Spotify Genre Classification

STATISTICAL LEARNING PROJECT - FELICE FRANCARIO

INTRODUCTION

- **OBJECTIVE:** Analyze the diverse landscape of music genres and develop a classification model capable of predicting the genre of a song based on its audio features.

DATASET

- The dataset was downloaded from Kaggle in a CSV format
- 21 columns, 11400 rows
- No missing values
- 5 nominal columns were removed
  - 16 columns remaining
    - 15 features and 1 response variable (genre)

Description of Features

- **popularity**: A score indicating how popular a track is on Spotify (ranging from 0 to 100).
- **duration_ms**: The duration of a track in milliseconds.
- **explicit**: Indicates whether a track contains explicit content (True or False).
- **Danceability**: Danceability measures how suitable a track is for dancing, ranging from 0 to 1. Tracks with high danceability scores are more energetic and rhythmic, making them ideal for dancing.
- **Energy**: Energy represents intensity and activity within a song on a scale from 0 to 1. Tracks with high energy tend to be more fast-paced and intense.

Description of Features

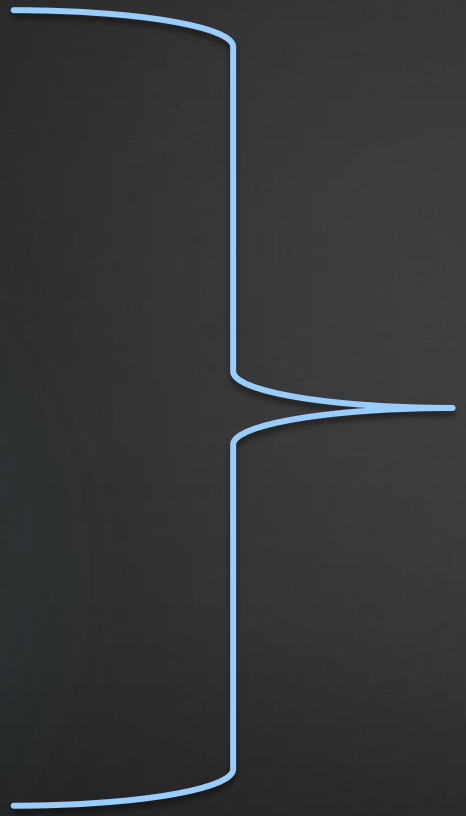
- **Loudness:** Loudness indicates how loud or quiet an entire song is in decibels (dB). Positive values represent louder songs while negative values suggest quieter ones.
- **Key:** Key refers to different musical keys assigned integers ranging from 0-11, with each number representing a different key. Knowing the key can provide insights into the mood and tone of a song.
- **mode:** The tonal mode of the track, represented by an integer value (0 for minor, 1 for major)
- **speechiness**: A score ranging from 0 to 1 that represents the presence of spoken words in a track.
- **acousticness**: A score ranging from 0 to 1 that represents the extent to which a track possesses an acoustic quality.

Description of Features

- **instrumentalness:** A score ranging from 0 to 1 that represents the likelihood of a track being instrumental.
- **Liveness:** A score ranging from 0 to 1 that represents the presence of an audience during the recording or performance of a track.
- **Valence:** Valence measures the musical positiveness conveyed by a track, ranging from 0 to 1. High valence values indicate more positive or happy tracks, while lower values suggest more negative or sad ones.
- **Tempo:** Tempo is the speed or pace of a song in beats per minute (BPM). It gives an idea about how fast or slow a track is.
- **time_signature**: The number of beats within each bar of the track.

Track Genres

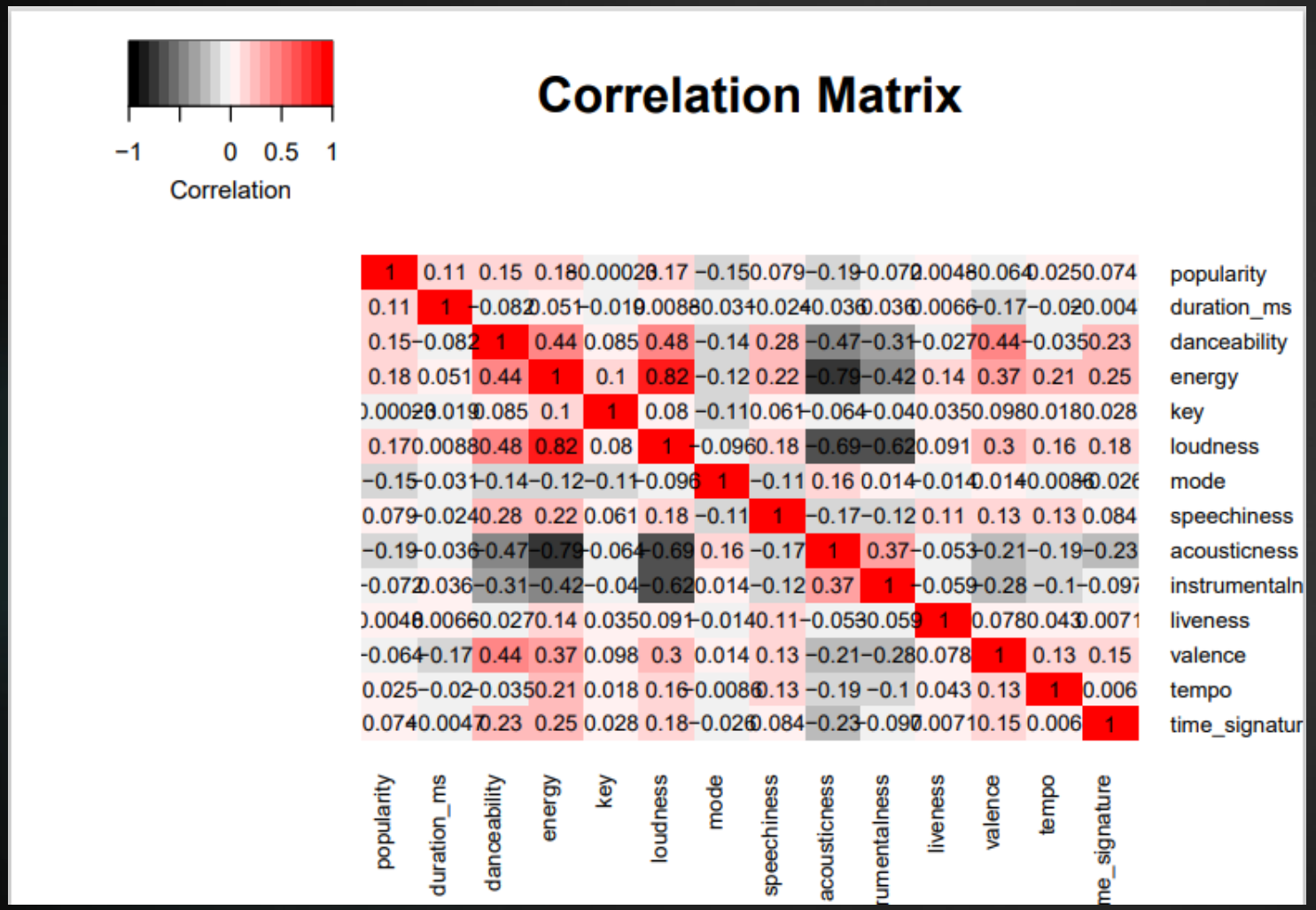
- Classical
- Country
- Electronic
- Hip-Hop
- Jazz
- Rock
- Pop
- Blues
- Reggae



9 Genres Selected

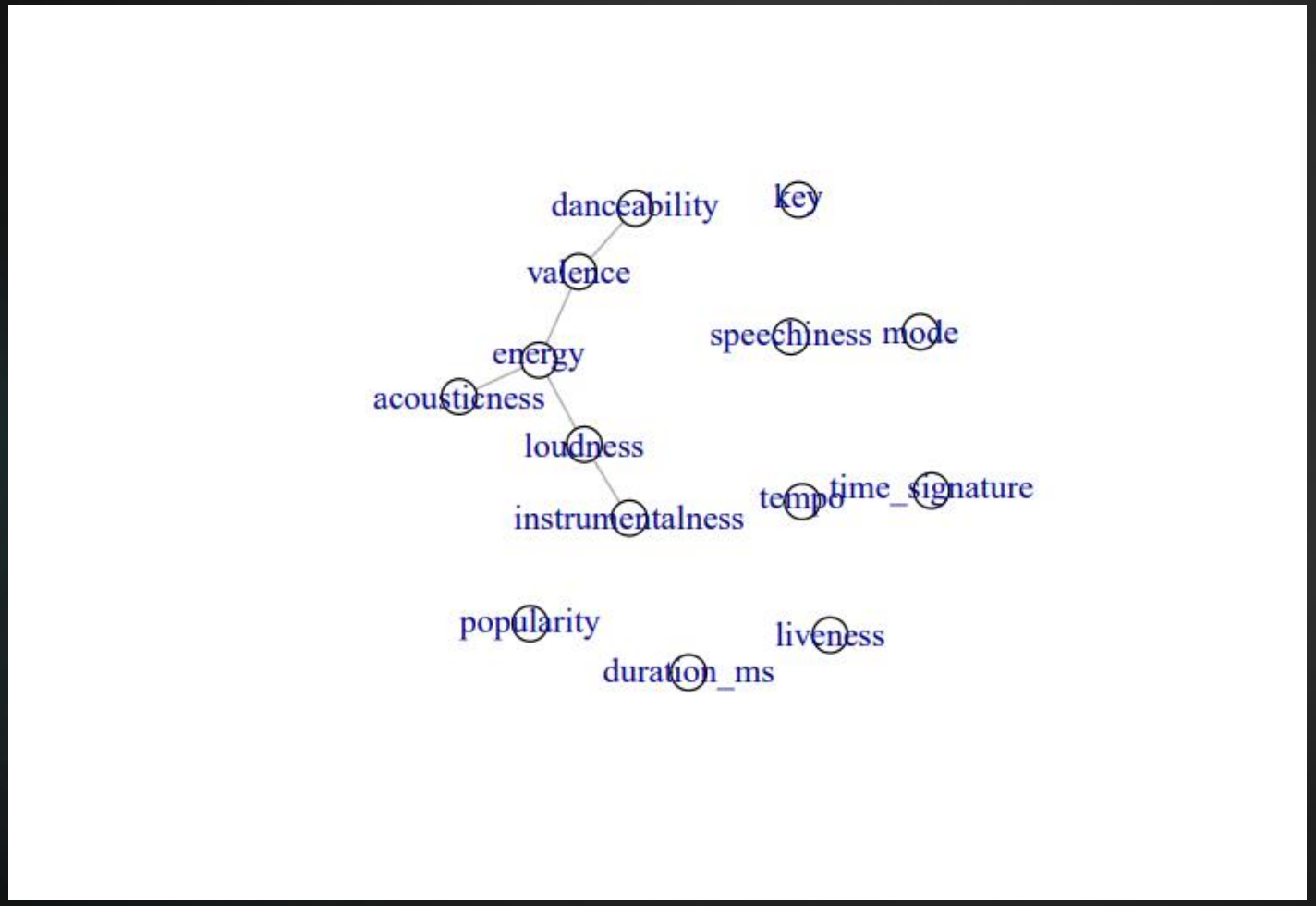
EXPLORATORY DATA ANALYSIS

Correlation Matrix



- Danceability, energy and loudness are significantly positively correlated with each other and negatively correlated with acousticness, instrumentalness
- Very high correlation of **0.82** between **energy** and **loudness.**
- Very high correlation between **energy** and **acousticness** of **-0.76**

