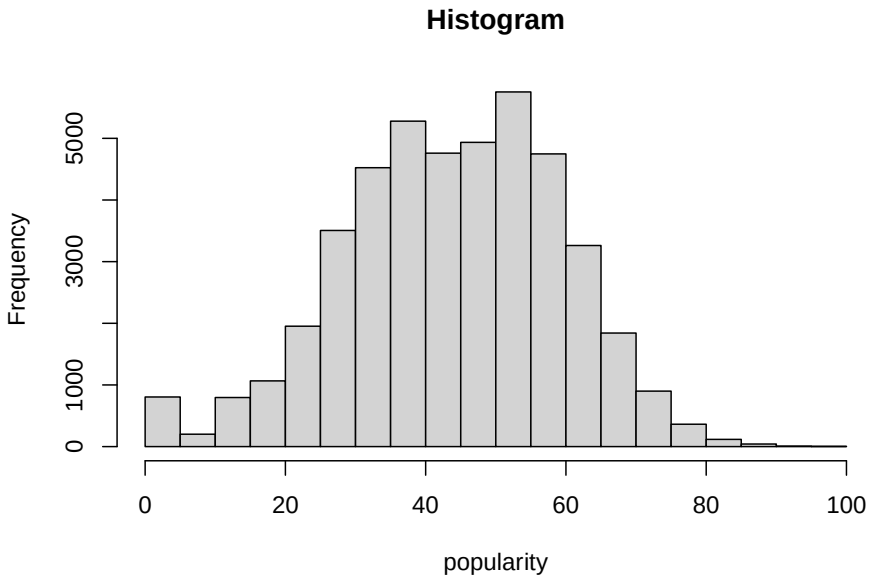
Musique

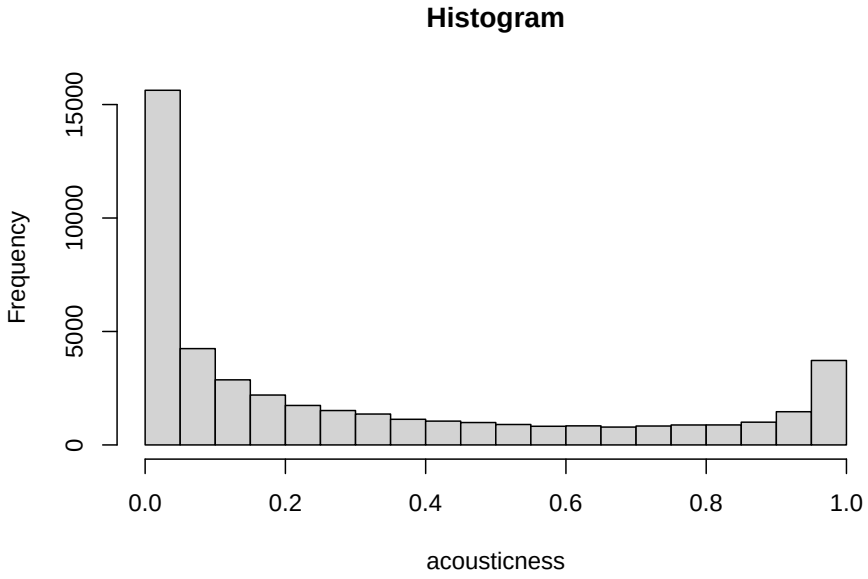
Visualisation

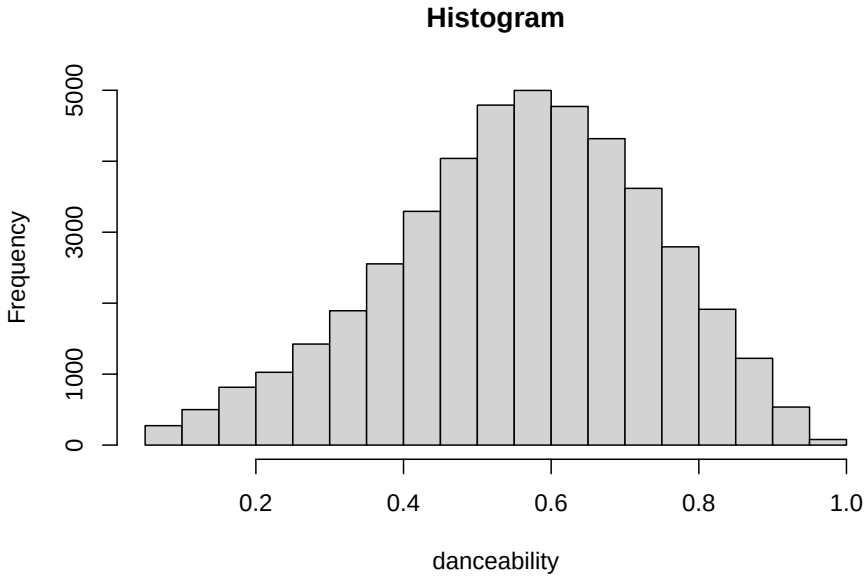
UNIVARIATE

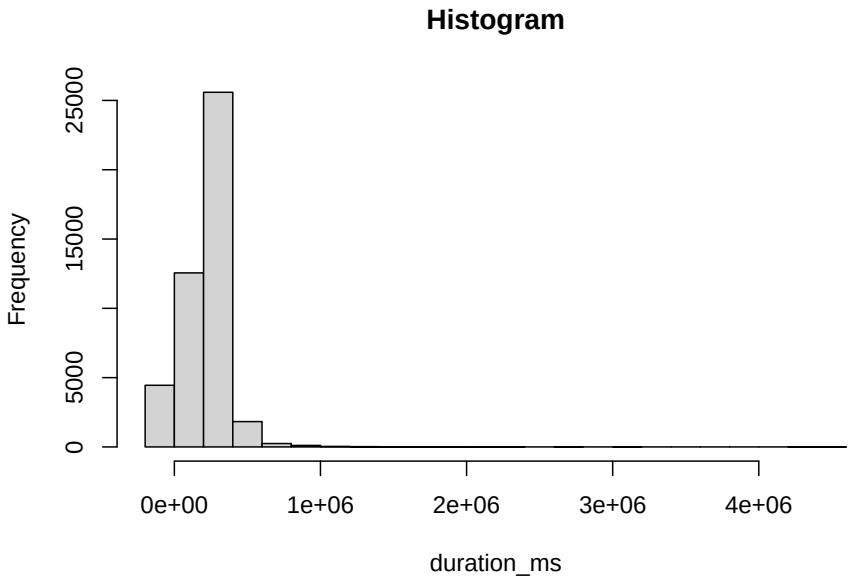
Histogram

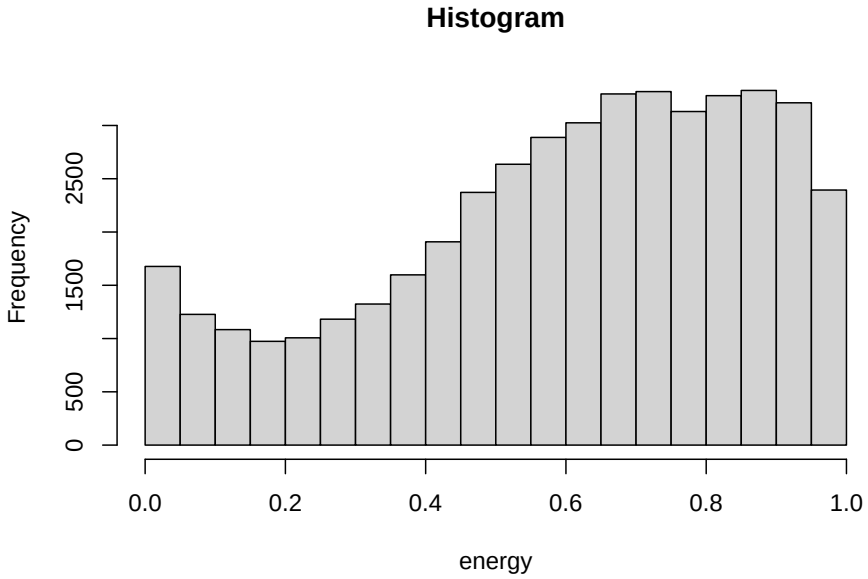
```
x <- musique[,c(4:9, 11:12, 14:15, 17)]

# Histogram: continous variables
for(i in 1:11){
  hist(x = x[,i],
       main = "Histogram",
       xlab = colnames(x)[i])
}
```

