Analyzing Song Popularity Using Spotify's Audio Features

Jennie Sun

11/18/2020

1. SUMMARY

Out of curiosity of different components of music as well as potential associations between music features and song popularity, this analysis investigates the relationship between music data - specifically audio features from the Spotify's database, such as tempo and loudness - and song popularity measured by Spotify's popularity index. The research analyzes over 160, 000 tracks from 1921 to 2020 with a variety of music related features collected from Spotify's database API. The analysis applies methods including but not limited to: preliminary feature screening using exploratory data analysis (EDA), model fitting using proportional odds and multinomial logistic models, model selection using AIC, deviance, and ANOVA test, and model validation using binned residuals plot, accuracy scores, and AUC. The final model shows that there is evidence suggesting certain audio features, such as acousticness and danceability, affect song popularity more than other elements, such as key or mode. This research can serve as a reference to Spotify as they continue to develop their song database and hit song analysis, which could aid in new value creation.

2. INTRODUCTION

This analysis discovers if there is any evidence that certain audio features have a higher influence on song popularity than others and, if so, to what extent. It also provides an estimated range statistics for the most influential features at each popularity level. In addition, it demonstrates evidence showing if the influence of certain features differ by popularity levels, and points out other interesting findings associated with song popularity. Specifically, for example, if a song is more suitable for dance or has more vocal in it, it generally has a higher popularity index on Spotify than a song with lower danceability score or fewer vocal component. A few other continuous and categorical variables also influence song popularity positively or negatively, which will be further explained in the following sections.

3. DATA

This analysis uses the 'Spotify Dataset 1921-2020, 160k+ Tracks' obtained from Kaggle. It contains 169, 909 songs collected from Spotify Web API, which includes data grouped by artists, year, and genre. Each row in the dataset represents a single track, and each column represents a field of the track (audio features and identifiers). Since the goal of the analysis is to identify features related to song popularity, the field `popularity`, a continuous variable with integer values from 0 to 100, is determined to be the response variable and is reassembled to 5 levels - Not Popular (`popularity`=0), Less Popular (0 < `popularity` < 25), Somewhat Popular (25 <= `popularity` < 50), More Popular (50 <= `popularity` < 75), and Popular (75 <= `popularity` <= 100). The factorized response variable `popularity_fac` is therefore used in the following analysis for better interpretebility and minimized effect of unbalanced data distribution based on popularity scores.

In addition, categorical variables `explicit`, `mode`, and `key` are converted into factor variables. Continuous variables `acousticness`, `danceability`, `energy`, `instrumentalness`, `speechiness`, `valence` are releved from a scale of 0 to 1 to a scale of 1 to 100 by multiplying 100, `duration_ms` is converted by `duration_s` by multiplying 1000, and `year` is centered to `year_c` by subtracting 1921 from each observation for easier interpretation. The dataset does not contain any missing values, so no missing value imputations is needed.

EXPLORATORY DATA ANALYSIS (EDA)

Since some of the fields in the Spotify dataset are the unique identifiers or contain repeated information of other variables, including `artists`, `id`, `name`, and `release_date`, they are excluded from this analysis. To start, we look at the number of observations in each level in the response variable `popularity_fac`. It is worth noticing that the last level - Popular - only has fewer than 2000 observations, while the rest of the levels all have more than ten times of this number. Therefore, it is likely that the final model could not fully capture the true relationship between song features in this level and its potential popularity score. The plots of the variables that deemed likely to affect song popularity can be found in the Appendix.

Below is a summary of the EDA for all of the audio features in the data:

`acousticness` describes how acoustic a song is. A score of 100 means the song is most likely to be an acoustic one. The boxplot shows that there is likely an inverse relationship between acousticness and popularity, suggesting that a song is less likely to receive a higher popularity index in Spotify if it is more acoustic. Therefore, this variable will be considered when fitting the model.

`danceability` describes how suitable a track is for dancing based on a combination of musical elements. The boxplot shows that there is likely a positive relationship between danceability and popularity, suggesting that a song is more likely to obtain a higher popularity index in Spotify if it is more suitable for dancing. Therefore, this variable will be considered when fitting the model.

`energy` represents a perceptual measure of intensity and activity. The boxplot shows that there is likely a positive relationship between energy and popularity, suggesting that a song is more likely to obtain a higher popularity index in Spotify if it is measured with a high energy score. Therefore, this variable will be considered when fitting the model.

`instrumentalness` represents the amount of vocals in the song. The boxplot shows that there is likely a weak negative relationship between instrumentalness and popularity, suggesting that a song is less likely to receive a higher popularity index in Spotify if it is more instrumental. However, we should also consider the fact that songs that are more instrumental tend to fall under the Not Popular group, as the plot demonstrates. We will consider this variable when fitting the model.

`liveness` describes the probability that the song was recorded with a live audience. The boxplot displays similar distributions and median values across all five popularity levels, suggesting that `liveness` is less likely to affect song popularity. Therefore, this variable will not be considered when fitting the model.

`speechiness` detects the presence of spoken words in a track. A score below 33 means the song does not have any speech. The boxplot displays, although similar median value, a small variation of distribution across all five popularity levels, suggesting that `speechiness` is likely to affect song popularity. Therefore, this variable will also be considered when fitting the model.

`valence` is a measure describing the musical positiveness conveyed by a track. The boxplot displays similar distributions and median values across all five popularity levels, suggesting that `valence` is less likely to affect song popularity. Therefore, this variable will not be considered when fitting the model.

`duration_s` is the length of the song measured in second. The point plot displays an upward trend in popularity index as the song gets longer. However, majority of the songs are similar in length, with very few outliers that are significantly longer in length than the rest of the observations. Therefore, this variable will be considered when fitting the model although it may turn out to be insignificant at some levels.

`key`: All keys on octave encoded as values ranging from 0 to 11, starting on C as 0, C# as 1 and so on. The boxplot shows some but not obvious variations in popularity distributions and median values across all 12 keys. Therefore, this variable will not be considered when fitting the model.

`loudness` measures average volume over a period of time that typically ranges from -60 to 0. The boxplot shows that there is likely a positive relationship between loudness and popularity, suggesting that a song is likely to receive a higher popularity index in Spotify if it is louder. Therefore, this variable will not be considered when fitting the model.

mode is a binary variable describing the tonality of a song (major or minor scale). The boxplot displays similar distributions and median values for both major and minor songs, suggesting that mode is less likely to affect song popularity. Therefore, this variable will not be considered when fitting the model.

tempo indicates the number of beats in one minute, typically ranges between 50 and 150. The boxplot displays similar distributions and median values across all five popularity levels, suggesting that tempo is less likely to affect song popularity. Therefore, this variable will not be considered when fitting the model either.

explicit: an explicit track is one that has curse words or language that is offensive in nature. The boxplot shows a higher distribution and median popularity index for songs that are marked as explicit, suggesting that a song is likely be more popular if it is an explicit one in Spotify. Therefore, this variable will be considered when fitting the model.

year_c_fac: release year (centered) by decade from 1921 to 2020. The boxplot shows the highest distributions and median popularity index for songs released in most recent years and lowest for those before 1959, suggesting that a song is likely to be more popular if it is newer. Therefore, this variable will be considered when fitting the model.

