



DATA VISUALISATION

BRUNEL UNIVERSITY LONDON

DEPARTMENT OF MATHEMATICS

Coursework CS5803/MA5673

Author:

Jean Charles Kouame' (CID: 2140558)

Date: April 18, 2022

1 Introduction

1.1 Data selection

The data set is named **Video Games Sales**, and contains the information about videogames sales between **1979 – 2016**. It is composed by **16,598** records and **11** features. It was generated by a scrape of vgchartz.com. and it was collected from Kaggle website. The data should also come with sufficient metadata and documentation A ‘data dictionary’ (i.e. a table listing variable names, descriptions and domains) of all variables relevant to your analysis should be included in your report.

Data dictionary:

- **Rank** - Ranking of overall sales
- **Name** - Title of the game [e.g. Super Mario Kart, Call of Duty]
- **Platform** - Platform of the games release [e.g. PC, PS4, GameBoy]
- **Year** - Year of the game’s release [1979-2016]
- **Genre** - Genre of the game [e.g. Action, Sport, Shooter]
- **Publisher** - Publisher of the game [e.g. Electronic Arts, ActiVision]
- **NA_Sales** - Sales in North America [in millions]
- **EU_Sales** - Sales in Europe [in millions]
- **JP_Sales** - Sales in Japan [in millions]
- **Other_Sales** - Sales in the rest of the world [in millions]
- **Global_Sales** - Total worldwide sales. [in millions]

1.2 User type/Persona specification

We have been asked by the **product manager** of a start-up active in computer games development to produce a dashboard that would give him an overview of the video games industry history.

The user is analytically trained enough to be able of handling medium-complex type of dashboards.

1.3 Question formulation

A product manager is the person who identifies the customer need and the larger business objectives that a product or feature will fulfill, articulates what success looks like for a product, and rallies a team to turn that vision into a reality^[4].

Therefore the user is interested in the following questions:

- **Question 1:** Who has been the **most successful company** in the video games industry?
- **Question 2:** What are the **most popular genres**?

Answering this questions will give the manager the directions that he needs to decide what will be the best way to employ the specific competences of his company.

1.4 User requirements specification

The main purpose of the dashboard emphasize the analysis. This dashboard should be able to give insights on the video games industry and effectively tracking the performances of different aspects of the industry, to deliver analytic recommendations for future strategies.

To accomplish this task a **tactical dashboard** is often utilized. A tactical dashboard is mainly used to track trends in relation to company goals and initiatives.

1.4.1 Functional requirements

Question 1: Who has been the most successful company in the video games industry?

- R1.1: Filtering: to ask Q1 is important to compare different views of the same criteria on different data: a filter by Platform is necessary
- R1.2: Comprehensive view: a time span view of the data will serve the user to better understand the evolution of different players in the industry
- R1.3: Highlighting data: by highlighting different part of the table the user will be able to compare how different companies perfomed with respect each other

Question 2: What are the most popular genres?

- R2.1: Filtering: to answer Q2 we need a second filter by Genre to
- R2.2: Comparative view: is useful for understanding market segmentation and market opportunities
- R2.3: Color coding: to be able to distinguish between different classes of the data a color encoding is necessary

1.4.2 Non-functional requirements

Non-functional requirements relate to relevant aspects of the user's background and context:

- R3.1: High level of detail: the user is a technical figure that is able to handle a relevant amount of information

- R3.2: Standard desktop resolution: the user will have at disposition a company PC where visualizing the dashboard

2 Design

2.1 Data exploration

The first thing to do is to obtain the structure of the data frame:

```
1 'data.frame': 16598 obs. of 11 variables:
2   Rank      : int  1 2 3 4 5 6 7 8 9 10 ...
3   Name       : chr "Wii Sports" "Super Mario" "Mario Kart Wii" ...
4   Platform   : chr "Wii" "NES" "Wii" "Wii" ...
5   Year       : chr "2006" "1985" "2008" "2009" ...
6   Genre      : chr "Sports" "Platform" "Racing" "Sports" ...
7   Publisher  : chr "Nintendo" "Nintendo" "Nintendo" "Nintendo" ...
8   NA-Sales   : num 41.5 29.1 15.8 15.8 11.3 ...
9   EU-Sales   : num 29.02 3.58 12.88 11.01 8.89 ...
10  JP-Sales   : num 3.77 6.81 3.79 3.28 10.22 ...
11  Other-Sales: num 8.46 0.77 3.31 2.96 1 0.58 2.9 2.85 2.26 ...
12  Global-Sales: num 82.7 40.2 35.8 33 31.4 ...
```

The data frame is composed by 16598 observations (rows) and 11 features (columns). We can start plotting the distribution of the numeric variable Global Sales:

```
1 hist((data$Global_Sales), breaks=5000, xlim = c(0,2))
```

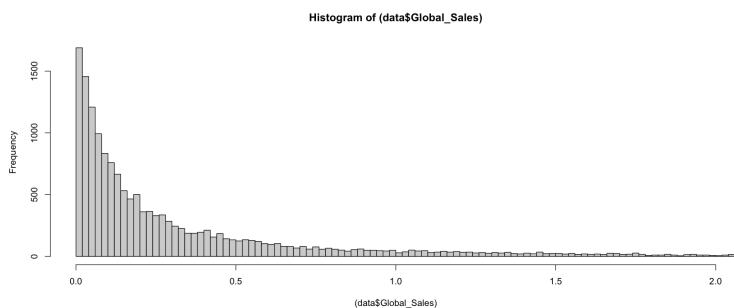


Figure 1: Global Sales distribution

We can extract some other interesting statistics of the data frame:

```
In [69]: table(data$Genre)
```

