Candidate Assignment

Chhandara Pech | October 6, 2022

This document details the methodology used to construct a *Crime + Poverty* index which can be utilized to assist the City of San Diego in identifying and prioritizing neighborhoods for streetlight repairs. Preliminary analysis conducted by the author on open cases in the City of San Diego found that residents in higher crime and higher poverty neighborhoods have, on average, longer wait days for repair compared to residents in lower crime and lower poverty neighborhoods. The differences were stark and Figure 1 and 2 display preliminary results. (For reference, the City's current average wait days for repair completion is 261.8 days).

Figure 1. Average Number of Days Since Case Submission by Neighborhood Crime Rate

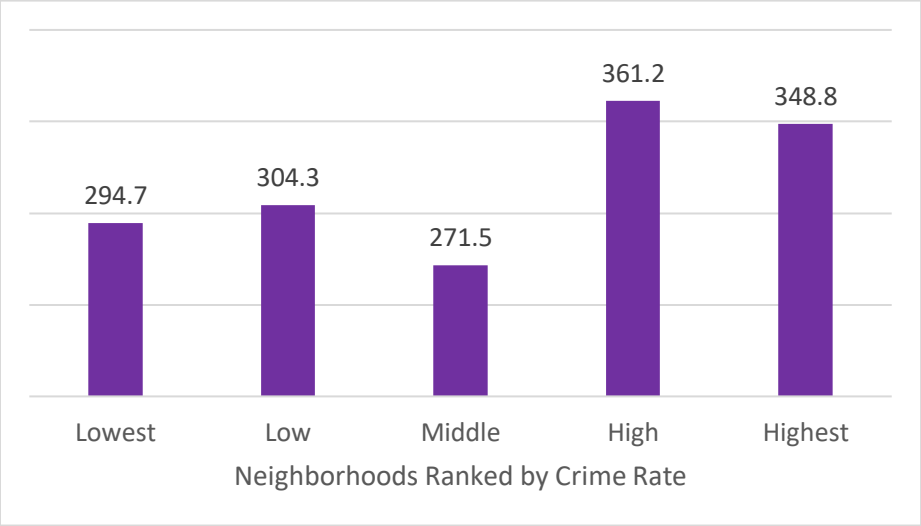
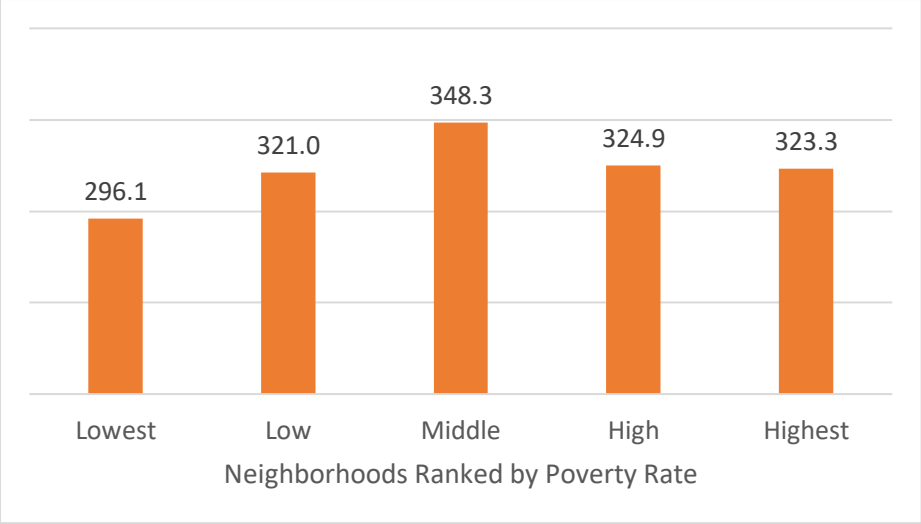


Figure 2. Average Number of Days Since Case Submission by Neighborhood Poverty Rate



Given these disparities, it is recommended that the City prioritize cases by considering neighborhood factors such as public safety and equity. This document details a potential indicator, *Crime + Poverty* index, that can be integrated into the relational database used by the City to process streetlight repairs.

