**INFO 5100 Project 1 Report**

**George Gu (lg576), Jerray Wu (sw994), Xiaohan Wang (xw642)**

**Part 1. Visualization**

**Part 2. Data**

For our project, we used the “Video Games Sales Dataset” from Kaggle (link). This data set consists of video games sales data as of December 22, 2016, alongside some game ratings data scraped from VzCharts. The columns “Name,” “Platform,” “Year\_of\_Release,” “Genre,” “Publisher,” “NA\_Sales,” “EU\_Sales,” “JP\_Sales,” “Other\_Sales,” and “Global\_Sales” represent the sales data, whereas the remaining columns “Critic\_Score,” “Critic\_Count,” “User\_Score,” “User\_Count,” “Developer,” and “Rating” refer to the game ratings data. Given the nature of combining two separate sources of data, we can see the original data set have a lot of empty cells on the game ratings side. Hence, we decided to focus on the sales part.

We discovered that the data set includes games from as many as 583 publishers, with over hundreds of games alone for the more famous publishers such as Nintendo, Sony Computer Entertainment, Activision, etc. As it seems unrealistic to plot all of them out, we chose to compare and contrast the top ten publishers with the most games by looking at their “average” global sales. And we opted for a unique approach to compute the “average” by only taking three games, the one with the most global sales, the one with the least global sales, and the one with median sales, for each publisher. Unlike the traditional way of calculating the mean, taking a closer look at specific games allowed us to zoom in and explore the individual stories behind these games. Then we compared sales from North America, Europe, Japan, and other parts of the world to find out if there was any interest within specific regions for each of the selected games.

**Part 3. Design Rationale**

In order to evaluate the overall performance of a publisher, we needed to compute some form of total or average sales. In this case, we decided to sum up the global sales of the most and least popular games and the one in the 50th percentile for each publisher. We believed that these three games could, to some extent, represent the publisher’s overall performance, especially since the ten companies we chose to plot are all well-known, large publishers with hundreds of games. Bar charts are easiest to present categorial data. In order to zoom in and plot the three games for each bar, we split each bar into stacks, with each stack representing a game. We also decided to color code the stacks to represent the category (top sales, least sales, or median sales) that the game belongs to. Finally, we added some legends and labels to make the visualization more informative. In short, the marks used in the stacked bar chart are colored rectangles. The visual channels are varying the vertical aligned length (i.e., publisher), horizontal aligned position (i.e., publisher), vertical unaligned length and position (i.e., different games for the same publisher), and color hue of rectangles.

To keep zooming into each of the games, we would like to compare and contrast regional sales to identify any trend or popularity amongst a certain population at a given time. To achieve this, we used pie charts to display regional sales by percentages, because the total sales for each game is different. Besides, a pie chart also allows us to compare the percentage for a game in a specific region to that of another game in that same region. In this case, it would be less intuitive to plot the sales by number. To sum up, the marks used in the pie chart are colored sectors. The visual channels used are varying the area (quantitative attribute) and color hue (categorial attribute) of sectors. Overall, the pie charts serve as an extension to the stacked bar chart above and enable us to extract more insights as we zoom in further

**Part 4. The Story**

The stacked bar chart presents the top 6 publishers with the largest sales of games in the world, including Electronic Arts, Activision, Namco Bandai Games, Ubisoft, Konami Digital Entertainment, and THQ. Among those publishers, three of them are from North America (the top 2 are all from the United States), two are from Japan, and one is from Europe, which shows the dominance of the United States in global game sales. Additionally, the stacked bar chart presents best-selling game around world, 75th, and 50th percentile global sales.

We originally wanted to select the games with the highest sales volume, the lowest sales volume and the median sales volume produced by each publisher to plot games in stacked bar chart, but we found that the sales volume of the games with the lowest sales volume was too small compared with the games with the highest sales volume, so that it could not be effectively displayed on stacked bar chart. Therefore, we choose games with median sales and 75th percentile sales to show the sales gap of different games of each publisher, and also analyze the game categories and customer groups that the publisher is good at producing according to the types of games with the highest sales. (For example, the game with the highest sales volume in EA is FIFA 16. This publisher is a major producer of sports games. Many famous sports games and action games are produced by this publisher.).

The figure below stacked bar chart is donut chart, which shows the geographical distribution of sales of the games with the highest sales volume produced by each top 6 manufacturer. Among them, green represents sales in North America, blue represents sales in Europe, orange represents sales in Japan, and purple represents sales in other regions. Through analysis, it is found that the sales volume in Japan is generally the lowest, which may be caused by the small population of Japan compared with other regions. In addition, we found that the geographical distribution of sales volume had little relationship with the location of the publisher. The top 6 manufacturers mainly sold in North America, followed by Europe. It can be seen that video games are very popular in those two regions. In addition, because our team members are all from China, we found that Chinese users are more inclined to play web games or multiplayer online competitive games than North America and Europe in analyzing the types of games that Chinese netizens love. One reason may be that Chinese users are not as active in buying switch or PS5 as those in Europe and North America. Most Chinese users will choose to play games on PCs instead of consoles. Based on our own experience, we can infer that the reason why the donut chart shows a one-sided trend in Europe and the United States is that other countries may have similar reasons (for example, the per capita GDP is not as high as that in Europe and the United States, so people do not want to spend money on the purchase of console devices and games), which leads to console games being unpopular in their countries.

**Part 5. Team Contributions**

|  |  |
| --- | --- |
| **Name** | **Contributions** |
| George Gu | · Tested several types of visualizations and finalized the appropriate types we used for our product.  · Parts of coding, including processing data and stacked bar chart.  · Project Report Write-Up (Part 2 Data, Part 3 Design Rationale, Part 5 Team Contributions). |
| Jerray Wu | · Came up with the visualization ideas and proposed demos of those ideas using Figma.  · Majority of coding, including processing data, filtering, stacked bar chart, and pie chart. |
| Xiaohan Wang | · Examined the video games dataset and analyzed the elements we can use for our final project as well as their correlations.  · Parts of coding, including pie chart.  · Project Report Write-Up (Part 4 The Story). |