Pokemon Visualization Explanation

Edward Liu │ Eric Wu │ Gregory Zeng

This document serves to explain the design of each of the visualizations on our website.

View on GitHub

<https://github.com/ericwuu11/DS4200-Pokemon-Project>

## Violin Plot (*Static*)

The first plot on the Pokemon visualization website are violin plots showing the stat-total distributions of *all* Pokemon for each generation. The mark is the area encapsulated within each violin, which shows the distribution. In this first violin plot, the first channel is the (ordinal) position on the x-axis, which indicates the generation. The next channel is the (continuous) y-position of a given stat-total. Finally, the width of the violin shows the relative frequency of a given stat-total for a given generation. Color was not added as a channel in that it would increase complexity without adding to the visual.

The first violin plot clearly shows a complete visualization of the entire dataset. The second violin plot expands on the first. In this second violin plot, we specifically filtered out “baby” Pokemon and “Not Fully Evolved” Pokemon. These tiers of Pokemon are not competitively viable and we wanted to compare the distributions of tiers both the complete and the competitive subsects of the dataset.

## Bar Plot (*Static*)

The next Bar Plot shows the percentage composition of each tier over time. The years on the bottom represent when new generations were released in Japan and the y-axis represents the total makeup of each generation over that timeframe. The sum of the bars in a given year is always 100%. While a stacked bar chart could be used to show the change over time, we wanted to instead focus on the relative proportion each generation takes up in a given year. Using this methodology, we can, for example, compare the relative performance of Gen4 over time and conclude that Gen4 was disproportionately popular in the OU tier from 2006-2016.

The marks in this graph are the bars. The channel the position on the x-axis, showing the year. Another channel is the length of the bars indicating the relative popularity of a Generation in a given year. Finally, the last channel is the color of the bars, indicating the Generation.

## Sankey Diagram (*Static*)

The Sankey diagram shows the flow of First-Generation Pokemon over time. As with the previous visualizations, we filtered out “baby” Pokemon and “Not Fully Evolved” Pokemon in that generally, there is very little variation over time. Additionally, these categories are generally non-competitive. As such, this could be considered “noise” in our overall analysis of the competitive landscape.

The marks in the Sankey are the lines used to represent the nodes in the diagram. We also use connection link marks to show the flow of these nodes over time. Our Sankey specifically uses the x-position channel to represent flow of time from left to right. It also uses the y-position channel to represent different discrete tiers. The tiers are ordered such that the “best” tiers are always at the top and the “worst” tiers are at the bottom. This provides a consistent view of the tier distribution across the x-axis.

## D3 Diagram (*Interactive*)

The D3 diagram displays how the unique average stats of Pokemon change over the generations. Each of the individual bars are color coded to match the key for simple and effective reference. The ability to hover over a bar and click on it to find the average without guessing based on the y-axis, or the total stats, makes statistical analysis easier. This visualization is key in leading to better understanding of statistical balance, where we can start asking questions regarding if the developers favor power creep or game balance. In combination with the other visualizations that were generated, this visualization is an important foundational piece in developing further conclusions.

The marks used within the D3 diagram are areas, as barplot tends to use rectangles of different areas to represent variations within the data. For example, you will notice in this visualization that generation 7 has a taller attack bar than the other generations, which indicates that numbers-wise, that generation outperforms the others. With regards to channels, position, color, and size (specifically length) are the key elements. Position is typically used to represent the quantitative data along one axis, while color, as previously mentioned, is used to help users identify what each of the bars signify based on the key provided. Lastly, the size can be explained in a similar fashion as for the area, where the overall length of each bar helps to determine variance in the visualization.

## Scatter Plot (*Interactive*)

The interactive scatter plot focuses on displaying the statistical factors influencing Pokemon prioritization across generations. In terms of some of the marks used to develop our visualization, we mainly used circles (‘mark\_circle’) to represent each individual Pokemon. For channels, the x-axis represents the generation of Pokemon (‘generation’) using an ordinal scale (‘O’). The y-axis represents the total points of each Pokemon (‘total\_points’) using a quantitative scale (‘Q’). There is a myriad of colors that represent each Pokemon tier (‘tier’) using a categorical scale (‘N’). The color of each circle corresponds to the tier of the Pokemon. The color scheme is custom-designed to differentiate between tiers. Finally, there is the tooltip, which displays additional information about each Pokemon when hovering over a circle, including the name, generation, tier, types, abilities, and base stats.

The reason our group decided to design this interactive visualization in this manner consists of a multitude of factors. The first being interactivity, allowing users to interact with the data. They can select specific Pokemon by name using a dropdown menu and highlight multiple Pokemon tiers by clicking on the legend. We also value comparison, enabling users to compare the total points of Pokemon across generations and see how different tiers are distributed. Then there is detail where the tooltip provides detailed information about each Pokemon, allowing users to get more context about the data points. Lastly, there is customization, the custom color scheme for tiers makes it easier to distinguish between different tiers visually.