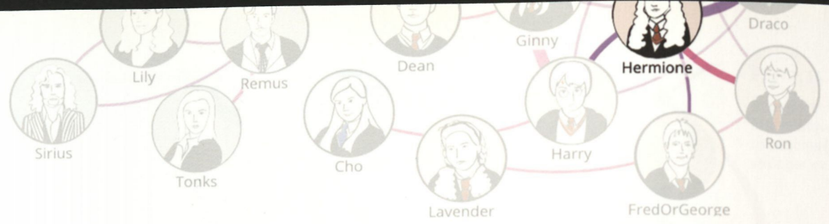
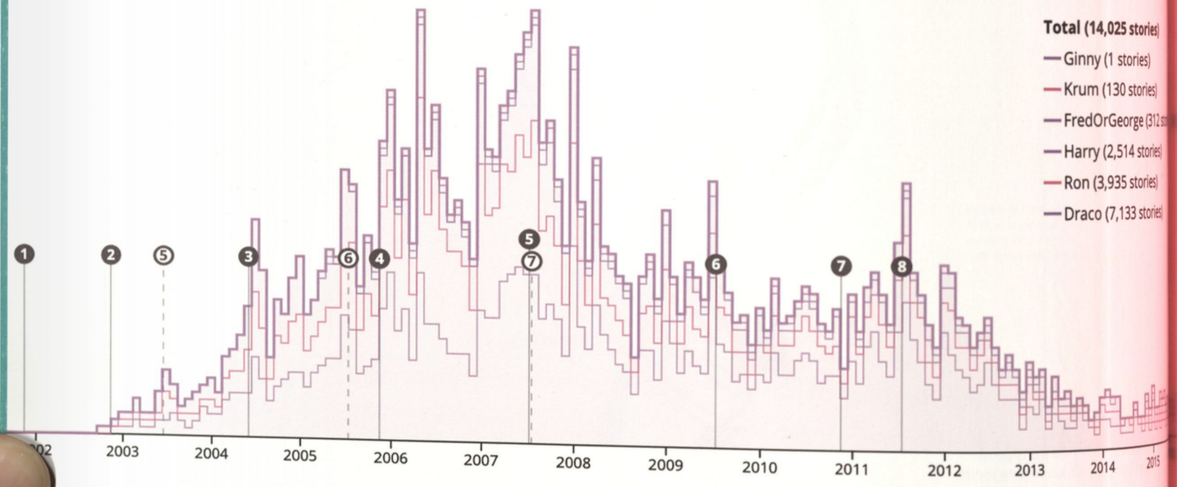
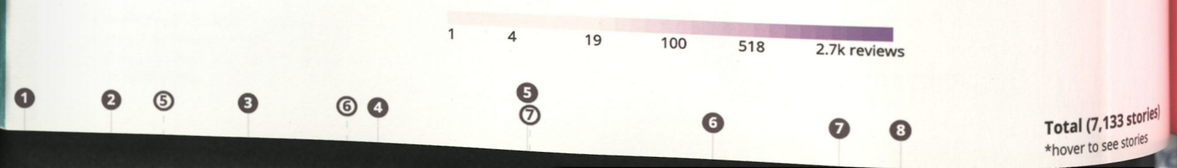


Hermione
1 st film 5 th book



Hermione ♥ Draco



The Most Popular of Them All

SHIRLEY

Alright, I'll be honest; I conned Nadieh into doing "Nostalgia" just so I could finally do a project exploring the *Harry Potter* universe. The Harry Potter book series (and *Toy Story*) will always have a special place in my heart: when the first movie came out, I was eleven years old and eagerly waiting for my Hogwarts letter. I'm the same age as actors Emma Watson and Daniel Radcliffe, who played Hermione Granger and Harry Potter, respectively. When the end credits rolled for the final movie in 2011, I had the distinct feeling that my childhood had ended.

Having said all that (sentimental stuff), I still went through a few different potential angles before settling on my final idea. At first, I wanted to do something on *Harry Potter and the Cursed Child*, but realized there were probably copyright issues with the script. I then bought a Marauder's Map hoping to visualize that somehow, but got stuck trying to figure out what kind of data I could extract from it.

Then one day, I suddenly remembered my original "Books" idea: to use Harry Potter fanfiction as a proxy for fan reactions to the movies and books.

Data

To create the dataset, I decided to scrape HarryPotterFanfiction.com because of its rich and Harry Potter specific metadata: in addition to categorizing by genres and ratings like many other fanfiction sites, it also organizes content by eras ("Hogwarts," "post-Hogwarts," "Marauders," "New Generation," etc.) and character pairings.

