Top 100 Billboards

Imports

```
library(tidyverse)
library(tidytuesdayR)
library(scales)
library(ggplot2)
library(tidytext)
library(widyr)
library(ggraph)
library(lubridate)
theme_set(theme_light())
```

Reading and cleaning the data

```
billboard <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/

billboard <- billboard %>%
mutate(
  week = mdy(week_id),
  year =  as.integer(format(as.POSIXct(week_id, format="%m/%d/%Y"), format="%Y")),
  decade = 10 * (year %/% 10))  %>%
  filter(year < 2021)

audio_features <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/

### Merge the two datasets for later analysis
billboard_songs <- merge(billboard, audio_features)
```

Analysis

Top 10 songs of all time? (Song stayed position 1 the longest)

```
billboard %>%
  filter(week_position  == 1) %>%
  count(performer, song, sort = TRUE) %>%
  head(10)
```

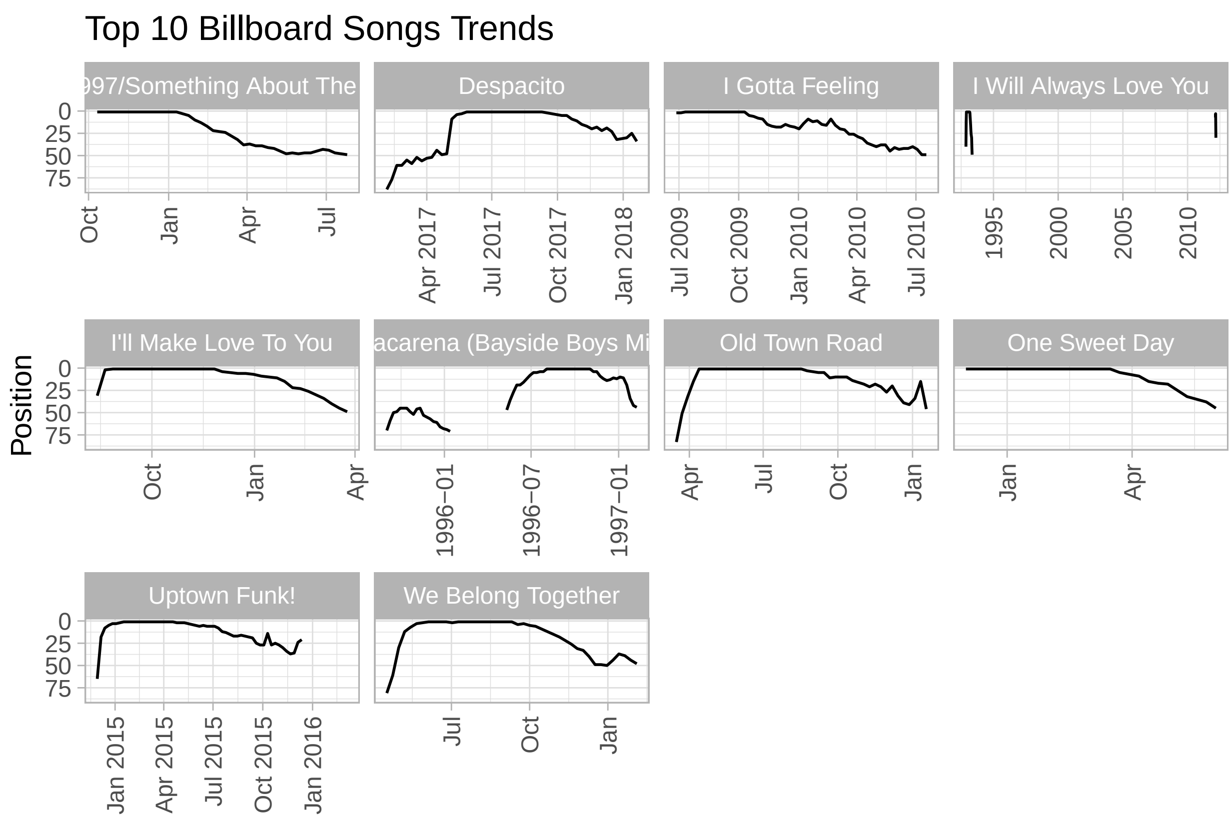
```
## # A tibble: 10 x 3
##    performer                                          song                       n
##    <chr>                                              <chr>                  <int>
##  1 Lil Nas X Featuring Billy Ray Cyrus                Old Town Road             19
```

```
##  2 Luis Fonsi & Daddy Yankee Featuring Justin Bieber Despacito            16
##  3 Mariah Carey & Boyz II Men                        One Sweet Day         16
##  4 Boyz II Men                                       I’ll Make Love To You    14
##  5 Elton John                                        Candle In The Wind 1~    14
##  6 Los Del Rio                                       Macarena (Bayside Bo~    14
##  7 Mariah Carey                                      We Belong Together       14
##  8 Mark Ronson Featuring Bruno Mars                  Uptown Funk!             14
##  9 The Black Eyed Peas                               I Gotta Feeling          14
## 10 Whitney Houston                                   I Will Always Love Y~    14
```

Trends of top 10 (Song stayed position 1 the longest) songs?

