Providence Geocoded Crime Incidents Documentation

Felicity Hade, Data Science Fellow, Brown University Library

January 18, 2024

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

Introduction

GeoData@SciLi's Providence Geocoded Crime Incidents dataset is an extension of the Providence Police Department's case logs, created to facilitate basic geospatial analysis and mapping of crime incidents. The dataset contains geocoded crime incidences across Providence beginning in May 2023, with separate files available by year. Incidents that could not be geocoded are published in a separate file. **Geocoded coordinates are approximations and do not represent the precise location where each incident 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).

Disclaimer: Every effort was made to ensure that the data, which was compiled from public sources, was processed and presented accurately. The creators and Brown University disclaim any liability for errors, inaccuracies, or omissions that may be contained therein or for any damages that may arise from the foregoing. Users should independently verify the accuracy and fitness of the data for their purposes.

Known limitations of this dataset:

1. The geocoded locations are approximations that represent the middle of a block, street intersection, or landmark. They are not the precise locations of where incidents occurred.
2. Not all incidents can be geocoded, if the location information is vague, or if the location cannot be identified in the data source used for finding coordinates. This may lead to some spatial bias, where certain locations are persistently missed.
3. The incidents are not equivalent in terms of seriousness. The dataset includes everything from the police logs: serious crimes, misdemeanors and minor infractions, and calls for service.
4. The coding of incidents is not standardized or controlled. We categorized serious crimes based on the FBI classification system to facilitate basic research, but users may need to do additional cleaning, categorizing, and filtering based on their research goals.

1 Dataset Features

All information from the original case logs is retained in the final dataset. The PPD case logs contain the following features, whose descriptions were copied from the associated documentation (linked under "Data Sources" below).

Column Name	Description
casenumber	Case number assigned by Providence Police records management system
location	Street Number[1] and Street Name or nearest intersection of reported incident
reported_date	Date and Time reported to Providence Police
month	Month of Reported Date (number)
year	Year of Reported Date
offense_desc	Description of offense type based on federal guidelines
statute_code	RI law or municipal statute code being violated
statute_desc	Description of statute code
counts	Number of counts associated with an offense
reporting_officer	First Initial and Last Name of person who entered the report

[1] The PPD now uses block numbers instead of street numbers.

The following features have been added to the dataset by the GeoData@SciLi team. Note that `latitude`, `longitude`, and `source` are only present for successfully geocoded crimes.

Column Name	Description
unique_id	Unique ID associated with the offense
latitude	Approximate latitude where the offense was committed (WGS 84)
longitude	Approximate longitude where the offense was committed (WGS 84)
source	Source of coordinates
violent_cat	Category of violent crime
property_cat	Category of property crime

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, 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 mirror that used by the FBI; more information about how the FBI classifies crimes can be found at `https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/topic-pages/offenses-known-to-law-enforcement`.
Please note that errors in the categorization process are possible due to limited or non-standardized offense information.

Unique IDs

Multiple offenses may belong to the same case if multiple laws are violated in a single incident. Each row of the dataset represents a single offense; therefore, case numbers are not unique identifiers. Unique IDs are assigned to an offense by adding three digits to the end of its case number. The first offense associated with a given case number is assigned a three-digit code of 001, the second uses a code of 002, and so on.

3 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. Documentation for the PPD case logs is available 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, a comprehensive dataset of addresses and other point locations that are used to coordinate emergency response. 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, which returns coordinates in the RI State Plane System (EPSG 3438). The geocoder is available at `https://risegis.ri.gov/gpserver/rest/services/E911_StreetRange_Locator/GeocodeServer/findAddressCandidates`.

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/`.