

Stat 315, Final Report: Video Game Sales

36-315 Statistical Graphics and Visualization

Summer 2020

Carnegie Mellon University

Julia Benson, Caleb Felton, Raymond Li, Ruiwen Zhang

Introduction

Problem and Motivation The video game industry is complex and produces *lots* of valuable data.

But how do we best use that data to *understand* the industry and predict which game will be the next big thing?

Interesting Relationships

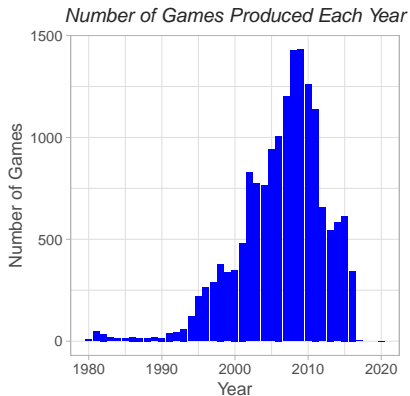
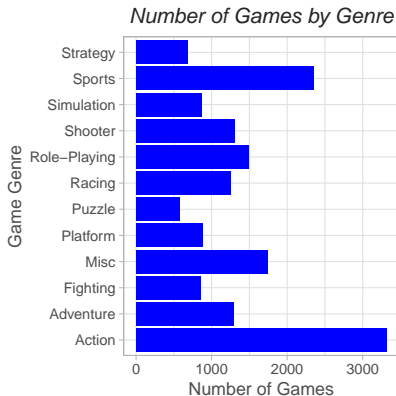
- ▶ Sales Across Video Game Genres (i.e. Action, Shooter, Puzzle, etc.)
- ▶ Sales Across Geographic Regions (i.e. North America, Japan, Europe, etc.)

The Dataset: Our Variables

Variable	Description
Rank	Rank of Overall Sales
Name	Video Game's Name
Platform	Platform Game Was Released On
Year	Year Released
Genre	Game's Genre
Publisher	Game's Publisher
NA_Sales	Sales in North America (in Millions of Copies)
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	Global Sales (in Millions)

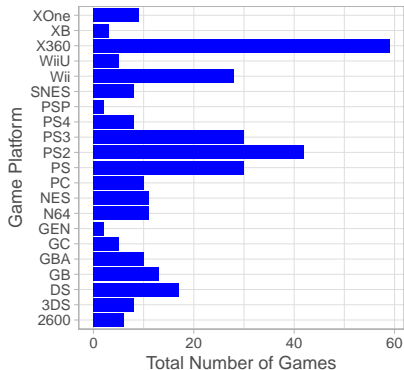
Introduction to Our Variables

Our dataset comes from Kaggle and includes 317 distinct video games that each sold over 2 million copies in North America.

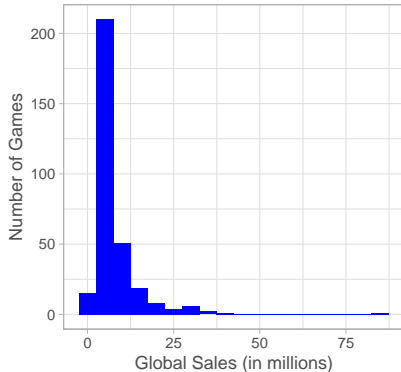


Introduction to Our Variables

Video Games by Platform



Global Sales by Game



Methods

There are several fascinating questions we seek to address with this data.

Our Hypotheses

Hypothesis 1: We hypothesize that video games that sell well in Japan also sell well in North America and Europe.

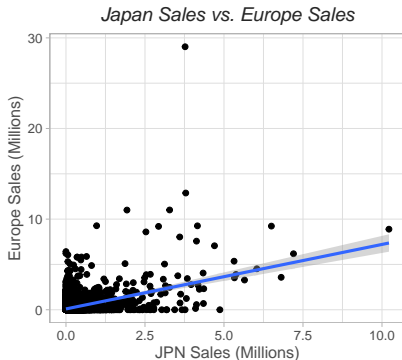
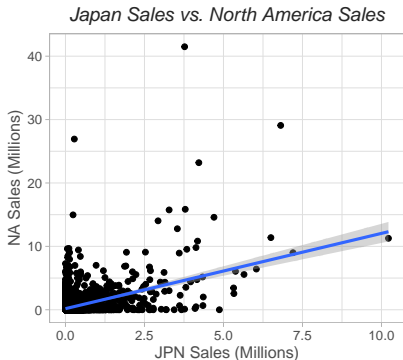
Hypothesis 2: We predict that action, shooting, and role-playing games have the highest sales.

Hypothesis 3: Action and role-playing games produce higher sales over time.

Hypothesis 4: We hypothesize that sales in Japan are the most predictive of worldwide sales.

