

Correlations of crime and land use in Chicago

Jorge Gerardo Acosta Montes

1 Introduction

This project explores the correlations between the amount and types of crimes committed in a community area and the venues located there.

Crime is a multi-factorial phenomena, and the study of only one component should acknowledge other variables. In order to have meaningful comparisons between two or more regions. The relations between crime and socioeconomic conditions are unclear. Often there are contradictory reports on whether difficult economic conditions create more crime or whether a booming economy that increases the availability of money and goods promotes crime. Nevertheless, it is important to do not dismiss possible connections between the economy and crime. In this project, I will try to find if the crimes committed in community areas with similar socioeconomic indicators have a relation with the land use. For example, let's suppose that A and B are community areas with similar economic conditions, but area A is much safer than area B. I want to see if the types of venues in area A are similar to those in area B. If they are not similar, then study what types of venues are present in the safer environment. The venues could be parks, liquor shops, restaurants among others.

Also, venues that reduce or increase crime in one type of area do not necessarily have an impact in crime in other types of areas. One possible hypothesis is that parks can be that kind of venue. In areas with a high rate of crime, social interactions can induce people to become criminals. Parks promote social interaction. Therefore parks there contribute to crime. On the other hand, parks will not affect crime in areas with low rate crime.

This project could help policymakers to design safer neighborhoods by building or closing venues to reduce crime.

2 Data

Cultural and geographical factors affect societies in complex ways. A study of crime among different cities with different cultures will need to take in account cultural indicators, and will require a score that already take those indicators to compare to compare two different cities. This is a modest study, and I will dismiss all cultural factors. To reduce the effects of those variables, only areas in the same city will be considered. Due to the availability of the data, I choose Chicago. The datasets of the crime and economic status of the neighborhoods in Chicago were already given in previous courses of this certification, and were made by the city of Chicago. The socioeconomic data can be downloaded in <https://ibm.box.com/shared/static/05c3415cbfbtfnr2fx4atenb2sd361ze.csv>. Crime data can be downloaded from <https://ibm.box.com/shared/static/svflyugsr9zbqy5bmowgswqemfpm1x7f.csv>.

Detailed explanations of the data can be found in <https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2> and <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>.

The dataset of socioeconomic indicators contain 9 parameters and divides the city in 77 areas. A description is provided in the ones that are not clear:

- **Community area number**
- **Community area name**
- **Percent of housing crowded:** Percentage of houses with more people than rooms.
- **Percent households below poverty:** Percentage of houses with an income less than the federal poverty level. The table contains data from 2008 to 2012. In 2012 it was an income of 23 050 for a family of 4.
- **Percent aged 16+ unemployed**
- **Percent aged 25+ without high school diploma**
- **Percent aged under 18 or over 64**
- **Per capita income**
- **Hardship index:** It is an score that combines the 6 socioeconomic indicators in this dataset.

The dataset about crime contains a wide description of parameters of crime incidents. I only use:

- **Primary type**
- **Community area number**
- **Latitude**
- **Longitude**

The land use was determined by the venues that are in a community area. The venues were obtained with the Foursquare API. The location of each community area was assigned as the mean latitude and longitude of the crimes committed in that area. For each area, I set a limit of 50 venues and a radius of 500 meters. I obtained 1219 venues. Of this dataset I only use:

- **Community area name**
- **Venue Category:** Examples include parks, restaurants, bars and recreation centers.

3 Methodology

The community areas were clustered by k-means using the economic indicators of the socioeconomic dataset with the exception of the Hardship index, because that index is already a combination of the other parameters. All the indicators but the per capita income are in a percentage form. The per capita income was transformed to a percentage form, in which the 100% corresponds to the average per capita income in Chicago.