Pokemon Types: How do They Matter

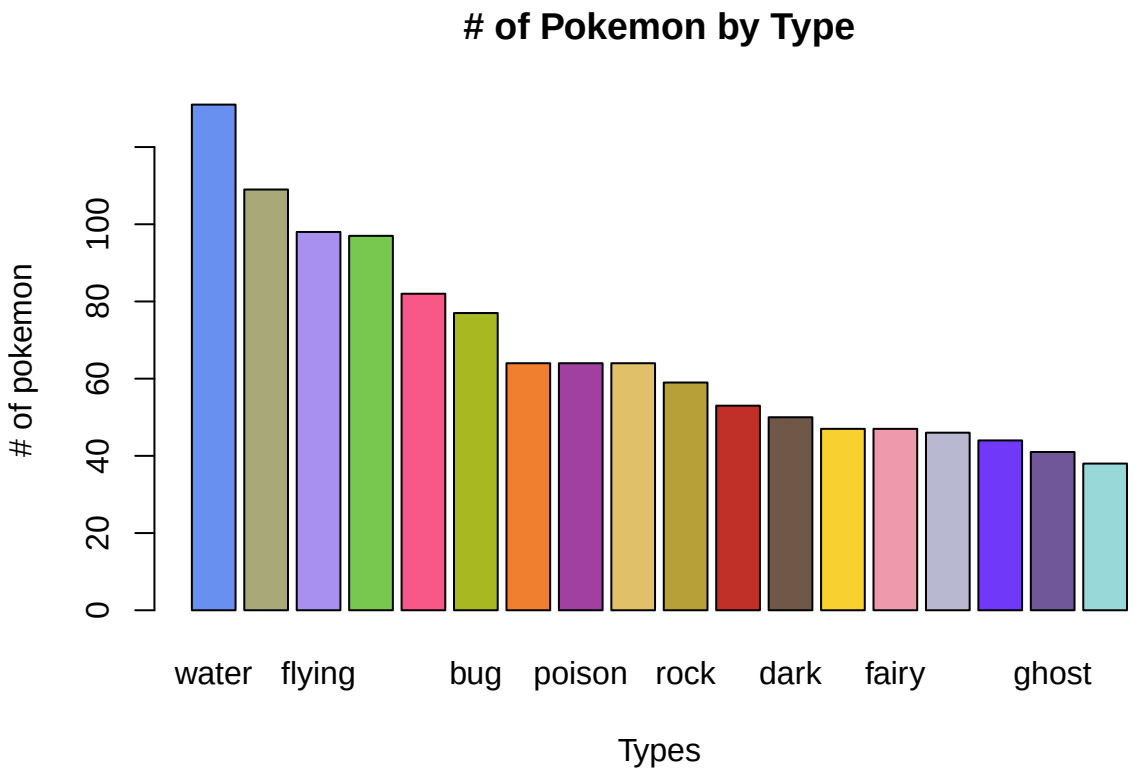
Pokemon Types and how they Affect Other Characteristics

The goal of this document is to see how the type of Pokemon matters and if all types are treated equally.

First Question: How many pokemon are there of each type?

There are many different types in Pokemon such as grass, water, and fire. These types could be analogous to their attribute or what they excel at. Water types excel at using water and so are strong against fire types (Water puts out Fire) and weak against grass (Plants soak up Water). What we want to look at is whether Gamefreak, the makers of Pokemon, split up the types across all pokemon evenly.

