Contents

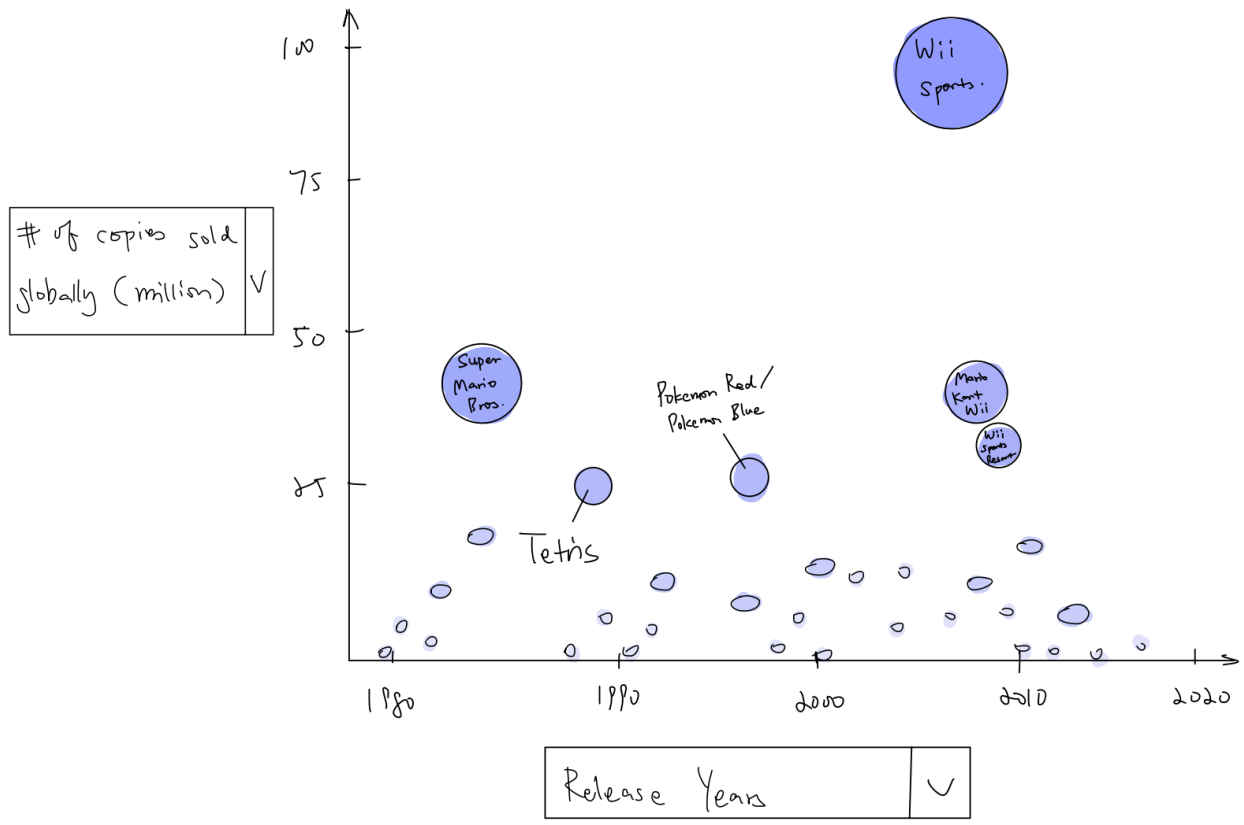
Dataset Description, Source, and Cleaning Procedure

For the final project, we used a public dataset titled “Video Game Sales”, available on Kaggle (Dataset). Our dataset is a table. It contains the video game sales data across platforms, genres, and regions from 1980 to 2020. Each item or row represents the name of a specific video game, such as Super Mario Bros. The dataset also includes attributes describing its release platform, year, genre, publisher, and sales figures in millions of units. The nine attributes include the **Platform** (categorical), which identifies the console the game was released on, such as PC or PS4. **Year** (ordinal) indicates the year the game was released. **Genre** (categorical) describes the category or style of the game, such as Action or Sports. **Publisher** (categorical) names the company that released the game. The sales figures are broken down by region: **NA_Sales** represents sales in North America (in millions of units), **EU_Sales** shows sales in Europe (in millions of units), **JP_Sales** captures sales in Japan (in millions of units), and **Other_Sales** includes sales in other parts of the world (in millions of units). Finally, **Global_Sales** provides the total worldwide sales for the game.

Our video game dataset can be expressed using several visualizations such as Bar charts and Scatter Plots. Bar charts are useful for comparing total sales both globally and by different regions across different categories such as genres, platforms, and publishers. They make it easy to see which genres or companies sold the most games and how different platforms performed against each other. Scatterplots help show how video game sales have changed over time. By plotting sales data against the year of release, we can identify patterns like years with very high sales, rising trends in the popularity of gaming, or years when the market slowed down.

During our cleaning process, we found a small number of missing values in the Year and Publisher attributes, marked as “N/A” in the CSV file. Out of 16,599 row entries, 329 rows contained such missing data — less than 2% of the total. Since the proportion of missing values was small, we decided to remove those entries entirely to preserve accuracy and make sure the trends we wanted to show were not misleading. After deleting the values, we were left with 16292 rows. We had 271 “N/A” values in the Year attribute, 58 “N/A” values in the Publisher attribute, and 22 “N/A” values that overlapped between the two, so the total values we deleted were 307 (271+58-22). Additionally, we noticed that some games were listed more than once due to being released on multiple platforms. We chose to treat each platform-specific release as a separate game. This allowed us to better highlight differences in performance across platforms. Our dataset also had a Rank attribute that was predetermined based on global sales. We decided to ignore the ranking data to make the tasks of the visualization more dynamic.