Like my "Movies" project, I used the Node.js *http* module to request for each page, but I also decided to use `html2json` to turn the page responses into JSON. I did this because with the "Movies" page responses I only needed to grab the ID of each movie, so I only had to write one regular expression. But for this project, I wanted to grab a lot of metadata and I didn't want to write regular expressions for all of them. With the HTML response converted to JSON, I only had to write a small function to loop through the children objects for each story and find the metadata I wanted (though I still had to do some work cleaning up the raw strings).

Because the website would only give me 25 stories per page, I had to loop through and request approximately 3,300 pages. I was worried that the server would time me out for making so many requests at once, but because I made those requests synchronously (asking for the next page only after the current page had responded and I had completely parsed through it), there were just enough seconds between each page request that I never had any issues.

I usually try to find the simplest and quickest way to gather a static dataset because I know I don't have to worry about performance, but this time that laziness actually paid off (๑•̀ㅂ•́)و✧

When I had all the data, I wanted to get a sense of it by logging some statistics to Terminal, the command line interface for Mac. I looked primarily at the publication years (Harry Potter fanfiction spiked in 2007 when the final book and fifth movie came out, but has been dwindling significantly ever since), and the most popular pairings (unsurprisingly, Harry/Ginny followed by James/Lily). I also looked at which characters appeared the most often in these stories (Harry was the most popular, followed by Hermione in second, the author's made-up "Original Character" in third, and Ron in fourth—which really says something about Ron's lack of popularity).

I had originally tried to log statistics to Terminal as a quicker way to analyze data (as opposed to building visualizations from scratch to test my hypotheses, like I did with my previous projects), but found that it didn't work out too well for me. My knowledge was limited to logging the min/max and some medians, and though they gave me a basic understanding of the data, they were very superficial and didn't help me find the deeper, more interesting insights I was seeking. And though I knew there were many more sophisticated approaches out there, I didn't know when or how to use them, so I decided to move onto the sketch section.

Fig.7.1

Logging some statistics to Terminal.

```
[ [ 'undefined', 18569 ],
  [ 'Other Pairing', 18416 ],
  [ 'Harry/Ginny', 15200 ],
  [ 'Ron/Hermione', 14171 ],
  [ 'James/Lily', 12057 ],
  [ 'Draco/Hermione', 8826 ],
  [ 'Sirius/OC', 6993 ],
  [ 'OC/OC', 5491 ],
  [ 'Remus/OC', 4322 ],
  [ 'Draco/OC', 3866 ],
```

```
[ [ 'Hogwarts', 23177 ],
  [ 'undefined', 14964 ],
  [ 'Marauders', 13299 ],
  [ 'Post-Hogwarts', 12914 ],
  [ 'Next Generation', 10270 ],
  [ 'Other', 4137 ],
  [ 'Pre-Hogwarts', 1879 ],
  [ 'Founders', 517 ],
  [ 'The Cursed Child', 7 ],
  [ 'Fantastic Beasts', 5 ] ]
```

Sketch

I decided to start my brainstorming by writing down all the metadata I had. From there, I identified genres, eras, and pairings as potentially interesting metadata that might correlate with a story's number of reviews, which I used to measure its popularity and "success." I also jotted down some questions for the data: which pairings were the most popular, and which other pairings did they usually appear with? Who did a character get paired with the most?

My first idea was to create a timeline of stories for each character and denote pairings for that character on top of the timeline. But it was around that time that *The Pudding*, a digital publication dedicated to visual and data-driven journalism, came out with a piece called "The Shape of Slavery"[1] (Figure 7.2). I liked how they used both the size and color of the dots on a map to encode two related dimensions (the size of Black population and their percent of total population). I was also visually inspired by the sequential color scale they used, which ranged from a deep blue to magenta to orange to yellow on a navy background.

