

Data 467 Project Code

Zach Kramer

2022-10-30

Data Cleanup

```
# Reads in data
main <- read.csv("./anime_table.csv")
dubbed <- read.csv("./anime_dubbed.csv")
ranking <- read.csv("./anime_ranking_table.csv")

# Cleans some of the unused large columns
main <- subset(main, select = -c(main_picture.medium, main_picture.large, alternative_titles.ja, altern

# Selects data where tm_ky == 11, the most recent data capture from the dataset
main <- main[main$tm_ky==11, ]
ranking <- ranking[ranking$tm_ky==11, ]
```

Data Editing

```
# Merges all tables together, 2nd is a Left Join
full <- merge(main, ranking, by = "mal_id")
full <- merge(full, dubbed, by = "mal_id", all.x = TRUE)
```

Turning Data Numeric

```
# Rating into numeric values
full$rating <- as.factor(full$rating)
full$rating_num <- unclass(full$rating)

# Turns dub status into 0 or 1, based on if it is dubbed
full$dub_status_de[full$dub_status_de == "dubbed, partially dubbed"] <- "dubbed"
full$dub_status_de <- as.factor(full$dub_status_de)
full$dub_status_num <- unclass(full$dub_status_de)
full$dub_status_num[is.na(full$dub_status_num)] <- 0
full$dub_status_num[full$dub_status_num==2] <- 0

# Status into numeric values
full$status <- as.factor(full$status)
full$status_num <- unclass(full$status)
```

```

# Start Season into numeric values
full$start_season.season <- as.factor(full$start_season.season)
full$start_season.season_num <- unclass(full$start_season.season)

# Studio into numeric values
full$studios_de <- as.factor(full$studios_de)
full$studios_num <- unclass(full$studios_de)

final <- subset(full, popularity > 7000)
final <- na.omit(final)

```

Linear Modeling

```

lmod <- lm(popularity~mean + rating + start_season.season + num_episodes
, data = final)
#removed variables "status" and "studio" because they were not significant to the model and to improve
summary(lmod)

```

```

##
## Call:
## lm(formula = popularity ~ mean + rating + start_season.season +
##     num_episodes, data = final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5656.5 -1355.2   113.8  1420.6  4583.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    18981.3390    605.0350   31.372 < 2e-16 ***
## mean          -1276.5695     93.3062  -13.682 < 2e-16 ***
## ratingmissing    -95.7711    376.3367   -0.254  0.799169
## ratingpg        -541.2683    145.1651   -3.729  0.000202 ***
## ratingpg_13     -1174.7419    125.4675   -9.363 < 2e-16 ***
## ratingr        -1676.2909    456.6873   -3.671  0.000253 ***
## ratingr+       -1598.6279    311.8888   -5.126  3.48e-07 ***
## start_season.seasonspring    50.1002    130.1698    0.385  0.700396
## start_season.seasonsummer   -59.4963    177.1983   -0.336  0.737113
## start_season.seasonwinter  -368.9930    153.8260   -2.399  0.016611 *
## num_episodes         1.1136     0.3568    3.121  0.001846 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1792 on 1135 degrees of freedom
## Multiple R-squared:  0.2112, Adjusted R-squared:  0.2042
## F-statistic: 30.38 on 10 and 1135 DF, p-value: < 2.2e-16

```

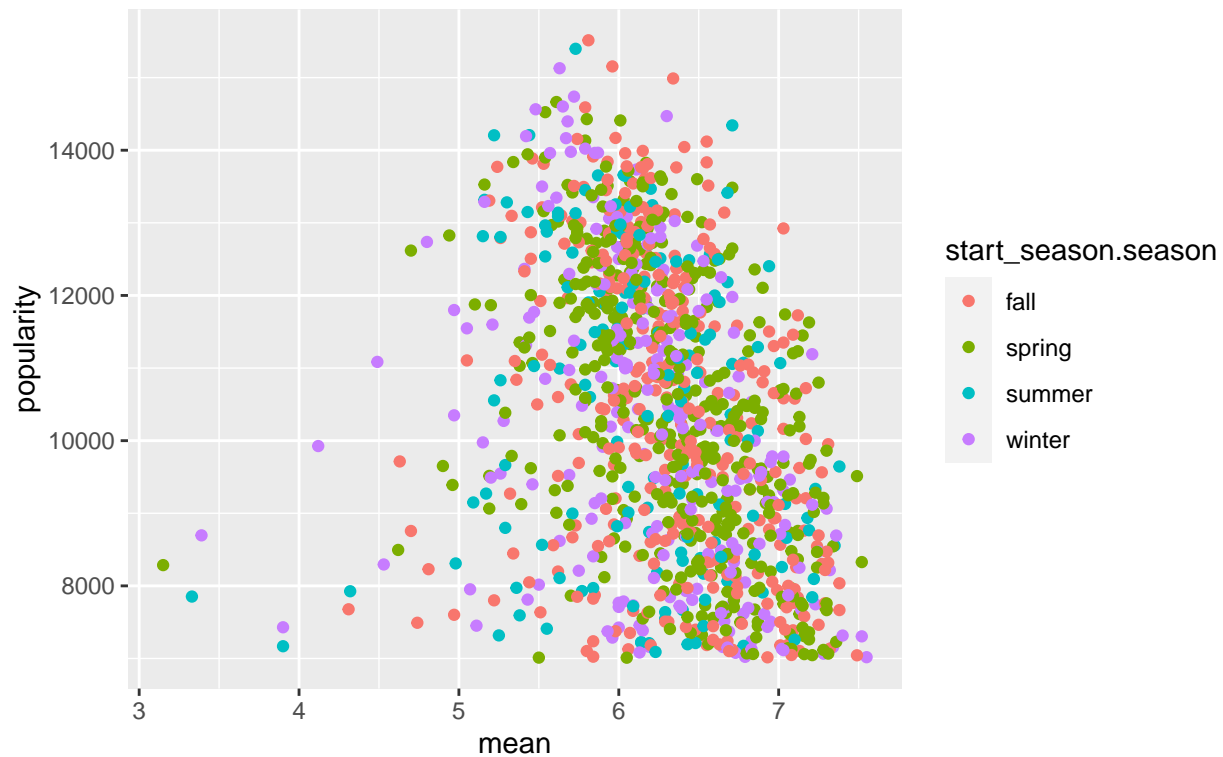
Results

```
library(ggplot2)
ggplot(aes(x=mean,y=popularity,color=rating),data=final) + geom_point() + facet_wrap(~rating) + ggtitle('Scatterplot of Anime Popularity vs. Mean')
```



```
ggplot(aes(x=mean,y=popularity,color=start_season.season),data=final) + geom_point() + ggtitle('Scatterplot of Anime Popularity vs. Mean')
```

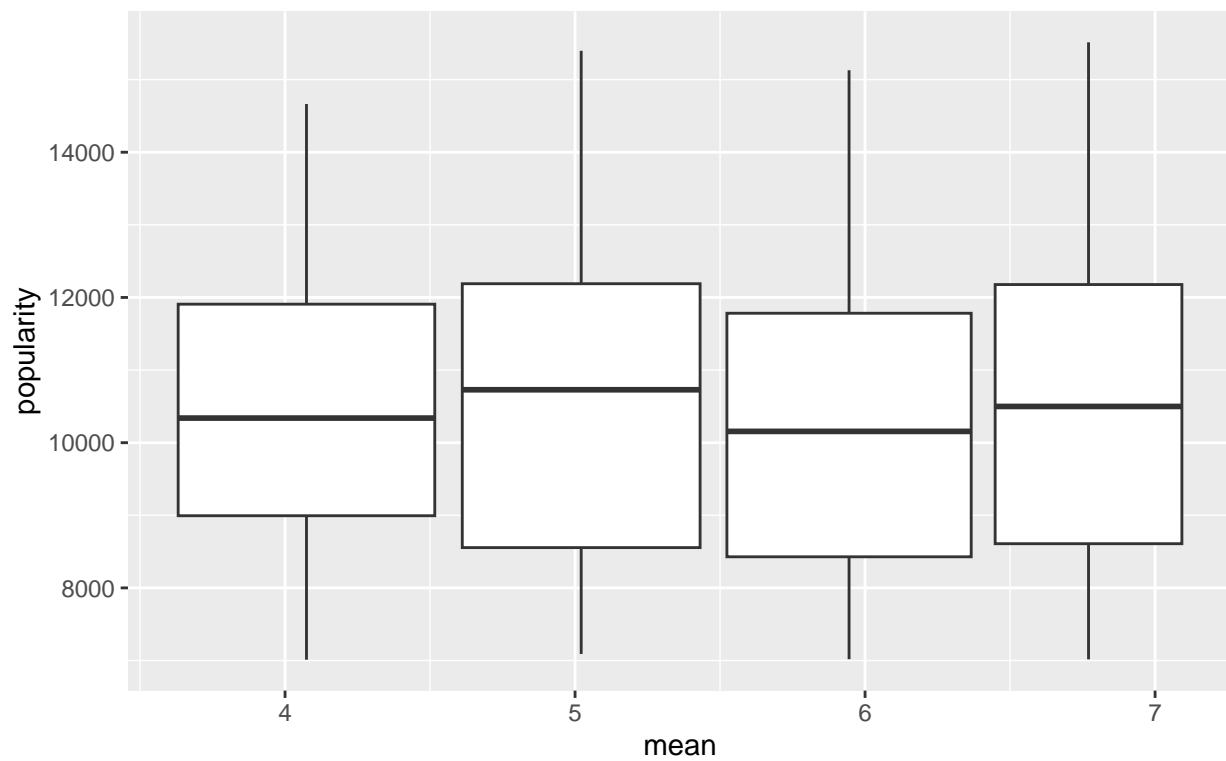
Scatterplot of Anime Popularity vs. Mean classed by Season
the First Season Was Released



```
ggplot(data=final)+geom_boxplot(aes(x=mean,y=popularity,group=start_season.season,color=rating))+ggtitle
```

