SteamViz

Process Book

Derin Arda Alpay, Ahmad Bilal Kakar,
Thai Nam Hoang

Introduction - Data

Choosing our topic

Each day, with faster growing technology, various kinds of entertainment options arise. One such option is gaming. Enhanced with the most cutting-edge research and hardware, each year is always exceeding the previous in terms of anticipation. As Computer Science students(and gamers), we asked ourselves, how much does gaming affect us? We might not feel we are spending too much time and money on games, but is it just us convincing ourselves?

Therefore, we as group SteamViz, decided to take a look at various data scraped from the most prominent game publishing platform, Steam, and visualize just how prominent gaming is, in our lives.

Finding a dataset

With our idea in mind, we ventured forward to find a suitable dataset. Here, we encountered our first issue: The gaming space is too broad. With tens of thousands of games, each having various information updated by the hour, scraping Steam or acquiring a full-sized dataset would be too difficult. Therefore, we opted to choose a smaller dataset with fewer per-game information, and kept a prominent website in mind, one that is the go-to place for Steam data and supports academic research by sharing data: SteamDB.

Definitely a big turning point for our project plans, SteamDB gave us ideas on just how big the data can get, and what kind of visualizations we could create from them. It also pushed us into thinking, with the prompt of our amazing TAs, to explore and look into unused parts of the data, consolidating our plans of visualizing addiction.

Therefore, setting our sights realistic, we chose a dataset published on Kaggle, while keeping SteamDB on the side if necessary.

Introduction - EDA and Website

EDA

An initial EDA showed us that most of the data is unnecessary for our project, and the necessary ones are not extremely precise, mostly involving ranges rather than specific numbers. Some fields are also not clear, raising the necessity of cleaning and preprocessing our data.

Website and visualizations from scratch

Having explored our data and set our objectives, we designed several visualizations; one for lightly discovering the data and the games within, two more for exploring the aspect of addiction. We made the decision to not overextend and choose a different type of technique for each visualization, so after careful consideration, we chose bubble plots as our main focus. The reason behind this choice was due to their dynamic ability to show relevance of categories, as well as having great transitions to other forms of plots.

Tools we have explored and used

Throughout development, we found several tools useful, some are explored but are not in our final product. Here are the ones we tried.

