

STATS 607

ASSIGNMENT 1

The FBI has gathered crime statistics from law enforcement agencies across the nation that have voluntarily participated in the Uniform Crime Reporting (UCR) Program since 1930. The UCR Program collects statistics on violent crime (murder and nonnegligent manslaughter, rape, robbery, and aggravated assault) and property crime (burglary, larceny-theft, and motor vehicle theft) (see more [here](#)).

I've collected the last year (2014) of available crime data from <http://www.ucrdatatool.gov/Search/Crime/State/RunCrimeOneYearofData.cfm> ('crime.csv' and 'crimeRates.csv' files) and converted each into a list of lists. All table columns (you can generate the table from the link above) except 'Legacy Rape' is included in the data. As you will see, each list within the main list has state info or total US info (last inside list). In other words, each inside list is one row of the table. Your first mission, case you choose to accept it, is to define a set of functions to validate and collect relevant summary statistics on this data. **You have until September 26 to upload your solutions via canvas.** Otherwise, it will be the end of the world as you know it.

Make sure you have the three provided files ('crime.csv', 'crimeRates.csv', 'assignment1_Data.py') for the assignment in your current working directory before starting to solve the problem. All functions should be defined in a module (file) named 'assignment1_[your name].py'. For example, my module is called 'assignment1_Marcio.py'. Your main script should be called 'test_assignment1_[your name].py'. Your script should import the module and properly call all functions so to demonstrate outcomes. Make sure your calls to all functions in the main script demonstrate the expected behavior, so you may have to call the same function more than once. You can make use of additional imports in your module file (you'll need them), but you should only use core or standard library Python modules. Note that your final solution should consist of only two .py files: one is the main script and the other is the module you created with the definition of your functions.

Good Luck!

Questions:

- 1.1 Create a script with name 'test_assignment1_[your name]' and import the assignment1_Data module 'as a1Data'. Then assign a variable 'crime' to the output of the function 'get_US_crime'.

- 1.2 Do the lists inside the crimes list have the same number of elements? Define a function 'equal_length' that outputs a Boolean.
- 1.3 Create one list of strings with the name of the states included in the crimes list. Note that the crimes list contains US totals. Define a function 'get_states' that outputs a list with the name of the states.
- 1.4 Show that the total number of violent crimes is equal to the 'Violent crime total' column and that the total number of property crimes is equal to the 'Property crime total'. Define functions 'equal_vc' and 'equal_pc' for the purpose. Both functions should return a Boolean.
- 1.5 Show that US total values (last element of the list) correspond to the sum of reported values for all states. Define a function 'equal_total' that returns a Boolean.
- 1.6 Create an identical list to crimes with the crime rate per 100,000 population. Define a function 'get_crime_rate' that returns a list of lists with the new values (rounded to one single decimal place) and assign the outcome to a variable named 'crimeRates'.
- 1.7 Assign a variable 'crimeRatesOriginal' to the output of the function 'get_US_crime_rates'. Show that your 'crimeRate' is identical to 'crimeRatesOriginal' by comparing n random sampled values. For the purpose, you should define a function named 'equal_rates' with at least one input, the number of comparisons 'n'. This function should return a tuple with three elements: the two lists with the compared values and a Boolean stating whether the lists have the same value.
- 1.8 Create a dictionary of the top 5 states with the highest violent crime rate. For the purpose, create a function named 'top5_violent_states' that returns the dictionary. The key and value pair of the dictionary should be the name of the state and the violent crime rate, respectively.
- 1.9 Expand the function in 1.8 to return a dictionary of the top n states with the highest given crime rate (you'll specify the crime using an index). This function should be named 'top_crime_states' and at least have as input 'n' and 'indexCrime'. 'indexCrime' should be within the range of possible values, otherwise, the function should return a warning message.
- 1.10 Calculate the average number, standard deviation and variance of a given crime rate in the US (you'll specify the crime using an index). Define a function 'crime_stats' that returns the three values for the given crime. As with 1.9, this function should verify if the index is proper.