After exploring a few combinations, no potential interaction is detected with the variables in the dataset. EDA only depicts the visual representation of the data, the actual significance of each variable needs to be assessed by statistical tests. Specifically, if a predictor has been deemed significant through EDA but is dropped through model selection, we will perform ANOVA tests to determine its significance.

4. MODEL

Variable & Model Selection

Since the categories of the response variable has a natural ordering from Not Popular to Popular, I first consider fitting a proportional odds model. From the EDA, the preliminary model consist of the following predictors: acousticness, danceability, energy, instrumentalness, speechiness, duration_s, loudness, explicit, and year_c_fac. This model achieves a Residual Deviance of 249307.7 and an AIC of 249347.7.

Model Assessment

I then perform ANOVA tests on this preliminary model by adding those predictors one at a time to determine the significance of each predictor. All ANOVA tests return a Chi-Square of 0, suggesting that all of the predictors stated above are significant so are worth keeping in the model.

Model Diagnostics & Validation

Next, I look at the overall binned residuals plot on each song popularity level for model diagnostic. The binned residuals plots are used to see if the relationship between the predictors and the response is being captured by looking at the points that fall inside and outside of the red curves. From the 4 plots, it looks like 95% of the observations fall inside the standard error bound for the 2 levels - More Popular and Popular. However, there are still some problems for the other 2 levels - Less Popular and Somewhat Popular.

In order to improve the binned residuals plots for the two other levels, I fit the same predictors to a multinomial logistic model this time because multinomial logistic regression is essentially a generalization of the binomial distribution. This model returns a residual deviance of 239806.5 and an AIC of 239950.5, which are about 10,000 units lower than the proportional odds model. The binned residuals plots, although still not perfect, seem to capture more observations within the 95% standard error bound compared to the ones from the proportional odds, especially for the 2 levels that had problems earlier. The observations inside the red curves are distributed in a more random fashion as well.

Although there are still a few observations that reside outside of the standard error bound, this is possibly the model with the best performance so far without excluding the information of song release year (year_c_fac). I choose to keep this predictor because by just adding this one to the model, the overall deviance residuals and AIC decreased by more than 100,000 units each. A hierarchical proportional odds was also under consideration. However, since the multinomial logistic regression already outperformed the proportional odds model, I will not proceed with exploring another proportional odds option further.

Accuracy and AUC are used as the metrics for model diagnostics. The overall accuracy is only 0.69 because it's hard to do across multiple groups. I also look at the ROC curve for each level of the response variable. The AUCs seem fine for all 4 levels, with the last level - Popular - being the highest. The rest of the three levels all have an AUC of over 80%, suggesting a relatively robust performance of the model.

Final Model

Since both binned residuals plot and AUC curves turn out to be reasonable and convincing for the multinomial logistic model, I decide to use it as my final model.

Final Model Equation:
This multinomial logistic regression is defined as a set of logistic regression models for each probability π_{ij} compared to the baseline, where $j \geq 2$. That is,

$log(\frac{\pi_{ij}}{\pi_{i1}}) = \beta_{0j} + \beta_{1j}x_{i1} + \beta_{2j}x_{i2} + ... + \beta_{pj}x_{ip},\ (j \geq 2)$
where j has five categories/levels with Not Popular as the baseline ($Y = 1$), along with Less Popular, Somewhat Popular, More Popular, and Popular; x_i are predictors `acousticness`, `danceability`, `energy`, `instrumentalness`, `speechiness`, `duration_s`, `loudness`, `explicit`, and `year_c_fac`. p is the number of predictors. It is also worth noticing that since `explicit` and `year_c_fac` are factor variables, they have different β and the number of that depends on the number of levels in the variable.

Therefore, there are $J - 1$, which is 4 here, separate logistic regressions in this analysis.

Model Interpretation

As we can see on the model output summary, almost all of the predictors are significant at p<0.05 level. Duration turns out to be insignificant for 2 levels, both in Somewhat Popular and More Popular. Since there are 17 predictors in each level, and 68 predictors in total, I will only interpret a few variables in each level due to the constraint of the paper length.

`acousticness`: With one unit increase in acousticness score, compared to the odds of being a Not Popular song (level 1/baseline), the odds of being a Less Popular song (level 2) are 1-exp(-0.0225) = 2.22% lower (95% CI: 2.13% to 2.31%); the odds of being a Somewhat Popular song (level 3) are 1-exp(-0.0288) = 2.84% lower (95% CI: 2.73% to 2.95%); the odds of being a More Popular song (level 4) are 1-exp(-3.15) = 2.98% lower (95% CI: 2.86% to 3.10%); the odds of being a Popular song (level 5) are 1-exp(-0.0327) = 3.22% lower (95% CI: 2.97% to 3.45%), holding other variables constant. This is consistent with what the EDA shows: there is an inverse relationship between acousticness and popularity, so a song is less likely to be a popular one as it gets more acoustic.

The same trend can be seen in `energy`, `instrumentalness`, and `speechiness`. Their coefficients are negative for all 4 levels, and are larger in absolute values from Less Popular (level 2) to Popular (level 5). For `explicit`, its coefficients are also negative for all 4 levels, but are smaller in absolute values from Less Popular (level 2) to Popular (level 5), suggesting that compared to not being marked as explicit in Spotify, the odds of an explicit song being a popular one decrease to a lesser extent as the popularity level goes up.

`danceability`: With one unit increase in danceability score, compared to the odds of being a Not Popular song (level 1/baseline), the odds of being a Less Popular song (level 2) are e^0.0138-1 = 1.39% higher (95% CI: 1.27% to 1.52%); the odds of being a Somewhat Popular song (level 3) are e^0.0156-1 = 1.57% higher (95% CI: 1.41% to 1.73%); the odds of being a More Popular song (level 4) are e^0.0220-1 = 2.22% higher (95% CI: 2.04% to 2.41%); the odds of being a Popular song (level 5) are e^0.0417-1 = 4.26% higher (95% CI: 3.91% to 4.60%), holding other variables constant. This is also consistent with what the EDA shows: there is a positive relationship between danceability and popularity, suggesting that a song is more likely to be a popular one if it is more suitable for dancing.

`loudness`: This is an interesting predictor. With one unit increase in loudness scale, compared to being a Not Popular Song (level 1/baseline), the odds of being a Less popular song (level 2) decrease by a little (can be seen from the negative coefficient in this level), but the odds of being a song in the rest of the 3 levels - Somewhat Popular (level 3), More Popular (level 4), and Popular (level 5) - actually increase (can be seen from the positive coefficients in these 3 levels) to a different extent, holding other variables constant. This is also fairly consistent with the EDA: there is a weak positive relationship between loudness and popularity, so a song is likely to be a popular one if it is louder.

If we look at the trends for release year by decade, generally speaking, compared to the odds of being a not popular song that was released between 1921 and 1929, the odds of being a popular one increase as the release year is newer, with some variations for songs in the last 2 to 3 decades, holding other variables constant, which is fairly consistent with the result from the EDA as well.

Duration is only significant at the Popular level, which is also consistent with the EDA. This is likely due to the fact that the data is insufficient for songs that are significant longer in length and the lack of observations in the Popular level, so increasing duration may not have much explanatory power in this analysis.

