Basic Statistics (STAT-202-001)

PROJECT

Analysis of Pokémon Generations

From: Samriddh Gupta

Introduction

The first episode of anime was released 1st of April 1997 and since then Pokémon have been all time favorites for everyone. Since then there has been a total of 1020 main series episodes, 8 TV specials (3 full-length, 5 normal-length), 25 side-story episodes,7 Winter Vacation shorts,8 ANA flights shorts, 20 movies and numerous games have been released. It is the only anime know by many people and is famous all around the world. There are more than 800 types of Pokémon made till now and have a lot of different stats and functions. For the project, we will be discussing the first 721 Pokémon we have and try to analysis the outliers in the Attack power of Pokémon VS the Defense power of the Pokémon. Furthermore, we will also analyze similarities in the generations of Pokémon (There is a total of 8 generations of Pokémon and each generation have its own new starter Pokémon along with different type of stats in every one of them.) For this project, we will be only considering till generation 6 as we only have data till that generation. Further, Considering the whole data, we will discuss the total stat of the Pokémon and try to see the nature of the graph and how it is changing in different places along with why it shows that type of nature.

DATA SOURCE

The data was taken from the taken from the Kaggle.com and is a genuine and can be trusted to give us the correct information. I do not face much problem while I was searching for data. The only problem which I face was the amount of data. As I said earlier there is a total of 8 generation of Pokémon but I could only find the data till 6 generation which is a total of 721 Pokémon which is not the complete data I was looking for. But the good part was that the data I found till 6 generation is complete and accurate and is arranged in the increasing order if the Pokedex number. Moreover, there is the repetition of the Pokedex number in the data. Means there are some Pokémon which have a similar Pokedex number. This is because of 2 reasons. First one is that in generation 6 they introduce the concept of mega evolution in Pokémon. In this case, a Pokémon can evolve and then is able to get back to the original look again. Since it happens for only short time and there is no permanent effect on the Pokémon they do not add any specific Pokedex number to them and make them share the same Pokedex number as the original one. In this condition, there is a change in the total stats of the Pokémon. The other reason that 2 or more Pokémon have the same Pokedex number is because these Pokémon shows a different type of changes according to their surroundings. But in this condition, their total stats remain same but individual power changes because of which they can't combine all of them. Unlike any other data, this data is neither a survey nor an experiment. This data is a retrospective data which shows all the data for each Pokémon along with their stats. Moreover, the data is arranged according to the number of the Pokedex which is issued by the official website of the Pokémon Makers. The data is not randomized or arranged in any manner. If the day is randomized, it will not be beneficial for my study as I need every data and entry of that Pokémon. Hence making it randomize does not make any sense for my project. Further If data is arranged in a manner (let's say according to the increasing order of attack power or defense power), but then the data will be

aligned in an increasing order and then there will be no outliers as we can see in figure 1 hence the whole purpose of the project will be nothing.

