Project Report - H517

Visualizing Chicago Crime Data

Anushri Jayaswal, Utkarsha Bhamare, Purvi Bawal

Motivation:

Chicago is one of the most populous cities in the United States. It is also a hub of technology, industry and many more. Chicago is also one tourist place visited by many people around the year from within the US and also has many international tourists for its architecture. Along with all these beautiful things in Chicago, we found that the crime rate here is very high. For residents' and tourist safety it is good to know about the prevalent crimes in different parts Chicago. These visualizations will help people know about the overall crime rate and the details of various crime types in different regions along with the time of year when most crimes happen.

We found a huge dataset with details of all crimes reported in Chicago from 2001 to 2018 (<u>https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2</u>). We also found the dataset that gave us co-ordinates of different wards in Chicago so we could plan for plotting the map and make our visualizations more effective (<u>https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Wards-2015-/sp34-6z76</u>). The main features we have used from the massive dataset are as below:

Data Column name	Туре		
ward	This column gave us ward number in Chicago		
Primary_type	This gave us broad category of crime like assault, theft, robbery, etc		
year	This gave us the year in which the crime happened. This was part of the date on		
	which crime was reported and then extracted using sql queries.		
month	This gave us the month in which the crime happened. This was part of the date		
	on which crime was reported and then extracted using sql queries.		
domestic	This column has true, false data for domestic violence cases.		
arrest	This was originally in the form of true, false. We have used this to calculate the		
	arrest ratio to crime reported in that ward for our visualizations.		

Below are the questions that we have tried to address in our project:

- How have crimes rates evolved over years from 2001 in Chicago?
- What are high crime areas in Chicago based on wards?
 - Which type of crime is more prevalent in this area?
- Which police district has maximum percentage of arrests made for total crimes in that district?
- Are there any specific areas where there are more domestic violence cases?
- What does the crime trend look like over time of months of year in Chicago?
 - Which crimes are higher during the peaks?

Design process:

Data Analysis and questions:

We analyzed csv files of crime data for each year and looked at multiple columns in the dataset. As there were files for each year, the obvious question we thought of was to look at the trend of crime count over years. Since we had details of ward co-ordinates and map, we wanted to look at count of crimes ward wise in a given year.

We wanted to see which kind of crime is more common in a given area. So, the next thing we could think of was to look at the kind of crimes happening in each ward based on primary type.

We then thought of going to the detail of month wise crimes to analyze which type of crime is more prevalent in a given month. This may be due to the effect of weather or festive season etc.

Since we also had a column related to domestic violence, we thought it would be a good idea to see which wards have more such cases.

Productivity of police department in an area may be judged based on arrest made compared to crimes reported. So we decided to look at the wards based on arrest ratio as well.

Data Preprocessing:

As the size of our dataset was so large, we loaded our dataset into MySQL database. We created MySQL database instance using local WAMP server. With the help of SQL queries, we retrieved all the required data to plot the graph for each identified question. We built comma separated (.csv) file containing data required for each graph.

User Interface Design:

Our original plan was to have a main page with title and a menu containing links to pages. These pages would have contained visualizations of each question as depicted below:

template		S 1		
titl	e	uinks.	*	on click change map.
MAP	VISUAL	91 92 93		
		Years		



We discussed this idea of layout with our professor and found that he would prefer a single page layout with minimum scrolling and more interactivity. He mentioned that its okay to reduce the number of graphs but have more interactivity. Based on this we decided to discard the multiple page layout and go for the single page layout shown below:



We then decided to have map of Chicago on top left corner and the timeline chart of crimes over years on the right side of the map. We tweaked the question requirement a little and displayed the gradient of colors on ward map based on count in that year. Similarly, we depicted arrest ratio and domestic violence case count using gradient of colors on map. To save some more space, we decided to plot ward specific crime bar chart in a tooltip (on click). Below this we displayed the month wise count bar chart. On click of each month, we displayed/updated pie chart with top 5 crimes and their percentages. We provided a reset link as well to reset the graphs to original state.

Data Visualization:

We worked on identifying the appropriate choice of graphs to answer the identified breadth and depth questions and then we identified various programmatical ways of providing interactivity between the graphs. In our project, we used topojson.js to plot the map of Chicago wards, and D3.js to plot the various graphs and provide interactivity between them.

Black Box Testing:

We tested visualizations on various browsers and screen sizes to ensure the positioning of the graphs. We also verified visualizations using various color blindness tools to ensure colorblind-friendly webpages.

