Question2

Ramzan Kamoto

2024-06-19

Question2

We use the rds function to load in the rds files and then bring in the csv's on their own.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
list.files('/Users/ramzankamoto/Documents/Masters/DS_EXAM/23550716/Question2/code', full.names = T, recu

custom_names <- c("spotify", "billboard_100")
directory <- "/Users/ramzankamoto/Documents/Masters/DS_EXAM/23550716/Question2/data/cold_v_met_rds"
read_rds_files(directory, custom_names)

coldplay <- read.csv("/Users/ramzankamoto/Documents/Masters/DS_EXAM/23550716/Question2/data/Coldplay_vs_
metallica <- read.csv("/Users/ramzankamoto/Documents/Masters/DS_EXAM/23550716/Question2/data/Coldplay_vs
```

Analysis

We want to create boxplots that compare the danceability of of the top 5 most popular metallica albums with the top 5 most popular cold play albums

We start by doing some cleaning

```
#renaming the albums column to easily append
coldplay1 <- coldplay %>%
    rename(album = album_name)
```

Dataframes

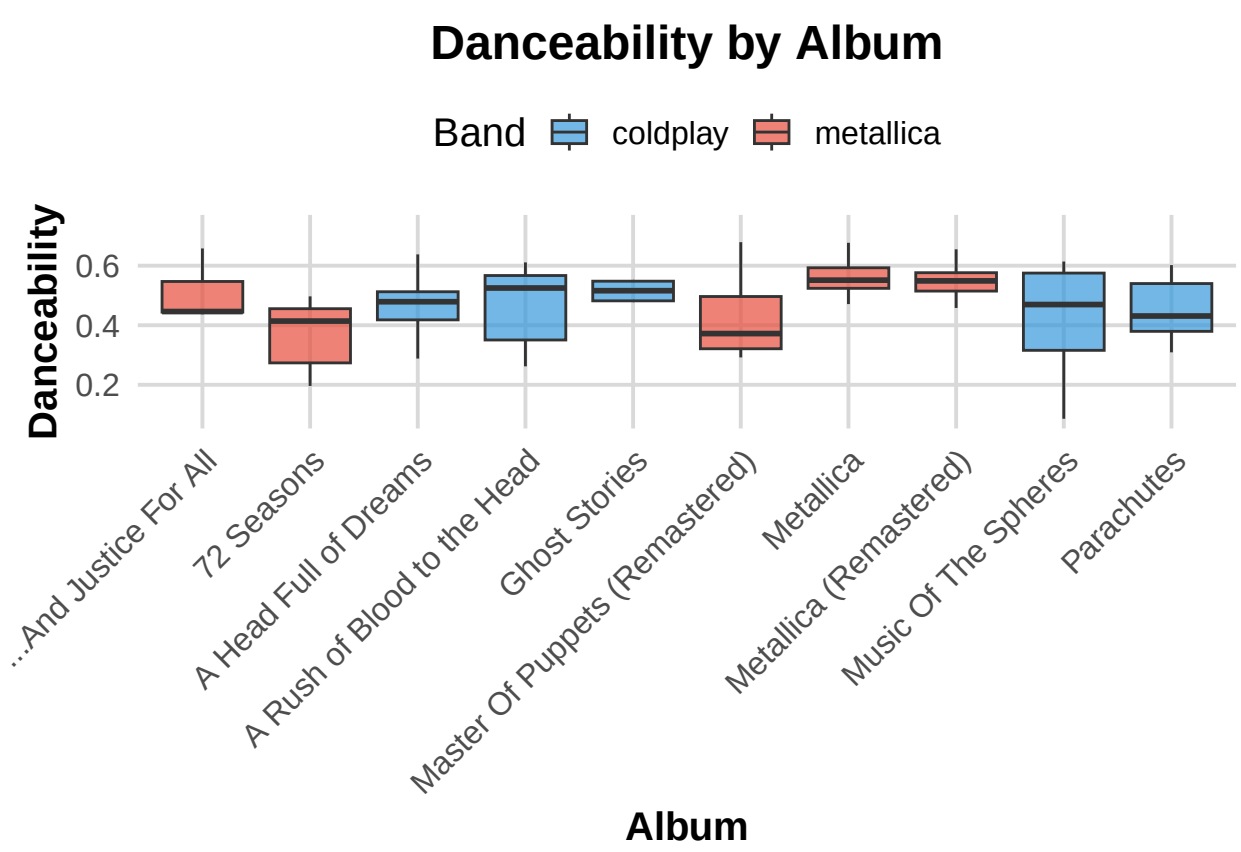
New data frames containing the songs in the top 5 albums

```
#select the variables that we will need to conduct some analysis
metallica_top_albums <- get_top_5_albums_song(metallica) %>%
    select(name, album, popularity, danceability)

metallica_top_albums$band <- "metallica"

coldplay_top_albums <- get_top_5_albums_song(coldplay1) %>%
    select(name, album, popularity, danceability)

#Create a column that has all its values as the band name
coldplay_top_albums$band <- "coldplay"

#bind the two data frames ready for plotting
top_albums <- bind_rows(metallica_top_albums, coldplay_top_albums)
```

Plot

Call function to plot danceability by album.

```
p <- create_danceability_boxplot(top_albums)
```



Summary

The two bands seem to be very comparable in terms of danceability. Most songs for both bands appear to be approximately 0.6, which when you consider these are their most popular albums may seem quite low. There are likely other factors that make these bands enjoyable which I do not explore here.