Visualization Prototype & Design Choice



Prototype 1: Scatterplot

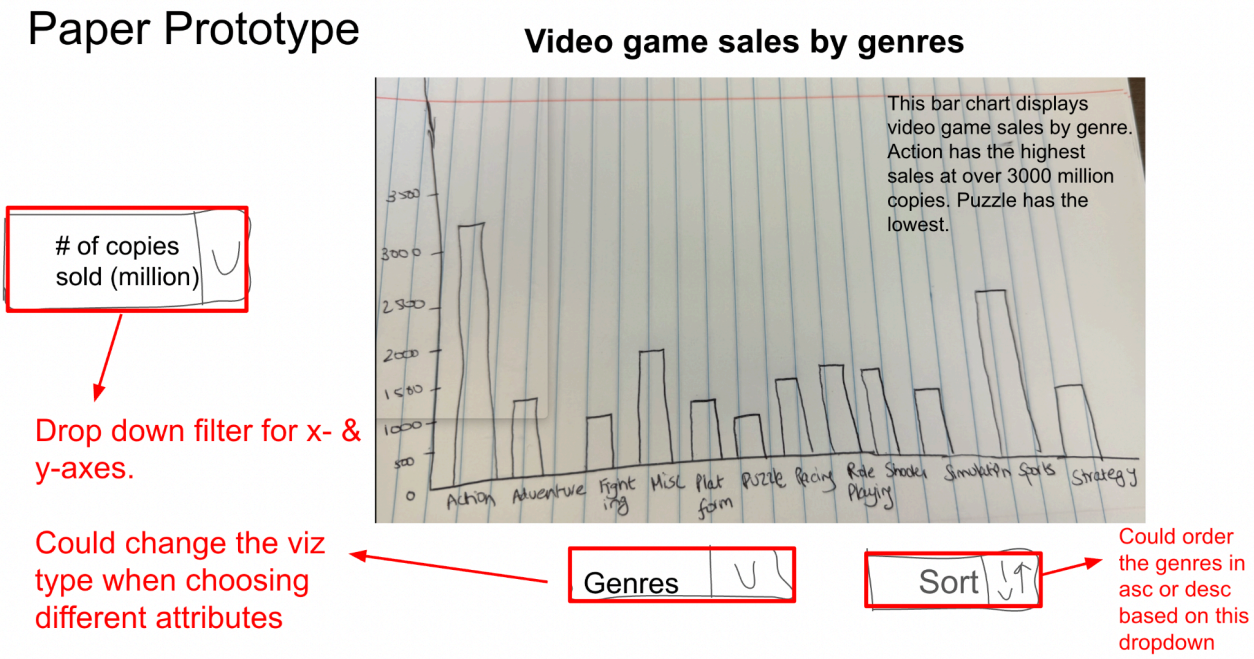
Our figure visualization prototype is an interactive scatter plot. Initially, when drawing the prototype, we designed it to show the top 20 most/least sold games as each individual bubble but then we decided to show each year as an individual bubble, displaying the total sales of the year depending on regions. This decision was made because we wanted to show an overall trend and distribution, instead of only 20 individual game items on the default scatter plot.

Overall, the x-axis represents the year in ordinal order, and the y-axis shows the number of sales in millions of copies sold. We chose a Scatterplot because it can display the overall distribution of video game sales across years.

For the interaction, we also incorporate a drop-down button for regional sales. For example, the user can click on the button and select regions such as Global, North America, Europe, Japan, and Other Regions to see the distribution of video game sales filtered by those regions across years.

In the final version, we incorporated the metric button for looking at our visualization at a different angle: the number of games *released*, instead of the number of games sold. Using this interaction, the user can view the visualization completely differently, switching from total video game sales to the number of video games released. We added extra data processing in our code when initiating the visualization to make sure the original data is aggregated to give us this artificial data – meaning that the data point for the number of games released each year did not exist until we processed it.

Beyond all of those above, the user can interact with the chart itself by hovering over the bubbles. The title would show sales/number of games released each year in total. Working through it, we went back to the original prototype and thought about adding individual game information for the tooltip. Therefore, we displayed the top 3 games sold in each year in the tooltip information for the sake of nostalgia for our original design and to give users more fun details to explore.



Prototype 2: Interactive Bar Chart

Our second prototype is an interactive bar chart that shows us the video game sales by different genres. The y-axis displays the number of video game copies sold in millions. The x-axis shows different genres of video games like Action, Sports, or Adventure. We chose a bar chart because it clearly displays comparisons between categories, making it easy to identify which types of games sold the most or least.

To make the chart more interactive, we included dropdown filters for both the x and y axes so that the user can select what they want to look at. For example, instead of genres on the x-axis, users can choose to see sales by platforms (like PlayStation or Xbox) or by publishers (like Nintendo or Electronic Arts (EA)). They can also change the global sales metric on the y-axis to look at regional sales in North America, Europe, Japan, or Other regions. There's also a sort button that helps organize the bars from highest to lowest and vice versa. This makes it easy for the users to spot the top five best-selling genres, or see which ones sold the least. To make the chart more accessible, we included a short written summary that explains what the chart is showing. This helps users who may not be able to read the chart easily or who use screen readers.

In the final version, we incorporated the metric button for looking at our visualization at a different angle: the number of games *released*, instead of the number of games sold. Using this interaction, the user can view the visualization completely differently, switching from total video game sales to the number of video games released. We added extra data processing in our code when initiating the visualization to make sure the original data is aggregated to give us this artificial data – meaning that the data point for the number of games released each year did not exist until we processed it.

Beyond all of those above, the user can interact with the chart itself by hovering over the bars. The title would show sales/number of games released each year in total. Working through it, we went back to the original prototype and thought about adding individual game information for the tooltip. Therefore, we displayed the top 3 sold games in each genre or by each publisher or on each platform into the tooltip information for the sake of nostalgia for our original design and to give users more fun details to explore.

Task Analysis

Users can perform three primary tasks using our video game sales visualizations, based on our scatter plot and bar chart paper prototypes. Each task is analyzed using marks and channels. We also include a detailed accessibility analysis to explain how low-vision users might accomplish these same tasks using non-visual cues.

Task 1: Compare Values

Question: Which genre sold more video game copies regarding Global Sales — Action or Puzzle?

- Visualization Used: Bar Chart
- Mark: Bar
- Channels: Vertical position (bar height)

How the user completes the task:
The user is cued to choose **Genre** in the attribute selection button, **Global** in the region selection one, and **Total Sales** in the metrics selection button. To compare the number of copies sold between two genres, the user first identifies the "Action" and "Puzzle" bars along the x-axis of the bar chart. The bar height (vertical position) encodes the number of copies sold. The user sees that the Action bar is significantly taller than the Puzzle bar, meaning more copies were sold in the Action genre. The user uses **bar height** (vertical position) as the primary visual cue. The user can also click on the bar and view the tooltips by hovering over the bars.

Task 2: Find the range of specific values and compare them

Question: Comparing the year regarding Europe sales, what is the difference between the year with the most number of games released and the year with the least?

- Visualization Used: Scatter Plot
- Mark: Bubble
- Channel: vertical positioning & color saturation

How the user completes the task:
First, the user needs to choose **Year** in the attribute selection button and **Europe** in the region selection one. The metric selection should be **Number of Games Released**. Each bubble represents the data point of the regional sales of video games of that specific year, corresponding to the **vertical positioning** and the x-axis values respectively. Each bubble is also encoded by **color saturation**, the more saturated the blue the more copies sold, and vice versa. The user needs to navigate through the bubbles first. Then, the user could locate the one with the most sales and the one with the least using not only the encoding channels but also tooltip information

by hovering on the bubbles. Eventually, the user will give the answer about the range between the most sold and least sold years regarding Europe sales.

Task 3: Order Games

Question: Order video game publishers from least to most total copies sold regarding the North America region.

