DataViz — Process Book

Olivier Cloux (236079) — Jonathan Collaud (235493)
Hugo Hueber (246095)

Debriefing

The objective of our course was to realize an interactive and dynamic visualization from a database of our choice. We wanted to choose interesting data to study, but above all to be able to make a fun and enriching visualization for the public.

By exploring the various databases available on the Web, we have pre-selected different themes, each with its own interests. The choice of databases was based first on the interests of all the members of the group, then on their apparent completeness.

We have therefore selected the following themes:

- Memes (From SNAP: Network datasets: 96 million Memetracker memes): A meme is an idea spreads by means of imitation. It acts as a unit for carrying cultural ideas, symbols, or practices. The memes are a major part of our present culture, and a study of their influence and/or origins could have been most interesting.
- Avocados (From Avocado Prices): An avocado (or avocado pear or alligator pear), is botanically a large berry containing a single large seed. Besides, it is commonly used as the unit of value for millenials, and the reason for them to be broke. Some controversies makes them the new “green gold”, as they are cheaper to produce and more cost-effective than most local roots.
- Pokémon (From Pokemon with stats): Pokémon is a Japanese media franchise centered on fictional creatures called "Pokémon". The influence of this universe, since 1989, is constantly increasing, and the universe is so vast that much research has already been done on the subject.
- Happiness (From World Happiness Report): What is happiness, and how do you calculate it? Interesting research, if ever there was one.
- Suicide rates (From Suicide Rates Overview 1985 to 2016): Suicide is a scourge of our modern society. Statistics on it exist, but the subject remains taboo in many circles. One of the objectives here would have been to study the data available to us, so that we could potentially find symptomatic links between suicide and the various corresponding environmental variables, with a possibly to prevent it.
- Google Play Store apps (From Google Play Store Apps): Many applications exist on the Google Play Store, with links between them and information about downloads, rating, etc. Finding the recipe for the best app would have been fun here.
- Video games (From Video Game Sales): We like video games. What makes a video game a commercial success will have interested us.

After talking among ourselves, we decided that a visualization of the Pokémon universe would be ideal. The complex universe was talking to all three members of our group, and we wanted to do something fun, yet educational. All that remained was to find a question, a proposal.

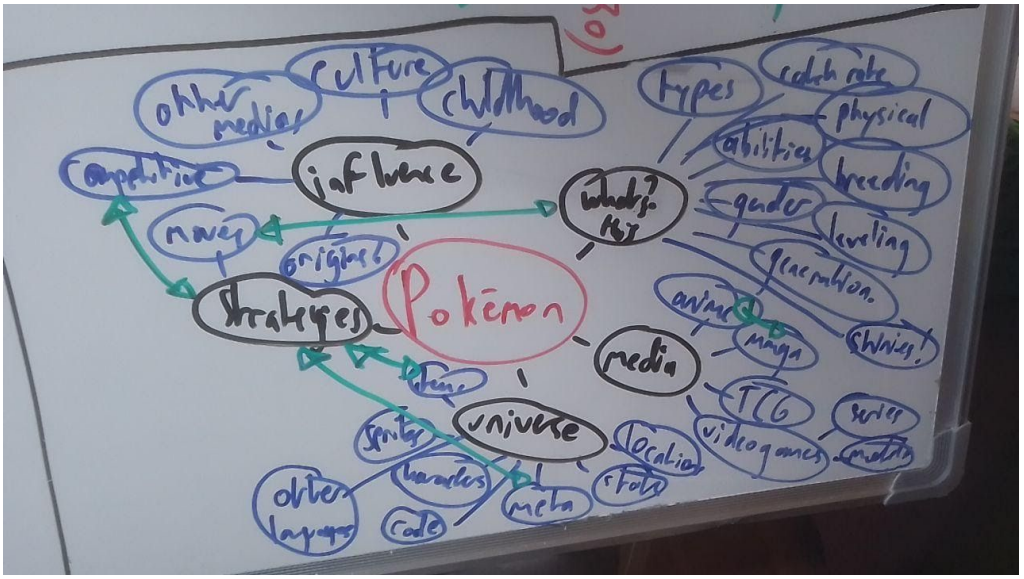


Fig. 1: Mindmap of one brainstorm about Pokémon.

