# Rated *'V'* for *Visualization*

Team Process Book

Submitted By:

Raihan Ahmed, Harrison Fackrell & Andrew Frost

December 4th, 2022

## OVERVIEW & MOTIVATION:

Our team shares an interest in social trends and media, and this shared interest was the driver for our exploration of the subject-matter. The nascent medium of video games was a natural choice to explore further, given the underdevelopment of the industry and its documentation relative to other industries like music and film. Our project has since expanded in scope, but our original idea was to document the proportion of video games with a given content rating sold within each year–and in-turn, the trends associated with that proportion over time.

Upon review, we discovered that visualizations of industry-wide release trends for the United States are available directly from the Motion Picture Association of America (MPAA) and Recording Industry Association of America (RIAA) for the mediums of film and music respectively. However, no similar organization exists for the medium of video games, and no similar visualizations are publicly available. We are motivated to fill this niche ourselves, using independently-sourced data, for the sake of completeness and discovery.

In this project, we will explore published video game data and use the techniques learned in the course as well as our imaginations, to create an interactive visualization to highlight major trends in video game history. Our goal is to effectively illustrate how video game genres, ratings, and platforms changed over the years and in such a way that the user(s) can experiment with the visualization to draw their own conclusions.

## PROJECT OBJECTIVES:

Our project aims to illustrate video game data in a broad sense, using interactive visualizations to plot industry-wide figures on a per-year basis. These visualizations should be easy-to-use, especially for users who are unfamiliar with the industry, and we intend to build our

visualization as a flexible tool. Web visualizations have, as an advantage over visualizations constrained to static media, the capacity for extensive interaction, and we intend to leverage this to a high degree: at any given time, the information immediately in-frame should be limited enough to not be overwhelming, and user-driven options to explore related information should be clearly available. Users could later modify the graphic to examine the usage ratio depending on other categories, such as rating and genre. A stacked bar chart with years on the x-axis and total number of publications on the y-axis will initially be shown based on different platforms. To allow users to examine the usage of the games exclusively for that particular category, we provided options inside each category. The user will be able to assess the popularity of video game years using a waffle chart and legends since we made each and every bar clickable. The description of the selected bar in the waffle chart will be displayed. Our first design was to have a pie chart, but we later realized that a waffle chart would be much more effective. A waffle chart displays the amount of progress achieved toward a goal or the percentage of completion. A wonderful technique to display data in relation to the whole is via waffle charts. By hovering over the chart's components, you'll be able to see the different patterns.

We learned how to create interactive graphics with various charts based on a small subset of data extracted from a large set of data. We were able to deal with a large dataset. We also discovered how to link human thinking with the images, turning this representation into an interactive one.

## RELATED WORK:

We were inspired by the form-factor of the RIAA US Sales Database, and by the visualizations shown in-class depicting multiple levels of brushable and proportional bar charts. Also, the brushing feature discussed in class as well as the assignment felt like a good idea to try out on our interactive visuals.

## QUESTIONS:

Our project aims to illustrate published video game data in a broad sense, using interactive visuals to plot published figures on a per-year basis. Once the visualization has been implemented, the user will be capable to see trends and answer possible questions like:

- *"Which gaming system had the most game releases in the year 1998?"*
- *"What trends (if any) are there for video game ratings?"*
- *"When did a particular gaming genre become popular?"*

We feel questions like these (potentially others) still stand for the user to figure out by interacting with our final design. Of course, as the project timeline continued we developed questions to ask ourselves in while we developed our final solution. We wanted to make sure we were being clear with what we were trying to get across and how we can do it. There was also another question on how we can clearly describe our findings to someone who may not be familiar with video games. To address these questions, we thought it would be helpful to add supporting text and labels to our charts. Many well-developed visualizations have great attraction in their design and complexity, but of course they also have simple, well-detailed supportive text available to fallback on for the user to remember what the visualization is telling them and what is the key takeaway.

## DATA:

The largest data-set for video game records that is both publicly-accessible and up-to-date is maintained by the VGChartz Network, which has a web portal for browsing records here.