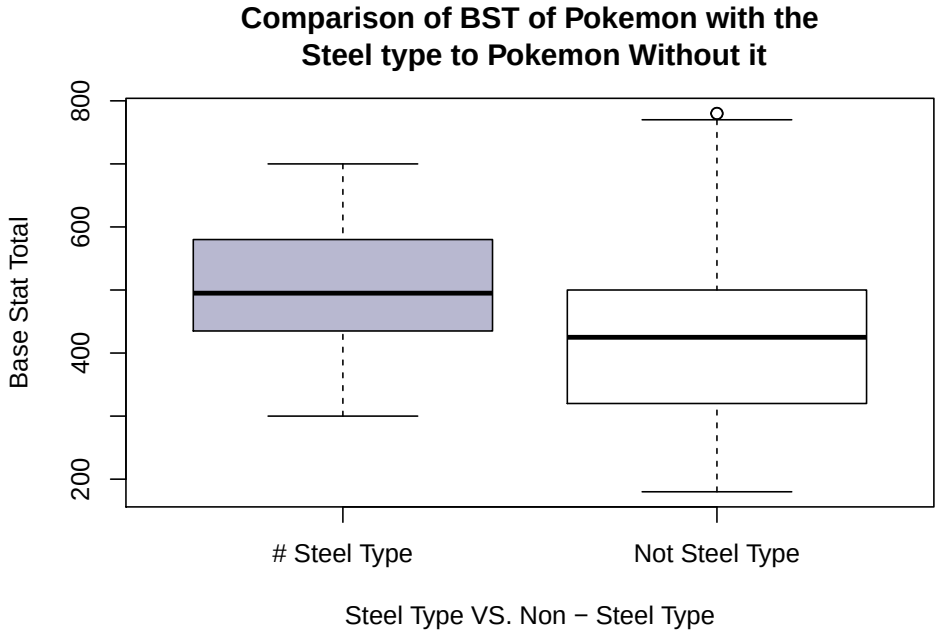
The distribution, counting pokemon with two types twice, once for each type, is as follows:

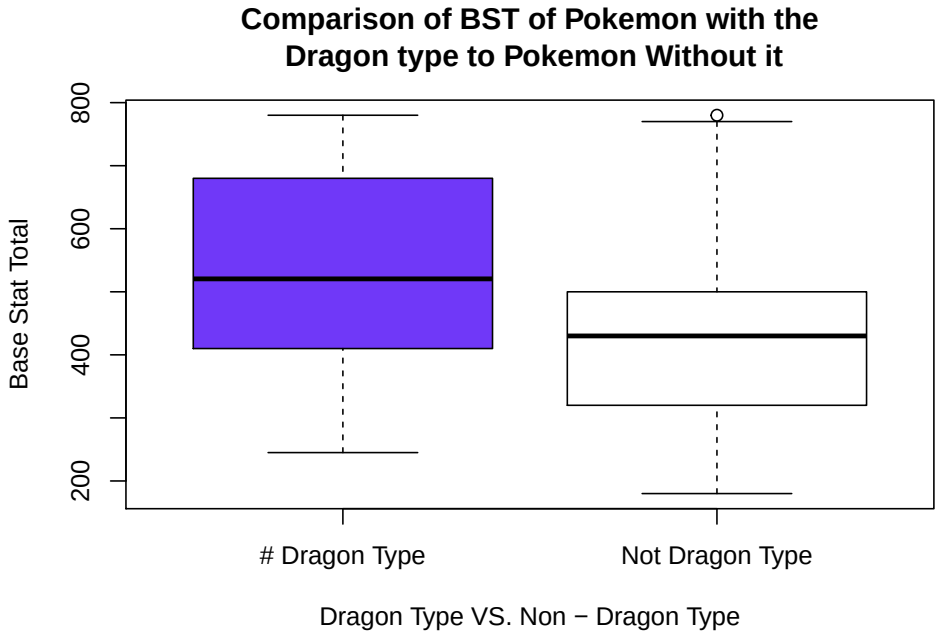


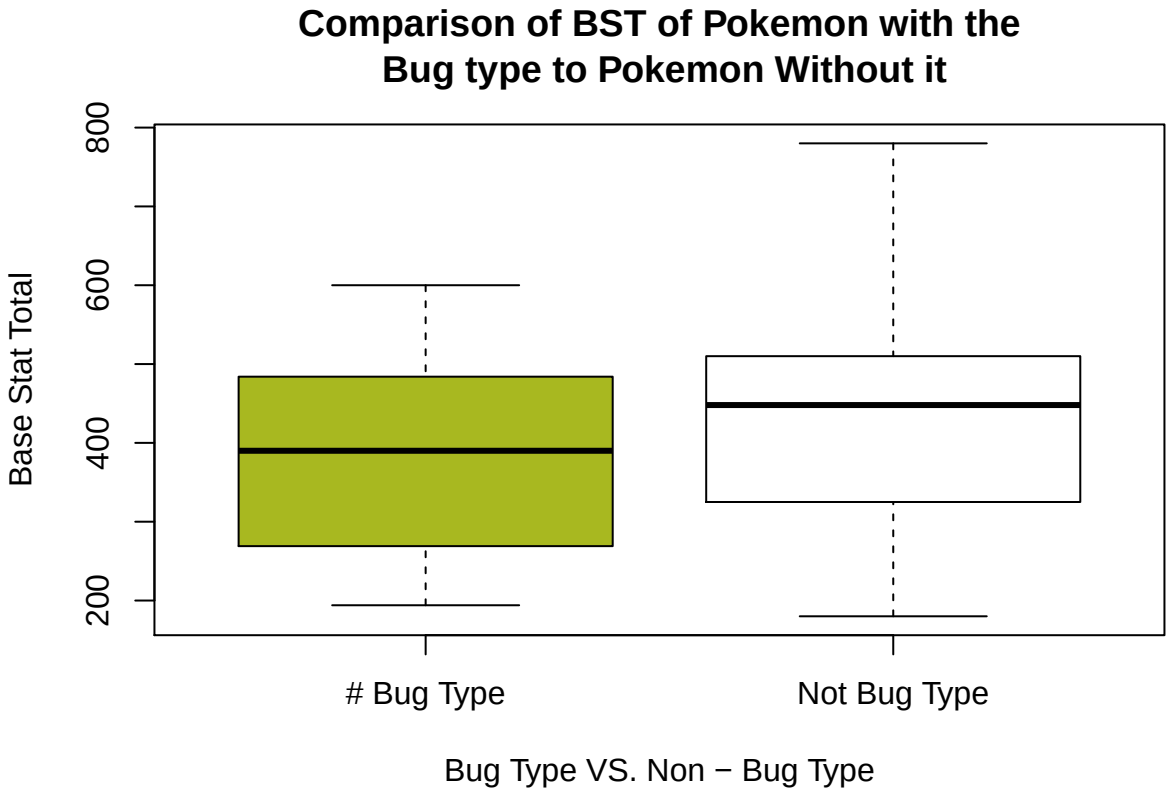
Here we see there is not an even distribution of pokemon across all the types. The most is of water pokemon and the least is of ice and ghost types. Some reasons could be that every pokemon region (area the game takes place) has an ocean/lake/sea while they don't all have a cemetery or mountain/glacier like region. Also there are only so many variations of a ghost type that could be made limiting design space for unique options.

Second Question: Is there a correlation between a Pokemon's type and it's Stats?

Every pokemon, like animals, has something they're good at. Cheetahs are fast, gorillas are strong, and turtles are defensive. In pokemon these attributes are called stats and they are as follows: speed, attack, defense, special attack, and special defense. Adding all 5 together forms the Base Stat Total (BST), which is what we are using for the overall strength of each pokemon. Baby pokemon would tend to have lower BST and legendary pokemon would tend to have higher BST. To find a correlation we would have to see the average BST for each type and compare it to the average BST of all pokemon. The following plots show the boxplots of BST for three specific types compared to all pokemon:





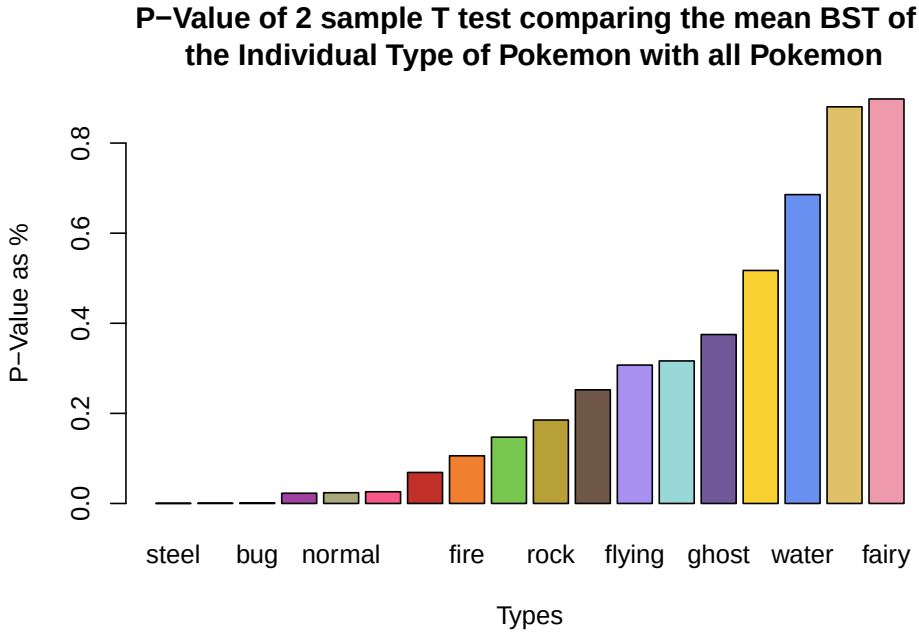


Boxplots show the median not the mean at the center but it's a better way to compare distributions of data while still getting information on the center of data. In this case Steel and Dragon have a median above the distribution of all pokemon while Bug type have a median below the distribution of all Pokemon. Simply, this means dragon types tend to be overstated and bug types tend to be understated relative to all pokemon. Now that we know the type of pokemon has an effect on its Stats, we want to understand how much this effect is for each type.

