**Problem Statement:**

The goal of this project is to analyze and visualize the trends in Spotify's "Top 200" and "Viral 50" charts from 2019 to 2021. By examining various attributes of the charts, such as song popularity, artist representation, and genre distribution, we aim to uncover insights into listening trends, song performance over time, and factors contributing to viral success.

**Dataset Description:**

The dataset consists of the complete collection of the "Top 200" and "Viral 50" charts published by Spotify globally from 2019 to 2021. It contains the following columns:

* **title:** The title of the song (string).
* **rank:** The rank of the song in the chart (integer).
* **date:** The date the chart was published (date).
* **artist:** The artist or group performing the song (string).
* **url:** The link to the song on Spotify (string).
* **region:** The region in which the chart was published (categorical).
* **chart:** The type of chart (e.g., "Top 200" or "Viral 50") (categorical).
* **trend:** The trend of the song (e.g., "up," "down," or "steady") (categorical).
* **streams:** The number of times the song has been streamed (numeric).

**Techniques Used:**

1. **Visualization Libraries:**
   * **ggplot2** for creating visualizations.
2. **Data Manipulation Libraries:**
   * **dplyr** for data manipulation tasks.
   * **tidyr** for reshaping data.

**Preprocessing:**

* **Handling Missing Data**:  
  Checked for any missing values, particularly in important columns like streams, region, and artist. Removed entries with missing values.
* **Converting Data Types**:  
  Converted date column to be in the date datatype.
* **Filtering Unnecessary Columns**:  
  Filtered out columns not necessary for visualization. Selected columns: artist, song, streams, region, date, trend.
* **Grouping and Aggregating Data**:  
  Aggregate data based on key dimensions like artist, region, song, and date for efficient analysis.

**Visualization:**

1. **Total Streams Over Time (Yearly)**

* **Preprocessing**:  
  Grouped data by year to calculate the total streams for each year from 2019 to 2021.
* **Analysis**:  
  The total streams each year reveal trends in music consumption. This allows for easy identification of peak streaming periods.
* **Visualization Understanding**:  
  A bar plot shows yearly total streams in millions, with values annotated. This provides a clear overview of how streaming volume changed year by year.

A graph of a number of blue squares

Description automatically generated with medium confidence

1. **Top 10 Artists by Total Streams**

* **Preprocessing**:  
  Data was aggregated by artist, and the top 10 artists were selected based on total streams.
* **Analysis**:  
  This analysis identifies the artists with the highest total streams, giving insight into the most popular musicians from 2019 to 2021.
* **Visualization Understanding**:  
  A horizontal bar chart shows the top 10 artists, with the number of streams displayed next to each bar. The use of a flipped axis makes artist names more readable.

A graph of different colored bars

Description automatically generated with medium confidence

1. **Top 10 Songs by Total Streams**

* **Preprocessing**:  
  Data was aggregated by song title, and the top 10 songs were selected based on total streams.
* **Analysis**:  
  It highlights the most streamed songs during the period, showcasing which tracks garnered the most attention globally.
* **Visualization Understanding**:  
  A horizontal bar plot lists the top 10 songs. Custom shortened titles are used to make the chart clean and easy to read.

A graph with different colored bars

Description automatically generated

1. **Total Streams by Region**

* **Preprocessing**:  
  Data was aggregated by region, excluding "Global" for more accurate regional comparison.
* **Analysis**:  
  Regional analysis helps in understanding which areas contribute the most to streaming volumes.
* **Visualization Understanding**:  
  A bar plot presents streams per region in millions. The regions are ranked, giving a straightforward view of which markets are largest.

A graph with a number of streams

Description automatically generated

1. **Map of Total Streams by Region**

* **Preprocessing**:  
  Data was grouped by region and merged with a world map to show the geographic distribution of streaming.
* **Analysis**:  
  Mapping the data gives a geographic context to streaming behavior, allowing for easy visual comparison between regions.
* **Visualization Understanding**:  
  A map visualizes streaming volumes by region with color coding. Darker colors indicate regions with higher total streams, creating a global view of music consumption.