HTML, CSS, JS

D3.js

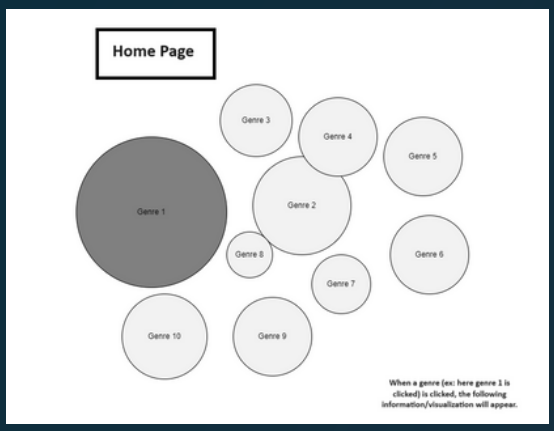
Tailwind CSS

React

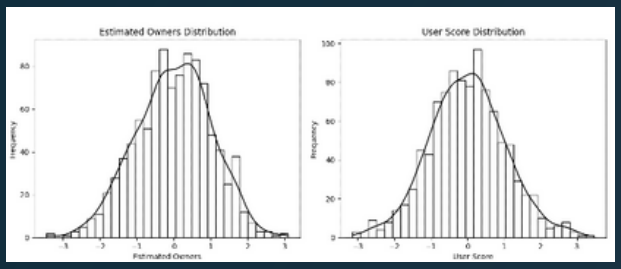
Implementation - Home

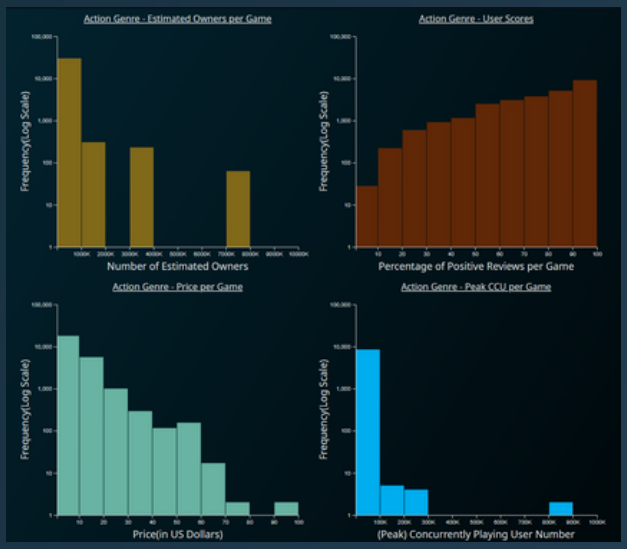
Welcome to SteamViz!

When you first land on our website, you are welcomed immediately by a bubble chart with different genres. Each bubble here is as large as the number of games tagged with it!



Clicking on any of them will bring up interesting information about them, such as how many games of a genre are owned by how many people, how is it rated among the community, how much do these games cost, and what is the largest number of people who have played it at the same time!





Implementation - Game Addiction

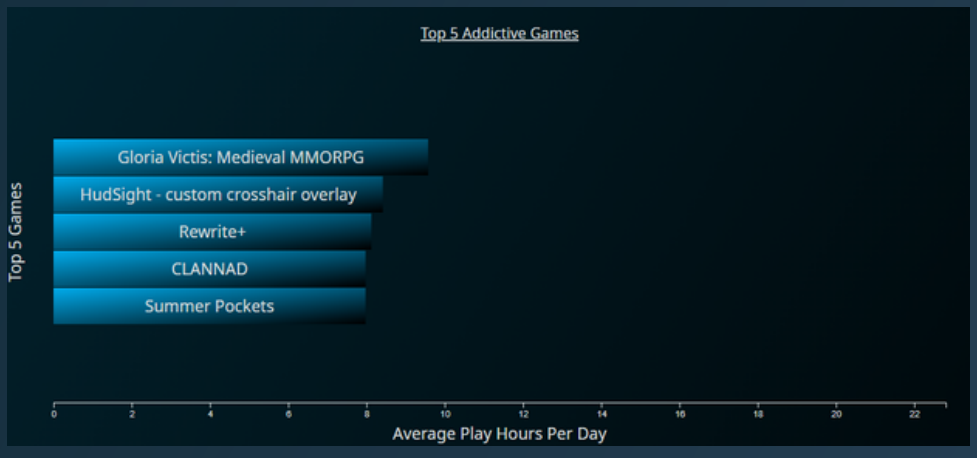
Game Addiction

In this page, we explore games individually according to how addictive they are. To make it easier, we remove a bunch of games with no playtime recorded whatsoever, and work on the remaining data. Here's how our bubbles look like:

And the table that shows various information about our games have changed. We decided to simply put how addicting some of the games within these categories are, instead of unrelated information.

Game	Price	Peak CCU	User Score
TBD	TBD	TBD	TBD
TBD	TBD	TBD	TBD
TBD	TBD	TBD	TBD
TBD	TBD	TBD	TBD
TBD	TBD	TBD	TBD

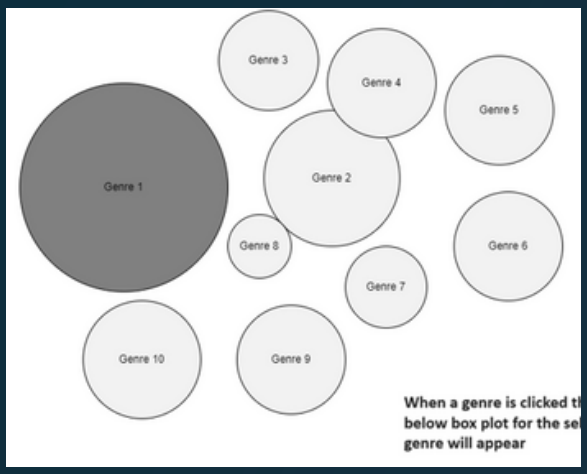
We've also kept the x-axis scale high, since some games fill it out entirely. To not cause disruption between different addiction stages, we decided to sacrifice some space.



