

COM-480 Milestone III: Process Book

Xinran Tao, Ahmed Ezzo, Guanqun Liu

June 3, 2022

1 Project Overview

Our project aims to provide a survival guide to game developers who focus more on technical issues and intend to learn more about remunerations and current market trends. Instead of illustrating the data statically, we implement our visualization components interactively to expand the range of information related to games and their developments to our audiences with increased readability and data comprehensiveness. Compared to our initial plan, we extend the time interval from the Covid period to recent years due to a lack of an accurate public database. In addition, we refer to more data sources to make our data story complete. The following sections will introduce our visualization components, the data source used, visualization design and challenges faced, brief part breakdowns to show each member's workload, and possible future improvements to our project.

2 Data Source

2.1 Kaggle

- **Games of All Time from Metacritic:** The list includes 8,831 games from Metacritics filtered with user ratings, description, developers, etc;
- **Steam Store Games:** Combines data of 27,000 games scraped from Steam and SteamSpy APIs until 2019, including game genres, platforms, release year, developers, etc;
- **Video Game Sales:** Contains a list of 3,817 video games with sales greater than 100,000 copies from 1976-2016, including game sales in million dollars;
- **Video Games List:** Contains a list of games for Windows from 1983 to 2021. including game title, release year, companies, and genres. Used data between 2011 and 2021.

2.2 Reference Websites

- **Metacritic & OpenCritic:** Websites that aggregate high-quality reviews of video games. They receive comments from game critics and commentators and convert them to a 100 point scale;
- **Games-Stats:** A website that collects game development information and revenues;
- **Glassdoor & Levels.fyi:** Websites that record employees' remuneration including famous game companies and studios;
- **Statistica:** A statistics portal for market data, where we collect sales data and popular games by country for 2020 and 2021.

3 Design Details

3.1 Development Tools

Tool	Usage
React.js	JS library for main user interface development
three.js	JS library for creating animated 3D computer graphics in web
d3.js	JS library for creating dynamic and interactive data visualization
Nivo	JS library built on d3.js that provides supercharged React components to build data visualization Apps
React-Globe	React components for globe data visualization using ThreeJS/WebGL
Pandas	Python library for data processing and analysis

3.2 Modifications

Regarding the visualization plans stated in former milestone reports, we modify the following contents in our final visualization project to present a common storyline of providing helpful information for developers seeking market trends, employment in big studios, and career in indie game development:

- Rearrange the time interval from the Covid period to the recent 10-20 years.
- Remove the top game ranking graph and studio-based statistics table since they fit closer to the player's angle. Overall statistics cannot reveal studios' genre preference and popularity.
- Modify the popularity-based (daily online players) game-studio bipartite graph into studio productivity-based (top 10 studios). Besides, we provide their HQ location and median salary (from websites in 2.2) of a standard game developer or designer.
- Combine the 3D animated scene with the general word cloud graph to provide regional gaming market sales and the most popular games. We modify the initial scene from sky-telescope to a zoomable and clickable earth that can show regional game trends when hovering over and clicking the country areas.

3.3 Challenges

The first challenge is the lack of the latest video game database. The best game sales database we found only has complete statistics until 2016. Besides, indie games and their developers have even fewer media exposures, which requires us to crawl all possible relevant information for the top indie games we selected.

The second challenge relates to the design of graph dimensions that can adequately render our idea of providing a "survival guide" to technical geeks. Solution of dilemmas such as the standard of an "out-standing" game, proper division of game genres, and data forms that fit our design intentions are discussed multiple times. Once we modify or remove some of the visualization components, we need to rearrange the data contents in all others to ensure our data story is complete and consistent.

The third challenge mainly includes the interactive design process. As we provide additional information through mouse interactions, we need to enrich our primary database for visualization. For instance, to show that successful indie games can also bring about considerable revenue for their developers, we search for the game sales, collect their earnings, and calculate their actual revenue. Besides, the coding-up process also takes much time to debug and conduct visual adjustments.

