

# Music & Unemployment in 2020

Tracy Ammann, Rosaicela Roman, Amber Royster

University of Arizona Continuing & Professional Education  
Data Visualization Bootcamp

Presented on May 1, 2021

# Hypothesis: Listening trends were impacted by unemployment in 2020.

Was there a correlation between streaming and changes in unemployment?  
Were people's genre tastes impacted throughout the course of 2020?

# Data Sources, Data Points Considered

## Music

- Spotify Charts Viral 50 (CSV)
  - Genres
- Spotify Charts Top 200 (CSV)
  - Genres, Stream Count
- Weekly charts
- Jan, Mar, May, Jul, Sep, Nov
- Spotify API

## Unemployment

- U.S. Department of Labor, Employment & Training Administration, Unemployment Insurance Weekly Claims Data (Excel converted to CSV)
  - Weekly number of claims (seasonally adjusted)

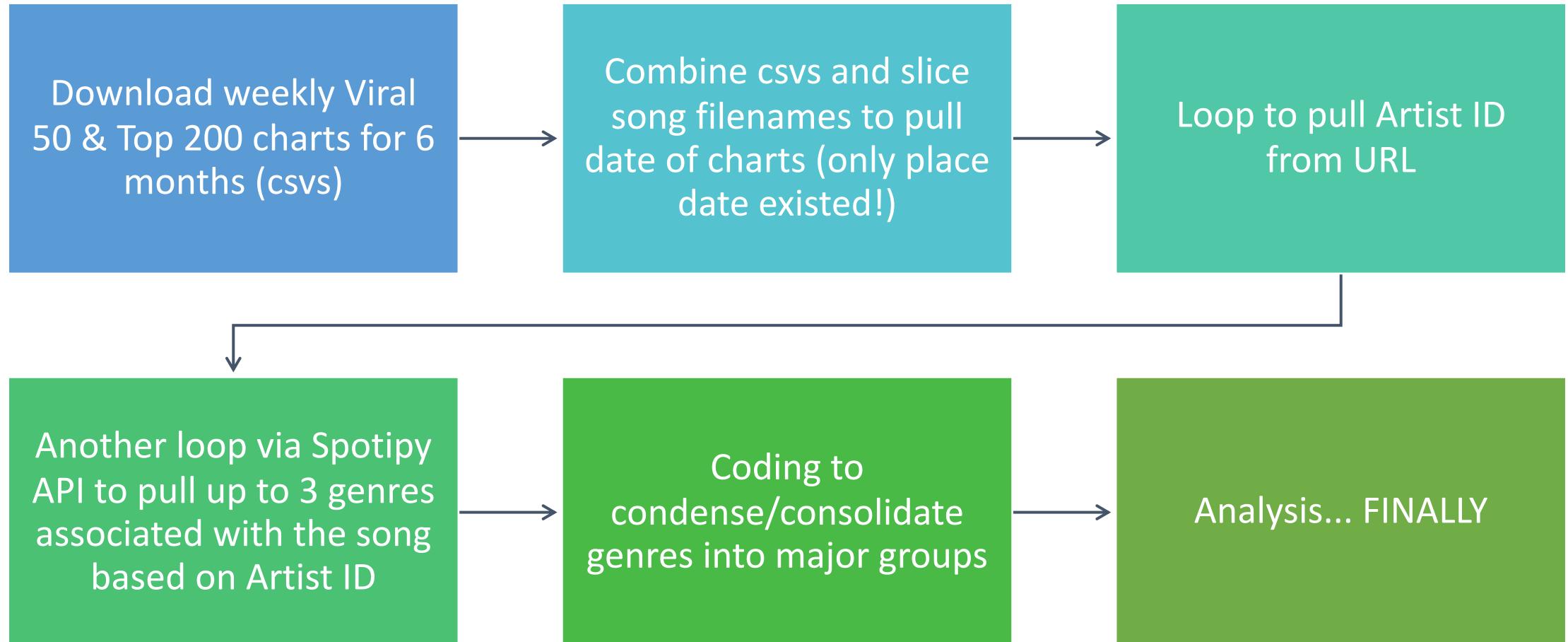
# Why those 6 months?

- January – beginning of the year – all is well ☺
- Mar – PANDEMIC – stay at home orders begin
- May – murder of George Floyd, Black Lives Matter demonstrations
- July – prolonged pandemic/midyear
- September – wildfires raging
- November – election...

# Limitations

- Originally wanted to incorporate podcast data; lacked streaming numbers over time
- Other music/podcast APIs (e.g. Apple Music, Listen API) were fee-based and/or very difficult to navigate
- Spotify Charts was not ideal; no user data available beyond our own; could look up individual songs but without stream data
- Genres were extremely diverse and needed to be grouped based on keyword/broad category (e.g. gauze pop = pop; ohio hip hop = hip hop)
- We limited song genres to (3); some songs had up to (5)
- Lack of time to identify and wrangle more data sources