Action	Adventure	Fighting	Misc	Platform	Puzzle
3316	1286	848	1739	886	582
Racing	Role-Playing	Shooter	Simulation	Sports	Strategy
1249	1488	1310	867	2346	681


```
In [70]: table(data$Platform)
```

2600	3D0	3DS	DC	DS	GB	GBA	GC	GEN	GG	N64	NES	NG	PC	PCFX	PS
133	3	509	52	2163	98	822	556	27	1	319	98	12	960	1	1196
PS2	PS3	PS4	PSP	PSV	SAT	SCD	SNES	TG16	WS	Wii	WiiU	X360	XB	XOne	
2161	1329	336	1213	413	173	6	239	2	6	1325	143	1265	824	213	


```
In [72]: sum(length(unique(data$Publisher)))
```

579


```
In [66]: sum(length(unique(data$Name)))
```

11493

We can find what are the frequency distribution of the years and genre variables in the data set:

```
1 # Years frequency
2 freq_year <- data.frame(cbind(Frequency = table(data$Year), Percent =
   prop.table(table(data$Year)) * 100))
3 freq_year <- freq_year[order(freq_year$Frequency, decreasing=TRUE), ]
4
5 # Gengre frequency
6 freq_genre <- data.frame(cbind(Frequency = table(data$Genre), Percent =
   prop.table(table(data$Genre)) * 100))
7 freq_genre <- freq_genre[order(freq_genre$Frequency, decreasing = T), ]
```

And we can plot the frequency histograms:

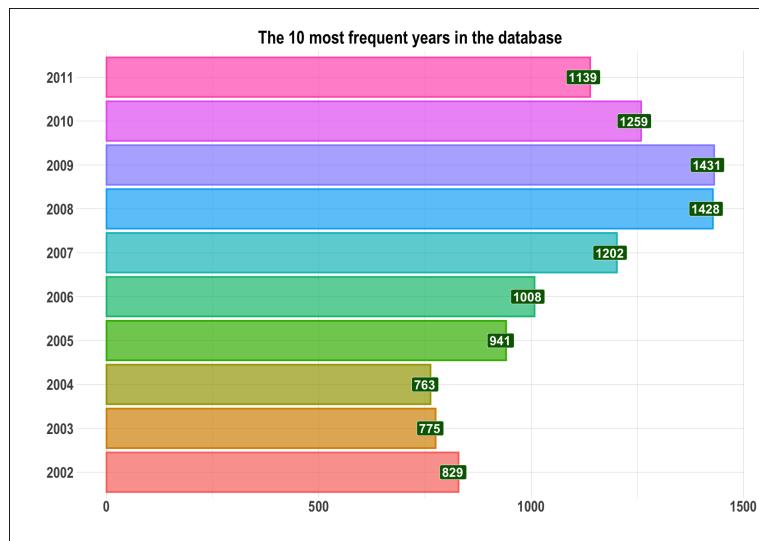


Figure 2: 10 Most Frequent years in the data set

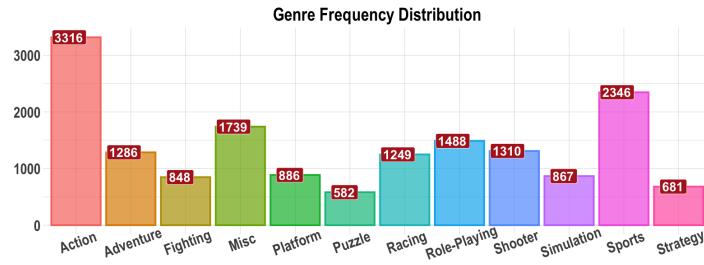


Figure 3: Genre frequency distribution

2.2 Paper landscape: unconstrained representation

The paper landscape is a powerful tool that help us to evaluate design alternatives and facilitate the discussion between stakeholders.

The solution aims to address the user requirements that we illustrated in paragraph 1.4. In this phase is only taken in consideration the desired output, without taking into consideration any specific feature of the implementation tool **Tableau** or **Google Design Studio**. It is assumed a standard screen resolution of 1080p, as the final user will work on a personal computer.



Figure 4: Prototype design

The solution contains 5 views, with different types of charts. Following we can find how each table will respond to a specific user requirement:

- Table 1, Area chart, requirement 1.2: this view can show how different Platform have performed against each other in term of sales, therefore giving a comprehensive view
- Table 2, Small multiple chart, requirement 2.2: this chart gives a comparative view. Different video games genres are compared on the same metric
- Table 3, Heat map, requirement 2.2: also this chart gives a comparative view. The area of the map highlight which genre has sold more copies.
- Table 4, Ranking table, requirement 1.3: this view is helpful to highlight particular important information coming from the data
- Table 5, Histogram, requirement 1 and 2: this view answer at more than 1 requirement as it is able to compare, highlight and sum up different information

2.3 Design principles

When you design your views, you should pay close attention to the basic principles of good data visualisation as taught in the lectures relating to visual encoding, representation and presentation. This includes (but is not limited to) Bertin's theory and Mackinlays guidance on choosing the correct visual encodings and also Few's guidance on choosing the correct representation for the relationship (e.g. comparison, part-whole, correlation etc.) implied by the question.