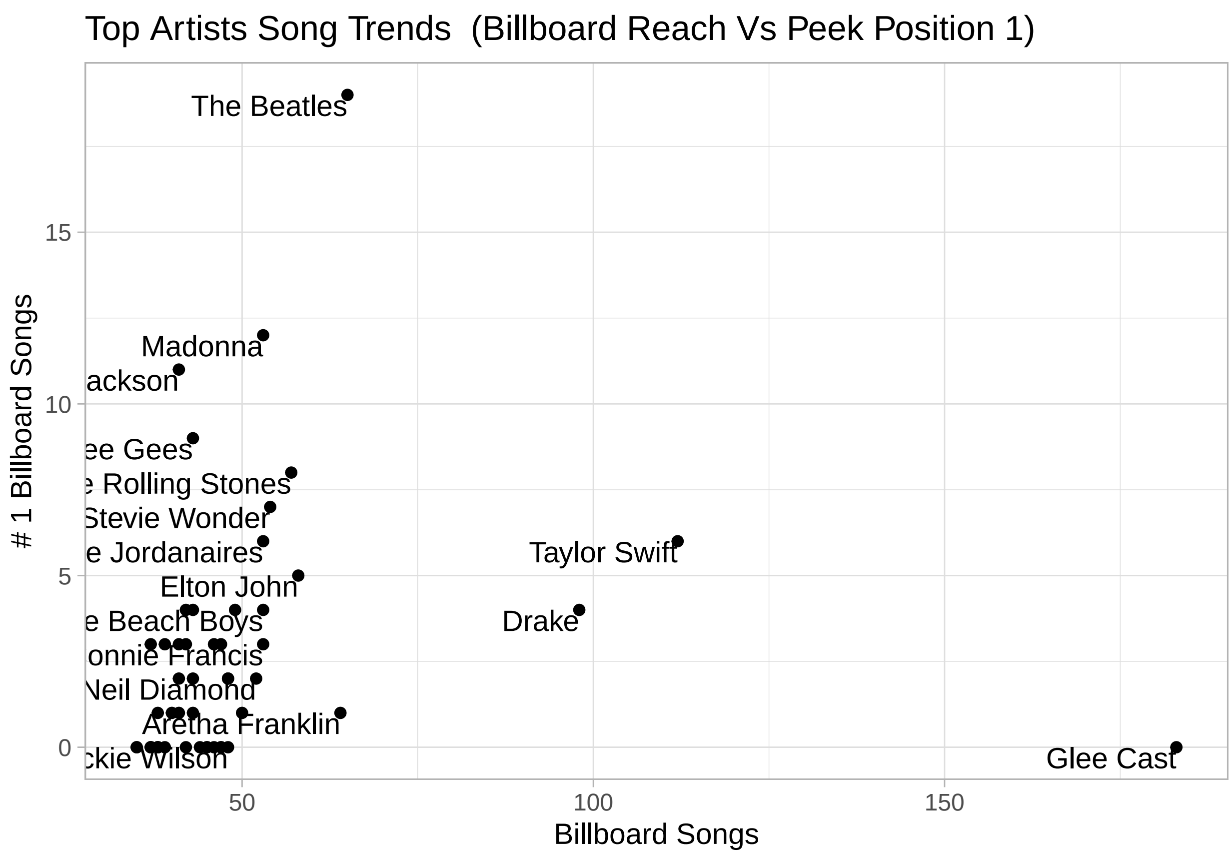
```
top_10_all_times_songs <-
  billboard %>%
  filter(week_position  == 1) %>%
  count(song_id, song, sort = TRUE) %>%
  head(10)

billboard %>%
  semi_join(top_10_all_times_songs, by = "song_id") %>%
  ggplot(aes(week, week_position, group = instance)) +
  geom_line() +
  facet_wrap(~ song, scales ="free_x")  +
  labs(title=str_to_title("Top 10 billboard songs trends"),
       x = "",
       y = "Position") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  scale_y_reverse()
```



Top 50 Billboard Performers Song Analysis (Billboard Reach vs Peek Position 1)

