Interim Report: Comédie-Française Exploratory Visualization Tool

Anna Sollazzo, Jack Belford, Ryan Samarajeewa, Thaddaeus Melaku

1 Introduction and Motivation

The Comédie Française is one of the oldest and most famous theatres that is still active in the world. It was originally created by a royal decree in 1680 when King Louis XIV forced two rival troupes to merge together by signing an Act of Association. To this day, the Comédie Française is recognized for having its own troupe even though it is uncommon in the modern day. The theatre is also very interesting because from its founding to the French Revolution, the theatre 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. These records were digitized by the Comédie Française Registers Project which primarily studies the history of the Comédie Française. The records include data about authors, plays, audience members, and sales and is the basis for our groups project. The main motivation for visualizing this data is that an effective exploratory tool can allow societal and academic trends throughout this time period to be more easily uncovered. Visualization tools have been created for this data, but our team feels like the visualizations are limited in scope and could be both easier to use and more effective for getting a sense of the data.

The starting point for this project is thus to develop a visualization tool for this database that will allow more effective exploration of this dataset. The benefits of this tool is that it will allow visual learners to navigate the data in a manner that reflects their own thought process. The principle of this visualization is a drill down approach that will echo the thought process of a user in that they will gain further information pertaining to a given database element by further probing that section of the visualization.
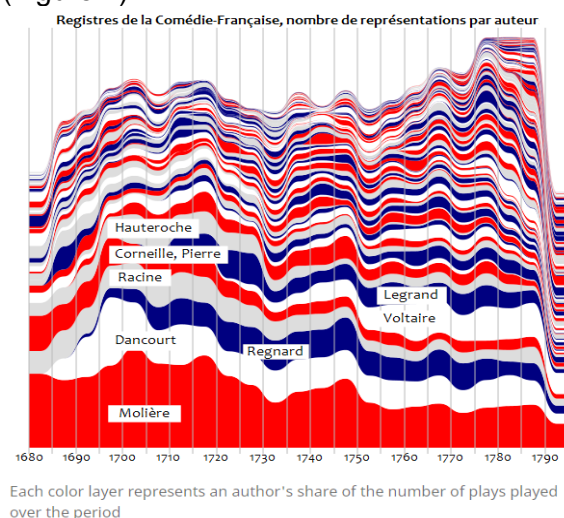
The goal for this project is to develop a visualization that is preferable to users when given the choice between using it and the current exploratory tool that is available on the Comédie Française Register project website.
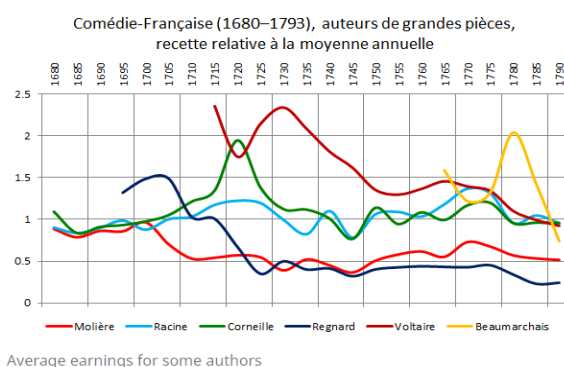
2 Related Work

There have been several visualizations created from the Comédie Française data, though not were exploratory tools, but rather static depictions of specific relationships. The works shown in these articles were sources of inspiration as well as references to sample data visualization projects in the common field. This section presents brief explanations of these articles. Much of these works were created by scholars and programmers during the Comédie-Française Registers Project (CFRP) international conferences in Paris (December 2015) and Cambridge MA (May 2016).

Computer engineer Frédéric Glorieux created several different visualizations in the form of graphs to display author popularity over time, as shown in his article *Un exemple de programmation, les registres de la Comédie-Française (1680-1793)* (2015) [2]. This project uses the same

database collection as the concerning project of this report. Each graph takes on a distinct viewpoint such as number of plays (Figure 1) and average earnings of authors (Figure 2).


Registres de la Comédie-Française, nombre de représentations par auteur

Each color layer represents an author's share of the number of plays played over the period

Figure 1. Each color layer represents and author's share of the number of plays played over the 1600's - 1700's period.

Comédie-Française (1680–1793), auteurs de grandes pièces, recette relative à la moyenne annuelle

Average earnings for some authors

Figure 2. Average earnings of authors.

