

# Spotify Analysis

Cam Smithers

2025-04-02

## Contents

<b>Cam's Spotify Streaming History</b>	<b>1</b>
Data Cleaning . . . . .	1
Personal Favorites . . . . .	2
Cumulative Counts (by Year and Overall) . . . . .	3
All Time Songs/Artists/Albums Plays . . . . .	4
Yearly Top 10 Songs/Artists/Albums (by number of plays) . . . . .	5
Joining Data . . . . .	6
Data Cleaning 2 . . . . .	6
Plotting Data . . . . .	6
<b>Listening Analysis</b>	<b>6</b>
All Songs/Artists/Albums . . . . .	6
Favorite Songs/Artists/Albums Statistics . . . . .	21
10 Most Played Songs/Artists/Albums (of All Time) Over Time . . . . .	39
Significant Dates . . . . .	54

## Cam's Spotify Streaming History

### Data Cleaning

1. Loading in data
2. Selecting variable I want to keep
3. Renaming variables
4. Dropping rows with missing values
5. Changing the date column from character to date
6. Creating a column for the specific year

```
streaming_data <- read_csv("/Users/camsmithers/Desktop/Data/combined_data.csv")

## Rows: 273776 Columns: 23
## -- Column specification -----
## Delimiter: ","
## chr  (12): platform, conn_country, ip_addr, master_metadata_track_name, mast...
## dbl  (2): ms_played, offline_timestamp
## lgl  (8): audiobook_title, audiobook_uri, audiobook_chapter_uri, audiobook...
## dtm  (1): ts
##
## 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.
```

```
streaming_data <- streaming_data %>%
  select(ts, ms_played, master_metadata_track_name,
         master_metadata_album_artist_name,
         master_metadata_album_album_name,
         spotify_track_uri, reason_start, reason_end,
         shuffle, skipped) %>%
  rename(
    "play_timestamp"="ts", "millisec_played"="ms_played",
    "song"="master_metadata_track_name",
    "album"="master_metadata_album_album_name",
    "artist"="master_metadata_album_artist_name") %>%
  drop_na() %>%
  mutate(play_timestamp = as.Date(play_timestamp)) %>%
  mutate(play_year = as.factor(year(play_timestamp)))
```

## Personal Favorites

Artists/Albums/Songs

1. Selecting needed vars
2. Filtering for favorite artists/albums/songs
3. Using unique() to only have one observation
4. Creating a column to specify it's a favorite of mine

```
favorite_artists <- streaming_data %>%
  select(artist) %>%
  filter(
    artist == "H.E.R." | artist == "Kehlani" | artist == "Drake" |
    artist == "Ariana Grande" | artist == "UMI" | artist == "Jenevieve" |
    artist == "Alina Baraz" | artist == "Lil Tecca" |
    artist == "Muni Long" | artist == "J. Cole" | artist == "NIKI" |
    artist == "Mac Ayres") %>%
  unique() %>%
  mutate(favorite_artist = "Yes")

favorite_albums <- streaming_data %>%
  select(album, artist) %>%
  filter(
    (album == "H.E.R." & artist == "H.E.R.") |
    (album == "Back of My Mind" & artist == "H.E.R.") |
    (album == "Forest in the City" & artist == "UMI") |
    (album == "Her Loss" & artist == "Drake") |
    (album == "Championships" & artist == "Meek Mill") |
    (album == "SweetSexySavage" & artist == "Kehlani") |
    (album == "blue water road" & artist == "Kehlani") |
    (album == "Public Displays Of Affection: The Album" &
     artist == "Muni Long") |
    (album == "Sweetener" & artist == "Ariana Grande") |
    (album == "It Was Divine" & artist == "Alina Baraz") |
    (album == "The Color of You" & artist == "Alina Baraz") |
    (album == "thank u, next" & artist == "Ariana Grande") |
    (album == "Rendezvous - EP" & artist == "Jenevieve") |
    (album == "Harry's House" & artist == "Harry Styles") |
```

```

      (album == "NEVER ENOUGH" & artist == "Daniel Caesar") |
      (album == "Fetty Wap" & artist == "Fetty Wap")) %>%
unique() %>%
mutate(favorite_album = "Yes")

favorite_songs <- streaming_data %>%
select(song, artist) %>%
filter(
  (artist == "H.E.R." & (song == "Changes" | song == "Gone Away" |
    song == "Rather Be" | song == "Jungle" | song == "My Own")) |
  (artist == "Meek Mill" &
    song == "Dangerous (feat. Jeremih and PnB Rock)") |
  (artist == "UMI" & (song == "moonlit room" |
    song == "Love Affair")) |
  (artist == "Eric Bellinger" & song == "Goat 2.0 (feat. Wale)") |
  (artist == "Drake" & song == "Spin Bout U") |
  (artist == "Metro Boomin" &
    song == "Trance (with Travis Scott & Young Thug)") |
  (artist == "21 Savage" & song == "prove it") |
  (artist == "Kehlani" & (song == "As I Am" | song == "melt" |
    song == "After Hours")) |
  (artist == "Lil Tecca" & song == "MONEY ON ME") |
  (artist == "Lil Baby" & song == "Go Hard") |
  (artist == "Jenevieve" &
    (song == "Love Quotes" | song == "Nxwhere")) |
  (artist == "Bryson Tiller" & song == "Years Go By") |
  (artist == "Carly Rae Jepsen" & song == "Run Away With Me") |
  (artist == "Harry Styles" & song == "Satellite")) %>%
unique() %>%
mutate(favorite_song = "Yes")

```

## Cumulative Counts (by Year and Overall)

### Yearly Cumulative Counts

1. Arrange by date
2. Group by...
  - Song, Artist, and Year Played
  - Artist and Year Played
  - Album, Artist, and Year Played
3. Running total of the number of times the artist/song/album was played
4. Remove the groupings

### All Time Cumulative Counts

1. Arrange by date
2. Group by...
  - Song and Artist
  - Artist
  - Album and Artist
3. Running total of the number of times the artist/song/album was played
4. Remove the groupings

```

#Total Count by Year for Songs, Artists, Albums
streaming_data_2 <- streaming_data %>%

```

```

#Yearly Running Total

##Song Count
arrange(play_timestamp) %>%
group_by(song, artist, play_year) %>%
mutate(yearly_song_cumsum = row_number(song)) %>%
ungroup() %>%
##Artist Count
arrange(play_timestamp) %>%
group_by(artist, play_year) %>%
mutate(yearly_artist_cumsum = row_number(artist)) %>%
ungroup() %>%
##Album Count
arrange(play_timestamp) %>%
group_by(album, artist, play_year) %>%
mutate(yearly_album_cumsum = row_number(album)) %>%
ungroup() %>%

#All Time Running Total

##Song Count
arrange(play_timestamp) %>%
group_by(song, artist) %>%
mutate(alltime_song_cumsum = row_number(song)) %>%
ungroup() %>%
##Artist Count
arrange(play_timestamp) %>%
group_by(artist) %>%
mutate(alltime_artist_cumsum = row_number(artist)) %>%
ungroup() %>%
##Album Count
arrange(play_timestamp) %>%
group_by(album, artist) %>%
mutate(alltime_album_cumsum = row_number(album)) %>%
ungroup()

```

## All Time Songs/Artists/Albums Plays

1. Group by...
  - Song and Artist
  - Artist
  - Album and Artist
2. Count the number of observations
3. Sort the values (high to low)

```

alltime_songs <- streaming_data_2 %>%
  group_by(song, artist) %>%
  summarize(alltime_song_count = n(), .groups = "drop") %>%
  arrange(desc(alltime_song_count))

alltime_artists <- streaming_data_2 %>%
  group_by(artist) %>%
  summarize(alltime_artist_count = n(), .groups = "drop") %>%
  arrange(desc(alltime_artist_count))

```

```
alltime_albums <- streaming_data_2 %>%
  group_by(album, artist) %>%
  summarize(alltime_album_count = n(), .groups = "drop") %>%
  arrange(desc(alltime_album_count))
```

## Yearly Top 10 Songs/Artists/Albums (by number of plays)

Yearly Top 10 Songs/Artists/Albums

1. Group by...
2. Get the maximum cumulative sum
3. Sort the observations by the year played, then by the number of plays
4. Group by the year played
5. Select the top ten values for each year
6. Column to specify if a song/artist/album was top ten in a respective year

Unique Songs/Artists/Albums

1. Using `distinct()` to remove duplicates that were top ten in multiple years

```
yearly_top10_songs <- streaming_data_2 %>%
  group_by(play_year, song, artist) %>%
  summarize(yearly_top10_songs_count = max(yearly_song_cumsum),
    .groups = "drop") %>%
  arrange(play_year, desc(yearly_top10_songs_count)) %>%
  group_by(play_year) %>%
  slice_head(n = 10) %>%
  mutate(year_x_top10_song = "Top 10 Song")

yearly_top10_artists <- streaming_data_2 %>%
  group_by(play_year, artist) %>%
  summarize(yearly_top10_artists_count = max(yearly_artist_cumsum),
    .groups = "drop") %>%
  arrange(play_year, desc(yearly_top10_artists_count)) %>%
  group_by(play_year) %>%
  slice_head(n = 10) %>%
  mutate(year_x_top10_artist = "Top 10 Artist")

yearly_top10_albums <- streaming_data_2 %>%
  group_by(play_year, album, artist) %>%
  summarize(yearly_top10_albums_count = max(yearly_album_cumsum),
    .groups = "drop") %>%
  arrange(play_year, desc(yearly_top10_albums_count)) %>%
  group_by(play_year) %>%
  slice_head(n = 10) %>%
  mutate(year_x_top10_album = "Top 10 Album")

top_unique_songs <- yearly_top10_songs %>%
  distinct(song, artist)

top_unique_artists <- yearly_top10_artists %>%
  distinct(artist)

top_unique_albums <- yearly_top10_albums %>%
  distinct(album, artist)
```

## Joining Data

Using left join to bring summary data frames into the main data frame

```
spotify_data <- streaming_data_2 %>%
  left_join(alltime_songs, by = c("song", "artist")) %>%
  left_join(alltime_artists, by = "artist") %>%
  left_join(alltime_albums, by = c("album", "artist")) %>%
  left_join(yearly_top10_songs, by = c("play_year", "song", "artist")) %>%
  left_join(yearly_top10_artists, by = c("play_year", "artist")) %>%
  left_join(yearly_top10_albums, by = c("play_year", "album", "artist")) %>%
  left_join(favorite_songs, by = c("artist", "song")) %>%
  left_join(favorite_artists, by = "artist") %>%
  left_join(favorite_albums, by = c("artist", "album"))
```

## Data Cleaning 2

Column to identify if a song/artist/album is its respective unique list

```
spotify_data_2 <- spotify_data %>%
  mutate(
    ever_top10_song = if_else(
      (artist %in% top_unique_songs$artist) &
      (song %in% top_unique_songs$song), "Yes", NA),
    ever_top10_artist = if_else(
      artist %in% top_unique_artists$artist, "Yes", NA),
    ever_top10_album = if_else(
      (artist %in% top_unique_albums$artist) &
      (album %in% top_unique_albums$album), "Yes", NA))
```

## Plotting Data

All Time Song/Artist/Album Plot Data: Joining main data to respective plot data.

```
alltime_songs_plot_data <- alltime_songs %>%
  left_join(spotify_data_2, by = c("song", "artist", "alltime_song_count"))

alltime_albums_plot_data <- alltime_albums %>%
  left_join(spotify_data_2, by = c("album", "artist", "alltime_album_count"))

alltime_artists_plot_data <- alltime_artists %>%
  left_join(spotify_data_2, by = c("artist", "alltime_artist_count"))
```

---

## Listening Analysis

### All Songs/Artists/Albums

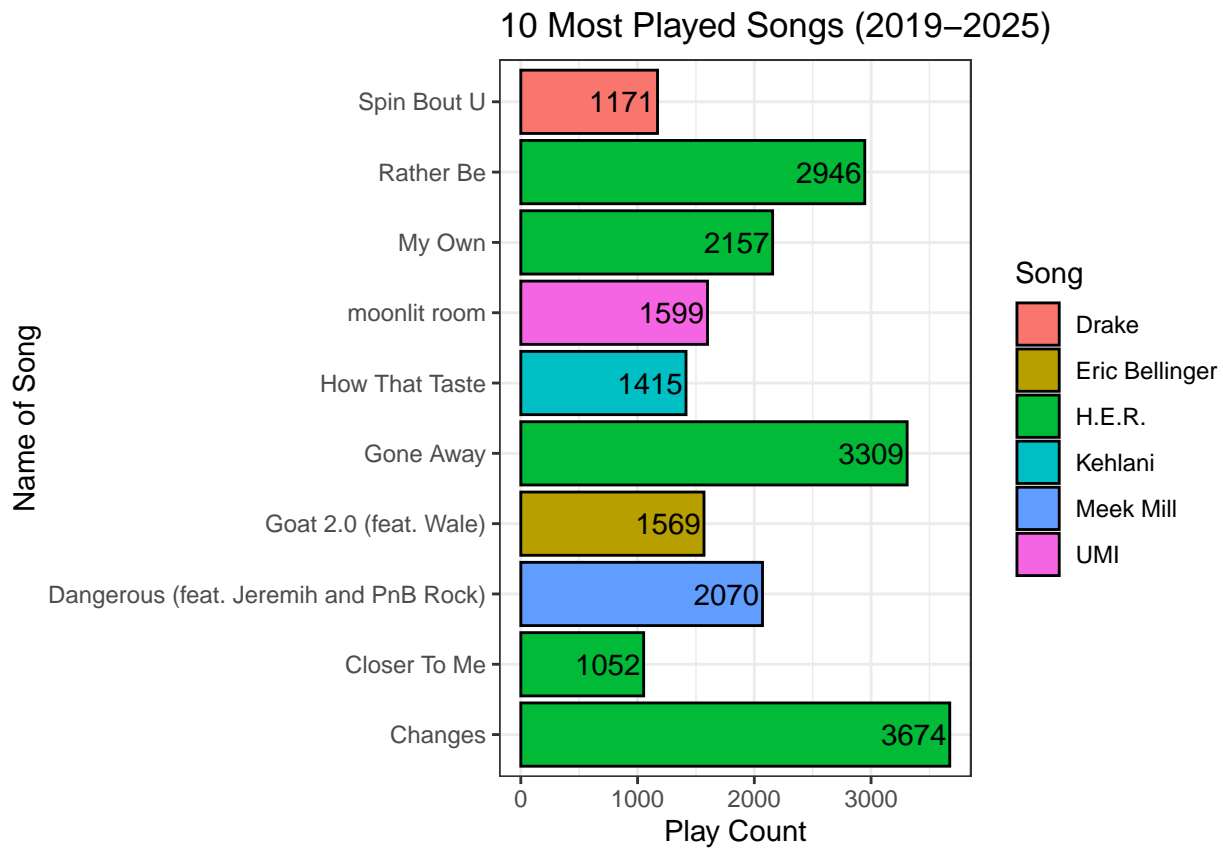
#### 10 Most Played Songs/Artists/Albums of All Time

```
#Most Listened to Songs of All Time
top10_songs_oat_bar <- ggplot(alltime_songs %>%
  slice_head(n = 10),
  aes(x = alltime_song_count, y = song,
  fill = artist)) +
```

```

geom_bar(stat = "identity", position = "dodge", color = "black") +
geom_text(aes(label = alltime_song_count),
          hjust = 1.05,
          size = 4) +
theme_bw() +
labs(
  title = "10 Most Played Songs (2019-2025)",
  fill = "Song",
  y = "Name of Song",
  x = "Play Count"
)
top10_songs_oat_bar

```



```

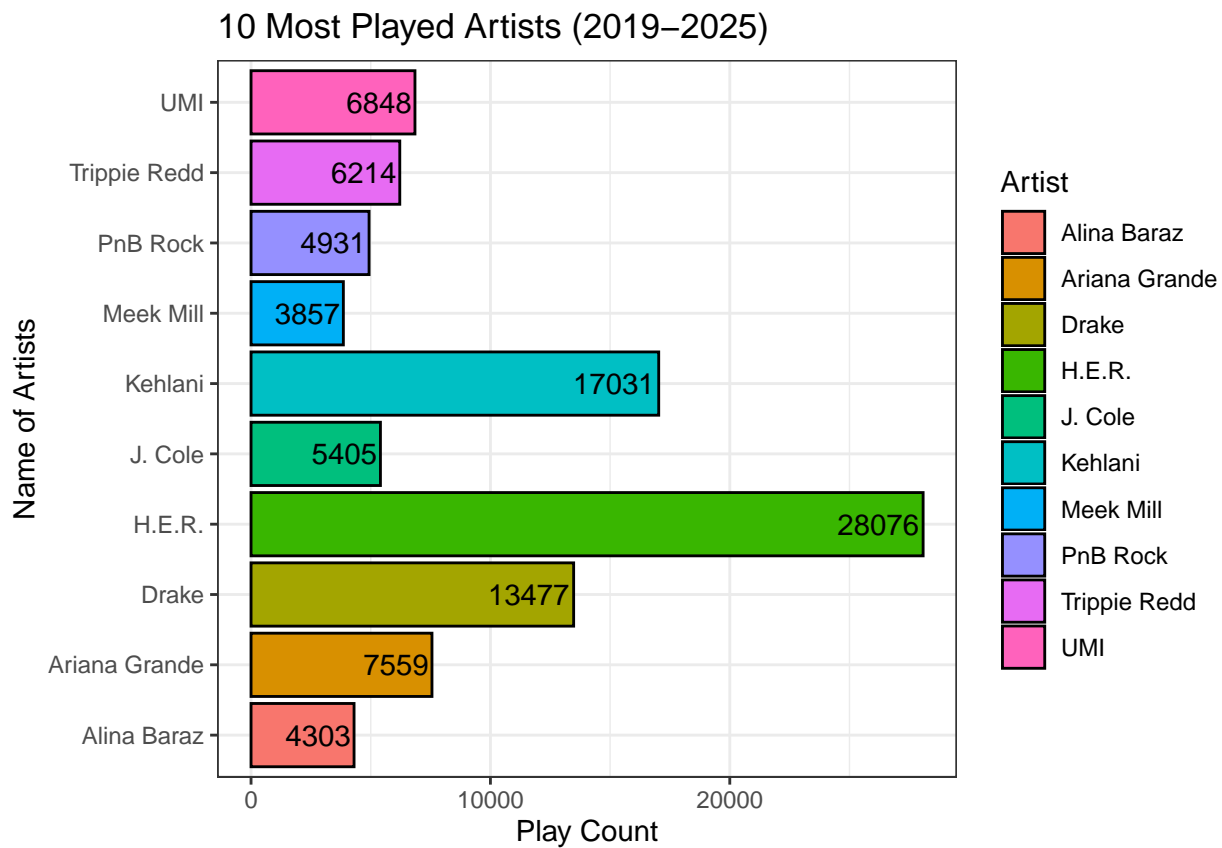
#Most Listened to Artists of All Time
top10_artists_oat_bar <- ggplot(alltime_artists %>%
  slice_head(n = 10),
  aes(x = alltime_artist_count, y = artist,
      fill = artist)) +
geom_bar(stat = "identity", position = "dodge", color = "black") +
geom_text(aes(label = alltime_artist_count),
          hjust = 1.05,
          size = 4) +
theme_bw() +
labs(
  title = "10 Most Played Artists (2019-2025)",
  fill = "Artist",
  y = "Name of Artists",

```

```

    x = "Play Count"
  )
top10_artists_oat_bar

```

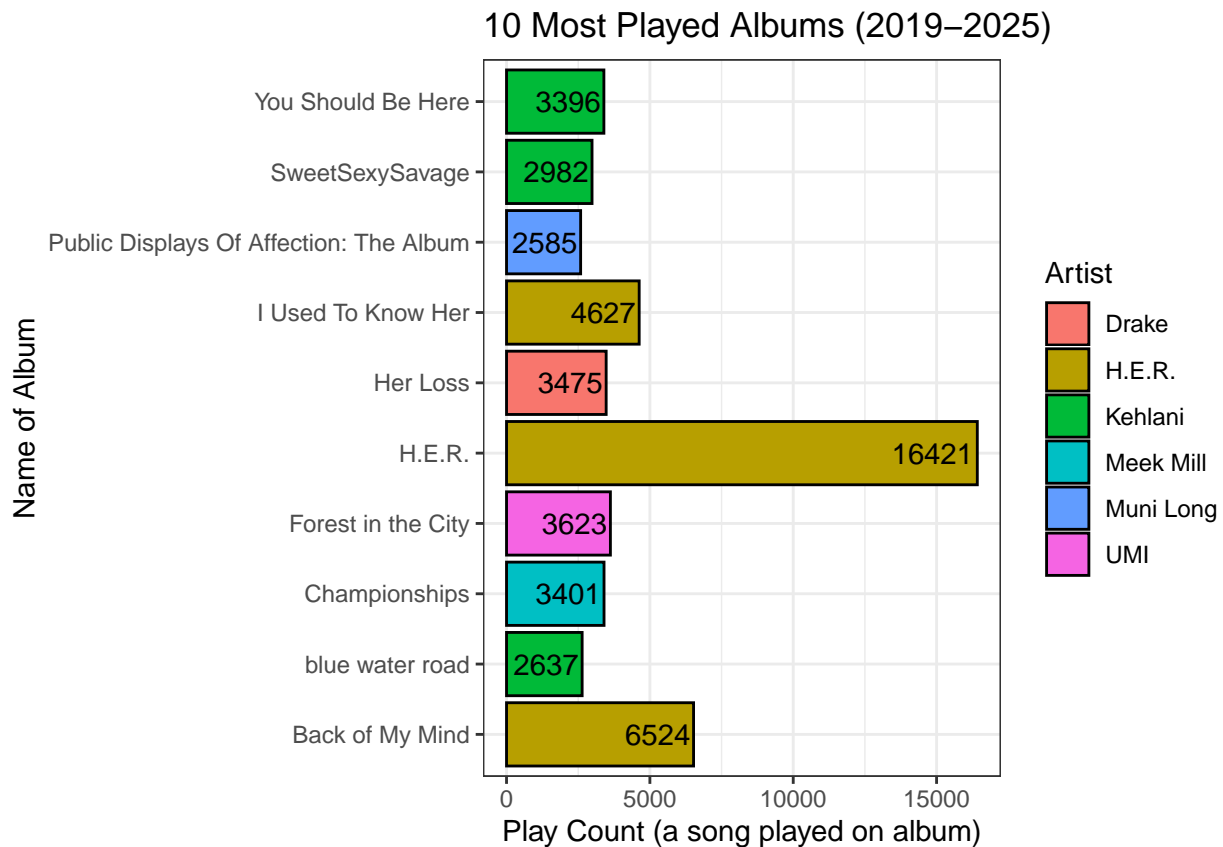


```

#Most Listened to Albums of All Time
top10_albums_oat_bar <- ggplot(alltime_albums %>%
  slice_head(n = 10),
  aes(x = alltime_album_count, y = album,
    fill = artist)) +
  geom_bar(stat = "identity", position = "dodge", color = "black") +
  geom_text(aes(label = alltime_album_count),
    hjust = 1.05,
    size = 4) +
  theme_bw() +
  labs(
    title = "10 Most Played Albums (2019–2025)",
    fill = "Artist",
    y = "Name of Album",
    x = "Play Count (a song played on album)"
  )
top10_albums_oat_bar

```



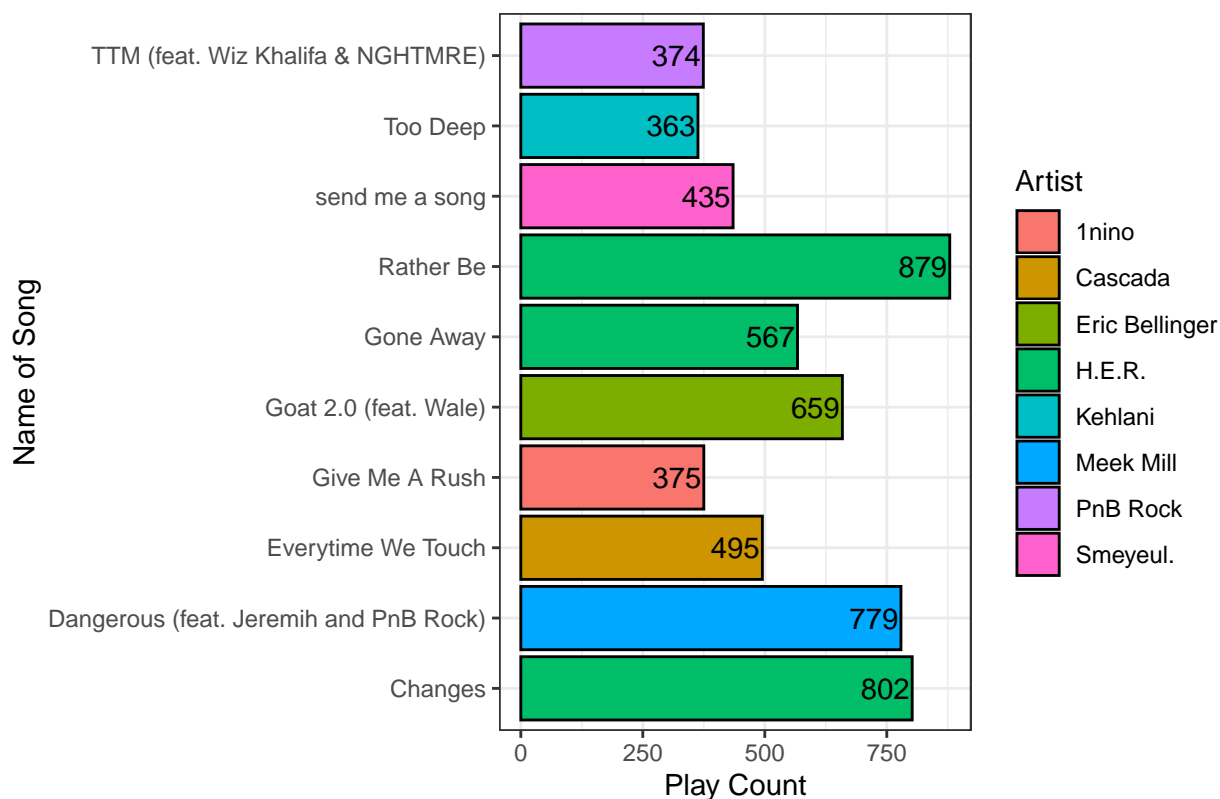


### 10 Most Played Songs by Year

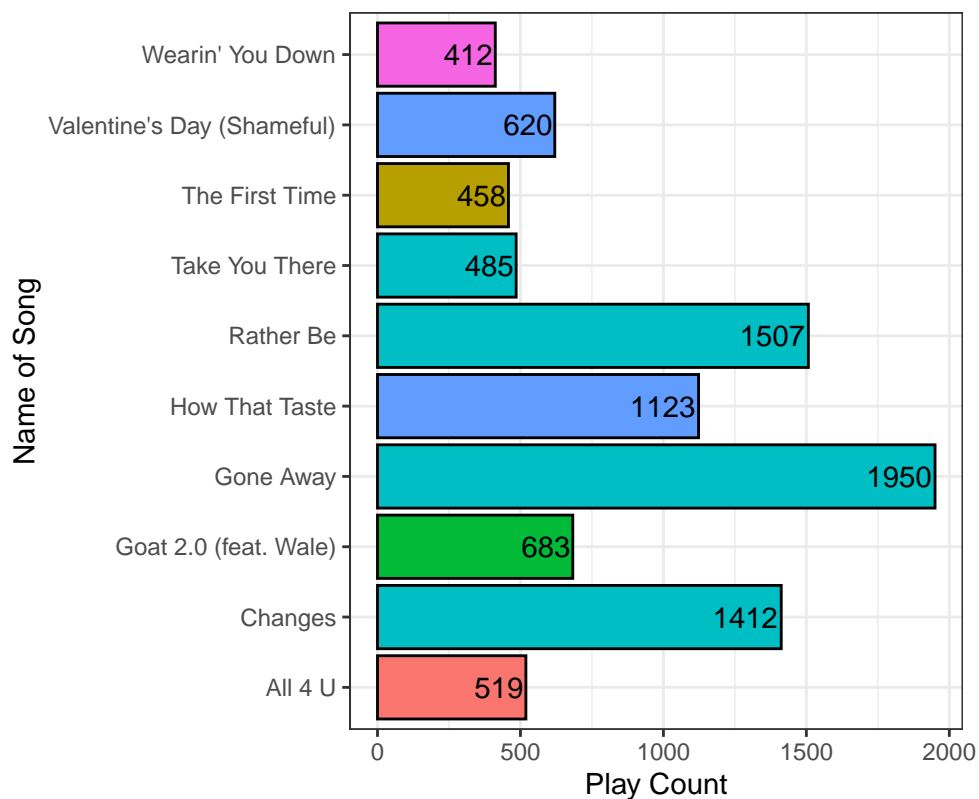
```
#Most Listened to Songs of All Time
music_years <- unique(spotify_data_2$play_year)

for (music_year in music_years) {
  yearly_top10_songs_bar <- ggplot(yearly_top10_songs %>%
    filter(play_year == music_year),
    aes(x = yearly_top10_songs_count, y = song,
        fill = artist)) +
    geom_bar(stat = "identity", position = "dodge", color = "black") +
    geom_text(aes(label = yearly_top10_songs_count),
      hjust = 1.05,
      size = 4) +
    theme_bw() +
    labs(
      title = paste("10 Most Played Songs in", music_year),
      fill = "Artist",
      y = "Name of Song",
      x = "Play Count"
    )
  print(yearly_top10_songs_bar)
}
```

