Milestone 3: SPRINGERs

Delivrables

GitHub repository with running instructions and screencast.

Process

To obtain the final result, we began by identifying a relevant and qualitative dataset. Because all members of the team are PhDs students in the privacy field, we were drawn to problems around privacy and concrete applications. After a scan of related datasets on the main platforms, we found the GDPR dataset on Kaggle, which aligns well with our research interest and with the requirements of the class.
During Milestone 1, we conducted exploratory analysis of the chosen dataset to understand its structure, the bigger patterns, and the limitations of the dataset. We found that the dataset is of high quality but medium quantity. We did not need to apply some preprocessing other than uniformization of the labels. We reviewed related work and found that the existing visualization of the data was purely tabular display (no point of view nor visualization). The TA validated our problem statement and allowed us to finalize our choice.
Throughout Milestones 2 and 3, we progressively built the website, integrating visualization. In Milestone 2, we selected the appropriate tools for the interactive visual components, and started sketching. Using this pre-selection, we selected the final choices and tools and built the final components of the website.
We made sure to iteratively test and refine our visualization based on team feedback and given our research questions. Once the visualizations were finalized and implemented, we used them to draw empirical conclusions, which we reported on the website.
Finally, we concluded with the creation of the screencast to demonstrate our tool.

Challenges we faced and design decisions

We did not encounter any major challenges during the project. The dataset size being small, there was no performance bottleneck. Using the class material as guidelines for the design decisions, we were able to map our requirements to solutions while ensuring accessibility. There are still technical implementation details that we would like to improve on in the website to improve on usability and correctness. However, these are minor and we have our main results.

Visualization Components

We planned to have the following visualizations: Minimal viable visualizations are shown in black, and extras to be implemented if time permits are shown in yellow.

Maps of European countries

Figure 1 shows our sketch from Milestone 2, showing the cumulative amount of GDPR fines per country. Figure 2 shows our final visualization. Our visualization is interactive, and allows the user to filter by country, article, and date range; it also shows summary statistics to the user (Fig. 3)

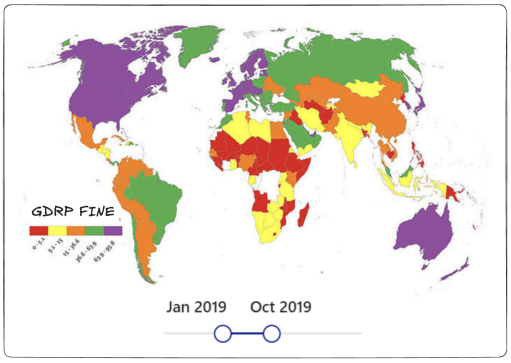

Figure 1: Sketch of map of countries showing GDPR fines (cumulative) given a time range.

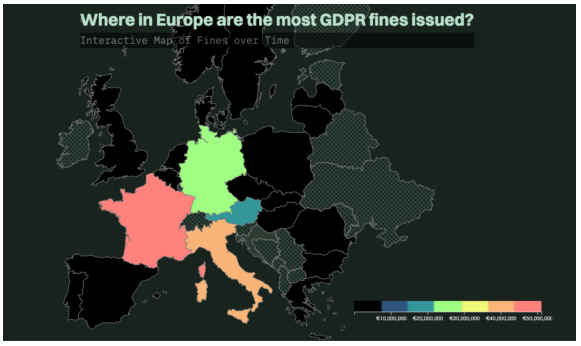

Figure 2: Final map of cumulative GDPR fines by country.