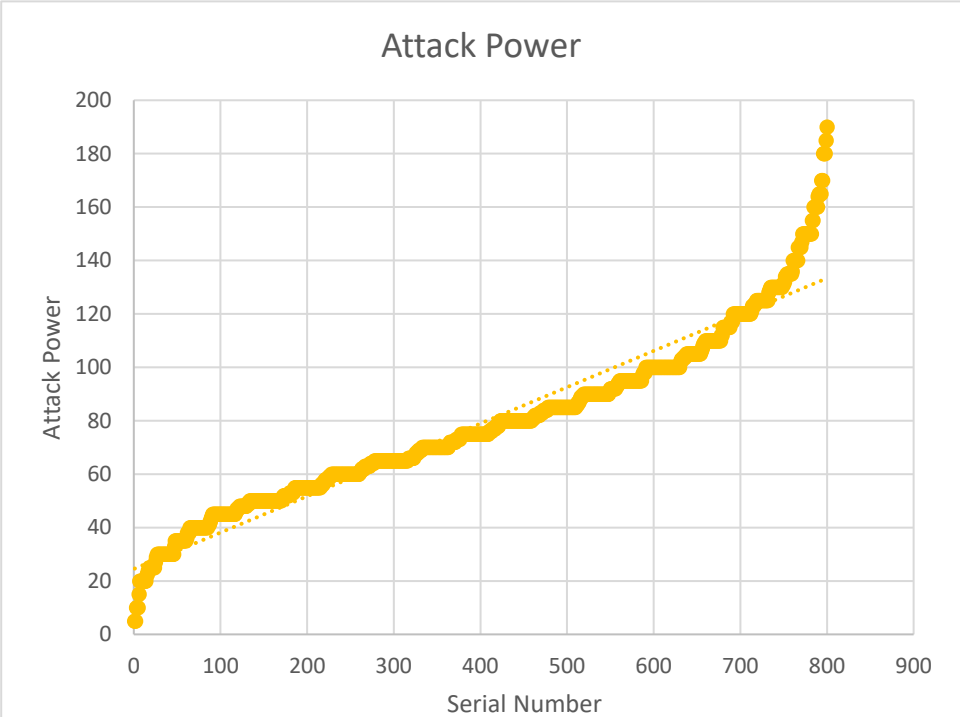


Figure 1. Scatter plot of the attack power of each pokemon arraged in the acceding order.
The darker line are the Scatter Plot and the ligher one is the linear…

DATA SUMMARY AND DATA ANALYSIS

The Pokemon data which I have contains a total of 10 variables and each is different from one another. The first 2 variables are the Type 1 and Type 2. These rows stores the data of the what type of that pokemon is. Each Pokemon are divided into various type and shows that type of power. We can see the there are a lot of empty spaces in type 2 because there are pokemon which only belong to one type while there are some which belong to 2 types. Next is the total power which is the sum of all the power pokemon have. It consists of attack, defense, speed, HP(Health of that pokemon), Sp. Atk(special attack) and Sp. Def(special defense). Next comes the attack power which is the amount of damage deal by that pokemon. Next is the defense which is the amount of defense it can sustain while under opponents attack. HP is basically the total health a pokemon have. It keeps on decreasing under a pokemon attack. When HP hit zero, It can no longer battle. Next is the speed which is how fast can pokemon attack or defense itself. If the speed is faster, the amount of damage can be minimized. Finally the Sp. Atk(special attack) and Sp. Def(special defense) is the damage and defense deal by a pokemon when it uses its special moves. The Scatterplot bellow shows us the comparison of the attack power vs defense power of the whole data and we can see that there is the total of 3 outliers.