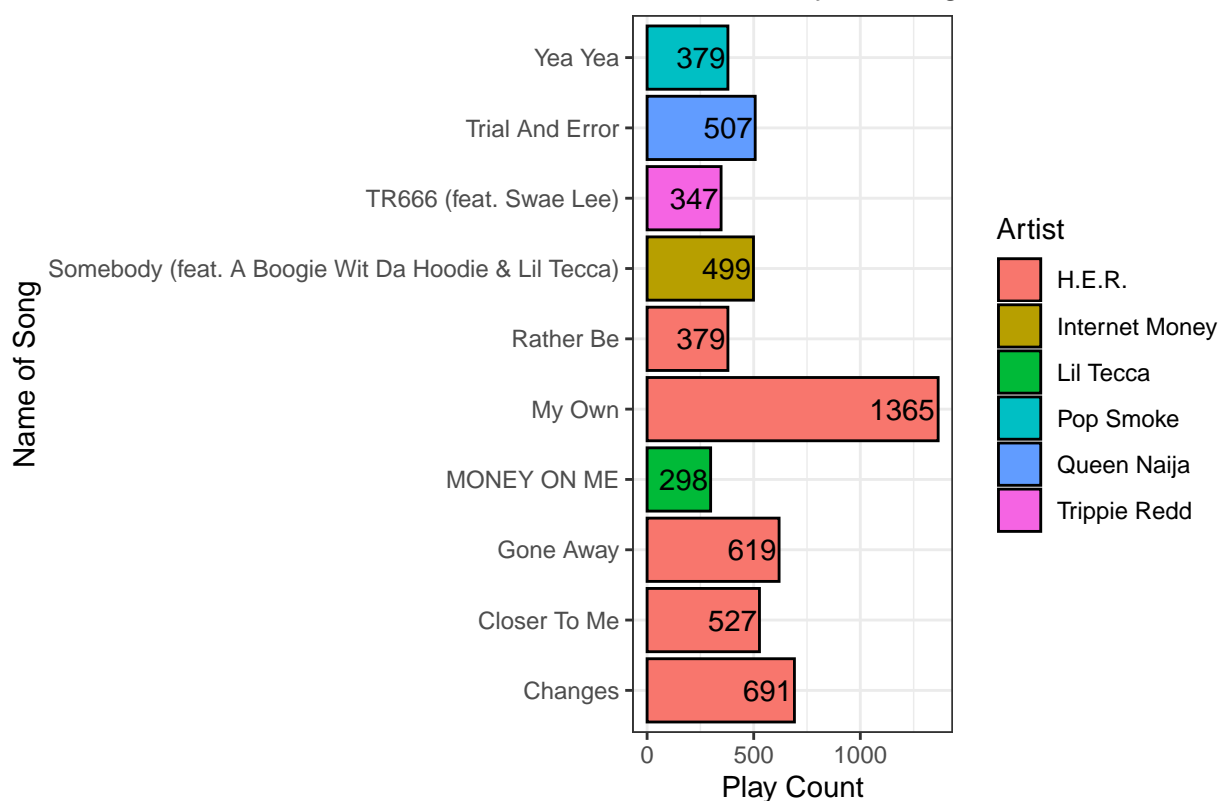
### 10 Most Played Songs in 2019



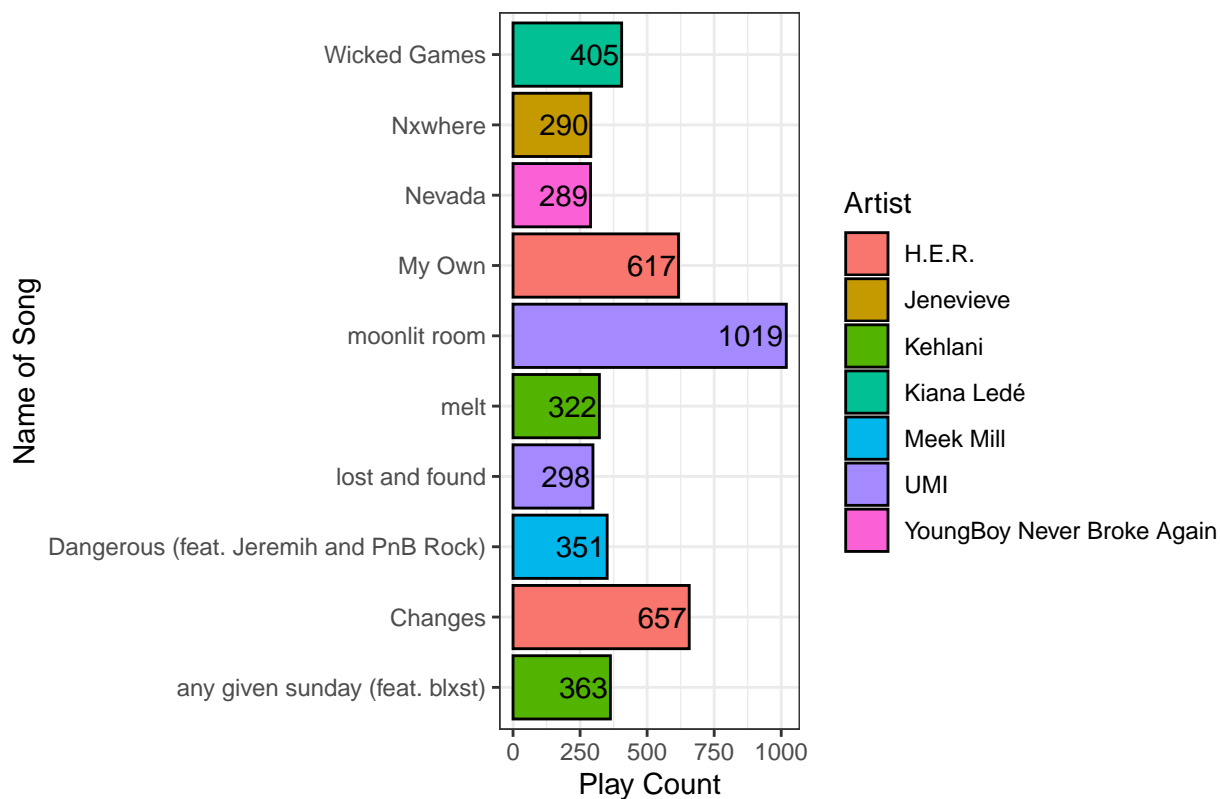
### 10 Most Played Songs in 2020



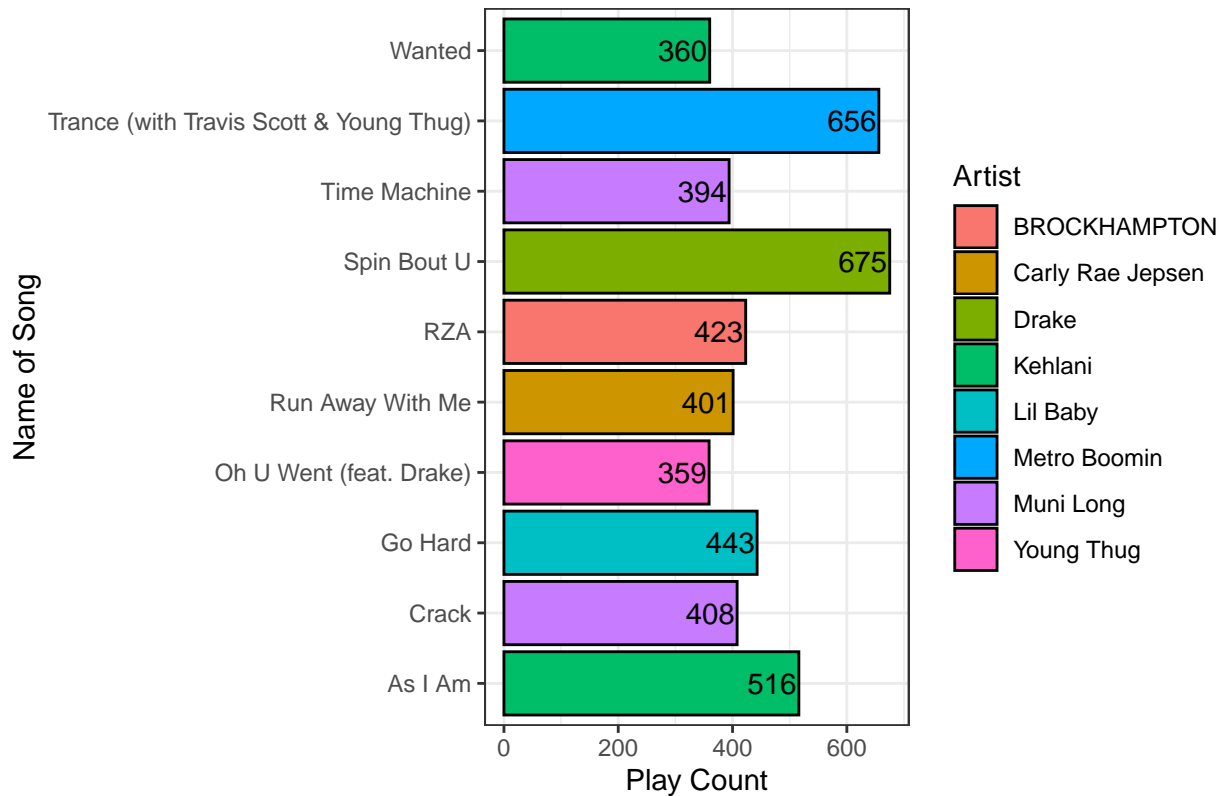
### 10 Most Played Songs in 2021



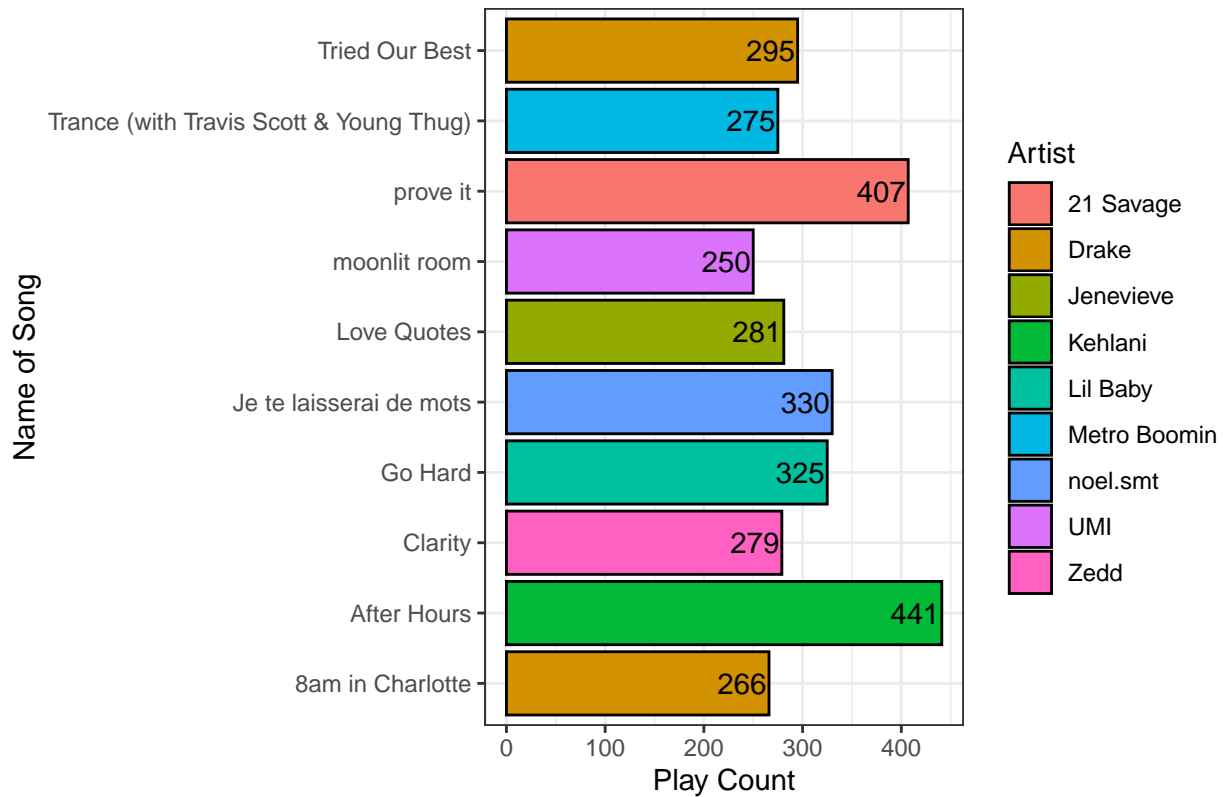
### 10 Most Played Songs in 2022

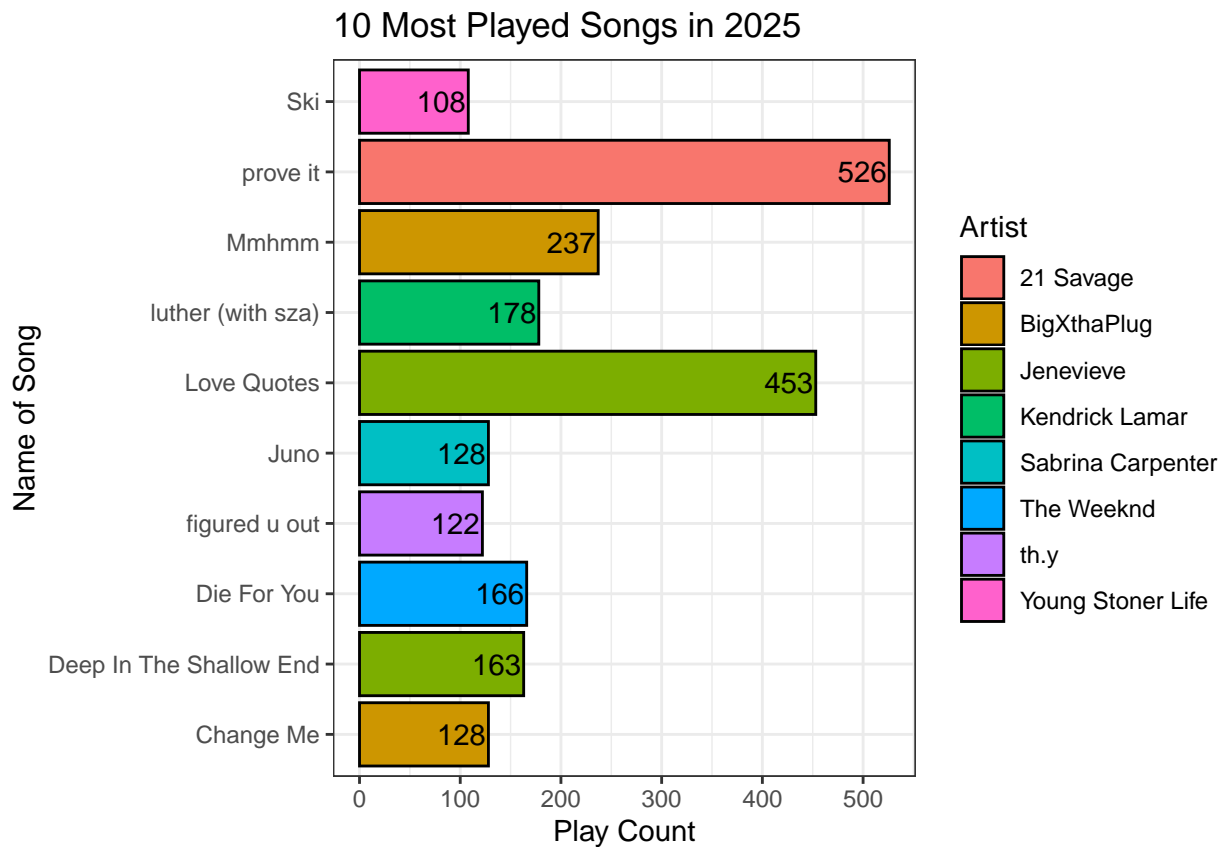


## 10 Most Played Songs in 2023



## 10 Most Played Songs in 2024

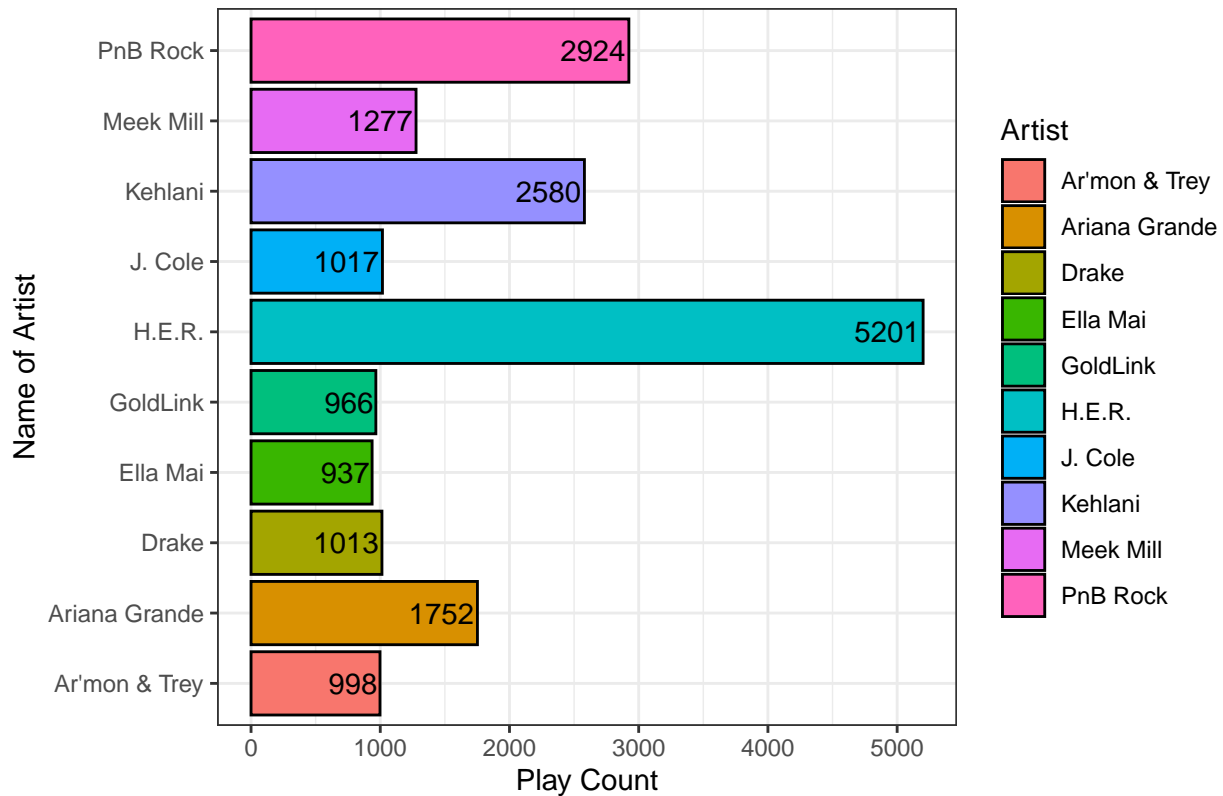




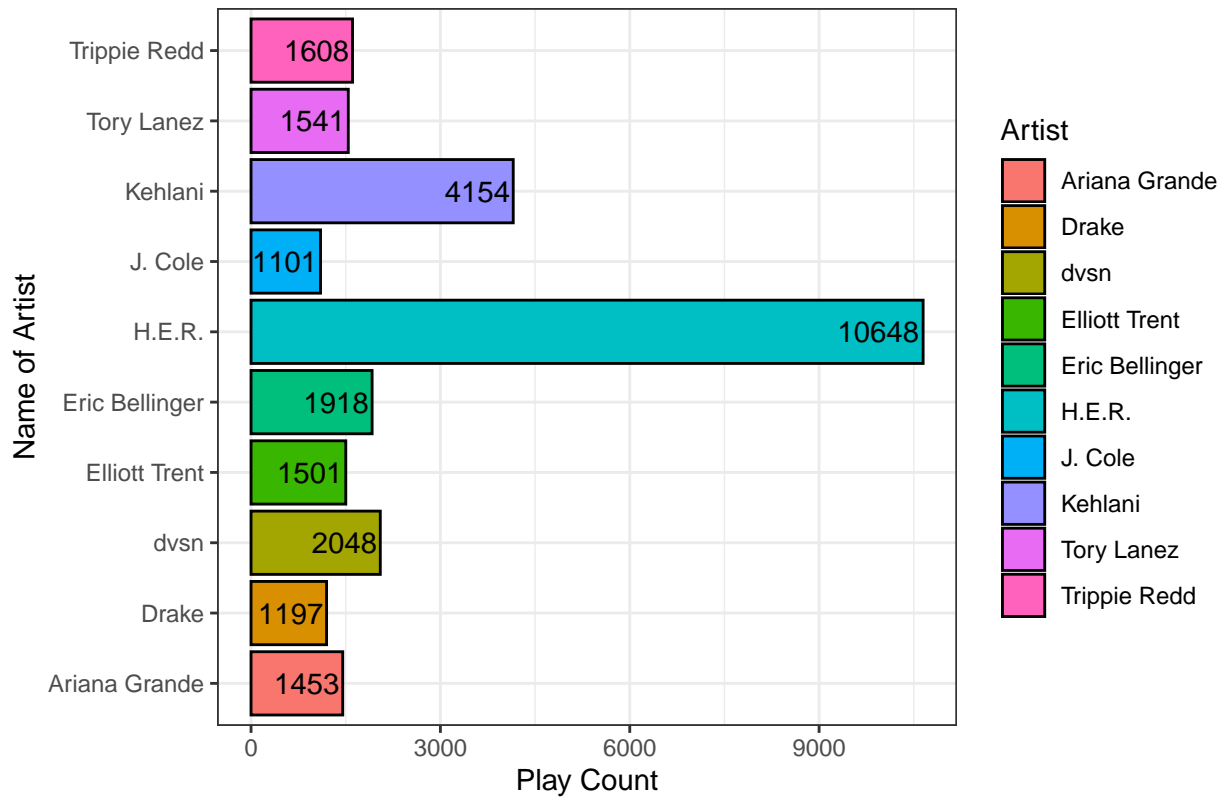
### 10 Most Played Artists by Year

```
for (music_year in music_years) {
  yearly_top10_artists_bar <- ggplot(yearly_top10_artists %>%
    filter(play_year == music_year),
    aes(x = yearly_top10_artists_count, y = artist,
        fill = artist)) +
    geom_bar(stat = "identity", position = "dodge", color = "black") +
    geom_text(aes(label = yearly_top10_artists_count),
      hjust = 1.05,
      size = 4) +
    theme_bw() +
    labs(
      title = paste("10 Most Played Artists in", music_year),
      fill = "Artist",
      y = "Name of Artist",
      x = "Play Count"
    )
  print(yearly_top10_artists_bar)
}
```

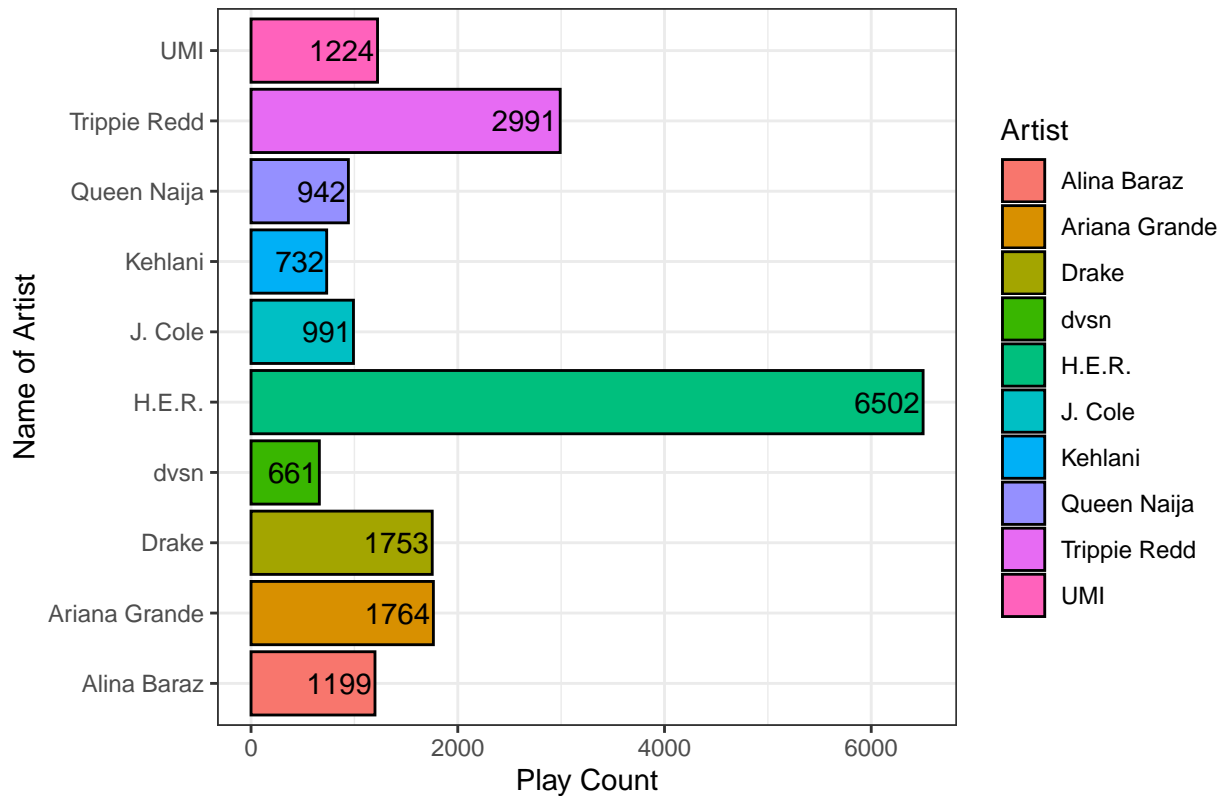
### 10 Most Played Artists in 2019



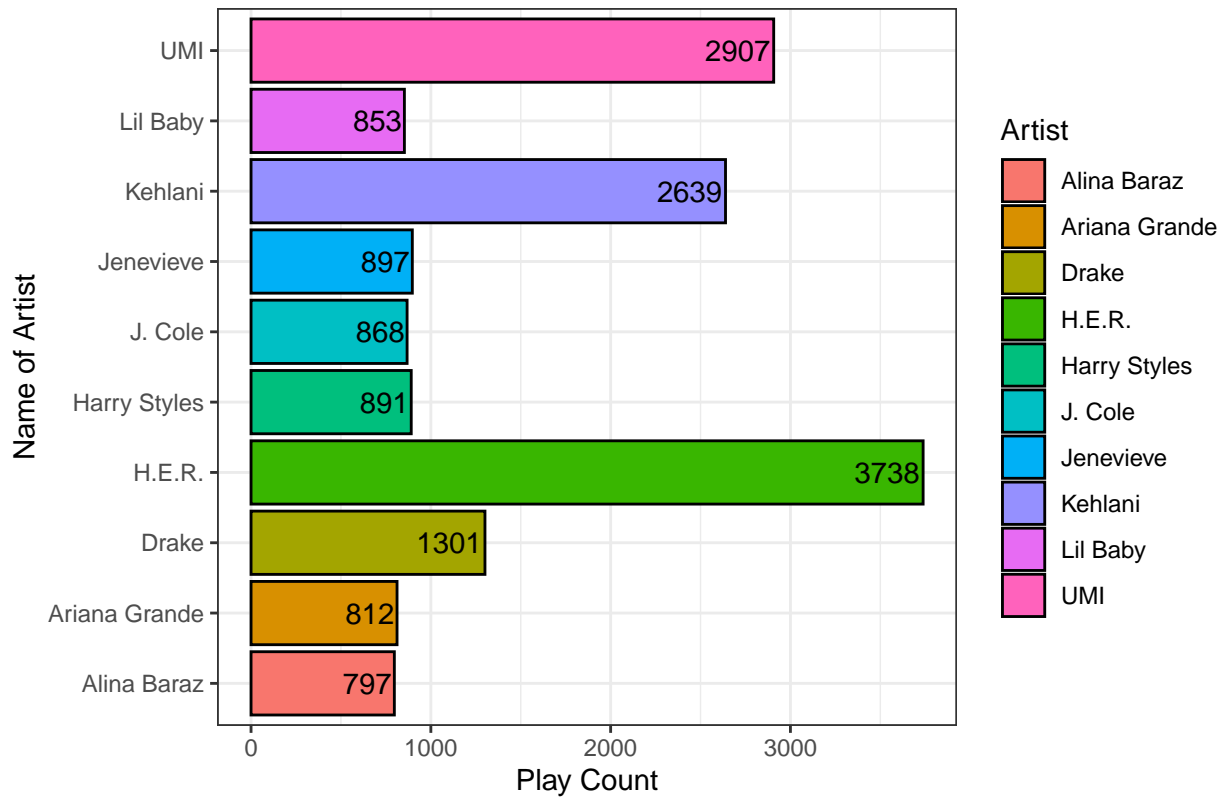
### 10 Most Played Artists in 2020



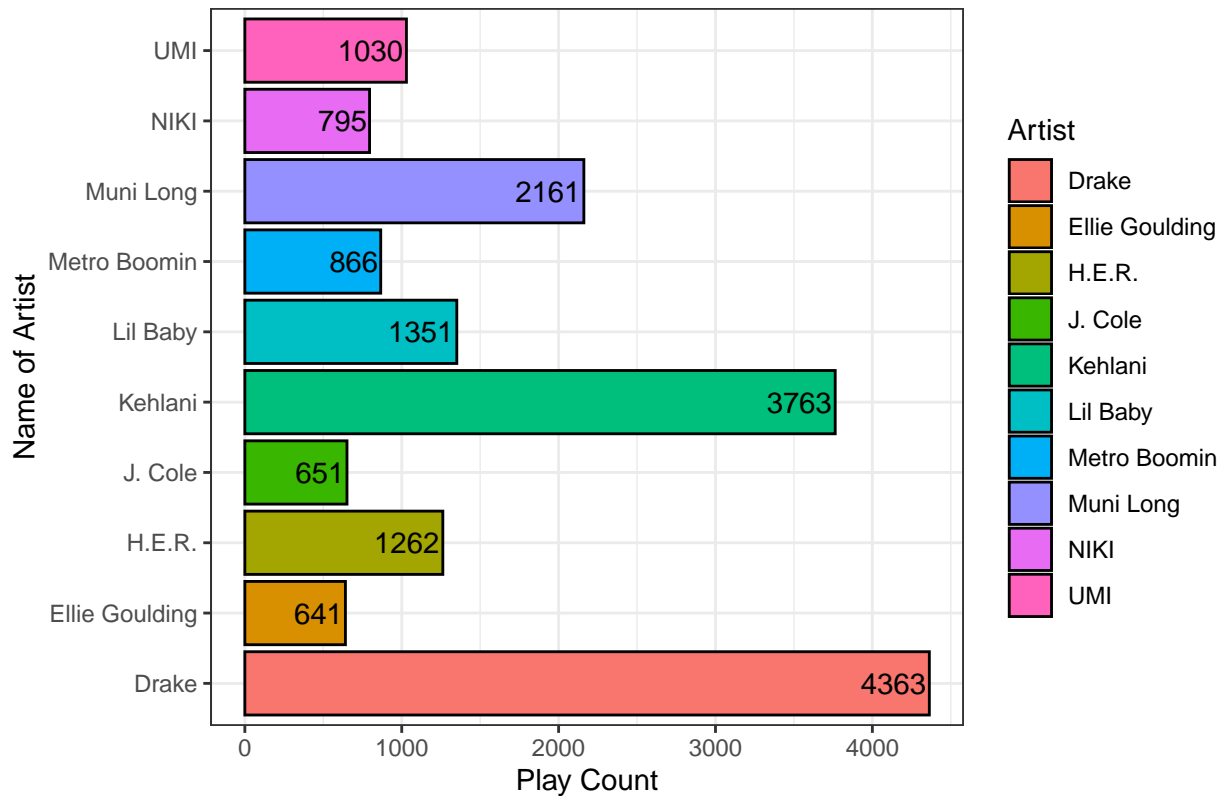
### 10 Most Played Artists in 2021



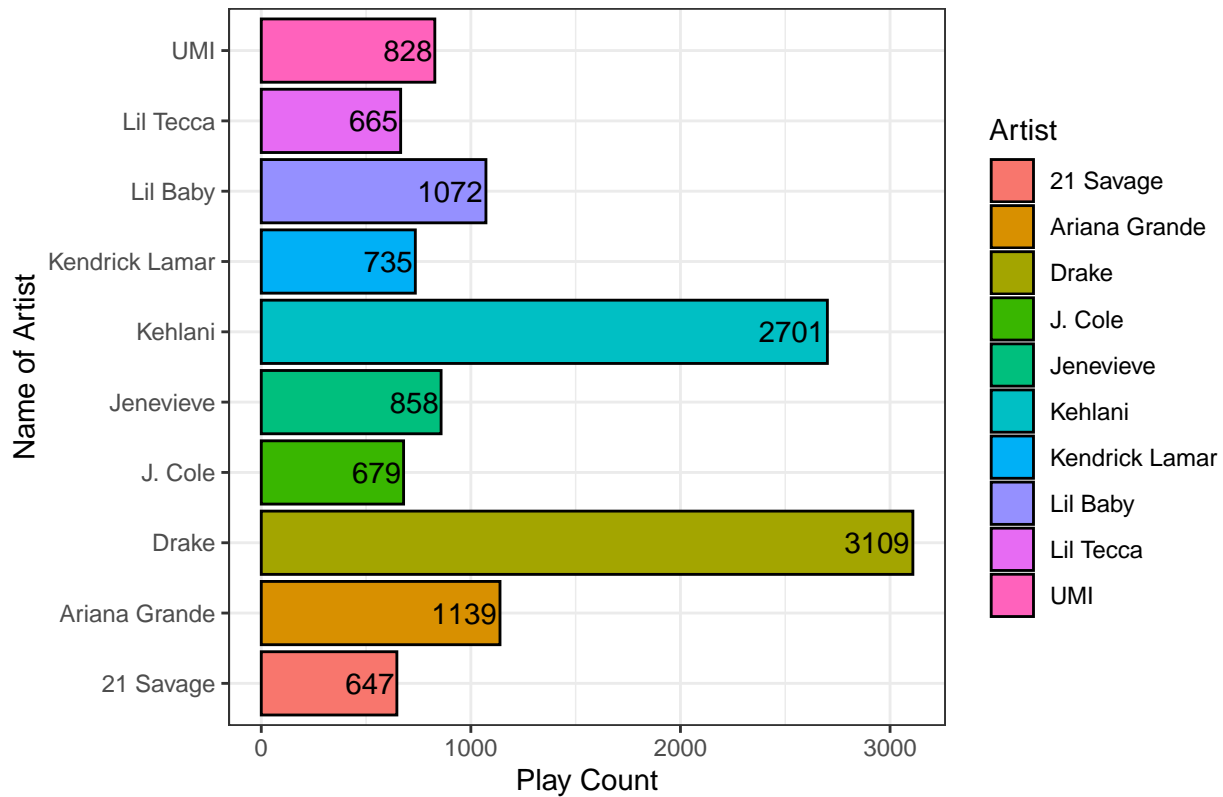
### 10 Most Played Artists in 2022



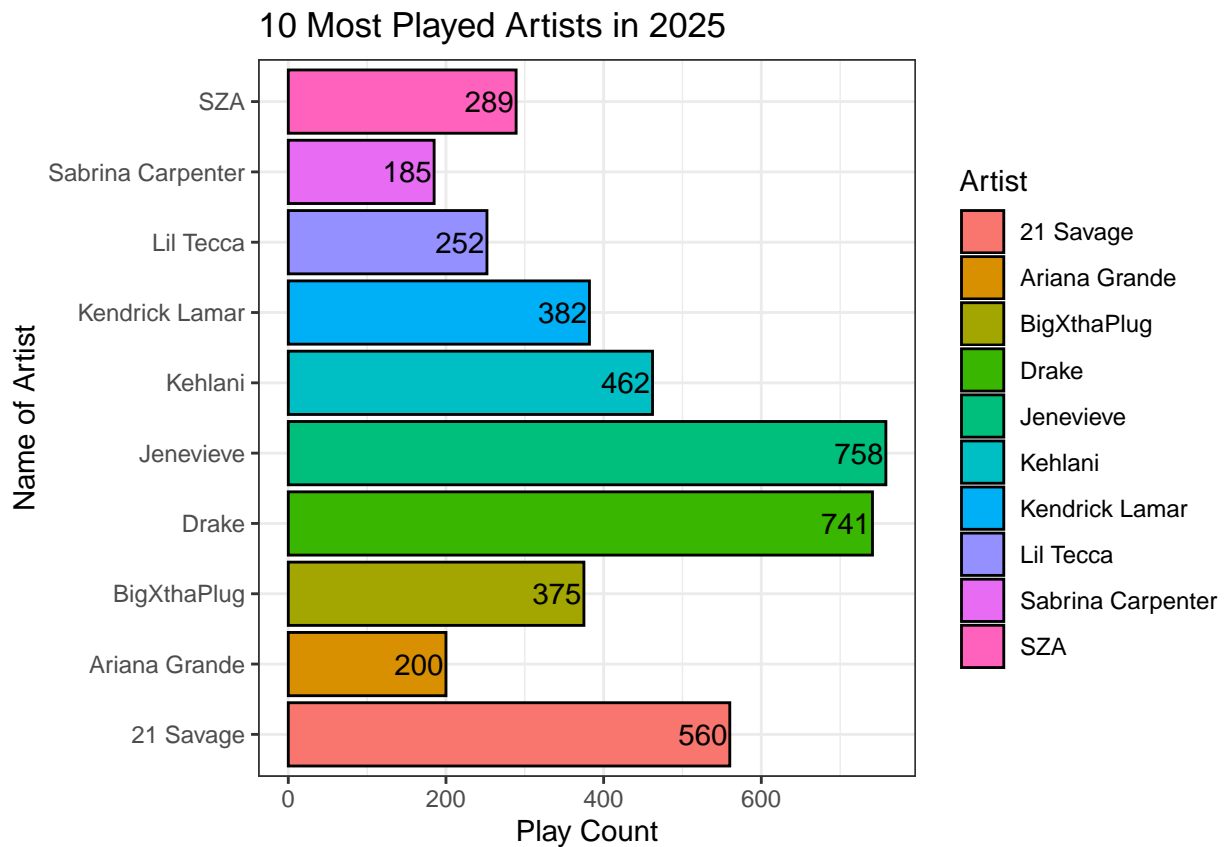
### 10 Most Played Artists in 2023



### 10 Most Played Artists in 2024



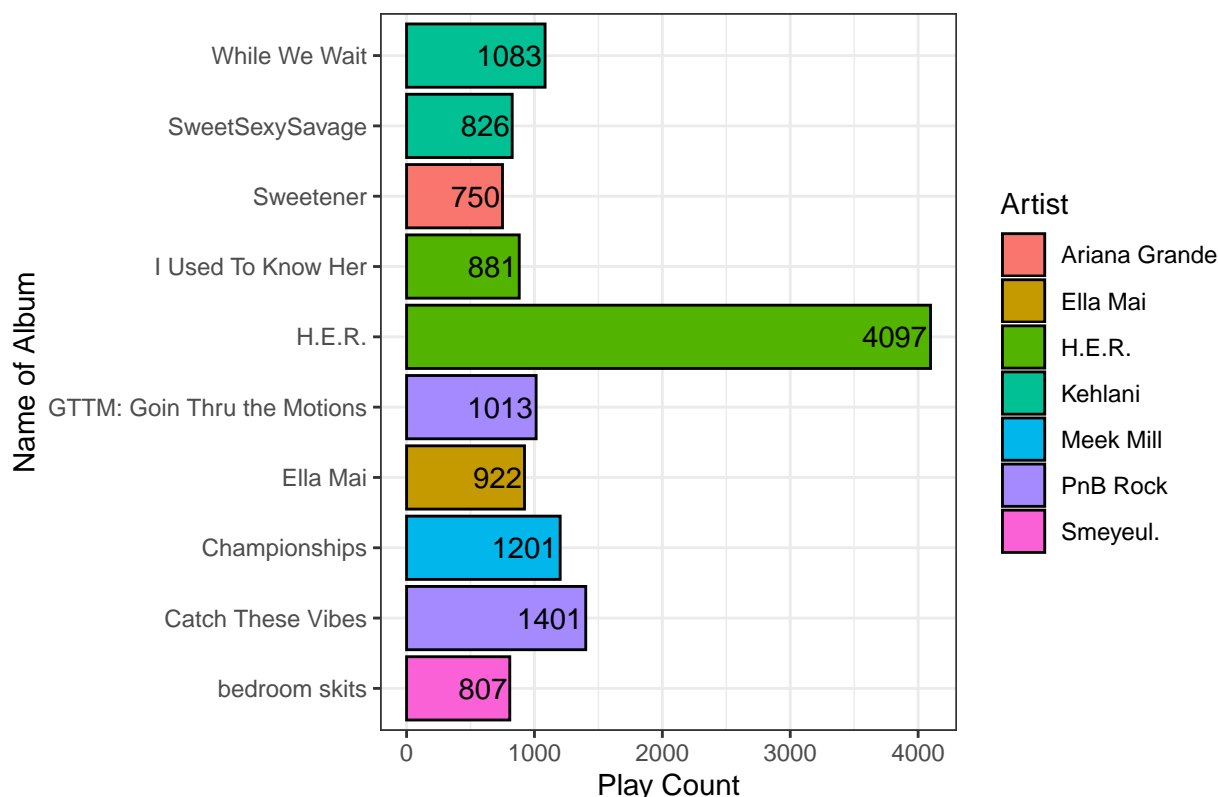




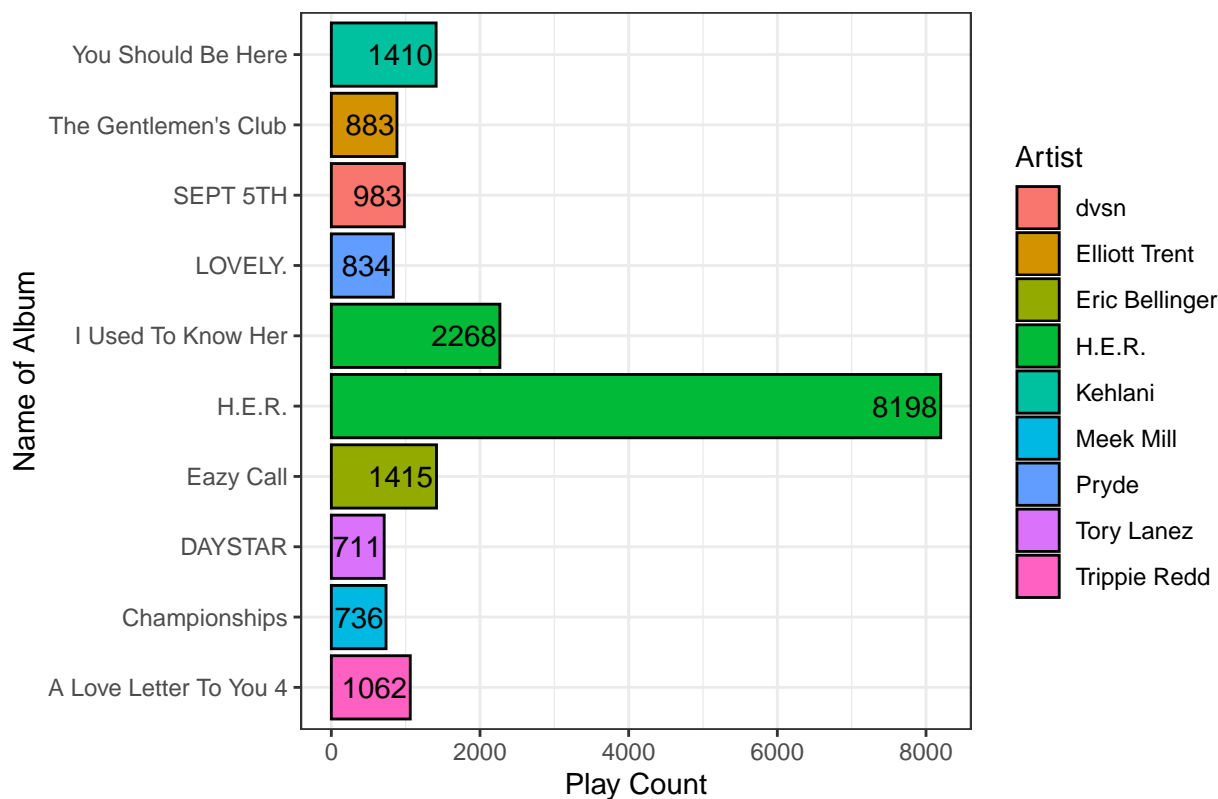
### 10 Most Played Albums by Year

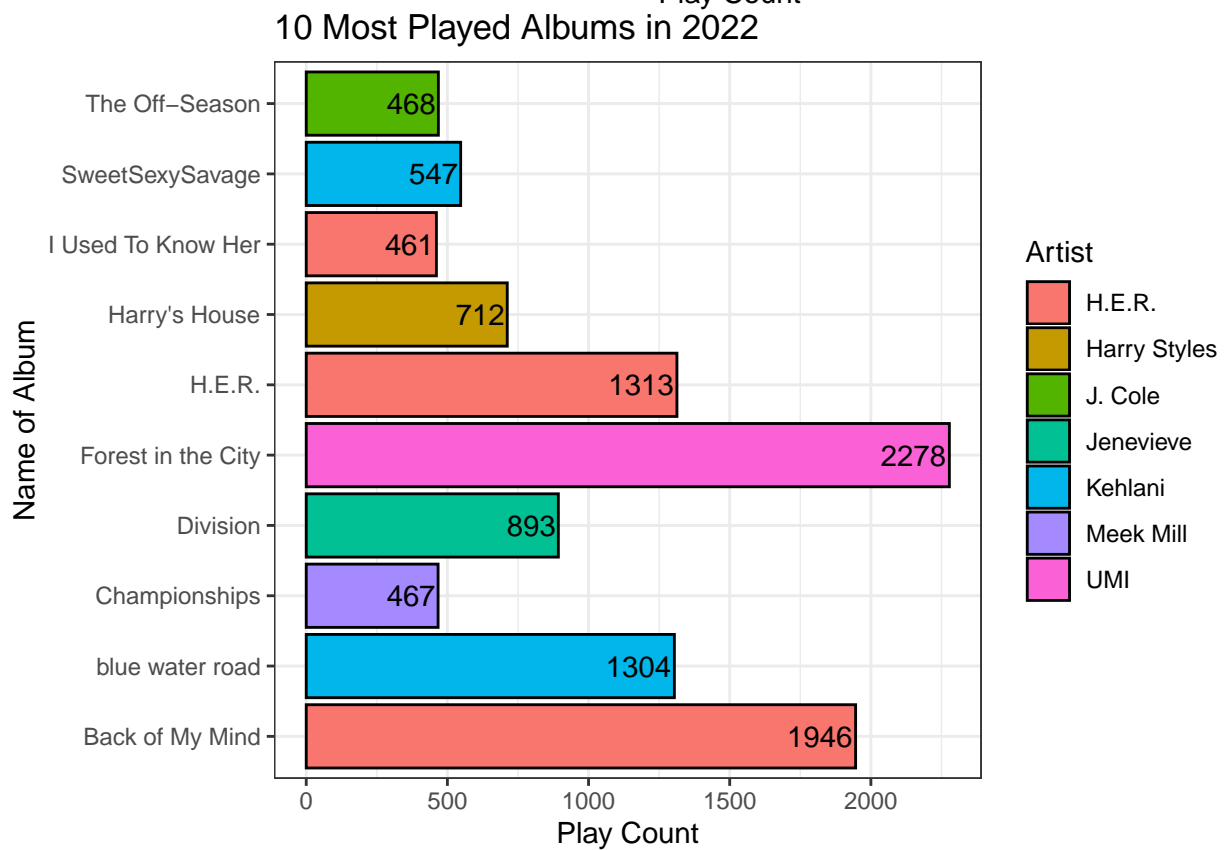
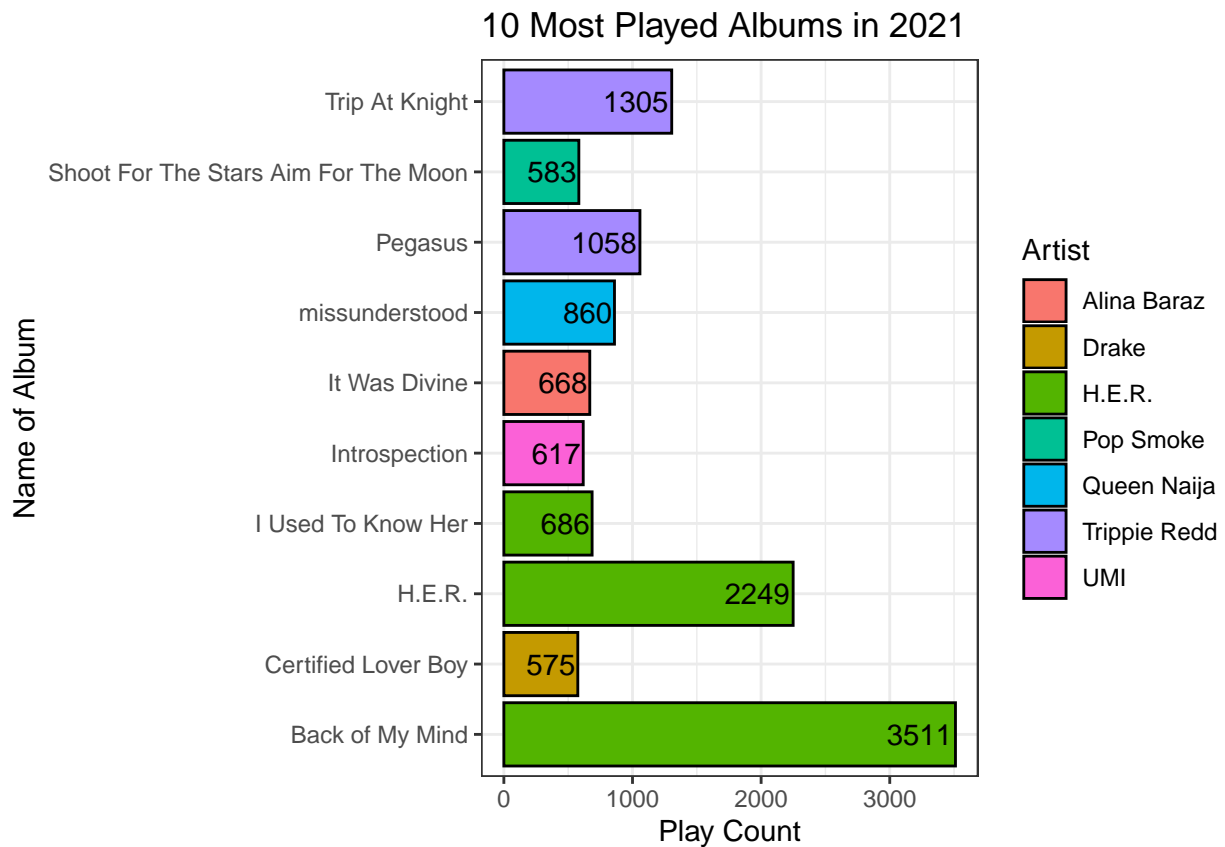
```
for (music_year in music_years) {
  yearly_top10_albums_bar <- ggplot(yearly_top10_albums %>%
    filter(play_year == music_year),
    aes(x = yearly_top10_albums_count, y = album,
        fill = artist)) +
    geom_bar(stat = "identity", position = "dodge", color = "black") +
    geom_text(aes(label = yearly_top10_albums_count),
      hjust = 1.05,
      size = 4) +
    theme_bw() +
    labs(
      title = paste("10 Most Played Albums in", music_year),
      fill = "Artist",
      y = "Name of Album",
      x = "Play Count"
    )
  print(yearly_top10_albums_bar)
}
```

