

Video Game Sales by Robert Rogness

I will be going over a dataset of video game sales. I downloaded the data from kaggle (<https://www.kaggle.com/gregorut/videogamesales>). The data was last updated on October 26, 2016.

Univariate Plots Section

```
## 'data.frame': 16598 obs. of 11 variables:  
## $ Rank      : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ Name       : Factor w/ 11493 levels "'98 Koshien",...: 10991 9343 5532 10993 7370 9707 6648 10989  
## $ Platform   : Factor w/ 31 levels "2600","3DO","3DS",...: 26 12 26 26 6 6 5 26 26 12 ...  
## $ Year       : Factor w/ 40 levels "1980","1981",...: 27 6 29 30 17 10 27 27 30 5 ...  
## $ Genre      : Factor w/ 12 levels "Action","Adventure",...: 11 5 7 11 8 6 5 4 5 9 ...  
## $ Publisher  : Factor w/ 579 levels "10TACLE Studios",...: 369 369 369 369 369 369 369 369 369 369 ...  
## $ 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 ...
```

Right off the bat I see a problem with this dataset. The dataset contains 16,598 games. A quick Google search finds (as of Feb 1, 2017: www.neogaf.com/threads/number-of-games-for-each-console.1339962/) that there has been over 60,000 games released. Now, the accuracy of a number pulled off of a web forum, which in turn was pulled off of Wikipedia (according to the forum poster), could easily be incorrect. However, 60,000 is a large jump from 16,598. Maybe this dataset does not include all platforms and regions.

```
## 2600 3DO 3DS DC DS GB GBA GC GEN GG N64 NES NG PC PCFX  
## 133 3 509 52 2163 98 822 556 27 1 319 98 12 960 1  
## PS PS2 PS3 PS4 PSP PSV SAT SCD SNES TG16 Wii WiiU WS X360 XB  
## 1196 2161 1329 336 1213 413 173 6 239 2 1325 143 6 1265 824  
## XOne  
## 213
```

First thing I notice is that there are two platforms that look like years (the 2600 is just the Atari 2600.). I think I have some data to clean up.

A quick search online shows these numbers are not correct. For example, on Steam (a digital distribution platform for PC games), on October 26, 2016, they had 10,079 PC games (I got this number by going to steams search page. Then I clicked on the ‘Games’ type under the ‘Show selected types’ filter category. Then I sorted by ‘Release date’. Then I went to the page where ‘Oct 26, 2016’ started, which was page 605 for me. Each page lists 25 games. There was a total of 1,009 pages. So I subtracted 606 from 1,008 and multiplied the result by 25. I then added the number of games on page 1,009 and the number of games released in October 26, 2016 on page 605.). The dataset’s 960 PC titles number is far short of the over 10,000 games on Steam, which does not contain all PC games ever sold.

Another easier to confirm example is the SNES (Super Nintendo Entertainment System). According to Wikipedia (en.wikipedia.org/wiki/List_of_Super_Nintendo_Entertainment_System_games), there were ‘1757 official releases, 721 were released in North America, 517 in Europe, 1,447 in Japan’. Any of those numbers is much larger than the dataset’s 239. Maybe the dataset is limited by some arbitrary year.

```
## [1] "1980" "1981" "1982" "1983" "1984" "1985" "1986" "1987" "1988" "1989"  
## [11] "1990" "1991" "1992" "1993" "1994" "1995" "1996" "1997" "1998" "1999"  
## [21] "2000" "2001" "2002" "2003" "2004" "2005" "2006" "2007" "2008" "2009"  
## [31] "2010" "2011" "2012" "2013" "2014" "2015" "2016" "2017" "2020" "N/A"
```

So there does seem to be an arbitrary date limit (1980 to 2020?). The first game console, the Magnavox Odyssey, was released in 1972. I don’t know if this dataset is a true sample, but I will treat it as such, since

it is obviously not a population.

There also seems to be some problems with the dataset. This dataset was made in 2016, yet there are games from 2017 and 2020! I also think there are two rows with values shifted to the left. I would like to take a look at the full rows for these strange values.

```
##      Rank                               Name Platform Year
## 5958 5959           Imagine: Makeup Artist       DS 2020
## 14391 14393 Phantasy Star Online 2 Episode 4: Deluxe Package     PS4 2017
## 16242 16244 Phantasy Star Online 2 Episode 4: Deluxe Package     PSV 2017
## 16439 16441 Brothers Conflict: Precious Baby     PSV 2017
##          Genre   Publisher NA_Sales EU_Sales JP_Sales Other_Sales
## 5958   Simulation   Ubisoft    0.27      0    0.00    0.02
## 14391 Role-Playing     Sega     0.00      0    0.03    0.00
## 16242 Role-Playing     Sega     0.00      0    0.01    0.00
## 16439 Action Idea Factory    0.00      0    0.01    0.00
##      Global_Sales
## 5958        0.29
## 14391        0.03
## 16242        0.01
## 16439        0.01
```

Like I thought. There seems to be two rows where the ‘Platform’ data merged with the ‘Name’ data. This seems to have caused everything to the right to shift left.

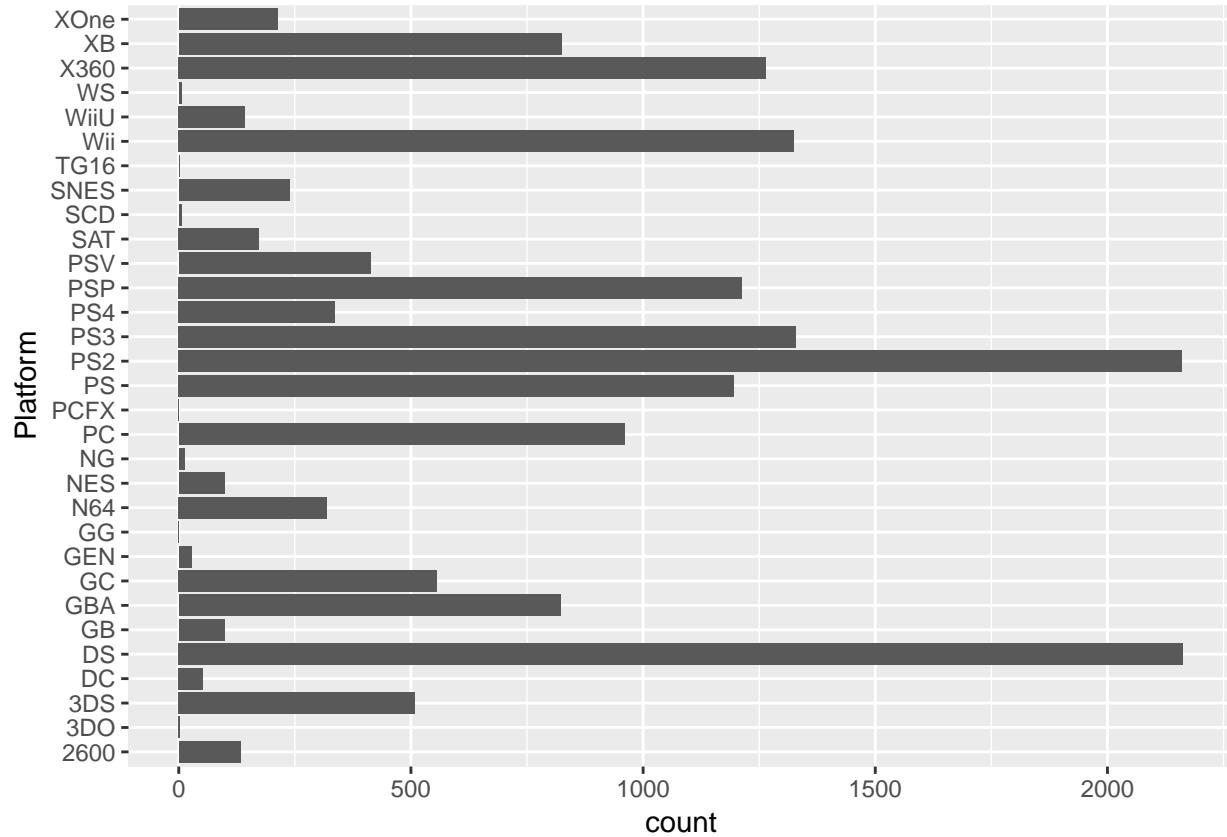
Looking at the sales numbers, the 2017 game years seem to be the planned release date for the US market. On October 26, 2016, those three games seemed to have been released in Japan, but not yet outside Japan.

The 2020 game seems to be a mistake. I Googled the title and it seems to have been released in 2009.

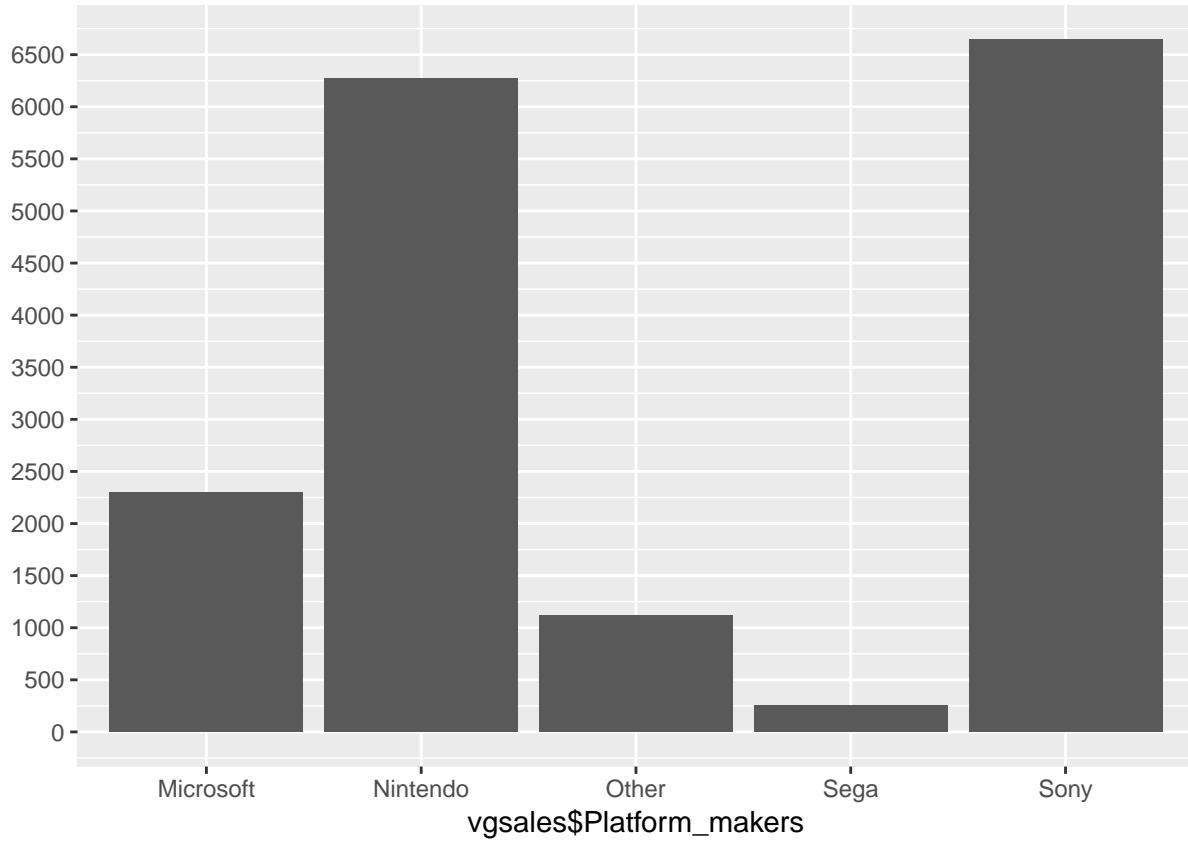
So I am going to have to fix the 2020 year and fix the two broken rows.

```
## 'data.frame': 16598 obs. of 11 variables:
## $ Rank      : int 1 2 3 4 5 6 7 8 9 10 ...
## $ Name       : Factor w/ 11493 levels "'98 Koshien",...: 10991 9343 5532 10993 7370 9707 6648 10989
## $ Platform   : Factor w/ 31 levels "2600","3DO","3DS",...: 26 12 26 26 6 6 5 26 26 12 ...
## $ Year       : int 2006 1985 2008 2009 1996 1989 2006 2006 2009 1984 ...
## $ Genre      : Factor w/ 12 levels "Action","Adventure",...: 11 5 7 11 8 6 5 4 5 9 ...
## $ Publisher  : Factor w/ 579 levels "10TACLE Studios",...: 369 369 369 369 369 369 369 369 369 369
## $ 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 ...
## 2600 3DO 3DS DC DS GBA GC GEN GG N64 NES NG PC PCFX
## 133   3 509 52 2163 98 822 556 27   1 319 98 12 960   1
## PS  PS2 PS3 PS4 PSP PSV SAT SCD SNES TG16 Wii WiiU WS X360   XB
## 1196 2161 1329 336 1213 413 173   6 239   2 1325 143   6 1265 824
## XOne
## 213
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.      NA's
##      1980    2003    2007    2006    2010    2017    271
```

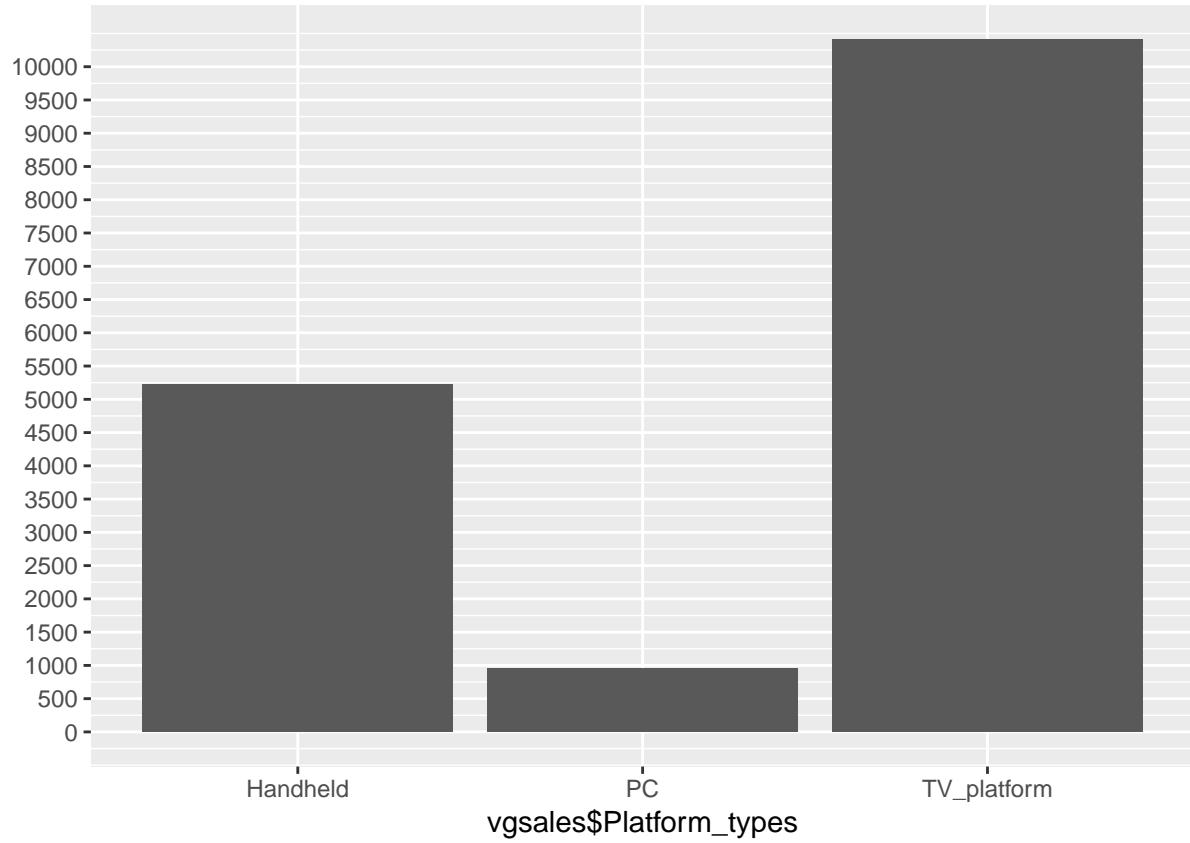
This looks much better. Now I can create some plots.



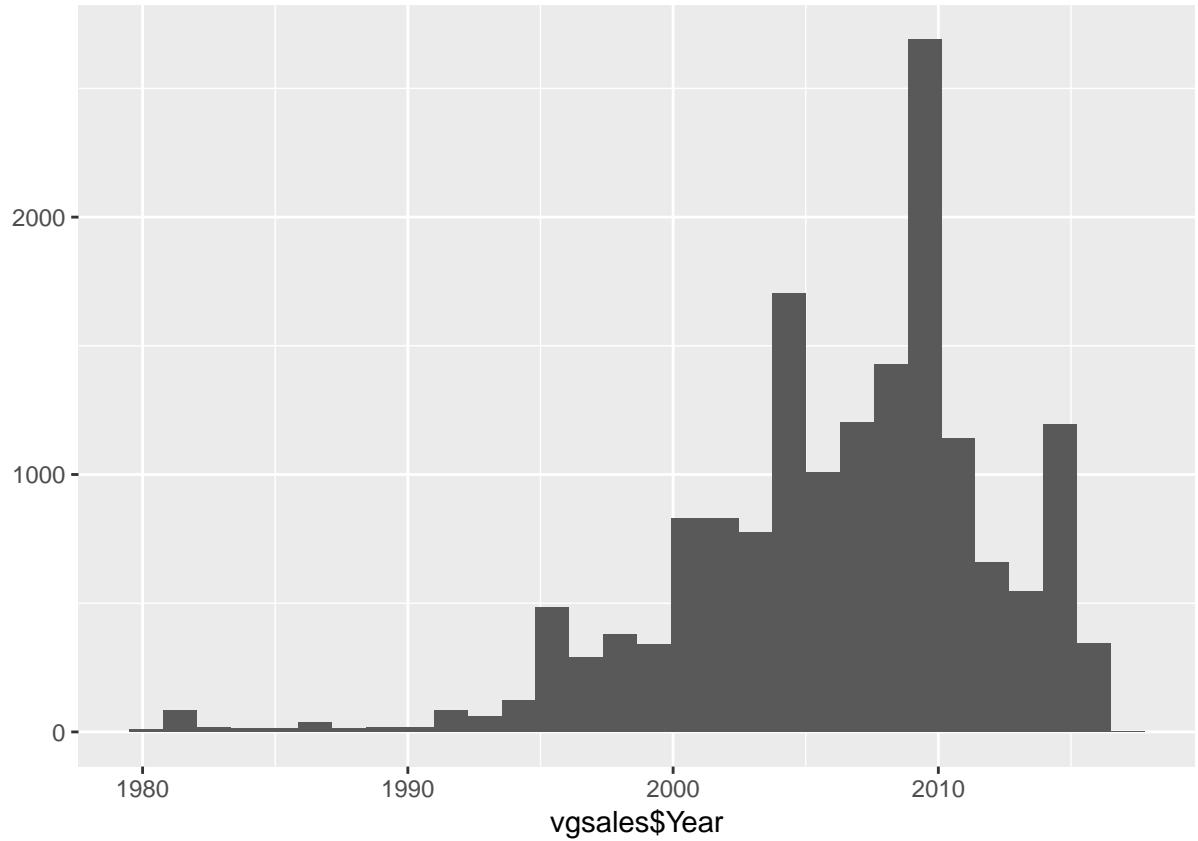
The first thing that sticks out is how many games the Nintendo DS and Sony PS2 have. They are leaps and bounds above every other system. This plot would be easier to read if the systems were grouped by company and ordered by age. I would also like to create a new variable that shows the maker of each platform.



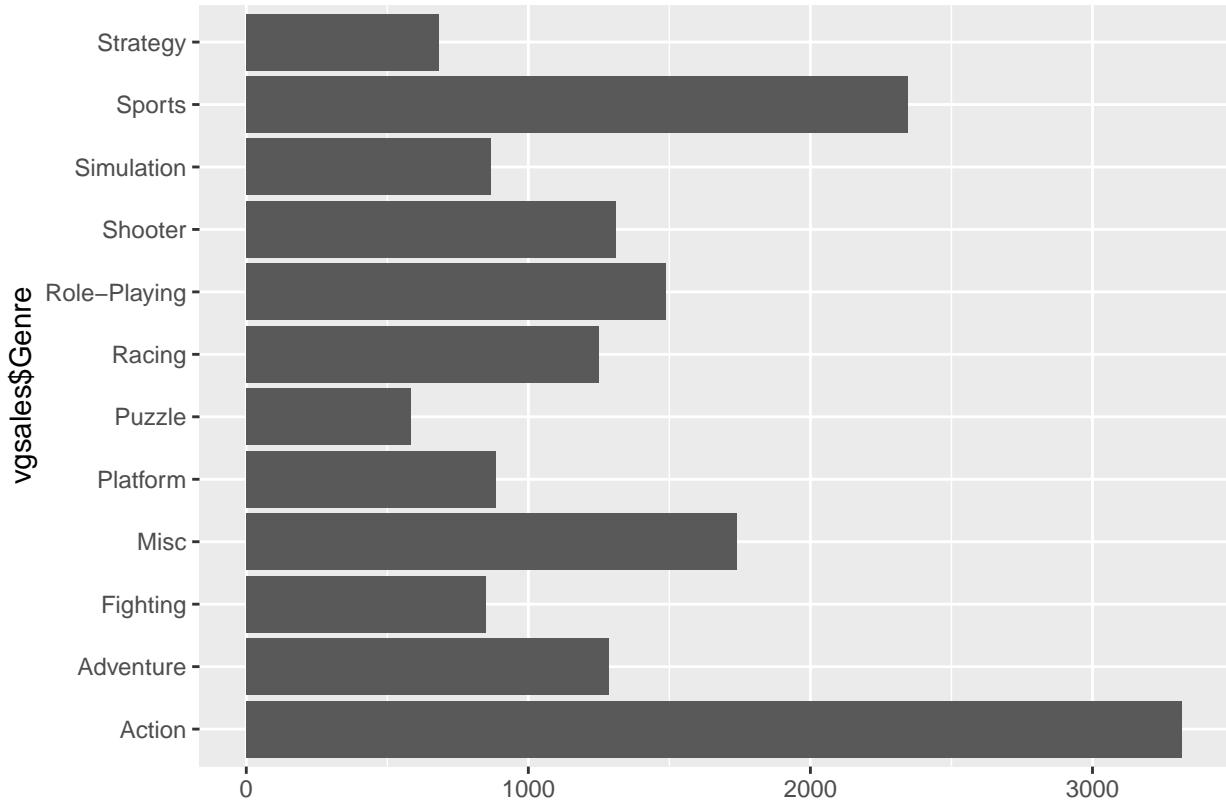
I find these numbers surprising. Despite Sony being a relatively new player on the scene, they lead with the most number of games on their systems. I am most surprised with Sega being so low. I quick use of Google tells me that they had at least more than 1,000 games on their systems. I would like to look at handheld platforms vs TV platforms vs PC. I will have to make another variable.



As said before, I doubt PC games are so few compared to TV and handheld platforms.



There were really a lot of games made in the years between 2005 to 2015.



