Crime Analysis in the U.S. (CAUS) and Involving Factors

Brisilda Ndreka (brisilda.ndreka@uconn.edu)
Min Lin (min.2.lin@uconn.edu)
Wenxiao Zhou (wenxiao.zhou@uconn.edu)

May 18$^{\text{th}}$, 2021

Contents

Abstract

For the purpose of our investigation, crime data during the years 1995-2019 is studied. This project takes different types of crime according to Criminal Justice Information Services Division in consideration. Our goal is not only to define the historical "trend" of crime during these years in the US but also to reveal the most dangerous cities in America, based on the number of violent and property crime per 100,000 residents. The location where crimes or activities occur and the relationship of those places to one another as well as to other information is an important factor in the analysis of crime.

Under these ideas we will work to visualize the information the data contains. Making use of various types of graphs, we provide an explanatory analysis of the temporal and spatial patterns of different crimes in U.S.

Another objective of this work is to provide information in a "live" application as is R Shiny, which is easy to use without any knowledge of the R language. We integrate HTML with R Shiny functions to make the application presentable. We reveal the underlying crime structure and potential risks associated with different areas, states, and regions.

Keywords: Crime; Location; R-shiny; Graph; Tendency; Estimation; Risk; Time.

1 Introduction

Based on the 2019 report from the Criminal Justice Information Services Division FBI website, an estimated 1,203,808 violent crimes occurred nationwide, a decrease of 0.5 percent from the 2018 estimate. Reports from FBI's Uniform Crime Reporting (UCR) Program have been providing this information during different years, but in a messy way. Under these statistics is where our analysis is based on. Other category related to crime, is the property crime and all specific subcategories that are included here are taken in consideration to complete the "puzzle" of the "trend" crime in U.S.

In this study we will process several nature of crime . In section 2, we provide a description of the data we used in the project. In section 3, we proceed with the analysis of the data. In this step of our work, we use graphical techniques in visualizing crime patterns data with three goals in mind:

- Analysing overall crime rate per state. Defining those states in America that are at the top and bottom of the "pyramid of crime".
- Historical focus: providing analysis to follow how the crime "trend" has changed through the years in the U.S. While working under spaghetti plots and bubble charts we will visualize these changes, determining the states where dramatic changes happen. Here the investigation has in focus the crime rate pattern for certain areas within states.
- Statistical analysis under a geographical aspect. Using map charts we categorize the states in U.S. with more potential "risk" in terms of overall crime level. The "Neighborhood effect" is another issue we point out here, how the crime risk impacted by nearby areas.

2 Description of Data

The data sets can be accessed by the https://ucr.fbi.gov/crime-in-the-u.s website. This compiled data set is pooled using the separated file linked by time and place to inspire crime trend. According to [1], we transform the original datasets into the tidy form, which is required by most of statistical models and the powerful ggplot2 package in R language. There are three interrelated rules which make a dataset tidy:

1. Each variable must have its own column.
2. Each observation must have its own row.
3. Each value must have its own cell.

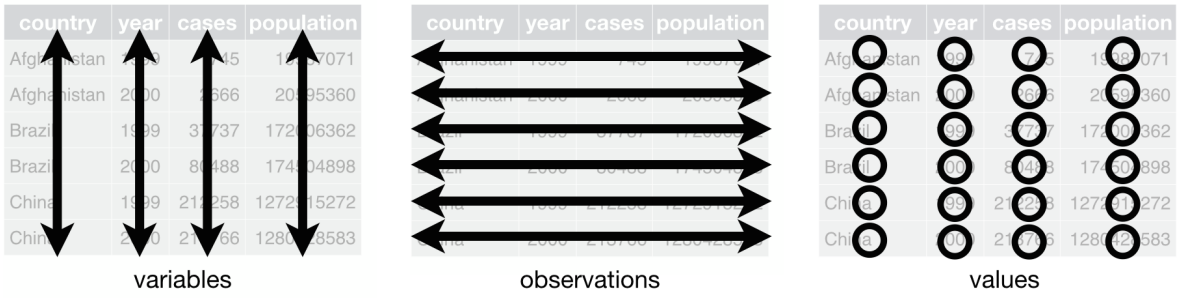

Figure 1: Three rules of the tidy data. Source: https://r4ds.had.co.nz/tidy-data.html

This data consists of observation characteristic variables, including State, Area, and Year, along with the number of several crime categories between 1995 and 2019. The variables to be used in the analysis are as follows:

Table 1: ID variables. A combination of {State, Area, Year} uniquely determines an observation, if not missing. There are 3597 observations.

