Assignment 1

Introduction to Spatial Data Science

Susan Paykin

10/24/2019

Introduction

This data represents one dimension of the agricultural landscape in the Hudson Valley region in New York State. The Hudson Valley is comprised of 10 counties that border the Hudson River Valley, located in the southern part of the state. I used publicly available data on county boundaries, agricultural districts, farmers markets, and population in those counties to explore and identify possible trends in the production and consumption of fresh food. I performed this analysis and created the maps using R.

Data Sources & Variables

I downloaded the New York Counties shapefile from the NYS GIS Clearing House. In addition to the county boundaries, this shapefile contained other variables of interest for this assignment: population and total area. I filtered this dataset down to the Hudson Valley counties, selected the variables I needed and saved this new variable. The New York State Agricultural Districts dataset was a shapefile from the Cornell University Geospatial Information Repository. This was a large dataset (36,926 observations) with spatial data for every agricultural district across the state. I filtered this dataset down to the Hudson Valley counties of interest for a total of 4,391 observations/districts, and saved this new variable. The Farmers' Market dataset was from the New York State Open Data Portal. This dataset was in a non-spatial .csv format, but it contained Latitude and Longitude variables so I was able to read it into R as a spatial points dataset. I also filtered this dataset down to the Hudson Valley counties of interest.

Mapping & Analysis

I transformed all the dataset projections into EPSG:32115 (New York East) to ensure they matched. Next, realizing that a simple count of the agricultural districts per county would not be a very meaningful comparison, I calculated the area of each agricultural district based on the Geometry of the spatial polygon. I used the *st_area()* function on the geography variable of the dataset to create a new variable of the area of each district in square meters (the standard measure of this projection) and converted it to square miles, as this is a more commonly used and understood area measurement in the United States.

In order to calculate the rates of density, I created new variables - density by population and density by area - by dividing each county's total area in square miles by its population and the total square miles of its agricultural districts. These are the variables I used for chloropleth maps comparing density (Figure 4).

I performed point-in-polygon calculations in order to obtain counts of agricultural districts and farmers markets in the counties. This required a spatial join using *st_join()* on the the Ag Districts and County datasets, then Farmers Markets and County datasets, by their shared *NAME* variables, then aggregated the count of the number of districts and markets per county. I conducted the same P-i-P analysis to obtain the total population and acreage of agricultural districts per county, both of which were used in my chloropleth mapping and analysis. See **Table 1** for complete aggregated variables.

```
## [1] ""
## [2] "Table 1: County Totals"
## [3] "=============================================================================="
## [4] "     Name     # Districts Districts (Acres) Districts (SqMi) # Markets Population"
## [5] "------------------------------------------------------------------------------"
## [6] "1   Albany        559        73,948.170         115.544         35       304204  "
```

```
##  [7] "2  Columbia       97        182,735.700        285.524           6        63096   "
##  [8] "3  Dutchess      1,262      197,308.800        308.295          18        297488  "
##  [9] "4   Greene       404        38,542.920         60.223           5        49221   "
## [10] "5   Orange      1,070       165,313.900        258.303          14        372813  "
## [11] "6   Putnam       76          3,961.006         6.189            4        99710   "
## [12] "7 Rensselaer     146        155,676.500        243.245          18        159429  "
## [13] "8   Ulster       661        70,035.010        109.430           8        182493  "
## [14] "9 Westchester    116         8,597.163         13.433          20        949113  "
## [15] "-------------------------------------------------------------------------------------"
```

Figure 1 (all figures below) is a points map shows the Farmers Markets (yellow points) and Ag Districts (red polygons) across counties in the region. We can observe that ag districts covering a lot of land in Columbia, Dutchess, Rensselaer and Orange Counties, while the markets are clustered down in Westchester, along the Hudson River in the lower half of the region, and then up in eastern Albany County. The only county where there are both a high number of markets and relatively large amount of ag lands is Albany County.