# An intense data wrangling process...



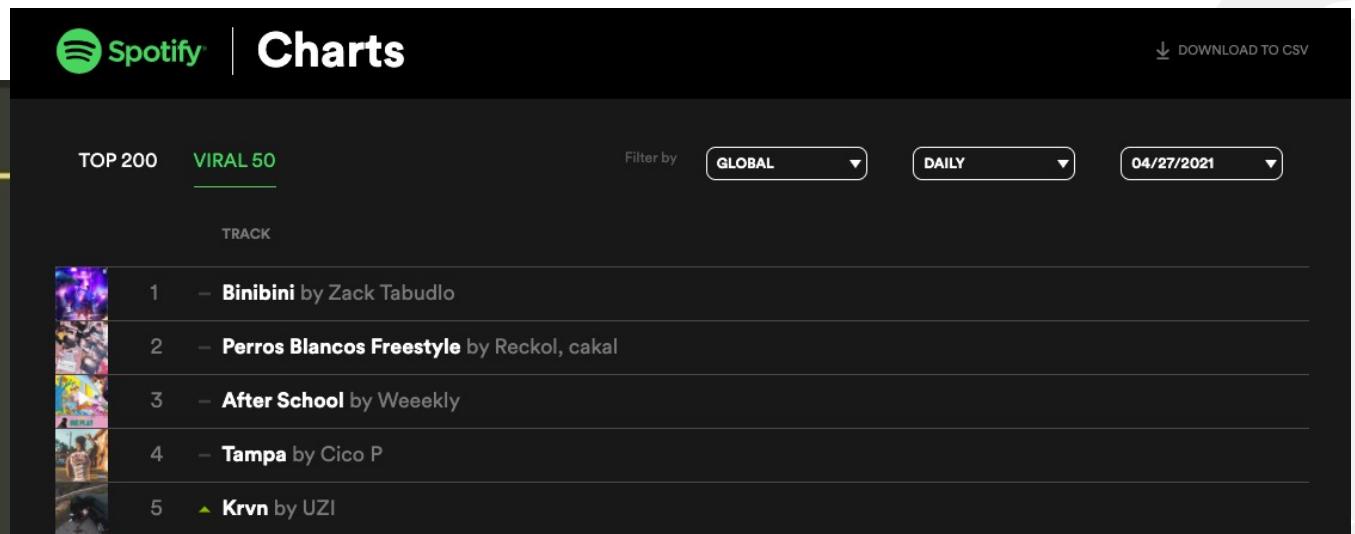
```
▶ ▶ Ml

# Import dependencies
import matplotlib.pyplot as plt
from matplotlib import cm
import pandas as pd
import numpy as np
import glob
import datetime
import time
import json
import spotipy

client_id = "9ab504ff3c144d03924e45bd7cae5168"
client_secret = "secret"

from spotipy.oauth2 import SpotifyClientCredentials

client_credentials_manager = SpotifyClientCredentials(client_id, client_secret)
sp = spotipy.Spotify(client_credentials_manager=client_credentials_manager)
```



▶ ▶ M↓

```
# merge viral 50 for January
files = glob.glob('/Users/amber/Desktop/GitHub/project-one/Data/viral_50/jan/*.csv')
jan_viral_50 = pd.concat([pd.read_csv(fp).assign(Filename=os.path.basename(fp)) for fp in files])

jan_viral_50['Month'] = ''
jan_viral_50['Week Number'] = ''
jan_viral_50['Type'] = ''

# Extract Date from Filename using slice
jan_viral_50['Date'] = jan_viral_50['Filename'].str.slice(16,26)

# Convert to datetime
jan_viral_50['Date'] = pd.to_datetime(jan_viral_50['Date'])
|
# Get Week Number from Date
jan_viral_50['Week Number'] = jan_viral_50['Date'].dt.isocalendar().week

for index, row in jan_viral_50.iterrows():
    jan_viral_50.loc[index, 'Month'] = 'January'
    jan_viral_50.loc[index, 'Type'] = 'Viral 50'

jan_viral_50.to_csv('/Users/amber/Desktop/GitHub/project-one/Data/viral_50/jan/jan_viral_50.csv')

jan_viral_50.head()
```

▶ ▶ M1

```
# create columns for Spotify data
jan_viral_50['Artist ID'] = ''
jan_viral_50['Genre 1'] = ''
jan_viral_50['Genre 2'] = ''
jan_viral_50['Genre 3'] = ''

# create list to collect artist ID
artist_id = []

# loop to go through dataframe and pull artist ID
for index, row in jan_viral_50.iterrows():
    track_id = row['URL']
    track_features = sp.track(track_id)
    jan_viral_50.loc[index, 'Artist ID'] = track_features['album']['artists'][0]['id']
    time.sleep(1.01)

# loop to go through dataframe and pull genres
for index, row in jan_viral_50.iterrows():
    genre_id = row['Artist ID']
    artist_features = sp.artist(genre_id)
    try:
        jan_viral_50.loc[index, 'Genre 1'] = artist_features['genres'][0]
        jan_viral_50.loc[index, 'Genre 2'] = artist_features['genres'][1]
        jan_viral_50.loc[index, 'Genre 3'] = artist_features['genres'][2]
    except (IndexError, KeyError):
        print(f'No genre found...skipping.')
    time.sleep(1.01)

jan_viral_50.head()
```

▶ ▶ M4

```
# consolidate genres for jan viral 50

jan_genre_data_50 = {
    'Genre': ['Rap/Hip Hop/Trap', 'R&B/Soul/Blues', 'Pop', 'Rock', 'Country', 'Latin', 'Disco', 'Indie', 'House/EDM', 'Alternative/Alt', 'Glitchcore', 'Brooklyn Drill', 'Halloween', 'Christmas', 'Classical'],
    'Count': [len(jan_viral_50[jan_viral_50['Genre 1']].str.contains('rap|hip hop|trap') | jan_viral_50[jan_viral_50['Genre 2']].str.contains('rap|hip hop|trap') | jan_viral_50[jan_viral_50['Genre 3']].str.contains('rap|hip hop|trap'))], len(jan_viral_50[jan_viral_50['Genre 1']].str.contains('r&b|soul|blues') | jan_viral_50[jan_viral_50['Genre 2']].str.contains('r&b|soul|blues') | jan_viral_50[jan_viral_50['Genre 3']].str.contains('r&b|soul|blues'))], len(jan_viral_50[jan_viral_50['Genre 1']].str.contains('pop') | jan_viral_50[jan_viral_50['Genre 2']].str.contains('pop') | jan_viral_50[jan_viral_50['Genre 3']].str.contains('pop'))], len(jan_viral_50[jan_viral_50['Genre 1']].str.contains('rock') | jan_viral_50[jan_viral_50['Genre 2']].str.contains('rock') | jan_viral_50[jan_viral_50['Genre 3']].str.contains('rock'))], len(jan_viral_50[jan_viral_50['Genre 1']].str.contains('country') | jan_viral_50[jan_viral_50['Genre 2']].str.contains('country') | jan_viral_50[jan_viral_50['Genre 3']].str.contains('country'))], len(jan_viral_50[jan_viral_50['Genre 1']].str.contains('latin') | jan_viral_50[jan_viral_50['Genre 2']].str.contains('latin') | jan_viral_50[jan_viral_50['Genre 3']].str.contains('latin'))], len(jan_viral_50[jan_viral_50['Genre 1']].str.contains('disco') | jan_viral_50[jan_viral_50['Genre 2']].str.contains('disco') | jan_viral_50[jan_viral_50['Genre 3']].str.contains('disco'))], len(jan_viral_50[jan_viral_50['Genre 1']].str.contains('indie') | jan_viral_50[jan_viral_50['Genre 2']].str.contains('indie') | jan_viral_50[jan_viral_50['Genre 3']].str.contains('indie'))], len(jan_viral_50[jan_viral_50['Genre 1']].str.contains('house|edm') | jan_viral_50[jan_viral_50['Genre 2']].str.contains('house|edm') | jan_viral_50[jan_viral_50['Genre 3']].str.contains('house|edm'))], len(jan_viral_50[jan_viral_50['Genre 1']].str.contains('alt|alternative') | jan_viral_50[jan_viral_50['Genre 2']].str.contains('alt|alternative') | jan_viral_50[jan_viral_50['Genre 3']].str.contains('alt|alternative'))], len(jan_viral_50[jan_viral_50['Genre 1']].str.contains('glitchcore') | jan_viral_50[jan_viral_50['Genre 2']].str.contains('glitchcore') | jan_viral_50[jan_viral_50['Genre 3']].str.contains('glitchcore'))], len(jan_viral_50[jan_viral_50['Genre 1']].str.contains('brooklyn drill') | jan_viral_50[jan_viral_50['Genre 2']].str.contains('brooklyn drill') | jan_viral_50[jan_viral_50['Genre 3']].str.contains('brooklyn drill'))], len(jan_viral_50[jan_viral_50['Genre 1']].str.contains('halloween') | jan_viral_50[jan_viral_50['Genre 2']].str.contains('halloween') | jan_viral_50[jan_viral_50['Genre 3']].str.contains('halloween'))], len(jan_viral_50[jan_viral_50['Genre 1']].str.contains('christmas') | jan_viral_50[jan_viral_50['Genre 2']].str.contains('christmas') | jan_viral_50[jan_viral_50['Genre 3']].str.contains('christmas'))], len(jan_viral_50[jan_viral_50['Genre 1']].str.contains('classical') | jan_viral_50[jan_viral_50['Genre 2']].str.contains('classical') | jan_viral_50[jan_viral_50['Genre 3']].str.contains('classical'))]
}

# create new dataframe
jan_genre_df = pd.DataFrame(jan_genre_data_50)
```

