An Analysis of Competitive Pokémon Usage on Pokémon Showdown

JSC370 - Final Project (https://github.com/tirangol/JSC370-Project)

Richard Yin

30 April 2025

Introduction

Pokémon (or Pokemon) is a role-playing video game series wherein player characters travel across a fictional region, capturing and training teams of creatures called Pokemon to use in battles against other Pokémon and Pokémon trainers to accomplish various objectives. Central to its gameplay is the turn-based battle system which has fostered a wide-reaching competitive scene, with tournaments like the Pokemon World Championships boasting prize pools of over $2,000,000 USD (The Pokémon Company, 2025).

There are many competitive formats to a Pokemon battle with various rules and clauses. In the common singles format, two opposing trainers field 1-6 Pokemon, which possess:

- An ability
- A level (which is usually set to the same for all players)
- An optional held item
- 1-4 unique moves
- 1-2 unique types (there are 18 total types, see Figure 1)
- Base statistics for hit points (HP), attack, special attack, defence, special defence, and speed.

As a battle begins, trainers send out a single Pokemon. Each turn, players can use a move from their current Pokemon or switch to a different Pokemon; they are unaware of their opponent's choice during this time. Once both sides have decided to act, events play out as such: switching occurs first, followed by moves with higher priority levels, then finally the Pokemon with higher speed acts first. If a Pokemon loses all of its HP ("faints") at the end of a turn, its trainer sends out another Pokemon and the next turn begins; if no more Pokemon can stand, the trainer loses the battle.

To win a battle, it is important for competitive players to build teams with a wide range of types and moves rather than simply those with the highest base statistics, since nuances in moves and Pokemon typing add significant complexity to the battle system. For instance, moves are labelled by types and may deal extra ("super effective") or reduced damage based on the defending Pokemon's type(s) (see Figure 1 below). Moves are further classified as "physical" or "special". Physical moves use the attacking Pokemon's attack against the defender's HP and defence to calculate damage, while special moves use special attack against HP and special defence. Finally, non-damaging moves have a variety of effects like temporarily changing a Pokemon's stats, healing, causing weather or status conditions (e.g. hail, rain, poison, burn) which can affect Pokemon on the field, and creating hazards that damage or weaken Pokemon upon switching in.

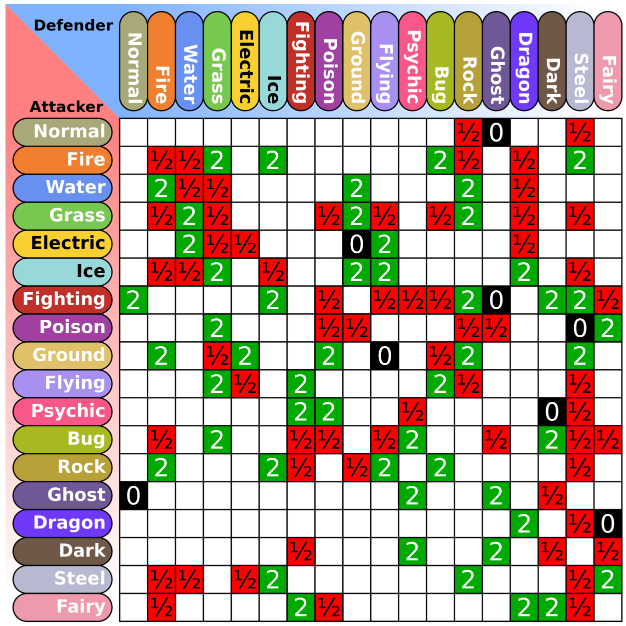

Figure 1. Pokemon type chart. Cell numbers (blank = 1) are damage multipliers for one type against another (e.g. the cell with row Fire and column Grass is 2, so Fire does x2 damage on Grass types). If a Pokemon has two types, the multipliers for both types are used (e.g. Fire does x2 damage on Grass types and Steel types, so it does x4 damage on Steel/Grass types). Source: Wikipedia

With this in mind, this project aims to uncover insights regarding Pokemon team building by answering the research question, **"What properties in a Pokemon most significantly influence its usage rate in competitive singles battles?"** Generation 7, the Alola region introduced by Pokemon Sun/Moon, was selected as the main focus because it has the largest pool of playable Pokemon. Furthermore, this analysis only considers Pokemon, base statistics, types, and moves, ignoring items and abilities. This is because items and abilities have wildly-differing effects, making it difficult to collapse into simpler categories; moreover, items are not unique to a Pokemon, while many abilities can uniquely identify individual Pokemon, which can lead to an overfitted analysis.

