

CS7DS4 / CSU44065 Data Visualization 2023-24

Assignment - 4

Student Name: Swetang Krishna

Student No: 21355087

DECLARATION

"I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at <http://www.tcd.ie/calendar>.

I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at <http://tcd.ie.libguides.com/plagiarism/ready-steady-write>"

Title: Enhanced Crime Data Visualization and Analysis

Abstract: This report discusses the implementation of an advanced data visualization for crime data analysis. The visualization employs various encoding channels and interactive features to support multiple tasks, providing a complex yet intuitive overview of crime patterns over time and demographics.

1. Introduction The visualization aimed at offering users a multifaceted view of crime statistics, aiding in identifying patterns and correlations that could inform policy and enforcement strategies. The dashboard offers the user different functionalities to analyze the reported crime by the US government. The following features are offered by the dashboard:

- Temporal Analysis:** With the "Total Crime Reported by Day" bar chart, one can analyze crime trends over time, identify patterns with respect to the days in a week.
- Demographic Analysis:** The doughnut chart titled "Total Crime Reported by Victim Sex" and the bar chart for "Victim Descent" provide insights into the demographics of crime victims. You can determine which groups are more frequently targeted and assess whether certain demographics are disproportionately affected by crime.
- Crime Severity:** The "Severity" filter along with associated data points on the dashboard can help in analyzing the breakdown of crimes by severity. This enables a focus on more severe crimes, potentially prioritizing resources and responses.
- Case Status:** By filtering the data with the "Case Status" dropdown, you can get a sense of how many cases are open, closed, or in other statuses, which is useful for managing the workload of investigative teams. This filter also has the ability to identify if the involved suspect is an adult or a juvenile.
- Geospatial Analysis:** The map visualization "Total Crime Reported by LAT and LON" offers a geographical analysis of where crimes are occurring. This can be critical for allocating police resources, planning patrols, and community safety initiatives.
- Victim Age Analysis:** The chart "Total Crime Reported by Victim Age" allows for understanding the age distribution of crime victims, which can inform prevention programs and other community support services.
- Weapon-Related Crimes:** The "Weapon Related Crimes" metric can inform policy and law enforcement strategies concerning weapon control and intervention programs.
- Effectiveness of Law Enforcement:** The "Solved Cases %" gives an overview of the effectiveness of the police force or investigative bodies in solving crimes. A lower percentage might indicate a need for more resources or changes in tactics.
- Correlations and Insights:** By interacting with the dashboard, one can explore correlations between different types of data, such as the link between crime severity and victim demographics, or the relationship between case status and crime type.
- Historical vs. Current Analysis:** The date filter allows for historical comparison to understand if crime rates are increasing or decreasing over time.

Total Crime Reported 259.20K	Weapon Related Crimes 148.27K	Severe Crimes % 100.00%	Solved Cases % 31.74%
--	---	-----------------------------------	---------------------------------

2. Tools and Data Pre-processing

- **Tools Used:**

- Power BI for data visualization and interaction
- Python for data cleaning and transformation
- About the DATASET,
link to the dataset:
<https://catalog.data.gov/dataset/crime-data-from-2020-to-present>

Dataset description:

This dataset reflects incidents of crime in the City of Los Angeles dating back to 2020. The crime reported until 13th Nov 2023 have been recorded in the dataset.

- **Pre-processing Steps:**

- Data cleaning: Missing values handled, outliers identified and corrected.
The data cleaning was performed in many applications such as MS-Excel, Python Jupyter Notebook and Power BI. Some of the data was manually removed as it was garbage data.
- Data transformation: Date and time attributes standardized; categorical data encoded.
The Date in 'Date reported' and 'Date Occurred' columns also included time; the time wasn't required for the analysis so it was removed. All the rows with empty cells were removed. For example, the rows with null value in the 'Victim Desc', 'Victim Sex', 'Victim Age' were removed as this would create a chaos while analyzing the data. Some of the columns were removed too such as 'Mocodes' and 'Cross Street' which were irrelevant for the visualization.
- Data aggregation: Summarization for daily crime totals and category breakdowns. The final summarization was done in Power Bi which provides filters for each column of the dataset. The final dataset was reduced to 93,546 KB from 213, 990 KB.

3. Dataset Overview

- **General Description:**

- A comprehensive compilation of crime reports over a specified period.
- Includes various attributes such as date, type of crime, victim demographics, and location coordinates

- **Detailed Analysis:**

- Data types: Numeric, categorical, datetime, and geospatial data, Victim Sex, Ethnicity.
- Attribute types: Discrete (e.g., crime description), continuous (e.g., age), ordinal (e.g., severity)
- Derived attributes: Calculations such as crime solved percentage, percentage of weapon-related crimes.

- **Complexity Justification:**

- Multidimensional data requiring sophisticated visualization techniques.
- Temporal, demographic, and spatial data interrelations necessitate an interactive platform (Dashboard created in Power BI) for effective exploration.

