DATA CLEANING PROJECT REPORT

(Project Semester Jan-May 2025)

Video Game Sales Analysis

Submitted by

Name: Divyanshu Singh

Registration No: 12415084

Programme: M.Tech. (Data Science and Analytics)

Section: K24DS

Course Code: INT553

Under the Guidance of

Savleen Kaur

(Faculty)

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

DECLARATION

I, **Divyanshu Singh** student of M.Tech. (Data Science and Analytics) Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12 April 2025 Signature:

Registration No: **12415084** Name of the Student: **Divyanshu**

ACKNOWLEDGEMENT

I would like to take this opportunity to show my sincere gratitude towards, Savleen Kaur, our subject teacher for course code INT553 as well as our university: Lovely Professional University whose constant help and guidance allowed me to complete this project in due time. I would also like to thank my classmates whose support was crucial in helping me finish this project.

List of Contents

S. No.	Title	Page
1	Declaration by Student	2
2	Acknowledgement	3
4	List Of Contents	4
5	Chapter-1 Introduction	5
6	Chapter-2 OVERVIEW AND SOURCE OF DATASET	7
7	Chapter-3 Data Preprocessing	9
8	Chapter-4 Analysis On Dataset	16
9	Chapter-5 DashBoard	43
10	Chapter-6 Code	46
11	Chapter-7 Conclusion and Future Scope	74
12	Chapter -8 Final Chapter- References	76

CHAPTER 1
INTRODUCTION

The video game industry has undergone a transformative evolution since the late 20th century, emerging as one of the most influential and profitable sectors in global entertainment. With annual revenues surpassing those of the film and music industries combined, the industry encompasses a wide range of platforms, genres, and publishers catering to diverse audiences across the globe. Understanding market behavior in this domain is critical for game developers, publishers, marketers, analysts, and researchers alike.

Video games are no longer confined to arcade systems or home consoles—they span across platforms such as consoles (e.g., PlayStation, Xbox, Nintendo Wii), handheld devices (e.g., Nintendo DS, PSP), personal computers, and increasingly mobile devices. This diversification has allowed the industry to reach broader demographics, from casual players using smartphones to hardcore gamers investing in high-end consoles and PCs.

Within the dataset explored in this project, we observe a rich collection of over 16,000 games released between 1980 and 2020, each tagged with metadata including the game's name, release year, genre, platform, publisher, and regional/global sales. These games span a variety of genres such as Action, Sports, Role-Playing, Shooter, Racing, and Simulation, reflecting evolving user interests and technological capabilities over time.

The platforms in the dataset reflect nearly all major gaming hardware ecosystems, including legacy systems like NES (Nintendo Entertainment System) and Atari 2600, to modern platforms like PlayStation 4, Xbox 360, and Nintendo Switch. Historically, platforms like PlayStation 2, Nintendo DS, and Wii have hosted a large volume of successful titles, making them critical in shaping industry sales dynamics.

On the publishing side, major players such as Nintendo, Electronic Arts (EA), Activision, Sony Computer Entertainment, and Ubisoft dominate global sales. These publishers are responsible for producing many iconic franchises—Mario, FIFA, Call of Duty, The Legend of Zelda, and Assassin's Creed—that have become household names and have generated billions in cumulative sales. The dataset allows us to track not just the total volume of games published, but also their geographical success, identifying which companies perform best in North America, Europe, Japan, and other regions.

To gain a structured understanding of this dynamic ecosystem, this project leverages R programming, the Shiny web framework, and Flexdashboard to create an interactive dashboard. This dashboard empowers users to explore and analyze the sales patterns of video games from multiple perspectives, including temporal trends, genre and platform popularity, publisher performance, and regional dominance. By leveraging data visualization and exploratory data analysis techniques, the dashboard enables users to explore trends in sales across different regions (North America, Europe, Japan, Others), platforms (e.g., PS2, X360, Wii), and genres (e.g., Action, Sports, Role-Playing).

The core objectives of this project are as follows:

- To visualize trends in the number of games released and sales performance over time.
- To assess the popularity of different genres and platforms in specific markets.
- To identify the top publishers globally and regionally.
- To explore correlations between regional and global sales for predictive insights.
- To provide a user-friendly, filterable interface that enables customizable analysis based on year, platform, and genre.

By integrating statistical summaries, visual storytelling, and interactivity, this analysis serves not only as a powerful business intelligence tool but also as a knowledge resource for academics, students, and enthusiasts interested in the video game economy.

CHAPTER 2

OVERVIEW AND SOURCE OF DATASET

The dataset used for this analysis, titled vgsales.csv, is a comprehensive compilation of global video game sales data collected across multiple gaming platforms and markets. It contains a total of 16,598 records, with each row representing a unique video game entry. The dataset includes key attributes essential for evaluating market performance, genre distribution, platform popularity, and regional consumer behavior.

Source : https://data.world/julienf/video-games-global-sales-in-volume-1983-2017

It reflects historical sales figures that span several decades, offering valuable insight into long-term trends in the gaming industry.

1. **Feature of the Dataset**
   - **Rank: Overall sales ranking of the game.**
   - **Name: Title of the video game.**
   - **Platform: The platform or console on which the game was released (e.g., Wii, PS4, X360, PC).**
   - **Year: Year of the game's release (some entries have missing values).**
   - **Genre: Genre classification (e.g., Sports, Action, Role-Playing).**
   - **Publisher: Name of the game publisher (some entries have missing values).**
   - **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 other regions (in millions).**
   - **Global_Sales: Total global sales across all regions (in millions).**

Out of the 16,598 original entries, a small number of records contained missing values in critical columns such as Year (271 missing) and Publisher (58 missing). These were removed during the preprocessing phase, resulting in a clean and analyzable dataset with minimal noise. Duplicate entries were also checked to ensure uniqueness and data quality.

Some of the most successful titles based on global sales in the dataset include:

- *Wii Sports* (Wii) – 82.74 million copies
- *Super Mario Bros.* (NES) – 40.24 million copies
- *Mario Kart Wii* (Wii) – 35.82 million copies
- *Wii Sports Resort* (Wii) – 33.00 million copies
- *Pokémon Red/Blue* (GB) – 31.37 million copies

These figures highlight the dominance of Nintendo platforms in global sales history.

Potential Applications:

- Identifying high-performing game genres and platforms.
- Analyzing publisher performance across regions.
- Evaluating regional market behaviors.
- Building predictive models for game success.
- Supporting marketing strategies and game development planning.

While the dataset provides a rich overview of video game sales trends, it is important to acknowledge its limitations:

- Some data points are estimates and may not match official publisher records.
- Sales figures reflect physical copies only and exclude digital downloads, in-app purchases, and subscription models, which have become dominant post-2016.
- The dataset does not account for re-releases, remastered editions, or bundle deals in a disaggregated way.

Despite these limitations, the dataset remains highly valuable for identifying macro-level trends, understanding publisher performance, comparing platform success, and performing educational data analysis.

CHAPTER 3

DATA PREPROCESSING

Prior to conducting any meaningful analysis, it is essential to ensure the dataset is clean, consistent, and suitable for visualization and modeling. The vgsales.csv dataset underwent a structured preprocessing pipeline using R and packages from the tidyverse ecosystem to address common data quality issues such as missing values, duplicate entries, inconsistent formats, and incorrect data types.

1. **Importing Required Libraries :**

```
# Import Packages and Libraries
install.packages("corrplot")
install.packages("tidyverse")
install.packages("dplyr")
install.packages("ggplot2")
install.packages("DT")
install.packages("rmarkdown")
install.packages("shiny")
install.packages("shinydashboard")
install.packages("flexdashboard")
install.packages("tidyversa")
library(flexdashboard)
library(ggplot2)
library(tidyverse)
library(dplyr)
library(DT)
```

2. **Reading Dataset :**

df <-
read.csv("C:/Users/d4i7v/OneDrive/Documents/Mtech/sem2/INT553DATACLEANINGR/completeRproject/vgsales.csv")
datatable(df)
View(df)
Data Overview
colnames(df)
as.list(colnames(df))
df$Name
df %>% select(Name, Platform, Year, Genre)
datatable(df %>% select(Name, Platform, Year, Genre, Publisher))

```
## Load Dataset ##

df <- read.csv("C:/Users/d4i7v/OneDrive/Documents/Mtech/sem2/INT553DATACLEANINGR/completeRproject/vgsales.csv")
datatable(df)
View(df)
### Data Overview ##
colnames(df)
as.list(colnames(df))

df$Name
df %>% select(Name, Platform, Year, Genre)
datatable(df %>% select(Name, Platform, Year, Genre, Publisher))
```

3. **Dataset Inspection :**
The dataset was first inspected using the glimpse(), summary(), and str() functions to understand the structure, data types, and the presence of anomalies or unexpected values.

head(df)
head(df,10)
tail(df)
sample_n(df, 15)

str(df)

```
> str(df)
Classes ‘tbl_df’, ‘tbl’ and 'data.frame':       16291 obs. of  11 variables:
 $ Rank        : num  1 2 3 4 5 6 7 8 9 10 ...
 $ Name        : chr  "Wii Sports" "Super Mario Bros." "Mario Kart Wii" "Wii Sports Resort" ...
 $ Platform    : chr  "Wii" "NES" "Wii" "Wii" ...
 $ Year        : num  2006 1985 2008 2009 1996 ...
 $ Genre       : chr  "Sports" "Platform" "Racing" "Sports" ...
 $ Publisher   : chr  "Nintendo" "Nintendo" "Nintendo" "Nintendo" ...
 $ NA_Sales    : num  41.5 29.1 15.8 15.8 11.3 ...
 $ EU_Sales    : num  29.02 3.58 12.88 11.01 8.89 ...
 $ JP_Sales    : num  3.77 6.81 3.79 3.28 10.22 ...
 $ Other_Sales : num  8.46 0.77 3.31 2.96 1 0.58 2.9 2.85 2.26 0.47 ...
 $ Global_Sales: num  82.7 40.2 35.8 33 31.4 ...
```

summary(df)

```
> summary(df)
      Rank           Name              Platform              Year         Genre             Publisher
 Min.   :    1   Length:16291       Length:16291       Min.   :1980   Length:16291       Length:16291
 1st Qu.: 4132   Class :character   Class :character   1st Qu.:2003   Class :character   Class :character
 Median : 8292   Mode  :character   Mode  :character   Median :2007   Mode  :character   Mode  :character
 Mean   : 8290                                         Mean   :2006
 3rd Qu.:12440                                         3rd Qu.:2010
 Max.   :16600                                         Max.   :2020
    NA_Sales          EU_Sales          JP_Sales         Other_Sales        Global_Sales
 Min.   : 0.0000   Min.   : 0.0000   Min.   : 0.00000   Min.   : 0.00000   Min.   : 0.0100
 1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.: 0.00000   1st Qu.: 0.00000   1st Qu.: 0.0600
 Median : 0.0800   Median : 0.0200   Median : 0.00000   Median : 0.01000   Median : 0.1700
 Mean   : 0.2656   Mean   : 0.1477   Mean   : 0.07883   Mean   : 0.04843   Mean   : 0.5409
 3rd Qu.: 0.2400   3rd Qu.: 0.1100   3rd Qu.: 0.04000   3rd Qu.: 0.04000   3rd Qu.: 0.4800
 Max.   :41.4900   Max.   :29.0200   Max.   :10.22000   Max.   :10.57000   Max.   :82.7400
```

4. Dataset Preprocessing:

The dataset contained missing values in several columns, most notably in the Year and Publisher fields. Since these fields are crucial for time-series and categorical analysis:

#ShowUnique Value

```
sapply(df, n_distinct)
as.list(sapply(df, n_distinct))
```

```
> sapply(df, n_distinct)
      Rank       Name   Platform       Year      Genre  Publisher   NA_Sales   EU_Sales
     16598      11493         31         40         12        579        409        305
  JP_Sales Other_Sales Global_Sales
       244        157        623
```

#remove duplicate

```
duplicate_value <- sum(duplicated(df))
cat("Number of Duplicate Value:", duplicate_value, "\n")
```

```
> duplicate_value <- sum(duplicated(df))
> cat("Number of Duplicate Value:", duplicate_value, "\n")
Number of Duplicate Value: 0
```

 # Check for missing values

```
df <-
read_csv("C:/Users/d4i7v/OneDrive/Documents/Mtech/sem2/INT553DATACLEANINGR/completeRproject/vgsales.csv", na = c("", "NA", "N/A", "n/a", "null", "NULL"))
missing_values <- colSums(is.na(df))
print(missing_values)
```

```
> missing_values <- colSums(is.na(df))
> print(missing_values)
      Rank       Name   Platform       Year      Genre  Publisher   NA_Sales   EU_Sales
         0          0          0        271          0         58          0          0
  JP_Sales Other_Sales Global_Sales
         0          0          0
```

df %>% filter(is.na(Year) | is.na(Publisher))

```
> df %>% filter(is.na(Year) | is.na(Publisher))
# A tibble: 307 x 11
    Rank Name                Platform  Year Genre Publisher NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales
   <dbl> <chr>               <chr>    <dbl> <chr> <chr>        <dbl>    <dbl>    <dbl>       <dbl>        <dbl>
 1   180 Madden NFL 2004     PS2         NA Spor… Electron…     4.26     0.26     0.01        0.71         5.23
 2   378 FIFA Soccer 2004    PS2         NA Spor… Electron…     0.59     2.36     0.04        0.51         3.49
 3   432 LEGO Batman: The Vid… Wii       NA Acti… Warner B…     1.86     1.02     0           0.29         3.17
 4   471 wwe Smackdown vs. Ra… PS2       NA Figh… NA            1.57     1.02     0           0.41         3
 5   608 Space Invaders      2600        NA Shoo… Atari         2.36     0.14     0           0.03         2.53
 6   625 Rock Band           X360        NA Misc  Electron…     1.93     0.34     0           0.21         2.48
 7   650 Frogger's Adventures… GBA       NA Adve… Konami D…     2.15     0.18     0           0.07         2.39
 8   653 LEGO Indiana Jones: … Wii       NA Acti… LucasArts     1.54     0.63     0           0.22         2.39
 9   713 Call of Duty 3      Wii         NA Shoo… Activisi…     1.19     0.84     0           0.23         2.26
10   784 Rock Band           Wii         NA Misc  MTV Games     1.35     0.56     0           0.2          2.11
# i 297 more rows
```

df %>% filter(is.na(Year) & is.na(Publisher)) %>% summarise_all(~sum(!is.na(.)

```
> df %>% filter(is.na(Year) & is.na(Publisher)) %>% summarise_all(~sum(!is.na(.)))
# A tibble: 1 x 11
   Rank  Name Platform  Year Genre Publisher NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales
  <int> <int>    <int> <int> <int>     <int>    <int>    <int>    <int>       <int>        <int>
1    22    22       22     0    22         0       22       22       22          22           22
```

#Drop null values

We have a total of 16,598 records, out of which 307 contain Na values for the Year and Publisher attributes.Since these features are not easily imputed or inferred from the remaining data, and given that the proportion of missing records is relatively small compared to the overall dataset,I have decided to eliminate them rather than attempt an imprecise replacement.

```
df <- df %>% drop_na(Year, Publisher)
colSums(is.na(df))
```

```
> colSums(is.na(df))
      Rank       Name   Platform       Year      Genre  Publisher   NA_Sales   EU_Sales
         0          0          0          0          0          0          0          0
  JP_Sales Other_Sales Global_Sales
         0          0          0
```

#Change DataType

```
df$Year <- as.numeric(df$Year)
str(df)
```

```
> df$Year <- as.numeric(df$Year)
> str(df)
tibble [16,291 x 11] (S3: tbl_df/tbl/data.frame)
 $ Rank        : num [1:16291] 1 2 3 4 5 6 7 8 9 10 ...
 $ Name        : chr [1:16291] "Wii Sports" "Super Mario Bros." "Mario Kart Wii" "Wii Sports Resort" ...
 $ Platform    : chr [1:16291] "Wii" "NES" "Wii" "Wii" ...
 $ Year        : num [1:16291] 2006 1985 2008 2009 1996 ...
 $ Genre       : chr [1:16291] "Sports" "Platform" "Racing" "Sports" ...
 $ Publisher   : chr [1:16291] "Nintendo" "Nintendo" "Nintendo" "Nintendo" ...
 $ NA_Sales    : num [1:16291] 41.5 29.1 15.8 15.8 11.3 ...
 $ EU_Sales    : num [1:16291] 29.02 3.58 12.88 11.01 8.89 ...
 $ JP_Sales    : num [1:16291] 3.77 6.81 3.79 3.28 10.22 ...
 $ Other_Sales : num [1:16291] 8.46 0.77 3.31 2.96 1 0.58 2.9 2.85 2.26 0.47 ...
 $ Global_Sales: num [1:16291] 82.7 40.2 35.8 33 31.4 ...
```

```
cat("Years covered:", min(df$Year), "to", max(df$Year), "\n")
cat("Total unique games:", n_distinct(df$Name), "\n")
cat("Unique Platforms:", n_distinct(df$Platform), "\n")
cat("Unique Genre:", n_distinct(df$Genre), "\n")
min(df$Year, na.rm = TRUE)
max(df$Year, na.rm = TRUE)
n_distinct(df$Year)
unique(df$Year)
unique(df$Genre)
```

```
> cat("Years covered:", min(df$Year), "to", max(df$Year), "\n")
Years covered: 1980 to 2020
> cat("Total unique games:", n_distinct(df$Name), "\n")
Total unique games: 11325
> cat("Unique Platforms:", n_distinct(df$Platform), "\n")
Unique Platforms: 31
> cat("Unique Genre:", n_distinct(df$Genre), "\n")
Unique Genre: 12
> min(df$Year, na.rm = TRUE)
[1] 1980
> max(df$Year, na.rm = TRUE)
[1] 2020
> n_distinct(df$Year)
[1] 39
> unique(df$Year)
 [1] 2006 1985 2008 2009 1996 1989 1984 2005 1999 2007 2010 2013 2004 1990 1988 2002 2001 2011 1998 2015 2012
[22] 2014 1992 1997 1993 1994 1982 2003 1986 2000 1995 2016 1991 1981 1987 1980 1983 2020 2017
> unique(df$Genre)
 [1] "Sports"       "Platform"     "Racing"       "Role-Playing" "Puzzle"       "Misc"         "Shooter"
 [8] "Simulation"   "Action"       "Fighting"     "Adventure"    "Strategy"
```

```
df %>%filter(NA_Sales == max(NA_Sales, na.rm = TRUE)) #top1
df %>%filter(EU_Sales == max(EU_Sales, na.rm = TRUE)) #top1
df %>%filter(JP_Sales == max(JP_Sales, na.rm = TRUE)) #top1
df %>%filter(Global_Sales == max(Global_Sales, na.rm = TRUE)) #top1
```

```
> df %>%filter(NA_Sales == max(NA_Sales, na.rm = TRUE)) #top1
# A tibble: 1 × 11
   Rank Name       Platform  Year Genre  Publisher NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales
  <dbl> <chr>      <chr>    <dbl> <chr>  <chr>        <dbl>    <dbl>    <dbl>       <dbl>        <dbl>
1     1 Wii Sports Wii       2006 Sports Nintendo      41.5     29.0     3.77        8.46         82.7
> df %>%filter(EU_Sales == max(EU_Sales, na.rm = TRUE)) #top1
# A tibble: 1 × 11
   Rank Name       Platform  Year Genre  Publisher NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales
  <dbl> <chr>      <chr>    <dbl> <chr>  <chr>        <dbl>    <dbl>    <dbl>       <dbl>        <dbl>
1     1 Wii Sports Wii       2006 Sports Nintendo      41.5     29.0     3.77        8.46         82.7
> df %>%filter(JP_Sales == max(JP_Sales, na.rm = TRUE)) #top1
# A tibble: 1 × 11
   Rank Name                  Platform  Year Genre Publisher NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales
  <dbl> <chr>                 <chr>    <dbl> <chr> <chr>        <dbl>    <dbl>    <dbl>       <dbl>        <dbl>
1     5 Pokemon Red/Pokemon B… GB        1996 Role… Nintendo      11.3     8.89     10.2           1         31.4
> df %>%filter(Global_Sales == max(Global_Sales, na.rm = TRUE)) #top1
# A tibble: 1 × 11
   Rank Name       Platform  Year Genre  Publisher NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales
  <dbl> <chr>      <chr>    <dbl> <chr>  <chr>        <dbl>    <dbl>    <dbl>       <dbl>        <dbl>
1     1 Wii Sports Wii       2006 Sports Nintendo      41.5     29.0     3.77        8.46         82.7
```

