Analysis of Gun Violence in 2019

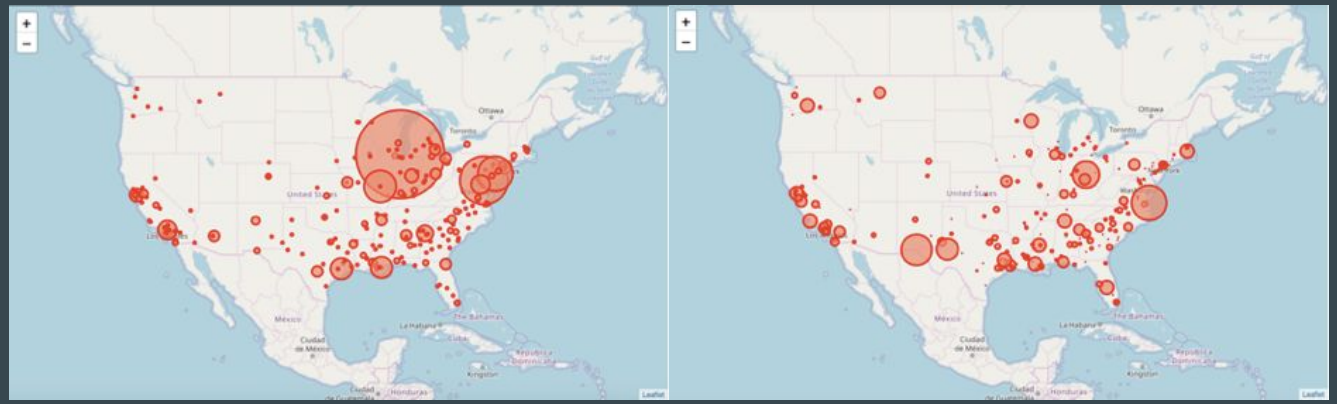
Lulu Ricketts -- March 2020

Overview

- Goal: to determine which locations are more at risk for violence and shootings
- Steps
    - 1. Understand data
    - 2. Cluster
        - via Per Capita Incomes
        - via location’s venues

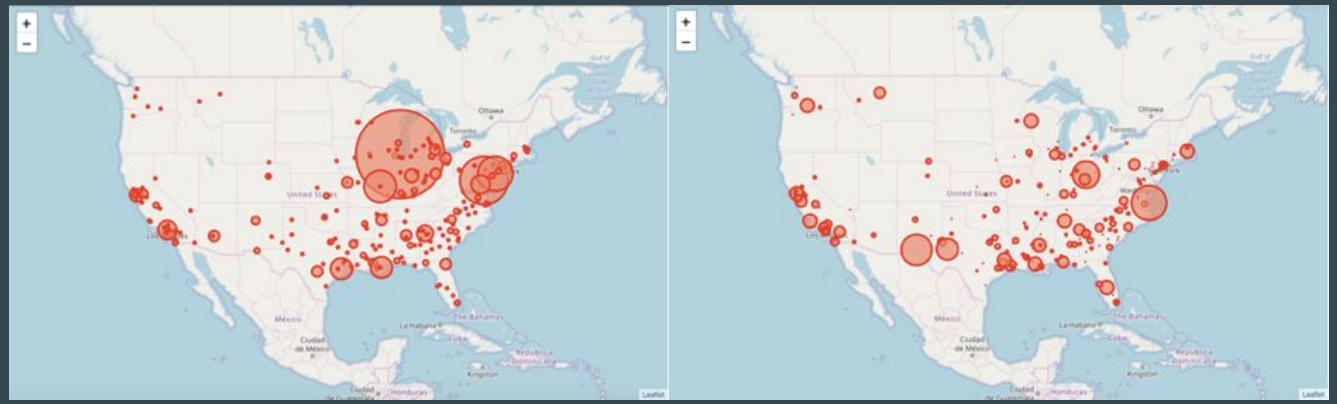
Understanding the Data

- The first step was to map the initial distributions in two ways:
    - 1: by the number of incidences for each location in 2019
    - 2: by the average number of people killed per incidence