Variable name	Type	Description
State	Categorical (52 levels)	50 U.S. states, District of Columbia, and Puerto Rico
Area	Categorical (3 levels)	Metropolitan Statistical Area, Cities outside metropolitan areas, and Nonmetropolitan counties
Year	Categorical (25 levels)	1995 - 2019

Table 2: Other variables for {State, Area, Year} observations. All are nonnegative count.

Variable name	Description
Population	Estimated population
ViolentCrime	Estimated number of violent crimes, comprising four type of crimes: murder and nonnegligent manslaughter, rape, robbery, and aggravated assault
Murder	Estimated number of murder and nonnegligent manslaughter crimes
Rape	Estimated number of rape crimes. Beginning from 2013, the it is estimated using the revised Uniform Crime Reporting (UCR) definition of rape.
Robbery	Estimated number of robbery crimes
AggravatedAssault	Estimated number of aggravated assault crimes
PropertyCrime	Estimated number of violent crimes, comprising three type of crimes: burglary, larceny theft, and motor vehicle theft
Burglary	Estimated number of burglary crimes
LarcenyTheft	Estimated number of larceny-theft crimes
MotorTheft	Estimated number of motor vehicle theft

Table 3: Variables used for the bubble chart (see section 3.4)

Variable name	Type	Description
Region	Categorical (7 levels)	Seven regions of the U.S, including New England, Mid-Atlantic, Southeast, Midwest, The Rocky Mountains, Southwest, and Pacific Coastal. According to an online poll where more than 2/3 participants (57 out of 83) votes for that District of Columbia belongs to the Mid-Atlantic region, we adopt the idea. But Puerto Rico is not classified into any region.
GDP_real_per_capital	Positive real number	State level real GDP divided by the estimated population

3 Exploratory Data Analysis

3.1 Crime rate in the United States in 2019, by state per 100,000 inhabitants

Crimes are "alive" in the United States. As a nation, the U.S is categorized to have relatively high crime rates; however, the rates have decreased significantly over the past 25 years. The American government categorizes crimes in two ways; violent crime or a property crime. Working under the latest data in our analysis, 2019 year, we focus on figuring out the states in the U.S. that are considered as the "safest" and those "riskest".

The dataset contains crime counts, which are not directly comparable among states, because of the huge population variability. Hence, we will focus on the crime rate per 100,000 inhabitants throughout this section, unless otherwise mentioned. The following two figures, Figure 2 and Figure 3, display how the two types of crime rates of each state in the America. The states with low levels of crime rate are located at the bottom of the "crime pyramid" and those with high levels of crime at the top. We will examine the potential causes in the later parts.

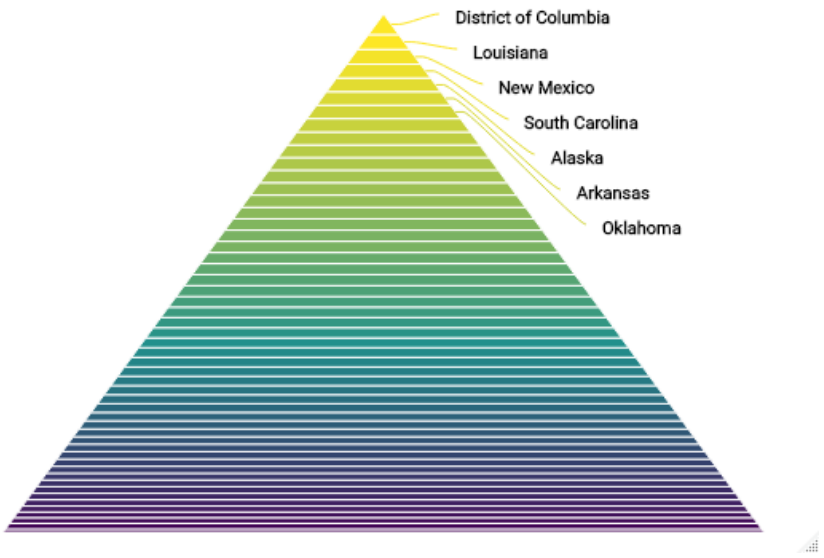

Figure 2: Property Crime 2019

Saying in the property crime rate for 2019 we have:

- **States with the Lowest Property Crime Rates:**
  - **Puerto Rico** has the lowest crime rate of 702.70 incidents per 100,000 people.
  - **Massachusetts** has the second-lowest crime rate in the United States for 2019 of 1,179.80 incidents per 100,000.
  - **New Hampshire** has a crime rate of 1,209.20 per 100,000 people, making it the state with the third-lowest crime rate.
  - **Idaho** has the fourth-lowest crime rate in the U.S. at 1,219.50 incidents per 100,000.
  - **Maine's** crime rate of 1,245.60 is the fifth-lowest nationwide.
- **States with the Highest Crime Rates:**
  - **The District of Columbia** has the highest crime rate in the United States of 4,367.1 incidents per 100,000 people.
  - **Louisiana** has the second-highest crime rate of 3,162.00 incidents per 100,000 residents.
  - **New Mexico** has the third-highest crime rate in the United States of 3,112.70 crimes per 100,000 people.
  - **South Carolina's** crime rate of 2,940.30 per 100,000 is the fourth-highest in the country.
  - **Alaska's** crime rate of 2,910.80 is the fifth-highest nationwide.

