STAT 360
Dr. Straight
Quiz 4, Due Tuesday, April 25

Note: *You are on your honor to do your own work and not help others!*
Submit a PDF file of your knitted R Notebook.

Do something interesting with unsupervised learning. Here are a couple of ideas using the Lahman package.

(1) Use the `Teams` data frame and some of the following variables: `Rank`, `DivWin`, `LgWin`, `WSWin`. Consider an appropriate range of seasons, and aggregate each variable (e.g., total 8 times the number of World Series wins). Use hierarchical clustering. What does the model yield? For example, perhaps the model is able to discern "successful" teams from unsuccessful ones.

(2) Use the `Teams` data frame and variables such as `R`, `RA`, `HR`, `HRA`. Consider an appropriate range of seasons, and aggregate each variable. Use hierarchical clustering. What does the model yield? For example, perhaps the model is able to discern teams that play in "hitter-friendly" stadiums from teams that play in "pitcher-friendly" ones.

(3) Use the `Pitchers` data frame and variables such as `W`, `L`, `G`, `GS`, `SV`. Consider an appropriate range of seasons, and aggregate each variable. Include only those pitchers with a large value of total `G`. Use *k*-means clustering. Produce an appropriate graph, such as a scatter plot, with `W + L` on one axis and `SV` on the other. What does the model yield? For example, perhaps the clusters show starting pitchers (high `W + L`, low `SV`), closers (low `W + L`, high `SV`), etc.