Logan Field

5/14/20

Final Project

Big Data

Section 1

* Nighttime light data
  + Clean data
  + Run general stats
* Urban Heat Island
  + Clean data
  + Run general stats

The average daytime high temp across all values is 36.65 Degrees Celsius

The highest value is 62.51 Degrees C which is over 140 degrees Fahrenheit and the low is 2.26 Degrees C

* Visualize

A star in the dark

Description automatically generated

* Discuss pattern

This is the view over the US of the UHI over laid onto NTL map. The values are very close, but the NTL data stretches out farther than the UHI data in some places such as Texas and near the Canadian border. These results may be due to various things such as neighborhoods that sprawl outside the city that do not produce as much heat. To the south of San Antonio and Austin, Texas, there is a band of lights stretching across the landscape but no cities in site. This pattern may be due to the large amounts of oil drills and pumps there are in this region. They may have lots of lights illuminating them for whatever reason causing the peak in NTL data but not UHI data.

Section 2

A close up of a map

Description automatically generated

* Use Azure ML to compare data

A screenshot of a cell phone

Description automatically generated

Section 3

* Discuss goal

The goal of this was to compare Nighttime light data with Urban Heat Island data and find the similarities and differences. Through file cleaning and comparing the two data files to each other, this project hopes to identify the where the data matches and more importantly where there are differences and if there is a pattern.

* Data description

Nighttime light data (NTL) is data of the intensity of light being emitted from the earth whether that be man made or natural light. This data is a good way to see industrialized areas where lights are on at night. In some regions, this data can be used as an indicator of population, urban sprawl, and many other indicators. Poorer or less industrialized places may be misrepresented in this data set as they might not have access to electricity or lighting. An example of this is in Africa where there may be massive cities, but many of its residents do not have access to this luxury so its NTL values are far less compared to cities of the same size in Europe or China for example.

Urban Heat Island (UHI) data monitors the varying heat levels in urban environments. A UHI occurs when a city or urban environment have many structures and people causing an increase in temperature compared to the less urban areas around it. For example, Denver Metro is a highly dense area with lots of roads, buildings, and people in it. The roads and buildings decrease the albedo (the amount of sun reflected back off a surface), which increases the temp. Humans doing daily activities such as driving and working also increases the temperature. Once you move to the west or east of the City, the infrastructure decreases dramatically, the albedo rises, and there are fewer people leading to lower temperatures.

* Steps preformed

1. The first step was locating both the NTL and UHI data and converting them into forms that can be uploaded to Azure notebooks and Machine Learning programs.
2. Next, I cleaned the data in python to remove the empty values and the black values in the nighttime light data.
3. I imported the data sets back into python to compare the values and where they match in location.
4. Finally, I compared the two data sets in Azure ML Using a linear regression model and comparing the error results.

* Conclusions and recommendations
  + Insights gained

By comparing these two data sets, it becomes easy to see similarities in where the light is and where urban heat islands exists. Using python to compare the data, it becomes much easier to find out how similar they actually are. Being able to use this method has been very interesting and taught me a lot. The biggest thing that held me back was a small oversight of converting the files into the right format using GIS programs and not any coding or 7zip functions on my end. If I did this project again, I would like to clean up the code and find more effective ways of linking the files into the code. I would also like to compare the data with other things like GDP.