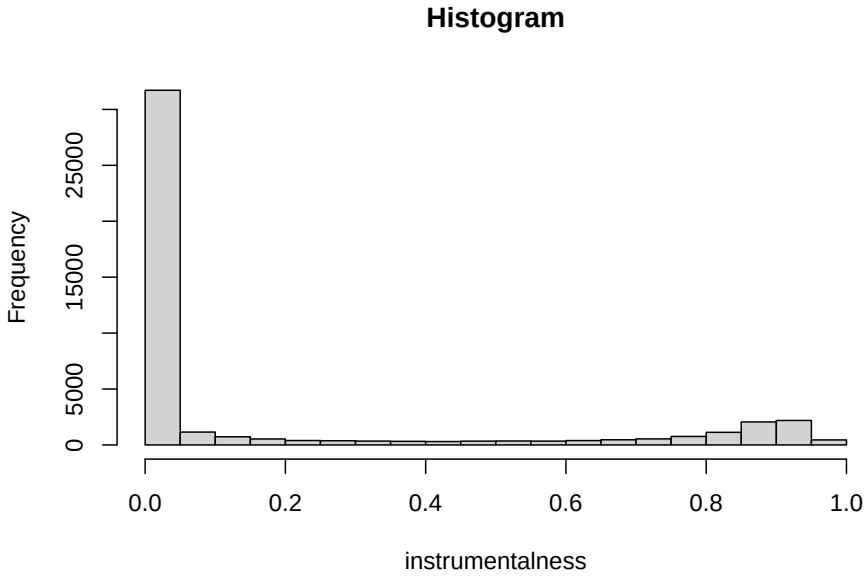


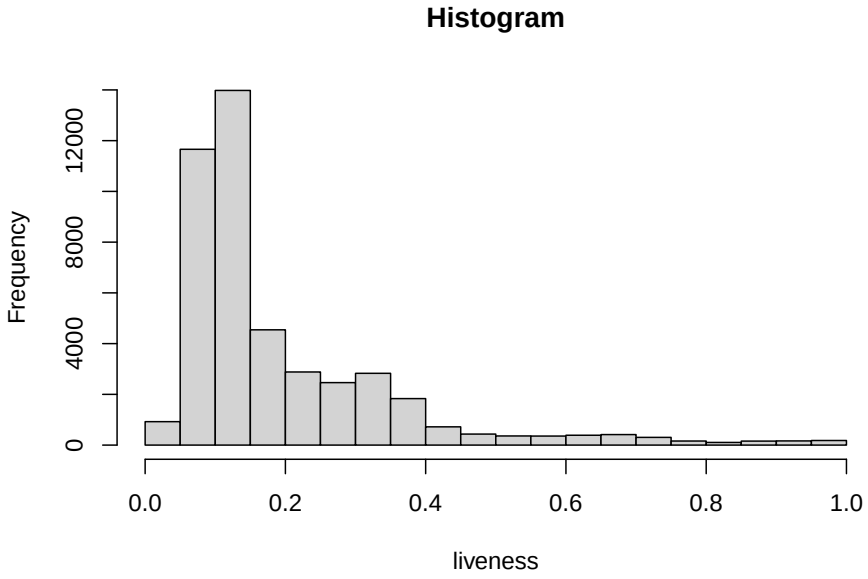

Histogram
Frequency
0
5000
10000
15000
0.0
0.2
0.4
0.6
0.8
1.0
acousticness

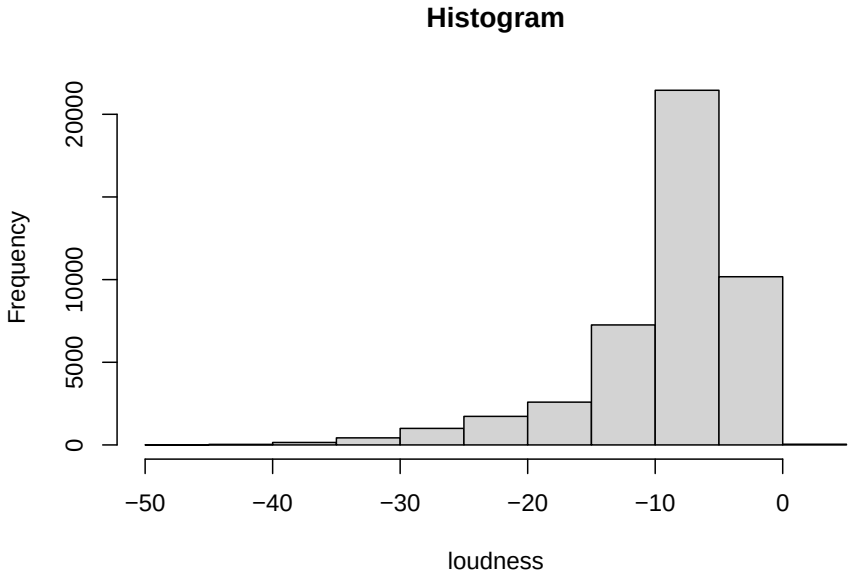

Histogram
Frequency
0
1000
3000
5000
0.2
0.4
0.6
0.8
1.0
danceability

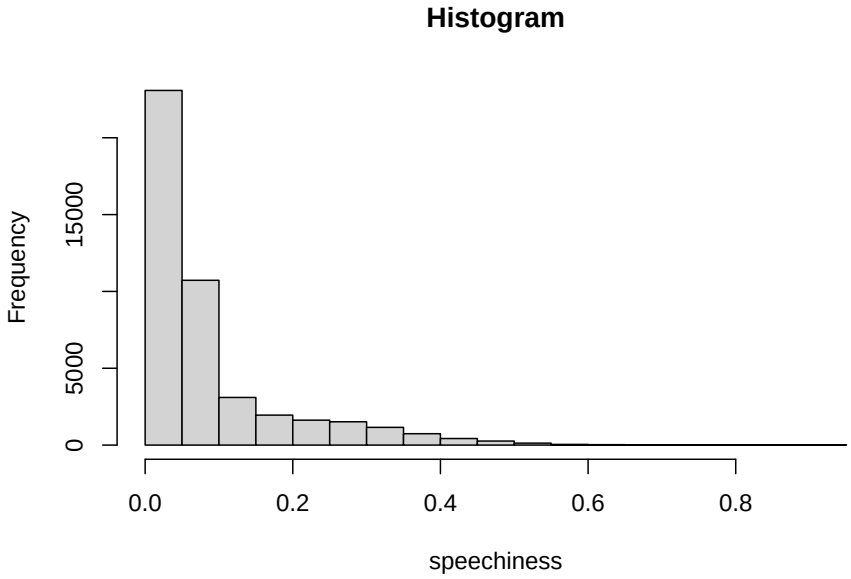

Histogram
Frequency
0
5000
15000
25000
0e+00
1e+06
2e+06
3e+06
4e+06
duration_ms

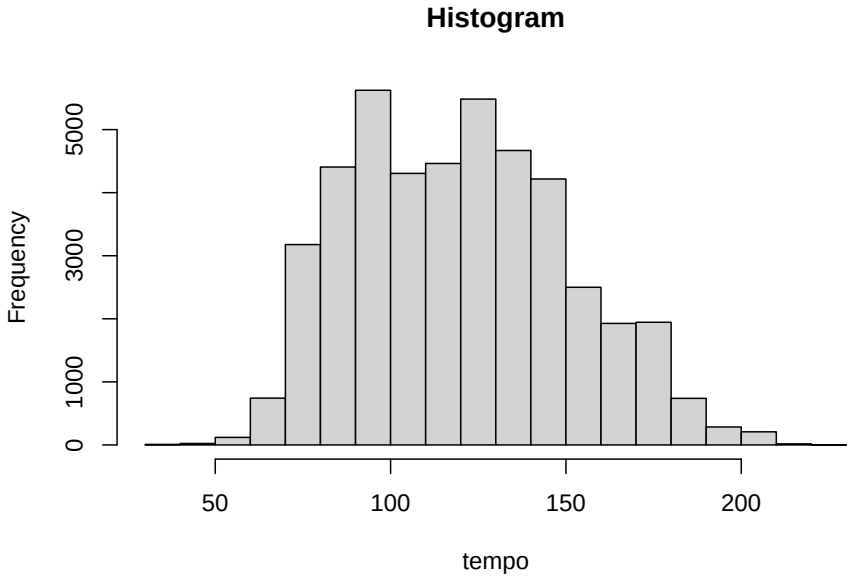

Histogram
Frequency
0
500
1500
2500
0.0
0.2
0.4
0.6
0.8
1.0
energy

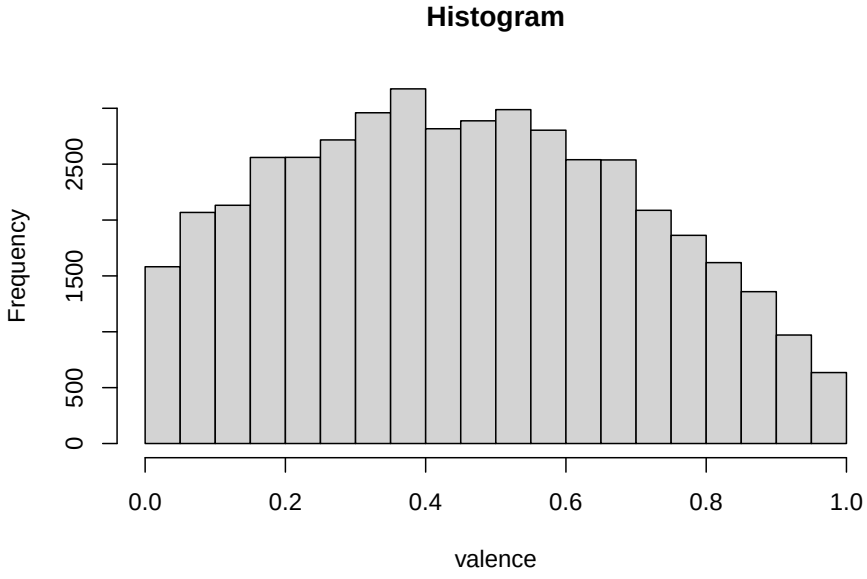

Histogram
Frequency
0
5000
15000
25000
0.0
0.2
0.4
0.6
0.8
1.0
instrumentalness


Histogram
Frequency
0
4000
8000
12000
0.0
0.2
0.4
0.6
0.8
1.0
liveness


Histogram
Frequency
0
5000
10000
20000
−50
−40
−30
−20
−10
0
loudness


Histogram
Frequency
0
5000
15000
0.0
0.2
0.4
0.6
0.8
speechiness





Table/barcharts

```
library(tidyverse)
```

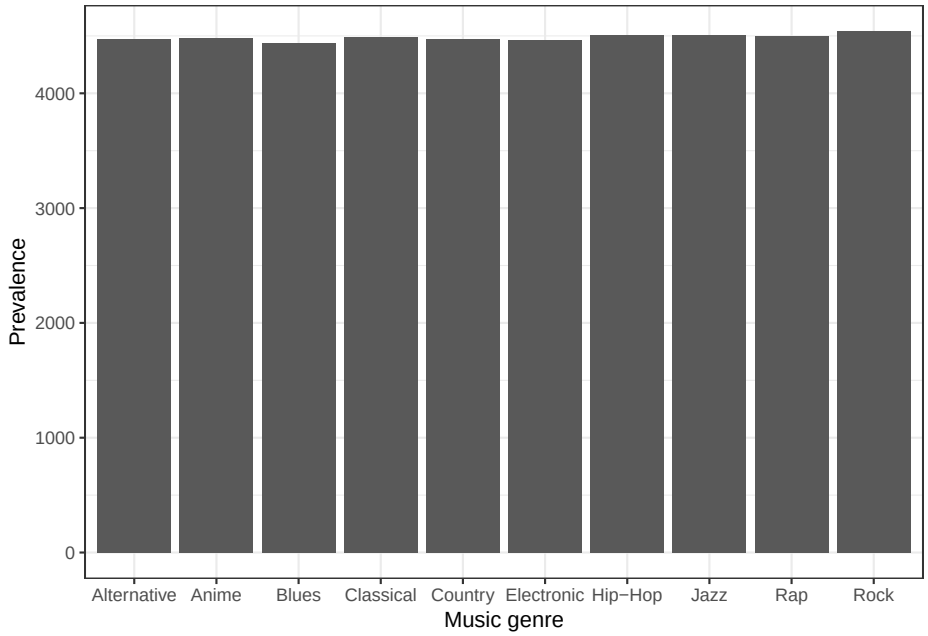
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
```

```
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
# Barchart: music_genre
tab1 <- table(musique$music_genre)
tab1[-1]
```

