Spotify Audio Feature Analysis

Elizabeth Brooks

Objectives

- Discover relationship between audio features and popularity
- Identify most influential track attributes
- Determine audience preferences for tracks and artists

Workflow

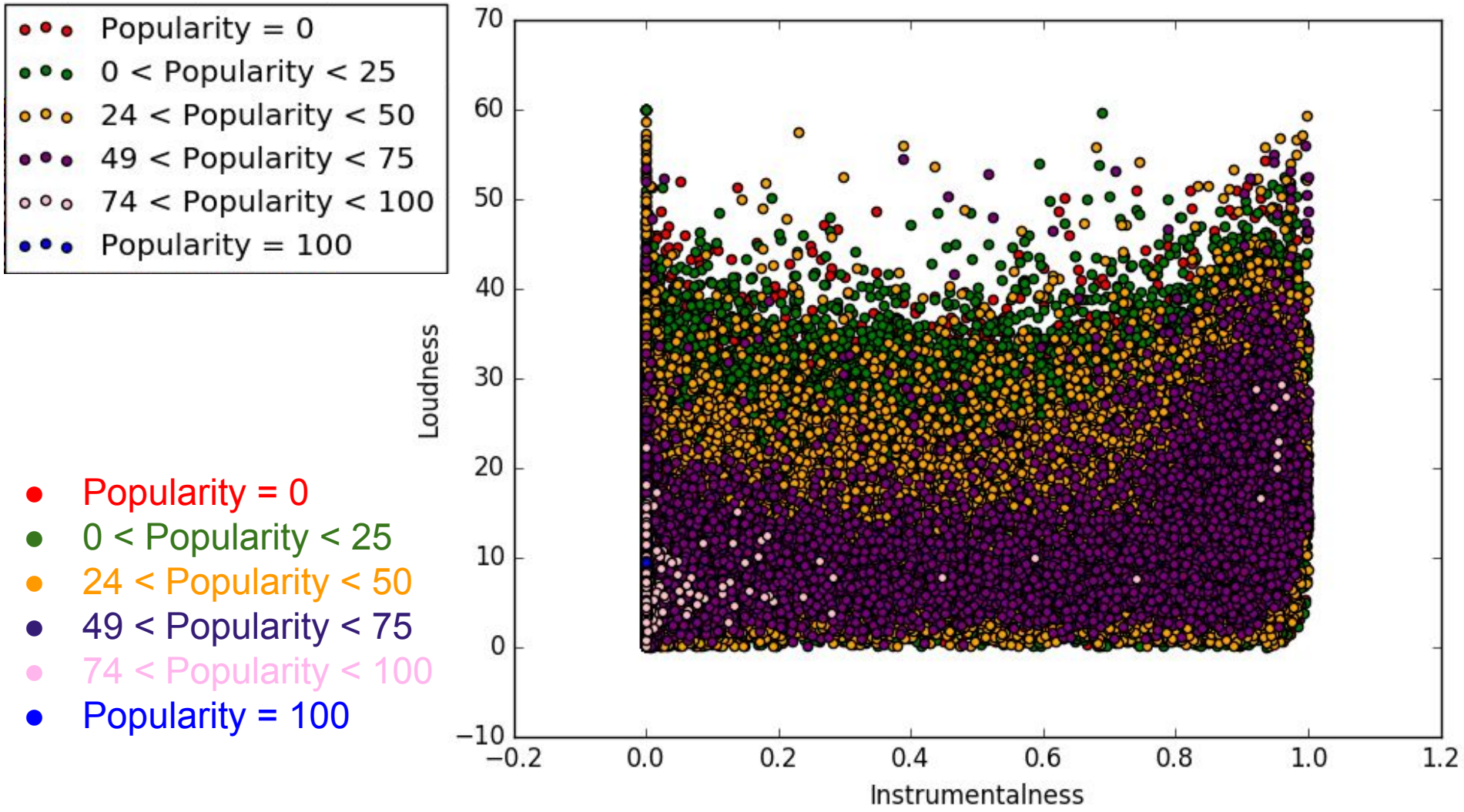
1. Get a List of Browse Categories -> Category IDs
2. Get a Category's playlists -> playlist IDs
3. Get a Playlist's Tracks -> artist IDs
4. Get an Artist's Related Artists -> related artist IDs
5. Get an Artist's Top Tracks -> track IDs
6. Get an Artist's Albums -> track IDs
7. Get Several Tracks -> track popularity
8. Get Audio Features for Several Tracks -> audio features

