Joseph Moravitz

Georgia Tech Data Analytics Boot Camp

Homework 6 – Python API’s

For this assignment we were given the task of attempting to prove and find evidence supporting the claim “It really does get hotter the closer you get to the equator.” To accomplish this, we were instructed to randomly select cities from around the world and request data from an API available to us about the current temperature for each of our chosen cities. Armed with this information, we then can create ourselves some scatterplots and run analysis on the data in order to see if there is any standing to the claim.

The first part of the task, randomly choosing the cities, turned out to be one of the more difficult steps for me. The starter code provided did it for most of my classmates without any thought required whatsoever however I was not satisfied with how the starter code runs for a number of reasons. I took it upon myself to heavily rewrite the code used to generate our list of cities.

Originally, the starter code would attempt to select an (x,y) point from a rectangular grid uniformly at random. Interpreting the grid as a map of the world these (x,y) points would be treated as longitude and latitude respectively. This, I feel, is a mistake as it would be just as likely to pick a point from a small grid square close to the pole as it is to pick a point from a small grid square near the equator since the squares are the same area when viewed as a part of the rectangular map. On the globe however, the area described by the grid square near the pole is much smaller than that of the grid square near the equator, giving an unfair bias away from the equator and towards the poles.

To remedy this, I opted to select points on the globe uniformly at random through the use of some multivariate distributions. Letting X,Y,Z each be independently normally distributed random variables, N(0,1), and treating the globe as the unit sphere (yes, the globe is technically a spheroid, not a sphere, this is a simplification), we have that (X,Y,Z) is equally likely to be pointing in any direction. Normalizing the vector (X,Y,Z) by dividing by the vector’s norm gives us a uniformly randomly selected point from the sphere without any bias towards the poles. I then converted these (x,y,z) points back to latitude,longitude coordinates for ease of use for the rest of the code.

The starter code would then take our randomly selected points and utilize a recommended package, Citipy, to determine what the closest city to our point is. Again, I feel this is an oversimplification that provides unfair bias to certain regions. The surface of the earth is mostly water, oceans taking up a considerable portion of the area. Every time a point we selected was from the ocean, Citipy would tell us the closest city would either be an island city or a coastal city.

To remedy this, I opted to ask “how far away is the ‘nearest city’ to my originally selected point?” This required looking up where the city that Citipy directed me towards is actually located and finding the distance between those points (which is easiest done using (x,y,z) coordinates rather than latitude-longitude coordinates requiring more conversions). If the distance was too far (currently set at 2% the radius of the earth in my code) then that would imply the point was too far in the middle of the ocean or in the middle of a desert or tundra to be of use and I would scrap that point to try to get another to fill it’s place.

Now, finally armed with a satisfactory list of randomly selected cities, I was able to query the API at openweathermap.org to find data about the current weather in each of those locations. The story the data shows is that it really does get warmer as you get closer to the equator… but the equator itself is not where the peak temperatures seem to occur most frequently. Of course, this should be expected. The earth spins on a tilt and depending on the season either the northern hemisphere or the southern hemisphere receives more sunlight on average. As it is still early fall, riding on the tail of summer, the northern hemisphere is still a bit warmer than the southern hemisphere. The scatterplot supports this with the peak appearing to occur somewhere around 10 degrees latitude north.

The data is still quite spread out for various latitudes. Of course, we queried for the current temperature, so cities which were in the middle of the night will have reported lower than usual temperatures while cities which were in the middle of the day would have reported higher than usual temperatures. Further, weather, cloudiness, and elevation play a big role in current temperatures as well. There was even a city around the 30 degrees north mark which reported freezing temperatures, likely from the mountain ranges.

Humidity and cloud cover both seemed to be spread evenly across most of the latitudes, however closer to the equator (or more accurately, closer to 10 degrees north) there is an empty area of the scatter plot, suggesting that the climate as it gets closer to the equator is generally less dry and more humid. It is difficult to see very well given the way the scatter plots are presented.