

Predicting Song Popularity - Model Testing

STAT 420, Summer 2023, UIUC - Final Data Project

Soumya Nanda, Jonas Jansen, Noam Isachar

Extracted this file for faster computations on model finding. The final models should be included into the report file.

```
library(knitr)
library(ggplot2)
library(GGally)

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
library(MASS)
library(gridExtra)
library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric

library(car)

## Loading required package: carData
```

Load prepared train dataset (currently 10% of all entries). Convert necessary variables into factors.

```
## 'data.frame': 115905 obs. of 17 variables:
## $ popularity      : int 35 0 28 14 0 2 64 7 16 17 ...
## $ year           : int 2010 2012 2000 2011 2010 2007 2019 2005 2007 2023 ...
## $ genre          : chr "spanish" "trip-hop" "alt-rock" "dancehall" ...
## $ danceability    : num 0.692 0.677 0.52 0.721 0.688 0.428 0.618 0.474 0.586 0.145 ...
## $ energy          : num 0.579 0.696 0.816 0.505 0.419 0.362 0.823 0.902 0.702 0.765 ...
## $ key             : Factor w/ 12 levels "0","1","2","3",...: 7 3 3 11 11 10 8 4 5 3 ...
## $ loudness        : num -7.08 -5.12 -6.57 -10.42 -12.13 ...
## $ mode            : Factor w/ 2 levels "major","minor": 2 2 1 2 2 1 1 2 2 2 ...
## $ speechiness     : num 0.243 0.0468 0.0299 0.0591 0.111 0.0337 0.0447 0.0728 0.0482 0.0506 ...
## $ acousticness    : num 0.355 0.48 0.00109 0.0826 0.155 0.843 0.0132 0.00182 0.553 0.00193 ...
## $ instrumentalness: num 0 0 0.000023 0.0003 0.887 0.0000174 0 0.00328 0.00312 0.301 ...
```

```

## $ liveness      : num  0.0835 0.0594 0.0704 0.0756 0.162 0.129 0.288 0.121 0.0943 0.115 ...
## $ valence       : num  0.738 0.288 0.488 0.782 0.301 0.253 0.58 0.467 0.69 0.319 ...
## $ tempo         : num  77.6 97.5 99 172.3 180.1 ...
## $ time_signature: Factor w/ 5 levels "0","1","3","4",...: 4 3 4 4 4 4 4 4 4 3 ...
## $ duration_m    : num  3.8 5.32 4.38 3.52 5.72 ...
## $ upper_genre   : Factor w/ 19 levels "acoustic","blues",...: 19 18 15 5 19 19 15 1 19 10 ...

## 'data.frame': 115905 obs. of 17 variables:
## $ popularity     : num  5.18 -5 4.74 3.48 -5 ...
## $ danceability   : num  -0.303 -0.317 -0.466 -0.275 -0.307 ...
## $ energy          : num  -0.401 -0.294 -0.18 -0.466 -0.54 ...
## $ loudness        : num  -8.07 -6.12 -7.57 -11.42 -13.13 ...
## $ speechiness     : num  -1.32 -2.64 -2.96 -2.46 -1.97 ...
## $ acousticness    : num  -0.935 -0.683 -3.722 -1.964 -1.556 ...
## $ instrumentalness: num  -18.42 -18.42 -10.68 -8.11 -0.12 ...
## $ liveness         : num  -2.2 -2.46 -2.33 -2.28 -1.66 ...
## $ valence          : num  -0.278 -0.877 -0.583 -0.229 -0.856 ...
## $ tempo            : num  28.6 33.8 34.2 51.1 52.7 ...
## $ duration_m       : num  1.53 1.98 1.72 1.43 2.09 ...
## $ year             : int  2010 2012 2000 2011 2010 2007 2019 2005 2007 2023 ...
## $ genre             : Factor w/ 82 levels "acoustic","afrobeat",...: 77 82 3 16 45 68 64 38 2 5 ...
## $ upper_genre       : Factor w/ 19 levels "acoustic","blues",...: 19 18 15 5 19 19 15 1 19 10 ...
## $ key               : Factor w/ 12 levels "0","1","2","3",...: 7 3 3 11 11 10 8 4 5 3 ...
## $ mode              : Factor w/ 2 levels "major","minor": 2 2 1 2 2 1 1 2 2 2 ...
## $ time_signature    : Factor w/ 5 levels "0","1","3","4",...: 4 3 4 4 4 4 4 4 4 3 ...

```

Model finding

Prepare test and Train data.

For an easier model fitting, let's create a data set without `artist_name` and `track_name`.

TODO: Split into test and train

Model Test Stats

Create a function, that is performing several test on the model and return the results.

```

check_model = function(model, test_data){
  sample_idx = sample(length(resid(model)), 5000)
  residuals_sample = resid(model)[sample_idx]

  # shapiro
  shapiro = shapiro.test(residuals_sample)$p.value

  # bp test
  bptest = bptest(model)$p.value

  # leverage
  high_leverage_count = sum(hatvalues(model) > 2 * mean(hatvalues(model)))

  # outliers
  outliers_count = length(rstandard(model)[abs(rstandard(model)) > 2])

  # influence
  influence_count = sum(cooks.distance(model) > 4 / length(cooks.distance(model)))
}

```

```

# loocv_rmse
loocv_rmse = sqrt(mean((resid(model) / (1 - hatvalues(model))) ^ 2))

# adjusted r squared
adjusted_r_squared = summary(model)$adj.r.squared

# TODO: Some test to test against the test_data

# Creating a data frame to store the results
results = data.frame(
  shapiro = shapiro,
  bptest = bptest,
  high_leverage_count = high_leverage_count,
  outliers_count = outliers_count,
  influence_count = influence_count,
  loocv_rmse = loocv_rmse,
  adjusted_r_squared = adjusted_r_squared
)

results
}

```

1. Full additive.

First try, fit an full additive model with popularity as response.

```

data = subset(data_raw, select = -c(upper_genre))

model_add_full = lm(popularity ~ ., data = data)

```

Let's do some first tests.

```

summary(model_add_full)

##
## Call:
## lm(formula = popularity ~ ., data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -67.22   -7.20   -1.56    5.64   68.14 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             -1.62e+03  9.63e+00 -167.99 < 2e-16 ***
## year                   8.17e-01  4.75e-03  171.96 < 2e-16 ***
## genreafrobeat          -1.04e+01  3.89e-01  -26.88 < 2e-16 ***
## genrealt-rock           2.07e+01  3.37e-01   61.34 < 2e-16 ***
## genreambient            5.91e+00  3.45e-01   17.15 < 2e-16 ***
## genreblack-metal        -4.88e+00  3.47e-01  -14.04 < 2e-16 ***
## genreblues               4.12e+00  3.38e-01   12.20 < 2e-16 ***
## genrebroadcast           -1.42e+01  3.93e-01  -36.25 < 2e-16 ***
## genrecantopop            -5.88e+00  3.59e-01  -16.37 < 2e-16 ***