The fact that action and sports are far and above the most common genres is not exactly unexpected. I would like to see the sales of the genres and the platforms they appear on.

```
##          Electronic Arts      1351
##          Activision        975
##          Namco Bandai Games  932
##          Ubisoft           921
##          Konami Digital Entertainment  832
##          THQ                 715
##          Nintendo           703
##          Sony Computer Entertainment  683
##          Sega                639
##          Take-Two Interactive    413
##          Capcom              381
##          Atari               363
```

```

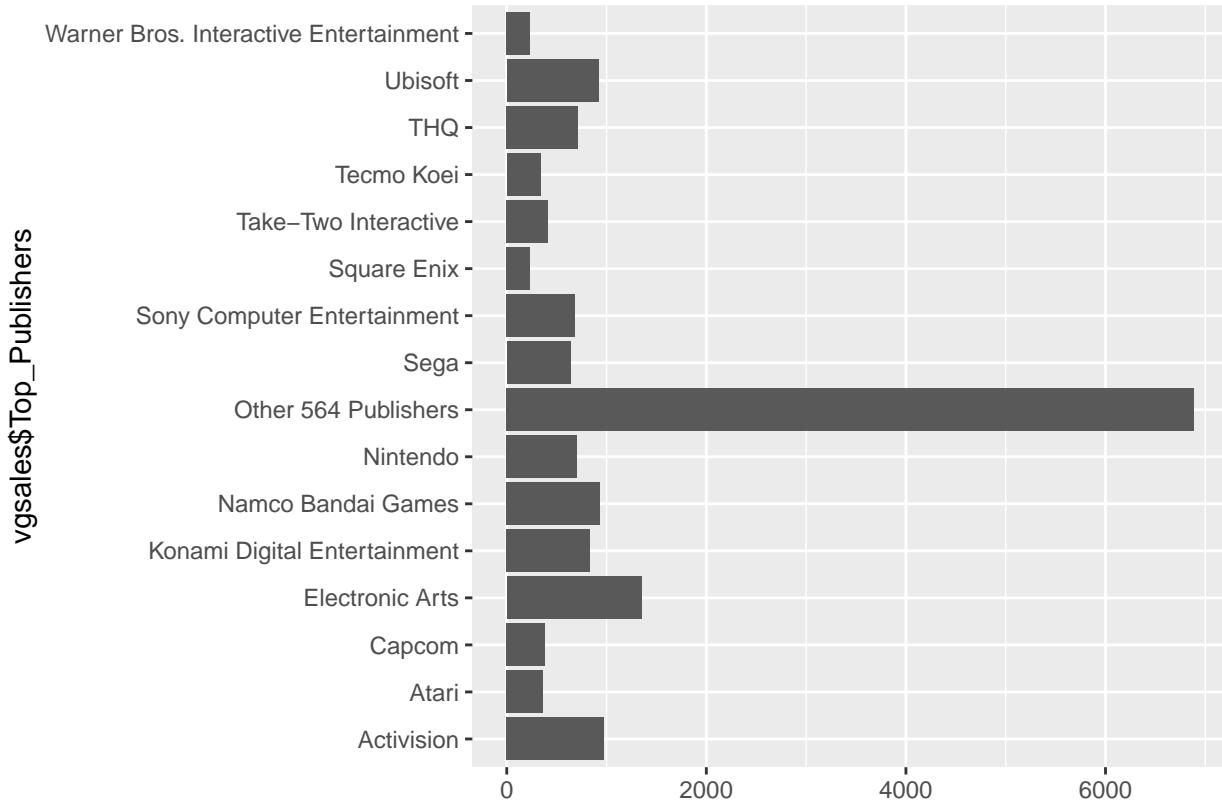
##           Tecmo Koei
##               338
##           Square Enix
##               233
## Warner Bros. Interactive Entertainment
##               232
## Disney Interactive Studios
##               218
##           Unknown
##               203
## Eidos Interactive
##               198
## Midway Games
##               198
##           (Other)
##               6070

```

There are far too many publishers to do a decent plot. There seems to be only a few publishers that make up the majority of the games.

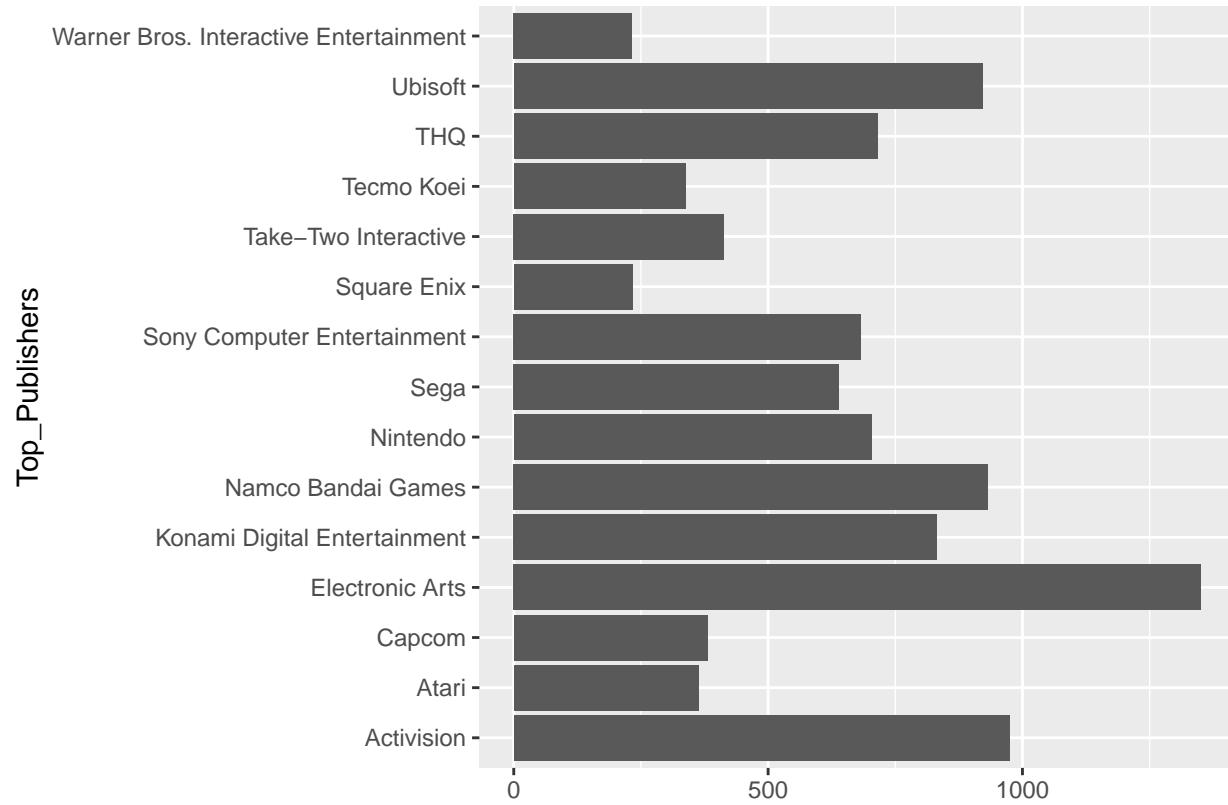
```
## [1] 579
```

There are 581 different publishers. Short of taking the top 10 or 20 and putting the rest in an ‘other’ category, I’m not sure what else there is to do at the moment. Though I would like to see how the sales of each publisher compares to each other. I would also like to see if there is a genre preference among each publisher.



I created a new variable that renames all the publishers ‘Other 564 Publishers’ unless they are in the top 15. The top 15 are the top 15 publishers from the previous summary list. However, the ‘Other’ takes up a large

amount of space on this list.



The top nine on this plot have a large number of games compared to the rest.

NA_Sales

```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0800  0.2647  0.2400 41.4900
```

JP_Sales

```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##  0.00000 0.00000 0.00000  0.07778  0.04000 10.22000
```

EU_Sales

```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0200  0.1467  0.1100 29.0200
```

Other_Sales

```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##  0.00000 0.00000 0.01000  0.04806  0.04000 10.57000
```

Global_Sales

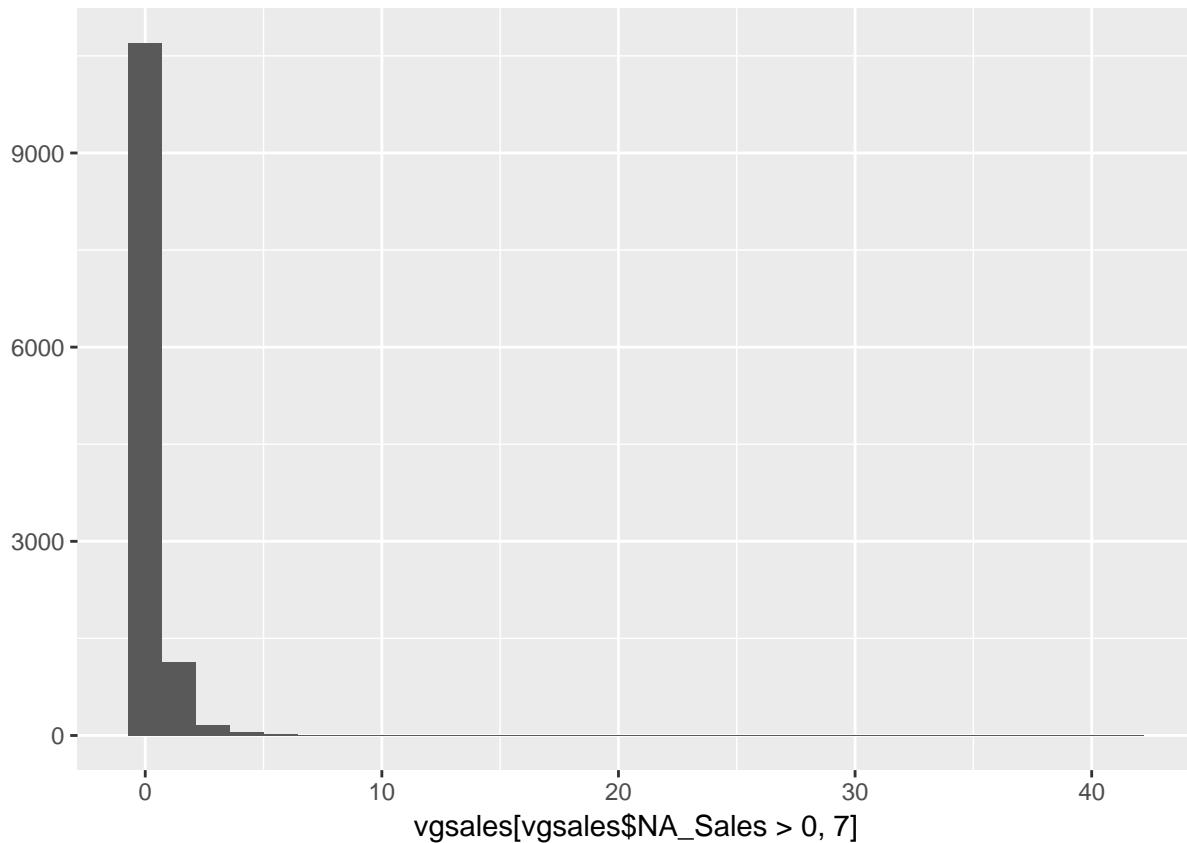
```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##  0.0100  0.0600  0.1700  0.5374  0.4700 82.7400
```

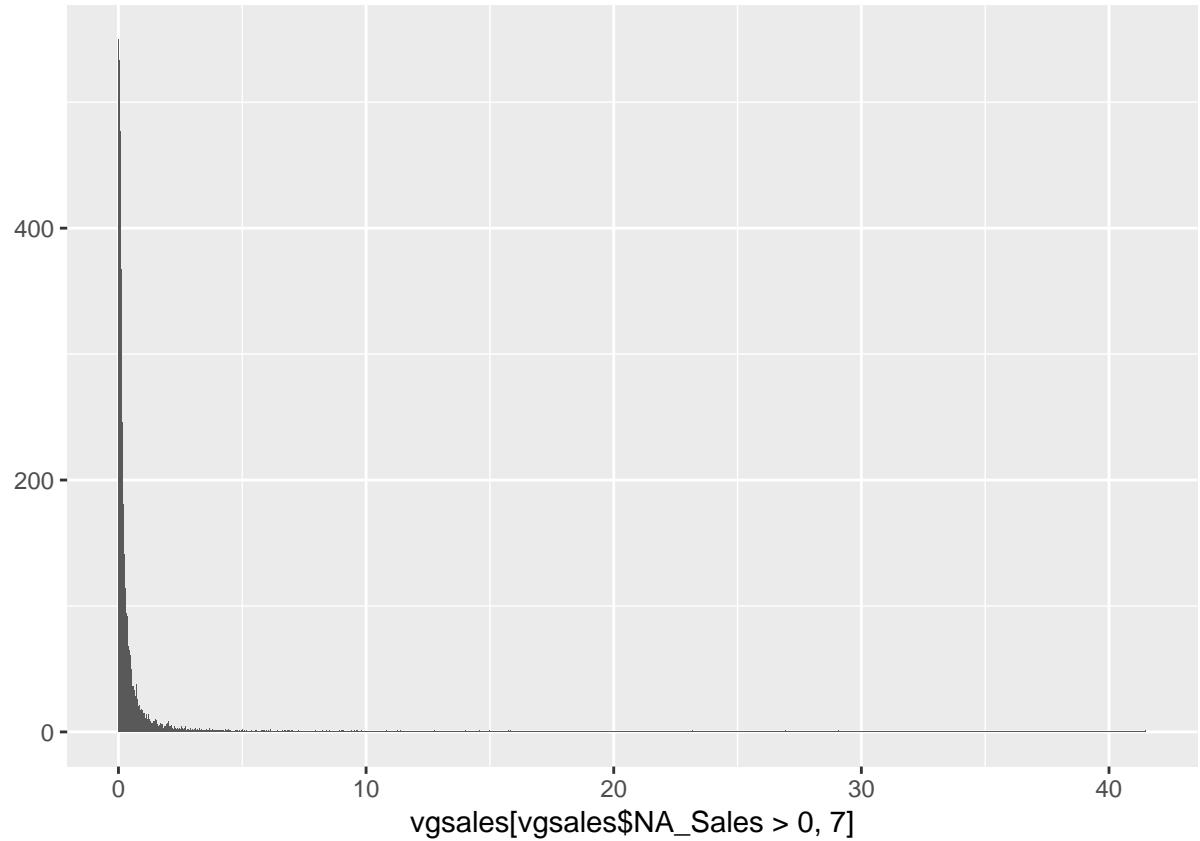
The fact that there is no region (other than global) that had all games released in it really pushes the mean and median down. A more accurate view would be to make a summary of only games released in that region. However, it could be that some 0 sales were just so small that it rounded to 0. I really need a more accurate

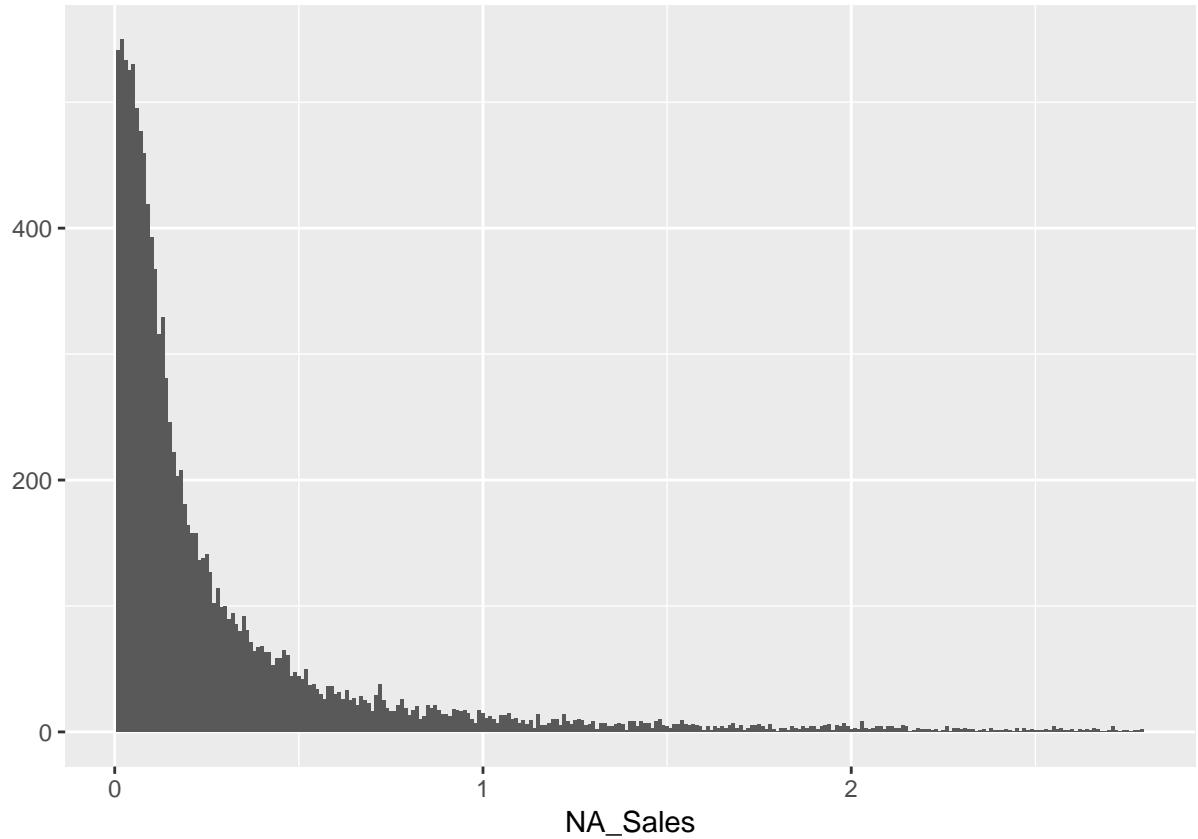
dataset. For now, I will assume (this is a big assumption on my part) that a 0 means it was not or has not been released in the corresponding region.

NA_Sales

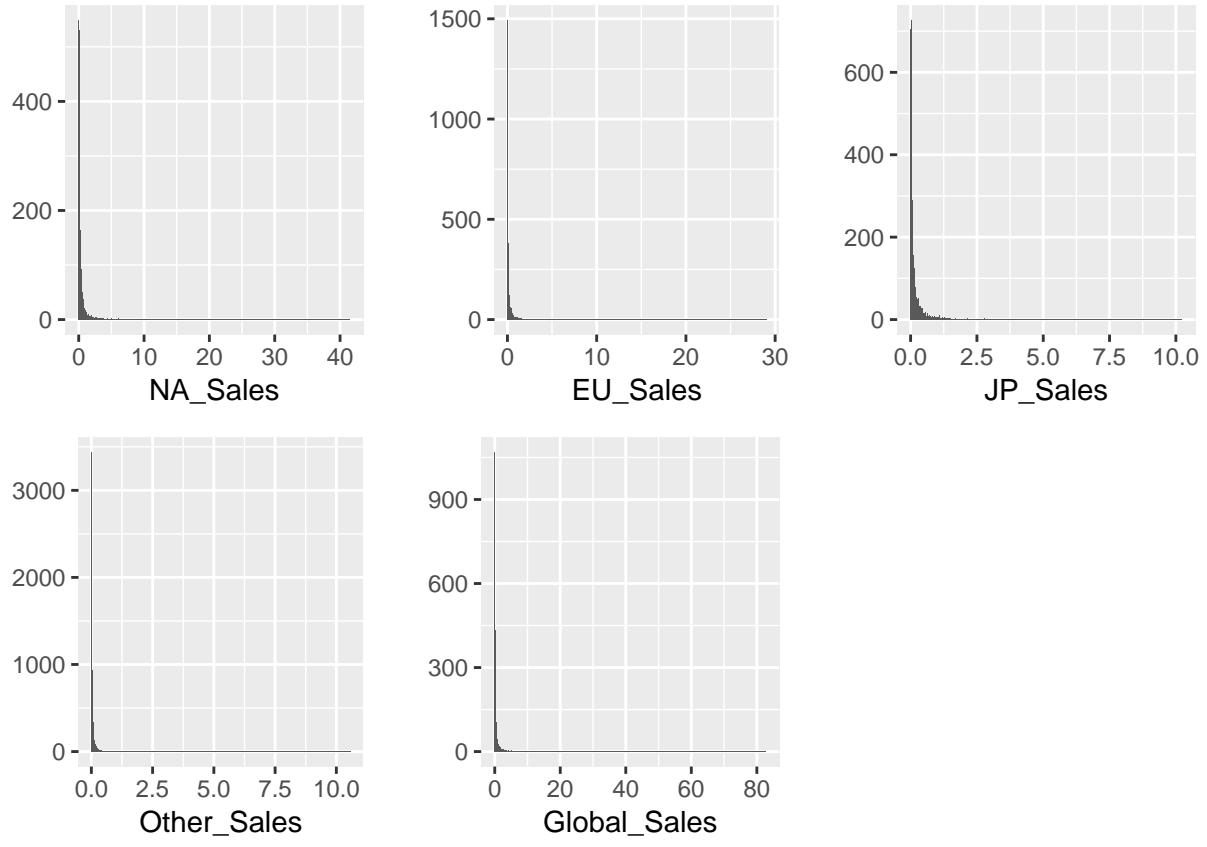
```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.  
## 0.0100 0.0600 0.1400 0.3631 0.3500 41.4900
```



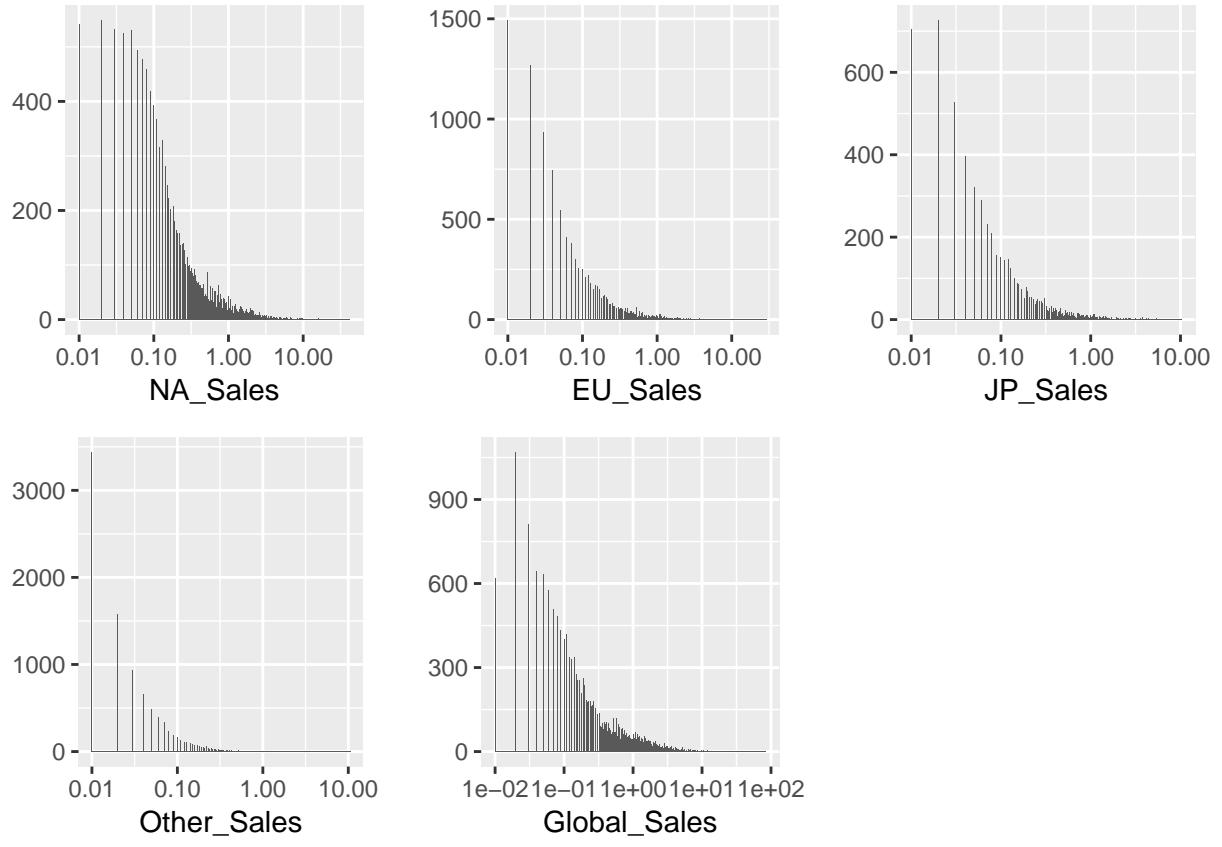




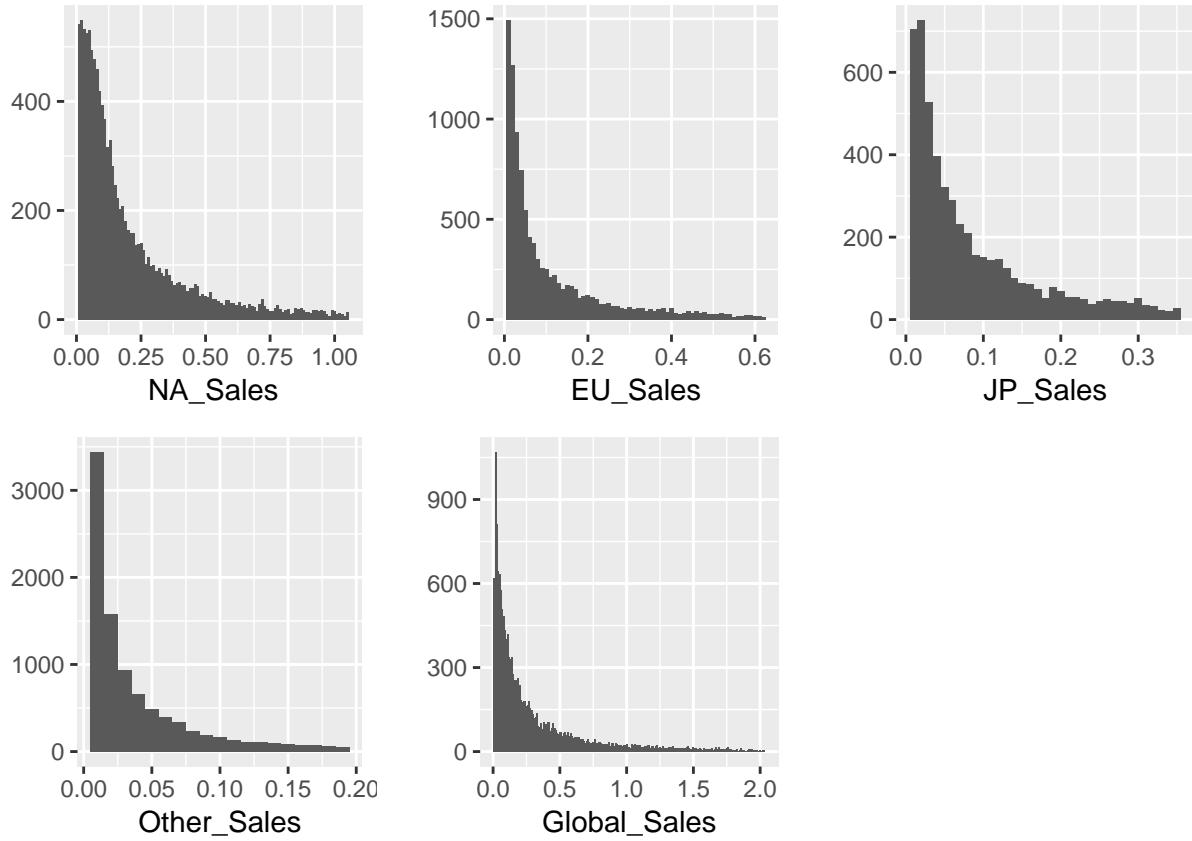
This ‘NA_Sales’ histogram is really right skewed, which is to be expected. Few games reach even one million sales. Let’s look at all the sales data together.