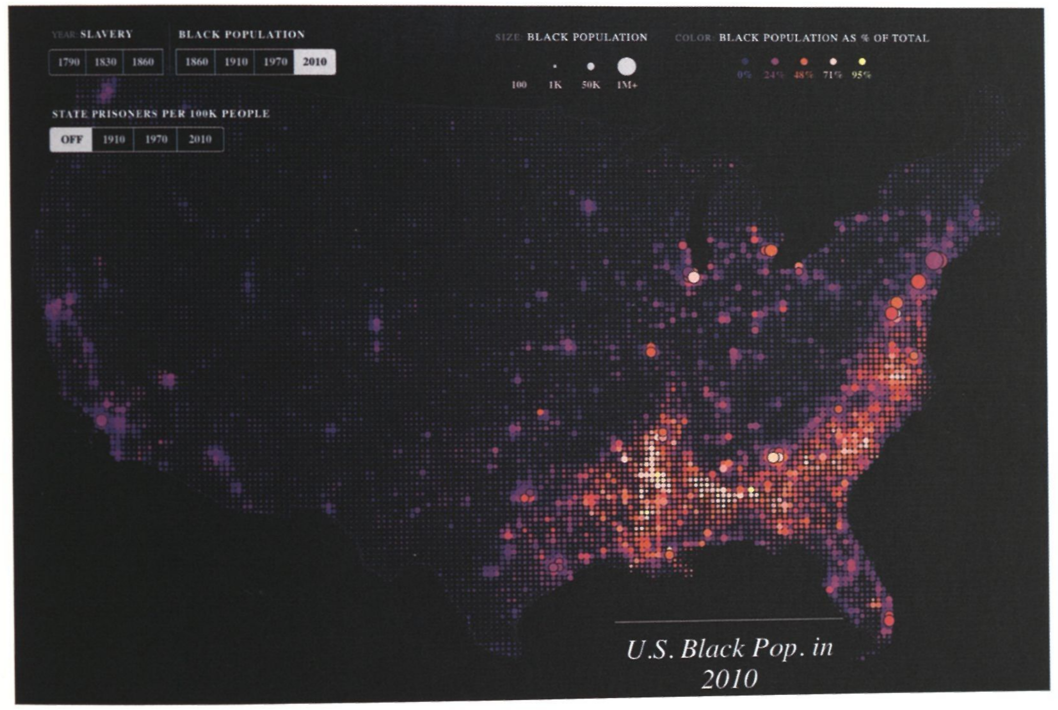


Fig.7.2

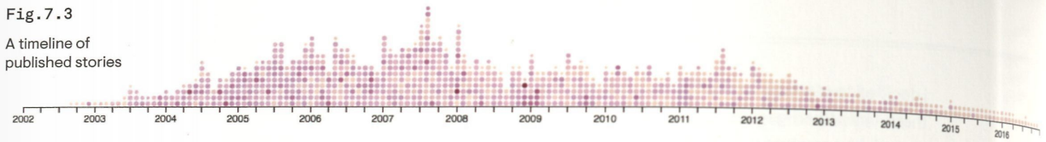
"The Shape of Slavery," a visual essay from *The Pudding*.

So, inspired by their visualization, I decided to give the dots a try. Since there were roughly 80,000 stories on HarryPotterFanfiction.com, I decided to have each dot represent up to 100 stories, color the dots by the average number of reviews (and thus the popularity) for those 100 stories, and place those dots over a timeline. I was keen on using a timeline because in my data exploration I noticed that there were definite spikes and declines in the number of stories across the years, and I knew a timeline would give good context. To place the dots, I grouped stories by the month of their publication (which determined their x-position), then stacked each set of 100 stories on top of each other. It ended up looking like a histogram made up of dots (Figure 7.3).

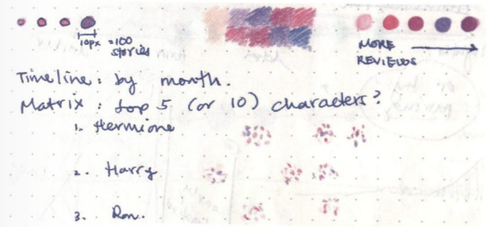
[1] "The Shape of Slavery" data visualization by The Pudding: https://pudding.cool/2017/01/shape-of-slavery/

Fig.7.3

A timeline of published stories



To get a sense of character pairings, I also sketched a matrix of the top five most popular characters along with their partners and placed all the stories at the intersection of those pairings (Figure 7.4, right).



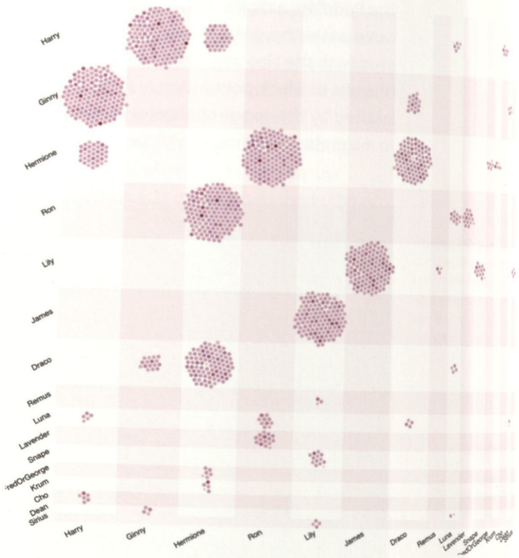


Fig.7.4

Sometimes my sketches are really simple, just jotting down some ideas (left), and the matrix idea implemented (right).

I quite liked the timeline. I could immediately see overall trends, including when the most stories were published and when those stories had the most reviews, which really showed when Harry Potter and the fandom peaked. I had mixed feelings about the matrix, though; on one hand, I liked being able to see who a character was paired with and in how many stories. On the other hand, I didn't like how much space the matrix took and how it mirrored itself across the diagonal and thus had redundant information (a drawback of matrices).

