Write Up

Jacob Shigeta

2024-02-02

Write Up

Introduction

I chose this data set because I have been a big fan of Pokemon for as long as I can remember as well as wanting to be able to visualize it and maybe take it up to a larger scale with a larger data set or more Pokemon to deal with too. I also chose it because I have not really done anything related to gaming in any R class so this gave me a nice test run at it.

As I was doing this project, I hoped to learn more about how to create better looking graphs as well as make them show what I wanted them to show. This is also partially why I decided to go with Pokemon for this not project because it's a topic I am comfortable with and know quite a lot about. I also wanted to make sure that I could get comfortable writing in an R file because most, if not all, the work I have done in these classes have been in RMarkdown.

I found this data more interesting than important mainly because it's just a hobby of mine so in the grand scheme of topics I could have picked, it is not that important. However, I always find Pokemon interesting, even if I already knew about it before so just being able to visualize this made it more interesting because I could see what I was thinking in ways I have not been able to find before.

Data Cleaning

When it came to cleaning the data, I did not find much challenges. The only real cleaning I did was changing the labels of some columns, changing the types of some columns, and filtering out the columns I did not need for what I wanted to visualize.

The only issue I encountered was when it came to changing the column names because I do not do that often so I did not know how to go about doing it. I went around this by just looking it up online and I found the solution rather quick.

Some of the key variables in the data set, at least out of the ones I used, are: - Primary_Type: which is the primary type of the Pokemon - Generation: which is the generation the Pokemon was introduced - HP: which is the HP stat of the Pokemon (same can be said for the rest of the variables I used and did not mention).