These plots have outliers that made it difficult to see any details.



I transformed the plots using a log10 transformation. This was better, but I thought just looking at the bottom 95% of the sales data might be better.



These histograms show basically the same pattern as one another. Though the size of the slop depends on the region.

Univariate Analysis

What is the structure of your dataset?

There are 16,598 games in this dataset. The 11 variables are Rank, Name, Platform, Year, Genre, Publisher, NA_Sales, EU_Sales, JP_Sales, Other_Sales, and Global_Sales. I create 3 additional variables: Platform_makers, Platform_types and Top_Publishers. Rank and Year are integers, Name, Genre, Publisher, Platform_makers, Platform_types and Top_Publishers are factors and the remaining variables are numeric.

What is/are the main feature(s) of interest in your dataset?

My main interests in this dataset are the various sales figures.

What other features in the dataset do you think will help support your investigation into your feature(s) of interest?

Platform, year and genre would be interesting to look at along with the sales data. If I could find a way to use publisher data, it could be interesting. But it would have to be changed to make it more usable. There are far too many different publishers.

Did you create any new variables from existing variables in the dataset?

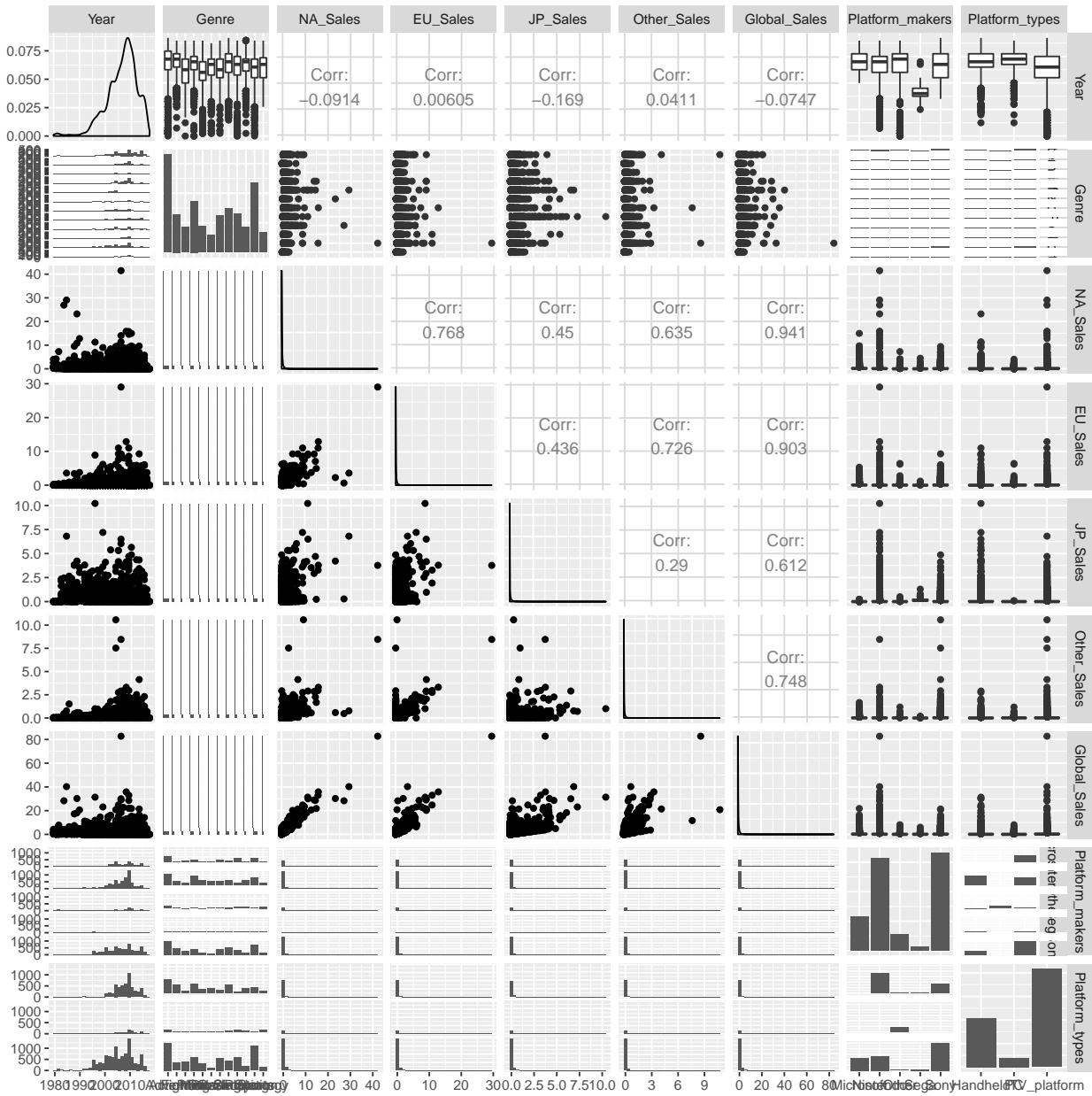
I created a top 15 publishers by number of games published variable. I also created a variable of platform makers and another variable of platform type.

Of the features you investigated, were there any unusual distributions?
 Did you perform any operations on the data to tidy, adjust, or change the form of the data? If so, why did you do this?

I did not notice any unusual distributions.

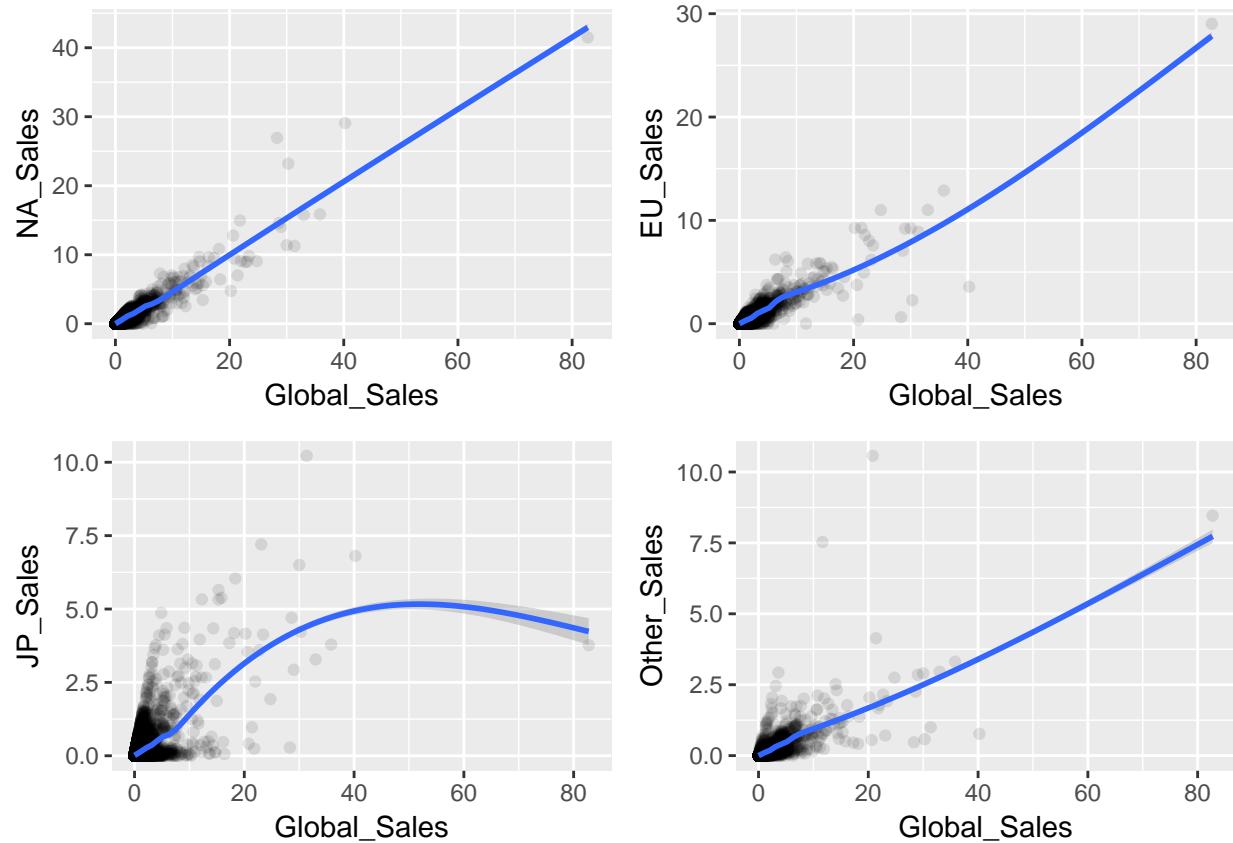
There were some obvious errors in the data that required fixing (though I am sure there are some or many not so obvious errors). I changed a year that said 2020 when it should have said 2009. Some errors in the csv file caused a couple of commas to go unnoticed. So I had to manually shift all the data to the right.

Bivariate Plots Section

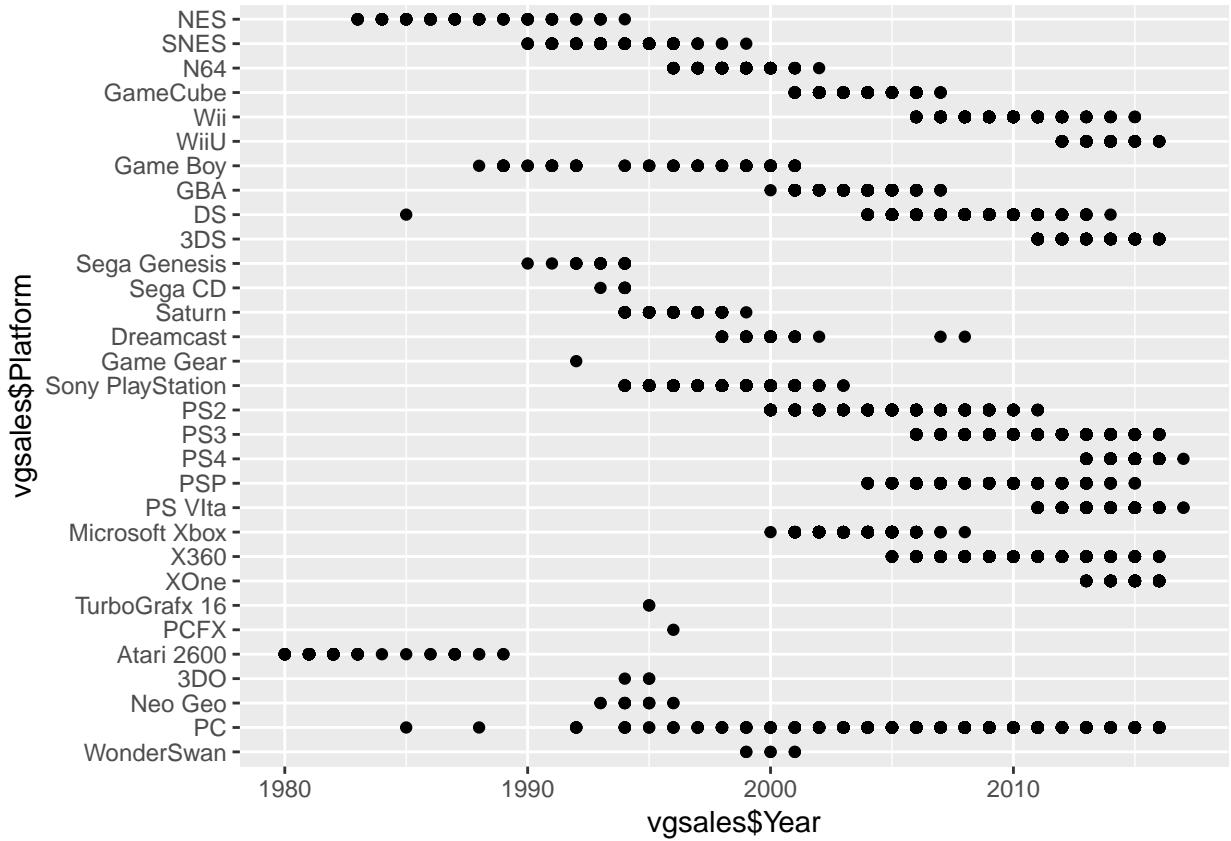


A matrix of plots on all 14 variables would be too cramped and unreadable, so I removed some variables and kept the variables that would be most interesting to see in a matrix of plots.

It seems that global sales are very correlated with North American and European sales. While Japan seems to be going its own way with no strong correlations.

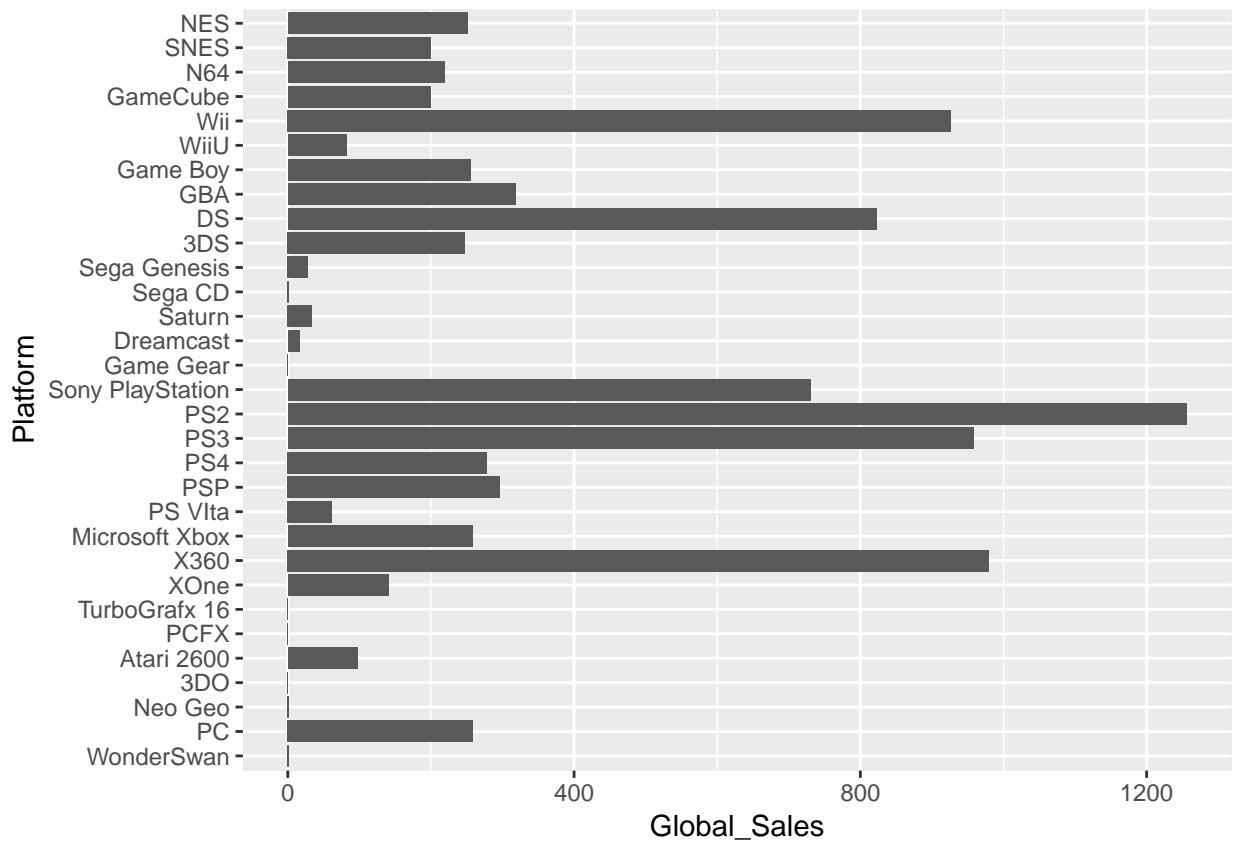


North American and EU sales seem to correlate well with global sales. But 'Other_Sales' and especially Japan's sales are really spread out.

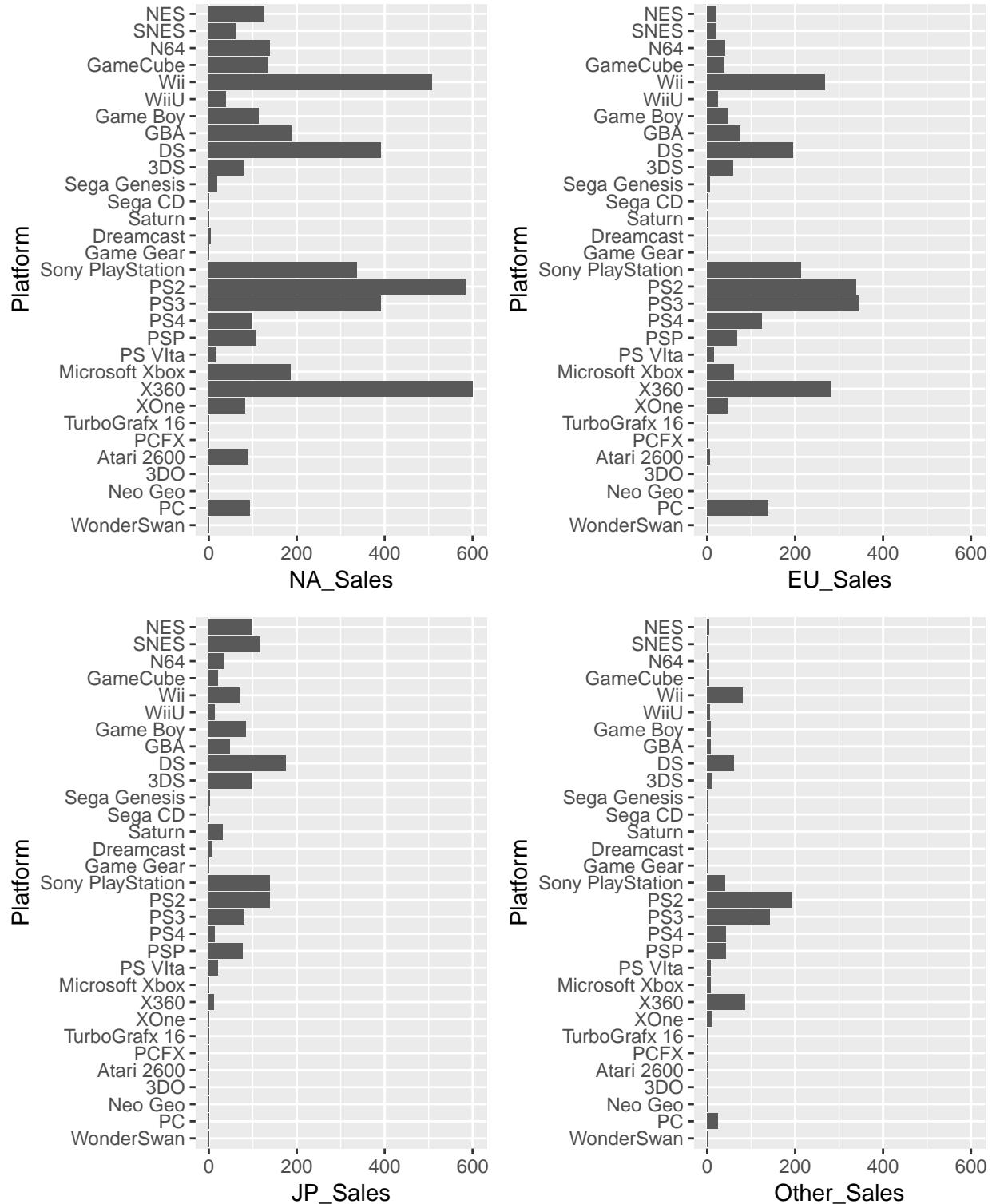


This is a good look at the lifespan of each platform. Though it shows another problem with the dataset. The Nintendo DS was not released until 2004, yet there is a game for it in the mid-1980s. When searching for the title on Google, the top results give the same mistaken date of 1986.