- Visualization Used: Bar Chart
- Mark: Bar
- Channels: Vertical position (copies sold), horizontal position (after sort)

How the user completes the task:
Given the task, the user should choose **Publisher** in the attribute selection button, **North America** in the region selection one, and **Total Sales** in the metrics selection one. To order the genres from least to most, the user uses the **Sort** dropdown to rearrange the bars on the x-axis in ascending order. Each bar represents a genre, and the height (vertical position) of the bar indicates how many millions of copies were sold. The bars now appear in left-to-right order based on increasing height. The user reads the publisher names from left to right, using **horizontal position** and **bar height** to confirm the ordering.

Accessibility Features

To increase accessibility, we incorporated several key features into our visualization design. Each chart—both the scatter plot and the bar chart—includes a brief text summary placed next to it, providing users with a quick overview of the key insights without requiring them to visually analyze it. Our choice of color is also considerate of the color-blinded community. We used a scale of blue saturation, which is also color-blind friendly even at grayscale.

Additionally, we implemented tooltips that appear as we hover over the bars and bubbles, allowing users to access detailed information about each video game, such as the top 3 games sold each year, of each genre, by each publisher, and on each platform. These tooltips ensure that the data remains accessible even when it is difficult to read axes or when visual marks are challenging.

We also checked our visualizations using the WAVE accessibility tool to make sure they work well with screen readers and follow web accessibility rules. In addition, we used colors that are friendly for people who are colorblind and made sure that important information is not shown using color alone.

To help with navigating and understanding the visualization overall with readiness, we incorporated an optional tutorial using the intro.js package. Within the tutorial, the first-time user would understand the features of all the buttons, essential features such as the charts and alt-text summaries.

We also implemented the intro.js hint feature from the intro.js package. We attached a hint bubble at the chart summary region in which the user could be cued to click on it. The message from the hint reinforces the option to use the chart summary for clarifying and helping to understand what the chart delivers. We believe that this implementation could further help with the readiness to navigate through our data visualization.

However, the hint was eventually taken away in the final version due to the lack of interaction. We decided to remove the hint feature although it could potentially be practically redundant based on our user-testing feedback.

Piloting

We conducted the pilot test on May 1st, 2025. Ryan was the speaker, guiding the session, while I (Anushri) was the note-taker. Jack Schwanewede was our participant, and he engaged with our visualization hosted on my GitHub.io website. We facilitated the pilot session following the given script:

Ryan: "Hey Jack, we are evaluating our visualization and are asking you, the participant, to complete some tasks using the visualization and then provide feedback about the visualization and experience. As a reminder, we are evaluating the visualization, not you as a participant, so you don't need to worry about being "right" as you complete these tasks. There are three tasks, followed by a brief feedback session. The whole pilot session should take around 5 minutes. Do you consent to participate?"

Observation (Anushri): Jack says "Yes."

Ryan: "Thank you for agreeing to participate. We will start with the three tasks. Please 'think aloud' as you complete the task, meaning voice what you are thinking as you work through the task. Your first task is: which genre sold more video game copies regarding Global sales — Action or Puzzle?"

Observation (Anushri): Jack opens the visualization hosted on my github.io website on his laptop, and Ryan repeats the task. Jack says "I see the Region and the Attribute Dropdown", and he proceeds to select the Global sales option from the region dropdown and the Genre option from the attributes dropdown. Then, he selects the total sales option in the metric dropdown and asks Ryan to repeat the question. Jack approaches the task by clicking on the Action bar and the Puzzle bar. He looks at their tooltip and says "Action has total sales of 1722.84 million and Puzzle has total sales of 242.21 million." Based on the bar height (vertical position) and tooltip description, Jack answers "Action."

Ryan: "Sweet! That's the right answer! Your second task is when you compare the year regarding Europe sales, what is the difference between the year with the most number of games released and years with the least?"

Observation (Anushri): Jack begins by switching from the total sales to the number of games released in the metric dropdown, Global to European sales in the region dropdown, and Genre to Year option in the attribute dropdown. He says that "the most number of games released is in 2009, and least is in 1980." Then he looks at the tooltip for each of those bubbles in the scatterplot, and says that "the number of games released in 2009 is 1431 and is 9 in 1980." Ryan asks a clarifying question.

Ryan: "What was the difference between the highest and lowest games released? What specifically made you choose those points?"

Observation (Anushri): Jack says "I looked at the highest value in the y-axis (vertical position), which gave me the year 2009 and similarly the smallest value in the y-axis." Jack appears a bit confused and says "Wait a second… 2017 has 0 number of games

released, which might be a bit strange, so the least number of games released would either be in 2017 or 1980."

Ryan: "The data points in 2017 as well as 2020 show that there were 0 number of games released and that seems unreasonable to me. Do you have any suggestions on that?"

Observation (Anushri): Jack says "Maybe there wasn't any game released in 2020 or there might be an issue with the dataset." Ryan reiterates that the problem is that even though 2017 has zero games, if we click on the tooltip we still see the top 3 games in 2017, which is contradictory, because in this case, the count should have been 3. Jack further suggests "Go into the dataset, find 2017, and check if there are any years missing. A possible solution could be truncating it and cutting the dataset past 2015."

Ryan: "Thank you for the suggestion. We will look into our aggregation logic again. Your third task is to order video game publishers from least to most total copies sold regarding the North America region."

Observation (Anushri): Jack begins by switching the attribute dropdown to Publisher and the region dropdown to North America. He clicks the sort button and chooses the ascending option to get the copies from least to most sold. The visualization we get shows a uniform distribution because some of the publishers sold 0 copies in North America so all of them get included in our dataset. When Jack tries to move the tooltips to see information about the publisher, the whole page starts jiggling. Looking at this, Jack suggests "There might be an issue with the tooltip and how it's placed."

Ryan: "Thank you for pointing that out. We also might look into our dataset and maybe exclude or limit the publishers that sold 0 video game copies. That is the end of the third task. For this last bit, we welcome any feedback you may have about the visualization or about your process for completing the tasks."

Observation (Anushri): Jack gave us his feedback after the whole pilot test. He really liked our tooltip description especially because it shows the user the top 3 games sold by the different attributes. He also found the labels we used for the dropdown really useful, especially the Region label. Ryan proceeds to ask Jack specific questions about the visualization to get his suggestions.

Ryan: "If you go back to the scatterplot, do you think the color coding/color scale here is obvious to you?"

Observation (Anushri): Jack responds "The color scale is very helpful because the darker shades of blue clearly represent higher sales compared to the lighter colors. However, I didn't really use the color scale for the question that y'all asked for task 2 because I'm used to looking at points without any color in graphs so I just used the vertical position to complete task 2."

Ryan: "Have you read the summary that we added for each graph on the right side? Did you find that helpful in your task analysis?"

Observation (Anushri): Jack says "I feel like I knew enough about the visualization without

the summary but if I was confused about the charts, I could read the summary and learn more about the visualization." Jack suggested adding a title for the summary- either saying "Alt text" or something like "What is this chart showing."

