Visualization for Data Science

CS-5630 / CS-6630

Fall 2022

An exploration into video game sales worldwide from 1982 to 2016

Process Book

**Karl Buckley (u1417778) - Chris Stevenson (u0679323) - Dominic Malouf (u1194321)**

Contents

Overview and Motivation.

When looking for a dataset to work with, the group was very open to suggestions as there were no overriding motivations or existing project ideas. We wanted to find a dataset that would be fun to work with, show interesting data and leave us with some innovative and different visualization style options. Having some background in video games meant that each team member would have some experience and frame of reference for the data and the project can be utilized by anyone wanting an insight into video game sales by region.

With the famous 1983 video game crash, many fledgling studios went bankrupt, and many people lost their careers. What effect did that have on global video game sales? How did the industry bounce back? These are the primary questions that we as a group wanted to investigate, however, this data also provides answers to a lot of other questions such as:" Which region had the most sales of X video game?" or "Which video game sold the most copies between 2002 and 2005?"

This data will enable us to build a picture of the Video Game sales landscape between 1982 and 2016 to provide trends and patterns that show which games sold the best, where they sold and maybe we will get some insights into why games sell differently in each region to build an idea of the video game markets globally.
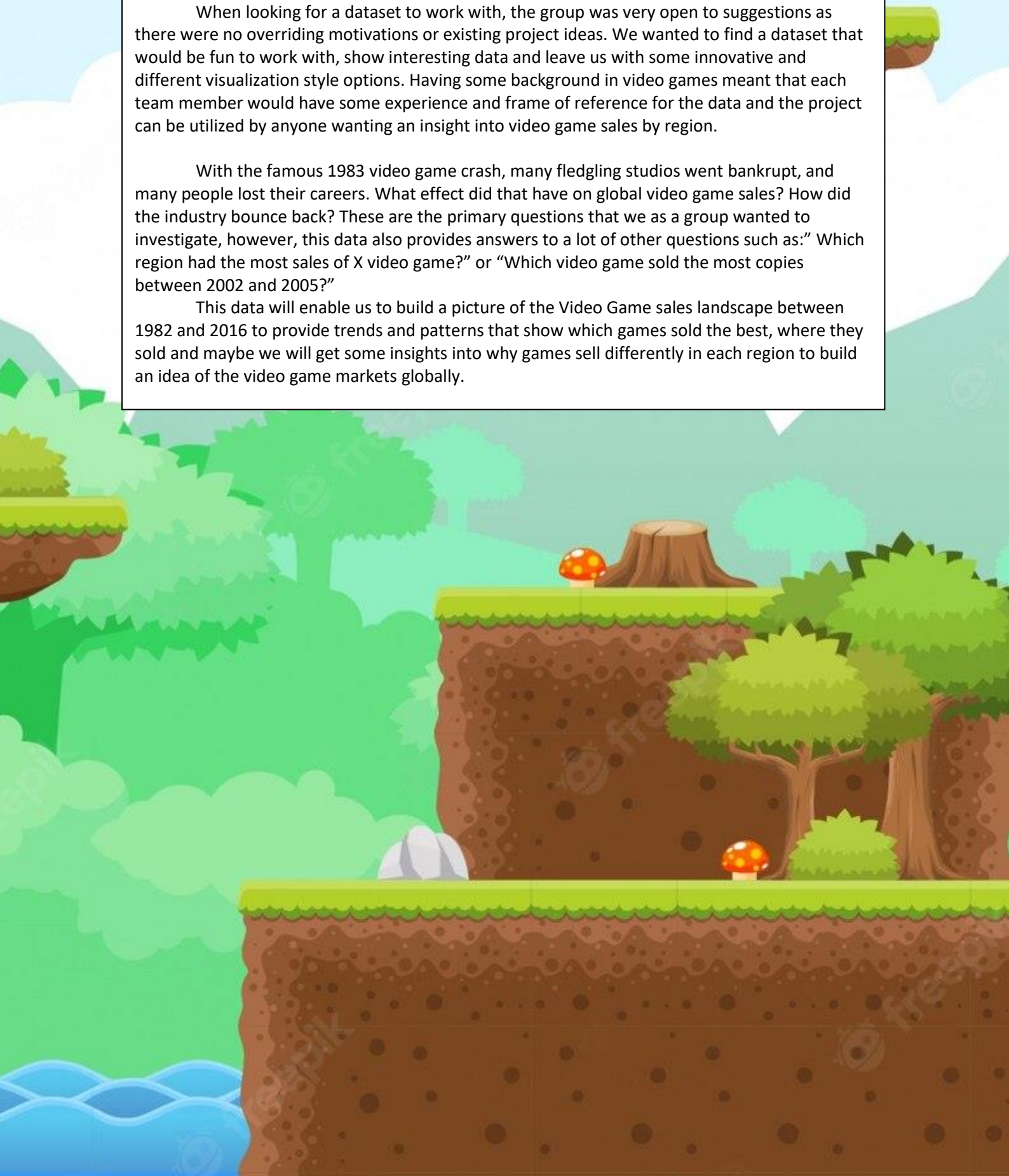
Related Work