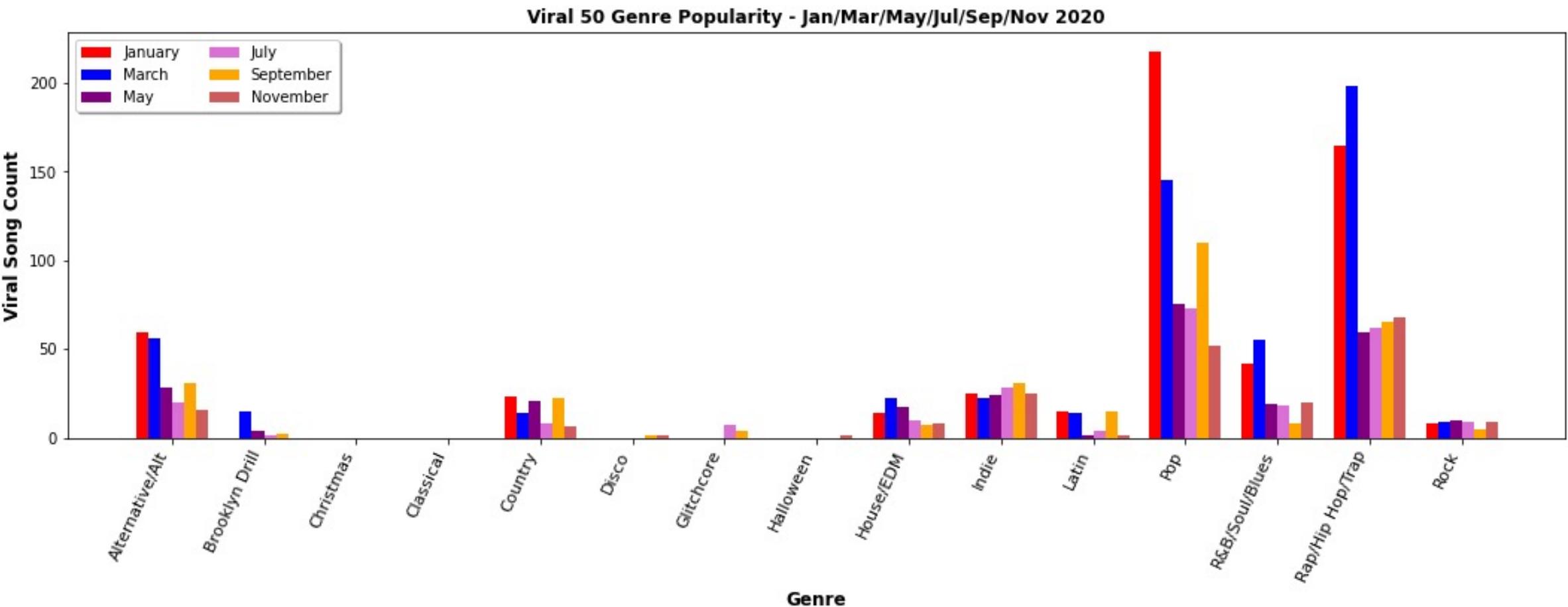
Could this have been easier? Yeah... probably...

# We basically did that 12 times...

Whereas the Unemployment data was just an Excel file we converted to CSV...

