****Simulation and Data Visualization Assignment****

Debaditya Bhattacharjee

K24087561

.

**ACADEMIC HONESTY AND INTEGRITY**

Students at King’s are part of an academic community that values trust, fairness and respect and actively encourages students to act with honesty and integrity. It is a College policy that students take responsibility for their work and comply with the university’s standards and requirements. Online proctoring / invigilation will not be used for our online assessments. By submitting their answers students will be confirming that the work submitted is completely their own. Misconduct regulations remain in place during this period and students can familiarize themselves with the procedures on the College website. **I agree to abide by the expectations as to my conduct, as described in the academic honesty and integrity statement.**

# Part 1: Analytics

# Exploratory Research Questions Proposed

Q1: Analyse the development of crimes over time. Are there detectable trends in types of crime (e.g., theft, assault) over the past ten years? This question focuses on the Crime dataset provided by the Metropolitan Police Service (MPS). Data ranges from February 2021 to January 2025. This will help detect long-term trends and anomalies over the years.

Q2: Which boroughs consistently report the highest crime rates for specific offences (e.g. theft, arson, burglary, etc.)? This question investigates geographical crime concentration. It investigates which boroughs are consistently affected by high crime levels in specific categories thereby helping identify hotspots. Secondly, it could be used to correlate with population density, income levels, or urban design. Enables targeted intervention (e.g., more patrols).

Q3: Are certain crime types likely to occur together in the same borough and month? This question analyses co-occurrence patterns of crime types. It attempts to identify combinations of offences that often happen together within the same space and time window. This is a crucial question as this reveals underlying social or behavioural correlations between crime types and supports building early-warning systems (if one offence spikes, another might follow).

# Required Data for Each Suggested Question

The provided London crime dataset contains monthly crime statistics reported across all London boroughs. Each row represents a specific crime type reported in a particular borough during a specific month. Key fields used in the analysis include:

* Month\_Year: This field combines the month and the year when the crime occurred. It was crucial for analysing how crimes change over time (Q1 and Q3).
* Area name ****/**** Borough\_SNT – These identify the borough where the crime was reported, used to compare crime distribution across locations (Q2 and Q3).
* Offence Group – A high-level category describing the type of crime (e.g., Theft, Robbery, Arson), used for filtering and grouping in all analyses.
* Count – The number of reported offences for a given category and borough in a given month. This was the core quantitative measure used for visual comparisons.

For this section the most appropriate attributes of the dataset were assessed for each question from part 1.a. Thus, data filtering was performed to focus only on raw offence counts, then grouped and aggregated values based on *time* (Month\_Year), *space* (Area name), and *category* (Offence Group). These operations allowed for *trend analysis* (Q1), *borough-wise comparison* (Q2), and *multi-dimensional pattern exploration (Q3),* all using just the original dataset without external sources.

Appropriateness:

* Q1: The dataset includes monthly crime counts by type and borough over multiple years — perfectly suitable for trend analysis over time.
* Q2: The dataset contains offence-level details for each borough and month. It can be easily grouped and filtered to find boroughs with high recurring offence rates.
* Q3: The structure of the dataset supports this question — crimes are already broken down by borough, offence group, and month, allowing co-occurrence analysis using aggregation techniques.

# Correlation

The dataset offers rich information along several dimensions — time, location, and crime type — which made it highly suitable for answering my research questions without needing additional data. The columns used were not only descriptive but also complemented each other, allowing meaningful grouping, filtering, and aggregation. Below is a summary of the key fields and why they were used:

|  |  |
| --- | --- |
| Attribute | **Role in Analysis** |
| Month\_Year | Enables time-series trend analysis (Q1, Q3) |
| Area name/Borough\_SNT | Enables spatial comparison across London (Q2, Q3) |
| Offence Group | Category-specific filtering. Grouping/Filtering (all) |
| Count | Incidents per offence, borough, and month. Core metric for analysis |

Together, these fields enabled:

* **Temporal Analysis** – Studying how crimes evolved month-by-month (Q1)
* **Spatial Patterns** – Identifying high-crime boroughs for specific offences (Q2)
* **Cross-type Relationships** – Comparing crime types in the same borough and month (Q3)

Since these columns already reflect **time**, **location**, **category**, and **count**, they provided the structure needed to perform a detailed analysis. The dataset's granularity and completeness helped answer all questions effectively using only internal attributes.

