Case Study 4: 2010-2018 WPD Data with a Census Data Pull

DSC 551

Background

- Police data on incidents, accidents, and arrests can be obtained from the local PD.
- There is an interest in understanding where arrests are occurring, how they are changing over time, and if there are any commonalities in these trends.
- PD’s have their own districts. However, we are not able to get census (or similar) variables at that level.

Data

- NC census tract shapefile
- The arrest data has already been reorganized for you into counts of arrests and other variables by tract and year.
- There's also a data set for the regression models

Objectives

1. Manipulate the shapefile
2. Calculate and map several measures for visualization of the data
3. Pull and work with Census data
4. Perform Poisson regression
5. Create a dashboard

Manipulating shapefiles

You already have the tools for this.

1. Read in the NC tracts shp.
2. Keep only those that are from New Hanover County
3. Reorder the spatialpolygonobject so that the tract labels are smallest to largest
4. Remove the last tract.

Calculating and mapping several measures

You also already have the tools for this!

Calculations:

1. Percent of total arrests
2. Arrests as a percent of the appropriate population
3. SIR of arrests

Census data pull

You can pull Census data through an API or directly from the Census webpage.

Find data to use in the Poisson regression using the following steps:

1. Filter by Geography for all tracts within NC's New Hanover County2. Filter by Topic: Housing
   Search
   Download 5 year ACS estimates per year (2010-2018) files for "Selected Housing Characteristics"
3. Filter by Topic: Housing and Vacancy
   Search
   Download 5 year ACS estimates per year (2010-2018) files for "Vacancy Status"
4. Filter by Topic: Populations and People
   Search
   Download 5 year ACS estimates per year (2010-2018) files for "SELECTED CHARACTERISTICS OF THE TOTAL AND NATIVE POPULATIONS IN THE UNITED STATES"

Census data pull

The data needs to be organized in the same way as the WPD data

- Total Housing Units: First pull, DP04_0001E
- Total Vacant Housing Units: First pull, DP04_0003E
- Total Vacant Housing Units Designated as Secondary Homes: Second pull, B25004_006E
- % Pop Aged 18-24: Third pull, S0601_C01_004E
- Median age: Third pull, S0601_C01_010E
- % Male Population: Third pull, S0601_C01_011E
- % Single population: Third pull, S0601_C01_028E
- % Poverty: Third pull, Pop$S0601_C01_049E
- % Education - < HS: Third pull, Pop$S0601_C01_033E
- % Education - HS: Third pull, S0601_C01_034E
- % Education - Some College: Third pull, S0601_C01_035E
- % Education - Bachelors, S0601_C01_036E
- % Education - Graduate school: Third pull, S0601_C01_037E

Census data pull

1. Create a matrix (dPop) with the appropriate number of rows and 20 columns for tract, year,
2. Give appropriate column names based on the variables listed in 1
3. Create the year column
4. Make a for loop to loop through the 9 years (2010-2018)
5. Read the pulled census data.
6. Fill in the correct matrix rows with the year-specific data
7. Calculate additional variables (secondary homes as a percent of total homes, secondary homes as a percent of vacant homes, education HS or less, education Bachelors or more)
8. Replace any missing or infinite values (from dividing by zero) with 0

Visualization

- View the data you've compiled thus far.
- I recommend corrplot::corrplot

Poisson regression

$$y_{ij} \sim Pois(\mu_{ij})$$
$$\mu_{ij} = \theta_{ij} e_{ij}$$
$$\log(\theta_{ij}) = X'_{ij}\beta + \phi_{ij}$$

- Used for count data
- Additional intro/review materials
- Use INLA again for model fitting.
- The purpose of ϕ_{ij} is to get an adjusted measure of risk at the tract level.

Poisson regression

- Step wise variable selection lead to the following list of predictors: % black, % poverty, % high education, % male, % secondary homes, and % aged 18-24
- Recommendation: Only perform variable selection for total arrests then use that model for black/white arrests.
- We use the inverse $\log(\exp())$ for the fixed effects because it leads to an interpretation as a multiplicative change in the relative risk (θ_{ij}) of arrests, e.g. for every one percent increase in % black population, the relative risk of total arrests is 1.03 times higher.

Poisson regression

- The random effect included here is spatio-temporal
- The “id” supplied to f() must reflect that
- Plot the random effect estimates like we’ve done before

Create a dashboard!

- I recommend attempting Shiny but you may use which ever you prefer.
- Shiny tutorial
- My Shiny app
- My Shiny app code