

House prices data analysis - 7CCSMSDV

Kavana Undiganalu Rangaswamy
Simulation and Data Visualization Assignment
King's College London
K23106657@kcl.ac.uk

Abstract—This report examines the dynamics of house prices and exploring underlying trends and factors influencing the real estate market in London. We use different data visualization methods to uncover patterns and insights into how prices fluctuate. By using statistical tools and visualization frameworks, we aim to pinpoint key trends like seasonal changes, price elasticity, and the influence of economic factors on housing costs. Our goal is to provide valuable information to stakeholders such as homebuyers, investors, and policy makers to help them make informed decisions in the real estate market.

I. ANALYTICS

A. Exploratory Research Questions Proposed

Q1. Analyze the development of house prices over time. Are there any detectable trends? How have house prices in London changed from 1992 to the most recent year in your dataset? This question will allow you to analyze long-term trends in house prices, identify periods of significant growth or decline, and potentially relate these trends to historical economic events or policy changes.

Q2. Impact of House Characteristics on Price: How do the characteristics of a house (such as type, size in sq ft, number of bedrooms, bathrooms, and receptions) affect its price? This could involve comparing different types of properties (e.g., flats vs. houses) and analyzing how physical attributes correlate with price variations across different areas.

Q3. How does the average area of properties in different Central London locations correlate with their average price, and are there any notable outliers where property size does not reflect expected price trends?

B. Explain what type of data could be used to answer each of the questions proposed. Assess the appropriateness of each dataset(s) you would potentially be using to answer the questions.

1) *Q1: Development of House Prices Over Time:* This analysis requires longitudinal data that tracks house prices over a significant period. The data should have timestamps or yearly records to show trends and patterns over time. The 'LondonBoroughs' dataset offers annual house price data for many years in a sequence. This dataset is perfect for our analysis because it helps us identify long-term trends, cycles, and any unusual changes in house prices. The thoroughness and chronological accuracy of this dataset allow us to analyze how house prices evolve in response to economic and policy shifts.

2) *Q2: Impact of House Characteristics on Price:* We need detailed information about each property, such as the type of house (e.g. flat, detached), square footage, and the number of bedrooms, bathrooms, and reception rooms. This data is essential for understanding how different factors affect property prices. The 'London' dataset provides in-depth details about every property listed, including type, size, and number of rooms. The depth of this dataset allows for a thorough analysis of how different physical attributes can influence the market value of properties in London, making it perfect for investigating this research question.

3) *Q3: Correlation of Property Size and Price in Central London:* We need data on property sizes in terms of Area and prices in Central London to analyze correlations and outliers. The focus on detailed property attributes and geographic specificity makes 'London' ideal for this investigation. It enables us to analyze the relationship between property size and price in Central London areas, uncovering pricing anomalies and trends.

C. Explain if and how datasets are or could be correlated.

1) Datasets Overview::

- London Dataset - This dataset provides longitudinal data on house prices across London, capturing trends over decades.
- London Boroughs Dataset - This dataset offers detailed property attributes such as type, size, and number of rooms, alongside pricing information for each borough.

2) Correlations::

- Temporal and Spatial Price Trends: The dataset for London, which tracks changes in prices over time, can be compared with the dataset for London Boroughs to study how specific property features have historically affected pricing trends in different boroughs. For example, one can analyze how an increase in prices over time within a certain borough is connected to changes in property types and sizes in that area.
- Property Attributes and Pricing: The connection between the physical characteristics of properties (from the LondonBoroughs dataset) and their pricing trends (from the London dataset) can be explored. This analysis can help determine which features (such as number of bedrooms or square footage) have the most significant impact.

II. DESIGN AND PROTOTYPING

A. *Propose and design visualizations that would answer the research questions proposed in Part 1*

Heat Maps of House Prices by London Borough over Time The visualization features heat maps divided into five London regions: Central, East, West, North, and South. These heat maps display the average house prices across various boroughs from 1995 to 2017. The data is reorganized using Tableau's 'PIVOT' function to consolidate all the years into one column and the corresponding values into a 'House Prices' column. Boroughs in London are categorized based on their location to make it easier to compare trends over time. To enhance these charts, a Compound Annual Growth Rate (CAGR) calculation has been added for each borough. This calculation offers a clear and standardized way to measure growth from 1995 to 2017. In addition to the Heatmap, a TreeMap is also utilized since it allows the viewer to quickly discern the relative house prices across the boroughs due to its area-based representation. This visual representation can efficiently communicate the distribution of house prices across London, making it straightforward to compare and contrast different areas at a glance.

Design Rationale Heat maps are selected for their ability to present complicated datasets in a clear visual format that is easy to understand. They excel at illustrating data density through color and size differences, making it easy to identify patterns and anomalies. The use of heat maps directly relates to Research Question 1 by allowing viewers to quickly see changes in house prices over time. In order to make the visualization accessible to a larger audience, including those with color vision impairments, a colorblind-friendly palette was implemented. This careful choice of colors ensures that the information is effectively communicated to all users.

The heat maps enhance the analysis by showing historical trends, going beyond just yearly snapshots. This helps understand how the housing market responds to economic changes over time. Additionally, interactive features like tooltips that show house prices and CAGR when hovering over a cell give users a more engaging experience. Users can explore the data in a personalized way, answering their specific questions and fostering a deeper connection with the housing market's history.

A TreeMap alongside the heatmap is included to show the difference in house prices in London boroughs from 1995 to 2017. The TreeMap is used because it can effectively display hierarchical data through varying sizes of rectangles, making it easy to compare property values at a glance. Larger rectangles indicate higher-priced boroughs such as Kensington and Chelsea, while smaller rectangles represent more affordable areas like Bexley and Sutton. Contrasting the 2017 TreeMap with the one from 1995 provides a clear picture of how the market has changed, highlighting which boroughs have seen the biggest changes in house prices.

Scatterplot of House Price vs. Area for Different House Types A scatterplot is used for analyzing the impact of house

characteristics on price, particularly focusing on the area of the property in square feet versus the price for different house types in London. Different types of houses such as duplexes, flats/apartments, houses, mews, new developments, penthouses, and studios are represented using a colorblind-friendly palette. Each type of house has a trend line showing the connection between size and price, making it easy to see the correlation visually.

Design Rationale The scatterplot visualization aims to answer Research Question 2: How does a house's characteristics impact its price? The design is based on visual analytics principles, examining the relationship between property size (measured in square feet) and the price of different types of houses in London.

Each property is denoted by a circle whose position is determined by its area and price, providing an intuitive depiction of the market. This simple yet intuitive visual helps us easily understand how size impacts cost. By adding trend lines for different types of houses, we can quickly see the overall pricing trends based on size and type. This information can guide buyers and investors in making decisions, highlighting certain property types that may have a premium regardless of size. Filters for bedrooms, bathrooms, and receptions allow us to conduct focused analysis, like comparing prices of properties with similar amenities. The interactive elements encourage user engagement, turning the visualization into an exploratory tool for personalized data interrogation.

The size encoding in the scatter plot points increases as the number of bedrooms increases, which enhances the visualization by showing the property scale beyond just its area. This choice not only adds more information to the plot but also maintains clarity, following the principle of maximizing data display. Using colors that are friendly to colorblind individuals makes the visualization more accessible to a wider audience, enabling those with color vision impairments to understand the data. Each type of house is assigned a distinct color, making it easier to distinguish and compare market trends across different property categories. In summary, from this visualization, we can observe that some house types have a steeper trend line, indicating a higher price per square foot, while others have a flatter trend line, suggesting less sensitivity to size in their pricing. This could imply a premium on certain property types regardless of size, possibly reflecting market preferences or scarcity.

Combined Bar and Line Chart for Property Area and Price in Central London The Combined Bar and Line Chart for Research Question 3 depicts the relationship between the average area of properties in square feet (displayed as bars) and their average price (£, shown as a line) across various Central London locations. The regions of Central London have been selectively filtered to focus the analysis on this specific geographic area. Each bar represents the average area of properties within a particular region, and the line overlaying the bars represents the average price of properties in those same regions.

