Spotify Data Analysis

Khoi Trinh

2023-02-21

Analyzing My Spotify Streaming History

All of the codes are adapted from this article

First, here are the required libraries

```
library(jsonlite)
library(lubridate)
library(gghighlight)
library(spotifyr)
library(tidyverse)
library(knitr)
library(ggplot2)
library(plotly)
```

Let's read in our data, you can find how to get your own Spotify data here

I have 5 of these history files, but you may have more or less.

```
streamHistory0 <- data.frame(fromJSON("StreamingHistory0.json", flatten = TRUE))
streamHistory1 <- data.frame(fromJSON("StreamingHistory1.json", flatten = TRUE))
streamHistory2 <- data.frame(fromJSON("StreamingHistory2.json", flatten = TRUE))
streamHistory3 <- data.frame(fromJSON("StreamingHistory3.json", flatten = TRUE))
streamHistory4 <- data.frame(fromJSON("StreamingHistory4.json", flatten = TRUE))
streamHistory5 <- data.frame(fromJSON("StreamingHistory5.json", flatten = TRUE))

streamHistory <- Map(c, streamHistory0, streamHistory1, streamHistory2,
                     streamHistory3, streamHistory4, streamHistory5)
```

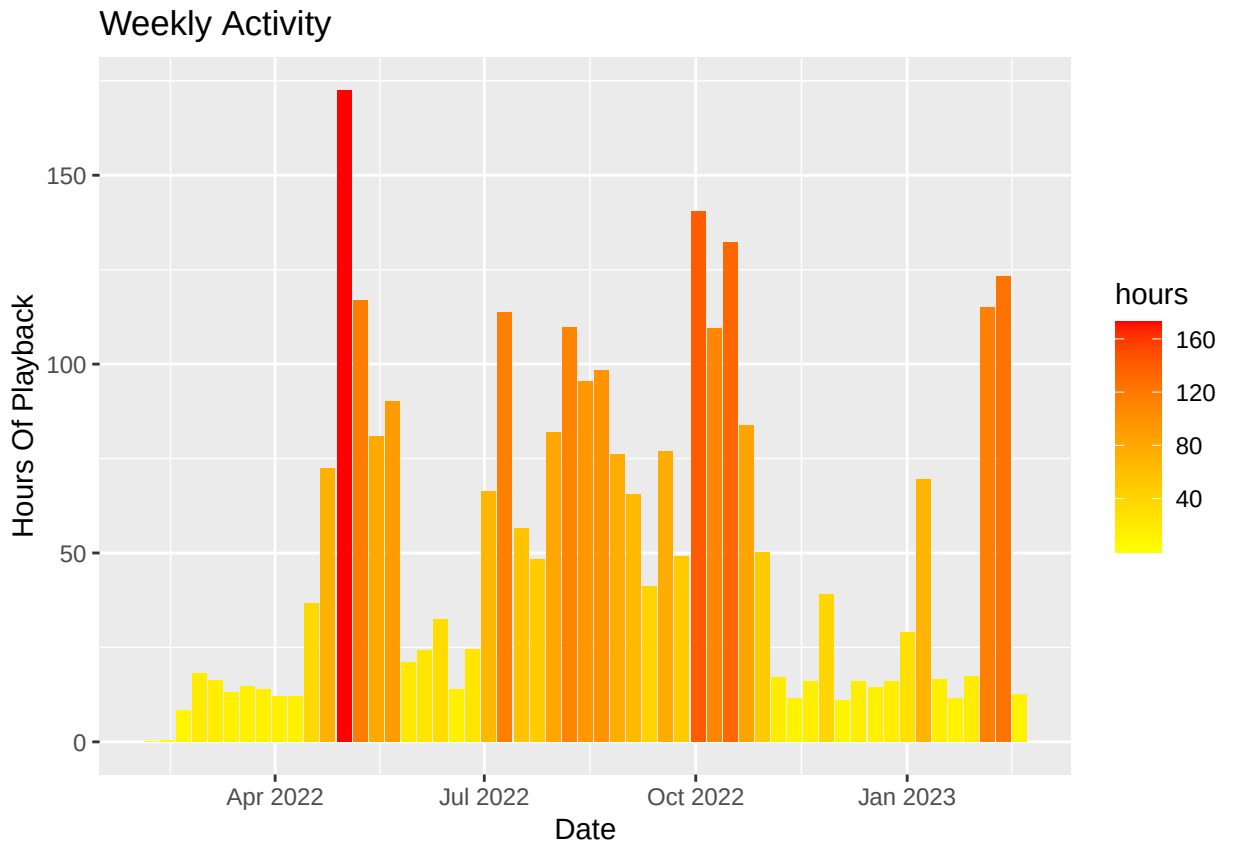
On what day did I listen to more or less music?

```
# Add date and time
mySpotify <- streamHistory %>%
  as_tibble() %>%
  mutate_at("endTime", ymd_hm) %>%
  mutate(endTime = endTime - hours(6)) %>%
  mutate(date = floor_date(endTime, "day") %>% as_date,
         seconds = msPlayed / 1000, minutes = seconds / 60)

# Playback activity per week and hours
streamingHours <- mySpotify %>%
  filter(date >= "2020-01-01") %>%
```

```
  group_by(date) %>%
  group_by(date = floor_date(date, "week")) %>%
  summarize(hours = sum(minutes) / 60) %>%
  arrange(date) %>%
  ggplot(aes(x = date, y = hours)) +
  geom_col(aes(fill = hours)) +
  scale_fill_gradient(low = "yellow", high = "red") +
  labs(x= "Date", y= "Hours Of Playback") +
  ggtitle("Weekly Activity")

streamingHours
```



