Introduction to Data Visualisation: Coursework

Alexander Smerdon
23031306

April 2024

Part 1: Analytics

I have chosen to take the House prices data analysis track.

Q1: "Analyse the development of house prices over-time. Are there any detectable trends?"

The provided continuous dataset shows the average price of houses over time in the UK from 1992 until 2023. This dataset shows the changing rate of house prices, and may show them rising (or falling) over time. I will be analysing the full time frame.

Q2: "Analyse regional house prices in the UK. Are there any differences between regions?"

The provided dataset also shows the breakdown of each UK region, allowing for observation of the variation in rate of change between regions. This may also require London specific (borough) continuous house price data [1], as there is likely to be massive variations in house prices across London. This question would be complementary to Question 1, as it would delve deeper into the regional differences in house prices from.

Q3: "Analyse household tenure in the UK. Are there any detectable trends?"

For this question, the Office of National Statistics (ONS) tenure date dataset [2] is needed. This categorical dataset shows the changes over time in rental vs ownership rate (and how this may change based upon region).

All of these datasets would be complimented by data showing the rate at which houses have been built across the UK, compared to the increase in population.

There may be a negative correlation between house price and ownership rate in the Q3 dataset (the more prices go up, the fewer people can afford to purchase a house). It may also be useful to look at the multiple home ownership rate in the UK.

There may be correlation between the regions. For example, house prices in London may drive up house prices in neighbouring regions, as the demand for housing is a lot higher.

There may be a positive correlation (dependency) between the London dataset and the regional dataset.

If a complimentary dataset showing GDP growth figures from each year were added, correlation between house prices and the growth rate could also be observed.

If a complimentary dataset showing population growth were added, positive correlation between house prices and population may be observed (as more demand leads to higher prices). Again, this would be complimented by data on house building rates in the UK.

Part 2: Design and Discussion

Q1: Design rationale:

The visualisation can be seen in Figure 1 of the appendix (and the initial sketch in Figure 2).

To best represent the development of UK house prices over time, I have chosen to use a time-series plot between 1992 and 2023. I believe this is the best way to display continuous data, and to capture temporal

trends whilst also allowing for regional differences to be highlighted. When a user hovers over a region's line, the price for that quarter will be shown.

Line encoding - utilising lines (1D) as the primary visual mark allows for clear representation of the house price trends. Each region of the UK is assigned a distinct colour. This ensures differentiation and allows for easy comparisons to be made. The UK average is marked with a dotted line.

Colour encoding - using colours to differentiate between regions enhances perceptual clarity. The colour choices made adhere to the principles outlined by Jacques Bertin in his work "Semiology of Graphics" [3]. The colours maximise visual discriminability whilst minimising ambiguity.

Positional encoding - having the position of data points the temporal axis allows for precise interpolation of temporal patterns and trend detection. Using Steven's Power Law [4], the placement of data points ensures proportional changes in house prices are accurately reflected.

Clarity and readability - the visualisation has a minimalist design to avoid visual clutter, and the axis labels (and titles) provide context and aid interpretation. The gridlines aid in aligning data points for accurate comparison.

Along with different regions, I have also derived an attribute by combining all the different regions together to create the average change in house price across the UK between the specified timeframe. This allows for a good baseline comparison, which makes regional differences in the dataset more apparent.

Finally, the scale of the graph has been designed with user perception in mind, as a user's ability to perceive differences in house prices might vary depending on the magnitude of the price changes (i.e., smaller changes might be more easily detected when house prices are lower compared to when they are higher).

Q2: Design rationale:

The visualisation can be seen in Figures 3 and 4 of the appendix.

To best represent the regional variation in house prices across the UK (also including London), I have chosen to use a choropleth map. This is because of its ability to encode quantitative information (i.e. average house price) through colour intensity. The user would be able to hover over a region with their cursor, and see the region's name and the average house price. Doing this separately (instead of showing all the information at once in each region) allows for more focus and clarity. When a user hovers over a region, that region's colour will become slightly lighter - again for clarity.