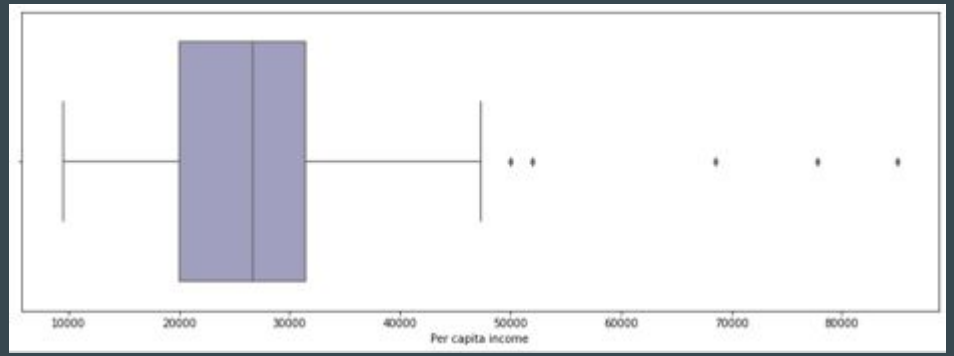
Understanding the Data

- Initial thoughts:
  - Number of incidences does not predict the violence of each incidence
  - “At risk” can mean either at risk of shootings or at risk of higher violence

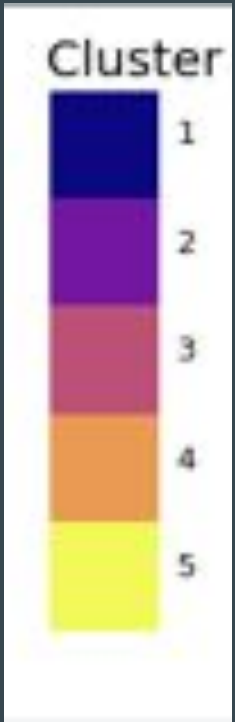


Cluster via Per Capita Incomes

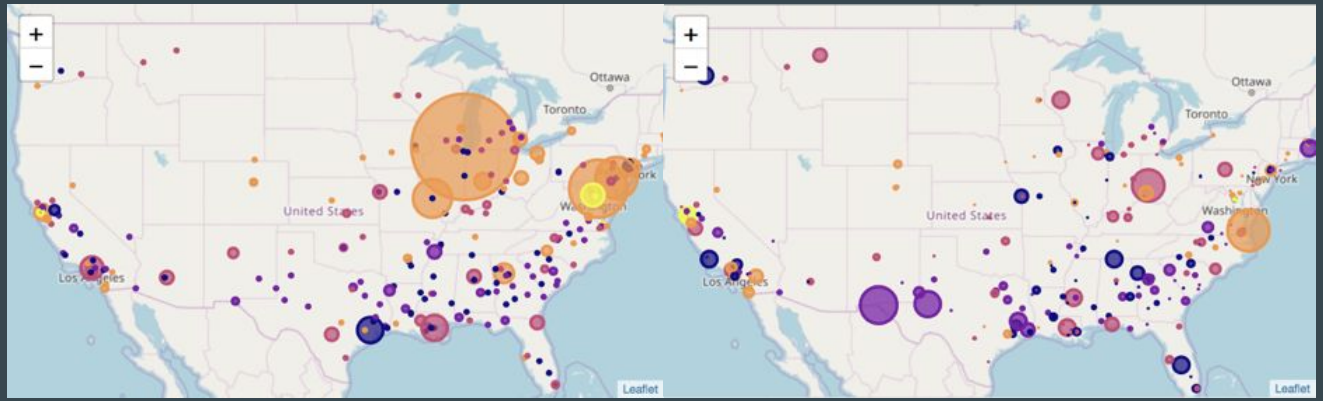
- 5 clusters created by quartiles of a boxplot
  - Cluster 1: first quartile, lower class
  - Cluster 2: second quartile, lower/mid class
  - Cluster 3: third quartile, middle class
  - Cluster 4: fourth quartile, upper middle class
  - Cluster 5: outlier, upper class



Cluster via Per Capita Incomes



- Higher income locations have more incidences
- Lower/mid income locations have more violence (people killed) per incidence