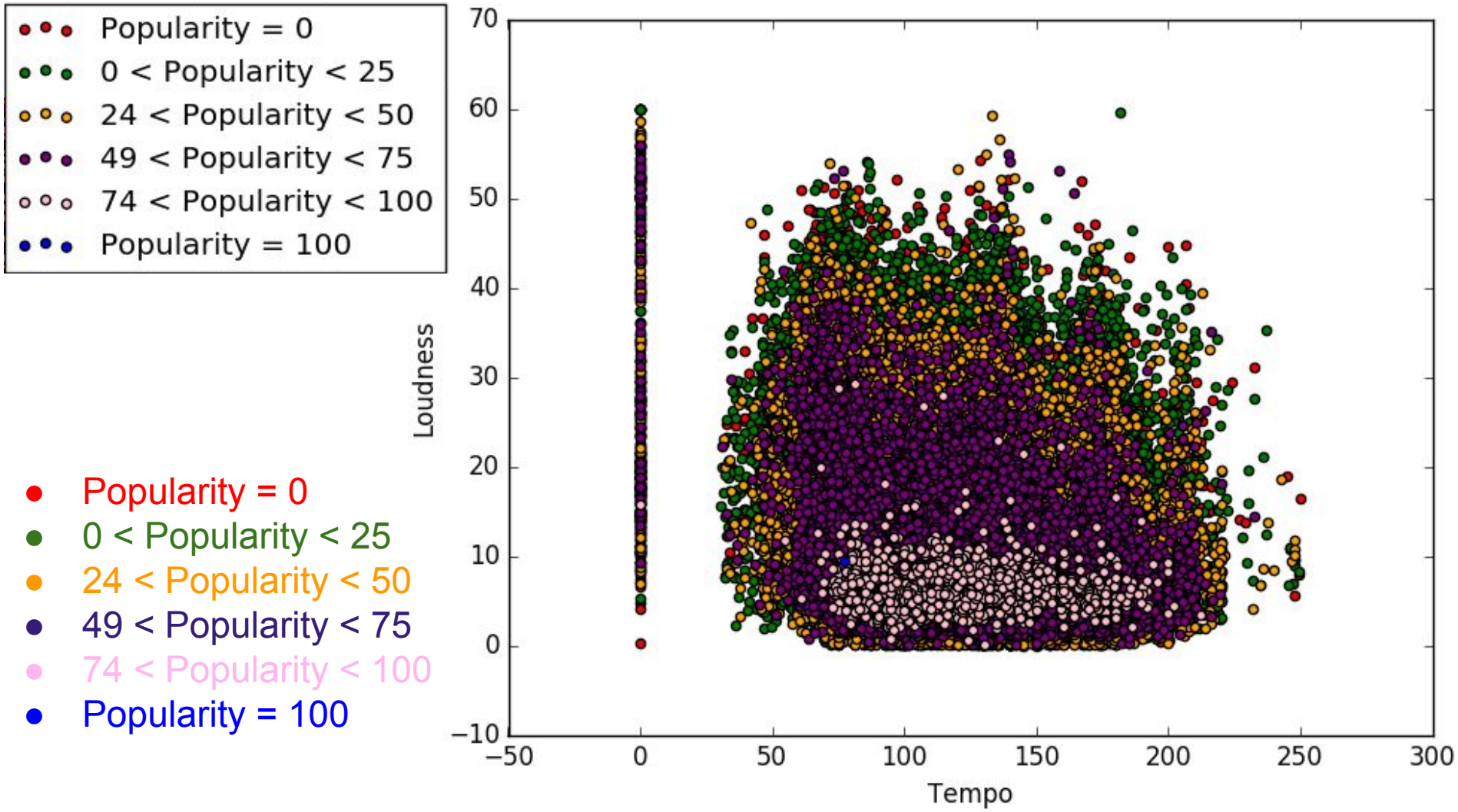
Audio Features

Key	Type	Description
acousticness	float	Measures whether the track is acoustic
danceability	float	Suitability for dancing
duration_ms	int	Track duration
energy	float	Perceptual measure of intensity and activity
instrumentalness	float	Presence of vocals
key	int	Musical key
liveness	float	Presence of an audience
loudness	float	Overall loudness
mode	int	Indicates modality
speechiness	float	Presence of spoken word
tempo	float	Overall estimated tempo
time_signature	int	Overall time signature
valence	float	Musical positiveness conveyed