Data in this study were collected largely from Smogon University (or Smogon), a website specializing in competitive Pokemon resources. Smogon has forums, damage calculators, as well as a Pokedex, a table of all Pokemon with base statistics, abilities, movesets, and community team-building strategies for each generation. Smogon also compiles monthly statistical analyses of Pokemon battles in the form of TXT files, which include raw counts of Pokemon usages as well as common abilities, items, and moves (data from January 2019 was used in this study). These battle statistics are sourced from Pokémon Showdown!, a free Pokemon battle simulator that allows for building Pokemon teams, playing against random opponents online, and competing with a ranked Elo system. Pokemon evolution data (explained later) were collected from Veekun, an independently-created Pokedex website with a more complex search system.

Methods

Data Retrieval and Cleaning

Pokemon and Moves Data

To obtain generic Pokemon data including base statistics, move details, and type effectiveness, Smogon's generation 7 Pokedex was scraped. This site was selected because its user interface was simplistic, its move descriptions were more standardized and technical than official in-game descriptions, and its Pokemon naming conventions would hopefully match the monthly battle statistics (since Smogon created both resources).

The displayed tabular data was dynamically injected from a Javascript object in a `<script>...</script>` tag and thus did not show up in the fetched HTML. Thus, the object was isolated and parsed into JSON, where it became clear the JSON contained a set of dataframes for Pokemon types and base statistics, moves, items, abilities, and offensive type matchups; only the first two dataframes were important for this study. Nonstandard Pokemon (i.e. fanmade Pokemon from Smogon's Create-A-Pokemon project) were filtered out, pokemon type was converted from a variable-length array to a string, then the type effectiveness dataframe was refitted into a matrix so both offensive and defensive matchups could be more easily indexed. Features were then extracted from the moves and Pokemon dataframes (see next section).

It should be noted that some Pokemon abilities (e.g. Levitate, Thick Fat) change the defensive type matchings, but these are not covered by this analysis because there also exists abilities which can nullify these effects (e.g. Mold Breaker) and the uncertainty around these extra defensive changes make it difficult to incorporate meaningfully into analysis.

Usage and Common Moves Data

To acquire the response variable – usage – one of Smogon's monthly battle statistics were scraped. The usage number of a Pokemon in a TXT file is defined as a Pokemon's number of appearances across all matches in the TXT file's specified generation, tier, and month; teams with the same Pokemon or opposing teams with matching Pokemon can add to the count, potentially resulting in > 1 Pokemon appearance per match. Data from January 2019 were selected in hopes of usage rates being relatively stable due to the time period being one of the last months before the next Pokemon generation.

This analysis included the tiers Ubers, OU (overused), UU (underused), RU (rarely-used), NU (never-used), PU, ZU (zero-used), and LC (little cup), which are defined in descending order by Smogon (n.d.) based on usage rates and community adjudication; Pokemon from lower (i.e. worse) tiers may be used in higher tiers, but not vice versa. Tiers were incorporated to create a richer analysis, since without them the vast majority of Pokemon have no utility and thus would never be used at all.

After web scraping the TXT file, regex was used to extract the rows corresponding to names, raw counts, and most common moves (a variable-sized list). Moves representing "Nothing" (if < 4 moves were chosen for the Pokemon) and "Other" were removed. For each Pokemon, its raw counts were summed across tiers, while the most common moves from the most-played tier were extracted. It was initially attempted to use the most common moves from the Pokemon's intended tier (specified by the Pokemon dataframe in Table 1), but some Pokemon had no usage for their specified tier, suggesting that the tier list changed since January 2019. This analysis focused on commonly-used moves instead of entire move pools because the former is more informative of a Pokemon's competitive strategy.

Pokemon Evolutions

Pokemon has an evolution mechanic in which certain Pokemon can change into new Pokemon under certain conditions (typically when their level reaches a threshold). Evolution chains are at most three Pokemon long;

Pokemon can thus be categorized as first-evolutions, middle-evolutions, final-evolutions, or non-evolving. As a potential confounding variable, evolution was included in this study as a control.

For every possible evolution status, a query was made to fetch the table returned by Veekun's Pokedex search, then the Pokemon names were combined with the evolution status to form a dataframe. To match Smogon naming conventions, many manual corrections were made:

- Adjusting capital letters
- Changing apostrophes to fit UTF encoding
- Removing unique names for "default" forms (e.g. "Normal Deoxys" to "Deoxys")
- Removing accents (e.g. "Flabébé" to "Flabebe")
- Hyphenating names (e.g. "Alolan Dugtrio" to "Dugtrio-Alola").
- For Pokemon with forms for each type (e.g. Arceus-Bug, Arceus-Grass, etc.), a row was created for each unique type

Data Wrangling

The moves dataframe was merged with the top 5 most common moves of each Pokemon; the 5 top was used instead of top 4 in order to account for small variations in play-style. Once feature engineering was performed (see next section), the move features were merged with the Pokemon dataframe and evolution dataframe. Inconsistent naming conventions were identified in the process and manually resolved:

- Hidden Power is a unique move whose type can vary across even the same Pokemon. It was notated in the TXT file by its actual type (e.g. 'Hidden Power Ice') but only as 'Hidden Power' in the moves dataframe, where new rows for each type had to be manually added.
- Differences in writing conventions existed between dataframe and TXT files for some multi-form Pokemon like 'NidoranM' and 'NidoranF' (to 'Nidoran-M' and 'Nidoran-F'), as well as 'Necrozma-Dusk-Mane' and 'Necrozma-Dawn-Wings' (to 'Necrozma-Dusk Mane' and 'Necrozma-Dawn Wings')

The Pokemon dataframe had 988 observations, which became 931 after merging. Three Pokemon genuinely saw zero usages – Tranquill, Gothita, and Scatterbug – while others were either not introduced in the mainline Pokemon game (e.g. Melmetal, Eevee-Starter, Pikachu-Starter) or were a specific Pokemon form. Some forms were only cosmetic in nature (e.g. Pikachu-Hoenn, Pikachu-Unova), so they were removed, their usages added to the base form; others were related to abilities that change a Pokemon's form in the middle of a game (e.g. Aegislash, Wishiwashi), which may affect some base statistics, resulting in extra rows created for this Pokemon's forms, which this analysis did not consider for the sake of simplicity.

No outliers and implausible values were detected in this final dataset.

Feature Engineering

Relevant initial features in all the scraped dataframes are shown in Table 1. Features that were manually engineered from the Pokemon dataframe's features are shown in Table 2. It should be noted that the `best_stat` feature is calculated based on the best statistic out of `atk`, `spa`, `bulk`, `spbulk`, and `spe`:

- 'spe' if it is `spe`
- 'atk' if it is `atk` or `spa`, where `atk` and `spa` differ by at most 10
- 'phatk' if it is `atk` and `atk > spa + 10`
- 'spatk' if it is `spa` and `spa > atk + 10`
- 'bulk' if it is `bulk` or `spbulk`, where `bulk` and `spbulk` differ by at most 10
- 'phbulk' if it is `bulk` and `bulk > spbulk + 10`
- 'spbulk' if it is `spbulk` and `spbulk > bulk + 10`
- 'none' if the difference between the best and worst stat is at most 10

Table 3 lists the features engineered from the move dataframe Apart from `priority` and `atk_{type}`, the logical-based features used string matching on the move description to filter for for a specific effect.

Table 1. Relevant initial features in the Pokemon (left) and moves (right) dataframe

Pokemon	Type	Description	Move	Type	Description
name	chr	Pokemon's name	name	chr	Move's name
hp	int	Hit points	category	chr	Either 'Physical', 'Special', or 'Non-Damaging'
atk	int	Attack	power	int	Move's damage, if it is damaging (0 otherwise)
def	int	Defence	accuracy	int	Probability of a move successfully executing
spa	int	Special Attack	priority	int	Move's priority level (-7 to 7)
spd	int	Special Defence	description	chr	Move's effects
spe	int	Speed	type	chr	Move's type
types	chr	Pokemon's types, comma-separated			
formats	chr	Pokemon's Smogon tier			

Table 2. Engineered features from the Pokemon dataframe

Feature	Type	Description
total	int	Sum of all base statistics
stdev	int	Standard deviation of all base statistics
bulk	num	(hp + def) / 2, a more well-rounded measure of defence
spbulk	num	(hp + spd) / 2, a more well-rounded measure of special defence
typecount	factor	Number of unique types, either 1 or 2
best_stat	chr	Estimation of the Pokemon's primary strength. Can be 'atk', 'phatk', 'spatk', 'bulk', 'phbulk', 'spbulk', 'spe', 'none'
weaknesses	num	Number of defensive weaknesses (where x2, x4)
resistances	num	Number of defensive resistances (where x0.5, x0.25, x0)
is_{type}	logi	Whether the pokemon is {type} (for all 18 types)
def_{type}	factor	The pokemon's defensive multiplier against {type} (for all 18 types). Can be 0, 0.25, 0.5, 1, 2, 4

Table 3. Engineered features from the moves dataframe. These features apply to the top 5 most common moves of a Pokemon.