Third Question: Can we quantify this correlation and see if it's statistically significant?

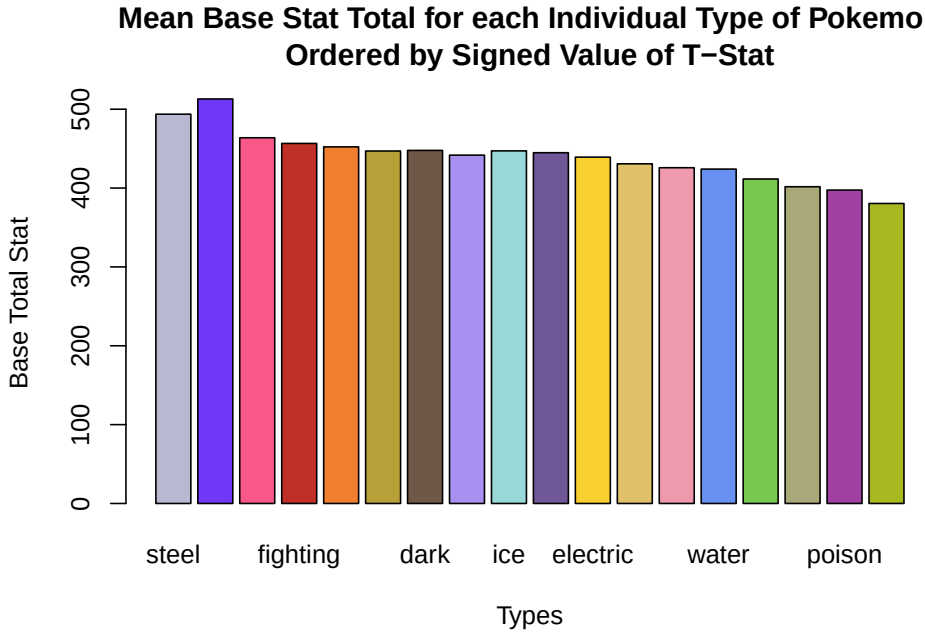
In Statistics the question 'how much?' is very important and has been studied rigourously. While there may be a difference in the mean between dragon types and all types, is this difference significant from a statistical point of view or is it insignificant. The question we are answering is "How much of difference between the mean of the group and the mean of the whole in order for the difference to indicate a meaningful change?" We accomplish this with a 2-sample t-test on each of the types compared to all Pokemon leading to 18 different P-values and 18 different T-Stats.

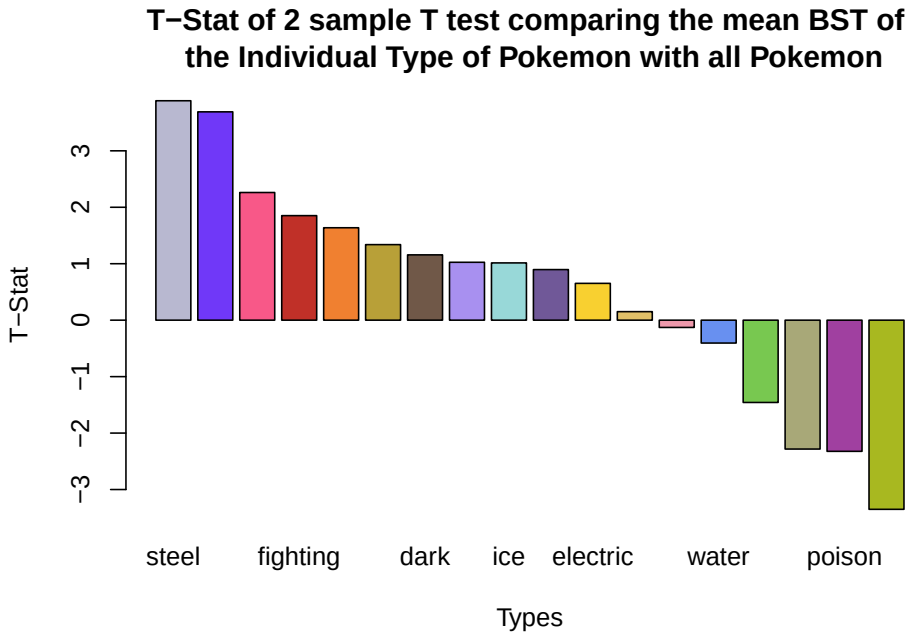
A P-Value is the probability a difference occurred due to random chance. This means we are looking for low probabilities (<5%) as they would show that the difference did not occur due to random chance. Ordered from left to right, we see the low percentages at the beginning and it looks like the first 6 are below 5%. We can confirm this by printing the p-values for the first 6 types and seeing they are all below 0.05 or 5%.



