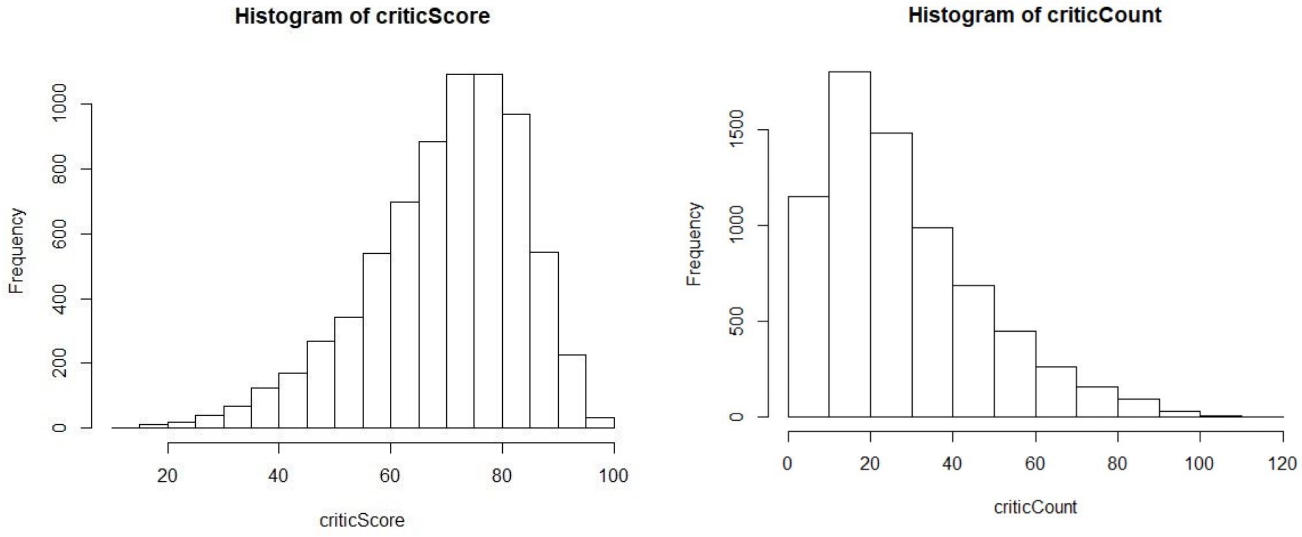
Team:Manipogo

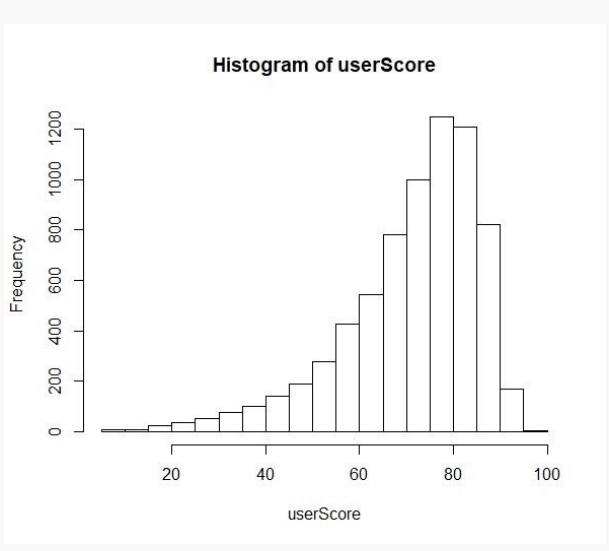
Group Members:	Richard Lee	rlee106@gmail.com
	Deannia Lucas	deannialucas80@gmail.com
	Caglar Cinar	cag.cin@gmail.com
	Hershel Don	hdon@mail.depaul.edu
	Hima Spandana Barla	spandanabarla26@gmail.com

