Final Project

Majorz

```
library(tidyverse)
library(tidymodels)
library(glmnet)
spotify <- read_csv("data/tf_mini.csv")
```

```
lm_all_popularity <- lm(us_popularity_estimate ~ duration + release_year + acousticness +
     beat_strength + bounciness + danceability + dyn_range_mean + energy +
     flatness + instrumentalness + key + liveness + loudness + mechanism +
     mode + organism + speechiness + tempo + time_signature + valence,
     data = spotify)
summary(lm_all_popularity)
```

```
Call:
lm(formula = us_popularity_estimate ~ duration + release_year +
    acousticness + beat_strength + bounciness + danceability +
    dyn_range_mean + energy + flatness + instrumentalness + key +
    liveness + loudness + mechanism + mode + organism + speechiness +
    tempo + time_signature + valence, data = spotify)

Residuals:
    Min      1Q  Median      3Q     Max
-9.5072  0.0474  0.3386  0.4694  1.8199

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       9.958e+01  1.031e+00  96.540  < 2e-16 ***
duration          5.881e-05  7.183e-05   0.819 0.412996
release_year      6.490e-04  5.060e-04   1.283 0.199672
acousticness      2.527e-01  6.647e-02   3.802 0.000144 ***
```

```
beat_strength    -2.685e-01  1.977e-01  -1.358 0.174358
bounciness        3.130e-01  2.609e-01   1.200 0.230321
danceability      1.830e-01  8.479e-02   2.158 0.030927 *
dyn_range_mean   -3.908e-02  1.011e-02  -3.866 0.000111 ***
energy           -5.313e-01  5.627e-02  -9.443  < 2e-16 ***
flatness         -1.496e-01  1.777e-01  -0.842 0.399619
instrumentalness -2.975e-01  2.774e-02 -10.726  < 2e-16 ***
key              -2.024e-03  1.411e-03  -1.435 0.151408
liveness         -1.856e-01  3.044e-02  -6.099 1.08e-09 ***
loudness          3.914e-02  2.283e-03  17.146  < 2e-16 ***
mechanism        -1.823e-01  1.062e-01  -1.715 0.086279 .
modeminor         1.389e-02  1.073e-02   1.295 0.195287
organism         -6.483e-01  1.563e-01  -4.149 3.34e-05 ***
speechiness       2.505e-01  5.194e-02   4.823 1.42e-06 ***
tempo            -3.939e-04  2.358e-04  -1.671 0.094793 .
time_signature    7.404e-03  1.270e-02   0.583 0.559833
valence          -2.702e-01  2.665e-02 -10.139  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.127 on 50683 degrees of freedom
Multiple R-squared:  0.0224,    Adjusted R-squared:  0.02202
F-statistic: 58.07 on 20 and 50683 DF,  p-value: < 2.2e-16
```

```r
spotify_mode <- spotify |>
  mutate(new_mode = if_else(mode == "major", 1, 0),
         new_mode = as.numeric(new_mode))

spotify_mode |> drop_na(new_mode)
```

```
# A tibble: 50,704 x 31
   track_id      durat~1 relea~2 us_po~3 acous~4 beat_~5 bounc~6 dance~7 dyn_r~8
   <chr>           <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
 1 t_a540e552-1~    110.    1950   100.    0.458   0.519   0.505   0.400    7.51
 2 t_67965da0-1~    188.    1950   100.    0.916   0.419   0.546   0.491    9.10
 3 t_0614ecd3-a~    161.    1951    99.6   0.813   0.426   0.508   0.492    8.37
 4 t_070a63a0-7~    175.    1951    99.7   0.397   0.401   0.360   0.552    5.97
 5 t_d6990e17-9~    370.    1951   100.    0.729   0.371   0.335   0.483    5.80
 6 t_fcb90952-0~    178.    1951   100.    0.186   0.549   0.579   0.744    8.67
 7 t_20675f8a-3~    166.    1952   100.    0.519   0.592   0.640   0.741    9.53
 8 t_7577ca53-5~    198.    1952    99.5   0.787   0.472   0.448   0.427    6.91
```

```
 9 t_8a461a4e-6~    215.    1954   100.    0.155   0.526   0.566   0.523    8.63
10 t_ae523005-8~    281.    1954    97.4   0.941   0.233   0.209   0.242    4.83
# ... with 50,694 more rows, 22 more variables: energy <dbl>, flatness <dbl>,
#   instrumentalness <dbl>, key <dbl>, liveness <dbl>, loudness <dbl>,
#   mechanism <dbl>, mode <chr>, organism <dbl>, speechiness <dbl>,
#   tempo <dbl>, time_signature <dbl>, valence <dbl>, acoustic_vector_0 <dbl>,
#   acoustic_vector_1 <dbl>, acoustic_vector_2 <dbl>, acoustic_vector_3 <dbl>,
#   acoustic_vector_4 <dbl>, acoustic_vector_5 <dbl>, acoustic_vector_6 <dbl>,
#   acoustic_vector_7 <dbl>, new_mode <dbl>, and abbreviated variable names ...
```

```r
lm_all_mode <- glm(new_mode ~ us_popularity_estimate + duration + release_year + acousticn
     beat_strength + bounciness + danceability + dyn_range_mean + energy +
     flatness + instrumentalness + key + liveness + loudness + mechanism +
       organism + speechiness + tempo + time_signature + valence,
     data = spotify_mode,
     family = "binomial")
summary(lm_all_mode)
```

```
Call:
glm(formula = new_mode ~ us_popularity_estimate + duration +
    release_year + acousticness + beat_strength + bounciness +
    danceability + dyn_range_mean + energy + flatness + instrumentalness +
    key + liveness + loudness + mechanism + organism + speechiness +
    tempo + time_signature + valence, family = "binomial", data = spotify_mode)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3569  -1.2543   0.7625   0.9493   1.8185

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)            32.2683808  2.3096693  13.971  < 2e-16 ***
us_popularity_estimate -0.0112941  0.0085642  -1.319 0.187249
duration               -0.0008868  0.0001370  -6.472 9.68e-11 ***
release_year           -0.0145826  0.0010562 -13.807  < 2e-16 ***
acousticness            0.4800550  0.1339125   3.585 0.000337 ***
beat_strength           2.3227249  0.3798220   6.115 9.64e-10 ***
bounciness             -4.2116774  0.5087117  -8.279  < 2e-16 ***
danceability            0.2508033  0.1611182   1.557 0.119556
dyn_range_mean          0.1188409  0.0200062   5.940 2.85e-09 ***
```

