

Assignment 2: Visual Design Justification Report

Student Name: Jabin Tasnim Khan Urmy

Date: November 2, 2025

Course: High Performance Data Analytics and Visualisation

Github Link: <https://github.com/JabinTasnimUrmy/Housing-Data-Interactive-Visualization>

1. Introduction

This report justifies the visual design of two synchronized interactive visualizations created for exploring the Housing dataset. The implementation uses React and D3.js to provide complementary analytical perspectives through:

1. **Scatterplot** - Showing the relationship between Price and Area with 2D brush selection
2. **Parallel Coordinates** - Displaying six dimensions simultaneously with click and hover interactions

Both visualizations are fully synchronized: selections made in either view automatically update the other, enabling seamless exploration from overview to detail.

2. Visual Encoding and Data Properties

2.1 Scatterplot: Price vs. Area

Visual Encoding Choices:

Visual Channel	Mapped to	Justification
X-position	Area (sq ft)	Position is the most accurate perceptual channel for continuous quantitative data
Y-position	Price (Unit)	Vertical position effectively shows the dependent variable
Color	Selection state	Red (#B91C1C) for selected houses, light blue (#93c5fd) for unselected
Opacity	Selection emphasis	High opacity (0.90) for selected, low (0.07) for unselected, base (0.28)
Stroke width	Selection emphasis	Thicker (2.6px) for selected, thinner (1.15px) for base
Mark type	Individual house	Circle - standard, recognizable mark for point data

How Encoding Fits Data Properties:

The Housing dataset contains 545 records with price ranging from 1.75M to 13.3M and area from 1,650 to 16,200 sq ft. These are the two most influential continuous variables with a clear positive correlation. Using position for both leverages human perception's highest accuracy channel. The scatterplot immediately reveals:

- **Correlation strength:** The overall upward trend shows how price increases with area
- **Outliers:** Points far from the trend line represent unusual properties (e.g., expensive small houses or cheap large houses)
- **Clusters:** Natural groupings of houses with similar characteristics
- **Variance:** The spread shows how much price varies for similar-sized properties

User Tasks Supported:

1. **Overview exploration:** Quickly grasp the general price-area relationship
2. **Pattern identification:** Find correlations, clusters, and trends
3. **Outlier detection:** Spot houses that don't follow expected patterns
4. **Subset selection:** Use 2D brush to select rectangular regions for focused analysis
5. **Comparison:** Compare different price/area segments

Interaction Design:

The 2D brush (d3.brush) allows users to drag and create rectangular selection regions. This interaction directly supports tasks like "find houses between 5-7M with 6000-9000 sq ft area." The brush selection immediately highlights corresponding lines in the parallel coordinates view.

2.2 Parallel Coordinates: Multi-dimensional View

Visual Encoding Choices:

Visual Channel	Mapped to	Justification
Vertical axes	Six dimensions (Price, Area, Bedrooms, Bathrooms, Stories, Parking)	Parallel placement enables comparison across multiple variables
Polyline	Individual house	Each line represents one house's values across all dimensions
Color	Selection state	Same color scheme as scatterplot for consistency
Opacity	Selection emphasis	Same opacity scheme for visual coherence
Stroke width	Selection and hover	Thicker for selected (2.6px), thickest for hover (3.5px with orange color)

How Encoding Fits Data Properties:

The Housing dataset is inherently multi-dimensional. While Price and Area are most important, users need to understand how bedrooms, bathrooms, stories, and parking relate to price and each other. Parallel coordinates excels at this by:

- **Showing all dimensions simultaneously:** Users see complete house profiles at once
- **Revealing multi-variate patterns:** Parallel line trajectories show correlations across multiple dimensions
- **Supporting individual inspection:** Users can examine specific houses by clicking their lines
- **Maintaining individual visibility:** Each of 545 houses remains a distinct, traceable line

User Tasks Supported:

1. **Profile comparison:** Compare complete attribute profiles of different houses
2. **Relationship discovery:** Identify which dimensions vary together across houses
3. **Individual inspection:** Click any line to highlight that specific house in both views
4. **Multi-select comparison:** Select multiple houses (using Ctrl+Click) to compare their profiles
5. **Pattern recognition:** Identify unusual combinations or typical patterns across dimensions

Interaction Design:

- **Line clicking:** Click any line to select that individual house; the corresponding point highlights in the scatterplot
 - **Multi-selection:** Hold Ctrl (Windows) or Cmd (Mac) while clicking lines to select multiple houses simultaneously
 - **Hover feedback:** Lines turn orange (#FF5722) and thicken (3.5px) on hover for easy identification and visual feedback
-

3. Synchronized Interaction

The key strength of this design is bidirectional synchronization:

Scatterplot → Parallel Coordinates:

- Brush a region in the scatterplot → Corresponding lines highlight in parallel coordinates
- Users see which multi-dimensional profiles belong to their selected price/area range

Parallel Coordinates → Scatterplot:

- Click any line in parallel coordinates → That house's point highlights in scatterplot
- Multi-select lines (Ctrl+Click) → Multiple corresponding points highlight in scatterplot
- Users see where their selected houses fall in price/area space

This two-way interaction creates a powerful exploratory workflow where discoveries in one view inform and guide exploration in the other.

4. Pros and Cons of Design Choices

4.1 Scatterplot Strengths and Limitations

Pros:

1. **Universal comprehension:** Scatterplots are widely understood, requiring no training
2. **Immediate pattern recognition:** Correlation, clusters, and outliers are instantly visible

3. **Precise selection:** 2D brush enables exact rectangular region selection
4. **Fast rendering:** 545 points display smoothly with no performance issues
5. **Effective for main relationship:** Perfectly suited for showing price-area correlation

Cons:

1. **Limited to two dimensions:** Cannot show other attributes without additional encoding
2. **No categorical display:** Binary variables (mainroad, basement, etc.) are not visible
3. **Potential overplotting:** Dense regions may obscure individual points (though not an issue with this dataset size)
4. **Static encoding:** Only position, color, and size are used; other channels remain unutilized

4.2 Parallel Coordinates Strengths and Limitations

Pros:

1. **Multi-dimensional analysis:** Shows six variables simultaneously in one view
2. **Individual house selection:** Direct clicking provides precise control over which houses to examine
3. **Multi-select capability:** Ctrl+Click enables custom comparison sets
4. **Immediate visual feedback:** Hover effects clearly indicate interaction possibilities
5. **Pattern visibility:** Parallel trends reveal correlations across dimensions

Cons:

1. **Learning curve:** Unfamiliar to many users; requires explanation of the concept
2. **Visual complexity:** 545 lines create visual clutter, especially in dense regions
3. **Limited filtering capability:** Cannot filter ranges of values, only select individual records
4. **Axis order dependency:** Different axis arrangements reveal different patterns; current order is arbitrary
5. **Scalability concerns:** Would become difficult to read with thousands of records

5. Conclusion

This design effectively integrates scatterplots and parallel coordinates to support comprehensive housing data exploration. The scatterplot enables intuitive area-based filtering, while parallel coordinates allow detailed inspection and comparison of individual houses. Their synchronized interactions bridge overview and detail, leveraging each visualization's strengths to deliver a cohesive, powerful exploratory experience.

References:

- Munzner, T. (2014). *Visualization Analysis and Design*. CRC Press.
- Inselberg, A. (2009). *Parallel Coordinates: Visual Multidimensional Geometry*. Springer.
- Cleveland, W.S., & McGill, R. (1984). *Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods*. JASA.