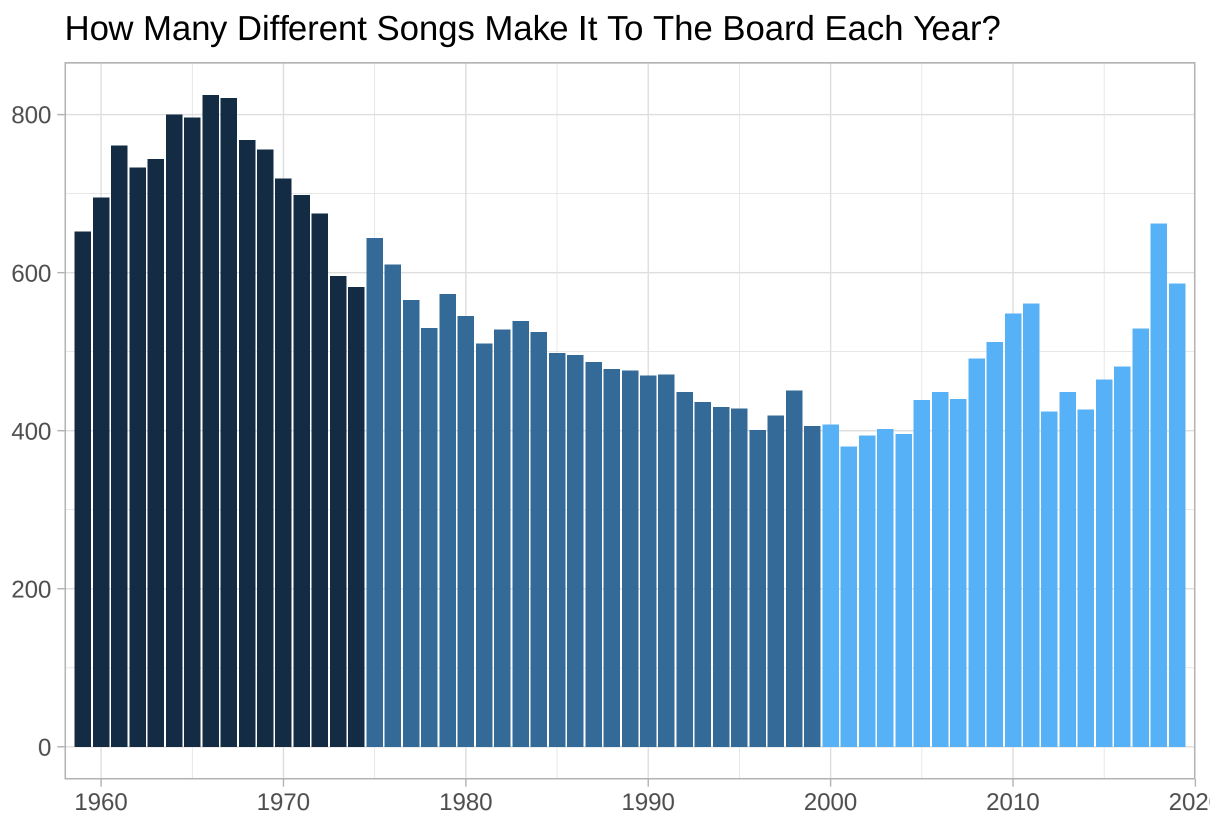
```
billboard %>%
  group_by(performer) %>%
  summarise(
    ocurrance = n(),
    distinct_songs_in_board = n_distinct(song_id),
    distinct_songs_pos_1_in_board = n_distinct(song_id[week_position == 1]),
    total_week_number_1 = sum(week_position == 1)
  )  %>%
  arrange(desc(distinct_songs_in_board)) %>%
  head(50) %>%
  ggplot(aes(distinct_songs_in_board, distinct_songs_pos_1_in_board))  +
  geom_point() +
  geom_text(aes(label = performer), check_overlap = TRUE, vjust = 1, hjust = 1) +
   labs(title=str_to_title("Top artists song trends  (Billboard Reach vs Peek Position 1)"),
       x = str_to_title("billboard songs"),
       y = str_to_title("# 1 billboard songs"))
```



Conclusion: The beetles for instance has few billboard songs but nearly half of them reached number 1 position. While on the other hand, Glee Cast have more than 150 songs but only 1 of them reached position 1 on the charts.

How many different songs make it to the board each year?