Ryan: "Thank you so much for your feedback. We are happy that you understood the visualization easily. We were worried that you might not be able to understand the attributes, region, metric, and sort dropdown. Maybe it will be better to give the users some instructions or include some annotations beside each dropdown button, explaining what each of them is used for, something like a tutorial."

After conducting the pilot test and receiving valuable feedback from Jack, we decided to make the following changes:

- We rechecked our aggregation logic for the number of games released in our code. The reason why there were 0 games in 2017 is because we had included a filter function in our code which was removing the games where sales were 0. We decided to remove that filter function. Now our visualization correctly shows 3 games in 2017.
- We decided to include a tutorial beside each dropdown button, which makes the visualization more accessible to the users because they can use the tutorials to understand what the dropdown buttons mean. Especially for users who could only navigate through all the buttons through the keyboard instead of the mouse (using tab to go over all the buttons), the tutorial could help them to have a better overview of what function the buttons and different areas have. This could potentially increase the readiness to understand the visualization and use it more intentionally.
- We added a title for the summary section beside each chart in our visualization. The title is "What does this chart do?" This helps screen readers clearly understand the function of the brief summary before they read it.
- Regarding the issue of having zero data points when the attribute is Publisher in ascending order, we eventually decided to keep the data as it is. We wanted to be loyal to the dataset as much as possible. A similar decision comes with incorporating data from 2017 to 2020, even though they may not be accurate as reflected in the scatter plot. However, we still decided to include all the data points as they are part of the original dataset.

User Testing

User #1

- Major: Classics

Script for Task Completion

Ryan: "We are evaluating our visualization and are asking you, the participant, to complete some tasks using the visualization and then provide feedback about the visualization and experience. As a reminder, we are evaluating the visualization, not you as a participant, so you don't need to worry about being "right" as you complete these tasks. There are three tasks, followed by a brief feedback session. The whole pilot session should take around 5 minutes. Do you consent to participate?"

Observer (Anushri): User 1 said "Yes, absolutely. My pleasure."

Ryan: "Sweet! Thank you for agreeing to participate. We have a built-in tutorial for all the first-time users, do you want to go through that first?"

Observer (Anushri): User 1 replied "Yes" and went through the tutorial.

Ryan: "Okay since you are done with the tutorial, we will ask you to complete the first task. Please 'think aloud' as you complete the task, meaning voice what you are thinking as you work through the task. Your first task is: Which Genre sold more video game copies regarding Global Sales — Action or Puzzle?"

Observer (Anushri): User 1 asked "For the whole time period?" Ryan asked them to explain their question. Then, User 1 started interacting with the dropdown buttons, from the Metric to the Region buttons. Ryan repeats the question. User 1 was more clear with the button to interact with after finding that the Attribute one contains Genre. User 1 clicked on it and after comparing the bar heights answered the question with the right answer: Action outweighed Puzzle in sales.

Ryan: "Good job there! Now, the second task is: when you compare the year regarding Europe sales, what is the difference between the year with the most number of games released and the year with the least?"

Observer (Anushri): User 1 started off by changing the attribute from Genre to Year, then selecting Europe in the region section and Number of Games Released for the metric. After seeing a scatter plot, User 1 quickly identified the most released year as 2009 and went through all the lowest bubbles across the x-axis. Eventually, User 1 located the Year 2020 by comparing the specific numbers in the tooltips across different bubbles. User 1 answered, "2009 has roughly 1400 games while 2020 has 0 games released."

Ryan: "Awesome! That is correct! You did a great job here. Now the third task is: order video game publishers from least to most total copies sold regarding the North America region."

Observer (Anushri): User 1 went on selecting North America for region first, ascending order in the sort button, and then changing Year to Publisher in attributes. However, user 1 was confused after seeing all the bars in the chart displaying zero values.

Ryan: "Yeah so this is one of the issues with our data visualization but largely due to our dataset. There are so many publishers that it is impossible to put all of them on one page. However, there are publishers that only launch games in, say, Europe or Japan, such that when you select other regions they may just display zero. However, for the sake of loyalty to the original dataset, we decided to still keep it this way. Back to the question, if you want to see the maximum value in this condition, maybe changing your sorting direction could be a good idea."

Observation (Anushri): After being cued by Ryan, User 1 went on to change the sorting direction from ascending to descending. User 1 answered, "The maximum value is Nintendo, producing a revenue of 815 million copies overall in North America."

Ryan: "Great! That is all the questions we have for you. Do you have any feedback or questions for us? What do you think is good? What do you think we need to improve?"

Observation (Anushri): User 1 said it was a great experience overall. However, initially, they were a bit confused by the function of the attribute button – user 1 did not expect that there would be that many selection options hidden within the dropdown button. A good suggestion that user 1 provided was to show all the options at once and highlight the selected options for each button.

Ryan: "That is great advice! What else do you have for us?"

Observation (Anushri): User 1 also thought we could rearrange the layout so that users might not need to scroll up and down every time they were manipulating the selection.

Ryan: "That is one thing we also noticed when working on it. However, since we want to make sure the chart area is big enough so that people can visually see the information clearly, we intentionally made the font and the graphic elements larger so that it is more easily perceived by users. That part of accessibility is a compromise to execution."

Observation (Anushri): User 1 agreed with the point and shared understanding. User 1 then complimented on the immediate change between charts whenever the selection option changes.

Ryan: "That is part of the design choice by us. We did implement some animation for transition in the earlier version, but then we decided to take it away because it did not look good and it might be misleading. Hearing that you like the sharp and crisp transitions really consolidated our decision. So, I have a few questions for you, too. First, what do you think about the color overall? Does it help you complete the tasks?"

Observation (Anushri): User 1 said "The color looks fine; there is no misleading information at all." In terms of helping with the task, user 1 admitted that the color did not help much for them individually, but having a color cue was still good.

Ryan: "Okay. The next question is, did you notice the chart summary right next to the chart? If so, did it help with task completion? Did you notice the hint bubble hanging on the chart summary title?"

Observation (Anushri): User 1 said that they noticed the chart summary during the tutorial, but they did not rely on it when completing the tasks. Also, the hint was not noticed, either.

Ryan: "Cool! Sounds good. Do you have any more questions for us?"

Observation (Anushri): User 1 said "no." The user testing for user 1 was completed.

Summary of performance & feedback

User 1 struggled with understanding the functions of all the dropdown buttons initially but adjusted quickly after getting the hint from Ryan. Overall, the completion of all three tasks was smooth, despite minor struggles that were resolved quickly by hints from the interviewers. User 1 displayed a more average, typical person's tendency to use the data visualization, with a few struggles involved. User 1 also gave valuable feedback to the interviewers in iterating the visualization to later versions:

- Display all the dropdown selection options in one place so that users can view them thoroughly. Highlight the selected options when displaying different charts.
- Rearrange the layout so that users do not need to scroll up and down to select combinations of options.
- Maybe adding a potential explanation in the visualization for the zero-valued charts.