One of our biggest influences and the closest match to our own site in terms of content was found at Statista.com. The site is full of statistics and visualizations for many different datasets, but the information provided for our subject area was very well displayed, current information being used and easy to read and understand. The simplicity of the bar charts paired well with the data as comparisons between data could be made very easily and the results are self-explanatory.
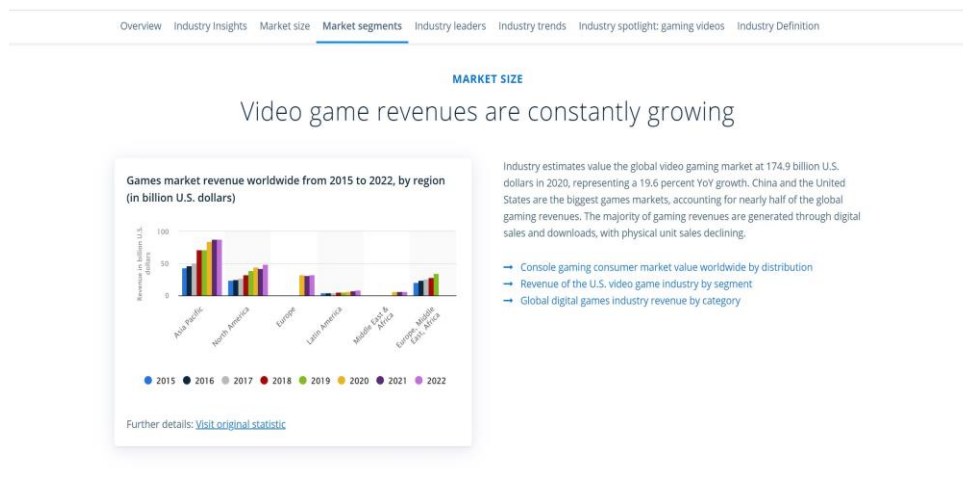
Borrowed kindly from ©Statista.com

Another website that we took some influence from was newzoo.com which has an article that details the global video game market in 2016. The article has many visualizations of datasets like ours, but this site uses a number of different techniques to show the dataset. These were all ideas that we had to choose from, which meant deciding on Pro's and Con's for each design.
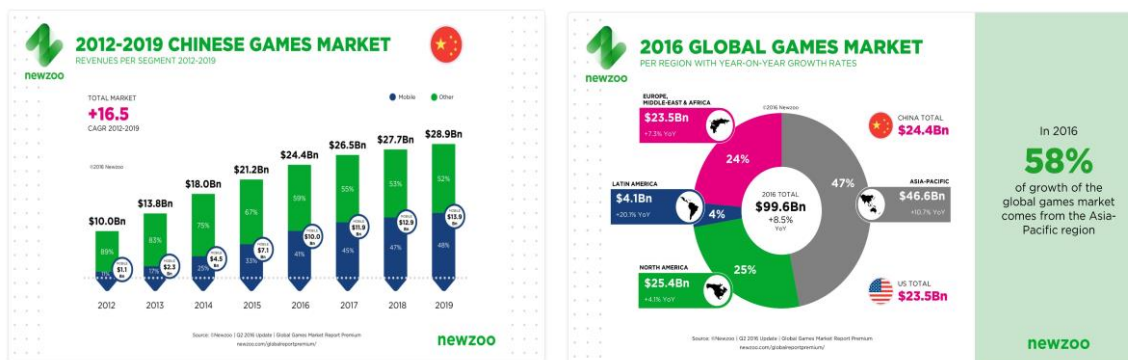
Borrowed kindly from ©newzoo.com