```
##
## Alternative       Anime       Blues   Classical     Country  Electronic
##        4473        4481        4436        4486        4477        4461
##     Hip-Hop        Jazz         Rap        Rock
##        4507        4507        4496        4539
```

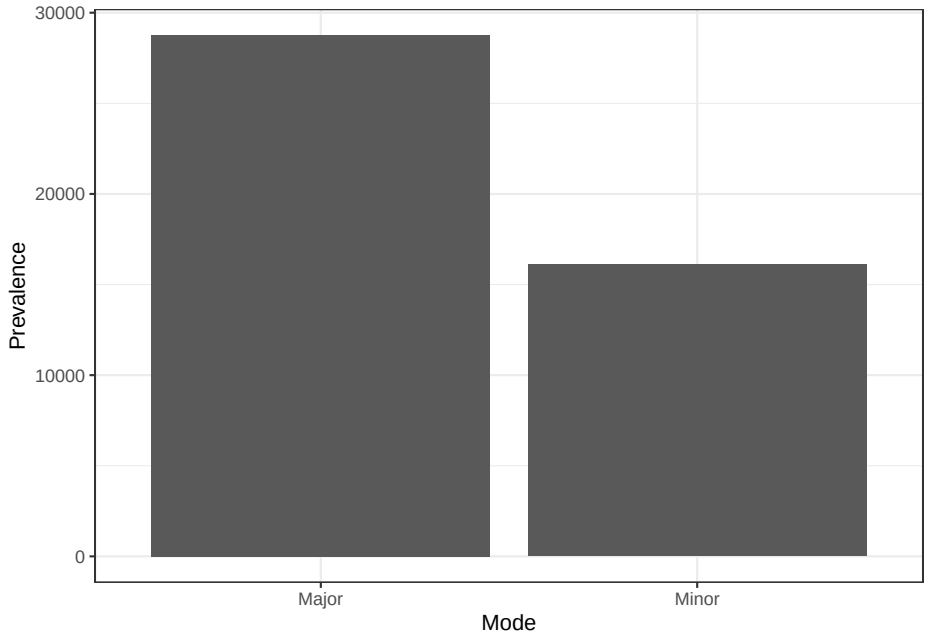
```
as.data.frame(tab1[-1]) %>% # must be a dataframe
  rename(Genre = Var1) %>%
  ggplot(aes(x = Genre, y = Freq)) +
  geom_bar(position = "dodge", stat = "identity") +
  scale_fill_viridis_d(option = "B", end = 0.8) +
  theme_bw() +
  labs(x = "Music genre",
       y = "Prevalence")
```




```
# Barchart: mode
tab2 <- table(musique$mode)
tab2[-1]
```

```
##
## Major Minor
## 28757 16106
```

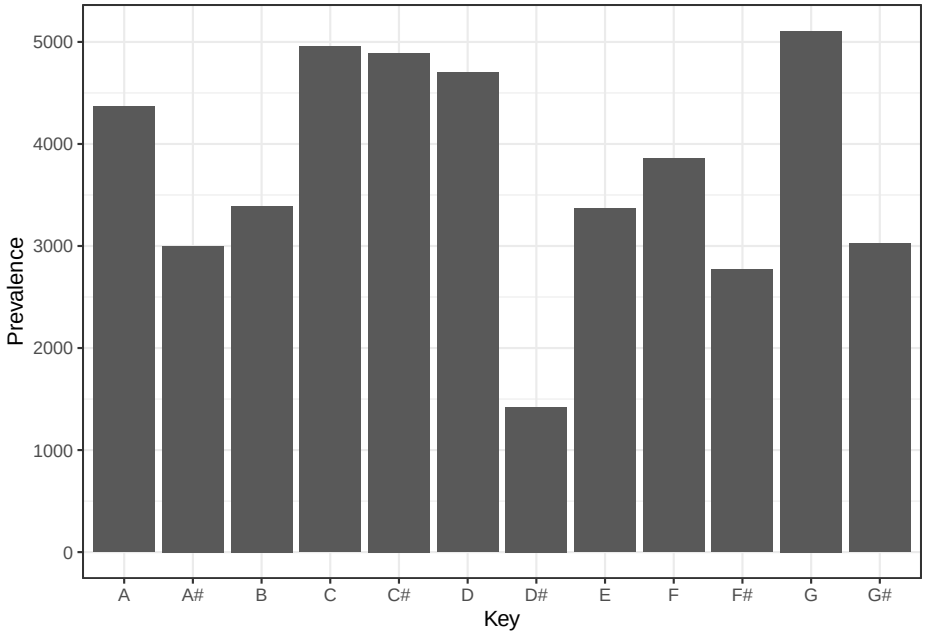
```r
as.data.frame(tab2[-1]) %>% # must be a dataframe
  rename(Mode = Var1) %>%
  ggplot(aes(x = Mode, y = Freq)) +
  geom_bar(position = "dodge", stat = "identity") +
  scale_fill_viridis_d(option = "B", end = 0.8) +
  theme_bw() +
  labs(x = "Mode",
       y = "Prevalence")
```