We performed a significant data cleanup, as our source did not provide its data in a use-friendly format. Fortunately, there exists an open-source web-scraping Python script for the VGChartz online database, which uses the BeautifulSoup library to deliver a csv-formatted file containing the following fields:

```
['Rank', 'Name', 'basename', 'Genre', 'ESRB_Rating', 'Platform', 'Publisher',
'Developer', 'VGChartz_Score', 'Critic_Score', 'User_Score', 'Total_Shipped',
'Global_Sales', 'NA_Sales', 'PAL_Sales', 'JP_Sales', 'Other_Sales', 'Year',
'Last_Update', 'url', 'status']
```

Unfortunately, this script was last updated in April of 2019, and it required substantive updates to be performed by us before it could be made usable again. Most notably, the proxy service it relies upon is now defunct, and the multithreading behavior tends to send requests too quickly, causing the scraper to be temporarily throttled. We ended up solving these problems by removing support for the proxies and multithreading altogether, which makes progress much slower, but much more reliable–and which has the arguable benefit of reducing the rate at which we make requests, lessening the strain on the server. Once we restored the scraper to working condition, we used it to collect 57,000 records. Data source files can be found here in the data folder, and the scraping source can be found here.

## EXPLORATORY DATA ANALYSIS:

Lacking any sophisticated visualization tools, we explored our data using a simple tabular spreadsheet program. We discovered that not every game has information available for every field, and some fields are inconsistently represented: for example, some games had platforms listed as “Series” or “All,” or simply had no information in the field. Similarly, some games had no year-of-publication on record, and there was a sharp drop-off in available records after 2018.

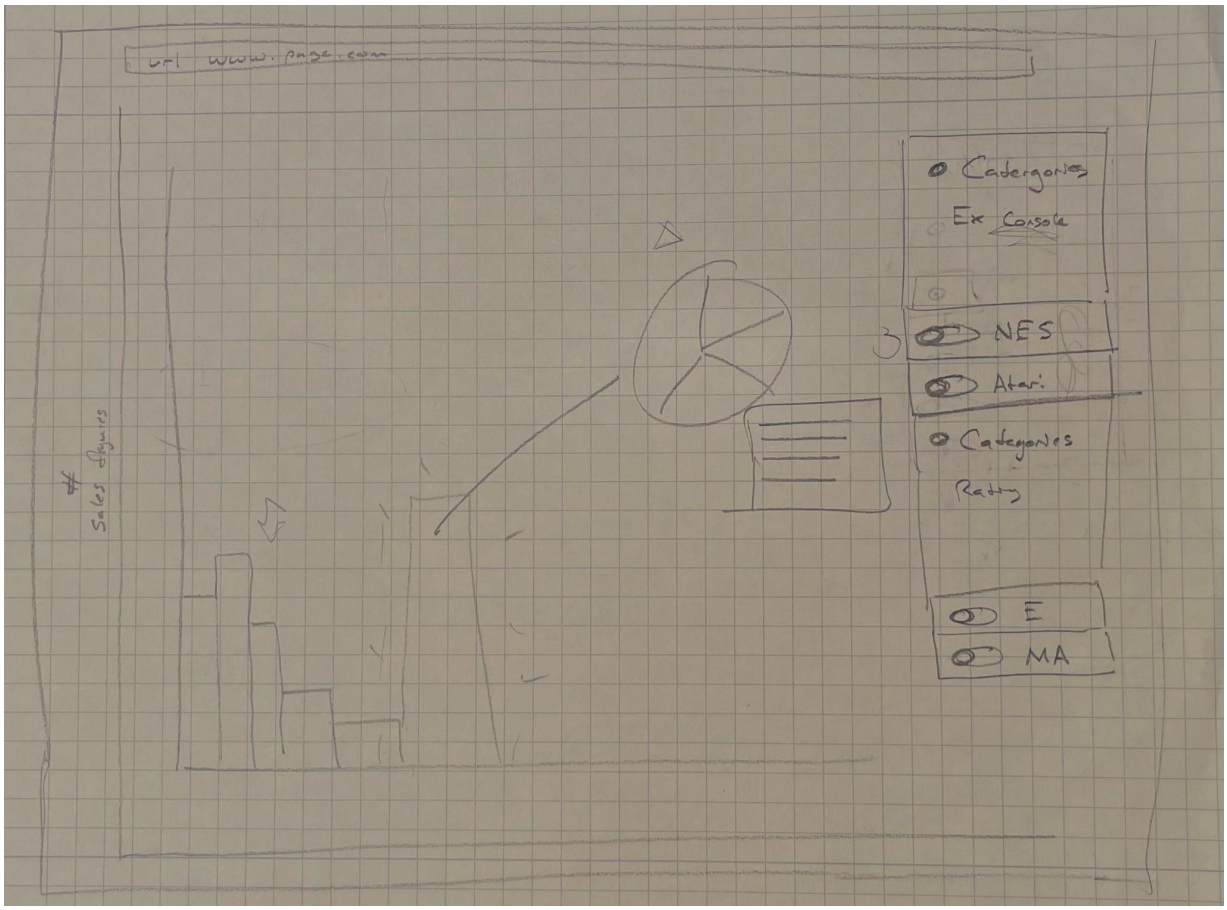
We decided to respond to this by clamping the span of time represented by our visualization: 1977 to 2018. We also excised any records that did not have accurate release-date or platform information. Upon review, most of these records appeared to be either duplicates of existing records (as was the case with multi-platform releases, which had records for “All” in addition to individual per-platform records) or too general to be useful (i.e., records for entire series, instead of individual titles).

