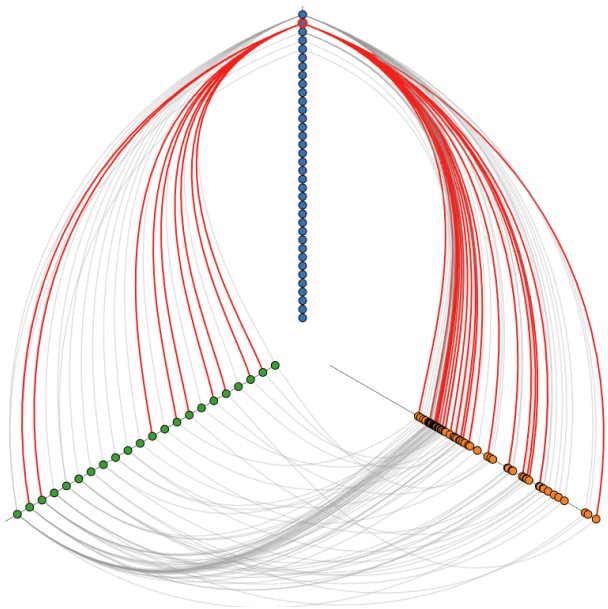
Comédie-Française Exploratory Visualization Tool

Anna Sollazzo, Jack Belford, Ryan Samarajeewa, Thaddaeus Melaku



Abstract—In this paper we will present an exploratory visualization tool for the French theater data of the Comédie Française Registers Project. The presented visualization tool is centered around a hive plot that expresses the theater′s plays, authors, and genres (nodes) and the relationships between them (links). Details about each node can be accessed through individual dashboards, which contain further visualizations spanning a range of unique idioms. The tool was implemented as a single-page web application with heavy usage of D3.js. Evaluation by an expert user yielded valuable feedback regarding as the tool′s major strengths, suggestions to improve content, and user interface. This feedback will be used as the basis for future work.

1 Introduction

La Comédie Française is one of the oldest and most prolific theatres in the world and is still active today. It was originally created by royal decree in 1680, when King Louis XIV forced two rival troupes to merge together by signing an Act of Association. To this day, la Comédie Française is recognized for having its own troupe even though it is uncommon in modern times. The theatre is also very interesting historically as, from its founding to the French Revolution, it kept a very detailed set of its box office attendance records. The troupe maintained these records because this was back when the formal arts were still considered a sinful profession, and in order for the actors to redeem themselves, they needed to prove they were each donating some of the profits to charity.

The Comédie Française Registers Project, started in 2009, is a Digital Humanities effort that is working on the digitization of all of the records. La Bibliothèque-Musée de la Comédie-Française in partnership with collaborators from two Paris universities, Harvard, MIT, and the University of Victoria (UVic) aims to make this data as understandable and available as possible to interested parties.

The records include data about authors, plays, audience members, and sales. They will soon be expanding to also include expenses, casting, and media response.

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

The goal of our project was to create an interactive visualization, with a thread of storytelling throughout, that would at once allow a user to view and understand broad trends in the data, but also extract more detailed information regarding specific facets of the data, through more focused visualizations.

2 Related Work

The most complete visualizations of the Comédie Française Registers Project data are the faceted browser and cross tab browser tools embedded in their website, which were both created by developers at MITs HyperStudio [4].

More of a database search tool than a visualization, the faceted browser [3] presents the user with broad search results and then allows them to refine the returned digital register records according to attributes such as season, author, genre, and day of the week (see Fig. 1).

The cross tab browser [2] targets domain experts, and is a visualization generation tool that allows users to explore the data in more detail, and generate tabular and line chart visualizations according to their specific needs (see Fig. 2).These charts can be made to map trends such as the number of performances or the revenue, based on attributes such as author or genre. Unless otherwise instructed, the chart automatically displays contrasting data for the four most popular authors (of all time).

The Comédie Française data has equally been the focus of two different hackathons at the CFRP conferences in December 2015 (Paris) and May 2016 (Cambridge, MA). Created by interested programmers in collaboration with scholars, these tools, in general , focused on hyper specialized slices of the data, rather than data overview. Additionally, many were static visualizations offering little to no interaction.

