

# Less Crime, More Dime

Disha Kedige Chandrashekarachar, Manan Rajdev, Posh Dave, and Rhushabh Vaghela

DSCI 554 Final Project  
University of Southern California,  
Los Angeles CA 90089, USA

**Abstract.** Los Angeles has been known to be one of the most dangerous and expensive cities in the United States. As a resident/incoming resident of Los Angeles, there is an extreme urgency to understand the economically safe neighborhoods within the Los Angeles County. The work in this paper focuses on providing residents a medium to compare the cost of living in the safe neighborhoods. The crime data and rent data are obtained from two different sources and are merged using the location coordinates. An interactive web application has been built to provide residents an overview of the Los Angeles Crime and Rent statistics over the years from 2010 to 2016. This work includes visualization on victim demographics, crime frequency in Los Angeles neighborhood, and safety score and rent for the neighborhoods. The results show a general inverse correlation between the crime score and rent prices and a direct correlation between the safety score and rent prices in Los Angeles.

**Keywords:** Police · Crime · Victim · Rent · Los Angeles · Neighborhood · Safety Precautions · Housing.

## 1 Introduction

### 1.1 Motivation

Rent prices in urban localities are affected by a number of factors: neighborhood, socioeconomic status of the residents, proximity to amenities, the quality of schooling, crime rates, etc. Los Angeles, a sprawling urban metropolis, thus serves as an ideal candidate for us to observe this cause-and-effect relationship at play. A major concern for residents and students living within and around the University of Southern California Parkside Campus is a comparably high crime rate, and as rent paying residents, we are particularly curious about how rent prices in this vicinity are impacted by the crime rate. This application will allow users to understand the crime patterns and get an overall sense of the economically safe neighborhood in Los Angeles.

### 1.2 Scope of Visualization

This work is addressed to residents, tenants, travelers and landlords of Los Angeles. It will be helpful for residents and tenants to gauge the rent prices and crime rates around Los Angeles during house search. A traveler can understand which areas are safe and economical to temporarily reside in. Finally, the landlords of Los Angeles can utilize this tool to optimally charge their tenants based on the crime indicated in the area. This is an interesting project since the amalgamation work of rent and crime in Los Angeles has not been undertaken yet and through our application, we will be able to visually represent this correlation in LA neighborhoods.

### 1.3 Related work

All the other visualizations till now have focused only either on the crime data/rent data. Currently, the work that has been done has heavily focused on crime data sets. In Los Angeles neighbourhoods[4], the Los Angeles Times has a database and its visualization of crime statistics, with forms of crime and their intensity identified on the map of the region chosen. It is up-to-date and delivers warnings in the selected area about recent violent crime. They have made the Homicide Report[6] on the map detailing murders in the last 12 months and all the victims, connected to the news report if they were reported. A web tool called, SpotCrime[9], also offers crime data with various forms of crime showing up on the map around the user's query area. Besides this, there has not been any visualization conducted for the Los Angeles housing data set. To enhance the already conducted visualization, this project aims at building cohesive and demonstrative illustrations for the combined data of both rent and crime.

## 1.4 Originality

This project work stands out from the rest of the work conducted till date because of the combined analysis of crime and rent information in Los Angeles. The two data sets have been merged using their geographical locations in order to establish a correlation between them. This type of merging required us to use latitude and longitude and pre-processing steps to identify standardized crime types and obtain crime scores. A grid system was built to merge the data sets using longitude and latitude. This combined data was then utilized to extract the correlation between the crime and rent in Los Angeles. Moreover, this data was also used to obtain the best and worst neighborhoods to live in based on the crime scores and rent.

## 1.5 Overview

The platform is separated into five sections to guide the user learning experience: the introduction (home-page), exploration of crime analysis, safety and rent correlation, quick overview of the neighborhoods in LA, and about page (creators' descriptions). The introduction explains in the detail about the datasets and data pre-processing steps that were undertaken prior to conducting the analysis and visualization.

# 2 Data

To understand the crime patterns in Los Angeles and the relationship between the rent price and the crime rate, we have used two different datasets.

## 2.1 Crime Data

The crime dataset[3] was made available through the City of Los Angeles, via the Office of the Mayor. The data available here was extremely detailed and comprehensive and was meticulously kept up to date by the city, i.e. any and all reported crime between 2010 and 2019 was logged here and included information such as Date and Time of Occurrence, Area, Victim Age/Sex/Race Description, Crime Description, Location, Weapon Used, Crime Code etc.

## 2.2 Rent Data

The dataset, Rent Price LA[8], prepared by USC Sol Price Center for Social Innovation provided median rent price data for the city of Los Angeles from the years 2006 to 2016. This dataset included information such as Year, Neighborhood, GEOID, Location, Date, Amount, Tract etc.

## 2.3 Data Pre-processing

### – Extracting Attributes

The first task was to combine the two data sets into a singular data set, removing redundant information, extracting that which would be required and then formatting it into a usable template. For pre-processing, the data was to read the Rent Price LA data and extract only the variables needed for the data frame from the available list, i.e. Year, Amount, Location and Neighborhood. Subsequently, the location data was split into two, i.e. Latitude and Longitude, and the Location column was then removed. The same step was conducted for the Crime data, where the date of occurrence, crime code description, latitude, longitude, victim age, victim gender and victim race were extracted.