Figure 2 is a points map overlaid on county population data. By overlaying the market points and district polygons onto the population chloropleth, we can start to observe how many people are interacting with markets and agricultural lands. Westchester has the highest population and a high number of farmers markets – it would make sense that demand would attract supply. However, Orange County has the next highest population and a much lower proportion of farmers markets. This points to other factors besides population size influencing the presence of farmers markets, such as socioeconomic or geographic differences between the counties. The fact that Westchester is one of the wealthiest counties in the country, combined with its close proximity to New York City, means there is likely a disproportionately high-demand market there. These questions could be explored in further analysis.

Figure 3 is a chloropleth map showing the number of farmers markets per county, using a Jenks natural break classification. To prepare this data for mapping, I merged the non-spatial markets count data with the county spatial data by the shared NAME variable. This map shows that the county with the most markets is Albany, while Putnam, Columbia and Greene are in the lowest bucket of only 4-6 markets.

Figure 4 shows two choropleth maps, also using natural breaks, illustrating the density of the agricultural districts compared with different measures: population and total area. Columbia County has the highest density in both of these categories, meaning it has both a large amount of agricultural lands in proportion to its population, as well as its total area. Putnam and Westchester both have the lowest densities. While Westchester has the highest population in the region, Putnam has one of the lowest, which indicates that Putnam County has the highest potential to add more agricultural districts – assuming that low-population areas are not already being used to other industries.

The magnitude and the range of density rates are also worth exploring further. The population density rate is much smaller than the area density rate by several orders of magnitude, which suggests that there is much more agricultural land per square mile than there is per person. This makes sense, since (of course) there is more than one person living per square mile. However, it also points to the ways in which this line of analysis could be further used in research or policy. For example, this comparison could be useful for exploring how productive agricultural lands can be expanded into more densely populated areas or identifying where land protection and conservation programs have been successful.