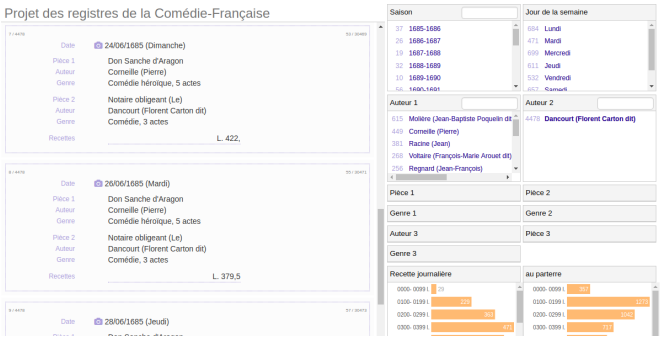


Fig. 1. The CFPR faceted browser

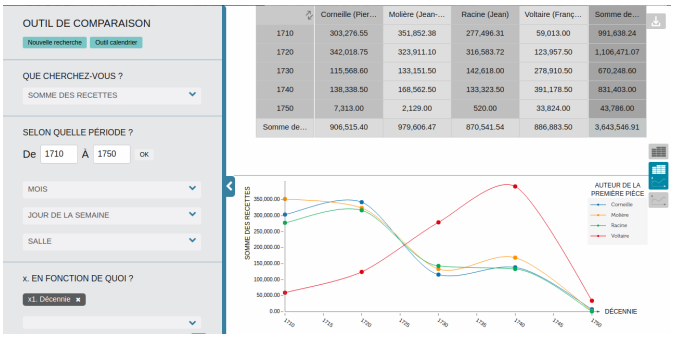


Fig. 2. The CFPR cross tab browser

In his essay *Un exemple de programmation, les registres de la Comédie Française* (1680-1793) (2015) [5] software engineer Frédéric Glorieux presents several summative visualizations which each explore popularity trends in a different facets of the data. Fig. 3 uses a streamgraph to show author popularity over time which also shows the general trends of theatre attendance; note the sudden drop of at the end- that is the disruption caused by the french revolution. Fig. 4 shows the attendance rates for days of the week over time.

Derek Miller also employed several different graph idioms in his visualizations of repertoire (repertoire, defined as a stock of plays) growth and variation over time. In his essay Four Perspectives on the Comédie Française Repertoire (2016) he presents four experimental visualizations , each with the repertoire being grouped by decade. Fig. 5 shows one of his attempts to further segment by play premiere date [9].

Christophe Schuwey and Christopher Morse used an interactive visualization to observe ticket sale patterns over time, as shown in their essay *Visualizing Theatre History* (2016) [10]. Their visualization shows ticket sales, segmented by seating tier, during the 1784-1785 season at the the Odeon-Theatre de l′Europe in Paris. The data is represented in a heat map (Fig. 6) - the hotter the colour, the busier the performance. The visualization includes a slider that allows users to advance time, which aids in spotting economic and cultural trends.

Raphaelle Lapotre created an interactive, dynamic table visualization with the goal of allowing users to, at first glance, understand the richness, density, and complexity of the data in mind. Her essay *Visualiser les donnés de la Comédie Française* (2017) [8], explains that the idiom was inspired by her approach to bibliographic metadata and catalogues.The interactive display, pictured in figure Fig. 7 allows users to tease out relationships in the data, between authors, plays, and seasons.

Hyperstudio also created a prototype visualization to capture the breadth and depth of the CFRP data.Their double axis graph [4] (see Fig. 8), allows users to mutate the visualization through data filtering and the addition of axes. These axes are connected by highlightable links ,in a set-up reminiscent of parallel coordinates, which can be used to infer a variety of relationships in the data, depending on the facets chosen for the axes to represent.

Our visualization aims to combine the ability to view overarching

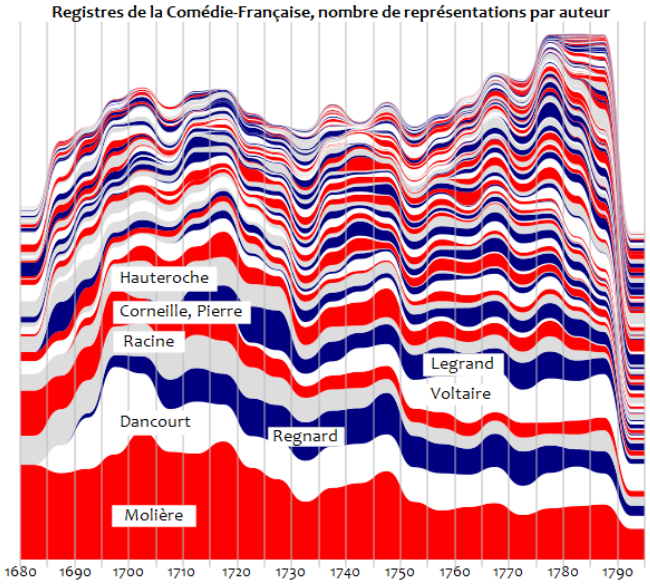


Fig. 3. Author popularity over time (by revenue)

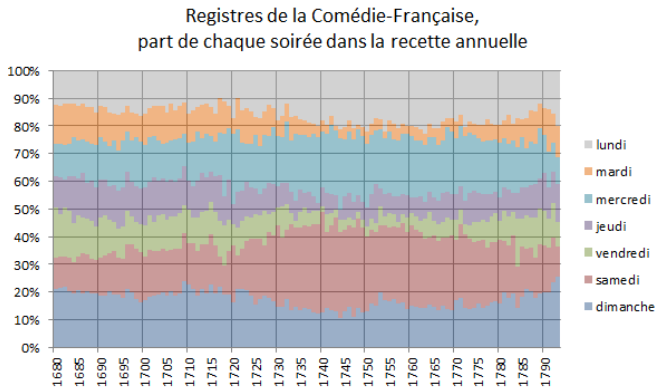


Fig. 4. Attendance patterns over time

themes in the data on a large scale , with more detailed analysis of data subsets pertaining to certain facets of the data.

3 Approach

We chose to implement the visualization in a top-down workflow, in that we first designed and created the key network visualization, the hive plot, before proceeding to implementing the entity specific dashboard views. This section elaborates this approach and each of the idioms involved in the visualization alongside screenshots from the tool.