The order changes when moving to the violent crime analysis. We see District of Columbia categorized with the highest crime rate in the U.S followed by Alaska, New Mexico, Tennessee etc. In our work, in the R Shiny application, we provide information that goes even deeper in analyzing this pattern for each category of property and violent crime during 1995-2019 years.

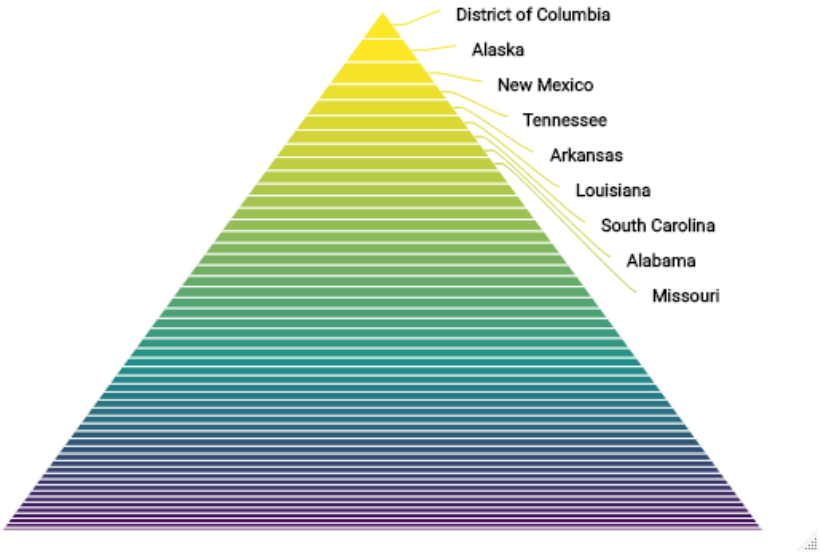

Figure 3: Violent Crime 2019

3.2 Historical structure of crime-rate (1995-2019)

The purpose of this subsection is to analyze the historical crime rates of states for each sub-types in two categories: burglary, larceny theft, motor theft under the property crime, and murder, rape, robbery, aggravated assault under the violent crime category. We analyze the trend of crime rates by the spaghetti plots, which also shows a special historical variations of rape from 1995 to 2019 especially after 2012. Here, we discuss on the historical structure of the crime rates.

In Figure 4, the historical evolution of crime rates is characterized by an almost constant decrease and almost a similar change for all states and crime categories during years. However, we notice two states have unusual crime trends, Alaska and District of Columbia.

There is a significantly higher rates of rape reports in Alaska than other states (except South Dakota in 2011 according to Table 4 in Appendix). District of Columbia shows dominance in crime levels in the U.S. Referring to the graph below and Table 4(Appendix), this dominance is not shown in burglary, rape and sometimes larceny-theft (2002-2007), but still it is in high level of crime.

