1. Basic Info

Project Title: Domain Expansion: Expanding Anime Horizons With Data!
Video: https://youtu.be/iepzYT3_NVc
Group Members:
Wonhak Choi 59765776
Danny Hu 94336237
Isabelle Tam 91309039
Jason Wei 26637397

2. Overview

Our visualization aims to provide insights into the anime industry by offering a comprehensive exploration of various facets. The visualization caters to diverse user needs, including comparing genre popularity, identifying critically acclaimed anime, studying age rating and score correlations, analyzing production fluctuations for historical insights, and examining the relationship between anime scores and completion rates. Tailored for industry experts and casual viewers, our visualization serves as a versatile tool for understanding market trends, guiding production decisions, and enhancing viewer recommendations.

3. Data and Data Preprocessing

Dataset: [Anime Recommendation Database 2020 (kaggle.com)]

This dataset contains around 18,000 anime reviewed by around 325,000 different users. We will be taking the top 1000 most “popular” anime sorted by their Score, which is the average of the users’ scores for each anime. The choice to only visualize the top 1000 anime was driven by the fact that we wanted to make use of a force-directed packed bubble chart in our visualization and too many elements present performance issues for force-directed graphs. The dataset has a total of 35 columns but we are only interested in 10 of the 35 columns. The columns we are interested in are as follows: [MAL_ID, Name, Score, Genres, Studios, Source, Rating, Completed, Dropped, Premiered].

3.1 Data preprocessing

We chose to use Python along with Pandas and Numpy to process the csv file containing all the animes (anime.csv). Below are the steps detailing the preprocessing pipeline:

1. Read anime.csv and convert to pandas dataframe
2. For every anime and for each of the 10 columns we are interested in, if any column is empty, not a number or is unknown we remove that anime
3. Remove all other columns that are not the 10 columns that we are interested in
4. We noted that the “Genres” column for each anime includes a list of the genres the anime belongs to sorted in order of descending significance. However, we want an anime to only belong to one distinct genre, therefore we decided to take the first Genre in the Genres list and assign the anime to that genre. For example, if an anime had genres: “Action, Comedy, Mystery” we changed this column to just include “Action”
5. Sort the animes in decreasing values of score
6. Get the first 1000 animes in the sorted list
7. Convert any numeric fields that are represented as strings to numeric values.
8. Convert the final dataframe back to a csv which will be used in for our visualizations

The final preprocessed csv contains 1000 rows and 10 columns. In addition to preprocessing done in Python, we also preprocess data in our D3 library such as grouping all animes into their genres or grouping anime by the year they were released.

3.2 Data descriptions

This is a table describing the attributes in our final csv that we will visualize.

Field Name	**Attribute Type**	**Cardinality /Range**	**Attribute Description**
MAL_ID	Categorical	1000 levels	This is a unique identifier used to identify the anime
Name	Categorical	1000 levels	Full name of Anime
Score	Ordinal	7.36 to 9.19	Average score of the anime given from all users in MyAnimelist database.
Genre	Categorical	10 levels	A string identifying the genre that the anime belongs to.
Studio	Categorical	170	Comma separated list of studios (e.g. Sunrise)
Source	Categorical	13	Source of anime eg. Manga, Light novel, Book, etc. (e.g Original)
Rating	Categorical	5	Age rating of the anime (e.g. R - 17+ (violence & profanity))
Completed	Quantitative	3 to 2182587	Number of users who have complete the anime. (e.g. 718161)
Dropped	Quantitative	122 to 174710	Number of users who have dropped the anime. (e.g. 26678)
Premiered	Ordinal	150 levels	The Season and Year that the anime was released (e.g. Summer 2021)

4. Tasks

#	Domain	Abstract
1	Compare the popularity of all the genres in the top 1000 anime using the metric of the number of anime in each genre. **Use case:** An anime industry analyst wants to understand market trends by assessing which genres are more prevalent in popular anime. They can use this information to advise production studios on which genres to explore for new projects.	{compare distribution}
2	A user may want to see the most critically acclaimed anime within a specific genre. **Use case:** A viewer is interested in finding the highest-rated anime within their favourite genre to ensure they watch quality content. They plan to use the visualization to pick their next anime to watch.	{Present extremes}
3	Discover if there is any correlation between the score and the age rating of animes in a specific genre. **Use case:** An analyst at an animation studio is studying whether more mature in comparison to kid's content tends to be better received, and wants to know if there's a relation between the age ratings of anime and their critical scores. They want to understand if higher-rated content tends to be more critically acclaimed to help decide which age-rating anime to produce next.	{discover correlation}
4	Determine the annual fluctuations in the production of anime across different genres to identify a specific year for more in-depth investigation. **Use case:** An academic is conducting a historical study on the anime industry and needs to identify years with significant changes in anime production to focus their research and understand the industry's evolution.	{explore distribution}
5	Discover what the correlation is between the score of the anime and its completed-to-dropped ratio. **Use case:** A streaming service algorithm designer wants to predict which anime will be completed by viewers based on its score. They are looking for patterns that indicate high viewer retention to improve recommendations.	{discover correlation}