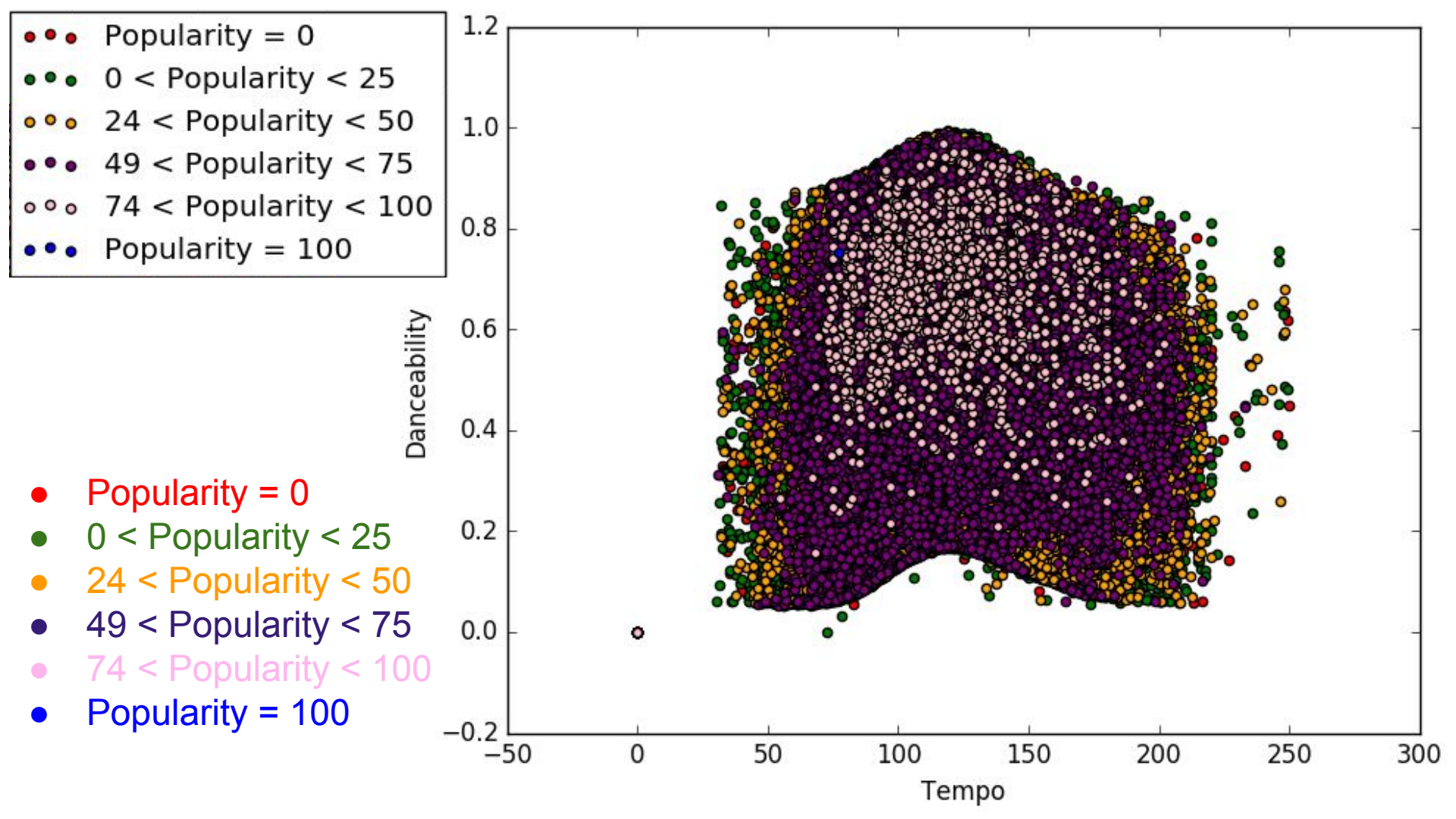
Track Feature Analysis

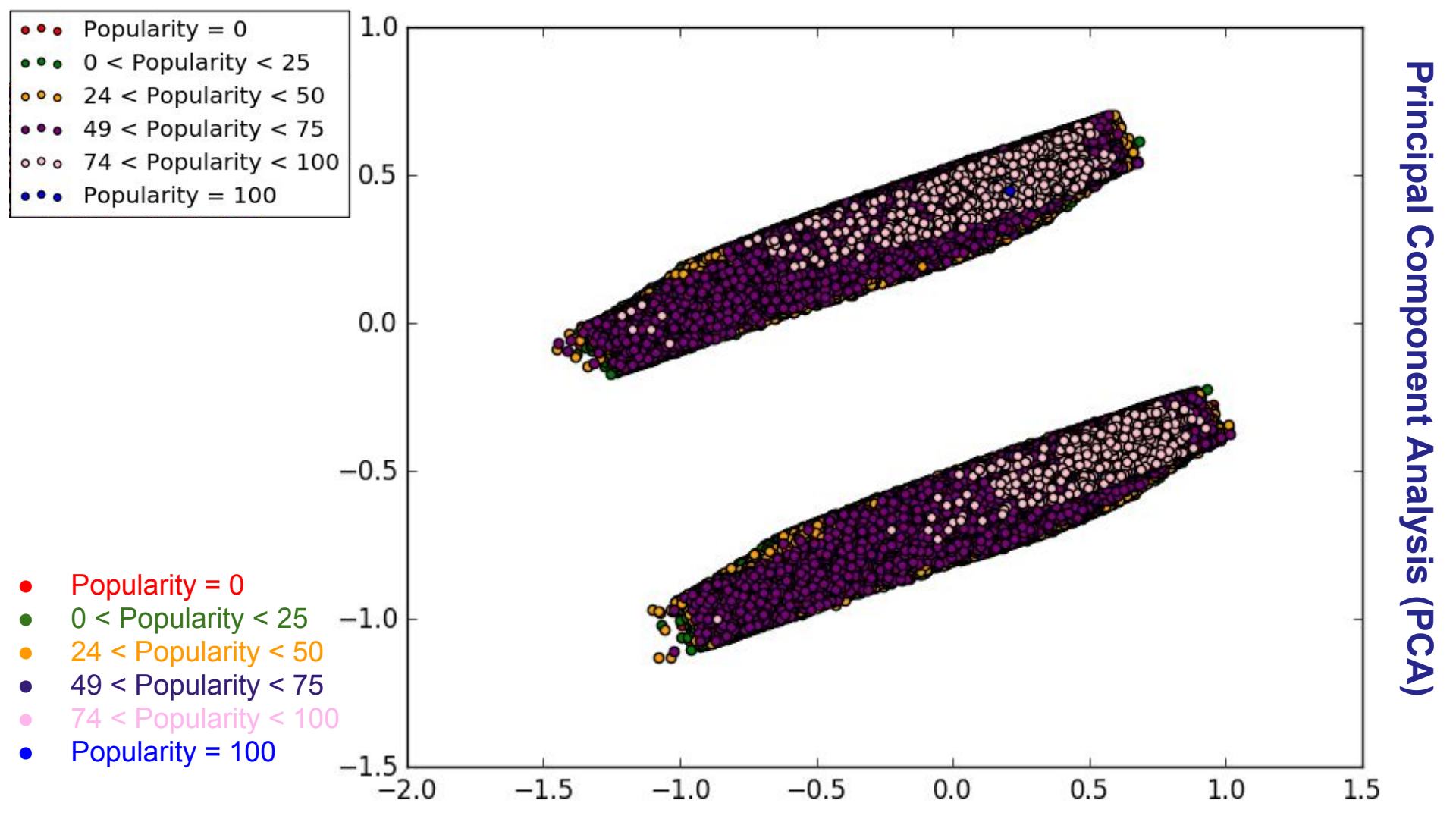
- Visualize multidimensional data -> Python
  - Two-dimensional slices
  - Binned popularity ranges
- Feature scaling -> Python
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
- Decision trees
  - Classification -> Python and R
  - Regression -> R

