

Analysis

Gulzina Kuttubekova, Meltem Ozcan

11/30/2019

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

ANOVA (use sales.csv)

In this section, we want to test if there are significant interactions between the categorical variables of our interest: (Genre &) Write the hypotheses for each ANOVA!!!

Write the general model here:

$$y = x'\beta + \epsilon$$

Assumptions:

- We will ignore all other explanatory variables, and test the interactions of our interest only!
- Even if find significant interactions, we should keep in mind that interaction may be significant only
- We'll still include interactions, which were found significant to our MLR prediction model. And compare

Hypotheses 1 (Genre vs Region):

- Write up the null/alternative
- Model fit

```
# since Sales has many zeros, log-transformation leads to -Inf. Hence we add some epsilon = 0.000001 to
sales$Sales[sales$Sales == 0] = 0.000001
```

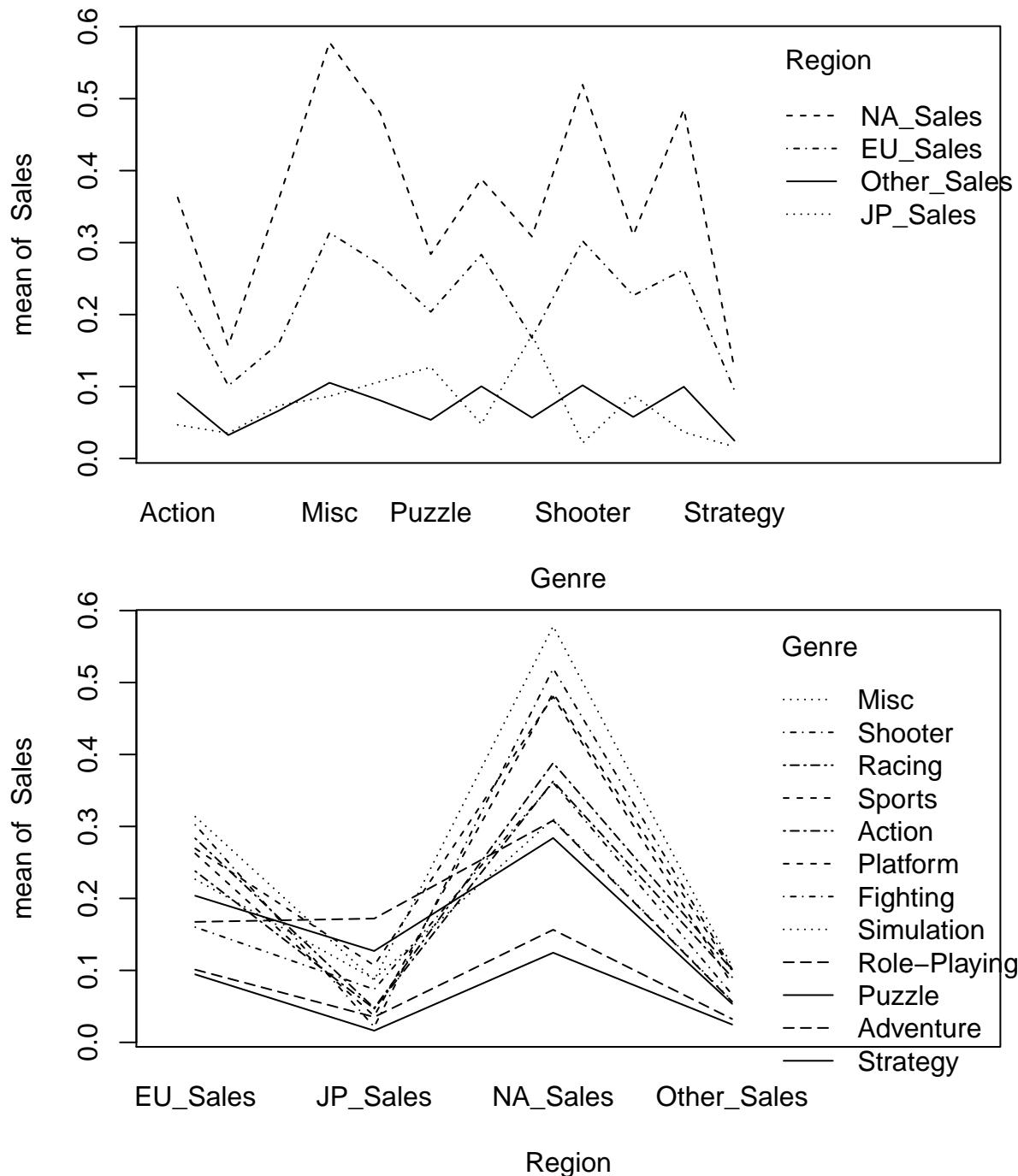
```
genre_region <- lm(log(Sales) ~ Genre*Region, data = sales)
anova(genre_region)
```

```
## Analysis of Variance Table
##
## Response: log(Sales)
##             Df Sum Sq Mean Sq F value    Pr(>F)
## Genre          11  6586     599   34.032 < 2.2e-16 ***
## Region         3 243201   81067 4607.801 < 2.2e-16 ***
## Genre:Region   33 16967      514   29.224 < 2.2e-16 ***
## Residuals     27252 479456       18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: F-value = 24.821 with 33 degrees of freedom and p-value = 2.2e-16. Hence we reject the null hypothesis and conclude that there is an interaction between genre and region (note that when only these two variables are included in the model).

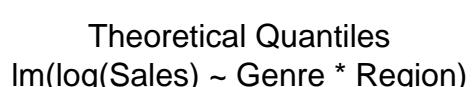
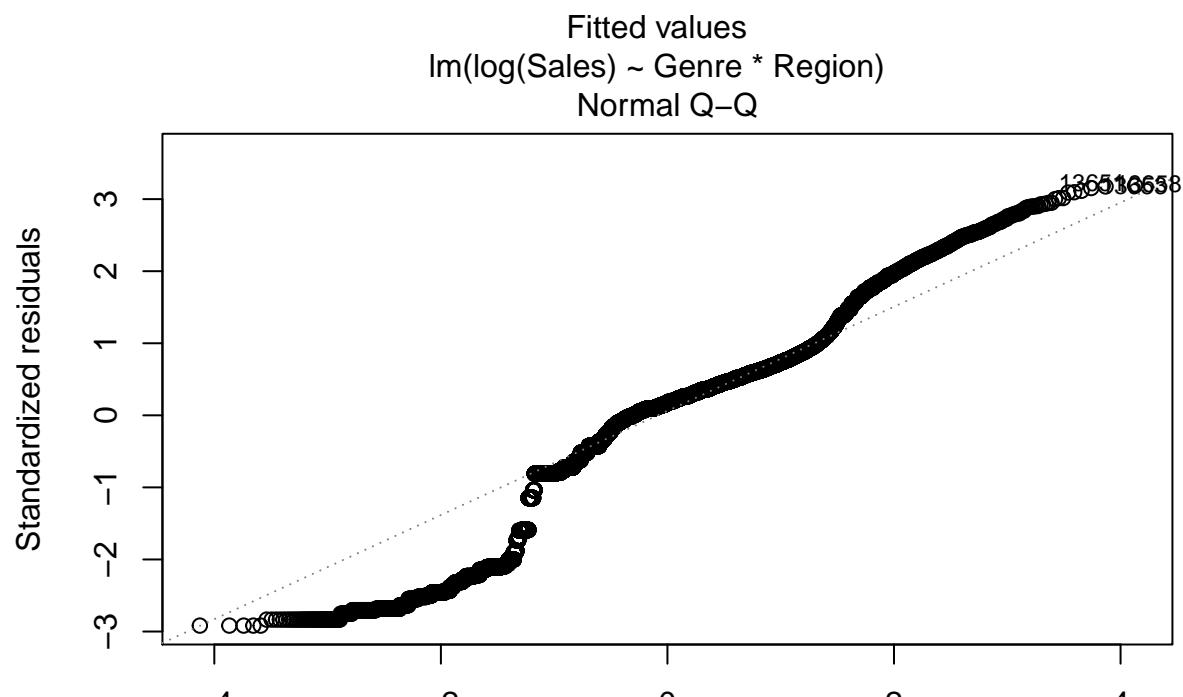
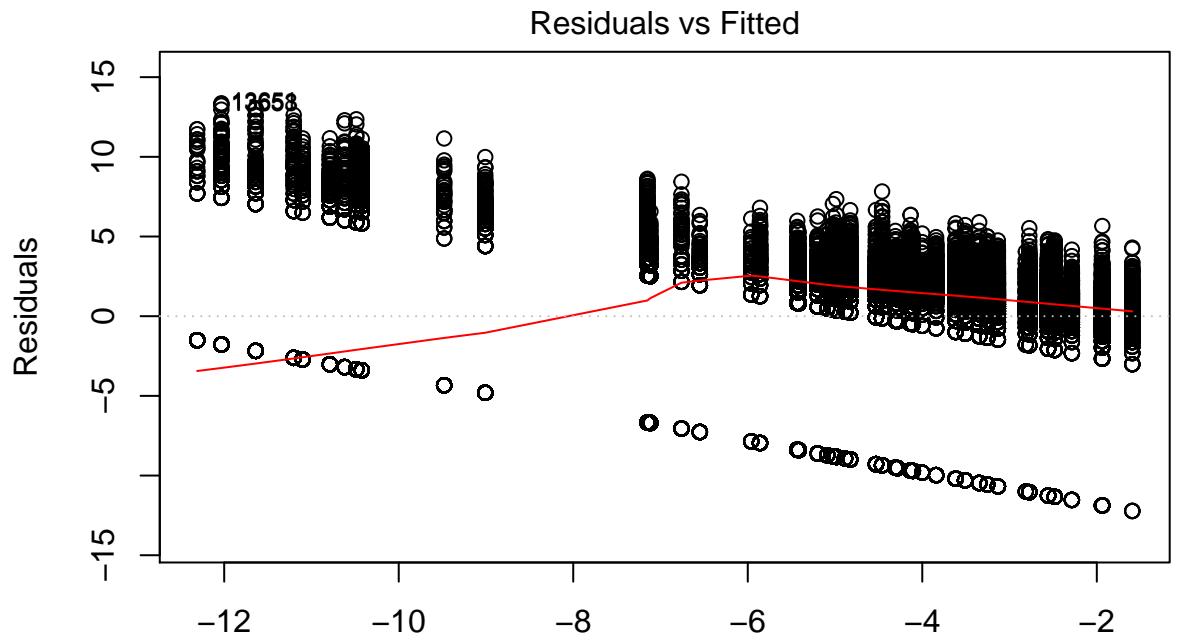
- Interactions plot:

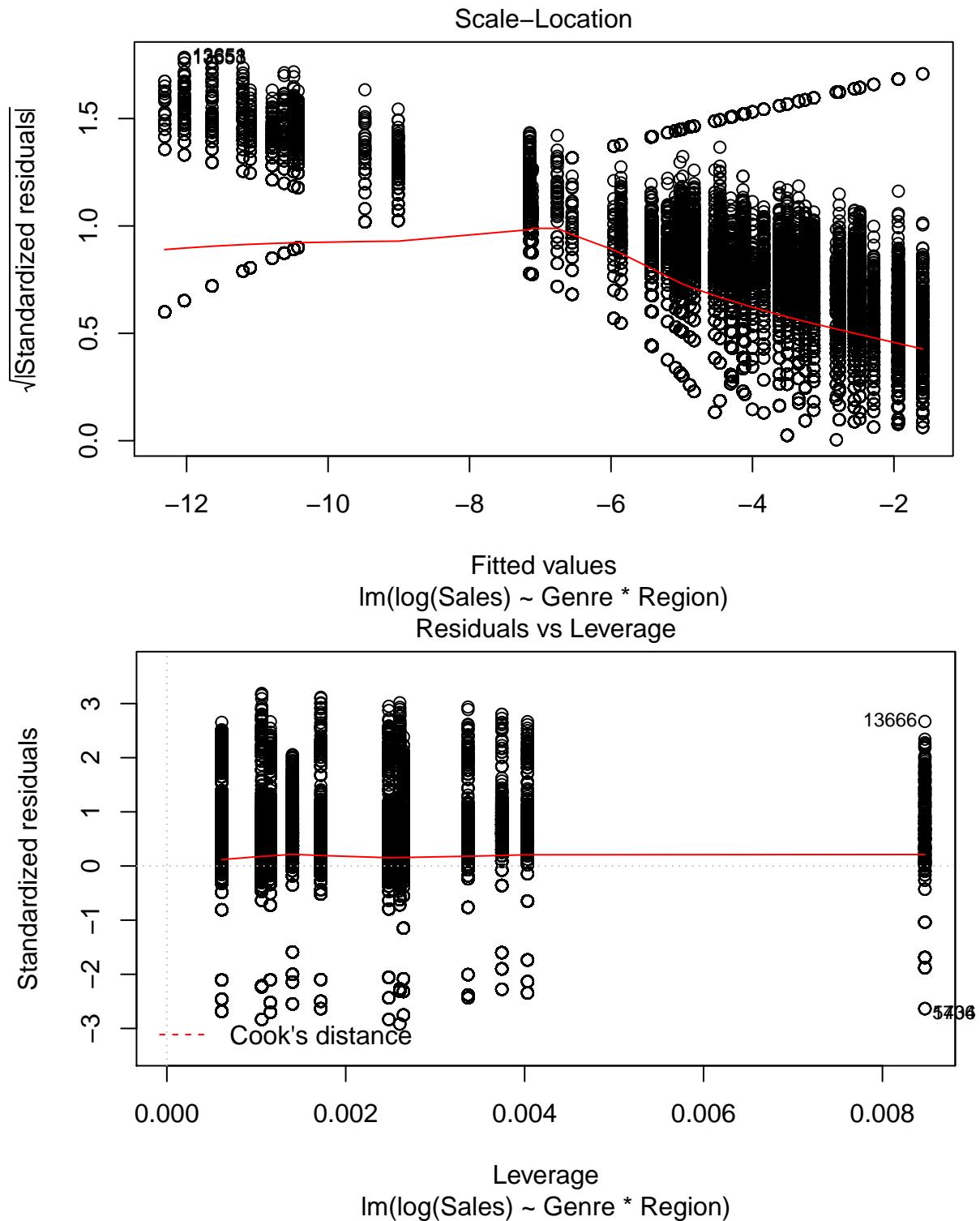
```
with(data=sales, expr={
  interaction.plot(Genre, Region, response=Sales)
  interaction.plot(Region, Genre, response=Sales)})
```



- Model checking:

