SOC 4650/5650: Lab-02

Christopher Prener, Ph.D.

*January 24*th*, 2017*

Directions

Please complete all steps below. Your final log-file and markdown file with answers should be uploaded to your GitHub assignment repository by 4:20pm on Tuesday, January 31st, 2017. This lab requires data distributed at the beginning of class on Tuesday, January 24th, 2017. A copy of these data are also available in the course data release.

Setting Atom Up

1. In Atom, go to File ▷ Settings ▷ Install. Search for the package `language-stata` and install it.

2. Once the installation is complete, use Atom to open the file from your `week-02` directory named `atomSnippets.cson`.

3. Go to File ▷ Snippets... . Copy and paste the contents of `atomSnippets.cson` into the `Snippets` file that opens. Save that file, and close it.

4. Fully exit Atom.

Creating a Do-file

5. Reopen Atom, create a new file, and change its language to "Stata".

6. Expand the snippet `headFull`.

7. *Without clicking anywhere*, begin using the Tab key to move through the fields that you are prompted to complete.[1] Fill them out as best you can. Your project name should be `listedLakes`. The raw data are named `MO_HYRDO_ImpairedLakes.csv`. Review your answers with your partner, and work together to improve your responses to these fields.

8. Save your do-file as `listedLakes.do`. Your do files should always be identical to your project name.

[1] If you lose the Tab functionality, you need to edit lines 3, 9, 44, 46, 47, 50, and 70. Note that these line numbers may change if you enter multiple lines of text in, for example, the description area. Look for `/*prompt*/` and replace all of that with your own text. Do not leave the `/*` and `*/` behind!

Working with Stata and Atom

9. In Stata, set the working directory to the `week-02` directory on your `Desktop`.

10. Create a plan with your partner for the commands you will need to answer each of the sub-questions under question 11.

11. Test the commands for the following tasks in Stata, and then copy the *working* commands into your Atom do-file beginning under the "Import/Open Data" header. When you copy the command in, include some narrative text below it that begins with the appropriate question number like so:[2]

```
/***
**11a.** This is the answer to question 11a.
***/
```

> [2] Look at the file `stlSchools.do`, included in the `week-02` directory, to see how it is formatted.

Your narrative text should explain what the command accomplished and also provide an answer to the question prompt.

(a) Import the raw data. Instead of specifying the file name, the import command should look like this:[3]

```
import delimited 'rawData', varnames(1)
```

> [3] This command works because, when you filled out the fields in question 7, you entered the filename of the listed lakes data and saved in the object (or, in Stata-speak, the local macro) named `rawData`. This improves the reproducibility of your work because it limits the number of times you have to enter the same filename in. If there is a change in the filename, you can make that update once and it will be applied to any area of your code where the object `rawData` is listed.

(b) Create a table that lists all the variables in the dataset. How many variables are stored as "string" data?

(c) Calculate descriptive statistics for the variable `yr`, which stores the year in which the body of water was first listed. What is the average year? What are the range of years where lakes have been listed?

(d) Create a frequency table of the variable `water_body`, which lists the individual lakes in the dataset. Which lake has the most entries currently?

(e) Create a frequency table of the variable `source_`, which lists the general source of pollutants. Which source is the most common?

(f) List the first ten observations for the variables `yr`, `water_body`, and `pollutant`. For "Bowling Green Lake", what are the three types of pollutants listed?

12. Save your do-file in Atom, and close the application. Open the do-file in Stata and execute it to create your final directories that contain a copy of your code as well as your log file, raw data, imported data, and your Markdown output. Your log file and Markdown output should be submitted into the `Lab-02` directory in your assignments repository.