CHAPTER 4

Analysis On Dataset

This section presents an in-depth examination of the dataset, structured into multiple analyses based on specific business questions. Each sub-analysis includes:

- **Introduction**: The rationale and importance of the question being explored.
- **General Description**: Summary of the approach taken.
- **Specific Requirements, Functions, and Formulas**: Code or logic applied.
- **Analysis Results**: Key observations from the data.
- **Visualization**: Graphical representation of the findings.

Each subsection is aimed at extracting targeted insights that contribute to a better understanding of video game market trends and dynamics. This section presents a comprehensive, objective-driven exploratory data analysis (EDA) of the video game sales dataset using the R programming language and multiple visualization libraries. Each visualization corresponds to a distinct analytical goal and is structured with an introduction, general description, formulas/functions used, the plot itself, and a detailed interpretation of the results.

1. Game Released Per Year

Introduction: Understanding the yearly trend in game releases provides valuable insight into the growth and shifts within the video game industry. An increasing number of game releases could indicate market expansion, while a decline might suggest industry consolidation or technological transition.

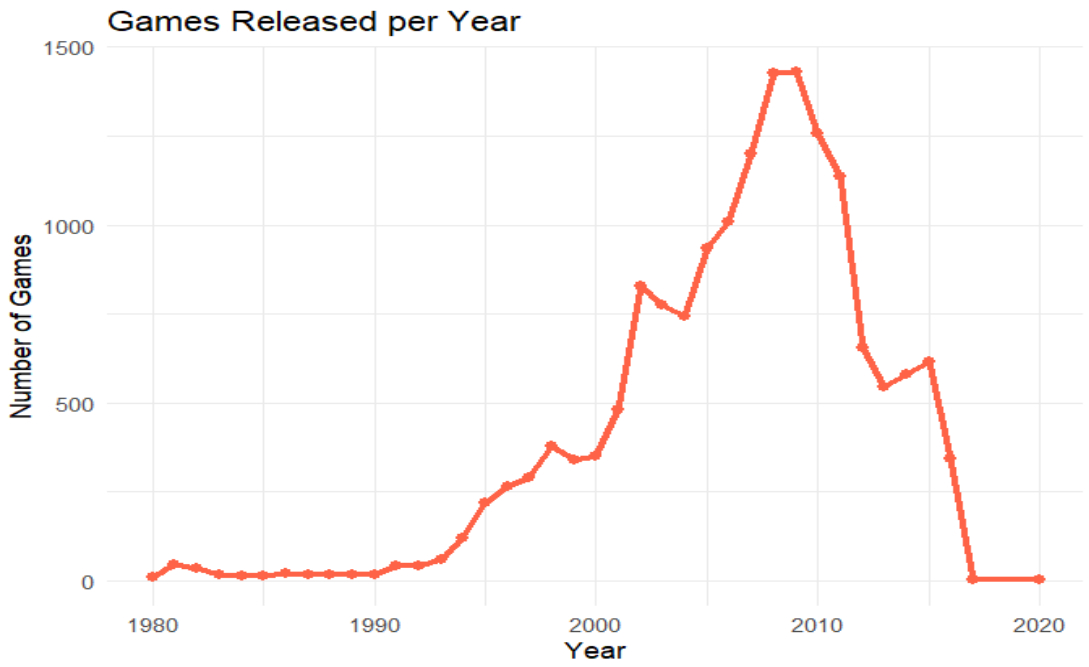
General Description: The analysis focuses on aggregating the number of games released per calendar year to visualize trends over time. This helps identify boom periods and potential slowdowns across the global video game market.

Specific Requirements, Functions, and Formulas:

```
df %>% filter(!is.na(Year)) %>% group_by(Year) %>% summarise(Count = n())
```

```r
df %>%
  filter(!is.na(Year)) %>%
  group_by(Year) %>%
  summarise(Count = n()) %>%
  ggplot(aes(x = Year, y = Count)) +
  geom_line(color = "tomato", size = 1.2) +
  geom_point(size = 2, color = "tomato") +
  labs(title = "Games Released per Year", x = "Year", y = "Number of Games") +
  theme_minimal()
```

Visualization: A line plot with dots representing yearly counts of game releases, clearly depicting the rise and fall in release volume over time.



Analysis Results: The number of games released increased significantly from the late 1990s and peaked between 2007 and 2010, indicating a golden age in physical game publishing. Post-2010, a noticeable decline suggests the rise of digital distribution platforms, which are not comprehensively represented in this dataset

2. Global Video Game Sales Trend Over Years

Introduction: Examining global video game sales trends over time offers a macroeconomic view of the industry's commercial performance. This analysis reveals periods of market expansion and potential saturation, enabling industry stakeholders to understand shifts in consumer behavior and technology adoption.

General Description: This analysis calculates the total global sales per year to identify high-performing years, economic cycles, and the impact of next-generation consoles on total revenue.

Specific Requirements, Functions, and Formulas:

```
## Load Dataset ##

df <- read.csv("C:/Users/d4i7v/OneDrive/Documents/Mtech/sem2/INT553DATACLEANINGR/completeRproject/vgsales.csv")
datatable(df)
View(df)
### Data Overview ##
colnames(df)
as.list(colnames(df))

df$Name
df %>% select(Name, Platform, Year, Genre)
datatable(df %>% select(Name, Platform, Year, Genre, Publisher))
```

Visualization: A time series line plot showing the total global sales in millions per year with annotated peaks during key years in the console lifecycle.

Analysis Results: The global video game market experienced steady growth during the early 2000s, reaching a peak around 2008–2009. This surge may be attributed to the popularity of platforms like the Nintendo Wii, Xbox 360, and PlayStation 3, as well as the introduction of

motion gaming and broader audience engagement. However, after 2009, the market began to decline, possibly due to the saturation of seventh-generation consoles, economic recession effects, and increased competition from mobile gaming platforms. The trend underscores the importance of innovation cycles and evolving consumer preferences.

3. Top 10 Games by Global Sales

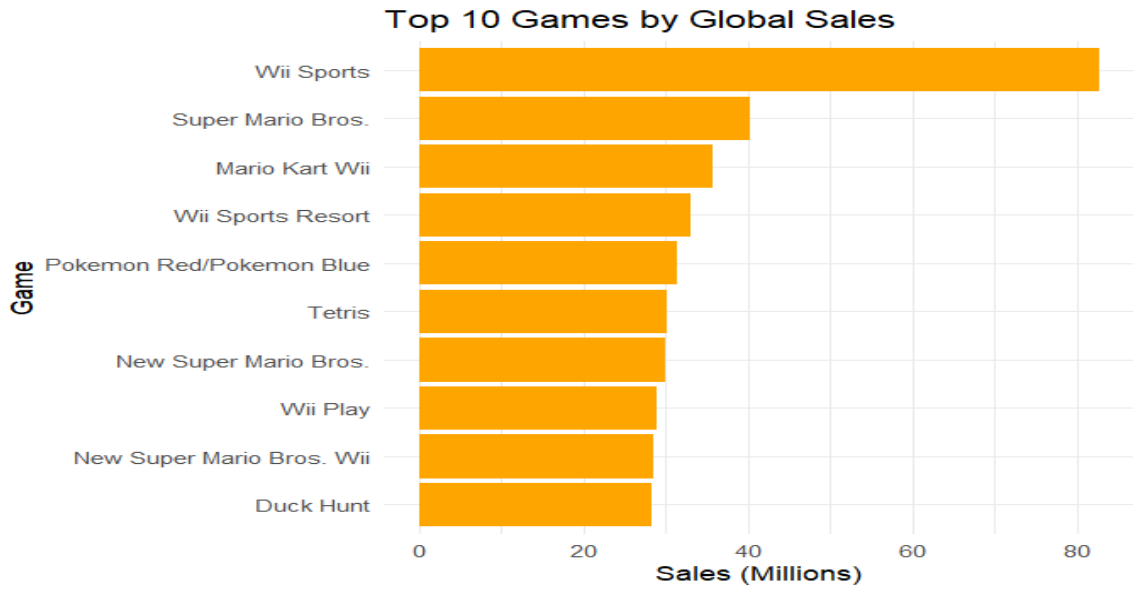
Introduction: Identifying the top-performing games provides insight into franchise success, publisher strategies, and user preferences. It also helps to contextualize which titles have historically driven the highest revenue and popularity across the gaming landscape.

General Description: The dataset is filtered to extract the top 10 video games based on their total global sales. These games are then displayed in descending order for comparison.

Specific Requirements, Functions, and Formulas:

```
# 3. Top 10 Games by Global Sales

df %>% arrange(desc(Global_Sales)) %>% select(Name,Year,Genre,Platform,Global_Sales) %>% head(10)

top10_games <- df %>%
  arrange(desc(Global_Sales)) %>%
  slice(1:10)

ggplot(top10_games, aes(x = reorder(Name, Global_Sales), y = Global_Sales)) +
  geom_col(fill = "Orange") +
  coord_flip() +
  labs(title = "Top 10 Games by Global Sales", x = "Game", y = "Sales (Millions)") +
  theme_minimal()
```

Visualization: A horizontal bar chart effectively displays the top 10 games with their respective global sales, allowing for clear comparative insight between leading titles.



Analysis Results: The analysis revealed that *Wii Sports* (Wii) dominates the global sales chart with over 82 million copies sold, followed by *Super Mario Bros.* (NES), *Mario Kart Wii* (Wii), and *Pokémon Red/Blue* (GB). Nintendo published a majority of these titles, reaffirming their stronghold in the industry, particularly with family-friendly and widely accessible game designs.

4. Popular Platforms by Number of Games Released

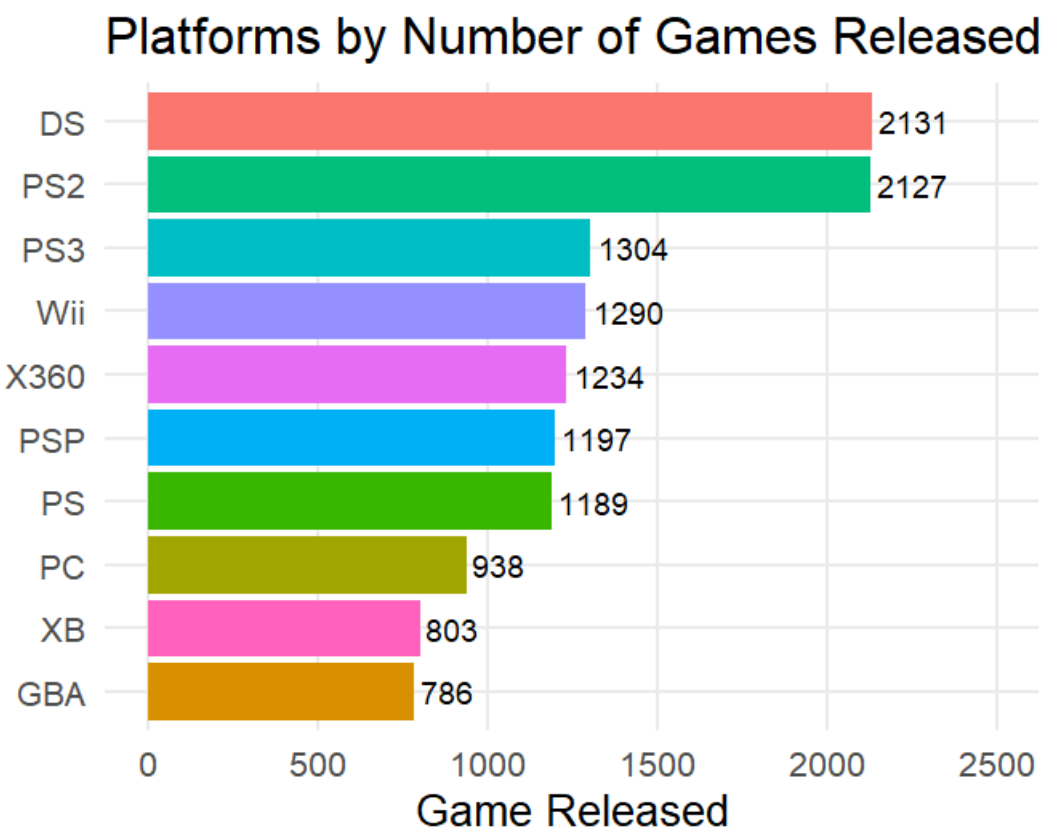
Introduction: Understanding which gaming platforms hosted the highest number of game releases helps identify where developers focused their attention and where gamers had the most choices. It also reflects the market dominance and ecosystem maturity of these platforms.

General Description: This analysis counts the number of games released per platform and ranks them to highlight the most developer-supported systems.

Specific Requirements, Functions, and Formulas:

```
# 4. Popular Platforms by Number of Games Released

unique(df$Platform)
df %>% group_by(Platform) %>% count(Platform) %>%arrange(desc(n)) %>% head()

platform_game<-df %>%
  group_by(Platform) %>%
  count() %>%
  arrange(desc(n)) %>% head(10) # for top 10 other wise remove head(10)

ggplot(platform_game, aes(x = n, y = reorder(Platform, n), fill = Platform)) +
  geom_col() +
  geom_text(aes(label = n), vjust = 0.5, hjust = -0.1, size = 4) +
  theme_minimal(base_size = 16, base_family = "Source Sans Pro") +
  theme(panel.grid.minor = element_blank()) +
  scale_x_continuous(limits=c(0, 2500)) +
  labs(x = "Game Released",
       y = element_blank(),
       title = "Platforms by Number of Games Released") +
  guides(fill = 'none')
```

Visualization: A pie chart is used to display the sales distribution among the top 5 platforms, with clear proportional representation and visual emphasis on the leading platform.



5. Top 5 Platforms by Total Sales

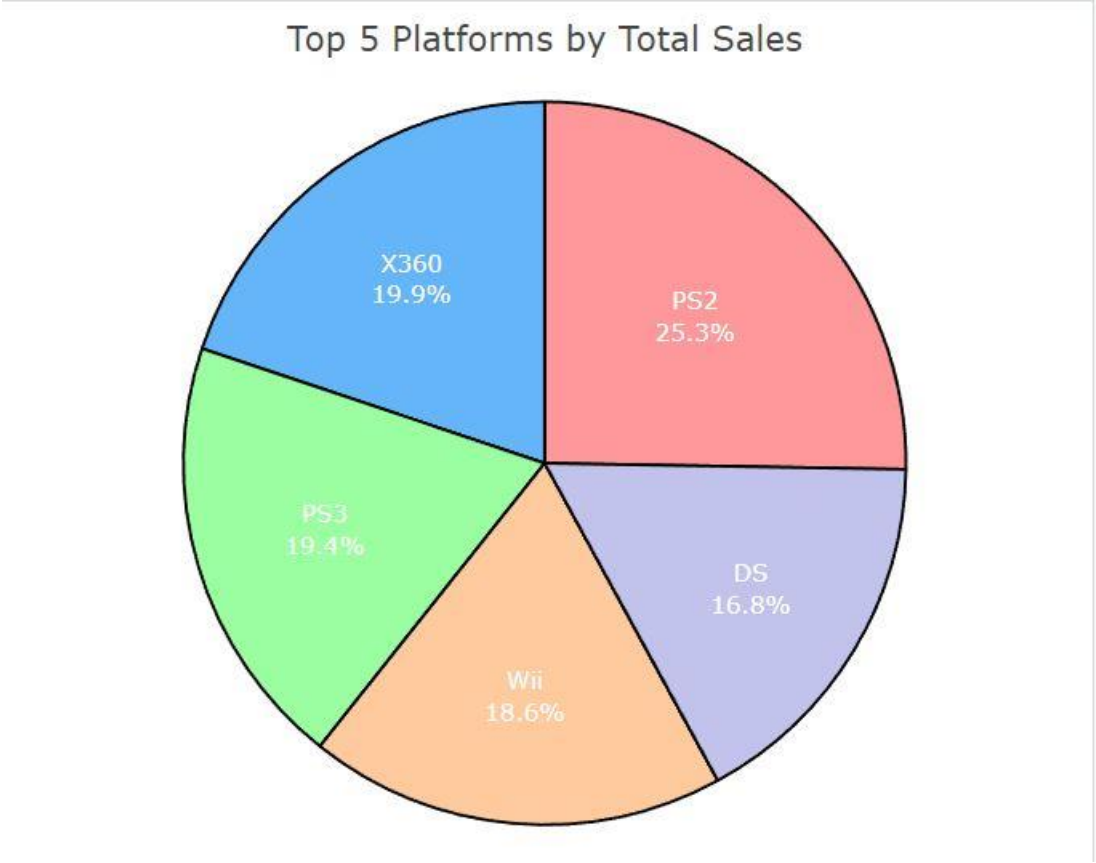
Introduction: Identifying which gaming platforms have generated the highest global sales provides insight into platform success and consumer adoption. This analysis highlights the dominance of specific consoles and the role they play in driving industry revenue.

