SPATIALLY ADJUSTED CORRELATIONS BETWEEN CRIME AND TOPOGRAPHIC VARIABLES: RELIEF, ROUGHNESS, AND DISTANCE TO THE NEAREST MOUNTAIN

A THESIS

PRESENTED TO THE

DEPARTMENT OF EARTH, ATMOSPHERIC,

AND GEOGRAPHIC INFORMATION SCIENCES

WESTERN ILLINOIS UNIVERSITY

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR DEPARTMENTAL HONORS

EMILY RICHARDSON

FALL 2024

Contents

Abstract

This study examines global and local spatially adjusted correlations between overall crime rates and three topographic variables—relief, elevation variability (roughness), and distance to the nearest mountain—in Los Angeles County, California. This research is novel in several ways: (1) it explores understudied connections between topography and crime within the field of environmental criminology; (2) it introduces a new topographic variable—distance to the nearest mountain—in crime-geography research; (3) it captures a new level of detail by defining topographic roughness (hilliness) at a different scale than previous studies; (4) it assesses relief, which is closely related to but distinct from slope steepness; and (5) it employs a spatially-adjusted correlation test that accounts for spatial clustering similarities. While global correlations between topographic variables and crime are generally weak or nonexistent, local Lee's L statistics reveal that topography's influence on crime varies significantly across space. For example, crime models incorporating topographic variables may be most beneficial in areas such as the Santa Monica Mountains and coastal regions near Malibu. Future research should investigate locally varying functional forms of these relationships, explore different spatial scales, and address challenges such as the modifiable areal unit problem (MAUP).

Keywords: crime, topography, terrain, relief, roughness, ruggedness, mountain, geographic information systems (GIS), bivariate relationships, spatially adjusted correlations, Los Angeles County

Introduction

A great deal of criminological research and expert opinion solely or primarily considers sociological (Cloward & Ohlin, 2013; Hirschi, 2002; Mills, 2000; Sutherland & Cressey, 1974), psychological (Gottfredson & Hirschi, 1990; Sinha, 2016), economic (Becker, 1968; Ehrlich & Liu, 2006), biological (Bedoya & Portnoy, 2023; Larregue, 2024; Ling et al., 2019; Raine, 2013), or some combination of these factors that are associated with criminal behavior. Advancements in the field of Geographic Information Science (GIS[1]) have led to substantial growth in the subdiscipline of environmental criminology. The first GIS emerged in 1963 with the creation of Roger Tomlinson's Canadian national land use management program. Today, improved design and increased computing power make complex spatial analyses, real-time data processing, and crime modeling more efficient for researchers and criminology hobbyists alike. While the body of environmental criminological research has grown rapidly in recent years, many approaches to the study of crime and place have been conceptualized in one of four ways as described by Bruinsma & Johnson (2018):

> …[1] the neighborhood-effects approach developed by the Chicago school of sociology in the 1920s; [2] modern environmental criminology that explains the geographic distribution of crime; [3] the criminology of place, which focuses on crime rates at specific places over time; and [4] a newer approach that attends to the perception of crime and disorder in communities.

Many researchers have explored the spatial dimension of crime as it relates to the built environment. Some studies explore pattern and trend detection methodologies through the mapping and modeling of crime hotspots and discrete objects in space. This body of research

[1] "GIS" may also reference one or more "Geographic Information Systems."

typically focuses on proximity to features of the built environment. Related research explores the relationship between quantity or quality of discrete objects related to infrastructure, street design, land use, building conditions, building density, natural surveillance, public spaces, and other man-made or man-modified aspects of the environment. Analysis is done to ascertain the statistical significance of spatial clustering or copatterning. Statistically significant results indicate an underlying relationship between two or more variables—a given variable or group of variables influences or is influenced by another—and justifies variable inclusion in crime modeling. For example, many studies have examined relationships between crime and schools (Harth et al., 2022; Murray & Swatt, 2013; Scribner et al., 2010; Willits et al., 2013), alcohol outlets (Cunradi et al., 2011; Fitterer et al., 2018; Gorman et al., 2013; Hobbs et al., 2020; Jones-Webb et al., 2008), shopping centers (Alharbi, 2022), parks (Iqbal & Wilhelmsson, 2018), and transport nodes (Ceccato & Uittenbogaard, 2014). Some studies also consider the arrangement of discrete objects in space (Chang, 2011; Perkins et al., 1993) or the density of the built environment (Ioannidis et al., 2024).

The built environment is often categorized into broader "types"—as defined according to variables of interest[2] (typically related to human geography)—and analyzed in relation to crime frequency or rate. The proximity, quantity, quality, or arrangement of discrete objects is not the primary interest, but spatial stratifications. For example, studies have explored relationships between crime and green spaces (Bogar & Beyer, 2016), urban poverty (Graif et al., 2014), economic activity (Hall, 2010), and other spatial stratifications (Chen et al., 2022; Giménez-Santana et al., 2018; Wells & Weisheit, 2004). Whether a researcher focuses on discrete objects

[2] Definitions of built environment categories are influenced by variables of interest. For example, the "type" of built environment can be characterized based on aspatial variable values, such as quantifiable social, economic, or health conditions and outcomes.

in space or the stratification of space itself (often according to functional, social, or cultural characteristics), the majority of environmental criminological research has remained centered around the built environment.

Addressing spatial aspects of criminology has largely been limited to analyses of influences of the built environment (particularly activity spaces) on criminal behavior and vice versa, with variables defined by or filtered through the lens of human geography. Significant gaps remain in understanding how aspects of the biophysical environment, specifically properties of physical geography, are related to crime. This research investigates global and local spatially-adjusted correlations between overall crime rate and three topographic variables—relief, elevation variability (roughness), and distance to the nearest mountain ("MD").

This report is organized into two main components. The first begins with a literature review that examines three existing studies relevant to this research and highlights their differences in methodological approach compared to this research. Objectives, research questions, hypotheses, and a discussion of the study area provide context, justifying the study area choice. The second component consists of methodology and results discussions. Tables, figures, and code snippets are included throughout and in the appendices.

Literature Review

Though interest in GIS applications for crime modeling has grown in recent years, Joseph Cohen's 1941 statement remains relevant: "[Students] have become less and less disposed to posit new hypotheses concerning the influences of *physical* geography upon crime." As of 2024, research on the physical geography-crime connection has expanded, but with few studies focused on topographic properties. Many studies concentrate on methods such as geographic profiling[3]

---

[3] Determines the most probable area of offender residence by analyzing the locations of a connected series of crimes (Rossmo 2014).

and crime pattern theory[4]. Cohen's (1941) overview discusses the works of Albert Leffingwell and E. G. Dexter from the 1890s, whose works mainly focused on weather-crime associations rather than properties of topography.

Recent studies, however, have begun to narrow the gap in understanding the relationship between crime and topography. For example, Haberman and Kelsay (2021) examined the impact of street block slope on robbery in Cincinnati, Ohio; negative binomial regression models revealed that "a 1% increase in street block slope was associated with roughly 4.5% fewer block robberies per foot of street block length." While this study focused on one of the four types of violent crimes and a different geographical context, the findings suggest a potential link between steepness and overall crime. This provides a rationale for further exploring slope and closely related properties of topography such as relief.

Kim and Wo (2023) explored how elevation differences (or hilliness), slope, and betweenness influence general crime patterns in San Francisco. Their study, which focused on overall crime rather than a specific type of crime, also applied a negative binomial regression modeling methodology. Kim and Wo concluded that a strong association between each independent variable and crime exists in their study area, though hilliness (elevation differences) was found to be the most influential factor. These findings justify the inclusion of elevation variability, also to referred to as "roughness", as a key variable in this study.

Breetzke's (2012) analysis of Tshwane, South Africa, examined the effect of altitude and slope on burglary patterns using ordinary least squares regression and geographically weighted regression, both ideal for linear relationships. Breetzke determined that neighborhoods at higher altitudes experience less burglary, while slope steepness seems to have no significant effect.

---

[4] "…provides explanations for the variation in the distribution of criminal events in space and time given…major components of the built and social environment…" (Brantingham and Brantingham 2021).

While burglary is a nonviolent property crime and makes up only a portion of overall crime, the findings indicate that certain properties of the broader physical environment are correlated with crime.

While there is a growing body of research exploring the relationship between crime and topography, further investigation is necessary. The purpose of this study is to contribute to the physical geography-crime literature by addressing this gap. This research is unique in that it (1) introduces a new topographic variable—distance to the nearest mountain, (2) defines topographic roughness (or hilliness) at a different scale than Kim and Wo (2023), (3) examines relief, which is closely-related but distinct from slope steepness, and (4) applies a new technique: spatially-adjusted correlation which accounts for both the degree of correlation and the similarity of spatial clustering.

Research Questions

Three independent analyses will be run with three independent topographic variables: relief, elevation variability (roughness), and distance to the nearest mountain. The variable of interest is overall crime rate. This study seeks to answer the following questions:

1. Does a significant correlation, adjusted for spatial autocorrelation, exist between overall crime rate and each topographic variable across the study area? If so, how strong is the correlation?
   a. Hypothesis 1a: Areas of lower relief will tend to have higher crime rates.
   b. Hypothesis 1b: Areas of lower roughness—smoother terrain—will tend to have higher crime rates.
   c. Hypothesis 1c: Areas farther from mountains will tend to have higher crime rates.

d. Hypothesis 1d: The expected correlation strength between crime rates and each topographic variable will be moderate to weak, reflecting the influence of other factors on crime.

2. If a significant global correlation exists, do some parts of the county have local spatial associations stronger or weaker than the global association?
    a. Hypothesis 2a: There will be spatial heterogeneity in the correlation strength between crime rates and topographic variables due to varying local effects of other types of variables, including cultural, social, economic, demographic, psychological, and environmental.

There are several reasons why each topographic variable might be correlated with crime rate. Urban areas, places of high population densities and crime rates, are typically located in low-relief, flat, smooth areas that facilitate residential and commercial development. Conversely, mountainous regions usually have lower population densities due to land management policies and regulations[5]. Steep-sloped, rugged terrain is less suitable for development due to construction limitations and logistical challenges in providing emergency response services. In addition, many wilderness areas are federally managed (e.g., Angeles National Forest and its five wilderness areas[6]) and protected as part of natural resource conservation and preservation efforts. These areas do not experience sustained high traffic as in urban areas, where the topography is smoother and flatter, and population centers (and residential areas[7]) tend to cluster. Lastly, specialized engineering techniques and materials required for construction in steep-sloped, rugged, mountainous areas often result in higher development costs. This might result in the

---

[5] "Nearly half of [LA County] is taken up by mountain chains" (Encyclopaedia Britannica 2024).
[6] Cucamonga, Magic Mountain, Pleasant View Ridge, San Gabriel, and Sheep Mountain
[7] In the U.S., violent crimes are most frequently reported in homes and apartment complexes, followed by highways, alleys, streets, sidewalks, parking garages or lots, and convenience stores (Dalton 2022).

creation of exclusive or affluent neighborhoods, where residents with higher socioeconomic statuses may have a lower propensity for engaging in criminal behavior.

Note that the expected correlation strength is moderate to weak because it is unlikely that topography is the main influence behind criminal activity. However, a significant correlation can reveal important insights beyond conventional criminological theory and support policymaking and law enforcement.

Study Area

The study area is the mainland region of the County of Los Angeles, California. LA County covers an area of 4,751 square miles (12,310 square kilometers). This region is an ideal setting for this study for several reasons. (See Appendix A for imagery, elevation, and population maps; see Figure C3 for crime rate map.)

1. It is the largest county in the nation by population with approximately 10 million residents in 2019; Illinois' Cook County was the next largest with a population total of just over 5 million (U.S. Census Bureau, 2019).
2. It has a complex and special topographic structure compared to other American cities and counties, with mountain chains[8] taking up nearly half the county, three waterways[9] passing through the region, an unincorporated area of more than 65 percent, 75 miles of coast, and additional geographical features such as wetlands, marshes, forests, lakes, valleys, basins, and canyons (County of Los Angeles, n.d.-a; County of Los Angeles, n.d.-b; Encyclopaedia Britannica, 2024). Within the city of

[8] Primarily the San Gabriel and San Bernardino mountains to the north and northeast, Santa Monica Mountains, Puente Hills, Repetto Hills, and San Jose Hills

[9] Santa Clara, Los Angeles, and San Gabriel rivers

Los Angeles itself, elevation varies widely, from sea level to a maximum of 1,550 meters (Encyclopaedia Britannica, 2024).

3. The population structure is clearly affected by the topographic structure, from densely populated urban centers in the Los Angeles Basin and the San Fernando and San Gabriel valleys to mountainous terrain and rural expanses encompassing wilderness areas and much of the 650,000-plus acre Angeles National Forest (National Forest Foundation, n.d.). “The city of Los Angeles [itself] is composed of a series of widely dispersed settlements loosely connected to downtown” (Encyclopaedia Britannica, 2024). The county can be divided into eight geographic regions (from north to south): Antelope Valley, Santa Clarita Valley, San Fernando Valley, San Gabriel Valley, Westside Cities, Central Los Angeles, Gateway Cities, and South Bay (County of Los Angeles, n.d.; Encyclopaedia Britannica, 2024).
4. It takes a special position in understanding America’s urban crimes. Outside of the infamous West Adams, Skid Row, and South Los Angeles neighborhoods of the “Gang Capital of America”, nearby cities such as Compton, Lancaster, Victorville, and Huntington Park also experience crime rates higher than the national average.

Boundaries

Note that Santa Catalina and San Clemente islands are excluded from this study. Both islands are geographically distinct from the mainland; the isolation could introduce confounding variables, potentially leading to inaccurate findings. San Clemente Island is also owned and controlled by the U.S. Navy. Therefore, the crime data reported to and by the Los Angeles County Sheriff’s Department may be inaccurate. Lastly, significant differences in demographic and socioeconomic characteristics could skew the crime data.

The borders for this study area are political rather than geographical due to the availability of county-level violent crime data from the L.A. County Sheriff's Department. This is because law enforcement agencies typically function within defined political jurisdictions. In addition, this study imposes temporal boundaries. Topographic characteristics such as terrain roughness and relief generally undergo gradual changes as a result of the slow rate of geological processes, but crime density is a function of population, which can change sporadically and drastically even within the short timeframe of one year. In addition, while GIS allows the characterization of geographic space by various diverse criteria, several types of boundaries are ignored in this study: ethnic, land use, and socioeconomic.

Methodology

Data Format and Suitability

Four datasets were used in this research: (1) 2019 crime records from the Los Angeles County Sheriff's Department in a CSV format, (2) block group population estimates from the U.S. Census Bureau (also in a CSV format), (3) a 30-meter Digital Elevation Model (DEM) raster from the U.S. Geological Survey (USGS), and (4) a land surface classification raster from the USGS (Sayre et al., 2010). The final output layer ("topocrime") used in analysis will be one polygon feature class with the following attribute fields: crime rate per 100 people, MD, relief, and roughness, where the three topographic variables are measured in meters. Mainland LA County will be divided into a grid of 300-by-300-meter polygons, each with an assigned variable value.