The intercepts can be interpreted as: compared to being a Not Popular Song (level 1/baseline) released between 1921 and 1929 that are not marked as explicit in Spotify, the odds of being a Less popular song (level 2) are eˆ0.5333-1 = 70.45% higher (95% CI: 70.20% to 70.71%); the odds of being a Somewhat Popular song (level 3) are 1-eˆ(-1.6096) = 80.00% lower (95% CI: 79.96% to 80.04%); the odds of being a More Popular song (level 4) are 1-eˆ(-3.8939) = 97.96% lower (95% CI: 97.96% to 97.97%); the odds of being a Popular song (level 5) are 1-eˆ(-21.6209) = 99.99% lower (95% CI: 99.99% to 99.99%), at average acousticness, danceability, energy, instrumentalness, speechiness, duration, and loudness scale.

5. CONCLUSION

To conclude, the results of the analysis show that following audio features are important in determining song popularity in Spotify: `acousticness`, `danceability`, `energy`, `instrumentalness`, `speechiness`, `explicit`, and release year by decade, as they all turn out to be significant at $p<0.05$ level. Additionally, some of these features affect song popularity differently in different response levels, as elaborated in model interpretation.

It is also important to notice that there are several limitations of this analysis. There are still a few potential outliers identified on the binned residuals plots so there is still room of improvement for model performance in order to better capture the relationship between the predictors and the response. However, given the constraint of time, this is probably the best outcome so far.

Moreover, the dataset is likely to be imbalanced, as Spotify's users base is dominated by Millennials, with over 50 percent aged under 34. Therefore, a few variables could dominate the outcome of the analysis. Also, since Spotify is launched in the United States in 2011, which was also when streaming services started to became popular, songs prior to this date may attract fewer listeners who are already used to listening music on other platforms, which may affect the popularity index subsequently.

In addition, the data is likely to be insufficient and biased due to the following reasons:

First, the response variable - popularity index - is the popularity metric determined by Spotify. So data from other music streaming platforms need to be considered for a more exhaustive analysis. Second, it is possible that a high popularity index is only calculated from a few listeners feedback or activities. Therefore, it is important that future analysis can also take other influencers, such as stream count, into consideration. Third, not everyone who listens to a song would actually rate the song, so the ratings information is likely to be insufficient in determining popularity.

Last but not least, each genre has its own popular characteristics so perhaps a more comprehensive analysis including genre's information will lead to a better model and understanding of the relationships in the data.

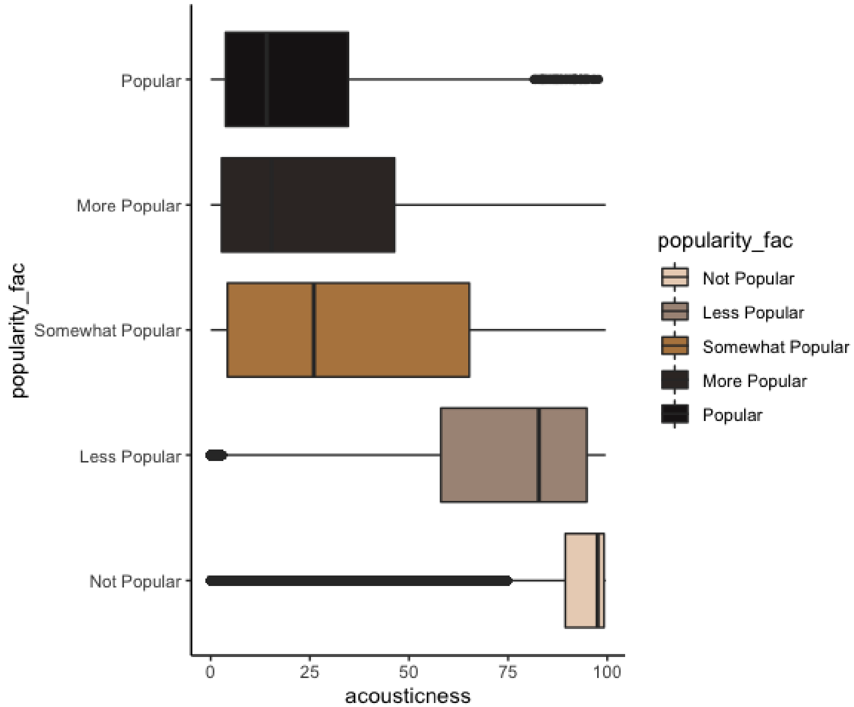
6. ACKNOWLEDGMENTS

I would like to express my deep gratitude to Professor Dr. Olanrewaju Michael Akande for his patient guidance, enthusiastic encouragement and constructive feedback of this project. Without his assistance and support I could not have chosen a project that was so fascinating to work on.