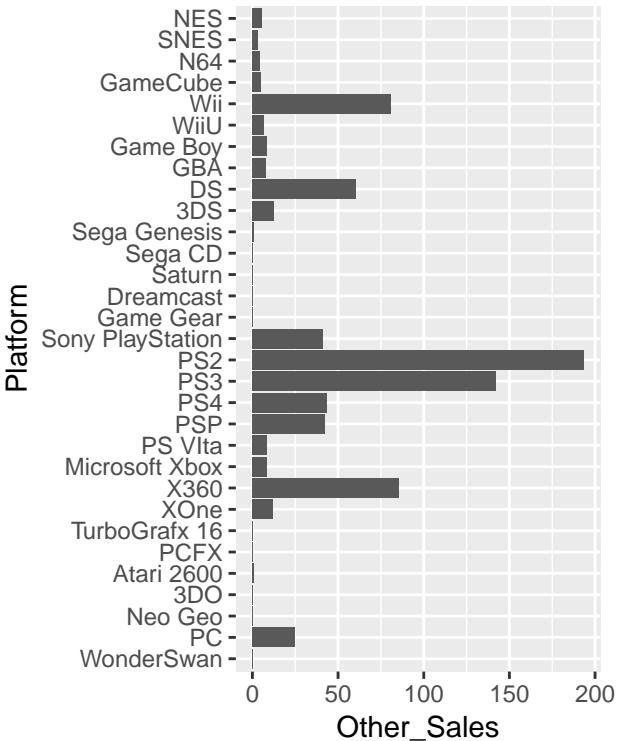
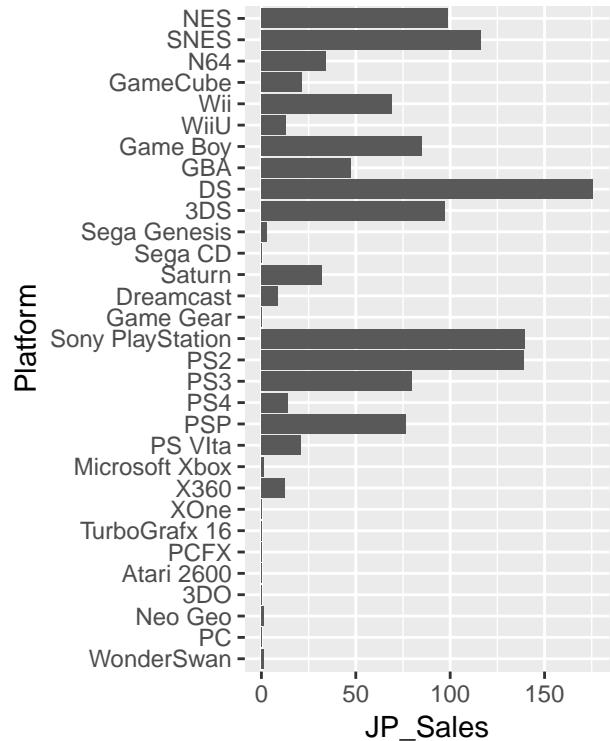
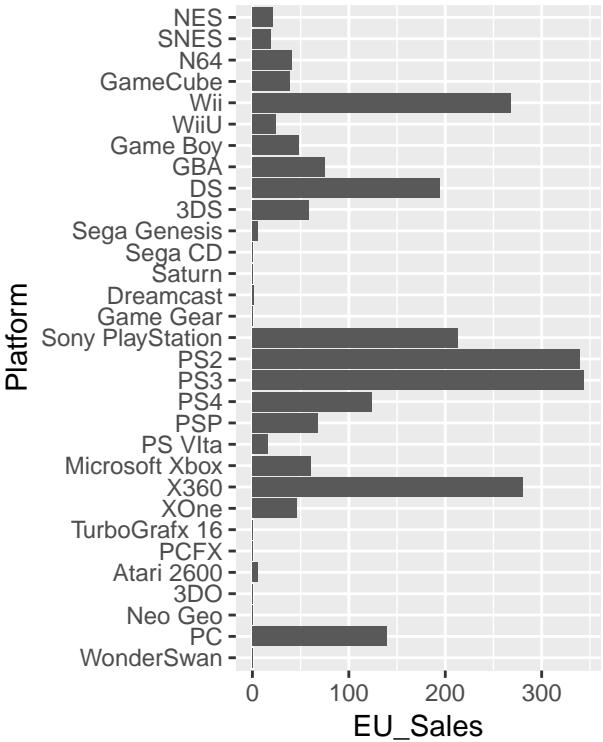
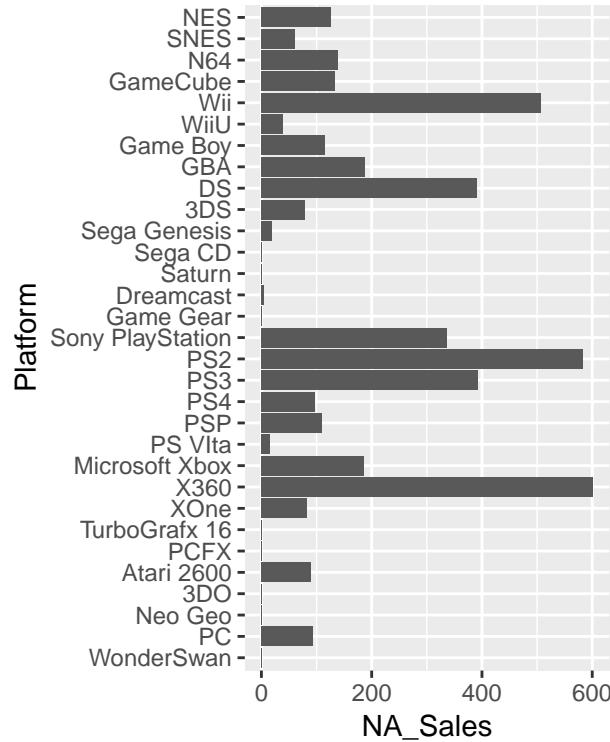
I would also like to see the number of games for each year via dot size.



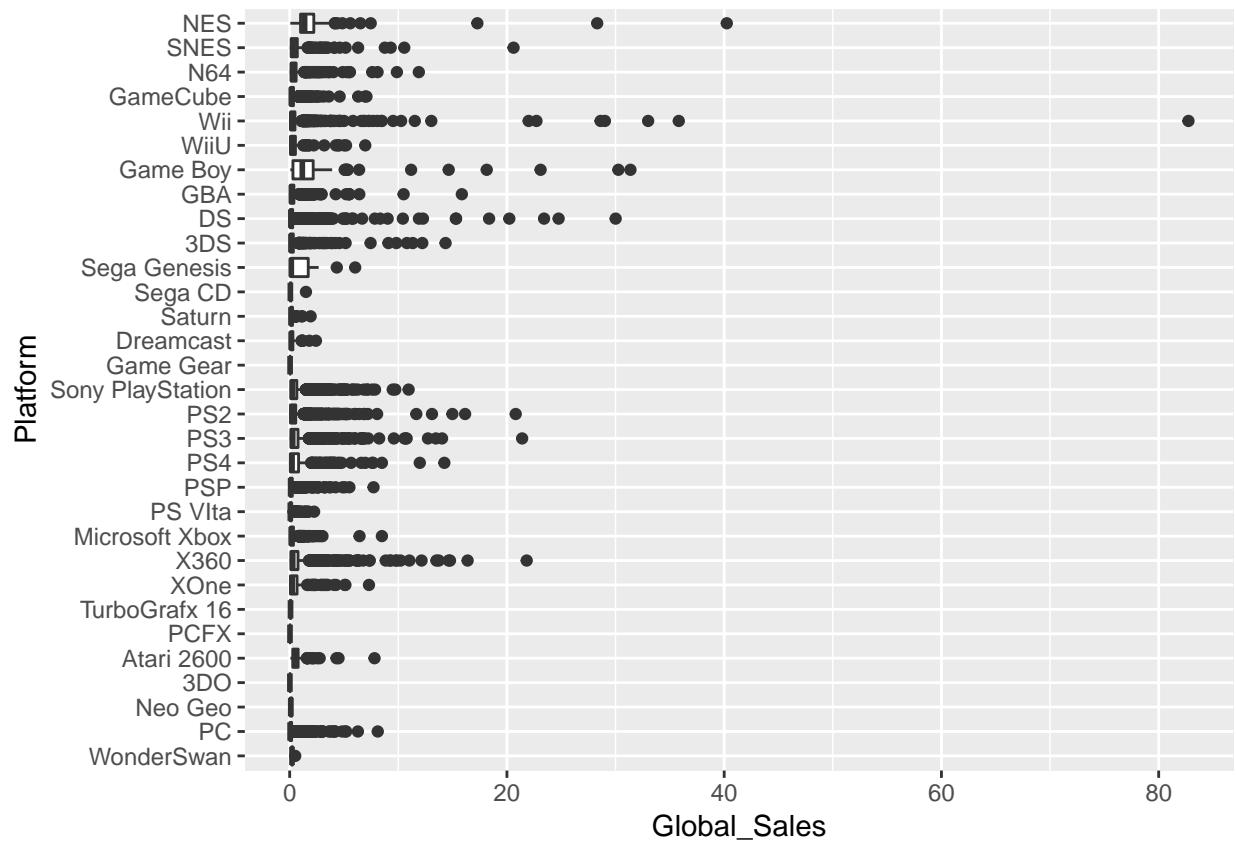
The only surprising thing here is the extent to which the PS2 out sold the other platforms. I wonder how each platform does regionally.



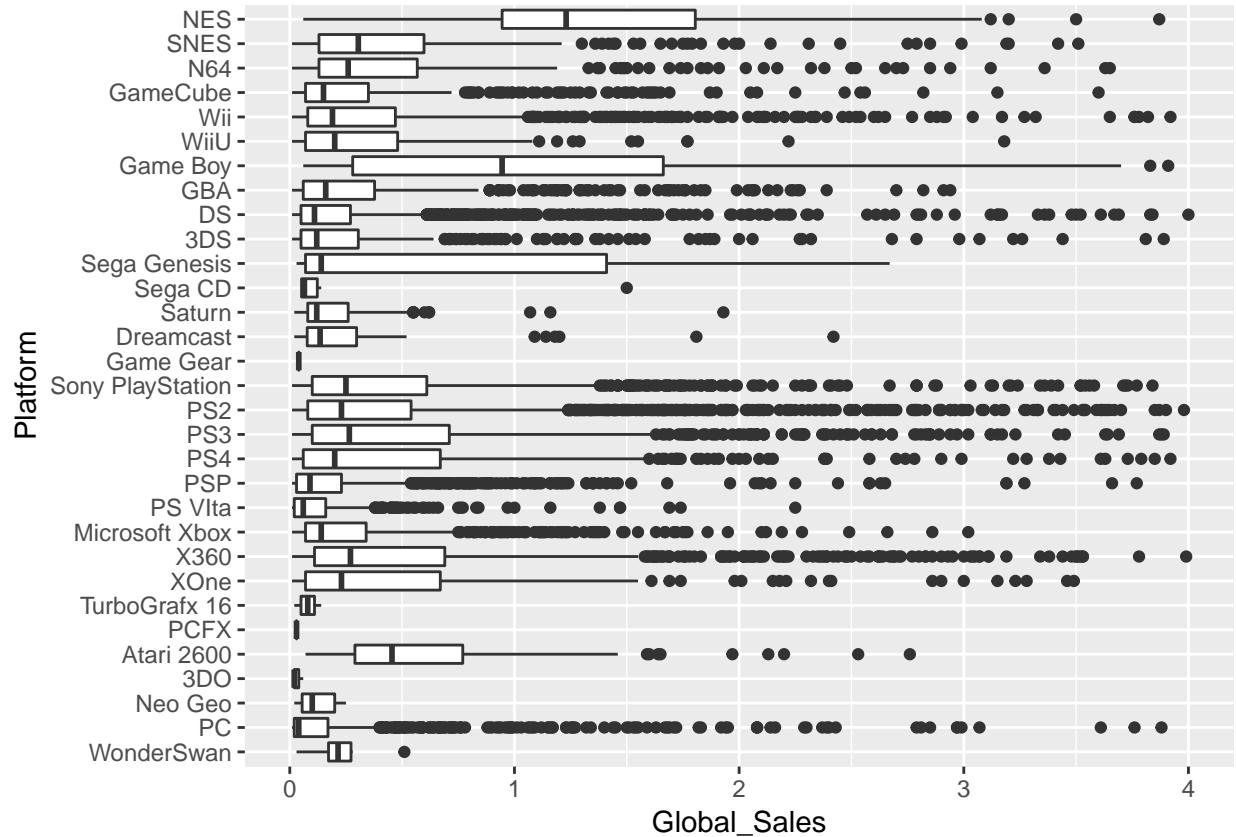
This shows how important the North American and EU sales are to the publishers of games. A single plot with overlapping bars and color coded by region would be easier to read I think.



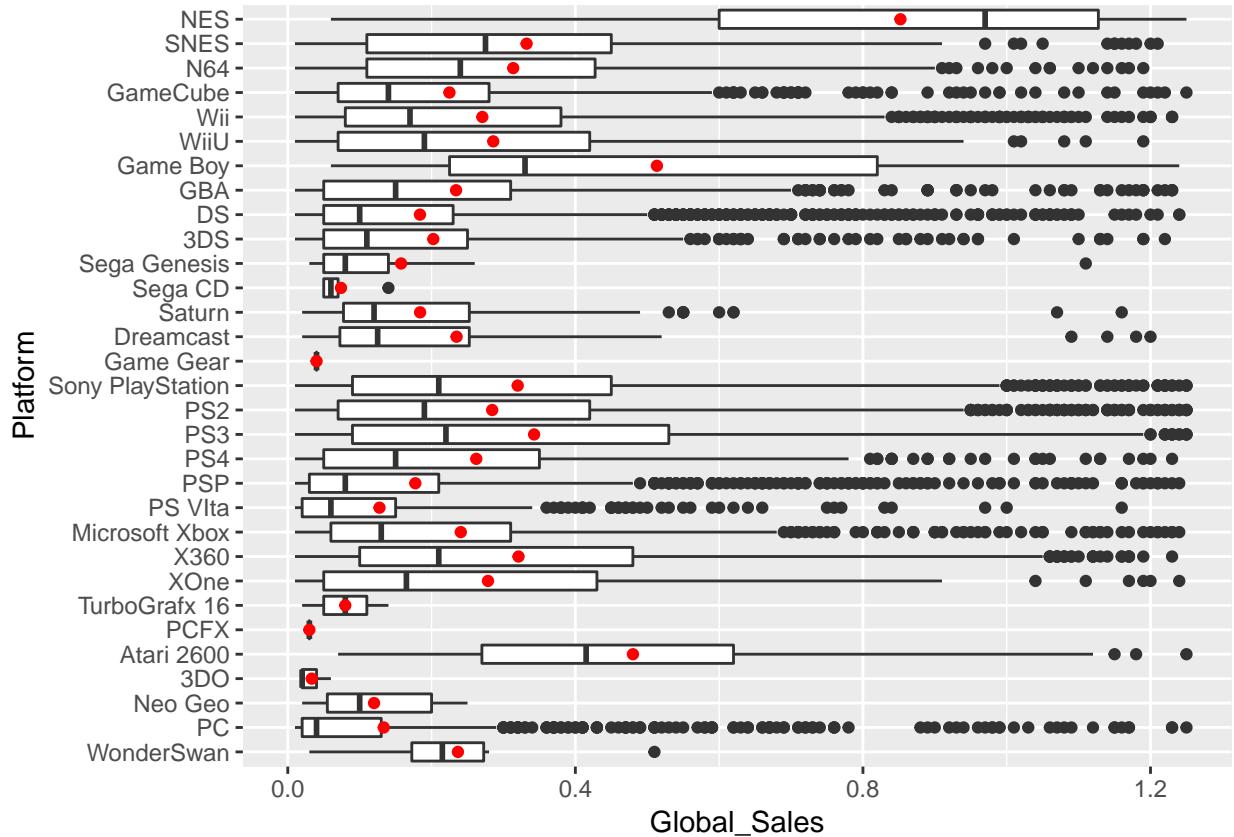
Looking closer at the data for each region gives a clearer look at regional platform preferences. The ratio of Nintendo games compared to other games is much bigger in Japan compared to other regions. And Microsoft seems to be fairly unimportant in Japan. However, Sony has a strong presence in all regions. I wonder what the mean sales looks like.



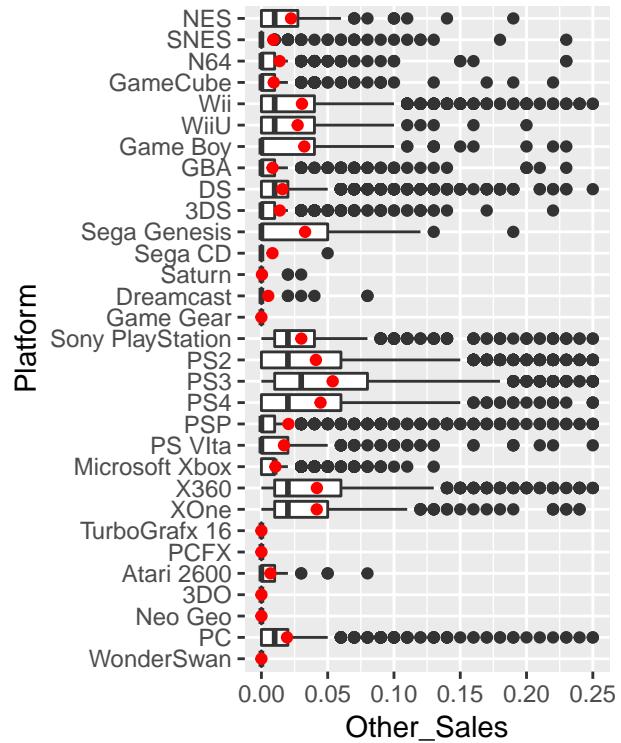
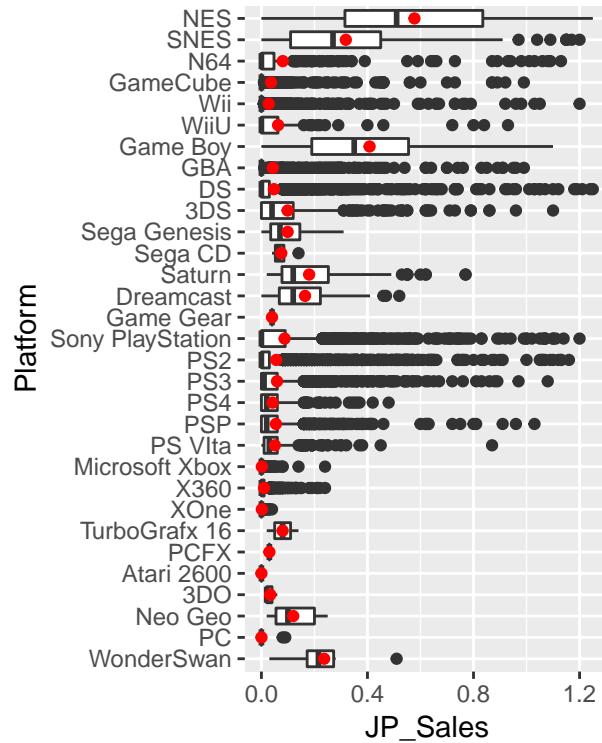
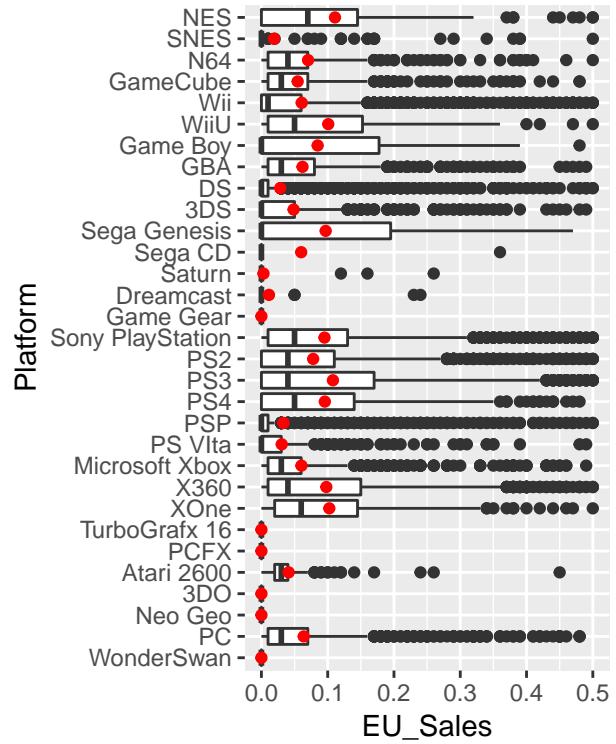
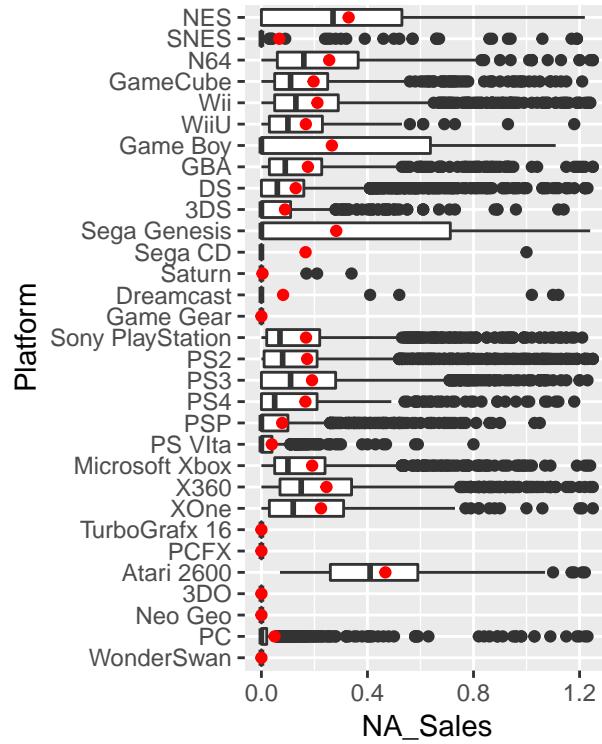
The upper end sales are making it difficult to see the sales' boxplots.



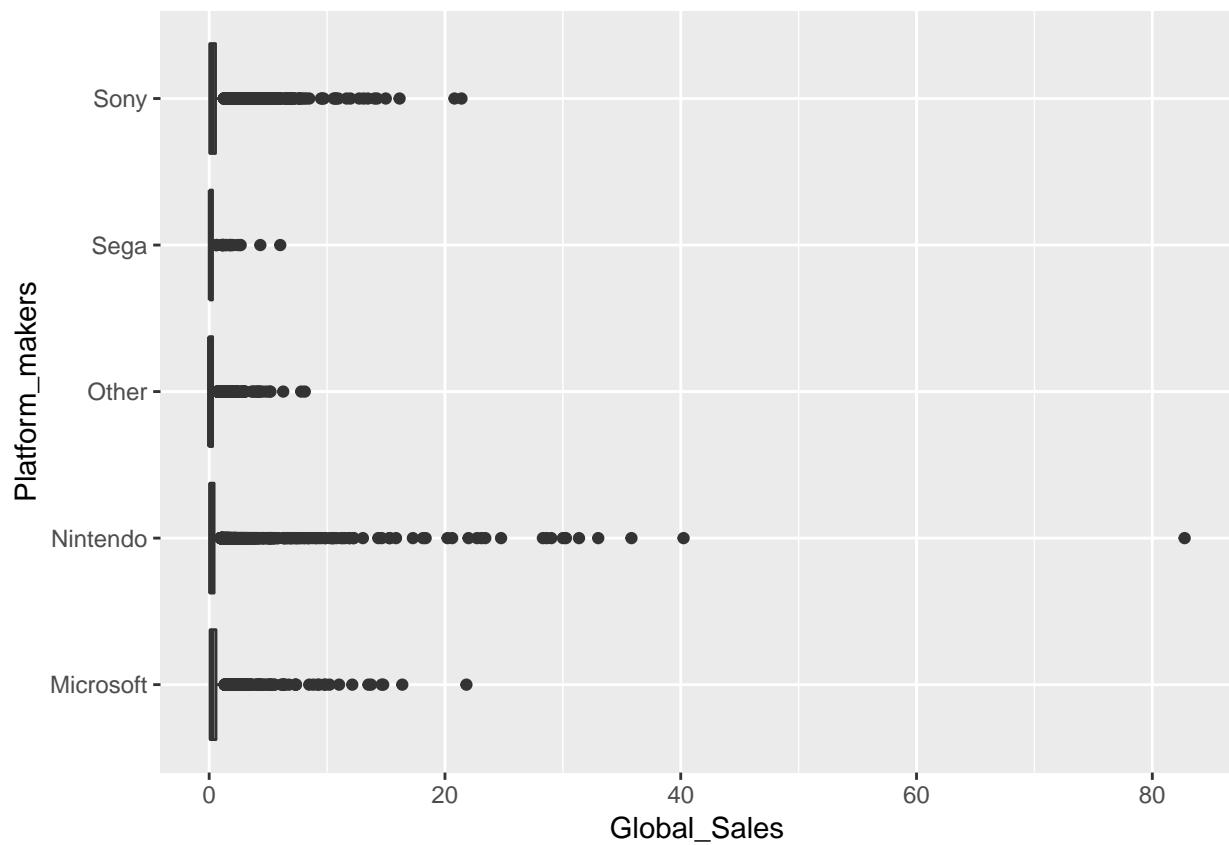
This is better, but I want to go even closer.