The crime data CSV is a comprehensive table of crimes that occurred in 2019 and were reported to the LASD, each record containing detailed attribute information such as type of violent offense, the date and time of occurrence, and location (coordinates and/or address).

Overall crime rate is the variable of interest rather than a subcategory of overall crime (e.g., violent crime) because starting with general crime analyses can help establish foundational knowledge and inform future studies on specific types of crime. In addition, making one type of crime the variable of interest would likely result in a need for a smaller study area or a multi-year study, which introduces confounding variables related to time. For example, using just one year's worth of crime data is much easier to work with, because time requires an account for evolving crime trends, shifting demographics, economic cycles, and policy changes. Therefore, this study uses 2019 crime data from the LASD. For any crime with a valid address or latitude and longitude value that occurred between 12:00 AM, January 1, 2019, and 11:59 PM, December 31, 2019, a point is created. The crime rate is calculated by dividing the total number of reported crimes by the total population of that area, then multiplying by a constant factor of 100. This study calculates crime rate using block-level population data from the 2019 American Community Survey. The year 2020 was not used as a temporal boundary due to the influence of COVID-19 pandemic lockdowns and restrictions on crime patterns.

The LASD is a reliable source and the crime data is suitable for this study, but four key challenges related to accuracy must be considered. First, there is the problem of over- and under-reporting of crime. Some areas, such as places with high gang activity, are more likely to experience crime. Other areas, such as tourist spots targeted by terrorists, are more vulnerable to crime. These places may have disproportionately high crime reports due to targeted policing. Conversely, geographically isolated or less accessible places may lead to less policing and result in under-reporting. Also, for various reasons, some types of crime (e.g., domestic violence or white-collar crime), go unreported. Second, crimes can be misclassified (intentionally or unintentionally) or inconsistently reported due to overlapping jurisdictions and differences in reporting practices. Third, data entry errors are also a concern, especially in the case of reporting precise coordinates or addresses, which makes spatial analysis impossible for crimes with no location information and reduces the accuracy of the analyses as a whole. Technological limitations may also result in data entry errors. Fourth, some crimes are not bound to one geographic location. For example, perpetrators and victims of cybercrimes, financial crimes, drug or human trafficking, or other mobile crimes may be located in two or more different jurisdictions. Where has the crime taken place? In two or more geographic locations? In digital space? Some crimes may result in response by multiple law enforcement agencies, resulting in multiple records for the same incident. Conducting spatial crime analysis can become complex and fail to capture the full extent or patterns of criminal activity. While these challenges make spatial crime analysis complex, the LASD crime data remains suitable for this study, given the study area size and relatively straightforward approach being used to analyze crime and topographic variables.

Block group population data comes from American Community Survey 5-year estimate data (U.S. Census Bureau, 2019). Despite these estimates not being as accurate as the decennial census, 2019 crime and population data is suitable for this research. Data from 2010 would be more accurate in terms of population totals, but fifteen-year-old crime data is sure to be very different from crime trends today. Data from 2020 would also be more accurate as far as population but could introduce additional confounding variables due to COVID-19 and related policy changes in California. At the time of this research, 2019 data is only five years old. It is still recent enough to be relevant, but enough time has passed to make major corrections or adjustments to ensure accuracy.

There are also problems of artificial boundaries when using census data. Block group boundaries may delineate demographic boundaries at some level of accuracy, but this accuracy is limited and often not extended to natural geographic or social divisions. These boundaries align more closely with human geography than physical geography. This research will bypass this problem. How one chooses to divide the study area is crucial, as it can significantly affect the results and interpretations of both traditional statistical analyses and spatial statistics. Geographic space can be divided in many ways, including but not limited to:

- geographic units defined by the U.S. Census Bureau (e.g., block, block group, or census tract);
- a grid of identically shaped and sized cells;
- topographic zones defined by the researcher; or
- any combination of these for use in a comparative analysis.

This research partitions the study area into a grid of identically shaped and sized cells (hereinafter referred to as "locations" with side lengths of 300 meters). This allows for the division of the county into uniform spatial units free from potential bias from census unit boundaries and geometry of form by social, economic, or demographic variables. Perhaps the most important factor is flexibility in scale. While census geographic units are nested within one another and therefore allow varying scales of study[10], researchers have more control over the scale and spatial resolution of the analysis when using a uniform grid, as cell size is only limited to practical considerations such as computational resources and the study area's extent.

[10] Census blocks make up block groups, which make up census tracts, and so on.

The 30-meter DEM[11] was published by the USGS, obtained from OpenTopography, and used to calculate relief and roughness (Sayre et al., 2010). This raster (as well as all other layers) was projected to a customized equal-area projected coordinate system (PCS). An equal-area PCS is most appropriate for this research because, while shape and distances are distorted, relative size (area) is preserved, making area-based measurements and visual comparisons as accurate as possible. See Table 1 for PCS properties. Higher-resolution DEMs would provide more detailed elevation data but are more appropriate for smaller-scale studies. Lower-resolution DEMs could provide a more general overview of the topography of this large region but lose some detail in the process. A 30-meter DEM is ideal for this research because it provides sufficient but not excessive spatial detail for evaluation and comparison of neighborhood topography.

Table 1

Projected Coordinate System Properties

Property	Value
Projection	Lambert Azimuthal Equal Area
Linear Unit	Meters (1.0)
Central Meridian	-118.2990
Latitude of Origin	34.2414
Geographic Coordinate System	NAD 1983
Angular Unit	Degree (0.0174532925199433)
Prime Meridian	Greenwich (0.0)
Spheroid	GRS 1980

Calculating MD requires mountain delineation, though the definition of mountain is subjective. Mountains and mountain boundaries are “fuzzy” geographical concepts—no universally accepted criteria exist to define mountains and clearly distinguish them from hills. This study uses a land surface classification raster (with approximately 34-meter resolution, ideal

[11] Min X: -119, Max X: -117.5, Min Y: 32.65, Max Y: 35.1; NAD83 horizontal coordinates and NAVD88 vertical coordinates

because the resolution is close to the DEM resolution of 30 meters) developed by NatureServe, which delineates landform classes by slope[12] and local relief[13] (Sayre et al., 2010). Using this systematic method makes the delineation of mountainous terrain more objective and enables easy comparison between this study and future research. The foothill, low mountain, and high mountain class features were extracted from the original shapefile and used to make this explanatory variable calculation via the Near tool. Thus, the variable value corresponds to the distance to the nearest foothill, low mountain, or high mountain.

Data Processing

Locations

The state-level geodatabase for California (“tlgdb_2019_a_06_ca.gdb”) contains seven feature classes, one of which—the county boundary layer—was used to create a uniform grid using ArcGIS Pro’s Create Fishnet tool (U.S. Census Bureau, 2019). This was done only after exploding the LA County multipart feature into three features (San Clemente Island, Santa Catalina Island, and mainland LA County) and creating a feature class with only mainland LA County. The fishnet tool was then configured to produce a polygon feature class of 300-by-300-meter polygons as well as a point feature class with points located at the center of each “location”, or polygon. Because the tool creates a location feature class that covers the spatial extent of another feature class (i.e., the county layer), all locations not intersecting the county boundary polygon were deleted. All points that did not intersect any fishnet location polygon were also deleted. Both output layers “locations” and “locations_label” (polygon and point, respectively) have a total of 121,678 features.

[12] Gently sloping or not gently sloping by a threshold of 8%
[13] Five classes: 0-15, 15-30, 30-90, 90-150, and greater than 150 meters

Population Estimates

Population data in a CSV format was obtained from the U.S. Census Bureau and joined to the census block group layer from the state-level geodatabase. A long integer field "Population" was created and populated before the join was removed. A population estimate was assigned to each location polygon using the Apportion Polygon tool. This approximation is based on the percentage of the location polygon that overlays the block group layer.

Crime Rate

A CSV containing all crime incident records from 2019 was obtained from the Los Angeles Sheriff's Department and cleaned using R (Los Angeles County Sheriff's Department, n.d.). See Appendix B for the code to create two CSVs: one with valid coordinates ("crimes19_XY") and one with invalid coordinates but valid addresses ("crimes19_gc"). The "crimes19_XY" and "crimes19_gc" CSVs were converted to point feature classes using ArcGIS Pro's XY Table to Point and Geocode[14] tools, respectively. After deleting all unmatched and tied records from the geocoded layer, the point feature classes were merged to create "crimes2019". After all points outside the study area were removed, the master 2019 crime layer contained 155,065 points.

Crime counts were calculated for each location by applying a one-to-one spatial join with the locations polygon feature class as the target features layer, "crimes2019" as the join features layer, and a match option of "contains". A double-precision type field "CrimeRateper100" was created and the overall crime rate (crimes per 100 people) for each location was calculated:

$$Crime\ Rate = \frac{Crime\ Count}{Total\ Population} \cdot 100$$

[14] Note that this step required the recreation of the Countywide Address Management System (CAMS) locator using the CAMS address point and road segment data and Create Locator tool (County of Los Angeles, 2024).

Relief and Roughness

The 30-meter DEM GeoTIFF was projected to the customized equal-area PCS using the bilinear interpolation resampling type. This technique was primarily chosen due to its faster processing time compared to cubic convolution. However, there are two additional reasons why bilinear interpolation is more appropriate for this study. First, this research works with 30-meter DEMs, which generally have less noise. Because noise reduction is not a strong concern, bilinear interpolation is more appropriate. Second, two variables of interest are relief and elevation variability. Cubic convolution is less suitable because some extreme values can be lost when edges are smoothed.

The Zonal and Focal Statistics tools were used to calculate relief and roughness variables, respectively. Relief is defined as the vertical height difference within a location. The location polygon feature class was used as the feature zone layer, the elevation raster as the in value raster, and relief was calculated using the "range" statistics type. Because the tool output was a raster, the Extract Values to Points tool was used to assign relief values to the location points layer. Topographic roughness (i.e., elevation variability) was calculated for each 30-meter cell within a 300-meter rectangular neighborhood via the Focal Statistics tool, using a rectangular neighborhood of 300 meters and a statistics type of "standard deviation." The center cell value of the output roughness raster was assigned to the corresponding location point using the Extract Multi Values to Points tool.

Distance to the Nearest Mountain (MD)

Distance is measured as the horizontal distance to the nearest mountain[15]. This research does not consider three-dimensional (slope) distance. With such a large such area divided into

[15] Breaks/foothills, low mountains, or high mountains/deep canyons as defined by the U.S. Geological Survey "Terrestrial Ecosystems of the Conterminous United States" dataset (Sayre et al., 2010).

over 100,000 locations, calculating two-dimensional distance is much faster to process. Also, in this foundational research, there is an advantage to starting with a simplified initial analysis. Basic relationships may be established, providing a baseline understanding. Subsequent research can incorporate complex three-dimensional distance measurements to account for changes in elevation.

A polygon feature class was created using the land surface classification raster. Only the three mountain classes were retained. Polygons were not simplified in order to preserve the exact raster cell edges, with no loss of information that could potentially distort analysis outputs and influence interpretations. The distance of each location polygon to the nearest mountain polygon was calculated using the Near tool. The geodesic method was used because the study area is rather large and the method takes into account the curvature of the spheroid. Even if significant curvature is not present, the geodesic method will produce accurate distance measurements.

Analysis

Traditional Descriptive Statistics

Descriptive statistics are important when determining an appropriate test for association or spatial statistics and are very easily found with ArcGIS Pro's data engineering tools. Table 2 shows descriptive statistics for all study variables.

Table 2

Descriptive Statistics

	Crime Rate	Relief	Roughness	MD
Nulls	49,661	0	0	0
Count	72,017	121,678	121,678	121,678
Minimum	0	0	0	0
Maximum	7,661.905	400	94.879	16,906.569
Mean	2.976	56.179	13.835	2,221.636
Standard Deviation	38.000	63.616	15.765	3320.790
Median	0	26	6.368	48.706

Mode	0	2	0	0
Outliers	14,530	1,249	1,423	6,958
Interquartile Range	0	93	22.812	3,732.324
Coefficient of Variation	12.769	1.132	1.140	1.495
Skewness	123.412	1.118	1.140	1.628
Kurtosis	23,255.257	3.415	3.457	5.014

Several key insights into crime rate are revealed in Table 2 as well as the histograms and quantile-quantile plots in Appendix C:

1. There are 121,678 locations. All have a valid numerical value for relief, roughness, and MD. However, 49,661 locations have a null crime rate. A null crime rate may occur due to missing population data or a population of zero. In addition, due to the method of population apportionment, some rural areas may have a population too close to zero which has resulted in a rounded integer population value of zero. Approximately 41% of the study area will *not* be considered in the final analysis.
2. The mean crime rate across the study area is approximately 3 per 100 people. Crime rate median and mode are both zero, which indicates most locations have a crime rate equal to or near zero, though some locations with extremely high crime rates are pulling the average up. A mean greater than the median indicates a positively skewed distribution, and in fact, there are 14,530 crime rate outliers.
3. With a skewness and kurtosis values of approximately 123 and 23,255, respectively, crime rate is highly positively skewed (expected, especially considering insight two) and the tails of the distribution are extremely heavy tails compared to the normal distribution. Crime rate is highly leptokurtic.

Applying standard parametric tests to measure linear relationships (e.g., Pearson correlation coefficient), monotonic relationships (e.g., Spearman rank correlation), or the strength of association (e.g., Kendall's Tau) would not be appropriate for this dataset. Assumptions of normality, outlier absence, and consistent direction of association cannot be met, and while Spearman's rank correlation coefficient could be calculated, converting data to ranks can lead to loss of information. While Kendall's Tau could also be computed to evaluate the strength of associations between crime rate and each topographic variable, this would be computationally intensive for a dataset of this size and would not account for the spatial nature of the data. Traditional statistical tests are aspatial and do not recognize special distributional aspects of spatial datasets.

Spatial Statistics

Many studies have shown that crime data almost always exhibit spatial autocorrelation (Andresen, 2011). The results of the Spatial Autocorrelation tool, which applies Global Moran's I to the crime data, show that this dataset exhibits statistically significant clustering across the county. The tool's conceptualization of spatial relationships parameter was set to "inverse distance squared" because of the strong (likely non-linear) influence of immediate spatial neighbors on crime patterns (Andresen, 2011). A default distance band of 300.03 meters was used. With a positive z-score of approximately 37 and a p-value of 0.000000, it is less than 1% likely that the spatial clustering of both high and low values is the result of random chance. Therefore, the data points are not independent of one another and the research questions require a statistical test that accounts for spatial autocorrelation.