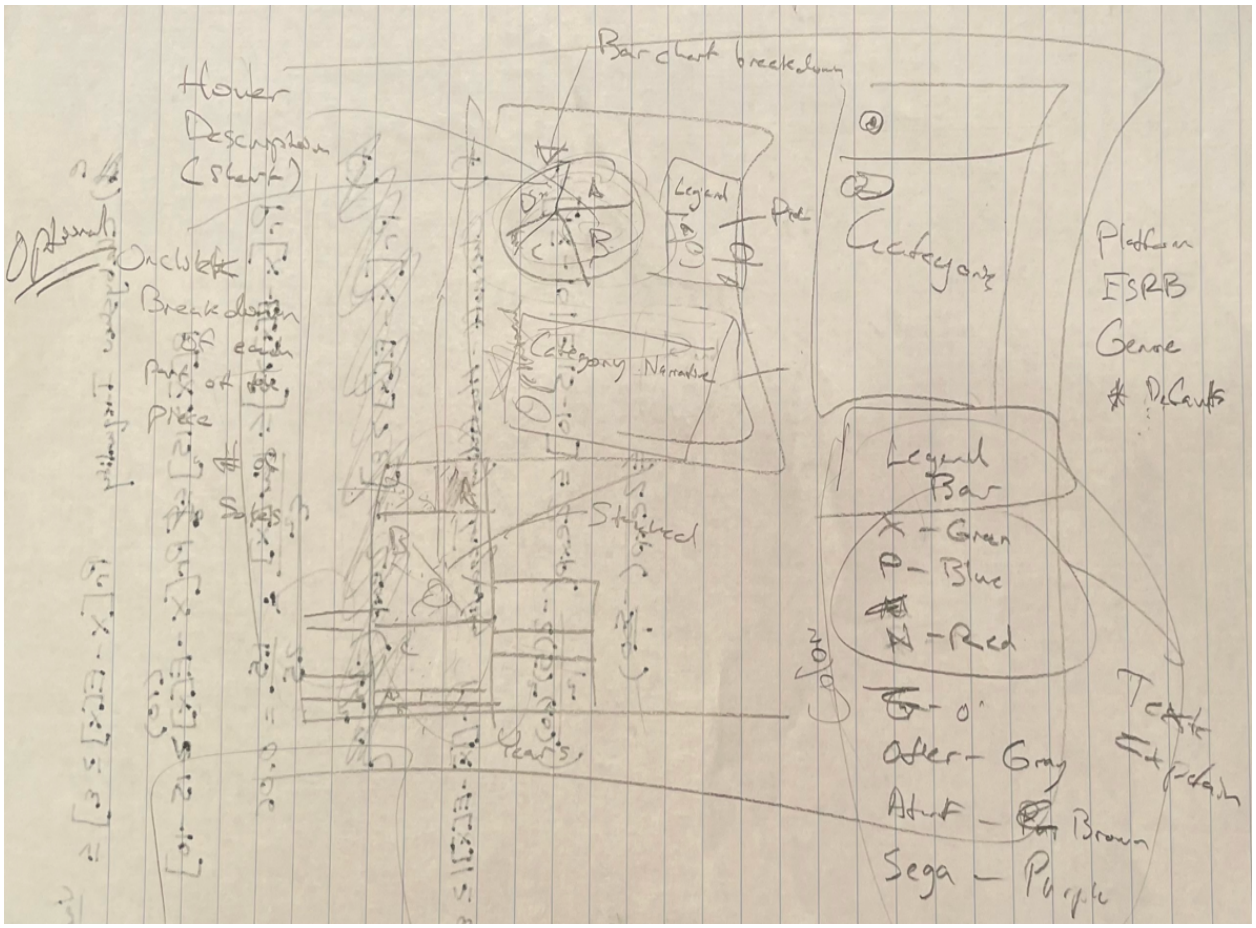
## VISUALIZATION DESIGNS:

Simplicity is the key to unlock dense information with visualization. In this project, the visual will be an interactive bar chart which can be interchanged based on the category selected by the user. Once the user has selected their desired data category, they can filter on one or multiple items that fall under the category.