There are many things that would have been interesting to study in Pokémon. We thought about studying the influence of the Pokémon universe on other media or culture in general, or just studying the universe itself. There were an incredible amount of statistics on just about everything in the world, but especially on the Pokémon themselves.

We were most interested in the links between the different Types, and how they influence each other. Influence of the types vs. the statistics, and analysis to answer the big "best starter" question (i.e. "Which is the best Pokémon to choose to start the game") will be a lot of fun as well.

Our problematic was thus found: **how to easily select the best Pokémon based on your preferences and style of play**.

Our approach attempted to interactively showcase Pokémon statistics, on the assumption that it will not be viewed by leading data scientists, but rather by ordinary people who want to discover the mathematical world of Pokémon. It was important to us to make this data accessible and understandable to as many people as possible.

Research

Once we had selected the original dataset, and had a rough idea of the subject, it was time to research the subject. There was two main subject for us to research:

1. Related datasets and information to use and display
2. Previous work in the subject

We wanted to use the former to enrich our analysis; while the original dataset was quite interesting, it only contained the statistics for the Pokémons. We looked at the latter to gain inspiration on what was possible, look for ideas that were not present in our original mindmap, and avoid copying someone else's work.

```
# Dataframe
df = pd.read_csv("../data/pokemon001-721.csv")
df.head()
```

	#	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation	Legendary
0	1	Bulbasaur	Grass	Poison	318	45	49	49	65	65	45	1	False
1	2	Ivysaur	Grass	Poison	405	60	62	63	80	80	60	1	False
2	3	Venusaur	Grass	Poison	525	80	82	83	100	100	80	1	False
3	3	VenusaurMega Venusaur	Grass	Poison	625	80	100	123	122	120	80	1	False
4	4	Charmander	Fire	NaN	309	39	52	43	60	50	65	1	False

Fig. 2: Sample of our data. The dataset offers many statistics to explore, including the types of Pokémon and some of their stats.

We encountered multiple interesting datasets.

- First one, that contains the images of pokemons, with a resolution of 120x120 px Pokemon Image Dataset
- We also found a related dataset, with slightly bigger images (about 256x256) Pokemon Images Dataset
- PokemonDataVisualization This one really inspired us when defining our project. We were amazed at how much information was present on it, all without being fuzzy.
- Pokemon List by generation contains the list of all the pokemons with a small icon of resolution 32x32 that we use for the lower level of pictures resolution.

Challenges

During the preliminary analysis and implementation, several challenges arose. Firstly, the Pokémons (in general) are fuzzy, and don't have a 1:1 mapping with a unique identifier. Some of them have Regional forms (notably Alolan and Galarian) (Regional form), that is a slightly different version of a same Pokémon (e.g. different skin, slightly different set of capacities or stats,...). Also, one pokémon might have multiple skins, based on other conditions (notably, Burmy Burmy (Pokémon) has 3 skins depending on the location where it last battled). Finally, the game version you play might have an influence too (for example, Deoxys Deoxys (Pokémon) has 4 different skins and slightly different stats, depending on the game you play). Finally, in recent versions was introduced the Mega evolution, a temporary evolution that will revert at the end of the match, making the Pokémon more powerful, and with a different skin. And to top that, the 120x120 images dataset was slightly incomplete.