```
plot(genre_region)
```





Comments: There is some pattern in the distribution of residuals, however the variance is stable, i.e. it's approximately the same for all. Hence, we conclude that the homogeneity assumption is not violated, however the normality assumption might be violated. It also can be noted from the QQ-plot. We'll test it by shapiro-wilk test:

```
shapiro.test(genre_region$residuals[1:5000])
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: genre_region$residuals[1:5000]  
## W = 0.67701, p-value < 2.2e-16
```

P-value < 2.2e-16 < 0.05, so the null hypothesis stating that residuals are normally distributed is rejected.

Hypotheses 2 (Developer company & Region)

- Write up the null/alternative
- Model fit

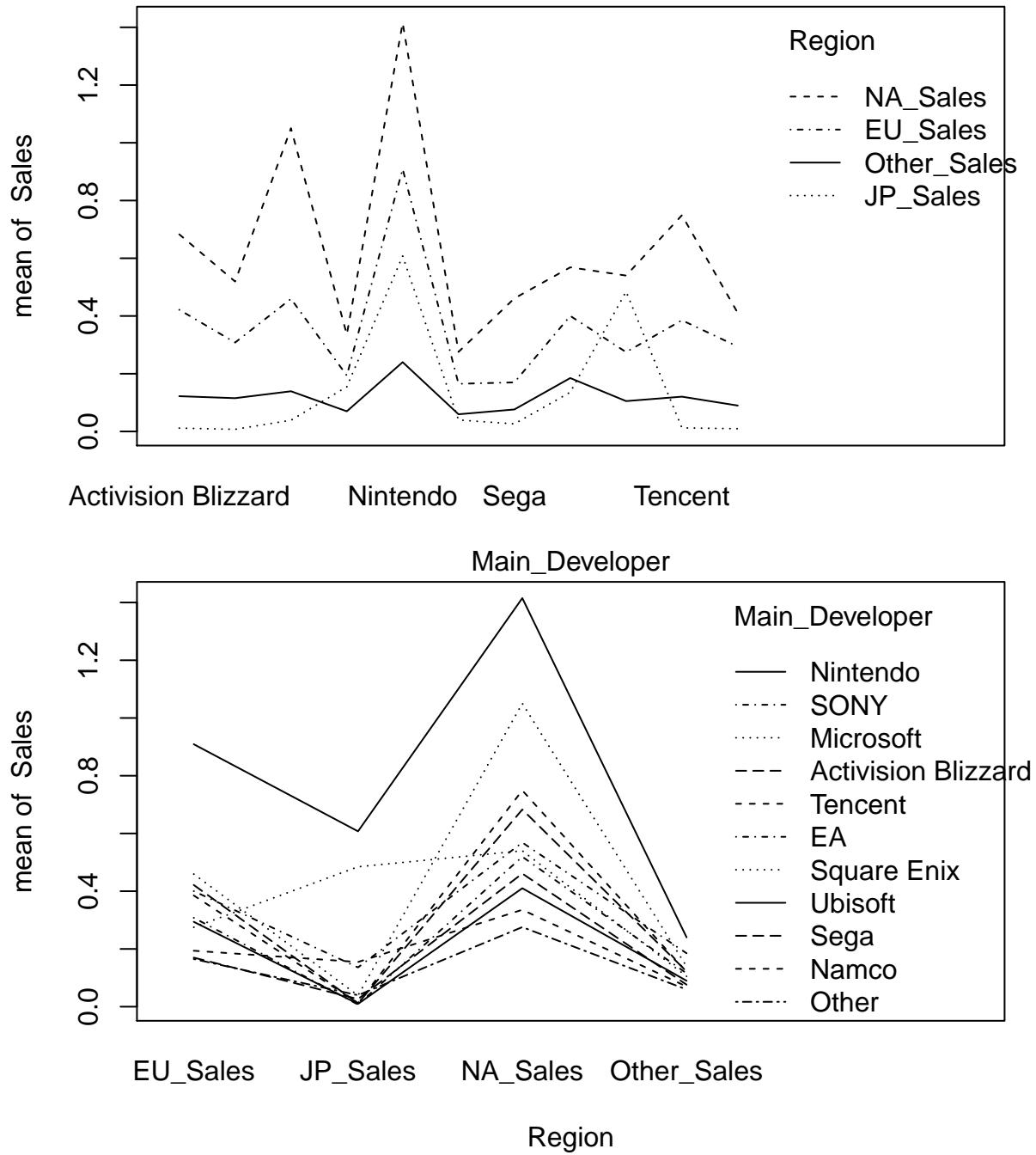
```
dev_region <- lm(log(Sales) ~ Main_Developer*Region, data = sales)  
anova(dev_region)
```

```
## Analysis of Variance Table  
##  
## Response: log(Sales)  
##  
## Df Sum Sq Mean Sq F value Pr(>F)  
## Main_Developer 10 15298 1530 88.180 < 2.2e-16 ***  
## Region 3 243201 81067 4672.672 < 2.2e-16 ***  
## Main_Developer:Region 30 14842 495 28.516 < 2.2e-16 ***  
## Residuals 27256 472869 17  
## ---  
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: F-value = 23.512 with 30 degrees of freedom and p-value < 2.2e-16 < 0.05. Hence we reject the null hypothesis and conclude that there is an interaction between genre and developer company (note that when only these two variables are included in the model).

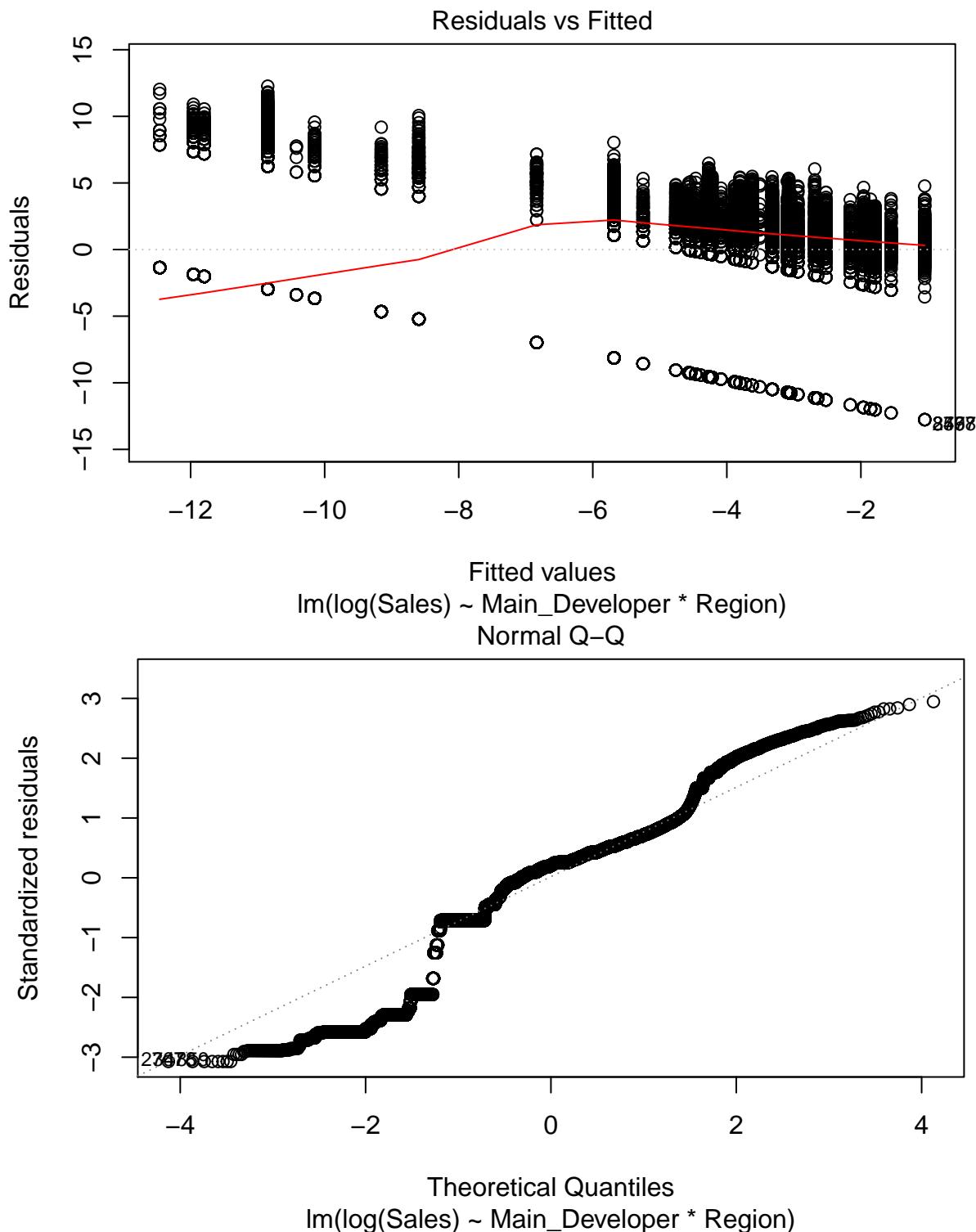
- Interaction plot:

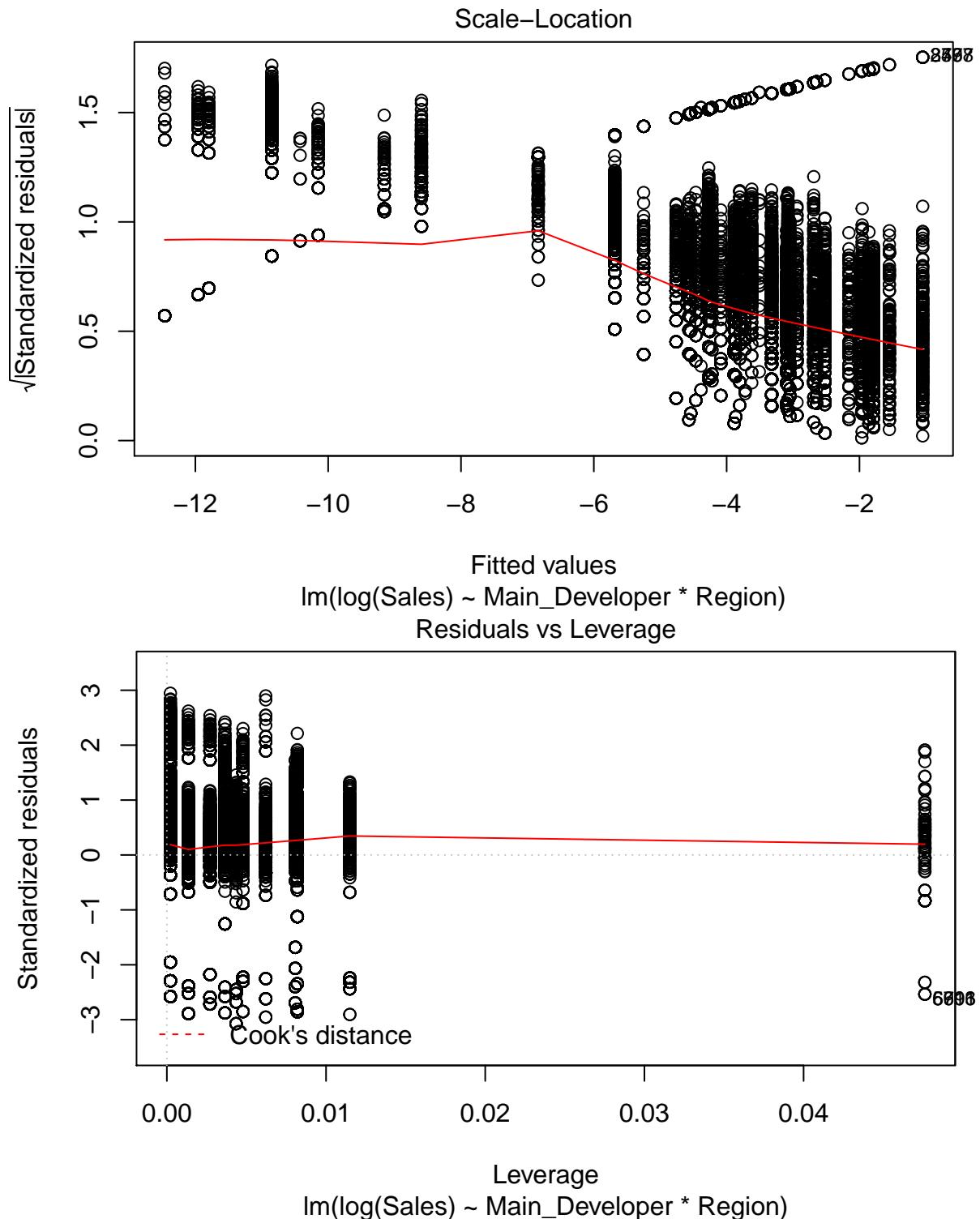
```
with(data=sales, expr={  
  interaction.plot(Main_Developer, Region, response=Sales)  
  interaction.plot(Region, Main_Developer, response=Sales)})
```



- Model checking:

```
plot(dev_region)
```





Comments: There is some pattern in the distribution of residuals, however the variance is stable, i.e. it's approximately the same for all. Hence, we conclude that the homogeneity assumption is not violated, however the normality assumption might be violated. It also can be noted from the QQ-plot. We'll test it by shapiro-wilk test:

```
shapiro.test(dev_region$residuals[1:5000])
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: dev_region$residuals[1:5000]  
## W = 0.62251, p-value < 2.2e-16
```

P-value < 2.2e-16 < 0.05, so the null hypothesis stating that residuals are normally distributed is rejected.

