

“Visualisation of censorship in literature around the world”

Candidate number: 1544586

I declare that, except where otherwise indicated, this mini-project is entirely my own work, and that it has not been previously submitted and/or assessed and is not due to be submitted in its entirety or in part for any other course, module or assignment.

1. Overview

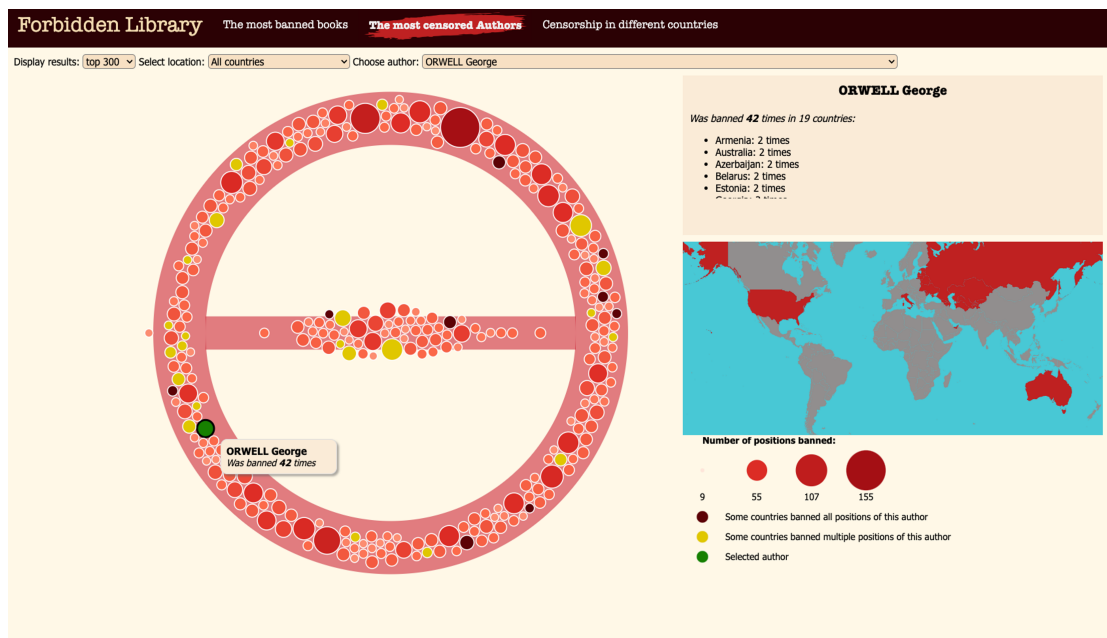


Figure 1.1: Teaser image

From ancient book burnings to modern-day novel bans, censorship has existed in the world of literature for centuries. While the first recorded book ban dates back to 1517 [8], censorship of literature persists today. In the US alone, 2,532 books were banned in 2021 [16]. Governments, religious institutions, and individuals can all seek to suppress ideas and perspectives that challenge their beliefs. Yet tracking the extent and impact of literary censorship can be a daunting task. Data on banned books is often incomplete or limited to individual countries [2, 11, 12, 14, 16, 18], making it difficult to gain a comprehensive understanding of the issue.

My project is, to the best of my knowledge, the first attempt to visualise censorship in literature worldwide. The website allows users to explore which authors and books have been banned throughout history and compare censorship trends across different countries. The visualisation can be used by literature or history lovers to explore and find interesting titles to read, as well as university students who would like to search for particular authors or check the censorship in a given country over the years.

2. Data

The data has been taken from the Die Kasseler Liste database (<https://www.kasselerliste.com/die-kasseler-liste/>). The dataset has been created and updated manually by many researchers. Each of more than 120 000 entries contains the "source" value indicating where the information has been taken from. It is also, to the best of my knowledge, the only dataset that includes data about censorship on a large, worldwide scale and contains detailed information about location and time of the bans. Therefore, I considered it the most valuable and reliable source for my visualisation.

2.1 Data overview and preprocessing

The dataset type is a flat table - there is one item per row, each column is a different attribute, and each cell holds value for an item-attribute pair. There are 120 309 data items defined by seven attributes: author, title, place, date of publication, location, year of the ban, year of the ban lifted, and source. Since all the data has been entered manually by multiple people, I focused on filtering incorrect or outdated values. There were 578 duplicated values that I removed. Table 2.1 shows the high-level overview of the properties used in the visualisation. Below I describe the pre-processing actions I took for each of them.

	Location	Author	Title	Year of the ban
Type	Categorical	Categorical	Categorical	Quantitative
Original cardinality	110 values	57 017 values	113 625 values	423 values, from 0 until 2022
Cardinality after processing	100 values	57 003 values	113 586 values	419 values, from 1517 until 2022

Table 2.1: Overview of the properties from Die Kasseler Liste dataset used in the project

2.1.1 Location

- I checked all the country codes (5 of them were incorrect), and I replaced all of them with the country names that also corresponded to the country names from the *countries – 50m.json* file that has been provided to us in Problem Sheet 4,
- I made sure that location values specified one country per item,
- There were many entries for countries that do not exist anymore (East Germany, Western occupation zone, Soviet Union, Yugoslavia). Since I wanted to show the accurate geographical position of each ban, I decided to replace them with the countries that currently exist on their territory.

- The Bible's location has been set to "various"; hence I replaced it with the 52 countries taken from the LovePackages page (<https://lovepackages.org/bibles-dangerous-illegal-covert/>).
- When the location specified the region of the country, I removed it to only display the accurate country location (e.g. Texas has been changed to the United States of America).

