

**DECLARATION:** I understand that this is an **individual** assessment and that collaboration is not permitted. I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at <http://www.tcd.ie/calendar>. I understand that by returning this declaration with my work, I am agreeing with the above statement.

## Tools/Technology used

The tools that were used for this visualization was mainly Tableau, and a bit of Excel for formatting. I've used Excel to export the top 100 ranking values to a separate csv to reduce the filtering complexity when handling the data. Another pre-processing step is set up the path values as separate file when combining it with the original dataset file to create a particular visualization in Tableau.

## Dataset

This is a table dataset which contains information on the global video game sales throughout the years. **Rank**, represents ranking of the game based on global sales. It is a quantitative ordinal attribute. **Name**, represents the name of the game sold. It is a qualitative nominal attribute. **Platform**, represents the platform the games was released on. It is a categorical attribute. **Year**, represents the year the game was released. It is a sequential attribute. **Genre**, represents the genre of the game. It is a categorical attribute. **Publisher**, represents the publisher of the game. It is a categorical attribute. **NA\_Sales**, represents the sales of the game in North America. It is a quantitative attribute. **EU\_Sales**, represents the sales of the game in Europe. It is a quantitative attribute. **JP\_Sales**, represents the sales of the game in Japan. It is a quantitative attribute. **Other\_Sales**, represents the sales of the game in other regions. It is a quantitative attribute. **Global\_Sales**, represents the sales of the game worldwide. It is a derived quantitative attribute. The **NA**, **EU**, **JP**, and **Other** region sales can also be derived from the data table and grouped to represent as category.

This dataset contains a decent amount of data which wouldn't be possible to visualize all of them using a single established idiom. There are potential correlations between certain data and there are many qualitative and quantitative attributes that should be taken into account to be able to see a better picture of the story.

## Tasks

- **Identify** which of the game genres had the most sales for a particular region.
- View the **Trend** on the amount of sales that's been made over time at certain period of the years.
- **Lookup** which of the platforms had the most variety of games.
- **Discover** how many sales each platform made when taken into account of the top 100 selling games throughout all platforms.
- **Browse** the most popular genres that had the most sales within a particular region.
- **Compare** the amount of sales made within certain regions and the genre associated with it.

## Encoding Channels and Idioms

The visualization as a whole uses a juxtapose facet which shows different parts of the visualizations where each uses a different idiom.

### Area Chart

**Colour** is used to represent the region of the sales made. Colour is suitable to represent regions categorically. **Position** is used to represent the amount of sales made, as well as the year that the sales were made at. **Size** of the area of particular colour represents the amount of sales made for a certain region which. This is a suitable chart to display the amount of video game sales that were

made in total throughout the year, but also roughly view the separated coloured areas to identify from which region certain amount of sales were made.

### Stacked Bar Chart

**Colour** is used to represent the region of the sales made. **Position** is used to split the genre category into each bar at the left, and the bottom represents the amount of sales made in total for games sold of particular genre. In that case the **size** of the bar represents the number of sales made. The total sales made from particular genre go in **descending order**. This is a simple but a neat visualization method to showcase the total games sold to see which genres are the most popular. Colour is used to split the bars in segments where each segment of colour represents the region where the portion of the sales were sold at.

### Pie Chart

**Colour** is used to represent the region of the sales made. **Size** is used to represent the number of sales made within the region. Colour is a suitable channel used to identify categories. Size is a suitable channel to showcase the amount of sales made. The larger the size, the more sales were made. It is a simple chart to quickly show which region gets certain amount of sales.

### Packed Bubble Chart

**Brightness** of the colour represents the number of unique games sold for a particular platform. The stronger the colour, the more games there are for that platform. **Size** is also used to represent the number of unique games. Each circle has labels to identify the platform it represents and the exact value of unique games it has. This particular chart helps the user to easily identify by size and brightness of colour to see which platform contains most amount of unique games.

### Treemap Chart

**Brightness** of the colour represents the number of unique games sold from a particular publisher. The stronger the colour, the more games it has published. **Size** is also used to represent the number of games published. Each square map has labels to identify the publisher it represents and the exact number of games it has published. Similar to packed bubble chart, this chart also helps the user to easily identify by size and brightness of colour to see which publisher has published the most amount of games.

### Dendrogram Chart

**Colour** represents the platform the game was released on. **Size** of the bar represents the number of sales made for a particular platform. Along with the chart, a table on the right aligned for every bar shows the platform it represents and the total amount of sales made for that platform. The chart is aligned in **descending order** starting from highest sold. This is a unique chart which helps to group up top 100 selling games and visually showcase which platform sold the most portion of the total sales.

