

A Bayesian Analysis of Spotify Data

Nathaniel Maxwell, Jessie Bierschenk

30 April, 2021

Introduction

For many musicians, the art of composing/performing/marketing a new song is an arduous process. Even after all the work has been completed and a song is ready to be played to the public, the biggest uncertainty still awaits: How will the song be received? Will it become a hit? Will it be a song that everyone skips over, or never becomes popular? The purpose of this analysis is to investigate which characteristics of a song (such as tempo, duration, mode, acousticness, etc.) would make it more “likeable,” less likely to be skipped, or more popular. Of course, music taste is a very subjective matter, and thus, there will be quite a bit of uncertainty around any variables that are deemed important/unimportant. What one person likes; another person may dislike. Therefore, looking at such musical characteristics through a Bayesian lens will help to quantify the uncertainty surrounding any of our findings. Through this analysis we hope to provide some conclusions that an aspiring musician (or even a well-established musician) can have at their disposal when creating new music.

Pre-Analysis

Data

Two datasets were utilized during this analysis.

1. The first dataset consists of 83,939 observations on Spotify of whether or not a track was skipped by users. In total, 65,417 different tracks were included in the dataset. Each track has the following characteristics:
 - (a) Release Year (Year the song was released)
 - (b) Duration (length of song in seconds)
 - (c) US Popularity Estimate (A popularity rating of song, on a scale 1-100)
 - (d) Acousticness (A confidence measure from 0-1 on whether the track is acoustic, where values near 1 represent high confidence that the track is acoustic)
 - (e) Beat Strength (The strength of the beat from 0-1, where 1 represents a very strong sense of beat)
 - (f) Bounciness (A rating of the bounciness from 0-1, where 1 represents a strong sense of bounciness)
 - (g) Danceability (A rating from 0-1 of how suitable the track is for dancing, where values near 1 represent high suitability)
 - (h) Energy (A rating from 0-1 representing a perceptual measure of intensity and activity, where values near 1 represent high energy)
 - (i) Instrumentalness (A rating from 0-1 that predicts whether a track has no vocals, where values close to 1 represent high confidence that there are no vocals)

- (j) Mode (Predicts whether or not a song is major or minor)
- (k) Speechiness (A rating from 0-1 that detects the presence of spoken words in a track, with values near 1 representing an exclusively speech-like track)
- (l) Tempo (The estimated tempo of the track in Beats Per Minute (BPM))
- (m) Valence (A rating from 0-1 that represents the positivity of the song, with 1 representing high positivity)
- (n) Skipped (Denotes whether or not that particular track was skipped or played the entire way through)

