Project 5. Working with Raster Data & Zonal Analysis

Abstract

This report will detail a raster analysis carried out using R. Large raster datasets such as a 30m DEM and the CONUS NLCD data will be used in this analysis.

1. Loading the Data

The data links provided by Dr. Shan for this project consisted of a 30m Digital Elevation Model (DEM) (raster), the CONUS National Land Cover (NLCD) data (raster), a counties shapefile, and a flood zones shapefile. The raster data was simply loaded into R using the raster packages raster function. The data was read in almost instantly despite the size given that raster reads from disk rather than memory. The shapefiles were read in as usual using st_read from the sf package.

2. Terrain Analysis

The next step in the analysis was to calculate the slope and map it as well as elevation. The elevation was already available in the raw DEM data in units of meters as seen in Figure 1a. Slope was calculated using the terrain function provided by raster in degrees as seen in Figure 1b. In order to develop an understanding of the relationship between the slope and elevation, the raster data for slope and elevation were converted to vectors in data frames. The vectorized slope (rather than a matrix) was sampled with a sample size of 250,000 points for both slope and elevation. The slope was plotted on the y-axis and the elevation on the x-axis for the sampled subset as seen in Figure 1c. No real relationship can be established.

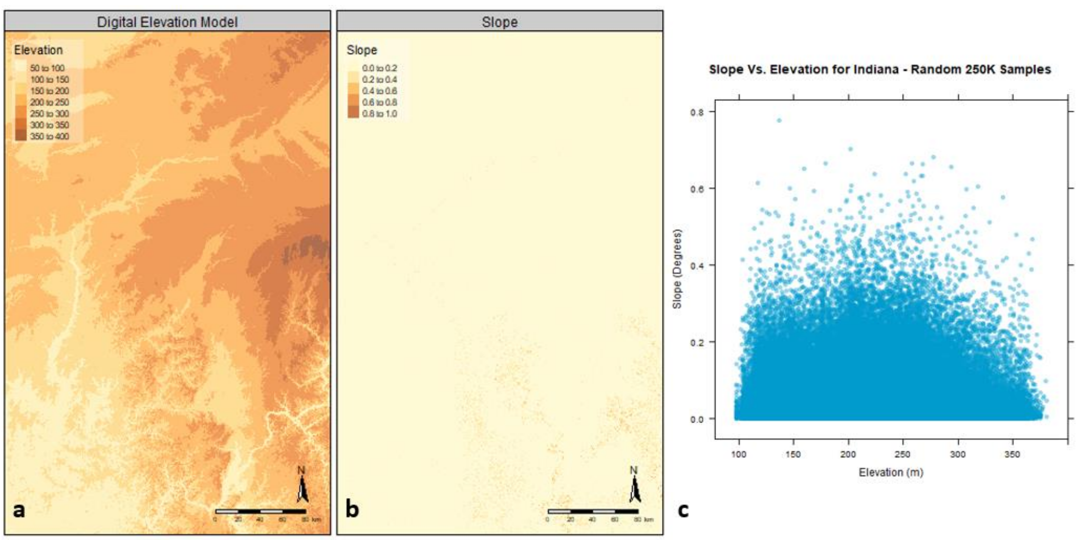

Figure 1: (a) Elevation in Meters. (b) Slope calculated in degrees. (c) Elevation vs Slope.

3. Zonal Analysis

To prepare the data sets for zonal analysis, all datasets were projected to the same projection as the DEM’s native projection, UTM Zone 16N, and cropped to the county data. The datum used was WGS84. This wasn’t a particularly difficult task but it was extremely slow. Given the NLCD data size, it was deemed unwise to project it and then crop it. The counties were instead projected to the projection provided by NLCD and then the NLCD data was cropped by this projected county data. After cropping the data, it was re-projected to UTM Zone 16N using the nearest neighbor method (due to it being categorical) and resolution set to 30 square meters in order to match it up with the DEM raster. Despite this attempt at speeding these operations up, projecting the subset of NLCD data still took roughly 40 minutes to complete. Given that the NLCD raster was a categorical representation, it was necessary to use the raster package’s ratify function on it in order to re-establish the factor levels. In order to display it with the NLCD’s official colors, a color picker tool in Google Chrome was used to lift the category colors from the web and were manually entered into a palette vector in R. The final NLCD figure was produced with a rasterVis levelplot. To prepare the flood data, it was re-projected to UTM Zone 16N and re-coded to use clearer codes for flood zones. The counties, NLCD, and flood data can be seen in Figure 2a, 2b, and 2c respectively.