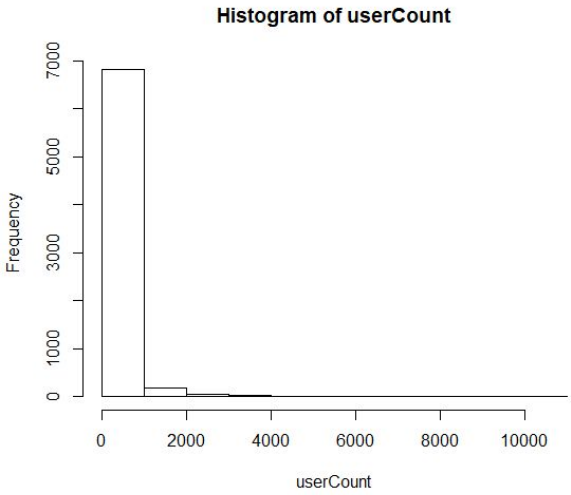
Topic: Video Games

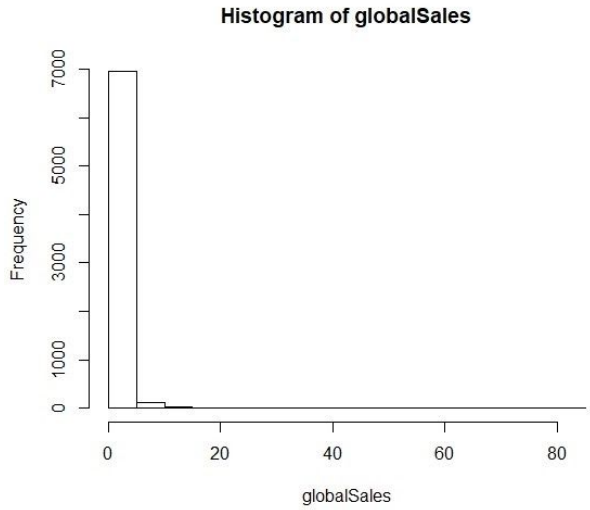
Dataset : Video Games Sales

Fig 1: Histogram plots of all the variables in the dataset




Histogram of userScore
Frequency
0
200
400
600
800
1000
1200
20
40
60
80
100
userScore


Histogram of userCount
Frequency
0
1000
3000
5000
7000
0
2000
4000
6000
8000
10000
userCount


Histogram of globalSales
Frequency
0
1000
3000
5000
7000
0
20
40
60
80
globalSales

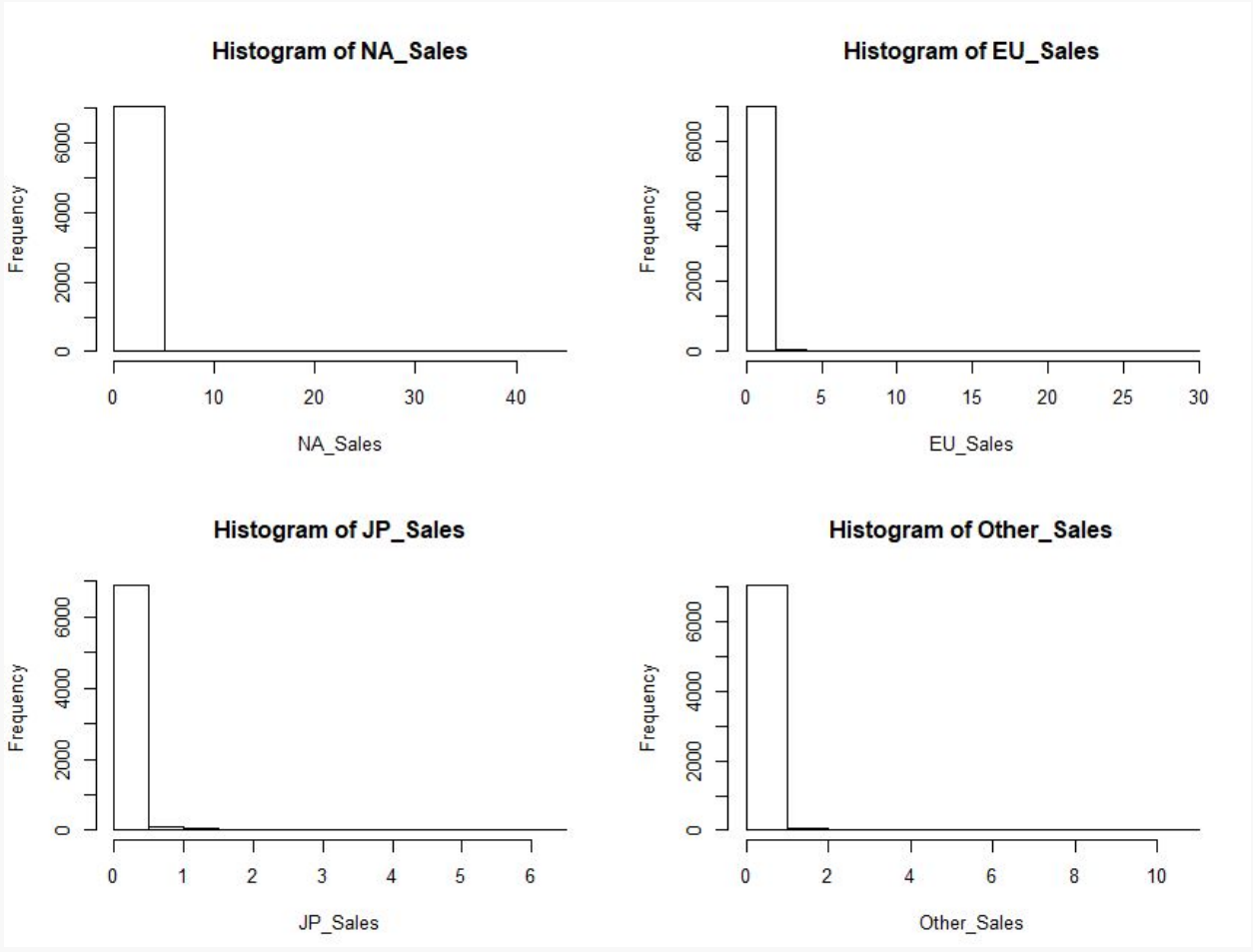
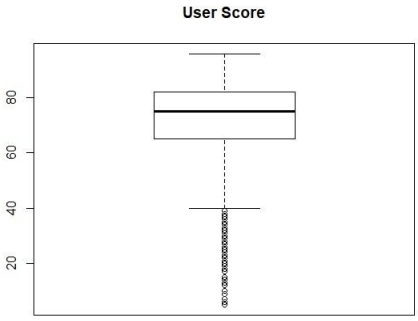
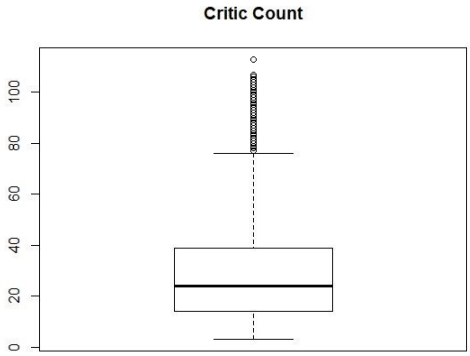