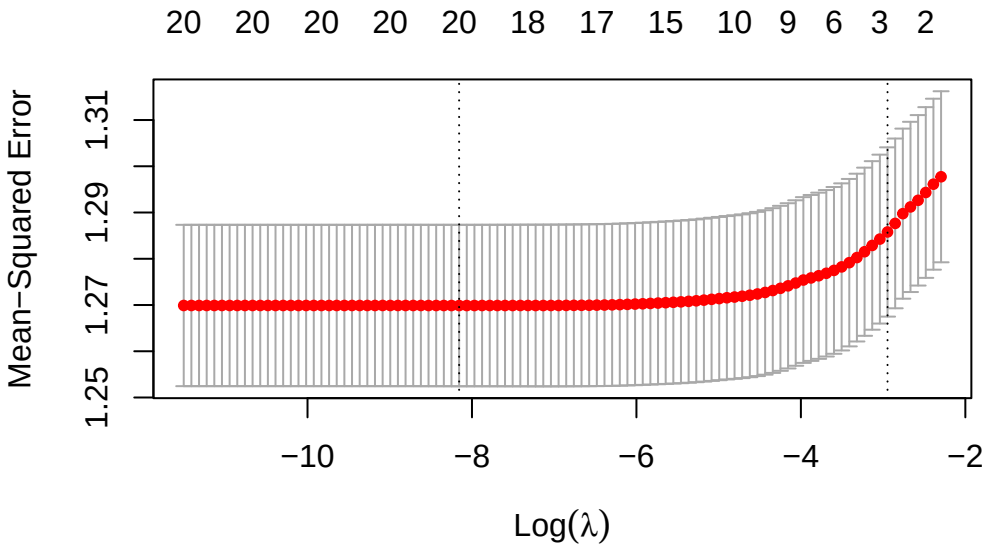
```
energy                -0.5804580  0.1072094  -5.414 6.15e-08 ***
flatness               0.7082200  0.3348900   2.115 0.034448 *
instrumentalness      -0.3421403  0.0522757  -6.545 5.95e-11 ***
key                   -0.0930592  0.0026793 -34.733  < 2e-16 ***
liveness               0.3261005  0.0588139   5.545 2.95e-08 ***
loudness               0.0223914  0.0043966   5.093 3.53e-07 ***
mechanism             -0.8263282  0.2122943  -3.892 9.93e-05 ***
organism              -0.3927748  0.3168700  -1.240 0.215144
speechiness           -1.0627013  0.0967583 -10.983  < 2e-16 ***
tempo                  0.0027563  0.0004504   6.120 9.37e-10 ***
time_signature        -0.2081995  0.0260103  -8.005 1.20e-15 ***
valence                0.5394631  0.0506272  10.656  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 66141  on 50703  degrees of freedom
Residual deviance: 63327  on 50683  degrees of freedom
AIC: 63369

Number of Fisher Scoring iterations: 4
```

```r
plot(lasso_sc)
```


20 20 20 20 20 18 17 15 10 9 6 3 2
Mean−Squared Error
1.25
1.27
1.29
1.31
−10
−8
−6
−4
−2
Log(λ)