A spatial method is necessary; however, it's important to first consider the extremely high proportion of zero values in the dataset. Typically, a transformation (likely logarithmic in this case) would be applied to one or both variables or a zero-inflated model would be used. However, despite the extreme skewness of the data, transformation is not ideal, even though it would reduce outlier influence. Applying a log transformation requires either the exclusion of zeros or the addition of a small constant. For all study variables, zeros are significant—not arbitrary—and choosing either option would result in a loss of information, misrepresentation, or misinterpretation of results. Therefore, this dataset will not be transformed. Zero-inflated models can also be useful when there are excess zeros in a dataset, though spatial variables (also continuous, rate data) require modification of conventional zero-inflated models that typically model count or binary data (Lee & Haran, 2024; Mutiso et al., 2024; Osei et al., 2022).

Geographically Weighted Regression (GWR) can answer some of the same questions as the Bivariate Spatial Association (Lee's L) tool in ArcGIS Pro. For example, "Do crime rates increase with proximity to mountains?" or "Do areas with higher topographic roughness tend to have lower crime rates?" However, because the primary purpose of this research is to determine the strength and significance of global and location spatial associations and additional factors proven to influence crime have not been prepared for modeling, this research uses Lee's L to assess the research questions. Lee's L integrates Pearson's r and Moran's I, thus quantifying both the correlation and copatterning of two spatial variables (Lee, 2001).

The Bivariate Spatial Association (BSA) tool was run three times with a fixed distance band neighborhood type, distance band of 1,501 meters, bisquare local weighting scheme, default kernel bandwidth of 1,501 meters, and the default number of permutations (999), which is recommended for 90% confidence tests. The distance band was identified using the Incremental Spatial Autocorrelation tool; it is at 1,501 meters that crime rate clustering is most prominent. BSA outputs included geoprocessing messages for the global statistics and a feature class with local p-values and significance levels.

Results

Spatial Patterns

Before interpreting the statistical results, visual exploration reveals interesting spatial patterns across the county. (See Appendix C for univariate maps and Appendix D for bivariate maps.) Large clusters of high-crime locations are observed in relatively flat, smooth, densely populated, urban areas, such as the cities of Los Angeles, Santa Clarita, Lancaster, and Palmdale. These areas are also the farthest from mountains. Surrounding many of these crime hotspots are moderately flat, smooth suburban regions with low crime rates. Interestingly, recreational areas like the coastline adjacent to the rugged, high-relief Santa Monica Mountains exhibit high crime rates. In contrast, while much of the Angeles National Forest was excluded from the study[16], most areas adjacent to this mountainous region exhibit very low crime rates.

Global Lee's L

A summary of global statistics is presented in Table 3. All three topographic variables have a global Lee's L value very close to zero, which indicates that the variables are very weakly or not spatially associated. This means the variable pairs are either uncorrelated or they are not spatially autocorrelated, though note that relief and roughness lean negative, while MD leans positive, as hypothesized. The Lee's L values are significant because the global p-values are less than 0.05. The spatial smoothing scalars for the topographic variables reveal that MD exhibits the strongest positive spatial autocorrelation, followed by relief and roughness. The spatial smoothing scalar for crime rate remains the same across all three runs and indicates positive spatial autocorrelation, though to a lesser degree than the topographic variables. The raw Pearson correlation values are less than the global Lee's L values and very close to zero, indicating very

[16] Reason: null crime rates, likely due to low or no population

weak negative or no linear aspatial correlations between each set of variables. Unsurprisingly, the correlation strength increases when taking geography into account.

In each run, the p-value is significant but the global Lee's L statistic and the correlation between neighborhood averages are very close to zero. Because the spatial smoothing scalars are also close to 1, this means it is likely there is little to no cross-correlation between the variables. This suggests that while the topographic variables exhibit some spatial patterning, the spatial association with crime is weak or non-existent. Rather, each variable is highly autocorrelated. It is unlikely that a significant correlation exists between crime and each topographic variable, though if a correlation does exist, it is very weak. It is likely that other factors, not topography alone—as hypothesized—are more influential in global crime patterns.

Table 3

Global Lee's L (BSA) Statistics

	Relief	Roughness	MD
Output feature class	bsa_v1	bsa_v2	bsa_v3
Global Lee's L	-0.0317	-0.0313	0.0250
Global p-value	0.0020	0.0020	0.0020
Spatial smoothing scalar	0.8096	0.7861	0.9898
Spatial smoothing scalar (Crime Rate)	0.0576	0.0576	0.0576
Pearson correlation (raw)	-0.0412	-0.0407	0.0260
Pearson correlation (neighborhood averages)	-0.1469	-0.1473	0.1049

Local Lee's L

Each relationship clearly exhibits varying levels of spatial association across the study area. See Table 4 for total locations in each local spatial association (LSA) category as well as the number of locations with a local Lee's L greater than the global Lee's L. See Appendix E for three maps symbolized by LSA category[17]. The least common LSA category is High-High. In a

[17] Not Significant, High-High, High-Low, Low-High, and Low-Low

small number of locations, both variables have a neighborhood average higher than the global average. This occurs in parts of Agoura Hills, Calabasas, Malibu, beaches adjacent to the Santa Monica Mountains, the Santa Monica Mountains (primarily near roads), and areas outside Santa Clarita. The majority of LA County has a LSA category of Low-Low, which means both variables have a neighborhood average lower than the global average. However, in urban areas such as Los Angeles, Lancaster, Palmdale, and Santa Clarita, where crime rates are highest, the LSA category is High-Low, which means crime rate neighborhood averages are higher than the global average, but topographic neighborhood averages are lower than the global average.

Table 4

BSA Output Statistics

LSA Category	Relief	Roughness	MD
High-High	2,319	2,335	6,510
High-Low	10,359	10,346	6,990
Low-High	21,020	21,021	12,827
Low-Low	24,595	24,424	34,280
Not Significant	13,724	13,891	11,410
Total	72,017	72,017	72,017
Locations with local L > global L	45,344	45,372	36,381

Discussion

Global correlations are unlikely due to the near-zero global Lee's L values for each relationship. Any global correlations the exist are very weak. However, local Lee's L statistics reveal nonstationary relationships across LA County. While relief, roughness, or distance to the nearest mountain are not suitable variables for large-scale models due to their weak global correlations, local Lee's L statistics suggest that in some areas, topographic variables have more influence on crime. For example, communities in the Santa Monica Mountains and the nearby coastal areas would likely see the most improvement in model accuracy with the inclusion of topographic variables.

Future studies should also explore the functional forms of local relationships. Perhaps in some locations, significant linear relationships exist, but in others, the relationship is exponential, quadratic, sinusoidal, or complex. Further research must be done to fully understand the highly complex relationships between physical geography and crime. This study may be improved or furthered in several ways. For example, BSA performs well when there are no or very few outliers, and the distribution of each variable is strongly influenced by the method of aggregation or study area partitioning. Spatial scale must be carefully defined, and the modifiable areal unit problem (MAUP) must be considered. For example, larger grid cells (e.g., a resolution of 3000 meters rather than 300) may better capture variation in crime rates. It might be more appropriate to aggregate the data at a larger spatial scale to reduce the number of zero observations. Or, aggregation by geographic unit defined by the U.S. Census Bureau would remove any inaccuracies in the apportionment of population counts while also allowing for easier inclusion of many other variables collected at these units.

Note additional influences of project design on final research outcomes. First, there is some subjectivity in defining topographic “roughness” (also referred to as “ruggedness”), as several academic studies have defined this characteristic differently. This study defines roughness as the variability of elevation values—calculated as the absolute standard deviation of elevation values—within each 300-meter-by-300-meter location. Second, sensitivity to methodologies may be especially significant in the estimation of total population per location as well as the delineation of mountains. Variations in parameter settings or methodologies may yield different outcomes. For instance, mountains do not have a widely accepted definition, though this study uses the definition used by some researchers: “A mountain is a landform that rises at least 1,000 feet (300 meters) or more above its surrounding area.” However, if a researcher were to change this definition and apply a multi-criteria selection process before delineation (perhaps considering relative or local relief, slope and elevation), the results might undergo a significant change.

Lastly, this study is limited by the residential population counts obtained from the U.S. Census Bureau. These estimates may not fully capture the transient population—hourly/daily population shifts—changes due commuting, tourism, or other human activity and traffic. Possible design improvements include (1) using violent crime count as the dependent variable and accounting for population as a confounding variable, or (2) using an alternative method of population estimation. Data sources and collection methods might include daytime population estimates (data might be obtained from transportation agencies, urban planning departments, or commercial datasets), mobile phone or social media data (analyze distribution and movement patterns over time), remote sensing data (e.g., detecting human activity indicators at certain times of day such as nighttime lights or vehicle movement patterns), or field studies and surveys.

This study offers valuable insights into the spatial patterns of crime. Crime is a complex phenomenon, and models become better representations of the real world when physical geography is considered. As studies bridge the gap in the crime-topography subfield of environmental criminology, insights beyond conventional criminological theory will be revealed. These insights will potentially support law enforcement agencies and policymakers as they explore crime patterns, identify high-risk areas, build predictive models, refine prevention strategies, and improve post-crime response by reallocating resources.

References

Alharbi, M. M. (2022). Shoplifting Crime Hotspots Are Influenced By Population Density And Proximity Of Shops To Police Station Locations In Jeddah City, Saudi Arabia. *European Journal of Social Sciences Studies*, *7*(6). https://doi.org/10.46827/ejsss.v7i6.1308

Andresen, M. A. (2011). Estimating the probability of local crime clusters: The impact of immediate spatial neighbors. *Journal of Criminal Justice*, *39*(5), 394–404. https://doi.org/10.1016/j.jcrimjus.2011.05.005

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy*, *76*(2), 169–217.

Bedoya, A., & Portnoy, J. (2023). Biosocial Criminology: History, Theory, Research Evidence, and Policy. *Victims & Offenders*, *18*(8), 1599–1629. https://doi.org/10.1080/15564886.2022.2133035

Bogar, S., & Beyer, K. M. (2016). Green Space, Violence, and Crime: A Systematic Review. *Trauma, Violence, & Abuse*, *17*(2), 160–171. https://doi.org/10.1177/1524838015576412

Breetzke, G. D. (2012). The effect of altitude and slope on the spatial patterning of burglary. *Applied Geography*, *34*, 66–75. https://doi.org/10.1016/j.apgeog.2011.10.017

Bruinsma, G. J. N., & Johnson, S. D. (Eds.). (2018). *The Oxford Handbook of Environmental Criminology* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190279707.001.0001

Ceccato, V., & Uittenbogaard, A. C. (2014). Space–Time Dynamics of Crime in Transport Nodes. *Annals of the Association of American Geographers*, *104*(1), 131–150. https://doi.org/10.1080/00045608.2013.846150

Chang, D. (2011). Social Crime or Spatial Crime? Exploring the Effects of Social, Economical, and Spatial Factors on Burglary Rates. *Environment and Behavior*, *43*(1), 26–52. https://doi.org/10.1177/0013916509347728

Chen, T., Bowers, K., Zhu, D., Gao, X., & Cheng, T. (2022). Spatio-temporal stratified associations between urban human activities and crime patterns: A case study in San Francisco around the COVID-19 stay-at-home mandate. *Computational Urban Science*, *2*(1), 13. https://doi.org/10.1007/s43762-022-00041-2

Cloward, R. A., & Ohlin, L. E. (2013). *Delinquency and Opportunity* (0 ed.). Routledge. https://doi.org/10.4324/9781315007274

Cohen, J. (1941). The Geography of Crime. *The ANNALS of the American Academy of Political and Social Science*, *217*(1), 29–37. https://doi.org/10.1177/000271624121700105

County of Los Angeles. (n.d.-a). *Beaches & Harbors*. Retrieved December 11, 2024, from https://lacounty.gov/residents/things-to-do/beaches-harbors/

County of Los Angeles. (n.d.-b). *Maps and Geography of Los Angeles County*. Retrieved December 11, 2024, from https://lacounty.gov/government/about-la-county/maps-and-geography/

Cunradi, C. B., Mair, C., Ponicki, W., & Remer, L. (2011). Alcohol Outlets, Neighborhood Characteristics, and Intimate Partner Violence: Ecological Analysis of a California City. *Journal of Urban Health*, *88*(2), 191–200. https://doi.org/10.1007/s11524-011-9549-6

Dalton, M. (2022). *Where Do Violent Crimes Commonly Occur in the U.S.?* Mickelsen Dalton LLC. https://www.mickelsendalton.com/blog/where-do-violent-crimes-commonly-occur-in-the-u-s/

Ehrlich, I., & Liu, Z. (Eds.). (2006). *The Economics of Crime*. E. Elgar.

Encyclopaedia Britannica. (2024). *Landscape of Los Angeles*. Encyclopaedia Britannica. https://www.britannica.com/place/Los-Angeles-California/Landscape

Fitterer, J. L., Nelson, T. A., & Stockwell, T. (2018). The Negative Effects of Alcohol Establishment Size and Proximity on the Frequency of Violent and Disorder Crime across Block Groups of Victoria, British Columbia. *ISPRS International Journal of Geo-Information*, *7*(8), 297. https://doi.org/10.3390/ijgi7080297

Giménez-Santana, A., Caplan, J. M., & Drawve, G. (2018). Risk Terrain Modeling and Socio-Economic Stratification: Identifying Risky Places for Violent Crime Victimization in Bogotá, Colombia. *European Journal on Criminal Policy and Research*, *24*(4), 417–431. https://doi.org/10.1007/s10610-018-9374-5

Gorman, D. M., Gruenewald, P. J., & Waller, L. A. (2013). Linking places to problems: Geospatial theories of neighborhoods, alcohol and crime. *GeoJournal*, *78*(3), 417–428. https://doi.org/10.1007/s10708-011-9425-7

Gottfredson, M. R., & Hirschi, T. (1990). *A General Theory of Crime*. Stanford University Press.

Graif, C., Gladfelter, A. S., & Matthews, S. A. (2014). Urban Poverty and Neighborhood Effects on Crime: Incorporating Spatial and Network Perspectives. *Sociology Compass*, *8*(9), 1140–1155. https://doi.org/10.1111/soc4.12199

Haberman, C. P., & Kelsay, J. D. (2021). The Topography of Robbery: Does Slope Matter? *Journal of Quantitative Criminology*, *37*(3), 625–645. https://doi.org/10.1007/s10940-020-09451-z

Hall, T. (2010). Economic geography and organized crime: A critical review. *Geoforum*, *41*(6), 841–845. https://doi.org/10.1016/j.geoforum.2010.06.004

Harth, L. de C., Xiao, B., & Jensen, S. T. (2022). *Spatial Analysis of the Association between School Proximity and Crime in Philadelphia* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2206.03387

Hirschi, T. (2002). *Causes of Delinquency*. Transaction Publishers.