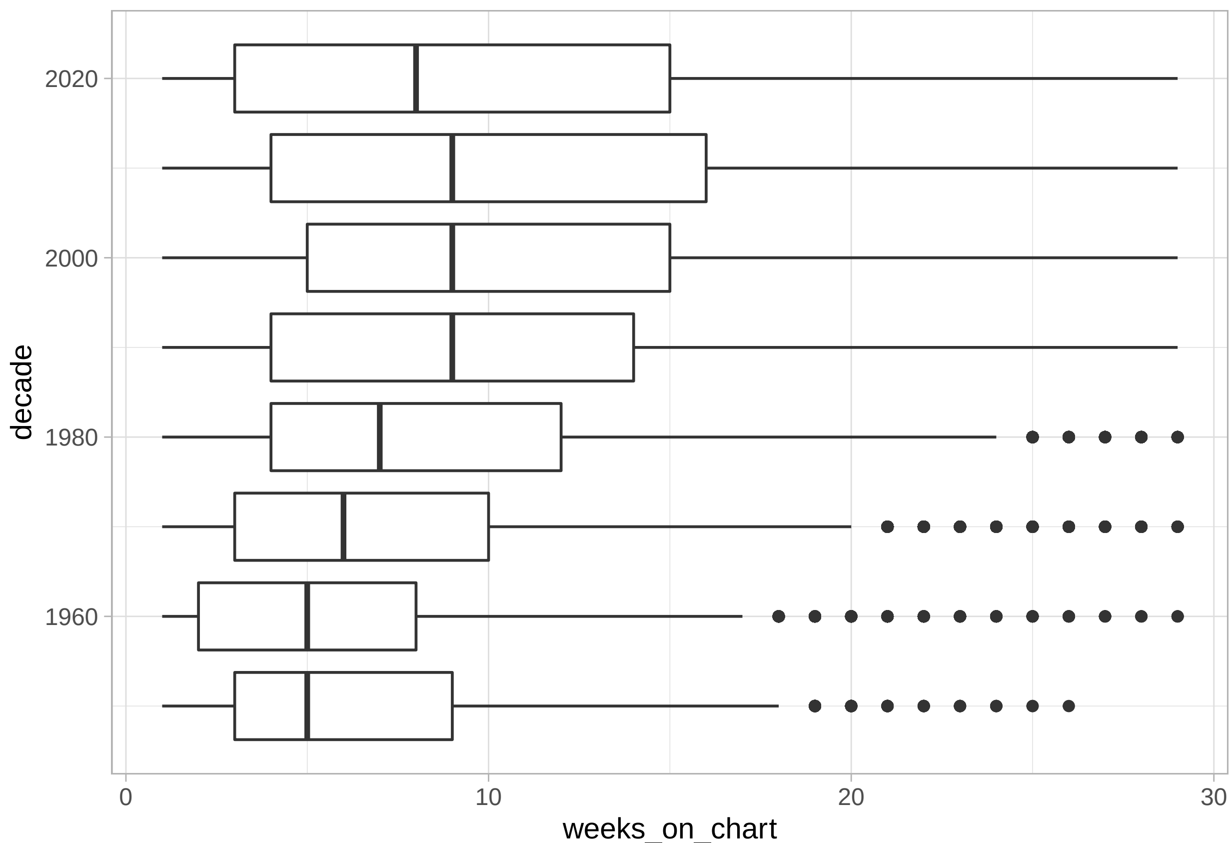
```
billboard %>%
  mutate(thirty_year_shading = 25 * (year %/% 25)) %>%
    filter(!is.na(year),
         instance == 1
         ) %>%
  distinct (year, thirty_year_shading, song_id) %>%
  count(year, thirty_year_shading) %>%
  ggplot(aes(x=year, y=n, fill = thirty_year_shading)) +
  geom_col() +
  theme(legend.position = "none") +
  labs(title=str_to_title("How many different songs make it to the board each year?"),
       x = "",
       y = "") +
 scale_x_continuous(limits = c(min(billboard$year), max(billboard$year)), expand = c(0,0), n.breaks = 10
```



Conclusion: Chart is dominated by less songs in later years.

How long a song stays on the chart through the years?

```
billboard %>%
  filter(weeks_on_chart < 30) %>%
  ggplot(aes(x=decade, y=weeks_on_chart, group = decade)) +
  geom_boxplot() +
  coord_flip()
```



