Studio Bibli

Welcome to the
Process Book of
Team Philippe !

Conception Journey

This whole project started with the crucial choice of the **dataset**. On first sight this may seem easy, but it requires to think about the story that you want to tell and whether the dataset's quality will allow it. Indeed, this course is all about data visualisation but processing big data can be greatly time consuming.

As it was important for us to work with **a subject that we like**, our search ended on a top 3 choice consisting of the huge Steam player dataset, the attracting BeerAdvocate dataset and the very complete BestBooksEver dataset. After some tests on the data and research on the kind of story we could tell, we decided to keep the book dataset. Our choice was also motivated by the fact one can find many websites that recommend their viewers film / music based on what they have already seen or heard. However, only few websites **suggest books** and even less do so in a modern manner.

How we would suggest books was the next challenge to decide. We started by identifying relevant parameters for the choice of a book that could be derived from the dataset. Our focus ended on the **number of pages**, **format**, and **price**. For the suggestion itself, we were not willing give basic suggestions based on book grades because we wanted to encourage discovery rather than pointing towards the already famous readings. We were curious to see if a **similarity metric** derived from the books' summary would also lead to insightful discoveries.

On top of our main book discovering purpose, we also wanted give the user a chance to find out by himself more awesome facts and insights from the literature world! Having additional interactive graphs such as one showing the links between **main genres** and other factors like the ones we derived earlier. Other interesting leads we considered were to create a circular timeline of books publications or highlight differences between the publishers. Indeed, we often don't pay attention to them and they remain quite mysterious. For example, discovering which ones usually publish the top-rated books could be a nice insight. However these last ideas would have required slightly more time!

Book Dataset

- Title
- Series
- Author
- Rating
- Description
- Language
- ISBN
- Genres
- Characters
- BookFormat
- Edition
- Nb pages
- Publisher
- PublishDate
- Awards
- Setting (Where the story take place)
- CoverImg
- Price

Ever experienced that strange emptiness feeling you get upon finishing a particularly great read? When it was so good and you don't know what to do now? Well, next time you have...

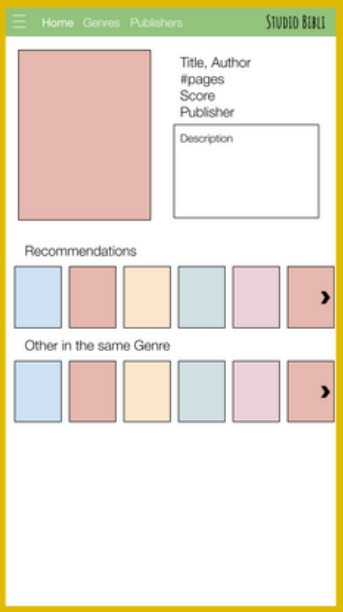
Go on Studio Bibli and find out what to read next!

Project Evolution

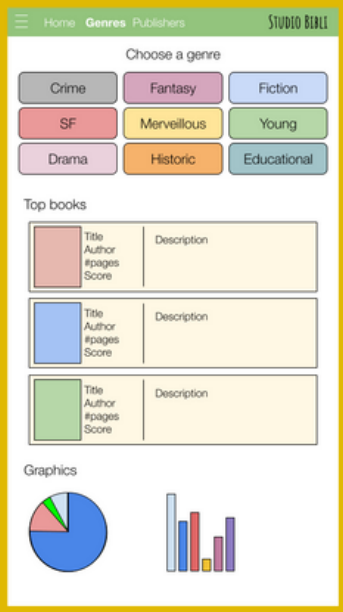
For milestone I, we thought it would be nice to make a kind of **Netflix for books**. The books would have been grouped by genre and you could have clicked on an arrow to scroll through them (page 1). Upon click, it would have displayed the details, recommendation as well as other books of the same genre underneath (page 2). Additional standard visualizations on book statistics could have also been displayed (page 3 and 4).

