

# Assignment 2: Exploring the Pokémon Dataset

For this assignment, I chose to investigate a dataset that contains information on the Pokémon from Gen 1 to present. You can find the dataset here [Pokémon Dataset](#). The notable features that we use in this analysis include Pokémon name, generation number, % male occurrence (vs female), attack, defense, hit points (hp), type (e.g., fire, water, normal, ghost) and is\_legendary (some Pokémon are rare and supposedly very good; they are called "legendary" Pokémon).

In this project, I explored the three main research questions:

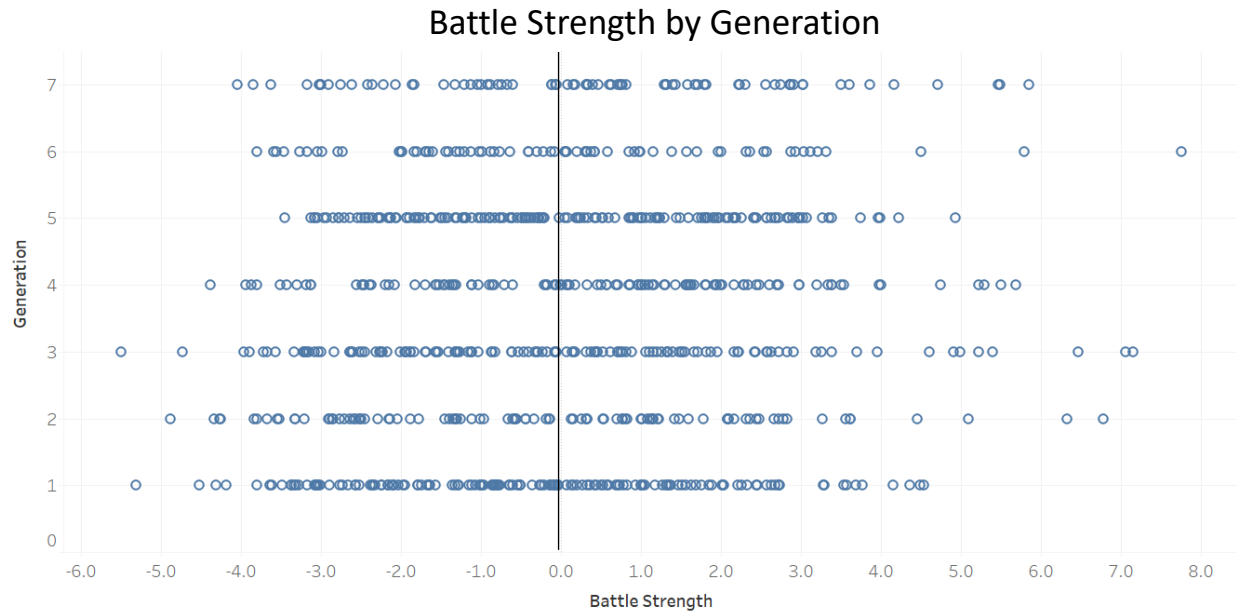
1. Does Power Creep occur in the Pokémon dataset?
2. Is the battle strength of Pokémon related to gender? (i.e., how sexist is Pokémon?)
3. What types co-occur the most frequently, and how did this change over time?



