

bla

Gulzina Kuttubekova

12/4/2019

```
games <- read.csv('games.csv')
sales <- read.csv('sales.csv')

# log-transform and remove zeros
sales_nonzero <- sales %>% select_all() %>%
  filter(Sales != 0)

set.seed(13)
sets <- stratified(sales_nonzero, c('Platform', 'Genre', 'Publisher',
  'Developer', 'Rating', 'Decade',
  'Platform_Company', 'Platform_Gen',
  'Family_Platform', 'Main_Developer',
  'Developer_Country', 'Main_Publisher',
  'Region'), 0.7, bothSets = TRUE)

train_sales <- sets$SAMP1
test_sales <- sets$SAMP2
```

MLR without interactions (for non-zero values)

```
mlr_without_nonzero <- lm(log(Sales) ~ Genre + Critic_Score + Critic_Count +
  Rating +
  Decade + Platform_Company + Platform_Gen + Family_Platform +
  Main_Developer + Main_Publisher + Region,
  data = train_sales)
mlr_without_nonzero %>% summary()

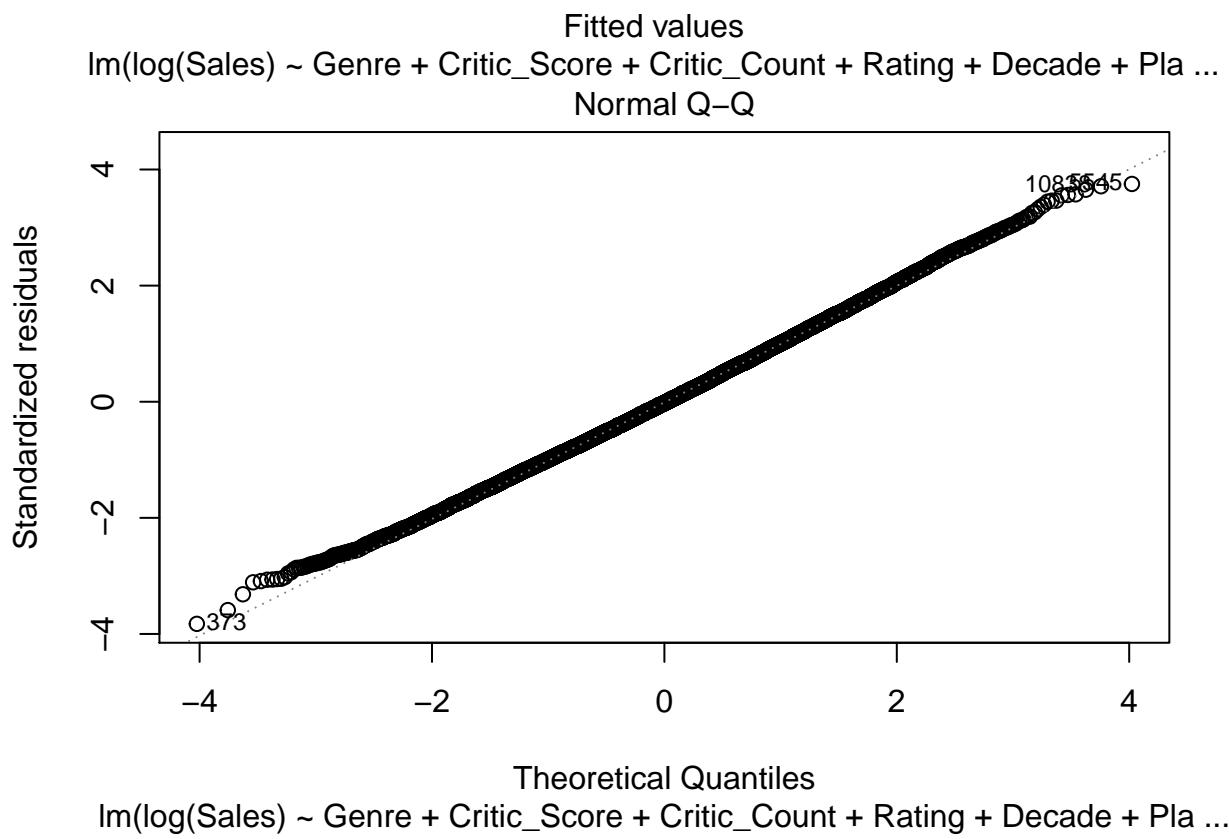
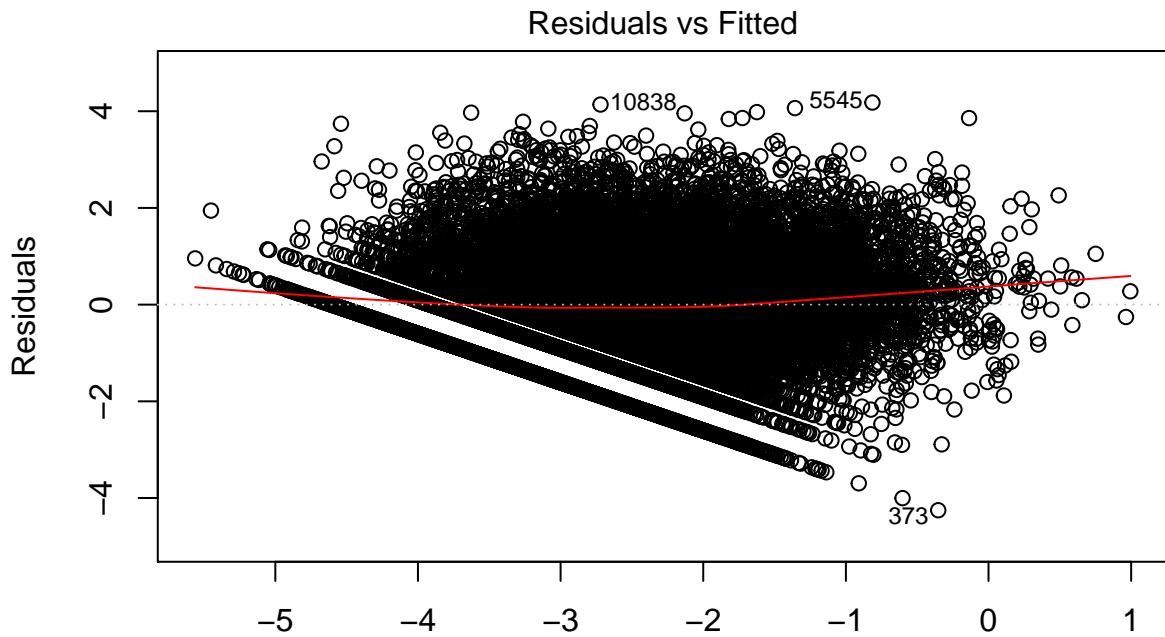
##
## Call:
## lm(formula = log(Sales) ~ Genre + Critic_Score + Critic_Count +
##   Rating + Decade + Platform_Company + Platform_Gen + Family_Platform +
##   Main_Developer + Main_Publisher + Region, data = train_sales)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -4.2528 -0.7674 -0.0225  0.7410  4.1798 
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -3.1253783  0.6556531 -4.767 1.89e-06 ***
## GenreAdventure -0.3528435  0.0478815 -7.369 1.79e-13 ***
## GenreFighting  0.0855878  0.0412036  2.077 0.037799 *  
## GenreMisc     0.1502997  0.0411569  3.652 0.000261 *** 
## GenrePlatform -0.0769589  0.0410885 -1.873 0.061085 .  
## GenrePuzzle   -0.5299455  0.0715502 -7.407 1.35e-13 ***
## GenreRacing   -0.0295724  0.0366395 -0.807 0.419610
```

