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.

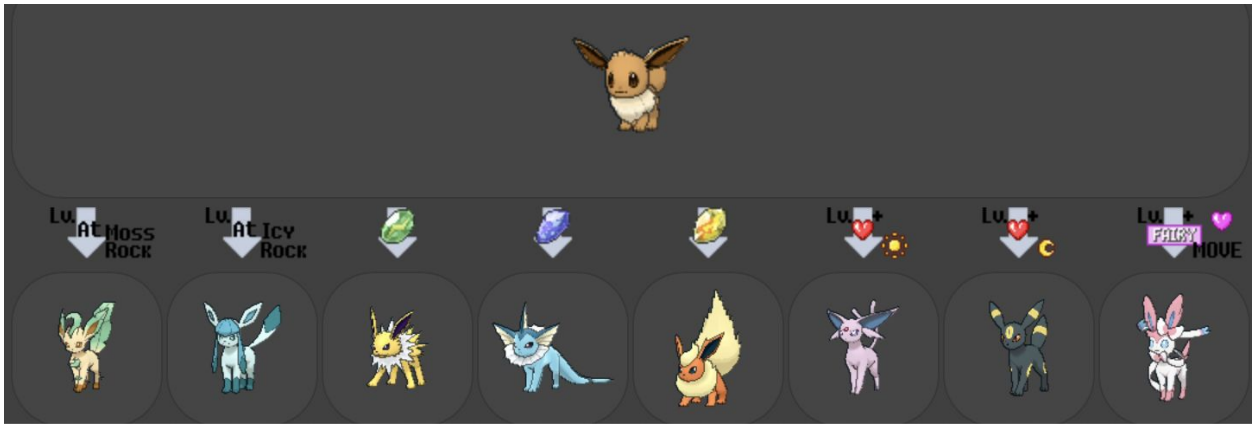
While there is no shortage of tools available online to help prospective trainers learn about the various Pokémon they might encounter on their journeys, it isn't always obvious how that Pokémon compares to others. Furthermore, locations for each creature are usually presented as a set of numbered locations. 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 we don't plan on having our regions be clickable, we do plan on highlighting 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. Though we have not integrated this yet, we have data-scraped all evolutions (with those like Eevee's needing manual correction) in preparation for including evolution trees into our project.



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:

- Dusklops 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?
- What Pokémon are similar in stats to Garchomp?

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

Our base dataset that we used. 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 will 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, we are manually drawing regions using an Online SVG Editor, then labeling using reference images (example). This is a long process, and is only partly complete.

Exploratory Data Analysis

The bulk of our time has been dedicated to obtaining data and cleaning it from many sources. However, we have taken some time to explore the viability of our Pokémon similarity network. Using python (because it's quick and easy to use), we wrote up an implementation that will

