Data Science Tools Workshop

CDCSC19

Project 2 – Data Analysis and Visualization using R

Ritesh Wadhwani (2022UCD2149)

Saharsh (2022UCD2109)

Aakash Arora (2022UCD2118)

Visualization of video game sales data using R

The project aims to develop a robust system for the visualization of video game sales data. The system is designed to operate efficiently on R, a programming language and free software environment for statistical computing and graphics. The dataset used for visualization consists of video game sales data.

Introduction

The project is designed to address the growing demand for data visualization in the video game industry. It aims to create an intelligent tool that can analyse and visualize video game sales data. This tool is particularly beneficial for game developers, marketers, and strategists who need to understand sales trends and patterns.

The project is designed to operate efficiently on R, a popular platform for data analysis and visualization. R's extensive package ecosystem and data handling capabilities make it an ideal choice for this project.

Objectives

● Develop a model capable of visualizing various aspects of video game sales data.

● Provide insights into sales trends and patterns.

● Implement the model in R for efficient and interactive visualizations.

Methodology

1) Data collection: A diverse dataset of video game sales was curated, containing sales data for many games and genres across multiple genres and regions.

2) Data cleaning and preprocessing: Data was cleaned and pre-processed for better and clearer analysis and visualization, using tinyverse, dplyr and tidyr.

3) Data analysis and visualization: Data was analysed and visualized using ggplot2, RColorBrewer, etc.

About the dataset

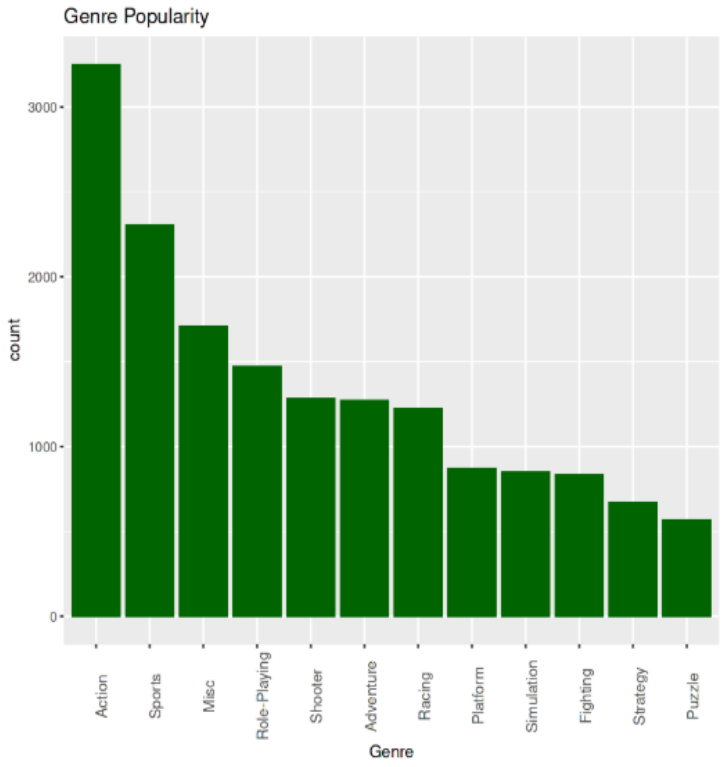
This dataset contains a list of video games with sales greater than 100,000 copies. It was generated by a scrape of vgchartz.com.