```

## genrechicago-house	-1.53e+01	5.40e-01	-28.40	< 2e-16 ***
## genrechill	6.99e+00	3.40e-01	20.54	< 2e-16 ***
## genreclassical	1.13e+01	3.52e-01	32.15	< 2e-16 ***
## genreclub	-8.27e+00	3.66e-01	-22.62	< 2e-16 ***
## genrecomedy	-7.46e+00	4.13e-01	-18.09	< 2e-16 ***
## genrecountry	1.51e+01	3.45e-01	43.86	< 2e-16 ***
## genredance	2.41e+01	3.50e-01	68.94	< 2e-16 ***
## genredancehall	-3.82e+00	3.46e-01	-11.04	< 2e-16 ***
## genredeath-metal	2.30e+00	3.58e-01	6.43	1.3e-10 ***
## genredeep-house	-3.29e-01	3.60e-01	-0.92	0.36014
## genredetroit-techno	-1.56e+01	5.86e-01	-26.64	< 2e-16 ***
## genredisco	1.69e-01	3.58e-01	0.47	0.63763
## genredrum-and-bass	-7.54e+00	3.80e-01	-19.86	< 2e-16 ***
## genredub	-1.72e+00	3.47e-01	-4.97	6.9e-07 ***
## genredubstep	-1.35e+01	5.42e-01	-24.91	< 2e-16 ***
## genreedm	6.60e+00	4.06e-01	16.25	< 2e-16 ***
## genreelectro	1.53e+01	3.96e-01	38.53	< 2e-16 ***
## genreelectronic	5.78e+00	4.26e-01	13.57	< 2e-16 ***
## genreemo	5.37e+00	3.35e-01	16.02	< 2e-16 ***
## genrefolk	1.55e+01	3.50e-01	44.19	< 2e-16 ***
## genreforro	-7.36e+00	3.46e-01	-21.25	< 2e-16 ***
## genrefrench	8.91e+00	3.45e-01	25.78	< 2e-16 ***
## genrefunk	8.01e+00	3.67e-01	21.81	< 2e-16 ***
## genregarage	-7.06e-01	3.50e-01	-2.02	0.04369 *
## genregerman	7.68e+00	3.51e-01	21.87	< 2e-16 ***
## genregospel	1.11e+00	3.30e-01	3.38	0.00074 ***
## genregoth	-5.87e+00	3.51e-01	-16.72	< 2e-16 ***
## genregrindcore	-1.52e+01	3.90e-01	-38.92	< 2e-16 ***
## genregroove	-5.09e+00	3.75e-01	-13.59	< 2e-16 ***
## genreguitar	-5.15e+00	3.53e-01	-14.58	< 2e-16 ***
## genrehard-rock	2.49e+00	3.79e-01	6.56	5.4e-11 ***
## genrehardcore	7.93e+00	3.67e-01	21.59	< 2e-16 ***
## genrehardstyle	-6.23e+00	3.92e-01	-15.89	< 2e-16 ***
## genreheavy-metal	-1.46e+01	3.74e-01	-38.96	< 2e-16 ***
## genrehip-hop	2.65e+01	3.67e-01	72.40	< 2e-16 ***
## genrehouse	7.93e+00	5.30e-01	14.96	< 2e-16 ***
## genreindian	-7.26e+00	3.36e-01	-21.61	< 2e-16 ***
## genreindie-pop	1.72e+01	4.05e-01	42.49	< 2e-16 ***
## genreindustrial	-6.58e+00	3.69e-01	-17.84	< 2e-16 ***
## genrejazz	1.31e+01	3.61e-01	36.23	< 2e-16 ***
## genrek-pop	9.65e+00	3.34e-01	28.85	< 2e-16 ***
## genremetal	2.03e+01	4.78e-01	42.45	< 2e-16 ***
## genremetalcore	-1.88e+00	4.94e-01	-3.80	0.00015 ***
## genreminimal-techno	-8.92e+00	3.81e-01	-23.38	< 2e-16 ***
## genrenew-age	-3.06e+00	3.42e-01	-8.94	< 2e-16 ***
## genreopera	-9.60e+00	3.60e-01	-26.63	< 2e-16 ***
## genreparty	-1.04e+01	4.15e-01	-25.02	< 2e-16 ***
## genrepiano	3.21e+00	3.71e-01	8.66	< 2e-16 ***
## genrepop	3.75e+01	4.90e-01	76.62	< 2e-16 ***
## genrepop-film	-5.38e-01	3.55e-01	-1.52	0.12962
## genrepower-pop	-1.24e+01	3.51e-01	-35.19	< 2e-16 ***
## genreprogressive-house	-1.45e+00	4.18e-01	-3.47	0.00052 ***
## genrepsych-rock	-1.11e+00	3.76e-01	-2.95	0.00319 **
## genrepunk	1.28e+01	5.01e-01	25.65	< 2e-16 ***

```

## genrepunk-rock      1.99e+00  4.56e-01   4.37  1.2e-05 ***
## genrerock          2.82e+01  6.53e-01   43.13 < 2e-16 ***
## genrerock-n-roll  -9.46e+00  3.59e-01  -26.32 < 2e-16 ***
## genreromance       -1.68e+01  4.84e-01  -34.60 < 2e-16 ***
## genresad           1.19e+01  5.01e-01   23.80 < 2e-16 ***
## genresalsa          -5.87e+00  3.54e-01  -16.58 < 2e-16 ***
## genresamba          -7.78e+00  3.45e-01  -22.57 < 2e-16 ***
## genresertanejo     1.84e+00  3.49e-01    5.27  1.4e-07 ***
## genreshow-tunes    -8.41e+00  3.99e-01  -21.06 < 2e-16 ***
## genresinger-songwriter 5.20e+00  3.63e-01   14.32 < 2e-16 ***
## genreska            -3.40e+00  3.76e-01  -9.06 < 2e-16 ***
## genresleep          1.15e+00  3.74e-01   3.08  0.00211 **
## genresongwriter     3.07e+00  1.45e+00   2.12  0.03430 *
## genresoul            1.22e+01  4.31e-01   28.29 < 2e-16 ***
## genrespanish         5.34e+00  3.39e-01   15.75 < 2e-16 ***
## genreswedish         2.15e+00  3.93e-01    5.48  4.3e-08 ***
## genretango           -1.47e+01  3.61e-01  -40.74 < 2e-16 ***
## genretechno          -2.68e+00  4.65e-01  -5.76  8.2e-09 ***
## genretrance          -1.12e+00  4.22e-01  -2.65  0.00815 **
## genretrip-hop        -1.11e+01  3.93e-01  -28.31 < 2e-16 ***
## danceability         4.43e+00  2.58e-01   17.14 < 2e-16 ***
## energy               -1.11e+00  2.62e-01  -4.22  2.4e-05 ***
## key1                 1.05e-01  1.39e-01   0.76  0.44922
## key2                 1.62e-01  1.34e-01   1.21  0.22570
## key3                 4.89e-01  2.00e-01   2.45  0.01446 *
## key4                 1.21e-01  1.47e-01   0.82  0.41005
## key5                 5.70e-02  1.46e-01   0.39  0.69602
## key6                 3.88e-01  1.56e-01   2.49  0.01276 *
## key7                 -1.48e-01  1.30e-01  -1.14  0.25537
## key8                 6.18e-01  1.58e-01   3.92  9.0e-05 ***
## key9                 -8.26e-02  1.36e-01  -0.61  0.54271
## key10                1.22e-01  1.56e-01   0.78  0.43500
## key11                1.40e-01  1.49e-01   0.94  0.34846
## loudness             2.02e-01  1.13e-02   17.88 < 2e-16 ***
## modeminor            4.77e-01  6.92e-02   6.89  5.8e-12 ***
## speechiness          -1.60e+00  3.63e-01  -4.40  1.1e-05 ***
## acousticness         6.42e-01  1.61e-01   4.00  6.4e-05 ***
## instrumentalness    -2.51e+00  1.18e-01  -21.20 < 2e-16 ***
## liveness              -2.40e-01  1.73e-01  -1.38  0.16616
## valence               -3.16e+00  1.64e-01  -19.20 < 2e-16 ***
## tempo                  1.43e-03  1.14e-03   1.26  0.20688
## time_signature1     -5.89e+00  1.05e+00  -5.62  2.0e-08 ***
## time_signature3     -5.68e+00  1.02e+00  -5.59  2.2e-08 ***
## time_signature4     -5.10e+00  1.01e+00  -5.04  4.8e-07 ***
## time_signature5     -5.69e+00  1.03e+00  -5.51  3.6e-08 ***
## duration_m          -4.14e-01  1.70e-02  -24.31 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.6 on 115796 degrees of freedom
## Multiple R-squared:  0.554, Adjusted R-squared:  0.553
## F-statistic: 1.33e+03 on 108 and 115796 DF, p-value: <2e-16

```

What we can see is: - p-value is very low - most variables are significant. Except some categorical keys. -

RSS? - Adjusted R => Okay, but could be better.

Model seems decent, but we can do better.

Trying to find a smaller good model. Using both AIC or BIC running backward did not lead to smaller models.

Let's have a look into the test results:

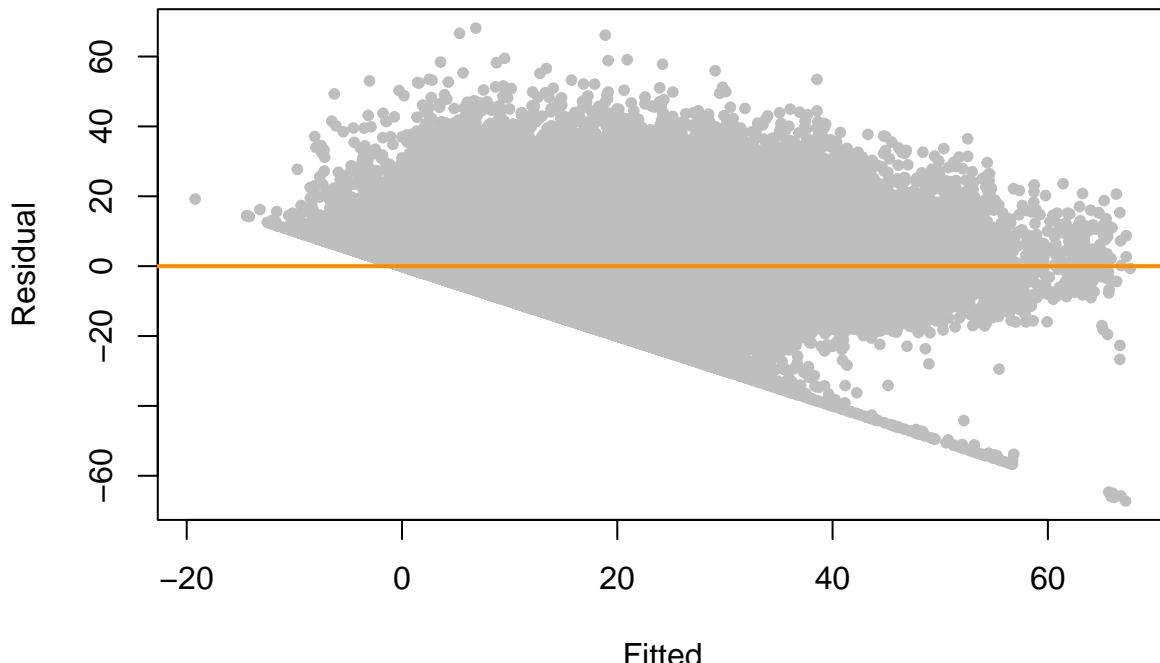
```
results = check_model(model_add_full, data_tst)
results

##      shapiro bptest high_leverage_count outliers_count influence_count
## BP 1.335e-34      0            3468        5793        5594
## loocv_rmse adjusted_r_squared
## BP      10.63       0.5532
```

The test results make show that the normality and constant variance and normal assumption. The Plots Fitted vs Residuals underline that.