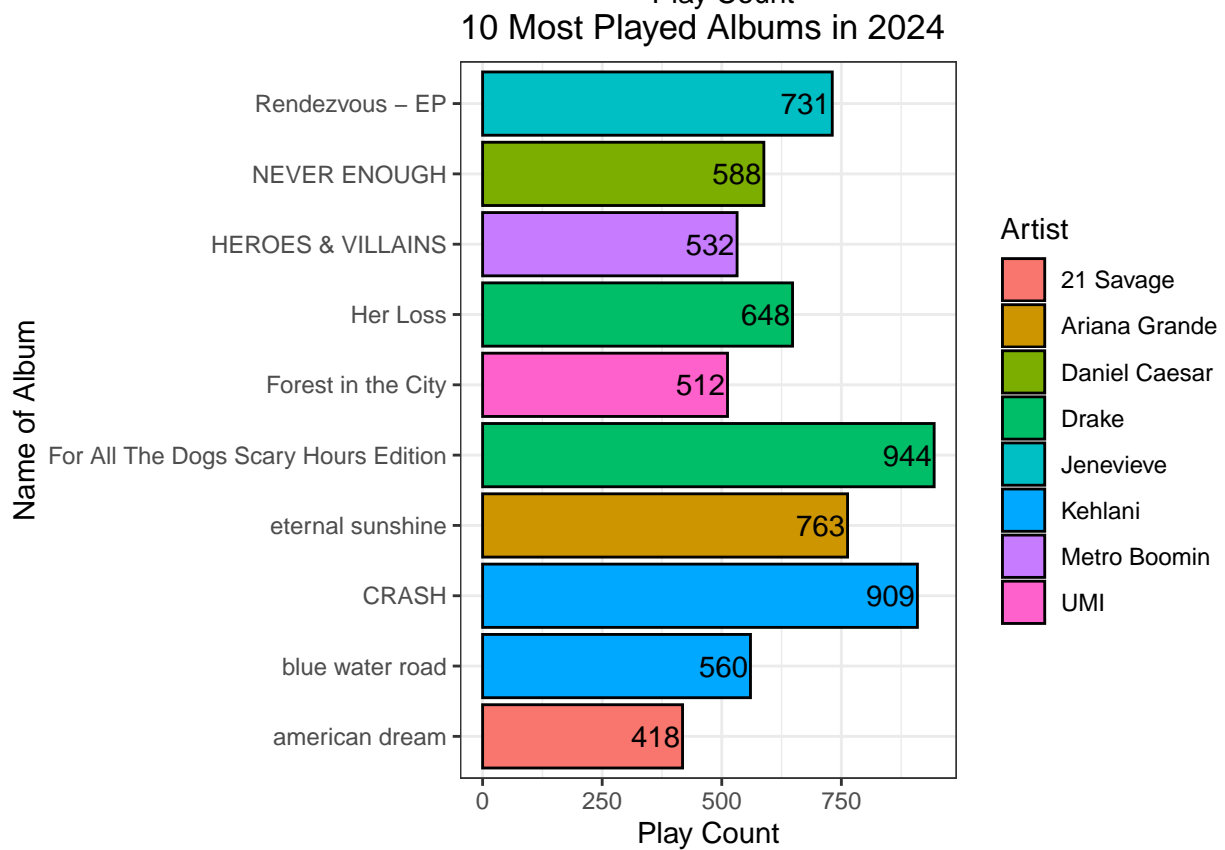
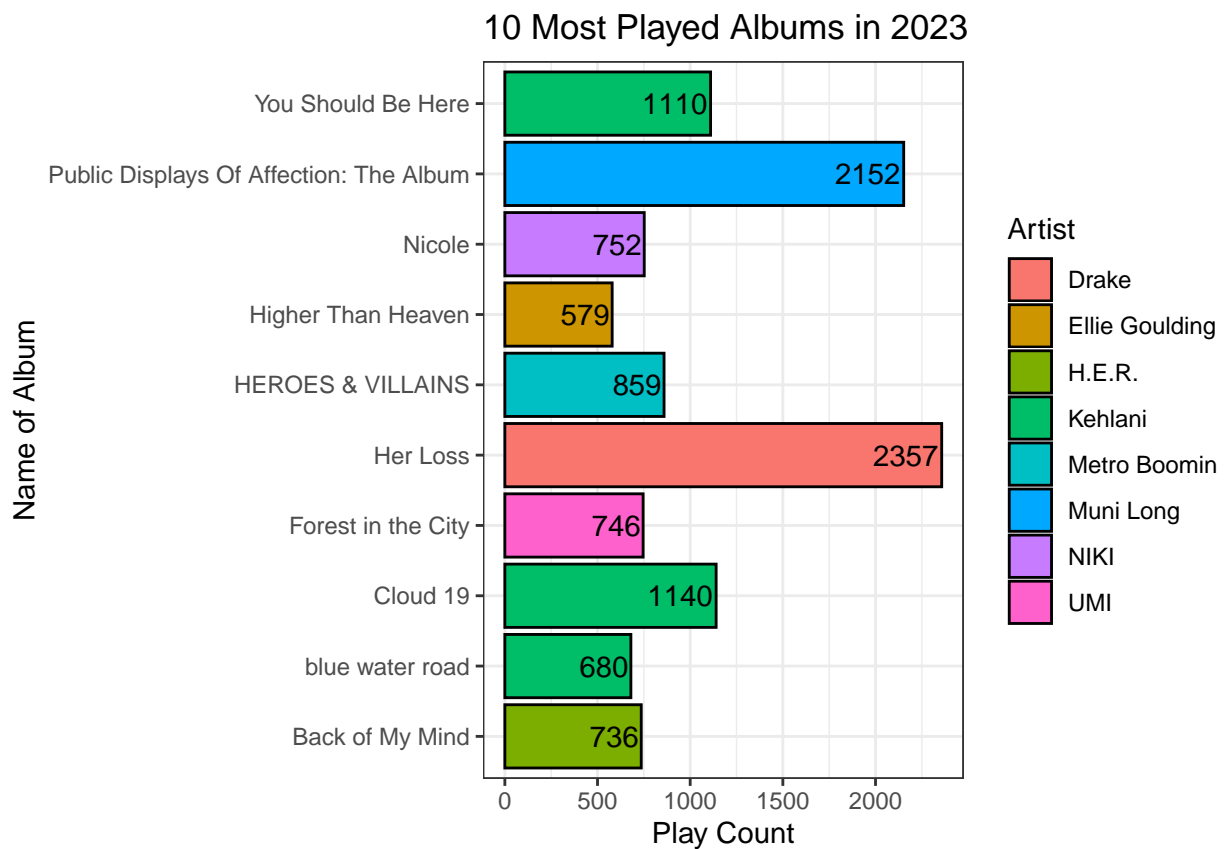
### 10 Most Played Albums in 2019

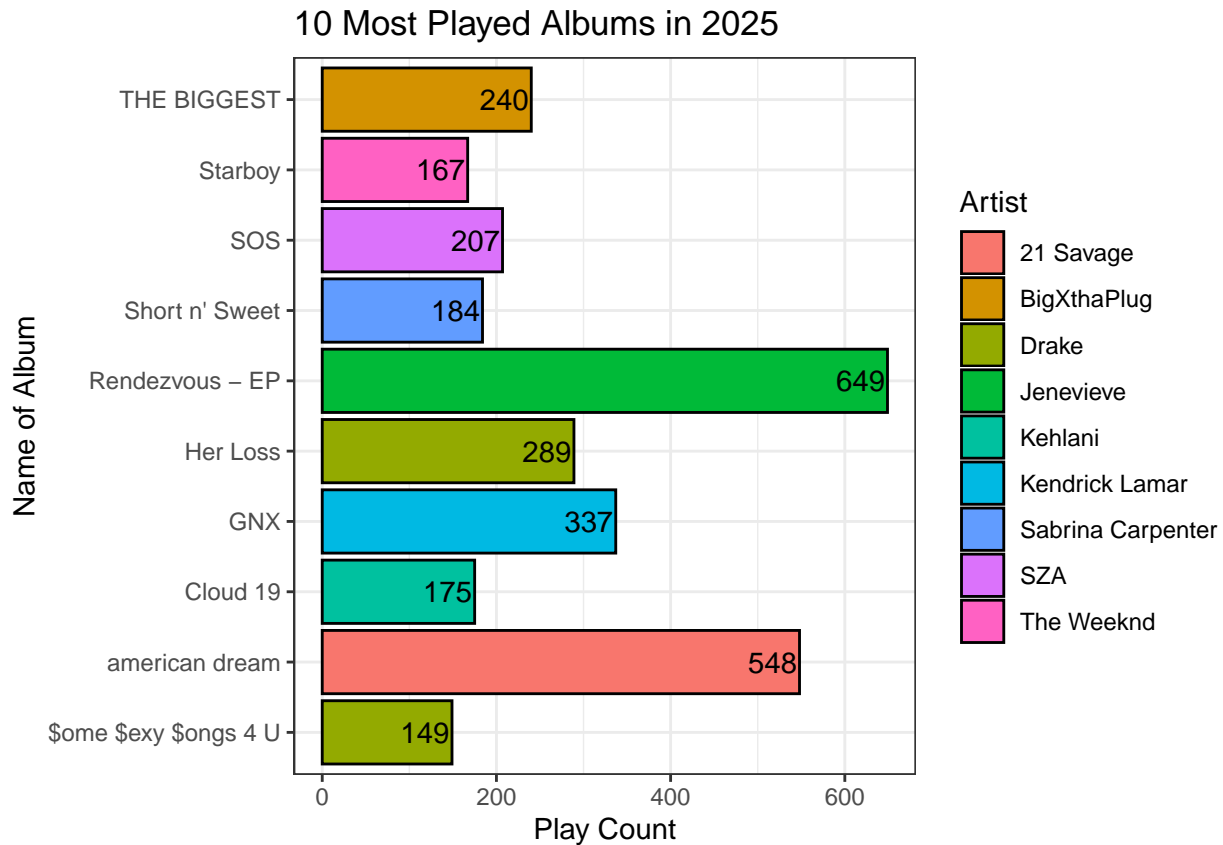


### 10 Most Played Albums in 2020





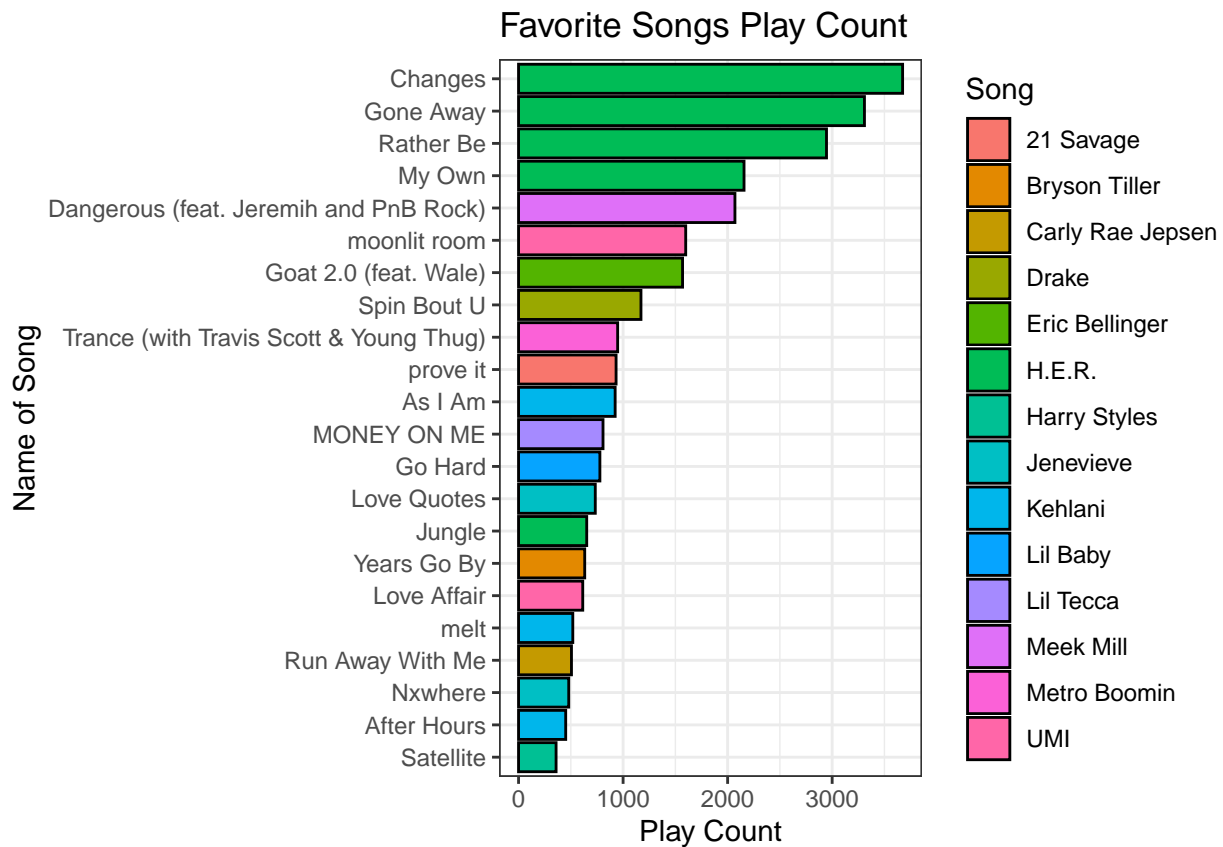




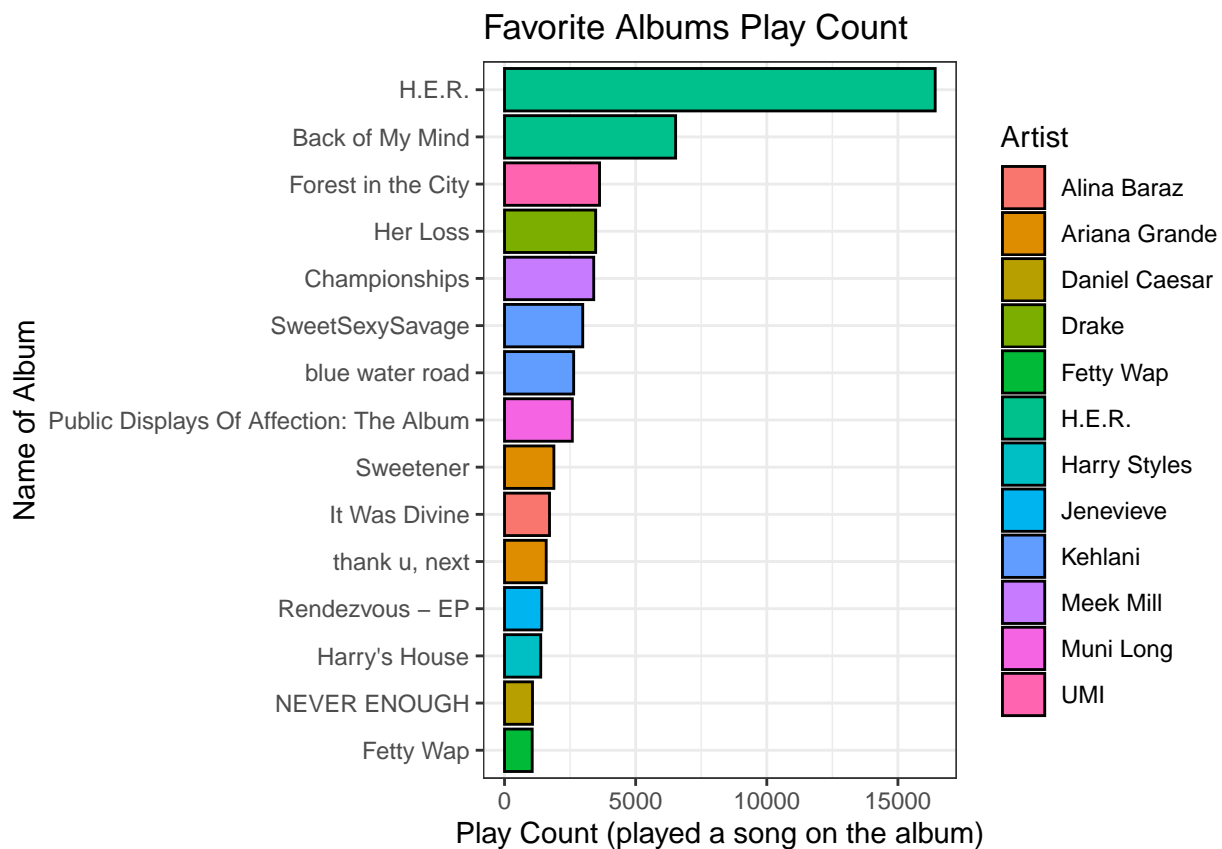
## Favorite Songs/Artists/Albums Statistics

### Play Count of Favorite Songs/Artists/Albums

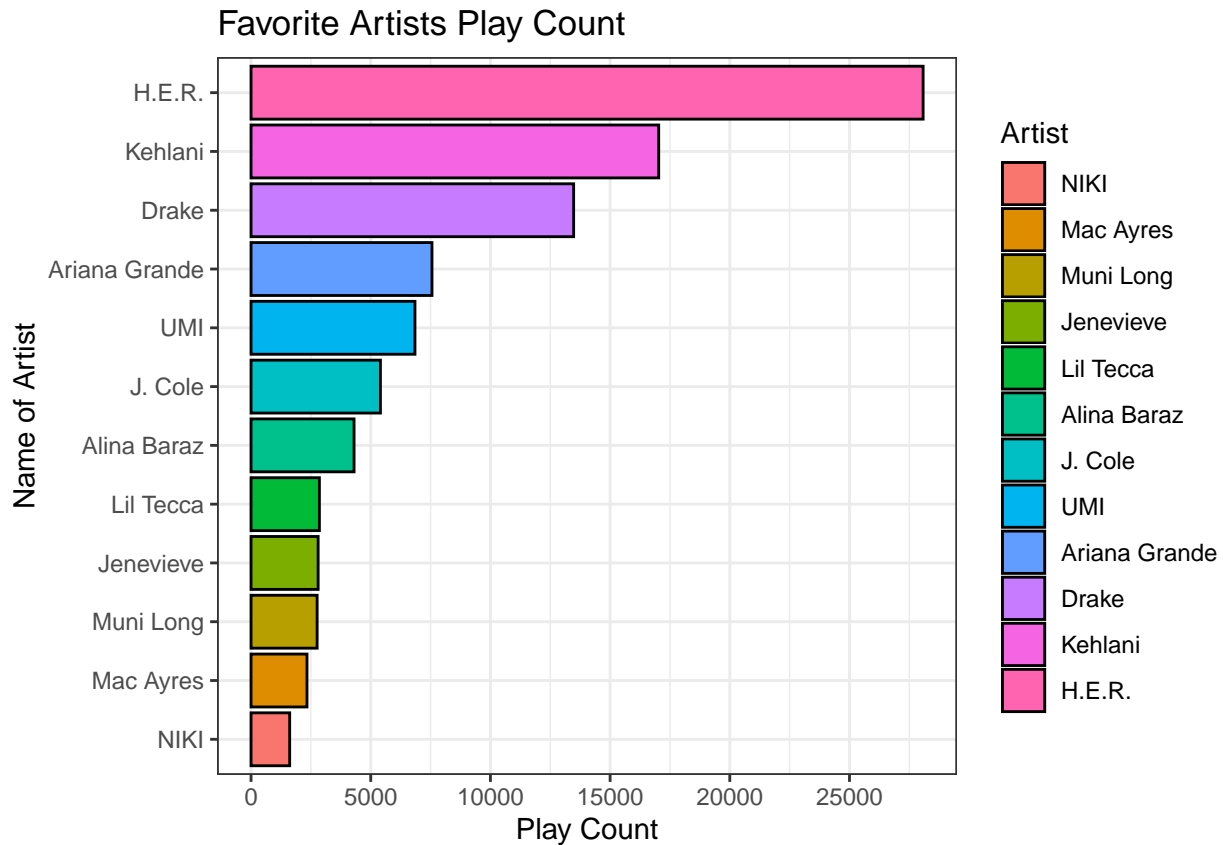
```
fav_songs_oat_bar <- ggplot(alltime_songs_plot_data %>%
  filter(favorite_song == "Yes") %>%
  mutate(
    song = fct_reorder(song, alltime_song_count)
  ) %>%
  distinct(song, artist, alltime_song_count),
  aes(x = alltime_song_count, y = song, fill = artist)) +
geom_bar(stat = "identity", position = "dodge", color = "black") +
theme_bw() +
labs(
  title = "Favorite Songs Play Count",
  fill = "Song",
  y = "Name of Song",
  x = "Play Count"
)
fav_songs_oat_bar
```



```
fav_albums_oat_bar <- ggplot(alltime_albums_plot_data %>%
  filter(favorite_album == "Yes") %>%
  mutate(
    album = fct_reorder(album, alltime_album_count)
  ) %>%
  distinct(album, artist, alltime_album_count),
  aes(x = alltime_album_count, y = album, fill = artist)) +
geom_bar(stat = "identity", position = "dodge", color = "black") +
theme_bw() +
labs(
  title = "Favorite Albums Play Count",
  fill = "Artist",
  y = "Name of Album",
  x = "Play Count (played a song on the album)"
)
fav_albums_oat_bar
```



```
fav_artists_oat_bar <- ggplot(alltime_artists_plot_data %>%
  filter(favorite_artist == "Yes") %>%
  mutate(
    artist = fct_reorder(artist,
                        alltime_artist_count)
  ) %>%
  distinct(artist, alltime_artist_count),
  aes(x = alltime_artist_count, y = artist,
      fill = artist)) +
  geom_bar(stat = "identity", position = "dodge", color = "black") +
  scale_x_continuous(
    breaks = seq(0, max(alltime_artists_plot_data$alltime_artist_count),
                 by = 5000)) +
  theme_bw() +
  labs(
    title = "Favorite Artists Play Count",
    fill = "Artist",
    y = "Name of Artist",
    x = "Play Count"
  )
fav_artists_oat_bar
```

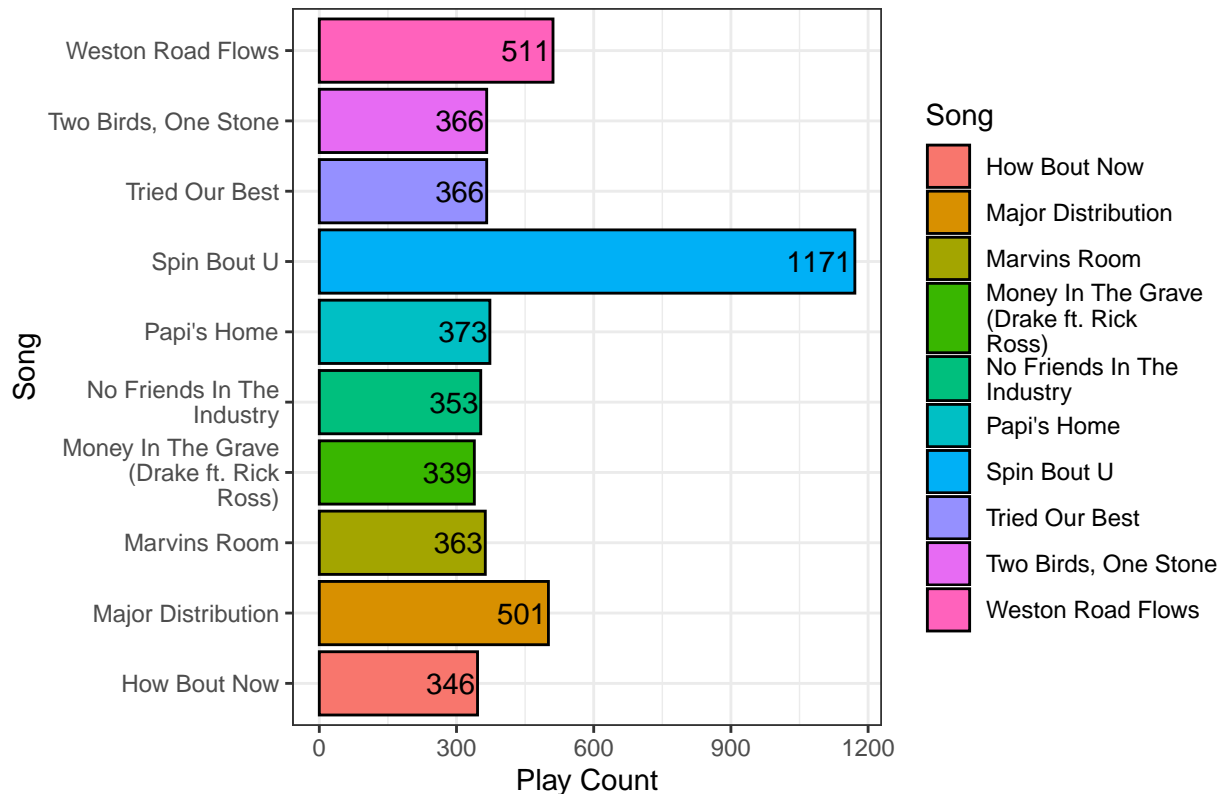


### 10 Most Played Songs for Favorite Artist

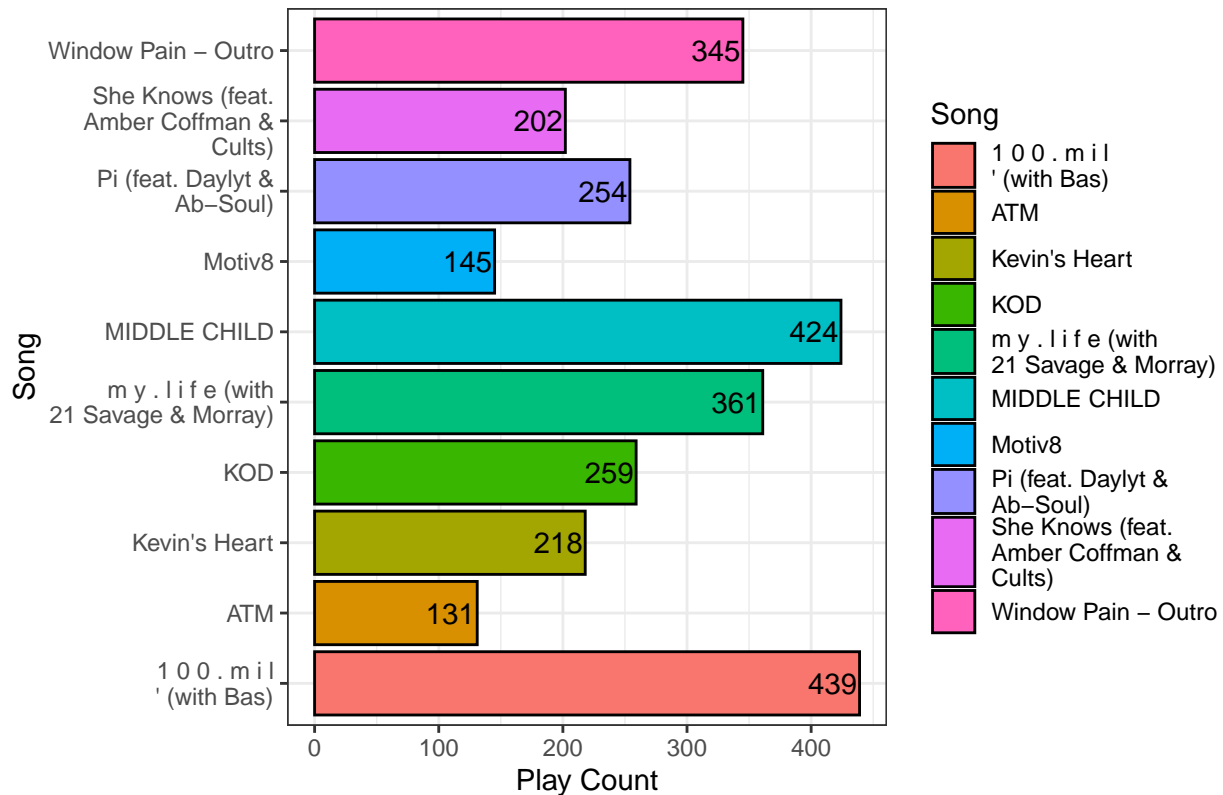
```
for (music_artist in favorite_artists$artist) {
  artist_top10_songs_bar <- ggplot(alltime_songs %>%
    mutate(song = str_wrap(song, width = 20)) %>%
    filter(artist == music_artist) %>%
    slice_head(n = 10),
    aes(x = alltime_song_count, y = song,
        fill = song)) +
  geom_bar(stat = "identity", position = "dodge", color = "black") +
  geom_text(aes(label = alltime_song_count),
    hjust = 1.05,
    size = 4) +
  theme_bw() +
  labs(
    title = paste(music_artist, "Top 10 Played Songs"),
    y = "Song",
    x = "Play Count",
    fill = "Song"
  )
  print(artist_top10_songs_bar)
}
```