General Description: This section creates an interactive pie chart that shows the share of total global sales among the top 5 platforms, making it easier to compare proportions visually.

Specific Requirements, Functions, and Formulas:

```
platform_sales <- df %>%
  group_by(Platform) %>%
  summarise(Global_Sales = sum(Global_Sales, na.rm = TRUE)) %>%
  arrange(desc(Global_Sales)) %>%
  slice_head(n = 5) %>%
  mutate(Platform = factor(Platform, levels = Platform),
         pct = Global_Sales / sum(Global_Sales) * 100,
         label = paste0(Platform, ": ", round(pct, 1), "%"))
plot_ly(platform_sales, labels = ~Platform, values = ~Global_Sales, type = 'pie',
        textinfo = 'label+percent',
        insidetextfont = list(color = '#FFFFFF'),
        marker = list(colors = c('#ff9999', '#66b3ff', '#99ff99', '#ffcc99', '#c2c2f0'),
                      line = list(color = '#000000', width = 1.5)),
        showlegend = FALSE) %>%
  layout(title = 'Top 5 Platforms by Total Sales',
         margin = list(l = 20, r = 20, b = 20, t = 50))
```

Visualization: A pie chart is used to display the sales distribution among the top 5 platforms, with clear proportional representation and visual emphasis on the leading platform.



Analysis Results: The top platforms by total global sales include PlayStation 2 (PS2), Xbox 360 (X360), Wii, and others. These platforms significantly outperformed their competitors due to their extensive game libraries, strong marketing strategies, and broad consumer appeal. The success of these platforms reflects their influence during key gaming generations.

6. Popularity of Game Genres by Global Sales

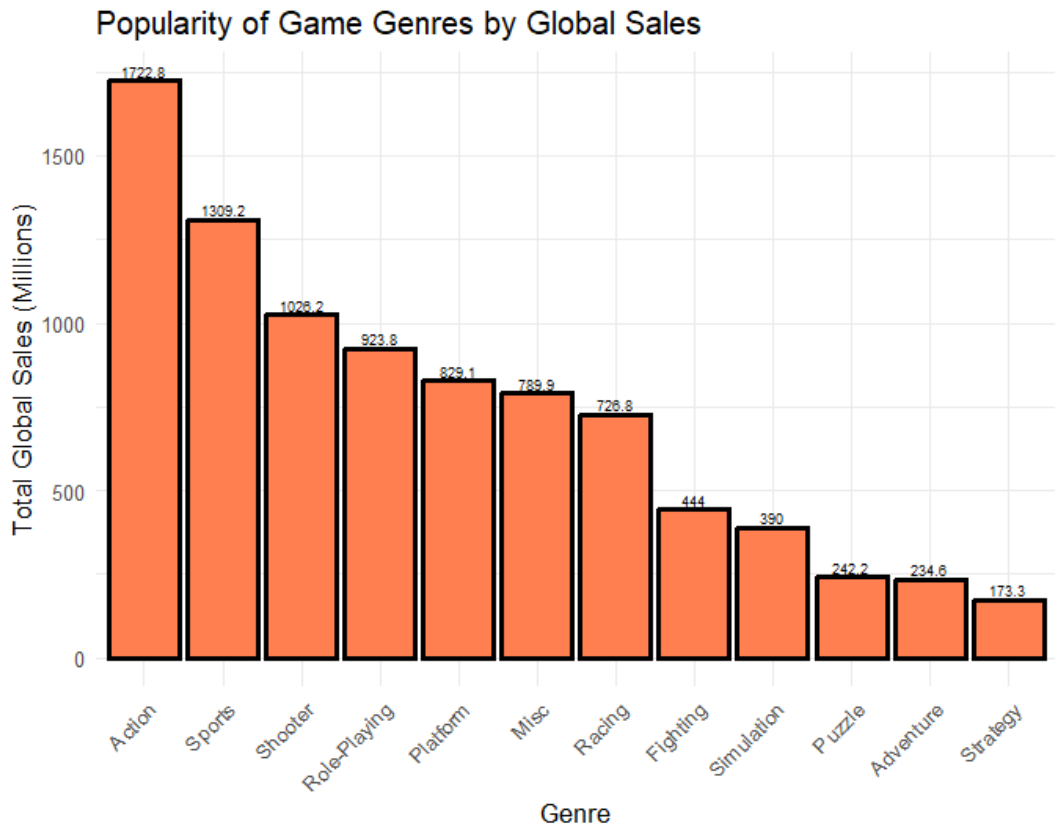
Introduction: Understanding which game genres generate the highest sales provides key insight into consumer preferences and market demand. This can help guide development strategies and marketing efforts.

General Description: Total global sales were aggregated by genre to identify the most commercially successful types of games.

Specific Requirements, Functions, and Formulas:

```
genre_sales <- df %>%
  group_by(Genre) %>%
  summarise(Global_Sales = sum(Global_Sales, na.rm = TRUE)) %>%
  arrange(desc(Global_Sales))

# Plot with text labels
ggplot(genre_sales, aes(x = reorder(Genre, -Global_Sales), y = Global_Sales)) +
  geom_col(fill = "coral", color = "black", linewidth = 1) +
  geom_text(aes(label = round(Global_Sales, 1)),
            vjust = -0.3, size = 2) +
  labs(title = "Popularity of Game Genres by Global Sales",
       x = "Genre",
       y = "Total Global Sales (Millions)") +
  theme_minimal(base_size = 10) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Visualization: A bar chart displays the total sales by genre with numerical labels above each bar, allowing for easy comparison and clear ranking of genre popularity.



Analysis Results: The genres “Action” and “Sports” emerged as the most popular in terms of total sales, followed by “Shooter,” “Role-Playing,” and “Platform.” Action games likely benefited from widespread appeal across age groups and demographics, while sports and shooter games have strong annual franchises and competitive appeal.

7. Number of Games by Genre (Count Plot)

Introduction

While global sales figures provide insights into financial success, it is also crucial to understand the frequency of releases by genre. This analysis focuses on how many games belong to each genre, helping identify industry production trends and genre saturation.

General Description

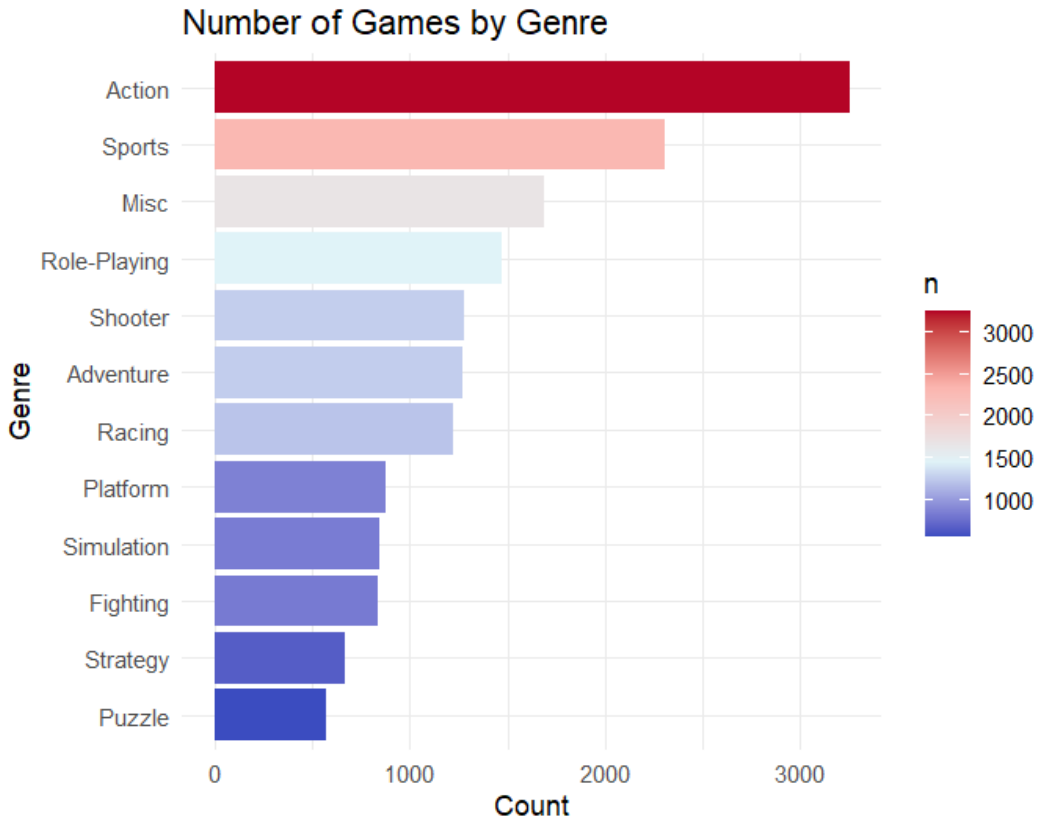
The analysis uses a count plot to visualize the distribution of games across different genres. This helps answer questions like which genres are most commonly produced and whether high sales align with genre popularity.

Specific Requirements, Functions, and Formulas

```
genre_count <- df %>% count(Genre)

# Apply a coolwarm gradient fill
ggplot(genre_count, aes(x = reorder(Genre, n), y = n, fill = n)) +
  geom_col() +
  coord_flip() +
  scale_fill_gradientn(colors = scales::gradient_n_pal
                       (c("#3b4cc0", "#e0f3f8", "#fbb4ae", "#b40426"))(seq(0, 1, length.out = 100))) +
  labs(title = "Number of Games by Genre", x = "Genre", y = "Count") +
  theme_minimal()
```

Visualization

A count plot shows the number of game entries per genre. The x-axis represents the genre, and the y-axis shows the count. Genres are ordered from most to least frequent, and color gradients help distinguish between categories visually.



Analysis Results

The most frequently produced genre is **Action**, followed by **Sports**, **Miscellaneous**, and **Role-Playing**. This aligns closely with genre-wise sales performance and reflects broader trends in consumer interest and ease of development. Action games, due to their wide appeal and flexible gameplay styles, are favored by many developers and publishers.

8. Trend of Count of Games Released Over the Years

Introduction

Exploring how the number of games released across different genres changes over time helps reveal evolving user preferences and industry focus. This can identify when specific genres gained or lost popularity.

General Description

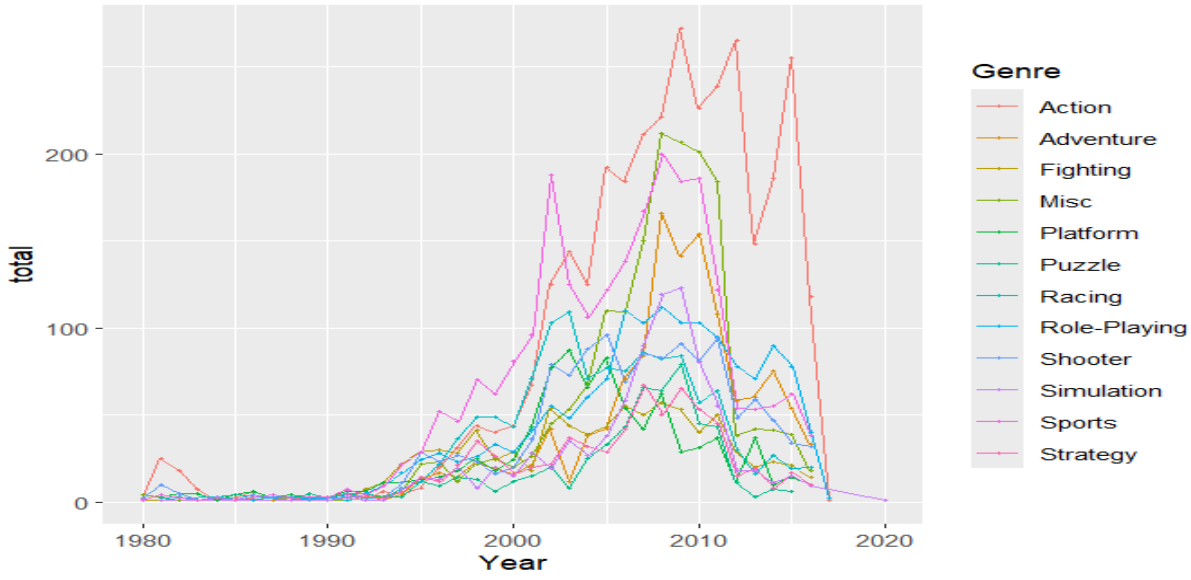
This analysis groups the data by both Year and Genre and counts the number of games released annually for each genre. The trends are then visualized to show the rise and fall of genre popularity over time.

Specific Requirements, Functions, and Formulas

```
# 8. By Trend of Count of games released over the years
genres <- subset(df, select = c(Year, Genre)) %>%
  filter(Year < 2021) %>%
  group_by(Year, Genre) %>%
  summarize (total = n())
ggplot(genres, aes(x = Year, y = total, color = Genre)) +
  geom_line(size = .5, position = position_dodge(width = 0.1)) + geom_point(size = .5)
```

Visualization

A horizontal bar chart is used to present the top six publishers by the number of games released. This visualization allows for easy comparison and emphasizes the scale of operations of leading publishers.



Analysis Results

The top publishers by volume include major industry players such as **Electronic Arts**, **Activision**, and **Namco Bandai Games**. These companies are known for their long-standing presence in the market, frequent franchise releases, and diversified game portfolios. Their dominance in terms of volume indicates an aggressive release strategy often accompanied by strong marketing and global distribution networks.

9. Genre Sales Trend Over the Years

Introduction

Tracking the sales performance of different genres over time helps us understand how player preferences have shifted. It also reflects the commercial success and lifecycle of certain genres.

General Description

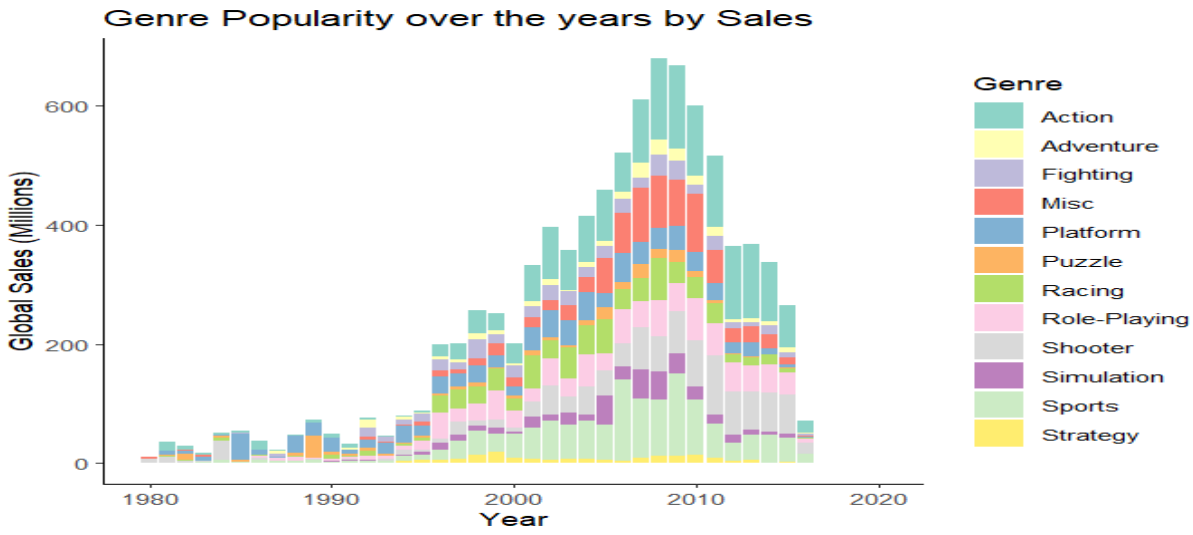
This analysis aggregates the total global sales for each genre annually to detect sales trends and key periods of genre performance.

Specific Requirements, Functions, and Formulas

