A comparative analysis of the evolution of Coldplay and Metallica Musical bands

Linda

2025-06-18

Abstract

This report explores how Coldplay and Metallica's music has evolved over the past twenty years by analyzing Spotify data from their studio albums. It focuses on core musical features such as tempo, energy, and danceability, revealing Coldplay's preference for a more moderate tempo in contrast to Metallica's consistently high-energy sound.

Introduction

Coldplay and Metallica have each carved out a distinct place in the music world, with unique artistic directions. Over the years, both bands have developed rich discographies that reflect changes in style, creative growth, and sustained fan loyalty. This report draws on Spotify data to explore their musical journeys through studio recordings, comparing their evolution within the wider context of the music industry. By examining elements such as tempo, energy, danceability, and lyrical themes, the analysis aims to understand how both bands have remained influential and relevant, leaving a lasting imprint on contemporary music culture.

Data

```
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.3
```

```
## Warning: package 'tibble' was built under R version 4.4.3
```

```
## Warning: package 'purrr' was built under R version 4.4.3
```

```
## -- 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   3.5.2     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- 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 error
```

```r
coldplay <- read.csv("C:/Users/sukol/Documents/Masters frst semester/Data Science Exam/Data-Science-Exam
metallica <- read.csv("C:/Users/sukol/Documents/Masters frst semester/Data Science Exam/Data-Science-Exam
spoitfy <- readRDS("C:/Users/sukol/Documents/Masters frst semester/Data Science Exam/Data-Science-Exam/
billboard_100 <- readRDS("C:/Users/sukol/Documents/Masters frst semester/Data Science Exam/Data-Science-

#renaming the albums column to be easily append
coldplay_clean <- coldplay %>%
    rename(album = album_name)
emetallica_clean <- metallica%>%
  filter(!str_detect( name , regex("live", ignore_case = TRUE)))
```

This dataset contains detailed audio and metadata for various songs, including duration, release date, and album name. It tracks whether lyrics are explicit and assigns a popularity score out of 100. Musical attributes such as tempo, danceability, energy, and valence describe the character and mood of each track. Acousticness, instrumentalness, and speechiness indicate the nature of sound, from acoustic instrumentation to spoken word content. Additional features like loudness, liveness, and genre suitability offer insight into how the music is experienced by listeners.After successifully loading the data, I cleaned it to make it suitable for the analysis.