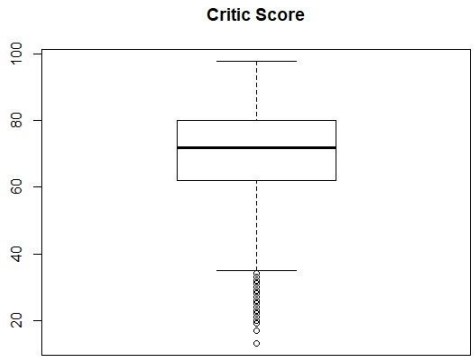
Table1:Summary of each Variable

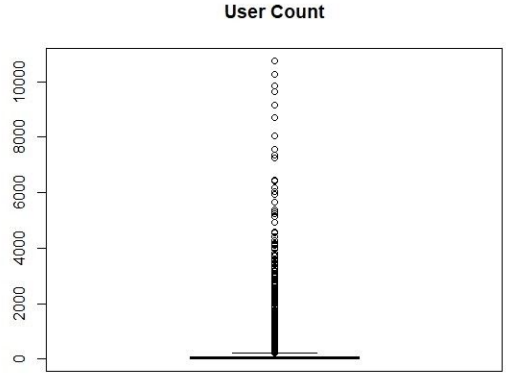
Variable	Min	1st Qu	Median	Mean	3rdQu	Max
Critic Score	13	62	72	70.23	80	98
Critic Count	3	14	24	28.68	39	113
User Score	0.5	6.5	7.5	7.176	8.2	9.6
User Count	3	14	24	28.68	39	113
Global_Sales	0.01	0.11	0.29	0.7653	0.7425	82.54
NA_Sales	0.0	0.06	0.15	0.3886	0.3900	41.360
EU_Sales	0.00	0.020	0.06	0.2325	0.2025	28.9600
JP_Sales	0.0	0.0	0.0	0.06265	0.010	6.500
Other_Sales	0.0	0.01	0.02	0.8135	0.07	10.570

