

# CHARACTERIZING COMMUNITIES FOR CRIME RATE INFERENCE

Kyle Morgan (VT), Sean Pili (VT), Claire Kelling (Penn State) with Joshua Goldstein and Gizem Korkmaz (SDAL)  
Sponsor: Captain Bruce Benson, Niki Levy, Arlington County Police Department

## Research Question

This project aims to characterize communities in Arlington County that are vulnerable to certain crime types. We focus on modeling three specific crime types that are of particular interest to police departments across the country: (i) Domestic Violence, (ii) Illegal Drugs, (iii) Mental Health Issues.

Our work builds on the recent study by Graif et. al (2016)<sup>1</sup> that studies crime rate inference at the neighborhood level using the following data sources:

- Demographic Data: 2000 Census
- Points-of-Interest (POI) Data: Foursquare
- Geographic Influence: boundary shapefiles
- Hypertlink by Taxi Flow: Illinois Freedom of Information Act

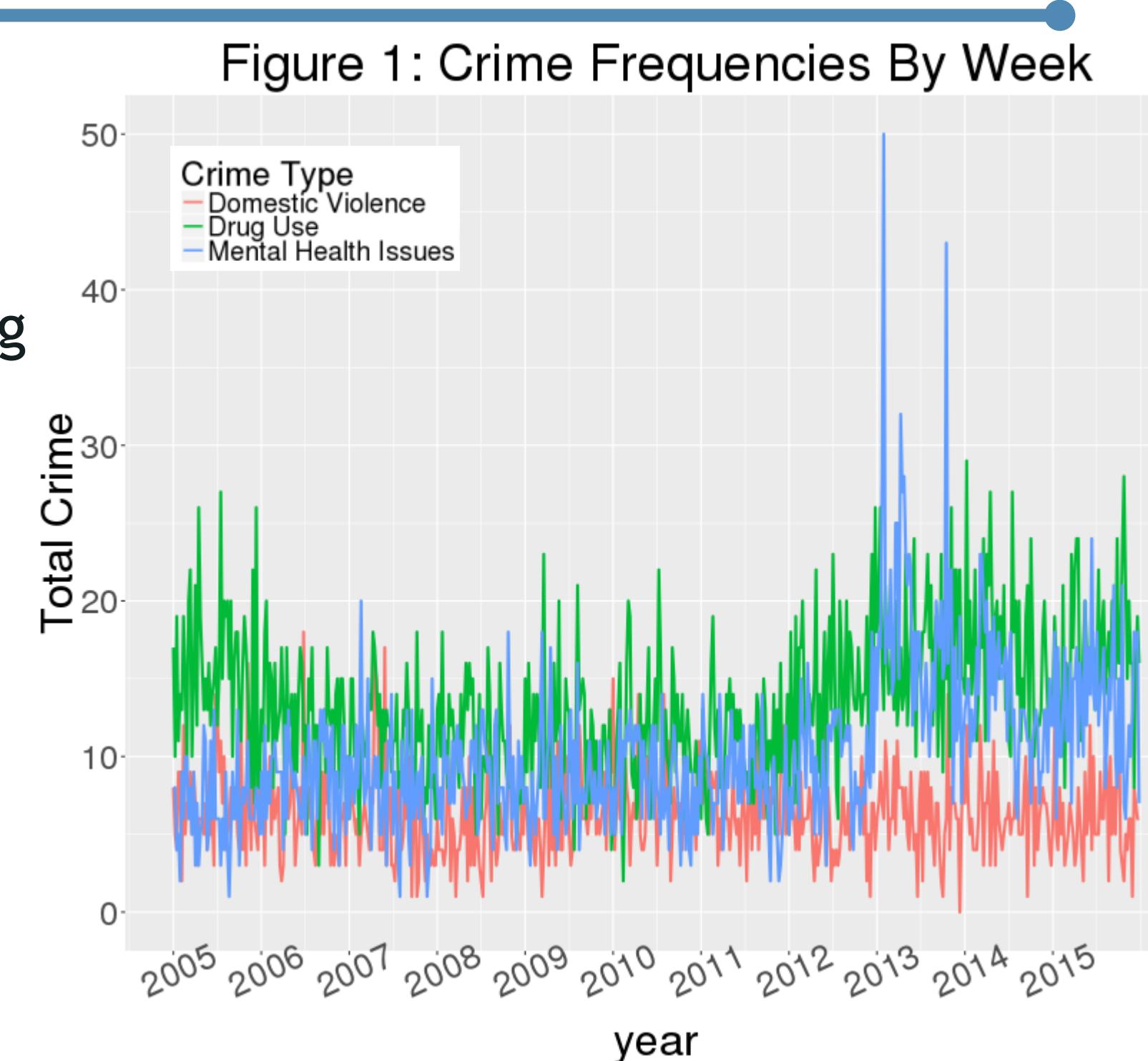
We use socioeconomic variables from American Community Survey (ACS) and POI data from various sources to characterize communities in Arlington County, VA.