Results and Analysis

In this section I 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.

```r
metallica_clean <- metallica %>%
  filter(!is.na(tempo), !is.na(danceability), !is.na(energy))


coldplay_features <- coldplay %>%
  select(name, tempo, danceability, energy)


metallica_features <- metallica_clean %>%
  select(name , tempo, danceability, energy)



# Combine data for plotting
combined_df<- bind_rows(
  mutate(coldplay_features, Band = "Coldplay"),
  mutate(metallica_features, Band = "Metallica")
)

# Comparison of Tempo between Coldplay and Metallica

g_tempo <- ggplot(combined_df, aes(x = Band, y = tempo, fill = Band)) +
  geom_boxplot() +
  labs(title = "Comparison of Tempo between Coldplay and Metallica",
```
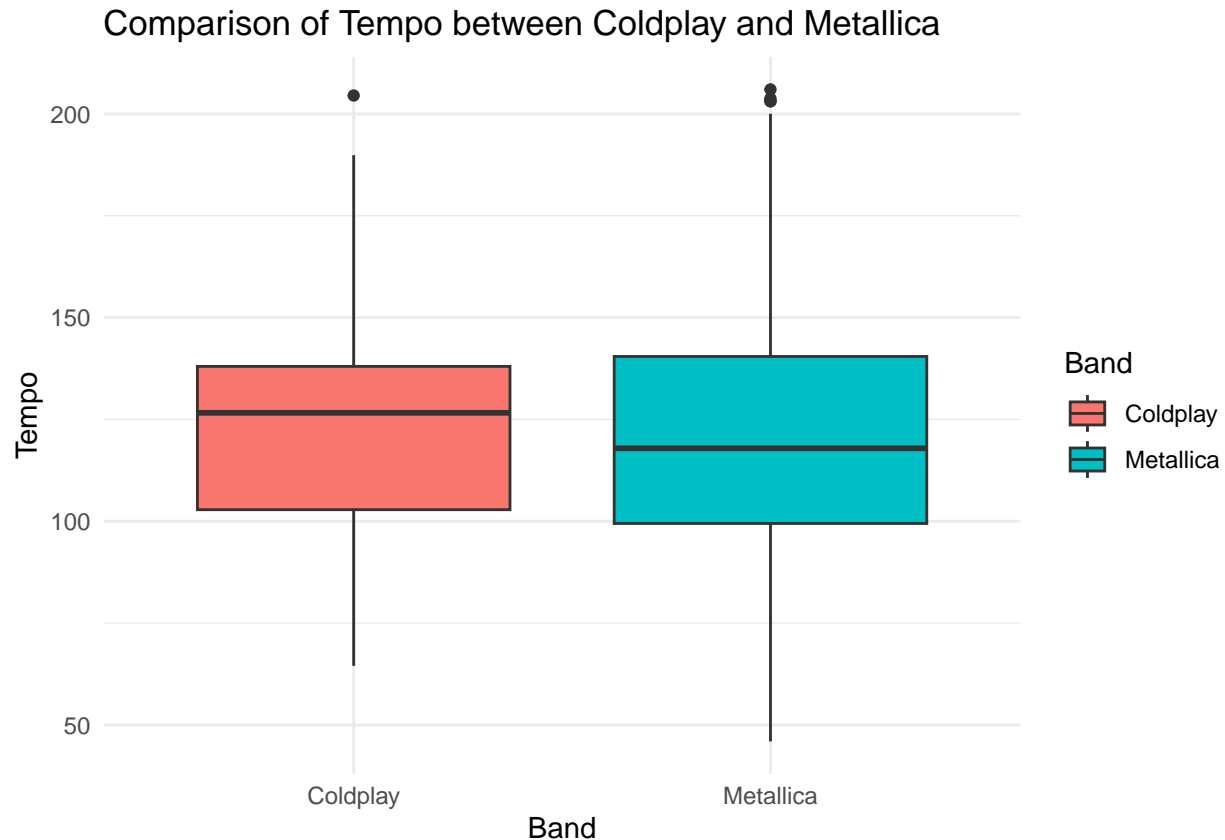
```
        x = "Band",
        y = "Tempo") +
  theme_minimal()
g_tempo
```

