# PROJECT 3 REPORT
# Harry Potter Visualization

Min Zheng mz474 | Yanfei Yu yy777 | Chen Li cl953 | Vivian Jiang kj322

---

## Work Distribution

**Min Zheng** has discovered datasets, processed the data and co-implemented the first plot about the relationships and basic demographics of the main characters in Harry Potter series, and the third plot about search trends of Harry Potter keywords from 2004 to 2018.

**Yanfei Yu** has discovered datasets, processed the data and co-implemented the first plot about the relationships and basic demographics of the main characters in Harry Potter series, and the third plot about search trends of Harry Potter keywords from 2004 to 2018.

**Chen Li** has implemented the second plot about popular spells, charms and curse used in Harry Potter series, and finalized the whole layout of the visualization.

**Vivian Jiang** has discovered datasets, drawn illustrations of all Harry Potter characters mentioned in the data visualization, designed the styling of the theme and finished writing up the final report.

## Data Description

### Relations and Demographics

The datasets used for the first plot about the relationships between main characters and basic demographics of main characters in Harry Potter series are “relation.csv” and “characters.csv.” Both datasets were downloaded from pottermore.com, which is a Harry Potter fan-based website.

In the dataset, “relation.csv,” the three variables are people1, people2, and relationship, where people1 and people2 are names of the characters and relationship variable is the relationship between theses two characters.

In addition to the relationship data, the dataset “characters.csv” describes the basic demographic information of each character in Harry Potter series. It has five variables which are name that is the name of the character, born that is the birth date of the character, died indicating the date when the character passed away, blood_status regarding whether the character is pure-blood, half-blood or muggle, marital_status and house.

In order to make the plot clearer and more explicit, we chose a total of 22 main characters that have the most data about them and hand-drew cartoon illustrations of those.


(Harry Potter Character Illustrations)

**Incantations**

For the second data visualization plot that shows the incantations used in Harry Potter series, we used two datasets which are “spell_sum.csv” and “people_spell.csv.”

Since we wanted to show Incantations that were used the most in the Harry Potter series, we downloaded a dataset called “spells.csv” from tableau. Because the original dataset only has the name, type and resulting effect but we would like to show which spell was used the most, we processed the original dataset in R by grouping same incantations and counting how many times each incantation appears in Harry Potter series, finally to a new data file called “spell_sum.csv” which contains two variables, spell and count.

In addition to showing how many times each incantation is used, we would like to connect this plot to the first plot that has the demographic information of each main character by presenting the three most popular incantations used by different characters. In this case, we grabbed a dataset called “people_spell.csv” from link that has the information of popular incantations used by different Harry Potter characters. In the file “people_spell.csv,” there are seven variables including name that is the name of each character, charms1, charms1_num, charms2, charms2_num, charms3, and charms3_num that are three most used incantations by each character and the number of times it was used.

**Search trends**

The data visualization plot demonstrating the search trends of keywords on Google related to Harry Potter series used two datasets: “trend.csv” and “markers.json.” We grabbed the search trends of Harry Potter keywords from Google Trends website and named it as “trend.csv.” For “markers.json” that documents dates of publications and screenings and important Harry Potter related news, we obtained data from pottermore.com and saved it as a json file.

The dataset, “trend.csv,” records the popularity of Harry Potter keywords from 2004 to 2018. The variables are month which indicates a timeline that the data is based on and Harry Potter keywords such as names of the characters. The highest popularity is normalized to 100, with all the numbers indicating the popularity of a certain keyword in this range.

“Markers.json” documents the descriptions of milestones related to Harry Potter series, including the dates and descriptions of book series publications, movies series screenings and some important news events.

The reason we chose these two datasets was that we would like to know if any of the new publications or news events has an impact on the search of keywords on Google.

## Data Mapping

**Relations and Demographics**

The chord diagram represents the relationships between selected characters. The relationships are represented as lines connecting two characters (the relationships are two-way and do not indicate the level of deepness). The number of relationship links depends on how many relations one character has with other selected characters. By hovering over the relationship link, users can explore the details of the relationship between selected Harry Potter characters.

Another interaction that can be achieved by the first plot is that when user hovers over a Harry Potter character, its basic demographic information will be displayed on the right of the chord plot. The basic demographics include date of birth, date of death, blood status, marital status, house, and a cartoon illustration of each character hand-drawn by us.

**Incantations**

The second plot of words shows incantations used by Harry Potter characters. The data we processed counts the times each incantation appears in the Harry Potter series. Therefore, we mapped the data on a plot of words where more-used incantations appear bigger than less-used incantations. Additionally, the data we used categorizes all incantations into three main categories: charm, spell and curse, so we used three different colors to show incantations, letting users know which type the incantation is.

By hovering over the character name in the first chord plot, users are able to see the top three most-used incantations of that character because the interaction highlights the three most-used incantations according to the data we used.

**Search Trend**

The line chart at the bottom displays the popularity of Harry Potter character keywords search from 2004 to 2018. The position of the line indicates the level of popularity of Harry Potter characters search on Google, whereas the dots demonstrating the milestones - either news events or publications/screenings - happening in this period of time.

Users can select two Harry Potter characters and compare the search popularity of these two characters on Google. Then, they are able to see if any of the Harry Potter milestones has an impact on the search results.

## The Story

Harry Potter is one of the most successful publication and movie series in the world, with a variety of characters and a complicated story. Due to the complexity of character relationships and the plot, readers or audience of Harry Potter series might experience hard time tracking each character’s story or differentiate characters. As a result, we created this data visualization that makes clear the relationship between main

characters and the basic demographics of these characters. To make our audience more aware of what each character looks like, we hand-drew cartoon illustrations of all the selected characters.

Incantations are another captivating element in Harry Potter series. Compared to characters, incantations are even more difficult to comprehend or remember. Therefore, our visualization categorizes incantations used in Harry Potter series and displays three most-used incantations by each character, eliciting memories of Harry Potter fans.

Popular series would be more likely to have a huge fanbase. In this case, our third plot gives a sense of the popularity of each main character on the Internet. One surprising thing we discovered through our visualization was that when we chose to see Severus Snape's search popularity chart, we found that the search peak of his name was neither on the publication nor the movie screening date of Harry Potter series. Instead, we searched on the Internet to figure out what happened on the date that Severus Snape's search peaked and found out that the time was the day he passed away. People were not that familiar with his real name so they searched his character name. Thanks to Alan Rickman for bringing this fascinating character to us.