```r
# Barchart: key
tab3 <- table(musique$key)
tab3[-1]
```

```
##
##    A   A#    B    C   C#    D   D#    E    F   F#    G   G#
## 4371 3000 3391 4954 4891 4702 1422 3367 3861 2773 5107 3024
```

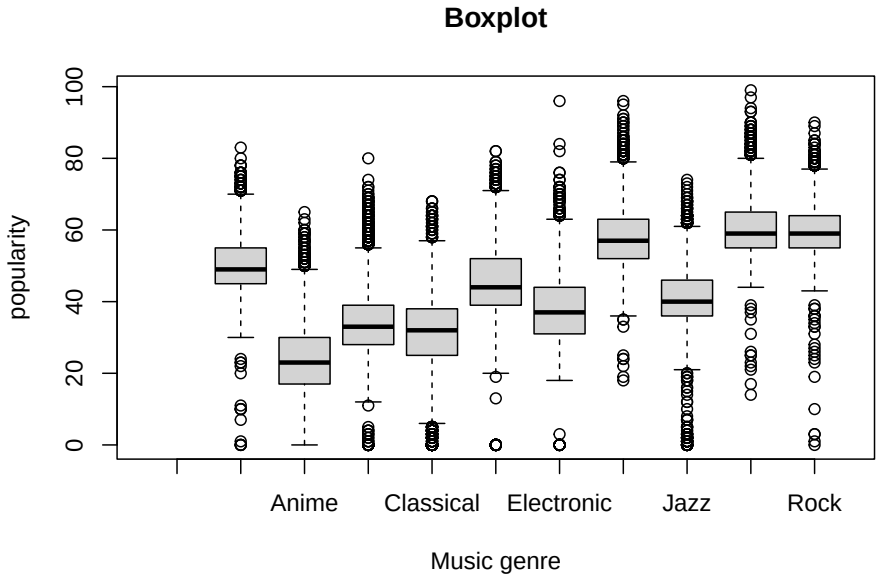
```r
as.data.frame(tab3[-1]) %>% # must be a dataframe
  rename(Key = Var1) %>%
  ggplot(aes(x = Key, y = Freq)) +
  geom_bar(position = "dodge", stat = "identity") +
  scale_fill_viridis_d(option = "B", end = 0.8) +
  theme_bw() +
  labs(x = "Key",
       y = "Prevalence")
```

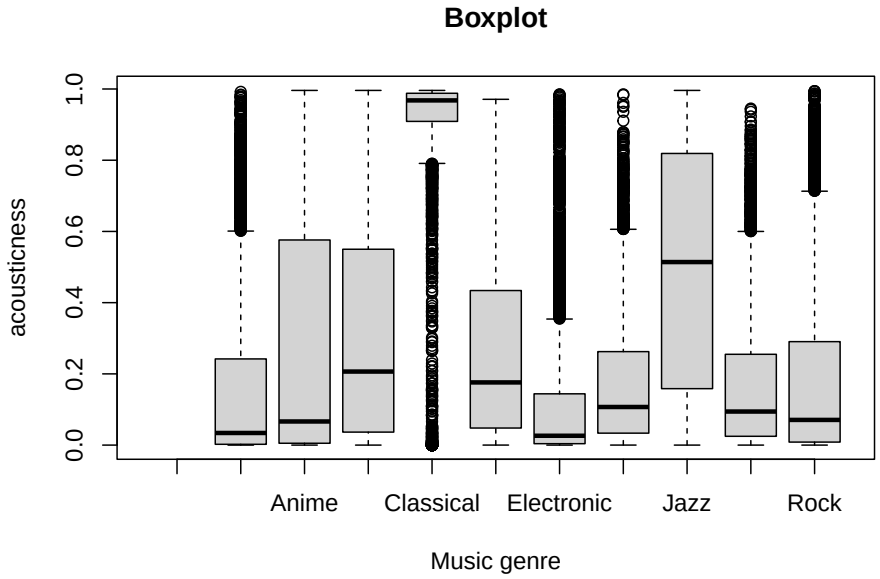


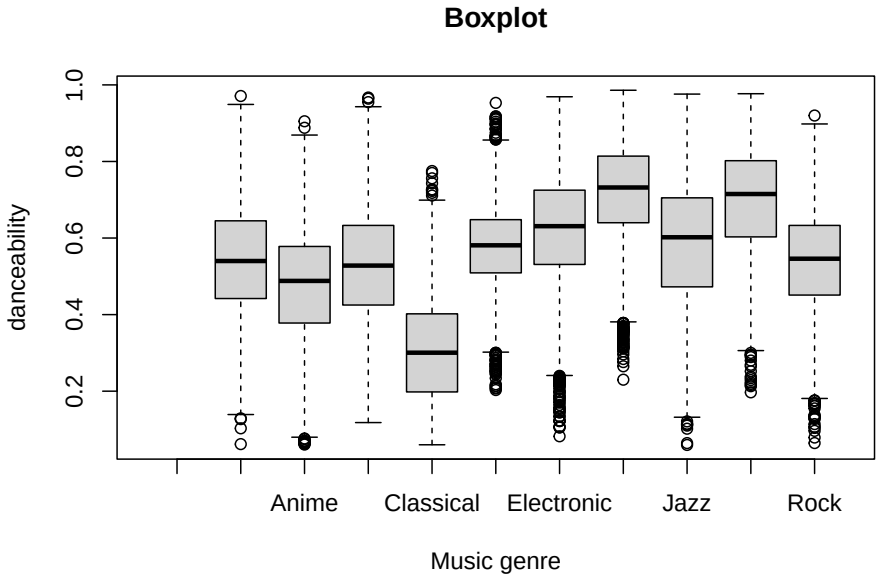
MULTIVARIATE

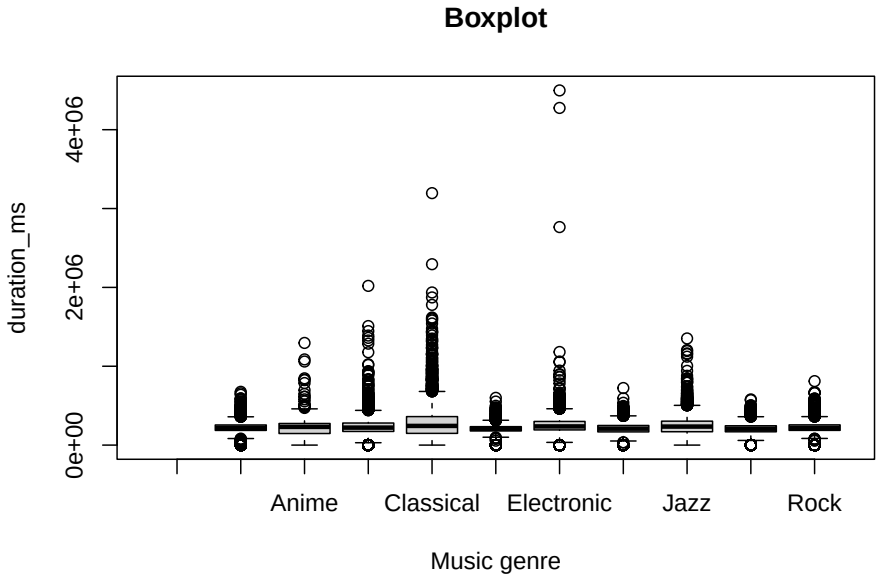
Boxplot

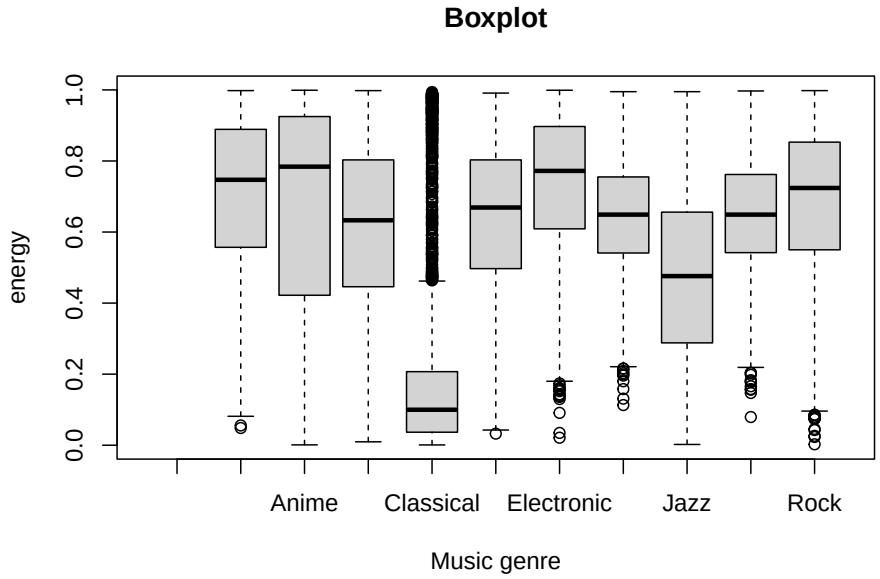
```
