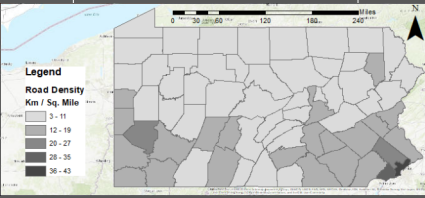# PENNSYLVANIAN ROAD DENSITY ERROR
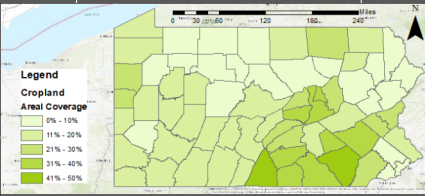
Assessing the Spatial Relationship of Road Density Factor Degrees
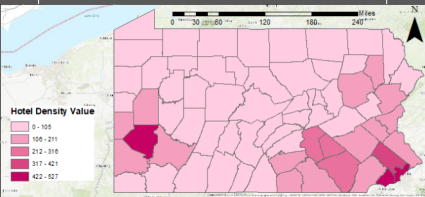
Road Density
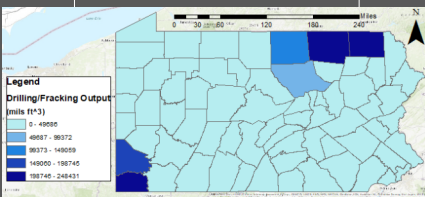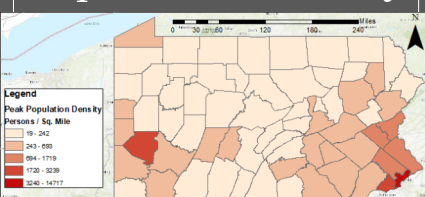
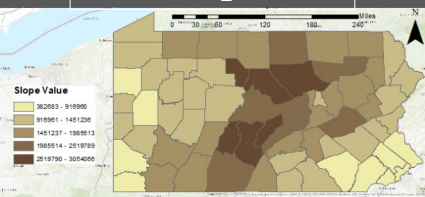

Cropland

Hotel Density

Oil & Gas

Population Density

Slope

Overview:

The purpose of this project is to see if there is a spatial relationship in the degrees to which certain factors affect an area's road density. In this project, data was collected on Pennsylvania's counties and used to determine correlations between identified factors and counties' road densities. Results were used to determine a formula that describes the average Pennsylvanian county's road density. Then, that formula was used to predict road density of other state's in order to see if there exists a spatial correlation in the percent error, which would indicate a spatial relationship in the degrees to which the identified factors affect an area's road density.

Data Collection & Methodology:

Data was collected on five factors: cropland (US Census Bureau), hotel density (ReferenceUSA), oil & gas extraction (USDA), population density (US Census Bureau), and slope (NED). The road network data raster file (US Census Bureau) was prepared using the map algebra and int tools. Then, the zonal statistics tool was applied to sum the road network along county and state lines and was spatially joined to an existing shapefile with census data including cropland using the raster to polygon, dissolve, and feature to point tools.
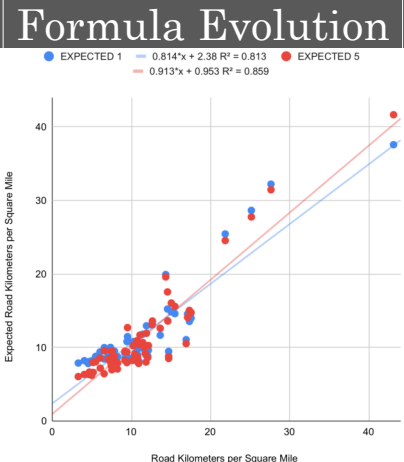
To attain slope data, the slope tool was applied to an elevation data raster file and the same process as above was followed. Additionally, county-level oil and gas extraction data was downloaded and added to a table.

Hotel data in each Pennsylvania county was downloaded and the kernel density tool was applied to create a hotel raster, and the same method from determining average slope was used to determine an average hotel density for each Pennsylvania county. For states, a simple number of hotels per unit area was applied to a table.

The data collected was used to find degrees of correlation and trendlines for each of these factors and road density values for each Pennsylvanian county. The trendline of the factor which had the highest correlation was made its trendline an expected value formula. Then, the absolute error from that expected value formula was correlated with each of the remaining factors, and the trendline from the most correlated factor was used to adjust the original expected value formula. This method was used until all factors were exhausted. The Formula Evolution chart shows how the formula became more accurate as more factors were considered.

Conclusion:

The spatial autocorrelation tool was run on the % error of the formula when applied to states and found a Z score of 5.3. Clearly the results are spatially correlated. To find out if they are correlated based on proximity Pennsylvania, the cluster and outlier analysis tool was run which found Pennsylvania to be included in a low error cluster of states with the high error cluster occurring in the Western half of the country. This means there exists a correlation in the distance to Pennsylvania and the degrees to which certain factors affect road density. If this project were repeated using other states, that data could possibly be used to create a spatial dataset for the affective degree of each road density factor. In order to understand why these results are occurring, research involving analyzing microdata such as local government archives would likely need to be conducted.

Formula Evolution

Pennsylvanian Formula Error