Feature	Type	Description
prop_physical	num	Proportion of moves that are physical
prop_special	num	Proportion of moves that are special
prop_nondamaging	num	Proportion of moves that are non-damaging
num_inaccurate	int	Number of moves with accuracy $<$ 100%
priority	logi	If there is a move with priority $>$ 0
stat_change	logi	If there is a non-damaging move that raises/lowers stats
multisetup	logi	If there is a non-damaging move that does something for multiple turns
heal	logi	If there is a move that heals
status	logi	If there is a non-damaging move that inflicts a status effect (i.e. burn, poison, paralysis, sleep, confusion)
hazards	logi	If there is a hazard-related move (hazards damage Pokemon on switch-in)
switch	logi	If there is a move that forces/prevents a Pokemon switch
recoil	logi	If there is a move, typically a powerful move, that also hurts the player
ohko	logi	If there is a move that involves instantly fainting the opponent or player
flinch	logi	If there is a move that can flinch the opponent, skipping their move

Feature	Type	Description
selfweaken	logi	If there is a move, typically a very powerful move, that lowers the player's stats after usage
multihit	logi	If there is a move that hits multiple times
protect	logi	If there is a move that protects the player from something for one turn
atk_{type}	logi	If there is a super-effective move against {type} (for all 18 types)
num_super_effective	int	Number of unique types that can be hit super-effectively from the moveset; a Pokemon's offensive coverage

Data Exploration

The following tools were utilized for data exploration:

- `dplyr` to calculate means and proportions of variable values.
- `ggplot2` to create histograms of Pokemon usages, box plots of usage against multiclass categorical variables (e.g. tier, best stat, number of types), scatter plots of usage against continuous variables (e.g. base statistics), 2D heatmaps/histograms of usage against small count variables (which would otherwise clump together in a regular plot, e.g. number of weaknesses/resistances), and a correlation matrix of base statistics to each other.
- `patchwork` to combine topically-related `ggplot` diagrams into a single figure.
- `kable` to display the dataset and results of ANOVA tests conducted with `anova`.
- `lm` to fit simple linear regression models of base statistics to usage.

Models and Tests

Descriptions of models/tests, training/testing

Statistical destails on models and test statistics (e.g. Rˆ2)

Results

Exploratory Data Analysis

Usage

The distribution of usages is extremely right-skewed, with a mean of 37,309 far exceeding the median of 7,981, and a logarithmic-scale histogram of usages appearing vaguely symmetrical (Figure 2). Being an example of count data, it seems appropriate to model the counts as a Poisson distribution (where $\mathrm{E}[X] = \mathrm{Var}[X]$), however the usage's variance of 7,486,205,609 suggests there is far more variance than expected, justifying the use of an overdispersed model like the Negative Binomial. Figure 2's appearance is also roughly normal, making a linear model fitted to `log(usage)` a reasonable approach.

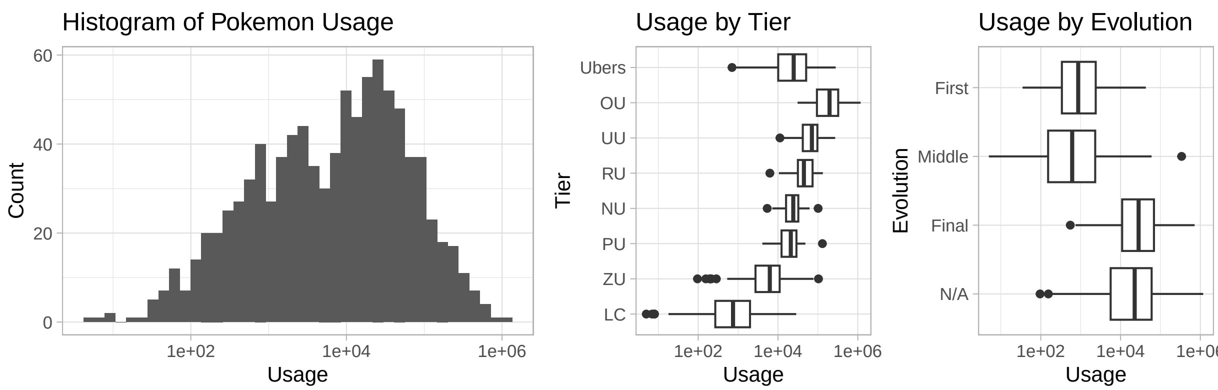

Figure 2. Histogram of Pokemon usage in the dataset independently (left), by tier (middle), and by evolution stage (right). The distribution of Pokemon is extremely right-skewed, with Pokemon in higher tiers being more commonly used. Non-evolving or final-evolution Pokemon are used more than not fully evolved Pokemon.

Table 4. Most and least commonly-used Pokemon in the dataset.

