

An overall introduction to your project, including the topic and tasks. (5 pts)

Crime is a multifaceted issue that affects all areas of society, and understanding its patterns is crucial for effective law enforcement and community safety strategies. Boston, being one of the major cities in the United States, presents a unique opportunity to analyze urban crime dynamics due to its diverse socio-economic landscape. By studying crime rates and their distribution across different neighborhoods, we aim to uncover patterns that could inform better policing policies, community outreach programs, and preventative measures. This analysis not only seeks to highlight areas with high crime rates but also to explore the underlying factors contributing to these incidences, such as economic disparities, educational opportunities, and urban development. Our project is significant as it provides insights that could help in crafting targeted interventions aimed at reducing crime and enhancing the quality of life for Boston's residents.

A paragraph to introduce the data, including the size, source and attributes. (5 pts)

The data we will be using for this project and analysis comes from the City of Boston's online data collection and we will be mainly referring to the Crime Incidents Report (August 2015 - Present). The Crime Incidents Report is provided by the Boston Police Department and it documents the initial details of any incident that the BPD is called for. This dataset contains ten years of data and is updated frequently to include data from current experiences. In terms of the size of our data, in the file containing data for 2023-Present there is currently over 16,330 rows of data and this specific file will continue to be updated as our project continues. In files for years before 2023, there are more than 7000 rows of data recorded. The source of the data, the city of Boston, is reputable and it is coming directly from their reporting site, Analyze Boston. We expect to use other data sets to support the primary Crime Incidents Report and create more in-depth analysis. Two datasets that we have already found are the Boston Shooting dataset and the Boston Shots Fired dataset. Both the Boston Shootings dataset and the Boston Shots Fired dataset come from the same source, Analyze Boston, and have data from 2015 to present that is updated on a rolling basis of about 7 days. The Shootings dataset contains information on shooting incidents where a victim was actually struck by a bullet either fatally or non-fatally. The Shots Fired dataset is similar to the shootings dataset but it is a collection of data on incidents where a shot was fired in the city of Boston under BPD jurisdiction and did not hit any victim.

- At least two links to the references. (5 pts)

[Understanding crime patterns using spatial data analysis](#): This paper discusses methodologies for analyzing crime spatially across urban areas, which will inform our approach to visualizing and interpreting crime data in Boston.

[Community policing and urban safety: reinventing security in city centers](#): This study explores the impact of community policing on crime rates in major cities, which will provide a comparative framework for our analysis of crime prevention effectiveness in Boston.

• We also need a paragraph to explain the takeaway of the visualization on the webpage, and a paragraph to explain the design idea for the visualization on the separated word document. (40 pts)

1. This chart plots the total number of reported crimes per month across several years, helping to illustrate any recurring seasonal peaks or valleys. In many of the years shown, we see a noticeable rise in crimes from late spring through early fall, with July and August showing around 10–15% higher reported incidents than the winter months. By comparing multiple years in one view, it's easy to confirm whether summer spikes are consistent patterns or if certain months only stood out in specific years. These trends can lead us to ask deeper questions about contributing factors. For example, do higher crimes in warmer months correlate with local festivals, holiday celebrations, or changes in policing resources?
 - a. Chart Type (Line Chart): We chose a multi-line chart to show how crime totals fluctuate by month for several years in one view. Line charts effectively convey trends over time and make it easy to compare one year's pattern against another. Color Palette: Each year is assigned a distinct color that stands out but is still part of a cohesive palette, helping viewers instantly see which line belongs to which year. Axis & Labels: The x-axis represents months (January–December), and the y-axis represents the number of crime incidents. We included a clear title, legend, and labeled axes so that users can quickly interpret the data.
2. Building upon the first visualization, the stacked bar chart breaks down the monthly totals into individual offense categories. A clear observation is that several crime types, like larceny and assault, often dominate, comprising a significant share of total offenses in certain months. For instance, in July and August, vehicle theft can sometimes surge to nearly double its wintertime rates, suggesting a seasonal vulnerability. Identifying such patterns helps law enforcement pinpoint exactly which offenses spike in specific months, guiding more focused interventions and resource distribution.
 - a. Chart Type (Multi-Line): A separate line for each day of the week helps compare daily patterns across the 24-hour cycle. Hourly intervals on the x-axis show how crime counts shift as the day progresses. Color Scheme: Each day (Monday through Sunday) is assigned a unique color so the lines don't blend together. We typically use a light-to-dark or rainbow approach to differentiate days while keeping the palette readable. Tooltip Interaction: Users can hover over any point on a line to see the specific hour and crime count for a given day, revealing if late-night Saturday spikes are significantly bigger than weekday evening peaks, for example. Gridlines & Axis Labels: Including hour markers on the x-axis and a y-axis for crime counts ensures clarity. Subtle gridlines can help trace a specific hour's crime volume across different days.
3. Focusing on the different years, which each had a different socio-economical effect on society, can give us insight on if monthly data is consistent through the years. In particular, the early months of 2020 show a dip in reported incidents, likely connected to

the onset of COVID-19 restrictions. By late 2021, however, crime figures begin to climb closer to pre-pandemic levels. These findings can spark discussion about how broader societal factors, like economic strains or changes in social behavior, might have influenced crime rates and whether current strategies are effectively addressing these new conditions.