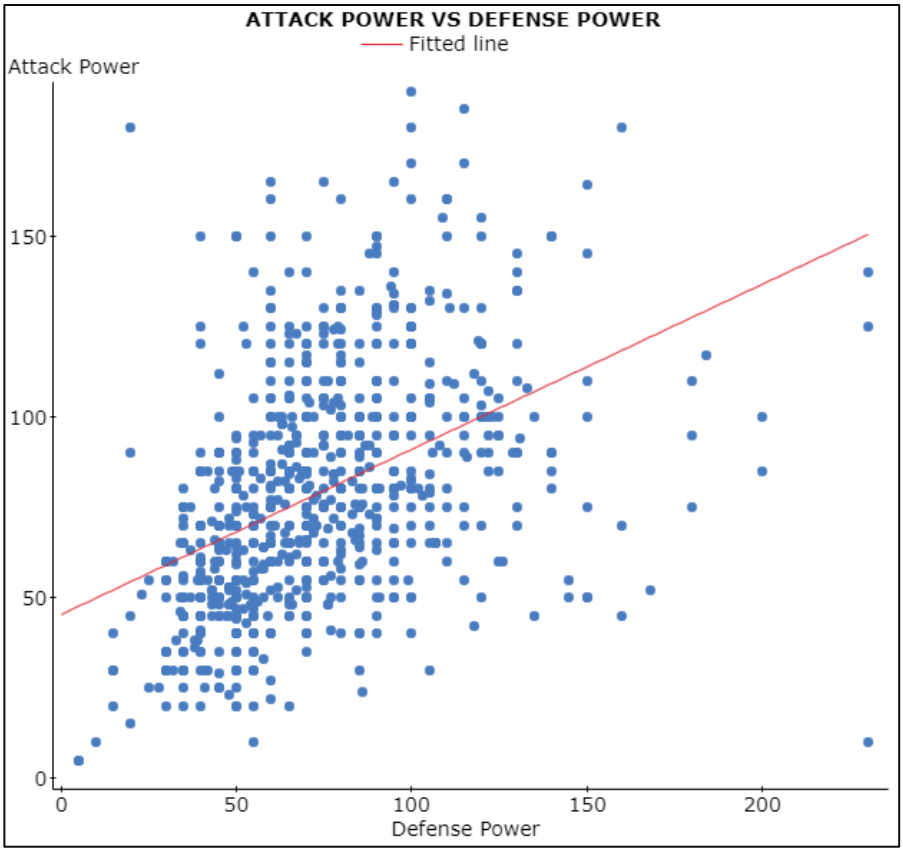
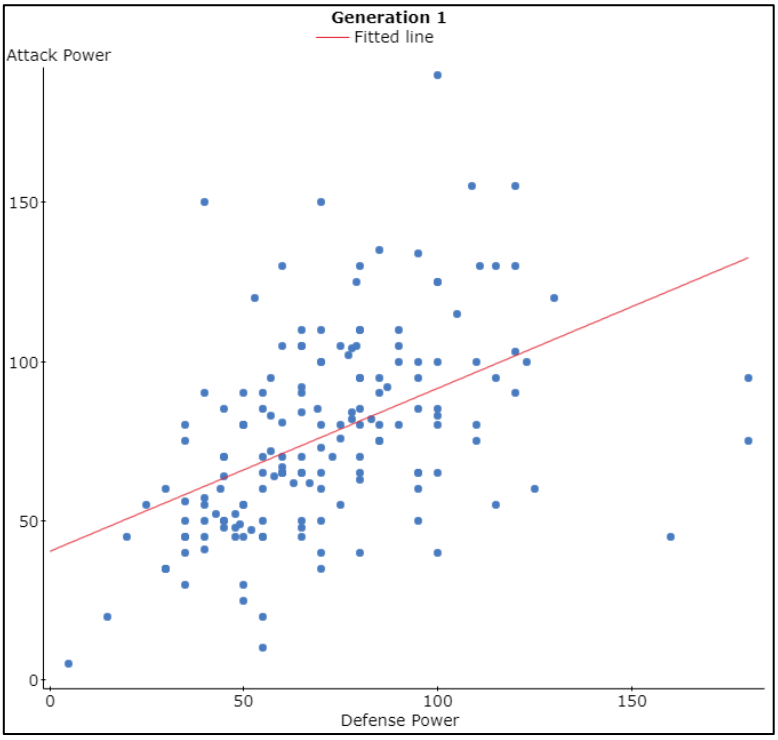