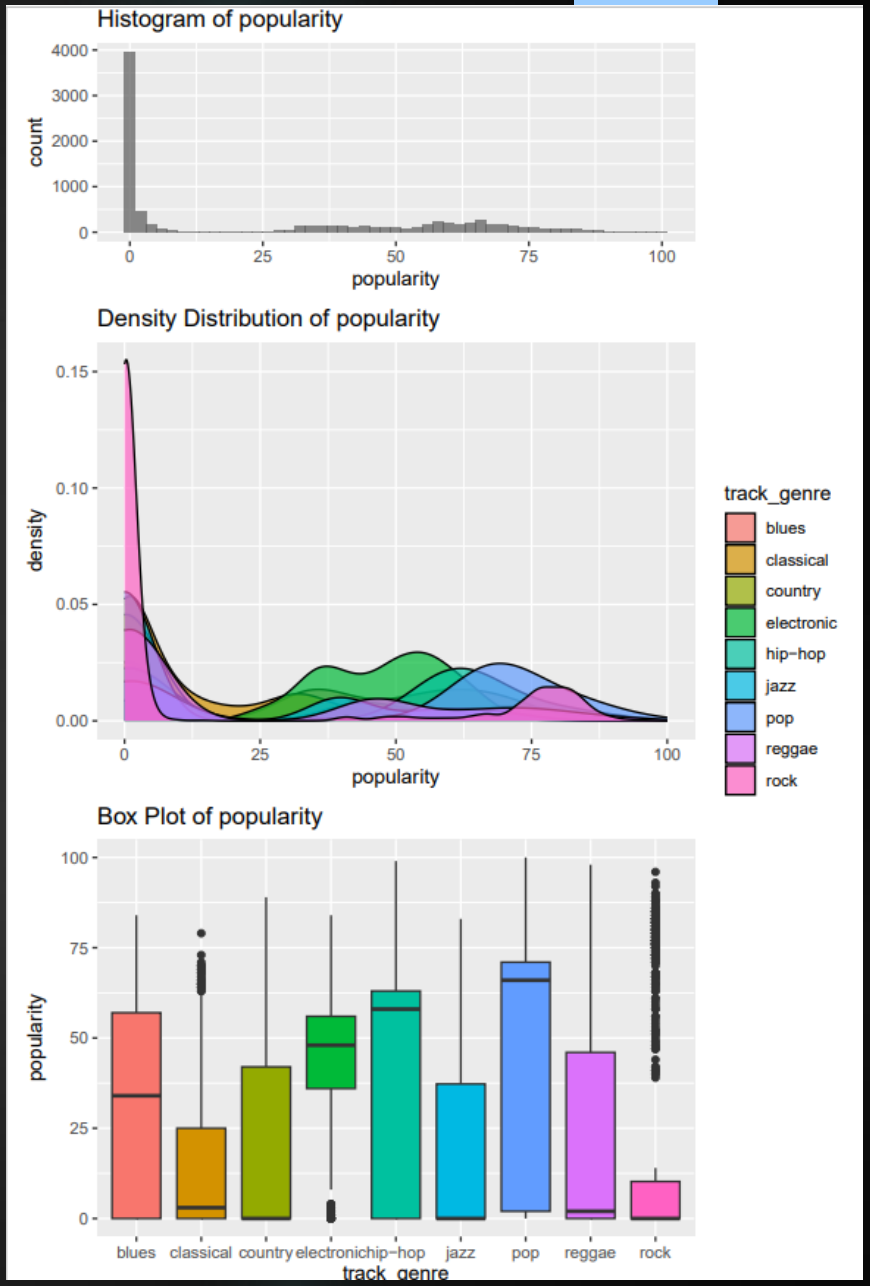
Partial correlation Graph



Threshold=0.3

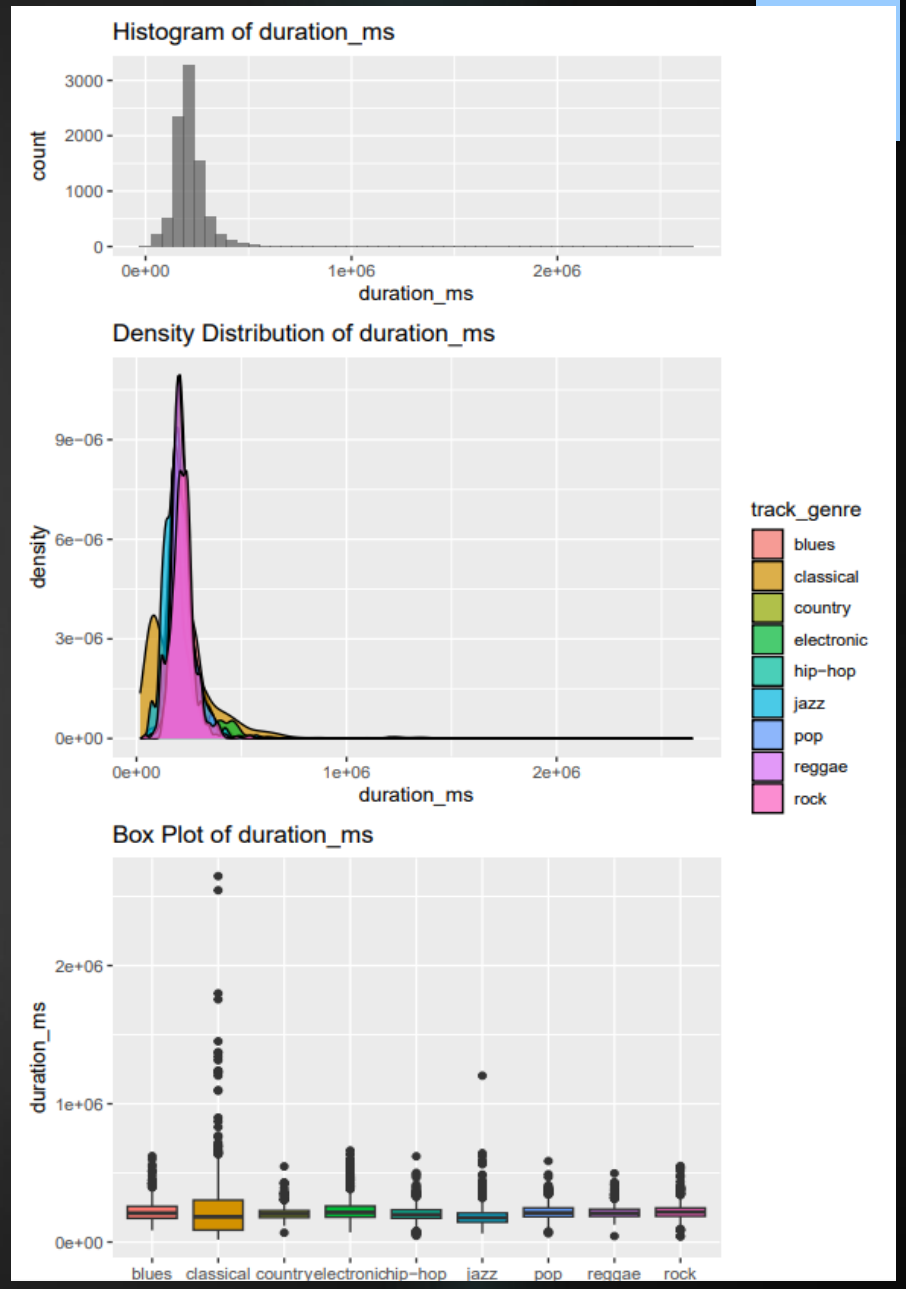
Popularity

- **Least popular genres**: Classical,country, jazz and reggae .
- The majority of the rock tracks have a particularly low popularity score , but with a lot of outliers with a very high score.
- **Most popular genres** :pop,hip-hop and electronic.
- Hip-hop and pop have the highest spread in popularity values,with a a very high median value but left skewed distributions



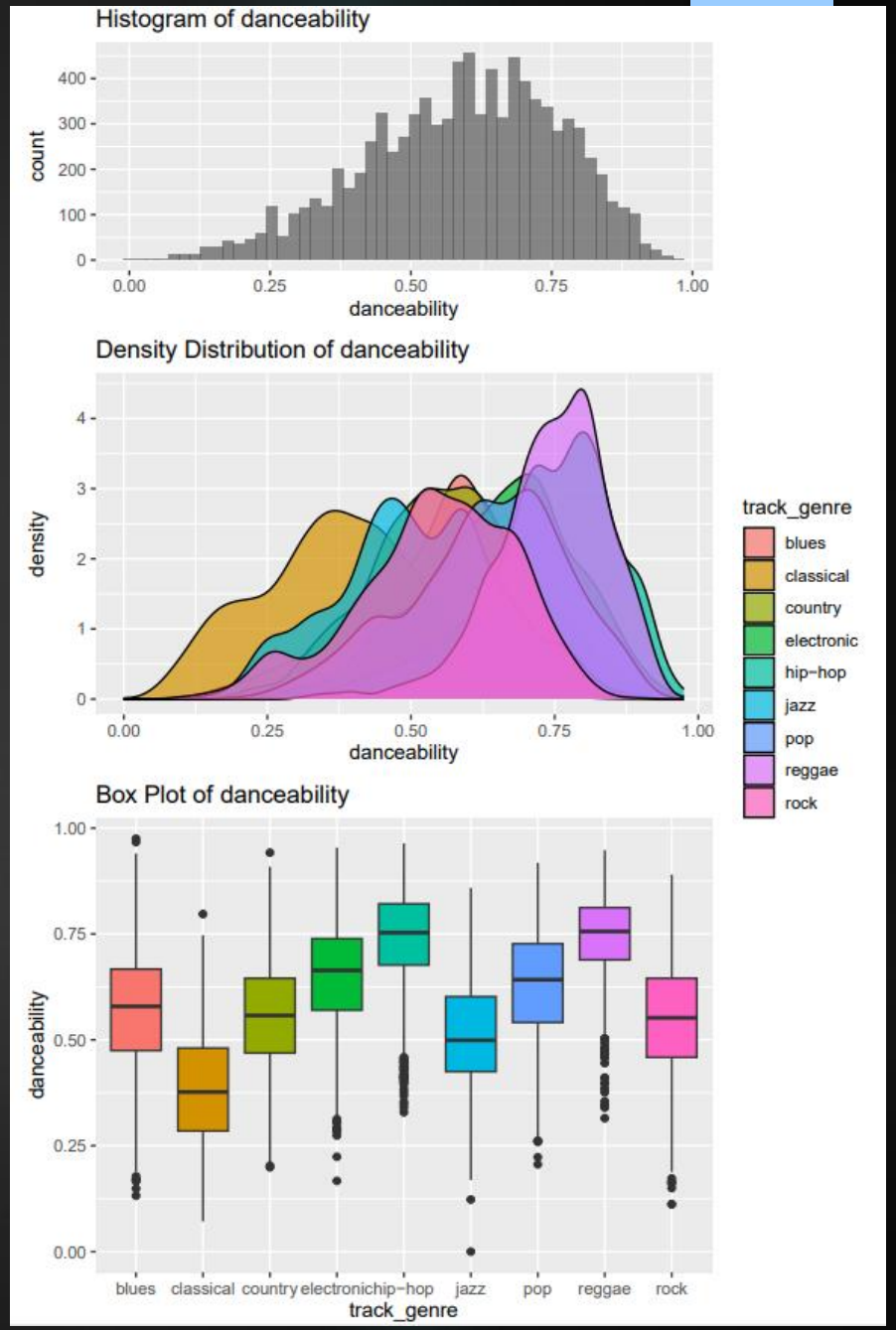
Duration (ms)

- All the genres have a relatively similar distribution except the **classical** genre which has the **highest variance and lowest median value.**



