“The Wall” Project: Crime and Border

CMPSC 310: An Introduction to Data Science - Spring, 2019

By Thao Phuong, Bakhytzhan Altynbayev

Abstract

Many citizens in the United States have been concerned with crimes, and worry that they will be victimized by [illegal] immigrants who enter the USA from Mexico. Additionally, whether the border wall is a critical factor for reducing the crime rates is still an ongoing controversial topic. This study explores the connection and relationship between the crime rates (per 100,000 people) of 83 major cities in the United States and their smallest distance from the Mexico—United States border. Violent crimes include homicide, rape, robbery, and assault. The study’s result suggests that there is no apparent relationship between the two variables.

Introduction

There is a continuous discussion about whether or not the [illegal] immigrants infiltrating the USA from Mexico positively affect the crime rate in the border neighborhoods. The Project’ goal is to cast some light on the situation using data available from Wikipedia.

Methods for data acquisition and processing

- Download and store the necessary html files in the local workspace for further processing purposes
- Search for the important data in the local html files using the imported module BeautifulSoup.
- Use the *find()* and *findall()* methods from the BeautifulSoup library to locate the following *Beaming indicators* (criterias which will help allocate the needed table):
    - “wikitable sortable” table
    - ‘tr’
    - ‘tbody’
    - ‘td’
- Use the *beaming indicators* to find the cities’ names, crime rates and links to each city’s html page.
- Extract the coordinates for each city from its separate html page
- Use the *smallest_distance()* to calculate and get the smallest distance from each city to default border cities.
- Once the data (cities’ names, crime rates and distances) is properly collected, calculated and processed, write the data into a CSV file.
- Calculate the coefficient of correlation based on collected crime rate and smallest distance parameters for each city.

Results

After calculating the coefficient of correlation based on crime rate and distance to border, the result was 0.31578.

```
In [1]: runfile('C:/Users/Bakhosh/Desktop/Python/Project The
Wall/Final_Version.py', wdir='C:/Users/Bakhosh/Desktop/Python/
Project The Wall')
The correlation between the city's crime rate and the city's
distance to border:
0.31578
```

Conclusion

The coefficient of 0.31578 is a positive weak correlation, which means that there is a direct relationship between two variables in which as one increases, the other can be expected to increase. Thus, if the distance from the city to the border increases, the crime rate should also increase, and vice versa. Additionally, since the coefficient is allocated closer to 0 rather than to 1, the relationship between variables is considerably weak. This founding indicates that the variables which were given to calculate the coefficient are possibly not dependent nor affecting each other. If there is any relationship between the two data set at all, it is still a direct relationship. Base on this supportive data, [illegal] immigrants infiltrating the USA from Mexico are not positively affecting the crime rate in the border neighborhoods.