Preprocess

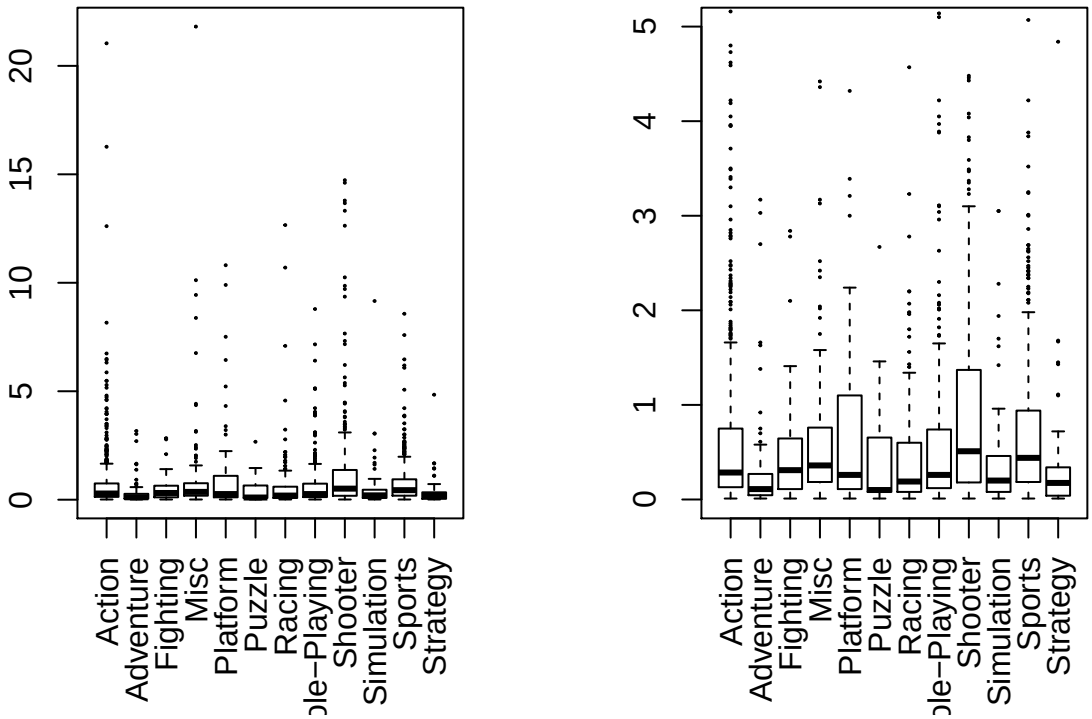
Group 5

7/22/2020

```
## Warning in `[.data.frame`(Vdata, as.numeric(as.character(Vdata$Year_of_Release))
## > : NAs introduced by coercion
```

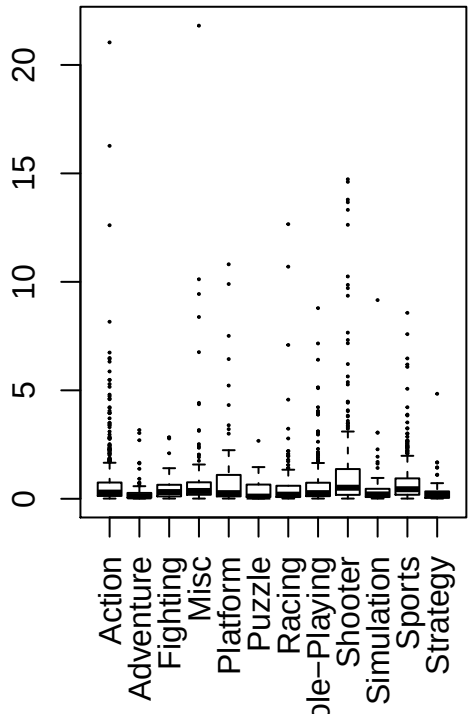
Test if group variables have no effect on Global_Sale

