CS5044

Information Visualisation Group Report

220011707

180020555

Part 1 Data Preparation

The Web-based link:
https://zz67.host.cs.st-andrews.ac.uk/version_final/information%20visulization_P2.html#

This VideoGames dataset, which was retrieved from Kaggle(I.e., https://www.kaggle.com/datasets/gregorut/videogamesales), a renowned online platform for data science enthusiasts, to present game-related information. The dataset contains a list of video games with sales greater than 100,000 copies. Its original data contains 11 features consisted of categorical and ordered data. Our group chose the 8 out of 11 features to do information visualize, as follows Table-1 shows.

Features	Describe
Platform	Platform of the games release (i.e. PC,PS4, etc.)
Year	Year of the game's release
Genre	Genre of the game
Publisher	Publisher of the game
NA_Sales	Sales in North America (in millions)
EU_Sales	Sales in Europe (in millions)
JP_Sales	Sales in Japan (in millions)
Global_Sales	Total worldwide sales.

Table-1 Data features

The VideoGames dataset is suitable as it contains both categorical and ordered data, with year acting as a perfect filter. Thus, we chose it for visual analysis and focused on the time period from 2000 to 2010.

There are several tasks and open-ended questions generated from the VideoGames.

Question	Description	Visual Analysis Task
Q1	What is the annual change in global sales over the years from 2001 to 2010?	Comparison analysis: Comparing the different areas' overall sales to global sales every year.
Q2	The number of global sales varies according to the different genres categorized by different platforms.	Distribution analysis: Understanding the distribution of global sales data set, based on genres and platform.
Q3	What does the number of sales of different categories vary according to different genres?	Comparison analysis: Comparing two or more data sets to see how they relate to each other, such as comparing sales figures for different products or regions.
Q4	What is the trend each area's sales and global sales varying over the from 2001 to 2010?	Trend analysis: Examining the overall trend in the data over time.

Table-2 Visual analysis task and question

Part 2 Visual Encoding and Classifying Visual Techniques

Attribute	Attribute Type	Visual Variable
Platform	nominal	Color
Year	ordinal	position
Genre	nominal	position
Publisher	nominal	Color
NA_Sales	quantitative	Shape and color
EU_Sales	quantitative	Shape and color
JP_Sales	quantitative	Shape and color
Global_Sales	quantitative	Shape and color

Table-3 Visual encoding table

We based our rule of thumb for classifying visual techniques on the number of attributes. With 8 attributes, we chose bubble charts as they can display multiple aspects of information. The horizontal and vertical axes show sales and genres, while the bubble color shows the platform. We added the time attribute after setting the filter for the year, and this applies to stacked bar charts as well.

We combine the year-to-year sales ratio to save space and align with human reading habits. This graph reflects the change of game sales over time, providing a general picture of the data that helps increase design accuracy.

To follow the principles of expressiveness and effectiveness, we chose bubble and stacked bar charts to display multiple attributes, in line with the effective principle. The use of size and color to represent categorical data increased the design's discriminability and separability.

Part 3 Interaction Manual

Our project support the selection, filtering and querying tasks out of basic interactivity and tasks as specified by Shneiderman(Steele,1983).

The universe filter

When the universe filter is clicked, and dropdown bar will be shown listing all year's information, which would act as trigger to change the corresponding pattern in year line, bubble chart and stacked bar.