5. Visualizations

5.1 View 1: Packed Bubble Chart [Unique View]

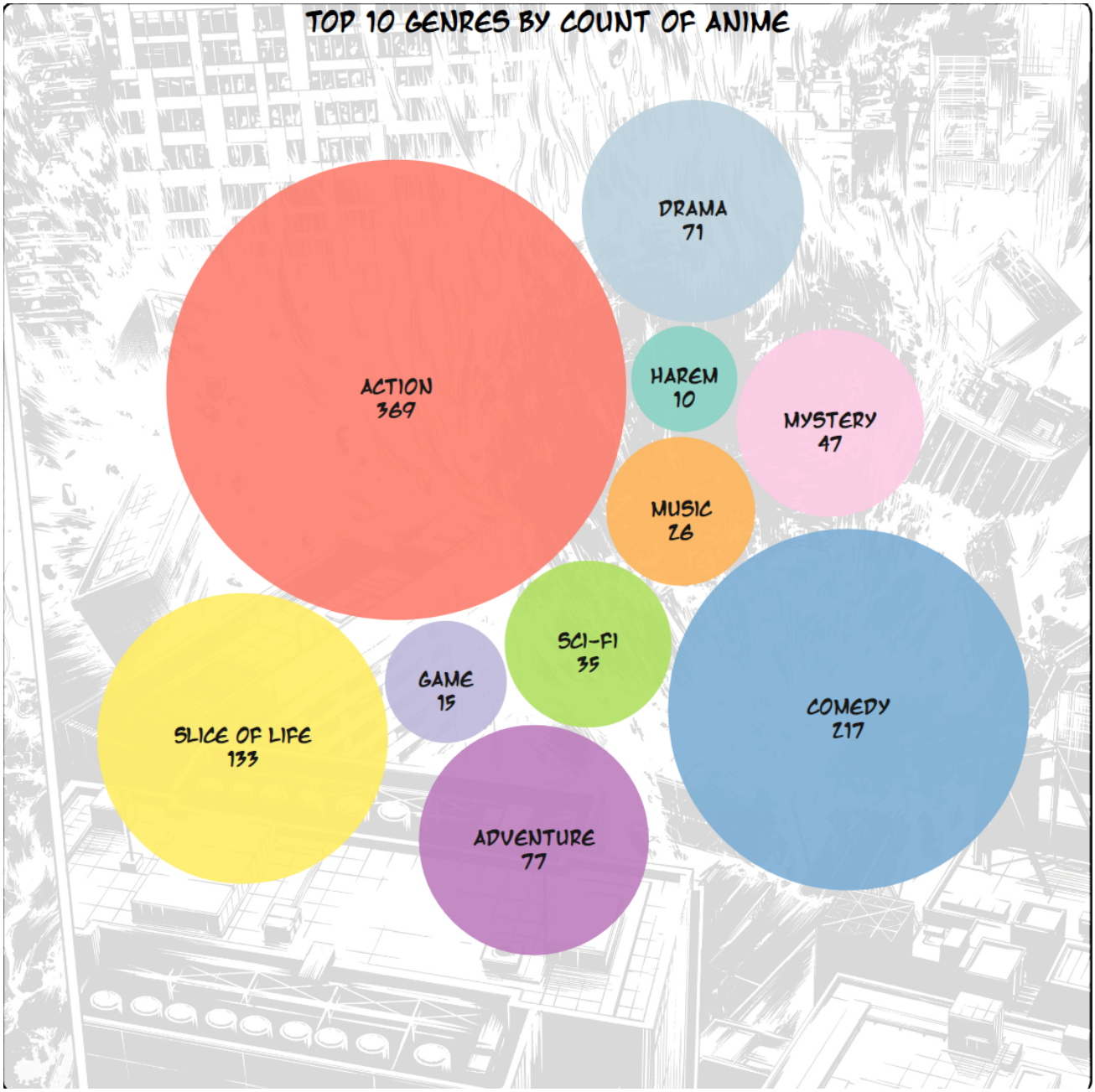


Figure 1. Top Level View

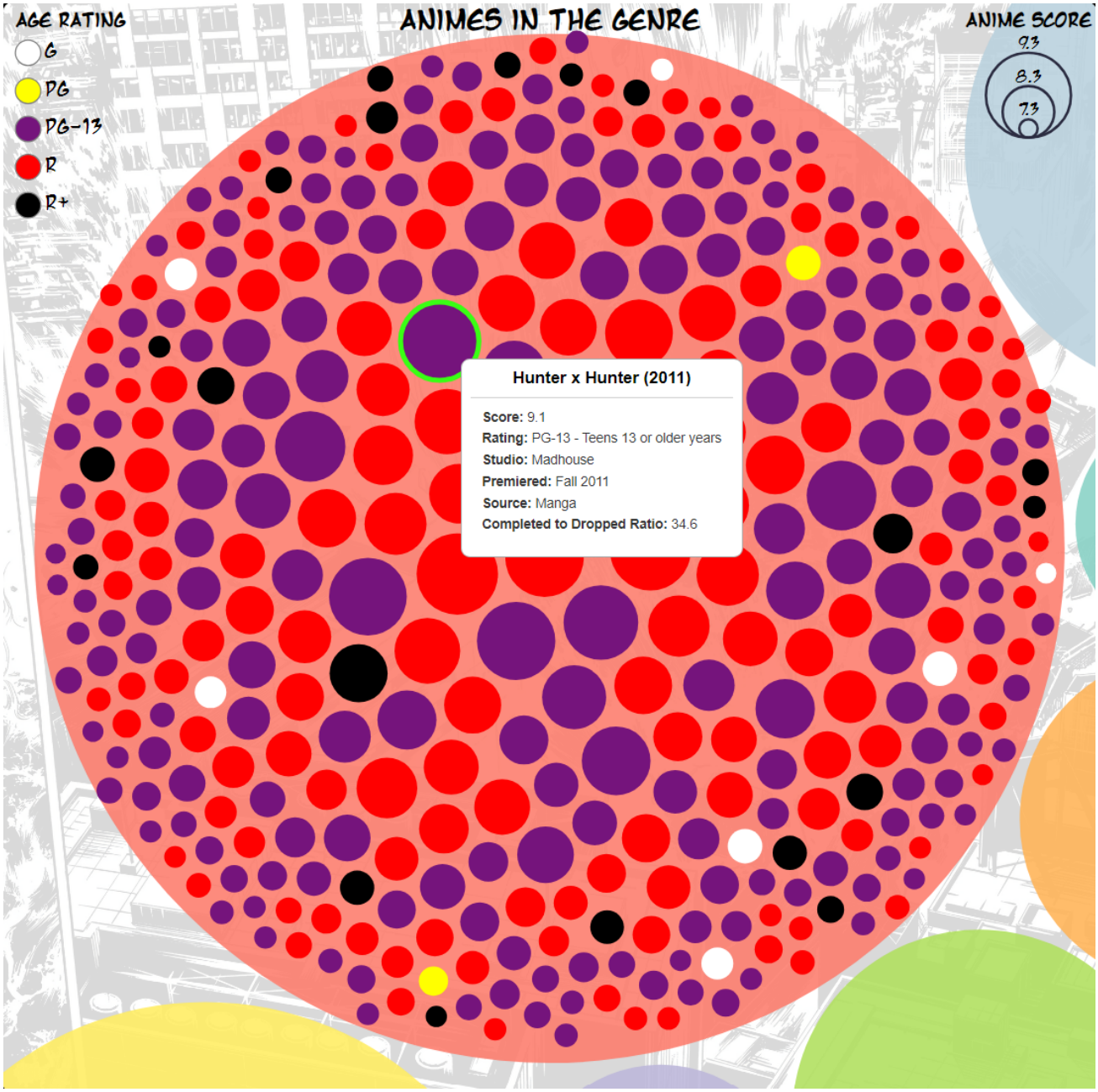


Figure 2. Drill Down View

5.1.1 Description of view:

View 1 is a hierarchical force-directed bubble chart. The top-level view in Figure 1 above presents the 10 unique genres in the top 1000 anime. A categorical color scale is used so that each genre has its unique color. By clicking on a specific bubble there is a transition that zooms in on the bubble and takes the viewer to the drill-down view of the genre displaying all the anime in that genre, the drill-down view uses force direction such that the larger bubbles are pulled towards the center of the graph whereas smaller bubbles are pushed outwards. The drill-down view also has a legend to describe what the colors represent and provides reference bubbles for the score of the anime. Furthermore, In the drill-down view when the user mouseovers a specific anime, a tooltip appears providing more specific information of the anime. There are also animated transitions from the

top-level view to the drill-down view, the drill-down view to another drill-down view and from the drill-down view back to the top-level view.

5.1.2 Rationale For View:

5.1.2.1 Top-level view rationale:

The top level (figure 1) addresses **task 1** which is comparing distributions. The force-directed packed bubble chart serves as an intuitive method to present the relative number of animes in the genre by utilizing interlocking 2D areas and area encoding. This encoding combined with the spatial grouping of the packed bubbles allows for an immediate visual comparison of genre popularity as it facilitates a more natural comparison of the relative bubble sizes. Another idiom that may be used to accomplish this task is using a bar chart, however, the bar chart is not as effective as the packed bubble chart for this task for various reasons. Firstly, the packed bubble chart is more visually impactful than a bar chart, the human eye is naturally drawn toward larger objects making it easier to identify which categories stand out and which are extremely small without the need for exact figures. Secondly, the packed bubble chart can reduce cognitive load by eliminating the need to compare lengths against a scale which is necessary in bar charts. Finally, a packed bubble chart in combination with force-direction providing animation is more aesthetic and attention-grabbing than a more traditional presentation with a bar chart. This can hold the viewer's attention for longer which will invite them to explore the data for longer rather than just reading off the values.

5.1.2.2 Drill-down view rationale:

The drill-down view in Figure 2 accomplishes **task 2** which is the presentation of extremes. We realize that anime scores are not quantitative and are ordinal instead which means that higher scores are far better than scores below them rather than just being a linear amount better. For example, an anime scored at 9 will be far better than an anime scored at 8 due to rating bias as users are more likely to reserve high ratings for very few select animes. Hence, the choice of using 2D interlocking areas and area encoding to encode the score better captures and presents the non-linearity of the score. Therefore, an anime that scored 9 has an area more than twice as large as an anime that scored 7. This choice supports the presentation of extremes because higher scores pop out due to being far larger than the rest. Furthermore, by using force direction to draw larger bubbles to the center of the screen we create a central focus on the largest bubbles creating a focal point for the viewers as human attention naturally gravitates towards the center of the screen. A packed bubble chart is effective also because of its efficient space utilization. A one-dimensional scatter plot may have overlapping points and hence may not be effective in presenting over 300 data points. Similarly, a bar chart with the x-axis being each anime would have scale issues as well with so many data points. Lastly, a packed bubble chart with force direction once again provides more aesthetic appeal over more common idioms which would again encourage users to explore and interact with the data.

The drill-down view also supports **task 3** which is the discovery of correlation. A categorical color scale is used to encode the age rating for the anime, this is effective as it is straightforward for people to recognize and process color and a quick scan across the chart can reveal patterns in color distribution. The size encoding for the score and having larger bubbles being drawn towards the center with smaller bubbles pushed towards the edges creates a radial gradient of scores with higher scores in the center. Hence, users can make a quick visual assessment by assessing whether there

are radial groupings of color (rings of color) to determine if there is some relation between the age rating and score. In the case of this task, other idioms could also be very effective such as a box plot for each category which provides the average score of animes in each genre. However, this would mean that we do not end up leveraging the ability of packed bubble charts to be extremely information-dense and the introduction of another chart meaning the user has to reference more visualizations leading to a more cluttered experience. Once again, a packed bubble chart does provide more visual engagement compared to a box plot of a bar chart.

5.1.3 Analysis of Marks and Channels:

5.1.3.1 Top level view marks and channels:

- Marks
    - **2D interlocking areas** (circle) represents a **genre**
- Channels
    - **Color**: differing hues encodes for the **specific genre of anime** this is an effective identity channel as it is scalable to our number of genres which is 10. Making it extremely quick to identify the genre of the bubbles
    - **Area**: area encodes for the **number of anime**. A square root scale is used to encode the sizes and this is effective because of the small number of bubbles (10) so it is very easy to see differences in the bubble.

5.1.3.2 Drill down view marks and channel:

- Marks
    - **2D interlocking areas** (circle) represents an **anime**
- Channels
    - **Color**: differing hues encodes for the **age rating** this is an effective identity channel as it is scalable to our number of genres which is just 5. Meaning that the visualization will not look too cluttered or busy with too many colours.
    - **Area**: area encodes for the **score**. The score was non-linear so rather than having linear increases in the area we wanted more of a quadratic increase and therefore area provides the most intuitive and visual way of capturing the non-linearity of increasing scores.
    - **Radial position**: This also serves as an encoding for the **score** given that our force direction pulls animes with larger scores closer to the center of the graph and animes with smaller scores will be pushed toward the edges. Thus a radial position that is closer to the center of the circle suggests a larger score.

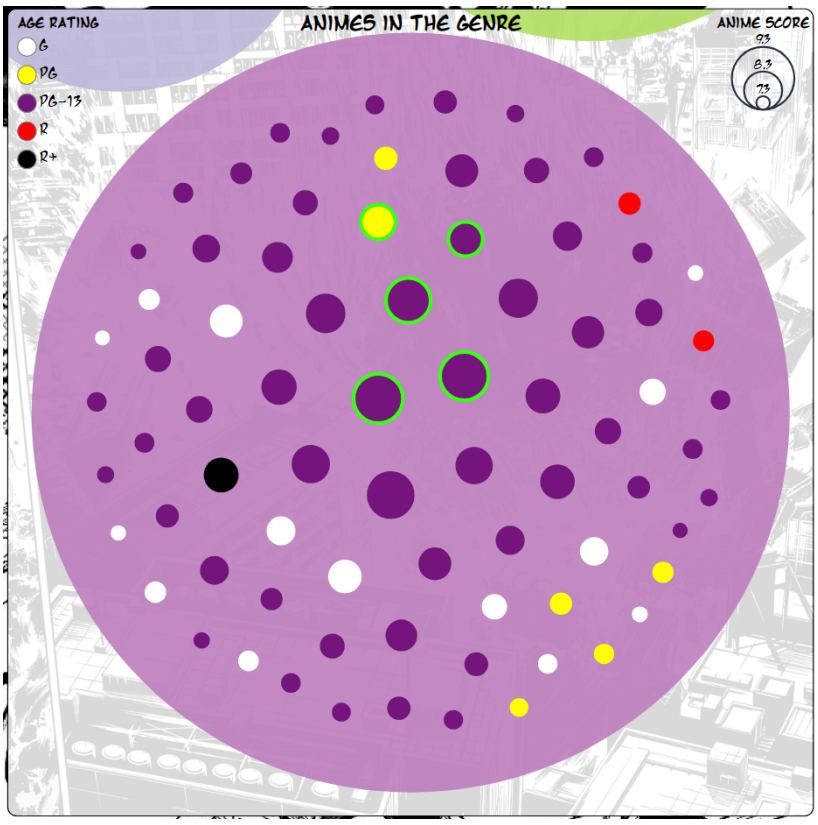
5.1.4 Interaction and Linkage:

- Selecting a specific genre in top-level view:

    - Clicking on a genre in the top-level will have a zoom animation in the chosen genre, displaying all the anime that belong to the selected genre.
    - Rationale: This zoom transition provides a sense of continuity by conveying the changing context of the visualization and this lines up with the concept of drilling down into the genre. This transition also engages users by drawing their attention with movement.

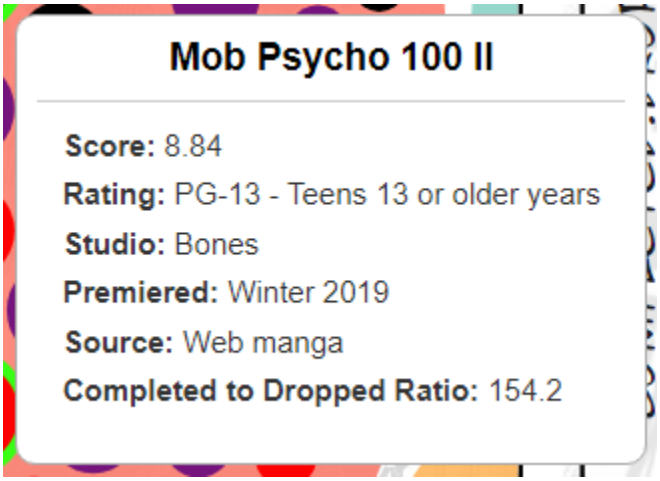
- Clicking on a blank area of the packed bubble chart resets zoom:
    - Clicking on a blank area of the char initiates a reverse zoom transition, guiding the user back to the top-level overview
    - Rationale: This interaction is designed to create an intuitive and seamless experience, symbolizing a step back from specific details to a broader perspective. The transition aids in reinforcing the user's understanding of the navigation structure, enabling them to easily toggle between detailed and general views of the data.

- Select multiple anime in the drill-down view



    - The user can click on multiple anime, and they will be highlighted by the neon green color on the drill-down view.
    - Rationale: The visual highlight will help the user focus on the selected anime. This lets the user select the anime they are interested in and allows them to compare their attributes without having to memorize which ones they find interesting.