User #2

- Major: Psychology

Script for Task Completion

Anushri: "We are evaluating our visualization and are asking you, the participant, to complete some tasks using the visualization and then provide feedback about the visualization and experience. As a reminder, we are evaluating the visualization, not you as a participant, so you don't need to worry about being "right" as you complete these tasks. There are three tasks, followed by a brief feedback session. The whole user testing session should take around 5 minutes. Do you consent to participate?"

Observation (Ryan): User 2 says "yes."

Anushri: "Thank you for agreeing to participate. We will start with the three tasks. Please 'think aloud' as you complete the task, meaning voice what you are thinking as you work through the task. We have a built-in tutorial for all the first-time users, do you want to go through that first?"

Observation(Ryan): User 2 says "yes", and they go through the tutorial.

Anushri: "Your first task is: which Genre sold more video game copies regarding Global sales — Action or Puzzle?"

Observation(Ryan): User 2 says "I first choose Genre in the Attributes dropdown, and then I'm selecting the global sales option from the region dropdown." The total sales option is selected by default in the metric dropdown, and they acknowledge that. User 2 then asks Anushri to repeat the question. User 2 answers "Action."

Anushri: "What made you choose action?"

Observation(Ryan): User 2 says "I looked at the y-axis, and the bar height of the Action genre was higher than Puzzle, so I selected Action."

Anushri: "Great! That's the right answer. You can also hover over the bars and see if that helps."

Observation(Ryan): After Anushri provides User 2 with the tooltip hint, they proceed to check it out and User 2 says "That's very helpful especially if the bar heights of the different genres were almost the same."

Anushri: "Your second task is when you compare the year regarding Europe sales, what is the difference between the year with the most number of games released and years with the least?"

Observation(Ryan): User 2 starts by first changing the region dropdown to Europe. User 2 takes some time to figure out where the Year attribute is, and then asks Anushri where it is located. Anushri doesn't explicitly want to give out the answer so she gives a hint.

Anushri: "It is similar to how you approached your first task. Think about it carefully."

Observation(Ryan): Anushri repeats the question, and after about a few seconds, User 2 finds and selects Year in the attribute dropdown. Then, they select the number of games released in the metric dropdown. User 2 clicks on the highest bubble and reads out the tooltip "2009 has 1431 number of games released so it's the highest." They proceed to click on the lowest bubbles, and since a lot of them have similar y-axis values, user 2 clicks on each of the bubbles and reads out the tooltip. After comparing each bubble, User 2 answers "2020 has only 1 game released so it's the lowest."

Anushri: "Great! That's right. I wanted to ask a follow-up question for this task: Did the color scale near the scatterplot help in the completion of this task?"

Observation(Ryan): User 2 answers "Yes, it helped because it clearly states that the higher the concentration of blue, the more the number of games released, and dark blue bubbles are on the top part of the scatterplot, whereas the lower concentrations or the white bubbles are at the bottom."

Anushri: "Perfect! Your third task is to order video game publishers from least to most total copies sold regarding the North America region."

Observation(Ryan): User 2 proceeds by changing the attribute dropdown to Publisher and the region dropdown to North America. Then, they select the total sales option in the metric dropdown and the ascending option in the sort dropdown. As mentioned before, the visualization we get has the same bar height which is 0. User 2 answers "0." However, user 1 was confused after seeing all the bars in the chart displaying zero values.

Anushri: Great job! For this third task, the reason why all these bars show 0 is because some of the publishers sold 0 copies in North America so all of them get included in our dataset, and since we had to choose from least to most, we see all publishers that sold 0 copies. Do you think we should add the zero data points or exclude it from the dataset?"

Observation(Ryan): User 2 answers "You should include the zero data points because that way you're representing your dataset accurately."

Anushri: "This is the end of the third task. For this last bit, we welcome any feedback you may have about the visualization or about your process for completing the tasks."

Observation(Ryan): User 2 really liked the tutorial at the beginning of the visualization. They said, "The tutorial helped me understand the different dropdown buttons and made completing the tasks easier."

Anushri: "Since you're not from this class, did you find it easy to finish all the tasks? Is there anything you found confusing or that you wanted to change?"

Observation(Ryan): User 2 answers "It was easier for me also because I have experience with visualizations in my psychology class, so it didn't take me too long to complete the tasks. Other than that, the different dropdowns, chart summary, and the tooltips were pretty self-explanatory."

Anushri: Thank you so much for your feedback. We are very happy that you understood our visualization and completed the tasks so effortlessly.

Summary of the participants' performance and their feedback.

User 2 performed all the tasks very effortlessly despite not being part of our class. They clearly described the steps needed to complete each task like changing the different dropdown options, comparing the y-axis values (vertical position channel), understanding the color saturation channel in the scatterplot, and analyzing the different tooltips. They also gave us valuable feedback that made us reflect on the changes we made after our pilot test with Jack. We noticed key improvements:

- Adding the tutorial before analyzing the visualization was extremely beneficial. User 2 found it easier to understand the dropdowns and complete all the tasks smoothly.
- The title added for the summary section was also useful because it helped user 2 clearly understand what information the different charts are supposed to convey.
- Regarding the issue of having zero data points when the attribute is Publisher in ascending order, we eventually decided to keep the data as it is based on User 2's feedback. We wanted to be loyal to the dataset as much as possible. A similar decision comes with incorporating data from 2017 to 2020, even though they may not be accurate as reflected in the scatter plot. However, we still decided to include all the data points as they are part of the original dataset.

User #3

- Major: Computer Science

Script for Task Completion

Ryan: "We are evaluating our visualization and are asking you, the participant, to complete some tasks using the visualization and then provide feedback about the visualization and experience. As a reminder, we are evaluating the visualization, not you as a participant, so you don't need to worry about being "right" as you complete these tasks. There are three tasks, followed by a brief feedback session. The whole pilot session should take around 5 minutes. Do you consent to participate?"

Observer (Anushri): User 3 said yes.

Ryan: "Sweet! Thank you for agreeing to participate. We have a built-in tutorial for all the first-time users, do you want to go through that first?"

Observer (Anushri): User 3 replied "Yes" and went through the tutorial.

Ryan: "Okay since you are done with the tutorial, we will ask you to complete the first task. Please 'think aloud' as you complete the task, meaning voice what you are thinking as you work through the task. Your first task is: Which Genre sold more video game copies regarding Global Sales— Action or Puzzle?"

Observer (Anushri): User 3 then selected Genre as the attribute and looking at the bar height, they recognized that Action clearly has higher sales than Puzzle.

Ryan: "Great. That is correct. Now, the second task is: when you compare the year regarding Europe sales, what is the difference between the year with the most number of games released and the year with the least?"

Observer (Anushri): User 2 changed the attribute to Year, region to Europe, and metric to the number of games released. Then, user 2 navigated to look for the highest bubble which is 2009. User 2 compared a series of the lowest-placed bubbles and by looking at the hovering tooltips, they also identified the year that released the least number of games was 2020.