Conclusion: Songs stays more on chart in later years.

```
billboard %>%
  filter(year == 1960) %>%
  group_by(year, performer, song) %>%
  summarise(max_week_on_chart = max(weeks_on_chart)) %>%
  arrange(desc(max_week_on_chart))  %>%
  head(10)
```

```
## # A tibble: 10 x 4
## # Groups:   year, performer [9]
##     year performer                       song                         max_week_on_cha~
##    <int> <chr>                           <chr>                                   <dbl>
##  1  1960 Johnny Preston                  Running Bear                               27
##  2  1960 Bobby Darin                     Mack The Knife                             26
##  3  1960 Hank Ballard And The Midnighters Finger Poppin' Time                       26
##  4  1960 Brenda Lee                      Sweet Nothin's                             24
##  5  1960 Connie Stevens                  Sixteen Reasons                            24
##  6  1960 Brenda Lee                      I'm Sorry                                  23
##  7  1960 Jim Reeves                      He'll Have To Go                           23
##  8  1960 Hank Locklin                    Please Help Me, I'm Falling                22
##  9  1960 Marty Robbins                   El Paso                                    22
## 10  1960 Marv Johnson                    You Got What It Takes                      22
```

```
billboard %>%
  filter(year == 2015) %>%
```

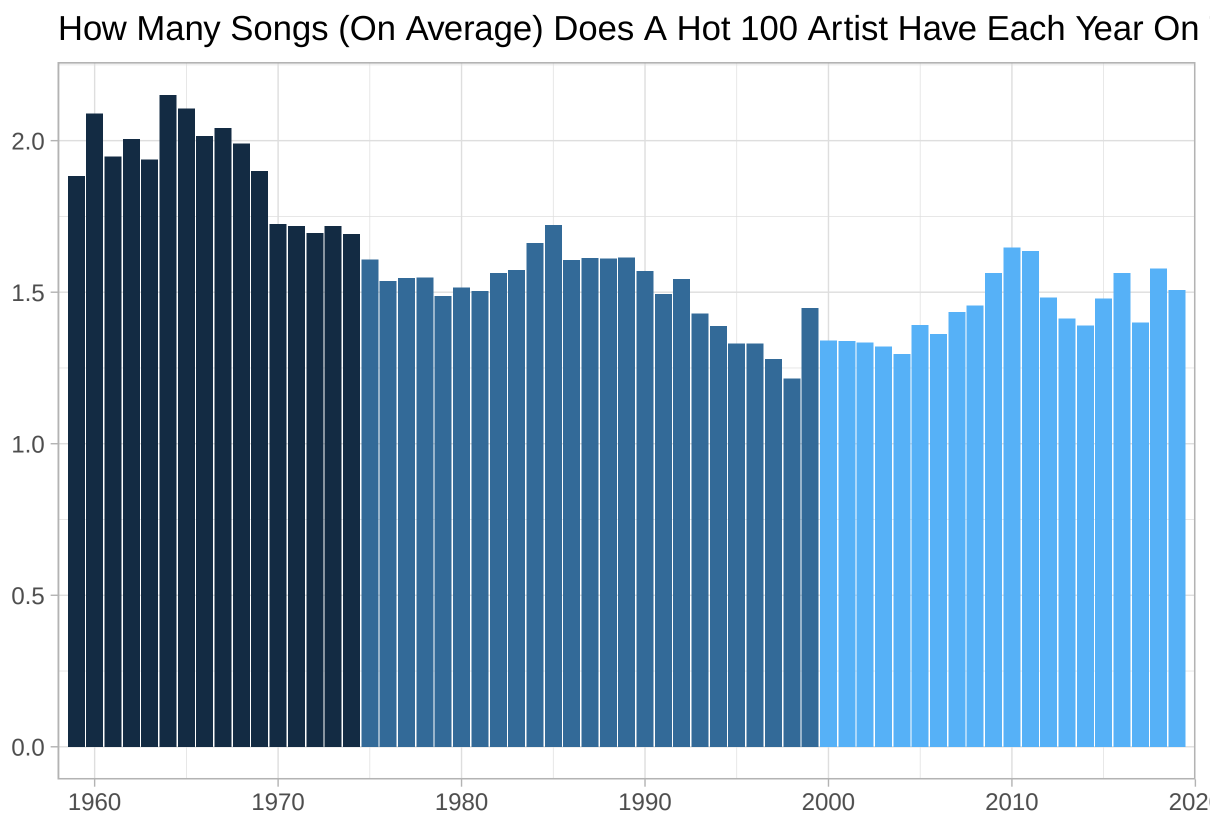
```r
  group_by(year, performer, song) %>%
  summarise(max_week_on_chart = max(weeks_on_chart)) %>%
  arrange(desc(max_week_on_chart))  %>%
  head(10)
```

```
## # A tibble: 10 x 4
## # Groups:   year, performer [10]
##     year performer                       song                             max_week_on_cha~
##    <int> <chr>                           <chr>                                       <dbl>
##  1  2015 Ed Sheeran                      Thinking Out Loud                              58
##  2  2015 Mark Ronson Featuring Bruno Mars Uptown Funk!                                  55
##  3  2015 Sam Smith                       Stay With Me                                   54
##  4  2015 WALK THE MOON                   Shut Up And Dance                              53
##  5  2015 Taylor Swift                    Shake It Off                                   50
##  6  2015 Fetty Wap                       Trap Queen                                     47
##  7  2015 Meghan Trainor                  All About That Bass                            47
##  8  2015 Sia                             Chandelier                                     46
##  9  2015 Vance Joy                       Riptide                                        44
## 10  2015 The Weeknd                      Earned It (Fifty Shades Of Grey)               43
```

