Draft

2023-07-28

Aditi Sunil

Introduction

In this project, we delve into the world of Pokémon to explore how the health of these creatures affect their prowess in battle. Pokémon is a role-playing game based around building a small team of monsters to battle other monsters in a quest to become the best. Each Pokémon have different strengths and weaknesses which affect their outcomes in battle. This data set includes a collection of 802 Pokémon from all generations, providing a rich and diverse population to investigate. We aim to answer the questions, does the Pokemon's health affect its average attack and average defense, and how are these affected by a Pokemon being legendary?

Through this analysis, we assert that a Pokémon's given characteristics, hit points and if they are legendary, significantly influence its combat effectiveness. By constructing a linear regression model and bar plot, we aim to provide compelling evidence and insights that support the correlation between specific traits and a Pokémon's ability to attack, defend, and it being legendary.

Background

Our data set chosen is the collection of all 802 Pokemon from all seven generations. This data set is a population, as it contains every official Pokemon. The key variables in this data set are: `name`(the name of the Pokemon), `height_m`(height of Pokemon in meters), `hp`(hit points a Pokemon has), `attack`(attack points a Pokemon has), `defense` (defense points a Pokemon has), `weight_kg`(weight in kilograms), `speed` (speed of Pokemon), `classification`(classification of Pokemon), `is_legendary` (tells if Pokemon is legendary), `generation` (generation of Pokemon [1-7]), `type1` (primary Pokemon type), and `type2` (some Pokemon are combinations of two types, `type2` is defined for such Pokemon). As new generations of Pokémon are introduced, the overall landscape of combat dynamics may shift. The introduction of new species, abilities, and mechanics could influence the patterns observed within our dataset. It is important to note that our analysis is based on a snapshot of Pokémon up to the point of dataset creation. Any subsequent changes or evolutions within the Pokémon franchise may impact the accuracy of our conclusions. The dataset used covers seven generations of Pokémon, which is not up to date with the current nine generations. While our dataset encompasses a vast array of Pokémon, it is essential to consider potential sampling biases. Certain generations or Pokémon types may be overrepresented or underrepresented, potentially affecting the generalizability of our findings.

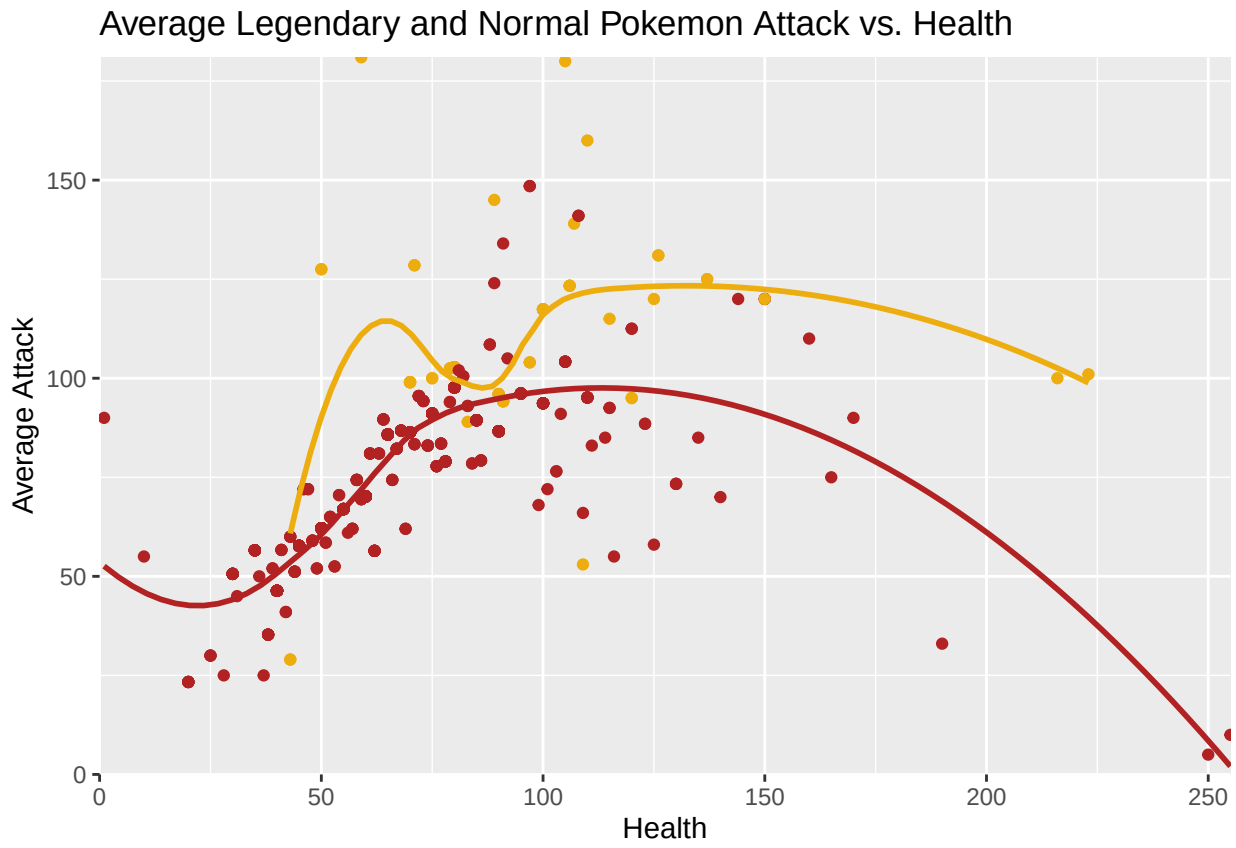
This data was scraped from http://serebii.net/ [1] which contains a database of all Pokemon. It was turned into a csv file by Rounak Banik and was posted onto Kaggle.

