STAT 406: HW5

- All computer code should be written using the language R. Type ALL your code into one PLAIN Text format file. Plain text format is available by default in R. Please do not use Microsoft Word .doc format or .rtf format of .pdf format. Inside your plain text file, make sure you identify each problem in a comment placed at the beginning of the problem. The file name should match your name as in 'JohnDoe.R'.
- Submit your R code file online (under Assignments) at or before the due date, and hand in a hard copy of the code as well as a printed copy of your answers to the questions. The hard copy is due at the beginning of your respective lab sessions.
- I recommend that before submitting your homework, you also create a new directory and run your R code, to make sure that it is self-contained and runs as you intended.

1. Using the baseball database, we want to investigate (very crudely) whether pitchers get better with age. A basic statistics to evaluate a pitcher is the walk to strike-out ratio (BB/SO), where BB and SO can be found in the Pitching table. Consider the year 2004. Group all the players that played in the 2004 season by age, and for each age group calculate the group average walk to strike-out ratio. Plot these average Walk to Strike-out ratios as a function of age. What is this plot suggesting?

2. We have seen how to process the NSF award files in class. Download the 2012 NSF awards files in XML formats from
   `https://www.nsf.gov/awardsearch/download.jsp`.
   Adapt and expand on the code written in class to find the top ten universities (only universities) which have received the most NSF funding in 2012.

3. The National Transportation Safety Board database contains information about civil aviation accidents within the United States. The data can be downloaded in XML format from

   `http://www.ntsb.gov/_layouts/ntsb.aviation/index.aspx`

   The data dictionary can be found at

   `http://www.ntsb.gov/_layouts/ntsb.aviation`
   `/AviationDownloadDataDictionary.aspx`

   Download the data file `AviationData.xml` and process it to build a dataframe that gives for each accident: the event date, event location, event country, and the total number of fatal injuries. Use the data to find the total number of fatal aviation injuries in the United States for each year. Make a plot of that series.