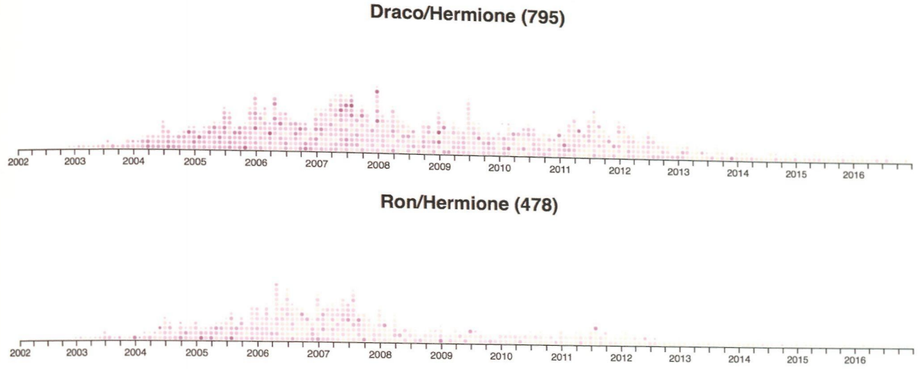
Looking back now, I do like how I varied the size of the row/columns for less popular characters so that the matrix didn't take up as much room.

I ultimately decided to keep the timeline, but removed the matrix. With that decision, I wasn't sure how to move forward with the visuals I had and what angle I wanted to approach the data from. It was around that (uncertain) time that I hung out with my friends RJ Andrews and Catherine Madden for an afternoon, and two beautiful things came out of that day.

First, while showing them screenshots of my progress, RJ noticed my timeline of Draco/Hermione was stacked on top of Ron/Hermione (Figure 7.5). He excitedly pointed out that one pairing was canon (Ron/Hermione, official in the books) and the other was non-canon (Draco/Hermione, a fan pairing), yet the non-canon had more stories. And thus we were able to narrow in on an interesting angle: how do canon and non-canon relationships compare in popularity? I re-thought all of my visuals around that question.

Fig.7.5

Timeline of Hermione/Draco and Hermione/Ron pairings.



One of the main things I decided to do was visually separate canon versus non-canon pairings using color: I chose pink for canon and purple for non-canon. I chose warmer colors since, to me, they represented relationships more, and I always loved the look of pink and purple together. Since I wanted to map those colors to the number of reviews, I called `chroma.js's scale` function with an array of colors to create a gradient that traveled through all of them.

`chroma.js` is an amazingly handy library for working with color.

And in that same afternoon, Catherine offered to draw the Harry Potter characters I needed in my visualization (Figure 7.7). Her illustrations really added another dimension of aesthetic and polish to the project, and I still can't thank Catherine enough for these beautiful icons. (*´∀`*)/

Fig.7.6

Sketch of idea comparing canon and non-canon pairings.

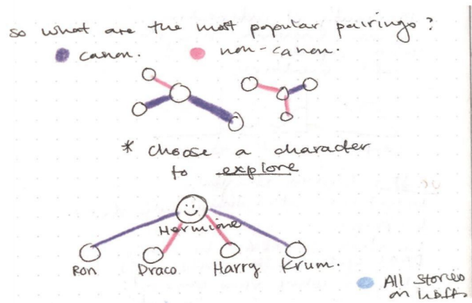
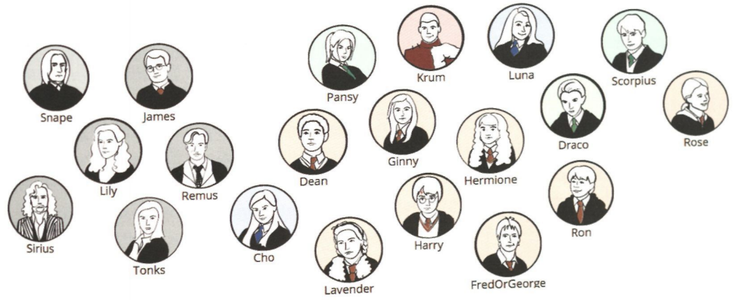


Fig.7.7

Catherine's gorgeous illustrations of Harry Potter characters.



Code

For this project, the hard part was really in the ideation; the coding was pretty straightforward. Once I sketched out the layout I liked, I got back to my timeline and added book release dates, film dates, and GIFs for context:

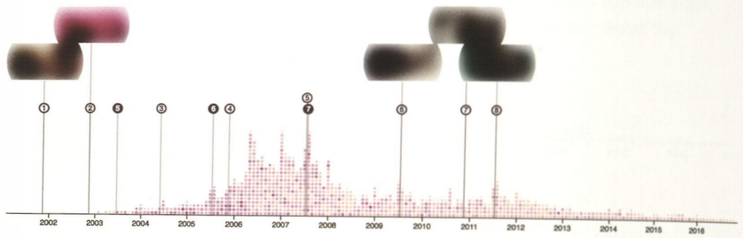


Fig.7.8

Timeline with book dates (filled circle), film dates (white circle), and GIFs for context (blurred for copyright reasons).