Data

We are sourcing our data from the online data repository Kaggle. It is authored by A. Ingle and there is no accreditation as to where the data was originally collected from. Here is the Kaggle link: https://www.kaggle.com/datasets/atharvaingle/video-games-dataset

The dataset is described as: "This dataset has the sale history of most popular video game year and region wise" and contains data structured as below, in a .csv file.

- Rank => Rank of the game
- Name => Name of the game
- Platform => Platform on which the game can run
- Year => Release year
- Genre => Genre of game
- Publisher => Publishing company
- NorthAmerica_Sales => Sales in North America
- EurpeanUnion_Sales => Sales in European Union
- Japan_Sales => Sales in Japan
- Other_Sales => Sales in other countries
- Global_Sales => Total Global sales

For this project, the group does not foresee needing to perform any data processing further than what has already been supplied. It may be required for us to remove the "Other" data if it shows during the data analysis that we only need the "Global" sales figures and the three main sales markets.
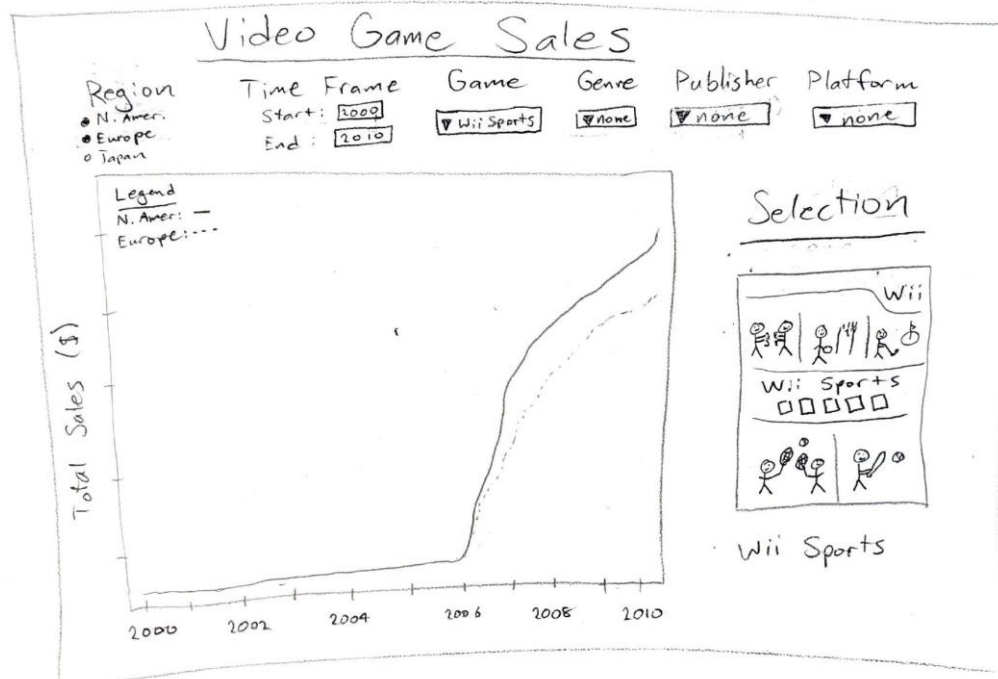
Design Evolution.

The visualization needs to be compelling as it displays statistics of sales numbers and given the subject matter, it should allow for bright, eye-catching visuals and interactions.

The most important feature of the visualization is that it must display a comparison of sales figures for each game against each region and each game. Once we have that main data, it is possible to display this data in a clear, concise but enjoyable format that encourages exploration of the data. Without this functionality, the project will have failed to educate the viewer on video game sales.