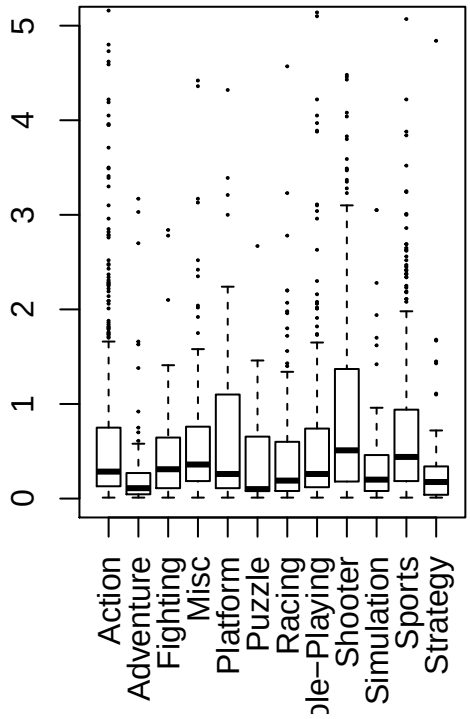
Genre




```
##                      Df Sum Sq Mean Sq F value   Pr(>F)
## data[[classifier]]   11    173  15.766   6.371 1.76e-10 ***
## Residuals          2300   5692   2.475
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

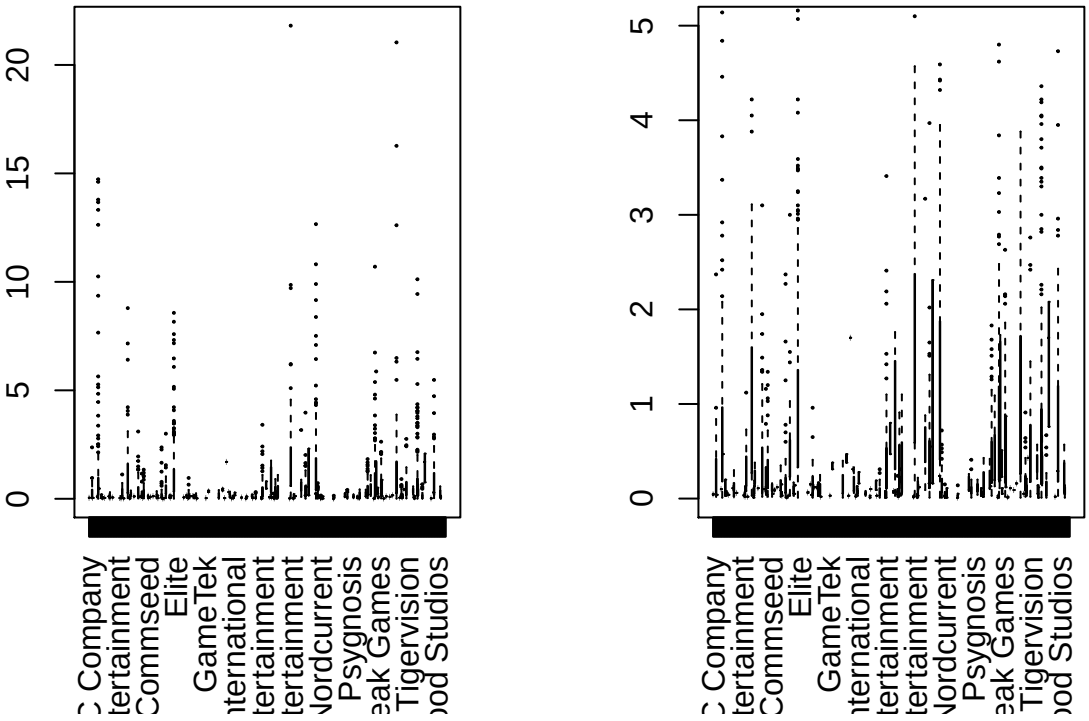
Platform





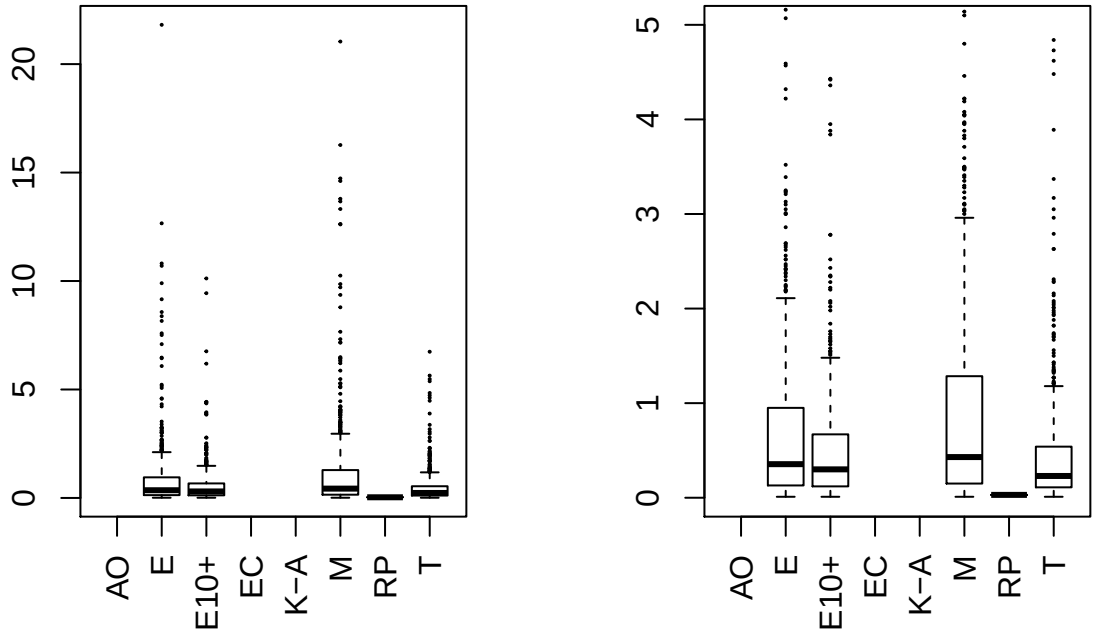