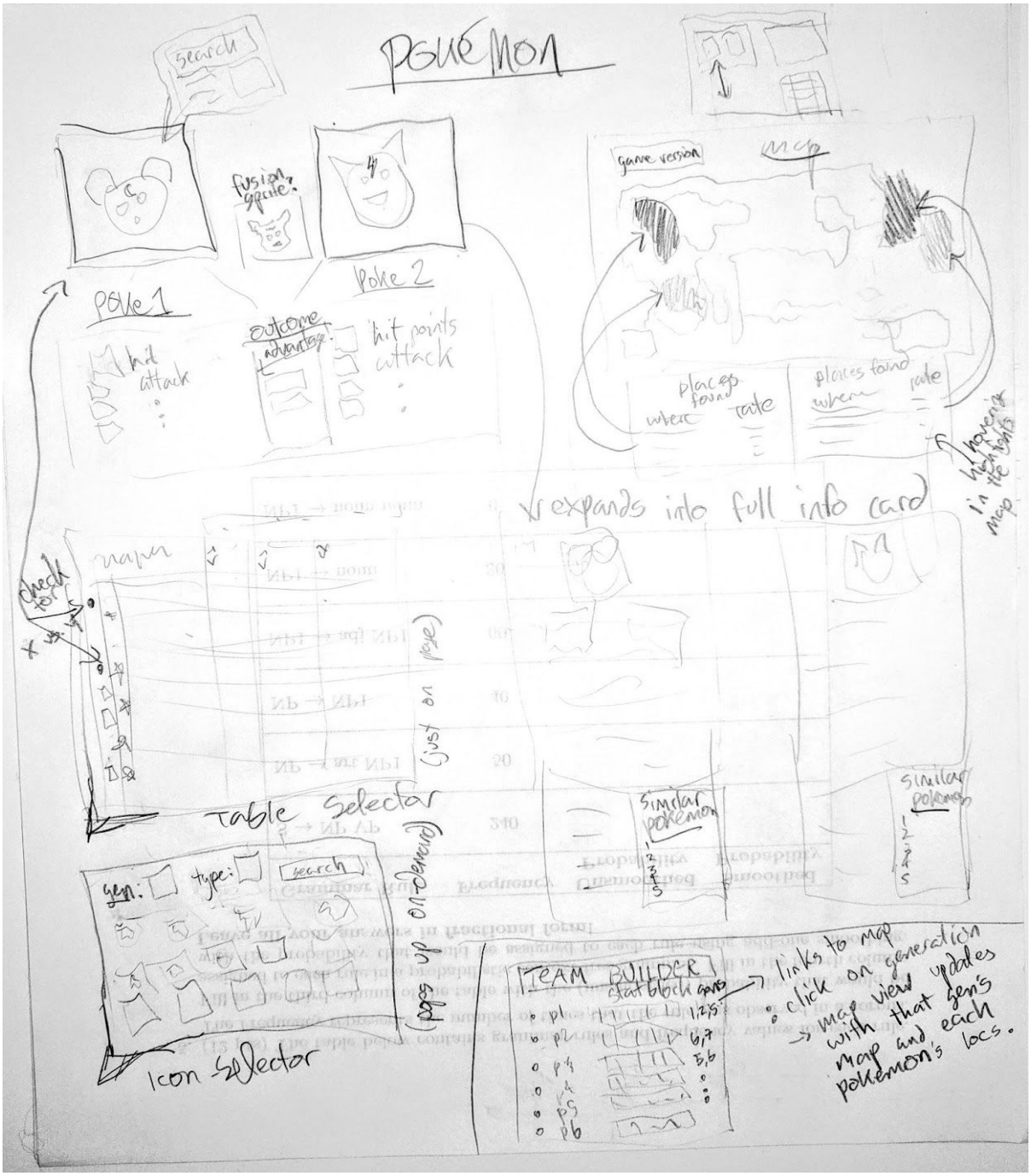
measure the “distance” between two Pokémon. Here, distance refers to the Euclidean distance between two different Pokémon’s stat vectors. Fortunately, since the dataset of all Pokémon is on the small side, it does not take long to calculate the distances between every combination of 2 Pokémon and summarize 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.