*Example:* The category selected is ‘Platform’. There are several ways a gamer can play games. The usual platform systems used are consoles, mobile phone/tablet, and personal computer (pc). The user can filter specifically on one system or multiple systems and see the visual output.

Hand drawn sketches:






(Note: Hand sketches were done to brainstorm the alternative prototypes prior to them being implemented in Microsoft Paint).

The following are <u>three</u> alternative prototype designs:

# DESIGN #1


For our first design, we wanted to start off simple, organized and direct. It was determined that a traditional bar chart visual would be the best approach to display different categorical data groups. We have in the top-right corner a ‘Categories’ menu that will have several options to choose from. Once a category is selected (radio button), a drop down widget will appear to present several sub-categorical options to filter down by a specific item/trait. Once the filtering has been completed by the user, the bar chart will update the results based on the combination of filters. Next, the user will have the ability to click/select a particular bar to see the makeup of the bar in a waffle chart. This chart will include a breakdown legend of the different pieces of the waffle chart as well as an accompanying information field containing narratives, stats, and key facts. This first design serves as a solid foundation to build on for the final design.

# DESIGN #2


For our second design, we were thinking of other possibilities for our sub-visualization. When the user wants to select a particular bar a bubble chart will appear, similar to the waffle chart in design #1. Also, hovering over the visual with the mouse cursor felt much more interactive than clicking in/out of individual bars.

# DESIGN #3


For design #3, a vertical-dashed line was rendered for additional interaction with the user. When the line is overlapping a desired bar the sub-visual would appear. However, the sub-visual has since been discarded. It was determined that a bubble chart would be a valiant option to use, but the limited size of the pop-up box would put a greater limit on the chart's scale. This possibly could put unneccesary strain on the viewer's gaze and possibly hide key takeaways for the selected bar.

# FINAL PROPOSED DESIGN


Key-strengths from each of the design prototypes were discussed and approved for the final design as well as adding a few new options. Stacked bars seemed the best way to distinguish the different gaming platforms' sales performances over the span of 34 years. A legend for the stacked bars will be included in the bottom-right corner accompanied with supportive text. This will support the visual with organization and informing the viewer before/during usage.

Like the other designs, hovering was deemed the simplest and most convenient way for a user to interact with the visualization. The rendered line was an idea for the third design, but it was deemed as an “over-need” for simplicity in the visual. A new feature added for the cursor is when the cursor clicks on a desired bar, the bar increases (smooth animated transition) and stands out in scale compared to the unselected bars. Next, the user can hover over the selected stacked bar and a pop-up will appear. The pop-up box will still include condensed, clear narrative information to educate the viewer of the determined findings. The box itself will increase in size and will overlap (brought forward) over the unselected bars in the chart. The sub-visual (waffle chart confirmed) will be inside the box once the hover is over the desired bar group. This visual will include features of correlating the same color of the bar with the leading driver, categorical percentages, and narrative information for each slice of the waffle. When the viewer clicks a slice of the waffle, a descriptive breakdown of subcomponents of that slice will appear in another pop-up box which contains additional sub data correlating to the parent categorical slice. These features will give a complex depth to the visual and give the viewer freedom to try various searching combinations.

## MUST-HAVE FEATURES:

We created a list of must-have features our final design must have. They are as follows:

1. The basic bar chart is drawn, with each bar representing a year, and its height representing overall video game sales figures within that year.
2. The bar chart is updated to include or exclude data-entries according to the selected toggles.

3. The individual bars feature stacked groups, based on the currently-selected categorical radio-button.
4. A dynamic legend is drawn and updated according to the currently-selected categorical radio button and included data entries.
5. The bar-chart is interactive: clicking on the bar chart displays a smaller sub-visualization that includes a waffle chart encoding of the information local to that bar.