Drawings of Milestone I

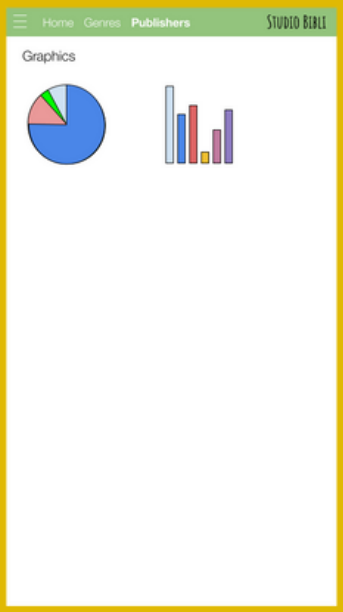
1



2

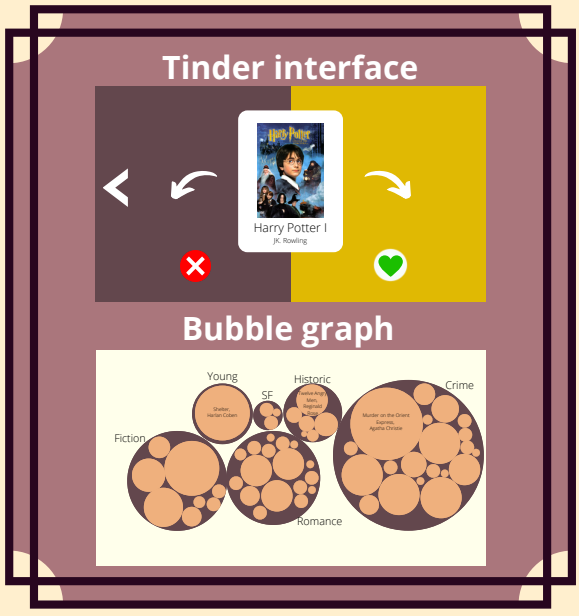


3



4

After some time, the idea of our website matured. We thought that this kind of web architecture was not innovative, and that the visualisation of the recommendations was poor. We wanted to spend more time on making a pretty website instead of mainly working on a big algorithms and data processing.



So we came up at milestone II with the idea of Philippe, a librarian cat that will guide the user through the website, telling him what kind of book he likes with a **Tinder interface** for books, and some widgets to specify the size, the type and the price of the book. After all these actions, the user ended-up on a **Bubble graph**, with all the recommended books grouped by genre.

The other visualisations to see the differences between the genres and between the publishers, that we thought were great to add if we had time would have been implemented just after the bubble graph as Philippe ask the user if he wants to learn more about books.

Tinder Interface

As explained in the previous section, we wanted to have a visualisation so that the user can easily tell Philippe **what books he likes**. To do so, we asked ourselves the following question: what kind of visualisation we are used to that serve the exact same purpose? That is how we ended-up with a tinder-like visualisation.

The main feature of this visualisation is that we have cards that we can **swipe** left of right depending on if we like them or if we don't know or dislike them. We added 2 buttons, a like and a cross button that trigger the swipe animation, to be quicker and to help users that may struggle with the swipe motion. Once a card is swiped, the one behind her move forward.

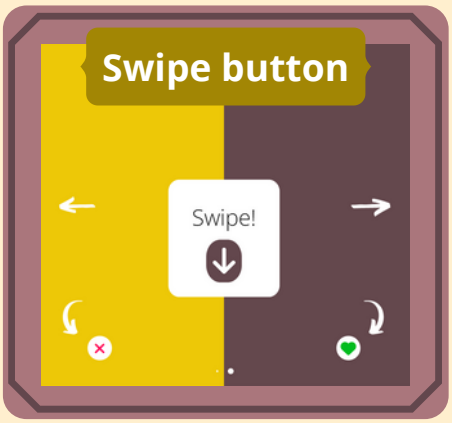
We also added some arrow in the background to help the user visually on what action he can do on this page, and a card behind the book to tell the user when he has finished that he must scroll to resume.

