Part 2 & Part 3: Project Scope, Task Analysis and Visualization Ideas

Visualizing Vancouver's AirBnB Dynamics Through Analyzing the Variables of Success and Guest Experiences

Through this project, we aim to uncover the relationships between property types, their pricing, availability, host features, and various review metrics in Vancouver's Airbnb listings. By understanding these dynamics, we hope to deduce property qualities that offer the best value, the effect of super hosts, and how property characteristics influence guests' experiences.

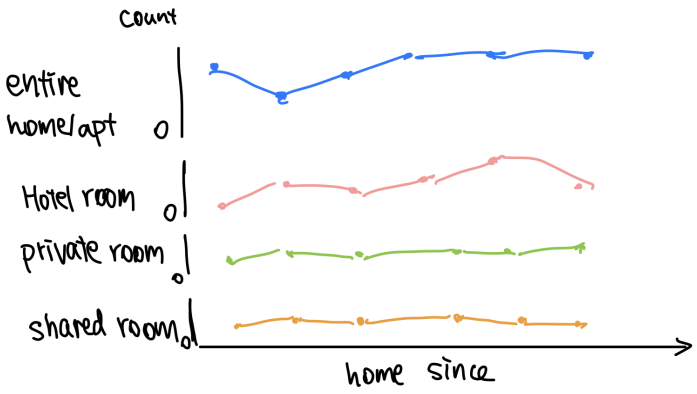
Intended Audience: AirBnB Existing and Potential Hosts.

In a crowded market, existing and potential hosts would want to improve and optimize their offerings, pricing strategy, and guest experiences. The visualizations would enable them to understand the dynamics of the market better, what works best, and how to stand out among the competitors. The benefits of enhancing the guest experience and pricing are reflected by positive reviews, directly relate to guest satisfaction and ultimately affect their profitability.

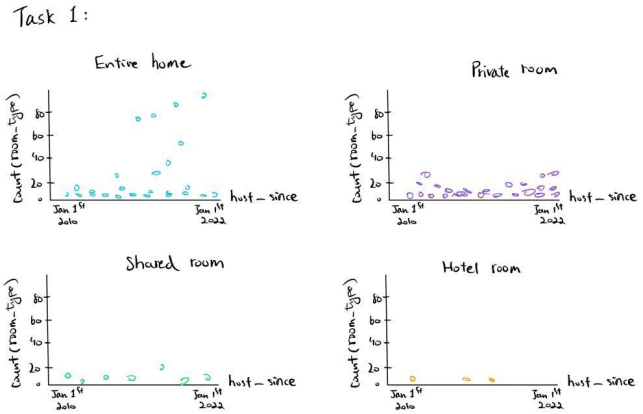
Using the visualization, the hosts can make more informed decisions. Effectively visualizing tools can simplify the tricky task of pricing, for example, displaying clusters of price in relation to factors like review scores. Moreover, spotting competitive strategies becomes more intuitive when top-performing listings are visualized and juxtaposed. Airbnb hosts can leverage this information to improve their services, pricing, and overall guest satisfaction.

Task Analysis:

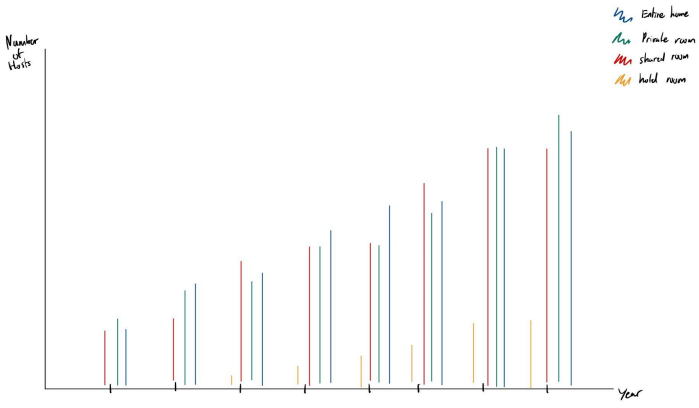
Task: Examine the influx of Airbnb hosts in relation to shifts in preferred room type throughout the years.
Attribute: host_since, room_type
Rationale: Assess the evolving competitive landscape by tracking both the growth in new hosts and any emerging trends in property type preferences.



(graph 1.1)



(graph 1.2)



(graph 1.3)

Graph 1.1: Line Graph

Expressiveness:

- The line graph aims to show trends over time, which is inherently expressive for indicating growth, decline, or stability in data. The choice of a line graph to show how different room types trend over a timeline is appropriate.
- The different colors effectively express distinct categories, allowing for quick visual distinction between trends.

Effectiveness:

- As a conceptual tool, the graph effectively shows the relative positions of different room types over time. It could suggest that entire homes/apartments have the highest count, followed by private rooms, and so on.

Graph 1.2: Scatter Plot Quadrants

Expressiveness:

- The scatter plots are expressive in demonstrating the distribution and density of data points. The spread of points can indicate variability and concentration within each category.
- The use of separate quadrants for each room type expresses a clear categorical distinction and allows each category to be individually analyzed without interference from the others.