## OPTIONAL FEATURES:

The optional features are listed below:

1. The y-axis may be toggled between two different, related quantities: number of games published, and number of units sold.
2. Hovering over a waffle chart segment displays a small block of text describing the category it represents.
3. Every waffle chart sub-visualization is split into smaller sub-categories, where applicable.

## DESIGN EVOLUTION:

As we continued to make progress to our final solution, we decided as a team to make changes when they felt appropriate. We deviated from the design proposal in the following ways:

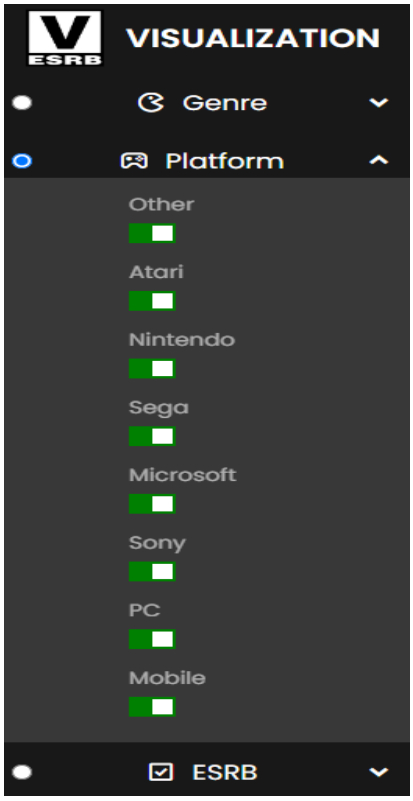
- The visualization's data-selection tools are now located on the left, instead of the right.

- There is now an additional bar-chart, above the primary one, which displays a proportional view of brushed bars in the primary chart.
- For our subvisual (when the user selects a bar to analyze), we decided not to pursue a pie chart. Instead, we felt a waffle chart would bring simplicity and clear visuals for the breakdown of our sub data. This was a helpful suggestion given by our assigned TA in our project milestone.

The largest change was the addition of a proportionality view chart, which we decided to implement after observing a similar visualization in class. The goal behind this change is to make proportional trends more visible by making it easier to compare the information encoded between years: when two marks are more closely aligned, it is easier to judge their relative lengths; and providing an option to view the data with a proportional encoding makes it much easier to see marks that would otherwise be very small.

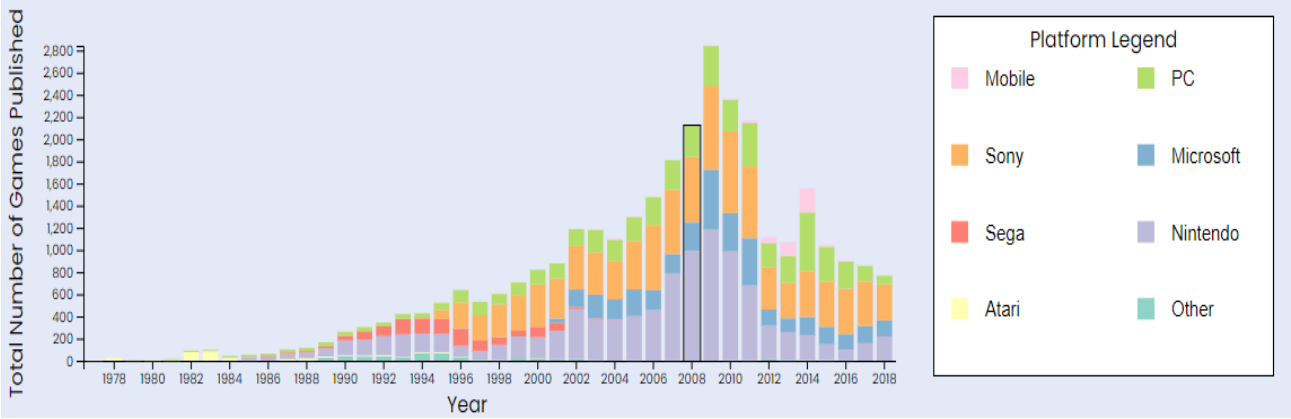
## IMPLEMENTATION:

Going back to what was mentioned back in the proposal, “simplicity is the key to unlock dense information with visualization”. A simple look with an effective interaction showing multiple levels of information is the intent of this visualization. First, we created a categorical menu.

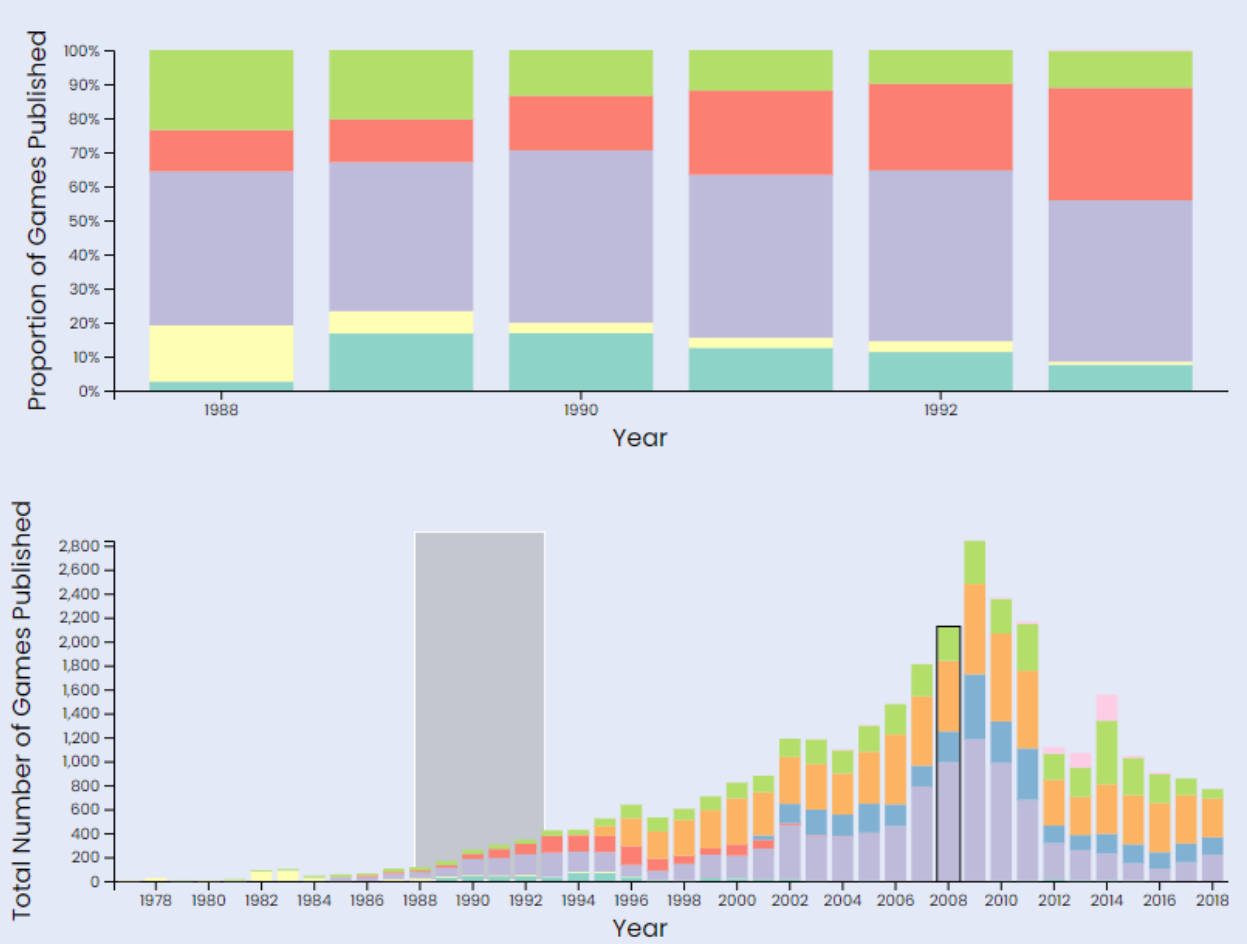



In the image above, categories are broken down by platform, video game genre, and Entertainment Software Rating Board (ESRB) rating. A radio button is included on the left side of each category header when selected, the data on the right side will change by the data output based on the category. Also, the user can toggle between different choices based on the category selected and filter by one or more options.

Now with a specific category selected and toggle filters specified, we then get a stack bar view of the data.