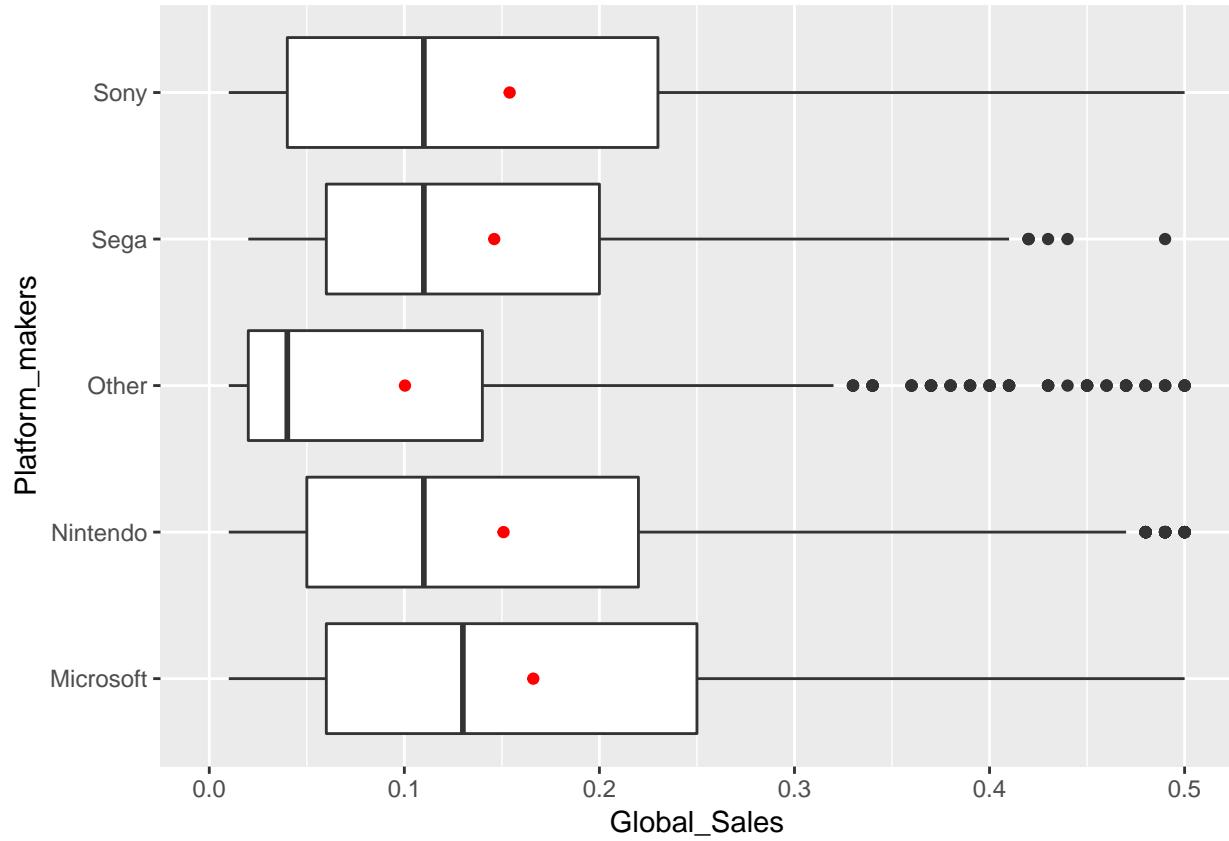


I added a red dot to represent the mean. Compared to total sales, Nintendo seems to have a strong showing in the mean sales. Atari 2600, despite being so old, seems to also have a strong mean.

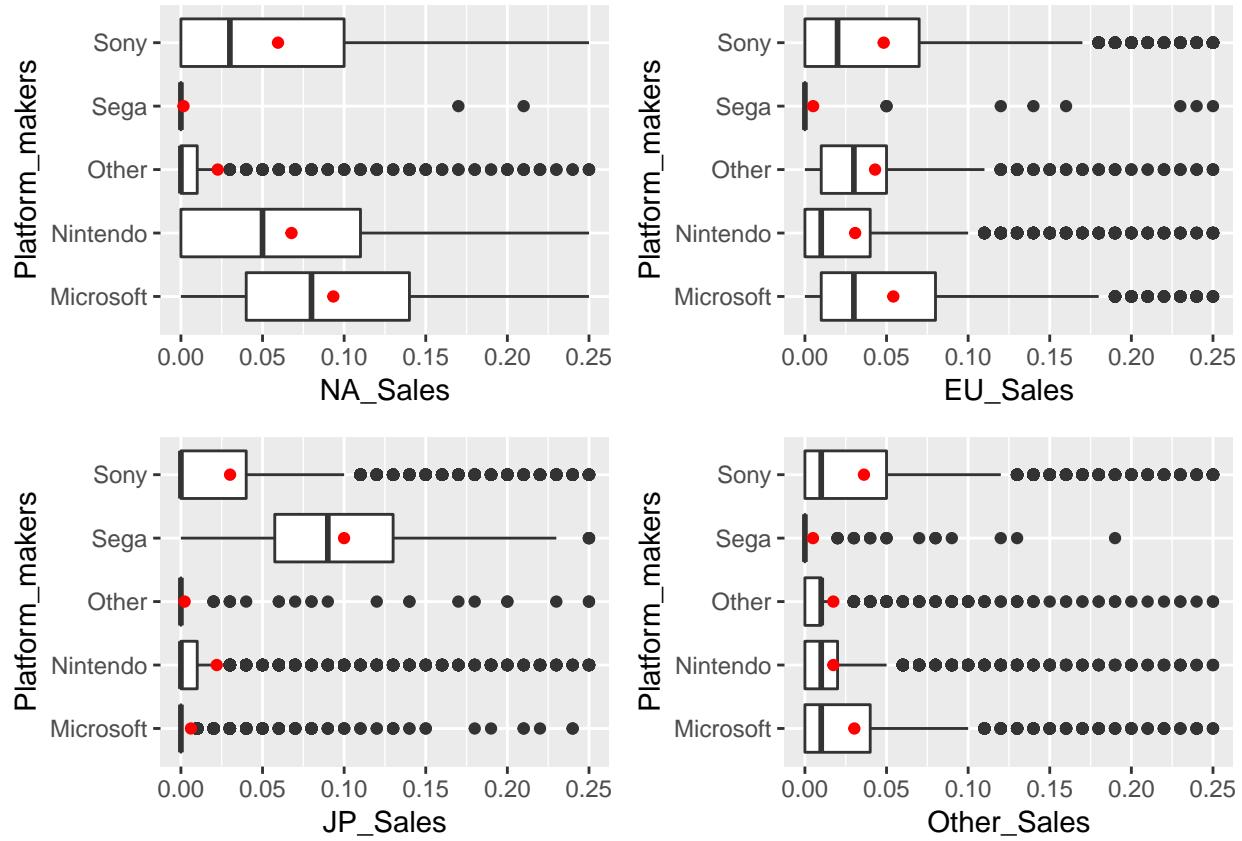


It is interesting that the Atari 2600 has such a strong mean presence in the North American market. Some other interesting data are the mean sales for the NES, SNES and Game Boy compared to the other platforms in the Japanese market.

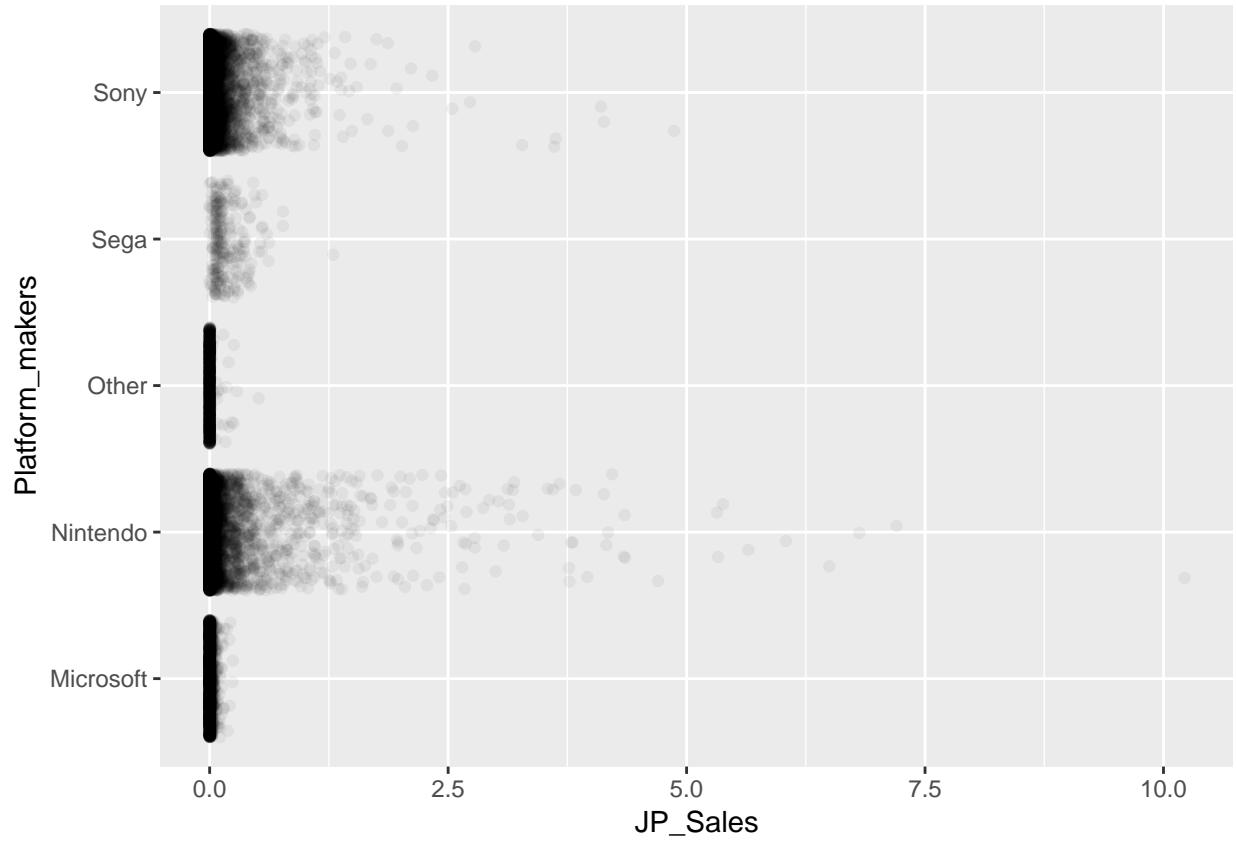


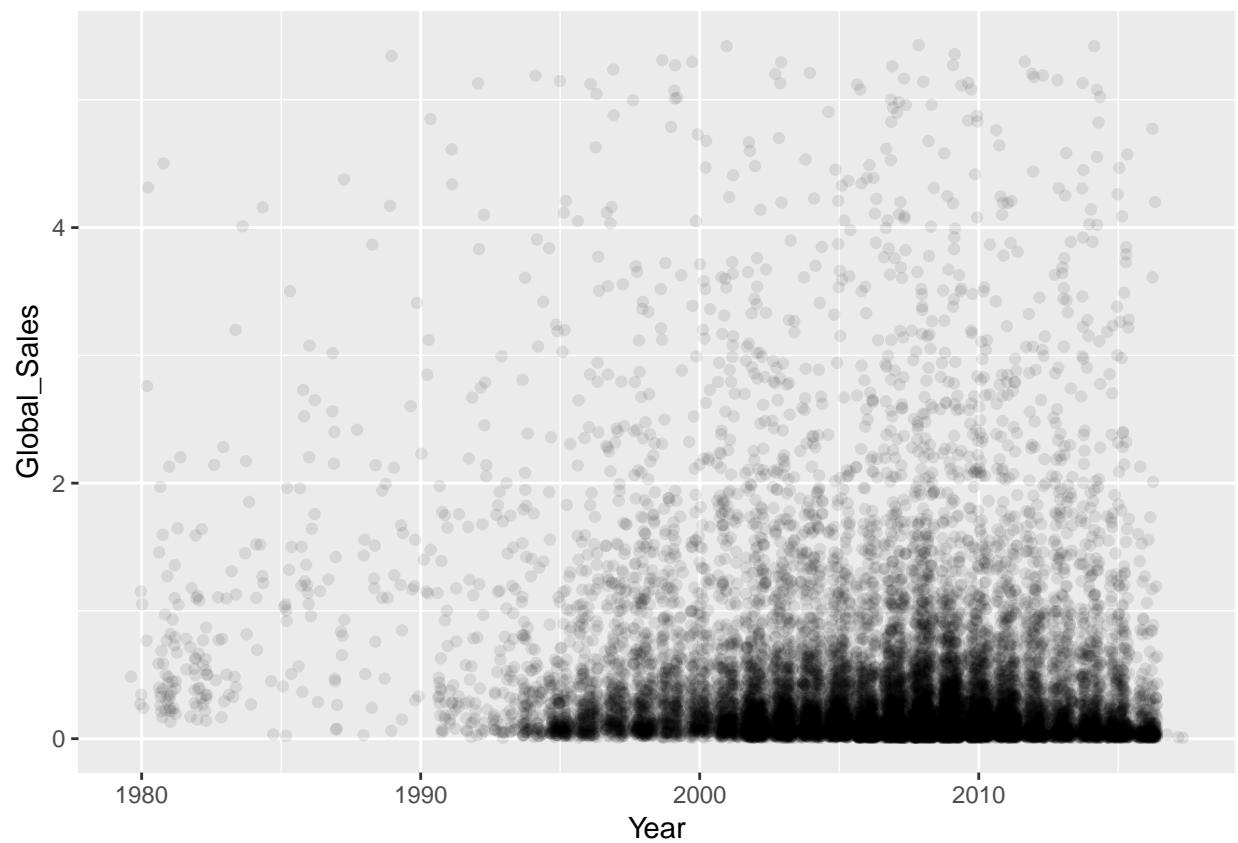


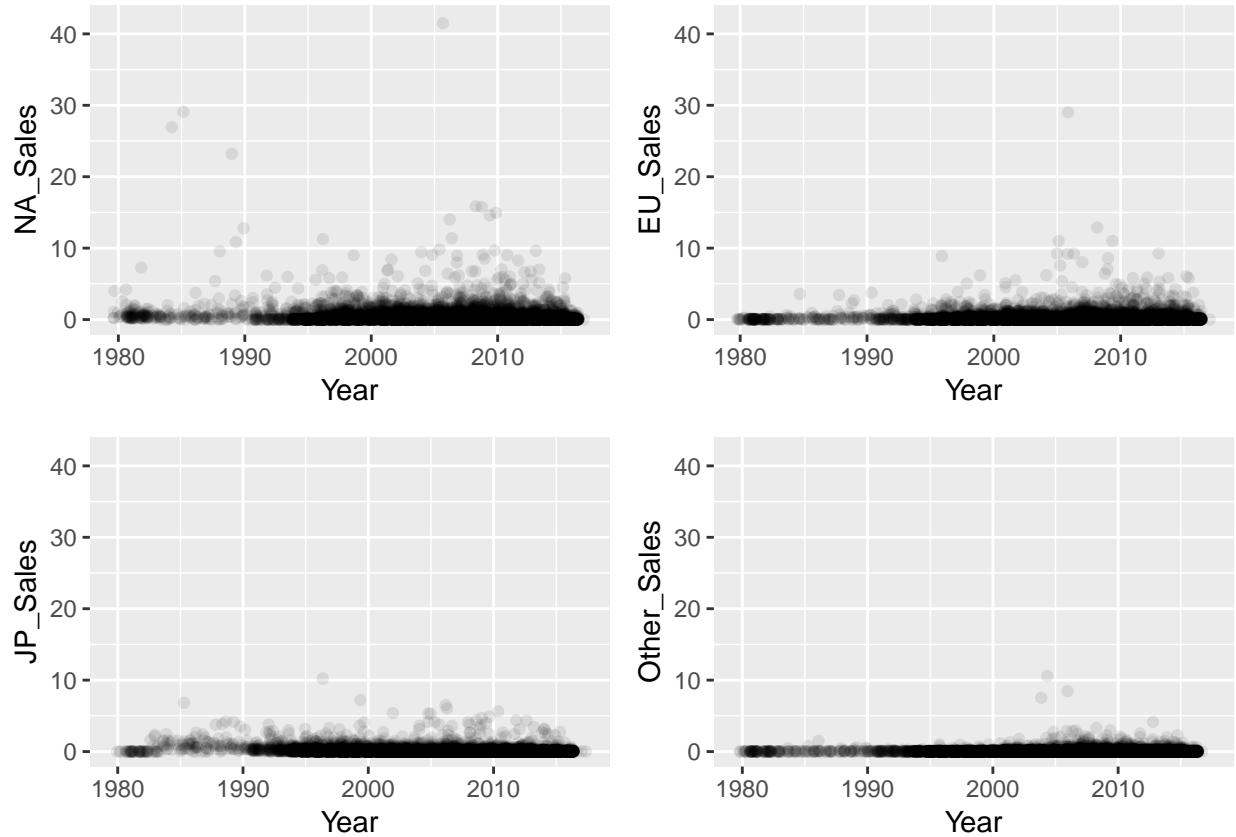
I'm surprised to see that Sony, Sega and Nintendo are neck and neck. Though I'm more surprised that Microsoft has a slight lead in mean and median sales.



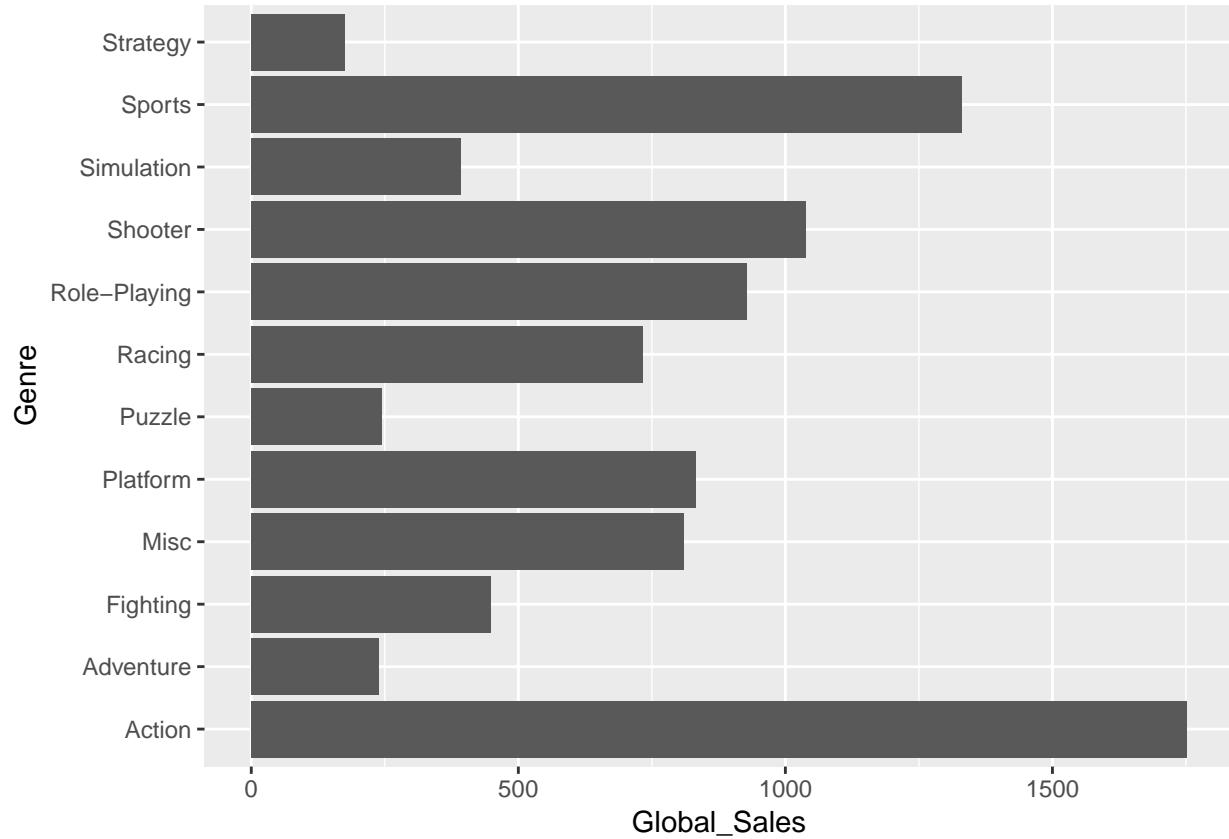
These regional means are very surprising and a little suspicious. The mean of Nintendo and Sony in Japan being so much lower than Sega is especially suspicious.