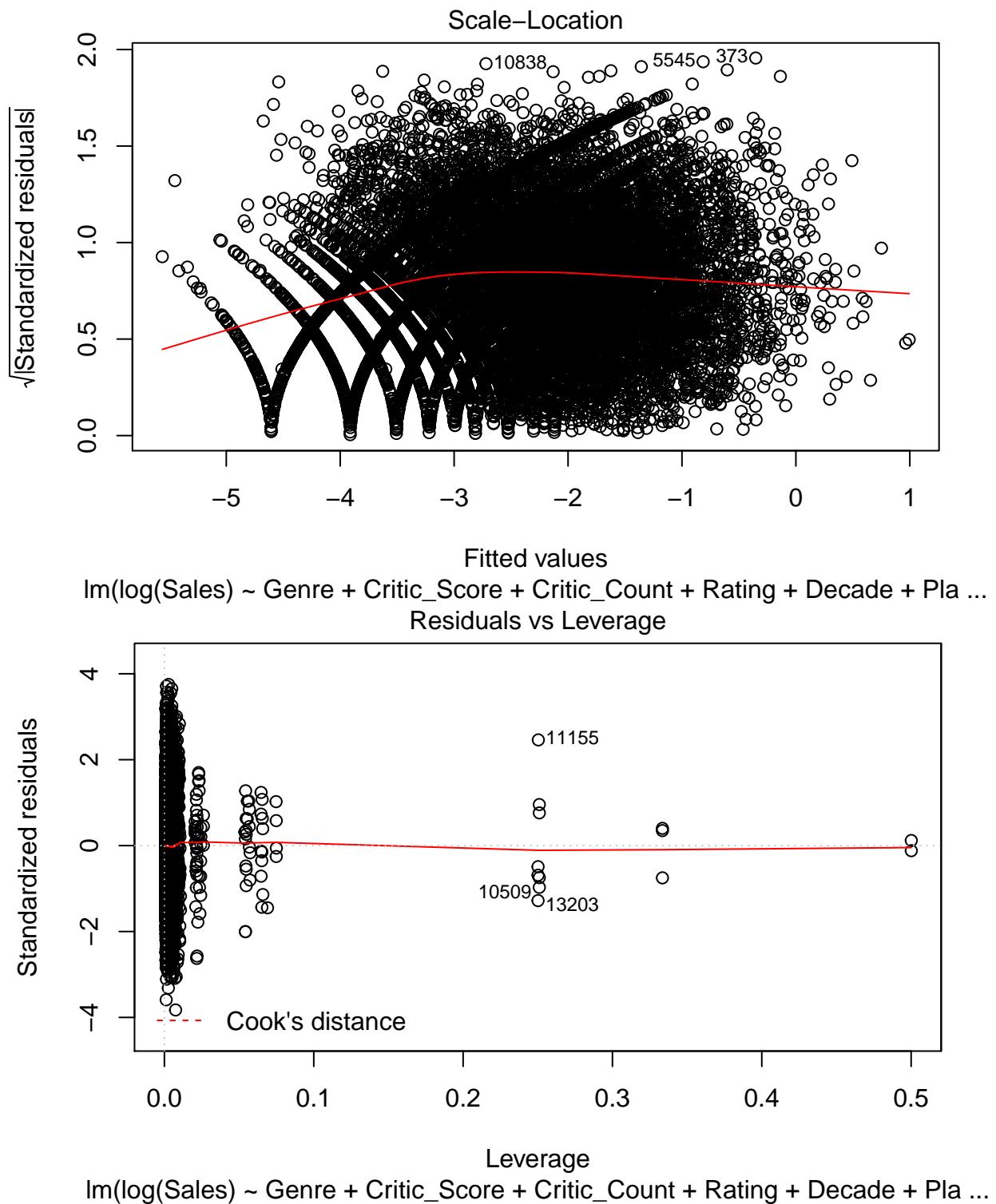
```

## GenreRole-Playing      -0.1811241  0.0326016  -5.556 2.81e-08 ***
## GenreShooter           -0.0753660  0.0299088  -2.520 0.011749 *
## GenreSimulation         0.1071648  0.0467540   2.292 0.021912 *
## GenreSports             -0.1532166  0.0367508  -4.169 3.07e-05 ***
## GenreStrategy           -0.5185585  0.0517040  -10.029 < 2e-16 ***
## Critic_Score            0.0204265  0.0007254   28.160 < 2e-16 ***
## Critic_Count            0.0173709  0.0005785   30.027 < 2e-16 ***
## RatingE                 -0.7418373  0.6465508  -1.147 0.251242
## RatingE10+              -0.8870206  0.6465843  -1.372 0.170126
## RatingK-A               -0.8689556  0.8584113  -1.012 0.311417
## RatingM                 -0.8562276  0.6460117  -1.325 0.185054
## RatingRP                0.0312343  1.0588554   0.029 0.976468
## RatingT                 -1.0194164  0.6461085  -1.578 0.114635
## Decade10s                0.2234664  0.0276938   8.069 7.53e-16 ***
## Decade80s                -0.8688264  0.5612188  -1.548 0.121615
## Decade90s                0.4603290  0.0734872   6.264 3.84e-10 ***
## Platform_CompanyNintendo 0.7029543  0.0935423   7.515 5.97e-14 ***
## Platform_CompanySega     1.6380007  0.2898484   5.651 1.62e-08 ***
## Platform_CompanySONY     0.5331792  0.0237971  22.405 < 2e-16 ***
## Platform_Gen6th_Gen      -0.1854338  0.0789325  -2.349 0.018821 *
## Platform_Gen7th_Gen      -0.1292237  0.0813791  -1.588 0.112322
## Platform_Gen8th_Gen      -0.5406296  0.0884470  -6.112 1.00e-09 ***
## Family_PlatformMisc      -0.6943693  0.0882704  -7.866 3.86e-15 ***
## Family_PlatformNintendo_DS -0.4544641  0.0982449  -4.626 3.76e-06 ***
## Family_PlatformNintendo_Wii -0.2378277  0.0980886  -2.425 0.015335 *
## Family_PlatformSONY_PS    NA          NA          NA          NA
## Main_DeveloperEA          -0.0367772  0.0815391  -0.451 0.651968
## Main_DeveloperMicrosoft   0.0506879  0.0977046   0.519 0.603916
## Main_DeveloperNamco       -0.1178182  0.0916522  -1.285 0.198638
## Main_DeveloperNintendo    0.2167704  0.0851471   2.546 0.010910 *
## Main_DeveloperOther        -0.1902076  0.0646546  -2.942 0.003266 **
## Main_DeveloperSega        -0.1673716  0.0874371  -1.914 0.055612 .
## Main_DeveloperSONY        -0.0066040  0.0984329  -0.067 0.946510
## Main_DeveloperSquare_Enix 0.2426033  0.1041767   2.329 0.019883 *
## Main_DeveloperTencent     -0.2428805  0.1697454  -1.431 0.152491
## Main_DeveloperUbisoft      0.2380149  0.0842517   2.825 0.004733 **
## Main_PublisherEA          -0.0784781  0.0564608  -1.390 0.164560
## Main_PublisherMicrosoft   -0.2668457  0.0796686  -3.349 0.000811 ***
## Main_PublisherNamco       -0.3780556  0.0532134  -7.105 1.25e-12 ***
## Main_PublisherNintendo    0.2155658  0.0647705   3.328 0.000876 ***
## Main_PublisherOther        -0.4553595  0.0399267  -11.405 < 2e-16 ***
## Main_PublisherSega        -0.2185968  0.0513189  -4.260 2.06e-05 ***
## Main_PublisherSONY        -0.4573247  0.0769590  -5.942 2.86e-09 ***
## Main_PublisherSquare_Enix -0.3457532  0.0729008  -4.743 2.12e-06 ***
## Main_PublisherTencent     -0.8216940  0.2868812  -2.864 0.004185 **
## Main_PublisherUbisoft      -0.4783102  0.0565504  -8.458 < 2e-16 ***
## RegionJP_Sales            -0.6409473  0.0317184  -20.207 < 2e-16 ***
## RegionNA_Sales             0.6775433  0.0216953   31.230 < 2e-16 ***
## RegionOther_Sales          -0.9113592  0.0223364  -40.801 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.116 on 17393 degrees of freedom
## Multiple R-squared:  0.4032, Adjusted R-squared:  0.4014