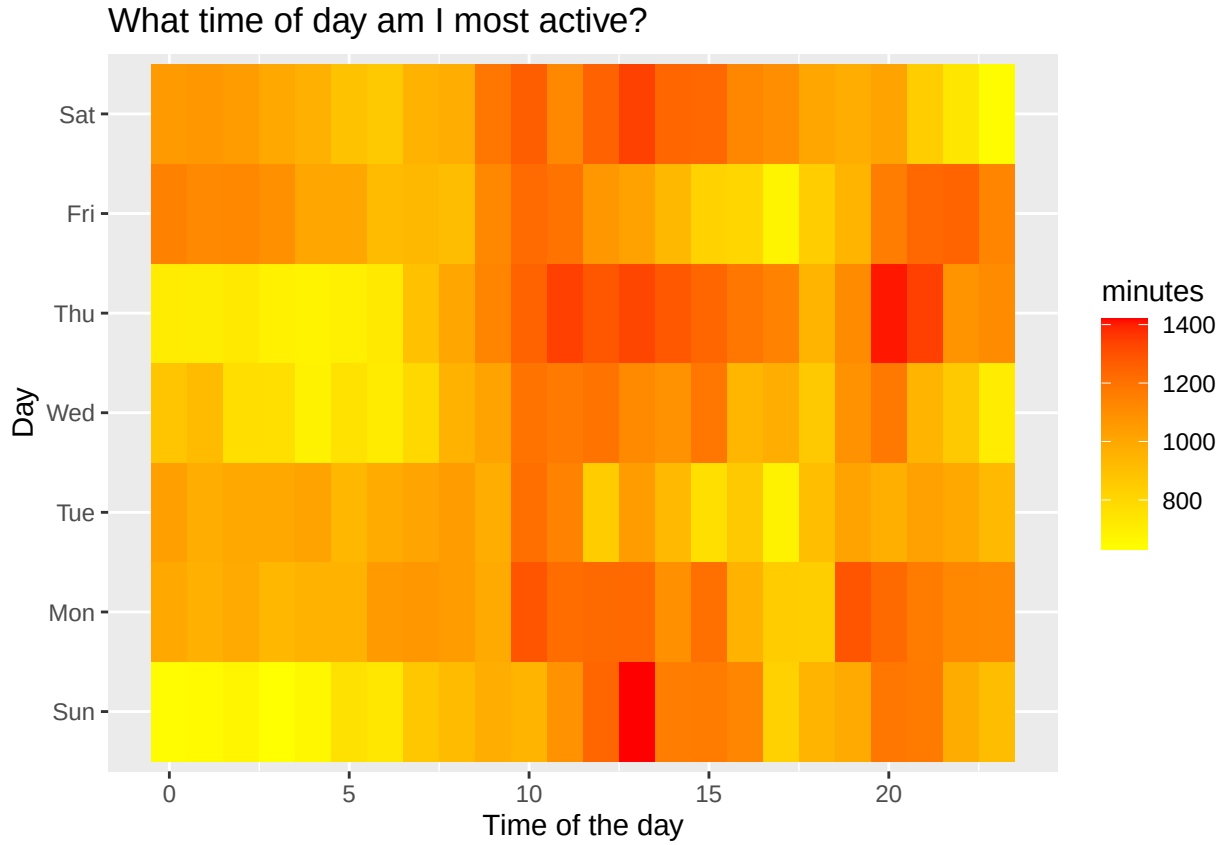
So, we know roughly what time of the year I listened to Spotify the most. Let's look at the data on a by-week basis.

```
hoursDay <- mySpotify %>%
  filter(date >= "2022-01-01") %>%
  group_by(date, hour = hour(endTime), weekday = wday(date, label = TRUE))%>%
  summarize(minutesListened = sum(minutes))
```

```
## `summarise()` has grouped output by 'date', 'hour'. You can override using the
## `.groups` argument.
```

```
hoursDay %>%
  group_by(weekday, hour) %>%
  summarize(minutes = sum(minutesListened)) %>%
  ggplot(aes(x = hour, weekday, fill = minutes)) +
  geom_tile() +
  scale_fill_gradient(low = "yellow", high = "red") +
  labs(x= "Time of the day", y= "Day") +
  ggtitle("What time of day am I most active?")
```

