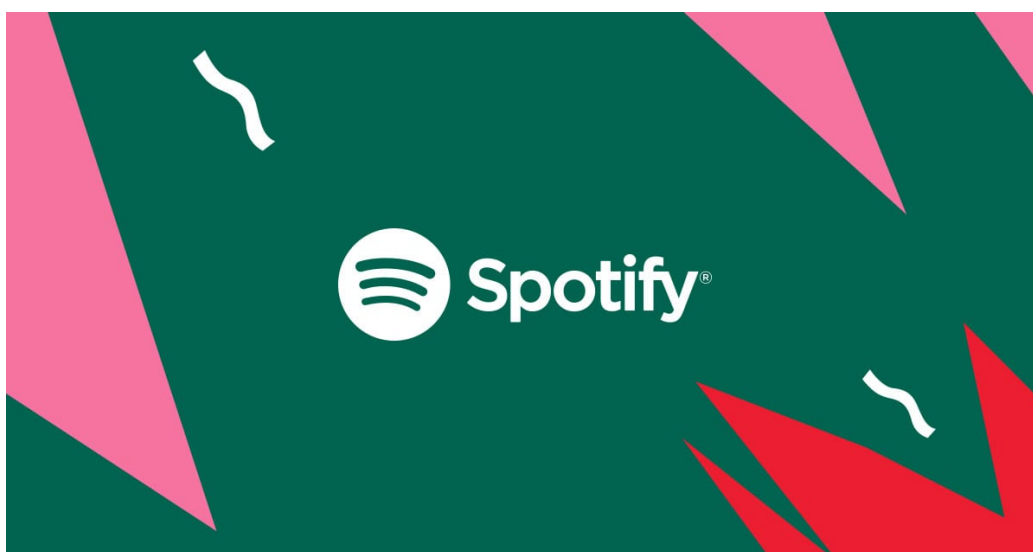


Najlepsze piosenki na imprezę w oparciu o dane ze Spotify

Edyta Łabędzka, Iga Świtalska

9 lutego 2022



1. Wprowadzenie

1.1. Opis danych

Dane, które będziemy analizować dotyczą cech utworów, pochodzących z notowania top 2000 najlepszych piosenek wszech czasów, dostępnego na platformie Spotify w 2019 roku. Dane "Top 2000" zawierają 15 kolumn oraz 1994 utwory. W dalszej części analizy będziemy się odnosić jedynie do podzbioru powyższych danych bez zmiennej "Index" oraz zmiennej "Speechiness". Kolumny "Index" nie wykorzystujemy, ponieważ R automatycznie numeruje dane. Natomiast kolumnę "Speechiness" zdecydowaliśmy się wykluczyć ze względu na trudności z jednoznaczną jej interpretacją. Dokładny opis zmiennych znajduje się w tabeli 1.1.

Zmienne	Opis	Typ	Wartości
Title	Nazwa utworu	Kategoryczna	1958 unikalnych tytułów
Artist	Nazwa wykonawcy	Kategoryczna	731 artystów
Top Genre	Gatunek utworu	Kategoryczna	149 gatunków
Year	Rok wydania	Kategoryczna	{1956, 1957, ..., 2019}
Beats Per Minute (BPM)	Tempo utworu (ilość uderzeń na minutę)	Ilościowa	{0, 1, ...}
Energy	Miara intensywności i aktywności piosenki	Ilościowa	{0, 1, ..., 100}
Danceability	Jak bardzo nadaje się utwór do tańca (dla wartości większych utwory bardziej taneczne)	Ilościowa	{0, 1, ..., 100}
Loudness (dB)	Głośność (w decybelach)	Ilościowa	\mathbb{Z}
Liveness	Opisuje prawdopodobieństwo, że utwór został nagrany na żywo (w procentach)	Ilościowa	{0, 1, ..., 100}
Valence	Opisuje pozytywność muzyczną (dla wyższych wartości utwory brzmią bardziej pogodnie)	Ilościowa	{0, 1, ..., 100}
Length (Duration)	Długość (w sekundach)	Ilościowa	{0, 1, ...}
Acousticness	Określa prawdopodobieństwo czy utwór jest akustyczny (w procentach)	Ilościowa	{0, 1, ..., 100}
Popularity	Popularność utworu	Ilościowa	{0, 1, ..., 100}

Tabela 1.1. Opis zmiennych

1.2. Cel projektu

Celem analizy będzie wybranie z analizowanej playlisty najlepszych utworów na przyjęcie. Rozważymy wiele czynników mogących wpływać na dobrą zabawę przy muzyce i postaramy się pomóc w doborze piosenek. Chcemy poruszyć poniższe problemy:

- Jakie są najlepsze piosenki do tańca na podstawie ich charakterystyk?
- Które gatunki najbardziej się nadają?
- Z jakich lat utwory wybrać, by była największa szansa, że ludziom się spodoba?
- Jeżeli nie chcemy szukać piosenek pojedynczo, muzykę jakiego wykonawcy puścić?

2. Przygotowanie danych

Importujemy potrzebne biblioteki.

```
library(xtable)
library(ggplot2)
library(corrplot)
library(stringr)
library(plyr)
library(nortest)
library(lmtest)
library(data.table)
```

Wczytujemy dane w R.

```
data <- read.csv("Spotify-2000.csv")
```

Najpierw zmienimy nazwy kilku zmiennych, by były bardziej czytelne.

```
colnames(data)[c(3,5,8,11)] <- c("Genre", "BPM", "Loudness", "Duration")
```

Sprawdzimy i przygotujemy dane pod dalszą analizę. Użyjemy do tego funkcji "summary" i wyniki zaprezentujemy w tabeli 2.1.

Title	Artist	Genre	Year	BPM	Energy	Danceability
Length:1994	Length:1994	Length:1994	Min. :1956	Min. : 37.0	Min. : 3.00	Min. :10.00
Class :character	Class :character	Class :character	1st Qu.:1979	1st Qu.: 99.0	1st Qu.: 42.00	1st Qu.:43.00
Mode :character	Mode :character	Mode :character	Median :1993	Median :119.0	Median : 61.00	Median :53.00
			Mean :1993	Mean :120.2	Mean : 59.68	Mean :53.24
			3rd Qu.:2007	3rd Qu.:136.0	3rd Qu.: 78.00	3rd Qu.:64.00
			Max. :2019	Max. :206.0	Max. :100.00	Max. :96.00
			NA's :19			
Loudness	Liveness	Valence	Duration	Acousticness	Popularity	
Min. :-27.000	Min. : 2.00	Min. : 3.00	Length:1994	Min. : 0.00	Min. : 11.00	
1st Qu.: -11.000	1st Qu.: 9.00	1st Qu.:29.00	Class :character	1st Qu.: 3.00	1st Qu.: 49.25	
Median : -8.000	Median :12.00	Median :47.00	Mode :character	Median :18.00	Median : 62.00	
Mean : -9.009	Mean :19.01	Mean :49.41		Mean :28.86	Mean : 59.53	
3rd Qu.: -6.000	3rd Qu.:23.00	3rd Qu.:69.75		3rd Qu.:50.00	3rd Qu.: 71.00	
Max. : -2.000	Max. :99.00	Max. :99.00		Max. :99.00	Max. :100.00	

Tabela 2.1. Podsumowanie danych

Okazuje się, że długość utworu jest zmienną typu "character" a powinna być oczywiście liczbą. Spowodowane jest to przez pojawiające się w liczbach przecinki tzw. "decimal points" typowe w tekstach anglojęzycznych. Usuwamy przecinki i zmieniamy typ na liczby całkowite.

```
data$Duration <- str_replace(data$Duration, ",", "")
class(data$Duration) <- "integer"
```

W tabeli 2.1 pojawiają się również braki danych w kolumnie "Year". Widzimy, że jest ich jedynie 19. Mając dostęp do Spotify (lub internetu) najrozsądniejszym rozwiązaniem jest sprawdzenie którym utworom i autorom odpowiadają, wyszukanie ich i uzupełnienie brakujących wartości.

```
data[is.na(data$Year),c(1,2,4)]
```

##		Title	Artist	Year
## 179		Another 45 Miles	Golden Earring	NA
## 195		Tumble and Fall	JOHAN	NA
## 415		A Night Like This	Caro Emerald	NA
## 463		Happiness	Jonathan Jeremiah	NA
## 526		Another Love	Tom Odell	NA
## 602		Nothing Really Matters	Mr. Probz	NA
## 665		I Just Can't Help Believin'	Elvis Presley	NA
## 709		Happier	Ed Sheeran	NA
## 938		Woman from Tokyo	Deep Purple	NA
## 1011		Fernando	ABBA	NA
## 1038		The Chain - 2004 Remaster	Fleetwood Mac	NA
## 1142		Off the Wall	Michael Jackson	NA
## 1253		I Heard It Through The Grapevine	Marvin Gaye	NA
## 1268		When The Lady Smiles	Golden Earring	NA
## 1323		Nikita	Elton John	NA
## 1398		Smooth Criminal - 2012 Remaster	Michael Jackson	NA
## 1627		Round Here	Counting Crows	NA
## 1792		I Can't Make You Love Me	George Michael	NA
## 1842		Folsom Prison Blues	Johnny Cash	NA

Znalezione brakujące lata, którymi uzupełnimy dane.

```
## [1] 2008 2001 2010 2011 2013 2014 1970 2017 1973 1976
## [1] 1977 1979 1968 1984 1985 1987 1993 1998 2012
```

Chcemy również sprawdzić czy są duplikaty.

```
sum(duplicated(data))
## [1] 0
```

Duplikatami w takim razie nie musimy się przejmować.

Ostatnią rzeczą którą zmieniamy będą lata. Połączymy je w przedziały po 10 lat, w ten sposób uzyskujemy zmienną kategoryczną, która pozwoli nam na analizowanie danych dekadami.

Tworzymy następujące przedziały:

```
## [1] "[1950,1960)" "[1960,1970)" "[1970,1980)" "[1980,1990)"
## [1] "[1990,2000)" "[2000,2010)" "[2010,2020)"
```

W tabeli 2.2 widzimy podsumowanie danych po wprowadzonych zmianach. Porównując tabele 2.1 i 2.2, widzimy nowe nazwy zmiennych, nie ma

już żadnych braków danych, obserwacje dla długości utworu są liczbami oraz rok wydania stał się zmienną kategoriyczną.

Title	Artist	Genre	Year	BPM	Energy	Danceability
Length:1994	Length:1994	Length:1994	Length:1994	Min. : 37.0	Min. : 3.00	Min. :10.00
Class :character	Class :character	Class :character	Class :character	1st Qu.: 99.0	1st Qu.: 42.00	1st Qu.:43.00
Mode :character	Mode :character	Mode :character	Mode :character	Median :119.0	Median : 61.00	Median :53.00
				Mean :120.2	Mean : 59.68	Mean :53.24
				3rd Qu.:136.0	3rd Qu.: 78.00	3rd Qu.:64.00
				Max. :206.0	Max. :100.00	Max. :96.00
Loudness	Liveness	Valence	Duration	Acousticness	Popularity	
Min. :-27.000	Min. : 2.00	Min. : 3.00	Min. : 93.0	Min. : 0.00	Min. : 11.00	
1st Qu.: -11.000	1st Qu.: 9.00	1st Qu.:29.00	1st Qu.: 212.0	1st Qu.: 3.00	1st Qu.: 49.25	
Median : -8.000	Median :12.00	Median :47.00	Median : 245.0	Median :18.00	Median : 62.00	
Mean : -9.009	Mean :19.01	Mean :49.41	Mean : 262.4	Mean :28.86	Mean : 59.53	
3rd Qu.: -6.000	3rd Qu.:23.00	3rd Qu.:69.75	3rd Qu.: 289.0	3rd Qu.:50.00	3rd Qu.: 71.00	
Max. : -2.000	Max. :99.00	Max. :99.00	Max. :1412.0	Max. :99.00	Max. :100.00	

Tabela 2.2. Podsumowanie danych po zmianach

3. Analiza danych

3.1. Jakie są najlepsze piosenki do tańca na podstawie ich charakterystyk?

Na dobrym przyjęciu nie może zabraknąć tańca, a do tańca potrzeba odpowiedniej muzyki. Spróbujmy taką znaleźć, zaczniemy od przyjrzenia się macierzy korelacji.

```
corrplot(cor(data[, -c(1, 2, 3, 4)]), method = "number")
```

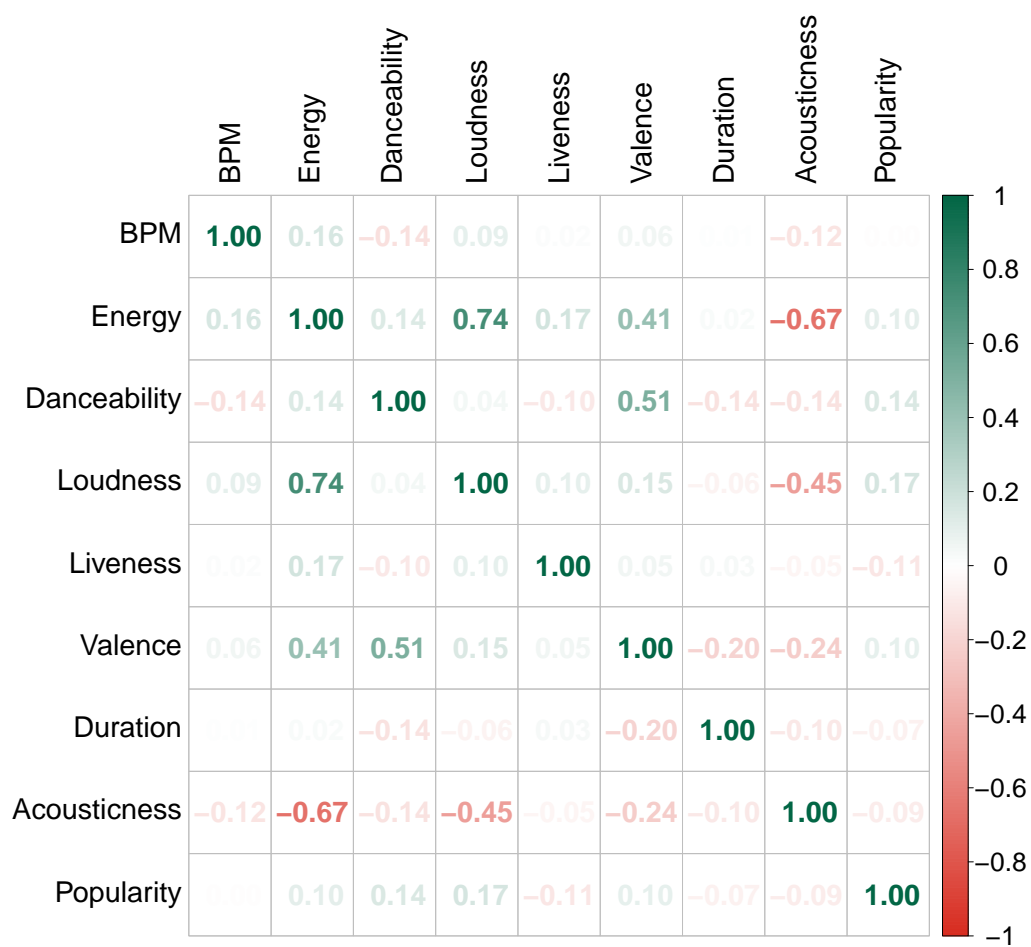
Na podstawie wykresu 3.1 widzimy, że "Danceability" jest dodatnio skorelowane z "Valence". Nie jest to zaskakująca obserwacja, lepiej się tańczy do radosnej muzyki. Na wykresie rozproszenia 3.2 dane wyglądają jakby układały się wzdłuż rosnącej prostej, spróbujemy dopasować do nich prostą regresji liniowej.

Najpierw sprawdzamy czy są spełnione założenia regresji liniowej. Przeprowadzamy analizę residuów:

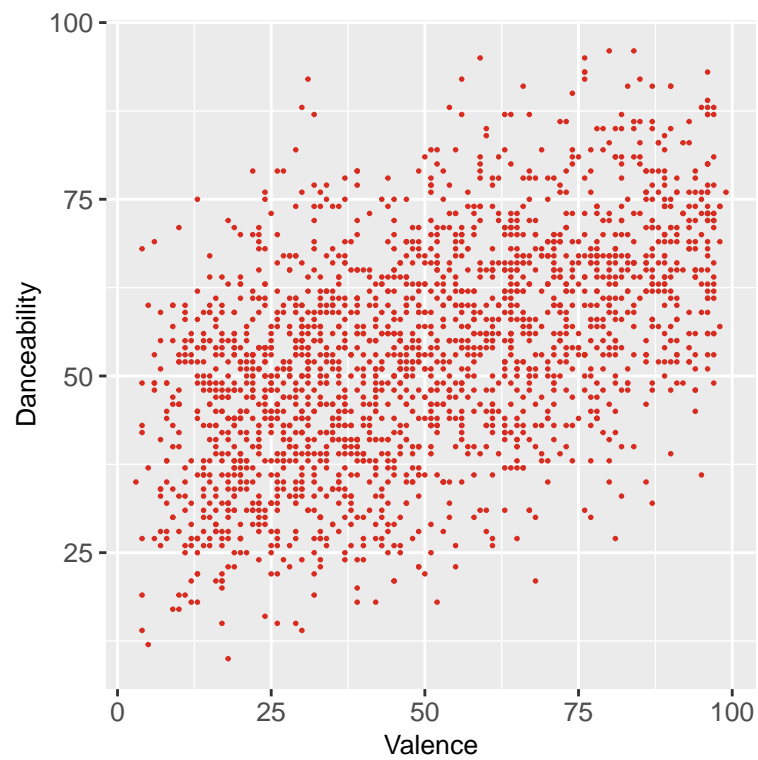
```
lmFit <- lm(data$Danceability~data$Valence)
```

```
##  
## Anderson-Darling normality test  
##  
## data: residuals  
## A = 0.67317, p-value = 0.07867  
##  
## Durbin-Watson test  
##  
## data: lmFit  
## DW = 1.8389, p-value = 0.0001603  
## alternative hypothesis: true autocorrelation is greater than 0
```

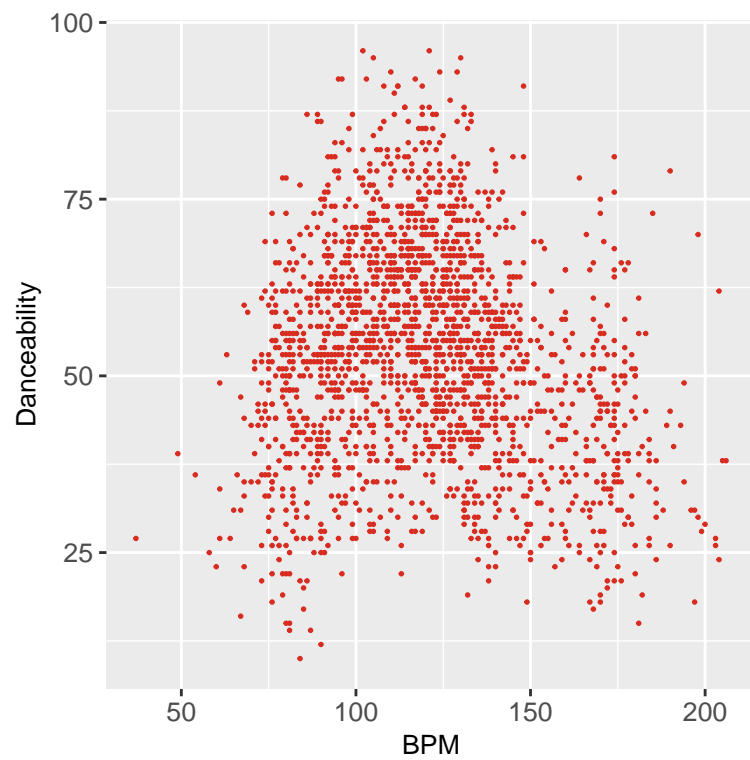
Z testu Andersona-Darlinga wynika, że nie mamy podstaw do odrzucenia założenia o rozkładzie normalnym residuów. Natomiast wynik testu Durбина-Watsona wskazuje na autokorelację między residuami przez co nie możemy zastosować regresji liniowej. Jedyne z czego możemy skorzystać po powyższej analizie to dodatnia korelacja. Wynika z niej, że radośniejsze piosenki bardziej nadają się do tańca.



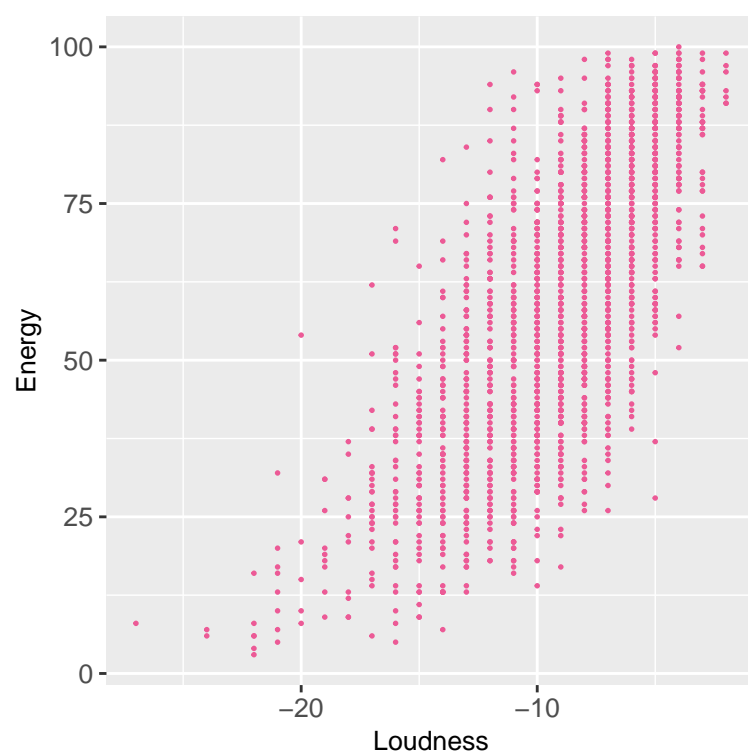
Rysunek 3.1. Macierz koleracji



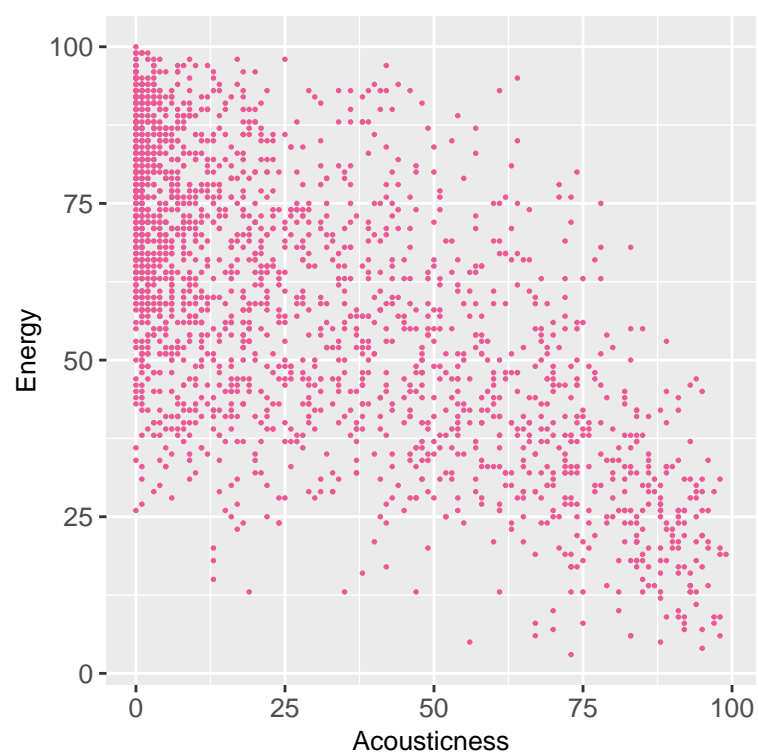
Rysunek 3.2. Wykres rozproszenia "Danceability" od "Valence"



Rysunek 3.3. Wykres rozproszenia taneczności od tempa



Rysunek 3.4. Wykres rozproszenia energiczności od głośności



Rysunek 3.5. Wykres rozproszenia energiczności od akustyczności

Intuicja podpowiada nam, że tempo utworu może mieć wpływ na to, czy piosenka nadaje się do tańca. Co nie koniecznie jest zależnością liniową, dlatego nie jest widoczne w macierzy korelacji. Zobaczmy wykres rozproszenia.

Na wykresie 3.3 taneczność przymiemy największe wartości dla tempa w okolicy 100 – 150 uderzeń na minutę. Najniższe natomiast dla wartości skrajnych tempa. Nasuwa to rozsądny wniosek, że do za szybkiej jak i zbyt wolnej muzyki źle się tańczy.

Uważamy, że piosenki do tańca na imprezie powinny być również energiczne, dlatego też przeanalizujemy zmienną "Energy". Podobnie jak poprzednio posłużymy się macierzą korelacji 3.1, na której jest widoczna duża dodatnia korelacja z "Loudness" oraz ujemna korelacja z "Acousticness". Zobaczmy te zależności na wykresach rozproszenia 3.4 i 3.5.

Po przyjrzeniu się wykresowi 3.4 możemy stwierdzić, że im głośniejsze piosenki tym bardziej energiczne. Natomiast na wykresie 3.5 wraz z rosnącą akustycznością, energiczność maleje. Nie ma utworów akustycznych, które byłyby bardzo energiczne, dlatego nie są one dobrym wyborem na przyjęcie.

3.2. Które gatunki najbardziej się nadają?

Chcemy uporządkować gatunki. Jest ich zbyt dużo, oraz wiele gatunków odpowiada tylko jednemu utworowi. Zobaczmy 5 najliczniejszych gatunków, którym jest przyporządkowanych najwięcej utworów i 5 najmniej licznych.

```
##          Genre freq
## 6      album rock 413
## 3    adult standards 123
## 85      dutch pop  88
## 13  alternative rock 86
## 72      dance pop  83
## 144     soft rock   1
## 145 stomp and holler 1
## 146     streektaal  1
## 147         trance  1
## 148         uk pop   1
## 149     yacht rock   1
```

Postanowiliśmy przyporządkować gatunki do kilku najbardziej popularnych kategorii:

```
## [1] "rock"    "pop"     "rap"     "blues"   "dance"   "country"
## [1] "metal"   "hip hop" "indie"   "latin"   "jazz"    "soul"
## [1] "mellow"
```

Rodzaj niektórych piosenek stanowi połączenie dwóch podstawowych gatunków, można więc przyporządkować je do więcej niż jednej kategorii. Na przykład gatunek "alternative pop rock" może być uznany jako pop i jako rock. Postanowiliśmy utworzyć kopię naszego zbioru danych ze zduplikowanymi

utworami, które należą do kilku rodzajów. Podjęliśmy taką decyzję, żeby nie utracić informacji wynikającej z tego, że utwory mogą być mieszanką gatunków i żeby równocześnie nie zaburzać pozostałych statystyk. W tej części analizy będziemy się odnosić jedynie do zmodyfikowanych danych.

Po wstępnej obróbce danych z 149 otrzymujemy 63 kategorii. Następnie piosenki z gatunku "hip-hop" przypisujemy do "rapu", a "electro house" i "electornica" do "dance ze względu na ich podobieństwo.

```
data2$Genre[data2$Genre == "hip hop"] <- "rap"
data2$Genre[data2$Genre %in% c("electro house", "electronica")] <- "dance"
```

Na koniec gatunkom które mają mniej niż 10 utworów przypisujemy kategorię "other", ponieważ analiza tak małej ilości utworów nie powie nam nic o tych gatunkach.

Ostatecznie mamy 20 następujących kategorii:

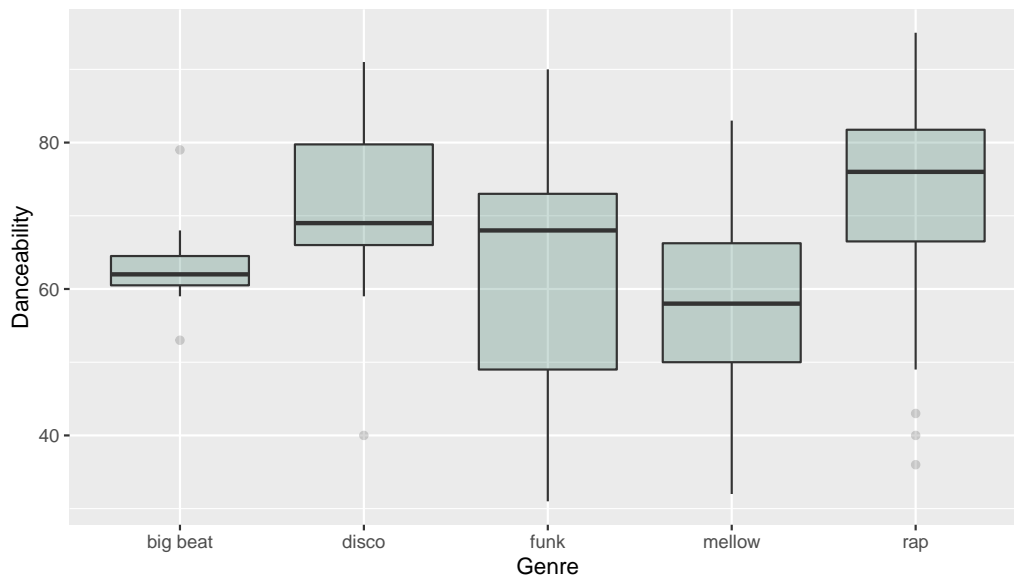
```
##          Genre freq
## 19         rock  857
## 17         pop  398
## 1  adult standards 123
## 15         other 121
## 14         metal  93
## 12         indie  79
## 7          dance  73
## 10  dutch cabaret  51
## 20         soul  45
## 16  permanent wave  38
## 4  british invasion  36
## 18          rap   30
## 13         mellow  28
## 3          blues  23
## 8          disco  18
## 11         funk   13
## 6         country  12
## 9  dutch americana  12
## 2          big beat  11
## 5  carnaval limburg  10
```

Spośród nich wybierzemy pięć o najwyższej średniej taneczności:

```
##      rap    disco big beat    funk    mellow
## 72.73333 70.50000 63.27273 62.00000 57.28571
```

Pięć o najwyższej średniej pozytywności:

```
##          disco          rap          blues          country
##      74.11111      64.43333      62.39130      58.75000
## british invasion
##      57.58333
```



Rysunek 3.6. Wykresy pudełkowe dla gatunków o najwyższej średniej taneczności

Pięć o najwyższej średniej energiczności:

```
## big beat    metal    disco    dance    rap
## 83.63636 74.70968 70.88889 70.06849 65.96667
```

Pięć o najwyższej średniej popularności:

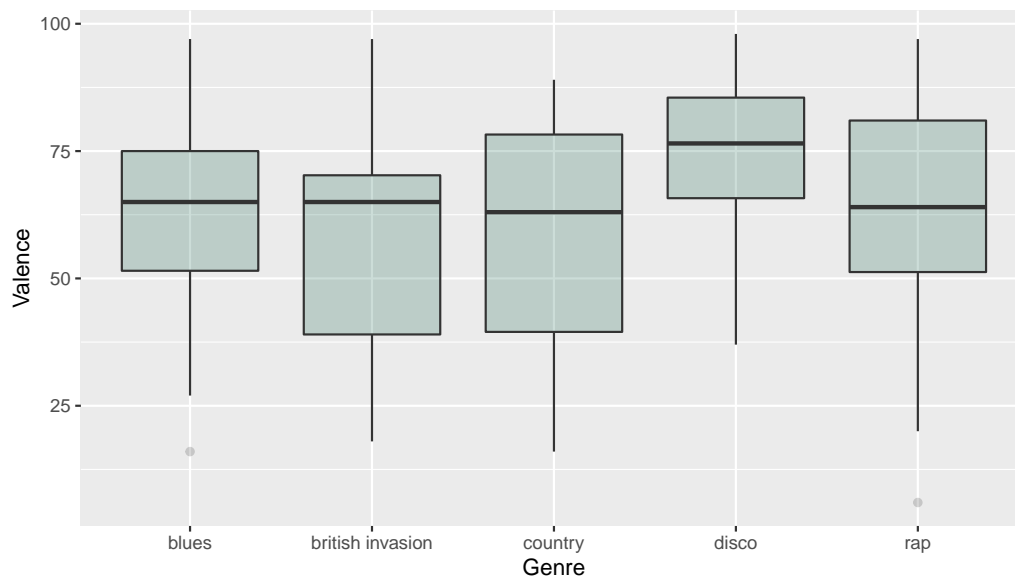
```
##          mellow british invasion permanent wave      soul
##          68.75000          68.08333          67.65789      67.37778
##          metal
##          66.03226
```

Na podstawie wykresów 3.6 możemy stwierdzić, że dla gatunków o najwyższej średniej taneczności najmniejszy rozstęp międzykwartyłowy mają gatunki big beat i disco. Oznacza to, że nie ma dużego rozrzutu wśród piosenek, do większości z nich dobrze się tańczy. Dlatego to właśnie te gatunki najbardziej nadają się na zabawę taneczną. Można również powiedzieć, że do większej ilości utworów disco będzie się tańczyć lepiej niż do big beat'u.

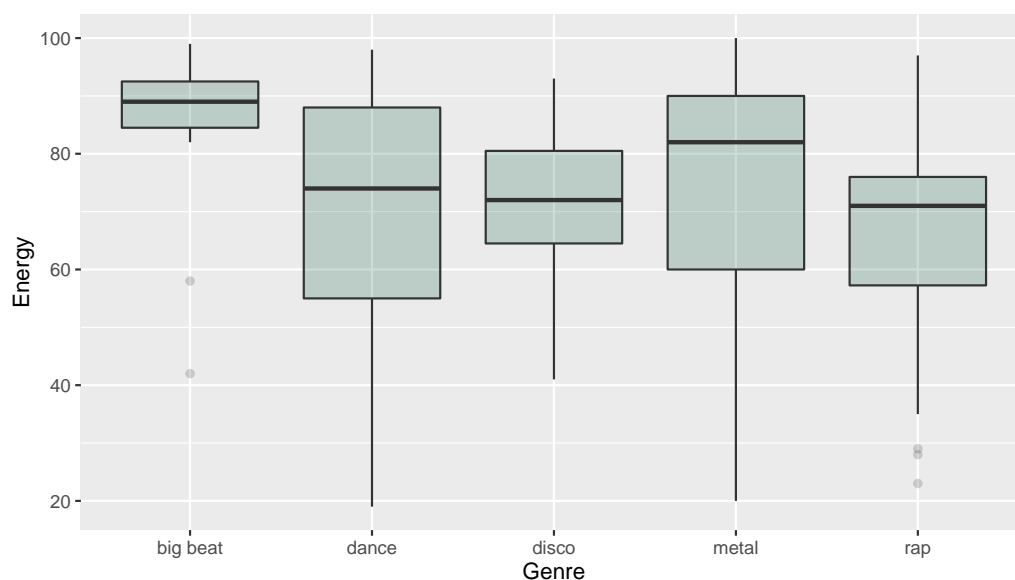
W przypadku wykresów pudełkowych 3.7 wszystkie gatunki mają duży rozrzut, jednak disco ma największą medianę i najmniejszy rozstęp międzykwartyłowy, więc można powiedzieć, że najbardziej się nadaje pod względem pozytywności muzycznej.

Patrząc na wykresy 3.8 widzimy, że piosenki z gatunków big beat mają najwyższą medianę, najwyższą średnią i najmniejszy rozstęp międzykwartyłowy, możemy zatem uznać ten gatunek za najbardziej energiczny.

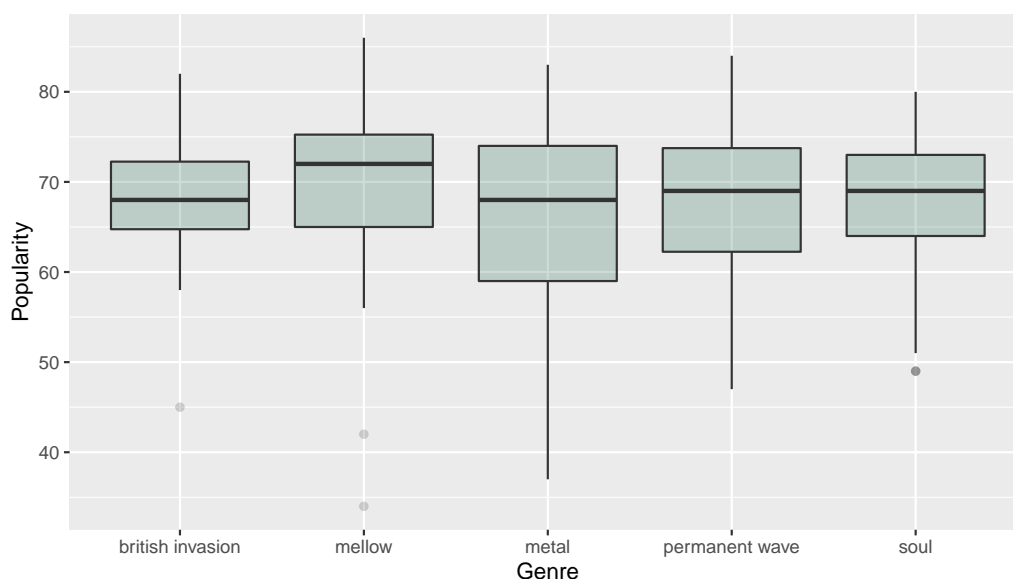
Na koniec rozpatrzmy popularność. W tym wypadku wszystkie gatunki mają zbliżone mediany oraz wielkości pudełek, przez co na podstawie wykresu 3.9 ciężko stwierdzić, który jest najpopularniejszy.



Rysunek 3.7. Wykresy pudełkowe dla gatunków o najwyższej średniej pozytywności



Rysunek 3.8. Wykresy pudełkowe dla gatunków o najwyższej średniej energiczności



Rysunek 3.9. Wykresy pudełkowe dla gatunków o najwyższej średniej popularności

Odpowiadając na zadane przez nas pytanie, zawężamy wybór do dwóch gatunków, a mianowicie disco i big beat. W zleżności czy zależy nam bardziej na pozytywności, czy energiczności, powinniśmy odpowiednio wybrać disco lub big beat.

3.3. Z jakich lat utwory wybrać by była największa szansa, że ludziom się spodoba?

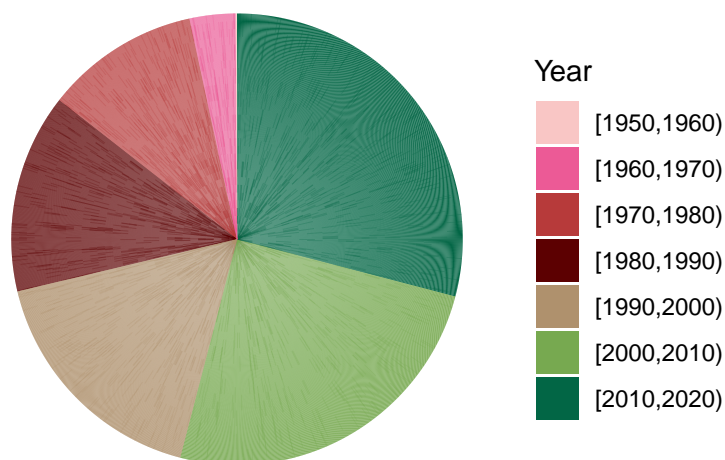
Ponieważ połączyliśmy lata w dekady, będziemy szukać dekady właśnie, z której piosenki były najbardziej nadające się do tańca, radosne, energiczne albo najpopularniejsze. Na początek sprawdźmy jak wygląda procentowa ilość utworów w danej dekadzie na wykresie kołowym.

Z wykresu kołowego 3.10 możemy odczytać, że próbek w pierwszej dekadzie jest bardzo mało, dlatego nie będziemy liczyć średnich z tego okresu (mogły by być zaniżone lub zawyżone przez pojedyncze piosenki). Dla pozostałych okresów zrobimy porównanie średnich, median oraz wariancji dla poszczególnych charakterystyk.

	[1960,1970)	[1970,1980)	[1980,1990)	[1990,2000)	[2000,2010)	[2010,2020)
Średnia	48.22	51.57	56.98	51.47	53.41	54.82
Mediana	48.00	51.00	58.00	52.00	53.00	55.00
Wariancja	179.47	240.19	234.20	253.24	238.82	209.44

Tabela 3.1. Średnie, mediany i wariancje taneczności utworów dla dekad

Analizując tabele 3.1 jak i 3.2 możemy stwierdzić, że wśród utworów z "Top 2000" średnio najwięcej piosenek do których się najlepiej tańczy jak i zarazem najbardziej pozytywnych/radosnych znajdziemy w latach 80.



Rysunek 3.10. Wykres kołowy dla dekad

	[1960,1970)	[1970,1980)	[1980,1990)	[1990,2000)	[2000,2010)	[2010,2020)
Średnia	52.90	54.25	57.12	42.91	46.23	45.17
Mediana	51.50	52.00	59.00	37.00	42.50	42.00
Wariancja	481.86	608.45	641.67	595.05	601.06	547.56

Tabela 3.2. Średnie, mediany i wariancje pozytywności utworów dla dekad

Stwierdzamy to na podstawie tego, że mediany jak i zarówno średnie są najwyższe, a wariancje dla wszystkich dekad mają podobną wartość.

Lata 2000 – 2010 charakteryzują się największym odsetkiem energicznych utworów, co możemy zobaczyć w tabeli 3.3. Wariancje ponownie osiągają duże wartości, więc bierzemy pod uwagę jedynie mediany i średnie.

Z tabeli 3.4 dowiadujemy się natomiast, że średnio najpopularniejsze utwory zostały wydane w latach 60. Odpowiada im także najmniejszą wariancja.

Korzystając z faktu, że analizujemy dane, które reprezentują top najpopularniejszych utworów na spotify. Możemy dowiedzieć się z jakiej dekady utwory są najpopularniejsze nie tylko z tych analizowanych ale z całego spotify. Biorąc to pod uwagę, będą to lata z których najwięcej piosenek znalazło się w danych. Korzystając z wykresu 3.10, widzimy, że są to lata: 2010 – 2020.

Podsumowując powyższe, najlepszym wyborem z "Top 2000" będą utwory z lat 1980 – 1990. Kierując się energicznością utworów warto wybrać, te z lat 2000. Jeżeli bardziej zależy nam na utworach będących hitami, lepszą opcją będą lata 60 lub playlisty na spotify z piosenkami z lat 2010 – 2020.

	[1960,1970)	[1970,1980)	[1980,1990)	[1990,2000)	[2000,2010)	[2010,2020)
Średnia	49.32	56.25	61.11	59.86	64.36	61.08
Mediana	47.50	56.50	65.00	60.00	67.00	64.00
Wariancja	431.23	463.39	554.17	503.89	454.16	429.77

Tabela 3.3. Średnie, mediany i wariancje energiczności utworów dla dekad

	[1960,1970)	[1970,1980)	[1980,1990)	[1990,2000)	[2000,2010)	[2010,2020)
Średnia	64.25	62.08	59.41	59.47	57.98	56.94
Mediana	67.00	64.00	60.00	61.00	58.00	59.00
Wariancja	116.42	129.40	162.29	168.95	193.74	371.51

Tabela 3.4. Średnie, mediany i wariancje popularności utworów dla dekad

3.4. Jeżeli nie chcemy szukać piosenek pojedynczo, muzykę jakiego wykonawcy puścić?

W tej sekcji skupimy się na najpopularniejszych artystach, czyli tych których utworów jest w danych najwięcej (bo dane to "Top 2000"). W końcu jest największa szansa, że to oni spodobają się większości. Oprócz tego, gdybyśmy bazowali na popularności utworów i robili średnią dla każdego artysty, mogliby przodować wykonawcy z jedną znaną piosenką, którzy nie mają więcej popularnych utworów.

Dlatego spośród wykonawców, do dalszej analizy wybierzemy tych, którzy mają co najmniej 10 utworów na playliście. To znaczy, że z 731 artystów zostanie nam 32. Policzmy dla nich średnią taneczność, pozytywność, energiczność i popularność. Pięć najwyższych wyników, dla każdej z charakterystyk zobaczyć można poniżej.

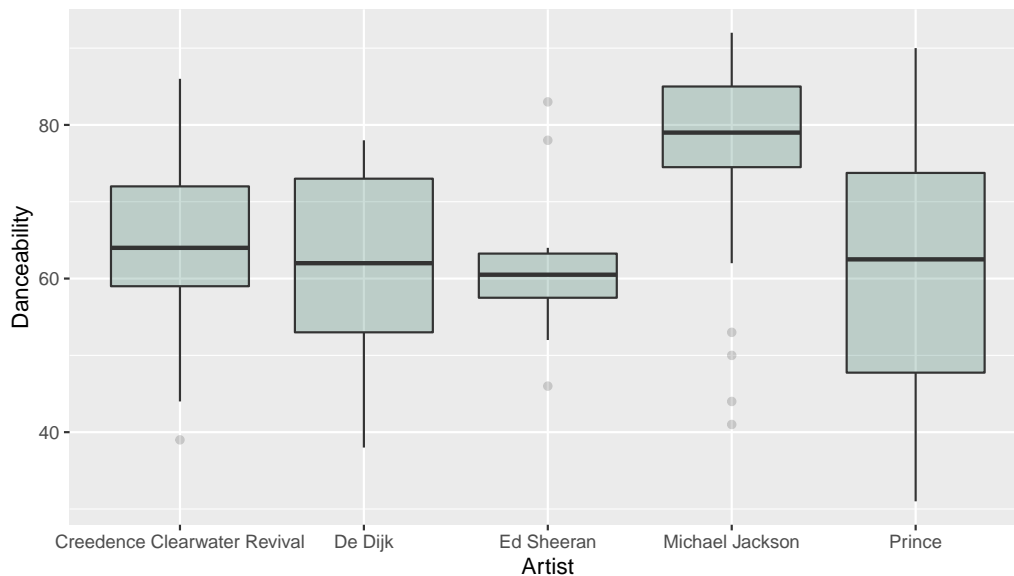
```
data_popular <- data[data$Artist %in% popular_artist,]
```

Taneczność:

```
##           Michael Jackson Creedence Clearwater Revival
##           74.73913                                63.53846
##           Ed Sheeran                                Prince
##           61.66667                                61.50000
##           De Dijk
##           61.38462
```

Pozytywność:

```
## Creedence Clearwater Revival                                ABBA
##           74.76923                                67.72727
##           Bee Gees                                Michael Jackson
##           63.08333                                62.30435
##           The Rolling Stones
##           62.16667
```

Rysunek 3.11. Wykresy pudełkowe dla artystów o najwyższej średniej taneczności

Energiczność:

##	Metallica	Red Hot Chili Peppers	Muse
##	78.18182	77.36364	74.93333
##	Golden Earring	Nirvana	
##	70.83333	70.72727	

Popularność:

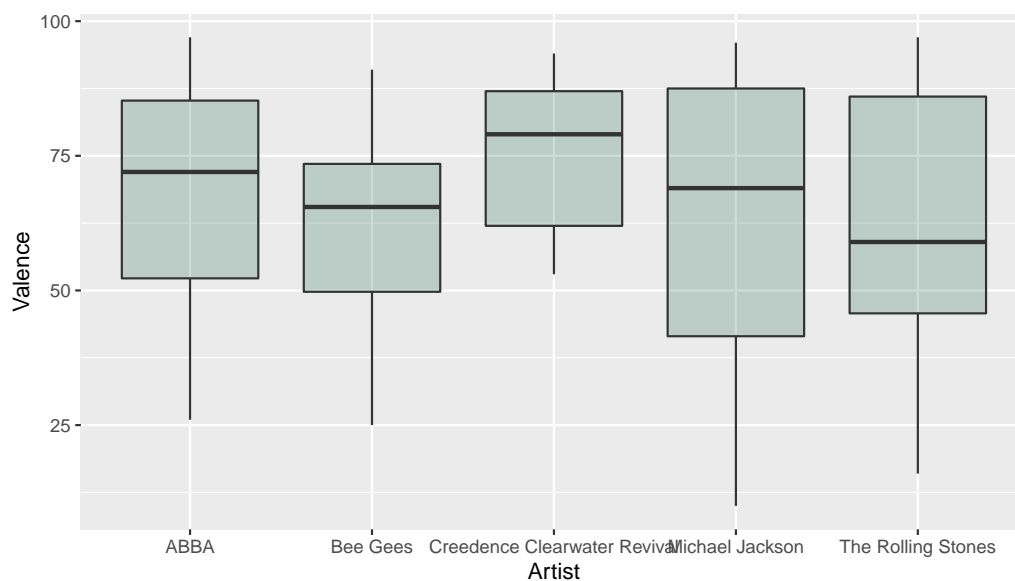
##	Ed Sheeran	Red Hot Chili Peppers	Coldplay
##	78.83333	76.00000	69.92593
##	Adele	Elton John	
##	68.57143	68.36364	

Nie jest to zaskoczeniem, że w kategorii najlepszych piosenek do tańca przoduje Michael Jackson. Widać, że pudełko wykresu (patrz wykres 3.11) jest ponad pozostałymi. Wartym uwagi jest również Ed Sheeran, który z pozostałych wykonawców, mimo podobnej mediany ma najmniejszy rozstęp międzykwartyłowy.

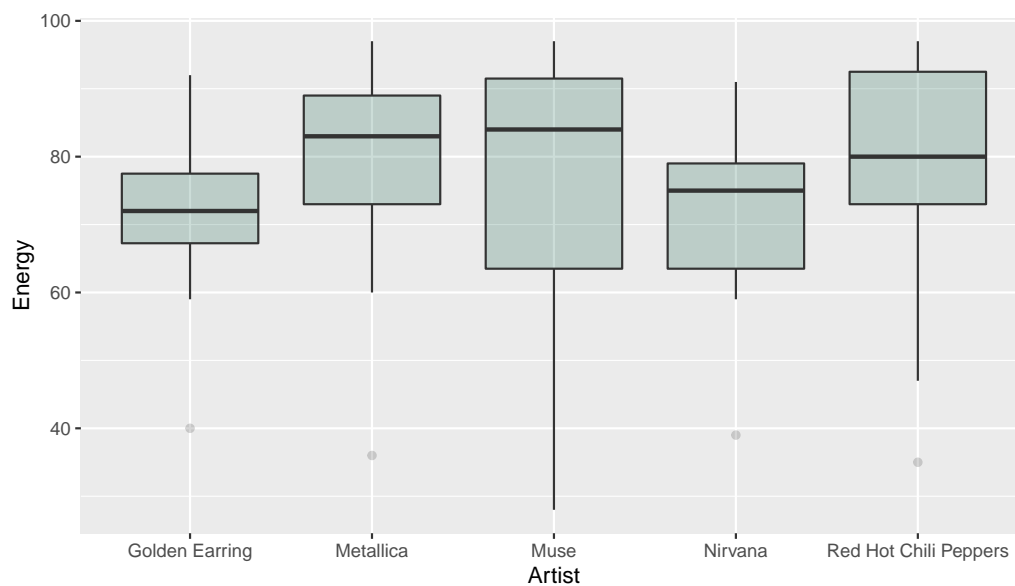
Po spojrzeniu na wykres 3.12 stwierdzamy, że najbardziej pozytywną muzykę ma wykonawca Creedence Clearwater Revival. Ma najkrótsze wąsy i najwyższą medianę.

Do najbardziej energicznych wykonawców, widocznych na wykresie 3.13 możemy zaliczyć Metallicę, Golden Earring oraz Nirvanę. Spośród pięciu artystów o najwyższej średniej energiczności mają oni najmniejszy rozrzut.

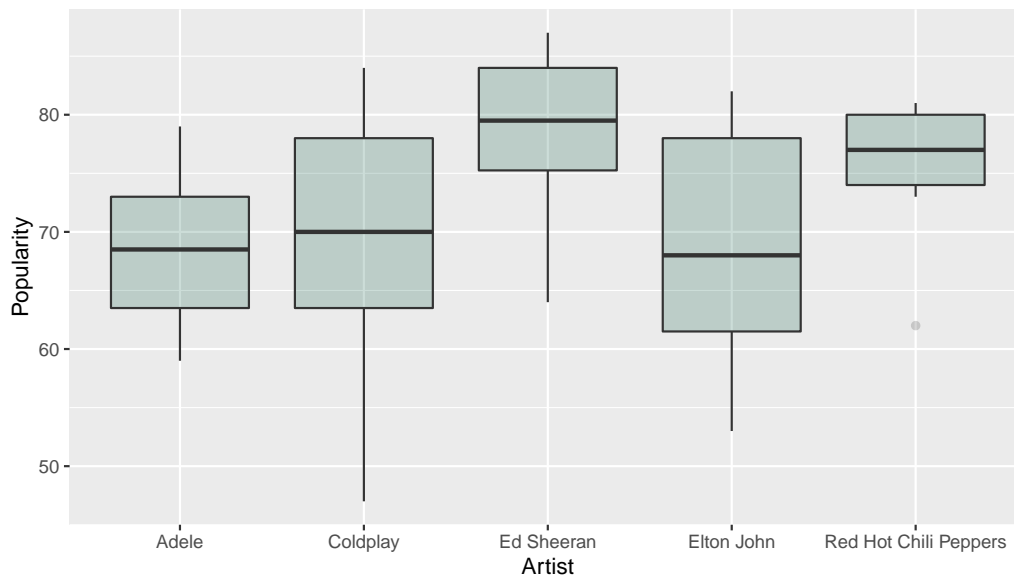
Na podstawie wykresu 3.14, jako najpopularniejszych możemy uznać Eda Sheerana i Red Hot Chili Peppers. Przy czym prawie wszystkie piosenki Red Hot Chili Peppers są popularne.



Rysunek 3.12. Wykresy pudełkowe dla artystów o najwyższej średniej pozytywności



Rysunek 3.13. Wykresy pudełkowe dla artystów o najwyższej średniej energiczności



Rysunek 3.14. Wykresy pudełkowe dla artystów o najwyższej średniej popularności

Możemy również, tak jak w przypadku dekad, określić najpopularniejszych artystów ogółem na spotify, a nie tylko w naszych danych. Korzystając z informacji, którzy wykonawcy mają najwięcej utworów. Zobaczmy 10 najpopularniejszych:

##	Artist	freq
## 495	Queen	37
## 612	The Beatles	36
## 126	Coldplay	27
## 697	U2	26
## 654	The Rolling Stones	24
## 91	Bruce Springsteen	23
## 415	Michael Jackson	23
## 9	ABBA	22
## 145	David Bowie	21
## 220	Fleetwood Mac	18

4. Podsumowanie

Podsumowując powyższą analizę, piosenek na imprezę powinniśmy szukać rozpatrując ich energiczność, popularność i to jak dobrze się do nich tańczy. Po przyjrzeniu się taneczności zauważyliśmy jej silny związek z pozytywnością muzyczną. Potwierdziło się to, gdy szukałyśmy zarówno najlepszych gatunków jak i dekad. Przy obu charakterystykach przodowało disco i lata 80. Najbardziej taneczne piosenki można też rozpoznać po ich tempie. Powinny mieć pomiędzy 100 a 150 uderzeń na minutę. Do arytsów, których piosenki charakteryzują się najwyższą tanecznością, zaliczamy Michaela Jacksona i Eda Sheerana. Eda Sheerana wybrałyśmy ze względu na to, że prawie wszystkie piosenki ma równie nadające się do tańca. W przypadku pozytywności wyróżniał się Creedence Clearwater Revival. Nie pokrywa się to z artystami o najlepszych do tańca utworach.

Piosenki na przyjęcie powinny być też energiczne. Wiąże się to z głośnymi utworami, które nie mogą być akustyczne. Najwięcej energicznych utworów powstało w latach 2000 – 2010. Gatunkiem najbardziej energicznym okazał się big beat, który nadaje się również do tańca. Jeżeli chodzi o wykonawców wśród najlepszych pojawili się Metallica, Golden Earring oraz Nirvana.

Wśród "Top 2000" średnio najbardziej popularna muzyka pochodzi z lat 60. Co ciekawe w tych latach powstała większość kawałków disco. Ponadto najwięcej popularnych kawałków znajdziemy u Eda Sheerana i Red Hot Chili Peppers.

Najwięcej utworów zostało wydanych w latach 2010 – 2020, a biorąc pod uwagę wykonawców największą ilość piosenek przypisujemy zespołom: Queen, The Beatles i Coldplay. Korzystając z faktu, że mamy top utwory ze Spotify, możemy powiedzieć, że lata 2010 – 2020 i wyżej wymienieni artyści są ogółem najpopularniejsi na tej platformie.