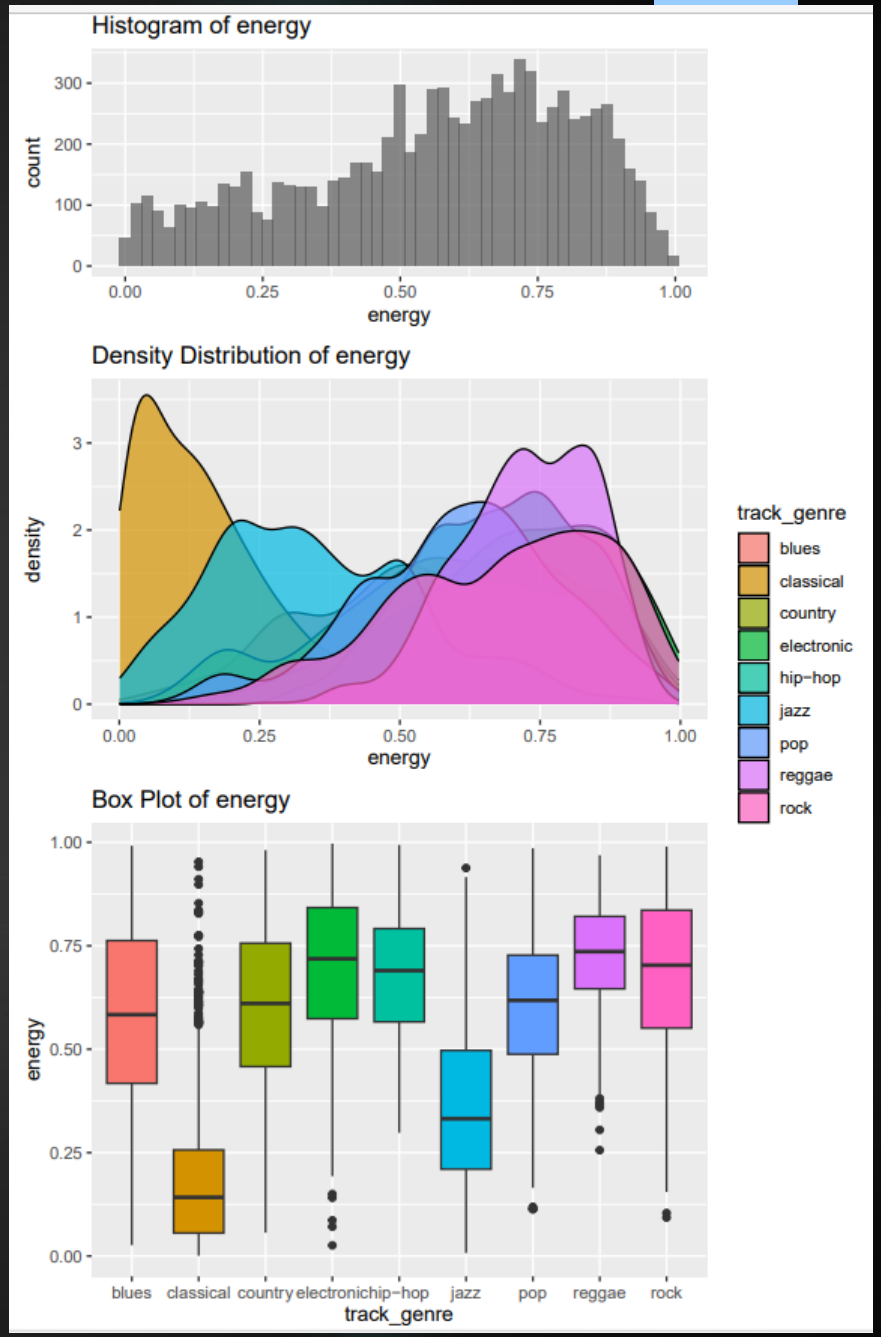
Danceability

- Highest Danceability: Hip-Hop and Reggae
- Hip-Hop has a slightly higher concentration of tracks near the very maximum value of danceability.
- Lowest Danceability: Classical and Jazz.



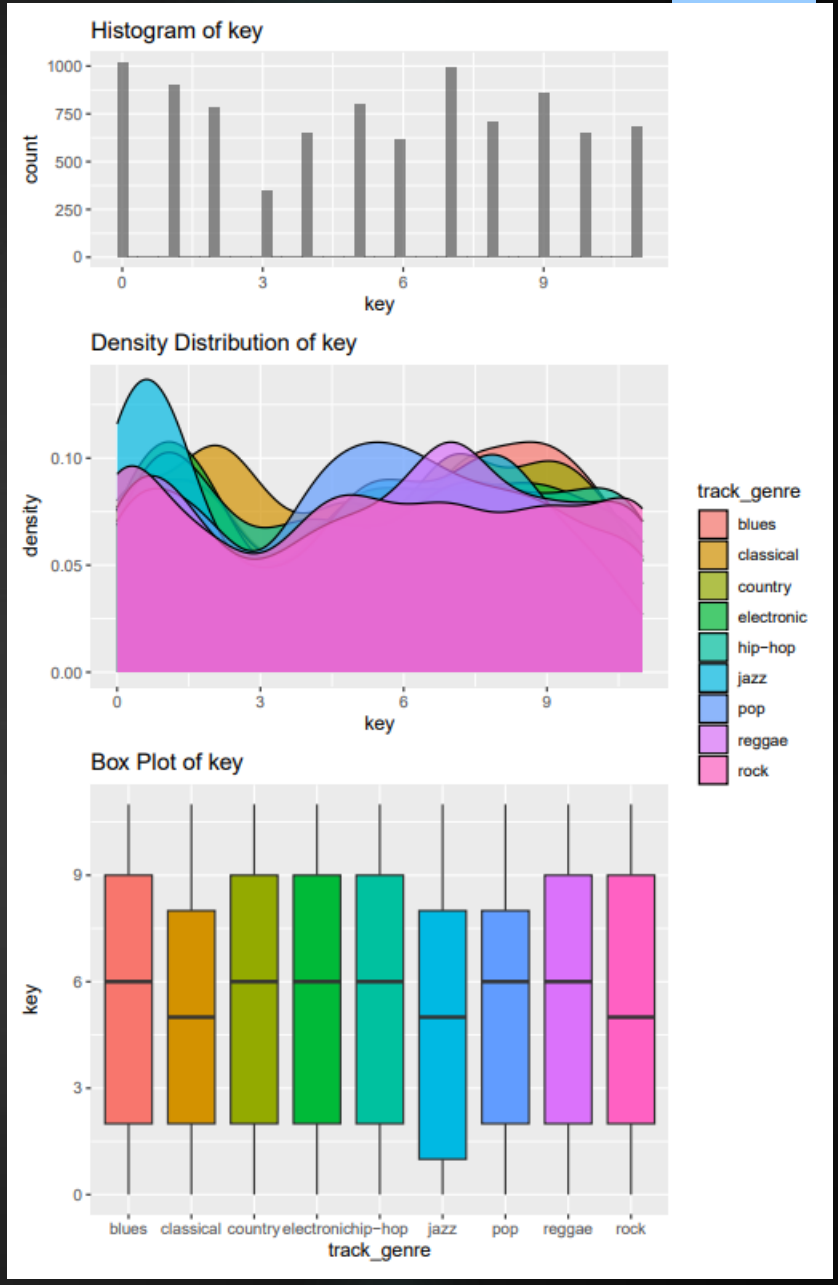
Energy

- The distributions of the energy scores are generally **similar to danceability** but with even lower scores for classical and jazz tracks.
- The main differences are that electronic tracks have a higher energy score than hip-hop and rock has a generally much higher energy score compared to danceability.



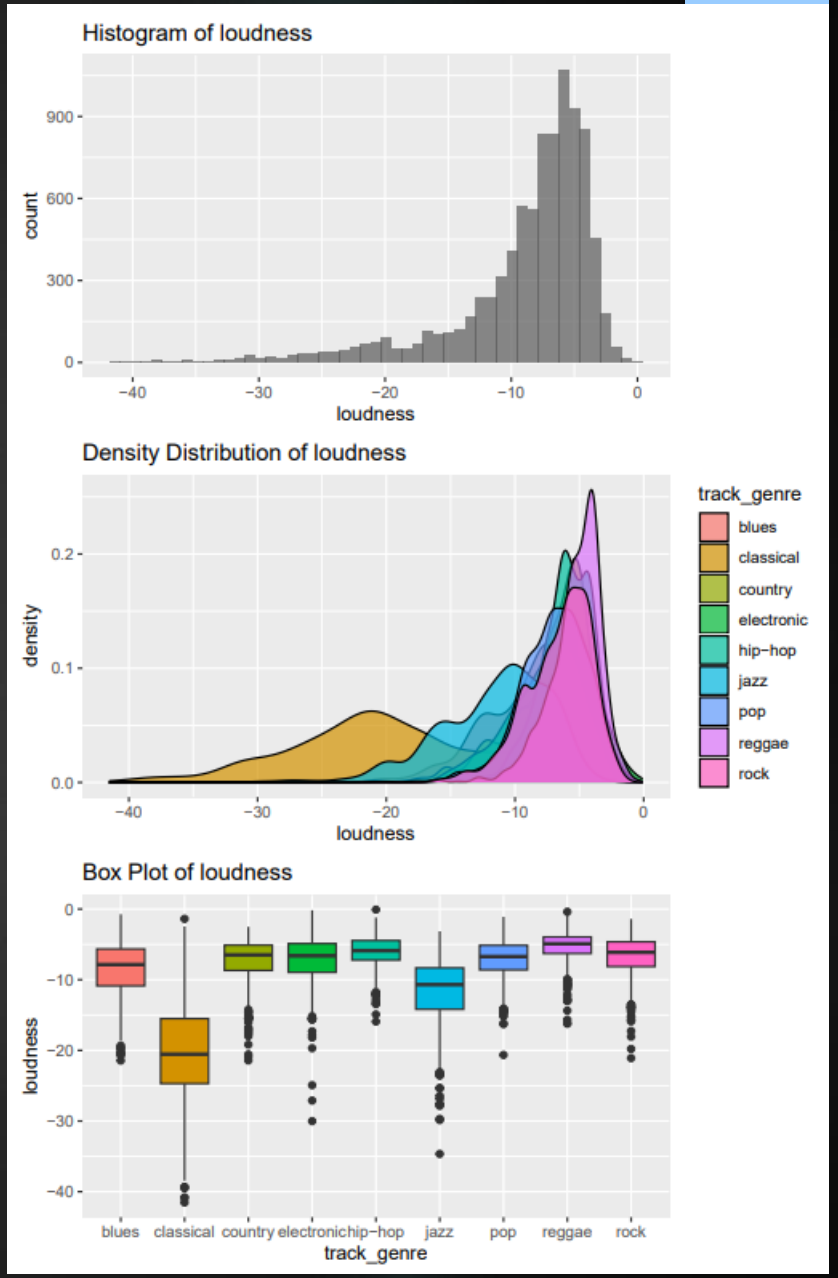
Key

- Visually doesn't have a big effect on the differences between the genres.
- Lowest Key median values: jazz,Classical,Rock