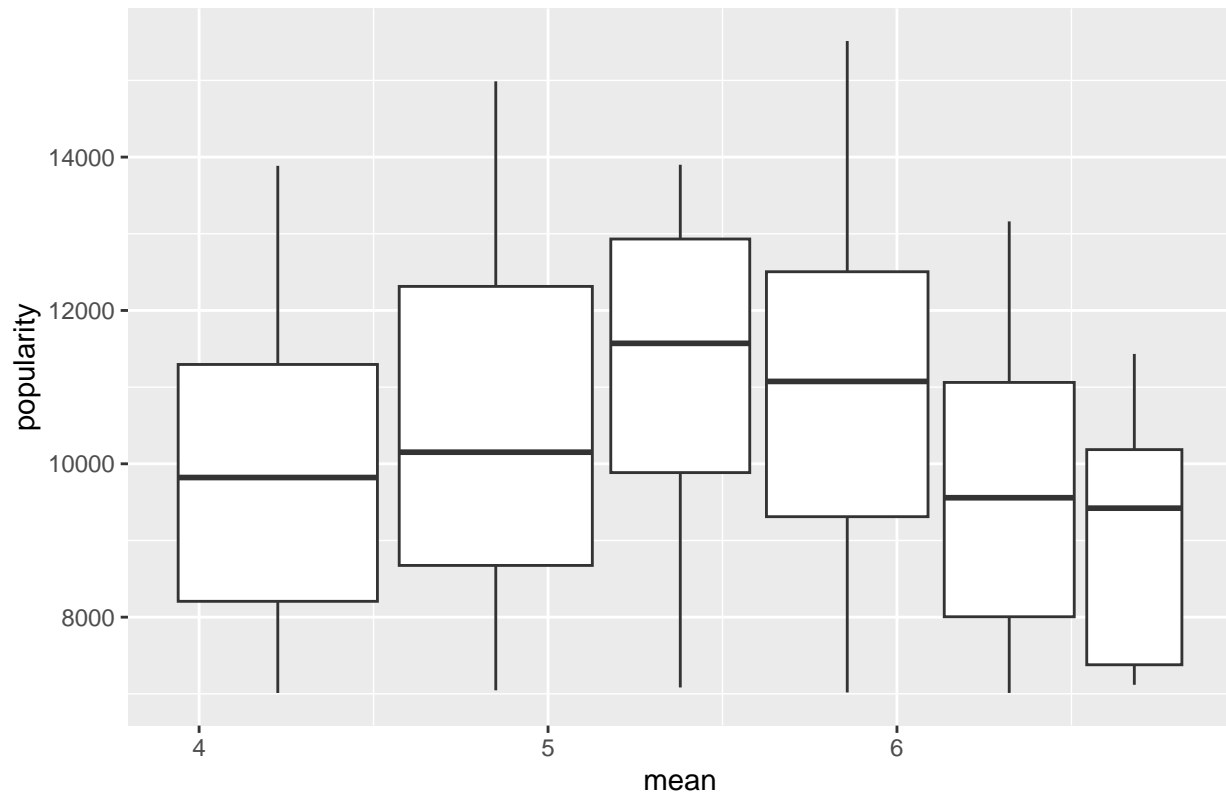
```
## Warning: The following aesthetics were dropped during statistical transformation: colour
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a 'group' aesthetic or to convert a numerical
##   variable into a factor?
```

Boxplot of Anime Popularity vs. Mean Score grouped by the Season the First Season Was Released



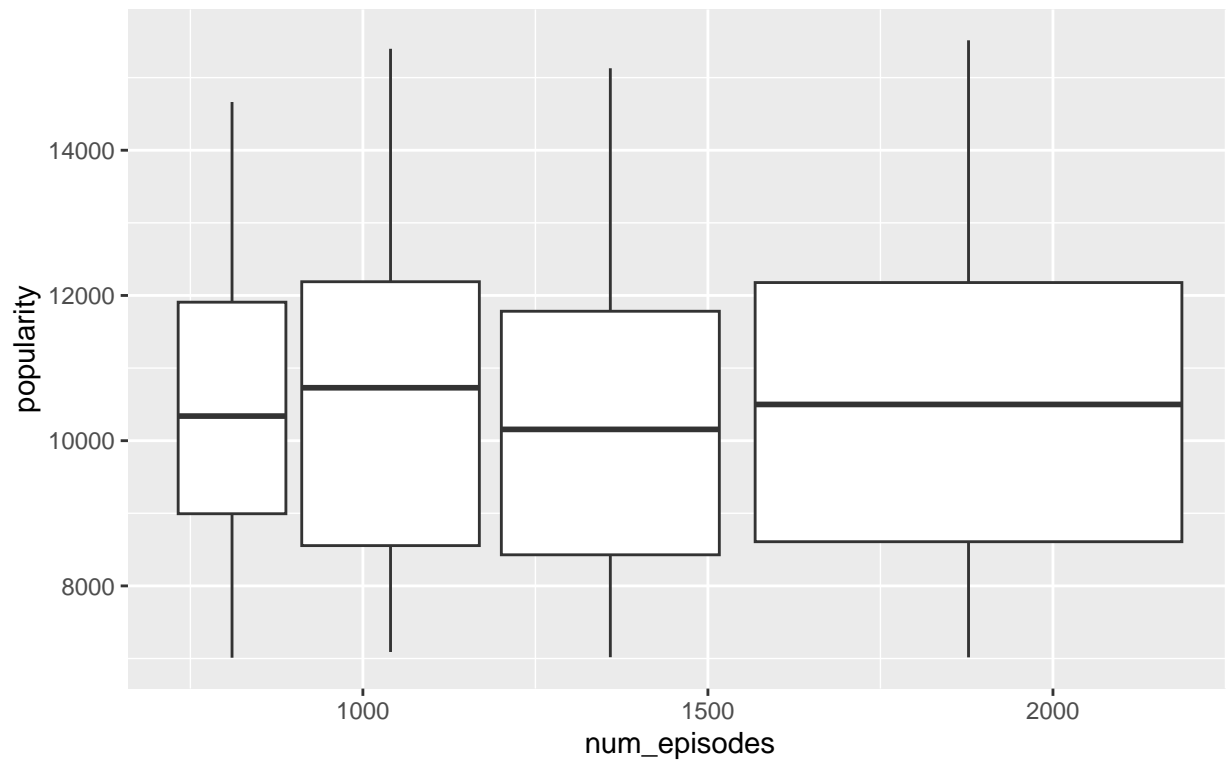
```
ggplot(data=final)+geom_boxplot(aes(x=mean,y=popularity,group=rating))+ggtitle('Boxplot of Anime Popularity vs. Mean Score grouped by the Season the First Season Was Released')
```

Boxplot of Anime Popularity vs. Mean Score grouped by MPAA Rating



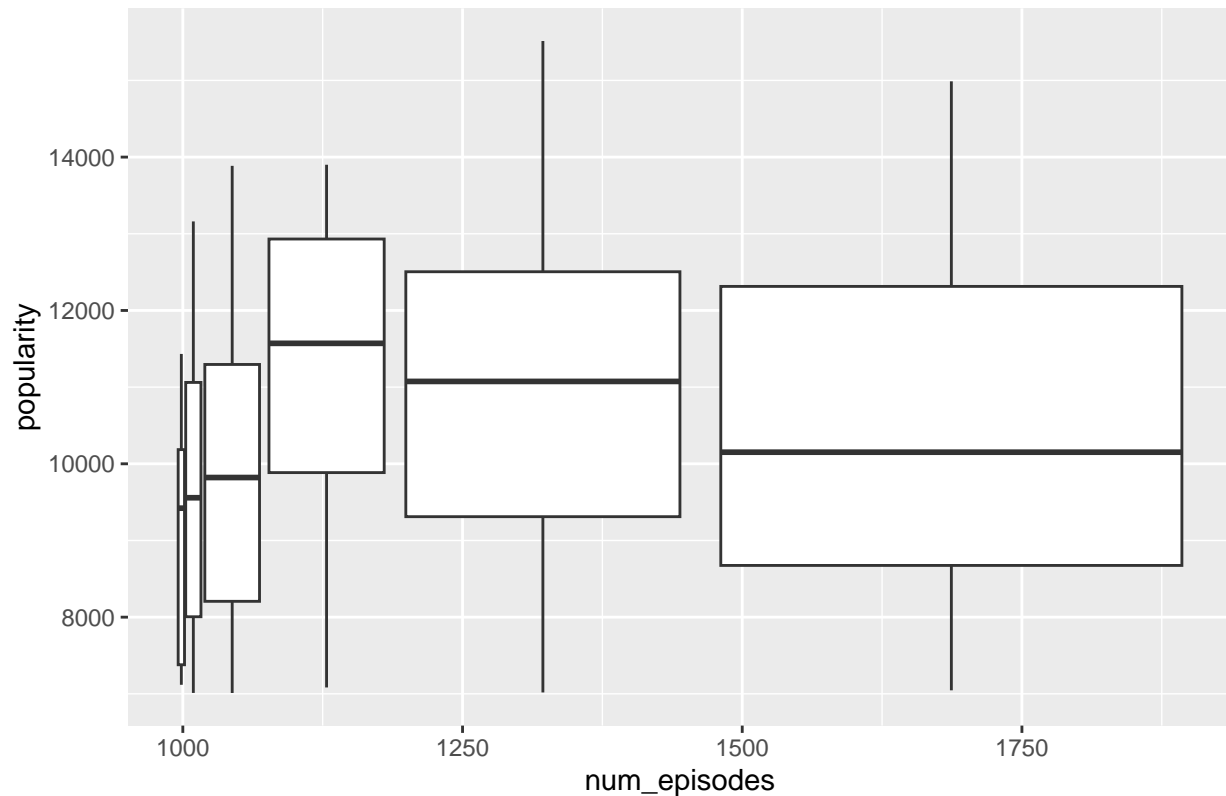
```
ggplot(data=final)+geom_boxplot(aes(x=num_episodes,y=popularity,group=start_season.season))+ggtitle('Boxplot of Anime Popularity vs. Mean Score grouped by MPAA Rating')
```

