**Homework #6 (Due Oct 8 11:59 PM)**

IST 3420 - Fall 2017, Chen

**Name**: \_\_\_\_\_\_\_\_Adam Forestier\_\_\_\_\_\_\_\_\_\_\_\_

**Weblog Analysis (20 points)**

**Instruction:** Finish the following tasks by using R Markdown. Then upload this document with your answers and your R Markdown file to “Homework 6” on Canvas.

1. Download the sample apache web server log file “sample\_apache.log”.
2. Read the log file into R (2 points).
3. Use dplyr::mutate() function to add the following columns (4 points):
   1. remote.host
   2. status
   3. request.datetime [use the format of POSIXct dates, refer to as.POSIXct() function]

*Note: mutate function does not support POSIXlt date, so choose the POSIXct date. Pay attention to the format parameter in the as.POSIXct() function.*

*This column should contain both date and time information (down to the second level).*

* 1. weekday [i.e., day of week, refer to weekdays() function]
  2. request.method
  3. request.uri

1. Use dplyr::select() function to only keep the above 6 columns in the dataset (2 points).
2. Use dplyr::count() function to create a data frame called “visit\_per\_IP” which shows the total number of visits for each unique IP address (). The data frame contains two columns (5 points):
   1. remote.host
   2. count\_of\_visit
3. Subset the “visit\_per\_IP” to create a data frame called “visit\_per\_IP2” which only contains information for unique IP addresses that have 2 or more visits. How many IP addresses have 2 or more visits? \_\_\_\_2306\_\_\_\_\_\_\_\_\_\_ (5 points)
4. Upload this document with your answers to “Homework 7” on Canvas.
5. Upload your R Markdown file to “Homework 7” on Canvas. (2 points)