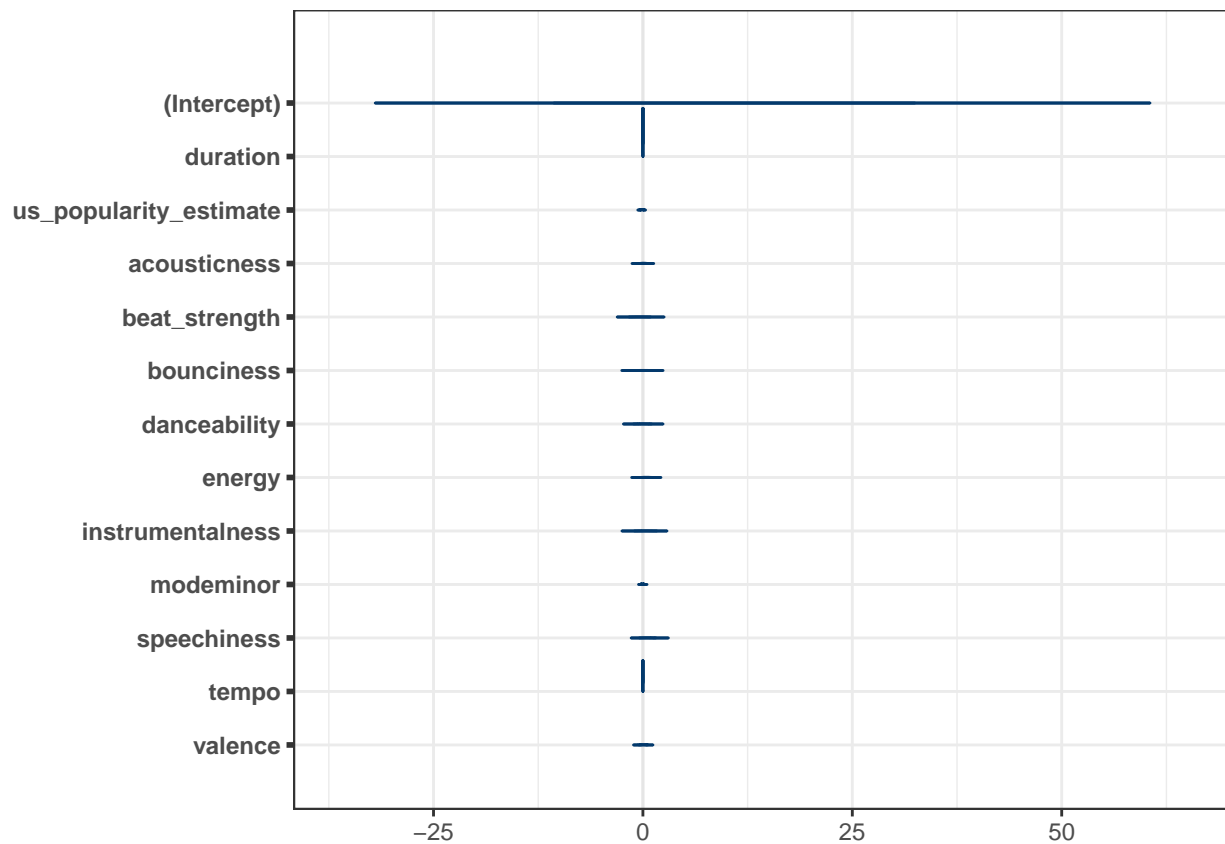
Note: in order to try to obtain tracks most representative of new music, only the following tracks were kept:

- (a) Tracks from 2010-present
 - (b) Tracks with a speechiness value ≤ 0.4 (filters out tracks that are mostly spoken, such as podcasts and ebooks)
 - (c) Tracks with an instrumentalness value ≤ 0.6 (filters out tracks that contain no vocals)
 - (d) Tracks with a duration ≤ 360 seconds (given that the average new song is 3-5 minutes, a cutoff of 6 minutes seemed appropriate)
2. The second dataset consisted of 2017 songs compiled by a single person, where a portion of the songs are songs that he likes, and the other portion are songs that he dislikes. This dataset includes similar variables as the first dataset, including:
- (a) Acousticness
 - (b) Danceability
 - (c) Duration
 - (d) Energy
 - (e) Instrumentalness
 - (f) Key (The particular grouping of chords and notes in a song)
 - (g) Liveness (rating from 0-1 of whether the track was performed live, with 1 representing high confidence the track was performed live)
 - (h) Loudness (Overall loudness of the track in decibels (dB))
 - (i) Mode
 - (j) Speechiness
 - (k) Tempo
 - (l) Time Signature (The way in which beats of the song are organized)
 - (m) Valence

Model Selection

For the first dataset, we wanted to estimate the values of the coefficients β for each of the variables to find out how they impact whether or not a track is skipped. We are assuming little knowledge about each variable's effect, so we propose a weakly informative prior for $[\beta]$: Using recommendations from Gelman, Jakulin, Pittau, and Su, we use a cauchy(0,2.5) prior for each scaled variable. Our response variable, \mathbf{Y} , will follow a logistic regression model, where 1 means the track was skipped. This is equivalent to the Bernoulli distribution $\mathbf{Y}|\theta \sim \text{Bern}(\theta)$, where $\text{logit}(\theta) = \eta$, and $\eta = \mathbf{x}^T \beta$, where \mathbf{x} is the covariate space for *textbfY*.

