Wrangling the Billboard Top 100

Kristen Lowe and Siboney Cardoso

2024-08-07

```
library(tidyverse)
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.3.3
```

Consider the data in billboard.csv containing every song to appear on the weekly Billboard Top 100 chart since 1958, up through the middle of 2021. Each row of this data corresponds to a single song in a single week. For our purposes, the relevant columns here are:

- performer: who performed the song
- song: the title of the song
- year: year (1958 to 2021)
- week: chart week of that year (1, 2, etc)
- week_position: what position that song occupied that week on the Billboard top 100 chart.

Use your skills in data wrangling and plotting to answer the following three questions.

Part A

Make a table of the top 10 most popular songs since 1958, as measured by the total number of weeks that a song spent on the Billboard Top 100. Note that these data end in week 22 of 2021, so the most popular songs of 2021 will not have up-to-the-minute data; please send our apologies to The Weeknd.

Your table should have 10 rows and 3 columns: performer, song, and count, where count represents the number of weeks that song appeared in the Billboard Top 100. Make sure the entries are sorted in descending order of the count variable, so that the more popular songs appear at the top of the table. Give your table a short caption describing what is shown in the table.

```
# import the data set
billboard <- read_csv('data/billboard.csv', show_col_types = FALSE)
```

```
## New names:
## • `` -> `...1`
```

```
billboard <- billboard %>%
  dplyr::select(performer, song, year, week, week_position)

top_10 <- billboard %>%
  group_by(performer, song) %>%
  summarise(count = n()) %>%
  arrange(desc(count)) %>%
  head(10)
```

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

```
kable(top_10, caption = "Top 10 Most Popular Songs Since 1958, as Measured by the Total
Number of Weeks that a Song Spent on the Billboard Top 100")
```

Top 10 Most Popular Songs Since 1958, as Measured by the Total Number of Weeks that a Song Spent on the Billboard Top 100

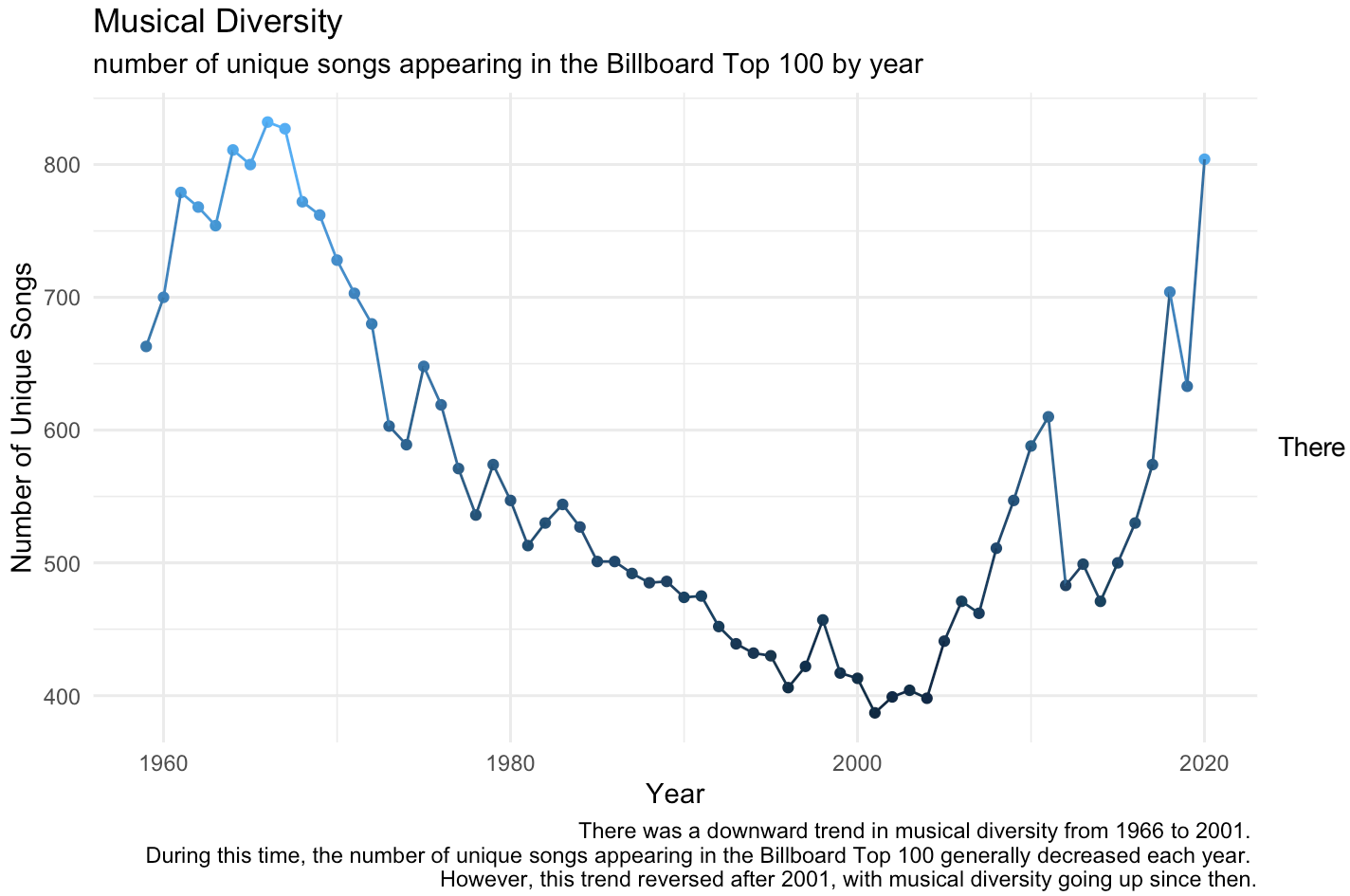
performer	song	count
Imagine Dragons	Radioactive	87
AWOLNATION	Sail	79
Jason Mraz	I’m Yours	76
The Weeknd	Blinding Lights	76
LeAnn Rimes	How Do I Live	69
LMFAO Featuring Lauren Bennett & GoonRock	Party Rock Anthem	68
OneRepublic	Counting Stars	68
Adele	Rolling In The Deep	65
Jewel	Foolish Games/You Were Meant For Me	65
Carrie Underwood	Before He Cheats	64