### – Building Merged Dataset

The next task was to merge the two data sets, by obtaining a list of names of all the 141 unique crime descriptions from the crime data set and adding each as a column to the Rent Price LA data frame. Following this, the data sets were iterated over year by year, starting off with 2010 and ending at 2016. For a given data point in the Rent Price LA data set, a filter of  $\pm 0.005$  degrees Latitude and Longitude of the given location was applied for the crime data set. Iterating over each incident in the crime data subset, the frequency of each type of crime for that rent listing was incremented. Once completed,

the data frame now provides a comprehensive view of all crimes in the vicinity of each listing as well as their frequencies. This data frame was saved as an intermediate database to be used for collective analysis.

#### – **Calculating Crime and Safety Scores**

The last step was to calculate the crime and safety scores for each neighborhood. The crime data set consists of a list of crimes with varying severity. Intuitively, it can be understood that the more serious the crime the higher it impacts the perception of security in a location. Based on [1], weights are assigned to every crime and an overall value is calculated for each location. Each crime in the crime data set was compared with the list of Notifiable Offence List Categories and given the Crime Severity Score. The Crime Severity Score (CSS) is the score assigned to a crime based on the mean sentence passed on those who were convicted of the offense. Each score is normalized and the final crime values for each location is calculated using the dot product of the crime weight and the crime frequency matrices. Using these crime scores, the safety scores are calculated by taking the negatives of the crime scores.

### 3 Approach

#### 3.1 Design Process and Considerations

This work has been conducted to build an interactive web application for an everyday technical user. Through this web application, the aim was to allow users to answer some of the basic questions related to crime and rent statistics in Los Angeles. Few of these questions include:

- 1) Which victim race is affected most in which neighborhood?
- 2) Which neighborhoods are safe to reside in? How expensive are these neighborhoods?
- 3) Are there neighborhoods where the rent is relatively low and crime is low as well?
- 4) How has the rent and crime changed over the years for Los Angeles and specific neighborhoods in Los Angeles?
- 5) How does age and gender impact the frequency of crime taking place?

To keep these questions into considerations as well as the technical capacity of a wide range of audience from residents to police to landlords, the work uses simpler, yet multidimensional chart techniques and incorporates interactivity to give users more detailed information if desired. Moreover, a standard color blind safe theme has been used throughout the web application to encompass various users. In order to provide a more personalized touch to the website, a logo has also been created using the same color theme.

#### 3.2 Visualization Wheel

Following visualization wheel provides a justification of our design choices to incorporate user requirements and questions that needed to be answered.

##### – **Figuration-Abstraction:**

The web application uses various symbols and texts rather than representing them photographically. This makes our website more abstract.

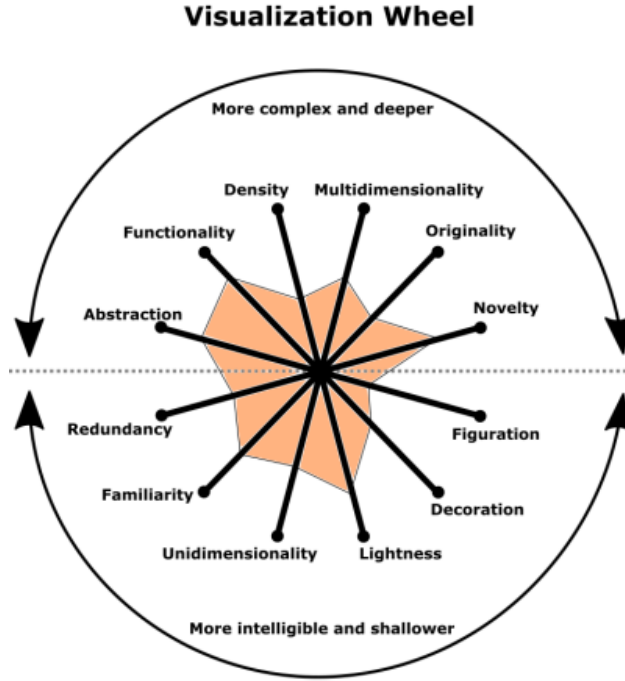
##### – **Decoration-Functionality:**

Since the focus of the web application was to answer the design questions in mind, there is a heavy weightage given to functionality. Through functionality, the users can depict the correlation and get more useful insights. Moreover, to maintain uniformity, the same color-blind safe color theme has been used throughout the web application as the decoration element.

##### – **Familiarity-Originality:**

The web application consists of word cloud, line chart, bar charts, pie charts, zoom able heat map, and circle packing. All these charts are relatively easier to understand for a non-technical user. This was a conscious choice made to consider all kinds of users.

Fig. 1. Visualization wheel.



– **Lightness-Density:**

By using multiple page functionality of the Vue Framework, the web application was able to maintain a structure for the ease of understanding. This allowed the pages to not be cluttered with multiple information on one page, yet gaining complete insight from the application.