3.1 Hive Plot

Upon opening the visualization tool, the user is greeted with a full view of a tri-axis hive plot (Fig. 9) accompanied by two dropdown menus containing years between 1680 and 1794.

A hive plot [7] is a rational visualization method for drawing networks. Nodes are mapped to and positioned on radially distributed linear axes this mapping is based on network structural properties. Edges are drawn as curved links.

In the case of our hive plot, green nodes represent authors (e.g. Molière), orange nodes represent plays (e.g. Agamemnon ou La Mort d′Agamemnon), and blue nodes represent genres (e.g. comédie). A curved link between two nodes indicates a unique connection between them. A link between a play and an author indicates that the author wrote that play. A link between a genre and a play indicates that the

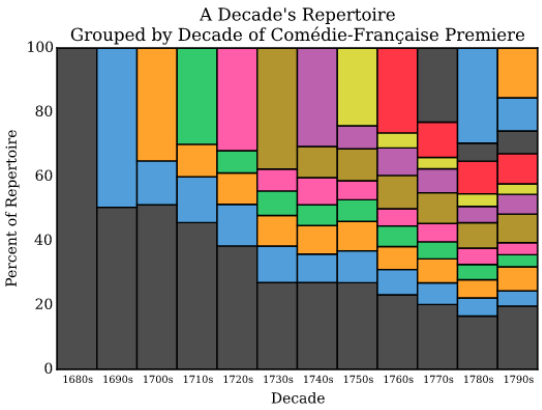


Fig. 5. Repertoire expansion over time

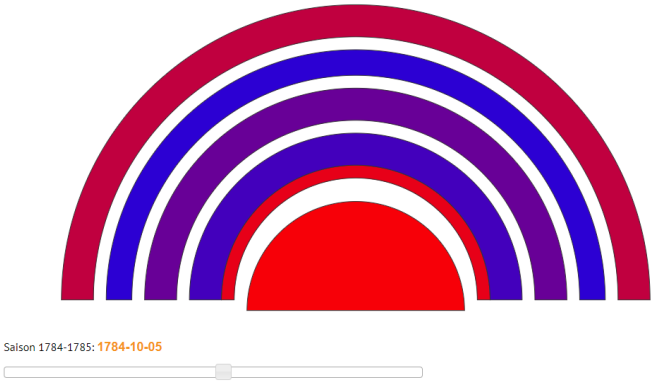


Fig. 6. Seating chart heat map of the theatre during a play in 1784

play is of that genre. A link between an author and a genre indicates that the author has written at least one play of that genre. The green author nodes are sorted along the axis (outer to inner) by the number of plays they had written. The blue genre nodes are sorted by the number of plays that belong to each genre. The orange play nodes are sorted chronologically by each play's premiere date.

The user can interact with this hive plot view in several ways. Firstly, they can hover their cursor over a node to see a label with a node's title (Fig. 10) and highlight its outgoing connections to the other types of nodes.

Hovering over a link highlights it and displays a label indicating the two nodes it connects (Fig. 11).

Selecting a date range from the filter configures the hive plot to display all plays that premiered on any date within that range. Consequently, the hive plot is also configured to display outgoing links from these plays. Widening the date range causes the hive plot to become more dense with more links and play nodes.

Clicking on a node opens its dashboard, as shown in the following sections.

Because the web application is intended to be an exploratory visualization tool, we chose to display a hive plot to provide the user with a high-level overview of the dataset upon entering the application. By doing so, the user has the freedom to browse the nodes or links to find an object of interest, whether it be a play, author, or genre. We also chose to implement a hive plot to emphasise the relationships between each node type. We felt that the tri-axis layout was intuitive for a user to interpret due to its organization of the nodes and its lack of occlusion. We chose to implement a date range filter to offer freedom for browsing plays of a particular time period.

When designing our top-level visualization, we have also considered an alternative idiom: a 3-way tree (Fig. 12).

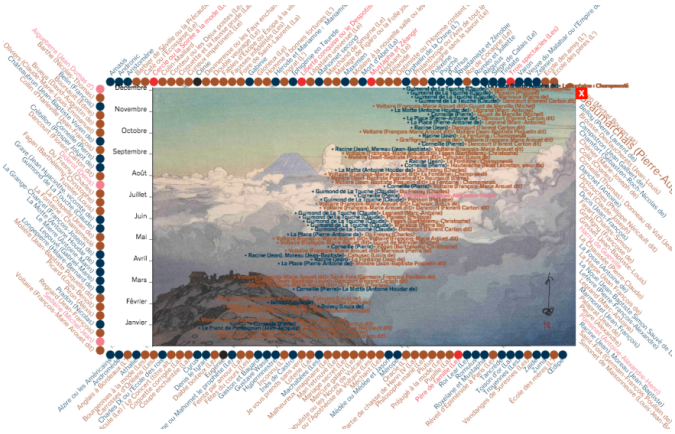


Fig. 7. Lapotres visualization showing prevalence of authors and their plays during each month of a selected year

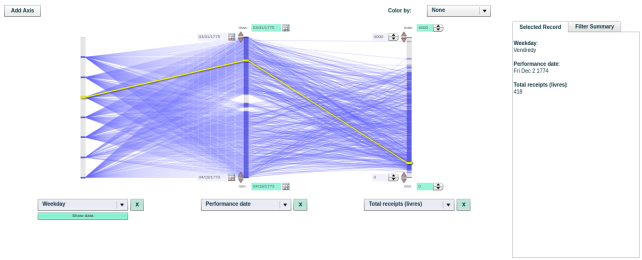


Fig. 8. Hyperstudios double axis graph

This design was inferior to the hive plot to our purpose as it is less scalable with increasing nodes. Implementing this design would require that a slider be implemented to horizontally scroll across the tree. It was also inferior to the hive plot as it would not have been possible to connect genres to authors. The tri-axis layout of the hive plot, however, allowed for this possibility.