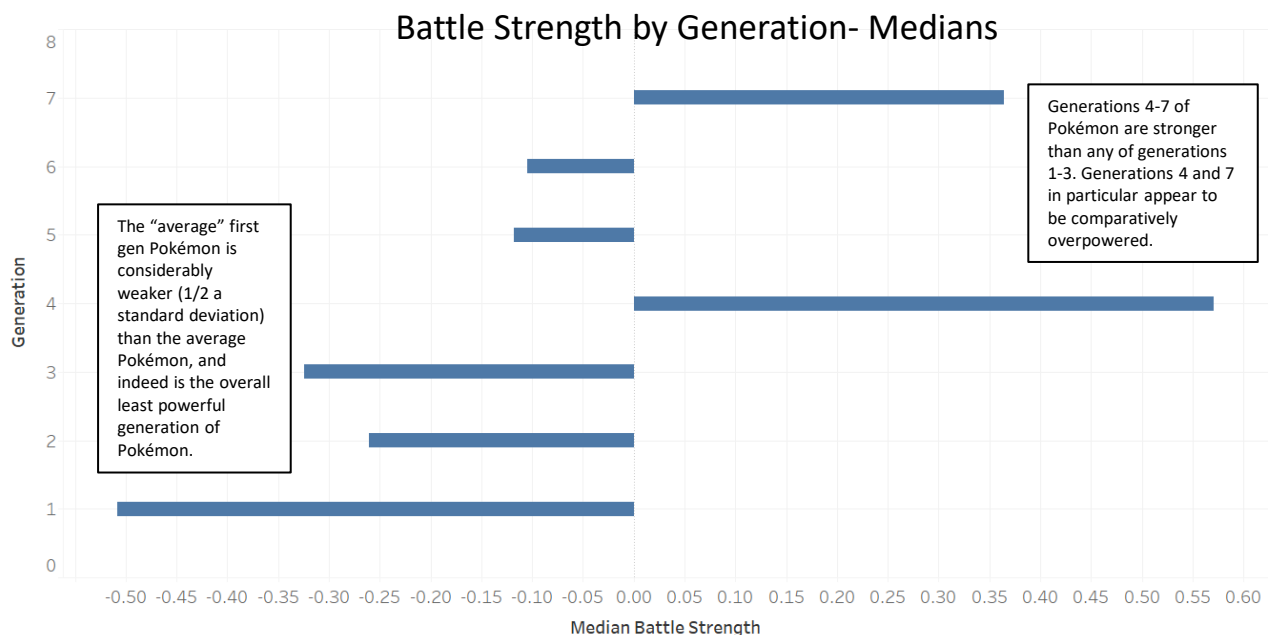
# RQ1: Does Power Creep occur in Pokémon?

Power creep is the phenomena in games where new patches or additions to a game are more powerful than those in previous patches/additions. In Pokémon terms, this maps to the question "are newer generation Pokémon stronger than older generations of Pokémon?"

As a metric for "strength" in the context of battling, I used a calculated field. In battle, the most important factors are how much damage you can do to your opponent, how strong your defense is, and how long you can stay alive (hit points). Therefore, I created a new field to proxy strength which is the sum of the Z-scores of the stats attack, defense, and hit points. To do so, for each stat, I subtracted the column mean from each datapoint and divided by the column's standard deviation. Thus, I created columns for how many std devs each datapoint was from the mean. After summing these three normalized measures, a Pokémon with a larger positive score is better than the "average" Pokémon, a Pokémon with a negative score is overall below "average", and a Pokémon with a score of 0 is "average".



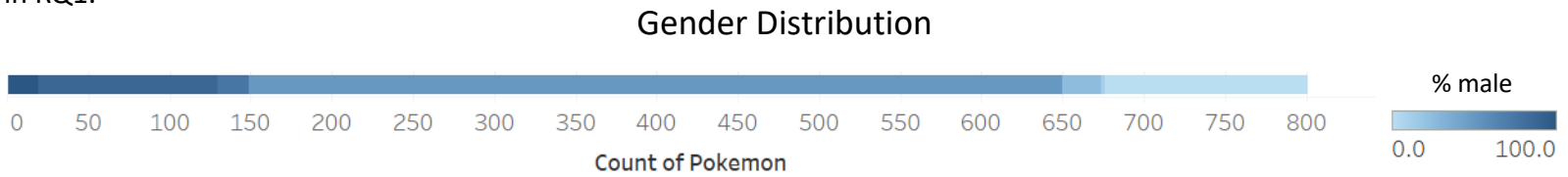
We see that, though there are a few outliers, the median and spread of the generations tends to shift to the right over time (more powerful Pokémon). To confirm, I plotted the median battle strength per generation, which is resistant to outliers.