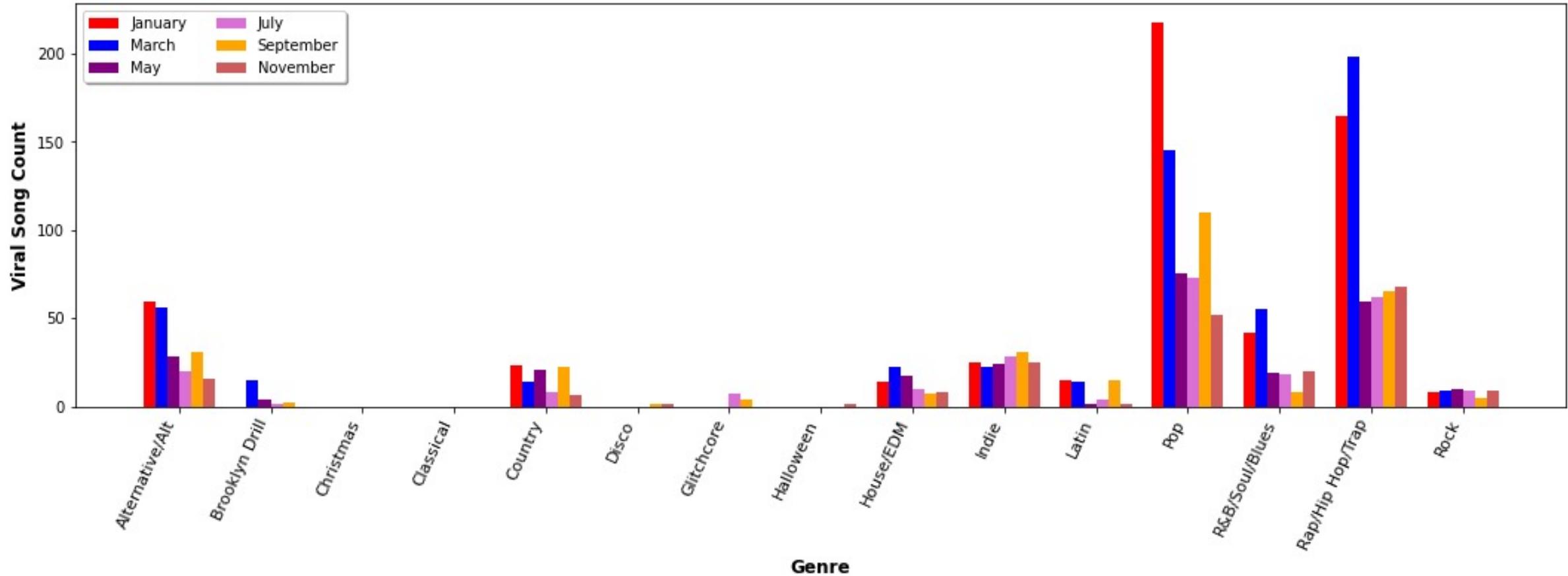
# Here are our results

- Pop and Rap/Hip Hop/Trap are clearly the top genres of 2020
- When Pop is top in a month, Rap/Hip Hop/Trap is next, and vice versa
- There is an interesting and sharp decline in Pop and Rap/Hip Hop/Trap genre popularity from March to May; Pop somewhat recovers in September

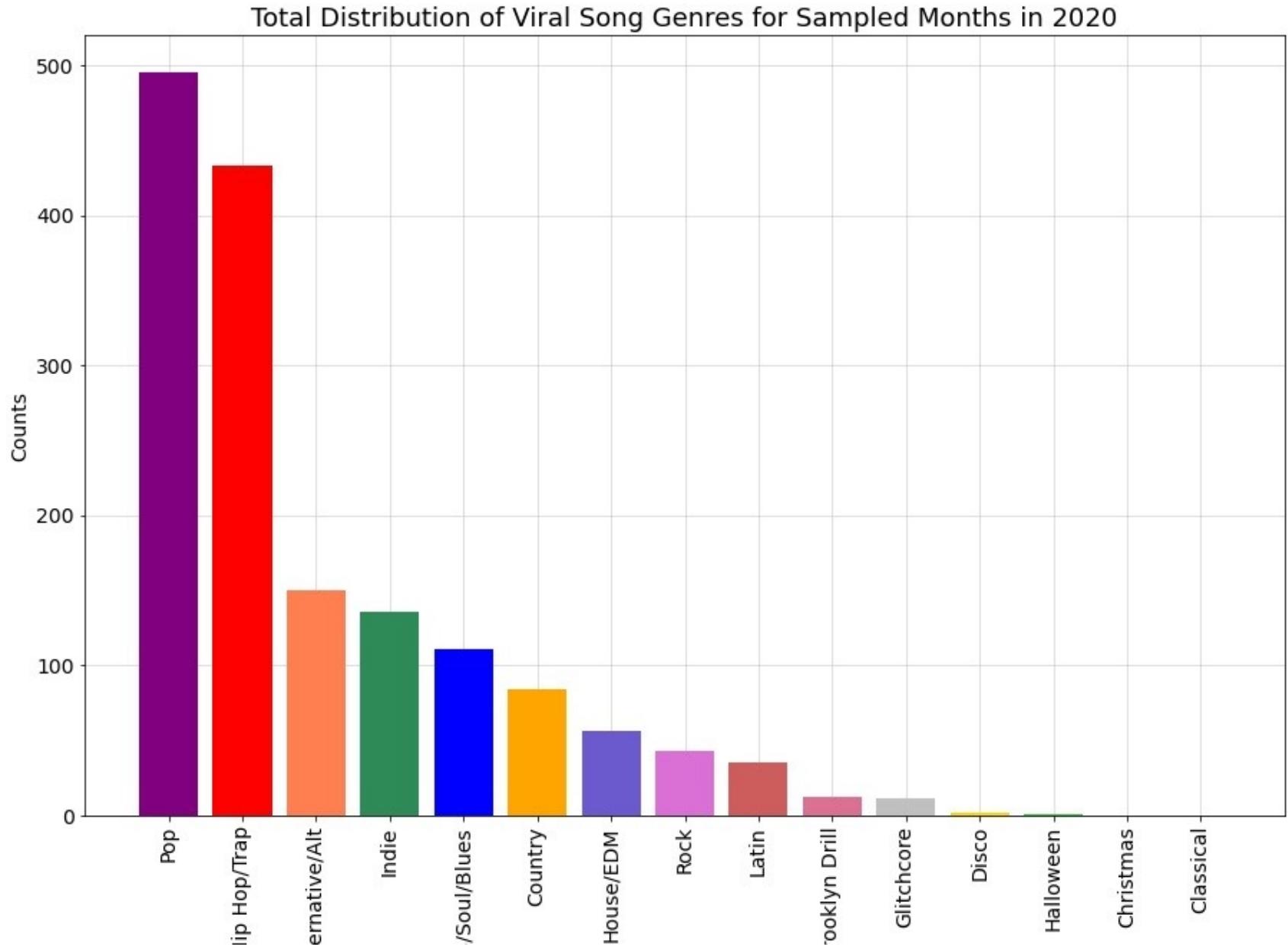


See? Not much difference.