Challenges

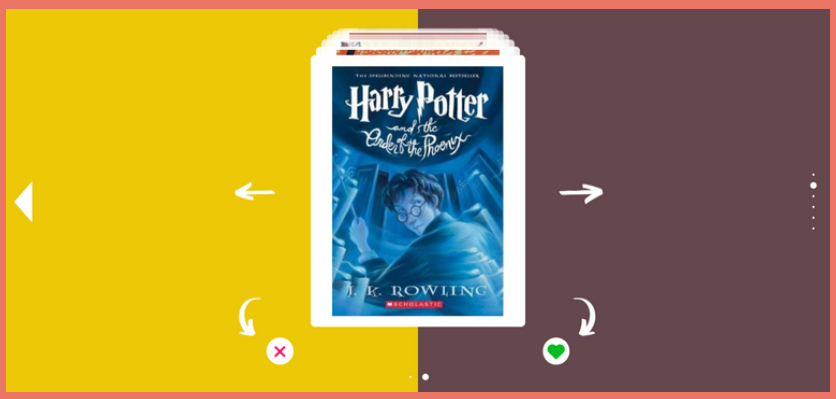
One of our main challenges in this page is to be able to **scale the content** of the window correctly when if the user has a different screen proportion or if he wants to resize his web window. On the right on this page, you can see how the content move when the screen changes. To achieve this goal, we hat to redefine all the objects of the screen according to the size of the screen, except the buttons, since reducing their size would make the navigation of the page more difficult.

Screen deformation

Swipe button



Tinder Interface - usual view



Widgets

Pictures of the widgets

How long do you like your books?

I don't want to deal with more than 50 pages!

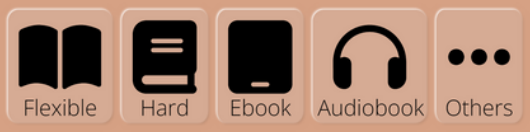
1

How long do you like your books?

Bring me the biggest book you have!

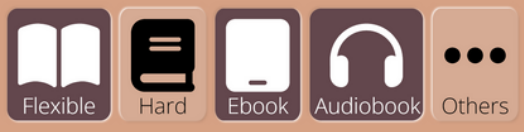
2

Select your favourite book formats



3

Select your favourite book formats



4

How much would you pay for a book?

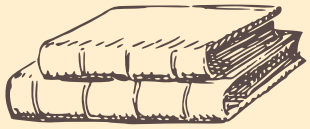
If not super cheap, I'm not in!

5

How much would you pay for a book?

Whatever the price, it will be mine!

6



After getting a better idea of the books the user finds attractive, we wanted to do a fine **filtering** on the number of pages, the type of cover and the cost of the book. These are the type of criteria anybody could have when going to a library.

To make this more fun and follow the storytelling aspect, we wanted to implement this idea through **interactive widgets**. For example, the user can select the maximum number of pages by widening a book (pictures 1 and 2).

To choose the type of cover, we implemented selectable icons (picture 3 and 4) as it seemed to us the most intelligible way for the user to see what he selected.

Finally, to select the price, we presented a stack of coins symbolizing the amount of money you are willing to spend on a book. One can add or remove coins of the pile (pictures 5 and 6). We thought this would be an understandable visualisation for all cultures.

Changes from milestone II

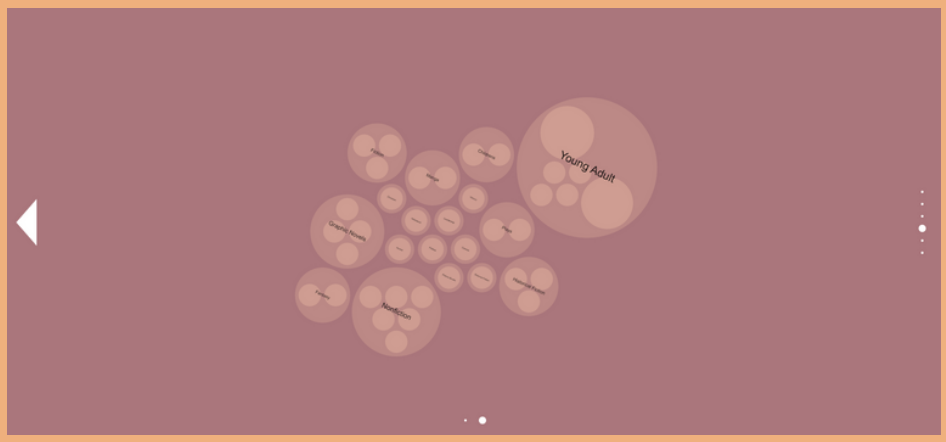
Compared to what initially planned, we did not display Philippe explaining each of the widgets in their page. We did this to simplify and make the general layout cleaner. We replaced the textual explanations with **intuitive layouts tricks** (highlight page on hoover, hollow interactive elements).