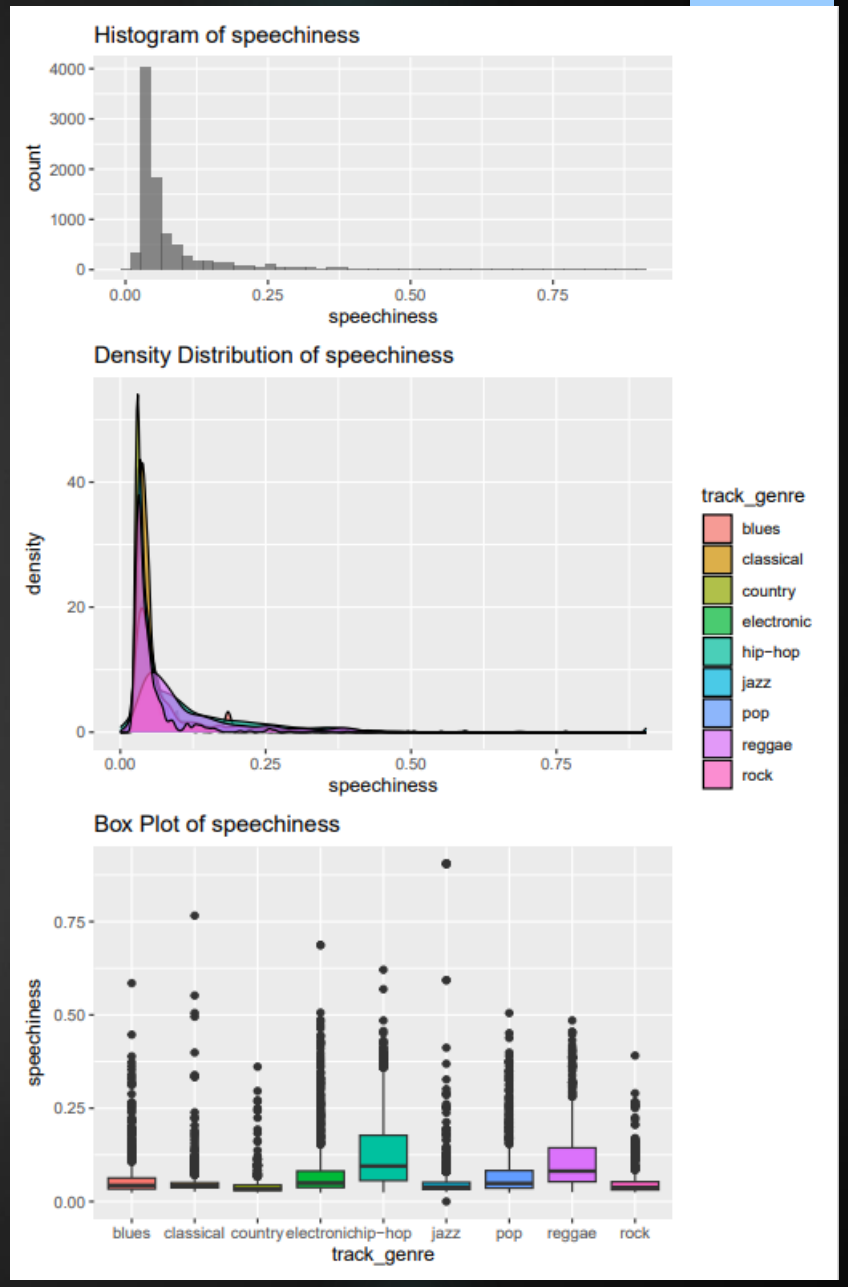
Loudness

- Loudest: Reggae and Hip-Hop
- Similar distribution to Energy(Highly correlated)



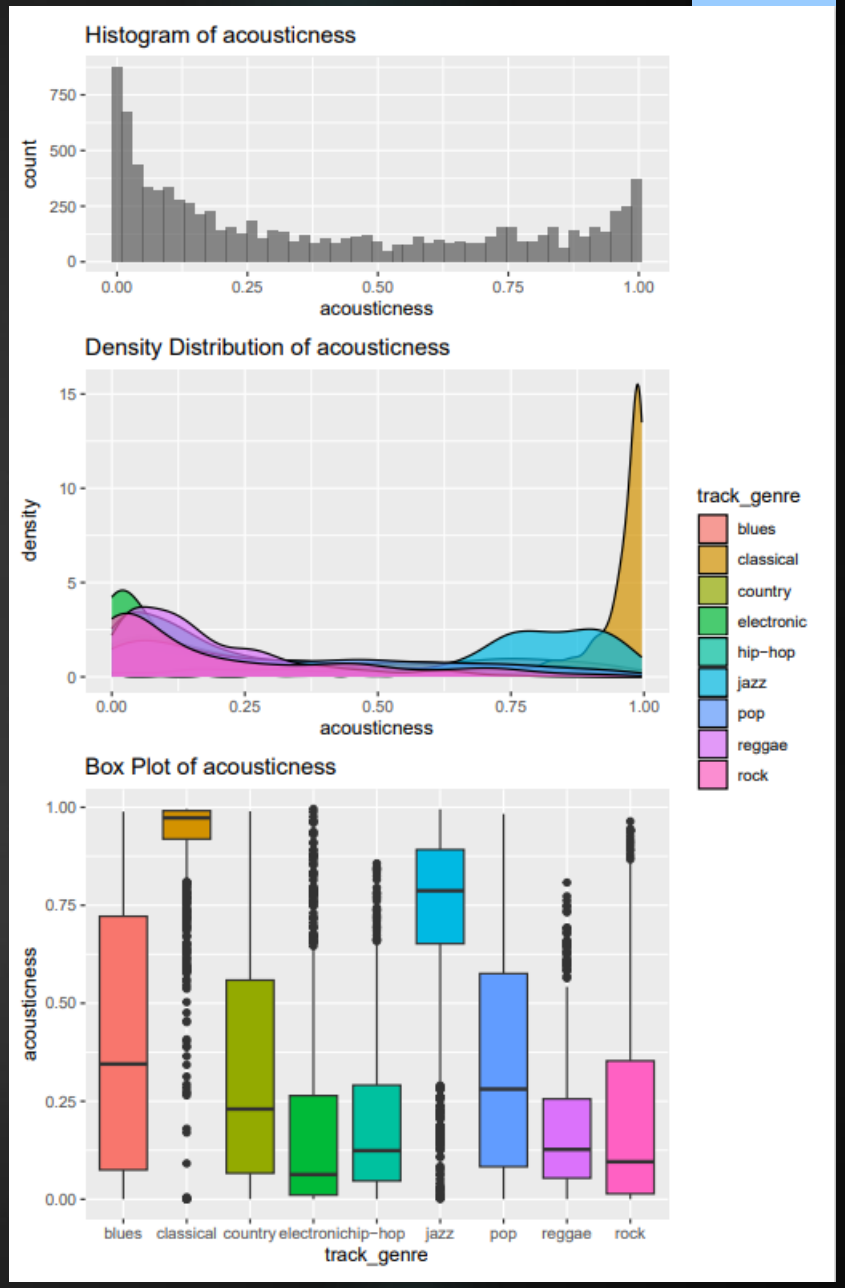
Speechiness

- All the genres have a majority of their tracks concentrated at a very low score
- country in has the lowest variance
- Hip-hop has the highest median value and variance.



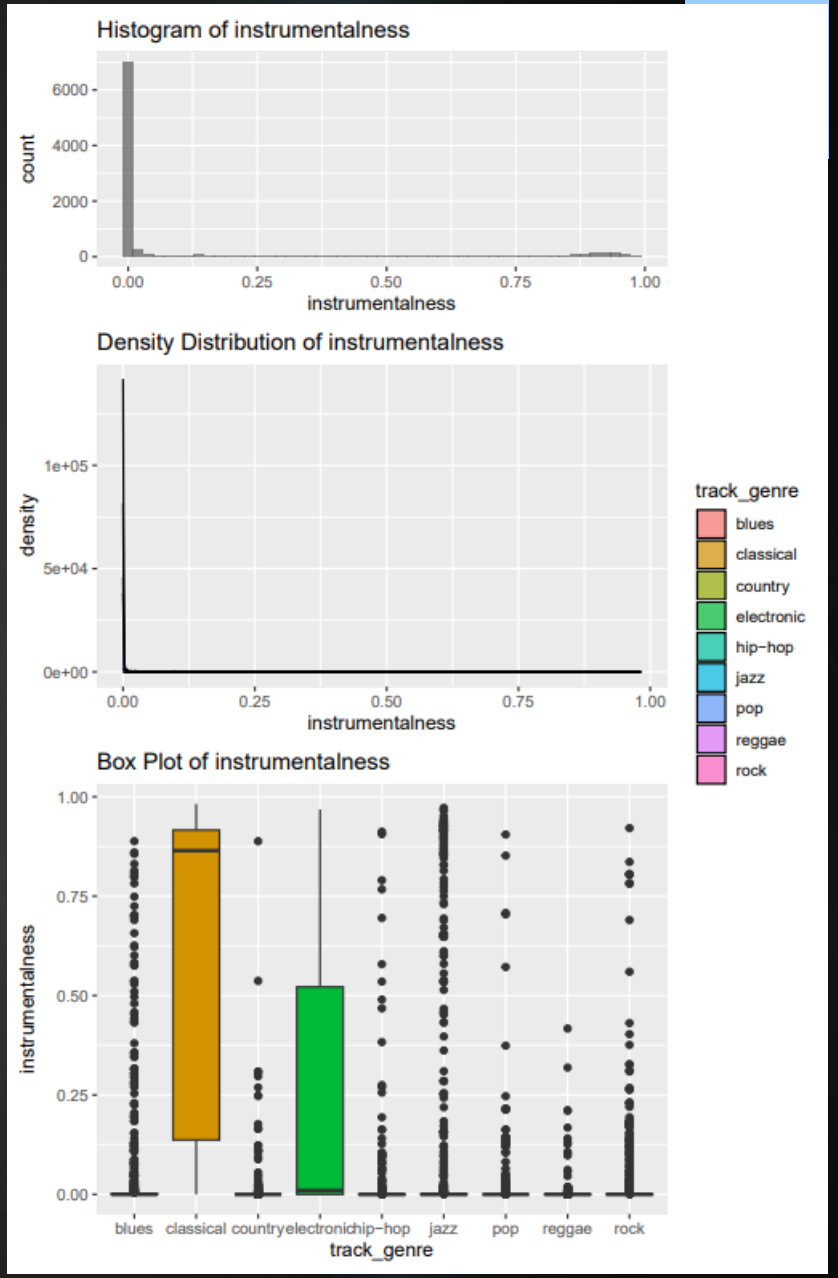
Acousticness

- This feature has a high variance for all the genres with the exception of the classical genre which is highly concentrated close to the maximum value of 1.
- Highest values: Classical,Jazz
- Lowest Values: Electronic, Hip-Hop, Regae



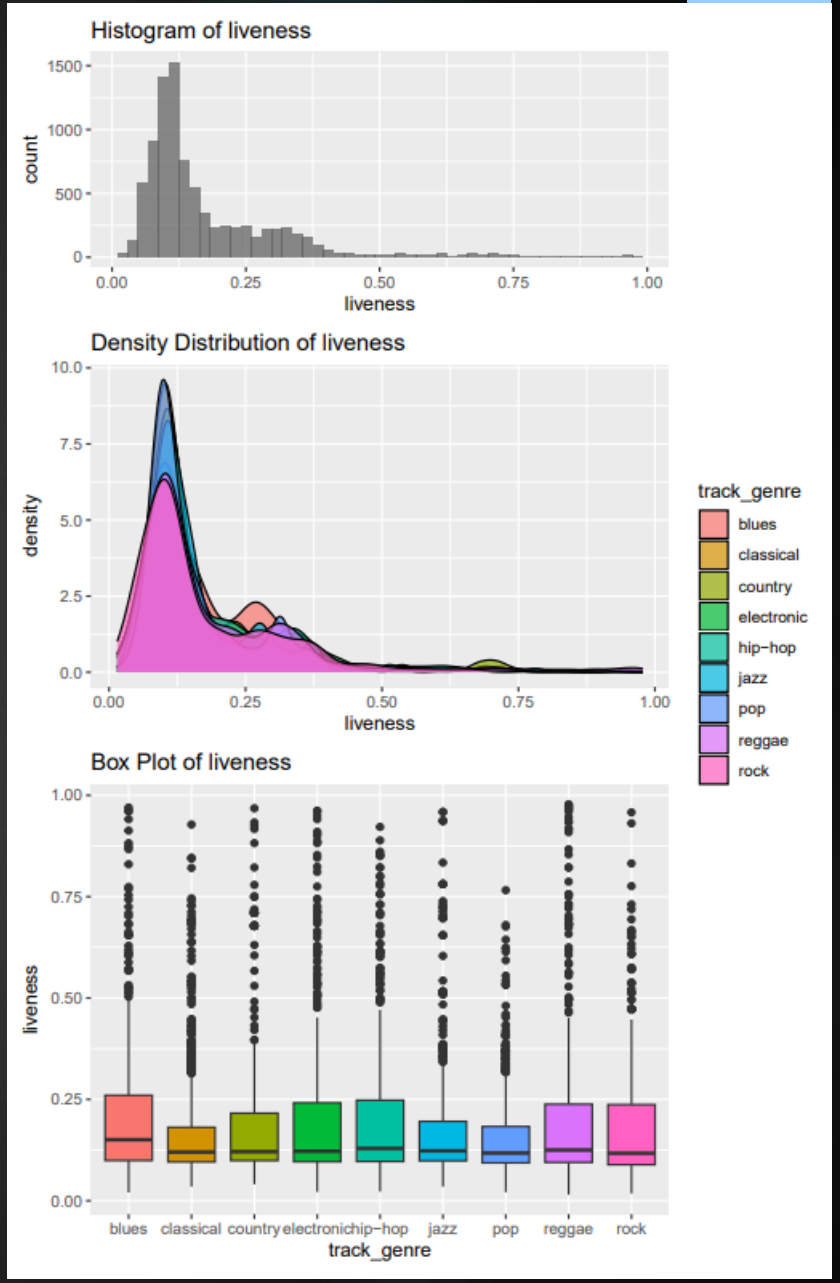
Instrumentalness

- Aside from classical and electronic, all the genres have an instrumentalness score close to 0.
- Classical has a heavily left skewed distribution with a high variance , and is the only genre with median significantly different from 0.
- Electronic has a very right skewed distribution with its median value close to 0 but a very high variance