The following sections expand upon each dashboard interface (author, genre, play) and the idioms within each.

3.2 Author Dashboard

The top two cards of the Author dashboard (Fig. 13) display number statistics regarding the author. The first number (left) indicates the popularity rank of the author. This value is calculated based on the total number of tickets sold for the plays written by that author.

The second number (right) indicates the total number of plays written by the author.

Below these two a the donut chart illustrating genre distribution (Fig. 14). Each colored slice represents a genre. When a user hovers their cursor over a slice/genre, the open space in the donut chart indicates the number of plays the author has written of that genre.

We chose to include this idiom as it is intuitive for a user to understand distribution of categorical values, in this case genres. It also offers the option for the user to find exactly how many plays within that genre the author has written.

To the right of the donut chart is a bar chart (Fig. 15) illustrating the success of the author's plays, based on the number of tickets sold per play. Only the author's five most successful plays are shown in the chart.

We chose to include this idiom to highlight the author's most recognized works and their success. The bar chart idiom was most appropriate for this purpose, as it explicitly indicates the number of tickets sold per play.

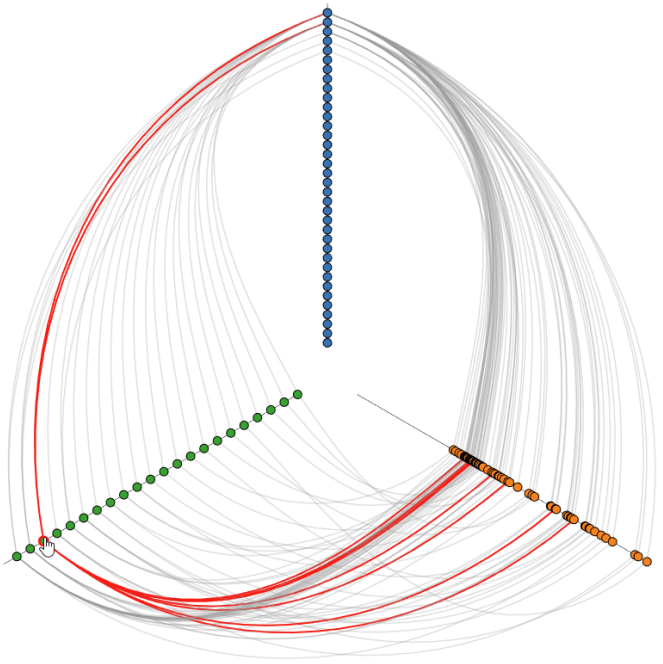


Fig. 9. The tri-axis hive plot, shown as the top level visualization upon entering the visualization tool

Fig. 10. Hovering over a node shows a label indicating its title and its outgoing connections

3.3 Genre Dashboard

The genre dashboard focused on aggregating data about a particular genre over the entirety of the dataset. In this dashboard there are four components that include information about the genre. These are: (1) most popular plays, (2) most popular authors, (3) distribution, and (4) popularity over time. An example of this dashboard can be seen in Fig. 16.

The ranking of popular plays and authors within the genre is based off of the total accumulated revenue of the play or author. The majority of genres were only performed a few times and so our dashboard also only includes the top five plays. However, future work could easily extend this limit.

The visualization of the distribution of total plays with respect to the genre is shown using a pie chart. The goal here is to compare the number of plays in the genre being examined genre to the total population, hence the use of a pie chart to show the part-whole relationship. In the

Fig. 11. Hovering over a link indicates the two nodes it connects.

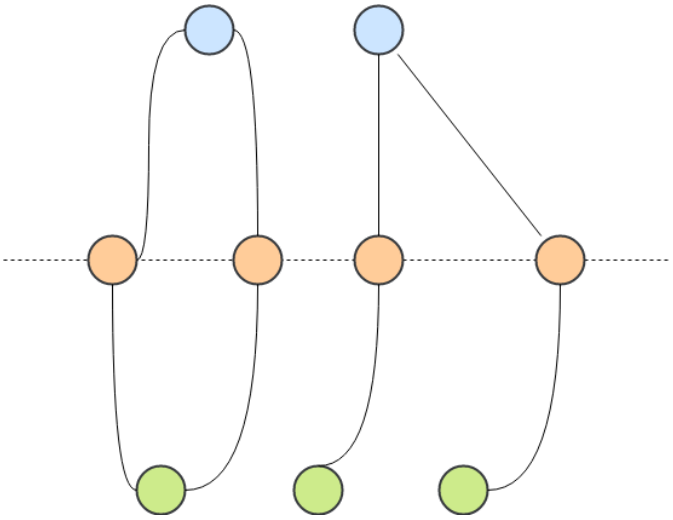


Fig. 12. A 3-way tree connecting genres (blue) to plays (orange), and plays to authors (genres).