Challenges

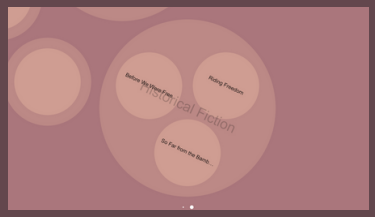
The widgets ended up being way more complex than we first expected. As we were not experienced in CSS layouts and SVG creation, some parts were painful to implement. The widgets are an important aspect of the modern UX we targeted, yet, if we had known in advance, we would probably have gone for simpler ones.

Bubble Graph

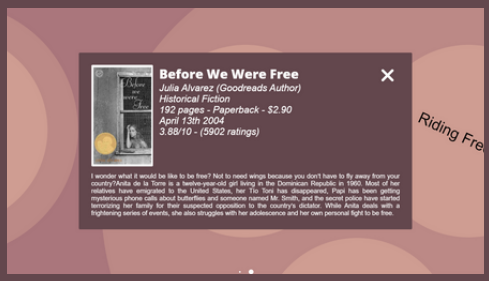
Bubble Graph with few books on display



'Historical Fiction' view



Book selection view



To visualize the suggested books, we wanted to first **group them by genre**, like in a real library. To do so in an interesting way, we thought about putting them into bubbles, one for each genre (over 365 in the dataset). Of course, a book can have several genres but to avoid showing them multiple times, we decided to stick only with the main one of each book. For the user to have smooth navigation, we implemented **multiple zoom levels** on the bubbles so that one can find the books inside even when there are lots. As when wandering in a bookshop, you can select a book if the name appeals to you and read its description and many additional information in the pop-up window. The size of each book bubble is proportional (non-linear scaling) to its **summary similarity** compared to the books liked in the Tinder visualization (self-tuned aggregate rule).

Computing similarities

There were multiple possibilities for computing and displaying book similarities on the graph. We chose to manually pick a small representative set of famous books (the ones in the Tinder) and, for each of them, compute its similarity with all other books. This allowed to keep a much lighter dataset compared to full matrix similarity and give reasonable value to the future similarity aggregate. The similarity itself is computed using term frequency-inverse document frequency (tf-idf) after a complete NLP pipeline of the book descriptions. In the end it revealed better than LSH methods as we did not need to compute similarities at runtime (can afford heavier processing) and we do not aim to provide a top-n but all similarities. I warmly encourage you to check the notebook for more details!

Challenges

Angry Philippe

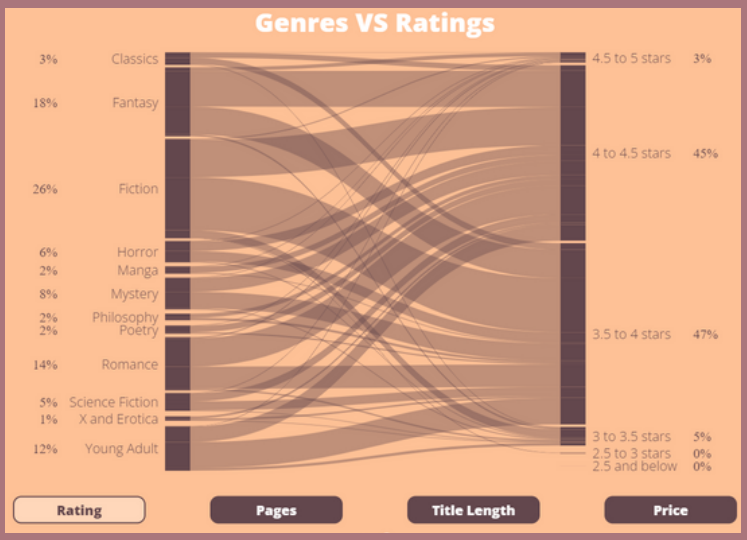
This visualisation presented several challenges. The first one was mathematical as we needed to make the transitions smooth between different zoom level. Then, we had to face the problem of the bubble population. Too many books inside make the titles overlap and clearly deteriorates the navigation experience. To solve this, we capped maximum number of books displayed to 10'000 and tilted the titles as well a cropping the ones too long. When no user selection, we decided to display an angry Phillipe to encourage the user to go back, check out the widgets and at least swipe a book.