And immediately, some interesting things jumped out: fanfiction really spiked in number between the publication of the sixth and seventh books (2005 and 2007 respectively) and also during May 2006 (I still haven't figured out why that month is significant). There were also spikes before and after a book or movie release and spikes during the winter holidays (presumably because people had more time to relax and write stories).

Next, I started looking into how I wanted to depict multiple pairings for a character. I started by overlapping a line chart for each pairing on top of each other. Unfortunately, the lines were very thin and hard to distinguish from each other, so I tried filling them to a baseline to make them appear more solid and hopefully easier to distinguish (Figure 7.9). I used D3.js's *area* module to accomplish this.

`d3.area()` works by passing it a set of points, and it draws a line between all of the points and fills the area below the lines down to the baseline.

But again, the area charts, while pretty, were still hard to read at a glance. I then tried stacking the area charts, which was only slightly easier to read; the pairings with the least stories blended into the others for most of the months. To try and mitigate this, I played around with a few of the curve options in D3.js and I ultimately decided on `d3.curveStep`, which had its downsides (when the lines are too close together, it's hard to tell where one ends and the next starts), but I still liked it the most because of how clean it looked (Figure 7.10).

By default, D3.js connects an array of points with just a straight line, but we can give it additional options to draw the line with different types of curves, which may help distinguish the lines that are closer together

Once the pairings timeline was done, I turned my attention to visualizing the genres for each pairing. I was curious to see if specific pairings leaned more heavily towards specific genres, and it turned out that they did. For example, Ron/Hermione stories tended to lean more towards drama, fluff, and humor, while top Draco/Hermione genres were drama and angst, with two or three times more in the horror/dark genre than all of Hermione's other pairings. James/Lily's top genre was humor, followed closely by drama, and Harry/Ginny's top genres were drama and action/adventure.

To show this, I initially considered displaying the cumulative numbers of stories for each genre, colored and placed according to how "dark and serious" or "light and fluffy" it was. I was also interested in seeing if specific genres rose or fell because of occurrences in the books or films, so I initially had plans to do a timeline along with a horizontal spectrum of the darkness/lightness of genres. I ended up scrapping that idea though, because I couldn't quite figure out how to quantify the darkness/lightness of a genre on a continuous spectrum (how much lighter was humor than drama, for example), and kept only the timeline (Figure 7.11).

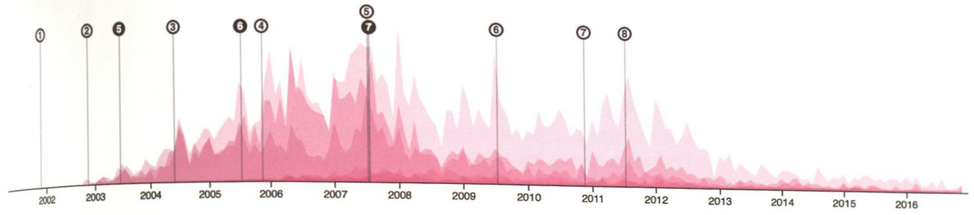


Fig.7.9

Timeline with pairings as area charts overlapping each other.

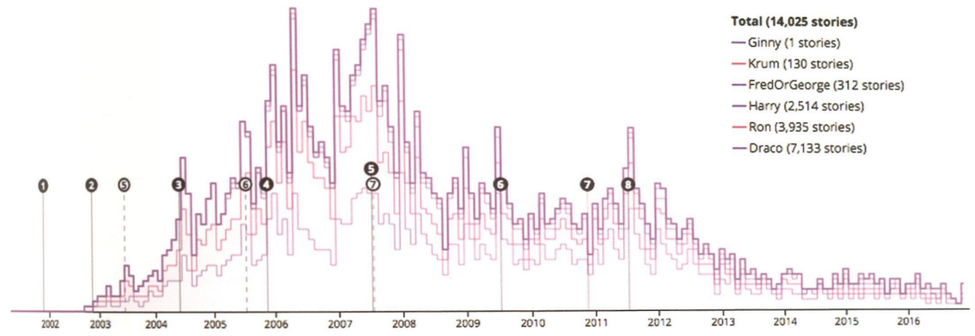


Fig.7.10

Timeline with pairings as area charts stacked on top of each other.

