

Access to Raw Data for Thesis: ‘Bayesian Analysis of Spatial Log-Gaussian Cox Processes’

Nadeen Khaleel

April 20, 2022

Introduction

This document details the access to the raw data that was used within the thesis ‘Bayesian Analysis of Spatial Log-Gaussian Cox Processes’ where the aggregated data, simulated data and code were archived, the raw could not be. Therefore, we provide details on access to the crime and socio-economic data called within the code. Some data will have been updated since the original data of access (discussed in the tables below), and so there may be differences with some changes occurring in the column names and organisation as well. Where necessary, we discuss how we accessed the data, and how it can be accessed now since, for example, the access for the Los Angeles 2015 data was made through the American FactFinder through the US Census Bureau website, which was retired not long after.

Tables 2, 3 and 4 are also discussed in Appendix F of the thesis.

Crime

The crime data was accessed through the Police Data Initiative, *Police Data Initiative (PDI)*, which linked to relevant departments for each city on interest.

1. Go to PDI website, policedatainitiative.org
2. Select “Data & Agencies” → “Datasets” and scroll to “All Datasets”
3. In search bar, search for the relevant city, e.g. “Los Angeles”
4. Select relevant link for data, e.g. “Incidents”. See tables below for the specific link for each cities data of interest and the website where the data is accessed.

Note that crime data may have been updated since the date of access highlighted in Tables 2, 3 and 4, and so there may be slight differences in the data. Additionally, for Portland city the ability to download a single file for the data is no longer available. Now we must select to download data for a particular year.

From these crime data sets we extracted the homicide and motor vehicle theft data only regardless of the year, removing any that had locations outside the city of interest, possibly due to erroneous input. These crime locations were used for some inference but were also aggregated into count data over the census tracts of the cities as well as onto discretisation grids of different resolutions.

Socio-Economic Variables

Census Tract level socio-economic data was accessed through the U.S. Census Bureau, American Community Survey (ACS), while the boundaries for this geographic support was also accessed through the U.S. Census Bureau, U.S. Department of Commerce, U.S. Census Bureau, Geography Division [2015].

It is important to note that the Los Angeles American Community Survey (ACS) data for the 2015 covariates were accessed through the American Fact Finder (AFF) Guided Search feature through the US Census Bureau (https://factfinder.census.gov/faces/nav/jsf/pages/guided_search.xhtml), however, this was retired in March 2020, and so while the AFF has been linked to in the below table, these links will not link to a webpage. The 2014 Los Angeles ACS data as well as the 2015 ACS data for New York and Portland were accessed through the new US Census Bureau API, with links provided. The access to the 2015 ACS data for Los Angeles can also be made through the new US Census Bureau API, through the following steps (as for the 2014 data and New York and Portland 2015 data):

1. Go to the US Census Bureau, <https://www.census.gov/>
2. Explore Data → “Explore Data Main”
3. Select: “Go to data.census.gov”
4. Select: “Advanced Search”
5. Search for the data table using the identifier, such as “B01003” for the total population
6. In filters:
 - (a) Geography
 - (i) Select “Tract”
 - (ii) Select relevant state, e.g. “California”
 - (iii) Select relevant counties, e.g. “Los Angeles County, California”
 - (iv) Select “All Census Tracts within *”, e.g. “All Census Tracts within Los Angeles County, California”. If the city of interest lies within multiple counties, you must select multiple counties as described to get all relevant census tracts. Table 1 contains the list of counties of interest for each US City we consider within this thesis.
 - (b) Years
 - (i) Select relevant year, e.g. “2015”
7. Then at the top of the page, select “Tables” to get relevant table results.

Note that there are additional differences in the output data between the LA 2015 from the AFF and the 2014 LA and 2015 Portland and New York, with respect to the locations of the necessary columns for the data. Additionally, the total population tables have two possible options, select “20X: ACS 5-Year Estimates Detailed Tables”, where X denotes the year of interest.

For interest, we set out the following steps to explain the original access for some of the socio-economic variables for Los Angeles from the American Fact Finder.

