Homework 3

Your Name Here

Due Date Here

Exercise 1 (STAT 612: 5 points; STAT 412: 7.5 points)

Baltimore City Crime Data:

1. Import the data `BPD_Part_1_Victim_Based_Crime_Data.csv`
2. Convert the given dates and times to date classes. For `CrimeTime`, not all of the rows conform to the "HH:MM:SS" format. If you cannot figure it out, remove those rows where the parsing failed.
3. Make `Location 1` into two columns `LocationLat` and `LocationLon`
4. Determine the % of crimes committed between midnight and 4:00 am.

Exercise 2 (STAT 612: 5 points; STAT 412: 7.5 points)

1. Read the description of the babynames dataset with (you might need to install babynames)

   ```
   library(help = "babynames")
   ```

   What are the data frames in this data set? What are the keys in each data frame?

2. Read the description of the nasaweather dataset with (you might need to install nasaweather)

   ```
   library(help = "nasaweather")
   ```

   What are the data frames in this data set? What are the keys in each data frame?

Exercise 3 (STAT 612: 5 points; STAT 412: 7.5 points)

This exercise concerns the Lahman dataset. You can read about it with:

```
library(Lahman)
```

```
help("Lahman-package")
```

For this exercise, we'll use the `Master`, `Batting`, `Pitching`, `Fielding`, `Teams`, and `Salaries` data frames.

1. Load these data frames into R and read about them.
2. Find all the names of the players who have ever had a stint (from the `Fielding` data frame) in the Red Sox (or the Boston Americans) in years where they made it to the World Series (so they won their leagues). Print out the first ten names (arranged in alphabetical order of last name). Note that the World Series was not played each year and began in 1903.

   Your output should look like this:

   ```
   ## # A tibble: 0 x 2
   ## # Groups:   yearID [0]
   ## # ... with 2 variables: yearID <int>, n <int>
   ```

   ```
   ##    nameFirst nameLast yearID
   ## 1    Alfredo   Aceves   2013
   ## 2      Jerry    Adair   1967
   ```

```
## 3      Terry    Adams   2004
## 4        Sam    Agnew   1916
## 5        Sam    Agnew   1918
## 6       Nick  Altrock   1903
## 7        Abe  Alvarez   2004
## 8      Jimmy Anderson   2004
## 9      Ernie   Andres   1946
## 10       Kim   Andrew   1975
```

3. Some players play on multiple teams each year. Construct a data frame that contains the total salary for each player for each year. Also construct a data frame that contains the total number of at bats and hits for each player in a year.

4. The batting average of a player is the number of Hits divided by the number of at bats. A larger value is good.

   Using the data frames you created in part 3, explore the marginal association between a player's batting average and their salary.

   Also explore if this association has changed over time (for example, because sports teams are getting more stats-savvy).

   Limit the pool of eligible payers to the years after 1985 (when salary information started being collected) and to players with a minimum of 400 at bats.

5. Find the salary of all players named "John" in even numbered years after 1985. Print the first ten values arranged in descending order of salary.

   Your output should look like this:

```
## # A tibble: 10 x 4
##    yearID nameFirst nameLast   salary
##     <int> <chr>     <chr>       <int>
##  1   2010 John      Lackey   18700000
##  2   2016 John      Lackey   16000000
##  3   2012 John      Lackey   15950000
##  4   2016 John      Danks    15750000
##  5   2014 John      Lackey   15250000
##  6   2014 John      Danks    14250000
##  7   2008 John      Smoltz   14000000
##  8   2004 John      Smoltz   11666667
##  9   2006 John      Smoltz   11000000
## 10   2000 John      Smoltz    8500000
```

Exercise 4 (STAT 612: 5 points; STAT 412: 7.5 points)

1. Load into R the list of acceptable (2015) Scrabble words from.

   Hint: "NA" is an actual word. It means "no" or "not".

2. How many words either begin or end in "X"?

3. How many words contain all of the vowels (A, E, I, O, and U)?

4. What are the shortest words that contain all of the vowels? (there should be five of them)

5. Switch the first and last letters of all of the words. How many of them are still words?

6. How many of the words that are still words after switching the fist and last letters have **different** first and last letters?

7. What are the longest words that are still words after switching the first and last letters **and** where the first and last letters are different? You should end up with six words (three pairs of words).

8. (Half a point extra credit) Find the words with the longest consecutive sequence of consonants (anything but A, E, I, O, and U). You should only get one word.

   Repeat this exercise where we also consider Y a vowel. You should again only get one word.