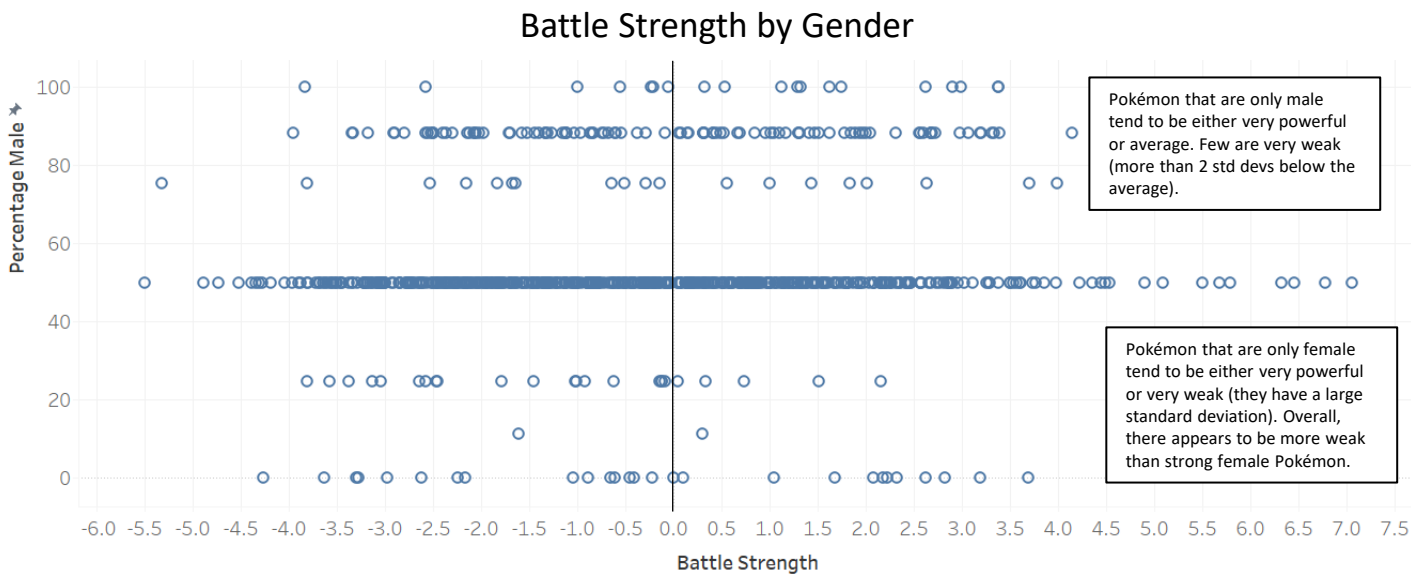
**Conclusion: evidence suggests that Power Creep is present in Pokémon.**