Name	Types	Total	Tier	Evolution	Usage
Landorus-Therian	Ground,Flying	600	OU	Non-Evolving	1177996
Greninja-Ash	Water,Dark	640	OU	Final Stage	717556
Ferrothorn	Grass,Steel	489	OU	Final Stage	684332
Heatran	Fire,Steel	600	OU	Non-Evolving	671964
Tapu Koko	Electric,Fairy	570	OU	Non-Evolving	592173
...	...	...	...	...	...
Skiploom	Grass,Flying	340	LC	Middle Stage	18
Luxio	Electric	363	LC	Middle Stage	8
Silcoon	Bug	205	LC	Middle Stage	8
Cascoon	Bug	205	LC	Middle Stage	7
Spewpa	Bug	213	LC	Middle Stage	5

Pokemon usage is also heavily dependent on tier and evolution: mid-to-high-level tiers like OU, UU, and RU being the most common tier by far and low-level tiers like LC and ZU being the least common; moreover, fully-evolved or non-evolving Pokemon are more popular than not fully-evolved Pokemon. Though tier and evolution do not directly affect a battle, the model may need to adjust for these confounders. The tier-usage trend also manifests in Table 4, which previews the dataset's extremes. The top 5 most common Pokemon are all fully-evolved or non-evolving dual-types in OU with high totals, while the top 5 least common Pokemon are monotypes with low totals in LC. This trend is further visualized in Figure A1/A3 in the report website, which show positive correlations between tier, evolution stage, total base statistics, and usage.

Base Statistics

Every base statistic appears to follow the same trend (Figure 3), being mostly distributed between 40 and 120 with more large outliers than small outliers. Plotted against usage, there is a noisy but somewhat nonlinear positive correlation that is roughly linear when the statistic is < 100, but flattens as the statistic increases, suggesting a GAM could model these variables well. Usage also correlates positively with total and standard deviation of all base statistics.

Fitting a linear model between logged usage and each base statistic, all slope parameters are significant ($p < 2 \cdot 10^{-16}$), with R^2 varying from 0.16 to 0.28, confirming the observed positive correlation.

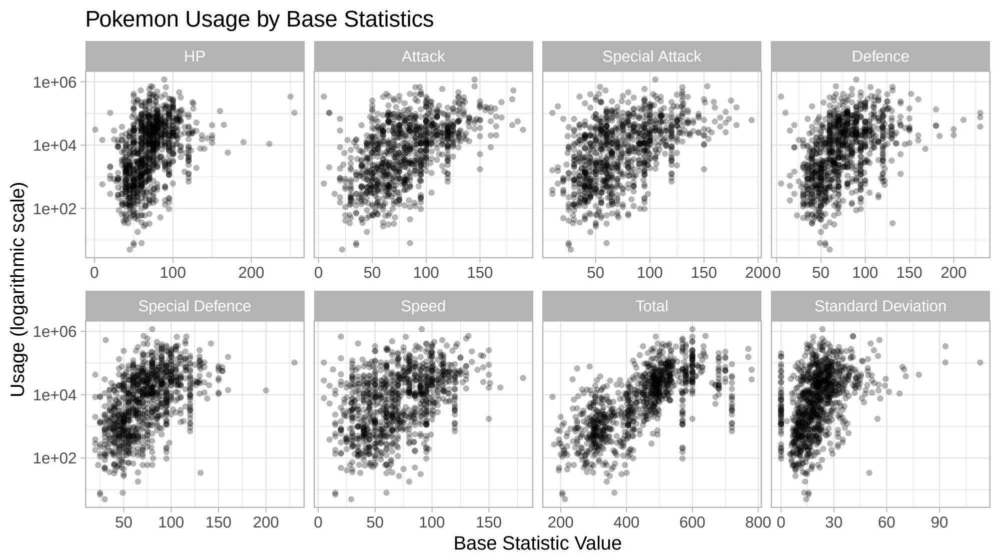

Figure 3. Scatter plots of a Pokemon's base statistics against usage. All follow a similar noisy but positive correlation.

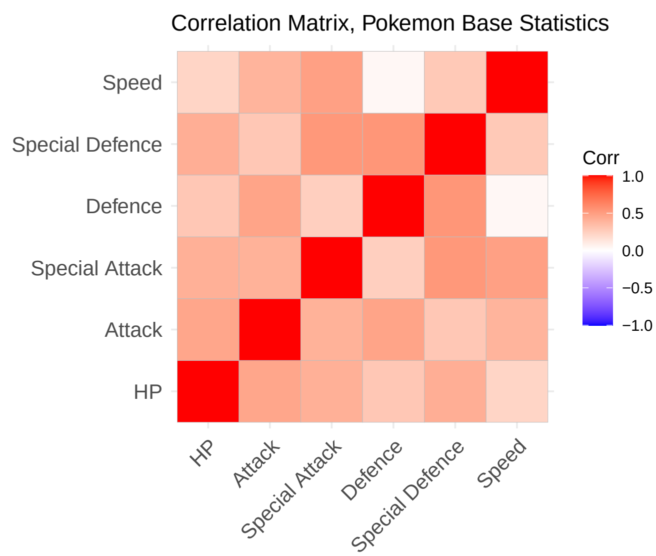

Figure 4. Correlation matrix of all Pokemon base statistics. All variables are correlated positively with each other, except for speed and defence which have no correlation.