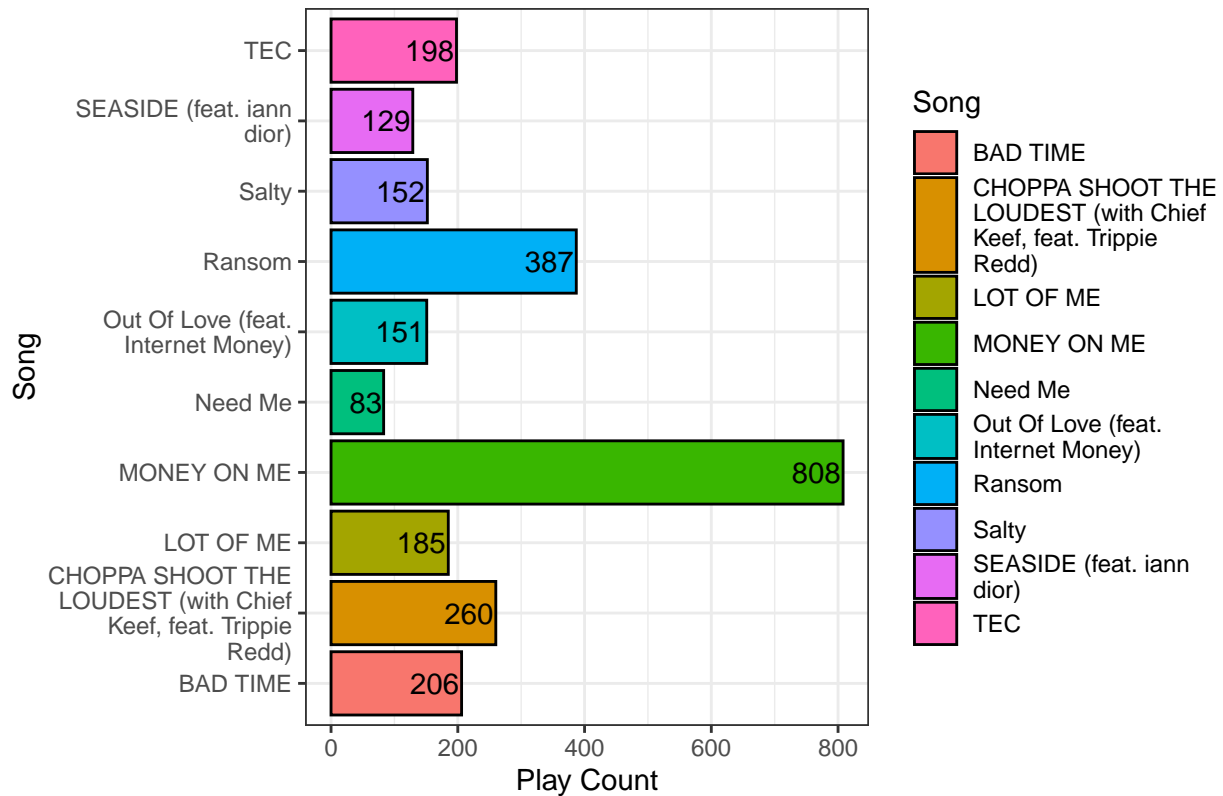
### Drake Top 10 Played Songs



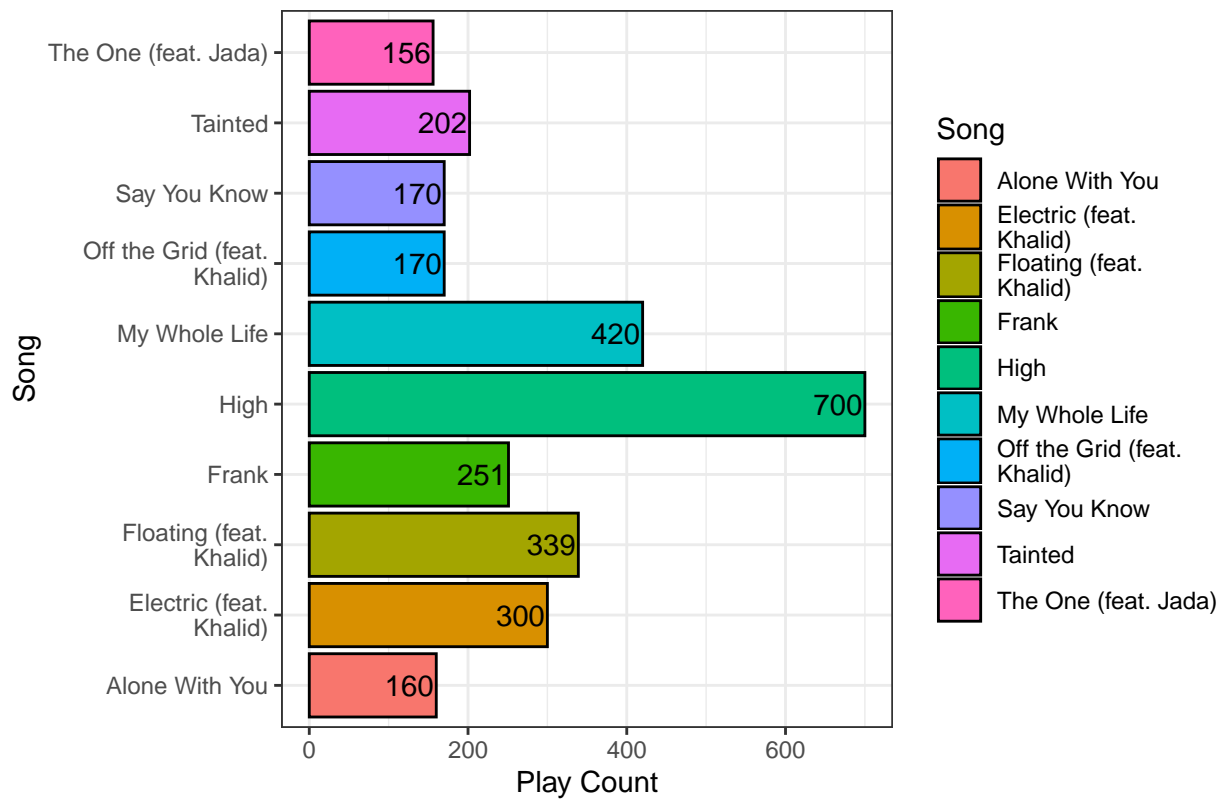
### J. Cole Top 10 Played Songs



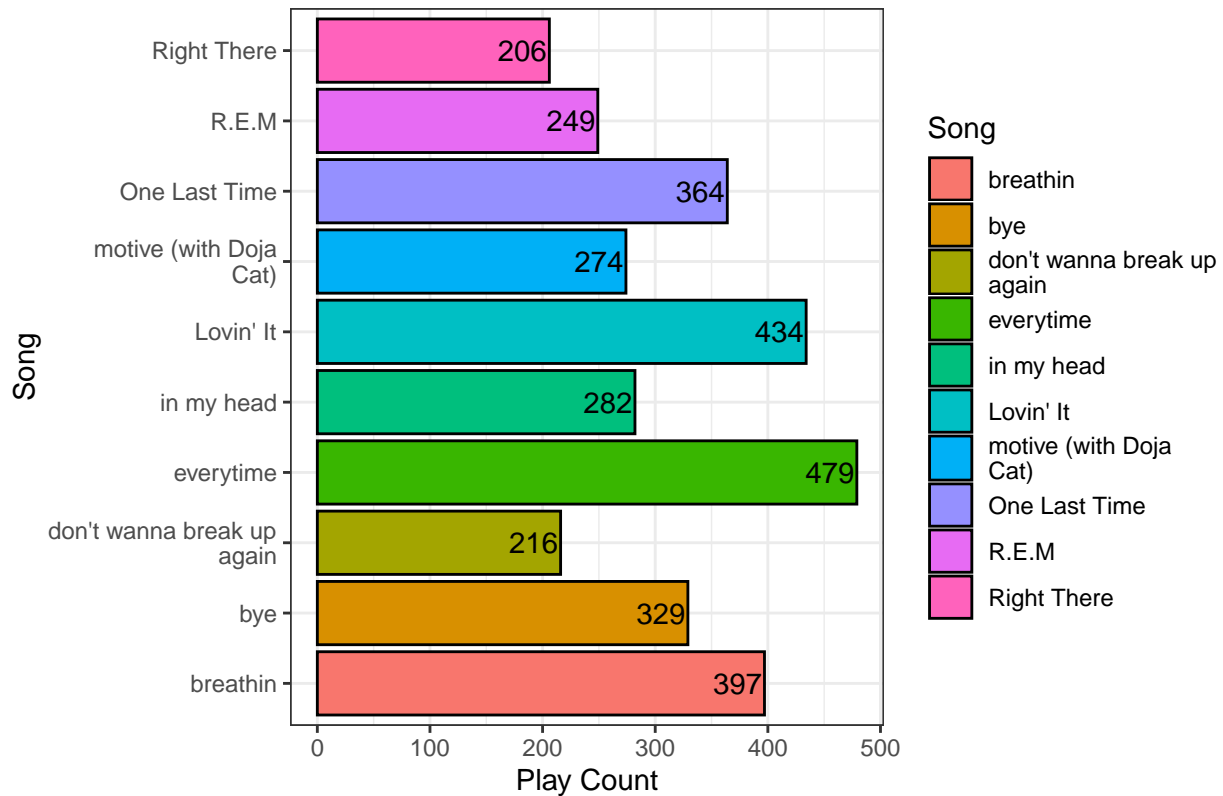
### Lil Tecca Top 10 Played Songs



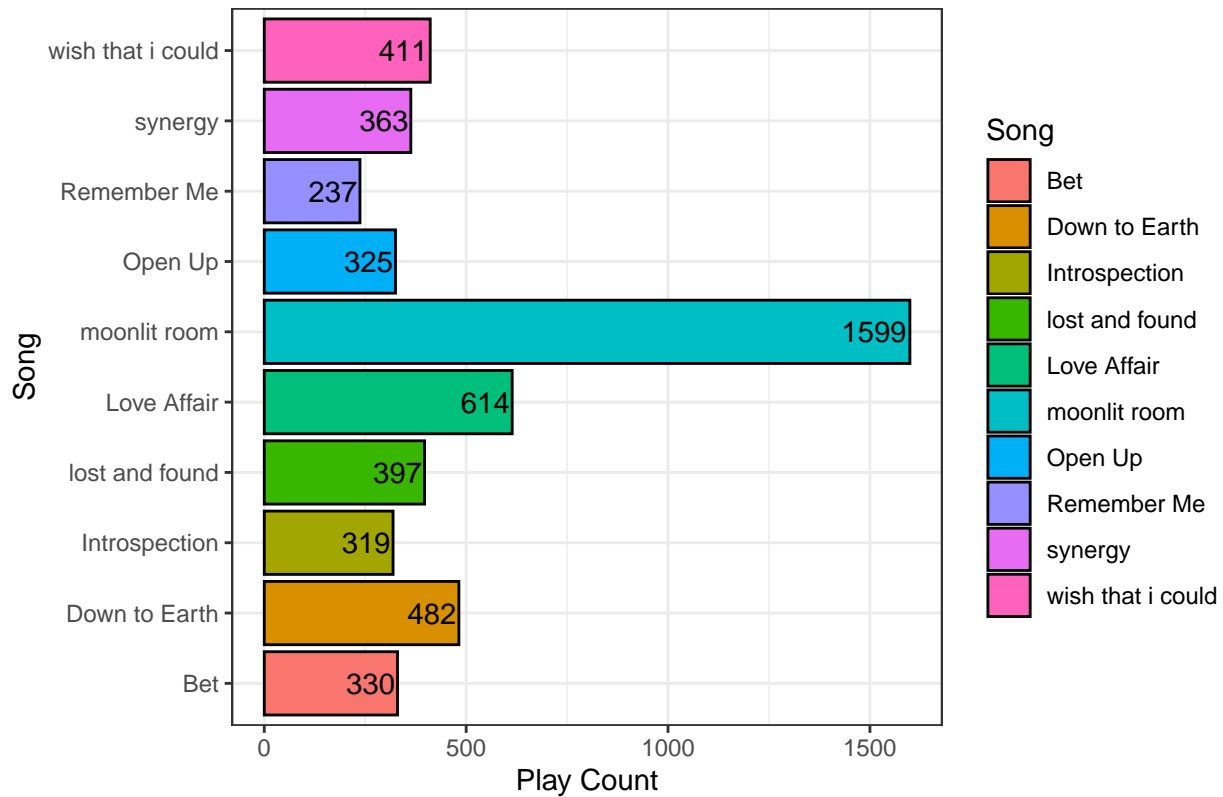
### Alina Baraz Top 10 Played Songs



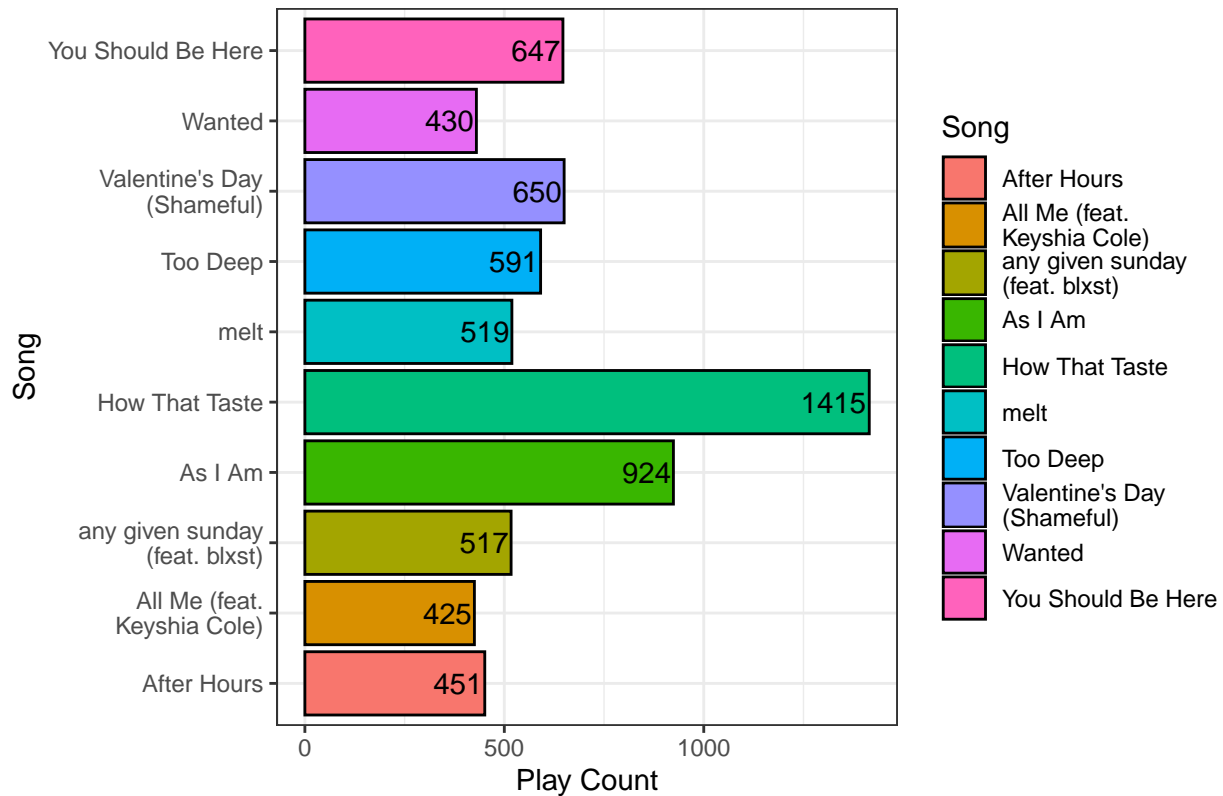
### Ariana Grande Top 10 Played Songs



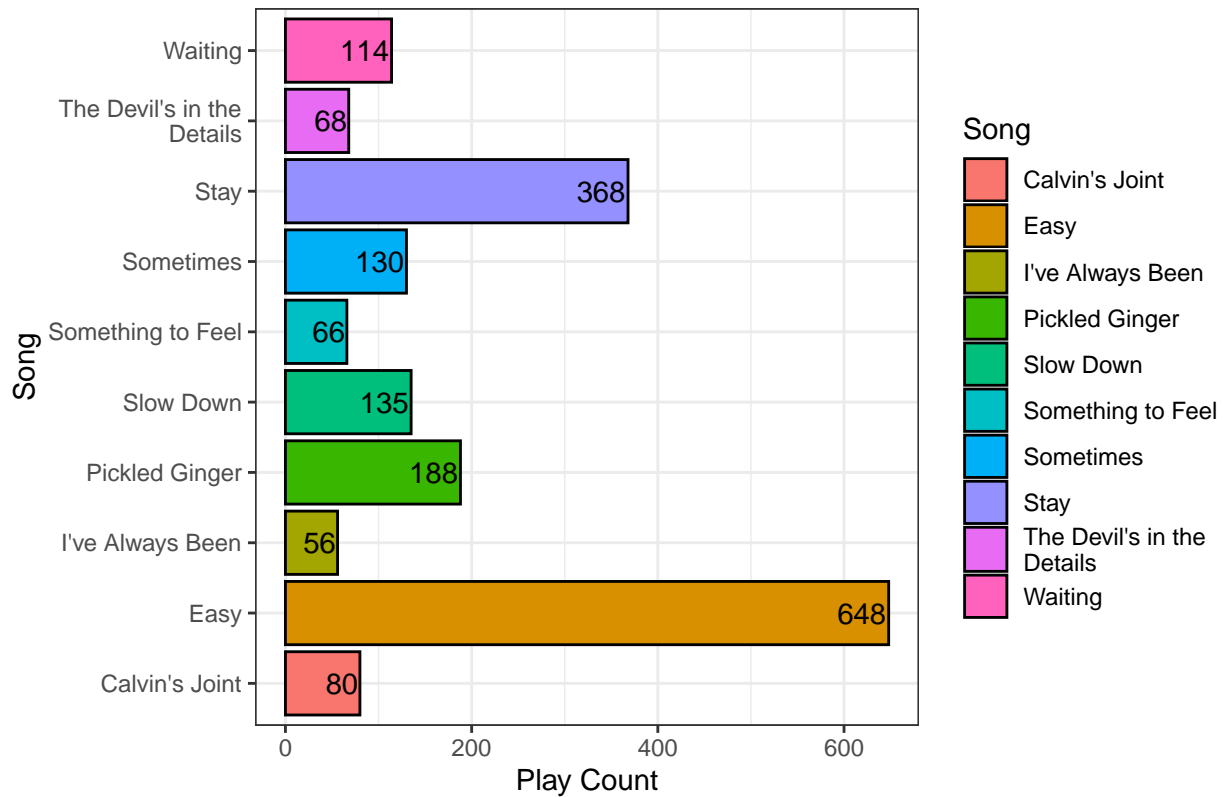
### UMI Top 10 Played Songs



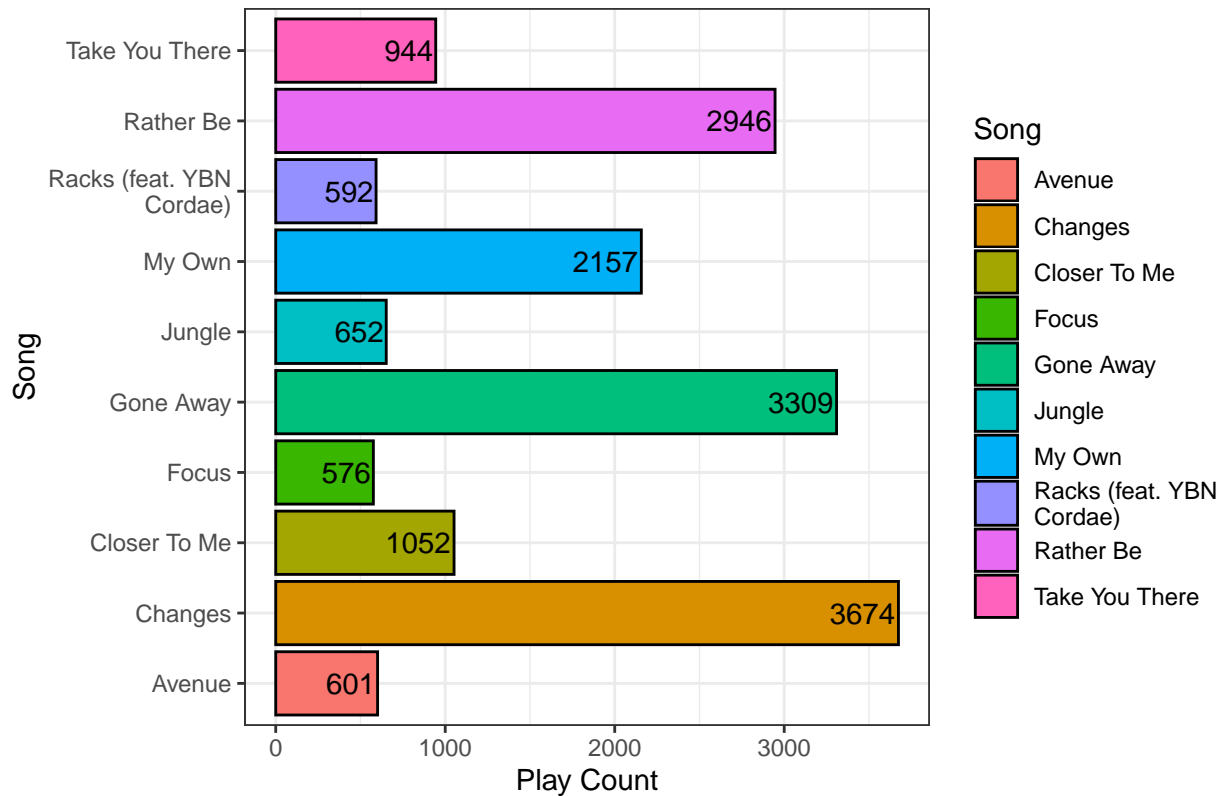
### Kehlani Top 10 Played Songs



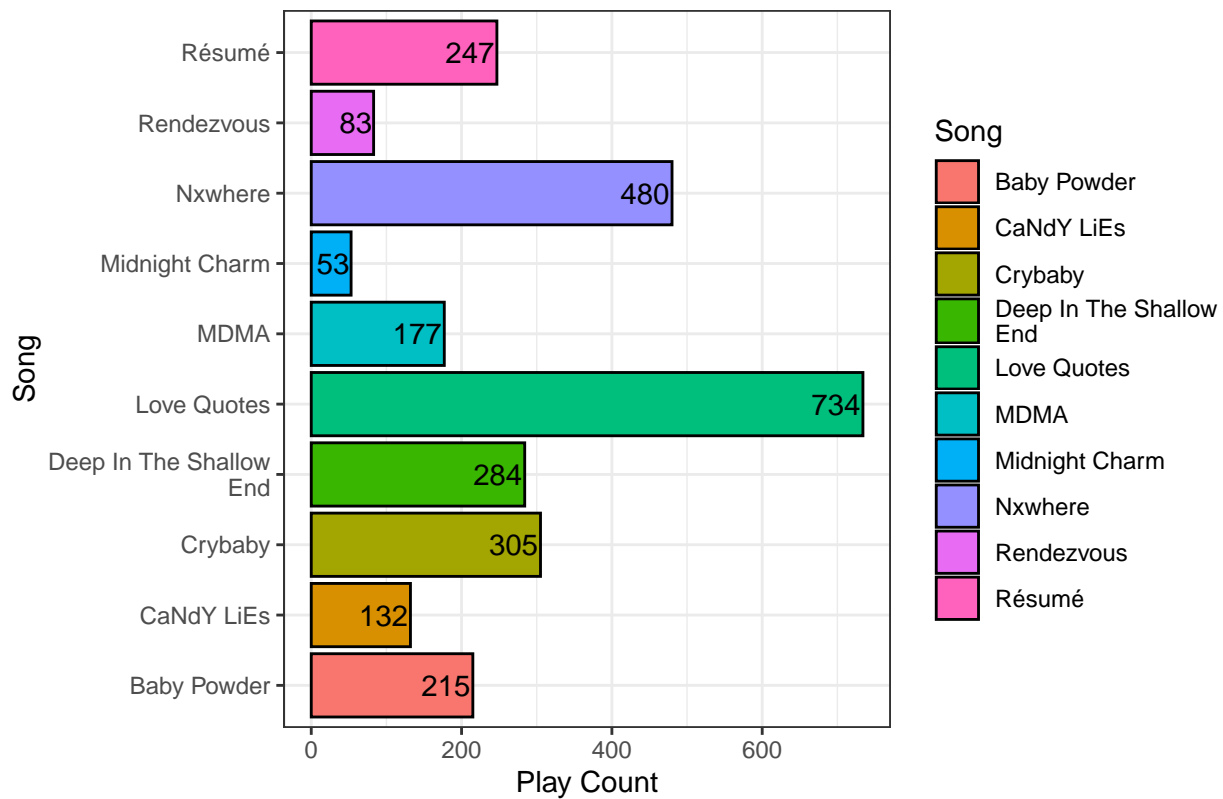
### Mac Ayres Top 10 Played Songs



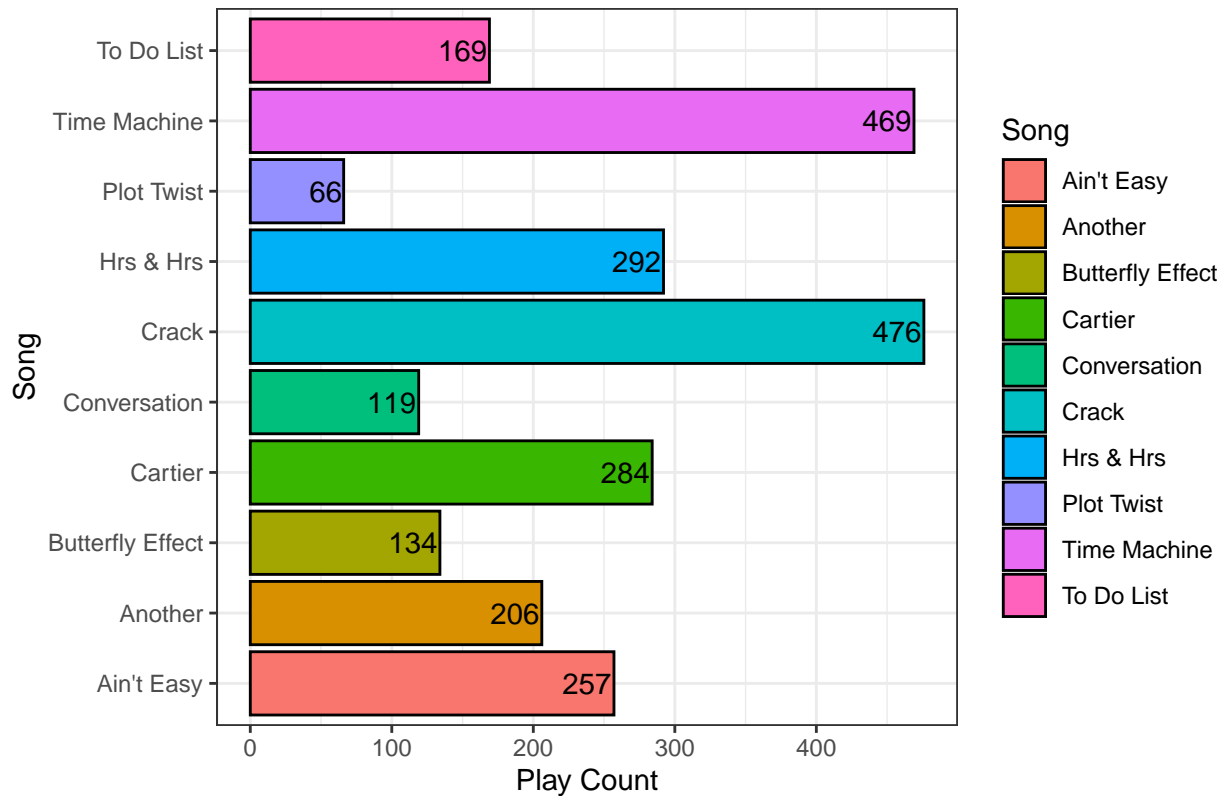
### H.E.R. Top 10 Played Songs



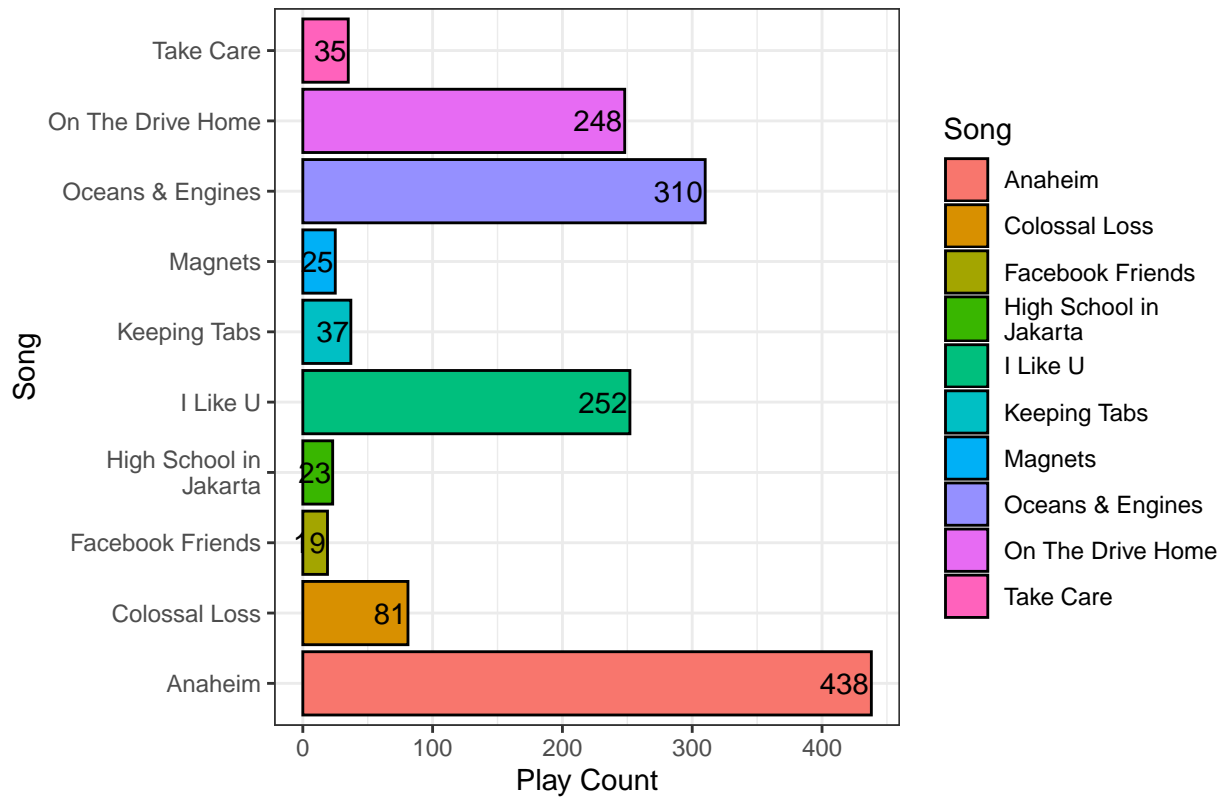
### Jenevieve Top 10 Played Songs



Muni Long Top 10 Played Songs



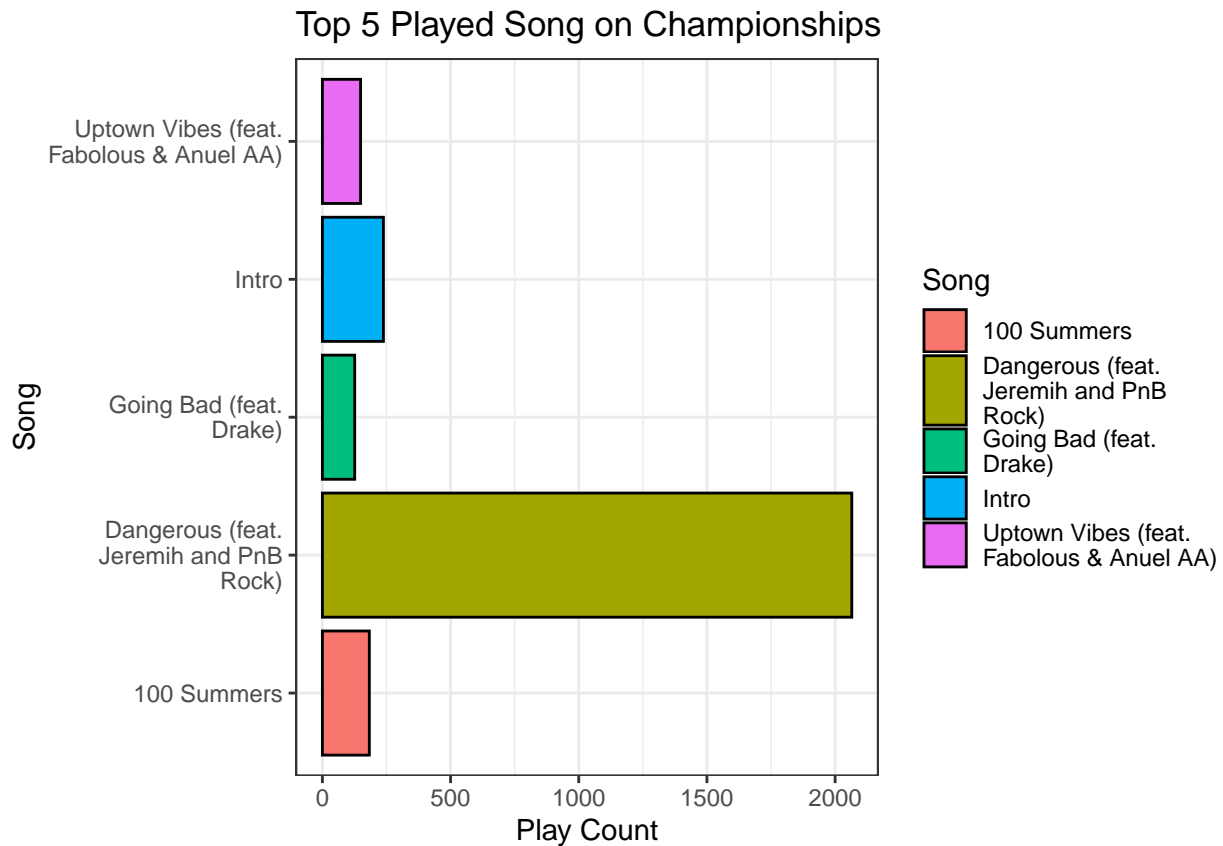
NIKI Top 10 Played Songs



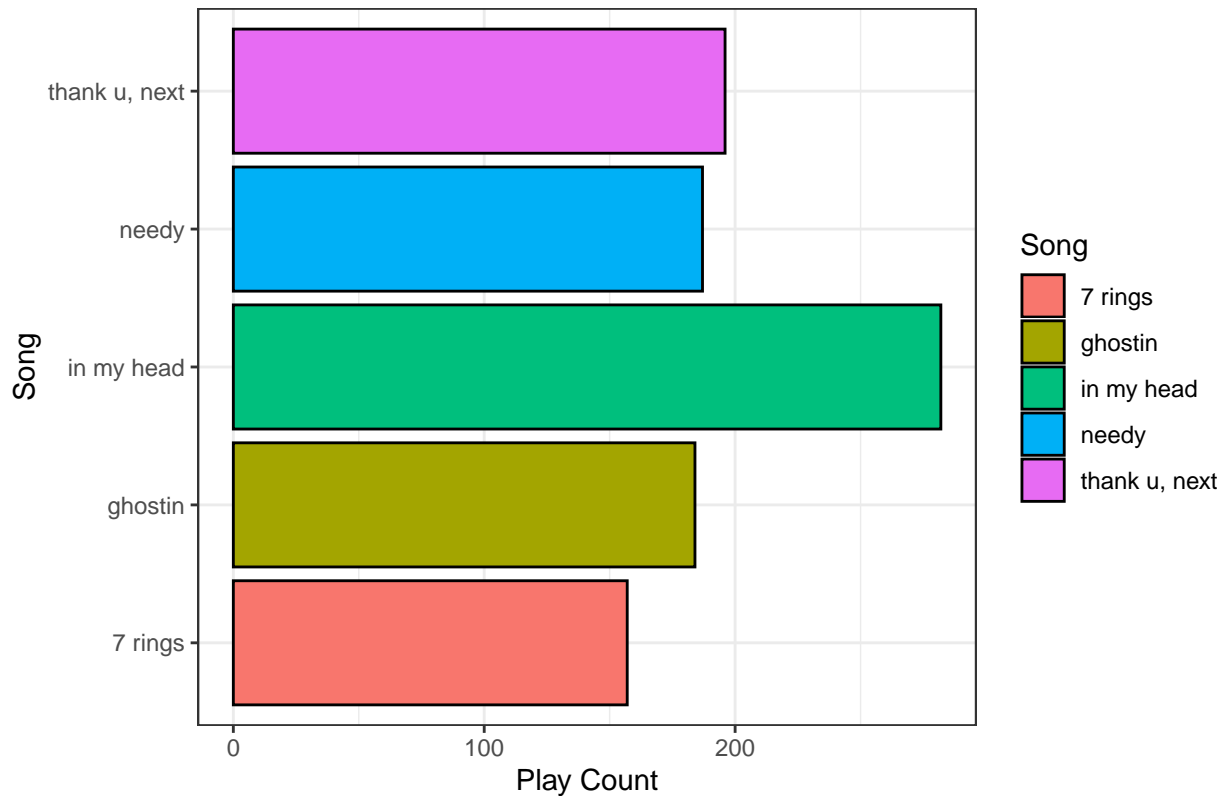
## Top 5 Songs from Favorite Albums

```
for (music_album in favorite_albums$album) {
  album_loop_data <- streaming_data_2 %>%
    filter(album == music_album) %>%
    group_by(song) %>%
    summarize(song_album_count = n(), .groups = "drop") %>%
    arrange(desc(song_album_count)) %>%
    slice_head(n = 5)

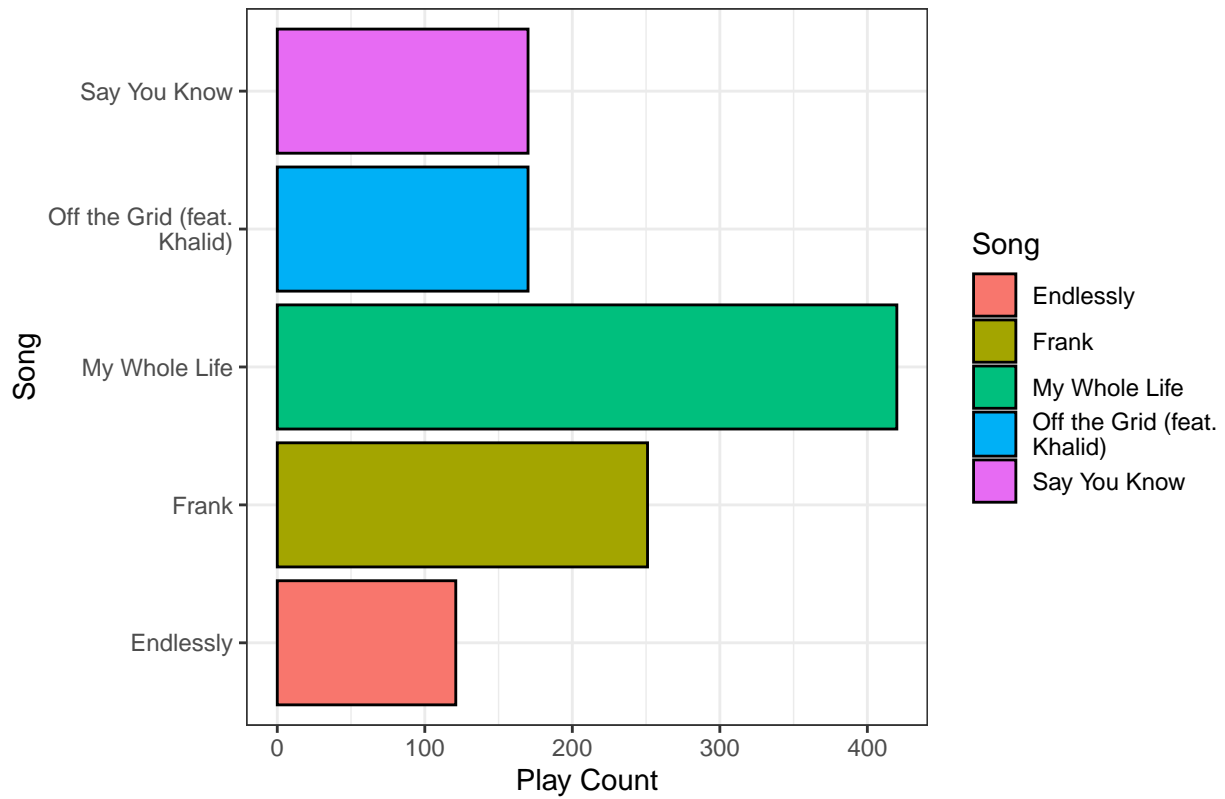
  top5_album_songs <- ggplot(album_loop_data %>%
    mutate(song = str_wrap(song, width = 20)),
    aes(x = song_album_count,
        y = song,
        fill = song)) +
    geom_bar(stat = "identity", position = "dodge", color = "black") +
    theme_bw() +
    labs(
      title = paste("Top 5 Played Song on", music_album),
      y = "Song",
      x = "Play Count",
      fill = "Song"
    )
  print(top5_album_songs)
}
```