Cluster via Location's Venues

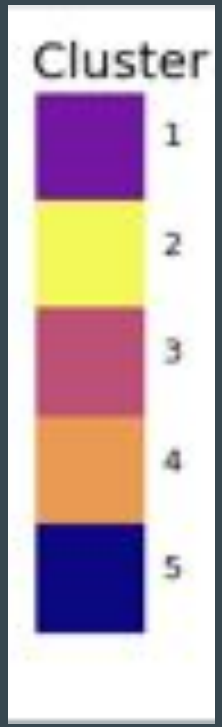
- Motivation: the venues in a location may give a description of that cluster and how it relates to per capita incomes
  - Could answer WHY some places are more at risk
- Strategy:
  - Use Foursquare Places API to get top 20 venues per location
  - One hot encode each unique venue
  - Use k-means clustering to form clusters

Cluster via Location's Venues

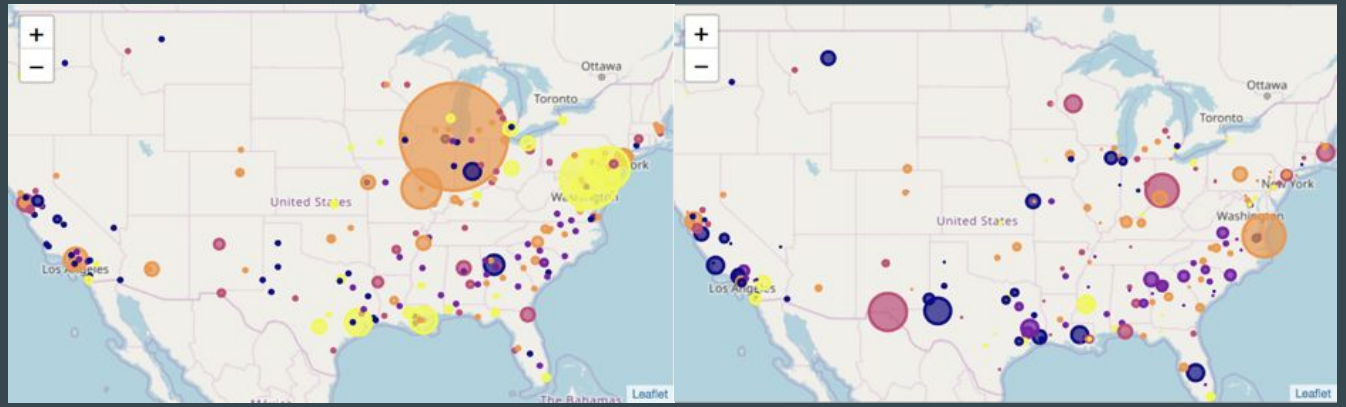
	Address	Accessories Store	Adult Boutique	African Restaurant	Airport	American Restaurant	Antique Shop	Aquarium	Arcade	Arepa Restaurant	...
0	Abbeville, South Carolina	0.0	0.0	0.0	0.0	0.00	0.0	0.00	0.0	0.0	...
1	Abington, Massachusetts	0.0	0.0	0.0	0.0	0.10	0.0	0.00	0.0	0.0	...
2	Akron, Ohio	0.0	0.0	0.0	0.0	0.05	0.0	0.00	0.0	0.0	...
3	Albany, Georgia	0.0	0.0	0.0	0.0	0.10	0.0	0.05	0.0	0.0	...
4	Albuquerque, New Mexico	0.0	0.0	0.0	0.0	0.05	0.0	0.00	0.0	0.0	...