Fig 2: Box plots showing the distributions of variables over two or more classes









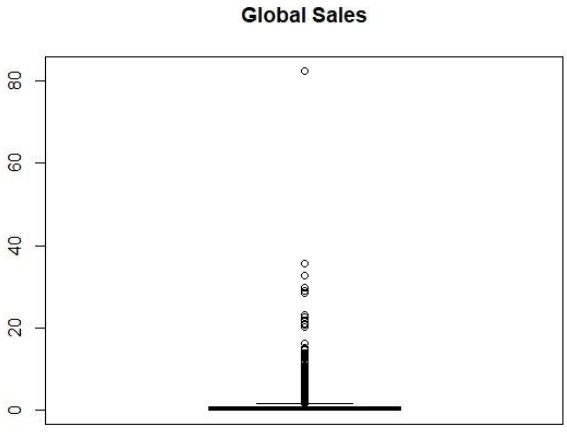
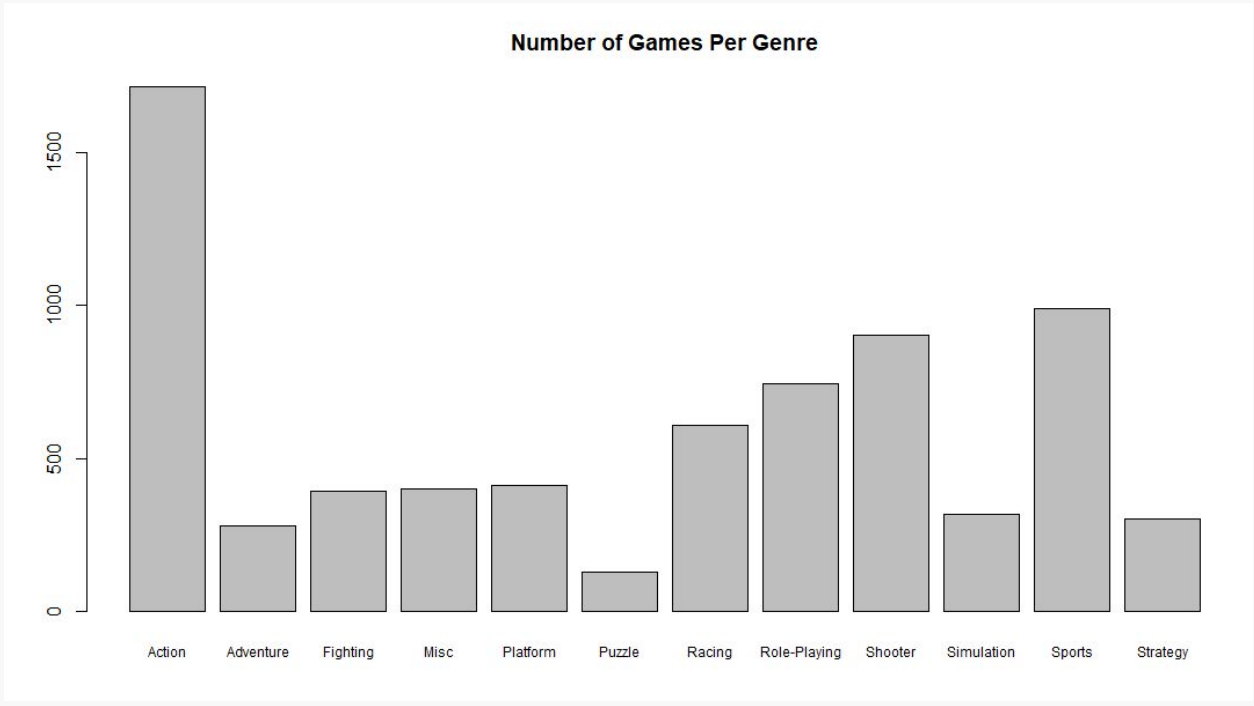
