Stat 135 - Ibser - Lab 3: Dungeness Crabs

The data is under files on bcourses. There are two data sets called crabmolt.csv and crabpop.csv. For question #8 of the lab, it refers to the crab population data, crabpop.csv.

Read the following sections of Ch 7 in Stat Labs (Dungeness Crabs):
Introduction, Data, Background, and Investigations.

You should submit your assignment with two files as in the first lab: 1) Use R Markdown. Submit the .Rmd file and the output file. The answers to questions in the lab should be in the output file. This is preferred format for those of you who are familiar with R Markdown.

2) Don't use R Markdown. Submit code in .R file, with answers to lab questions in comments. Also submit raw output from running code (for example using the sink() function, producing a .txt file).

1. Examine the two data sets. Be curious, make plots and look at summary statistics. After you explore a bit, and maybe after you do later parts of the lab,
   (a) Submit an interesting plot and explain what it tells you.
   (b) Summarize one relevant variable using numbers and words.
2. Develop a procedure for predicting the premolt size of a female adult crab, given only its postmolt size. That is, build a linear regression model where premolt size is the response variable, and postmolt size is the explanatory variable.
3. Make a scatterplot of premolt sizes (along the y-axis) versus postmolt sizes (along the x-axis). Plot the regression line on this scatterplot.
4. What is your interpretation of this regression line? What percent of the variation of the premolt sizes is explained by the model?
5. Examine the residuals and describe what you find.
6. Test the null hypothesis that the postmolt size has no predictive value; that is, test that the slope of the true regression line is 0.
7. Now construct a new linear model, by dropping the juvenile crabs from the data set. (These are all the crabs with premolt carapace widths under 100mm.) Plot the regression line on the same scatterplot (from part (3)), in a different color. Comment on this regression.
8. Use your procedure (from (2)) to describe the premolt size of the molted crabs collected immediately following the 1983 molting season (the data set in Crab Population). Provide a numerical summary and make a histogram for the size distribution prior to the molting season of the crabs caught in 1983. Use different colors to distinguish the crabs that molted from those that did not molt. (The histogram that you get should be similar to Figure 7.2 on page 144 of Stat Labs.)