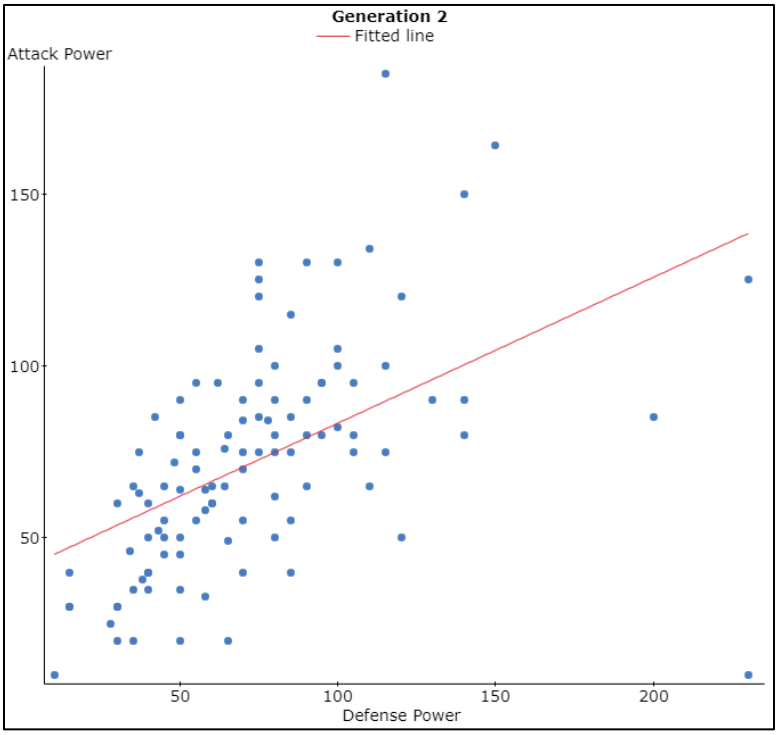
Figure 2. Shows the attack power of Pokémon VS its defense power.

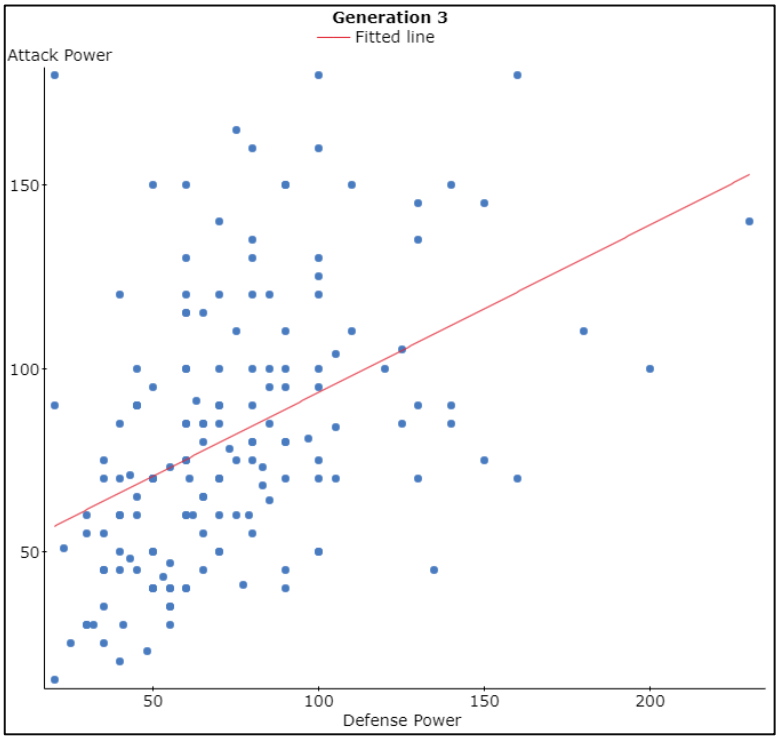
The first one is at row 232 which is pokemon name Shuckle. It has the attack power of 10 and defense power 230. It is one of the outliers because of its ability which has high defense stands and it does not have any much of attack. The second one is at row 426 whose name GroudonPrimal Groudon. It is one of the legendary pokemon(Which means it is only one of its kind and very rare to find in the pokemon world.) The GroudonPrimal Groudon can be seen as Mega evolve form of the Groudon. As explained earlier, the mega evolution is only the evolution stage of pokemon which can be reversed and pokemon can come back to its original form. Hence during the mega Evolution the power of pokemon increases. It has the attack power of 180 and