2.3.1 Visual encoding principles implemented:

- **Small multiple** (on table 2): this method is encouraged by Edward Tufte as it allows a quick visual comparison of multiple series. A chart with many series shown on a single pair of axes that can often be easier to read when displayed as several separate pairs of axes placed next to each other[5].
- **Color hue** (on all tables): Hue is the visual perceptual property corresponding in humans to the categories called red, green, blue, and others[1]. To respond at this principle the variables "Platforms" will be encoded in 5 main groups and at each group will be assigned a correspondent color.
- **Length/Area comparison** (on table 3 and 5): this principle will be implemented with the use of a histogram and a heat map.

2.3.2 Interaction principles

- **Feedback** (on filters and clicking on charts): a tight coupling between action and reaction is part of Tableau functionalities. Updates should be dynamic; near real-time (< 25 ms). Table allows a quick update between table.
- **Brushing and linking** (on all charts): Specifically, linking refers to a change of parameters (for example a data filter) in one data representation being reflected in other connected data representations. Brushing refers to highlighting, for example selected data, in one view, in other connected data representations [2]. In the Tableau implementation all chart will be connected to each other and will be possible to click on certain data of a chart and update instantly the other ones.

2.3.3 User experience (UX) principles

- **Consistency:** Consistency throughout an interface will ensure that the least amount of brain effort is required when using your product [3]. In the implementation we will keep a consistent color code between charts and a similar approach to graphical signs.

3 Implementation

The following instructions will present the implementation process.

A number of new variable has to be defined before starting the implementation.

3.1 View 1: Number of Games grouped by Platforms in temporal order

This table represent the sum of number of games titles (Names) grouped by Platforms in temporal order. It means that from this table we can observe how different platforms have performed against each other along the years in terms of number of games released.

3 IMPLEMENTATION

3.2 View 2: Life span of different Platform

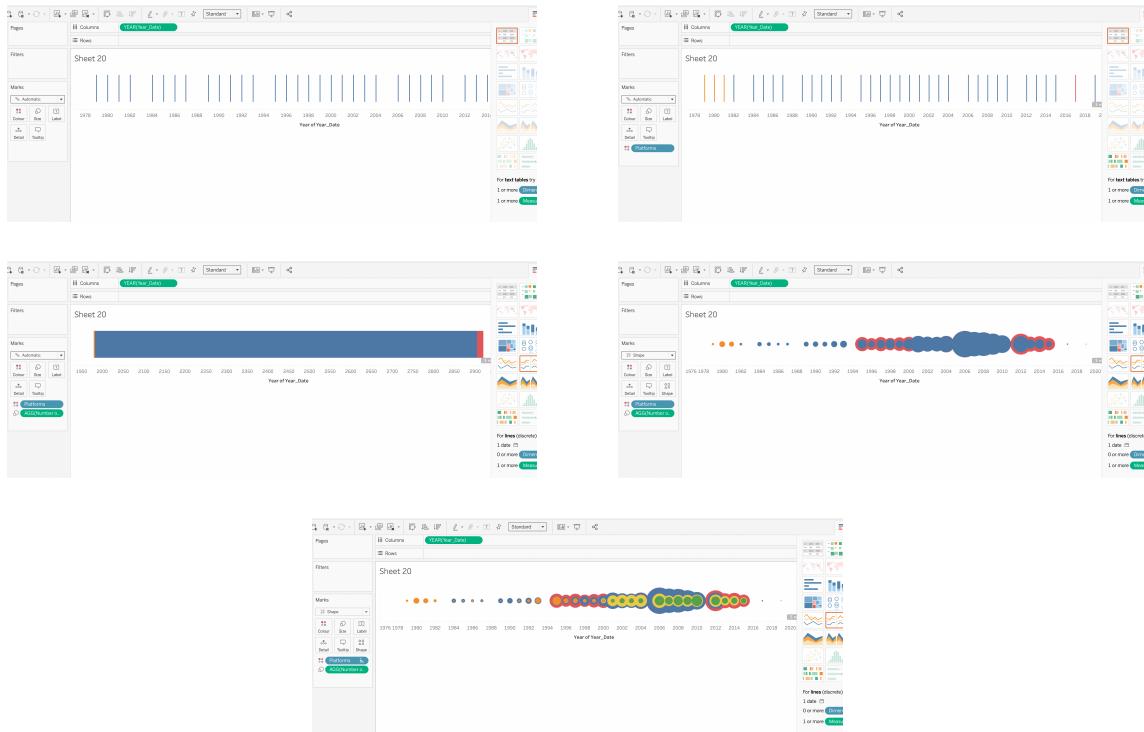


Figure 5: implementation of view 1

1. Drag "Year Date" variable to Columns and change type to Continuous
2. Drag "Platforms" in Marks under "Colour"
3. Drag Number of games variable in marks under Size
4. Select Shape from marks drop down menu and chose full circle
5. Select Descending order from toolbar menu

3.2 View 2: Life span of different Platform

The following is the most complex view: it illustrates the life span of each platform with a Gantt Bar type of chart.