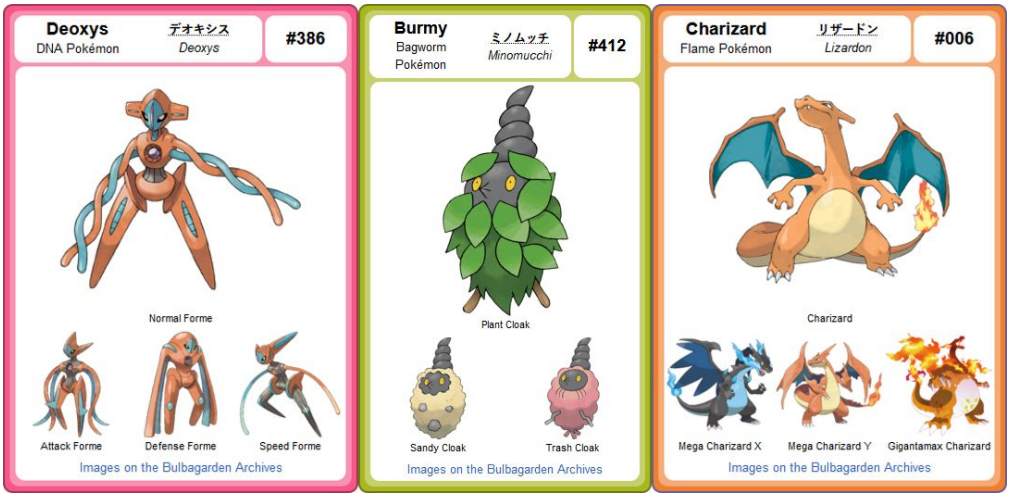


Fig. 3: Examples of multiple variations. Source: Bulbapedia.

All of that made working with the images challenging. We identified the missing images in 120x120, and created versions of them by copying and resizing 256x256 images. We decided to deal with Pokémons that have multiple skins by choosing one “normal” skin. We wished to present all skins to the user, with some input from them; sadly, this proved to be too challenging, and seemed out of our original scope of comparing statistics. We thus decided to keep one image per Pokémon.

Furthermore, no member of our team had good knowledge of web development and in particular of D3.js. Getting familiar with this framework consumed a solid share of our time. We struggled to make the brush work, and failed to implement the zoom, by lack of experience of the framework.

Finally, concerning the evolutions, we were unable to find a preformatted dataset with usable information of the evolutions. We therefore had to scrape Pokémon evolution charts to obtain list of evolutions, using BeautifulSoup and a quick python script.

Implementation

For the implementation of our visualization, we relied on two main technologies, Bootstrap and D3.js. **Bootstrap** is a free and open-source CSS framework for front-end web development, with CSS JavaScript design templates. **D3.js** is a JavaScript library for producing dynamic, interactive data visualizations in web browsers.

The base of our website uses Bootstrap. We made a basic template, based on the one-page template, with four areas. The first zone is an introduction zone on our toy, with a short explanatory text. The next zone is the visualization, which we will talk about below. The third area contains simple instructions on how to use the visualization. Finally, the fourth zone contains a detail about the team.

We made a floating menu using Bootstrap, so that the menu can be constantly accessible to the user. We have also implemented autoscroll, so that the user can go directly to the viewer - we leave it up to the user to explore further.

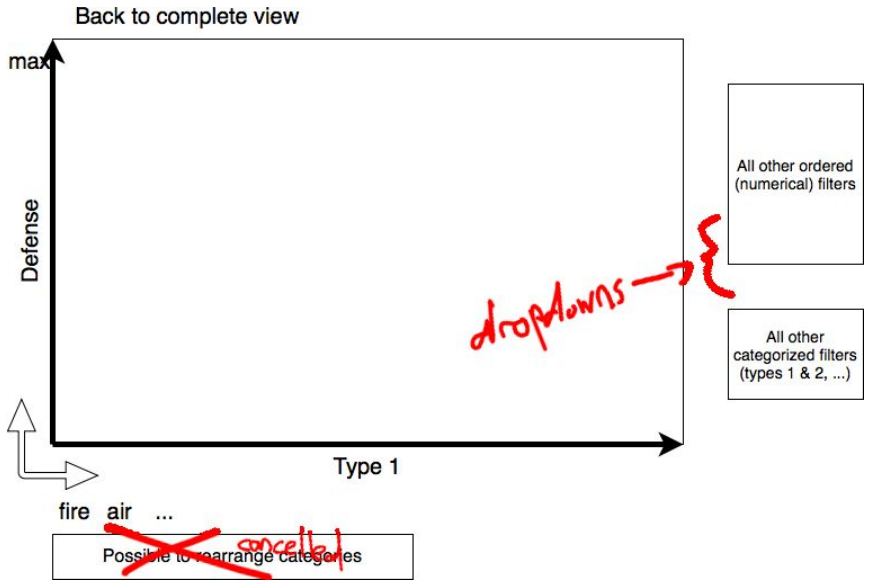


Fig. 4: Minimal Viable Product, updated.