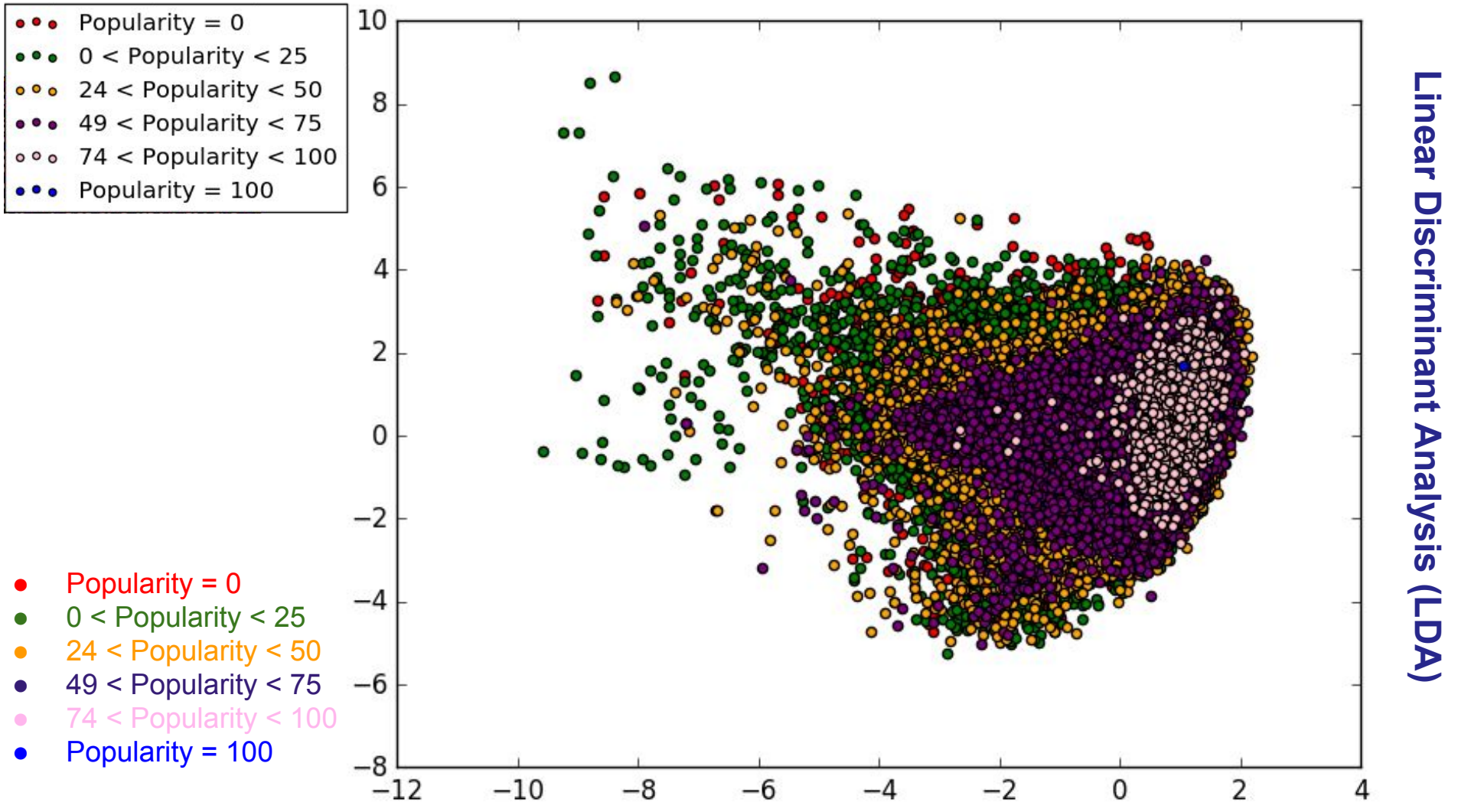

Popularity = 0
0 < Popularity < 25
24 < Popularity < 50
49 < Popularity < 75
74 < Popularity < 100
Popularity = 100
Loudness
Instrumentalness
70
60
50
40
30
20
10
0
−10
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2