Hobbs, M., Marek, L., Wiki, J., Campbell, M., Deng, B. Y., Sharpe, H., McCarthy, J., & Kingham, S. (2020). Close proximity to alcohol outlets is associated with increased crime and hazardous drinking: Pooled nationally representative data from New Zealand. *Health & Place*, *65*, 102397. https://doi.org/10.1016/j.healthplace.2020.102397

Ioannidis, I., Haining, R. P., Ceccato, V., & Nascetti, A. (2024). Using remote sensing data to derive built-form indexes to analyze the geography of residential burglary and street thefts. *Cartography and Geographic Information Science*, 1–17. https://doi.org/10.1080/15230406.2023.2296598

Iqbal, A., & Wilhelmsson, M. (2018). Park proximity, crime and apartment prices. *International Journal of Housing Markets and Analysis*, *11*(4), 669–686. https://doi.org/10.1108/IJHMA-04-2017-0035

Jones-Webb, R., McKee, P., Hannan, P., Wall, M., Pham, L., Erickson, D., & Wagenaar, A. (2008). Alcohol and Malt Liquor Availability and Promotion and Homicide in Inner Cities. *Substance Use & Misuse*, *43*(2), 159–177. https://doi.org/10.1080/10826080701690557

Kim, Y.-A., & Wo, J. C. (2023). Topography and crime in place: The effects of elevation, slope, and betweenness in San Francisco street segments. *Journal of Urban Affairs*, *45*(6), 1120–1144. https://doi.org/10.1080/07352166.2021.1901591

Larregue, J. (2024). *Hereditary: The Persistence of Biological Theories of Crime*. Stanford University Press.

Lee, B. S., & Haran, M. (2024). A class of models for large zero-inflated spatial data. *Journal of Agricultural, Biological and Environmental Statistics*. https://doi.org/10.1007/s13253-024-00619-9

Lee, S.-I. (2001). Developing a bivariate spatial association measure: An integration of Pearson's r and Moran's I. *Journal of Geographical Systems*, *3*(4), 369–385. https://doi.org/10.1007/s101090100064

Ling, S., Umbach, R., & Raine, A. (2019). Biological Explanations of Criminal Behavior. *Psychology, Crime & Law*, *25*(6), 626–640. https://doi.org/10.1080/1068316X.2019.1572753

Los Angeles County Sheriff's Department (LASD). (n.d.). *2019 Part 1 and Part 2 Crime Data* [CSV]. Los Angeles County Sheriff's Department. https://lasd.org/transparency/part1and2crimedata/

Mills, C. W. (2000). *The Sociological Imagination*. Oxford University Press.

Murray, R. K., & Swatt, M. L. (2013). Disaggregating the Relationship Between Schools and Crime: A Spatial Analysis. *Crime & Delinquency*, *59*(2), 163–190. https://doi.org/10.1177/0011128709348438

Mutiso, F., Pearce, J. L., Benjamin-Neelon, S. E., Mueller, N. T., Li, H., & Neelon, B. (2024). A Marginalized Zero-Inflated Negative Binomial Model for Spatial Data: Modeling COVID-19 Deaths in Georgia. *Biometrical Journal*, *66*(5), e202300182. https://doi.org/10.1002/bimj.202300182

National Forest Foundation. (n.d.). *Angeles National Forest*. National Forest Foundation. Retrieved December 11, 2024, from https://www.nationalforests.org/our-forests/find-a-forest/angeles

Osei, F. B., Stein, A., & Andreo, V. (2022). A zero-inflated mixture spatially varying coefficient modeling of cholera incidences. *Spatial Statistics*, *48*, 100635. https://doi.org/10.1016/j.spasta.2022.100635

Perkins, D. D., Wandersman, A., Rich, R. C., & Taylor, R. B. (1993). The physical environment of street crime: Defensible space, territoriality and incivilities. *Journal of Environmental Psychology*, *13*(1), 29–49. https://doi.org/10.1016/S0272-4944(05)80213-0

Raine, A. (2013). *The Anatomy of Violence: The Biological Roots of Crime*. Pantheon/Random House.

Sayre, R., Comer, P., Cress, J. J., & Warner, H. (2010). *Conterminous United States Terrestrial Ecosystems* [Dataset]. U.S. Geological Survey. https://doi.org/10.5066/P94VQO8E

Scribner, R. A., Mason, K. E., Simonsen, N. R., Theall, K., Chotalia, J., Johnson, S., Schneider, S. K., & DeJong, W. (2010). An Ecological Analysis of Alcohol-Outlet Density and Campus-Reported Violence at 32 U.S. Colleges. *Journal of Studies on Alcohol and Drugs*, *71*(2), 184–191. https://doi.org/10.15288/jsad.2010.71.184

Sinha, S. (2016). Personality Correlates of Criminals: A comparative study between normal controls and criminals. *Industrial Psychiatry Journal*, *25*(1), 41. https://doi.org/10.4103/0972-6748.196058

Sutherland, E. H., & Cressey, D. R. (1974). *Criminology* (9th ed). Lippincott.

United States Geological Survey (USGS). (2021). *United States Geological Survey 3D Elevation Program 1/3 arc-second Digital Elevation Model* [Dataset]. OpenTopography. https://doi.org/10.5069/G98K778D

U.S. Census Bureau. (2019a). *SEX BY AGE* [CSV]. American Community Survey, ACS 5-Year Estimates Detailed Tables, Table B01001. https://data.census.gov/table/ACSDT5Y2019.B01001?t=Populations%20and%20People&g=050XX00US06037$1500000&y=2019

U.S. Census Bureau. (2019b). *TIGER/Line State-Level Geodatabase for California* [Dataset]. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-geodatabase-file.html

Wells, L. E., & Weisheit, R. A. (2004). Patterns of Rural and Urban Crime: A County-Level Comparison. *Criminal Justice Review*, *29*(1), 1–22. https://doi.org/10.1177/073401680402900103

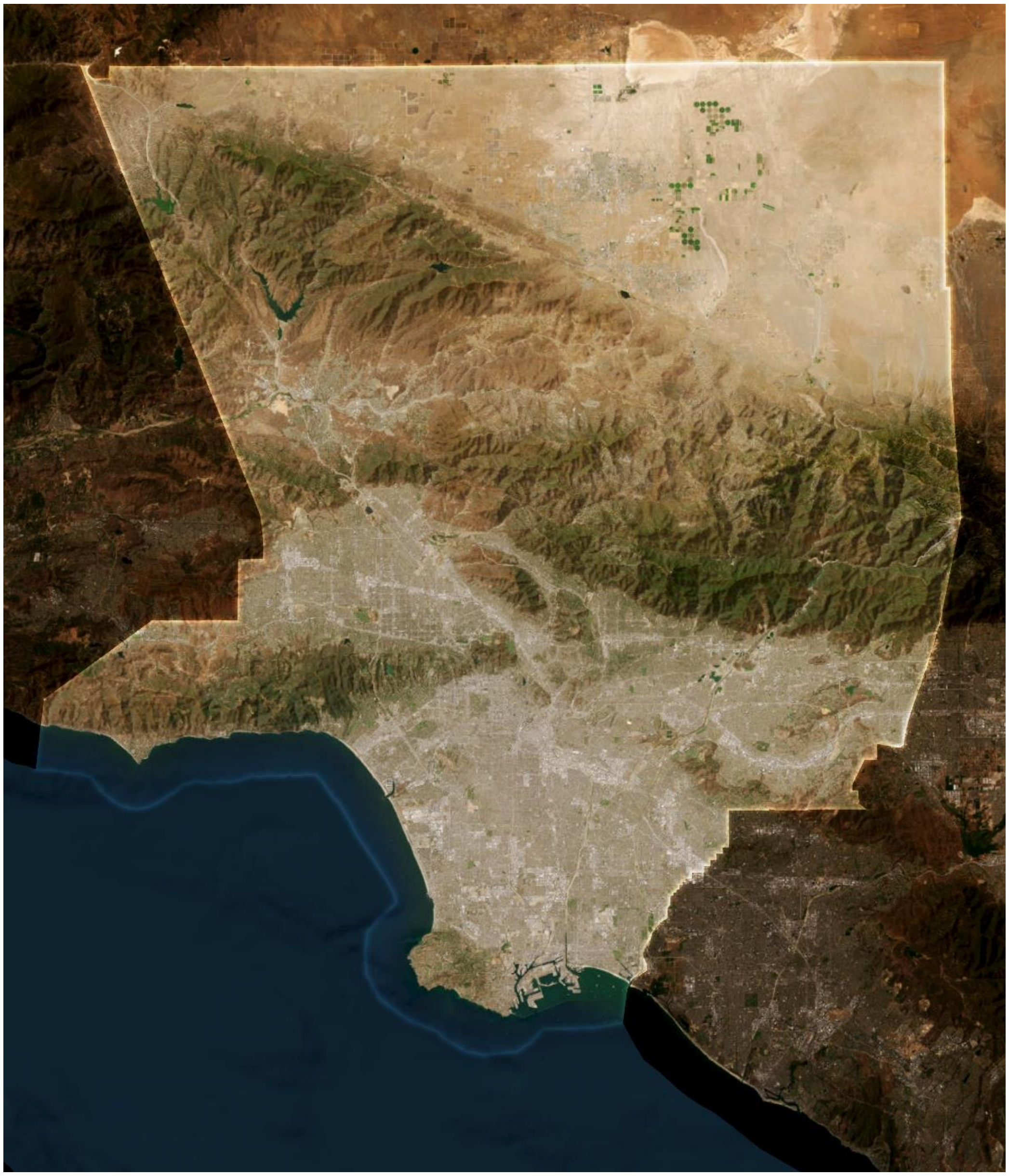
Willits, D., Broidy, L., & Denman, K. (2013). Schools, Neighborhood Risk Factors, and Crime. *Crime & Delinquency*, *59*(2), 292–315. https://doi.org/10.1177/0011128712470991

Appendix A

Study Area Maps

Figure A1

Los Angeles County, California



Note. This high-resolution imagery shows the varied landscape of the study area. Adapted from Esri, Maxar, Earthstar Geographics, and the GIS User Community, 2024.

Figure A2

Digital Elevation Model (DEM) Raster

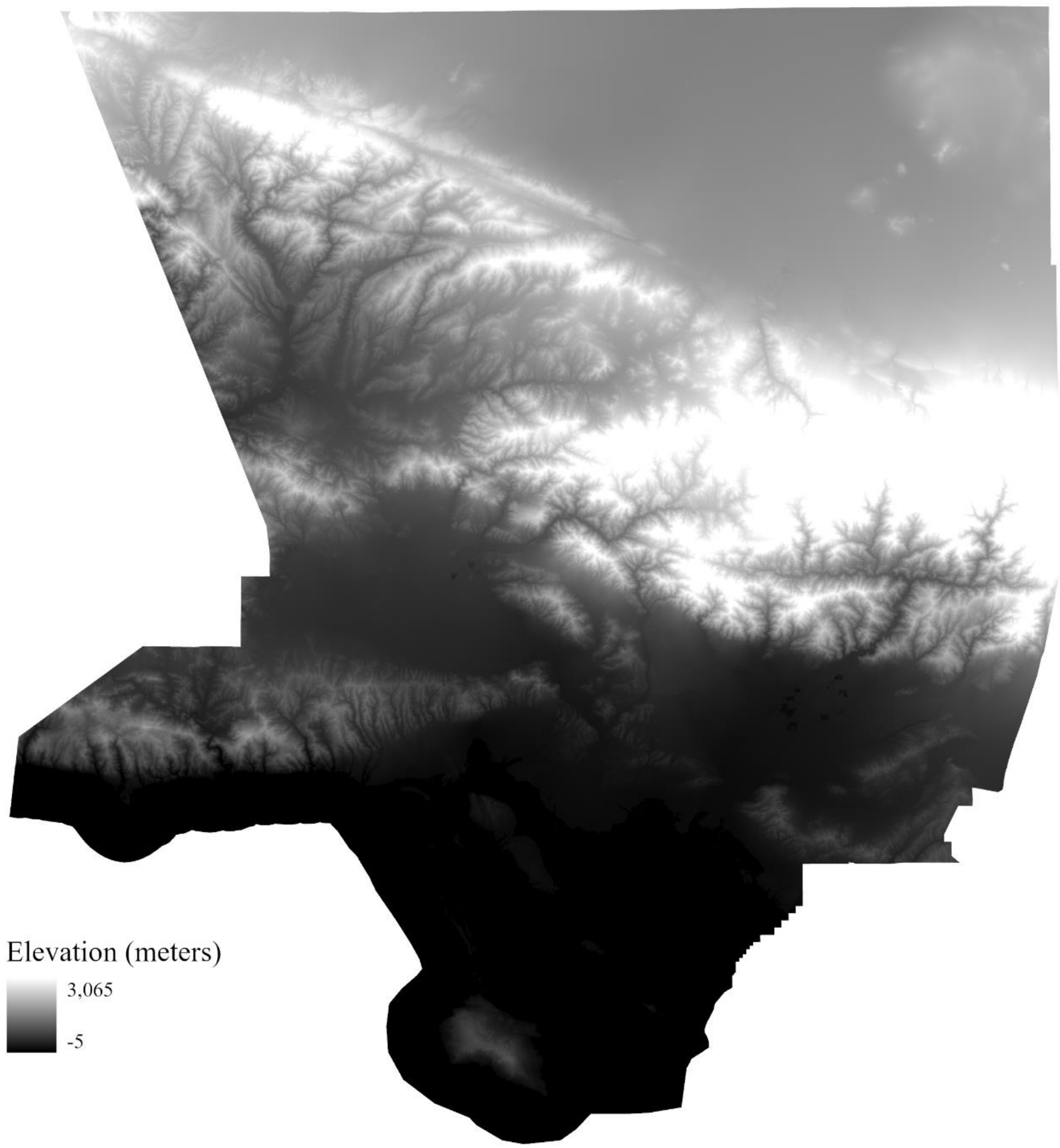
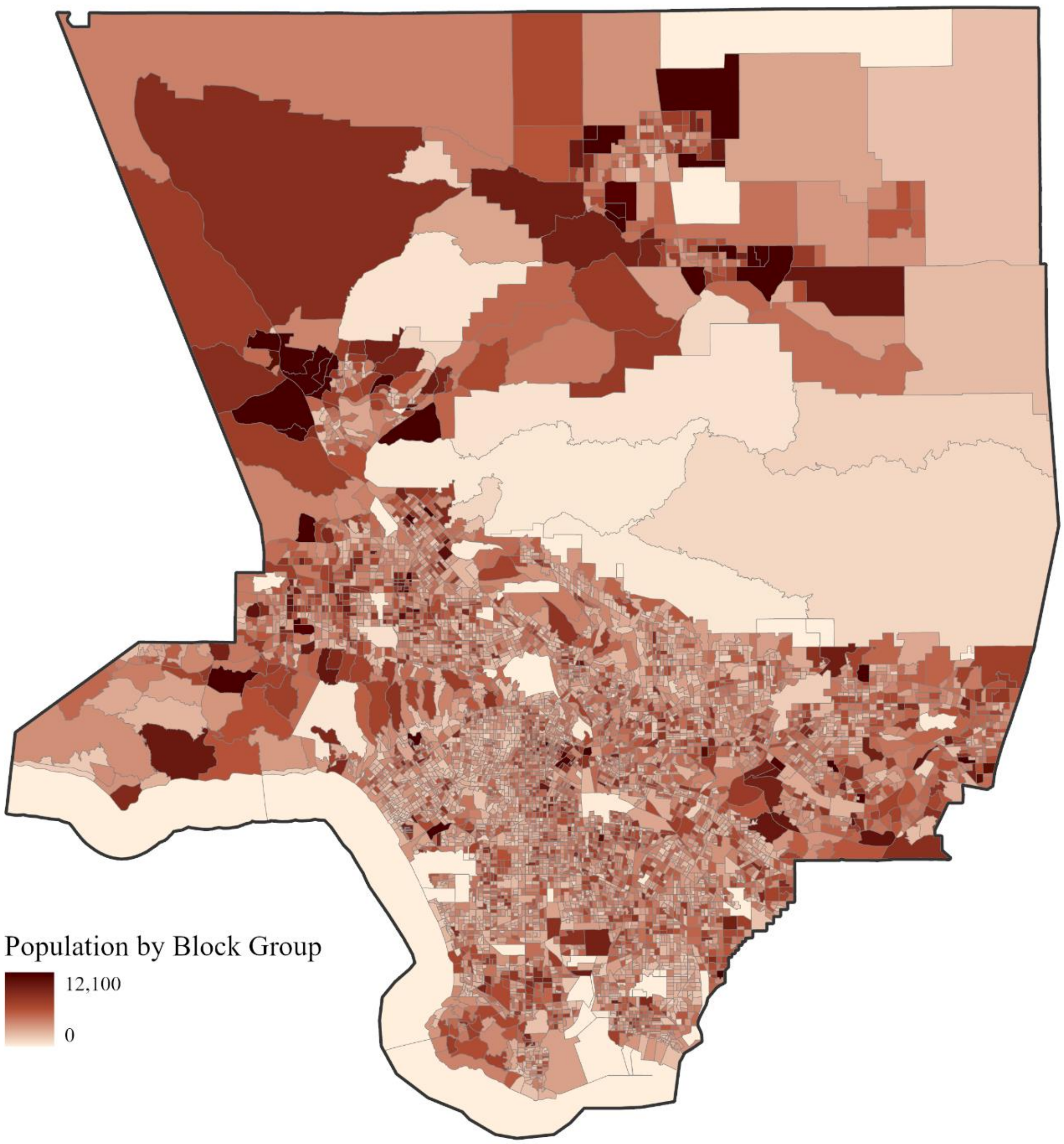