Comparison of Tempo between Coldplay and Metallica

The boxplot compares the tempo of songs between Coldplay and Metallica. Coldplay has a slightly higher median tempo, indicating that their songs generally tend to be a bit faster than Metallica's on average. However, Metallica shows a wider range and greater variability in tempo, suggesting that their music spans a broader spectrum of speeds. Coldplay's tempo distribution is more concentrated, with one notable outlier above 200 BPM, while Metallica's tempos range from around 50 to nearly 200 BPM. Overall, Coldplay's songs appear to have more consistent tempos, whereas Metallica explores a greater variety of tempos in their music.

```
get_top_5_albums_song <- function(df) {
  df %>%
    group_by(album) %>%
    summarise(mean_popularity = mean(popularity, na.rm = TRUE)) %>%
    arrange(desc(mean_popularity)) %>%
    slice_head(n = 5)
}
get_top_5_albums_song <- function(df) {
  top_albums <- df %>%
    group_by(album) %>%
    summarise(avg_popularity = mean(popularity, na.rm = TRUE)) %>%
    arrange(desc(avg_popularity)) %>%
    slice_head(n = 5) %>%
```

```r
    pull(album)  # just get the top 5 album names

  df %>%
    filter(album %in% top_albums)  # return full data for those albums
}



#selecting the variables that I will need to conduct this analysis
metallica_top_albums <- get_top_5_albums_song(metallica_clean) %>%
    select(name, album, popularity, danceability)

metallica_top_albums$band <- "metallica"


get_top_5_albums_song <- function(df) {
  df %>%
    group_by(album_name) %>%
    summarise(avg_popularity = mean(popularity, na.rm = TRUE)) %>%
    arrange(desc(avg_popularity)) %>%
    slice(1:5) %>%
    inner_join(df, by = "album_name")
}

coldplay_top_albums <- get_top_5_albums_song(coldplay) %>%
    select(name, album_name, 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)



#Call function to plot danceability by album.
library(ggplot2)

create_danceability_boxplot <- function(df) {
  ggplot(df, aes(x = album_name, y = danceability)) +
    geom_boxplot(fill = "skyblue") +
    labs(
      title = "Danceability by Album",
      x = "Album",
      y = "Danceability"
    ) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
}

#Then call it
p <- create_danceability_boxplot(coldplay_top_albums)
print(p)
```
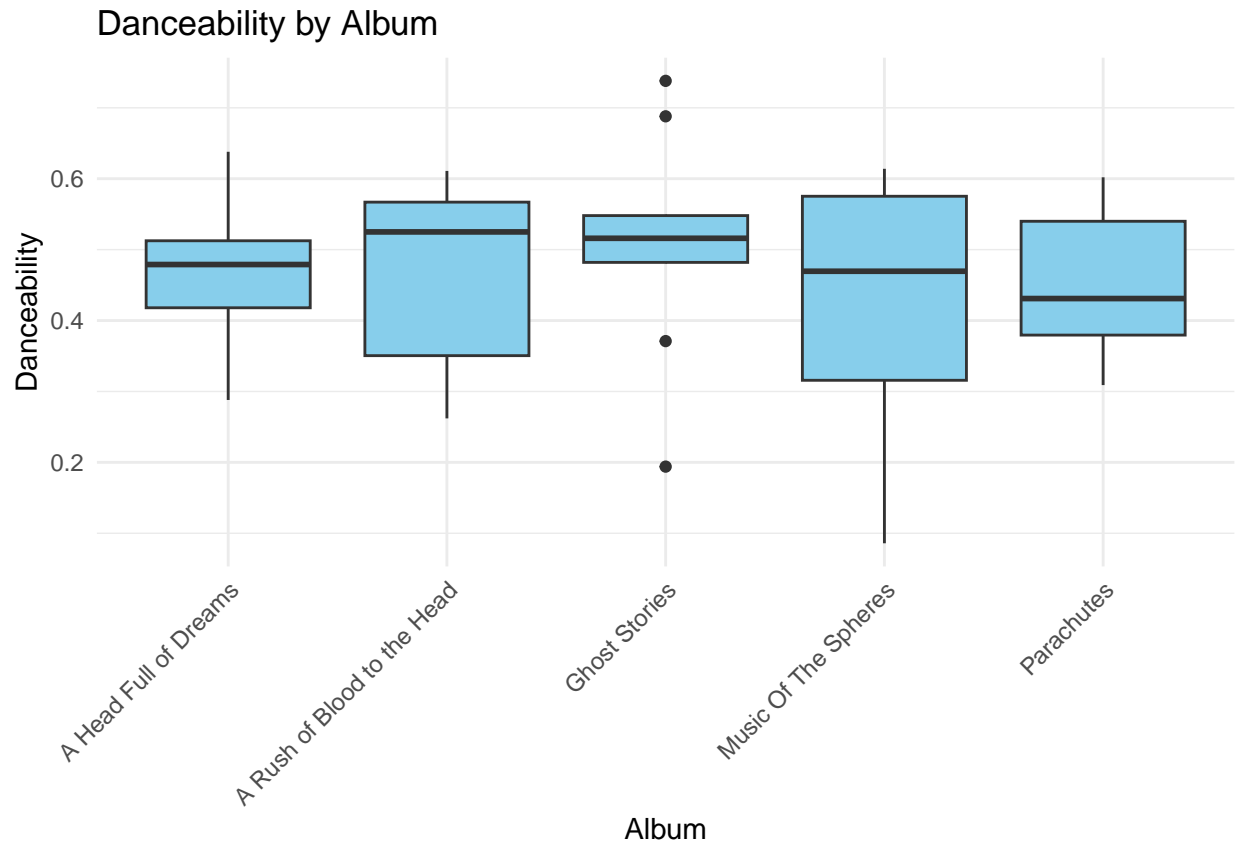
Danceability by Album