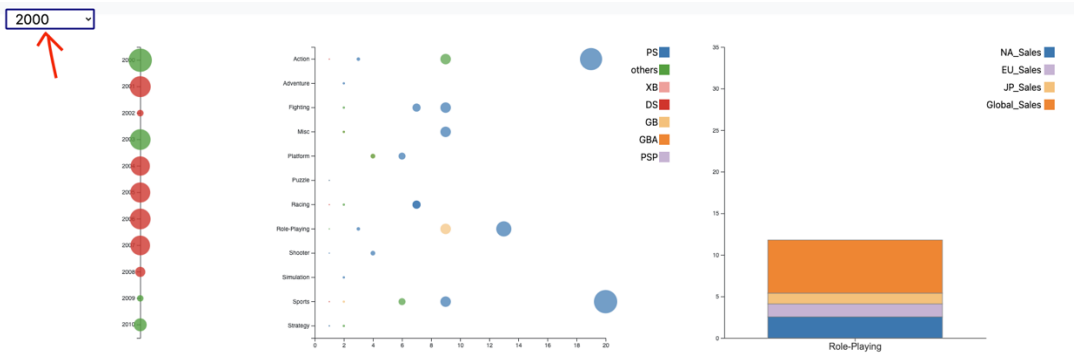

Fig-1 The filter dropdown bar

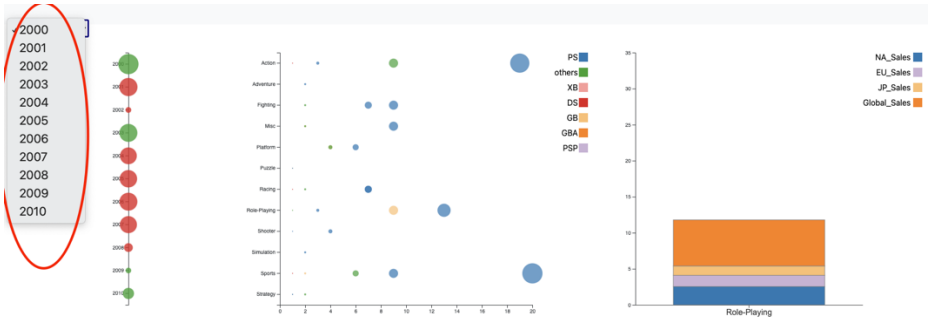

Fig-2 The filter dropdown bar opened

Hovering

The aim of this function is to better show the detail of the data, gaining the exploration capability for visualization.

The shape showed the scale of the global scale, and the color show the different publishers in Fig-3.

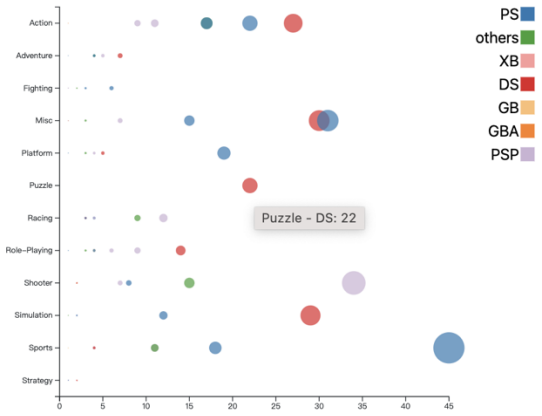

Fig-3 The hovering for bubble chart

Hovering in stack bar, the shape showed comparison of different sales using color as the distinguishing method.

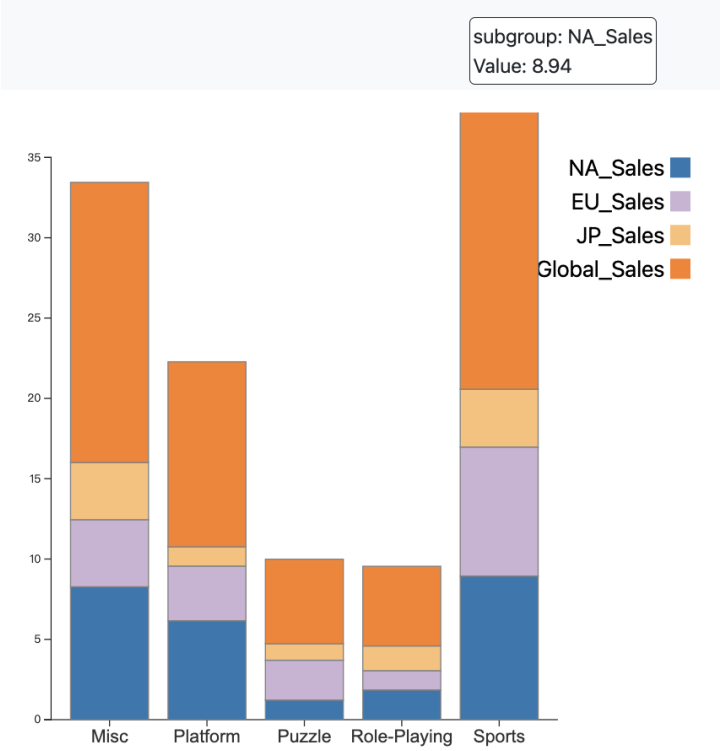

Fig-4 The hovering for stacked bar

Slider bar

Clicking the slider bar, the trend of the global sales would be shown to indicate the change of sales in different regions as Fig-5 shows.