Design Rationale This combined bar and line chart is

an effective method for presenting two related dimensions of data. By plotting average area and price for each Central London region, the visualization allows for an immediate visual comparison between two potentially correlated factors: property size and value.

The reason for using bars to represent area and a line to represent price is intentional; it takes advantage of the bar's natural ability to show quantity, making it intuitive for viewers to connect bar height with size. The line graph, known for illustrating trends and changes, effectively displays price fluctuations, making it easy to see how prices change in relation to property size. By incorporating dual axes, different scales can be compared, recognizing that while area and price are connected, they are measured in different units and do not have a direct one-to-one relationship. The distinct visual elements – solid bars for area and a continuous line for price – help avoid any confusion between the two sets of data.

The visualization helps us understand if larger average property areas are linked to higher prices. It also helps us spot outliers, areas where the expected correlation doesn't apply, suggesting other factors might be at play. This visualization improves on current models by combining these metrics into one visual display, giving a better insight into real estate trends in Central London. It's a valuable tool for investors, homebuyers, and policy makers, revealing the complexities of property values. The inclusion of Central London regions caters to a segment known for its economic significance, providing insights specific to an area where even slight variations in space can have a substantial impact on price. This visualization also prompts viewers to consider the influence of non-size-related factors on property prices, such as location desirability, historical value, and access to amenities. Regions where prices are high despite smaller average property areas might indicate a premium placed on location over space, an insight valuable for market analysis.

III. IMPLEMENTATION

In addressing Research Question 02, a sophisticated interactive scatterplot visualization has been constructed, employing D3.js to scrutinize the influence of house characteristics on market price. The dataset, hailing from a curated online JSON GitHub repository, providing a thorough look at the varied London real estate market.

Visualization Layout and Data Processing: The scatterplot graphically represents properties, with axes calibrated to denote 'Price' against 'Area in sq ft' to effectively represent individual data points that correspond to various properties. Notably, the points are color-coded using a palette designed to be color-blind friendly, ensuring the scatterplot is accessible and interpretable by a wider audience, including those with color vision deficiencies. The attributes such as the number of bathrooms, bedrooms, and receptions are extracted to populate corresponding filter controls. Different house types are easily distinguishable through this scheme, enhancing the user's ability to discern and analyze the data.

User Interaction: The visualization is designed with the following interactive elements to engage users effectively:

- **Dynamic Filtering:** Utilizing dropdown selectors, users can filter the dataset based on specific characteristics, such as the number of bathrooms, bedrooms, receptions, or house type. This interactive filtering causes the scatterplot to update in real-time, allowing users to dissect the dataset and observe how selected house features influence pricing.
- **Point Highlighting and De-emphasis:** A click on any dot in the scatterplot accentuates it by increasing its opacity, simultaneously dimming other points to a lower opacity. This highlight can be toggled off by re-clicking the selected point, thus allowing users to focus on and compare properties of interest with ease.
- **Informational Tooltips:** When hovering over a data point, a tooltip emerges, presenting detailed information such as the property's name, area, and price. This feature provides an immediate, in-depth view of data for each property without introducing visual clutter.

Description: This interactive scatterplot is an investigative tool that elucidates the correlation between house prices and their defining features in the London property market. Users are afforded a level of interaction that enables the direct manipulation of the visualization to reflect varying scenarios and housing configurations. The legend assists users in discerning the house types at a glance, while tooltips supply a granular level of detail per property, enhancing the data narrative. The careful integration of interactive elements empowers users to engage with the visualization not just as a static representation, but as a dynamic model for inquiry. The user-driven exploration afforded by the visualization facilitates a nuanced understanding of how specific house characteristics—such as type, size, and number of rooms—bear upon property values. This serves as an invaluable aid in real estate assessment, portfolio management, and market trend analysis.

IV. ACCESS TO IMPLIMENTATION:

- Github <https://github.com/KavanaRangaswamy/DataVisualization>
- CodePen <https://codepen.io/KavanaRangaswamy/pen/OJGdYxw>

V. REFERENCES

Datasets can be accessed through the following links::

- **London Dataset:** <https://www.kaggle.com/datasets/arnavkulkarni/housing-prices-in-london>
- **London Boroughs Dataset:** <https://data.london.gov.uk/dataset/average-house-prices>

A. Journals

- A.Glivinska.The25bestdatavisualizationsof2020[examples],2021.
- <https://clauswilke.com/dataviz/time-series.html>
- <https://nycdatascience.com/blog/student-works/key-insights-ames-iowa-housing-data-multiple-factors-behind-house-price/>
- <https://www.hindawi.com/journals/>
- <https://www.mdpi.com/2073-445X/12/4/740>
- <https://www.mdpi.com/2073-445X/11/3/334>
- N.LLC.3dbarcharts,2021
- U.Unknown.Uksummary,2022.
- <https://data.un.org/>

ACKNOWLEDGMENT

The Author expresses gratitude to Prof.Rita Borgo for providing the opportunity to write on the subject of data visualization in the context of development of house prices over time.

Appendix

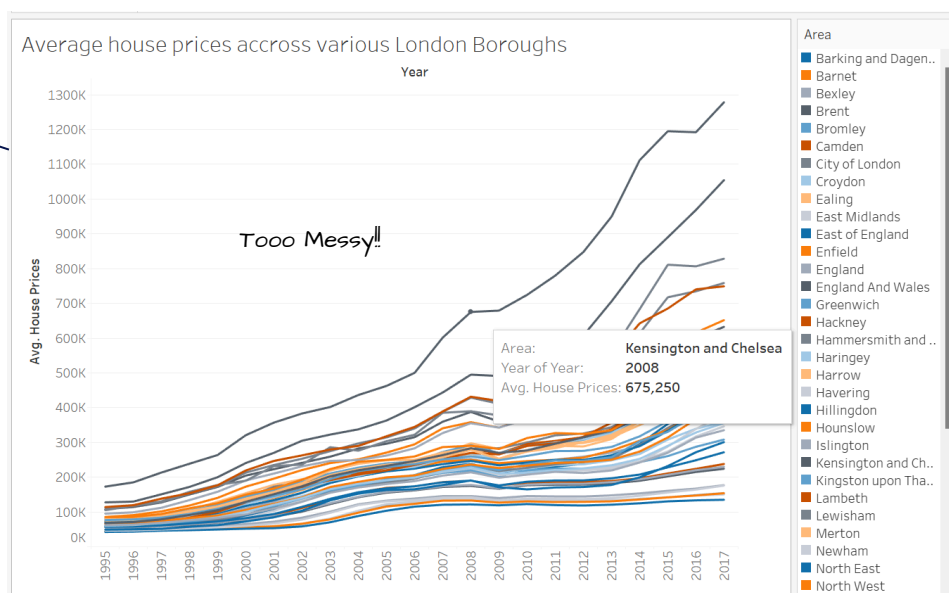
FRIENDLY NOTE: All the visualizations are made color-blind friendly ensuring the visualizations are accessible and interpretable by wider audience

01. Analysing Development of House Prices Over Time across various London Boroughs



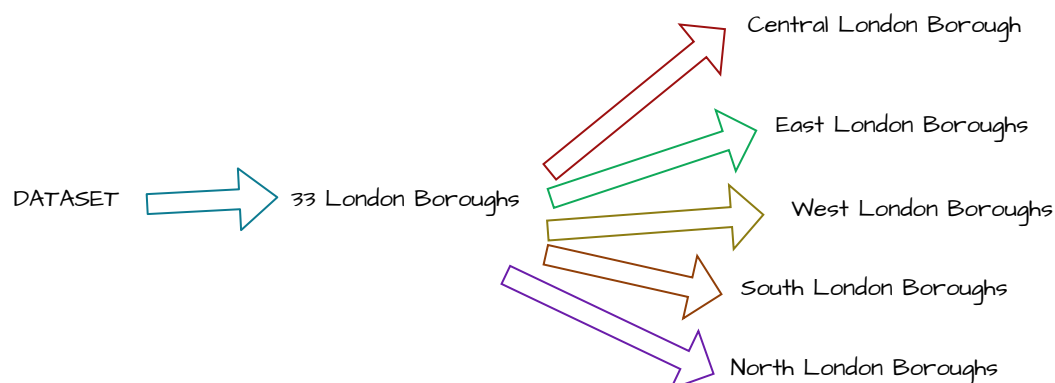
Too difficult to visualize!

