Data Exploration of Video Game Sales

Author: Colin Beran

Dataset source: <https://www.kaggle.com/gregorut/videogamesales>

Kaggle notebook: <https://www.kaggle.com/colinberan/data-exploration-of-video-game-sales>

This is an exploratory data analysis of video game titles with over 100,000 total sales that I retrieved from the above listed source.

I began by loading a few useful libraries and importing the dataset. As a preprocessing step, I removed all rows with N/A values to avoid working with any incomplete data.

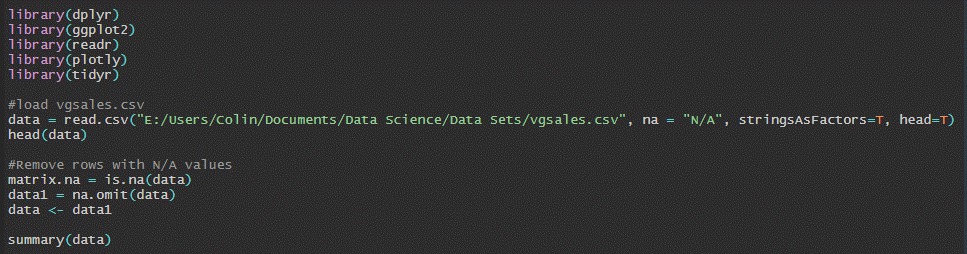


Fig. 1: R code for importing data and slight preprocessing.

I grouped all video games by the release year and noticed incomplete data for 2017 and 2020. Additionally, I wanted to focus on games released in the 21st century, so I created a subset of the data, filtering “Year” for 2000-2016.

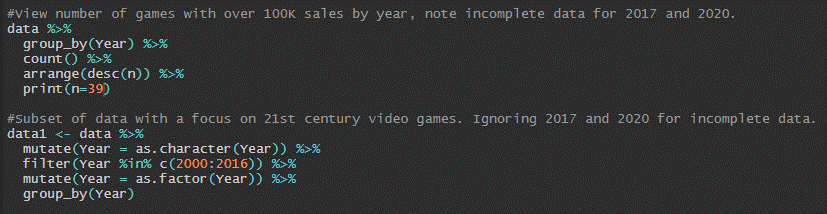
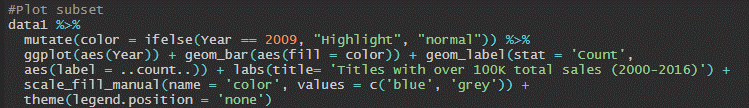


Fig. 2: R code for subset of data filtered to 2000-2016.

I then plotted this subset visually, highlighting 2009 as it was the year with the highest count of video games sold.

Fig. 3: R code plotting 2000-2016 subset by year.

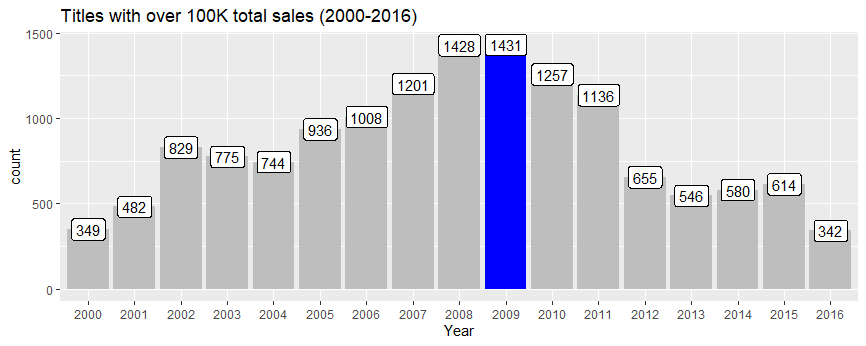


Fig. 4: Plot of video games with over 100,000 total sales by year.

Note that there is likely incomplete data for 2016 as well.

I created a second subset using the data from 2009 and created a plot to highlight the top performing platforms at the time.