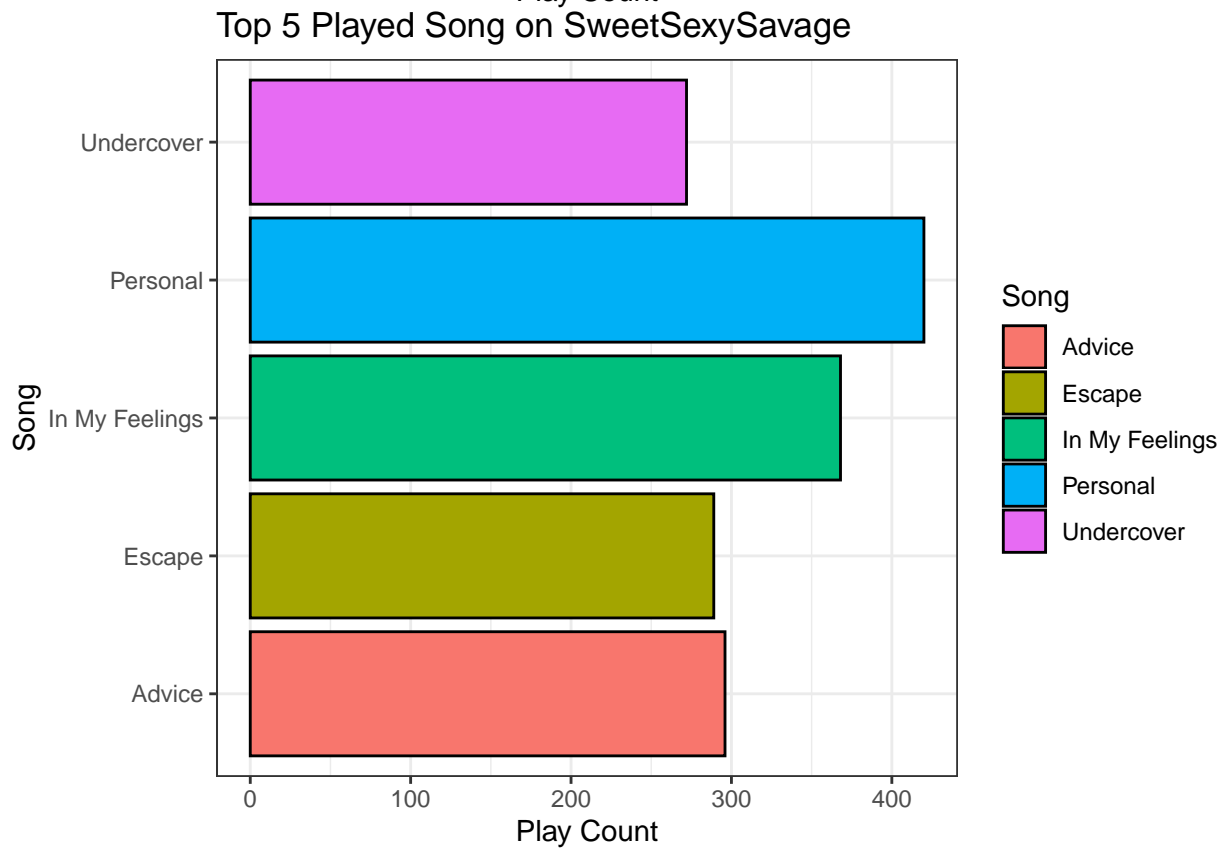
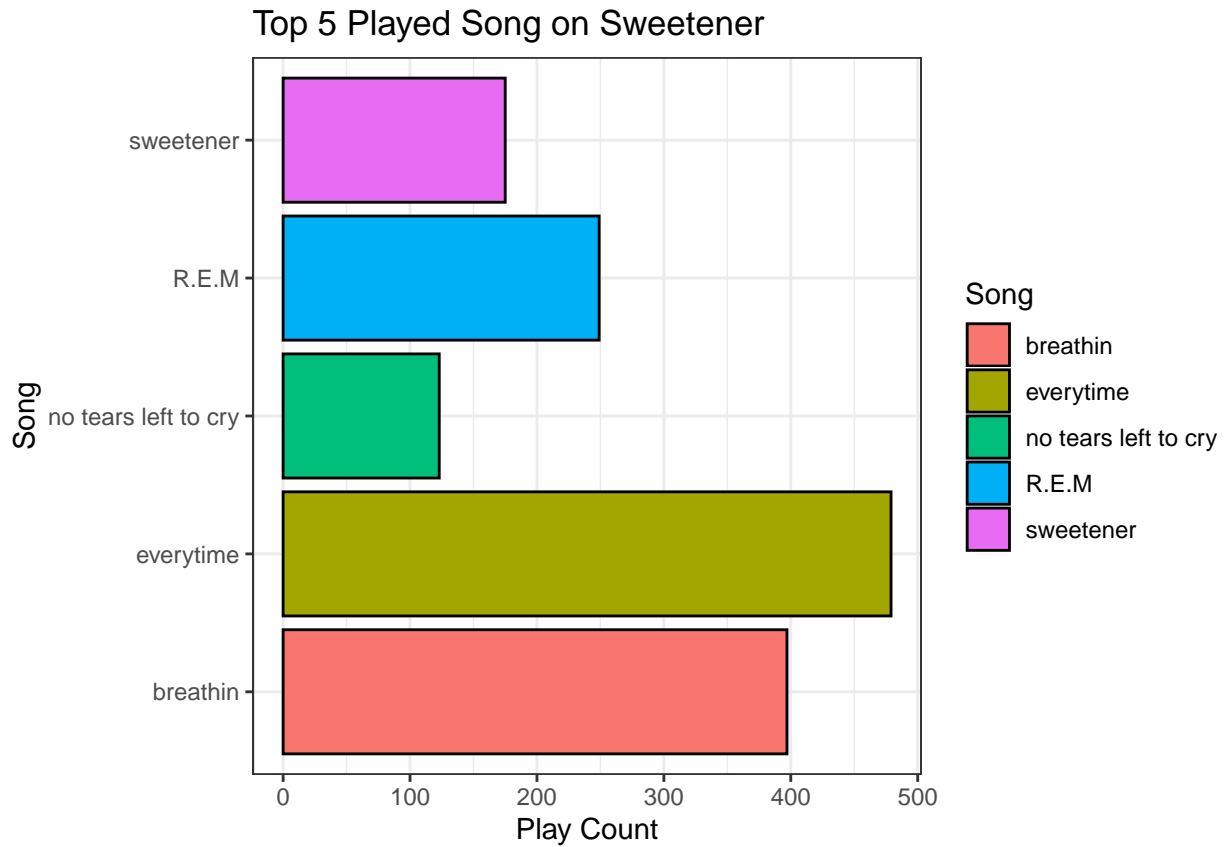
Top 5 Played Song on thank u, next



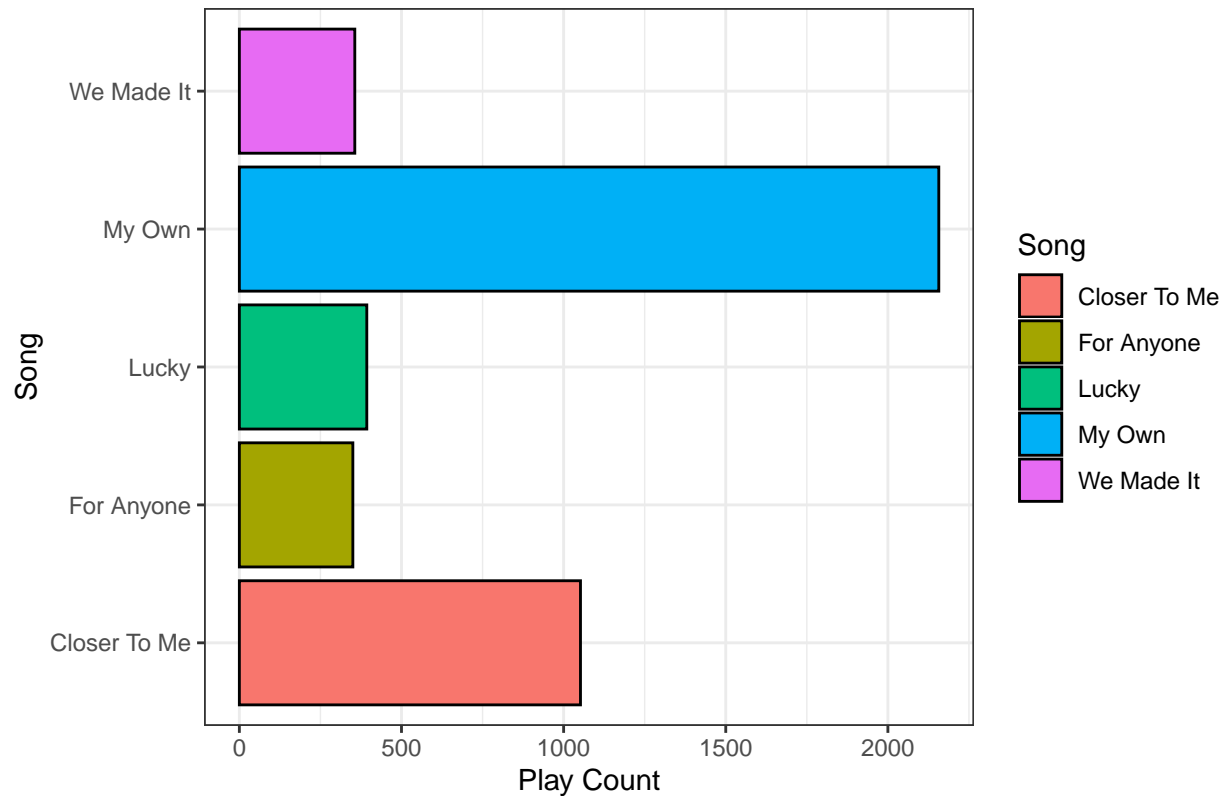
Top 5 Played Song on It Was Divine



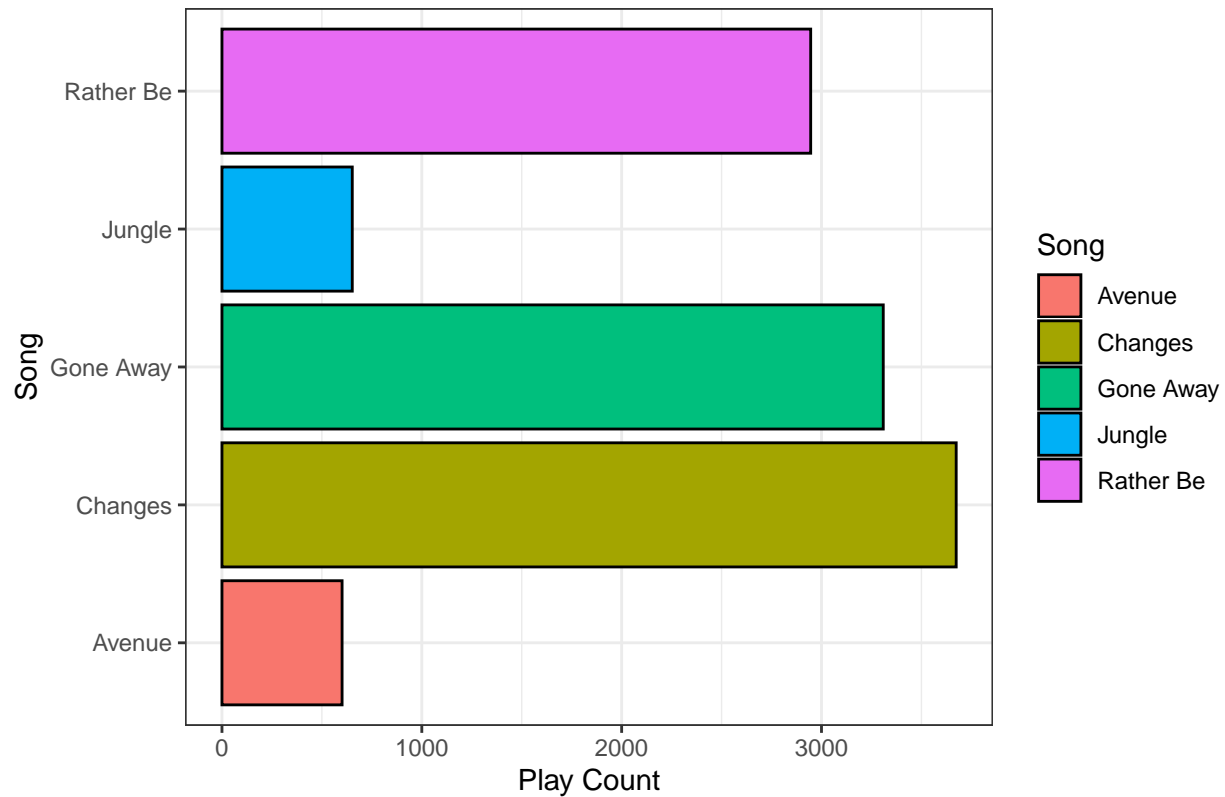




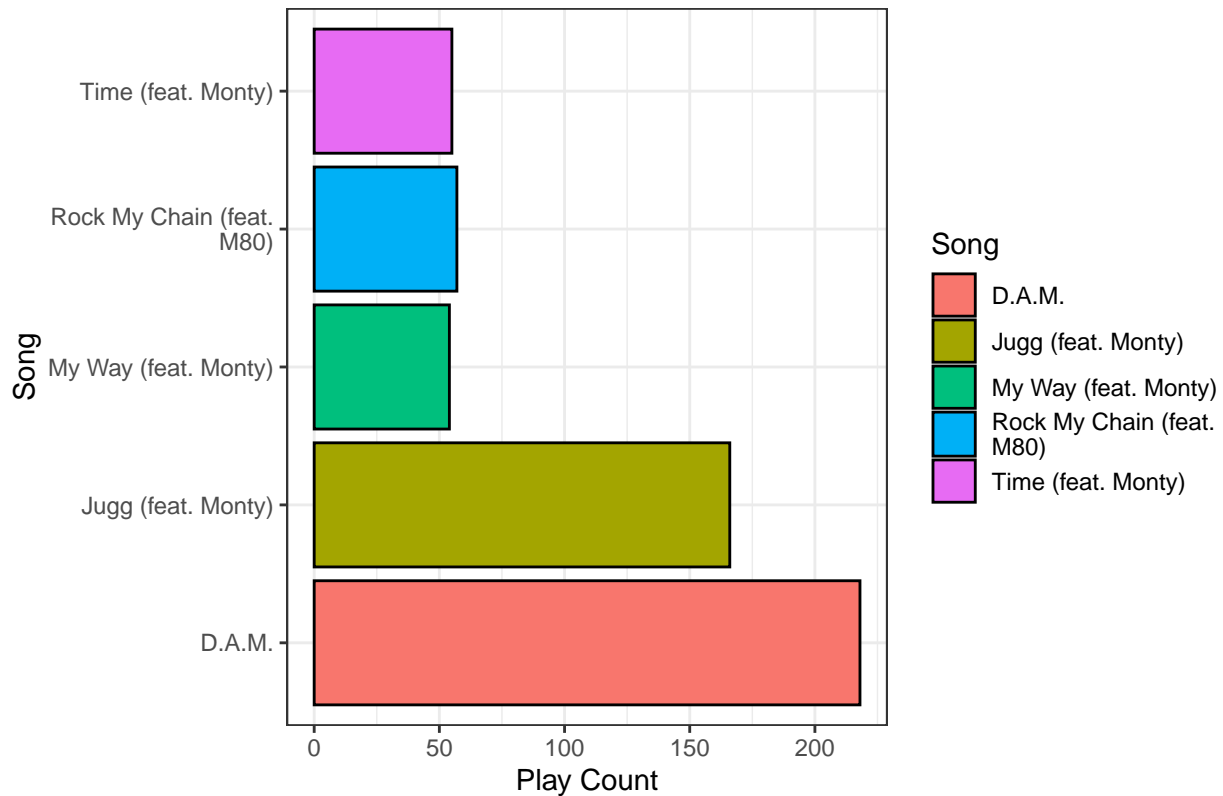
Top 5 Played Song on Back of My Mind



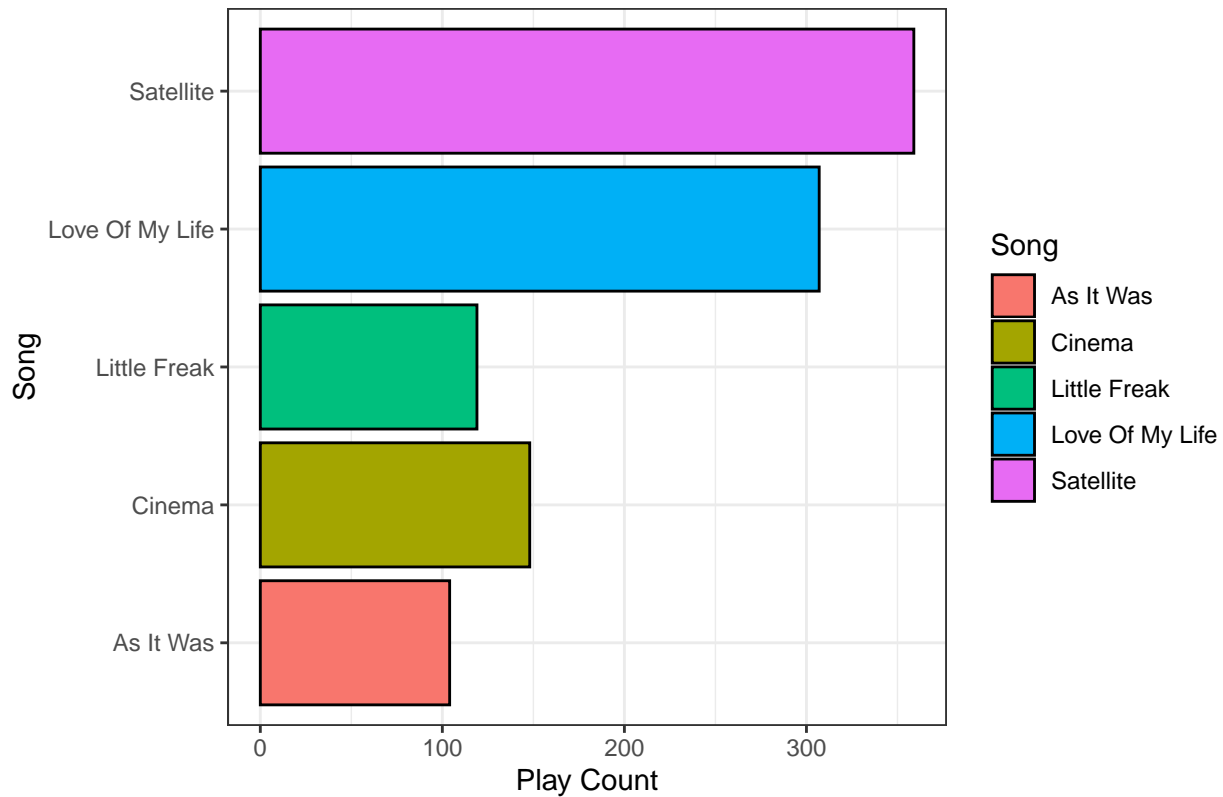
Top 5 Played Song on H.E.R.



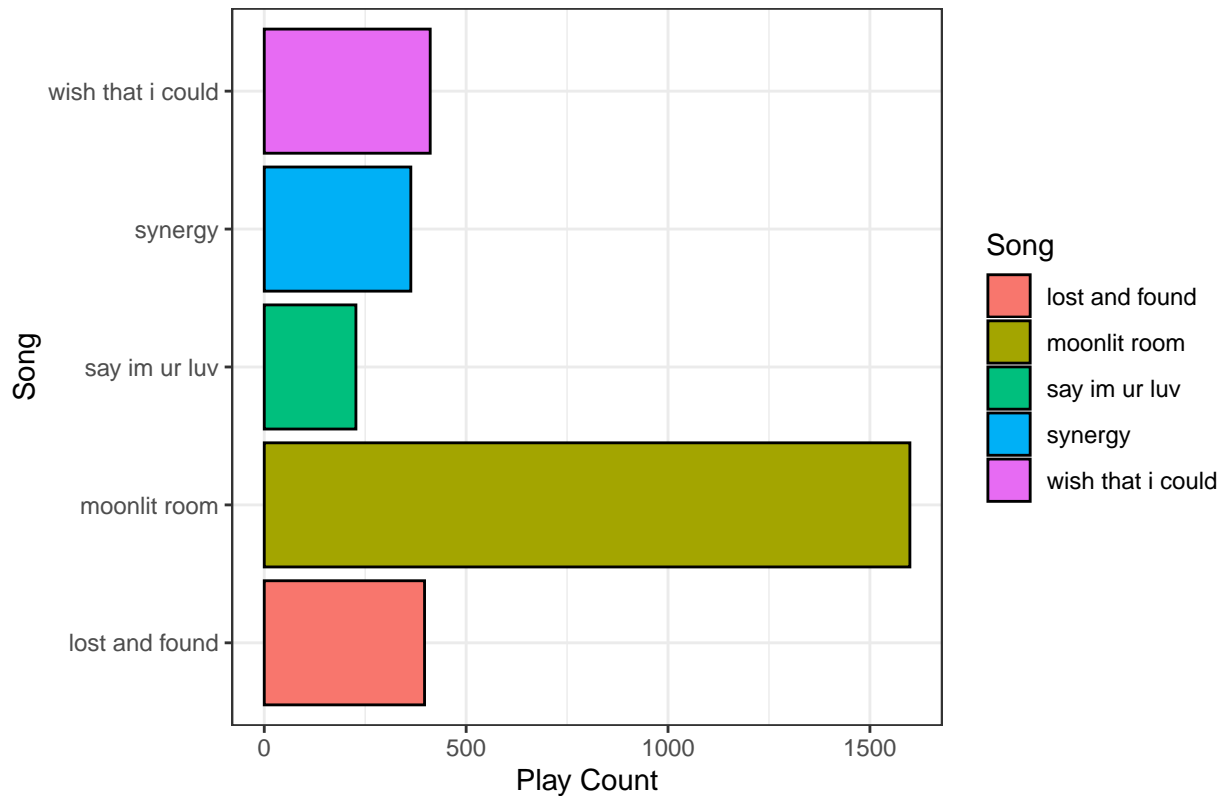
Top 5 Played Song on Fetty Wap



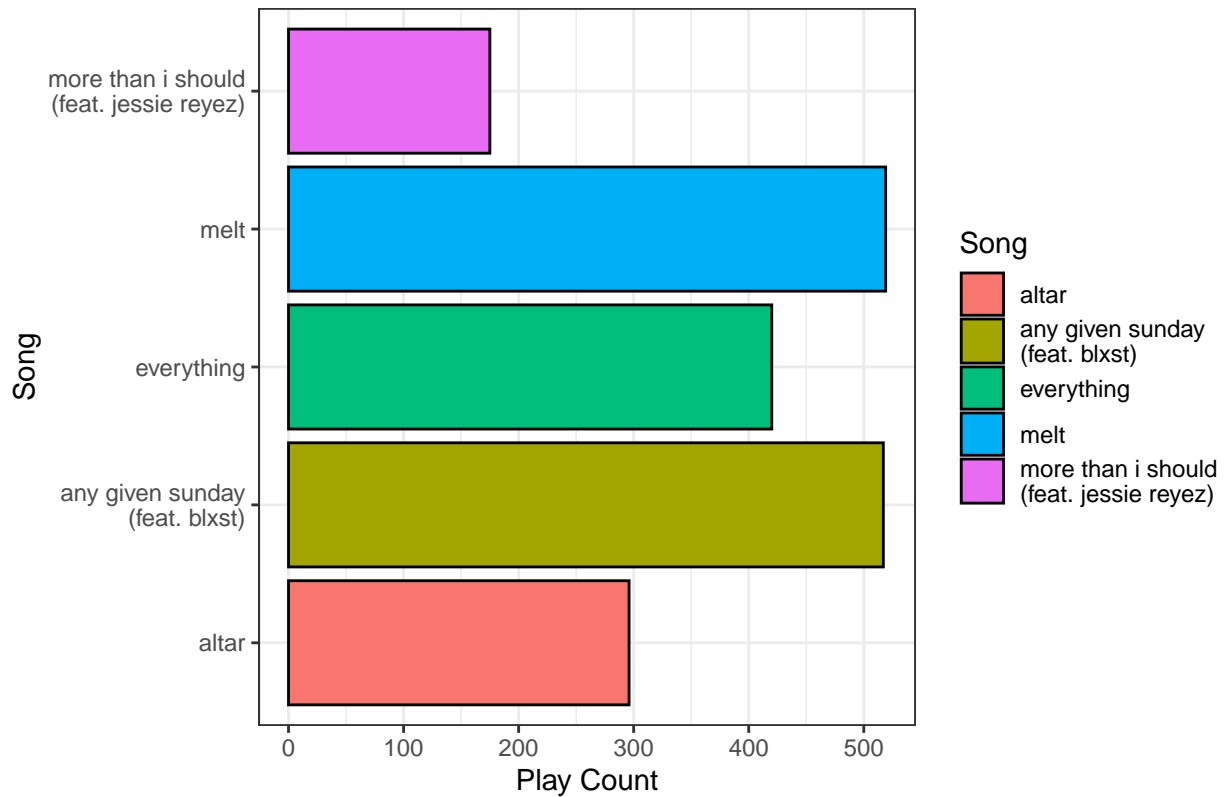
Top 5 Played Song on Harry's House



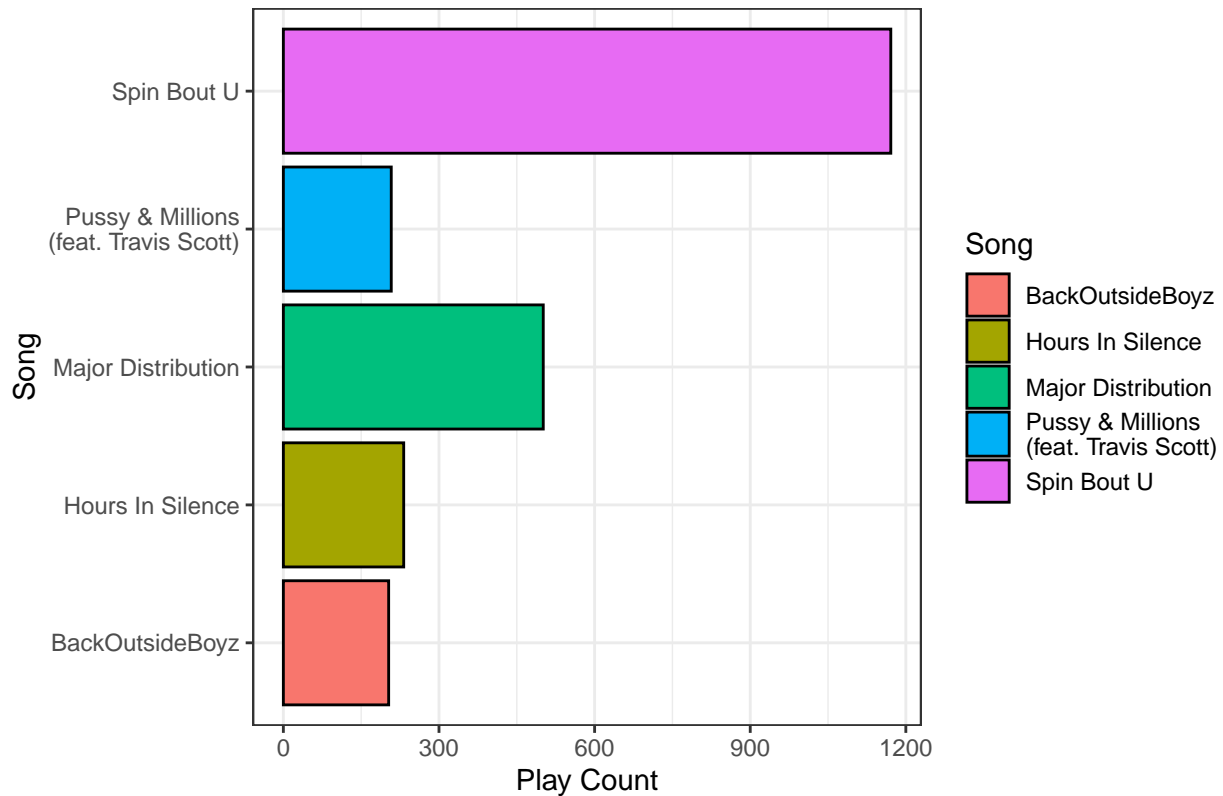
Top 5 Played Song on Forest in the City



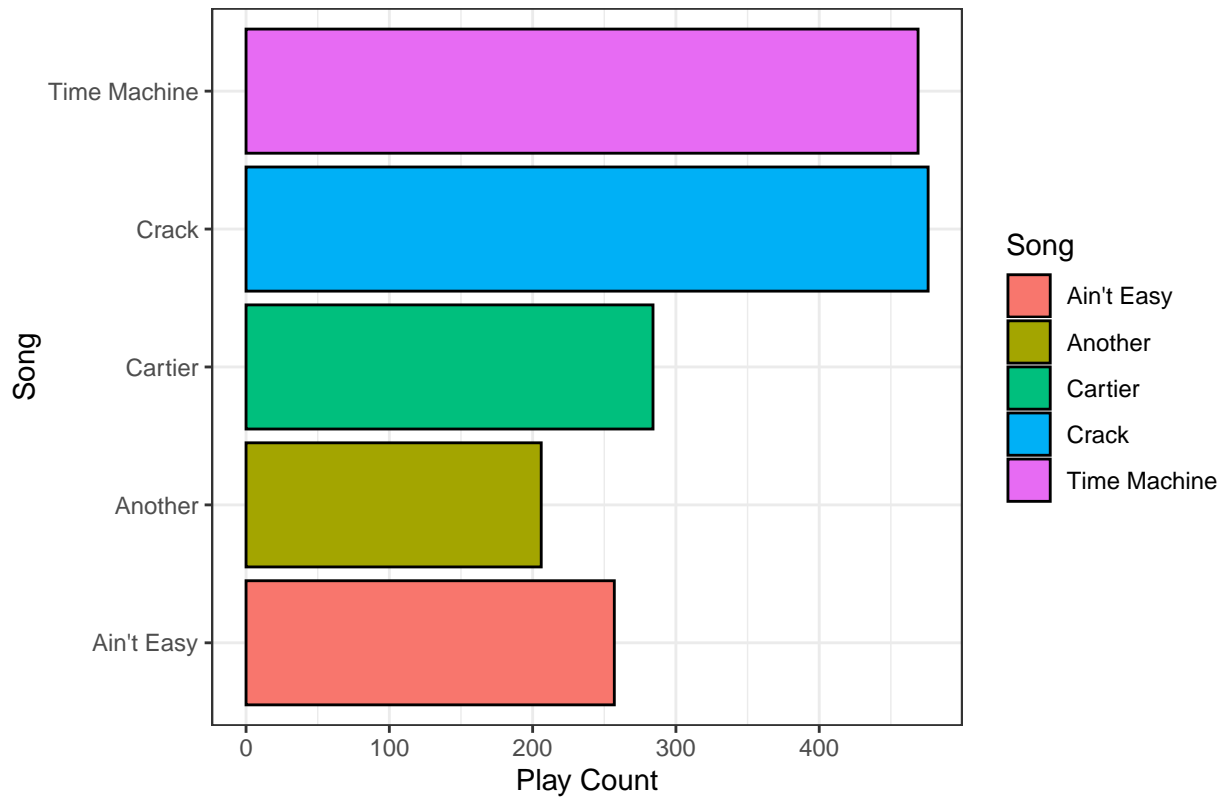
Top 5 Played Song on blue water road



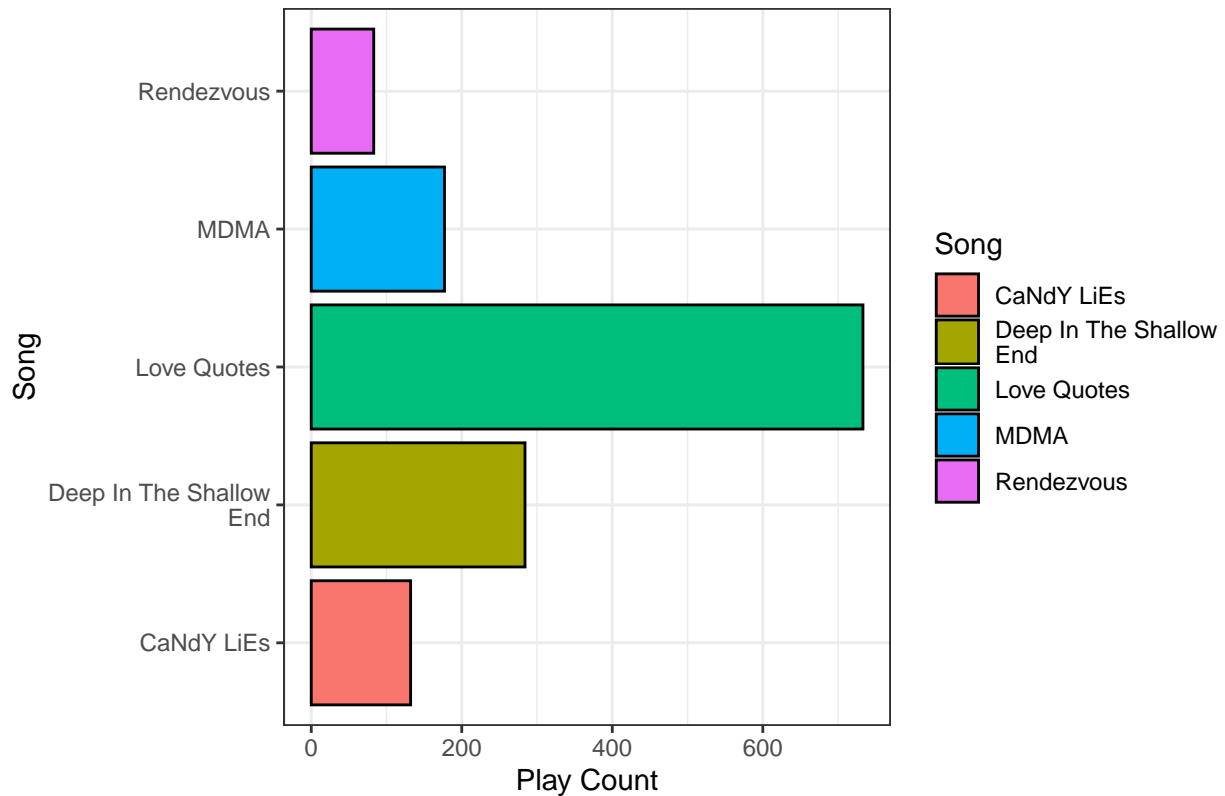
Top 5 Played Song on Her Loss



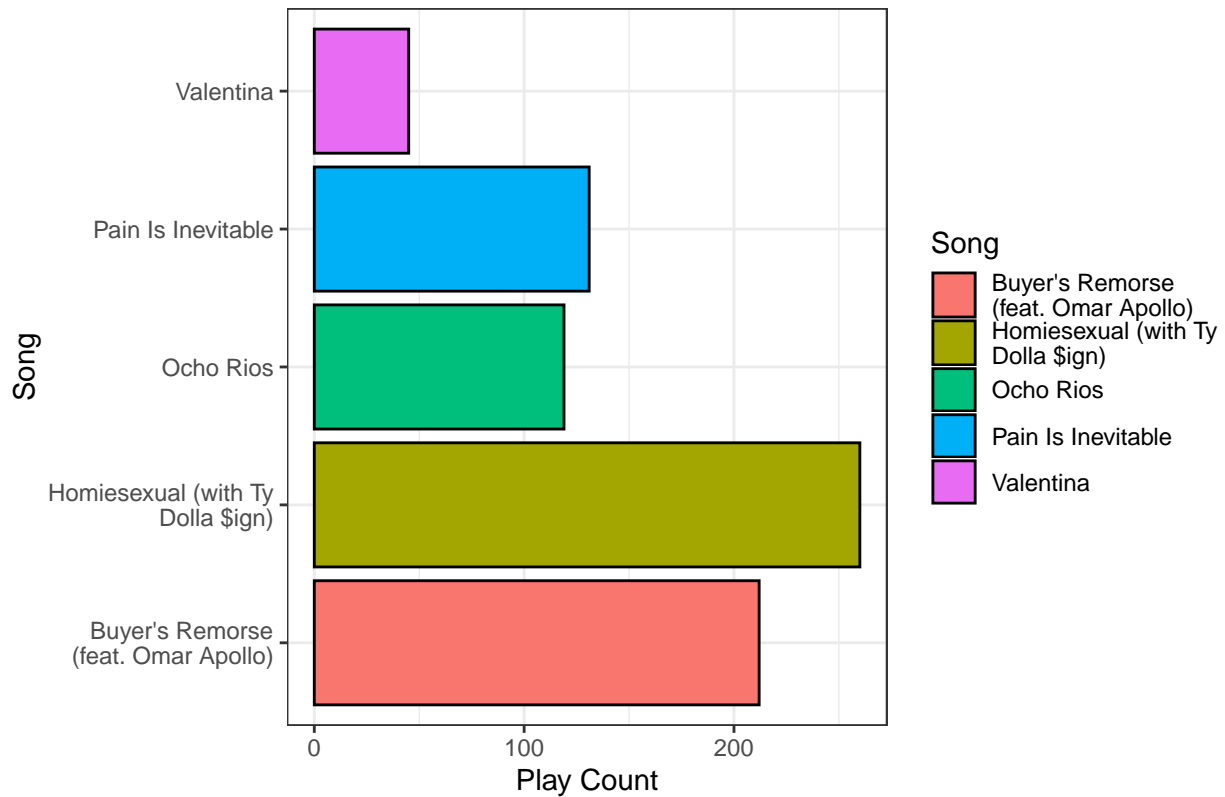
Top 5 Played Song on Public Displays Of Affection: The Album