## Crime Rates in Arlington County

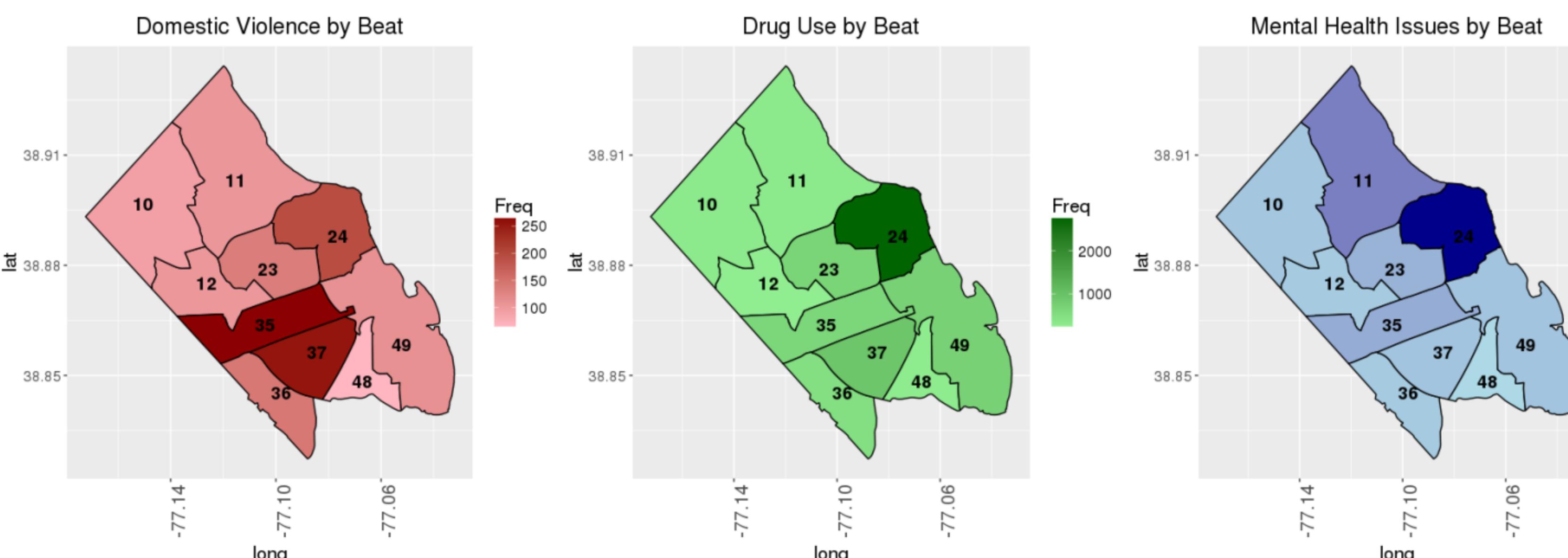
We use incident-level data provided by the Arlington County Police Department over a 10 year time window (2005-2015) for the following incident types:

- Domestic Violence
- Drug-Related
- Mental Health

We use the incidents based on the call/incident type (not the original call).