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



```
##      (Intercept)      duration us_popularity_estimate
##      9.721      0.003      -0.101
##      acousticness      beat_strength      bounciness
##      0.082      -0.372      0.097
##      danceability      energy      instrumentalness
##      -0.062      0.421      0.295
##      modeminor      speechiness      tempo
##      -0.062      0.561      0.001
##      valence
##      0.070
```

```
##      5%      95%
## (Intercept)      -10.625 32.464
```

```
## duration          0.001  0.006
## us_popularity_estimate -0.330  0.103
## acousticness      -0.434  0.631
## beat_strength     -1.649  0.906
## bounciness        -1.096  1.300
## danceability       -1.114  1.021
## energy            -0.330  1.171
## instrumentalness   -1.008  1.651
## modeminor         -0.282  0.164
## speechiness       -0.422  1.550
## tempo             -0.003  0.005
## valence           -0.440  0.595
```

```
(loo1 <- loo(posterior1, save_psis = TRUE))
```

```
##
## Computed from 4000 by 1000 log-likelihood matrix
##
##           Estimate   SE
## elpd_loo    -652.5 10.0
## p_loo         10.2  0.6
## looic       1304.9 19.9
## -----
## Monte Carlo SE of elpd_loo is 0.0.
##
## All Pareto k estimates are good (k < 0.5).
## See help('pareto-k-diagnostic') for details.
```

```
post0 <- stan_glm(skipped ~ 1, data = Track_features_a,
                  family = binomial(link = "logit"),
                  prior = normal(0,1), prior_intercept = normal(0,1),
                  seed = seed,
                  refresh = 0)
(loo0 <- loo(post0, save_psis = T))
```

```
##
## Computed from 4000 by 1000 log-likelihood matrix
##
##           Estimate   SE
## elpd_loo    -648.5  9.3
## p_loo         1.0  0.0
## looic       1297.0 18.7
## -----
## Monte Carlo SE of elpd_loo is 0.0.
##
## All Pareto k estimates are good (k < 0.5).
## See help('pareto-k-diagnostic') for details.
```

```
rstanarm::loo_compare(loo0, loo1)
```

```
##           elpd_diff se_diff
## post0           0.0      0.0
## posterior1     -4.0      3.3
```

