**Workshop Exercise 6**

**Exercise VI: Graphing**

**Dataset = Baseball.sas7bdat**

In this exercise we will use the permanent SAS file baseball.sas7bdat, in the ExerciseDataSets folder. The codebook for this dataset can be found in the Appendix.

1. Make a boxplot of SALARY for the American and National League players (LEAGUE). Describe the distribution of SALARY for the two leagues based on these boxplots.
2. Make a paneled boxplot of SALARY for each DIVISION, paneling by LEAGUE.
3. Make a paneled histogram of SALARY for the American and National League players using Panels for the two LEAGUES.
4. Make a Scatterplot with a single regression line showing the relationship between SALARY as Y and years in the Major Leagues (YR\_MAJOR) as X. Include 95% confidence limits for the predicted regression line.
5. Make a paneled scatterplot, paneling by LEAGUE. Within each panel include separate regression lines for the relationship between SALARY and YR\_MAJOR for each DIVISION. Edit this graph to add some interesting things to it.
6. Generate a scatterplot matrix including the following variables: YR\_MAJOR, CR\_ATBAT, CR\_HITS, CR\_HOME, CR\_RUNS. How do these variables relate to each other?
7. If you have time, calculate the natural log of salary (call your new variable LOGSAL) and redo questions 1 through 5 using LOGSAL, rather than SALARY.