All visualizations for this question are created using Tableau



Lets break down the dataset, into smaller subsets and visualize for better understanding

DATA PRE-PROCESSING:



Central London Boroughs House Prices for the period 1995-2017

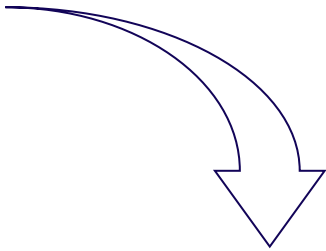
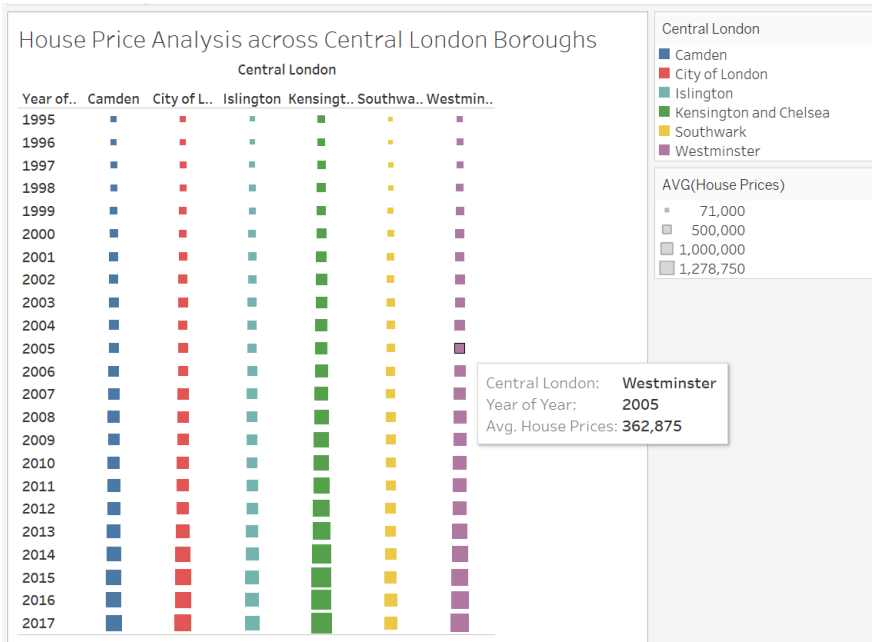
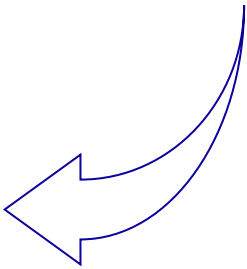
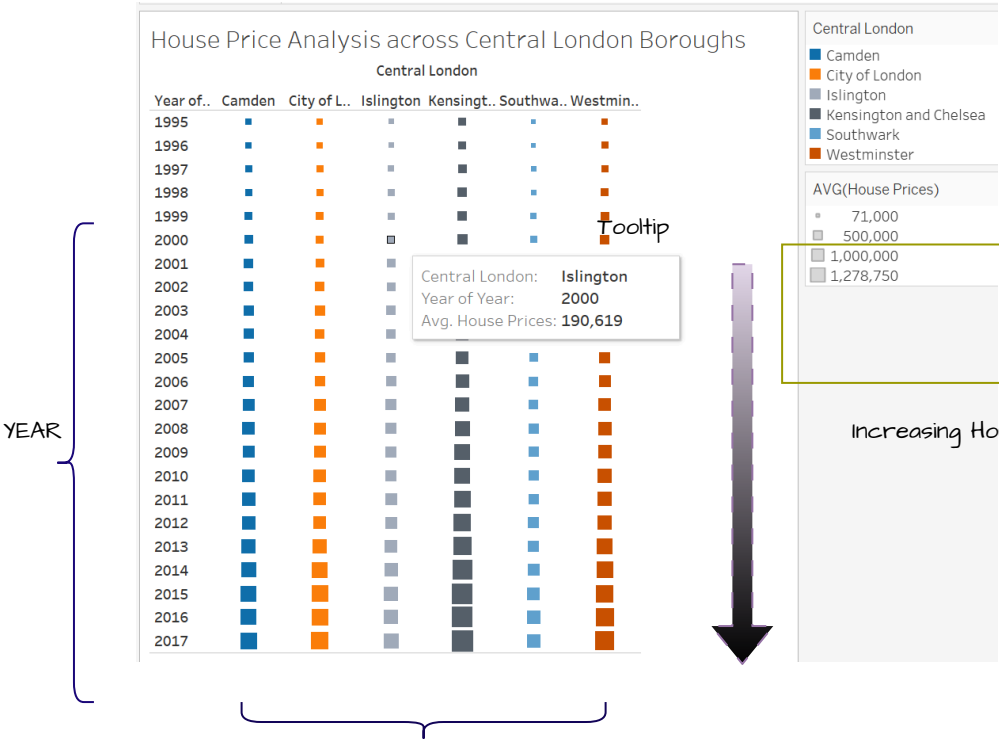
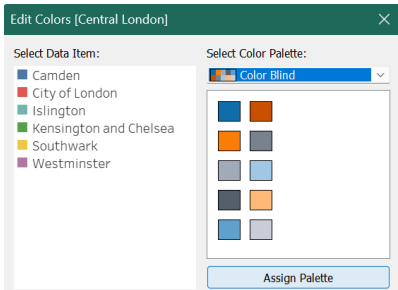


Tableau builtin Colorblind Palette



Average House Prices in terms of size of each square

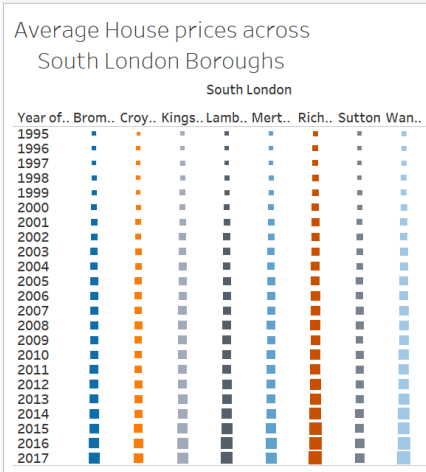
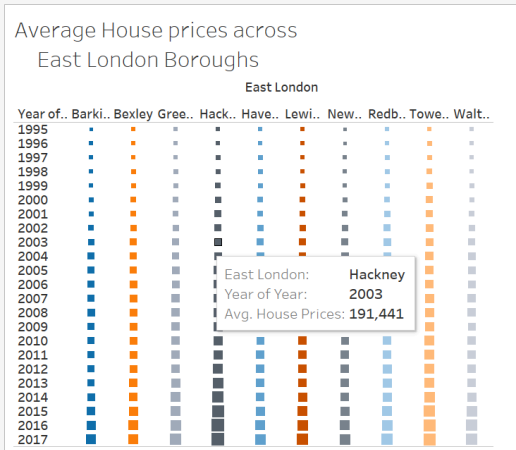
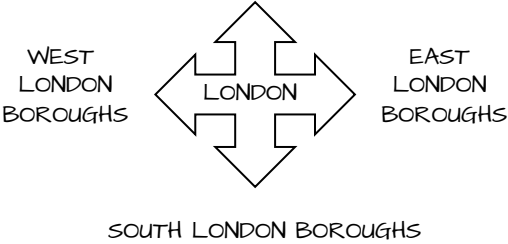
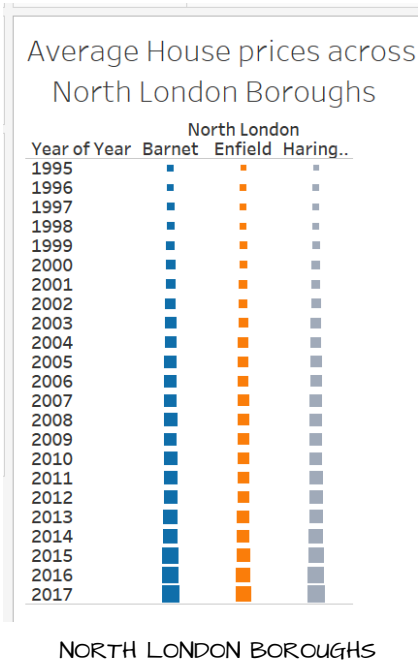
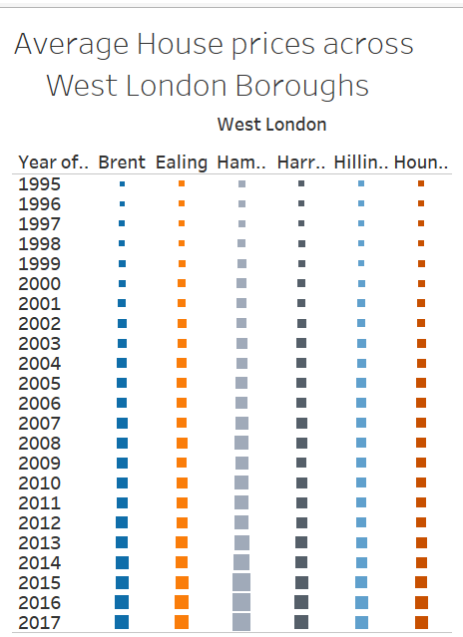
Increasing House Prices over the years