Top 5 Played Song on Rendezvous – EP



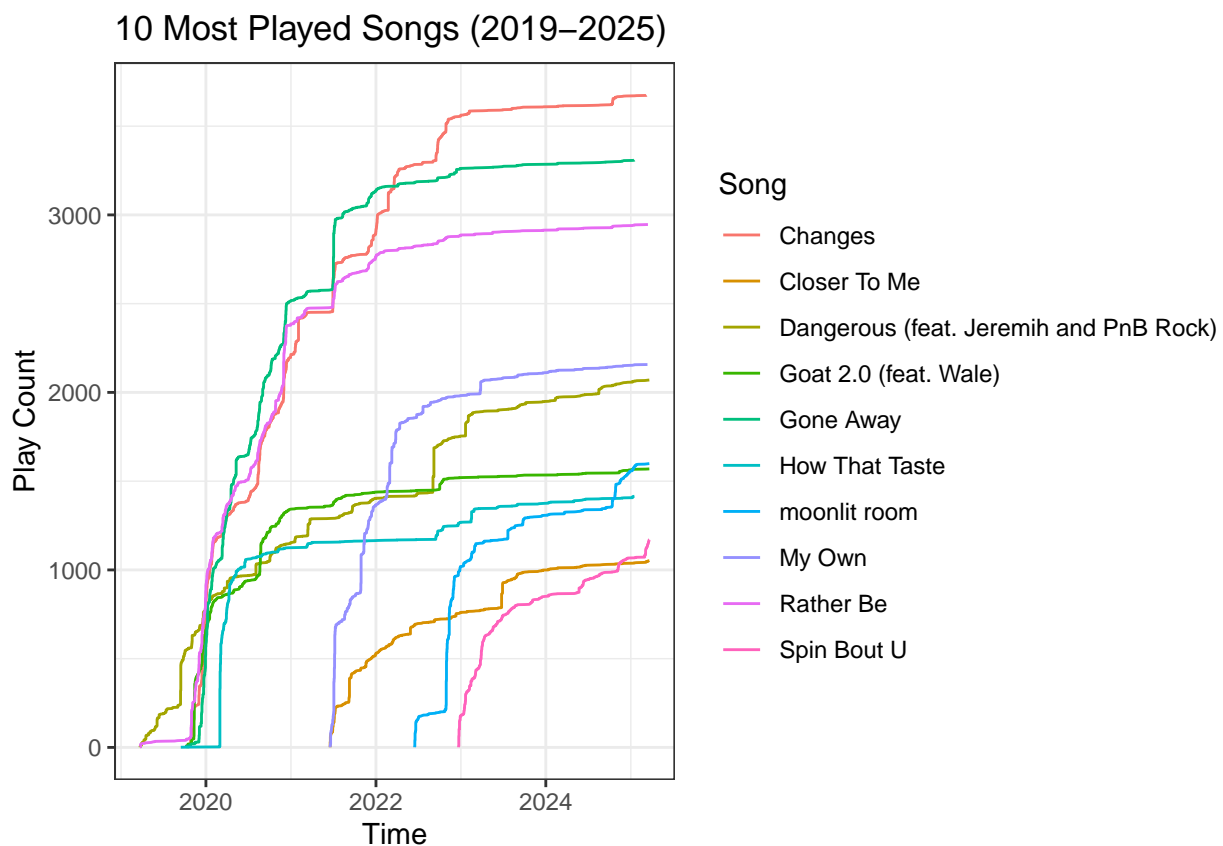
Top 5 Played Song on NEVER ENOUGH



## 10 Most Played Songs/Artists/Albums (of All Time) Over Time

```
#Most Listened to Songs of All Time
alltime_songs_2 <- alltime_songs %>%
  arrange(desc(alltime_song_count)) %>%
  slice_head(n = 10)

top10_songs_oat_bytime <- ggplot(alltime_songs_plot_data %>%
  filter(
    song %in% alltime_songs_2$song &
    artist %in% alltime_songs_2$artist &
    alltime_song_count %in%
    alltime_songs_2$alltime_song_count),
  aes(x = play_timestamp,
    y = alltime_song_cumsum,
    color = song)) +
  geom_path() +
  theme_bw() +
  labs(
    title = "10 Most Played Songs (2019–2025)",
    color = "Song",
    x = "Time",
    y = "Play Count"
  )
top10_songs_oat_bytime
```



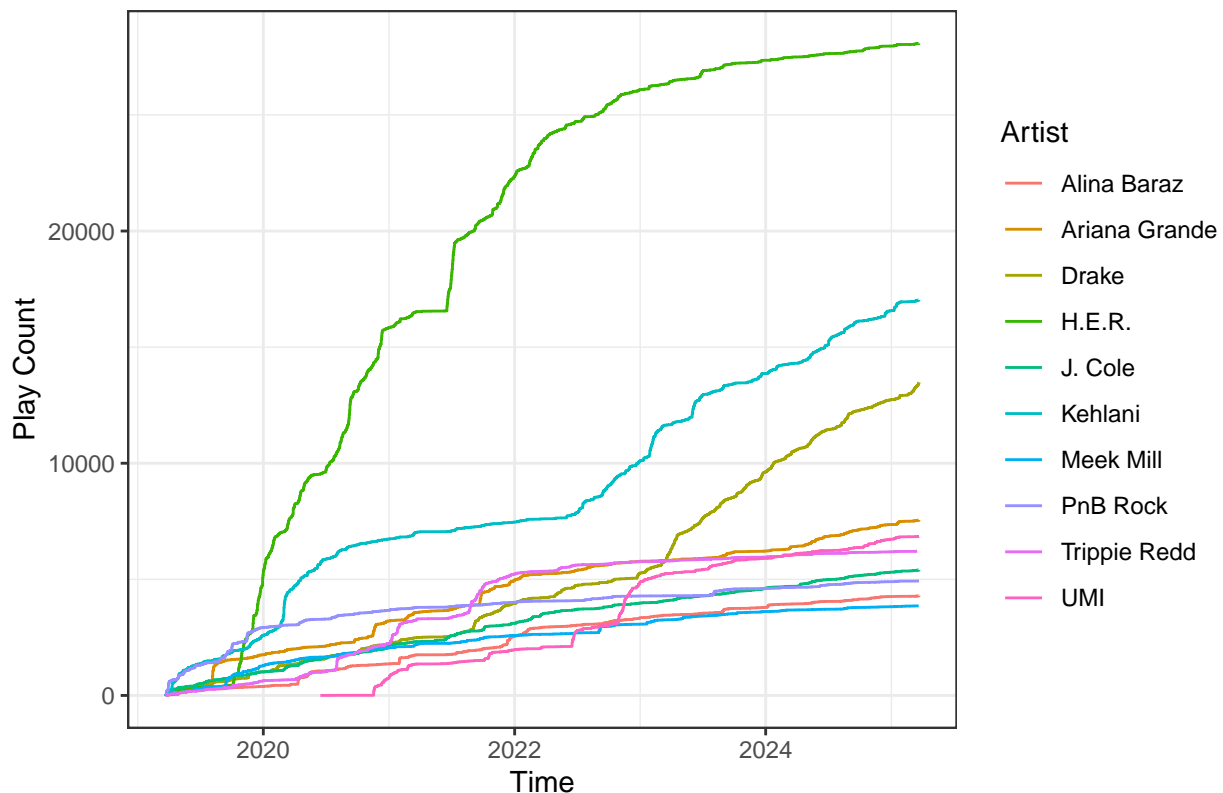
```

#Most Listened to Artists of All Time
alltime_artists_2 <- alltime_artists %>%
  arrange(desc(alltime_artist_count)) %>%
  slice_head(n = 10)

top10_artists_oat_bytime <- ggplot(alltime_artists_plot_data %>%
  filter(
    artist %in% alltime_artists_2$artist &
    alltime_artist_count %in%
    alltime_artists_2$alltime_artist_count),
  aes(x = play_timestamp,
    y = alltime_artist_cumsum,
    color = artist)) +
  geom_path() +
  theme_bw() +
  labs(
    title = "10 Most Played Artists (2019-2025)",
    color = "Artist",
    x = "Time",
    y = "Play Count"
  )
top10_artists_oat_bytime

```

10 Most Played Artists (2019–2025)



```

#Most Listened to Albums of All Time
alltime_albums_2 <- alltime_albums %>%
  arrange(desc(alltime_album_count)) %>%
  slice_head(n = 10)

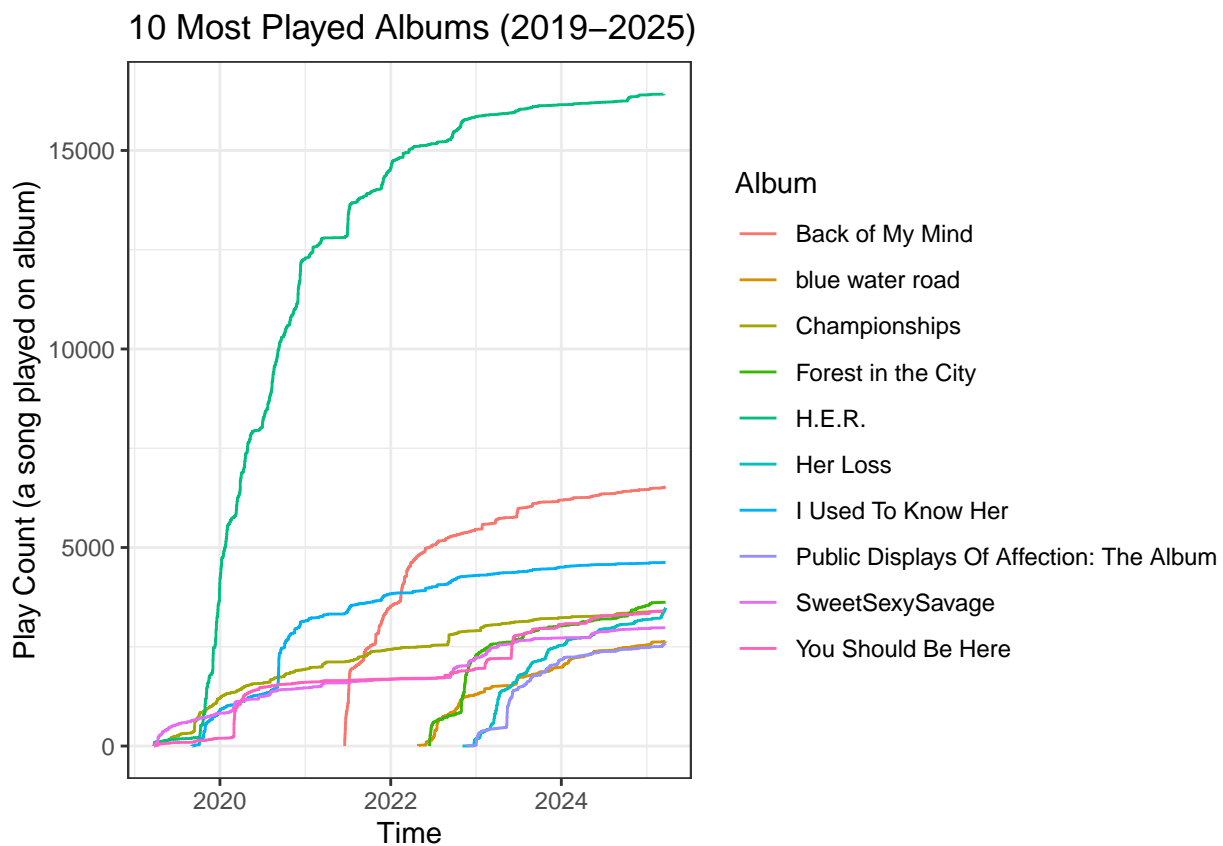
```



```

top10_albums_oat_bytime <- ggplot(alltime_albums_plot_data %>%
                                filter(
                                  album %in% alltime_albums_2$album &
                                  artist %in% alltime_albums_2$artist &
                                  alltime_album_count %in%
                                  alltime_albums_2$alltime_album_count),
                                aes(x = play_timestamp,
                                    y = alltime_album_cumsum, color = album)) +
  geom_path() +
  theme_bw() +
  labs(
    title = "10 Most Played Albums (2019-2025)",
    color = "Album",
    x = "Time",
    y = "Play Count (a song played on album)"
  )
top10_albums_oat_bytime

```



#### 10 Most Played Songs (in Each Year) Over the Year

```

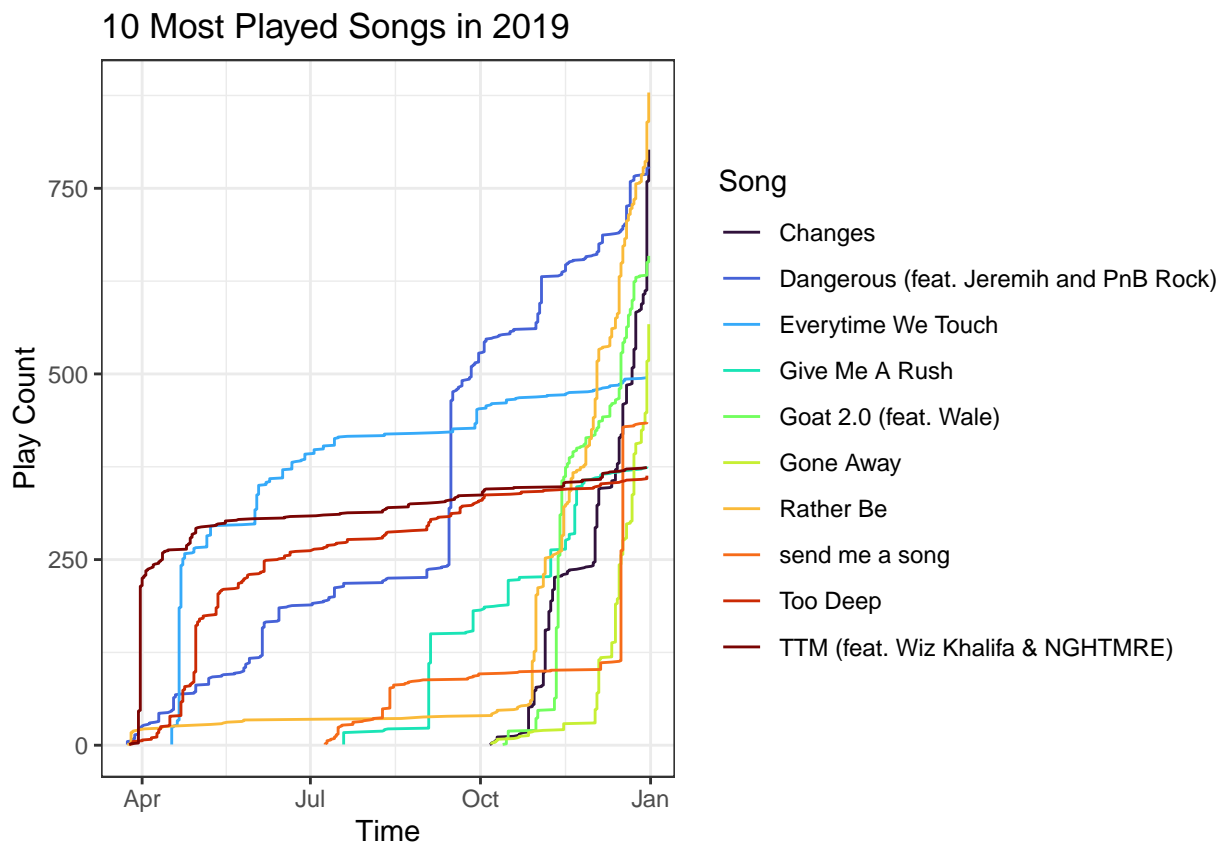
for (music_year in music_years) {
  yearly_top10_songs_bytime <- ggplot(alltime_songs_plot_data %>%
                                      filter(play_year == music_year &
                                             year_x_top10_song == "Top 10 Song"),
                                      aes(x = play_timestamp,
                                          y = yearly_song_cumsum,

```

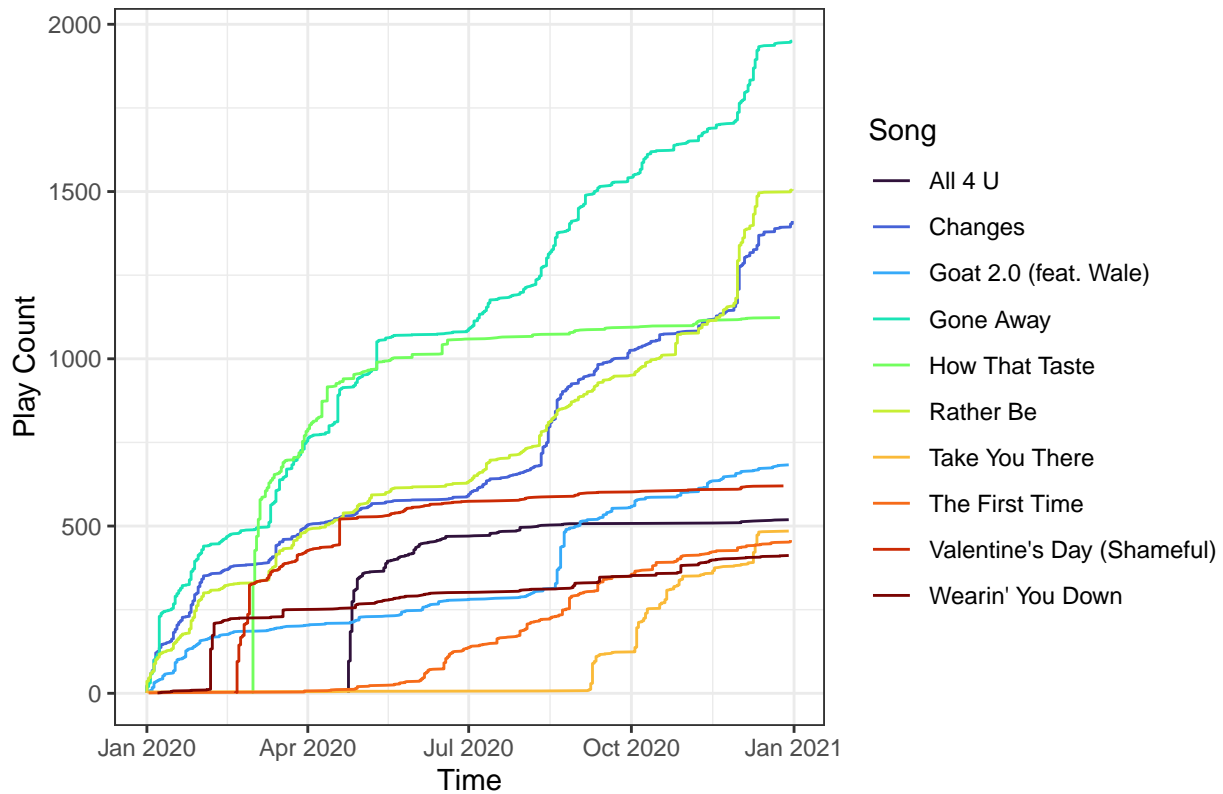
```

                                color = song)) +
  geom_path() +
  scale_color_viridis_d(option = "turbo") +
  theme_bw() +
  labs(
    title = paste("10 Most Played Songs in", music_year),
    color = "Song",
    x = "Time",
    y = "Play Count"
  )
print(yearly_top10_songs_bytime)
}

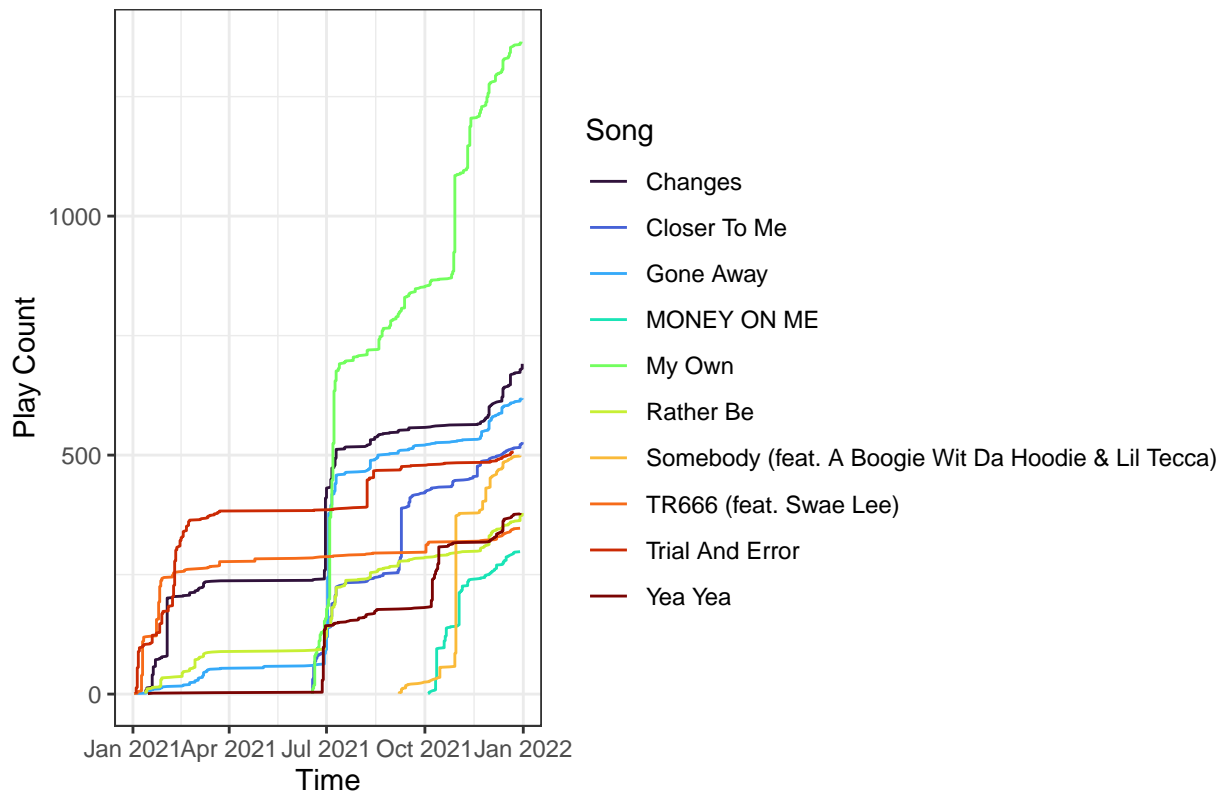
```



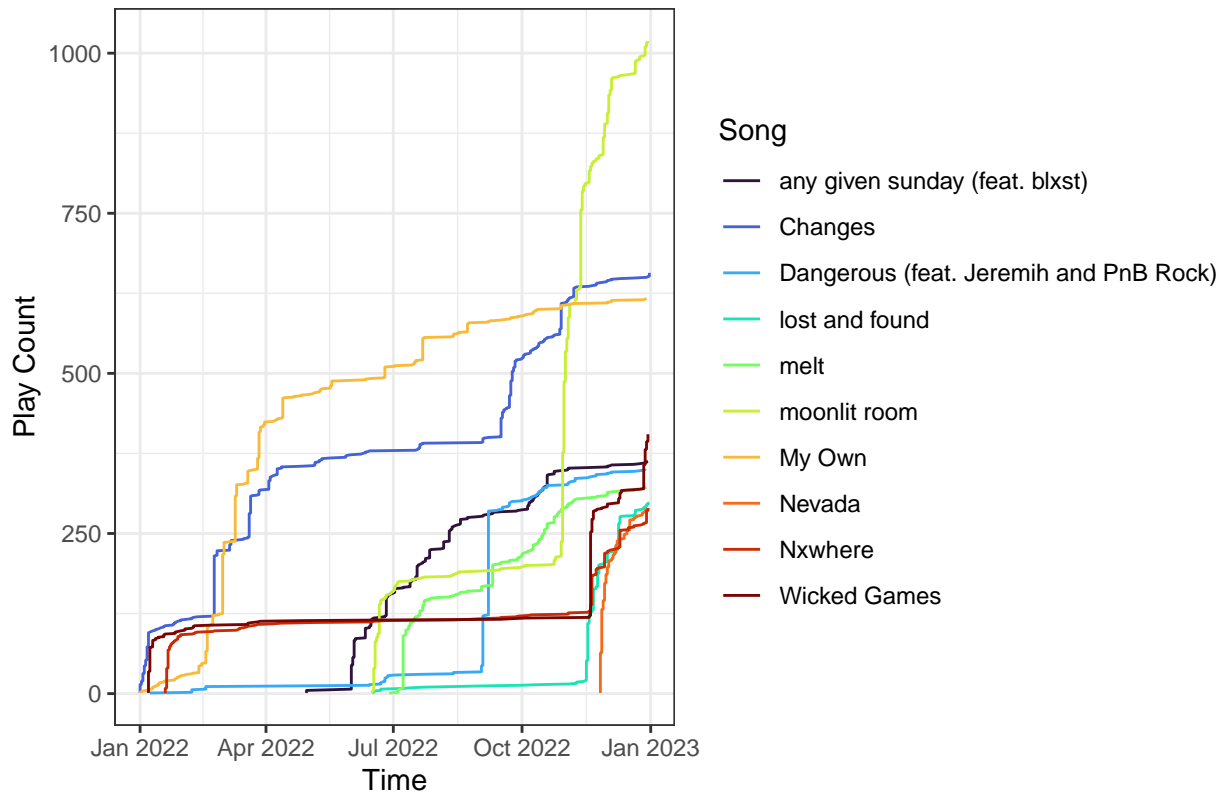
### 10 Most Played Songs in 2020



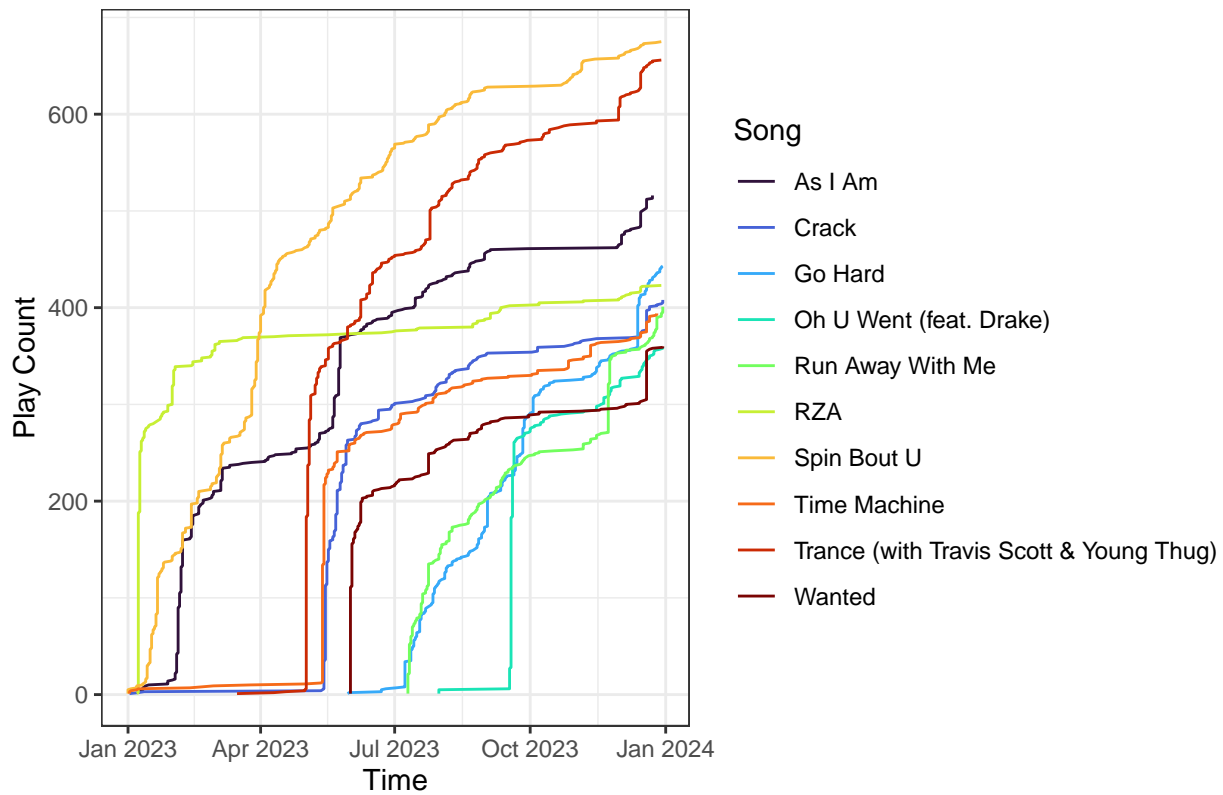
### 10 Most Played Songs in 2021



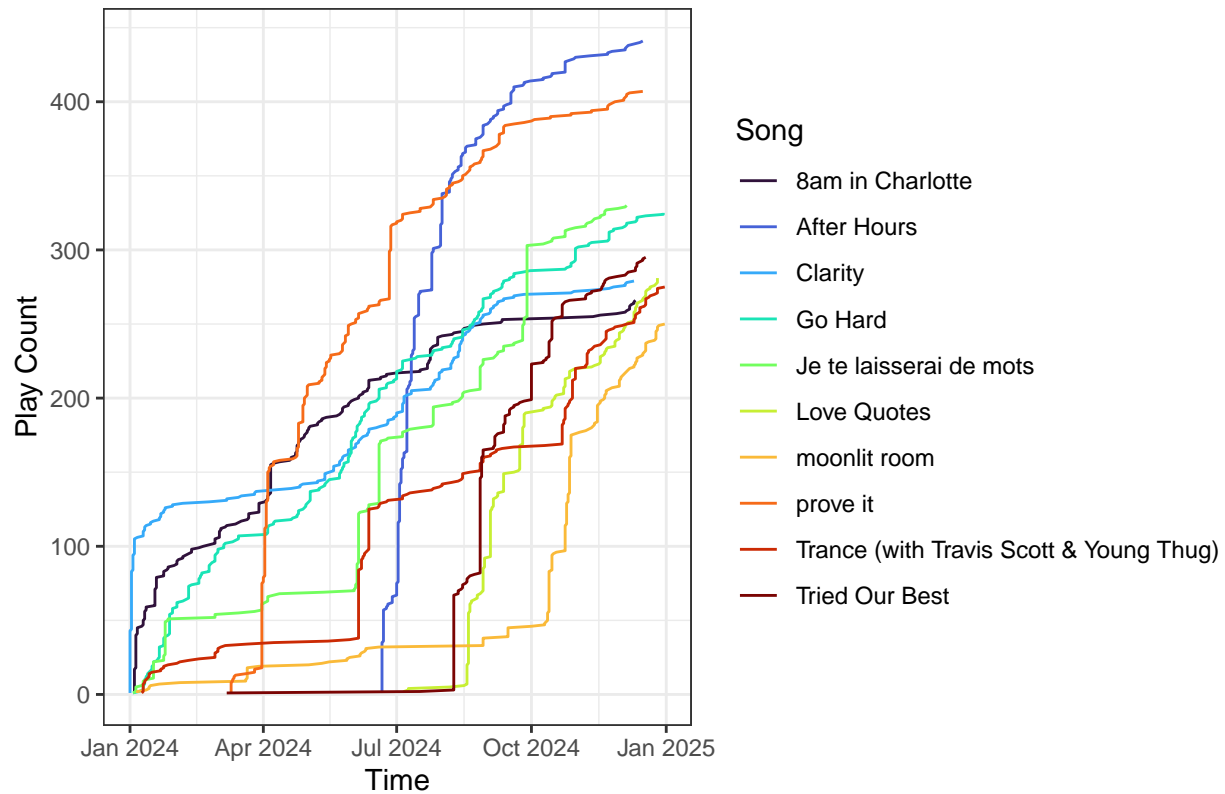
### 10 Most Played Songs in 2022



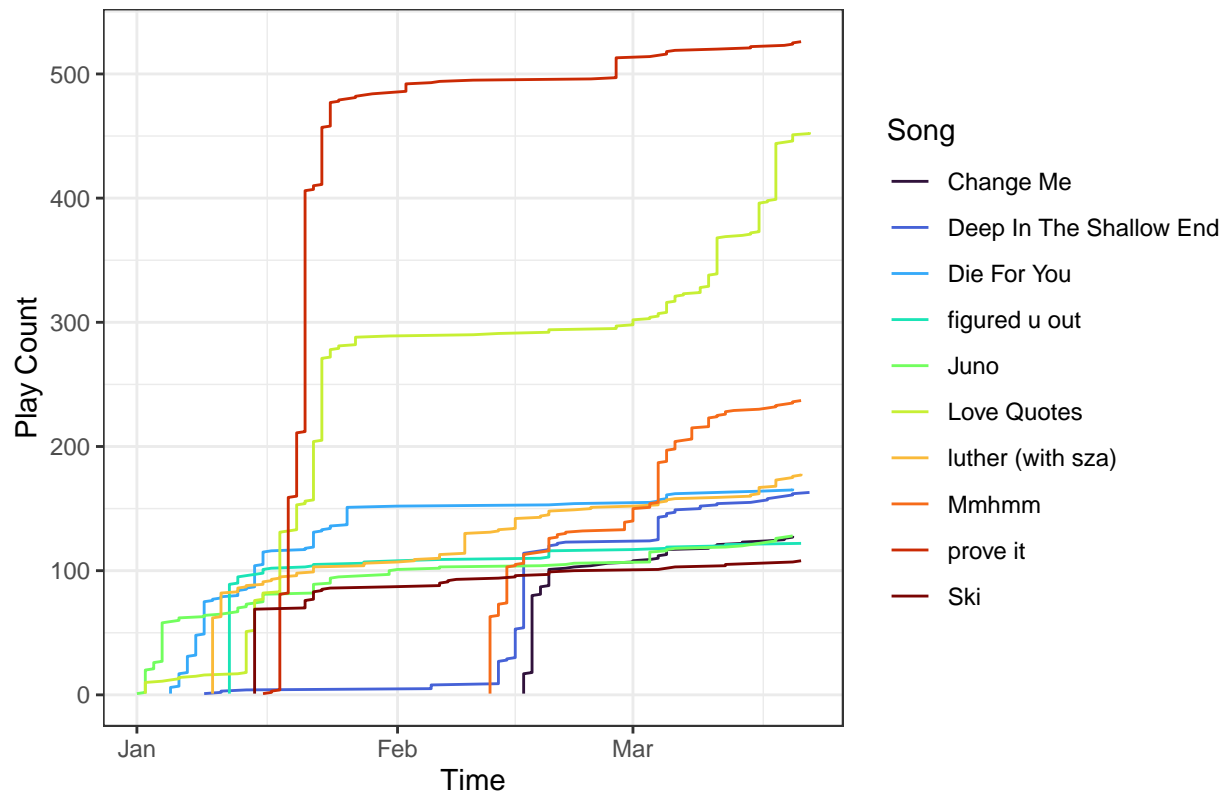
### 10 Most Played Songs in 2023



### 10 Most Played Songs in 2024



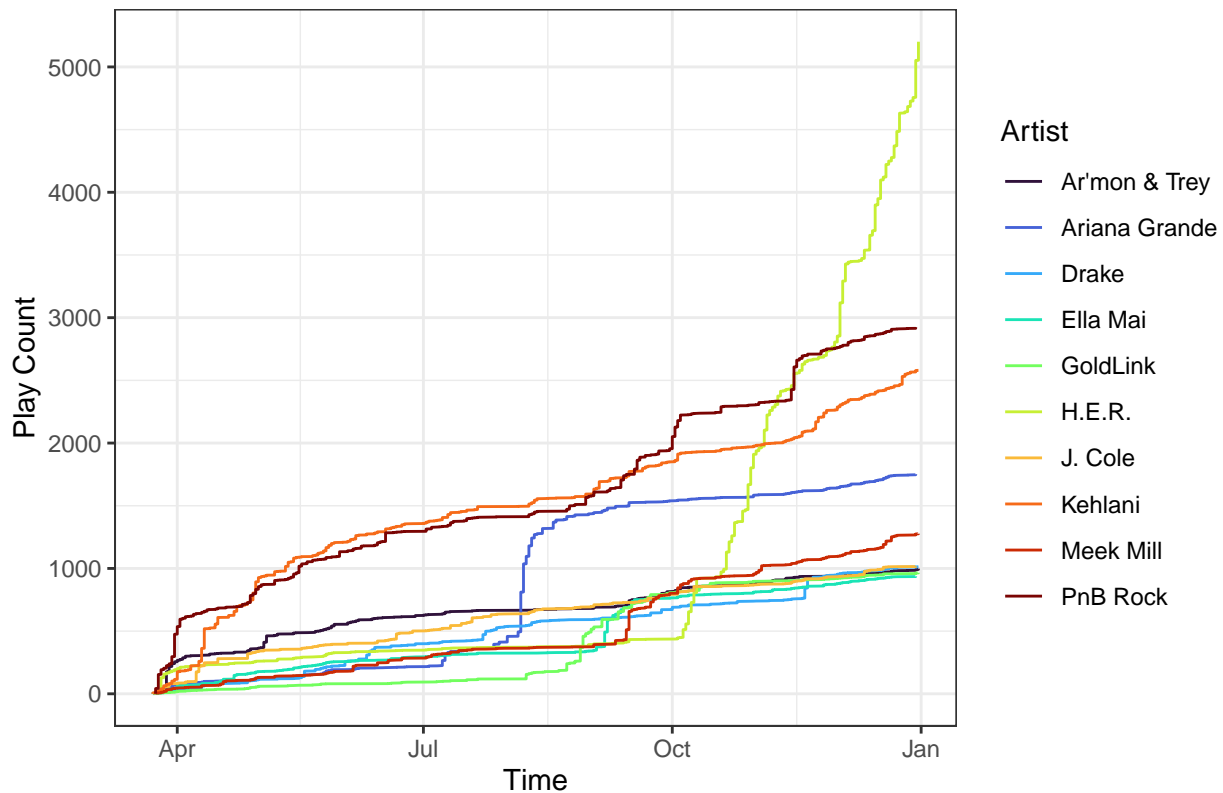
### 10 Most Played Songs in 2025



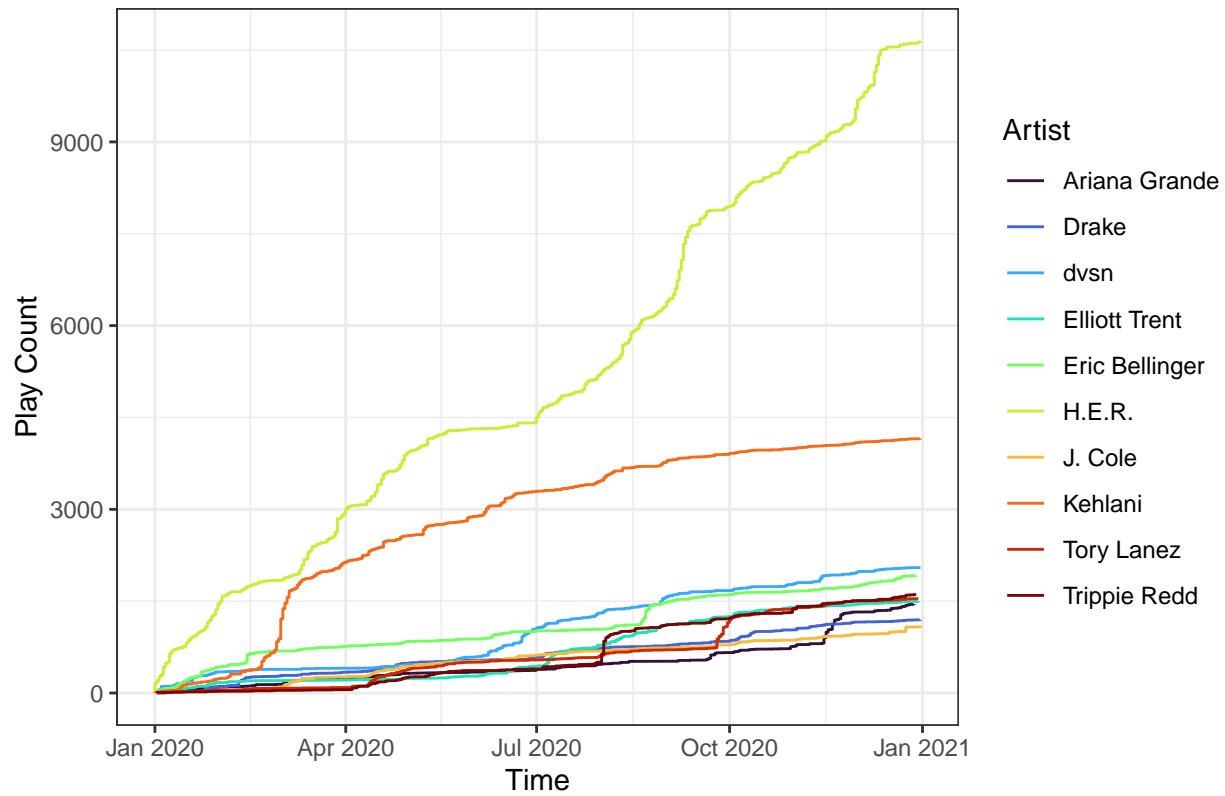
## 10 Most Played Artists (in Each Year) Over the Year

```
for (music_year in music_years) {
  yearly_top10_artists_bytime<- ggplot(alltime_artists_plot_data %>%
    filter(
      play_year == music_year &
      year_x_top10_artist == "Top 10 Artist"),
    aes(x = play_timestamp,
        y = yearly_artist_cumsum,
        color = artist)) +
  geom_path() +
  scale_color_viridis_d(option = "turbo") +
  theme_bw() +
  labs(
    title = paste("10 Most Played Artists in", music_year),
    color = "Artist",
    x = "Time",
    y = "Play Count"
  )
  print(yearly_top10_artists_bytime)
}
```