For detailed information on particular graphs, the user also has an option to interact with the graphs by opening the Tableau workbook file and inspect the full information such as exact number of sales made for a certain year and region in the area chart.

## Novelty

Due to a certain number of attributes available it was challenging to start because there was some analysis on the data that was needed to be done to see which of the data would be particularly useful to showcase. Data would have to be meaningful in the visualization to show correlation and tell a

story to the user. Some of the data could also be derived as a category itself such as region sales which adds up to the amount of attributes we are dealing with.

The best way to deal with the data was to create a juxtaposition facet to showcase the explanatory data using different idioms to represent a particular data for specific attributes. Dendrogram chart was a challenging chart to create as there were no built in methods or extensions which would create this particular visualization. Besides the implementation challenge, visually it helps to showcase the information to the user in an intuitive manner by showcasing the total top 100 sales which then splits into different paths which can be easily identified and perceived. The size of the screen was also a challenging limitation to the overall visualization as there was needed to be a good balance of information, but at the same time avoid over cluttering the space with diagrams making it look overwhelming and feel claustrophobic to the user when analysing the data. The balance of white space and the information displayed would also be affected depending on the visualization method used. Some methods can be perceived even if it's small, and some data need to be of certain minimum size to easily discern information the visualization is trying to show. So because of those concerns there was some brainstorming and trial and error to see which method of visualization would be ideal but at the same time appropriate for particular data attributes being used.

## Critical Analysis

The visualization is able to effectively communicate to the user on the data is being shown without much straining. It tells a story of how the sales have been made overtime and discover particular trends between the data being viewed. The colour of the diagrams besides the dendrogram chart remain fairly consistent which puts less strain and makes it easy to follow the data presented between different idioms.

The stacked bar chart is a simple yet effective idiom to use to showcase the genres and use colour to represent the portion of sales the regions were sold in. The pie chart is easy to interpret and understand but it can be sometimes hard to interpret the exact value it represents without the assistance of labels. Using both of these charts side by side can be effective to compare region sales and the total sales overall for all genres between all regions. The issue with this is it can be sometimes hard to interpret the exact value the bar chart represents when it is further away from the sales reference at the bottom. The pie chart can also be distracting if you are trying to interpret the stacked bar chart more accurately.

The area chart provides a good visual indication on how many games were sold per year along with the colour indication of the regions of where the sales were made. In terms of accuracy it can be hard to tell the exact amount of sales that were made, especially if trying to decipher the total sales made within certain region that doesn't start from the bottom of the chart.

The packed bubble chart and treemap chart are both effective visualizations to showcase the ranking of platforms and publishers using size and brightness of colour. It's easy to visually spot the largest object and allows easier time to comprehend the games quantity difference between different platforms and publishers.

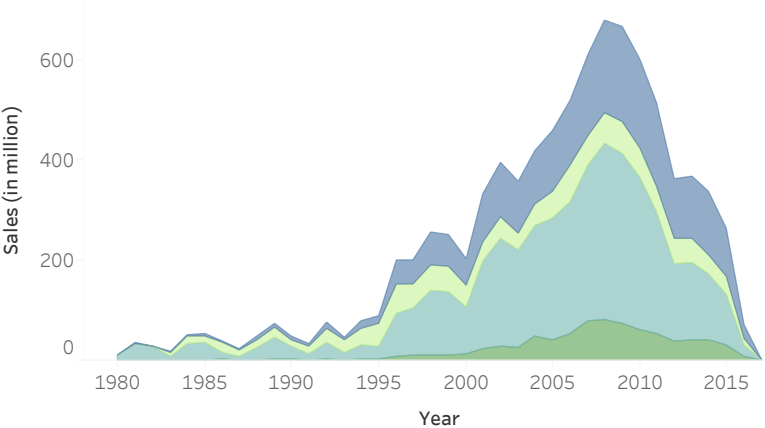
The dendrogram chart is able to effectively communicate in an intuitive way of the total sales made for top 100 games split between the portion of the platforms it represents. It is able to communicate the story effectively to the user based on the chart sizes represented for each platform. The colour variations for this chart could also be downside to this visualizations. A lot of colours can sometimes be very hard to interpret and remember which colour represents a particular platform without double checking. The label on the right of the chart is the main assistance for this chart and colour at this point could mainly be for aesthetics.