3.2 View 2: Life span of different Platform

3 IMPLEMENTATION



Figure 6: Implementation of view 2

1. Drag "Year" variable to Columns and change Measure to Minimum
2. Drag "Min Year" variable to Columns and modify expression subtracting 0.25
3. Drag Platforms to Rows
4. By selecting in columns the Min Year tab drag in Marks: Platforms in colour, Year Range in Size, Min Year and Max Year in Detail, Year Range in Label
5. Change Marks type in Gant Chart 6. By selecting in columns the second Min Year tab drag in Marks: Platforms in colour and in Size, then drag Min Year in label
7. Change Marks type in Shape and select any shape
8. Right click on both axis and adjust range between 1979-2018
9. Right click on the axis and select Dual Axis
10. Set Desending order

3 IMPLEMENTATION.3 View 3: Small Multiple: Global Sales divided by Genres

3.3 View 3: Small Multiple: Global Sales divided by Genres

This view illustrates how different Platforms encoded by colour, have performed in term of Global Sales in each Genre and each Year.

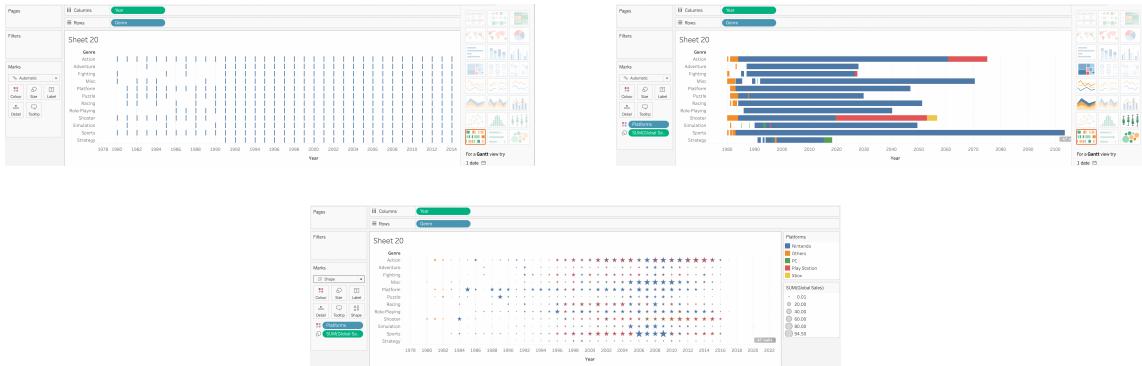
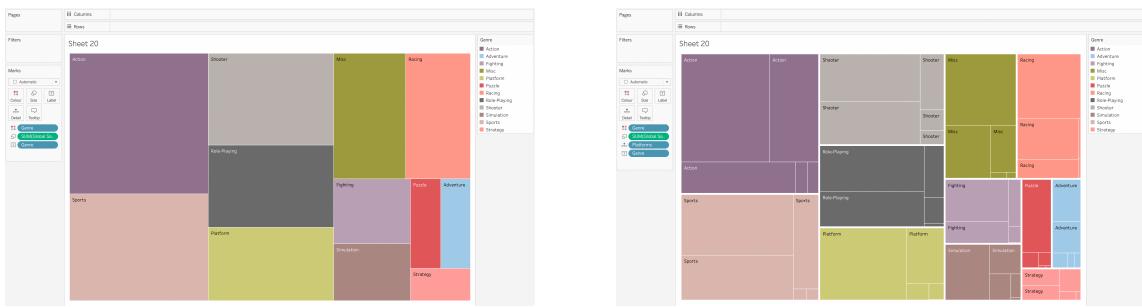


Figure 7: implementation of view 3

1. Drag "Year Date" variable to Columns and change type to Continuous
2. Drag "Genre" in Columns
3. Drag Platforms variable in Marks under Colour
4. Drag Global Sales in Size under Marks
5. Select Shape type in the Marks drop down bar
6. Adjust the size of the elements

3.4 View 4: Genres heatmap with detail by Platforms

This view shows the Global Sales size as the size of each rectangle, and code each genre by colour. Also it gives a detail on how different platforms have performed in each Genre.



1. Drag Genre in Colour under Marks
2. Drag "Global Sales" in Marks under "Size"
3. Drag "Platforms" in Marks under "Detail"

3.5 View 5: Platforms rank by Global Sales

This view shows the Platforms rank by Global Sales.



1. Drag "Platform" under Colour and Text in Marks
2. Drag "Global sales" under text in Marks
3. Drag "Platforms" in Filters and filter the top 5 values by Global Sales

3.6 Dashboard

The dashboard is the collection of the different views.