Direct manipulation of the data will be the main feature that elevates this project, allowing the comparison of different data to provide insights into sales trends across regions. Allowing for the easy selection of data and providing an interaction that not only provides entertainment but also an insight into the data is the aim of the project. Being able to select multiple games and seeing how they perform against one another in each region will provide insights into player populations, spending habits and gaming trends.
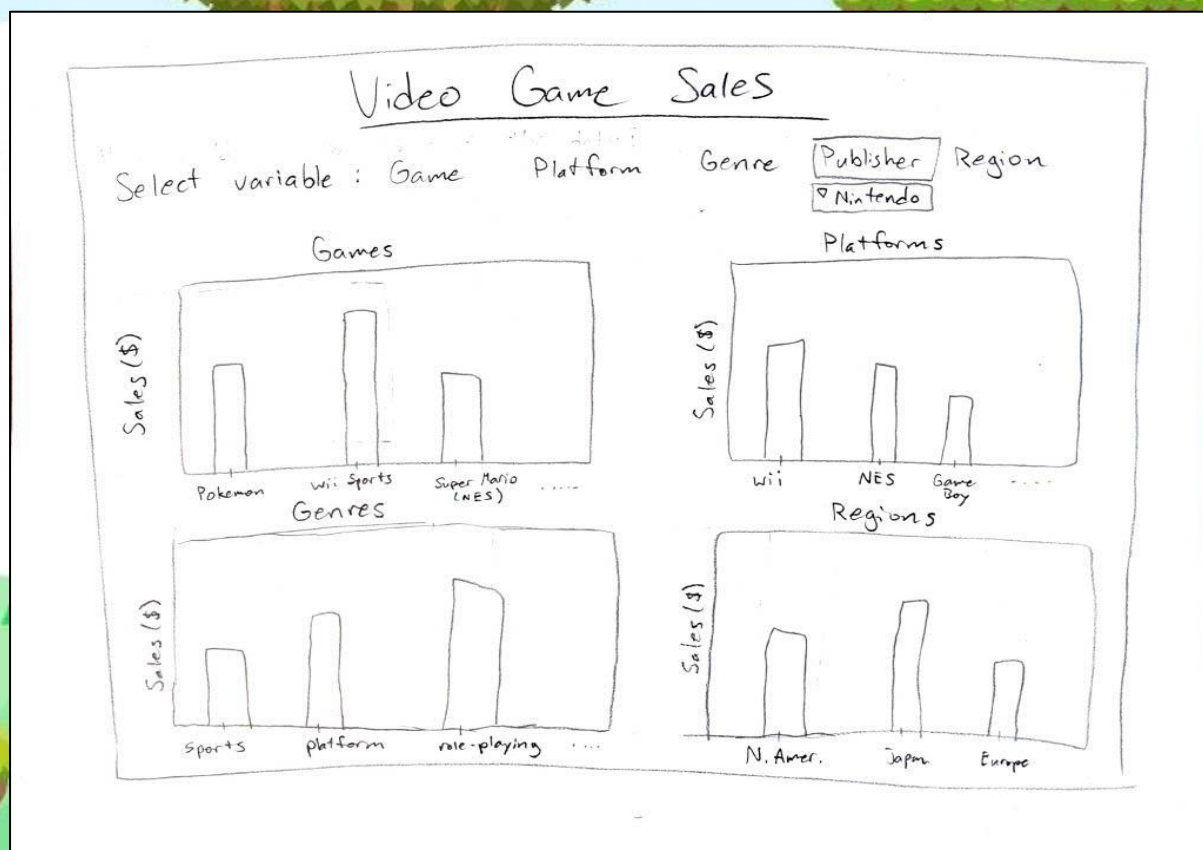
Design 1

Video Game Sales

Region
• N. Amer.
• Europe
○ Japan

Time Frame
Start: 2000
End: 2010

Game
▼ Wii Sports

Genre
▼ none

Publisher
▼ none

Platform
▼ none

Legend
N. Amer: —
Europe: ----

Total Sales ($)