1. Go to the US Census Bureau, <https://www.census.gov/>
2. Explore Data → “Explore Data Main”
3. Select “Data Tools and Apps”

4. Select “American FactFinder”
5. Select “Guided Search”
6. Select “I’m looking got information about people”, select “Next”
7. For example, for:
 - (a) Total Population: Select “People” → “Basic Count/Estimate” → “Population Total”, select “Next”
 - (b) Average Income: Select “People” → “Income & Earnings” → “Income/Earnings (Individuals)”, select “Next”
8. Under Geographic type select “Census Tract - 140”
9. Under Select a state, search for required state, e.g. “California”
10. Under Select a county, search for required county, e.g. “Los Angeles County”
11. Select “All Census Tracts within *,*”, e.g. “All Census Tracts within Los Angeles County, California”, select “Add to your selections”
12. For “Race/Ethnic Groups”, select “Skip this step”
13. Select the relevant data set for the years required, such as “TOTAL POPULATION” or “MEAN INCOME IN THE PAST 12 MONTHS (IN 2015 INFLATION-ADJUSTED DOLLARS)” for 2015.

City	Counties the City Lies Within
Los Angeles City	Los Angeles County
New York City	New York County (Manhattan), Bronx County, Kings County (Brooklyn), Queens County and Richmond County (Staten Island)
Portland	Multnomah County, Washington County and Clackamas County

Table 1: Counties that contain Census Tracts for US Cities

Using the results from the shapefile extraction below, the values for the variables for the relevant census tracts were extracted and missing data was imputed, in particular for the average income variables. For the imputation, neighbourhood matrices were used and the missing data was taken to be the average of the values of the variable in the neighbouring census tracts.

Note: in addition to the Total Population (B01003) and Average Household Income (S1902) variables, we also accessed files for Age and Sex (S0101), Food Stamps/SNAP (S2201), Tenure (B25003). Except for the 2015 Los Angeles Food Stamps and Tenure data, these were accessed at the same time as the other variables (as detailed in Tables 2, 3 and 4) for the relevant city and year. Where 2015 LA Food Stamps and Tenure data were instead accessed on 26/09/2019.

Shapefiles

For the TIGER/Line Shapefiles for the Census Tract Shapefiles, we can find the relevant state’s census tracts through the following:

1. Go to the US Census Bureau, <https://www.census.gov/>
2. Browse by Topic → “Geography”
3. Select: “TIGER/Lines Shapefiles”
4. Select relevant year, e.g. “2015”
5. Select one of either “Web Interface” or “FTP Archive”
 - (a) “Web Interface”
 - (i) Select relevant year from “Select year” drop down, e.g. “2015”
 - (ii) Select “Census Tracts” from “Select a layer type” drop down
 - (iii) Select “Submit”
 - (iv) Select relevant state from “Select a State drop down”, e.g. “California”
 - (v) Select “Download”
 - (b) “FTP Archive”
 - (i) Select “TRACT”
 - (ii) Select `t1_20X*_tract.zip` where X denotes the year and * denotes the relevant FIPs code for the state of interest, e.g. for California in 2015 `t1_2015_06_tract.zip`

The city boundaries can be downloaded through the links provided in Tables 2, 3 and 4.

From these shapefiles, the census tract shapefiles data for only the relevant cities were extracted to generate the census tract count data as well as to produce the aggregated count data over grids. For each city the individual census tracts were additionally merged to form single polygonal windows that were used in our inference, such as producing a mesh for the INLA-SPDE algorithm. Additionally, the city boundary shapefiles were used to remove any data points from the crime data that were not contained within the city of interest, possibly due to erroneous input.