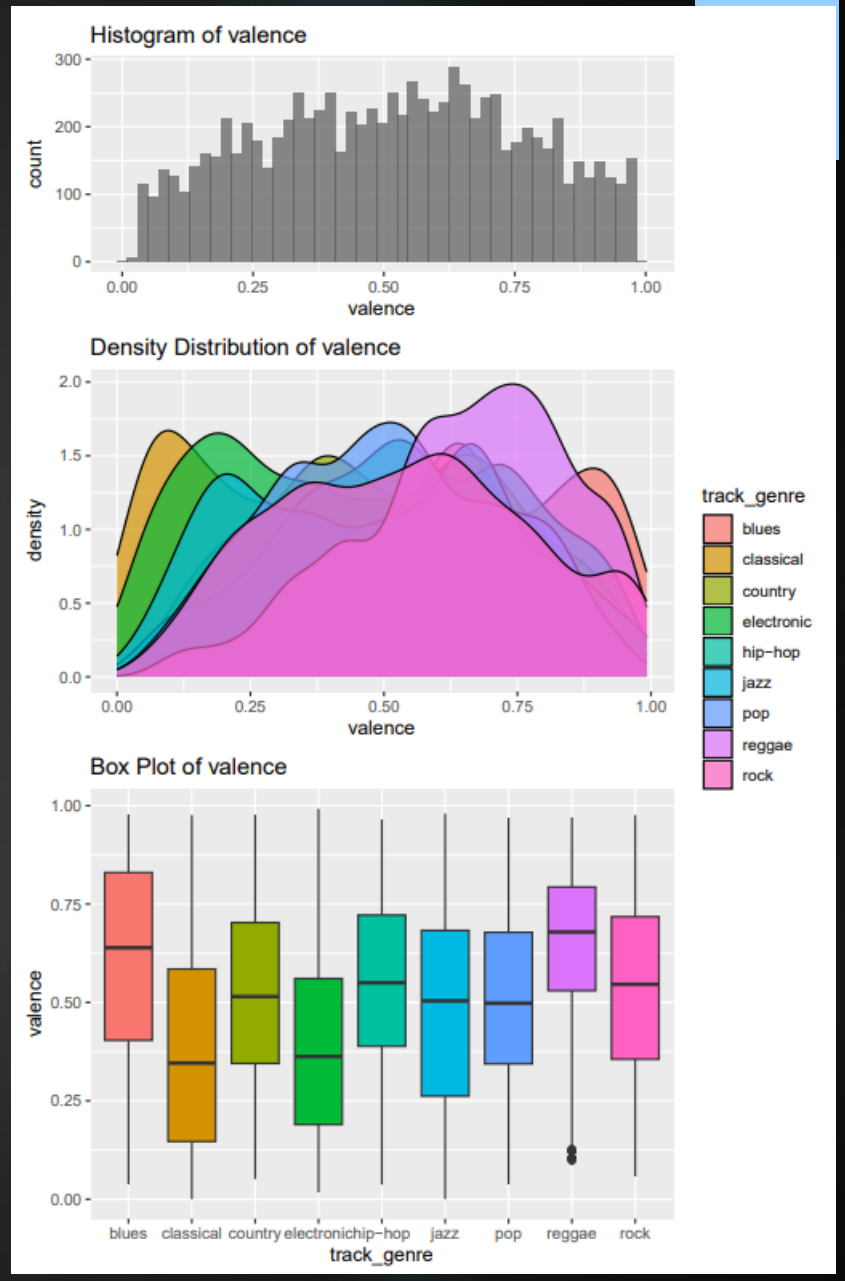
Liveness

- Visually there doesn't seem to be much difference in the liveness distribution of all the genres.
- They all have a right skewed distribution with median around 0.12 with the exception of the **blues** genres which has a slightly **higher median value.**
- There a lot of outliers for each genre.



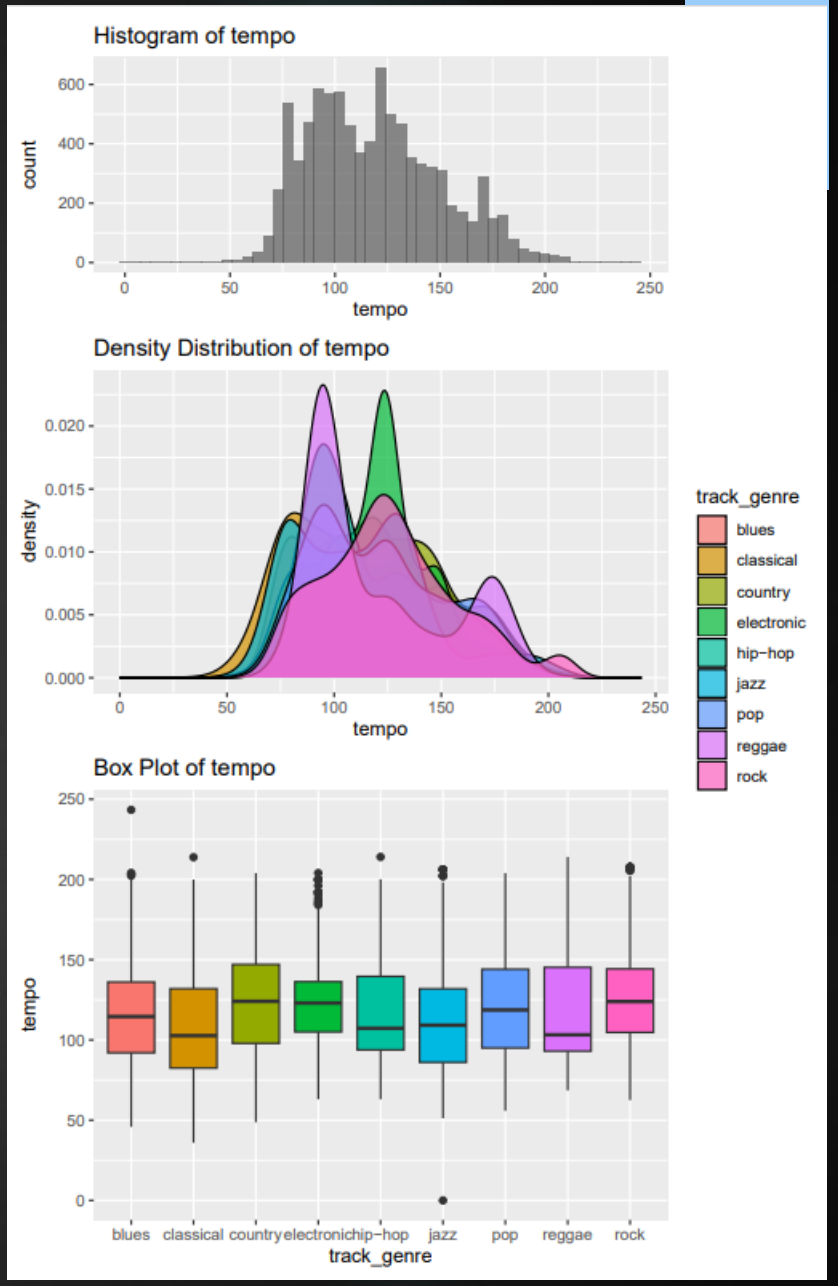
Valence

- Highest Valence: Reggae, Blues
- Lowest Valence: Classical, Electronic



Tempo

- The genre distribution are multimodal and have peaks around distinct values.
- Reggae and electronic have the most identifiable distributions.



MODELS

Multinomial Logistic Regression

- **Base Model**
  - Accuracy =49.55% (on test set)
  - Lowest Class precision: Pop 34.4%, Highest Class precision: Classical 86.7%
- **Subset Selection**
  - Foward/Backward selection by AIC
    - Only ‘key’ feature was removed from full set of variables
    - Accuracy=48.88%
    - ‘Blues’ precision most negatively affected (38% -> 35%)
  - Backward selection by BIC
    - Same as the above
  - Forward selection by BIC
    - 3 features were removed: key,liveness and mode
    - Accuracy=49.39%

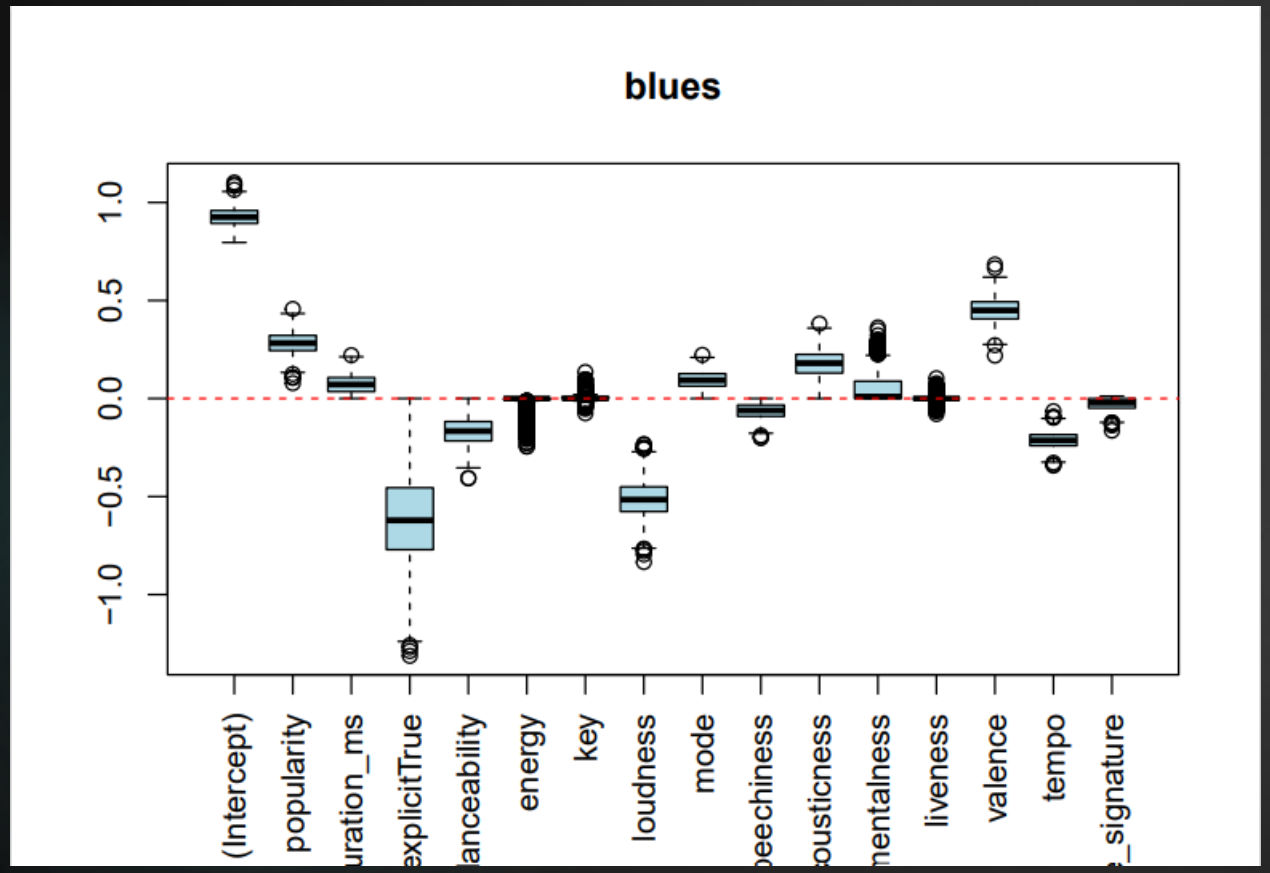
Regularized Multinomial Logistic Regression

- **Ridge Regularization**
  - Minimum error Lambda chosen from ten fold cross-validation
  - Accuracy=49.44%
  - ‘Blues’ precision falls from 37% to 26%
  - Increase in pop precision from 34% to 38% (previous lowest)
- **Lasso Regularisation**
  - Lambda chosen with 1 standard error rule from ten fold cross-validation
  - Accuracy=49.38%
- **Elastic Net**
  - Lambda and alpha from chosen from ten fold cross-validation
  - Accuracy=49.56%

Lasso Coefficients

- Applied Bootstrap approach to assess the stability and variability of Lasso coefficients across different samples.
- Aim to identify important predictors that consistently contribute to the classification of each class while accounting for the variability introduced by the bootstrap resampling process.