```
# 9. genre sales  trend over the year
unique(df$Year)
unique(df$Genre)
df %>%
  group_by (Genre, Year) %>%
  summarize(total_sales = sum(Global_Sales)) %>%
  ggplot(aes(x = Year, y = total_sales, fill = Genre)) +
  geom_col() + scale_fill_brewer(palette = "Set3") +
  theme_classic() + labs(title = "Genre Popularity over the years by Sales", y = "Global Sales (Millions)")
```

Visualization

A stacked bar chart showing genre-wise sales trends year-over-year, enabling a comparative look at the rise and fall of each genre's popularity.



Analysis Results

The chart reveals how Action and Shooter genres experienced significant growth during the late 2000s. Other genres such as Role-Playing and Sports show consistent performance, while niche genres like Strategy and Puzzle display fluctuating or declining trends.

10. Genre Sales Trend Over the Years

Introduction

Analyzing genre-based sales performance across regions provides insight into regional preferences. Certain genres perform better in specific markets due to cultural preferences, local publishers, and gaming habits.

General Description

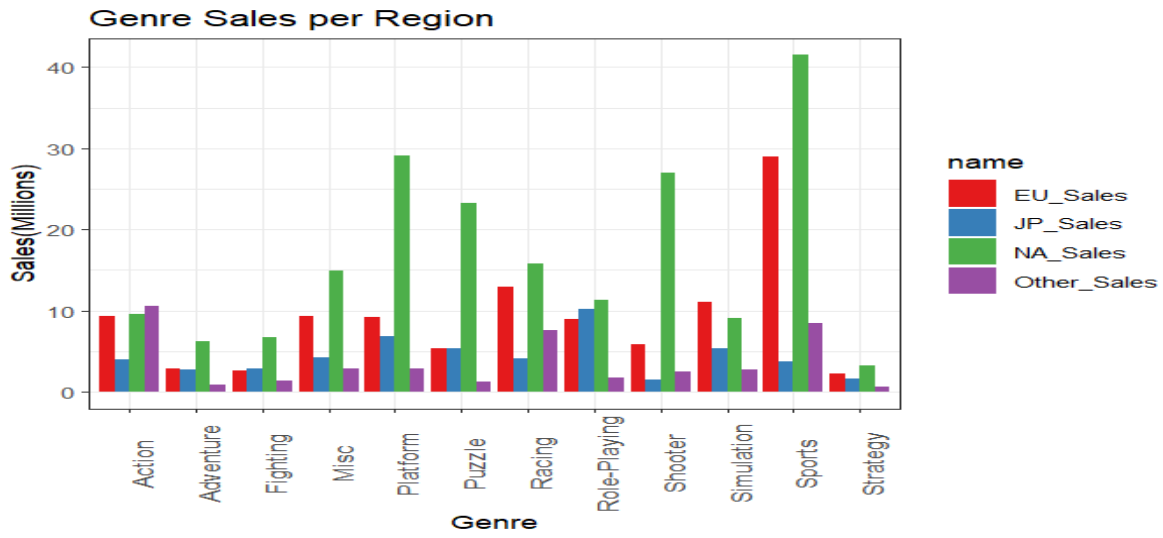
Tracking the sales performance of different genres over time helps us understand how player preferences have shifted. It also reflects the commercial success and lifecycle of certain genres.

Specific Requirements, Functions, and Formulas

```
# 10. genre Sales Per Region
sale = df %>%
  subset(select = c(Genre, NA_Sales, EU_Sales, JP_Sales, Other_Sales)) %>%
  group_by (Genre) %>%
  pivot_longer(c("NA_Sales", "EU_Sales", "JP_Sales", "Other_Sales"), values_to = "Sales")
ggplot(sale, aes(x = Genre, y = Sales, fill = name)) + geom_col(position = "dodge") +
  labs(title = "Genre Sales per Region", x= "Genre", y = "Sales(Millions)") +
  theme_bw() + theme(axis.text.x= element_text(size =10, angle = 90)) + scale_fill_brewer(palette = "Set1")
```

Visualization

A horizontal bar chart is used to present the top six publishers by the number of games released. This visualization allows for easy comparison and emphasizes the scale of operations of leading publishers.



Analysis Results

The top publishers by volume include major industry players such as **Electronic Arts**, **Activision**, and **Namco Bandai Games**. These companies are known for their long-standing presence in the market, frequent franchise releases, and diversified game portfolios. Their dominance in terms of volume indicates an aggressive release strategy often accompanied by strong marketing and global distribution networks.

11. **Top 10 Publishers with Most Games Released**

Introduction:

Understanding which publishers have released the most games helps identify the most active and prolific entities in the video game industry. This metric reflects not only company output but also market influence.

General Description

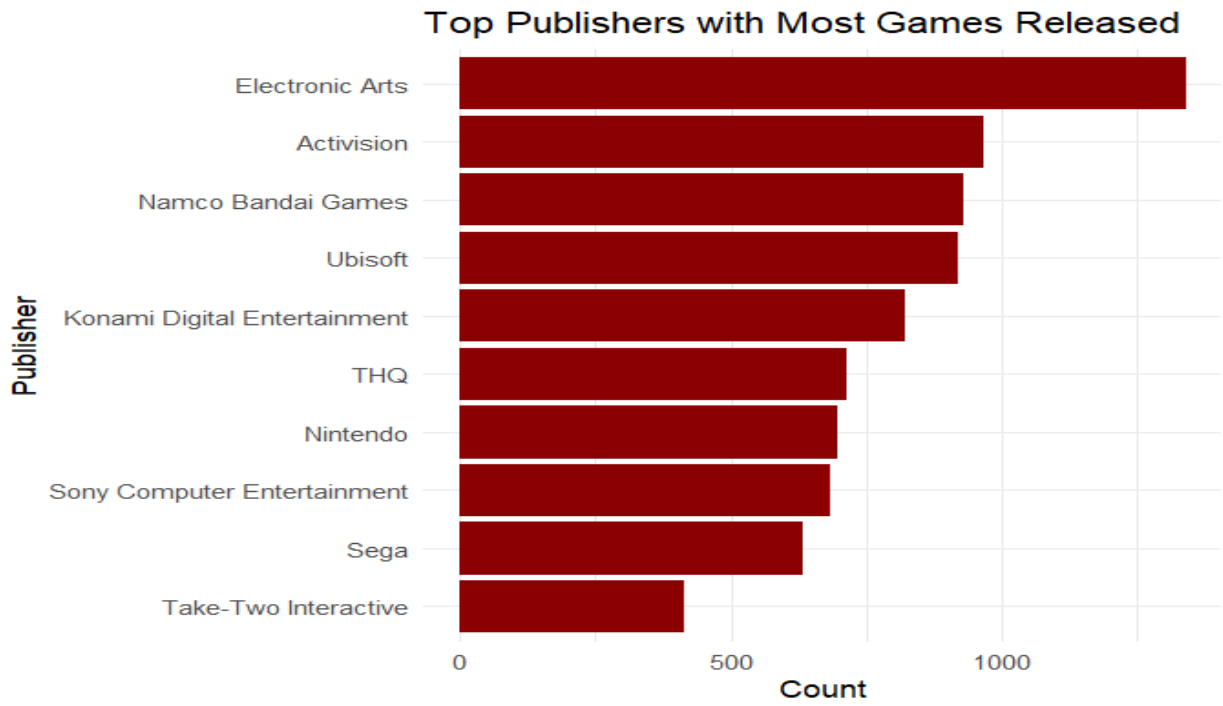
This analysis counts the number of games published by each company and ranks the top 10 based on frequency.

Specific Requirements, Functions, and Formulas

```
# 11. Top 10 Publishers with Most Games Released
as.list(unique(df$Publisher))
top_publishers_count <- df %>%
  count(Publisher, sort = TRUE) %>%
  slice(1:10)

ggplot(top_publishers_count, aes(x = reorder(Publisher, n), y = n)) +
  geom_col(fill = "darkred") +
  coord_flip() +
  labs(title = "Top Publishers with Most Games Released", x = "Publisher", y = "Count") +
  theme_minimal()
```

Visualization

A horizontal bar chart that displays the top 10 publishers by number of games released, helping to identify industry leaders in terms of volume.



Analysis Results

The results show that **Nintendo** is the clear leader in global sales, with a significant margin over other publishers. Other top performers include **Electronic Arts**, **Activision**, and **Sony Computer Entertainment**, all of which have launched high-profile franchises across various console generations. These publishers have leveraged both popular IPs and consistent release strategies to secure high cumulative sales figures.

12. Top Publishers by Global Sales

Introduction

While the volume of game releases is a useful metric, total global sales better reflect a publisher's market impact and revenue performance. This analysis identifies which publishers have amassed the highest global sales, showcasing the commercial strength of industry leaders.

General Description

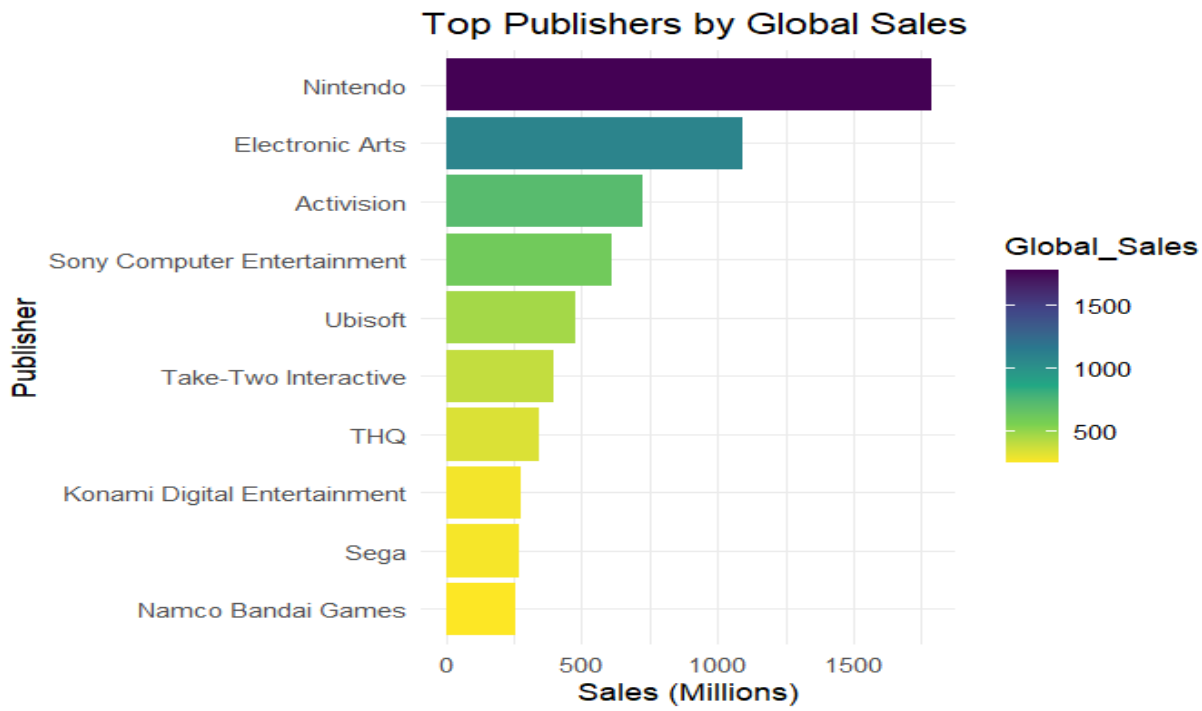
The data is grouped by Publisher, and the total Global_Sales per publisher is summed. The top 10 publishers are then visualized to compare their total contribution to the global market.

Specific Requirements, Functions, and Formulas

```r
top_publishers_sales <- df %>%
  group_by(Publisher) %>%
  summarise(Global_Sales = sum(Global_Sales, na.rm = TRUE)) %>%
  arrange(desc(Global_Sales)) %>%
  slice(1:10)

ggplot(top_publishers_sales, aes(x = reorder(Publisher,Global_Sales), y = Global_Sales, fill= Global_Sales))+
  geom_col() +
  scale_fill_viridis_c(option = "flare", direction = -1) +
  coord_flip() +
  labs(title = "Top Publishers by Global Sales", x = "Publisher", y = "Sales (Millions)") +
  theme_minimal()
```

Visualization

A horizontal bar chart effectively presents the total global sales for the top 10 publishers. This visualization underscores the disproportionate dominance of a few key players in the industry and highlights the gap between Nintendo and its competitors.



Analysis Results

The results show that **Nintendo** is the clear leader in global sales, with a significant margin over other publishers. Other top performers include **Electronic Arts**, **Activision**, and **Sony Computer Entertainment**, all of which have launched high-profile franchises across various console generations. These publishers have leveraged both popular IPs and consistent release strategies to secure high cumulative sales figures.

13. Publisher Sales Tend Over Time

Introduction

Studying how publisher sales evolve over time reveals which companies sustained performance across console generations. It also shows how market leadership shifts in the gaming industry.

General Description

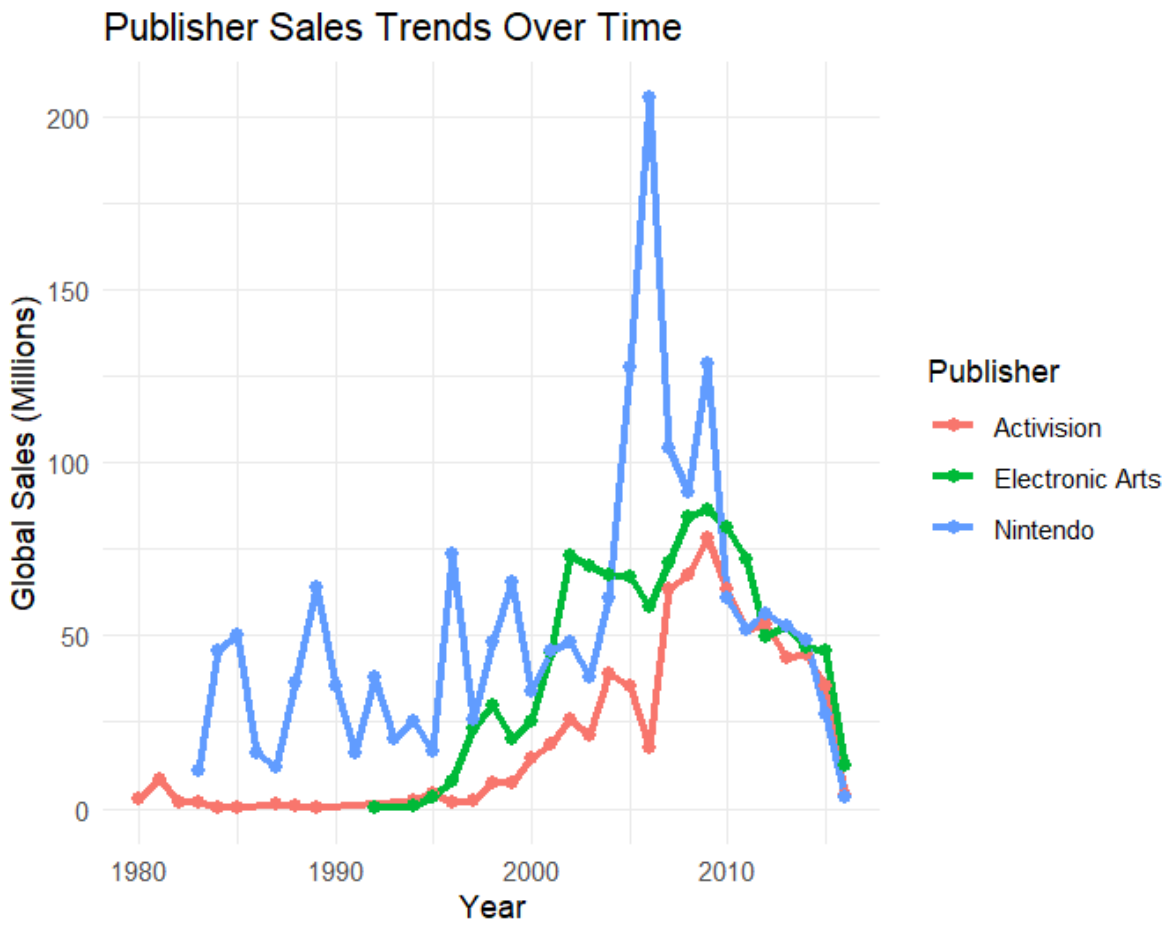
This analysis calculates annual sales for each publisher and filters for the top 3 highest-grossing publishers overall to highlight their performance trends over time.

Specific Requirements, Functions, and Formulas