```

```
## F-statistic: 217.6 on 54 and 17393 DF, p-value: < 2.2e-16  
plot(mlr_without_nonzero)
```





Check the real normality. On the 0.05 significance level, normality condition is still violated. But at least the Residual and QQ plots look nicer.

```
shapiro.test(mlr_without_nonzero$residuals
             [sample(1:length(train_sales$Platform), 5000)])  
  
##  
## Shapiro-Wilk normality test  
##  
## data: mlr_without_nonzero$residuals[sample(1:length(train_sales$Platform), 5000)]  
## W = 0.99883, p-value = 0.001195  
  
However, heteroskedasticity is still present.  
  
bpptest(mlr_without_nonzero)  
  
##  
## studentized Breusch-Pagan test  
##  
## data: mlr_without_nonzero  
## BP = 1042.2, df = 54, p-value < 2.2e-16
```

get the AIC and BIC

```
AIC(mlr_without_nonzero)  
  
## [1] 53396.81  
BIC(mlr_without_nonzero)  
  
## [1] 53831.76  
  
# rmse on train data  
sqrt(sum((mlr_without_nonzero$fitted.values - train_sales$Sales)^2) /  
     length(mlr_without_nonzero$fitted.values))  
  
## [1] 2.900504  
  
predicted_sales <- predict(mlr_without_nonzero,  
                           test_sales[,-c(1,2,4,7,8,9,11,17,20)])  
  
## Warning in predict.lm(mlr_without_nonzero, test_sales[, -c(1, 2, 4, 7, 8, :  
## prediction from a rank-deficient fit may be misleading  
  
# rmse on test data  
sqrt(sum((predicted_sales - test_sales$Sales)^2) / length(predicted_sales))  
  
## [1] 2.8137
```

MLR with interactions (for non-zero values)

```

mlr_with_nonzero <- lm(log(Sales) ~ Genre + Critic_Score + Critic_Count + Rating +
Decade + Platform_Company + Platform_Gen + Family_Platform +
Main_Developer + Main_Publisher + Region +
Genre:Region + Main_Developer:Region +
Main_Publisher:Region + Genre:Decade,
data = train_sales)
mlr_with_nonzero %>% summary()

## 
## Call:
## lm(formula = log(Sales) ~ Genre + Critic_Score + Critic_Count +
##     Rating + Decade + Platform_Company + Platform_Gen + Family_Platform +
##     Main_Developer + Main_Publisher + Region + Genre:Region +
##     Main_Developer:Region + Main_Publisher:Region + Genre:Decade,
##     data = train_sales)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.4429 -0.7347 -0.0212  0.7054  4.5416
## 
## 
## Coefficients: (13 not defined because of singularities)
##                                     Estimate Std. Error t value
## (Intercept)                   -3.2033939  0.6382784 -5.019
## GenreAdventure                -0.2187686  0.0883509 -2.476
## GenreFighting                  0.0498200  0.0777668  0.641
## GenreMisc                      0.1607246  0.0758056  2.120
## GenrePlatform                  0.0790213  0.0730732  1.081
## GenrePuzzle                    -0.6000973  0.1421340 -4.222
## GenreRacing                     0.1605363  0.0631554  2.542
## GenreRole-Playing               -0.3540385  0.0636403 -5.563
## GenreShooter                    -0.0524898  0.0547675 -0.958
## GenreSimulation                 0.0976189  0.0823046  1.186
## GenreSports                     -0.2574540  0.0620030 -4.152
## GenreStrategy                   -0.3892434  0.0867004 -4.490
## Critic_Score                     0.0211208  0.0007059 29.921
## Critic_Count                     0.0181544  0.0005627 32.263
## RatingE                         -0.7261993  0.6241235 -1.164
## RatingE10+                      -0.8617537  0.6241544 -1.381
## RatingK-A                       -1.4409883  0.8887025 -1.621
## RatingM                         -0.8409591  0.6236207 -1.349
## RatingRP                        -0.3890384  1.0444686 -0.372
## RatingT                          -0.9988588  0.6236920 -1.602
## Decade10s                        0.2132118  0.0395893  5.386
## Decade80s                        -0.8012359  0.5432330 -1.475
## Decade90s                        1.0744635  0.1362895  7.884
## Platform_CompanyNintendo        0.6723659  0.0919073  7.316
## Platform_CompanySega            1.7602437  0.2903136  6.063
## Platform_CompanySONY            0.5494334  0.0230476 23.839
## Platform_Gen6th Gen              -0.1087839  0.0781055 -1.393
## Platform_Gen7th Gen              -0.0423111  0.0804456 -0.526
## Platform_Gen8th Gen              -0.4280647  0.0872819 -4.904

