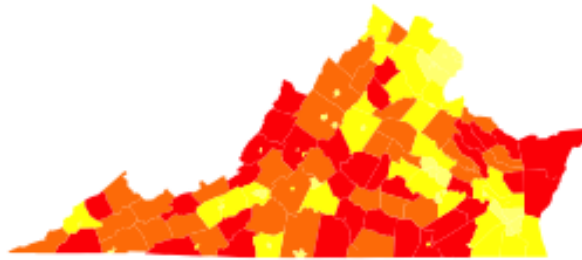


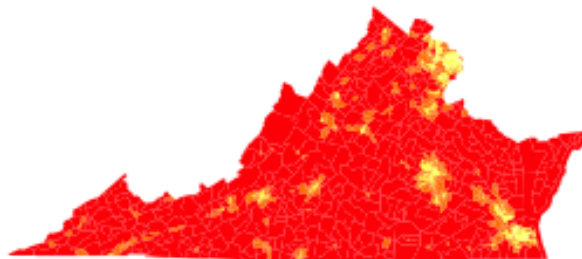
Lab 2. Data Science Basics: Packages, Datasets and Spatial Objects

Due by noon on Friday, September 11th

County Densities



Tract Densities



Block Group Densities



"Everything is related to everything else, but near things are more related than distant things."

Tobler, 1970



Laboratory in Brief:

The purpose of this laboratory is to introduce students to working with spatial data by statistically describing it and quantitatively analysing it in an increasingly disaggregate manner. You will be introduced to the statistical programming framework R, by learning how to install a package and execute some of its basic functions. Initially you will use the `chloropleth` package to statistically describe a single demographic characteristic of one American state at the county level. Following this, you will use the `USCensus2010` package to create a spatial R object and increase the disaggregate level of analysis from county to sub-county areas. Finally, you will learn how to remotely access the Census Bureau's Application Program Interface (API) directly from within R, create a sub-county area spatial object, and statistically describe that area by one or more demographic characteristics.

Specific Objectives:

1. To install the `chloropleth` package in R and create a map that statistically and spatially describes a demographic characteristic of one American State, county by county
2. To add a dataset, create a spatial object in R and create a map that statistically and spatially describes a demographic characteristic of several bordering counties within a state, tract by tract
3. To obtain data from the Census bureau Application Program Interface (API), create a R object and create a map that statistically and spatially describes a demographic characteristic of a county, block by block

Software and Resources you will need or will be helpful:

1. The R Project for Statistical Computing otherwise known simply as R. The R framework can be downloaded from <https://www.r-project.org/>
2. Quick R, an online resource developed by Rob Kabacoff, a statistical consultant and methodologist, that is purposed towards introducing students to the power of R. <http://www.statmethods.net/>
3. Introduction to R, a pdf presentation created by Phil Spector, Applications Manager and Associate Adjunct Professor at Cal-Berkeley. The pdf is more advanced than the the Quick-R online learning resource, but it covers many of R's basics with regard to functionality and emphasis on data management. <https://www.stat.berkeley.edu/~spector/Rcourse.pdf>
4. Packages and datasets as instructed during the laboratory



Grading

This lab will be graded in 4 parts. **First**, you create a map that spatially describes your chosen state according some demographic characteristic, count by county (5%). **Second**, you will create a map that spatially describes two or more adjoining counties within your state according to some demographic characteristic, tract by tract (10%). **Third**, you will create a map that spatially describes a sub-county area of your state according to some demographic characteristic, block by block (15%). **Finally**, you will use each of these three maps in a report that spatially and statistically describes your state and its dissaggregate parts in accordance with one or more demographic or economic characteristics.

Your lab report should include the following elements.

1. all three maps, including description an analysis of each one
2. integrated into your report, a description of the code you used, in a manner that demonstrates your knowledge of how the code functioned and operated
3. an analysis of how increased disaggregation effects the statistical description of spatial data

The highest grades will be reserved for work that not only spatially describes your chosen area in a statistically rigorous manner, but also uses quantitative analysis to suggest and support inferential conclusions.

"Data of geographic units are tied together, like bunches of grapes, not seperate, like balls in an urn. Of course, mere contiguity in time and space does not of itself indicate independence between units in a relevant variable or attribute, but in dealing with social data, we know that by very virtue of their social character, persons, groups, and their characteristics are interrelated and not independent. Sampling error formulas may yet be developed which are applicable to these data, but until then, the older formulas must be used with great caution. Likewise, other statistical measures must be carefully scrutinised when applied to these data"

Stephan, 1934