2.1.2 Title

- I changed: "all works except", "all works, vol." and "all works from" to "multiple works",
- I changed "all works" and "all works of and about" to "all works",
- Unfortunately, a big part of the data values was in German. I translated some of the most banned volumes, but in the end, I decided that users can translate the interesting titles themselves if they want to.

I noticed that some authors have "all works" or "multiple works" banned, and they only appear in the dataset once (even though they refer to multiple titles); therefore, I decided to highlight these authors differently in my visualisation.

2.1.3 Author

- I unified the spelling of some authors (e.g. Dostoyewski → Dostojewski).

2.1.4 Year of the ban

- I removed the invalid years (0, 1, 939, 193).

2.2 Data processing

To make the visualisation faster and more efficient, I decided to further process data before using it. I generated 3 datasets using Excel and Python, that I used for different views.

2.2.1 authors.csv

For performance reasons, I decided not to use all the 120,000 values in the visualisation; but only the top banned authors. The generated dataset contains data about the 1000 most banned authors. Its attributes are: author, count, all_works_forbidden, multiple_works_forbidden, and 100 locations (each corresponding to a different country). The author's field is described in Table 2.1. The remaining fields are analysed below.

	count	all_works _forbidden	multiple_works _forbidden	location attributes
Type	Quantitative	Categorical	Categorical	Quantitative
Description	No of positions banned worldwide	Whether any country forbid all works of this author	Whether any country forbid multiple works of this author	No of bans of that author in a given country
Cardinality	9-155	True, False	True, False	empty, 1, 2

Table 2.2: Overview of the properties in *authors.csv*

2.2.2 books.csv

I considered using only the top banned positions, but I wanted to accurately display all the banned titles for any selected author; therefore, I used all the data points. The attributes in this dataset are: author, title, count and 100 location columns. The only significant difference from the other dataset is the changed cardinality of count which now ranges from 1 to 52.

2.2.3 countries.csv

This dataset contains information about the number of banned books in each country over the years. It consists of: country, count and 419 attributes for each specified year.

	country	count	year attributes
Type	Categorical	Quantitative	Quantitative
Description	Country name	Number of all book bans in that country	Number of book bans in that year
Cardinality	100 unique values	0-34369	0-4782

Table 2.3: Overview of the properties in *countries.csv*

3. Goals and Tasks

The visualisation should be enjoyable and useful to multiple groups of users - those with no knowledge of censorship or literature, book lovers, and people who already know about censorship. Therefore, users should be able to explore the most censored books (explore extremes) to familiarise themselves with the topic. They should also be able to compare trends easily - for example, to compare the location and number of bans between different titles. The same applies to the most censored authors - users can explore extremes and compare trends to learn about the authors and to be able to find similarities between them. If users are already familiar with the topic of censorship in literature, they could use the visualisation to search where a given author or book has been banned (locate outliers). Finally, the visualisation should also provide a high-level overview of the topic. Therefore, users could compare the censorship in different countries - find countries with the most severe censorship or periods with the highest number of book bans (compare trends). Therefore, the four main goals for the visualisations are as follows:

1. Explore the most censored books,
2. Learn about the most censored authors,
3. Search where the given author or book has been banned,
4. Compare the censorship in different countries.

4. Visualisation

I wanted the visualisation to be engaging, intuitive and accessible for every user.

4.1 Accessibility

Before describing each view, I list the most important high-level rules I followed when designing and implementing the visualisation to ensure it is accessible to as many users as possible.

4.1.1 Accessible colour palette

I ensured that the visualisation was clear for people with any kind of colour blindness. I used the reds palette from ColorBrewer2 [4], and I created an 8-colour palette from the D3 schemeCategory10 [5]. These colour schemes were designed to be accessible, but I double-checked them with Coblis — Color Blindness Simulator [3]. Figure 4.1 shows what the colours in the visualisation look like for Red-Blind users.

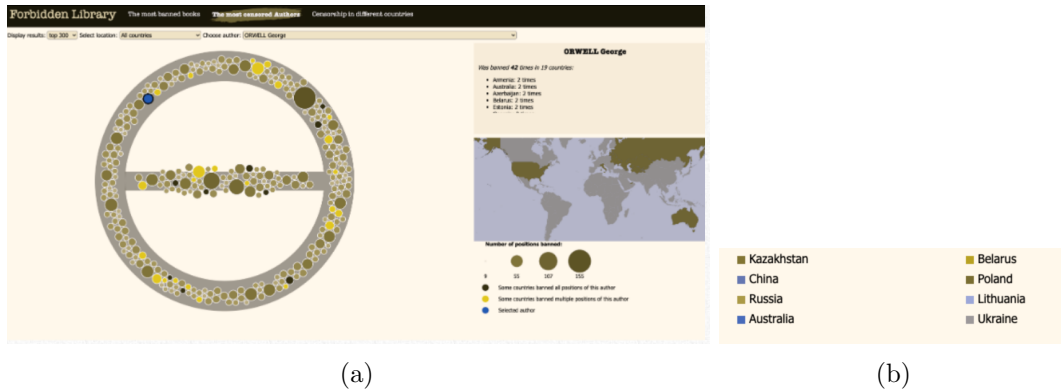


Figure 4.1: Screenshots of the visualisation checked in Coblis [3]

4.1.2 Readable text

When displaying text, I ensured that the contrast ratio between text and background meets the AAA WCAG Accessibility Criteria [10, 21]. I also used readable and web-safe fonts [15] to ensure the visualisation will look the same in every browser.