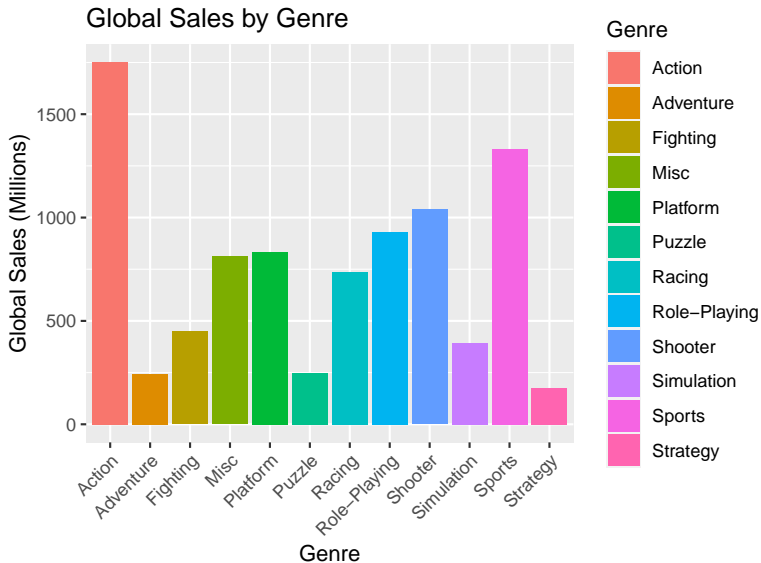
Hypothesis 1

We hypothesize that video games that sell well in Japan also sell well in North America and Europe.



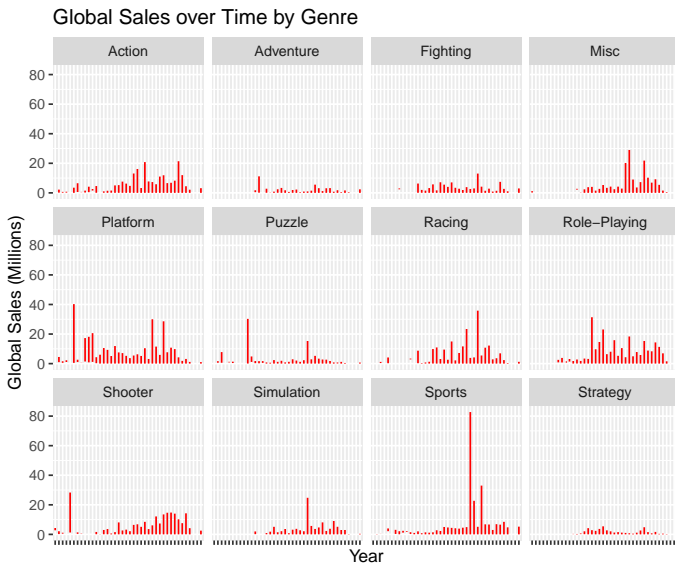
Hypothesis 2

We predict that action, shooting, and role-playing games have the highest sales.



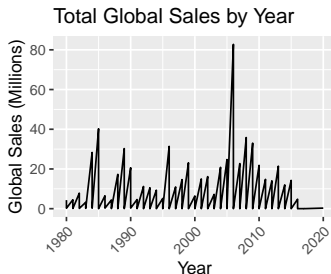
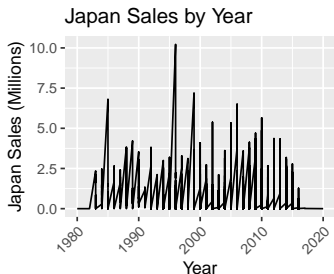
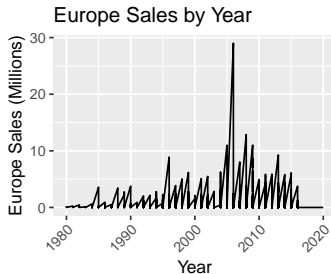
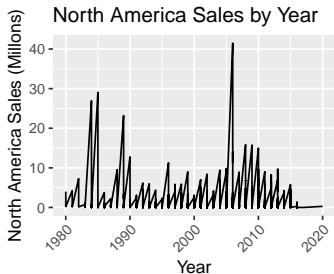
Hypothesis 3

Action and role-playing games produce higher sales over time.



Hypothesis 4

We hypothesize that sales in Japan are the most predictive of worldwide sales.



Hypothesis 4

We hypothesize that sales in Japan are the most predictive of worldwide sales.

	NA_Sales	EU_Sales	JP_Sales	Global_Sales
NA_Sales	1.00	0.77	0.45	0.94
EU_Sales	0.77	1.00	0.44	0.90
JP_Sales	0.45	0.44	1.00	0.61
Global_Sales	0.94	0.90	0.61	1.00

Results

- ▶ Two of our predictions were correct, while 2 were wrong.
- ▶ Video games sales in Japan were correlated with sales in NA and EU
 - ▶ NA market is the best predictor for the global market
- ▶ Action, Sports, Shooter, and Role Playing were the best sold genres
- ▶ Popularity of game genres occur cyclically, rather than consistently increasing or decreasing

Discussion

Limitations of Data

- ▶ Is data complete?
- ▶ Credibility issues with data source (Kaggle)
- ▶ Most games were produced between 1990 and 2010
- ▶ How do other variables influence our analyses?

Future Work

- ▶ Relationship between sales in different regions over time
- ▶ Broader changes in video game industry
- ▶ How does competition work in different video game genres?
- ▶ Will the increase of sales of shooter games result in the decrease of puzzle games?