A possible reason for the identical trends across base statistics could be that they are highly-correlated. Plotting a correlation matrix in Figure 4 confirms this hypothesis, which shows positive correlations between every value with each other (with the exception of speed to defence, with no relationship). This correlation is likely explained by the fact that evolutions almost always result in increases to base statistics, resulting in a distribution with positively-correlated base statistics. To reduce complexity and multicollinearity, using the total instead of individual base statistics may suffice.

Out of the 952 Pokemon in the dataset, 503 (52.8%) can be said to be attack-based (with 255 physical attackers, 169 special attackers, and 79 general attackers), 271 (28.5%) are defence-based (with 58 physical defenders, 42 special defenders, and 171 general defenders), and 178 are speed-based (18.7%). Analyzed against usage (Figure 5), it appears that primarily fast-moving Pokemon are the least likely to be used. Special-based Pokemon are more likely to be used than physical-based Pokemon, and physical/special defenders/attackers are more likely to be used than general defenders/attackers. This may be explained by there simply being more physical-based Pokemon to choose from.

The box-and-whisker plots are wide and the categories significantly overlap, raising the question of statistical significance. It was also discovered there were Pokemon with very well-rounded stats (e.g. Arceus, Sunkern) which this method would fail on, indicating this predictor needs additional engineering. However, an ANOVA fit between logged usage and `best_stat` reveals a p-value of 1.813×10^{-5}, confirming there is evidence for predictive power in this feature.

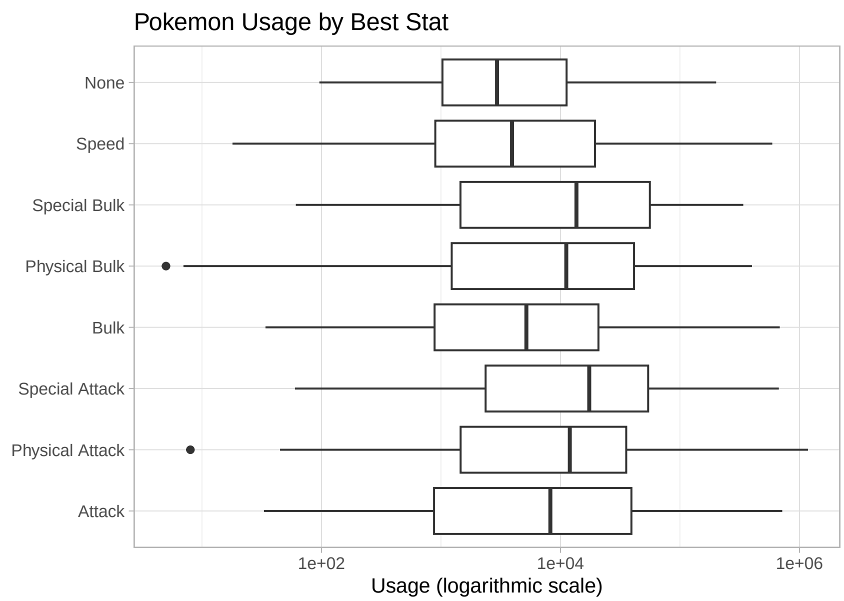

Figure 5. Box and whisker plots of a Pokemon's best stat type against usage. Special-based Pokemon are used most, followed by physical-based Pokemon, followed by general attackers/defenders, followed by speed-based Pokemon.

Typing

Figure 6 shows various 2D histograms and a box-whisker plot to assess the predictive power of a type-related features. It was hypothesized that Pokemon with less weaknesses, more resistances, and more super-effective coverage would be more useful competitively and thus used more often. However, apart from the number of resistances, there does not appear to be a strong correlation, if there exists one at all, between usage and coverage or between usage and number of weaknesses. Perhaps having many weaknesses and poor coverage is not problematic if a Pokemon is not primarily defensive in nature, while having many resistances is good regardless of the Pokemon's primary use cases. It may be necessary to combine `best_stat` with model weaknesses and coverage to properly model the hypothesized effect.

