My Spotify Activity - 1

Ananya Mantravadi

Synopsis

This project involves exploring and visualizing my activity using my personal data from Spotify with R. It aims at answering the following:
1. What is my streaming activity measured in hours per week between June ’20 to June ’21?
2. How do my times listening to various artists compare?
3. Which artists have I listened to for more than 5 hours?
4. How is my streaming activity at different times of the day between June ’20 to June ’21?
5. How is my streaming activity at different times of the day?
6. How is my streaming activity at different times of the day during the days of the week?
7. How do my times listening in weekdays and weekends compare?

Importing Libraries and Reading Data

```r
#Importing libraries and reading the data

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

s1 <-fromJSON("StreamingHistory0.json",flatten = TRUE)
s2 <-fromJSON("StreamingHistory1.json",flatten = TRUE)

streaming_history <- rbind(s1,s2)
```

Sample of the data:

```r
head(streaming_history,10)
```

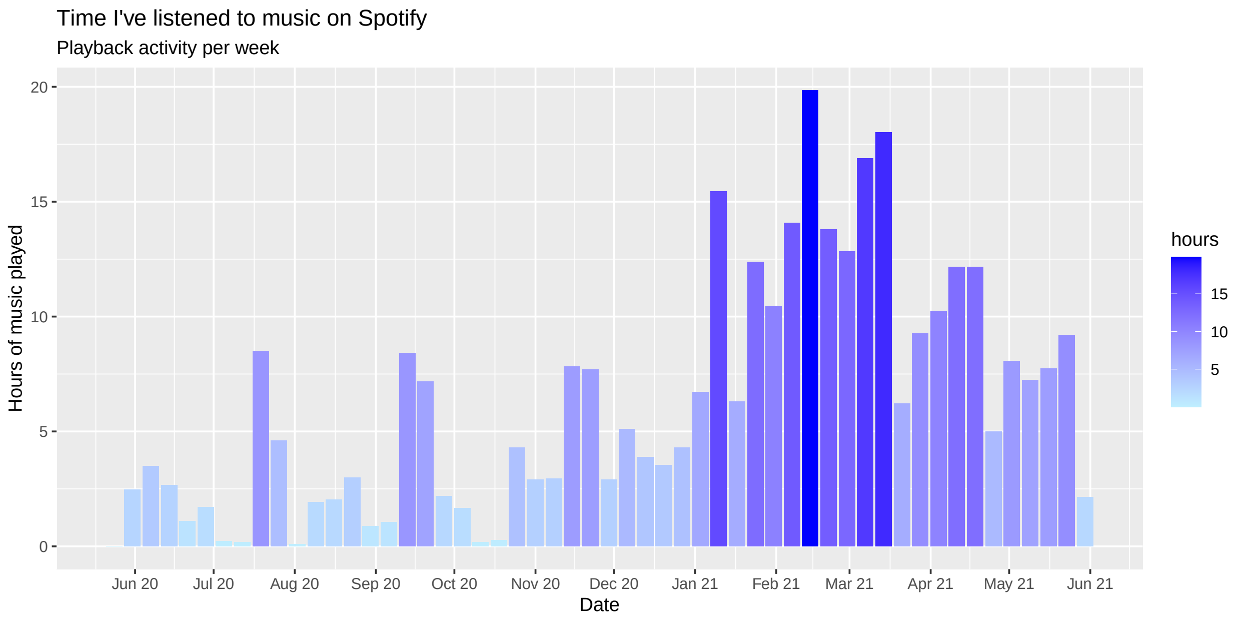
```
##              endTime   artistName                       trackName msPlayed
## 1  2020-05-30 14:51 The Longshot                    Devil’s Kind    10770
## 2  2020-06-04 13:06    Green Day          21st Century Breakdown     4466
## 3  2020-06-04 13:08    Green Day                  99 Revolutions     3585
## 4  2020-06-04 13:08 The Longshot                    Devil’s Kind   120783
## 5  2020-06-04 13:15    Green Day                     Forever Now   412346
## 6  2020-06-04 13:21    Green Day                     Brutal Love   294746
## 7  2020-06-04 13:23    Green Day                Father of All...   151362
## 8  2020-06-04 13:31    Green Day Extraordinary Girl / Letterbomb   460000
## 9  2020-06-04 13:34    Green Day                Church on Sunday   198466
## 10 2020-06-04 13:39    Green Day   Oh Love - Otis Big Guitar Mix   292177
```