# Boxplot: continous variables vs music_genre
for(i in 1:11){
  boxplot(formula = x[,i] ~ musique$music_genre,
       main = "Boxplot",
       xlab = "Music genre",
       ylab = colnames(x)[i])
}
```

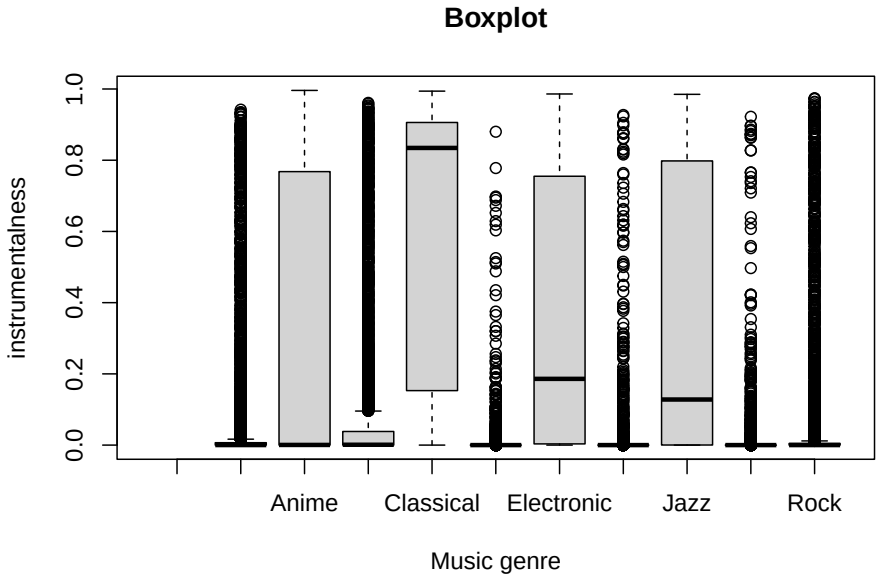

Boxplot
popularity
0
20
40
60
80
100
Anime
Classical
Electronic
Jazz
Rock
Music genre

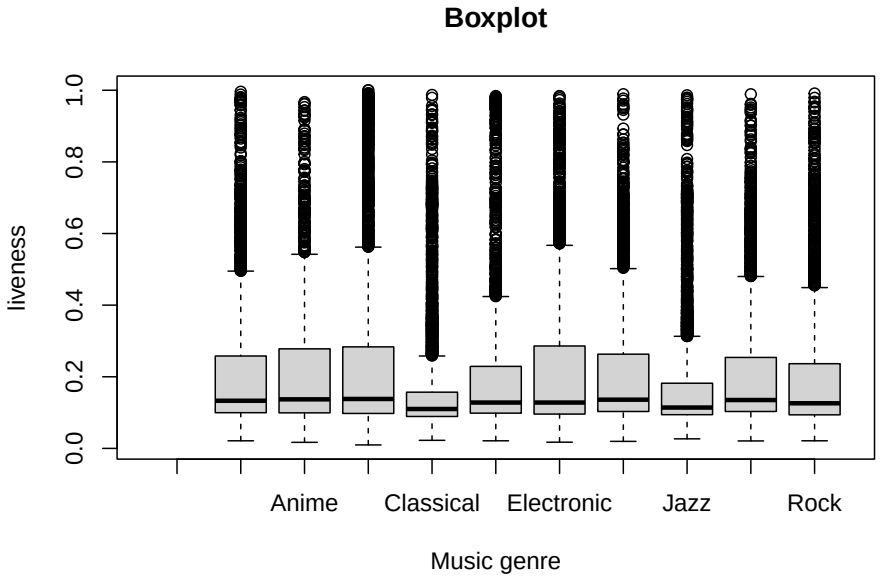

Boxplot
acousticness
0.0
0.2
0.4
0.6
0.8
1.0
Anime
Classical
Electronic
Jazz
Rock
Music genre

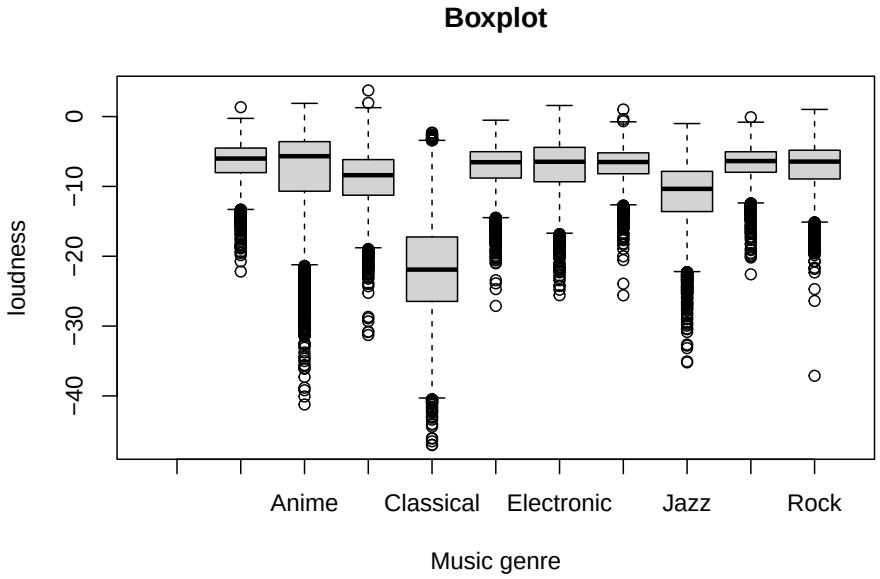

Boxplot
danceability
0.2
0.4
0.6
0.8
1.0
Anime
Classical
Electronic
Jazz
Rock
Music genre

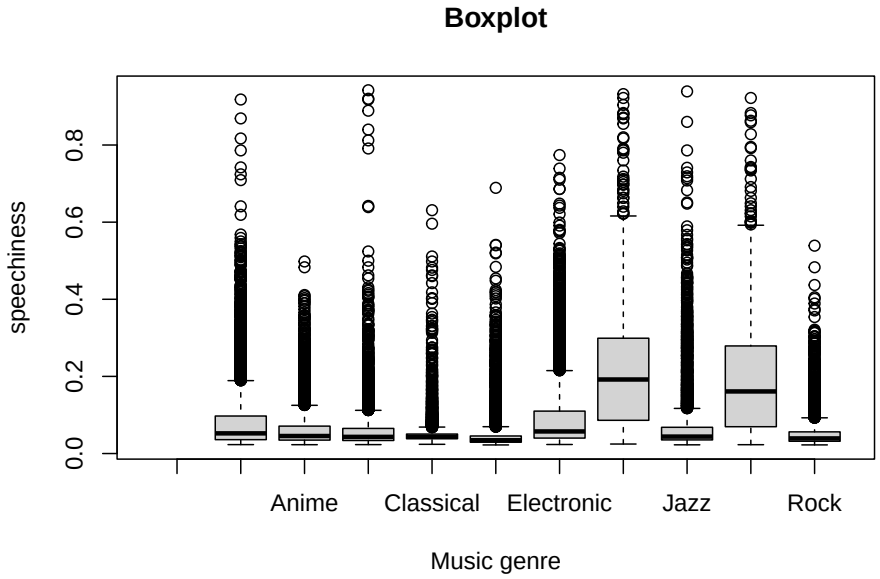

Boxplot
duration_ms
0e+00
2e+06
4e+06
Anime
Classical
Electronic
Jazz
Rock
Music genre

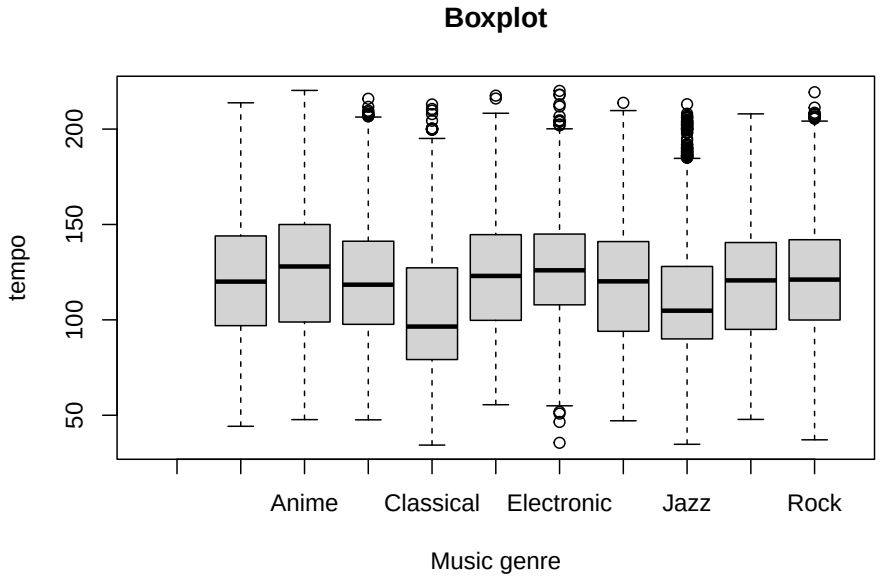