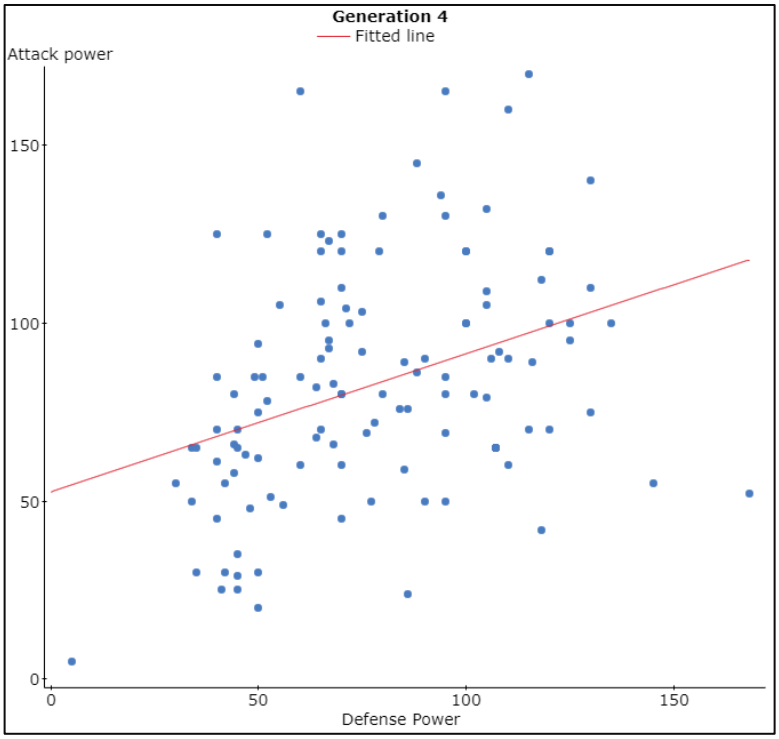
defense of 160. The third one is at row 431 whose name is Deoxys (Attack Forme). Again Deoxys is also a legendary pokemon. But it has some special forms. It can change its form according to its surroundings. There are four such forms which include the normal forme, DeoxysAttack Forme, DeoxysDefense forme and speed forme. Since in the attack form it has highest attack power and lowest defenses, it has high attack power and separates it from others. Although it can increase it power according to its will, still its total power remains the same which is a total of 600.

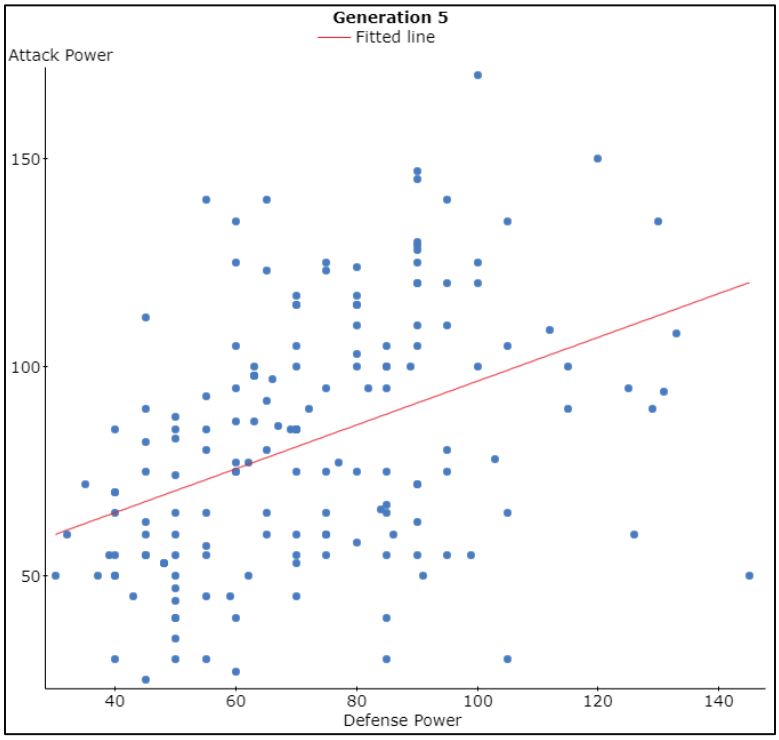
Now if we compare them by the generation we will get the following graphs











