Temporal Trends in Popular Music Genres: A Data-Driven Analysis

Elizabeth Hall - Data Analysis Project

Elizabeth Hall

2024-02-23

# 1. Part 1

## 1.1 The Data

The data I have so far is a list of the Billboard Hot 100s from 1958 to 2023. (Which can be found here <https://github.com/HipsterVizNinja/random-data/blob/main/Music/hot-100/README.md>)

The data can also be found in the “raw-data” folder in this GitHub repo. It is titled “Hot\_100s.csv”.

This dataset includes observations for chart position, chart date, song name, performer name, song id, instance, time on chart, consecutive weeks, previous week, peak position, worst position, chart debut date, and chart url. The data was collected from the Billboards Hot 100

I will be adding the genre category to this dataset myself by cross-referencing the artists with the Spotify API (found here -<https://developer.spotify.com/documentation/web-api>) to obtain their musical genre. Though I am not entirely sure how the data was collected, I know that the Billboard Hot 100s keeps weekly records on their website, so this is where the data came from.

Here is some code to display a little info on the data:

chart\_position chart\_date song performer  
1 84 1990-05-05 "B" Girls Young And Restless  
2 78 1990-05-12 "B" Girls Young And Restless  
3 68 1990-05-19 "B" Girls Young And Restless  
4 60 1990-05-26 "B" Girls Young And Restless  
5 58 1990-06-02 "B" Girls Young And Restless  
6 55 1990-06-09 "B" Girls Young And Restless  
 song\_id instance time\_on\_chart consecutive\_weeks  
1 "B" GirlsYoung And Restless 1 1 NA  
2 "B" GirlsYoung And Restless 1 2 1  
3 "B" GirlsYoung And Restless 1 3 2  
4 "B" GirlsYoung And Restless 1 4 3  
5 "B" GirlsYoung And Restless 1 5 4  
6 "B" GirlsYoung And Restless 1 6 5  
 previous\_week peak\_position worst\_position chart\_debut  
1 NA 84 84 1990-05-05  
2 84 78 84 1990-05-05  
3 78 68 84 1990-05-05  
4 68 60 84 1990-05-05  
5 60 58 84 1990-05-05  
6 58 55 84 1990-05-05  
 chart\_url  
1 https://www.billboard.com/charts/hot-100/1990-05-05  
2 https://www.billboard.com/charts/hot-100/1990-05-12  
3 https://www.billboard.com/charts/hot-100/1990-05-19  
4 https://www.billboard.com/charts/hot-100/1990-05-26  
5 https://www.billboard.com/charts/hot-100/1990-06-02  
6 https://www.billboard.com/charts/hot-100/1990-06-09

Data Summary:

'data.frame': 338395 obs. of 13 variables:  
 $ chart\_position : int 84 78 68 60 58 55 54 57 67 75 ...  
 $ chart\_date : chr "1990-05-05" "1990-05-12" "1990-05-19" "1990-05-26" ...  
 $ song : chr "\"B\" Girls" "\"B\" Girls" "\"B\" Girls" "\"B\" Girls" ...  
 $ performer : chr "Young And Restless" "Young And Restless" "Young And Restless" "Young And Restless" ...  
 $ song\_id : chr "\"B\" GirlsYoung And Restless" "\"B\" GirlsYoung And Restless" "\"B\" GirlsYoung And Restless" "\"B\" GirlsYoung And Restless" ...  
 $ instance : num 1 1 1 1 1 1 1 1 1 1 ...  
 $ time\_on\_chart : int 1 2 3 4 5 6 7 8 9 10 ...  
 $ consecutive\_weeks: num NA 1 2 3 4 5 6 7 8 9 ...  
 $ previous\_week : num NA 84 78 68 60 58 55 54 57 67 ...  
 $ peak\_position : int 84 78 68 60 58 55 54 54 54 54 ...  
 $ worst\_position : int 84 84 84 84 84 84 84 84 84 84 ...  
 $ chart\_debut : chr "1990-05-05" "1990-05-05" "1990-05-05" "1990-05-05" ...  
 $ chart\_url : chr "https://www.billboard.com/charts/hot-100/1990-05-05" "https://www.billboard.com/charts/hot-100/1990-05-12" "https://www.billboard.com/charts/hot-100/1990-05-19" "https://www.billboard.com/charts/hot-100/1990-05-26" ...