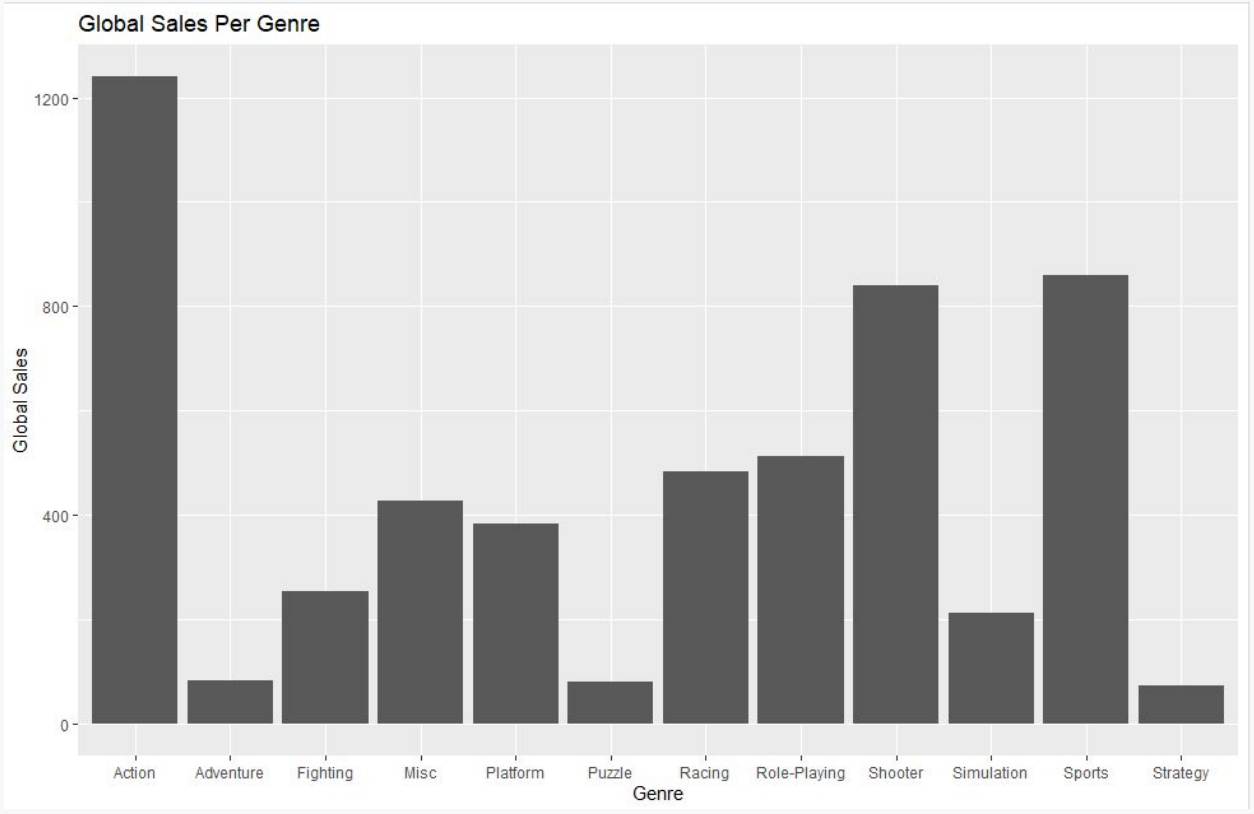
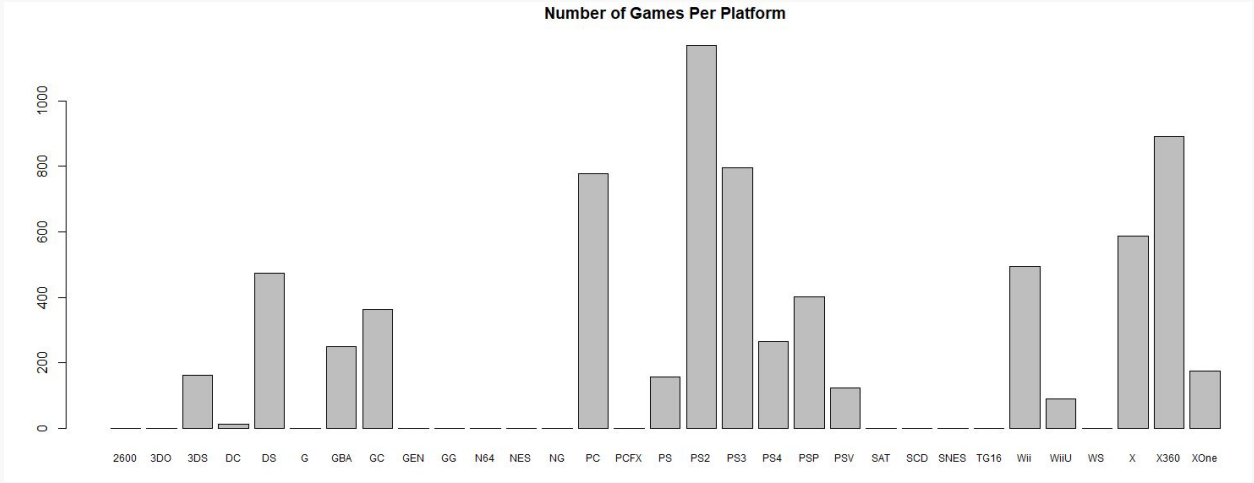


Fig 3: Bar charts showing the distribution of categorical variables.