Effectiveness:

- The scatter plot could be effective in showing the lack of a clear trend within each room type over time.
- It's also effective in showing the potential outliers or clusters within the data

Graph 1.3: Bar Chart

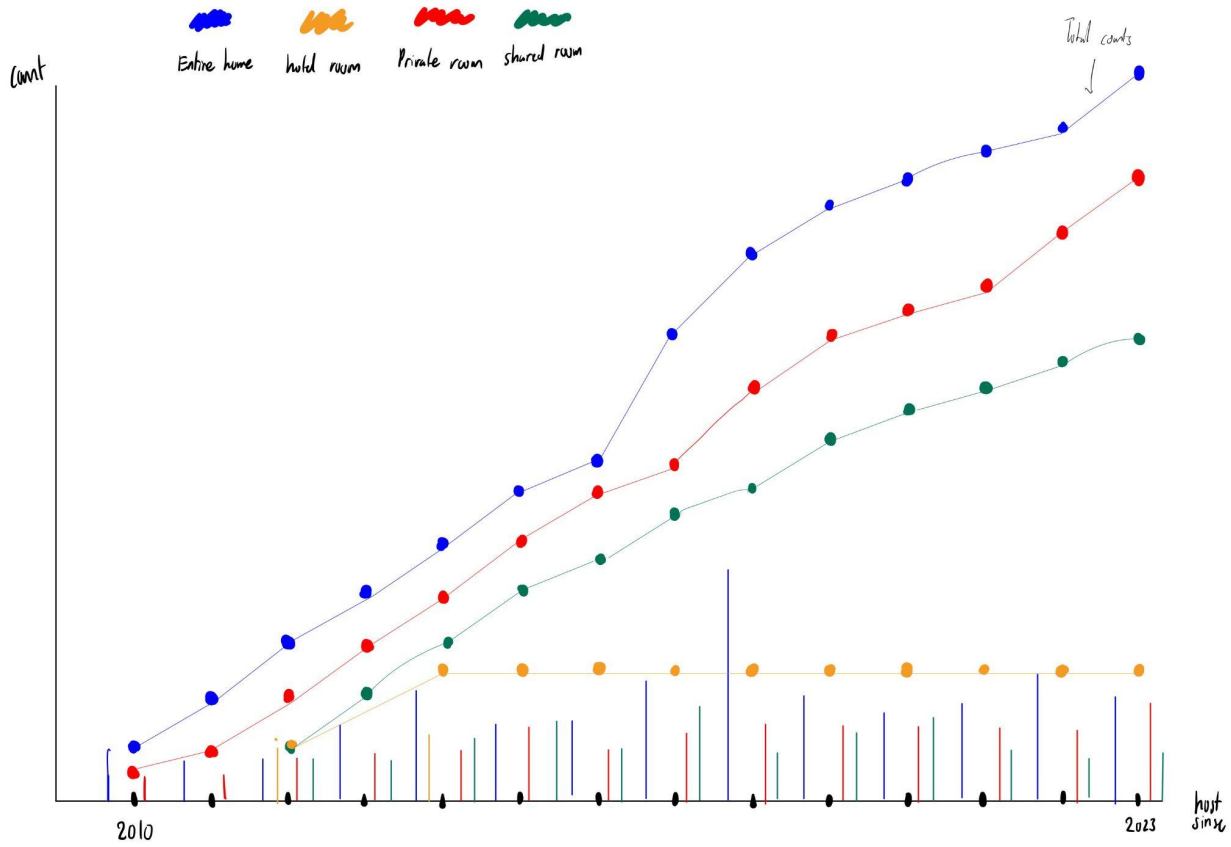
Expressiveness:

- Bar charts are inherently expressive when it comes to showing comparative quantities. The use of vertical bars allows for an intuitive understanding of more vs. less.
- The color-coding continues to be expressive of distinct categories, tying this graph to the previous ones.

Effectiveness:

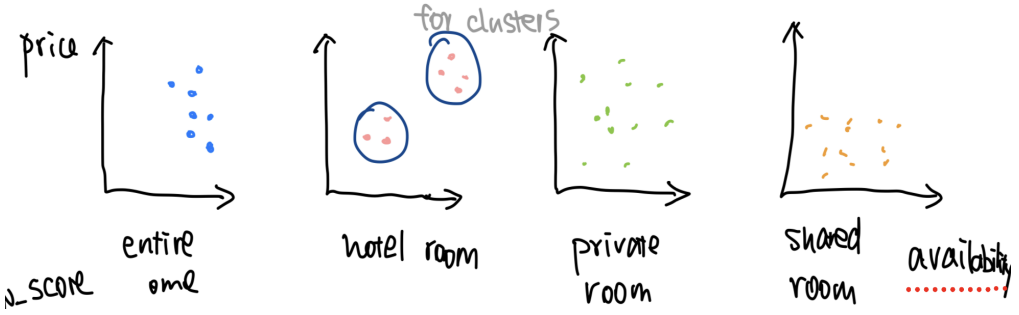
- Assuming each bar represents a year or another unit of time, the bar chart effectively shows the growth or changes in the number of hosts across the different room types.
- The separation of room types by color in adjacent positions is effective for quick side-by-side comparisons within a specific time frame.

High Fidelity Sketch:

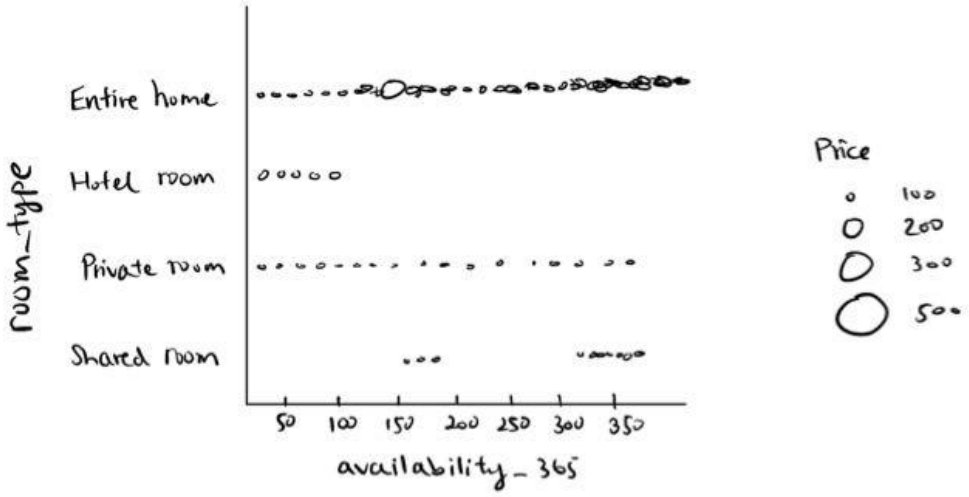


The finial high fidelity sketch is a used inspiration from graph 1.1 and 1.3 the sketch used a combination of bar and point plot. The bar shows the number of the host that have join Airbnb in that particular year for each room type. The point shows the total increase of the of each room type. In the final design, attention to detail would be necessary to ensure that the dual elements of bars and points do not create a cluttered visualization. Moreover, integrating these components should aim to highlight meaningful insights, such as peaks, troughs, in the data, and guide the viewer to understand the significance of trends in the context of Airbnb's host growth and room type popularity.

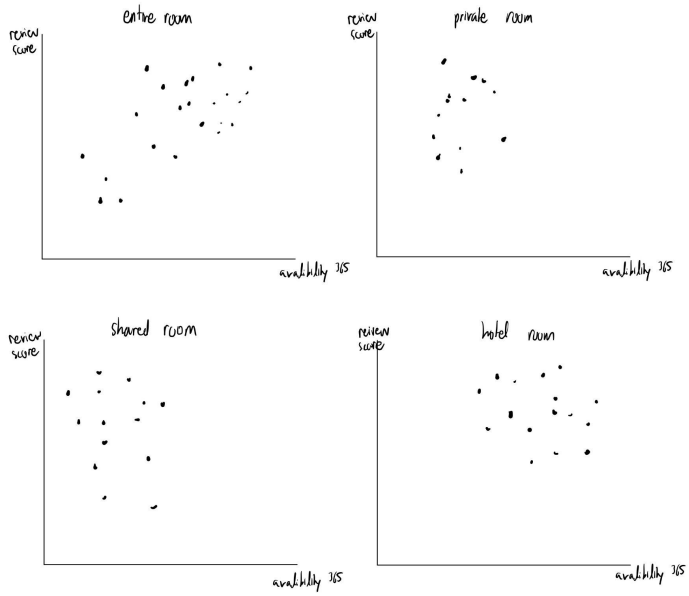
Task: Is there a cluster of the typical price and availability based on the room type?
Attribute: availability_365, price, room_type
Rationale: Determine whether certain room types tend to have similar price ranges and availability, and whether these groupings or patterns can be easily identified or segregated.



(graph 2.1)



(graph 2.2)



(graph 2.3)

Graph 2.1: Multiple Scatter Plots

Expressiveness:

- The use of different colours for each room type aids in distinguishing them visually.
- The graphs are presented alongside one another, allowing for effective comparison. However, comparison is not the goal of this task.

Effectiveness:

- The illustration effectively communicates that clusters exist within the data for different room types, indicating variations in price and availability within each category.

Graph 2.2: Scatter Plot

- This graph uses points to encode individual listings, and size to represent the price of each listing.
- For each variable in the y-axis, the denser the points are, it corresponds to having more days of availability.
- The area of the circle makes it very hard to distinguish the specific price, and the vertical positioning in the y-axis only helps with comparing the distribution of the availability of each room_type, but hard to compare the price between each room_type.

Graph 2.3: Multiple Scatter Plots

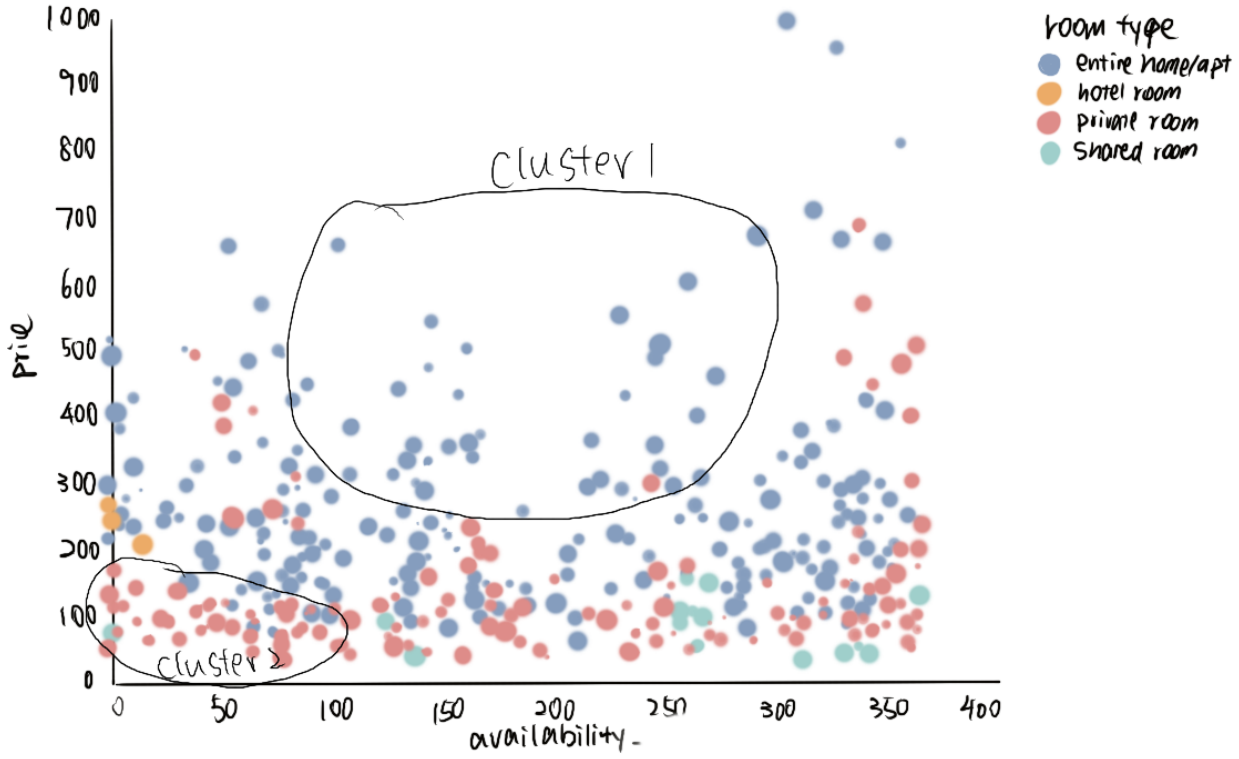
Expressiveness:

- By breaking down the scatter plots by room type, the graph expresses individual distributions of price for each room type, which can be indicative of each category's market positioning.
- The lack of a trend line or overlaid information focuses the viewer on the raw distribution of data points, which is expressive of the natural spread in the market.

Effectiveness:

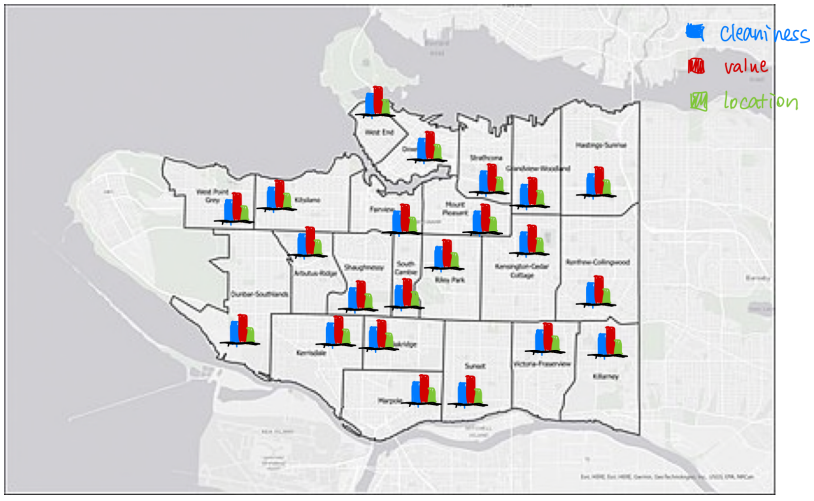
- This series of scatter plots effectively isolates the behavior of each room type concerning price, making it easier to compare them qualitatively.
- It is effective in showing that the concentration of data points and thus, pricing strategies, vary by room type.

High Fidelity Sketch:

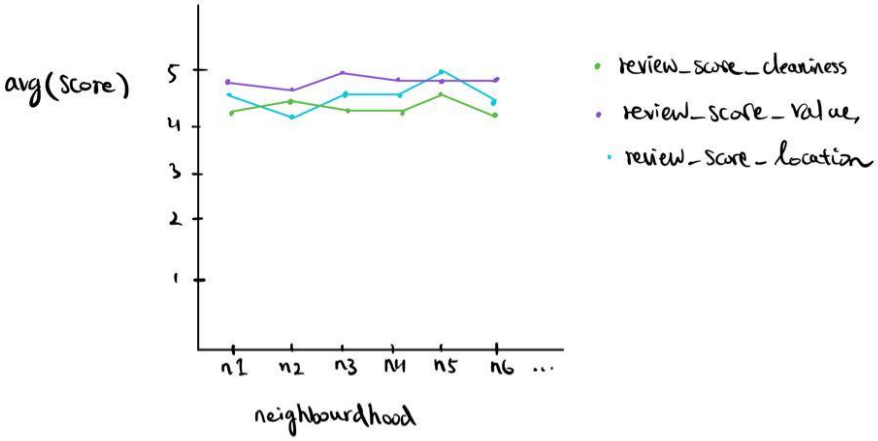


We chose to use a scatter plot for this task. Using color hue to encode room type is effective since it is nominal and only has a cardinality of 4. The scatter plot is effective at visualizing the distribution of the listings regarding price and availability. Although most of the data points are overlapping, there is still some cluster that can be identified as shown in the graph. Since the task is to determine whether certain room types tend to have similar price ranges and availability, using faceted graphs is not suitable for the task.

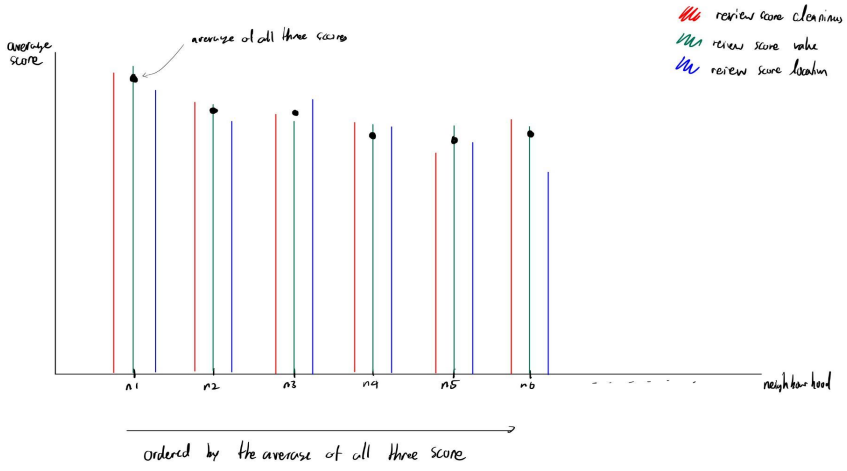
Task: Compare the review_score_cleaniness, review_score_value, review_score_location of each neighbourhood.
Attribute: review_score_cleaniness, review_score_value, review_score_location, neighbourhood_cleansed
Rationale:



(graph 3.1)



(graph 3.2)



(graph 3.3)

Graph 3.1:
This is a symbol map of aggregated mean scores of cleanliness, value and location of each neighborhood in Vancouver. The symbols in this map are the bar charts, with color and horizontal position encoding the types of scores (cleanliness, value and location), and y encoding the score.

Effectiveness:

- The map accurately captures the shape, area, and geographic position of each neighborhood, which is easy to understand.
- The necessity to fit bar charts within the neighborhoods means that some are too small, impacting the visualization's effectiveness. Smaller symbols can make it challenging to perceive differences in bar height, crucial for interpreting the data.
- The visualization does not effectively facilitate the intended task, which is to compare scores between neighborhoods. It is more effective at comparing different score types within each neighborhood.

Expressiveness:

- The choice of bar charts inherently expresses a comparison among different data points, aligning with the nature of the dataset.
- The bar charts compare cleanliness, value, and location scores within each neighborhood, expressing the relative strengths or weaknesses in these areas.
- The utilization of color in the bar charts is effective and expressive, with distinct hues representing different scores and a grey-scaled base map ensuring no interference with the bar chart colors.

Graph 3.2
This is a line chart with horizontal position encoding neighborhoods, color hue encoding types of scores and vertical position encoding the mean scores.

Effectiveness

- The chart effectively allows for comparison of average scores across different neighborhoods
- By using a line chart, the visualization effectively shows trends and patterns in scores across neighborhoods, which could indicate consistent issues or strengths in specific areas.
- The vertical axis clearly denotes the average score, allowing for a straightforward assessment of performance in each category across neighborhoods.

Expressiveness:

- The use of different color hues to represent various review scores (cleanliness, value, location) is expressive and helps to distinguish between the data series at a glance.
- Utilizing lines to connect data points representing neighborhoods implies a sequential or inherent order where none exists, as neighborhood data is nominal and unordered. This connection might inadvertently suggest a relationship or trend between neighborhoods that could mislead the viewer, as no ordinal significance is intended among the categories presented.

Graph 3.3
This is a bar chart, with horizontal position encoding neighbourhoods sorted by overall score,, color hue encoding each type of scores, vertical position encoding the score. With a dot representing the overall scores in each neighbourhood.

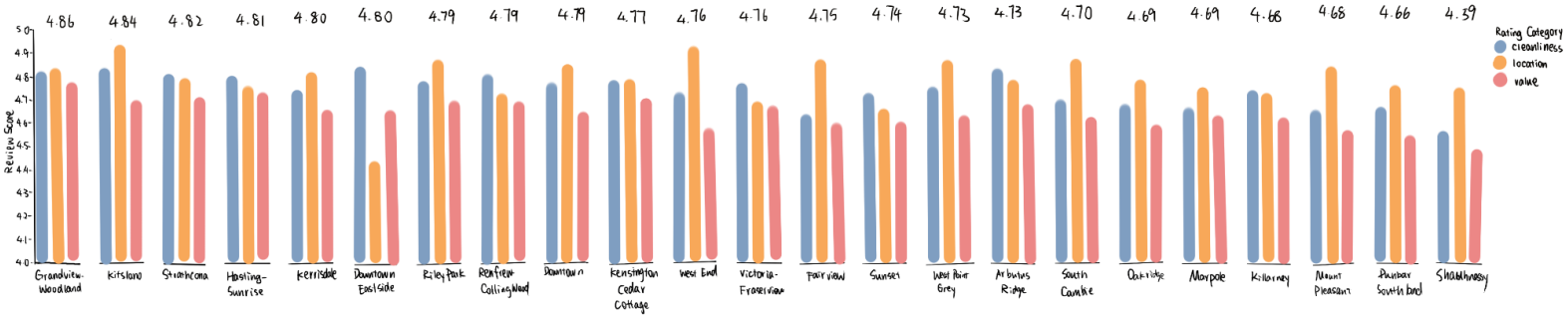
Effectiveness

- Grouped bars allow for a direct comparison of cleanliness, value, and location scores within each neighborhood.
- Ordering by the average score helps to quickly identify which neighborhoods perform better or worse on average, which may be useful for certain analytical purposes.
- Different colors are used to represent the different types of scores, which makes distinguishing between them straightforward.

Expressiveness:

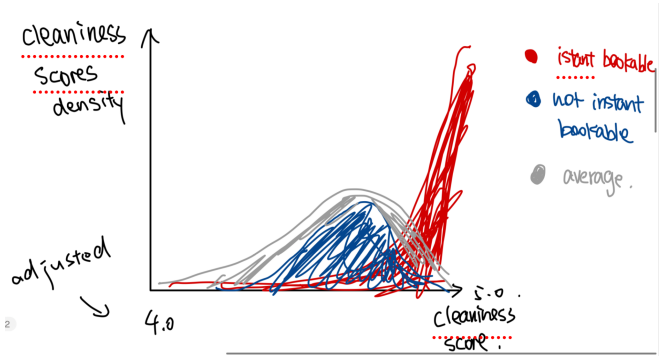
- The sequence from left to right reflects an intuitive increase or decrease, adhering to conventional reading patterns.
- The grouping and spacing of bars, along with the consistent color scheme, create a visually balanced chart.
- The line indicating the average score across all three categories provides an additional layer of information, giving a sense of overall performance without cluttering the graph.