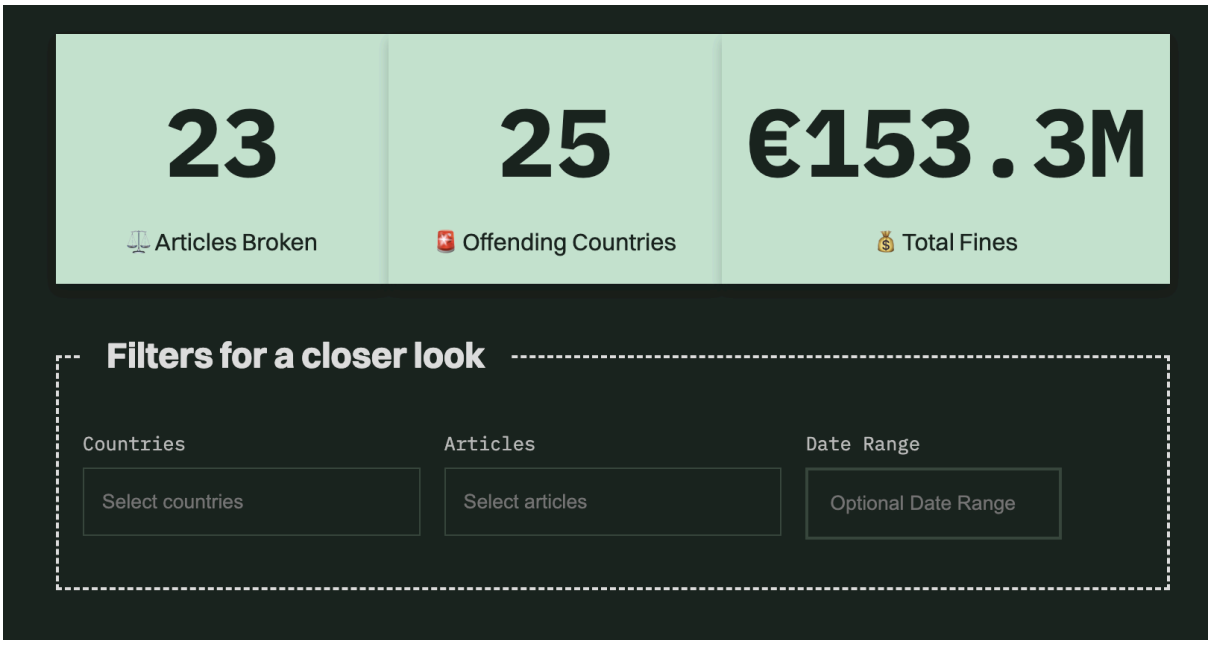

Figure 3: Summary of statistics and interactive filtering interface.

Correlation plots

Figure 4 shows our sketch for correlation metrics, and Figure 5 shows our final visualization. We show the correlation between fines by articles. The visualization is also interactive, and can be filtered using the same interface as the main map (Fig. 2).

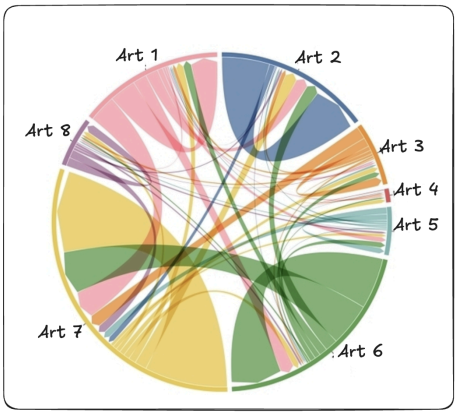

Figure 4: Sketch of a Sankey Diagram showing correlation between articles citations, from Milestone 1.

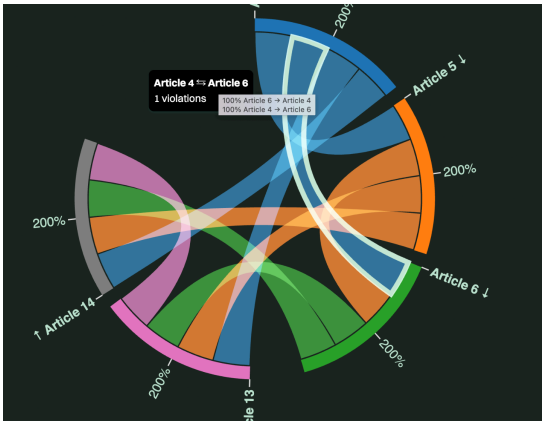

Figure 5: Sankey Diagram in our final visualization.

Peer Assessment

Milestone 1: All team members contributed to the selection of the dataset and design of the problematic and research questions. The exploratory analysis was done by Malo, and the related work exploration by Saiid.
Milestone 2: The visualization component plan and sketches were finalized by Christian and Mathilde. The tools were pre-selected from the class by Malo. The prototype was built by Saiid.
Milestone 3: The final website was built by Saiid, with minor adjustments by Mathilde. This document was produced by Christian and Mathilde. The screencast was done by Mathilde.
For each milestone, the whole team gave feedback on and reviewed the deliverables, i.e., documents and prototype.