4.1.3 Accessible display of data

The common rule in most accessibility tutorials is that the colour on the graph should not be the only visual cue [9, 10, 13]. I followed it by adding pattern, opacity or border-width changes, on top of the colour.

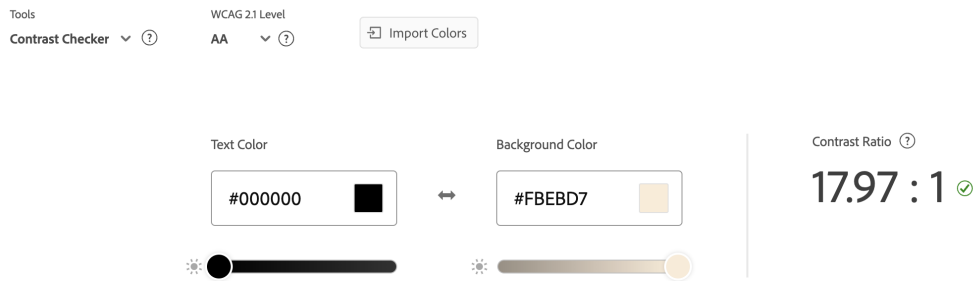


Figure 4.2: Adobe Color Contrast Ratio [1]
check of the text and background colours

4.1.4 Macro white space

According to keen.io [13], macro white space allows the brain to process information faster, especially for those with spatial reasoning difficulties. To follow it, I added a small padding to each item in the visualisation.

4.1.5 Understandable visualisation

I added labels and legends to make my graphs more readable and self-explanatory. I also provided a high-level context and explained visualisations by adding the description boxes on the right side of the screen.

4.1.6 Zooming-in

The graphs contain multiple data points, and the screen size may vary drastically between users. Therefore, to ensure that everyone can see and select particular elements, all the charts can be zoomed-in.

4.2 Views

The project has three views in total, each presenting different information about censorship in literature.

4.2.1 "The most censored authors" view

I consider it to be my most innovative view. It shows 1000 of the the most banned authors in history. Users can filter the view by the number of results displayed (100-1000), the country of the bans, or the name of the author.

To make the view engaging and symbolise censorship, I displayed authors on a shape resembling the "no entry" sign. I created an animated bubble chart where each point (mark) corresponds

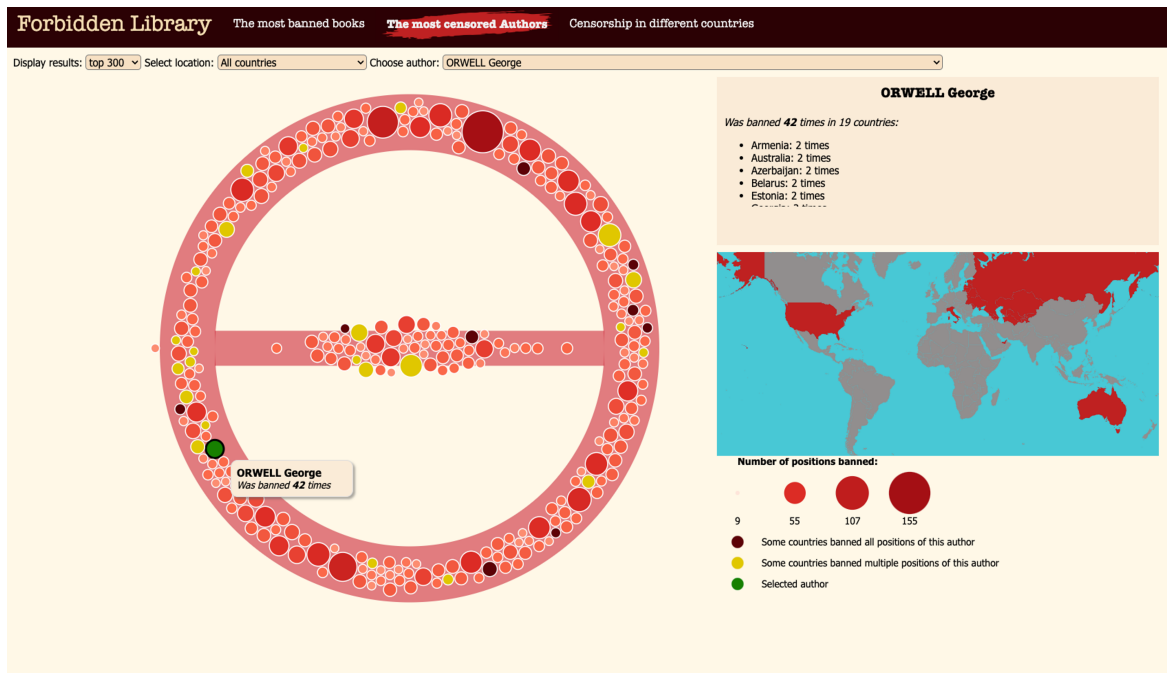


Figure 4.3: "The most censored authors" view

to an author. The goal of this visualisation was to encourage the user to browse/explore different banned authors in an engaging and fun way. Therefore, instead of displaying data on more conventional graphs like bar charts, I focused on the unique shape and animations of the graph.

