CS5803 Data Visualisation 2023/4

Student Id: 2349128

Introduction

The dataset we picked is a Crimes dataset by Los Angeles Police Department. This dataset represents occurrences of criminal activity in the City of Los Angeles starting from the year 2020. The data provided is a transcription of the original crime reports, which were recorded on paper in such a way that it preserved privacy. This dataset contains over 900K rows and around 28 columns. The dataset gets updated bi-monthly. This dataset was huge for our use case, so we simply cleaned it and sampled 10K rows from year 2023 and 2024 each. For the cleaning, we had to remove certain negative values from the age and drop the null values present in the impactful columns. The data includes various attributes such as the division of records number, date and time of occurrence, geographic area, and much more which is covered in detail in the data dictionary below.

You can also find the data at: https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/about_data

Data Dictionary

Variable Name	Description	Datatype
DR_NO	Division of records number: Official File number	Numeric
Date Rptd	Date the file was reported on. Format : MM/DD/YYYY	Date
DATE OCC	Date the crime occurred on. Format: MM/DD/YYYY	Date
TIME OCC	Time the crime occurred in 24 hr format	String
AREA	Number of Geographic areas within LA, numbered 1-21	Numeric
AREA NAME	Name of the Geographic area within LA	String
Rpt Dist No	A four-digit code that represents a sub-area within a Geographic Area	Numeric
Part 1-2		
Crm Cd	Crime code which represents the type of crime committed	Numeric
Crm Cd Desc	Describe the code of the crime committed	String

Mocodes	Modus Operandi: Activities associated with the suspect in commission of the crime	Numeric
Vict Age	Age of the victim	Numeric
Vict Sex	Sex of the victim. F- Female, M- Male, X- unknown	String
Vict Descent	Ethnic or racial background of the victim. A- Other Asian, B- Black, C- Chinese ….	String
Premis Cd	The code of type of structure/location/vehicle the crime took place	Numeric
Premis Desc	Describes the premise code above	String
Weapon Used Cd	Code for type of weapon used	Numeric
Weapon Desc	Describe the code for type of weapon used	String
Status	Status of the Case. IC is Investigation Continued	String
Status Desc	Describe the code of the status	String
Crm Cd 1	Indicates the crime committed. Crime Code 1 is the primary	Numeric
Crm Cd 2	May contain code of additional crime	Numeric
Crm Cd 3	May contain code of additional crime	Numeric
Crm Cd 4	May contain code of additional crime	Numeric
LOCATION	Street address of crime incident rounded to the nearest hundred block to maintain anonymity.	String
Cross Street	Cross Street of rounded Address	String
LAT	Latitude	Numeric
LON	Longitude	Numeric
	Custom Variables	
AgeGroups	Grouping different ages into groups like ‘0-12(Children)’	String
TypeOfArea	Grouping different areas into groups like ‘Commercial’ ’	String
TypeOfCrime	Grouping Crimes into broader group like ‘Assault & Battery’	String

Violent Or Non-Violent Crimes	Groups Crimes into either violent or non-violent crime	String

Persona

A well-constructed persona serves as an easily understood and practical depiction of the underlying user group. (Jansen, Salminen and Jung, 2020). Jansen, Salminen and Jung (2020) discuss the importance of personas and how they make the data easier to interpret and what the persona would look/aim for. Choosing a persona who is relevant and to whom the data would actually be useful is crucial.

The persona I selected is that of a Crime Analyst. In this role, he is responsible for examining crime patterns, identifying trends, and providing actionable insights to create and implement strategies to reduce crime rates and improve community safety. As a crime analyst. I work closely with law enforcement officers and detectives to provide them with insights that help in solving cases and preventing future crimes.

Questions

While formulating questions for a crime analysis point of view, I'll also keep in mind that the visualizations should be following certain principles. I

1) How does the frequency of reported crimes vary throughout the day across different types of premises in Los Angeles? [Time-Series, looking for the time when the crime is highest]

R1) The graph depicts the variation in reported crimes over the course of a day, segmented by different types of premises where the crimes occurred. By examining this graph, we can gain insights into the temporal patterns of crime occurrence in different settings within the city. Each line on the graph represents a specific type of premise. The interactivity enables you to use the whole dashboard more efficiently.

