GROUP PROJECT - DSCI 320 with Dr. Kemi Ola

Group name: W vizz

Participants: Sahil Vashist, Manav Kakar, Prabhjot Singh

Project Focus (High-Level Goal):

The project aims to provide a comprehensive exploration and understanding of the real estate landscape in Ames City. By leveraging data visualization techniques, the focus is on extracting meaningful insights related to housing trends, investment opportunities, and neighbourhood characteristics.

Intended Audience:

The intended audience for this project includes real estate investors, prospective homebuyers, city planners, and anyone interested in gaining insights into the housing market dynamics in Ames City.

Task 1: Visualize houses in different neighbourhoods on the Ames City Map with options to filter based on Sale Price, Lot Area, and House style.

Goal: Explore the geographical distribution of houses in different neighbourhoods of Ames City.
Benefit: Provides a spatial understanding of where houses are located, helping users identify regions of interest.

Taks 2: Finding the best neighbourhood to invest in by evaluation current mean Sale Price, mean Lot Area and analysing Sale Price trends.

Goal: Evaluate the best neighbourhood for real estate investment by considering mean Sale Price, mean Lot Area, and analysing Sale Price trends.
Benefit: Aids real estate investors in making informed decisions by highlighting neighbourhoods with favourable investment potential.

Task 3: How does Lot Area and number of rooms in a house varies depending on Garage Type and the number of floors in a house.

Goal: Explore how Lot Area and the number of rooms in a house vary based on Garage Type and the number of floors.
Benefit: Provides insights into the relationships between housing attributes, guiding users in understanding preferences and patterns.

Task 4 : Identifying top 5 neighbourhoods based on sales price per unit area and comparing various house styles with their respective conditions.

Goal: Find the top 5 neighbourhoods with the best cost-to-area ratio (cheapest sales price per unit area) and compare various house styles with their respective conditions.
Benefit: Assists homebuyers in identifying cost-effective neighbourhoods and understanding housing styles and conditions in those areas.

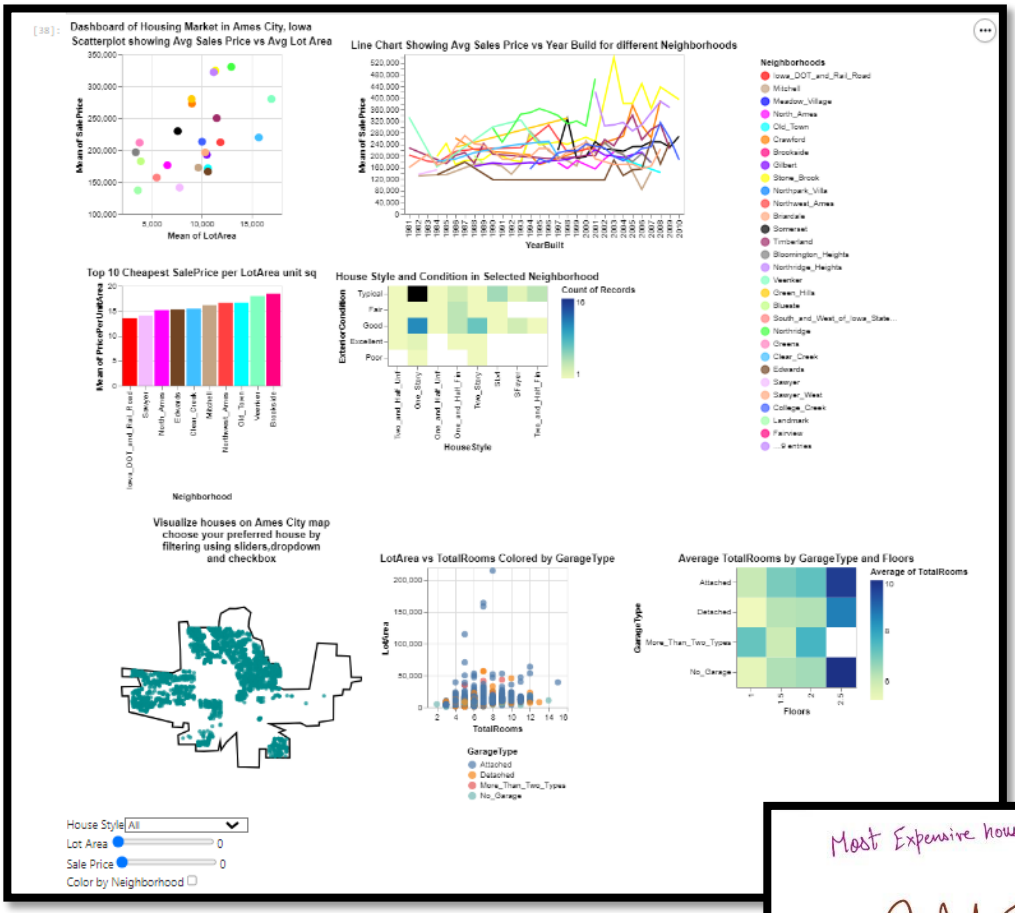
Task 5(Hand sketch): Plotting the Most Expensive Houses in Top 5 Expensive Neighbourhoods

