**DS4200 Final Submission**

**Website:** [**https://jasminechiou.github.io/ds4200\_final\_project-/index.html**](https://jasminechiou.github.io/ds4200_final_project-/index.html)

**Static Plot 1**

For our bar chart on species richness in Sequoia, Joshua Tree, and Redwood national parks, the marks are the bar lengths that represent the number of unique species there are in each national park. One of the channels for this graph is the length which is the height of each bar in which it shows the number of unique species in each park. Another one is the position in which the national parks are represented on the x-axis, while the y-axis shows the number of species richness. A bar chart was chosen for this visualization because it allows whoever is viewing this graph to quickly compare the species richness for all three national parks. The use of only green in this visualization is to put an importance on the number of unique species instead of the different groupings.

**Static Plot 2**

For the second bar chart on species counts across parks by category, the marks are the grouped bars. A channel for this graph is the length which is the bar heights which show the number of unique species for each family. Another channel is the position, in which the x-axis represents the species categories, while the y-axis shows the number of species. The third channel is the different colors to differentiate the national parks. Green for Redwood, orange for Joshua Tree, blue for Sequoia. A clustered bar chart was chosen for this visualization because it allows the data to be compared across species categories and national parks. The use of three different colors is to make it easier for the viewer to compare the three national parks across each category.

**Static Plot 3**

For the box plot on combined vouchers and references by park and category, the marks are the box-and-whisker shapes. A channel for this graph is the length which is the height of the boxes. The height of these boxes represents the IQR, while the whiskers and points show variability. Another channel is the position which is the x-axis that represents the parks, and the y-axis that has the combined count of vouchers and references. The third channel is different shades of blue which show the variability. A box plot was chosen for this visualization because it shows the distribution of data as it shows median and variability primarily. The shades of blue in the boxplot give the viewer an idea of the variability and the distribution overall. It also helps them understand the spread of the data.

**Interactive Plot 1**

The heatmap uses rectangles as a mark to represent the intersection of abundance and nativeness categories. The X-axis shows abundance levels, while the Y-axis shows nativeness. A gradient color scheme (channel), ranging from light to dark blue, indicates the number of species in each category. Darker shades represent higher counts. Interaction is included through tooltips, which display exact abundance, nativeness and number of species values when you hover over a cell. This allows users to access detailed information without cluttering the visualization. The heatmap is chosen because it is ideal for identifying patterns and anomalies in species distribution. It highlights areas with high or low counts of native and non-native species. The blue color gradient emphasizes variations clearly and is accessible for users, including those with colorblindness. The interactive tooltips make it easy to explore trends while accessing precise data. This design helps ecologists and park managers monitor biodiversity and address conservation needs effectively.

**Interactive Plot 2**

The Sankey Diagram shows how non-native abundant species are connected across different categories, orders, and parks. The nodes (marks) represent the species categories (like Vascular Plant, Amphibian, Mammal), biological orders (like Poales, Apiales), and parks (like Redwood National Park). The flows between the nodes (channels) show which species belong to which orders and which parks they’re found in. The thickness of each flow reflects the relative abundance of these species. This diagram is interactive, so users can hover over or click on nodes to highlight connections and get more information. You can even drag the nodes around to explore specific parts of the data in more depth. This type of visualization works well because it breaks down complex relationships and makes it easy to see how non-native species are distributed across categories, orders, and parks. The interactive features let users focus on what interests them most without cluttering the visual. Dragging nodes adds another layer of flexibility, helping users uncover patterns and connections they might not have noticed otherwise. The flows and their widths clearly show the dominance of certain species, like Vascular Plants, in the non-native and abundant group. The color and flow thickness highlight important trends, like the prevalence of Poales (a group of flowering grasses) in these parks. This is a key insight, as these plants can negatively impact wildlife, water cycles, and even agriculture. By focusing on these details, the diagram makes it clear where park managers should start addressing invasive species. It’s a practical tool for conservation and planning efforts.