2) How does the distribution of victim descent vary across different geographic areas of Los Angeles over quarterly intervals? [Distribution, where is the highest concentration of a specific demographic]

R4) The geo-map visualizes the spread of victim descent across various geographic areas of Los Angeles over quarterly intervals. Each geographic area is represented on the map, with color-coding indicating the predominant victim descent group reported in that area during a specific quarter. The map allows us to discern spatial patterns in the distribution of victim descent within the city over time. By examining changes in colour and distribution across different quarters, we can identify any shifts or trends in the demographic composition of crime victims across various regions of Los Angeles. Geospatial navigation is important for traversing and interacting with geo-maps.

3) How do the types of weapons used in reported crimes vary across different premises in Los Angeles ? [ranking, which type of premises are most in danger]

R3) This is a tree map visualization which provides an overview of the types of weapons used in reported crimes across various premises in Los Angeles. Each rectangle within the tree map represents a specific premise, with the size of the rectangle proportional to the number of reported incidents in that location. Additionally, the color-coding of each rectangle indicates the predominant category of weapon used in crimes occurring at that premise. By examining the tree map, viewers can quickly discern patterns and trends in weapon usage across different types of premises.

Complex Question

1) How does the distribution of reported crimes vary across different types of crime and which crimes are prominent in certain premises?

   R1 This bar graph illustrates the distribution of reported crimes in Los Angeles, categorized by both the type of crime and the premises where the crimes occurred. Each bar represents the number of reported incidents for a specific combination of crime type and premises type. By examining this graph, we can gain insights into the prevalence of different types of crimes across various locations within the city. We'll look at the Top 5 premises in Los Angeles, as these are the hotspots for crimes and need more focus or manpower.

Requirements

Functional Requirements

R1: Interactive Exploration (implemented in all graphs):

Weber and Hauser (2014) discuss how interactive exploration leads to effective knowledge discovery. It leads to ease of access in extracting useful information and insights.

Users should be able to interact with the visualizations to explore detailed information about reported crimes, such as hovering over specific data points to view tooltips with additional details.

R2: Filtering Options (implemented in all graphs):

Users should have the ability to filter the data based on different criteria, such as crime type, premises type and time intervals, allowing them to focus on specific subsets of the data for analysis.

Even though all the graphs can filter, only the geo-map and the tree map have external list for filtering.

R3: Comparative Analysis (implemented in all graphs):

Yoshizumi et al. (2020)

The system should support comparative analysis features, enabling users to compare trends, patterns, and distributions across different categories, such as comparing crime frequencies across quarters or weapon types across premises.

R4: Geospatial Navigation (implemented using the geo-map):

For geo-map visualizations, users should be able to zoom in/out, pan across the map, and click on specific geographic regions to view detailed information about crime statistics and victim demographics for those areas.

Non-functional Requirements

Usability:

The goal of usability is to detect, articulate, and fulfil user demands (Grinstein et al, 2005). Even when comparing and contemplating the Usability vs Utility factor, Grinstein et al (2005) has the stance that both are important in their respective ways. Usability is thus an important aspect which should be covered as a requirement.

The user interface should be intuitive and user-friendly, with clear navigation pathways, informative labels, and consistent design patterns, facilitating ease of use and minimizing the learning curve for users interacting with the system.

Colour Scheme and Visualization Design:

Bartram, Patra and Stone (2017) carried out studies to examine how different colours affect their interpretations. We learn that a colour scheme can have a major impact on how easy the data is to interpret and the consistency.

A carefully chosen colour scheme should be used to enhance readability and interpretability of visualizations with Bertin's Theory in mind. Visualizations should employ appropriate chart types and graphical elements to effectively convey information without overwhelming the user.