Crime Type	Selected Incident Types	Frequency	Top Original Call Types
Domestic Violence	ASSAULT FAMILY , DOMESTIC FAMILY DISTURBANCE, etc.	3,715	Assault With Injuries (32), 911 Suspicious Circumstances (30), Assault Just Occurred (20)
Drug-Related	NARCOTICS, COCAINE - SELLING, etc.	7,714	Traffic Stop (2456), Subject Stop (369), Suspicious Persons (209)
Mental Health	MENTAL CASES, ATTEMPT SUICIDE, etc.	5,867	Admin Service (1,528), Overdose (426), Check on Welfare (210)



## The Model & Findings

We develop a Poisson model for domestic violence, and Quasi-Poisson models<sup>2</sup> for drug use and mental health issues (based on the dispersion estimates).

For the Poisson model, we let  $E(y) = \text{Var}(y) = \mu$ . We assume  $y_i \sim \text{Poisson}(\mu_i)$ .

For Quasi-Poisson models, we let  $E(y) = \mu$  and  $\text{Var}(y) = \theta\mu$ . We assume that  $y_i \sim \text{Poisson}(\mu_i, \theta)$ .

For both models, we let the mean  $\mu_i$  for the  $i^{\text{th}}$  observation vary as a function of the p covariates, and the mean function is as follows:

$$\mu_i = \exp(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_p x_{p,i})$$

where socioeconomic factors, POIs, and the year are used as covariates.

The table below summarizes our model results. We report the relation of significant variables to the dependent variable: frequencies of domestic violence, drug-related crime, and mental health, respectively.

Domestic Violence	Drug-Related	Mental Health
Poisson Model (dispersion 1.37)	Quasi Poisson Model (dispersion 3.68)	Quasi Poisson Model (dispersion 7.94)
Population(+), Income per capita(-), %Married(-), %White(-), Year(-)	Income per capita(-), Poverty rate(-), %Mobility(-), %Male(-), %Married(-), %Black(+), Year(+)	Population(+), %Mobility(+), %Married(-), Year(+)
Preschool(+), Library(+), University(-)	Preschool(+), School(-), Library(+), University(+)	Preschool(+)
Homeless shelter(-), Church(+), Community/Nature center(+), County office/center, including jails(+)	County office/center(+), Fire station(-), Cooperative playgroup(-), Community/Nature center(+), Church(-), Child development center(+)	County office/center, including jails(+), Homeless shelter(+)
Public mental health facility(+), Hospital(-), Family child-care center(+)	Private mental health facility(+), Hospital(-)	Private mental health facility(+), Hospital(+), Family child-care center(+)
Park(-), Bus station(+), Restaurant(+), Post office(+)	Park(-), Bar(+), Bus station(+), Metro(+), Restaurant(+), Post office(-), Swimming pool(+)	Park(-), Metro(+), Swimming pool(+)

[1] Graif, C., Kifer, D., Li, Z., & Wang, H. (2016). Crime Rate Inference with Big Data. KDD.