```

## Family_PlatformMisc	-0.6707379	0.0869522	-7.714
## Family_PlatformNintendo_DS	-0.4300212	0.0964561	-4.458
## Family_PlatformNintendo_Wii	-0.2030624	0.0963288	-2.108
## Family_PlatformSONY_PS	NA	NA	NA
## Main_DeveloperEA	-0.2823871	0.1393368	-2.027
## Main_DeveloperMicrosoft	0.1386939	0.1683739	0.824
## Main_DeveloperNamco	-0.3469747	0.1621589	-2.140
## Main_DeveloperNintendo	-0.1058644	0.1531911	-0.691
## Main_DeveloperOther	-0.3731206	0.1104984	-3.377
## Main_DeveloperSega	-0.4606128	0.1539473	-2.992
## Main_DeveloperSONY	-0.1527853	0.1732731	-0.882
## Main_DeveloperSquare_Enix	-0.1728579	0.1921266	-0.900
## Main_DeveloperTencent	-0.3244892	0.2808916	-1.155
## Main_DeveloperUbisoft	0.2743993	0.1431073	1.917
## Main_PublisherEA	0.1076641	0.0967409	1.113
## Main_PublisherMicrosoft	-0.2441393	0.1332839	-1.832
## Main_PublisherNamco	-0.3772831	0.0928872	-4.062
## Main_PublisherNintendo	0.1591664	0.1152073	1.382
## Main_PublisherOther	-0.3712722	0.0689580	-5.384
## Main_PublisherSega	-0.1112654	0.0887337	-1.254
## Main_PublisherSONY	-0.3516413	0.1346494	-2.612
## Main_PublisherSquare_Enix	-0.2750108	0.1324498	-2.076
## Main_PublisherTencent	-0.4474868	0.3848260	-1.163
## Main_PublisherUbisoft	-0.4808896	0.0961003	-5.004
## RegionJP_Sales	-1.8247138	0.2593612	-7.035
## RegionNA_Sales	0.7139688	0.1371959	5.204
## RegionOther_Sales	-1.0733336	0.1414518	-7.588
## GenreAdventure:RegionJP_Sales	0.4283220	0.1742467	2.458
## GenreFighting:RegionJP_Sales	0.1218748	0.1306968	0.933
## GenreMisc:RegionJP_Sales	0.0609321	0.1548583	0.393
## GenrePlatform:RegionJP_Sales	0.0435059	0.1407311	0.309
## GenrePuzzle:RegionJP_Sales	0.2461829	0.2324997	1.059
## GenreRacing:RegionJP_Sales	-0.4438643	0.1322761	-3.356
## GenreRole-Playing:RegionJP_Sales	0.8659633	0.1000989	8.651
## GenreShooter:RegionJP_Sales	-0.6065397	0.1053520	-5.757
## GenreSimulation:RegionJP_Sales	0.3832064	0.1563291	2.451
## GenreSports:RegionJP_Sales	0.1653619	0.1306288	1.266
## GenreStrategy:RegionJP_Sales	-0.0390293	0.2097064	-0.186
## GenreAdventure:RegionNA_Sales	-0.1037754	0.1157335	-0.897
## GenreFighting:RegionNA_Sales	0.1892732	0.1000777	1.891
## GenreMisc:RegionNA_Sales	0.1346233	0.0974204	1.382
## GenrePlatform:RegionNA_Sales	0.0179714	0.0941844	0.191
## GenrePuzzle:RegionNA_Sales	0.1290012	0.1776509	0.726
## GenreRacing:RegionNA_Sales	-0.1923644	0.0826364	-2.328
## GenreRole-Playing:RegionNA_Sales	0.0529884	0.0817174	0.648
## GenreShooter:RegionNA_Sales	-0.0783314	0.0709198	-1.105
## GenreSimulation:RegionNA_Sales	-0.0232097	0.1117418	-0.208
## GenreSports:RegionNA_Sales	0.1151045	0.0787622	1.461
## GenreStrategy:RegionNA_Sales	-0.3076052	0.1201713	-2.560
## GenreAdventure:RegionOther_Sales	-0.0260605	0.1206647	-0.216
## GenreFighting:RegionOther_Sales	0.0159046	0.1036843	0.153
## GenreMisc:RegionOther_Sales	0.0712270	0.1002960	0.710
## GenrePlatform:RegionOther_Sales	-0.0992523	0.0983602	-1.009
## GenrePuzzle:RegionOther_Sales	0.0527178	0.1855626	0.284

## GenreRacing:RegionOther_Sales	-0.0811027	0.0849311	-0.955
## GenreRole-Playing:RegionOther_Sales	0.0230373	0.0841201	0.274
## GenreShooter:RegionOther_Sales	-0.0594192	0.0733048	-0.811
## GenreSimulation:RegionOther_Sales	-0.0211938	0.1129485	-0.188
## GenreSports:RegionOther_Sales	0.0306987	0.0806006	0.381
## GenreStrategy:RegionOther_Sales	-0.1143540	0.1218524	-0.938
## Main_DeveloperEA:RegionJP_Sales	-0.2836340	0.3781707	-0.750
## Main_DeveloperMicrosoft:RegionJP_Sales	-0.8577414	0.3916702	-2.190
## Main_DeveloperNamco:RegionJP_Sales	0.4975186	0.3729659	1.334
## Main_DeveloperNintendo:RegionJP_Sales	0.3689160	0.3567741	1.034
## Main_DeveloperOther:RegionJP_Sales	0.1861431	0.3170168	0.587
## Main_DeveloperSega:RegionJP_Sales	0.0787709	0.3819926	0.206
## Main_DeveloperSONY:RegionJP_Sales	-0.2256702	0.3941305	-0.573
## Main_DeveloperSquare Enix:RegionJP_Sales	0.8881121	0.3927475	2.261
## Main_DeveloperTencent:RegionJP_Sales	-0.9175135	0.6421188	-1.429
## Main_DeveloperUbisoft:RegionJP_Sales	-1.2984706	0.3948482	-3.289
## Main_DeveloperEA:RegionNA_Sales	0.3093916	0.1947082	1.589
## Main_DeveloperMicrosoft:RegionNA_Sales	0.0084670	0.2383787	0.036
## Main_DeveloperNamco:RegionNA_Sales	0.0954335	0.2248432	0.424
## Main_DeveloperNintendo:RegionNA_Sales	0.4902555	0.2114115	2.319
## Main_DeveloperOther:RegionNA_Sales	0.1833709	0.1531792	1.197
## Main_DeveloperSega:RegionNA_Sales	0.6365962	0.2119279	3.004
## Main_DeveloperSONY:RegionNA_Sales	0.1816978	0.2413145	0.753
## Main_DeveloperSquare Enix:RegionNA_Sales	0.2779680	0.2624075	1.059
## Main_DeveloperTencent:RegionNA_Sales	0.1863988	0.4175186	0.446
## Main_DeveloperUbisoft:RegionNA_Sales	-0.1090300	0.2002878	-0.544
## Main_DeveloperEA:RegionOther_Sales	0.4086295	0.2004667	2.038
## Main_DeveloperMicrosoft:RegionOther_Sales	0.0586101	0.2441000	0.240
## Main_DeveloperNamco:RegionOther_Sales	0.2570633	0.2332151	1.102
## Main_DeveloperNintendo:RegionOther_Sales	0.2923450	0.2164198	1.351
## Main_DeveloperOther:RegionOther_Sales	0.2559045	0.1585139	1.614
## Main_DeveloperSega:RegionOther_Sales	0.2861282	0.2178524	1.313
## Main_DeveloperSONY:RegionOther_Sales	0.2890826	0.2447815	1.181
## Main_DeveloperSquare Enix:RegionOther_Sales	0.3452694	0.2719715	1.270
## Main_DeveloperTencent:RegionOther_Sales	0.1809602	0.4120875	0.439
## Main_DeveloperUbisoft:RegionOther_Sales	0.1822348	0.2063693	0.883
## Main_PublisherEA:RegionJP_Sales	0.0351116	0.2560439	0.137
## Main_PublisherMicrosoft:RegionJP_Sales	0.6483350	0.2956947	2.193
## Main_PublisherNamco:RegionJP_Sales	1.5118067	0.2301904	6.568
## Main_PublisherNintendo:RegionJP_Sales	1.7030215	0.2400325	7.095
## Main_PublisherOther:RegionJP_Sales	0.9767261	0.2004146	4.874
## Main_PublisherSega:RegionJP_Sales	0.5442566	0.2342463	2.323
## Main_PublisherSONY:RegionJP_Sales	1.0747727	0.2881252	3.730
## Main_PublisherSquare Enix:RegionJP_Sales	1.0957757	0.2638433	4.153
## Main_PublisherTencent:RegionJP_Sales	NA	NA	NA
## Main_PublisherUbisoft:RegionJP_Sales	1.2655134	0.2666489	4.746
## Main_PublisherEA:RegionNA_Sales	-0.2025348	0.1351079	-1.499
## Main_PublisherMicrosoft:RegionNA_Sales	0.0976879	0.1906696	0.512
## Main_PublisherNamco:RegionNA_Sales	-0.2241895	0.1283593	-1.747
## Main_PublisherNintendo:RegionNA_Sales	-0.1375254	0.1569990	-0.876
## Main_PublisherOther:RegionNA_Sales	-0.2898164	0.0945790	-3.064
## Main_PublisherSega:RegionNA_Sales	-0.3859133	0.1225258	-3.150
## Main_PublisherSONY:RegionNA_Sales	-0.4195550	0.1876590	-2.236
## Main_PublisherSquare Enix:RegionNA_Sales	-0.3181816	0.1854930	-1.715

