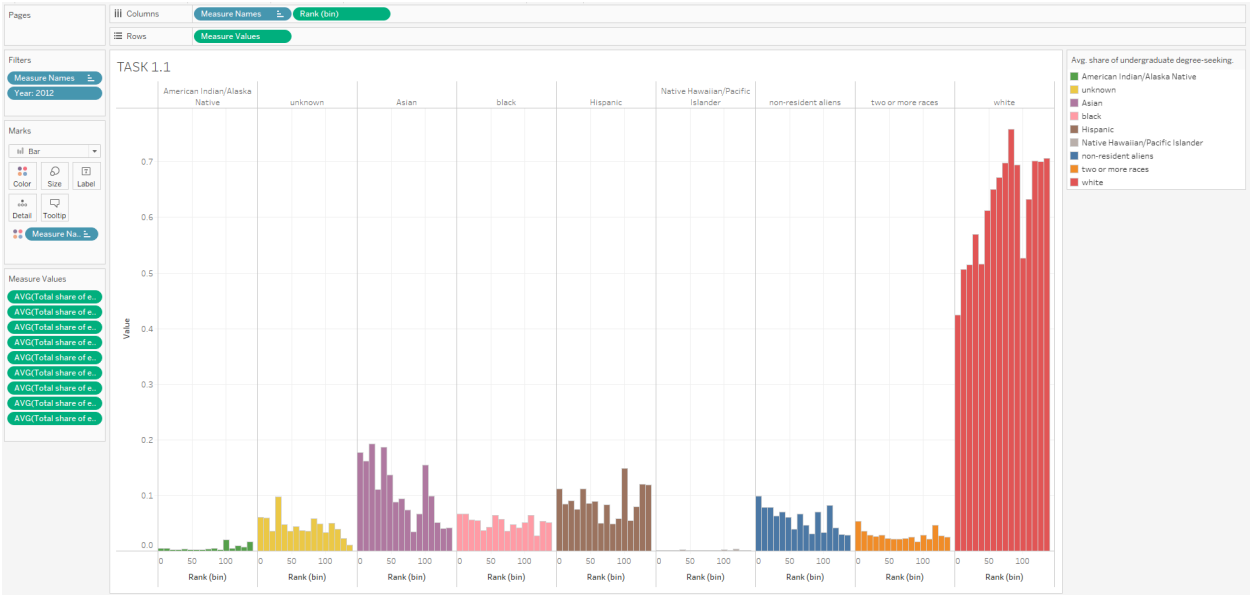
Project Deliverable 1

TASK 1

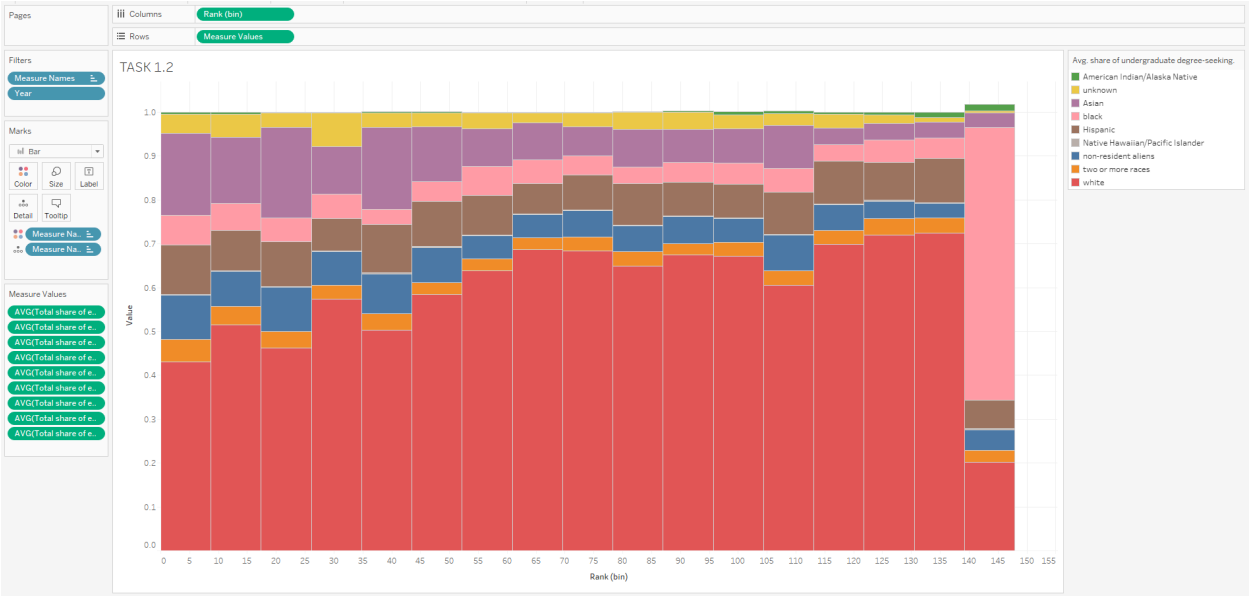
How the diversity percentage among undergraduate students varies across all ranked universities.

Task 1.a - Bar Chart



This bar graph represents the average percentage of each race category across all university ranks(bin). We can easily identify the majority race here which is ‘white’ across all ranks. This helps us to see trends in each race across the ranks but identification of one rank(bin) within all the races is difficult in this plot.

Task 1.b - Stacked Bar Chart



This stacked bar chart presents each university's diversity percentages as segments of a whole, where each segment's length is proportional to the subgroup's size. This visualization effectively shows the relative size of each race across all ranks of the university.

Critique:

In both the visualizations we are trying to compare the diversity across all the ranks of the university. In the bar graph we are using a design similar to juxtaposition and in the stacked bar chart we are using superposition by placing all the races in the same space. In the stacked bar chart it is easy to visualize the part of the whole relationship compared to the bar chart as we are comparing diversity with respect to rankings.

From the stacked bar chart we can see that top-ranked universities show a broader range of diversity compared to low-ranked universities. As we move from top-ranked universities to low-ranked universities, the percentage of one race (white) increases and the other race's percentage decreases.

