Process Book

Yanni Zhang 336558
Xiaoyu Chen 337263
Sixiao Xu 338094

Introduction

From daily living to national-level decisions, the era of data explosion is refreshing the way people explore the unknown. It's only natural that as we grow older, we desire to learn more about ourselves. We often have deep questions in our hearts, such as, "Who am I?" What kind of individual am I? Is it true that I'm the only one of my kind? This is the driving force behind our MBTI personality research.

The Myers–Briggs Type Indicator (MBTI) identifies psychological tendencies that influence how people see the environment and make decisions. Introversion or extraversion, sensing or intuition, thinking or feeling, judging or perceiving are the four characteristics assigned by the MBTI.

We strive to present insightful visuals from three MBTI datasets to enable individuals better understand themselves in this context. In addition, we want to delve deeper and compare specific features of each personality in order to uncover clues. We also put a premium on providing an interactive user experience, so the visualizations we present here are not only instructive but also attractive to the eye. Here is the link to our official website: https://github.com/com-480-data-visualization/datavis-project-2022-nomorebugs.

Problem Statement

Our inspiration comes from the Myers-Briggs Type Indicator® (MBTI®) personality inventory. Its purpose was to make the theory of psychological types understandable and useful in people's lives. We explored three related datasets, hoping to gain insights from different perspectives. People can take a test on 16Personalities, then get more inspiration from this project.

The main axis we developed are the following:

- How are different personality types distributed around the world? What is the dominant type in each country? We also give a detailed introduction of each MBTI personality.
- Do different types of people post on Twitter with different styles? What are their main topics? The frequent words different people use may reflect their personality.
- What are the personality types of some well-known movie characters? Are there any predispositions to characters in different movies genres?
- Is there any connection between personality type and birthday(season)? Maybe the people born in summer are more outgoing?

Dataset

We found 4 datasets on kaggle about different aspects related to the MBTI test.

How the MBTI types are distributed around the world: MBTI-TYPES Data

This dataset contains 2 files:

- *countries.csv*: compares distribution of MBTI-Types in 158 countries
- types.csv

  : contains following information:

  - Type: type (as index)
  - Description: a to-the-point description of type
  - Nickname: nickname associated with personality
  - Definition: a formal definition of type
  - Celebrities: celebrities of that type
  - E/I/T/F/S/N/J/P: abbreviation of features like "extroverted", "sensing", "judgmental" etc.
  - and other features like strengths and weaknesses, romantic relationships...

We are only interested in *countries.csv*.

Posts by people of different MBTI types: (MBTI) Myers-Briggs Personality Type Dataset

This dataset contains over 8600 rows of data, on each row is a person's:

- Type (This persons 4 letter MBTI code/type)
- A section of each of the last 50 things they have posted (Each entry separated by "|||" (3 pipe characters))

MBTI types of movie characters: Movie Character MBTI Dataset

This dataset contains over 18700 rows of data, on each row is a character's:

- stat: number of people who voted/commented for this character, as his type is determined by the votes of users on the forum.
- mbti: 4 letter MBTI code/type
- enneagram: the enneagram type, which is another typology that divides humans into nine interrelated (non-psychological) personalities.
- role: name of the character
- movie: name of the movie in which this character was portrayed
- img_url: url of an image of the character

We mainly focus on `mbti`, `role` and `movie`.

MBTI types and birthdays: MBTI and Birthdays

This dataset contains about 300 rows of data, on each row is a person's:

- timestamp: timestamp when this record was created
- birthday
- Time of Birth (not required)
- MBTI type

We are not interested in the time of birth, so we simply drop it.

Tools for visualization

This project consists of two main parts: drawing charts and web design. We used the following two main tools to accomplish this.

- Plotly: An interactive, open-source, and browser-based graphing library for Python. It ships with over 30 chart types, including scientific charts, 3D graphs, statistical charts, SVG maps, financial charts, and more.
- Vue.js: An open-source model–view–view model front end JavaScript framework for building user interfaces and single-page applications. To make it more elegant, we also used Element UI and AOS to unify the UI style and add lovely animation effects.
- Element UI: A component library based on Vue 2.0 for developers to use. It provides efficient and uniform UI components to brighten up the web style.
- Countup.js: A dependency-free, lightweight Javascript class that can be used to quickly create animations that display numerical data in a more interesting way.
- AOS: Animate On Scroll Library.

Lectures and Tutorials used

- Javascript: We use javascript to complete the code logic in the website.
- Interaction: User interaction is important, and we include elements that can be interacted with both on the page and in the chart.
- Perception colors: Since we have many images, it is very important to unify the tone of each image.
- Designing viz: We actually followed the advice from this lecture to design the skeleton.
- Map: A large part of the site is a map showing the distribution of various personality types in the world.