Conclusion: Songs stays DOUBLE number of weeks in later days

How many songs does a hot 100 Artist have each year?

```r
billboard %>%
  distinct(year, performer, song_id) %>%
  mutate(thirty_year_shading = 25 * (year %/% 25)) %>%
    filter(!is.na(year)) %>%
  count(year, thirty_year_shading, performer) %>%
  group_by(year, thirty_year_shading) %>%
  summarise(
    average_songs_per_artist = mean(n)
  ) %>%
  ggplot(aes(x=year, y=average_songs_per_artist, fill = thirty_year_shading)) +
  geom_col() +
  theme(legend.position = "none") +
  labs(title=str_to_title("How many songs (On average) does a hot 100 Artist have each year on the char
       x = "",
       y = "") +
  scale_x_continuous(limits = c(min(billboard$year), max(billboard$year)), expand = c(0,0), n.breaks = 
```



Conclusion: Artist had more variety of songs to dominate the chart in earlier days

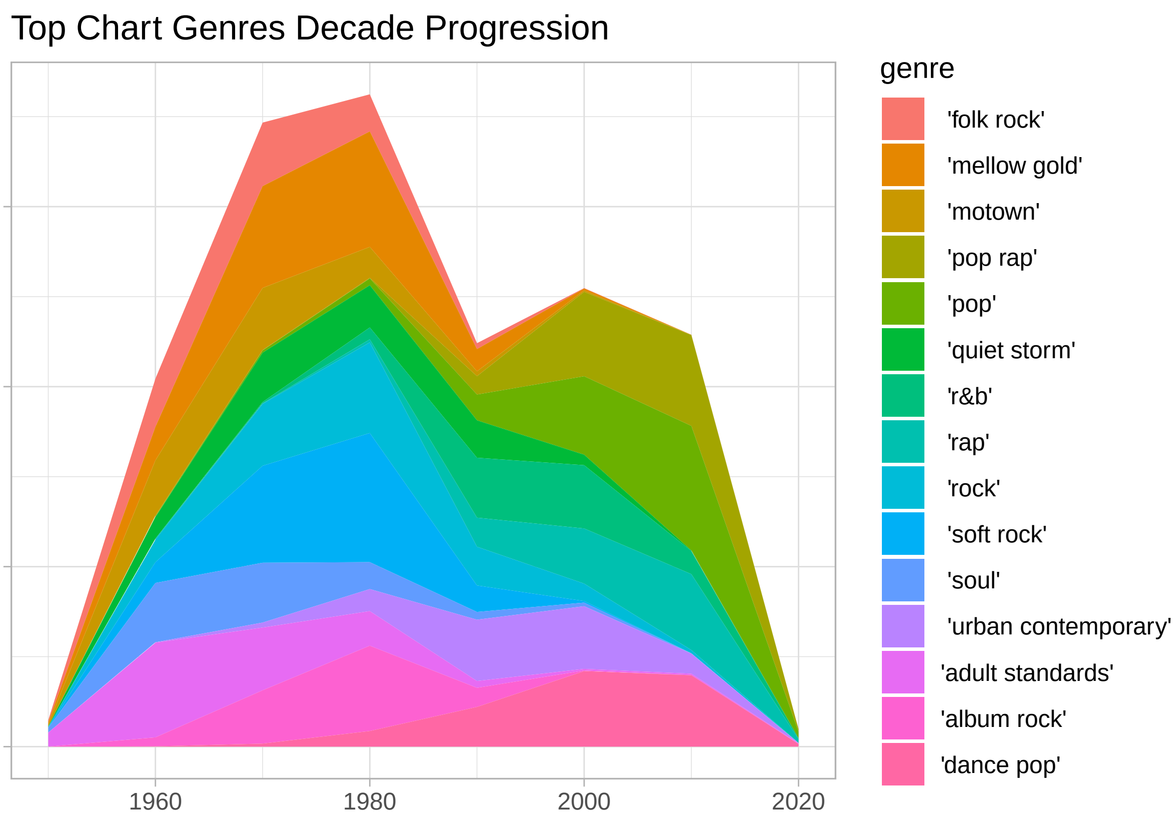
Now we'll tackle the audio_features dataset to attain more insights

With the added info, we'll try to figure out the reasons of some music trends each year.