```
##                       Df Sum Sq Mean Sq F value   Pr(>F)
## data[[classifier]]    11    173  15.766   6.371 1.76e-10 ***
## Residuals           2300   5692   2.475
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Publisher




```
##                       Df Sum Sq Mean Sq F value   Pr(>F)
## data[[classifier]]  138    729   5.286   2.236 9.91e-14 ***
## Residuals          2173   5136   2.363
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

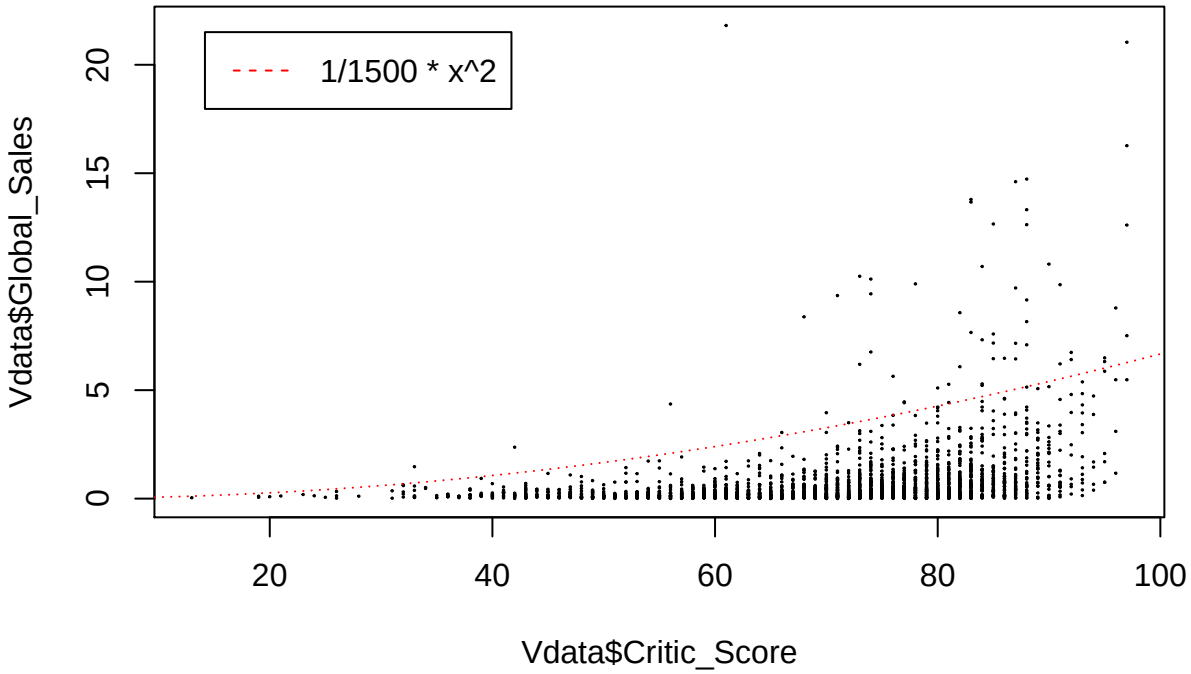
Rating

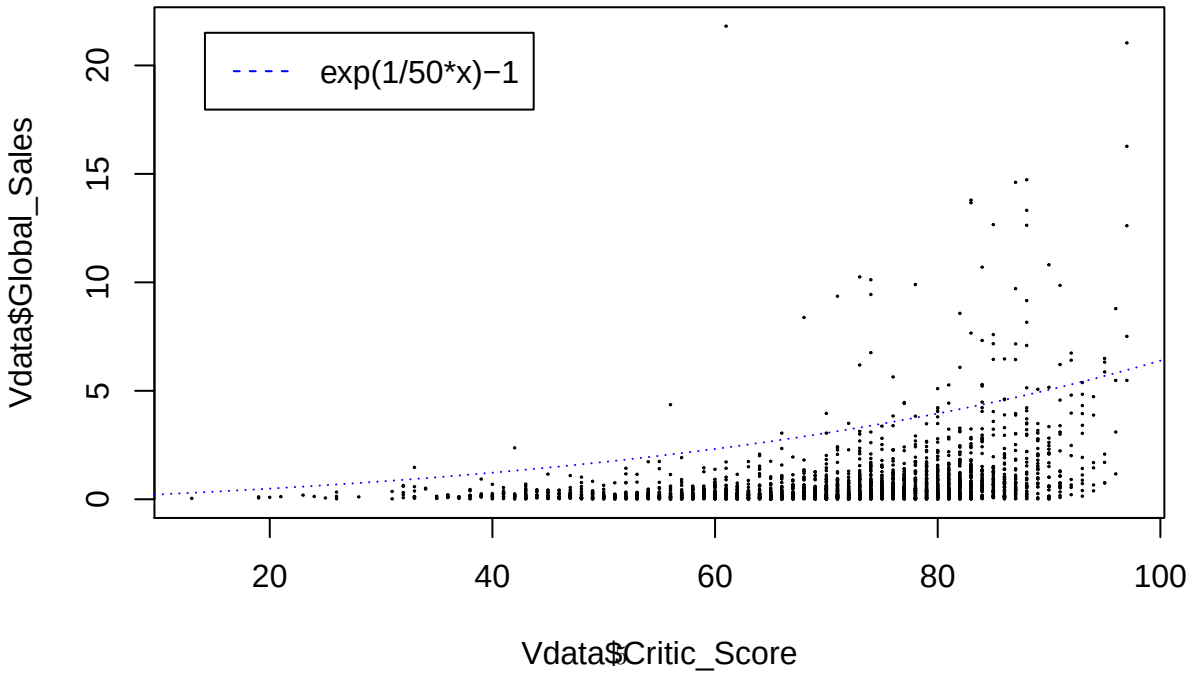



