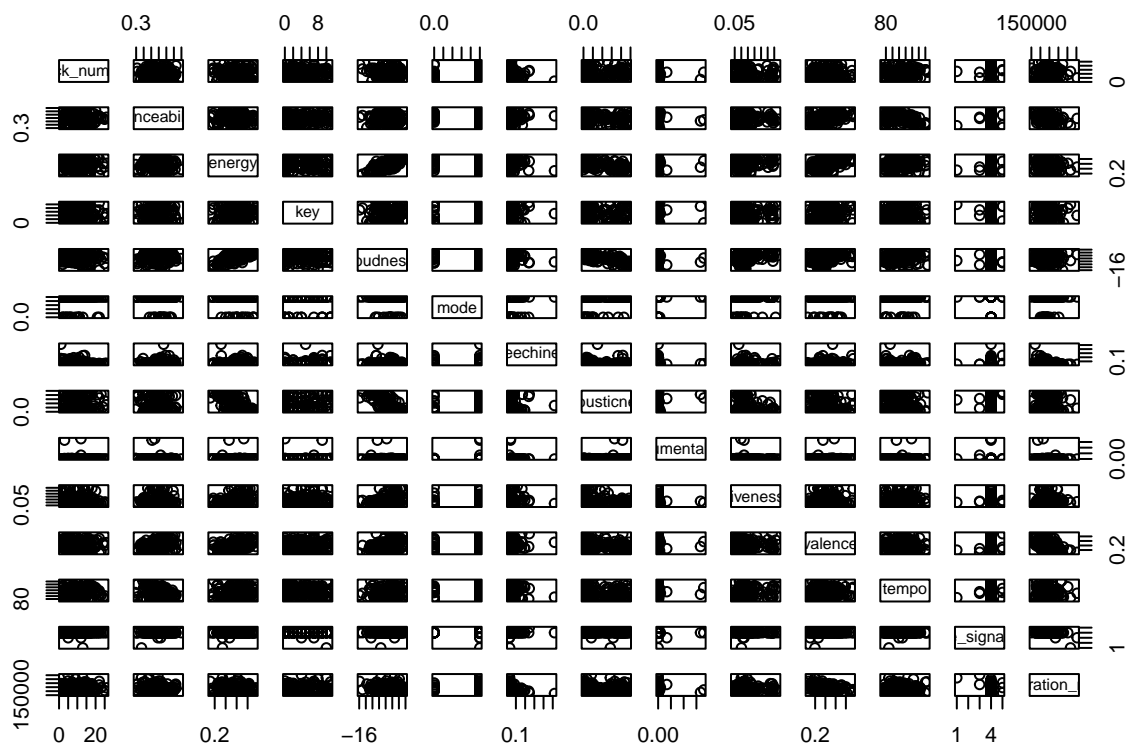# Rgroup Project

Project instructions

R Markdown

#Initialize

```
album_songs <- as.data.frame(taylor::taylor_album_songs)
all_songs <- as.data.frame(taylor::taylor_all_songs)
albums <- as.data.frame(taylor::taylor_albums)
```

Clean up data

```
x<-"lyrics"
y<-"key_name"
albumsongs_raw<-album_songs
albumsongs<-album_songs[,sapply(album_songs, is.numeric)]
# albumsongs = subset(taylor::taylor_album_songs, select = -c(lyrics) )
pairs(albumsongs)
```

```
cor(albumsongs)
```

```
##                 track_number danceability       energy          key
## track_number     1.00000000 -0.054580273 -0.125315481 -0.027174633
## danceability     -0.05458027  1.000000000  0.078777963  0.006267286
## energy           -0.12531548  0.078777963  1.000000000  0.061850353
## key              -0.02717463  0.006267286  0.061850353  1.000000000
## loudness         -0.21385680  0.043242112  0.782270708  0.069415465
## mode              0.07204943 -0.137084889 -0.108700832 -0.084001489
## speechiness      -0.13488221  0.155938800  0.038239268  0.115259661
## acousticness      0.15857765 -0.161002788 -0.691889642 -0.062206985
## instrumentalness -0.04358477 -0.050354882 -0.004066416  0.002704464
## liveness         -0.05858942 -0.049066384  0.241770124  0.151629277
## valence          -0.08128997  0.376213245  0.495620616  0.154996318
## tempo             0.13612146 -0.251625217  0.195689430 -0.108810640
## time_signature   -0.00807437  0.043788915  0.077593624 -0.081435837
## duration_ms       0.07801149 -0.279690687 -0.198340940 -0.074530139
##                     loudness         mode  speechiness acousticness
## track_number     -0.213856798  0.07204943 -0.134882207   0.15857765
## danceability      0.043242112 -0.13708489  0.155938800  -0.16100279
## energy            0.782270708 -0.10870083  0.038239268  -0.69188964
## key               0.069415465 -0.08400149  0.115259661  -0.06220699
## loudness          1.000000000 -0.05446063 -0.106734905  -0.75748204
## mode             -0.054460629  1.00000000 -0.151223008   0.09397666
## speechiness      -0.106734905 -0.15122301  1.000000000  -0.02000852
```

```
## acousticness       -0.757482043  0.09397666 -0.020008518   1.00000000
## instrumentalness  -0.100171458  0.03847064 -0.031232603   0.15012147
## liveness           0.275375513  0.06273754  0.004533237  -0.21991050
## valence            0.331714155 -0.05343266  0.162702472  -0.22306075
## tempo              0.178689935  0.05622053  0.031796714  -0.12299185
## time_signature    -0.001130676 -0.01429885  0.110445072   0.04902469
## duration_ms       -0.089420027  0.03003156 -0.317570919   0.10380429
##                    instrumentalness     liveness     valence       tempo
## track_number            -0.043584767 -0.058589424 -0.08128997  0.13612146
## danceability            -0.050354882 -0.049066384  0.37621324 -0.25162522
## energy                  -0.004066416  0.241770124  0.49562062  0.19568943
## key                      0.002704464  0.151629277  0.15499632 -0.10881064
## loudness                -0.100171458  0.275375513  0.33171416  0.17868993
## mode                     0.038470639  0.062737539 -0.05343266  0.05622053
## speechiness             -0.031232603  0.004533237  0.16270247  0.03179671
## acousticness             0.150121467 -0.219910495 -0.22306075 -0.12299185
## instrumentalness         1.000000000 -0.055719234  0.02173494  0.04066497
## liveness                -0.055719234  1.000000000  0.01015398  0.09398635
## valence                  0.021734944  0.010153982  1.00000000 -0.04097463
## tempo                    0.040664975  0.093986352 -0.04097463  1.00000000
## time_signature          -0.028750117  0.019275167  0.14902619  0.05657931
## duration_ms             -0.090320475 -0.058541564 -0.44128386  0.04007797
##                    time_signature duration_ms
## track_number          -0.008074370  0.07801149
## danceability           0.043788915 -0.27969069
## energy                 0.077593624 -0.19834094
## key                   -0.081435837 -0.07453014
## loudness              -0.001130676 -0.08942003
## mode                  -0.014298852  0.03003156
## speechiness            0.110445072 -0.31757092
## acousticness           0.049024689  0.10380429
## instrumentalness      -0.028750117 -0.09032047
## liveness               0.019275167 -0.05854156
## valence                0.149026188 -0.44128386
## tempo                  0.056579313  0.04007797
## time_signature         1.000000000 -0.31759712
## duration_ms           -0.317597120  1.00000000
```