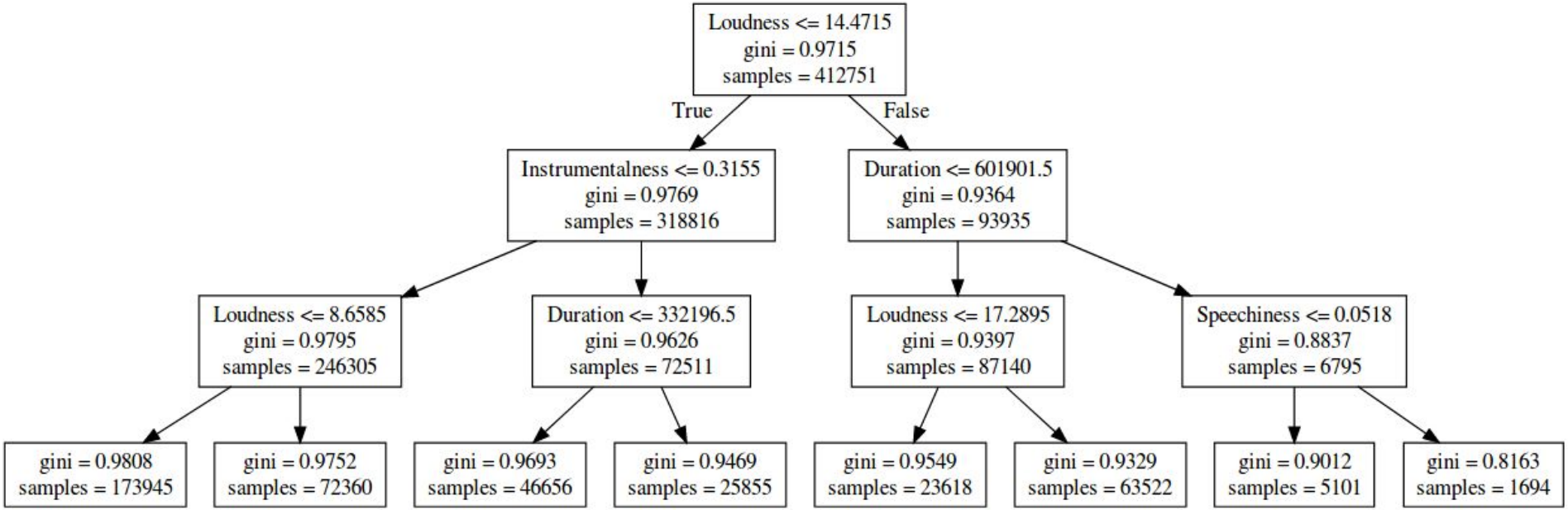


- Popularity = 0
- 0 < Popularity < 25
- 24 < Popularity < 50
- 49 < Popularity < 75
- 74 < Popularity < 100
- Popularity = 100

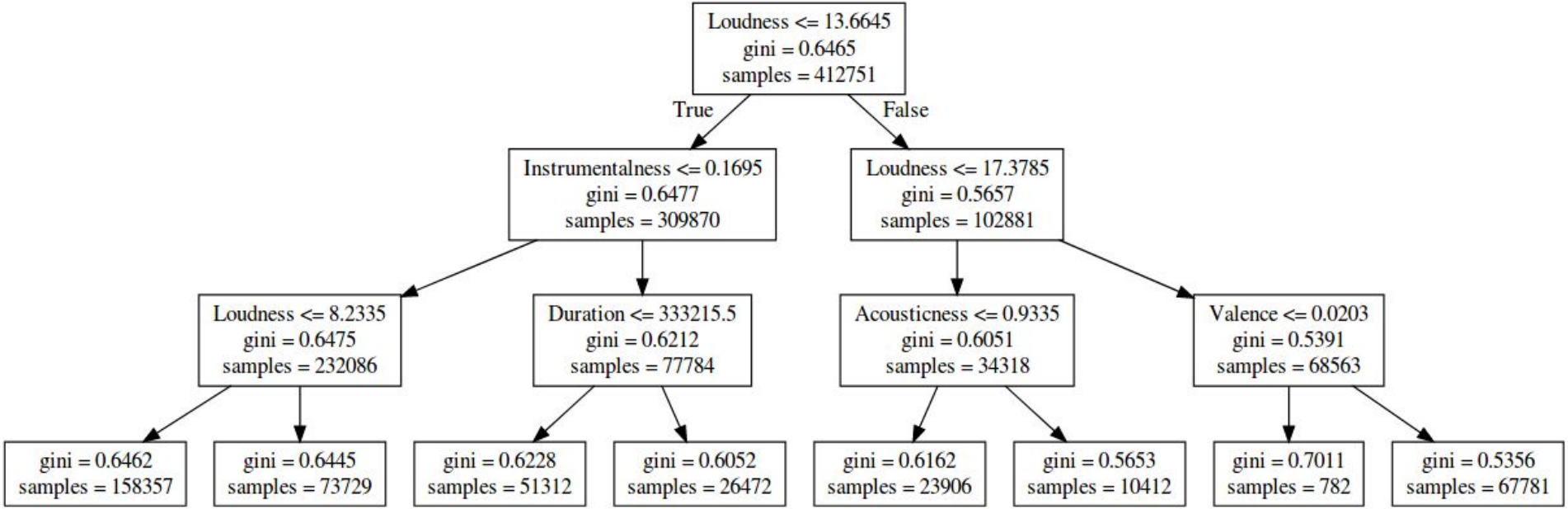

Popularity = 0
0 < Popularity < 25
24 < Popularity < 50
49 < Popularity < 75
74 < Popularity < 100
Popularity = 100
Danceability
Tempo
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−50
0
50
100
150
200
250
300


Popularity = 0
0 < Popularity < 25
24 < Popularity < 50
49 < Popularity < 75
74 < Popularity < 100
Popularity = 100
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Principal Component Analysis (PCA)


Popularity = 0
0 < Popularity < 25
24 < Popularity < 50
49 < Popularity < 75
74 < Popularity < 100
Popularity = 100
10
8
6
4
2
0
−2
−4
−6
−8
−12
−10
−8
−6
−4
−2
0
2
4
Linear Discriminant Analysis (LDA)