Hypotheses 3 (Publisher company & Region)

- Write up the null/alternative
- Model fit

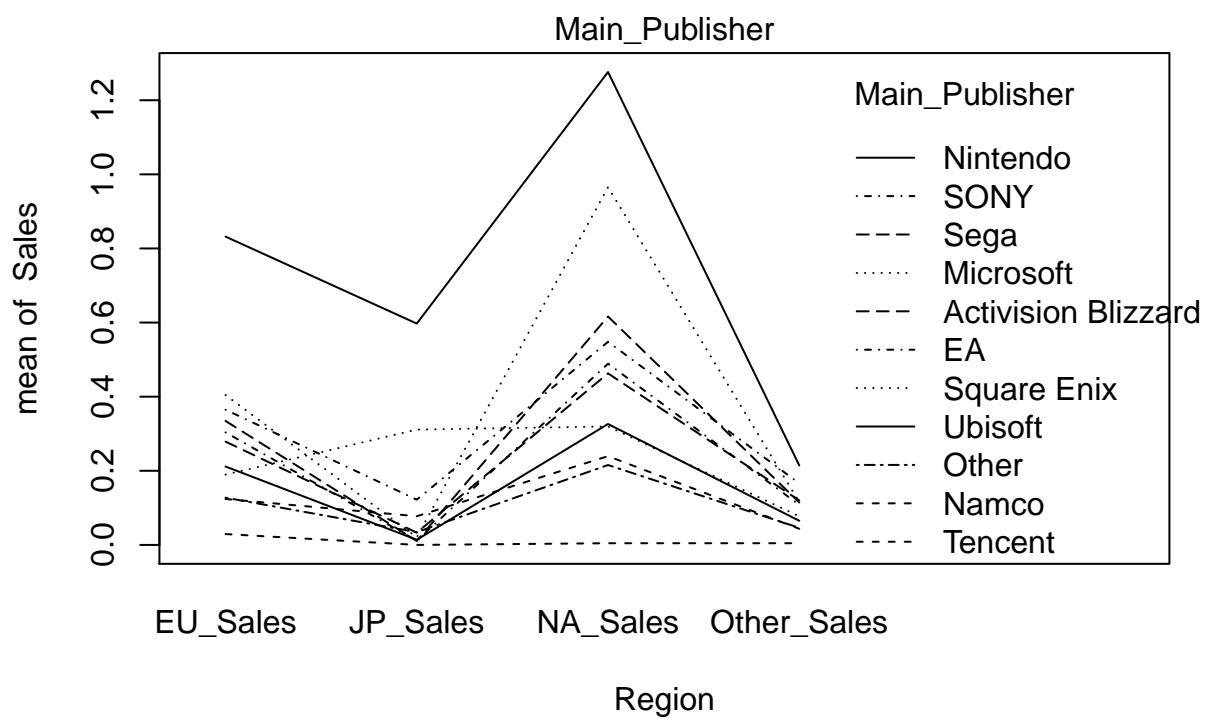
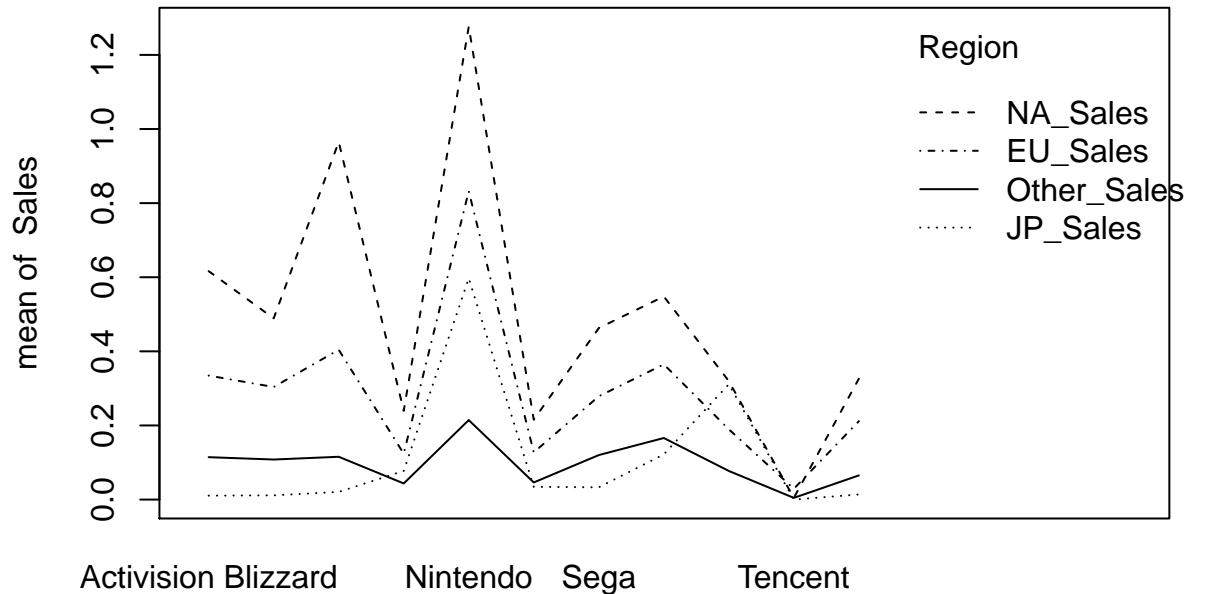
```
publish_region <- lm(log(Sales) ~ Main_Publisher*Region, data = sales)  
anova(publish_region)
```

```
## Analysis of Variance Table  
##  
## Response: log(Sales)  
##  
## Df Sum Sq Mean Sq F value Pr(>F)  
## Main_Publisher 10 22276 2228 132.319 < 2.2e-16 ***  
## Region 3 243201 81067 4815.478 < 2.2e-16 ***  
## Main_Publisher:Region 30 21888 730 43.339 < 2.2e-16 ***  
## Residuals 27256 458846 17  
## ---  
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Conclusion: F-value = 36.592 with 30 degrees of freedom and p-value < 2.2e-16 < 0.05. Hence we reject the null hypothesis and conclude that there is an interaction between genre and publisher company (note that when only these two variables are included in the model).

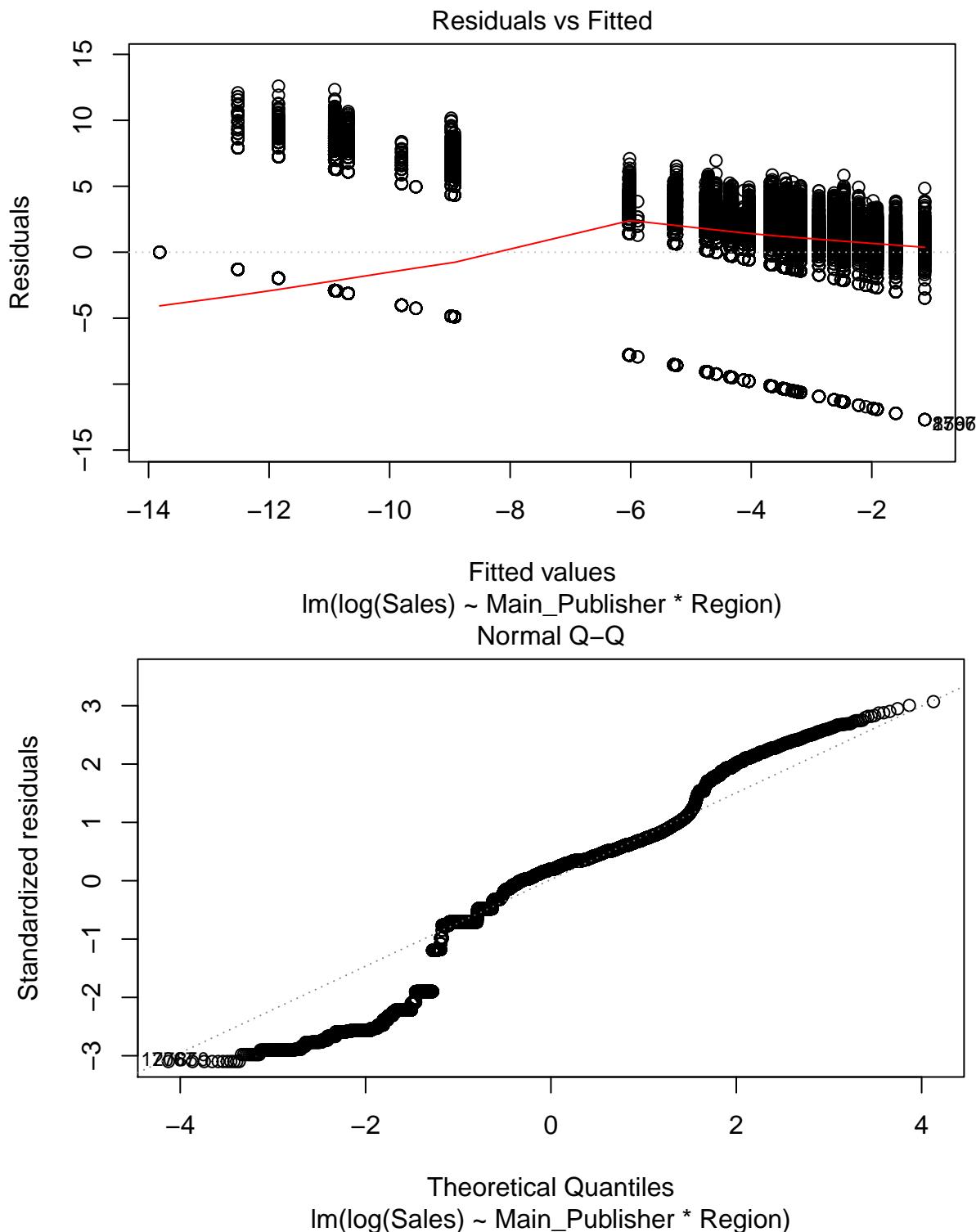
- Interaction plot:

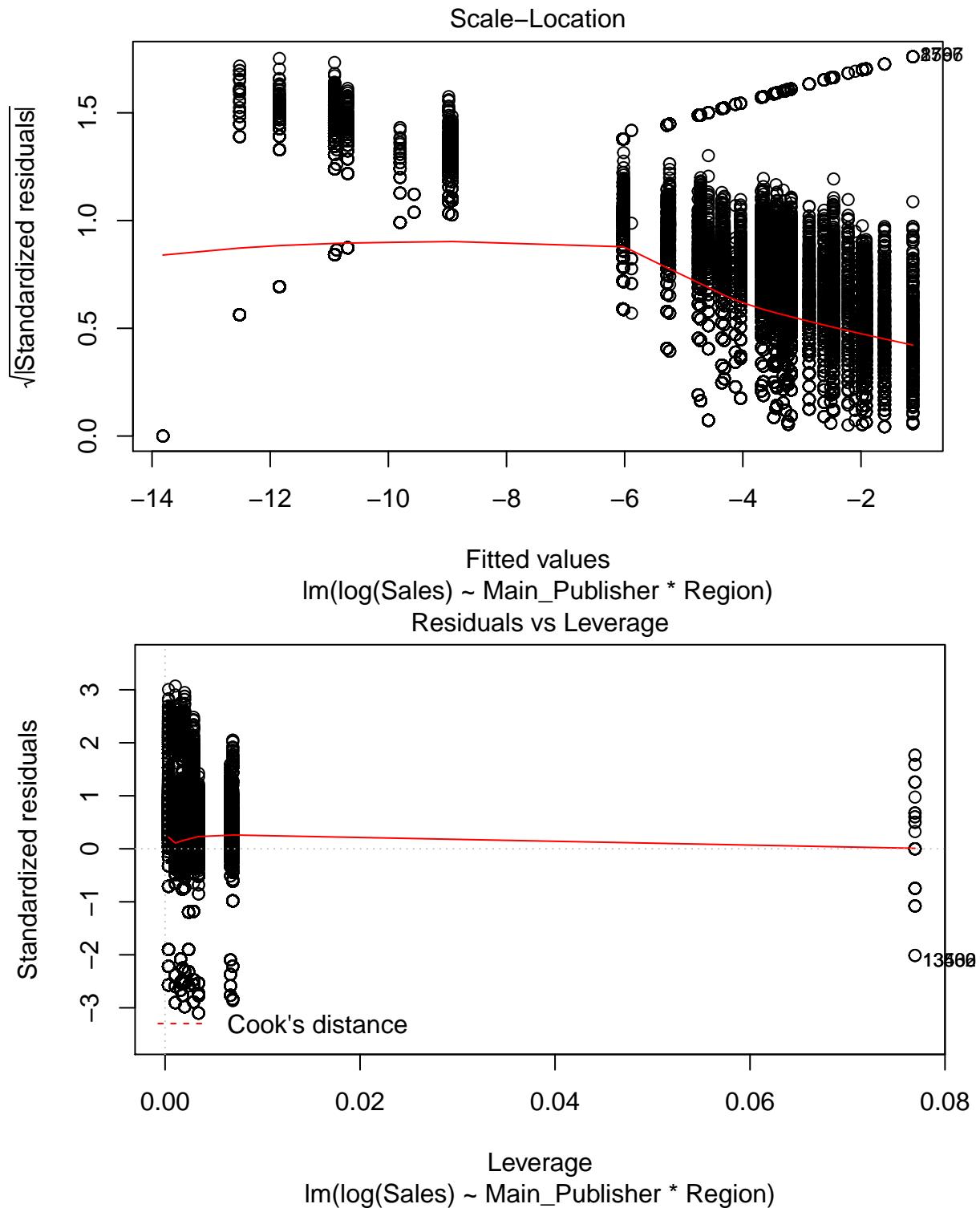
```
with(data=sales, expr={  
  interaction.plot(Main_Publisher, Region, response=Sales)  
  interaction.plot(Region, Main_Publisher, response=Sales)})
```



- Model checking:

```
plot(publish_region)
```





Comments: There is some pattern in the distribution of residuals, however the variance is stable, i.e. it's approximately the same for all. Hence, we conclude that the homogeneity assumption is not violated, however the normality assumption might be violated. It also can be noted from the QQ-plot. We'll test it by shapiro-wilk test:

```
shapiro.test(publish_region$residuals[1:5000])
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: publish_region$residuals[1:5000]  
## W = 0.6287, p-value < 2.2e-16
```

P-value < 2.2e-16 < 0.05, so the null hypothesis stating that residuals are normally distributed is rejected.

Hypotheses 4 (Decade & Genre)

- Write up the null/alternative
- Model fit

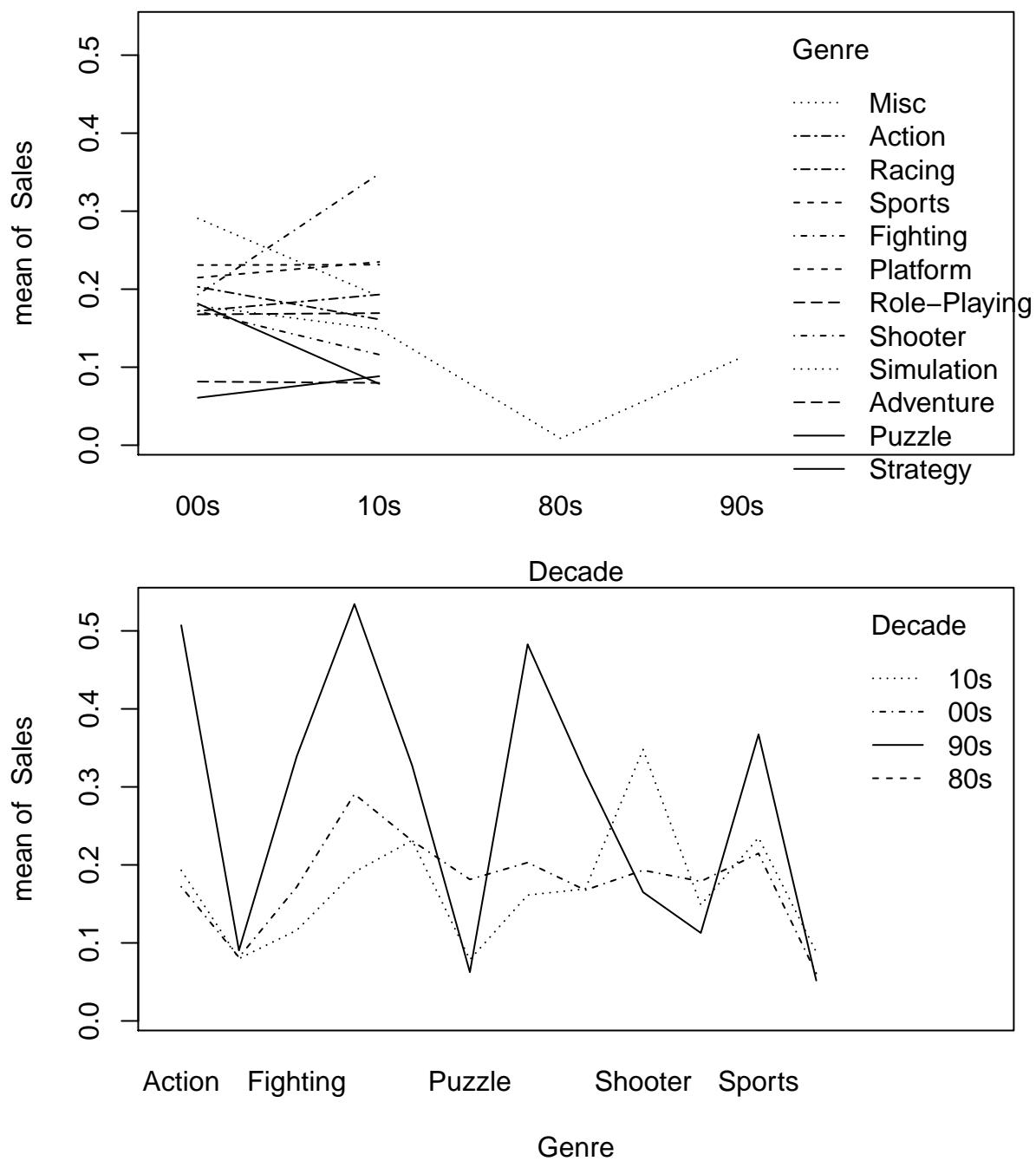
```
decade_genre <- lm(log(Sales) ~ Decade*Genre, data = sales)  
anova(decade_genre)
```

```
## Analysis of Variance Table  
##  
## Response: log(Sales)  
##           Df Sum Sq Mean Sq F value    Pr(>F)  
## Decade       3   3709  1236.27 45.8978 < 2.2e-16 ***  
## Genre        11   6043   549.32 20.3941 < 2.2e-16 ***  
## Decade:Genre  22   2122    96.45  3.5808 2.655e-08 ***  
## Residuals   27263  734337    26.94  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Conclusion: F-value = 3.2986 with 22 degrees of freedom and p-value = 2.668e-07 < 0.05. Hence we reject the null hypothesis and conclude that there is an interaction between genre and decade (note that when only these two variables are included in the model).

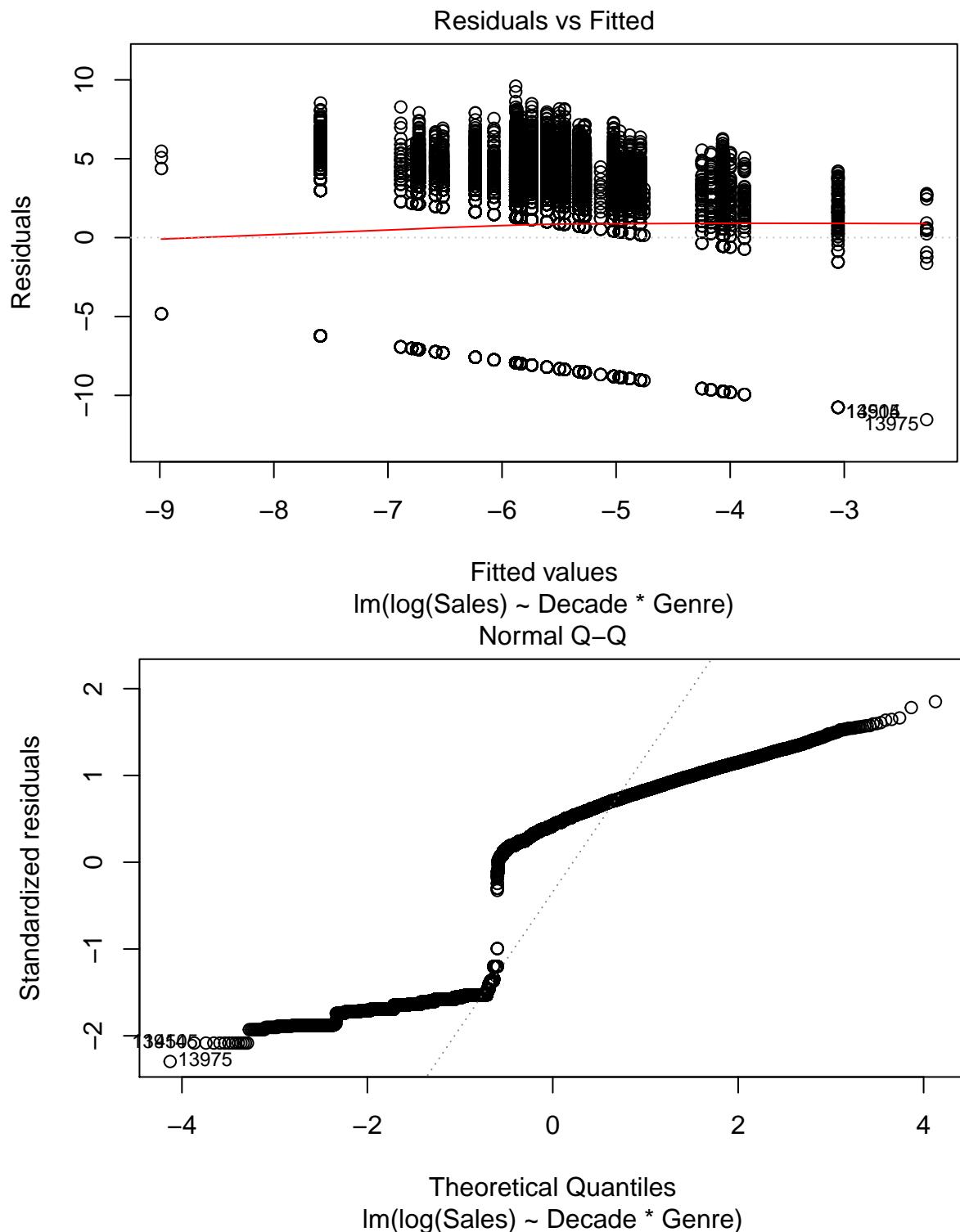
- Interaction plot:

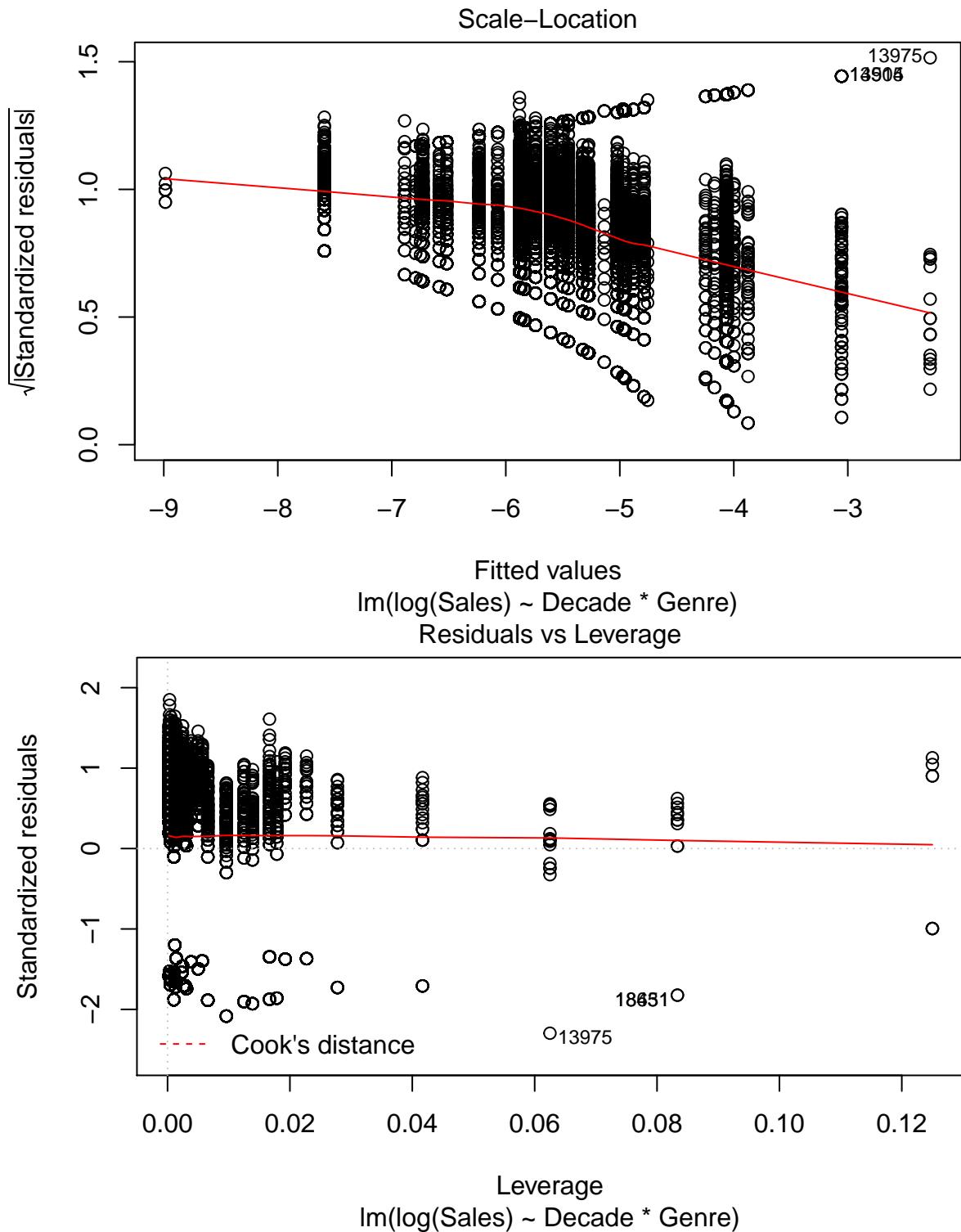
```
with(data=sales, expr={  
  interaction.plot(Decade, Genre, response=Sales)  
  interaction.plot(Genre, Decade, response=Sales)})
```



- Model checking:

```
plot(decade_genre)
```





Comments: There is some pattern in the distribution of residuals, however the variance is stable, i.e. it's approximately the same for all. Hence, we conclude that the homogeneity assumption is not violated, however the normality assumption might be violated. It also can be noted from the QQ-plot. We'll test it by shapiro-wilk test:

```
shapiro.test(decade_genre$residuals[1:5000])
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: decade_genre$residuals[1:5000]  
## W = 0.65772, p-value < 2.2e-16
```

P-value < 2.2e-16 < 0.05, so the null hypothesis stating that residuals are normally distributed is rejected.