BLUES



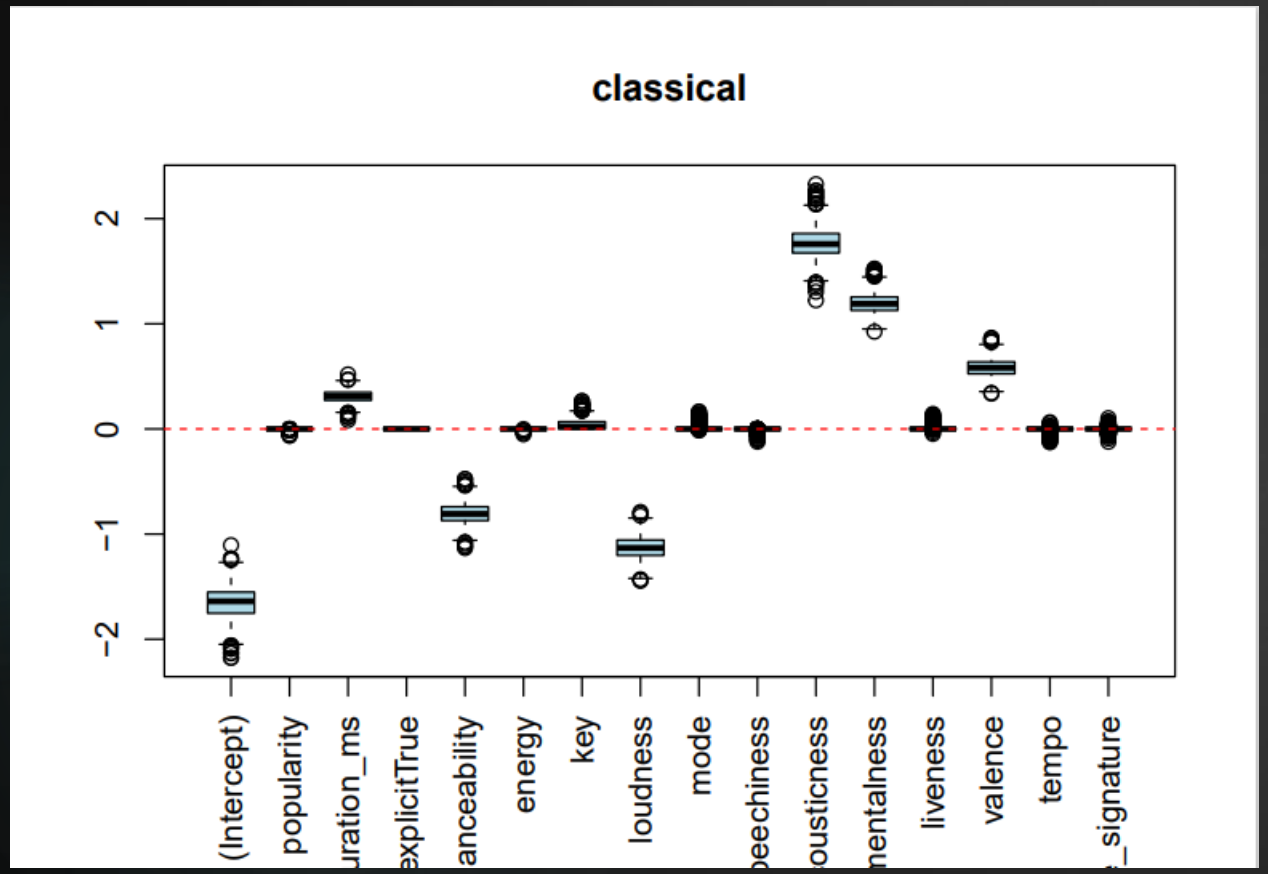
Zero coefficients:

- Energy
- Key
- Liveness

Important Coefficients:

- Explicit (-)
- Valence (+)
- Loudness (-)

Classical



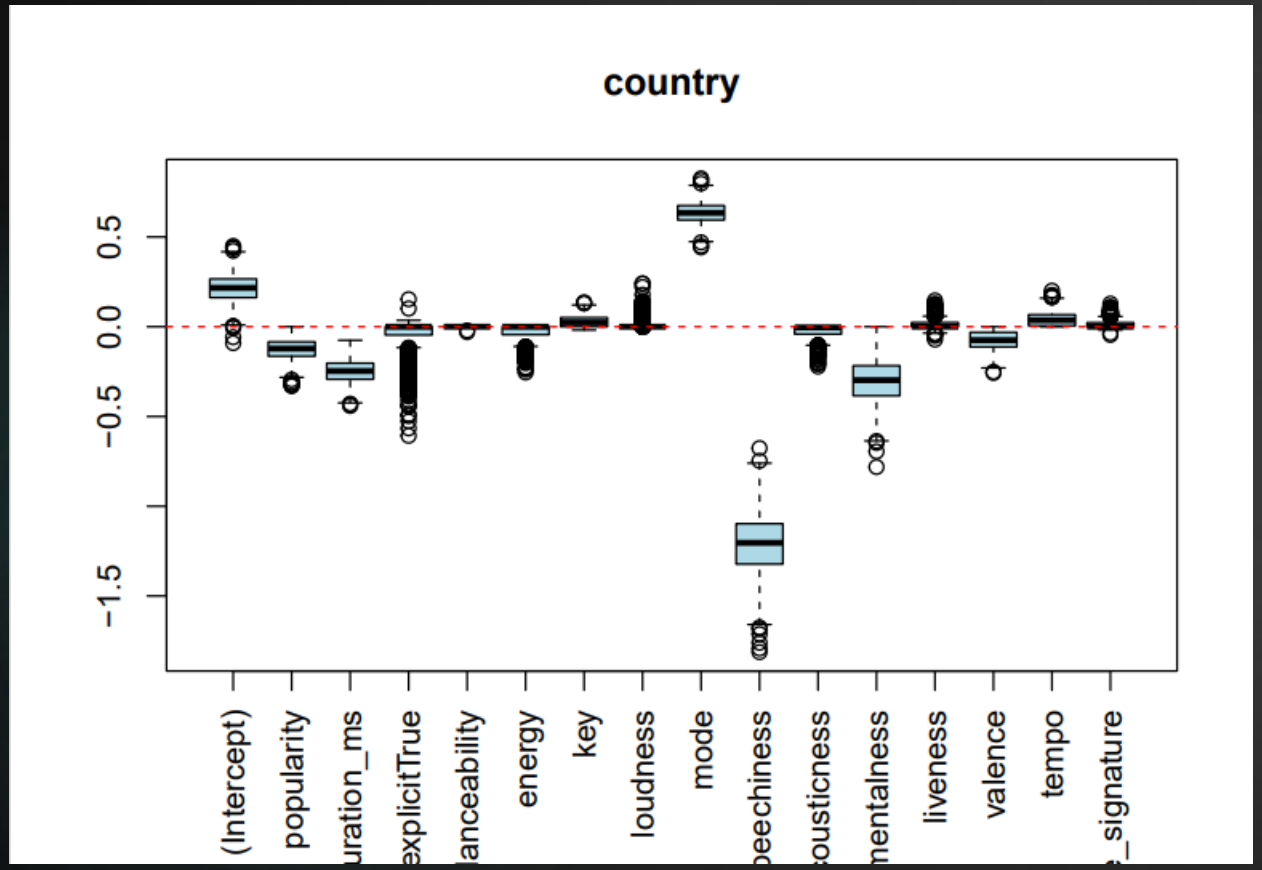
Zero coefficients:

- Popularity
- Energy
- Key
- Mode
- Speechiness
- Liveness
- Tempo
- Time signiture

Important Coefficients:

- Acousticness (+)
- Instrumentalness (+)
- Loudness (-)

Country



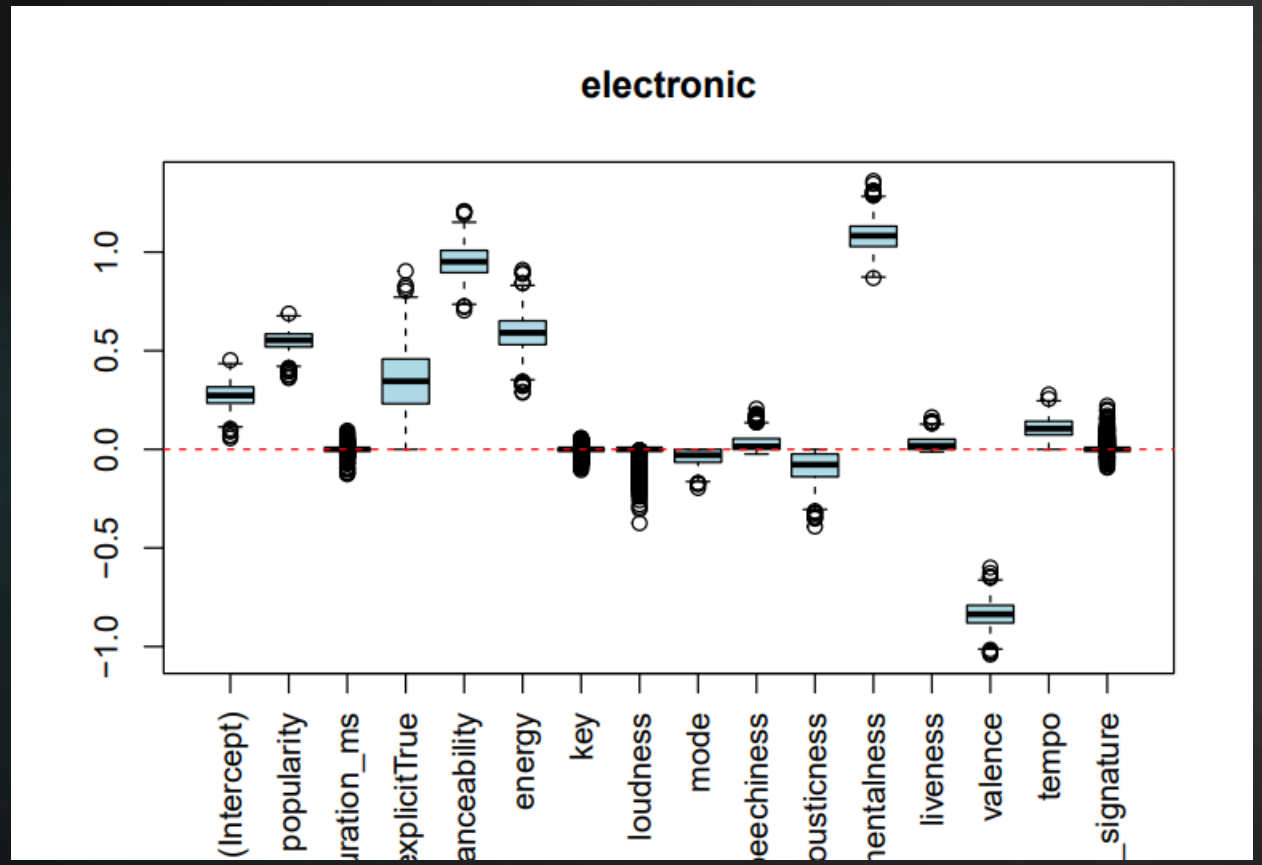
Zero coefficients:

- Energy
- Danceability
- Loudness
- Liveness
- Time signiture

Important Coefficients:

- Speechiness(-)
- Mode(+)

Electronic



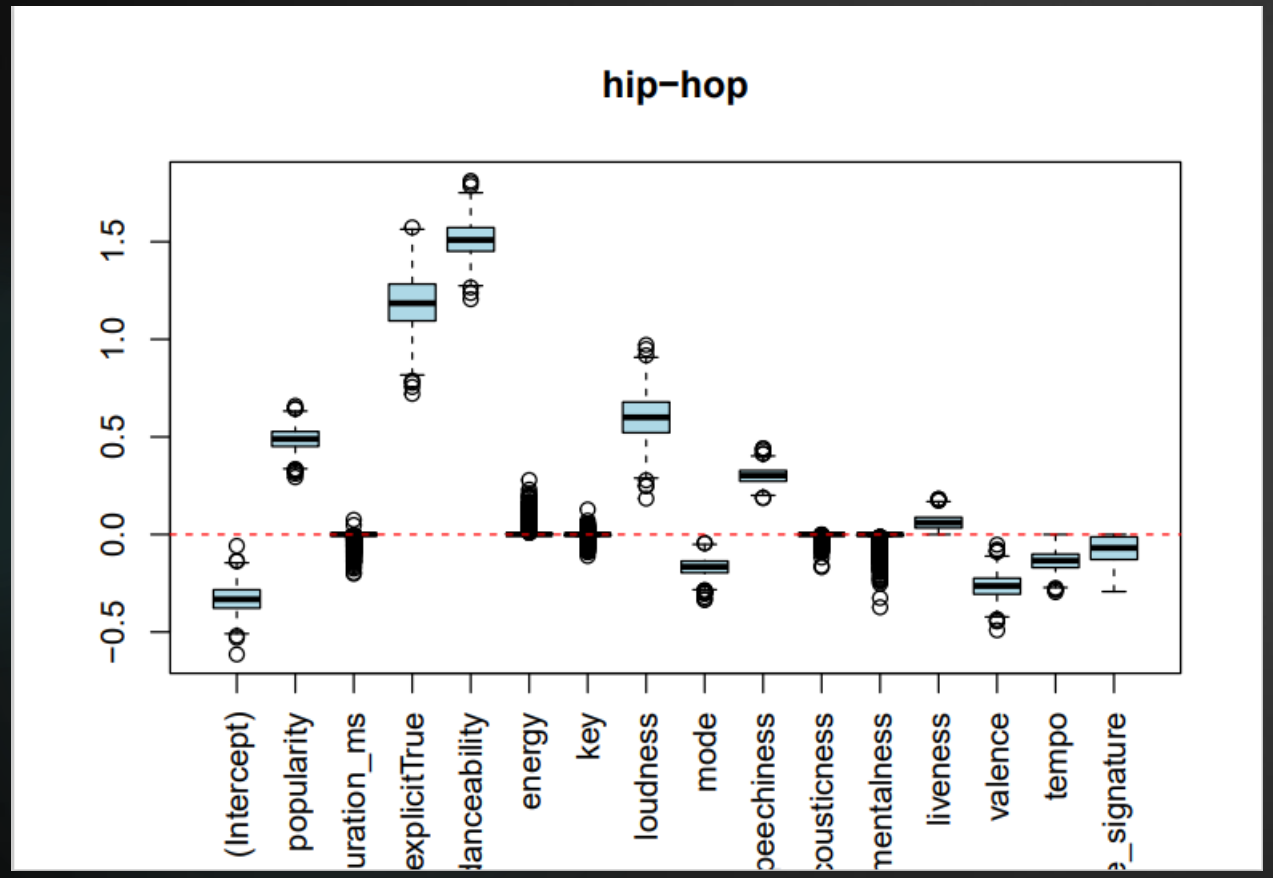
Zero coefficients:

- Duration
- Time signiture
- Loudness
- Key

Important Coefficients:

- Instrumentalness(+)
- Danceability(+)

Hip-Hop



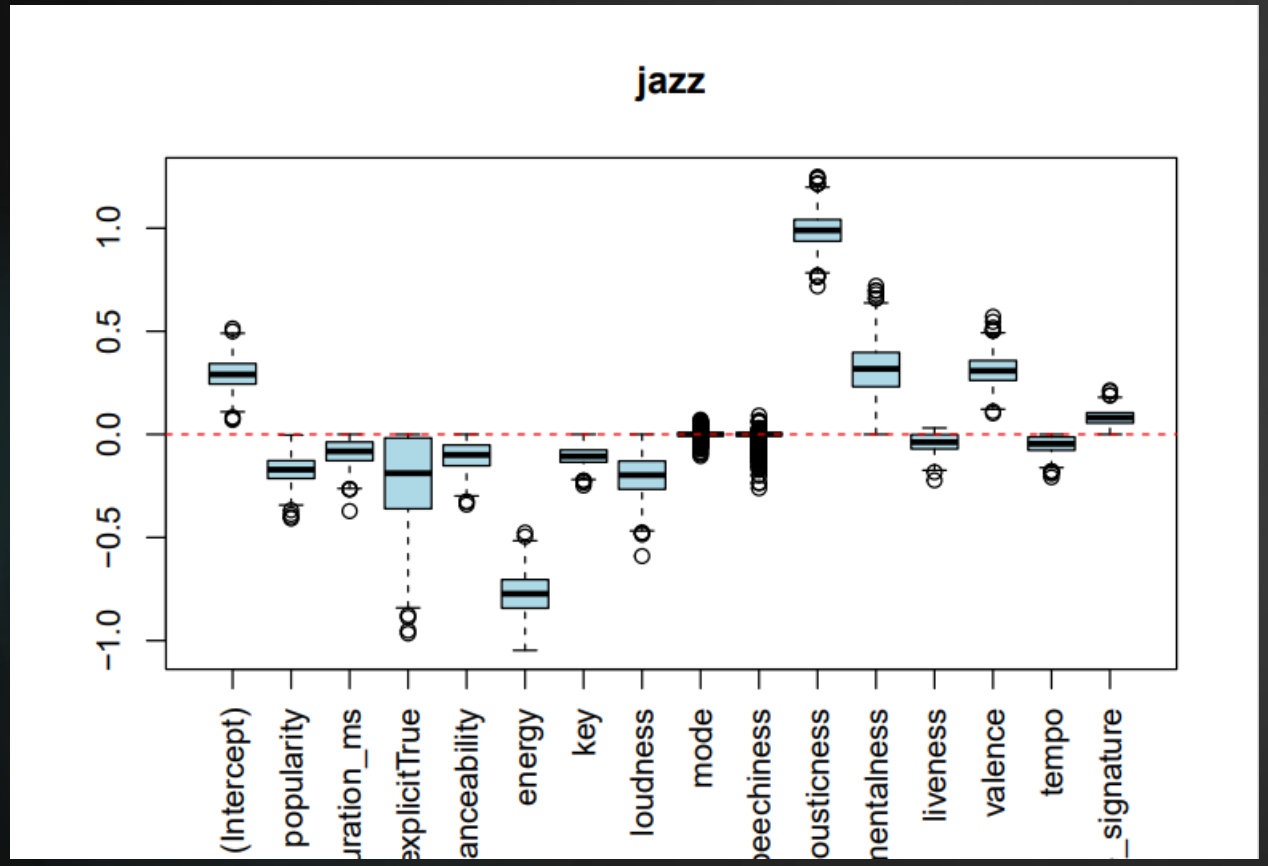
Zero coefficients:

- Duration
- Energy
- Key
- Acousticness
- Instrumentalness

Important Coefficients:

- Explicit(+)
- Danceability(+)
- Loudness(+)

Jazz



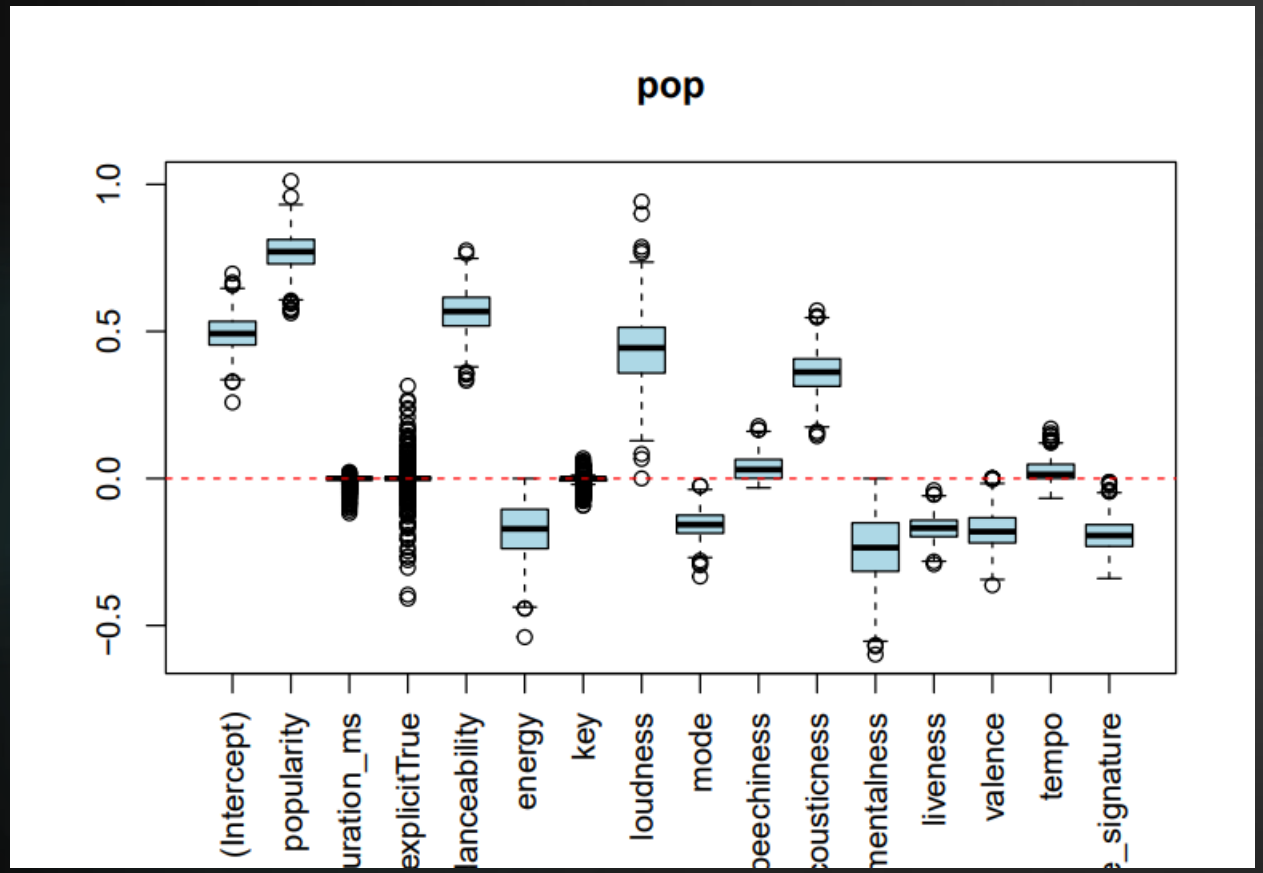
Zero coefficients:

- Mode
- Speechiness

Important Coefficients:

- Energy(-)
- Acousticness(+)

