Chicago Crimes Visualization

André Brandão
Department of Computer Science
Faculty of Sciences of the University of Porto
Porto, Portugal
up201908477@fc.up.pt

Vítor Bernardes
Department of Computer Science
Faculty of Sciences of the University of Porto
Porto, Portugal
up201908581@fc.up.pt

***Abstract*—Data visualization allows faster communication, to better find patterns that would be impossible to see just by looking at the numbers and also provides the possibility to interact with the data itself.**

In this report we use data visualization techniques to answer 4 questions about crimes in Chicago.

***Index Terms*—Data Visualization, Chicago Crimes, Demographics**

I. Introduction

A data set containing the crimes committed in Chicago from 2001 to 2017 was given us for analysis. In this report we present the process of transforming this and other data sets from external sources into the visualizations that will answer these four questions:

- Does education influence the number of crimes?
- What type of crimes happens more often near schools? And is there a difference between public and private schools?
- What types of socioeconomic factors influence the amount and types of crimes? Study the special case of the 2008 crash.
- Study the correlation between the gun laws in Chicago and the incidence of crimes.

II. The Data Set

The data set contains reported incidents of crime and data about each incident, mostly related to location, date, and type of crime, as well as if an arrest was made.

The initial step in our investigation was performing an exploratory data analysis (EDA) on the Chicago data set so as to discover patterns, spot anomalies, and test potential hypotheses about the data. The EDA also helps us make sense of the data, understand how it is structured and identify potential relationships between variables of interest. We used Pandas in Jupyter notebooks for performing the EDA.

During this step, a lot of cleaning was also necessary so as to get the data in order for analysis. We discovered the data from 2001 to 2004 was particularly problematic, with several community data missing in the period of 2001–2002. In addition to that, there was an unusual spike in the number of crimes in 2001 and a sudden drop during the second quarter of 2004, which led us to obtain the data for the period of 2001–2004 from the Chicago data portal. There was also a large number of duplicate entries from 2006 to 2011, which were easily handled by the Pandas library. In addition to that, some other minor cleaning was performed, mainly related to removing missing data.

With the cleaned data set, we came up with straightforward questions that could be easily answered with some plots and would help us make sense of the data we would be dealing with. Those included the evolution of the number of crimes through the years, the number of crimes per type of location, per hour of day, and per community, among others. Figure 1 shows a couple of examples from the EDA. The plot with the number of crimes per community shows that community 25 (Austin) is isolated with a much larger number of crimes than any other community, a fact that would help our investigation on section III-C.

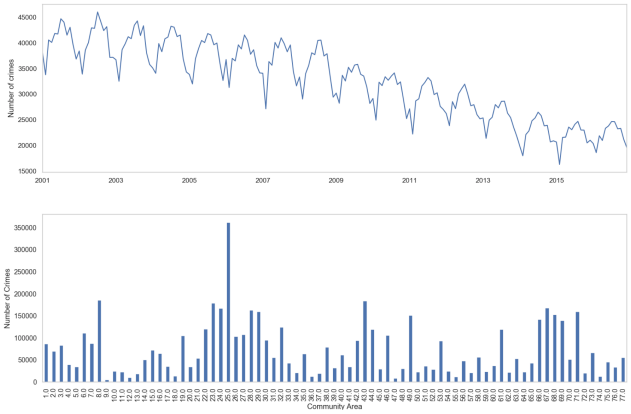

Fig. 1. Examples of plots created during the exploratory data analysis.

III. Questions

A. Does education influence the number of crimes?

With this question we intend to know if the level of education of people in a community correlates to the amount of crimes and if communities with a higher percentage of low education levels have more crimes. Our hypothesis is that areas with higher percentage of people with high level education present less crimes.

To answer this question we need information about education levels per community and number of crimes per community as well the population for each community. For the

education levels and population number per community we obtained the data from [1], from the educational attainment table that contains the population number divided by tracts as well as the education levels (Less than 9^{th} grade, 9^{th} to 12^{th} grade, High school graduate, Some college (no degree), Associate's degree, Bachelor's degree, Graduate or professional degree) from 2010 to 2017. Using a Python script and a data set from [2] that associates each tract to the corresponding community code we grouped the data from US census in communities by adding the respective values of each tract. The information about number of crimes per community was obtained from the Chicago crimes data set.

