Intermediate Python **Homework 2**
Due: 11:59 PM Sunday, November 15 **This is an individual assignment.**

1. Download the file 'pop.csv' for this assignment (it's the same file from Homework 1). Note: Puerto Rico and the District of Columbia should be considered states for this assignment.

Read the file 'pop.csv' into a pandas DataFrame named pop. For the problems below, you should use pandas functionality, not regular Python, to do the data processing – that is, don't write a for loop to do the job. Recall the sort_values( ) function:

<dataframe>.sort_values(by=[<columnName>]) # Sorts on a column's values

Write a Python ***program*** named hw21.py that presents a menu of options and prompts the user for a numeric choice until they choose to quit. Put the menu code into a function named 'menu'; it should return a valid int menu choice, 0-6 (don't worry about bad conversions to int; assume the user enters some int). If the use does not enter 0-6, display an error message and prompt again. The main program should open the file, then loop until the menu( ) function returns 0. For each choice, print an appropriate label with the data display. The menu choices are:

1. Display the entire table
2. Display the total population of all the states.
3. Prompt for the name of a state. Display its population.
4. Display the table sorted by state name
5. Display the table grouped by region
6. Display the table sorted by population, largest to smallest
0. Quit

You should do this incrementally – don't do the entire menu at once – build up to the solution. Also, remember that for a program (instead of a script), *all* code must be inside a function. The main( ) function is the one that runs first. You need this code at the end of your file for it to work correctly:

```
if __name__ == '__main__':
    main()
```

2. Download the file 'nstData.csv'. It is a superset of the data contained in pop.csv in problem 1. Write a script (not a full program) that does the following; name it hw22.py. Read the file into a DataFrame named allPop.

a. Display the first two lines of allPop. Display the last five lines of allPop. (When you do this in the interpreter, you won't see all the columns – no problem).

b. Display all the statistics for allPop using describe(). Then display all the statistics but just for the column ESTIMATESBASE2010. Then display just the mean from that column.

c. Create a Series named states of the state names from allPop and display it. Delete the values that are regions, not states (the first 5 rows) by slicing and re-assigning. Display it.

d. Redo part c by creating states2 from allPop without the rows whose state number is 0. Display it. Display states == states2.

e. Create a DateFrame called myPop from allPop that is exactly the same as pop in problem 1; use the POPESTIMATE2010 field for the population. This may take multiple steps. You'll need to:

- capture just the relevant 4 columns of allPop
- remove the non-state regions (like in part c)
- rename the population column as 'POPULATION'
- change the Puerto Rico's region field to '5'
- change the 'REGION' field's type to int. Do this: myPop['REGION'] = pd.to_numeric(myPop['REGION']). This seems to be the most straightforward way to do this; there's no simple way to use int( ), unless you write your own for loop – which you can try if you'd like
- change myPop's index to 0-51instead of 5-56. Do this: myPop.reset_index(drop=True,inplace=True). This drops the old index; by default, you get the range for the new index.

Display myPop. Then show that it really is the same as pop, which you'll have to read as you did in problem 1.