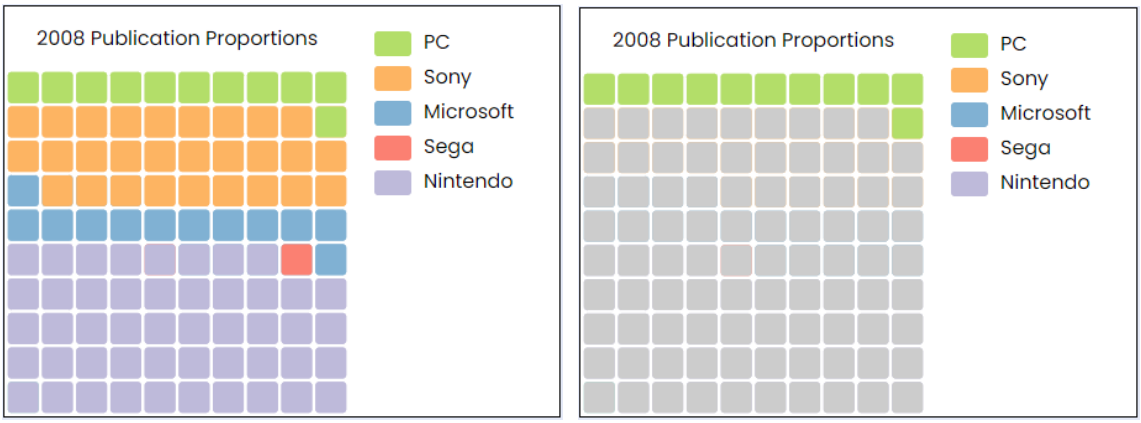



In the image above, the data ranges from the late 1977 to 2018 in a stacked bar chart. Each year is broken up by the number of video games released and are identified with a color scheme on the right side. The legend will update based on the category selected and different color schemes will be used to identify each group based on platform, genre and rating. Above this stacked bar chart, a second visual will appear once the user brushes over any specific number of bars in the stacked bar chart.




This second visual is another way to interpret the proportion (also in stacked bar chart form) of the selected data. As we see in the image above, the brush selected the years 1988 through 1993. We can see that these years don't have as much data compared to later years. However, the user can identify the percent distributions of game releases. This gives the user an opportunity to see trends in releases and how they change, and also allows the user to more clearly observe bars that would otherwise be difficult to see, on account of their size.

The third visual is the waffle chart we mentioned that breaks down the percentages of both the total and proportion charts.




When a particular year is selected (ex. 2008), the waffle chart updates with the breakdown of percentages of the legend items. Also, if a user wants to zoom in on a particular item (ex. PC), the user can hover over that item's color in the legend and then all other items in the legend will be grayed out to focus on the desired field.

## EVALUATION:

In the conclusion of this project, we feel confident in our design to be aligned with the vision we had for this project. While implementing the design, we could see the questions we had for ourselves going into this project being answered with every change. We found ways to address the interaction experience concerns a user may have with our visualization. Including our axis labels and appropriate titles to each view, a user will know what to do and have the expectations of what will happen before doing so. We also can see several trend collections between different views (Platform, ESRB and Genre) of our stratified data. These discovered trends may answer some of the questions a user may have while exploring the visual.

Since the ESRB's inception, the relative popularity of all ratings has grown in a "bottom-up" fashion, with lower ratings being widely adopted first, and higher ratings being

adopted later. Interestingly, since ~2009, we've seen a relative *decrease* in games with content rated E for Everyone, and a relative increase in Unrated and Mature-Rated games. Some of this can be explained by the introduction of the E10 rating in 2005–a rating that provides an additional graduation between E for Everyone and T for Teen–but even the combined proportion of E- and E10-rated publications is lesser in 2018 than the proportion of E-rated publications alone in 2008.

In our genre view, we can see steady growth of several video game genres. For example, we can see the popularity of sports games over the years. Starting in the early 1980's and going forward, sport games steadily grew for releases on several popular platforms. Interesting enough, in 1982 Electronic Arts (EA) was founded. EA is known for the production of popular sporting games like Madden and FIFA.

We also saw from the visualization the rise and retirement of popular video game platforms. We could see clearly when popular platforms dominated the market like Atari in the 1970s and 1980s following with Nintendo and Sega in the 1990s. A further improvement to our visualization could be adding interactive timelines to see when a certain platform was introduced and when it fell out of touch with consumers. Same would be applied with genre and ratings. A user can compile these findings on their own, but additional breakdowns would be helpful in speeding up their results.