- Tooltips



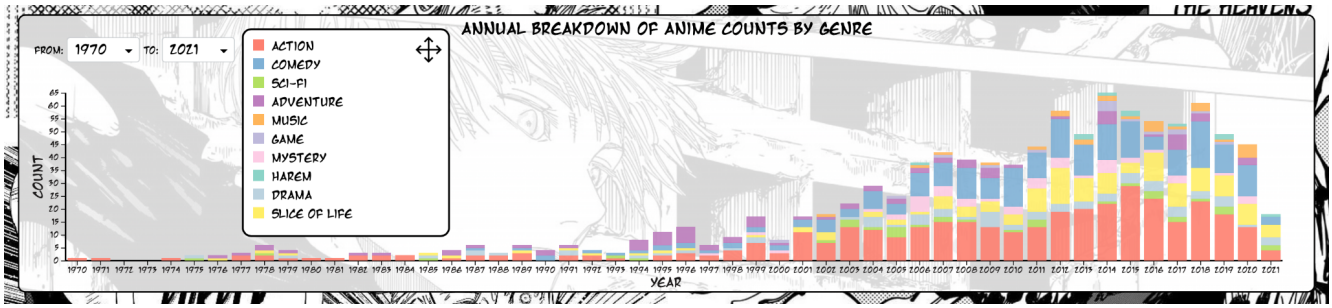
    - Hovering over an anime shows more anime information such as studio, premiere date, source, and completed-to-drop ratio
    - Rationale: This includes more detailed data for the academic to explore more attributes associated with the anime.

- Linkage with other views (described in detail in section 5.4)
    - Clicking on a bubble in the top-level view will filter the scatter chart.
    - Clicking on a genre will filter the bar chart to show only anime in that genre.
    - In the drill-down view selecting an anime on the main view will also select the same one on the scatter view and vice-versa. A user may select as many anime to compare as they want.
    - Clicking on blank space in the drill-down view globally resets the genre filter.

5.2 View 2: Stacked Bar Chart

5.2.1 Description of view:

View 2 is a stacked bar chart and it serves as a dynamic and insightful representation of annual anime production across different genres. Each bar encapsulates a specific year, segmented into distinct sections, each corresponding to a particular anime genre. The x-axis displays time, and the y-axis represents the total number of anime released during that time. The use of color differentiation within each bar allows for quick identification of genres, while the length of each segment directly reflects the quantity of anime produced for that genre in the given year. Users can interact with the chart through a dropdown menu to select specific years of interest, providing a focused view of production trends. This visualization is particularly valuable for industry analysts seeking to understand genre-specific production changes over time, enabling them to identify significant years for more in-depth investigation and contributing to a nuanced comprehension of the anime industry's historical evolution.



5.2.2 Rationale For View:

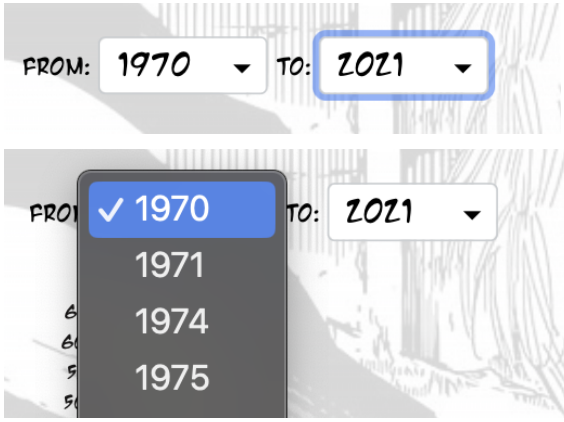
The choice of a stacked bar chart aligns with the principles covered in this course, particularly when dealing with part-to-whole relationships over time. The visualization encapsulates the total number of anime produced in a given year, providing a comprehensive view of the industry's annual output, supporting **task 4**. The segmentation of each bar into genres allows for a nuanced exploration of how different genres contribute to the overall production landscape. This design choice effectively captures the multifaceted nature of anime production, enabling users to discern patterns and trends across genres within specific years.

The incorporation of a dropdown menu for year selection enhances the interactivity of the visualization, aligning with principles of user engagement and exploration. This feature allows users to focus on specific years of interest, providing a more targeted view of production trends. This interactive element is particularly valuable for tasks such as historical studies, enabling academics to pinpoint significant years for in-depth investigation. The dropdown menu balances the need for a broad overview of the data with the ability to zoom in on specific time periods, ensuring flexibility in exploration.

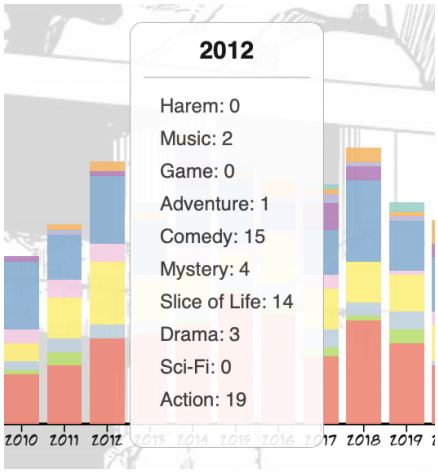
In considering alternatives, such as a grouped bar chart, the stacked design was preferred due to its ability to showcase both the total number of anime and the genre-specific distribution within a given year. A grouped bar chart, while suitable for part-to-whole relationships, could potentially introduce clutter and readability issues, especially when dealing with a large number of genres. The stacked bar chart, with its seamless integration of genres within each bar, proved more effective in presenting a comprehensive yet clear overview of the data.

5.2.3 Interaction and View Linkage:

- Year Selection Dropdown:
    - Users have the option to choose a "from" and "to" year using the dropdown menu to concentrate on changes within those specific years.
    - Rationale: This interaction allows the academic to examine selected years for significant changes or trends while maintaining the ability to explore the dataset as a whole.

- 

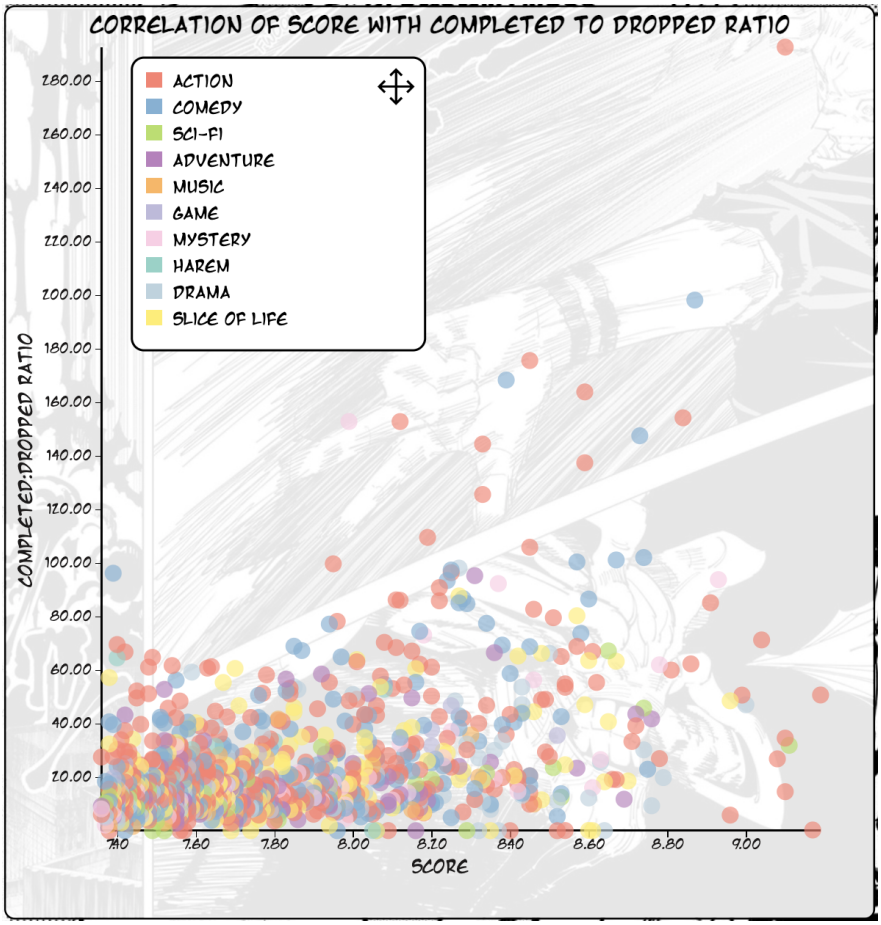
- Tooltip
  - Hovering over a genre segment provides detailed information, including the exact number of anime produced and the genre name.
  - Rationale: This feature enhances the user's ability to gather precise data, supporting the academic in identifying specific genres and years for further investigation.



5.3 View 3: Scatter plot

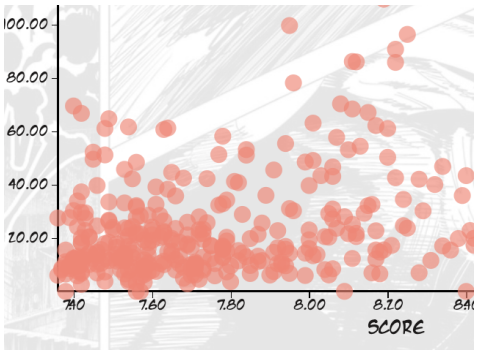
5.3.1 Description of view:

View 3 is a scatter plot designed to explore potential correlations between an anime's score and the ratio of viewers who have completed it versus those who dropped it. Each data point on the scatter plot represents an individual anime, with the x-axis indicating the anime's score and the y-axis portraying the completed-to-dropped ratio. The genres of the anime are visually distinguished through categorical colors. A tooltip feature is integrated, providing detailed information about each anime upon hovering, including its name, score, age rating, studio, premiere date, media source, and the completed-to-dropped ratio. Notably, the points on the scatter plot are slightly opaque, allowing users to observe data density and patterns, especially in areas where correlations between score and completion ratios are more pronounced. Furthermore, the scatter plot can also act as a global filter as clicking on specific points will filter all views to display anime with only the selected anime’s genre.

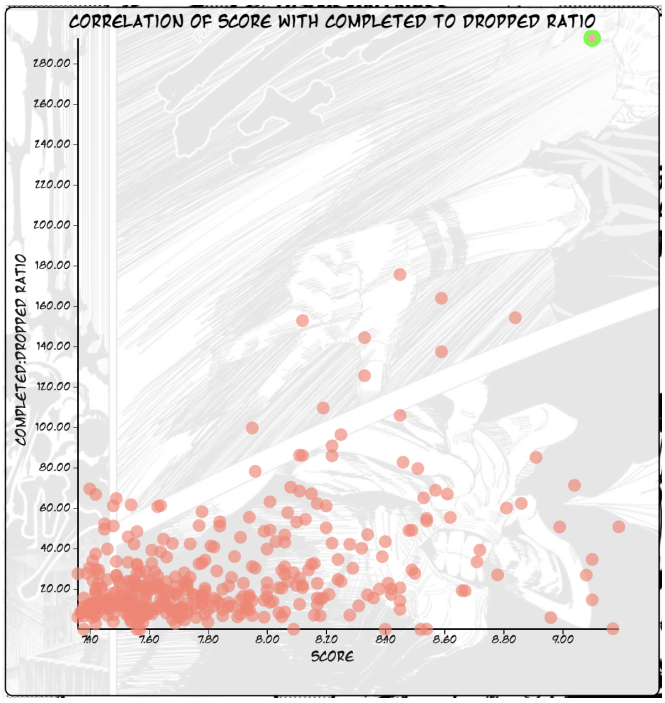


5.3.2 Rationale For View:

The scatter plot helps indicate trend correlating between the score and ratio of watchers who have completed vs. those who dropped the anime. With its x-axis representing the score and the y-axis indicating the completed-to-dropped ratio, it provides an effective means to explore potential correlations, supporting **task 5**. Users can visually inspect the arrangement of data points and identify trends or patterns that indicate the relationship between an anime's score and its completion ratio. The 2D space allows for an intuitive analysis of how changes in one variable correlate with changes in the other. This visual representation enhances the user's ability to uncover meaningful connections between the two variables. Furthermore, the slightly opaque points in the scatter plot serve a crucial role in supporting this task. The opacity allows users to gain insights into the density and distribution of data points. In areas where there is a concentration of points, users can quickly identify clusters or patterns, especially focusing on regions where higher opacity indicates a greater density of anime. This feature aids in identifying genres or specific score ranges with a higher concentration of critically acclaimed anime, making it easier for users to pinpoint areas of interest within the scatter plot.