Effectiveness: In the bar chart we have mapped all the important attributes to all the highly ranked channels (Position for race, percentage, and ranks(bin)), and it satisfies the effectiveness principle. In the stacked bar chart we have mapped the percentage of races to the 'Area' channel while it is ranked low compared to position. That's why it is not very effective. But shows part of the whole relationship between races across all ranks(bin).

Expressiveness: The stacked bar chart accurately expresses the relative proportions of the students from each race using the bar heights across ranks. However, in the bar chart the portion of individual race is clearly plotted but comparison of races across the ranks is not easy.

Separability: The bar graph has better separability as it allows viewers to distinguish and compare each demographic group's percentages independently from the others. In contrast, the stacked bar chart's separability is limited by the overlaying of data, which can lead to difficulty in isolating information about individual groups.

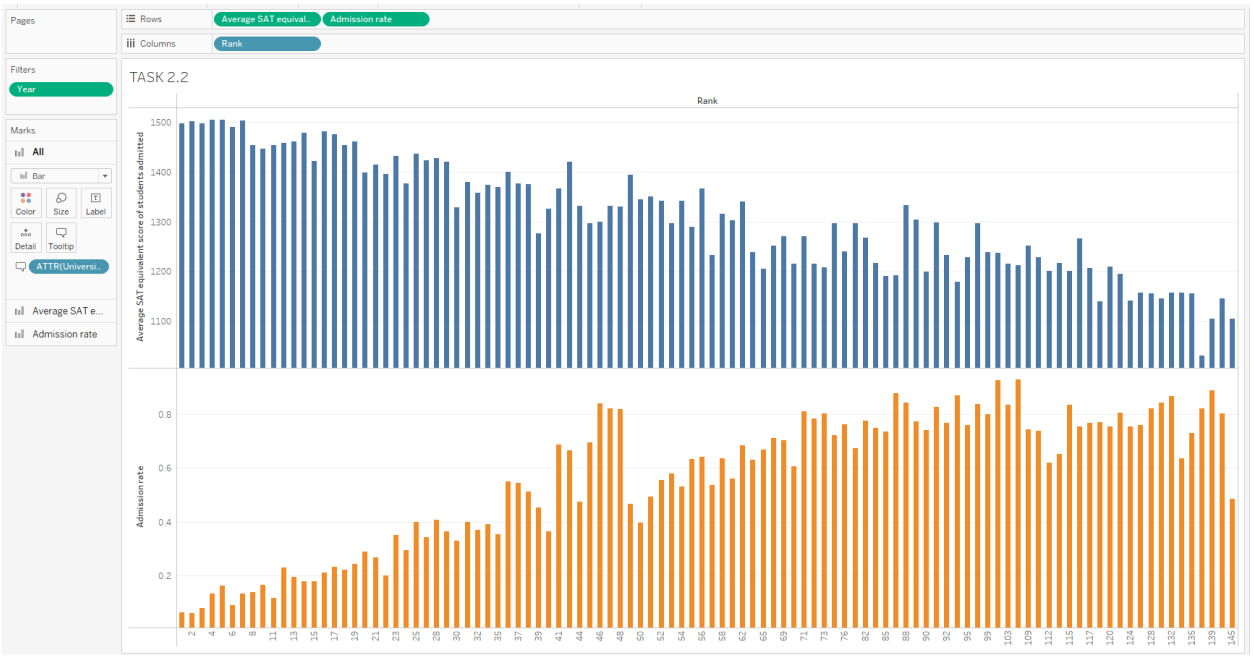
Conclusion:

Overall the stacked bar chart is more effective for comparing diversities across ranks as it represents the part of the whole relationship and emphasizes the intended task of comparison.

TASK 2

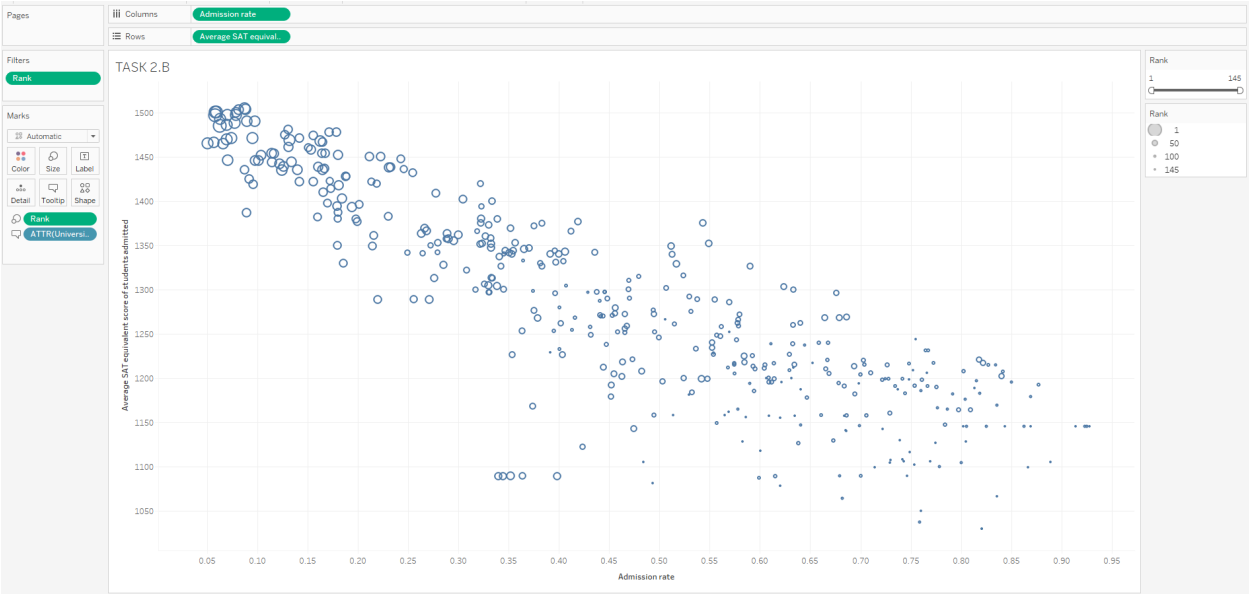
How do university rankings change with respect to Admission Rate versus the average SAT score?

Task 2.a - Bar Chart



This visualization represents the changes in university rankings based on Admission Rate and average SAT scores using a bar chart. The bar chart presents data in a way that represents average SAT scores and admission rates as two distinct sets of bars, making it easy to compare these two variables for each university individually.