# RQ2: Is the battle strength of a Pokémon related to gender?

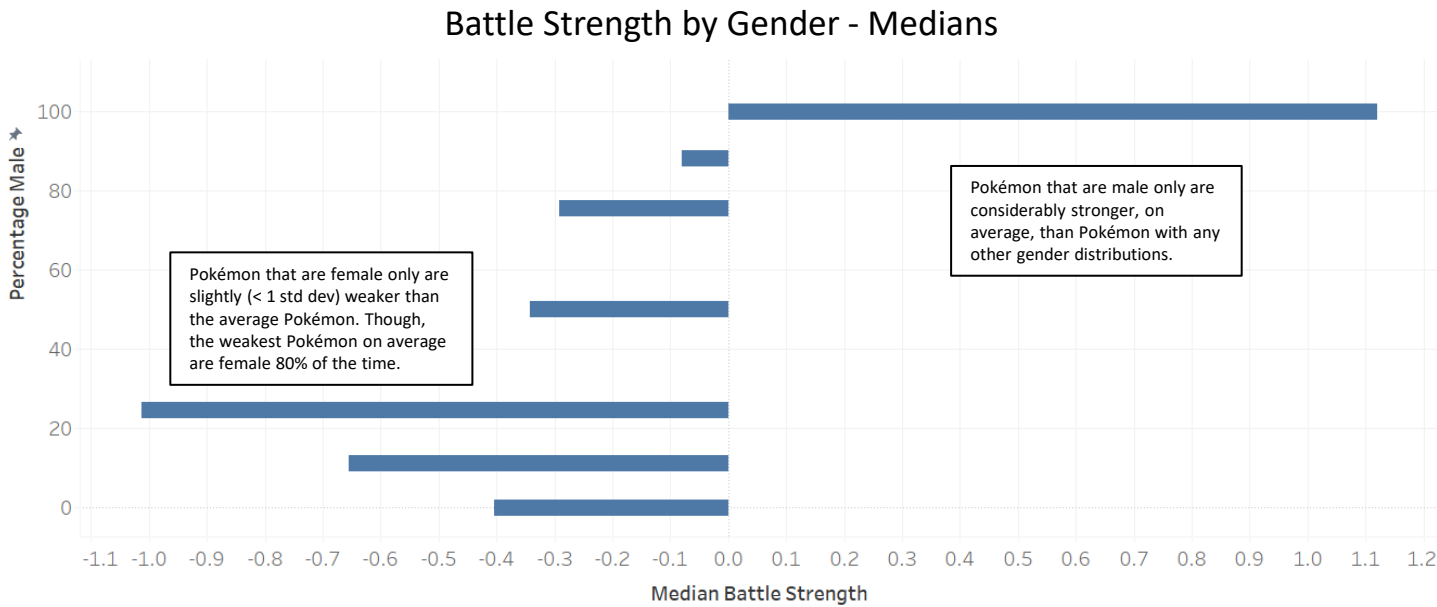
Pokémon can be one of two genders: female or male. Some Pokémon can only come in one gender, while others follow a probability distribution (e.g., 50% chance a certain Pokémon in the wild is female). With this question, I aim to see if male Pokémon are generally "stronger in battle" than female Pokémon. To quantify "battle strength" I use the same metrics as in RQ1.



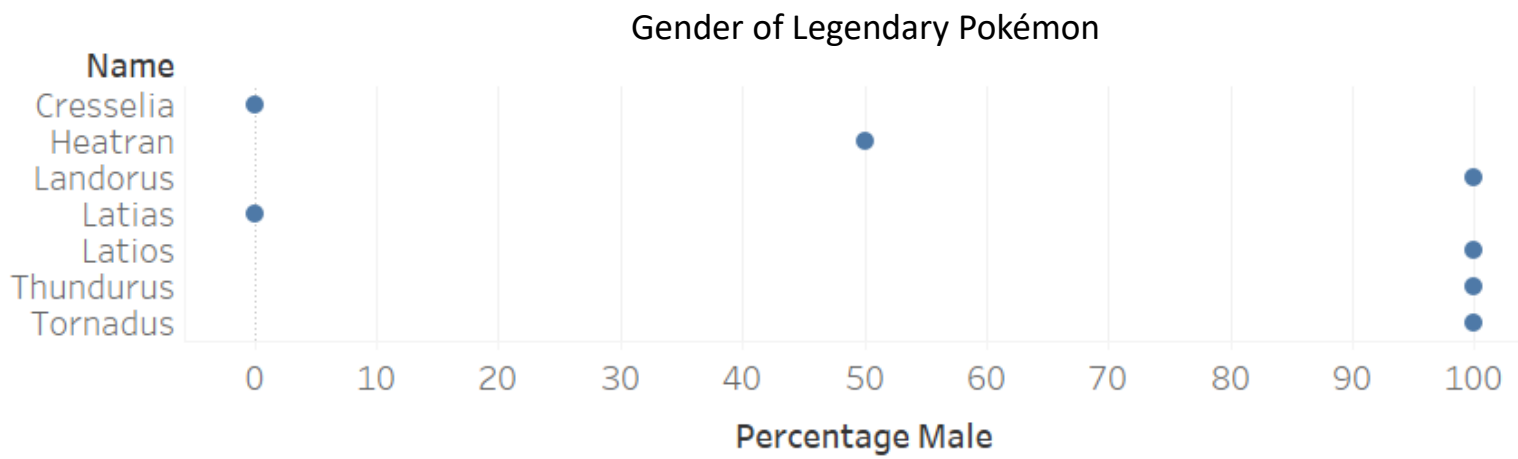
First, I plotted the overall distribution of gender. Deep blue is Pokémon that are male only, the palest blue is Pokémon that are female only. The majority of Pokémon have a 50% chance of being male or female. Though, there are ~150 Pokémon that are male only and ~125 that are female only.