Descriptive Statistics:

chart\_position chart\_date song performer   
 Min. : 1.0 Length:338395 Length:338395 Length:338395   
 1st Qu.: 25.5 Class :character Class :character Class :character   
 Median : 50.0 Mode :character Mode :character Mode :character   
 Mean : 50.5   
 3rd Qu.: 75.0   
 Max. :100.0   
   
 song\_id instance time\_on\_chart consecutive\_weeks  
 Length:338395 Min. : 1.000 Min. : 1.000 Min. : 1.00   
 Class :character 1st Qu.: 1.000 1st Qu.: 4.000 1st Qu.: 3.00   
 Mode :character Median : 1.000 Median : 7.000 Median : 7.00   
 Mean : 1.086 Mean : 9.251 Mean : 8.79   
 3rd Qu.: 1.000 3rd Qu.:13.000 3rd Qu.:12.00   
 Max. :15.000 Max. :91.000 Max. :89.00   
 NA's :33667   
 previous\_week peak\_position worst\_position chart\_debut   
 Min. : 1.00 Min. : 1.00 Min. : 1.00 Length:338395   
 1st Qu.: 23.00 1st Qu.: 13.00 1st Qu.: 74.00 Class :character   
 Median : 47.00 Median : 38.00 Median : 86.00 Mode :character   
 Mean : 47.56 Mean : 40.77 Mean : 80.64   
 3rd Qu.: 71.00 3rd Qu.: 65.00 3rd Qu.: 94.00   
 Max. :100.00 Max. :100.00 Max. :100.00   
 NA's :33667   
 chart\_url   
 Length:338395   
 Class :character   
 Mode :character

First few rows of the Data Frame:

chart\_position chart\_date song performer  
1 84 1990-05-05 "B" Girls Young And Restless  
2 78 1990-05-12 "B" Girls Young And Restless  
3 68 1990-05-19 "B" Girls Young And Restless  
4 60 1990-05-26 "B" Girls Young And Restless  
5 58 1990-06-02 "B" Girls Young And Restless  
6 55 1990-06-09 "B" Girls Young And Restless  
 song\_id instance time\_on\_chart consecutive\_weeks  
1 "B" GirlsYoung And Restless 1 1 NA  
2 "B" GirlsYoung And Restless 1 2 1  
3 "B" GirlsYoung And Restless 1 3 2  
4 "B" GirlsYoung And Restless 1 4 3  
5 "B" GirlsYoung And Restless 1 5 4  
6 "B" GirlsYoung And Restless 1 6 5  
 previous\_week peak\_position worst\_position chart\_debut  
1 NA 84 84 1990-05-05  
2 84 78 84 1990-05-05  
3 78 68 84 1990-05-05  
4 68 60 84 1990-05-05  
5 60 58 84 1990-05-05  
6 58 55 84 1990-05-05  
 chart\_url  
1 https://www.billboard.com/charts/hot-100/1990-05-05  
2 https://www.billboard.com/charts/hot-100/1990-05-12  
3 https://www.billboard.com/charts/hot-100/1990-05-19  
4 https://www.billboard.com/charts/hot-100/1990-05-26  
5 https://www.billboard.com/charts/hot-100/1990-06-02  
6 https://www.billboard.com/charts/hot-100/1990-06-09

I may also use other datasets to supplement or do further analysis but I do not have those currently.

## 1.2 The Question

The question I want to answer is: How has the popularity of different music genres evolved over time?

The primary outcome of interest is to see what if any temporal trends exist in the data. Ex: Was pop less popular in the 50-90s and more popular in the 2000s? Did rock become less popular after the 80s?

Some specific predictors I want to look at are time(years), genre, and artists. I may also look into societal and cultural events, or outside factors that could influence music trends. This would not be 100% reliable, but it would be interesting to see if any events line up with trends in the data. The main relationships/patterns that I will be looking for in the data are patterns/trends in genre popularity over time. Though if I find more interesting datasets this may expand and change some.

