**Pokemon Data Analysis**

Tony Schmitt Payton Beans

[aschmi5@clemson.edu](mailto:aschmi5@clemson.edu) [pbeans@clemson.edu](mailto:pbeans@clemson.edu)

C36755261 C21285765

<https://github.com/Tschmitt99/datavizgroup16>

**Background and Motivation**

Pokemon is a game that focuses on the collection, training and battling of powerful creatures called Pokemon in a turn based team battle. Currently throughout the pokemon franchise there are 898 unique species of pokemon. In the game each Pokemon has a collection of stats and attribute data that play into the game’s battle mechanics such as a single or double typing and a set of numerical stats.

We chose our data set because Pokemon is a complex game that was a part of many people’s childhoods. When we were young we didn’t explore the data behind Pokemon much at all and most choices in the game made by kids are more focused on what Pokemon we thought were interesting at the time. Now we would like to go back and find a better understanding of the game and it’s wide array of creatures through data analysis.

**Project Objectives**

Our goal with this project is to use the data gathered about each Pokemon to create representations and assumptions about their stats. Conclusions can be made through analyzing trends and relationships between various Pokemon based on multiple different metrics and data points. Through the various core stats, their moves, evolutions, and types we can create benefits for players and those that are interested in exploring how Pokemon are created. Our goal is for current and potential Pokemon players to be able to potentially use our data analyses to perform at a higher level while playing and to also be able to create a sort of objective metric for which Pokemon are comparatively “better” than others of similar caliber and tier.

**Data**

<https://data.world/steveinatx/pokemon-index/workspace/file?filename=pokemon.xlsx>

The source for our data comes from an international enterprise data catalog Data.World that allows people from across the world to upload datasets. Our data consists of the following 4 sections that we will be able to digest in various forms.

*Pokemon:*

This section of the data consists of all 898 unique species of Pokemon from across the generations of the games. This includes their name, their elemental type (one of 18 possible types), their health points or HP, attack stat value, defense stat value, speed stat value, as well as their special attack and special defense stat values.

*Moves:*

This section consists of all of the possible moves that Pokemon can perform. There are 608 total moves. Each value includes the move’s name; its elemental type (much like the Pokemon themselves); its category (physical, special, or status); its power stat value; its accuracy from 1-100; its power points of PP (which is a numerical value representing how many times in a battle that a Pokemon can use that move); whether or not a technical machine or TM is required to gain access to the move; its ‘effect’ which is a description of what the move does; and the probability percentage that the effect it has (if applicable) hits or not.

*Evolution***:**

This portion of the data includes which Pokemon evolve and what they can evolve into. It includes the names of both the evolving Pokemon and its evolution name, the level that it needs before it can evolve, any special required conditions for the evolution to occur, and the evolution type (level, stone, trade, happiness, other).

**TypeChart**

The last section of the data is a very important part of the Pokemon battle mechanics; the TypeChart. In the Pokemon games, certain elements or “types” are considered to be more effective against others, which then provides multipliers to move stats when used in-game if the appropriate types are in play. This part of the data consists of the attacking Pokemon’s type, the defending Pokemon’s type, its effectiveness against the defender (normal, not very effective, no effect, or super effective), and the stat multiplier.

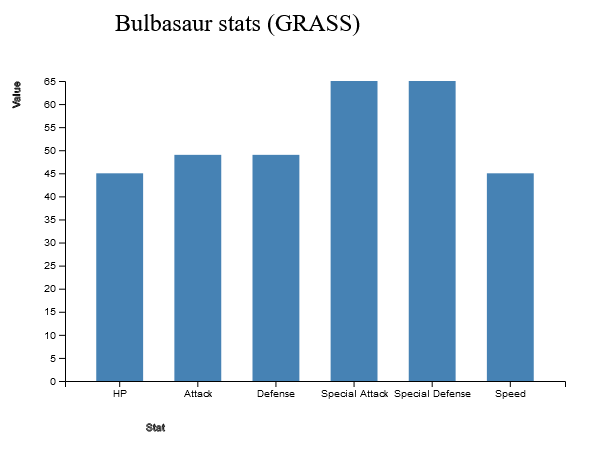
**Data processing**

In order to get our data into a consistent format and address an outlying information in the data set we did have to do some processing work. The pokemon games are released in generations and currently there are 8 unique generations of the game that each introduce new pokemon and game mechanics. In order to keep our data more consistent across generations we will remove Pokemon that only exist through the use of a game mechanic in later generations. An example of this is Mega Evolutions, These are upgraded versions of existing pokemon and, while they have unique stats and typings compared to the non Mega Evolved counterparts, not all Pokemon have a Mega Evolution.