Boxplot of Anime Popularity vs. Number of Episodes grouped by the Season the First Season Was Released



```
ggplot(data=final)+geom_boxplot(aes(x=num_episodes,y=popularity,group=rating))+ggtitle('Boxplot of Anime
```

Boxplot of Anime Popularity vs. Number of Episodes grouped by MPAA R



Linear Models

```
#written model and coefficients

#linear model with coefficients
lmod_interaction <- lm(popularity~(mean + rating + start_season.season + num_episodes)^2,
                       data = final)
summary(lmod_interaction)
```

```
##
## Call:
## lm(formula = popularity ~ (mean + rating + start_season.season +
##   num_episodes)^2, data = final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5492.0 -1335.4    17.5  1323.5  4291.2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.466e+04  1.447e+03  17.041  < 2e-16
## mean        -2.158e+03  2.278e+02  -9.470  < 2e-16
## ratingmissing -1.160e+04  4.291e+03  -2.704  0.006948
## ratingpg     -5.881e+03  1.633e+03  -3.602  0.000329
```


## ratingpg_13	-9.665e+03	1.402e+03	-6.893	9.15e-12
## ratingr	-3.422e+03	6.693e+03	-0.511	0.609180
## ratingr+	-1.402e+04	3.650e+03	-3.840	0.000130
## start_season.seasonspring	4.761e+02	1.534e+03	0.310	0.756425
## start_season.seasonsummer	-4.149e+03	1.858e+03	-2.233	0.025722
## start_season.seasonwinter	-1.425e+03	1.667e+03	-0.855	0.392819
## num_episodes	-1.616e+00	8.703e+00	-0.186	0.852702
## mean:ratingmissing	1.867e+03	6.734e+02	2.772	0.005662
## mean:ratingpg	8.777e+02	2.555e+02	3.435	0.000615
## mean:ratingpg_13	1.241e+03	2.196e+02	5.651	2.03e-08
## mean:ratingr	-1.723e+02	1.065e+03	-0.162	0.871501
## mean:ratingr+	1.957e+03	5.977e+02	3.275	0.001090
## mean:start_season.seasonspring	-7.711e+01	2.396e+02	-0.322	0.747610
## mean:start_season.seasonsummer	6.340e+02	2.960e+02	2.142	0.032398
## mean:start_season.seasonwinter	1.423e+02	2.595e+02	0.548	0.583609
## mean:num_episodes	3.485e-01	1.511e+00	0.231	0.817597
## ratingmissing:start_season.seasonspring	9.480e+01	9.167e+02	0.103	0.917656
## ratingpg:start_season.seasonspring	-4.833e+02	3.427e+02	-1.410	0.158712
## ratingpg_13:start_season.seasonspring	4.348e+02	3.028e+02	1.436	0.151261
## ratingr:start_season.seasonspring	1.484e+03	1.204e+03	1.233	0.217771
## ratingr+:start_season.seasonspring	1.214e+03	7.946e+02	1.528	0.126682
## ratingmissing:start_season.seasonsummer	-1.092e+03	1.051e+03	-1.038	0.299266
## ratingpg:start_season.seasonsummer	-6.043e+01	4.939e+02	-0.122	0.902644
## ratingpg_13:start_season.seasonsummer	4.569e+02	4.136e+02	1.105	0.269569
## ratingr:start_season.seasonsummer	9.493e+02	1.338e+03	0.709	0.478226
## ratingr+:start_season.seasonsummer	1.088e+03	1.068e+03	1.019	0.308373
## ratingmissing:start_season.seasonwinter	-1.560e+03	1.380e+03	-1.131	0.258272
## ratingpg:start_season.seasonwinter	-7.278e+01	4.193e+02	-0.174	0.862219
## ratingpg_13:start_season.seasonwinter	5.990e+02	3.560e+02	1.683	0.092730
## ratingr:start_season.seasonwinter	1.174e+03	1.406e+03	0.835	0.404141
## ratingr+:start_season.seasonwinter	3.855e+02	9.196e+02	0.419	0.675172
## ratingmissing:num_episodes	4.386e+00	3.129e+00	1.402	0.161279
## ratingpg:num_episodes	1.455e-01	7.768e-01	0.187	0.851422
## ratingpg_13:num_episodes	1.023e+01	3.173e+00	3.225	0.001296
## ratingr:num_episodes	7.855e+01	4.134e+01	1.900	0.057649
## ratingr+:num_episodes	-7.283e+00	1.552e+01	-0.469	0.639080
## start_season.seasonspring:num_episodes	-7.709e-01	1.694e+00	-0.455	0.649141
## start_season.seasonsummer:num_episodes	1.346e+00	1.502e+00	0.896	0.370301
## start_season.seasonwinter:num_episodes	8.621e-01	8.404e-01	1.026	0.305173
##				
## (Intercept)	***			
## mean	***			
## ratingmissing	**			
## ratingpg	***			
## ratingpg_13	***			
## ratingr				
## ratingr+	***			
## start_season.seasonspring				
## start_season.seasonsummer	*			
## start_season.seasonwinter				
## num_episodes				
## mean:ratingmissing	**			
## mean:ratingpg	***			
## mean:ratingpg_13	***			

```

## mean:ratingr
## mean:ratingr+                **
## mean:start_season.seasonspring
## mean:start_season.seasonsummer      *
## mean:start_season.seasonwinter
## mean:num_episodes
## ratingmissing:start_season.seasonspring
## ratingpg:start_season.seasonspring
## ratingpg_13:start_season.seasonspring
## ratingr:start_season.seasonspring
## ratingr+:start_season.seasonspring
## ratingmissing:start_season.seasonsummer
## ratingpg:start_season.seasonsummer
## ratingpg_13:start_season.seasonsummer
## ratingr:start_season.seasonsummer
## ratingr+:start_season.seasonsummer
## ratingmissing:start_season.seasonwinter
## ratingpg:start_season.seasonwinter
## ratingpg_13:start_season.seasonwinter .
## ratingr:start_season.seasonwinter
## ratingr+:start_season.seasonwinter
## ratingmissing:num_episodes
## ratingpg:num_episodes
## ratingpg_13:num_episodes          **
## ratingr:num_episodes              .
## ratingr+:num_episodes
## start_season.seasonspring:num_episodes
## start_season.seasonsummer:num_episodes
## start_season.seasonwinter:num_episodes
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1743 on 1103 degrees of freedom
## Multiple R-squared:  0.2752, Adjusted R-squared:  0.2476
## F-statistic:  9.97 on 42 and 1103 DF, p-value: < 2.2e-16

```

```
#residual values
```

```
lmod_r <- resid(lmod)
```

```
#fitted values
```

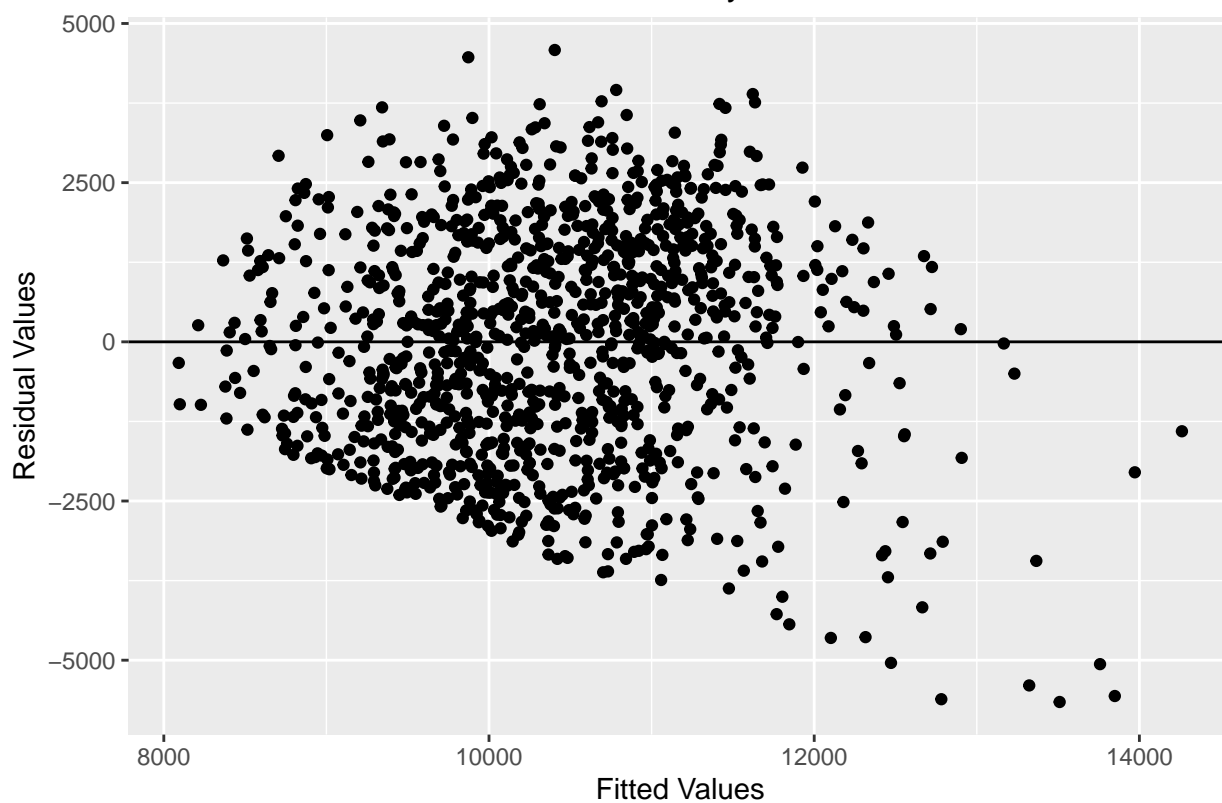
```
lmod_f <- fitted(lmod)
```

```
#residual vs. fitted plot -
```