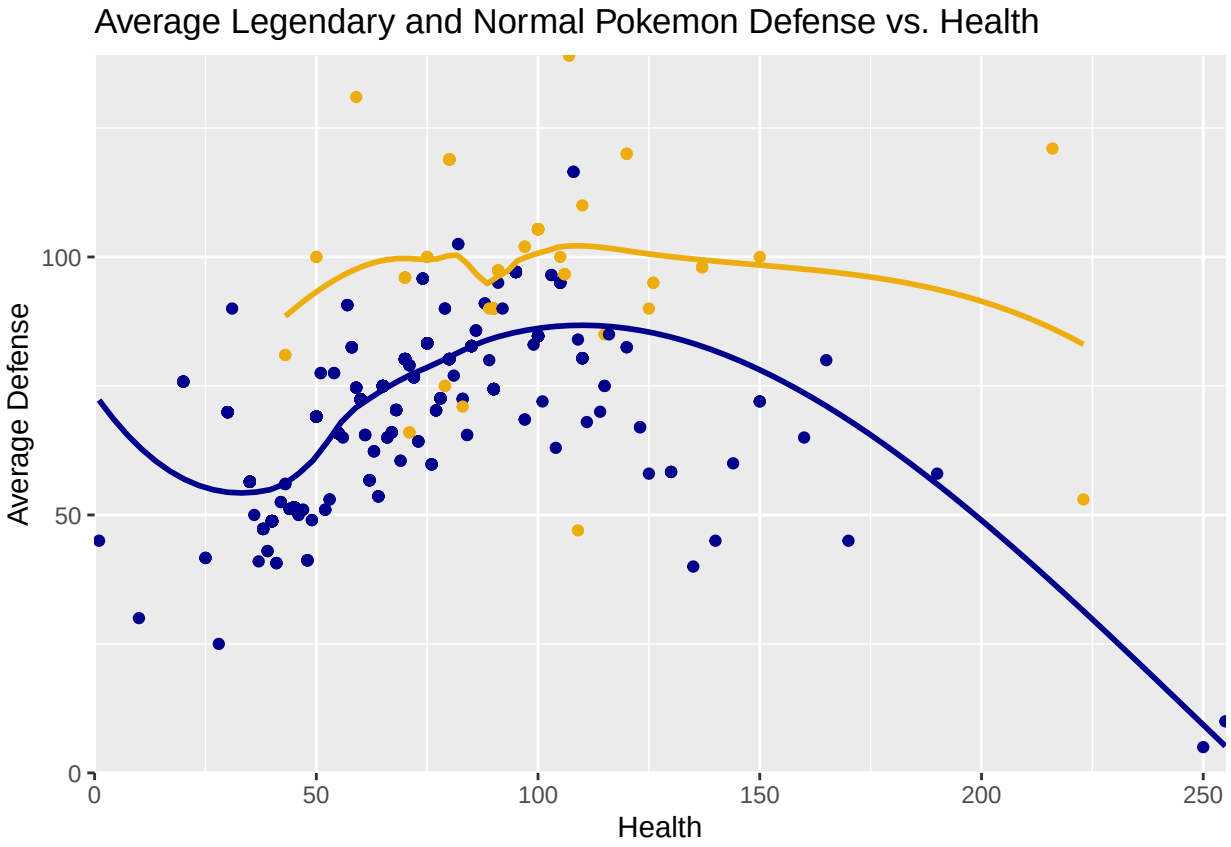
The rest of the report will analyse our Pokemon data in order to answer our main questions and display them through graphs.