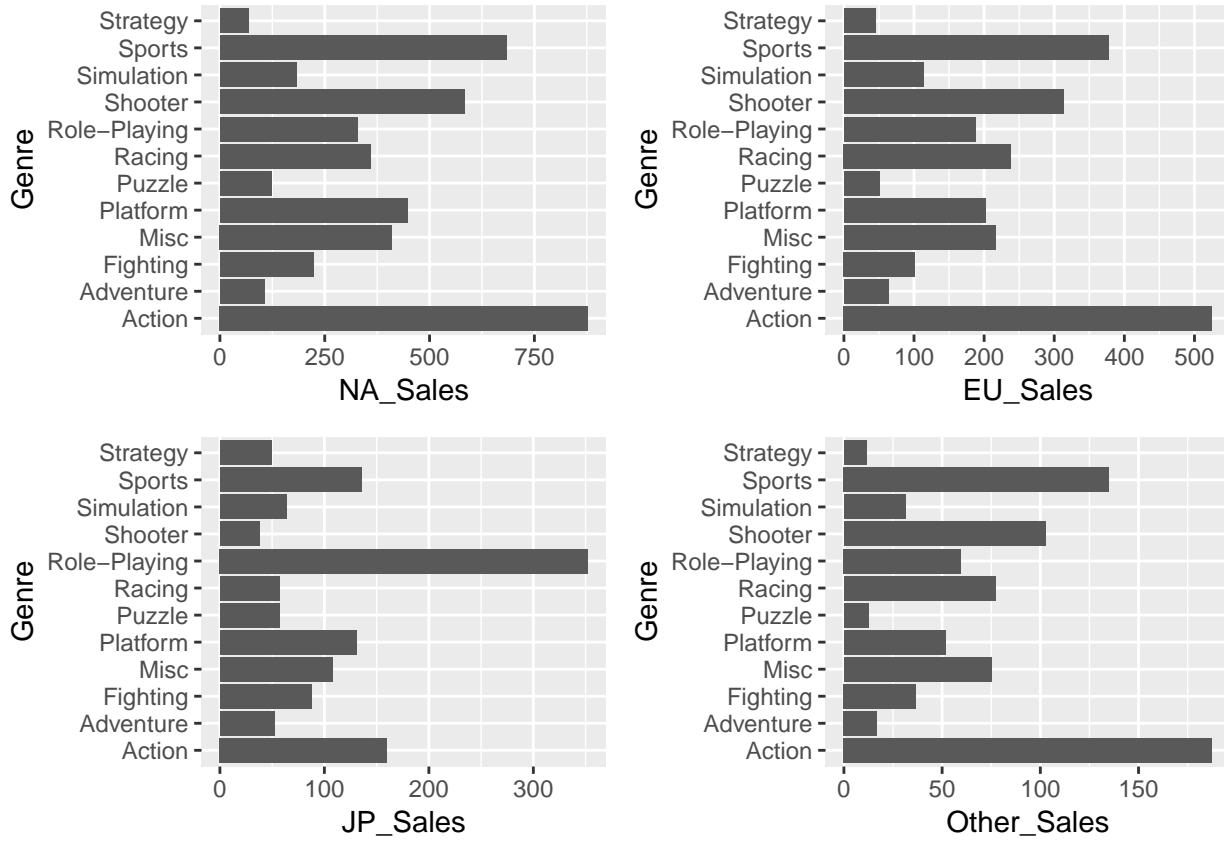




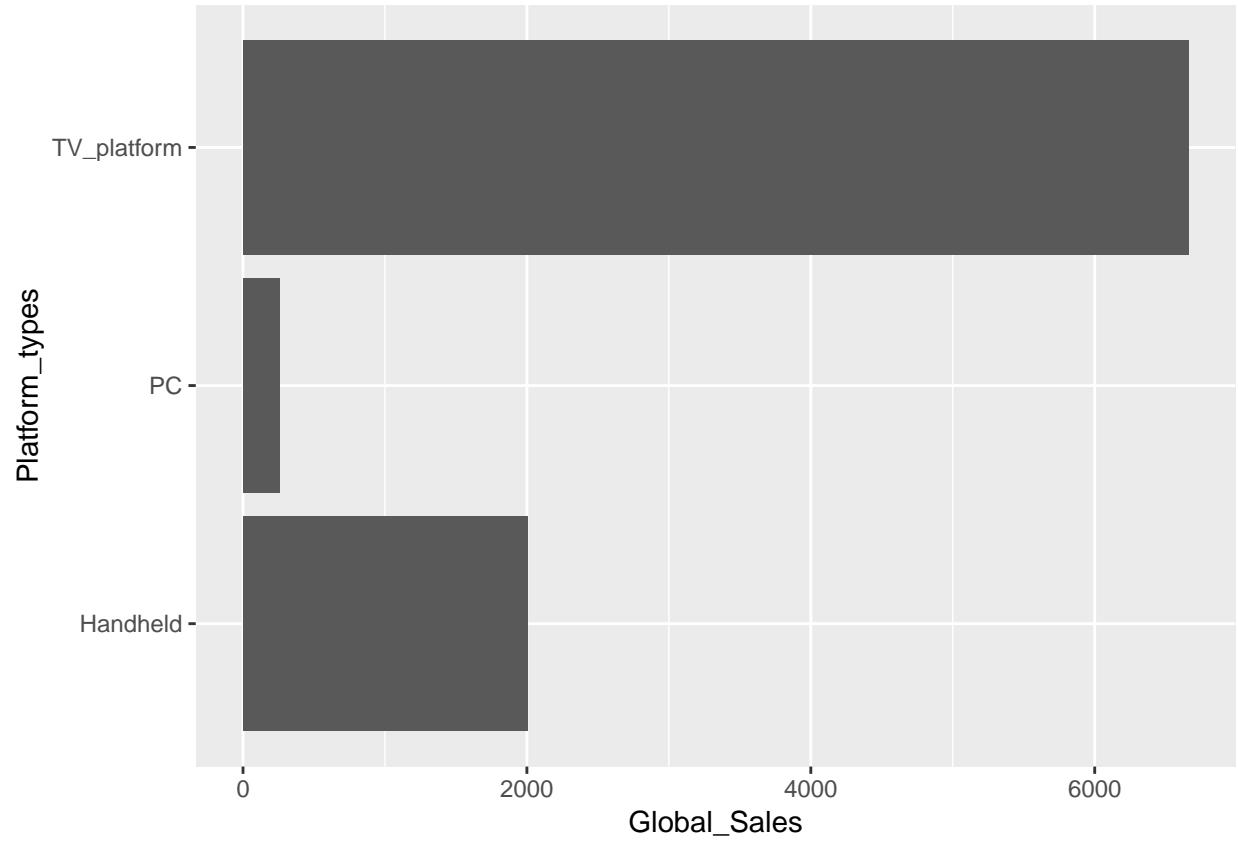


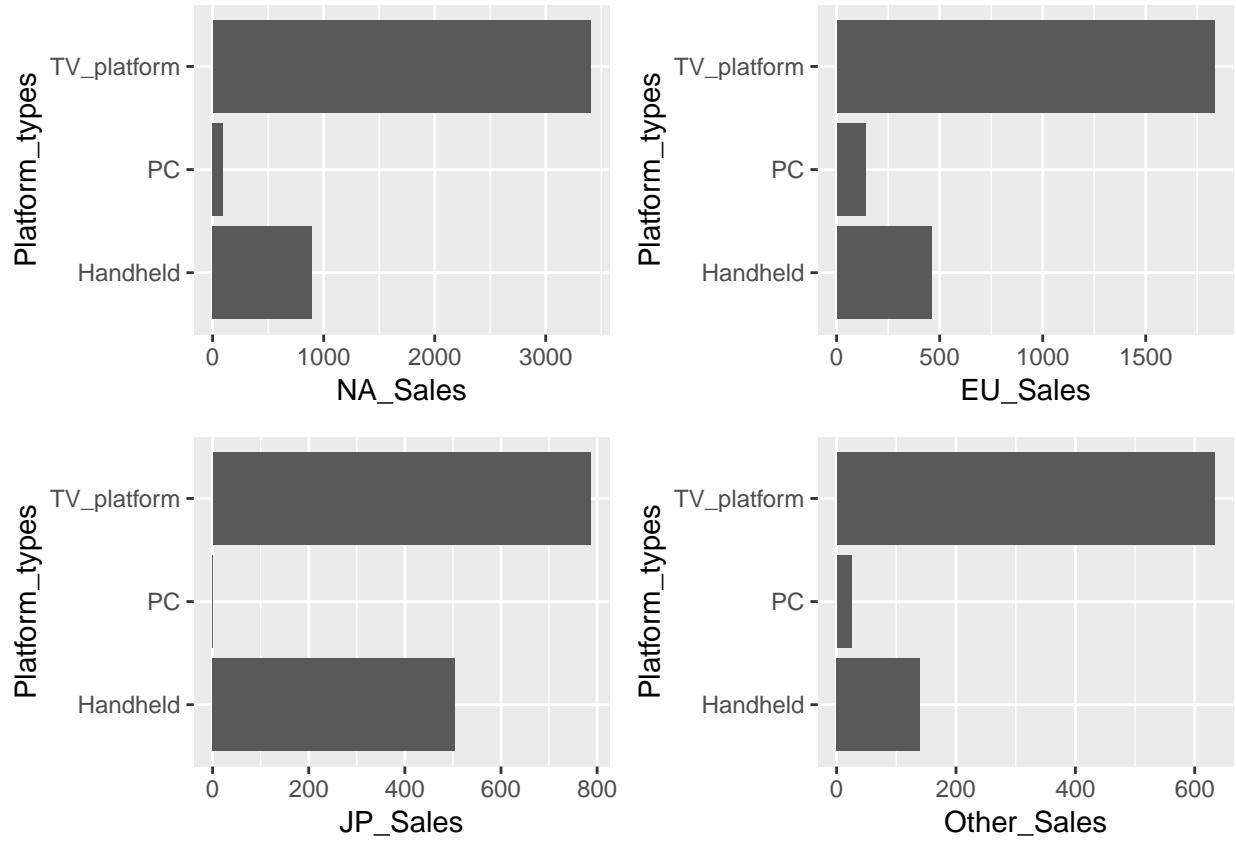
It is interesting the semi-uniformity of the sales in the North American and Japanese regions while the EU and other regions lean more towards the 2000s and 2010s. Though that would make sense if you take into account the increase of people that can now afford the luxury of buying videogames.



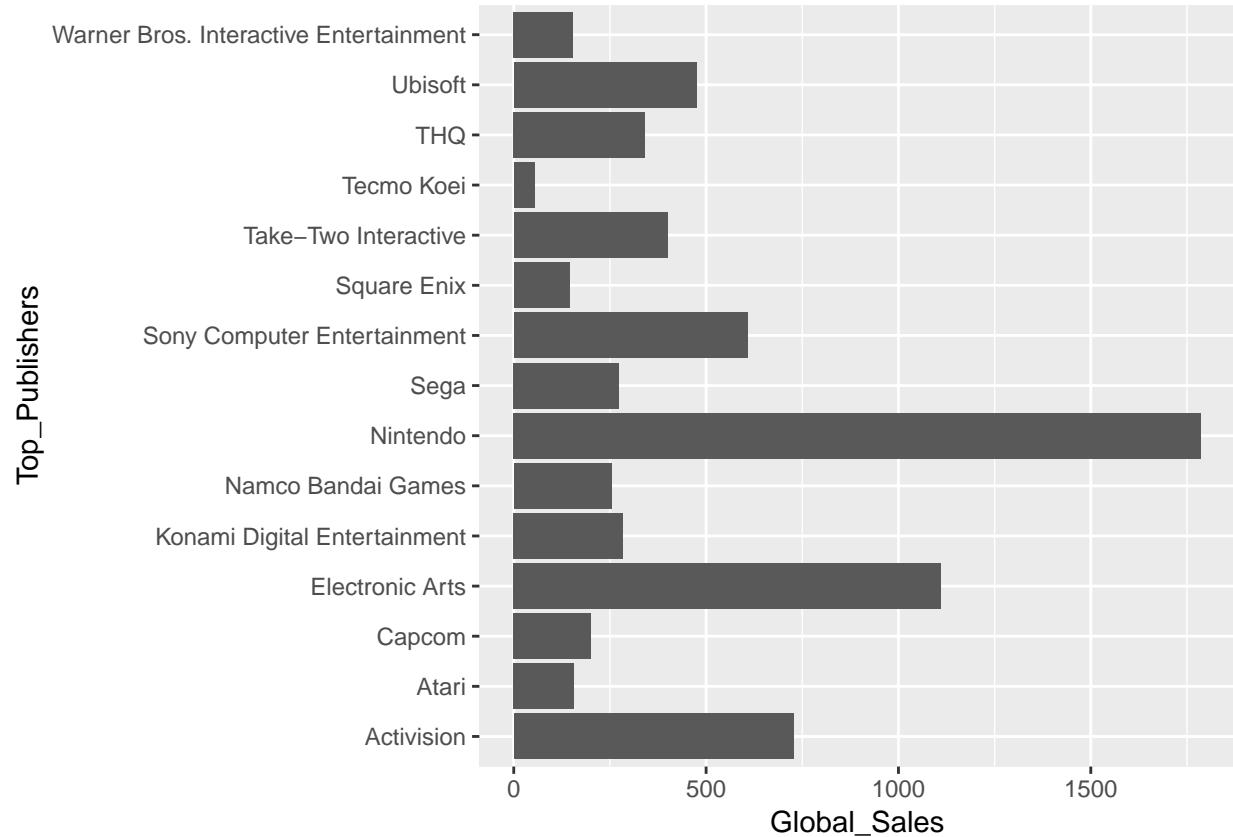


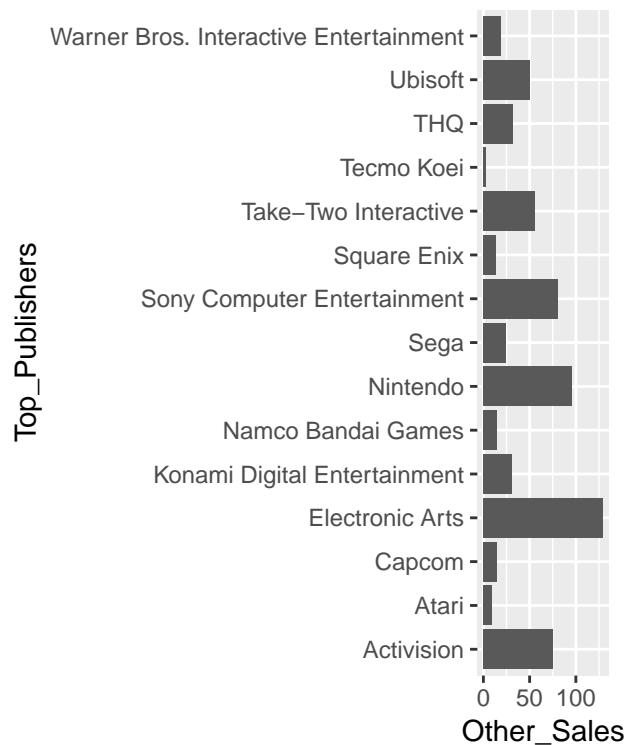
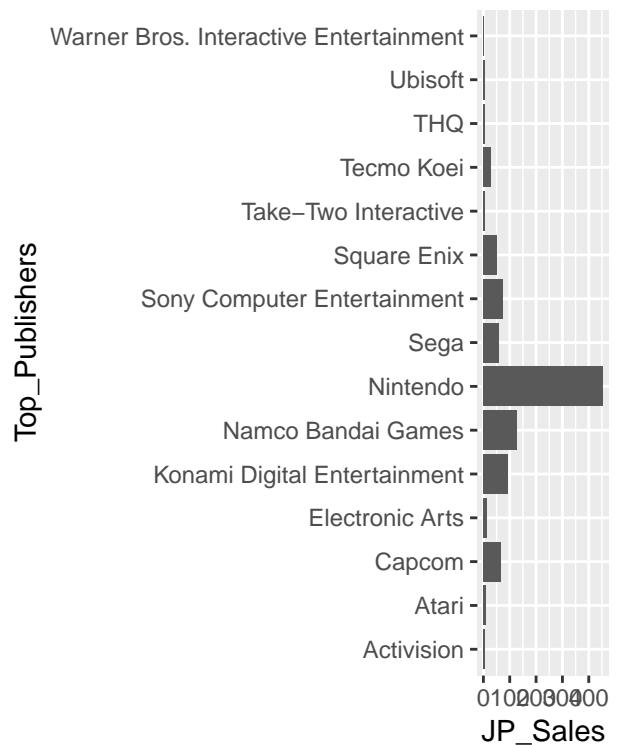
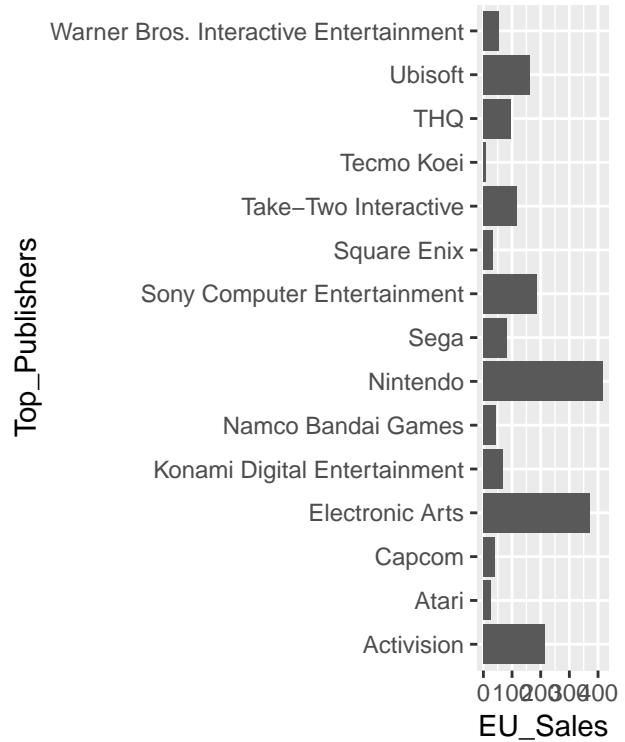
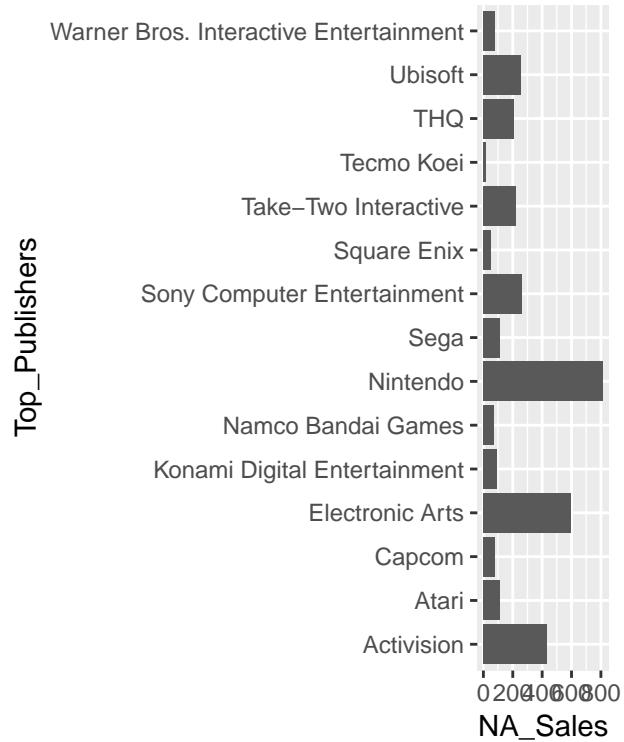
It is interesting that the ratio of total genre sales compared to one another is similar in each region except Japan. RPGs vastly outsell sports, shooters and action games, which are more popular in the other regions.





PC gaming seems to be a much smaller market in Japan compared to the other platforms. It looks as though there were no PC sales in Japan, but according to the dataset, there were 0.17 million dollars worth of PC games sold in Japan.





While Nintendo has strong sales in all regions, it seems to be without equal in Japan.

Bivariate Analysis

Talk about some of the relationships you observed in this part of the investigation. How did the feature(s) of interest vary with other features in the dataset?

Global sales correlated strongly with North American and EU sales.

The ratio of total sales between categories within the platform, genre, platform type, top publishers, and platform makers share a similar bar pattern between the North American, EU and other regions. Japan seems to always have its own pattern.

Within Japan, Japanese based publishers and platform makers seem to do better overall than companies based from outside of Japan. On the other hand Microsoft does very poorly in Japan.

While the role-playing genre is 4th in sales globally, it is 6th in ‘other’ and 7th in North America and the EU. However, it is 1st by a wide margin in Japan. Which probably contributes to its global prominence.

The PC has almost no sales in Japan. And while handheld platform sales are about a quarter of TV platform sales in regions outside Japan, handheld is about two-thirds of TV with Japan.

The only place that Nintendo is not the top publisher is in the ‘other_sales’ category. While Electronic Arts takes the top publisher spot in ‘other_sales’, it still retains a close second in North America and the EU. However, it is only a sliver in Japan.

Did you observe any interesting relationships between the other features (not the main feature(s) of interest)?

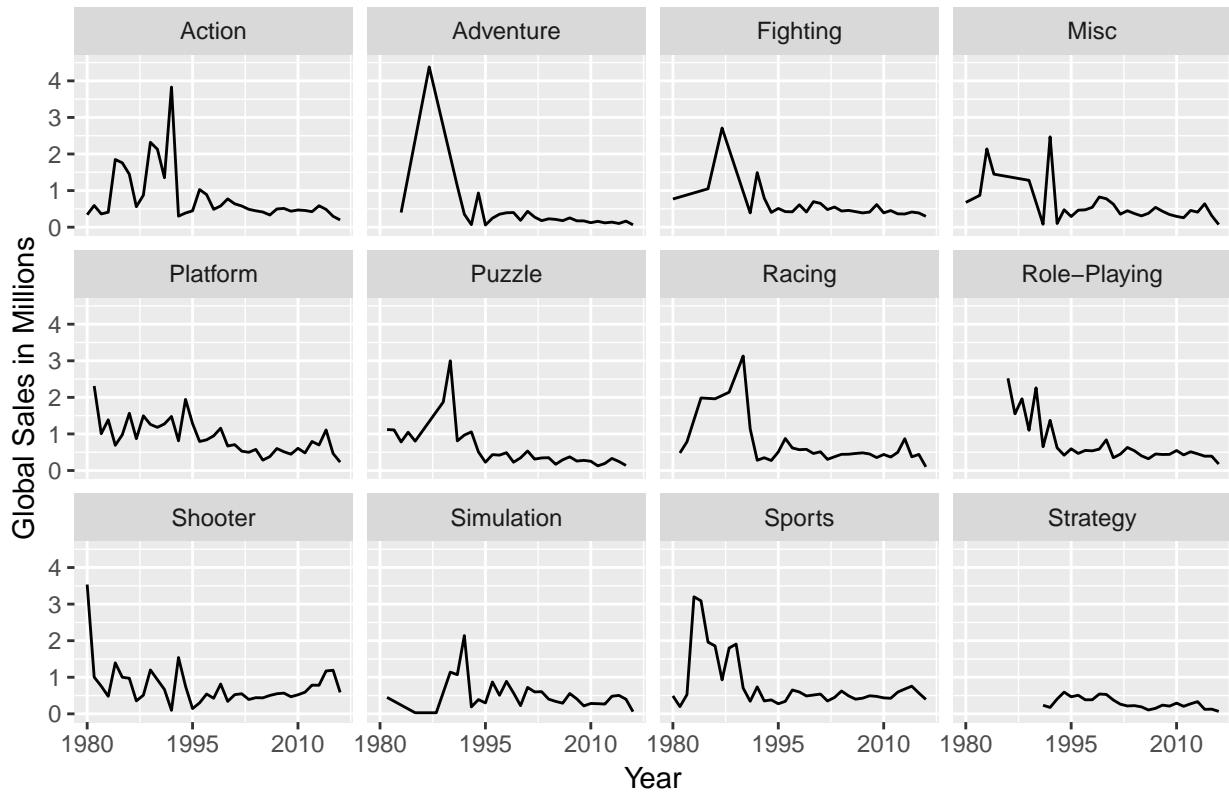
While the ratio of total sales in various categories tend to be very similar in regions outside Japan, the ratio of mean sales (at least for platform makers) seems to be more varied.

What was the strongest relationship you found?

With a correlation coefficient of 0.941, as well as the total sales of any variable being overwhelmingly higher in North America, North American sales really seem to have the strongest force on global sales.

Multivariate Plots Section

Average Global Sales per Genre

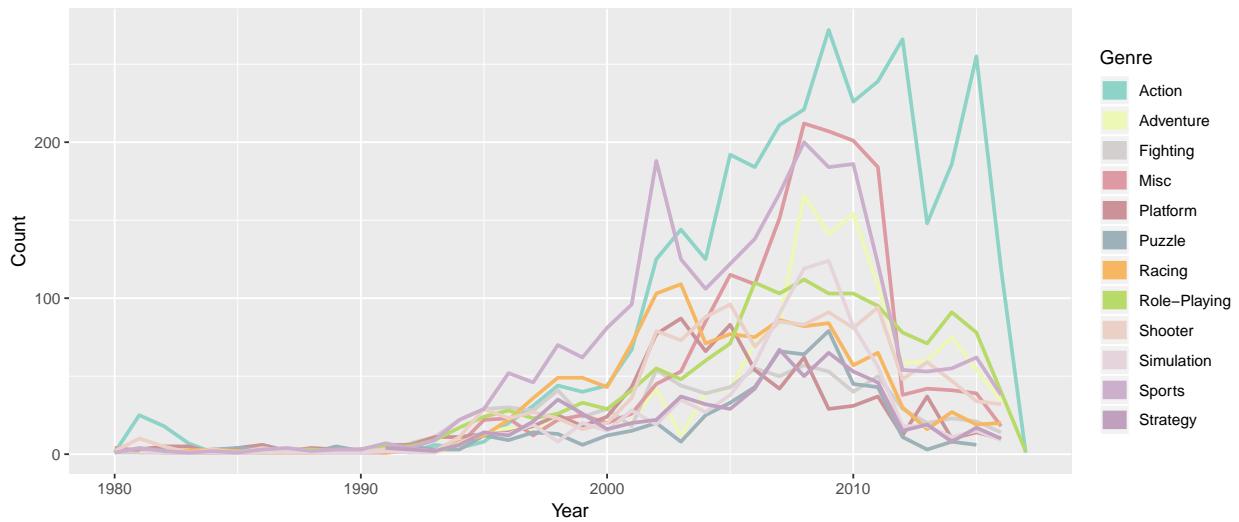


This plot shows the yearly mean of various genres excluding the lower and upper 1% of global sales numbers when calculating the mean. I feel this exclusion gives a more accurate view of the yearly mean sales.

It seems that shooters, and to a lesser extent platformers and sports, are making a slight resurgence.