To visualize how education influences the number of crimes we created a demographic map that maps the color of each community to the education level and number of crimes per capita, however in order to compare the two we decided to plot both side by side (figure 2).

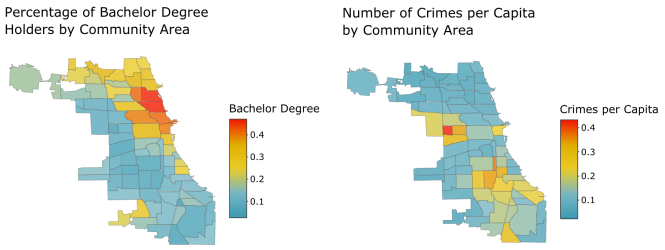

Fig. 2. Demographics maps for education level and crimes per capita by community.

This is not sufficient since we can only see one education level, so we decided to allow the user to change between education levels (9^{th} or Less, High School, Bachelor Degree). This data might change along the years, therefore we divided the data in years, from 2010 to 2017, and added a slider so the user can choose which year they want to see (figure 3).

Fig. 3. The user can change the education level and year to visualize.

Now it's possible to see a choropleth map of education level by community and year and compare it to the number of crimes per capita in that given year. We also decided to add a tooltip to the maps so the user can see the actual value of each community (figure 4).

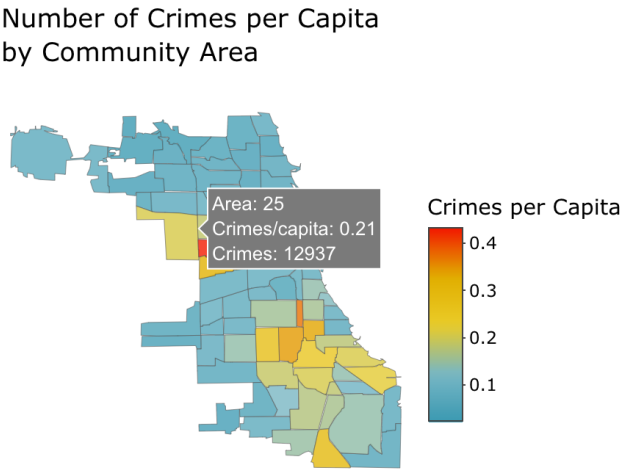

Fig. 4. The user can hover the communities to get the corresponding information about them.

Looking at the plots generated we can see that in fact the top right corner of Chicago has a very high percentage of bachelor degree holders and is one of the communities with less crime. However we cannot conclude that the communities with higher percentage of lower education contain more crimes. After seeing that the upper right corner of Chicago, mainly Lake View Area has a high percentage of bachelor degree holders and low number of crimes we decided to investigate it. In the visualization made to answer the question from section III-C we will see that this area is also one of the areas with the highest income per capita. After searching on the internet about Lake View, trying to understand why its statistics were so good, the only distinguishable characteristics found about this area were:

- Lake Area contains one of the biggest LGBT+ communities in the nation.
- More than 85% of the population is white (less than 50% of the population in Chicago is white).
- Supports the Democratic party – more than 80% of the population voted for Hillary and more than 70% for Barack Obama.

However we cannot conclude that these characteristics are influencing the low rate crime and high income since we don't have a way to compare it with other communities.

B. What type of crimes happens more often near schools? And is there a difference between public and private schools?

With this question we want to understand if there is a high number of crimes near schools and if the places close to private schools have different types of crimes than the ones close to public schools. Our hypothesis is that the most common crime near schools is narcotraffic and it happens more often near public schools.

To answer this question we need the public and private schools geographical locations, the geographical locations of the crimes as well the types of crimes.