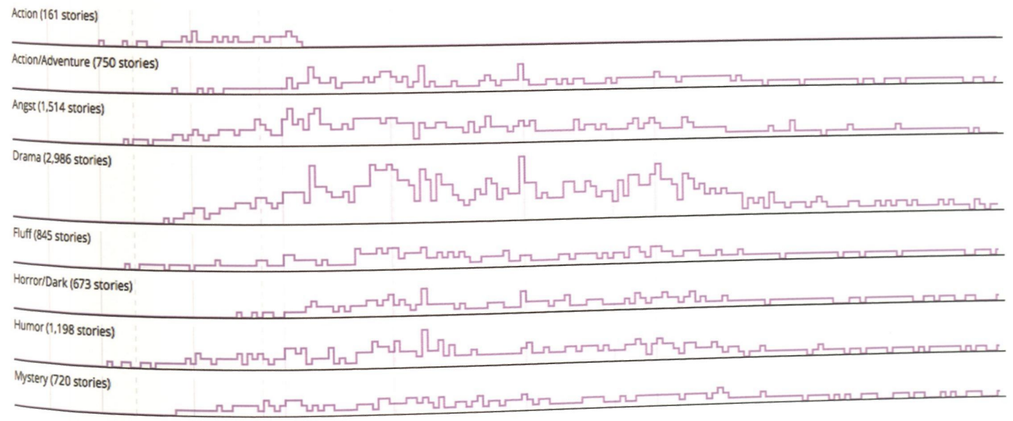


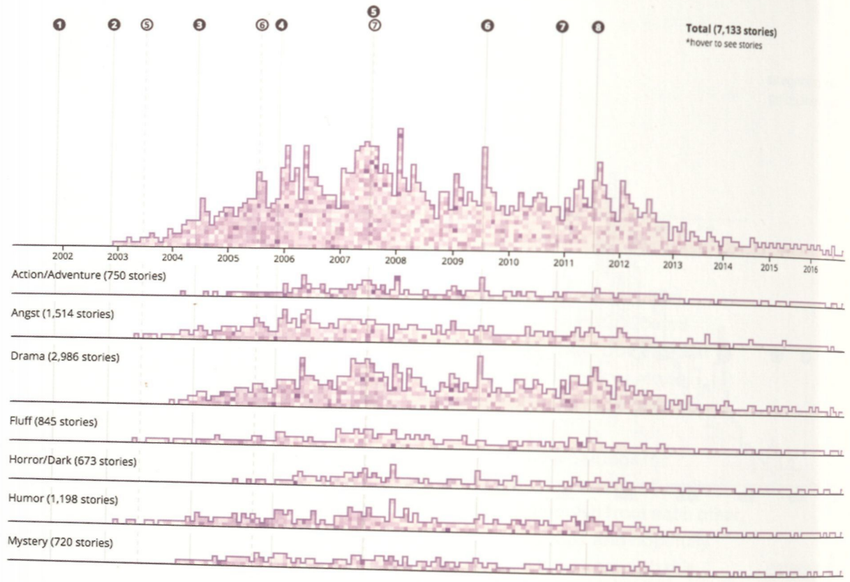
Fig.7.11

Genres organized into timelines.

I lined the genre timelines up with the pairing timeline and used a canvas element to color each block by number of story reviews (I converted my original dots into blocks to fit in the step area chart I ended up creating). I added a legend above the timelines to map the color gradient back to the number of reviews. I also added a hover interaction for each block of stories to show the titles, authors, and number of reviews for the stories within that block:

Fig.7.12

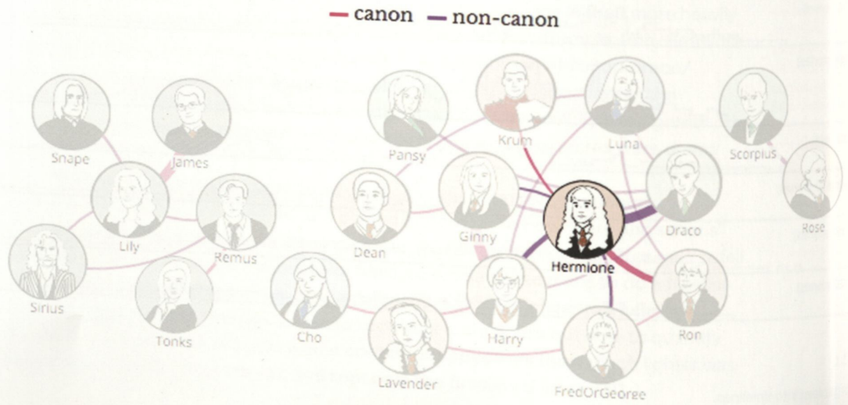
Final visualization for each pairing.



Finally, I put in the graph of characters linked by their pairings and had it double as navigation for selecting and exploring each character (Figure 7.13). It defaults to Hermione, who has the most romance stories with the most suitors. (o≧∀≦)o

Fig.7.13

Graph of the characters, linked by their relationships.



Reflections

I really like the overall look of the project (Catherine's character illustrations really added to the aesthetics of the visualization), but I have mixed feelings about the fact that every visual is some form of a timeline. I really wonder if I should have changed up the visuals to break up the monotony. I'm also dissatisfied that though I was able to eventually see some interesting trends between genres and pairings, that I didn't have the time to annotate those findings.