**Top 200 Genre Popularity - Jan/Mar/May/Jul/Sep/Nov 2020**

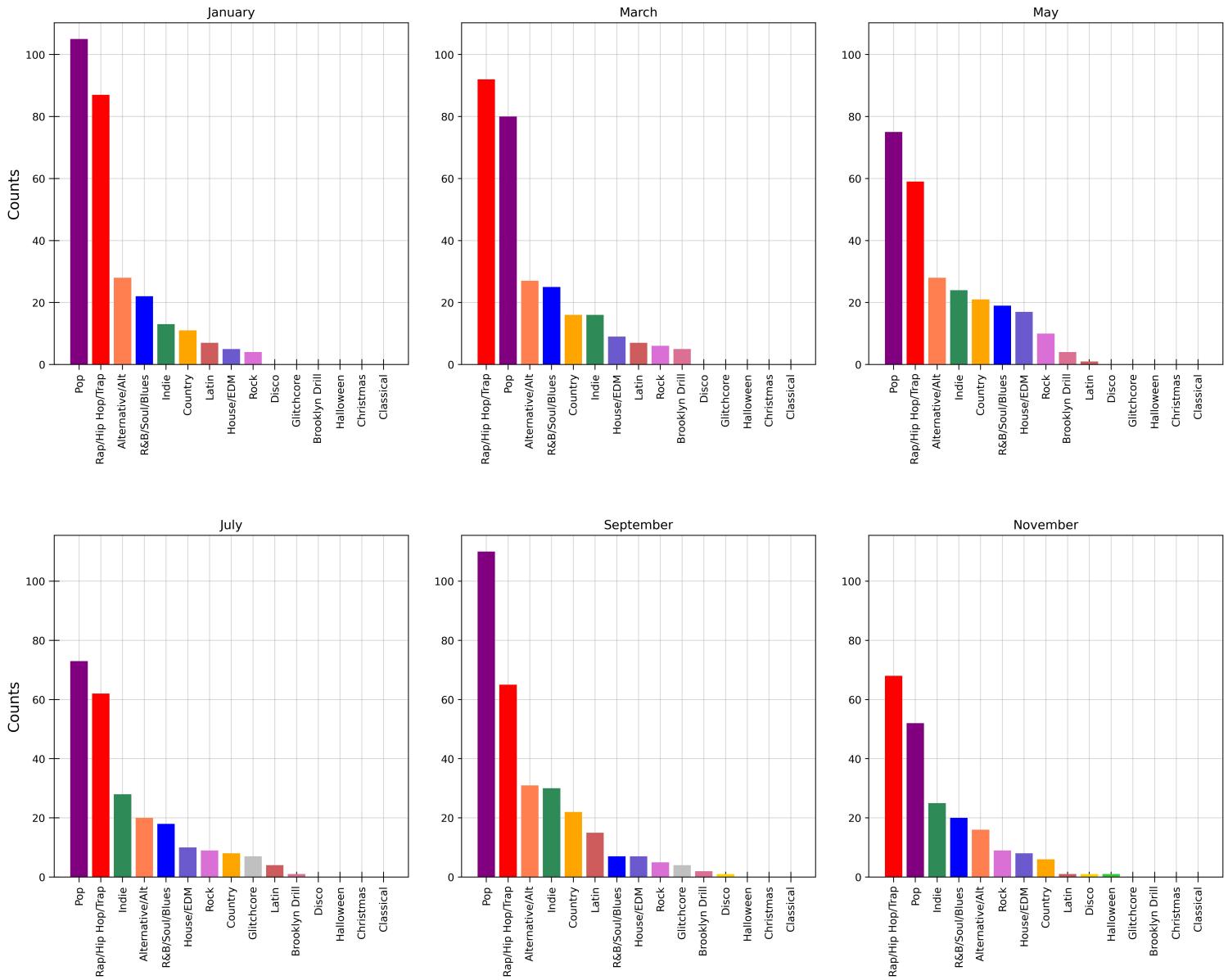


- This plot shows the total distribution of Viral 50 song genres
- Pop is king, followed by Rap/Hip Hop/Trap
- Classical not so much



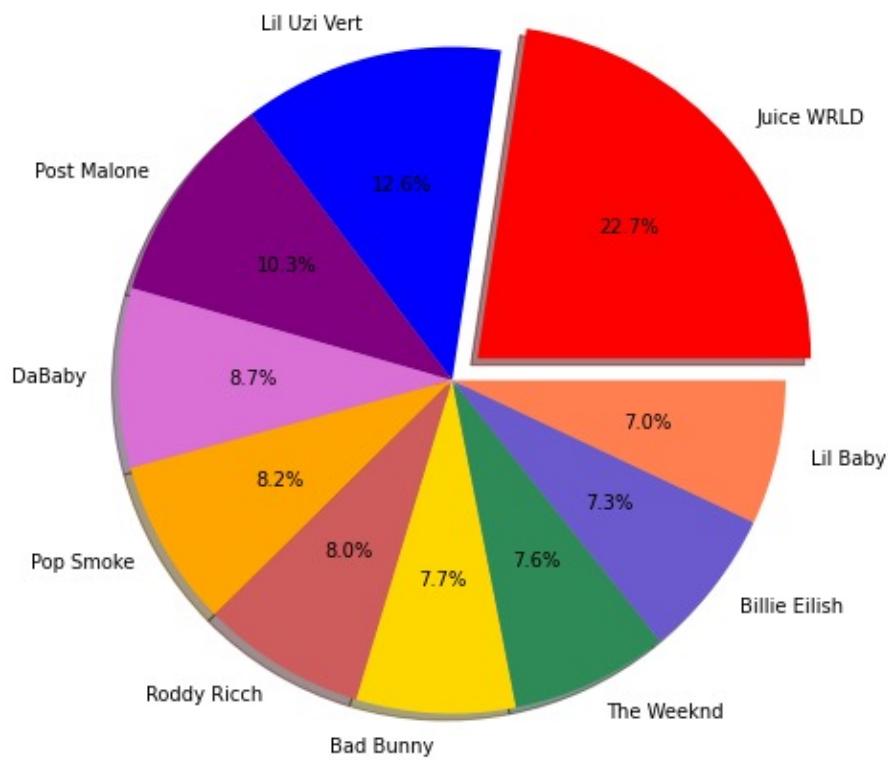
Distribution of Viral Song Genres in 2020