The boxplot compares the danceability of Coldplay's top five albums. Ghost Stories stands out with consistently high danceability and minimal variation, suggesting a uniform rhythm style. A Rush of Blood to the Head and Music of the Spheres show wider variability, indicating a mix of both highly and less danceable tracks. Parachutes and A Head Full of Dreams fall in the middle range, with moderate danceability. Overall, the graph highlights how the band's rhythmic style shifts across albums.

```r
# Example: Extract audio features (adjust based on available columns)

coldplay_features <- coldplay %>%
  select(name, tempo, danceability, energy)

metallica_features <- metallica_clean %>%
  select(name , tempo, danceability, energy)


# Combine data for plotting (example)

# Combine data for plotting (example)
combined_df<- bind_rows(
  mutate(coldplay_features, Band = "Coldplay"),
  mutate(metallica_features, Band = "Metallica")
)

# Example Plot: Comparison of Tempo between Coldplay and Metallica

g_tempo <- ggplot(combined_df, aes(x = Band, y = tempo, fill = Band)) +
```

```
  geom_boxplot() +
  labs(title = "Comparison of Tempo between Coldplay and Metallica",
       x = "Band",
       y = "Tempo") +
  theme_minimal()
###tempo comparison graph works fine


coldplay_df <- data.frame(coldplay)

metallica_df <- data.frame(metallica_clean)

gscatter <- ggplot(combined_df, aes(x = tempo, y = danceability, color = Band)) +
  geom_point(alpha = 0.7) +
  labs(title = "Tempo vs. Danceability",
       x = "Tempo",
       y = "Danceability",
       color = "Band") +
  theme_minimal()

gscatter
```
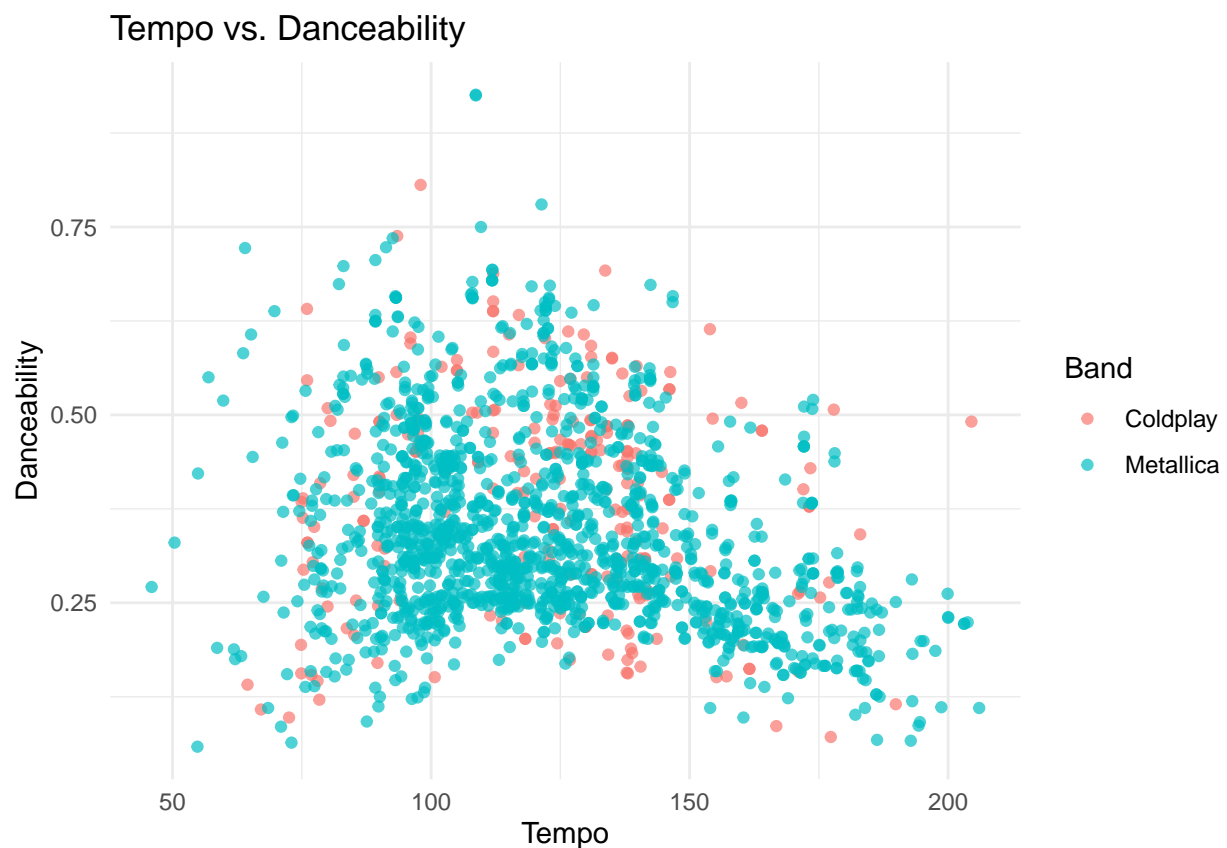