Playback activity per week in hours:

```
# Changing into Indian Standard Time Zone, creating Date and Time columns

data <- streaming_history %>%
  as_tibble() %>%
  mutate_at("endTime",ymd_hm) %>%
  mutate(endTime = endTime + minutes(330)) %>%
  mutate(date = floor_date(endTime,"day") %>%
  as_date, seconds = msPlayed/1000, minutes = seconds/60)
```

```
# Playback activity per week and hours

streaming_hrs <- data %>%
  group_by(date = floor_date(date,"week")) %>%
  summarise(hours = sum(minutes/60)) %>%
  arrange(date) %>%
  ggplot(aes(x = date, y= hours)) + geom_col(aes(fill = hours)) +
  scale_x_date(date_labels= "%b %y",date_breaks  = "1 month")+
  scale_fill_gradient(low = 'lightblue1',high = 'blue') +
  labs(x = "Date", y = "Hours of music played") +
  ggtitle("Time I've listened to music on Spotify", "Playback activity per week")
```

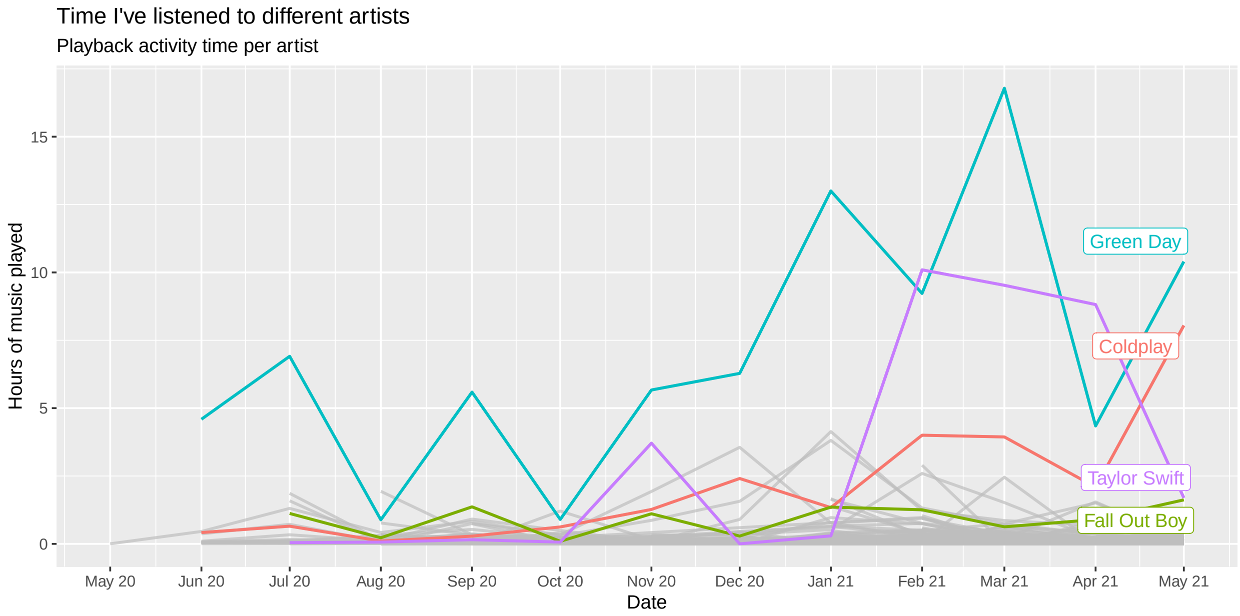
```
streaming_hrs
```



Playback Activity per artist in hours:

```r
# Playback activity time per artist

artist_hrs <- data %>%
  group_by(artistName, date = floor_date(date,"month")) %>%
  summarise(hours = sum(minutes)/60) %>%
  ggplot(aes(x = date, y = hours, group = artistName)) +
  labs(x = "Date", y = "Hours of music played") +
  scale_x_date(date_labels = "%b %y",date_breaks  = "1 month")+
  ggtitle("Time I've listened to different artists", "Playback activity time per artist") +
  geom_line(aes(date,hours,color = artistName),lwd = 0.8) +
  gghighlight(artistName == "Fall Out Boy" || artistName == "Coldplay" || artistName == "Taylor Swift"
```