Feedback and Interactivity:

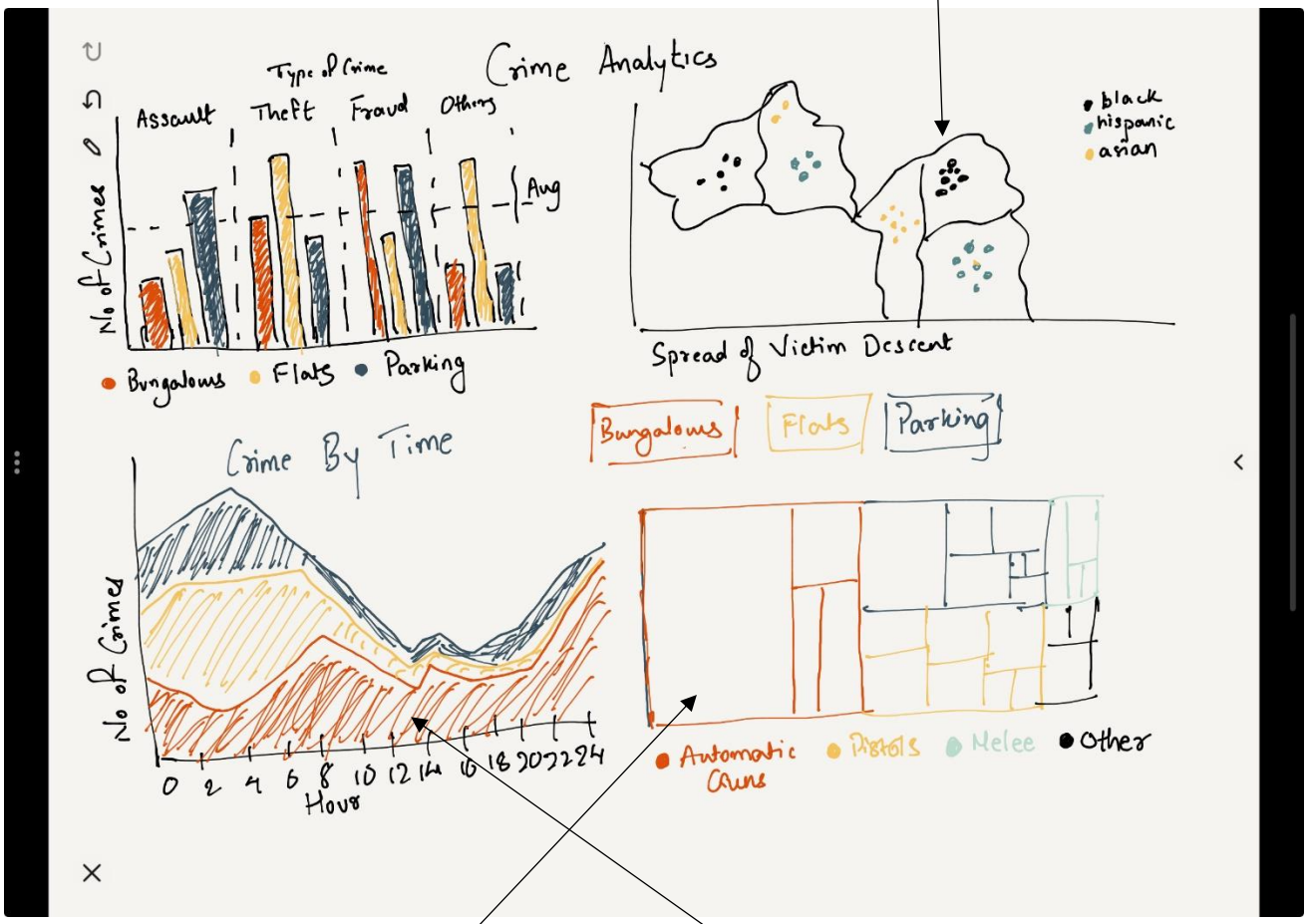
Saket et al. (2018) delves deeply into the concept of implementing the user interactions directly into the visual representation and its effectiveness. We learn that interactivity is complex on its own but varies a lot on the application. Which makes it a vital factor to carefully choose and work on.

The user interface should provide feedback to user interactions, such as highlighting selected data points or displaying tooltips with additional information, to enhance the user's understanding and engagement with the system.