Despite all that, I'm happiest about the fact that I managed to do this whole piece with gradients of only two colors: pink and purple. Generally, I'm quite over-reliant on color in my visuals and wanted to try something more subtle in comparison, and I feel I was quite successful on that front!

The Most Popular of Them All

↳ shirleywu.studio/projects/hpff

The Most Popular of Them All: a look at fanfiction's favorite ships

BY SHIRLEY WU

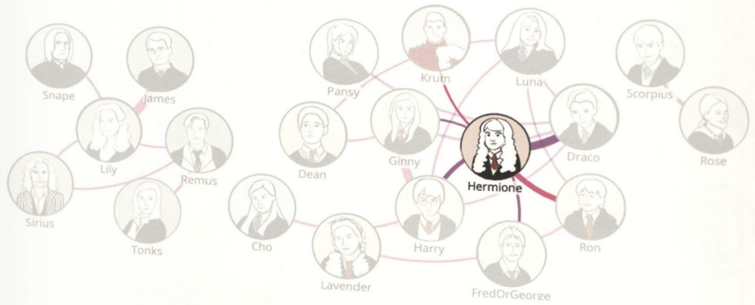
There are many aspects to the *Harry Potter* franchise: books, movies, a new play, studio tours and theme parks. But when it comes to gauging fan reaction, there's probably no better place than fanfiction; with ~760,000 stories on fanfiction.net, ~110,000 on Archive of Our Own, and ~80,000 on Harry Potter Fanfiction - that's a lot of reactions.

For this visualization, I looked into Harry Potter Fanfiction because of its more manageable numbers, meaningful metadata, and (I'll be real) relatively family-friendly content. First interesting find: the number of stories peaked in **2007** at **12,613** stories, held steady at around 7,000 until **2011**, and have been rapidly declining since. Second interesting (though not unexpected) find: at **52,908** stories, **Romance** is the most popular genre - almost twice as popular as the next biggest genre, **Drama (29,429)**.

So of course, I was curious about all the ships: who does the fandom want to see with whom? The top two - **Lily** and **James**, and **Ginny** and **Harry** - are definitely canon, but the third - **Hermione** and **Draco** - is non-canon. (I have this theory that Dramione blew up after Hermione punched Draco in the third movie, but alas, the data is inconclusive.) The most popular leading lady is undoubtedly **Hermione**, with a whopping **six** suitors.

Explore Hermione's stories, or select another character to see theirs:

— canon — non-canon



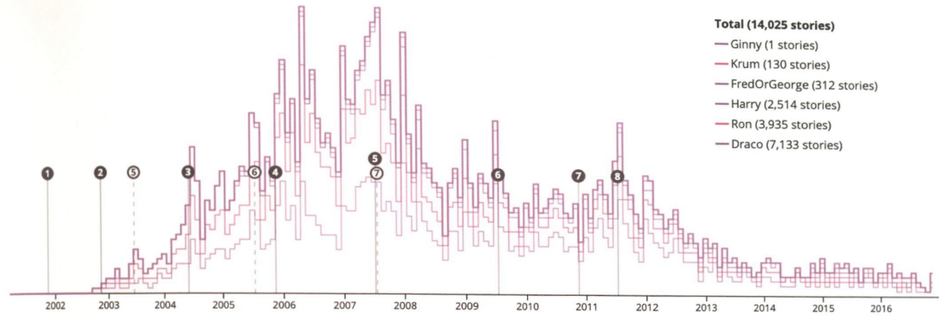
♥♥

Fig.7.14

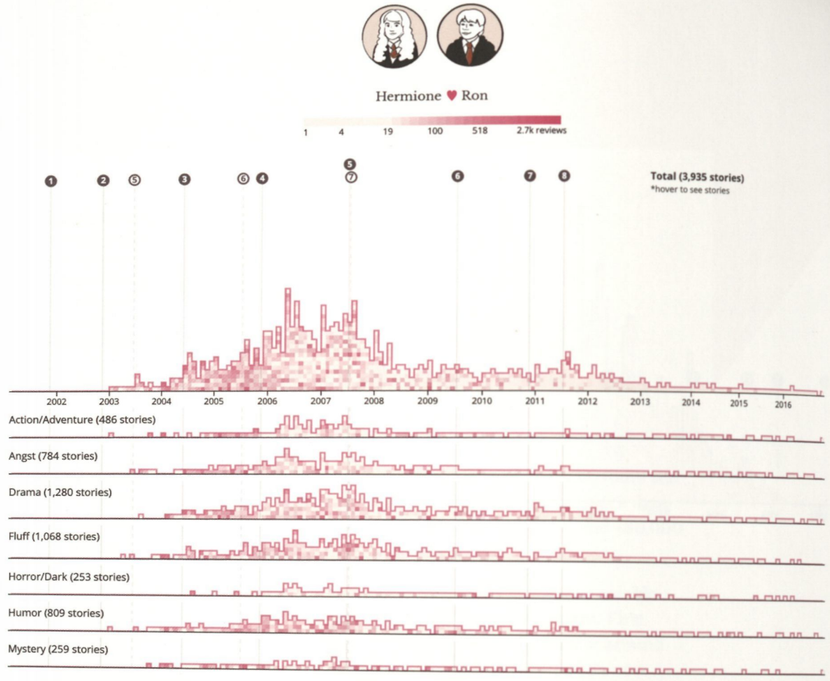
Selecting a character in the graph switches the visualization to center around that character instead.