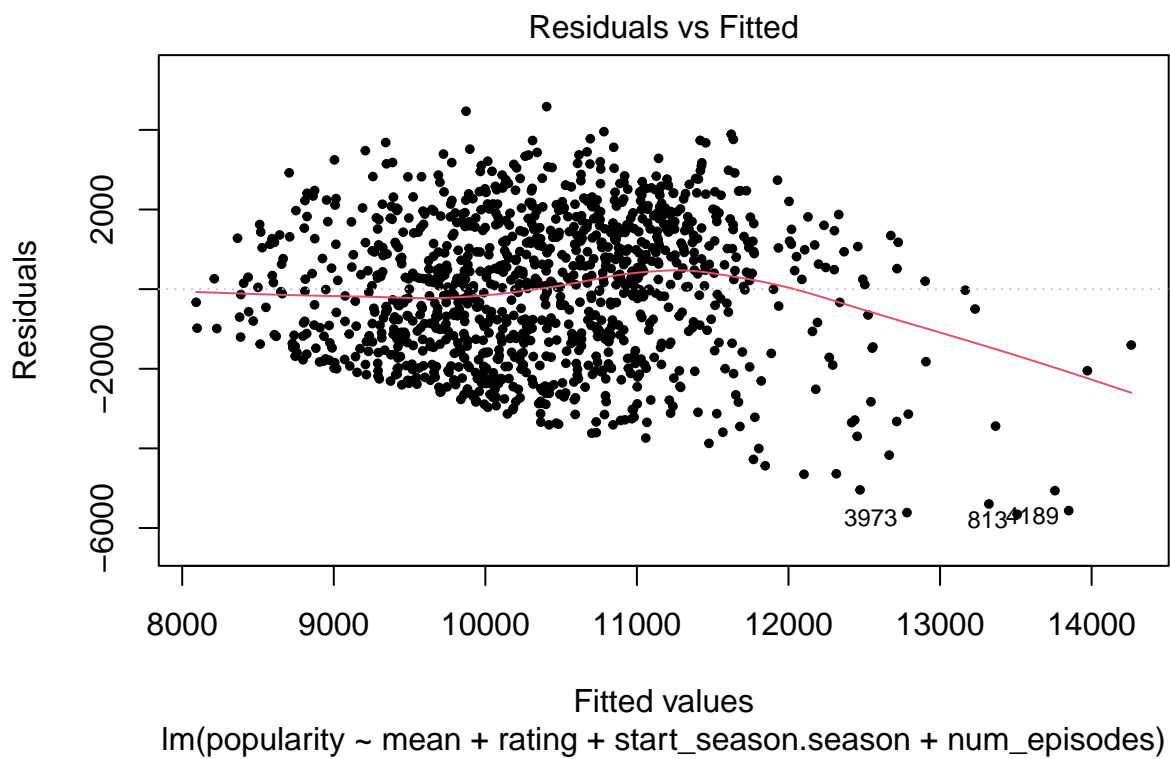
```
library(ggplot2)
```

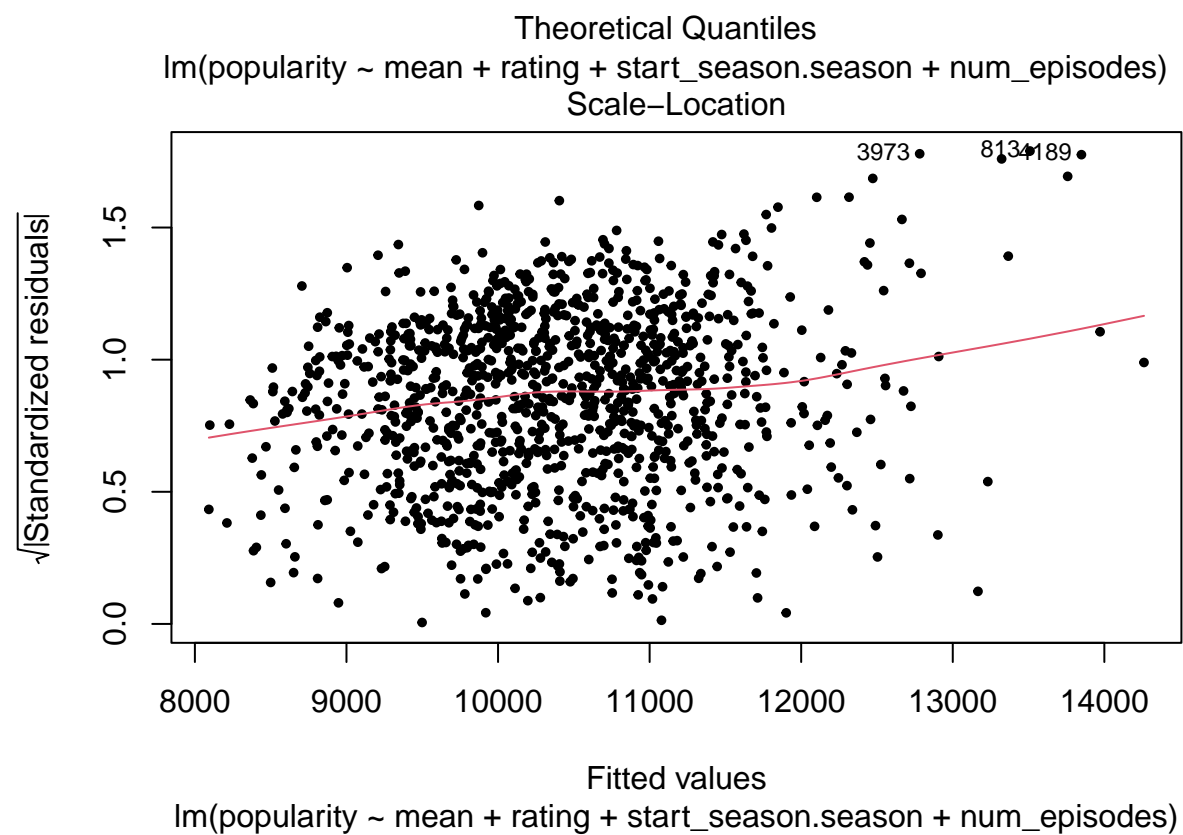
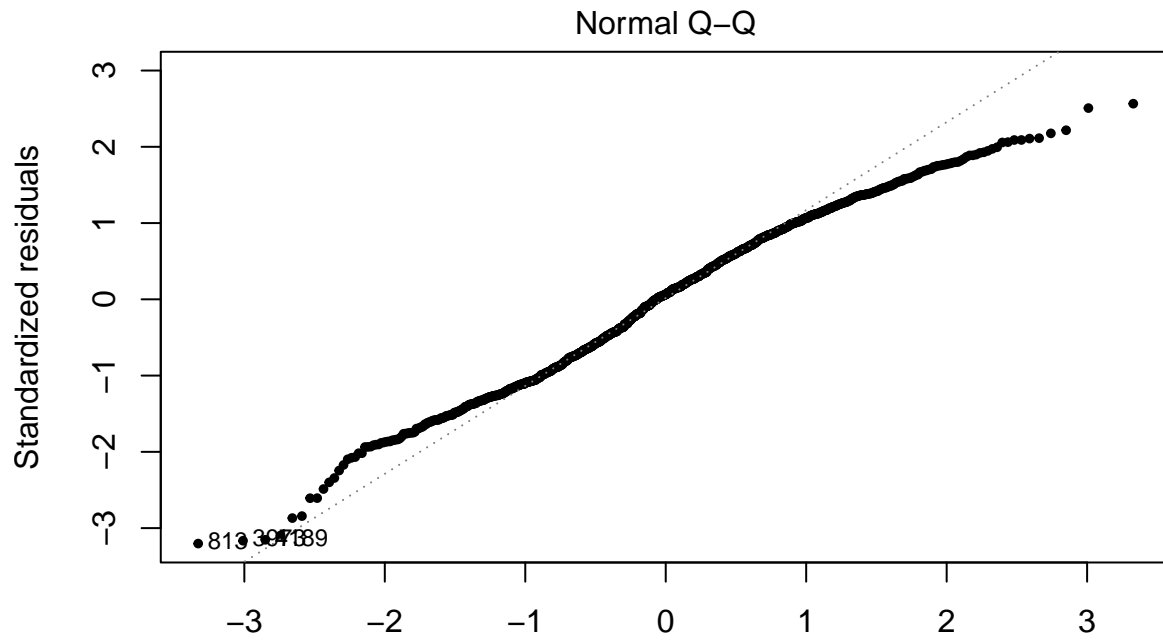
```
ggplot()+geom_point(aes(x=lmod_f,y=lmod_r))+geom_hline(yintercept=0)+xlab('Fitted Values')+ylab('Residuals')
```