Design

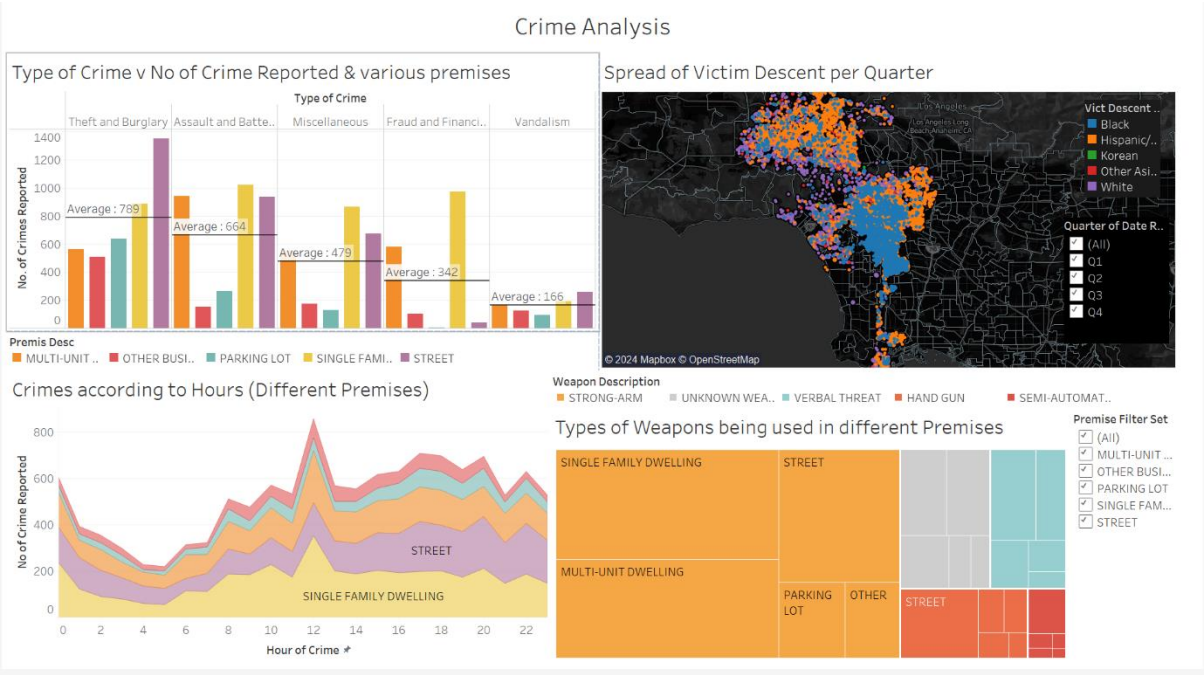
Paper Landscape

Filter Buttons, whenever you press them, the whole dashboard adapts to those options.



Will have tooltips shows the exact no of crimes and the label of Premises.

All areas are clickable and act as filters for the rest of the charts, same for every chart



There are few differences in dashboard and the paper landscape:

First of all, the bars are more in number instead of just 3.As more were feasible, it works well.

Additionally, I added a quarter filter in the geo-map, as there are too many data values, you can see the concentration even after filtering.

I added 1 more filter for premise which is a shared interactive filter, that filter is made for all the charts in the dashboard.

The Filter buttons were too complex to make and also wouldn't make much sense if I wanted multiple selections.

Implementation:

Tableau:

After performing basic data cleaning and formatting, the data is prepared for visualization. Just make sure, the dates are correctly formatted as dd-mm-yyyy.

You'll need to make a few variables which are broader categories which make visualisations much cleaner and easier to interpret. Simply make a calculated field named TypeOfCrime and write switch statement or if else and contain all assault crime into Assault category and so on. Drag Type of Crime and Premise Desc into Columns and Count of Dr no in rows which represents the number of crimes reported. Also assign premise desc to colour and you're good to go. Apply filtering on premise desc to reduce the number of bars on the screen. Typically, I would go for top 5 or top 10.

For the Treemap, simply drag count of drno, weapon desc and premise desc in any row or column and select Treemap in the show me pane. That's all you must do, again filter out top 5 or top 10 premises and weapons. I specifically removed no weapons as I didn't that for graph.

For the line area graph, Use Date & Time Occurred and select Hours, place count of dr no in rows. Add premise desc to colour and label both. Apply basic filtering to achieve readability, I filtered out few of the top premises.