Top genres trends

```
billboard_songs %>%
  # Remove empty and NULL genres
  filter(!is.na(spotify_genre),  spotify_genre != "[]") %>%
  # Remove brackets
  mutate(spotify_genre = substr(spotify_genre,2,nchar(spotify_genre)-1)) %>%
  # Seperate genres
  unnest_tokens(output="genre", input=spotify_genre, token = 'regex', pattern=",")  %>%
  # Get top 10 genres
  mutate(genre = fct_lump(genre, 15))   %>%
  # Remove others
  filter(genre != "Other")  %>%
  # We want only to trace the top 20
  count(decade, genre, sort = TRUE) %>%
  ggplot(aes(x=decade, y=n, fill = genre)) +
  geom_area() +
  # facet_wrap(~genre)  +
```

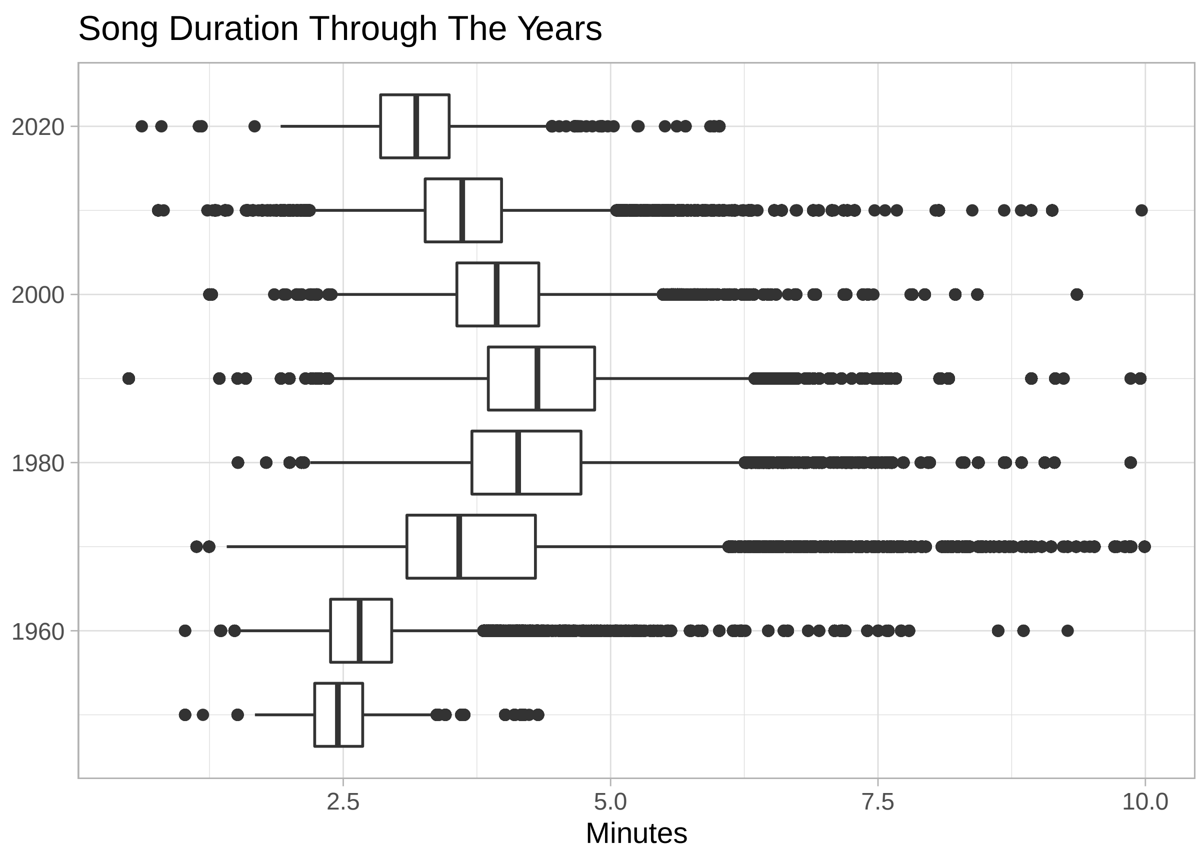
```
  theme(axis.text.y = element_blank()) +
  labs(title=str_to_title("Top chart genres decade progression"),
     x = "",
     y = "")
```



The progression of spotify's top (10) genres throughout the years.