Increasing trend over the years

- Type: Categorical Heatmap
- Size: Square marks, increase with increase in price
- Colour: Colorblind-friendly palette for distinct categories
- Scale: Ordinal scale for years, nominal for boroughs, and quantitative scale for average house prices
- Mark: Square marks representing data points
- Channels: Color saturation for average house prices, spatial positioning for years and boroughs

Increasing Trends across all the Boroughs:

Similar trends across all other subsets:
North London Boroughs
South London Boroughs
East London Boroughs
West London Boroughs

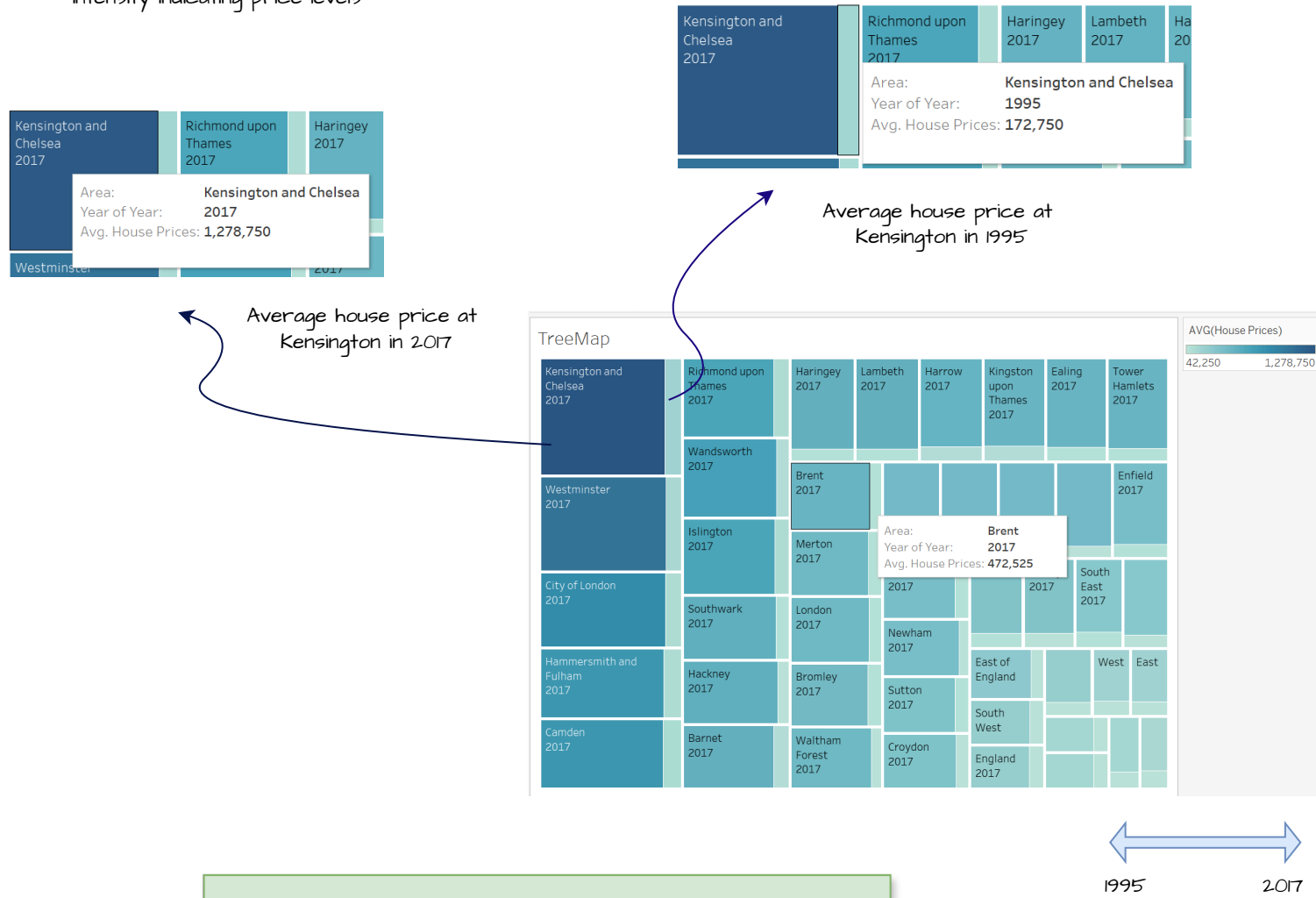


Average house prices across London INCREASE over the time.

Design improvement: To showcase all the Boroughs in a single visualization for wider range of analysis,

How much change is too much change?

TreeMap to show changes compares average house prices in various regions for the years 1995 and 2017, with each box divided to represent the two different years and color intensity indicating price levels



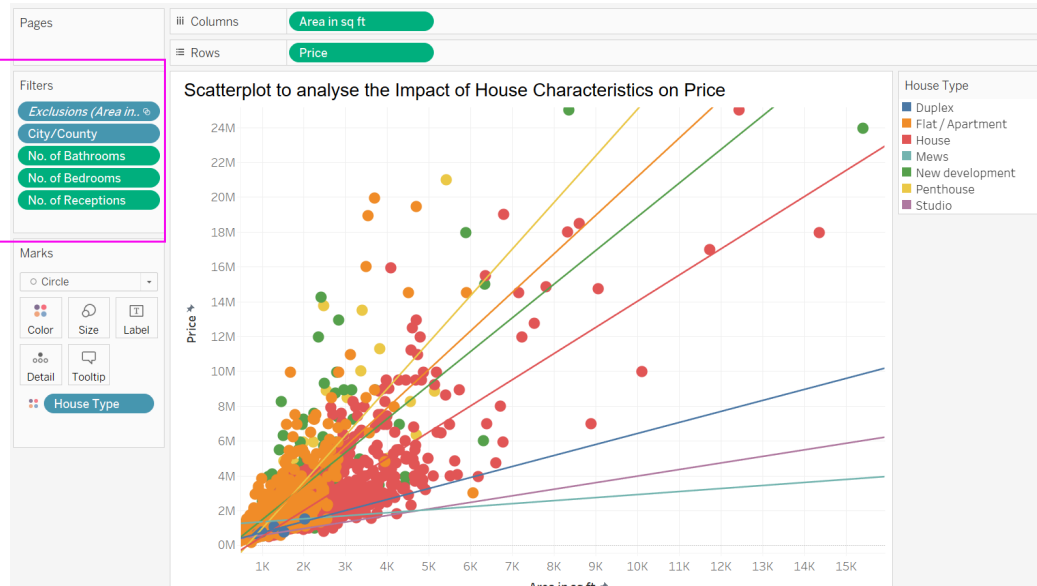
The TreeMap analysis reveals a pronounced escalation in London's housing prices from 1995 to 2017, accentuating the capital's expanding real estate market and the increased value of properties in this time.