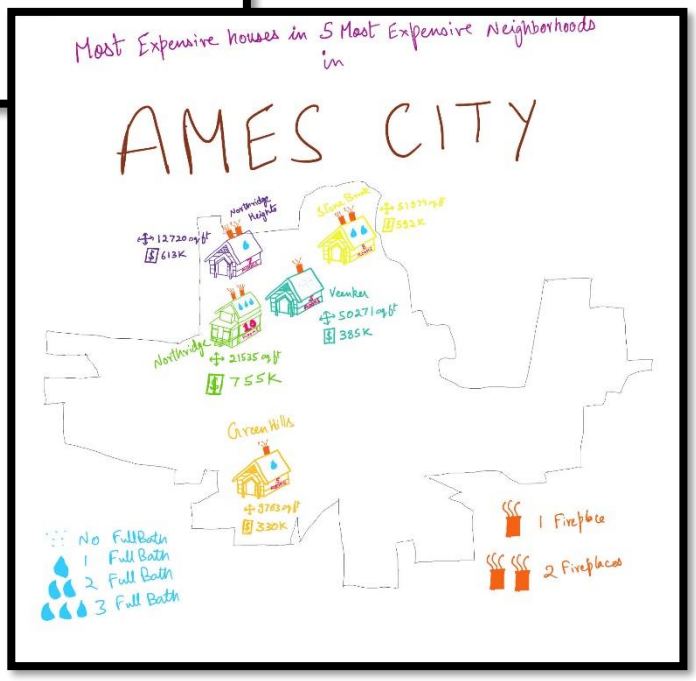
Goal: Identify the top costliest neighbourhoods (with the highest average sales price) and plot the most expensive house in each of those neighbourhoods, displaying detailed configurations.
Benefit: Offers a detailed view of premium real estate locations and the characteristics of the most expensive houses, aiding in luxury home market exploration.

Overall, these visualizations will help one understand:

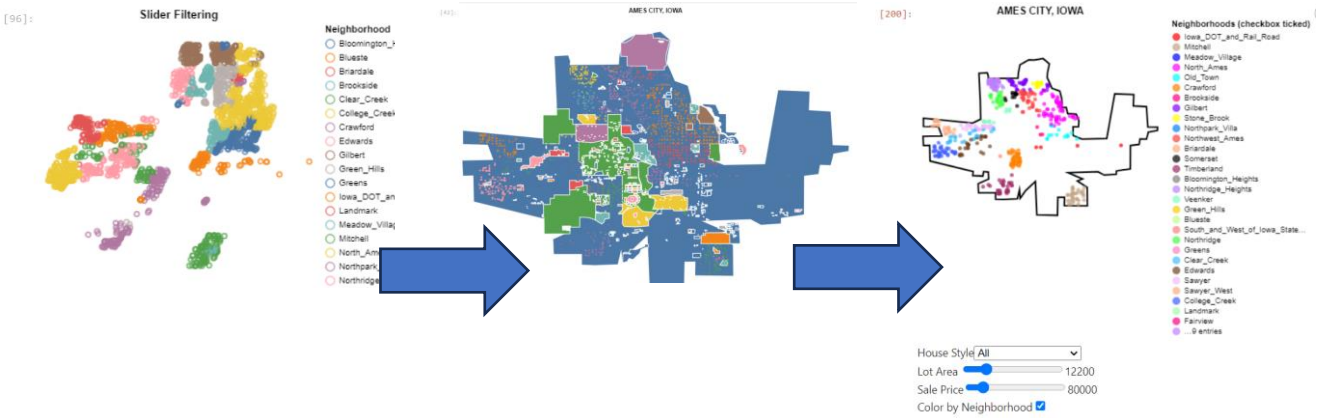
- **Diversity of Neighbourhoods**: Understand the characteristics of different neighbourhoods in terms of pricing, size, and investment potential.
- **Trends and Patterns**: Identify trends and patterns in housing attributes, helping users make data-driven decisions.
- **Cost-Effective Options:** Pinpoint neighbourhoods with favourable cost-to-area ratios, aiding homebuyers in finding affordable options.
- **Luxury Real Estate Market:** Explore premium real estate locations and gain insights into the configurations of high-end houses for those interested in luxury properties.