Boxplot
energy
0.0
0.2
0.4
0.6
0.8
1.0
Anime
Classical
Electronic
Jazz
Rock
Music genre

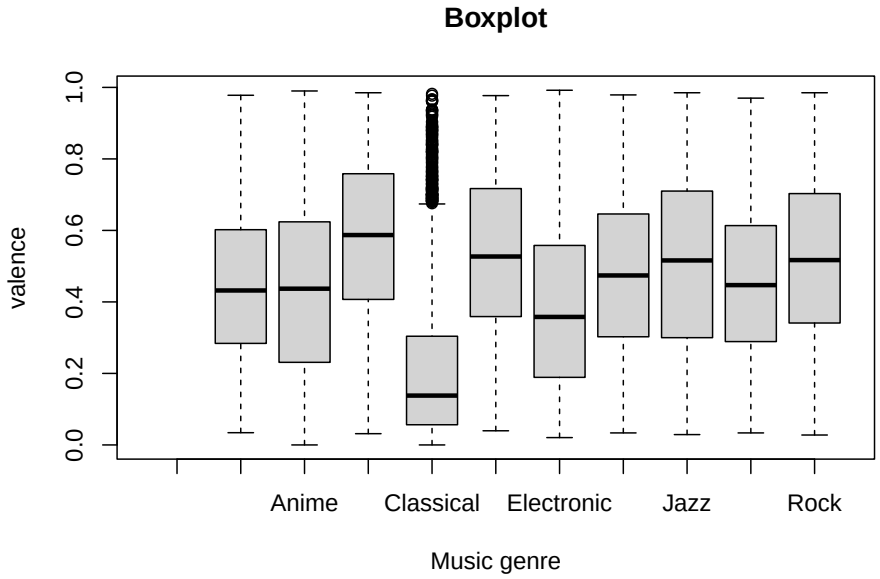

Boxplot
instrumentalness
0.0
0.2
0.4
0.6
0.8
1.0
Anime
Classical
Electronic
Jazz
Rock
Music genre


Boxplot
liveness
0.0
0.2
0.4
0.6
0.8
1.0
Anime
Classical
Electronic
Jazz
Rock
Music genre


Boxplot
loudness
−40
−30
−20
−10
0
Anime
Classical
Electronic
Jazz
Rock
Music genre


Boxplot
speechiness
0.0
0.2
0.4
0.6
0.8
Anime
Classical
Electronic
Jazz
Rock
Music genre


Boxplot
tempo
50
100
150
200
Anime
Classical
Electronic
Jazz
Rock
Music genre



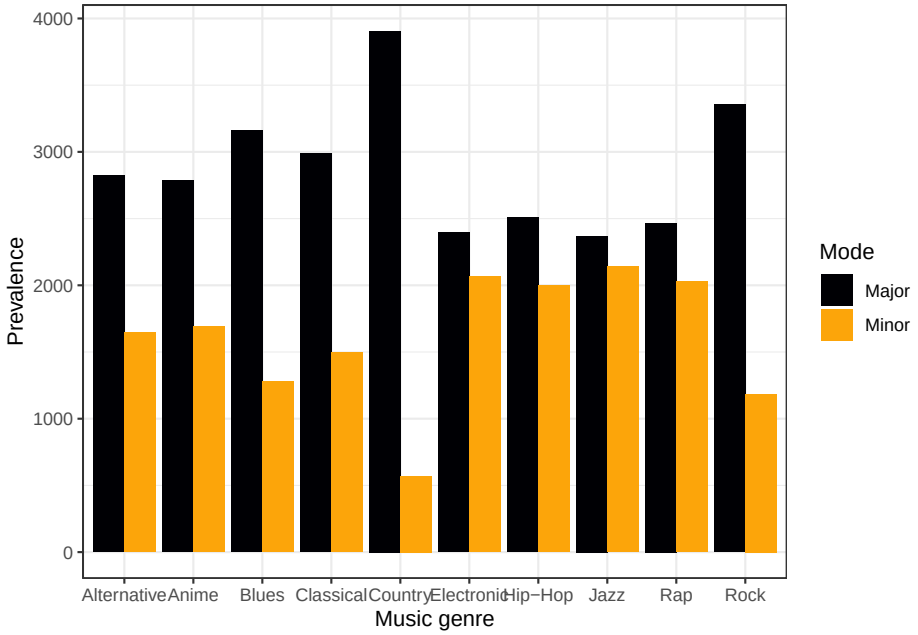
Table/barcharts

- Mode$major: gai, lumineux et ouvert;
- Mode$minor: triste, gris et mélancolique.