Ryan: "Good job on that last one. Now the third task is: order video game publishers from least to most total copies sold regarding the North America region."

Observer (Anushri): User 3 went on selecting North America for region first, ascending order in the sort button, and then changing Year to Publisher in attributes. However, user 3 was confused after seeing all the bars in the chart displaying zero values.

Ryan: "Yeah so this is one of the issues with our data visualization but largely due to our dataset. There are so many publishers that it is impossible to put all of them on one page. However, there are publishers that only launch games in, say, Europe or Japan, such that when you select other regions they may just display zero. However, for the sake of loyalty to the original dataset, we decided to still keep it this way. Back to the question, if

you want to see the maximum value in this condition, maybe changing your sorting direction could be a good idea."

Observation (Anushri): After being cued by Ryan, user 3 went on to change the sorting direction from ascending to descending. User 3 answers "The maximum value is Nintendo, producing a revenue of 815 million copies overall in North America."

Ryan: "Great! That is all the questions we have for you. Do you have any feedback or questions for us? What do you think is good? What do you think we need to improve?"

Observer (Anushri): User 3 asked if it was possible to go through the tutorial again. Although the answer was no, user 3 suggested that we could add a button that allows users to rewatch the tutorial just in case they might need it.

Ryan: "That is great advice, thank you! What else do you have for us?"

Observer (Anushri): User 3 said maybe adding a little explanation whenever the zero-value scenario was met would be helpful for getting rid of confusion.

Ryan: "That is also a good idea. I think we can add an annotation on the chart to mark it. I have some questions for you now. First, how is the color overall to you?"

Observer (Anushri): User 3 said the color is fine. It is easily distinguishable for them.

Ryan: "Great! My second question is, did you use the chart summary to assist with task completion? Did you notice the hint on the chart summary?"

Observer (Anushri): User 3 said that the chart summary was not used to help them complete the task but keeping it is helpful. They also did not notice the hint for the chart summary.

Ryan: "Thank you so much for your participation! If there is no further question, then this is the end of the user testing session."

<u>Summary of performance & feedback</u>

User 3 performed exceptionally well in completing all three tasks. They reacted quickly to all the features of our visualization and gave meaningful feedback on our improvement of the final version. Despite no struggle with understanding the functionality of the visualization, user 3 did great in navigating the data points and locating the target ones. Here is a list of feedback that could be used for our improvement:

- Make sure that the tutorial can be rewatched again. User 3 suggested adding a button that could refresh the page so that the tutorial would pop up again without resetting the visualization to the default setup.
- Add an annotation for an explanation of the zero-valued scenario. This could largely alleviate users' confusion when operating the visualization.

Final Visualization

Changes we made following the user testing in the final version:

1. We added a new button for rewatching the tutorial. After clicking the button, the webpage will refresh itself, therefore generating the tutorial again for the user who needs to go through it again.
2. We got rid of the intro.js hint because it did not get noticed well. Since we already added the tutorial button mentioned above, we realized that the hint button was redundant.
3. We added an annotation for zero-value charts (Publisher + Total Sales + Ascending + any non-Global region) to explain why the data makes the chart look like that. We hope it will help users understand our intention of presenting the data in such a way.

Link: https://anushrimodi.github.io/

Images:

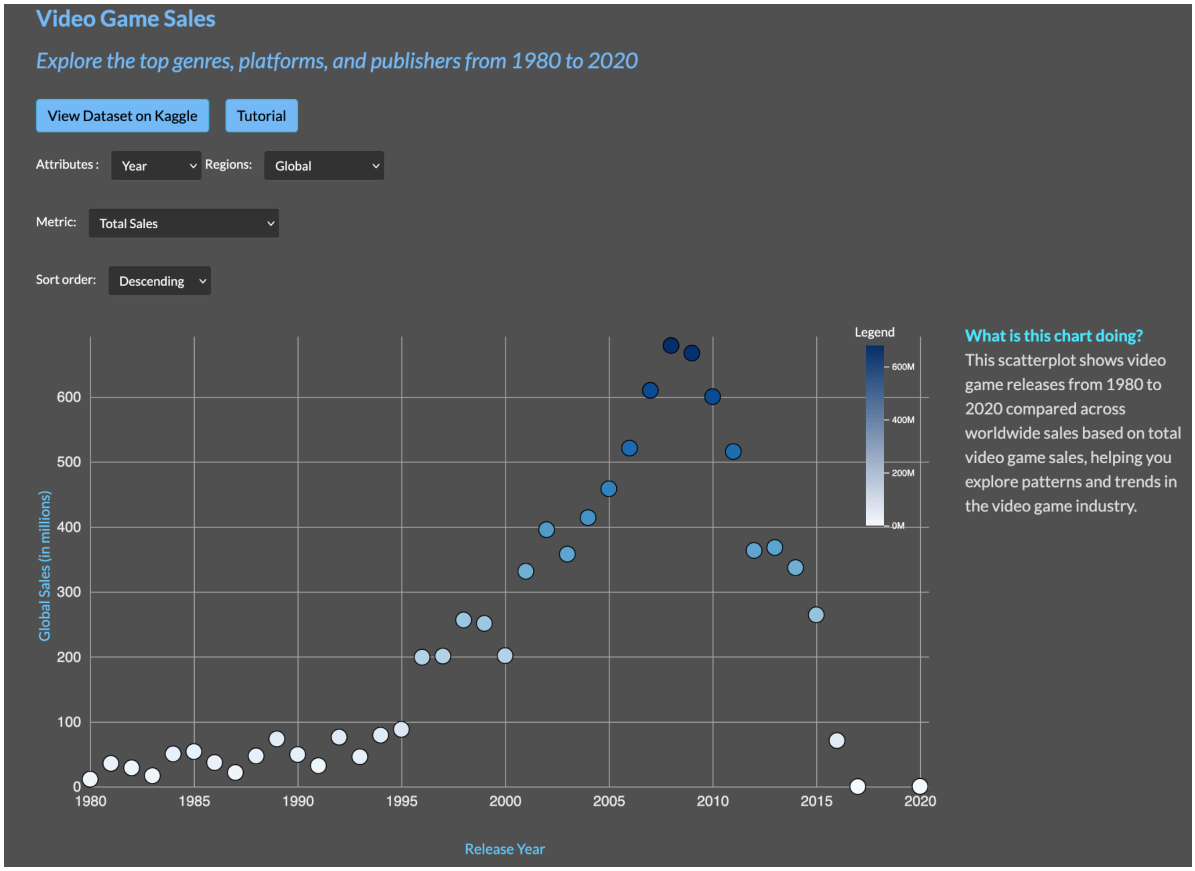


Figure 1. Default scatter plot when entering the visualization for the first time after the tutorial. The attribute is set to be Year by default.