The number of banned books for each author is shown through 2 channels - the area and colour saturation of the point. It is not an ideal solution, as according to data visualisation theory, each channel should only display one data property to avoid confusing the user. However, because of the limited dataset, I did not have another property that I could show using saturation. I considered making each circle a pie chart that would display the distribution of bans in different countries, but I rejected the idea because the visualisation was already quite big and complex. I decided to use 2 channels because of accessibility reasons. If I only used one channel, it would be difficult to differentiate between different circles, and the graph would not be readable. Figure 4.4 shows how the graph would look if I used either just the area or the colour saturation of the point. I made sure this design would not confuse the user by displaying a clear legend next to the chart.

I also used colour saturation to distinctly show which authors had all or multiple titles banned in at least one country. I decided to do it because in the dataset, the "all works" or "multiple works" was counted as a single data point; therefore, the total count of their banned work was inaccurate. Another difficulty was an uneven distribution of data. The number of books banned varied from 8 to 155, but only 19 out of 1000 authors had more than 50 banned books.

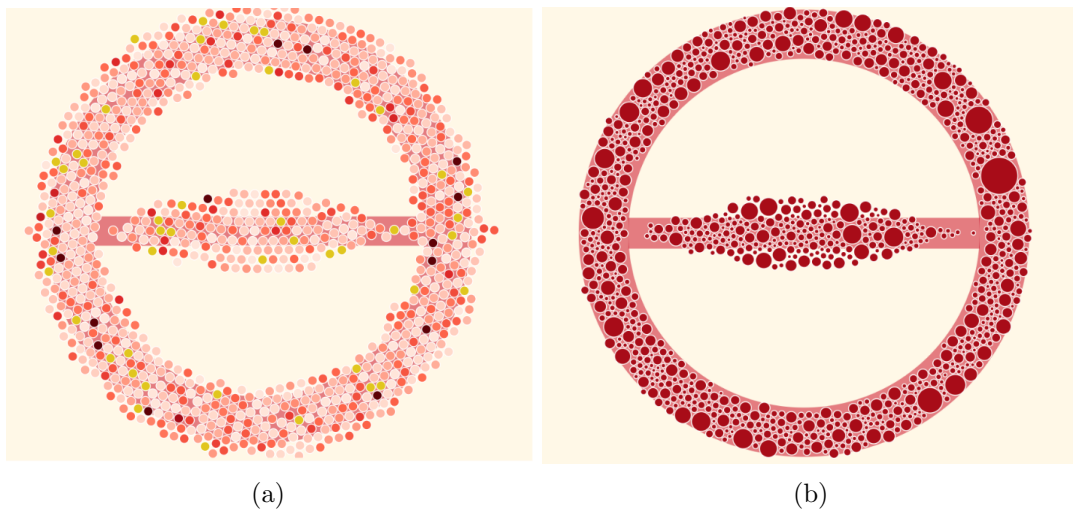


Figure 4.4: Alternative design: displaying information using a single channel

Therefore the linear colour scale changes the most from 8 to 50 to help differentiate between the median values. Following the rules described in the previous section, each circle has a small padding and a thick white border to help users separate points quicker.



Figure 4.5: A panel that displays the legend and detailed author's information

On hover, the number of bans and the author's name are displayed in the tooltip near the point. Since up to the 1000 data points are displayed in one view, it was impossible to fit all the author names on the screen. Adding this information on a tooltip seemed like a perfect solution as it appears immediately when the user needs it and does not take much space nor make the visualisation less readable. If the user clicks on the circle, they can see more information about the author and the number of their books banned in each country. I displayed that information on a list and a small map on the right side of the screen. To differentiate between the selected author and the rest of the points, I changed the colour of the selected circle to blue and thickened its border. To save space (in this and the remaining views), I displayed chart instructions and details about the author on the same scrollable panel on the right side of the screen (Figure 4.5).

4.2.2 "The most banned books" view

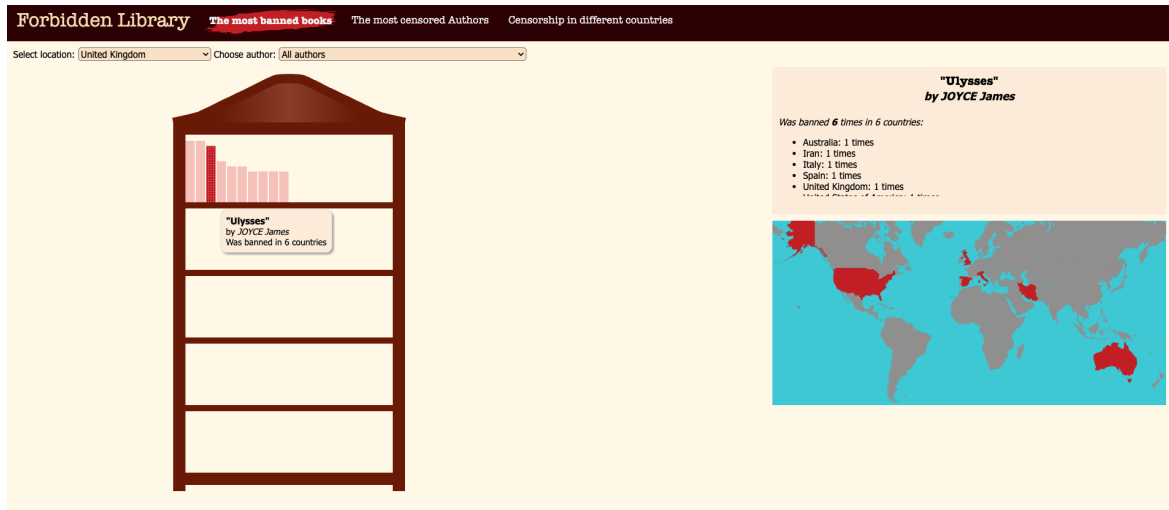


Figure 4.6: "The most banned books" view with "United Kingdom" filter

In this view, users can see the top 100 most censored books of all time worldwide and for each country. They can also filter them by the author's name. Each book is represented by a line. The main channel is the position on a common scale to display the number of times the books have been banned. The visualisation is a simple bar chart, but to make it more engaging, I created an illusion of a book cabinet (Figure 4.8). The chart has been cut into five parts to resemble books fitted on five bookshelves. As the main goal of this visualisation is the exploration of the most censored books, the users need to know the "relative" numbers (which books were banned more frequently than others) rather than the exact ones. Because of that reason, the chart legend and axis were not necessary, so I decided not to display them to keep the illusion of a book cabinet.

