Projet D'Analyse de Données. Professeur: Mme Fanny Villiers

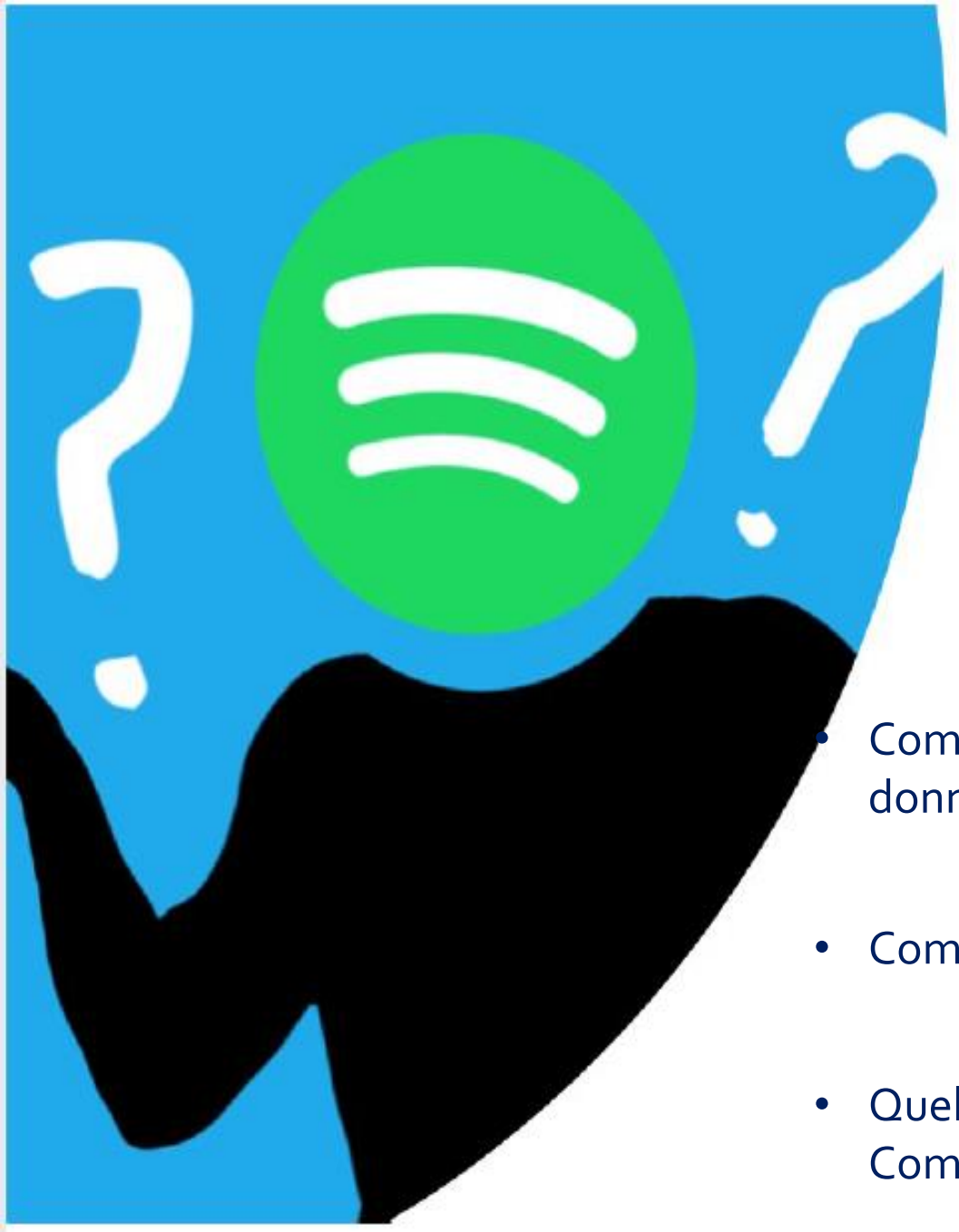
«Spotify Audio Features»

Analyse d'une base de données musicale

POLYTECH SORBONNE

SCIENCES SORBONNE UNIVERSITÉ

Najwa Moursli, Paolo Conti



Quel est le secret de la popularité d'un artiste?

Pourquoi une chanson serait plus célèbre qu'une autre?

STRATEGIES:

- Comment travailler sur un *grand* jeu de données? → **Analyse Descriptive et AFC/ACP**
- Comment identifier les *genres musicaux* ? → **Classification non Supervisée**
- Quels critères déterminent la popularité? Comment peut-on prédire cette popularité? → **Régression linéaire et Prédiction**

DATASET

#chansons: n = 130326
#variables: p = 17

- **3 VARIABLES DESCRIPTIVES:**

Artist name, Track ID, Track Name.

- **2 VARIABLES QUALITATIVES:**

Key, Mode.

- **12 VARIABLES QUANTITATIVES:**

Acousticness, Energy, Danceability, Loudness, Speechiness, Instrumentalness, Liveness, Valence, Tempo, Duration, Time signature, **Popularity**.