2000   2002   2004   2006   2008   2010

Selection

Wii

Wii Sports

Wii Sports

This design uses variable selectors to modify a single graph. The tabs at the top will drop down and allow for the user to select which variable they would like to be visualized. When one is selected it could make the other menus unavailable (for example if a single game is selected, then a publisher could not be selected because each game has only one publisher). Some tabs can also have multiple variables selected at once (in this case the graph will add more lines and update the legend accordingly)
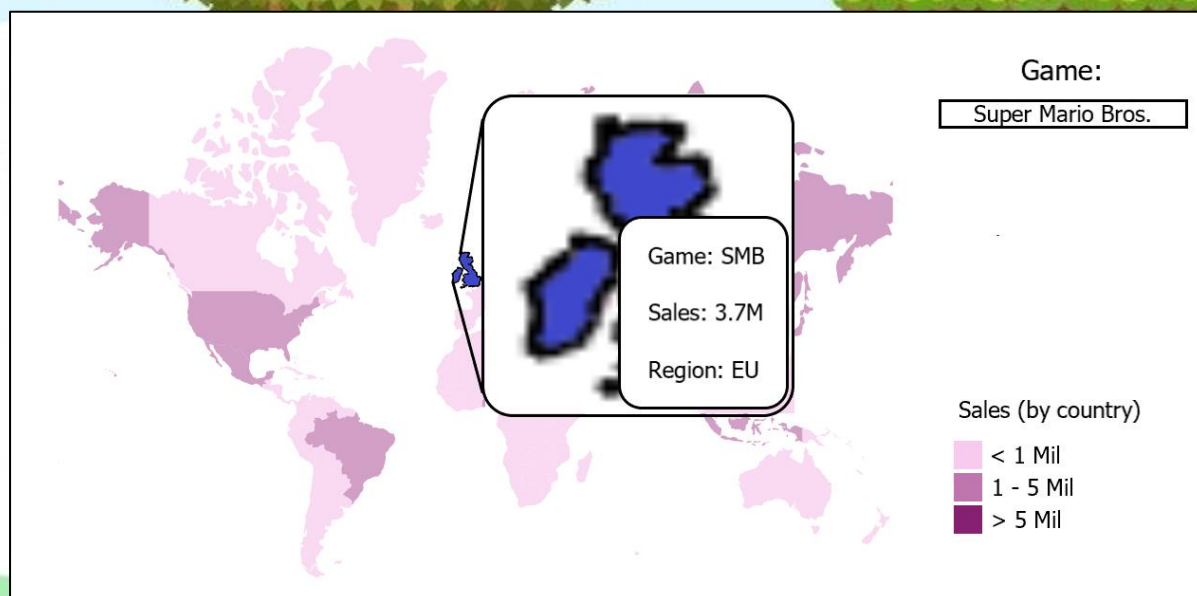
Design 2