- This plot shows the distribution of all (6) months analyzed in a single frame

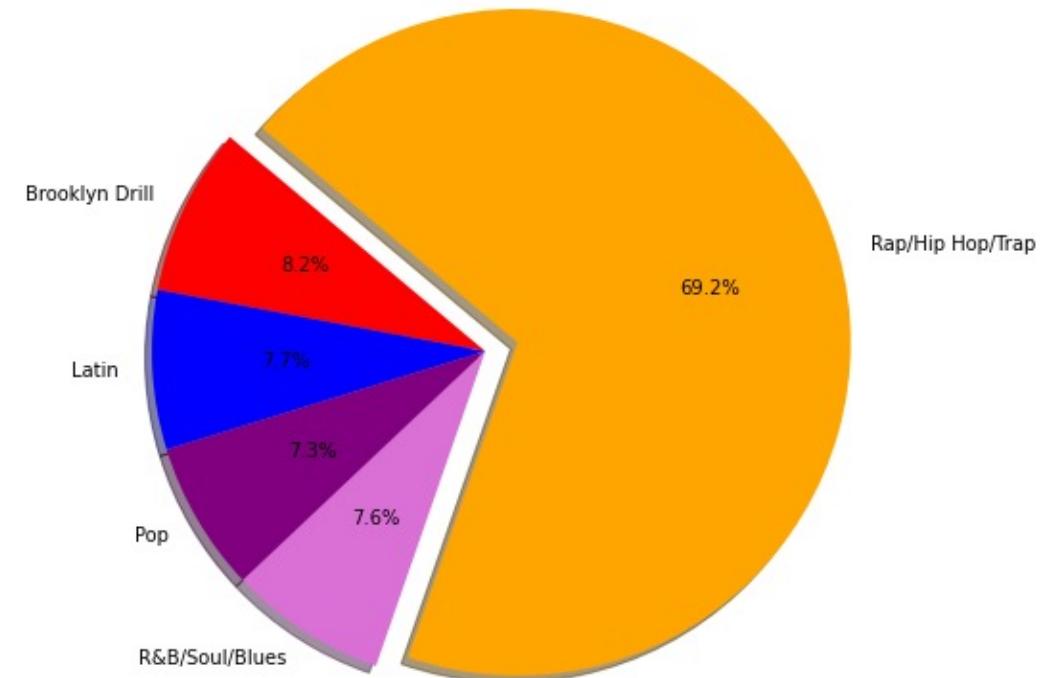


Here's a look at the Top 10 Artists of 2020 and their genres...