Fig. 13. Quick stats in the Author dashboard

example at Fig. 16 we can see that the comédie genre accounts for the majority of plays.

The final component in the dashboard is the popularity over time. This is shown using a simple line chart which immediately allows one to recognize trends. This is helpful for those conducting research that would benefit from cross referencing these trends with outside factors that may have influenced them.

3.4 Play Dashboard

Following the style of the author dashboard, in an effort to create a cohesive look, the play dashboard is also headed up by rankings- this time among plays of the same genre, and by the same author, respectively. These rankings are decided by ticket sale revenue (see Fig. 17).

Following the quick stats section there follow several visualizations examining finer grained details of the data pertaining to the given play.

The first is a pie chart showing the distribution of attendance - measured by ticket sale revenue- across days of the week. We felt a pie chart was an appropriate idiom for displaying this kind of part-whole relationship.This chart could aid researchers in understanding the role a work played in society by examining when it was most popular and inferring what kind of audience may have been present on that day (see Fig. 18).

The next visualization is very specific to the proceedings of la Comédie Française. Most often, an evening theatre session featured two plays, and this bar chart (Fig. 19) allows users to see which five plays were most frequently played alongside the subject of the dashboard. The y axis measures the number of co-occurrences, and the colour channel encodes the genre of the plays displayed.

Finally, we used a grouped line chart (Fig. 20) to display ticket sale trends over time. Its utility is twofold; in addition to showing the general popularity trend over time, the separation of the different types of tickets being sold into separate lines could allow a user to infer the socioeconomic makeup of the audience. The line chart idiom was chosen as it allows for clear examination of trends, and the colour channel is used to distinguish ticket varieties.

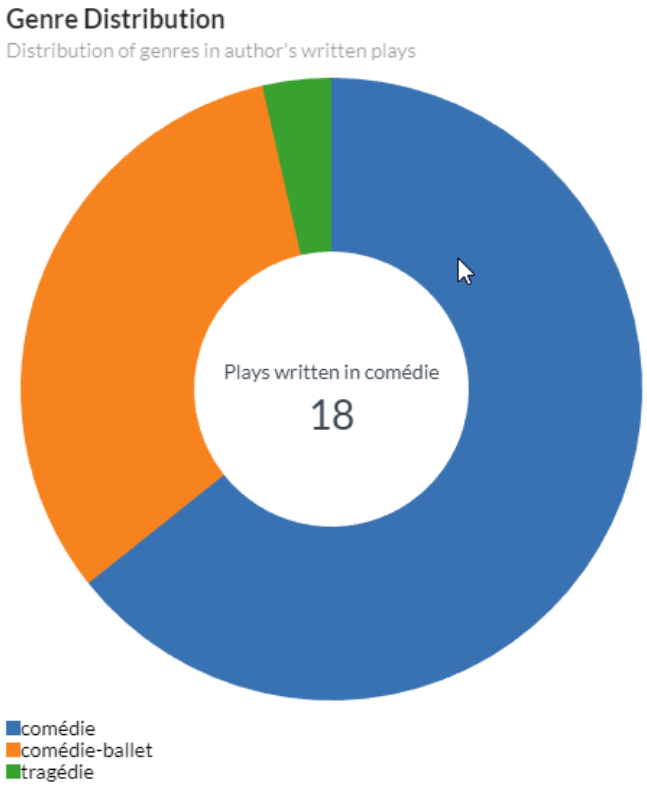


Fig. 14. Genre distribution donut chart in the Author dashboard. Hovering over a slice indicates the number of plays the author had written for that slices genre

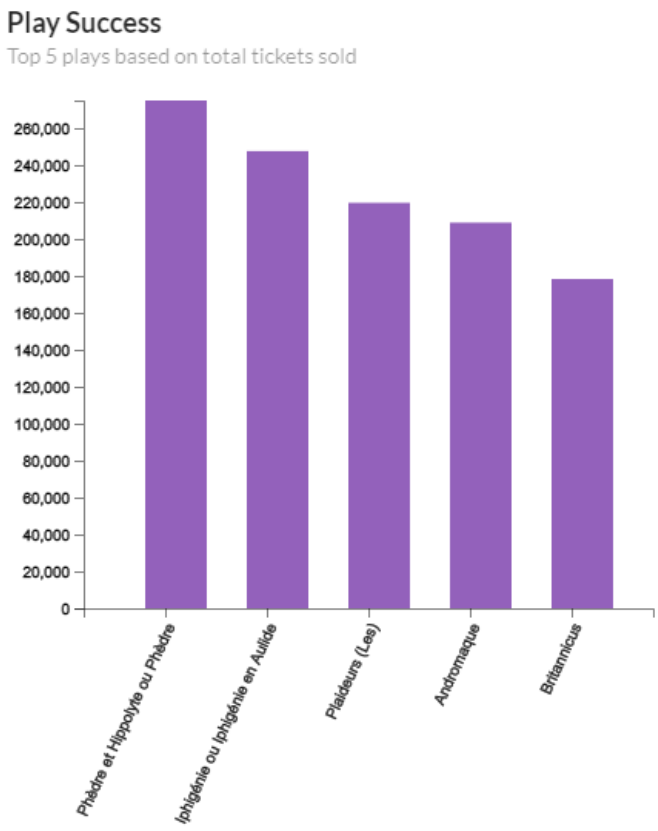


Fig. 15. Play success bar chart in the Author dashboard.

