

---

# Orange Osceles Geospatial API Project

Chris Mulvey, Jake Randolph, Vahideh Rasekhi

---

---

# Getting and Cleaning Data

- About 50% of our time was spent obtaining the data we needed.
- After getting the data, another 30% of the time was cleaning the data and putting it in the format we needed.

---

# Maps

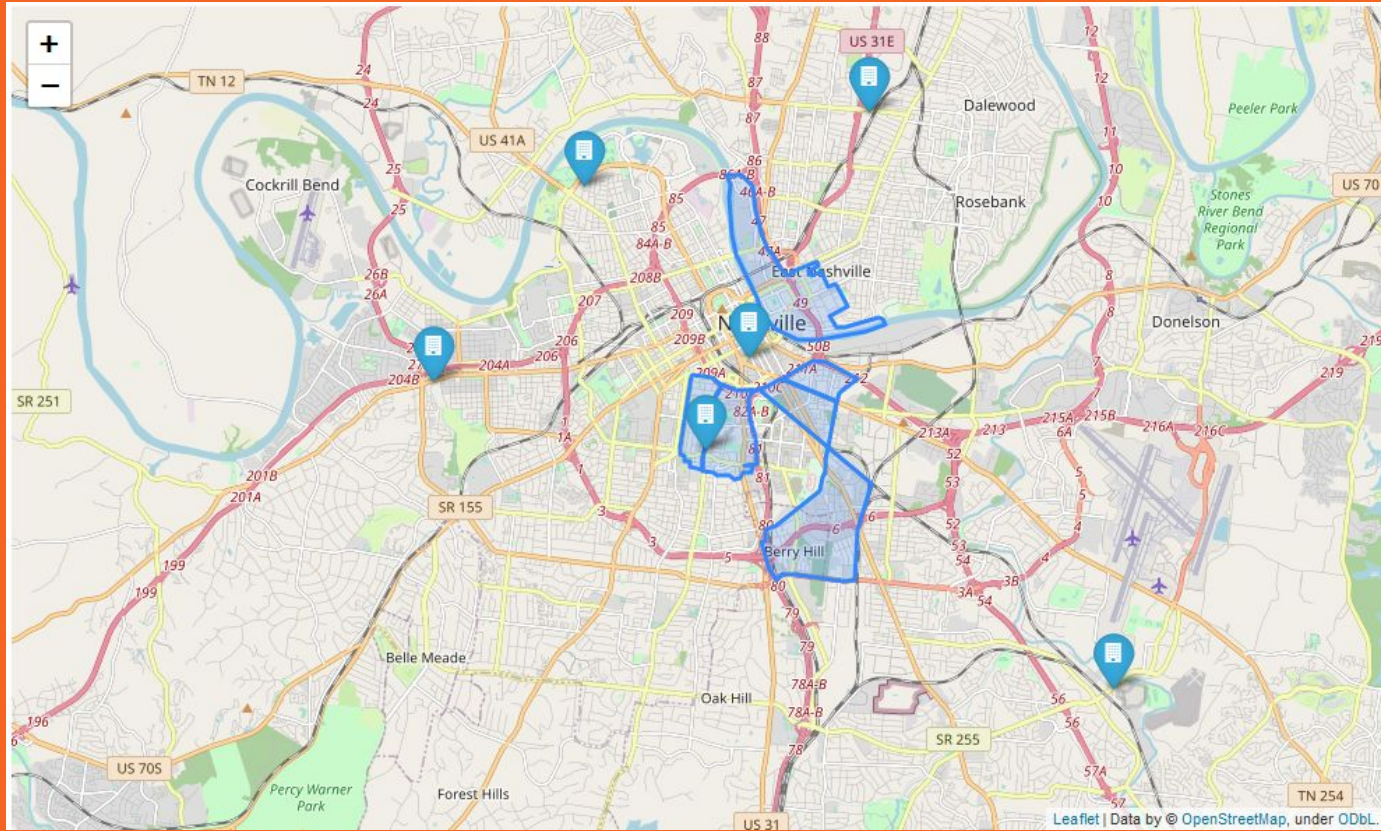
- Started with Choropleth Map
  - Police Station proximity
  - Comparison of burglary rates and median income
-

