Providence Crime Dataset Documentation

Felicity Hade, Data Science Fellow, Brown University Library

January 5, 2023

`https://github.com/Brown-University-Library/geodata_pvdcrime`

Introduction

GeoData@SciLi's Providence Crime dataset is an extension of the Providence Police Department's case logs. The dataset contains geocoded crime incidences across Providence from May 2023 onward, divided by year. Offenses are classified as violent crimes, property crimes, or neither. Geocoded coordinates are approximations and are not meant to represent the precise location where each crime occurred. Shapefiles use the Rhode Island State Plane coordinate system (EPSG 3438), whereas `latitude` and `longitude` column values are provided in WGS 84 (EPSG 4326).

1 Data Sources

PPD case logs
The Providence Police Department provides a 180-day summary of recent crime incidents across the city. This data is accessed regularly via API to construct the annual, geocoded output files in this repository. More information about the PPD case logs can be found at `https://data.providenceri.gov/Public-Safety/PPD-Arrest-and-Case-Logs-FAQ/4t25-ekcs/about_data`.

E-911 sites dataset
Rhode Island GIS (RIGIS) provides a point layer of E-911 sites. This dataset contains latitude and longitude coordinates in WGS 84 for points of interest across Rhode Island. E-911 files are used to calculate the coordinates of block centers in the final dataset. The point layer can be visualized and downloaded at `https://www.rigis.org/datasets/e-911-sites/explore`.

RIDOT geocoder
Intersections are geocoded using RIDOT's "Find Address Candidates" geocoder, available at `https://risegis.ri.gov/gpserver/rest/services/E911_StreetRange_Locator/GeocodeServer/findAddressCandidates`. The geocoder returns coordinates in the RI State Plane System (EPSG 3438).

Open Street Map
Open Street Map provides coordinates in WGS 84 for points of interests via manual search. Landmark coordinates are sourced here at `https://www.openstreetmap.org/`.

2 Methodology

PPD case logs are pulled via API at least every 180 days to construct an ongoing record of Providence crime.

Geocoding

Crime incidents are sorted into one of three categories according to the format of the `location` attribute:

- Block locations (e.g. "800 Block Chalkstone Ave")

- Intersections (e.g. "N Main St & Thomas St")

- Landmarks (e.g. "Roger Williams Park")

Block locations are geocoded using the E-911 site dataset. All E-911 addresses which belong to a particular block are identified and ordered by address number. The center of the block is approximated by averaging the coordinates of the block's endpoints, or the buildings with the highest/lowest address numbers on that block. Intersections are geocoded via API. Landmarks are geocoded using a manually constructed Excel spreadsheet (`landmarks.xlsx`); the spreadsheet maps landmark names to their respective coordinates and is updated as new landmarks appear.

Offense Categorization

Offenses are categorized as violent crimes, property crimes, or neither based on the value of the `offense_desc` attribute. Types of violent crime are murder and nonnegligent manslaughter, rape, robbery, and aggravated assault. Types of property crime are burglary, larceny-theft, motor vehicle theft, and arson. Offenses which are neither violent crimes nor property crimes are assigned `violent_cat` and `property_cat` values of "None". `crime_cats.xlsx` maps offense descriptions to their respective categorizations and is updated as new offense descriptions appear. This categorization scheme is meant to mirror that used by the FBI; more information about how the FBI classifies crimes can be found at `burglary`, `larceny-theft`, `motorvehicletheft`, `andarson`.