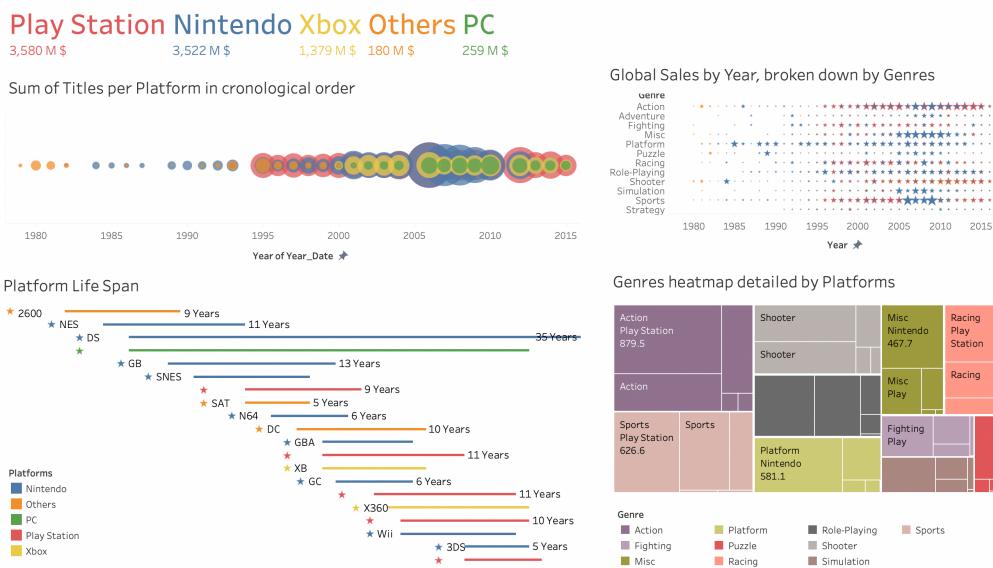


Figure 8: Dashboard

1. Positioning the views:

when adding each view at the dashboard it is important to select Floating from the Object box. In this way it will be possible to resize and position each element manually. A 4x4 grid has been chosen to display the elements.

2. Transparent background:

by right-clicking on each object and selecting Format, it has been possible to select a transparent color to the Sheet Shading option.

3.7 Google Data Studio Implementation

The comparative implementation has been conducted in **Google Data Studio**: the tool is part of Google free products and has an intuitive and responsive interface.

Also in this case have been implemented 4 views: the main goal is to visualise the market segmentation and the magnitude in terms of Global Sales.

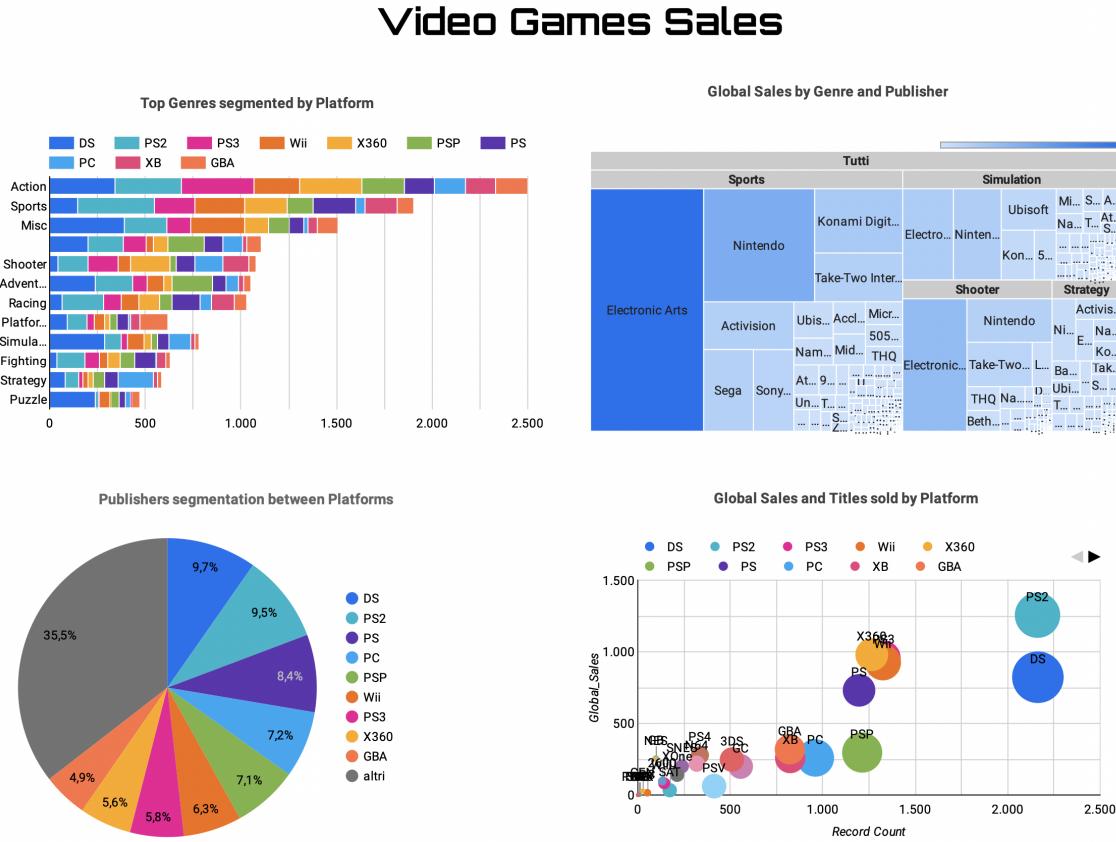


Figure 9: Alternative implementation: Google Data Studio

The first view, **Top Genres segmented by Platform** shows the ranking of the most sold genres in the video game industry. Each bar is segmented by platform.

The second view **Global Sales by Genre and Publisher** illustrates how the genres are segmented by publisher. The most active publisher is Electronic Arts, followed by Nintendo and Activision.

The third view **Publisher segmentation between Platform** illustrates how publishers are segmented by platform. All the platforms have very similar percentages, at exception of the category "Other", meaning that the publishers are equally developing games for all kind of platform.

The forth view **Global Sales and Titles by Platform** shows how the number of games and the total revenue are linearly correlated. In fact on the platform with more titles (Playstation 2 and Nintend DS) correspond also a higher value of total sales, while for the platforms with less title correspond a lower value of revenues. One key issue with respect to Tableau, was that has not been possible to group the

Platforms under more general categories.

4 Walkthrough

4.1 Activating the Dashboard

The main feature of the dashboard is the **Highlighting functionality**. By **clicking on the name of the platform** we select the platform from view 5, and we highlight all the content relative to that specific Platform group.