```
# 13. Publisher Sales Trends Over Time


publisher_trend <- df %>%
  group_by(Year, Publisher) %>%
  summarise(Global_Sales = sum(Global_Sales, na.rm = TRUE)) %>%
  ungroup()

top_publishers <- df %>%
  group_by(Publisher) %>%
  summarise(Total_Sales = sum(Global_Sales, na.rm = TRUE)) %>%
  top_n(3, Total_Sales) %>%
  pull(Publisher)

filtered_trend <- publisher_trend %>%
  filter(Publisher %in% top_publishers)


ggplot(filtered_trend, aes(x = Year, y = Global_Sales, color = Publisher)) +
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  labs(title = "Publisher Sales Trends Over Time",
       x = "Year",
       y = "Global Sales (Millions)") +
  theme_minimal()
```

Visualization

A grouped bar chart presents the top publishers for each region along with their total regional sales. The chart highlights market-specific dominance and reinforces the strategic importance of regional differentiation for publishers.



Analysis Results

The analysis reveals that **Nintendo** leads sales in North America and Japan, leveraging strong brand loyalty and culturally resonant titles. **Sony** performs best in Europe, likely benefiting from localized marketing and a strong PlayStation presence. **Take-Two Interactive** dominates the "Other" category, likely due to blockbuster franchises like Grand Theft Auto that appeal to a broad international audience.

14. Top Publisher by Region

Introduction

Understanding which publishers dominate specific regions (North America, Europe, Japan, and Others) offers insights into regional preferences and publisher strategies. This regional view complements global performance metrics and highlights market segmentation.

General Description

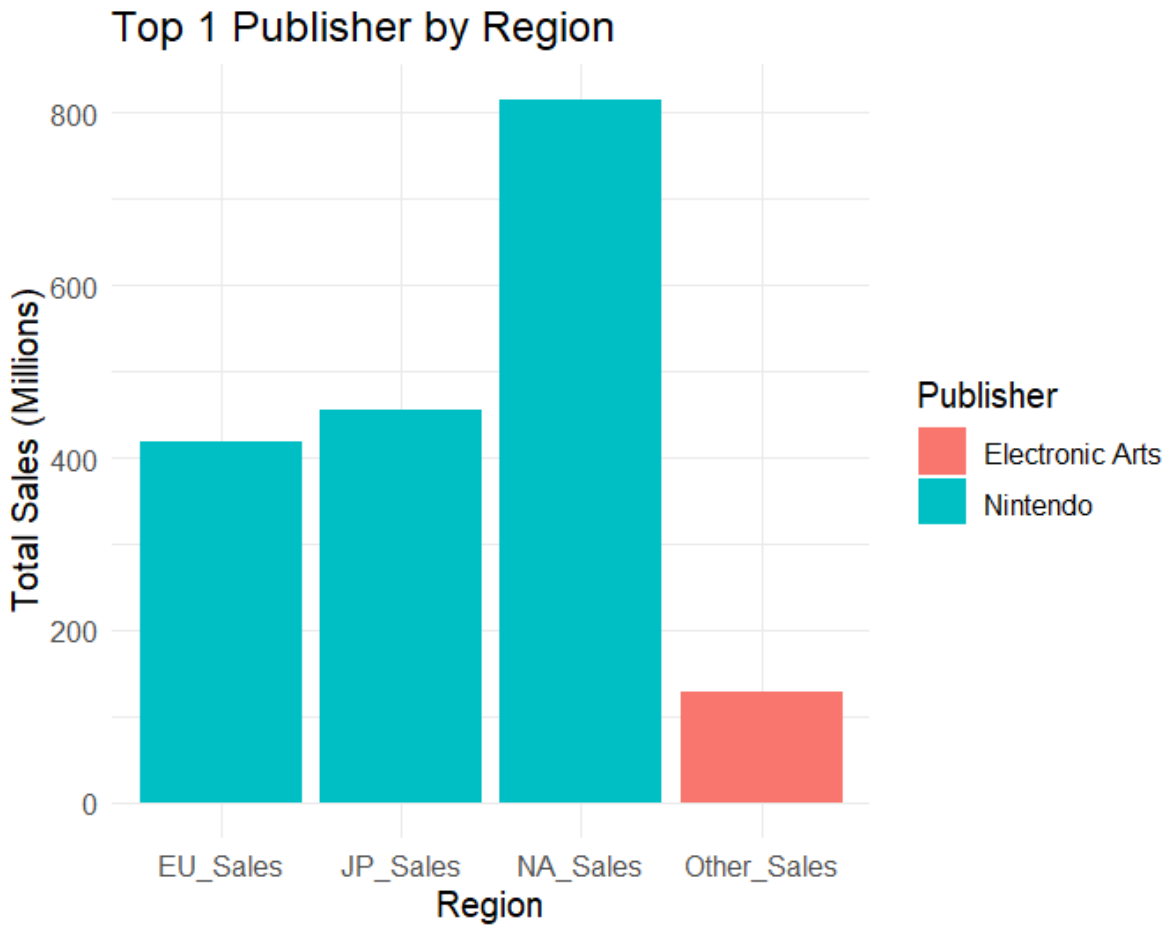
Sales data is grouped by Publisher and by each regional sales column. The top publisher in each region is identified by sorting the summed sales in descending order. A summary DataFrame is created to store and visualize the top publisher for each region.

Specific Requirements, Functions, and Formulas

```
# 14. Top Publisher by each Regions

publisher_sales <- df %>%
  group_by(Publisher) %>%
  summarise(
    NA_Sales = sum(NA_Sales, na.rm = TRUE),
    EU_Sales = sum(EU_Sales, na.rm = TRUE),
    JP_Sales = sum(JP_Sales, na.rm = TRUE),
    Other_Sales = sum(Other_Sales, na.rm = TRUE)
  )
sales_long <- publisher_sales %>%
  pivot_longer(cols = -Publisher, names_to = "Region", values_to = "Sales")
top_publishers <- sales_long %>%
  group_by(Region) %>%
  slice_max(order_by = Sales, n = 1)
ggplot(top_publishers, aes(x = Region, y = Sales, fill = Publisher)) +
  geom_col() +
  labs(title = "Top 1 Publisher by Region",
       x = "Region",
       y = "Total Sales (Millions)",
       fill = "Publisher") +
  theme_minimal() +
  theme(text = element_text(size = 12))
```

Visualization

A grouped bar chart presents the top publishers for each region along with their total regional sales. The chart highlights market-specific dominance and reinforces the strategic importance of regional differentiation for publishers.



Analysis Results

The analysis reveals that **Nintendo** leads sales in North America and Japan, leveraging strong brand loyalty and culturally resonant titles. **Sony** performs best in Europe, likely benefiting from localized marketing and a strong PlayStation presence. **Take-Two Interactive** dominates the "Other" category, likely due to blockbuster franchises like Grand Theft Auto that appeal to a broad international audience.

s

15. Correlation Between Regional and Global Sales

Introduction

Correlation analysis between regional and global sales helps identify which markets contribute most significantly to worldwide performance. It also reveals interdependencies among regional trends and can guide region-focused strategies for future releases.

General Description

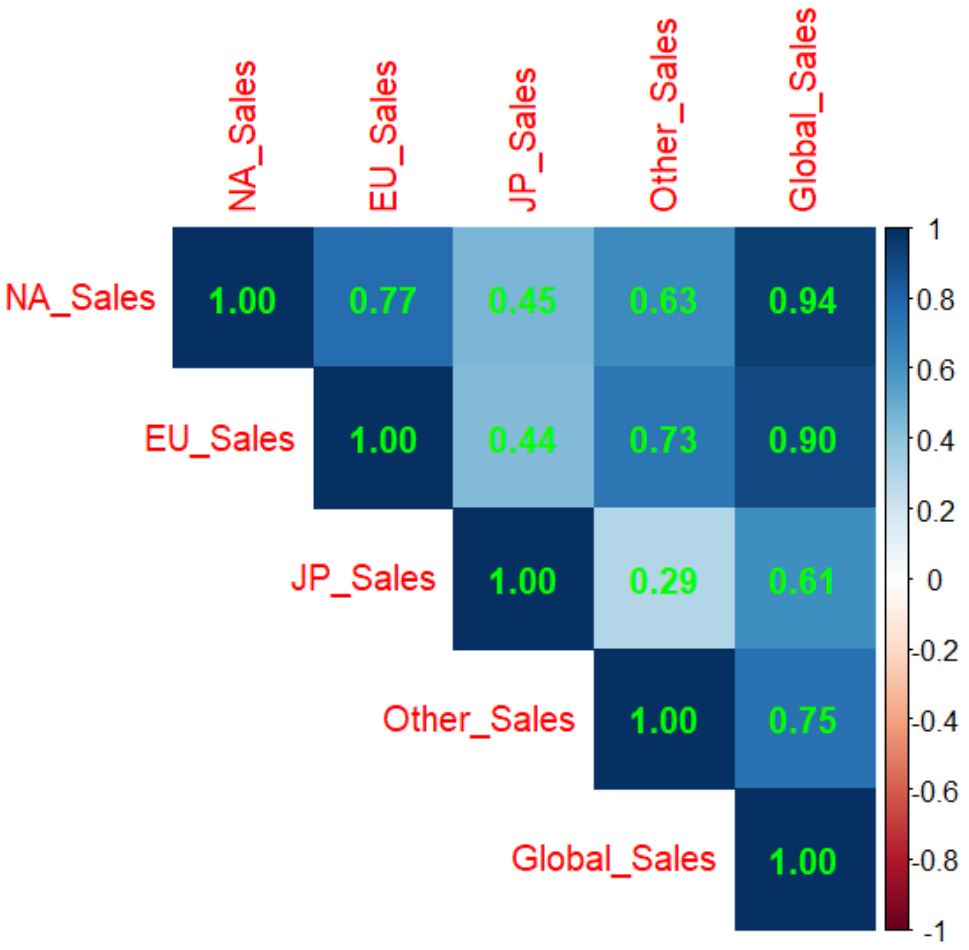
A correlation matrix was computed using sales figures from each region (North America, Europe, Japan, Other) along with total global sales. This correlation matrix was then visualized as a heatmap to detect strong or weak linear relationships.

Specific Requirements, Functions, and Formulas

```
# 15. Correlation Between Regional and Global Sales
sales_data <- df %>%
  select(NA_Sales, EU_Sales, JP_Sales, Other_Sales, Global_Sales)
cor_matrix <- cor(sales_data, use = "complete.obs")
corrplot(cor_matrix, method = "color", type = "upper", addCoef.col = "green")
get_top_publishers <- function(region_col, region_name) {
  df %>%
    group_by(Publisher) %>%
    summarise(Sales = sum(.data[[region_col]], na.rm = TRUE)) %>%
    arrange(desc(Sales)) %>%
    slice_head(n = 5) %>%
    mutate(Region = region_name)
}
```

Visualization

A correlation heatmap visually displays how strongly each regional sales figure is associated with global sales and with other regions. Color gradients from blue (low) to red (high) and numerical annotations provide a clear, interpretable format.



Analysis Results

- **North America (NA_Sales)** and **Europe (EU_Sales)** show strong positive correlations with **Global Sales**, indicating their major influence on total performance.
- **Japan (JP_Sales)** demonstrates a weaker correlation, suggesting regional differences in game preferences or purchasing behavior.
- **Other Sales** contributes moderately to global totals and reflects a diverse but less concentrated market.

These findings support the prioritization of North American and European markets in global launch strategies, while highlighting the need for region-specific approaches in Japan.

16. Top 5 Publishers by Region (Including Global Sales)

Introduction

Analyzing the top publishers in each region provides a comparative look at who leads the market in North America, Europe, Japan, Other regions, and globally. This helps evaluate cross-regional brand presence and strategic success.

General Description

This section defines a reusable function to extract the top publishers per region. These results are combined and displayed using a faceted chart.

Specific Requirements, Functions, and Formulas

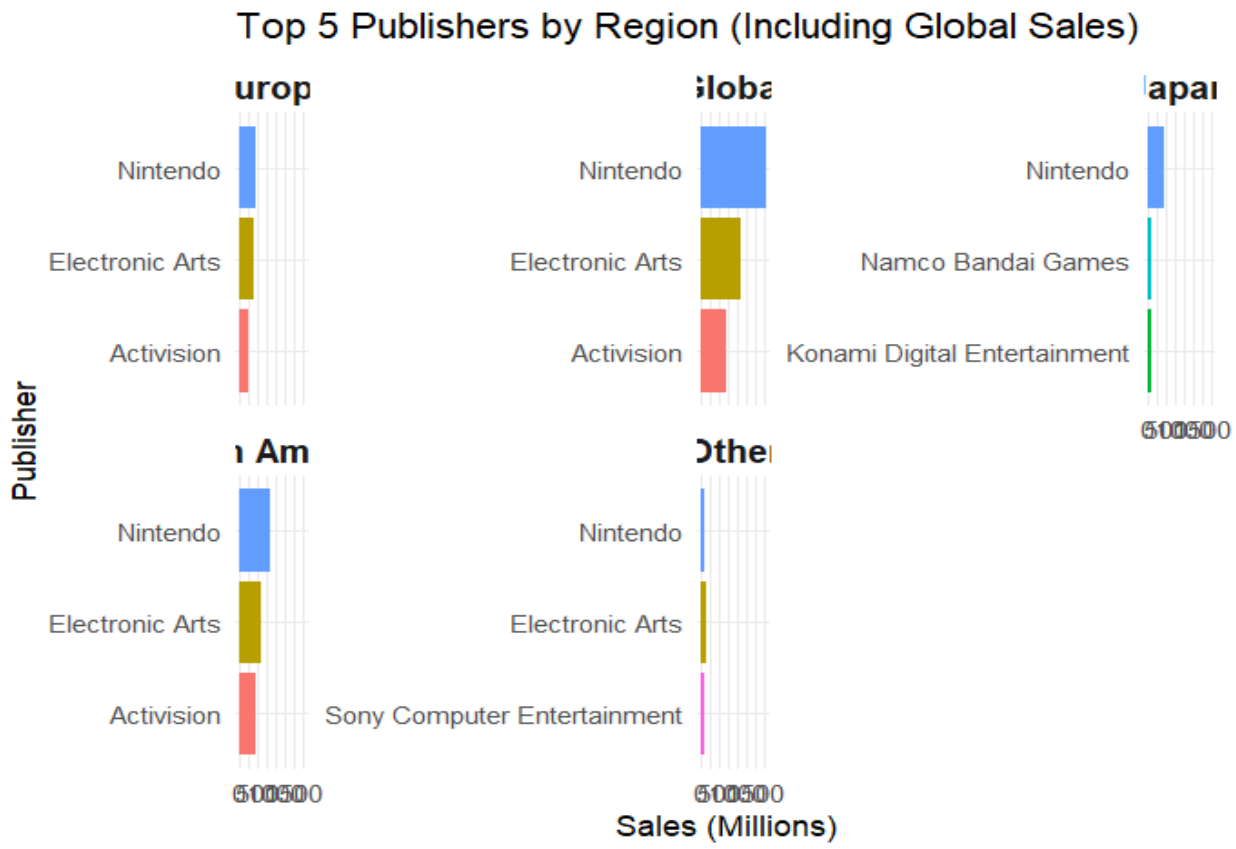
```
# 16. Top 5 publisher by Region
get_top_publishers <- function(region_col, region_name) {
  df %>%
    group_by(Publisher) %>%
    summarise(Sales = sum(.data[[region_col]], na.rm = TRUE)) %>%
    arrange(desc(Sales)) %>%
    slice_head(n = 5) %>%
    mutate(Region = region_name)
}

top_na     <- get_top_publishers("NA_Sales", "North America")
top_eu     <- get_top_publishers("EU_Sales", "Europe")
top_jp     <- get_top_publishers("JP_Sales", "Japan")
top_other  <- get_top_publishers("Other_Sales", "Other")
top_global <- get_top_publishers("Global_Sales", "Global")

top_publishers <- bind_rows(top_na, top_eu, top_jp, top_other, top_global)

ggplot(top_publishers, aes(x = reorder(Publisher, Sales), y = Sales, fill = Publisher)) +
  geom_col(show.legend = FALSE) +
  coord_flip() +
  facet_wrap(~ Region, scales = "free_y") +
  labs(title = "Top 5 Publishers by Region (Including Global Sales)",
       x = "Publisher", y = "Sales (Millions)") +
  theme_minimal() +
  theme(strip.text = element_text(face = "bold", size = 12))
```

Analysis Results:

Nintendo leads in Japan and globally, while EA and Activision are dominant in North America and Europe. This visualization illustrates both regional specialization and global brand leadership.

Visualization

Faceted bar charts for each region showing the top 5 publishers based on sales, allowing for side-by-side comparison of regional market leaders.



17. Publisher Sales Trends Over Time

Introduction

Analyzing how sales evolve over time for major publishers provides insight into their long-term performance, strategic timing, and adaptability across console generations. This time series analysis focuses on the top three publishers in terms of global sales.

General Description

The data is filtered to include only the top 3 publishers by total global sales. For each of these publishers, yearly sales are aggregated across regions (NA, EU, JP, Other), and line plots are generated to visualize regional trends.

Specific Requirements, Functions, and Formulas