Task 2.b - Scatter Plot



This visualization represents how university rankings correlate with Admission Rate and average SAT scores using a scatter plot.

The scatter plot places each university as a point on a two-dimensional plane defined by SAT scores and admission rates. This design excels in showing the correlation between the two variables, making it highly effective for this task. It is expressive in its ability to display a large number of universities on a single plot while showing the trend of the relationship.

Critique:

Considering the comparison between scatter plots and bar charts, for the analysis of how university rankings shift concerning Admission Rate versus the average SAT score, a preference leans towards the scatter plot over the bar chart.

Effectiveness: When comparing the average SAT score and admission rate across university rankings, scatter plots work well at representing the relationship between

continuous variables. The attributes are encoded using position. Based on the effectiveness ranking of channels, these are the most effective magnitude channels. Scatter plots offer a clearer representation of the relationship between these variables than bar charts, which may clutter the visual with multiple rank categories.

Expressiveness: The ability of the scatter plot to visually relate individual data points offers an in-depth understanding of the relationship between average SAT score, admission rate, and university rankings. Bar charts, on the other hand, could find it difficult to represent every aspect of this relationship resulting in oversimplifying the data.

Separability: The separability for the bar chart is high, as the two metrics are clearly delineated, but this separation can hinder the understanding of their relationship. For the scatter plot, the separability is lower because it presents both variables in a single unified view, which is advantageous for analyzing the relationship but may make it harder to evaluate each variable independently.

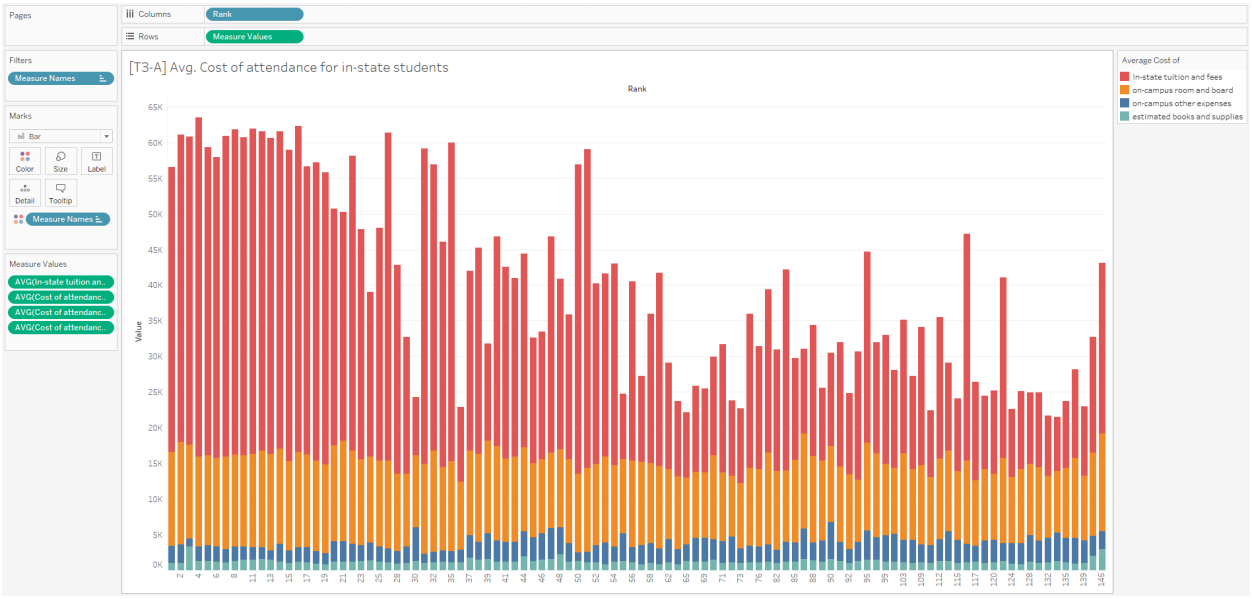
Conclusion:

Overall, the scatter plot is the most effective choice for visualizing the relation between admission rate, average SAT score, and university rankings because of its effectiveness in displaying the relationship between continuous variables and its ability to convey details of university rankings.

TASK 3

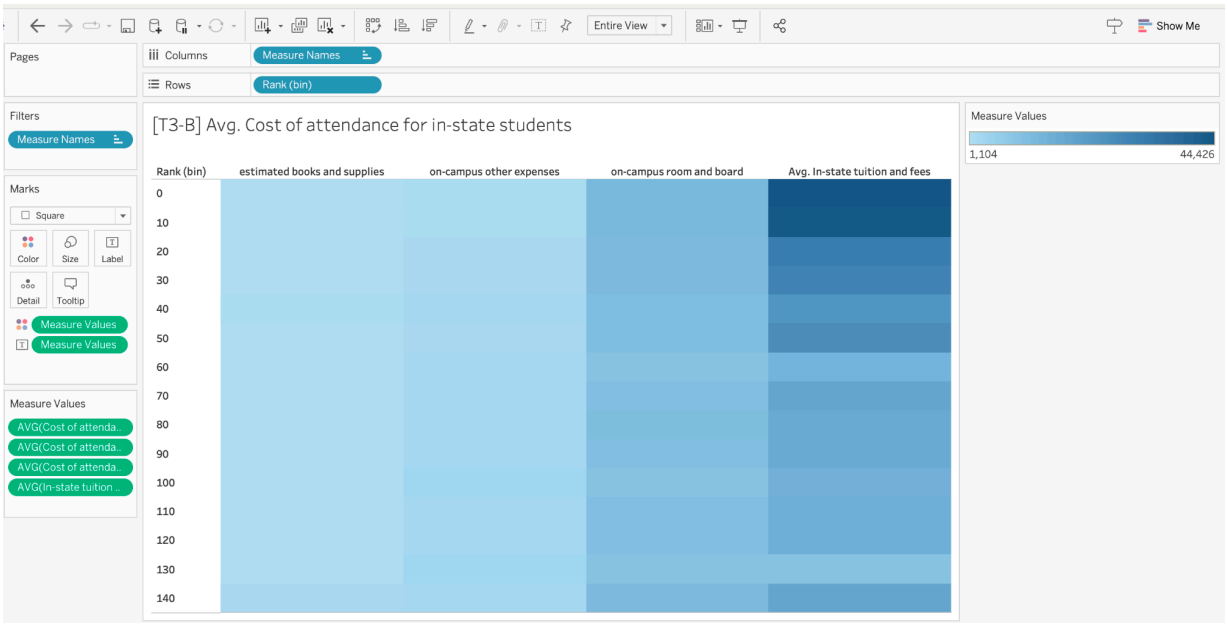
How do the various components of university costs contribute to the overall expenses of in-state students across different ranking tiers?

