STAT 406: HW4

- 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.

Use the Baseball database to do the following.

1. Use the table `Salaries` to compute the average payroll of the league (total salaries paid in the league in a given year divided by the total number of teams in the league in that year) for each year after 1984. It is ok to solve this with two SQL queries.
2. Extract from the database the payroll of the World Series winner for each year after 1984. Compare this with the league average obtained above (by plotting the two series on the same plot). The binary variable `WSWin` (which takes values 'Y' or 'N') in Table `Teams` gives the winner of the World Series. Explain how you handle the fact that there was no world series in 1994.
3. `Left-handed versus right-handed players.` In Baseball, the Batting Average is a commonly used statistics to compare hitters. It is given by the number of hits divided by the number of times at bats.

These numbers can be found in the table `Batting` of the Baseball database. The number of hits (for each player, and for each year) is given in the attribute `H`, whereas the number of times at bats (for each player, and for each year) is given in `AB`.

For each given year, compute the average Batting Average (the first average is over over players, the second is part of the definition of the variable) for all the left-handed players. Repeat the calculations for the right-handed players. Plot the two time-series on the same plot. Whether a player is left/right/both-handed is available in the table *Master*.

4. For each player who has played after 1984, compute their average (over the years) annual salaries and their average (over the years) batting average. Do a scatter plot of salaries as function of batting average.