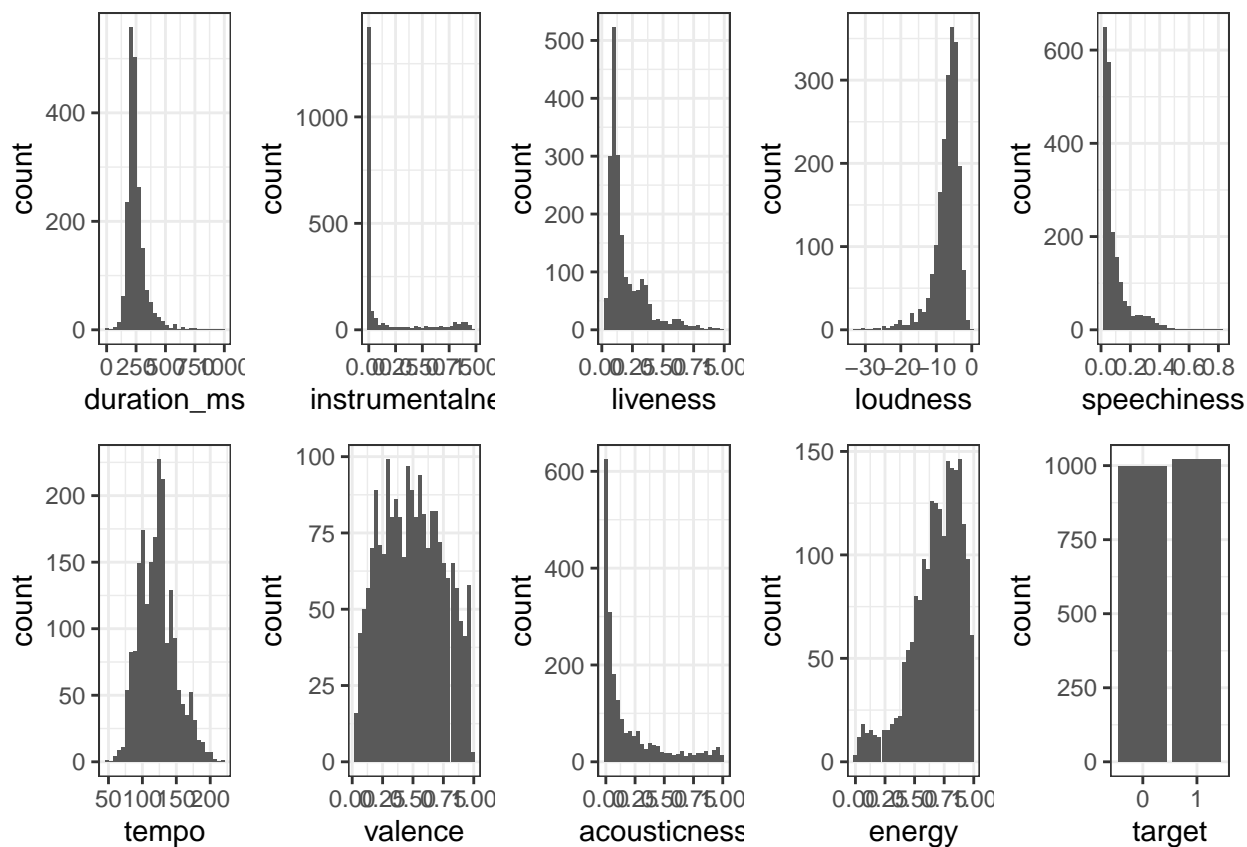
####New Data

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## -- Column specification -----
## cols(
##   X1 = col_double(),
##   acousticness = col_double(),
##   danceability = col_double(),
##   duration_ms = col_double(),
##   energy = col_double(),
##   instrumentalness = col_double(),
##   key = col_double(),
##   liveness = col_double(),
##   loudness = col_double(),
##   mode = col_double(),
##   speechiness = col_double(),
##   tempo = col_double(),
##   time_signature = col_double(),
##   valence = col_double(),
##   target = col_double(),
##   song_title = col_character(),
##   artist = col_character()
## )
```

