

## WEBSCRAPING DAY 2

From our last activity, you should have an R-notebook (.RMD) file that has the appropriate code to webscrape at least 4 columns of data from one of the NCAA sites listed in option #1 of the *Webscraping Activity* on our D2L site and save the data in a dataframe. We will build on this in today's activity. Similar to our previous activity, we will try to explain scoring.

1. Start by webscraping scoring and at least four other offensive statistics from the sport of your choosing. (It could be the same sport as the previous activity.)
2. Clean up the data and create a dataframe containing the data.
3. Do multiple regression with the four offensive statistics trying to explain scoring.
4. Determine which of the offensive statistics are significant, and which are insignificant.
5. Create a new variable (statistic) which includes the correct ratio of significant statistics.
6. Create a linear model that uses your new variable to explain scoring. Explicitly give the equation,  $R^2$ , and standard error of this model.

Turn your R-notebook into the appropriate Gradescope assignment. It is due at the end of class.