This design uses a similar mechanic to the first design, but here only one variable is selected, and the unselected variables are shown as graphs. If a specific platform were to be selected, the graph containing platforms would switch to publishers. This design benefits from the data not being crowded to one graph as well as having the user interact with all graphs using the top menu.
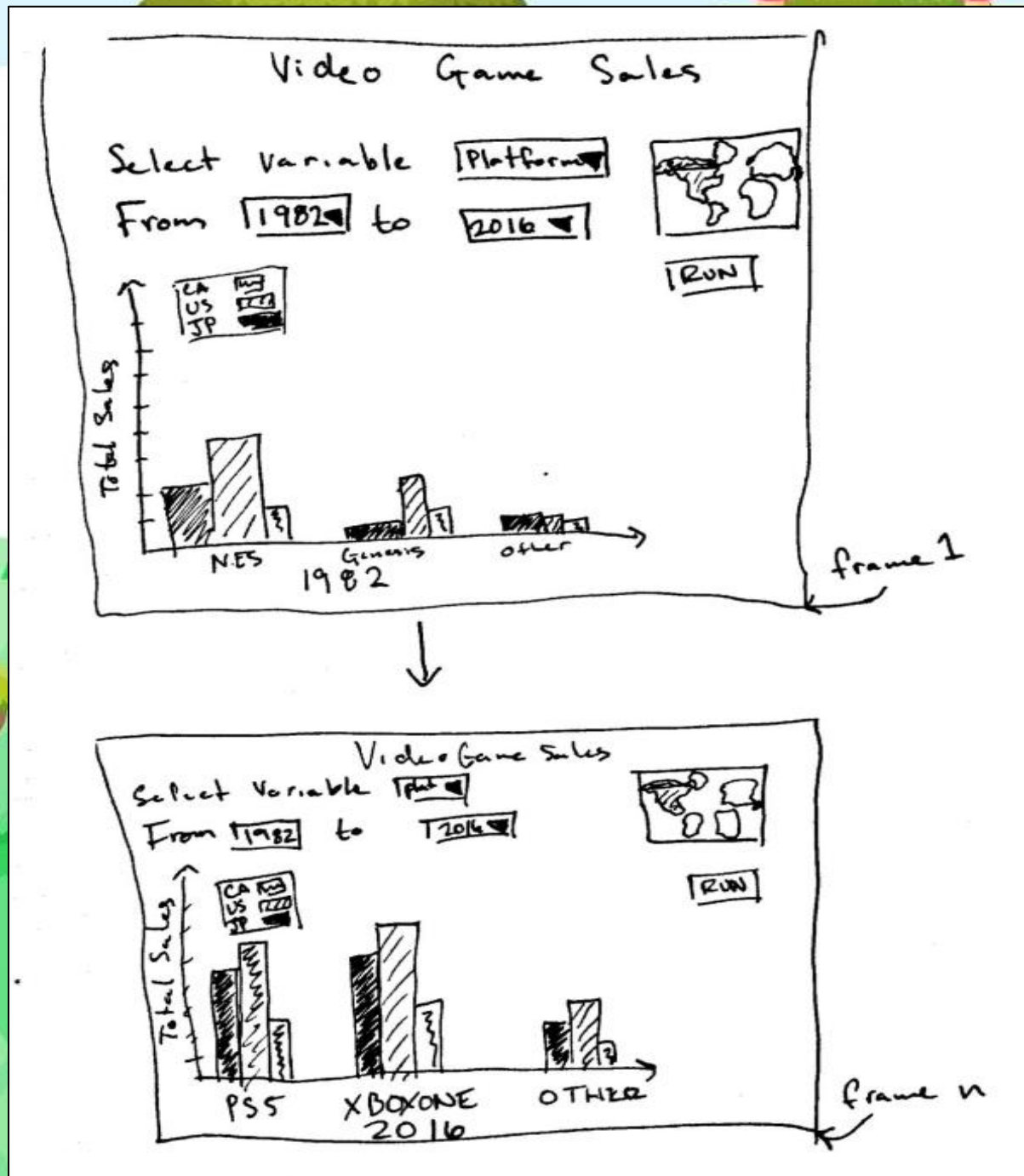
Design 3

Game:

Super Mario Bros.

Game: SMB

Sales: 3.7M

Region: EU

Sales (by country)

< 1 Mil

1 - 5 Mil

> 5 Mil

This design has been developed to display the world map with context sensitive bubbles that show further information regarding sales figures for each country. The design here is to promote exploration rather than tell a narrative about the data as it allows each viewer to drill down into the data for each game by publisher/year and number of sales. There exists the opportunity to expand this design further with a time scale slider to allow data changes over time and further comparison between publishers and individual titles.

Design 4

Video Game Sales

Select Variable [Platform▼]

From [1982▼] to [2016▼]

[RUN]

CA
US
JP

Total Sales

NES   Genesis   other

1982

frame 1

↓

Video Game Sales

Select Variable [plat▼]

From [1982] to [2016▼]

[RUN]