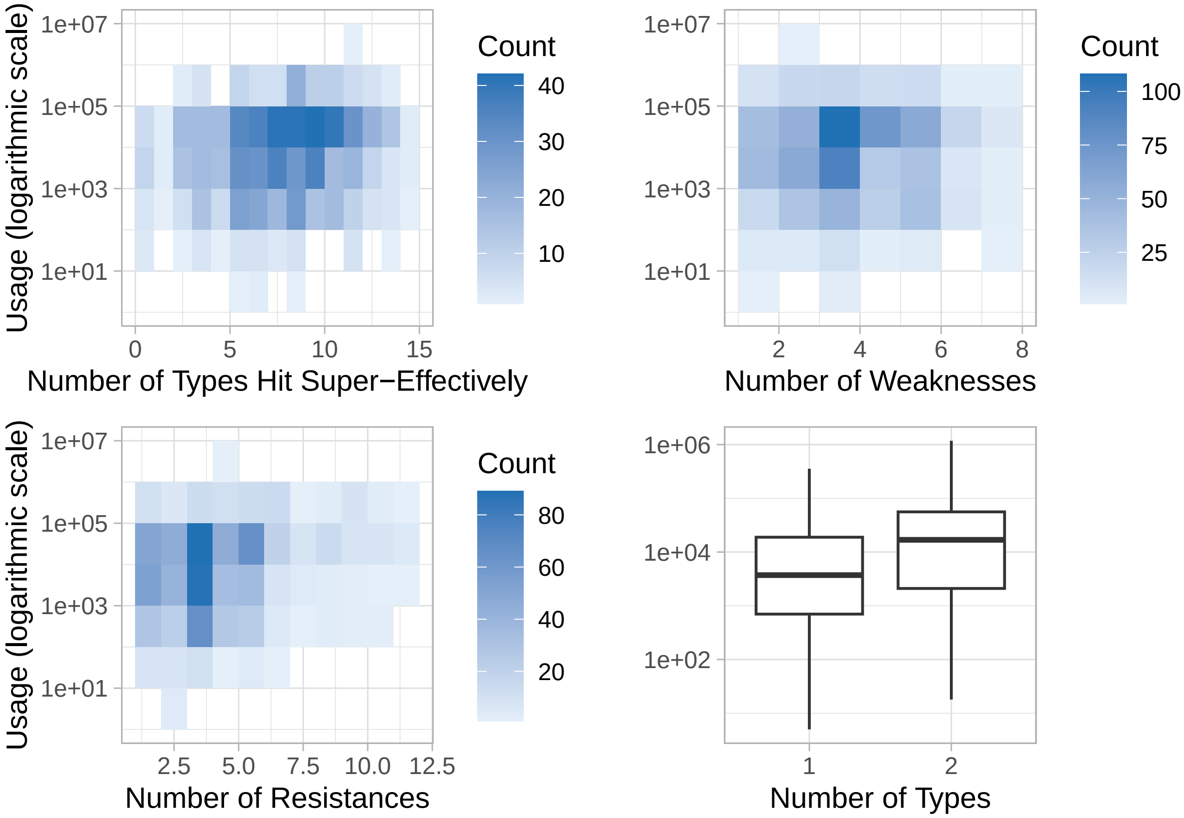

Figure 6. 2D histogram plots and box-plot of usage against type-related features, including the number of types a Pokemon's top 5 most common moves could hit super-effectively, plus a Pokemon's number of types, weaknesses, and resistances. Only the number of types and resistances shows a positive correlation with usage.

Another noticed trend is that having more types leads to increased usage. This is difficult to understand, but could be another byproduct of evolutionary mechanics, a confounder variable. Pokemon commonly gain an additional type as they evolve into stronger (and thus more-used) Pokemon.

An alternative and more in-depth method to analyzing the effect of Pokemon typing on usage is to analyze the effect for each type separately. Table 5's three ANOVA tests indicate that, to a 0.05 level,

- There are 10 types whose presence or absence significantly determine usage: normal, grass, fighting, flying psychic, bug, dragon, dark, steel, and fairy.
- Defensive matchups against 7 types significantly determine usage: normal, grass, electric, poison, ground, flying, psychic
- Offensive matchups against 11 types significantly determine usage, including all types except grass, fighting, poison, bug, rock, steel, and fairy.

Since many of the variables are not significant, it is probably better to use at most one of the three sets of variables in the final model to prevent overfitting. Furthermore, at this level of granular detail, it may also be difficult to explain why specific types are more important than others for determining usage.

Table 5. Significance tables for three separate ANOVA tests for logged usage against the presence of a type in a Pokemon (left), the defensive type multiplier against an attacking type (middle), and the offensive type multiplier against a defending type (right).