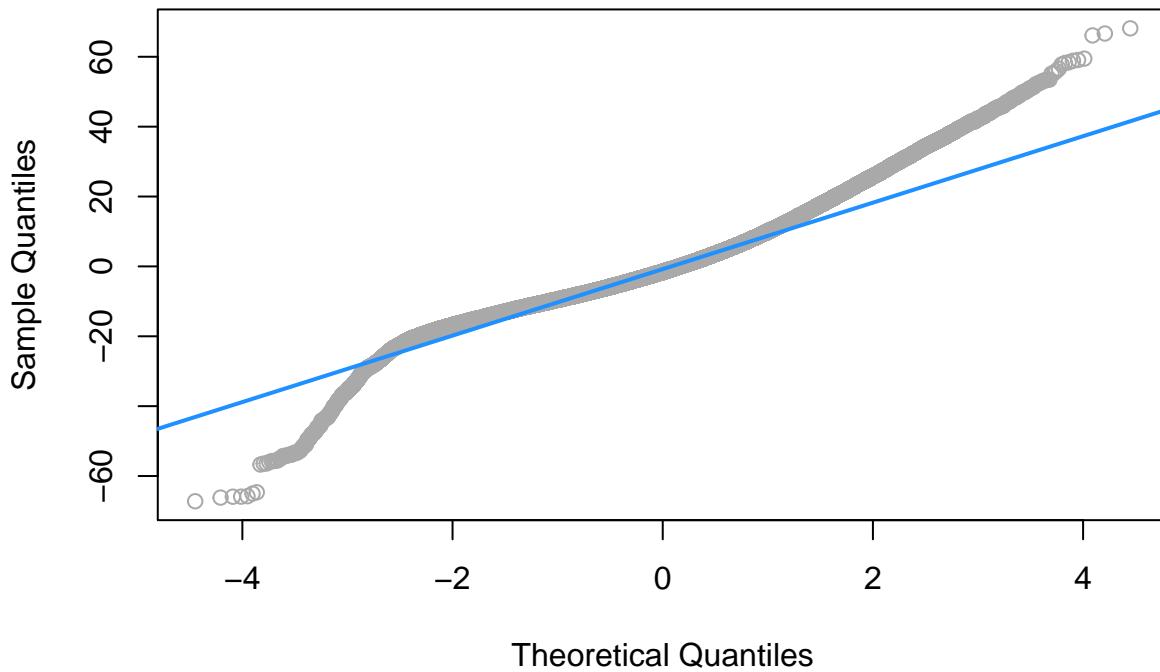
```
plot(fitted(model_add_full), resid(model_add_full), col = "grey", pch = 20,
      xlab = "Fitted", ylab = "Residual",
      main = "Full Additive Model - Fitted vs Residuals")
abline(h = 0, col = "darkorange", lwd = 2)
```

Full Additive Model – Fitted vs Residuals



```
qqnorm(resid(model_add_full), col = "darkgrey")
qqline(resid(model_add_full), col = "dodgerblue", lwd = 2)
```

Normal Q-Q Plot



The model needs adjustments. Transformations could be helpful to fulfill the assumptions.

2. Smaller additive models, but more explainable

In terms of finding a better explainable model, we leave out the year and the genre.

```
### TEMP
data = subset(data_raw, select = -c(upper_genre))
### TEMP

model_add_noyear = lm(popularity ~ . -year, data = data)
model_add_nogenre = lm(popularity ~ . -genre -year, data = data)
model_add_nogenre = lm(popularity ~ . -genre, data = data)
```

Let's plot the model's summary

```
summary(model_add_noyear)

##
## Call:
## lm(formula = popularity ~ . - year, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -59.48    -8.21   -1.82    6.81   68.04 
##
## Coefficients:
## (Intercept) 28.33941   1.17957   24.03 < 2e-16 ***
## genreafrobeat -7.47974   0.43496  -17.20 < 2e-16 ***
## genrealt-rock 20.55754   0.37722   54.50 < 2e-16 ***
## genreambient  6.16510   0.38618   15.96 < 2e-16 ***
```

## genreblack-metal	-5.07420	0.38928	-13.03	< 2e-16 ***
## genreblues	4.98158	0.37868	13.16	< 2e-16 ***
## genrebreakbeat	-13.18303	0.43989	-29.97	< 2e-16 ***
## genrecantopop	-5.96595	0.40243	-14.82	< 2e-16 ***
## genrechicago-house	-12.62818	0.60472	-20.88	< 2e-16 ***
## genrechill	7.41760	0.38118	19.46	< 2e-16 ***
## genreclassical	11.58007	0.39475	29.33	< 2e-16 ***
## genreclub	-6.80286	0.40942	-16.62	< 2e-16 ***
## genrecomedy	-7.31787	0.46222	-15.83	< 2e-16 ***
## genrecountry	15.19842	0.38702	39.27	< 2e-16 ***
## genredance	23.46692	0.39234	59.81	< 2e-16 ***
## genredancehall	-4.10094	0.38812	-10.57	< 2e-16 ***
## genredeath-metal	1.67582	0.40122	4.18	3.0e-05 ***
## genreddeep-house	0.30612	0.40290	0.76	0.44738
## genredetroit-techno	-13.17945	0.65593	-20.09	< 2e-16 ***
## genredisco	0.04114	0.40101	0.10	0.91828
## genredrum-and-bass	-7.20492	0.42558	-16.93	< 2e-16 ***
## genredub	-2.33999	0.38848	-6.02	1.7e-09 ***
## genredubstep	-14.56842	0.60704	-24.00	< 2e-16 ***
## genreedm	6.45635	0.45536	14.18	< 2e-16 ***
## genreelectro	13.09064	0.44384	29.49	< 2e-16 ***
## genreelectronic	6.29104	0.47777	13.17	< 2e-16 ***
## genreemo	4.85381	0.37544	12.93	< 2e-16 ***
## genrefolk	15.97470	0.39196	40.76	< 2e-16 ***
## genreforro	-6.67356	0.38789	-17.20	< 2e-16 ***
## genrefrench	7.96734	0.38701	20.59	< 2e-16 ***
## genrefunk	8.34252	0.41123	20.29	< 2e-16 ***
## genregarage	-0.66498	0.39191	-1.70	0.08974 .
## genregerman	7.63520	0.39339	19.41	< 2e-16 ***
## genregospel	1.66369	0.36993	4.50	6.9e-06 ***
## genregoth	-5.65866	0.39350	-14.38	< 2e-16 ***
## genregrindcore	-15.76204	0.43651	-36.11	< 2e-16 ***
## genregroove	-5.26606	0.41985	-12.54	< 2e-16 ***
## genreguitar	-6.01189	0.39597	-15.18	< 2e-16 ***
## genrehard-rock	3.11525	0.42443	7.34	2.2e-13 ***
## genrehardcore	7.29706	0.41155	17.73	< 2e-16 ***
## genrehardstyle	-4.22108	0.43883	-9.62	< 2e-16 ***
## genreheavy-metal	-13.60278	0.41872	-32.49	< 2e-16 ***
## genrehip-hop	26.54443	0.41081	64.62	< 2e-16 ***
## genrehouse	5.04464	0.59385	8.49	< 2e-16 ***
## genreindian	-6.33330	0.37617	-16.84	< 2e-16 ***
## genreindie-pop	16.81476	0.45341	37.09	< 2e-16 ***
## genreindustrial	-5.83133	0.41323	-14.11	< 2e-16 ***
## genrejazz	13.31734	0.40410	32.96	< 2e-16 ***
## genrek-pop	9.49605	0.37458	25.35	< 2e-16 ***
## genremetal	21.71747	0.53559	40.55	< 2e-16 ***
## genremetalcore	-2.07516	0.55381	-3.75	0.00018 ***
## genreminimal-techno	-8.04177	0.42723	-18.82	< 2e-16 ***
## genrenew-age	-3.14804	0.38362	-8.21	2.3e-16 ***
## genreopera	-8.91310	0.40380	-22.07	< 2e-16 ***
## genreparty	-9.72652	0.46461	-20.93	< 2e-16 ***
## genrepiano	4.34333	0.41510	10.46	< 2e-16 ***
## genrepop	37.13582	0.54897	67.65	< 2e-16 ***
## genrepop-film	-0.05991	0.39740	-0.15	0.88017

```

## genrepower-pop      -12.37122   0.39368  -31.42 < 2e-16 ***
## genreprogressive-house -1.05304   0.46843  -2.25  0.02458 *
## genrepsych-rock     -0.69887   0.42094  -1.66  0.09686 .
## genrepunk            13.51078   0.56117  24.08 < 2e-16 ***
## genrepunk-rock       2.30075   0.51118   4.50  6.8e-06 ***
## genrerock             28.58645   0.73199  39.05 < 2e-16 ***
## genrerock-n-roll     -8.84461   0.40273  -21.96 < 2e-16 ***
## genreromance          -16.97063   0.54246  -31.28 < 2e-16 ***
## genresad              17.51150   0.55971  31.29 < 2e-16 ***
## genresalsa             -4.74502   0.39662  -11.96 < 2e-16 ***
## genresamba             -7.10591   0.38645  -18.39 < 2e-16 ***
## genresertanejo        1.98266   0.39069   5.07  3.9e-07 ***
## genreshow-tunes        -7.55477   0.44744  -16.88 < 2e-16 ***
## genresinger-songwriter 5.70952   0.40719  14.02 < 2e-16 ***
## genreska              -2.50017   0.42121  -5.94  2.9e-09 ***
## genresleep              2.51420   0.41869   6.00  1.9e-09 ***
## genresongwriter         -4.88451   1.62673  -3.00  0.00268 **
## genresoul               12.82605   0.48259  26.58 < 2e-16 ***
## genrespanish            5.16168   0.38000  13.58 < 2e-16 ***
## genreswedish            1.84675   0.44068   4.19  2.8e-05 ***
## genretango              -14.68761   0.40392  -36.36 < 2e-16 ***
## genretechno             -2.92184   0.52137  -5.60  2.1e-08 ***
## genretrance              -2.07728   0.47235  -4.40  1.1e-05 ***
## genretrip-hop            -10.30490   0.44047  -23.40 < 2e-16 ***
## danceability             8.07750   0.28821  28.03 < 2e-16 ***
## energy                  -1.13499   0.29406  -3.86  0.00011 ***
## key1                     0.38549   0.15534   2.48  0.01308 *
## key2                     -0.06889   0.15009  -0.46  0.64623
## key3                     0.60787   0.22391   2.71  0.00663 **
## key4                     -0.08407   0.16511  -0.51  0.61060
## key5                     0.20879   0.16345   1.28  0.20147
## key6                     0.58671   0.17439   3.36  0.00077 ***
## key7                     -0.38010   0.14543  -2.61  0.00896 **
## key8                     0.99577   0.17679   5.63  1.8e-08 ***
## key9                     -0.41479   0.15210  -2.73  0.00639 **
## key10                    0.09451   0.17465   0.54  0.58843
## key11                    0.23741   0.16696   1.42  0.15504
## loudness                 0.29242   0.01262  23.17 < 2e-16 ***
## modeminor                0.82611   0.07753  10.66 < 2e-16 ***
## speechiness              -0.54851   0.40659  -1.35  0.17733
## acousticness              0.77666   0.17994   4.32  1.6e-05 ***
## instrumentalness         -2.10215   0.13258  -15.86 < 2e-16 ***
## liveness                  -0.45047   0.19432  -2.32  0.02044 *
## valence                  -7.67127   0.18181  -42.19 < 2e-16 ***
## tempo                      0.00573   0.00127   4.51  6.6e-06 ***
## time_signature1           -6.63081   1.17589  -5.64  1.7e-08 ***
## time_signature3           -6.55362   1.13724  -5.76  8.3e-09 ***
## time_signature4           -5.80790   1.13434  -5.12  3.1e-07 ***
## time_signature5           -6.15193   1.15730  -5.32  1.1e-07 ***
## duration_m                 -0.80754   0.01892  -42.69 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.9 on 115797 degrees of freedom