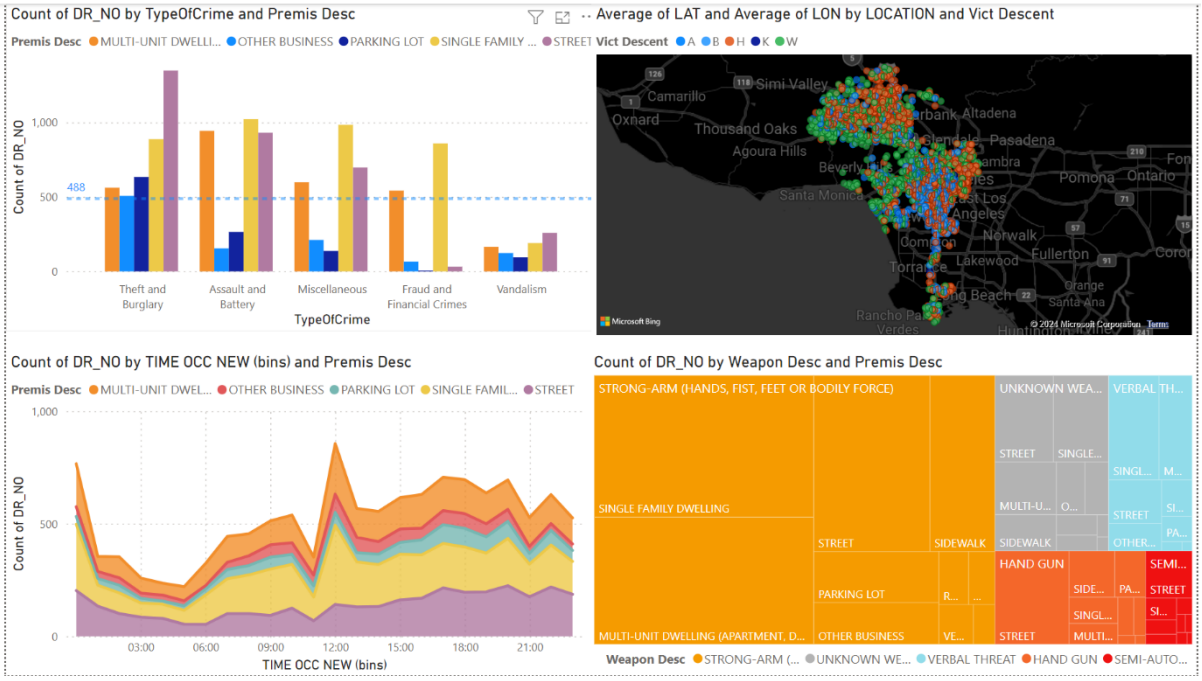
For the geo map, we have location, longitude and latitude available to us from the dataset. Drag vict descent to colours and count of vict descent in size. Apply filtering on the victim descent, I removed the Others and Unknowns from Descent.

Power BI:

The PowerBI implementation was a whole different process. Instead of dragging, simply clicking on the checkbox was much easier option.

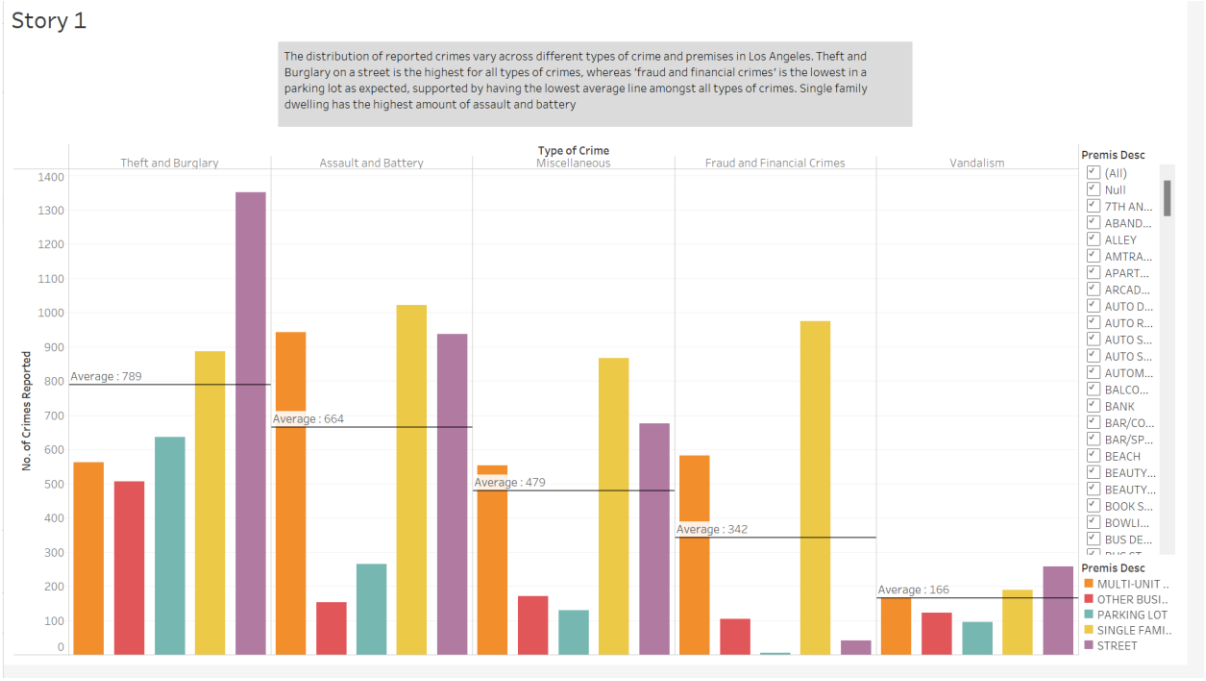
Select bar charts in visualisation and select relevant features as we did above for you to create the first bar graph. Make sure to filter