CA
US
JP

Total Sales

PS5   XBOXONE   OTHER

2016

frame n

This design uses an animation to display a selected time series. The user may select the variable for the x-axis and select a collection of countries. If no country is selected, then the continents (minus Antarctica) will be displayed instead. The animation transitions the bars in the chart depending on their value from year to year.
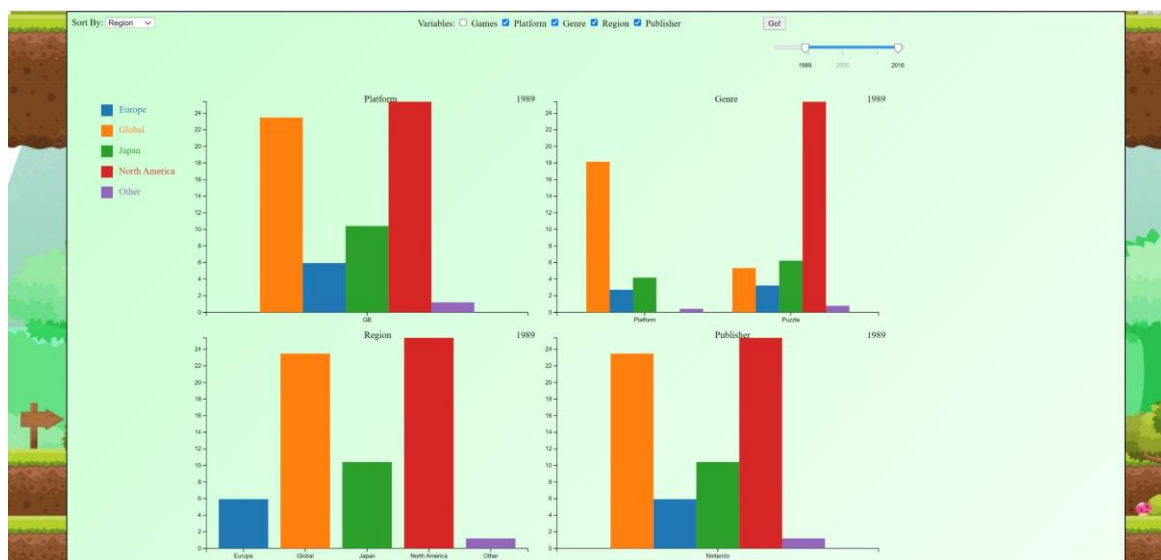
Final Design
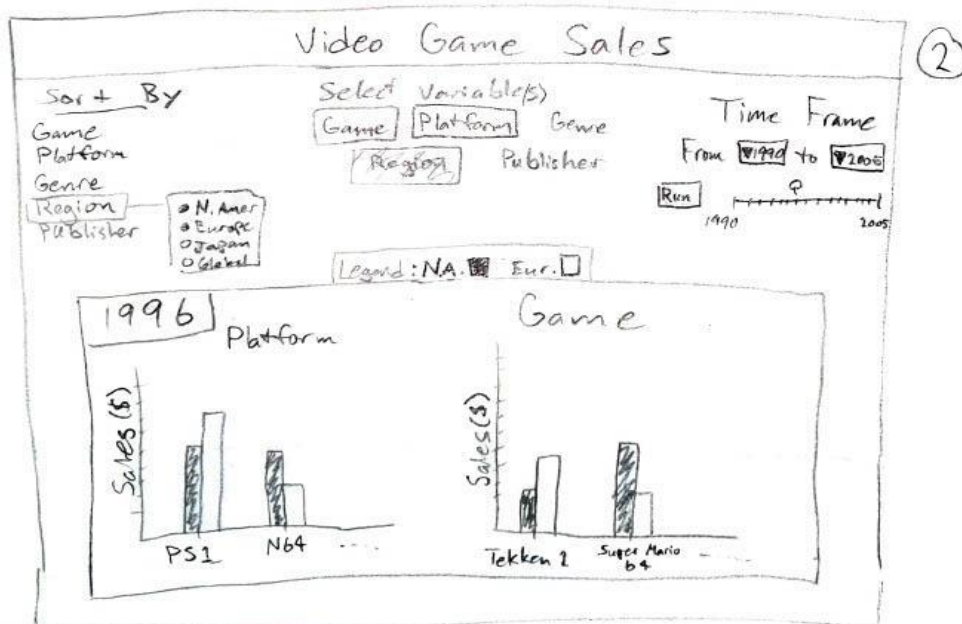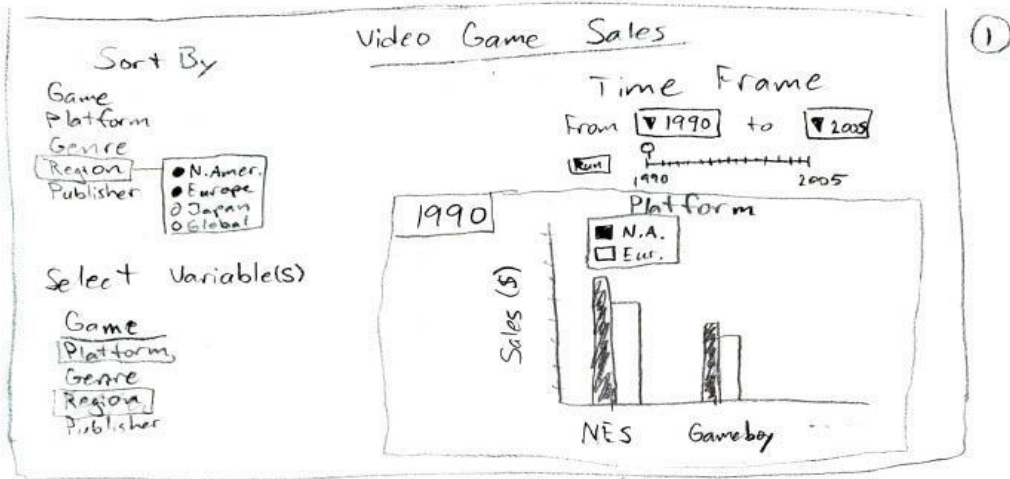
This design lets the user select a main variable to graph, in this case region. Graphs are then created by selecting from the remaining variables, each variable appends a new graph to the screen. All graphs are controlled by the Time Frame section. The year can be selected manually or hitting Run will animate the graphs from the time frame selected (same as design 4). The difference in frame 1 from frame 2 is that the controls are moved. These are possible spots for the variable selectors, but their spots on the screen are open to change. Same as with the section of the screen that houses the graphs and the legend.