Concerning the implementation of the Minimal Viable Product (MVP), we chose to cancel the possibility to rearrange the categories and only use an auto sorting feature. The menu on the right was made out of dropdowns, so that the user can choose what axis they would like to use.

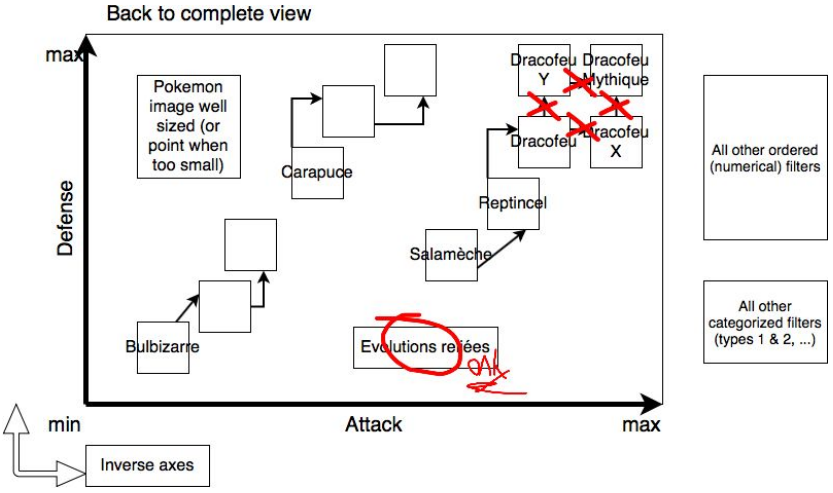
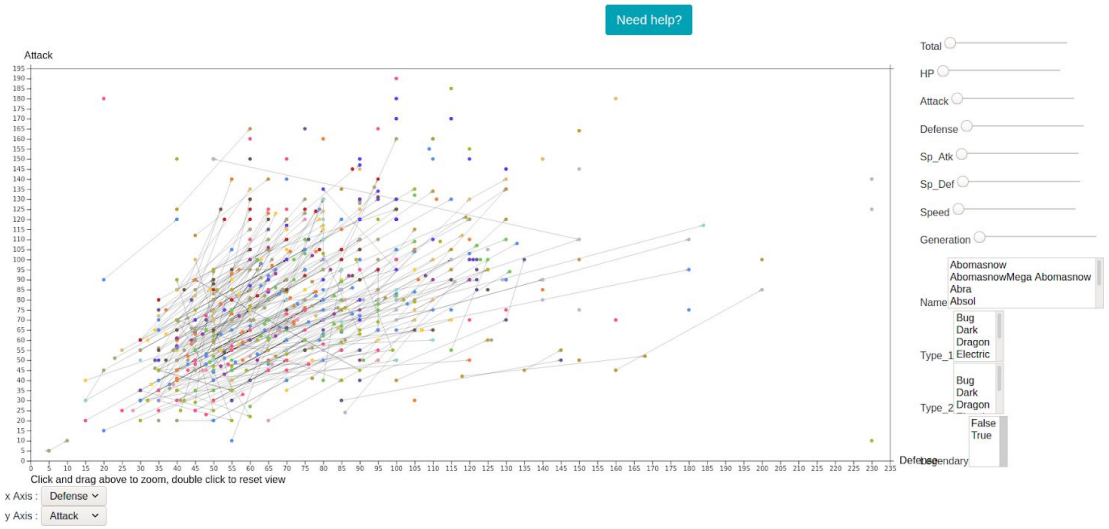


Fig. 5: (Not so) Complex links.