We obtained the schools' address data at [3]. Since there were a lot of schools in Chicago we decided to keep only the public and private high schools, then we created a script in Python to obtain the geographical coordinates from the addresses using the google geocode API (https://developers.google.com/maps/documentation/geocoding/intro). All the information about the crimes is available in the Chicago crimes data set.

To understand if the schools influence the number of crimes a heatmap of the crimes was made on top of the Chicago map with circles representing the schools with a 1km radius (figure 5). The **gmplot** library was used [4] with a few changes to achieve the intended result.

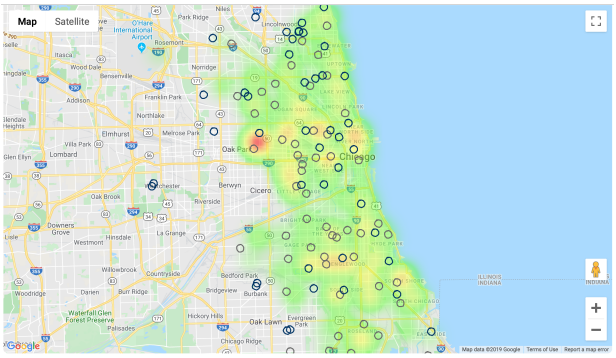

Fig. 5. Heatmap of crimes and schools representation.

Looking at this picture we can see that there are schools that have more crimes nearby than others, so we cannot confirm that the schools influence the number of crimes. In order to understand if there are more crimes near public schools or near private schools a public / private schools crime ratio is also presented. We also wanted to check if there was a difference between the years, so instead of plotting everything at once we gave the user a slider so we could choose the year to show the respective data.

From figure 5 it's not possible to understand what types of crimes happen more often near schools. In order to answer this question we add two bar plots with the 5 most common crimes for public and private schools, where these plots also change according to the year chosen by the user using the slider (figure 6).

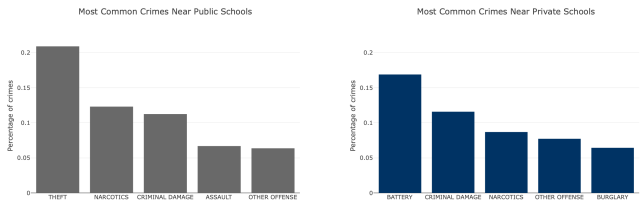

Fig. 6. Barplot with the five most common crimes near public and private schools.

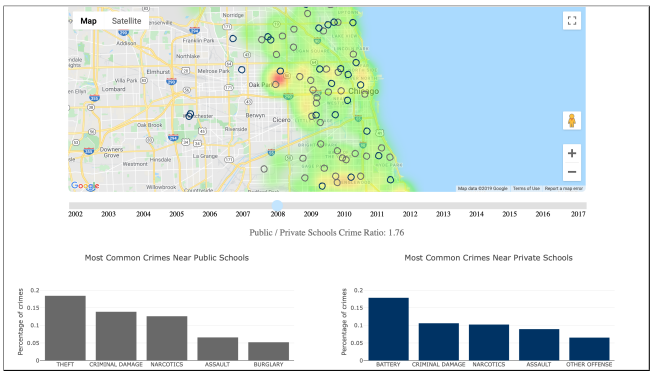

Fig. 7. Full visualization.

Looking at the full visualization (figure 7) and the public / private schools crime ratio it's possible to conclude that there are more crimes near public schools than near private schools. The most common crimes that happen near schools are theft, battery, criminal damage and narcotics, having narcotics related crimes happening more frequently near public schools.

C. What types of socioeconomic factors influence the amount and types of crimes? Study the special case of the 2008 crash.

For this question, we try to understand how socioeconomic indicators are related to the amount and types of crimes in Chicago. We explore selected indicators in different regions and look for a potential correlation between the indicators and crime activity in that region. Our hypothesis is that poor socioeconomic indicators are related to an increased crime activity. Also, the types of crime that will increase the most in these situations are theft and related crimes.

In order to investigate these potential relationships, we needed to associate our crime data for Chicago with socioeconomic data from external sources. We obtained data released by the Illinois Department of Public Health (IDPH) and the U.S. Census Bureau [3] for several indicators spanning the period of 2007–2011. The indicators were calculated by the Chicago Department of Public Health (CDPH) using U.S. Census Bureau 2000 census, 2010 census, and 2006-2010 American Community Survey, and are presented on a community-level.