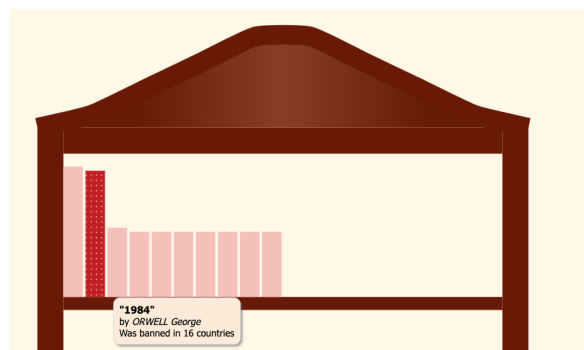


Figure 4.7: Selected book with changed colour and pattern

Similarly to the other view, on hover, the book details and the number of times the book has been banned are displayed in a tooltip. On book click, the details of its ban are displayed

on the right, in the form of a list and a map. To keep the views consistent, I used the same map and panel for showing detailed information. The selected book changes colour, and the dot pattern is added to it (Figure 4.7). Users can filter the information using two dropdown menus. They can select the country to see which books were the most banned in that location. They can also check the books of a specific author by selecting their name in the dropdown. The options in the dropdown are dynamic - if the location changes, the dropdown only allows one to choose the authors that were banned in a given country. If a user selects the author and then changes the country where that author has not been banned - the author selection is automatically removed.

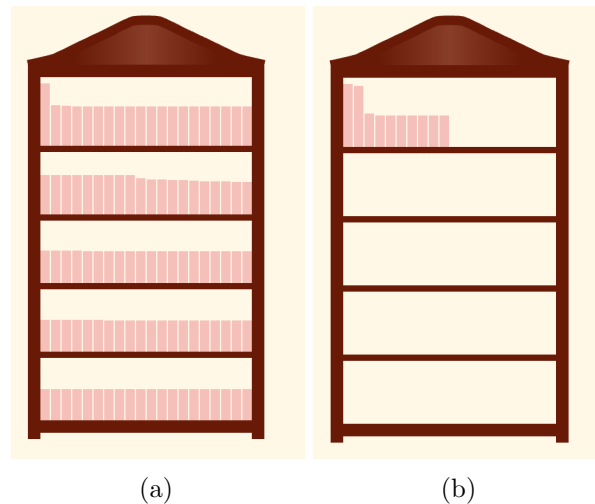


Figure 4.8: Book cabinet with different number of books

4.2.3 "Censorship in different countries" view

The last view provides a high-level summary of censorship worldwide. It comprises a choropleth map and a line chart. The number of books banned is displayed using the the colour saturation of every country on the map. On a country selection, users can see a detailed timeline of when the bans have been imposed.

To make it more accessible and faster to use, users can select a country in 2 ways - by using a dropdown list or by clicking on the country area. On country selection, to highlight the user's choice, its border thickens, and the other countries' opacity decreases. The data is also added to the line chart. The countries that have no known bans in the dataset are grey and cannot be selected. Their names do not appear in the dropdown.

The line chart changes dynamically - users can select up to 8 countries to compare the censorship trends between them. If they try to select more or if there is no data about the exact years of each ban, a pop-up message is displayed on the top of the screen (see: Figure