Derek Miller also employed several different graph idioms in his visualizations of repertoire (repertoire, defined as a stock of plays) growth over time, as shown in his article *Four Perspectives on the Comédie-Française Repertoire* (2016) [3]. His article outlines four of his experimental attempts to visualize the stability and variety of the troupe's repertoire in the time period in question. One such attempt is the illustration of each decade's repertoire, segmented by the decade of each play's premiere date.

A Decade's Repertoire
Grouped by Decade of Comédie-Française Premiere
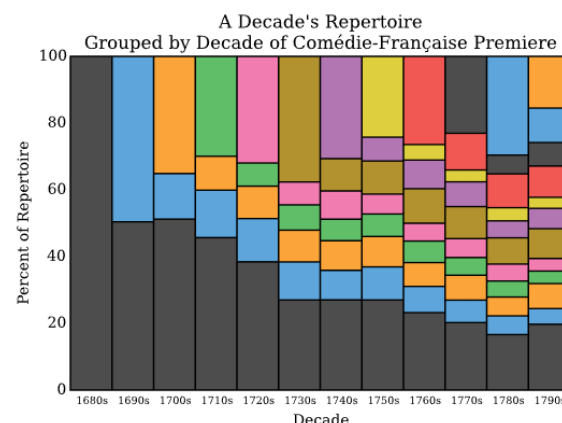
**Figure 3**. A decade's performances by decade of premiere

Christophe Schuwey and Christopher Morse used an interactive visualization to observe ticket sale patterns over time, as shown in their essay *Visualisation des Billets Vendus* (2016) [4]. This interactive visualization depicts the ticket sales (corresponding to seating) to plays at the Odéon-Théâtre de l'Europe in Paris during the 1784-1785 season. The data is represented in a heat map (Figure 4) - the hotter the colour, the busier the performance.
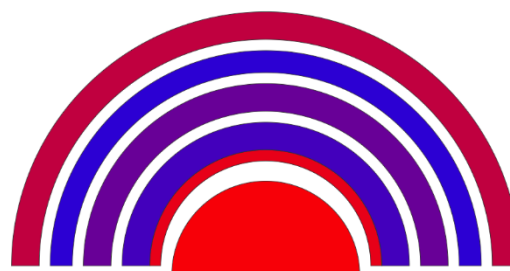
Saison 1784-1785: 1784-10-05

Figure 4. Heat map (seating chart) of the theatre during a play in 1784.

Raphaëlle Lapôtre created an interactive, dynamic table of sorts to allow users to tease out relationships in the data, as shown in his blog post *Visualiser les données des registres de la Comédie-Française* (2016) [5]. The tool was influenced by his approach to bibliographic metadata and catalogs.
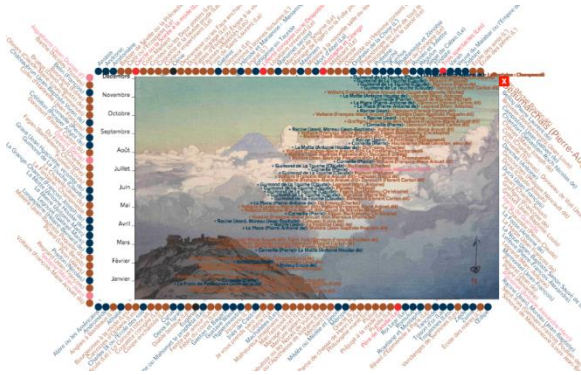
Figure 5. Lapotre's visualization showing prevalence of authors and their plays during each month of a selected year.

We also aim to find data exploration tools (whether within a similar field or outside) further in the project to guide our interface design.

3 Approach and Milestones

We have chosen to implement the visualization in a 'top-down' workflow, first creating the key network visualization before moving on to the entity specific elements. This section elaborates this approach and sets a timeline with milestones for the design project.

The key module of this visualization is the network graph. This network graph takes the form of a 3-way tree layout which connects authors to plays to genres, organized by season. Each node in the tree can be selected to display further details regarding that item in the form of a dashboard. For instance, if a play is selected the visualization tool will activate a popover containing further details, possibly with further visualizations, regarding that play. The implementation of this graph is described in the next section. As shown in Table 1, we plan to prioritize this work and deliver the network graph as a minimum viable product (MVP) by only constructing a basic HTML and JavaScript implementation.

It is not planned that the detailed dashboards are implemented in the MVP.

The dashboard visualization is prioritized after the MVP as it adds significant interactivity to the solution. This dashboard may include pie charts, bar graphs, and other graphs alongside annotations and text to provide further details regarding a node. The motivation behind the interactive dashboard is that the user would learn more about a specific item or determine previously unknown questions.