## 1.3 Analysis

I will likely be analyzing time(years) vs genre vs popularity. Popularity will be some measure based on time spent on chart, peak position, and consecutive weeks. I am not 100% sure how this system will work yet but that is the vague idea. Genre analysis, as mentioned earlier, will be done using the Spotify API (found here - <https://developer.spotify.com/documentation/web-api>) to obtain genre per artist.

# 2. Part 2

## 2.1 Data Cleaning and EDA

The original dataset was already quite clean, however, more processing was done in order to condense the dataset to make it more managable. This was especially important when utilizing the Spotify Web API, as a large dataset can create issues.

The dataset was split into week, month, and year, and then ranked using peak position as the primary factor and time spent on chart as the secondary factor.

After that, the set was condensed even more to only contain unique entries, this was for the purposes of making things easier when using the API.

The bulk of the processing came from using the Spotify Web API to fetch song id, artist id, and genres for each unique entry. Then genres were sorted and simplified into main genres, which will help gain an overall picture of genre trends over time.

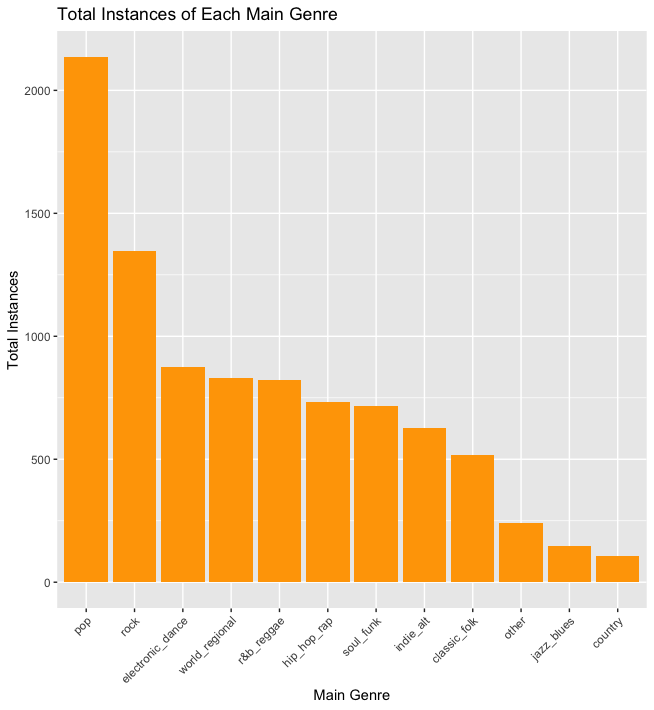
## 2.2 Exploratory Data Analysis

The exploratory part included a stats summary and then two figures. The two figures focused mostly on genre as that is my main focus for this project. Here is the stats summary and the two figures:

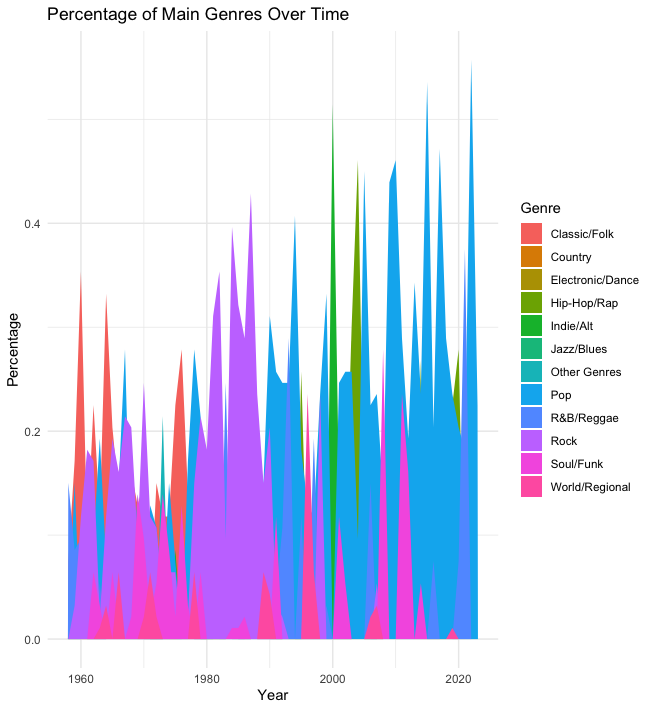
Stats Summary for Hot 100s processed data:

