September 23rd, 2020

PBCAR Seminar 1 Take-Home Exercise

The purpose of this exercise is to get you thinking like a programmer and asking questions like a programmer. Accordingly, you can absolutely google the answers (as literally every single programmer today does when they run into trouble). We highly recommend using stackoverflow. We will take this exercise up at the start of our next seminar, where we will discuss topics like calculating descriptive statistics, visualizing data, and other cool things. See you then :)

Exercise Instructions:

1.) Download and install RStudio on your desktop/laptop - use this link to help

2.) Download the *pbcar_df_all.csv* file from the PBCAR official GitHub account - use this link to help

3.) Import the *pbcar_df_all.csv* file into RStudio and using what you have learned today (and online resources), apply the following transformations:

   a. Rename the 'cudit' and 'gad' columns to 'cudit_1' and 'gad_1'

   b. Arrange the columns in the following order from left to right:

      'ID', 'age_group', 'drinks_1', 'drinks_2', 'cudit_1', 'cudit_2', 'gad_1', 'gad_2'

   c. Sort the data frame by its ID column in ascending order (So the row where the entry in the ID column is 1 comes first, followed by the row where ID = 2, then 3 ... until the last row is the one where ID = 150)

When all is said and done, your output should look like the *pbcar_df_all_solved.csv* data set (also available on the GitHub)

IMPORTANT: There's no one/right way of applying the three transformations we specified above to the data. We anticipate that you will all generate different answers... and that is okay (and if you could not get an answer, then that is okay too). Learning R can be frustrating and difficult. Once you know it, you can exercise extreme control over your data and perform most (if not all) your analyses with a single software. Please be kind to yourself when doing this exercise.

P.S. If you feel brave, you can try to visualize some of the data in the *pbcar_df_all_solved.csv* too!