WHO'S THAT POKÉMON?

Team Members: Jennifer Nelson (u1060401) & Matthew Christensen (u0899251)
Email: jennymarienelson@gmail.com   u0899251@utah.edu
Project Repository: https://github.com/jenniNelson/dataviscourse-pr-christennelsenson

Overview and Motivation

With a new generation of Pokémon just around the corner, trainers around the world are getting ready to start their adventures again. From the outside, Pokémon seems like a basic rpg game series. However, it should be noted that lifelong fans often take these games **very** seriously. Depending on who is approaching the game, they may want to fly by the seat of their pants or plan every decision out in advance meticulously.

While there is no shortage of tools available online to help prospective trainers learn about the various Pokémon they might encounter on their journeys, it isn't always obvious how that Pokémon compares to others. Furthermore, locations for each creature are usually presented as a set of numbered locations. For someone who isn't as familiar with every game, it can be difficult to visualize where each Pokémon is on the map. On top of all of that, most databases don't make it obvious if an individual Pokémon can even be obtained in a certain game.

All of these factors and more can make it difficult to plan out a team of 6 Pokémon to take on a journey through a game. We are creating a tool that will bring together information from several large community databases for the purpose of helping new or veteran trainers learn about and locate Pokémon to bring to their team.

Related Visualizations

In compiling data from our two primary web sources, Serebii.net and bulbapedia.net, we've drawn inspiration from the map view on serebii, and the evolution trees that both sites use. The map view on serebii is a large map containing regions from five generations of Pokémon games, where every in-game section is a clickable link. Though our regions aren't clickable, our tool highlights the regions a Pokémon can be found in.

Evolution trees are a common way to visualize what other Pokémon a particular Pokémon can evolve from/into. Most are straightforward chains, but some can evolve in different ways, depending on in-game conditions. Eevee is a notorious example, with eight possible evolutions. Our tool is capable of displaying all evolutions of Pokémon, including the unique cases. This can make it easier to locate Pokémon that have previous evolutions that can be found in a particular region.
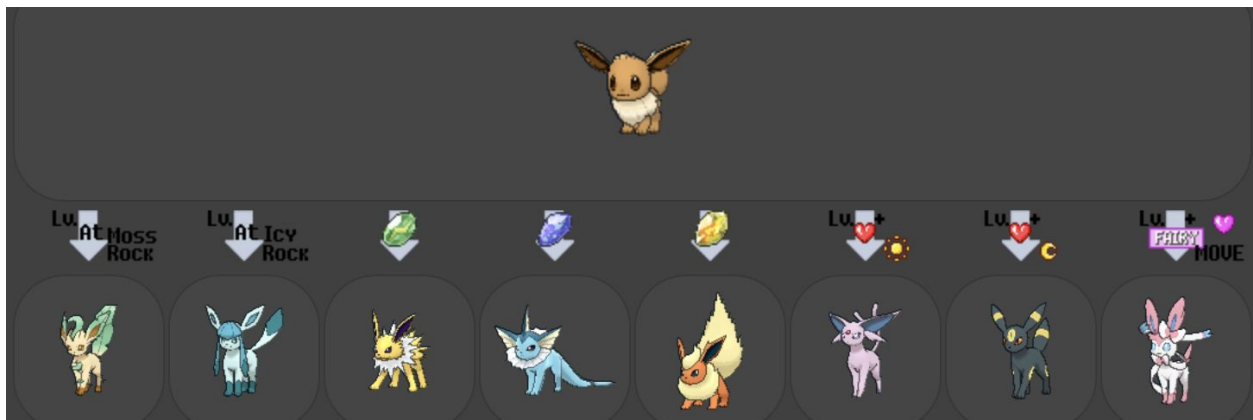
Project Objectives

The primary objective of our tool is to help users learn practical information about Pokémon. Specifically, we are creating a tool that will make locating Pokémon in each game as quick and intuitive as possible. In addition, our tool will assist users in the team building process (e.g. providing basic stat and type advantage information for a team of 6 Pokémon that the user suggests). As well as locating specific Pokémon, our tool will also provide information on which ones are available in any given location on each of the games' maps.

Some example questions that the user will be able to answer:
- Dusclops is my favorite! Is it possible to get one in Generation VII?
- I have 6 Pokémon already for my team. Are there any types that I should watch out for when battling?
- What Pokémon are available near the start of the game in Generation IV?
- How would Seadra fare against Scizor?
- How would *the worst* Pokémon do against *the best* Pokémon?
- I'm really struggling with the electric type gym in Lumiose City. Are there any ground types nearby I can use for help?

Data

Pokémon is an incredibly popular video game franchise, and numerous data sources are publicly available. Our dataset is based off "The Complete Pokemon Dataset" found on Kaggle, providing the basic stats for each pokemon. In addition to the stats, we have also scraped Pokémon sprites, evolution trees, some fixed stats, and Pokémon locations from Serebii. We've pulled map images from Bulbapedia.