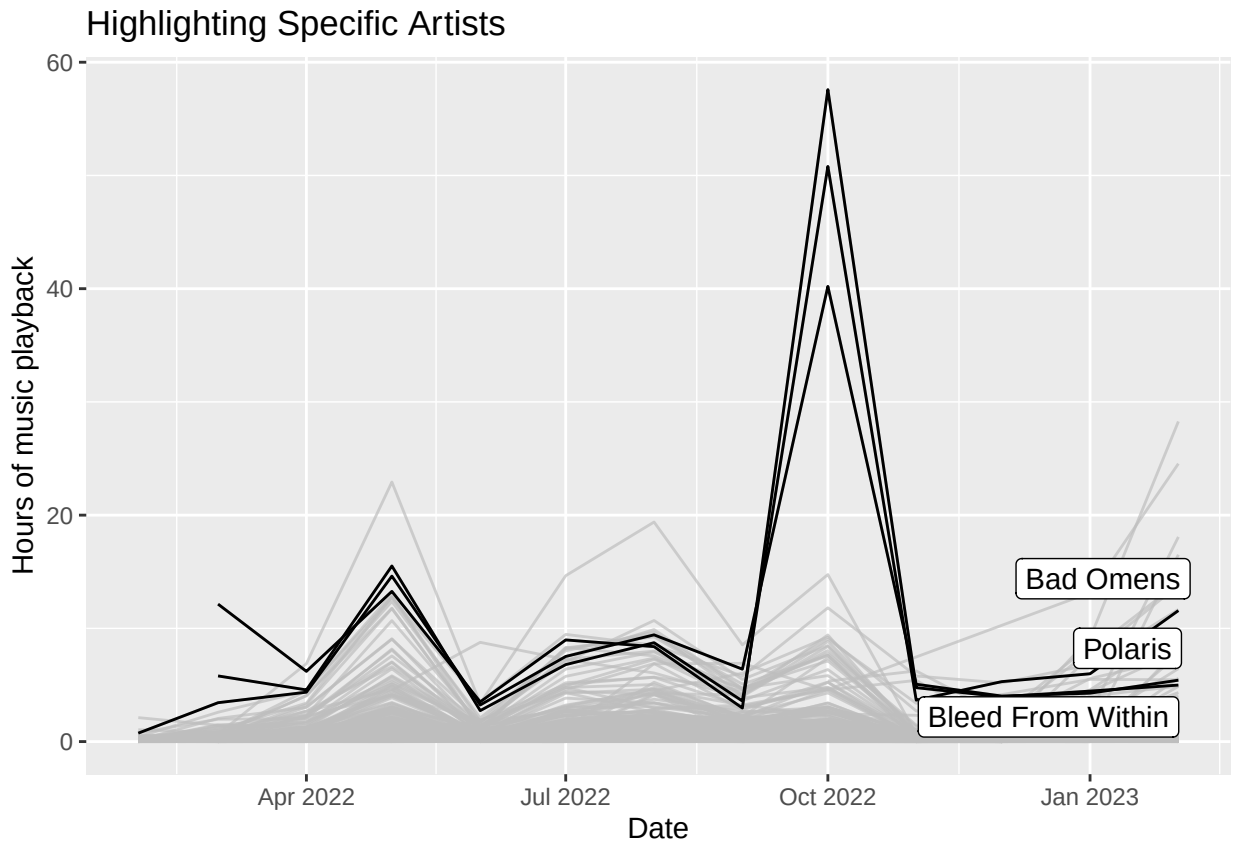
```
## `summarise()` has grouped output by 'weekday'. You can override using the
## `.groups` argument.
```



How about streaming time by a specific artist? I know I listened to a lot of Bad Omens, Bleed From Within, and Polaris.

```
hoursArtist <- mySpotify %>%
  group_by(artistName, date = floor_date(date, "month")) %>%
  summarize(hours = sum(minutes) / 60) %>%
  ggplot(aes(x = date, y = hours, group = artistName)) +
  labs(x= "Date", y= "Hours of music playback") +
  ggtitle("Highlighting Specific Artists") +
  geom_line() +
  gghighlight(artistName == "Bad Omens" || artistName == "Bleed From Within"
              || artistName == "Polaris")
```

