Virginia Cooperative Extension Community Dashboard

Luke Kim (VT), Cory Kim (VT) and Devika Nair (American University)
SDAL: Aaron Schroeder   Affiliations: Michael Lambur (Virginia Cooperative Extension)

SDAL SOCIAL & DECISION ANALYTICS LABORATORY
DATA SCIENCE FOR THE PUBLIC GOOD

Background

The Virginia Cooperative Extension is a network of over 100 local and regional offices operating out of the public universities, Virginia State & Virginia Tech. Extension agents, VCE's public servants, work with counties to provide educational programs, develop research, and administer resources with a particular focus on agriculture, natural resources, and environmental science. Agents are tasked with solving local problems, but often do not have insight into how their population's needs vary throughout their assigned region. The goal of this project is to identify and translate meaningful datasets to a sub-county level of detail to better guide agents' decisions around program investments and resource allocation.

The project began with of a phase of data discovery and evaluation, and ultimately, we decided to leverage federally published datasets around several domains of natural resources:

- EPA - Air Quality Index, Radon
- USGS - Soil composition
- MSHA - Mines
- USDA - Crop Coverage

Once datasets were identified, our team used web-scraping techniques to access and store the data locally. Much of this data was provided at the county level, rather than at the locality level of detail (see Figure 1 at right). Different methods were applied to transform our data to a more granular level of detail, to facilitate analysis for extension agents.

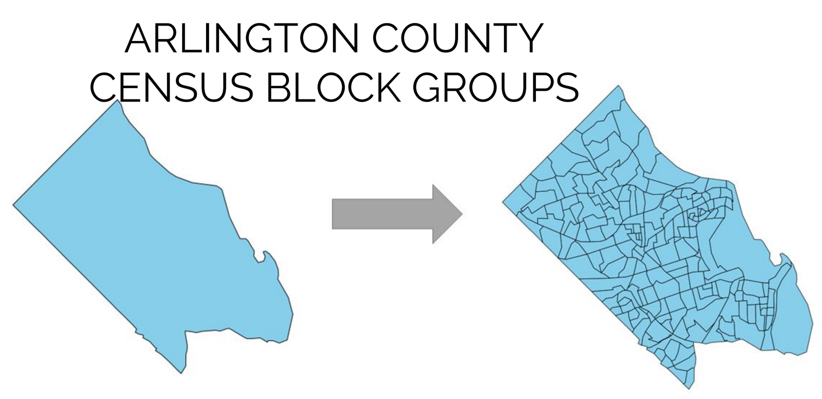


Figure 1: Arlington County divided into its Census Block Groups

Methods

Centroid
Centroid locations were calculated for each census block group of Virginia, using the sf package and st_centroid function.

Inverse Distance Weighting (IDW):
A data interpolation method in which unknown values at specific locations are approximated through weighting values at known locations. Weight formula is outlined below, with w as weight, x and y as locations, and d as the distance between them.

$$w(x,y) = \frac{1}{d(x,y)^2}$$

Greater distances lead to a lower assigned weight. For air quality, this means the further away an air quality monitor is from a census block, the less influence it has on the corresponding AQI approximation.

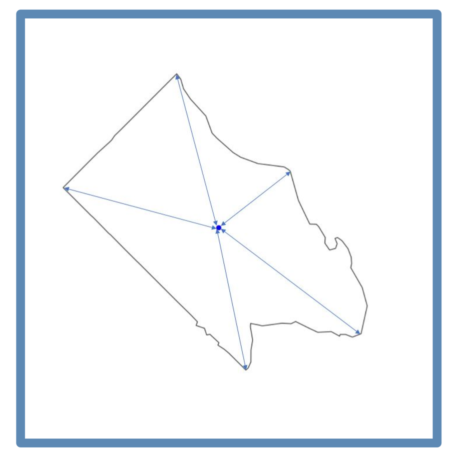


Figure 2: Centroid coordinate example of Arlington County

Productivity Calculations:
Soil taxonomy is given in a number of layers: order, suborder, great group, sub group, and texture. These were translated into point values to better assess and evaluate soil fertility and this calculation is given below.