Figure 10: User Action to activate the dashboard view

Once this view is activated the user will be able to have an overview on different aspect of the video game industry, filtered by the highlighted platform. The most interesting aspect is that the highlighted parts are in context with respect the rest of the industry.

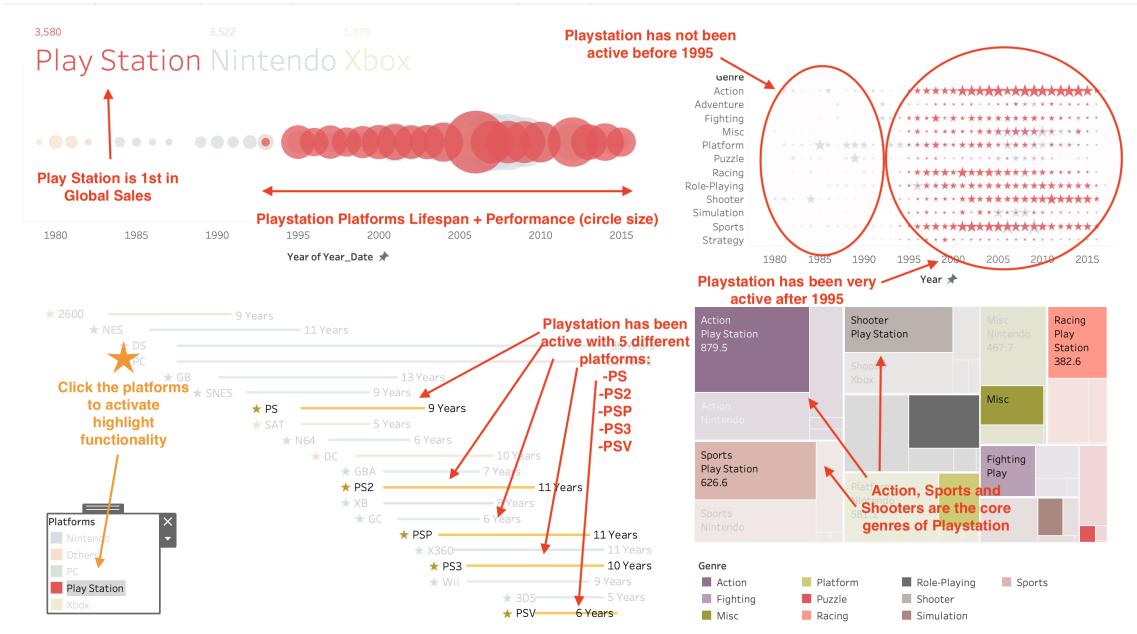


Figure 11: Walkthrough

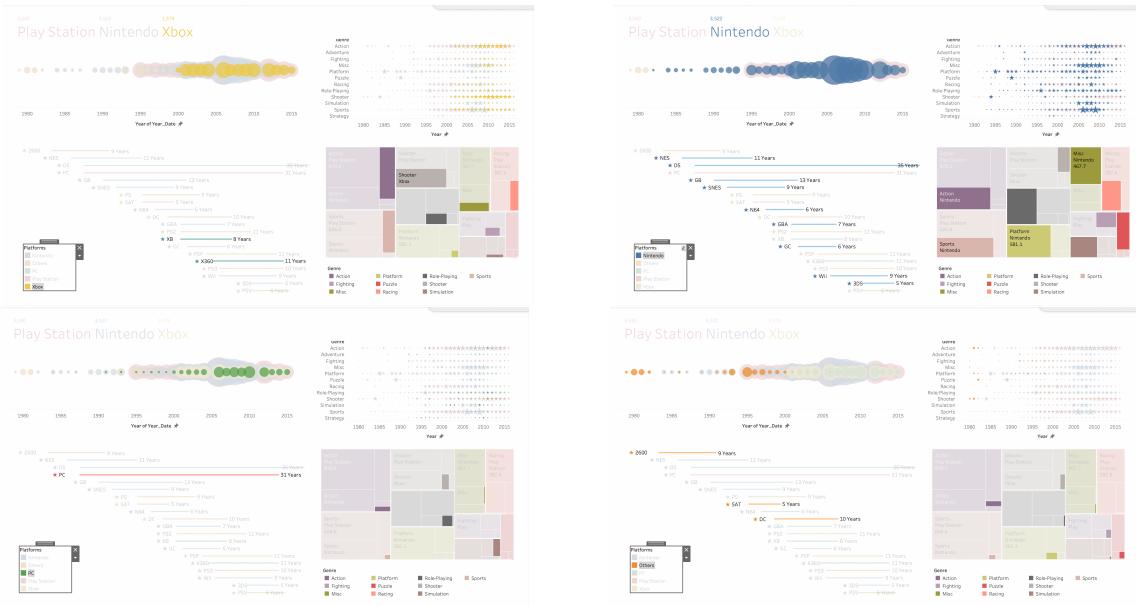


Figure 12: a, b, c, d. Different views depending on Platforms

4.2 Question 1:

Who has been the most successful company in the video games industry?

4.2.1 View 5

To answer this question we have the view in fig.13 that shows how Playstation has been the industry leader, with a total of 3580 Million \$ in Global Sales. Playstation Sales are very close to the one of Nintendo, with 3522 Million \$. XBox is at the third place, but very far from Playstation and Nintendo.