4 Implementation

This section provides a description of the implementation details of the visualization tool. This includes details of the target platform, tools, languages and libraries used. Also described are some of the serious implementation roadblocks we encountered and how we overcame them.

4.1 Technologies

Target platform. The visualization tool was built as a static web application that targets all modern web browsers including, but not limited to, Google Chrome, Mozilla Firefox, Microsoft Edge, Safari, and Opera. During development the web application was run in a Python web server provided by its built-in SimpleHTTPServer module.

Tools. Each member of the team employed their preferred text editors for development, varying between popular editors such as Sublime Text, Visual Studio Code, and Atom. Git was employed for version control. Collaboration took place in a private GitHub repository, which contained all source code and data files.

Languages. Traditional front-end web development languages were used to construct the application: Hypertext Markup Language (HTML), Cascading Stylesheets (CSS), and JavaScript. The HTML and CSS defined the layout and styling of objects in the application. The Javascript code implemented parsing, filtering, reducing, and mapping of the JSON data. It also specifies what events occur when a user clicks or hovers their cursor over certain objects. Additionally, the JavaScript code manipulated the HTML objects to reflect this data with the D3.js library, which is described as follows.

Libraries. We have used three primary libraries to build the web application: D3.js, Semantic UI, and jQuery. D3.js is a JavaScript library for manipulating documents based on data [1]. We took use of the D3.js library to introduce visualization elements such as the hive plot nodes, hive plot links, pie charts, donut charts, line charts, and bar charts. The JavaScript code reads and modifies the given JSON data and maps them to D3.js elements to realize the mentioned visualization idioms. We have also incorporated Semantic UI, a web development framework that provides predefined layouts such as buttons, animations, modals, and stylized text. jQuery [6] is JavaScript library that makes HTML document traversal and manipulation, event handling, animation, and Ajax much simpler with an easy-to-use API that works across a multitude of browsers. We used jQuery to simplify JavaScript development, thus jQuery syntax is abundant across the source code.

4.2 Challenges

An initial roadblock in our development process was the learning curve of the Angular web application stack. Because some of our team members had little to no web application development experience, it was overwhelming to newly build a web application with such a complex stack. To overcome this obstacle, we chose not to build the visualization tool by following this stack. Instead, we decided to build it as a