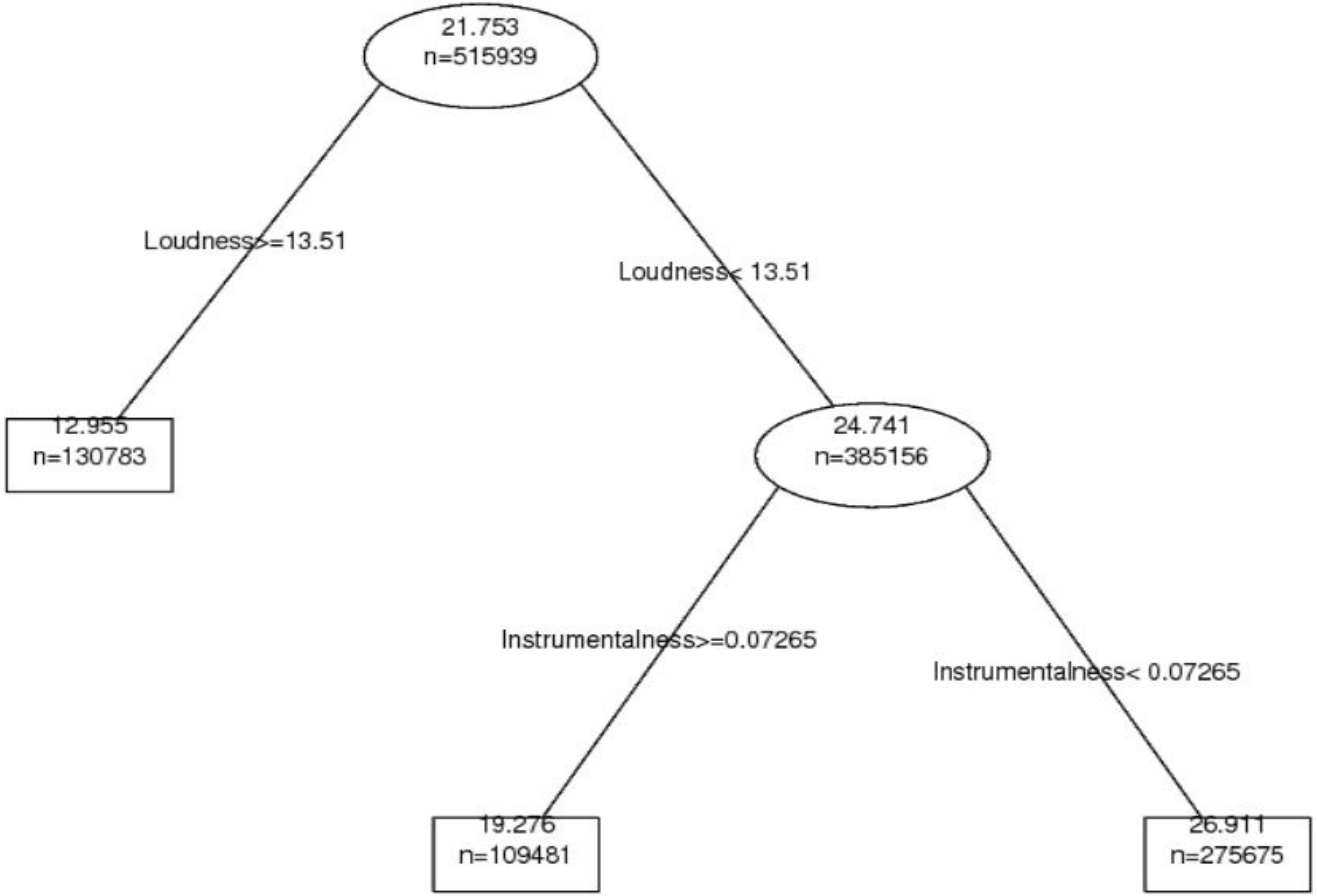


Regression Tree - Gini



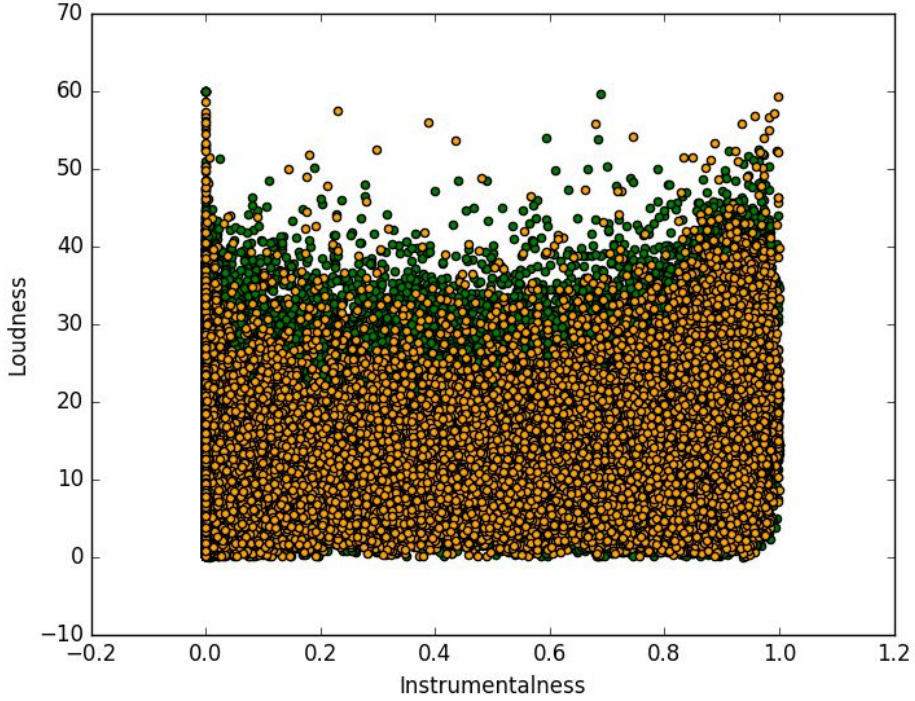
Classification Tree - Gini

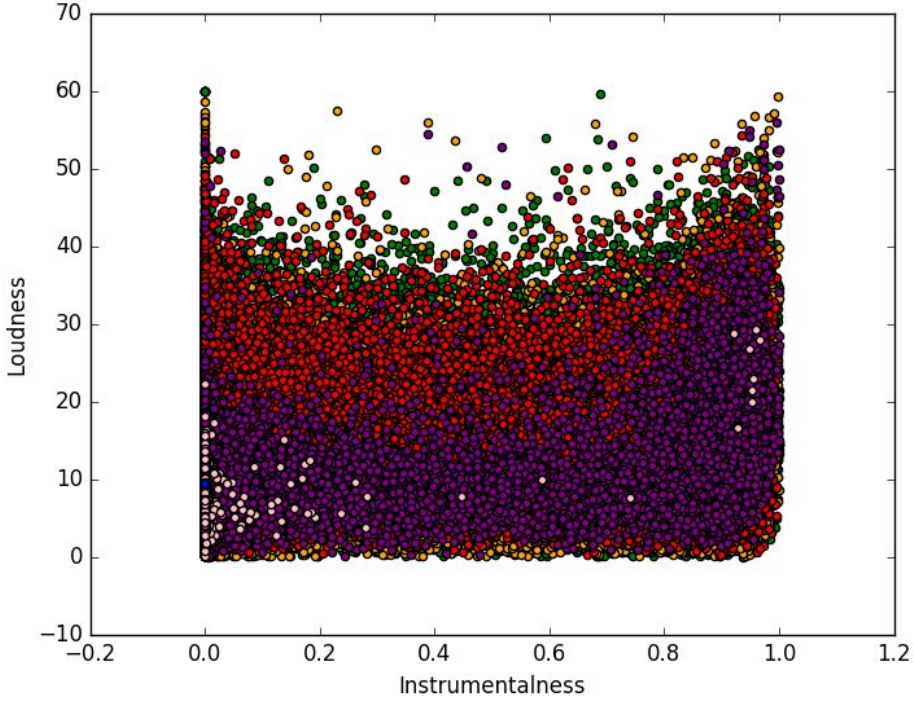
Regression Tree for Track Popularity



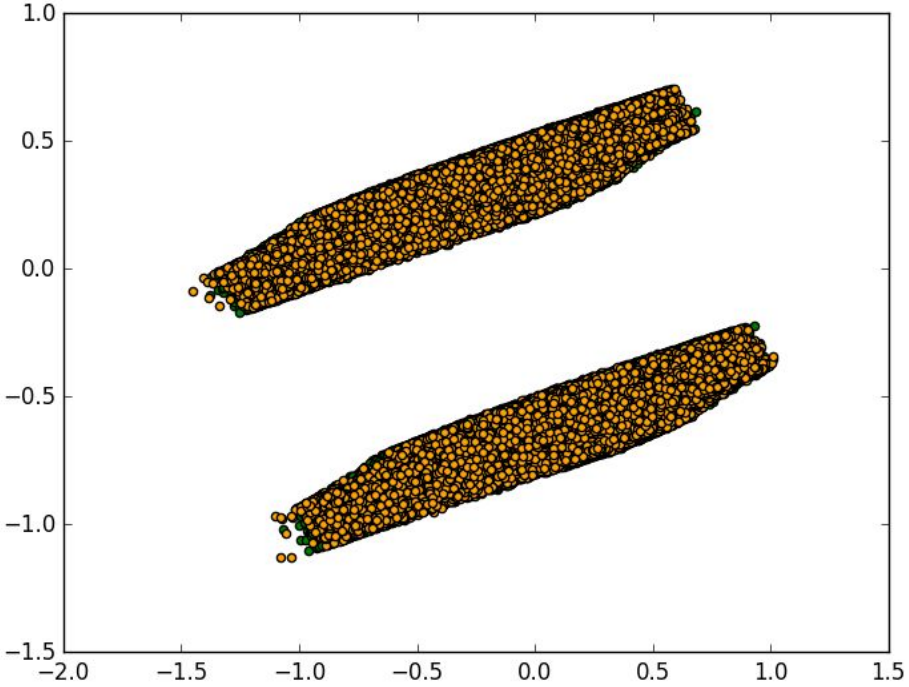
Regression Tree Analysis

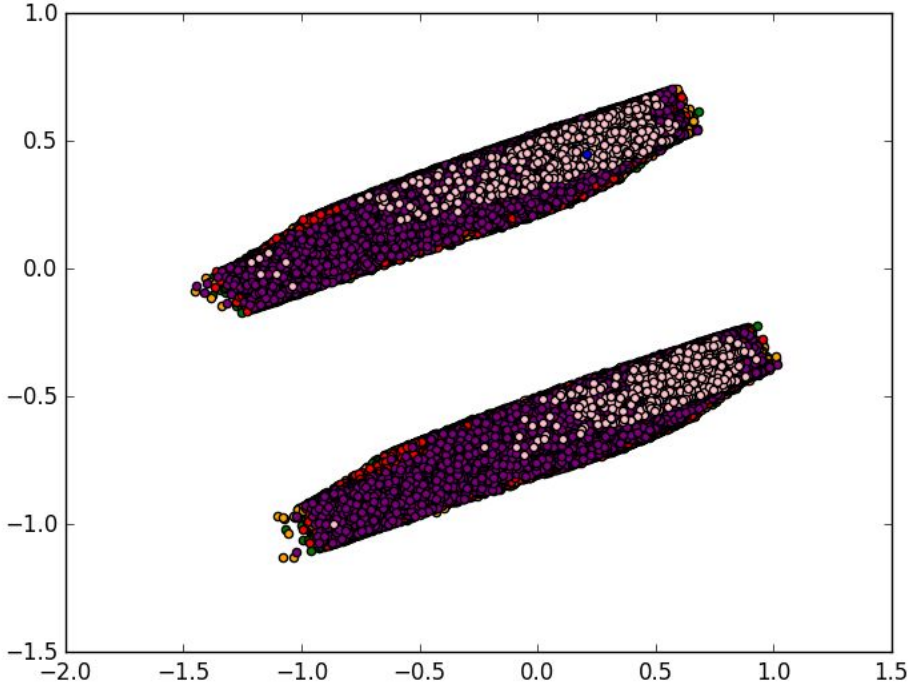
```
Variable importance
        Loudness           Energy     Acousticness Instrumentalness
              31               21               17               15
    Danceability          Valence         Duration
               8                7                2

Node number 1: 515939 observations,    complexity param=0.08452563
  mean=21.75331, MSE=310.9656
  left son=2 (130783 obs) right son=3 (385156 obs)
  Primary splits:
      Loudness         < 13.5145   to the right, improve=0.08452563, (0 missing)
      Acousticness     < 0.8335    to the right, improve=0.07047104, (0 missing)
      Energy           < 0.2975    to the left,  improve=0.06770697, (0 missing)
      Instrumentalness < 0.02945   to the right, improve=0.06201646, (0 missing)
      Danceability     < 0.4195    to the left,  improve=0.04075925, (0 missing)
  Surrogate splits:
      Energy           < 0.2805    to the left,  agree=0.916, adj=0.670, (0 split)
      Acousticness     < 0.8615    to the right, agree=0.885, adj=0.545, (0 split)
      Danceability     < 0.3315    to the left,  agree=0.807, adj=0.238, (0 split)
      Valence          < 0.1075    to the left,  agree=0.797, adj=0.200, (0 split)
      Instrumentalness < 0.8665    to the right, agree=0.779, adj=0.129, (0 split)
```