02. Analysing the Impact of House Characteristics on Price

A Comparative Scatterplot of Area vs. Price Across Varied House Types

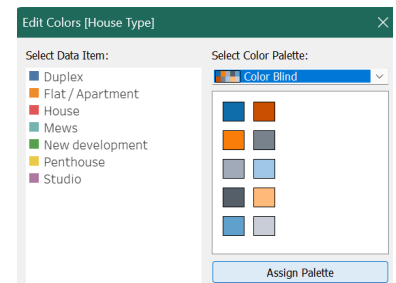


Filters to filter out various House Characteristics

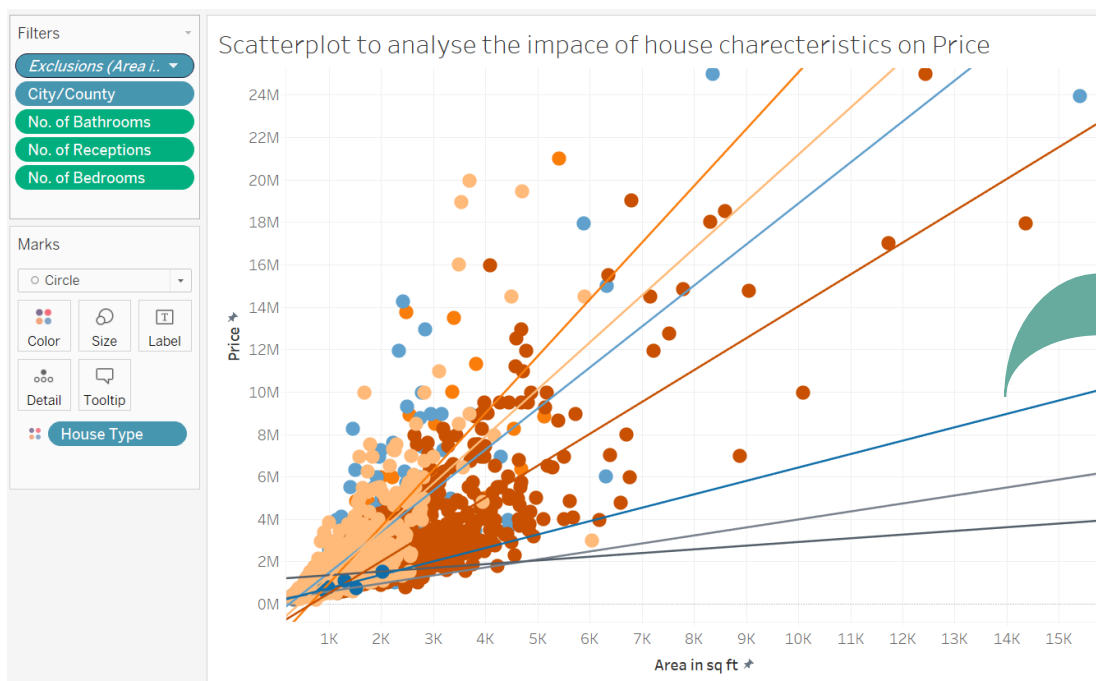


All visualizations for this question are created using Tableau

Tableau builtin Colorblind Palette

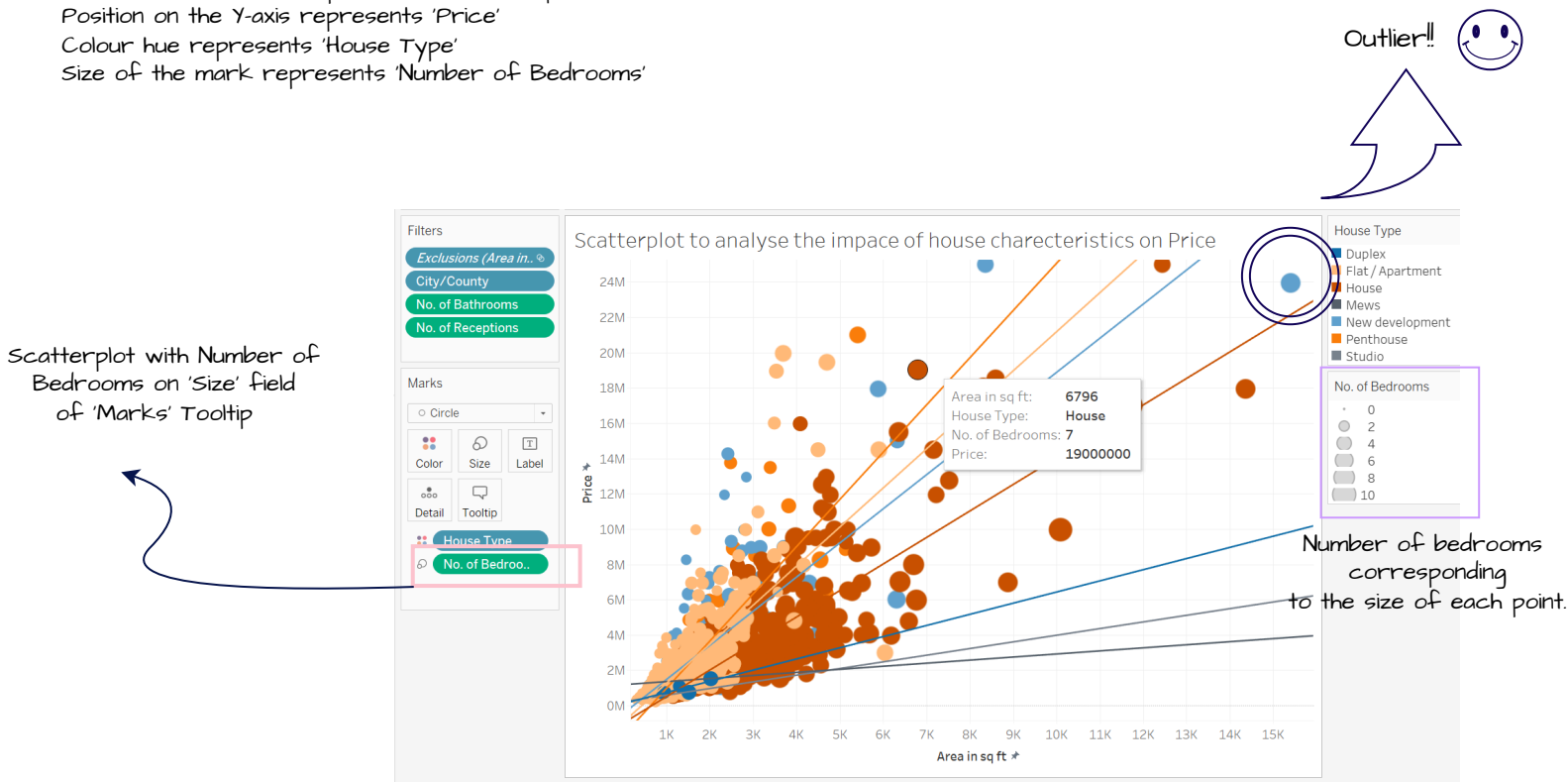


Filter to Select the House Type



The trend lines illustrate a positive correlation between area and price for each house type indicating that larger properties tend to command higher prices in the market.