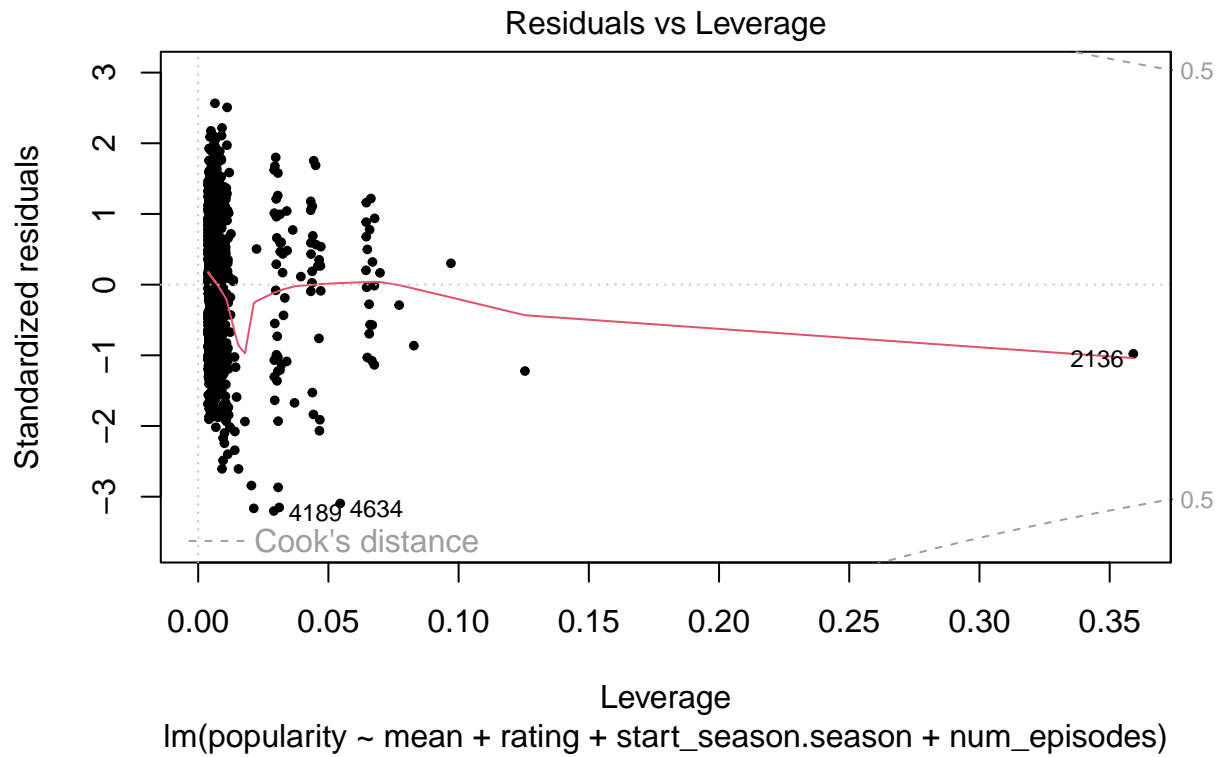
Residual Plot of the Fitted Values for My Anime List Data



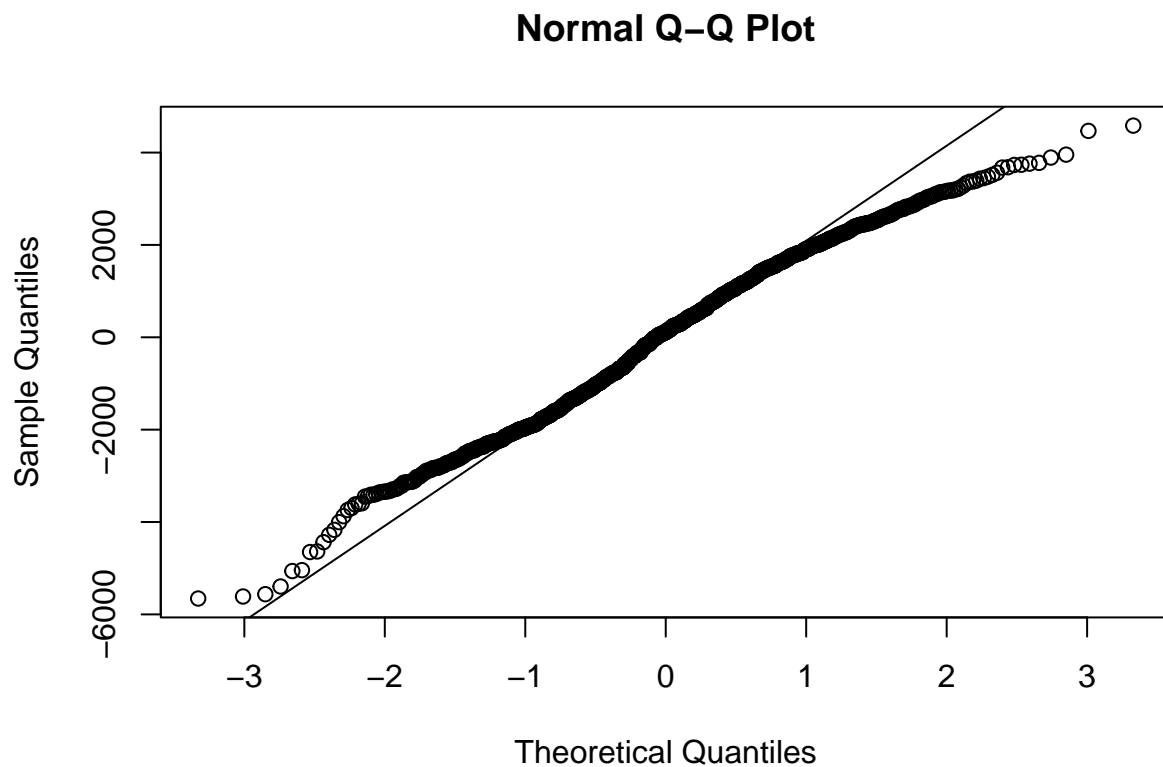
```
plot(lmod,cex=.8,pch=20)
```







```
#qq plot
qqnorm(lmod_r)
qqline(lmod_r)
```



```
#parallel slope
#need to include hypotheses and conclusion
anova(lmod,lmod_interaction)
```

```
## Analysis of Variance Table
##
## Model 1: popularity ~ mean + rating + start_season.season + num_episodes
## Model 2: popularity ~ (mean + rating + start_season.season + num_episodes)^2
##   Res.Df      RSS Df Sum of Sq    F    Pr(>F)
## 1    1135 3646429951
## 2    1103 3350604246 32 295825704 3.0433 4.24e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#will need to use the full model with significant interaction terms because the slopes are not parallel

Assumptions Based on the diagnostics completed, one out of the four assumptions have been met completely. The first assumption that is fulfilled is normality of the errors. Based on the QQ plot, our data points seem to follow a normal distribution for the most part. There is a cause for slight concern when looking at the tails; they deviate slightly, but the center of the graph precisely follows the QQ line. Based on the diagnostics, there is cause for concern when it relates to the constant variance and parallel slopes. When looking at the residual plot of the fitted values, there seems to be a degree of heteroscedacity in the error. Even with cleaning the data, the fitted vs. residual values continued to follow the same pattern. This causes limits to our analysis that will discussed later. The other assumption that is not fulfilled is parallel slopes between the full and reduced model. After completing the ANOVA test, we were able to reject the null hypothesis (p-value: $<2.2e-16 < 0.05$); therefore, we were able to conclude that the model containing the interaction terms does not have parallel slopes with the model without interaction terms. Thus, we are inclined to use a revised model that contains the interaction terms that are significant to the model.

Hypothesis Testing

```
lmod_revised <- lm(popularity~mean + rating + start_season.season + num_episodes + mean*num_episodes + 
summary(lmod_revised) #will need to explain why interaction terms that were selected were selected
```

```
##
## Call:
## lm(formula = popularity ~ mean + rating + start_season.season + 
##     num_episodes + mean * num_episodes + rating * num_episodes + 
##     mean * rating, data = final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5661.7 -1302.8    96.1  1340.4  4586.6
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.382e+04  1.036e+03  22.998 < 2e-16 ***
## mean         -2.038e+03  1.647e+02 -12.375 < 2e-16 ***
## ratingmissing -1.533e+04  3.869e+03  -3.963 7.87e-05 ***
## ratingpg      -5.607e+03  1.612e+03  -3.477 0.000526 ***
## ratingpg_13   -9.729e+03  1.339e+03  -7.264 7.00e-13 ***
## ratingr       -4.514e+03  5.892e+03  -0.766 0.443723
## ratingr+      -1.284e+04  3.330e+03  -3.855 0.000122 ***
```