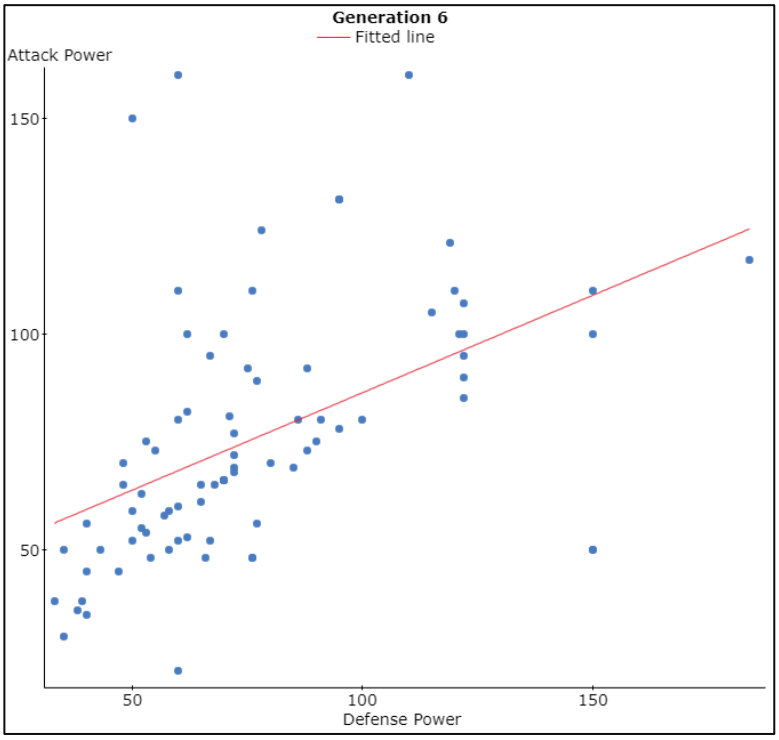


Figure 3,4,5,6,7,8 are Scatter plot of 6-different generation of Pokémon and shows the Attack VS Difference power of the Pokémon.

All these graphs though have a different number of samples, but still quite similar and are almost similar. Each graph has at least 2 outliers and most of the points are in the cluster and are near to line. The value of the R2 lies near 20% but there are not exactly equal. Though these look similar still there are the difference too. There is different sample space in each of them. Generation 4 has the most number of the outliers while generation 2 has the least number of outliers.

If we see the data as whole them and consider the data without doing any change in the arrangement we will obtain the following graph.