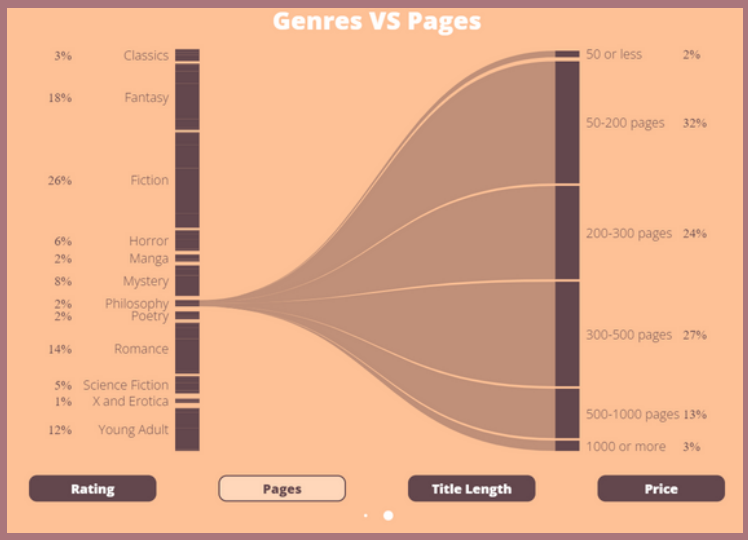
Bipartite Graph

The timings were getting short but we really wanted to bring **more data insights to the users** so that he can learn new things from the website experience. One way to do so while using d3 was to come up with a bipartite graph matching proportions between the 12 main book genres to other key statistics (ratings, pages, price, etc...). With the zooming features, it is possible to dive into particular relations for both sides of the graph. Even without diving, the bipartite graph displays valuable information such as the fact the main genres are not uniformly distributed.

To have a better understanding on what we see, we chose to display the percentages of the different topics shown on the graph. We also added buttons to change easily he graph between the different fields that were "Rating", "nb. Pages", "Title length", and "Price".

Rating Bipartite Graph

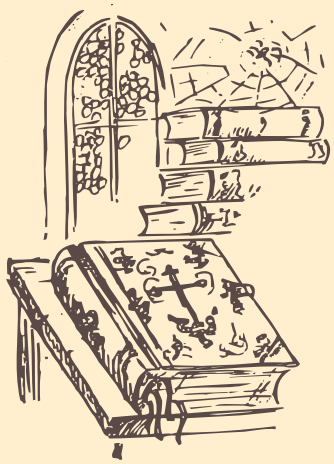




Challenges

We were happy to find a nice library which allows to create bipartite graphs. However, it required an older version of d3 which created conflicts of all sorts. Figuring out a clean way to import both versions was somehow challenging.

Lastly, a small issue was to display the statistics in the right order as the library automatically sorted them alphabetically. This led the labels to be placed all the way around breaking simple visualization logic. We had to find a small workaround by processing the data differently (adding one char before the label, then after using the bp library only display the intended characters).



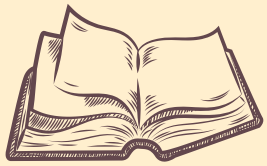
Future improvements

There is always room for improvement in projects. For example, in our case, we can imagine adding sounds on the pages were Philippe is talking to improve the immersive experience of the website. We can also create more images of Philippe to make his mouth move while he is speaking.

For the tinder, we can add a "Pass" button for the books that the user doesn't know, to be able to do a double similarity search: one for the books he like and one for the books he don't like. Like that, the suggestion will even better! Currently, we choose to display only 30 books, but we can imagine to show an infinite number of book (within the limits of the dataset), to have an ever better suggestion.

In terms of widget, we could add new ones to filter the publisher date or the rating for example. Like that, the user can have more impact on the king of suggestion that are given to him.

For the visualisations, we can add more insight of the dataset in terms of publishers. Indeed, for now it is difficult to group them as one publisher can be written in more than 10 different ways, but the information given by them could be very interesting.



Contributions

All team members contributed in each of the milestones and the final work on the website, process book and screencast.

Guillaume

- Data Processing
- Coin widget
- Bipartite graph
- Website navigation
- Process Book
- Readme

Roxane

- Cover buttons
- Tinder page
- Webs-design
- Photoshop png pictures
- Process Book
- Readme

Arnaud

- Data statistics
- Book widget
- Bubble graph
- Website skeleton
- Readme
- Video

Icons made by Flaticon and Canva