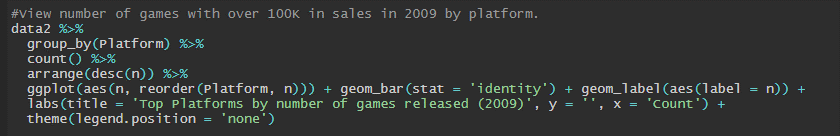


Fig. 5: R code plotting top platforms in 2009.

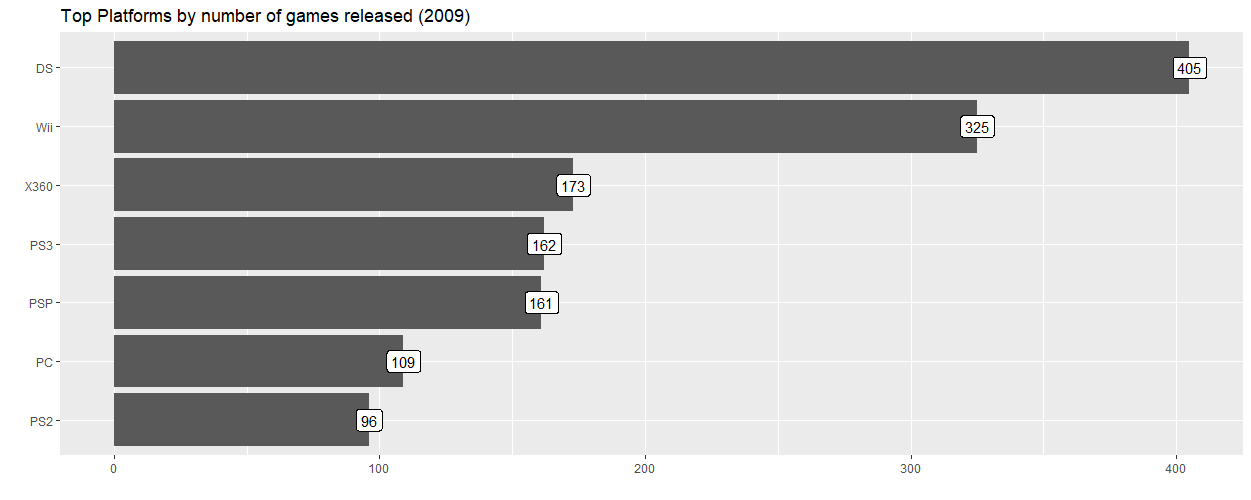


Fig. 6: Plot of top platforms in 2009 by number of games released with over 100,000 total sales.

To explore the data further, I used these two subsets to also create a plot to highlight the top 20 video game publishers during 2000-2016 and the highlighted year 2009. Note that the “head” function can be changed to alter the number of top publishers highlighted.

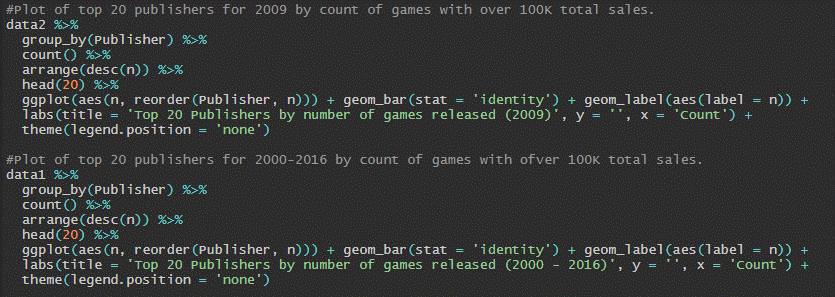


Fig. 7: R code plotting top 20 publishers for 2009 as well as the 2000-2016 range.