## Main_PublisherTencent:RegionNA_Sales	-1.0156233	0.6714797	-1.513
## Main_PublisherUbisoft:RegionNA_Sales	-0.0433505	0.1340721	-0.323
## Main_PublisherEA:RegionOther_Sales	-0.2221615	0.1386560	-1.602
## Main_PublisherMicrosoft:RegionOther_Sales	-0.1247873	0.1970267	-0.633
## Main_PublisherNamco:RegionOther_Sales	-0.0544976	0.1339396	-0.407
## Main_PublisherNintendo:RegionOther_Sales	-0.3832147	0.1599319	-2.396
## Main_PublisherOther:RegionOther_Sales	-0.0477360	0.0976609	-0.489
## Main_PublisherSega:RegionOther_Sales	-0.0270185	0.1260385	-0.214
## Main_PublisherSONY:RegionOther_Sales	-0.0683987	0.1893203	-0.361
## Main_PublisherSquare Enix:RegionOther_Sales	-0.1263471	0.1874003	-0.674
## Main_PublisherTencent:RegionOther_Sales	-0.2950966	0.6206715	-0.475
## Main_PublisherUbisoft:RegionOther_Sales	-0.0313557	0.1381995	-0.227
## GenreAdventure:Decade10s	-0.3464049	0.1035500	-3.345
## GenreFighting:Decade10s	-0.2148313	0.0843120	-2.548
## GenreMisc:Decade10s	-0.3357885	0.0901284	-3.726
## GenrePlatform:Decade10s	-0.4792501	0.0873193	-5.488
## GenrePuzzle:Decade10s	0.0697890	0.2054271	0.340
## GenreRacing:Decade10s	-0.2525192	0.0817505	-3.089
## GenreRole-Playing:Decade10s	-0.0013155	0.0654309	-0.020
## GenreShooter:Decade10s	0.3792741	0.0605535	6.263
## GenreSimulation:Decade10s	0.0526545	0.1127240	0.467
## GenreSports:Decade10s	0.2087735	0.0694666	3.005
## GenreStrategy:Decade10s	0.2689191	0.1249998	2.151
## GenreAdventure:Decade80s	NA	NA	NA
## GenreFighting:Decade80s	NA	NA	NA
## GenreMisc:Decade80s	NA	NA	NA
## GenrePlatform:Decade80s	NA	NA	NA
## GenrePuzzle:Decade80s	NA	NA	NA
## GenreRacing:Decade80s	NA	NA	NA
## GenreRole-Playing:Decade80s	NA	NA	NA
## GenreShooter:Decade80s	NA	NA	NA
## GenreSimulation:Decade80s	NA	NA	NA
## GenreSports:Decade80s	NA	NA	NA
## GenreStrategy:Decade80s	NA	NA	NA
## GenreAdventure:Decade90s	-1.0225932	0.2858518	-3.577
## GenreFighting:Decade90s	-0.4274167	0.1945988	-2.196
## GenreMisc:Decade90s	-0.0315904	0.3501130	-0.090
## GenrePlatform:Decade90s	-0.8032622	0.2101077	-3.823
## GenrePuzzle:Decade90s	-1.0858157	0.3694511	-2.939
## GenreRacing:Decade90s	-0.5657196	0.1844896	-3.066
## GenreRole-Playing:Decade90s	-0.4659967	0.1637806	-2.845
## GenreShooter:Decade90s	-1.0720206	0.2243537	-4.778
## GenreSimulation:Decade90s	-1.1609413	0.2439400	-4.759
## GenreSports:Decade90s	-0.7091303	0.2168934	-3.269
## GenreStrategy:Decade90s	-1.0427253	0.2431566	-4.288
##	Pr(> t)		
## (Intercept)	5.25e-07	***	
## GenreAdventure	0.013291	*	
## GenreFighting	0.521770		
## GenreMisc	0.034002	*	
## GenrePlatform	0.279535		
## GenrePuzzle	2.43e-05	***	
## GenreRacing	0.011033	*	
## GenreRole-Playing	2.69e-08	***	