Similarly line graph and treemaps are also easy to select from the visualization and simply click on the relevant data columns.

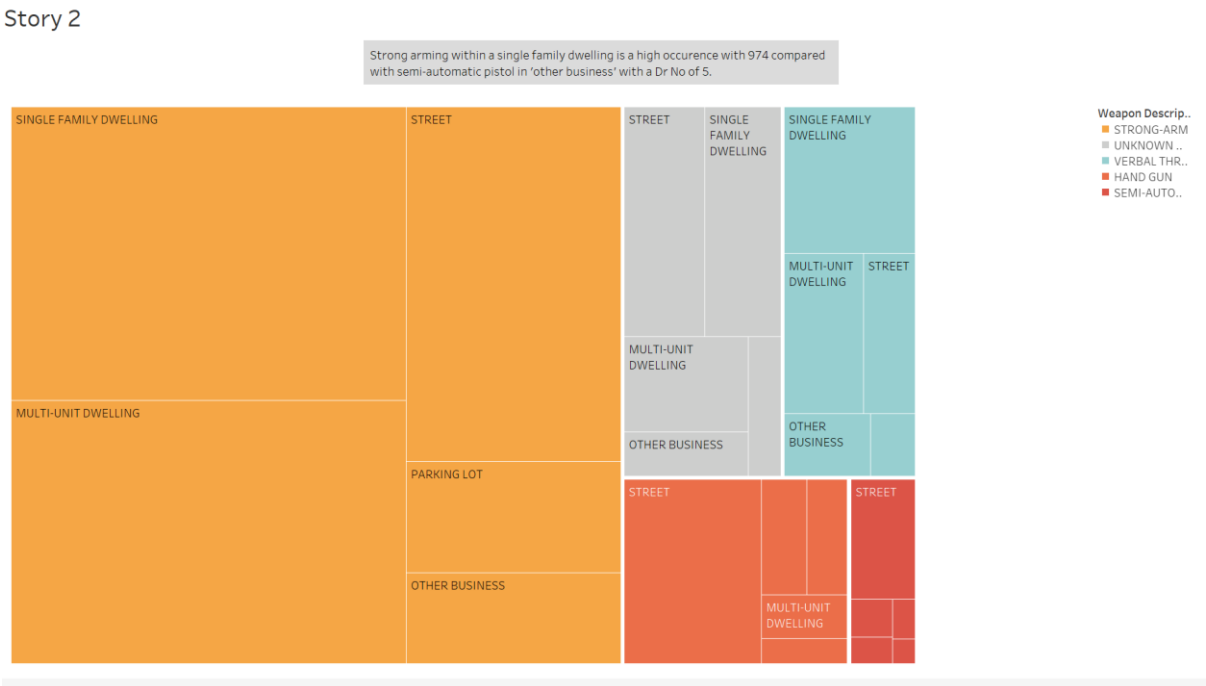


Walkthrough:

From the stories we gain some insightful information. From the 1st Story, we find facts like Theft and burglary are highest in Streets . The visualizations make it easy to find information and the average line helps us identify the variation present.

