Wizard's First Rule - Process Book

website

Introduction

Our group shares a common interest in video games, and we embarked on a quest to find an intriguing dataset. After thorough research, we concluded that the "Video Game Sales with Ratings" dataset would be the most fitting choice. This dataset spans from 1998 to 2016. The objective of our website is to delve into the functioning of video games across various regions. We acknowledge that cultural disparities exist between these regions, and we aim to explore how these differences impact video games, which are primarily designed for a universal audience. Additionally, we intend to analyze the evolution of video games based on the consoles they are played on. Our analysis also extends to changes in interest over the years and determining whether they are global or localized phenomena.

To meaningfully explore these questions, we designed the following graphs:

- **Violin plot**: This plot serves to provide users with a comprehensive overview of the dataset, while offering valuable insights into the evolution of platforms, categories, and publishers over time and across regions.
- **Barchart Race**: This animated graph evolves through time. It is a horizontal barchart-like graph which represents the top most items in a category at a specific epoch. It moves from one epoch to another by rearranging its bars such that they are ordered for that specific epoch. Its horizontal axis also grows with the animation because the bars are the cumulative sum until the present epoch.
- **Map of Genres:** This custom tailored graph represents the differences in video game genres preferences across regions and through time. The visualization, represented as an ocean of genres, conveys the message in a straightforward manner, while still providing detailed information in the tooltip.
- **Sankey Diagram:** a Sankey diagram is composed of ordered nodes and links between themes. Each link represents the flow from its left node to its right one, and the link width is proportional to the amount of flow from one node to the other. In our case, the Sankey diagram shows the flow of money from region to genre to platform, back to region, so we can have the money flow between each of those categories.

Violin Plot

Initially, we envisioned a simple representation of the data, as depicted in Figure 1, to provide users with basic information. However, we realized that this plot had a significant drawback: while it allowed for comparison of violin widths, it did not convey the meaning behind these widths. To address this limitation, we incorporated a tooltip that displays the information associated with each width when the user hovers over a violin (as shown in the top right of Figure 2).

Furthermore, we opted to enhance the visual appeal of the plot by utilizing a color palette throughout our website, specifically the `interpolateRainbow` palette, offered by D3, rather than relying on a monochromatic scheme.

During milestone 2, we toyed with the idea of including images inside the violins, such as genre-specific game images, console logos, or publisher logos. However, we found that the violins were too small to effectively display these images in a visually pleasing manner. As a result, we ultimately decided to employ a simple color scheme instead.