Statistiques Descriptives

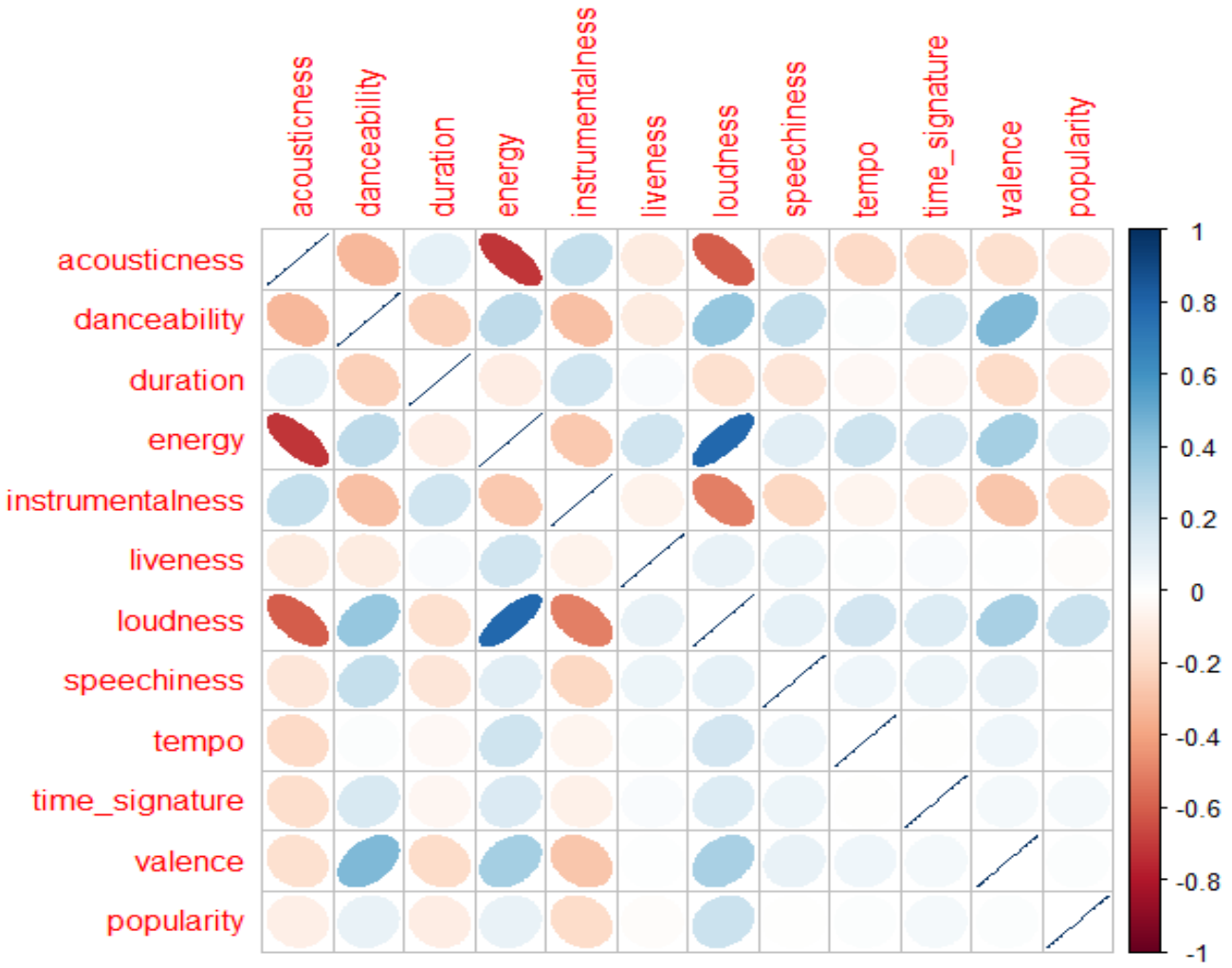


Table des Corrélations

→ **Analyse bivariée exploratoire ++**

→ **Pas assez d'information sur POPULARITE**

PASSAGE OBLIGATOIRE ACP/AFC !!!