Interaction that span multiple vizzes: The visualization incorporates sliders for setting Sales Price and lot area, enabling users to filter houses based on their budget and preferences. The dropdown refines searches by house type, while the radio button allows viewing neighbourhoods. These interactions work simultaneously for optimal filtering. In another visualization, selecting a neighbourhood in a scatterplot filters its sales price trends in a line chart, aiding focused examination. Another interaction links a bar chart and heatmap, offering insights into house styles and conditions in neighbourhoods with affordable housing. The scatterplot and heatmap have a bidirectional interaction, allowing seamless selection and highlighting specific house configurations for in-depth market trend exploration. Static visualizations lack this dynamic filtering capability.





TASK 1: Visualize houses in different neighbourhoods on the Ames City Map with options to filter based on Sale Price, Lot Area, and House style.



Map of Ames City: A geoJSON-based map, marked with geoshapes, provides a geographical context for the data. Properties are shown as circles on the map.

Critique and application of theoretical principles:

- Marks: Geoshapes for the map and circles for properties.

- Channels: Spatial channels for property locations; color for neighbourhood distinction.

- Effectiveness: Provides a geographic context for property data.

- Expressiveness: Direct mapping of property locations to real-world geography.

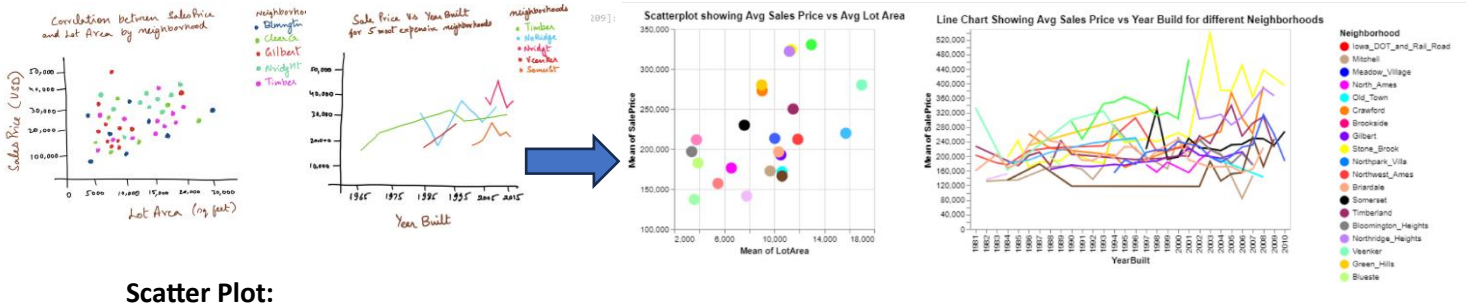
- Map Clarity: While the map provides valuable context, the clarity of individual data points might be compromised in areas with high property density.

Gestalt Principles are applied in the visualization design through techniques like Common Region, where outlined or color-shaded neighborhood boundaries create distinct areas. The Enclosure principle is implemented with subtle lines or shading around neighbourhoods, reinforcing groupings. Figure-Ground principles ensure a clear visual contrast between data points and the map background, facilitating focused attention on properties with essential geographical context. These principles collectively contribute to an effective representation of neighborhood data.

Interaction:

We have utilized sliders to set the Sales Price and lot area. This allows the user to establish their preferred sale amount and lot area, enabling them to filter out all the houses within their set budget and with their preferred lot area. The dropdown allows the user to filter out the type of house they are looking for and also provides an option to display all houses. The radio button allows the user to view separate neighbourhoods in which those houses are located and also provides an option to turn it off. All of the sliders, button and the dropdown work simultaneously to provide the best filtering that the user desires based on their preferences. These interactions are justified as they allow the user to filter housing based on their preferences, making it easier to find their ideal house. The static version of this visualization would not have allowed filtering based on preferences, as all the points would have been simply laid out on the map, with most separated by neighbourhoods and houses using different marks.

Task 2: Finding the best neighbourhood to invest in by evaluation current mean SalePrice,mean Lot Area and analyzing Sale Price trends.



