Homework 3

For this homework you will create an R Markdown file and output (PDF) and upload both to wolfware. Be sure to include text explaining your thought process/what you are doing with your questions.

The purpose of this homework is to get practice manipulating data with packages from the `tidyverse`. We'll utilize the `Lahman` package (you may need to install this).

Note: Use chaining where possible!

1. Print the "Pitching" data set out after first converting it to a tibble.
2. Print only columns ending in "ID" and the W, L, and ERA columns.
3. What is a major advantage of using the `select` function over `Base R` methods for selecting columns?
4. Add to your code from the previous question above (with chaining) to then subset in order to only include rows from 2010 to 2015.
5. Continuing your result, now create add a new column that is the Win loss percentage (wins divided by wins and losses).
6. Continuing, now sort your data set by Win/Loss Percentage in descending order.
7. Continuing, now find the average ERA for each team for each of the years 2010 through 2015. Remove NA values when finding the mean. Save this result as a new object.
8. Take the output from question 7 and put the data into wide form with years each being a column (you should have 31 rows). Print all of these rows out in your document (hint: use the print function and specify `n = nrow(.)`.)
9. Challenge! You should notice 6 `NA` values in your data. This is because the Florida Marlins changed their name to the Miami Marlins. Combine the `MIA` and `FLO` rows into one and reprint the entire data frame using the `kable` function. Make the new `teamID` '`MIA/FLO`'. Note the `teamID` column is a factor so you will need to mess around with the levels.
10. Lastly, explain the idea of coercion and give an example of explicit and implicit coercion.