Required software: Microsoft Excel, Statistical Software (e.g. SAS, R), and GIS (e.g. ArcGIS, QGIS)

Key Datasets: Streetlight Maintenance Open Cases, American Community Survey 5-year average, and Crime Data

Data and Methodology

The *Crime + Poverty* index utilizes two dimensions to identify priority neighborhoods. High priority neighborhoods are those with: 1) disproportionately high rates of crime, and 2) high rates of poverty.

Geographic Unit of Analysis

The basic geographic unit of analysis for *Crime + Poverty* index is the census tract, which serves as a reasonable proxy for neighborhood bounds. The Bureau of the Census defines census tracts as "a relatively homogenous area with respect to population characteristics, economic status and living conditions." The average population of a census tract is 4,000 people (ranging from 2,500 to 8,000) and approximately 1,500 housing units. Census tracts are geographic units embedded within a county and are smaller than ZIP Code Tabulation Areas (ZCTA)/ZIP Codes.

Operationalizing the *Crime + Poverty* Index

The *Crime + Poverty* index utilizes two components: crime rate and poverty rate. A description of how each component is constructed is discussed below.

Crime Rate

Constructing a measure of crime rate at the census tract level involves multiple steps that can largely be divided into four major parts. The first is downloading and assembling the crime data. The second is cleaning and geocoding the data. The third is spatially joining the data to census tracts and summarizing the number of crimes per tract. The final step is constructing the crime rate. Each step is furthered detailed below.

<u>Downloading/Assembling Crime Data</u>

The first step involves downloading and assembling the crime dataset. Crime data comes from the San Diego Association of Governments (SANDAG). The data is publicly available and can be downloaded from the Automated Regional Justice Information System (ARJIS), a division of SANDAG's Technical Services Department.[1] The data represent crimes that occurred in the county of San Diego within the last 180 days. (The data is the same crime data that is displayed on crimemapping.com).

[1] To access and download the crime dataset, visit: https://www.sandag.org/index.asp?classid=14&subclassid=21&projectid=446&fuseaction=projects.detail

Crimes that are shown in the dataset are limited to: arson, assault, burglary, drugs/alcohol violations, DUI, fraud, homicide, motor vehicle theft, robbery, theft/larceny, vandalism, vehicle break-in/theft, and weapons. Sensitive data such as domestic violence, suicides and sex crimes are excluded.

To construct the *Crime + Poverty* index, a decision needs to be made as to which crimes should be captured in the index. For the purpose of this assignment, it is recommended that we only focus on violent and property crimes. Table 1. reports the frequency of each crime type in the dataset downloaded from SANDAG. It further notes which crimes were considered as violent crimes and which were property crimes. The remaining crimes are categorized into a residual category called "Other".

Table 1. Crime Types in SANDAG's Dataset

CM_LEGEND							
CM_LEGEND	Frequency	Percent	Cumulative Frequency	Cumulative Percent	Violent	Property	Other
ARSON	160	0.29	160	0.29		X	
ASSAULT	8,617	15.46	8,777	15.74	X		
BURGLARY	3,443	6.18	12,220	21.92		X	
DRUGS/ALCOHOL VIOLATIONS	10,520	18.87	22,740	40.79			X
DUI	2,316	4.15	25,056	44.95			X
FRAUD	2,922	5.24	27,978	50.19			X
HOMICIDE	44	0.08	28,022	50.27	X		
MOTOR VEHICLE THEFT	5,317	9.54	33,339	59.81		X	
ROBBERY	1,246	2.24	34,585	62.04	X		
SEX CRIMES	1,204	2.16	35,789	64.2			X
THEFT/LARCENY	8,216	14.74	44,005	78.94		X	
VANDALISM	4,563	8.19	48,568	87.13		X	
VEHICLE BREAK-IN/THEFT	6,225	11.17	54,793	98.29		X	
WEAPONS	952	1.71	55,745	100			X

Geocode Crime Data

The second step involves geocoding the location of each crime data. Geocoding is a process by which a physical address is converted into geographic coordinates (longitude and latitude or x and y coordinates). This can be done using a GIS software (e.g. ArcGIS, QGIS). The addresses represent where the crime occurred (some addresses may be approximated). Table 2 provides a snapshot of the crime dataset from SANDAG. It displays the variables included in the dataset. Some addresses may need to be "cleaned" and formatted for geocoding. (For example, in the screenshot displayed in Table 2, the word "Block" is contained in each record, and should be removed for geocoding). Some observations may not be able to geocode. Most of these occurrences are usually due to errors in the address such as street names being spelled wrong, an extra digit in the address number etc. As such, multiple rounds of cleaning and geocoding may be required to successfully geocode each observation.