Figure A3

Population by Block Group, 2019



Appendix B

R Code for Crime CSV Preprocessing

The following R code was used to clean the CSV ("2019-PART_I_AND_II_CRIMES") obtained from the Los Angeles Sheriff's Department.

```
# Obtain 2019 crime data from LASD
crimes19 <- read_csv("http://shq.lasdnews.net/CrimeStats/CAASS/2019-
PART_I_AND_II_CRIMES.csv")
# Remove unnecessary variables and reorder the remaining columns
crimes19 <- crimes19 %>%
  select(INCIDENT_DATE,
         PART_CATEGORY,
         CATEGORY,
         STAT_DESC,
         LONGITUDE,
         LATITUDE,
         ADDRESS)
# Convert INCIDENT_DATE strings to date-times (mm/dd/yyyy hh:mm:ss AM/PM)
mutate(crimes19, INCIDENT_DATE = trimws(INCIDENT_DATE))
# Trim spaces for other character fields:
mutate(crimes19, CATEGORY = trimws(CATEGORY), STAT_DESC = trimws(STAT_DESC),
ADDRESS = trimws(ADDRESS))
# Convert to date-time format
crimes19 <- crimes19 %>%
  mutate(DATE = mdy_hms(INCIDENT_DATE,
                        tz = "America/Los_Angeles",
                        quiet = TRUE))
```

```
# Retain and reorder necessary columns (variables)
crimes19 <- crimes19 %>%
  select(DATE,
         PART_CATEGORY,
         CATEGORY,
         STAT_DESC,
         LONGITUDE,
         LATITUDE,
         ADDRESS)
# Rename DATE to INCIDENT_DATE
crimes19 <- rename(crimes19, INCIDENT_DATE = DATE)
# Remove observations that occurred before or after 2019
crimes19 <- crimes19 %>%
  filter(
    INCIDENT_DATE >= as.POSIXct("2019-01-01 12:00:00", tz =
"America/Los_Angeles"),
    INCIDENT_DATE <= as.POSIXct("2019-12-31 23:59:59", tz =
"America/Los_Angeles")
  ) %>%
  arrange(INCIDENT_DATE)
# Rename LATITUDE and LONGIUDE to LAT and LONG
crimes19 <- rename(crimes19, LONG = LONGITUDE, LAT = LATITUDE)
# Create crimes19_XY: incidents with valid coordinates
crimes19_XY <- filter(crimes19, !is.na(LONG) | !is.na(LAT))
# Remove observations with invalid LAT
crimes19_XY <- filter(crimes19_XY, LAT > 33)
```

```
# Create crimes19_gc: incidents without valid coordinates
crimes19_gc <- filter(crimes19, is.na(LONG) | is.na(LAT) | LAT < 33)
# Remove 1,792 observations with missing address information
crimes19_gc <- filter(crimes19_gc, !is.na(ADDRESS))
# Remove 640 observations with missing street information (i.e., ADDRESS
value begins with a comma)
crimes19_gc <- crimes19_gc %>%
  filter(!str_starts(ADDRESS, ","))
# Remove 138 observations containing "UNK" (i.e., unknown) address values
crimes19_gc <- crimes19_gc %>%
  filter(!str_detect(ADDRESS, "UNK"))
# Note:
  # There are 160,129 total incident records obtained from the LASD
  # 152,058 records can be used with the XY Table to Point tool
  # 5,639 records will be geocoded with the CAMS locator
  # 2,432 records do not have valid location information
# Write two CSVs to use in ArcGIS Pro
write_csv(crimes19_gc, "crimes19_gc.csv"); write_csv(crimes19_XY,
"crimes19_XY.csv")
```

Appendix C

Univariate Visualizations

Figure C1

Histogram: Crime Rate (per 100 people)

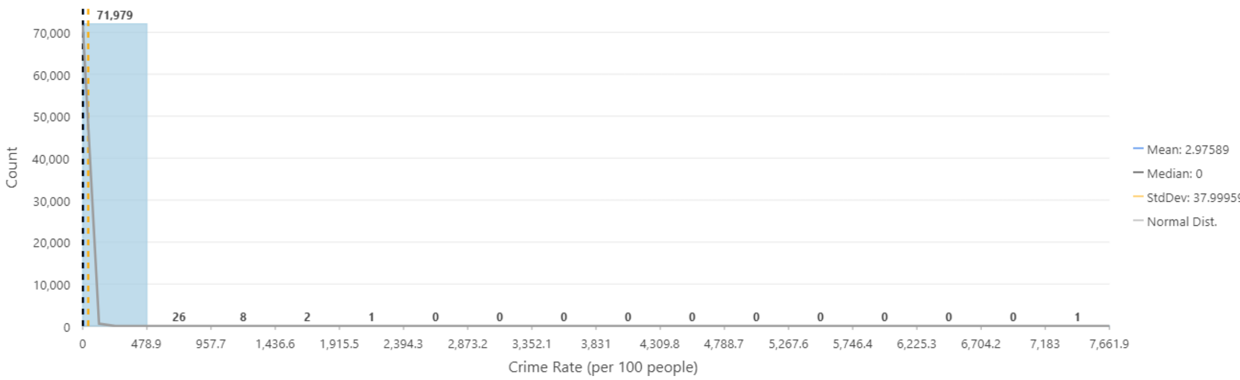


Figure C2

QQ Plot: Crime Rate (per 100 people)

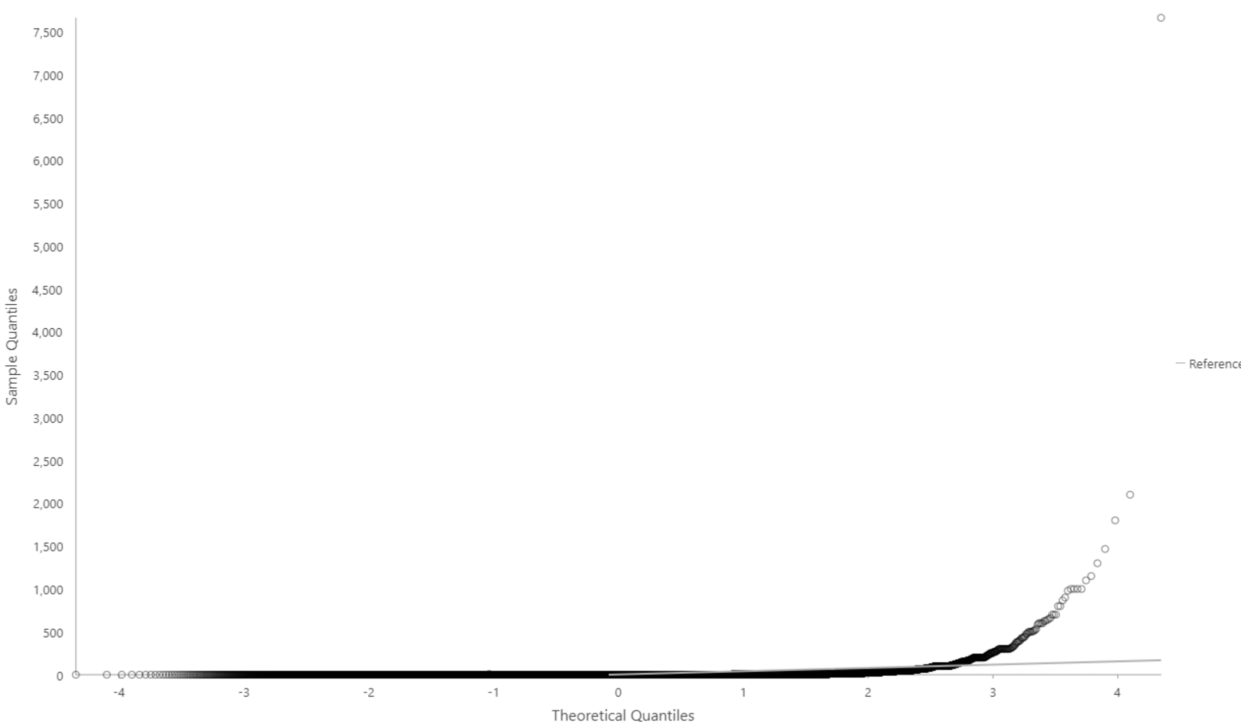


Figure C3

Crime Rate (per 100 people)

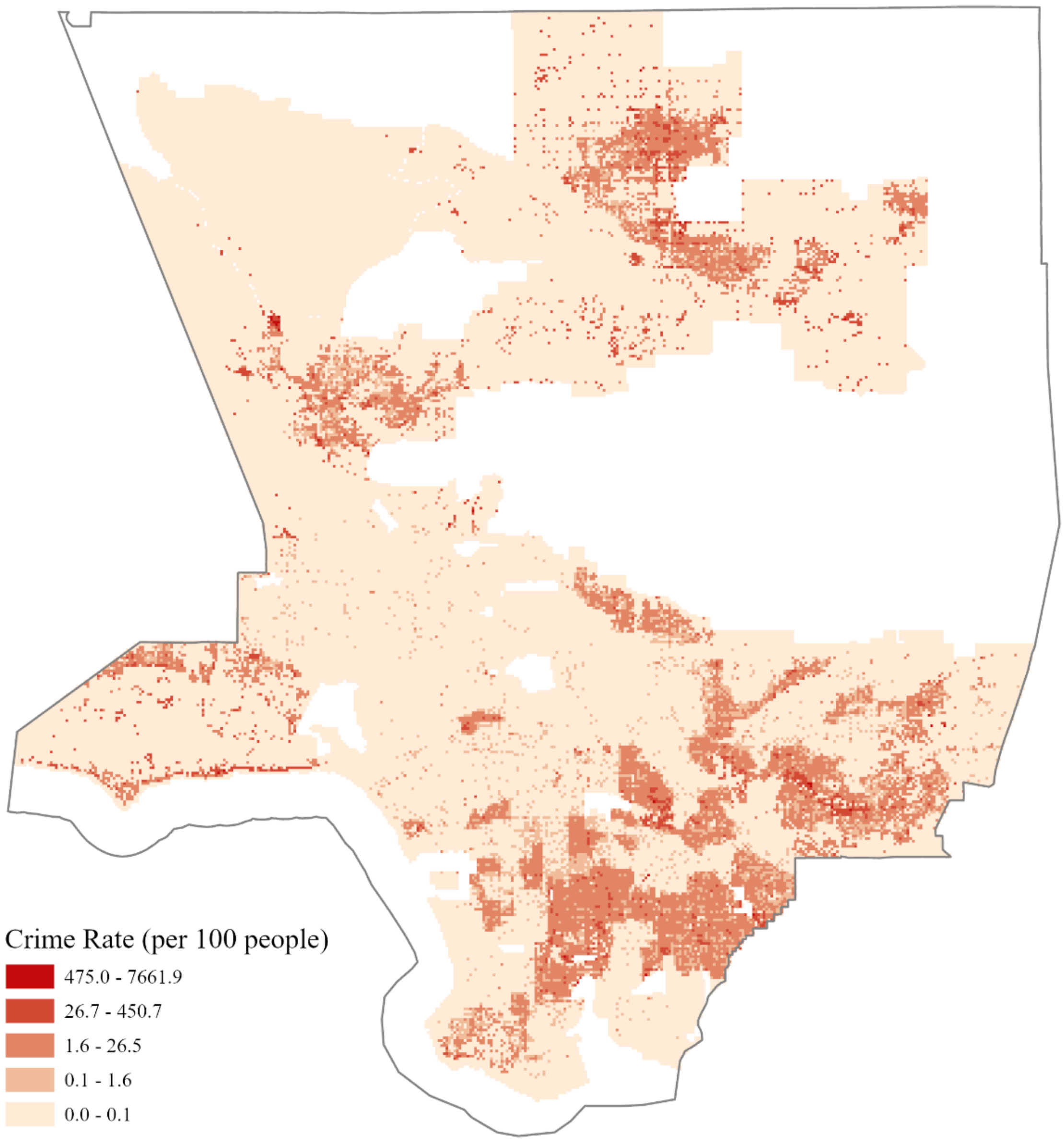


Figure C4

Histogram: Relief (meters)