chart\_position.x chart\_date.x song artist   
 Min. : 1.00 Length:3383 Length:3383 Length:3383   
 1st Qu.: 17.00 Class :character Class :character Class :character   
 Median : 32.00 Mode :character Mode :character Mode :character   
 Mean : 36.07   
 3rd Qu.: 47.00   
 Max. :100.00   
   
 instance.x time\_on\_chart.x consecutive\_weeks.x previous\_week.x  
 Min. : 1.000 Min. : 1.00 Min. : 1.00 Min. : 1.00   
 1st Qu.: 1.000 1st Qu.:18.00 1st Qu.:16.00 1st Qu.:12.00   
 Median : 1.000 Median :24.00 Median :22.00 Median :25.00   
 Mean : 1.156 Mean :27.77 Mean :25.04 Mean :29.61   
 3rd Qu.: 1.000 3rd Qu.:35.00 3rd Qu.:32.00 3rd Qu.:41.00   
 Max. :13.000 Max. :91.00 Max. :89.00 Max. :99.00   
 NA's :10 NA's :10   
 peak\_position.x worst\_position.x year.x month.x   
 Min. :1 Min. : 1.00 Min. :1958 Min. : 1.000   
 1st Qu.:1 1st Qu.: 58.00 1st Qu.:1974 1st Qu.: 4.000   
 Median :1 Median : 80.00 Median :1990 Median : 7.000   
 Mean :1 Mean : 73.27 Mean :1990 Mean : 6.524   
 3rd Qu.:1 3rd Qu.: 92.00 3rd Qu.:2007 3rd Qu.:10.000   
 Max. :1 Max. :100.00 Max. :2023 Max. :12.000   
   
 week.x chart\_position.y chart\_date.y instance.y   
 Min. : 1.0 Min. : 1.0 Length:3383 Min. : 1.000   
 1st Qu.:14.0 1st Qu.: 5.0 Class :character 1st Qu.: 1.000   
 Median :27.0 Median : 15.0 Mode :character Median : 1.000   
 Mean :26.6 Mean : 20.9 Mean : 1.119   
 3rd Qu.:40.0 3rd Qu.: 30.0 3rd Qu.: 1.000   
 Max. :53.0 Max. :100.0 Max. :10.000   
 NA's :38 NA's :38   
 time\_on\_chart.y consecutive\_weeks.y previous\_week.y peak\_position.y  
 Min. : 1.00 Min. : 1.00 Min. : 1.00 Min. :1   
 1st Qu.:16.00 1st Qu.:14.00 1st Qu.: 4.00 1st Qu.:1   
 Median :20.00 Median :19.00 Median :10.00 Median :1   
 Mean :22.71 Mean :21.15 Mean :16.08 Mean :1   
 3rd Qu.:28.00 3rd Qu.:27.00 3rd Qu.:22.00 3rd Qu.:1   
 Max. :59.00 Max. :57.00 Max. :99.00 Max. :1   
 NA's :38 NA's :62 NA's :62 NA's :38   
 worst\_position.y year.y month.y week.y   
 Min. : 1.00 Min. :1958 Min. : 1.000 Min. : 1.00   
 1st Qu.: 56.00 1st Qu.:1974 1st Qu.: 4.000 1st Qu.:15.00   
 Median : 78.00 Median :1990 Median : 7.000 Median :29.00   
 Mean : 70.55 Mean :1990 Mean : 6.805 Mean :27.73   
 3rd Qu.: 90.00 3rd Qu.:2006 3rd Qu.:10.000 3rd Qu.:41.00   
 Max. :100.00 Max. :2022 Max. :12.000 Max. :53.00   
 NA's :38 NA's :38 NA's :38 NA's :38   
 spotify\_song\_id spotify\_artist\_id genres main\_genres   
 Length:3383 Length:3383 Length:3383 Length:3383   
 Class :character Class :character Class :character Class :character   
 Mode :character Mode :character Mode :character Mode :character

Total Instances of Each Main Genre:



Percentage of Main Genres Over Time:



This particular graph is the kind of analysis I am interested in as it shows the progression of genre popularity over time. This is something I will explore further during the analysis portion.