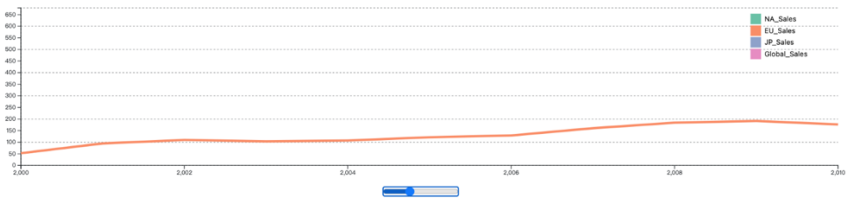

Fig-5 Click of slider bar

Part 4 Insight

Insight 1:

The Fig-1 graph shows global changes in game sales over time, with red indicating an increase and green indicating a decrease. The year-on-year growth rate is represented by the circle's radius, and hovering over the circles shows the exact annual change. Overall, game sales had a continuous growth rate except for 2000, 2003, and 2009-2010, where sales decreased.

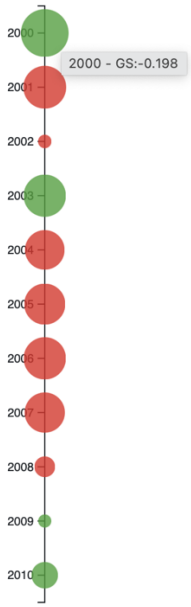

Fig-6 Year-on-year rate by time

Insight 2:

The bubble plot displays sales volume (million) on the horizontal axis, game genres on the vertical axis, and different game platforms as color circles. The graph shows that the best-selling platforms for different game genres change over time, such as PS being the top-selling platform for action and sports games in 2000, while Wills became the top-selling platform for sports games in 2007.

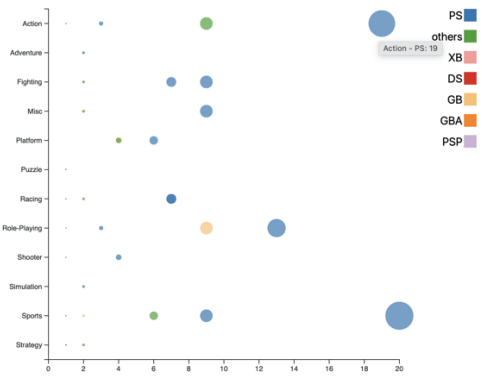

Fig-7 Year 2000 sales distribution

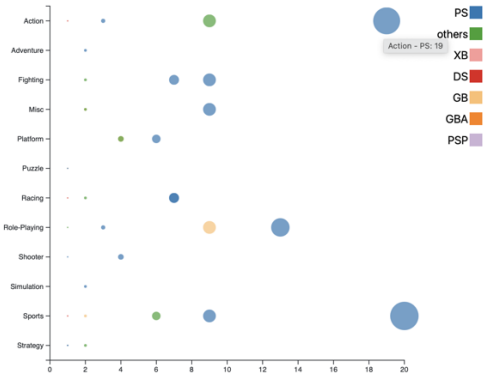

Fig-8 Year 2007 sales distribution

Insight 3:

The stacked bar of different genres illustrates the comparison between varied types of sales of same genres over time, those which not only able to observe annual change in global sales, but also to see the comparison to the other sales. Additionally, the comparison between different genres within one particular year can be spotted clearly through the height of the bar. Those can be achieved by clicking and choosing the wanted year in universal filter. However, during the data cleaning, some data had eliminated, so some bias might conjure up, on top of that, if the added data is bigger than the 35, the bar might beyond the y-axis while cannot be shown on the chart.

Insight 4

The line graph overtime would be the best way to show the change over time for different sales individually. Trend of sales is extremely important for exploration, which could tell the peak and bottom for different trends, and further explore collaborating with other graphs.

Part 5 Critical Reflection

Benefit:

Good visualization of the data both effective and expressive. Over 1000 dots and 8 attribute are used and good in color Platte.

Limitations:

There are some limitations in terms of exploration. One is the position of the filter might change when loading the web different screen sizes. Also due to the design is initially carried on entirely based on the size of the lab computer (I.e., 1920 – 1080), the layout would change when the size of the screen changes.

Future:

There is still room for improvement. In the future, we want to use some fancy techniques for visualization and to be more creatable.

References

D3. (2021a). *ine plot with dropdown to select group in d3.js*. Office.net. https://statics.teams.cdn.office.net/evergreen-assets/safelinks/1/atp-safelinks.html

D3. (2021b). *Stacked barplot with tooltip*. Office.net. https://statics.teams.cdn.office.net/evergreen-assets/safelinks/1/atp-safelinks.html

GREGORYSMITH. (2018, January 1). *Video Game Sales*. Www.kaggle.com. https://www.kaggle.com/datasets/gregorut/videogamesales

Steele, A. C. (1983). An Experimental Foundation for a Task Taxonomy. *The Journal of Data Education*, *24*(1), 25–27. https://doi.org/10.1080/00220310.1983.11646283