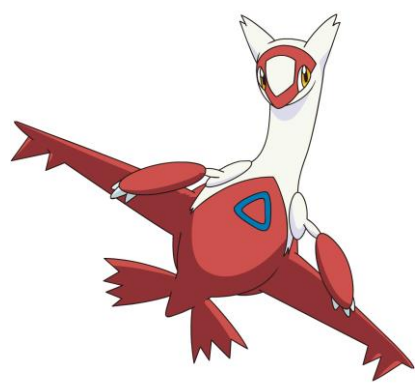
To confirm our suspicions that male Pokémon tend to be more powerful than female Pokémon, I plotted the median battle strength with gender.



Additionally, I looked at the gender distribution for legendary Pokémon, which have a reputation of being very rare and powerful (the `is_legendary` attribute from the dataset). Unfortunately, of the 70 legendary Pokémon, only 7 had gender information, which is shown below.



This is far too small of a sample to conclude anything concrete, but if this small sample is somewhat representative, then we might again be seeing that male Pokémon are stronger than female Pokémon.



Latias



Tornadus

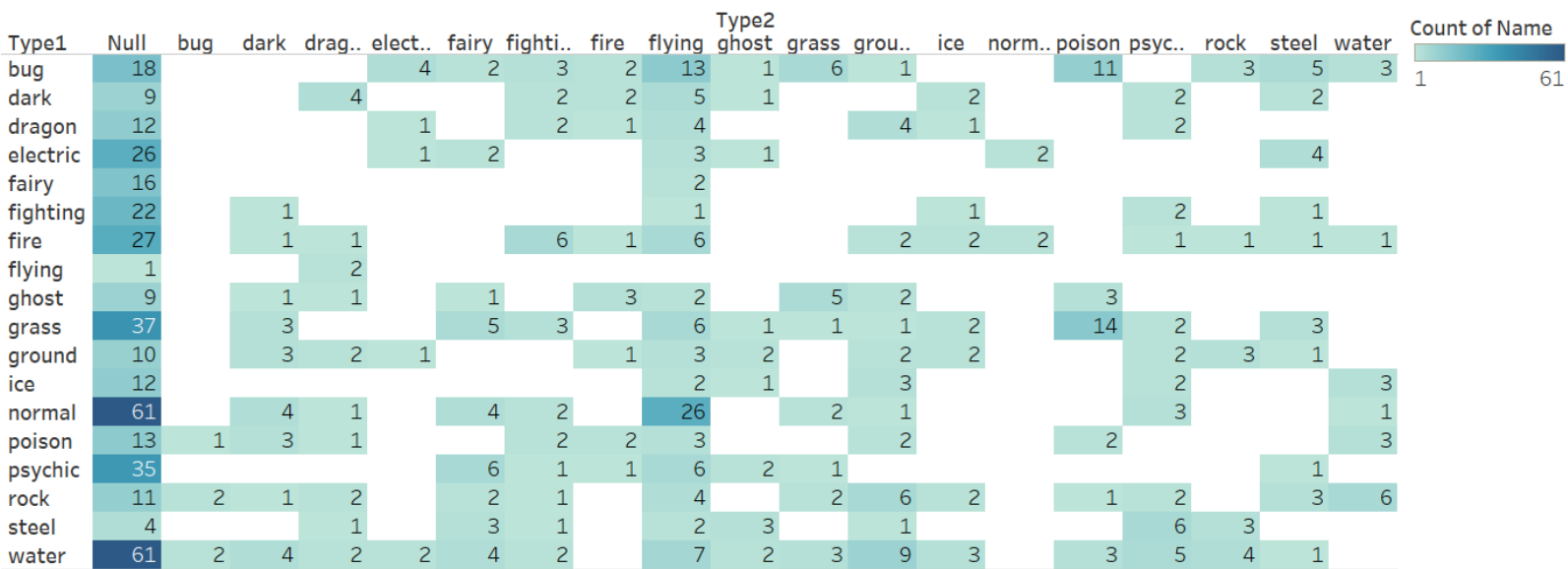
**Conclusion: evidence suggests that male Pokémon are better in battle than female Pokémon (which I argue is a bit sexist).**