Pop



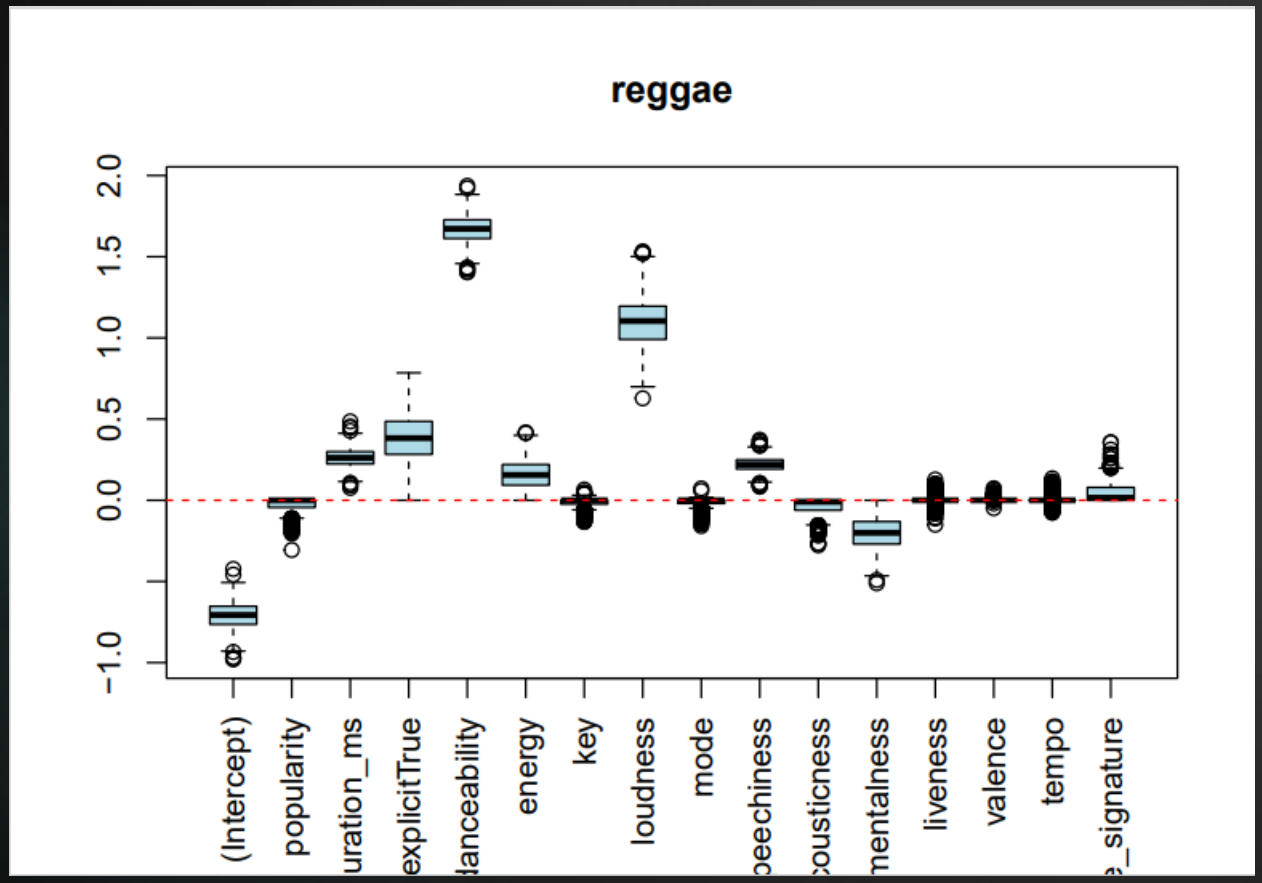
Zero coefficients:

- Duration
- Explicit
- Key

Important Coefficients:

- Popularity (+)
- Danceability (+)
- Loudness (+)

Reggae



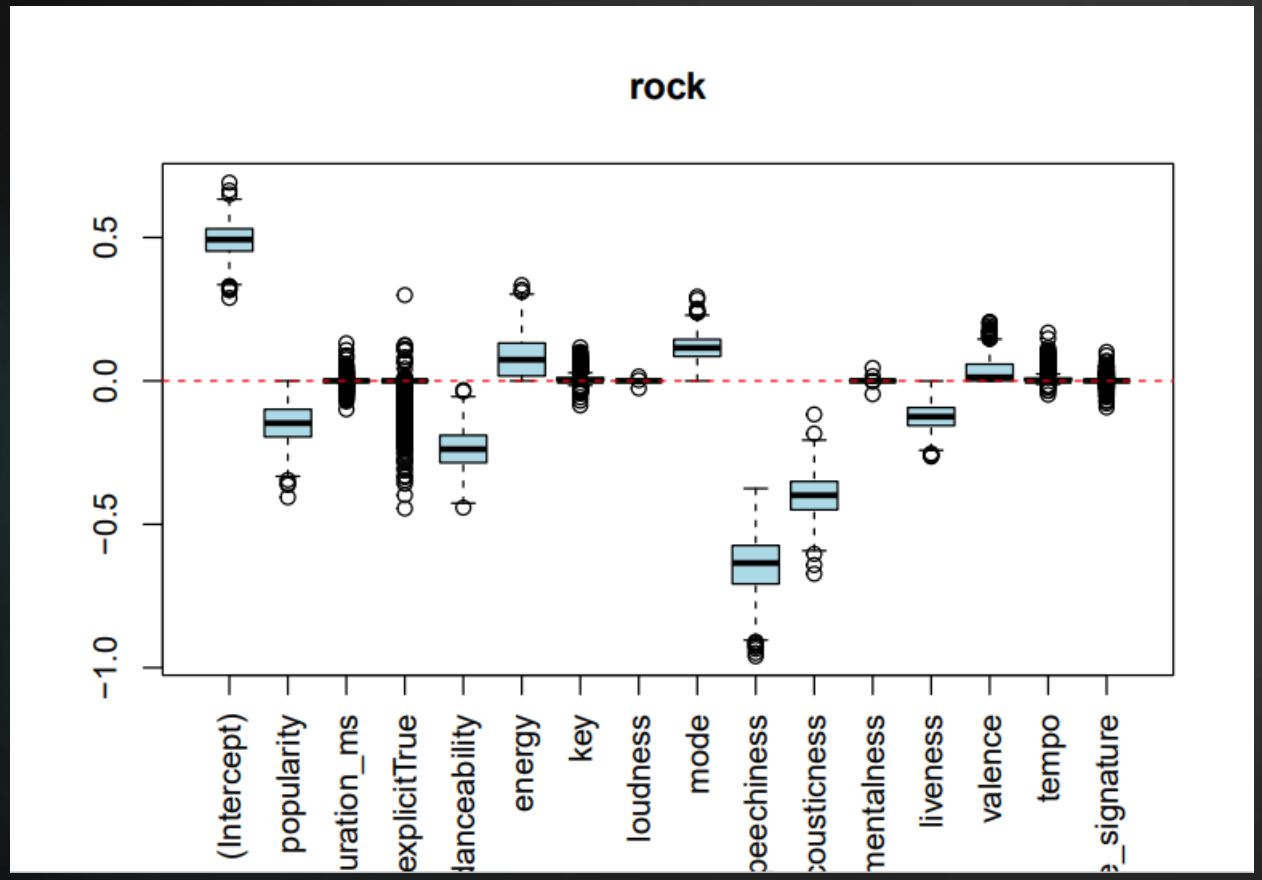
Zero coefficients:

- Key
- Mode
- Liveness
- Valence
- Tempo

Important Coefficients:

- Danceability (+)
- Loudness (+)

Rock



Zero coefficients:

- Duration
- Key
- Loudness
- Instrumentalness
- Tempo
- Time signiture

Important Coefficients:

- Speechiness (-)
- Acousticness (-)

K-Nearest Neighbours

- The optimal K found from cross-validation was K=1
- Accuracy= 54.39%
- Low precision of around 40% for Blues, Pop and Hip-Hop
- Very High precision of 87% for classical

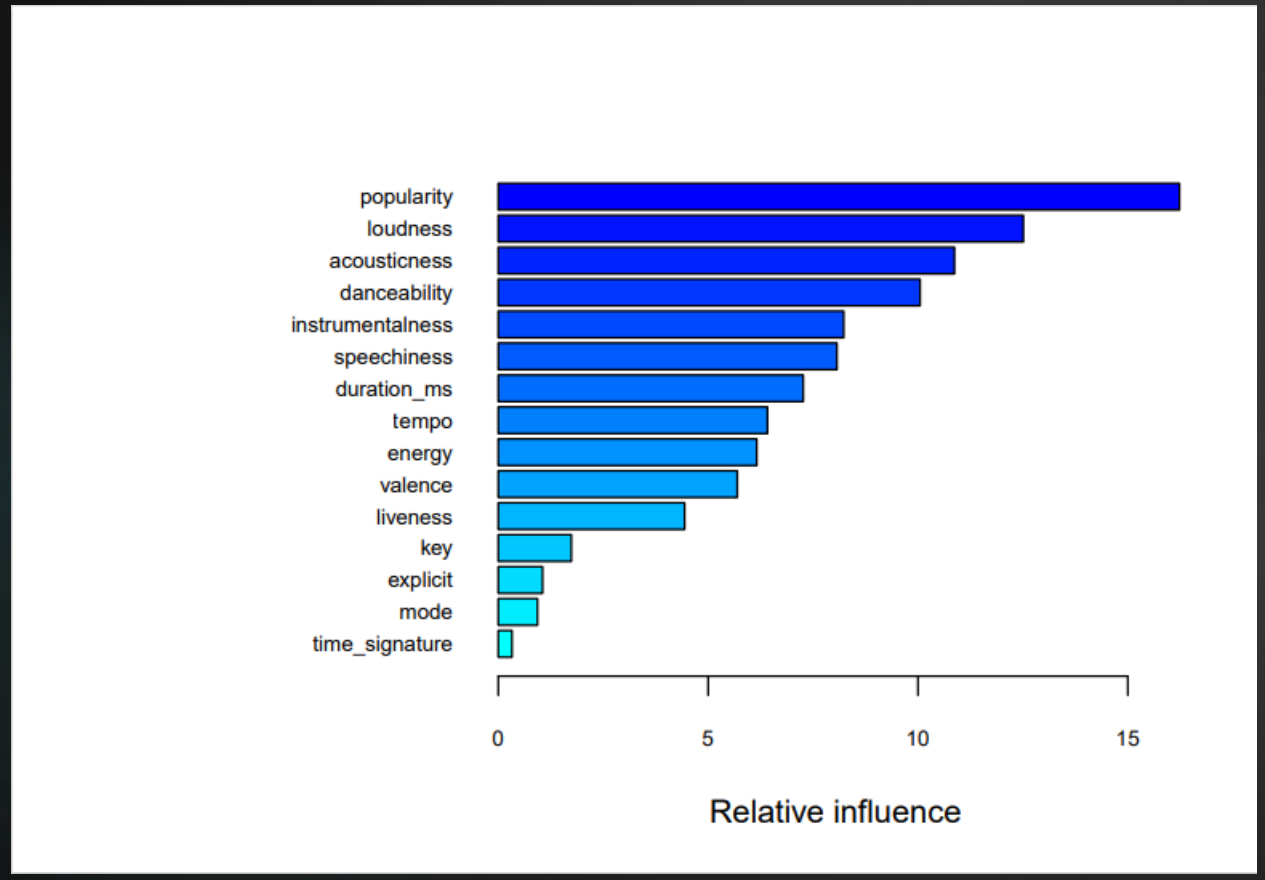
Random Forest

- Applied cross-validation to find the optimal number of variables randomly sampled as candidates at each split when building each tree in the forest: mtry=8
- Accuracy=67.78%
- Min class precision of 50%: Blues increased by around 10% compared to KNN
- Max class precision (classical) is basically the same: 87.67%

Gradient Boosting

- Applied cross-validation to select number of trees=2000, interaction depth=10 and shrinkage=0.01
- **Accuracy = 90%**
- Lowest class precision is 84.21% for Reggae
- Blues precision =88.04% (Lowest in most other models)
- Highest precision class Classical = 99.5%

Variable Relative Influence



THANK YOU FOR YOUR ATTENTION