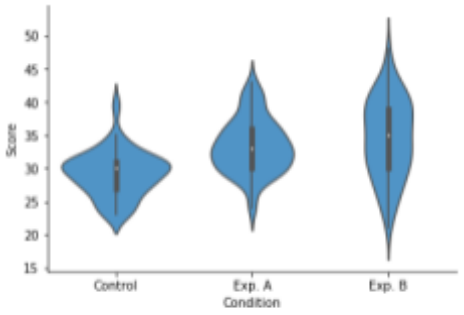


Fig. 1 - sketch of violin plot

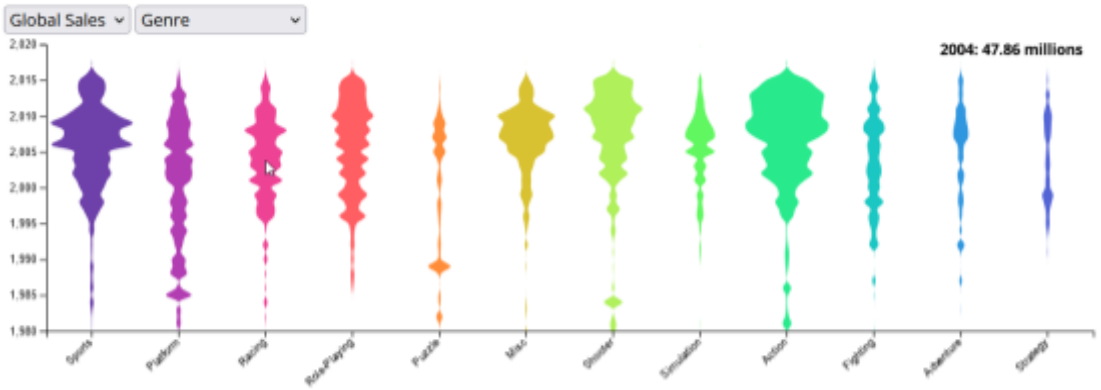
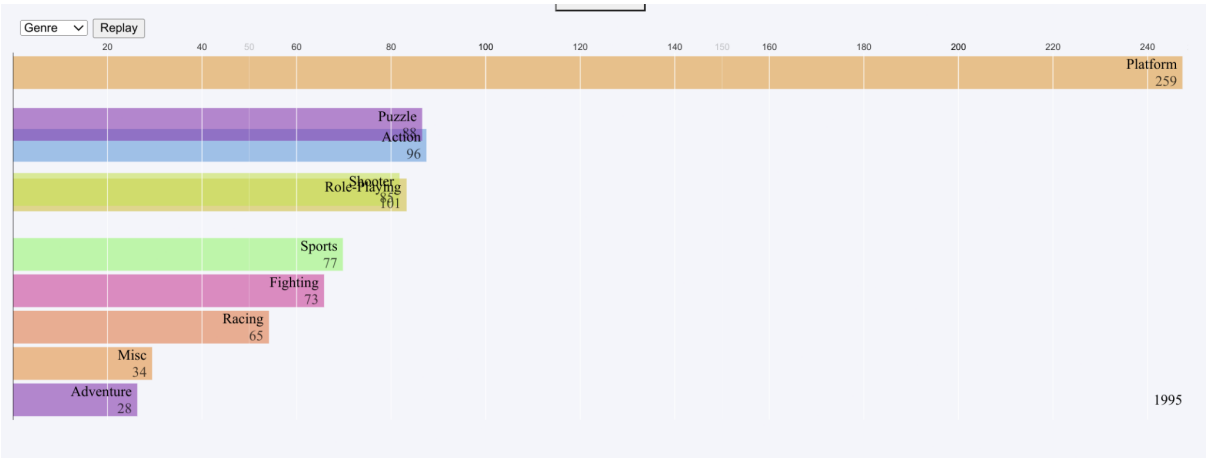


Fig. 2 - Final violin plot

Race barchart

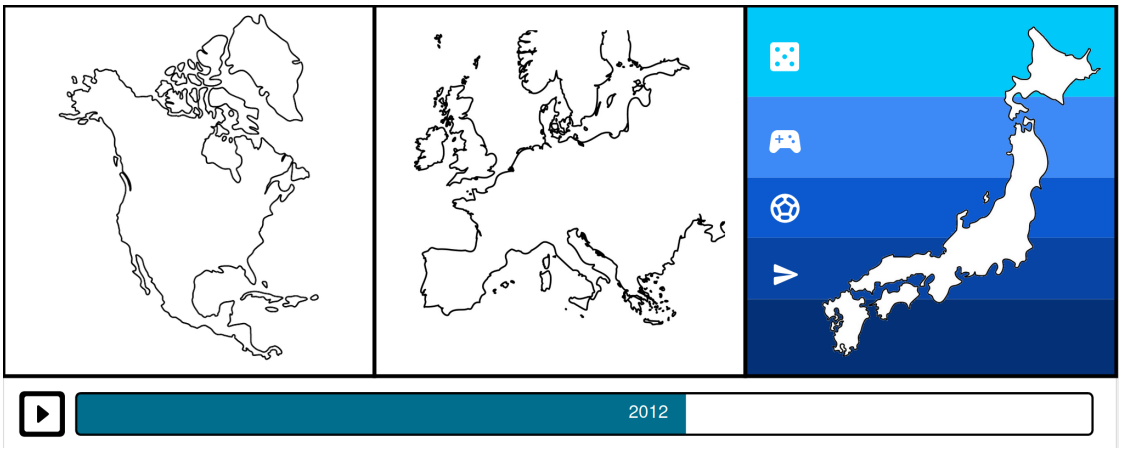
This graph, which classifies the candidates for a given time epoch, provides some hint about how the interest for a specific item evolves through time. It actually represents the amount covered by a horizontal line growing up from zero in the Violin plot presented just above. Many deductions can be done from such a graph, the three main are the amount of time required for an item to reach top position, the duration it stays to its position, and how far it is from the other item.

Our dataset only contains the year of release of each game. We chose to account for each revenue related to a game for its release year. And as we only have one value per year, we chose to linearly interpolate intermediate points bringing smoothness to the graph animation. The graph grows two years per seconds.