AFC / ACP

qualitatives

KEY

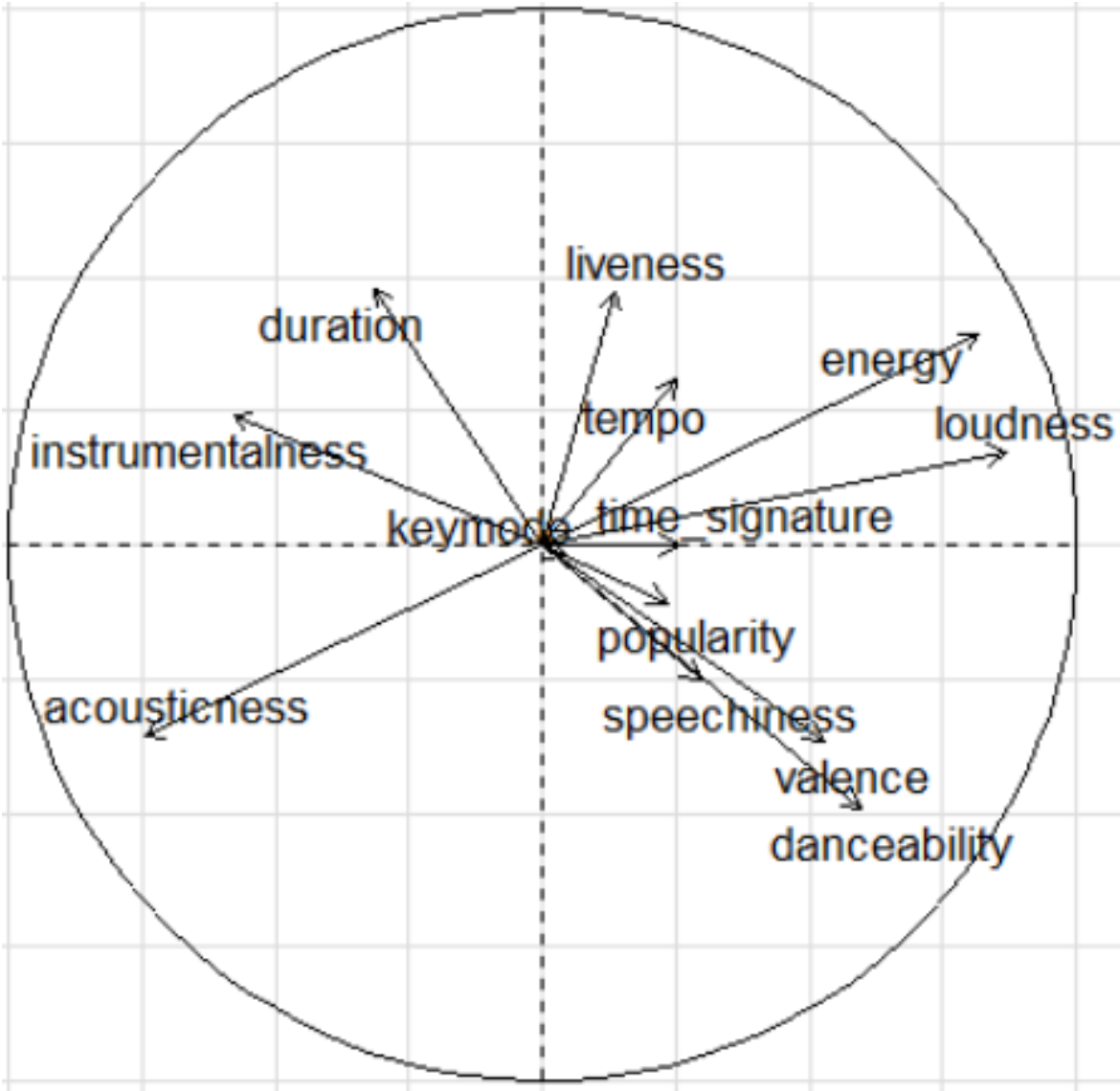
MODE

AFC

quantitative

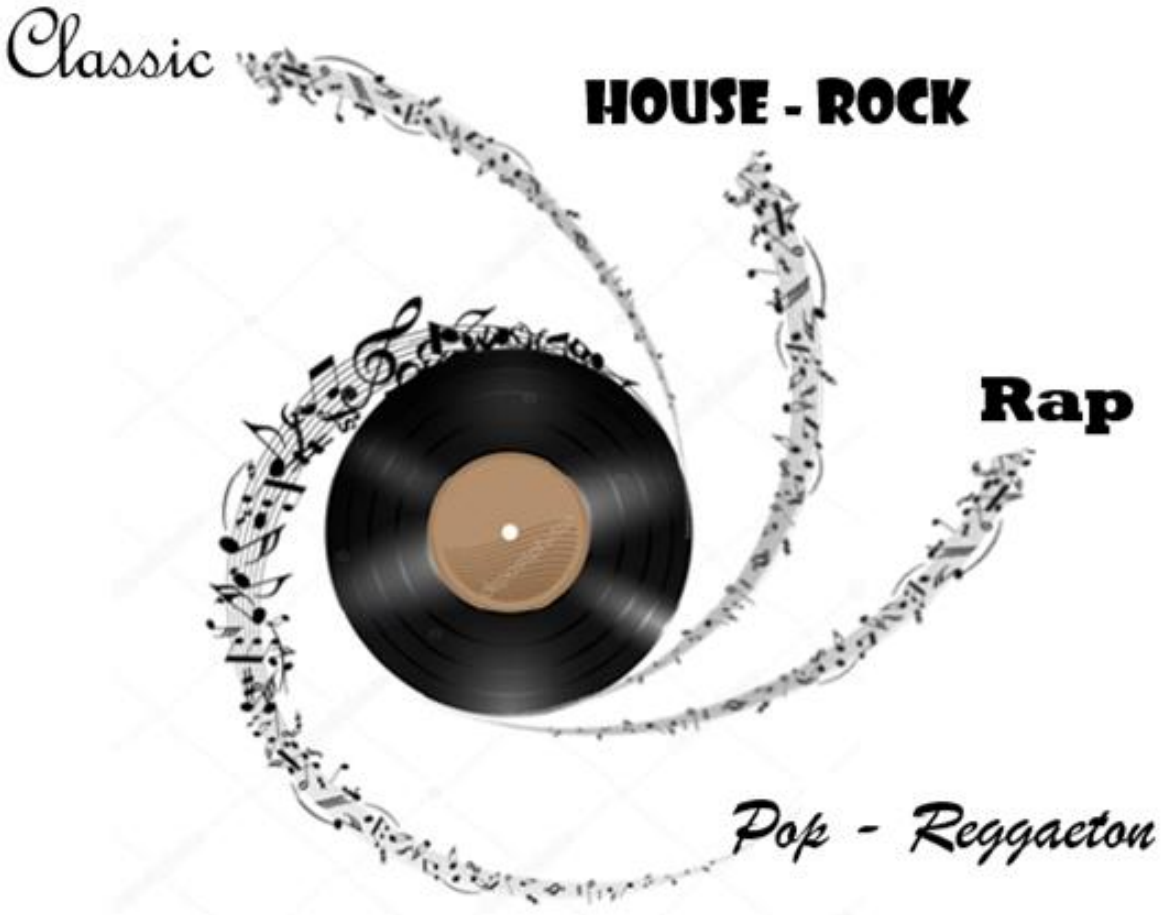
KEYMODE

ACP



ACP axes 1-2

