Pokemon Data Visualization and Analysis with R

Jihong Zhang

2025-04-10

Motivation

Pokemon is a global icon. It is a TV series that expanded into video games, Pokemon cards/collectibles, and merchandise that made it into a multimillion dollar franchise. Growing up, Pokemon was one of my favorite shows and to this day, it remains one of the fondest memories from my childhood.

Research Question

Before I begin my project, I should know what I am doing and how to do it. This quarter, I am in my first *Introduction to R* course so my project directly applies the course material. I utilized my lecture notes to learn the technicalities of coding it and the mathematics behind each line I wrote.

After doing my research, I took this approach to solve my own problem: **Could I predict battle outcomes?**

My report is organized into the following section:

1. Loading the Data Set
2. Data Visualization
3. Linear Regression.

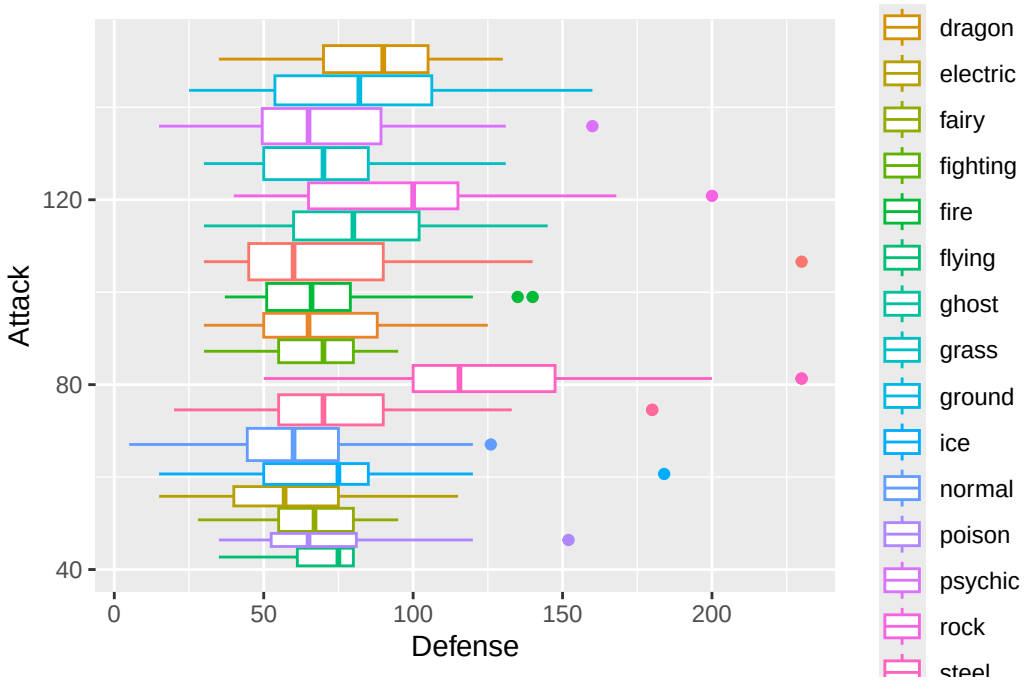

Figure 1: Boxplot of denfense and attack by their types

Data Visualization

Data visualization is a powerful tool. Figure 1 shows the characteristics by their Types.

Second, I check out the relationships among six dimensions of information of Pokemon using pairwise scatterplot. As shown in Figure 2, I did not see clear relationships.

Linar Regression Results

Table 1: Linear regression of six indices on HP

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	29.98	2.69	11.13	0.00
Speed	-0.11	0.03	-3.88	0.00
Attack	0.31	0.03	10.84	0.00
Defense	-0.07	0.03	-2.30	0.02
SpAttack	0.14	0.03	4.98	0.00
SpDefense	0.24	0.03	7.01	0.00

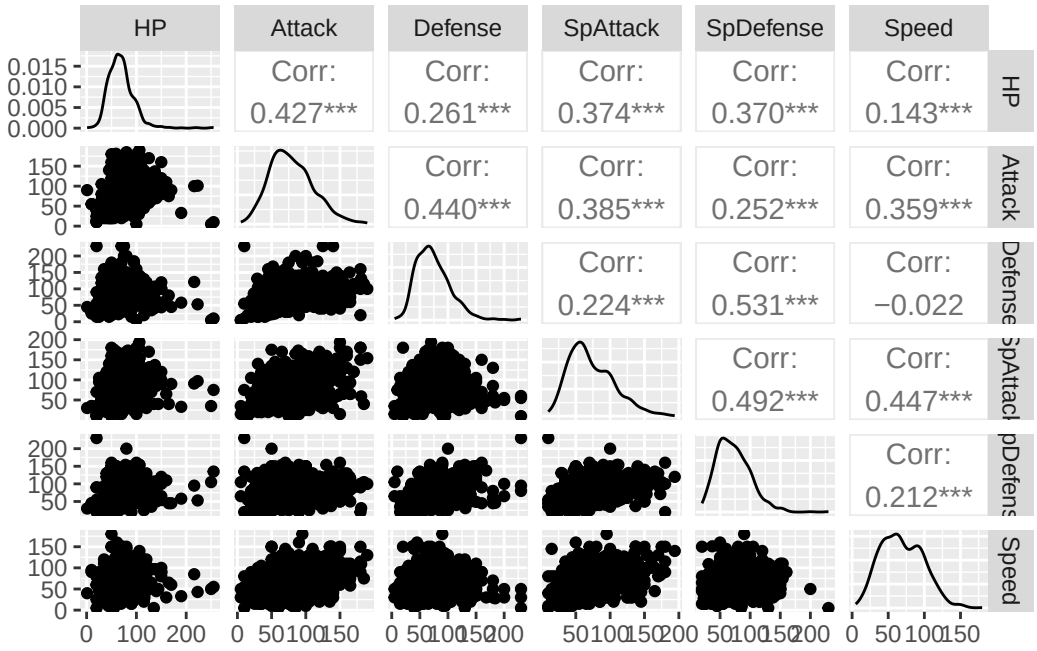

Figure 2: Pairwise scatterplot

As Table 1 shows, *Speed* and *Defense* have negative effects on *HP*, while *Attack* has positive effect on Pokemon's HP. *Defense* has no significant relationship with *HP*.

Conclusion

With this project, I achieved a lot of firsts. It was the first time that I worked with such a large dataset. Not only were there 800 observations, but there were also 12 different variables. This meant I had a lot to play around with and explore, but also there was room for a lot of error.

It was also the first time I have worked with multiple linear regression. I am accustomed to performing linear regression analysis on the relationship between two variables, so working with eight was a new challenge.