Data Visualizations
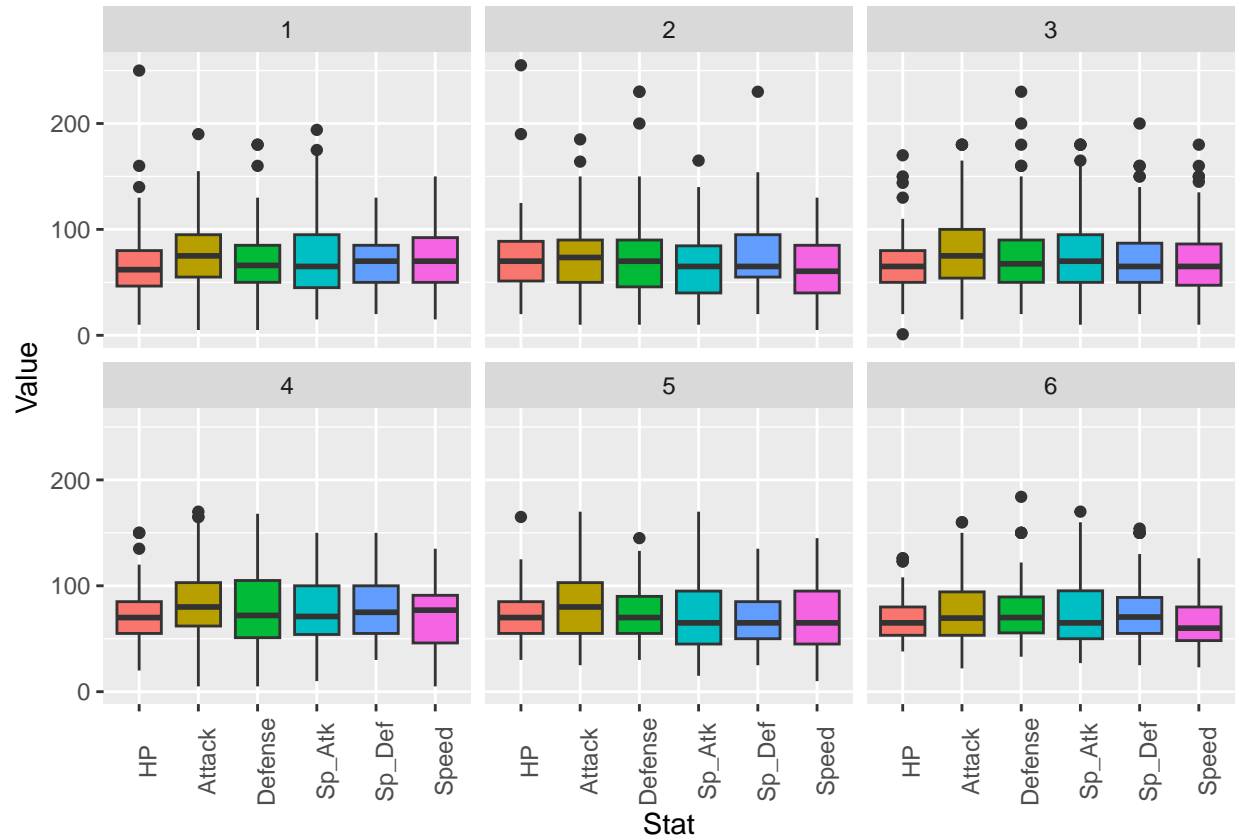
The graph above shows every main stat for every Pokemon in the first six generations grouped by Generations. This data shows that as the generations go on, there tend to be less outliers when it comes to a single stat. Because of this, we can determine different things, such as generation 4 having some of the faster mons in the data set and generation 6 having some of the slowest. It can also let us see which attacking stat to prioritize while playing through those games, such as generation 6. In those games, we can tell that the Defense is way lower than the Special Defense, meaning that we should be building our teams around Attackers rather than Special Attackers so we can have an easier time clearing the game.
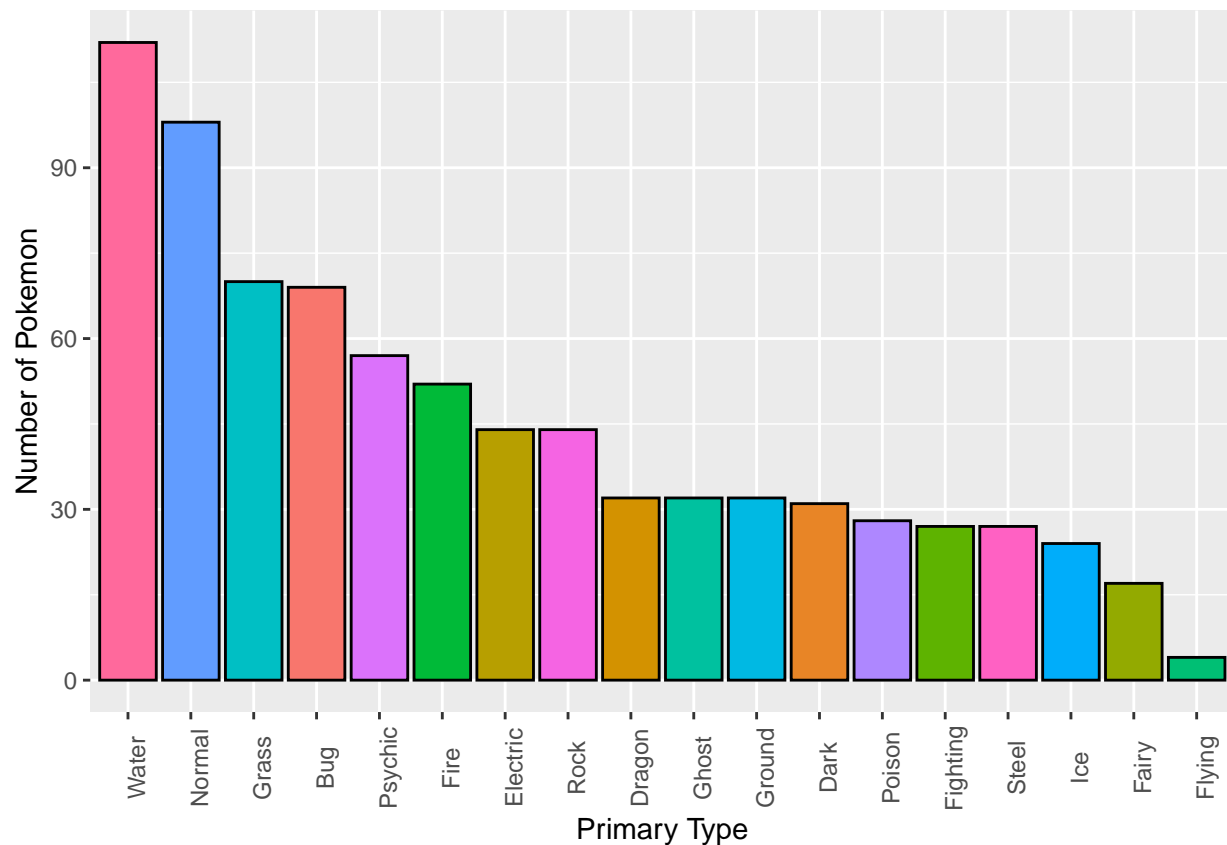
This graph shows the Primary type of all the Pokemon from the first 6 generations of the game's lifespan. Looking at the data, we can assume that the Water and Normal types will remain the most common types for a long while. We can also assume that the Flying and Fairy types will be getting more additions to catch them up with the rest of the types. However, these are just primary types and because a lot of the Flying and Fairy types have them as Secondary types, this graph could change if I were to implement them.

Conclusion

If I had more time, I would want to figure out how to make a graph that contains the Pokemon's total stats in groups by 100 (0-100, 101-200, etc.) and graph how many Pokemon are in those groups. It would be nice if there was also some way to split the legendaries and non-legendaries in that scenario to maybe make it a multi-bar chart or maybe even a stacked bar chart since I feel like that will be easier to read. I would also want to figure out if there is a way to add the Pokemon sprites to the boxplots but only have them on the outliers. That way we can see which Pokemon are outliers in their generation for their stats.

When it comes to analyzing, I knew I would have an issue explaining what the graph shows. This is because I find explaining things hard, even when I am passionate about. Another challenge was figuring out what to talk about when analyzing it. This is because, while my second graph is straight forward, there is a lot I can talk about in my first graph. With that, it means that I could tend to repeat what I am talking about (as I tend to do). Despite having a lot of things to talk about in my top graph, I found it hard to pick certain things to talk about. This is because a lot of the ranges for the stats were surprisingly similar and therefore, harder to talk about what stands out.

Github Repository

Mmaxs808/NOT_Project_1