The scatter plot shows the relationship between tempo and danceability for Coldplay and Metallica songs. Both bands cluster around a tempo range of 80–140 BPM, with no strong correlation between tempo and danceability. Meltica's songs (red) appear slightly more spread out at higher danceability values. Overall, both bands have songs that vary widely in how danceable they are, regardless of tempo.

```r
  # Filter out non-studio recordings based on song names
clean_coldplay <- coldplay %>%
  filter(!grepl(" live", name , ignore.case = TRUE))  # Adjust based on actual column names

clean_metallica <- metallica_clean %>%
  filter(!str_detect( name , regex("live", ignore_case = TRUE)))

#Feature Extraction
#Extract relevant features or metrics for comparison. This could include audio features like tempo, dan

#Extract audio features (adjust based on available columns)

coldplay_features <- clean_coldplay %>%
  select(name, tempo, danceability, energy)

metallica_features <- clean_metallica %>%
  select(name , tempo, danceability, energy)


# Combine data for plotting

combined_df<- bind_rows(
  mutate(coldplay_features, Band = "Coldplay"),
  mutate(metallica_features, Band = "Metallica")
)

# Example Plot: Comparison of Tempo between Coldplay and Metallica

g_tempo <- ggplot(combined_df, aes(x = Band, y = tempo, fill = Band)) +
  geom_boxplot() +
  labs(title = "Comparison of Tempo between Coldplay and Metallica",
       x = "Band",
       y = "Tempo") +
  theme_minimal()
###tempo comparison graph works fine


coldplay_df <- data.frame(coldplay)

metallica_df <- data.frame(metallica_clean)

# Density plot of Danceability by Band

gdens <- ggplot(combined_df, aes(x = danceability, fill = Band)) +
  geom_density(alpha = 0.5) +
  labs(title = "Density of Danceability by Band",
       x = "Danceability",
       fill = "Band") +
  theme_minimal()
gdens
```
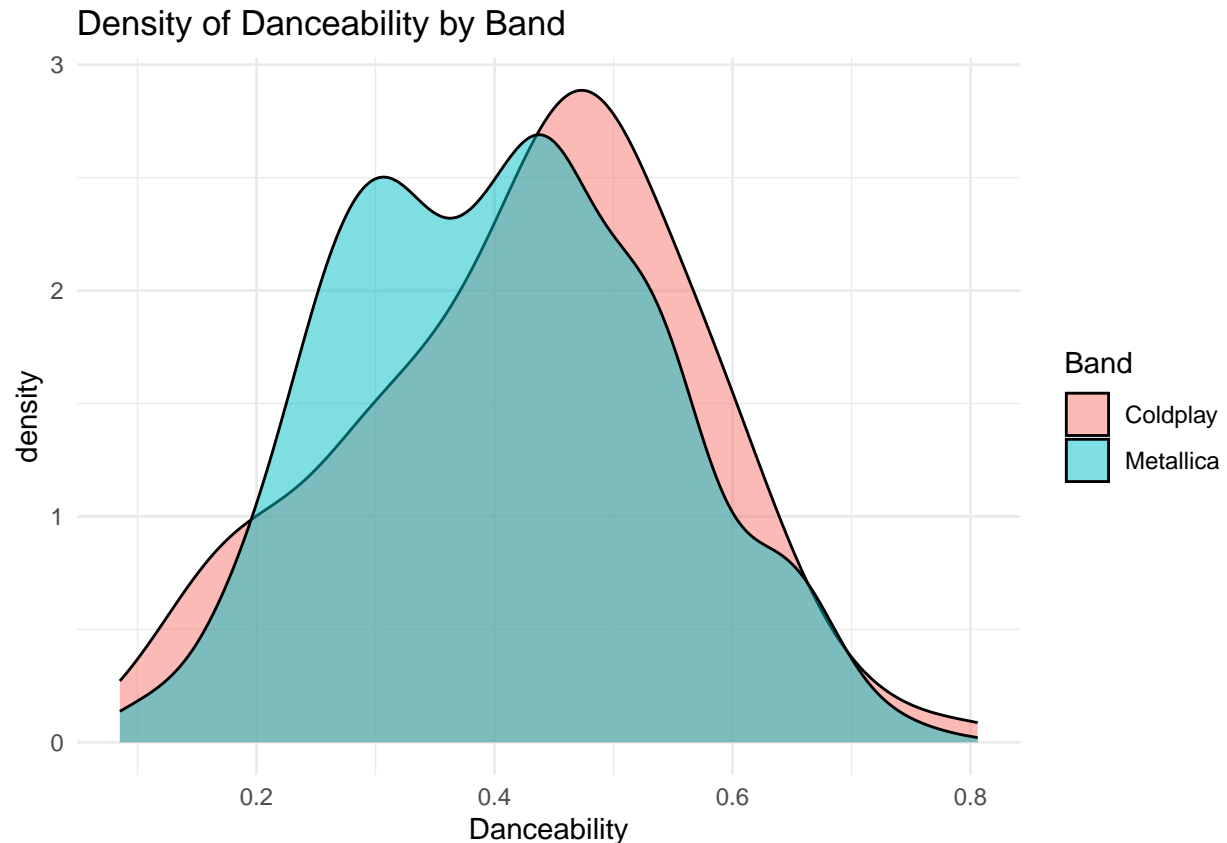
Density of Danceability by Band