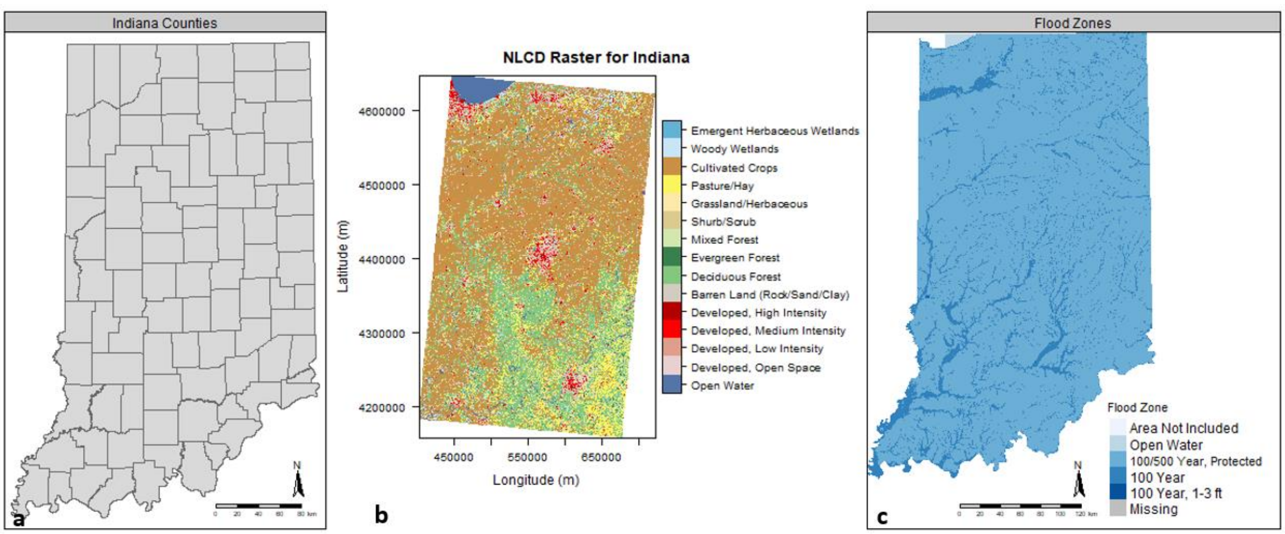

Figure 2: (a) Indiana county polygons. (b) NLCD data by classification. (c) Flood Zones by FEMA classifications. All data uses the WGS84 Datum and has been projected to the UTM Zone 16N projection.

With the data ready for zonal analysis, the package exactextractr was used to speedily extract the average and standard deviation for elevation and slope by county. The results can be seen in Figure 3a, 3b, 3c, and 3d. While elevation doesn't change much in Indiana, it is interesting to see the significant change in slopes toward the southern part of the state.

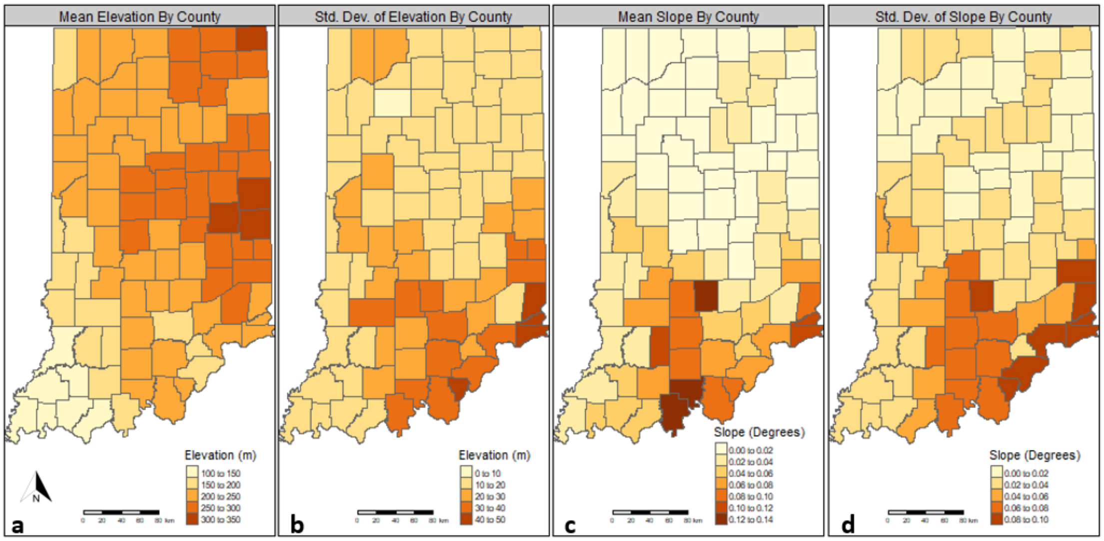

Figure 3: Mean and Standard Deviation for elevation and slope by Indiana County.

The next step in zonal analysis was to determine the mean and standard deviation of slope and elevation by land cover class. To accomplish this the dem and slope rasters were resampled to match the nlcd data. These are shown in Figure 4a-4d of the appendix. Finally, zonal statistics were carried out for the Flood zones. This proved to be somewhat trickier as there were more than five-thousand polygons. Class discussions proved to be fruitful and pointed to the use of the aggregate function to sum up the zones once the zonal statistics were calculated. The results can be found in Table 1.

Table 1: Zonal analysis of flood zones.