```
#Drop un-needed variables
spotify1 <- spotify[-c(1,16,17)]
#View(spotify1)
spotify1$target <- factor(spotify1$target)
spotify1$mode <- factor(spotify1$mode)
spotify1$key <- factor(spotify1$key)
spotify1 <- spotify1 %>%
  mutate(duration_ms = duration_ms / 1000)
```

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

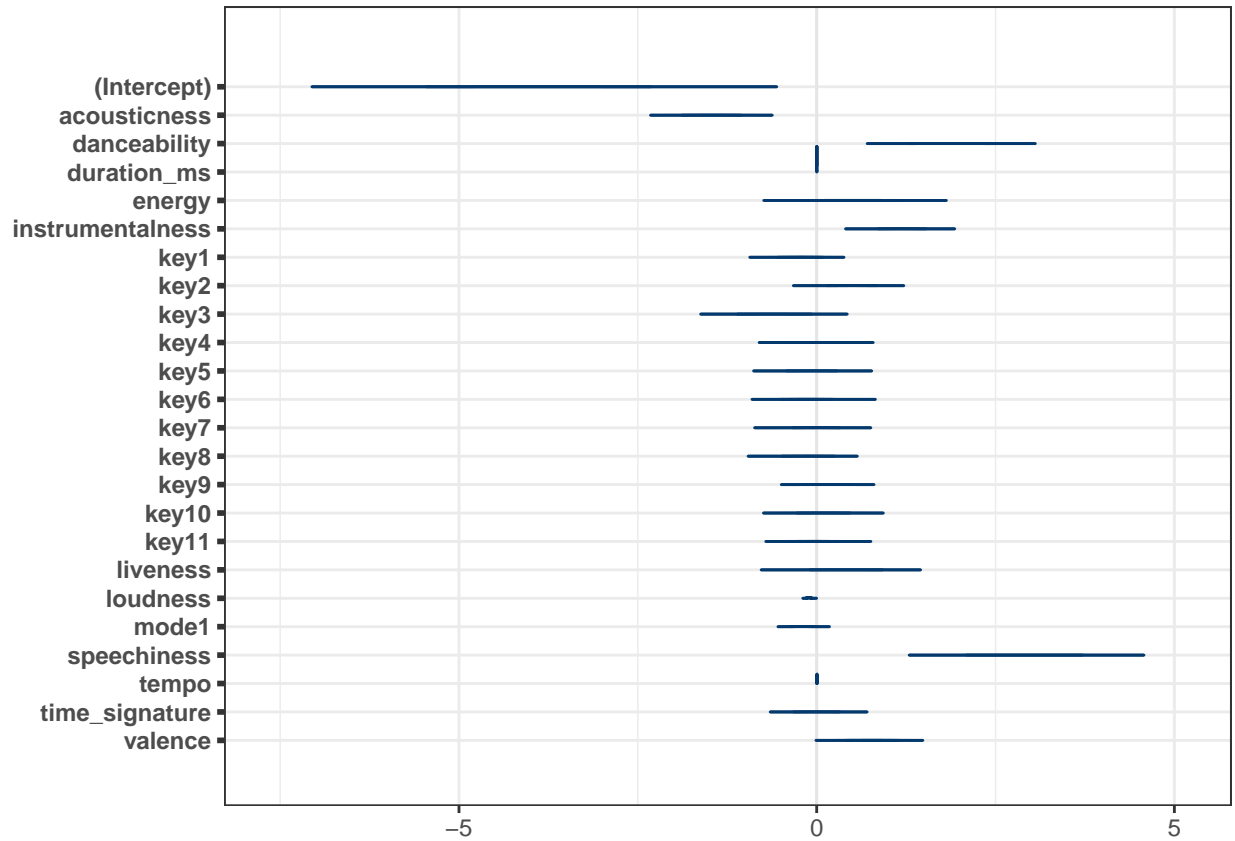


```
##
## Model Info:
## function:      stan_glm
## family:        binomial [logit]
## formula:       target ~ .
## algorithm:     sampling
## sample:        4000 (posterior sample size)
## priors:        see help('prior_summary')
## observations:  2017
## predictors:    24
##
## Estimates:
##              mean    sd  10%  50%  90%
## (Intercept)  -3.9    1.0 -5.1  -3.9  -2.7
## acousticness -1.5    0.3 -1.8  -1.5  -1.1
## danceability  1.9    0.3  1.4   1.9   2.3
## duration_ms   0.0    0.0  0.0   0.0   0.0
## energy        0.5    0.4  0.0   0.5   1.0
## instrumentalness 1.2    0.2  0.9   1.2   1.5
## key1         -0.2    0.2 -0.5  -0.2   0.0
## key2          0.5    0.2  0.2   0.5   0.8
## key3         -0.6    0.3 -1.0  -0.6  -0.2
## key4          0.0    0.2 -0.3   0.0   0.3
## key5         -0.1    0.2 -0.3  -0.1   0.2
## key6         -0.1    0.2 -0.4  -0.1   0.1
## key7          0.0    0.2 -0.3   0.0   0.2
```