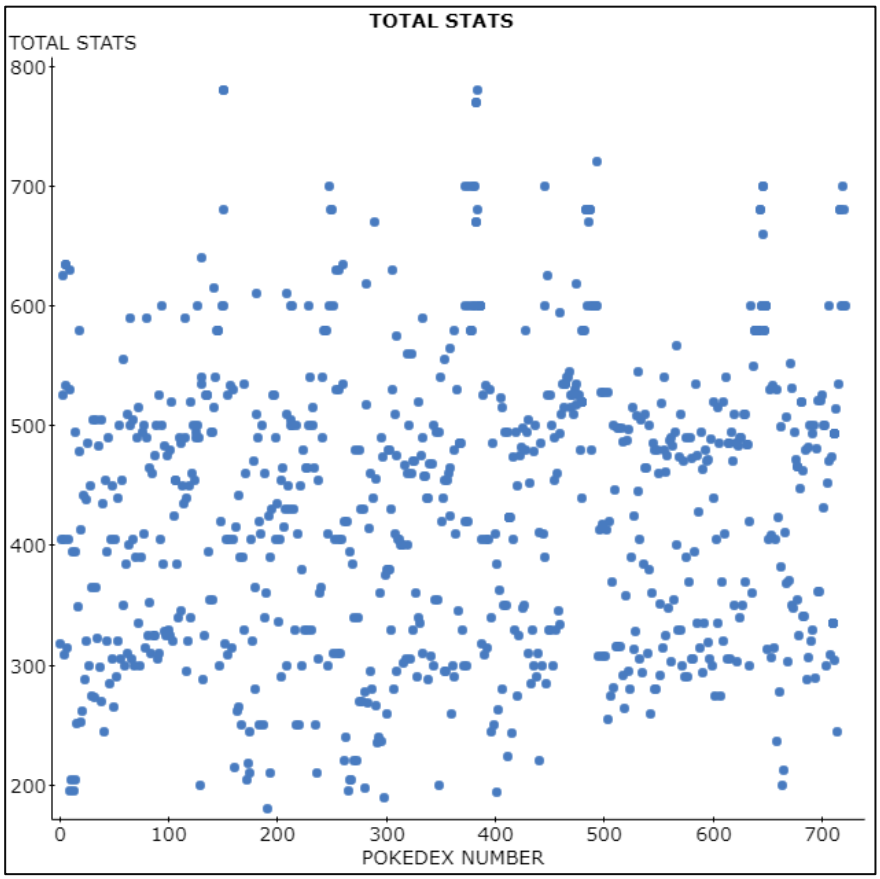


Figure 9 shows the scatter Plot of Total stats VS Pokedex Number

And it does not give us any clue about what is going on in the data. Hence to get us better representation of the data we use the histogram which gives a better representation of the data. The following is the histogram of the data we have.

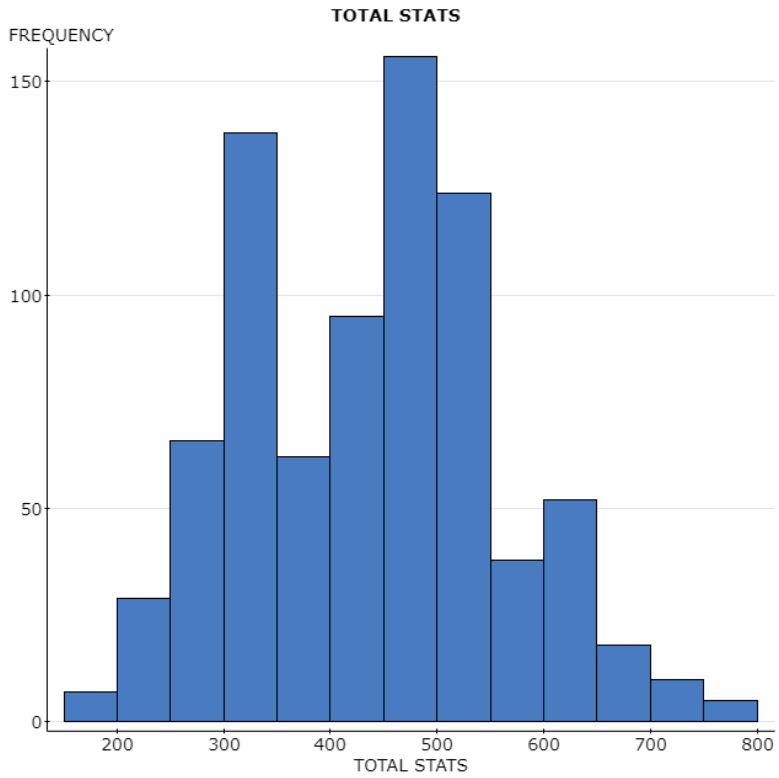


Figure10 shows the histogram of the total stats of the Pokémon's.

Here each bar represents the data of 50 interval and the bars start at 150 and end at 800. The data is Skew to Left and is unimodal and have most of the data in range of 450 t0 500. It has a median of 450 and median 435.35.

CONCLUSTION

As seen in the above data there are lot of outliers in the data but it can easily have eliminated we arrange data according to the specific order or using specific functions. Furthermore, there is close pattern in each generation and we might be able to predict what the 7^{th} or 8^{th} generation graph looks like based on the data we have till now.