	Address	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0	Abbeville, South Carolina	Gas Station	Discount Store	Fast Food Restaurant	Grocery Store	Mexican Restaurant	Fried Chicken Joint	Gym
1	Abington, Massachusetts	Coffee Shop	Liquor Store	Ice Cream Shop	Bar	American Restaurant	Pizza Place	Fast Food Restaurant
2	Akron, Ohio	Music Venue	Italian Restaurant	Performing Arts Venue	Deli / Bodega	Art Museum	Speakeasy	Food & Drink Shop
3	Albany, Georgia	American Restaurant	Donut Shop	BBQ Joint	Hotel	University	Plaza	Playground
4	Albuquerque, New Mexico	Bar	Café	Hotel	Design Studio	Plaza	Diner	Pizza Place

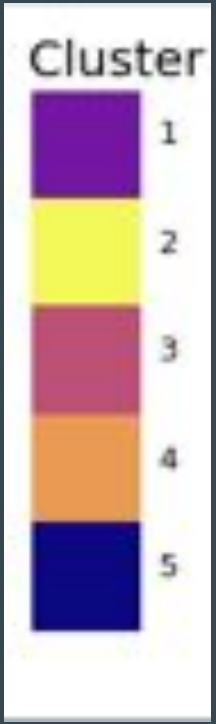
Cluster via Location's Venues



- Very similar to per capita income clusters
- BUT what do these clusters mean?



Cluster via Location's Venues



- Analyzed the top 3 venues per cluster, and these labels arose:

Cluster 1: Low income, smaller town

Cluster 2: Tourist City

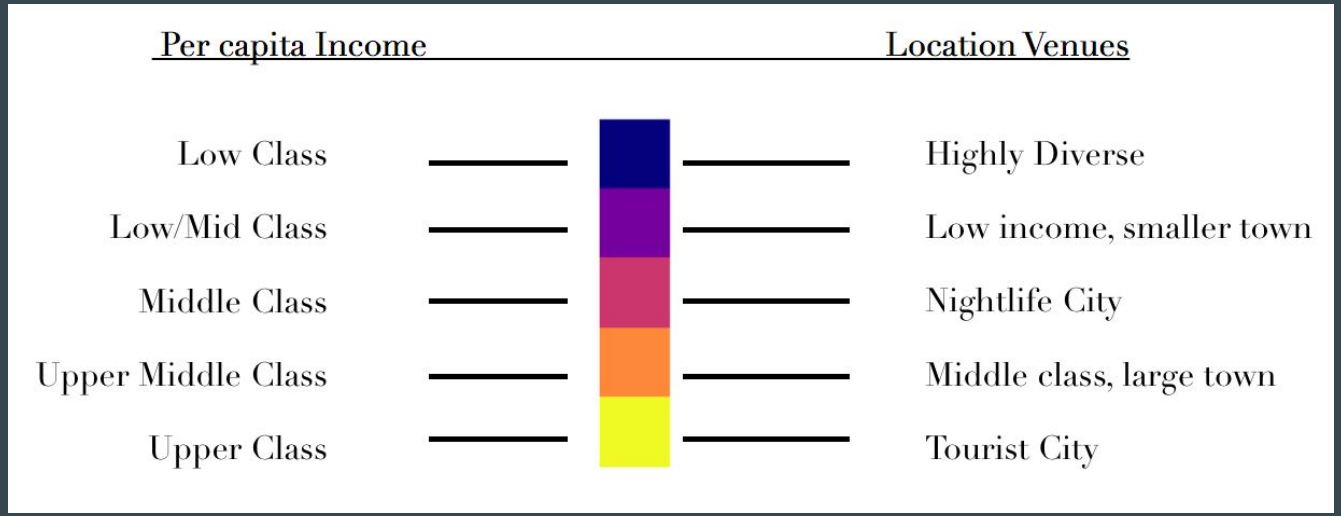
Cluster 3: Nightlife City

Cluster 4: Middle class, large town

Cluster 5: City with high diversity

Relation Between Clusters

- Clusters align by color:
  - Align pretty intuitively with about what we would expect for each income class



Relation Between Clusters - Key Takeaways

- Higher per capita income correlates with:
  - More incidences per year
  - Less violence per incidence
  - More “At Risk” for shootings
- Lower per capita income correlates with:
  - Less incidences per year
  - More violence per incidence
  - More “At Risk” for fatalities
- These could possibly be due to population size as well

Future Direction

- Analyze correlations with population size
- More reliable dataset for per capita incomes
  - Maybe use median incomes
- Account for income discrepancies and fluctuations within large cities
  - Eg. Chicago has very rich and very poor areas