# PART 2: DESIGN AND DISCUSSION

For all visualizations, the following core design concepts were consistently applied to support readability, interpretation, and user interaction:

* **Color:** A sequential color palette was used to indicate varying intensity of values. Color was synchronized with axis ticks to reflect accurate mappings. A ***colorblind-friendly*** palette option was added to support accessibility, using perceptually distinct hues.
* **Readability:** Layouts were kept clean with high-contrast text, appropriate label sizing, and well-positioned legends. Visual elements were optimized to be easily interpretable, even by novice users with no data analysis background.
* **Interactivity:** Dark-mode, dropdown menus, tooltips, and animated transitions were used to enhance user experience. These allow users to explore data dynamically without overwhelming static elements.
* **Consistency & Minimalism:** All charts were generated keeping in mind the basics and core ideas of data visualization—grid alignment, font sizes, color scales, and spacing to ensure a cohesive look consistently across visualizations.

# Question 1: Analyze the development of crimes over time.

*Visualization 1: Multi-line Time Series Chart*

This visualization presents a temporal encoding of various crime types using a multi-line chart. Each line represents an offence group, color-coded using a custom palette to ensure perceptual separability and avoiding **JND** (just noticeable difference) as much as possible. The x-axis encodes time in monthly intervals, while the y-axis encodes the number of reported offences. This allowed month-by-month fluctuations in total offences to be visualized clearly across a continuous time axis. This format facilitates easy comparison between crime categories, helping detect overall trends, seasonal fluctuations, and outlier events (e.g., policy changes, lockdown effects). This visualization supports comparative pattern discovery across types. The use of continuous temporal encoding combined with a rich color scheme helps even novice users to visually extract both short-term spikes and long-term trends.

A graph of different colored lines

AI-generated content may be incorrect.

# Question 2: Which boroughs report the highest crime rates for specific offences?

*Visualization 2: Horizontal Bar Chart*

This horizontal bar chart ranks the top 10 boroughs by average offence count for a selected offence group (e.g., ARSON, THEFT, BURGLARY, etc.). It uses the x-axis to represent offence count and the y-axis for boroughs. The horizontal orientation improves label readability, especially for long borough names. Color is used to indicate severity or normalized ranking. This chart excels in simplicity and clarity. *It focuses attention on high-frequency areas and can be paired with interactivity (dropdowns or sliders) for exploring multiple crime types*. Compared to map visualizations, this chart offers exact numerical comparison and is less dependent on geographic context, thus improving accessibility and reducing clutter. This visualization provides a basis for identifying persistent crime hotspots.

