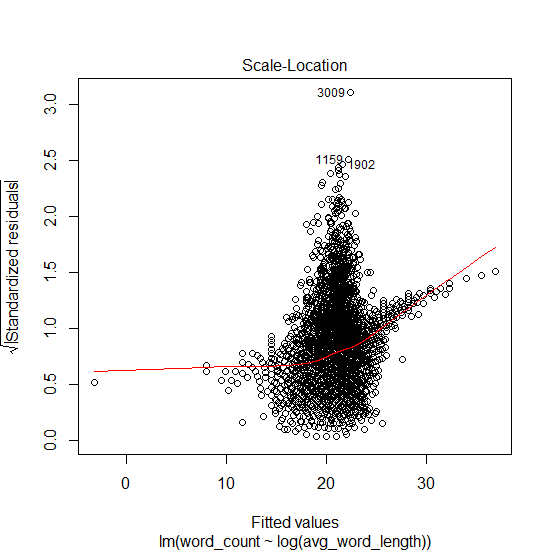
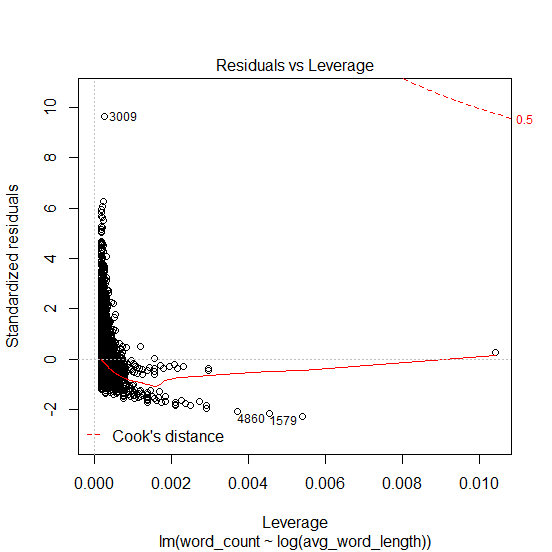
Spencer Swartz

1. In this problem, you will create and apply a function that rates cities based on how appealing they are for you to live in.
   1. The scoring system for CityScore is as fallows:
      1. Percent of people of age 25 or older that have a Bachelor’s degree or higher/10
         1. Example: 45.6%/10= 4.56
      2. If average travel time to work is less than 20 min then score +5 else if average travel time to work is less than 25 score+3
      3. If the max temp is less than 100 and the min temp is greater than 20 then score+5 else if min temp is greater than 20 then score+3
      4. If average household income divided by average selected monthly owner costs (with a mortgage) is greater than 26 then score+5
   2. 
      1. the data was transformed into a more intuitive dataset where the cities are represented by rows in the data and the attributes are identified by columns. This was done in excel using copy/past transpose
   3. 
      1. Austin City would be the highest rated city based on my criteria. This seems fairly accurate. If there were anything that I would change it would be more lenient on the criteria as all scores are fairly low so it is mostly based on the first criteria of % with Bach. degree or higher.
2. Can we use statistical analysis of word lengths to identify the author of an anonymous essay? In Homework 7, you wrote a Python function that counted the lengths of words in the 1770 essay by “A Mourner”. Analysis of other articles published in The Boston Gazette and Country Journal in early 1770 finds that John Hancock wrote a 121-word article with a mean word length of 4.69 and standard deviation of 2.60.
   1. 
      1. 
   2. To be able to draw a conclusion to the test above we are assuming that this one article by John Handcock is representative of all of John Handcock works which there is no guarantee of.
3. In this problem, you will investigate whether sentences with more words tend to contain longer words.
   1. 
      1. A linear model will likely not fit well because of some of the outliers on the x axis. This will be dealt with by taking the log of avg\_word\_length
   2. See 2.a.i
      1. 
   3. 
      1. Based on this linear model we can see with a p-values of less than 2.2e-16 that there is an increase in average word length as sentences increase in size.
   4. 
      1. As show with the linear model there is an increasing slope as word\_count increases so does the log of avg\_word\_length.
      2. 
      3. The Residuals v Fitted plot shows that the residuals do not have a linear patter and that they are not evenly distributed.
      4. 
      5. There is a pretty string deviation from the line showing that the residuals are not normally distributed.
      6. 
      7. We can tell here since the line is not horizontal that the residuals are not spread equally along the rages of predictors.
      8. 
      9. We do not see any outliers in this chart showing that there are no high leverage points that are affecting the linear model.