Task 3.a - Stacked Bar Chart



The stacked bar chart represents each university as a bar, with the overall height of the bar indicating the total cost of attendance. Each bar is subdivided into colored segments proportional to the costs they represent: estimated books and supplies, on-campus other expenses, on-campus room and board, and average in-state tuition and fees. The universities are ordered along the horizontal axis by their ranking tier.

Task 3.b - Heatmap



The heatmap uses color to represent the magnitude of costs, with each cell corresponding to a cost component for a particular university. The universities are ordered vertically by their ranking tier, and the cost components are laid out horizontally. The color saturation within each cell reflects the cost amount, with darker or more intense colors indicating higher costs.

Critique

Effectiveness: The stacked bar chart effectively communicates the total cost and the proportion of each cost component within that total. This makes it possible to compare the total costs as well as individual components across rankings. Based on the effectiveness ranking of channels, the best channels are used, hence satisfying the effectiveness principle. The heatmap quickly communicates the relative magnitude of costs across different universities and cost components through color intensity, which is not the best channel and does not satisfy the effectiveness principle.

Expressiveness: The stacked bar chart is expressive in showing the relationship between the individual cost components and the total cost. The meaning of the colors and segments is clear and intuitive. While the heatmap is less expressive when it comes to understanding the precise value of each component or how they add up to the total cost, as there are no clear boundaries or segments.

Separability: While the total costs are easy to compare at a glance in the stacked bar chart, the individual segments can be difficult to compare between universities, especially if they are thin or similar in size. In heatmap, it is difficult to separate and compare exact values for each component, especially between universities with similar cost structures.

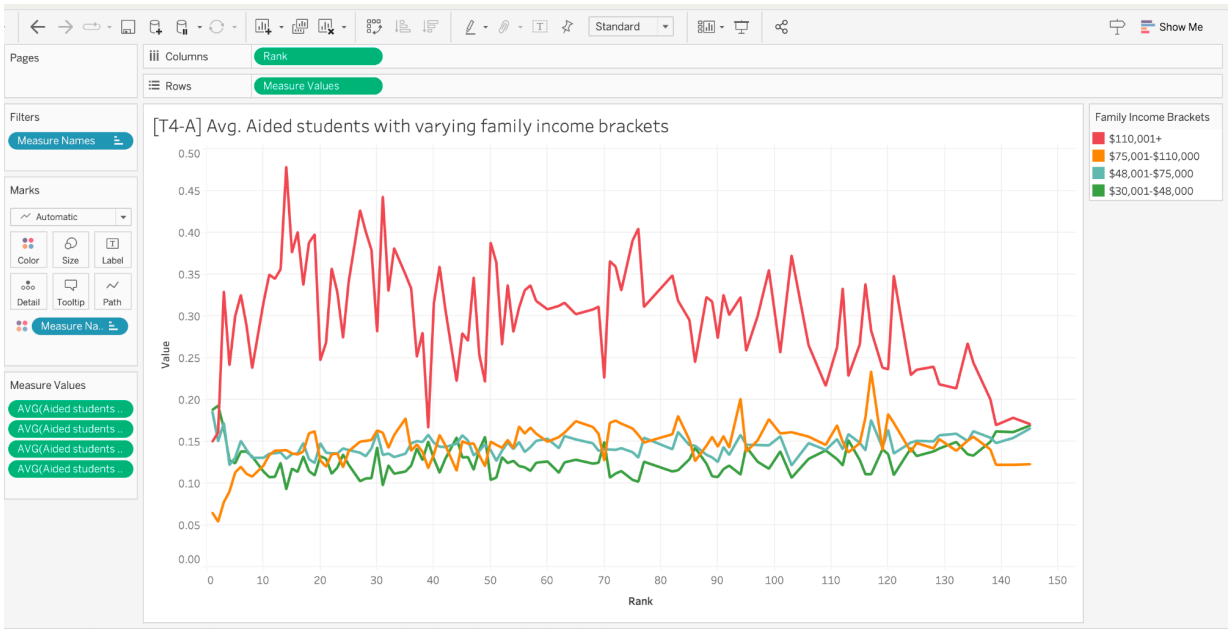
Conclusion

Between the two, the stacked bar chart is superior in terms of expressiveness and separability for this specific task (T3). It more clearly communicates the exact contribution of each cost component to the total and allows for comparison of these contributions across the ranking tiers.

TASK 4

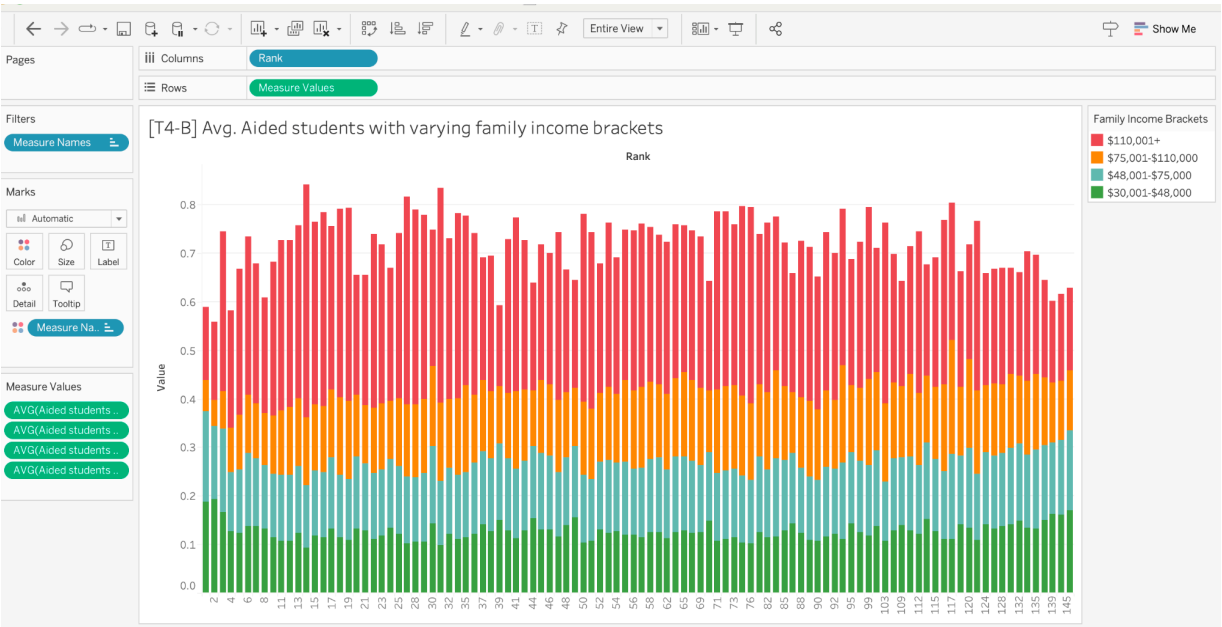
How do the financial aid profiles of universities, including the distribution of aid across different family income brackets relate to their rankings?