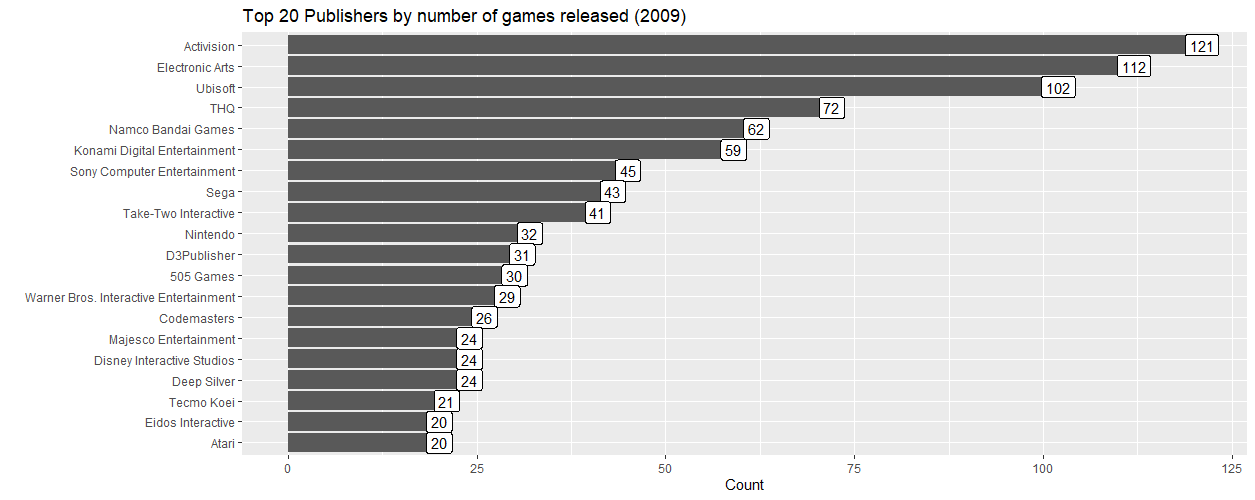


Fig. 8: Plot of top 20 publishers for 2009 by number of games with over 100,000 total sales.

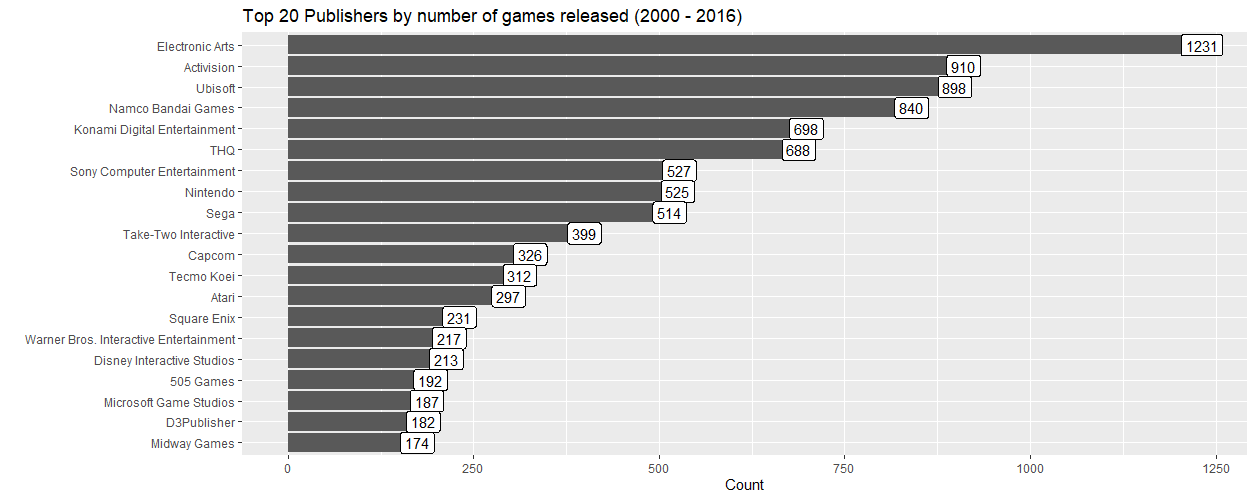


Fig. 9: Plot of top 20 publishers for 2000-2016 by number of games with over 100,000 total sales.