$$\text{Productivity Index} = \text{Order} \pm \text{Suborder} \pm \text{Great Group} \pm \text{Subgroup} \pm \text{Texture}$$

Image Raster:

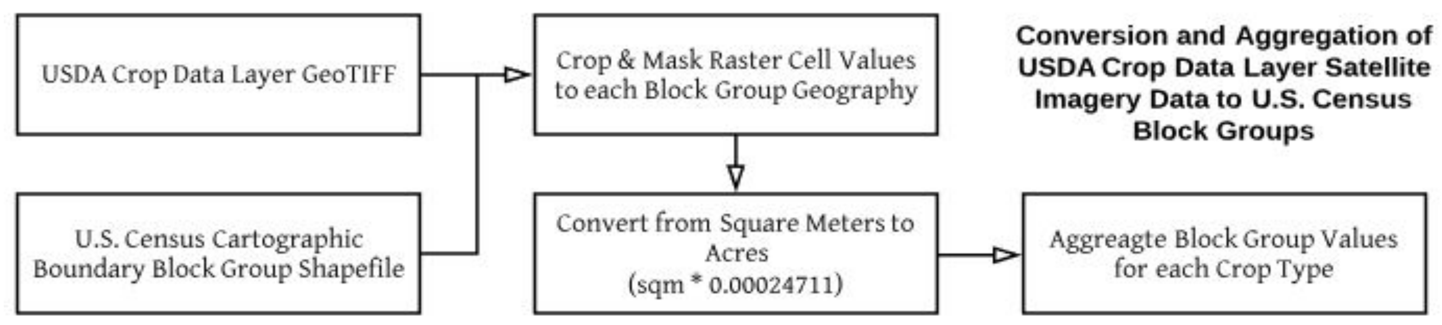
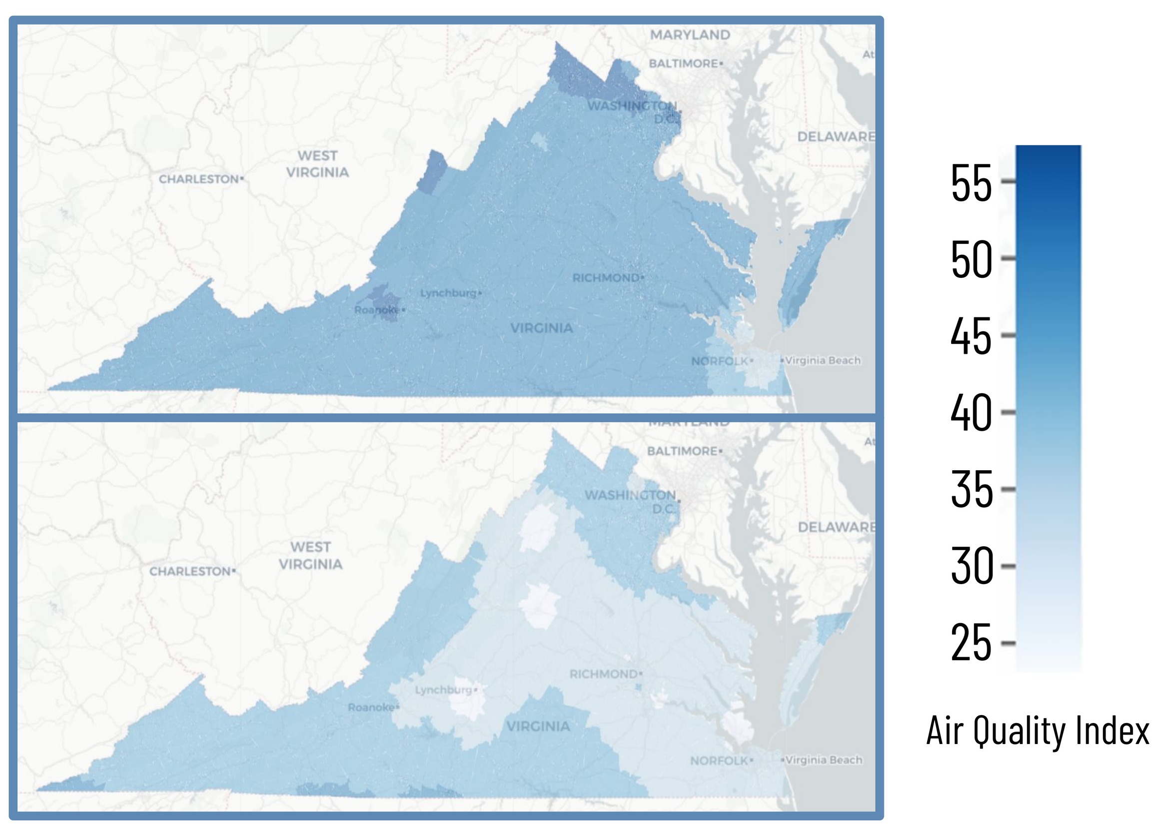