In a scatter plot, each data point represents an individual anime, with the x-axis representing the score and the y-axis representing the completed-to-dropped ratio. The use of 2D space allows for a visual representation where data points that deviate significantly from the general pattern or trend can be easily identified, supporting **task 3**. Outliers, which may have extreme values in one or both dimensions, stand out visually as points that are distant from the main cluster. The 2D space provides a contextual framework for understanding the position of each data point relative to others. This is crucial for identifying outliers that may not only have extreme values but also exhibit unique relationships between their score and completion ratios. For example, an anime with an exceptionally high score and an unusually low completion ratio or vice versa, can be quickly spotted and examined within the scatter plot. The ability to see outliers in 2D space is not just about spotting anomalies; it's about gaining actionable insights. Users can identify specific anime that deviate significantly from the norm, allowing for focused investigation into the characteristics, reasons, or trends associated with these outliers. This is essential for making informed decisions or drawing attention to unique and noteworthy cases.



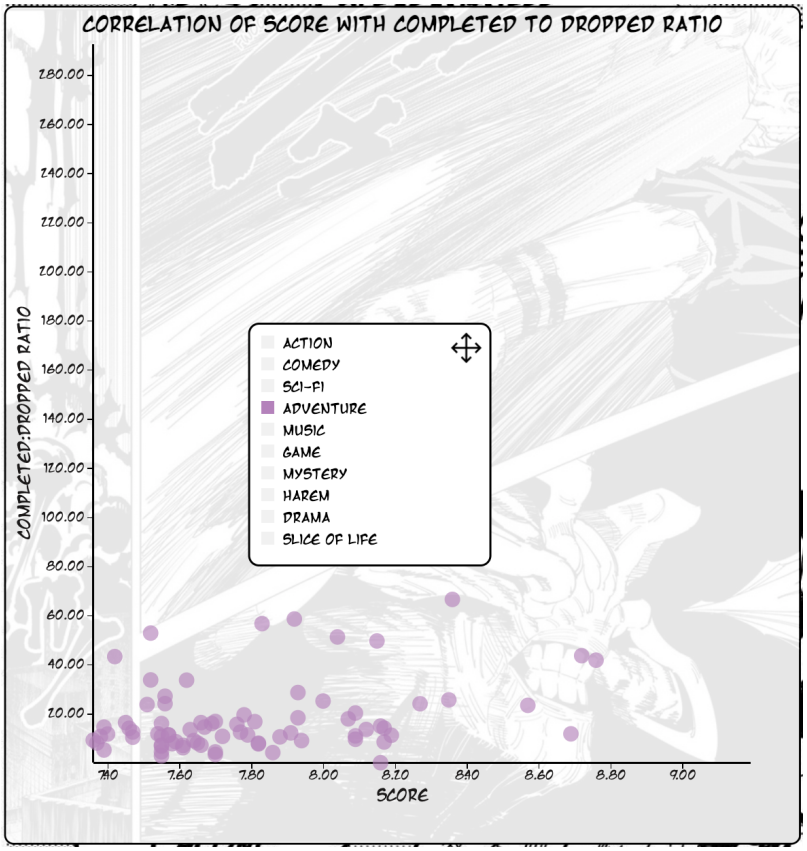
The scatter plot's utilization of 2D space for correlation analysis and the incorporation of opacity for insight into density collectively empower users to accomplish the specified tasks efficiently. The visual representation facilitates a nuanced exploration of the dataset, enabling users to make informed

observations about the relationship between an anime's score and its completed-to-dropped ratio while also providing insights into the overall density and distribution of data points.

Lastly, we opted for a scatter plot instead of another type of chart such as a bar chart, as we wanted to be able to visualize the overall trend, without losing sight of outliers. Thus, we believe that this is the easiest and most standard way people expect to see this information.

5.3.3 Interaction and Linkage:

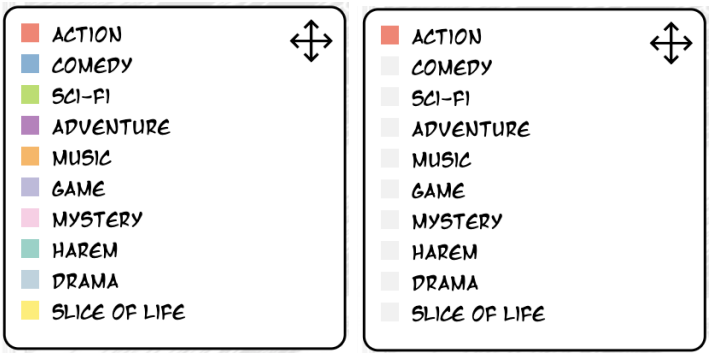
- Genre Filtering:
    - Similar to other views, a global legend is incorporated to enable users to filter data based on specific genres. But by clicking on a specific point in the scatter plot, it also sets the global filter to filter to only the same anime. This feature transforms the scatter plot into a genre-specific visualization. (Described in detail in section 5.4)
    - Rationale: Allows users to focus on the relationship between score and completion ratios within a particular genre.



- Tooltip for Detailed Information:
    - The tooltip feature enhances the scatter plot's functionality by providing detailed information about each data point.
    - Rationale: Users can gather comprehensive details about an anime by simply hovering over its corresponding point, supporting a more in-depth analysis.

5.4 Linkage:

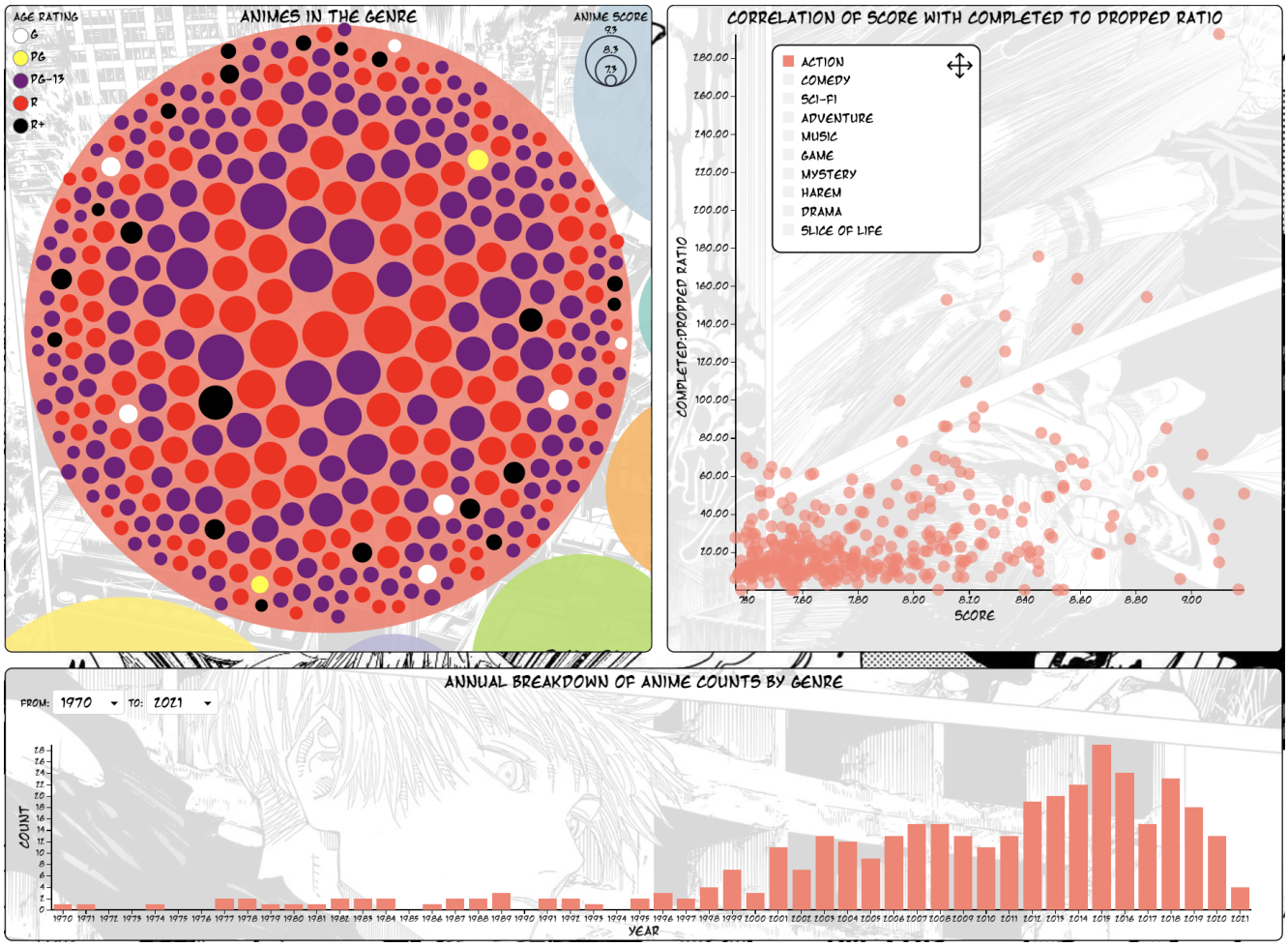
5.4.1 Global Legend:



The global legend can control the genre selection of all 3 views and reset the selection.

Rationale: The global legend serves as a reference point for the colors of the genres which are shared across all 3 views, thus the global legend provides a method to globally select a genre and reset the selection.

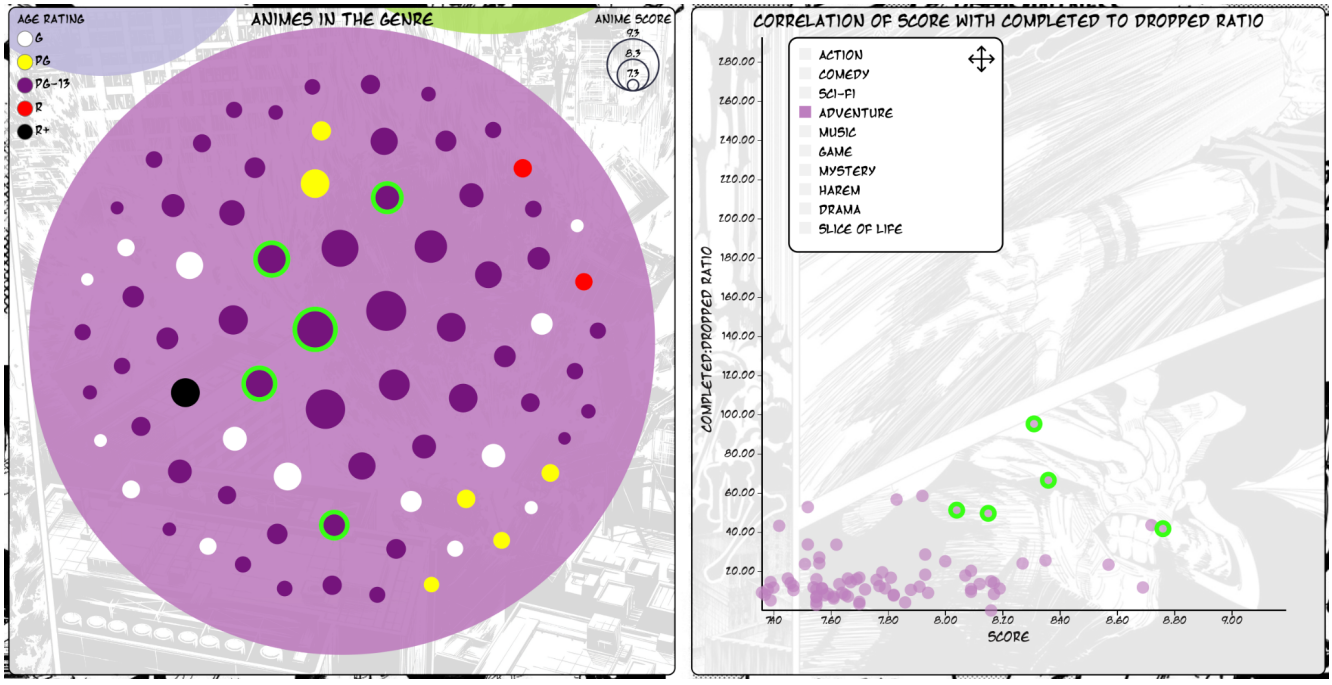
5.4.2 Global Filtering



Our global legend has control over all 3 of our views. When the user clicks on a particular genre on the global filter, our main view, stacked bar chart, and scatter plot will only show data within the selected genre. Additionally, when the user clicks on the white space of the legend, the global legend will reset. All of our views will reflect accordingly and show data including all the genres. The scatter plot also acts as a global filter. By clicking on any of the points on the scatter plot, the views will change to display only the data of the selected genre. Similar to the global legend, clicking on the background of the scatterplot after a point is selected will bring back all the genres.