	Name	Source	Date Downloaded	URL	Notes
LA Data					
Crime Point Pattern	“Crime Data from 2010 to Present”	Los Angeles Police Department (http://data.lacity.org) through the PDI	18/11/2018 (last updated: 13/11/2018)	https://data.lacity.org/A-Safe-City/Crime-Data-from-2010-to-Present/y8tr-7khq	Through PDI, search for Los Angeles Police Department then select Incidents. Select “Crime Data from 2010 to Present”. Now data can possibly be found under “Crime Data from 2010 to 2019”. At time of access, licence was CC0.
Shapefiles: Boundary	Los Angeles County Boundary	Los Angeles County Department of Public Works	06/12/2018	https://controllerdata.lacity.org/dataset/City-Boundaries-for-Los-Angeles-County/sttr-9nxz	Select ‘Export’ and under ‘Download Geospatial Data’ select ‘Shapefile’.
Shapefiles: CT	Los Angeles City Census Tracts	TIGER/Line through US Census Bureau	24/09/2019	https://www2.census.gov/geo/tiger/TIGER2015/TRACT/	FIPs code for California is ‘06’

Total Population (2015) on Census Tracts	“B01003 TOTAL POPULATION”	American Community Survey (ACS) through American FactFinder from the US Census Bureau	04/09/2019	https://factfinder.census.gov/faces/nav/jsf/pages/guided_search.xhtml	American FactFinder was retired end of March 2020 so URLs are not longer correct
Average Income (2015) on Census Tracts	“S1902 MEAN INCOME IN THE PAST 12 MONTHS (IN 2015 INFLATION-ADJUSTED DOLLARS)”	American Community Survey (ACS) through American FactFinder from the US Census Bureau	04/09/2019	https://factfinder.census.gov/faces/nav/jsf/pages/guided_search.xhtml	American FactFinder was retired end of March 2020 so URLs are not longer correct
Total Population (2014) on Census Tracts	“B01003 TOTAL POPULATION”	American Community Survey (ACS) through the US Census Bureau API	24/07/2020	https://data.census.gov/cedsci/table?g=0500000US06037.140000&t=Population\%20Total&y=2014&tid=ACSDT5Y2014.B01003&hidePreview=false&vintage=2014&layer=VT_2014_040_00_PP_D1&cid=DP05_0001E	

Average Income (2014) on Census Tracts	“S1902 MEAN INCOME IN THE PAST 12 MONTHS (IN 2014 INFLATION-ADJUSTED DOLLARS)”	American Community Survey (ACS) through the US Census Bureau API	24/07/2020	https://data.census.gov/cedsci/table?g=0500000US06037.140000&t=Income%20%28Households,%20Families,%20Individuals%29&y=2014&tid=ACSST5Y2014.S1902&vintage=2014&hidePreview=false&cid=S1901_C01_001E	
--	--	--	------------	---	--

Table 2: Details about Los Angeles crime and socio-economic variables data as well as shapefiles data.

	Name	Source	Date Downloaded	URL	Notes
NYC Data					
Crime Point Pattern	“NYPD Complaint Data Historic”	New York Police Department through the PDI	18/11/2018 (last updated: 16/11/2018)	https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i	Through PDI, search for New York Police Department then select Incidents. Select “NYPD Complaint Data Historic”.
Shapefiles: Boundary	Borough Boundaries NYC	Department of City Planning (DCP)	06/12/2018 (last updated: 09/11/2016)	https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm	Select ‘Export’ and under ‘Download Geospatial Data’ select ‘Shapefile’.
Shapefiles: CT	New York City Census Tracts	TIGER/Line through US Census Bureau	24/09/2019	https://www2.census.gov/geo/tiger/TIGER2015/TRACT/	FIPs code for New York is ‘36’

Population (2015) on Census Tracts	“B01003 TOTAL POPULATION”	American Community Survey (ACS) through the US Census Bureau API	15/06/2020	https://data.census.gov/cedsci/table?g=0500000US36061.140000,36047.140000,36005.140000,36085.140000,36081.140000&tid=ACSDT5Y2015.B01003&vintage=2015&y=2015&t=Population\%20Total&hidePreview=false&layer=VT_2015_140_00_PY_D1&cid=S0101_C01_001E	
Average Income (2015) on Census Tracts	“S1902 MEAN INCOME IN THE PAST 12 MONTHS (IN 2015 INFLATION-ADJUSTED DOLLARS)”	American Community Survey (ACS) through the US Census Bureau API	15/06/2020	https://data.census.gov/cedsci/table?g=0500000US36061.140000,36047.140000,36005.140000,36085.140000,36081.140000&y=2015&t=Income\%20and\%20Earnings&cid=S1901_C01_001E&tid=ACSST5Y2015.S1902&vintage=2015&hidePreview=false	

Table 3: Details about New York City crime and socioeconomic variables data as well as shapefiles data.

