Assignment 4

Lacey Hartigan

due 2/23/21

First code chunk - set the file structure and open your libraries.

For this assignment, you'll need to open up, clean and save the following datasets, using the tools we've gone over in class. For each dataset, make sure that when you're done you have a nice, neatly labeled dataset that would be easy for you or another analyst to open and analyze. Save each dataset in an RData file using the names provided. Include a display of the "head" of each dataset after it's cleaned up. You need to turn in an .Rmd script file named 04-assignment_.Rmd that cleanly creates each dataset requested AND one knit file.

1. Read in the panel data for OECD countries. Clean it as needed and display the "head" of it. Save it as oecd.RData. link - http://www.wiley.com/legacy/wileychi/baltagi/supp/Gasoline.dat

2. Read in data from "Statistical Discrimination or Prejudice? Large Sample Field Experiment. Use the mainData.csv file. Clean it as needed and display its"head." Save it as maindata.RData. link - https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/26410&studyListingIndex=1_f1a615c2cbee97fb79c9d575ae73

3. Read in the Lalonde dataset, covering work experiences in the Panel Study of Income Dynamics (psid). Use the psid_controls.txt file. Clean it as needed and dispaly its "head." Save it as psid.RData. link - http://users.nber.org/~rdehejia/nswdata2.html

4. A dataset of your choosing from kaggle (https://www.kaggle.com/datasets). Note, the first time you use kaggle, you will need to register (for free). Read in a dataset that looks interesting to you and give me a few sentences about what the dataset contains. As with the datasets above, clean the data as needed, dispaly its "head," and save it as an Rdata file.