Statistics 3080
Homework 3
Due: Wednesday, February 14

Complete the following problems in a R Markdown file and submit your compiled PDF.

Problem 1): The data file *nym2002.txt* contains the finishing time and some demographics for select finishers of the 2002 New York City Marathon.

(a) Read in the file and save the data in a data frame called `nym2002`.

(b) Determine the number of finishers' times that are contained in this data set.

(c) Determine the fastest and slowest finishing times given in the data.

(d) Rewrite the times determined in part (c) in the folowing form: XX hours and XX minutes.

(e) Determine the number of men in the data who finished after the slowest woman in the data.

(f) Determine the home state or country of the fastest finisher in the data.

(g) Determine the age of the slowest male finisher in the data.

(h) Determine the finishing position of the fastest finisher in the data.

(i) Determine number of finishers in the data whose home country is not the U.S. (include U.S. territories as the U.S.)

Problem 2: The data file *tv.txt* contains the number of minutes that subjects spent sitting at home in the evening (x) and the number of minutes that subjects spent watching tv at home in the evening (y). The variable z represents whether the subject regularly exercises (2) or not (1).
Note: You do not need to build upon your code from the previous parts.

(a) Plot these data using a point that has only an outline.

(b) Modify the graphic created in part (a) such that it is made up of rectangles that vary in colors showing where the data are instead of plotting individual points.

(c) Modify the graphic created in part (a) to distinguish those who exercise regularly by color using blue and orange.

(d) Modify the graphic created in part(c) to make the points transparent enough that all of them can be seen.

(e) Add just a regression line for each group of subjects to the graphic created in part (d) (no confidence bands).

(f) Add the confidence bands to the graphic created in part (e) that are shaded the same color as the line and mostly transparent.

(g) Create a plot similar to the graphic created in part (e) with just one regression line for all of the subjects.