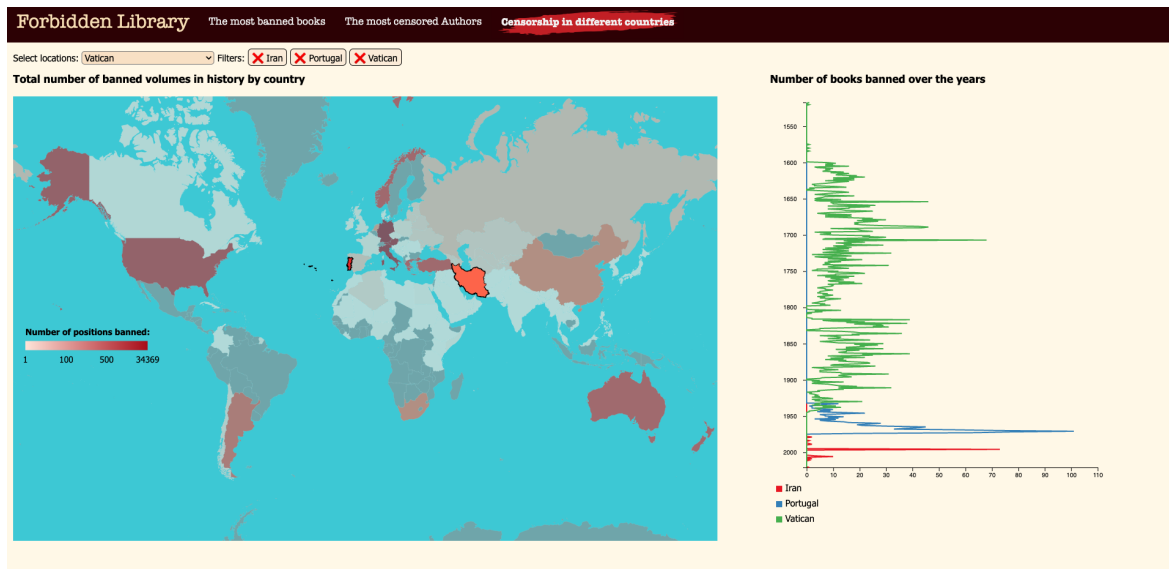


Figure 4.9: "Censorship in different countries" view

4.10). The ranges of the axis change dynamically, always displaying minimum and maximum keys and values of the selected data. The existing lines on the chart adjust smoothly to the new axis ranges with an animation. I decided to use a line chart as it allows comparing trends of ordered attributes on multiple values and keys levels. To fit the line chart into a limited space next to the map, I decided to flip its values and keys - unconventionally, the years are displayed on the y-axis while the number of bans is on the x-axis.

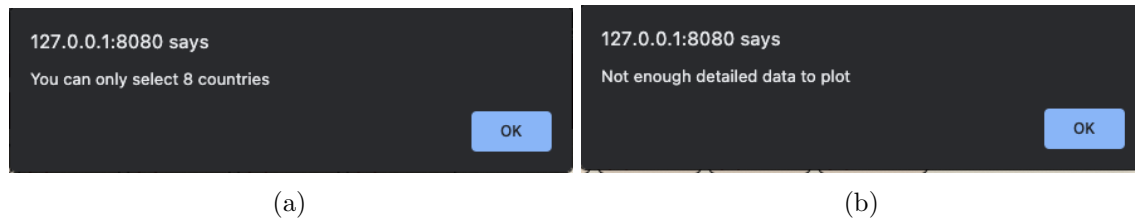


Figure 4.10: Pop-up warnings

4.3 How are the views linked?

I ensured that all the views had the same styling and colour palettes. The style and the amount of information on the tooltip are consistent between views. On click events, the new information is always displayed on the right side of the screen.

Authors' and books' views are closely linked - they have the same dropdown filters that work between views - e.g. if the user selects location and/or the specific author in one view, the selection also works in the other view.

5. Usage Scenario

The user is a literature enthusiast that wants to learn more about censorship.

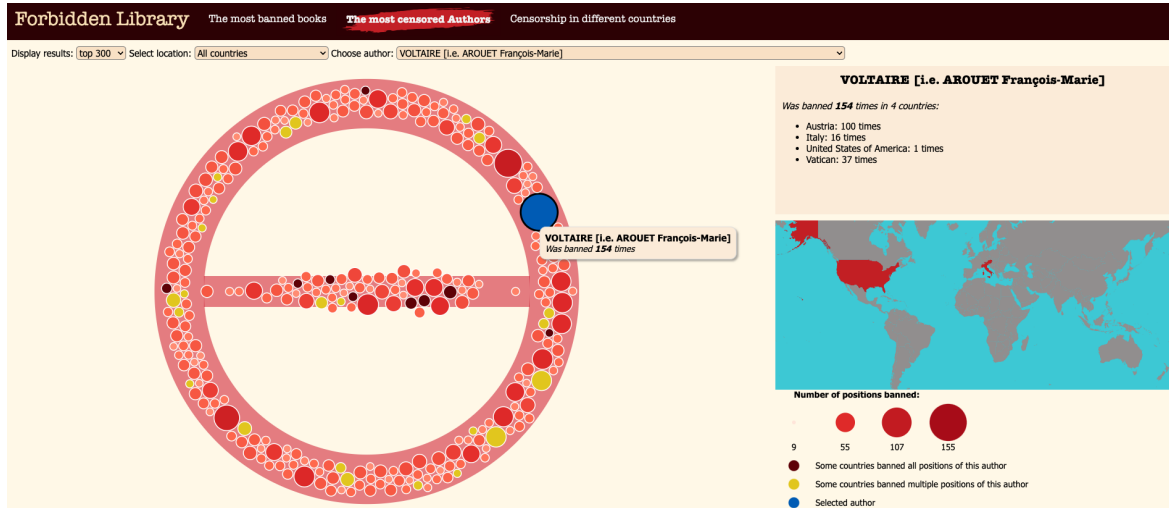


Figure 5.1: Step 1: "The most censored authors" with the largest circle selected

Step 1: They first go to the "the most censored authors" tab, where they see the graph. They select the largest circle to check the author.