# Choropleth Map

Burglary Rate By Census Tract Per 1000 People, 2021



# Police Precinct Proximity

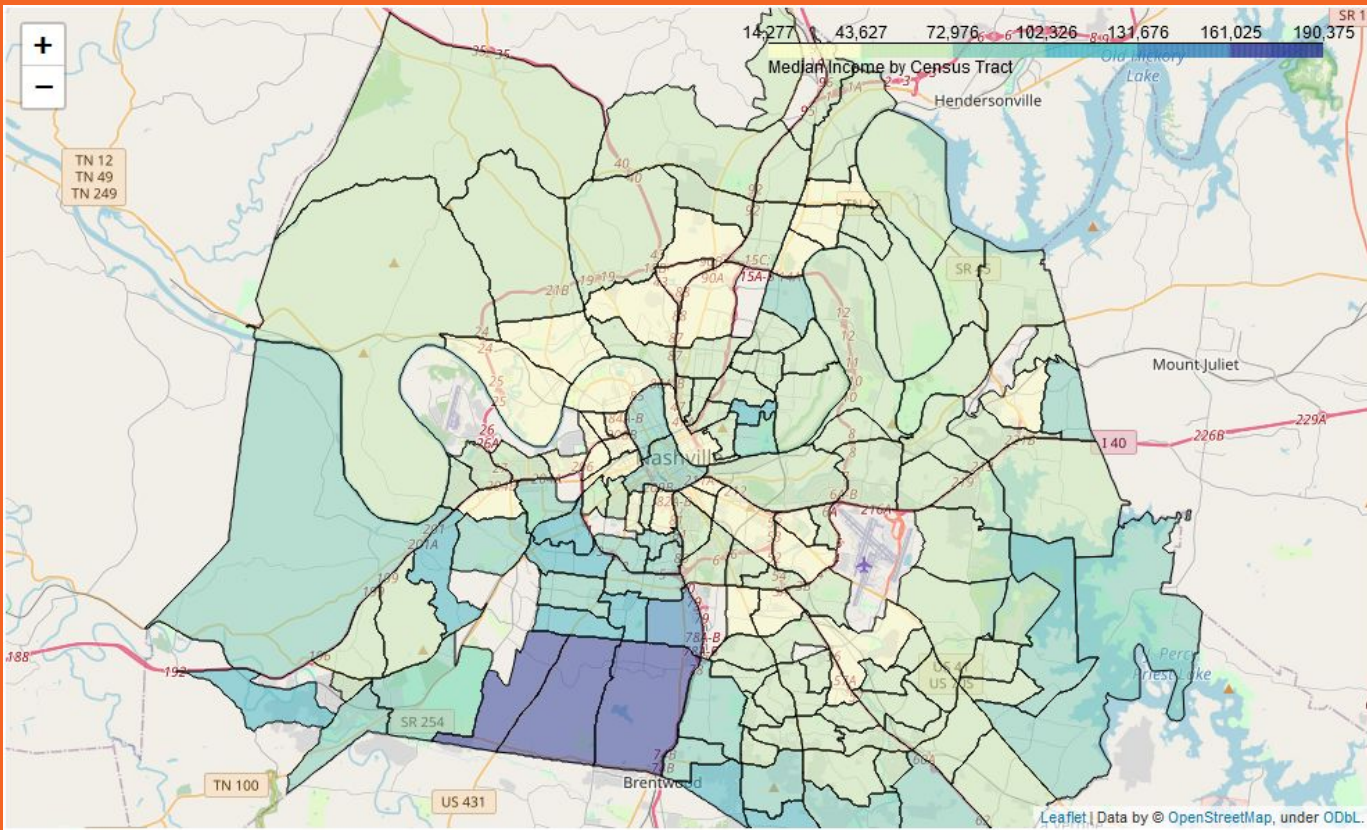


Tracts shown are all tracts with over 5 burglaries per 1000 residents





# Median Income By Tract



# Poisson regression base model

Target variable: *rate of burglaries per census tract*

## Generalized Linear Model Regression Results

<b>Dep. Variable:</b>	burglaries_per_1000	<b>No. Observations:</b>	146
<b>Model:</b>	GLM	<b>Df Residuals:</b>	145
<b>Model Family:</b>	Poisson	<b>Df Model:</b>	0
<b>Link Function:</b>	log	<b>Scale:</b>	1.0000
<b>Method:</b>	IRLS	<b>Log-Likelihood:</b>	-307.95
<b>Date:</b>	Sat, 09 Oct 2021	<b>Deviance:</b>	317.23
<b>Time:</b>	10:33:47	<b>Pearson chi2:</b>	907.
<b>No. Iterations:</b>	5		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	z	P> z	[0.025	0.975]
<b>const</b>	-1.0922	0.068	-16.124	0.000	-1.225	-0.959

The estimated mean of the distribution of burglary per 1000 population is **-1.0922**.

The estimated variance of the distribution is **1**.



# Poisson regression model

Target variable: rate of burglaries/ predictor variable: median income

## Generalized Linear Model Regression Results

<b>Dep. Variable:</b>	burglaries_per_1000	<b>No. Observations:</b>	146
<b>Model:</b>	GLM	<b>Df Residuals:</b>	144
<b>Model Family:</b>	Poisson	<b>Df Model:</b>	1
<b>Link Function:</b>	log	<b>Scale:</b>	1.0000
<b>Method:</b>	IRLS	<b>Log-Likelihood:</b>	-279.19
<b>Date:</b>	Sat, 09 Oct 2021	<b>Deviance:</b>	259.72
<b>Time:</b>	10:34:05	<b>Pearson chi2:</b>	682.
<b>No. Iterations:</b>	5		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	z	P> z	[0.025	0.975]
const	0.3222	0.198	1.630	0.103	-0.065	0.710
median_income	-2.449e-05	3.55e-06	-6.907	0.000	-3.14e-05	-1.75e-05