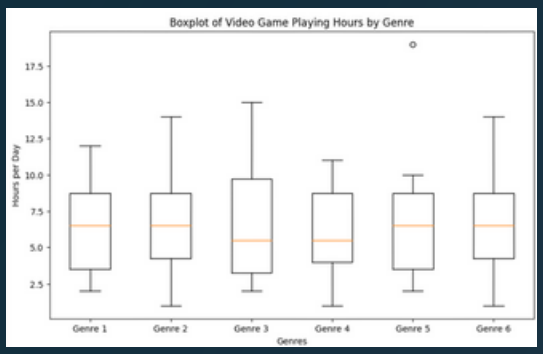
Implementation - Genre Addiction

Genre Addiction

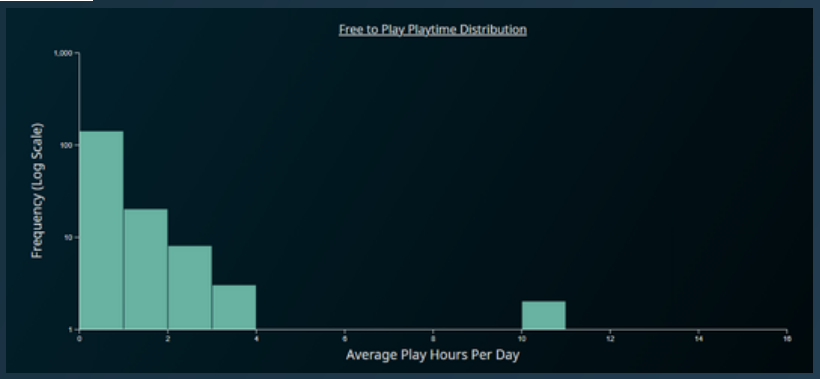
In this page, we explore various genres once more, but this time according to how addicting they are! Once more, we remove a bunch of games with no recorded playtime. Here's how our bubbles look like:



Although our initial plans were about having a boxplot, after trying it, we unfortunately saw a lot of outliers and other issues with our data. To not give a faulty representation, we decided to swap back to a regular old histogram of gameplay time:



Same as our Game Addiction page, we decided to keep our histograms in scale with each other. This causes empty spaces to appear as usual, but the subject is worth these sacrifices.



Challenges

Issues we've faced...

Here is a small list of what we had trouble with:

- None of us knew how to code frontend projects. Therefore, learning some components took a long time, and our inexperience led to writing unoptimized code.
- We lacked a certain coordination, mainly due to what will be explained in "Peer Assessment" section.
- The data we've collected had inherent limitations regarding what we wanted to show.

... and issues we've solved.

- Solving the first issue was easy, we just studied and worked. Although our code being unoptimized most likely remains as a fact, we've improved on it.
- Sadly, the second issue was unsolved, all we could do was to split the entire work among the remaining two of us. On the bright side, we'd like to believe this forced us to learn a bit more about our project than usual.
- We addressed the last one by changing our perspective on "addiction", and exploring other options without being tethered to any singular one. However, the limitations that remained were harsh.

Peer Assessment

Whodunnit?

Below, you can find the contributions of each member throughout our project!

Derin Arda Alpay

For M1, wrote the report and did part of EDA. For M2, created the skeleton of the website and added finishing touches to the report. For M3, completed the homepage, improved visuals and design of the website, wrote the process book and recorded the screencast.

Ahmad Bilal Kakar

For M2, wrote the report and designed the interactions and visualizations. For M3, created the main visualization pipeline each page uses, and completed the Game Addiction and Genre Addiction pages.

Thai Nam Hoang

For M1, did part of EDA.

Issues with the third teammate

As it's obvious to see, our last team member has not done anything since M1. We haven't heard from him since before M2, despite numerous attempts. Therefore, the huge majority of this project has been done with just two people. Please contact derin.alpay@epfl.ch for more details, if you'd like.