```r
#17. Publisher-wise Regional Sales Trends
top_publishers <- df %>%
  group_by(Publisher) %>%
  summarise(Global_Sales = sum(Global_Sales, na.rm = TRUE)) %>%
  arrange(desc(Global_Sales)) %>%
  slice_head(n = 3) %>%
  pull(Publisher)
df_top3 <- df %>%
  filter(Publisher %in% top_publishers) %>%
  group_by(Year, Publisher) %>%
  summarise(
    NA_Sales = sum(NA_Sales, na.rm = TRUE),
    EU_Sales = sum(EU_Sales, na.rm = TRUE),
    JP_Sales = sum(JP_Sales, na.rm = TRUE),
    Other_Sales = sum(Other_Sales, na.rm = TRUE),
    .groups = "drop"
  )
df_long <- df_top3 %>%
  pivot_longer(cols = c("NA_Sales", "EU_Sales", "JP_Sales", "Other_Sales"),
               names_to = "Region", values_to = "Sales")
library(ggplot2)
ggplot(df_long, aes(x = Year, y = Sales, color = Publisher)) +
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  facet_wrap(~ Region, scales = "free_y", ncol = 1,
             labeller = as_labeller(c(
               NA_Sales = "NA Sales Over Time",
               EU_Sales = "EU Sales Over Time",
               JP_Sales = "JP Sales Over Time",
               Other_Sales = "Other Sales Over Time"
             ))) +
  labs(title = "Regional Sales Trends Over Time (Top 3 Publishers)",
       x = "Year", y = "Sales (Millions)", color = "Publisher") +
  theme_minimal(base_size = 14) +
  theme(strip.text = element_text(size = 14, face = "bold"),
        axis.title = element_text(face = "bold"),
        axis.text = element_text(size = 12),
        legend.title = element_text(size = 12),
        legend.text = element_text(size = 11))
```

Analysis Results

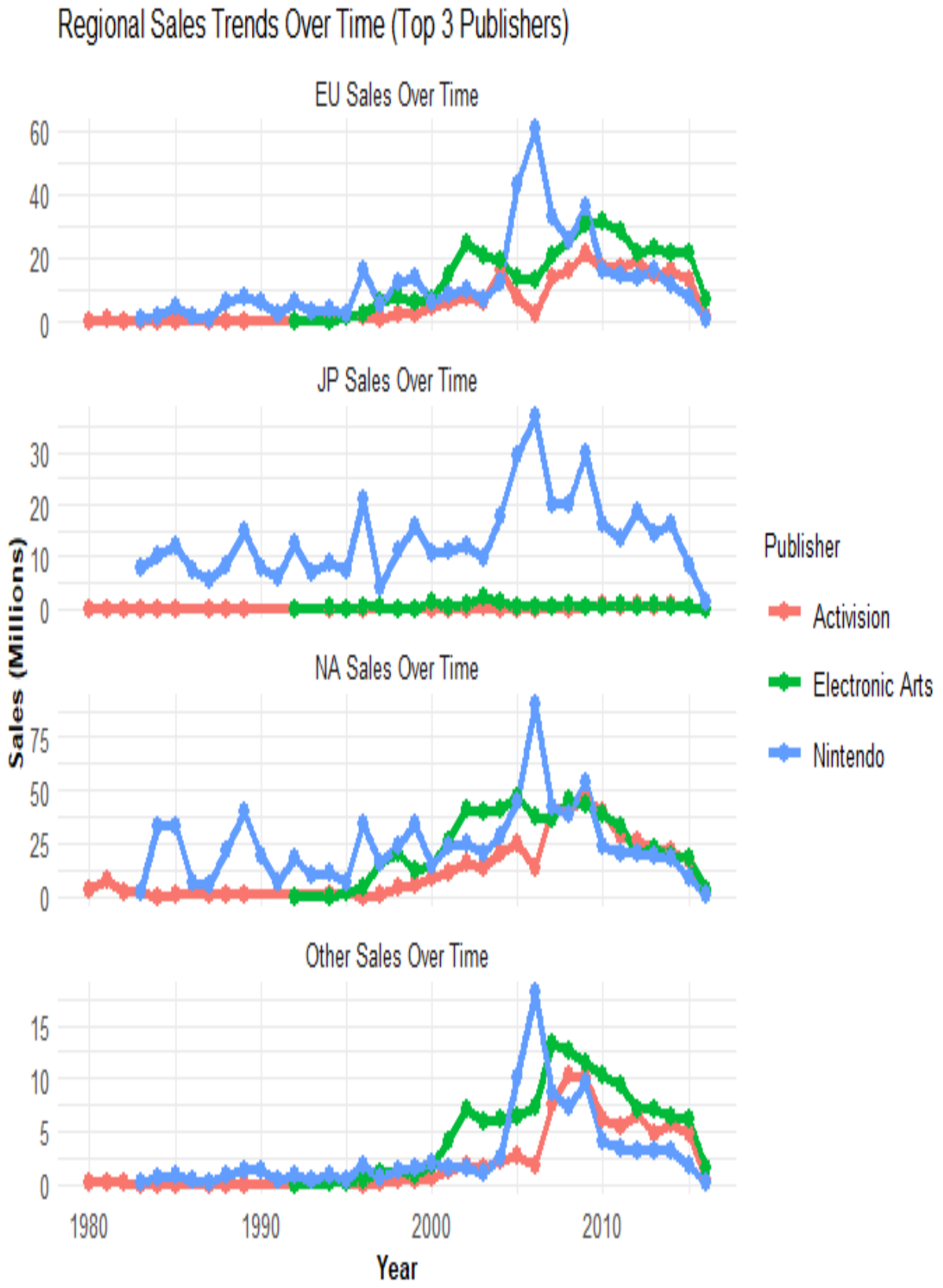
The line plots show:

- **Nintendo** exhibits consistently strong sales across all regions with prominent peaks during the release of major consoles like Wii and Switch.
- **Electronic Arts** performs robustly in North America and Europe but shows lower engagement in Japan.
- **Activision** demonstrates periodic spikes aligned with blockbuster franchises like Call of Duty.

Overall, the visualization captures regional strengths and the ability of publishers to sustain sales momentum over different timeframes.

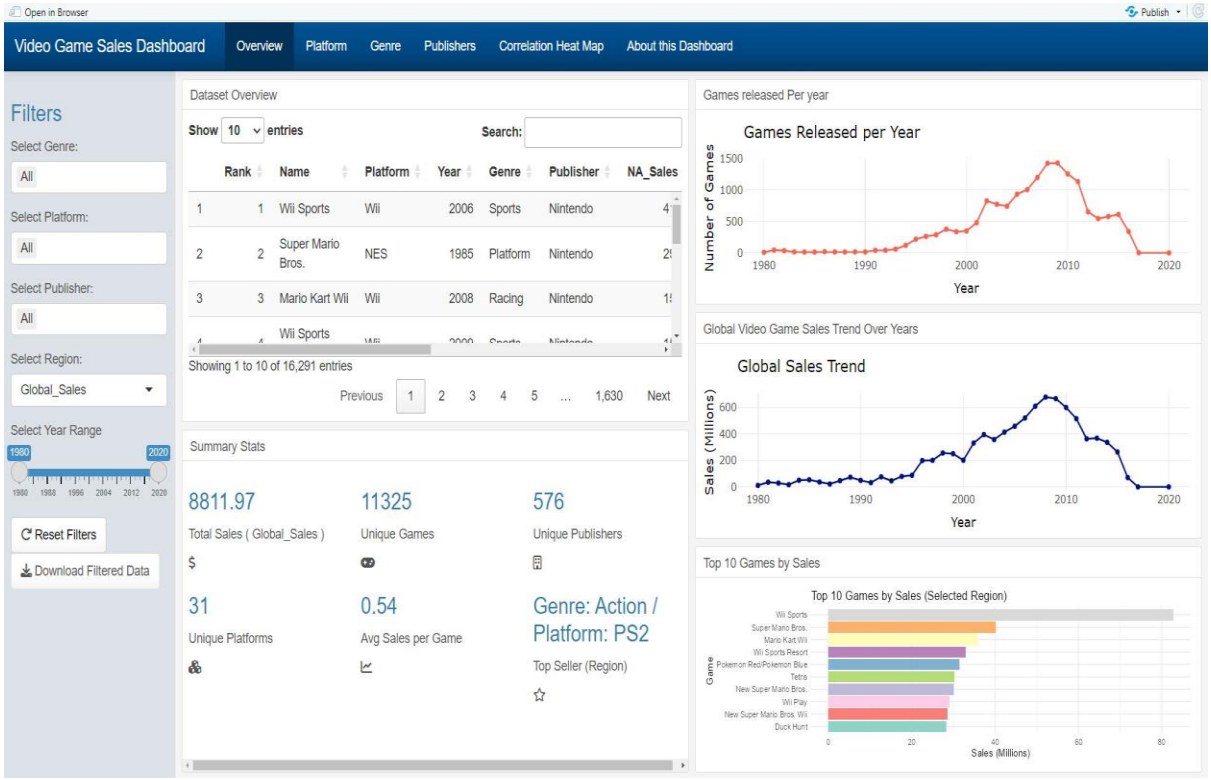
Visualization

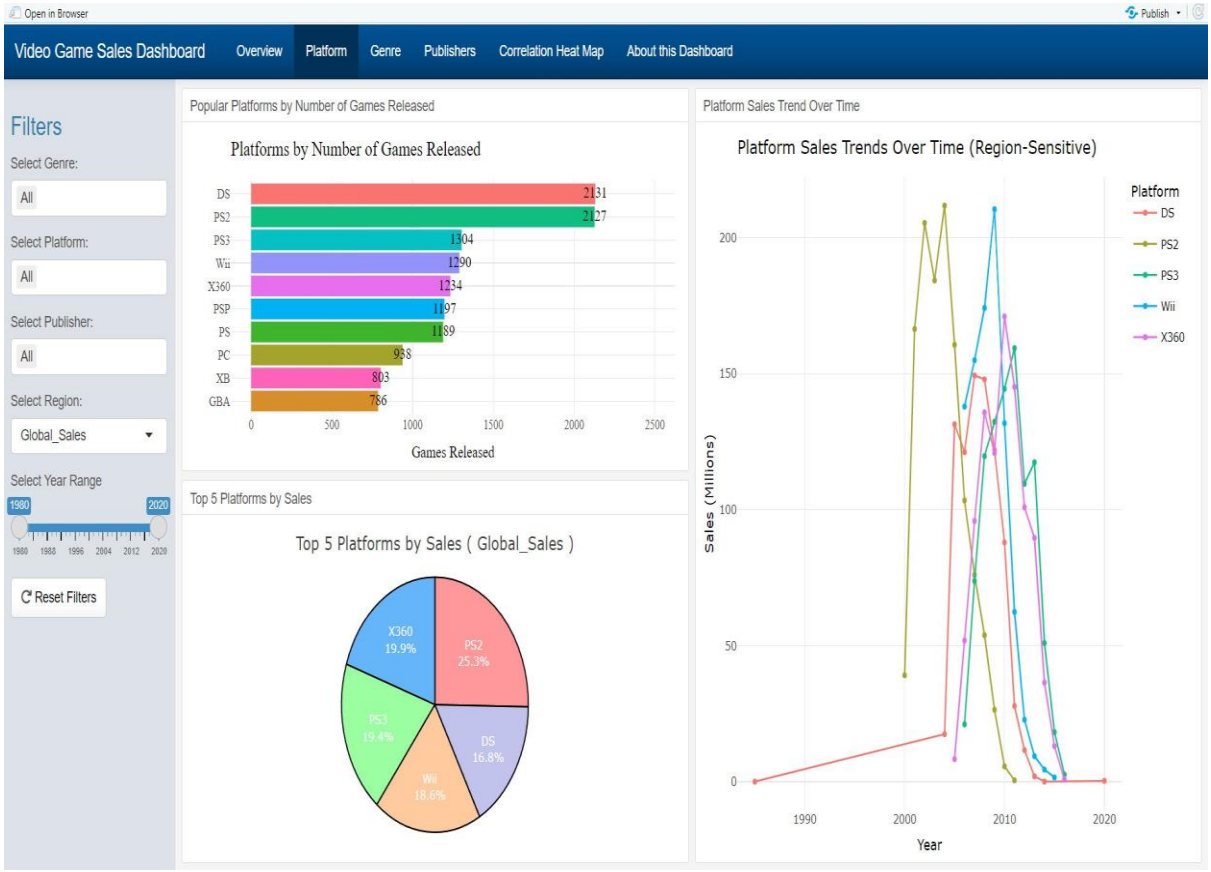
Four line plots are presented—one for each major region—highlighting yearly sales trends for the top three publishers. These plots enable direct comparison of performance across regions and over time.

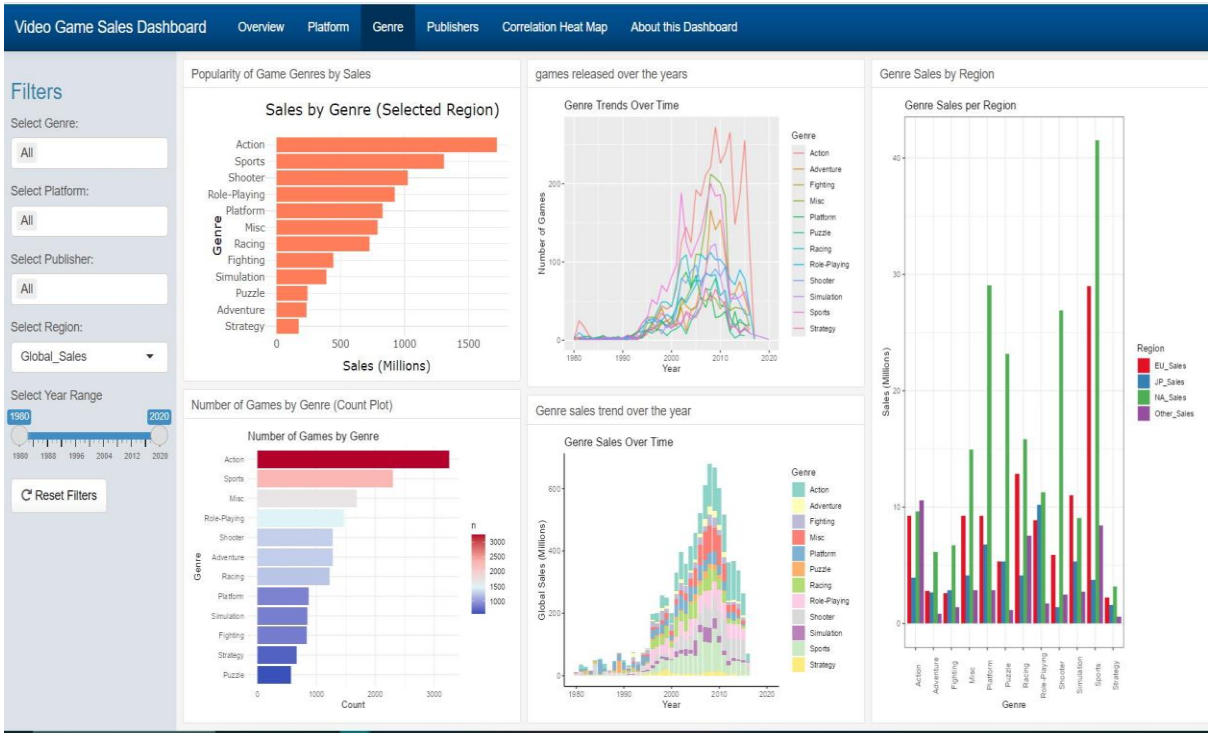

Regional Sales Trends Over Time (Top 3 Publishers)
EU Sales Over Time
JP Sales Over Time
NA Sales Over Time
Other Sales Over Time
Sales (Millions)
Year
Publisher
Activision
Electronic Arts
Nintendo

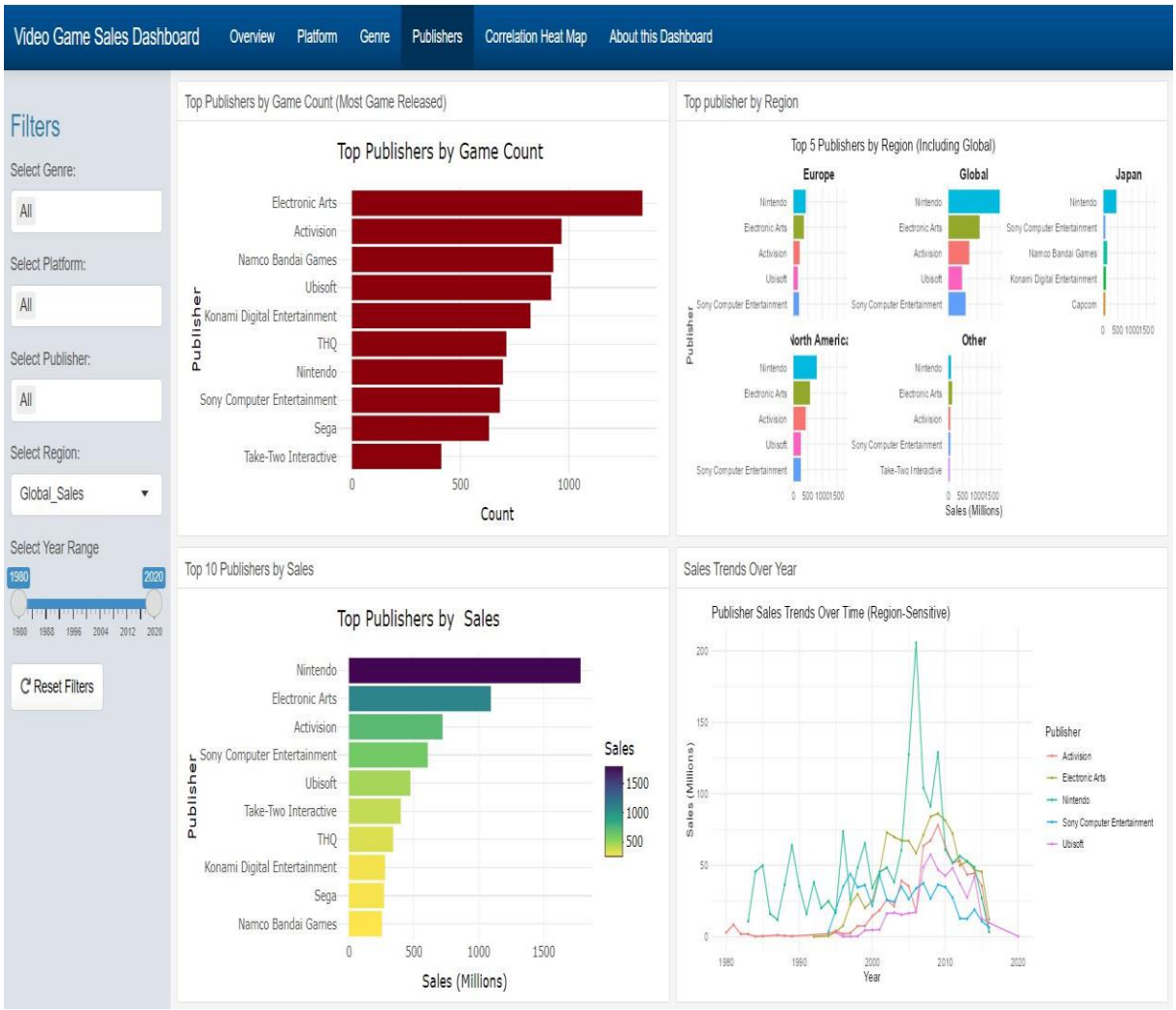
CHAPTER 5

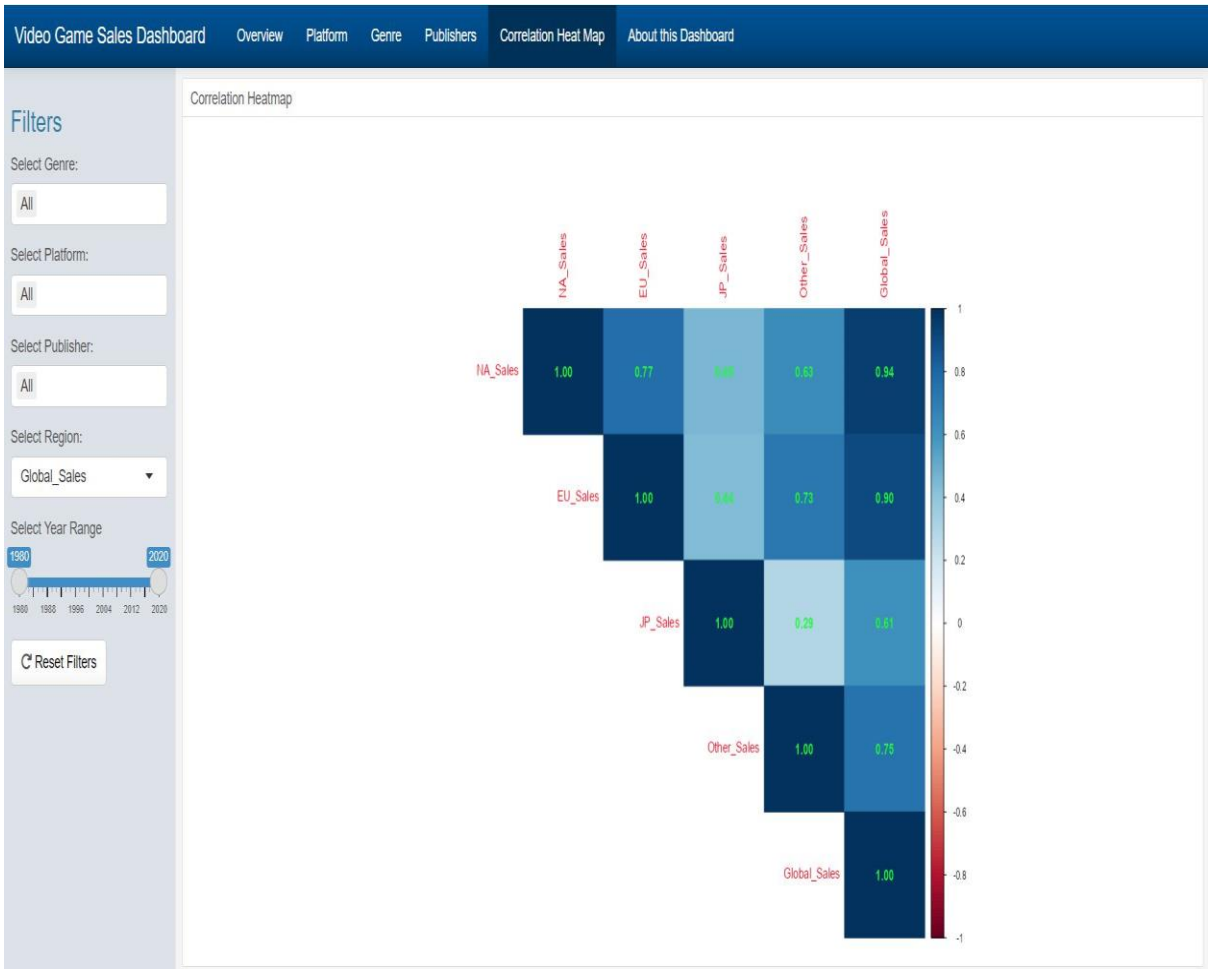
Dashboard

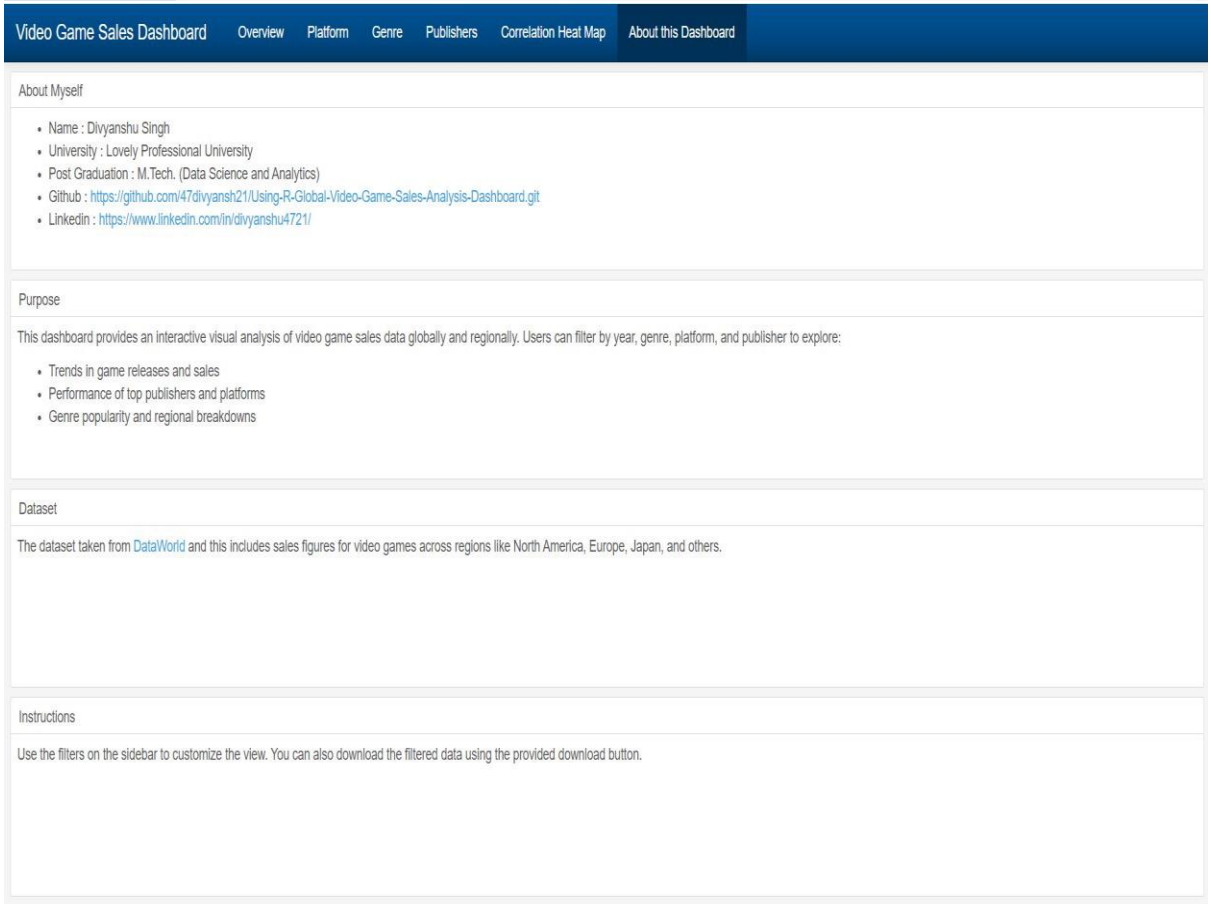





Video Game Sales Dashboard
Overview
Platform
Genre
Publishers
Correlation Heat Map
About this Dashboard
Filters
Select Genre:
All
Select Platform:
All
Select Publisher:
All
Select Region:
Global_Sales
Select Year Range
1980
2020
Reset Filters
Popularity of Game Genres by Sales
Sales by Genre (Selected Region)
Genre
Sales (Millions)
games released over the years
Genre Trends Over Time
Number of Games
Year
Genre Sales by Region
Genre Sales per Region
Sales (Millions)
Region
EU_Sales
JP_Sales
NA_Sales
Other_Sales
Number of Games by Genre (Count Plot)
Count
Genre sales trend over the year
Genre Sales Over Time
Global Sales (Millions)


Video Game Sales Dashboard
Overview
Platform
Genre
Publishers
Correlation Heat Map
About this Dashboard
Filters
Select Genre:
All
Select Platform:
All
Select Publisher:
All
Select Region:
Global_Sales
Select Year Range
1980
2020
Reset Filters
Top Publishers by Game Count (Most Game Released)
Top Publishers by Game Count
Publisher
Count
Top publisher by Region
Top 5 Publishers by Region (Including Global)
Europe
Global
Japan
North America
Other
Sales (Millions)
Top 10 Publishers by Sales
Top Publishers by Sales
Sales
Sales (Millions)
Sales Trends Over Year
Publisher Sales Trends Over Time (Region-Sensitive)
Year


Video Game Sales Dashboard
Overview
Platform
Genre
Publishers
Correlation Heat Map
About this Dashboard
Filters
Select Genre:
All
Select Platform:
All
Select Publisher:
All
Select Region:
Global_Sales
Select Year Range
1980
2020
1980
1988
1996
2004
2012
2020
Reset Filters
Correlation Heatmap
NA_Sales
EU_Sales
JP_Sales
Other_Sales
Global_Sales
1.00
0.77
0.63
0.94
1.00
0.73
0.90
1.00
0.29
0.61
1.00
0.75
1.00
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


Video Game Sales Dashboard
Overview
Platform
Genre
Publishers
Correlation Heat Map
About this Dashboard
About Myself
Name : Divyanshu Singh
University : Lovely Professional University
Post Graduation : M.Tech. (Data Science and Analytics)
Github : https://github.com/47divyansh21/Using-R-Global-Video-Game-Sales-Analysis-Dashboard.git
Linkedin : https://www.linkedin.com/in/divyanshu4721/
Purpose
This dashboard provides an interactive visual analysis of video game sales data globally and regionally. Users can filter by year, genre, platform, and publisher to explore:
Trends in game releases and sales
Performance of top publishers and platforms
Genre popularity and regional breakdowns
Dataset
The dataset taken from DataWorld and this includes sales figures for video games across regions like North America, Europe, Japan, and others.
Instructions
Use the filters on the sidebar to customize the view. You can also download the filtered data using the provided download button.

CODE