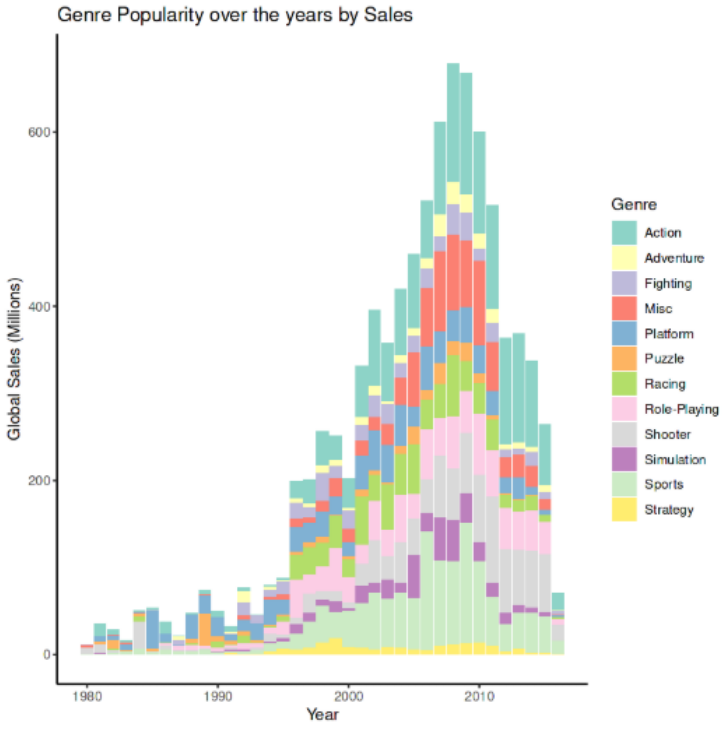
Fields include

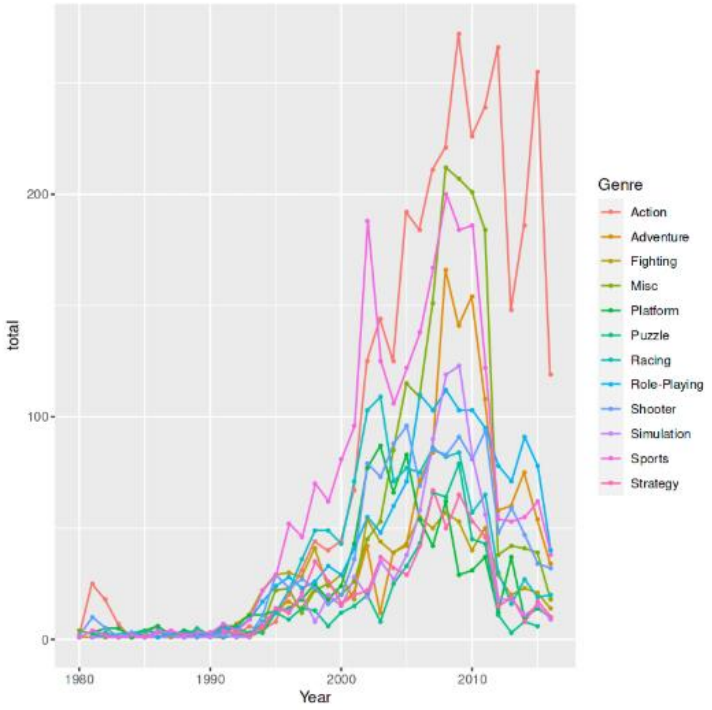
- Rank - Ranking of overall sales
- Name - The games name
- Platform - Platform of the games release (i.e. PC,PS4, etc.)
- Year - Year of the game's release
- Genre - Genre of the game
- Publisher - Publisher of the game
- NA_Sales - Sales in North America (in millions)
- EU_Sales - Sales in Europe (in millions)
- JP_Sales - Sales in Japan (in millions)
- Other_Sales - Sales in the rest of the world (in millions)
- Global_Sales - Total worldwide sales.