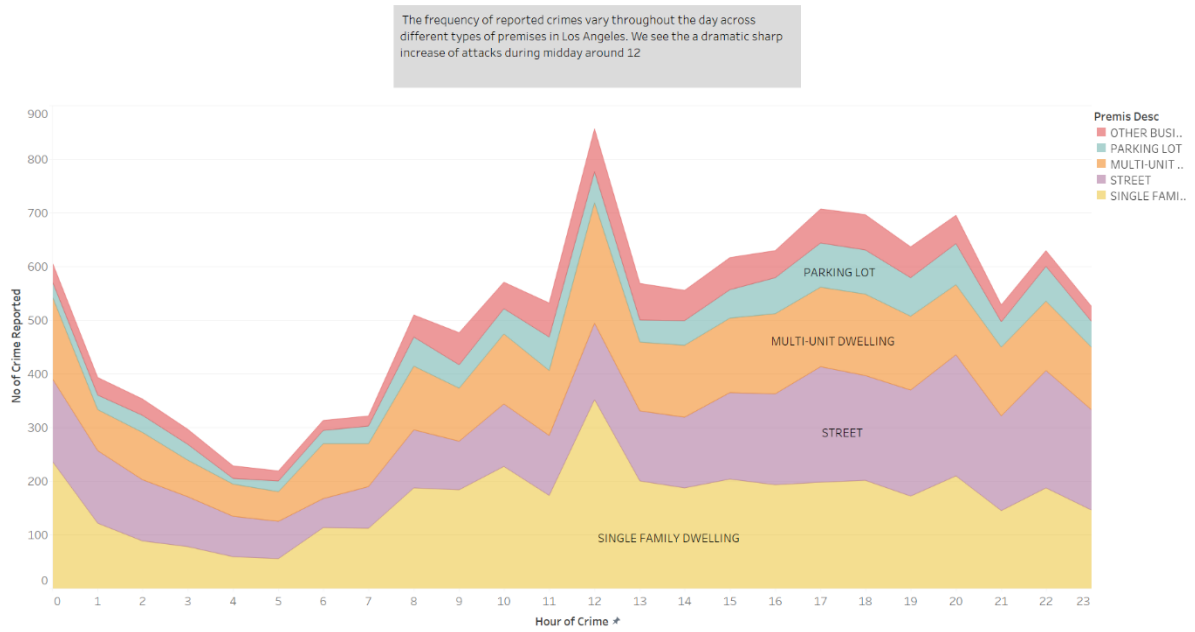


From the 2nd Story, we find how some crimes take up huge space meaning there is significant more amount of crime happening compared to other crimes.



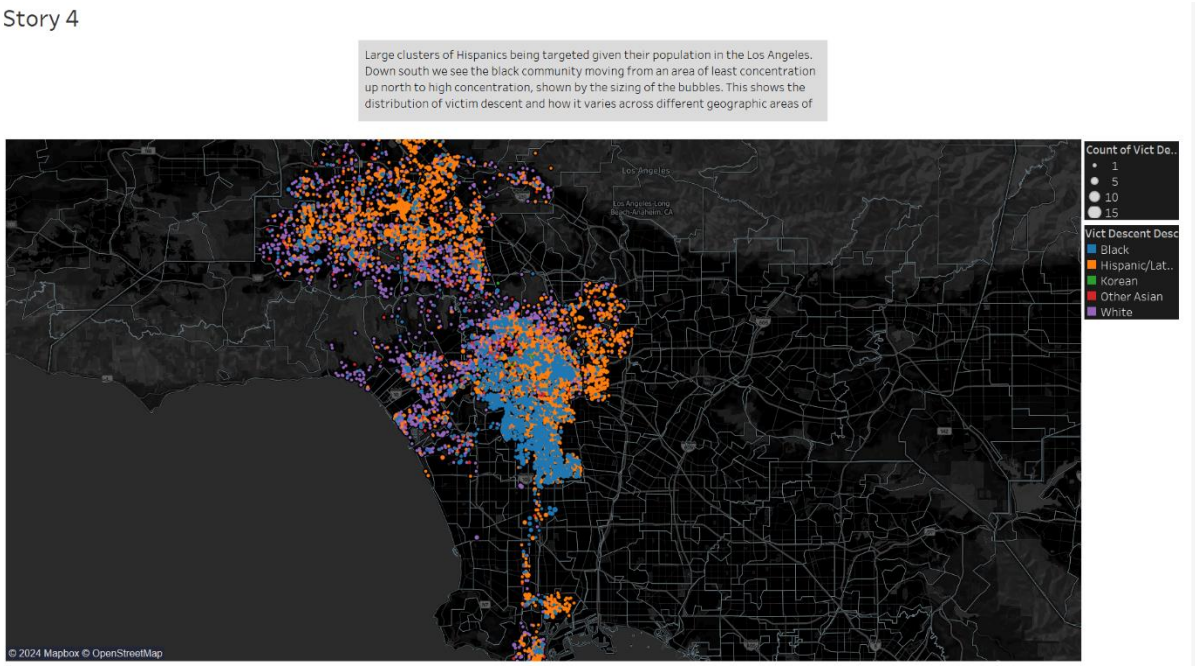
In the 3rd Story, we learn how time affects the rate of crime and from the spike in the middle we learn that crimes increase around midday. We can even compare the thickness of premises and compare which ones higher and which ones are lower.

