

analiza wstepna

Krzysztof Kurek

20 02 2020

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(reshape2)  
library(zoo)
```

```
##  
## Attaching package: 'zoo'  
  
## The following objects are masked from 'package:base':  
##  
##   as.Date, as.Date.numeric
```

```
library(lubridate)
```

```
##  
## Attaching package: 'lubridate'  
  
## The following object is masked from 'package:base':  
##  
##   date
```

```
library(ggplot2)
```

```
dane <- read.csv("../train_set.csv", stringsAsFactors = FALSE, na.strings="")
```

```
#Wstepna analiza danych
```

```
head(dane)
```

```
##          X          track_id      track_name      track_artist
## 1 20378 7uH27oIt4a6cIFCA8ZPcyG Baila Baila Baila          Ozuna
## 2  3549 6GunRGvhXUyvREwRnnKbDC      Last Dance          Covenant
## 3 28314 4IyVra117kbef0wCEkKpb2      Activate          Sandro Silva
## 4   149 0lYBSQXN6rCTvUZvg9S0lU  Let Me Love You          DJ Snake
## 5  1931 27L8sESb3KR79asDUBu8nW      Stacy's Mom Fountains Of Wayne
## 6  5541 7tgjDlQsMR8Rvkd1NCJA58      6 Kiss          Trippie Redd
## track_popularity      track_album_id      track_album_name
## 1          65 6xQurnsfSvF3hQwU5EIqB8      Baila Baila Baila
## 2           0 0UHnC3KThaYwuDoDh4HRzB      Last Dance
## 3          31 21l4YlhJ5FYFvGDeaW9jrZ      Activate
## 4          81 02sEJTj1syelJaqxqpcSCp      Encore
## 5          73 6TZp52tXShLQbq8yNMxqNT Welcome Interstate Managers
## 6          83 0tKX7BLXiiRgXUKYdJzjEz      A Love Letter To You 4
## track_album_release_date
## 1      2019-01-05
## 2      2013-06-07
## 3      2019-12-04
## 4      2016-08-05
## 5      2003-01-01
## 6      2019-11-22
##
##                                playlist_name
## 1                                Latest Latin American Hits 2020
## 2 Gothic / Industrial / Mittelalter / EBM / Futurepop / Gothik / Electropop
## 3                                Big Room EDM
## 4                                Dance Pop
## 5                                Dr. Q's Prescription Playlistdž'SŠ
## 6                                RapCaviar
## playlist_id playlist_genre playlist_subgenre danceability energy
## 1 51LvliKED7oZNJR8XCiPip      latin      latin hip hop      0.816 0.754
## 2 53CmFroG6MWR5re00XJX6B      pop      electropop      0.611 0.988
## 3 3S03NwbWkrR1e7DPH9rf9Z      edm      big room      0.641 0.874
## 4 37i9dQZF1DWZQaaqNMbbXa      pop      dance pop      0.649 0.716
## 5 6jAPdgY9XmxC9cgkXAVmVv      pop      post-teen pop      0.774 0.750
## 6 37i9dQZF1DX0XUsuxWHRQd      rap      hip hop      0.687 0.449
## key loudness mode speechiness acousticness instrumentalness liveness valence
## 1 11 -2.750 0 0.1110 0.03610 0.00e+00 0.2170 0.533
## 2 10 -3.971 1 0.1020 0.00254 7.84e-01 0.2220 0.269
## 3 1 -4.991 1 0.0746 0.13600 2.44e-06 0.1990 0.140
## 4 8 -5.371 1 0.0349 0.08630 2.63e-05 0.1350 0.163
## 5 11 -4.927 0 0.0413 0.00210 1.85e-05 0.0808 0.925
## 6 2 -10.061 1 0.1540 0.10900 0.00e+00 0.2820 0.229
## tempo duration_ms
## 1 99.966 158400
## 2 136.015 378120
## 3 127.922 177187
## 4 99.988 205947
## 5 118.015 197987
## 6 142.048 199437
```

```
tail(dane)
```

```
##          X          track_id
## 26261 23727 4b8hQaPTM7Fhp00AkKBj5A
## 26262 20376 6o8ogTrSX8HOF66Pd3Tsc
## 26263 15792 2G7SBK5Tr4H1N6L3N3gGzp
## 26264 13877 2rxJabc4yx6mAIsTmxg1Gv
## 26265  2084 4bb2UdCvi9LynD4rWvE06T
## 26266 30895 4Uem7GbHwFucQQUS2f01dU
##
##          track_name      track_artist
## 26261          Brioschi      Franco126
## 26262      Baby Girl (feat. Lalo Ebratt) Mario Bautista
## 26263          Kano Kalokeria          Kim
## 26264      Trampled Rose      Tom Waits
## 26265      Turn Up the Music      Adam Hicks
## 26266 Collide (feat. Collin McLoughlin) - Radio Edit Laidback Luke
##      track_popularity      track_album_id      track_album_name
## 26261          61 5f2LuFGKudyDDp9d93Qvvc      Stanza Singola
## 26262          75 0B2NMC13p92JyFvXudy69Q Baby Girl (feat. Lalo Ebratt)
## 26263          21 1MEtRPBLvZoAiJSQo6gzxt      Kano Kalokeria
## 26264          34 6F17Mb1oTeRoXs7JI3lTHy      Real Gone (Remastered)
## 26265          56 6Vsm16sziVKBHxbMEp5wgQ      Lemonade Mouth
## 26266          27 4jg5WP3YK3DAYKYOmJr4Px      Collide
##      track_album_release_date
## 26261      2019-01-25
## 26262      2018-08-17
## 26263      2016-05-30
## 26264      2004-10-05
## 26265      2011-01-01
## 26266      2014-01-27
##
##          playlist_name
## 26261 Musica Italiana 2020 - Playlist Pop & Hip-Hop (Canzoni Italiane 2020)
## 26262      Latest Latin American Hits 2020
## 26263      Rock Hubspot
## 26264      Permanent Wave
## 26265      post teen pop
## 26266      Happy EDM
##      playlist_id playlist_genre playlist_subgenre danceability
## 26261 6kVFIQBhLT4003iw2WWEv1      r&b      hip pop      0.768
## 26262 51LvliKED7oZNR8XCiPip      latin      latin hip hop      0.804
## 26263 5hvf4yL105Rw6cvCP6AMjM      rock      hard rock      0.711
## 26264 5Go0Jsxj1UnsU70m841BEo      rock      permanent wave      0.379
## 26265 2l1tt5ouiE0301o2jbxn30d      pop      post-teen pop      0.531
## 26266 3d2JFEbvcx7p2CHkBaQeMZ      edm      pop edm      0.443
##      energy key loudness mode speechiness acousticness instrumentalness
## 26261 0.562  5  -7.790  0  0.2150      0.4720      0.00e+00
## 26262 0.705  1  -5.272  1  0.1990      0.0275      1.38e-04
## 26263 0.747  7  -5.115  0  0.0400      0.2720      1.35e-03
## 26264 0.396 10  -9.274  0  0.0276      0.8430      8.97e-04
## 26265 0.827  7  -3.700  1  0.0810      0.0957      1.84e-06
## 26266 0.948  4  -3.290  1  0.1080      0.0124      3.99e-06
##      liveness valence      tempo duration_ms
## 26261  0.102  0.737  92.013      219293
```

```
## 26262    0.202    0.450 103.009    218470
## 26263    0.105    0.620  96.979    191862
## 26264    0.121    0.450 130.709    237453
## 26265    0.074    0.801 183.977    176600
## 26266    0.131    0.341 127.928    200435
```

```
summary(dane)
```

```
##          X          track_id      track_name      track_artist
## Min.   :    0      Length:26266      Length:26266      Length:26266
## 1st Qu.: 8184      Class :character      Class :character      Class :character
## Median :16440      Mode  :character      Mode  :character      Mode  :character
## Mean   :16421
## 3rd Qu.:24634
## Max.   :32832
## track_popularity track_album_id      track_album_name
## Min.   :  0.00      Length:26266      Length:26266
## 1st Qu.: 24.00      Class :character      Class :character
## Median : 45.00      Mode  :character      Mode  :character
## Mean   : 42.39
## 3rd Qu.: 62.00
## Max.   :100.00
## track_album_release_date playlist_name      playlist_id
## Length:26266      Length:26266      Length:26266
## Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character
##
##
## playlist_genre      playlist_subgenre      danceability      energy
## Length:26266      Length:26266      Min.   :0.0000      Min.   :0.0118
## Class :character      Class :character      1st Qu.:0.5630      1st Qu.:0.5800
## Mode  :character      Mode  :character      Median :0.6720      Median :0.7210
##                                     Mean   :0.6551      Mean   :0.6981
##                                     3rd Qu.:0.7610      3rd Qu.:0.8400
##                                     Max.   :0.9830      Max.   :1.0000
##
##          key          loudness          mode          speechiness
## Min.   : 0.000      Min.   : -46.448      Min.   :0.000      Min.   :0.0000
## 1st Qu.: 2.000      1st Qu.: -8.191      1st Qu.:0.000      1st Qu.:0.0410
## Median : 6.000      Median : -6.164      Median :1.000      Median :0.0625
## Mean   : 5.395      Mean   : -6.725      Mean   :0.565      Mean   :0.1068
## 3rd Qu.: 9.000      3rd Qu.: -4.644      3rd Qu.:1.000      3rd Qu.:0.1320
## Max.   :11.000      Max.   :  1.275      Max.   :1.000      Max.   :0.8690
## acousticness      instrumentalness      liveness      valence
## Min.   :0.0000      Min.   :0.000000      Min.   :0.0000      Min.   :0.0000
## 1st Qu.:0.0150      1st Qu.:0.000000      1st Qu.:0.0926      1st Qu.:0.3300
## Median :0.0798      Median :0.000017      Median :0.1270      Median :0.5110
## Mean   :0.1755      Mean   :0.085842      Mean   :0.1893      Mean   :0.5099
## 3rd Qu.:0.2550      3rd Qu.:0.005020      3rd Qu.:0.2470      3rd Qu.:0.6920
## Max.   :0.9940      Max.   :0.994000      Max.   :0.9960      Max.   :0.9910
## tempo      duration_ms
## Min.   :  0.00      Min.   :  4000
## 1st Qu.: 99.95      1st Qu.:187805
## Median :121.98      Median :216050
```

```
## Mean      :120.85    Mean      :225836
## 3rd Qu.   :133.90    3rd Qu.   :253733
## Max.      :220.25    Max.       :517810
```

```
glimpse(dane)
```

```
## Observations: 26,266
## Variables: 24
## $ X                <int> 20378, 3549, 28314, 149, 1931, 5541, 19225...
## $ track_id         <chr> "7uH27oIt4a6cIFCA8ZPcyG", "6GunRGvhXUyvREw...
## $ track_name       <chr> "Baila Baila Baila", "Last Dance", "Activa...
## $ track_artist     <chr> "Ozuna", "Covenant", "Sandro Silva", "DJ S...
## $ track_popularity <int> 65, 0, 31, 81, 73, 83, 73, 59, 47, 51, 74,...
## $ track_album_id   <chr> "6xQurnsfSvF3hQwU5EIqB8", "0UHnC3KThaYwuDo...
## $ track_album_name <chr> "Baila Baila Baila", "Last Dance", "Activa...
## $ track_album_release_date <chr> "2019-01-05", "2013-06-07", "2019-12-04", ...
## $ playlist_name    <chr> "Latest Latin American Hits 2020", "Gothic...
## $ playlist_id      <chr> "51LvliKED7oZNR8XCiPip", "53CmFroG6MWR5re...
## $ playlist_genre   <chr> "latin", "pop", "edm", "pop", "pop", "rap"...
## $ playlist_subgenre <chr> "latin hip hop", "electropop", "big room",...
## $ danceability     <dbl> 0.816, 0.611, 0.641, 0.649, 0.774, 0.687, ...
## $ energy           <dbl> 0.754, 0.988, 0.874, 0.716, 0.750, 0.449, ...
## $ key              <int> 11, 10, 1, 8, 11, 2, 11, 6, 0, 3, 7, 2, 2,...
## $ loudness         <dbl> -2.750, -3.971, -4.991, -5.371, -4.927, -1...
## $ mode             <int> 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, ...
## $ speechiness      <dbl> 0.1110, 0.1020, 0.0746, 0.0349, 0.0413, 0....
## $ acousticness     <dbl> 0.036100, 0.002540, 0.136000, 0.086300, 0....
## $ instrumentalness <dbl> 0.00e+00, 7.84e-01, 2.44e-06, 2.63e-05, 1....
## $ liveness         <dbl> 0.2170, 0.2220, 0.1990, 0.1350, 0.0808, 0....
## $ valence          <dbl> 0.533, 0.269, 0.140, 0.163, 0.925, 0.229, ...
## $ tempo            <dbl> 99.966, 136.015, 127.922, 99.988, 118.015,...
## $ duration_ms      <int> 158400, 378120, 177187, 205947, 197987, 19...
```

```
dim(dane)
```

```
## [1] 26266    24
```

Weryfikacja wartości brakujących

```
na_dane <- sapply(dane, function(y)sum(length(which(is.na(y)))))
na_dane <- as.data.frame(na_dane); na_dane
```

```
##                na_dane
## X                0
## track_id         0
## track_name       4
## track_artist     4
## track_popularity 0
## track_album_id   0
```

```
## track_album_name      4
## track_album_release_date 0
## playlist_name         0
## playlist_id           0
## playlist_genre        0
## playlist_subgenre     0
## danceability          0
## energy                0
## key                   0
## loudness              0
## mode                  0
## speechiness           0
## acousticness          0
## instrumentalness      0
## liveness              0
## valence               0
## tempo                 0
## duration_ms           0
```

```
dane[is.na(dane$track_name),]
```

```
##           X               track_id track_name track_artist track_popularity
## 150      9283 5TTzhRSWQS4Yu8xTgAuq6D      <NA>      <NA>              0
## 15546 19811 69gRFGOWY90MpFJgFol1u0      <NA>      <NA>              0
## 19676  8151 69gRFGOWY90MpFJgFol1u0      <NA>      <NA>              0
## 25781 19568 3VKFip30dAvv40fNTgFWeQ      <NA>      <NA>              0
##           track_album_id track_album_name track_album_release_date
## 150      3luHJEPw434tvNbme3SP8M      <NA>      2017-12-01
## 15546 717UG2du6utFe7CdmpuUe3      <NA>      2012-01-05
## 19676 717UG2du6utFe7CdmpuUe3      <NA>      2012-01-05
## 25781 717UG2du6utFe7CdmpuUe3      <NA>      2012-01-05
##           playlist_name           playlist_id playlist_genre
## 150      GANGSTA Rap 5GA8GDo7RQC3JEanT81B3g      rap
## 15546      latin hip hop 3nH8aytdqNeRbcRCg3dw9q      latin
## 19676      HIP&HOP 5DyJsJZ0pMJh34WvUrQzMV      rap
## 25781 Reggaeton viejitođŹ"Ą Osi5tw70PIgPkY1Eva6V8f      latin
##           playlist_subgenre danceability energy key loudness mode speechiness
## 150      gangster rap      0.465 0.820 10 -5.907 0 0.3070
## 15546      latin hip hop      0.714 0.821 6 -7.635 1 0.1760
## 19676      southern hip hop      0.714 0.821 6 -7.635 1 0.1760
## 25781      reggaeton      0.675 0.919 11 -6.075 0 0.0366
##           acousticness instrumentalness liveness valence tempo duration_ms
## 150      0.0963      0.00000 0.0888 0.505 86.839 206465
## 15546      0.0410      0.00000 0.1160 0.649 95.999 282707
## 19676      0.0410      0.00000 0.1160 0.649 95.999 282707
## 25781      0.0606      0.00653 0.1030 0.726 97.017 252773
```

Weryfikacja gatunków muzycznych

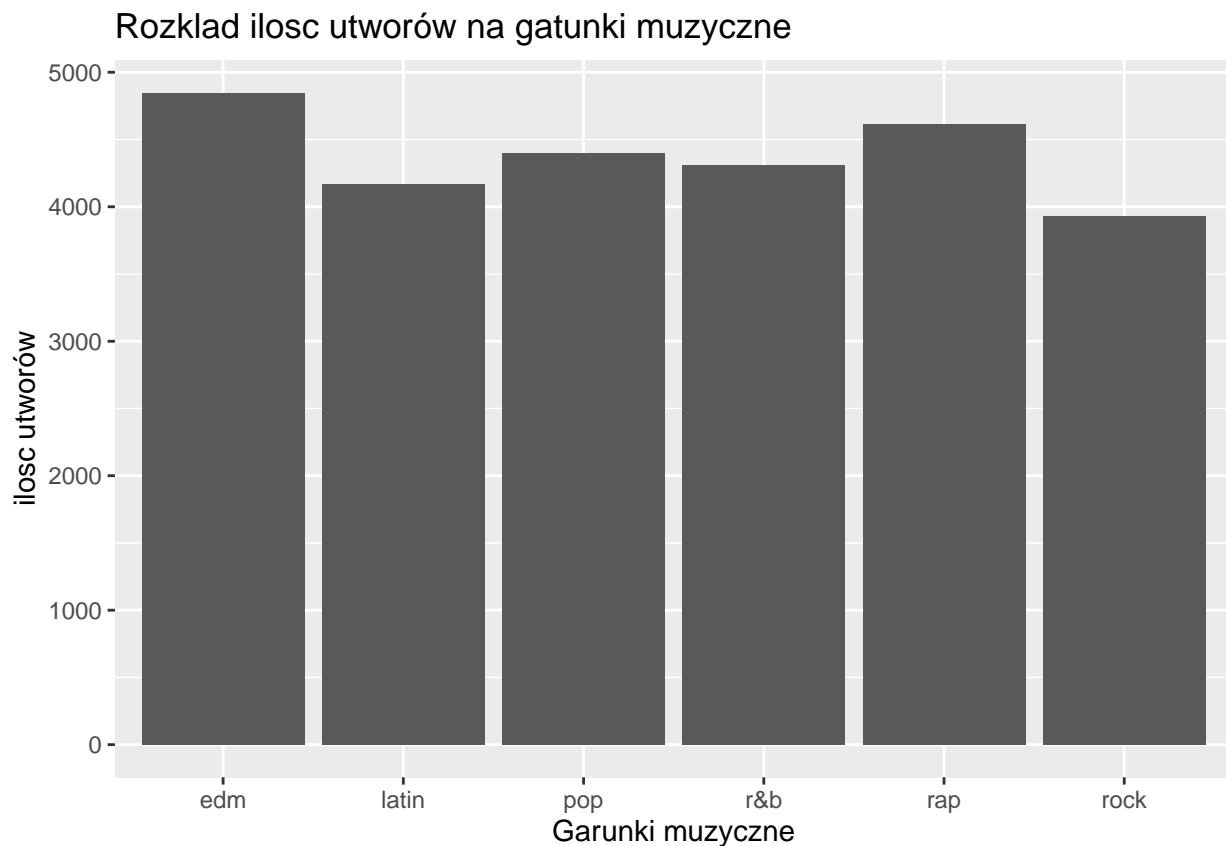
```
unique(dane$playlist_genre)
```

```
## [1] "latin" "pop"    "edm"    "rap"    "r&b"    "rock"
```

```
dane%>%  
  group_by(playlist_genre) %>%  
  summarise(ilosc_utworow = n())
```

```
## # A tibble: 6 x 2  
##   playlist_genre ilosc_utworow  
##   <chr>          <int>  
## 1 edm            4848  
## 2 latin          4165  
## 3 pop            4400  
## 4 r&b            4306  
## 5 rap            4616  
## 6 rock           3931
```

```
dane%>%  
  group_by(playlist_genre) %>%  
  summarise(ilosc_utworow = n()) %>%  
  arrange(desc(ilosc_utworow)) %>%  
  ggplot() + geom_col(aes(x = playlist_genre, y = ilosc_utworow)) +  
  labs(title = "Rozkład ilość utworów na gatunki muzyczne",  
        x = "Garunki muzyczne",  
        y = "ilość utworów")
```



```
# Widzimy, że naliczniejszą grupą utworów jest grupa edm, następnie rap i pop
```

```
dane$track_album_release_date <- ymd(dane$track_album_release_date)
```

```
## Warning: 1498 failed to parse.
```

```
# 1498 rekordów nie posiada odpowiedniego formatu daty
```

```
dane$track_album_release_date <- as.Date(as.yearmon(dane$track_album_release_date, "%Y%m"))
```

```
dane %>%
```

```
  group_by(track_album_release_date, playlist_genre) %>%
```

```
  filter(track_album_release_date >= "1963-01-01") %>%
```

```
  summarise(ilosc_utworow = n()) %>%
```

```
  ggplot(aes(x = track_album_release_date, y = ilosc_utworow, color = playlist_genre, fill = playlist_g
```

```
  scale_x_date(date_breaks = "10 year") +
```

```
  labs(title = "Rozkład dodawanych utworów na przestrzeni lat z podziałem na gatunki",
```

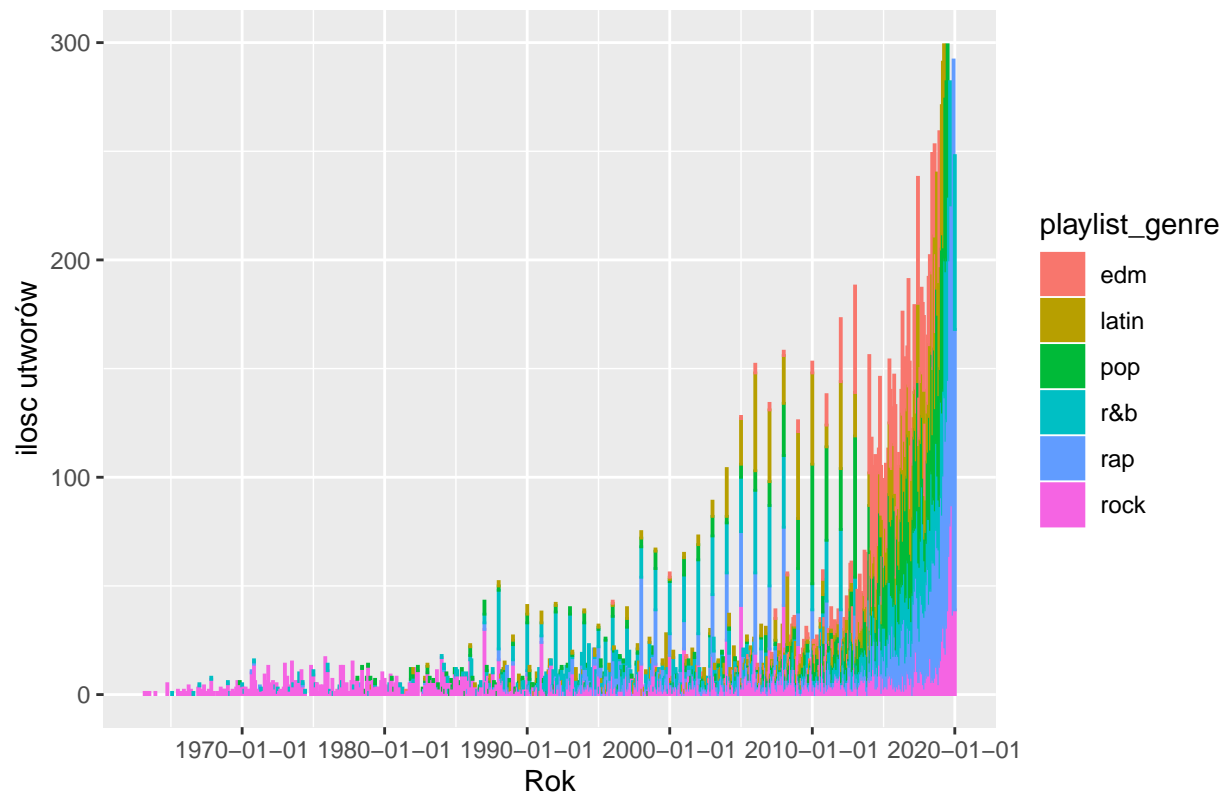
```
        x = "Rok",
```

```
        y = "ilość utworów") +
```

```
  ylim(0,300)
```

```
## Warning: Removed 34 rows containing missing values (geom_col).
```

Rozkład dodawanych utworów na przestrzeni lat z podziałem na gatunki

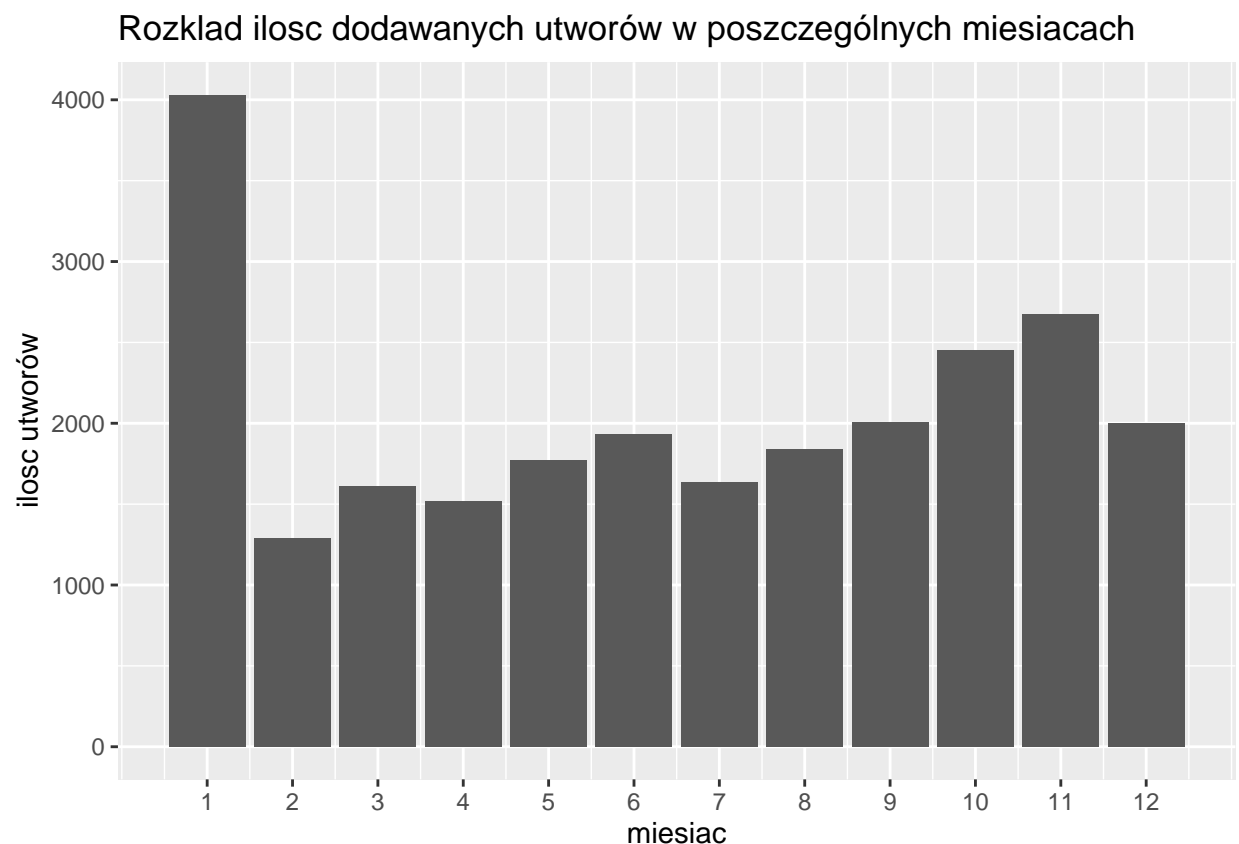



```
# Do roku 1990 najczęściej dodawanym gatunkiem muzycznym był rock i , później zaczęto dodawać rap, laty
```

Ilość dodawanych utworów w poszczególnych miesiącach

```
dane$miesiac <- month(dane$track_album_release_date)

dane%>%
  group_by(miesiac) %>%
  summarise(ilosc_utworow = n()) %>%
  filter(!is.na(miesiac)) %>%
  arrange(desc(ilosc_utworow)) %>%
  ggplot() + geom_col(aes(x = miesiac, y = ilosc_utworow)) +
  labs(title = "Rozkład ilość dodawanych utworów w poszczególnych miesiącach",
       x = "miesiąc",
       y = "ilość utworów")+
  scale_x_continuous(breaks = seq(1,12,1))
```



```
# Tabela dodawanych utworów na przestrzeni lat z podziałem na gatunek muzyczny
```

```
dcast(dane, track_album_release_date ~ playlist_genre )
```

```
## Using miesiac as value column: use value.var to override.
```

Aggregation function missing: defaulting to length

##	track_album_release_date	edm	latin	pop	r&b	rap	rock
## 1	1957-01-01	0	0	0	1	0	0
## 2	1958-03-01	0	0	0	0	0	1
## 3	1963-03-01	0	0	0	0	0	1
## 4	1963-05-01	0	0	0	0	0	1
## 5	1963-07-01	0	0	0	0	0	1
## 6	1963-12-01	0	0	0	0	0	1
## 7	1964-10-01	0	0	0	0	0	5
## 8	1965-02-01	0	0	0	1	0	0
## 9	1965-07-01	0	0	0	0	0	2
## 10	1965-08-01	0	0	0	0	0	1
## 11	1965-11-01	0	0	0	0	0	1
## 12	1965-12-01	0	0	0	0	0	2
## 13	1966-01-01	0	0	0	0	0	1
## 14	1966-04-01	0	0	0	0	0	4
## 15	1966-06-01	0	0	0	0	0	2
## 16	1966-08-01	0	0	0	1	0	0
## 17	1966-12-01	0	0	0	0	0	3
## 18	1967-01-01	0	0	0	1	0	4
## 19	1967-03-01	0	0	0	0	0	2
## 20	1967-05-01	0	0	0	0	0	4
## 21	1967-06-01	0	0	0	0	0	1
## 22	1967-08-01	0	0	0	0	0	1
## 23	1967-09-01	0	0	0	0	0	2
## 24	1967-10-01	0	0	0	0	0	1
## 25	1967-11-01	0	0	0	1	0	7
## 26	1968-01-01	0	0	0	0	0	1
## 27	1968-04-01	0	0	0	0	0	1
## 28	1968-06-01	0	0	0	0	0	1
## 29	1968-07-01	0	0	0	0	0	3
## 30	1968-08-01	0	0	0	0	0	2
## 31	1968-10-01	0	0	0	0	0	2
## 32	1968-11-01	0	0	0	0	0	2
## 33	1968-12-01	0	0	0	0	0	1
## 34	1969-01-01	0	0	0	1	0	5
## 35	1969-03-01	0	0	0	1	0	0
## 36	1969-04-01	0	0	0	0	0	1
## 37	1969-05-01	0	0	0	0	0	2
## 38	1969-07-01	0	0	0	0	0	1
## 39	1969-08-01	0	0	0	0	0	3
## 40	1969-09-01	0	0	0	0	0	6
## 41	1969-10-01	0	0	0	0	0	5
## 42	1969-11-01	0	0	1	0	0	5
## 43	1969-12-01	0	0	0	0	0	3
## 44	1970-01-01	0	0	0	1	0	4
## 45	1970-02-01	0	0	0	0	0	3
## 46	1970-03-01	0	0	0	1	0	1
## 47	1970-04-01	0	0	0	0	0	3
## 48	1970-05-01	0	0	0	0	0	1
## 49	1970-06-01	0	0	0	0	0	2
## 50	1970-07-01	0	0	0	0	0	1
## 51	1970-08-01	0	0	0	0	0	2

## 52	1970-09-01	0	0	0	0	1	10
## 53	1970-10-01	0	0	0	0	0	2
## 54	1970-11-01	0	0	0	1	1	14
## 55	1970-12-01	0	0	0	0	0	6
## 56	1971-01-01	0	0	0	0	0	2
## 57	1971-03-01	0	0	0	0	0	1
## 58	1971-04-01	0	0	0	0	1	3
## 59	1971-05-01	0	0	0	1	0	2
## 60	1971-07-01	0	0	0	0	0	1
## 61	1971-08-01	0	0	0	0	0	9
## 62	1971-09-01	0	0	0	0	0	6
## 63	1971-10-01	0	0	0	0	0	2
## 64	1971-11-01	0	0	0	0	0	13
## 65	1971-12-01	0	0	0	0	0	5
## 66	1972-01-01	0	0	0	0	0	4
## 67	1972-03-01	0	0	0	0	0	6
## 68	1972-05-01	0	0	0	1	0	3
## 69	1972-06-01	0	0	0	0	0	5
## 70	1972-07-01	0	0	0	2	0	1
## 71	1972-08-01	0	0	0	0	0	1
## 72	1972-09-01	0	0	0	0	0	2
## 73	1972-11-01	0	0	0	0	0	1
## 74	1972-12-01	0	0	0	0	0	3
## 75	1973-01-01	0	0	0	0	0	14
## 76	1973-02-01	0	0	0	0	0	1
## 77	1973-03-01	0	0	0	0	0	8
## 78	1973-04-01	0	0	0	0	0	4
## 79	1973-05-01	0	0	0	1	0	2
## 80	1973-07-01	0	0	0	0	0	15
## 81	1973-08-01	0	0	0	1	0	4
## 82	1973-09-01	0	0	0	0	0	1
## 83	1973-10-01	0	0	0	0	0	6
## 84	1973-11-01	0	0	0	0	0	4
## 85	1973-12-01	0	0	0	0	0	6
## 86	1974-01-01	0	0	0	4	0	6
## 87	1974-03-01	0	0	0	0	0	11
## 88	1974-04-01	0	0	0	1	1	4
## 89	1974-05-01	0	0	0	0	0	1
## 90	1974-06-01	0	0	0	1	0	1
## 91	1974-11-01	0	0	0	0	0	13
## 92	1975-01-01	0	0	0	2	0	9
## 93	1975-02-01	0	0	0	0	0	5
## 94	1975-03-01	0	0	0	1	0	4
## 95	1975-04-01	0	0	0	0	0	5
## 96	1975-05-01	0	0	0	0	0	1
## 97	1975-07-01	0	0	0	0	0	4
## 98	1975-08-01	0	0	0	0	0	3
## 99	1975-09-01	0	0	0	0	0	4
## 100	1975-10-01	0	0	0	0	0	1
## 101	1975-11-01	0	0	0	0	0	17
## 102	1976-01-01	0	0	0	3	0	11
## 103	1976-03-01	0	0	0	1	0	1
## 104	1976-04-01	0	0	0	0	0	1
## 105	1976-05-01	0	0	0	0	0	7

## 106	1976-06-01	0	0	0	1	0	0
## 107	1976-07-01	0	0	0	0	0	1
## 108	1976-08-01	0	0	0	0	0	1
## 109	1976-09-01	0	0	0	0	0	2
## 110	1976-10-01	0	0	0	0	0	1
## 111	1976-11-01	0	0	0	1	0	3
## 112	1976-12-01	0	0	0	0	0	13
## 113	1977-01-01	0	0	0	2	0	4
## 114	1977-02-01	0	0	0	0	0	13
## 115	1977-03-01	0	0	0	0	0	2
## 116	1977-04-01	0	0	0	0	0	1
## 117	1977-05-01	0	0	0	0	0	5
## 118	1977-07-01	0	0	0	0	0	1
## 119	1977-09-01	0	0	0	0	0	6
## 120	1977-10-01	0	0	0	0	0	15
## 121	1977-11-01	0	0	0	1	0	3
## 122	1977-12-01	0	0	0	0	0	1
## 123	1978-01-01	0	0	0	0	0	6
## 124	1978-02-01	0	0	1	0	0	7
## 125	1978-03-01	0	0	0	0	0	2
## 126	1978-05-01	0	0	0	0	1	4
## 127	1978-06-01	0	0	1	0	0	12
## 128	1978-07-01	0	0	0	0	0	1
## 129	1978-08-01	0	0	0	0	0	2
## 130	1978-09-01	0	0	1	0	0	1
## 131	1978-10-01	0	0	1	0	0	7
## 132	1978-11-01	0	0	1	0	0	13
## 133	1978-12-01	0	0	0	1	0	0
## 134	1979-01-01	0	0	1	0	0	7
## 135	1979-02-01	0	1	0	0	0	3
## 136	1979-03-01	0	0	0	0	0	6
## 137	1979-05-01	0	0	2	1	0	4
## 138	1979-06-01	0	0	0	0	0	3
## 139	1979-07-01	0	0	0	0	0	2
## 140	1979-08-01	0	0	0	0	0	3
## 141	1979-10-01	0	0	0	0	0	6
## 142	1979-11-01	0	0	0	0	0	10
## 143	1980-01-01	0	0	0	0	0	7
## 144	1980-02-01	0	0	0	0	0	4
## 145	1980-03-01	0	0	0	0	0	3
## 146	1980-04-01	0	0	0	0	0	3
## 147	1980-05-01	0	0	0	3	0	2
## 148	1980-06-01	0	0	0	0	0	9
## 149	1980-07-01	0	0	1	0	0	7
## 150	1980-08-01	0	0	0	0	0	2
## 151	1980-09-01	0	0	1	0	0	2
## 152	1980-10-01	0	0	0	0	0	5
## 153	1980-11-01	0	0	0	0	0	8
## 154	1980-12-01	0	0	0	0	0	1
## 155	1981-01-01	0	0	0	0	0	5
## 156	1981-02-01	0	0	0	1	0	4
## 157	1981-04-01	0	0	0	2	0	1
## 158	1981-07-01	1	0	0	0	0	1
## 159	1981-08-01	0	0	0	0	0	1

## 160	1981-09-01	0	0	0	0	0	1
## 161	1981-10-01	0	0	0	0	0	5
## 162	1981-11-01	0	0	5	0	1	6
## 163	1981-12-01	0	0	0	0	0	1
## 164	1982-01-01	0	2	2	2	0	6
## 165	1982-04-01	0	0	0	0	0	12
## 166	1982-05-01	0	0	2	0	0	11
## 167	1982-09-01	0	0	1	0	0	2
## 168	1982-10-01	0	0	0	0	1	1
## 169	1982-11-01	0	1	1	0	1	5
## 170	1983-01-01	0	1	3	0	0	10
## 171	1983-02-01	0	0	0	0	0	5
## 172	1983-03-01	0	0	2	0	0	6
## 173	1983-04-01	0	0	0	0	0	2
## 174	1983-05-01	0	0	0	0	0	6
## 175	1983-06-01	0	0	1	0	0	6
## 176	1983-08-01	0	0	0	0	0	1
## 177	1983-09-01	0	0	0	0	0	11
## 178	1983-10-01	0	0	1	0	1	3
## 179	1983-11-01	0	0	1	0	0	3
## 180	1983-12-01	0	0	0	0	0	1
## 181	1984-01-01	0	0	0	1	0	17
## 182	1984-02-01	0	0	1	0	0	2
## 183	1984-03-01	0	0	0	2	1	11
## 184	1984-04-01	0	0	1	0	0	1
## 185	1984-05-01	0	0	1	0	0	9
## 186	1984-06-01	0	0	0	2	0	6
## 187	1984-07-01	0	0	0	1	0	1
## 188	1984-08-01	0	0	2	0	0	0
## 189	1984-09-01	0	0	1	0	0	6
## 190	1984-10-01	0	0	1	0	0	3
## 191	1984-11-01	0	0	2	0	0	3
## 192	1984-12-01	0	0	2	0	0	10
## 193	1985-01-01	0	0	0	3	1	8
## 194	1985-02-01	0	0	0	2	0	9
## 195	1985-05-01	0	0	1	0	0	5
## 196	1985-06-01	0	0	3	0	0	9
## 197	1985-07-01	0	0	0	1	0	9
## 198	1985-08-01	0	0	0	0	0	6
## 199	1985-09-01	0	0	1	0	0	0
## 200	1985-10-01	0	0	0	1	0	9
## 201	1985-11-01	0	0	0	0	1	1
## 202	1985-12-01	0	0	0	1	0	7
## 203	1986-01-01	0	1	4	2	0	16
## 204	1986-02-01	0	0	0	12	0	4
## 205	1986-03-01	0	0	1	0	0	7
## 206	1986-05-01	0	1	0	0	0	4
## 207	1986-06-01	0	0	0	0	0	1
## 208	1986-07-01	0	0	0	2	0	0
## 209	1986-08-01	0	0	0	0	0	2
## 210	1986-09-01	0	0	1	1	0	2
## 211	1986-10-01	0	0	0	2	0	4
## 212	1987-01-01	0	0	6	4	3	30
## 213	1987-02-01	0	0	0	4	0	1

## 214	1987-03-01	0	0	2	4	0	7
## 215	1987-04-01	0	0	0	0	0	8
## 216	1987-05-01	0	0	1	0	0	1
## 217	1987-06-01	0	2	1	3	0	3
## 218	1987-07-01	0	1	0	2	1	3
## 219	1987-08-01	0	0	0	2	0	10
## 220	1987-09-01	0	0	2	1	0	3
## 221	1987-10-01	0	0	1	0	1	3
## 222	1987-11-01	0	0	0	0	0	4
## 223	1987-12-01	0	0	1	0	0	0
## 224	1988-01-01	0	2	2	27	5	16
## 225	1988-02-01	0	0	0	0	0	2
## 226	1988-03-01	0	1	0	0	1	7
## 227	1988-04-01	0	0	1	4	1	5
## 228	1988-05-01	0	1	1	3	0	7
## 229	1988-06-01	0	3	1	8	1	2
## 230	1988-08-01	0	0	0	0	12	1
## 231	1988-09-01	0	0	0	0	3	3
## 232	1988-10-01	0	0	3	2	1	1
## 233	1988-11-01	0	3	0	0	0	2
## 234	1988-12-01	0	0	0	0	0	1
## 235	1989-01-01	0	2	2	7	2	14
## 236	1989-03-01	0	0	2	0	0	1
## 237	1989-04-01	0	0	1	0	0	2
## 238	1989-05-01	0	0	0	1	0	4
## 239	1989-06-01	0	2	1	4	0	2
## 240	1989-07-01	0	3	0	1	0	2
## 241	1989-08-01	0	1	0	0	0	3
## 242	1989-09-01	0	0	1	6	0	1
## 243	1989-10-01	0	0	0	3	0	3
## 244	1989-12-01	0	0	0	1	0	0
## 245	1990-01-01	0	4	4	22	2	9
## 246	1990-02-01	0	0	0	1	0	3
## 247	1990-03-01	0	2	4	2	2	3
## 248	1990-04-01	0	0	0	2	1	1
## 249	1990-05-01	0	0	2	2	1	0
## 250	1990-06-01	0	0	1	0	0	4
## 251	1990-07-01	0	1	0	0	0	0
## 252	1990-08-01	0	0	0	1	0	2
## 253	1990-09-01	0	4	1	0	2	3
## 254	1990-10-01	0	0	1	2	0	4
## 255	1990-11-01	0	0	0	12	0	0
## 256	1990-12-01	1	0	0	0	0	0
## 257	1991-01-01	0	5	0	6	3	24
## 258	1991-02-01	0	0	0	0	0	1
## 259	1991-03-01	0	0	1	3	0	3
## 260	1991-04-01	0	0	0	6	0	2
## 261	1991-05-01	0	0	0	8	1	1
## 262	1991-06-01	0	0	1	0	0	0
## 263	1991-07-01	0	1	0	9	1	0
## 264	1991-08-01	0	3	0	0	1	12
## 265	1991-09-01	0	0	0	2	0	14
## 266	1991-10-01	0	0	0	0	2	2
## 267	1991-11-01	0	0	0	6	1	4

## 268	1991-12-01	0	0	1	2	0	1
## 269	1992-01-01	0	1	3	29	2	7
## 270	1992-02-01	0	0	0	6	0	1
## 271	1992-03-01	0	0	0	4	2	3
## 272	1992-04-01	0	0	0	0	1	3
## 273	1992-05-01	0	0	1	1	9	0
## 274	1992-06-01	0	0	1	5	0	7
## 275	1992-07-01	0	1	0	0	5	0
## 276	1992-08-01	0	0	0	7	0	2
## 277	1992-09-01	0	0	0	1	2	0
## 278	1992-10-01	0	0	0	13	0	0
## 279	1992-11-01	0	0	1	5	5	3
## 280	1993-01-01	0	0	3	26	4	7
## 281	1993-02-01	0	0	0	2	3	2
## 282	1993-03-01	0	1	1	1	7	10
## 283	1993-04-01	0	0	0	0	0	4
## 284	1993-05-01	0	0	0	2	0	0
## 285	1993-06-01	0	5	0	3	2	2
## 286	1993-07-01	0	0	0	2	2	0
## 287	1993-08-01	0	0	0	4	0	1
## 288	1993-09-01	0	0	1	1	2	2
## 289	1993-10-01	0	0	0	4	2	3
## 290	1993-11-01	0	0	0	5	13	1
## 291	1993-12-01	0	1	0	5	4	1
## 292	1994-01-01	0	1	5	21	5	7
## 293	1994-02-01	0	0	1	1	1	2
## 294	1994-03-01	0	0	0	2	1	3
## 295	1994-04-01	0	0	1	1	9	3
## 296	1994-05-01	0	0	0	14	4	5
## 297	1994-06-01	0	0	0	5	5	2
## 298	1994-07-01	0	0	1	0	1	3
## 299	1994-08-01	0	0	0	2	1	3
## 300	1994-09-01	0	0	0	0	21	0
## 301	1994-10-01	0	3	0	1	3	14
## 302	1994-11-01	0	0	0	8	3	0
## 303	1995-01-01	0	1	1	8	13	9
## 304	1995-02-01	0	0	0	2	0	1
## 305	1995-03-01	0	0	3	0	5	8
## 306	1995-04-01	0	1	0	2	4	0
## 307	1995-05-01	0	1	1	0	1	0
## 308	1995-06-01	0	0	2	4	2	2
## 309	1995-07-01	0	0	1	16	8	1
## 310	1995-08-01	1	2	0	0	1	0
## 311	1995-09-01	0	0	0	4	0	2
## 312	1995-10-01	0	2	0	0	6	6
## 313	1995-11-01	0	2	0	4	7	0
## 314	1995-12-01	0	0	0	0	0	2
## 315	1996-01-01	1	1	5	15	12	9
## 316	1996-02-01	0	1	1	4	15	1
## 317	1996-03-01	0	0	0	0	0	2
## 318	1996-04-01	0	0	0	12	4	4
## 319	1996-05-01	0	0	0	2	1	0
## 320	1996-06-01	0	0	1	1	2	0
## 321	1996-07-01	0	0	1	1	14	2

## 322	1996-08-01	0	0	0	1	10	2
## 323	1996-09-01	0	0	2	7	0	2
## 324	1996-10-01	0	0	1	10	3	2
## 325	1996-11-01	0	0	2	3	8	0
## 326	1997-01-01	0	5	4	17	5	9
## 327	1997-02-01	0	0	0	9	1	2
## 328	1997-03-01	0	0	0	7	10	3
## 329	1997-04-01	0	0	0	5	0	6
## 330	1997-05-01	0	0	1	0	0	8
## 331	1997-06-01	0	0	0	4	2	0
## 332	1997-07-01	0	4	0	0	0	1
## 333	1997-08-01	0	1	0	1	0	5
## 334	1997-09-01	0	0	1	10	1	3
## 335	1997-10-01	0	2	2	9	0	10
## 336	1997-11-01	0	4	0	4	5	2
## 337	1997-12-01	0	0	0	0	0	10
## 338	1998-01-01	0	3	4	14	40	14
## 339	1998-02-01	0	2	0	0	0	0
## 340	1998-03-01	0	0	0	0	0	4
## 341	1998-04-01	0	1	0	0	2	0
## 342	1998-05-01	0	1	0	4	4	2
## 343	1998-06-01	0	0	0	4	2	1
## 344	1998-07-01	0	0	1	0	0	0
## 345	1998-08-01	0	2	0	16	4	4
## 346	1998-09-01	0	4	0	1	5	2
## 347	1998-10-01	0	1	0	3	7	5
## 348	1998-11-01	0	0	3	1	9	9
## 349	1998-12-01	0	0	0	1	0	2
## 350	1999-01-01	0	1	8	19	25	14
## 351	1999-02-01	0	1	2	5	1	3
## 352	1999-03-01	0	0	0	7	2	1
## 353	1999-04-01	0	0	0	1	4	7
## 354	1999-05-01	0	1	0	4	2	2
## 355	1999-06-01	1	0	0	2	1	7
## 356	1999-07-01	0	2	0	2	10	2
## 357	1999-08-01	0	0	0	2	4	1
## 358	1999-09-01	0	0	0	5	1	2
## 359	1999-10-01	0	16	0	4	3	5
## 360	1999-11-01	0	0	0	2	8	2
## 361	1999-12-01	0	0	0	1	0	1
## 362	2000-01-01	2	1	1	24	10	18
## 363	2000-02-01	0	1	0	1	2	3
## 364	2000-03-01	1	1	1	0	1	4
## 365	2000-04-01	0	3	0	16	0	1
## 366	2000-05-01	0	0	1	0	1	3
## 367	2000-06-01	0	0	0	2	8	2
## 368	2000-07-01	0	0	1	4	0	4
## 369	2000-08-01	0	0	0	6	3	0
## 370	2000-09-01	0	0	0	1	2	3
## 371	2000-10-01	0	0	0	7	5	6
## 372	2000-11-01	0	0	0	4	0	1
## 373	2000-12-01	0	2	0	0	2	0
## 374	2001-01-01	0	2	8	21	13	21
## 375	2001-02-01	0	1	0	0	1	0

## 376	2001-03-01	1	3	3	3	5	3
## 377	2001-04-01	0	2	0	0	1	2
## 378	2001-05-01	0	0	2	2	3	9
## 379	2001-06-01	0	0	0	2	1	2
## 380	2001-07-01	0	0	2	1	1	4
## 381	2001-08-01	0	7	0	11	0	3
## 382	2001-09-01	0	0	1	1	0	9
## 383	2001-10-01	0	0	2	10	4	3
## 384	2001-11-01	0	0	2	3	1	4
## 385	2001-12-01	0	2	0	4	7	0
## 386	2002-01-01	0	4	7	34	14	14
## 387	2002-02-01	0	0	0	4	3	5
## 388	2002-03-01	0	0	0	2	1	1
## 389	2002-04-01	0	0	2	2	2	0
## 390	2002-05-01	1	1	1	3	4	0
## 391	2002-06-01	0	4	1	4	2	1
## 392	2002-07-01	0	1	0	3	2	2
## 393	2002-08-01	0	2	2	0	2	1
## 394	2002-09-01	0	0	0	6	1	1
## 395	2002-10-01	1	0	0	1	2	3
## 396	2002-11-01	0	2	1	14	5	8
## 397	2002-12-01	0	0	0	6	2	0
## 398	2003-01-01	0	7	9	27	28	18
## 399	2003-02-01	0	0	0	6	17	3
## 400	2003-03-01	0	2	0	3	0	9
## 401	2003-04-01	0	0	2	1	0	6
## 402	2003-05-01	0	0	1	0	2	2
## 403	2003-06-01	0	1	2	9	4	0
## 404	2003-07-01	0	4	0	6	1	0
## 405	2003-08-01	0	4	0	1	5	2
## 406	2003-09-01	0	0	1	10	4	5
## 407	2003-10-01	0	0	4	8	1	4
## 408	2003-11-01	0	4	2	1	5	2
## 409	2003-12-01	1	1	0	3	2	0
## 410	2004-01-01	0	22	3	23	31	25
## 411	2004-02-01	0	0	0	3	3	2
## 412	2004-03-01	0	5	2	11	9	10
## 413	2004-04-01	0	3	0	2	1	6
## 414	2004-05-01	0	5	1	0	2	1
## 415	2004-06-01	1	3	1	1	0	5
## 416	2004-07-01	0	10	0	0	8	4
## 417	2004-08-01	0	0	1	15	0	1
## 418	2004-09-01	1	2	2	5	0	2
## 419	2004-10-01	1	0	1	6	1	3
## 420	2004-11-01	0	1	1	4	9	0
## 421	2004-12-01	0	2	1	1	9	0
## 422	2005-01-01	1	21	6	25	34	41
## 423	2005-02-01	0	2	1	1	0	3
## 424	2005-03-01	0	0	0	3	12	6
## 425	2005-04-01	0	2	0	4	5	3
## 426	2005-05-01	0	1	0	6	1	8
## 427	2005-06-01	0	2	1	9	6	9
## 428	2005-07-01	1	3	1	0	8	5
## 429	2005-08-01	0	2	3	0	1	6

## 430	2005-09-01	0	1	0	3	4	14
## 431	2005-10-01	1	4	0	8	1	6
## 432	2005-11-01	0	2	3	6	6	4
## 433	2005-12-01	0	1	0	5	10	10
## 434	2006-01-01	4	45	9	38	37	19
## 435	2006-02-01	0	0	0	0	3	2
## 436	2006-03-01	2	0	1	3	5	3
## 437	2006-04-01	3	1	1	0	1	3
## 438	2006-05-01	0	6	0	0	1	5
## 439	2006-06-01	0	5	1	5	11	10
## 440	2006-07-01	0	2	0	3	1	1
## 441	2006-08-01	0	2	4	4	2	2
## 442	2006-09-01	1	2	3	3	2	2
## 443	2006-10-01	0	3	8	7	5	8
## 444	2006-11-01	0	8	5	0	3	1
## 445	2006-12-01	2	1	1	4	2	0
## 446	2007-01-01	3	33	11	37	29	21
## 447	2007-02-01	0	1	2	2	1	3
## 448	2007-03-01	2	3	2	3	8	1
## 449	2007-04-01	2	0	3	3	1	6
## 450	2007-05-01	1	3	1	4	0	9
## 451	2007-06-01	4	10	2	2	9	12
## 452	2007-07-01	1	1	1	1	2	3
## 453	2007-08-01	0	4	1	1	5	2
## 454	2007-09-01	2	3	2	6	3	7
## 455	2007-10-01	1	4	1	4	2	5
## 456	2007-11-01	2	4	1	0	3	8
## 457	2007-12-01	0	15	4	4	4	5
## 458	2008-01-01	2	22	24	33	36	41
## 459	2008-02-01	1	1	0	7	10	2
## 460	2008-03-01	1	1	2	4	0	3
## 461	2008-04-01	1	32	3	5	3	12
## 462	2008-05-01	3	3	5	2	1	4
## 463	2008-06-01	2	3	4	4	1	5
## 464	2008-07-01	2	7	1	0	1	16
## 465	2008-08-01	6	3	11	6	1	5
## 466	2008-09-01	3	7	5	8	3	9
## 467	2008-10-01	4	1	5	14	0	12
## 468	2008-11-01	0	7	5	6	2	4
## 469	2008-12-01	7	4	1	3	2	0
## 470	2009-01-01	5	40	23	20	15	23
## 471	2009-02-01	3	1	0	2	2	5
## 472	2009-03-01	3	5	4	3	2	1
## 473	2009-04-01	2	2	0	2	1	0
## 474	2009-05-01	3	5	3	3	6	1
## 475	2009-06-01	5	7	4	1	4	4
## 476	2009-07-01	2	2	6	9	2	3
## 477	2009-08-01	7	5	10	1	3	5
## 478	2009-09-01	2	2	4	7	4	9
## 479	2009-10-01	1	0	4	2	1	3
## 480	2009-11-01	1	5	4	1	3	2
## 481	2009-12-01	0	1	1	1	3	9
## 482	2010-01-01	5	42	55	12	21	18
## 483	2010-02-01	1	3	11	3	5	2

## 484	2010-03-01	3	4	8	2	1	4
## 485	2010-04-01	4	4	4	3	4	0
## 486	2010-05-01	5	3	10	3	2	4
## 487	2010-06-01	1	6	6	4	1	6
## 488	2010-07-01	1	3	2	2	2	3
## 489	2010-08-01	1	3	0	9	1	21
## 490	2010-09-01	2	3	14	3	4	2
## 491	2010-10-01	4	7	8	15	8	15
## 492	2010-11-01	3	8	10	6	2	6
## 493	2010-12-01	1	5	5	2	3	1
## 494	2011-01-01	14	10	43	28	19	24
## 495	2011-02-01	1	3	4	7	3	2
## 496	2011-03-01	1	5	4	5	12	3
## 497	2011-04-01	1	0	3	8	7	5
## 498	2011-05-01	9	2	8	11	6	4
## 499	2011-06-01	3	8	8	2	0	4
## 500	2011-07-01	5	5	1	6	3	3
## 501	2011-08-01	6	9	6	7	2	4
## 502	2011-09-01	3	3	10	5	3	7
## 503	2011-10-01	3	3	14	3	1	8
## 504	2011-11-01	6	6	9	10	2	6
## 505	2011-12-01	2	2	2	9	1	7
## 506	2012-01-01	29	40	28	37	30	9
## 507	2012-02-01	7	3	5	1	2	3
## 508	2012-03-01	2	2	7	9	0	5
## 509	2012-04-01	5	5	1	11	6	7
## 510	2012-05-01	8	11	6	6	1	3
## 511	2012-06-01	10	6	17	6	4	2
## 512	2012-07-01	3	2	6	1	6	0
## 513	2012-08-01	7	5	1	2	8	6
## 514	2012-09-01	12	9	18	8	4	9
## 515	2012-10-01	9	4	19	6	20	3
## 516	2012-11-01	11	4	14	8	1	2
## 517	2012-12-01	14	3	9	3	7	2
## 518	2013-01-01	49	20	65	17	24	13
## 519	2013-02-01	7	3	5	4	3	3
## 520	2013-03-01	22	3	9	6	4	4
## 521	2013-04-01	17	6	7	7	3	0
## 522	2013-05-01	21	8	15	5	2	4
## 523	2013-06-01	17	4	7	8	5	5
## 524	2013-07-01	17	8	3	8	4	7
## 525	2013-08-01	10	9	2	6	7	7
## 526	2013-09-01	20	7	21	9	3	6
## 527	2013-10-01	25	6	13	9	5	7
## 528	2013-11-01	20	7	15	10	7	2
## 529	2013-12-01	22	14	11	3	4	0
## 530	2014-01-01	54	15	43	21	16	7
## 531	2014-02-01	34	9	6	4	5	6
## 532	2014-03-01	51	13	17	9	21	7
## 533	2014-04-01	49	4	14	14	2	7
## 534	2014-05-01	44	7	15	16	11	12
## 535	2014-06-01	46	9	19	11	16	9
## 536	2014-07-01	25	12	13	6	7	3
## 537	2014-08-01	27	14	10	12	10	3

## 538	2014-09-01	44	17	29	14	5	4
## 539	2014-10-01	44	18	42	11	17	14
## 540	2014-11-01	22	10	27	21	16	9
## 541	2014-12-01	21	9	12	11	9	1
## 542	2015-01-01	18	17	33	13	10	8
## 543	2015-02-01	21	9	19	15	13	3
## 544	2015-03-01	29	17	29	8	19	4
## 545	2015-04-01	37	15	28	14	5	4
## 546	2015-05-01	45	16	26	13	10	3
## 547	2015-06-01	29	33	40	28	16	8
## 548	2015-07-01	35	25	28	24	19	9
## 549	2015-08-01	28	26	32	29	12	5
## 550	2015-09-01	27	23	36	13	24	8
## 551	2015-10-01	32	24	50	23	15	3
## 552	2015-11-01	40	8	43	15	23	4
## 553	2015-12-01	32	12	21	3	16	2
## 554	2016-01-01	22	11	21	32	15	4
## 555	2016-02-01	25	26	26	16	15	3
## 556	2016-03-01	24	13	25	19	9	5
## 557	2016-04-01	35	25	33	25	19	3
## 558	2016-05-01	47	27	52	20	28	2
## 559	2016-06-01	39	33	20	18	24	9
## 560	2016-07-01	38	23	38	21	33	2
## 561	2016-08-01	33	25	42	28	10	5
## 562	2016-09-01	29	31	55	27	15	3
## 563	2016-10-01	49	23	58	30	21	10
## 564	2016-11-01	31	42	40	16	16	8
## 565	2016-12-01	28	21	20	15	25	5
## 566	2017-01-01	23	37	28	16	10	8
## 567	2017-02-01	21	25	35	22	24	0
## 568	2017-03-01	40	21	43	33	38	4
## 569	2017-04-01	37	19	46	28	37	11
## 570	2017-05-01	27	22	31	23	34	20
## 571	2017-06-01	58	36	59	40	34	11
## 572	2017-07-01	21	21	33	30	25	6
## 573	2017-08-01	32	31	28	34	19	6
## 574	2017-09-01	43	28	45	25	35	11
## 575	2017-10-01	32	20	32	54	32	10
## 576	2017-11-01	34	40	21	41	33	5
## 577	2017-12-01	32	20	10	26	33	3
## 578	2018-01-01	29	15	18	18	33	6
## 579	2018-02-01	35	23	37	19	43	8
## 580	2018-03-01	58	22	31	23	49	9
## 581	2018-04-01	41	33	47	39	33	9
## 582	2018-05-01	40	48	37	28	31	17
## 583	2018-06-01	55	37	55	54	42	6
## 584	2018-07-01	62	30	32	27	54	7
## 585	2018-08-01	42	52	69	29	55	6
## 586	2018-09-01	35	35	47	25	52	8
## 587	2018-10-01	68	53	71	40	63	13
## 588	2018-11-01	67	63	74	33	58	10
## 589	2018-12-01	69	52	48	23	62	5
## 590	2019-01-01	113	46	63	33	41	13
## 591	2019-02-01	110	66	70	44	75	16

```
## 592      2019-03-01 125      79 82 47 58 25
## 593      2019-04-01 74      93 66 39 70 31
## 594      2019-05-01 111     154 74 63 115 22
## 595      2019-06-01 133     107 87 78 98 19
## 596      2019-07-01 102     116 99 74 95 31
## 597      2019-08-01 158     124 97 83 106 39
## 598      2019-09-01 160     116 92 82 136 64
## 599      2019-10-01 199     184 113 99 146 78
## 600      2019-11-01 293     241 134 128 277 86
## 601      2019-12-01 271     190 111 107 255 37
## 602      2020-01-01 140     108 124 80 129 39
## 603      <NA> 61      159 157 262 226 633
```

#Analiza składowych podgrup

Selekcja danych

```
dane_podgrupy <- dane %>%
  group_by(playlist_genre, playlist_subgenre) %>%
  select_if(is.numeric)
```

```
head(dane_podgrupy)
```

```
## # A tibble: 6 x 17
## # Groups:   playlist_genre, playlist_subgenre [6]
##   playlist_genre playlist_subgenre~ X track_popularity danceability energy
##   <chr>          <chr>          <int>          <int>          <dbl> <dbl>
## 1 latin         latin hip hop      20378           65      0.816 0.754
## 2 pop           electropop        3549            0      0.611 0.988
## 3 edm           big room         28314           31      0.641 0.874
## 4 pop           dance pop         149             81      0.649 0.716
## 5 pop           post-teen pop     1931            73      0.774 0.75
## 6 rap           hip hop          5541            83      0.687 0.449
## # ... with 11 more variables: key <int>, loudness <dbl>, mode <int>,
## #   speechiness <dbl>, acousticness <dbl>, instrumentalness <dbl>,
## #   liveness <dbl>, valence <dbl>, tempo <dbl>, duration_ms <int>,
## #   miesiac <dbl>
```

```
tail(dane_podgrupy)
```

```
## # A tibble: 6 x 17
## # Groups:   playlist_genre, playlist_subgenre [6]
##   playlist_genre playlist_subgenre~ X track_popularity danceability energy
##   <chr>          <chr>          <int>          <int>          <dbl> <dbl>
## 1 r&b           hip pop          23727           61      0.768 0.562
## 2 latin         latin hip hop     20376           75      0.804 0.705
## 3 rock          hard rock        15792           21      0.711 0.747
## 4 rock          permanent wave   13877           34      0.379 0.396
## 5 pop           post-teen pop    2084            56      0.531 0.827
## 6 edm           pop edm          30895           27      0.443 0.948
## # ... with 11 more variables: key <int>, loudness <dbl>, mode <int>,
## #   speechiness <dbl>, acousticness <dbl>, instrumentalness <dbl>,
## #   liveness <dbl>, valence <dbl>, tempo <dbl>, duration_ms <int>,
## #   miesiac <dbl>
```

```
unique(dane_podgrupy$playlist_subgenre)
```

```
## [1] "latin hip hop"      "electropop"
## [3] "big room"           "dance pop"
## [5] "post-teen pop"      "hip hop"
## [7] "reggaeton"          "new jack swing"
## [9] "album rock"         "progressive electro house"
## [11] "hip pop"            "permanent wave"
## [13] "electro house"      "latin pop"
## [15] "trap"               "classic rock"
## [17] "gangster rap"       "hard rock"
## [19] "tropical"           "urban contemporary"
## [21] "indie popitism"     "southern hip hop"
## [23] "neo soul"           "pop edm"
```

widzimy, że dane podgrupy poszczególnych gatunków są wymieszane między sobą.

```
dane_podgrupy %>%
  group_by(playlist_genre, playlist_subgenre) %>%
  summarise(ilosc = n()) %>%
  arrange(desc(ilosc))
```

```
## # A tibble: 24 x 3
## # Groups:   playlist_genre [6]
##   playlist_genre playlist_subgenre      ilosc
##   <chr>          <chr>          <int>
## 1 edm            progressive electro house 1456
## 2 latin          latin hip hop      1355
## 3 pop            indie popitism     1355
## 4 rap            southern hip hop   1339
## 5 r&b            neo soul          1299
## 6 edm            electro house      1212
## 7 edm            pop edm           1207
## 8 rock           hard rock          1193
## 9 rap            gangster rap       1171
## 10 pop           electropop         1125
## # ... with 14 more rows
```

widzimy, że największą podgrupą gatunkową jest progressive elctro house, następnie indie poptisimism

```
dane_numeryczne <- dane %>%
  group_by(playlist_genre) %>%
  select_if(is.numeric)

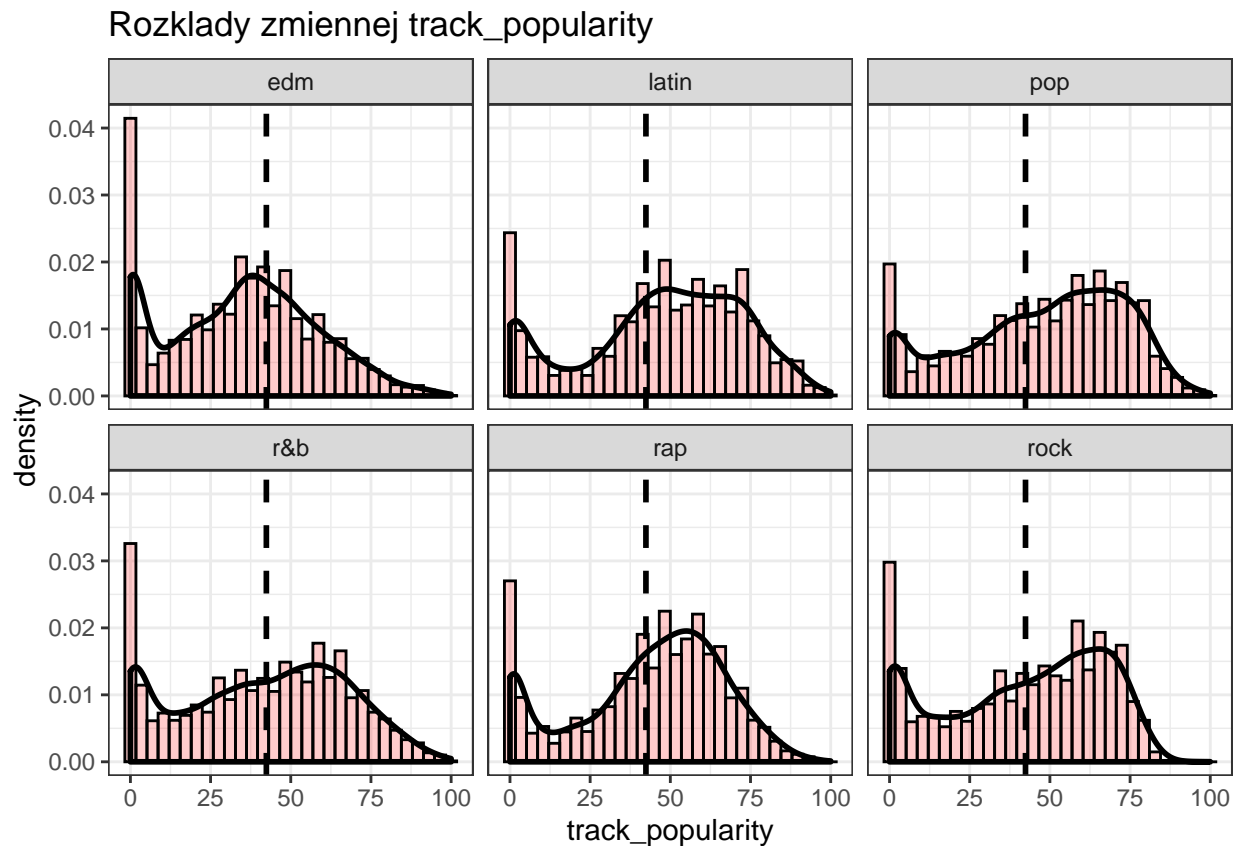
colnames(dane_numeryczne)
```

```
## [1] "playlist_genre" "X"          "track_popularity" "danceability"
## [5] "energy"         "key"        "loudness"         "mode"
## [9] "speechiness"    "acousticness" "instrumentalness" "liveness"
## [13] "valence"        "tempo"      "duration_ms"      "miesiac"
```

```
#track_popularity
```

```
ggplot(dane_numeryczne, aes(track_popularity))+  
  geom_histogram(aes(y = ..density..), color = "black", alpha = 0.2, fill = "red")+  
  geom_density(color = "black", size = 1 )+  
  geom_vline(xintercept = mean(dane_numeryczne$track_popularity), size = 1, linetype = 2)+  
  facet_wrap(~playlist_genre)+  
  theme_bw()+  
  labs(title = "Rozkłady zmiennej track_popularity")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

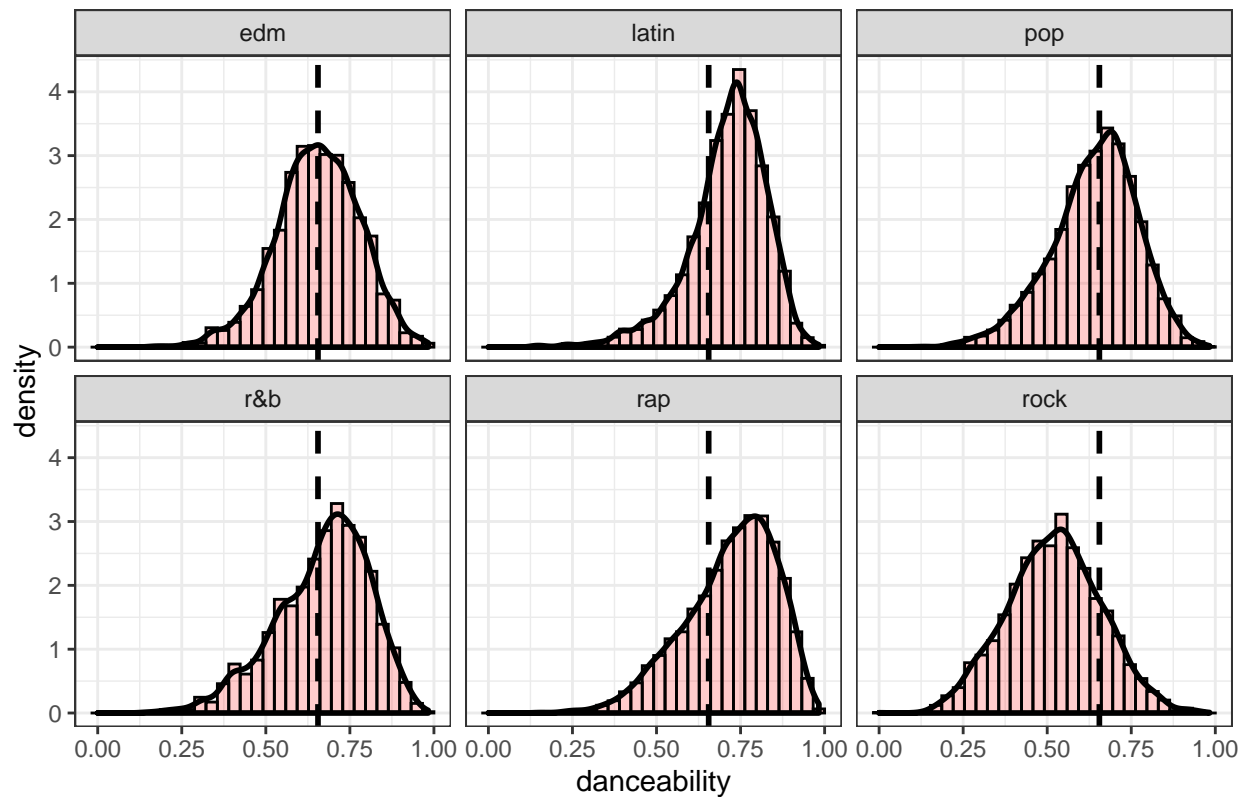


```
#danceability
```

```
ggplot(dane_numeryczne, aes(danceability))+  
  geom_histogram(aes(y = ..density..), color = "black", alpha = 0.2, fill = "red")+  
  geom_density(color = "black", size = 1 )+  
  geom_vline(xintercept = mean(dane_numeryczne$danceability), size = 1, linetype = 2)+  
  facet_wrap(~playlist_genre)+  
  theme_bw()+  
  labs(title = "Rozkłady zmiennej danceability")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Rozkłady zmiennej danceability

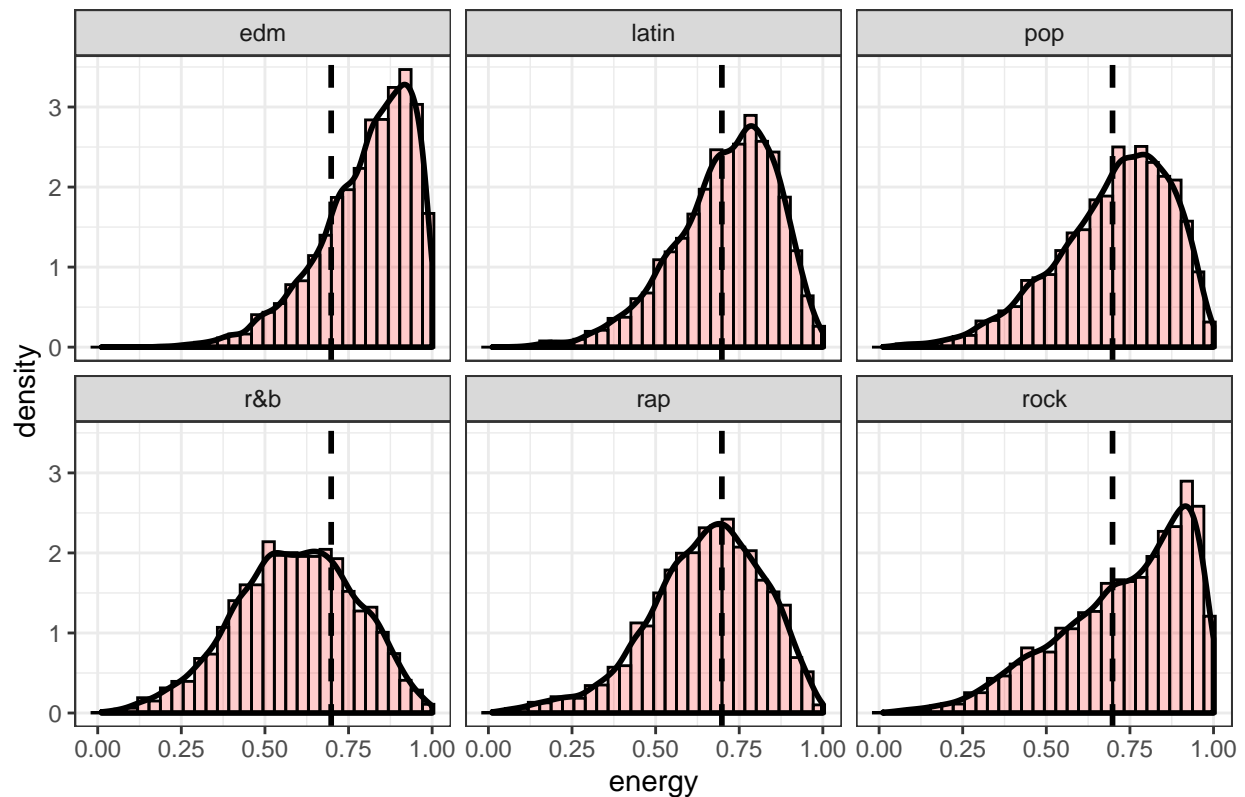


#Energy

```
ggplot(dane_numeryczne, aes(energy))+
  geom_histogram(aes(y = ..density..), color = "black", alpha = 0.2, fill = "red")+
  geom_density(color = "black", size = 1)+
  geom_vline(xintercept = mean(dane_numeryczne$energy), size = 1, linetype = 2)+
  facet_wrap(~playlist_genre)+
  theme_bw()+
  labs(title = "Rozkłady zmiennej energy")
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Rozkłady zmiennej energy

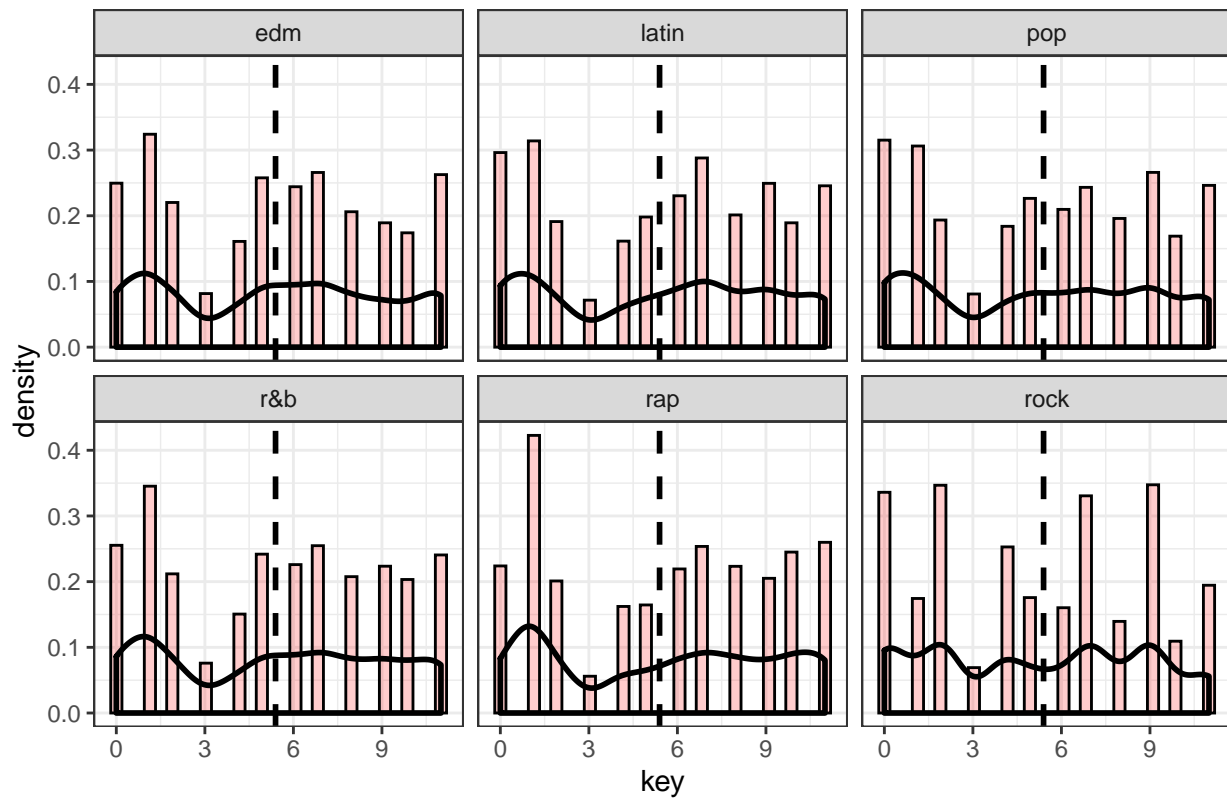


#key

```
ggplot(dane_numeryczne, aes(key)) +
  geom_histogram(aes(y = ..density..), color = "black", alpha = 0.2, fill = "red") +
  geom_density(color = "black", size = 1) +
  geom_vline(xintercept = mean(dane_numeryczne$key), size = 1, linetype = 2) +
  facet_wrap(~playlist_genre) +
  theme_bw() +
  labs(title = "Rozkłady zmiennej key")
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Rozkłady zmiennej key

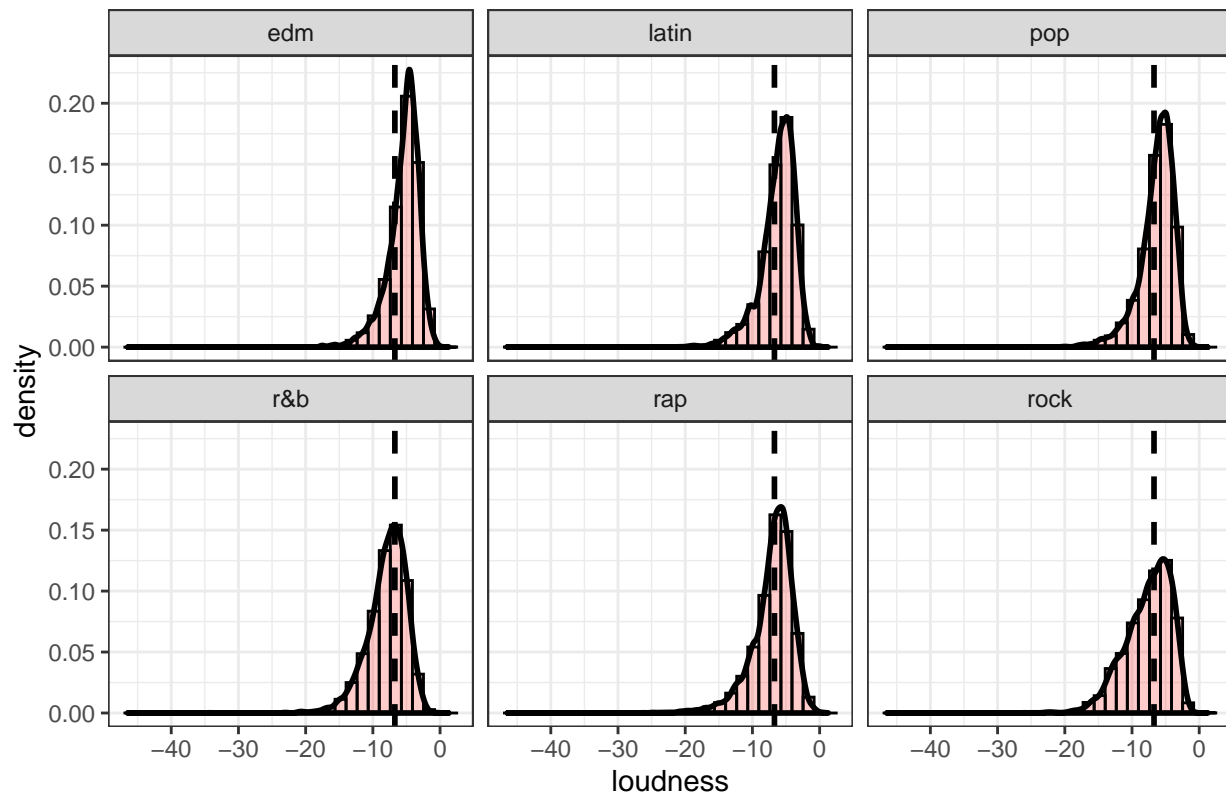


#loudness

```
ggplot(dane_numeryczne, aes(loudness))+
  geom_histogram(aes(y = ..density..), color = "black", alpha = 0.2, fill = "red")+
  geom_density(color = "black", size = 1)+
  geom_vline(xintercept = mean(dane_numeryczne$loudness), size = 1, linetype = 2)+
  facet_wrap(~playlist_genre)+
  theme_bw()+
  labs(title = "Rozkłady zmiennej loudness")
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Rozkłady zmiennej loudness

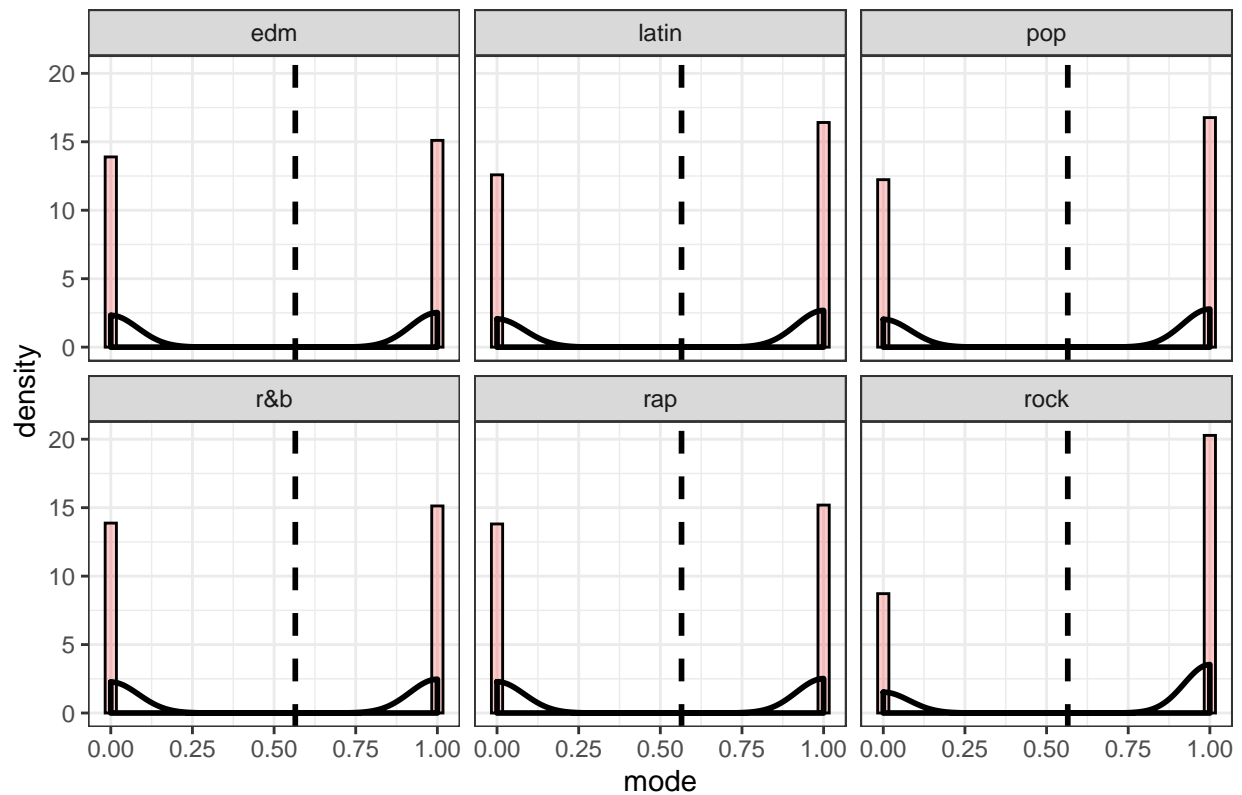


#mode

```
ggplot(dane_numeryczne, aes(mode)) +
  geom_histogram(aes(y = ..density..), color = "black", alpha = 0.2, fill = "red") +
  geom_density(color = "black", size = 1) +
  geom_vline(xintercept = mean(dane_numeryczne$mode), size = 1, linetype = 2) +
  facet_wrap(~playlist_genre) +
  theme_bw() +
  labs(title = "Rozkłady zmiennej mode")
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Rozkłady zmiennej mode

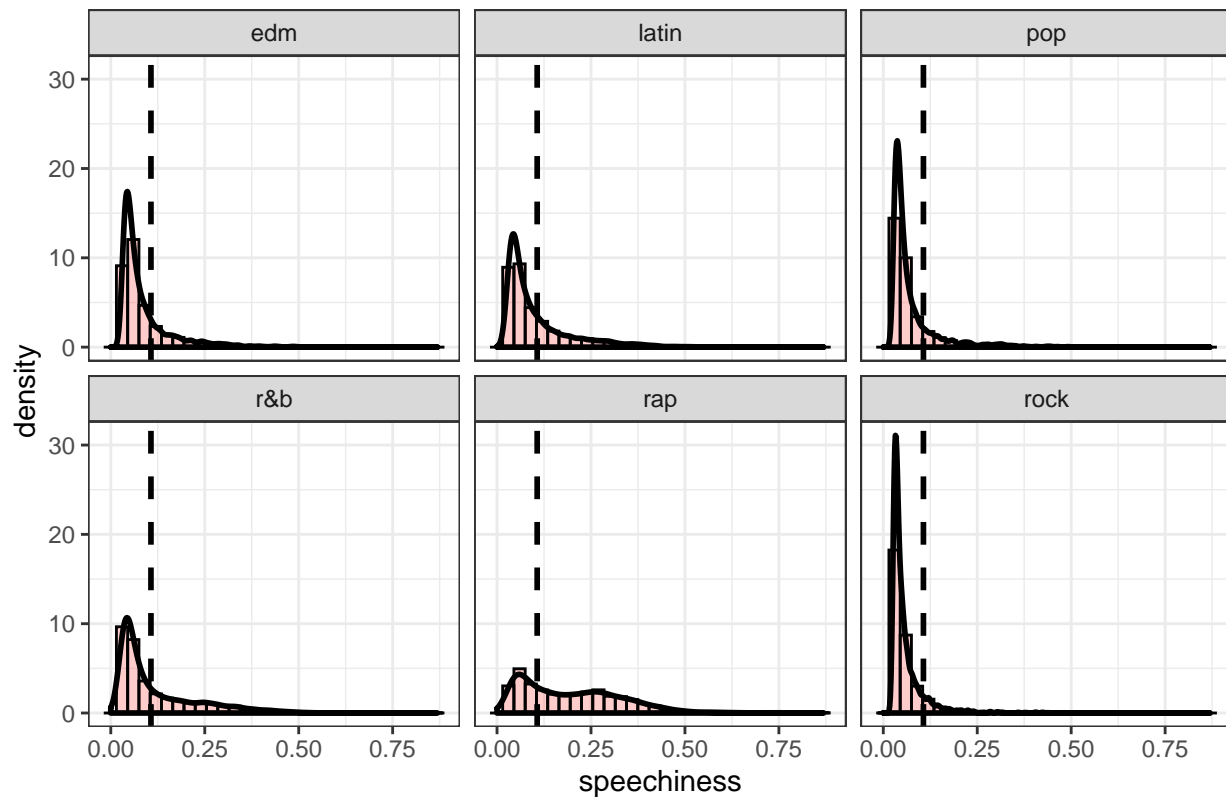


#Speechiness

```
ggplot(dane_numeryczne, aes(speechiness))+
  geom_histogram(aes(y = ..density..), color = "black", alpha = 0.2, fill = "red")+
  geom_density(color = "black", size = 1 )+
  geom_vline(xintercept = mean(dane_numeryczne$speechiness), size = 1, linetype = 2)+
  facet_wrap(~playlist_genre)+
  theme_bw()+
  labs(title = "Rozkłady zmiennej speechiness")
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Rozkłady zmiennej speechiness

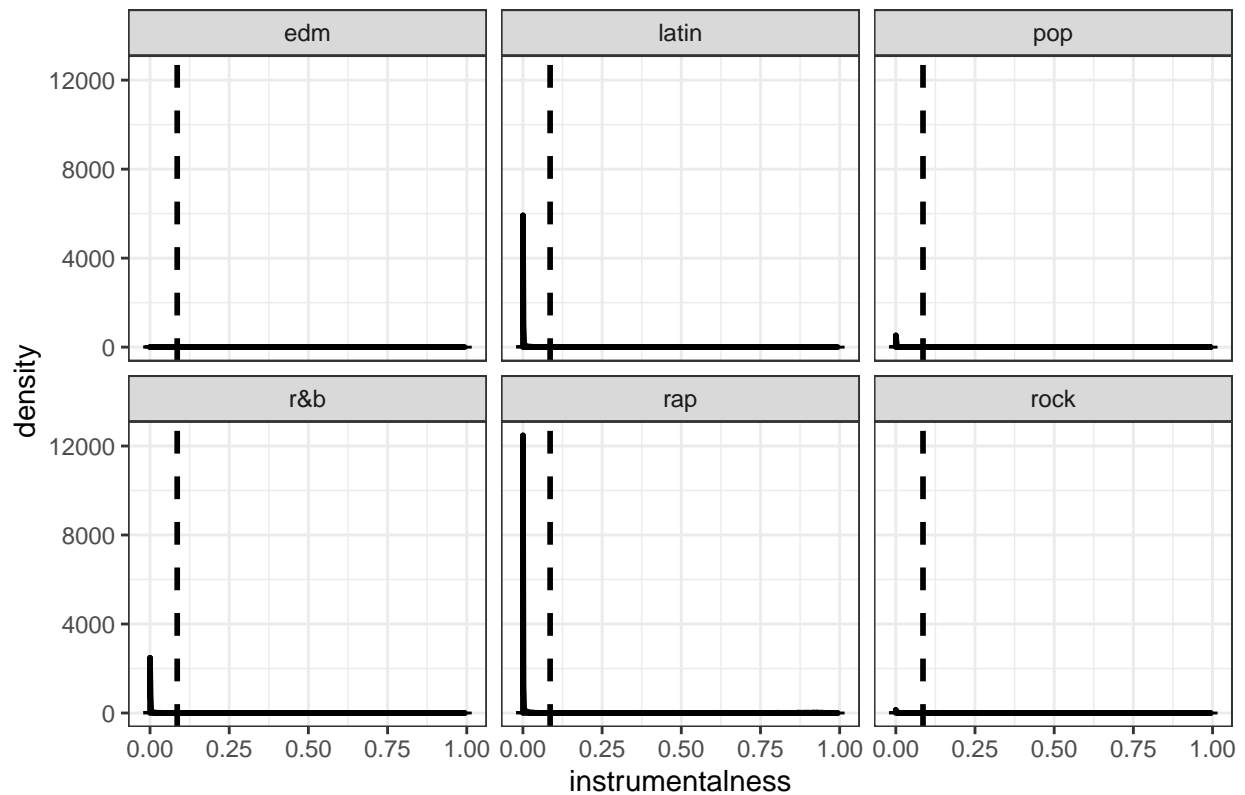


#instrumentalness

```
ggplot(dane_numeryczne, aes(instrumentalness))+
  geom_histogram(aes(y = ..density..), color = "black", alpha = 0.2, fill = "red")+
  geom_density(color = "black", size = 1 )+
  geom_vline(xintercept = mean(dane_numeryczne$instrumentalness), size = 1, linetype = 2)+
  facet_wrap(~playlist_genre)+
  theme_bw()+
  labs(title = "Rozkłady zmiennej instrumentalness")
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Rozkłady zmiennej instrumentalness



```
range(dane_numeryczne$instrumentalness)
```

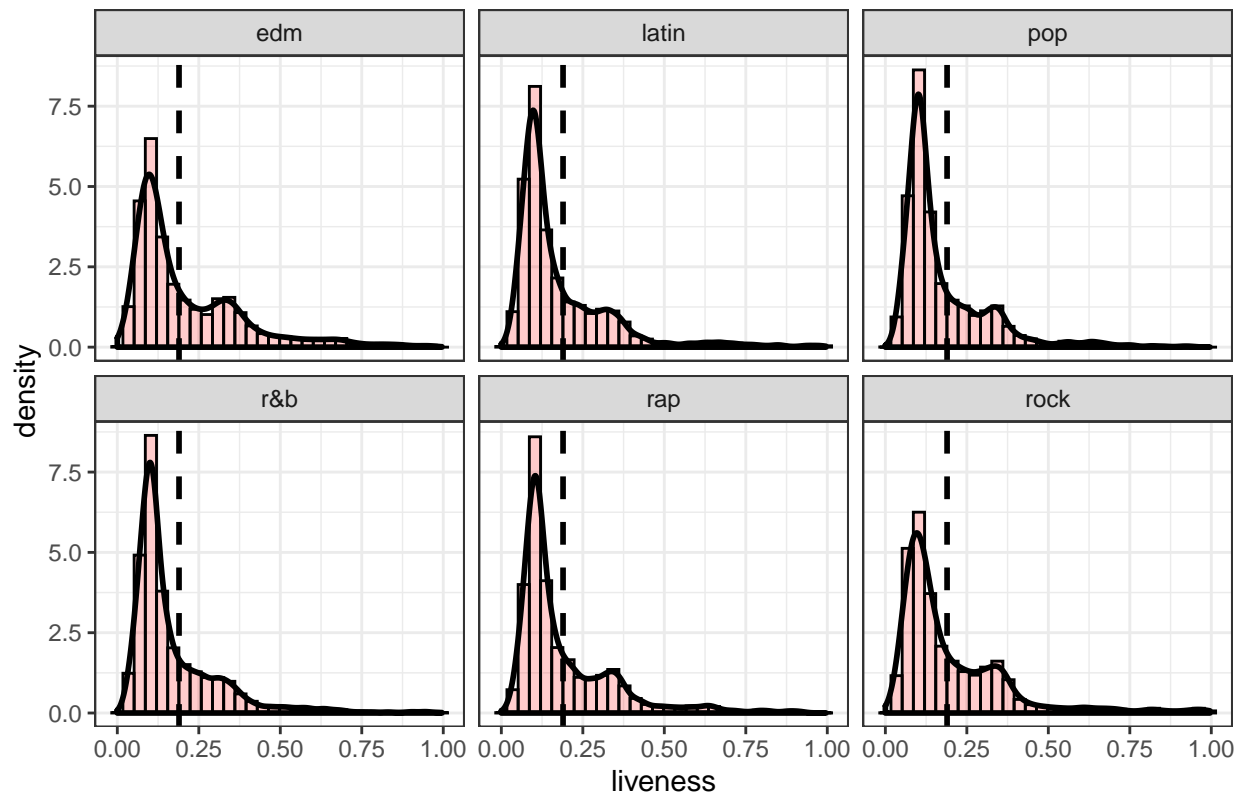
```
## [1] 0.000 0.994
```

```
#liveness
```

```
ggplot(dane_numeryczne, aes(liveness))+
  geom_histogram(aes(y = ..density..), color = "black", alpha = 0.2, fill = "red")+
  geom_density(color = "black", size = 1 )+
  geom_vline(xintercept = mean(dane_numeryczne$liveness), size = 1, linetype = 2)+
  facet_wrap(~playlist_genre)+
  theme_bw()+
  labs(title = "Rozkłady zmiennej liveness")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Rozkłady zmiennej liveness

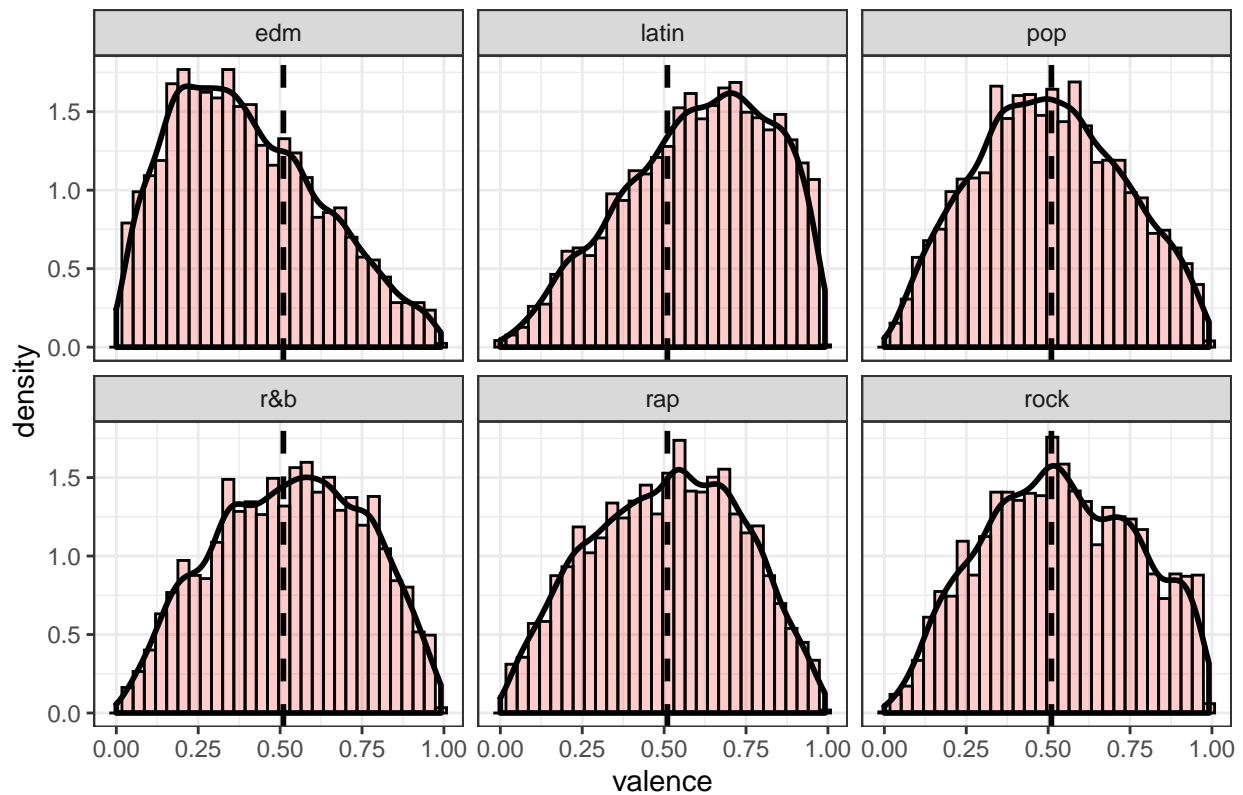


#valence

```
ggplot(dane_numeryczne, aes(valence))+
  geom_histogram(aes(y = ..density..), color = "black", alpha = 0.2, fill = "red")+
  geom_density(color = "black", size = 1 )+
  geom_vline(xintercept = mean(dane_numeryczne$valence), size = 1, linetype = 2)+
  facet_wrap(~playlist_genre)+
  theme_bw()+
  labs(title = "Rozkłady zmiennej valence")
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Rozkłady zmiennej valence

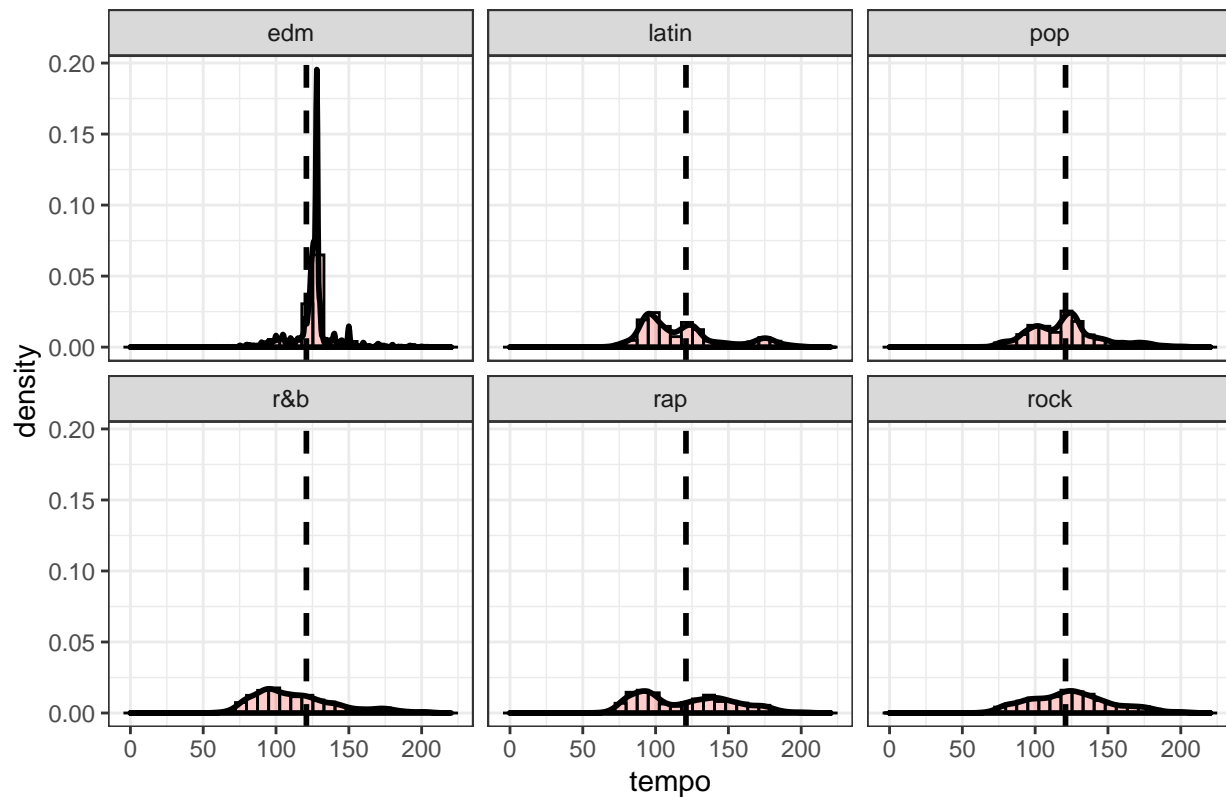


#tempo

```
ggplot(dane_numeryczne, aes(tempo))+
  geom_histogram(aes(y = ..density..), color = "black", alpha = 0.2, fill = "red")+
  geom_density(color = "black", size = 1)+
  geom_vline(xintercept = mean(dane_numeryczne$tempo), size = 1, linetype = 2)+
  facet_wrap(~playlist_genre)+
  theme_bw()+
  labs(title = "Rozkłady zmiennej tempo")
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Rozkłady zmiennej tempo



#duration

```
ggplot(dane_numeryczne, aes(duration_ms))+
  geom_histogram(aes(y = ..density..), color = "black", alpha = 0.2, fill = "red")+
  geom_density(color = "black", size = 1)+
  geom_vline(xintercept = mean(dane_numeryczne$duration_ms), size = 1, linetype = 2)+
  facet_wrap(~playlist_genre)+
  theme_bw()+
  labs(title = "Rozkłady zmiennej duration_ms")
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Rozkłady zmiennej duration_ms