In the end, we decided to keep the links only between Pokémon evolutions. This is because the different forms (mega, alternative, skins, etc.) can only be obtained in exchange, not from an original Pokémon. It was therefore not interesting for us to see them appear explicitly, as the user can't actually use this information.

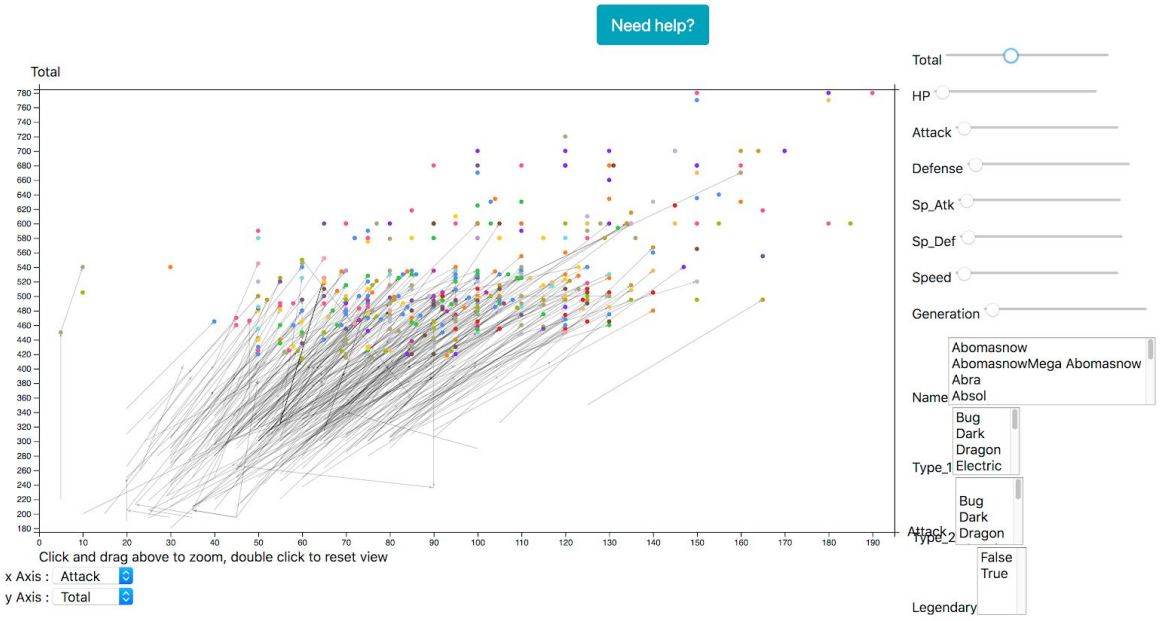
Results

The basic features of the visualization are working properly. The 2D view plots the data points according the features selected in the corresponding dropdown menus. A change in some axis feature trigger the transition to the new view, adapting axis minimum and maximum, as well as all the data points.