Scatter Plot:

- Marks: Points represent individual neighbourhoods, effectively marking their mean SalePrice and LotArea.

- Channels: Spatial Positioning: Encodes mean SalePrice (y-axis) and mean LotArea (x-axis).

- Color: Distinguishes neighbourhoods, facilitating identification and comparison.

- Effectiveness: Clearly displays a comparative view of neighbourhoods based on key investment metrics.

- Expressiveness: Directly maps quantitative investment data (SalePrice and LotArea) to visual elements.

Time Series Line Chart:

- Marks: Lines represent the SalePrice trend over time for each neighbourhood.

- Channels: Encodes the passage of time (x-axis) and the fluctuation of SalePrice (y-axis).

- Color: Maintains consistency with the scatter plot, linking the same neighbourhoods across both visualizations.

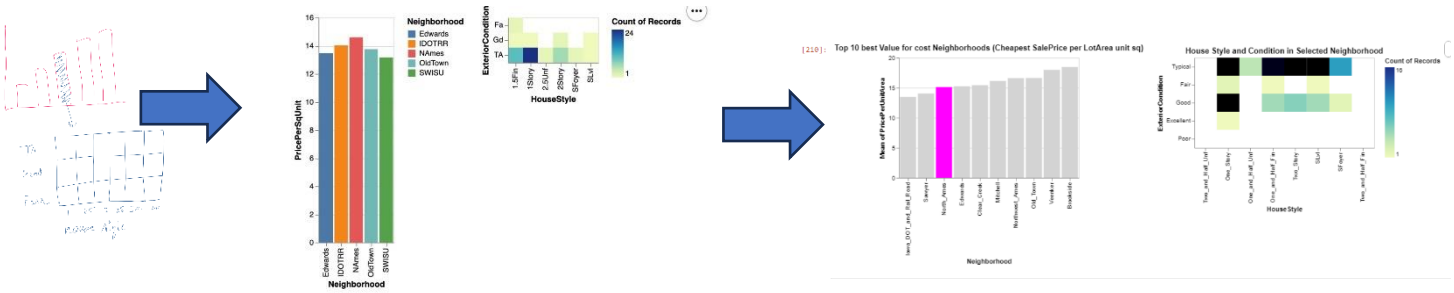
- Effectiveness: Illustrates historical SalePrice trends, providing insights into market dynamics.

- Expressiveness: The line chart format intuitively communicates temporal changes in SalePrice.

The principle of proximity in the scatter plot helps discern relationships between closely positioned points, indicating neighbourhoods with similar property characteristics. Both visualizations benefit from the principle of similarity, as similar colors across plots visually link neighbourhoods, enhancing overall comprehension. Additionally, the Figure-Ground principle ensures a clear focus on the distinct foreground, comprising data points and lines, against the background of the chart area in both visualizations.

From the scatterplot, we have the capability to select a specific neighbourhood, which then filters out that neighbourhood in the line chart. This functionality allows us to observe how the sales price of that particular neighbourhood has evolved over the last 30 years, providing valuable insights into the changes in the housing market for that area. The interaction is justified as it enables us to isolate a particular neighbourhood from the scatterplot, facilitating a focused examination of its trends over time. While it might be challenging to discern a single neighbourhood's line in the line chart without this interaction, it helps us concentrate on it individually and subsequently compare it with different neighbourhoods. Such a nuanced analysis would not be feasible with a static chart.

Task 3: Identifying Top 10 Neighbourhoods Based on Sales Price per Unit Area and Comparing Various House Styles with Their Respective Conditions



Bar Chart:

- Marks: Vertical bars.

- Channels: Spatial positioning on the y-axis for mean price per unit area; color-coded x-axis for neighbourhoods.

- Effectiveness: Clearly displays neighbourhood rankings by price per unit area.

- Expressiveness: Directly maps quantitative values to visual elements. Simple yet effective.

Heatmap:

- Marks: Rectangular markers.

- Channels: Spatial positions on x- and y-axes for house style and external condition, respectively; frequency encoded in color.

