DSCI401 - Homework 6

Due: November 10, 2024
Homework should be submitted as an R Markdown file with links to Google colab notes where necessary. Homework should be turned in on Sakai.

Answer all questions below with R AND Python.

1. Answer the questions below:

   (a) Use the HELPrct data from the mosaicData to calculate the mean of all numeric variables (be sure to exclude missing values).

   (b) Find the mean of all the numeric variables stratified by sex and age group where age groups are defined as ranges of 10 years (i.e. 0-10, 10-20, 20-30, etc).

   (c) Using the data set created in the previous problem, create a set of line plots with the average age of the age group on the x-axis and each of other numeric variables on the y-axis in separate plots stratified by sex. (Note: You are not allowed to use a for loop here or simply copy-and- paste 20 times!)

```
library(mosaicData)
head(HELPrct[,1:6])

##   age anysubstatus anysub cesd d1 daysanysub
## 1  37            1    yes   49  3        177
## 2  37            1    yes   30 22          2
## 3  26            1    yes   39  0          3
## 4  39            1    yes   15  2        189
## 5  32            1    yes   39 12          2
## 6  47            1    yes    6  1         31
```

2. The team IDs corresponding to Brooklyn baseball teams from the Teams data frame from the Lahman package are listed below. Use map_int() to find the number of seasons in which each of those teams played by calling a function called count_seasons.

```
library(Lahman)
bk_teams <- c("BR1", "BR2", "BR3", "BR4", "BRO", "BRP", "BRF")
```