```
## Import Packages and Libraries
install.packages("corrplot")
install.packages("tidyverse")
install.packages("dplyr")
install.packages("ggplot2")
install.packages("DT")
install.packages("rmarkdown")
install.packages("shiny")
install.packages("shinydashboard")
install.packages("flexdashboard")
install.packages("tidyversa")
library(flexdashboard)
library(ggplot2)
library(tidyverse)
library(dplyr)
library(DT)
library(flexdashboard)
library(ggplot2)
library(tidyverse)
library(dplyr)
library(plotly)
library(DT)
library(corrplot)
library(shiny)
library(scales)
```

```
## Load Dataset ##

df <-
read.csv("C:/Users/d4i7v/OneDrive/Documents/Mtech/sem2/INT553DATACL
EANINGR/completeRproject/vgsales.csv")
datatable(df)
View(df)
### Data Overview ##
colnames(df)
as.list(colnames(df))

df$Name
df %>% select(Name, Platform, Year, Genre)
datatable(df %>% select(Name, Platform, Year, Genre, Publisher))

## Data Inspection ##

head(df)
head(df,10)
tail(df)
sample_n(df, 15)

str(df)
summary(df)



## DATA PREPROCESSING ##
```

```
sapply(df, n_distinct)

as.list(sapply(df, n_distinct))


duplicate_value <- sum(duplicated(df))

cat("Number of Duplicate Value:", duplicate_value, "\n")



# Missing values in each column

df <-
read_csv("C:/Users/d4i7v/OneDrive/Documents/Mtech/sem2/INT553DATACL
EANINGR/completeRproject/vgsales.csv", na = c("", "NA", "N/A", "n/a",
"null", "NULL"))

missing_values <- colSums(is.na(df))

print(missing_values)

datatable(df[!complete.cases(df), ])




datatable(df %>% filter(is.na(Year)))

datatable(df %>% filter(is.na(Publisher)))

datatable(df %>% filter(is.na(Year) | is.na(Publisher)))

datatable(df %>% filter(is.na(Year) & is.na(Publisher)))

datatable(df %>% filter(is.na(Year) & is.na(Publisher)) %>%
summarise_all(~sum(!is.na(.))))


df %>% filter(is.na(Year) | is.na(Publisher))

df %>% filter(is.na(Year) & is.na(Publisher)) %>%
summarise_all(~sum(!is.na(.)))
```

```
# Remove rows with NA in 'Year' or 'Publisher'
df <- df %>% drop_na(Year, Publisher)

# Confirm missing values removed
colSums(is.na(df))
str(df)

## Data TYPE change##

df$Year <- as.numeric(df$Year)
str(df)
head(df, 50)



#  Global Overview
cat("Years covered:", min(df$Year), "to", max(df$Year), "\n")
cat("Total unique games:", n_distinct(df$Name), "\n")
cat("Unique Platforms:", n_distinct(df$Platform), "\n")
cat("Unique Genre:", n_distinct(df$Genre), "\n")
min(df$Year, na.rm = TRUE)
max(df$Year, na.rm = TRUE)
n_distinct(df$Year)
unique(df$Year)
unique(df$Genre)

View(df)

#Top1 in each region
```

```
df %>%filter(NA_Sales == max(NA_Sales, na.rm = TRUE)) #top1
df %>%filter(EU_Sales == max(EU_Sales, na.rm = TRUE)) #top1
df %>%filter(JP_Sales == max(JP_Sales, na.rm = TRUE)) #top1
df %>%filter(Global_Sales == max(Global_Sales, na.rm = TRUE)) #top1

#Which genres and platforms are the top 10 most popular in terms of sales in each region and worldwide?
#Top 10 sales in North America
df %>% arrange(desc(NA_Sales)) %>%
select(Name,Year,Genre,Platform,NA_Sales) %>% head(10)
#Top 10 sales in Europe
df %>% arrange(desc(EU_Sales)) %>%
select(Name,Year,Genre,Platform,EU_Sales) %>% head(10)
#Top 10 sales in Japan
df %>% arrange(desc(JP_Sales)) %>%
select(Name,Year,Genre,Platform,JP_Sales) %>% head(10)
#Top 10 sales for the rest of the world
df %>% arrange(desc(Other_Sales)) %>%
select(Name,Year,Genre,Platform,Other_Sales) %>% head(10)
#Top 10 worldwide sales (Golbal sales)
df %>% arrange(desc(Global_Sales)) %>%
select(Name,Year,Genre,Platform,Global_Sales) %>% head(10)

#Game that Released to the Most Platforms
df %>% group_by(Name) %>% count() %>% arrange(desc(n)) %>% head(1)
df %>% filter(Name == "Need for Speed: Most Wanted") %>% arrange(Year)
```

#DashBoard Code

```
---
title: "Video Game Sales Dashboard"
output:
  flexdashboard::flex_dashboard:
    orientation: columns
    vertical_layout: fill
    theme: cerulean
    runtime: shiny
---

```{r setup, include=FALSE}
library(flexdashboard)
library(ggplot2)
library(tidyverse)
library(dplyr)
library(plotly)
library(DT)
library(corrplot)
library(shiny)
library(scales)
library(shinydashboard)

# Load Data
df <-
read_csv("C:/Users/d4i7v/OneDrive/Documents/Mtech/sem2/INT553DATACLEANINGR/completeRproject/vgsales.csv", na = c("", "NA", "N/A", "n/a", "null", "NULL"))
```
```

```
df <- df %>% drop_na(Year, Publisher)
df$Year <- as.numeric(df$Year)

# Predefine Regions
regions <- c("All", "NA_Sales", "EU_Sales", "JP_Sales", "Other_Sales",
"Global_Sales")

# Reactive Filtered Data
filtered_data <- reactive({
  data <- df %>%
    filter(Year >= input$year[1], Year <= input$year[2])

  if (!is.null(input$genre) && !"All" %in% input$genre) {
    data <- data %>% filter(Genre %in% input$genre)
  }
  if (!is.null(input$platform) && input$platform != "All") {
    data <- data %>% filter(Platform == input$platform)
  }
  if (!is.null(input$publisher) && !"All" %in% input$publisher) {
    data <- data %>% filter(Publisher %in% input$publisher)
  }
  if (!is.null(input$region) && !"All" %in% input$region) {
    # Only the first selected region will be used for analysis
    data <- data %>% mutate(Selected_Sales = .data[[input$region[1]]])
  } else {
    data <- data %>% mutate(Selected_Sales = Global_Sales)
  }
```

```
  data
})

observeEvent(input$reset, {
  updateSelectizeInput(session, "genre", selected = "All")
  updateSelectizeInput(session, "publisher", selected = "All")
  updateSelectizeInput(session, "region", selected = "Global_Sales")
  updateSelectInput(session, "platform", selected = "All")
  updateSliderInput(session, "year", value = c(1980, 2020))
})
```

Overview
==========

Column {.sidebar}

Filters

```{r}
selectizeInput("genre", "Select Genre:", choices = c("All", unique(df$Genre)), selected = "All", multiple = TRUE)

selectizeInput("platform", "Select Platform:", choices = c("All", unique(df$Platform)), selected = "All", multiple = TRUE)

selectizeInput("publisher", "Select Publisher:", choices = c("All", unique(df$Publisher)), selected = "All", multiple = TRUE)

selectInput("region", "Select Region:", choices = regions, selected = "Global_Sales")
```

```
sliderInput("year", "Select Year Range", min = 1980, max = 2020, value = c(1980, 2020), sep = "")

actionButton("reset", label = tags$span(style = "color:black;", "Reset Filters"), icon = icon("redo"), style = "background-color:white; border-color:#ccc;")

# Download Button

downloadButton("downloadData", "Download Filtered Data")


```


```{r}
# Outputs for Download and KPIs
output$downloadData <- downloadHandler(
  filename = function() { paste("filtered_data", ".csv", sep="") },
  content = function(file) {
    write.csv(filtered_data(), file, row.names = FALSE)
  }
)


output$total_sales <- renderValueBox({
  valueBox(
    value = paste0(round(sum(filtered_data()$Global_Sales), 2), " M"),
    subtitle = "Total Global Sales",
    icon = icon("dollar-sign"),
    color = "blue"
  )
})


output$total_games <- renderValueBox({
```

```
  valueBox(
    value = nrow(filtered_data()),
    subtitle = "Total Games",
    icon = icon("gamepad"),
    color = "purple"
  )
})

output$unique_publishers <- renderValueBox({
  valueBox(
    value = n_distinct(filtered_data()$Publisher),
    subtitle = "Unique Publishers",
    icon = icon("users"),
    color = "green"
  )
})
```

Column {data-width=650}

Dataset Overview

```{r}
renderDataTable({ datatable(filtered_data()) })
```

Summary Stats

```{r}
renderUI({
  top_genre <- filtered_data() %>%
    group_by(Genre) %>%
    summarise(Sales = sum(Selected_Sales, na.rm = TRUE)) %>%
    arrange(desc(Sales)) %>%
    slice_head(n = 1) %>%
    pull(Genre)

  top_platform <- filtered_data() %>%
    group_by(Platform) %>%
    summarise(Sales = sum(Selected_Sales, na.rm = TRUE)) %>%
    arrange(desc(Sales)) %>%
    slice_head(n = 1) %>%
    pull(Platform)

  fluidRow(
    valueBox(round(sum(filtered_data()$Selected_Sales, na.rm = TRUE), 2), paste("Total Sales (", input$region, ")"), icon = icon("dollar-sign"), color = "blue"),
    valueBox(n_distinct(filtered_data()$Name), "Unique Games", icon = icon("gamepad"), color = "green"),
    valueBox(n_distinct(filtered_data()$Publisher), "Unique Publishers", icon = icon("building"), color = "yellow"),
    valueBox(n_distinct(filtered_data()$Platform), "Unique Platforms", icon = icon("cubes"), color = "orange"),
    valueBox(round(mean(filtered_data()$Selected_Sales, na.rm = TRUE), 2), "Avg Sales per Game", icon = icon("chart-line"), color = "purple"),
    valueBox(paste("Genre:", top_genre, "/ Platform:", top_platform), "Top Seller (Region)", icon = icon("star"), color = "teal")
```