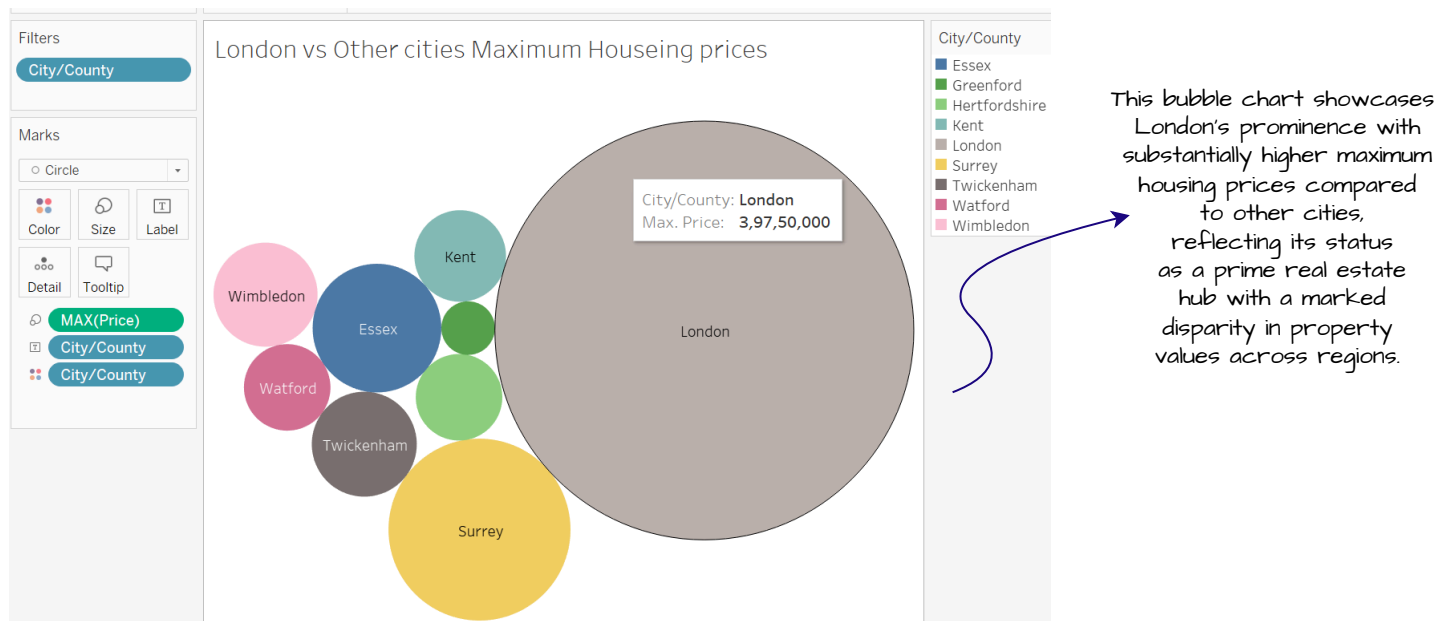
Type: Quantitative Scatterplot
Size: Data points vary based on the filter conditions on Size on 'Marks' field
Colour: Colourblind-friendly palette (distinct colours for different house types)
Scale: Linear scale for both 'Price' and 'Area in sq ft'
Mark: Circles to represent data points
Channels:
 Position on the X-axis represents 'Area in sq ft'
 Position on the Y-axis represents 'Price'
 Colour hue represents 'House Type'
 Size of the mark represents 'Number of Bedrooms'



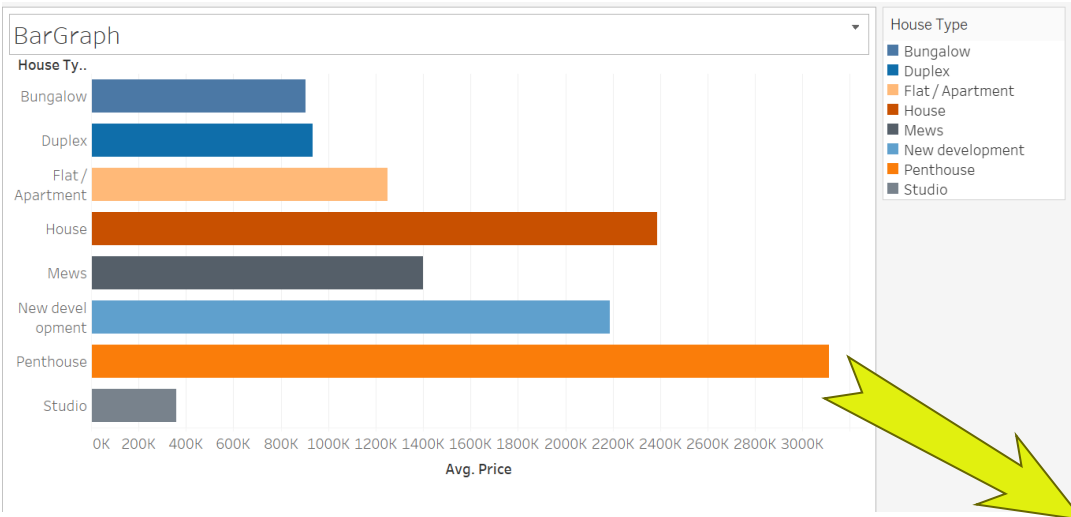
The visualization illustrates a general positive correlation between house characteristics such as area, no. of bedrooms, bathrooms and reception and their market prices, underscoring the significant role these features play in determining property value.

Design improvement: To showcase each Housetype in individual visualization for detailed analysis,

Fancy a comparison between Prices in London and Other Cities?



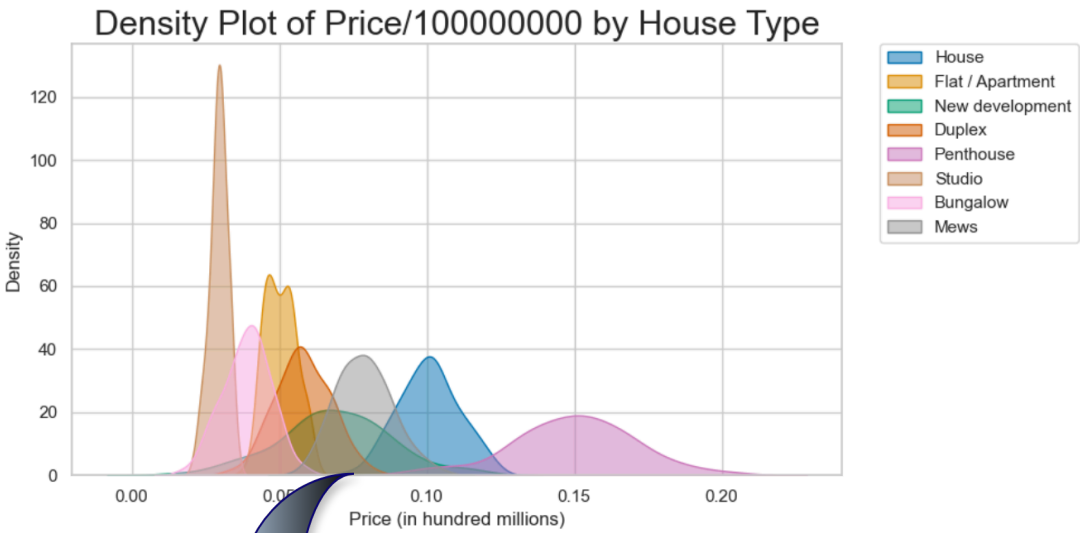
Bargraph to illustrate diversity in housing prices across different types of property.



Deeper insight into the price distribution within each category

Distribution of house prices by type

Penthouse and Houses commanding higher average prices



Flats/apartments and new developments showing a wider spread in values, indicating a diverse range of prices within these categories