Figure 6: Crop Satellite Imagery method

Air Quality



Figures 3a and 3b: Air Quality Index Interpolations for the (a) Summer months of 2013 (above) and (b) Non-Summer months of 2013 (below). Lower AQI values indicate healthier air quality levels.

Crop and Mine Locations

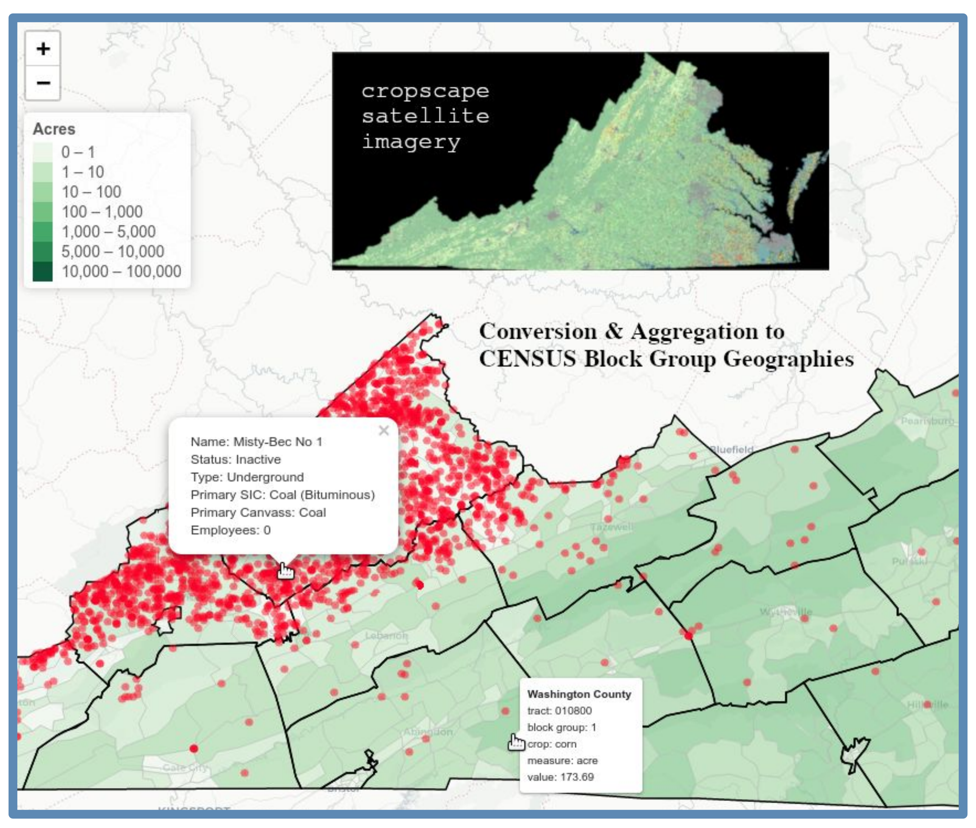


Figure 4: Coal Mines and Cropscape Satellite Imagery

Soil Productivity

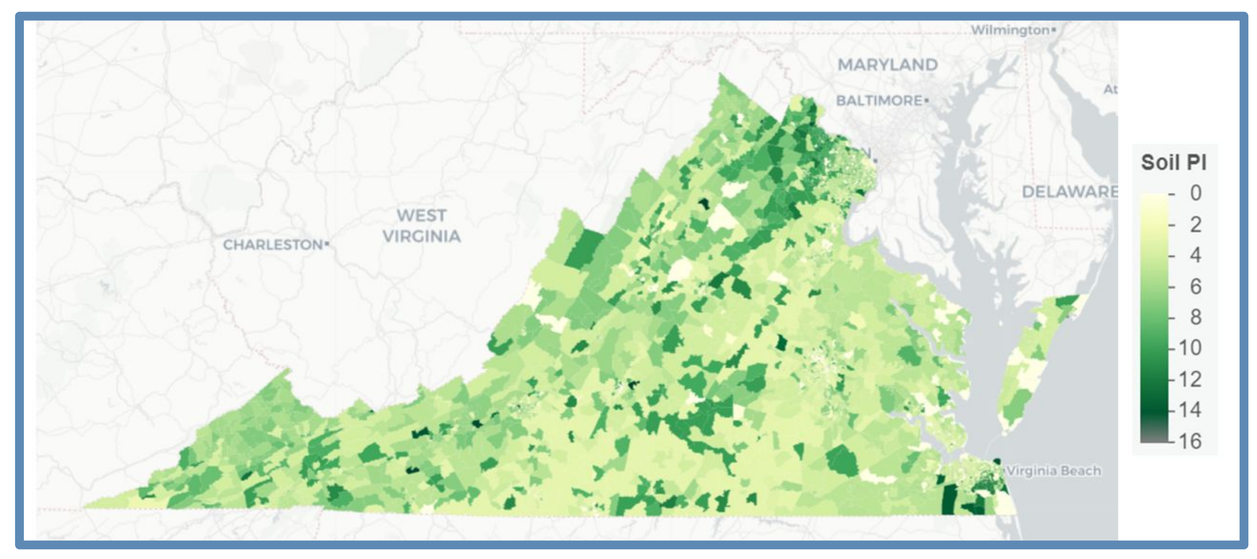


Figure 5: Soil Productivity Index of Virginia by census block group

Findings and Conclusions

We sought to provide a more dimensional view of a given region of Virginia by translating broader datasets about the area's natural resources into granular site-specific maps. These multidimensional maps enable agents to make more informed decisions about their programs by combining the socio-demographic information given in the census with regional data about the locality's natural resources.

Air quality is one such resource. Our maps reveal that AQI levels shift significantly over time. Virginia sees higher AQI levels during summer months as compared to the rest of the year. Also, AQI has shown a steady decline over time. Analyzing trends in AQI can help agents forecast crop output and public health concerns.

Soil fertility is another such resource. Measuring and visualizing soil productivity can help agents maximize crop yield and target environmental efforts by focusing programs that support the best possible use and reconstruction of soil. Our map highlights productive and non-productive block groups (reasons ranging from geography to climate to urban land), allowing agents to target vulnerable areas and support productive areas.

References | Acknowledgements

Shepard, Donald. 1968. "A Two-Dimensional Interpolation Function for Irregularly-Spaced Data." In , 517–24. ACM Press. DOI: 10.1145/800186.810616

Schaetzl, *et al.* 2012. "A Taxonomically Based Ordinal Estimate of Soil Productivity for Landscape-Scale Analyses." *Soil Science*. Lippincott Williams & Williams. DOI: 10.1097/SS.0b013e3182446c88

BIOCOMPLEXITY INSTITUTE
bi.vt.edu/sdal
Stay Connected: facebook.com/biocomplexity
twitter.com/biocomplexityvt
youtube.com/c/biocomplexity
google.com/+biocomplexity

Virginia Cooperative Extension
Virginia Tech • Virginia State University

VIRGINIA TECH