Colour encoding - using colour intensity to represent house prices allows for readers to immediately understand regional disparities. Utilising an ordered colour map (with lighter shades of blue representing lower house prices and darker shades of blue representing higher house prices), the transition between prices is easy to interpret and enhances visual clarity. When it comes to accessibility, I have experimented with the colour palette to make sure that people with visual impairments (such as colour-blindness) are also easily able to interpret the choropleth map. I felt the changing of the colour blue provided the best clarity.

Geographic encoding - using the geographic boundaries of regions of the UK (and the boroughs of London) gives immediate context to the visualisation and aids in the spatial analysis of house prices. Each region is outlined and filled with colour to indicate its average house price level.

Contrast and differentiation - the selection of colours ensures contrast between regions, which allows for easy differentiation and comparison.

Projection selection - although the Mercator projection is often used for geographic data visualisation, its distortion of spatial areas may be less relevant with this dataset as it only applies to the national-level. This choice of projection however should still maintain an accurate representation of the UK's geographic features.

Q3: Design rationale:

The visualisation can be seen in Figure 5 of the appendix.

To best represent the trends of UK housing tenure over time, the visualisation I have chosen is a stacked bar chart. This would allow for the comparison of multiple categories (i.e. renting, mortgage, own-outright) within several different time periods, while also highlighting changes over time. Unlike the previous visualisation (choropleth map), the names of the tenure type will be shown on the stacked bar chart. This is because there are far fewer categories, and so the reader will not be so overwhelmed with information.

Length and position encoding - utilising the length of each bar segment to represent the proportion of households with a specific tenure type allows for an easy comparison between the categories. Positioning each bar to a specific year allows for the visualisation to easily show changes in the tenure distribution over

time (for example, the proportion of home-owners could fall from 30% to 25% over the course of the plotted years).

Colour encoding - having different colours represent the tenure types within each time period aids in distinguishing categories and enhances visual clarity. Following the principles outlined by Cleveland and McGill [5], contrasting colours are chosen to maximise perceptual discriminability. Similarly to the choropleth map in Q2, the colour palette is designed to make sure people with visual impairments are able to interpret the visualisation.

Gestalt principles - using Gestalt principles such as proximity and similarity, the stacked bar chart organises the tenure types into distinct groups. This makes it easier for viewers to perceive patterns and trends. Grouping similar tenure types together enhances the chart's readability and interpretability.

References

[1] London Datastore. Average house prices by borough dataset. Accessed March 2024.

[2] Office for National Statistics (ONS). Household characteristics by tenure, England and Wales: Census 2021. Accessed March 2024.

[3] J. Bertin. *Semiology of Graphics: Diagrams, Networks, Maps*, 1967.

[4] S. S. Stevens. Neural events and psychophysical law. *Science*, Volume 172, Issue 3982, pp. 502, April 1971.

[5] W. S. Cleveland and R. McGill. *Journal of the American Statistical Association*, September 1984.

Appendix

Figures:

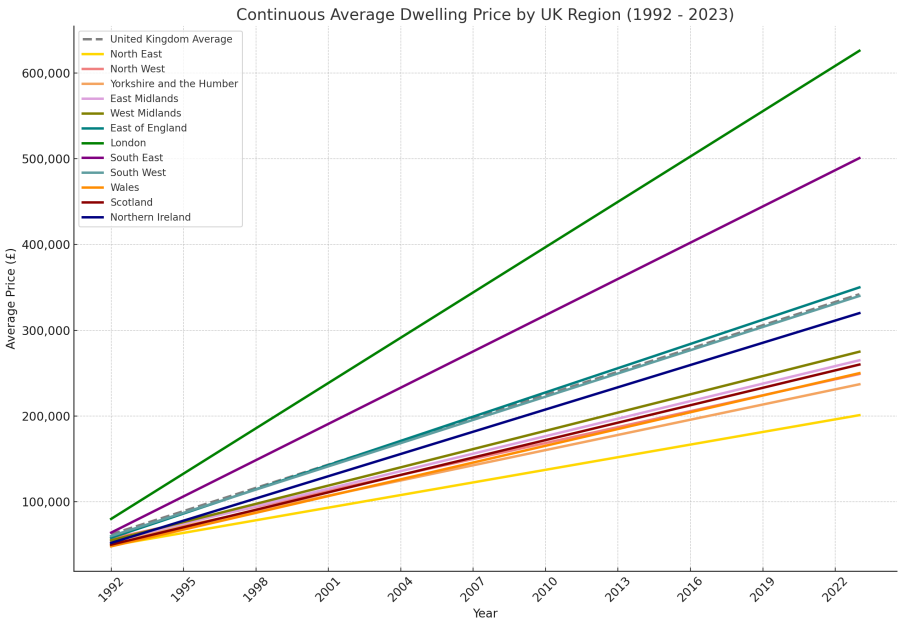

Figure 1: Q1 Visualisation

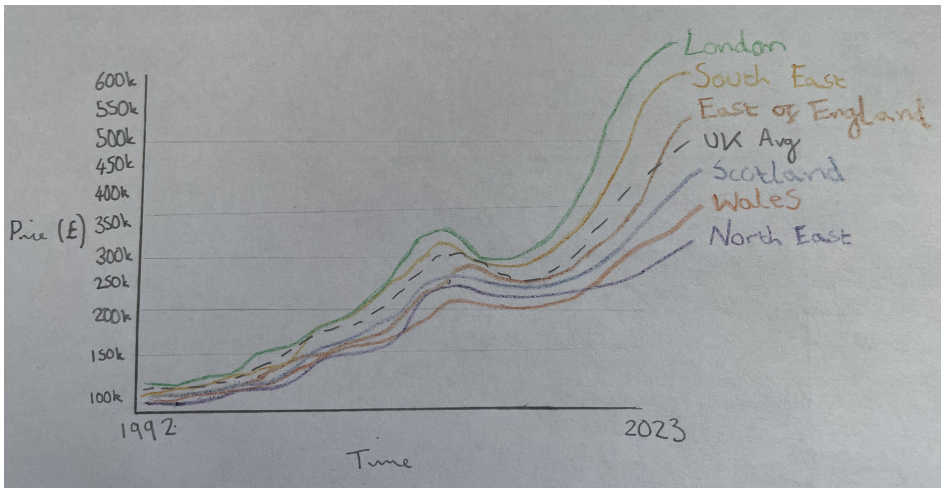

Figure 2: Q1 Sketch

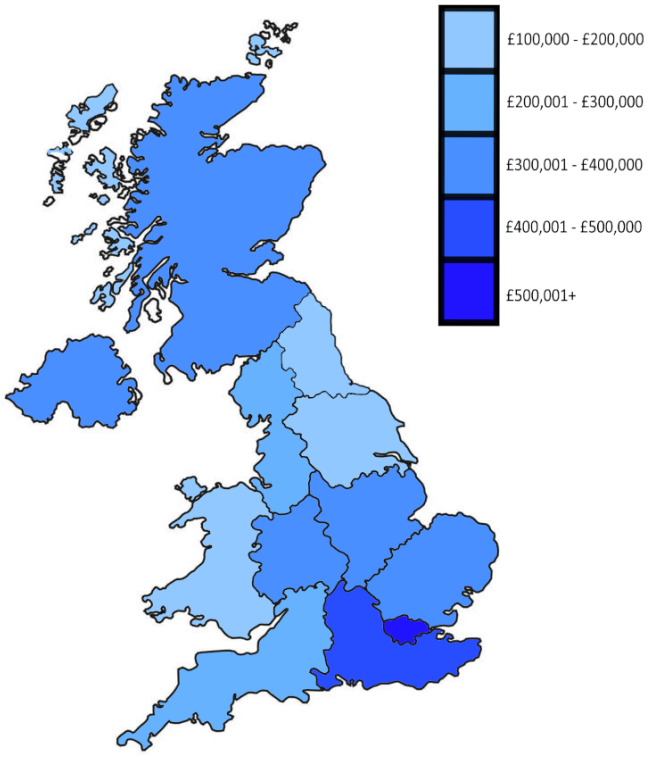