4. Supported Tasks

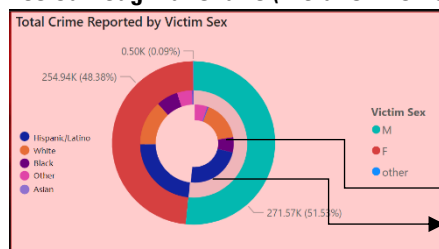
- Temporal analysis of crime trends over time, day.
- Demographic breakdown of crime victims. (Age, Sex, Ethnicity)
- Spatial distribution of crime hotspots. (Location on Map)
- Correlation identification between crime types and victim demographics. (Severity of Crime correlated with the location)

5. Encodings and Idioms

- **Bar Chart ("Total Crime Reported by Day"):**

- Position (X-axis): Represents days of the week, effectively depicting the temporal distribution of crimes.
- Length: The length of the bars indicates the quantity of crime reported, allowing for a quick comparison across days.

- **Nested Doughnut Charts ("Total Crime Reported by Victim Sex" and the "Ethnic Group they belong"):**



- Angle and Area: Each segment's angle and area represent the proportion of crimes reported by victim sex, providing a clear visual distinction between categories.
- Hue: Different colors signify different sexes, utilizing categorical hues to distinguish data points easily.
- The first inner doughnut chart is used to identify the ethnic group of the male victims. The second inner doughnut chart represents the ethnic group of the female victims.

- **Map Visualization ("Total Crime Reported by LAT and LON"):**
 - Position: Geographic location of crimes, intuitively representing where incidents have occurred.
 - Bubble Intensity: Could represent the density or severity of crimes, with a greater number of bubbles indicating higher frequencies or more severe crimes.
- **Tree Map ("Total Crime Reported by Victim Descent"):**
 - Hue and Area: Utilized to differentiate between ethnic groups, with distinct colors aiding in quick recognition. The area signifies the number of victims per ethnic group.
- **Area Chart ("Total Crime Reported by Victim Age"):**
 - Position (X-axis): Represents age groups, effectively mapping a continuous variable.
 - Area: The size of the area under the curve correlates with the number of crimes, allowing for the perception of trends across age groups.

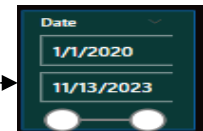
6. Novelty and Complexity

The dashboard employs various visualization idioms—bar charts, nested doughnut charts, area charts, and tree map, Aerial style Map, each chosen for their strength in encoding specific data types and relationships. The combination of these idioms presents a complex dataset in an accessible and information-rich format. Despite the potential complexity, the design maintains a balance, avoiding clutter and facilitating ease of interpretation.

Dynamic elements enhance the utility and interactivity of the dashboard:

Tooltips: Implemented across all visualizations, tooltips provide additional data such as exact numbers or percentages when the user hovers over any data point. This design choice effectively minimizes on-screen clutter by replacing the need for static data labels.

Date Range Slider: This interactive element allows users to refine the displayed data to a specific period of interest. It aids in temporal analysis, helping to discern patterns or trends over time without the need to navigate away from the dashboard.



Filtering Options: Filters for severity, case status, victim sex, and ethnicity provide a means to dynamically customize the data presentation. Users can isolate specific subsets of the data for a more focused analysis, such as examining crime trends within a particular demographic group or case type.

7. Critical Analysis

Strengths of the Dashboard

- **Multifaceted Data Presentation:** The dashboard successfully combines different data visualizations to present a holistic view of crime statistics.
- **Interactivity:** Features like tooltips and date sliders engage users and provide additional context without cluttering the visual space.
- **Visual Hierarchy:** The dashboard has a clear layout with a visual hierarchy that guides the viewer through the data in a logical sequence.
- **Data Filtering:** The ability to filter by date, severity, case status, and victim demographics allows for dynamic data exploration.
- **Nested Charts:** The nested doughnut charts help the user to analyze the crime against a Male and their ethnic groups.

Weaknesses of the Dashboard

- **Cognitive Load:** The amount of information and variety of visualizations may overwhelm users, especially those unfamiliar with data analysis.
- **Potential for Clutter:** The map may become cluttered if many crimes are concentrated in a small area, which could make it difficult to discern individual incidents.
- **Data Density:** Some visualizations may be too dense or too sparse depending on the applied filters, potentially leading to misinterpretation of the data.

Reference:

Images: <https://pixabay.com/images/search/crime%20background/>

Dataset: <https://catalog.data.gov/dataset/crime-data-from-2020-to-present>

Tutorials:

<https://learn.microsoft.com/en-us/power-bi/fundamentals/service-get-started>

<https://www.process.st/how-to/create-a-measure-in-power-bi/>

<https://zebrabi.com/guide/how-to-add-filter-in-power-bi/>

<https://data-flair.training/blogs/power-bi-card-visualizations/>