# RQ3: What types of Pokémon are likely to co-occur? Does the co-occurrence matrix change over time (i.e., generations)?

Pokémon have 1-2 "types" (e.g., fire, water, normal, flying). First, we analyzed the overall co-occurrence matrix for these traits for the full dataset:

Type 1 vs Type 2 in all Pokémon



We see that Pokémon are more likely to have only one type than two; this data is represented in the first column. There is a wide variety of type combinations (most of the squares in the matrix are shaded), and if we look at the sparse columns or rows, we see that some types are more likely to be listed as type 1 (e.g., bug, water, normal) and others are more likely to be listed as type 2 (e.g., flying, fighting).

Overall, the types that are most likely to occur are normal-flying, bug-flying, grass-poison, and bug-poison



Pidgy  
normal-flying



Ledyba  
bug-flying

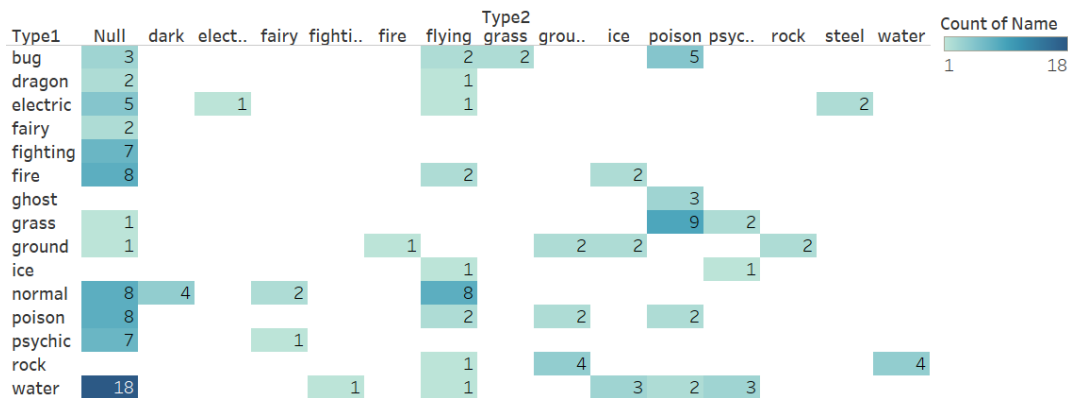


Oddish  
grass-poison

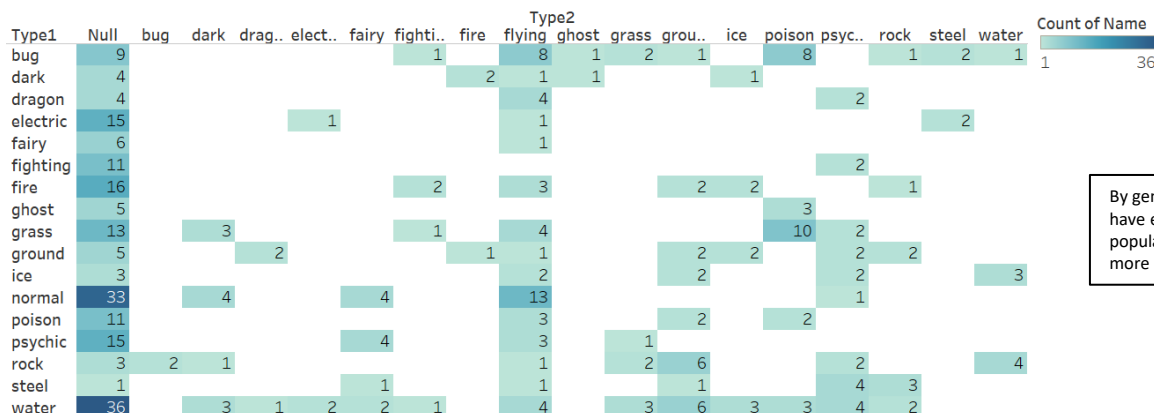


