

**Data Source (URL web address with hyperlink):**

- <https://www.billboard.com/charts/billboard-global-200>

**Context of Data and Variables:**

- The data scraped from the billboard is the rank of the most popular songs, and the following variables are obtained for my research.

**Output from str() function applied to the data object (apply a monospaced font like “Courier New” to the output):**

- > str(rankstibble)
- tibble [200 x 4] (S3: tbl\_df/tbl/data.frame)
- \$ names : chr [1:200] "Good 4 U" "Levitating" "Kiss Me More" "my.life" ...
- \$ last : num [1:200] NA 3 5 NA 1 2 4 NA NA 6 ...
- \$ peak : num [1:200] 1 2 3 4 1 1 1 8 9 3 ...
- \$ weeks\_onchart: num [1:200] 1 33 6 1 8 9 20 1 1 17 ...

**Research Questions to be explore:**

1. The relationship of the variable “peak” and “weeks\_onchart”.
2. Not ensure but try to find the relationship of the trend .

## Statistical Analysis Plan

**Population**

- All popular songs in the sequence rule of Billboard.

**Primary Objective:**

- Estimate the change of rank in one week and find if the peak is higher and the weeks\_onchart is higher too.

**Secondary Objectives:**

- Estimate the relationship between the trend(peak minus last) and the weeks\_onchart.

**Data Collection methods:**

- Popular songs in Billboard rank.

**Variables under consideration:**

- Names: The name of song.
- Last: The rank of the song in last week.
- Peak: The best rank for the song.
- Weeks\_onchart : How long the song exists in the rank.

**Missing data procedures:**

- If data on any of the following are missing, set them as “201”.(Because it means that the song is new and it entered the top 200 in a short time.)

**Numerical and graphical summaries to be presented:**

- The scatterplot which terms peak as axis X and weeks\_onchart as axis Y.
- The scatterplot which terms last as axis X and peak as axis Y.

**Models to be fitted:**

- Primary objective:
- General linear model:  
weeks\_onchart~peak + last +possible interactions