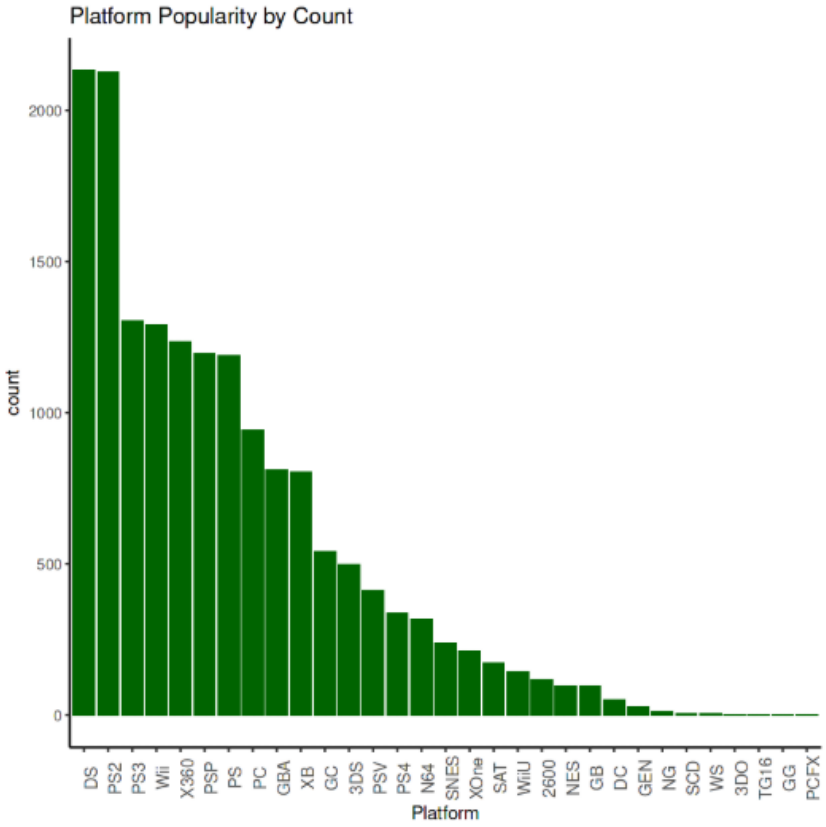
The script to scrape the data is available at https://github.com/GregorUT/vgchartzScrape.
It is based on BeautifulSoup using Python.
There are 16,598 records.