Number of Games Per Platform
0
200
400
600
800
1000
2600 3DO 3DS DC DS G GBA GC GEN GG N64 NES NG PC PCFX PS PS2 PS3 PS4 PSP PSV SAT SCD SNES TG16 Wii WiiU WS X X360 XOne

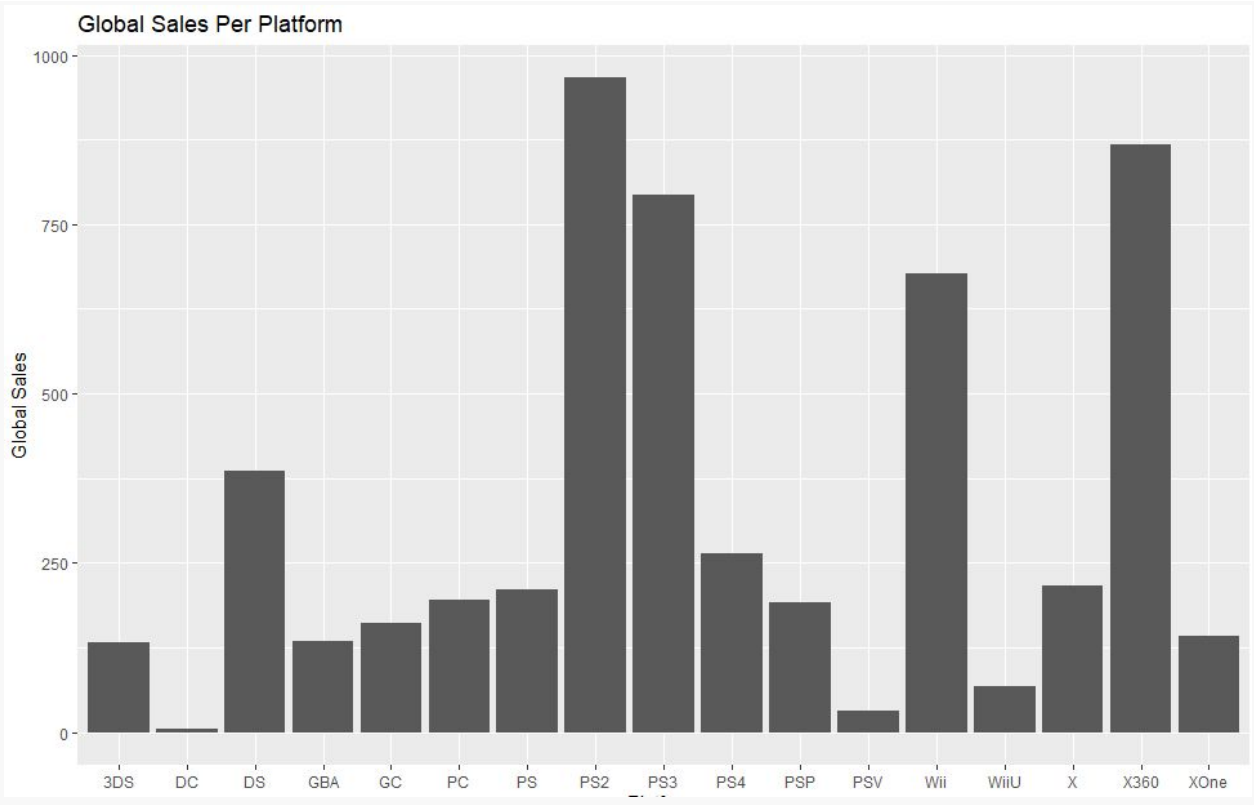
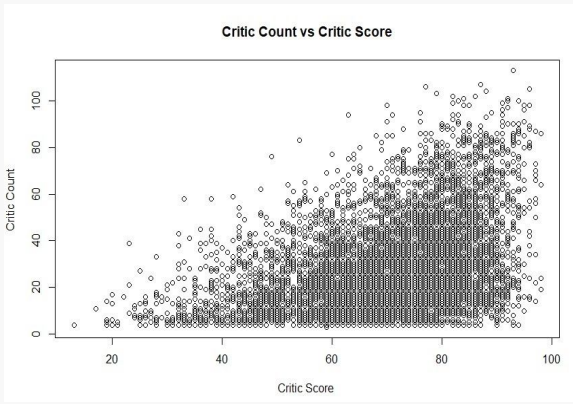
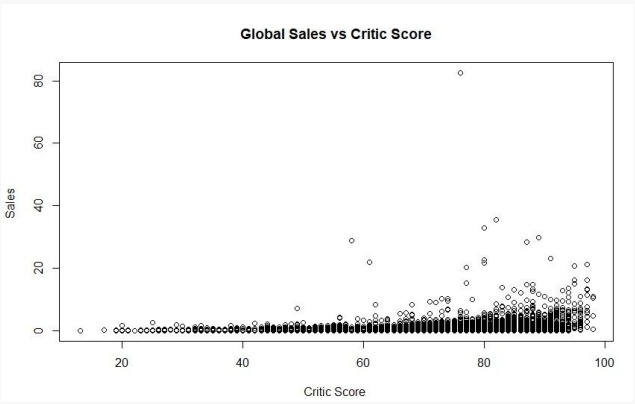
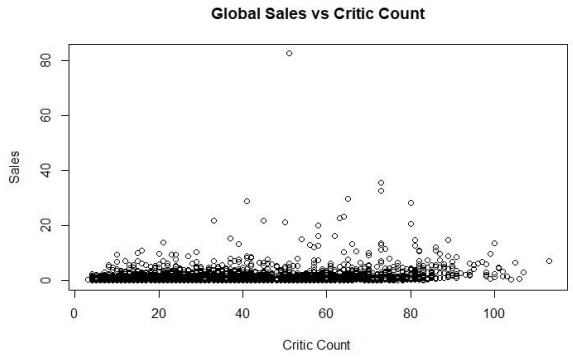