summary(albumsongs)

```
##   track_number      danceability        energy             key
##  Min.   : 1.000   Min.   :0.2920   Min.   :0.1180   Min.    : 0.000
##  1st Qu.: 5.000   1st Qu.:0.5155   1st Qu.:0.4695   1st Qu.: 2.000
##  Median :10.000   Median :0.5940   Median :0.6090   Median : 5.000
##  Mean   : 9.877   Mean   :0.5867   Mean   :0.5909   Mean    : 4.601
##  3rd Qu.:14.000   3rd Qu.:0.6520   3rd Qu.:0.7320   3rd Qu.: 7.000
##  Max.   :26.000   Max.   :0.8970   Max.   :0.9500   Max.    :11.000
##     loudness           mode         speechiness       acousticness
##  Min.   :-15.910   Min.   :0.0000   Min.   :0.02310   Min.    :0.000191
##  1st Qu.: -8.828   1st Qu.:1.0000   1st Qu.:0.03030   1st Qu.:0.032950
##  Median : -6.719   Median :1.0000   Median :0.03720   Median :0.156000
##  Mean   : -7.215   Mean   :0.8957   Mean   :0.05419   Mean    :0.313234
##  3rd Qu.: -5.352   3rd Qu.:1.0000   3rd Qu.:0.05445   3rd Qu.:0.648500
##  Max.   : -2.098   Max.   :1.0000   Max.   :0.51900   Max.    :0.971000
```

```
##  instrumentalness       liveness            valence           tempo
##  Min.   :0.000e+00   Min.   :0.03570   Min.   :0.0499   Min.   : 68.53
##  1st Qu.:0.000e+00   1st Qu.:0.09295   1st Qu.:0.2795   1st Qu.: 96.54
##  Median :2.210e-06   Median :0.11400   Median :0.4160   Median :122.08
##  Mean   :2.603e-03   Mean   :0.14064   Mean   :0.4201   Mean   :124.94
##  3rd Qu.:6.610e-05   3rd Qu.:0.15400   3rd Qu.:0.5410   3rd Qu.:148.01
##  Max.   :1.790e-01   Max.   :0.38200   Max.   :0.9420   Max.   :207.48
##  time_signature   duration_ms
##  Min.   :1.000   Min.   :150440
##  1st Qu.:4.000   1st Qu.:211833
##  Median :4.000   Median :234147
##  Mean   :3.988   Mean   :237338
##  3rd Qu.:4.000   3rd Qu.:253568
##  Max.   :5.000   Max.   :403920
```

Strong relation between Energy and loudness, energy and accustoics, energy and valence, energy and tempo

Lyrics Analysis

```
while (length(ids) > 0) {
  for (id in ids) {
    tryCatch({
      allLyrics <- rbind(get_lyrics_id(id), allLyrics)
      successful <- unique(allLyrics$song_id)
      ids <- ids[!ids %in% successful]
      print(paste("done - ", id))
      print(paste("New length is ", length(ids)))
    }, error = function(e){})
  }
}


# Create empty dataframe to house them
allLyrics <- data.frame()
for (id in ids) {
  allLyrics <- rbind(get_lyrics_id(id), allLyrics)
}



```r
#most of the libraries needed
library(dplyr) #data manipulation

##
## 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
```

```r
library(ggplot2) #visualizations
library(gridExtra) #viewing multiple plots together
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##      combine
```

```r
library(tidytext) #text mining
```

```
## Warning: package 'tidytext' was built under R version 4.1.2
```

```r
library(wordcloud2) #creative visualizations
```

```
## Warning: package 'wordcloud2' was built under R version 4.1.2
```