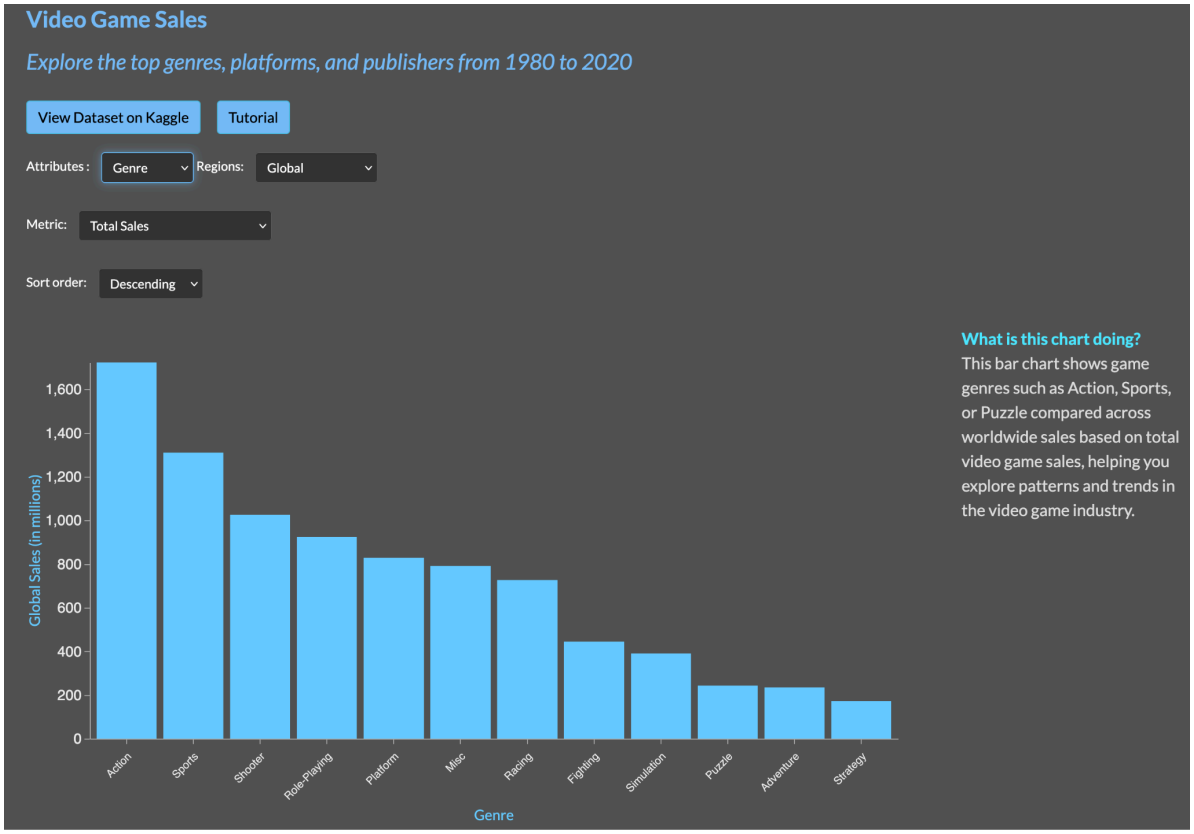


Figure 2. The chart changes to a bar chart when the attribute is not Year. Here the attribute is Genre.

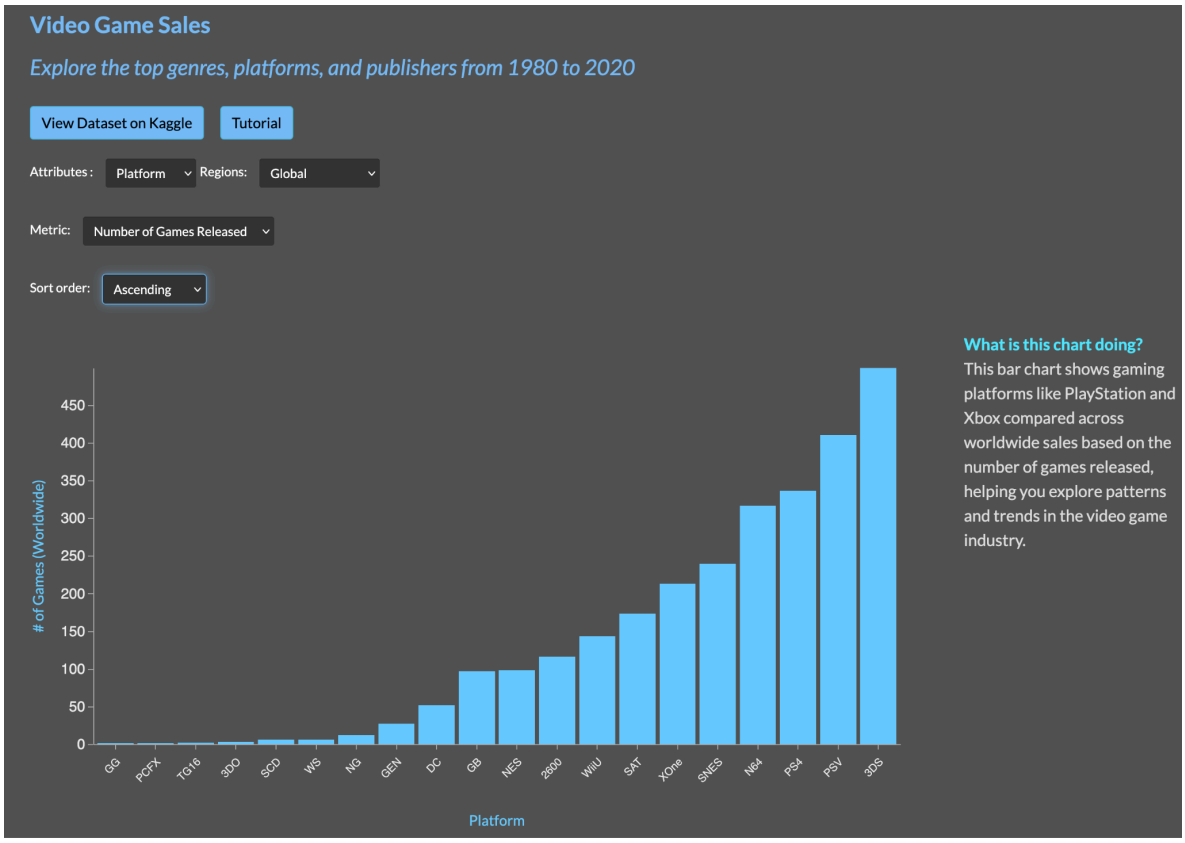


Figure 3. Metric and sort order could also be selected to display other options. Here, the metric is Number of Games Released in ascending order.