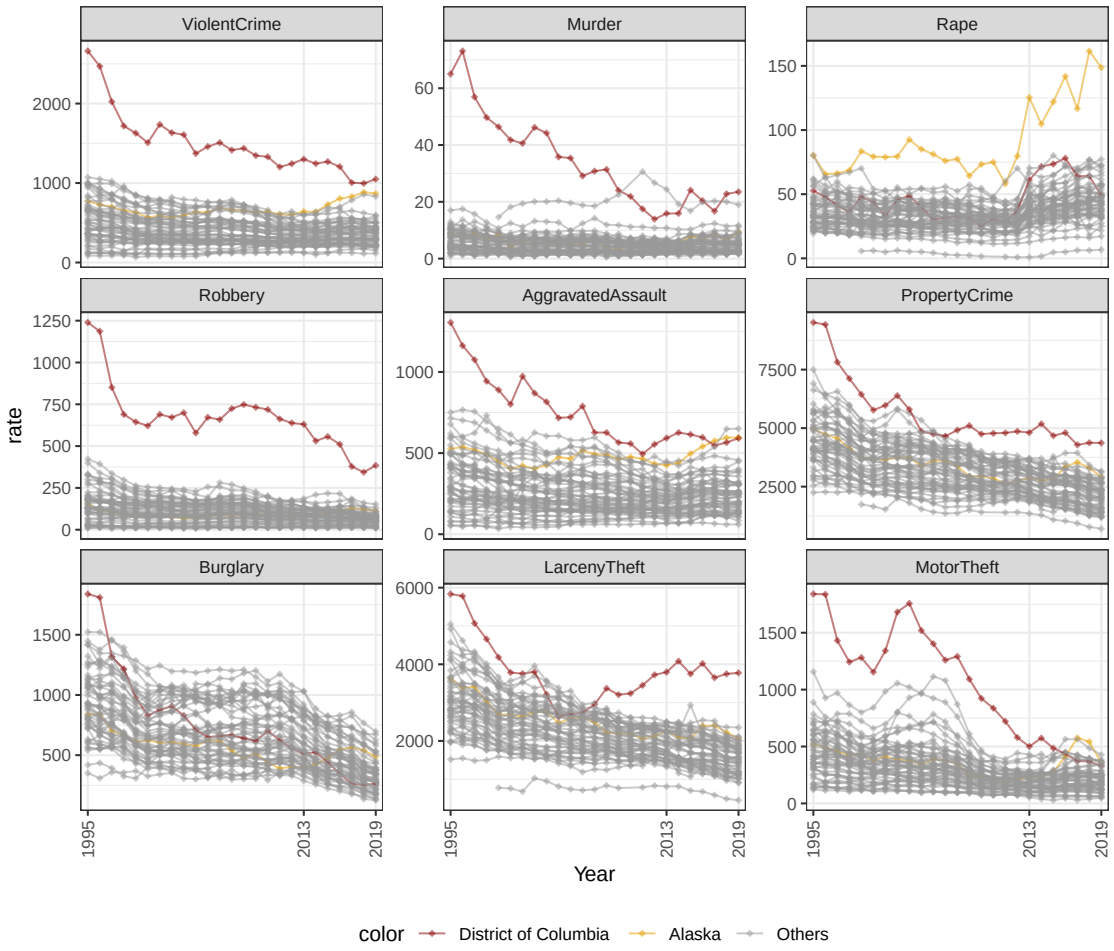

Figure 4: Spaghetti plot for crime rates per 100k inhabitants for each state from 1995 to 2019

To refine the analysis for crime rate, we process the spaghetti plot now including the area factor. Within states, the comparison among areas is best carried out by crime counts rather than rates. Figure 5 shows the rape counts for three different areas: metropolitan statistical area, cities outside metropolitan area, and non-metropolitan area for all states in the U.S. We notice approximately flat curves from 1995-2012 and an increasing trend of crimes start in 2013 up to 2019. One point to emphasize here: that can cause the changes in the trend seen in Figure 5, is the revision of the definition of rape. After 2012, FBI UCR Program started using the revised definition to collect and estimate the rape data.