Part B

Is the “musical diversity” of the Billboard Top 100 changing over time? Let’s find out. We’ll measure the musical diversity of given year as the number of unique songs that appeared in the Billboard Top 100 that year. Make a line graph that plots this measure of musical diversity over the years. The x axis should show the year, while the y axis should show the number of unique songs appearing at any position on the Billboard Top 100 chart in any week that year. For this part, please filter the data set so that it excludes the years 1958 and 2021, since we do not have complete data on either of those years. Give the figure an informative caption in which you explain what is shown in the figure and comment on any interesting trends you see.

```
musical_diversity <- billboard %>%
  filter(year != 1958) %>%
  filter(year != 2021) %>%
  dplyr::select(-week, -week_position) %>%
  distinct(performer, song, year) %>%
  group_by(year) %>%
  summarise(num_unique_songs = n())

ggplot(musical_diversity, aes(x = year, y = num_unique_songs, color = num_unique_songs))
+
  geom_point() +
  geom_line() +
  theme_minimal() +
  labs(x = "Year",
       y = "Number of Unique Songs",
       title = "Musical Diversity",
       subtitle = "number of unique songs appearing in the Billboard Top 100 by year",
       caption = "There was a downward trend in musical diversity from 1966 to 2001.
       During this time, the number of unique songs appearing in the Billboard Top 100 g
enerally decreased each year.
       However, this trend reversed after 2001, with musical diversity going up since th
en.") +
  theme(legend.position="none")
```



There was a downward trend in musical diversity from 1966 to 2001. During this time, the number of unique songs appearing in the Billboard Top 100 generally decreased each year. However, this trend reversed after 2001, with

musical diversity going up since then.