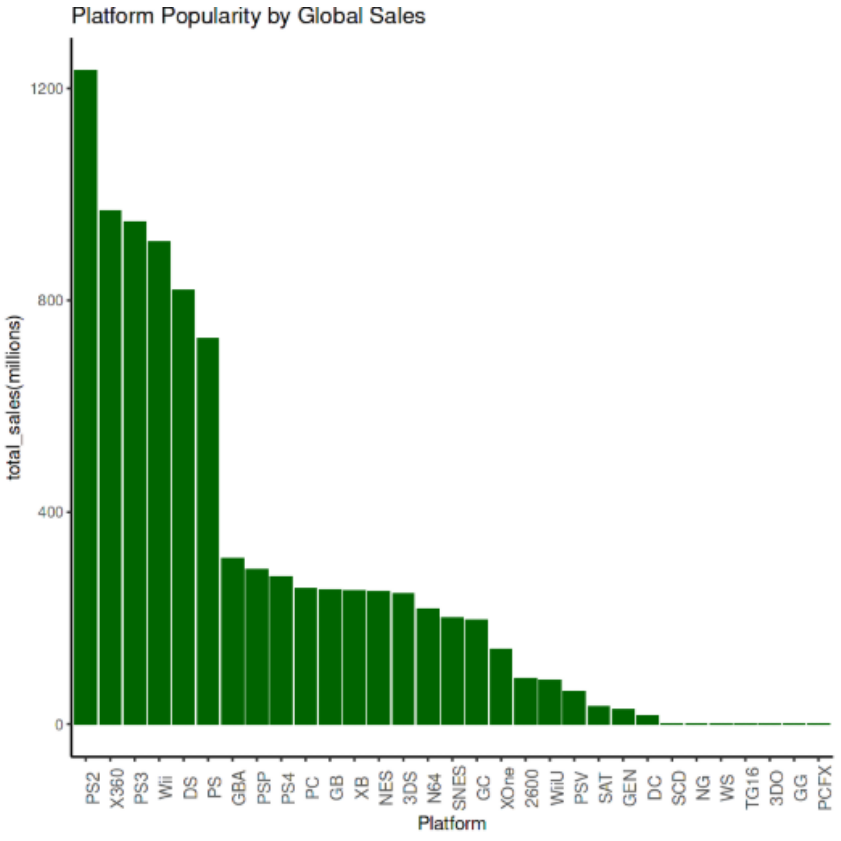
Results of visualization

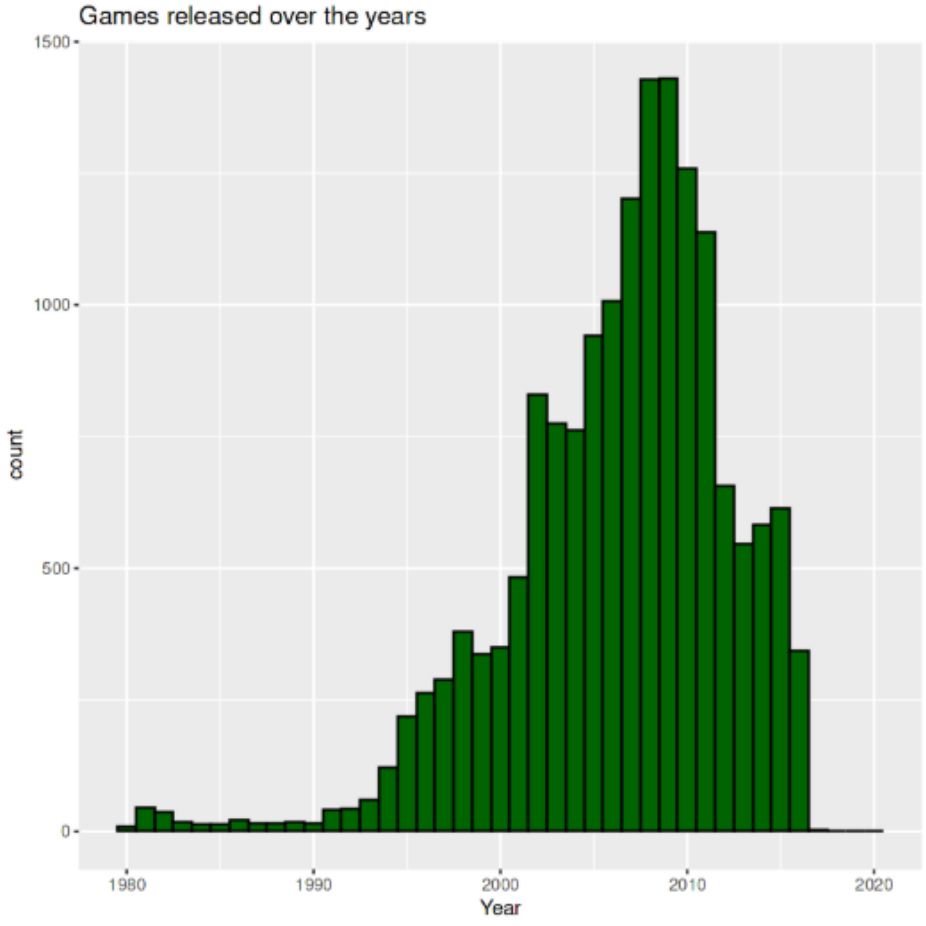


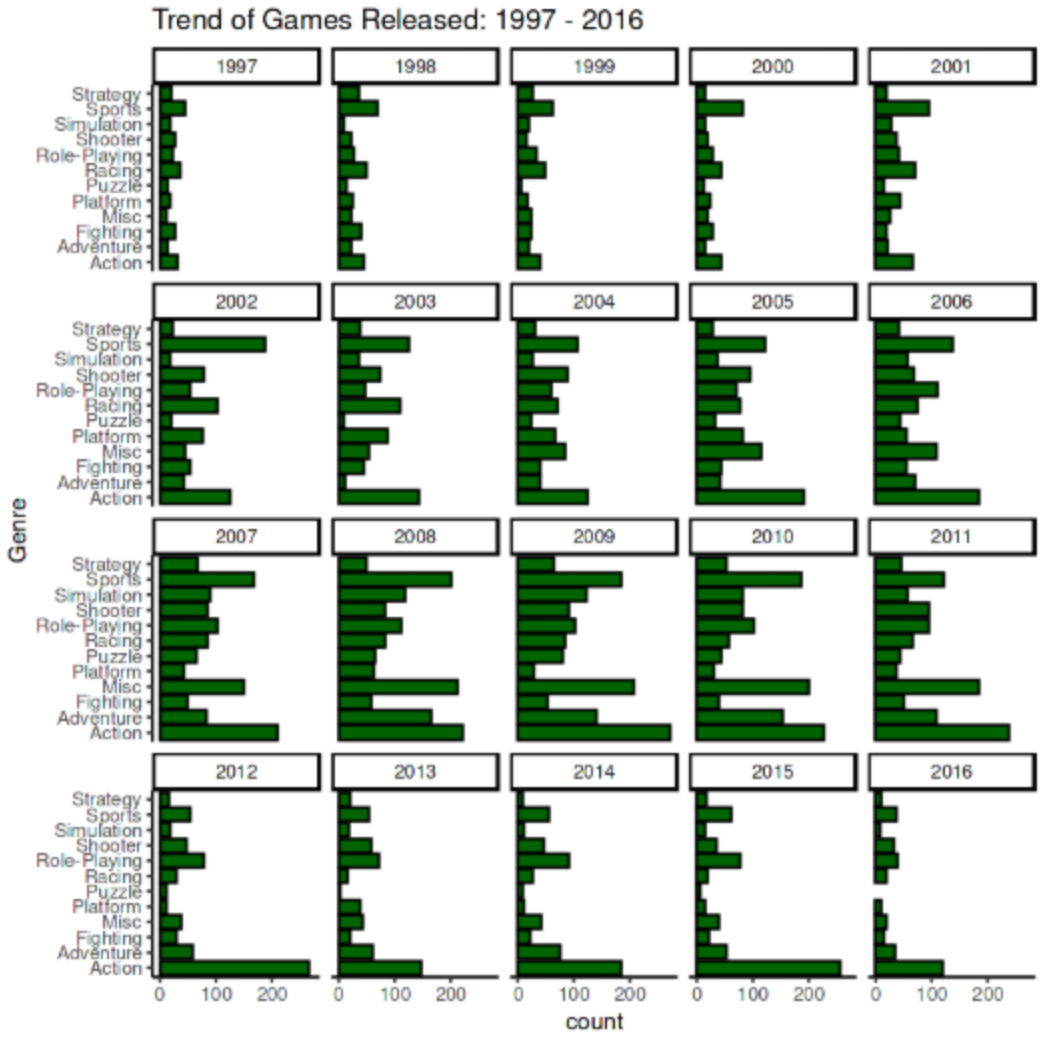



total
200
100
0
1980
1990
2000
2010
Year
Genre
Action
Adventure
Fighting
Misc
Platform
Puzzle
Racing
Role-Playing
Shooter
Simulation
Sports
Strategy


Platform Popularity by Count
count
2000
1500
1000
500
0
DS
PS2
PS3
Wii
X360
PSP
PS
PC
GBA
XB
GC
3DS
PSV
PS4
N64
SNES
XOne
SAT
WiiU
2600
NES
GB
DC
GEN
NG
SCD
WS
3DO
TG16
GG
PCFX
Platform


Platform Popularity by Global Sales
total_sales(millions)
1200
800
400
0
PS2
X360
PS3
Wii
DS
PS
GBA
PSP
PS4
PC
GB
XB
NES
3DS
N64
SNES
GC
XOne
2600
WiiU
PSV
SAT
GEN
DC
SCD
NG
WS
TG16
3DO
GG
PCFX
Platform


Games released over the years
count
1500
1000
500
0
1980
1990
2000
2010
2020
Year



Many more visualization results are there, and can be found at the following Google Colab link:

Video game sales analysis using R

The model demonstrated high efficiency in visualizing various aspects of video game sales data. The visualizations provided valuable insights into sales trends and patterns, aiding decision-making in the video game industry.

Challenges and solutions:

Complex and unclean Data: Handling and visualizing complex and unclean sales data posed challenges, addressed by leveraging the capabilities of R and its various packages.

Diverse Game Genres: The vast array of video game genres required careful curation of the dataset to ensure the model's ability to visualize a broad spectrum of games.

Conclusion:

The successful implementation of the project showcases the potential of using R for data visualization. The efficiency and interactivity of the visualizations open avenues for data-driven decision-making in the video game industry.