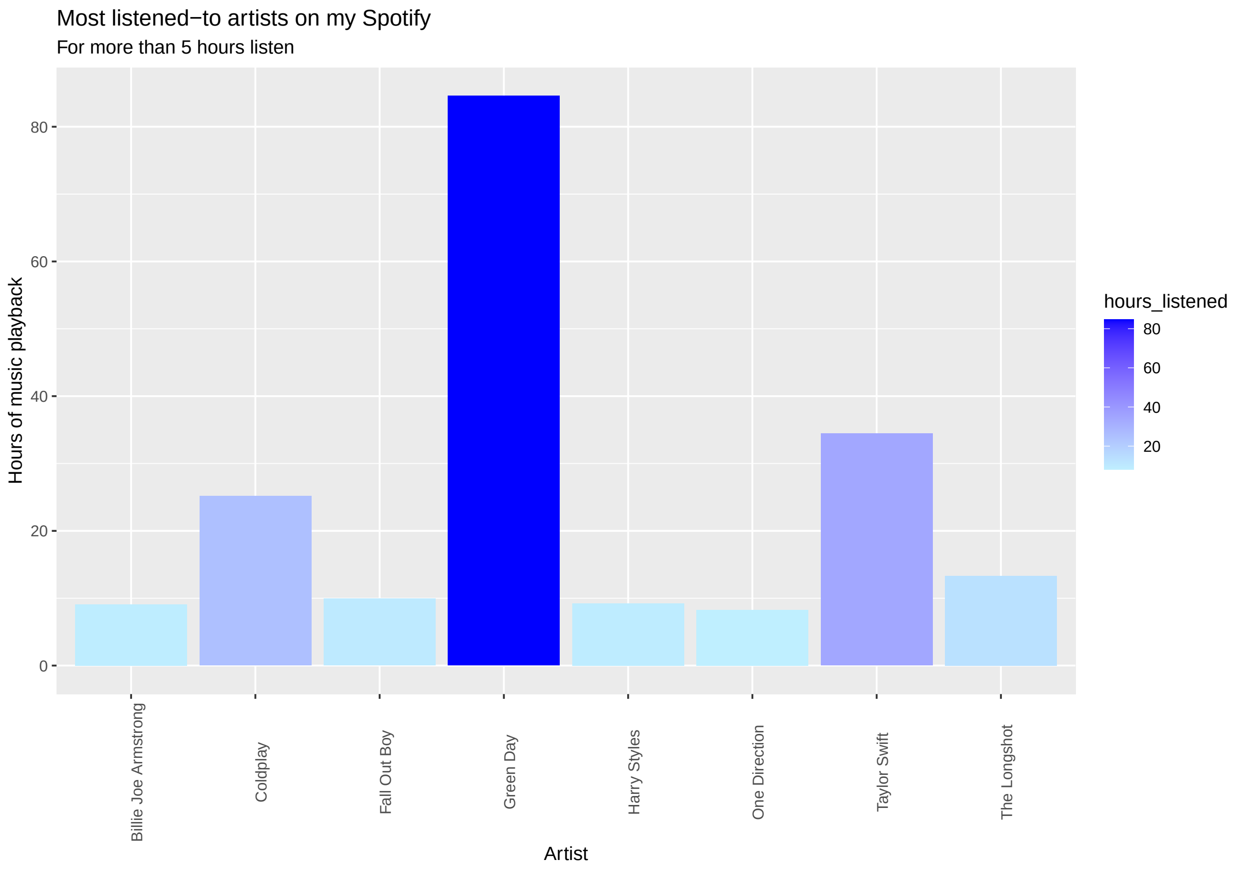
```r
artist_hrs
```



Most listened-to Artists (>5 hours)

```r
# Playback activity of most listened-to artists

most_listened <- data %>%
  group_by(artistName) %>%
  summarize(hours_listened = sum(minutes)/60) %>%
  filter(hours_listened >= 5) %>%
  ggplot(aes(x = artistName, y = hours_listened)) +
  geom_col(aes(fill = hours_listened)) +
  scale_fill_gradient(low = "lightblue1", high = "blue") +
  labs(x = "Artist", y = "Hours of music playback") +
  ggtitle("Most listened-to artists on my Spotify", "For more than 5 hours listen") +
  theme(axis.text.x = element_text(angle = 90))
```

