HW3- CMDA

Submit your script to Dropbox by Monday, September 22, 5pm. Name your script ("YourName_HW3.r"). Use the textbooks, class notes and scripts, as well as R help and internet resources. Ask questions in the Forum for your colleagues to answer. Work in teams, but submit your individual file with your work.

Part I

1) Load your project DATA.
2) Decide which variables are numeric and which ones are best set as factor.
3) Get numerical summaries for all your variables. Comment on the most common data issues, as listed in class.
4) Get visualizations. Use at least three types of graphs to describe both individual variables and relationship between variables. Describe further data issues that you might see or just general patterns.
5) Comment on and perform any needed treatment for missing values.
6) Comment on and perform any needed data transformations.
7) Save a new version of your data set as TransformedData.RData.
8) Update your project repository on Git with a new version of the R script and the new RData file. Include a comment of what you did in the description of the commit.

PART II

Python

- Install Enthought Canopy, by following the book instructions.
- https://store.enthought.com/#canopy-academic
- Choose the "full" installation that comes for free if you are in the academic field; make sure you sign up with your vt email address; that will give you access to the "full" version of Enthought Canopy.
- Follow the instructions at pages 8-9 of your textbook, for Windows and Mac respectively.
- We are using the Python 2.7 version.