Exploring Python

The intent of this homework is to give you some initial exposure into the Python programming language.

1. Assuming that you are in the lab or have a Python distribution installed (Anaconda or Python(x,y)) on a Windows machine:

Start → command:
spyder (run python with an IDE)
[ipython (select ipython notebook server)
This command will allow you to run python notebooks.]

2. Read Python: A Tool for the Practical Data Scientist

3. Follow the official Python tutorial. You can also skim through Learn Python the Hard Way (LPtHW) or use it as a reference guide. The Python bootcamp will give you an introduction to Python. Note that we are using a slightly different setup than described in LPtHW, namely a terminal and Anaconda python. However, the suggested notepad applications (notepad++, textpad and textwrangler) are both applicable and can be installed unless you prefer an alternate text editor (emacs, vim, sublime, nano, etc) or a more complex development environment like eclipse or visual studio.

4. **Installing Python**- We will be using the Anaconda python distribution. Look at this quick guide.

One of the primary criticisms of python are related to difficulties installing all the libraries that make python such a powerful tool. Anaconda encapsulates many libraries in a single install. If you want to start programming python in an interactive environment, type ipython in your terminal to bring up a special, very powerful python terminal. Try this command:
print 'hello world!'

A very simple python program! If you're eager to learn more, check out learnpython.org.

Problem:

For this homework, as well as for every programming homework, you should submit the following files: report, python code, log and README file that solve this problem. This homework is based on the file income1.data which is described in the file income1.info.

1. Indicate the number of lines in the file.
2. Indicate the number of lines in the file after eliminating those lines that have fields characterized by unavailable (NA) data.
3. Indicate the most common education level (the fifth column corresponds to education level).
4. Indicate the level of income for households with some graduate school.