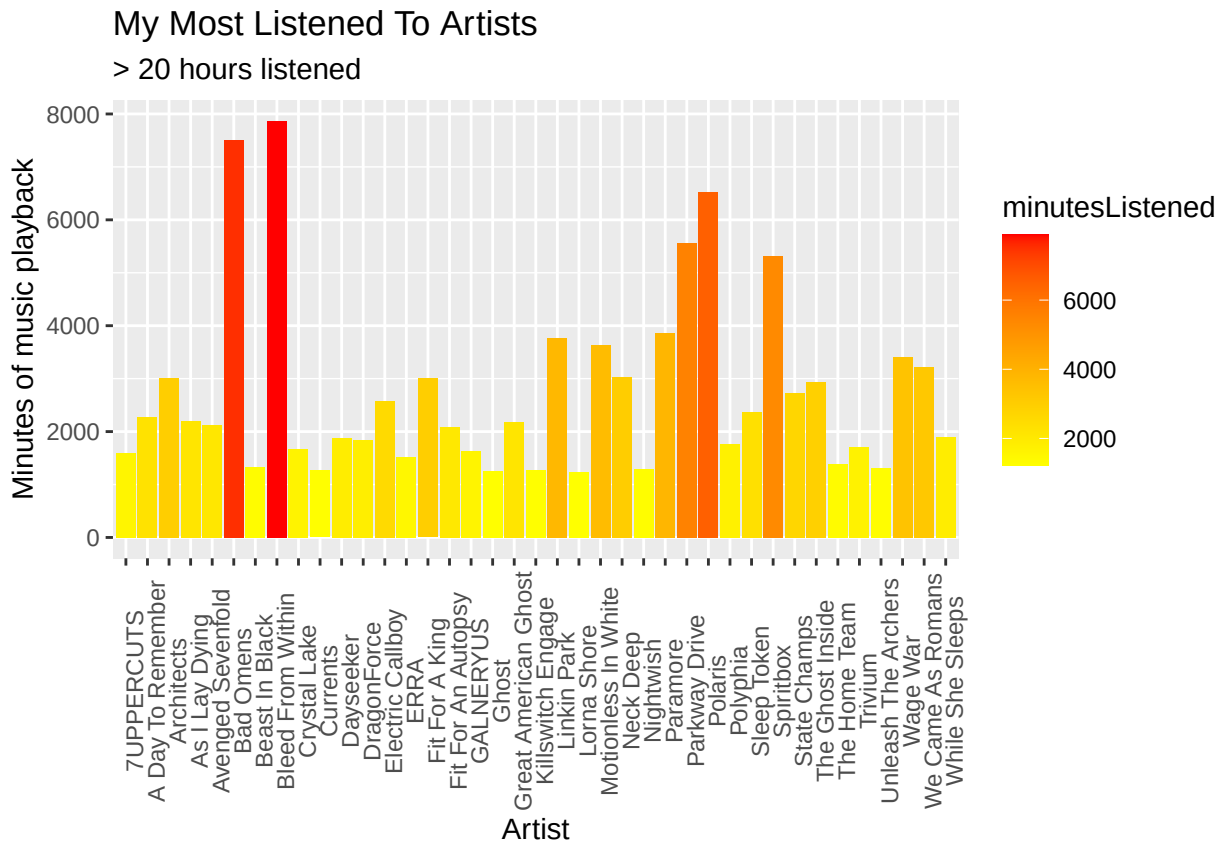
```
hoursArtist
```



Finally, let's get my most listened to artist(s) in 2022

```
topArtists <- mySpotify %>%
  filter(date >= "2022-01-01") %>%
  group_by(artistName) %>%
  summarize(minutesListened = sum(minutes)) %>%
  filter(minutesListened >= 1200) %>%
  ggplot(aes(x = artistName, y = minutesListened)) +
  geom_col(aes(fill = minutesListened)) +
  scale_fill_gradient(low = "yellow", high = "red") +
  labs(x= "Artist", y= "Minutes of music playback") +
  ggtitle("My Most Listened To Artists", "> 20 hours listened") +
  theme(axis.text.x = element_text(angle = 90))

topArtists
```


My Most Listened To Artists
> 20 hours listened
Minutes of music playback
8000
6000
4000
2000
0
7UPPERCUTS
A Day To Remember
Architects
As I Lay Dying
Avenged Sevenfold
Bad Omens
Beast In Black
Bleed From Within
Crystal Lake
Currents
Dayseeker
DragonForce
Electric Callboy
ERRA
Fit For A King
Fit For An Autopsy
GALNERYUS
Ghost
Great American Ghost
Killswitch Engage
Linkin Park
Lorna Shore
Motionless In White
Neck Deep
Nightwish
Paramore
Parkway Drive
Polaris
Polyphia
Sleep Token
Spiritbox
State Champs
The Ghost Inside
The Home Team
Trivium
Unleash The Archers
Wage War
We Came As Romans
While She Sleeps
Artist
minutesListened
6000
4000
2000