Feature	F value	Pr(>F)	Feature	F value	Pr(>F)	Feature	F value	Pr(>F)
is_normal	6.8186	0.0092	def_normal	7.2968	0.0001	atk_normal	13.6854	0.0002
is_fire	1.2489	0.2641	def_fire	2.9892	0.0182	atk_fire	6.7616	0.0095
is_water	0.2350	0.6279	def_water	1.0265	0.3925	atk_water	5.1385	0.0236
is_grass	11.6467	0.0007	def_grass	4.1519	0.0025	atk_grass	0.0642	0.8000
is_electric	0.2677	0.6050	def_electric	5.3555	0.0001	atk_electric	12.6827	0.0004
is_ice	0.2265	0.6342	def_ice	3.3982	0.0091	atk_ice	8.3388	0.0040
is_fighting	8.6138	0.0034	def_fighting	1.7823	0.1303	atk_fighting	1.8040	0.1796
is_poison	0.1099	0.7403	def_poison	2.6977	0.0198	atk_poison	0.1917	0.6616
is_ground	1.9756	0.1602	def_ground	3.8989	0.0017	atk_ground	3.6020	0.0580
is_flying	3.5543	0.0597	def_flying	8.0661	0.0000	atk_flying	3.6633	0.0559
is_psychic	4.0443	0.0446	def_psychic	4.9650	0.0002	atk_psychic	3.5742	0.0590
is_bug	10.9199	0.0010	def_bug	2.2655	0.0605	atk_bug	1.5702	0.2105
is_rock	0.0046	0.9457	def_rock	1.1881	0.3145	atk_rock	0.1691	0.6811
is_ghost	2.8441	0.0921	def_ghost	1.3774	0.2399	atk_ghost	3.1195	0.0777
is_dragon	9.7100	0.0019	def_dragon	0.7830	0.5035	atk_dragon	5.0074	0.0255
is_dark	15.1775	0.0001	def_dark	0.5365	0.6574	atk_dark	4.6385	0.0315
is_steel	49.5947	0.0000	def_steel	1.6312	0.1643	atk_steel	1.2351	0.2667
is_fairy	33.1631	0.0000	def_fairy	1.5313	0.1910	atk_fairy	0.0667	0.7962

Move Details

Finally, the specific mechanisms of moves in a Pokemon's common movepool are analyzed against usage using ANOVA. From Table 6, it appears that the presence of special moves, multi-turn setup moves, status-inflicting moves, hazard-related moves, switching-related moves, self-weakening moves, and multi-hitting moves significantly impacts usage. The signifance of having special moves appears to align with the previous observation of special-based Pokemon being more widely-used. These results suggest that a subset of these predictors can be used in the final model.

Table 6. ANOVA test for a Pokemon's usage against whether their common movepool contains moves with specific properties.

Feature	Df	Sum Sq	Mean Sq	F value	Pr(>F)
prop_physical > 0	1	10.7445	10.7445	2.1897	0.1393
prop_special > 0	1	36.7202	36.7202	7.4833	0.0063
prop_nondamaging > 0	1	2.2064	2.2064	0.4496	0.5027
num_inaccurate > 0	1	6.7457	6.7457	1.3747	0.2413
priority	1	0.4851	0.4851	0.0989	0.7533
stat_change	1	0.9653	0.9653	0.1967	0.6575
multisetup	1	36.1473	36.1473	7.3666	0.0068
heal	1	6.3018	6.3018	1.2843	0.2574
status	1	20.7928	20.7928	4.2374	0.0398
hazards	1	23.0007	23.0007	4.6874	0.0306
switch	1	27.0524	27.0524	5.5131	0.0191
recoil	1	1.1971	1.1971	0.2440	0.6215
ohko	1	0.5110	0.5110	0.1041	0.7470
selfweaken	1	78.9179	78.9179	16.0830	0.0001
multihit	1	27.0918	27.0918	5.5211	0.0190
protect	1	4.6874	4.6874	0.9553	0.3286
Residuals	919	4509.4640	4.9069	NA	NA

Linear Regression

Decision Trees

Summary

This analysis has hitherto demonstrated that a Pokemon's competitive usage is significantly affected by its base statistics, typing, and top 5 most common moves. Specifically, larger base statistics, dual-types, special-attacking Pokemon, and better defensive typings appear to correlate with increased competitive usage. This work suggests that tiers and Pokemon evolution are confounding variables, and that analysis by individual base stat does not lead to very clear and insightful conclusions. The relationship of typing and moveset to usage was found to be less straightforward compared to base statistics and may require further exploration with predictor interactions.

The scope of this work has included exploring the widest possible variety of variables and conducting tests of group of similar variables together. Future work will require simplifying the set of predictors to an efficient subset, testing for variable interactions, and finding the best way to present a feature (e.g. performing a shallower type analysis by aggregating counts of weaknesses/resistances or including a predictor per type). Furthermore, this work only conducted preliminary ANOVA tests and linear regressions using small subset of variables at a time. The final project will include a full linear regression model, as well as more complex models like GLMs, GAMs, and decision tree models.

Bibliography

The Pokémon Company. (2025). *2025 Pokémon World Championships*. Pokémon. Retrieved 16 March, 2025, from https://www.pokemon.com/us/play-pokemon/pokemon-events/championship-series/2025/world-championships

Smogon University. (n.d.) *An Introduction to Smogon's Tier System*. Retrieved 16 March, 2025, from https://www.smogon.com/bw/articles/bw_tiers