– **Unidimensionality-Multidimensionality:**

The major goal of this project was to show the correlation between the crime and rent data of Los Angeles. Since, this required using multiple data sources, most of the charts are multi-dimensional charts. However, to also understand the significance of single attribute, the work includes uni-dimensional charts as well to give holistic picture.

– **Redundancy-Novelty:**

To gain the best insight from each attribute, there is no redundant use of any attribute. This was, again, a conscious choice to give users appropriate amount of information in every visualization.

## 4 Development, Insights and System

### 4.1 Crime Data Analysis

In this page, the focus was on the initial analysis of the crime data. From the crime data, the aim was to understand the relationship between the different attributes such as neighborhoods, victim demographic and the crime frequency. The users can interact with the visualizations to better infer the patterns between crime frequency and victim demographic in different neighborhoods in Los Angeles. The first chart indicates the frequency of crime taking place by age and gender. Through this, the user can understand which gender is hit worse in every age group. This is an interactive chart which allows user to hover over either the bar/pie chart to obtain dynamic information and update the charts accordingly. The second chart is a circle packing with additional zoomable functionality which shows the crime by neighborhood and victim's race. It allows users to determine the race that is affected most in every neighborhood by clicking on the neighborhood.

However, it is important to note that this chart indicate absolute crime frequencies taking place for the race and not ratio of the population of the race in the neighborhood. The last chart on this page is a mapbox with a heatmap which shows the intensity of crime frequency taking place in all the neighborhoods of LA. This chart gives an overview of safe and unsafe neighborhoods relative to each other on a map.

## 4.2 Safety and Crime Correlation

To understand the relationship between the rent and crime more cohesively, a new attribute, called, “Safety Score” was introduced. This was calculated to consider that different crime types have different severity levels and hence, should be given higher weightage even if their frequency is low. Therefore, the crime score was first calculated by aligning each crime type to it’s severity. After calculating the crime score, the safety score were calculated using the negatives of the crime score. Detailed explanation can be found in the data pre-processing section. In the choropleth map on this page, the colors indicate the average rent price for each neighborhood. When hovered over a neighborhood, the average rent over the years in that neighborhood is shown in the tooltip. Through the interactive feature, which allows users to select a specific neighborhood on the map, one can dive deeper into the trends of both safety score and average rent over the years for that specific neighborhood. In general, the user can observe that as the safety score increases, the rent price increases.

## 4.3 Quick Overview

To understand various neighborhoods of Los Angeles, four criterions were selected and a word cloud was built using the top 10 neighborhoods. Through this word cloud, users can directly check the best ten and worst ten neighborhoods to live in as well as the safest ten and most dangerous ten neighborhoods to live in. The best neighborhoods represent the lowest rent and lowest crime rates and the worst represent the highest rent and highest crime rate areas. The safe neighborhoods are the neighborhoods with the lowest crime rates while dangerous neighborhoods are the neighborhoods with the highest crime rates. Moreover, an option of selecting the information per year has also been included.

## 4.4 System Technologies

The visualization platform has been built using HTML, CSS, D3 Javascript library[5] to manipulate data and build interactive charts. Vue framework[10] has been used to bundle up the application and deploy it. Besides this, python was used to conduct data pre-processing.

## 5 Conclusion

In conclusion, the interactive web application showcases the best and worst neighborhoods to live in Los Angeles over the time frame of 2010 to 2016. Moreover, the users can observe an inverse correlation between the crime score and rent prices and a direct correlation between the safety score and rent prices. This indicates that as the crime in a neighborhood increases, the rent decreases. Since this tool was inspired from personal difficult experiences to find safe housing options in Los Angeles, our hope is that users can get a clearer picture regarding the crime and rent in Los Angeles.

## 6 Future Scope

The current work can be further developed in two different directions. The first direction will be the focus on building dynamic interactive visualizations prompting users to guide their own learning journeys by exploring the illustrations. This will allow users to explore the data more in depth and understand the correlation further. Moreover, the map box functionality can be improved with the addition of interactivity using a pop-up window showcasing specific data. The second direction will be to collaborate with LAPD, local apps such as ”Citizen” [2], ”LiveSafe” [7], etc, and landlord to obtain real time information on crime and rent to represent current data.

## References

1. Ashby, M.P.J.: Comparing methods for measuring crime harm/severity. *Policing: A Journal of Policy and Practice* **12**(41), 439–454 (2018)
2. Citizen. <https://citizen.com/>
3. Crime data from 2010 to present - los angeles police department. <https://data.lacity.org/A-Safe-City/Crime-Data-from-2010-to-2019/63jg-8b9z/data>
4. L.a. crime maps - los angeles times. <http://maps.latimes.com/crime/2>
5. d3js. <http://d3js.org>
6. Victims in the last 12 months - the homicide report - los angeles times. <http://homicide.latimes.com/>
7. Live safe. <https://www.livesafemobile.com/>
8. Rent price (la). <https://usc.data.socrata.com/Los-Angeles/Rent-Price-LA-/4a97-v5tx>
9. Spotcrime. <https://spotcrime.com/>
10. Vue - the progressive javascript framework. <https://vuejs.org/>