Design Evolution

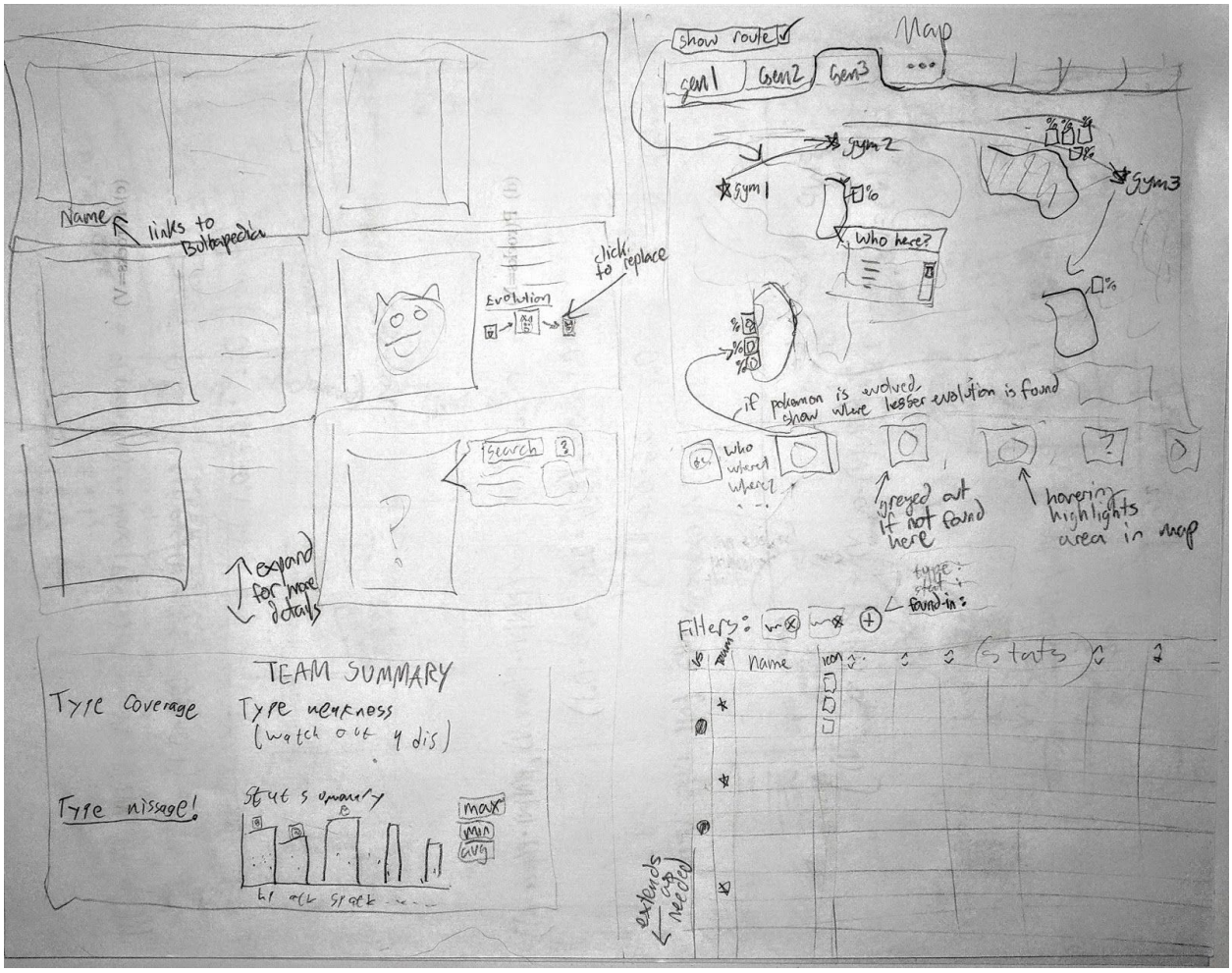
Original Concept



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 underway, and has not significantly changed from our original vision. As we progress, however, some of the optional features are falling away. In data collection, we stuck to just one sprite per Pokémon, rather than animated sprites specific to each generation. The 8th generation will likely not have all the data we need online in time for the project deadline. 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). We are choosing to say “Not obtainable in the wild” or “Not found in this generation” for special Pokémon.

Still, many of our optional features are making progress. We have obtained most of the data needed for evolution trees, and have a good idea of how to implement them. This will allow us to locate where prior-evolutions are found on the map view. The “right-click-to-search” feature has been implemented as a searchable drop-down menu. We have a script to create Pokémon stat similarity data, which we will use to add a “top-five-most-similar” section for each Pokémon card. We anticipate being able to make the similar-Pokémon and Pokémon in evolution trees clickable to make Pokémon selection easier.