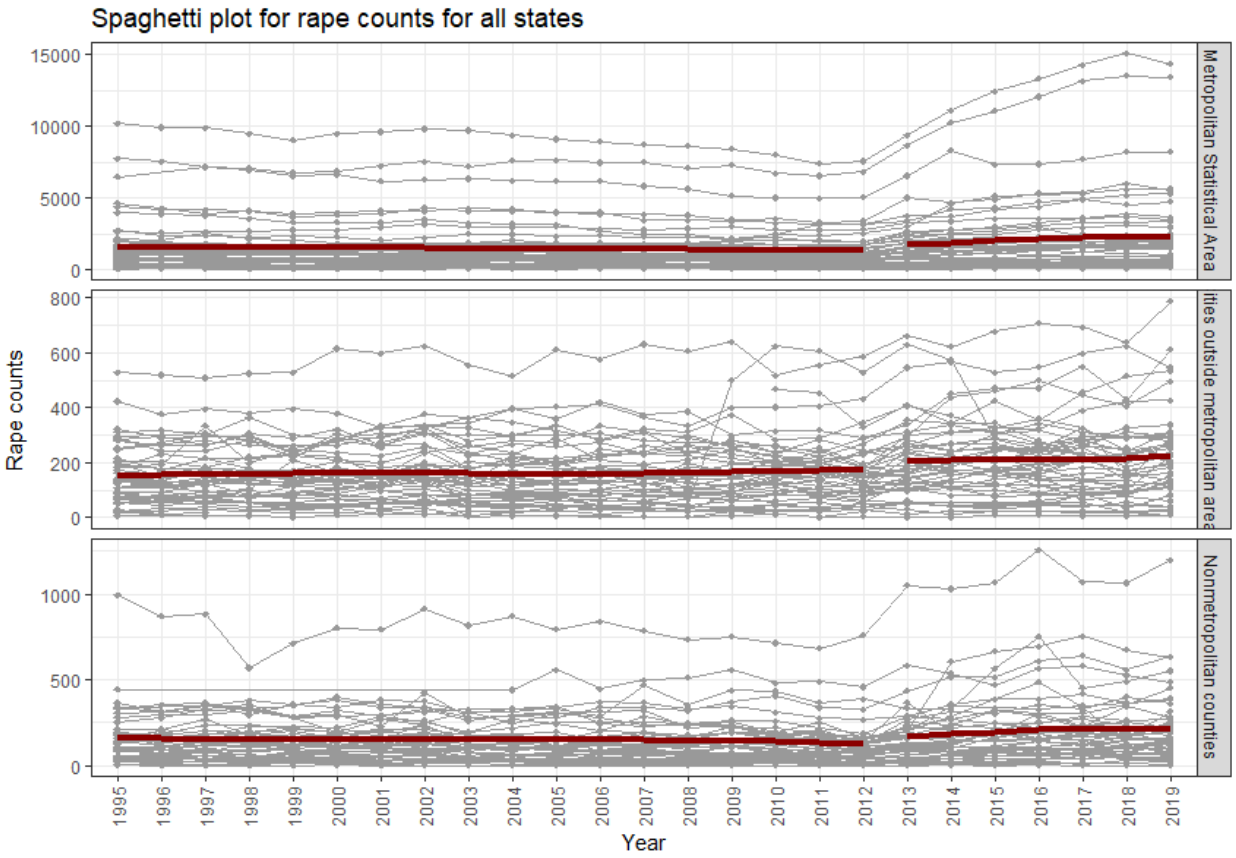

Figure 5: Rape reports from 1995 to 2019 for all states (50 U.S. states, District of Columbia, and Puerto Rico). The dark-red curve is fitted by loess method.

3.3 Crime heat-map

The following chart is an outline of the history of crime mapping pattern on 2019. In this part of the report, the work will continue towards a better understanding of global and regional crime patterns in U.S within two main categories, property and violent crime. An interesting approach is to observe any neighborhood impacts between states close to each other.

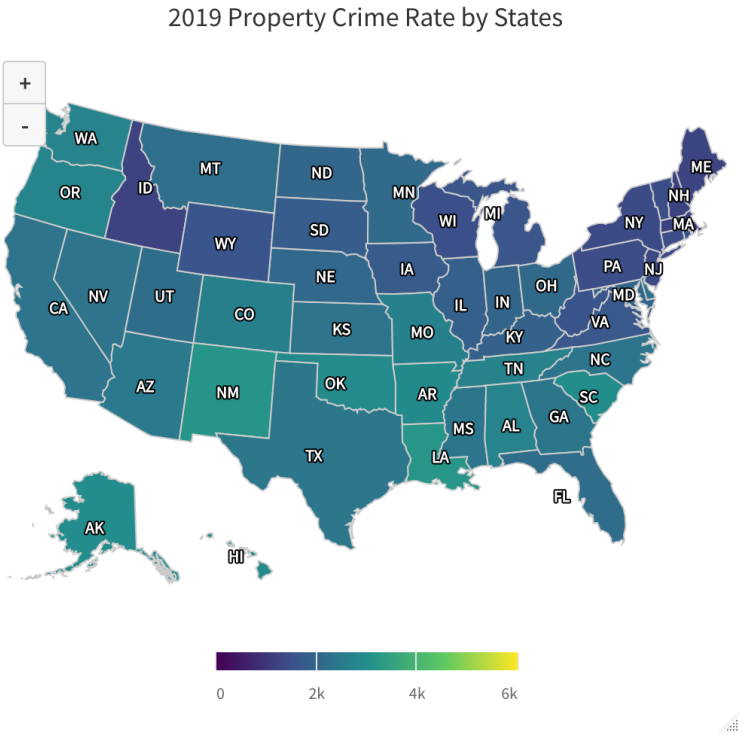

Figure 6: Property crime map

Considering two crime maps, we conclude the prevalence of property crime vs. violent crime in U.S. For a regional analysis of the crime distribution we have the Northeast part of America which can be considered as the "safest" zone followed by Midwest, leaving behind the Southern and Western parts. Referring to the analysis done in part one this was expected, recalling that states as Alaska, New Mexico, Tennessee etc, crime rate is at high level. In the two maps, the influence of the states close to each other can be seen. We do not have dramatic changes for states close to each other.