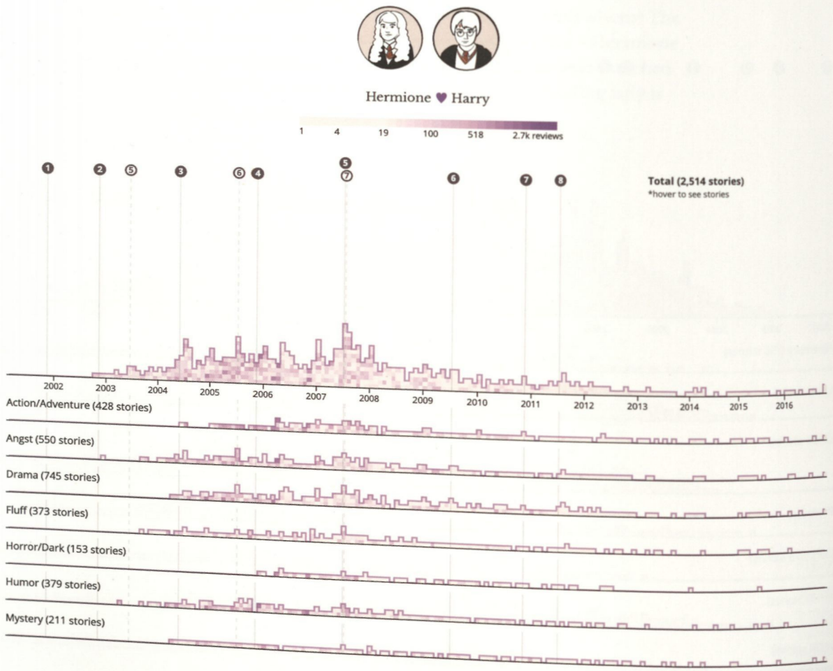
Hermione

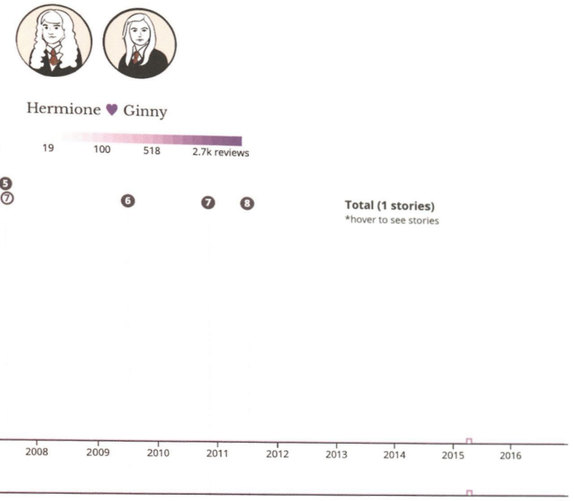
❶ st film ⑤ th book



Hermione ♥ Draco







A tribute

My family immigrated to America when I was 10 years old, and I spoke no English. I studied every day, motivated by the need to understand what was going on around me. My dictionary-reading paid off and half a year later, I was deemed "fluent enough". I was able to read most grade-school books by then, but the Harry Potter series eluded me; I had checked out *Sorcerer's Stone* many times, but found it too difficult every time.

When *Goblet of Fire* came out that summer, my parents came back from Costco with the hardcover. Money was still tight in our household back then, and I had never owned such a beautiful book; I was touched and determined to get through the whole book. When I did, I was ecstatic. I binged the first three books, and was proud when I joined my first Harry Potter conversation.

I have only good memories associated with Harry Potter, from that first hardcover (my parents made it a tradition to buy me every hardcover since), to the 3am screenings with my college floormates, to the full-theatre standing ovation as the credits rolled on the final film. It's given me a whole other world to daydream about, and a sense of belonging in this one.

This year is the 20th anniversary of *Philosopher's Stone*.
With all my love, thank you JKR ♥

many many thank you's to Catherine Madden for
the beautiful illustrations of every character (and of me) 😍

made with 💖 for January: datasketch|es
a monthly collaboration between Nadieh Bremer and Shirley Wu