Task 4.a - Line Graph



A line graph is used to represent the average proportion of aided students across different income brackets at universities with varying ranks. Each income bracket is represented by a different colored line, and the x-axis shows the university ranks.

Task 4.b - Stacked Bar Chart



The stacked bar chart shows the composition of financial aid across different income brackets for each university rank. Each income bracket is represented by a different color, and the bars are segmented proportionally to show how much aid goes to each bracket within a given university rank.

Critique

Effectiveness: Even though the line graph is effective in showing trends, making it easier to spot general trends and outliers for each income bracket, the overlap in the lines makes it difficult to discern individual trends clearly. The stacked bar chart is more effectively displaying the distribution of aid for each rank in a way that is easily comparable across ranks and also based on the effectiveness ranking of channels, the best channels are used, hence satisfying the effectiveness principle.

Expressiveness: The stacked bar chart is more expressive in expressing the part-to-whole relationship of aid distribution within each rank. The line graph with multiple overlapping lines is less expressive.

Separability: The separability of the line graph suffers due to the potential for line overlap. Whereas, the stacked bar chart allows for clear separability of data for individual ranks.

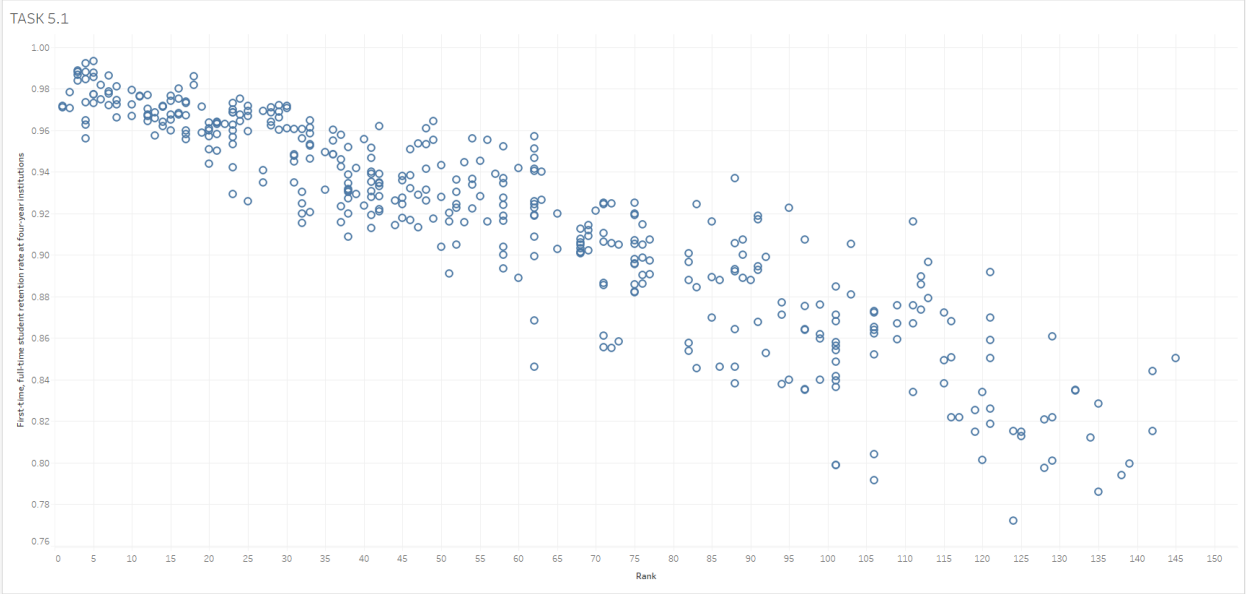
Conclusion

Overall the stacked bar chart offers a better overall visual assessment of the distribution of aid across ranks than the line graph.