Finally, we plan to implement filtering functionality as the last technical deliverable of the visualization project. The user would be able to filter the network graph and its nodes based on season, author, or genre to suit their interests. A project report is to be finalized as the last deliverable of the project.

Date	Expected Work Completed
March 12	Complete MVP JavaScript network graph
March 22	Complete dashboard visualizations for specific nodes
April 1	Complete filtering functionality
April 12	Finalize project report

Table 1. Project schedule

4 Implementation and Technology Choice

The visualization that has been discussed is to be built using the Data-Driven Documents JavaScript library (D3.js). The benefits of this library is that many of the difficult aspects of SVG rendering for web apps can be abstracted away and allow our group to focus on purely design. Currently the team is focused on the building of the graph component of the application. This component will be a D3 force-directed graph. The current idea is that authors of plays will be displayed in a top row of

nodes. These authors will be connected to another row of nodes which would be plays. The plays would then be connected to a third set of nodes which would be genres. The nodes that are currently displayed will be filtered by the timespan that can be configured by the user. For example, the user can specify to view only data relevant to the years 1685-1690 if they so choose. Each of these nodes in the graph will also be clickable by the user. Clicking a node will bring up a modal screen that will show statistics about the data. The minimum viable product for our design is to be able to navigate the data using the graph visualization and click on nodes to see more information about them. Ideally our group will also have time to expand on this model, including more visualizations within the statistics for particular nodes. These are mostly to be chart visualizations for aggregated data that is related to the data. The implementation of the statistical portions of this visualization will be influenced by the tools seen in the Comédie-Française Register Project as the main problem our visualization is attempting to address is the exploratory aspect of this dataset. Other expansions of this project could include the configuration of nodes that will be displayed in the graph. That is, the graph will not just be limited to authors, plays, and genres but also allow other features. At present, our team prefers to prioritize simple configuration to ensure the tool is easy to use. Another technology we are using for building this application is the Angular framework. This framework allows our team to break up our application into smaller moving parts to allow easier collaboration on the codebase for our project. This is accomplished by utilizing the component nature of Angular and splitting up the functionality between a few components and service classes.

5 Proposed Evaluation Steps

We hope to test the utility of the visualization on a small set of users, some of whom have prior knowledge of the domain, and some who do not. Since the aim of the visualization is twofold, so too will be the evaluation.

We plan to first give the users a set of questions pertaining to the data and record the time it takes them to answer the questions. An example of a type of question we may ask is to find the number of plays that a specific actor was a part of. Another question could be to find what genre of play was the most popular during this time period. By looking at the results of things like, how long it takes users to complete tasks compared to what the expected time is, or what tasks tended to take the longest to complete will help us in evaluated the tool that we have created.

The second set of questions would be more exploratory rather than quantitative, and after giving the users time to explore with and without the visualization, we will interview them about which tool they felt served them better and why. With the feedback we get from the users, it should help us find the strengths and weaknesses of our tool. One question that we might ask is which tool they found was easier to use to complete the tasks. Another question we could ask the user what they to be difficult or unclear about the tool, or if they struggled with something at first but eventually figured it out.

6 Risks

Some of the risks that we may face is dealing with incomplete data. For example, in the data set for actors there are a number of which, though they all have their own unique actor id, many do not have the info for a first name or even a last name. They may only have a pseudonym.This may be come up as an issue, when wanting to visualize the data and finding a way to identify actors in a more meaningful way

than simply their actorid. Also, this may lead to difficulty in extracting relationships since there is a lot of variation in how data within the same field is represented.

Another risk is the lack of knowledge about the field may be an obstacle in distinguishing the things are more significant from those that are less significant. One thing that we are definitely considering is consulting an expert. This would allow us to gain some perspective from someone that is familiar with this type of data and will be able to interpret it and give us some insight. They will be able distinguish the categories that we should focus on, and highlight the more important relationships that should be shown. Talking to an expert may also allow us to focus our energy more effectively, especially since there is a large amount of data and we are relatively new to the field.

References

[1] http://cfregisters.org/en/project-history/about-the-com%C3%A9die-fran%C3%A7aise
[2] http://resultats.hypotheses.org/598
[3] http://cfregisters.org/cf-repertoire/
[4] http://darthcrimson.org/visualisation-des-billets-vendus/
[5] https://omelettebio.wordpress.com/2017/12/16/visualiser-les-donnees-des-registres-de-la-comedie-francaise/