Table 2. Snapshot of SANDAG's Crime Data

ID	CM_LEGEND	agency	Address	ZipCode	City	State
1	FRAUD	SAN DIEGO	4600 BLOCK MOUNT BIGELOW DRIVE	92111	SAN DIEGO	CA
2	VEHICLE BREAK-IN/THEFT	SAN DIEGO	800 BLOCK B STREET	92101	SAN DIEGO	CA
3	SEX CRIMES	CHULA VISTA	600 BLOCK 03RD AVENUE	91910	CHULA VISTA	CA
4	MOTOR VEHICLE THEFT	SAN DIEGO	4100 BLOCK IDAHO STREET	92104	SAN DIEGO	CA
5	ROBBERY	SAN DIEGO	1200 BLOCK IMPERIAL AVENUE	92101	SAN DIEGO	CA
6	VEHICLE BREAK-IN/THEFT	CHULA VISTA	0 BLOCK (1100 BLOCK) OCALA AVENUE	91911		CA
7	VANDALISM	SHERIFF	2500 BLOCK SWEETWATER ROAD	91977	SPRING VALLEY	CA
8	VANDALISM	SHERIFF	100 BLOCK AMMUNITION ROAD	92028	FALLBROOK	CA
9	DRUGS/ALCOHOL VIOLATIONS	SAN DIEGO	300 BLOCK 05TH AVENUE	92101	SAN DIEGO	CA
10	ASSAULT	SAN DIEGO	900 BLOCK 06TH AVENUE	92101	SAN DIEGO	CA
11	VANDALISM	OCEANSIDE	1200 BLOCK MISSOURI AVENUE	92054	OCEANSIDE	CA
12	BURGLARY	SAN DIEGO	14500 BLOCK PENASQUITOS DRIVE	92129	SAN DIEGO	CA
13	DUI	ESCONDIDO	1200 E BLOCK GRAND AVENUE	92027	ESCONDIDO	CA
14	VANDALISM	SAN DIEGO	300 S BLOCK MEADOWBROOK DRIVE	92114	SAN DIEGO	CA
15	ROBBERY	SAN DIEGO	10700 BLOCK WESTVIEW PARKWAY	92126	SAN DIEGO	CA
16	DRUGS/ALCOHOL VIOLATIONS	SHERIFF	1000 BLOCK 12TH STREET	91932	IMPERIAL BEACH	CA

Spatially Assign Crime Data to Census Tracts and Summarized

Once geocoded, each crime data is spatially joined to their corresponding census tract using the geographic coordinates that was generated from the geocoding process. With each crime now assigned to a census tract, we can summarize the total number of crimes in each tract. This step can be accomplished using a statistical software (e.g. SAS, R). In the end, each tract should now have information on the total number of crimes that occurred in each tract. We then merge population data from the American Community Survey into the summarized dataset by tracts and then calculate the crime rate.

Calculate Crime Rate

The crime rate is read as X number of crimes per 1,000 people and each tracts' crime rate is calculated as follows:

$$\frac{\text{(Total crime)}}{\text{(Total Population/1000)}}$$

Total crime = violent crime + property crime

Poverty Rate

The second component needed for the *Crime + Poverty* index is the neighborhood's poverty rate. Data on poverty rates at the tract level comes from the 5-year American Community Survey (tract level data is only reported in the 5-year ACS due to sample size concerns). The current 5-year ACS is the 2016-2020 dataset. It should be noted that the 2016-2020 dataset is the first 5-year ACS that uses the new 2020 Census boundary (census tract boundaries change every 10 years with the Census enumeration).

The poverty rate is the ratio of the number of people whose income falls below the federal poverty line (FPL). The poverty rate is calculated as follows:

$$\frac{\text{Number of people below FPL}}{\text{Number of people for whom poverty status is determined}}$$

Table # C17002: "Ratio Of Income To Poverty Level In The Past 12 Months" from the ACS includes the necessary variables needed to calculate the neighborhood's poverty rate.