Check interaction all together with all categorical variables:

Tried it. It takes hell of a time!!!! SO, let's include the interactions we already checked by ANOVA.

MLR (use sales.csv)

Before fitting MLR, we did some EDA and already eliminated and added some new variables:

- Platform: NOT used, since there are many categories in it. Categories in contrast, were grouped and never interacted
- Year_of_Release: NOT used (Inductive bias)
- Genre: USED (EDA)
- Publisher: NOT used, since there are many categories in it. Categories in contrast, were grouped and never interacted
- Critic_Score : USED (have some relationship from EDA)
- Critic_Count: USED (have some relationship from EDA)
- User_Score: NOT used (collinear relationship with Critic_Score from EDA)
- Used_Count: NOT used (EDA & inductive bias)
- Developer: NOT used, since there are many categories in it. Categories in contrast, were grouped and never interacted
- Rating: USED
- Year_since_Release: NOT used (EDA)
- Decade : USED
- Platform_Company: USED
- Platform_Gen: USED
- Family_Platform: USED
- Main_Developer: USED
- Developer_Country: NOT used (EDA & inductive bias: dependence with other categorical variables)
- Main_Publisher: USED
- Region: USED

Regress all USED variables on *sales* numerical response variable!

Since we want to know **which** properties of games are significant in predicting the sales, region can also be treated as explanatory variable. In this part, our main goal is to

- Identify significant variables
- Assess MLR prediction model on the test data

Assumptions:

- Sales variable has bell-shaped curve (required for iid error terms)
- $y = x^{\top} \beta + \epsilon$ where $y_* = \text{sales}$, and x_i is an explanatory variables
- $\epsilon \sim N(0, \sigma^2)$ iid with fixed variance (homogeneity)
- Add some significant interactions from ANOVA results
- Some variables were omitted due to *inductive bias* or *EDA*

1. MLR with no interactions

Divide the dataset into training (70%) and test(30%) sets. We used stratified partitioning method, so that each level of each categorical variable is contained in equal proportion in each set.

```
set.seed(13)
sets <- stratified(sales, 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
```

Fit the model

```
mlr_without <- 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 %>% 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
## -15.2373  -1.7923   0.6192   2.1935  14.6729
##
## Coefficients: (1 not defined because of singularities)
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -7.789929  1.952487 -3.990 6.63e-05 ***
## GenreAdventure              -0.782735  0.134634 -5.814 6.18e-09 ***
## GenreFighting                0.149153  0.122881  1.214 0.224835
## GenreMisc                   -0.007137  0.120409 -0.059 0.952736
## GenrePlatform               -0.036884  0.120380 -0.306 0.759306
## GenrePuzzle                  -1.360606  0.200145 -6.798 1.08e-11 ***
## GenreRacing                  -0.060492  0.106640 -0.567 0.570543
## GenreRole-Playing             -0.231475  0.096798 -2.391 0.016795 *
## GenreShooter                 -0.282404  0.088288 -3.199 0.001382 **
## GenreSimulation               0.456367  0.135219  3.375 0.000739 ***
## GenreSports                  -0.527069  0.106751 -4.937 7.97e-07 ***
## GenreStrategy                -0.757938  0.141233 -5.367 8.10e-08 ***
## Critic_Score                  0.034511  0.002094 16.483 < 2e-16 ***
## Critic_Count                  0.053240  0.001743 30.547 < 2e-16 ***
## RatingE                      0.517005  1.925515  0.269 0.788315
## RatingE10+                   0.274479  1.925569  0.143 0.886651
## RatingK-A                     2.296814  2.731439  0.841 0.400423
## RatingM                      0.701295  1.924024  0.364 0.715492
## RatingRP                     2.672176  2.788355  0.958 0.337904
## RatingT                      0.303554  1.924264  0.158 0.874655
## Decade10s                     1.045059  0.081621 12.804 < 2e-16 ***
```

```

## Decade80s           -1.006963  1.367020 -0.737  0.461365
## Decade90s            1.454672  0.219418  6.630  3.44e-11 ***
## Platform_CompanyNintendo 3.329039  0.270706 12.298 < 2e-16 ***
## Platform_CompanySega   -1.327560  0.638143 -2.080  0.037504 *
## Platform_CompanySONY    1.303916  0.069820 18.675 < 2e-16 ***
## Platform_Gen6th_Gen     -0.434422  0.238054 -1.825  0.068030 .
## Platform_Gen7th_Gen     -0.553254  0.244025 -2.267  0.023386 *
## Platform_Gen8th_Gen     -1.662718  0.264225 -6.293  3.17e-10 ***
## Family_PlatformMisc    -3.432954  0.256968 -13.359 < 2e-16 ***
## Family_PlatformNintendo_DS -2.730487  0.286443 -9.532 < 2e-16 ***
## Family_PlatformNintendo_Wii -2.188675  0.286458 -7.640  2.24e-14 ***
## Family_PlatformSONY_PS      NA        NA        NA        NA
## Main_DeveloperEA         -0.224181  0.238361 -0.941  0.346965
## Main_DeveloperMicrosoft  -0.311667  0.290956 -1.071  0.284098
## Main_DeveloperNamco      -0.203673  0.273168 -0.746  0.455917
## Main_DeveloperNintendo   0.253633  0.257477  0.985  0.324601
## Main_DeveloperOther       -0.505696  0.188170 -2.687  0.007205 **
## Main_DeveloperSega        -0.007586  0.255538 -0.030  0.976319
## Main_DeveloperSONY        -0.358200  0.296180 -1.209  0.226522
## Main_DeveloperSquare_Enix 0.198156  0.317884  0.623  0.533055
## Main_DeveloperTencent     -0.486796  0.496434 -0.981  0.326807
## Main_DeveloperUbisoft     0.047806  0.243980  0.196  0.844658
## Main_PublisherEA          0.193195  0.165572  1.167  0.243287
## Main_PublisherMicrosoft   0.065633  0.235488  0.279  0.780470
## Main_PublisherNamco       -0.200566  0.154243 -1.300  0.193504
## Main_PublisherNintendo    1.762508  0.199396  8.839 < 2e-16 ***
## Main_PublisherOther        -0.419744  0.115391 -3.638  0.000276 ***
## Main_PublisherSega         -0.096964  0.149319 -0.649  0.516100
## Main_PublisherSONY         -0.711120  0.232778 -3.055  0.002254 **
## Main_PublisherSquare_Enix 0.624233  0.224359  2.782  0.005402 **
## Main_PublisherTencent     -1.387416  0.631074 -2.199  0.027923 *
## Main_PublisherUbisoft     -0.338035  0.162272 -2.083  0.037250 *
## RegionJP_Sales             -6.517479  0.069944 -93.181 < 2e-16 ***
## RegionNA_Sales              1.330765  0.069944 19.026 < 2e-16 ***
## RegionOther_Sales           -1.181419  0.069944 -16.891 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.838 on 24029 degrees of freedom
## Multiple R-squared:  0.4637, Adjusted R-squared:  0.4625
## F-statistic: 384.8 on 54 and 24029 DF,  p-value: < 2.2e-16

```

According to results: the model with no interactions and all possible explanatory variables included has R-adj = 0.4625, s.t. only 46.25% of variation in log(sales) is explained by the the variables {Genre, Critic_Score, Critic_Count, Rating, Decade, Platform_Company, Platform_Gen, Family_Platform, Main_Developer, Main_Publisher, Region}. Residual standard error or *RSE* = 3.838 with 24029 degrees of freedom and *RMSE* = 3.833312. The results table shows that intercept, Critic_Score, Critic_Count, along with many levels of categorical variables are statistically significant. We would also like to analyze the ANOVA table to see the significance of variables as a whole:

ANOVA

```

mlr_without %>% anova()

## Analysis of Variance Table
##
```

```

## Response: log(Sales)
##                               Df Sum Sq Mean Sq F value    Pr(>F)
## Genre                      11   5619    511   34.684 < 2.2e-16 ***
## Critic_Score                 1  18614   18614 1263.851 < 2.2e-16 ***
## Critic_Count                 1  25503   25503 1731.613 < 2.2e-16 ***
## Rating                       6   1017     169   11.507 6.719e-13 ***
## Decade                      3   3151    1050   71.311 < 2.2e-16 ***
## Platform_Company              3  18013    6004   407.694 < 2.2e-16 ***
## Platform_Gen                  3   9145    3048   206.984 < 2.2e-16 ***
## Family_Platform                3   4060    1353   91.898 < 2.2e-16 ***
## Main_Developer                 10  4143     414   28.130 < 2.2e-16 ***
## Main_Publisher                 10  2998     300   20.359 < 2.2e-16 ***
## Region                        3 213781   71260 4838.447 < 2.2e-16 ***
## Residuals                     24029 353897      15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

According to the ANOVA table and individual F-test p-values, each covariate is significant at 0.05 significance level given all other variables in the model. However, we still extract this model fit's p-value and do a stepwise selection/elimination of explanatory variables to achieve the *best* model according to AIC measure:

Do stepwise selection of covariates:

```
step(mlr_without)
```

```

## Start:  AIC=64834.75
## log(Sales) ~ Genre + Critic_Score + Critic_Count + Rating + Decade +
##   Platform_Company + Platform_Gen + Family_Platform + Main_Developer +
##   Main_Publisher + Region
##
##                               Df Sum of Sq    RSS    AIC
## <none>                           353897 64835
## - Rating                      6      568 354465 64861
## - Main_Developer                10     731 354628 64864
## - Platform_Gen                  3     1775 355673 64949
## - Genre                        11     2360 356257 64973
## - Main_Publisher                 10     2997 356894 65018
## - Decade                       3     3024 356921 65034
## - Platform_Company               2     3637 357534 65077
## - Family_Platform                 3     3675 357572 65078
## - Critic_Score                   1     4001 357898 65104
## - Critic_Count                   1     13743 367640 65750
## - Region                        3    213781 567678 76210
##
## 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)
##
## Coefficients:
## (Intercept)          GenreAdventure
## -7.789929            -0.782735
## GenreFighting        GenreMisc
## 0.149153             -0.007137
## GenrePlatform         GenrePuzzle

```

```

##          -0.036884          -1.360606
##      GenreRacing      GenreRole-Playing
##          -0.060492          -0.231475
##      GenreShooter      GenreSimulation
##          -0.282404          0.456367
##      GenreSports      GenreStrategy
##          -0.527069          -0.757938
##      Critic_Score      Critic_Count
##          0.034511          0.053240
##          RatingE          RatingE10+
##          0.517005          0.274479
##          RatingK-A          RatingM
##          2.296814          0.701295
##          RatingRP          RatingT
##          2.672176          0.303554
##      Decade10s          Decade80s
##          1.045059          -1.006963
##      Decade90s          Platform_CompanyNintendo
##          1.454672          3.329039
##      Platform_CompanySega      Platform_CompanySONY
##          -1.327560          1.303916
##      Platform_Gen6th_Gen      Platform_Gen7th_Gen
##          -0.434422          -0.553254
##      Platform_Gen8th_Gen      Family_PlatformMisc
##          -1.662718          -3.432954
##  Family_PlatformNintendo_DS      Family_PlatformNintendo_Wii
##          -2.730487          -2.188675
##  Family_PlatformSONY_PS      Main_DeveloperEA
##          NA          -0.224181
##  Main_DeveloperMicrosoft      Main_DeveloperNamco
##          -0.311667          -0.203673
##  Main_DeveloperNintendo      Main_DeveloperOther
##          0.253633          -0.505696
##  Main_DeveloperSega          Main_DeveloperSONY
##          -0.007586          -0.358200
##  Main_DeveloperSquare_Enix      Main_DeveloperTencent
##          0.198156          -0.486796
##  Main_DeveloperUbisoft      Main_PublisherEA
##          0.047806          0.193195
##  Main_PublisherMicrosoft      Main_PublisherNamco
##          0.065633          -0.200566
##  Main_PublisherNintendo      Main_PublisherOther
##          1.762508          -0.419744
##  Main_PublisherSega          Main_PublisherSONY
##          -0.096964          -0.711120
##  Main_PublisherSquare_Enix      Main_PublisherTencent
##          0.624233          -1.387416
##  Main_PublisherUbisoft          RegionJP_Sales
##          -0.338035          -6.517479
##  RegionNA_Sales          RegionOther_Sales
##          1.330765          -1.181419

```

The stepwise model selection suggests the original model where the AIC = 64834.75.

2. MLR with interactions

The previous MLR_without model is still not the final model, so we would like to include interactions which were found (each pair individually) significant in the beginning ANOVA part:

```

mlr_with <- 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 %>% 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
## -14.2999 -1.4199  0.5633  2.0847 14.8089
##
## Coefficients: (12 not defined because of singularities)
##                                     Estimate Std. Error t value
## (Intercept)                 -7.0410952  1.8901204 -3.725
## GenreAdventure              -0.5530552  0.2703817 -2.045
## GenreFighting               -0.5900121  0.2432510 -2.426
## GenreMisc                   -0.6928972  0.2337338 -2.964
## GenrePlatform                0.2239380  0.2308953  0.970
## GenrePuzzle                  -3.1938065  0.3815989 -8.370
## GenreRacing                  0.5627560  0.2013752  2.795
## GenreRole-Playing            -1.5060875  0.1933036 -7.791
## GenreShooter                 -0.3391489  0.1740092 -1.949
## GenreSimulation              0.8663837  0.2613215  3.315
## GenreSports                  -1.0492481  0.1921406 -5.461
## GenreStrategy                0.3320477  0.2717006  1.222
## Critic_Score                  0.0354142  0.0020179 17.550
## Critic_Count                  0.0531238  0.0016767 31.684
## RatingE                      0.4633340  1.8439719  0.251
## RatingE10+                    0.2534288  1.8440263  0.137
## RatingK-A                     -0.3965742  2.8181586 -0.141
## RatingM                      0.6363457  1.8425638  0.345
## RatingRP                     2.2403846  2.6840841  0.835
## RatingT                      0.2808421  1.8427620  0.152
## Decade10s                     0.9806318  0.1171880  8.368
## Decade80s                     -1.0007691  1.3103685 -0.764
## Decade90s                     2.4160187  0.4280163  5.645
## Platform_CompanyNintendo     3.1357537  0.2624671 11.947
## Platform_CompanySega          -1.6054687  0.6187039 -2.595
## Platform_CompanySONY          1.2999294  0.0669748 19.409
## Platform_Gen6th Gen           -0.2556462  0.2319945 -1.102
## Platform_Gen7th Gen           -0.3843653  0.2376975 -1.617
## Platform_Gen8th Gen           -1.3820419  0.2573738 -5.370
## Family_PlatformMisc           -3.2828698  0.2496390 -13.150

```

## Family_PlatformNintendo_DS	-2.5267276	0.2776036	-9.102
## Family_PlatformNintendo_Wii	-1.9887849	0.2776535	-7.163
## Family_PlatformSONY_PS	NA	NA	NA
## Main_DeveloperEA	-0.9035524	0.4556812	-1.983
## Main_DeveloperMicrosoft	-0.7396379	0.5559635	-1.330
## Main_DeveloperNamco	-1.3101151	0.5216858	-2.511
## Main_DeveloperNintendo	-1.0185033	0.4920350	-2.070
## Main_DeveloperOther	-0.9926555	0.3589381	-2.766
## Main_DeveloperSega	-1.4990889	0.4880480	-3.072
## Main_DeveloperSONY	-1.7133570	0.5640235	-3.038
## Main_DeveloperSquare_Enix	-1.2865758	0.6066533	-2.121
## Main_DeveloperTencent	0.0715244	0.9487512	0.075
## Main_DeveloperUbisoft	-0.2543729	0.4658189	-0.546
## Main_PublisherEA	0.4770721	0.3159792	1.510
## Main_PublisherMicrosoft	0.7601060	0.4458696	1.705
## Main_PublisherNamco	-0.3810062	0.2945824	-1.293
## Main_PublisherNintendo	0.6261959	0.3691821	1.696
## Main_PublisherOther	-0.8004780	0.2194129	-3.648
## Main_PublisherSega	-0.0733221	0.2838075	-0.258
## Main_PublisherSONY	-0.9826779	0.4396417	-2.235
## Main_PublisherSquare_Enix	0.3977099	0.4276201	0.930
## Main_PublisherTencent	-0.3463002	1.1635696	-0.298
## Main_PublisherUbisoft	-0.0825459	0.3090217	-0.267
## RegionJP_Sales	-9.6908164	0.4536169	-21.363
## RegionNA_Sales	1.0132516	0.4536177	2.234
## RegionOther_Sales	-1.5454250	0.4536176	-3.407
## GenreAdventure:RegionJP_Sales	-0.0040066	0.3629839	-0.011
## GenreFighting:RegionJP_Sales	1.5688832	0.3237983	4.845
## GenreMisc:RegionJP_Sales	-0.3937440	0.3150384	-1.250
## GenrePlatform:RegionJP_Sales	-0.5476305	0.3080963	-1.777
## GenrePuzzle:RegionJP_Sales	2.4903831	0.5218743	4.772
## GenreRacing:RegionJP_Sales	-1.4348196	0.2689095	-5.336
## GenreRole-Playing:RegionJP_Sales	3.4734044	0.2570152	13.514
## GenreShooter:RegionJP_Sales	-0.3604365	0.2314582	-1.557
## GenreSimulation:RegionJP_Sales	-0.0160929	0.3532026	-0.046
## GenreSports:RegionJP_Sales	-0.4265643	0.2540193	-1.679
## GenreStrategy:RegionJP_Sales	-1.1887946	0.3672620	-3.237
## GenreAdventure:RegionNA_Sales	-0.0646230	0.3629839	-0.178
## GenreFighting:RegionNA_Sales	0.9299026	0.3237984	2.872
## GenreMisc:RegionNA_Sales	1.9495089	0.3150376	6.188
## GenrePlatform:RegionNA_Sales	0.6713842	0.3080963	2.179
## GenrePuzzle:RegionNA_Sales	3.0643608	0.5218744	5.872
## GenreRacing:RegionNA_Sales	-0.3369421	0.2689101	-1.253
## GenreRole-Playing:RegionNA_Sales	0.8134207	0.2570158	3.165
## GenreShooter:RegionNA_Sales	-0.0198123	0.2314583	-0.086
## GenreSimulation:RegionNA_Sales	-1.1474200	0.3532028	-3.249
## GenreSports:RegionNA_Sales	1.5236688	0.2540202	5.998
## GenreStrategy:RegionNA_Sales	-2.1091104	0.3672629	-5.743
## GenreAdventure:RegionOther_Sales	-0.2282881	0.3629841	-0.629
## GenreFighting:RegionOther_Sales	0.5715755	0.3237983	1.765
## GenreMisc:RegionOther_Sales	1.3611506	0.3150374	4.321
## GenrePlatform:RegionOther_Sales	-0.3306548	0.3080963	-1.073
## GenrePuzzle:RegionOther_Sales	2.3450481	0.5218745	4.494
## GenreRacing:RegionOther_Sales	-0.2845912	0.2689102	-1.058

## GenreRole-Playing:RegionOther_Sales	0.8888850	0.2570157	3.458
## GenreShooter:RegionOther_Sales	-0.2636937	0.2314588	-1.139
## GenreSimulation:RegionOther_Sales	-0.3572546	0.3532026	-1.011
## GenreSports:RegionOther_Sales	0.9100572	0.2540211	3.583
## GenreStrategy:RegionOther_Sales	-1.1736023	0.3672629	-3.196
## Main_DeveloperEA:RegionJP_Sales	0.7934877	0.6435348	1.233
## Main_DeveloperMicrosoft:RegionJP_Sales	0.3653145	0.7850432	0.465
## Main_DeveloperNamco:RegionJP_Sales	3.0362243	0.7366393	4.122
## Main_DeveloperNintendo:RegionJP_Sales	2.2484896	0.6948846	3.236
## Main_DeveloperOther:RegionJP_Sales	0.9065253	0.5067833	1.789
## Main_DeveloperSega:RegionJP_Sales	2.8876983	0.6892558	4.190
## Main_DeveloperSONY:RegionJP_Sales	2.6054829	0.7952018	3.277
## Main_DeveloperSquare Enix:RegionJP_Sales	3.7002101	0.8544090	4.331
## Main_DeveloperTencent:RegionJP_Sales	-0.5330404	1.3392642	-0.398
## Main_DeveloperUbisoft:RegionJP_Sales	0.4822698	0.6575946	0.733
## Main_DeveloperEA:RegionNA_Sales	0.5352601	0.6435385	0.832
## Main_DeveloperMicrosoft:RegionNA_Sales	0.3680635	0.7850431	0.469
## Main_DeveloperNamco:RegionNA_Sales	0.6520135	0.7366420	0.885
## Main_DeveloperNintendo:RegionNA_Sales	1.3920305	0.6948875	2.003
## Main_DeveloperOther:RegionNA_Sales	0.5366285	0.5067850	1.059
## Main_DeveloperSega:RegionNA_Sales	1.3690892	0.6892571	1.986
## Main_DeveloperSONY:RegionNA_Sales	1.0301884	0.7952017	1.296
## Main_DeveloperSquare Enix:RegionNA_Sales	1.7200756	0.8544061	2.013
## Main_DeveloperTencent:RegionNA_Sales	-1.7972460	1.3392641	-1.342
## Main_DeveloperUbisoft:RegionNA_Sales	0.0450394	0.6575954	0.068
## Main_DeveloperEA:RegionOther_Sales	1.2629823	0.6435355	1.963
## Main_DeveloperMicrosoft:RegionOther_Sales	0.9457722	0.7850431	1.205
## Main_DeveloperNamco:RegionOther_Sales	0.8027185	0.7366394	1.090
## Main_DeveloperNintendo:RegionOther_Sales	1.5520409	0.6948857	2.234
## Main_DeveloperOther:RegionOther_Sales	0.5001005	0.5067838	0.987
## Main_DeveloperSega:RegionOther_Sales	1.7265898	0.6892574	2.505
## Main_DeveloperSONY:RegionOther_Sales	1.7743170	0.7952018	2.231
## Main_DeveloperSquare Enix:RegionOther_Sales	0.4189766	0.8544070	0.490
## Main_DeveloperTencent:RegionOther_Sales	-0.0105046	1.3392652	-0.008
## Main_DeveloperUbisoft:RegionOther_Sales	0.6355513	0.6575942	0.966
## Main_PublisherEA:RegionJP_Sales	0.4563650	0.4457965	1.024
## Main_PublisherMicrosoft:RegionJP_Sales	1.1644328	0.6278483	1.855
## Main_PublisherNamco:RegionJP_Sales	2.8510913	0.4158026	6.857
## Main_PublisherNintendo:RegionJP_Sales	7.1597642	0.5155789	13.887
## Main_PublisherOther:RegionJP_Sales	2.3361162	0.3093367	7.552
## Main_PublisherSega:RegionJP_Sales	1.2982281	0.3999974	3.246
## Main_PublisherSONY:RegionJP_Sales	2.0416849	0.6177750	3.305
## Main_PublisherSquare Enix:RegionJP_Sales	3.5537343	0.6031558	5.892
## Main_PublisherTencent:RegionJP_Sales	1.7689752	1.6230352	1.090
## Main_PublisherUbisoft:RegionJP_Sales	0.7536662	0.4356913	1.730
## Main_PublisherEA:RegionNA_Sales	-0.7983365	0.4457964	-1.791
## Main_PublisherMicrosoft:RegionNA_Sales	-1.7610963	0.6278460	-2.805
## Main_PublisherNamco:RegionNA_Sales	-0.8647817	0.4158027	-2.080
## Main_PublisherNintendo:RegionNA_Sales	-1.3977739	0.5155790	-2.711
## Main_PublisherOther:RegionNA_Sales	-0.2712843	0.3093368	-0.877
## Main_PublisherSega:RegionNA_Sales	-1.0136068	0.3999976	-2.534
## Main_PublisherSONY:RegionNA_Sales	-0.9529198	0.6177749	-1.543
## Main_PublisherSquare Enix:RegionNA_Sales	-1.9423174	0.6031559	-3.220
## Main_PublisherTencent:RegionNA_Sales	-3.9632654	1.6230424	-2.442