The density plot below supports the interpretation of the scatter plot. Metallica's distribution is skewed, indicating that many of their songs tend to be more danceable, likely due to higher energy levels in the tracks. In contrast, Coldplay's density plot appears more normally distributed, showing a more even spread of danceability. This points to a consistent and varied approach in producing music that remains engaging and accessible to listeners.

```r
# Filter out non-studio recordings based on song names
clean_coldplay <- coldplay %>%
  filter(!grepl(" live", name , ignore.case = TRUE))  # Adjust based on actual column names

clean_metallica <- metallica_clean %>%
  filter(!str_detect( name , regex("live", ignore_case = TRUE)))


# Extract audio features (adjust based on available columns)

coldplay_features <- clean_coldplay %>%
  select(name, tempo, danceability, energy)

metallica_features <- clean_metallica %>%
  select(name , tempo, danceability, energy)


# Combine data for plotting
```

```r
combined_df<- bind_rows(
  mutate(coldplay_features, Band = "Coldplay"),
  mutate(metallica_features, Band = "Metallica")
)

# Comparison of Tempo between Coldplay and Metallica

g_tempo <- ggplot(combined_df, aes(x = Band, y = tempo, fill = Band)) +
  geom_boxplot() +
  labs(title = "Comparison of Tempo between Coldplay and Metallica",
       x = "Band",
       y = "Tempo") +
  theme_minimal()
#tempo comparison graph