A map of the world

Description automatically generated

1. **Top 10 Regions by Average Streams**

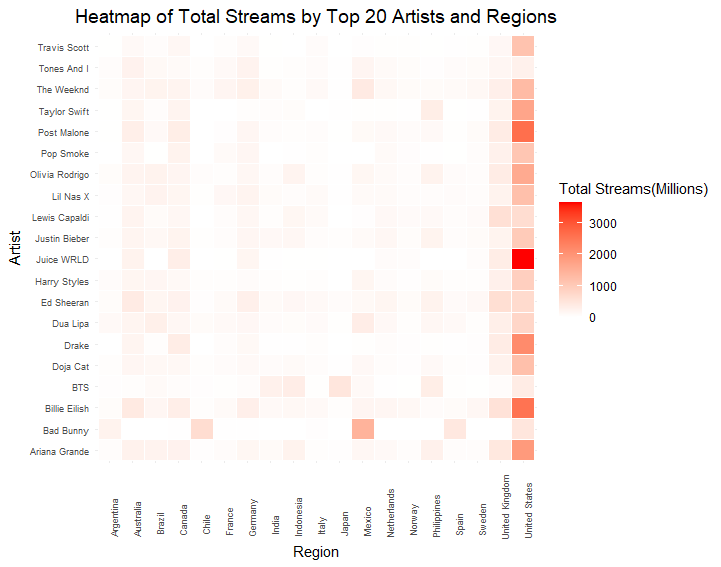
* **Preprocessing**:  
  Data was grouped by region, and the average streams per region were calculated.
* **Analysis**:  
  This analysis highlights the regions with the highest average streams, offering insights into per-region listening intensity.
* **Visualization Understanding**:  
  A bar chart shows the top 10 regions by average streams. This comparison is useful for identifying markets where songs are consistently highly streamed.

A graph of a number of regions

Description automatically generated

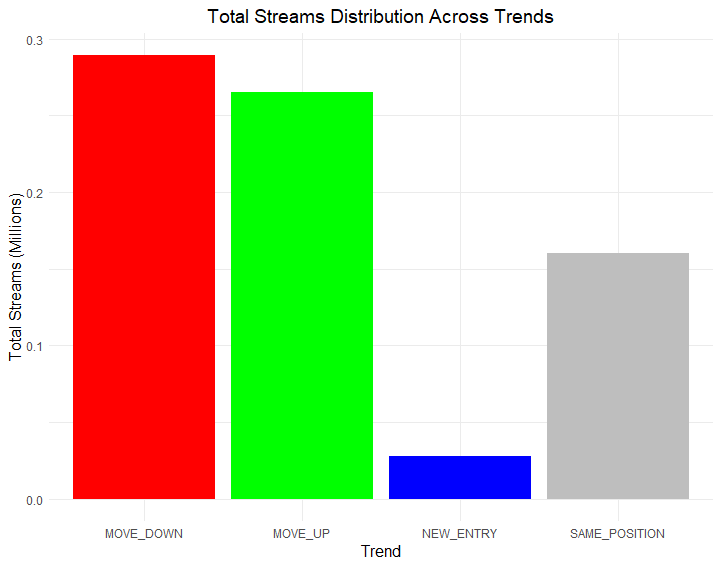
1. **Heatmap of Total Streams by Top 20 Artists in Top 20 Regions**

* **Preprocessing**:  
  Data was aggregated by both artist and region, with a focus on the top 20 artists and top 20 regions for total streams.
* **Analysis**:  
  This analysis explores how different artists perform across various regions, revealing global and regional patterns in popularity.
* **Visualization Understanding**:  
  A heatmap shows the distribution of streams by artist and region, with color gradients indicating streaming intensity. This plot offers a high-level overview of artist performance across regions.



1. **Total Streams Distribution Across Trends**

* **Preprocessing**:  
  Data was aggregated based on the trend category (e.g., MOVE\_UP, MOVE\_DOWN, etc.), and total streams were calculated for each trend.
* **Analysis**:  
  Trend analysis helps in understanding the dynamics of how songs move in the charts, such as which categories (new entries, moving up, or down) accumulate the most streams.
* **Visualization Understanding**:  
  A bar plot visualizes total streams for each trend category, using distinct colors to differentiate them. This shows how different song trends perform in terms of streaming.



1. **Time Series Plot of Total Streams (Monthly)**

* **Preprocessing**:  
  Data was filtered to include only entries up to August 2021 and aggregated by month for total streams.
* **Analysis**:  
  Monthly trends in total streams provide insight into seasonality or any sharp rises in streaming during specific months.
* **Visualization Understanding**:  
  A line plot displays the total streams per month. The time series format highlights month-by-month fluctuations in streaming, showing growth or dips over the years.

A graph showing a line of blue lines

Description automatically generated with medium confidence

**Conclusion:**

The analysis of Spotify's Top 200 and Viral 50 datasets (2019-2021) reveals several key insights:

1. **Streaming Growth**: There was a significant increase in music streaming during the pandemic, with a clear upward trend over time, signaling the growing dominance of digital platforms.
2. **Artist Dominance**: A small group of global superstars, like Bad Bunny and Drake, consistently dominate streaming charts, highlighting the concentration of streams among top-tier artists.
3. **Regional Disparities**: Regions such as the US, Brazil, and the UK lead in streaming, with notable regional variations in musical preferences and activity.

Overall, the findings emphasize the importance of a few key artists and regions in shaping global music consumption trends.