```
## start_season.seasonspring 3.272e+01 1.274e+02 0.257 0.797420
## start_season.seasonsummer -2.703e+01 1.738e+02 -0.156 0.876400
## start_season.seasonwinter -3.044e+02 1.504e+02 -2.023 0.043268 *
## num_episodes 4.215e+00 8.031e+00 0.525 0.599774
## mean:num_episodes -5.962e-01 1.391e+00 -0.429 0.668184
## ratingmissing:num_episodes 4.448e+00 2.502e+00 1.778 0.075653 .
## ratingpg:num_episodes 5.312e-02 7.291e-01 0.073 0.941940
## ratingpg_13:num_episodes 1.011e+01 3.086e+00 3.277 0.001083 **
## ratingr:num_episodes 7.880e+01 3.966e+01 1.987 0.047194 *
## ratingr+:num_episodes -1.358e+01 1.477e+01 -0.919 0.358107
## mean:ratingmissing 2.408e+03 6.196e+02 3.886 0.000108 ***
## mean:ratingpg 8.021e+02 2.536e+02 3.163 0.001603 **
## mean:ratingpg_13 1.307e+03 2.138e+02 6.113 1.35e-09 ***
## mean:ratingr 1.322e+02 9.624e+02 0.137 0.890729
## mean:ratingr+ 1.879e+03 5.697e+02 3.299 0.001001 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1746 on 1124 degrees of freedom
## Multiple R-squared: 0.2591, Adjusted R-squared: 0.2453
## F-statistic: 18.72 on 21 and 1124 DF, p-value: < 2.2e-16
```

#mechanics

```
test1 <- lm(popularity~mean + start_season.season + num_episodes + mean*num_episodes,
            data = final) #removed "rating"
anova(lmod_revised, test1)
```

Analysis of Variance Table

```
##
## Model 1: popularity ~ mean + rating + start_season.season + num_episodes +
##          mean * num_episodes + rating * num_episodes + mean * rating
## Model 2: popularity ~ mean + start_season.season + num_episodes + mean *
##          num_episodes
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1    1124 3424659779
## 2    1139 3989222299 -15 -564562521 12.353 < 2.2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
test2 <- lm(popularity~mean + rating + start_season.season + mean*rating,
            data = final) #removed "num_episodes"
anova(lmod_revised, test2)
```

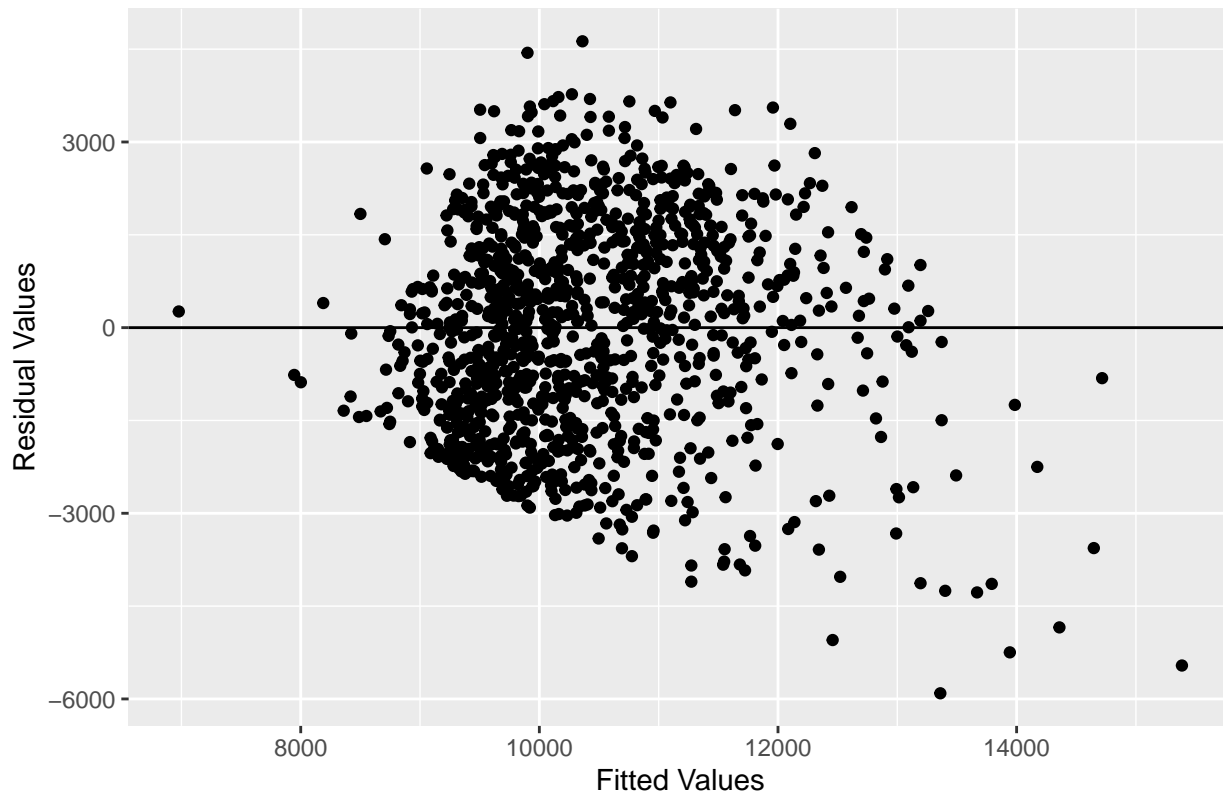
Analysis of Variance Table

```
##
## Model 1: popularity ~ mean + rating + start_season.season + num_episodes +
##          mean * num_episodes + rating * num_episodes + mean * rating
## Model 2: popularity ~ mean + rating + start_season.season + mean * rating
##   Res.Df      RSS Df Sum of Sq      F    Pr(>F)
## 1    1124 3424659779
## 2    1131 3507328304 -7 -82668526 3.8761 0.0003518 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Smaller Model

```
#removing start season because it is the only predictor that is not apart of the interaction terms; the
small_lmod <- lm(popularity~mean + rating + num_episodes + mean*num_episodes + rating*num_episodes + me
small_res <- resid(small_lmod)
small_fit <- fitted(small_lmod)
ggplot()+geom_point(aes(x=small_fit,y=small_res))+geom_hline(yintercept=0)+xlab('Fitted Values')+ylab('')
```

Residual Plot of the Fitted Values for Anime Data



```
#qqnorm(small_res) + qqline(small_res)
```

F-Test

```
#mechanics
anova(lmod_revised, small_lmod)
```

```
## Analysis of Variance Table
##
## Model 1: popularity ~ mean + rating + start_season.season + num_episodes +
##      mean * num_episodes + rating * num_episodes + mean * rating
## Model 2: popularity ~ mean + rating + num_episodes + mean * num_episodes +
##      rating * num_episodes + mean * rating
##   Res.Df      RSS Df Sum of Sq    F Pr(>F)
## 1    1124 3424659779
## 2    1127 3442824595  -3  -18164816 1.9873 0.1141
```

Confidence Interval


```
#(tci<-TukeyHSD(popularity~mean+rating+start_season.season+num_episodes, data = final))
#plot(tci)
head(predict(lmod_revised,level=.99,interval='prediction'))
```

```
## Warning in predict.lm(lmod_revised, level = 0.99, interval = "prediction"): predictions on current d
```

```
##          fit      lwr      upr
## 13 10332.092 5812.580 14851.60
## 122 10417.451 5889.326 14945.58
## 159 10325.406 5810.507 14840.31
## 196  9326.976 4803.993 13849.96
## 218 10569.108 6042.797 15095.42
## 220  9712.050 5188.091 14236.01
```

General Linear Model

```
#using a mixed effects model because the response variable does not meet the assumptions needed to comp
```

```
library(ggplot2)
#Fixed Effect Model
options(contrasts=c("contr.sum","contr.poly"))
lmod_fixed<-lm(popularity~num_episodes + mean + rating, final)
summary(lmod_fixed)
```

```
##
## Call:
## lm(formula = popularity ~ num_episodes + mean + rating, data = final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5592.8 -1352.1   162.4  1432.0  4631.1
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17941.3821   591.4582  30.334 < 2e-16 ***
## num_episodes     1.0963     0.3573   3.068  0.0022 **
## mean        -1254.7834     92.7927 -13.522 < 2e-16 ***
## rating1         831.9991    131.7992   6.313 3.92e-10 ***
## rating2         782.7935    320.0976   2.445  0.0146 *
## rating3         313.8185    148.9843   2.106  0.0354 *
## rating4        -330.3490    135.6825  -2.435  0.0151 *
## rating5        -844.2483    384.0480  -2.198  0.0281 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1797 on 1138 degrees of freedom
## Multiple R-squared:  0.2052, Adjusted R-squared:  0.2003
## F-statistic: 41.97 on 7 and 1138 DF, p-value: < 2.2e-16
```

```
#Random Effect Model
library(lme4)
```

```
## Loading required package: Matrix
```

```
lmod_mixed<-lmer(popularity~1+(1|start_season.season),final)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
summary(lmod_mixed)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: popularity ~ 1 + (1 | start_season.season)
## Data: final
##
## REML criterion at convergence: 20673.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.68128 -0.84391 -0.00853  0.80495  2.55061
##
## Random effects:
## Groups              Name      Variance Std.Dev.
## start_season.season (Intercept)    0      0
## Residual                      4037166  2009
## Number of obs: 1146, groups: start_season.season, 4
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 10389.14      59.35     175
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
#Large/Overall Model
```

```
mixed_mod<-lmer(popularity~num_episodes + mean +(1|rating),final)
summary(mixed_mod)
```