- a. Focus on Recent Data: By limiting the timeframe to 2019–2023, we could create a less cluttered line chart that highlights more current shifts and makes year-to-year comparisons simpler.

Minimalist Color Scheme: Because we're displaying fewer years, a smaller but distinct set of colors helps each line stand out clearly.

Highlight Significant Periods: Some implementations might add subtle shading or markers to indicate major societal events (for example, a gray band for COVID-19 lockdown months), so viewers can quickly see where real-world events align with spikes or dips.

Tooltips & Axis Labels: We kept the design similar to previous graph for consistency: months on the x-axis, crime counts on the y-axis, and tooltip popups for precise data points.

4. This chart ranks the ten most frequent offenses each year, making it straightforward to see if certain categories are expanding or contracting over time. For example, vandalism may account for up to 20–25% of the total in some months of a particular year, while drug-related offenses can spike in another. Seeing categories like “Investigate Person” or “Simple Assault” consistently near the top also helps stakeholders understand which problems remain persistent. If vandalism shows a noticeable jump two years in a row, it could be a sign that community outreach programs or specialized policing might be needed to address property damage concerns.

- a. Chart Type (Stacked Bar for Ranked Categories): We chose a stacked bar to let users compare both the absolute and relative size of the top 10 offenses in each year.

Ordering & Ranking: By highlighting only the top 10 categories, we avoid an overly busy chart and direct attention to the most impactful or frequent offenses.

Bars are typically sorted from the most frequent at the bottom to the least frequent at the top (or vice versa), making year-to-year patterns more apparent.

Color Palette for Categories: Each category has a distinct hue consistent with the palette used in the previous graph, ensuring continuity across all offense-type visualizations.

Axis & Legends: A clear legend helps viewers map each color to an offense category, while x-axis labels indicate the year.

5. This graph shows the relationship of when crime data meets external socioeconomic variables, such as median income, population density, or educational attainment, to identify potential correlations. Notably, several districts with lower average incomes display higher crime volumes, suggesting a strong relationship between economic conditions and certain types of offenses. At the same time, a few outliers show moderate or high incomes but still elevated crime levels, which causes questions to arise about

other possible contributors, such as proximity to nightlife areas or major transportation hubs. By drilling down to specific points in this scatter plot, viewers can explore local nuances, like a district's recent gentrification or targeted community initiatives, that may explain why it doesn't fit the broader trend.

- a. Chart Type (Scatter Plot): A scatter plot is great for displaying relationships between two continuous variables, crime counts (or rates) and a socioeconomic measure (like median income).

Axis Configuration: Typically, the x-axis might show the socioeconomic indicator (like the income), while the y-axis represents crime volume. This layout helps emphasize whether there's a linear or more complex correlation.

Color-Coding or Symbol Variation: Each district can be uniquely colored or shaped, making outliers stand out immediately. Alternatively, a single color gradient could reflect a third variable (like population density).

Interactive Elements: Dropdowns allow users to switch between different socioeconomic variables (ex: unemployment rate, education level). Hover-tooltips display exact district names, crime counts, and the chosen socioeconomic metric, helping viewers delve deeper into any point of interest.

6. This view breaks down daily crime by hour, illustrating when offenses are most common throughout the week. A clear pattern emerges on Friday and Saturday nights, where reported incidents remain elevated later into the evening, often peaking around 10 PM to midnight, possibly reflecting nightlife activity and larger gatherings. During weekdays, there can be a smaller midday spike, particularly around lunchtime, correlating with high foot traffic in business districts. Such findings can be invaluable for scheduling police patrols and for community organizations looking to run awareness campaigns during the hours when incidents are most likely to occur.

- a. Chart Type (Multi-Line): A separate line for each day of the week helps compare daily patterns across the 24-hour cycle. Hourly intervals on the x-axis show how crime counts shift as the day progresses.

Color Scheme: Each day (Monday through Sunday) is assigned a unique color so the lines don't blend together. We typically use a light-to-dark or rainbow approach to differentiate days while keeping the palette readable.

Tooltip Interaction: Users can hover over any point on a line to see the specific hour and crime count for a given day, revealing if late-night Saturday spikes are significantly bigger than weekday evening peaks, for example.

Gridlines & Axis Labels: Including hour markers on the x-axis and a y-axis for crime counts ensures clarity. Subtle gridlines can help trace a specific hour's crime volume across different days.

• A summary for your findings, including what you have learned from the data visualization and what can be done in the future. (10 pts)

From our visual exploration of crime data in Boston, we observed clear seasonal fluctuations, with incidents tending to peak during the summer months and drop off in winter. Certain offenses, such as vandalism, theft, and assault, consistently appeared among the top reported

crimes year after year, suggesting specific areas where increased attention and resources might have a strong impact. Zooming in on the 2019–2023 window revealed that broader societal events, like the COVID-19 pandemic, could temporarily suppress crime rates, though incident counts gradually rebounded afterward. We also found that while lower-income districts typically experienced higher crime, some neighborhoods defied this pattern, implying that additional factors, such as local community initiatives or proximity to nightlife hubs, play a role in shaping outcomes. Going forward, we can enhance our insights by integrating other data points, like weather conditions, major public events, and more granular socioeconomic metrics, to pinpoint why certain neighborhoods see sudden spikes. Finally, applying predictive modeling techniques would allow policymakers and community organizations to forecast potential hotspots and craft more targeted, proactive interventions that address Boston’s evolving public safety needs.