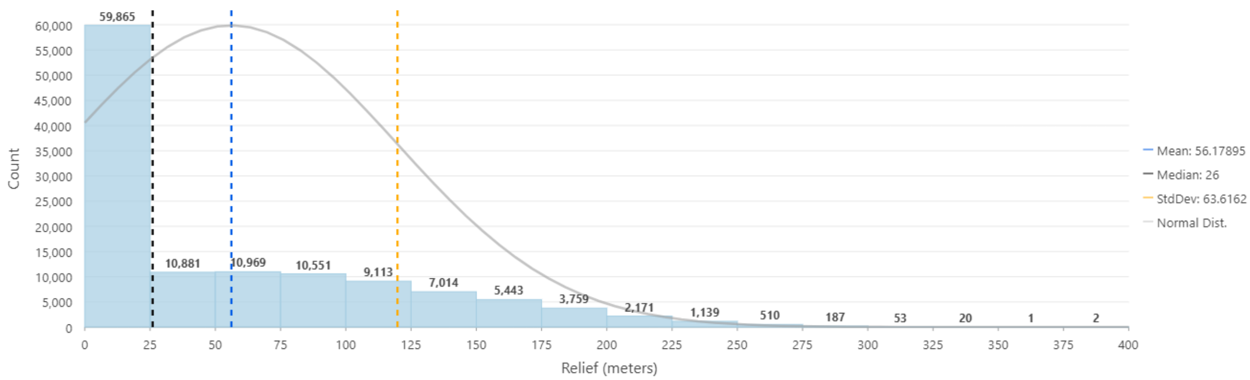


Figure C5

QQ Plot: Relief (meters)

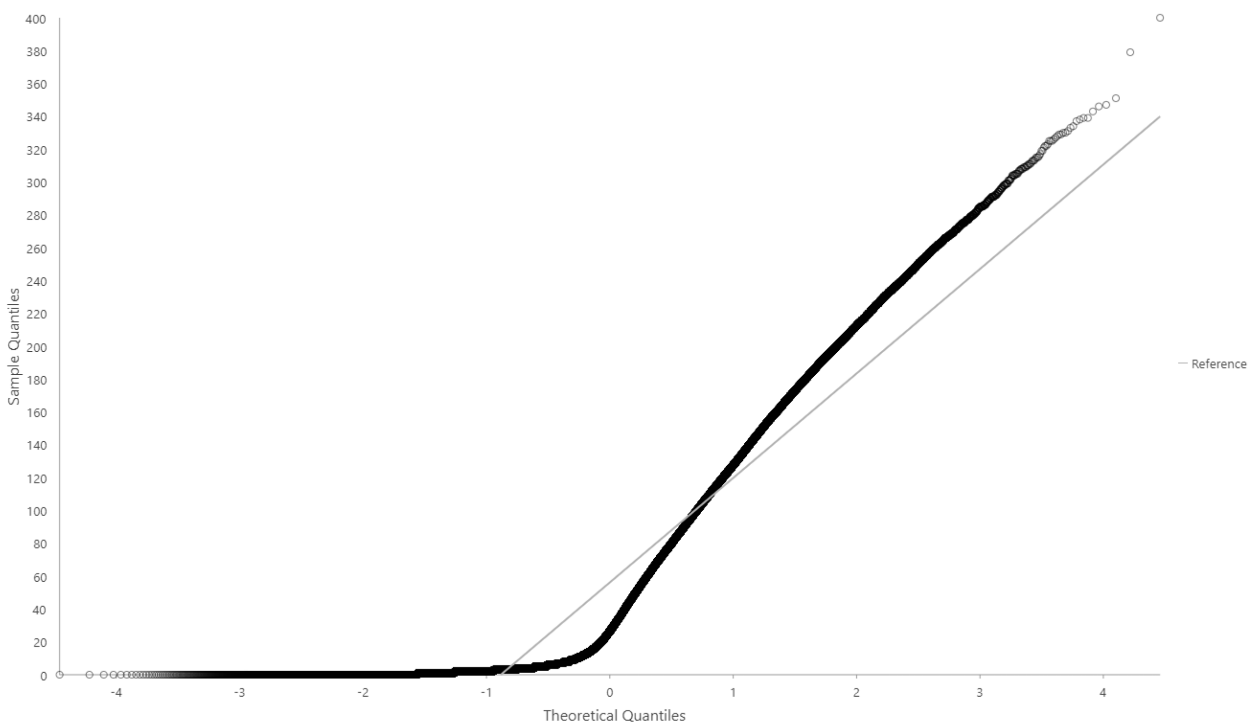


Figure C6

Relief (meters)

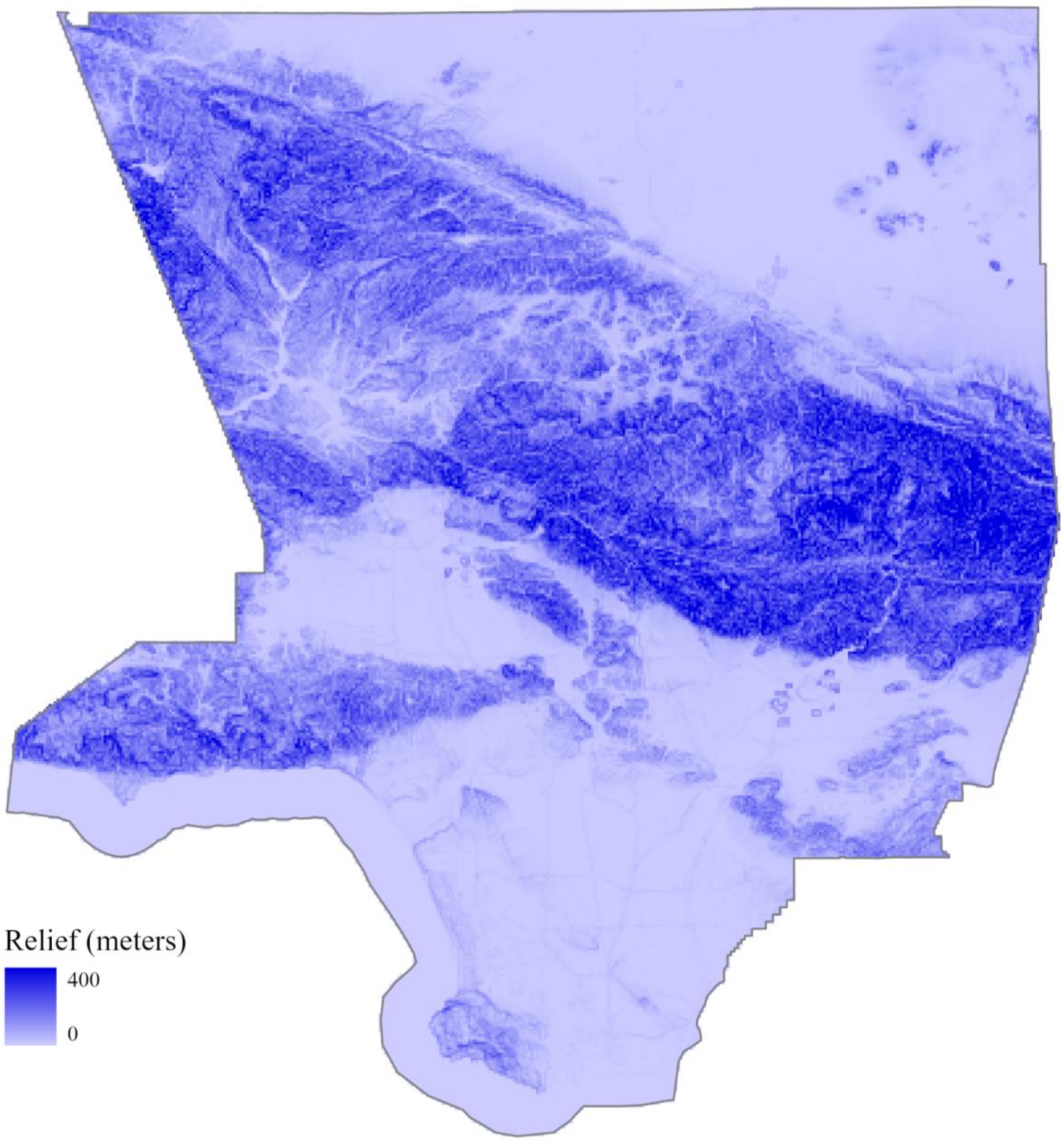


Figure C7

Histogram: Roughness (meters)

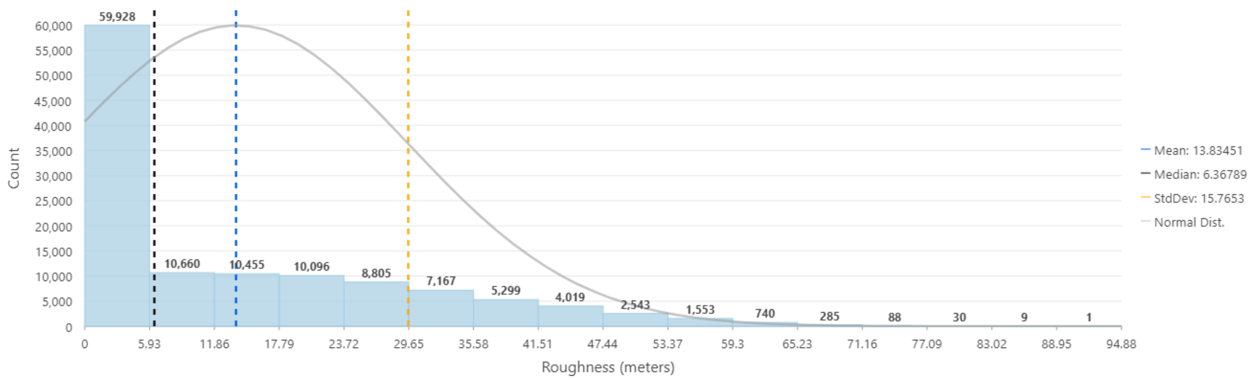


Figure C8

QQ Plot: Roughness (meters)

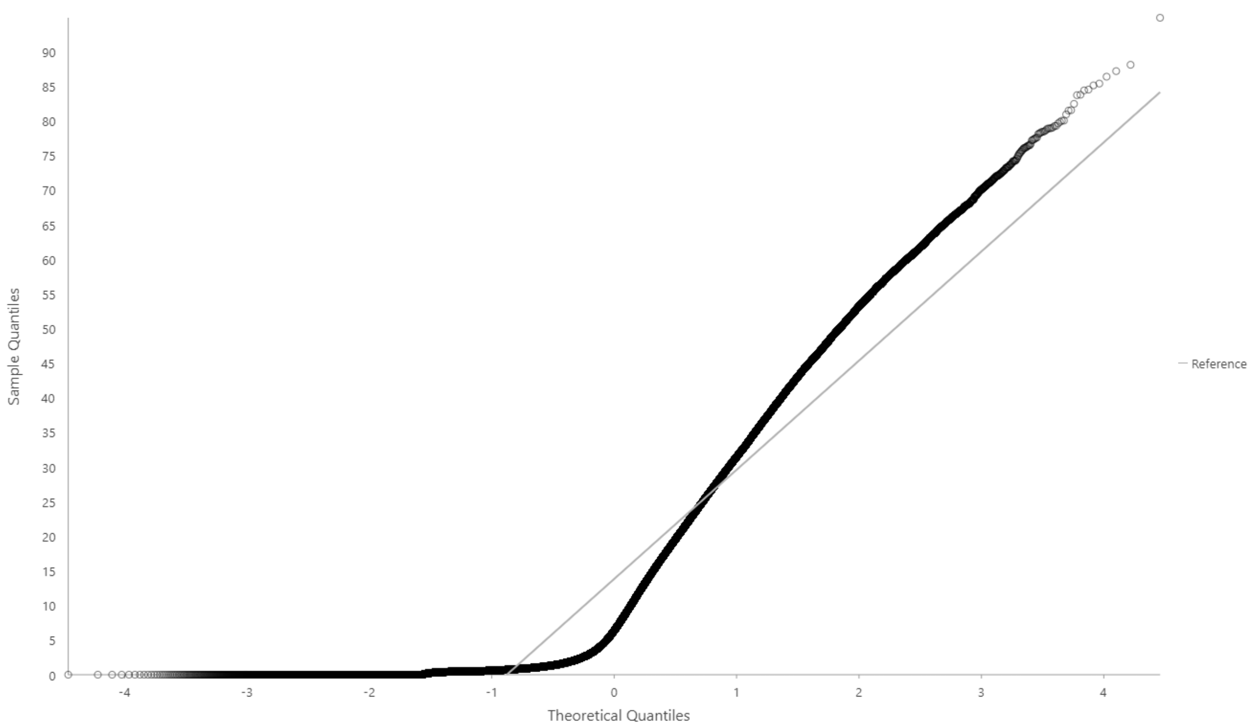


Figure C9

Roughness (meters)