```

```

## Multiple R-squared:  0.44,   Adjusted R-squared:  0.439
## F-statistic:  849 on 107 and 115797 DF,  p-value: <2e-16
summary(model_add_noyear_nogenre)

##
## Call:
## lm(formula = popularity ~ . - genre - year, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max 
## -32.73 -12.35 -2.61 10.10 70.59 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 33.96902  1.46628  23.17 < 2e-16 ***
## danceability 10.69407  0.31159  34.32 < 2e-16 ***
## energy      -7.83096  0.35307 -22.18 < 2e-16 *** 
## key1         0.63738  0.19873   3.21  0.0013 ** 
## key2        -0.15999  0.19231  -0.83  0.4054  
## key3         1.24677  0.28687   4.35  1.4e-05 *** 
## key4        -0.01144  0.21144  -0.05  0.9568  
## key5         0.42237  0.20956   2.02  0.0439 *  
## key6         1.07901  0.22338   4.83  1.4e-06 *** 
## key7        -0.75025  0.18645  -4.02  5.7e-05 *** 
## key8         1.57541  0.22646   6.96  3.5e-12 *** 
## key9        -0.63039  0.19482  -3.24  0.0012 ** 
## key10        0.30888  0.22376   1.38  0.1675  
## key11        0.49057  0.21384   2.29  0.0218 *  
## loudness     0.30838  0.01457  21.16 < 2e-16 *** 
## modeminor    0.69466  0.09781   7.10  1.2e-12 *** 
## speechiness  -4.64145  0.39197 -11.84 < 2e-16 *** 
## acousticness -3.21679  0.20564 -15.64 < 2e-16 *** 
## instrumentalness -6.21284  0.14796 -41.99 < 2e-16 *** 
## liveness      -2.77485  0.24363 -11.39 < 2e-16 *** 
## valence       -9.33052  0.21482 -43.43 < 2e-16 *** 
## tempo          0.00215  0.00159   1.35  0.1773  
## time_signature1 -3.44926  1.49780  -2.30  0.0213 *  
## time_signature3 -3.12957  1.44655  -2.16  0.0305 *  
## time_signature4 -1.96210  1.44216  -1.36  0.1737  
## time_signature5 -3.12074  1.47367  -2.12  0.0342 *  
## duration_m     -1.04424  0.02272 -45.95 < 2e-16 *** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.3 on 115878 degrees of freedom
## Multiple R-squared:  0.0767, Adjusted R-squared:  0.0765
## F-statistic:  370 on 26 and 115878 DF,  p-value: <2e-16
summary(model_add_nogenre)

##
## Call:
## lm(formula = popularity ~ . - genre, data = data)
##

```

```

## Residuals:
##      Min     1Q Median     3Q    Max
## -34.65 -10.63  -2.15   9.13  65.05
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -1.49e+03  1.29e+01 -115.79 < 2e-16 ***
## year                  7.59e-01  6.37e-03  119.10 < 2e-16 ***
## danceability          7.03e+00  2.96e-01   23.78 < 2e-16 ***
## energy                -8.18e+00 3.33e-01  -24.54 < 2e-16 ***
## key1                  3.79e-01  1.88e-01    2.02  0.0435 *
## key2                  6.31e-02  1.82e-01    0.35  0.7283
## key3                  1.11e+00  2.71e-01    4.10  4.2e-05 ***
## key4                  1.98e-01  2.00e-01    0.99  0.3218
## key5                  2.54e-01  1.98e-01    1.28  0.1991
## key6                  9.18e-01  2.11e-01    4.36  1.3e-05 ***
## key7                 -5.51e-01  1.76e-01   -3.13  0.0017 **
## key8                  1.22e+00  2.14e-01    5.69  1.3e-08 ***
## key9                 -3.03e-01  1.84e-01   -1.65  0.0992 .
## key10                 2.71e-01  2.11e-01    1.28  0.1996
## key11                 4.07e-01  2.02e-01    2.02  0.0439 *
## loudness              2.45e-01  1.38e-02   17.76 < 2e-16 ***
## modeminor             4.02e-01  9.24e-02    4.35  1.3e-05 ***
## speechiness           -5.21e+00 3.70e-01  -14.08 < 2e-16 ***
## acousticness          -3.56e+00 1.94e-01  -18.32 < 2e-16 ***
## instrumentalness     -6.54e+00 1.40e-01  -46.83 < 2e-16 ***
## liveness              -2.62e+00 2.30e-01  -11.38 < 2e-16 ***
## valence               -5.34e+00 2.06e-01  -25.98 < 2e-16 ***
## tempo                 -2.38e-03 1.51e-03   -1.58  0.1142
## time_signature1       -2.22e+00 1.41e+00   -1.57  0.1161
## time_signature3       -1.76e+00 1.37e+00   -1.29  0.1965
## time_signature4       -6.57e-01 1.36e+00   -0.48  0.6294
## time_signature5       -2.09e+00 1.39e+00   -1.50  0.1339
## duration_m            -7.05e-01 2.16e-02  -32.57 < 2e-16 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.4 on 115877 degrees of freedom
## Multiple R-squared:  0.177, Adjusted R-squared:  0.177
## F-statistic:  926 on 27 and 115877 DF, p-value: <2e-16

```

Let's do some first tests.

```
check_model(model_add_noyear, data_tst)
```

```

##      shapiro bptest high_leverage_count outliers_count influence_count
## BP 3.624e-31      0          3567         5565        5449
## loocv_rmse adjusted_r_squared
## BP      11.91      0.4391

```

```
check_model(model_add_noyear_nogenre, data_tst)
```

```

##      shapiro bptest high_leverage_count outliers_count influence_count
## BP 1.494e-36      0          6052         4584        4292
## loocv_rmse adjusted_r_squared
## BP      15.28      0.07651

```

```

check_model(model_add_nogenre, data_tst)

##      shapiro bptest high_leverage_count outliers_count influence_count
## BP 8.915e-29      0          5996           4806           4472
## loocv_rmse adjusted_r_squared
## BP      14.42        0.1772

```

3. Using upper genre

Fit the same models but using upper_genre instead.

```

data = subset(data_raw, select = -c(genre))

model_upper_genre_add_full = lm(popularity ~ ., data = data)
model_upper_genre_add_noyear = lm(popularity ~ . -year, data = data)
model_upper_genre_add_nogenre = lm(popularity ~ . -upper_genre -year, data = data)
model_upper_genre_add_nogenre = lm(popularity ~ . -upper_genre, data = data)

```

Test these models

```

check_model(model_upper_genre_add_full, data_tst)

##      shapiro bptest high_leverage_count outliers_count influence_count
## BP 9.062e-29      0          11239           5266           4445
## loocv_rmse adjusted_r_squared
## BP      13.57        0.2717

```

```
check_model(model_upper_genre_add_noyear, data_tst)
```

```

##      shapiro bptest high_leverage_count outliers_count influence_count
## BP 5.703e-37      0          11648           4976           4314
## loocv_rmse adjusted_r_squared
## BP      14.48        0.1701

```

```
check_model(model_upper_genre_add_nogenre, data_tst)
```

```

##      shapiro bptest high_leverage_count outliers_count influence_count
## BP 4.934e-37      0          6052           4584           4292
## loocv_rmse adjusted_r_squared
## BP      15.28        0.07651

```

```
check_model(model_upper_genre_add_nogenre, data_tst)
```

```

##      shapiro bptest high_leverage_count outliers_count influence_count
## BP 1.691e-29      0          5996           4806           4472
## loocv_rmse adjusted_r_squared
## BP      14.42        0.1772

```

3. Transformed data

```

### TEMP
data = data_trans_raw
### TEMP

model_trans_add_full          = lm(popularity ~ ., data = data)
model_trans_genre_add         = lm(popularity ~ . -upper_genre, data = data)
model_trans_genre_add_noyear = lm(popularity ~ . -upper_genre -year, data = data)

```

```

model_trans_genre_add_nogenre      = lm(popularity ~ . -upper_genre -genre, data = data)
model_trans_genre_add_noyear_nogenre = lm(popularity ~ . -upper_genre -year -genre, data = data)

model_trans_upper_genre_add        = lm(popularity ~ . -genre, data = data)
model_trans_upper_genre_add_noyear = lm(popularity ~ . -genre -year, data = data)
model_trans_upper_genre_add_nogenre = lm(popularity ~ . -genre -upper_genre, data = data)
model_trans_upper_genre_add_noyear_nogenre = lm(popularity ~ . -genre -year -upper_genre, data = data)

print_checks = function(model) {
  print(summary(model))
  print(check_model(model))

  plot(fitted(model), resid(model), col = "grey", pch = 20,
    xlab = "Fitted", ylab = "Residual",
    main = "Full Additive Model - Fitted vs Residuals")
  abline(h = 0, col = "darkorange", lwd = 2)

  qqnorm(resid(model), col = "darkgrey")
  qqline(resid(model), col = "dodgerblue", lwd = 2)
}

print_checks(model_trans_add_full)

##
## Call:
## lm(formula = popularity ~ ., data = data)
##
## Residuals:
##    Min     1Q   Median     3Q    Max 
## -12.876 -0.736   0.330   1.453   9.673 
##
## Coefficients: (18 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -2.86e+02  2.28e+00 -125.34 < 2e-16 ***
## danceability 1.63e-02  6.53e-02   0.25  0.80251  
## energy       2.20e-01  6.77e-02   3.25  0.00115 ** 
## loudness     5.02e-02  2.59e-03  19.40 < 2e-16 ***
## speechiness -1.27e-01  1.61e-02  -7.91  2.6e-15 ***
## acousticness 7.42e-02  8.71e-03   8.52 < 2e-16 *** 
## instrumentalness -1.33e-02  1.45e-03  -9.17 < 2e-16 *** 
## liveness     5.73e-02  1.26e-02   4.56  5.1e-06 *** 
## valence      -5.07e-01  2.69e-02  -18.84 < 2e-16 *** 
## tempo        5.89e-04  1.14e-03   0.52  0.60394  
## duration_m   -2.37e-01  1.48e-02  -16.05 < 2e-16 *** 
## year         1.43e-01  1.12e-03  127.44 < 2e-16 *** 
## genreafrobeat -2.74e+00  9.19e-02  -29.84 < 2e-16 *** 
## genrealt-rock 2.21e+00  7.95e-02   27.76 < 2e-16 *** 
## genreambient  7.75e-01  8.11e-02   9.55 < 2e-16 *** 
## genreblack-metal -8.72e-01  8.27e-02  -10.55 < 2e-16 *** 
## genreblues    1.15e+00  7.99e-02  14.40 < 2e-16 *** 
## genrebroadcast -4.70e+00  9.32e-02  -50.46 < 2e-16 *** 
## genrecantopop -6.90e-01  8.51e-02  -8.10  5.4e-16 *** 
## genrechicago-house -4.99e+00  1.28e-01  -39.02 < 2e-16 *** 
## genrechill     2.54e-01  8.04e-02   3.16  0.00155 **
```

## genreclassical	1.98e+00	8.33e-02	23.75	< 2e-16 ***
## genreclub	-2.49e+00	8.64e-02	-28.83	< 2e-16 ***
## genrecomedy	-1.19e+00	8.68e-02	-13.68	< 2e-16 ***
## genrecountry	1.88e+00	8.17e-02	23.00	< 2e-16 ***
## genredance	2.38e+00	8.27e-02	28.77	< 2e-16 ***
## genredancehall	-6.46e-01	8.22e-02	-7.86	3.8e-15 ***
## genredeath-metal	6.55e-01	8.59e-02	7.62	2.5e-14 ***
## genreddeep-house	-9.21e-01	8.51e-02	-10.83	< 2e-16 ***
## genredetroit-techno	-4.89e+00	1.38e-01	-35.34	< 2e-16 ***
## genredisco	3.32e-01	8.45e-02	3.92	8.7e-05 ***
## genredrum-and-bass	-1.73e+00	9.00e-02	-19.24	< 2e-16 ***
## genredub	-3.47e-01	8.21e-02	-4.22	2.4e-05 ***
## genredubstep	-4.34e+00	1.28e-01	-33.86	< 2e-16 ***
## genreedm	-6.00e-02	9.61e-02	-0.62	0.53246
## genreelectro	2.08e+00	9.37e-02	22.15	< 2e-16 ***
## genreelectronic	9.95e-01	1.01e-01	9.87	< 2e-16 ***
## genreemo	6.66e-01	7.93e-02	8.39	< 2e-16 ***
## genrefolk	1.92e+00	8.28e-02	23.16	< 2e-16 ***
## genreforro	-2.26e+00	8.21e-02	-27.52	< 2e-16 ***
## genrefrench	1.38e+00	8.17e-02	16.92	< 2e-16 ***
## genrefunk	1.38e+00	8.68e-02	15.91	< 2e-16 ***
## genregarage	-2.72e-01	8.28e-02	-3.29	0.00101 **
## genregerman	1.36e+00	8.30e-02	16.42	< 2e-16 ***
## genregospel	6.49e-01	7.83e-02	8.29	< 2e-16 ***
## genregoth	-8.28e-01	8.32e-02	-9.95	< 2e-16 ***
## genregrindcore	-5.04e+00	9.38e-02	-53.71	< 2e-16 ***
## genregroove	-1.61e+00	8.88e-02	-18.18	< 2e-16 ***
## genreguitar	-1.12e+00	8.32e-02	-13.43	< 2e-16 ***
## genrehard-rock	8.32e-01	8.98e-02	9.27	< 2e-16 ***
## genrehardcore	1.38e+00	8.73e-02	15.83	< 2e-16 ***
## genrehardstyle	-1.95e+00	9.28e-02	-21.05	< 2e-16 ***
## genreheavy-metal	-4.55e+00	8.90e-02	-51.11	< 2e-16 ***
## genrehip-hop	2.46e+00	8.69e-02	28.27	< 2e-16 ***
## genrehouse	1.03e+00	1.25e-01	8.23	< 2e-16 ***
## genreindian	-3.25e+00	7.91e-02	-41.01	< 2e-16 ***
## genreindie-pop	1.95e+00	9.57e-02	20.37	< 2e-16 ***
## genreindustrial	-5.94e-01	8.76e-02	-6.78	1.2e-11 ***
## genrejazz	1.92e+00	8.53e-02	22.49	< 2e-16 ***
## genrek-pop	1.17e+00	7.92e-02	14.83	< 2e-16 ***
## genremetal	1.92e+00	1.13e-01	16.96	< 2e-16 ***
## genremetalcore	-4.33e-01	1.17e-01	-3.69	0.00022 ***
## genreminimal-techno	-2.42e+00	8.95e-02	-27.06	< 2e-16 ***
## genrenew-age	-1.75e-01	8.06e-02	-2.17	0.02998 *
## genreopera	-2.98e+00	8.52e-02	-34.98	< 2e-16 ***
## genreparty	-2.86e+00	9.82e-02	-29.15	< 2e-16 ***
## genrepiano	4.70e-01	8.74e-02	5.38	7.6e-08 ***
## genrepop	2.99e+00	1.16e-01	25.79	< 2e-16 ***
## genrepop-film	-1.34e+00	8.39e-02	-16.03	< 2e-16 ***
## genrepower-pop	-3.41e+00	8.31e-02	-41.07	< 2e-16 ***
## genreprogressive-house	-1.22e+00	9.88e-02	-12.31	< 2e-16 ***
## genrepsych-rock	4.28e-01	8.89e-02	4.82	1.5e-06 ***
## genrepunk	1.69e+00	1.19e-01	14.26	< 2e-16 ***
## genrepunk-rock	8.56e-01	1.08e-01	7.93	2.2e-15 ***
## genrerock	2.56e+00	1.54e-01	16.57	< 2e-16 ***

```

## genrerock-n-roll      -1.88e+00  8.50e-02 -22.11 < 2e-16 ***
## genreromance        -6.85e+00  1.14e-01 -59.93 < 2e-16 ***
## genresad             8.71e-01  1.18e-01   7.35 2.0e-13 ***
## genresalsa           -8.69e-01  8.37e-02 -10.38 < 2e-16 ***
## genresamba           -2.10e+00  8.16e-02 -25.69 < 2e-16 ***
## genresertanejo       -1.99e-01  8.27e-02  -2.40  0.01626 *
## genreshow-tunes      -2.46e+00  9.45e-02 -26.06 < 2e-16 ***
## genresinger-songwriter 1.09e+00  8.59e-02  12.71 < 2e-16 ***
## genreska              1.68e-01  8.89e-02   1.89  0.05878 .
## genresleep            -1.17e+00  8.85e-02 -13.20 < 2e-16 ***
## genresongwriter       1.46e+00  3.44e-01   4.24  2.2e-05 ***
## genresoul              1.60e+00  1.02e-01  15.70 < 2e-16 ***
## genrespanish          9.01e-01  8.02e-02  11.23 < 2e-16 ***
## genreswedish          8.14e-01  9.30e-02   8.76 < 2e-16 ***
## genretango             -4.85e+00  8.53e-02 -56.83 < 2e-16 ***
## genretechno           -9.74e-01  1.10e-01  -8.86 < 2e-16 ***
## genretrance            -5.45e-01  9.97e-02  -5.47  4.6e-08 ***
## genretrip-hop          -2.65e+00  9.29e-02 -28.47 < 2e-16 ***
## upper_genreblues       NA      NA      NA      NA
## upper_genreclassical/opera  NA      NA      NA      NA
## upper_genrecountry      NA      NA      NA      NA
## upper_genreelectronic    NA      NA      NA      NA
## upper_genrefolk          NA      NA      NA      NA
## upper_genrefunk/disco     NA      NA      NA      NA
## upper_genrehip-hop       NA      NA      NA      NA
## upper_genrejazz          NA      NA      NA      NA
## upper_genremetal         NA      NA      NA      NA
## upper_genrenew-age       NA      NA      NA      NA
## upper_genreother         NA      NA      NA      NA
## upper_genrepop           NA      NA      NA      NA
## upper_genrepunk          NA      NA      NA      NA
## upper_genrerock          NA      NA      NA      NA
## upper_genreska           NA      NA      NA      NA
## upper_genresoul          NA      NA      NA      NA
## upper_genretrip-hop      NA      NA      NA      NA
## upper_genreworld         NA      NA      NA      NA
## key1                     4.08e-02  3.28e-02  1.24  0.21416
## key2                     3.51e-02  3.17e-02  1.11  0.26786
## key3                     4.42e-02  4.72e-02  0.94  0.34913
## key4                     8.74e-04  3.49e-02  0.03  0.97998
## key5                     3.08e-02  3.45e-02  0.89  0.37135
## key6                     8.49e-02  3.68e-02  2.31  0.02109 *
## key7                     -3.81e-02 3.07e-02 -1.24  0.21483
## key8                     1.17e-01  3.73e-02  3.14  0.00166 **
## key9                     -2.11e-02 3.21e-02 -0.66  0.51163
## key10                    1.97e-02  3.69e-02  0.53  0.59297
## key11                    3.01e-02  3.53e-02  0.85  0.39285
## modeminor                9.99e-02  1.64e-02  6.10  1.0e-09 ***
## time_signature1          1.20e+00  2.77e-01  4.32  1.6e-05 ***
## time_signature3          1.26e+00  2.69e-01  4.69  2.7e-06 ***
## time_signature4          1.40e+00  2.69e-01  5.20  2.0e-07 ***
## time_signature5          1.24e+00  2.74e-01  4.52  6.1e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

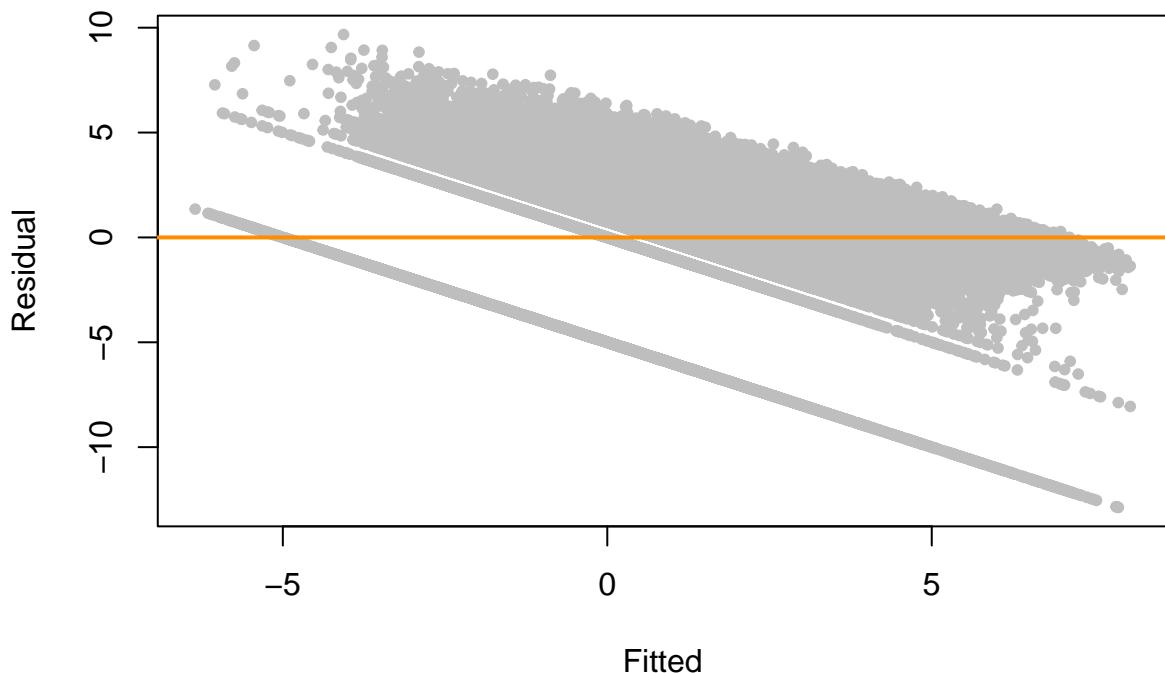
```

```

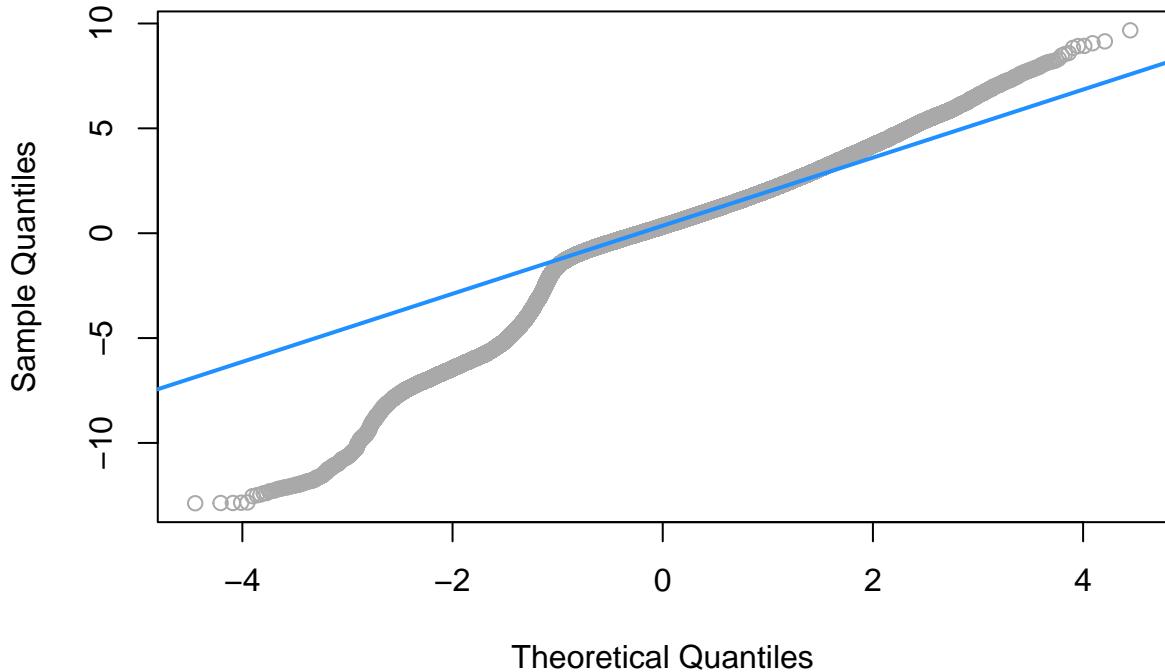
##
## Residual standard error: 2.51 on 115796 degrees of freedom
## Multiple R-squared:  0.438, Adjusted R-squared:  0.437
## F-statistic:  834 on 108 and 115796 DF, p-value: <2e-16
##
##      shapiro bptest high_leverage_count outliers_count influence_count
## BP 2.525e-46      0          3433         9184        8890
##      loocv_rmse adjusted_r_squared
## BP      2.514           0.437

```

Full Additive Model – Fitted vs Residuals



Normal Q-Q Plot



```
print_checks(model_trans_genre_add)
```

```
##  
## Call:  
## lm(formula = popularity ~ . - upper_genre, data = data)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max  
## -12.876  -0.736   0.330   1.453   9.673  
##  
## Coefficients:  
##                               Estimate Std. Error t value Pr(>|t|)  
## (Intercept)              -2.86e+02  2.28e+00 -125.34 < 2e-16 ***  
## danceability             1.63e-02  6.53e-02    0.25  0.80251  
## energy                   2.20e-01  6.77e-02    3.25  0.00115 **  
## loudness                 5.02e-02  2.59e-03   19.40 < 2e-16 ***  
## speechiness              -1.27e-01  1.61e-02   -7.91  2.6e-15 ***  
## acousticness              7.42e-02  8.71e-03    8.52 < 2e-16 ***  
## instrumentalness         -1.33e-02  1.45e-03   -9.17 < 2e-16 ***  
## liveness                  5.73e-02  1.26e-02    4.56  5.1e-06 ***  
## valence                  -5.07e-01  2.69e-02   -18.84 < 2e-16 ***  
## tempo                     5.89e-04  1.14e-03    0.52  0.60394  
## duration_m                -2.37e-01  1.48e-02   -16.05 < 2e-16 ***  
## year                      1.43e-01  1.12e-03   127.44 < 2e-16 ***  
## genreafrobeat             -2.74e+00  9.19e-02   -29.84 < 2e-16 ***  
## genrealt-rock              2.21e+00  7.95e-02    27.76 < 2e-16 ***  
## genreambient               7.75e-01  8.11e-02     9.55 < 2e-16 ***  
## genreblack-metal           -8.72e-01  8.27e-02   -10.55 < 2e-16 ***  
## genreblues                 1.15e+00  7.99e-02    14.40 < 2e-16 ***  
## genrebroadcast              -4.70e+00  9.32e-02   -50.46 < 2e-16 ***
```

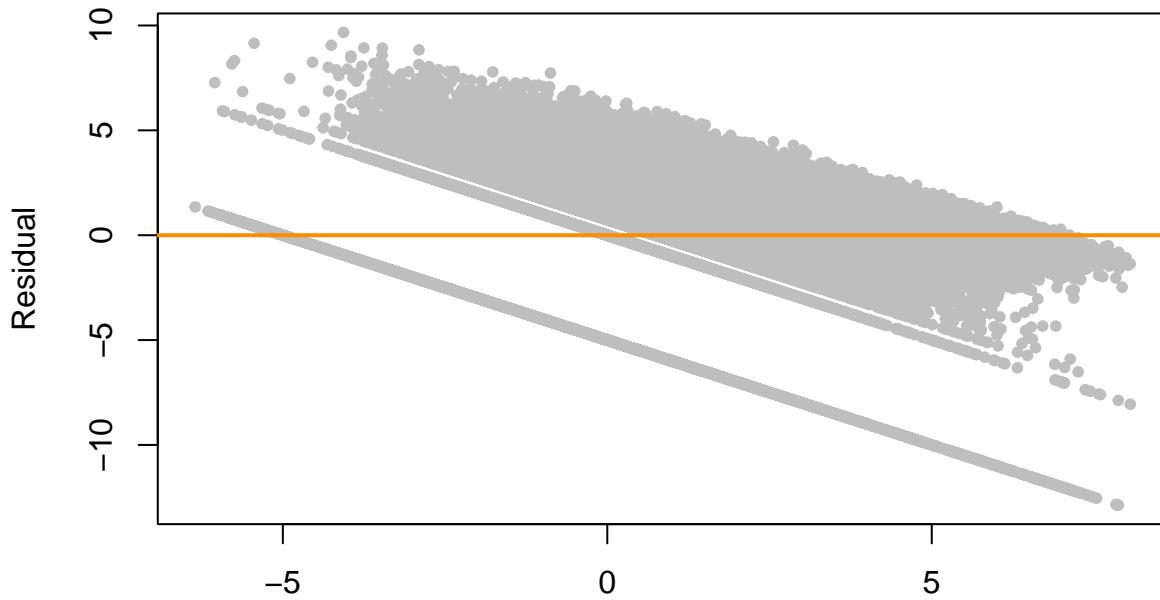
## genrecantopop	-6.90e-01	8.51e-02	-8.10	5.4e-16	***
## genrechicago-house	-4.99e+00	1.28e-01	-39.02	< 2e-16	***
## genrechill	2.54e-01	8.04e-02	3.16	0.00155	**
## genreclassical	1.98e+00	8.33e-02	23.75	< 2e-16	***
## genreclub	-2.49e+00	8.64e-02	-28.83	< 2e-16	***
## genrecomedy	-1.19e+00	8.68e-02	-13.68	< 2e-16	***
## genrecountry	1.88e+00	8.17e-02	23.00	< 2e-16	***
## genredance	2.38e+00	8.27e-02	28.77	< 2e-16	***
## genredancehall	-6.46e-01	8.22e-02	-7.86	3.8e-15	***
## genredeath-metal	6.55e-01	8.59e-02	7.62	2.5e-14	***
## genredeep-house	-9.21e-01	8.51e-02	-10.83	< 2e-16	***
## genredetroit-techno	-4.89e+00	1.38e-01	-35.34	< 2e-16	***
## genredisco	3.32e-01	8.45e-02	3.92	8.7e-05	***
## genredrum-and-bass	-1.73e+00	9.00e-02	-19.24	< 2e-16	***
## genredub	-3.47e-01	8.21e-02	-4.22	2.4e-05	***
## genredubstep	-4.34e+00	1.28e-01	-33.86	< 2e-16	***
## genreedm	-6.00e-02	9.61e-02	-0.62	0.53246	
## genreelectro	2.08e+00	9.37e-02	22.15	< 2e-16	***
## genreelectronic	9.95e-01	1.01e-01	9.87	< 2e-16	***
## genreemo	6.66e-01	7.93e-02	8.39	< 2e-16	***
## genrefolk	1.92e+00	8.28e-02	23.16	< 2e-16	***
## genreforro	-2.26e+00	8.21e-02	-27.52	< 2e-16	***
## genrefrench	1.38e+00	8.17e-02	16.92	< 2e-16	***
## genrefunk	1.38e+00	8.68e-02	15.91	< 2e-16	***
## genregarage	-2.72e-01	8.28e-02	-3.29	0.00101	**
## genregerman	1.36e+00	8.30e-02	16.42	< 2e-16	***
## genregospel	6.49e-01	7.83e-02	8.29	< 2e-16	***
## genregothing	-8.28e-01	8.32e-02	-9.95	< 2e-16	***
## genregrindcore	-5.04e+00	9.38e-02	-53.71	< 2e-16	***
## genregroove	-1.61e+00	8.88e-02	-18.18	< 2e-16	***
## genreguitar	-1.12e+00	8.32e-02	-13.43	< 2e-16	***
## genrehard-rock	8.32e-01	8.98e-02	9.27	< 2e-16	***
## genrehardcore	1.38e+00	8.73e-02	15.83	< 2e-16	***
## genrehardstyle	-1.95e+00	9.28e-02	-21.05	< 2e-16	***
## genreheavy-metal	-4.55e+00	8.90e-02	-51.11	< 2e-16	***
## genrehip-hop	2.46e+00	8.69e-02	28.27	< 2e-16	***
## genrehouse	1.03e+00	1.25e-01	8.23	< 2e-16	***
## genreindian	-3.25e+00	7.91e-02	-41.01	< 2e-16	***
## genreindie-pop	1.95e+00	9.57e-02	20.37	< 2e-16	***
## genreindustrial	-5.94e-01	8.76e-02	-6.78	1.2e-11	***
## genrejazz	1.92e+00	8.53e-02	22.49	< 2e-16	***
## genrek-pop	1.17e+00	7.92e-02	14.83	< 2e-16	***
## genremetal	1.92e+00	1.13e-01	16.96	< 2e-16	***
## genremetalcore	-4.33e-01	1.17e-01	-3.69	0.00022	***
## genreminimal-techno	-2.42e+00	8.95e-02	-27.06	< 2e-16	***
## genrenew-age	-1.75e-01	8.06e-02	-2.17	0.02998	*
## genreopera	-2.98e+00	8.52e-02	-34.98	< 2e-16	***
## genreparty	-2.86e+00	9.82e-02	-29.15	< 2e-16	***
## genrepiano	4.70e-01	8.74e-02	5.38	7.6e-08	***
## genrepop	2.99e+00	1.16e-01	25.79	< 2e-16	***
## genrepop-film	-1.34e+00	8.39e-02	-16.03	< 2e-16	***
## genrepower-pop	-3.41e+00	8.31e-02	-41.07	< 2e-16	***
## genreprogressive-house	-1.22e+00	9.88e-02	-12.31	< 2e-16	***
## genrepsych-rock	4.28e-01	8.89e-02	4.82	1.5e-06	***

```

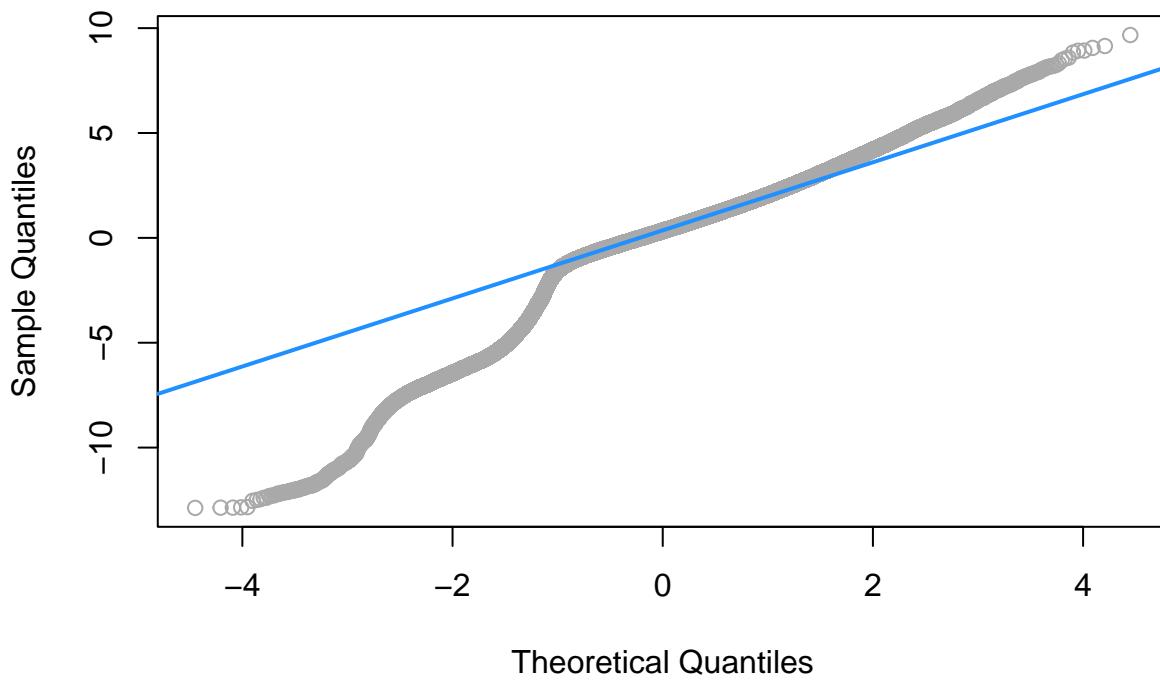
## genrepunk      1.69e+00  1.19e-01  14.26 < 2e-16 ***
## genrepunk-rock 8.56e-01  1.08e-01   7.93 2.2e-15 ***
## genrerock      2.56e+00  1.54e-01  16.57 < 2e-16 ***
## genrerock-n-roll -1.88e+00 8.50e-02 -22.11 < 2e-16 ***
## genreromance    -6.85e+00 1.14e-01 -59.93 < 2e-16 ***
## genresad        8.71e-01  1.18e-01   7.35 2.0e-13 ***
## genresalsa      -8.69e-01 8.37e-02 -10.38 < 2e-16 ***
## genresamba      -2.10e+00 8.16e-02 -25.69 < 2e-16 ***
## genresertanejo -1.99e-01 8.27e-02  -2.40 0.01626 *
## genreshow-tunes -2.46e+00 9.45e-02 -26.06 < 2e-16 ***
## genresinger-songwriter 1.09e+00 8.59e-02 12.71 < 2e-16 ***
## genreska         1.68e-01  8.89e-02   1.89 0.05878 .
## genresleep       -1.17e+00 8.85e-02 -13.20 < 2e-16 ***
## genresongwriter 1.46e+00  3.44e-01   4.24 2.2e-05 ***
## genresoul        1.60e+00  1.02e-01  15.70 < 2e-16 ***
## genrespanish     9.01e-01  8.02e-02  11.23 < 2e-16 ***
## genreswedish     8.14e-01  9.30e-02   8.76 < 2e-16 ***
## genretango       -4.85e+00 8.53e-02 -56.83 < 2e-16 ***
## genretechno      -9.74e-01 1.10e-01  -8.86 < 2e-16 ***
## genretrance      -5.45e-01 9.97e-02  -5.47 4.6e-08 ***
## genretrip-hop   -2.65e+00 9.29e-02 -28.47 < 2e-16 ***
## key1            4.08e-02  3.28e-02   1.24 0.21416
## key2            3.51e-02  3.17e-02   1.11 0.26786
## key3            4.42e-02  4.72e-02   0.94 0.34913
## key4            8.74e-04  3.49e-02   0.03 0.97998
## key5            3.08e-02  3.45e-02   0.89 0.37135
## key6            8.49e-02  3.68e-02   2.31 0.02109 *
## key7            -3.81e-02 3.07e-02  -1.24 0.21483
## key8            1.17e-01  3.73e-02   3.14 0.00166 **
## key9            -2.11e-02 3.21e-02  -0.66 0.51163
## key10           1.97e-02  3.69e-02   0.53 0.59297
## key11           3.01e-02  3.53e-02   0.85 0.39285
## modeminor       9.99e-02  1.64e-02   6.10 1.0e-09 ***
## time_signature1 1.20e+00  2.77e-01   4.32 1.6e-05 ***
## time_signature3 1.26e+00  2.69e-01   4.69 2.7e-06 ***
## time_signature4 1.40e+00  2.69e-01   5.20 2.0e-07 ***
## time_signature5 1.24e+00  2.74e-01   4.52 6.1e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.51 on 115796 degrees of freedom
## Multiple R-squared:  0.438, Adjusted R-squared:  0.437
## F-statistic:  834 on 108 and 115796 DF, p-value: <2e-16
##
##      shapiro bptest high_leverage_count outliers_count influence_count
## BP 3.666e-46      0          3433         9184        8890
##      loocv_rmse adjusted_r_squared
## BP      2.514          0.437

