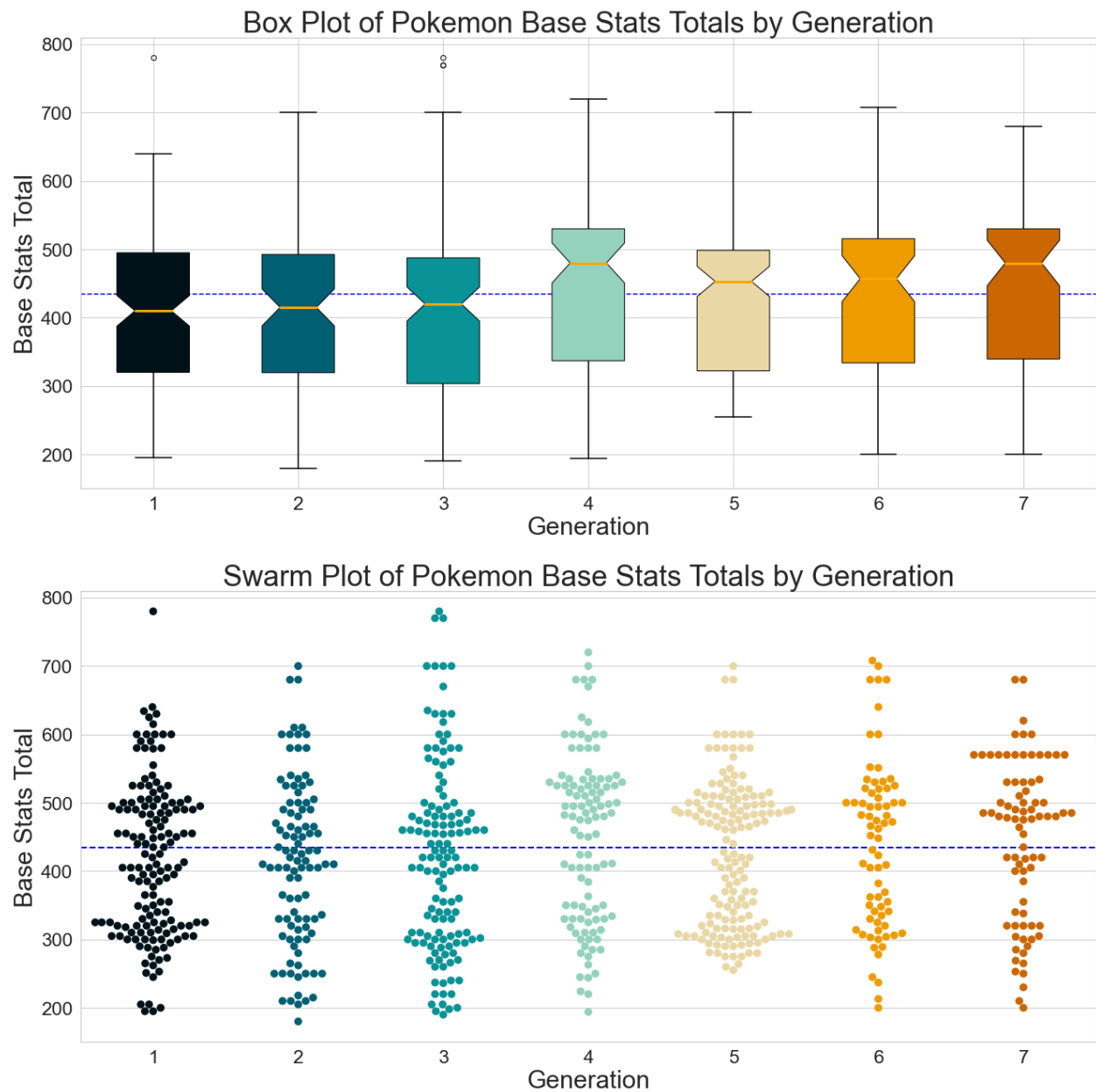
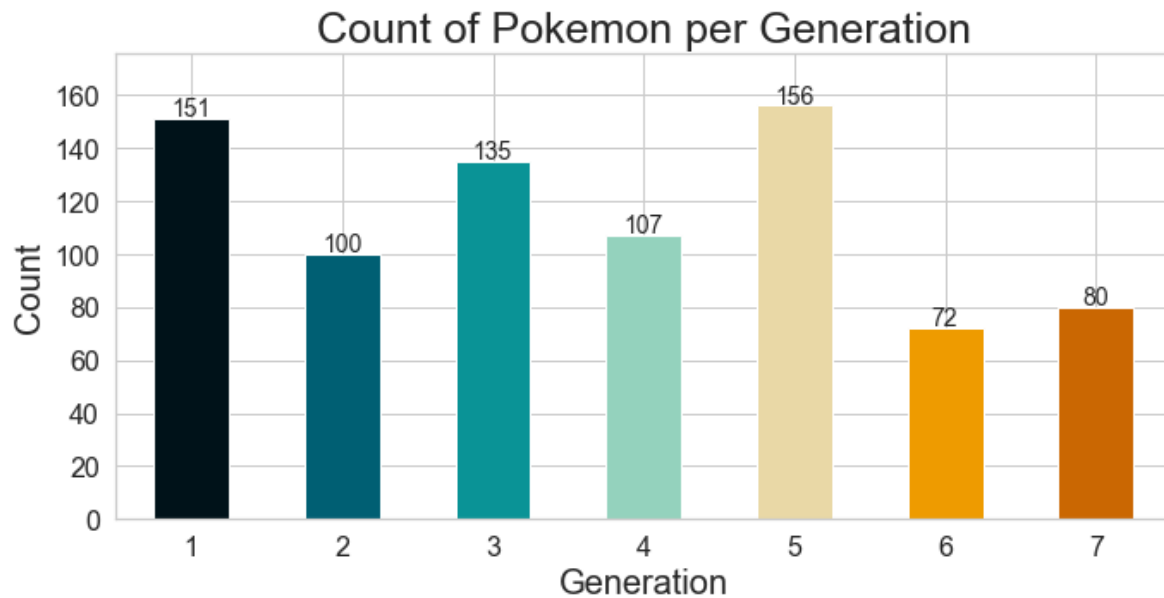