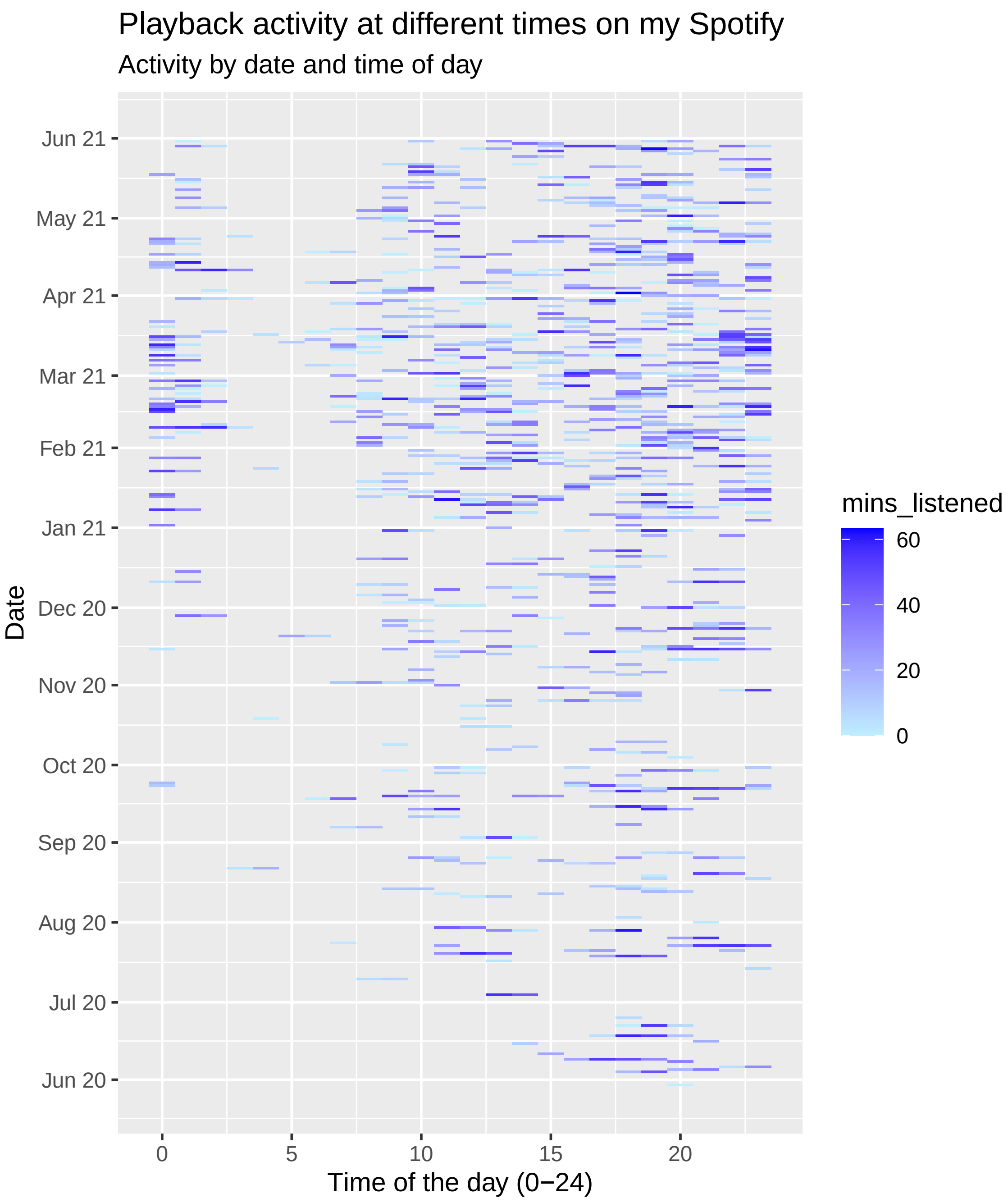
```
most_listened
```



Playback Activity by time of day with dates

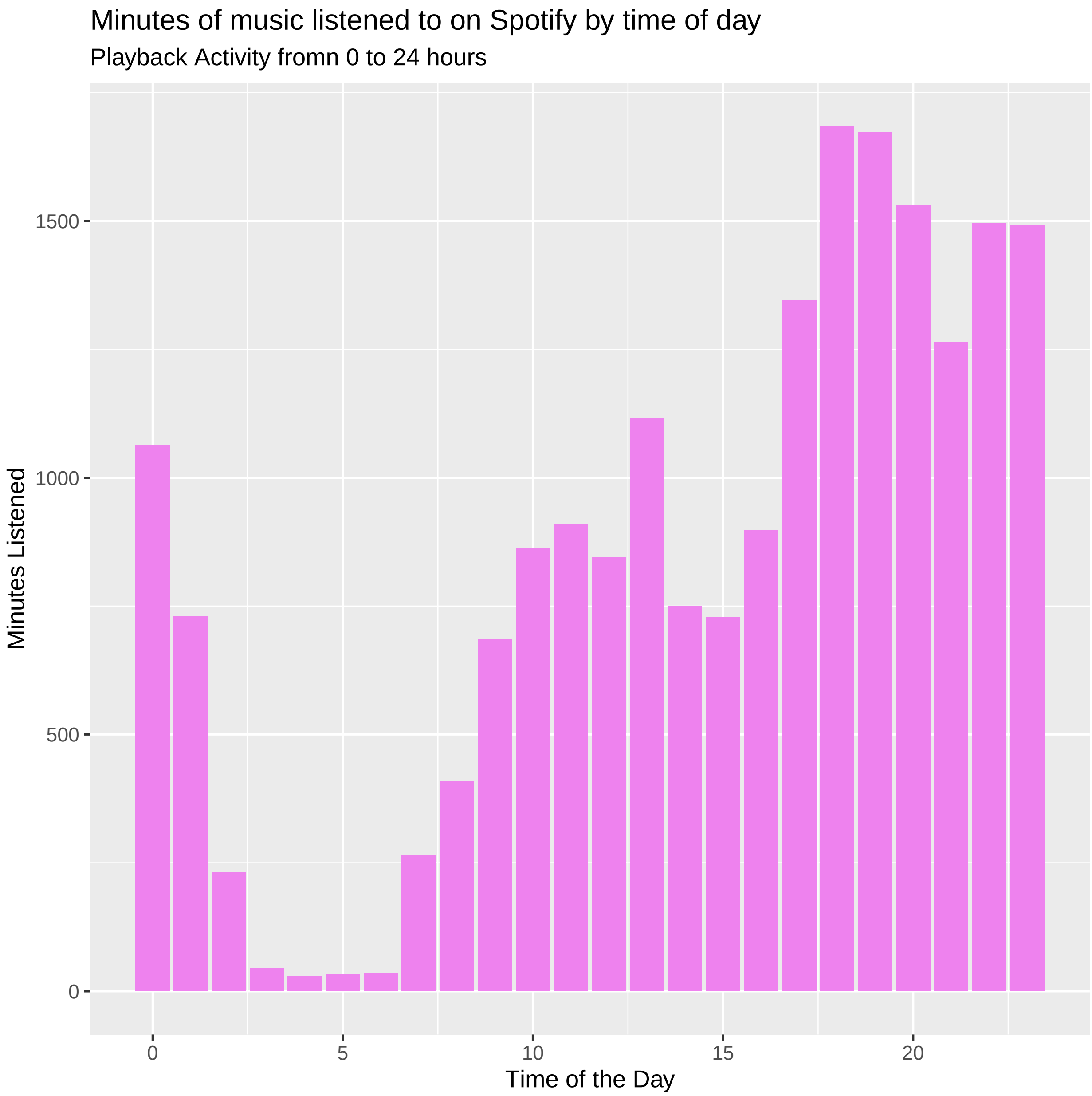
```
# Activity by date and time of day

times <- data %>%
  group_by(date, hour = hour(endTime))  %>%
  summarize(mins_listened = sum(minutes)) %>%
  ggplot(aes(x = hour, y = date, fill = mins_listened)) +
  geom_tile() +
  labs(x = "Time of the day (0-24)", y = "Date") +
  scale_y_date(date_labels="%b %y",date_breaks  ="1 month") +
  ggtitle("Playback activity at different times on my Spotify", "Activity by date and time of day") +
  scale_fill_gradient(low = "lightblue1", high = "blue")
```

```
times
```