Crime + Poverty Index

To generate the *Crime + Poverty* index, we first transform the two components (crime rate and poverty rate). The individual components tend to be nonlinear and skewed and have different coefficients of variance (a measure of the spread in value across tracts); therefore, each variable is transformed into ordinal ranking. Each component has the same weight, and the two rankings are summed up to produce an overall *Crime + Poverty* score.

Neighborhoods with the highest score, that is, those with high crime rates and high poverty rates, should be prioritized for streetlight maintenance repair.

Illustration Using ZCTAs/ZIP Codes

To illustrate how the *Crime + Poverty* index can be used to identify neighborhoods for prioritization, examples are provided using ZCTAs/ZIP Codes as the geographic unit of analysis. As previously mentioned, it is recommended that the index be constructed at the census tract level because of heterogeneity within ZCTA/ZIP codes, and to help further narrow prioritization down to a smaller geographic unit. ZCTA/ZIP codes are geographically larger than tracts. (Constructing the index at the tract level for this assignment would require cleaning and geocoding over 50,000 records from the crime dataset. ZIP Code information was already included in the dataset and was used to summarized the number of crimes by ZIP code).

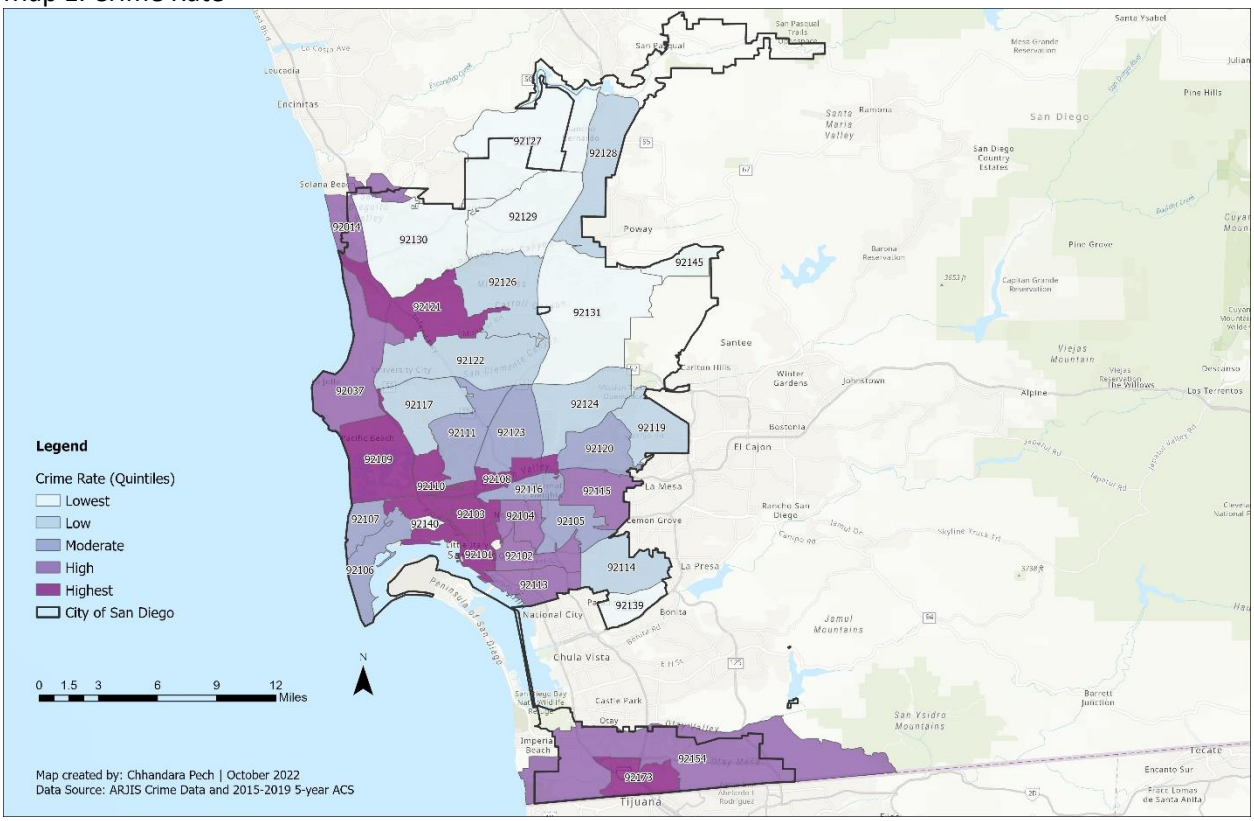
For analytical purposes, the City of San Diego's ZCTAs/ZIP codes are assigned into five hierarchical groups (quintiles) based on their *Crime + Poverty* score. The ranking ranges from neighborhoods with the lowest score to the highest score. Each group, or quintile, includes roughly 20 percent of all ZCTAs/ZIP codes in the city. Table 3 reports the neighborhood's averages (mean) of the of the two variables used to generate the *Crime + Poverty* index. As expected, higher priority neighborhoods have more crime per population and more residents who live in poverty. Although not included in the index, Table 3 also shows that the share of people of color (those who do not identify as non-Hispanic white) is higher in priority neighborhoods.