Jerry Huang
Information Visualization
Dr. Lingfei Wu
Spring 2023

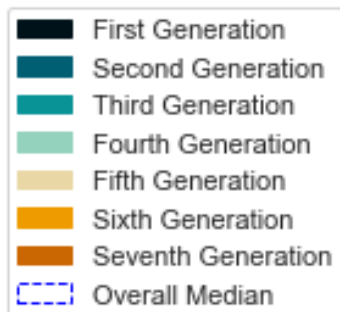
Pokemon Stats Comparison Across Generations in the Pokemon Universe Report

main figure





Legend



- Each generation of pokemon is presented by a different color while the overall base stat total value is presented by a dotted blue line.

Findings

- The base stats totals are skewed slightly more on the left side, the earlier generations of Pokemon, as the right side, the later generations main box body were higher up.
- The box is relatively short which shows that the base stats had a small interquartile range between the Q3 and Q1 of the box plot.
- The top whisker is longer than the bottom whisker which shows that there is a larger spread in the upper range of base stats in pokemon than in the lower range.

- Only Generation 1 and 3 had extreme outliers that were not able to be reflected in the box plot with the standard box and whiskers and had to involve the use of small dots to represent extreme outliers.
- The generations that have a higher average of higher base stats tend to have more variance in base stats (lowest vs highest)
- The generation that had the highest average base stats total is the 4th generation of pokemon
- The more recent generations generally have fewer amount of pokemon released in their given generation

Data and Method

The dataset was from the public Kaggle dataset by Alberto Barradas seven years ago. The dataset contained information from the first 7 generations of Pokemon of the 9 current generations. At first, there was a consideration of giving weight to each particular stat of a Pokemon so that stats like speed were worth more for fighting since Pokemon is a turn-based game the person who attacks first depends on which Pokemon's speed is higher. However, after considering that some people do not particularly care for player versus player I decently leave each base stat untouched and use the total base stat value that was already generated by the maker of the dataset. The main graph features a box plot for the 7 generations that were covered in this project with an overall median base stat line going across the graphs to give some clear comparative metrics to connect the box plots more. Median was used since there are a handful of outliers in the lower and upper sections which could have skewed the line if we were to use mean. Also, I created a simple bar chart for the amount of Pokemon that were released in each generation as there was never a set amount of new Pokemon being made each generation that comes. Additionally, plots all followed a general 2.2 to 1 ratio that was suggested by Edward Tufte along with the use of Sans-serif font is generally easier and clearer to read on screens

Significance Statement

The presented figure based on base stat totals across 7 generations of Pokemon is significant because it provides a graphical insight into the Pokemon universe and how it changes from every generation. With information like this box and swamp plots, game developers can easily identify a level of power or weakness they should be creating in future generations so that there is no one generation so out of tune compared to the rest. Another demographic of people that benefits from this is the competitive Pokemon players as they will potentially be able to choose which generation of Pokemon is the most optimal generation to craft a team out of for a higher chance of winning at events when Pokemon statistics play an important role in performance in a competition.

