

# Analyse descriptive du jeu de données Spotify

## Projet en Statistique descriptive

### **Membres**

LOULIDI Younes

PHAM Tuan Kiet

VO Van Nghia

### **Date**

17 Mars, 2021

# Table des matières

Table des matières	i
<b>1 Statistiques descriptives unidimensionnelle et bidimensionnelle</b>	<b>1</b>
1.1 La nature des jeux de données . . . . .	1
1.1.1 Des jeux de données . . . . .	1
1.1.2 Des variables statistiques . . . . .	1
1.1.3 Charger le jeu de données dans R . . . . .	2

# 1 Statistiques descriptives unidimensionnelle et bidimensionnelle

## 1.1 La nature des jeux de données

### 1.1.1 Des jeux de données

Ces jeux de données se composent de 10000 chansons extraites de la base de données Spotify.

Chaque ligne contient 11 variables statistiques comme suit:

- **year**: année de sortie du morceau,
- **acousticness**: métrique relative interne de l'acoustique morceau,
- **duration**: durée du morceau en millisecondes (ms),
- **energy**: métrique relative interne de l'intensité, des rythmes du morceau,
- **explicit**: vaut 1 si le morceau contient des vulgarités, et 0 sinon,
- **key**: tonalité en début de morceau,
- **liveness**: proportion du morceau où l'on entend un public,
- **loudness**: mesure relative du volume du morceau (en décibels, dB)
- **mode**: mode du morceau (0 si la tonalité est mineure, et 1 si la tonalité est majeure),
- **tempo**: le tempo du morceau, en battement par minute (bpm),
- **pop.class**: la popularité du morceau.

### 1.1.2 Des variables statistiques

Ici, nous précisons la nature de chaque variable et son format dans R.

Nom de variable statistique	Type de variable	Format dans R
<b>year</b>	qualitative ordinale	<b>integer</b>
<b>acousticness</b>	quantitative continue	<b>numeric</b>
<b>duration</b>	quantitative continue <sup>1</sup>	<b>numeric</b>

Nom de variable statistique	Type de variable	Format dans R
energy	quantitative continue	numeric
explicit	qualitative nominale	logical
key	qualitative nominale	factor
liveness	quantitative continue	numeric
loudness	quantitative continue	numeric
mode	qualitative nominale <sup>2</sup>	logical
tempo	quantitative continue	numeric
pop.class	qualitative nominale	factor

### 1.1.3 Charger le jeu de données dans R

```
LoadDataset <- function(fname) {
  colclasses <- c(
    "integer", "numeric", "numeric",
    "numeric", "integer", "factor", "numeric",
    "numeric", "integer", "numeric", "factor"
  )
  dataframe <- read.csv(fname, colClasses = colclasses)
  dataframe$explicit <- as.logical(dataframe$explicit)
  dataframe$mode <- as.logical(dataframe$mode)
  return(dataframe)
}

daf <- LoadDataset("dataset.csv")
str(daf)
```

<sup>1</sup>On choisi son nature est de quantitative continue parce que.

<sup>2</sup>On pose FALSE si la tonalité est mineure et TRUE si non.

```
## 'data.frame':    10000 obs. of  11 variables:
## $ year          : int  1998 1992 1973 1969 2008 2015 1935 1928 2013 1945 ...
## $ acousticness: num  0.147 0.193 0.388 0.733 0.979 0.0742 0.99 0.995 0.000506 0.98 .
## $ duration      : num  148520 189800 289267 170267 438907 ...
## $ energy        : num  0.74 0.389 0.856 0.454 0.494 0.766 0.42 0.211 0.53 0.106 ...
## $ explicit      : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
## $ key           : Factor w/ 12 levels "A","Ab","B","Bb",...: 7 5 1 10 11 11 2 6 12 12 .
## $ liveness      : num  0.0452 0.154 0.139 0.0889 0.123 0.0827 0.13 0.106 0.0477 0.237
## $ loudness      : num  -8.16 -11.64 -8.4 -8.12 -10.65 ...
## $ mode          : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
## $ tempo         : num  157.1 85.3 101.3 82.4 156.3 ...
## $ pop.class     : Factor w/ 4 levels "A","B","C","D": 3 3 3 3 3 1 4 4 2 4 ...
```