```
# Barchart: mode vs music_genre
tab4 <- table(musique$music_genre, musique$mode)
tab4[-1,2:3]
```

```
##
##                Major Minor
##   Alternative   2824  1649
##   Anime         2787  1694
##   Blues         3160  1276
##   Classical     2989  1497
##   Country       3906   571
##   Electronic    2394  2067
##   Hip-Hop       2508  1999
##   Jazz          2369  2138
##   Rap           2467  2029
##   Rock          3353  1186
```

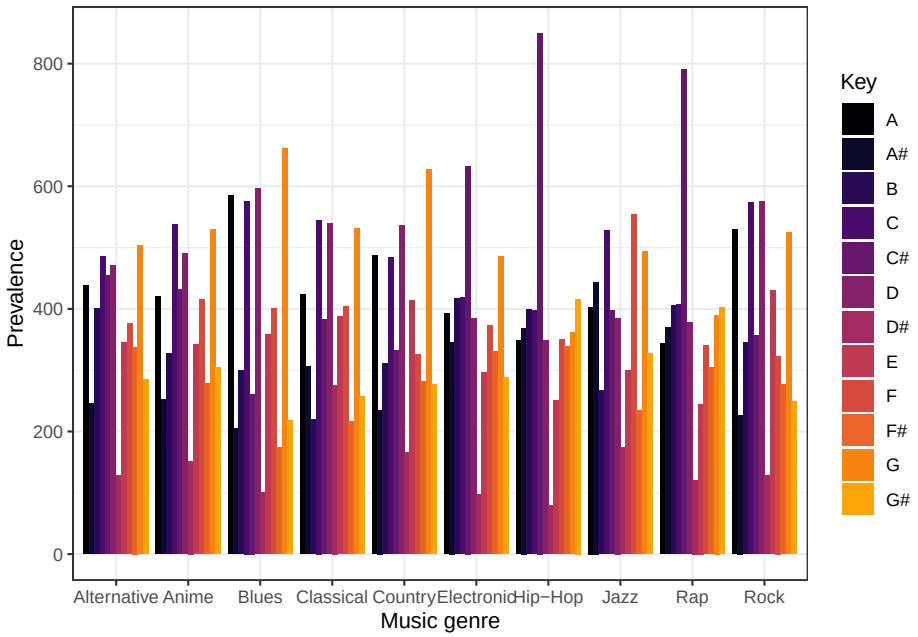
```
as.data.frame(tab4[-1,2:3]) %>% # must be a dataframe
  group_by(Var2) %>%
  rename(Genre = Var1,
         Mode = Var2) %>%
  ggplot(aes(x = Genre, y = Freq, fill = Mode)) + # colored by mode
  geom_bar(position = "dodge", stat = "identity") +
  scale_fill_viridis_d(option = "B", end = 0.8) +
  theme_bw() +
  labs(x = "Music genre",
       y = "Prevalence")
```




```
# Barchart: key vs music_genre
tab5 <- table(musique$music_genre, musique$key)
tab5[-1,-1]
```

