



Geospatial Data Specialist Exercise

Summary

This is a short assignment to assess your geospatial data-wrangling skills. You will be given a set of data inputs and are asked to transform this data into new data products. Use scripting (python, CLI, notebooks, etc.) as much as possible, and make sure to **document every step** of your work. Reproducibility is an essential part of this role.

Data Sources

Available [here](https://drive.google.com/file/d/1VGUNksqVAd1iGCvexBRbgPz18Yj1bdY/) (https://drive.google.com/file/d/1VGUNksqVAd1iGCvexBRbgPz18Yj1bdY/).

- A. **pvsystems_3857.gpkg**: a point layer (in GeoPackage format) identifying the (fictitious) location of solar panels in Massachusetts.
- B. **ma.geojson**: a polygon layer (in GeoJSON format) for the boundary of the state.
- C. **ma_zipcodes.shp**: a polygon layer (in shapefile format) delimiting all zip codes in MA.
- D. **ghi_jan.tif** and **ghi_jul.tif**: two raster layers (in GeoTIFF format) providing the January and July average global horizontal irradiance (GHI), in watts per square meters, at an approximate 0.04-degree latitude by 0.04-degree longitude resolution.
- E. **charlestown.laz**: a LiDAR point cloud, in LAZ format, of the Charlestown zip code (02129). To quickly visualize this file, go to plas.io and drag it into your browser window.
- F. **charlestown_buildings.geojson**: a polygon layer (in GeoJSON format) of building footprints for the Charlestown zip code.

⚠ Note: these datasets use different spatial reference systems!

Expected Outputs

Using the above datasets, please produce:

- 1. **zip_stats.csv**: a CSV file tabulating (1) the number of solar panels and (2) the average difference between the July and January GHI in every zip code.
- 2. **high_jan_ghi.geojson**: a polygon layer delimiting the areas within the state that have above-average GHI for January.
- 3. [BONUS] **charlestown.tif**: a [DSM](#) GeoTIFF at 2m X 2m resolution, derived from the LiDAR point cloud.
- 4. [BONUS] **charlestown_buildings_heights.csv**: use your results in Output 3 to make a CSV file tabulating the building height of each building found in dataset F.

Delivery

Please submit a zipped folder containing the data inputs you used, the resulting outputs, and any code, script or other documentation needed to reproduce and understand your work. Feel free to make any simplifying assumptions if you are blocked, but make sure to explain them.