**Assignment 1: Introduction to Spatial Analysis in Python**

Due February 5 (add/drop deadline!)

## Part 1

This assignment will take a significant amount of time to make it through this and truly understand/ingest it (I anticipate 10-20 hours for those who don’t have experience in python). Those who are unable to complete this assignment will be unable to continue the course. Please start early and dedicate a lot of time to this.

NOTE: if you have significant python experience, set up a time to discuss with me. I may ask you to share and explain some of your python code to me. If I feel that your experience is sufficient, I may waive part or all of Part 1.

Go through the following tutorial <https://courses.spatialthoughts.com/python-foundation.html>. After each exercise, take a screenshot of the code you created and the output produced. After each jupyter notebook, save the notebook to include with your submission.

You will submit a word file with screenshots for each of the sections below. You will also submit a zipped file with all of the jupyter notebooks that you have modified for the sections below.

* Variables
* Data Structures
* String Operations
* Loops and Conditionals
* Functions
* The Python Standard Library
* Third-party Modules
* Using Web APIs
* Assignment
* Reading Files
* Reading CSV Files
* Working with Pandas
* Working with GeoPandas
* Creating Spatial Data
* Introduction to NumPy
* Working with RasterIO
* Writing Standalone Python Scripts

Describe the following:

* What is the difference between a shapefile and raster in geospatial analysis? Give three examples of datasets for each (i.e. three datasets that are rasters, and three datasets that are shapefiles).
* What is a coordinate reference system? Write a sample line of code you could use to change coordinate reference systems, using python, for both a raster and vector dataset.

## Part 2

Your assignment is to create a time series using remotely sensed data. PRISM data from 1890 to 2024 are located on Canvas. Download all of these datasets.

Open a command window and activate your anaconda environment. Then run the following to install rasterstats:

conda install -c conda-forge rasterstats

Open a python editor such as jupyterlabs or spyder. In the python editor, write code that will do the following:

* Read a HUC8 watershed using geopandas
* Project the HUC8 watershed to the same coordinate reference system as the PRISM rasters
* Create two empty arrays. Each should have a length equal to the number of years that you have PRISM data; one for temperature, one for precipitation
* Create an array of the same length as the two empty arrays whose values represent each year that PRISM data is available, i.e. [1890, 1891, 1892, etc]
* Create a for loop that iterates through all years and does the following:
  + Define a variable that contains the exact filename of a PRISM raster (.bil extension) for the given year
  + Use the zonalstats function within rasterstats to calculate the average temperature within your watershed
    - Before starting the for loop, you should have imported the function needed using something like this: from rasterstats import zonal\_stats
    - Within the for loop, you will use something like this:
      * precip\_1yr = zonal\_stats(shp,prism\_fname) where shp is the projected shapefile with your watershed, and prism\_fname is the filename of the prism raster you want to calculate statistics of
      * *Hint:* I find it easier to work with the output of zonal\_stats by converting it to a dataframe, i.e. precip\_1yr\_df = pd.DataFrame(precip\_1yr)
  + Record this value to your temperature array
  + Use the zonalstats function within rasterstats to calculate the average precipitation within your watershed
  + Record this value to your precipitation array
* Plot the average annual precipitation and temperature data over time in two separate plots using matplotlib