Statistics in Personal Spotify Data

Creating Your Own Spotify Wrapped

Emma Johnson

University of Minnesota-Duluth

October 15, 2024

Outline

Image Credit: Spotify (2024a)

Available Personal Info from Apps

Free for Download:

- ▶ Facebook
- ▶ X (Twitter)
- ▶ Instagram
- ▶ TikTok
- ▶ LinkedIn
- ▶ Spotify***

Image Credit: The Daily Guardian (2024)

Downloading Spotify Data

▸ Link for download instructions

1. Download by going to web version of Spotify > Account > Security & Privacy > Account Privacy > Download Your Data > Select **Account Data*** > Request Download
2. Follow confirmation steps
3. Receive data in about 5 business days

*Note: Extended Streaming History also available for a longer wait time

Spotify Wrapped Background Info

- ▶ Summarizes the past year of streaming
- ▶ Top Songs, Top Artists, Total Minutes Listened, Monthly Activity

Image Credit: Lions Digest (2024)
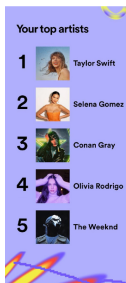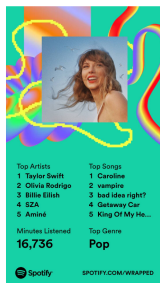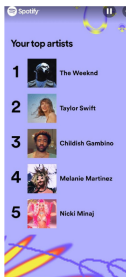
Analysis in R

Image Credit: R-bloggers (2020)

Tidyverse!!!

```
## Streaming History (from 9/12/2023 - 9/11/2024)

hist.all <- bind_rows(hist1, hist2, hist3)
str(hist.all)
```

```
## 'data.frame':    24985 obs. of  4 variables:
##  $ endTime   : chr  "2023-04-29 20:47" "2023-09-12 14:20" "2023-09-12 14:21" "2023-09-12 14:23" ...
##  $ artistName: chr  "Nirvana" "Don Toliver" "Famous Dex" "070 Shake" ...
##  $ trackName : chr  "Breed" "Way Bigger" "JAPAN" "Web" ...
##  $ msPlayed  : int  111124 196011 2716 126151 401572 143818 211582 140800 80404 80466 ...
```

Top Songs

```
## Total Time Spent Listening

# Minutes:

hist.sort.min <- hist.all %>%
  group_by(trackName, artistName) %>%
  summarise(min = sum(msPlayed)/60000, .groups = 'drop') %>%
  arrange((desc(min)))

hist.sort.min %>% print(n = 50)
```

```
## # A tibble: 6,189 × 3
##    trackName                              artistName     min
##    <chr>                                  <chr>        <dbl>
##  1 Everlong                               Foo Fight…   168.
##  2 Cherry Waves                           Deftones     160.
##  3 EXCALIBUR                              DUCKBOY      151.
##  4 GHOST!                                 Kid Cudi     141.
##  5 Whatsername                            Green Day    138.
##  6 Innerbloom                             RÜFÜS DU …   135.
##  7 Afraid To Feel                         LF SYSTEM    133.
##  8 Alive                                  Pearl Jam    131.
##  9 The Color Violet                       Tory Lanez   128.
## 10 Crumbled                               ThxSoMch     124.
## 11 You Get Me So High                     The Neigh…   120.
## 12 Not Even Ghosts Are This Empty         $uicidebo…   120.
## 13 Creep - 2017 Remaster                  Stone Tem…   119.
## 14 You and I                              d4vd         117.
## 15 Heart Racing                           Kanii        117.
```

Total Time Spent Listening

```r
# Total minutes spent listening to music:
sum(hist.sort.min$min)
```

```
## [1] 47525.81
```

```r
# hours:
sum(hist.sort.hr$hr)
```

```
## [1] 792.0968
```

```r
# total number of(consecutive) days spent listening:
(sum(hist.sort.hr$hr))/24
```

```
## [1] 33.00403
```