Peut-on créer des clusters représentants des *GENRES MUSICAUX?*



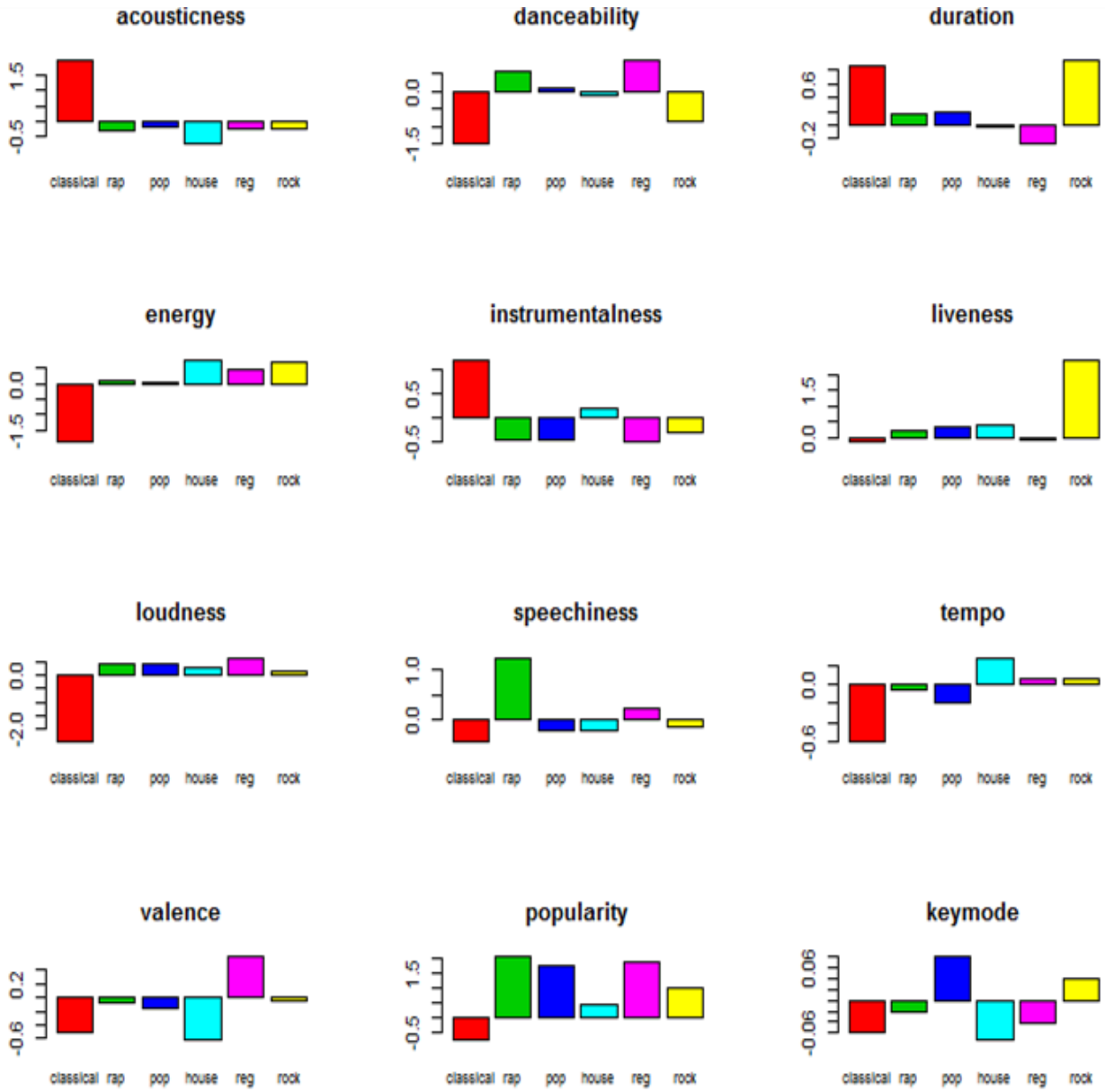
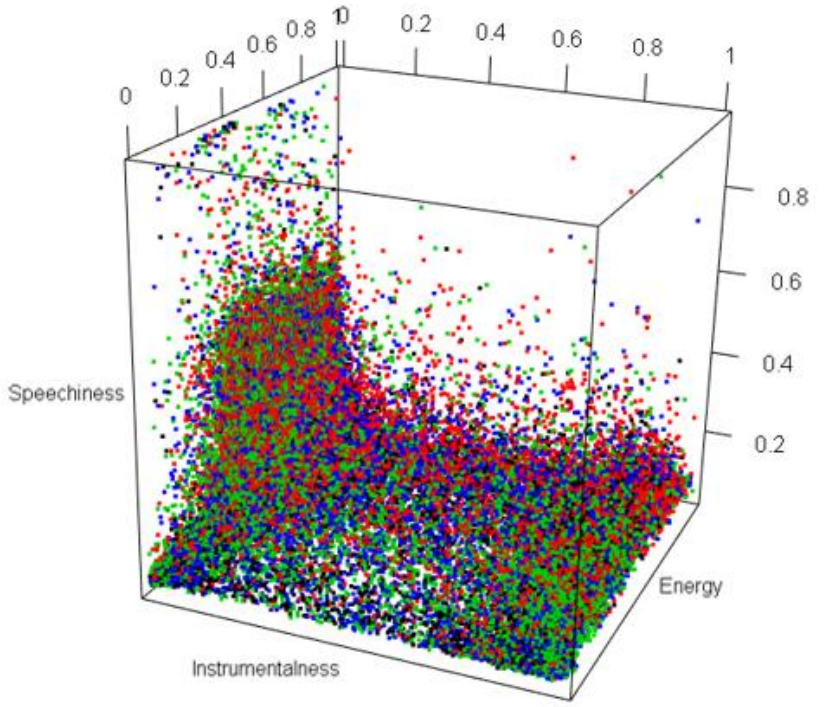
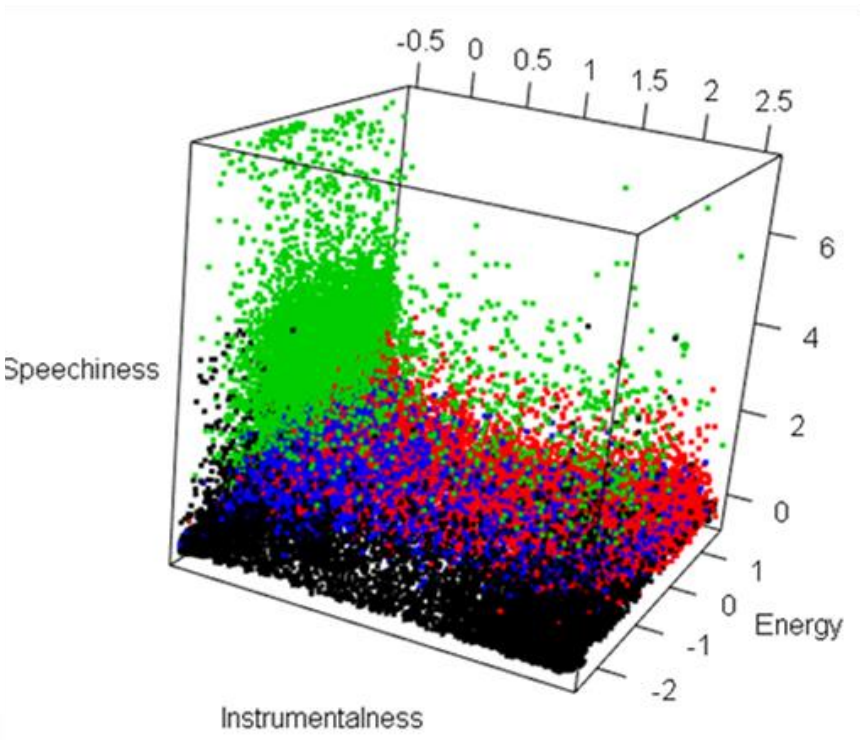