	Name	Source	Date Downloaded	URL	Notes
Portland Data					
Crime Point Pattern	"Portland Police Bureau - Incidence"	Portland Police Bureau through the PDI	26/11/2018	https://www.portlandoregon.gov/police/71978	Through PDI, search for Portland Police Bureau then select Incidents. Then select "Download Open Data". Note that due to changes, the open crime data can now only be downloaded on a year-by-year basis.

Shapefiles: Boundary	Portland City Boundary	Metro Data Re- source Center	06/12/2018	http://gis-pdx. opendata.arcgis. com/datasets/city- boundaries?geometry=- 124.785\%2C44.956\%2C- 120.665\%2C46.301	This database and GIS data are made available under the Open Database License (ODbL v1) (https://opendatacommons.org/licenses/odbl/1-0/) and Database Content License (DbCL v1) (https://opendatacommons.org/licenses/dbcl/1-0/) at last access to the metadata, 22/10/2021. On left-hand menu, third item down should open a menu for downloads, scroll to 'Shapefile' and download from there.
Shapefiles: CT	Portland City Census Tracts	TIGER/Line through US Census Bureau	24/09/2019	https://www2.census. gov/geo/tiger/ TIGER2015/TRACT/	FIPs code for Portland is '41'

Population (2015) on Census Tracts	“B01003 TOTAL POPULATION”	American Community Survey (ACS) through the US Census Bureau API	13/06/2020	https://data.census.gov/cedsci/table?layer=VT_2015_140_00_PY_D1&g=0500000US41051.140000,41067.140000,41005.140000&vintage=2015&y=2015&tid=ACSDT5Y2015.B01003&t=Population\%20Total&hidePreview=false&cid=S0101_C01_001E	
Average Income (2015) on Census Tracts	“S1902 MEAN INCOME IN THE PAST 12 MONTHS (IN 2015 INFLATION-ADJUSTED DOLLARS)”	American Community Survey (ACS) through the US Census Bureau API	13/06/2020	https://data.census.gov/cedsci/table?layer=VT_2018_140_00_PY_D1&g=0500000US41051.140000,41067.140000,41005.140000&y=2015&t=Income\%20and\%20Earnings&tid=ACSST5Y2015.S1902&vintage=2015&hidePreview=false&cid=S1901_C01_001E	

Table 4: Details about Portland crime and socio-economic variables data as well as shapefiles data.

References

- American Community Survey (ACS). U.S. Census Bureau, *2011-2015 American Community Survey 5-Year Estimates*. U.S. Census Bureau. URL <https://data.census.gov/cedsci/advanced>. U.S. Census Tracts ACS Data: Los Angeles, New York and Portland.
- Department of City Planning (DCP). New York City Borough Boundary. New York City Department of City Planning. URL <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm>. Borough Boundaries NYC Shapefiles.
- Los Angeles County Department of Public Works. City Boundaries for Los Angeles County. Los Angeles City Controller's Office. URL <https://controllerdata.lacity.org/dataset/City-Boundaries-for-Los-Angeles-County/sttr-9nxz>. Los Angeles City's Boundary Shapefiles.
- Metro Data Resource Center. City Boundaries. Metro Data Resource Center, Metro GIS. URL <http://gis-pdx.opendata.arcgis.com/datasets/city-boundaries?geometry=-124.785%2C44.956%2C-120.665%2C46.301>. Portland City Boundary Shapefiles.
- Police Data Initiative (PDI)*. Police Data Initiative (PDI). URL <https://www.policedatainitiative.org/>. PDI Data Sets for All Participating Law Enforcement Agencies.
- U.S. Department of Commerce, U.S. Census Bureau, Geography Division. Tiger/line shapefile, 2015. prepared by the U.S. Census Bureau, 2015. URL <https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2015.html>. Los Angeles, New York and Portland Census Tracts Shapefiles.