$\approx$ Nonstop listening
from now until
November 17

Top Artists

```
## Top artists

hist.sort.art <- hist.sort.min %>%
  group_by(artistName) %>%
  summarise(minutes = sum(min), .groups = 'drop') %>%
  arrange((desc(minutes)))

hist.sort.art %>%
  print(n = 50)
```

```
## # A tibble: 2,251 × 2
##    artistName          minutes
##    <chr>                 <dbl>
##  1 $uicideboy$          1639.
##  2 Nirvana              1035.
##  3 Kanye West            871.
##  4 Mac Miller            809.
##  5 Tame Impala           806.
##  6 Green Day             793.
##  7 Kid Cudi              751.
##  8 Tory Lanez            686.
##  9 Tyler, The Creator    625.
## 10 A$AP Rocky            608.
## 11 Travis Scott          589.
## 12 Red Hot Chili Peppers 529.
## 13 Frank Ocean           523.
## 14 Riovaz                523.
## 15 Lil Peep              490.
## 16 Pearl Jam             477.
## 17 d4vd                  469.
## 18 Foo Fighters          463.
## 19 ThxSoMch              456.
## 20 Kali Uchis            433.
```

Top Artists

```
# plotting top 10 artists minutes spent listening
ggplot(data = hist.sort.art[1:10,], aes(artistName, minutes)) +
  geom_point(size = 2.5) +
  labs(x = "Artist", y = "Minutes", title = "Time Spent Listening to Top 10 Artists of 2024") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```


Time Spent Listening to Top 10 Artists of 2024
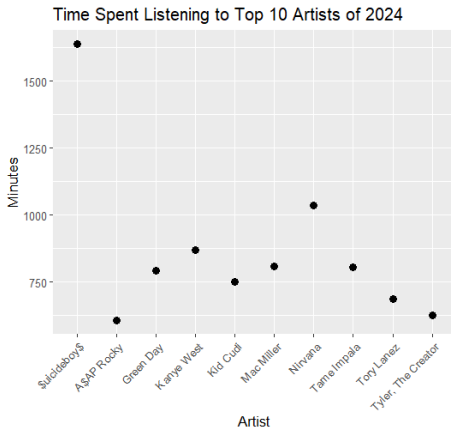
Monthly Breakdown

Creating a data frame based on minutes listened per month

```
## Monthly Breakdown

sep23.hist <- hist.all %>%
  filter(grepl("2023-09", endTime)) %>%
  group_by(trackName, artistName) %>%
  summarise(min = sum(msPlayed)/60000, .groups = 'drop') %>%
  arrange((desc(min))) # starts midway through the month (9/12/23)
oct23.hist <- hist.all %>%
  filter(grepl("2023-10", endTime)) %>%
  group_by(trackName, artistName) %>%
  summarise(min = sum(msPlayed)/60000, .groups = 'drop') %>%
  arrange((desc(min)))
nov23.hist <- hist.all %>%
  filter(grepl("2023-11", endTime)) %>%
  group_by(trackName, artistName) %>%
  summarise(min = sum(msPlayed)/60000, .groups = 'drop') %>%
  arrange((desc(min)))
```

```
# Creating data frame for monthly listening

# minutes
month.min <- c(
  sep23.hist$min,
  oct23.hist$min,
  nov23.hist$min,
  dec23.hist$min,
  jan24.hist$min,
  feb24.hist$min,
  mar24.hist$min,
  apr24.hist$min,
  may24.hist$min,
  jun24.hist$min,
  jul24.hist$min,
  aug24.hist$min,
  sep24.hist$min)

# month
month <- as.factor(c(rep("sep23", length(sep23.hist$min)),
                     rep("oct23", length(oct23.hist$min)),
                     rep("nov23", length(nov23.hist$min)),
                     rep("dec23", length(dec23.hist$min)),
                     rep("jan24", length(jan24.hist$min)),
                     rep("feb24", length(feb24.hist$min)),
                     rep("mar24", length(mar24.hist$min)),
                     rep("apr24", length(apr24.hist$min)),
                     rep("may24", length(may24.hist$min)),
                     rep("jun24", length(jun24.hist$min)),
                     rep("jul24", length(jul24.hist$min)),
                     rep("aug24", length(aug24.hist$min)),
                     rep("sep24", length(sep24.hist$min))))