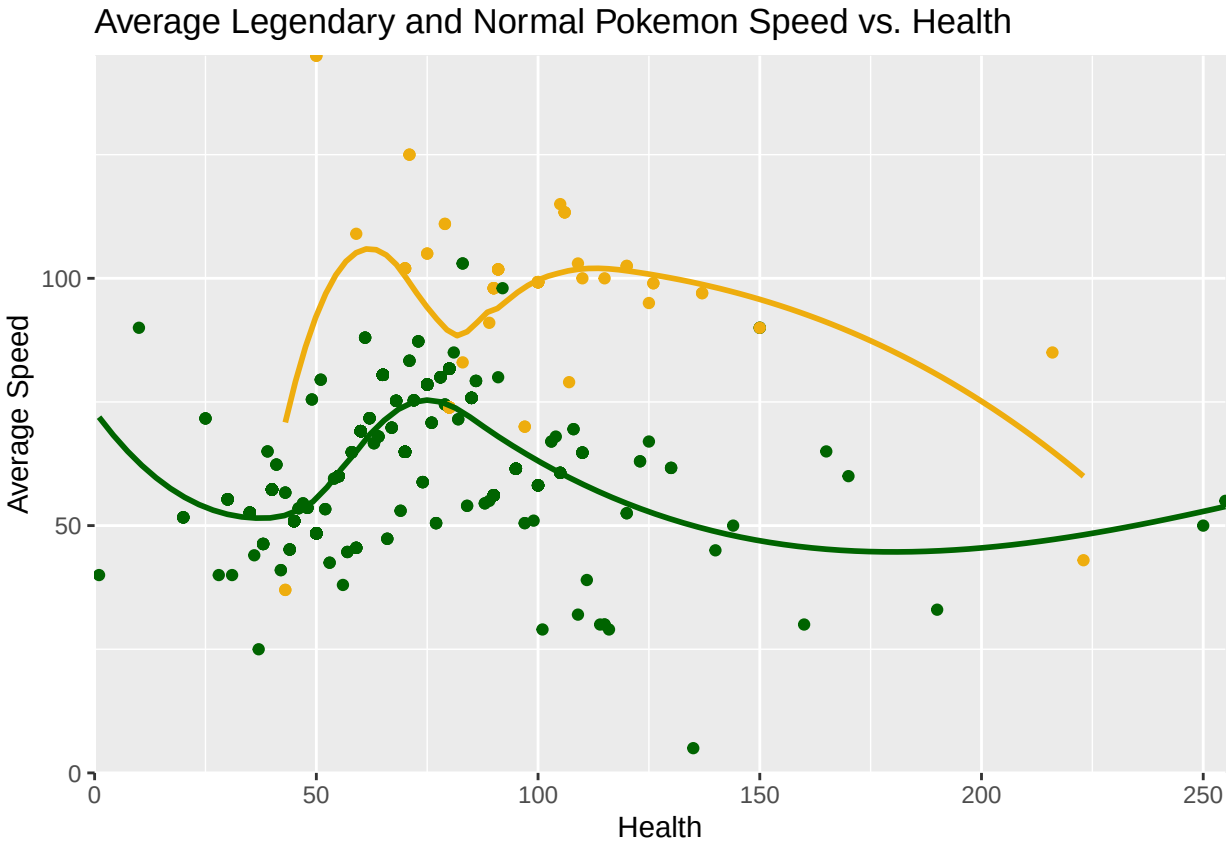
[1] http://serebii.net/

Analysis

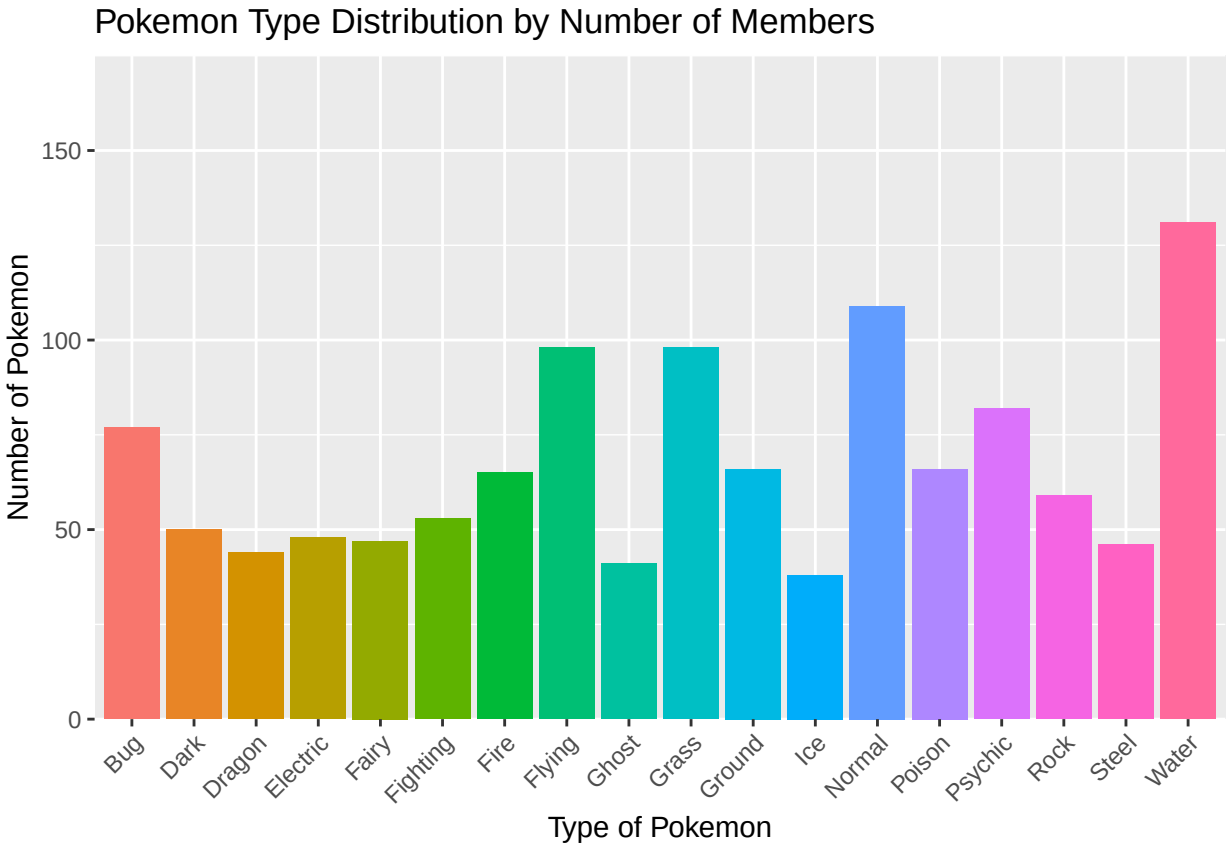
Group each Pokemon by their HP and whether or not they are legendary, and for multiple Pokemon with the same HP, compute the average of their attack, defense, and speed. Create scatter plot to visualize the relationship between HP and each of the statistics to compare normal Pokemon averages to those of legendary Pokemon. For each grouping, plot a smooth trendline to illustrate the relationship between legendary and normal Pokemon averages of the four primary factors (HP, attack, defense, and speed) determining the effectiveness of the Pokemon.