TASK 5

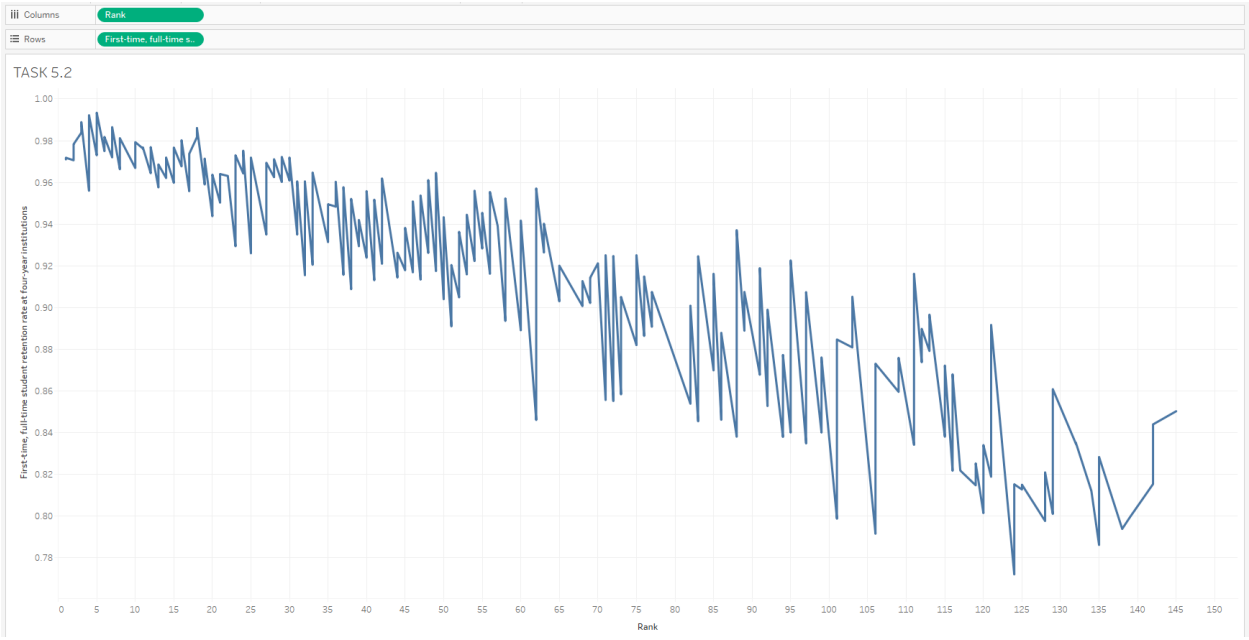
How does the retention rate of first-time, full-time students at four-year institutions vary across universities with different rankings?

Task 5.a - Scatter Plot



The scatterplot maps the retention rates of first-time, full-time students at four-year institutions against their respective university rankings. Each data point on the graph represents an individual university, with the horizontal axis (X-axis) indicating the rank of the university and the vertical axis (Y-axis) representing the retention rate.The plot reveals a wide distribution of retention rates across various ranks, with a slight tendency for higher-ranked universities to have higher retention rates.

Task 5.b - Line Chart



This visualization shows the relationship between retention rates of first-time, full-time students at four-year institutions and university rankings through a line graph.

Critique:

Effectiveness: The scatter plot displays individual universities as points, allowing for the identification of outliers and patterns across the rank continuum. It effectively shows the distribution and variance of retention rates and also satisfies the effectiveness principle. The line chart connects points along the rank axis, which could imply a sequential relationship that doesn't actually exist between university ranks. It is less effective for showing individual university data

Expressiveness: The scatterplot expresses the relationship between two quantitative variables: rank and retention rate. Each point represents a university, and the axes clearly communicate the data. The line chart suggests continuity in the data that may not be present, which could lead to misinterpretation and misleading trends or relationships between the ranks of universities, violating the principle of expressiveness.

Separability: The scatterplot offers better separability as each university's data is represented as a distinct point, which, barring overplotting issues, allows for individual

assessment without the influence of adjacent data points. In contrast, the line chart diminishes separability by emphasizing the trend over individual data points.

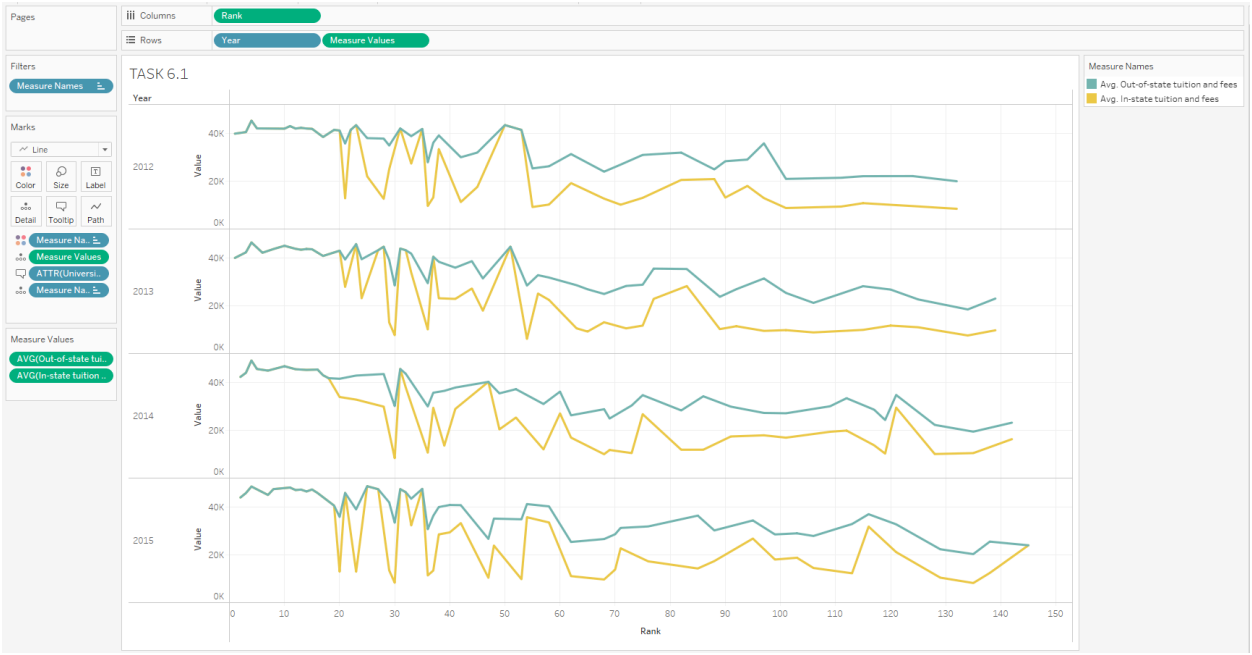
Conclusion:

For task T5, the scatterplot is superior because it more effectively and expressively represents the variability of retention rates across universities with different rankings and maintains separability between data points to a reasonable extent. It adheres well to the principle of expressiveness by accurately representing the nature of the data without implying relationships that don't exist.