Table 3. Components of the Crime + Poverty Index

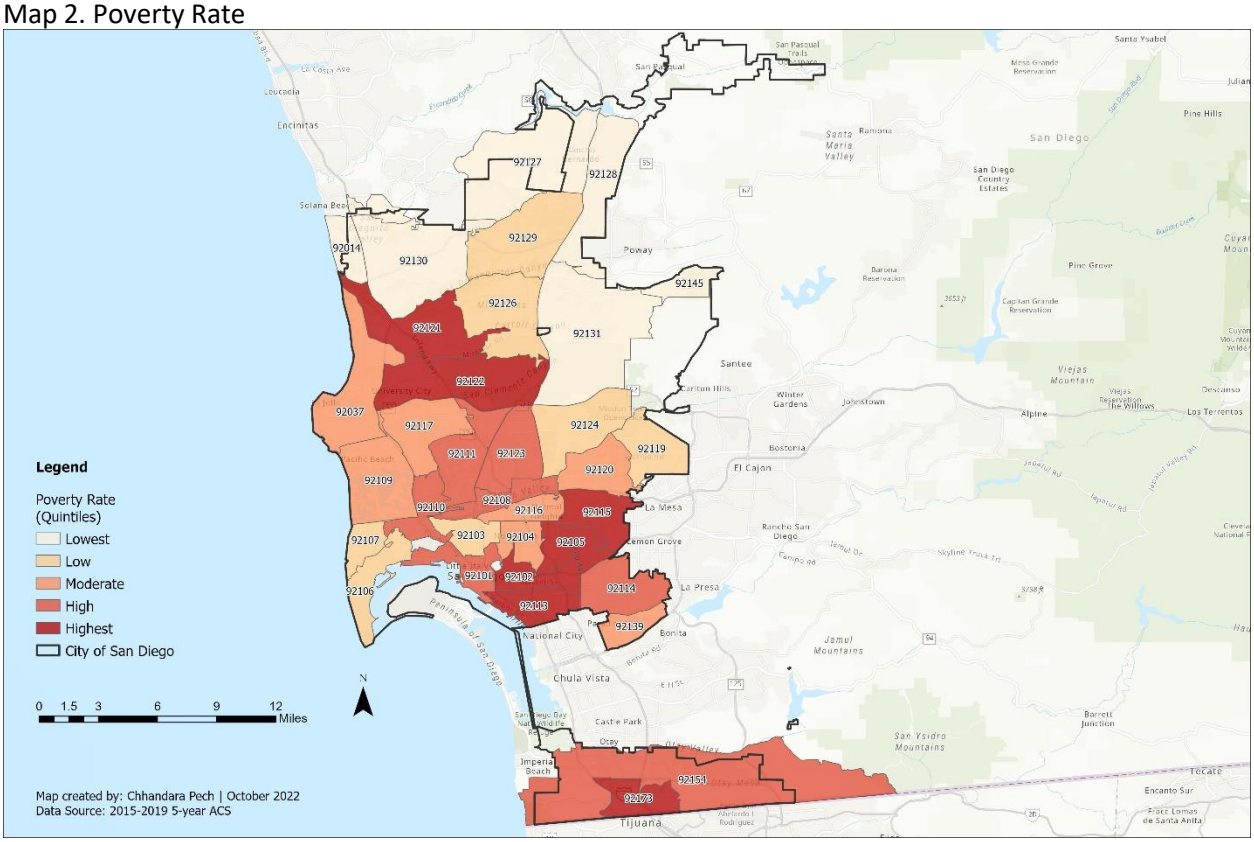
				Not included in index
Neighborhoods Ranked by C+P Score	**Crime + Poverty Score**	**Crime Rate (per 1K people)**	**% Poverty**	**% People of Color**
Lowest	10	4	4%	47%
Low	24	8	7%	46%
Moderate	36	10	11%	47%
High	48	15	16%	59%
Highest	59	35	19%	60%