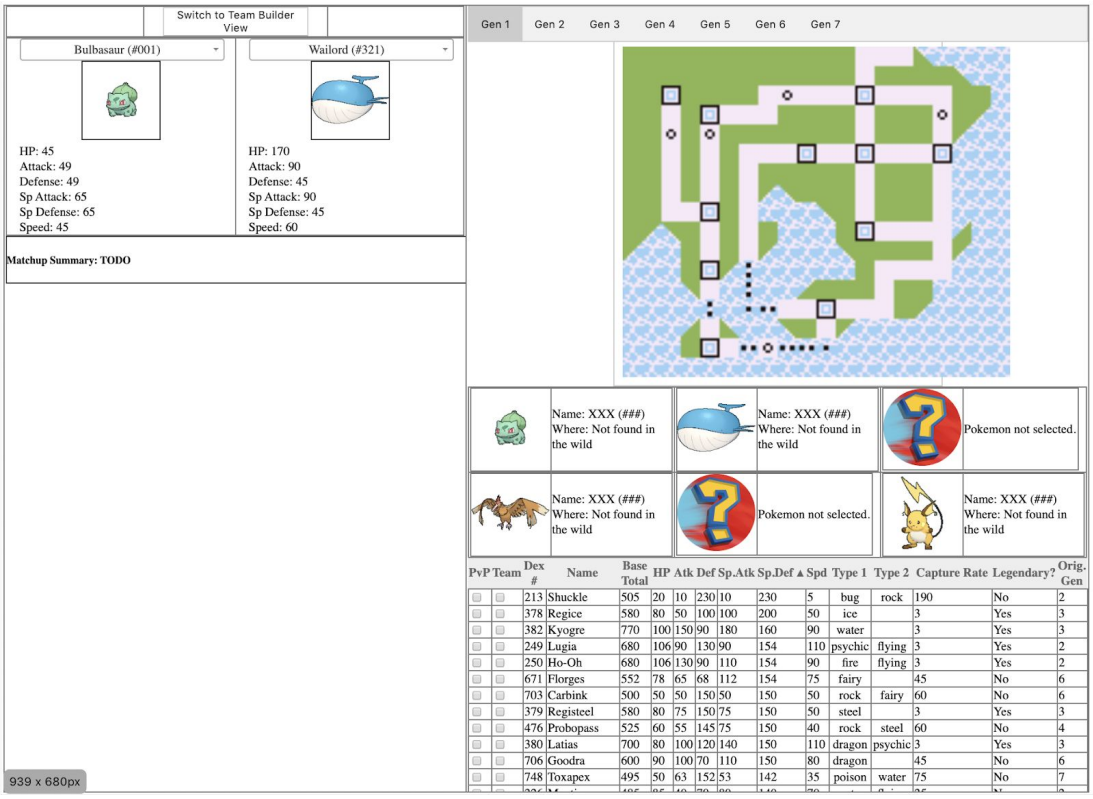
Still to Go

As of this Project Milestone, we’ve met all the scheduled deadlines from our Project Proposal. These mostly involved data acquisition and a basic prototype. Below is a table of what’s done, and what’s left

	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



Pokémon v. Pokémon / Team Builder

Currently, only the Pokémon v. Pokémon view is sketched out in html. There are searchable dropboxes ready to go for when we use d3 to fill them with data. The visualization that is currently in place is rather bare-bones. Instead of using the numerical value for each stat, we intend to use a bar-chart to show how that stat compares to average (Castform).
In addition, we want to flesh out the matchup summary with a mathematical analysis of how the two Pokémon's stats would compare in battle (factoring in Type advantage and other elements). While it isn't currently interactive, the sprite and information for each card will update when the dropdown menu selects a Pokémon.
The Team Builder view will be very similar to the Matchup view. Instead of displaying cards for 2 Pokémon, it will show 6.

Pokédex Table

This is implemented as a table, and made sortable via the sorttable library. The leftmost two columns contain checkboxes for adding a Pokémon to either the PvP view or Team Builder view. This is not yet implemented, but will eventually only allow 2 (PvP) or 6 (Team Builder) checkboxes checked at a time. New checks will place that Pokémon in the first available slot, or replace the second (PvP) or sixth (Team Builder) slot if all slots are full.

[Located in js/pokedex.js]

Map View

The map view is implemented as seven tabs, one per Pokémon generation. Tabs were implemented copying W3School's method, with some modifications. Under each tab is an image of the in-game map for each generation of games. Below the maps is a set of currently-selected Pokémon, each listing the areas where they can be found on the map. Since many Pokémon are found in many places, this will be a scrollable table.
This view is not interactive yet, but will eventually respond to changes in which Pokémon are selected, and whether the visualization is in PvP or Team Builder Mode. When the map above changes, so too will the locations listed. Hovering over a Pokémon's sprite will eventually highlight all the regions that Pokémon can be found in.
[Located in js/maps.js]

PokéWeb

This is unimplemented currently, save for a script to generate data. This will eventually be a 2-way scrollable window, containing sprites linked to each other via d3 force layouts. This will be located either as a tab within the PvP/Team Builder view, or as a separate linked page.

Evaluation

Not done yet! Without all of the interactive elements, it is difficult to draw conclusions of Pokémon comparisons. So far, the best insights have come from our data collection phase.

- As was stated above, Castform is the closest on average to all other Pokémon using our Euclidean distance.
- Shuckle is the furthest from other Pokémon.
- There is actually a Pokémon named "Type: Null", because I guess the game devs hate us.
- Magikarp is the most common Pokémon across all games because it can be fished out of nearly every body of water in every game.
- Generation 7 doesn't actually have an in-game map (at least, not a super helpful one)
- Generation 1 Pokémon tend to be more common than others, as they are more likely to appear in subsequent games.