High Fidelity Sketch:

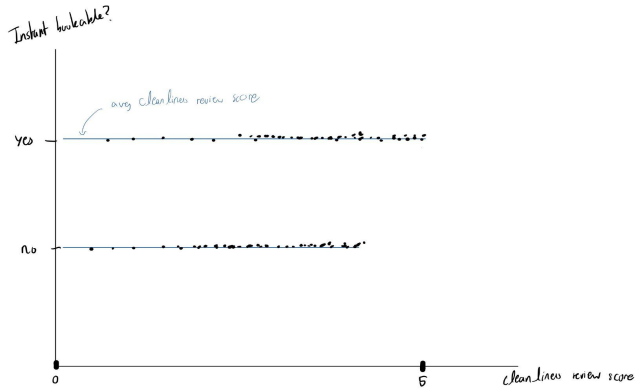


We chose to use Graph 3.3. Using a grouped sorted bar chart is not only effective at comparing the different types of rating in each neighbourhood, but also effective at comparing the overall review score rating between each neighbourhood. We labeled the overall review score rating for each neighbourhood instead of using the dot of over all review score rating in the sketch for accuracy.

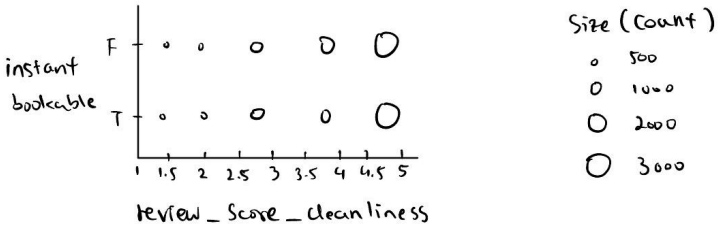
Task: Investigate if listings with instant booking enabled have different cleanliness scores than those that don't
Attribute: Instant_bookable, review_scores_cleanliness
Rationale: Check if the convenience of instant booking affects the cleanliness rating by visualizing the potential trade-offs between booking convenience and property upkeep.



(graph 4.1)



(graph 4.2)



(graph 4.3)

Graph 4.1 (Density Plot with Overlays)
Expressiveness:

- The graph uses colour-coded lines to differentiate between listings with instant booking (red) and those without (blue), along with a black line to denote the average. This distinction is essential for the task at hand.
- The density plots could be expressive in showing the distribution of cleanliness scores for both categories.

Effectiveness:

- The effectiveness is hampered by the overlay of many lines, which can make it difficult to discern individual distributions or identify the central tendencies of each group.

Graph 4.2 (Horizontal Line Graph)
Expressiveness

- The graph displays the average cleanliness as a comparison of the individual review_score_cleanliness.
- We could determine the distribution of the cleanliness_score easily by looking at the density of the points.

Effectiveness:

- The horizontal layout of each individual point makes it hard to track the exact review_score_cleanliness.
- We don't have a convenient way to visualize how many listings of each category are different from the average cleanliness score.

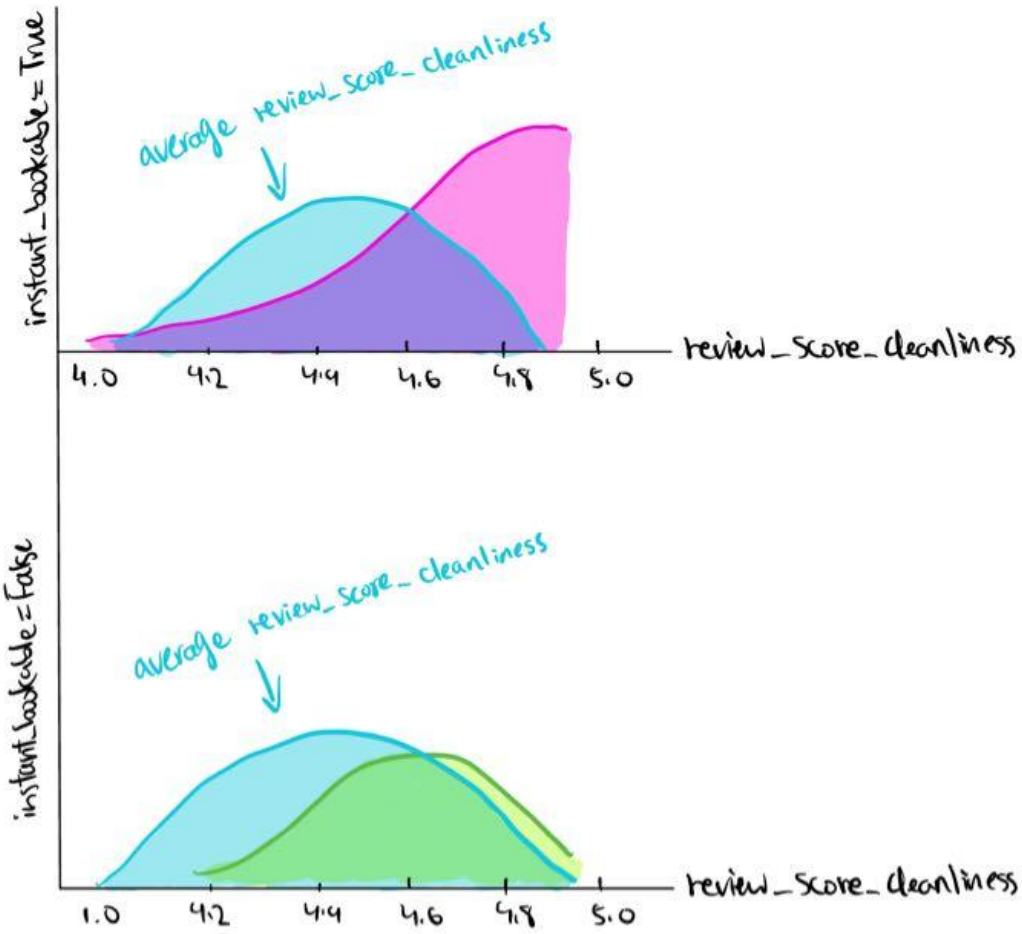
Graph 4.3 (Bubble Chart)
Expressiveness:

- The graph utilizes bubble sizes to indicate the count of listings at each cleanliness score level, differentiated by instant bookable (top) and non-instant bookable (bottom) listings.
- The visual difference in bubble sizes is expressive in suggesting the quantity of listings at specific cleanliness levels.

Effectiveness:

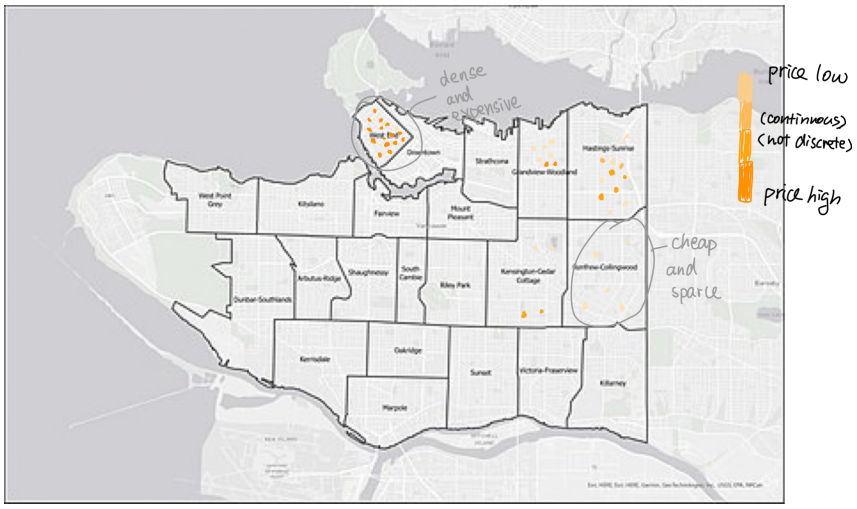
- The chart is effective in showing the distribution of cleanliness scores for both types of listings at a glance.
- However, it does not provide information about the distribution's shape or spread beyond the count of listings. Also, without an exact scale for the bubble sizes, it's difficult to accurately compare the counts between scores

High Fidelity Sketch:

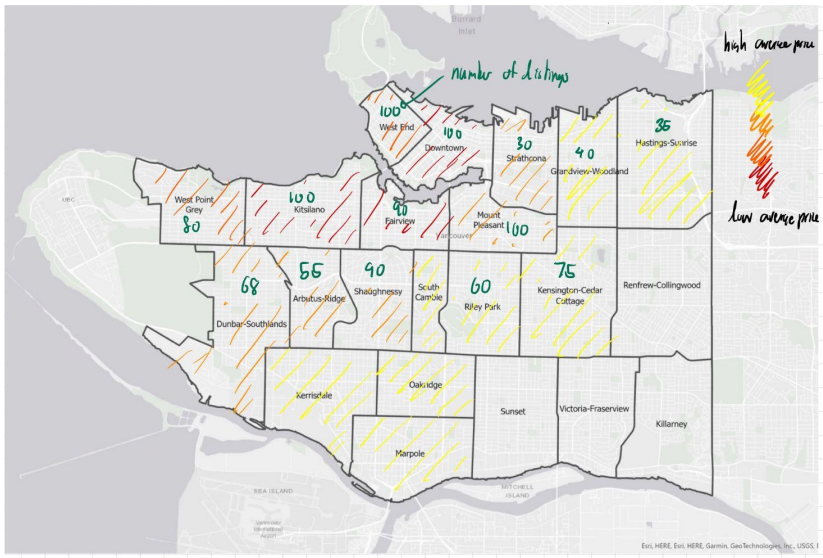


The vertical position indicates if the airbnb is instant bookable or not. And the x-axis indicates the review_score_cleanliness. The color of the average density plot is blue, the color for the instant_bookable = True is red, and the color of the instant_bookable = False is green. Using different colors, we could separately compare the average cleanliness score with the score of whether the airbnb is instant bookable or not. Moreover, by looking at the overlap of the average cleanliness score and the cleanliness score of instant_bookable or not, we could distinguish how different each category is from the mean cleanliness score.

Task: Which neighbourhood has the most listing and the highest price
Attribute: price, neighbourhood_cleansed



Graph 5.1



Graph 5.2

Graph 5.3

Graph 5.1
This is a dot density map, where each dot is a listing, and the color saturation channel encodes the price of each listing.

- It effectively visualizes the distribution of the listings spatially.
- It is not effective to find the extremum because it is not effective at comparison.
- The continuous scale of color saturation to indicate price may not clearly express the actual price levels.
- The density of dots to indicate listing concentration could be misinterpreted without a key—clusters of dots suggest more listings, but it's not quantitative.