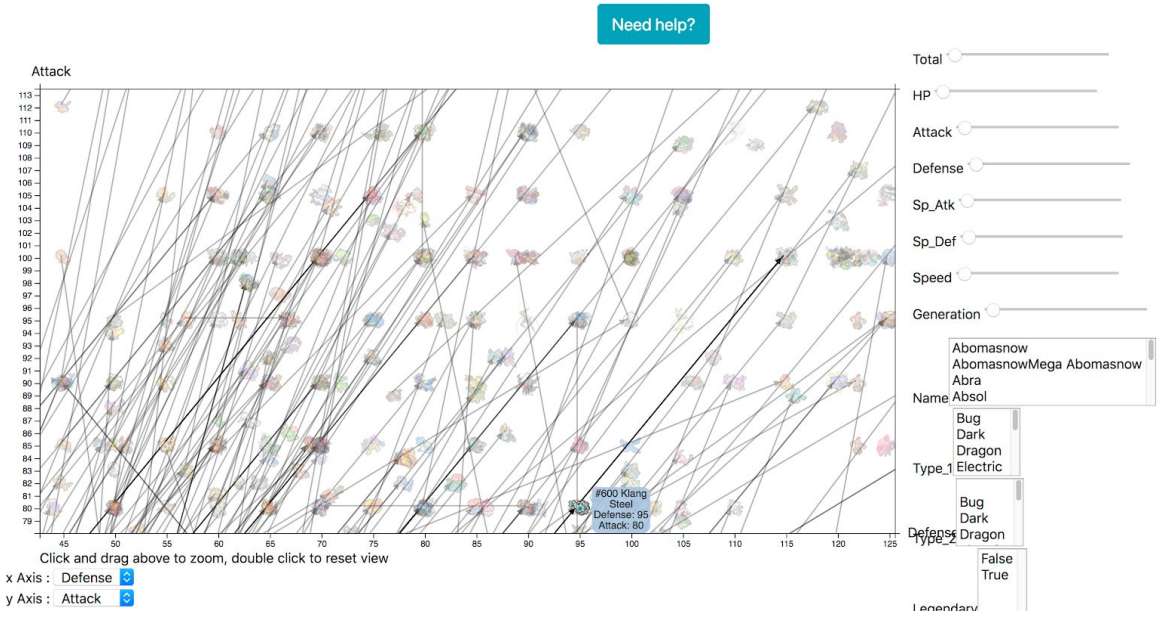
The brush tool allows the user to zoom in the selected region, and dynamically adapt the point size. When the points are bigger than a certain limit, they are replaced by a visual representation of the pokement. Again, there are three different resolutions of pictures that used according to the zoom factor.

The multiple selection filters allow the user to display only a part of the dataset, to focus the visualization on the parts that interest itself. However, the end of the semester was quite full of different delay for all the member of the team and the filter feature is still in a beta version, with only the “Total” filter that works. Also, while the points/images disappear, the links don’t, and the scale doesn’t change accordingly.



When one looks closely to the points, it will see that their color represents the Type_1 feature of the pokemon. There are also arrows that represent the “evoluates in” relationship.

Finally, a function highlights the hovered pokemon and its evolution relationships (both parent and child), and displays different features with a tooltip too.



Peer Assessment

Olivier Cloux

Super Nerd **Olivier Cloux**
"Strong Pokemon. Weak Pokemon. That is only the selfish perception of people. Truly skilled trainers should try to win with all their favorites." – Karen

- Favorite Pokémon: **Alakazam**
- Best move: **Jackpot**
- Fighting style: **Full Frontal**

I took part in some preliminary analysis of the dataset we had at hand. Later on, I laid out the structure of the process book, and wrote some part of it. Then, I took take of the the evolutions, meaning scraping them and displaying them. Visualization-wise, I implemented the tooltips and the evolutions links, as well as most of the "hover" features.

Jonathan Collaud

Engineer **Jonathan Collaud**
"We do have a lot in common. The same earth, the same air, the same sky. Maybe if we started looking at what's the same, instead of looking at what's different, well, who knows?" – Meowth

- Favorite Pokémon: **Pidgeot**
- Best move: **Splash, *but nothing happened.***
- Fighting style: **Action, Reaction**

I implemented the first version of the chart, including default axis values, brushing, zooming and "point vs. circle" management. After that, I chose to implement the different filter options. Separate numeric vs. textual filter attributes was quite easy, but dealing with the filters and linking their outputs to the data points make me lost a lot of hairs.

Hugo Hueber

Hex Maniac **Hugo Hueber**
"It's more important to master the cards you're holding than to complain about the ones your opponent was dealt." – Grimsley

- Favorite Pokémon: **Absol**
- Best move: **Requiem**
- Fighting style: **Endgame Havoc**

In the project, I took care of the consistency and completeness of the reports, including writing, verification of sources and spell checking. I also took care of the preliminary study of the database and the formatting of our visualization. Concerning the visualization itself, I implemented dynamic data selection (sliders, conditions, …), and refactored the code along the way for each step.