```

## Main_PublisherUbisoft:RegionNA_Sales      -0.8084949  0.4356907 -1.856
## Main_PublisherEA:RegionOther_Sales        -0.6635794  0.4457966 -1.489
## Main_PublisherMicrosoft:RegionOther_Sales -2.0989122  0.6278499 -3.343
## Main_PublisherNamco:RegionOther_Sales     -1.0602315  0.4158028 -2.550
## Main_PublisherNintendo:RegionOther_Sales   -1.1128013  0.5155792 -2.158
## Main_PublisherOther:RegionOther_Sales      -0.4194465  0.3093369 -1.356
## Main_PublisherSega:RegionOther_Sales       -0.5394534  0.3999974 -1.349
## Main_PublisherSONY:RegionOther_Sales       0.0806539  0.6177753  0.131
## Main_PublisherSquare Enix:RegionOther_Sales -0.5063326  0.6031561 -0.839
## Main_PublisherTencent:RegionOther_Sales    -1.7836246  1.6230419 -1.099
## Main_PublisherUbisoft:RegionOther_Sales    -0.7327943  0.4356911 -1.682
## GenreAdventure:Decade10s                  -0.6778059  0.2887825 -2.347
## GenreFighting:Decade10s                 -0.2520382  0.2532764 -0.995
## GenreMisc:Decade10s                      -0.3000875  0.2664968 -1.126
## GenrePlatform:Decade10s                  -0.8263652  0.2592600 -3.187
## GenrePuzzle:Decade10s                   -1.3468849  0.5578319 -2.414
## GenreRacing:Decade10s                   -0.4808509  0.2361747 -2.036
## GenreRole-Playing:Decade10s            -0.1217285  0.1952029 -0.624
## GenreShooter:Decade10s                  1.0016639  0.1812309  5.527
## GenreSimulation:Decade10s             -0.0005734  0.3217425 -0.002
## GenreSports:Decade10s                  0.2596887  0.2014806  1.289
## GenreStrategy:Decade10s                0.4878773  0.3453719  1.413
## 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              0.3469493  0.8609766  0.403
## GenreFighting:Decade90s            -0.2329749  0.6088991 -0.383
## GenreMisc:Decade90s                 1.8063927  1.1373763  1.588
## GenrePlatform:Decade90s            -1.4738620  0.6515461 -2.262
## GenrePuzzle:Decade90s              -0.4332273  1.1476791 -0.377
## GenreRacing:Decade90s              -1.0525427  0.5752694 -1.830
## GenreRole-Playing:Decade90s        -0.6315650  0.5147124 -1.227
## GenreShooter:Decade90s              -2.6445377  0.6352692 -4.163
## GenreSimulation:Decade90s          -1.6511042  0.7432408 -2.221
## GenreSports:Decade90s              -1.0051856  0.6660901 -1.509
## GenreStrategy:Decade90s            -1.9072300  0.6977454 -2.733
## Pr(>|t|)                           0.000196 *** 
## (Intercept)                         0.040820 *  
## GenreAdventure                       0.015294 *  
## GenreFighting                        0.003035 ** 
## GenreMisc                            0.332122
## GenrePlatform                         < 2e-16 ***
## GenrePuzzle                          0.005201 ** 
## GenreRacing                           6.90e-15 ***
## GenreRole-Playing                     0.051304 .

```

## GenreSimulation	0.000917 ***
## GenreSports	4.79e-08 ***
## GenreStrategy	0.221678
## Critic_Score	< 2e-16 ***
## Critic_Count	< 2e-16 ***
## RatingE	0.801608
## RatingE10+	0.890690
## RatingK-A	0.888092
## RatingM	0.729828
## RatingRP	0.403899
## RatingT	0.878871
## Decade10s	< 2e-16 ***
## Decade80s	0.445035
## Decade90s	1.67e-08 ***
## Platform_CompanyNintendo	< 2e-16 ***
## Platform_CompanySega	0.009468 **
## Platform_CompanySONY	< 2e-16 ***
## Platform_Gen6th Gen	0.270495
## Platform_Gen7th Gen	0.105884
## Platform_Gen8th Gen	7.96e-08 ***
## Family_PlatformMisc	< 2e-16 ***
## Family_PlatformNintendo_DS	< 2e-16 ***
## Family_PlatformNintendo_Wii	8.13e-13 ***
## Family_PlatformSONY_PS	NA
## Main_DeveloperEA	0.047394 *
## Main_DeveloperMicrosoft	0.183409
## Main_DeveloperNamco	0.012035 *
## Main_DeveloperNintendo	0.038465 *
## Main_DeveloperOther	0.005687 **
## Main_DeveloperSega	0.002132 **
## Main_DeveloperSONY	0.002386 **
## Main_DeveloperSquare Enix	0.033951 *
## Main_DeveloperTencent	0.939907
## Main_DeveloperUbisoft	0.585018
## Main_PublisherEA	0.131102
## Main_PublisherMicrosoft	0.088250 .
## Main_PublisherNamco	0.195893
## Main_PublisherNintendo	0.089867 .
## Main_PublisherOther	0.000265 ***
## Main_PublisherSega	0.796138
## Main_PublisherSONY	0.025415 *
## Main_PublisherSquare Enix	0.352352
## Main_PublisherTencent	0.765997
## Main_PublisherUbisoft	0.789379
## RegionJP_Sales	< 2e-16 ***
## RegionNA_Sales	0.025511 *
## RegionOther_Sales	0.000658 ***
## GenreAdventure:RegionJP_Sales	0.991193
## GenreFighting:RegionJP_Sales	1.27e-06 ***
## GenreMisc:RegionJP_Sales	0.211374
## GenrePlatform:RegionJP_Sales	0.075504 .
## GenrePuzzle:RegionJP_Sales	1.83e-06 ***
## GenreRacing:RegionJP_Sales	9.60e-08 ***
## GenreRole-Playing:RegionJP_Sales	< 2e-16 ***

## GenreShooter:RegionJP_Sales	0.119426
## GenreSimulation:RegionJP_Sales	0.963659
## GenreSports:RegionJP_Sales	0.093115 .
## GenreStrategy:RegionJP_Sales	0.001210 **
## GenreAdventure:RegionNA_Sales	0.858699
## GenreFighting:RegionNA_Sales	0.004084 **
## GenreMisc:RegionNA_Sales	6.19e-10 ***
## GenrePlatform:RegionNA_Sales	0.029331 *
## GenrePuzzle:RegionNA_Sales	4.37e-09 ***
## GenreRacing:RegionNA_Sales	0.210221
## GenreRole-Playing:RegionNA_Sales	0.001553 **
## GenreShooter:RegionNA_Sales	0.931787
## GenreSimulation:RegionNA_Sales	0.001161 **
## GenreSports:RegionNA_Sales	2.02e-09 ***
## GenreStrategy:RegionNA_Sales	9.43e-09 ***
## GenreAdventure:RegionOther_Sales	0.529407
## GenreFighting:RegionOther_Sales	0.077539 .
## GenreMisc:RegionOther_Sales	1.56e-05 ***
## GenrePlatform:RegionOther_Sales	0.283184
## GenrePuzzle:RegionOther_Sales	7.04e-06 ***
## GenreRacing:RegionOther_Sales	0.289923
## GenreRole-Playing:RegionOther_Sales	0.000544 ***
## GenreShooter:RegionOther_Sales	0.254603
## GenreSimulation:RegionOther_Sales	0.311801
## GenreSports:RegionOther_Sales	0.000341 ***
## GenreStrategy:RegionOther_Sales	0.001398 **
## Main_DeveloperEA:RegionJP_Sales	0.217582
## Main_DeveloperMicrosoft:RegionJP_Sales	0.641690
## Main_DeveloperNamco:RegionJP_Sales	3.77e-05 ***
## Main_DeveloperNintendo:RegionJP_Sales	0.001215 **
## Main_DeveloperOther:RegionJP_Sales	0.073662 .
## Main_DeveloperSega:RegionJP_Sales	2.80e-05 ***
## Main_DeveloperSONY:RegionJP_Sales	0.001052 **
## Main_DeveloperSquare Enix:RegionJP_Sales	1.49e-05 ***
## Main_DeveloperTencent:RegionJP_Sales	0.690626
## Main_DeveloperUbisoft:RegionJP_Sales	0.463331
## Main_DeveloperEA:RegionNA_Sales	0.405561
## Main_DeveloperMicrosoft:RegionNA_Sales	0.639185
## Main_DeveloperNamco:RegionNA_Sales	0.376103
## Main_DeveloperNintendo:RegionNA_Sales	0.045162 *
## Main_DeveloperOther:RegionNA_Sales	0.289661
## Main_DeveloperSega:RegionNA_Sales	0.047009 *
## Main_DeveloperSONY:RegionNA_Sales	0.195158
## Main_DeveloperSquare Enix:RegionNA_Sales	0.044107 *
## Main_DeveloperTencent:RegionNA_Sales	0.179620
## Main_DeveloperUbisoft:RegionNA_Sales	0.945395
## Main_DeveloperEA:RegionOther_Sales	0.049708 *
## Main_DeveloperMicrosoft:RegionOther_Sales	0.228316
## Main_DeveloperNamco:RegionOther_Sales	0.275855
## Main_DeveloperNintendo:RegionOther_Sales	0.025524 *
## Main_DeveloperOther:RegionOther_Sales	0.323745
## Main_DeveloperSega:RegionOther_Sales	0.012252 *
## Main_DeveloperSONY:RegionOther_Sales	0.025672 *
## Main_DeveloperSquare Enix:RegionOther_Sales	0.623876

```

## Main_DeveloperTencent:RegionOther_Sales      0.993742
## Main_DeveloperUbisoft:RegionOther_Sales      0.333814
## Main_PublisherEA:RegionJP_Sales              0.305984
## Main_PublisherMicrosoft:RegionJP_Sales       0.063660 .
## Main_PublisherNamco:RegionJP_Sales           7.21e-12 ***
## Main_PublisherNintendo:RegionJP_Sales         < 2e-16 ***
## Main_PublisherOther:RegionJP_Sales            4.44e-14 ***
## Main_PublisherSega:RegionJP_Sales             0.001174 **
## Main_PublisherSONY:RegionJP_Sales             0.000951 ***
## Main_PublisherSquare_Einx:RegionJP_Sales      3.87e-09 ***
## Main_PublisherTencent:RegionJP_Sales          0.275760
## Main_PublisherUbisoft:RegionJP_Sales          0.083676 .
## Main_PublisherEA:RegionNA_Sales               0.073336 .
## Main_PublisherMicrosoft:RegionNA_Sales        0.005036 **
## Main_PublisherNamco:RegionNA_Sales            0.037556 *
## Main_PublisherNintendo:RegionNA_Sales         0.006711 **
## Main_PublisherOther:RegionNA_Sales            0.380503
## Main_PublisherSega:RegionNA_Sales             0.011282 *
## Main_PublisherSONY:RegionNA_Sales             0.122965
## Main_PublisherSquare_Einx:RegionNA_Sales      0.001282 **
## Main_PublisherTencent:RegionNA_Sales          0.014618 *
## Main_PublisherUbisoft:RegionNA_Sales          0.063514 .
## Main_PublisherEA:RegionOther_Sales            0.136626
## Main_PublisherMicrosoft:RegionOther_Sales     0.000830 ***
## Main_PublisherNamco:RegionOther_Sales          0.010783 *
## Main_PublisherNintendo:RegionOther_Sales       0.030910 *
## Main_PublisherOther:RegionOther_Sales          0.175127
## Main_PublisherSega:RegionOther_Sales          0.177465
## Main_PublisherSONY:RegionOther_Sales          0.896128
## Main_PublisherSquare_Einx:RegionOther_Sales   0.401213
## Main_PublisherTencent:RegionOther_Sales        0.271806
## Main_PublisherUbisoft:RegionOther_Sales        0.092599 .
## GenreAdventure:Decade10s                      0.018927 *
## GenreFighting:Decade10s                      0.319692
## GenreMisc:Decade10s                          0.260158
## GenrePlatform:Decade10s                     0.001437 **
## GenrePuzzle:Decade10s                        0.015764 *
## GenreRacing:Decade10s                        0.041762 *
## GenreRole-Playing:Decade10s                 0.532897
## GenreShooter:Decade10s                       3.29e-08 ***
## GenreSimulation:Decade10s                   0.998578
## GenreSports:Decade10s                        0.197445
## GenreStrategy:Decade10s                      0.157782
## 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          0.686973
## GenreFighting:Decade90s         0.702007
## GenreMisc:Decade90s            0.112252
## GenrePlatform:Decade90s        0.023700 *
## GenrePuzzle:Decade90s          0.705819
## GenreRacing:Decade90s          0.067314 .
## GenreRole-Playing:Decade90s    0.219825
## GenreShooter:Decade90s         3.15e-05 ***
## GenreSimulation:Decade90s      0.026327 *
## GenreSports:Decade90s          0.131291
## GenreStrategy:Decade90s        0.006273 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.675 on 23914 degrees of freedom
## Multiple R-squared:  0.5107, Adjusted R-squared:  0.5072
## F-statistic: 147.7 on 169 and 23914 DF, p-value: < 2.2e-16

```

Comments: type something as in previous case. $RMSE = 3.661722$.

ANOVA:

```
anova(mlr_with)
```

```

## Analysis of Variance Table
##
## Response: log(Sales)
##                               Df Sum Sq Mean Sq   F value   Pr(>F)
## Genre                           11  5619   511   37.8286 < 2.2e-16 ***
## Critic_Score                     1 18614  18614  1378.4469 < 2.2e-16 ***
## Critic_Count                     1 25503  25503  1888.6209 < 2.2e-16 ***
## Rating                          6   1017   169   12.5502 3.509e-14 ***
## Decade                          3   3151   1050   77.7772 < 2.2e-16 ***
## Platform_Company                 3   18013   6004   444.6605 < 2.2e-16 ***
## Platform_Gen                     3   9145   3048   225.7521 < 2.2e-16 ***
## Family_Platform                  3   4060   1353   100.2304 < 2.2e-16 ***
## Main_Developer                   10  4143    414   30.6805 < 2.2e-16 ***
## Main_Publisher                   10  2998    300   22.2051 < 2.2e-16 ***
## Region                          3 213781  71260  5277.1569 < 2.2e-16 ***
## Genre:Region                      33 14570    442   32.6968 < 2.2e-16 ***
## Main_Developer:Region             30  8209    274   20.2643 < 2.2e-16 ***
## Main_Publisher:Region              30  6551    218   16.1703 < 2.2e-16 ***
## Genre:Decade                      22  1644     75   5.5327 9.463e-16 ***
## Residuals                        23914 322923     14
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

All possible variables and interactions are significant!!!