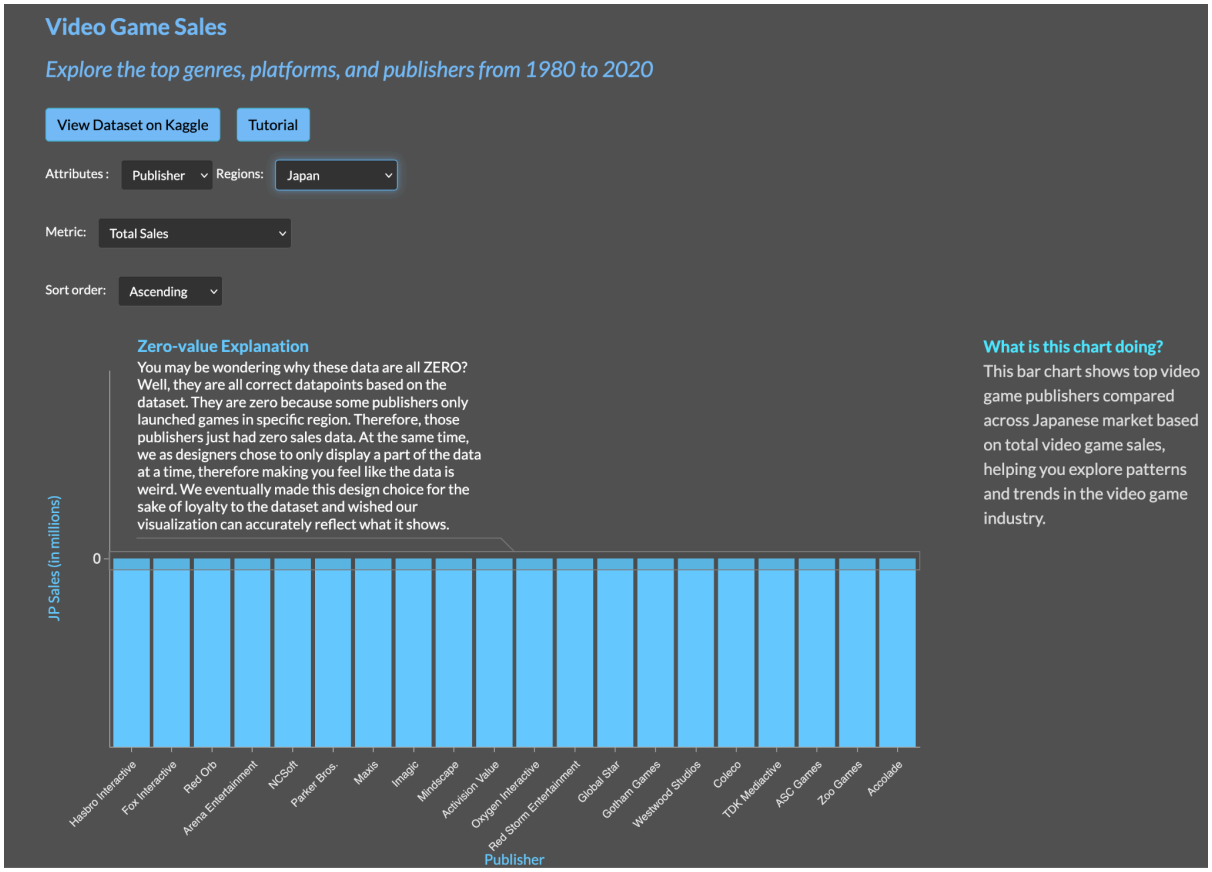


Figure 4. Regions could also be changed. Here, the region is Japan. Notice that it is one of the zero-value scenarios, and the annotation is for explanation.

Personal Reflection

Link to the final visualization and write-up: https://anushrimodi.github.io/

Ryan and I went through a lot of datasets initially before choosing the video game sales one. For example, we looked at an air quality index dataset and a COVID-19 dataset, but they both lacked a lot of data points and required too much cleaning. So we decided to go with the video game dataset instead. It's also a topic both of us enjoy because we like playing video games, and luckily, this dataset didn't need much cleaning. I have experience with data preprocessing from my previous internship, and I really enjoy analyzing and cleaning large datasets, so I took the lead on that part. I first counted the number of missing values in Excel and deleted them since there were very few and they didn't affect our dataset much. I also created several pivot tables that helped us figure out what we wanted our visualization prototype to look like. It made it easier to see which attributes like genre or platform we wanted to plot against which sales region.

The next step was sketching our prototypes. Our initial idea was to use a bar chart, similar to what we did in Lab 5 with the coffee dataset that had different x-axis dropdowns. We wanted to include dropdowns not just for the x-axis but also for the y-axis to let users choose different attributes and sales metrics. But since our dataset was pretty large, Ryan suggested we also use a scatterplot as our default chart to show the number of copies sold per year. This helped avoid having too many bars of the same height for the Year attribute, which would've made the chart harder to read. So our sketches ended up being very close to what our final visualization looks like. We also added accessibility features to our sketches like a short summary beside each chart.

Interviewing and conducting user testing on the different participants was probably the most fun part of the process because it helped us see whether our visualization actually worked. All the participants we interviewed had different opinions, but all of them were able to complete the tasks we gave them, which felt great. Even participants who were not from our class had such valuable feedback for us, and that made this process even more fun. Using the different perspectives, we were able to create a final visualization that is extremely accessible to readers and at the same time helps them learn something about video game sales.

I also learned so much through this project. Ryan and I really wanted to include everything we've learned in class, from tooltips and transitions to dynamic label positioning and axes. We even added new features like a tutorial for accessibility and user guidance, a button to explain why we included publishers that sold zero copies, an interactive chart summary, and a very detailed tooltip. Plus, I learned a lot about GitHub too because we created countless branches and fixed countless conflicts, which was not the most fun part.

Lastly, my collaboration with Ryan was very smooth. He was super approachable, and we divided the tasks equally, which made it easy to work with him. We both learned a lot from each other too and definitely made a great team. Overall, I really enjoyed working on this final project.

Works Cited

GeeksforGeeks. “How to Create Dictionary and Add Key–Value Pairs Dynamically?” *GeeksforGeeks*, 28 May 2020, www.geeksforgeeks.org/how-to-create-dictionary-and-add-key-value-pairs-dynamically/. Accessed 5 May 2025.

Holtz, Yan. “Building Tooltips with D3.Js.” *D3-Graph-Gallery.com*, d3-graph-gallery.com/graph/interactivity_tooltip.html.

“Intro.js Documentation – Intro.js.” *Introjs.com*, 2024, introjs.com/docs. Accessed 5 May 2025.

Lu, Susie. *Susielu.com*, 2024, d3-annotation.susielu.com/.

Navetat, Vincent. “Building a Simple Tooltip Component That Never Goes off Screen.” *Medium*, Carwow Product, Design & Engineering, 14 May 2020, medium.com/carwow-product-engineering/building-a-simple-tooltip-component-that-never-goes-off-screen-c7039dcab5f9. Accessed 5 May 2025.

“Workbook.” *The Chartability Workbook*, chartability.github.io/POUR-CAF/.