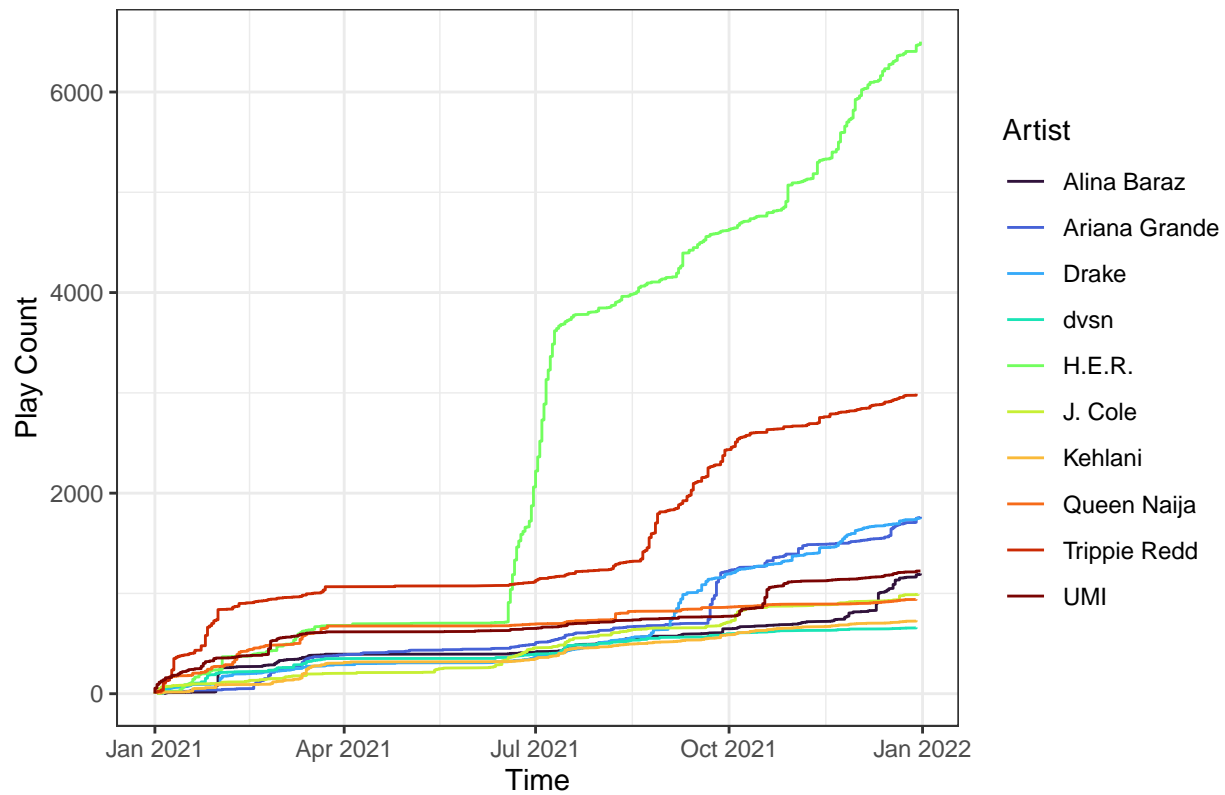
### 10 Most Played Artists in 2019



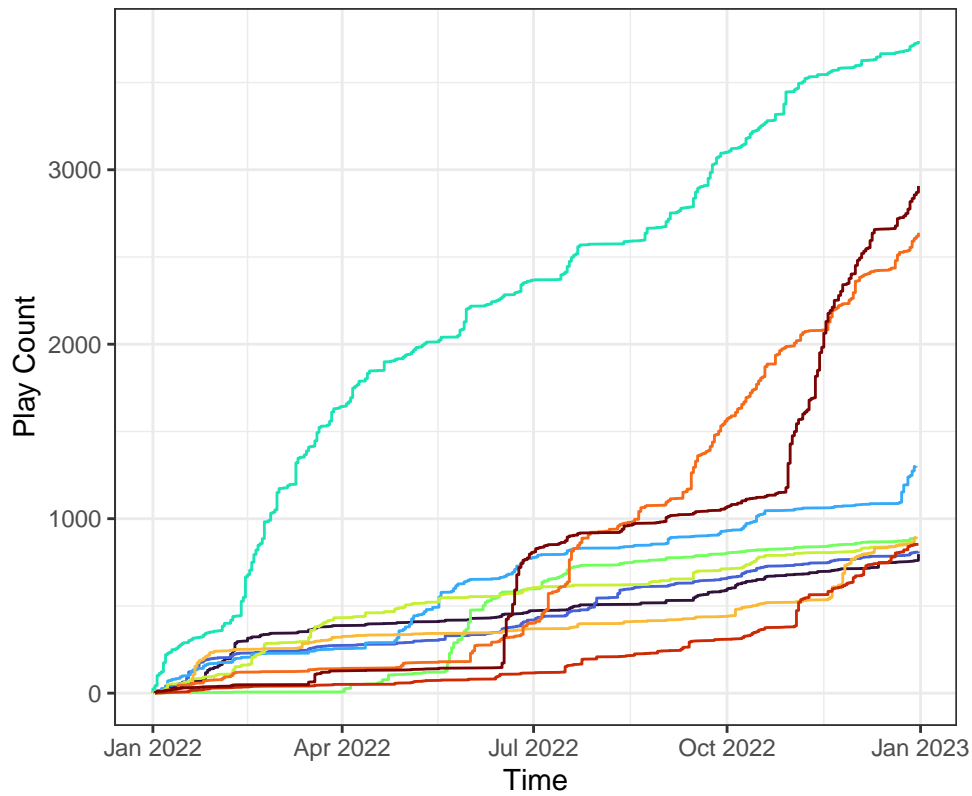
10 Most Played Artists in 2020



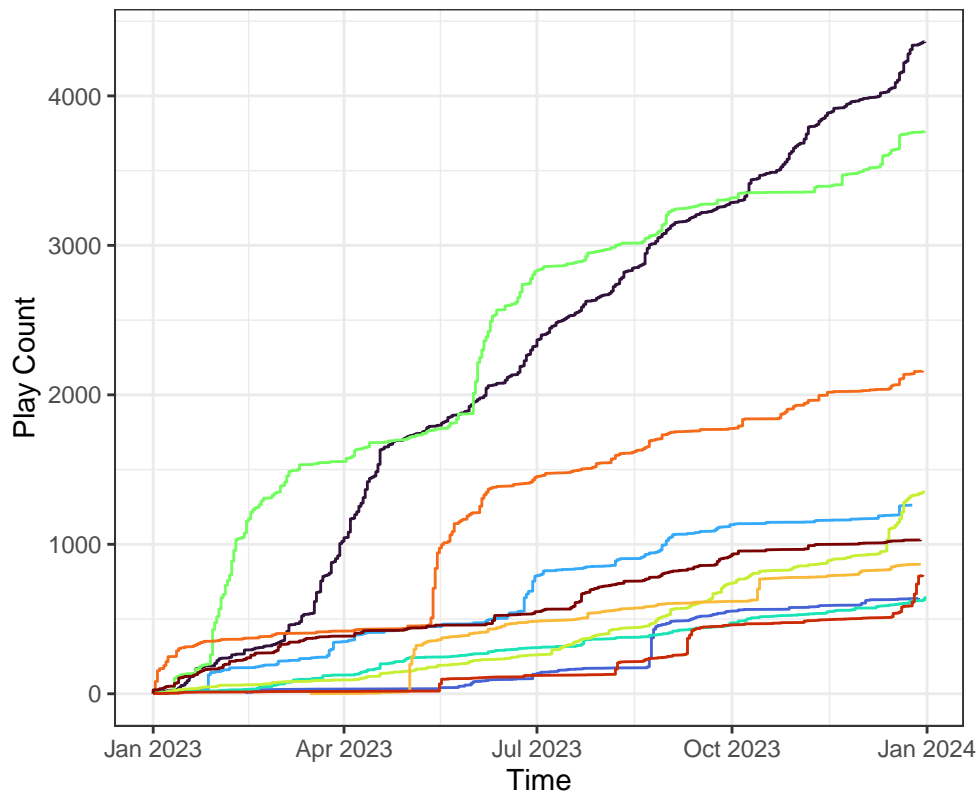
10 Most Played Artists in 2021



10 Most Played Artists in 2022

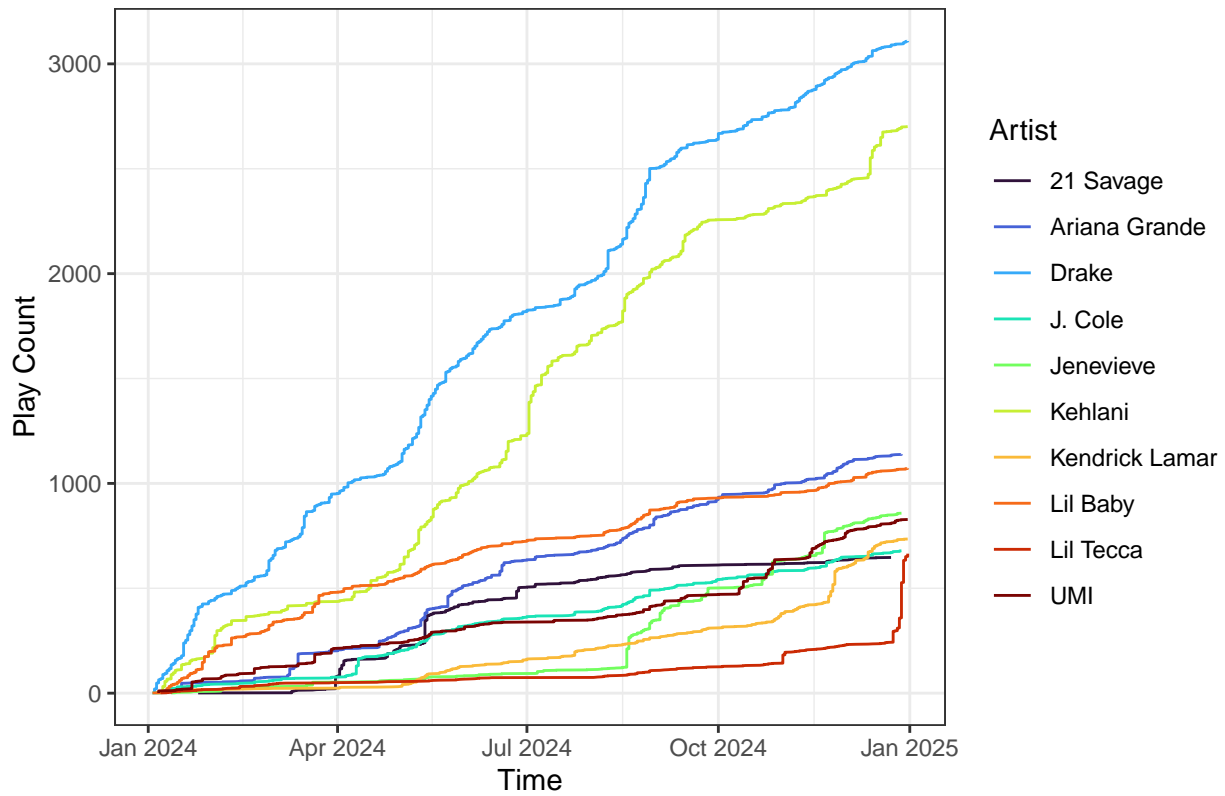


10 Most Played Artists in 2023

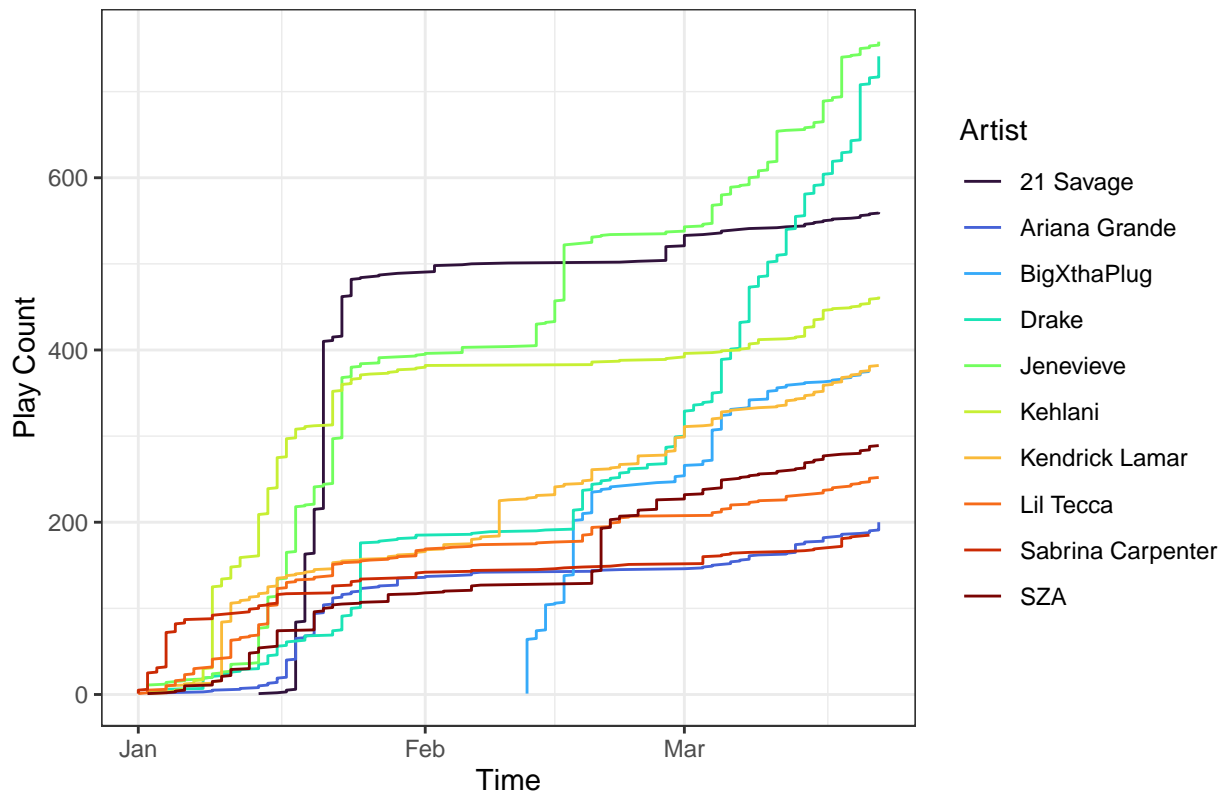




10 Most Played Artists in 2024



10 Most Played Artists in 2025

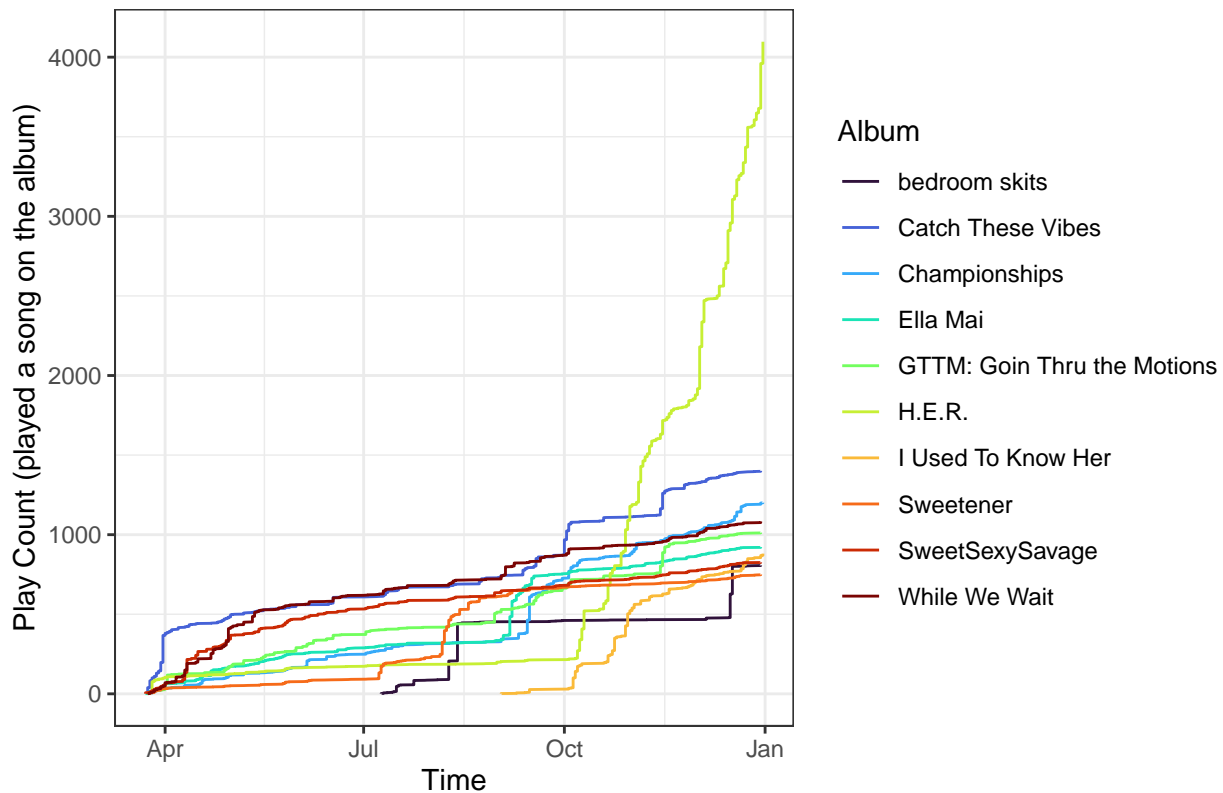


## 10 Most Played Albums (in Each Year) Over the Year

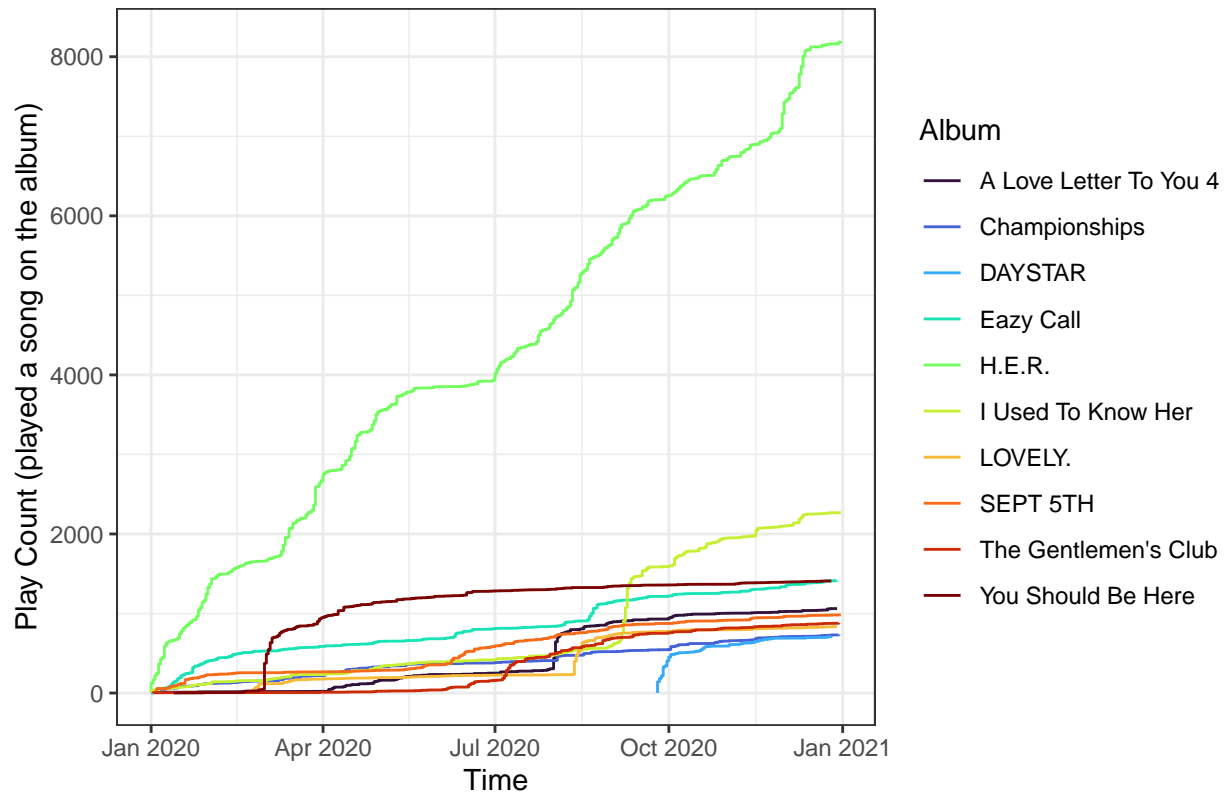
```
for (music_year in music_years) {
  yearly_top10_albums_bytime<- ggplot(alltime_albums_plot_data %>%
    filter(
      play_year == music_year &
      year_x_top10_album == "Top 10 Album"),
    aes(x = play_timestamp,
      y = yearly_album_cumsum,
      color = album)) +

  geom_path() +
  scale_color_viridis_d(option = "turbo") +
  theme_bw() +
  labs(
    title = paste("10 Most Played Albums in", music_year),
    color = "Album",
    x = "Time",
    y = "Play Count (played a song on the album)"
  )
  print(yearly_top10_albums_bytime)
}
```

