Homework 2: Video Game Sales Data and Ratings Exploration

Link to dataset: https://www.kaggle.com/datasets/thedevastator/video-game-sales-and-ratings?select=Video_Games.csv

Link to code: https://github.com/sid-valecha/stat436-hw2-shiny/blob/main/app.R

Introduction

This project uses Shiny to explore video game sales data, focusing on game titles, platforms, release years, genres, and critic/user scores. The goal is to provide users with an interactive way to analyze trends in video game sales across regions and compare critic and user ratings. The design offers flexibility for users to filter and visualize data dynamically, ensuring a high level of customization for data exploration.

Interesting Findings

- **Global Sales Trends**: Video game sales show a rise starting in the mid-1990s, peaking around 2010. The industry saw steady growth in the 1980s and early 1990s, followed by a rapid growth in the mid-2000s, indicating the height of gaming popularity during that period.
- **Sales by Region**: North America consistently leads in sales, followed by Europe, Japan, and other regions. This indicates that North America is the most dominant market, though Japan remains influential in certain genres, particularly role-playing games.
- **Critic vs User Scores**: The relationship between critic and user scores shows a positive correlation, but is not perfectly linear. The regression line suggests that higher critic scores tend to align with higher user scores, however, there is a wide spread in user ratings, suggesting that users may be more divided in their opinions than critics. For instance, some games with lower critic scores still received relatively high user scores, especially in niche genres. This suggests that user perception doesn't always align with professional critic reviews**.**
- **Critic Scores**: The box plots for critic scores across different genres revealed that critic ratings tend to be relatively consistent across genres, with most games falling within a similar score range. There are fewer outliers and variability in critic scores, indicating that critics generally agree on the quality of games across different genres.

Interface Creation

The interface was built to allow users to:

- **Multi-Select Dropdown for Genre**: Users can filter by one or more genres, allowing for focused exploration of specific genres or broader trends across multiple categories (or not pick any, which selects all genres by default).
- **Slider for Year Selection**: Users can specify a range of years to analyze, making it easy to examine trends over specific time periods.
- **Dynamic Visualizations**: The app includes multiple visualizations, including sales trends, regional sales distribution, critic score distributions, and critic vs. user score comparisons. The dynamic nature of the app allows users to customize the data view based on their selections, with real-time updates to the plots..

Data Preparation:

- **Removed rows** where genre was missing
- **Missing values** for Critic_Score and User_Score were removed.
- **Non-numeric values** such as "tbd" in the User_Score and Critic_Score fields were handled by converting them to NA and filtering them out.

Interface and Design Choices:

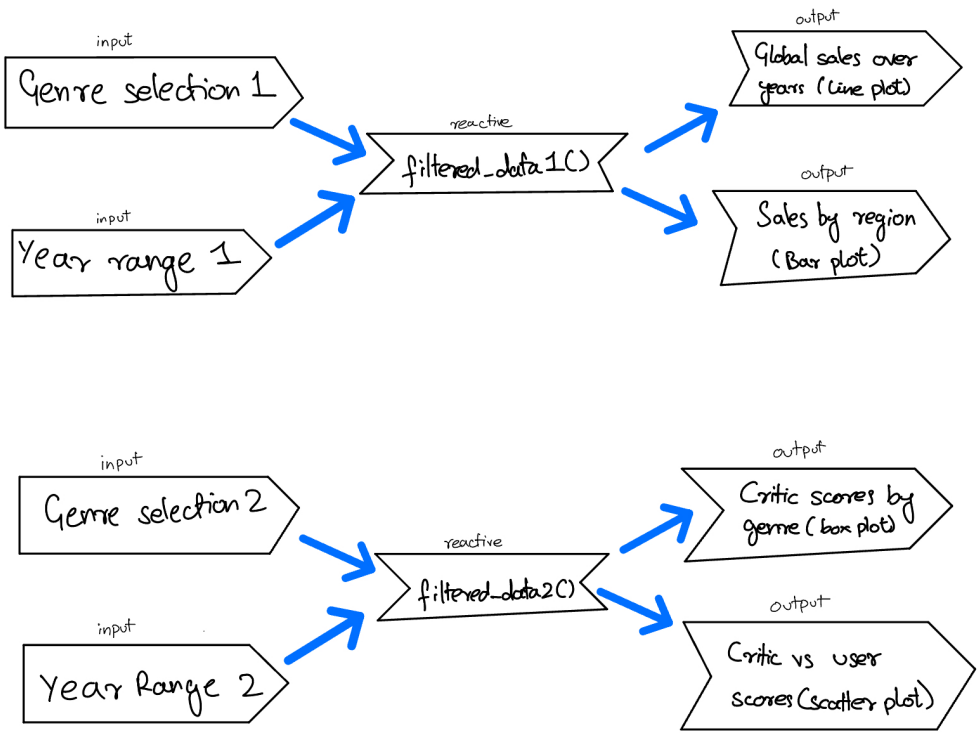
The app provides four key visualizations:

1. **Global Sales Over Years**: A line plot that tracks global sales trends for selected genres over time. This allows users to see how game sales evolved across different time periods and identify peak periods for the gaming industry.
2. **Sales by Region**: A bar chart that breaks down total sales by region—North America, Europe, Japan, and Others. This chart shows that North America and Europe dominate video game sales, with Japan following, especially in role-playing genres.
3. **Critic Scores by Genre**: A box plot that shows the distribution of critic scores across genres. This highlights the consistency of critic scores, revealing that most genres receive similar ratings with few major deviations.
4. **Critic vs User Score**: A scatter plot comparing critic and user scores. This plot reveals a positive but non-linear relationship between critic and user ratings, where critics tend to be more consistent, while user ratings show more variability.

Reactive Graph Structure:

The app uses **reactive expressions** to dynamically update the visualizations based on user input. When users adjust the genre selection(s) or change the year range, the underlying data is re-filtered, and the visualizations are regenerated in real time. This reactive structure ensures the app remains responsive and interactive.

The **sales trends**, **critic score distributions**, and **critic vs. user score comparisons** are all dependent on the user's input, making the app super customizable for different analysis needs.



Conclusion

This project provided me with a great opportunity to explore Shiny's capabilities for data visualization. The app's flexible interface allows users to easily interact with the data and extract meaningful insights about video game sales trends, regional preferences, and the relationship between critic and user scores. The results of the visualizations revealed important trends in the video game industry, in terms of sales as well as ratings.