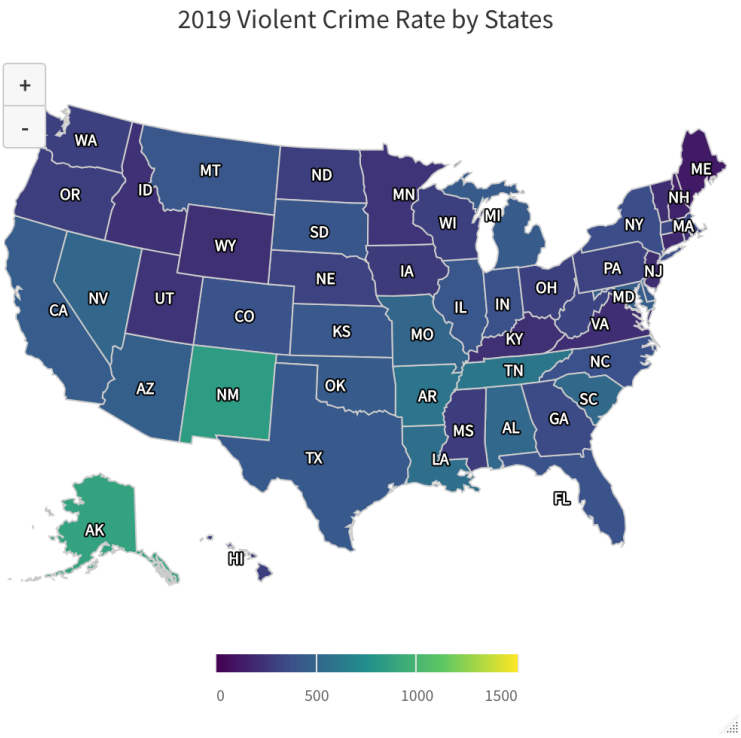

Figure 7: Violent crime map

The "crime pattern" becomes more interesting if this analysis is processed through each sub-type of property and violent crime that we have. This is an analysis that is done in R Shiny application, for all years.

3.4 Relationship detection: real GDP and region information

In this subsection, we introduce the animation bubble chart on the 'Relationship Detection' page in our R Shiny app. The 50 U.S. states can be classified into one of seven regions, and we classify District of Columbia into Mid-Atlantic (refer to Table 3). In the following chart, we have seven facets, each corresponding to one specific region. Each bubble represents one state (50 U.S. states along with District of Columbia). The size of the bubble reflects the amount of per capital real GDP. A larger bubble means greater GDP. The horizontal and vertical axes stand for, respectively, the Property crime score (0-100) and Violent crime score (0-100). The higher the score is, the higher the risk is.

The score is attributed by the following method. Firstly, we compute the crime rate per 100k inhabitants for both violent and property crimes. Next, we apply the Box-Cox transformation to the two rates across all states and years, in order to reduce the slowness for better visualization. Last, since the two Box-Cox transformed rates are of different scales and ranges, we further normalize them into the range of 0-100.