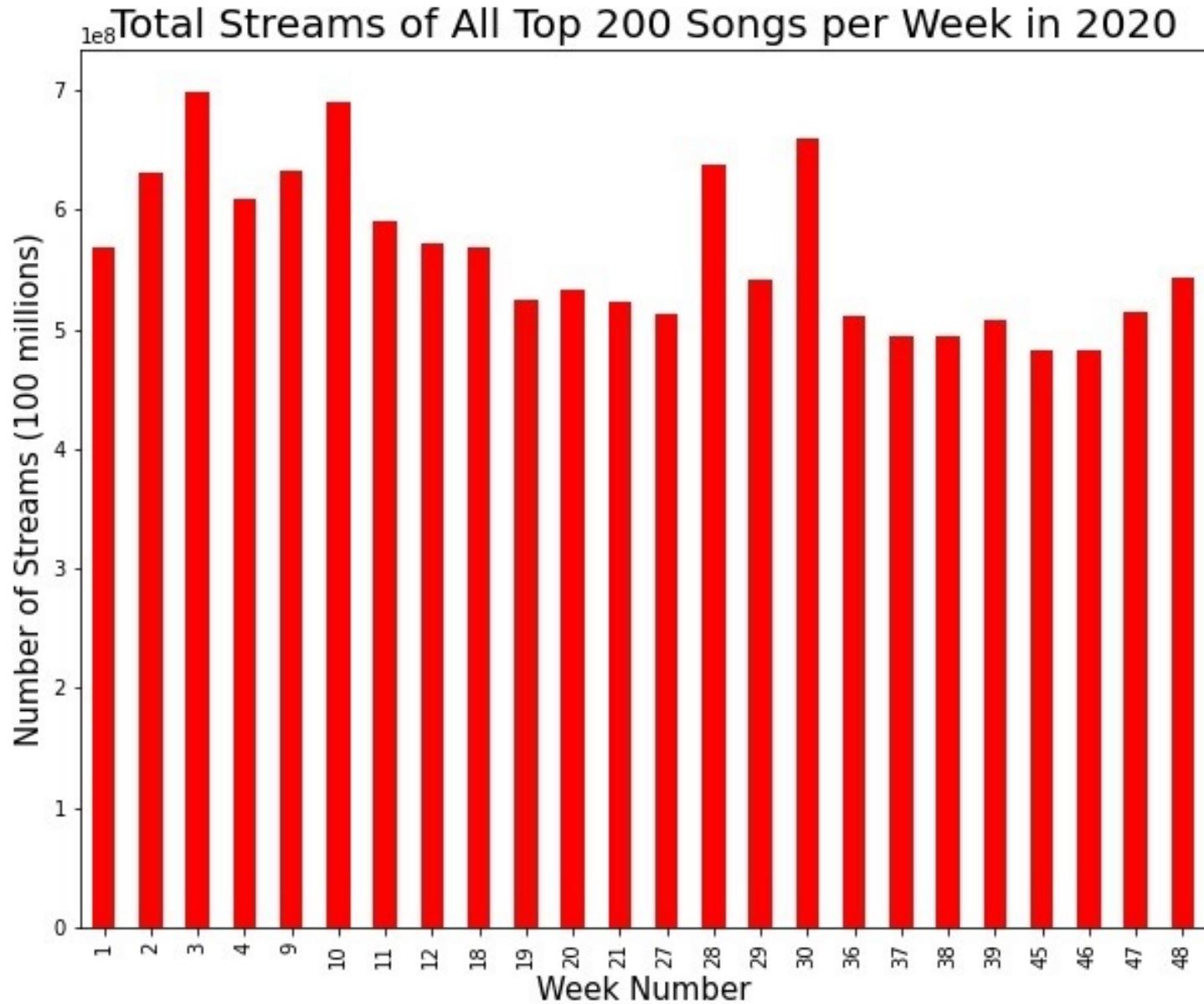
Top 10 Artists of 2020



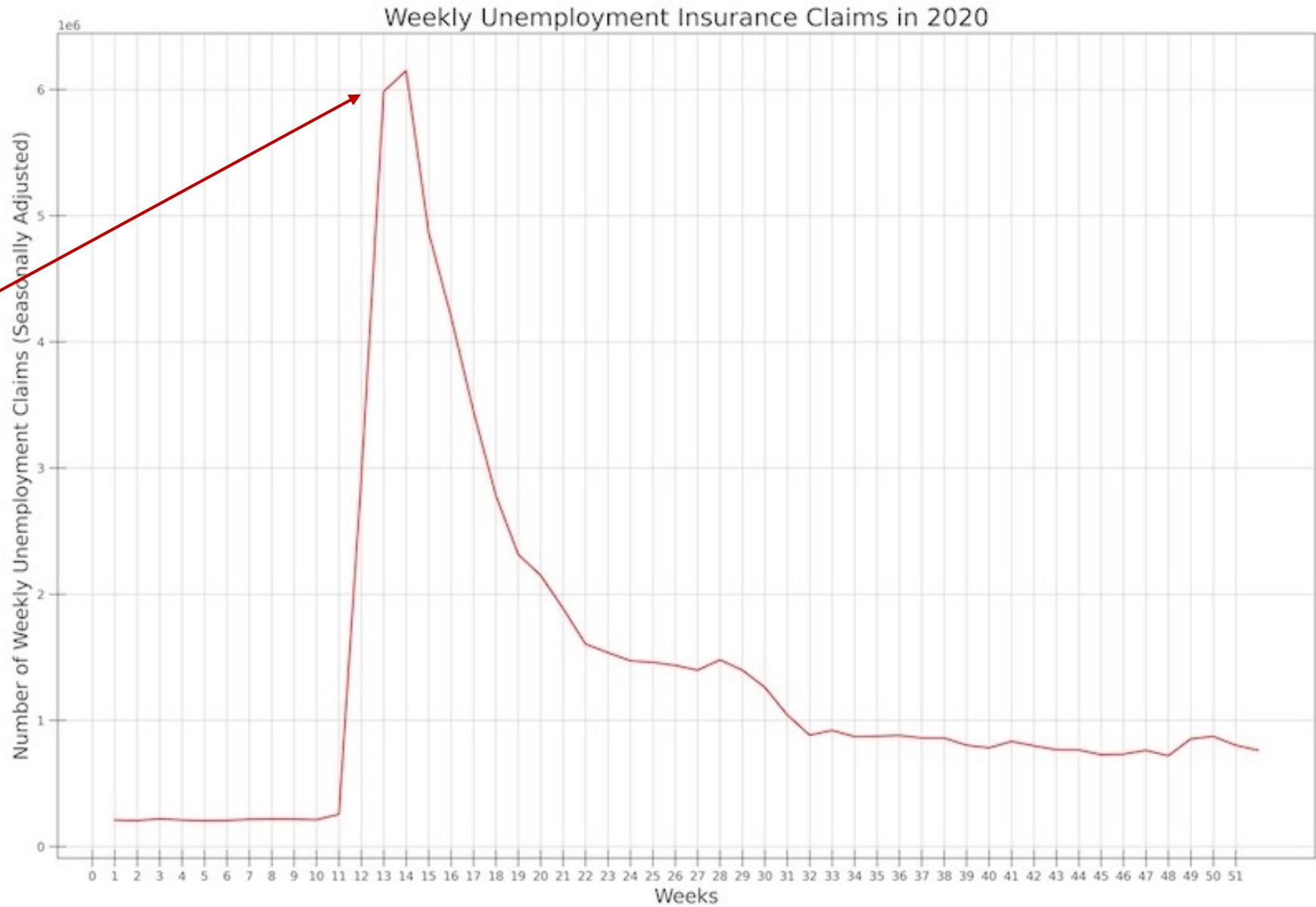
Top 10 Artists Genre Makeup



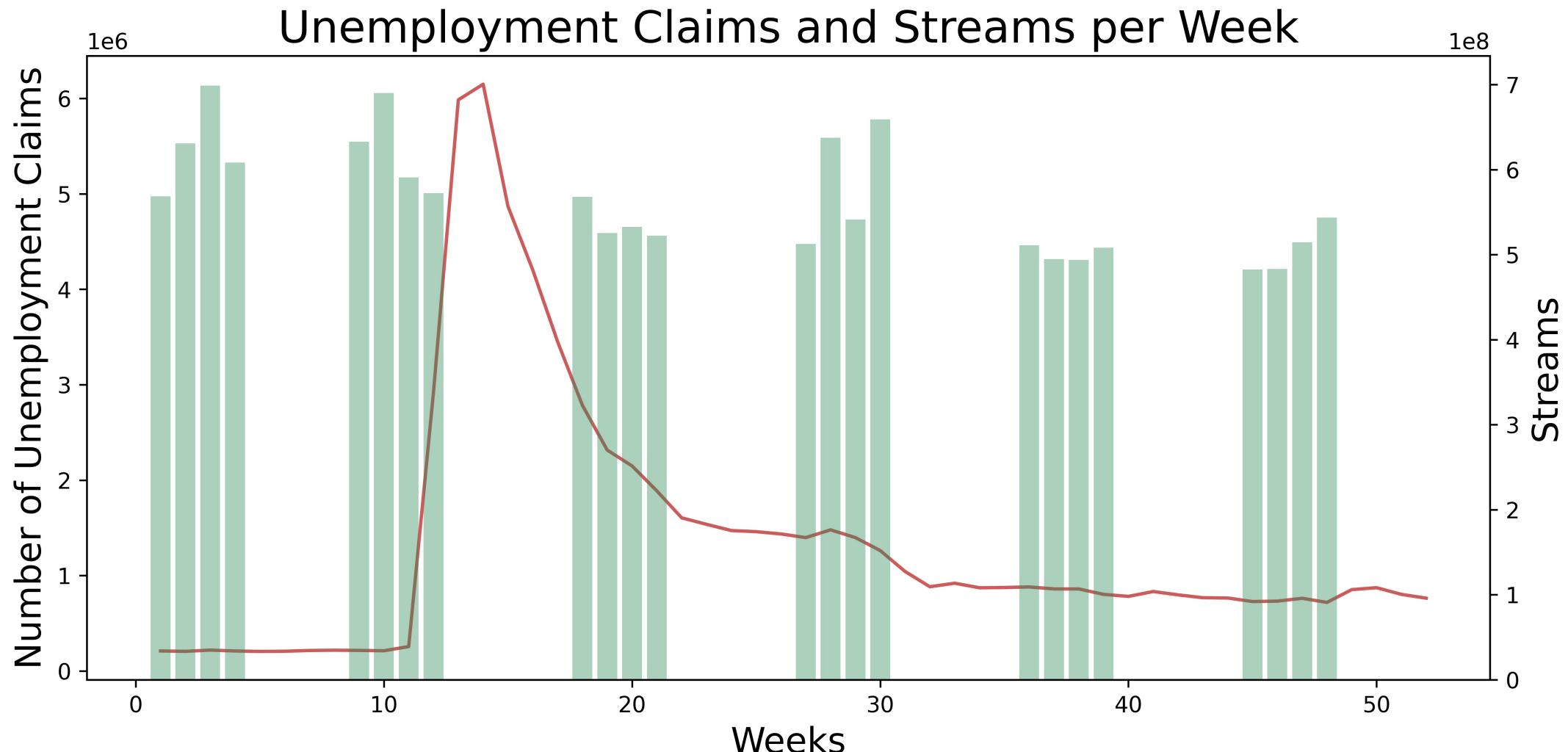
- There is not significant variation in streams per week; however, it does appear that streams tend to increase toward the middle to end of each month.