TASK 6

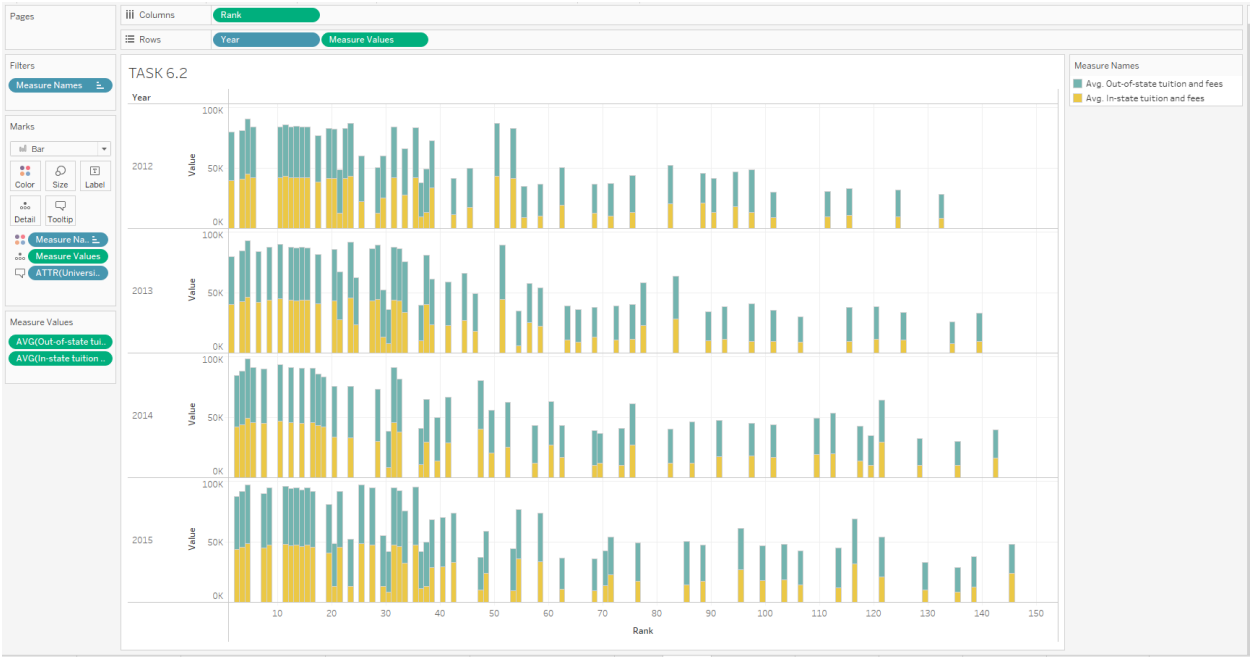
How do the in-state and out-of-state fees of universities vary in relation to their ranks over the years?

Task 6.a - Line Chart



This visualization represents the relationship between university fees and ranks for four years, showcasing distinct trends for in-state and out-of-state fees across ranks through line charts.

Task 6.b - Bar Chart



This visualization displays how the cost of fees at various universities has evolved over the years, organized by university rank. X-axis represents the university ranks, and the bar graphs within each row illustrate the variation in fees across different years. This representation enables a straightforward comparison of fee trends based on the ranking of universities, offering insights into how fee structures have changed over time within different ranks(bin).

Critique:

Effectiveness: The line chart is more effective than a bar chart because it shows how university fees change across ranks and fee types over time. It's easier to see gradual trends with lines compared to separate bars. This clarity helps us understand how fees change and relate to each other over different periods and ranks. Line chart shows how top ranked universities have the same tuition fee for both in-state and out-state students. Overall, the line chart is better for getting a clear picture of fee trends in universities.

Expressiveness: The line chart is better at showing fee trends over time and ranks compared to bar graphs. Its lines connect the fee data points, making it easy to see patterns and changes. This makes it more useful for understanding how fees change in universities. On the other hand, bar graphs might not show these trends as clearly because

they separate the data into bars. Overall, the line chart is a clearer and more helpful way to visualize university fee trends.

Separability: The line chart clearly separates in-state and out-of-state fees, making it easy to analyze each fee type's trends over time. On the other hand, bar graphs may not show this separation as clearly because they combine different fee types into single bars, which can make it harder to see each fee type's individual trends over the years.

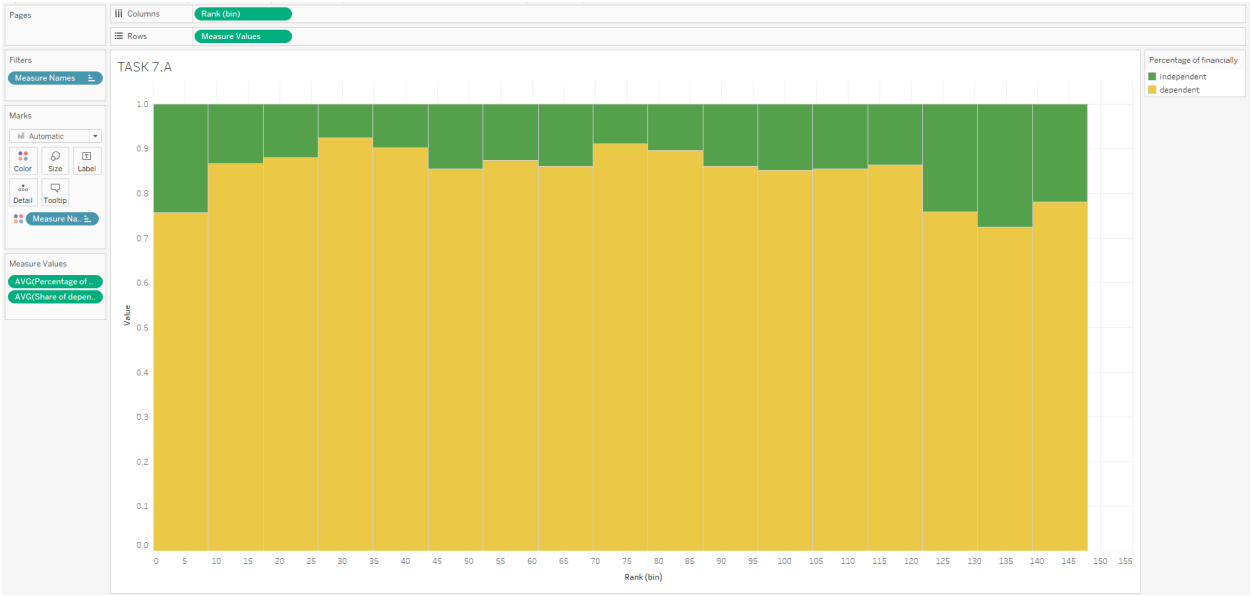
Conclusion:

In conclusion, the line charts used in Task 6 prove to be a better visualization choice due to the clear depiction of distinct trends between in-state and out-of-state university fees over time. This visualization effectively showcases the independent evolution of these fees, enhancing our understanding of their dynamics and trends. Such insights are valuable for decision-makers and stakeholders in comprehending the factors shaping university fee structures and their implications for students and institutions.