Professor Review Session & Changes Made Since Milestone 1

After Milestone 2, we met with Professor Vuillon in person, and he offered us very positive and helpful feedback.

He proposed that, instead of the four selection bars we had planned in the skeleton (so that the user could choose between 4*4 personalities), we could consider using a long bar that the user can directly drag between 16 personalities. users can see the evolution of the visualizations between different personalities by doing so. We include his suggestions into our final design.

He also recommended normalization. Therefore, for birthday visualization, we present the proportion within one month for each personality type. In other words, we standardize the data for each month to make it easier to understand.

We also shared our main problem with the professor at that meeting: rather than using a single large dataset, we used multiple smaller datasets with few columns, and we wanted a visualization that could highlight the relationships between them. But at that time we didn’t come up with ideas about this. The professor agreed that finding linkages between different datasets is difficult. He patiently addressed the issue with us, and at the conclusion of that discussion, we decided to solve the problem using a Sankey diagram.

Project structure

Below is the file structure in the main branch.

```
.
├── README.md
├── data
│   ├── DataAnalysis.ipynb           // data analysis for milestone 2
│   ├── DistributionDataset.ipynb    // draw distribution graphs
│   ├── Donut_Diagram.ipynb          // draw birthday distribution graphs
│   ├── Parellel.ipynb               // data processing for coordinate graph
│   ├── Sankey_Diagram.ipynb         // data processing for sankey diagram
│   ├── World_Cloud.ipynb            // data processing and draw world cloud
│   ├── birthday                     // birthday dataset
│   │   └── Birthdays.csv
│   ├── distribution                 // distribution dataset
│   │   ├── countries.csv
│   │   └── types.csv
│   ├── movie                        // movie dataset
│   │   └── mbti.csv
│   └── post                         // post dataset
│       ├── 16statistics.csv
│       ├── mbti_1.csv
│       ├── split.csv
│       └── split_statistics.csv
└── report
    ├── ProcessBook.pdf
    └── Milestone2.pdf
```

In addition to the main branch, several other branches are mainly used for:

- gh-page: website (build)
- gh-page-old branch: prototype website (build)
- website: website source code

Visualization

Our website has two main pages, Introduction and Personalities, which can be selected at the top.

Index Page

This is the first page of our website. To begin, we provide a brief overview of the MBTI and provide a link to the test for users to take. Second, we demonstrate in a playful manner that this analysis is based on a large enough data set. Then there are three graphs with generic data: a histogram of birth months for various personality kinds, a sankey graph of birth months versus continents, and a Sunburst graph of personality types for characters from both the Marvel and Harry Potter books. We conducted a rudimentary analysis of the information represented in each graph.

Introduction to MBTI

In personality typology, the Myers–Briggs Type Indicator (MBTI) is an introspective self-report questionnaire indicating differing psychological preferences in how people perceive the world and make decisions. The test attempts to assign four categories: introversion or extraversion, sensing or intuition, thinking or feeling, judging or perceiving. One letter from each category is taken to produce a four-letter test result, such as "INTJ" or "ESFP".[1]

You can take a test HERE if you want.

About our datasets

Our research is based on 4 datasets on kaggle about different aspects related to the MBTI test. Following the wrangling of the original dataset, we now have:

158
different countries

8,675
different posts

18,741
movie characters

306
birthday records

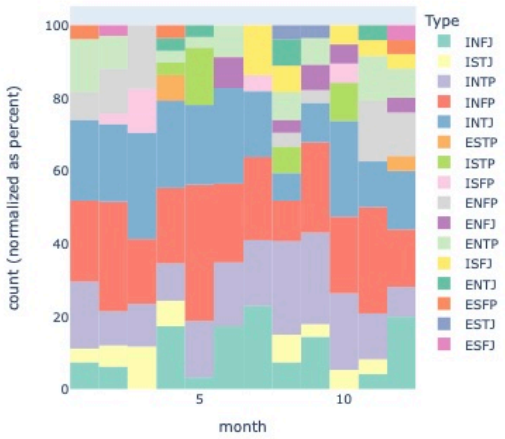
This anaysis has a high level of credibility since they are based on sufficient data.

Data analysis

Birth month distribution

The bar chart on the right depicts the monthly distribution of birth dates for the 16 personalities. We applied normalization to the original dataset to get the result. We found the INTJ and INFP types were frequent in all months except August. Other types like ESFP, ESTJ, and ESFJ are rare.