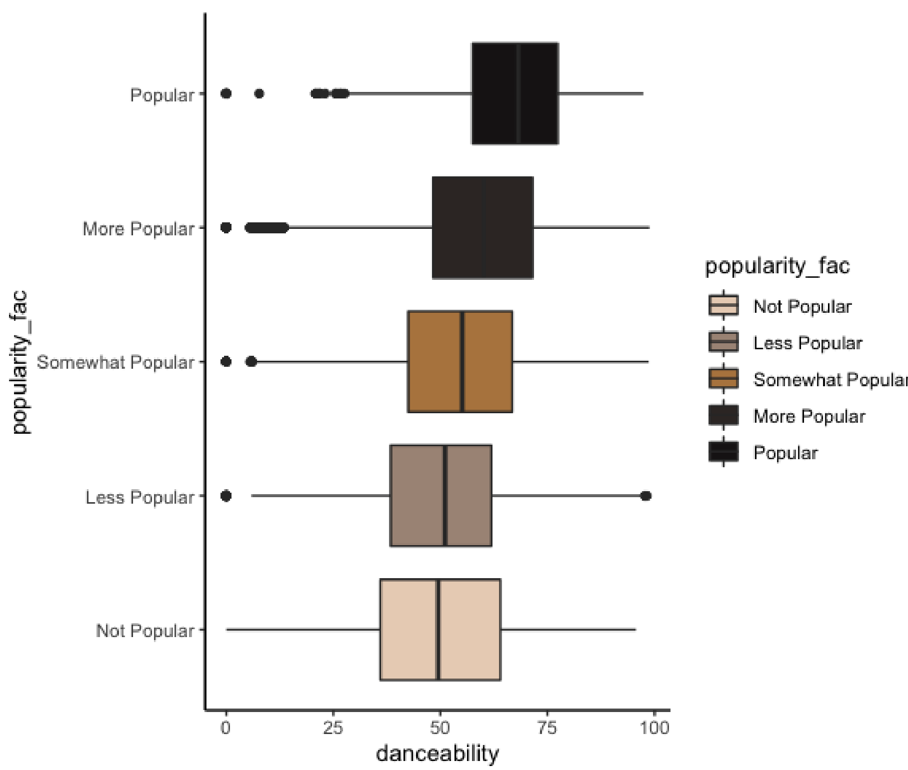
APPENDIX

EDA

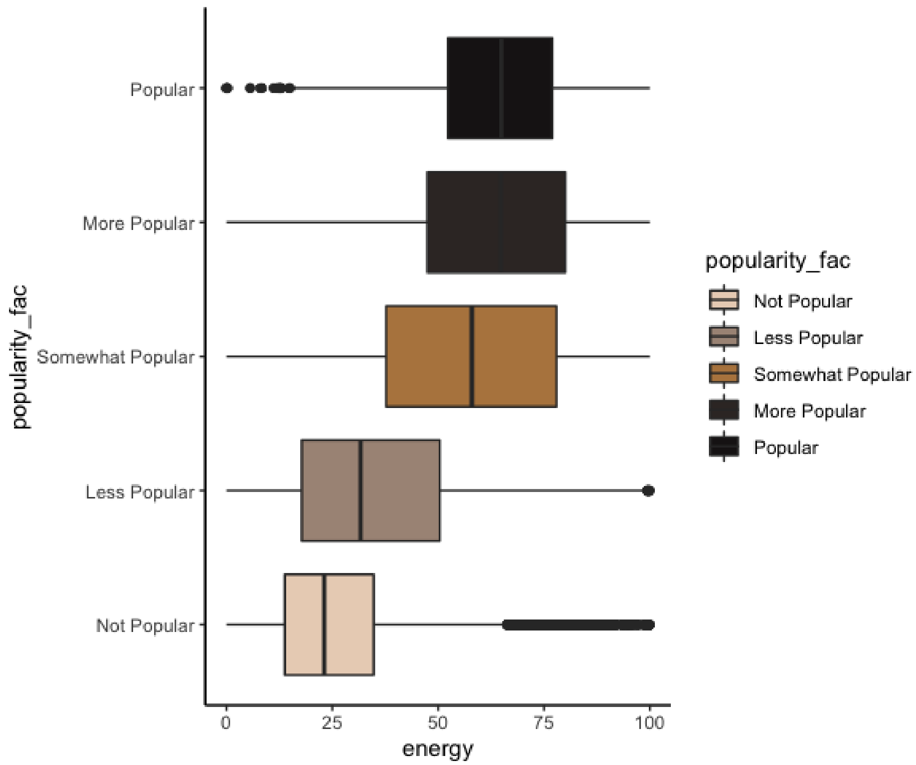
popularity vs acousticness



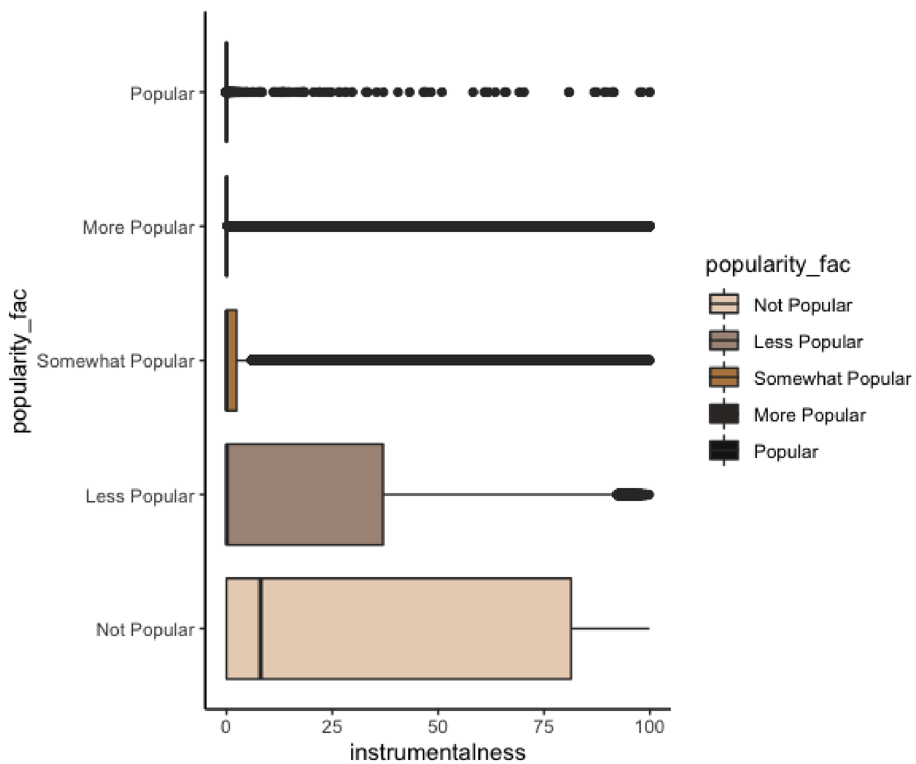
popularity vs danceability



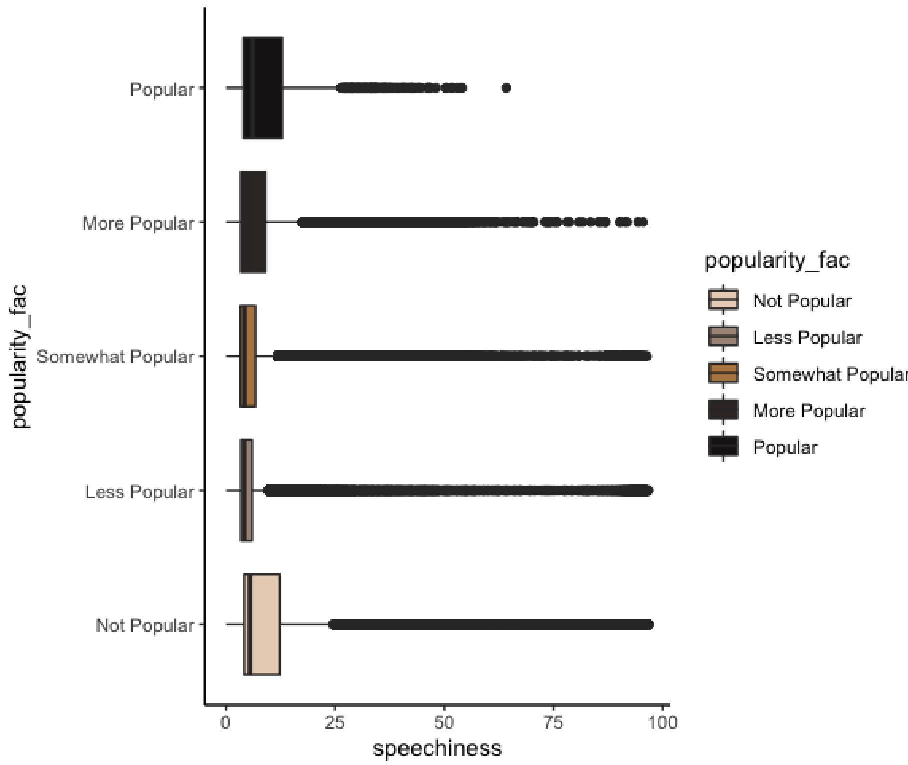
popularity vs energy