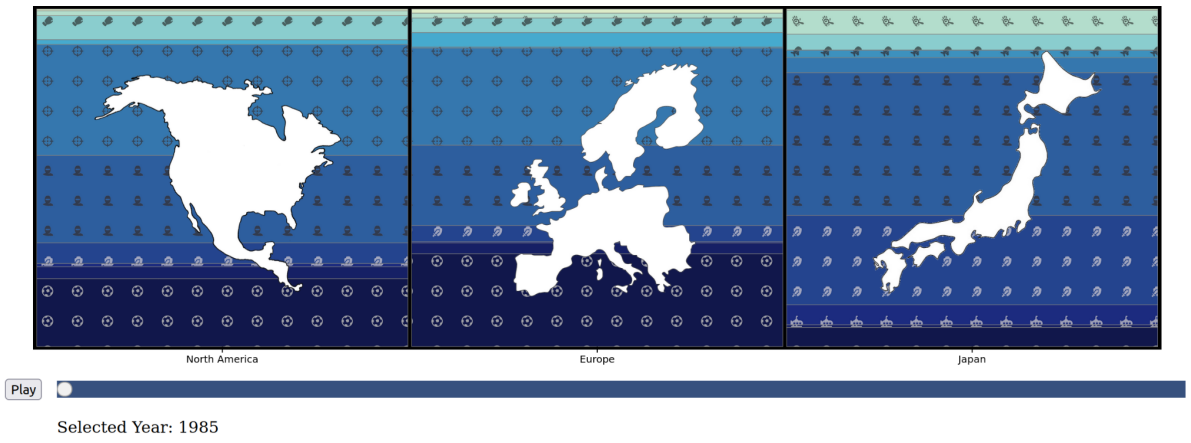
As explained above, this graph is animated through time, which is not so convenient for such a medium as this static pdf. But here is a screenshot of a replacement of bars, in between two epochs. A good equilibrium between the duration of animation, the size of the bars, the numbers of items accounted and size of the overall graph have been reached by playing with such parameters and discussing between teammates for which, and what for such combination it preferred over another one. This graph somehow perfectly fit our desire expressed in milestone 2, except for the year selection feature: its development was mostly straightforward.

Map of genre preferences



Original sketch from Milestone 2

The goal of this graph was to display the differences in video game genre preferences across the available world regions. The initial idea was to have the map of these regions with the proportions of each genre displayed inside of the occupied surface. This idea was quickly scrapped as it would've been very difficult to implement and harder to read. The idea of an "Ocean of Genres" came as we explored the variety of these cultural differences, and thought it could be represented as the ocean surrounding each of these regions.



Final graph as presented on the website

The achieved result successfully displays the intentions we had when doing milestone 2, but with a few changes. Firstly, the bars had to be implemented in a way that represents the proportion of genre per region for each year which was a relatively simple task.

The goal for the color and icon choices was to have an organic movement, simulating the effect of waves in the ocean while still being able to discern the different sections.

After that, a tooltip was added to textually display the genre and proportion represented by a section to be able to get more explicit details about the data. One idea that was scrapped was to display the main video game titles that represent a section. This was due to the complexity of the task and the limitations of the resources at hand.

The final addition is the year slider and play button which enable the navigation of the graph through time. Once again this task was complex due to the previously added functionalities and the need for all of them to work together.

The final results conveys the message in a simple and visually pleasing manner. The tooltips also help with the comprehension of the displayed results.