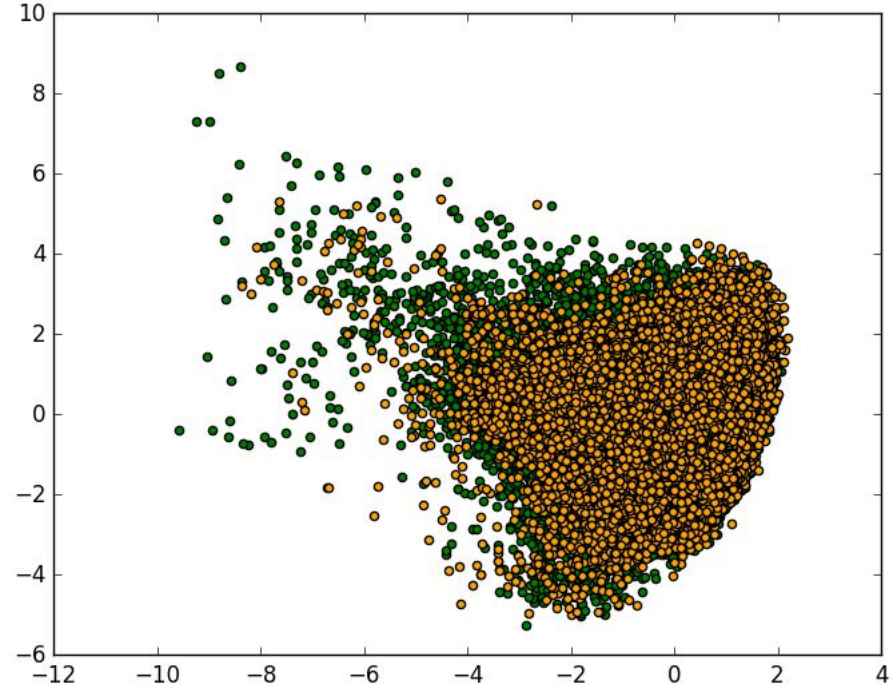


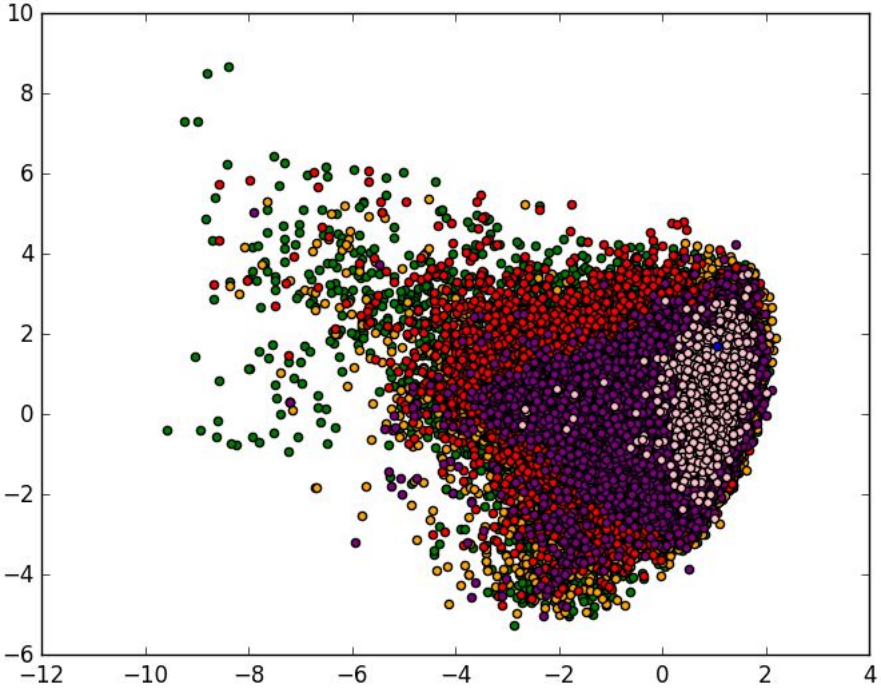
Instrumentalness vs Loudness





Principal Component Analysis





Linear Discriminant Analysis

Conclusions - Future Work

- Key features:
  - Loudness
  - Instrumentalness
- Reprocess track dataset to include genres, number of followers, and related artist popularity
- Identify and remove erroneous entries

Questions?