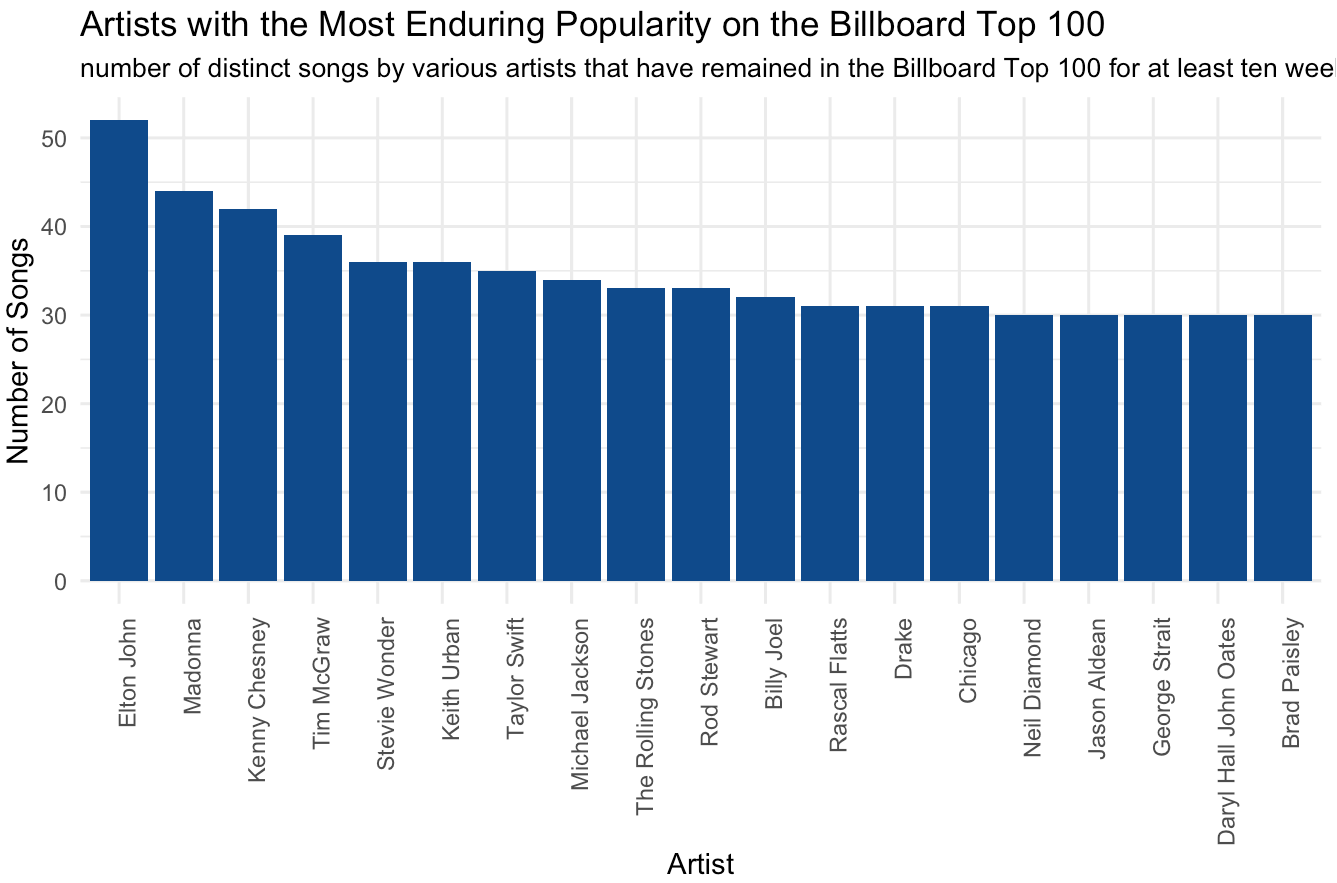
Part C

Let's define a "ten-week hit" as a single song that appeared on the Billboard Top 100 for at least ten weeks. There are 19 artists in U.S. musical history since 1958 who have had at least 30 songs that were "ten-week hits." Make a bar plot for these 19 artists, showing how many ten-week hits each one had in their musical career. Give the plot an informative caption in which you explain what is shown.

```
ten_week_hits <- billboard %>%
  group_by(performer, song) %>%
  summarise(num_weeks = n()) %>%
  filter(num_weeks >= 10) %>%
  group_by(performer) %>%
  summarise(num_popular_songs = n()) %>%
  filter(num_popular_songs >= 30) %>%
  arrange(desc(num_popular_songs))
```

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

```
ggplot(ten_week_hits, aes(x = fct_rev(fct_reorder(performer, num_popular_songs)),
                          y = num_popular_songs)) +
  geom_bar(stat = "identity", fill = "dodgerblue4") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust=1),
        plot.subtitle = element_text(size=10)) +
  labs(x = "Artist",
       y = "Number of Songs",
       title = "Artists with the Most Enduring Popularity on the Billboard Top 100",
       subtitle = "number of distinct songs by various artists that have remained in the
Billboard Top 100 for at least ten weeks",
       caption = "This bar chart shows the number of songs by artists that remained on t
he charts for at least ten
       weeks, with Elton John and Madonna leading the list, highlighting significant lon
g-term chart success.")
```



This bar chart shows the number of songs by artists that remained on the charts for at least ten weeks, with Elton John and Madonna leading the list, highlighting significant long-term chart success.