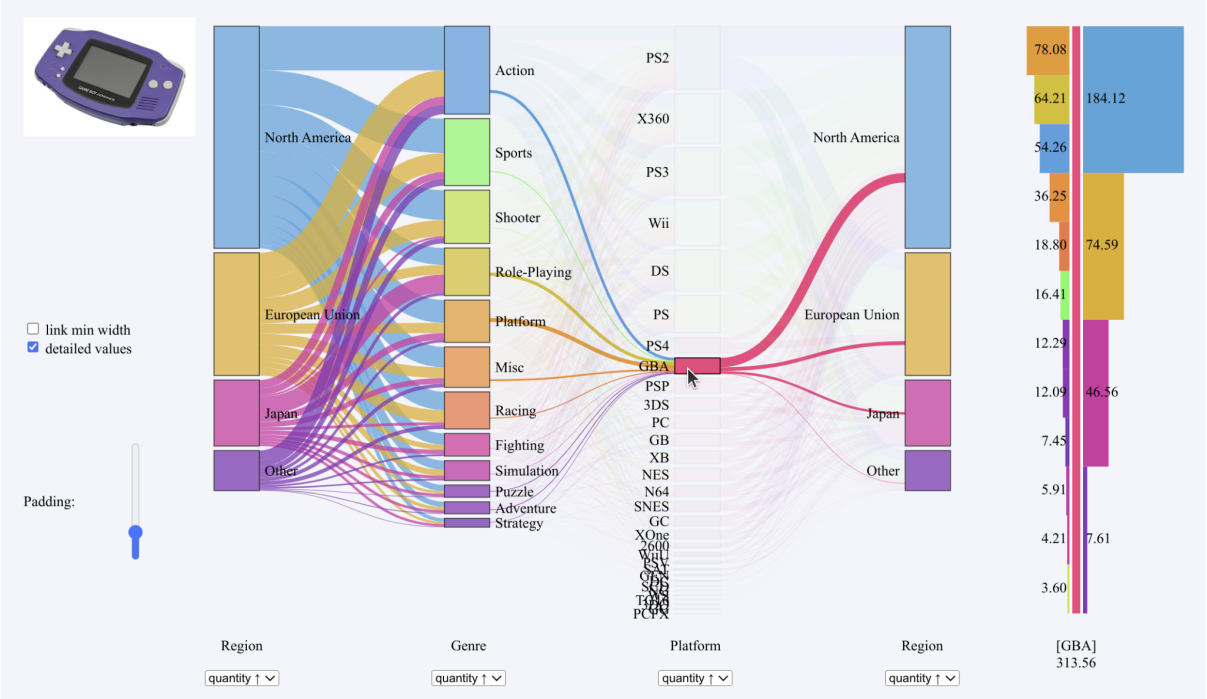
Sankey diagram

Sankey Diagrams development is based on the `d3-sankey` package offered by Data Driven Document. Sankey diagrams represent the flow between nodes; so the first step is to choose what nodes should represent. We ended up choosing three categories: regions, genres and platforms.

The used package requires the data to be in a specific format, so after cleaning a bit the raw data, we construct the structured data object.

This part has been done in two approaches, the first one was in a pre-processing approach: we pre-compute the structured data using python's pandas library and save the output in a .json file, which is then only read each time the site loads.

The second approach is to clean and create the object with javascript methods on site loading. As the second approach did not slow down the website in a noticeable way, we chose to keep this path for the beauty of the exercise. With the constructed object, we can then create a raw diagram, which still requires to be shaped and for which interaction can enhance the experiment a lot.



The first interaction implemented was to darken the surroundings on hovering of a node, in other words, enlightening the hovered node and its incoming and outgoing links. Then once this has been implemented, based on the opacity of the links, we remark that, due to the high money flow difference, some of the links were too tight, thus invisible. So the possibility to set a minimal width for links brings some value to the graph, even if it tricks the user a bit.

Platform names are not speaking to everybody, thus we collect wikipedia links and a descriptive image, so that on hovering labels or nodes, the image appears, and clicking it brings the user to the related wikipedia article. One small trick there was using timeout for image hiding so that the image does not disappear directly when exiting the hovering zone, otherwise it would not be possible to click on it.

By default, nodes are ordered in the same order as the structured object. To have meaningful representation, we chose to order nodes by value or by name, letting the choice to the user. Ordering by value is great for global scope: the graph is cleaner and a good overview of the total flow tendency can be seen. Ordering by name provides some local hint, especially for platforms. It is, for example, easier to compare all the playstations platforms by hovering in a name ordering.

Sankey diagrams have a `padding` attribute for their nodes. The graph is somehow more pleasant with no padding, but as some platforms have really low input/output flow, the node's labels overlap with neighbors', thus impossible to read. So we chose to enable the user to add some padding shaping the graph so that lapels are readable. Let's notice here that the

hovered node’s label also appears on the bottom right of the graph.
Our Sankey diagram is not adapted to show the links widths values. The implemented solution is to plot a horizontal barchart corresponding to the incoming and outgoing value of the hovered node. One central vertical bar, filled with the color of the hovered node, acts as the backbone for the other values. The values on the left are the incoming value (on the node’s left side in the Sankey), and on the right the outgoing ones. The flow value of the links (their width) are represented by the bar length, and the number of bars correspond to the number of incoming/outgoing links. Each bar's colors correspond to the other extremity’s node colors. This barchart allows the user to easily see whether the node flow corresponds to the overall flow tendency or not. In other words, when the bars are in the same order as the nodes, the hovered node follows the overall tendency, but if there is a lot of entropy in the bars colors, the node is somehow against the grain. For example, North America is the biggest region’s node, and every genre but the role-playing have the top incoming node being the blue one (North america). As role-playing top incoming node is Japan, it does not follow the main tendency. Colors often tell enough, but for sake of completeness, it is possible to tick a checkbox for printing the link value in arabic numbers.

Peer Assessment

Dayan Massonnet:

- Violin Plot
- Web text for Introduction, for Violin and for Barchart Race graphs

Paul Keller:

- Barchart Race
- Sankey Diagram
- Web text for Sankey Diagram

Eloi Eynard:

- Website design
- Map of genre preferences
- Video recording/editing/publishing
- Web text for Map of Genre and Conclusion

Overall, we discussed the main design decisions together, and the team had a great group dynamic. We are pleased with the result we presented.