Aakash Suryavanshi and Jason Nair

Final Project Write Up

1) Link to video –
   - Used College.csv data set

2) What is the title of your visualization?
   - College Student Population by Region

3) Why did you choose this particular visualization?
   - Bar Chart
     - The bar chart was chosen because it effectively utilizes length, a preattentive attribute, to visually encode the number of students at each college. This makes comparisons of student populations across colleges within the same region both easy and intuitive.
     - The horizontal layout ensures scalability and clarity, aligning with class concepts of effective data encoding for categorical and quantitative variables.
     - Colors enhance visual distinction between the colleges, making it easier for users to differentiate them at a glance.
   - Pie Chart and Tooltip
     - The pie chart and tooltip were added to provide detail on demand, allowing users to explore the racial demographics of individual colleges without overwhelming the bar chart with too much information.
     - The pie chart was designed for users who prefer visual representations, while the tooltip caters to those who prefer precise numerical details. Together, these interactive elements ensure the visualization caters to a wide audience.
   - Concept
     - The concept behind the visualization was to create an intuitive and interactive tool that simplifies the process of exploring and understanding demographic data about colleges. By integrating a high-level overview of student populations with detailed insights into racial diversity, the visualization empowers users to identify trends, compare colleges, and delve into specific details.
   - Audience
     - The primary audience includes students, parents, and educators researching colleges. For these users, understanding student population sizes and racial demographics are crucial factors in evaluating institutions. The interactive features make the data accessible and relevant to these groups.

4) Why did you choose the variables and the encodings?
   - Variables
     - Region: Serves as a filter, allowing users to select a specific region when searching for colleges.
     - College Name: Displayed as categories on the y-axis of the bar chart, enabling easy comparison between colleges.
     - Undergrad Population: Represents the number of undergraduate students at each college, shown along the x-axis of the bar chart for clear visual comparison.

    - %White, %Black, %Asian, %Hispanic, %Other: Displays the racial demographics of each college, presented interactively through the tooltip and pie chart for detailed insights.
- Encodings
    - The bar chart uses length to encode the number of students, making it easy to compare different colleges. Colors are applied to enhance visual clarity and distinguish between the bars.
      The pie chart and tooltip encode racial demographic proportions. The pie chart provides a visual representation for users who prefer graphical insights, while the tooltip offers precise numerical details for those who prefer text-based information. Colors in the pie chart make it easy to differentiate between the various racial groups.

5) Describe some insights—say a few things that you learned from the data through the visualization

- The University of Central Florida stands out as the college with the largest undergraduate student population in the dataset.
- Georgia Tech ranks third among colleges in the Southeast for the highest percentage of Asian students at 17.5%, following Duke University in second place with 21.2%, and Emory University in first place with 22%.
- Across most colleges in the United States, White students constitute the majority of the population. However, colleges in California deviate from this trend, with Hispanic students forming the largest demographic group.
- One particularly interesting observation is the inclusion of Puerto Rico as a separate region in the dataset. This "Outlying Areas" region predominantly features colleges with student populations that are at 100% or nearly 100% Hispanic