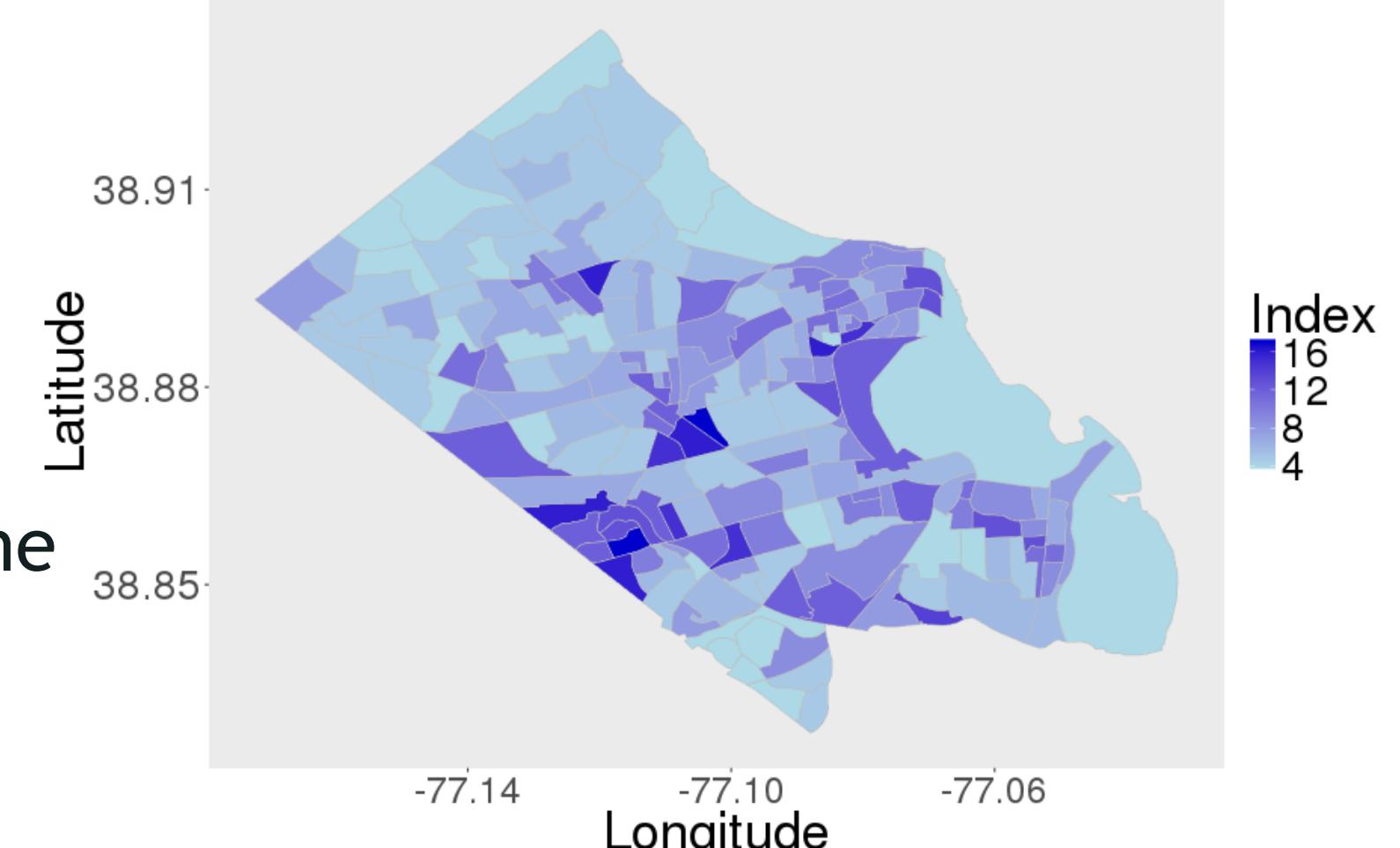
[2] Ver Hoef, J. M. and Boveng, P. L. (2007), Quasi-Poisson vs Negative Binomial Regression: How Should We Model Overdispersed Count Data?. Ecology, 88: 2766-2772. doi:10.1890/07-0043.1

## Economic Vulnerability

We develop an index of economic vulnerability to illustrate (Fig. 2) the distribution in Arlington. We use a quantile ranking of the following ACS variables:

- %population receiving food stamps
- %population with no access to a personal vehicle
- %population using over 50% of their income to pay rent/mortgage
- Poverty Rate

Figure 2: Index of Economic Vulnerability (Higher Index = more vulnerable)



## Points-of-Interest (POI)

POIs are relevant venues that provide information about neighborhood functions.

We identify 23 different types of POI using the following data sources:

- Public Arlington Open Data Portal:
  - Hospitals/Metro/Schools,
  - Childcare Centers
- Substance Abuse and Mental Health Administration (SAMHSA):
  - Public Mental Health Providers
- PsychologyToday.com:
  - Private Mental Health Providers
- Google Maps API Radar Search:
  - Bars, Churches, Restaurants, etc.

Points-of-Interest (source: Arlington Data Portal)

