CSCI N317 Project Two

**Data processing using R**

Find a large data set with at least 3000 rows and 5 columns, using R to perform data analysis, data processing or visualization. Think of a purpose, and design some tasks that use at least 8 of the following techniques.

1. Select rows
2. Select columns
3. Perform some computation
4. Create new columns
5. Change column types
6. Apply some functions on data
7. Data cleaning
8. Merging
9. Reshaping data
10. Binning
11. Visualization
12. Sampling
13. Counting

**Rubric**

1. Data set meets the requirement. (10’)
2. Use comments in the script file(s) to specify your name, the purpose for each task, and purpose for each code block. (10’)
3. Label your file(s) or tasks with the numbers in the above list. (10’)
4. Your codes match with your task descriptions and generate correct results. (60’)

60 points will be evenly distributed to each task picked from the list, e.g. if you select 10 out of 13, each task is worth 6 points.

1. Submit a zip file that contains your data file and scripts. File name format is lastname\_project2.zip.