```
  )
})
```

Column {data-width=650}

Games released Per year

```{r, height=400}
renderPlotly({
  filtered_data() %>%
    group_by(Year) %>%
    summarise(Count = n()) %>%
    ggplot(aes(x = Year, y = Count)) +
    geom_line(color = "tomato", size = 0.5) +
    geom_point(size = 1, color = "tomato") +
    labs(title = "Games Released per Year", x = "Year", y = "Number of Games")
+
    theme_minimal()
})
```

Global Video Game Sales Trend Over Years

```{r, height=400}
renderPlotly({
  filtered_data() %>%
```

```
    group_by(Year) %>%
    summarise(Sales = sum(Selected_Sales, na.rm = TRUE)) %>%
    ggplot(aes(x = Year, y = Sales)) +
    geom_line(color = "darkblue", size = 0.5) +
    geom_point(size = 1, color = "darkblue") +
    labs(title = "Global Sales Trend", x = "Year", y = "Sales (Millions)") +
    theme_minimal()
})
```

Top 10 Games by Sales

```{r, height=400}
renderPlot({
  filtered_data() %>%
    arrange(desc(Selected_Sales)) %>%
    slice_head(n = 10) %>%
    ggplot(aes(x = reorder(Name, Selected_Sales), y = Selected_Sales, fill = Name)) +
    geom_col(show.legend = FALSE) +
    coord_flip() +
    scale_fill_brewer(palette = "Set3") +
    labs(title = "Top 10 Games by Sales (Selected Region)", x = "Game", y = "Sales (Millions)") +
    theme_minimal()
})
```

Platform

==========

Column {.sidebar}

Filters

```{r}

selectizeInput("genre", "Select Genre:", choices = c("All", unique(df$Genre)), selected = "All", multiple = TRUE)

selectizeInput("platform", "Select Platform:", choices = c("All", unique(df$Platform)), selected = "All", multiple = TRUE)

selectizeInput("publisher", "Select Publisher:", choices = c("All", unique(df$Publisher)), selected = "All", multiple = TRUE)

selectInput("region", "Select Region:", choices = regions, selected = "Global_Sales")

sliderInput("year", "Select Year Range", min = 1980, max = 2020, value = c(1980, 2020), sep = "")

actionButton("reset", label = tags$span(style = "color:black;", "Reset Filters"), icon = icon("redo"), style = "background-color:white; border-color:#ccc;")
```

Row

Popular Platforms by Number of Games Released

```{r, height=400}
renderPlotly({
  platform_game <- filtered_data() %>%
    group_by(Platform) %>%
    count() %>%
    arrange(desc(n)) %>%
    head(10)

  ggplot(platform_game, aes(x = n, y = reorder(Platform, n), fill = Platform)) +
    geom_col() +
    geom_text(aes(label = n), vjust = 0.5, hjust = -0.1, size = 4) +
    theme_minimal(base_size = 12, base_family = "Source Sans Pro") +
    theme(panel.grid.minor = element_blank()) +
    scale_x_continuous(limits = c(0, 2500)) +
    labs(x = "Games Released", y = NULL, title = "Platforms by Number of Games Released") +
    guides(fill = "none")
})
```

Top 5 Platforms by Sales

```{r}
renderPlotly({
  platform_sales <- filtered_data() %>%
    group_by(Platform) %>%
    summarise(Sales = sum(Selected_Sales, na.rm = TRUE)) %>%
    arrange(desc(Sales)) %>%
```

```
    slice_head(n = 5) %>%
    mutate(pct = Sales / sum(Sales) * 100,
           label = paste0(Platform, ": ", round(pct, 1), "%"))

  plot_ly(platform_sales, labels = ~Platform, values = ~Sales, type = 'pie',
          textinfo = 'label+percent',
          insidetextfont = list(color = '#FFFFFF'),
          marker = list(colors = c('#ff9999', '#66b3ff', '#99ff99', '#ffcc99',
'#c2c2f0'),
                        line = list(color = '#000000', width = 1.5)),
          showlegend = FALSE) %>%
    layout(title = paste('Top 5 Platforms by Sales (', input$region, ')'),
           margin = list(l = 20, r = 20, b = 20, t = 50))
})

```

Row

Platform Sales Trend Over Time
```{r, height=650}
renderPlotly({
  platform_trend <- filtered_data() %>%
    group_by(Year, Platform) %>%
    summarise(Sales = sum(Selected_Sales, na.rm = TRUE), .groups = "drop")

  top_platforms <- filtered_data() %>%
```

```
  group_by(Platform) %>%
  summarise(Total_Sales = sum(Selected_Sales, na.rm = TRUE)) %>%
  arrange(desc(Total_Sales)) %>%
  slice_head(n = 5) %>%
  pull(Platform)

filtered_platform_trend <- platform_trend %>%
  filter(Platform %in% top_platforms)

ggplot(filtered_platform_trend, aes(x = Year, y = Sales, color = Platform)) +
  geom_line(size = 0.5) +
  geom_point(size = 1) +
  labs(title = "Platform Sales Trends Over Time (Region-Sensitive)",
       x = "Year", y = "Sales (Millions)", color = "Platform") +
  theme_minimal()
})
```

Genre

==========

Column {.sidebar}

Filters

```{r}
selectizeInput("genre", "Select Genre:", choices = c("All", unique(df$Genre)), selected = "All", multiple = TRUE)
```

```
selectizeInput("platform", "Select Platform:", choices = c("All", unique(df$Platform)), selected = "All", multiple = TRUE)

selectizeInput("publisher", "Select Publisher:", choices = c("All", unique(df$Publisher)), selected = "All", multiple = TRUE)

selectInput("region", "Select Region:", choices = regions, selected = "Global_Sales")

sliderInput("year", "Select Year Range", min = 1980, max = 2020, value = c(1980, 2020), sep = "")

actionButton("reset", label = tags$span(style = "color:black;", "Reset Filters"), icon = icon("redo"), style = "background-color:white; border-color:#ccc;")
```

Row

Popularity of Game Genres by Sales

```{r, height=400}
renderPlotly({
  filtered_data() %>%
    group_by(Genre) %>%
    summarise(Sales = sum(Selected_Sales, na.rm = TRUE)) %>%
    ggplot(aes(x = reorder(Genre, Sales), y = Sales)) +
    geom_col(fill = "coral") +
    coord_flip() +
    labs(title = "Sales by Genre (Selected Region)", x = "Genre", y = "Sales (Millions)") +
    theme_minimal()
})
```

Number of Games by Genre (Count Plot)

```{r, height=400}
renderPlot({
  genre_count <- filtered_data() %>% count(Genre)

  ggplot(genre_count, aes(x = reorder(Genre, n), y = n, fill = n)) +
    geom_col() +
    coord_flip() +
    scale_fill_gradientn(colors = scales::gradient_n_pal(
      c("#3b4cc0", "#e0f3f8", "#fbb4ae", "#b40426"))(
      seq(0, 1, length.out = 100))) +
    labs(title = "Number of Games by Genre", x = "Genre", y = "Count") +
    theme_minimal()
})
```

Row

---

games released over the years

```{r, height=400}
renderPlot({
  filtered_data() %>%
    group_by(Year, Genre) %>%
    summarise(total = n()) %>%
    ggplot(aes(x = Year, y = total, color = Genre)) +
```

```
    geom_line(size = 0.5) +
    labs(title = "Genre Trends Over Time", x = "Year", y = "Number of Games")
})
```

Genre sales trend over the year

```{r, height=400}
renderPlot({
  filtered_data() %>%
    group_by(Genre, Year) %>%
    summarise(total_sales = sum(Global_Sales)) %>%
    ggplot(aes(x = Year, y = total_sales, fill = Genre)) +
    geom_col() +
    scale_fill_brewer(palette = "Set3") +
    theme_classic() +
    labs(title = "Genre Sales Over Time", y = "Global Sales (Millions)")
})
```

Row

Genre Sales by Region

```{r, height=400}
renderPlot({
  sale <- filtered_data() %>%
```

```
    select(Genre, NA_Sales, EU_Sales, JP_Sales, Other_Sales) %>%
    group_by(Genre) %>%
    pivot_longer(cols = c(NA_Sales, EU_Sales, JP_Sales, Other_Sales),
names_to = "Region", values_to = "Sales")


  ggplot(sale, aes(x = Genre, y = Sales, fill = Region)) +
    geom_col(position = "dodge") +
    labs(title = "Genre Sales per Region", x= "Genre", y = "Sales (Millions)") +
    theme_bw() +
    theme(axis.text.x= element_text(angle = 90)) +
    scale_fill_brewer(palette = "Set1")
})
```

Publishers

==========

Column {.sidebar}

Filters

```{r}
selectizeInput("genre", "Select Genre:", choices = c("All", unique(df$Genre)),
selected = "All", multiple = TRUE)
selectizeInput("platform", "Select Platform:", choices = c("All",
unique(df$Platform)), selected = "All", multiple = TRUE)
selectizeInput("publisher", "Select Publisher:", choices = c("All",
unique(df$Publisher)), selected = "All", multiple = TRUE)
```

```
selectInput("region", "Select Region:", choices = regions, selected = "Global_Sales")

sliderInput("year", "Select Year Range", min = 1980, max = 2020, value = c(1980, 2020), sep = "")

actionButton("reset", label = tags$span(style = "color:black;", "Reset Filters"), icon = icon("redo"), style = "background-color:white; border-color:#ccc;")
```

Row

Top Publishers by Game Count (Most Game Released)

```{r, height=400}
renderPlotly({
  filtered_data() %>%
    count(Publisher, sort = TRUE) %>%
    slice_head(n = 10) %>%
    ggplot(aes(x = reorder(Publisher, n), y = n)) +
    geom_col(fill = "darkred") +
    coord_flip() +
    labs(title = "Top Publishers by Game Count", x = "Publisher", y = "Count") +
    theme_minimal()
})
```

Top 10 Publishers by Sales

```{r, height=400}
```

```
renderPlotly({
  top_publishers_sales <- filtered_data() %>%
    group_by(Publisher) %>%
    summarise(Sales = sum(Selected_Sales, na.rm = TRUE)) %>%
    arrange(desc(Sales)) %>%
    slice(1:10)


  ggplot(top_publishers_sales, aes(x = reorder(Publisher, Sales), y = Sales, fill = Sales)) +
    geom_col() +
    scale_fill_viridis_c(option = "flare", direction = -1) +
    coord_flip() +
    labs(title = "Top Publishers by  Sales", x = "Publisher", y = "Sales (Millions)") +
    theme_minimal()
})
```

Row

Top publisher by Region

```{r, height=400}
renderPlot({
  get_top_publishers <- function(region_col, region_name) {
    df %>%
      filter(
        (Genre == input$genre | input$genre == "All"),
        (Platform == input$platform | input$platform == "All"),
```

```
      Year >= input$year[1], Year <= input$year[2]
    ) %>%
    group_by(Publisher) %>%
    summarise(Sales = sum(.data[[region_col]], na.rm = TRUE)) %>%
    arrange(desc(Sales)) %>%
    slice_head(n = 5) %>%
    mutate(Region = region_name)
}

top_na     <- get_top_publishers("NA_Sales", "North America")
top_eu     <- get_top_publishers("EU_Sales", "Europe")
top_jp     <- get_top_publishers("JP_Sales", "Japan")
top_other  <- get_top_publishers("Other_Sales", "Other")
top_global <- get_top_publishers("Global_Sales", "Global")

top_publishers_all <- bind_rows(top_na, top_eu, top_jp, top_other,
top_global)

ggplot(top_publishers_all, aes(x = reorder(Publisher, Sales), y = Sales, fill =
Publisher)) +
  geom_col(show.legend = FALSE) +
  coord_flip() +
  facet_wrap(~ Region, scales = "free_y") +
  labs(title = "Top 5 Publishers by Region (Including Global)",
       x = "Publisher", y = "Sales (Millions)") +
  theme_minimal() +
  theme(strip.text = element_text(face = "bold", size = 12))
})
```

```

### Sales Trends Over Year

```{r}
renderPlot({
  publisher_trend <- filtered_data() %>%
    group_by(Year, Publisher) %>%
    summarise(Sales = sum(Selected_Sales, na.rm = TRUE), .groups = "drop")

  top_publishers <- filtered_data() %>%
    group_by(Publisher) %>%
    summarise(Total_Sales = sum(Selected_Sales, na.rm = TRUE)) %>%
    arrange(desc(Total_Sales)) %>%
    slice_head(n = 5) %>%
    pull(Publisher)

  filtered_trend <- publisher_trend %>%
    filter(Publisher %in% top_publishers)

  ggplot(filtered_trend, aes(x = Year, y = Sales, color = Publisher)) +
    geom_line(size = 0.5) +
    geom_point(size = 0.5) +
    labs(title = "Publisher Sales Trends Over Time (Region-Sensitive)",
         x = "Year", y = "Sales (Millions)", color = "Publisher") +
    theme_minimal()
})
```
```

Correlation Heat Map

==========

Column {.sidebar}

Filters

```{r}
selectizeInput("genre", "Select Genre:", choices = c("All", unique(df$Genre)), selected = "All", multiple = TRUE)

selectizeInput("platform", "Select Platform:", choices = c("All", unique(df$Platform)), selected = "All", multiple = TRUE)

selectizeInput("publisher", "Select Publisher:", choices = c("All", unique(df$Publisher)), selected = "All", multiple = TRUE)

selectInput("region", "Select Region:", choices = regions, selected = "Global_Sales")

sliderInput("year", "Select Year Range", min = 1980, max = 2020, value = c(1980, 2020), sep = "")

actionButton("reset", label = tags$span(style = "color:black;", "Reset Filters"), icon = icon("redo"), style = "background-color:white; border-color:#ccc;")
```

Row

Correlation Heatmap

```{r, height=400}
renderPlot({
```

```
  sales_data <- filtered_data() %>%
    select(NA_Sales, EU_Sales, JP_Sales, Other_Sales, Global_Sales)
  cor_matrix <- cor(sales_data, use = "complete.obs")
  corrplot(cor_matrix, method = "color", type = "upper", addCoef.col = "green")
})
```

About this Dashboard

==========

Column

About Myself

- Name : Divyanshu Singh

- University : Lovely Professional University

- Post Graduation : M.Tech. (Data Science and Analytics)

- Github : https://github.com/47divyansh21/Using-R-Global-Video-Game-Sales-Analysis-Dashboard.git

- Linkedin : https://www.linkedin.com/in/divyanshu4721/

Purpose

This dashboard provides an interactive visual analysis of video game sales data globally and regionally. Users can filter by year, genre, platform, and publisher to explore:

- Trends in game releases and sales

- Performance of top publishers and platforms

- Genre popularity and regional breakdowns

Dataset

The dataset taken from [DataWorld](https://data.world/julienf/video-games-global-sales-in-volume-1983-2017) and this includes sales figures for video games across regions like North America, Europe, Japan, and others.

Instructions

Use the filters on the sidebar to customize the view. You can also download the filtered data using the provided download button.

CHAPTER 7

CONCLUSION AND FUTURE SCOPE

1. CONCLUSION

This analysis provided an in-depth, data-driven exploration of the global video game market using a diverse range of visualizations and analytical techniques. By leveraging the VGChartz dataset, we identified critical trends, patterns, and insights across multiple dimensions including genre, platform, region, and publisher.

The data revealed that:

- **Action and Sports** are the most popular genres by both release count and global sales.
- **PlayStation 2**, **Nintendo DS**, and **Wii** stand out as the most prolific and top-selling platforms.
- **Nintendo**, **Electronic Arts**, and **Activision** consistently rank as the highest-grossing and most active publishers.
- **Regional market preferences** vary significantly, with Japan favoring Role-Playing games, while North America and Europe show strong affinity for Action and Sports genres.
- **Sales correlations** confirm that North America and Europe contribute significantly to global sales, while Japan exhibits unique market behavior.

The integration of interactive filters and dynamic plots made it possible to drill down into specific subsets of data, enhancing interpretability and encouraging user exploration. Each visualization was designed to not only present numerical data but also narrate meaningful market stories.

In summary, this project offers a comprehensive overview of the historical landscape of video game sales. It serves as a valuable reference for stakeholders aiming to understand commercial patterns and regional dynamics in the gaming industry.

2. FUTURE SCOPE

While the current dashboard offers valuable descriptive insights, there is significant potential for further enhancement and application of predictive analytics:

- **Incorporating Digital Sales**: Since this dataset mostly tracks physical sales, future updates could integrate data from digital distribution platforms like Steam, Xbox Live, and PlayStation Network for a more holistic analysis.
- **Real-Time Dashboarding**: By linking to live APIs or updated databases, the dashboard can be extended into a real-time performance monitor for publishers and genres.
- **Forecasting and Trend Prediction**: Machine learning models such as time series forecasting (ARIMA, Prophet) could be applied to predict future sales by genre, region, or publisher.

- **Sentiment and Review Integration**: Combining sales data with user reviews or sentiment analysis from Metacritic, IGN, or Reddit could yield insights into how public perception impacts commercial success.
- **Platform Lifecycle Analysis**: A deep dive into platform release and retirement cycles could offer strategic insights for hardware manufacturers and developers.
- **Regional Focus Dashboards**: Creating region-specific dashboards (e.g., for Japan, NA, EU) would help local stakeholders understand territory-specific trends.
- **Cloud Gaming and Subscription Trends**: Future work could explore the growing market of Game Pass, PS Plus, and cloud-based gaming platforms and their impact on traditional sales models.

These directions not only extend the value of the current analysis but also align it with emerging shifts in the gaming industry and consumer behavior.

CHAPTER 8

References

- Flexdashboard - https://rmarkdown.rstudio.com/flexdashboard/
- Plotly for R - https://plotly.com/r/
- Corrplot package documentation - https://cran.r-project.org/web/packages/corrplot/
- ggplot2: Elegant Graphics for Data Analysis - https://ggplot2.tidyverse.org/
- https://data.world/julienf/video-games-global-sales-in-volume-1983-2017