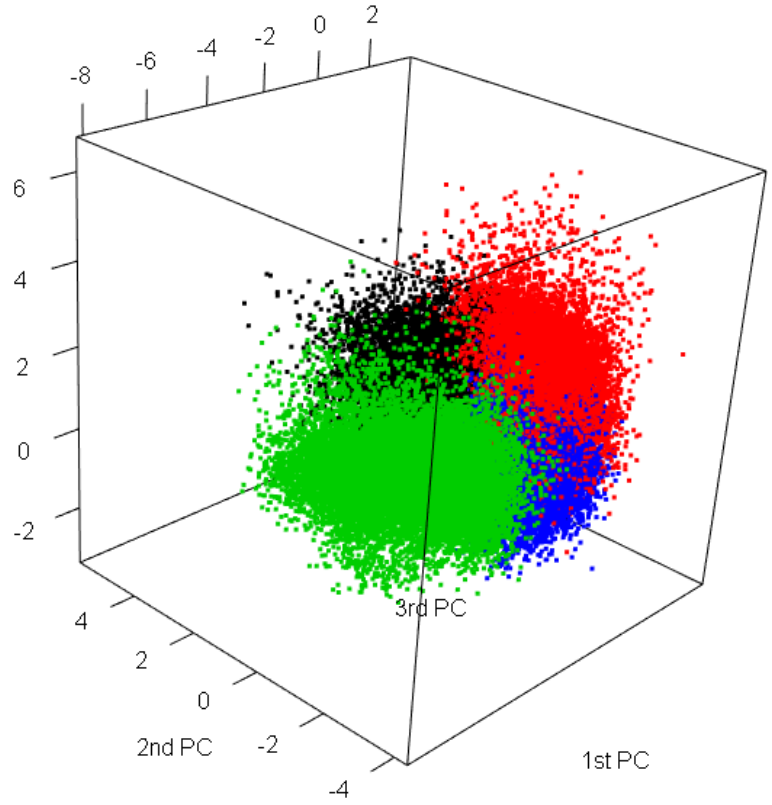
Diagramme en bâtons des différentes variables pour chaque genre