```
##
##                  A  A#   B   C  C#   D  D#   E   F  F#   G  G#
##   Alternative 438 246 401 486 455 471 128 345 376 338 504 285
##   Anime       420 252 327 538 432 490 152 342 415 278 530 305
##   Blues       585 205 300 576 261 596 101 358 401 174 661 218
##   Classical   423 306 219 545 383 539 276 387 404 216 531 257
##   Country     487 235 311 484 332 536 166 414 325 282 628 277
##   Electronic  393 346 417 418 633 384  97 296 373 331 485 288
##   Hip-Hop     348 369 399 398 850 348  79 250 350 339 361 416
##   Jazz        403 444 267 528 398 385 174 300 554 234 493 327
##   Rap         344 370 405 407 790 378 121 245 340 304 390 402
##   Rock        530 227 345 574 357 575 128 430 323 277 524 249
```

```
as.data.frame(tab5[-1,-1]) %>% # must be a dataframe
  group_by(Var2) %>%
  rename(Genre = Var1,
         Key = Var2) %>%
  ggplot(aes(x = Genre, y = Freq, fill = Key)) + # colored by key
  geom_bar(position = "dodge", stat = "identity") +
  scale_fill_viridis_d(option = "B", end = 0.8) +
  theme_bw() +
  labs(x = "Music genre",
       y = "Prevalence")
```



Scatterplots

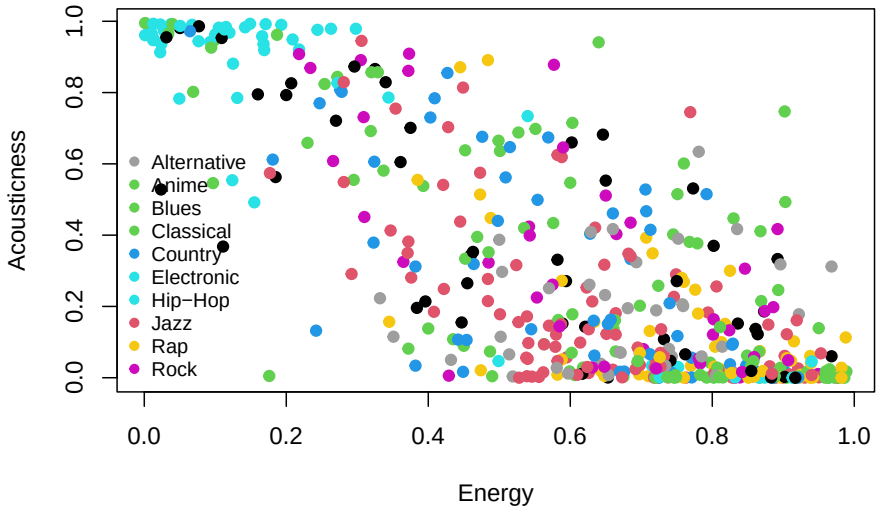
```
sample_ind = sample(1:nrow(musique),500)
musique_red = musique[sample_ind,]

abs(cor(musique_red[,c(4:9, 11:12, 14:15, 17)])) > 0.6
```

```
##                  popularity acousticness danceability duration_ms energy
## popularity             TRUE        FALSE        FALSE       FALSE  FALSE
## acousticness          FALSE         TRUE        FALSE       FALSE   TRUE
## danceability          FALSE        FALSE         TRUE       FALSE  FALSE
## duration_ms           FALSE        FALSE        FALSE        TRUE  FALSE
## energy                FALSE         TRUE        FALSE       FALSE   TRUE
## instrumentalness      FALSE        FALSE        FALSE       FALSE  FALSE
## liveness              FALSE        FALSE        FALSE       FALSE  FALSE
## loudness              FALSE         TRUE        FALSE       FALSE   TRUE
## speechiness           FALSE        FALSE        FALSE       FALSE  FALSE
## tempo                 FALSE        FALSE        FALSE       FALSE  FALSE
## valence               FALSE        FALSE        FALSE       FALSE  FALSE
##                  instrumentalness liveness loudness speechiness tempo valence
## popularity                  FALSE    FALSE    FALSE       FALSE FALSE   FALSE
## acousticness                FALSE    FALSE     TRUE       FALSE FALSE   FALSE
## danceability                FALSE    FALSE    FALSE       FALSE FALSE   FALSE
## duration_ms                 FALSE    FALSE    FALSE       FALSE FALSE   FALSE
## energy                      FALSE    FALSE     TRUE       FALSE FALSE   FALSE
## instrumentalness             TRUE    FALSE    FALSE       FALSE FALSE   FALSE
## liveness                    FALSE     TRUE    FALSE       FALSE FALSE   FALSE
## loudness                    FALSE    FALSE     TRUE       FALSE FALSE   FALSE
## speechiness                 FALSE    FALSE    FALSE        TRUE FALSE   FALSE
## tempo                       FALSE    FALSE    FALSE       FALSE  TRUE   FALSE
## valence                     FALSE    FALSE    FALSE       FALSE FALSE    TRUE
```

```
cor(musique_red$energy, musique_red$acousticness)
```

```
## [1] -0.7822474
```

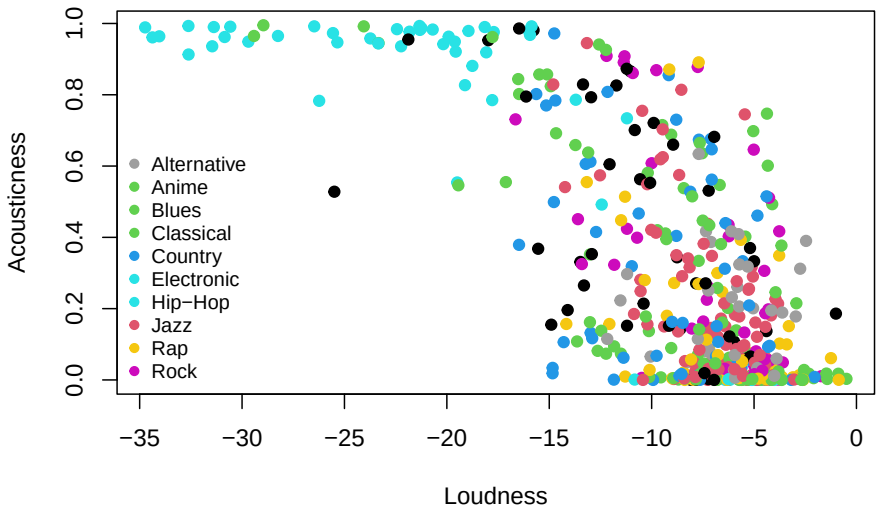
```
plot(musique_red$energy, musique_red$acousticness,
     col=musique_red$music_genre,
     xlab="Energy",ylab="Acousticness", pch=19)
legend("bottomleft",levels(musique_red$music_genre)[-1],
       col=musique_red$music_genre,pch=19, cex=0.8,bty="n")
```




```
cor(musique_red$loudness, musique_red$acousticness)
```

```
## [1] -0.7269124
```

```
plot(musique_red$loudness, musique_red$acousticness,
     col=musique_red$music_genre,
     xlab="Loudness",ylab="Acousticness", pch=19)
legend("bottomleft",levels(musique_red$music_genre)[-1],
       col=musique_red$music_genre,pch=19, cex=0.8,bty="n")
```




```
cor(musique_red$loudness, musique_red$energy)
```

```
## [1] 0.8355379
```

```
plot(musique_red$loudness, musique_red$energy,
     col=musique_red$music_genre,
     xlab="Loudness",ylab="Energy", pch=19)
legend("topleft",levels(musique_red$music_genre)[-1],
       col=musique_red$music_genre,pch=19, cex=0.8,bty="n")
```