## GenreShooter	0.337869
## GenreSimulation	0.235612
## GenreSports	3.31e-05 ***
## GenreStrategy	7.18e-06 ***
## Critic_Score	< 2e-16 ***
## Critic_Count	< 2e-16 ***
## RatingE	0.244622
## RatingE10+	0.167397
## RatingK-A	0.104939
## RatingM	0.177512
## RatingRP	0.709544
## RatingT	0.109279
## Decade10s	7.31e-08 ***
## Decade80s	0.140247
## Decade90s	3.37e-15 ***
## Platform_CompanyNintendo	2.67e-13 ***
## Platform_CompanySega	1.36e-09 ***
## Platform_CompanySONY	< 2e-16 ***
## Platform_Gen6th Gen	0.163704
## Platform_Gen7th Gen	0.598923
## Platform_Gen8th Gen	9.46e-07 ***
## Family_PlatformMisc	1.29e-14 ***
## Family_PlatformNintendo_DS	8.32e-06 ***
## Family_PlatformNintendo_Wii	0.035044 *
## Family_PlatformSONY_PS	NA
## Main_DeveloperEA	0.042713 *
## Main_DeveloperMicrosoft	0.410107
## Main_DeveloperNamco	0.032391 *
## Main_DeveloperNintendo	0.489537
## Main_DeveloperOther	0.000735 ***
## Main_DeveloperSega	0.002775 **
## Main_DeveloperSONY	0.377919
## Main_DeveloperSquare Enix	0.368288
## Main_DeveloperTencent	0.248020
## Main_DeveloperUbisoft	0.055199 .
## Main_PublisherEA	0.265762
## Main_PublisherMicrosoft	0.067010 .
## Main_PublisherNamco	4.89e-05 ***
## Main_PublisherNintendo	0.167123
## Main_PublisherOther	7.38e-08 ***
## Main_PublisherSega	0.209886
## Main_PublisherSONY	0.009022 **
## Main_PublisherSquare Enix	0.037877 *
## Main_PublisherTencent	0.244915
## Main_PublisherUbisoft	5.67e-07 ***
## RegionJP_Sales	2.06e-12 ***
## RegionNA_Sales	1.97e-07 ***
## RegionOther_Sales	3.41e-14 ***
## GenreAdventure:RegionJP_Sales	0.013976 *
## GenreFighting:RegionJP_Sales	0.351091
## GenreMisc:RegionJP_Sales	0.693977
## GenrePlatform:RegionJP_Sales	0.757217
## GenrePuzzle:RegionJP_Sales	0.289682
## GenreRacing:RegionJP_Sales	0.000794 ***

```

## GenreRole-Playing:RegionJP_Sales < 2e-16 ***
## GenreShooter:RegionJP_Sales 8.69e-09 ***
## GenreSimulation:RegionJP_Sales 0.014245 *
## GenreSports:RegionJP_Sales 0.205569
## GenreStrategy:RegionJP_Sales 0.852358
## GenreAdventure:RegionNA_Sales 0.369904
## GenreFighting:RegionNA_Sales 0.058606 .
## GenreMisc:RegionNA_Sales 0.167026
## GenrePlatform:RegionNA_Sales 0.848676
## GenrePuzzle:RegionNA_Sales 0.467757
## GenreRacing:RegionNA_Sales 0.019932 *
## GenreRole-Playing:RegionNA_Sales 0.516712
## GenreShooter:RegionNA_Sales 0.269389
## GenreSimulation:RegionNA_Sales 0.835459
## GenreSports:RegionNA_Sales 0.143919
## GenreStrategy:RegionNA_Sales 0.010484 *
## GenreAdventure:RegionOther_Sales 0.829010
## GenreFighting:RegionOther_Sales 0.878089
## GenreMisc:RegionOther_Sales 0.477610
## GenrePlatform:RegionOther_Sales 0.312955
## GenrePuzzle:RegionOther_Sales 0.776339
## GenreRacing:RegionOther_Sales 0.339630
## GenreRole-Playing:RegionOther_Sales 0.784194
## GenreShooter:RegionOther_Sales 0.417620
## GenreSimulation:RegionOther_Sales 0.851160
## GenreSports:RegionOther_Sales 0.703301
## GenreStrategy:RegionOther_Sales 0.348019
## Main_DeveloperEA:RegionJP_Sales 0.453255
## Main_DeveloperMicrosoft:RegionJP_Sales 0.028541 *
## Main_DeveloperNamco:RegionJP_Sales 0.182237
## Main_DeveloperNintendo:RegionJP_Sales 0.301135
## Main_DeveloperOther:RegionJP_Sales 0.557097
## Main_DeveloperSega:RegionJP_Sales 0.836629
## Main_DeveloperSONY:RegionJP_Sales 0.566938
## Main_DeveloperSquare Enix:RegionJP_Sales 0.023754 *
## Main_DeveloperTencent:RegionJP_Sales 0.153056
## Main_DeveloperUbisoft:RegionJP_Sales 0.001009 **
## Main_DeveloperEA:RegionNA_Sales 0.112078
## Main_DeveloperMicrosoft:RegionNA_Sales 0.971666
## Main_DeveloperNamco:RegionNA_Sales 0.671247
## Main_DeveloperNintendo:RegionNA_Sales 0.020409 *
## Main_DeveloperOther:RegionNA_Sales 0.231284
## Main_DeveloperSega:RegionNA_Sales 0.002670 **
## Main_DeveloperSONY:RegionNA_Sales 0.451490
## Main_DeveloperSquare Enix:RegionNA_Sales 0.289479
## Main_DeveloperTencent:RegionNA_Sales 0.655282
## Main_DeveloperUbisoft:RegionNA_Sales 0.586196
## Main_DeveloperEA:RegionOther_Sales 0.041526 *
## Main_DeveloperMicrosoft:RegionOther_Sales 0.810250
## Main_DeveloperNamco:RegionOther_Sales 0.270365
## Main_DeveloperNintendo:RegionOther_Sales 0.176770
## Main_DeveloperOther:RegionOther_Sales 0.106459
## Main_DeveloperSega:RegionOther_Sales 0.189064
## Main_DeveloperSONY:RegionOther_Sales 0.237626