```
##                       Df Sum Sq Mean Sq F value Pr(>F)
## data[[classifier]]     5    215   43.03   17.56 <2e-16 ***
## Residuals           2306   5650    2.45
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Find some relationship

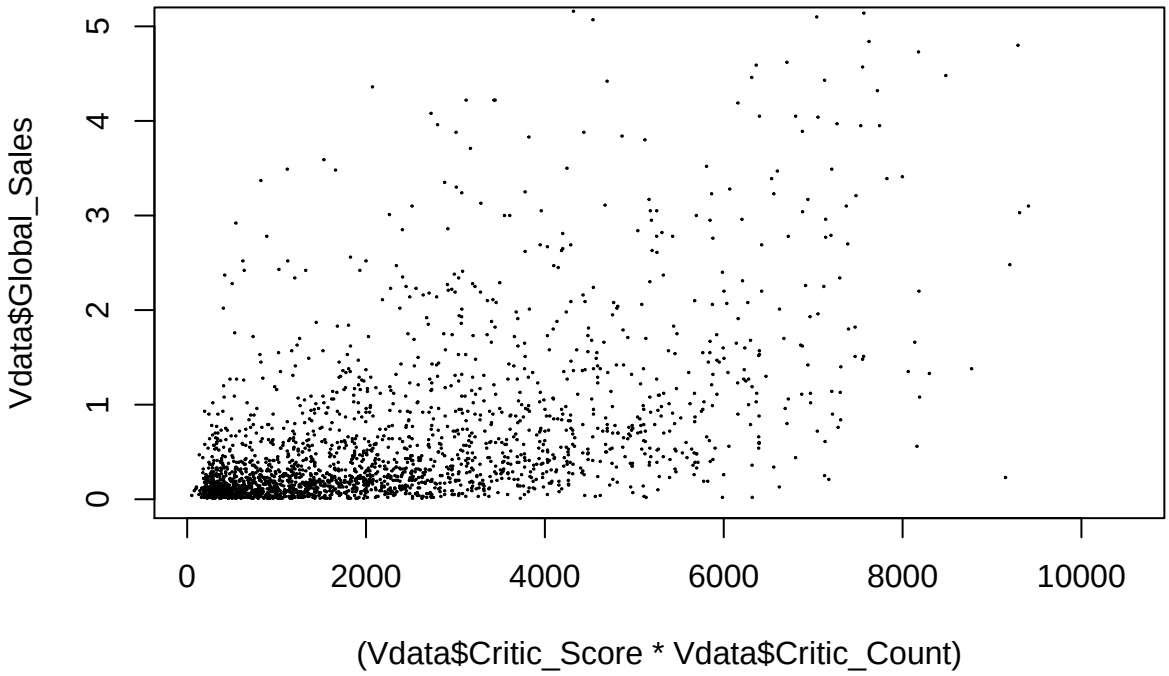
Critic_Score defines an upperbound



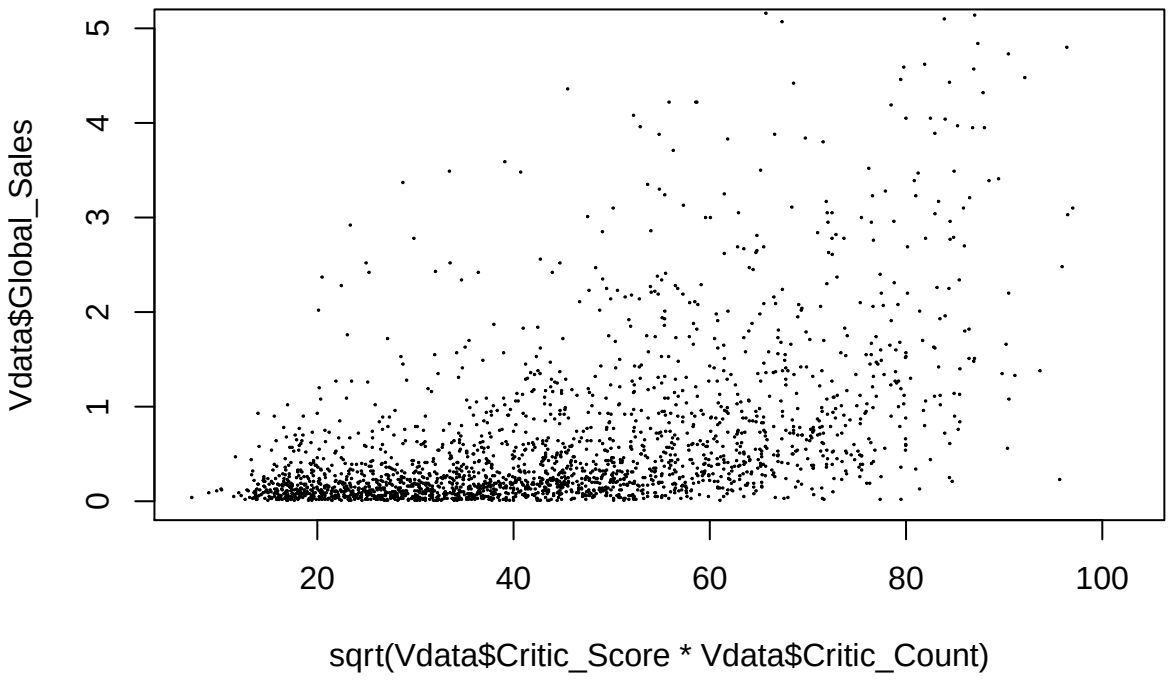


Critic_Score * Critic_count no useful finding

