Introduction to Data Science
Homework 5: Due Friday October 5 at 2:00pm

Exercises:

1. Using an appropriate choice of functions in the `dplyr` package, modify the `Salaries.csv` dataset to create a new variable called `hp_df` that is a data.frame that contains the highest paid player on each team in each year.

2. In the `flights` dataset, currently `dep_time` and `sched_dep_time` are convenient to look at, but hard to compute with because they're not really continuous numbers. Convert them to a more convenient representation of number of minutes since midnight. Use `mutate` to add this as a new column.

3. What trigonometric functions does R provide? Make plots of each of the trigonometric functions over an appropriate period.

4. Brainstorm at least 5 different ways to assess the typical delay characteristics of a group of flights. Consider the following scenarios:

   - A flight is 15 minutes early 50% of the time, and 15 minutes late 50% of the time.
   - A flight is always 10 minutes late.
   - A flight is 30 minutes early 50% of the time, and 30 minutes late 50% of the time.
   - 99% of the time a flight is on time. 1% of the time it's 2 hours late.

   Which is more important: arrival delay or departure delay?

5. Go through the first five sections of chapter 7 on Exploratory Data Analysis in R for Data Science (the sections and chapter numbers correspond to the online version). Use what you learn to conduct an EDA for the `gapminder` dataset contained in the `gapminder` package.

6. Read the R Markdown chapter of R for Data Science. (This is chapter 27 of the online version). Use what you learn to create the beginning of a notebook for your semester project. Start with a section titled "Problem Description and Objectives," once you have written it, copy and paste the background, problem description, and objective into this section. Also include a section titled "Data Description," and copy and paste the data description that you already submitted into this section of your notebook.

7. Create an R Project and corresponding folder (directory) that contains the notebook you just created, and also a subfolder called Data that contains any data files that you have obtained for your semester project.

8. **Optional:** If you want to use git or github to maintain your project files (highly recommended), see the course instructor if you want help with this.