Global Sales Per Platform
Global Sales
0
250
500
750
1000
3DS DC DS GBA GC PC PS PS2 PS3 PS4 PSP PSV Wii WiiU X X360 XOne

Fig 4:Scatter plots showing the relationship between variables.







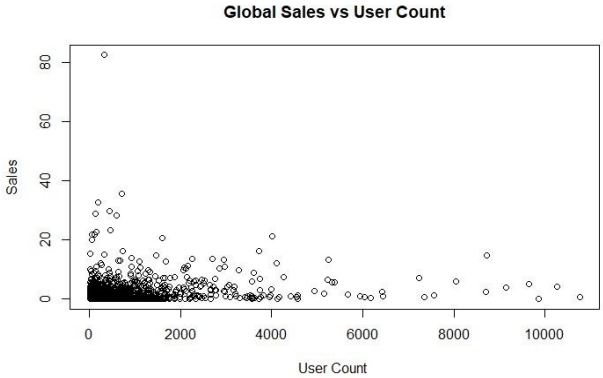
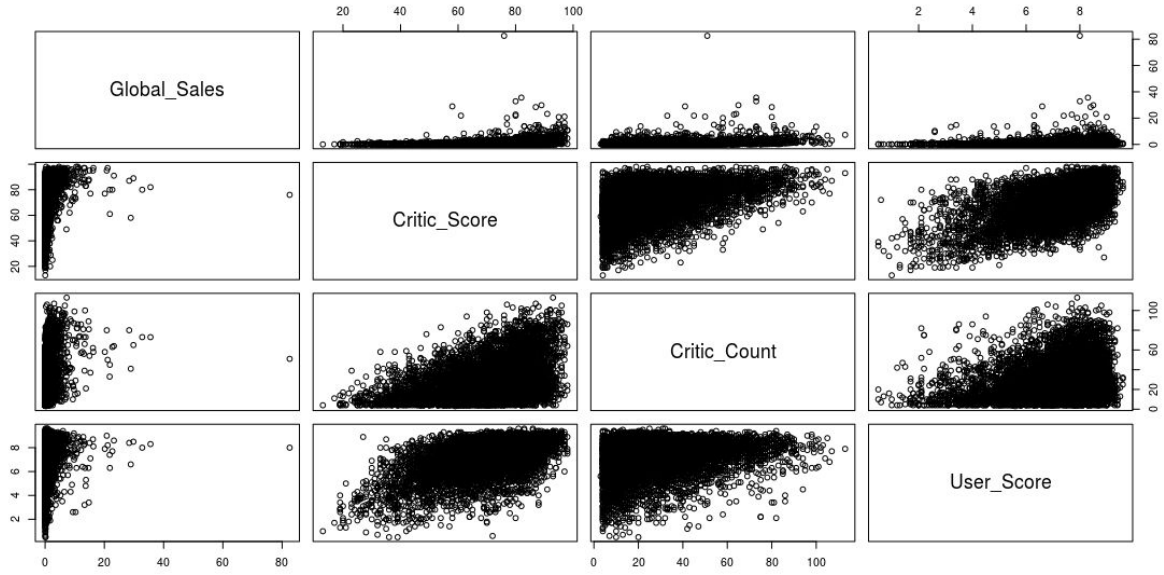