The following maps displays the components of the *Crime + Poverty* index at the ZCTA/ZIP code level.

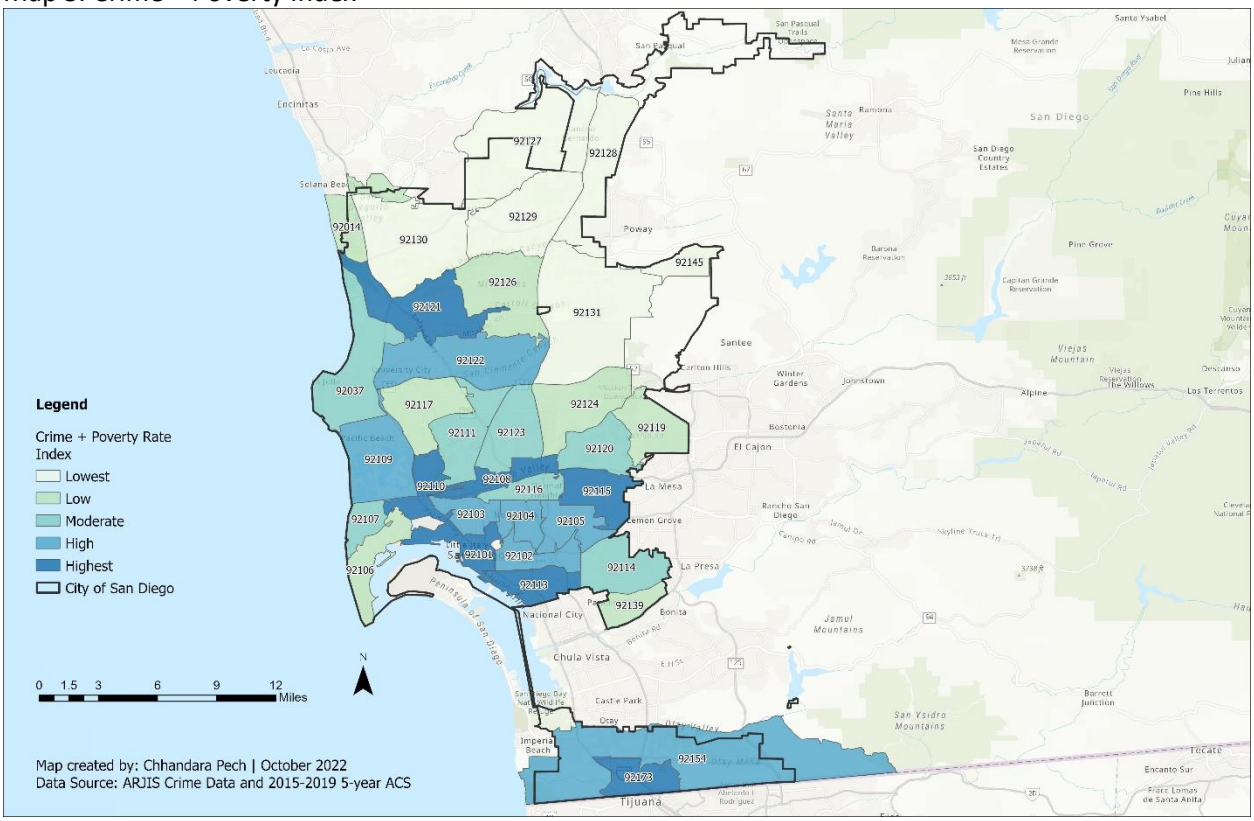
Map 1. Crime Rate



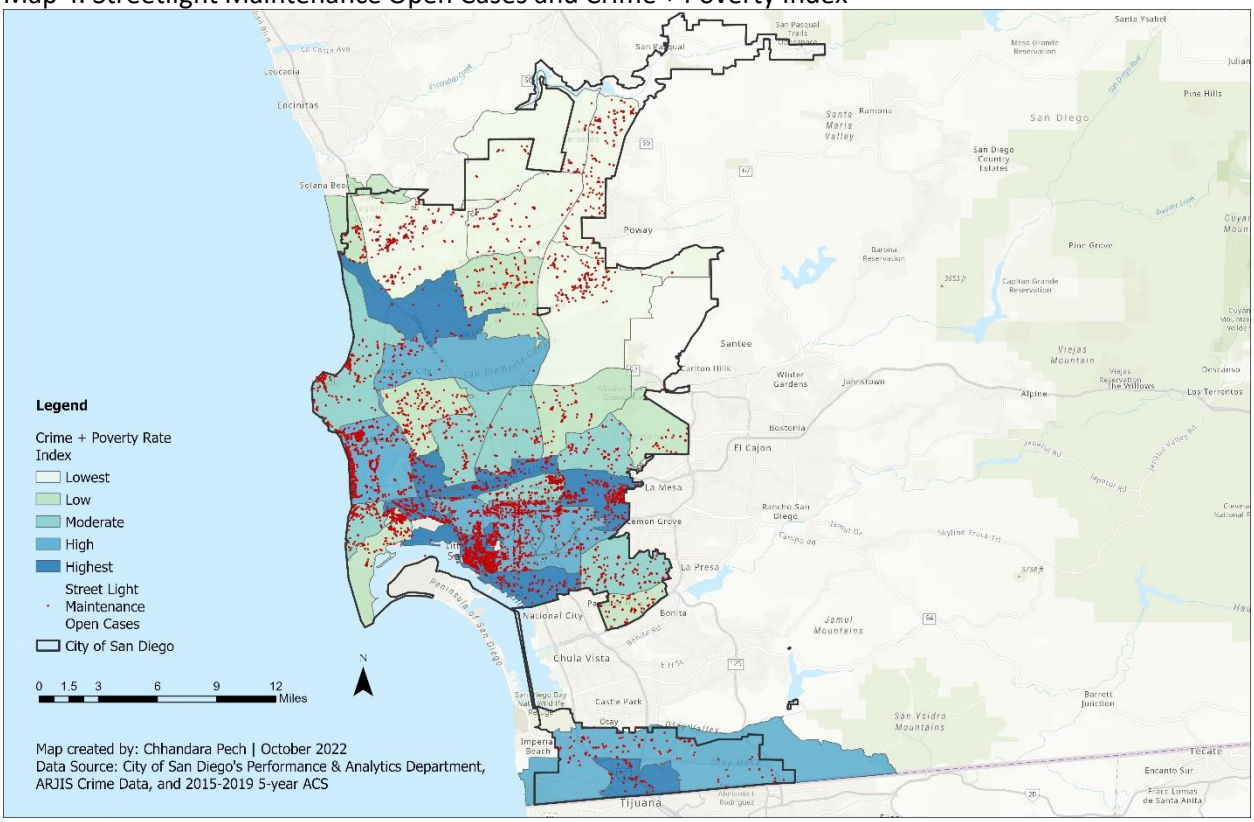
Map 2. Poverty Rate



Map 3. Crime + Poverty Index



Map 4. Streetlight Maintenance Open Cases and Crime + Poverty Index



Recommendations

- Incorporate public safety and equity measures to prioritize open cases for streetlight repairs
    - Develop a *Crime + Poverty* index at the census tract level for all tracts in the county; rank neighborhoods by *Crime + Poverty* index from lowest to highest score
- The City should prioritize cases according to the neighborhood that yields the most personal and societal benefits
    - This step will integrate the neighborhood information into the relational database used by the City to process streetlight repairs
- Prioritize cases that are in areas with high crime and high poverty
    - In addition to prioritizing neighborhoods by *Crime + Poverty* index, the City can further prioritize open cases by how many days since opened
- Monitor and perform near real-time assessment
    - This will enable the City to determine if the pilot program is meeting its targets in terms of addressing streetlight repairs in priority neighborhoods. If the assessment identifies any deficits relative to desired targets, then the City can make mid-course programmatic changes to address the problem.