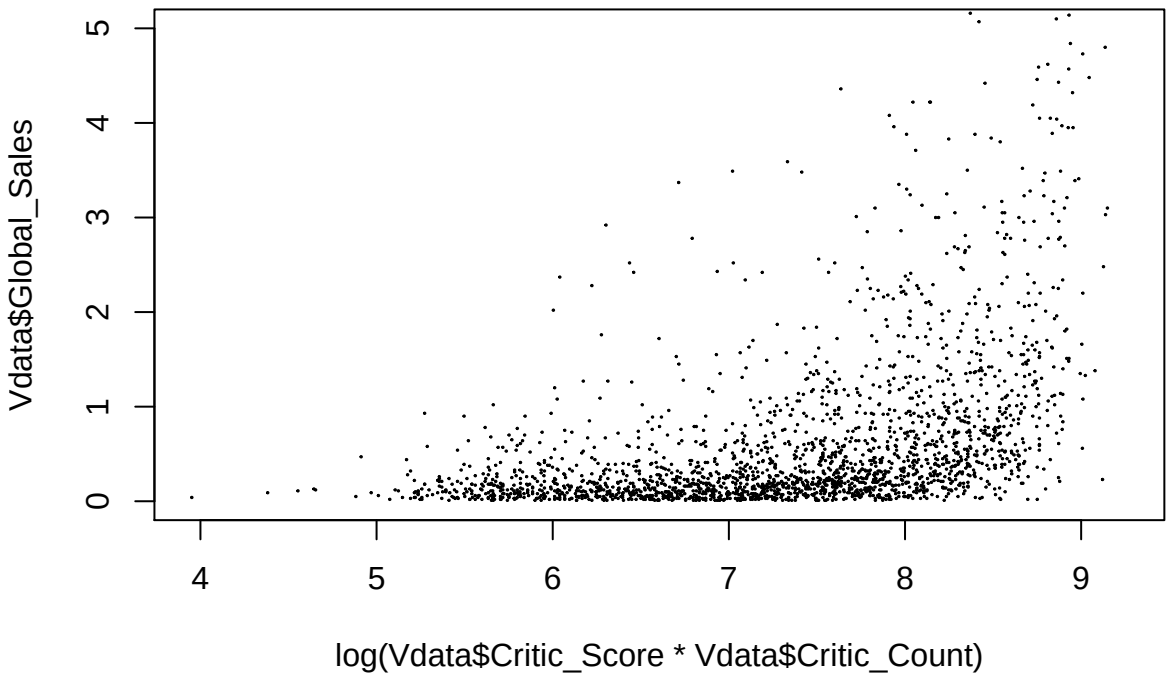
```
plot((Vdata$Critic_Score*Vdata$Critic_Count), ylim=c(0, 5),Vdata$Global_Sales, cex=0.1)
```




```
plot(sqrt(Vdata$Critic_Score*Vdata$Critic_Count), ylim=c(0, 5),Vdata$Global_Sales, cex=0.1)
```




```
plot(log(Vdata$Critic_Score*Vdata$Critic_Count), ylim=c(0, 5),Vdata$Global_Sales, cex=0.1)
```


Vdata$Global_Sales
0
1
2
3
4
5
4
5
6
7
8
9
log(Vdata$Critic_Score * Vdata$Critic_Count)