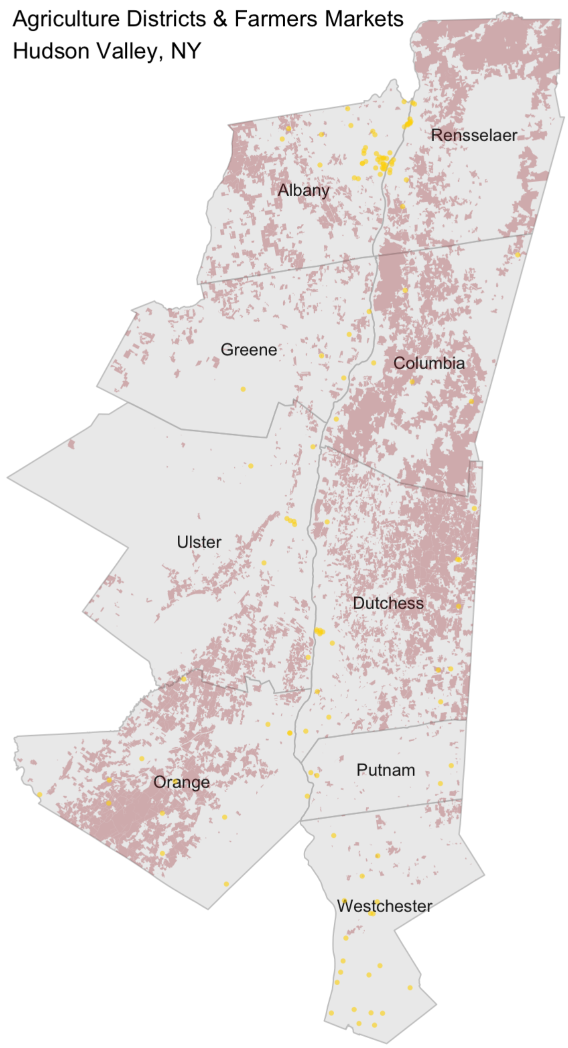

Figure 1: Agricultural Districts and Farmers Markets in the Hudson Valley, NY

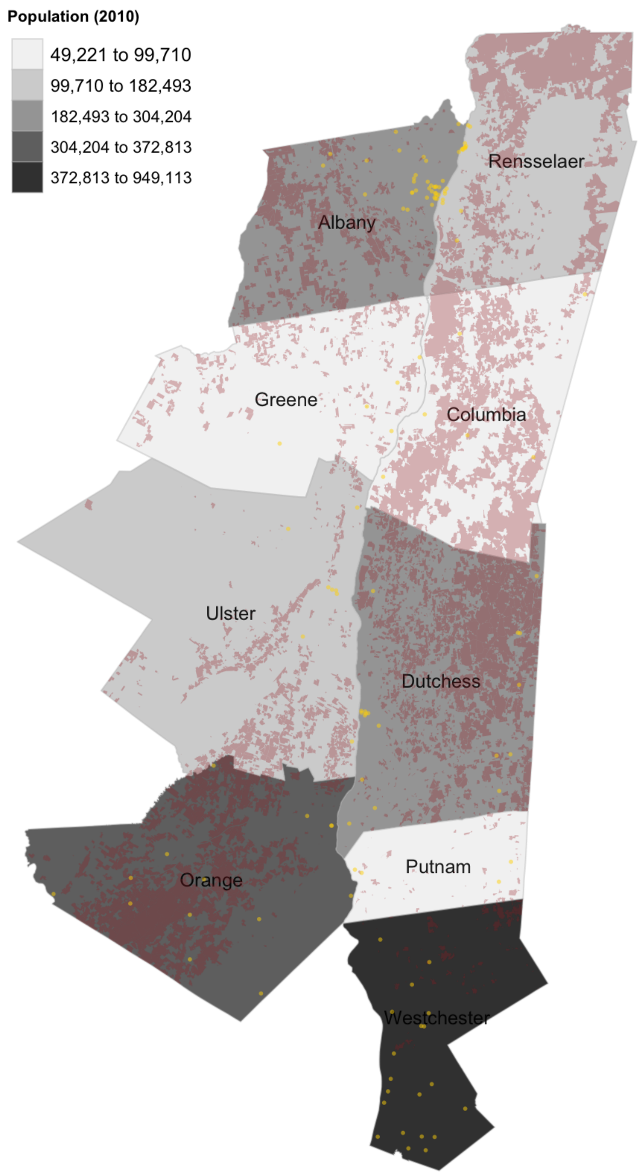