Rationale: The linkage enables users to seamlessly transition between different views, allowing them a more nuanced understanding of the data by examining it from various perspectives. We realize that it's not beneficial to compare anime ratings for anime in different genres and want to limit anime to anime comparisons to within genres. This is because audience preferences and criteria for what makes an anime outstanding can vary significantly between genres hence the comparison of anime in two different genres has little meaning.

5.4.3 Bidirectional Linked Views



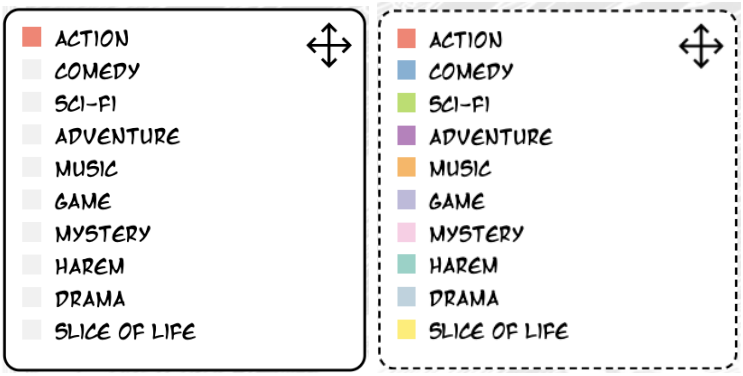
The packed bubble chart (View 1) and the scatter plot (View 2) are bidirectionally linked. By clicking on any one of the genre circles in the packed bubble chart, it filters the scatter plot to display only anime with that genre. In the other direction, by clicking on a specific point in the scatter plot, the scatter plot will display only anime of that genre, and the packed bubble chart will display only the packed bubble chart for that specific anime. Furthermore, when selecting a specific anime in the packed bubble chart, a neon green circle highlights the selected point. The same anime on the scatter plot is then also highlighted in neon green. Likewise, when selecting a specific point on the scatter plot, the point will be outlined in neon green, with the same anime's bubble in the packed bubble chart also being outlined in neon green.

Rationale: Bidirectionally linked views in our exploration tool seamlessly connects the packed bubble chart and scatter plot, enabling users to navigate genres and individual anime with ease. Clicking on genre circles refines the scatter plot, focusing on specific genres, while selecting an anime in the packed bubble chart highlights it in both visualizations. The consistent neon green highlighting enhances user interaction, providing a cohesive and intuitive experience for exploring and comparing anime data across different dimensions.

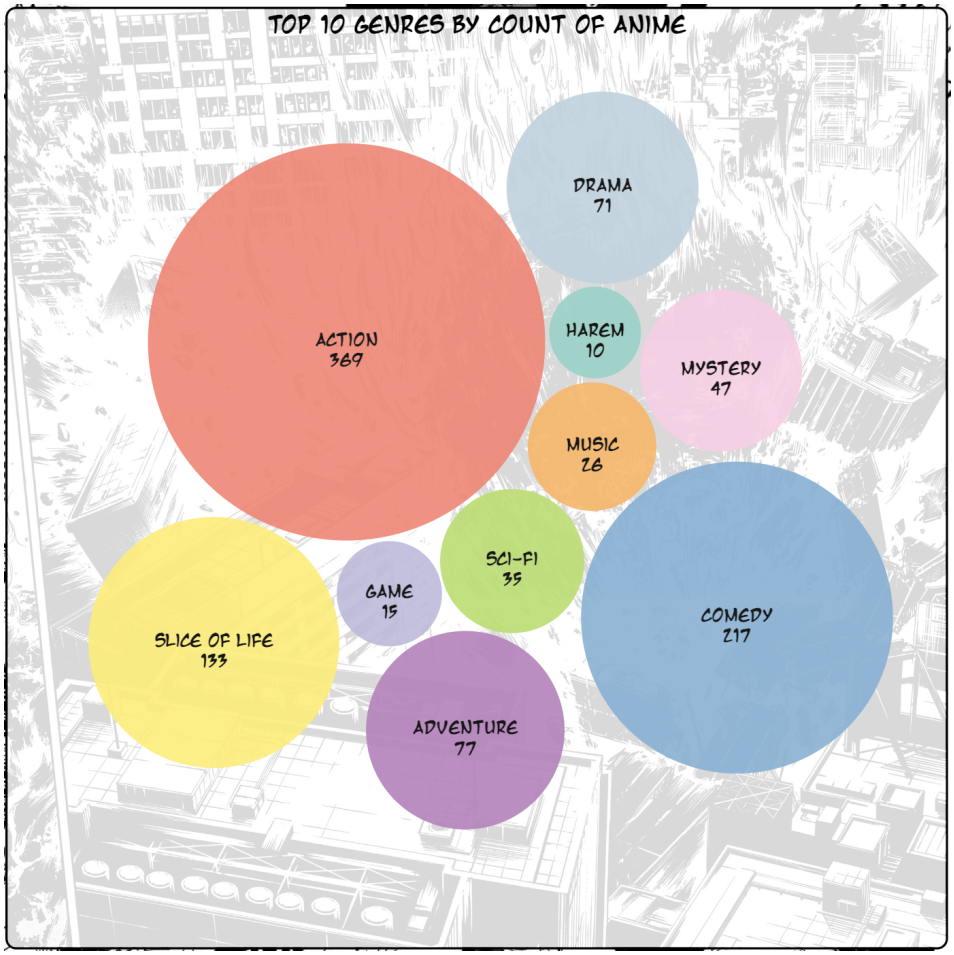
6. Usage Scenarios

Sarah, an anime enthusiast who has recently finished watching her favourite anime series of all time, is now on the lookout for something new to binge watch. She is looking for a platform that not only displays anime based on her preferences but also provides insights into different anime genres, production studios, and other features. Sarah visits the anime recommendation application, and immediately notices View 1, a vibrant bubble chart with the names of anime genres, as well as the number of animes with that genre as their "main" genre present in the top 1000 anime in MyAnimeList.

Sarah then notices that there is a draggable global filter that allows her to filter all three views. She clicks on the colour icon for action as that is her favourite genre, and notices that by clicking on the action colour icon again, all the genres return to all the views.

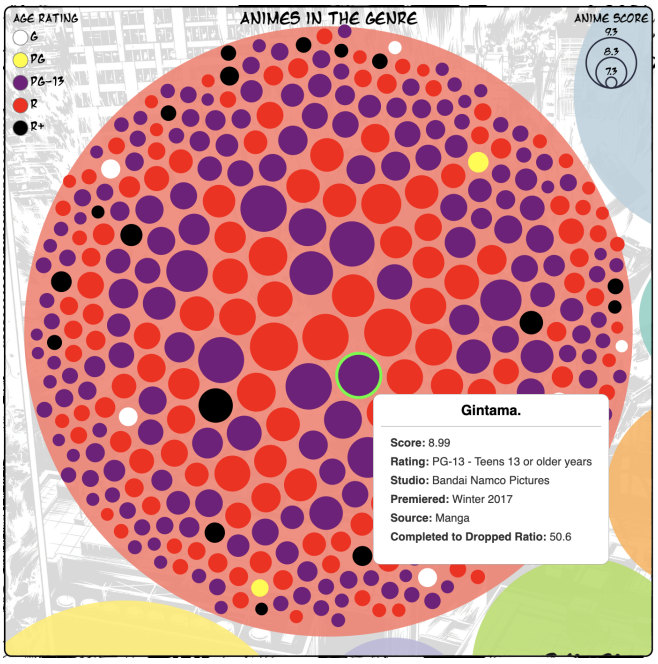


Now with curiosity, she hovers on the “Action” bubble as that is her favourite genre and sees the number of action anime there are in the top 1000 anime.

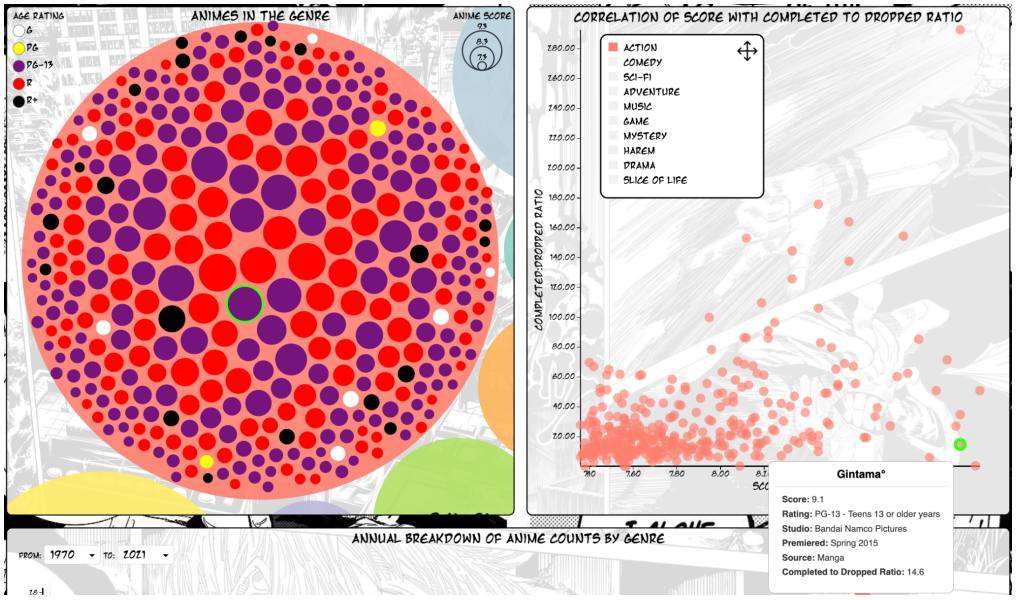


After clicking on the “Action” bubble, the bubble chart (View 1) drills down into a packed bubble chart displaying radially organized bubbles, with the largest being in the center. After noticing that the

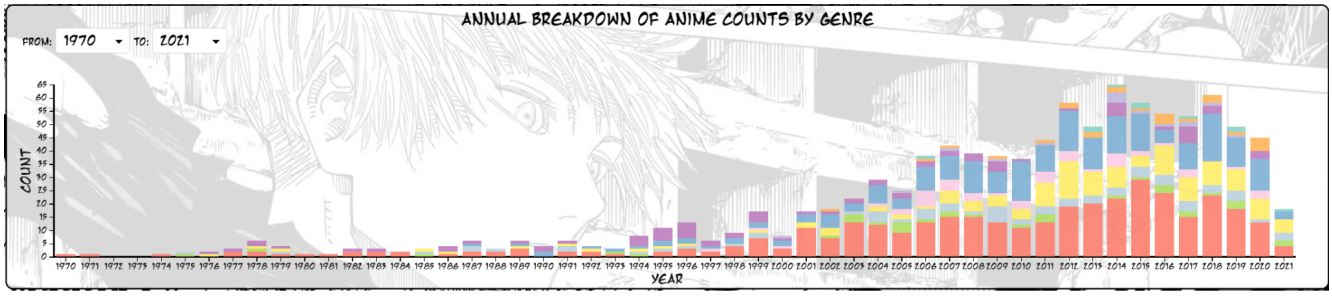
largest bubbles are in the center, she sees the legend indicating that the size of the bubbles are representative of the score of the anime. However, Sarah is only interested in PG-13 anime, so she looks to the packed bubble chart's (View 1) legend in the top left corner and sees that PG-13 anime are indicated by the colour purple.