Figure 13: Ranking: View 5

By analyzing the dashboard we can extract many different information as showed in the following walkthrough:

4.2.2 View 1

If click on the Playstation name, we highlight all red colored content in the dashboard, as in fig.10, and by starting from view 1, we can see that Playstation has

been active since 1995, and immediately became one of the biggest players in the industry, as the size of the red bubble suggest with respect the other companies.

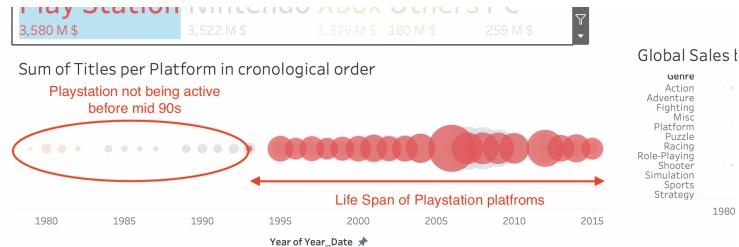


Figure 14: View 1

4.2.3 View 2

Also by looking at the Gantt-chart, Playstation has been active with 5 different platform. Only Nintendo (fig.12b) had an bigger number of **active platforms** (9), while Xbox had only 2 platforms with a life span larger than 5 years.

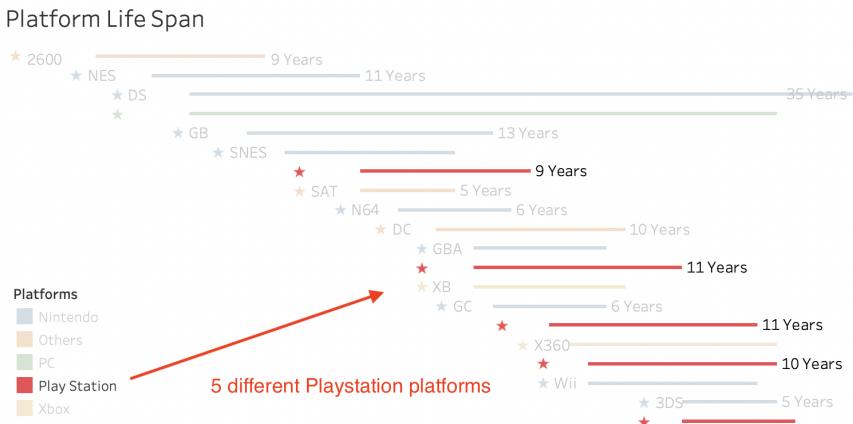
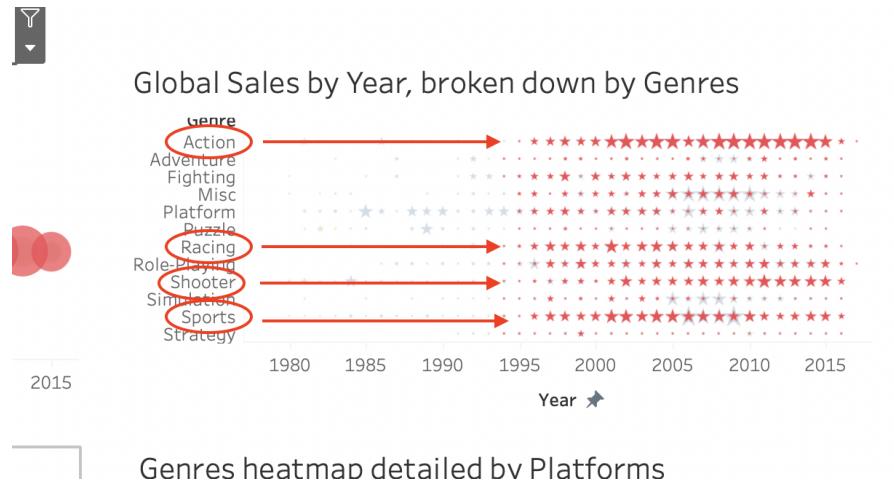


Figure 15: View 2

4.2.4 View 3

The fundamental chart to understand who is the industry leader is the view number 2 (top-right corner of the dashboard). Here we can see how Playstation has been very active on a large variety of video games genres, being particularly strong in the **Action** category. This explain that even if Playstation had less active platforms with respect Nintendo, the sales volume have been constantly higher.

**Figure 16: View 3**

It is also interesting to see how **before 1995** the video games industry was not as much diffused as much after 1995. In that era only the **Atari platform** and the **PC** were active in the market.

It is important to notice that the PC has been one of the most resistant platforms, having a life-span of **31 years** of activity.

4.3 Question 2:

What are the most popular genres?

4.3.1 View 4

To answer this question we can analyze the heat map in view 5. The chart box sizes are proportioned with the **Global Sales** and in the top-left corner we can find the most successful genres, while at the right-bottom we find the less popular games.

It is immediate to see that the most popular category has been **Action**, followed by **Sports** and **Shooters**. It is also possible to see a detail of how different Platforms perform in each genre: **Playstation is the Leader in all the top three categories**, while Nintendo is specialized in the category **Miscellaneous**, and XBox has a good position in the **Shooter** games.