3.4 Design Decisions

According to the above challenges discussed, we finalized our visualization decision as follows:

- Use complete and up-to-date datasets as many as possible. For the game sales data, we still use the one up to 2016 since we cannot find a more complete and latest version. As for other graphs, we use datasets that at least contains games published until 2019. We also use a web crawler to collect information on studio team size and development duration for our scatter-bubble graph of indie games.
- Follow the criterion and division of games if recognized publicly (such as Wikipedia and Fandom). Keep all graph dimensions to attributes included in our prepared datasets and avoid using composite dimensions (such as an index combining multiple standards).
- Implement our graphs using the same library and design our graph interactions in similar visual patterns.

4 Visualization Components

4.1 Stacked Diagram: Game Sales by Genre

Game sales reveal the market trends in a very intuitive sense and it is critical to developers of measuring market success. By showing the ups and downs of game sales, audiences can capture an outlook of the game industry. Considering our audiences may care more about sales under different genres, we dig deeper into the game market by showing the changing trends of game sales by different genres from 2000 to 2016. We gathered our data from *Video Game Sales* combining the genre information from *Games of All Time* and *Steam Store Games*.

4.2 Sankey Diagram: Big Head Companies

After showing the detailed game sales in the global market, we show game developers of top game companies' preferences over game genres within the recent ten years. Audiences can easily see the preference over the width of links and obtain an exact number of game products by hovering over the links and side brackets. We also provide company locations and the median salary as a standard game developer to our audiences as a career reference.

4.3 Scatter-Bubble Diagram: Small Studio, Amazing Games

However, some game developers may intend to launch their career in developing indie games, allowing them to show their creative game ideas. We care about their thoughts on whether a large development team is necessary for a successful game. Top indie games can also bring considerable revenue to keen developers as a reward. With node size representing the revenue, we want to convey our idea that small studios can still make phenomenal indie games and bring career success.

4.4 3D Earth & Wordcloud: Gaming Markets and Trends of Different Regions

Finally, as a bonus part of our visualization project, we created a 3D animated scene of rolling earth that shows the amount of gaming market sales and most popular games in different regions in 2020 and 2021. For game developers who want to explore regional game markets, we provide them with a fully interactive scene. Audiences can see regional gaming market sales by hovering around the countries. We also make the earth "zoomable" and "draggable", similar to a funny game which conforms to our main topic of a guide for game developers. We will show the corresponding most popular games in a word cloud form by clicking on a country's area. We combine all published games as recorded in "Games of All Time" and "Steam Store Games" and select the most popular ones by data from Statistica if provided.

5 Project Breakdowns

Our workload distribution follows the flowchart as shown in figure 1:

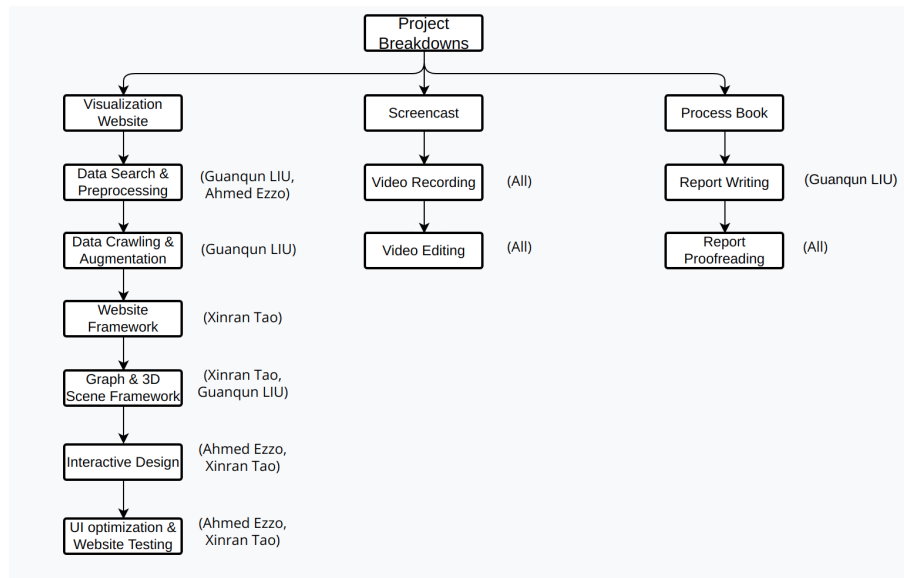


Figure 1: Flowchart of Project Breakdowns