Figure 2: Points Overlaid on Population

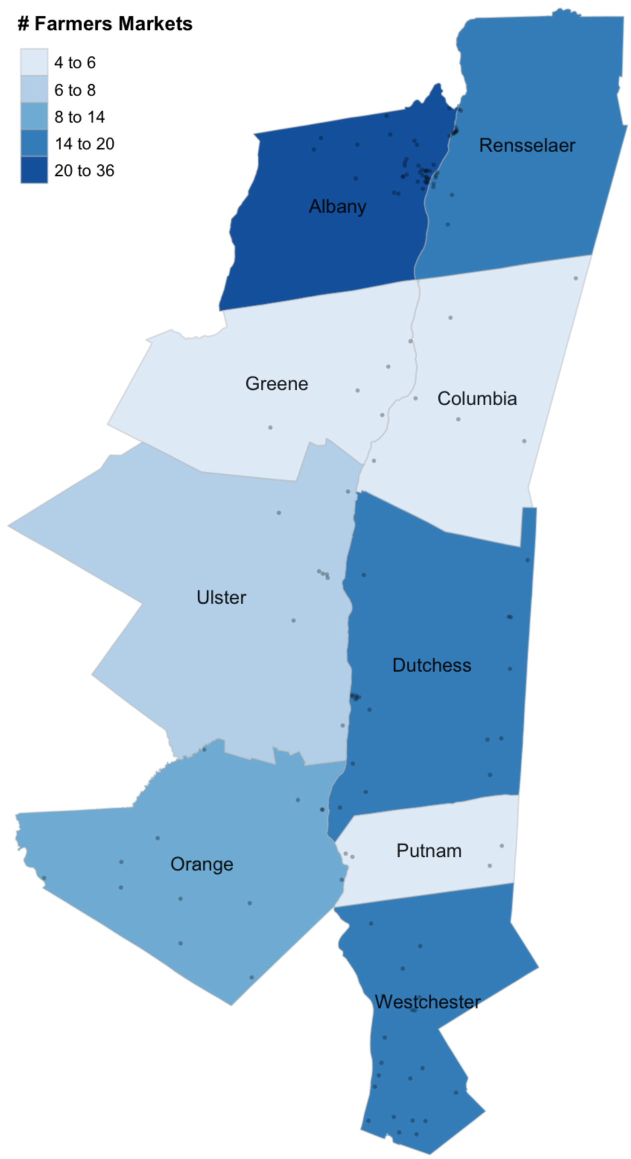

Figure 3: Farmers Markets by Count

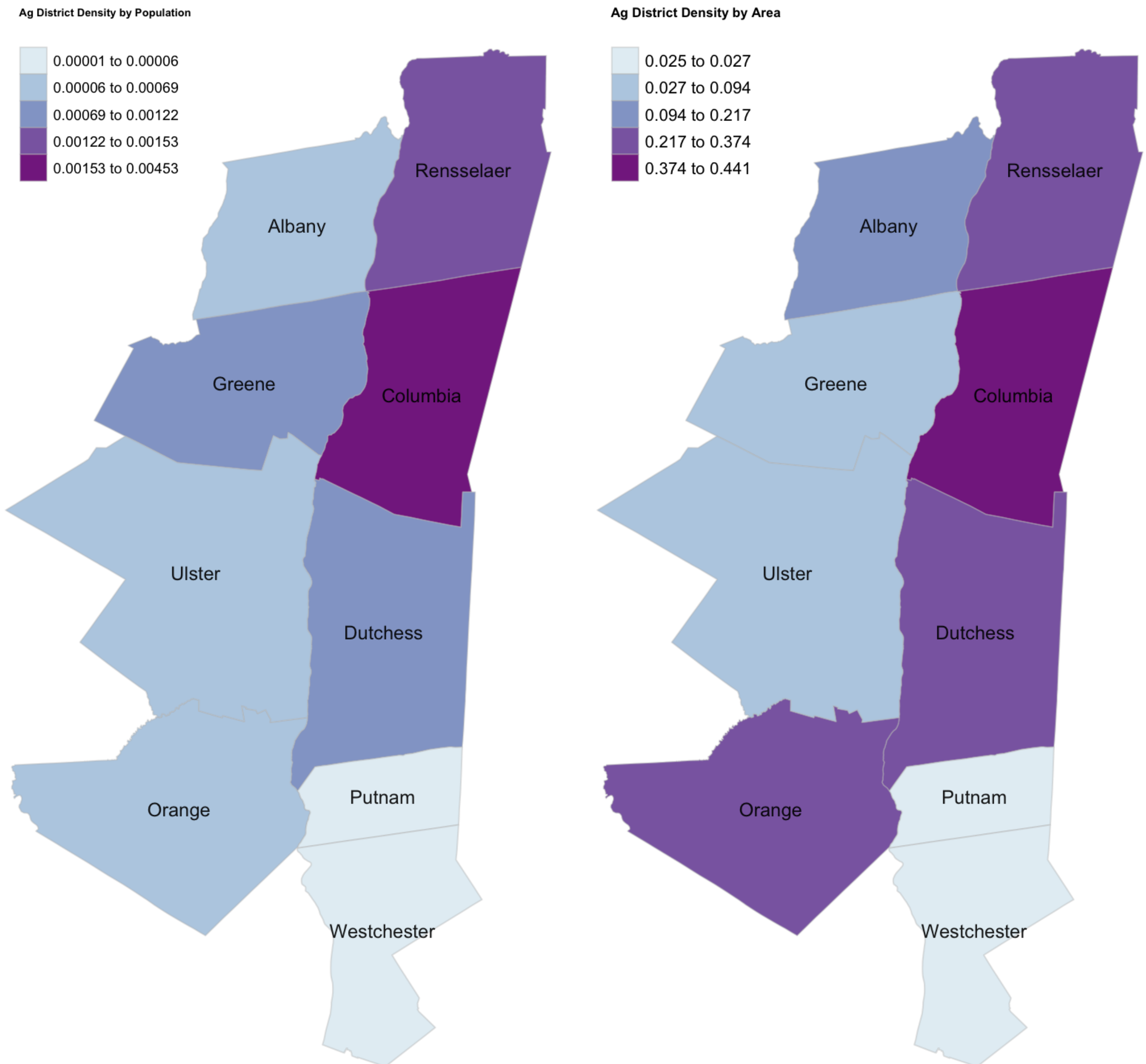

Figure 4: Density Ratios