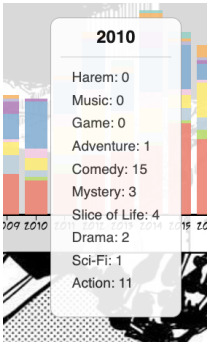
Sarah clicks on the largest purple circle she can see in the chart and notices a neon green outline appear around it, as well as a neon green circle appear around a specific point in View 3 (scatter plot). Now, Sarah can find the top-rated anime in the "Action" genre that are PG-13 like she wanted to. Sarah also noticed that by clicking the "Action" bubble, the other two views, a bar chart (View 2) and a scatter plot (View 3), now also only display data corresponding to the action genre.



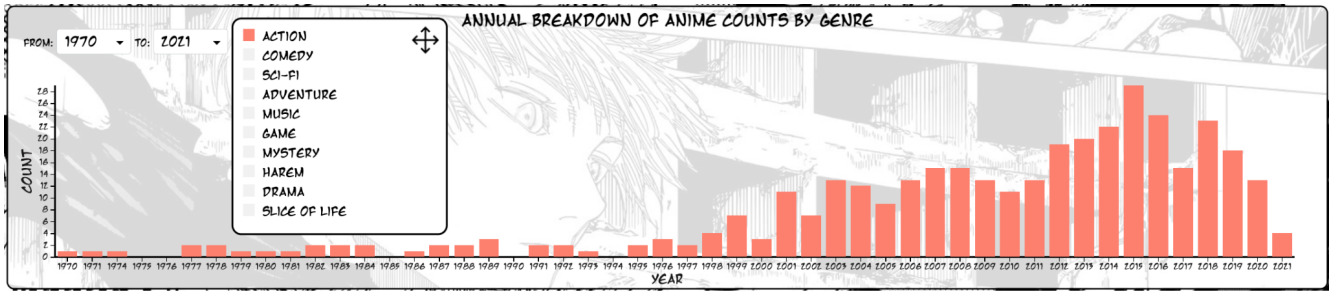
Sarah is also interested in exploring how anime tastes have changed over time for her favourite genre and see if it has gotten more popular over the years. She can refer to View 2 in the application, a bar chart,

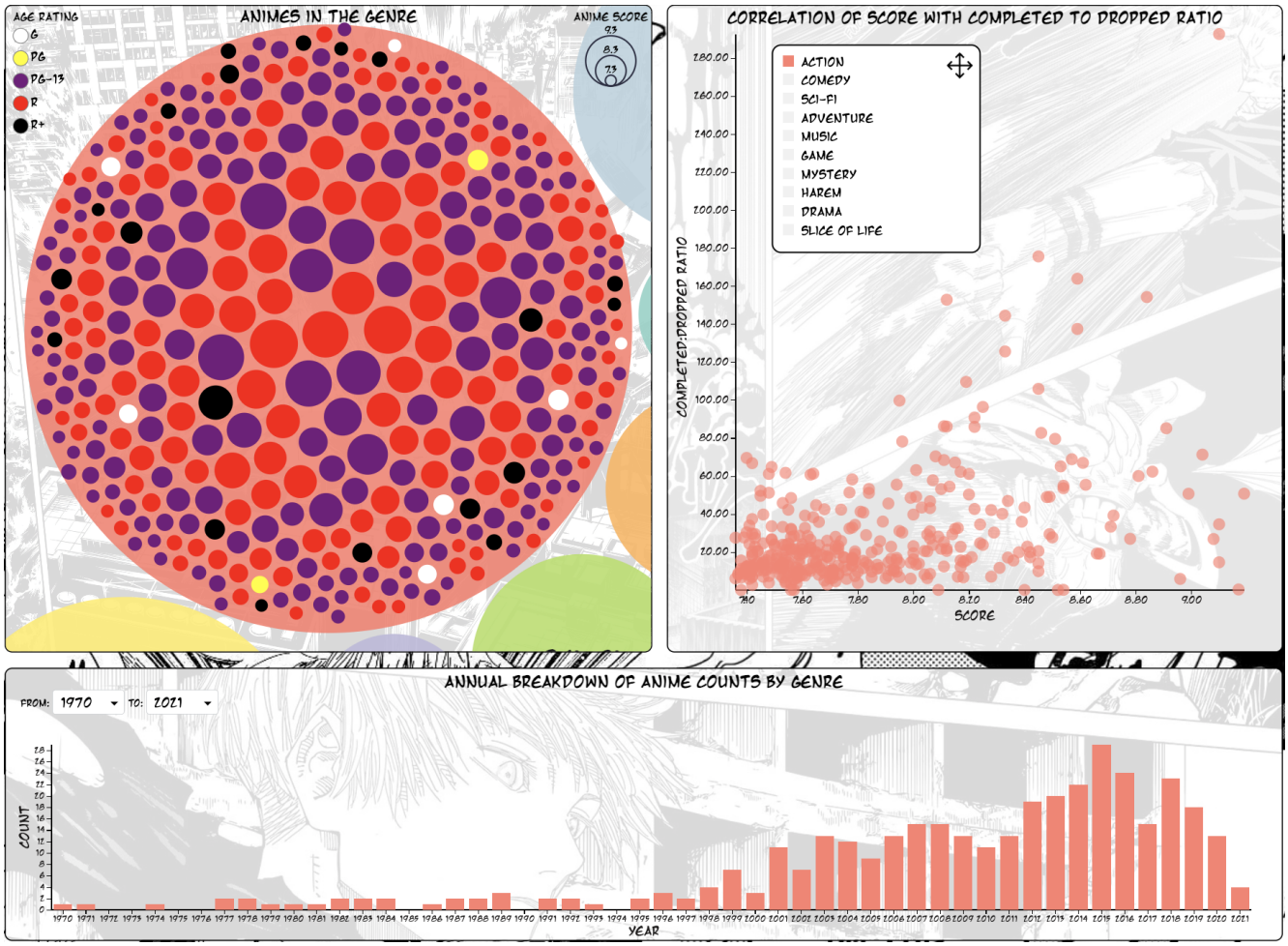


Sarah is curious about anime in 2010, so she hovers her cursor over the bar in 2010 and sees a box pops up providing detailed information, including the exact number of anime produced and the genre name.

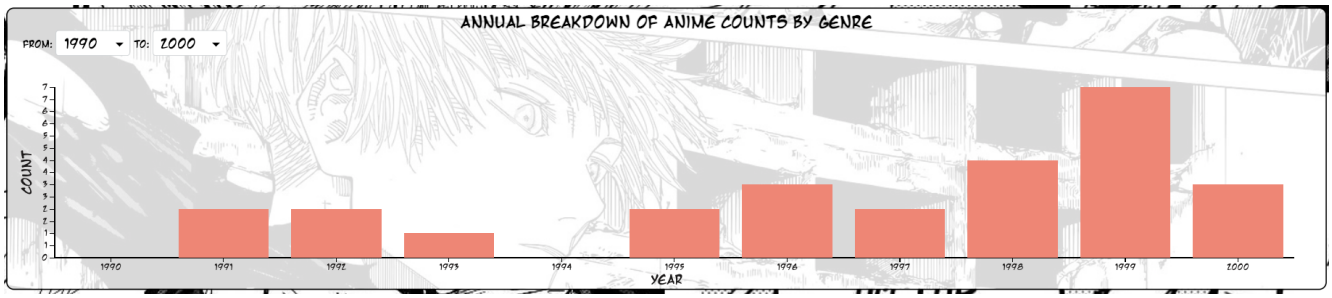


She then selects her favourite genre from the global filter which follows the user around, allowing her to explore how anime preferences have changed over time.

