spotify-code

German Bautista, Eric Bayer, Quentin Wetzel

11/03/2021

```
# Read data
spotify_data = readr::read_csv("spotify-data.csv")
```

```
## Rows: 100 Columns: 14
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (3): title, artist, top genre
## dbl (11): year, beats.per.minute, energy, danceability, loudness.dB, livenes...
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Clean data
names(spotify_data)[names(spotify_data) == "beats.per.minute"] = "bpm"

# INSERT annual stream rate column
# INSERT total number of streams
# INSERT years since release

# anyNA(spotify_data) FALSE
head(spotify_data, 10)
```

```
## # A tibble: 10 x 14
##    title artist `top genre`  year   bpm energy danceability loudness.dB liveness
##    <chr> <chr>  <chr>       <dbl> <dbl>  <dbl>        <dbl>       <dbl>    <dbl>
##  1 Blin~ The W~ canadian c~  2020   171     73           51          -6        9
##  2 Wate~ Harry~ pop          2019    95     82           55          -4       34
##  3 Mood~ 24kGo~ cali rap     2021    91     72           70          -4       32
##  4 Some~ Lewis~ pop          2019   110     41           50          -6       11
##  5 Perf~ Ed Sh~ pop          2017    95     45           60          -6       11
##  6 Beli~ Imagi~ modern rock  2017   125     78           78          -4        8
##  7 love~ Billi~ electropop   2018   115     30           35         -10       10
##  8 Circ~ Post ~ dfw rap      2019   120     76           70          -3        9
##  9 Shap~ Ed Sh~ pop          2017    96     65           83          -3        9
## 10 Memo~ Maroo~ pop          2021    91     33           78          -7        8
## # ... with 5 more variables: valance <dbl>, length <dbl>, acousticness <dbl>,
## #   speechiness <dbl>, popularity <dbl>
```

```
# MAKE a dataframe of pooled data by groups (genre)
# We need to decide how to make a pooled estimate. The most obvious is to take a weighted average
```

PLAN:

- Follow IMRaD structure: (Introduction, Methods, Results, Analysis, Discussion, Appendix)
- Clean and adjust data set
-
-

```
#summary(spotify_data)
```

Try different priors: an uninformative prior and another appropriate prior. Check how the posterior distributions differ (Brief sensitivity analysis)

```
# Markov Chain Monte Carlo (MCMC)
```

- Introduction
- Methods
- Results
- Analysis
- Discussion
- Appendix