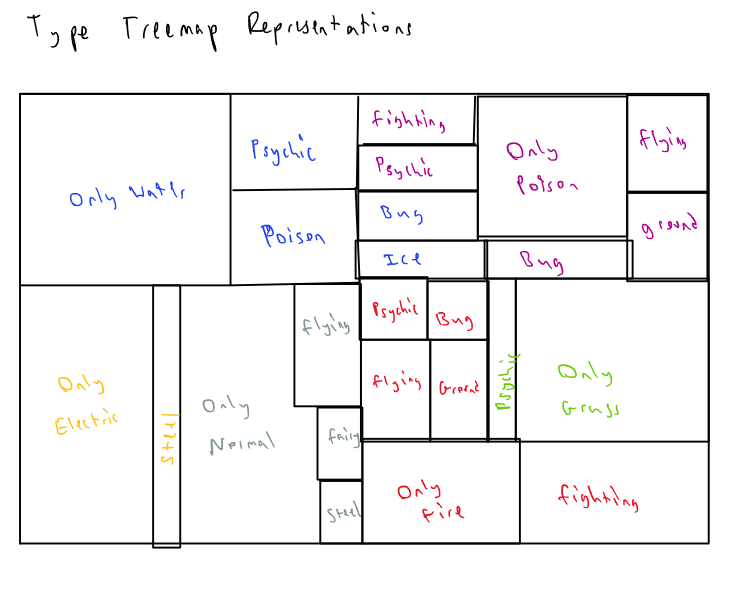
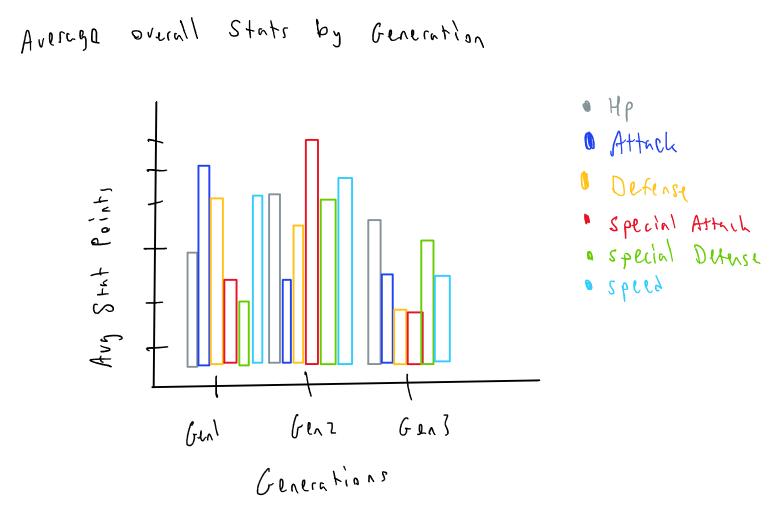
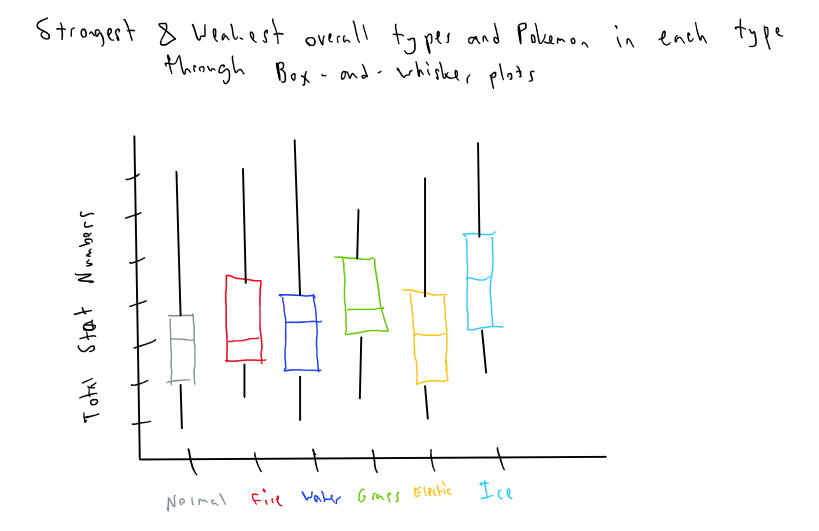
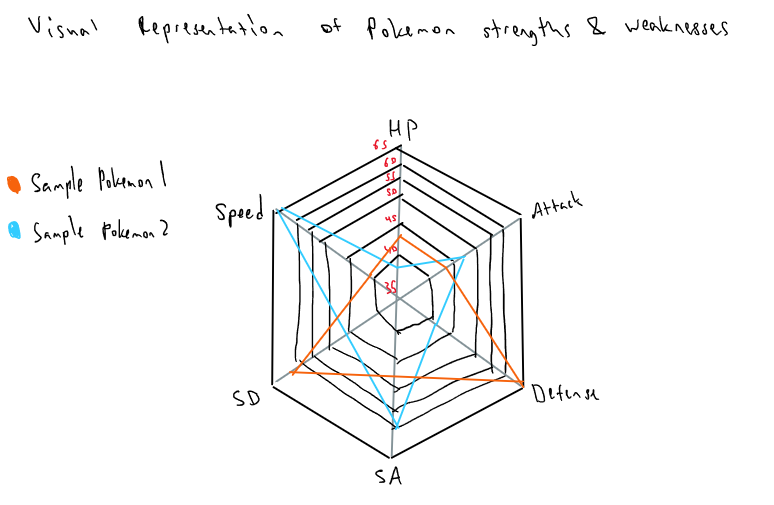
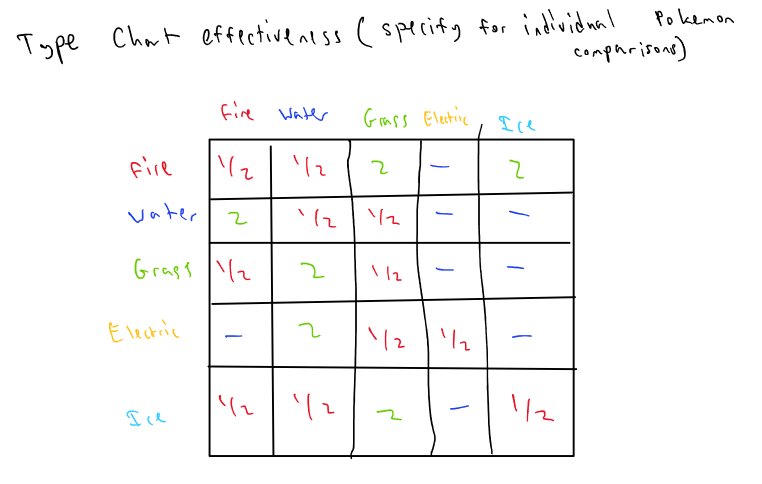
We’ve also had to standardize the stats that each pokemon has. In the first generation of Pokemon each Pokemon had only one stat that contributed to attack or defence appropriately named “attack” and “defense”, later generations divided each of these stats into a normal and special attack/defense stat. The first generation used the more simple stat spread for both cases of normal/special. To standardize the data we have added a special attack and special defence to each of the Pokemon from the first generation that is just a duplicate value of the original non-special stat. This wouldn’t change any aspect of the game play and creates a more uniform data set for analysis.

**Visualization Design**

Bulbasaur solo stats prototype



Below are some screenshots of concepts that we would like to develop in similar ways



**Must have features**

* A comparison of pokemon stat totals over each generation to identify power creep in the game
* An analysis of the most common and most unique typings of Pokemon
* A ranking of the strongest Pokemon in each stat category
* A graph that places Pokemon based a relation of their defensive and health stats
* A graph that places Pokemon based a relation of their attack and speed stats

**Optional Features**

* A selection box to bring up statistics on individual Pokemon
* Comparing multiple user-specified Pokemon to each other

**Project Schedule**

Oct. 22 - Complete project proposal

Oct. 25 - Complete data cleaning

Nov. 1 - Drafts of visuals in D3

Nov. 7 - Complete prototype

Nov. 13 - Polish for peer evals

Nov. 20 - First final draft

Nov. 26 - Revised draft

Nov. 30 - Fixes for oral presentation

Dec. 5 - Final delivery