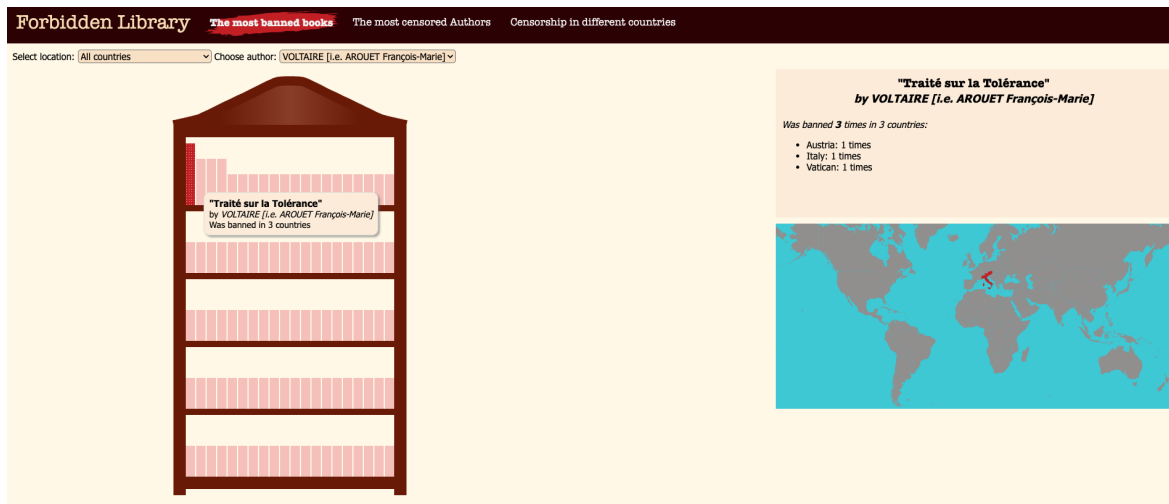


Figure 5.2: Step 2: "The most banned books" with an author automatically selected

Step 2: They want to know more about the work of this author; therefore, they go to the "the most banned books" tab.

Step 3: They notice that many books have been banned in Vatican City. They want to know more about censorship there and understand the reason behind these bans, so they go to the third tab.

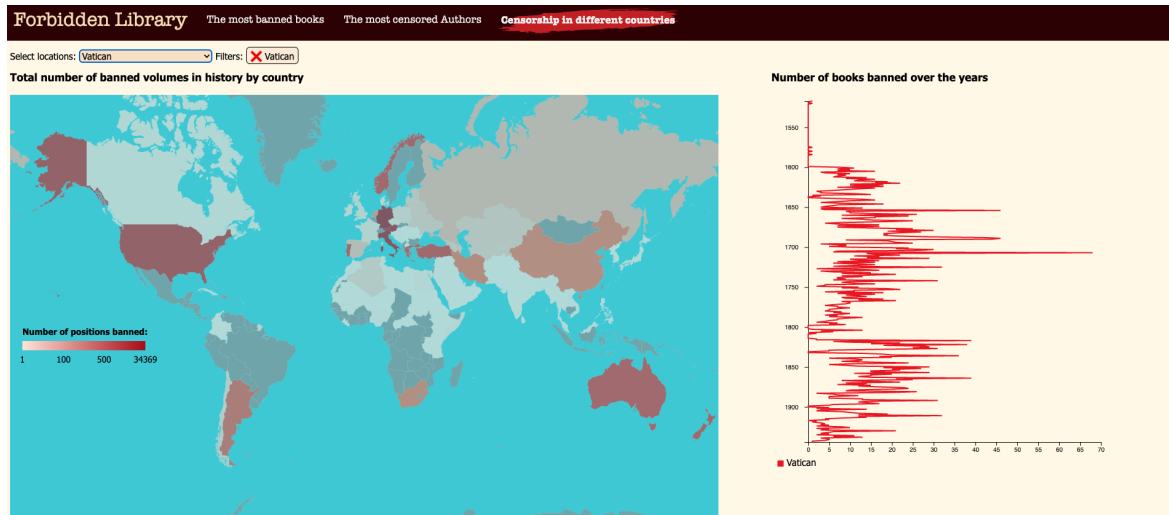


Figure 5.3: Step 4: "Censorship in different countries" with Vatican selected

Step 4: They select the Vatican from the dropdown, and they can see that the Vatican has been consistently banning many books throughout history.

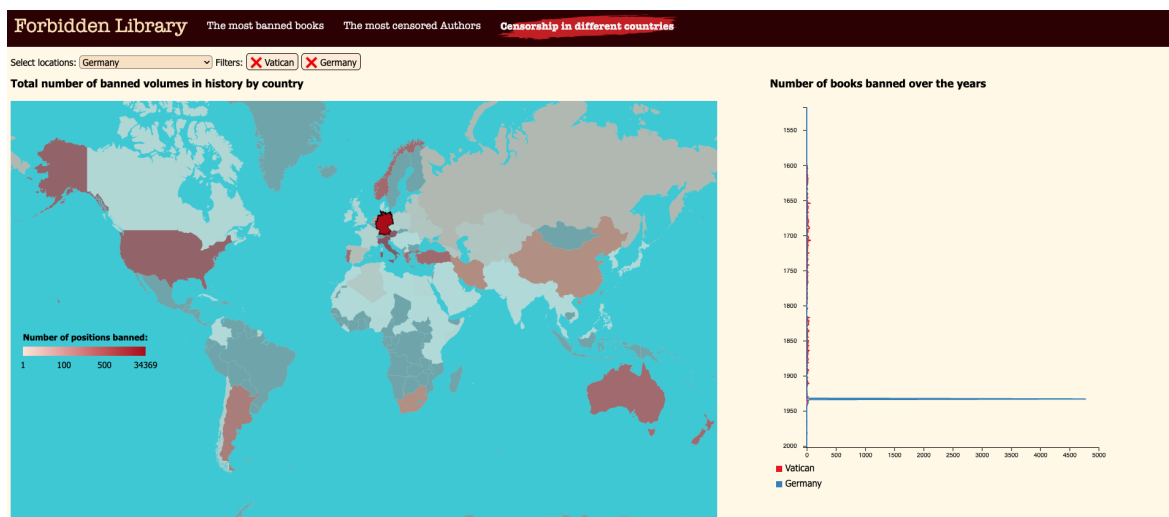


Figure 5.4: Step 5: "Censorship in different countries" with Vatican and Germany selected

Step 5: They wonder if that is the case with other countries. They select "Germany" and notice a huge spike in bans during the Second World War. They find it interesting how historical events can affect censorship, so they explore more to see other examples of that.