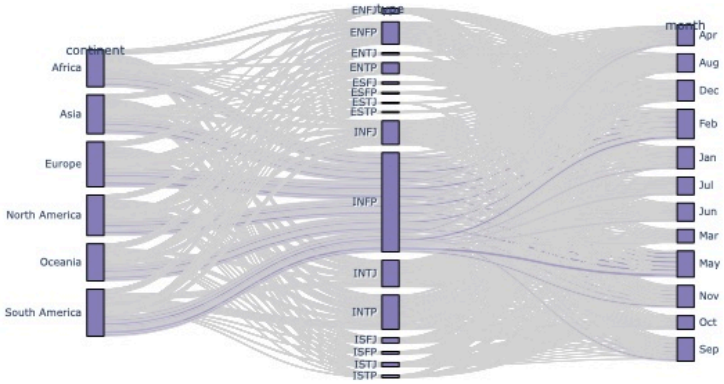
In personality types, E is for extrovert and I is for introvert. In the graph, above the grey line is extrovert and below is introvert. We found that, overall, there are more introverts. But in August, the proportion of extroverts was larger than in other months. This is in line with our expectation that people born in summer are more outgoing.



Birth month and continent

We integrated the two datasets and created the parallel categories diagram below to investigate the association between continents and birth month. The INFP personality type

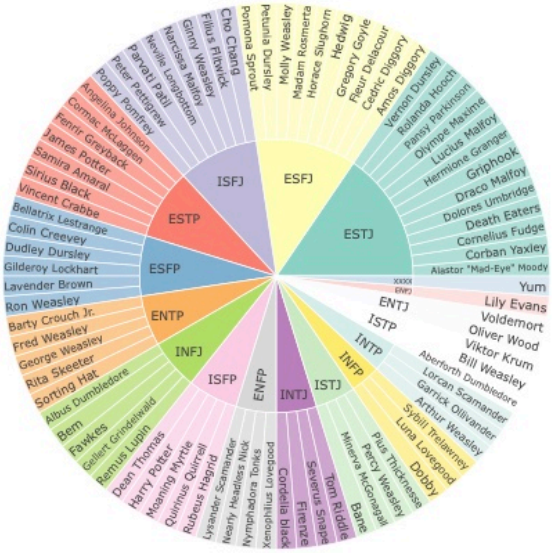
was discovered to be the most frequent of all personality types. There are also personality types that only appear in a few months. This may be due to the size of the dataset, where some cases are not covered.



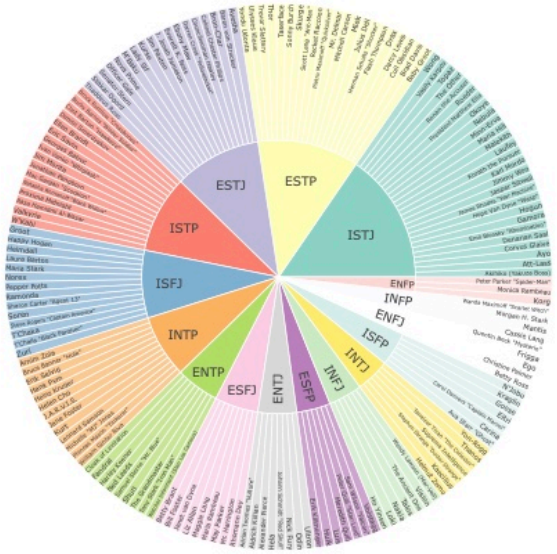
Personalities in cinematic universes

Have you ever wondered what the personalities of the characters in the cinematic universes are like? We explored the distribution of personality types for characters in the Marvel and Harry Potter films and came up with the following two Sunburst graphs. We found that the ESTJ (Executive) type is the most in Harry Potter, and the ISTJ (Logistician) is the most in Marvel movies.