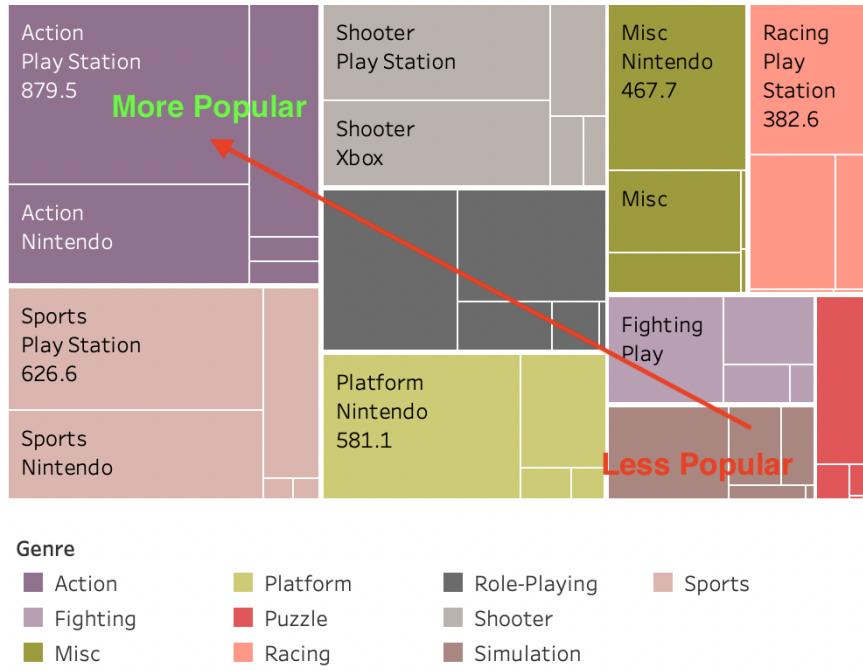


Figure 17: Heat map direction

4.4 Principles implementation:

4.4.1 Color encoding: Platforms vs Genres

The dashboard chart are full of colors. Because the number of the many different codes it was fundamental to tune two very different color palette, to avoid overlapping in colors that might cause confusion.

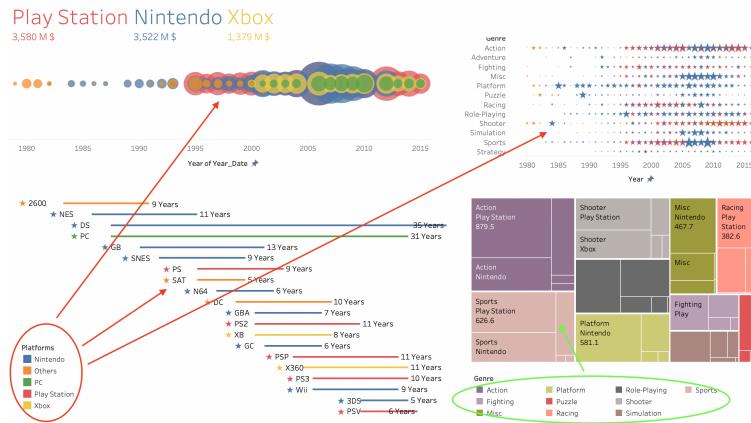


Figure 18: Color coding

4.4.2 Pop-out/highlighting

The pop out highlighting is fundamental in the implementation of this dashboard, as it was important to look at the data in the context of the global measurements.

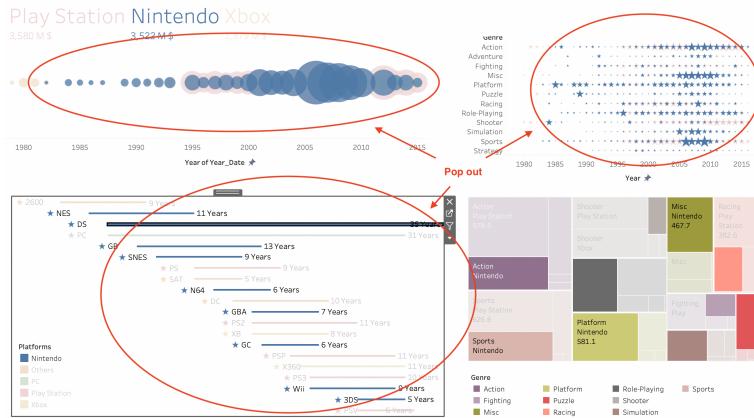


Figure 19: Pop-Out/Highlighting

5 Reflective Discussion

The implementation in Tableau has been successful, as the dashboard is able to answer both the question we formulated in the design phase.

The result is an interactive platform, and all the views talk to each other. The information are visualized with respect the context, and this makes a strong point of this implantation.

The weakest point of Tableau has been the customization of the chart point shapes and the rigid choice of colors palette. Overall Tableau is a good data visualization tool able to answer the needs of the majority of professionals.

Another notable weak point of Tableau is the limited choice of pre-installed charts. Many types of charts are missing, such as Lollipop-charts or the Half-circle charts.

With respect to Google Data Studio, Tableau is a much more flexible solution. The possibility of creating separate Worksheets and Dashboards, increases the level of usability of the platform. Also Tableau has a greater number of features with respect Google Data Studio

References

- [1] Jacque Bertin. **Semiology of Graphics**. University of Wisconsin Press; English Edition. 2001.
- [2] Helmut Doleisch. **Visual analysis of complex simulation data using multiple heterogenous views**. IEEE. 2004.
- [3] Jakob Nielsen. **Usability Engineering - Interactive Technologies**. IEEE. 1994.
- [4] Atlassian SHERIF MANSOUR. **Product Manager: The role and best practices**. <https://www.atlassian.com/agile/product-management/product-manager>. 2020.

- [5] Edward tufte. **The Visual Display of Quantitative Information**. Graphics Pr; 2° edition ASIN 0961392142. 2001.