# Homework 2
# Data Science, PCS – UConn, 2022
# Due Date: July 13, 2022 by 8pm

1. On HuskyCT you will find a file with the season's best results for men's long-jump, since 1960 (longjump1.txt). Data were retrieved from Wikipedia: `https://en.wikipedia.org/wiki/Long_jump#Men_5` Use the following command to read the data:

   ```
   longjump <- read.csv("longjump1.txt",sep="\t",header=TRUE)
   ```

   If you get an error that the file is not found, make sure you set the working directory.

   (a) Use the commands you learned in class to create a frequency table, by the top athlete's nationality.

   (b) Use the commands you learned in class to find the minimum, maximum, and mean result across all years.

   (c) Create a plot with the year on the x-axis and the best result on the y-axis. Create an aesthetically pleasing plot with the options you saw in the notes. Comment on the plot and describe the most striking observations which can be seen in the plot.

2. Run the following code to generate a data frame called ucb:

   ```
   gender <- rep(c("female","male"),c(1835,2691))
   admitted <- rep(c("yes","no","yes","no"),c(557,1278,1198,1493))
   dept <- rep(c("A","B","C","D","E","F","A","B","C","D","E","F"),
               c(89,17,202,131,94,24,19,8,391,244,299,317))
   dept2 <- rep(c("A","B","C","D","E","F","A","B","C","D","E","F"),
                c(512,353,120,138,53,22,313,207,205,279,138,351))
   department = c(dept,dept2)
   ucb <- data.frame(gender,admitted,department)
   rm(gender,admitted,dept,dept2,department)
   ls()
   ```

   (a) Use one of the functions you saw in class and the notes to create a summary of each variable.

   (b) Create a contingency table of gender by department called GenderDept, and include the row and column sums.

   (c) Using the table you created, plot a spinogram.

   (d) Dose there seem to be a relationship between department and gender?

3. "airquality" is a built-in data set in R. It has 154 observations and 6 variables. Read the description by yourself by typing ?airquality

   (a) Get summary statistics on all 6 variables in "airquality".

   (b) Draw a boxplot of the Temp variable.

   (c) Plot the kernel density graph of Wind by each month. And compare them on one graph. Add necessary titles, and legend.

   (d) Are there any conclusions you can draw from the plots and tables? Feel free to analyze additional variable.