Harry Potter



Marvel

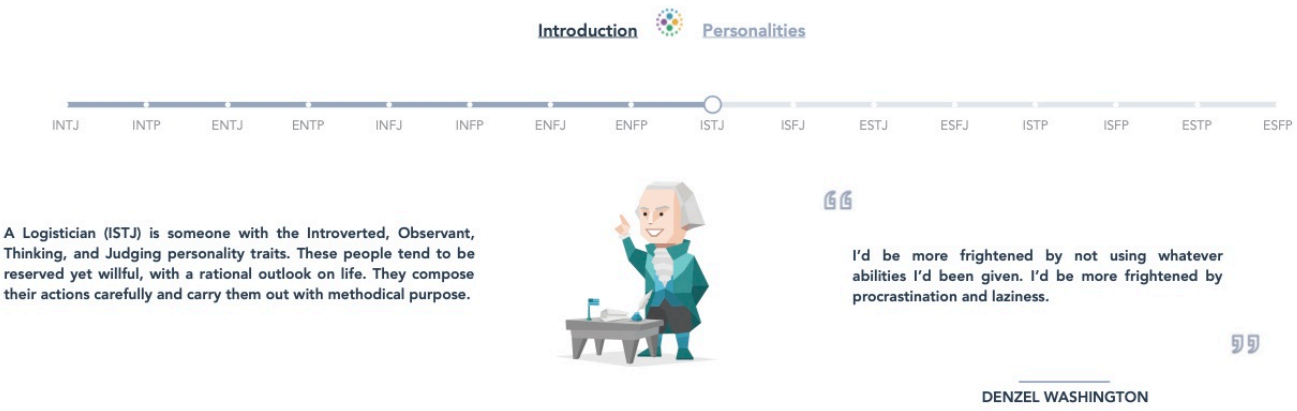


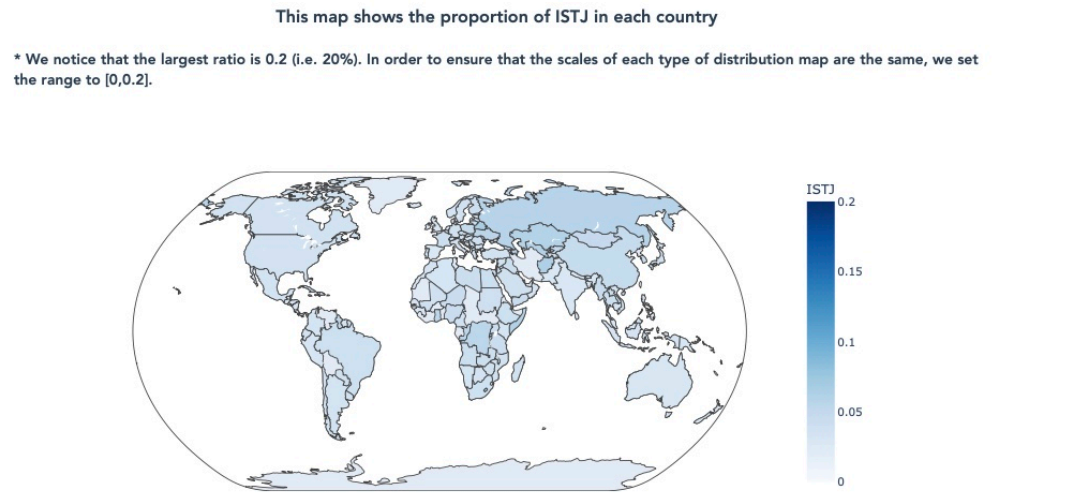
Next...

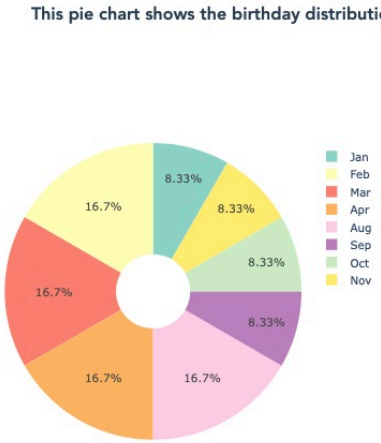
Now, you can go to the Personalities page to explore each personality.

Personality Page

There is a discrete slider at the top of the page that can be used to select a personality type. Below the slider is an introduction to the personality type, a typical picture, and a quote from a celebrity of the type. Then there is a world map showing the percentage of that personality type in each country. Finally, there's a birthday distribution on the left, and a word cloud of what this type of person likes to say on the right.







Challenges and Design Choices

Our main challenge comes from the datasets we choose. We use four datasets to construct our website, trying to give relatively comprehensive insights regarding different aspects of MBTI. But it is kind of hard to find relationships and connections between several datasets. And it is also not intuitive in a visualization that the information we convey are independent from each other.

To solve this problem, the design choice we made here is to build a Sankey diagram where users can get overall information on different aspects of MBTI. In this diagram, we focus on three attributes: personality types, continents, and birthday(all come from distinct dataset). We normalized the data to make more sense of the continent distribution.

Peer Assessment

Each team member focused on their own visualizations and data processing. We all had a hand in the general visualization concepts and decisions that led to the final website.

Sixiao Xu: Website design; World distribution diagram (data processing, visualization, and animations); Sankey diagram (data processing, visualization, and animations)

Xiaoyu Chen: Website design; Donut diagram (data processing, visualization, and animations); Birthday distribution diagram (data processing, visualization, and animations)

Yanni Zhang: Website design; Word cloud (data processing, visualization, and animations); Sankey diagram (data processing)

Conclusion

We propose numerous visualizations of MBTI personalities in this project. We created a web application with two pages: an overview page and a page for each personality's detailed depiction. Vue, Plotly, Element UI, Countup, AOS, and other well-known current web technologies were used to construct the project.

We hope that users will like our website and have a good time interacting with all of the interactive visualizations we provide. We also hope users can gain visual insight from our work and know better about themselves.