CLASSIFICATION NON SUPERVISÉE : *ALGORITHME DES K-MEANS*

Choix du nombre de classes: 4
Choix des centres: centres aléatoires

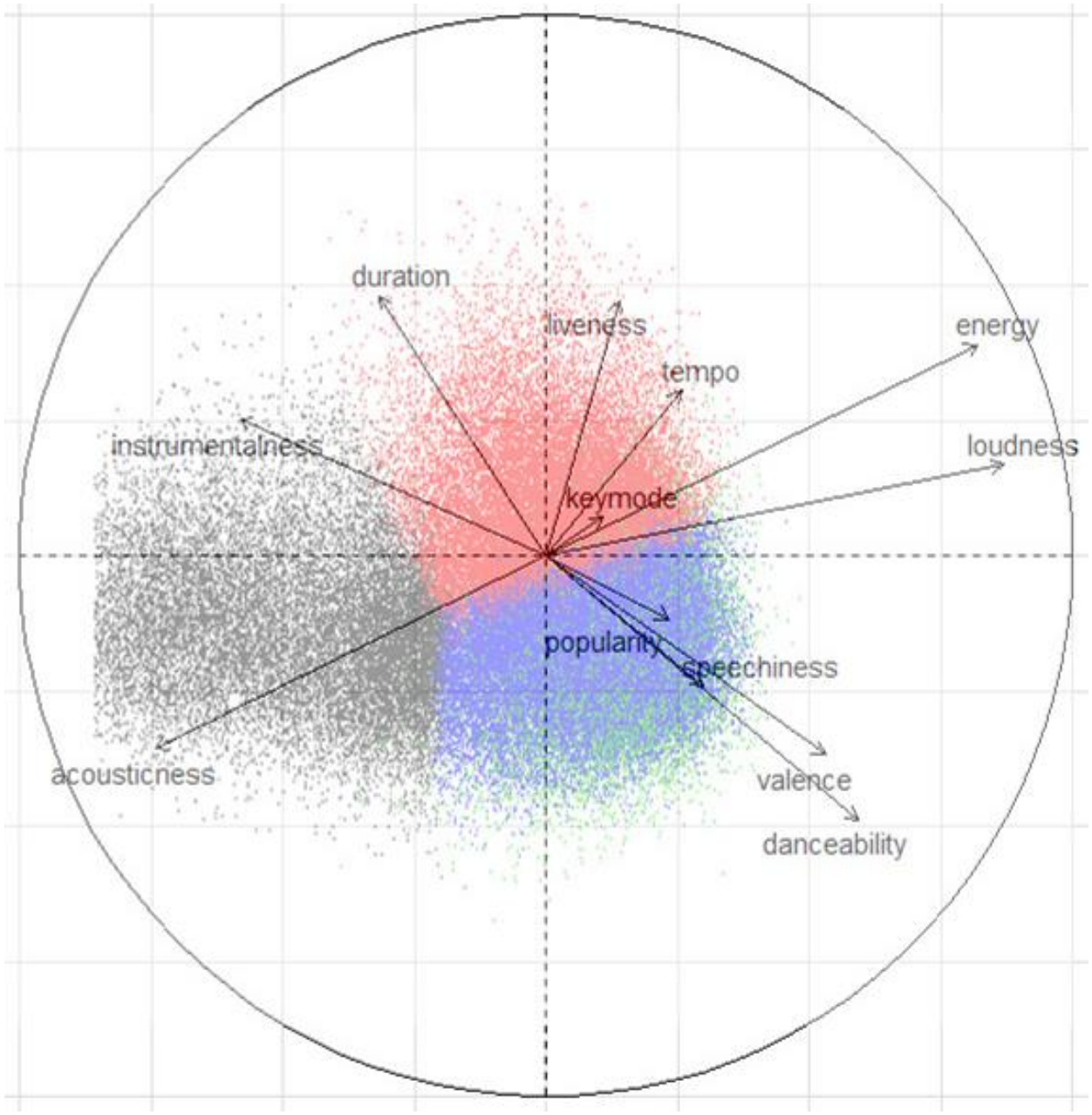




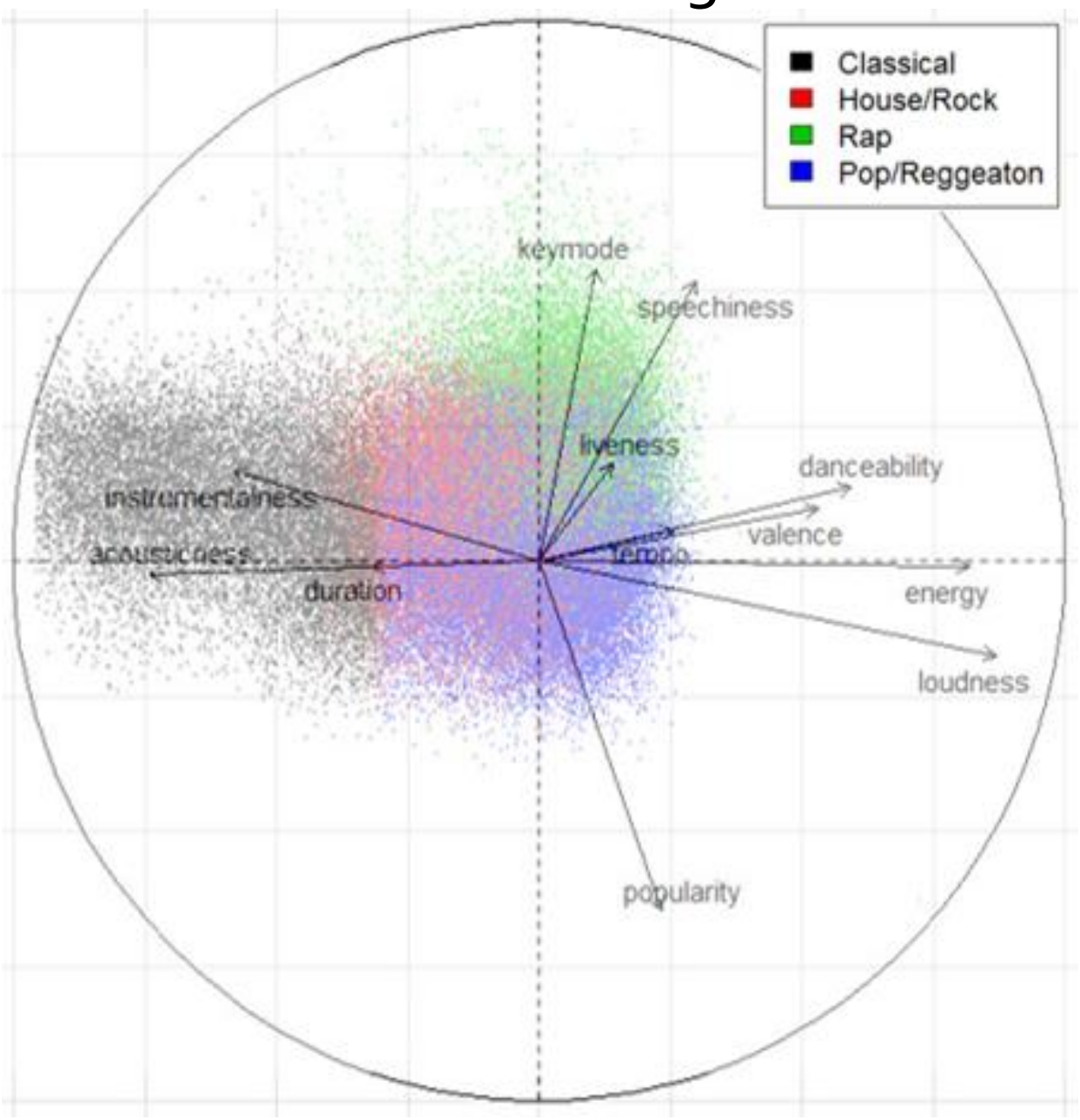