03. Analysing the correlation of Property Size and Price in Central London

PRE-PROCESSING: Segregate Central London regions from the overall dataset

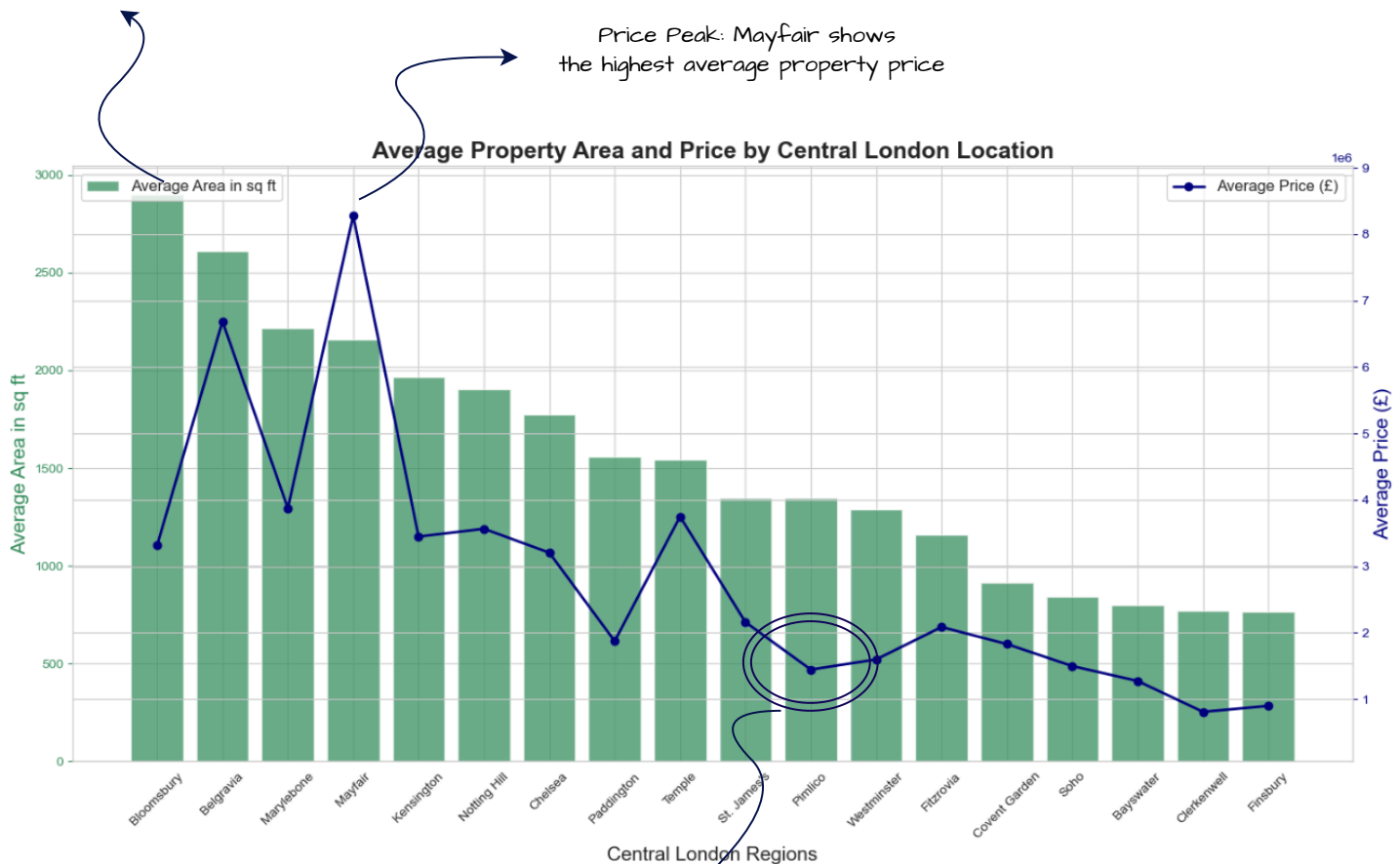
Bar and Line graph to analyse
Correlation of Property Size and Price in Central London



All visualizations for this question are created using Python and its libraries

Maximum Area: Bloomsbury boosts the largest average property sizes within Central London

Price Peak: Mayfair shows the highest average property price



Contrasting Trends

Compact Living: Smaller average property areas in regions like Soho, Bayswater, Finsbury could indicate a higher density of flats or apartments.

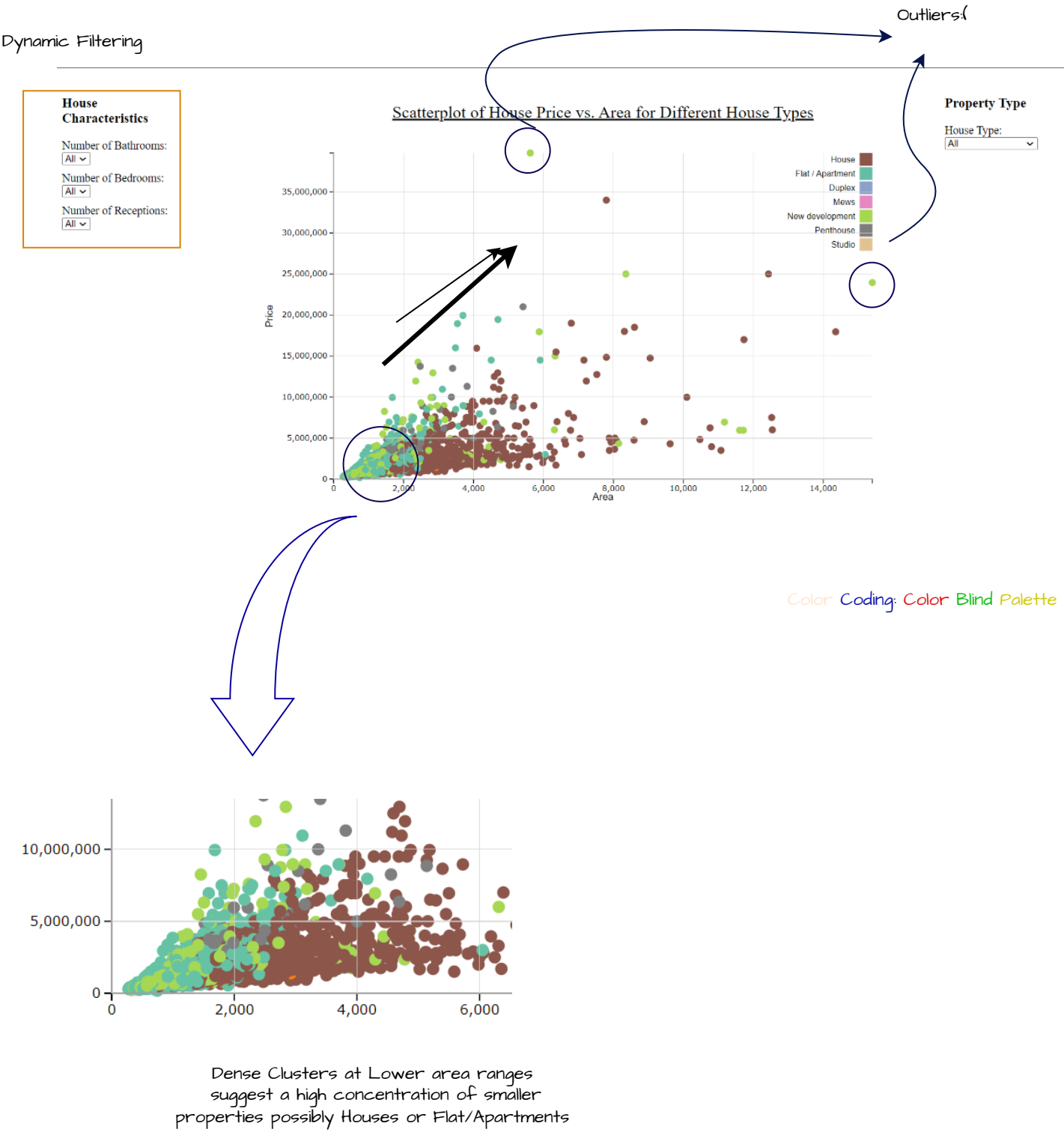
Type: Combined bar and line chart.
Size: Variable bar and line heights
Colour: Colourblind palette
Scale: Dual quantitative
Mark: Bars and line
Channels:
X-axis categorical (regions)
Y-axis dual-scaled (area and price)
Colour distinction (area and price indicators)

we can infer trends such as whether larger average property areas correlate with higher prices

Design improvement: Create visualization to showcase all regions of London and neighbouring cities.