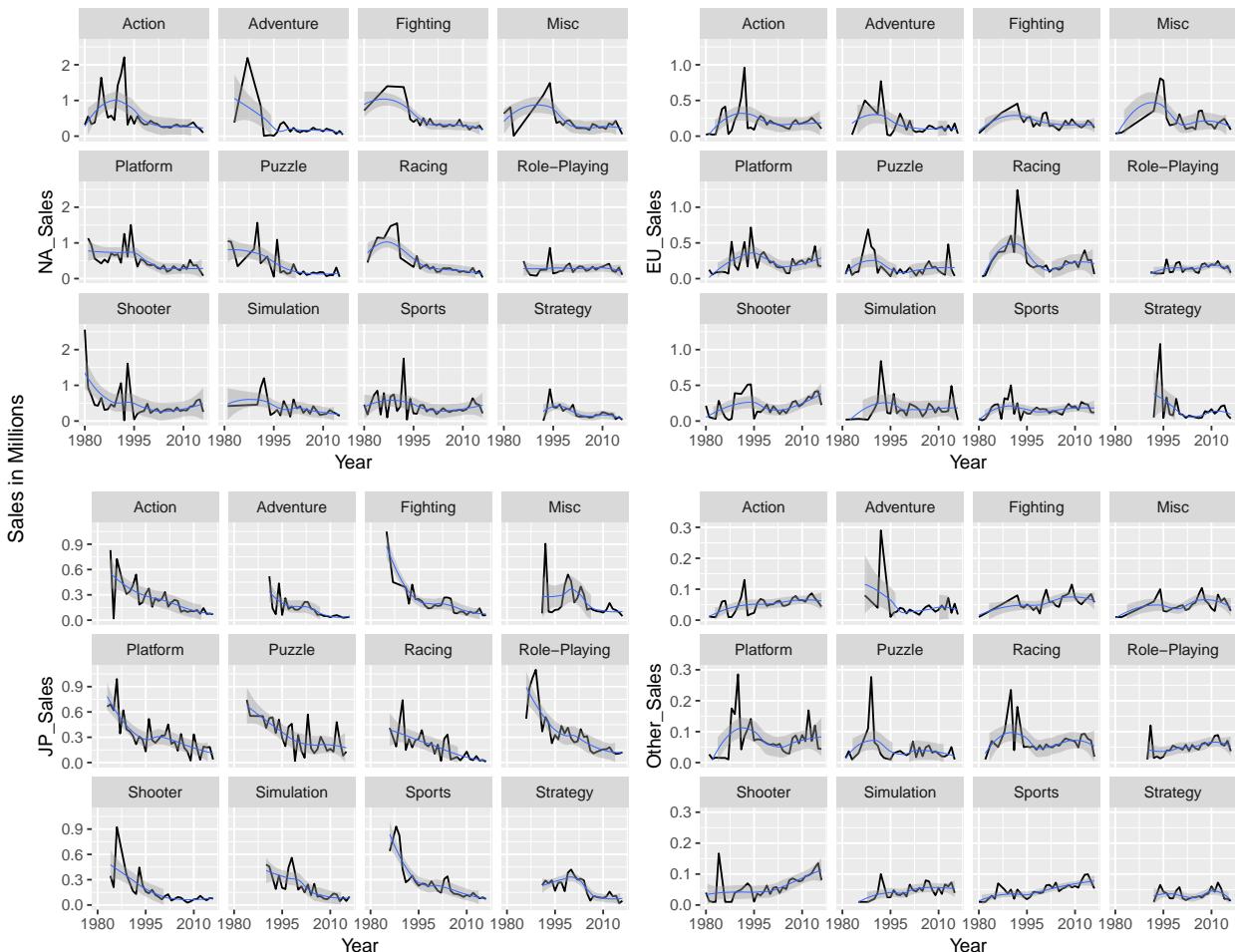
It also seems the early years of global video game sales had a much higher average. I wonder if it has to do with market saturation. Because there seems to be more choices of games on the market now than in the past.

Total Number of Games Released Globally per Genre



This plot shows the total number of games that were released each year. It does seem to show that there was a greater saturation of games in the market, especially in the mid to late 2000s and early 2010s. It looks like there has been a slight decline in all genres, except for action, in the past 6 years or so.

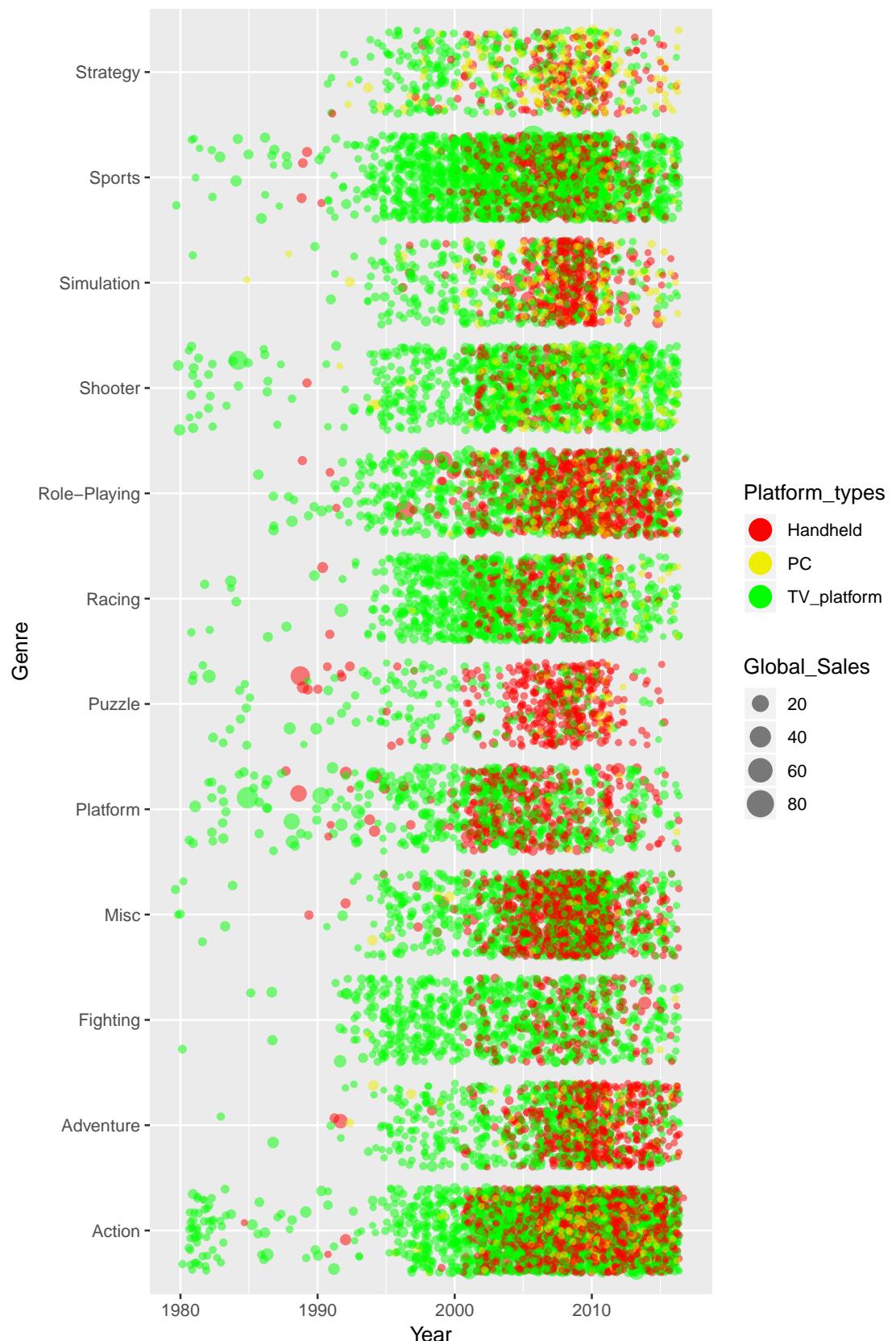
Average Sales per Region per Genre



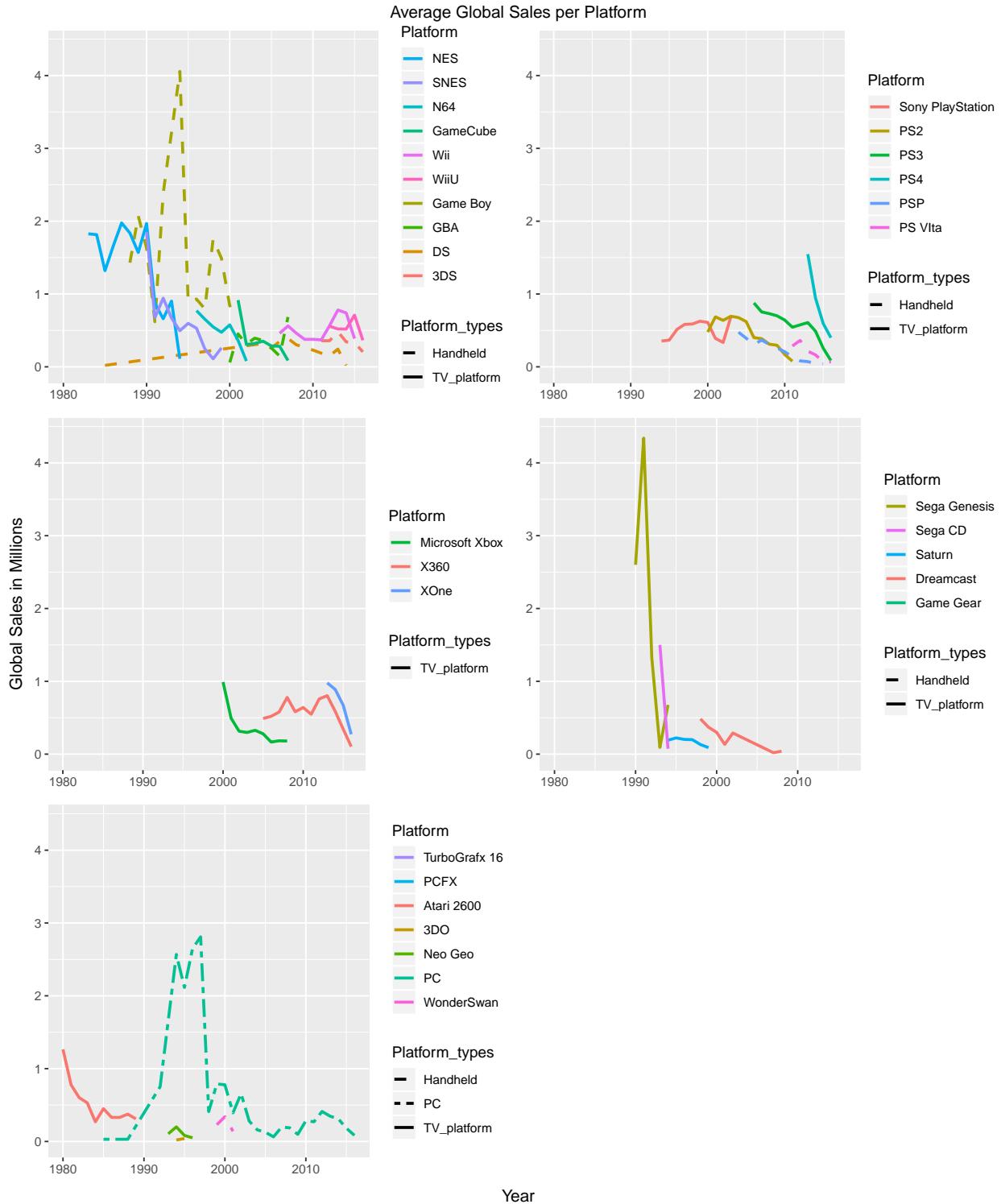
It looks as though the ‘Other’ regions have been experiencing a growth in average sales for the past decade and a half. Which could be an indicator of future sales growth. To a lesser extent the EU has seen growth in some genres while other genres have remained relatively steady. In North America, like the global sales, sales of genres such as shooters, platformers and sports are increasing ever so slightly. Every other genre is relatively steady. Finally in Japan it seems that the only sign of possible growth is in the puzzle genre. Every other genre is remaining steady or declining. Especially the adventure and racing genres.

I’m curious to compare platform type and maker to genre and year.

Global Sales of Platform Types by Genre and Year

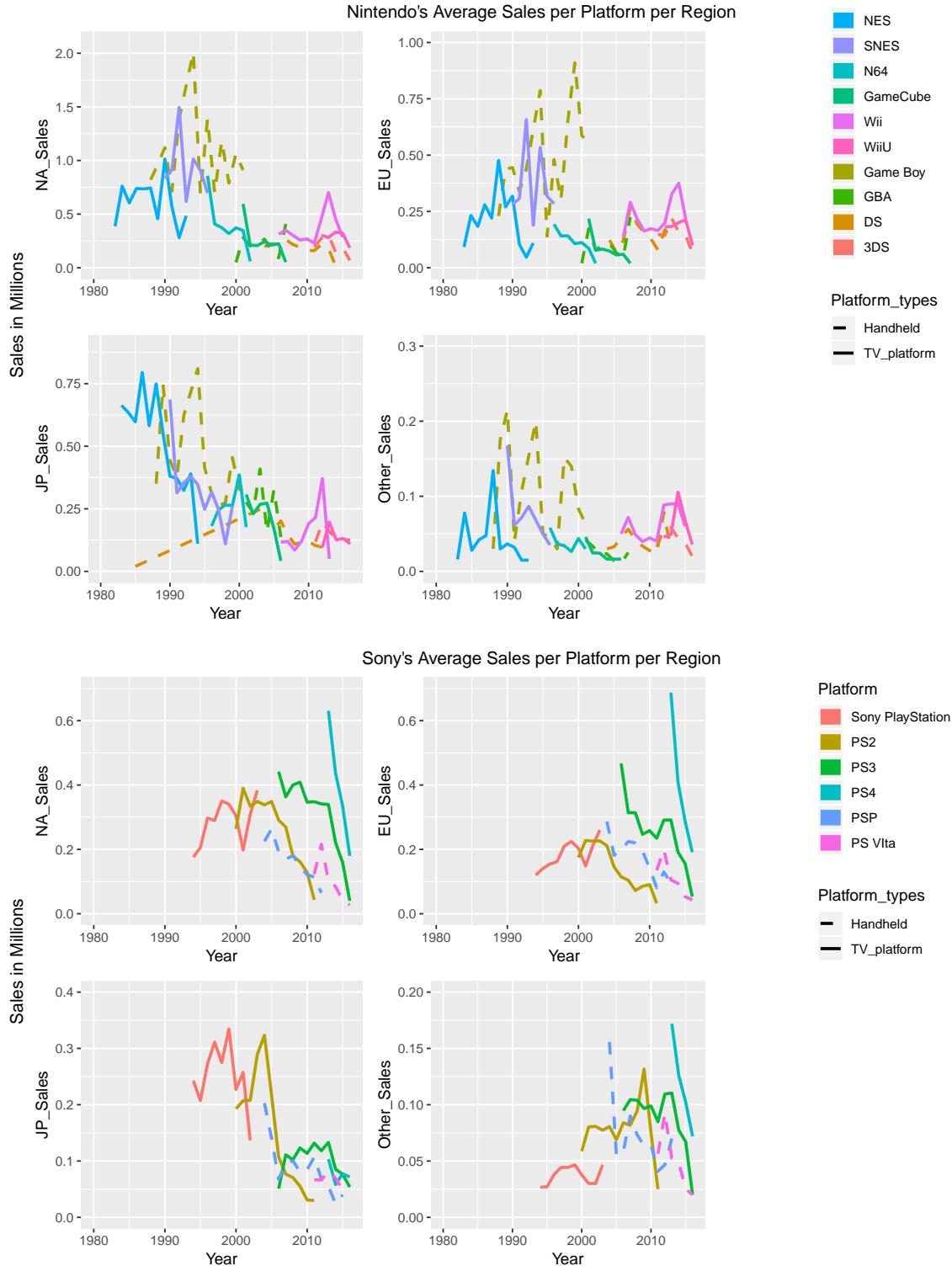


This plot really shows that the 2000s and 2010s are the era of handheld platforms. There also seems to be a decent number of games with a higher global sales in the handheld category. Especially in the puzzle, platform and role-playing genres. PC games, while much fewer than handheld and TV based platforms, have a much larger showing in the strategy, shooter and simulation genres. The strategy genre seems to be the only genre with an almost even balance between the three platform types.

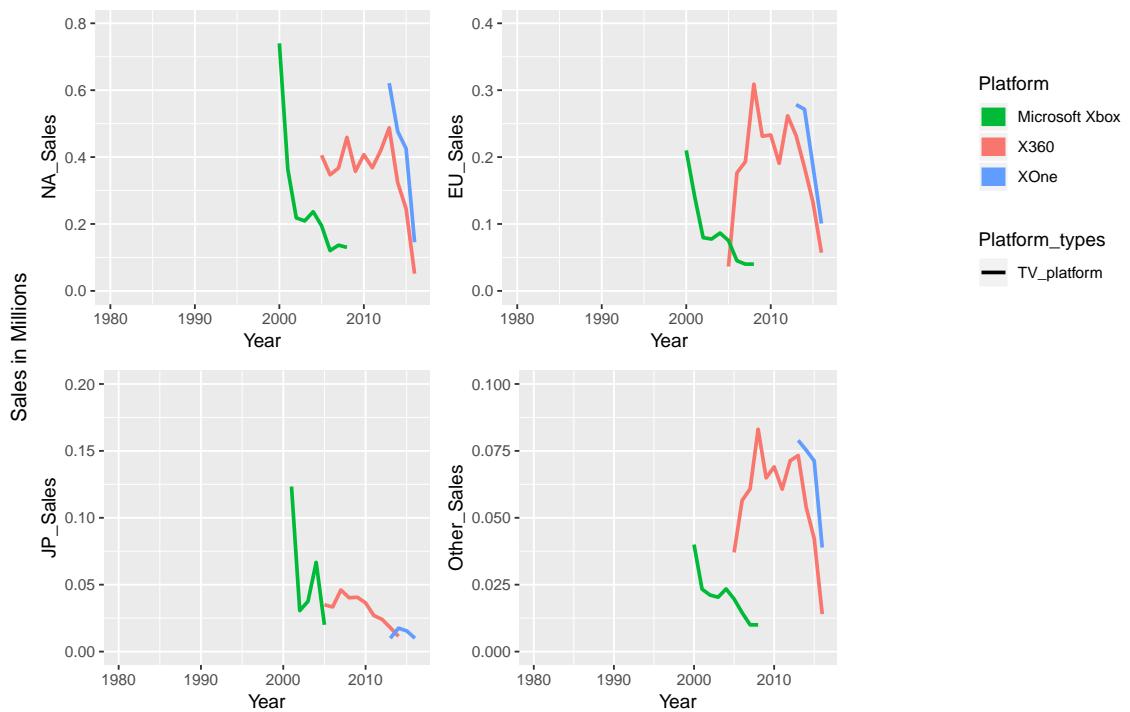


This is a plot of global mean game sales for each platform over the past few decades. From this plot it seems

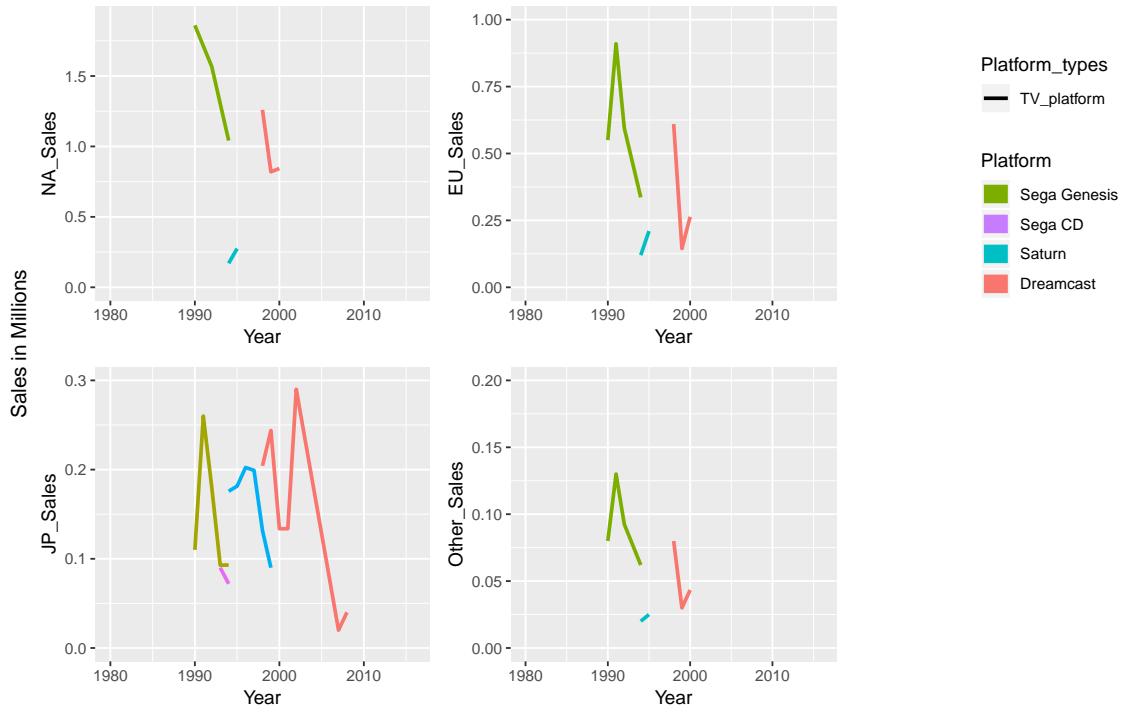
to predict stronger average sales earlier in the platform's lifetime. Though some platforms, like the NES and Game Boy, have strong sales well into the middle part or later part of its lifetime.