```

## key8          -0.1    0.2 -0.4  -0.1    0.2
## key9          0.2     0.2 -0.1   0.2     0.4
## key10         0.1     0.2 -0.2   0.1     0.4
## key11         0.0     0.2 -0.3   0.0     0.2
## liveness      0.4     0.3  0.0   0.4     0.8
## loudness      -0.1    0.0 -0.1  -0.1    -0.1
## model1        -0.2    0.1 -0.3  -0.2    -0.1
## speechiness   2.9     0.5  2.3   2.9     3.5
## tempo         0.0     0.0  0.0   0.0     0.0
## time_signature 0.0     0.2 -0.3   0.0     0.3
## valence       0.8     0.2  0.5   0.8     1.1
##
## Fit Diagnostics:
##           mean    sd   10%   50%   90%
## mean_PPD 0.5     0.0  0.5   0.5   0.5
##
## The mean_ppd is the sample average posterior predictive distribution of the outcome variable (for de
##
## MCMC diagnostics
##           mcse Rhat n_eff
## (Intercept) 0.0  1.0  4989
## acousticness 0.0  1.0  4846
## danceability 0.0  1.0  4350
## duration_ms  0.0  1.0  3965
## energy       0.0  1.0  3140
## instrumentalness 0.0  1.0  5027
## key1         0.0  1.0  1889
## key2         0.0  1.0  2246
## key3         0.0  1.0  3447
## key4         0.0  1.0  2721
## key5         0.0  1.0  2315
## key6         0.0  1.0  2215
## key7         0.0  1.0  2402
## key8         0.0  1.0  2519
## key9         0.0  1.0  2207
## key10        0.0  1.0  2399
## key11        0.0  1.0  2371
## liveness     0.0  1.0  5539
## loudness     0.0  1.0  3396
## model1       0.0  1.0  5168
## speechiness  0.0  1.0  4780
## tempo        0.0  1.0  4846
## time_signature 0.0  1.0  5077
## valence      0.0  1.0  4466
## mean_PPD     0.0  1.0  4960
## log-posterior 0.1  1.0  1758
##
## For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective sample

```



```
##      (Intercept)      acousticness      danceability      duration_ms
##      -3.903      -1.478      1.853      0.003
##      energy instrumentalness      key1      key2
##      0.502      1.197      -0.227      0.497
##      key3      key4      key5      key6
##      -0.588      -0.016      -0.061      -0.134
##      key7      key8      key9      key10
##      0.002      -0.121      0.159      0.096
##      key11      liveness      loudness      mode1
##      -0.046      0.409      -0.109      -0.201
##      speechiness      tempo      time_signature      valence
##      2.893      0.004      0.001      0.764
```

```
##      5%      95%
## (Intercept)      -5.458 -2.318
## acousticness      -1.886 -1.056
## danceability      1.277  2.442
## duration_ms      0.002  0.004
## energy      -0.133  1.170
## instrumentalness  0.862  1.532
## key1      -0.544  0.087
## key2      0.149  0.836
## key3      -1.109 -0.071
## key4      -0.397  0.373
## key5      -0.420  0.278
```



```
## key6          -0.498  0.225
## key7          -0.333  0.323
## key8          -0.486  0.250
## key9          -0.179  0.481
## key10         -0.274  0.466
## key11         -0.388  0.298
## liveness      -0.094  0.921
## loudness      -0.145 -0.073
## model         -0.376 -0.030
## speechiness    2.093  3.715
## tempo         0.001  0.007
## time_signature -0.323  0.321
## valence        0.395  1.156
```

```
(loo3 <- loo(post2, save_psis = TRUE))
```

```
##
## Computed from 4000 by 2017 log-likelihood matrix
##
##           Estimate   SE
## elpd_loo  -1268.3 15.7
## p_loo      23.5  0.6
## looic      2536.6 31.5
## -----
## Monte Carlo SE of elpd_loo is 0.1.
##
## All Pareto k estimates are good (k < 0.5).
## See help('pareto-k-diagnostic') for details.
```

```
##
## Computed from 4000 by 2017 log-likelihood matrix
##
##           Estimate   SE
## elpd_loo  -1398.9 0.5
## p_loo      1.0 0.0
## looic      2797.9 1.0
## -----
## Monte Carlo SE of elpd_loo is 0.0.
##
## All Pareto k estimates are good (k < 0.5).
## See help('pareto-k-diagnostic') for details.
```

```
##           elpd_diff se_diff
## post2      0.0      0.0
## post4 -130.6     15.7
```

```
## Instead of posterior_linpred(..., transform=TRUE) please call posterior_epred(), which provides equi
```

```
## [1] 0.674
```

```
## [1] 0.661
```