Rated *'V'* for *Visualization*

Team Process Book

Submitted By:

Raihan Ahmed, Harrison Fackrell & Andrew Frost

December 2nd, 2022

OVERVIEW & MOTIVATION: Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.

Our team shares an interest in social trends and media, and this shared interest was the driver for our exploration of 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 sales 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, but 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.

RELATED WORK: Anything that inspired you, such as a paper, a website, visualizations we discussed in class, etc.

QUESTIONS: What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?

Our project aims to illustrate video game sales data in a broad sense, using interactive

visualizations to plot industry-wide sales figures on a per-year basis. Once the visualization has been implemented, the user will be able to answer certain questions like:

- "Which gaming system had the most sales in the year 1998?"
- "What trends (if any) are there for Entertainment Software Rating Board (ESRB) ratings?"

DATA: Source, scraping method, cleanup, etc.

The largest data-set for video game sales 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 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 10,000 records. We plan to use it to collect an additional 52,000 records before submitting our final visualization, but this is

Data source files can be found here, in the data folder.

EXPLORATORY DATA ANALYSIS: What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?

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). After extracting these records, we were left with approximately 9,750.

DESIGN EVOLUTION: What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?

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, beneath the primary one, which displays a proportional view of brushed bars in the primary chart.

The largest change was the addition of the extra "zoom 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: Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.

EVALUATION: What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

At-present, our visualization features a narrow view of ESRB rating trends. We have added support in our internal bar-chart class for visualizations of arbitrary data, but our navigation bar has yet to receive support for toggling between them. Because of this, the conclusions we can draw from our current product are limited.

Nevertheless, some clear trends are apparent. 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.