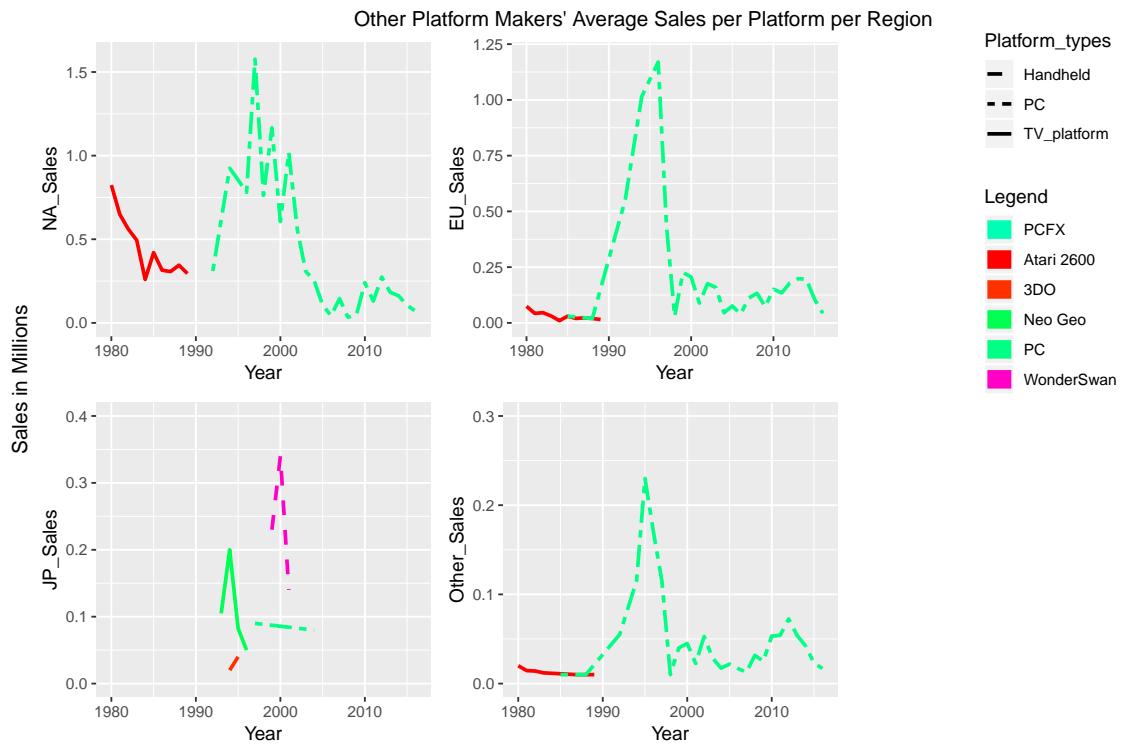


Microsoft's Average Sales per Platform per Region

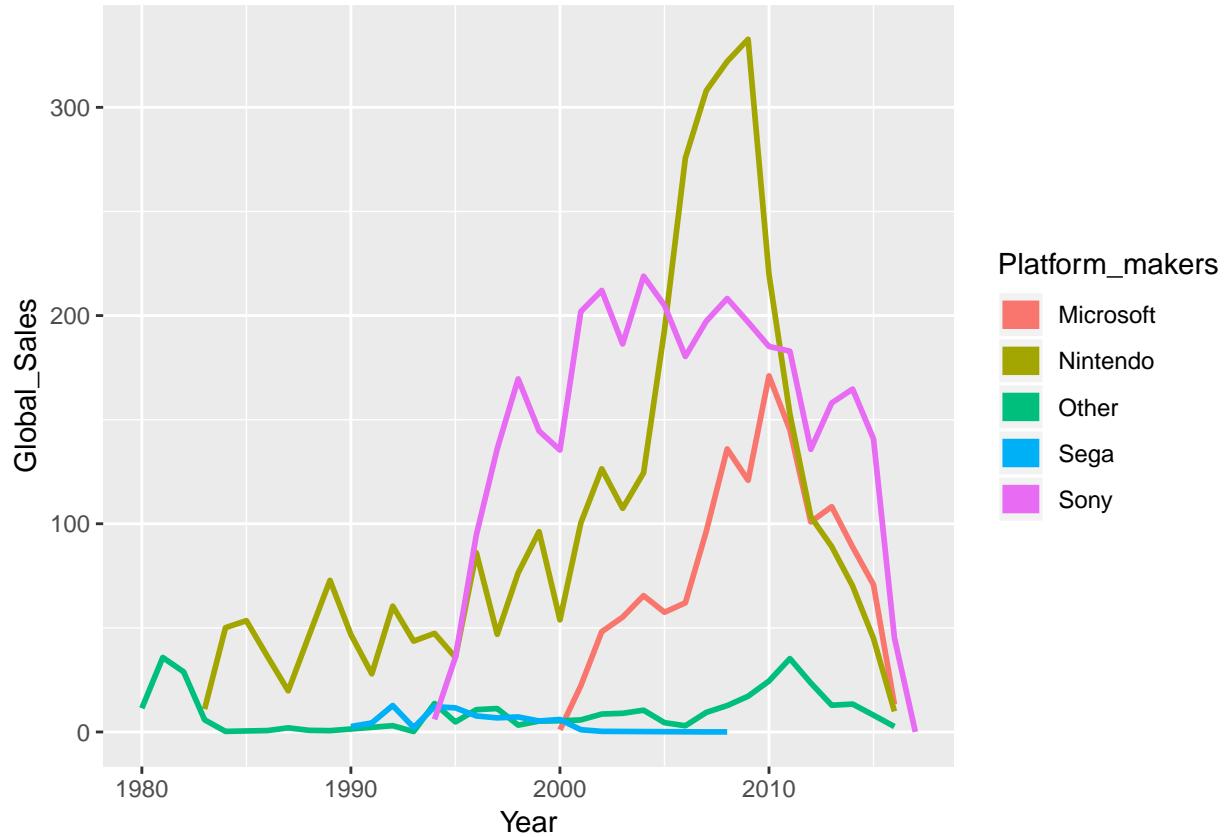


Sega's Average Sales per Platform per Region



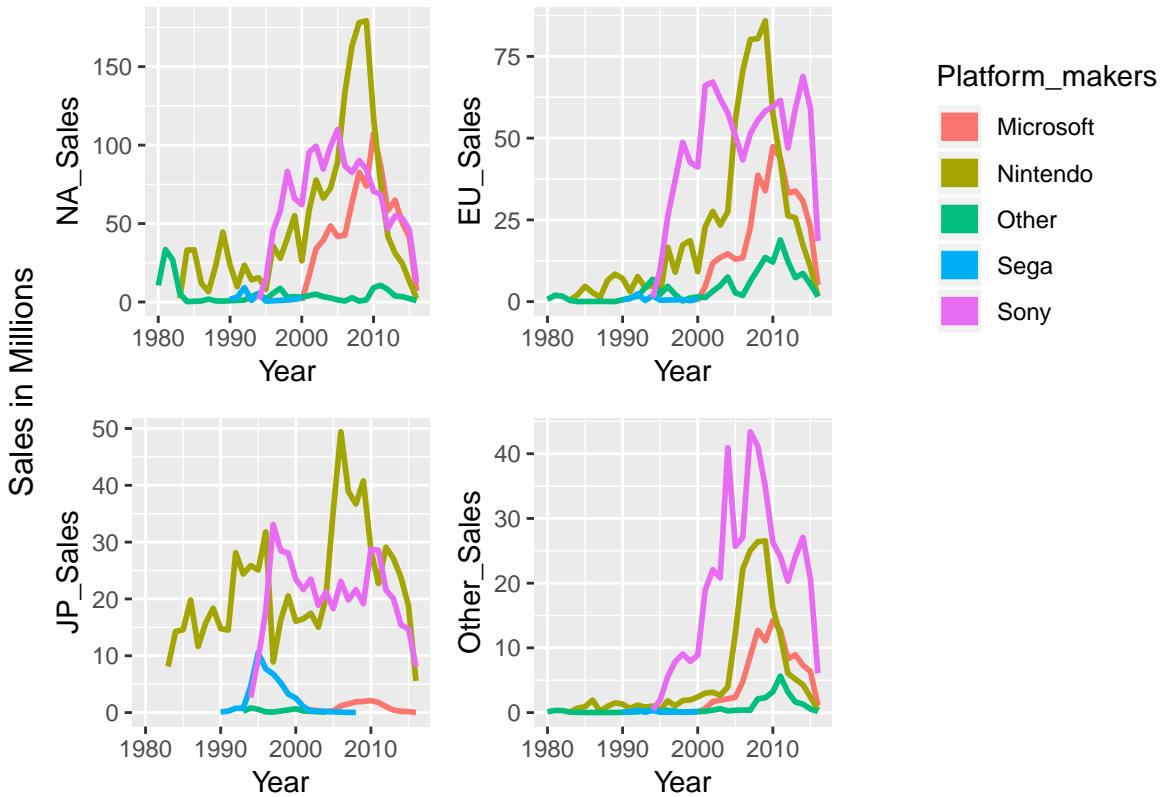


While overall the regional plots seem to follow the same pattern as the global sales, there are some interesting differences. The SNES had a longer lifetime in Japan compared to other regions. Also in Japan, compared to Japanese made platforms, non-Japanese platforms seem to have weaker to no sales in Japan.



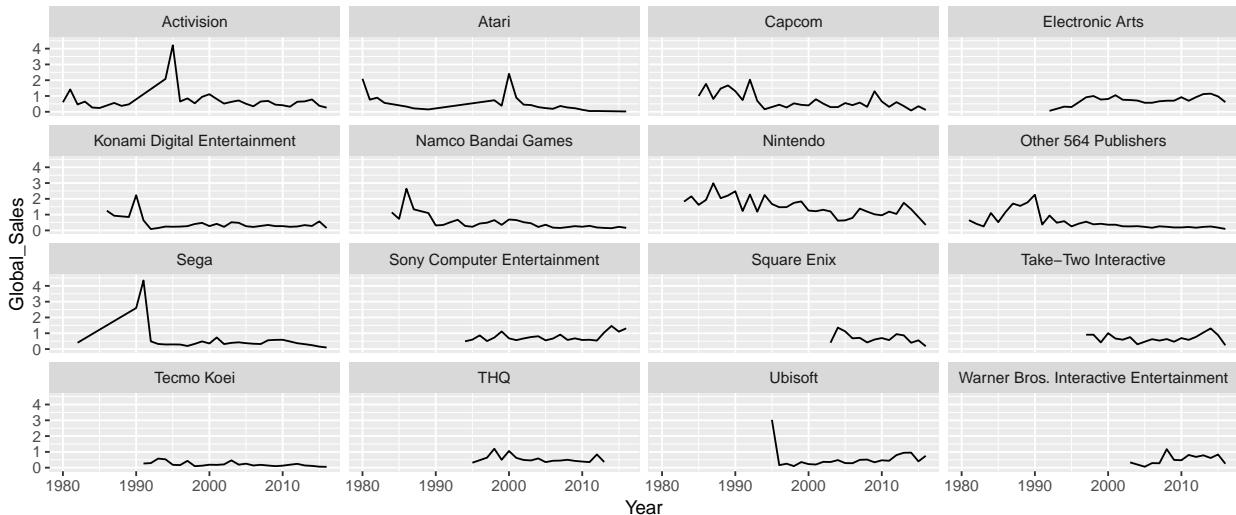
This plot shows the total global game sales for each year for each platform maker. Despite Nintendo having a longer history, games on Sony's platforms seem to have done better overall. Though Nintendo had a stronger peak in the late 2000's.

Average Sales per Platform Maker per Region

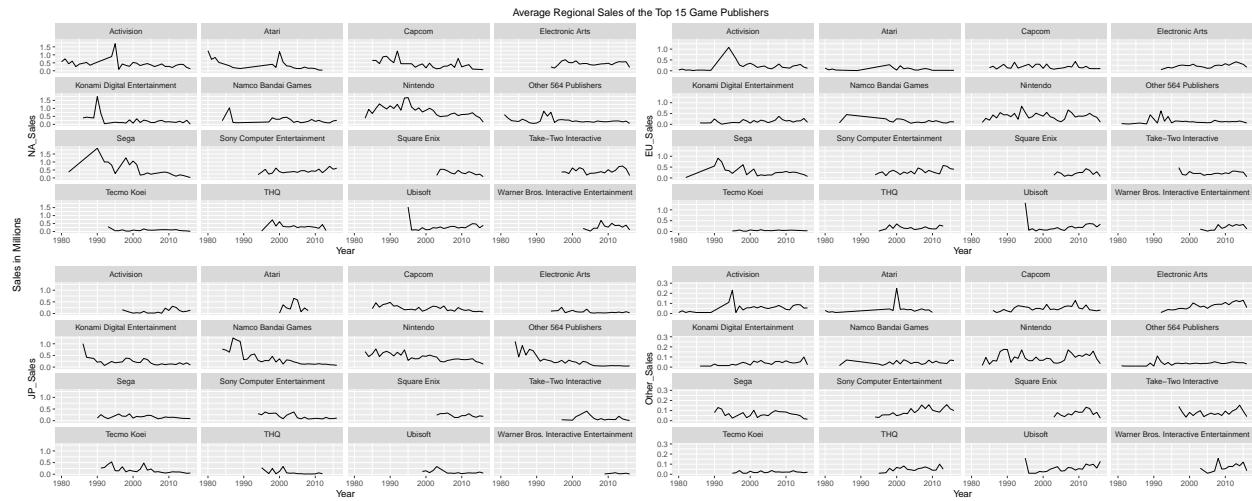


The Japanese plot shows how Japanese customers prefer games on Japanese platforms. Microsoft seems to have no difficulty in regions outside Japan.

Average Global Sales of the Top 15 Game Publishers



Of the top 15 most prolific game publishers, if one was to invest in some of them, it looks as though Electronic Arts, Nintendo, Sony and Take-Two Interactive would be safe bets. Ubisoft and Warner Bros. also look like they are growing into strong companies as well.



Publishers that seem to be growing, outside of Japan at least, are Activision, Electronic Arts, Namco Bandai, Nintendo, Sony, Take-Two Interactive, Ubisoft and Warner Bros. Interestingly, within Japan, there does not seem to be as much growth. Only Konami, Nintendo and Square Enix show signs of growth.

Because this is a list of the top 15 publishers globally, if I had the time, I would like to make a list of the top 15 publishers for each region.

Multivariate Analysis

Talk about some of the relationships you observed in this part of the investigation. Were there features that strengthened each other in terms of looking at your feature(s) of interest?

When looking at the Japanese region, it seems that games published by Japanese companies and/or games published on Japanese platforms seem to do much better sales. Also, total sales seem to be better in the 2000s. While average sales seems to be stronger in the 1980s.

Were there any interesting or surprising interactions between features?

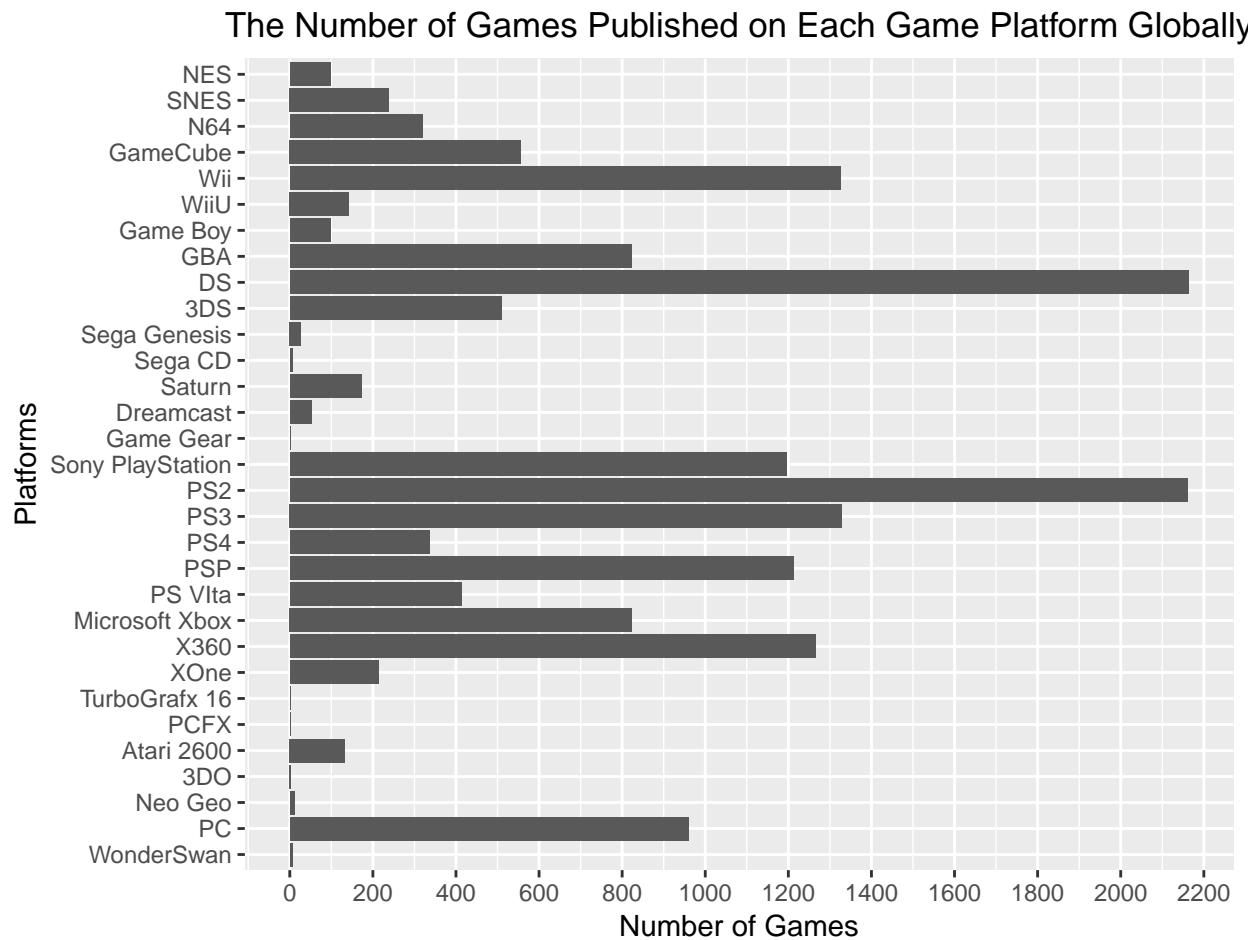
Despite handheld platforms being available since the mid-late 1980s, it was surprising to see that it didn't start rivaling TV platforms until the 2000s.

OPTIONAL: Did you create any models with your dataset? Discuss the strengths and limitations of your model.

I did not. I wish I had the time to.

Final Plots and Summary

Plot One

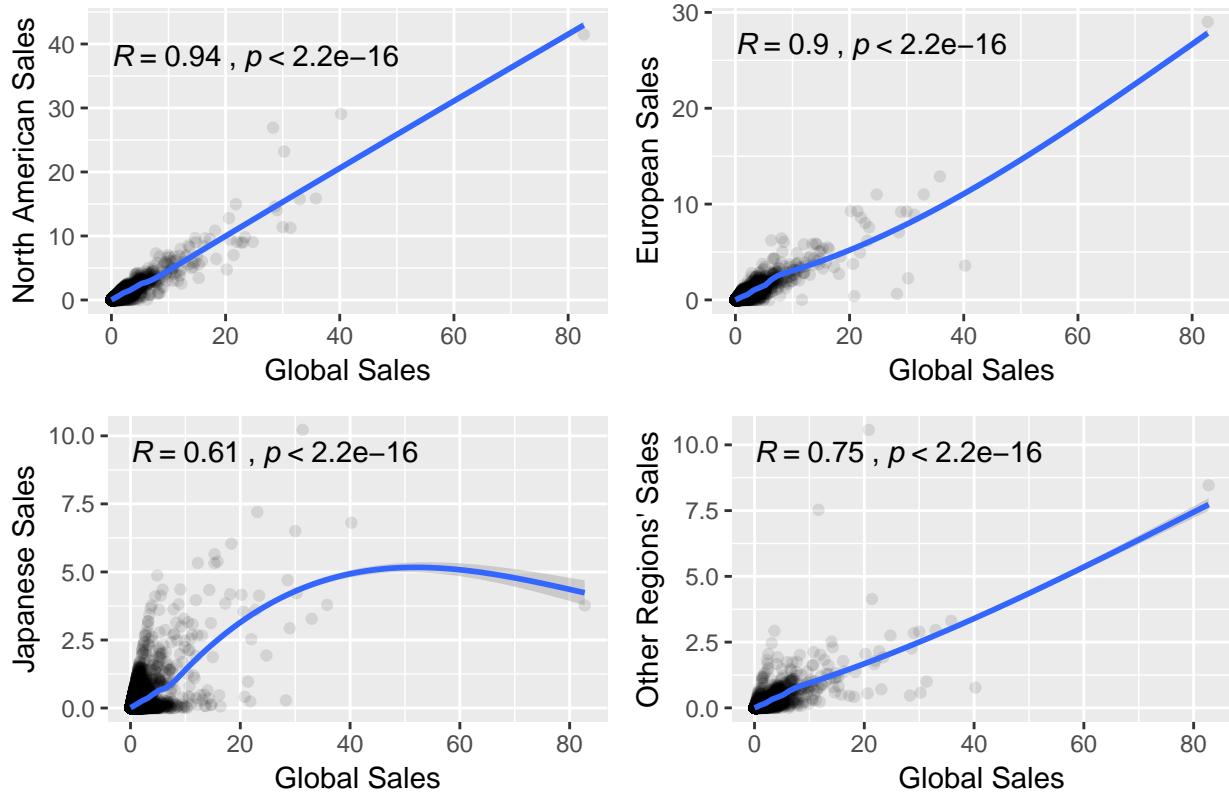


Description One

It is amazing to see how popular the DS and PS2 platforms were for game publishers. However, this plot shows the problem with the dataset. There are too few games. For example, a quick Google search shows there were 714 games released on the NES. Yet the plot shows less than 100.

Plot Two

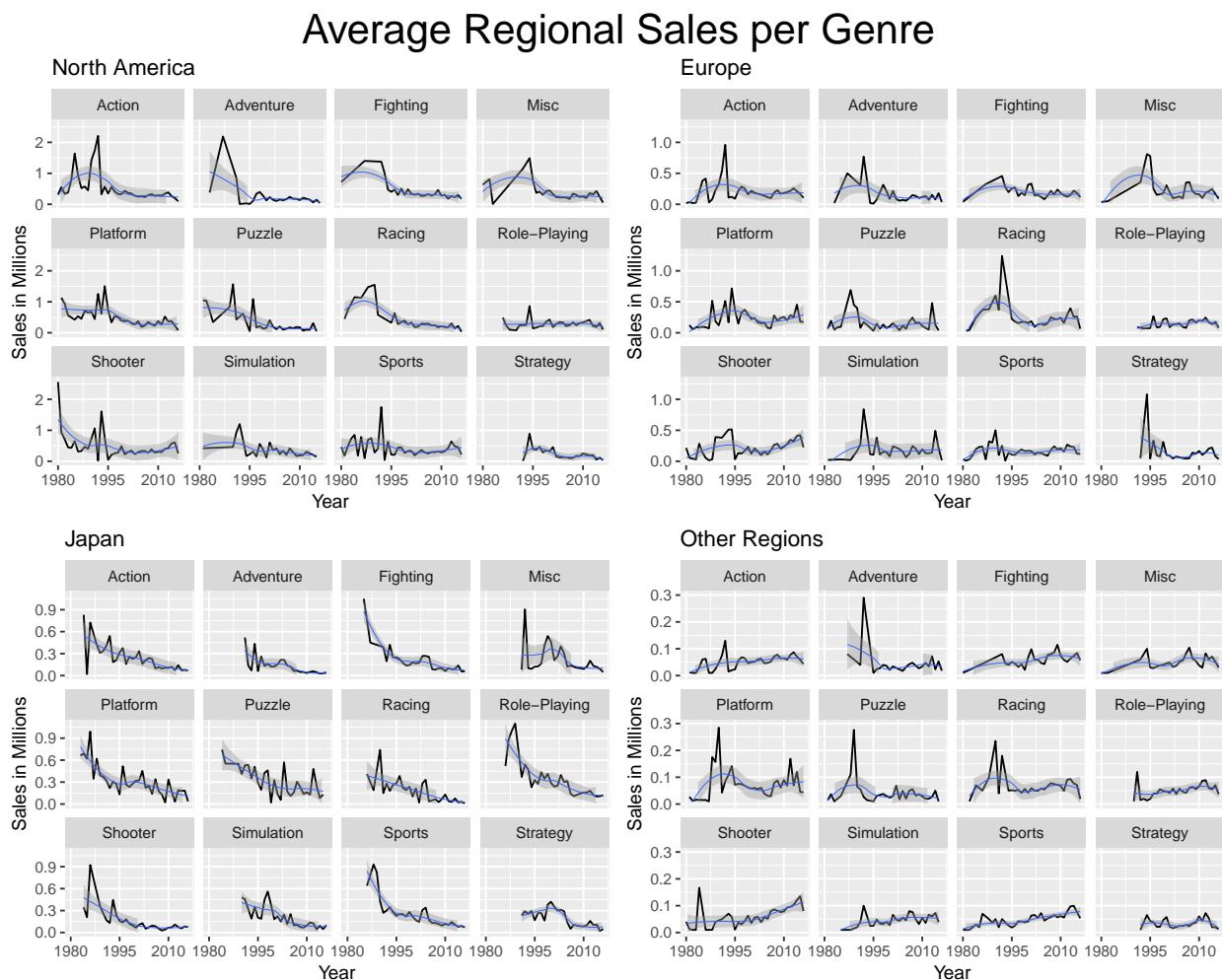
Correlation of Regional Sales to Global Sales (in Millions)



Description Two

North American sales and, to a slightly lesser extent, European sales are strongly correlated with global sales. I interpreted these correlations as game sales in these two regions having a strong effect on global sales figures. On the other hand, Japan seems to have little effect on global sales.

Plot Three



Description Three

Of all the plots I have looked at with this dataset, I think comparing regional differences has been the most interesting. This plot is the most clear example of this. North America's and Europe's lines follows similar patterns. And to a lesser extent, Other Regions' lines too. But Japan seems to follow its own trends. The clearest example of this is comparing the average sales of role-playing games in Japan to the other regions. My interpretation of this is that Japan is less influenced by outside markets than the other regions are. It is neither a strong influencer or influencee.

Reflection

When looking for a dataset to work on, this dataset seemed to be a perfect fit for me. It was interesting, I have some background in the topic and it didn't require too much cleaning to use. However, after creating my first few plots, I realized that there was a large problem with this dataset. It was incomplete. It is not even a very good sample of games released. For example, the Game Gear had 363 games released worldwide. Yet there was only 1 game for the Game Gear in this dataset. A single game is not a good sample of 363 games.

So why did I continue with this dataset? This project's main objective is to 'learn skills to frame and present data'. To complete that objective, this dataset is sufficient. It has over 16,000 games with 14 variables (11

original variables and 3 additional variables I created). The only cleaning I had to do to the dataset was to shift some rows that had not properly been created when running read.csv and fix an incorrect value.

When creating different plots, I found the regional differences to be most interesting. Since North America is the largest and strongest market, global sales seemed to mirror any trends that appeared in the North American sales. And for whatever reason, the European and other region sales data also followed similar patterns as North America. But the most interesting region was Japan. It has the lowest correlation coefficient of the four regions when compared to global sales. When looking at genre, platform, platform maker, etc., regions outside Japan often looked similar visually. But Japan almost always seems to follow its own trends.

In the future, new sales variables that take inflation into account could return interesting results. A developer variable would also be interesting to look at. A more complete dataset, or at least one that represents a better sample, could possibly completely change all my conclusions reached here.