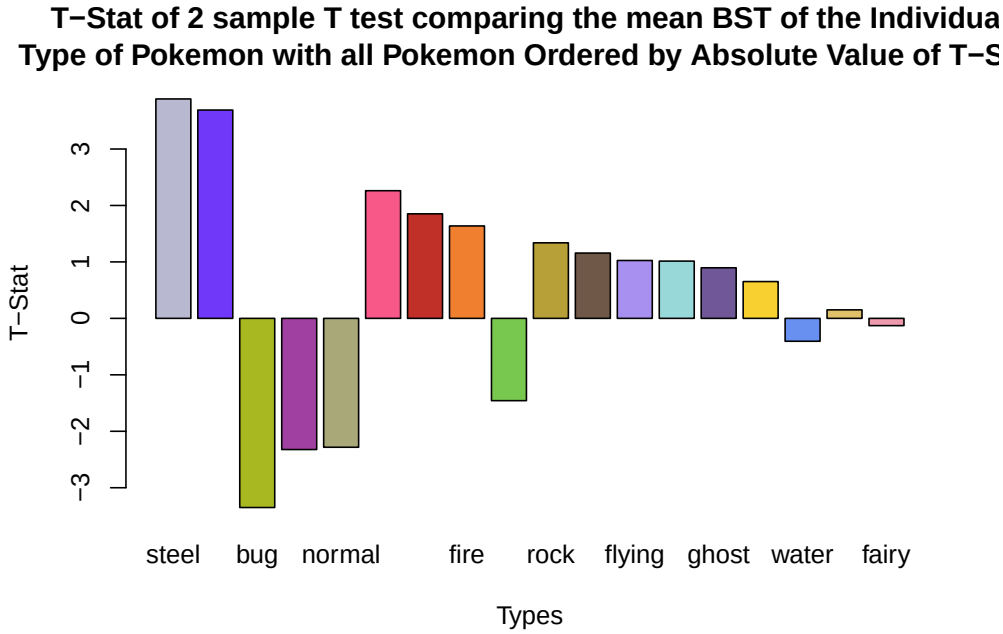

```
##      Types      P_Value
## 17   steel 0.0002929438
## 3   dragon 0.0005910994
## 1      bug 0.0011743341
## 14  poison 0.0227519072
## 13  normal 0.0238848133
## 15 psychic 0.0260415403
```

The next two graphs first show the average BST per type then the T Stat from the 2 sample t Test of that type compared to all Pokemon. As the T stat is correlated with the mean of what is being measured, BST, you'll notice a trend of the T-Stat dropping along with the Mean BST as we move from left to right. For the T Stat we want the absolute value to be over 2. We choose 2 for the same reason we choose 5% for the P-Value, it is based on what is deemed just low enough that we don't commit a Type I error, we mistake a type of having a significant difference when the difference was due to randomness, and it is just high enough that we don't commit a type II error, we miss a type that had a significant difference because we attributed it to randomness. The absolute value means we will be dealing with some negatives in the second graph which make it hard to find those that are over 2/under -2 but we fix that with the third graph.





Here we sort by the absolute value of the T-Stat instead of in order from positive to negative. Now we can can clearly see which types are over 2/under -2. If the T stat is over 2 then the mean BST for that type is greater than the overall mean for all pokemon, that type is statistically overstated. If the T stat is under -2 then the mean BST for that type is less than the overall mean for all pokemon, that type is statistically understated.



So we can quantify this correlation, with the T stats and P-Values, and the following types have a statistically significant difference, Steel, Dragon, and Psychic have a positive difference and Bug, Poison, Normal have a negative difference.

Thank you for reading through this report, please find more of my work at: https://github.com/ShanmanT/Data_Analytics_Projects