**Lab 0: Geographic data 101***Geog2011: Introduction to GIScience*

**Introduction**

This mini-lab gets you started working with geographic data on climate across the continental United States. You’ll be working with climate data in Excel, understanding the variables you’re working with and doing some basic analysis. You’ll also take a look at a shapefile version of that data and understand how the file is put together. Lastly, you’ll develop a workplan for your data over the semester. The primary outcomes of this lab should be an improved ability to describe a spatial dataset we’re using and a better understanding of how geographic data is put together.

**Downloading Data**

You’ll need two data files for this lab, all available in the Lab 0 folder on ELC

* Daymet\_2015.xls provides numerical data on climate for over 3,000 U.S. counties.
* Daymet\_2015\_map.zip is a zipped folder that contains a shapefile, which is the most common format for geographic data. In addition to the climate data in the Excel spreadsheet, it also contains information on boundaries for each county in the U.S., which allows the data to be mapped.

Download these files. You should put them in a working folder that allows you to keep all project materials together. There’s lots of ways to do this--an online backup tool like Dropbox or a flash drive copied to a laptop are two examples. Whatever you choose, it should match the following criteria:

* **Self-contained**: Should contain all materials related to your project
* **Secure**: It should be in a place no one else can access it without your permission
* **Backed up**: You should be able to access these data even if your primary storage method fails (e.g., laptop gets lost, flash drive falls in water, etc.)
* **Easy access:** It should provide a convenient way for you to work on your project at a time that works for you.

You’ll have to summarise this at the end of this assignment. Save the files to your working folder once you’ve figured out a system that works for you.

**Inspecting the data of the data**

Open the “Daymet\_2015.xls” file in Excel (or a similar program--Google Sheets or Excel online can also work). There’s two tabs on this spreadsheet--one with the data and one with the metadata. The latter includes variable code names and descriptions and a link for more information. Take a look at the metadata to see what variables are included in this dataset. You’ll need this for the questions at the end of the lab.

Each county in the dataset already has its region listed (the *Region* variable), so we can use Excel to filter by this variable to calculate statistics for each row using the Filter tool on the data tab. See [this tutorial](https://support.office.com/en-us/article/filter-data-in-a-range-or-table-01832226-31b5-4568-8806-38c37dcc180e) from Microsoft for an example of how to do this.

For this short assignment, filter the counties so that you just see those in the South Atlantic Division. Select and copy those over to a ***new spreadsheet.*** (This is important--the formula below won’t work right if you apply it to a spreadsheet with a filter.)

**Summarizing the data**

For this lab, you’ll also need to use Excel to calculate basic descriptive statistics with these data. You can do so using formulas. ([Here’s a link](https://support.office.com/en-us/article/overview-of-formulas-in-excel-ecfdc708-9162-49e8-b993-c311f47ca173) to Microsoft’s help page on using formulas.)

For example, to calculate the mean value for maximum temperature (column F), you’d type this into an empty cell:

* =AVERAGE(F:F)

Calculate the mean, median (MEDIAN is the function), and standard deviation (STDEV is the function) for max. temperature (column F) and precipitation (column I).

**Opening the geographic data**

We will work with the shapefile version of this data more in Lab 1, but for this lab just unzip the file and take a look at what’s inside. The easiest way to do this on both Windows and Macs is to double click the folder. Open another window with your working folder and drag the files in this zipped folder there. Take a look at these files, which together are ***one*** shapefile. Then do some googling to try to understand the different file types. You’ll write more about that below.

**Lab deliverables**

*Note: Please save your files as: LastName\_Lab0.doc*

In a Word compatible document, provide the following:

1. (2 points) What are the mean, median, and standard deviation for the max temperature and precipitation variables **in the South Atlantic Division**?
2. (2 points) Which county had the ***lowest maximum temperature*** in the South Atlantic Division during this time period? How many standard deviations below the mean was this county? To answer this question, subtract the temperature from the mean value and divide by the standard deviation. (This is called a “[Z-score](https://www.khanacademy.org/math/ap-statistics/density-curves-normal-distribution-ap/measuring-position/v/z-score-introduction)”).
3. (1 point) Using the filter and/or sort tool, identify a county from the South Atlantic Division that’s ***more than two standard deviations above the mean*** ***for precipitation***. Give the county and state name for that county. What was its precipitation?
4. (2 points) Do some internet sleuthing and figure out the purpose of the shp, prj, and dbf files in the shapefile you have. Summarise them *in your own words* and provide a link to the resource you used to figure this out.
5. (2 points) In class we will talk about vector data and raster data. Which one is the county data in this shapefile? If it’s vector data, what ***kind*** of vector data would it be--points, lines, or polygons? Give a justification for your answers.
6. (1 point) Briefly summarise your data management plan for the semester and how it satisfies the four criteria listed above.

Submit this document to ELC once you’re finished.