Flood Zone	Mean Elevation	Std. Dev. Elevation	Mean Slope	Std. Dev. Slope
Area Not Included	262.24	92.31	0.01	0.01
Open Water	176.38	16.29	0.00	0.00
100/500 Year, Protected	228.79	NA	0.03	NA
100 Year	227.17	NA	0.04	NA
100 Year, 1-3 ft	187.02	NA	0.01	NA

4. Conclusion

This project has shown that working with raster data can be quite time intensive and requires careful consideration for how to carry out raster analytics must be taken in order to arrive at a solution efficiently. Perhaps with more time, newer packages that leverage C/C++ code like exactextractr will become more normal in R for raster analytics and speed up these tasks.

References

DEM – Provided by Dr. Shan – 22 October 2019

Flood Rate Insurance Map (n.d.). Retrieved October 22, 2019, from https://maps.indiana.edu/previewMaps/Hydrology/Floodplains_FIRM.html.

IndianaMap. (n.d.). Layer Gallery. Retrieved October 22, 2019, from https://maps.indiana.edu/layerGallery.html?category=Census.

NLCD Data. (n.d.). Retrieved October 22, 2019, from https://www.mrlc.gov/data.

Appendix

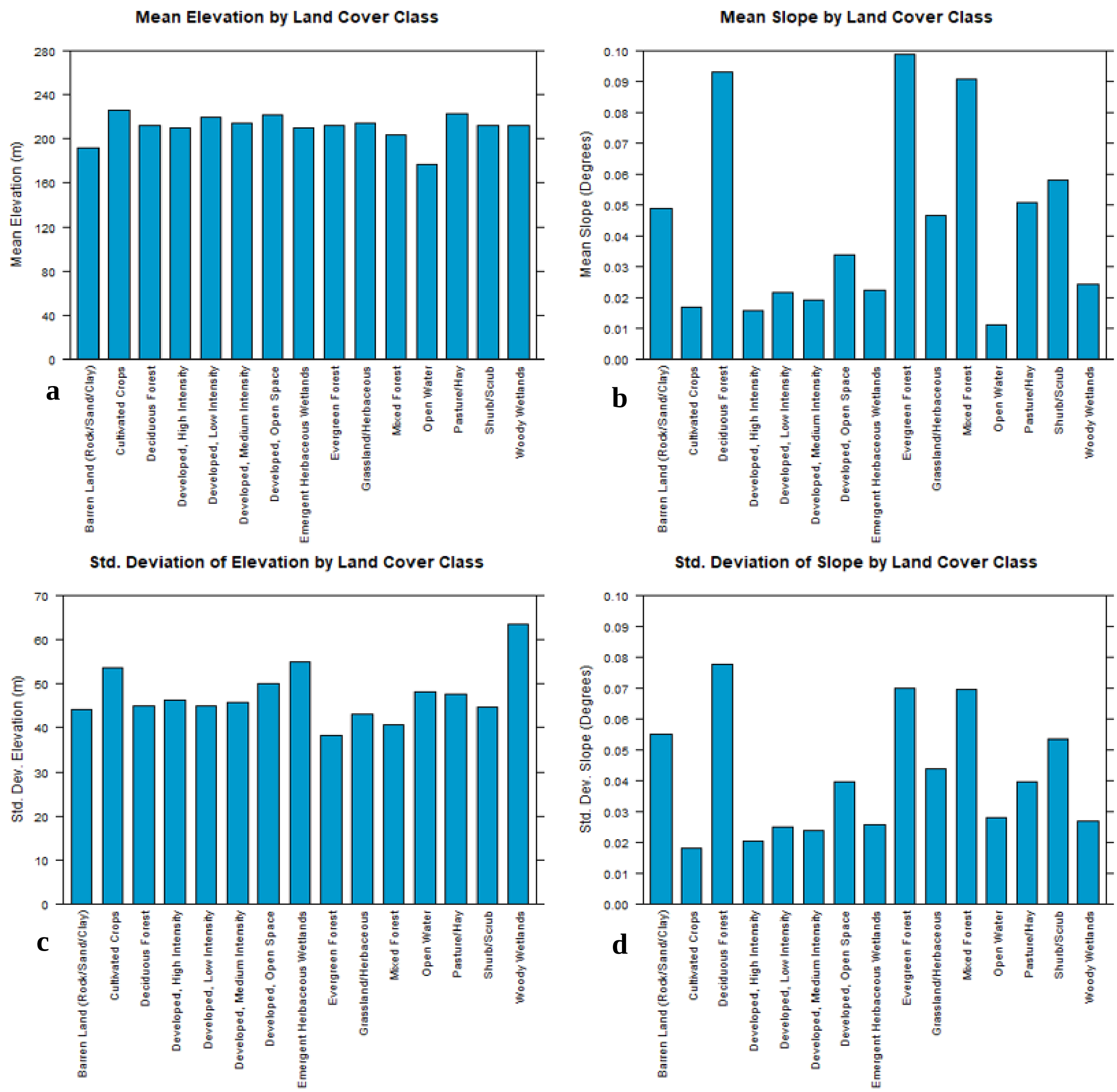

Figure 4: Zonal analysis results by land cover categories.