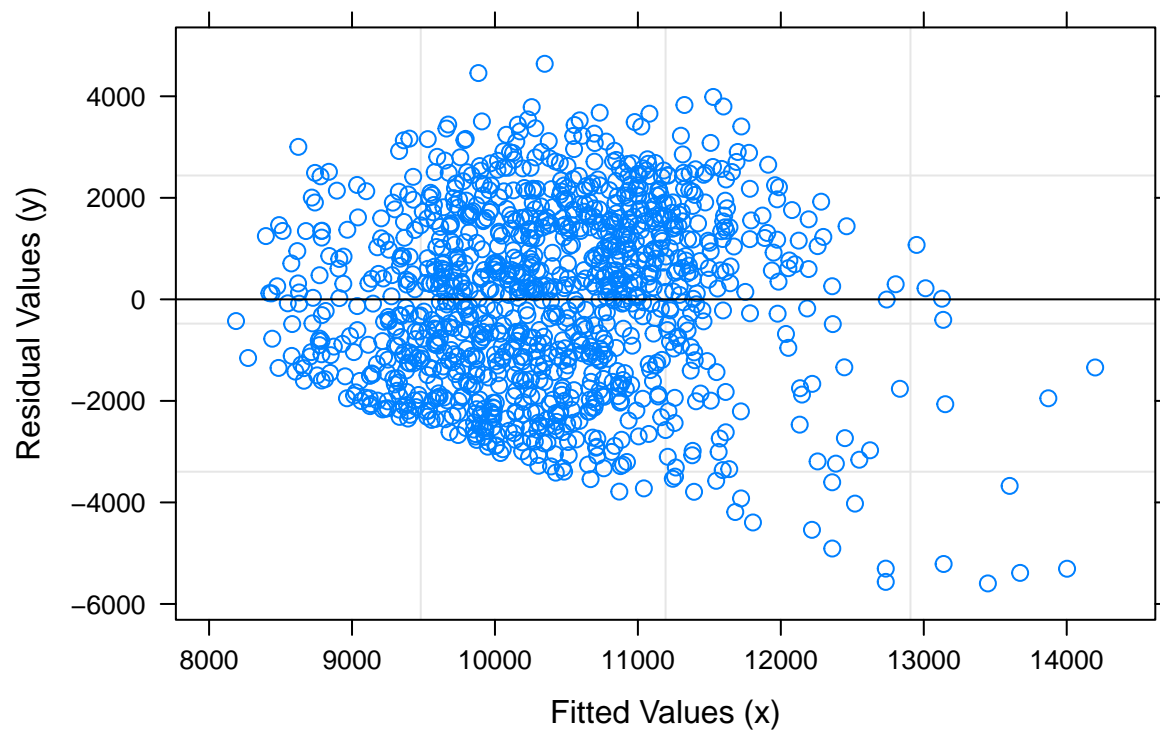
```
## Linear mixed model fit by REML ['lmerMod']
## Formula: popularity ~ num_episodes + mean + (1 | rating)
## Data: final
##
## REML criterion at convergence: 20417.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.11391 -0.76413  0.08842  0.79511  2.58206
##
## Random effects:
## Groups      Name      Variance Std.Dev.
## rating      (Intercept) 448388   669.6
## Residual          3228996  1796.9
## Number of obs: 1146, groups: rating, 6
##
## Fixed effects:
##              Estimate Std. Error t value
```

```
## (Intercept) 17968.9561 650.0086 27.644
## num_episodes 1.1146 0.3572 3.121
## mean -1253.4889 92.7275 -13.518
##
## Correlation of Fixed Effects:
## (Intr) nm_psd
## num_episodes -0.060
## mean -0.892 0.038
```

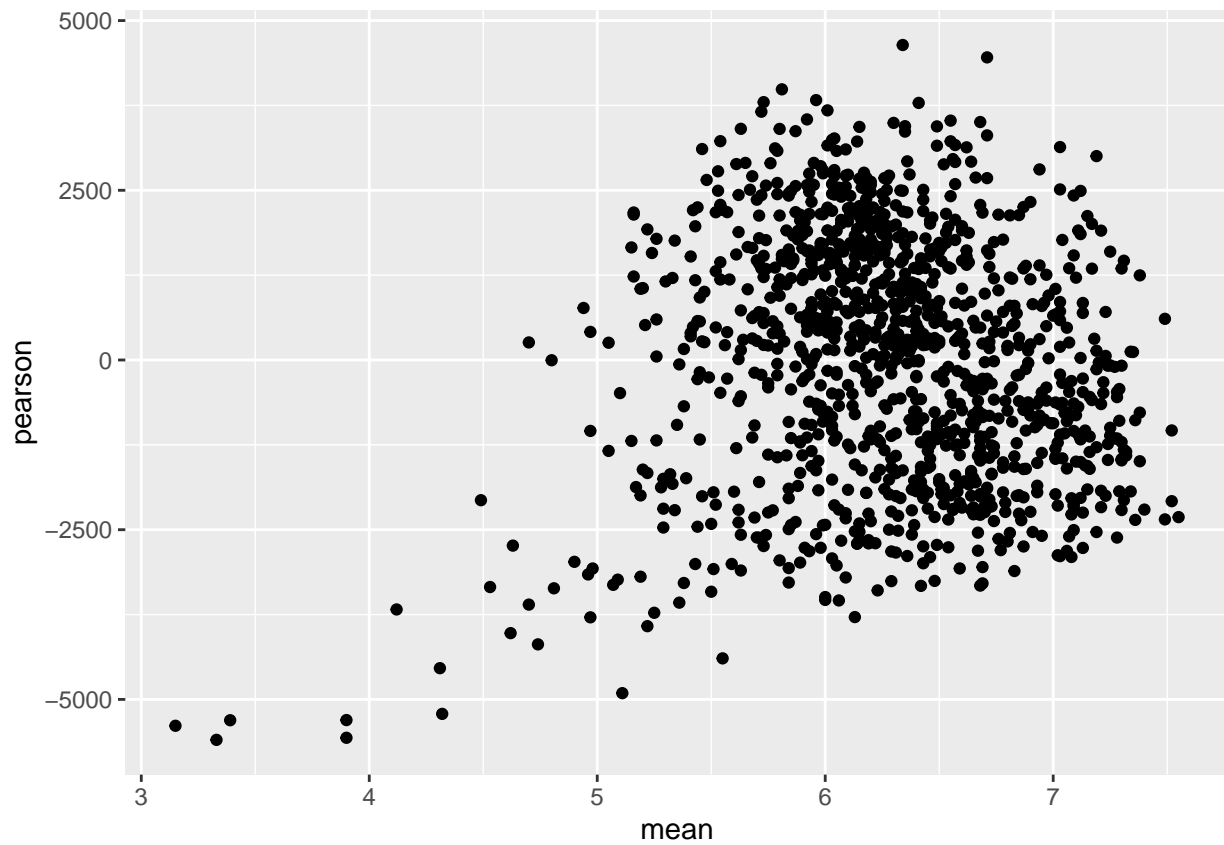
#Diagnostics

```
revised <- subset(final, num_episodes <= 1000)
plot(mixed_mod,xlab='Fitted Values (x)',ylab='Residual Values (y)',main='Scatterplot of the Residual Va
```

Scatterplot of the Residual Values as a Function of the Fitted Values

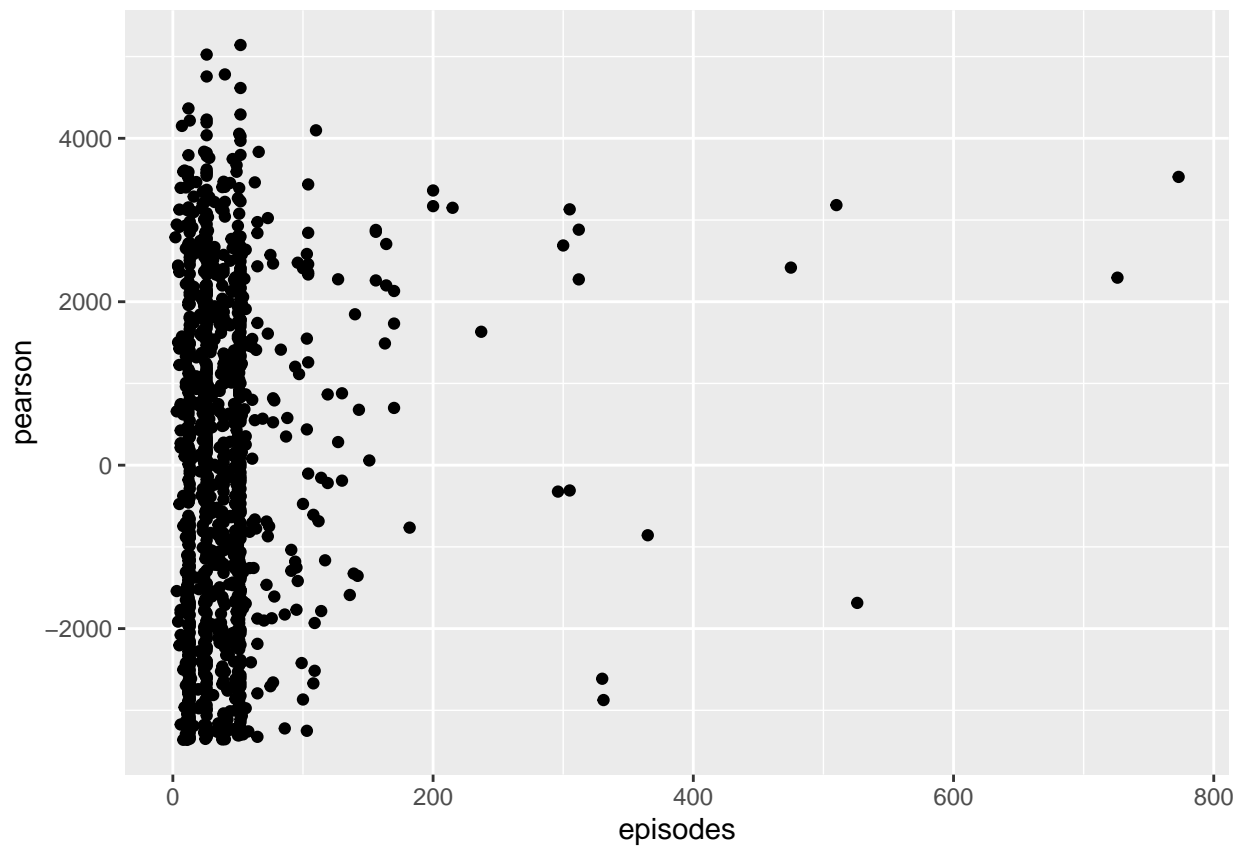


```
ggplot(data.frame(mean=final$mean,pearson=residuals(mixed_mod,type="pearson")),
  aes(x=mean,y=pearson)) +
  geom_point()
```

