

0.1 Datasets

Statistical GIS Boundary Files for London: This file, offered by Greater London Authority (GLA) and available at London Datastore, provided a range of key GIS boundary files for ESRI and Map Info covering Greater London, including shape files of London Wards and London Boroughs. This data set includes Name: Name of ward, District: Name of borough, Geometry: Pair of longitude and latitude.

Crime levels by borough: This data set, offered by Metropolitan Police Service and available at London Datastore, counted the number of crimes at three different geographic levels of London (borough, ward, LSOA) per month, according to crime type from 2010 to 2021. This data set contain variables like LookUp_BoroughName: Name of Borough and X201004: count of crimes during April 2010.

Population by Ward and Borough: This file, offered by Greater London Authority (GLA) and available at London Datastore, provided population for 2001 to 2050 for London wards and boroughs. This data set includes variables like Name: Borough name, Year: population of the year, Population: population of the borough at that year.

Trees: The trees dataset, available as a built-in dataset in R, offers measurements from 31 felled black cherry trees and provides insights into the relationship between a tree's girth, its height, and the volume of timber it can produce. The dataset contains 3 variables: Girth: The diameter of the tree, Height: The height of the tree, Volume: The volume of timber that the tree can produce.

0.2 Network Graphs

Definition and Utility: Network graphs, often referred to as graphs or networks, are a powerful data visualisation method used to depict relationships between entities. These entities, known as nodes, are interconnected by edges or links, which represent relationships, connections, or interactions. Network graphs find extensive utility in various fields, such as social network analysis, transportation systems, and even biological networks like protein-protein interactions. They excel at revealing complex dependencies and structures, making them a critical tool for understanding relational data.

0.2.1 The Mathematics behind Network Graphs:

Constructing network graphs involves several mathematical intricacies. Here we present just a few of the many concepts that play a role in the creation of such graphs:

1. **Nodes and Edges:** Mathematically, a network graph, G , is defined as $G = (V, E)$, where V represents the set of nodes and E represents the set of edges connecting these nodes.
2. **Node Degree:** The degree of a node is the number of edges connected to it. In a directed graph, nodes can have both in-degrees and out-degrees.
3. **Centrality Measures:** Centrality metrics like degree centrality, betweenness centrality, and closeness centrality provide insights into the relative importance or influence of nodes within a network.

4. **Graph Metrics:** Graph theory concepts like shortest paths, connected components, and clustering coefficients are used to analyse the network's structure.

Formulas used in Network Graphs:

1. **Degree of a Node (Undirected Graph):**

$$\text{Degree}(v) = \sum_{w \in V} A(v, w)$$

where $A(v, w)$ is the adjacency matrix element, indicating whether there is a connection between nodes v and w .

2. **Degree of a Node (Directed Graph):**

$$\text{In-Degree}(v) = \sum_{w \in V} A(w, v)$$

$$\text{Out-Degree}(v) = \sum_{w \in V} A(v, w)$$

3. **Betweenness Centrality (for unweighted graphs):**

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the number of shortest paths from node s to t , and $\sigma_{st}(v)$ is the number of those paths passing through node v .

0.2.2 Network Graphs in Practice

```
# Plot the graph
#plot(lesmis_graph, layout = layout, vertex.label.cex = 0.7, main = "Character Interactions in Les Misérables")
```

d

0.3 Geographic Maps and Spatial Data Visualisation

A **geographical maps** is a visual representation of an area—a symbolic depiction highlighting relationships between elements of that space, such as objects, regions, or themes. Maps have been used for centuries to navigate and explore the world, and they play a crucial role in understanding our environment, both locally and globally. These maps serve as canvases on which spatial data is painted, allowing for a visual comprehension of information that might otherwise remain abstract.

Before moving on to the spatial data visualisation. It is essential to understand how we map the the earth on a plane. The Earth, a three-dimensional spheroid, can be transformed on a plane through map projections. Each projection offers a different way to "flatten" the Earth, and as a result, each has its strengths and distortions. For instance, the Mercator projection preserves angles but distorts areas as you move towards the poles. Beyond projections, coordinate systems, like the commonly used latitude and longitude, provide a standardized way to pinpoint any location on Earth.

- **Latitude** measures the angle between a point on the Earth's surface and the equator, moving north or south. And latitude values range from -90 degrees (South Pole) to +90 degrees (North Pole). The equator, which divides the Earth into the Northern and Southern Hemispheres, is at 0 degree latitude.
- **Longitude** measures the angle between a point on the Earth's surface and the prime meridian, moving east or west. And longitude values range from -180 degrees to +180 degrees. The prime meridian, which is at 0 degree longitude, runs from the North Pole through Greenwich, England, to the South Pole. It divides the Earth into the Eastern and Western Hemispheres.

Together, lines of longitude and latitude create a grid system over the Earth's surface. By providing both a latitude and longitude value, one can specify an exact location on the Earth's surface. For example, the coordinates (0° N, 0° E) would indicate the intersection of the equator and the prime meridian, located in the Gulf of Guinea off the west coast of Africa.

Spatial data visualisation

Spatial data visualisation are powerful tools that transform raw, often complex datasets into visual representations, revealing patterns, relationships, and insights rooted in location. At their core, maps provide a spatial context, allowing us to see the world's intricate web of interconnectedness. Today, with the surge in big data and advanced visualisation tools, spatial data visualisation is not just about presenting information but also about telling compelling stories, guiding decision-making, and predicting future trends based on geographical patterns.

Here, we are going to construct a geographical map of the Greater London and show each ward.

```
# Plot London map by ward
ggplot(data = london_boroughs) + geom_sf(fill = "lightblue", color = "black") +
  theme_minimal() + labs(y="Latitude", x="Longitude")
```

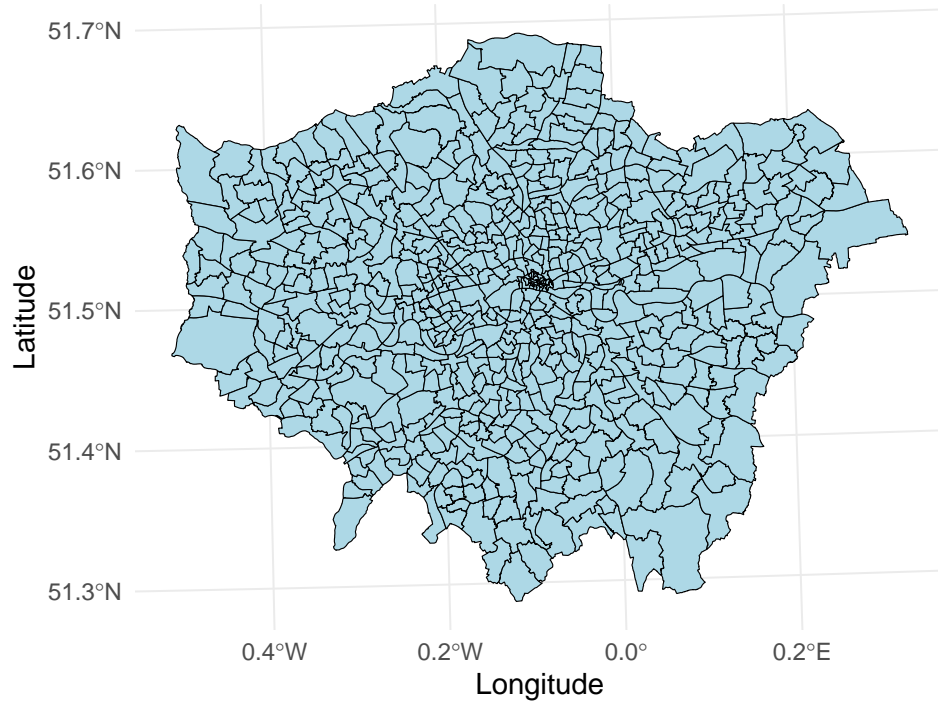


Figure 1: London Wards Map

From Figure 1, we are introduced to a detailed layout of all wards within Greater London, pinpointed by their geographical coordinates. While this map provides a clear depiction of location and boundaries, it offers limited insight into the dynamics of crime distribution.

Hence, we will use a more informative visualisation: a choropleth map showcasing crime rates by borough. Here, crime rates are calculated by dividing the total crime count in each borough by its respective population as of 2020. This per capita approach normalizes the data, ensuring comparability across boroughs with different population sizes.

```
# Plot the Crime rate in London by boroughs
ggplot(data=aggregated_data) + geom_sf(aes(fill=Crime_rate)) +
  geom_sf_text(aes(label = DISTRICT, geometry = centroid),
    size = 2.5, check_overlap = TRUE) +
  scale_fill_gradient(low="Green", high="red") + theme_minimal() +
  labs(fill="Crime rate")
```

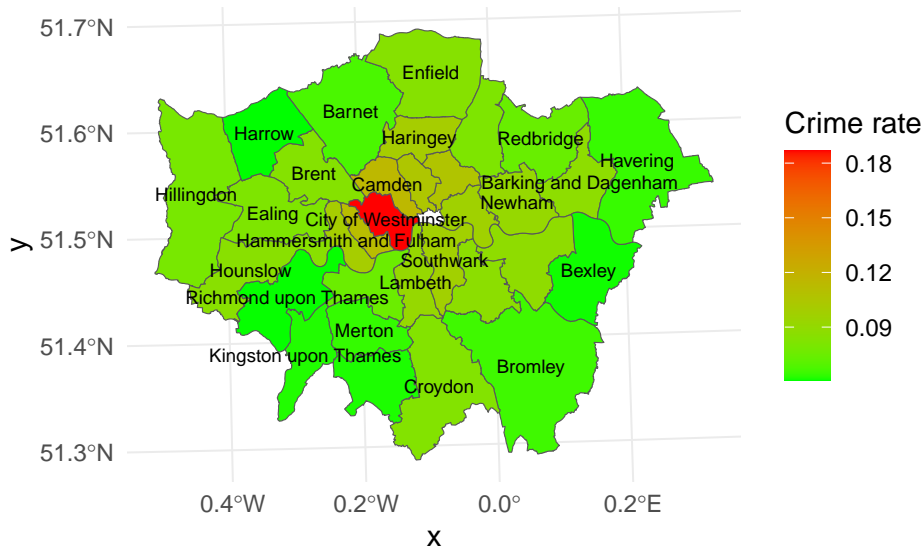


Figure 2: Crime rate by boroughs in London in 2020

From Figure 2, Westminster stands out, labeled in red, indicating a higher crime rate. This visual marker highlights Westminster as an area of particular concern regarding criminal activity. This could be attributed to factors like its status as a central, densely populated area with significant tourist traffic and commercial activity, which often correlate with higher crime rates.

In contrast, the map reveals a trend where rural or less densely populated areas tend to be safer, as evidenced by their lighter coloration. These areas typically experience lower crime rates, possibly due to factors such as smaller populations, less anonymity for potential offenders, and different socio-economic dynamics compared to urban centers.

The choropleth map of London's crime rates elucidates key spatial patterns in crime distribution. It highlights the variance in crime rates from urban centers to rural areas, accentuating the higher crime rates in central, densely populated areas like Westminster, depicted in red, compared to the lighter shades marking safer, rural boroughs. This visual tool facilitates an understanding of how urbanization, socio-economic status, and population density impact crime rates across the boroughs.

0.4 3D and Interactive Visualisations

ggplot2 is one of the most popular data visualisation libraries in R, but it is primarily designed for 2D data visualisation. Directly creating 3D views with ggplot2 can be challenging.

R provides several packages for 3D visualisation, such as rgl, plot3D, rayshader, and others, which are specifically designed for three-dimensional data. These packages offer the capability to create 3D scatter plots, surface plots, heat maps, contour maps, and more.

rgl: This is one of the most popular R packages for creating interactive 3D charts. It supports various types of 3D graphics including points, lines, and surfaces, and allows users to interactively rotate, zoom, and pan the view.

scatterplot3d: This package provides a function to create 3D scatter plots. It does not support interactive manipulation, but the generated graphics are well-suited for display in static reports.

3D data visualisation is an approach that employs three-dimensional graphics to represent complex data structures, allowing for an immersive exploration of information. Unlike traditional 2D visualisations (like bar graphs or line charts), 3D visualisations can convey an additional dimension of data, making them particularly valuable in specific contexts.

Our first example will be a scatter plot. We can use scatterplot3d package to help us for data visualisation.

```
# Generate colors based on the Volume variable
colors <- colorRampPalette(c("blue", "red"))(length(unique(trees$Volume)))
color_assign <- colors[as.numeric(as.factor(trees$Volume))]
# Create 3d scatter plot with colors
scatterplot3d(trees$Girth, trees$Height, trees$Volume,
              color=color_assign,
              main="3D Scatterplot of trees data",
              xlab="Girth (inches)",
              ylab="Height (ft)",
              zlab="Volume (cubic ft)")
```

3D Scatterplot of trees data

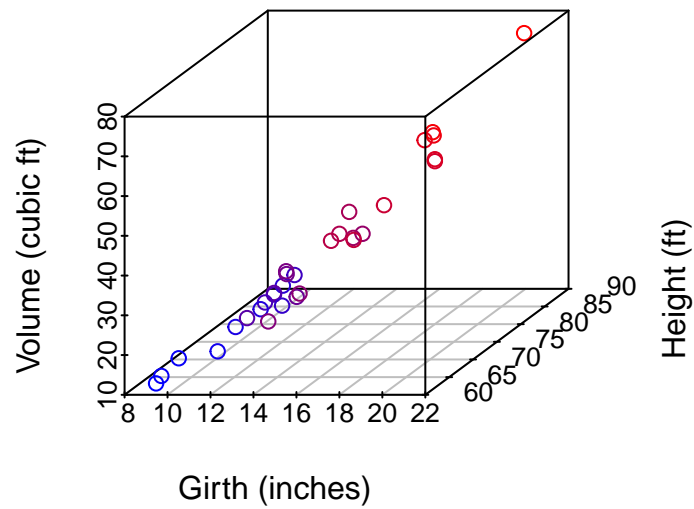


Figure 3: 3d scatter plot

1 Trees Dataset in R

XXX

1.1 Advanced Visualisation Techniques

XXX

2 Practical Implementations

XXX

3 Case Studies

3.1 Market Analysis Dashboards

XXX

3.2 Healthcare Data Visualisation

XXX

4 State-of-the-Art Approaches

XXX

5 Conclusion

XXX