Fig 5:Correlation matrices showing how variables are related.

	Global_Sales	Critic_Score	Critic_Count	User_Score
Global_Sales	1.00000000	0.2368053	0.2934365	0.08821127
Critic_Score	0.23680526	1.0000000	0.3906699	0.58270534
Critic_Count	0.29343651	0.3906699	1.0000000	0.19361887
User_Score	0.08821127	0.5827053	0.1936189	1.00000000



Scatter plot observation

We can see on the scatterplots that global sales vs users score shows a weak positive association. The sales of video games decreases as user score increases. There is also the same association between global sales and critic score. The association between critic count and critic score shows a positive relationship. The critic count increases as the critic score increases.

Based on these scatter plot observations we believe that there is a positive correlation between critic scores and global sales as a video game that is given a high score does well in sales relative to its high score. On the other hand the correlation is not the same between user scores and global sales. It looks like there is a weak positive correlation but it then dips down towards the end. So this tells us that some video game sales with good user scores may not have done well based on some other variable. We have also done an analysis based on the critic count vs the critic score which can correlate to the global sales of video games. Based on the scatterplot, the data represents a positive correlation but not relatively a strong one.

Box plot observation

We also can see that based on the observations for the boxplots, that there is an overwhelming amount of data far outside every variable boxplot. What this indicates is that there is a large amount of outliers being shown within each boxplot variable. In particular, the extreme's which are abnormal are evident within the global sales and the user count variable box plots. For these extreme cases, we can see that the box is almost non-existent on the image and that it mostly consists of data outside of the box which are outliers far exceeding where the box is visible.

Correlation matrix

This shows correlations between the negative and positive relationships observed between variables. In our example, the main observable pattern shows that each variable perfectly correlates with itself. User score and critic score are the only two variables that show a high degree of correlation.

Bar Plot Observation

The Bar plot for the genre clearly shows that there are a lot more games available in some genres than others, such as there are more than twice the amount of games with Action genre than games with Role-Playing, and the amount of games with Role-Playing tag is more than twice of the ones with Strategy. That might be because some games sell a lot more than the others, thus companies make more of one genre than the other. Also you can see that some games with certain genres get sold less even though there are more games available. E.g. Adventure, Strategy.

Similarly, there are a lot more games available to some platform than the others. And this gap might even get bigger if we combine some platforms together such as PS2 and PS3. Also probably the games that are only available to certain platforms affect the sales.

Histogram Observation

Histograms show the distribution of relevant numeric variables. Sales are very heavily skewed with not much of a noticeable tail, in every region. The scores and counts from users and critics, respectively, appear to be more like normal distributions. The scores are left-skewed, while the counts are right-skewed. The count of user reviews for most games are relatively low, shown in the highly asymmetric histogram.

Predictions

We think that User & Critic Scores as well as the number of each will all correlate positively with sales in a multiple regression model. We will examine models for global sales as well as sales for each region specified. Perhaps differences will emerge between regions with respect to the strength or number of opinions of users and critics. Within the genre side of data, we think that the number of specific genre games correlates to that same genre in global sales. For example there is a high number of action/shooter genre games which can correlate to the high number of sales within action/shooter genre games. Whereas strategy, there is a low number of strategy games which correlates to a low amount of global sales within the strategy genre.