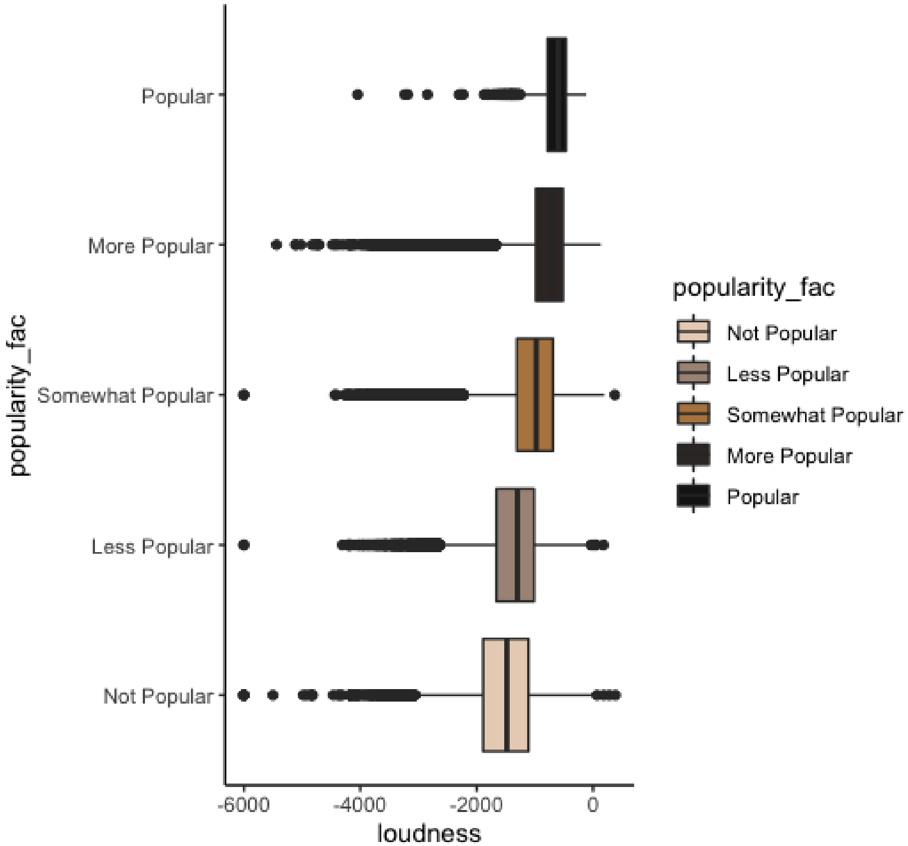
popularity vs instrumentalness



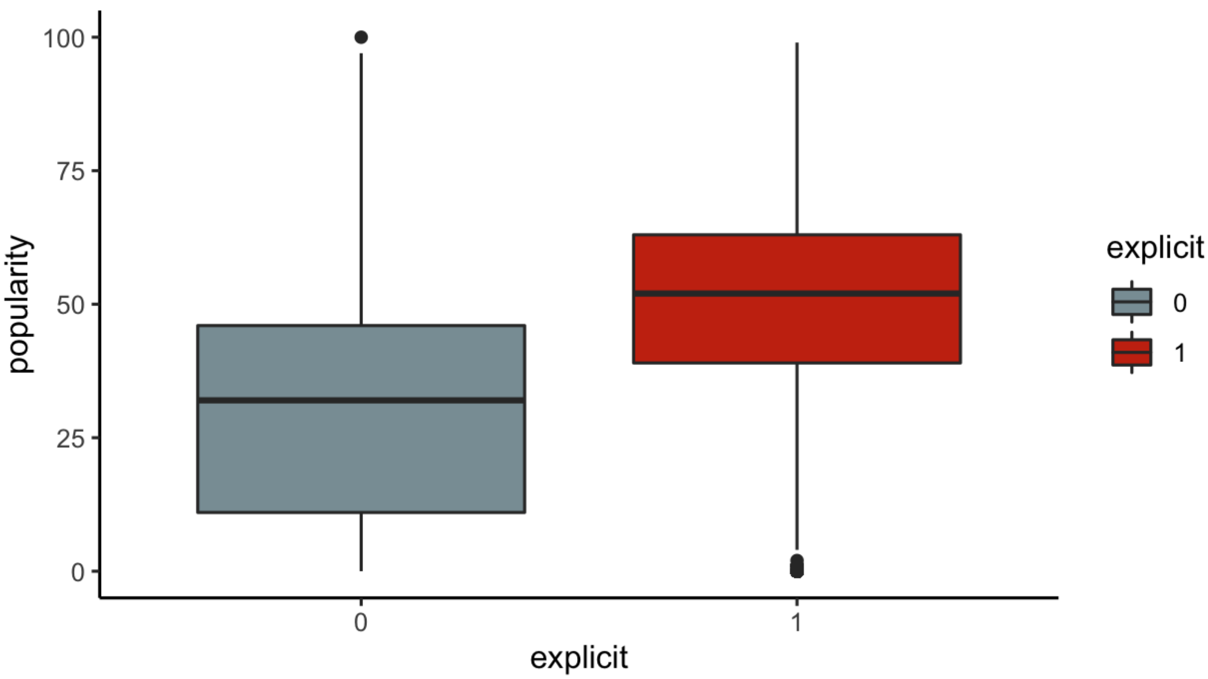
popularity vs speechiness



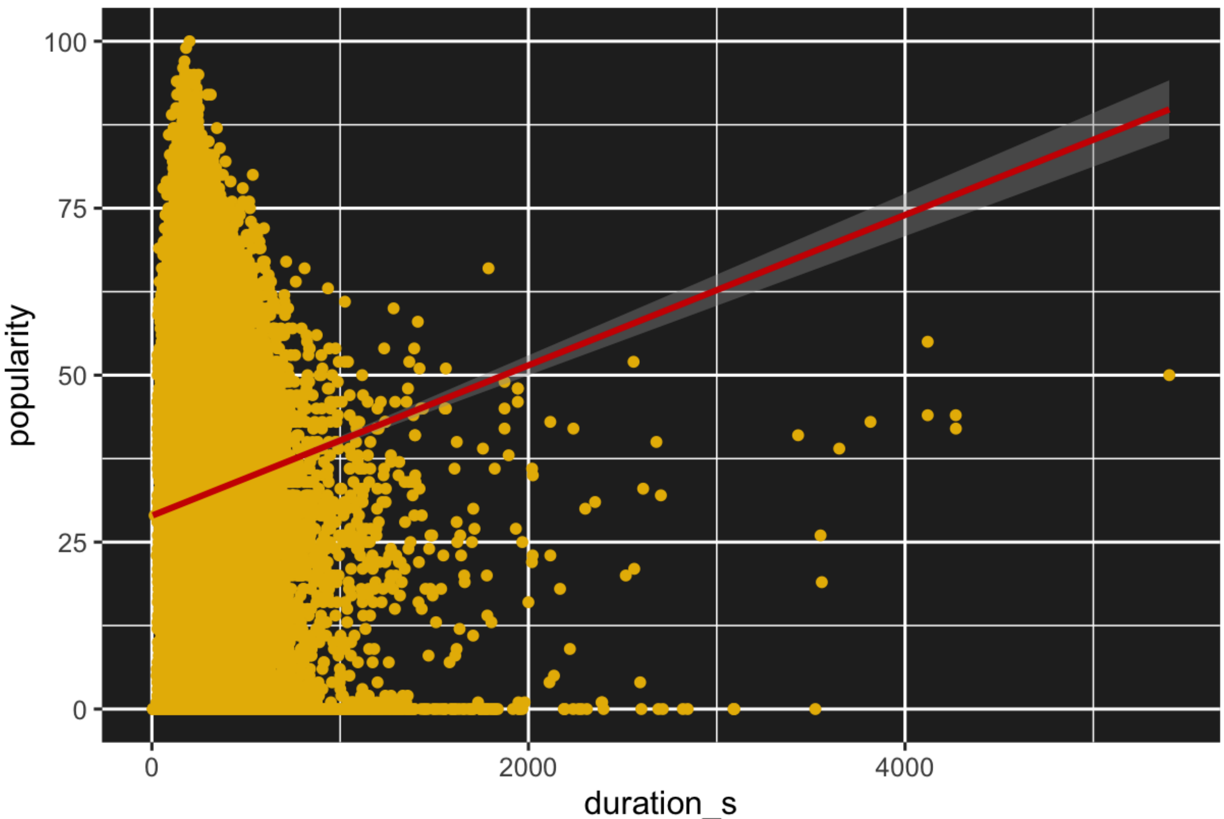
popularity vs loudness



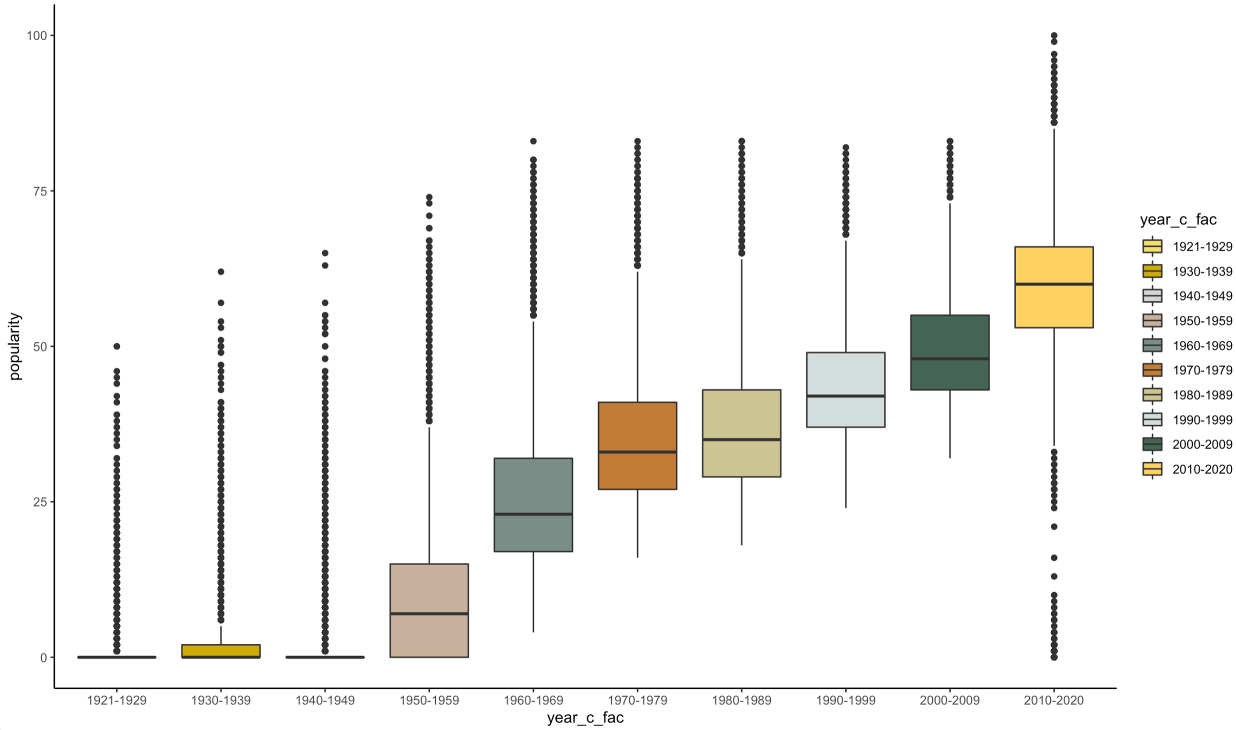
popularity vs explicit



popularity vs duration