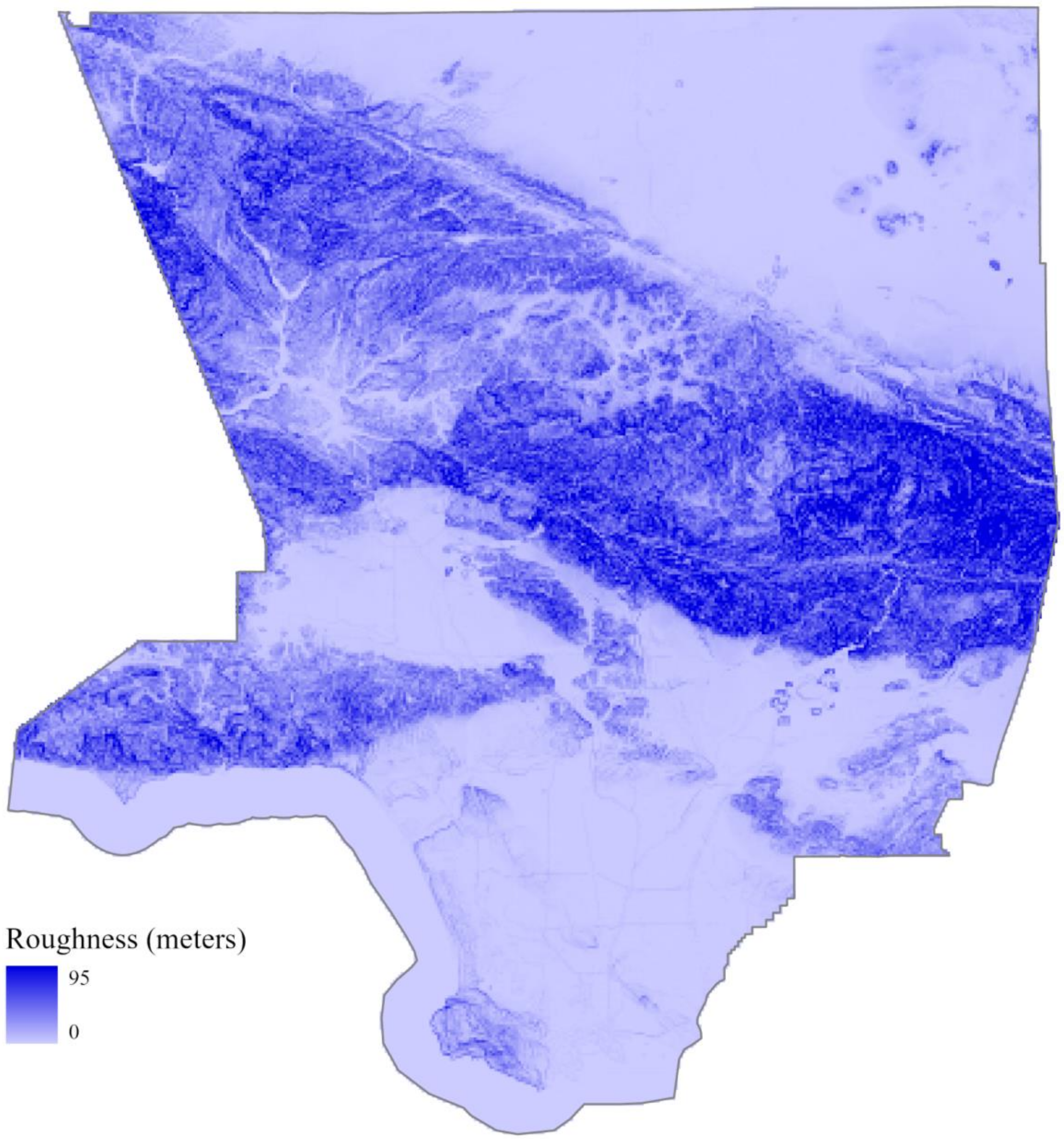


Figure C10

Histogram: Distance to the Nearest Mountain (meters)

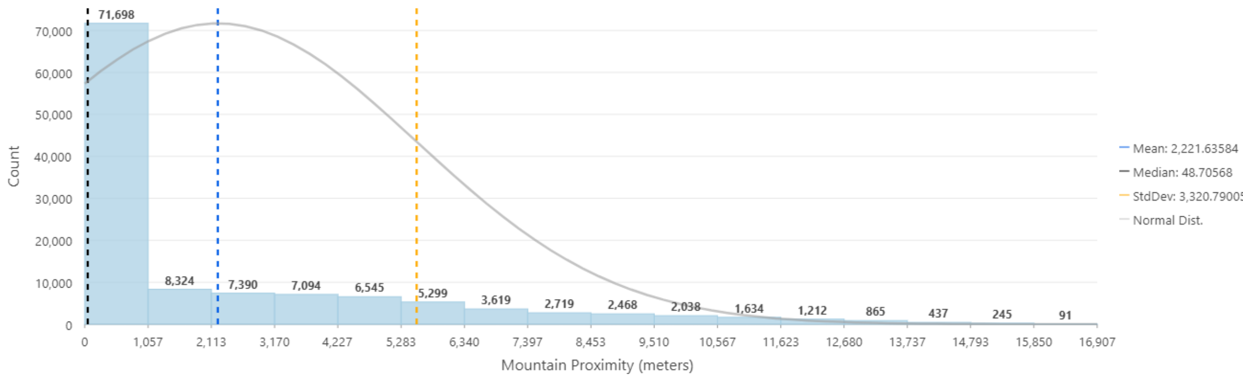


Figure C11

QQ Plot: Distance to the Nearest Mountain (meters)

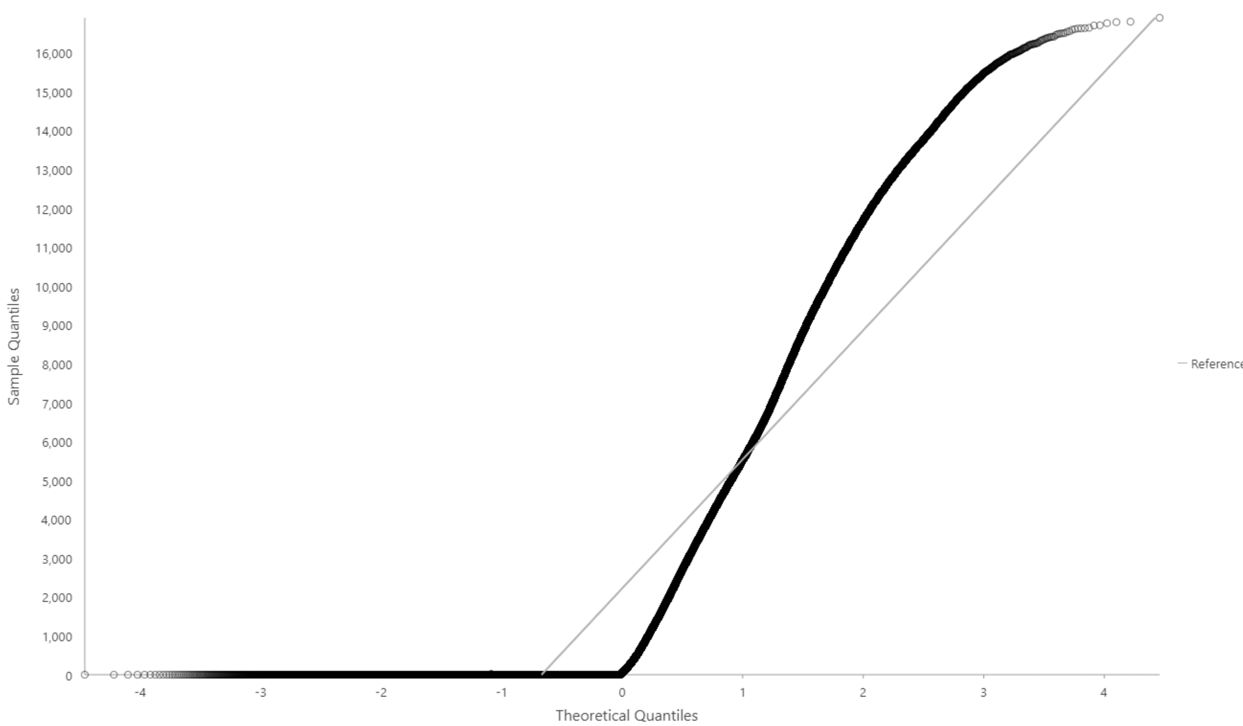
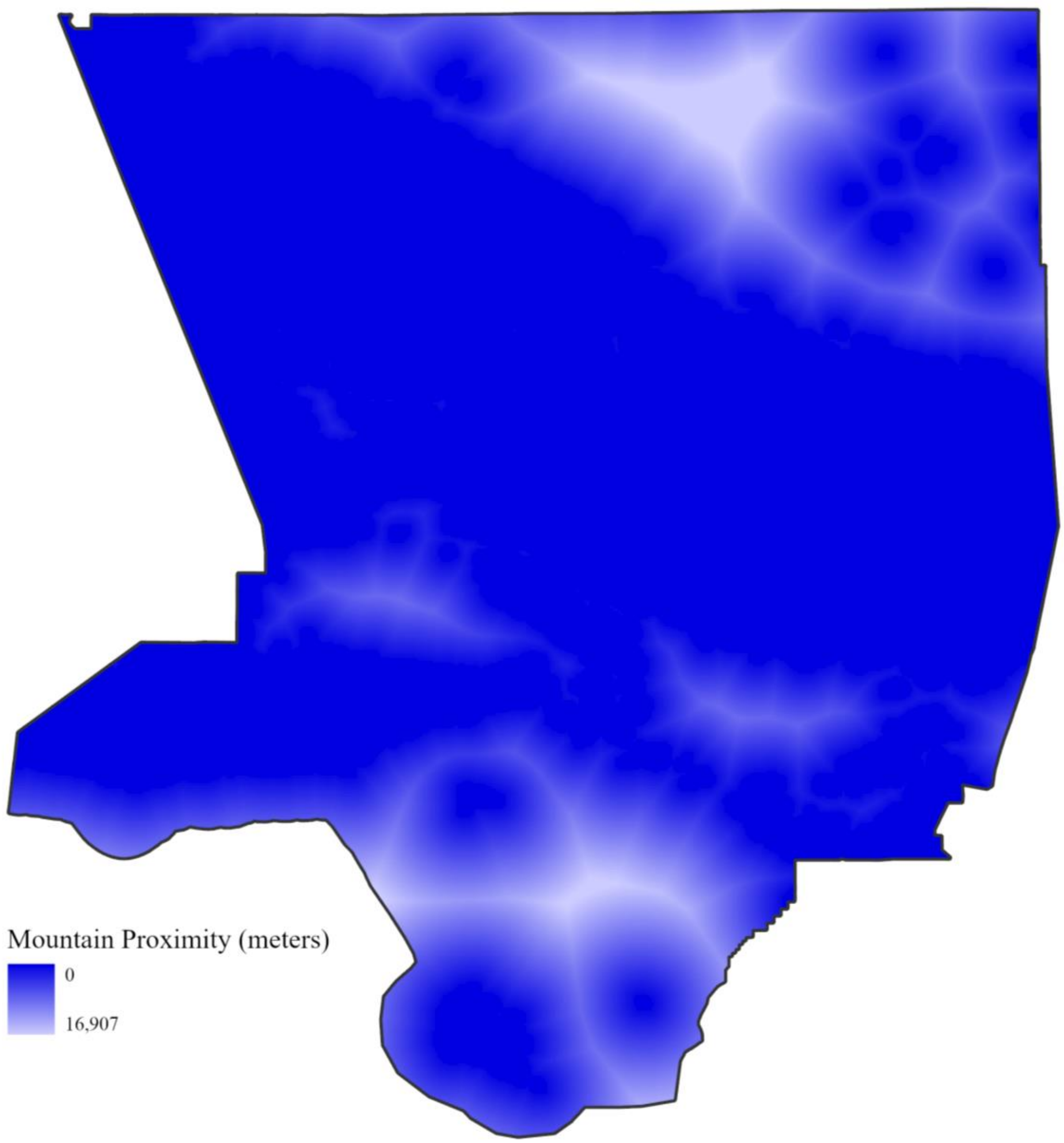


Figure C12

Distance to the Nearest Mountain (meters)