Duration throughout the years (Mean)

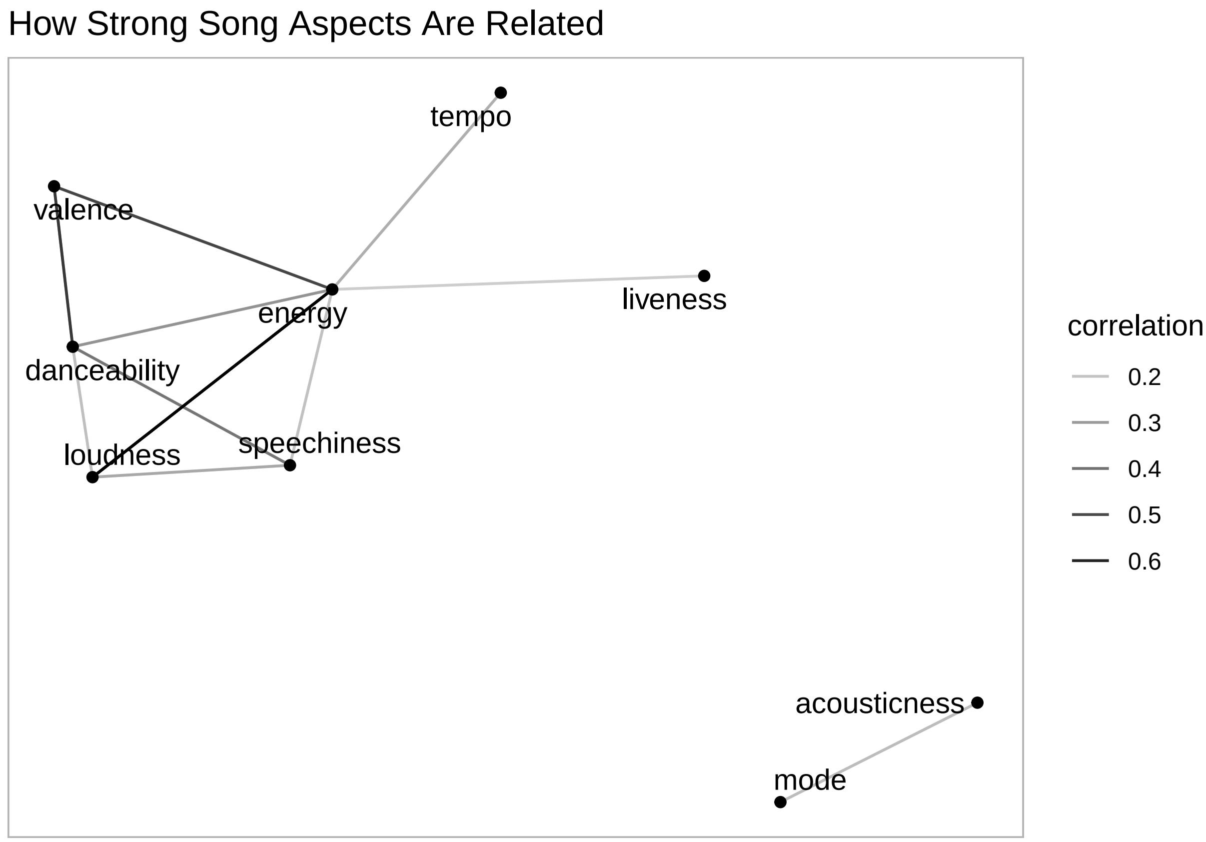
```
billboard_songs %>%
  filter(!is.na(spotify_track_duration_ms)) %>%
  mutate(duration_minutes = ((spotify_track_duration_ms/1000)/60)) %>%
  filter(duration_minutes<= 10) %>%
  ggplot(aes(x=decade, y=duration_minutes, group = decade)) +
  geom_boxplot() +
  coord_flip() +
  labs(title=str_to_title("Song duration through the years"),
     x = "",
     y = str_to_title("Minutes"))
```



Conclusion: Songs reached peak duration in the early 90s then started to decline going through the 2000s.

Song aspects coorelation

```
audio_features %>%
    pivot_longer(danceability:tempo, names_to = "metric", values_to = "value") %>%
    filter(!is.na(value)) %>%
    pairwise_cor(metric, song_id, value, sort = TRUE) %>%
    filter(correlation >= 0.1)  %>%
    igraph::graph_from_data_frame()  %>%
    ggraph(layout = "fr") +
    geom_edge_link(aes(alpha =correlation)) +
    geom_node_point() +
    geom_node_text(aes(label=name),repel = TRUE)  +
    labs(title=str_to_title("How strong song aspects are related"),
     x = "",
     y = "")
```



Conclusion: Energy dectitates other aspects of the song. Strong correlation between energy and loudness and valence.