CMU MSP 36601   HW2: EDA   Due: 10PM Wednesday Sept. 13

Download "nhanes2003.csv" from Canvas. This file has data from the National Health and Nutrition Survey for 2003. You should read about each variable in the on-line documentation. The variables seqn, riagendr, ridageyr, ridreth2, and dmdeduc are described in https://wwwn.cdc.gov/Nchs/Nhanes/2003-2004/DEMO_C.htm. The variables cvxparc and cvdvomax are described in https://wwwn.cdc.gov/Nchs/Nhanes/2003-2004/CVX_C.htm. The variables bpxsy1 and bpxdi1 are described in https://wwwn.cdc.gov/Nchs/Nhanes/2003-2004/BPX_C.htm.

Choose three variables other than "seqn". Perform some non-graphical univariate, graphical univariate, non-graphical multivariate, and graphical multivariate analyses on your variables.

Upload to Canvas two files named nhanes.R (or nhanes.Rmd) and nhanes.pdf.

Requirements for uploading:

- Use exactly these file names.
- Do not add your name to the filenames.
- Do not zip the files together.

Requirements for R (or Rmd) file:

- The R (or Rmd) file should run on any computer using `source("nhanes.R")` (or knit). This means that there is no setwd() command in your file. Test your file on a clear workspace. If you use extra packages use the "if (require)" syntax from the EDA1 handout.
- The first several lines of the file should be comments containing your name, the date and name of the assignment, and the setup (i.e., nhanes2003.csv is in the working directory).
- If you use .R, you must use `print()` around commands for graphical output; otherwise the `source()` command will produce nothing.
- Follow good style, e.g., meaningful variable names, appropriate indenting, lines no longer than about 70 characters.

Requirements for the pdf file:

- Include your name at the top of the file.
- Write a one paragraph introduction describing the variables you are studying (normally this would also include the goal, but that is vague here).
- Include two or three examples of each of the four EDA types in four sections.
- Round the non-graphical output to no more than 3 significant figures.
- Use meaningful labels in graphs and tables, not the original, cryptic variable names.
- End each of the four sections with a clear statement of what you found.

It is OK to discuss general questions with classmates, but all of the work should be your own.

Feel free to ask Howard for help in person or by email.

This should take approximately 2-4 hours.