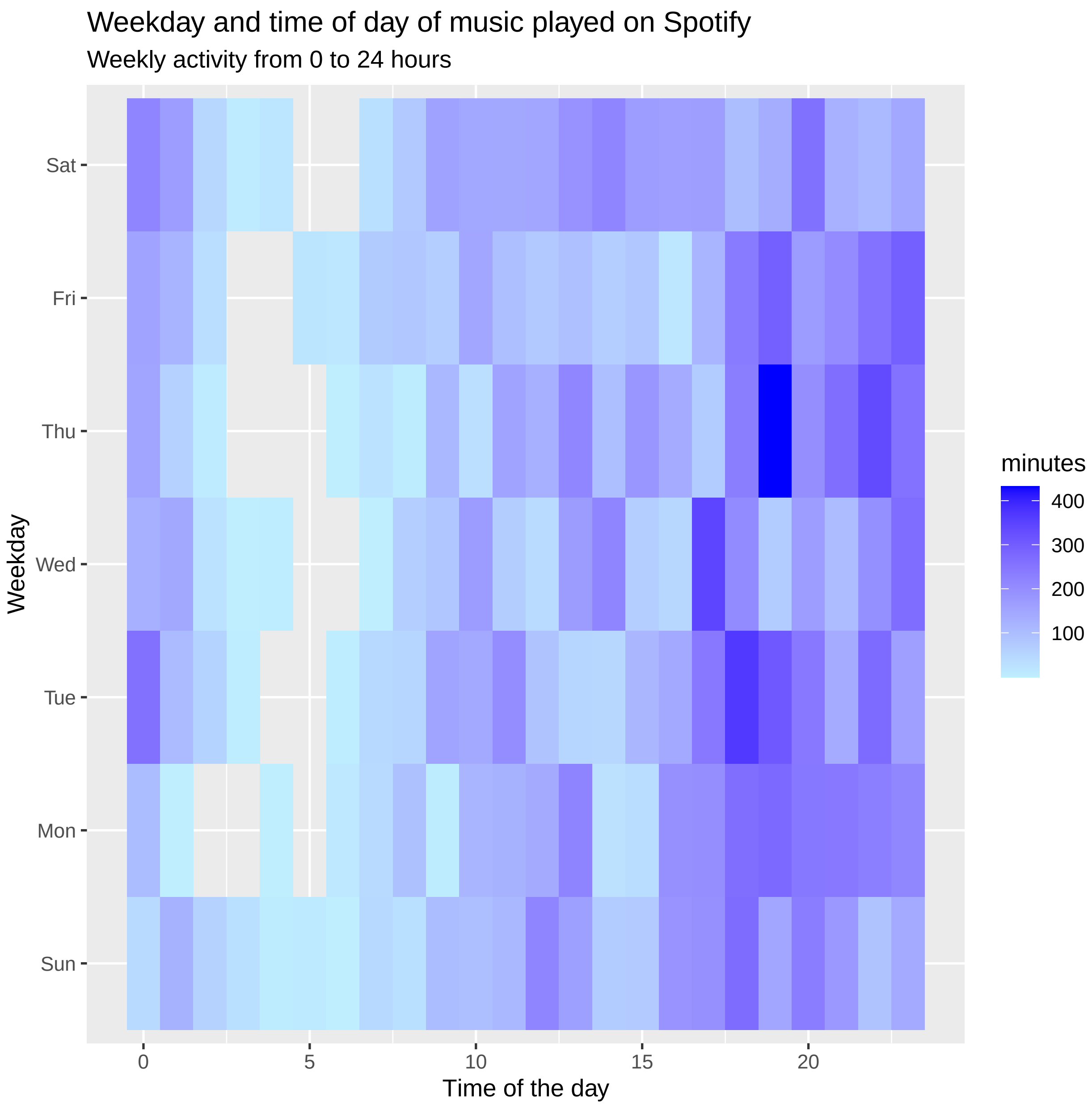
Playback activity by time of day

```
# Playback Activity from 0 to 24 hours

day_times <- data %>%
  group_by(date, hour = hour(endTime), weekday = wday(date, label = TRUE)) %>%
  summarise(mins_listened = sum(minutes))

day_times %>%
  ggplot(aes(x = hour, y = mins_listened, group = date)) +
  geom_col(fill = "violet") +
  labs(x = "Time of the Day", y = "Minutes Listened") +
  ggtitle("Minutes of music listened to on Spotify by time of day", "Playback Activity fromn 0 to 24 hou
```