Story 3



In the 4th Story, we see patterns of people of similar descent suffer from crimes in those specific areas. Theres a possibility that people from certain descent get targetted in those areas.

Story 4



Discussion:

Reflecting on the crime analysis project as a whole, it has been an enriching experience that has provided valuable insights into both the complexities of crime data analysis and the capabilities of visualization tools.

Insightful Analysis: The project allowed for in-depth analysis of crime data, uncovering patterns, trends, and correlations that contribute to a better understanding of crime dynamics in Los Angeles.

Tool Constraints: The limitations of the visualization tools used, such as restricted customization options or performance issues with handling large datasets, may have constrained the flexibility and scalability of the visualizations. Especially in powerbi

I look forward to using more of tableau and the new features they will be coming up with. Not much excited about PowerBI though. But new applications might be promising.

Conclusion:

After all the visualizations and analysis, we find answer to multiple of our questions which had various requirements. The project deepened our understanding of crime dynamics, including factors influencing crime occurrence, spatial clustering of incidents, and demographic patterns of victimization. By examining crime data from multiple perspectives, we gained a more holistic understanding of the challenges and opportunities for enhancing public safety in Los Angeles. Visualizations proved to be powerful tools for communicating complex crime data in a clear and intuitive manner.

While the project yielded valuable insights, we acknowledge the need for continuous improvement in data quality, visualization design, and analytical techniques. This was actually fun.

References

- Bartram, L., Patra, A. and Stone, M. (2017) “Affective Color in Visualization,” in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM.

- Grinstein G, Kobsa A ,Plaisant C ,Stasko J. (2005) “Which comes first, usability or utility?,” in IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control. IEEE.

-Jansen, B. J., Salminen, J. O. and Jung, S.-G. (2020) “Data-driven personas for enhanced user understanding: Combining empathy with rationality for better insights to analytics,” Data and Information Management, 4(1), pp. 1–17. doi: 10.2478/dim-2020-0005.

- LAPD (2024) Crime data from 2020 to present: Los Angeles - open data portal, Crime Data from 2020 to Present | Los Angeles - Open Data Portal. Available at: https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8 (Accessed: 24 March 2024).

- Saket, B., Arjun Srinivasan, E., Ragan, A., Endert (2018) “Evaluating interactive graphical encodings for data visualization,” IEEE transactions on visualization and computer graphics, 24(3), pp. 1316–1330. doi: 10.1109/tvcg.2017.2680452.

- Weber, G. H. and Hauser, H. (2014) “Interactive Visual Exploration and Analysis,” in Mathematics and Visualization. London: Springer London, pp. 161–173.

-Yoshizumi A., Coffer M., Collins E., Gaines M., Gao X., Jones K., McGregor I., McQuillan K., Perin V., Tomkins L., Worm T., Tateosian L. (2020) “A Review of Geospatial Content in IEEE Visualization Publications,” in 2020 IEEE Visualization Conference (VIS). IEEE. Available at : https://dl.acm.org/doi/10.1145/1103900.1103931