Now choose model by the AIC:

```
step(mlr_with)
```

```

## Start:  AIC=62858.87
## 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
##
##                                Df Sum of Sq    RSS   AIC
## <none>                           322923 62859
## - Rating                      6     429.4 323353 62879
## - Main_Developer:Region       30    1247.2 324171 62892
## - Genre:Decade                22    1643.6 324567 62937
## - Platform_Gen                 3    1378.6 324302 62955
## - Platform_Company             2    3322.4 326246 63101
## - Family_Platform              3    3470.5 326394 63110
## - Critic_Score                  1    4159.0 327082 63165
## - Main_Publisher:Region       30    6550.8 329474 63283
## - Genre:Region                  33    8758.7 331682 63437
## - Critic_Count                  1   13555.6 336479 63847
##
## 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)
##
## Coefficients:
##                               (Intercept)
##                               -7.0410952
##      GenreAdventure
##                               -0.5530552
##      GenreFighting
##                               -0.5900121
##      GenreMisc
##                               -0.6928972
##      GenrePlatform
##                               0.2239380
##      GenrePuzzle
##                               -3.1938065
##      GenreRacing
##                               0.5627560
##      GenreRole-Playing
##                               -1.5060875
##      GenreShooter
##                               -0.3391489
##      GenreSimulation
##                               0.8663837
##      GenreSports
##                               -1.0492481
##      GenreStrategy
##                               0.3320477
##      Critic_Score
##                               0.0354142
##      Critic_Count
##                               0.0531238
##      RatingE
##                               0.4633340

```

```

##                               RatingE10+
##                               0.2534288
##                               RatingK-A
##                               -0.3965742
##                               RatingM
##                               0.6363457
##                               RatingRP
##                               2.2403846
##                               RatingT
##                               0.2808421
##                               Decade10s
##                               0.9806318
##                               Decade80s
##                               -1.0007691
##                               Decade90s
##                               2.4160187
## Platform_CompanyNintendo
##                               3.1357537
## Platform_CompanySega
##                               -1.6054687
## Platform_CompanySONY
##                               1.2999294
## Platform_Gen6th_Gen
##                               -0.2556462
## Platform_Gen7th_Gen
##                               -0.3843653
## Platform_Gen8th_Gen
##                               -1.3820419
## Family_PlatformMisc
##                               -3.2828698
## Family_PlatformNintendo_DS
##                               -2.5267276
## Family_PlatformNintendo_Wii
##                               -1.9887849
## Family_PlatformSONY_PS
##                               NA
## Main_DeveloperEA
##                               -0.9035524
## Main_DeveloperMicrosoft
##                               -0.7396379
## Main_DeveloperNamco
##                               -1.3101151
## Main_DeveloperNintendo
##                               -1.0185033
## Main_DeveloperOther
##                               -0.9926555
## Main_DeveloperSega
##                               -1.4990889
## Main_DeveloperSONY
##                               -1.7133570
## Main_DeveloperSquare_Enix
##                               -1.2865758
## Main_DeveloperTencent
##                               0.0715244

```

```

##          Main_DeveloperUbisoft      -0.2543729
##          Main_PublisherEA        0.4770721
##          Main_PublisherMicrosoft  0.7601060
##          Main_PublisherNamco     -0.3810062
##          Main_PublisherNintendo   0.6261959
##          Main_PublisherOther      -0.8004780
##          Main_PublisherSega       -0.0733221
##          Main_PublisherSONY       -0.9826779
##          Main_PublisherSquareEnix 0.3977099
##          Main_PublisherTencent    -0.3463002
##          Main_PublisherUbisoft    -0.0825459
##          RegionJP_Sales         -9.6908164
##          RegionNA_Sales         1.0132516
##          RegionOther_Sales       -1.5454250
##          GenreAdventure:RegionJP_Sales -0.0040066
##          GenreFighting:RegionJP_Sales 1.5688832
##          GenreMisc:RegionJP_Sales   -0.3937440
##          GenrePlatform:RegionJP_Sales -0.5476305
##          GenrePuzzle:RegionJP_Sales  2.4903831
##          GenreRacing:RegionJP_Sales -1.4348196
##          GenreRole-Playing:RegionJP_Sales 3.4734044
##          GenreShooter:RegionJP_Sales  -0.3604365
##          GenreSimulation:RegionJP_Sales -0.0160929
##          GenreSports:RegionJP_Sales   -0.4265643
##          GenreStrategy:RegionJP_Sales -1.1887946
##          GenreAdventure:RegionNA_Sales -0.0646230
##          GenreFighting:RegionNA_Sales  0.9299026

```

```

##          GenreMisc:RegionNA_Sales
##                                1.9495089
##          GenrePlatform:RegionNA_Sales
##                                0.6713842
##          GenrePuzzle:RegionNA_Sales
##                                3.0643608
##          GenreRacing:RegionNA_Sales
##                                -0.3369421
##          GenreRole-Playing:RegionNA_Sales
##                                0.8134207
##          GenreShooter:RegionNA_Sales
##                                -0.0198123
##          GenreSimulation:RegionNA_Sales
##                                -1.1474200
##          GenreSports:RegionNA_Sales
##                                1.5236688
##          GenreStrategy:RegionNA_Sales
##                                -2.1091104
##          GenreAdventure:RegionOther_Sales
##                                -0.2282881
##          GenreFighting:RegionOther_Sales
##                                0.5715755
##          GenreMisc:RegionOther_Sales
##                                1.3611506
##          GenrePlatform:RegionOther_Sales
##                                -0.3306548
##          GenrePuzzle:RegionOther_Sales
##                                2.3450481
##          GenreRacing:RegionOther_Sales
##                                -0.2845912
##          GenreRole-Playing:RegionOther_Sales
##                                0.8888850
##          GenreShooter:RegionOther_Sales
##                                -0.2636937
##          GenreSimulation:RegionOther_Sales
##                                -0.3572546
##          GenreSports:RegionOther_Sales
##                                0.9100572
##          GenreStrategy:RegionOther_Sales
##                                -1.1736023
##          Main_DeveloperEA:RegionJP_Sales
##                                0.7934877
##          Main_DeveloperMicrosoft:RegionJP_Sales
##                                0.3653145
##          Main_DeveloperNamco:RegionJP_Sales
##                                3.0362243
##          Main_DeveloperNintendo:RegionJP_Sales
##                                2.2484896
##          Main_DeveloperOther:RegionJP_Sales
##                                0.9065253
##          Main_DeveloperSega:RegionJP_Sales
##                                2.8876983
##          Main_DeveloperSONY:RegionJP_Sales
##                                2.6054829

```

```

##      Main_DeveloperSquare Enix:RegionJP_Sales
##                                         3.7002101
##      Main_DeveloperTencent:RegionJP_Sales
##                                         -0.5330404
##      Main_DeveloperUbisoft:RegionJP_Sales
##                                         0.4822698
##      Main_DeveloperEA:RegionNA_Sales
##                                         0.5352601
##      Main_DeveloperMicrosoft:RegionNA_Sales
##                                         0.3680635
##      Main_DeveloperNamco:RegionNA_Sales
##                                         0.6520135
##      Main_DeveloperNintendo:RegionNA_Sales
##                                         1.3920305
##      Main_DeveloperOther:RegionNA_Sales
##                                         0.5366285
##      Main_DeveloperSega:RegionNA_Sales
##                                         1.3690892
##      Main_DeveloperSONY:RegionNA_Sales
##                                         1.0301884
##      Main_DeveloperSquare Enix:RegionNA_Sales
##                                         1.7200756
##      Main_DeveloperTencent:RegionNA_Sales
##                                         -1.7972460
##      Main_DeveloperUbisoft:RegionNA_Sales
##                                         0.0450394
##      Main_DeveloperEA:RegionOther_Sales
##                                         1.2629823
##      Main_DeveloperMicrosoft:RegionOther_Sales
##                                         0.9457722
##      Main_DeveloperNamco:RegionOther_Sales
##                                         0.8027185
##      Main_DeveloperNintendo:RegionOther_Sales
##                                         1.5520409
##      Main_DeveloperOther:RegionOther_Sales
##                                         0.5001005
##      Main_DeveloperSega:RegionOther_Sales
##                                         1.7265898
##      Main_DeveloperSONY:RegionOther_Sales
##                                         1.7743170
##      Main_DeveloperSquare Enix:RegionOther_Sales
##                                         0.4189766
##      Main_DeveloperTencent:RegionOther_Sales
##                                         -0.0105046
##      Main_DeveloperUbisoft:RegionOther_Sales
##                                         0.6355513
##      Main_PublisherEA:RegionJP_Sales
##                                         0.4563650
##      Main_PublisherMicrosoft:RegionJP_Sales
##                                         1.1644328
##      Main_PublisherNamco:RegionJP_Sales
##                                         2.8510913
##      Main_PublisherNintendo:RegionJP_Sales
##                                         7.1597642

```

```

##          Main_PublisherOther:RegionJP_Sales
##                                2.3361162
##          Main_PublisherSega:RegionJP_Sales
##                                1.2982281
##          Main_PublisherSONY:RegionJP_Sales
##                                2.0416849
##          Main_PublisherSquare Enix:RegionJP_Sales
##                                3.5537343
##          Main_PublisherTencent:RegionJP_Sales
##                                1.7689752
##          Main_PublisherUbisoft:RegionJP_Sales
##                                0.7536662
##          Main_PublisherEA:RegionNA_Sales
##                                -0.7983365
##          Main_PublisherMicrosoft:RegionNA_Sales
##                                -1.7610963
##          Main_PublisherNamco:RegionNA_Sales
##                                -0.8647817
##          Main_PublisherNintendo:RegionNA_Sales
##                                -1.3977739
##          Main_PublisherOther:RegionNA_Sales
##                                -0.2712843
##          Main_PublisherSega:RegionNA_Sales
##                                -1.0136068
##          Main_PublisherSONY:RegionNA_Sales
##                                -0.9529198
##          Main_PublisherSquare Enix:RegionNA_Sales
##                                -1.9423174
##          Main_PublisherTencent:RegionNA_Sales
##                                -3.9632654
##          Main_PublisherUbisoft:RegionNA_Sales
##                                -0.8084949
##          Main_PublisherEA:RegionOther_Sales
##                                -0.6635794
##          Main_PublisherMicrosoft:RegionOther_Sales
##                                -2.0989122
##          Main_PublisherNamco:RegionOther_Sales
##                                -1.0602315
##          Main_PublisherNintendo:RegionOther_Sales
##                                -1.1128013
##          Main_PublisherOther:RegionOther_Sales
##                                -0.4194465
##          Main_PublisherSega:RegionOther_Sales
##                                -0.5394534
##          Main_PublisherSONY:RegionOther_Sales
##                                0.0806539
##          Main_PublisherSquare Enix:RegionOther_Sales
##                                -0.5063326
##          Main_PublisherTencent:RegionOther_Sales
##                                -1.7836246
##          Main_PublisherUbisoft:RegionOther_Sales
##                                -0.7327943
##          GenreAdventure:Decade10s
##                                -0.6778059

```

```

##          GenreFighting:Decade10s
##                                -0.2520382
##          GenreMisc:Decade10s
##                                -0.3000875
##          GenrePlatform:Decade10s
##                                -0.8263652
##          GenrePuzzle:Decade10s
##                                -1.3468849
##          GenreRacing:Decade10s
##                                -0.4808509
##          GenreRole-Playing:Decade10s
##                                -0.1217285
##          GenreShooter:Decade10s
##                                1.0016639
##          GenreSimulation:Decade10s
##                                -0.0005734
##          GenreSports:Decade10s
##                                0.2596887
##          GenreStrategy:Decade10s
##                                0.4878773
##          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
##                                0.3469493
##          GenreFighting:Decade90s
##                                -0.2329749
##          GenreMisc:Decade90s
##                                1.8063927
##          GenrePlatform:Decade90s
##                                -1.4738620
##          GenrePuzzle:Decade90s
##                                -0.4332273
##          GenreRacing:Decade90s
##                                -1.0525427

```

```

##          GenreRole-Playing:Decade90s
##                               -0.6315650
##          GenreShooter:Decade90s
##                               -2.6445377
##          GenreSimulation:Decade90s
##                               -1.6511042
##          GenreSports:Decade90s
##                               -1.0051856
##          GenreStrategy:Decade90s
##                               -1.9072300

```

KEPP this model with AIC=62858.87. AIC even decreased from the previous case. At this step, we answered our **first problem's question:** which properties of games are significant in prediction the global sales? First, we did include our subjective view, then eliminated variables by EDA (multicollinearity, dependence) and found all other variables significant. **Subsequent question (part of first):** do the sales differ regionally? According to the ANOVA, yes they differ.

We can further analyze which regions differ by using pairwise comparison tests on the one-way ANOVA:

```

regional_diff_sales <- lm(log(Sales) ~ Region, data = train_sales)
TukeyHSD(aov(regional_diff_sales))

```

```

##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = regional_diff_sales)
##
## $Region
##            diff      lwr      upr p adj
## JP_Sales-EU_Sales -6.515768 -6.717341 -6.3141945 0
## NA_Sales-EU_Sales  1.331520  1.129947  1.5330932 0
## Other_Sales-EU_Sales -1.181040 -1.382613 -0.9794668 0
## NA_Sales-JP_Sales   7.847288  7.645714  8.0488610 0
## Other_Sales-JP_Sales  5.334728  5.133155  5.5363010 0
## Other_Sales-NA_Sales -2.512560 -2.714133 -2.3109867 0

```

All differences are significant. For instance, there are significantly less number of video games sold between 1985 and 2016 in Japan compared to Europe, North America and other part of the world (which is actually intuitive. I'd really wonder if it were the opposite). There are more video games sold in North America than in the other part of the world. Finally, second most selling region is Europe.

3. Prediction using the model we chose

Least squares coefficient estimates are found in the final: —— model. So the prediction (fitted) equation is given as follows:

$$kadjd$$

write it here/or not if very bigggggg

- We know the *RMSE* on the training data: 3.661722

Assess accuracy on the test data by finding \sqrt{MSE} on the test data:

```

predicted_sales <- predict(mlr_with, test_sales[,-c(1,2,4,7,8,9,11,17,20)])

```

```

## Warning in predict.lm(mlr_with, test_sales[, -c(1, 2, 4, 7, 8, 9, 11, 17, :
## 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] 6.650445
```

- RMSE on the test data = 6.650445

Conclusions

1. By EDA, ANOVA and MLR model fit we derived the *best* in our case, model according to AIC and BIC criteria.
2. Model accuracy on the test set:
3. Model assumptions violations – new challenges
4. Old challenge: non-generalizable model (look at the years ~)
5. Room for improvement: