

MA678 Final Project

Jiahao Liu

2022-11-18

```
library(tidyverse)

## — Attaching packages — tidyverse
1.3.2 —
## ✓ ggplot2 3.4.0      ✓ purrr  0.3.5
## ✓ tibble  3.1.8      ✓ dplyr  1.0.10
## ✓ tidyr   1.2.1      ✓ stringr 1.4.1
## ✓ readr   2.1.3      ✓ forcats 0.5.2

## Warning: package 'ggplot2' was built under R version 4.2.2
## Warning: package 'tidyr' was built under R version 4.2.2
## Warning: package 'readr' was built under R version 4.2.2
## Warning: package 'purrr' was built under R version 4.2.2

## — Conflicts —
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()

library(dplyr)
library(ggplot2)
library(grid)
library(gridExtra)

##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##   combine

#Library(rstanarm)
library(lmerTest)

## Warning: package 'lmerTest' was built under R version 4.2.2
## Loading required package: lme4
## Warning: package 'lme4' was built under R version 4.2.2
## Loading required package: Matrix
```

```
## Warning: package 'Matrix' was built under R version 4.2.2

##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
##
## Attaching package: 'lmerTest'
##
## The following object is masked from 'package:lme4':
##
##     lmer
##
## The following object is masked from 'package:stats':
##
##     step

library(lme4)
library(merTools)

## Warning: package 'merTools' was built under R version 4.2.2

## Loading required package: arm
## Loading required package: MASS
##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##     select
##
##
## arm (Version 1.13-1, built: 2022-8-25)
##
## Working directory is C:/Users/Jiahao Liu/Desktop

library(correlation)

## Warning: package 'correlation' was built under R version 4.2.2

##
## Attaching package: 'correlation'
##
## The following object is masked from 'package:arm':
##
##     display

library(corrplot)
```

```
## corrrplot 0.92 loaded
##
## Attaching package: 'corrrplot'
##
## The following object is masked from 'package:arm':
##
##     corrrplot
library("PerformanceAnalytics")
## Loading required package: xts
## Warning: package 'xts' was built under R version 4.2.2
## Loading required package: zoo
## Warning: package 'zoo' was built under R version 4.2.2
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
##
## Attaching package: 'xts'
##
## The following objects are masked from 'package:dplyr':
##
##     first, last
##
## Attaching package: 'PerformanceAnalytics'
##
## The following object is masked from 'package:graphics':
##
##     legend
```

Data Processing

```
## song_name      song_popularity song_duration_ms  acousticness
## Length:13070   Min.    : 0.00   Min.    : 12000   Min.    :0.000001
## Class :character 1st Qu.: 37.00   1st Qu.: 182967   1st Qu.:0.025000
## Mode  :character Median : 51.00   Median : 211486   Median :0.147000
##                Mean   : 48.49   Mean   : 218627   Mean   :0.277770
##                3rd Qu.: 63.00   3rd Qu.: 244506   3rd Qu.:0.479750
##                Max.   :100.00   Max.   :1799346   Max.   :0.996000
## danceability    energy      instrumentalness    key
## Min.    :0.0000   Min.    :0.00107   Min.    :0.000000   Min.    : 0.00
## 1st Qu.:0.5242   1st Qu.:0.49000   1st Qu.:0.000000   1st Qu.: 2.00
```

```
## Median :0.6370 Median :0.66700 Median :0.000022 Median : 5.00
## Mean :0.6250 Mean :0.63572 Mean :0.096672 Mean : 5.32
## 3rd Qu.:0.7410 3rd Qu.:0.81500 3rd Qu.:0.005910 3rd Qu.: 8.00
## Max. :0.9870 Max. :0.99900 Max. :0.997000 Max. :11.00
## liveness loudness audio_mode speechiness
## Min. :0.0119 Min. : -38.768 Min. :0.000 Min. :0.0000
## 1st Qu.:0.0933 1st Qu.: -9.537 1st Qu.:0.000 1st Qu.:0.0373
## Median :0.1210 Median : -6.859 Median :1.000 Median :0.0544
## Mean :0.1805 Mean : -7.790 Mean :0.633 Mean :0.1008
## 3rd Qu.:0.2230 3rd Qu.: -5.041 3rd Qu.:1.000 3rd Qu.:0.1150
## Max. :0.9860 Max. : 1.585 Max. :1.000 Max. :0.9410
## tempo time_signature audio_valence
## Min. : 0.00 Min. :0.000 Min. :0.0000
## 1st Qu.: 98.07 1st Qu.:4.000 1st Qu.:0.3340
## Median :120.03 Median :4.000 Median :0.5290
## Mean :121.15 Mean :3.952 Mean :0.5286
## 3rd Qu.:139.96 3rd Qu.:4.000 3rd Qu.:0.7300
## Max. :242.32 Max. :5.000 Max. :0.9840
```

Variable Selection

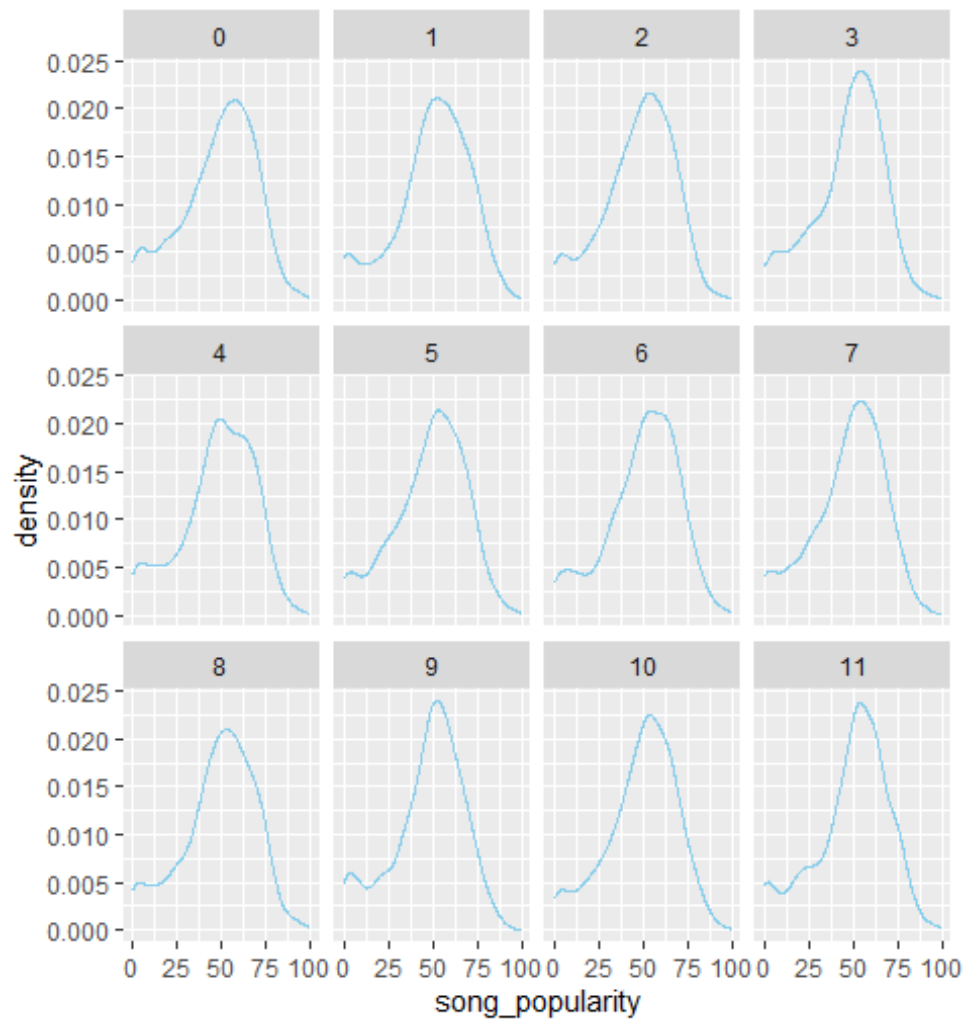
```
## # A tibble: 12 × 2
##   key      n
##   <dbl> <int>
## 1     0 1484
## 2     1 1404
## 3     2 1202
## 4     3  394
## 5     4  954
## 6     5 1113
## 7     6  911
## 8     7 1457
## 9     8  915
## 10    9 1252
## 11   10  922
## 12   11 1062

## # A tibble: 5 × 2
##   time_signature      n
##   <dbl> <int>
## 1     0      3
## 2     1     57
## 3     3    626
## 4     4 12208
## 5     5    176

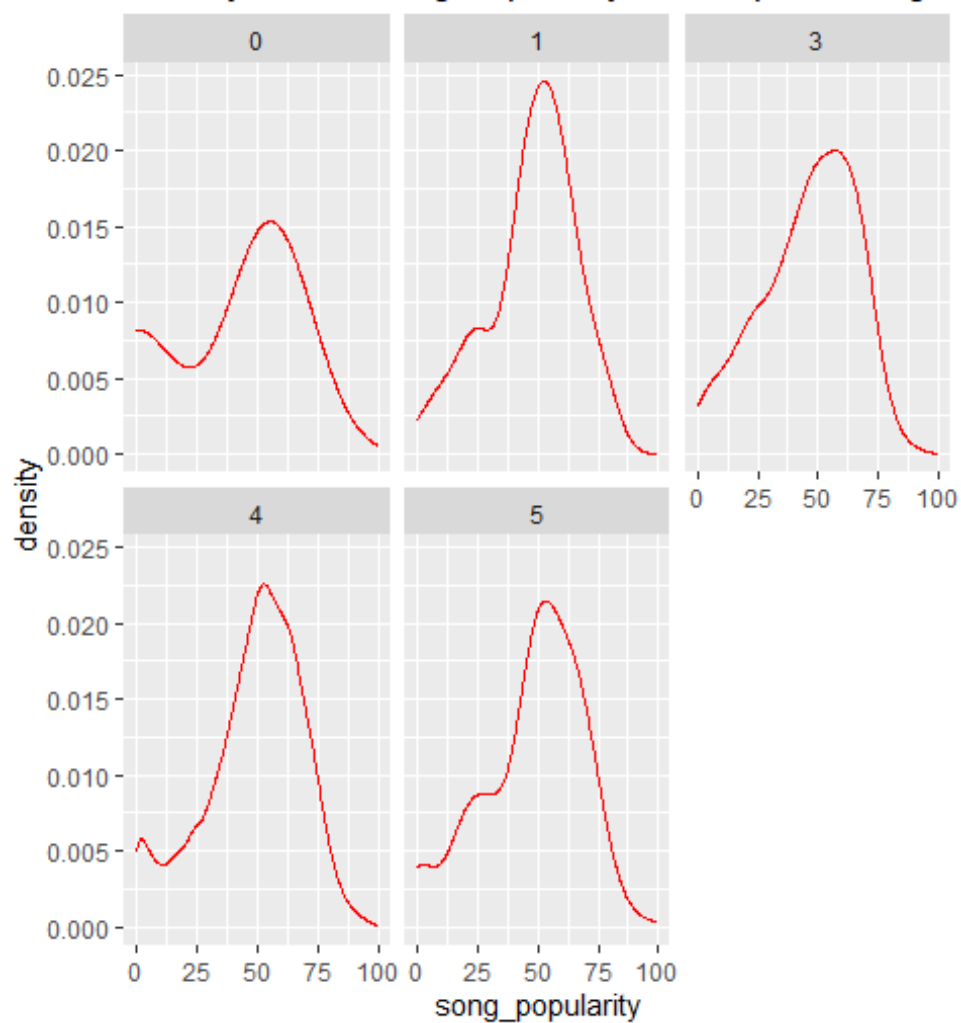
## # A tibble: 2 × 2
##   audio_mode      n
##   <dbl> <int>
```

## 1	0	4797
## 2	1	8273

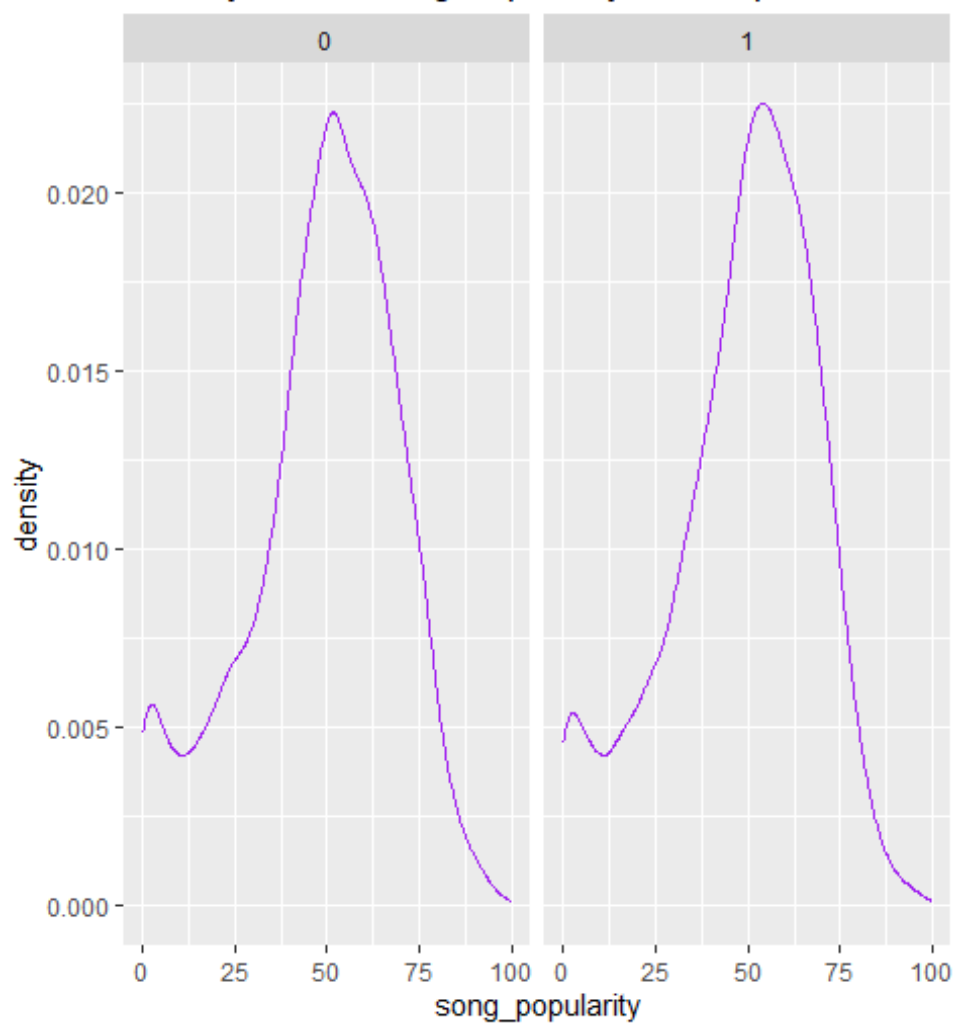
Density Plot of Song Popularity in Group Key



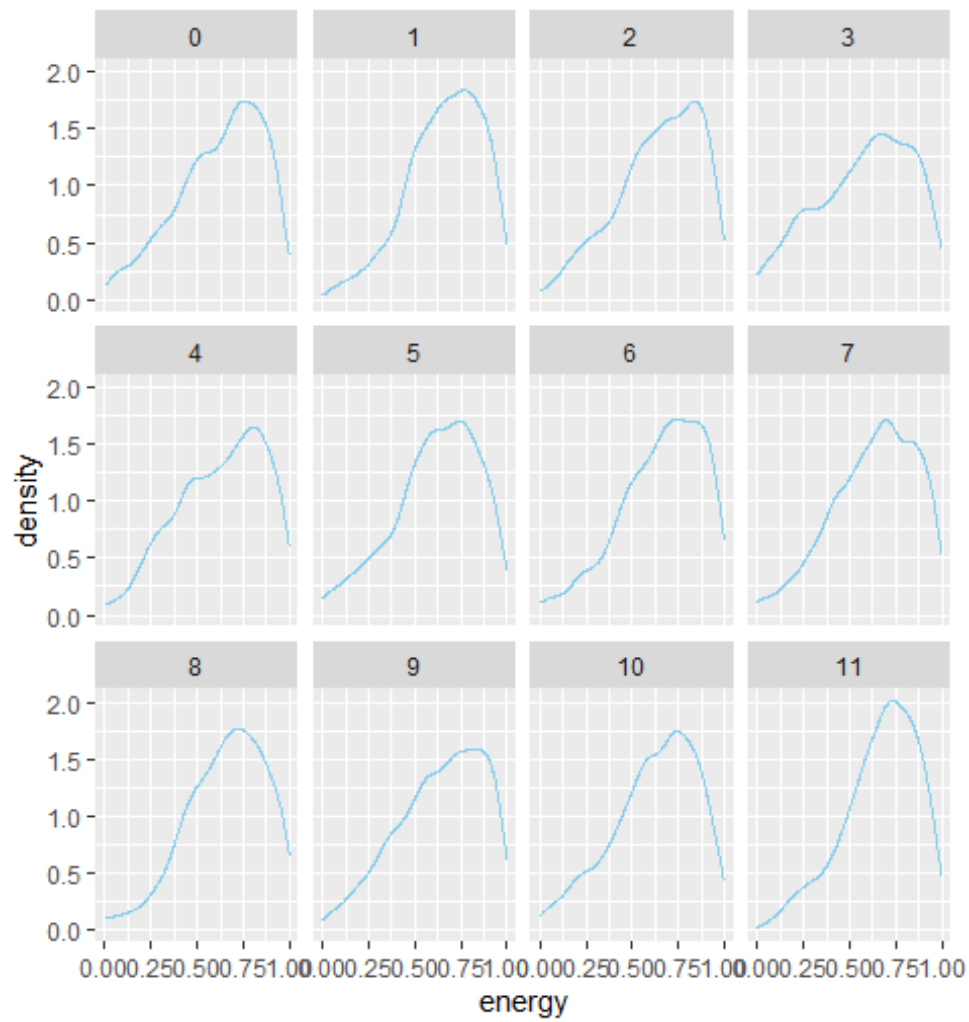
Density Plot of Song Popularity in Group Time Signa



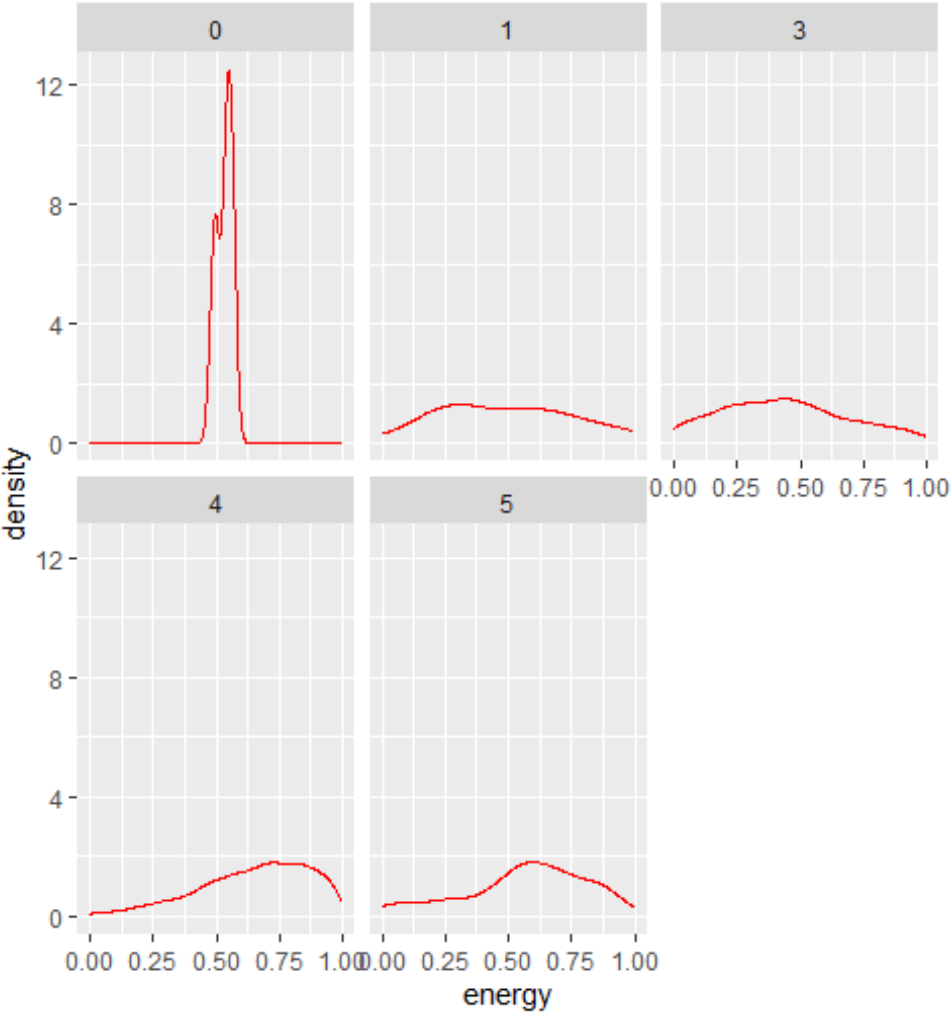
Density Plot of Song Popularity in Group Audio Mode



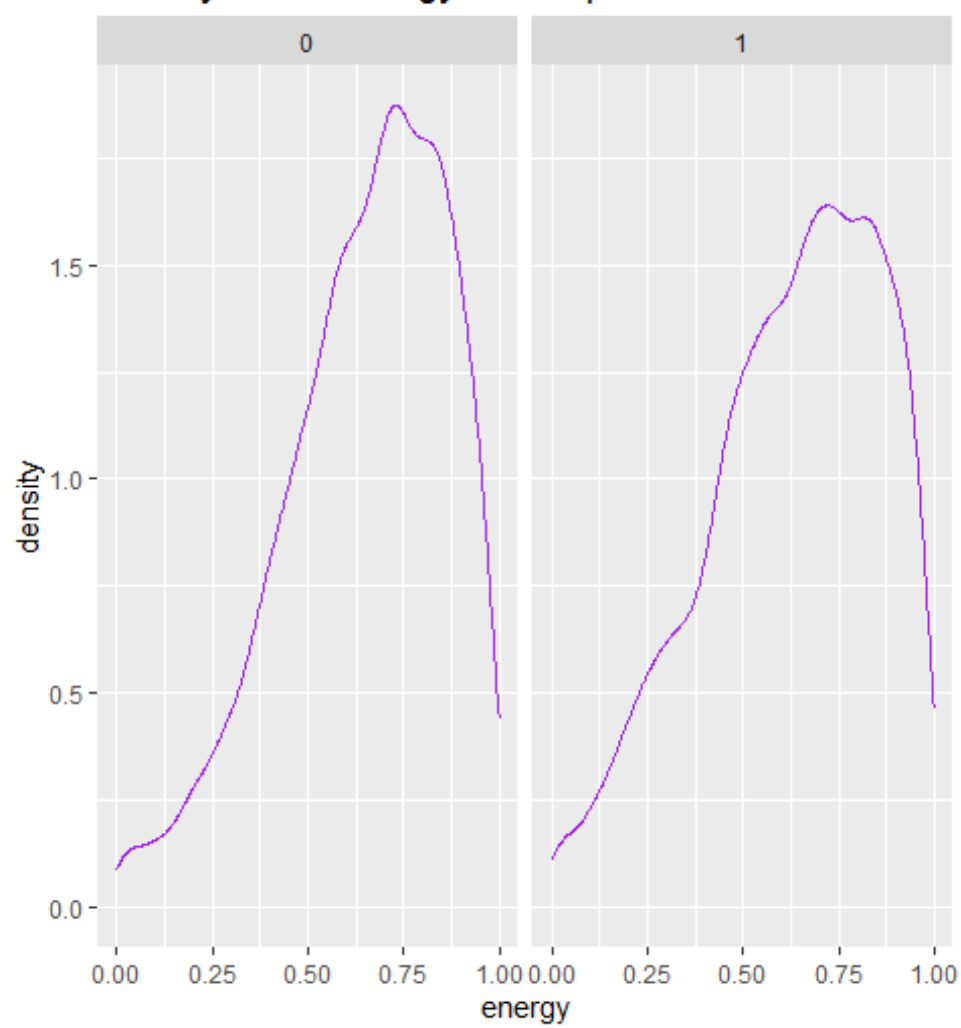
Density Plot of Energy in Group Key



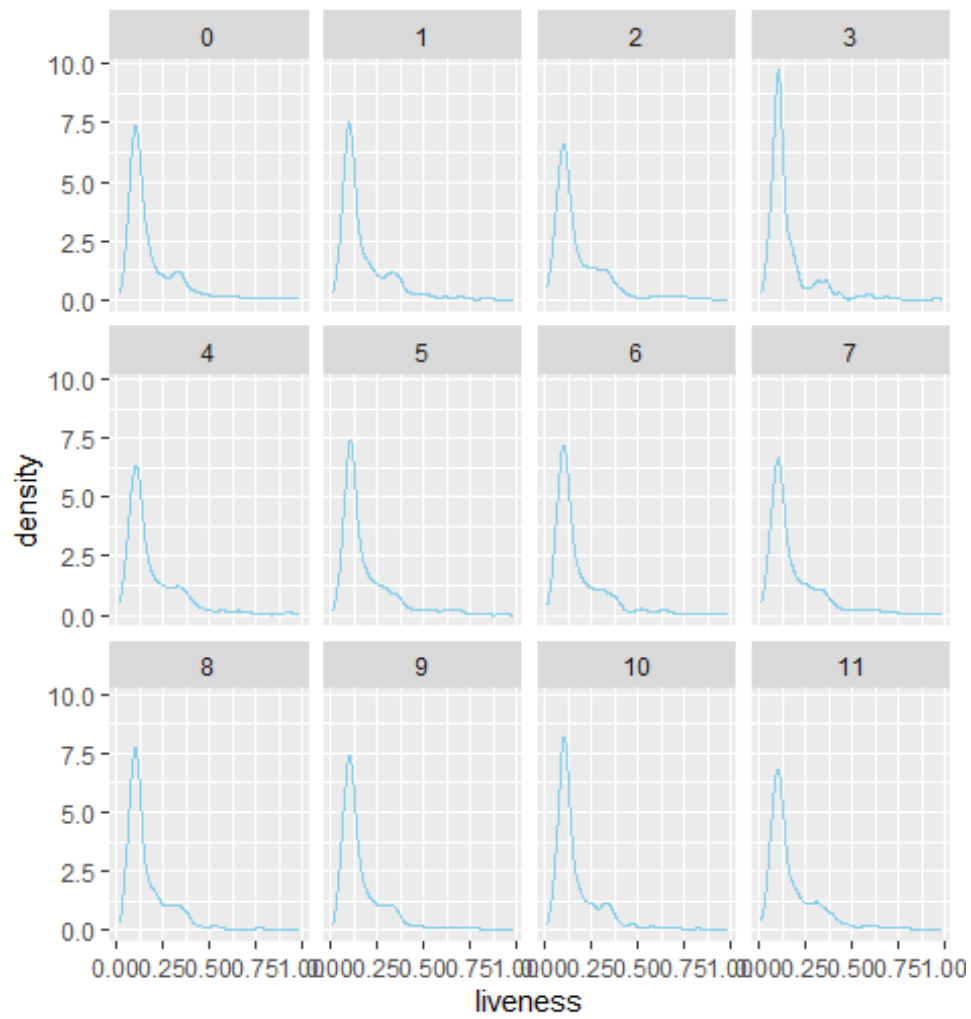
Density Plot of Energy in Group Time Signature



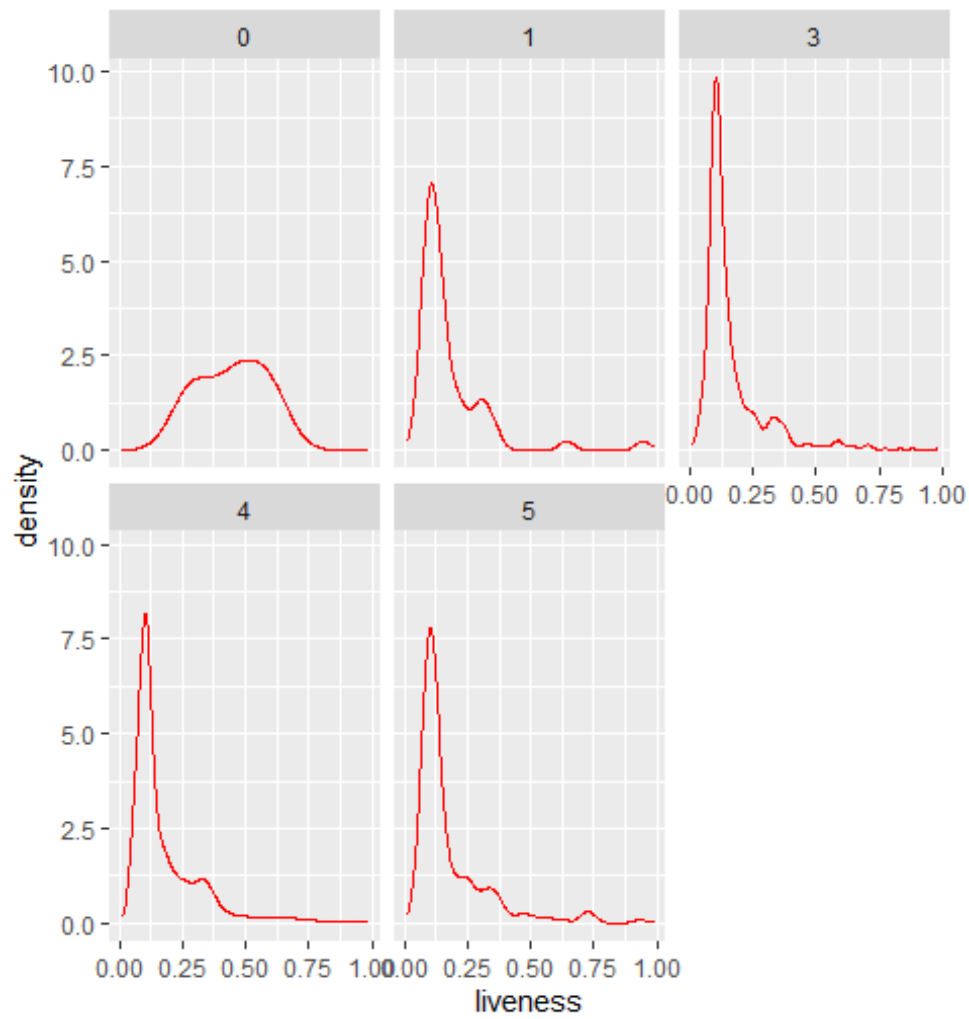
Density Plot of Energy in Group Audio Mode



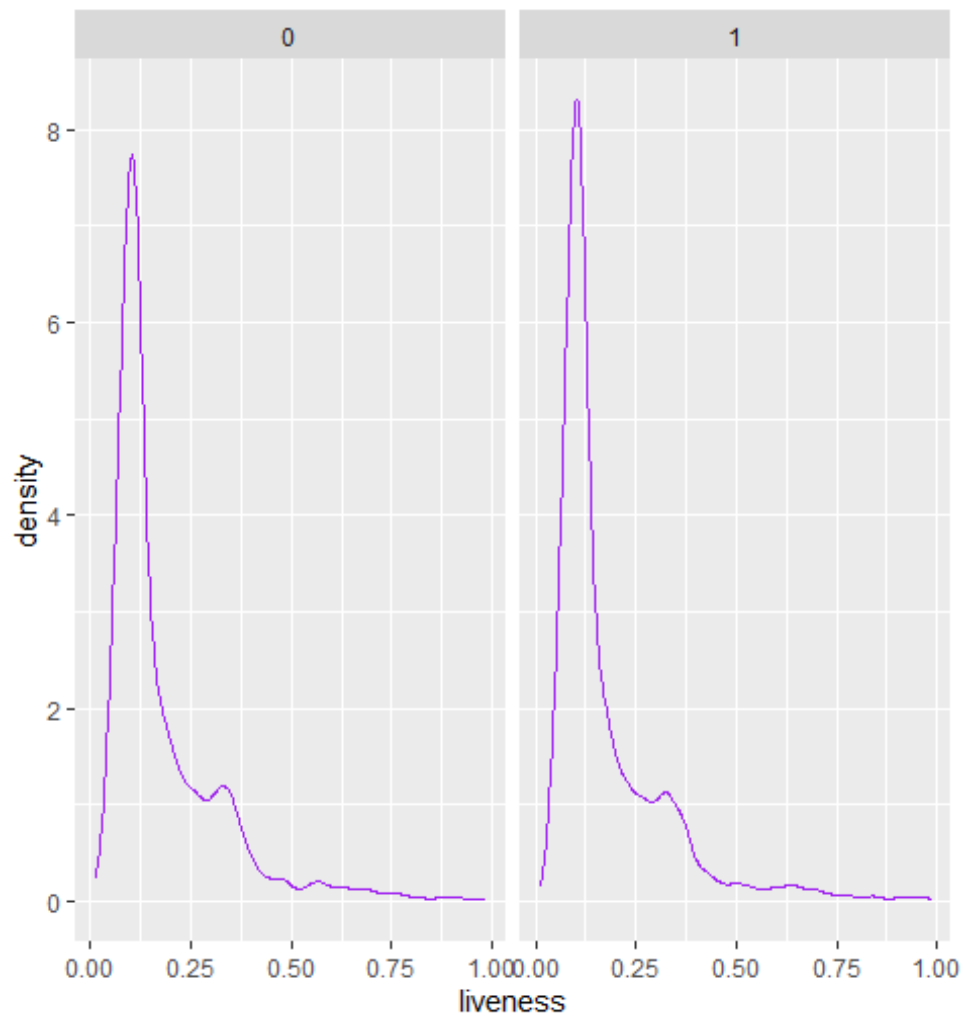
Density Plot of Liveness in Group Key



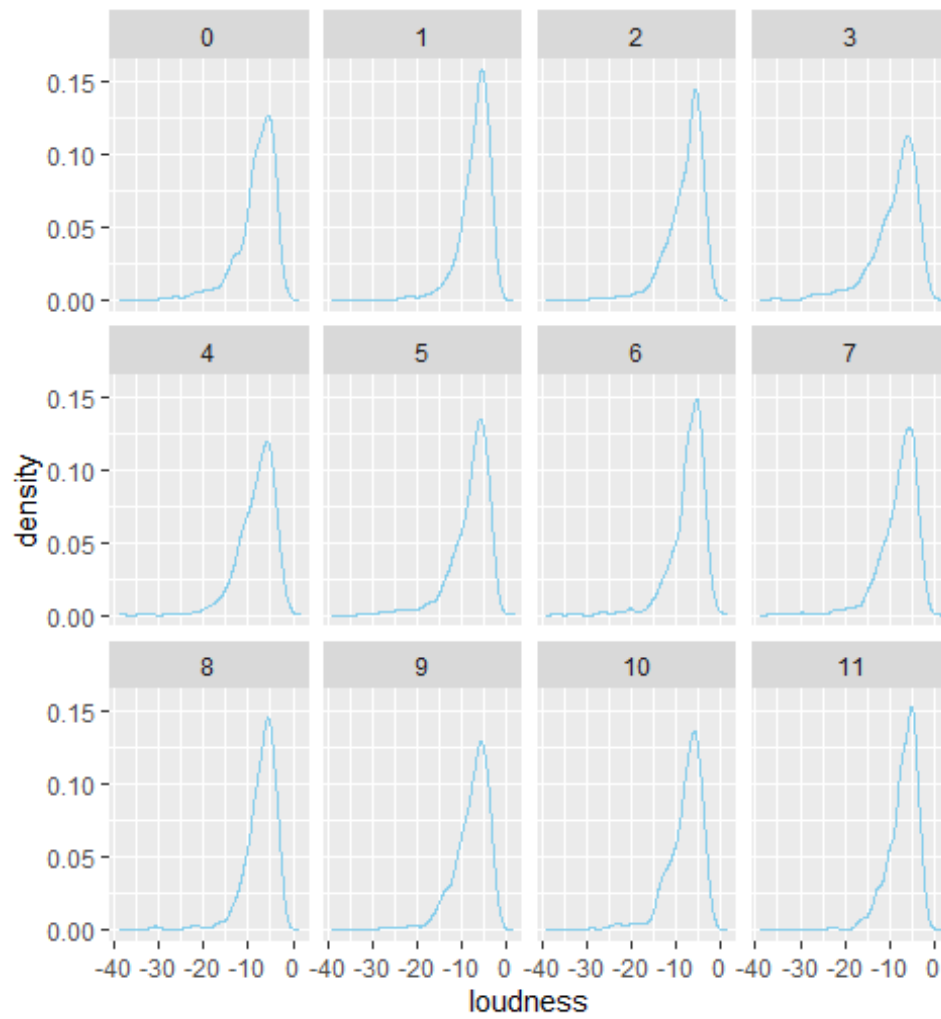
Density Plot of Liveness in Group Time Signature



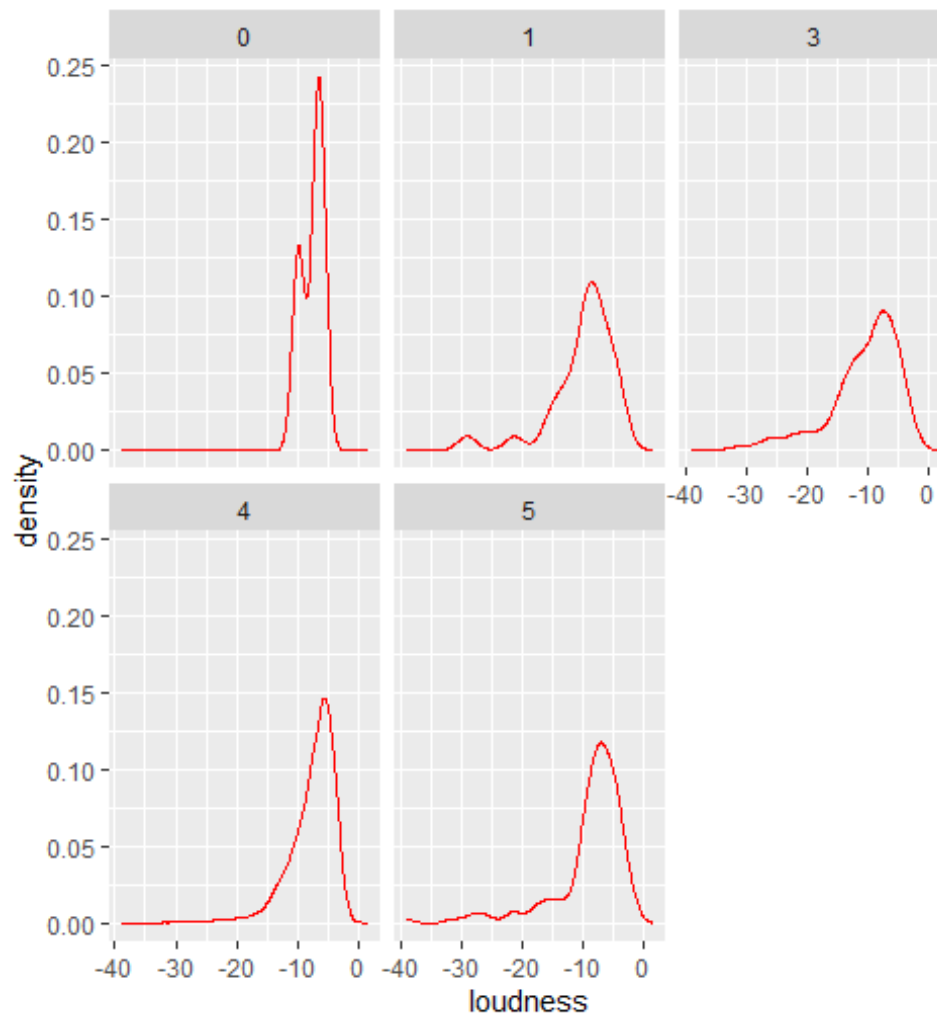
Density Plot of Liveness in Group Audio Mode



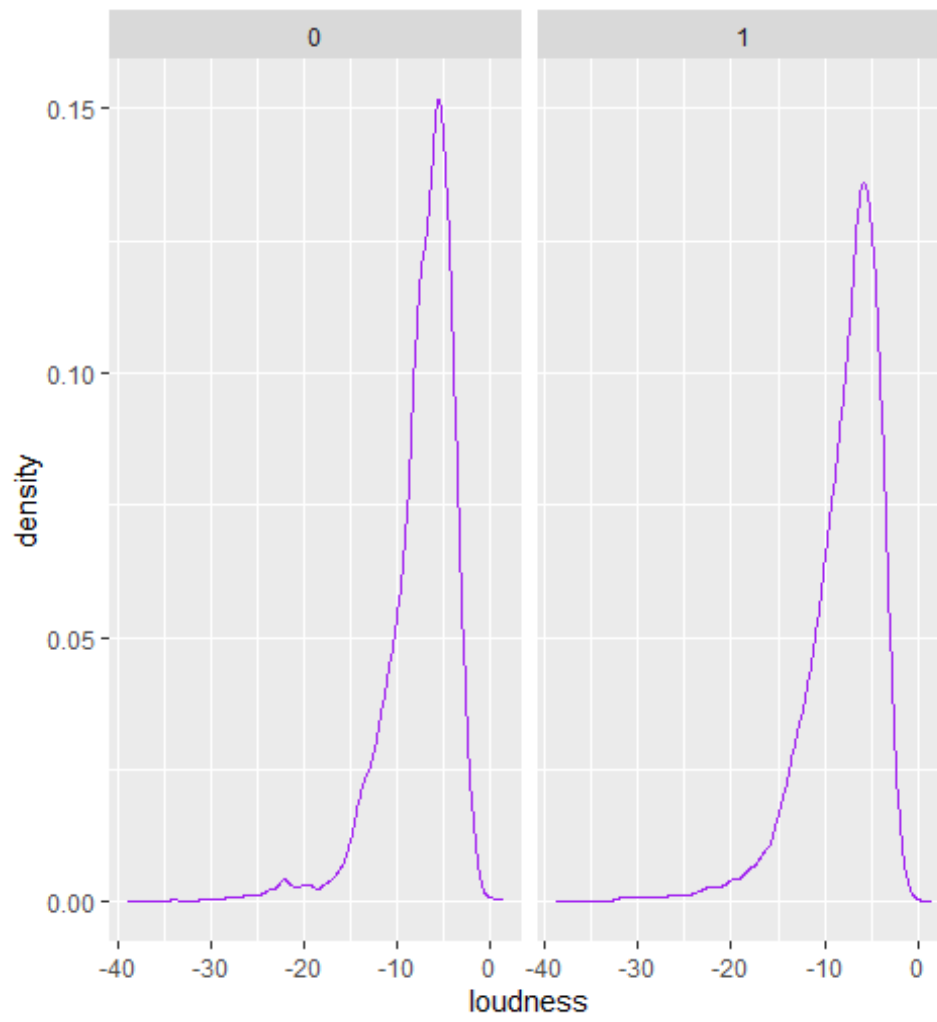
Density Plot of Loudness in Group Key



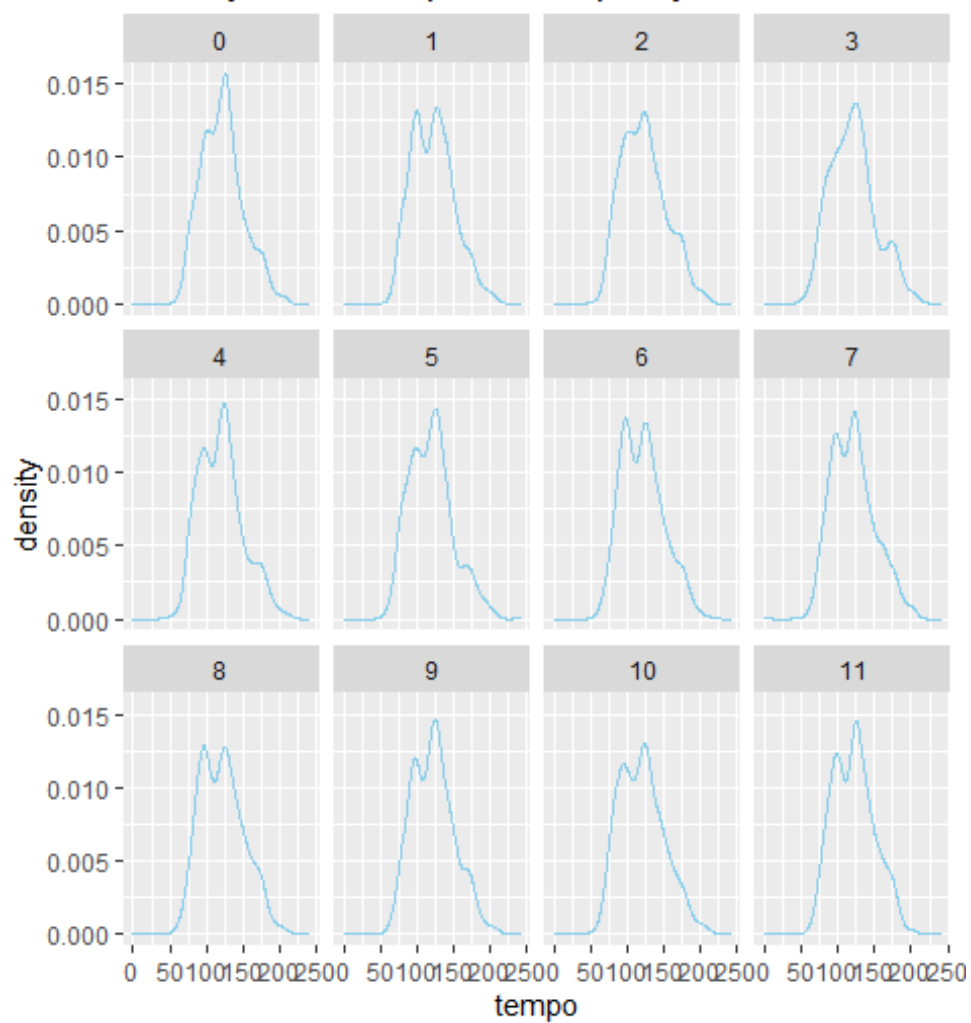
Density Plot of Loudness in Group Time Signature



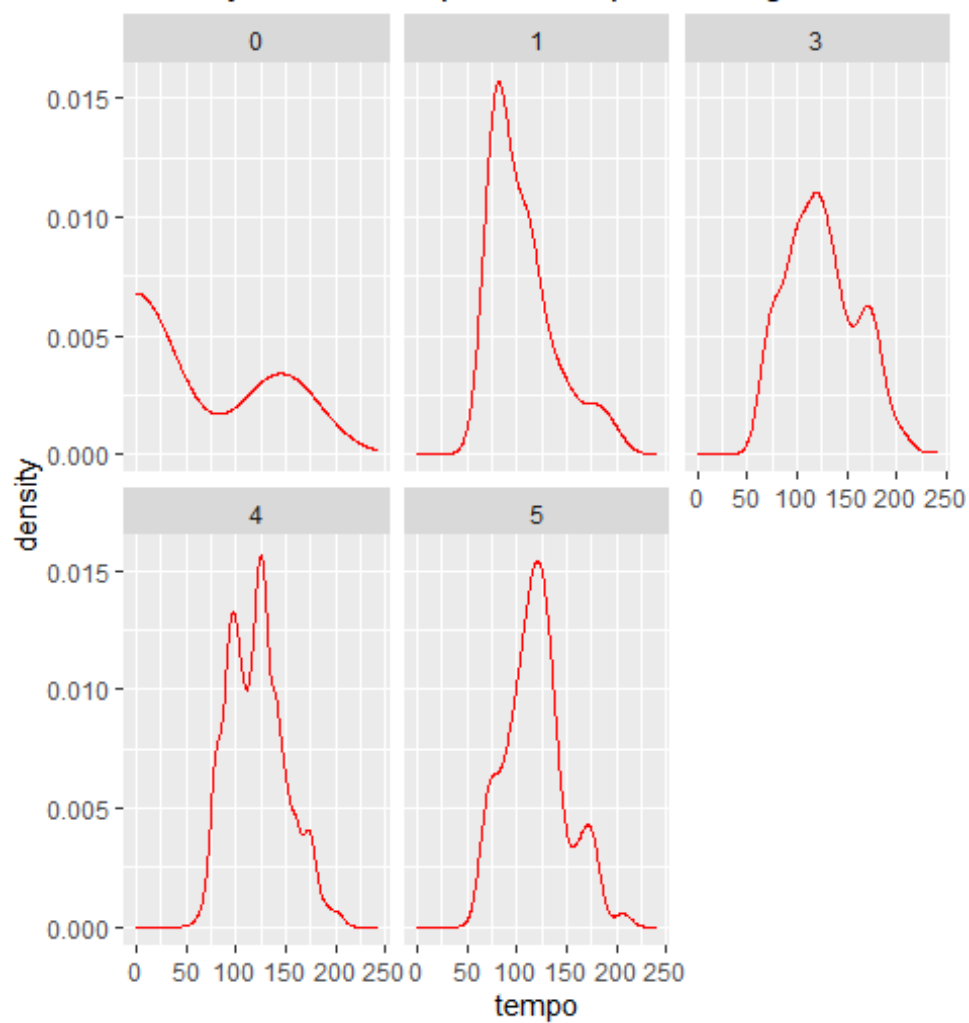
Density Plot of Loudness in Group Audio Mode



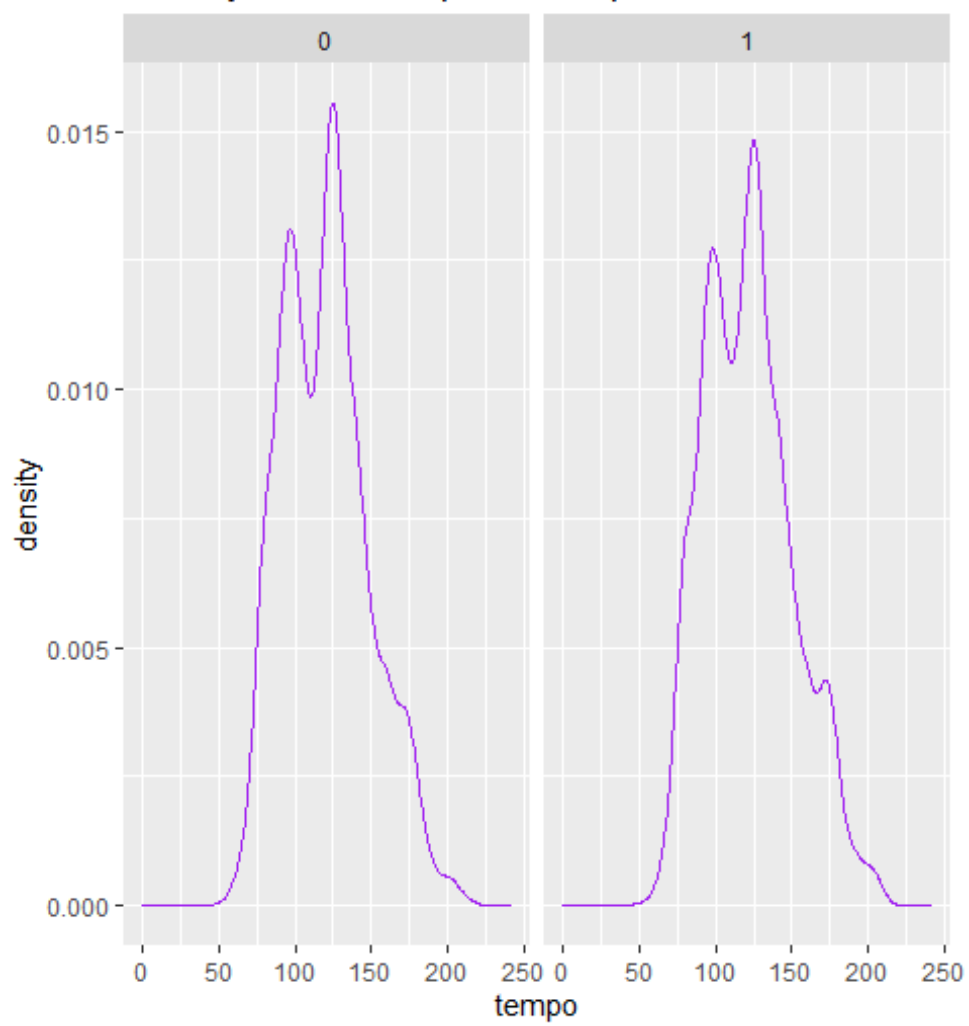
Density Plot of tempo in Group Key



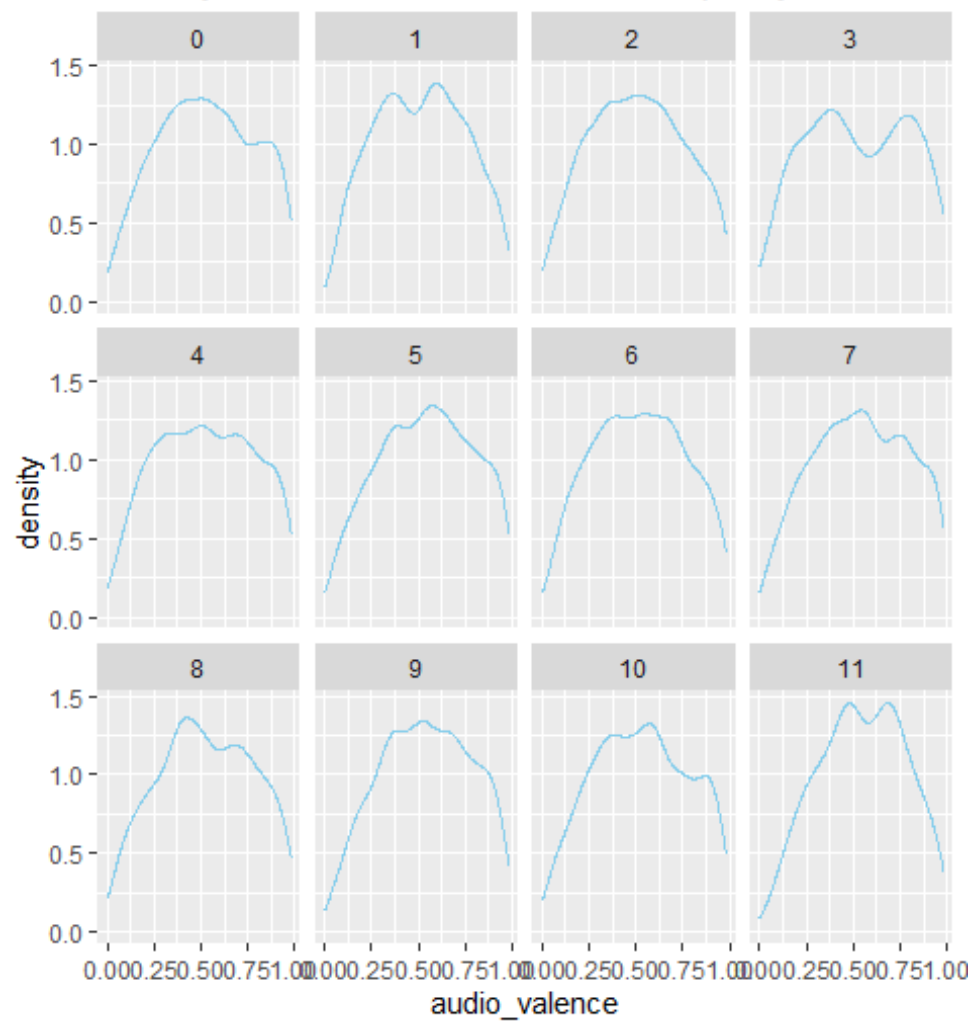
Density Plot of Tempo in Group Time Signature



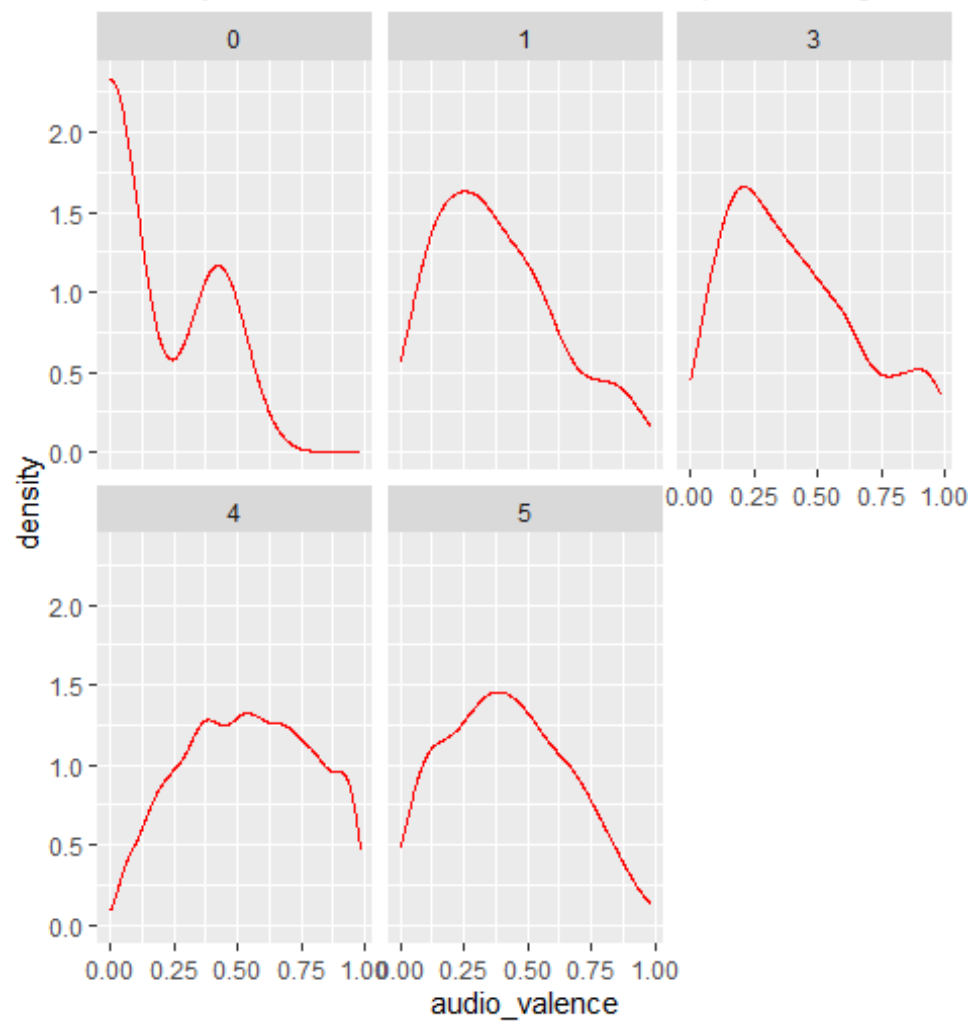
Density Plot of Tempo in Group Audio Mode



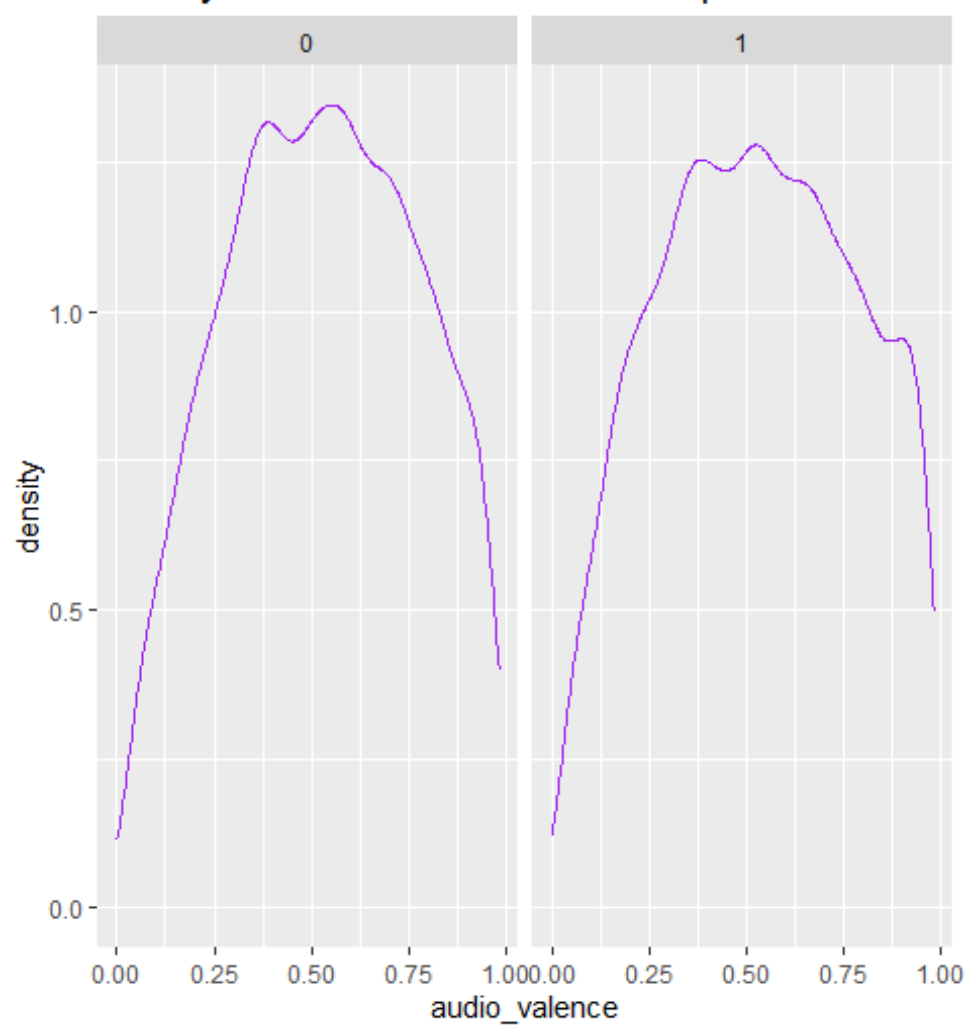
Density Plot of Audio Valence in Group Key



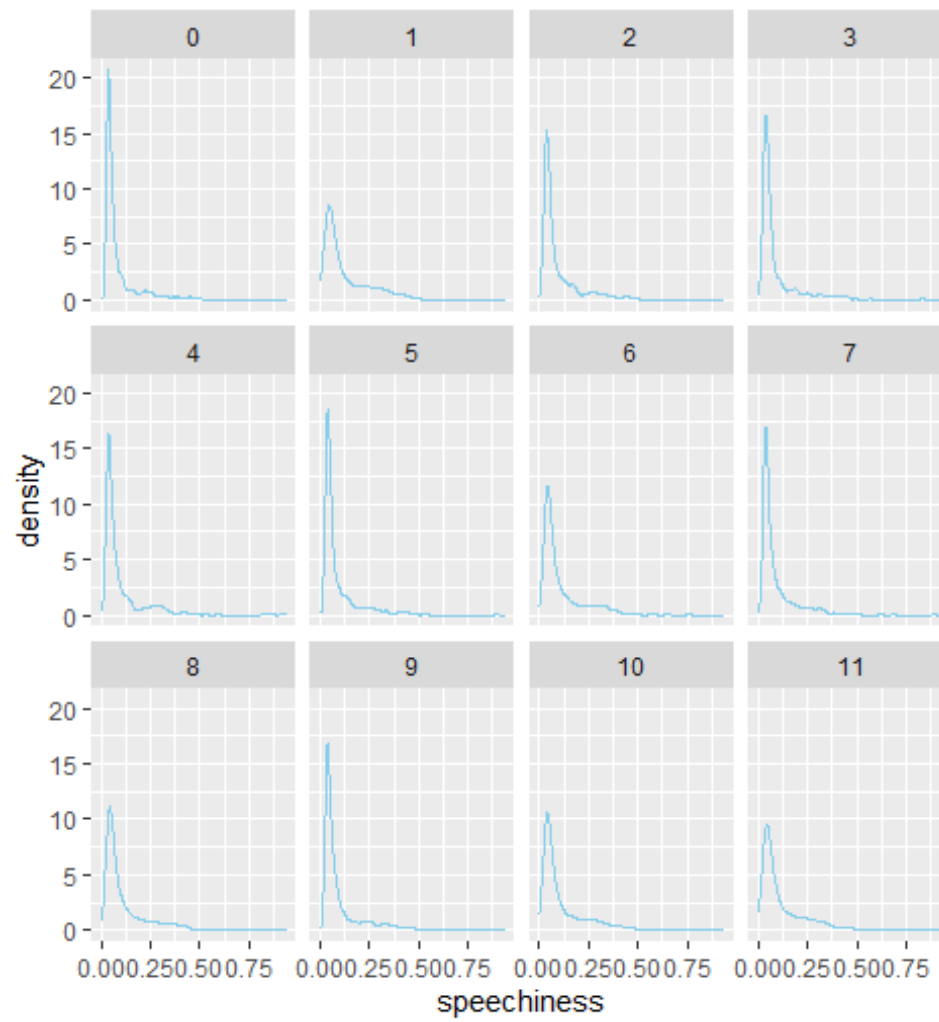
Density Plot of Audio Valence in Group Time Signature



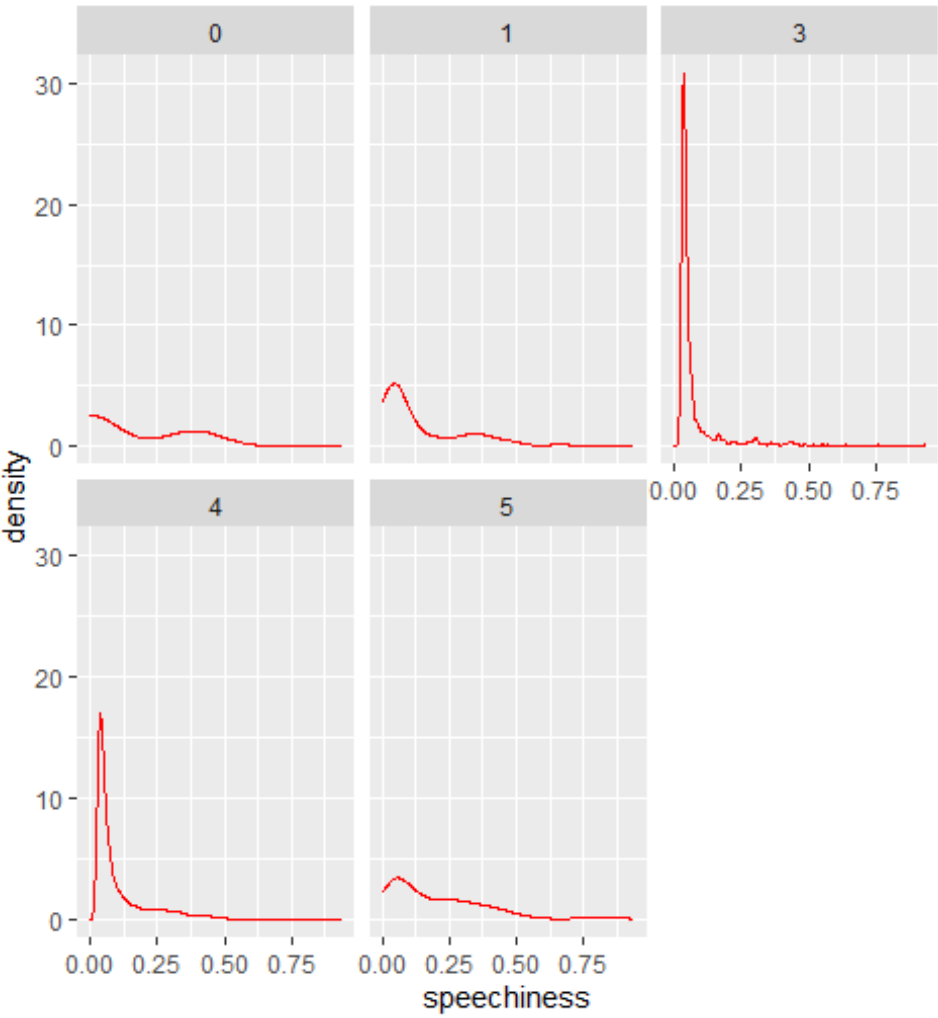
Density Plot of Audio Valence in Group Audio Mode



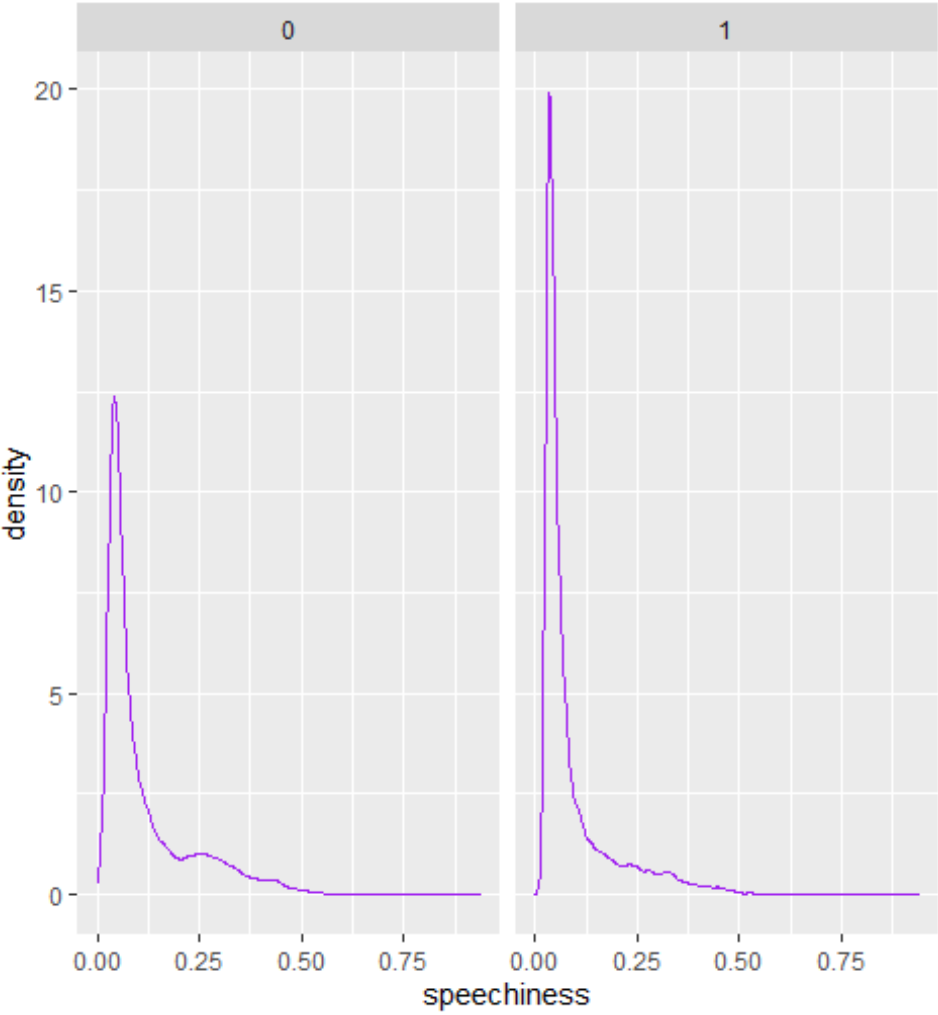
Density Plot of Speechiness in Group Key



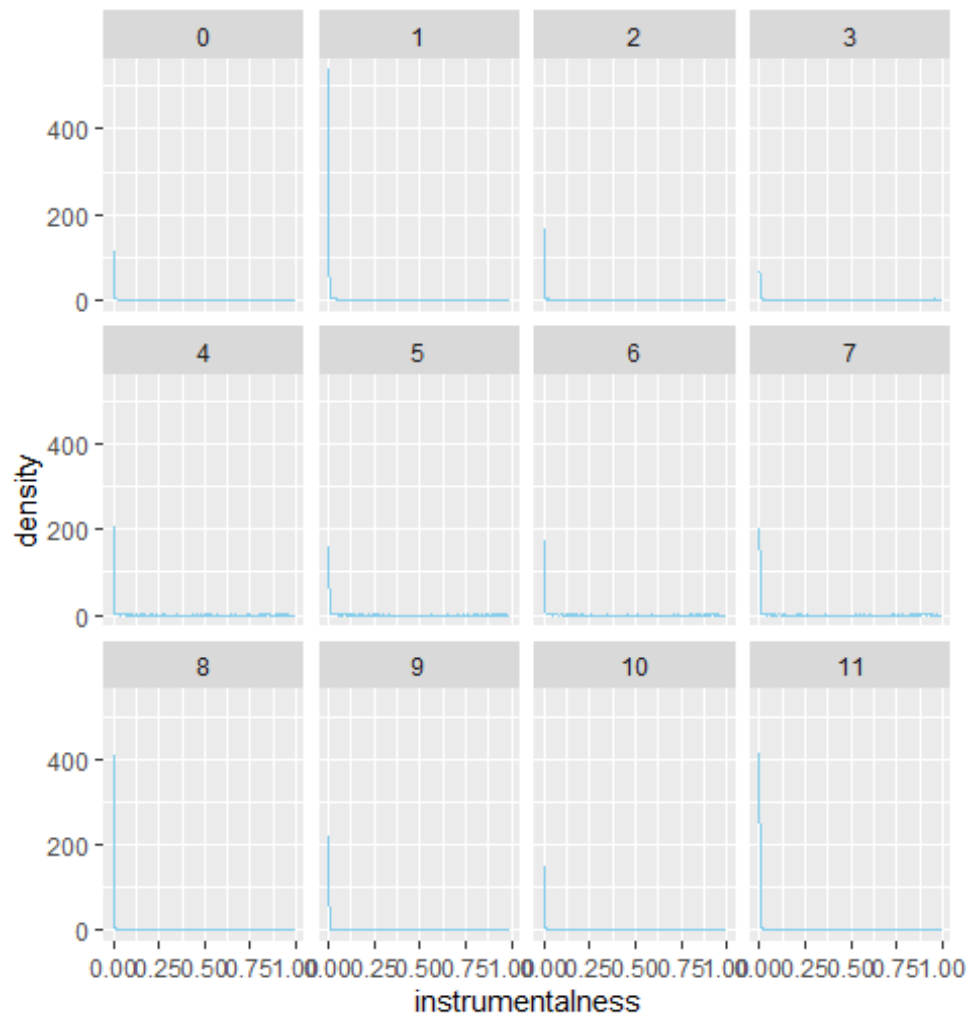
Density Plot of Speechiness in Group Time Signature



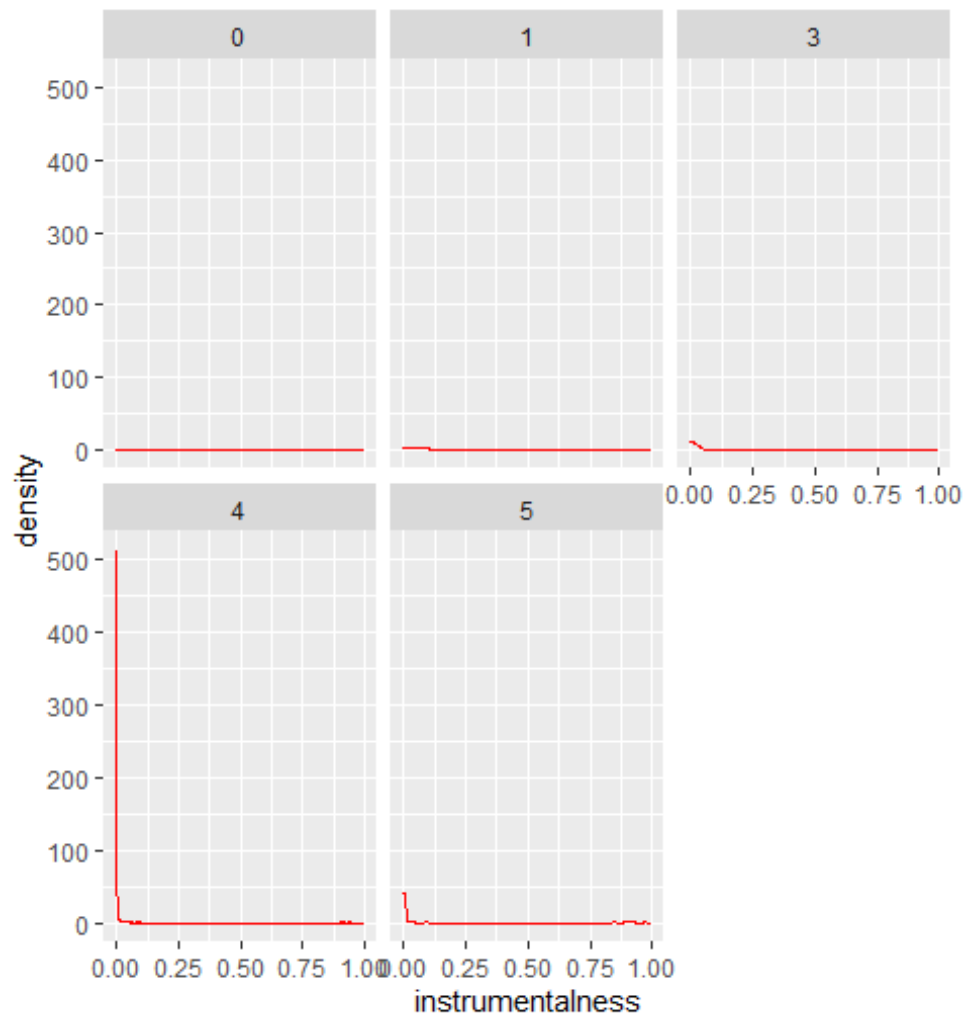
Density Plot of Speechiness in Group Audio Mode



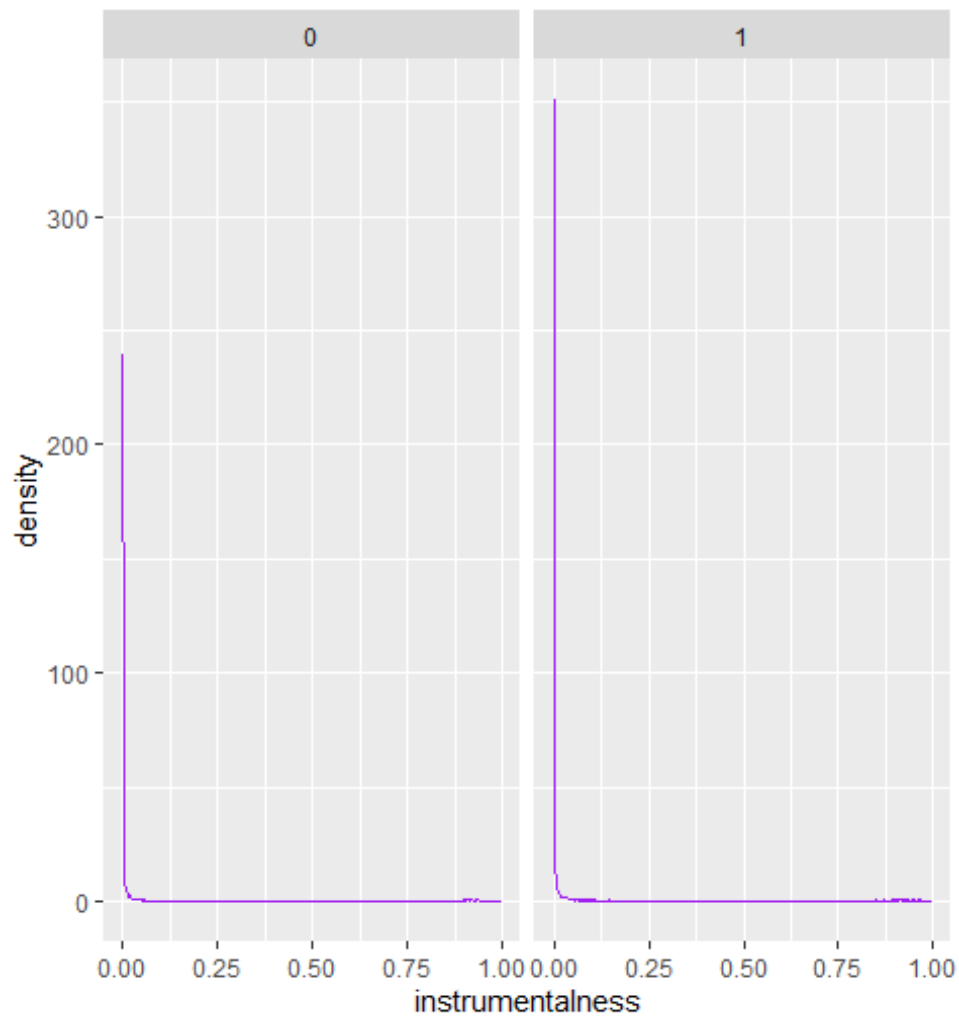
Density Plot of Instrumentalness in Group Key



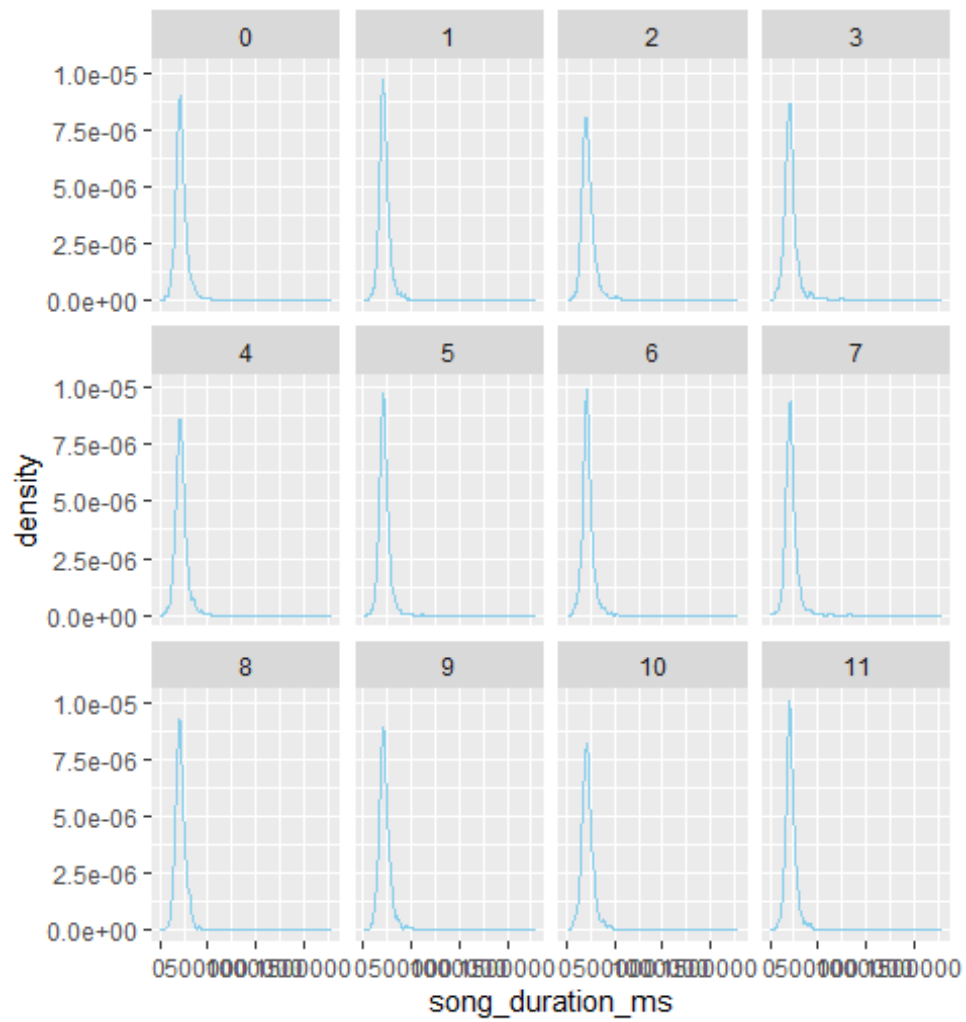
Density Plot of Instrumentalness in Group Time Signa



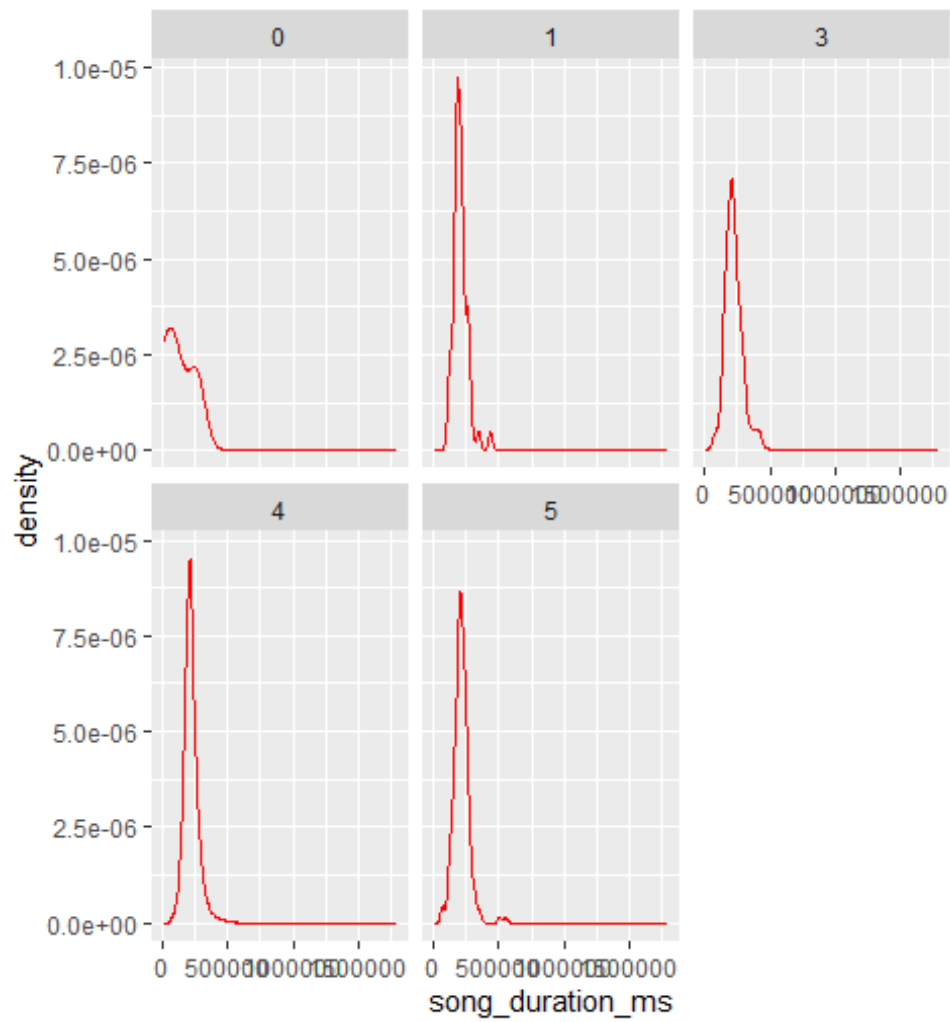
Density Plot of Instrumentalness in Group Audio Mode



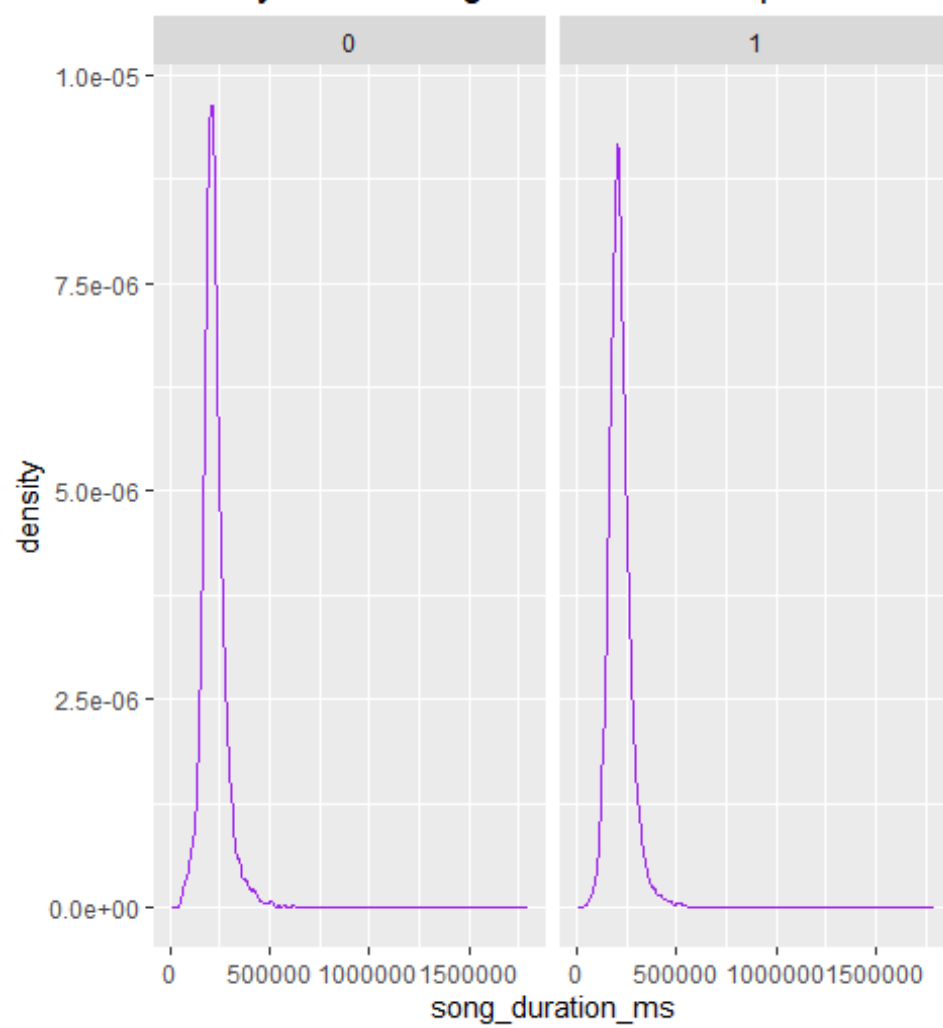
Density Plot of Song Duration in Group Key



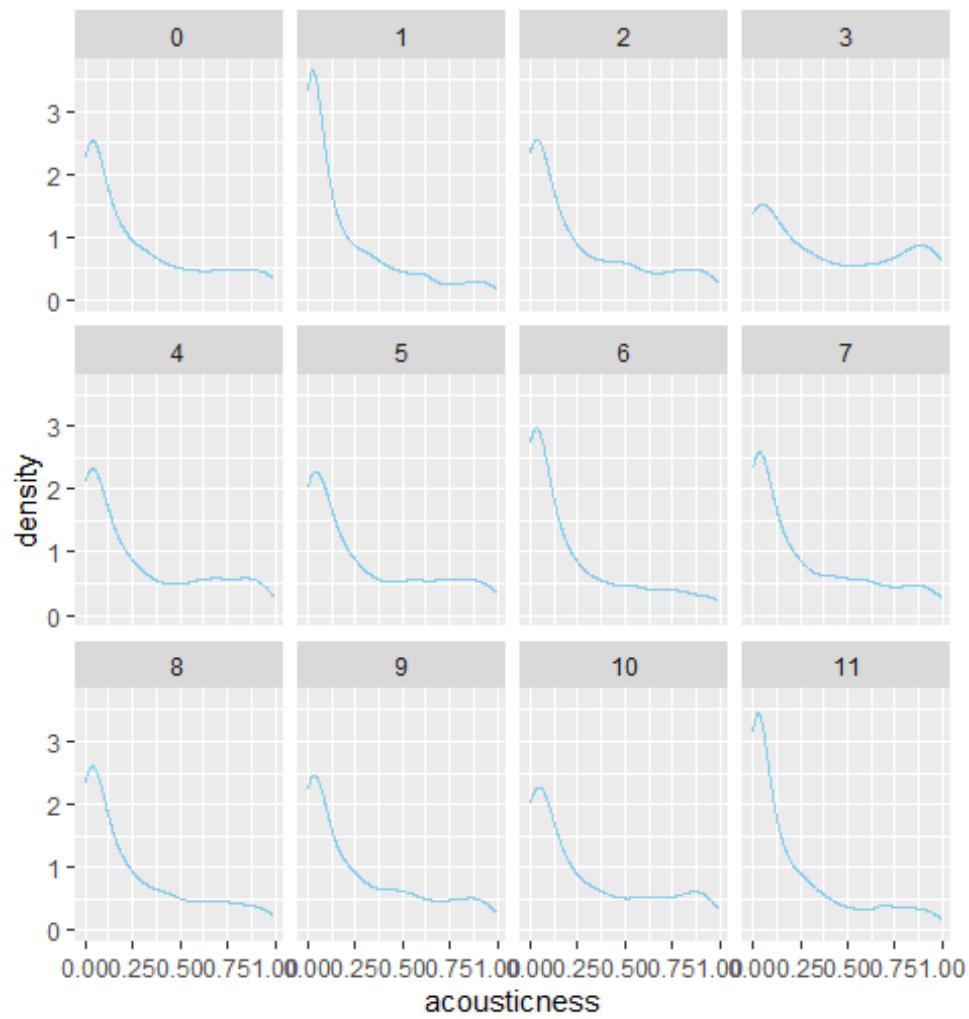
Density Plot of Song Duration in Group Time Signa



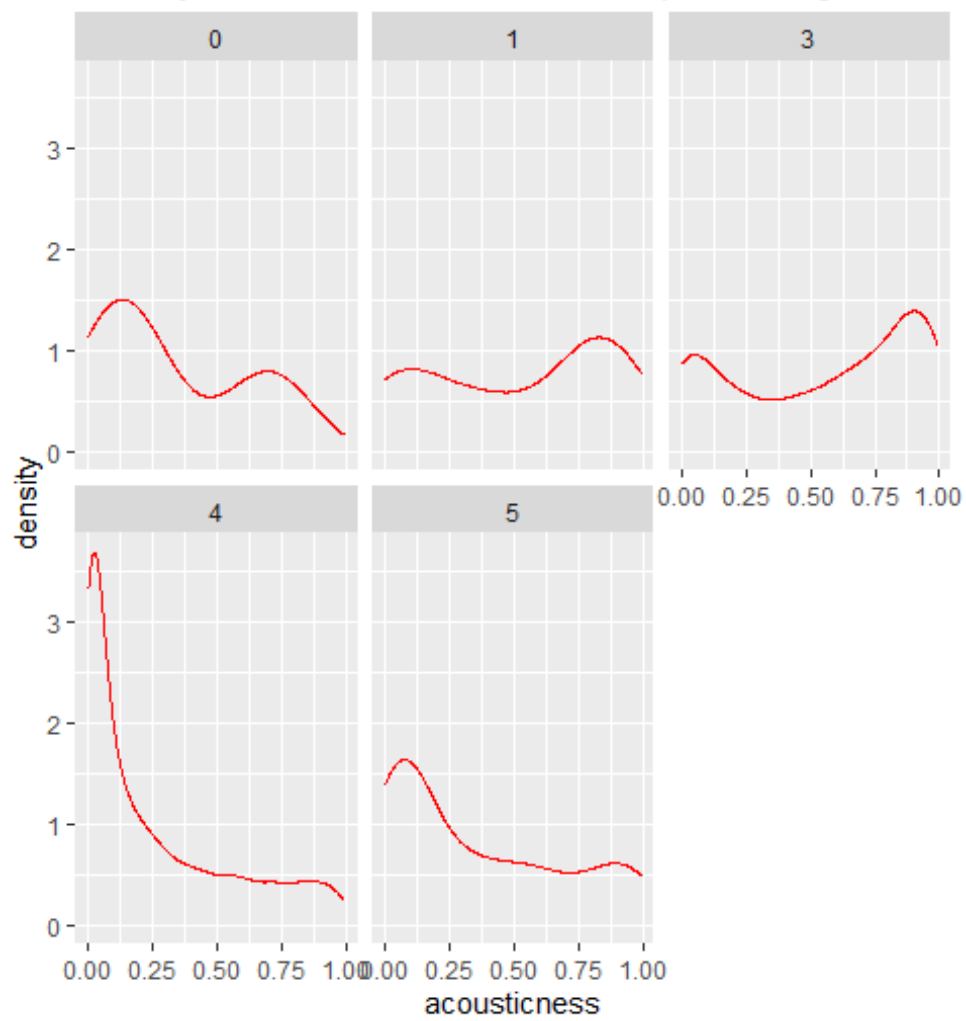
Density Plot of Song Duration in Group Audio Mode



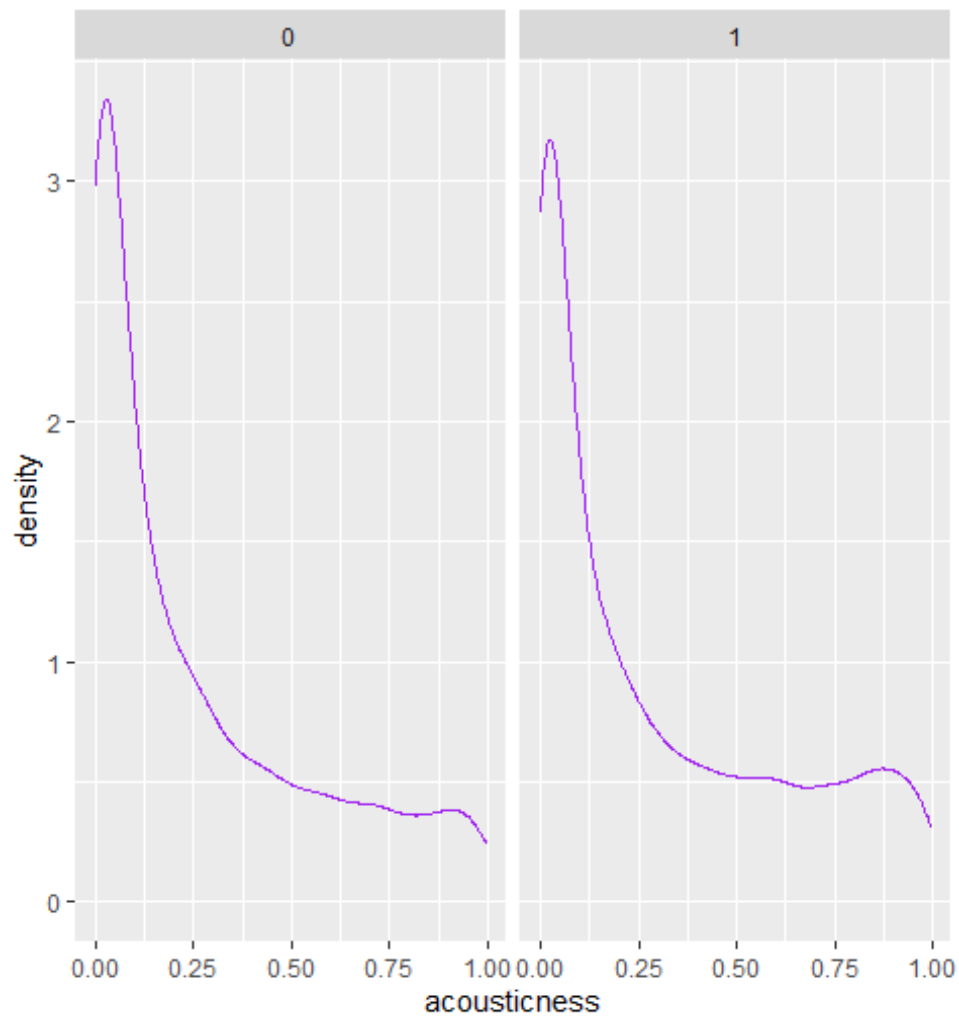
Density Plot of Acousticness in Group Key



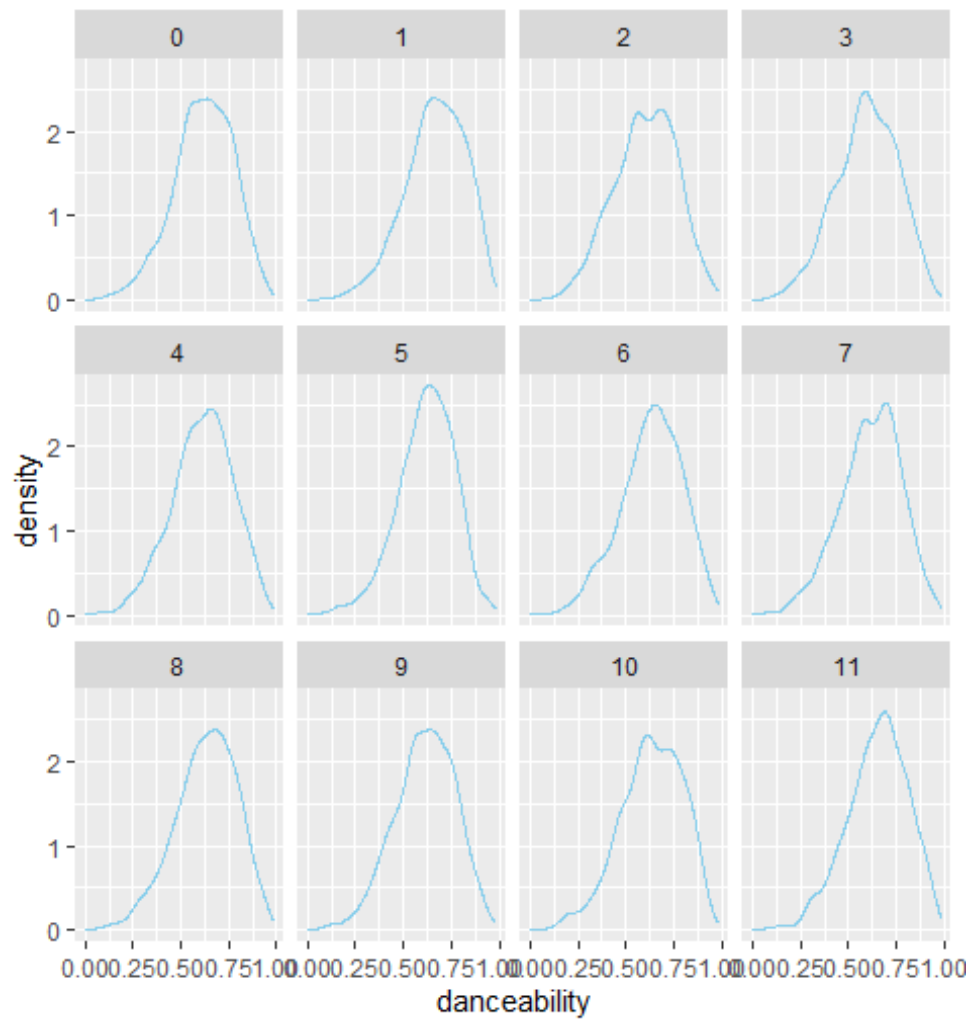
Density Plot of Acousticness in Group Time Signature



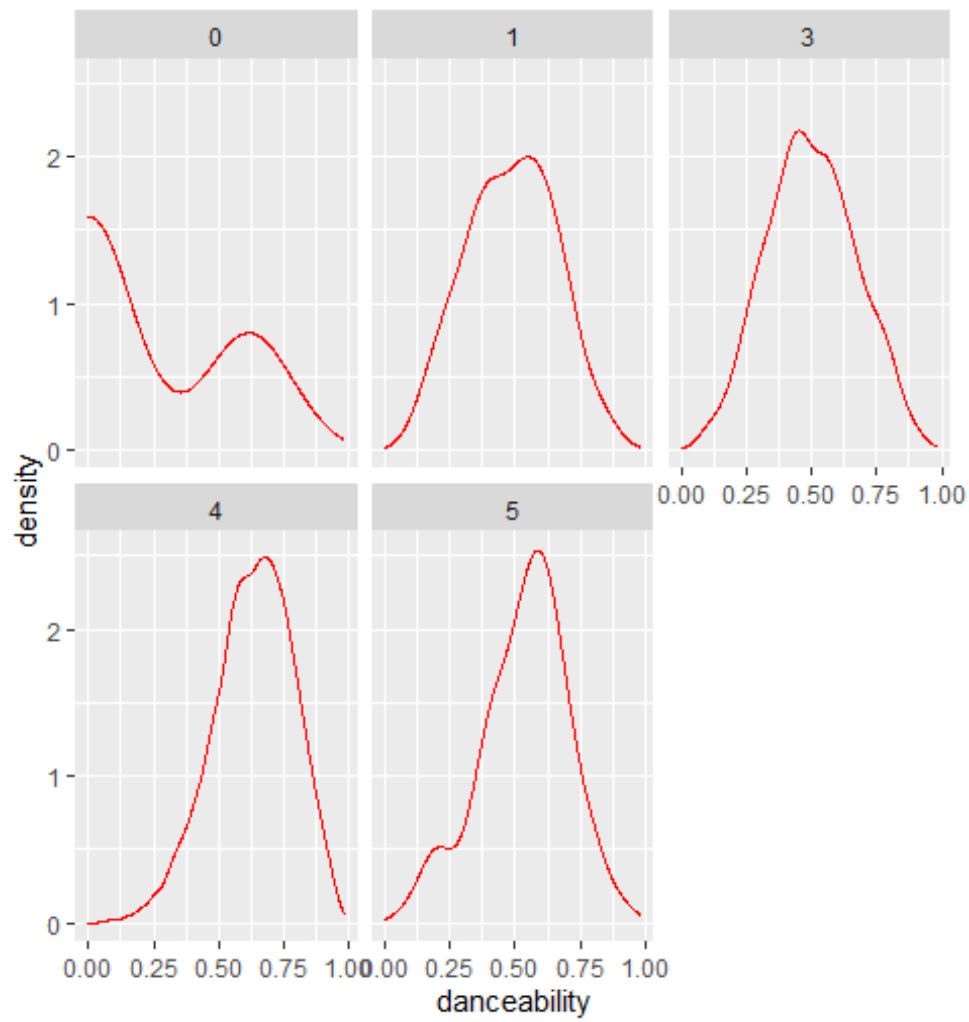
Density Plot of Acousticness in Group Audio Mode

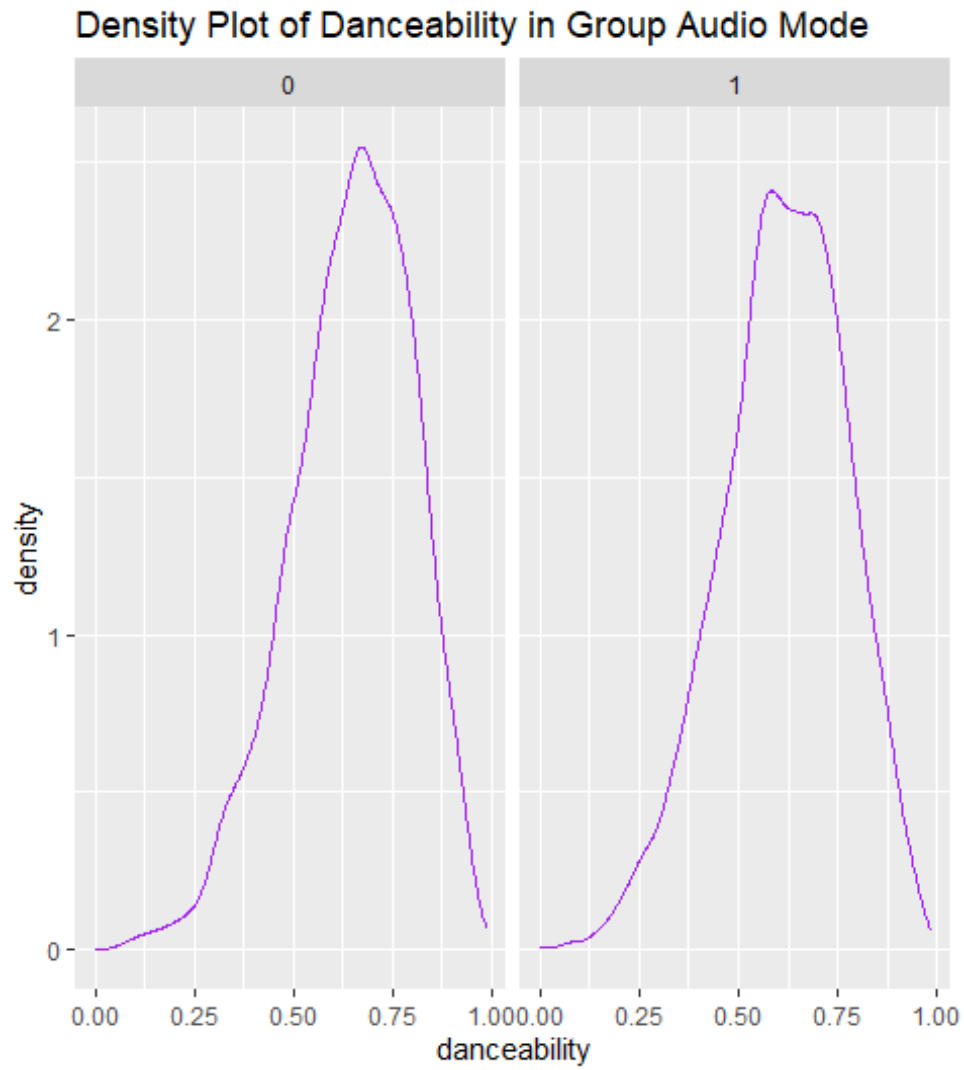


Density Plot of Danceability in Group Key

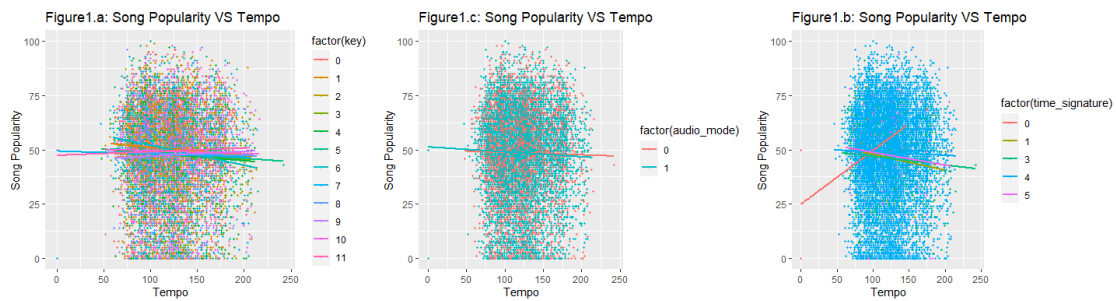


Density Plot of Danceability in Group Time Signature





Variable Selection



```
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
```

Figure2.a: Song Popularity VS Log(audio valence)

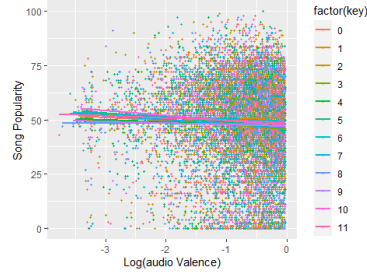


Figure2.c: Song Popularity VS Audio Valence



Figure2.b: Song Popularity VS Audio Valence

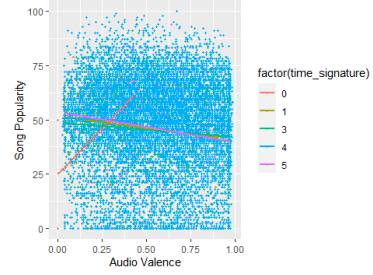


Figure3.a: Song Popularity VS Log(energy)

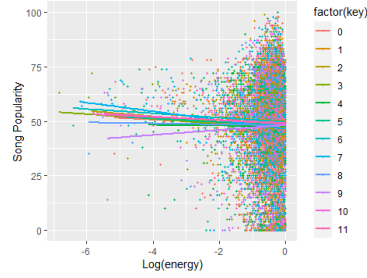


Figure3.c: Song Popularity VS Energy

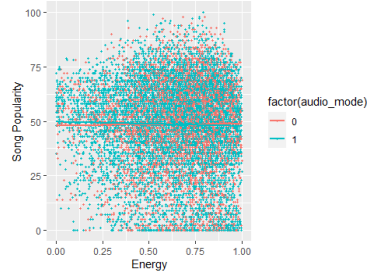


Figure3.b: Song Popularity VS Energy

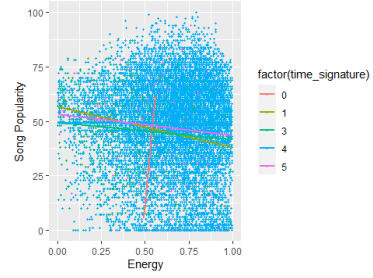


Figure4.a: Song Popularity VS Loudness

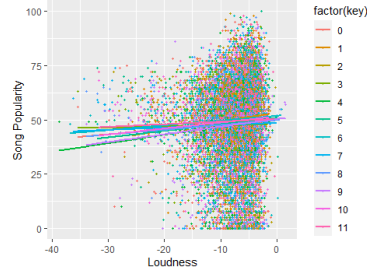


Figure4.c: Song Popularity VS Loudness

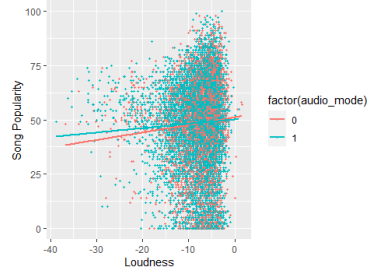


Figure4.b: Song Popularity VS Loudness

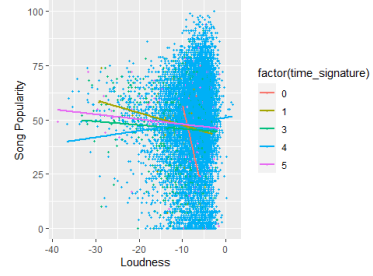


Figure5.a: Song Popularity VS Danceability

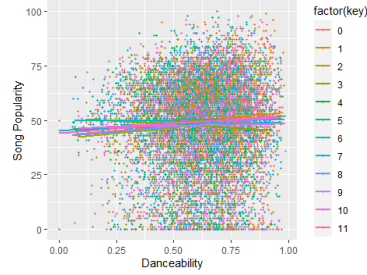


Figure5.c: Song Popularity VS Danceability

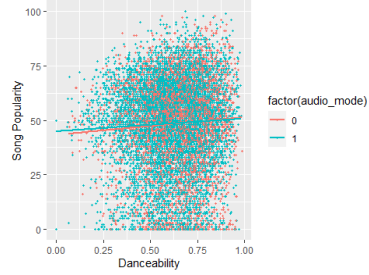


Figure5.b: Song Popularity VS Danceability

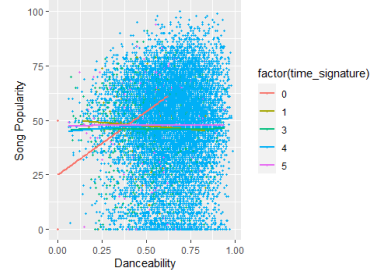


Figure6.a: Song Popularity VS Duration

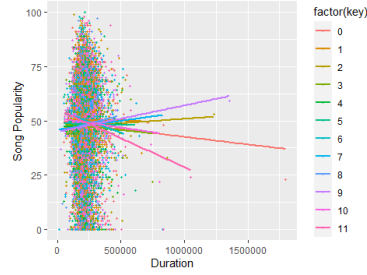


Figure6.c: Song Popularity VS Duration

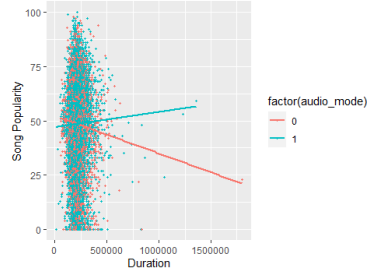
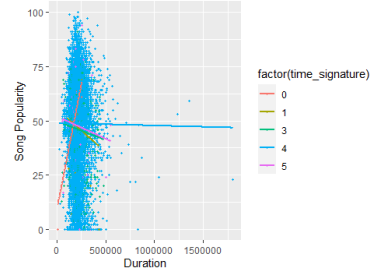
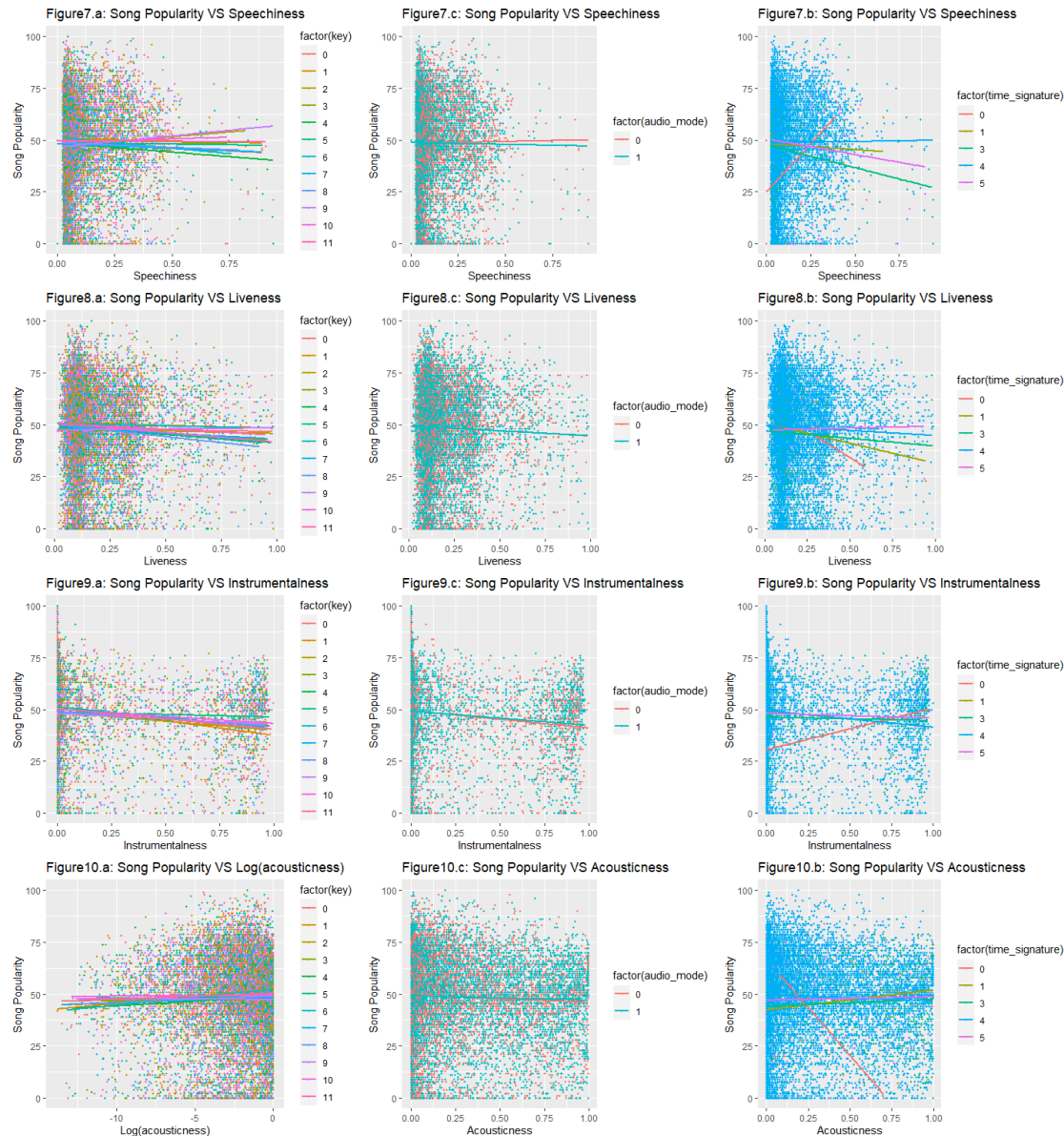


Figure6.b: Song Popularity VS Duration





Correlation

```
music_data_cor <- music_data[, c(3,4,5,6,7,9,10,12,13,15)]
correlation(music_data_cor)
```

```
## # Correlation Matrix (pearson-method)
```

```
##
```

```
## Parameter1 | Parameter2 | r | 95% CI | t(13068) |
```

```
p
```

```
## -----
```

```
## song_duration_ms | acousticness | -0.12 | [-0.14, -0.11] | -14.05 |
```

```
< .001***
```

```
## song_duration_ms | danceability | -0.09 | [-0.10, -0.07] | -9.86 |
```



```

< .001***
## song_duration_ms |          energy |  0.11 | [ 0.09,  0.12] |  12.15 |
< .001***
## song_duration_ms | instrumentalness | -0.03 | [-0.04, -0.01] |  -3.08 |
0.012*
## song_duration_ms |          liveness |  0.02 | [ 0.01,  0.04] |   2.69 |
0.028*
## song_duration_ms |          loudness |  0.04 | [ 0.02,  0.05] |   4.05 |
< .001***
## song_duration_ms |        speechiness | -0.08 | [-0.10, -0.06] |  -9.27 |
< .001***
## song_duration_ms |          tempo |  0.02 | [ 0.00,  0.03] |   1.95 |
0.152
## song_duration_ms |    audio_valence | -0.07 | [-0.09, -0.05] |  -7.84 |
< .001***
## acousticness     |    danceability | -0.18 | [-0.20, -0.16] | -20.88 |
< .001***
## acousticness     |          energy | -0.68 | [-0.69, -0.67] | -106.22 |
< .001***
## acousticness     | instrumentalness |  0.19 | [ 0.17,  0.20] |  21.68 |
< .001***
## acousticness     |          liveness | -0.09 | [-0.11, -0.07] | -10.32 |
< .001***
## acousticness     |          loudness | -0.57 | [-0.58, -0.56] | -79.81 |
< .001***
## acousticness     |        speechiness | -0.09 | [-0.11, -0.07] | -10.51 |
< .001***
## acousticness     |          tempo | -0.15 | [-0.16, -0.13] | -17.00 |
< .001***
## acousticness     |    audio_valence | -0.12 | [-0.14, -0.11] | -14.14 |
< .001***
## danceability      |          energy |  0.06 | [ 0.05,  0.08] |   7.31 |
< .001***
## danceability      | instrumentalness | -0.13 | [-0.14, -0.11] | -14.65 |
< .001***
## danceability      |          liveness | -0.09 | [-0.11, -0.08] | -10.83 |
< .001***
## danceability      |          loudness |  0.18 | [ 0.17,  0.20] |  21.12 |
< .001***
## danceability      |        speechiness |  0.20 | [ 0.19,  0.22] |  23.81 |
< .001***
## danceability      |          tempo | -0.13 | [-0.14, -0.11] | -14.61 |
< .001***
## danceability      |    audio_valence |  0.35 | [ 0.34,  0.37] |  42.95 |
< .001***
## energy            | instrumentalness | -0.23 | [-0.25, -0.21] | -27.14 |
< .001***
## energy            |          liveness |  0.18 | [ 0.16,  0.19] |  20.63 |
< .001***
## energy            |          loudness |  0.77 | [ 0.76,  0.78] | 138.08 |

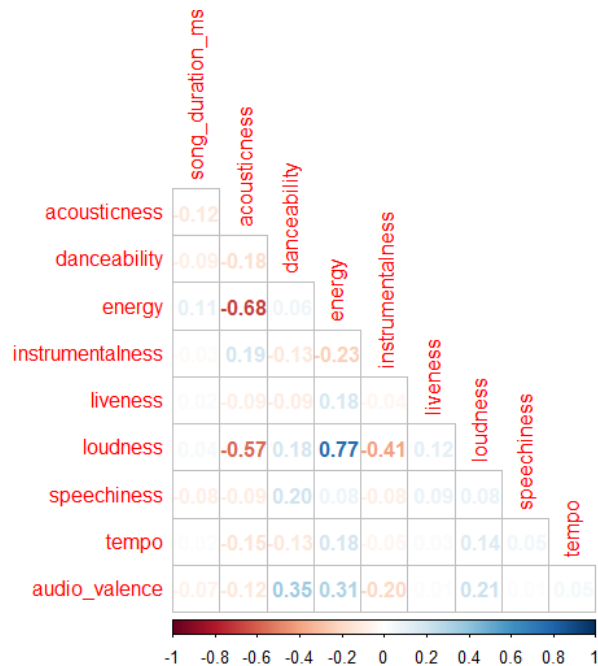
```

```

< .001***
## energy          |      speechiness |  0.08 | [ 0.06,  0.09] |    8.91 |
< .001***
## energy          |      tempo       |  0.18 | [ 0.16,  0.20] |   21.08 |
< .001***
## energy          |    audio_valence |  0.31 | [ 0.30,  0.33] |   37.57 |
< .001***
## instrumentalness |      liveness    | -0.04 | [-0.06, -0.03] |   -5.02 |
< .001***
## instrumentalness |      loudness     | -0.41 | [-0.42, -0.39] |  -50.68 |
< .001***
## instrumentalness |      speechiness  | -0.08 | [-0.10, -0.07] |   -9.45 |
< .001***
## instrumentalness |      tempo        | -0.05 | [-0.06, -0.03] |   -5.22 |
< .001***
## instrumentalness |    audio_valence  | -0.20 | [-0.22, -0.19] |  -23.56 |
< .001***
## liveness        |      loudness     |  0.12 | [ 0.10,  0.13] |   13.29 |
< .001***
## liveness        |      speechiness  |  0.09 | [ 0.07,  0.11] |   10.26 |
< .001***
## liveness        |      tempo        |  0.03 | [ 0.01,  0.04] |    2.97 |
0.015*
## liveness        |    audio_valence  |  0.01 | [ 0.00,  0.03] |    1.63 |
0.206
## loudness        |      speechiness  |  0.08 | [ 0.07,  0.10] |    9.59 |
< .001***
## loudness        |      tempo        |  0.14 | [ 0.13,  0.16] |   16.43 |
< .001***
## loudness        |    audio_valence  |  0.21 | [ 0.20,  0.23] |   25.01 |
< .001***
## speechiness     |      tempo        |  0.05 | [ 0.04,  0.07] |    6.06 |
< .001***
## speechiness     |    audio_valence  |  0.01 | [-0.01,  0.03] |    1.33 |
0.206
## tempo           |    audio_valence  |  0.05 | [ 0.03,  0.07] |    5.53 |
< .001***
##
## p-value adjustment method: Holm (1979)
## Observations: 13070

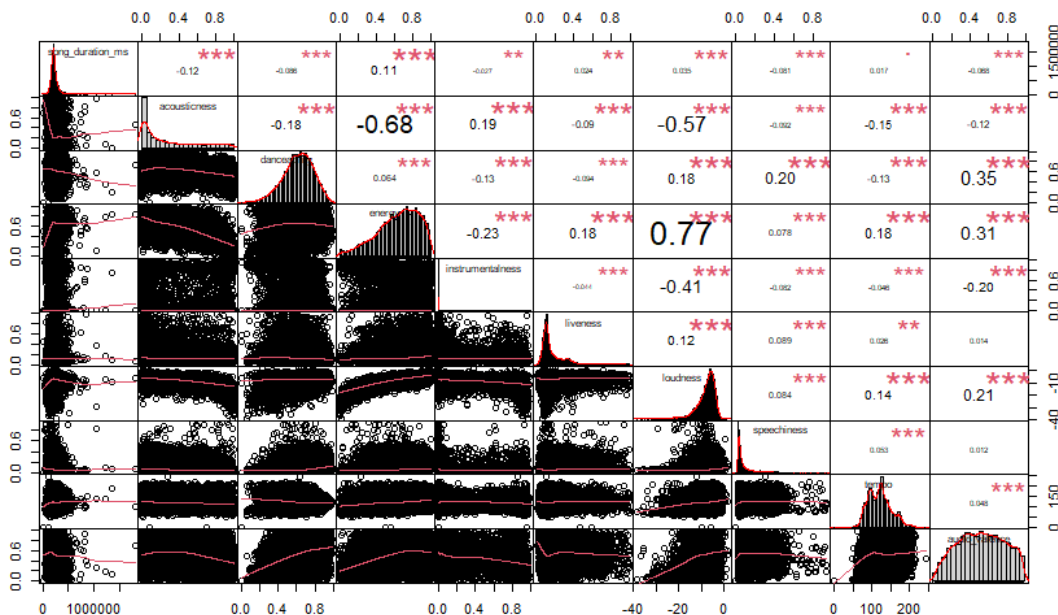
corr_music <- cor(music_data_cor)
corrplot(corr_music, method = "number", diag = FALSE, type = "lower")

```

[illegible]

[illegible]

```
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
```



Fit Model

```
model <- lmer(song_popularity ~ tempo + danceability + audio_valence +
liveness + loudness + (1 + tempo + danceability|key) + (1|time_signature) +
(1|audio_mode), data = music_data)
```

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

```
## Warning: Model failed to converge with 4 negative eigenvalues: -2.0e+00 -
9.1e+00
```

```
## -3.4e+02 -1.2e+03
```

```
summary(model)
```

```

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: song_popularity ~ tempo + danceability + audio_valence + liveness
+
##      loudness + (1 + tempo + danceability | key) + (1 | time_signature) +
##      (1 | audio_mode)
##      Data: music_data
##
## REML criterion at convergence: 115396.8
##
## Scaled residuals:
##      Min      1Q  Median      3Q      Max
## -2.7740 -0.5528  0.1424  0.7216  2.4867
##
## Random effects:
##      Groups          Name          Variance Std.Dev.  Corr
##      key              (Intercept)  1.603e+00  1.265996
##              tempo              1.507e-05  0.003882 -1.00
##              danceability        1.869e+00  1.367055 -0.76  0.76
##      time_signature (Intercept)  2.530e+02  15.904768
##      audio_mode      (Intercept)  3.940e+02  19.850453
##      Residual              3.988e+02  19.969965
## Number of obs: 13070, groups:  key, 12; time_signature, 5; audio_mode, 2
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  5.036e+01  1.592e+01  1.294e+04   3.164  0.00156 **
## tempo        -1.627e-02  6.278e-03  6.830e+02  -2.591  0.00977 **
## danceability  7.505e+00  1.299e+00  3.775e+01   5.778  1.17e-06 ***
## audio_valence -6.840e+00  7.664e-01  1.267e+04  -8.925  < 2e-16 ***
## liveness      -4.771e+00  1.220e+00  1.306e+04  -3.910  9.27e-05 ***
## loudness       3.043e-01  4.520e-02  1.286e+04   6.732  1.74e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) tempo  dncblt ad_vln livnss
## tempo        -0.054
## danceability -0.048  0.201
## audio_valenc -0.006 -0.073 -0.312
## liveness      -0.024  0.009  0.116 -0.029
## loudness       0.041 -0.154 -0.117 -0.141 -0.125
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')

ranova(model)

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

## Warning: Model failed to converge with 1 negative eigenvalue: -2.3e+02

```

```

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

## Warning: Model failed to converge with 3 negative eigenvalues: -1.5e+00 -
5.2e+00
## -8.0e+02

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

## Warning: Model failed to converge with 2 negative eigenvalues: -1.5e+00 -
1.1e+01

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

## Warning: Model failed to converge with 4 negative eigenvalues: -6.1e+00 -
1.2e+01
## -1.3e+03 -4.7e+03

## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## song_popularity ~ tempo + danceability + audio_valence + liveness +
loudness + (1 + tempo + danceability | key) + (1 | time_signature) + (1 |
audio_mode)
##
##                                npar logLik    AIC
LRT Df
## <none>                                15 -57698 115427
## tempo in (1 + tempo + danceability | key)    12 -57689 115402 -
19.2986  3
## danceability in (1 + tempo + danceability | key) 12 -57701 115426
5.1712  3
## (1 | time_signature)                        14 -57697 115421 -
3.4402  1
## (1 | audio_mode)                            14 -57698 115424 -
0.8801  1
##
##                                Pr(>Chisq)
## <none>                                1.0000
## tempo in (1 + tempo + danceability | key)    0.1597
## danceability in (1 + tempo + danceability | key) 1.0000
## (1 | time_signature)                        1.0000
## (1 | audio_mode)                            1.0000

VarCorr(model)

## Groups      Name      Std.Dev.  Corr
## key         (Intercept)  1.2659959
##              tempo       0.0038822 -1.000
##              danceability 1.3670545 -0.761  0.761
## time_signature (Intercept) 15.9047685
## audio_mode    (Intercept) 19.8504527
## Residual      19.9699651

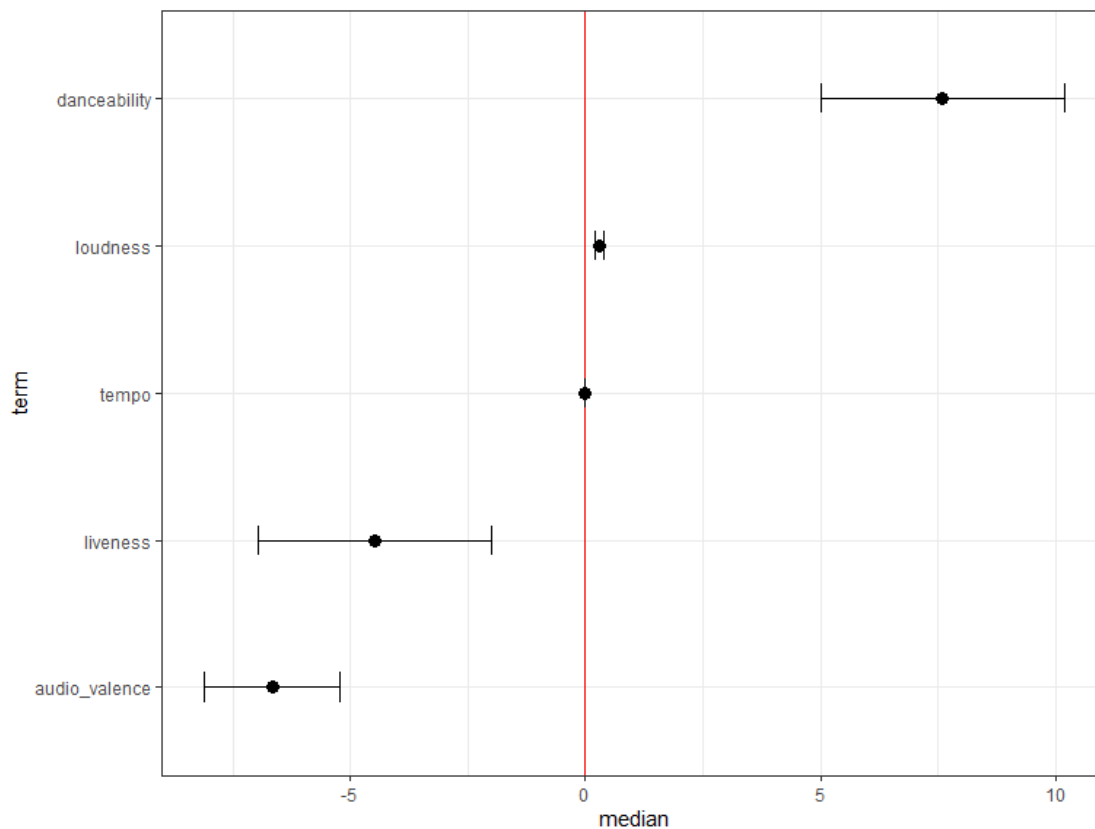
```

```
# Check the confidence interval to random effects
confint.merMod(model, method = 'Wald')
```

```
##              2.5 %      97.5 %
## .sig01         NA         NA
## .sig02         NA         NA
## .sig03         NA         NA
## .sig04         NA         NA
## .sig05         NA         NA
## .sig06         NA         NA
## .sig07         NA         NA
## .sig08         NA         NA
## .sigma         NA         NA
## (Intercept)  19.16256998  81.553999815
## tempo       -0.02856893 -0.003961337
## danceability  4.95938199 10.050901362
## audio_valence -8.34266364 -5.338253987
## liveness     -7.16307452 -2.379751467
## loudness      0.21569048  0.392860235
```

```
# Fixed Effect
```

```
plotFEsim(FEsim(model))
```



```
# Random Effect
```

```
round(ranef(model)$key, digits = 4)
```



```
##      (Intercept)      tempo danceability
## 0      -0.4415    0.0014      0.1306
## 1       0.2288   -0.0007      0.8507
## 2     -0.8040    0.0025      0.2677
## 3       0.2578   -0.0008     -0.6062
## 4       0.1129   -0.0003     -0.4721
## 5       0.1656   -0.0005      0.0153
## 6       1.3714   -0.0042     -0.4203
## 7     -0.1107    0.0003     -0.5004
## 8     -0.3975    0.0012     -0.0023
## 9     -0.3423    0.0010     -0.3820
## 10      0.2703   -0.0008      0.2386
## 11     -0.3110    0.0010      0.8806

round(ranef(model)$time_signature, digits = 4)

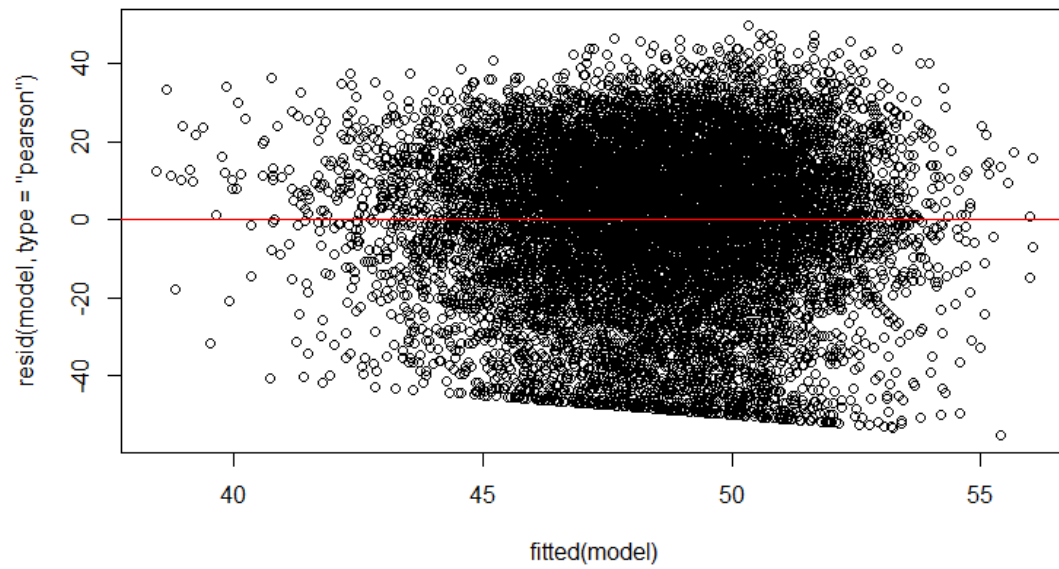
##      (Intercept)
## 0      -5.7595
## 1       1.4134
## 3       0.5875
## 4       2.2786
## 5       1.4799

round(ranef(model)$audio_mode, digits = 4)

##      (Intercept)
## 0      -0.2717
## 1       0.2717
```

Model Check

```
# Residuals VS Fitted value
plot(fitted(model), resid(model, type = "pearson"))
abline(0,0, col="red")
```



```
# Q-Q Plot
qqnorm(resid(model))
qqline(resid(model), col = "red")
```