Fig. 16. The genre dashboard

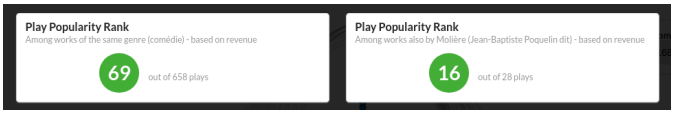


Fig. 17. Play rank cards

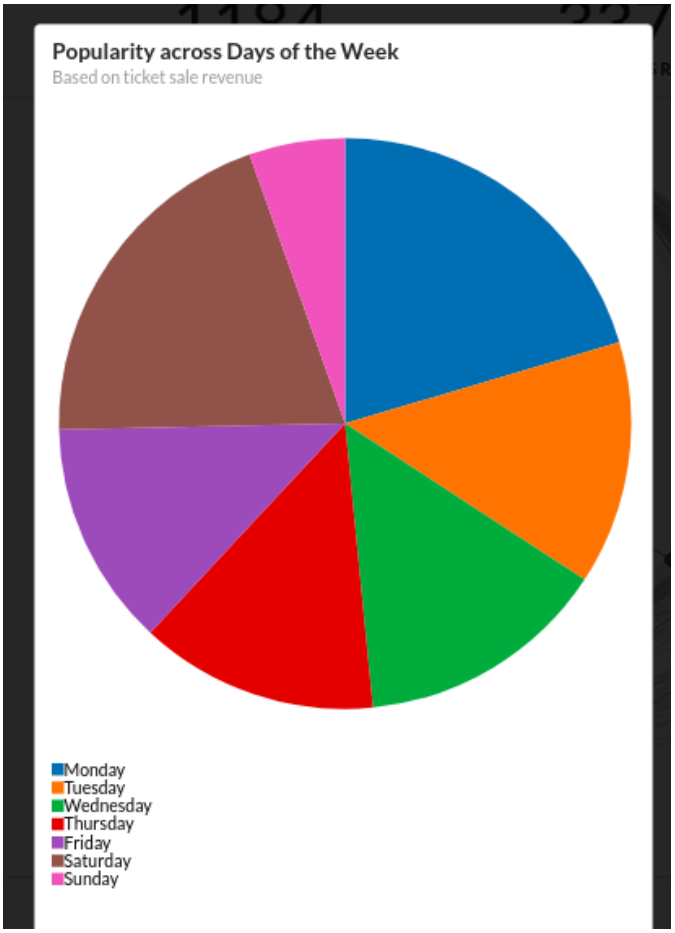


Fig. 18. Attendance distribution across days of the week

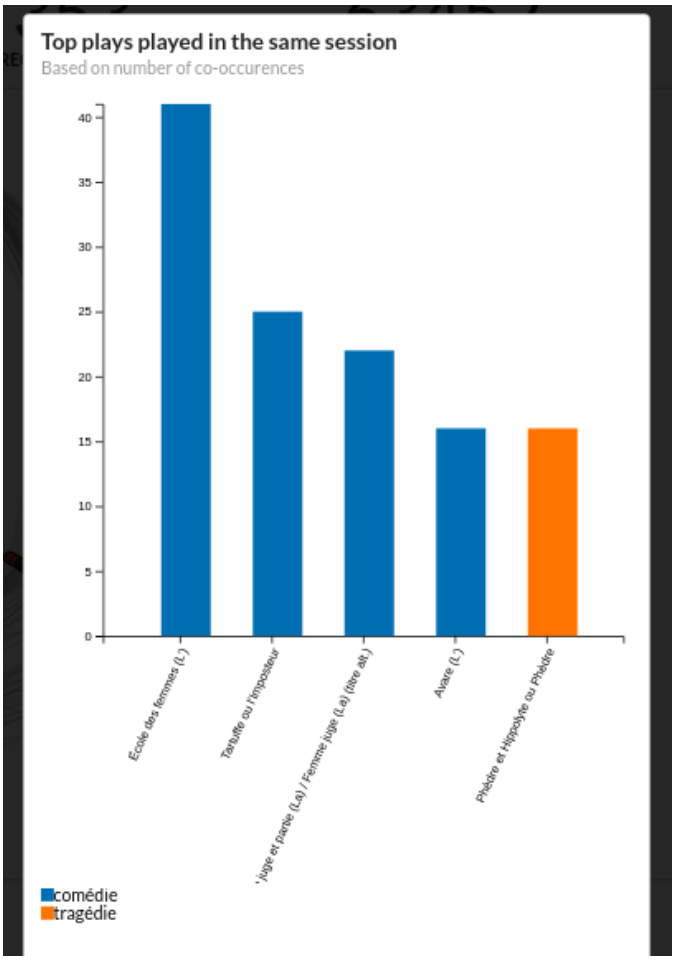


Fig. 19. Co-occurrence bar chart

single-page, static web application only requires traditional JavaScript, HTML and CSS with use of the libraries listed in the previous section.

Although less severe, another roadblock during development was the learning of the D3.js library. All of our team members had no previous development experience using D3.js, thus initial implementations of visualizations such as bar charts, pie charts, donut charts and line charts were challenging. We were able to overcome the learning curve by examining various example implementations, browsing documentation, and forum platforms such as StackOverflow.

The data itself also provided some unique challenges. Being drawn from century old records, it was often incomplete or inconsistently formatted. We minimized this issue by performing data cleaning operations to remove non viable entities before beginning to work with the data. Additionally, not being experts in 17th and 18th century French theatre, understanding the significance of the data required some additional research.

5 Findings

For tool evaluation, our team was fortunate to work with CFRP principal investigator and UVic French department faculty member, Dr. Sara Harvey.

As the visualization is designed to be an exploratory tool, we asked her to click through and explore the tool, and provide feedback regarding clarity to someone outside of software,relevance of data displayed, and general usability. Additionally, we were curious as to how she might use the tool as a researcher and as an educator.

After evaluating the tool for herself, Dr Harvey provided suggestions around use case clarification, overall navigability, and potential future work.

5.1 Scenarios

We identified three primary use-case scenarios that would benefit from this tool. The first is as a tool to quickly browse the dataset in the context of examining the connections between items such as authors, genres and plays. This scenario is more likely to be from those who are inexperienced with the data or who are interested in seeing what the visualization may offer, as opposed to the digital records themselves.

Another potential scenario that was offered by the specialist is as an effective way to visualize the data during presentations of the Comdie-Franaise dataset, perhaps in a pedagogical context. A tool such as this would make it easy to identify interesting pieces of data and present it to an audience that is less experienced.

The third potential scenario for the dataset is as a teaching tool during workshops involving the data. She suggested using it at the beginning and end of an educational workshop to provide users with some understanding of the complexity of the data and overarching trends.

5.2 Performance

The Comédie Française dataset includes millions of entries of information and performing various aggregations on the client-side impacted performance. As such, an early design decision was to perform various

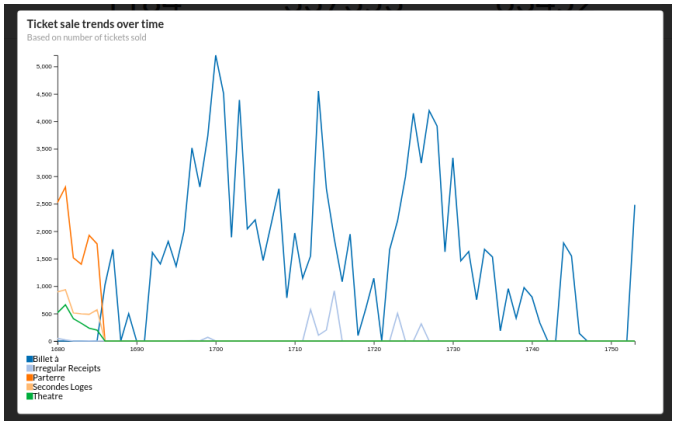


Fig. 20. Ticket sale grouped line chart

preprocessing on the JSON files containing the data to reduce file size and increase accessibility of relevant data. This allowed simpler logic for rendering the visualizations as well as improved performance. However, a great deal of data is still loaded on the client side. In total this amounts to over 100MB of data that would need to be sent to a user who wanted to browse this information online. Future work for this aspect of the application would require further trimming of the data to reduce the bandwidth required by users to run the application. This can be easily accomplished by storing the aggregations that are currently executed in the client and only loading these.