Playback activity by time of day and day of week

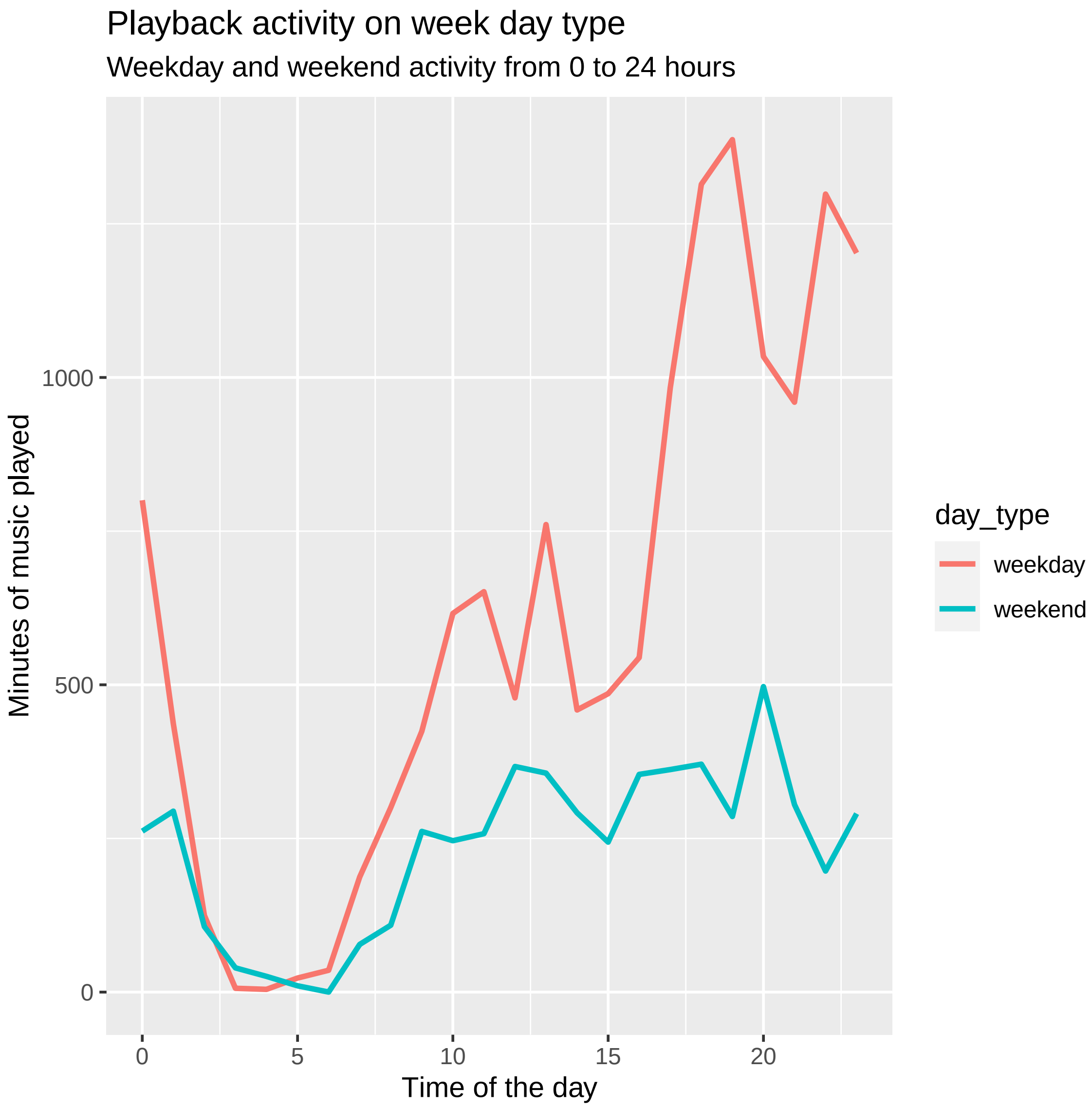
```
# Weekly playback activity from 0 to 24 hours

day_times %>%
  group_by(weekday, hour) %>%
  summarize(minutes = sum(mins_listened)) %>%
  ggplot(aes(x = hour, weekday, fill = minutes)) +
  geom_tile() +
  scale_fill_gradient(low = "lightblue1", high = "blue") +
  labs(x = "Time of the day", y = "Weekday") +
  ggtitle("Weekday and time of day of music played on Spotify", "Weekly activity from 0 to 24 hours")
```



Playback activity on weekdays and weekends

```r
# Weekday and weekend activity from 0 to 24 hours
day_type <- day_times %>%
  mutate(day_type = if_else(weekday %in% c("Sat", "Sun"), "weekend", "weekday")) %>%
  group_by(day_type, hour) %>%
  summarize(minutes = sum(mins_listened)) %>%
  ggplot(aes(x = hour, y = minutes, color = day_type)) +
  geom_line(size = 1) +
  labs(x = "Time of the day", y = "Minutes of music played") +
  ggtitle("Playback activity on week day type", "Weekday and weekend activity from 0 to 24 hours")
```

```r
day_type
```


Playback activity on week day type
Weekday and weekend activity from 0 to 24 hours
Minutes of music played
Time of the day
day_type
weekday
weekend