```

Full Additive Model – Fitted vs Residuals



Fitted
Normal Q–Q Plot



```
print_checks(model_trans_upper_genre_add)
```

```
##  
## Call:  
## lm(formula = popularity ~ . - genre, data = data)  
##
```

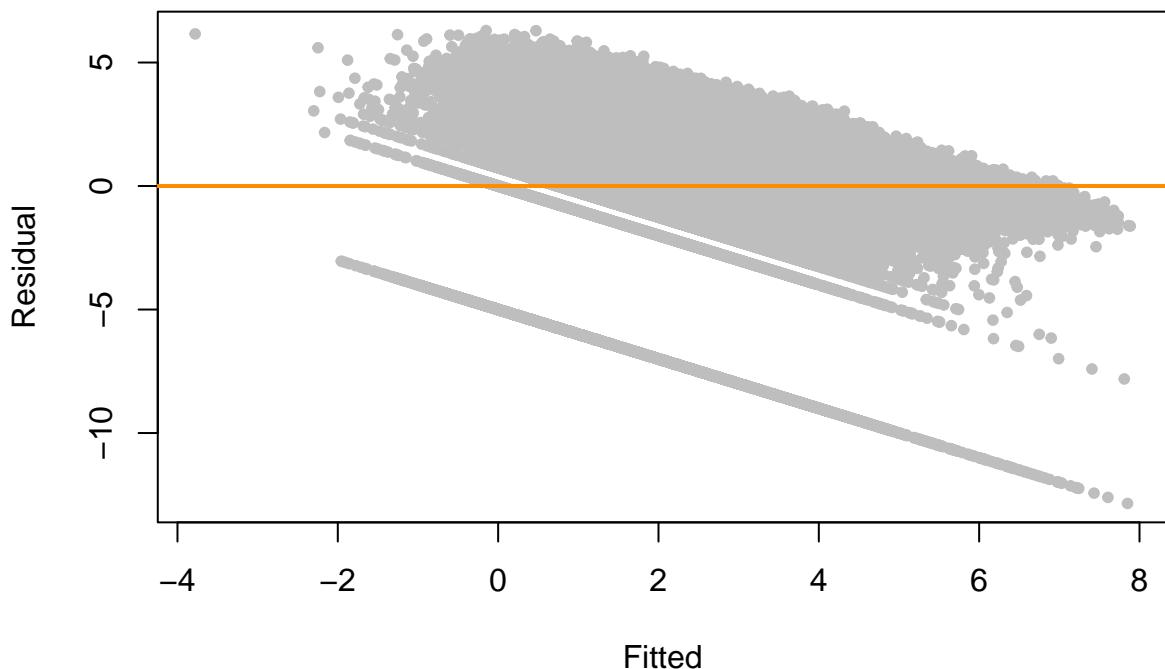
```

## Residuals:
##      Min     1Q Median     3Q    Max
## -12.851 -0.705  0.763  1.965  6.294
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -263.48106   2.72660 -96.63 < 2e-16 ***
## danceability              0.36787   0.07372   4.99 6.0e-07 ***
## energy                   0.59772   0.07706   7.76 8.8e-15 ***
## loudness                  0.02943   0.00281  10.47 < 2e-16 ***
## speechiness                -0.21740   0.01744 -12.46 < 2e-16 ***
## acousticness                 0.08370   0.00992   8.44 < 2e-16 ***
## instrumentalness            -0.05863   0.00159 -36.83 < 2e-16 ***
## liveness                   -0.01915   0.01490   -1.28 0.19882
## valence                     -0.96755   0.03115 -31.06 < 2e-16 ***
## tempo                      -0.00360   0.00135   -2.67 0.00758 **
## duration_m                  -0.20329   0.01660 -12.25 < 2e-16 ***
## year                        0.13068   0.00134   97.34 < 2e-16 ***
## upper_genreblues             1.14173   0.07823  14.59 < 2e-16 ***
## upper_genreclassical/opera   -0.46445   0.06419   -7.24 4.7e-13 ***
## upper_genrecountry              1.56895   0.08131  19.30 < 2e-16 ***
## upper_genreelectronic            -0.90675   0.04319 -21.00 < 2e-16 ***
## upper_genrefolk                  1.77009   0.08263  21.42 < 2e-16 ***
## upper_genrefunk/disco              0.74624   0.06677  11.18 < 2e-16 ***
## upper_genrehip-hop                  2.12415   0.08793  24.16 < 2e-16 ***
## upper_genrejazz                  1.93277   0.08575  22.54 < 2e-16 ***
## upper_genremetal                 -0.89014   0.05061 -17.59 < 2e-16 ***
## upper_genrenew-age                 -0.24714   0.07736   -3.19 0.00140 **
## upper_genreother                  -1.63961   0.04970 -32.99 < 2e-16 ***
## upper_genrepop                  -0.60167   0.04969 -12.11 < 2e-16 ***
## upper_genrepunk                  1.12338   0.09216  12.19 < 2e-16 ***
## upper_genrerock                  0.35932   0.05059   7.10 1.2e-12 ***
## upper_genreska                  0.10651   0.09124   1.17 0.24309
## upper_genresoul                  1.36901   0.10934  12.52 < 2e-16 ***
## upper_genretrip-hop                -2.59385   0.09629 -26.94 < 2e-16 ***
## upper_genreworld                  -1.24051   0.04328 -28.66 < 2e-16 ***
## key1                         0.04258   0.03955   1.08 0.28174
## key2                         -0.03673   0.03822   -0.96 0.33661
## key3                         0.10296   0.05700   1.81 0.07090 .
## key4                         -0.07031   0.04204   -1.67 0.09447 .
## key5                         0.03899   0.04164   0.94 0.34899
## key6                         0.14718   0.04440   3.31 0.00092 ***
## key7                         -0.12404   0.03706   -3.35 0.00082 ***
## key8                         0.18549   0.04503   4.12 3.8e-05 ***
## key9                         -0.08041   0.03873   -2.08 0.03791 *
## key10                        0.00381   0.04448   0.09 0.93175
## key11                        0.01590   0.04252   0.37 0.70838
## modeminor                     0.13613   0.01957   6.96 3.5e-12 ***
## time_signature1                  2.83626   0.32841  8.64 < 2e-16 ***
## time_signature3                  3.01697   0.31903  9.46 < 2e-16 ***
## time_signature4                  3.22504   0.31760  10.15 < 2e-16 ***
## time_signature5                  3.03891   0.32526  9.34 < 2e-16 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
##  
## Residual standard error: 3.03 on 115859 degrees of freedom  
## Multiple R-squared:  0.179, Adjusted R-squared:  0.179  
## F-statistic:  562 on 45 and 115859 DF, p-value: <2e-16  
##  
##      shapiro bptest high_leverage_count outliers_count influence_count  
## BP 3.971e-54      0          10820         10258          5863  
## loocv_rmse adjusted_r_squared  
## BP      3.035          0.179
```

Full Additive Model – Fitted vs Residuals



Normal Q-Q Plot