5.3 Feedback

Our team received feedback from both classmates and our domain specialist, Dr. Harvey.Our domain expert appreciated that the visualization′s simplicity makes it an enjoyable environment to explore the data. The visualization works well because it does not show too much information at once, despite some occlusion when larger time slices are examined. It was also appreciated that the visualization presented data that would otherwise be hidden, such as how there are actually over 30 genres of plays despite only a few of them accounting for majority of plays performed.

Some constructive feedback for improving the app was also provided, some of which - such as giving context for the rankings- we have immediately implemented.

One of the most prevalent issues was the lack of visual cues to aid users in understanding data layout. It was suggested that we clarify when the time constraints are applied to the visualizations, and when the data being viewed is comprehensive, and also that we indicate the significance of node position along the axes of the hive plot.

Additionally, as the CFRP is an international effort, it was suggested that it would be useful for our tool to be bilingual.

The most significant issue our visualization faces, however, is occlusion for play and author nodes on their axis; a problem that our expert mentioned was common to many other visualizations of the data. It was suggested to include a magic lens with magnifying glass like functionality that would allow the high-level overview, which effectively communicated the density of the data, to coexist with the ability to fine tune navigation.

6 Discussion

Presented below is a discussion regarding the project overall.

6.1 Strengths and weaknesses

One strength of our approach was the use of a hive plot, as it allows the user to see the connections between genres, plays and authors in one view. This is something that we felt was important, and using a hive plot was an effective way of displaying it.

One weakness of our approach is that when looking at the hive plot, the time constraint plays a major role in the number of links and nodes shown. When constrained to one year the number of links and is manageable, but as the time constraint increases the number of links increases dramatically.

6.2 Lessons learned

As mentioned in the implementation section, we first were faced with having to learn the Angular Web application stack. Because we felt like it would have a steep learning curve we felt like working with the D3.js JavaScript library was a better option. This actually ended up working really well for us, and taught us the lesson that if one approach does not seem feasible it is most likely a good idea to try a different approach

6.3 Future work

Since we have only had a specialist that understands the data use and evaluate out visualization, we think it would be beneficial to have users with little to no domain knowledge evaluate the tool.

As far as additional features, we would like to implement some flavour of solution to the occlusion problem. One possibility for this is adding a magnifying lens embedding, as suggested by our expert. When presented with a dense amount of information it would allow you to focus and understand it by looking at smaller specific areas, while at the same time understanding the magnitude and scale of the information.

Another idea is to add visual cues to aid users in understanding data placement choices. These visual cues would be used to indicate the significance of node arrangement along axis, as well as to clarify when the time constraints are being applied to the data.

The specialist also suggested that we configure our visualization to be fully bilingual. Our specialist was enthusiastic about adding our visualization to the CFRP website, but emphasized that having both French and English would be essential.

Finally, as the project will soon be expanding to include new data about casting, expenses and media reaction, it would be an interesting exercise to see if we could expand our visualization to include these new fields.

References

[1] M. Bostock. *D3.js*. Available: `https://d3js.org/`, Accessed: 2018-04-12.

[2] The Comédie Française Registers Project. *CFPR Cross Tab Browser*. Available: `http://cfregisters.org/en/the-data/faceted-browser`, Accessed: 2018-04-12.

[3] The Comédie Française Registers Project. *CFPR Faceted Browser*. Available: `http://cfregisters.org/en/the-data/faceted-browser`, Accessed: 2018-04-12.

[4] Digital Humanities at MIT. *HyperStudio*. Available: `http://hyperstudio.mit.edu/`, Accessed: 2018-04-12.

[5] F. Glorieux. *Un exemple de programmation, les registres de la Comédie Française*. Available: `https://resultats.hypotheses.org/598`, Accessed: 2018-04-12.

[6] The jQuery Foundation. *jQuery*. Available: `https://jquery.com/`, Accessed: 2018-04-12.

[7] M. Krzywinski. *Hive Plots*. Available: `http://egweb.bcgsc.ca/`, Accessed: 2018-04-12.

[8] R. Lapotre. *Visualiser les donnés de la Comédie Française*. Available: `https://omelettebio.wordpress.com/2017/12/16/visualiser-les-donnees-des-registres-de-la-comedie-francaise/`, Accessed: 2018-04-12.

[9] D. Miller. *Four Perspectives on the Comédie Française Repertoire*. Available: `http://cfregisters.org/cf-repertoire/`, Accessed: 2018-04-12.

[10] C. Schuwey. *Visualisation des Billets Vendus*. Available: `http://vps311427.ovh.net/cfrp-heatmap/odeon.html`, Accessed: 2018-04-12.