Chapter 3

Purpose: provide descriptive information on data cleaning, i.e. inclusion/exclusion criteria

Our primary unit of analysis is songs. We will be using all the songs in the Tidy Tuesday Spotify datasets. The only exclusion criteria we will apply is to remove duplicate songs, indicated by track_id.

The following code will be used to remove duplicate songs, indicated by track_id:

```
spotify_songs <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytu
spotify_songs = spotify_songs[!duplicated(spotify_songs$track_id),]
length(spotify_songs$track_id)
```

```
[1] 28356
```

This reduces the sample size from 32833 songs to 28356 songs.