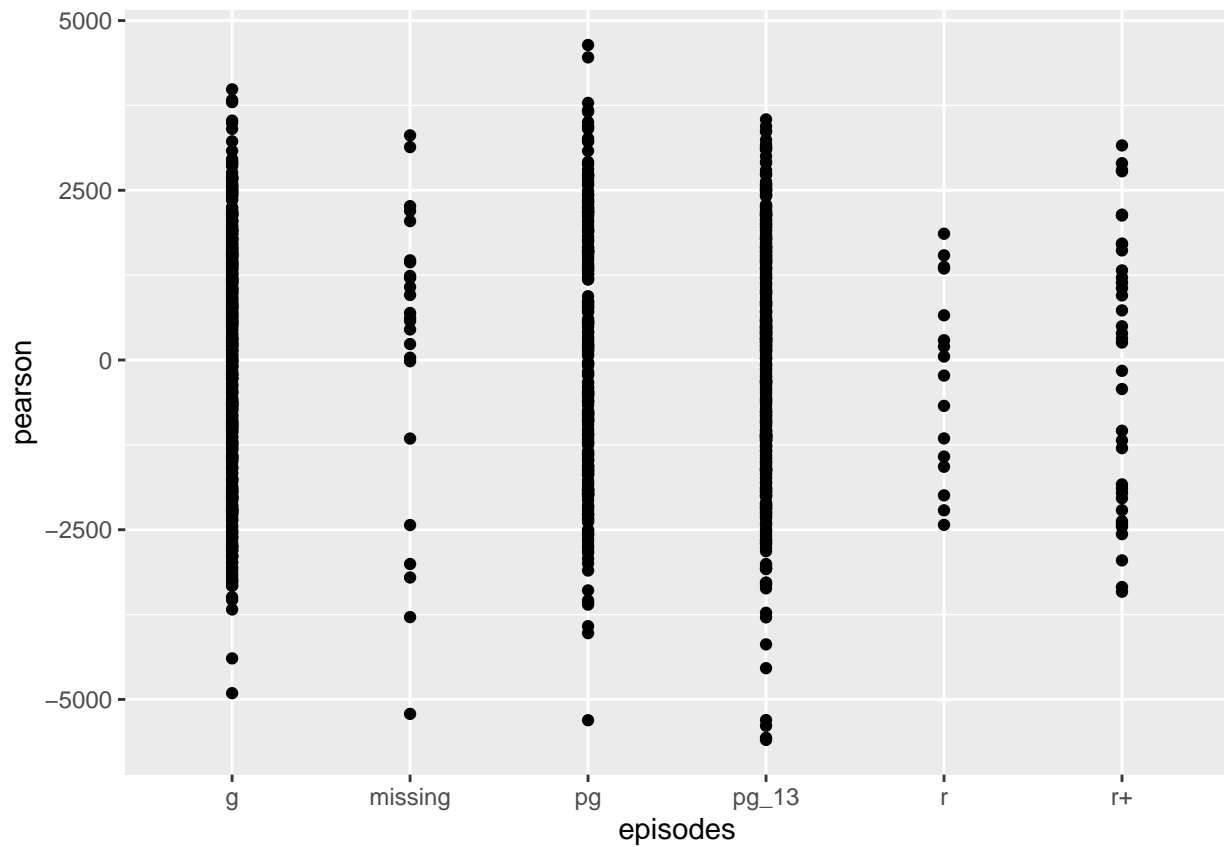


```
ggplot(data.frame(episodes=revised$num_episodes,pearson=residuals(lmer(popularity~1+(1|start_season.sea
  aes(x=episodes,y=pearson)) +
  geom_point()
```

```
## boundary (singular) fit: see help('isSingular')
```

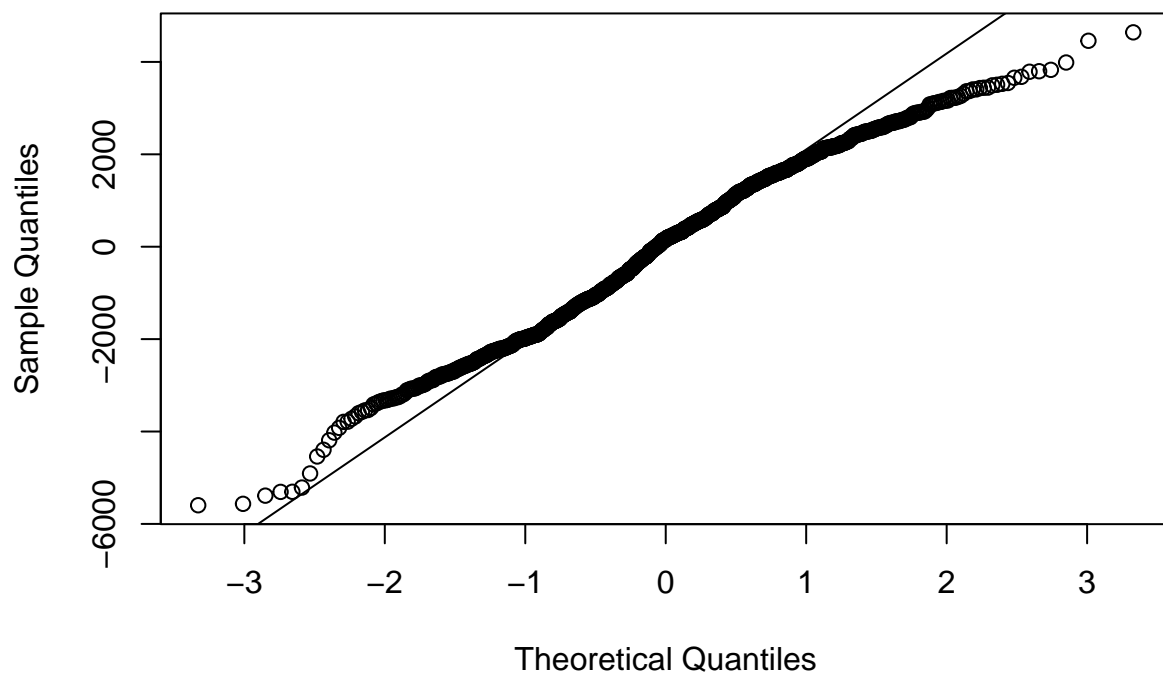


```
ggplot(data.frame(episodes=final$rating,pearson=residuals(mixed_mod,type="pearson")),  
  aes(x=episodes,y=pearson)) +  
  geom_point()
```



```
qqnorm(resid(mixed_mod))
qqline(resid(mixed_mod))
```

Normal Q-Q Plot



#Inference

```
red_mixed <- lmer(popularity~num_episodes + mean + (1+num_episodes|rating),final)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
## Model failed to converge with max|grad| = 8.04421 (tol = 0.002, component 1)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unidentifiable: large eigenvalue ratio  
## - Rescale variables?  
## - Rescale variables?
```

```
summary(red_mixed)
```

```
## Linear mixed model fit by REML ['lmerMod']  
## Formula: popularity ~ num_episodes + mean + (1 + num_episodes | rating)  
## Data: final  
##  
## REML criterion at convergence: 20409.5  
##  
## Scaled residuals:  
##      Min       1Q   Median       3Q      Max   
## -3.08727 -0.75913  0.07239  0.79602  2.56719   
##  
## Random effects:  
## Groups Name Variance Std.Dev. Corr  
## rating (Intercept) 1898572.9 1377.89  
## num_episodes 101.8 10.09 -0.98  
## Residual 3178744.7 1782.90  
## Number of obs: 1146, groups: rating, 6  
##  
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 18006.826 815.808 22.072  
## num_episodes 7.227 4.315 1.675  
## mean -1295.404 92.524 -14.001  
##  
## Correlation of Fixed Effects:  
## (Intr) nm_psd  
## num_episodes -0.663  
## mean -0.708 -0.015  
## optimizer (nloptwrap) convergence code: 0 (OK)  
## Model failed to converge with max|grad| = 8.04421 (tol = 0.002, component 1)  
## Model is nearly unidentifiable: very large eigenvalue  
## - Rescale variables?  
## Model is nearly unidentifiable: large eigenvalue ratio  
## - Rescale variables?
```

```
anova(red_mixed,mixed_mod)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: final
```

```
## Models:
## mixed_mod: popularity ~ num_episodes + mean + (1 | rating)
## red_mixed: popularity ~ num_episodes + mean + (1 + num_episodes | rating)
##      npar   AIC   BIC logLik deviance Chisq Df Pr(>Chisq)
## mixed_mod    5 20451 20476 -10220    20441
## red_mixed    7 20448 20483 -10217    20434  6.97  2    0.03065 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

revised <- subset(final, num_episodes <= 1000)
```