The Complete Pokemon Dataset

This is our base dataset. Out of the box, it contains names, ids, stats, types, height, weight, as well as a few other less important facts about 801 Pokémon species. We stored this dataset in /data/pokemon_data/pokemon.csv in the project. While working with the dataset, Matt noticed some errors in the stats for certain Pokémon. He wrote a web-scraping script in Python to collect the proper stats for every Pokémon. We use these corrected stats in our final implementation. The corrected stats are contained in /data/pokemon_data/stats.csv in the project.

Evolution Trees

Most Pokémon are part of an evolution family that involves 1 or 2 other Pokémon. Some families contain as many as 8, but these are in the vast minority. Our base dataset does not

include this information, so we also wrote a web-scraping tool to help us collect the information. Our scraper was able to collect information on most evolution families without an issue, but there are some families that aren't straightforward. For instance, some Pokémon (like Eevee) can evolve along 2 or more different paths. After scraping the data, we manually had to clean the individual entries of Pokémon with non-standard evolution families. Luckily, the number of such cases is relatively small.

Sprites

Serebii uses a standardized path for all its Pokémon sprites: "serebii.net/{game}/{pokedex_num}.png". This made it easy to programmatically download all sprites, from Pokédex #001 to Pokédex #809. We chose to only use sprites from generation seven, for consistency. To represent non-selected Pokémon, we use a cropped image from the "Who's That Pokémon?" segment (surrounding commercials for the Pokémon anime). We keep all of the sprites in /data/pokemon_data/sprites

Locations

This was the most involved scraping process. Jenni found a URL that contained location information for a given Pokémon for a given generation. The problem was figuring out how to format the data from that page. Not only will a Pokémon have different locations in different generations, but it also may have different locations within the *same* generation depending on which game is chosen. To account for this, we wrote a scraper that could analyze every generation page for every Pokémon. It organized every location in the following syntax: game/region/location.
The locations we scraped are kept in /data/pokemon_data/stats.csv

Scraping Methods

Our scraping tools were written in python (the files are included in the repo). We used several libraries to assist us: primarily BeautifulSoup4, urllib, and wget. We used the csv library to read and write to all of our csv files.

Sample URLS:
Evolution and Stats: https://www.serebii.net/pokedex-sm/013.shtml
Location: https://www.serebii.net/pokedex-sm/location/013.shtml
Sprites: https://www.serebii.net/sunmoon/pokemon/013.png

Map Data

We use map images pulled from Bulbapedia, preferring in-game maps over artistic renderings. To interactively color in-game regions, Jenni manually drew regions using an Online SVG Editor, then labeled using reference images (example). This took an extremely long time. Many Pokémon can be found in areas not annotated on the map, especially in the Kanto map, because the region appears in several different games, with slight variations. We chose to still list these locations for completeness, even if they aren't findable on the map.

Exploratory Data Analysis

Using python (because it's quick and easy to use), we wrote up an implementation that will measure the "distance" between two Pokémon. Here, distance refers to the Euclidean distance between two different Pokémon's stat vectors. Fortunately, since the dataset of all Pokémon is on the small side, it does not take long to calculate the distances between every combination of 2 Pokémon and summarize the results. We've noticed some interesting things so far.

First, there is a small subset of Pokémon that are extremely distant from all of the others. Usually, these are Pokémon with extremely outsized stats (e.g. the Pokémon Shuckle (#213) has the highest defense stats, but all of its other stats are the lowest).

On a whim, we decided to see which Pokémon was the most "average"; i.e. which Pokémon has the shortest average distance to all others. It turns out that Castform (#351) fits the bill by a decent margin. It's stats are exactly 70 across the board. If one were imagining all Pokémon as points in 6-D space, it may be helpful to imagine Castform as existing at the origin, and all others being based off of that.

Both of these informed us when we were making our bar charts to include some information to help our users. Since the range of stat values is so wide, but most Pokémon have comparatively low stats, we decided to add a dotted line representing the "average" stat value across the board. This way it's a little easier to see if an individual stat is larger or smaller than one might expect.

We also took time to create a full table of all of the Pokémon with their stats. When we took this table to a TA for our midpoint, he brought up that a flat table is quite boring. This inspired us to add a bit of color and visual bars to each row. Also, the sheer size of the table served as impetus to add filtering tools to help users narrow down their search.

Design Evolution

Original Concept
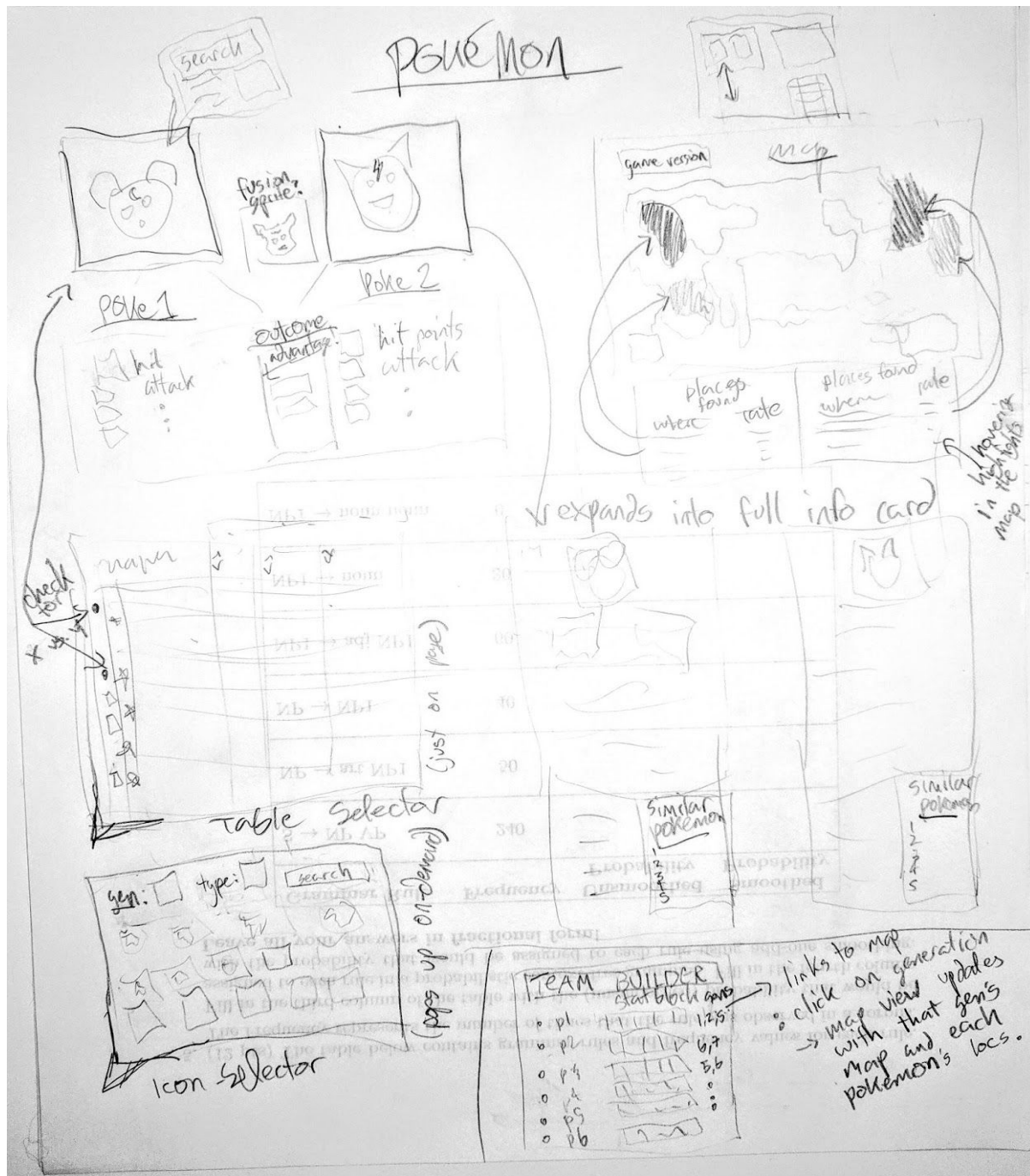
After deciding we wanted to do Pokémon, we brainstormed three different visualizations we wanted. First we thought of a 'Vs' view, comparing two Pokémont to each other, battle-style.

Then we thought of a map view, showing what Pokémon could be found where, per game. Then we came up with a rough idea of a "team building" mode. Though first just a small card, we later expanded on this view.
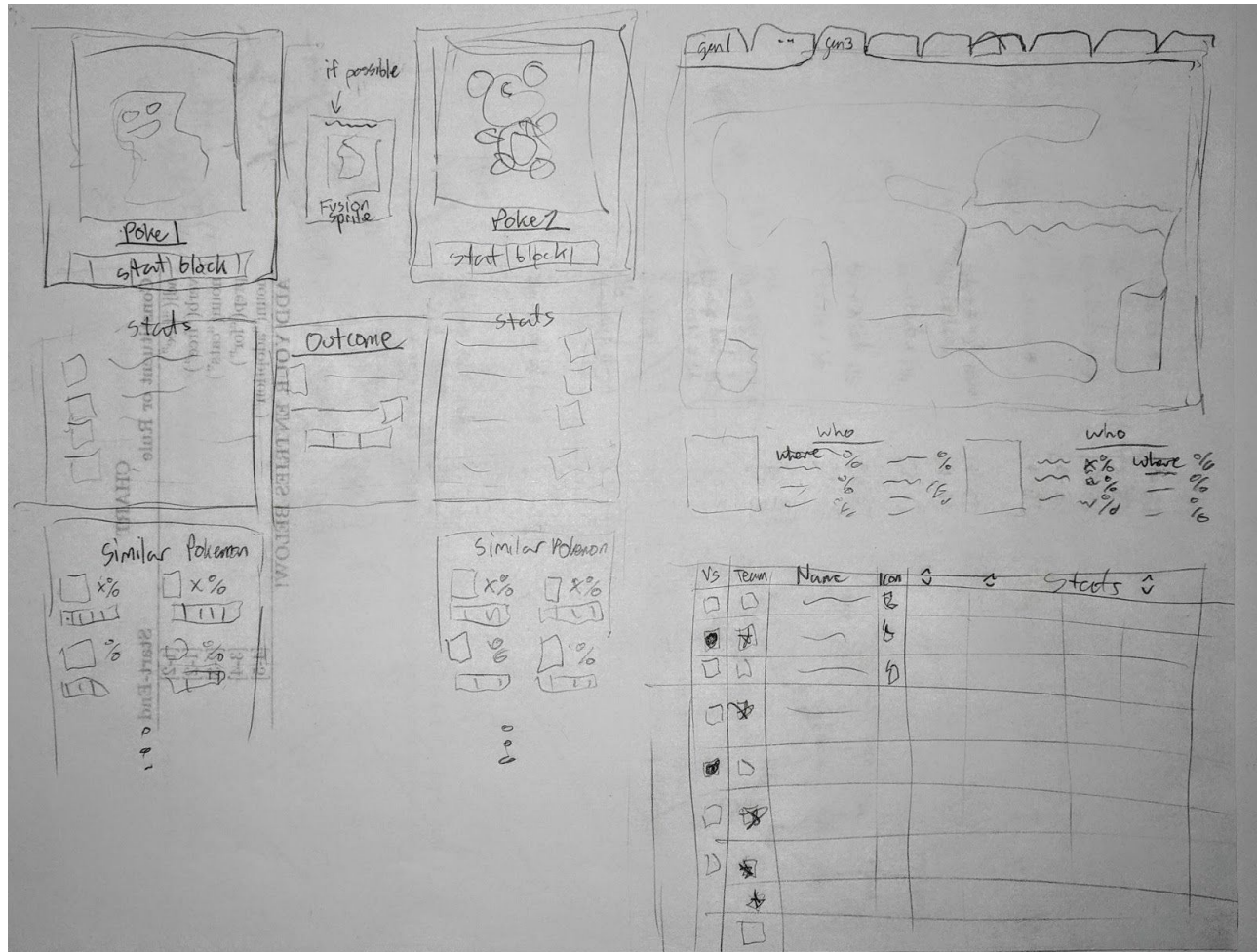
Second Iteration of the 'Vs' tab. The cards for the two selected Pokémon are still in the top left, but we decided to provide additional information for each of the Pokémon below their cards, including the matchup information and nearest stat 'neighbors'. We also moved the full interactive list to the bottom right beneath the Map panel. The tabs at the top of the Map panel will switch between the regions for each of the 7 generations.
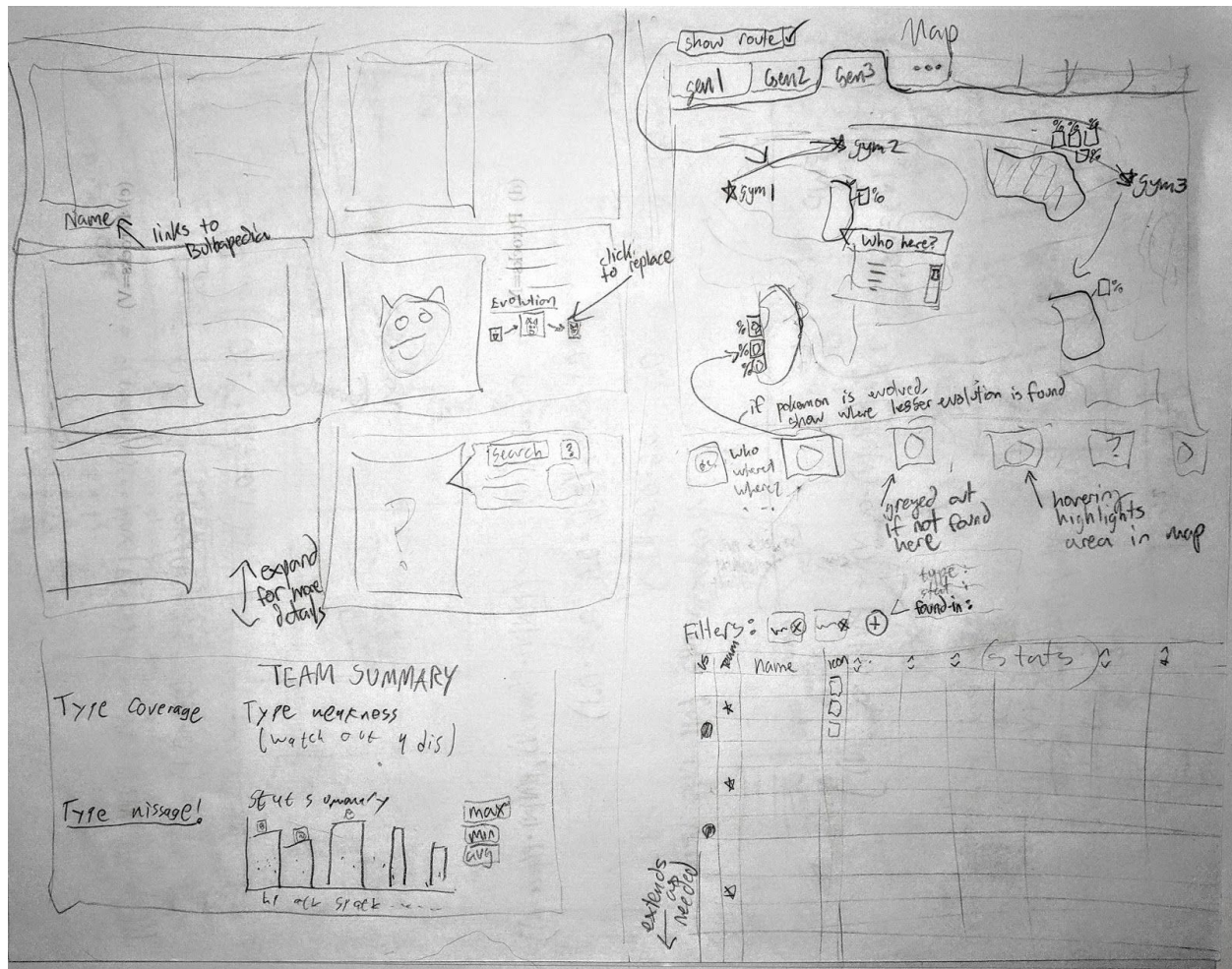
A draft of our Team Builder view. In the top left, there are 6 cards for each team member that the user can fill in. When the user clicks on a card, a dialog will pop up allowing them to search for their Pokémon by name. Alternatively, they can search through the interactive list that also appears in this view on the bottom right. By clicking a checkbox next to each Pokémon's row, that Pokémon will be added to the team builder. In the bottom-left, there is a Team Summary that will be updated every time a Pokémon is added to the Team Builder cards. The Summary will contain basic information that the user will need to evaluate their team. In the top right Map panel, the locations of all 6 Pokémon will be marked on the map with their corresponding icon.

At this time we also theorized about doing a PokéWeb visualization, linking sprites of every Pokémon to their five 'most-similar' neighbors, where similarity is defined by the euclidean distance between their stats.

Core Visualizations: Initial Idea

- 'Vs' View
    - Two side-by-side cards show two Pokémon sprites/names/Pokédex index.

- Expanded stats blocks below each, with an "Outcome" block between.
    - "Outcome" block estimates the probable results of a battle between the two Pokémon.
    - "Top Five Similar Pokémon" cards below each stats block. Similarity is defined by the d2 distance of two Pokémons' stat vectors.
- Team Builder View
    - Six cards in a grid, each describing a Pokémon and its basic stats.
    - A "Team Analysis" block below the cards
        - Type Coverage (i.e. the set of types this team is effective against)
        - Type Weakness (i.e. the types which 2 or more Pokémon on the team are weak to)
        - Type Mis-Coverage (i.e. types for which no Pokémon on the team is effective against)
        - Stat Summary (bar chart summarizing the best Pokémon for each individual stat)
- Map View
    - Maps from each generation's game, selectable via tabs
    - Maps are "the best we can find", hand annotated with regions
    - Two ('Vs' view) or six (Team Builder view) Pokémon sprites below the map, each with a list of the locations they appear in, and spawn rate per location.
        - Hovering over a sprite highlights those areas on the map
- Pokédex table
    - All Pokémon, sortable by name or stat
    - Column to select which two Pokémon are included in 'Vs' view
    - Column to select which six Pokémon are included in Team Builder view

Optional Features: Initial Ideas

- Animated sprites/icons (for generations with animated sprites/icons)
- Gen 8 (Comes out Nov. 15th)
- "Right-click-to-search" feature for selecting a Pokémon for any card
- Bonus 'Vs. Mode' Features
    - "Fusion" sprite view, with images from a generator.
    - Show evolution tree. Other evolutions of that Pokémon are clickable to switch Poke1 (or Poke2) to that Pokémon.
- Bonus Team Builder Features
    - Cards can expand to include advanced stats and the five "most similar" Pokémon
    - Clickable evolution tree (as in Vs. mode)
- Bonus Map Features
    - Map annotations for typical path, gyms, etc.
    - Annotate which Pokémon are obtained in special ways
    - Click a region to show what Pokémon can be obtained in that region

- Small icons for the two or six selected Pokémon appear near/in/on all regions that Pokémon appears in.
    - The Pokémon Spiderweb (Pokéweb)
        - A Network view of all Pokémon, linking each Pokémon to its top 5 (or other number) "most similar" other Pokémon. Similarity is defined by the euclidean distance of two Pokémons' stat vectors.

Dreams Meet Reality

Our core visualization is complete, and has not significantly changed from our original vision. The further we got, however, the more we realized that even our basic plan was ambitious. As a result, some of the optional features didn't make the cut. In data collection, we stuck to just one sprite per Pokémon, rather than animated sprites specific to each generation. The 8th generation did not have all the data we needed online in time for the project deadline (though it is a great game). Map annotations beyond region highlighting would make an already tedious task unbearably so. It's also not straightforward to programmatically obtain data about specially-obtained Pokémon (such as those that can only be obtained from certain NPCs).

The "right-click-to-search" feature has been implemented as a searchable drop-down menu. Evolution data is recorded on the card view, and enables quicker navigation up and down evolution trees.
All views are synchronized, so changes in one view affect all of the others.

Schedule Goal

Our project mostly went according to plan. We weren't able to complete a PokéWeb view, but all of our required and core functionality is in place. We were also able to implement some of our bonus features, such as clickable evolution trees.

	Due Date	Component	Team Member
✔	Oct-25-2019	Project Proposal	Jenni/Matt
✔	Nov-1-2019	Get Stat Data	Jenni/Matt
✔		Get Map/Sprite assets	Jenni
✔	Nov-8-2019	Functional Prototype	Jenni/Matt
✔		Get Evolution Tree Data	Matt
✔		Get Location Data	Matt
✔		Generate Stat neighbor Data	Matt
✔	Nov-15-2019	Card Views Complete	Matt

✔		Map Annotations Complete	Jenni
✔		Map Functional	Jenni
✔	Nov-18-2019	Interactivity Working	Jenni/Matt
		PokéWeb prototype: decide if continuing	Jenni/Matt
✔	Nov-22-2019	Feature Complete	Jenni/Matt
✔		Website Running	Jenni/Matt
✔	Nov-25-2019	Beautification	Jenni/Matt
✔	Nov-26-2019	Screencast Recording	Matt's sweet buttery Voice / Jenni's bullseye mouse maneuvering
✔	Nov-27-2019	Final Project Due	Jenni/Matt

Implementation

Team Builder/Map

The Team Builder is now the first view displayed on the screen. To the left, there is a table of 6 cards that enables the user to select (using searchable dropdown tables) a team of 6 Pokémon. Their selection is reflected in the smaller cards on the right side beneath the map. In addition to displaying each Pokémon's stats in the card view, the locations where that Pokémon can be found are displayed on the right. When the user hovers over the location list, those locations will highlight themselves on the map above, making it easier to navigate exactly to where the Pokémon is.

Since there are many different generations of Pokémon, each with a different map, it is possible to switch map views between each generation. When switching views, the locations next to each Pokémon will update to their corresponding locations in the new map.

In addition to searching through the dropdown, each Pokémon card on the left has information on its evolutions. By clicking the little icon, the view will update so that the evolution's stats and locations are displayed. Regardless of the method used, changing the team of 6 in the card view will also update the other views on screen.

Team Summary
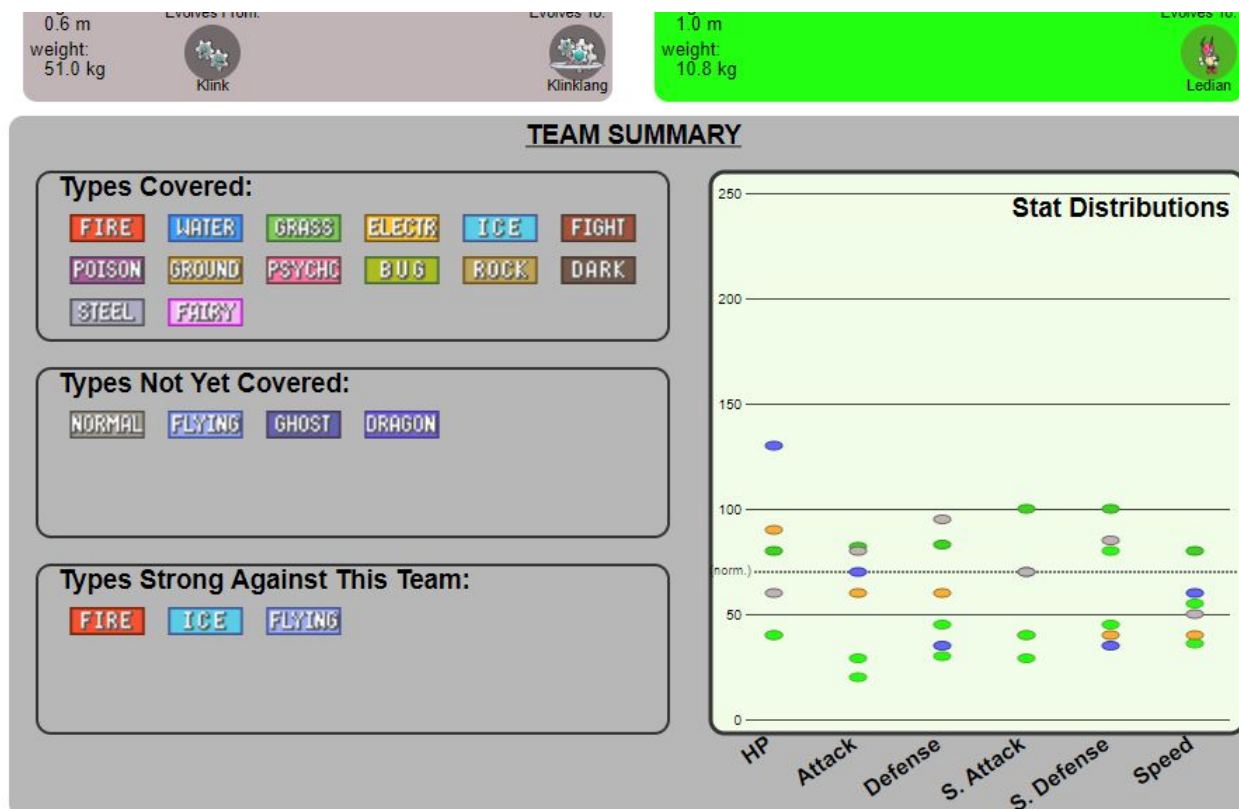
On the left beneath the 6 cards, there is now a summary view of the selected team. This view contains valuable information for what it would be like to use the team. To the left, there is a summary of type coverage. Type coverage refers to the ability of a given Pokémon to deal with opponents of other types. For instance, this particular team does not yet have a member who is

Effective against Normal, Flying, Ghost, or Dragon type Pokémon. "Types that are strong against this team" refers to types that many members of the team are weak to, and thus should be avoided.

To the right, there is a summarized version of the Team's stats. If the user hovers over any point on the chart, a small tooltip will pop up with that Pokémon's name, and stat value for that category. In addition, lines will connect all of the stats associated with that particular member, making it easy to see their contribution. At a glance, it is easy to see which team member is the best in each stat category, and which categories the team is weaker in.

Pokédex Table

| Vs	TB	#..	Name	Bas...	HP	Atk	Def	S.Atk	S.Def	Spd	T1	T2	G..
☐ ☐	☐☐☐ ☐☐☐	798	Kartana	570	59	181	131	59	31	109	GRASS	STEEL	7
☐ ☐	☐☐☐ ☐☐☐	409	Rampardos	495	97	165	60	65	50	58	ROCK		4
☑ ☐	☐☐☐ ☐☐☐	289	Slaking	670	150	160	100	95	65	100	NORMAL		3
☐ ☐	☐☐☐ ☐☐☐	486	Regigigas	670	110	160	110	80	110	100	NORMAL		4
☐ ☐	☐☐☐ ☐☐☐	383	Groudon	670	100	150	140	100	90	90	GROUND		3
☐ ☐	☐☐☐ ☐☐☐	384	Rayquaza	680	105	150	90	150	90	95	DRAGON	FLYING	3
☐ ☐	☐☐☐ ☐☐☐	386	Deoxys	600	50	150	50	150	50	150	PSYCHO		3
☐ ☐	☐☐☐ ☐☐☐	644	Zekrom	680	100	150	120	120	100	90	DRAGON	ELECTR	5
☐ ☐	☐☐☐ ☐☐☐	612	Haxorus	540	76	147	90	60	70	97	DRAGON		5
☐ ☐	☐☐☐ ☐☐☐	464	Rhyperior	535	115	140	130	55	55	40	GROUND	ROCK	4
☐ ☐	☐☐☐ ☐☐☐	534	Conkeldurr	505	105	140	95	55	65	45	FIGHT		5
☐ ☐	☐☐☐ ☐☐☐	555	Darmanitan	480	105	140	55	30	55	95	FIRE	FIRE	5
☐ ☐	☐☐☐ ☐☐☐	567	Archeops	567	75	140	65	112	65	110	ROCK	FLYING	5

The Pokédex Table view is a complete table of all Pokémon. We've upgraded the table significantly in many ways. For starters, wherever possible we've augmented stat numbers with bars allowing for quick comparison between Pokémon on the list. In addition, the list is sortable by any criteria.

Of particular note are the filter buttons at the top. Using these buttons, it is possible to filter down the list by any categorical criteria, such as type, generation, evolution status, and even whether or not each Pokémon is "Legendary." A last minute addition is the ability to filter the list

by numerical data as well. For instance, a user can determine what Pokémon have a speed stat greater than 100.

On the left side of the table, there is a series of check boxes. These check boxes correspond to both the Vs and Team Builder selections, making it one more way that a user can make the selection for their team. The purpose of this table is to provide a more systematic search tool for users, and the filters and sorting allow for just that.

Vs View
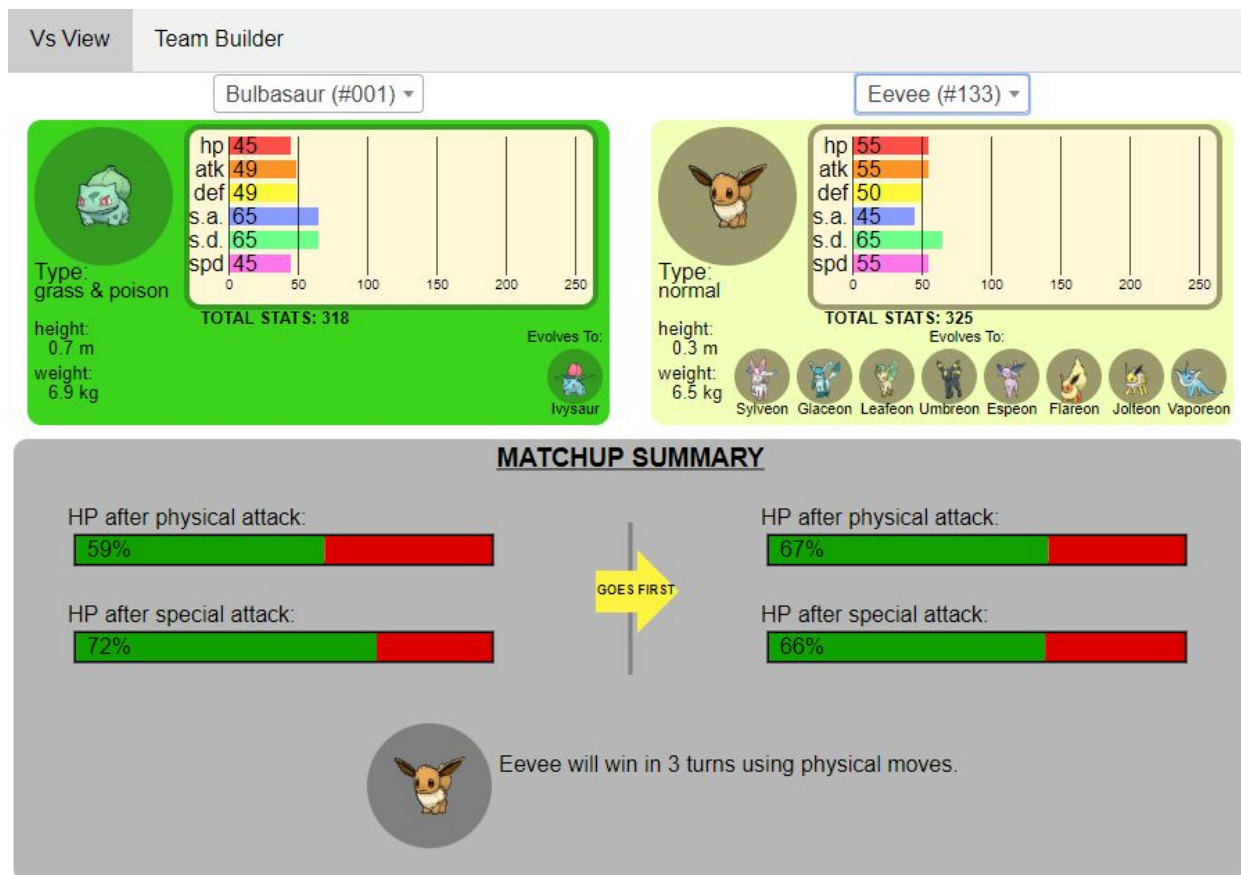
The Vs tab on the top left will bring the user to a Vs view. The purpose of this view is to give the user a good idea of how a simulated battle would play out between two Pokémon. There is a lot of complicated math behind how much damage Pokémon will do to each other, so we wanted to make sure that users wouldn't have to do anything complicated.

The winner in this simulated battle is whichever one would last longer in a 1 on 1 battle (where speed is the tiebreaker.

The card view makes a return in this view as well, because it makes it so easy to navigate between different Pokémon. Of course, users can also select these using the table view.

Evaluation

We're Done! This visualization took a long time, and a lot of effort, but we're really pleased with what we've made.

What we learned:
- Pokémon is a slightly messy dataset. Each new generation introduces some new exception or set of exceptions that need to be accounted for.
- Manual map annotation is extremely time consuming.
- It's not easy to get a searchable drop-table in html, but libraries were our friend.
- Data scraping is fun!

Summary:
Our visualization works very well, and we've stamped out as many bugs as we could find. Selecting Pokémon is quite easy, and the view shifts quickly whenever changes are detected.

I feel the graphs, summaries, and maps answer all of the questions that we were asking quite thoroughly, and make for a fun and informative tool to explore.