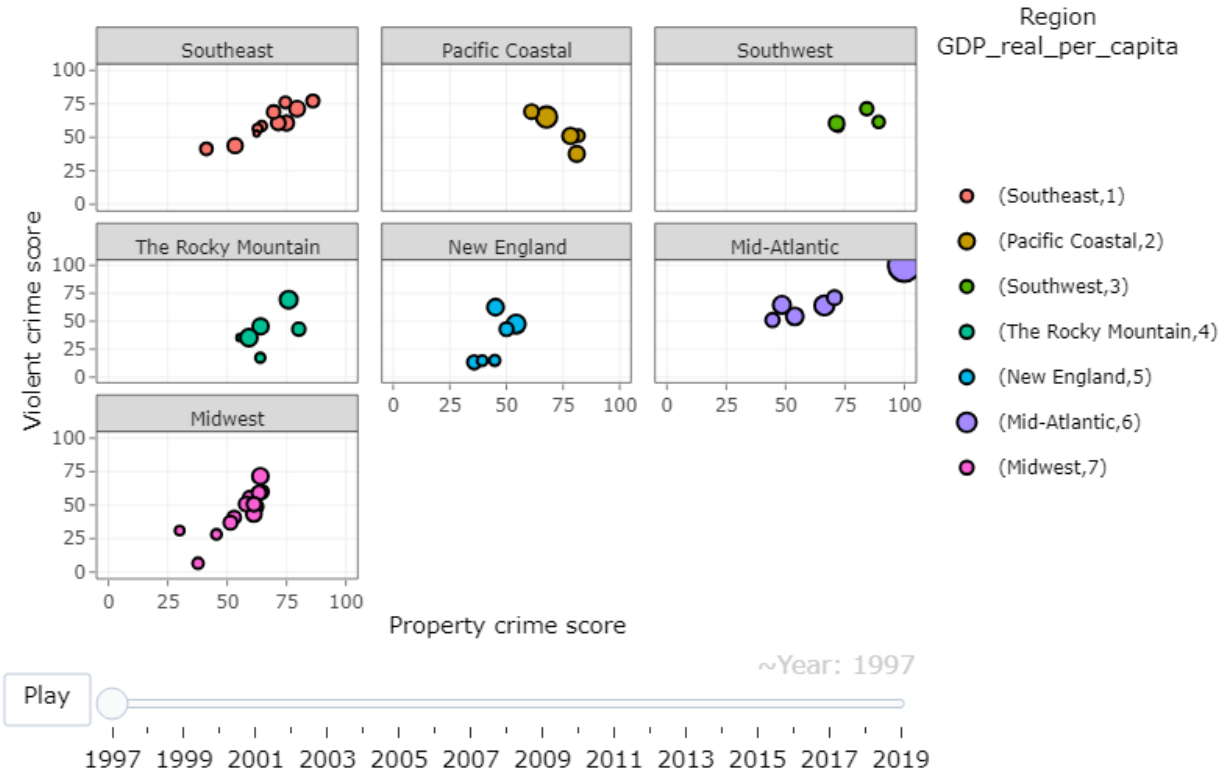

Figure 8: Bubble chart 1997

Clearly, the crime risk scores show a strong association within each region, which provides

evidence for geographical relationship of crimes. Furthermore, the largest bubble in Figure 8 (District of Columbia) dominate all the others in both scores: two 100 risk. Consequently, the per capital real GDP can also be an informative explanatory variable for the crime risk.

The following shows the 2019 bubble chart. As expected, both violent and property crime risks are lower than the 1997 counterpart. The GDP and regional relationship are still noticeable. For more information, please refer to our R Shiny app.

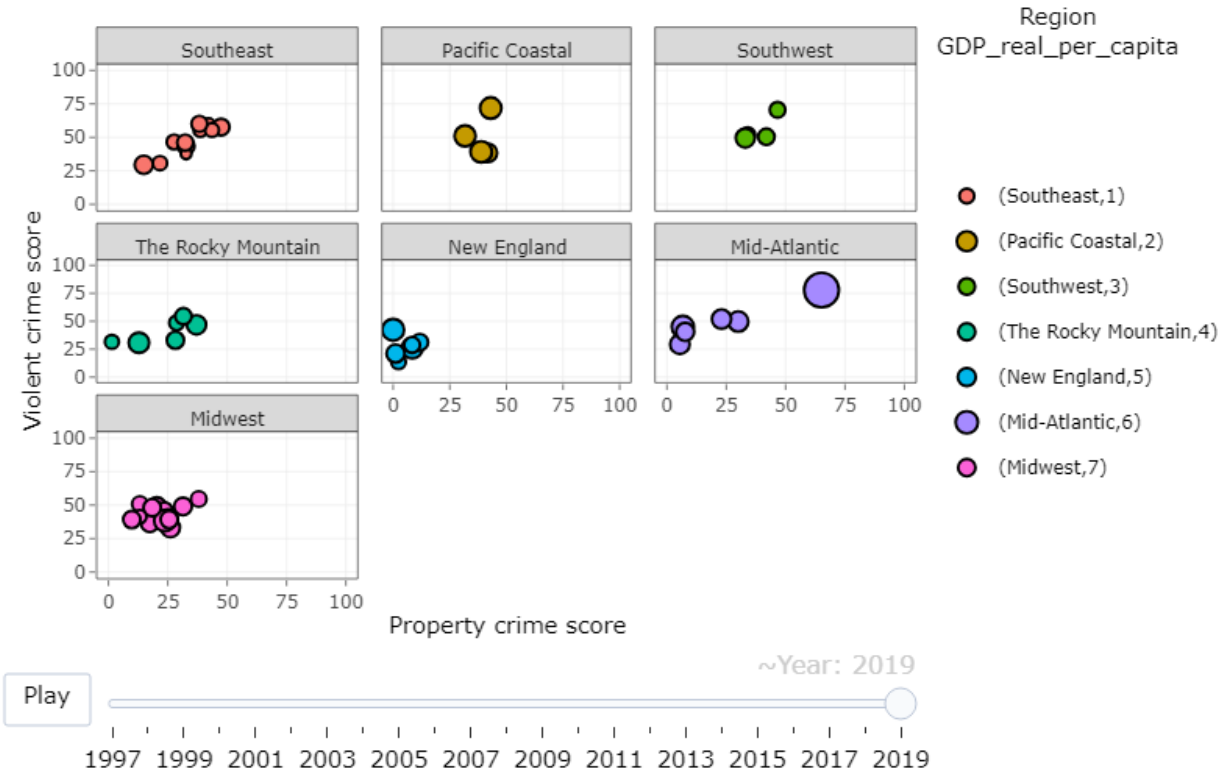

Figure 9: Bubble chart 2019

4 Summary

Through this study we worked to elaborate the question: How much is the seriousness of crimes in the U.S? Visualizing the information from a general point of view gradually up to a detailed one, we provide the evidence for risk of different areas, states, and regions in the U.S. The goal is to provide information in an easier form that could be absorbed by everyone. What we can answer for sure is that in the America, the overall crime trend is going down through the years after adjusting for several factors: geographical and economic ones. We can spot a strong correlation within each region, and there is also evidence for the association between per capita real GDP and crime rates.