Venonat  
bug-poison

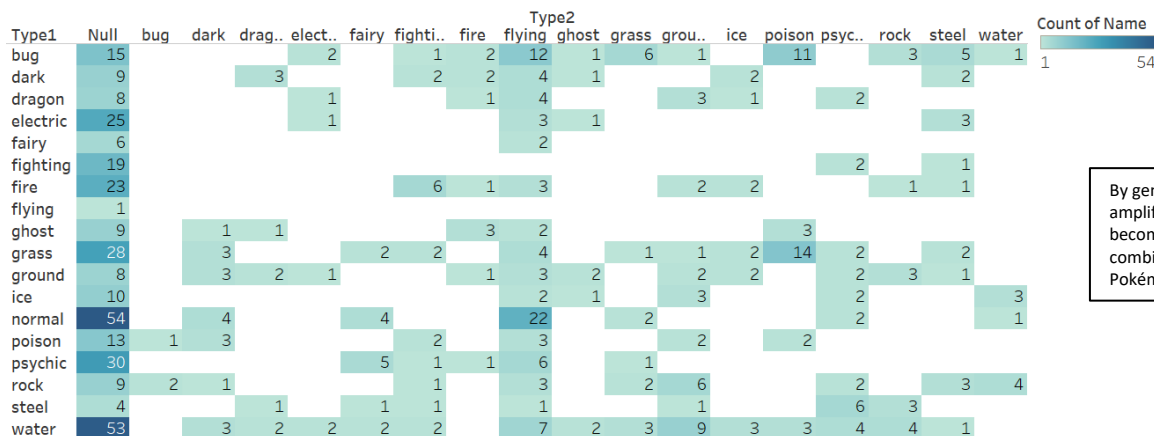
# Type 1 vs Type 2 cumulative occurrence in generations 1, 3, 5, 7



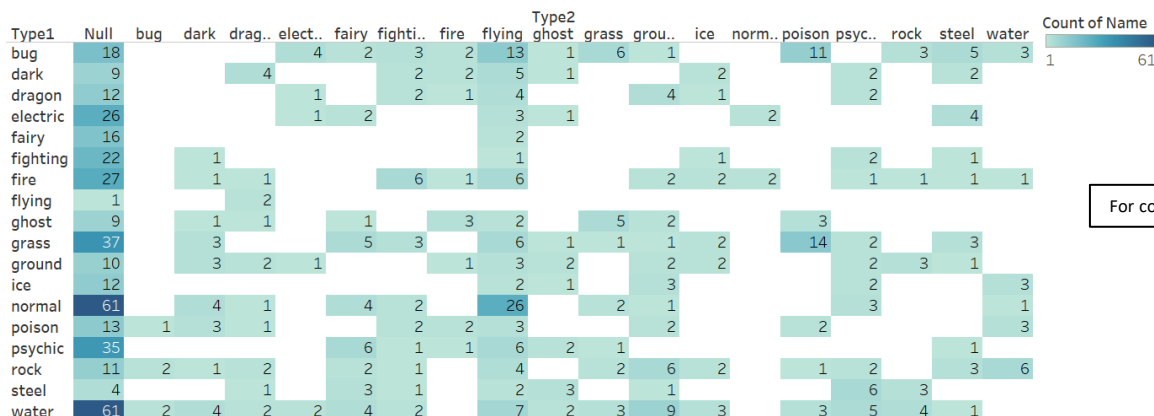
In gen 1, the matrix is rather sparse. Flying and Poison are the two most common types to co-occur. We see that normal-flying, grass-poison, and bug-poison are already popular combinations.



By gen 3, water only and normal only have established themselves as the most popular types. Bug-flying starts becoming more popular as well.



By gen 5, the trends from gen 3 are amplified. The matrix is also starting to become less sparse, as more creative combinations of types are used in Pokémon.



For comparison, we present the gen 1-7 data.

**Conclusion:** while Pokémon are receiving more creative combinations of types as time passes, the prevalence of the types water only, normal only, normal-flying, bug-flying, grass-poison, and bug-poison were established early on in Pokémon.