Figure 3: Q2 Visualisation (UK)

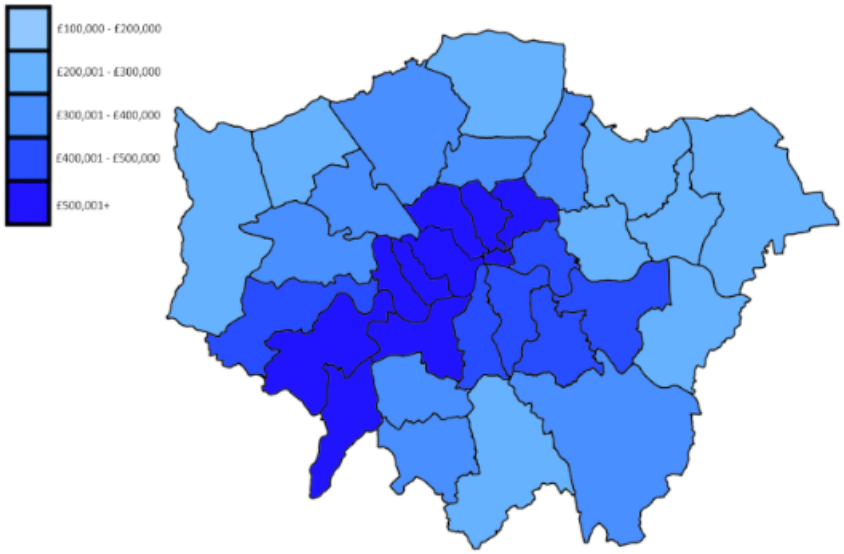

Figure 4: Q2 Visualisation (London Boroughs)

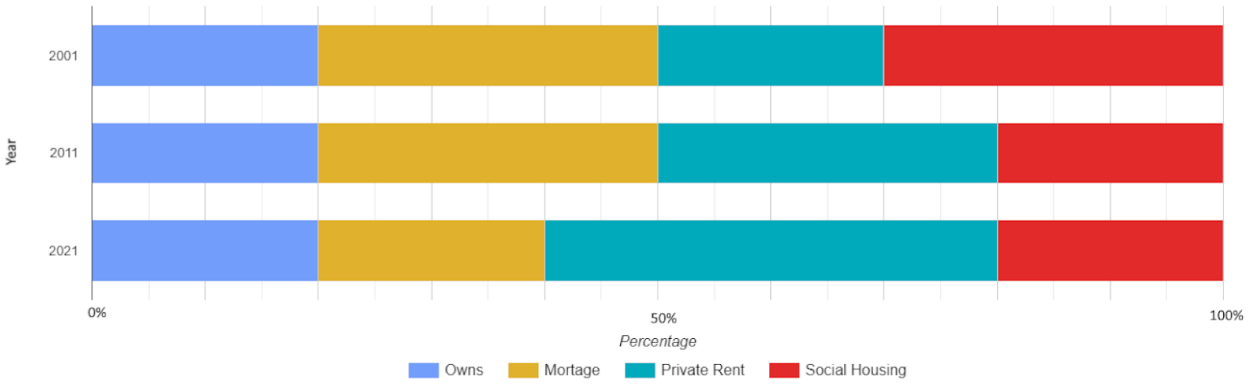

Figure 5: Q3 Visualisation