For median income of  $t$ , the estimated value of the mean is  $\exp(0.3222 + (-2.449e-05t))$

Rate of burglary decreases by  $-2.449e-05$  times the median income

# Negative binomial model

## Generalized Linear Model Regression Results

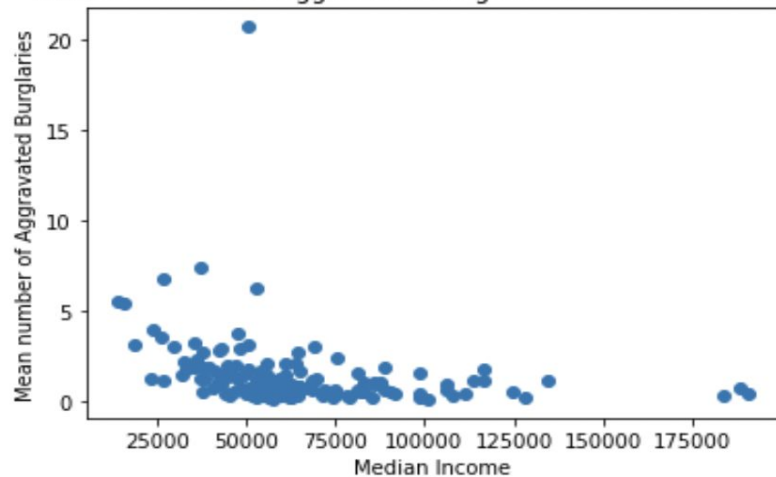
<b>Dep. Variable:</b>	burglaries_per_1000	<b>No. Observations:</b>	146
<b>Model:</b>	GLM	<b>Df Residuals:</b>	144
<b>Model Family:</b>	NegativeBinomial	<b>Df Model:</b>	1
<b>Link Function:</b>	log	<b>Scale:</b>	1.0000
<b>Method:</b>	IRLS	<b>Log-Likelihood:</b>	-258.77
<b>Date:</b>	Mon, 11 Oct 2021	<b>Deviance:</b>	94.035
<b>Time:</b>	18:25:42	<b>Pearson chi2:</b>	278.
<b>No. Iterations:</b>	7		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	z	P> z	[0.025	0.975]
<b>Intercept</b>	-6.5958	0.303	-21.788	0.000	-7.189	-6.002
<b>median_income</b>	-2.1e-05	4.9e-06	-4.283	0.000	-3.06e-05	-1.14e-05

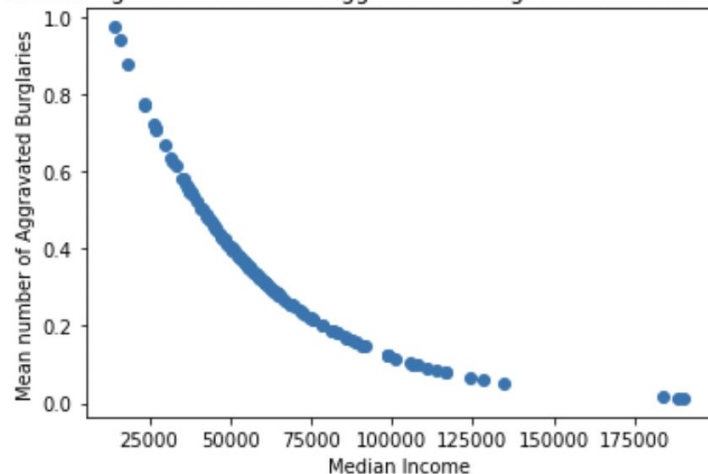
For a one unit change in the predictor variable (median income), the difference in the logs of expected counts of the response variable (rate of burglary) is expected to change by -.000021.

# Graphical Comparisons

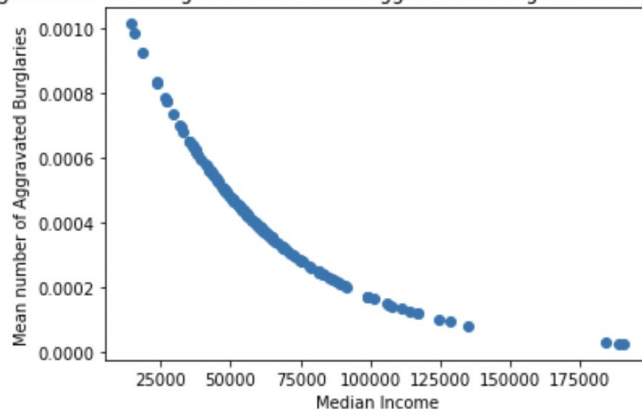
Median Income vs. Aggravated Burglaries in Davidson Counties



Poisson Regression Model for Aggravated Burglaries in Davidson County



Negative Binomial Regression Model for Aggravated Burglaries in Davidson County



---

# Akaike Information Criterion (AIC)

AIC attributes:

Base Poisson Regression: **617.89**

Poisson Regression: **562:37**

Negative Binomial Regression: **521.53**

---

# Scatterplot of Incidents and Income