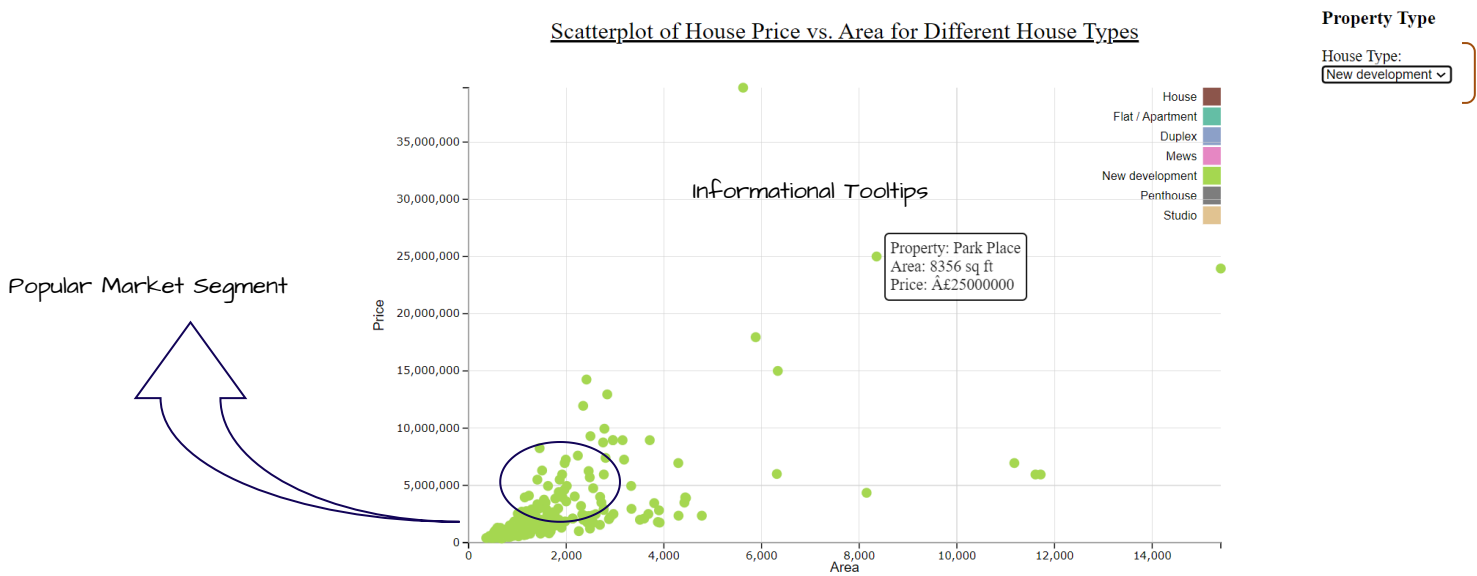


Impact of House Characteristics on Price

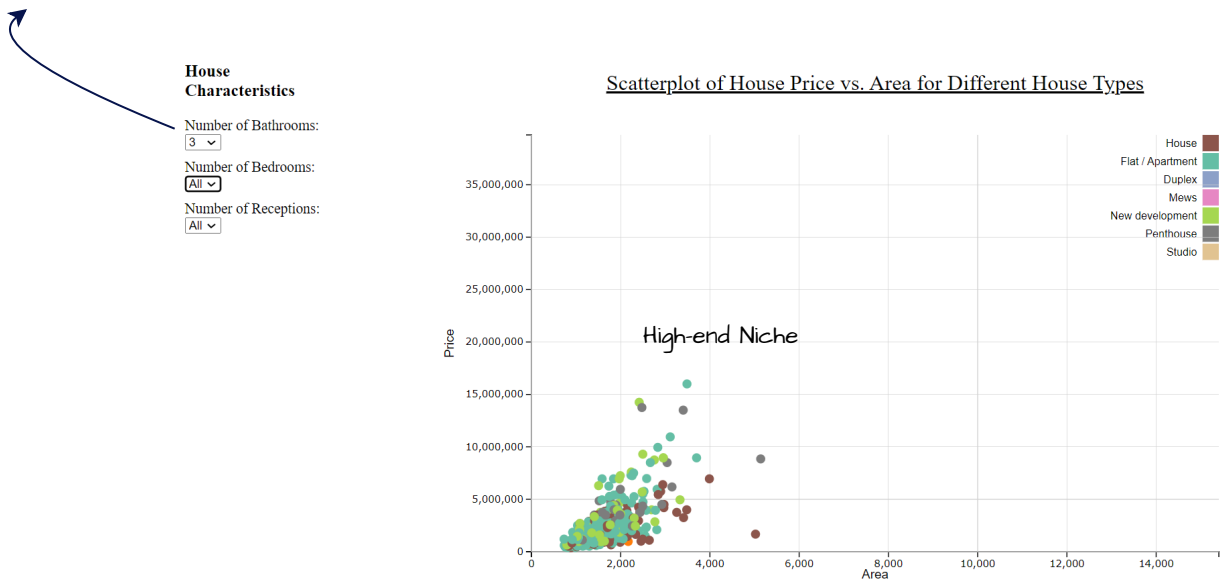


Filter the House type by New Development

House prices versus area for new developments

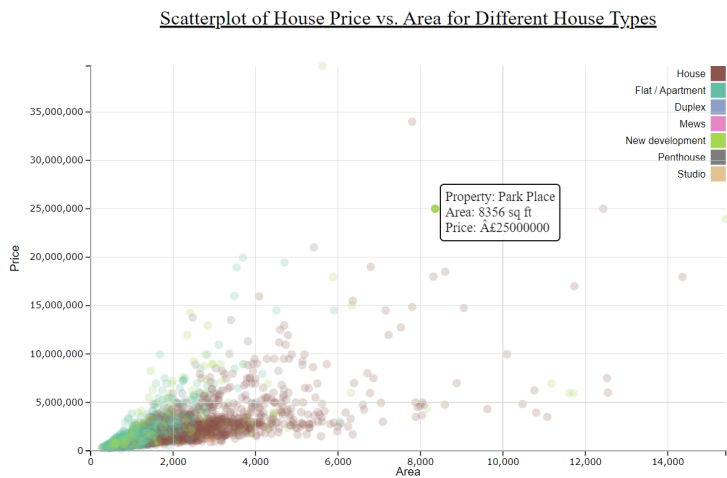
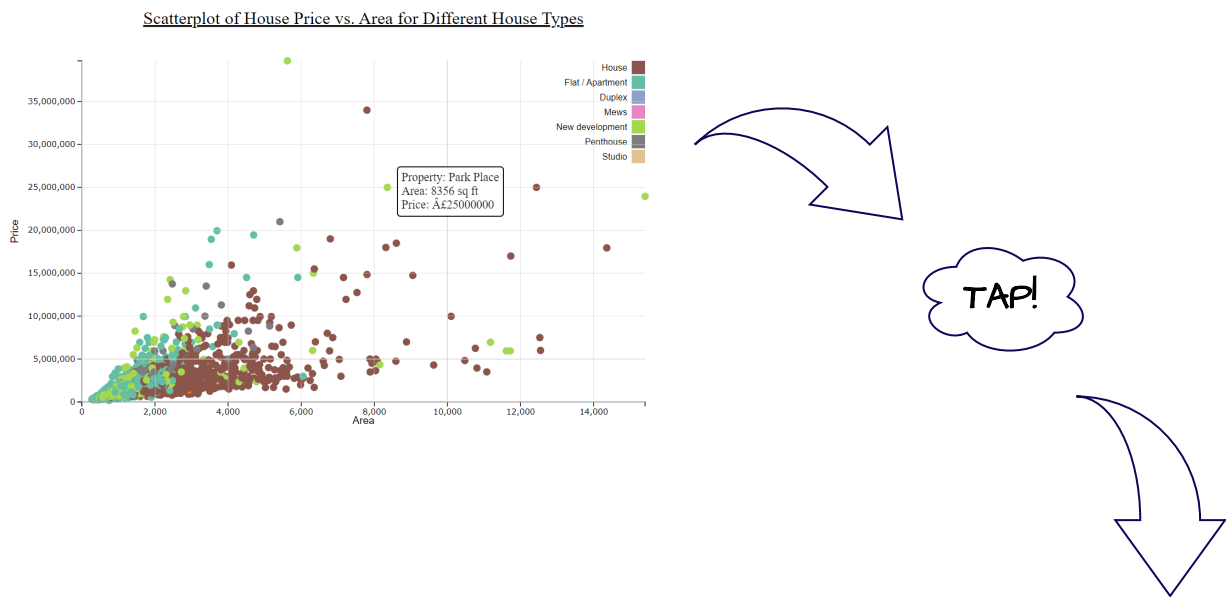


Filter applied for Number of Bathrooms=3



Price Variation: Even with the same number of bedrooms, prices vary widely indicating the influence of other factors like locations or property age

Point Highlighting and De-emphasis



In conclusion, this scatterplot reveals varying price sensitivities across different House types. Steeper trend for some indicates a higher premium per square foot. This suggests that market preferences or scarcity significantly influence property values regardless of size.

Thankyou!