coldplay_df <- data.frame(coldplay)

metallica_df <- data.frame(metallica_clean)

# Boxplot of Energy by Band

gbox <- ggplot(combined_df, aes(x = Band, y = energy, fill = Band)) +
  geom_boxplot() +
  labs(title = "Energy Distribution by Band",
       x = "Band",
       y = "Energy",
       fill = "Band") +
  theme_minimal()

gbox
```
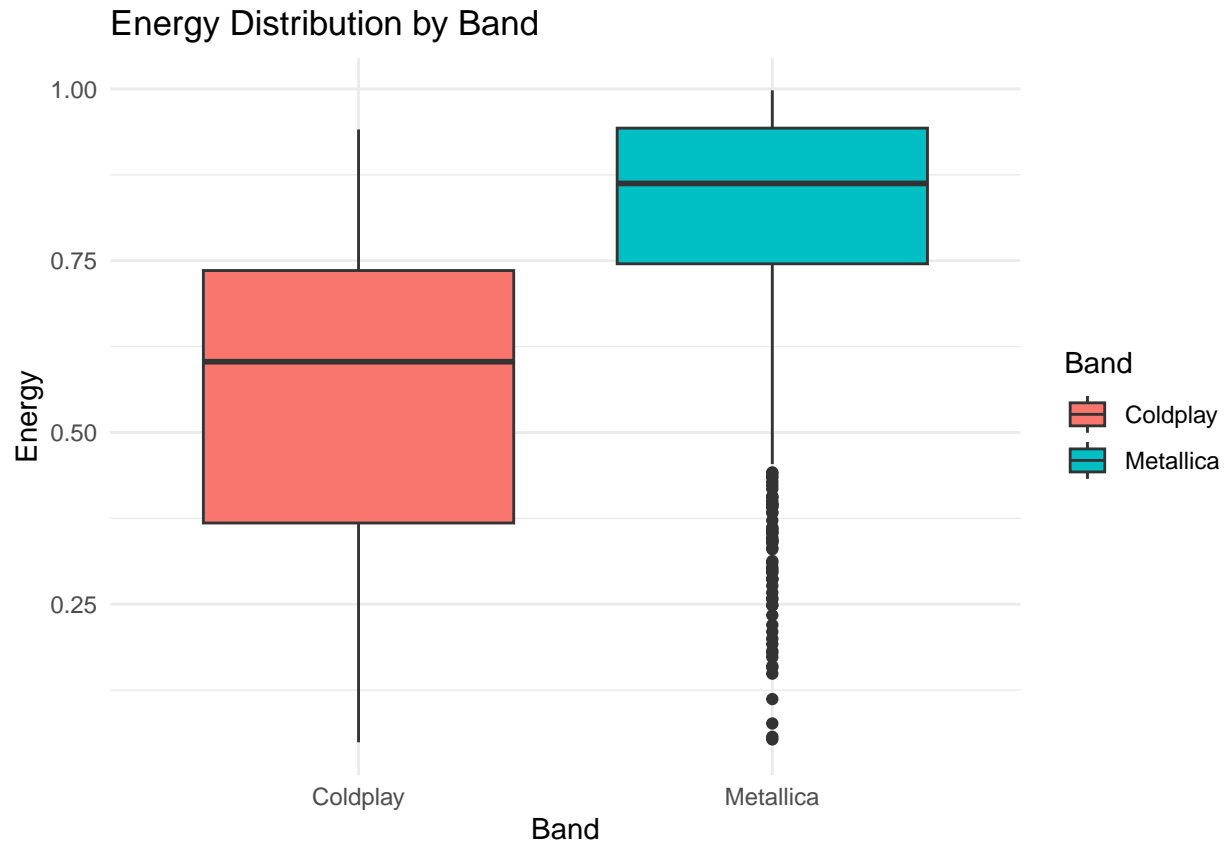
Energy Distribution by Band

The box plot highlights Metallica's strong energy levels; the long upper whiskers point to outliers with exceptionally high intensity. A higher median suggests that Metallica consistently releases high-energy tracks, emphasizing their energetic and dynamic sound. The compact box shows limited variation in energy, reinforcing the band's distinct style of producing powerful, intense music. In contrast, Coldplay's box plot shows shorter tails and a lower median, indicating a broader range of energy levels and a more balanced musical output.

Conclusion

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. Metallica's elevated levels of energy, danceability, and tempo align well with the tastes of younger audiences, who tend to favor vibrant and intense musical experiences. On the other hand, Coldplay's softer, more laid-back sound is likely to resonate with older listeners who value melody and reflective songwriting. These differences illustrate how musical tastes can shift between generations, influenced by preferences in style and themes.