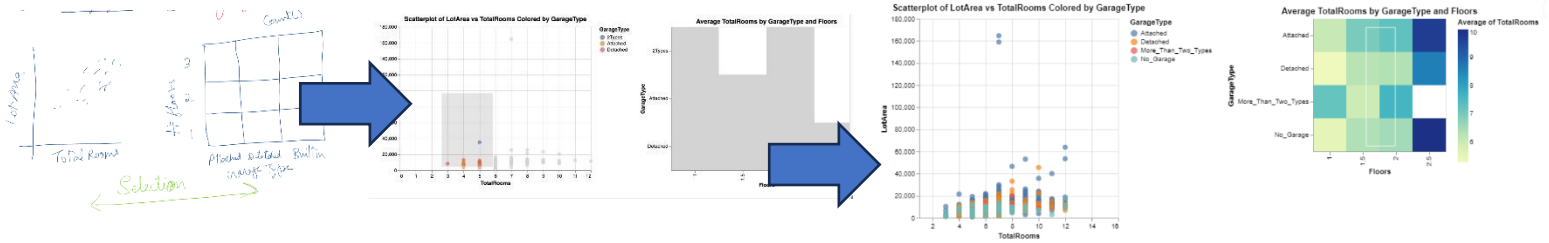
- Effectiveness: Illustrates distribution of house conditions and styles effectively.

- Expressiveness: Color intensity variations provide quick insight into frequency.

Proximity and common region principles are evident in closely positioned bars on the same axis, indicating shared categories. The continuous arrangement of bars contributes to a narrative of neighbourhood rankings, supported by the orderly presentation facilitated by the symmetrical grid layout. Additionally, the figure and ground principle are applied through different color intensities, distinguishing individual data points from the overall heatmap, ensuring clarity in visualization. Improving legend clarity with explicit instructions on the color scale would enhance usability.

The interaction between the bar chart and the heatmap allows us to select a single neighbourhood from the bar chart and observe the count of different styles of houses in that neighbourhood, along with their external conditions. This interaction streamlines the process of gathering information about various types of neighbourhoods within limited space. Additionally, it aids in identifying possibly why these neighbourhoods have the most affordable houses per square foot, investigating the number of smaller one-story houses and houses in poor condition, providing a comprehensive view of the housing market. While a static visualization could potentially achieve this, it would require a considerable amount of space to display separate graphs for numerous neighbourhoods, making it less convenient for analysis and exploration.

Task 4: How does Lot Area and number of rooms in a house varies depending on Garage Type and the number of floors in a house.



Scatter Plot:

- Marks: Circle marks represent individual data points.

- Channels: Spatial positions encode 'TotalRooms'(x-axis) and 'LotArea'(y-axis), and color differentiates 'GarageTypes'.

- Effectiveness: Provides a direct and effective representation of individual data points, allowing for immediate perception of correlations and clusters.

- Expressiveness: Directly maps quantitative variables to spatial positions, making it intuitive to interpret relational patterns.

Heatmap:

- Marks: Rectangular markers indicate aggregated data points.

- Channels: Spatial positions encode 'Floors'(x-axis) and 'GarageType'(y-axis), and color intensity represents average 'TotalRooms'.

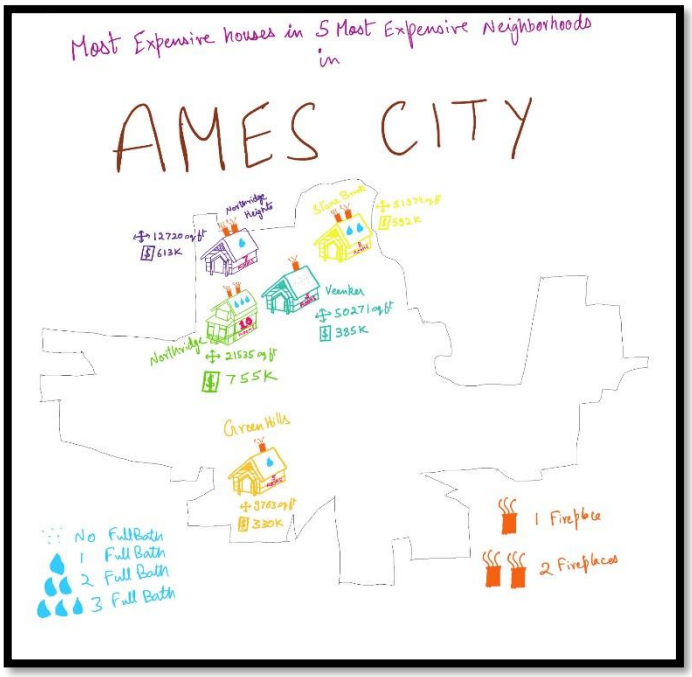
- Expressiveness: Utilizes color intensity for quick, pre-attentive processing of average values, aiding in rapid comparison across categories.

- Effectiveness: Effectively handles data density, presenting information in a compact form.