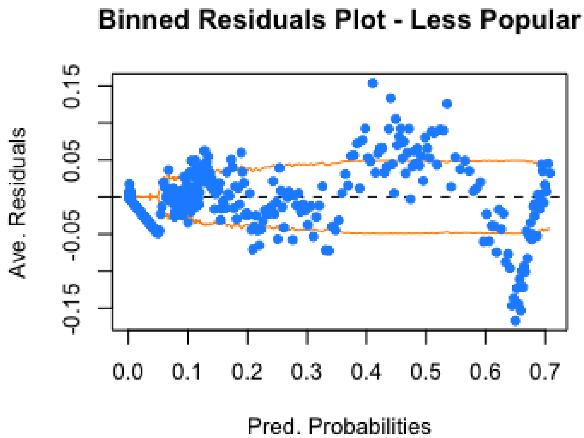
popularity vs year by decade

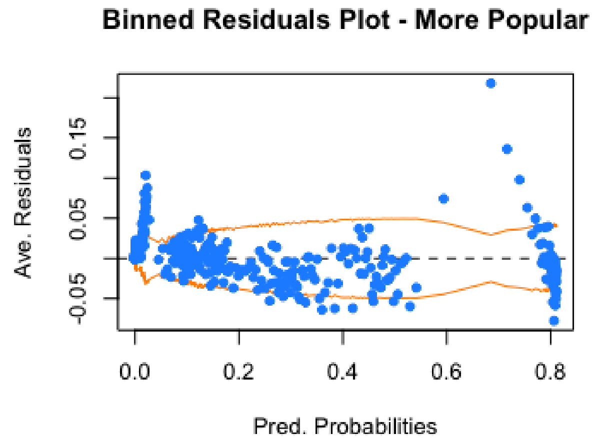


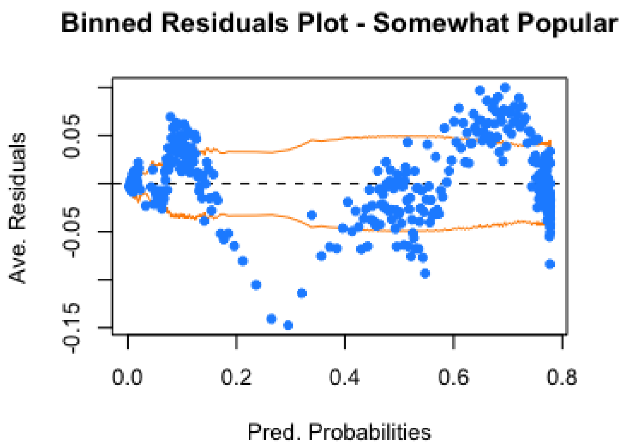
Model Diagnostics

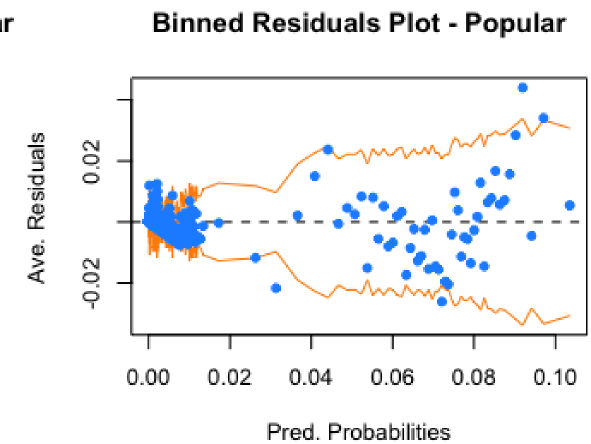
Porportional Odds Model

Binned Residuals Plots:



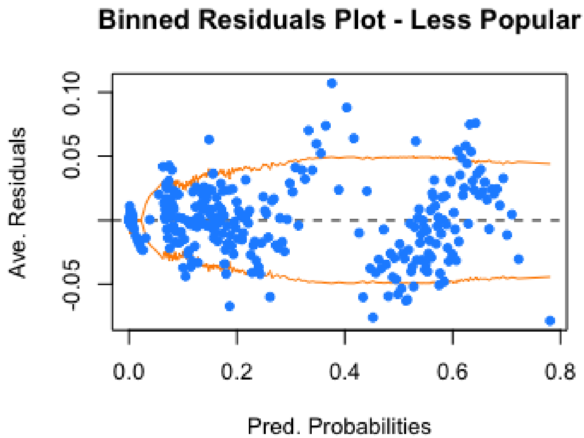


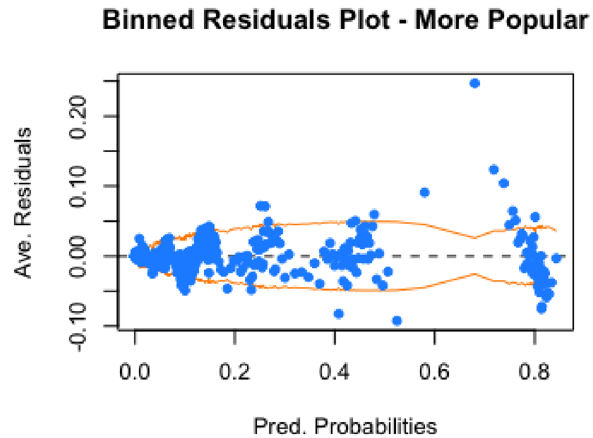


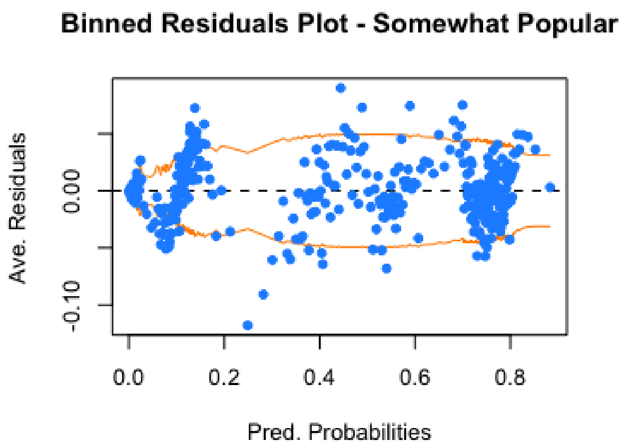


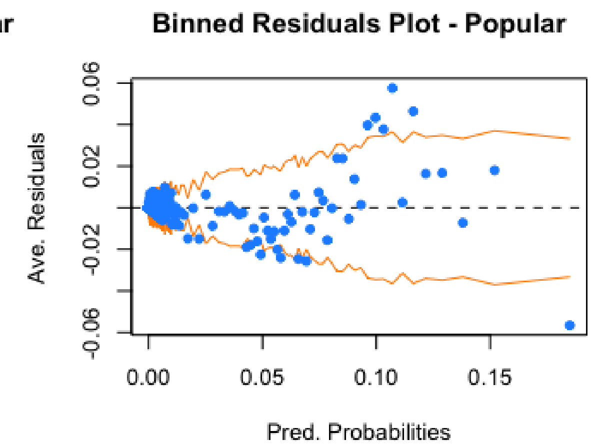
Multinomial Logistic Model (Final Model)

Binned Residuals Plots:

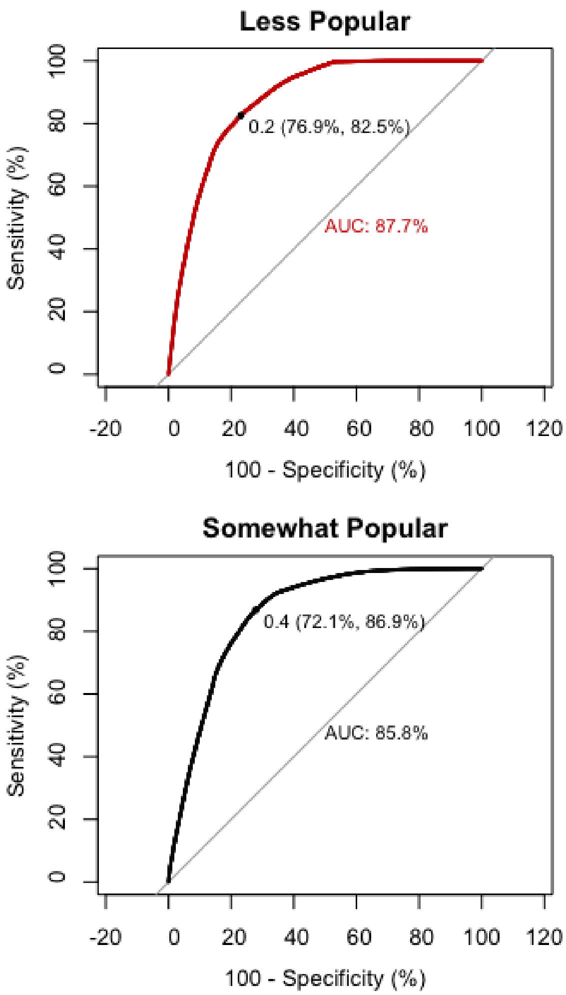








Individual ROC curves for the different levels:




```
# weights:  95 (72 variable)
initial  value 273457.986264
iter  10 value 202315.641796
iter  20 value 196931.556747
iter  30 value 163138.368365
iter  40 value 152101.328851
iter  50 value 147235.108912
iter  60 value 133771.443336
iter  70 value 128117.596060
iter  80 value 120615.134920
iter  90 value 119980.348845
iter 100 value 119922.194205
iter 110 value 119906.694418
iter 120 value 119905.416780
iter 130 value 119905.183303
iter 140 value 119903.526962
final  value 119903.229600
converged
```

```
 Accuracy
0.6931769
```

	Sensitivity	Specificity
Class: Not Popular	0.7822495	0.9485521
Class: Less Popular	0.6280248	0.8839623
Class: Somewhat Popular	0.8175104	0.7617465
Class: More Popular	0.4853747	0.9609761
Class: Popular	0.0000000	1.0000000