Appendix D

Bivariate Maps

Figure D1

Crime Rate vs. Relief

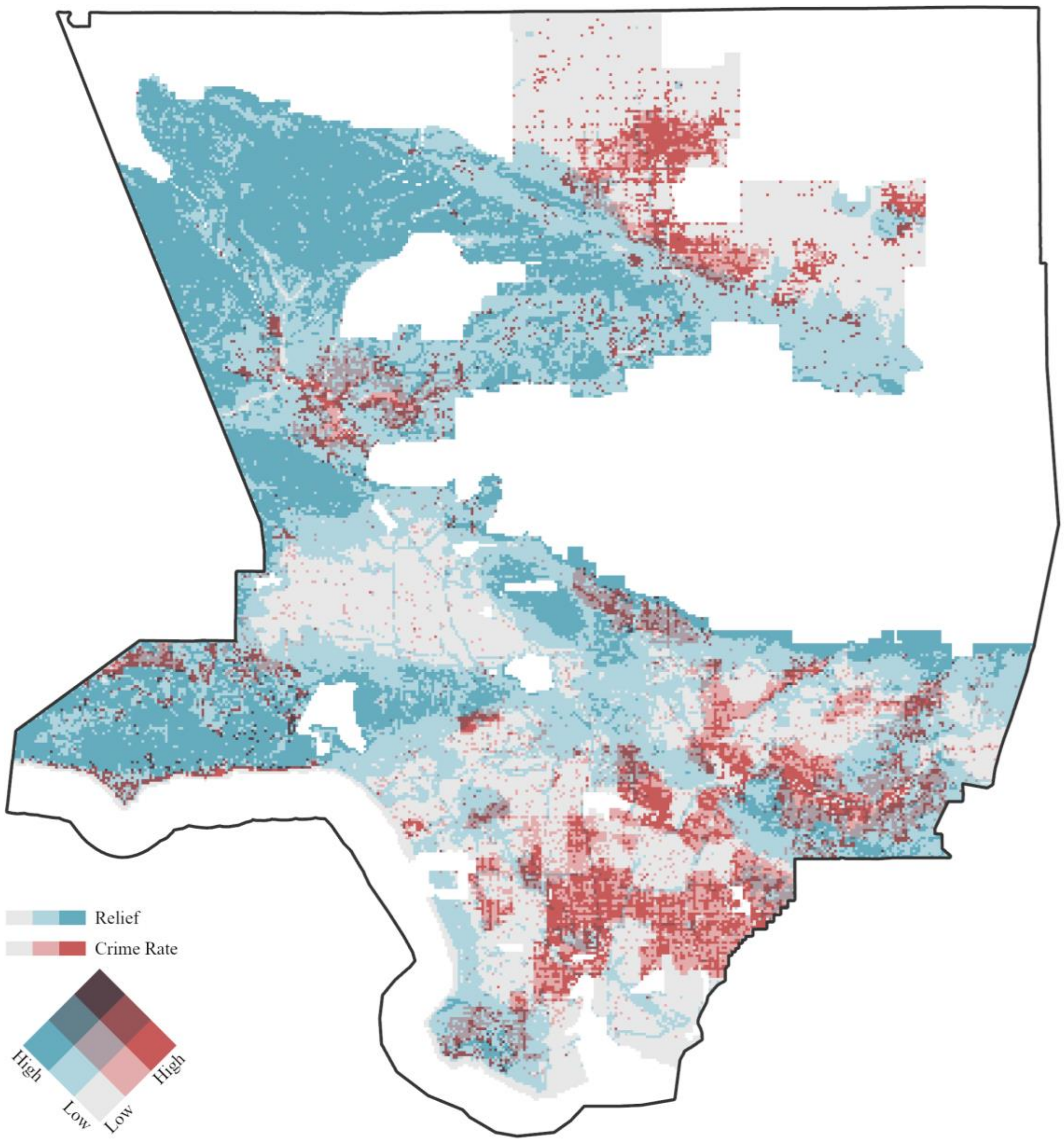


Figure D2

Crime Rate vs. Roughness

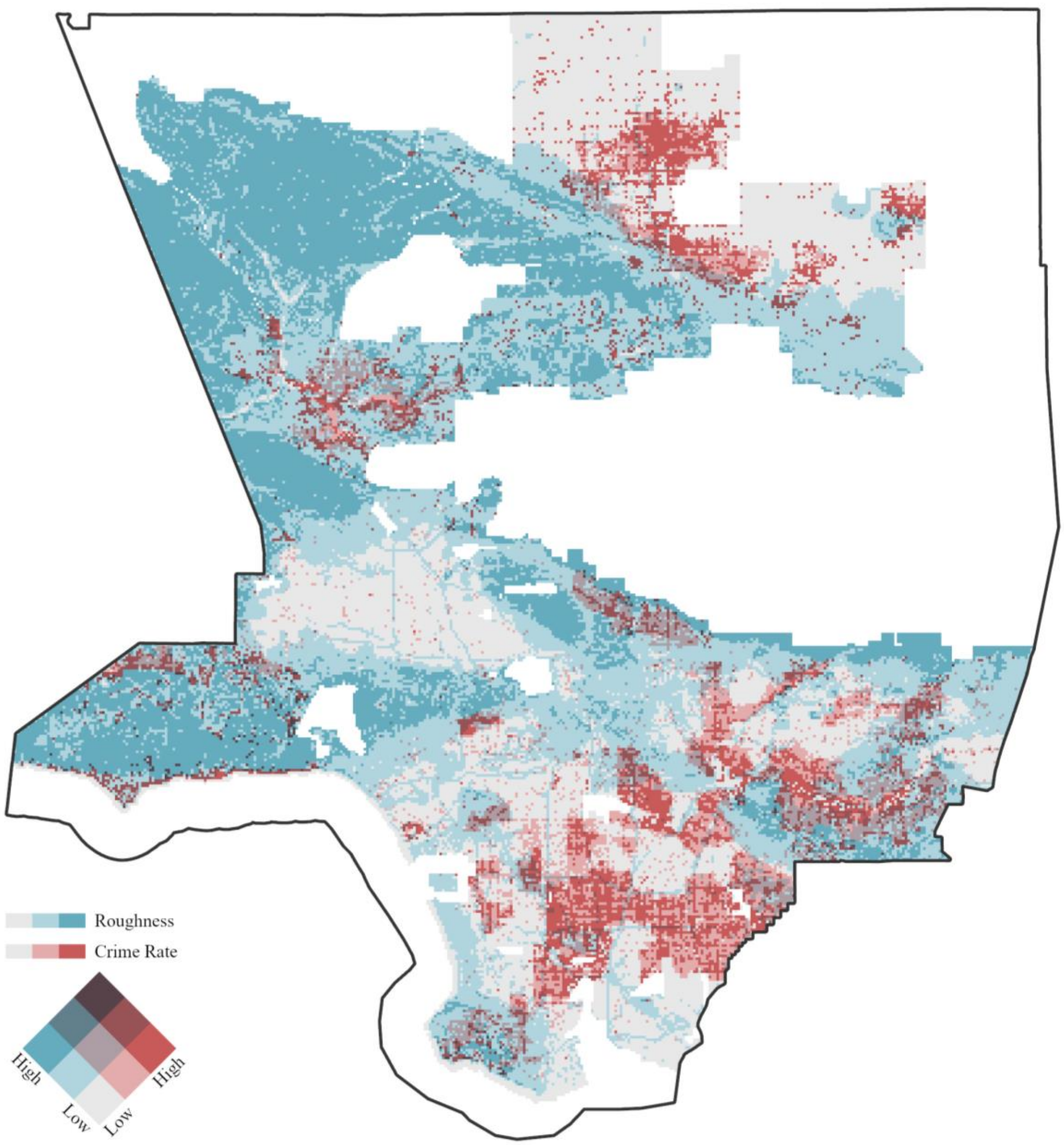
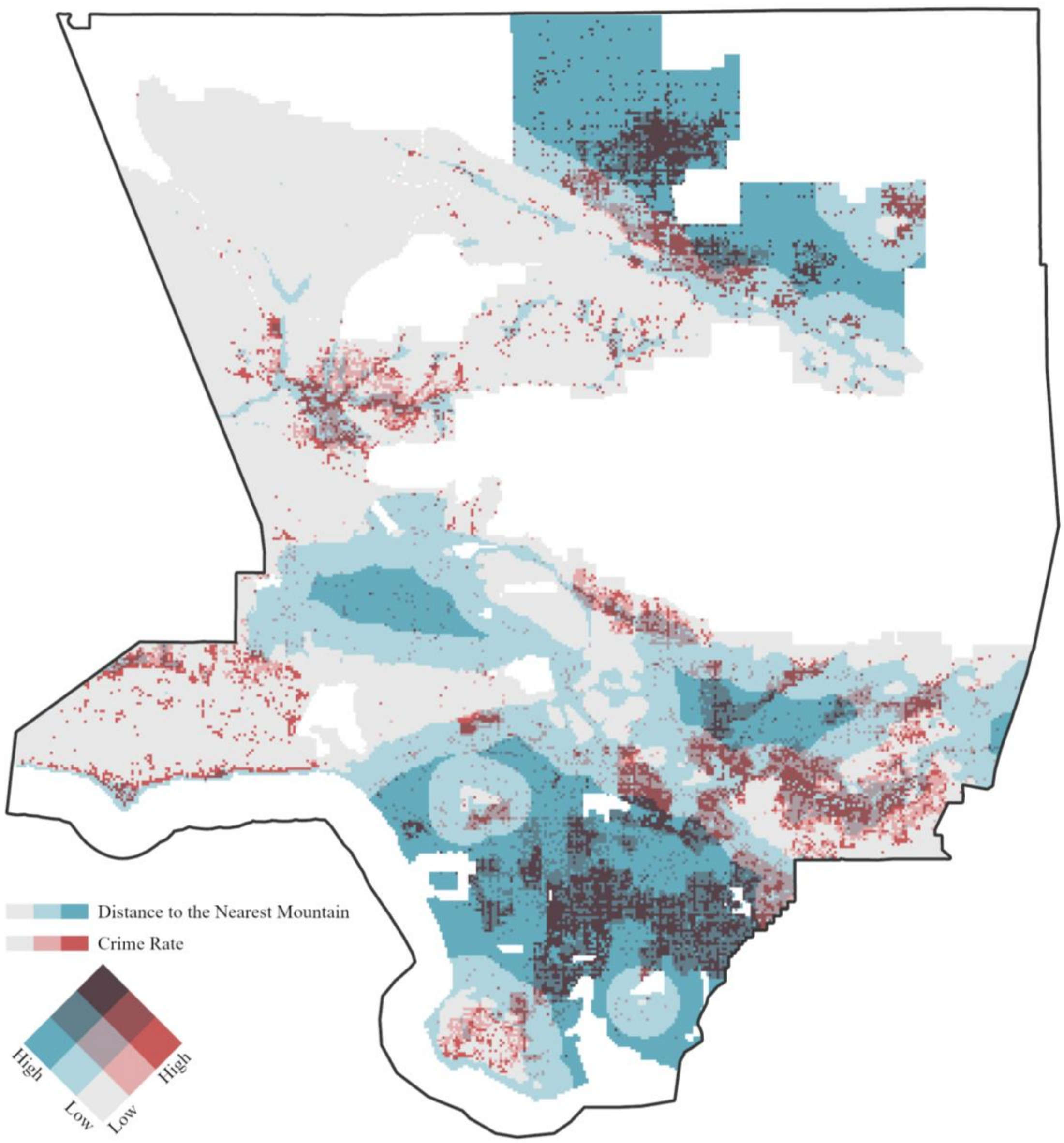


Figure D3

Crime Rate vs. Distance to the Nearest Mountain



Appendix E

Bivariate Spatial Association Maps

Figure E1

Crime Rate vs. Relief

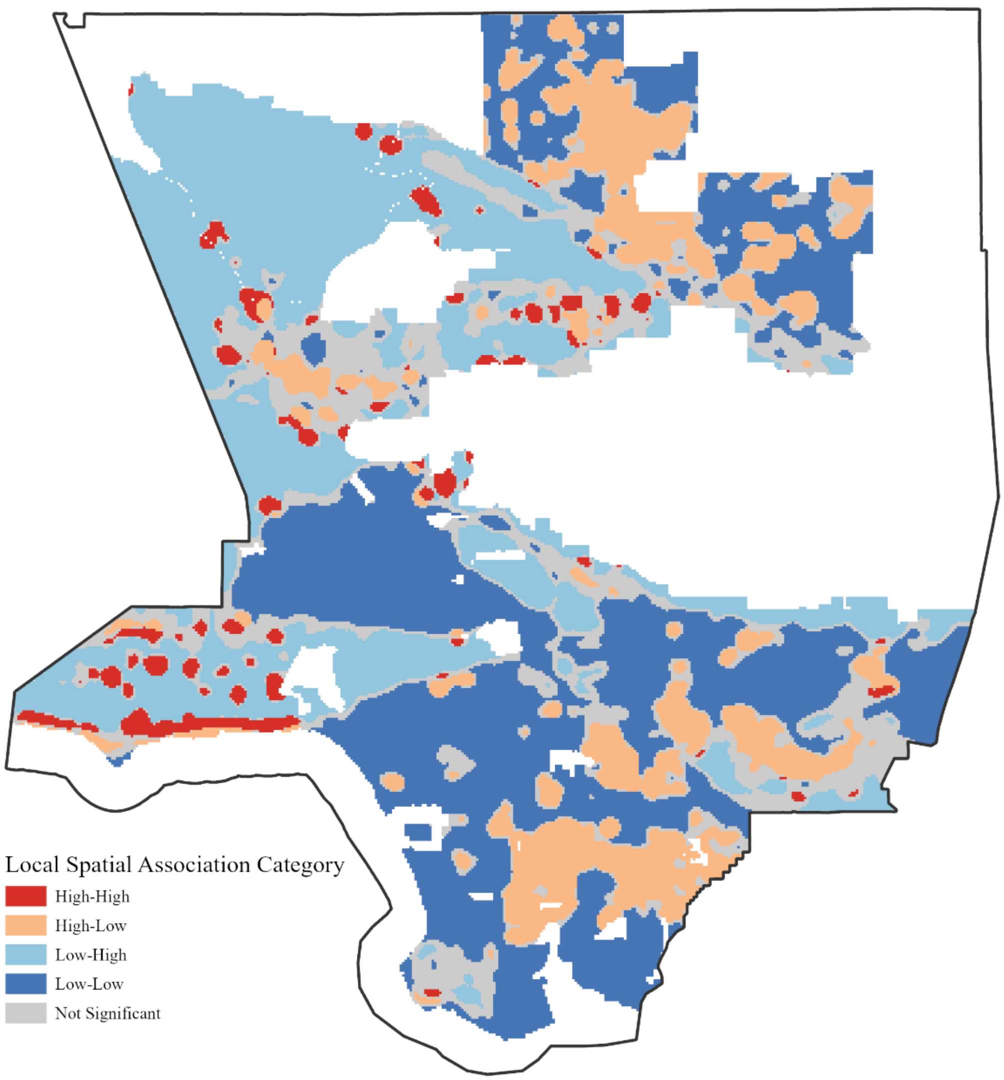


Figure E2

Crime Rate vs. Roughness

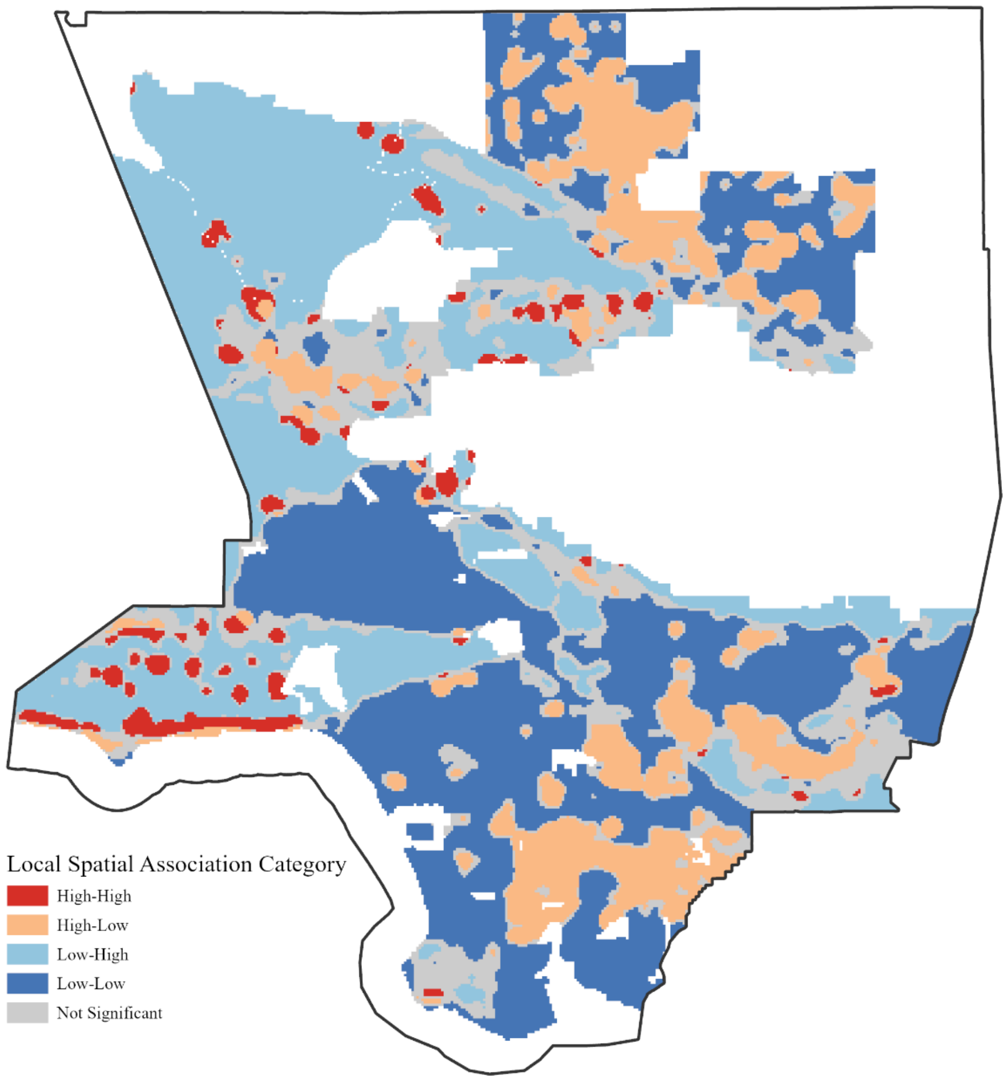


Figure E3

Crime Rate vs. Distance to the Nearest Mountain