The use of whitespace in the overall visualization makes it feel less claustrophobic for the user, but it can sometimes feel cluttered and unordered. Some sort of grids or boxes around the graphs can make it look cleaner but would be difficult to do with certain visualizations like the dendrogram chart without overlapping the other graphs.

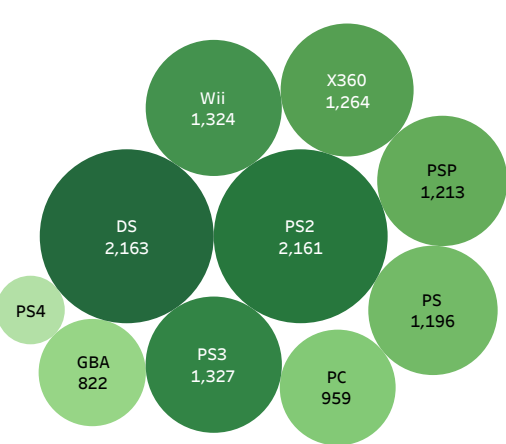
## References

[1] Video Game Sales Data Source: <https://www.kaggle.com/datasets/thedevastator/global-video-game-sales/data>

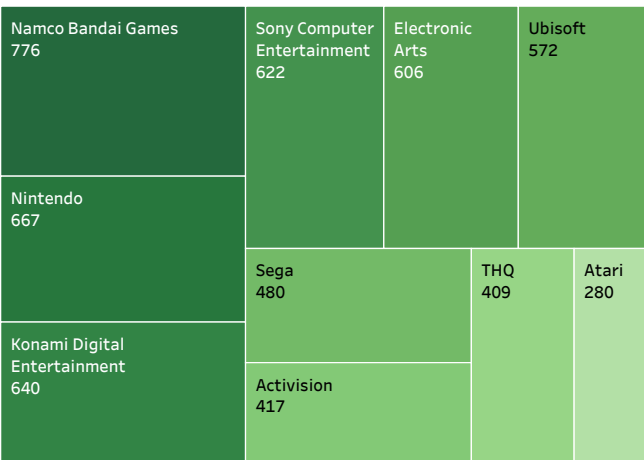
Video Game Sales by Year



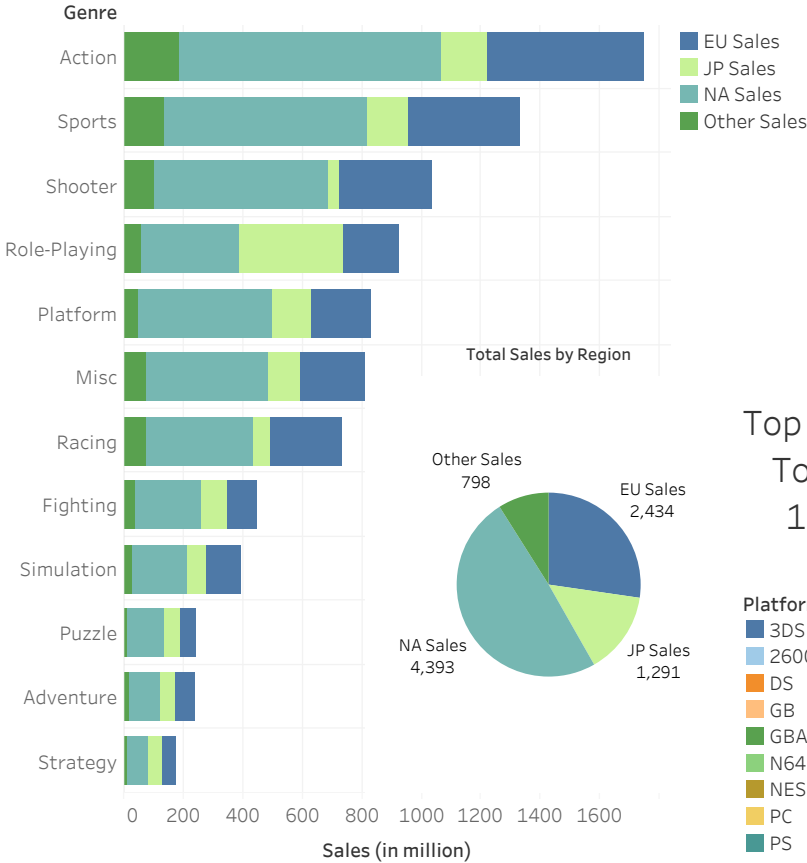
Top 10 Platform by Games Quantity



Top 10 Publishers by Games Quantity



Video Game Sales by Genre



Top 100 Games  
Total Sales:  
1,462 mil.