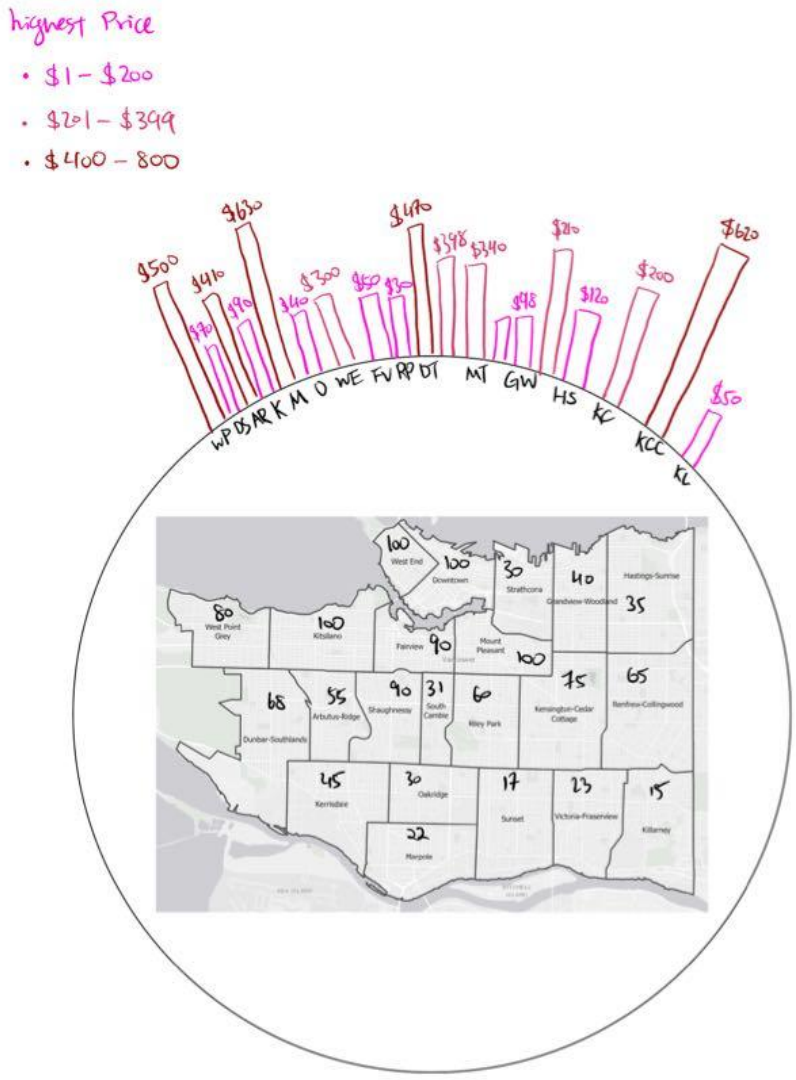
Graph 5.2
This is a choropleth map with color hue encoding the average listing price of that neighbourhood, and a label of the number of the listings in the neighbourhood.

- Uses actual numbers to indicate the count of listings, which directly provides the quantitative data needed for the task.
- Color hue is perceptually unordered and perceptually nonlinear. Using it to encoding a quantitative attribute is not effective.

Graph 5.3
The mark of this visualization are circles, representing neighborhoods. The color hue channel encodes the neighbourhood. The size channel encodes the number of listings. The horizontal position channel encodes the average price of the listings in a neighbourhood.

- Since color hue has a limited number of discriminable bins, using it for nominal attributes with more than 5 bins are not effective.
- The interference between the size of the circles and their horizontal placement can compromise the accuracy with which the average price and number of listings are visualized.

High Fidelity Sketch:



Each neighbourhood has a number of listings in the map, which is easier for comparison in a map layout. And the bar chart on the sphere represents the highest price of a certain neighborhood, labelled by the neighbourhood name. The neighbourhood of the bar chart is displayed according to the position of the neighbourhood on the map. For example, the west point grey neighbourhood is located on the upper left in the map; therefore, “WP” which is an abbreviation of west point grey neighbourhood is displayed in the left of the bar chart. The color represents a price range. By looking at the color, we could get a more distinguished visualization of the price distribution. Using this high fidelity graph, we could clearly compare which neighbourhood has the highest price according to the bar chart, and which neighbourhood has the most number of listings in the map.

Part 4: Next Steps:

Project Feedback (Nov 8th):

1. Rachel will write the dataset detail, prepare for answering question for visualizations that she has covered for milestone 1
2. Chloe will prepare for answering questions for visualizations that she covered for milestone 1
3. Tak will prepare for answering questions for visualization that he covered for milestone 1, and make the overall slide looks prettier.
4. We will start preparing the slides after submitting the milestone 1 on Sunday, then meet up on Monday in class to share what we've got for the slide.

Implementation:

1. Brainstorm on two views (after discussion of the slide on Monday, brainstorm together as a group)
    a. How are we supporting mouse-based interaction *(e.g., hover, click, drag, tooltip, scrollytelling)*
    b. Which widget we can use (eg. dropdown, radio button, range slider, calendar).
    c. What advanced interaction techniques such as brushing and linking or detailed views we can use.
2. Decide which ones to code and which ones to draw. (after class Nov 8th, brainstorm together as a group)
3. Implement visualization with Altair in a Jupyter Notebook (Chloe and Rachel start to do it on Nov 9th).
4. Hand draw at least one high fidelity visualization (Tak starts to do it when the visualization implemented in jupyter notebook is almost done) .

Report:

1. Write the summary of the tasks we focused on (Rachel's task)
2. Put together the final prototype for each view (as well as preliminary sketches) (Chloe's task)
3. Write justification of visualization choices (Tak's task)