Base de Données initiale → **Données échelonnées** → **Composantes Principales**

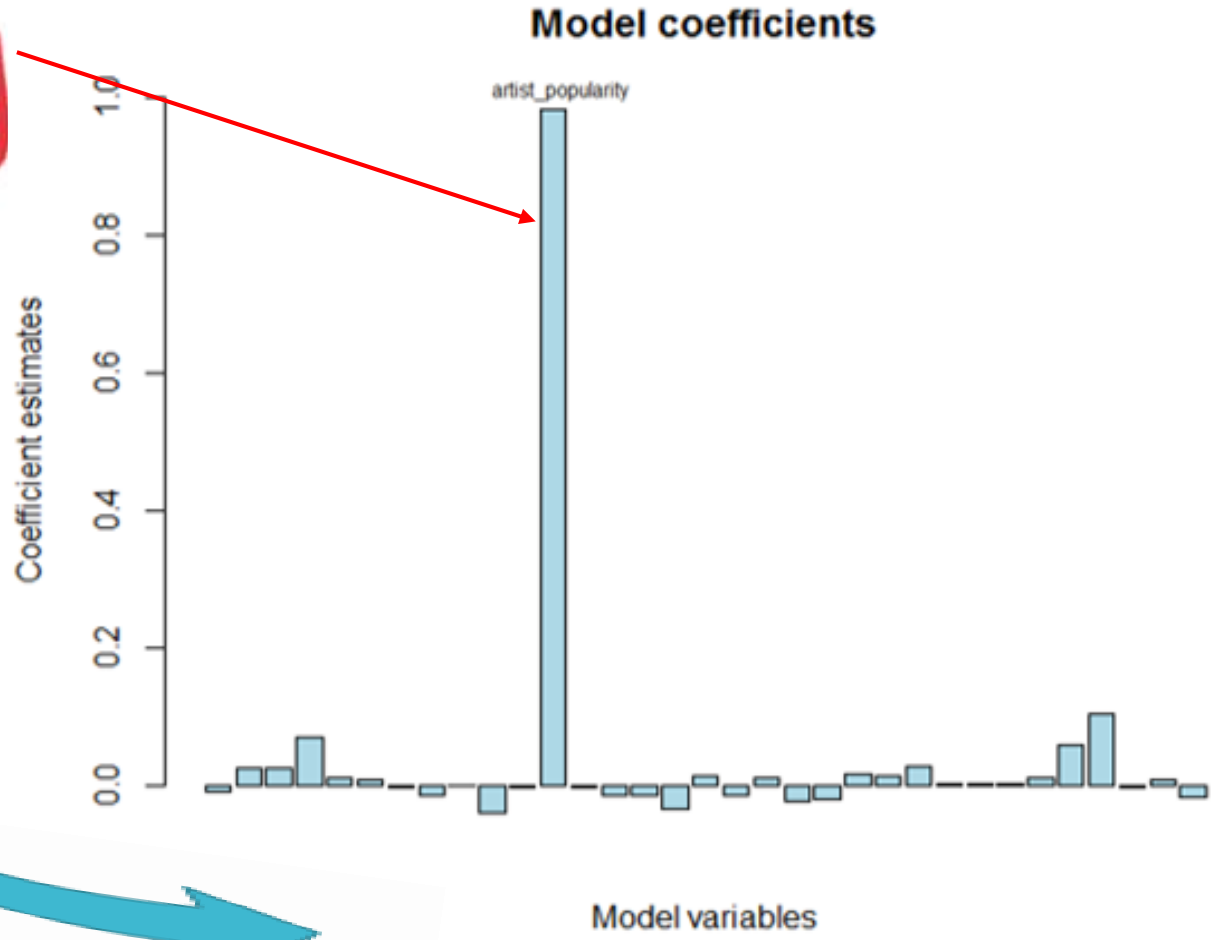
ACP axes 1-2



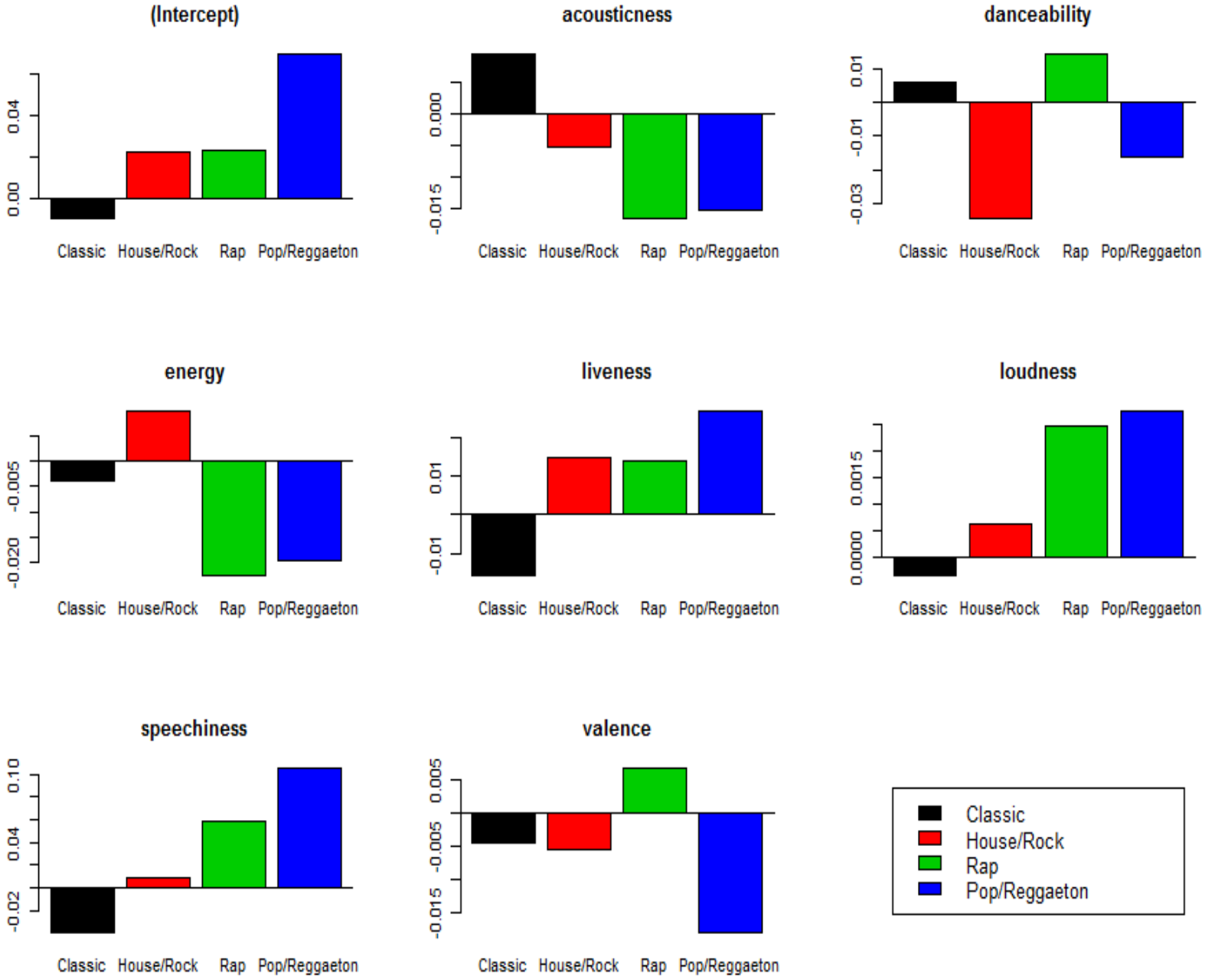
ACP axes 1-3



Nouvelle variable:
«POPULARITÉ DE L'ARTISTE»