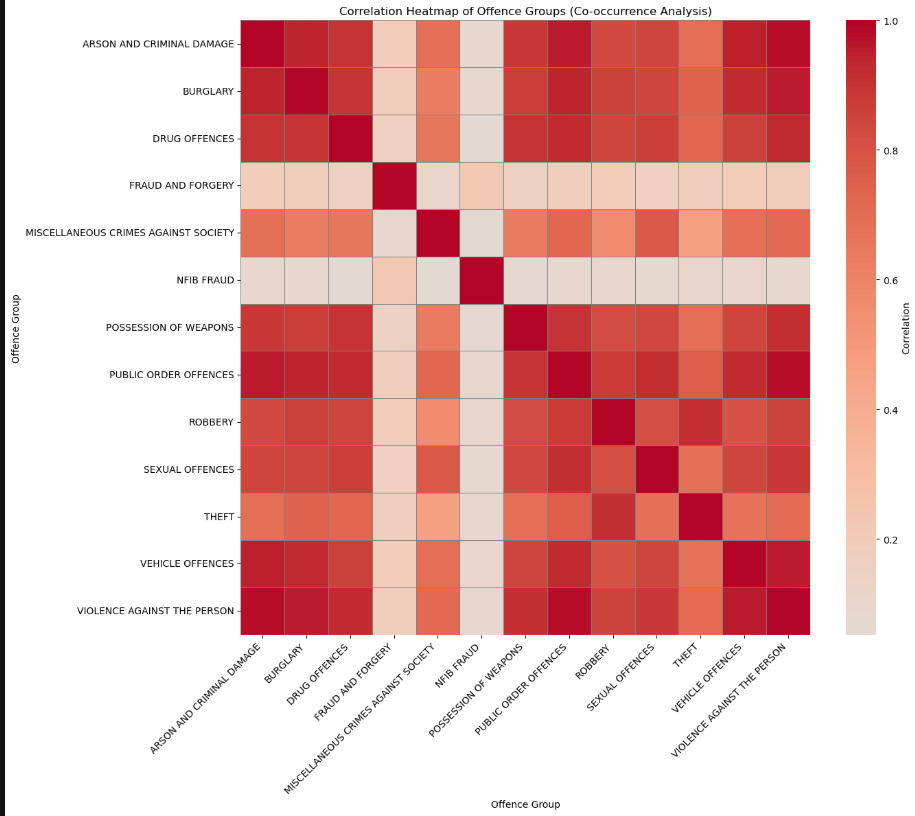
A graph of different colored bars

AI-generated content may be incorrect.

# Question 3: Are certain crime types more likely to occur together in the same borough and month?

*Visualization 3 – Correlation Heat map*

This visualization explores the relationship between offence groups using a correlation heat-map. Data is transformed such that each (borough, month) is a unit of observation, and each crime type becomes a variable. Pairwise correlations are calculated and displayed in a square matrix where both axes represent offence groups. Color encodes the strength and direction of correlation (e.g., lighter shade = negative, darker shade = positive). This helps identify which crime types tend to spike together — offering insight into potential *social* or *behavioral* links. It surpasses side-by-side charts or tables by summarizing multiple relationships compactly. The design facilitates pattern discovery for both technical and non-technical viewers. Labels, color scales, and a reference key ensure interpretability. This visualization provides a foundation for future co-occurrence modeling or policing strategies.



# Part 3: Implementation

The visualization implemented corresponds to research Question 2, which explores which boroughs consistently report the highest crime rates for specific offences in London.

# Data Processing

The data used is exclusively sourced from the ***London Crime Dataset (CSV)*** provided in the coursework materials. The dataset contains monthly records of reported crimes across all London boroughs, categorized by offence group and accompanied by numerical counts.

**To prepare the data:**

* Only records where the ‘*Measure*’ field equals "*Offences*" were retained to ensure the focus remained on actual reported crimes.
* The data was grouped by ‘*Offence Group*’and‘*Area name*’, and the‘*Count*’ values were aggregated to compute the total number of offences per borough for each offence type.
* For each offence type, the **Top 10 boroughs** were selected based on the highest total counts.
* Data was dynamically mapped to a *linear color scale* reflecting offence intensity, aligned with the x-axis scale for consistency and accessibility.

# Web Implementation in D3

The visualization was built using *D3.js v7*, embedded in a standalone HTML file. It features:

* A *horizontal bar chart* with boroughs as categories and offence counts as lengths.
* An *interactive dropdown menu* to select the offence type.
* A *synchronized legend* and color scale for readability.
* Support for *dark mode* and a *colorblind-friendly* palette switch, improving accessibility.
* *Tooltips* and *animated transitions* to enhance interactivity and engagement.

This interactive design enables users to explore borough-level crime distributions by offence type and supports exploratory analysis even for non-technical audiences.

**Acknowledgements**

The author acknowledges the generous support and help from the lecturer and teaching assistants at King’s College London, Department of Informatics

**References**

1. Metropolitan Police Service (MPS) – London Crime Data, accessed via KEATS, 2025.
2. Bostock, M. (2012). Data-Driven Documents (D3.js). Retrieved from: https://d3js.org
3. ColorBrewer2.0 – Color advice for cartography, Cynthia A. Brewer.
4. Ware, C. (2012). Information Visualization: Perception for Design.
5. Tableau Public. Best Practices for Bar Charts.
6. Kant, I. Critica della ragion pura, p.1. Ed. Laterza, 1949. (for mark-channels theory)
7. Borgo, R. (2024). Lecture Slides & Resources on Visual Encoding, King’s College London, MSc DV Module.
8. Vazquez Gonzalez, C. (2022). Simulation and Data Visualisation
9. AM Charts: https://www.amcharts.com/javascript-charts
10. Holtz, Y. (2018). *The R-Graph Gallery*: https://r-graph-gallery.com
11. Anonymous. Line Charts, Bar Charts, and Heat Map Charts. Anychart Gallery: https://www.anychart.com/products/anychart/galleryC
12. Curtis, H. (2021). Simulation and Data Visualisation