TASK 7

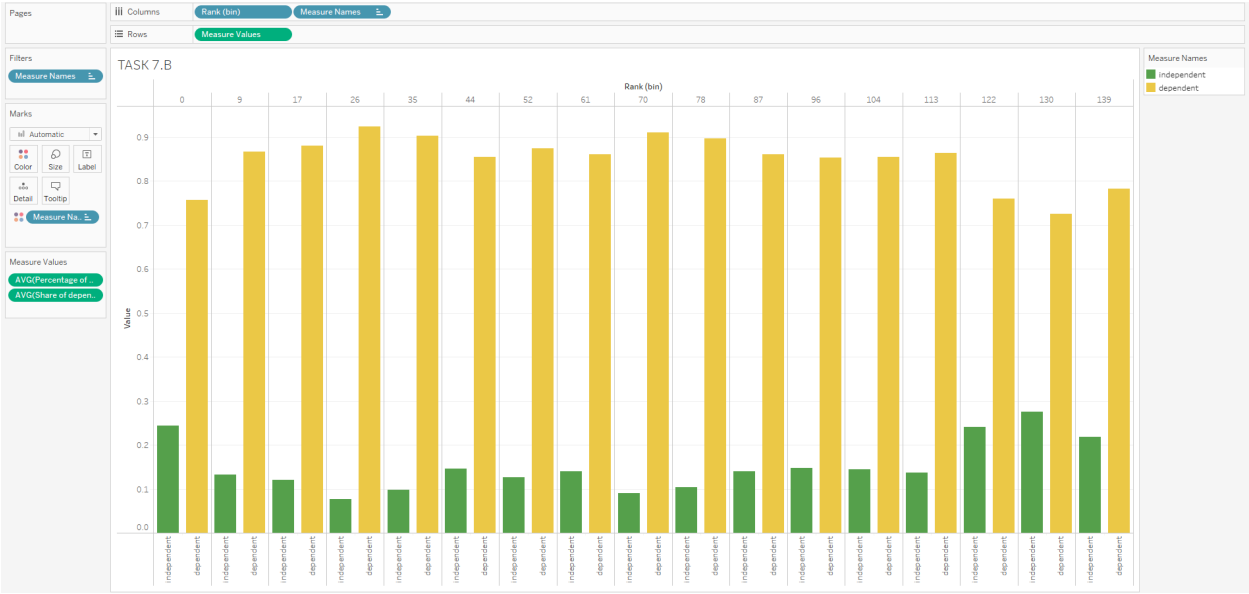
Compares the proportion of independent and dependent students across all ranked universities

Task 7.a - Stacked Bar Chart Scaled to 100%



This stacked bar chart visualizes the distribution of independent and dependent students across ranked universities. Each column (X-axis) represents the ranks(bin) of the universities category, while the rows (Y-axis) depict the proportion of independent and dependent students. The height of the stacked bars in each row shows the percentage of each student type at universities within that rank category, enabling a comparison of student demographics across different ranks.

Task 7.b - Side-By-Side Bar Chart



This visualization shows how many students are dependent versus how many are financially independent at different ranked universities. The columns show university ranks, and each row represents either dependent (yellow) or independent (green) students. It helps see how student financial status changes across different ranked universities.

Critique:

Effectiveness: This Side-by-side bar chart satisfies the effectiveness principle, because for all the attributes (independent, dependent, ranks(bin), and share) we are using position channel. Position is a highly-ranked channel. On the other hand, stacked bar charts use area or length to represent share, which is ranked lower in terms of effectiveness compared to using position directly.

Expressiveness: For this question we have to compare the proportion of independent and dependent students across all ranked universities, which we can clearly infer from the scaled stacked bar chart because we have only two categories and stacked bar charts with fewer categories are best to show part of the whole relationship. However, in side-by-side the bar chart it is easy to identify the share of independent or dependent students for each bin of rank but it is not expressive for showing the portion of each category across the ranks(bin).

Separability: Side-by-side bars do this well by showing dependent and independent students separately for each university rank. Stacked bars combine these groups, making it easier to compare them directly with respect to ranks(bins). So, this stacked bar chart helps us quickly understand how student types vary across university ranks.

Conclusion:

For the Task 7 In conclusion, this scaled stacked bar chart is more effective for comparing independent and dependent student proportions across university ranks. They effectively show the portion of financially dependent and independent students across university rankings. This Scaled stacked bar chart provides a straightforward view of how student demographics vary across different university ranks(bins).

CONTRIBUTION STATEMENT :

For this deliverable, we structured seven tasks, with each task being tackled by a pair of team members. This approach allowed us to help each other and develop suitable visualizations through shared expertise.

Task Assignments :

Task 1 : Denil Jain and Anurag Ramadugu
Task 2 : Shalini Sharma and Anurag Ramadugu
Task 3 : Jyothsna Kaamala and Shalini Sharma
Task 4 : Jyothsna Kaamala and Denil Jain
Task 5 : Shalini Sharma and Jyothsna Kaamala
Task 6 : Denil Jain and Anurag Ramadugu
Task 7 : Denil Jain and Jyothsna Kaamala

Throughout this Assignment, we held regular team meetings to review each member's progress and provide constructive critiques that enhanced our final outputs. We also used a shared document to track contributions and revisions, ensuring transparency and collective responsibility in our project development.

- Document Link: Contribution-Tracking

The deliverable-1 report was collaboratively written; team members responsible for each task elaborated on their techniques, offered critiques, and justified their design choices and established principles of data visualization. This collaborative effort not only ensured the quality of our visualizations but also enriched our learning experiences through peer feedback and shared problem-solving.

Signed By :

1. Anurag Ramadugu - ar2697
2. Denil Jain - dj325
3. Jyothsna Kaamala - jk734
4. Shalini Sharma - ss4923