```

```

## Main_DeveloperSquare Enix:RegionOther_Sales 0.204278
## Main_DeveloperTencent:RegionOther_Sales 0.660572
## Main_DeveloperUbisoft:RegionOther_Sales 0.377220
## Main_PublisherEA:RegionJP_Sales 0.890929
## Main_PublisherMicrosoft:RegionJP_Sales 0.028351 *
## Main_PublisherNamco:RegionJP_Sales 5.26e-11 ***
## Main_PublisherNintendo:RegionJP_Sales 1.34e-12 ***
## Main_PublisherOther:RegionJP_Sales 1.11e-06 ***
## Main_PublisherSega:RegionJP_Sales 0.020167 *
## Main_PublisherSONY:RegionJP_Sales 0.000192 ***
## Main_PublisherSquare Enix:RegionJP_Sales 3.30e-05 ***
## Main_PublisherTencent:RegionJP_Sales NA
## Main_PublisherUbisoft:RegionJP_Sales 2.09e-06 ***
## Main_PublisherEA:RegionNA_Sales 0.133876
## Main_PublisherMicrosoft:RegionNA_Sales 0.608419
## Main_PublisherNamco:RegionNA_Sales 0.080728 .
## Main_PublisherNintendo:RegionNA_Sales 0.381062
## Main_PublisherOther:RegionNA_Sales 0.002185 **
## Main_PublisherSega:RegionNA_Sales 0.001637 **
## Main_PublisherSONY:RegionNA_Sales 0.025382 *
## Main_PublisherSquare Enix:RegionNA_Sales 0.086303 .
## Main_PublisherTencent:RegionNA_Sales 0.130421
## Main_PublisherUbisoft:RegionNA_Sales 0.746444
## Main_PublisherEA:RegionOther_Sales 0.109119
## Main_PublisherMicrosoft:RegionOther_Sales 0.526512
## Main_PublisherNamco:RegionOther_Sales 0.684100
## Main_PublisherNintendo:RegionOther_Sales 0.016581 *
## Main_PublisherOther:RegionOther_Sales 0.624994
## Main_PublisherSega:RegionOther_Sales 0.830264
## Main_PublisherSONY:RegionOther_Sales 0.717891
## Main_PublisherSquare Enix:RegionOther_Sales 0.500187
## Main_PublisherTencent:RegionOther_Sales 0.634474
## Main_PublisherUbisoft:RegionOther_Sales 0.820514
## GenreAdventure:Decade10s 0.000824 ***
## GenreFighting:Decade10s 0.010841 *
## GenreMisc:Decade10s 0.000195 ***
## GenrePlatform:Decade10s 4.11e-08 ***
## GenrePuzzle:Decade10s 0.734067
## GenreRacing:Decade10s 0.002012 **
## GenreRole-Playing:Decade10s 0.983960
## GenreShooter:Decade10s 3.85e-10 ***
## GenreSimulation:Decade10s 0.640427
## GenreSports:Decade10s 0.002656 **
## GenreStrategy:Decade10s 0.031462 *

## GenreAdventure:Decade80s NA
## GenreFighting:Decade80s NA
## GenreMisc:Decade80s NA
## GenrePlatform:Decade80s NA
## GenrePuzzle:Decade80s NA
## GenreRacing:Decade80s NA
## GenreRole-Playing:Decade80s NA
## GenreShooter:Decade80s NA
## GenreSimulation:Decade80s NA
## GenreSports:Decade80s NA

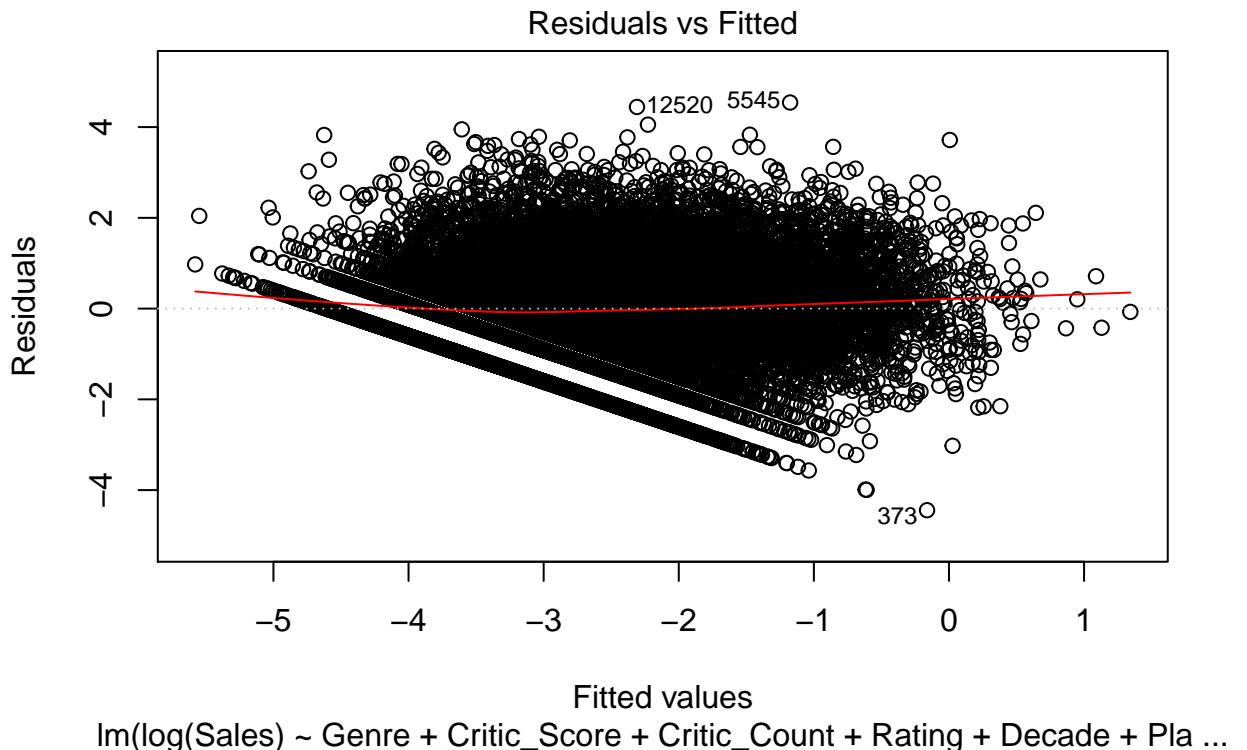
```