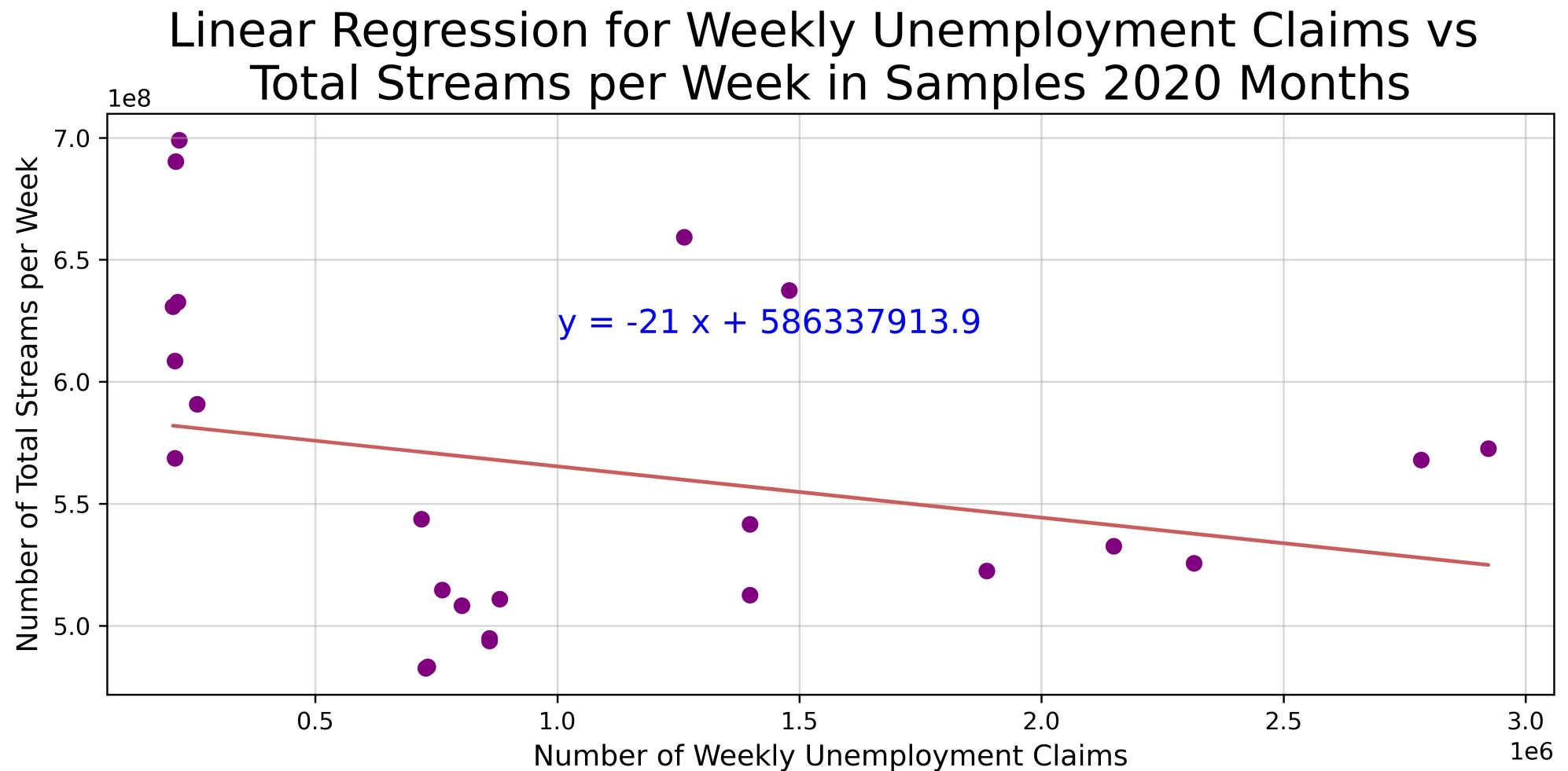
- This is unemployment claims data for 2020 – that spike you see is when the stay-at-home orders started



- Finally, here is our plot showing Top 200 streaming numbers (100 million) against unemployment claims per week (1 million); this data does not indicate any consistent correlation between Top 200 streaming trends and unemployment claims in 2020.



- Last, we ran a linear regression of Weekly Unemployment Claims vs Total Streams per week of our sample months... while there is a slight negative correlation, the r-squared value of 0.07 indicates our model does not explain data variability.



# Summary

## Conclusions

- Hypothesis: We were wrong.
- Based on the data we had available and could wrangle, it does not appear that streaming trends were influenced by unemployment in 2020.
- While there are interesting changes in genre popularity, it is not possible to correlate those changes to unemployment or key events in 2020.

## Lessons Learned

- Spend more time on project organization early to ensure a smooth final merge (e.g. naming conventions)
- Perhaps assess data availability before designing an amazing project we all love and then can't really do
- Consistency with datasets is key; difficult to assess trends or correlations with datasets that differ in time periods covered