Implementation.

As we've described previously, the main concept we wanted to stick to was that of an easy-to-use design that displays relevant data for the options chosen in a clear and understandable format. The theme was chosen to convey the fun nature of video games, so we used bright colors and have a background showing a scene from a videogame. Each of the website elements are made from the same theme and really push the gaming aesthetic.

The tool was implemented as bar graphs, which were chosen for their simplicity to understand, good fit for the datatype and easy manipulation, having seen them used effectively on Statista.com. Each bar graph represents one element we are exploring, and they are dynamically created as needed, being removed from the page when not relevant to the most recent search.

Implementing the search parameters for the dataset to display combined three different interactive elements: a drop-down selection box, tick boxes and a dual slider bar. Using three different elements lets the user decide what data they want to see and how.

The drop-down box element allows the user to select a main variable to graph, in this case region and sort all results by this choice. Graphs are then created by selecting from the remaining two elements, a tick box selection and a dual slider bar. With each variable selected, the tool appends a new graph to the screen.

Any graphs that have been created by the variable selection are controlled by the Time Frame dual slider bar. The years for the data range can be selected manually by dragging the sliders to the years requested and the graphs will update accordingly. Also, hitting the Run button will animate the graphs for the time frame selected.

Evaluation.

As this was a data exploration project, the main objective was to display the video game sales data so that subsequent studies and projects can use the tool as a reference. The finished project does fulfil the brief and our requirements and delivers clear visualizations of the dataset.

Results from the project are hard to quantify as the exploration of the data creates emergent questions and stories. There are some starting points that can be used to get an initial start to the process, such as: "What are the worldwide sales figures, by region, of the most popular video games between 1982 and 2016?" The results of this examination were of no surprise to any gamer and included famous titles, such as "Super Mario Bros" and "Halo".

This data also provides answers to a lot of other questions such as:" Which region had the most sales of X video game?" or "Which video game sold the most copies between 2002 and 2005?" which successfully builds a picture of the Video Game sales landscape between 1982 and 2016 and allows us to investigate further trends and patterns that show which games sold the best, where they sold.

The visualization does what the project set out to do. It uses the multiple variable inputs to return easy-to-understand results. Unfortunately, the data was unable to show us any indications as to why video games sell in each region and so we were unable to make any predictions.