6. Credits

The data visualisation called "Horried?" made by Mohamad Wake [20] was the biggest inspiration for this project. It gave me the idea to use charts in a shape that relates to the topic of the visualisation (bubble chart in "The most censored authors" view and bar chart in "The most banned books" view).

Implementation-wise, the project has been built on the code provided and developed in problem sheets from this module. The bubble chart has been developed using tutorial on D3-force by Ben Tanen [19] (major functionality additions to merge two types of force). D3 in Depth [6] and StackOverflow [17] have been used as a reading material to help understand the logic behind D3-Force, but no code has been reused.

For the line chart with multiple data groups, an example from D3.js Graph Gallery [7] has been used as an inspiration, with major functionality additions made to the source code to make the chart dynamically update with new data.

Bibliography

- [1] ADOBE COLOR. Contrast checker. <https://color.adobe.com/create/color-contrast-analyzer>. [Online]. Accessed: 14 Apr 2023.
- [2] BANNED & CHALLENGED BOOKS. Top 10 most challenged books lists. <https://www.ala.org/advocacy/bbooks/frequentlychallengedbooks/top10>. [Online]. Accessed: 16 Apr 2023.
- [3] COLBLINDOR. Coblis — color blindness simulator. <https://www.color-blindness.com/coblis-color-blindness-simulator/>. [Online]. Accessed: 14 Apr 2023.
- [4] COLORBREWER 2.0. color advice for cartography: 6-class reds. <https://colorbrewer2.org/#type=sequential&scheme=Reds&n=6>. [Online]. Accessed: 14 Apr 2023.
- [5] D3. d3-scale-chromatic: d3.schemecategory10. <https://github.com/d3/d3-scale-chromatic#schemeCategory10>. [Online]. Accessed: 14 Apr 2023.
- [6] D3 IN DEPTH. D3 force layout. <https://www.d3indepth.com/force-layout/>. [Online]. Accessed: 16 Apr 2023.
- [7] D3.JS GRAPH GALLERY. Line plot with several groups. https://d3-graph-gallery.com/graph/line_several_group.html. [Online]. Accessed: 16 Apr 2023.
- [8] DIE KASSELER LISTE. Die kasseler liste: search the list. <https://www.kasselerliste.com/die-kasseler-liste/>. [Online]. Accessed: 14 Apr 2023.
- [9] FOSSHEIM, S. L. An intro to designing accessible data visualizations. <https://fossheim.io/writing/posts/accessible-dataviz-design/>. [Online]. Accessed: 16 Apr 2023.
- [10] GIV.UK. Dos and don'ts on designing for accessibility. <https://accessibility.blog.gov.uk/2016/09/02/dos-and-donts-on-designing-for-accessibility/>. [Online]. Accessed: 16 Apr 2023.
- [11] GOMEZ, B. 10 reasons books are challenged and banned. <https://bannedbooksweek.org/10-reasons-books-are-challenged-and-banned/>. [Online]. Accessed: 16 Apr 2023.

- [12] ISLINGTON. Banned books week 2021. <https://www.islington.gov.uk/~media/sharepoint-lists/public-records/leisureandculture/information/factsheets/20212022/20210907bannedbooks20211.pdf>. [Online]. Accessed: 16 Apr 2023.
- [13] KEENBLOG. Accessibility considerations in data visualization design. <https://keen.io/blog/accessibility-in-data-vis/>. [Online]. Accessed: 14 Apr 2023.
- [14] LONDON LIBRARIES / STELLAR LIBRARIES. Banned books 2011. <http://www.banned-books.org.uk/index.html>. [Online]. Accessed: 16 Apr 2023.
- [15] OSMAN, M. The ultimate list of web-safe html and css fonts. <https://blog.hubspot.com/website/web-safe-html-css-fonts>. [Online]. Accessed: 14 Apr 2023.
- [16] PEN AMERICA. Banned in the usa: The growing movement to censor books in schools. <https://pen.org/report/banned-usa-growing-movement-to-censor-books-in-schools/>. [Online]. Accessed: 16 Apr 2023.
- [17] STACK OVERFLOW. Getting d3 force to work in update pattern. <https://stackoverflow.com/questions/51153379/getting-d3-force-to-workin-update-pattern>. [Online]. Accessed: 16 Apr 2023.
- [18] STYLIST. 50 banned books from recent history (including harry potter). <https://www.stylist.co.uk/life/50-books-that-were-banned/47882>. [Online]. Accessed: 16 Apr 2023.
- [19] TANEN, B. A tutorial to using d3-force from someone who just learned how to use it. <https://observablehq.com/@ben-tanen/a-tutorial-to-using-d3-force-from-someone-who-just-learned-ho>. [Online]. Accessed: 16 Apr 2023.
- [20] WAKE, M. Horrified? <https://alhadaqa.com/wp-content/uploads/2020/04/horrified.html>. [Online]. Accessed: 16 Apr 2023.
- [21] WEB ACCESSIBILITY INITIATIVE (WAI). Web content accessibility guidelines (wcag) 2 level aa conformance. <https://www.w3.org/WAI/WCAG2AA-Conformance>. [Online]. Accessed: 16 Apr 2023.