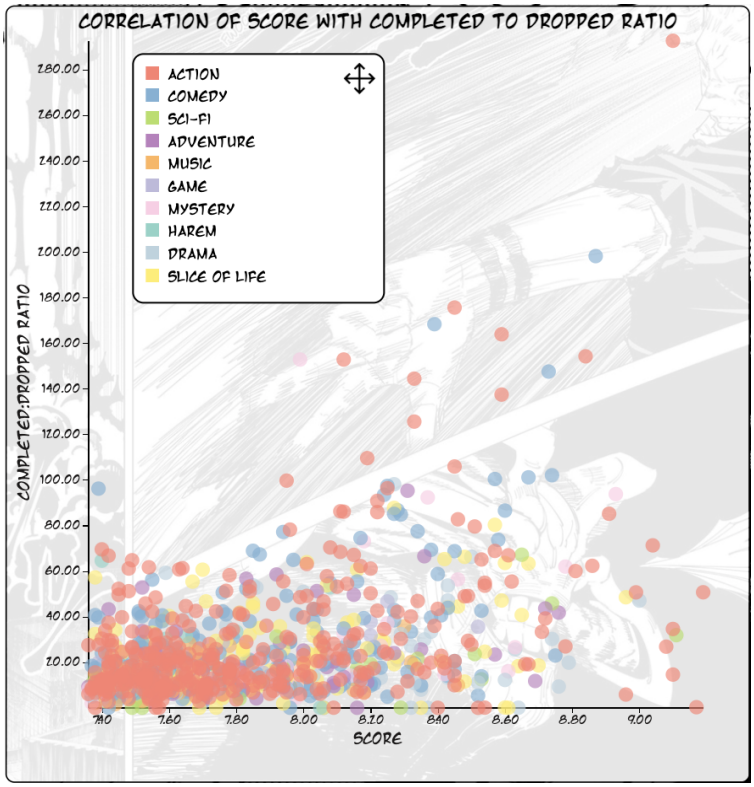




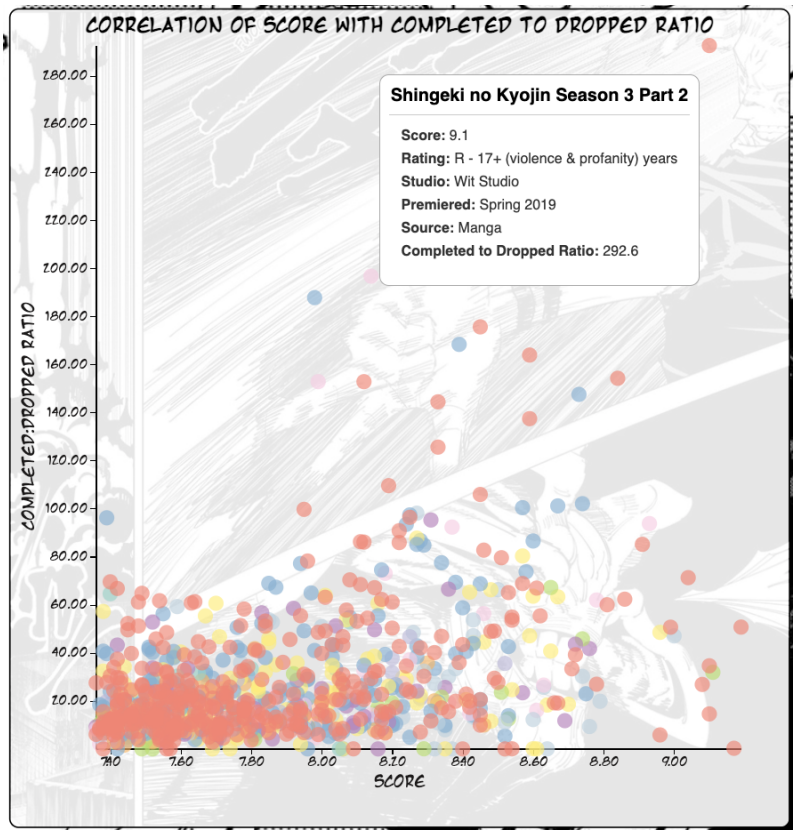
However, Sarah is interested in the anime produced during the years 1990 to 2000 and is curious if she can take a better look. Sarah then uses the year drop down UI widgets to adjust the year range in the bar chart. Satisfied with the fact that she can now see each year more clearly, she takes a closer look.



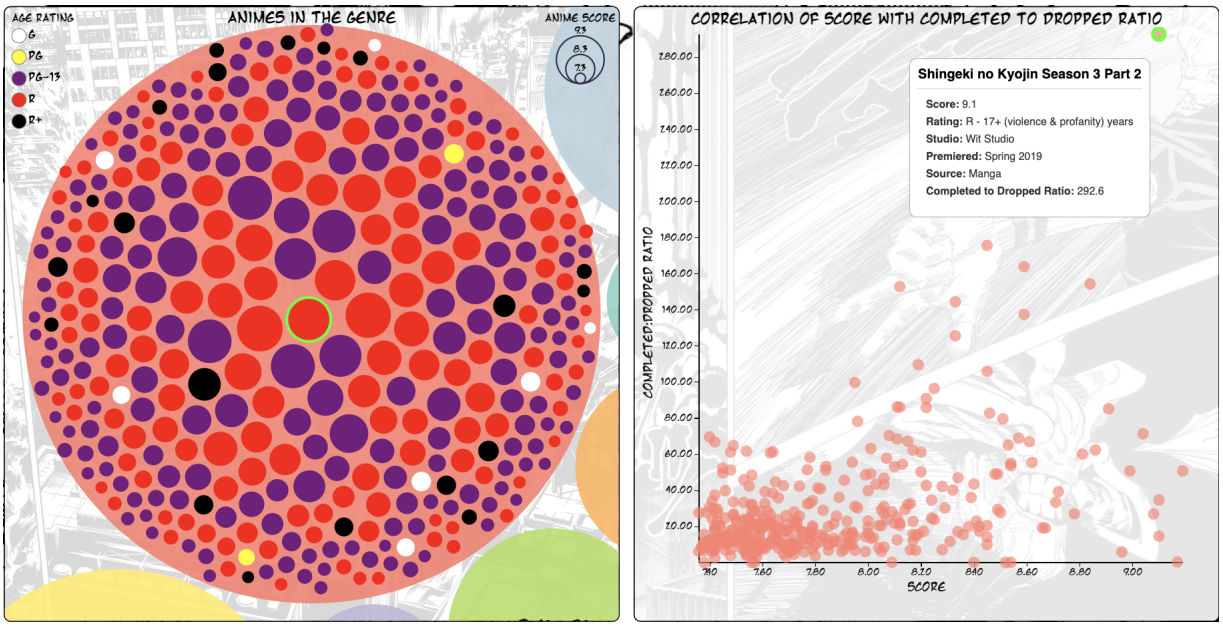
Sarah has now narrowed down her next anime to the action genre, but wants to make sure that the anime she chooses to watch next not only has a high score but also has a lot of people completing it which would indicate that the anime's story is consistently good throughout all its episodes compared to if people were to drop it, which would indicate the story weakening throughout the runtime of the anime. To do this, she can refer to View 3, a scatter plot which plots the completed/dropped ratio against the score.



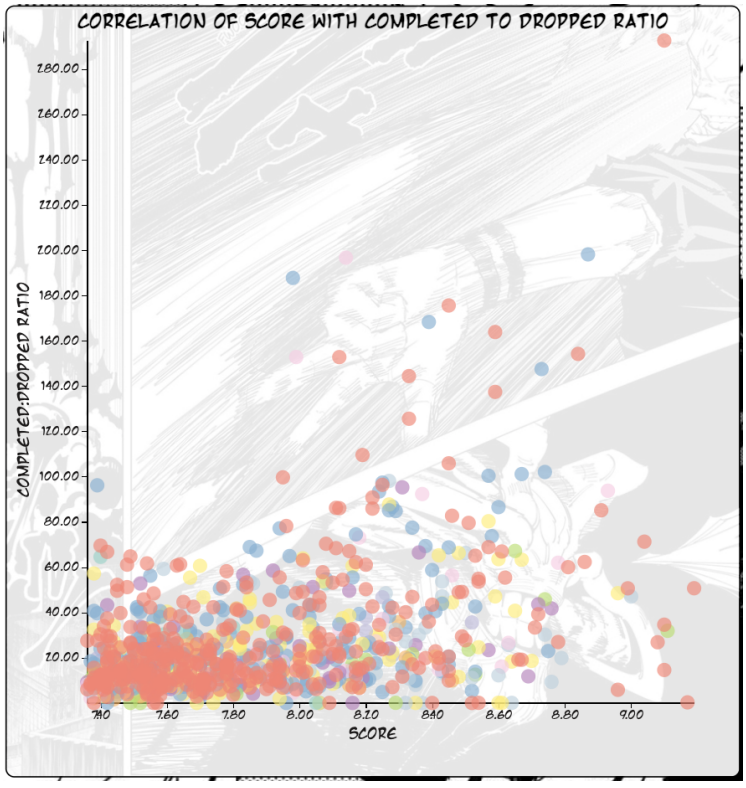
Sarah hovers over a point that shows that many users have finished the anime and have also rated it highly and sees a box pop up displaying the anime's name, score, age rating, studio, premiere date, media source, as well as the completed to dropped ratio.



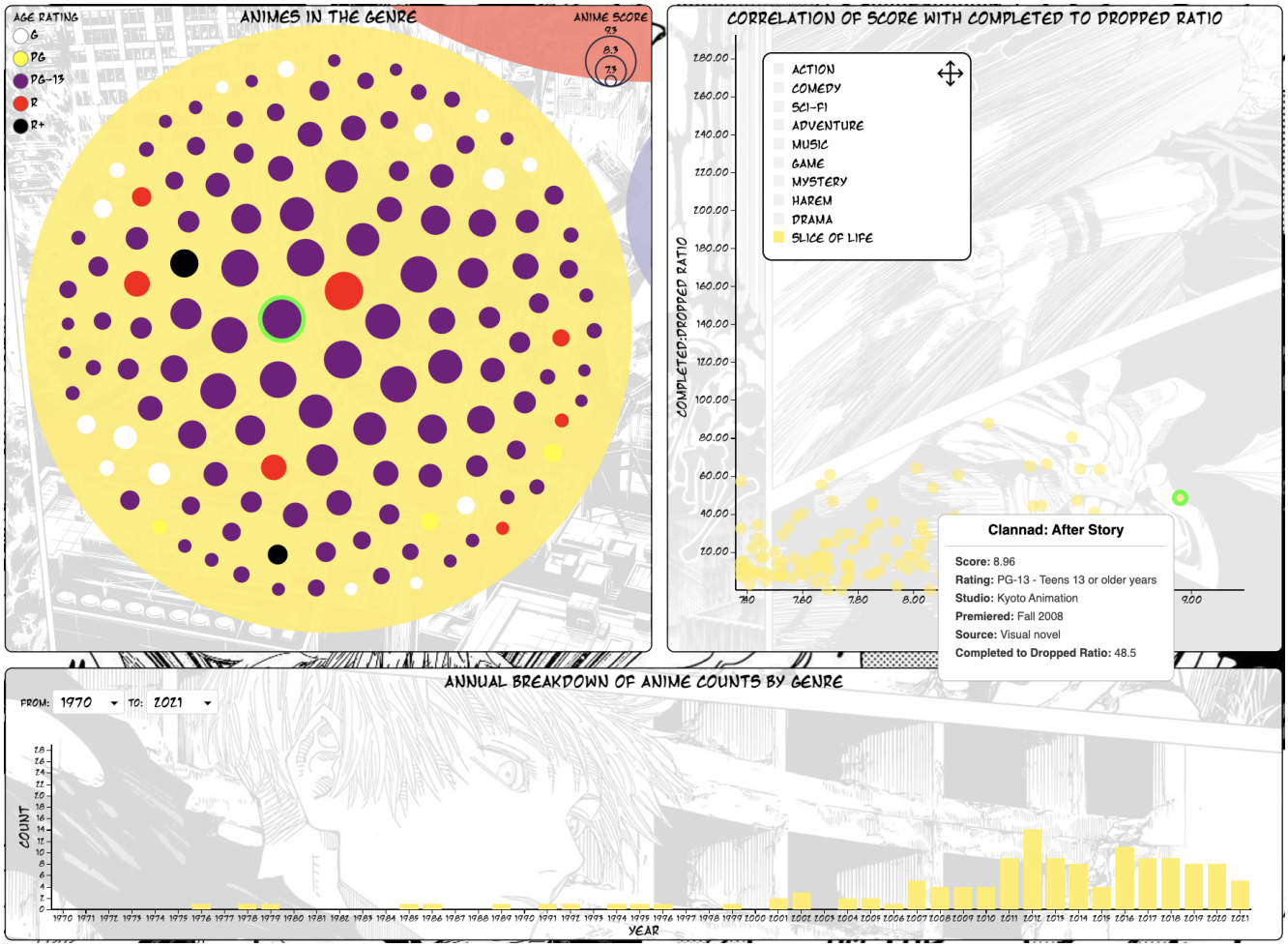
Sarah then clicks on that point and notices that there is a neon green highlight around this point. Sarah then also notices there is a corresponding a circle with a large neon green outline near the center in the packed bubble chart (View 1), which allows her to see how the score compares to other anime in that genre.



Sarah clicks on the background of the scatter plot because she wants to see how all anime compare against each other and thinks that she might want to watch some slice of life anime next.



She then clicks on a light yellow point in the scatter plot, and it filters both view 1 and view 2 to display that specific genre as well.



7. Reflection

7.1 Describe how your project has developed from your initial proposal, through your first submission, to your final product.