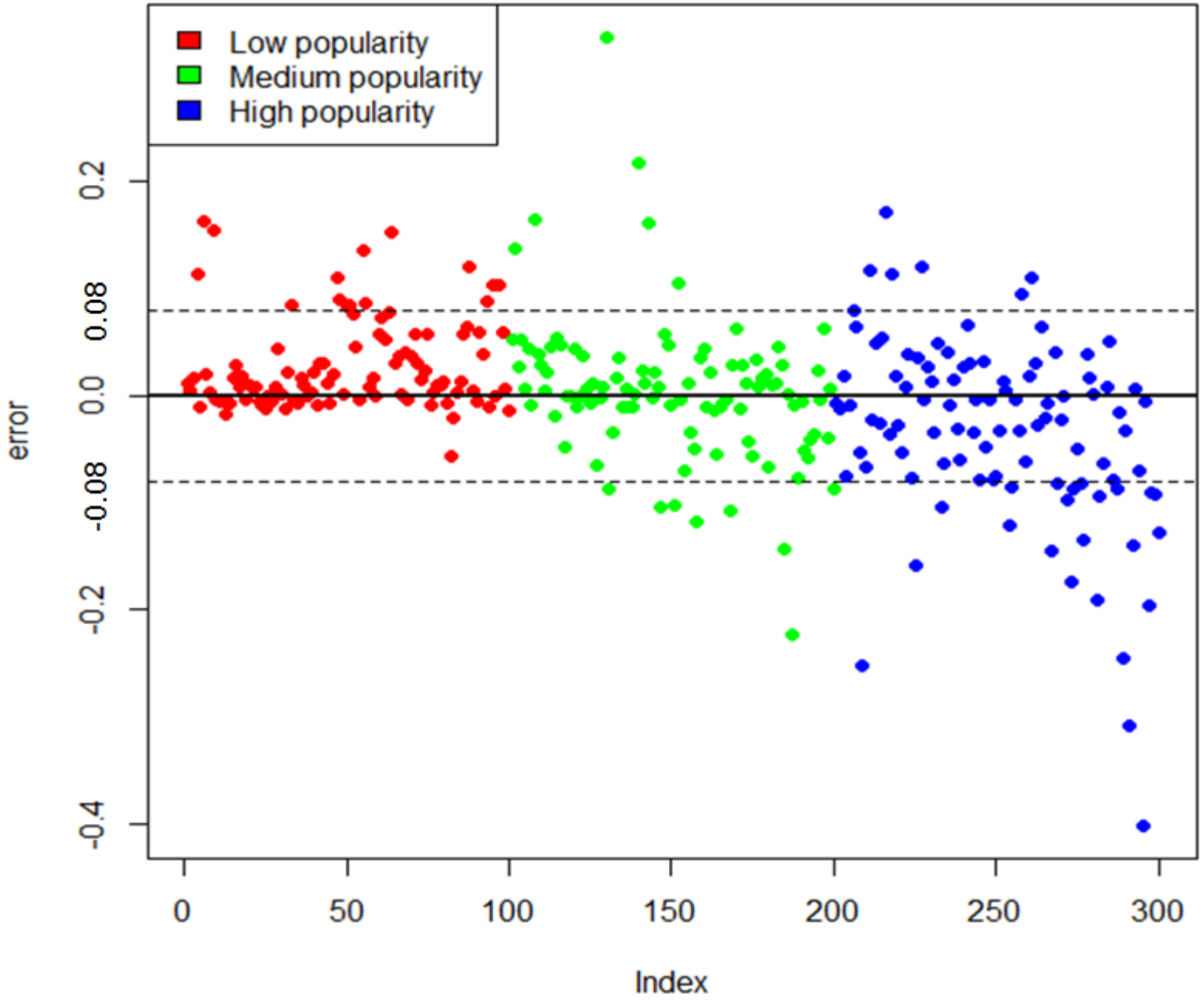


COEFFICIENTS DU MODÈLE



PRÉDICTION

erreur moyenne: 0.05 (~5%)
variance: 0.003



Merci de votre attention!

Dataset source: Kaggle "Spotify Audio Features"
https://www.kaggle.com/tomigelo/spotify-audio-features