

## **Homework #1: Exploring Data through Visualization (100 points)**

In this homework, your task is to use Python's matplotlib module to *visualize* and *explore* the provided “Census Income” data set and *identify at least three interesting trends* within it. Specifically, you must submit the following **two** items:

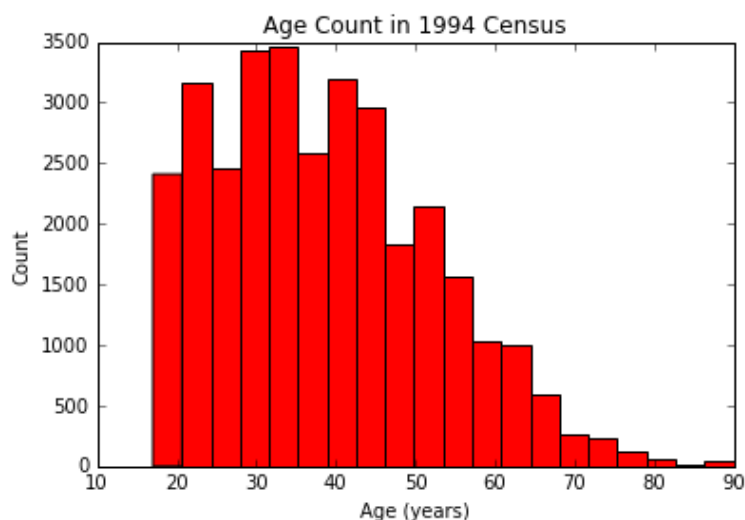
1. A *plain-English* document (with figures) explaining the 3 discovered **trends**. E.g.,
  - What trends did you look for?
  - What trends did you find?
  - *How* did you find them?
2. All Python code used to *process* the provided data and *generate* all figures.
  - Python 2.7 is preferred
  - Please comment your code judiciously

### **NOTES:**

- **adult.names** contains information about the dataset
- **adult.data** contains the actual comma-separated data
- Download and install the **Anaconda** package for Python<sup>1</sup>
  - Anaconda has everything you need to get started.

### **EXAMPLE:**

- One (not so) interesting trend within this data set is the presence of the “baby boom” generation.
  - In 1994, the “baby boom” generation (born between 1946 and 1964) would range between 30 and 48 in age.
  - A histogram of participant ages shows the presence of the “baby boom” generation.



<sup>1</sup> <https://www.continuum.io/downloads>