Our project has changed significantly from our initial proposal. Initially, we planned to create an anime recommendation program to compare and contrast the features of two anime so that a user can find an anime similar to the one they are interested in. The main view has changed significantly because we realized that the dataset is too hard to visualize with the efficient use of screen space if we put all the anime in different genres on the same view, resulting in 1000 data points showing at the same time. The linkage will be hardly visible. Instead, we decided to create two views, the top level bubbles categorized by genre and a drill-down view for each specific genre that enables the exploration of different anime in the genre.

The interaction between the main view and the scatter chat has changed. Selecting an anime on the scatter chart will zoom in on the genre of the selected anime on the main view and highlight the

selected anime in neon green color on both charts. It will also hide the anime in other genres on the scatter plot.

We added a year range filter to enable the ability to display anime count by genre in a stacked view in different year ranges. This gives the user more visual details of the trend in years that they are interested in.

7.2 How have your visualization goals changed?

We decided to change our original goal of showing all the anime on a flat hexagon shaped grouped by genre to allow comparison between different anime, aiming to recommend anime that the user is interested in. We implemented a top level force directed bubble chart for all the genre with the count of anime produced in each and a drill-down force direct bubble chart to show the anime in that specific genre. The user is able to compare the popularity between genres by looking at the size of each genre bubble, which is determined by the number of anime produced in the genre.

In the drill-down view, we changed the channel encoding for the anime from using color to represent genre to represent the age rating of each anime within the genre following the motion picture content rating system. Using the force directed bubble chart, the higher rated anime circles are grouped in the middle, and the lower score ones radiate outward in a radial fashion.

The user will transit from top level to drill down by clicking on the genre, or the anime on the scatter plot to identify its properties such as score, age rating, and the popularity of the genre over a number of years with the linking between three views. We added a year range filter on the bar chart to give the user the ability to focus on specific year ranges and the ability to see the overall trend.

7.3 How have your technical goals changed?

Our technical goals are adjusted to fit our new visualization goal of drawing a new top level bubble chart and a drill-down bubble chart. We have to create more animation to allow the transition between top level to drill down view on the unique chart. Implementing the force directed graph is more math intensive in calculating the force involved with the bubble movement on the chart. We added animation to the bar chart, aiding the transition on the bars and axis to adjust for different year range inputs.

7.4 How realistic was your original proposal in terms of what is technically possible in D3?

We would say that our proposal was relatively realistic in terms of the technical aspects. The standard visualizations, such as bar charts and scatter plots, were well within the capabilities of D3, and we were given many examples and guidance on how to create them.

However, our unique view had more challenges and required additional workarounds.

7.5 Was there anything you wanted to implement that you ultimately couldn't figure out how to do? If so, then what workarounds did you employ, or did you abandon your original idea?

In our unique view, we initially used the d3.pack library to create the bubbles for different anime genres and anime within each bubble. We run into an issue related to the size of each bubble drawn is not on the consistent scale. d3.pack() calculates its radius based on the number of data points in each category and the drawing space available, which results in the genre with a low number of anime being drawn with a larger radius. This creates a visual illusion that the anime is higher scored among others in a different genre that has more anime. Our first approach is to provide an adaptive legend that is used as a reference for the score of each anime, and it resizes based on different genres. After some research, it was proven rather difficult to draw circles with a given input number using the default d3 scale to create such a legend. Therefore, we decided to define a radius scale using scaleSqrt() to determine the radius for the individual anime bubbles. This allows us to ensure a uniform radius scaling across different genres, despite the number of anime in a specific genre. It also gives us the ability to create a fixed legend with a certain score input for reference.

We decided to split our unique visualization into two parts, a top-level and drill-down view grouped by genre, to present the data in a more detailed and organized fashion. This is different compared to our original proposal of rendering all 1000 anime on the same view in a hexagon shape and grouping anime in each genre by one edge in the graph. The shortcoming of the initial proposal is that some genre has way more data points to render than other genres. There will be clustering hotspots on the graph, making it look crowded in certain areas of the graph visually. Our workaround is to represent each genre in bubbles, whose sizes are determined by the number of anime in that genre. After the user selects a genre, the anime bubbles are rendered based on the chart space to be equally spread out over the entire chart area to avoid data clustering.

7.6 If you were to make the project again from scratch (or any other interactive visualization), what would you do differently?

- Try out different ideas early on to see what works best for showing information and getting feedback.
- Explore many different visualization techniques or layouts, especially for unique views or complex datasets.
- Test out various types of charts, colors, and interactive features to make sure users can easily understand and make use of the visualization.

8. Work Breakdown and Schedule

We met on a weekly basis and hosted ad-hoc meetings to discuss visualization ideas and used pull requests to merge our code and resolve any merge conflicts. We also helped each other out to troubleshoot and bounce off ideas in making the changes from our initial proposal to meet the requirements of the project better.

Milestone 1

Task	Team Member	Completion Date	Hours Taken	Status
Overview writeup	D, W, I, J	Oct 1, 2023	0.5hr	Completed
Finding dataset	D, W, I, J	Oct 1, 2023	1hr	Completed
Data pre-processing	D, W, I, J	Oct 1, 2023	1hr	Completed
Tasks	D, W, I, J	Oct 2, 2023	1hr	Completed
Team Communication plan writeup	D, W, I, J	Oct 2, 2023	0.5hr	Completed

Milestone 2

Task	Team Member	Completion Date	Hours Taken	Status
Adding abstract tasks	D, W, I, J	Oct 22, 2023	1hr	Completed
Visualization sketch ups/High level descriptions	D, W, I, J	Oct 22, 2023	1hr	Completed
Usage scenarios	D, W, I, J	Oct 22, 2023	0.5hr	Completed
Work breakdown	D, W, I, J	Oct 22, 2023	0.5hr	Completed

Work remaining:

General Task	Task	Team Member	Expected Completion	Actual Completion	Expected Hours	Actual Hours	Notes
	Process and cleanup	D	Nov 1	Nov 5	6	4	
	layout design	D, I	Nov 5	Nov 29	6	5	
	art assets	I	Nov 16	Nov 26	4	4	
	global legend	D, I, J, W	Nov 15	Nov 26	8	8	
	font and page title design	I	Nov 23	Nov 23	6	6	
	Add page scaling	D	Nov 30	Dec 1	4	4	
	Write up	D, W, I, J	Dec 3	Dec 3	10	10	

Unique View	**Task**	**Team Member**	**Expected Completion**	**Actual Completion**	**Expected Hours**	**Actual Hours**	**Notes**
	Define chart shape, area	D	Nov 5	Nov 10	6	4	change in m3
	top level bubble chart	D	Nov 14	Nov 15	5	3	grouped by genre
	drill down bubble chart	D	Nov 14	Nov 15	8	4	show animes
	title info for top level	D, J	Nov 15	Nov 21	5	5	add count
Static	legend for drill down view	D, J	Nov 15	Nov 23	5	5	
	transition from top level to drill down	D	Nov 16	Nov 16	6	6	
	transition from drill down to top level	D	Nov 17	Nov 17	6	6	
	text fade with transition	J, D	Nov 16	Nov 18	2	2	
Animation	legend fade	J	Nov 14	Nov 23	2	2	
	top level bubble force direction	D	Nov 11	Nov 14	3	3	
Force direction	drill down bubble force direction	D	Nov 12	Nov 15	4	4	Have to adjust charge strengths for each genre
	Highlight selected bubble	J	Nov 20	Nov 23	3	3	select and deselect
	Linking with scatter	J	Nov 25	Nov 28	8	8	dispatcher to states
Interaction/ Hover	Tooltips	J, D	Nov 27	Nov 27	3	2	p2 tooltips
Debug	Optimize transition animations and fix edge cases	D	Nov 28	Nov 30	6	8	Various edge cases have to be handled due to the number of transitions there are
Test	User testing	D, J	Nov 26	Nov 29	10	15	

Bar Chart	**Task**	**Team Member**	**Expected Completion**	**Actual Completion**	**Expected Hours**	**Actual Hours**	**Notes**
	Define chart shape, area	W	Nov 7	Nov 11	6	4	
	Draw x and y axis	W	Nov 5	Nov 11	2	2	
	Draw the bars	W	Nov 10	Nov 14	3	12	stacked by genre. Initially included both grouped and stacked bars

Static

	Create legend	W	Nov 15	Nov 15	2	0	dropped to use global
Interaction/ Hover	add year range filter	W	Nov 20	Dec 1	4	4	changed in M3
	Linking with unique chart	W	Nov 22	Nov 28	5	10	
Animation	axis and chart animation	W	Nov 26	Nov 26	6	4	
	Manual testing	W	Nov 24	Nov 24	3	3	
	Debugging	W	Nov 29	Dec 1	10	8	

Scatter Plot	**Task**	**Team Member**	**Expected Completion**	**Actual Completion**	**Expected Hours**	**Actual Hours**	**Notes**
Static	Setup chart area	I	Oct 30	Nov 2	1	1	
	Create X and Y axes and titles	I	Nov 1	Nov 5	1	1	
	Plot the points	I	Nov 2	Nov 10	2	4	
	Create legend	I	Nov 5	Nov 16	1	2	
Interaction/ Hover	Select genre and display data for that genre	I	Nov 8	Nov 17	4	8	
	Interactive Legend	I	Nov 10	Nov 11	2	4	
	Filter via points	I	Nov 12	Nov 11	4	4	
	Tooltip	I	Nov 17	Nov 17	6	2	
	Manual testing	I	Nov 20	Nov 29	4	6	
	Bug fixes	I	Dec 1	Dec 2	10	6	
	Auto-scaling	I	Dec 1	Dec 1	2	2	

Team Member	**Total hours per member**
Danny	4 (M1) + 3(M2) + 55(View 1 + General) = 62
Jason	4 (M1) + 3(M2) + 50(View 1 + General) = 57
Isabelle	4 (M1) + 3 (M2) + 55(View 3 + General tasks) = 62
Wonhak	4 (M1) + 3 (M2) + 50(View 2 + General) = 57
Total Hours:	238

9. Credits

9.1 Packed bubble chart

Force directed packed bubble chart inspiration:
https://observablehq.com/@sharad/force-directed-bubble-chart
The general structure of the code to achieve the force direction was used however, we used a square root scale rather than using the radius provided in d3.pack() to draw the sizes of the circles. We also manually tuned the forces and charge strengths to achieve the functionality we wanted.

ChatGPT was referenced to accomplish the effect of drawing the larger bubbles towards the center and pushing the smaller bubbles out towards the edges.

9.2 Bar chart

The idea of having both a stacked and a bar chart was given to us by one of the TAs.
Link: https://observablehq.com/@d3/stacked-to-grouped-bars?intent=fork
We drew inspiration from this example and initially opted for a transition from stacked to grouped bars. However, we later chose to stick only with the stacked bar chart and incorporated a dropdown for year control because we encountered readability challenges with the grouped bar chart. Despite similarities in code structure with the source, substantial modifications and the addition of functionalities were necessary due to differences in the dataset and our desire for extra features.

9.3 Scatter plot

I followed the CPSC 447 scatter plot D3 tutorial to get me started:
https://github.com/UBC-InfoVis/447-materials/blob/23Sep/d3-examples/d3-interactive-scatter-plot/
This allowed me to set up the frame of my scatter plot. After this, I was able to tweak and specify it to my needs (ie. axis, ticks and lines, points, scaling etc.)

I also used ChatGPT throughout the development of my feature to help with finding and correcting syntax for minor code blocks.

9.4 Art

I was heavily inspired by modern manga and in particular, Gege Akutami's series Jujutsu Kaisen. All of the background illustrations are taken from the manga. The title is also inspired by a move in Jujutsu Kaisen called "Domain Expansion."