The aim for this visualization was to convey the potential relationships to the reader as easily as possible. So as not to overwhelm the reader with too many relationships to analyze, we decided to work with a small subset of available indicators. We chose indicators that are easily understood and have a direct economic significance for citizens: unemployment rate, per capita income and poverty level.

All initial data handling was carried out using the Pandas data analysis library for Python on Jupyter notebooks. Since the economic indicators we worked with represent average yearly estimates for 2007–2011, we aggregated our crime data accordingly. In order to investigate the relationship between the number of crimes and each of those indicators, we com-

puted the average number of crimes committed per year for that same period.

The most direct way to visually investigate the potential correlation between two measures is the scatter plot, so we created scatter plots for the three chosen indicators against the number of crimes computed. The initial analysis showed a moderate positive correlation between the amount of crimes and both unemployment rate and poverty level. The correlation between per capita income and the number of crimes was negative, as expected, but smaller.

However, one outlier present in the data led the next step in the investigation. We found that the community of Austin had more than five times the average number of crimes per year during that period, which is made more understandable due to the fact that Austin is the second largest Chicago community area by population.

So we computed the number of crimes per capita and remade the scatter plots. This time, the correlations were much clearer, with a particularly stronger positive correlation for unemployment and poverty level.

This finding was the basis for the visualization we designed. We opted for a mix of explanatory and exploratory visualization, in that there was a specific finding we wanted to convey through the explanatory concept of the visualization, but we also wanted to leave room for the readers to explore different parameters and arrive at their own conclusions.

One exploratory feature was the inclusion of color to the final scatter plots to display an additional measure. We chose to encode the proportion of certain characteristics as the color of points representing each community. In light of the theme of the visualization, we decided to offer three options of proportions the user could choose from: arrests, violent crimes, and crimes motivated by financial gain.

Other features could have been incorporated into the visualization, such as using a bubble chart to display other measures encoded as the bubble size, but since our design principle was to keep it simple and not overwhelm the reader, we opted to only add color.

Finally, since we are dealing with data related to different communities in Chicago, we decided to illustrate the geographical aspect with a map of communities beside the scatter plots. Again, we decided not to encode any new information in the map and keep it only so the user could relate the points in the scatter plot to the community location by clicking them.

The end result, shown in figure 8, brings together all of these aspects to present the user with the information that crime activity is highly correlated to some socioeconomic indicators.

Since the data set we are analyzing spans the period of the 2007–2008 financial crisis, we decided to study that special case and create a visualization that would showcase an interesting finding about it. Inspired by the relationship between the socioeconomic indicators and crime, uncovered on the previous visualization, we decided to relate those two dimensions.

For this visualization we wanted to observe the evolution of the relationship between the unemployment rate and the

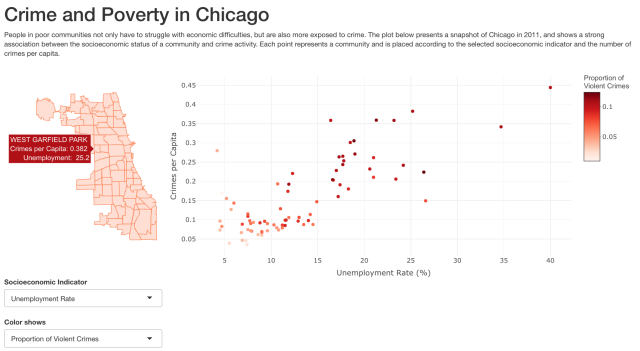

Fig. 8. Final visualization displaying the correlation between crime per capita the socioeconomic indicators.

number of certain types of crimes committed each year. The previous plot leads us to consider there is a positive correlation between those two measures. We chose a subset of crimes primarily related to financial gain, and that a study [5] conducted by the United Nations Office on Drugs and Crime has shown that are likely to increase during periods of economic turmoil. These mainly include theft, robbery and burglary. The unemployment rate data was retrieved from the Federal Reserve Bank of St. Louis, using data from the U.S. Bureau of Labor Statistics [6].

Although the traditional approach for this type of visualization would be to plot a line graph over time, this solution would possibly require using two Y-axis scales with vastly different orders of magnitude, which we wanted to avoid. We looked for an alternative and more integrated way to display the evolution in both these measures over time, so we opted to use a connected scatter plot, with the unemployment rate on the X-axis and the number of crimes per year on the Y-axis. On this type of plot, the time dimension is displayed on the segments connecting the points, which represent the values for both measures for each year.

Several steps were taken to aid the reader in interpreting the visualization. We displayed the year as a label for each point and also added arrows to the segments in order to make navigating through the plot easier. An annotation was added relating important findings. Also, as an additional aid to comprehension, color was carefully used to direct the reader's attention to a conclusion highlighted in the annotation and to connect elements of the plot, shown in text and segments, that are related to the crisis.

The finalized plot displayed in 9 clearly shows that the number of economically-motivated crimes increased during the peak of the financial crisis, and that the downward trend in unemployment rate was also reverted. However, while unemployment more than doubled from 2007 to 2010, the number of crimes decreased year after year starting from 2008, a trend that only showed signs of potentially reverting on 2016. Therefore, while the previous plot showed a high correlation between regions with high unemployment and crime activity, and this plot shows an increase in the number of crimes during

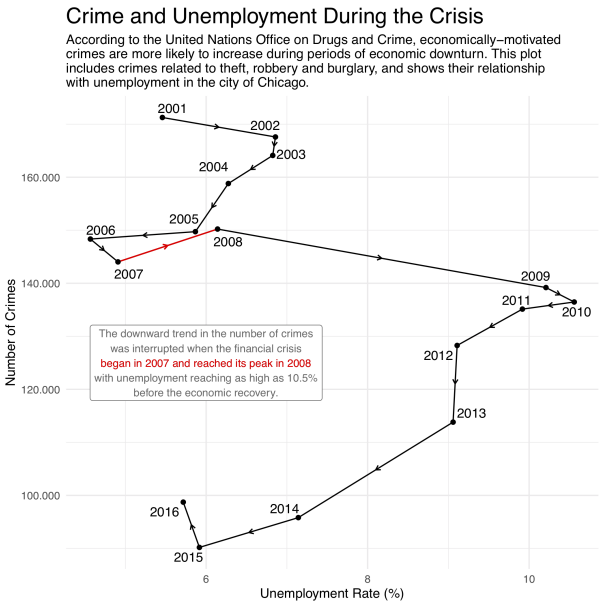

Fig. 9. Connected scatter plot displaying the relationship between unemployment and theft, robbery, and burglary.

the peak of the crisis, we do not observe the same clear correlation between periods of high unemployment and high crime activity in the city of Chicago.

D. Study the correlation between the gun laws in Chicago and the incidence of crimes.

This question is related to the effectiveness of gun restriction laws in reducing crime, and if we can observe their effect in Chicago.

As discussed, our data set contains data from 2001 to 2017, so we studied the changes in Chicago laws during that period regarding gun control. The state of Illinois, and particularly the city of Chicago, have historically had strict regulations related to the sale, possession, and use of firearms. Since the late 1980s, several Illinois municipalities had banned the possession of handguns. Chicago required the registration of all firearms but did not allow handguns to be registered, which effectively outlawed their possession. In our investigation, we uncovered two main points in time when there were significant changes in gun control law. The first was on June 2010, when the handgun ban was reverted and handguns started to be allowed again. The second one was in 2013, when concealed carry was permitted and the requirement of registration of firearms and obtaining a city-issued firearm owners' permit was repealed.

As was done on the previous investigations, we used a subset of crimes, directly related to firearm use in this case. We used the crime type categories established in the Chicago Police Department CLEARMAP application [7] and extracted the codes for all crime types related to handguns or firearms, which we used with our original data set to extract only those types. We then proceeded to aggregate the number of crimes over time at different frequencies, both monthly and yearly. Since the monthly data displayed a lot of variation, with a clear seasonality, we opted to work with the yearly aggregated data, which was smoother and suppressed the seasonality which was not relevant to this specific investigation.

We approached this investigation in a more traditional way, using line graphs to plot the evolution of the number of crimes over time, and comparing the periods with stricter and looser gun restriction laws. A clear downward trend in the total number of crimes per year can be seen, something which can not be observed in the number of firearm-related crimes, which present a lower variation. With that in mind, we decided it would be more interesting to investigate the proportion of firearm-related crimes in relation to the overall number of crimes committed.

We explored the idea of using a stacked area chart, so we could observe the evolution over time of the number of firearm-related crimes in direct comparison with the overall number of crimes. However, since the proportion of firearm-related crimes is very small (from around 2% to 6%), the resulting plot was dominated by the area of the other types of crime, making it hard to see any meaningful information regarding firearm-related crimes.

Therefore we decided to stick with the more traditional approach and create a line plot depicting the variation in the proportion of firearm-related crimes. We had a few options as to how to indicate the changes in gun control laws, such as annotations, boxes and arrows. We decided to change the background color in the periods after the gun control laws were made less strict, increasing the color intensity when further changes were made which made the laws even less strict. In order to provide the reader with context, we used annotations describing the law changes.

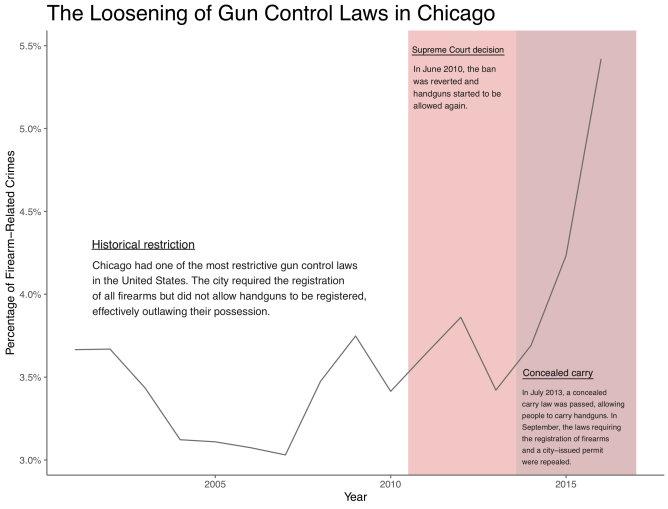

Fig. 10. Plot showing the evolution of firearm-related crimes in Chicago.

The final plot (figure 10) shows a particularly sharp increase in the number of firearm-related crimes after the second law change was made in July 2013. The period after the first law change, from 2010 to 2013, does not display a significantly

different variation from the past. In general, the plot does seem to indicate that when gun control was loosened, the number of firearm-related crimes spiked. However, it should be noted that the relationship between gun control and crime is a complex one, with numerous factors, and we cannot draw strong conclusions from a graph that only takes into account a small part of that process.

IV. Conclusion

With this work we were able to conclude that data visualization indeed helps to draw conclusions from the data allowing the users, even with low experience in data analysis, to observe patterns and retrieve information from them.

We were able to answer all the questions we set out to answer and also some that were not intended but ended up emerging from the initial questions.

References

[1] https://data.census.gov, last accessed 13/12/2019.
[2] www.chicago.gov, last accessed 13/12/2019.
[3] https://data.cityofchicago.org, last accessed 13/12/2019.
[4] https://github.com/vgm64/gmplot, last accessed 13/12/2019.
[5] Malby, S.; Davis, P. Monitoring the impact of economic crisis on crime. Global Pulse, New York, USA (2011) 70 pp.
[6] U.S. Bureau of Labor Statistics, Unemployment Rate in Chicago-Naperville-Elgin, IL-IN-WI (MSA) [CHIC917URN], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CHIC917URN, December 13, 2019.
[7] http://gis.chicagopolice.org/clearmap_crime_sums/crime_types.html, last accessed 13/12/2019.