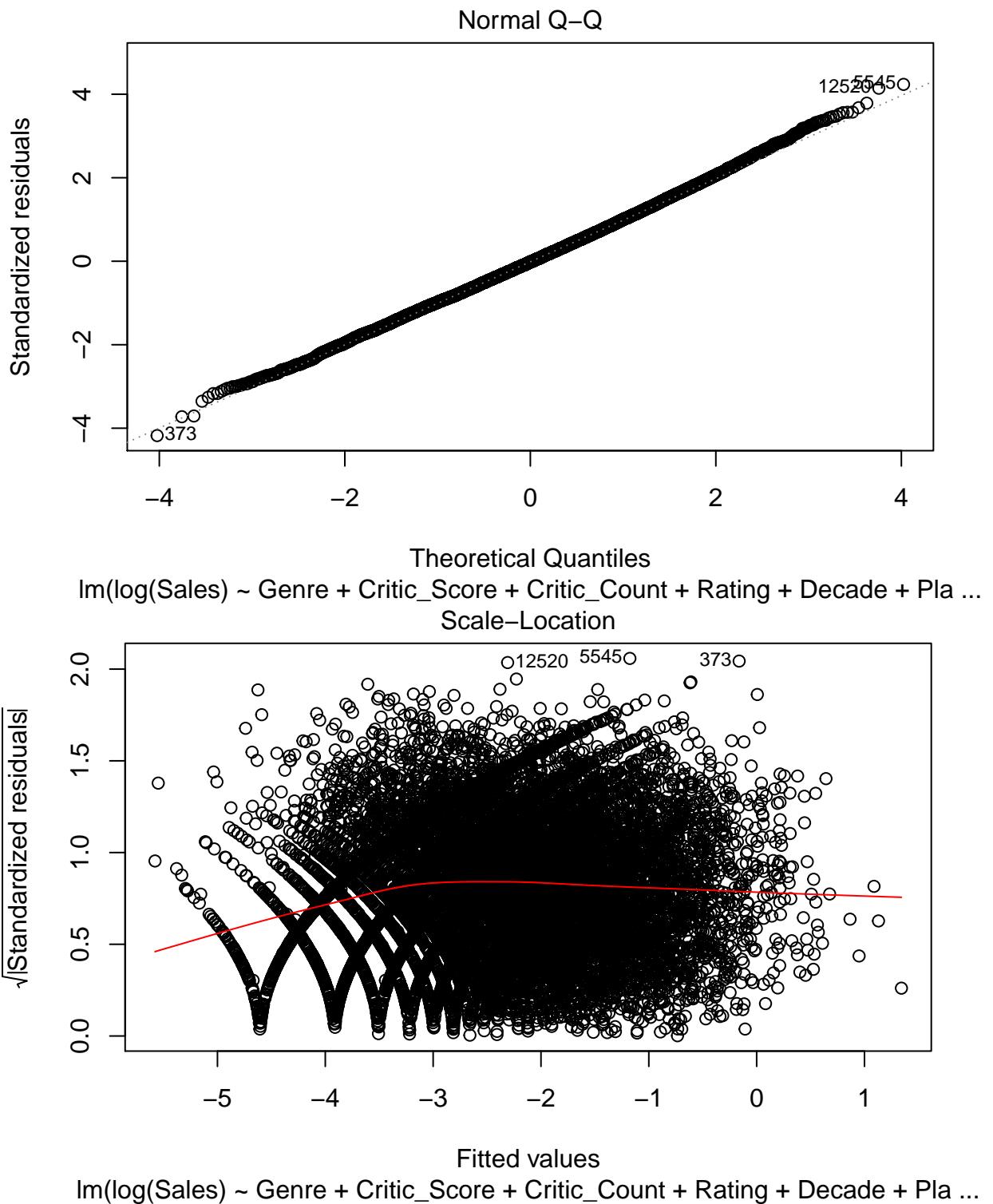
```

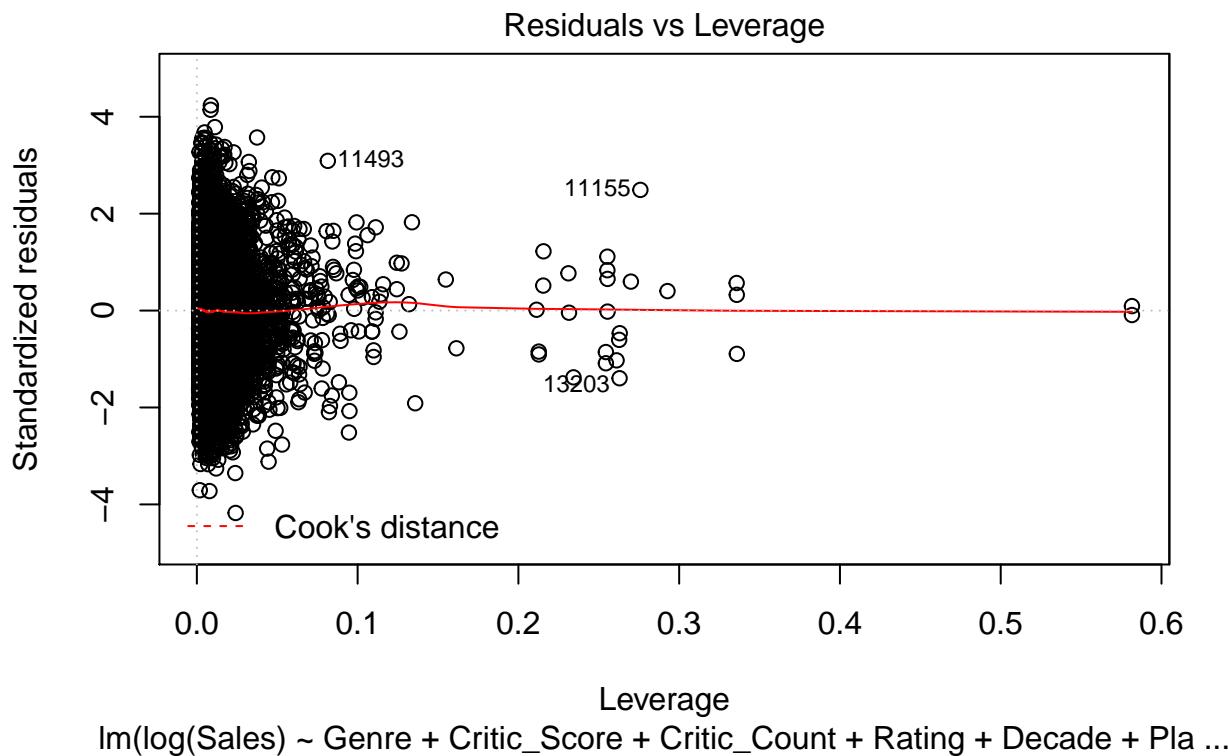
## GenreStrategy:Decade80s NA
## GenreAdventure:Decade90s 0.000348 ***
## GenreFighting:Decade90s 0.028077 *
## GenreMisc:Decade90s 0.928106
## GenrePlatform:Decade90s 0.000132 ***
## GenrePuzzle:Decade90s 0.003297 **
## GenreRacing:Decade90s 0.002170 **
## GenreRole-Playing:Decade90s 0.004443 **
## GenreShooter:Decade90s 1.78e-06 ***
## GenreSimulation:Decade90s 1.96e-06 ***
## GenreSports:Decade90s 0.001080 **
## GenreStrategy:Decade90s 1.81e-05 ***

## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.077 on 17279 degrees of freedom
## Multiple R-squared: 0.4479, Adjusted R-squared: 0.4425
## F-statistic: 83.43 on 168 and 17279 DF, p-value: < 2.2e-16
plot(mlr_with_nonzero)

```







Check the real normality. On the 0.05 significance level, normality condition is still violated. But at least the Residual and QQ plots look nicer.

```
shapiro.test(mlr_with_nonzero$residuals
             [sample(1:length(train_sales$Platform), 5000)])
```

```
##
## Shapiro-Wilk normality test
##
## data: mlr_with_nonzero$residuals[sample(1:length(train_sales$Platform), 5000)]
## W = 0.99909, p-value = 0.008808
```

However, heteroskedasticity is still present.

```
bptest(mlr_with_nonzero)
```

```
##
## studentized Breusch-Pagan test
##
## data: mlr_with_nonzero
## BP = 1061.3, df = 168, p-value < 2.2e-16
```

get the AIC and BIC

```
AIC(mlr_with_nonzero)
## [1] 52267.82
BIC(mlr_with_nonzero)
## [1] 53588.21
# rmse on train data
sqrt(sum((mlr_with_nonzero$fitted.values - train_sales$Sales)^2) /
    length(mlr_with_nonzero$fitted.values))
## [1] 2.908342
predicted_sales <- predict(mlr_with_nonzero,
                           test_sales[,-c(1,2,4,7,8,9,11,17,20)])
## Warning in predict.lm(mlr_with_nonzero, test_sales[, -c(1, 2, 4, 7, 8, 9, :
## prediction from a rank-deficient fit may be misleading
# rmse on test data
sqrt(sum((predicted_sales - test_sales$Sales)^2) / length(predicted_sales))
## [1] 2.806784
```