Design Justification:

We plotted a map of Chicago using boundary coordinates. The purpose of this project is to identify crime prone wards and the type of crime which is more prevalent in a specific ward. The map will help people to understand and explore the trends of crime over wards. There is facility of zoom-in/zoom-out along with panning of the map, which will help user to explore the wards and data more accurately.

On the map for each ward, at mouse click event, a tooltip bar chart will open. It demonstrates the top five crime in that particular ward. Bar charts are the type of graph that are used to display and compare the number for different discrete categories of data. Along with this, bar charts are simple to create and very easy to interpret. This is the reason we moved with bar charts to show type of crime in each wards with its count. To save the space on the screen, we created this bar chart in the tooltip.

We plotted a time line chart which shows the crime trend for each year from 2001-2018. As we all know that time line chart is an effective way to visualize a time trend. We had to show the total crime happened in that particular year. Y axis is able to demonstrate this, but to show the exact number, we created a tooltip on mouse hover which shows the total count of crime happened. Along with the total crime, we thought of showing the affected wards also in that particular year. As we were having a very large counts of each wards' crime, we decided 5 ranges of crime counts, so that we can easily assign a color gradient for each range. Fewer range of color gradient is easily interpretable to relate. We assigned lighter shade for less crime-prone wards and darker the more crime prone wards.

As per our plan, to analyze the domestic violence it needs to be reflected on the map for each ward. Because we had to only identify the impacted wards on map, we didn't plot anything for it rather than gave a link. On the click of that link, domestic violence will be reflected in the map by filling colors. There were so many wards, where domestic violence has not been reported, for those wards, we didn't change the original color of the map. For this also, we used color gradients with the same concept i.e. the darker the more domestic violence prone wards. Similarly, to demonstrate the arrest ratio on the map for each ward, we gave another link. On click of that link, the arrest ratio range will be reflected in the map for each ward.

We used bar chart to show the crime trend over months in Chicago as this is the suitable graph type to show the changes over time. Again, bar chart is simple and easy to interpret the relationship between each bar. As we have to identify months, which is significantly affected by crime, the bar chart was the most suitable chart for this purpose.

We used pie chart in order to represent top 5 contributing crimes from a month as it's generally used to represent proportional data by each category. We thought this would be the best way to represent top 5 crimes in Chicago. We clubbed rest of the crime types into 'others' category. We used color coding to represent the top crimes, crime type with higher crime count has darker color shade than others.

There are legends for each graph, which helps users to interpret accurately. Sometimes user need to refresh everything in-between the analysis, at the same time using link of reset, user can easily reload the page and can start analysis all over again.

Findings:

Below are the findings of our visualization :

- 1. Crime department from west side of the Chicago area made highest number of arrests out of total crimes cases reported.
- 2. Maximum number of arrests were made in the ward 8, 4 and 12.
- 3. Crime department from south side, west side, far southeast side and far southwest side area of Chicago has maximum number of domestic violence cases reported. Maximum number of domestic violence cases were reported in the ward 25.
- 4. We found that highest number of crimes were happened in the year of 2003, and then crime-count gradually decreased till the year of 2018.
- 5. In the year of 2003, maximum number of crimes were reported in the ward 3, 42, 20, 27 and 24 ward. Wards like 27,42,28 and 1 are more prone to crimes as they appeared frequently in the higher crime area.
- 6. We observed some spike in crime-count during the year 2016 and 2017. As the sufficient data was not available for the year 2001, it has very low crime count.
- 7. Highest number of crimes were reported in month of July (crime count: 630314). Second highest number of crimes were reported in the month of August (crime count: 624185). Lowest number of crimes were reported in the month of February (crime count: 463816).
- 8. Theft, battery, criminal damage, narcotics and assault were the frequently occurred crime types in the Chicago.

Reflection:

We learnt that the best way to keep any visualization informative is to keep it simple and clear but with enough information. Adding a lot of connected plots may cause confusion in some cases. It is very important to design the visualizations with knowledge about background of audiences.

We also learnt about using correct kind of graph for effective communication. Like line chart is best for showing trend, pie chart is best for showing percentage contribution and usage of single. color gradient.

There are many more things that can be explored using this dataset and use the data columns like street, location, beat, district, community area.

The street level crime analysis can be done but it could not have been appropriate for a small screen. We can do this using zooming and can be considered as scope for future work.

There is scope of analyzing arrest ratio per police beat instead of wards to see which beat has maximum/minimum arrest ratio and maybe relate it to productivity.

One thing we wished we had done was to connect the month chart with year timeline chart to have three-way connectivity of graphs that could have enabled a different way to analyze crimes.