Final Model Output

y.level	term	estimate	std.error	statistic	p.value	conf.low	conf.high
Less Popular	(Intercept)	0.53	0.00	687.34	0.00	0.53	0.53
Less Popular	acousticness	-0.02	0.00	-47.32	0.00	-0.02	-0.02
Less Popular	danceability	0.01	0.00	21.27	0.00	0.01	0.02
Less Popular	energy	-0.01	0.00	-12.17	0.00	-0.01	-0.01
Less Popular	instrumentalness	0.00	0.00	-12.93	0.00	0.00	0.00
Less Popular	speechiness	-0.02	0.00	-30.65	0.00	-0.02	-0.02
Less Popular	duration_s	0.00	0.00	-2.57	0.01	0.00	0.00
Less Popular	loudness	0.00	0.00	-3.14	0.00	0.00	0.00
Less Popular	explicit1	-2.95	0.00	-12590.56	0.00	-2.95	-2.95
Less Popular	year_c_fac1930-1939	0.49	0.02	24.36	0.00	0.45	0.53
Less Popular	year_c_fac1940-1949	0.18	0.02	9.99	0.00	0.15	0.22
Less Popular	year_c_fac1950-1959	1.91	0.01	135.43	0.00	1.89	1.94
Less Popular	year_c_fac1960-1969	13.30	0.01	1080.86	0.00	13.28	13.32
Less Popular	year_c_fac1970-1979	12.94	0.01	971.93	0.00	12.92	12.97
Less Popular	year_c_fac1980-1989	14.02	0.02	914.66	0.00	13.99	14.05
Less Popular	year_c_fac1990-1999	5.75	0.00	1672417.49	0.00	5.75	5.75
Less Popular	year_c_fac2000-2009	0.53	0.00	44119896.74	0.00	0.53	0.53
Less Popular	year_c_fac2010-2020	1.90	0.00	6782.14	0.00	1.90	1.90
Somewhat Popular	(Intercept)	-1.61	0.00	-1603.12	0.00	-1.61	-1.61
Somewhat Popular	acousticness	-0.03	0.00	-50.67	0.00	-0.03	-0.03
Somewhat Popular	danceability	0.02	0.00	19.16	0.00	0.01	0.02
Somewhat Popular	energy	-0.01	0.00	-14.51	0.00	-0.01	-0.01
Somewhat Popular	instrumentalness	-0.01	0.00	-14.17	0.00	-0.01	-0.01
Somewhat Popular	speechiness	-0.04	0.00	-30.05	0.00	-0.04	-0.04
Somewhat Popular	duration_s	0.00	0.00	-1.18	0.24	0.00	0.00
Somewhat Popular	loudness	0.00	0.00	3.97	0.00	0.00	0.00
Somewhat Popular	explicit1	-2.10	0.01	-150.33	0.00	-2.12	-2.07
Somewhat Popular	year_c_fac1930-1939	0.68	0.00	837.40	0.00	0.68	0.68
Somewhat Popular	year_c_fac1940-1949	0.07	0.00	136.39	0.00	0.07	0.07
Somewhat Popular	year_c_fac1950-1959	3.20	0.02	202.34	0.00	3.17	3.24
Somewhat Popular	year_c_fac1960-1969	15.90	0.01	1255.67	0.00	15.88	15.93
Somewhat Popular	year_c_fac1970-1979	17.30	0.01	1609.25	0.00	17.28	17.32
Somewhat Popular	year_c_fac1980-1989	18.95	0.01	1661.06	0.00	18.92	18.97
Somewhat Popular	year_c_fac1990-1999	18.01	0.01	2126.23	0.00	17.99	18.02
Somewhat Popular	year_c_fac2000-2009	18.05	0.01	2258.62	0.00	18.04	18.07
Somewhat Popular	year_c_fac2010-2020	8.10	0.01	829.02	0.00	8.08	8.12
More Popular	(Intercept)	-3.89	0.00	-3640.19	0.00	-3.90	-3.89
More Popular	acousticness	-0.03	0.00	-47.71	0.00	-0.03	-0.03
More Popular	danceability	0.02	0.00	24.61	0.00	0.02	0.02
More Popular	energy	-0.02	0.00	-19.50	0.00	-0.02	-0.01
More Popular	instrumentalness	-0.01	0.00	-18.01	0.00	-0.01	-0.01
More Popular	speechiness	-0.05	0.00	-33.62	0.00	-0.05	-0.05
More Popular	duration_s	0.00	0.00	0.06	0.95	0.00	0.00

y.level	term	estimate	std.error	statistic	p.value	conf.low	conf.high
More Popular	loudness	0.00	0.00	11.59	0.00	0.00	0.00
More Popular	explicit1	-1.84	0.02	-122.41	0.00	-1.87	-1.81
More Popular	year_c_fac1930-1939	0.56	0.00	14521.53	0.00	0.56	0.56
More Popular	year_c_fac1940-1949	0.63	0.00	13619.63	0.00	0.63	0.63
More Popular	year_c_fac1950-1959	3.59	0.00	4344.86	0.00	3.59	3.60
More Popular	year_c_fac1960-1969	16.54	0.02	894.53	0.00	16.50	16.57
More Popular	year_c_fac1970-1979	18.06	0.01	1288.03	0.00	18.03	18.09
More Popular	year_c_fac1980-1989	19.80	0.01	1437.68	0.00	19.78	19.83
More Popular	year_c_fac1990-1999	19.48	0.01	2308.92	0.00	19.46	19.50
More Popular	year_c_fac2000-2009	20.34	0.01	2551.20	0.00	20.32	20.35
More Popular	year_c_fac2010-2020	12.44	0.01	1203.74	0.00	12.42	12.46
Popular	(Intercept)	-21.62	0.00	-408204.08	0.00	-21.62	-21.62
Popular	acousticness	-0.03	0.00	-26.06	0.00	-0.04	-0.03
Popular	danceability	0.04	0.00	24.92	0.00	0.04	0.04
Popular	energy	-0.03	0.00	-22.52	0.00	-0.04	-0.03
Popular	instrumentalness	-0.02	0.00	-7.68	0.00	-0.02	-0.01
Popular	speechiness	-0.05	0.00	-19.52	0.00	-0.06	-0.05
Popular	duration_s	0.00	0.00	-6.74	0.00	0.00	0.00
Popular	loudness	0.00	0.00	18.79	0.00	0.00	0.00
Popular	explicit1	-1.71	0.00	-1380.04	0.00	-1.71	-1.71
Popular	year_c_fac1930-1939	13.40	0.00	317002370.92	0.00	13.40	13.40
Popular	year_c_fac1940-1949	13.62	0.00	278368998.11	0.00	13.62	13.62
Popular	year_c_fac1950-1959	11.47	0.00	3278377280.35	0.00	11.47	11.47
Popular	year_c_fac1960-1969	32.70	0.00	307052.57	0.00	32.70	32.70
Popular	year_c_fac1970-1979	34.58	0.00	293985.33	0.00	34.58	34.58
Popular	year_c_fac1980-1989	36.04	0.00	219077.21	0.00	36.04	36.04
Popular	year_c_fac1990-1999	34.86	0.00	420776.60	0.00	34.86	34.86
Popular	year_c_fac2000-2009	35.72	0.00	190199.70	0.00	35.72	35.72
Popular	year_c_fac2010-2020	29.40	0.00	56057.05	0.00	29.39	29.40