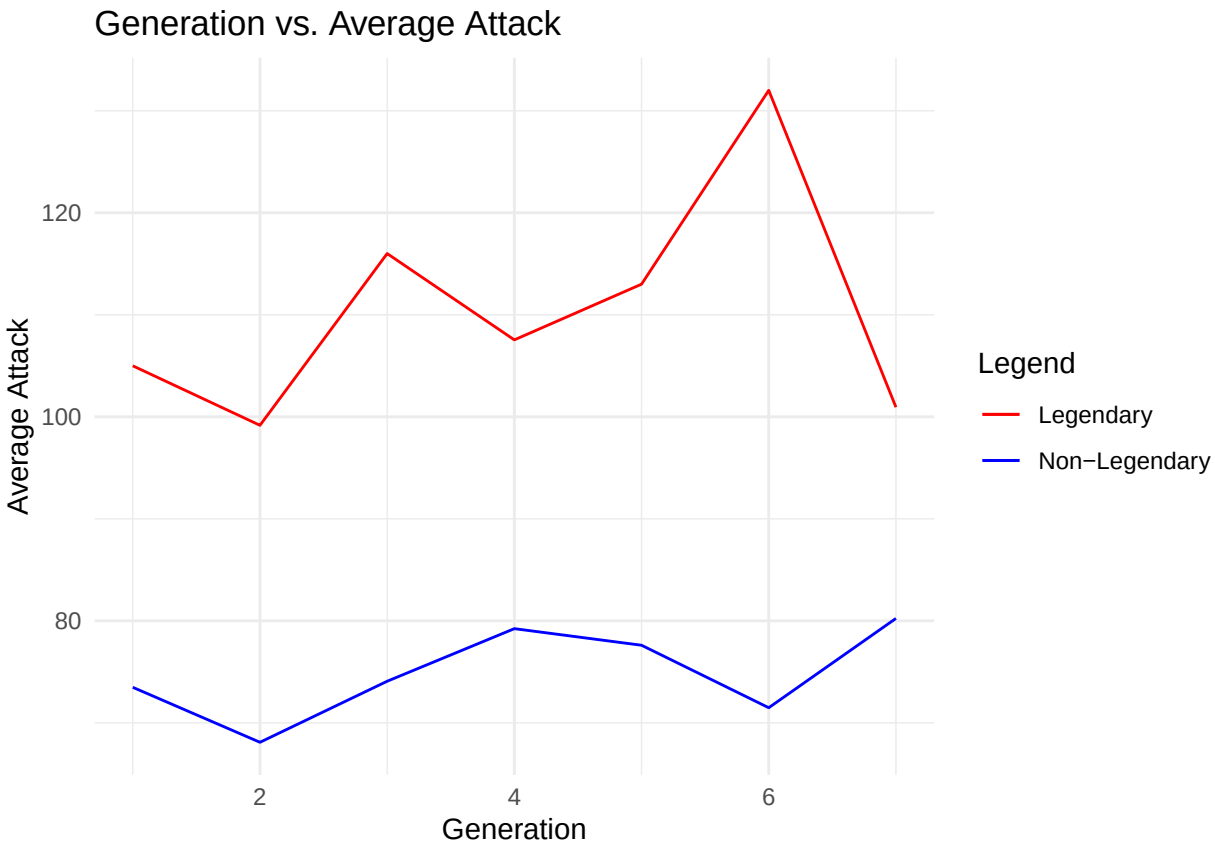

Average Legendary and Normal Pokemon Defense vs. Health
Average Defense
Health
0
50
100
150
200
250
0
50
100