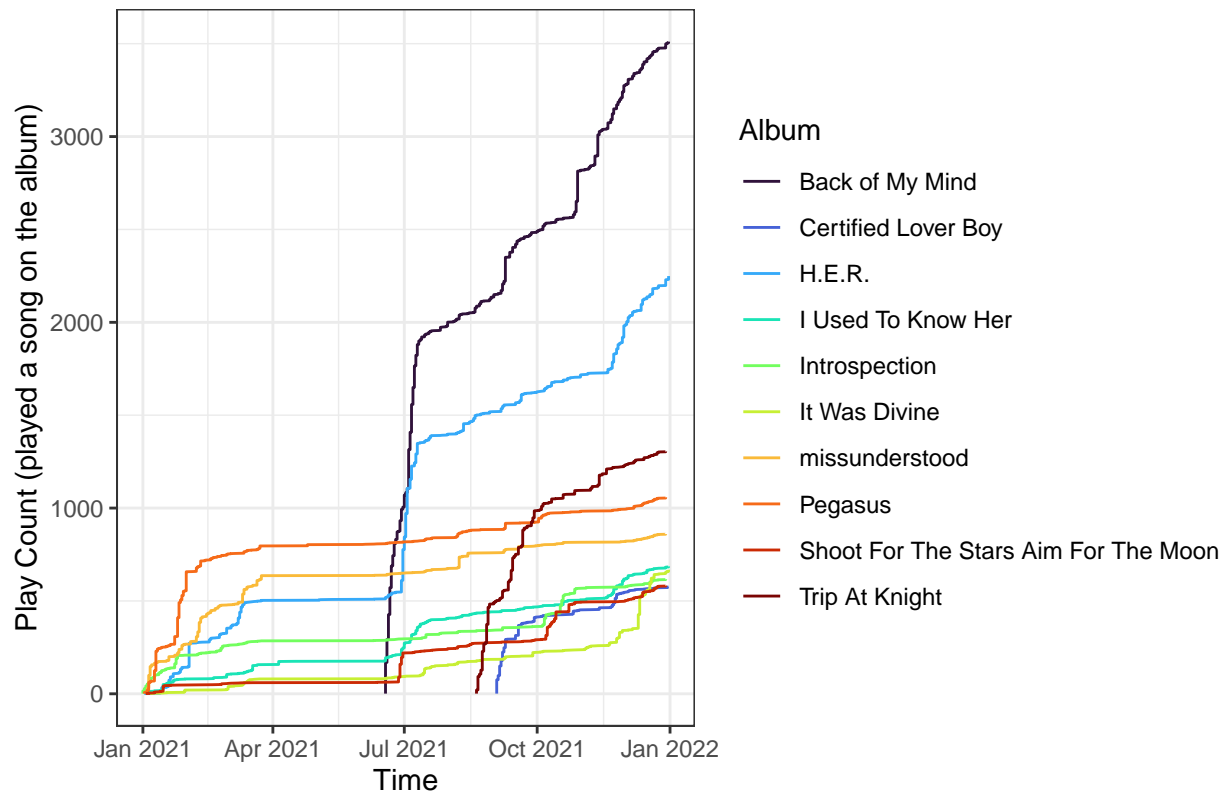
### 10 Most Played Albums in 2019



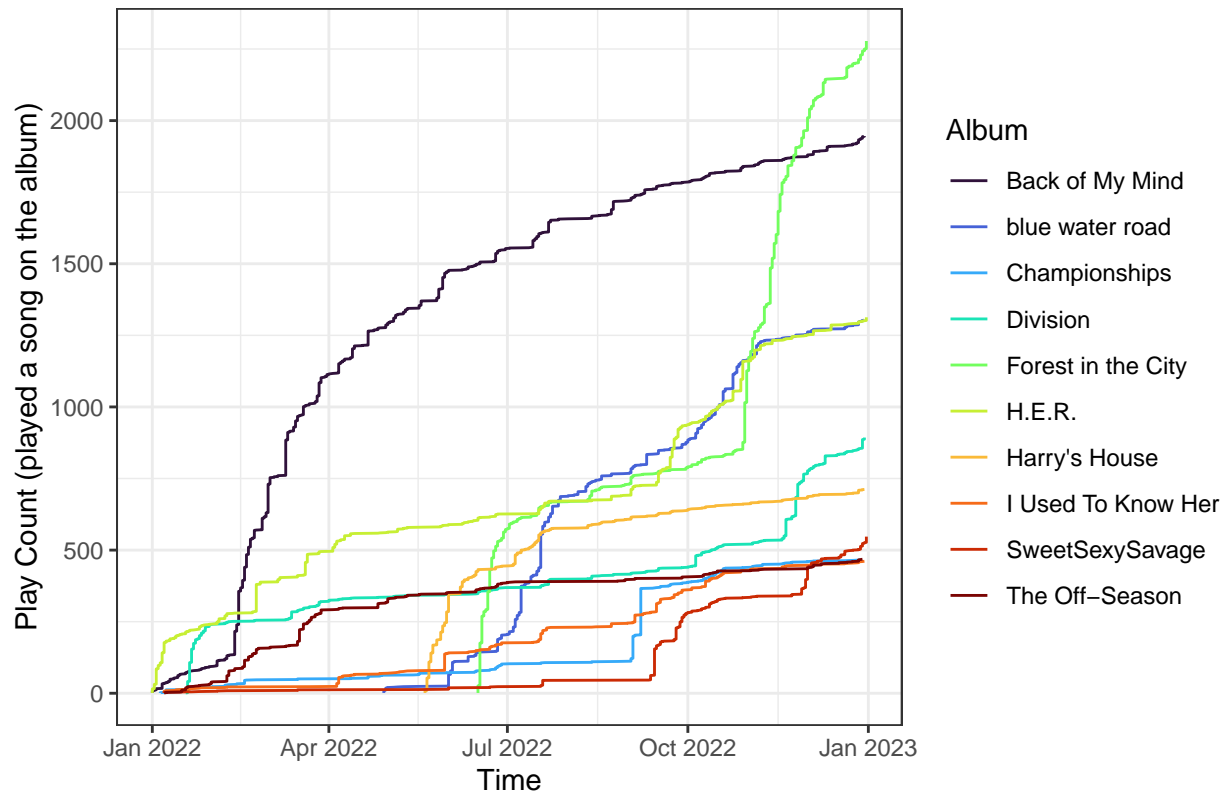
### 10 Most Played Albums in 2020



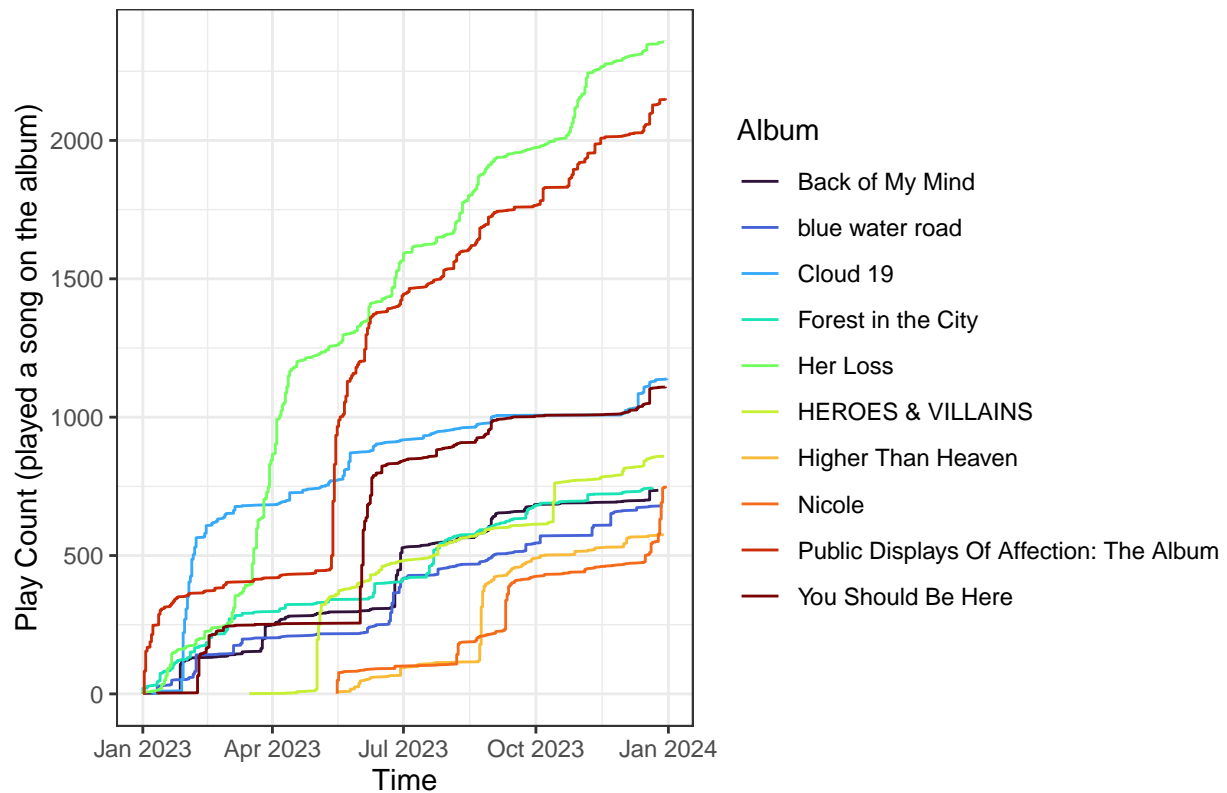
### 10 Most Played Albums in 2021



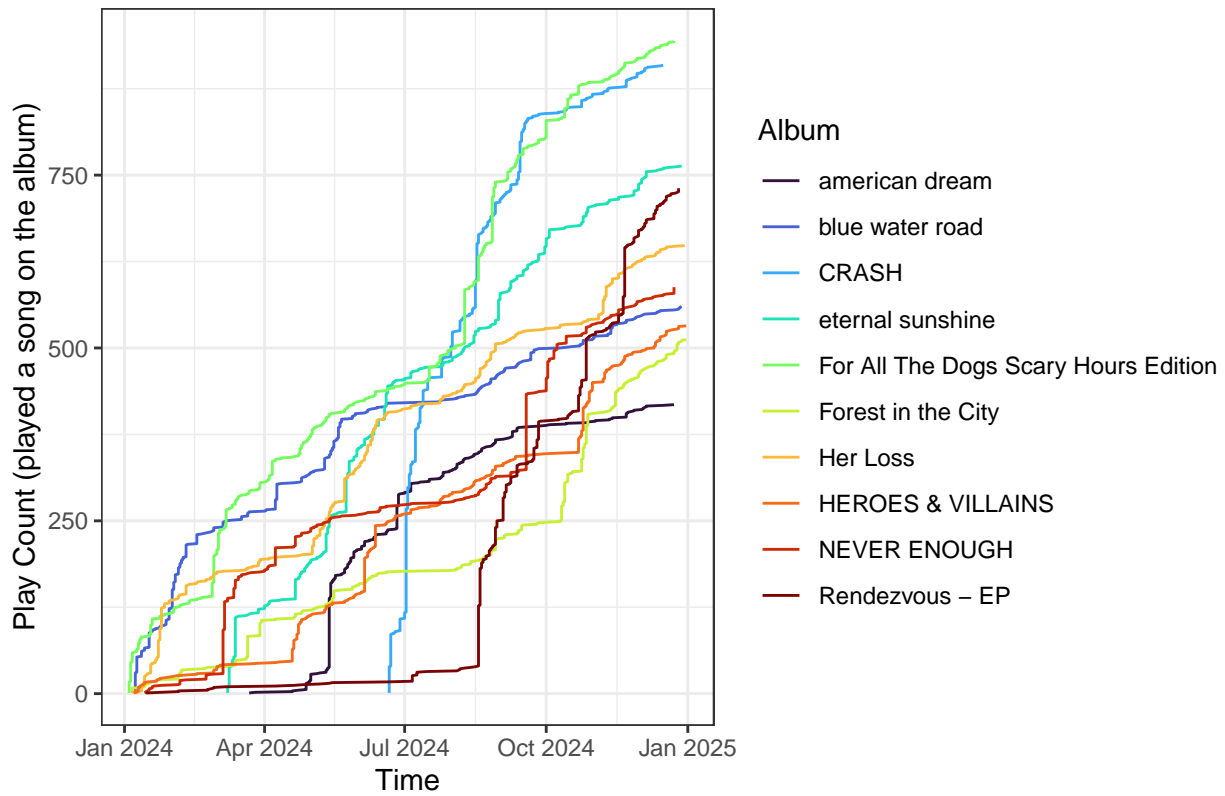
## 10 Most Played Albums in 2022



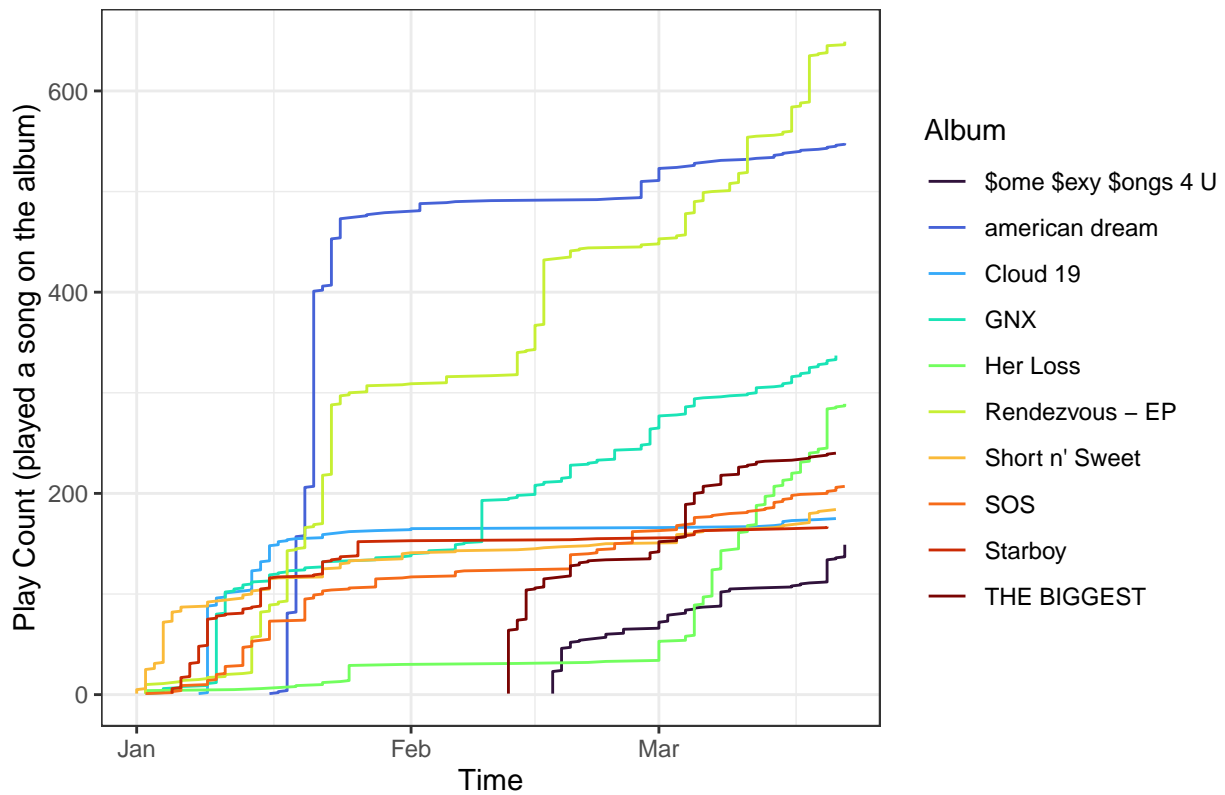
## 10 Most Played Albums in 2023



## 10 Most Played Albums in 2024



## 10 Most Played Albums in 2025



## Significant Dates

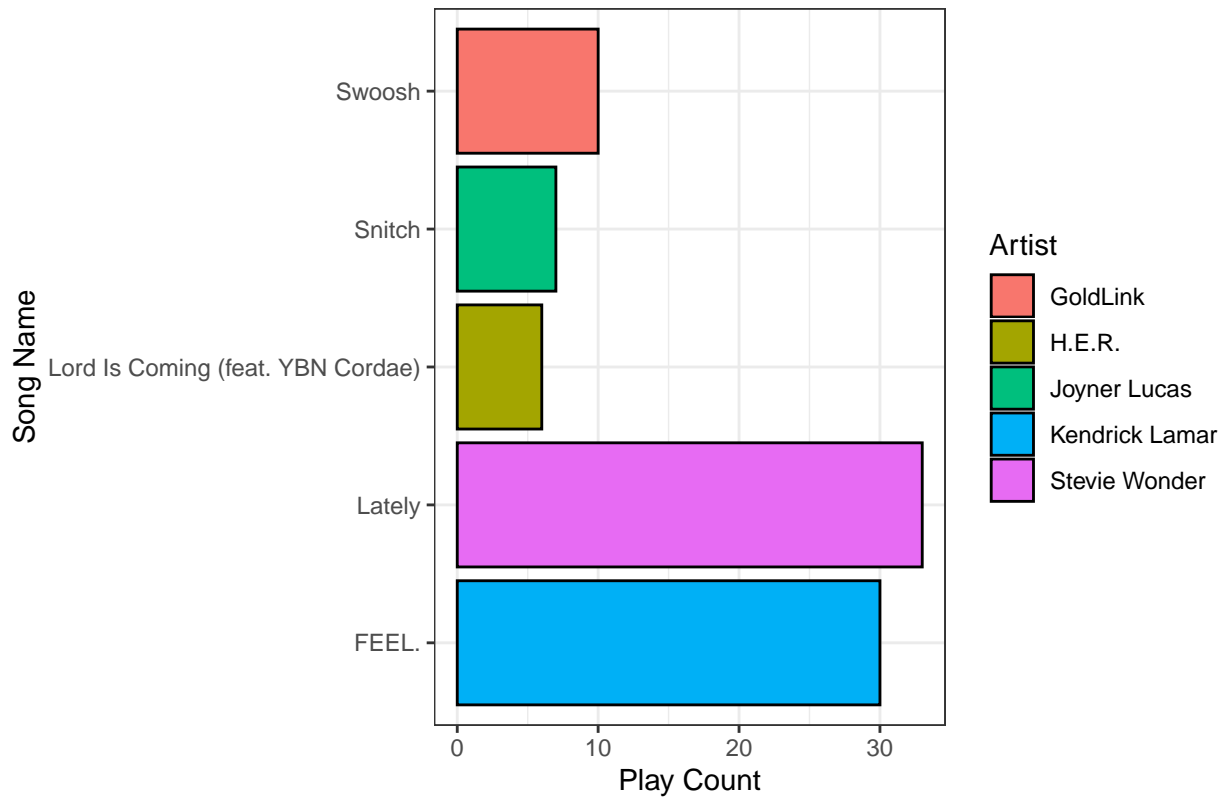
Music has been something I've always loved, most days I listen to music for hours upon hours. When I'm feeling certain emotions or I'm going through something significant, I tend to turn to music that aligns with those emotions.

```
significant_dates <- c("2020-10-23", "2021-06-26", "2021-12-25", "2022-02-14",
                      "2022-05-11", "2022-07-10", "2022-09-07", "2022-11-12",
                      "2022-12-24", "2024-11-05", "2025-02-14", "2025-03-11")
significant_dates <- as.Date(significant_dates)

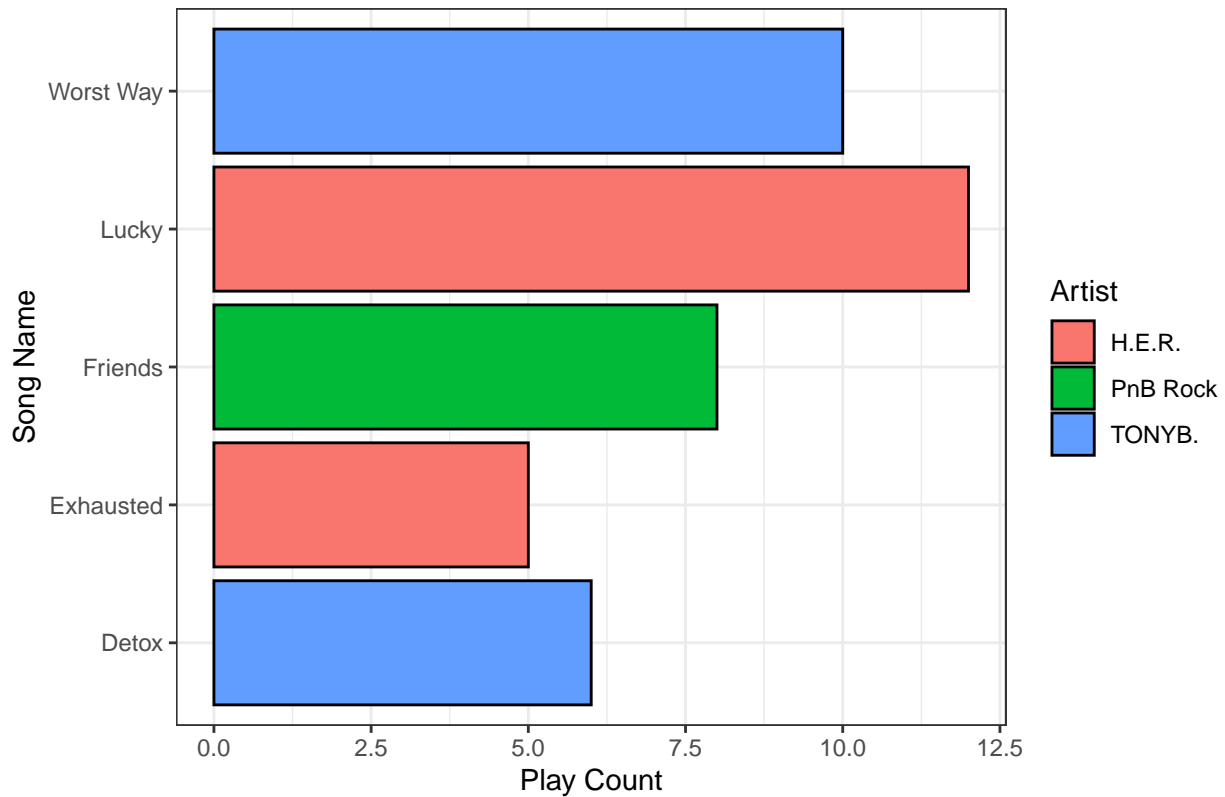
daily_songs <- streaming_data_2 %>%
  group_by(play_timestamp, song, artist) %>%
  summarize(daily_song_count = n(), .groups = "drop") %>%
  arrange(play_timestamp, desc(daily_song_count)) %>%
  group_by(play_timestamp) %>%
  slice_head(n = 5)

for (sig_date in significant_dates) {
  sig_date_plot <- ggplot(daily_songs %>%
    filter(play_timestamp == sig_date),
    aes(x = daily_song_count, y = song, fill = artist)) +
    geom_bar(stat = "identity", position = "dodge", color = "black") +
    theme_bw() +
    labs(
      title = paste("Top 5 Songs On", as.Date(sig_date),
                    ifelse(
                      as.Date(sig_date) %in%
                        as.Date(c("2021-06-26", "2021-12-25",
                                "2022-07-10", "2022-02-14",
                                "2022-05-11", "2025-02-14",
                                "2025-03-11")),
                      "(Favorite Days)", "(Worst Days)")),
      y = "Song Name",
      x = "Play Count",
      fill = "Artist"
    )
  print(sig_date_plot)
}
```

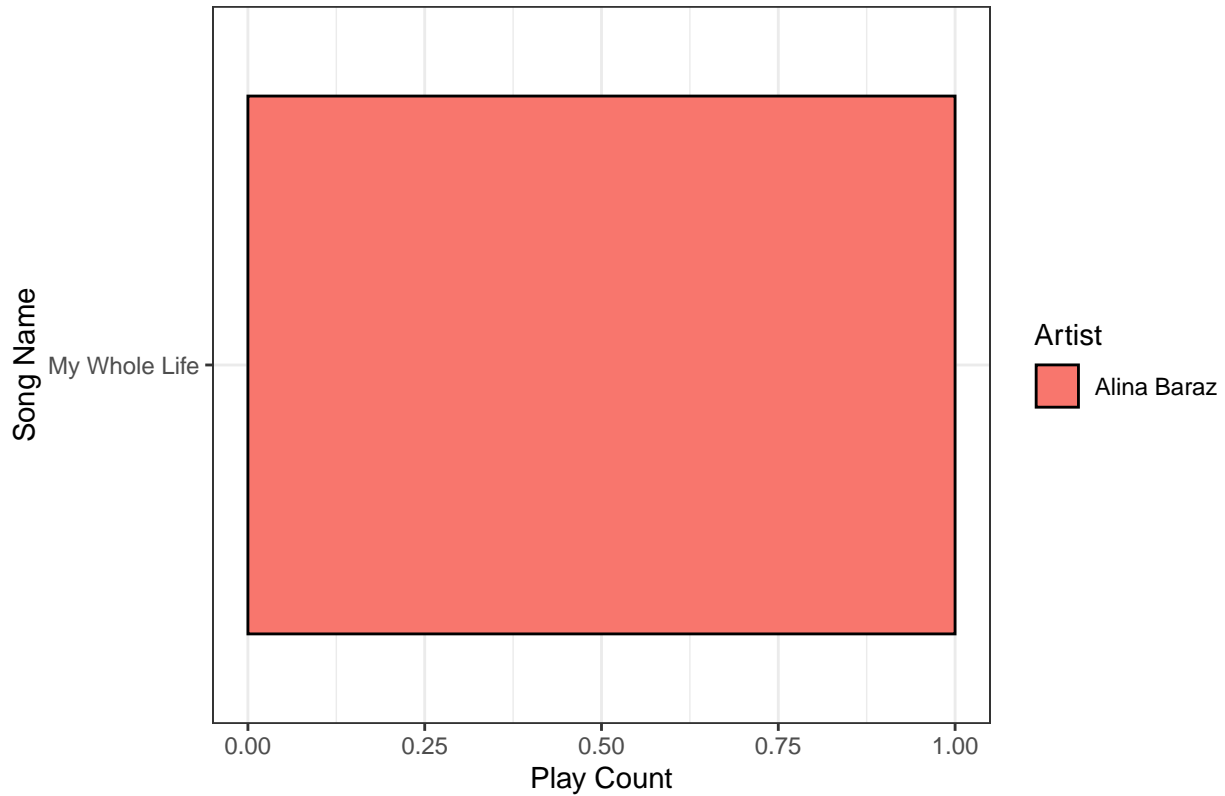
Top 5 Songs On 2020-10-23 (Worst Days)



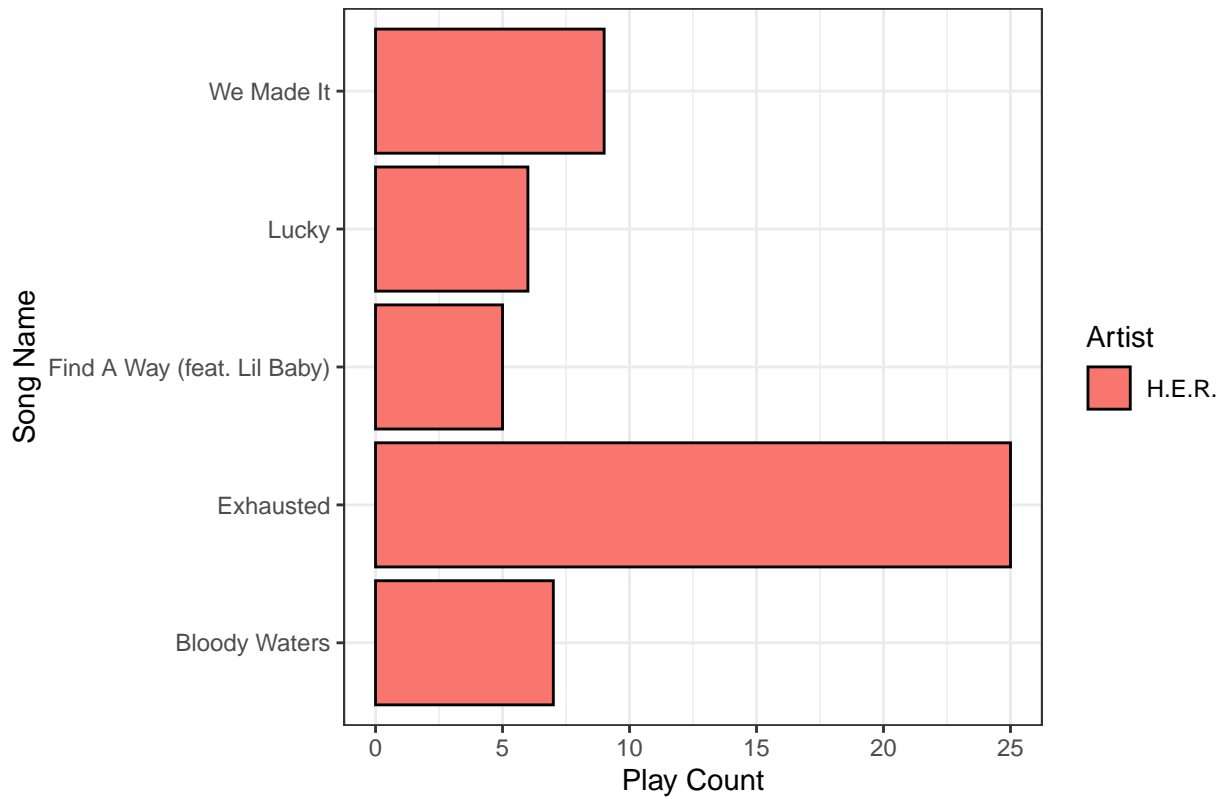
Top 5 Songs On 2021-06-26 (Favorite Days)



Top 5 Songs On 2021-12-25 (Favorite Days)

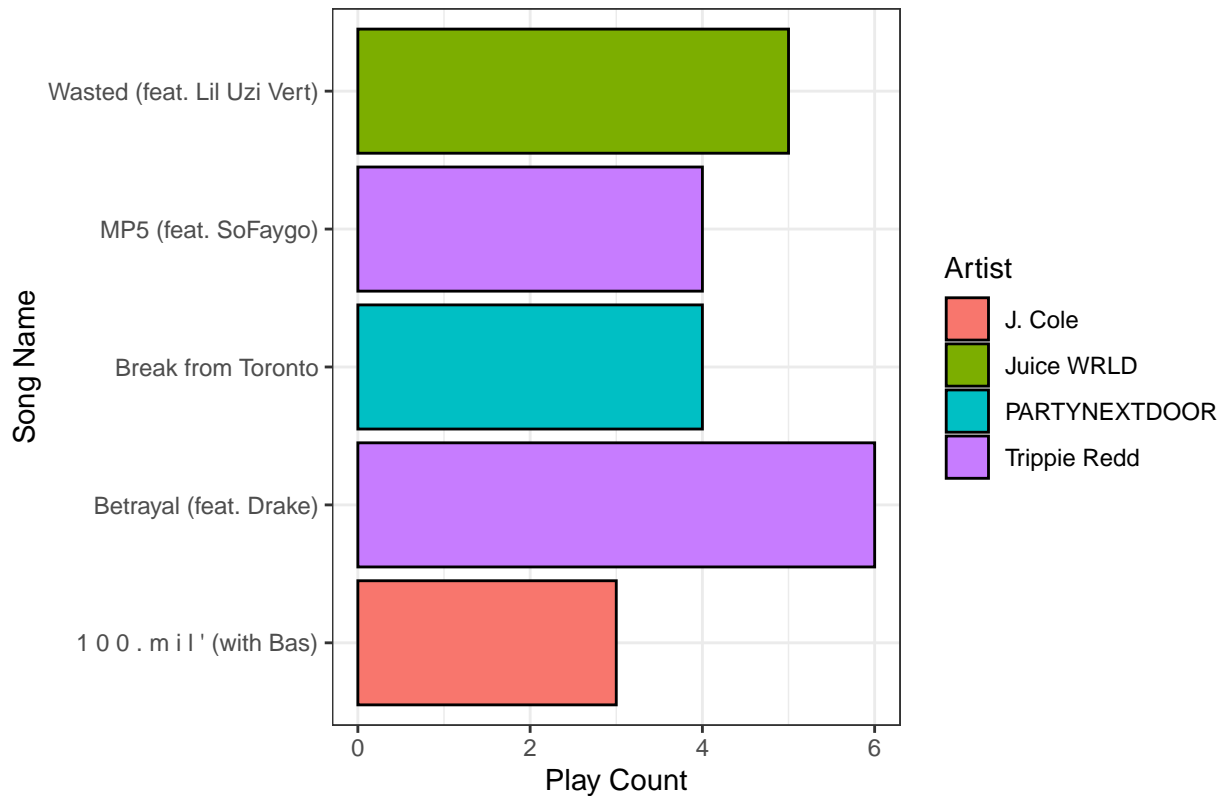


Top 5 Songs On 2022-02-14 (Favorite Days)

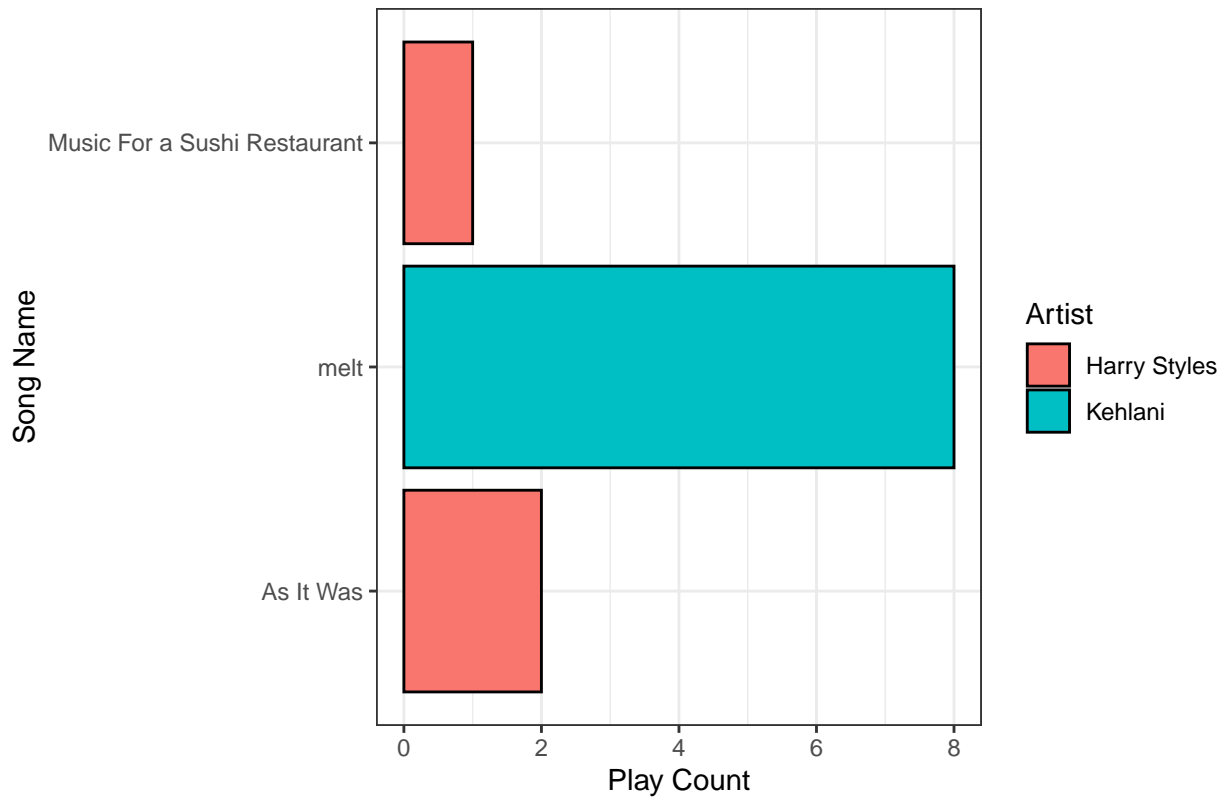




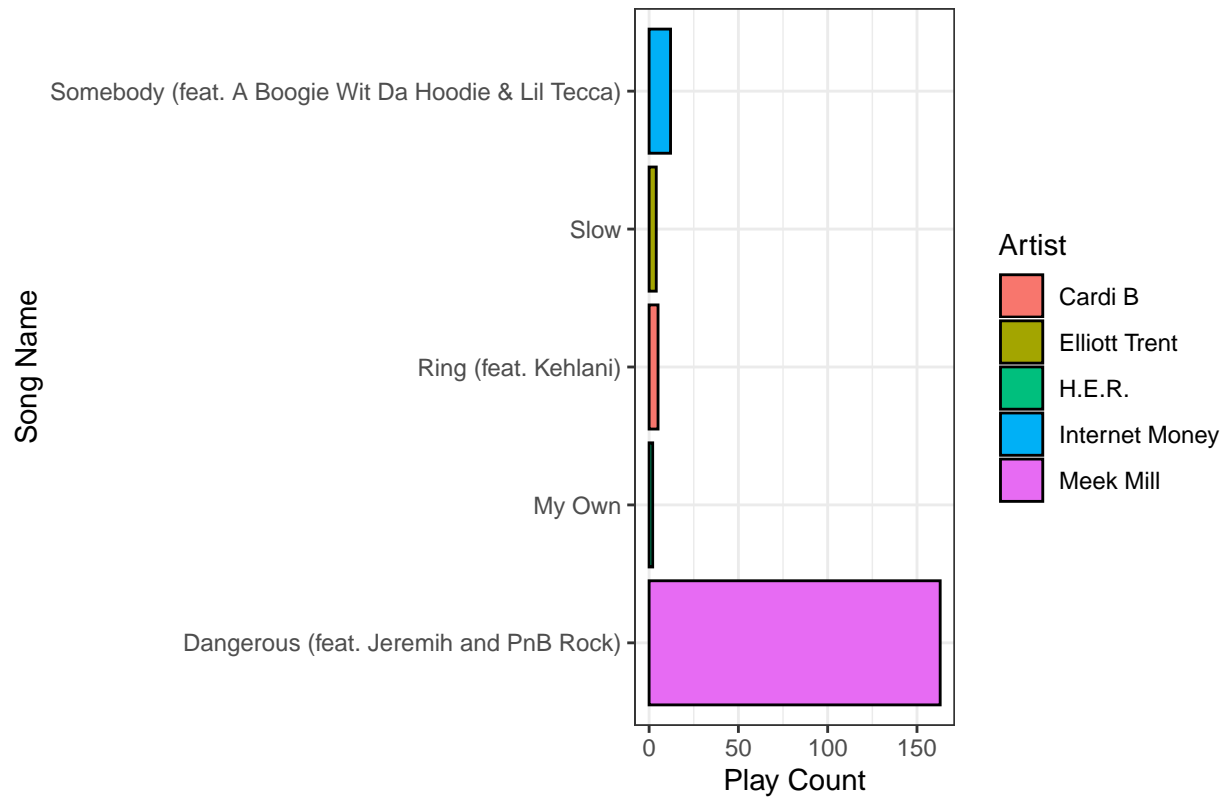
Top 5 Songs On 2022-05-11 (Favorite Days)



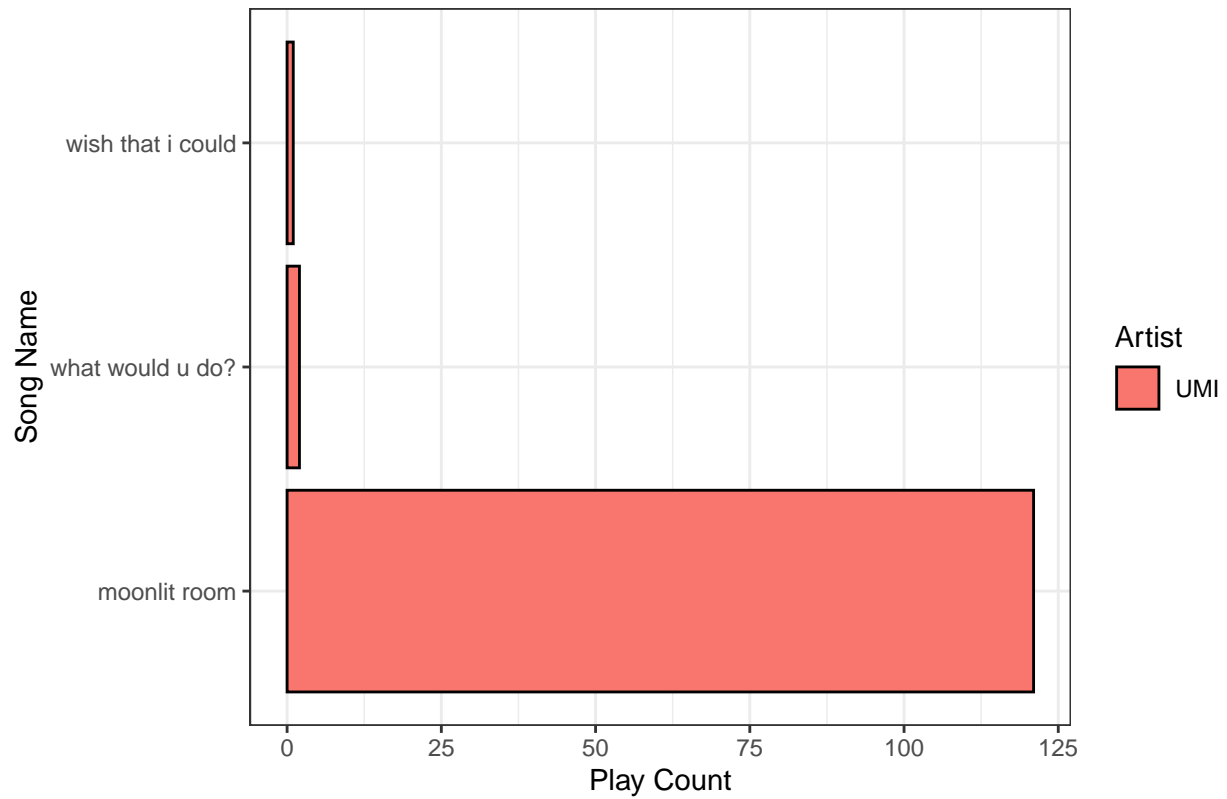
Top 5 Songs On 2022-07-10 (Favorite Days)



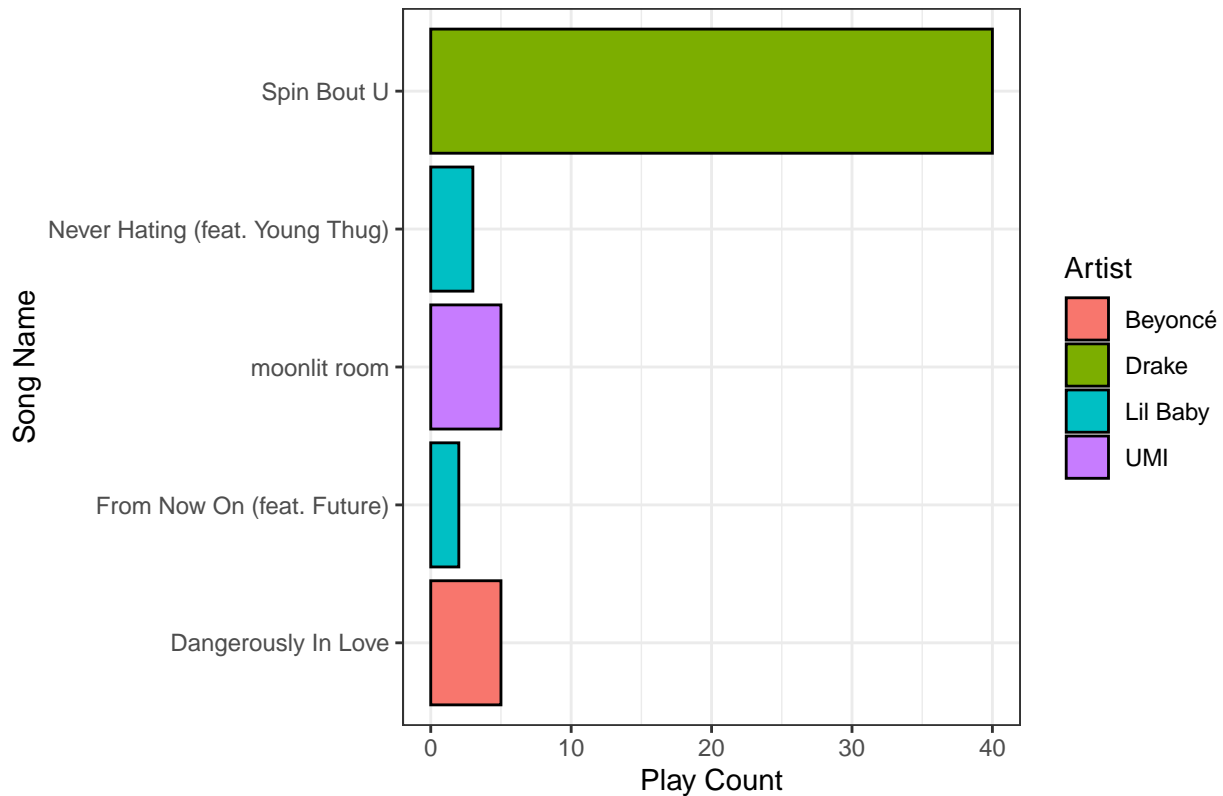
Top 5 Songs On 2022-09-07 (Worst Da



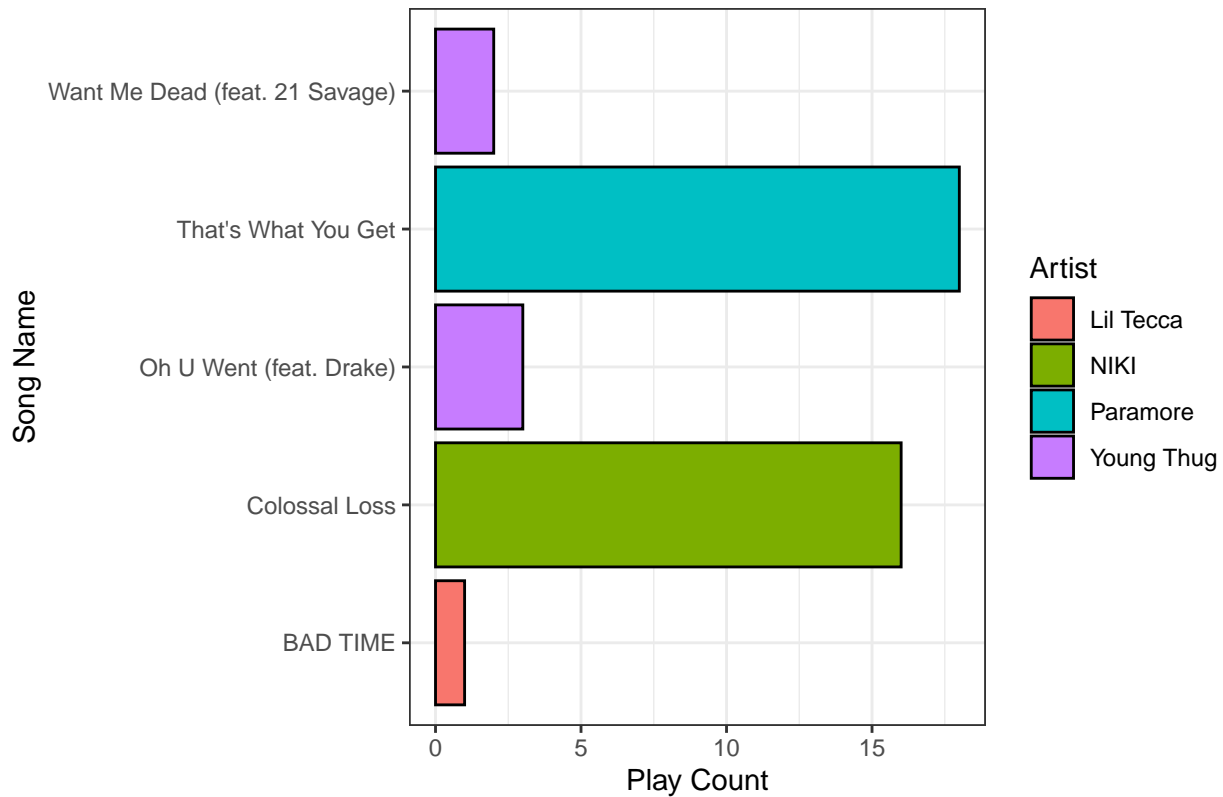
Top 5 Songs On 2022-11-12 (Worst Days)



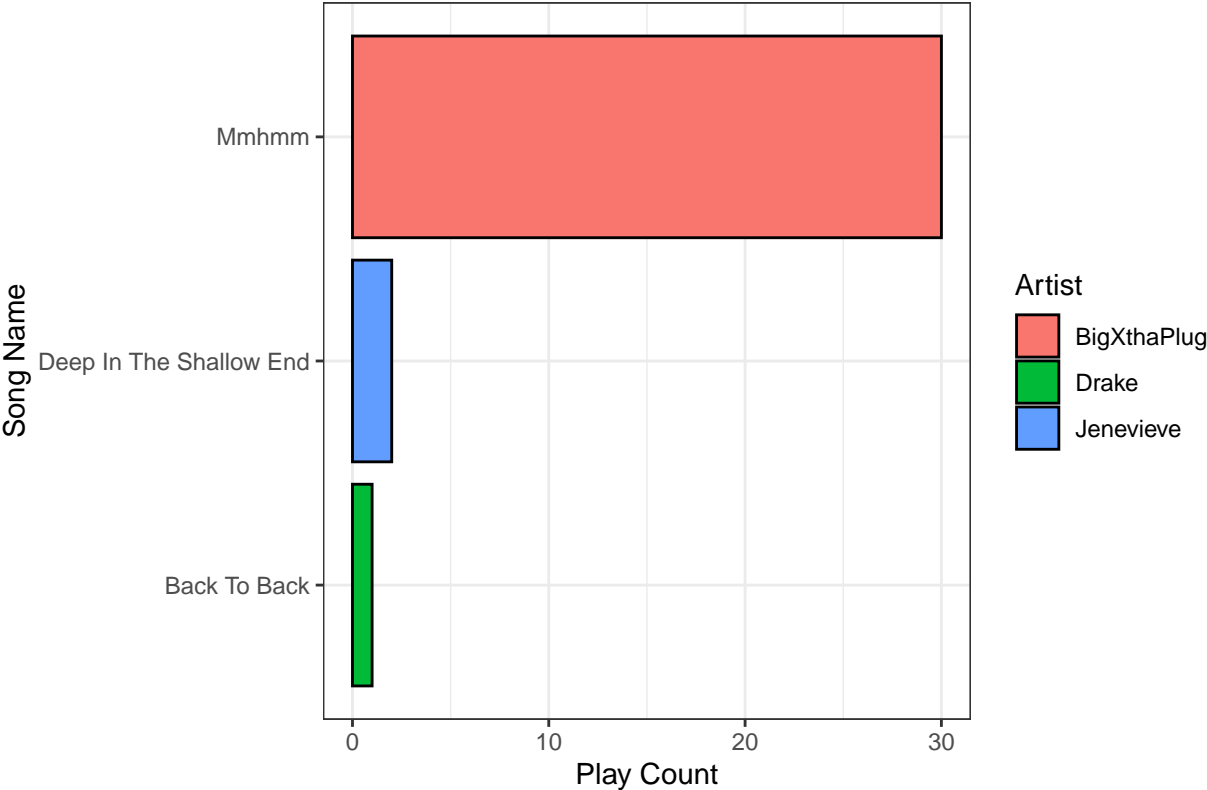
Top 5 Songs On 2022-12-24 (Worst Days)



Top 5 Songs On 2024-11-05 (Worst Days)



Top 5 Songs On 2025-02-14 (Favorite Days)



Top 5 Songs On 2025-03-11 (Favorite Days)

