Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ STT 3850 Lab Activity 1: Basic Exploratory Data Analysis

Go to our class AsULearn page and download the file **StudentGuessingData** to your working directory in **R Studio**. This file contains information on students in two sections of a summer Stats class.

Create an R Markdown/Notebook document that will contain the R codes and answers to the following:

1. If the reason for collecting this data is for the instructor to describe and compare the students from the two sections, should the data be treated as a population data or a sample data? How many cases or observational units are there? Which of the variables are numerical and which are categorical?
2. Construct the frequency table for the WorkStat variable using the **table()** command. What percent of the students are working students?
3. Describe the distribution of the GPAs of all the students by describing its shape and discussing statistics that measure the center and the spread of the data distribution. Attach an appropriate graph that shows this shape and estimate the average GPA from this graph.
4. Repeat #3 for the working students.
5. Repeat #3 for non-working students.
6. Which group of students has higher average GPA, the working students or the non-working students? Which group’s GPAs have more variability?
7. Which section has higher GPA on average, is it section 1 or section 2? Justify your answer.
8. Compare and discuss the number of hours working students are registered compared with non-working students. Attach descriptive statistics to justify your answer.
9. Do working students expect to get a higher grade than non-working students? Justify your answer.
10. Students with GPAs above 3.5 are eligible for a certain scholarship award. How many students from each section are eligible for this award?

Upload the resulting .html document at the Activity 1 Submission folder in AsULearn by midnight on Wednesday, Jan. 31, 2018.