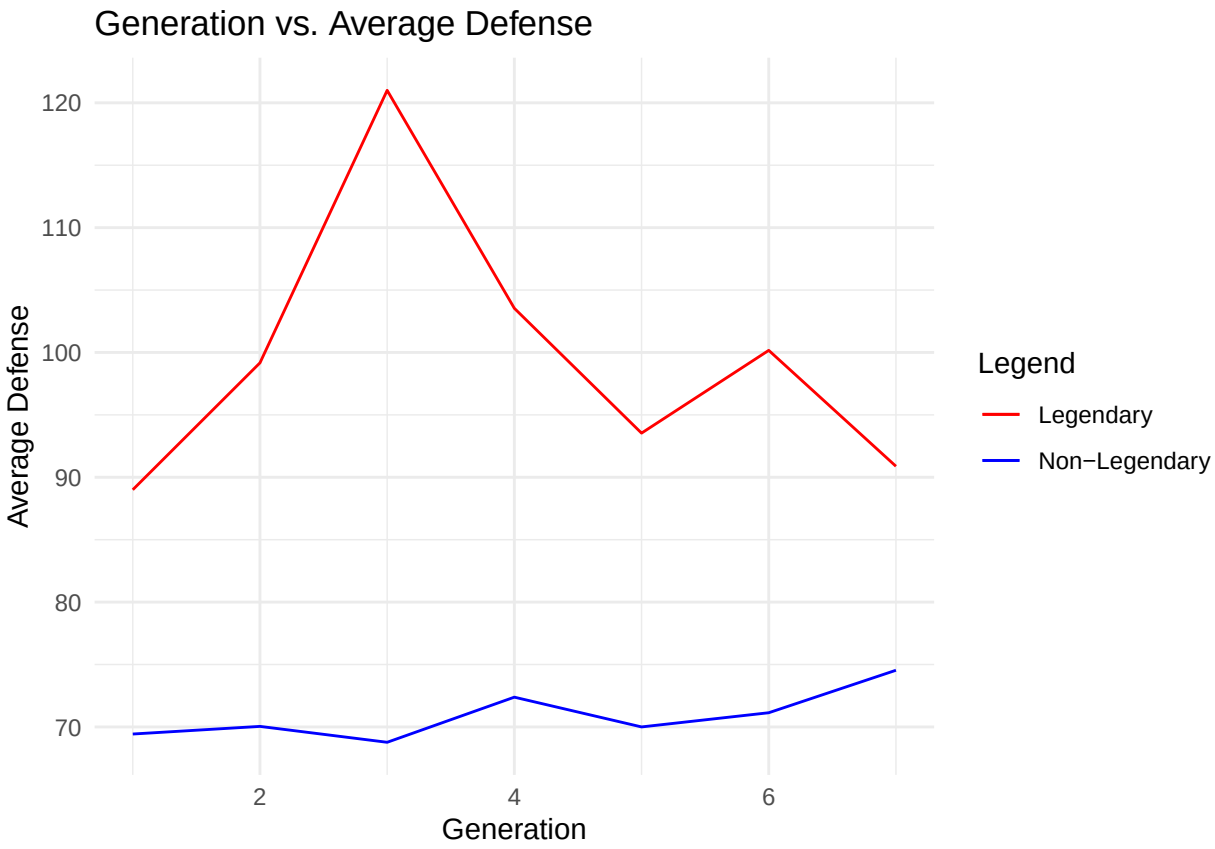


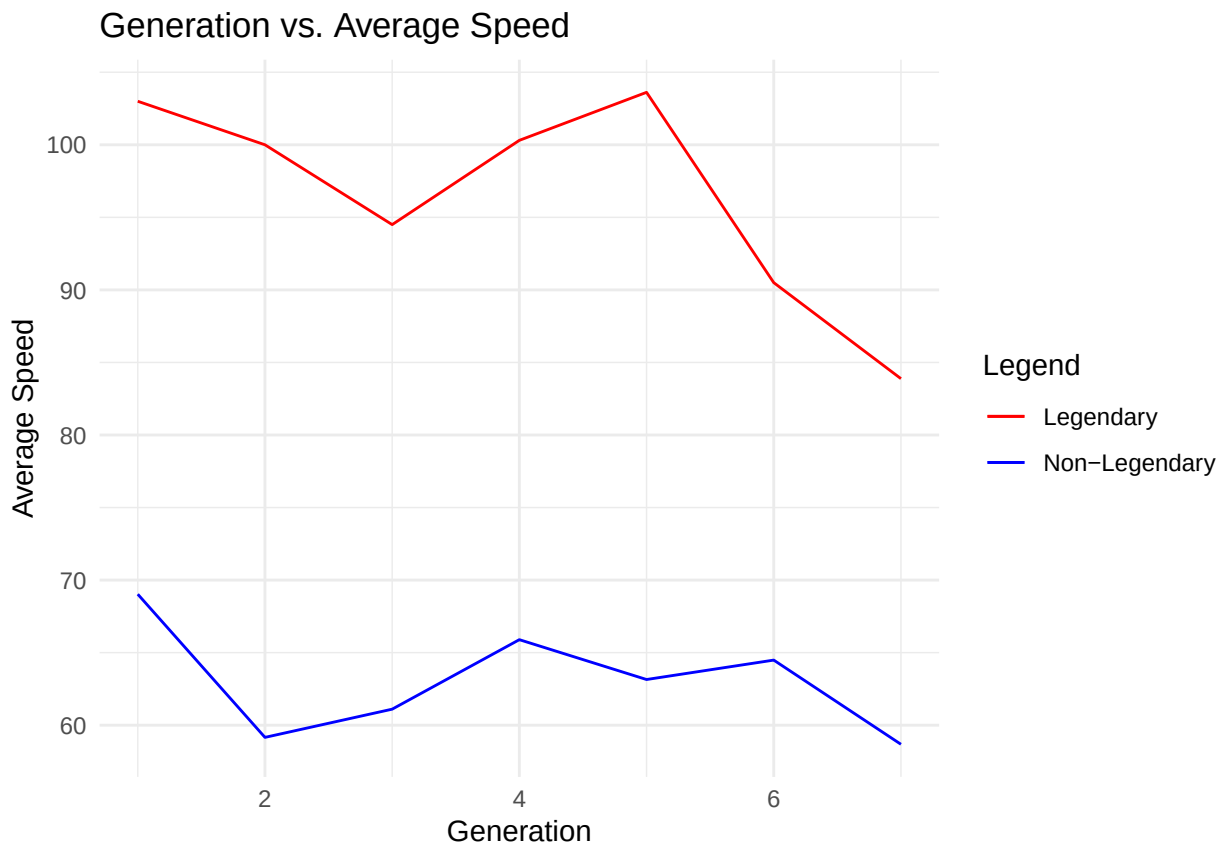
We can survey the distribution of Pokemon per type and calculating the size of each individual class:



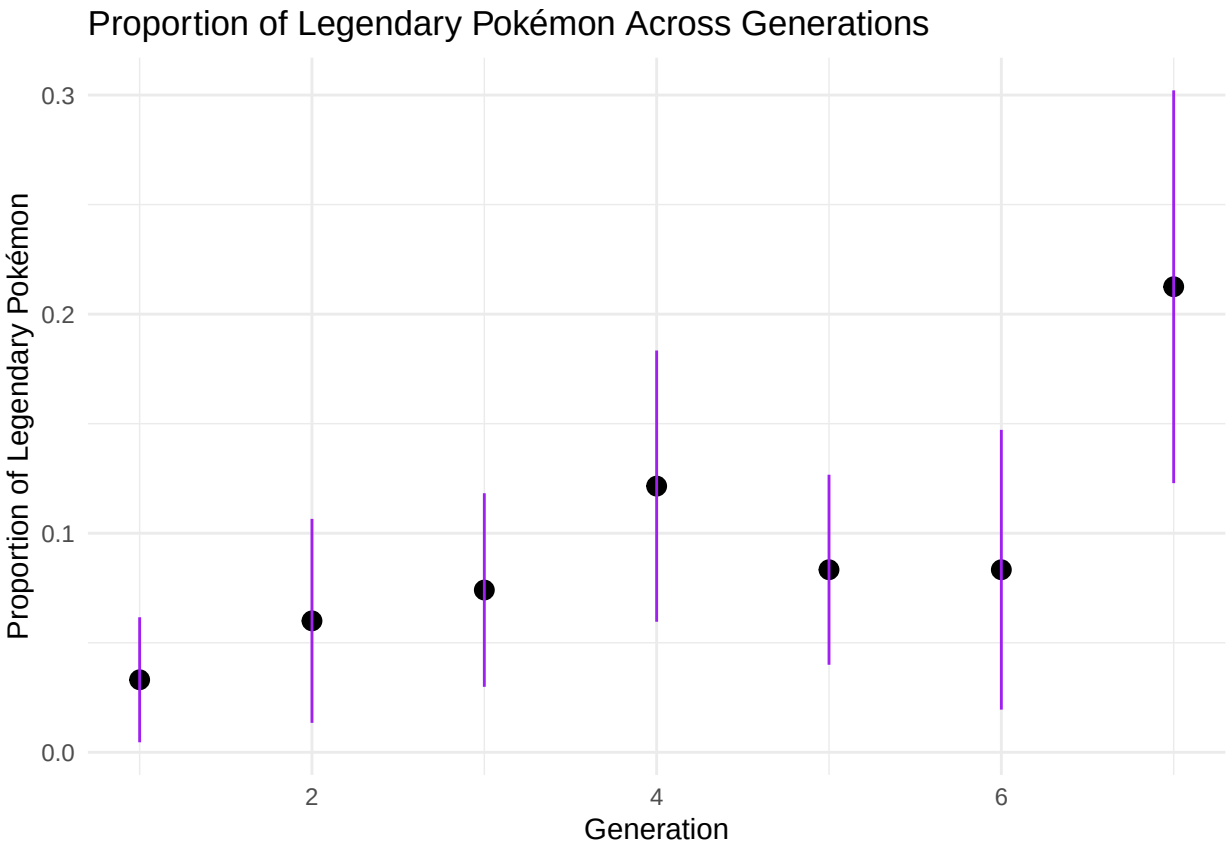
Additionally, we can create scatter plots for generation and average attack, defense, and speed to determine if more recent generations have stronger characteristics.


Generation vs. Average Attack
Average Attack
120
100
80
2
4
6
Generation
Legend
Legendary
Non−Legendary


Generation vs. Average Defense
Average Defense
120
110
100
90
80
70
2
4
6
Generation
Legend
Legendary
Non−Legendary



Below is a numerical analysis graphic to visualize confidence intervals of 95% for the proportion of pokemon which are legendary across each generation.



Discussion

From our plots, we can notice that average attack and average defense against health follow the same trend, they start in a decreasing pattern until health is around 25 and increase until around 125 before decreasing again. This shows that those with the largest average attack and defense have a health of between 25 and 125. Average speed also somewhat follows this trend, where its highest average speed is between around 50 to 75, before decreasing again. The trend for legendary Pokemon is different for average speed, attack, and defense as for each category legendary Pokemon's average is above the non-legendary average. We then compared average attack, defense, and speed to each generation of Pokemon. Average attack and defense had a minor increase, while speed had a minor decrease. As we saw before, the legendary Pokemon was much more than the non-legendary Pokemon. All in all, those who had health between 25 and 125 seemed to have the most average speed, attack, and defense.

With the information we have collected, we have noticed that the legendary Pokemon are more powerful than non-legendary Pokemon. It would be interesting to dive further into the characteristics of the legendary Pokemon and compare them to non-legendary Pokemon. We could try. I think our choice of using a scatter plot to demonstrate the relationship between average attack, defense, and speed was good, however we could have converted it into a histogram to give an idea of the spread of data. Additionally we could use a density plot instead of a violin plot for Pokemon Height and Weight Distribution, especially if our audience is unfamiliar with more complex plots. We have information about the Pokemon itself, but we could gather more information about its popularity to further look into if its average attack, defense, or speed is influenced by its popularity, or vice versa.

It's noteworthy to mention that the dataset covers seven generations of Pokémon, up to the point of its creation. However, as of the current date, there have been more generations introduced. This discrepancy could potentially impact our analysis if new generations exhibit different characteristics.

References