# data frame for monthly listening:
year <- data.frame(month.min, month)
```
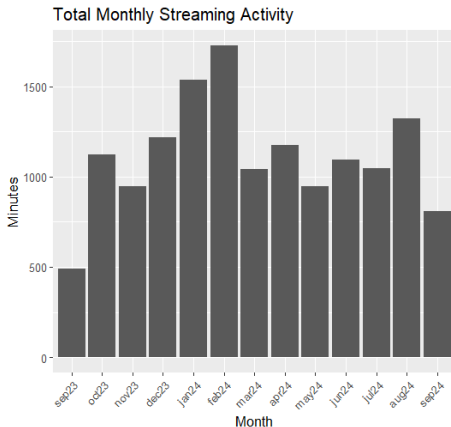
Monthly Breakdown

```r
# Plot of Total Monthly Streaming Activity
ggplot(data = year, aes(x = month)) +
  geom_bar() +
  labs(x = "Month", y = "Minutes", title = "Total Monthly Streaming Activity") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```


Total Monthly Streaming Activity

Testing for significant differences across months

Single Factor ANOVA

▶ H_0 : Each monthly average will be equal to the expected minutes listened per month.

▶ H_a : At least one monthly average will be different.

```
anova(lm(month.min ~ month, data = year))
```

```
## Analysis of Variance Table
##
## Response: month.min
##              Df Sum Sq Mean Sq F value    Pr(>F)
## month        12   2666 222.141  11.291 < 2.2e-16 ***
## Residuals 14464 284577  19.675
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Significant
```

Which months are significantly different

▶ Family-wise significance level = 0.05

▶ Z-score comparisons

```
# Finding which months are significantly different than the total monthly average:
results
```

```
##      Group_Mean   Difference    Z_score  Significant    Month
## 1     2699.635  -1260.69476  4.53091435          Yes    oct23
## 2     2688.358  -1271.97225  4.57144542          Yes    nov23
## 3     4646.827    686.49694  2.46725766           No    dec23
## 4     5694.127   1733.79774  6.23123793          Yes    jan24
## 5     5471.446   1511.11615  5.43092432          Yes    feb24
## 6     4036.005     75.67574  0.27197724           No    mar24
## 7     4123.457    163.12712  0.58627594           No    apr24
## 8     3233.713   -726.61720  2.61144916           No    may24
## 9     3974.966     14.63634  0.05260273           No    jun24
## 10    3211.513   -748.81698  2.69123478           No    jul24
## 11    4111.102    150.77265  0.54187421           No    aug24
## 12    3632.808   -327.52148  1.17710632           No  sep2324
```

Further Suggestions

Your latest Wrapped data, if Wrapped is a feature in your market. This includes (where available):

Wrapped data

1. Number of unique artists listened to for the year.
2. Top artists for the year.
3. Milliseconds spent listening to the number 1 artists.
4. Top percentage fan for the top artist.
5. Number of genres listened to for the year.
6. Top genres for the year.
7. Top podcasts for the year.
8. Milliseconds spent listening to the top podcast.
9. Top percentage fan for the top podcast.
10. Total milliseconds spent listening to podcasts for the year.
11. Top tracks for the year.
12. Amount of plays for the top track of the year.
13. First date the top track was played for the year.
14. Total distinct tracks played for the year.
15. Total milliseconds listened on Spotify for the year.
16. Day with the most time spent listening for the year.
17. Minutes of content listened on the top listening day for the year.
18. Top percentage of worldwide listeners for the year.
19. One-off Wrapped stories for that year.

Image Credit: Spotify (2024b)

Bibliography

Lions Digest (2024). Spotify wrapped grips users for another year.

R-bloggers (2020). 10 must-know tidyverse features.

Spotify (2024a). Logo and brand assets.

Spotify (2024b). Understanding my data.

The Daily Guardian (2024). Top social media platforms: A deep dive into today's most influential choices.