The Scatter Plot effectively displays individual data points but may face challenges with clutter, particularly in densely populated areas. The Heatmap, utilizing the average 'TotalRooms,' offers a broad overview but risks oversimplifying data, potentially obscuring outliers or unique distributions within categories. In the Scatter Plot, the principles of Similarity and Proximity are evident, with points of similar colors and close proximity indicating related data and emphasizing groupings. The Heatmap employs the Figure-Ground principle through different color intensities, aiding in distinguishing the foreground (data) from the background (grid) and facilitating focused examination of specific data segments. Lastly, the Scatter Plot's continuous arrangement of data points contributes to the representation of trends and relationships.

The scatterplot and the heatmap exhibit a bi-directional interaction, allowing for seamless selection between them. This functionality aids in selecting houses with the user's preferred lot area and number of rooms in the scatterplot, with the corresponding floor and type of garage information displayed in the heatmap. Conversely, users can investigate how the trends in total rooms and house area change with varying numbers of floors and types of garages by making selections in the heatmap, with the corresponding houses highlighted in the scatterplot. This interactive approach enhances our ability to comprehend market trends by facilitating the exploration of specific house configurations. Static visualizations, in contrast, lack the capacity to dynamically select and filter.

Task 5(Hand sketch): Task 5:Identify and plot the top 5 costliest neighbourhoods by average sales price. Create a sketch displaying the most expensive house in each neighbourhood, highlighting key configurations including house style, sales price, lot area, number of rooms, fireplaces.



Spatial House Map plotted on the outlines map of Ames City Iowa

- Marks: Houses

- Channels: Number of floors based on the actual number of floors, Fireplace icon for fireplaces, water icon for number of full-baths, colour for neighbourhoods. Also tells lot area and total rooms as text, position on map as approximate latitude and longitude.

- Effectiveness: Displays House Attributes accurately.

- Expressiveness: Directly maps quantitative values to visual elements.

The challenge of difficulty in portraying detail arises due to the increasing complexity of encoding various attributes in a single house figure. To address this, Gestalt Principles are applied in the visualization. Proximity and Common Region principles are implemented by placing houses on a map, allowing for a more organized arrangement. The Figure & Ground principle is employed through different intensities in house structures, aiding in distinguishing individual data points from the overall map background. These principles enhance the overall clarity and organization of the visualization, mitigating the difficulties associated with portraying intricate details within the house figures.

We tried making it as visually appealing as possible while still encoding useful information.

REFLECTION

3 strengths of your project:

- Includes a lot of different interaction including sliders, radio button, dropdown, unidirectional linking, bi-directional linking. All the tasks have interactive visualizations.
- We have encoded a lot of different attributes (encoded 11 attributes in all our tasks combined). This helps provide even better understanding of the Ames housing market.
- We spent a lot of time in data wrangling, and we have a very good dataset and around 2972 values with no missing values in any columns. On top of that each of our row can be plotted on the geographic map. This helped us get good and reliable visualizations.

2 weaknesses of your project:

- We could have used color better. Some of our visualization lack color harmony principle.
- We could have had 3 or 4 visualizations interacting with each other i.e., changing something in one viz changes something in at least 2 more viz. Right now, we only have bi-directional linking.

2 things you would have done differently:

- We could have paid more attention in framing the task questions. It felt we just wanted to make visually appealing visualizations; however, we weren't solving any question/task meaningfully through that. So, we had to change a lot of visualizations and tasks due to this.
- Not leaving things to last minute. One of our plots which we had to code last minute included boxplot interactions however we didn't know about the bug in Altair 5 that does not let you select individual boxplots in a boxplot chart, so we had to change our idea.

WORK DISTRIBUTION

	HIGH	LEVEL	OVERVIEW	
Project Milestone	Sahil Vashist	Prabhjot Singh	Manav Kakar	
Milestone 1	48	42	10	
Milestone 2	50	42	8	

Breakdown

Milestone Description	Sahil Vashist	Prabhjot Singh	Manav Kakar	Percentage of Total Work for the deliverable
Milestone 1 Task 1 - Finding Data	40	40	20	5
Milestone 1 - EDA & Data Wrangling	60	40	0	30
Visual Summaries	40	60	0	10
Formulating tasks	65	35	0	20
Visualizations	45	40	15	30
Next steps	50	40	0	5
MileStone 2				
Re-evalutaing tasks	70	30	0	10
Novel Sketch	40	30	30	8
Implementation	50	40	10	50
Analysis and critique of viz in Report	50	50	0	30
Reflection	60	40	0	2