The following questions relate to Exercise 25 on page 147 of your text where a random sample of $n = 2584$ Americans with paying jobs were asked about their income level and educational level in 2006. Educational level was simplified into 5 distinct categories. You must report your answers to this problem in the format according to the Syllabus and Writing a Statistical Report available on the course website. The report, not including the Appendix that contains your R-code and R-output and figures and any tables, should not exceed two pages. Your grade will be determined by how well you answer the questions and by the organization and clarity of your write-up.

1. Load in the data for this problem. In R:
   ```r
   library(Sleuth3)
   d = ex0525
   summary(d)
   ```

2. Be sure to indicate the sampling plan and the study design in the introduction to your report. In the scope of inference part of your report you will indicate how the sampling plan and the study design affects the applicability (or lack of applicability) of your conclusion.

3. Plot side-by-side boxplots of the income of the educational groups in R. Also plot side-by-side boxplots of the log_{10}-transformed income of the educational groups in R. The education categories are initially displayed alpha-numerically, so re-order the categories before you plot.
   ```r
   d$Educ = factor(as.character(d$Educ),levels = c("<12","12","13-15","16",">16"))
   boxplot(Income2005 ~ Educ,data=d)
   boxplot(log10(Income2005) ~ Educ,data=d)
   ```
   Which of these two plots do you prefer for presenting these data to a reader? Why?

4. Update the R-code in #3 to include the `boxplot()` options `main="Informative Title"`, `xlab="Informative x-axis label"` and `ylab="Informative y-axis label"` to label plot appropriately. Include whichever of the two boxplots that you think represents the data best in your report.

5. Although you may have liked displaying these data on the log-scale, it is not necessary to transform these data prior to statistical analyses (either for a permutation test or an ANOVA). Why? Put your answer in the Statistical Procedures Used section of your report.

6. Apply a permutation test of the hypothesis that the mean income of a worker is associated with the worker’s level of education. In HW1 and in the Chapter 1 notes you performed a randomization test using a difference in means as the test statistic. This was a reasonable test statistic when comparing only two groups. When comparing more than 2 groups in a randomization or permutation test as in this problem (where 5 groups are compared) the variance of the means can be used as the test statistic (see the R-code in the Chapter 5 notes).

7. In an Appendix to your report, include a histogram of the test statistic over many permutations of the data and report the test statistic value calculated from the data.

8. Be sure to check the assumption that the permutation test is appropriate. Report the $p$-value in the Summary of Statistical Findings section of your report. Put your R-code and R-output in the Appendix.

9. Apply an ANOVA to test the hypothesis that the mean income of an employee is associated with the employee’s level of education (see Chapter 5 notes for R code). Be sure to check the assumptions that the ANOVA is appropriate. Report the $p$-value in the Summary of Statistical Findings section of your report. Put the R-code and R-output in the Appendix.

10. Compare the results from the permutation test to the ANOVA. Are they similar? Did you expect the results to be similar? Why or why not?