As expected, the property crime is more common than the violent crime. So far the most common from of property crime in 2019 was theft, followed by burglary and motor vehicle theft. Among violent crimes, aggravated assault was the most common offense, followed by robbery, rape, and murder. The seriousness of crimes are of the inverse proportion of the number of crimes.

How the crime rates in the U.S. differ geographically, is another question that study takes care of. There are significant differences in violent and property crime rates from state to state. Even in cities withing the same states crime rates can vary in a considerable level. In 2019, there were more than 1000 violent crimes per 100,000 residents in District of Columbia, compared with fewer than 200 per 100,000 people in Maine and New Hampshire, according to the FBI. A refined inferential study may be carried out based on our explanatory analysis.

References

[1] Hadley Wickham. Tidy Data. *Journal of Statistical Software*, 59(10), 2014.

[2] Haim Bar. Lecture Notes.

Appendix

Table 4: States with maximum real GDP per capita and maximum crime rates from 1995 to 2019. We use "State" to refer to 50 states, D.C. (District of Columbia), and Puerto Rico. GDP data are not available for 1995 and 1996, and hence the values are "NA" (not applicable).

Year	GDP_real_per_capita	Murder	Rape	Robbery	AggravatedAssault	Burglary	LarcenyTheft	MotorTheft
1995	*NA*	D.C.	Alaska	D.C.	D.C.	D.C.	D.C.	D.C.
1996	*NA*	D.C.	Alaska	D.C.	D.C.	D.C.	D.C.	D.C.
1997	D.C.	D.C.	Alaska	D.C.	D.C.	Florida	D.C.	D.C.
1998	D.C.	D.C.	Alaska	D.C.	D.C.	New Mexico	D.C.	D.C.
1999	D.C.	D.C.	Alaska	D.C.	D.C.	North Carolina	D.C.	D.C.
2000	D.C.	D.C.	Alaska	D.C.	D.C.	North Carolina	D.C.	D.C.
2001	D.C.	D.C.	Alaska	D.C.	D.C.	North Carolina	D.C.	D.C.
2002	D.C.	D.C.	Alaska	D.C.	D.C.	North Carolina	Hawaii	D.C.
2003	D.C.	D.C.	Alaska	D.C.	D.C.	North Carolina	Hawaii	D.C.
2004	D.C.	D.C.	Alaska	D.C.	D.C.	North Carolina	Arizona	D.C.
2005	D.C.	D.C.	Alaska	D.C.	D.C.	North Carolina	Hawaii	D.C.
2006	D.C.	D.C.	Alaska	D.C.	D.C.	North Carolina	Hawaii	D.C.
2007	D.C.	D.C.	Alaska	D.C.	D.C.	North Carolina	Hawaii	D.C.
2008	D.C.	D.C.	Alaska	D.C.	D.C.	North Carolina	D.C.	D.C.
2009	D.C.	D.C.	Alaska	D.C.	D.C.	Arkansas	D.C.	D.C.
2010	D.C.	D.C.	Alaska	D.C.	D.C.	Arkansas	D.C.	D.C.
2011	D.C.	Puerto Rico	South Dakota	D.C.	D.C.	Arkansas	D.C.	D.C.
2012	D.C.	Puerto Rico	Alaska	D.C.	D.C.	Arkansas	D.C.	D.C.
2013	D.C.	Puerto Rico	Alaska	D.C.	D.C.	Arkansas	D.C.	D.C.
2014	D.C.	Puerto Rico	Alaska	D.C.	D.C.	New Mexico	D.C.	D.C.
2015	D.C.	D.C.	Alaska	D.C.	D.C.	Mississippi	D.C.	D.C.
2016	D.C.	D.C.	Alaska	D.C.	D.C.	New Mexico	D.C.	New Mexico
2017	D.C.	Puerto Rico	Alaska	D.C.	Alaska	New Mexico	D.C.	Alaska
2018	D.C.	D.C.	Alaska	D.C.	New Mexico	New Mexico	D.C.	Alaska
2019	D.C.	D.C.	Alaska	D.C.	New Mexico	New Mexico	D.C.	New Mexico