**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. Many difficulties were presented by the volume of the data.

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, which was counterintuitive given the size of both datasets. What was discovered by Isis Chagas during class presentations was that the raster data is not being brought into memory but is rather being indexed and read from disk. The raster data staying on disk during operations is nice but also significantly slows down computation. The shapefiles were read in as usual using st\_read from the sf package.

1. **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 and could quickly be viewed through with tmap INSERT FIGURE. Slope was calculated using the terrain function provided by raster in degrees. This operation was relatively quick and was also visualized with tmap INSERT FIGURE. 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 as a matrix) was sampled with a sample size of 250,000 points for both slope and elevation. With the help of lattice, the slope was plotted on the y-axis and the elevation on the x-axis for the sampled subset as seen in INSERT FIGURE.

1. **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 was an extremely slow portion of this analysis. 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 NLCD 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, as seen in INSERT FIGURE, was produced with the rasterVis levelplot.

1. **Summary/Conclusion/Concluding Remarks**

Texts..

**Acknowledgement**

**References**