r-data-better-than-yours

Stella Traylor, Sam Boyce, Lexi Field



Topic: What Makes a Hit Song?

Research Question:

How do song tempo and artist popularity influence song popularity, and how do these relationships differ across musical genres and over time?

Motivation:

Music is a major part of our everyday lives, but what actually makes a song popular? We want to explore which factors are most closely associated with a song's success.

Our Data (Better Than Yours)

The Million Song Dataset is a collaboration between the Echo Nest and LabROSA.



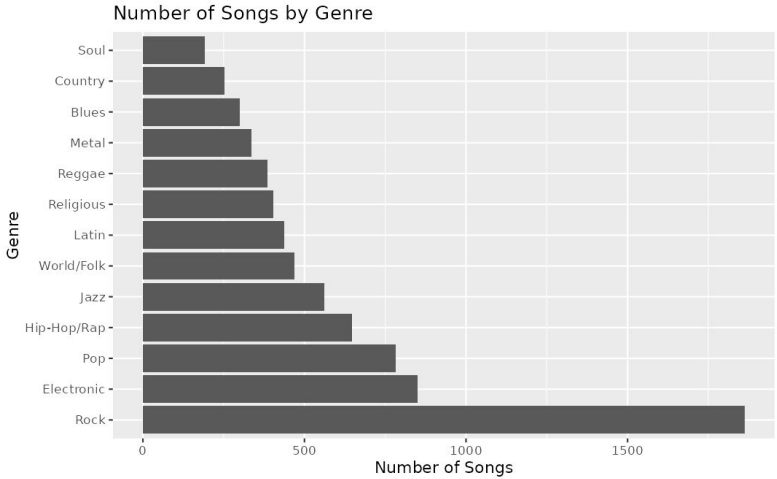
There are 10,000 observations each representing a song, and 35 variables.

The variables most relevant to our proposed analysis are: artist_familiarity, artist_popularity, song_popularity, artist, genre, tempo, and year.

Genre: Cleaning and Grouping

```
music <- music |>
    mutate(
        genre = case_when(
            artist.terms %in%
                c(
                    "blues-rock",
                    "country rock",
                    "post-grunge",
                    "pop rock",
                    "rock",
                    "hardcore punk",
                    "post-hardcore",
                    "dance rock",
                    "hard rock",
                    "folk rock",
                    "soft rock",
                    "experimental rock",
                    "grunge",
                    "new wave",
                    "indie rock",
                    "glam rock",
                    "modern rock",
                    "rock 'n roll",
                    "oi",
                    "heartland rock",
                    "roots rock",
                    "ska punk",
                    "alternative rock",
                    "art rock",
                    "pop punk",
                    "instrumental rock",
                    "freakbeat",
                    "southern rock",
                    "rockabilly"
                ) ~ "Rock",
```

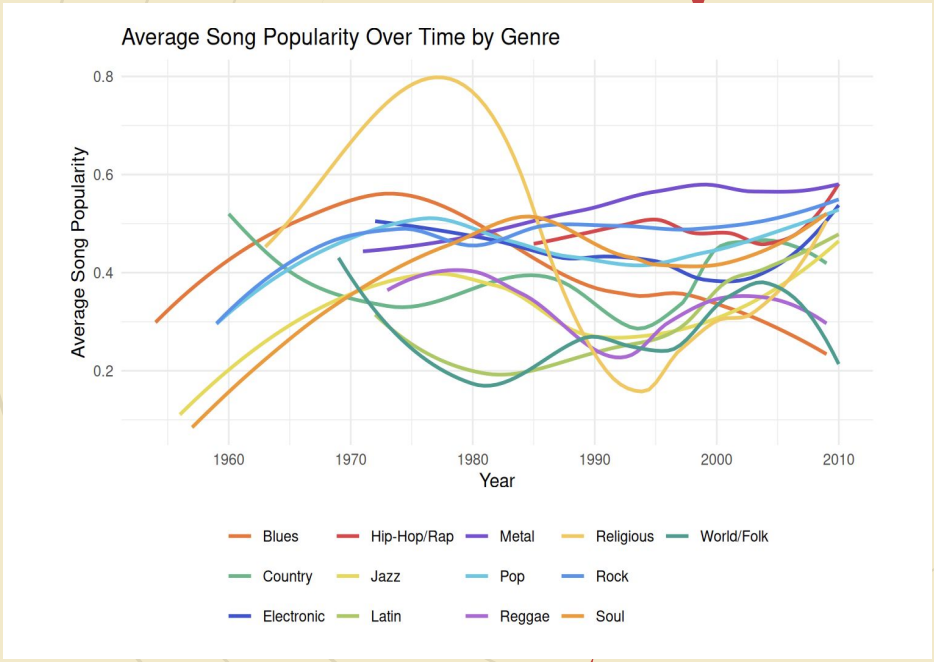
We grouped the artist.terms variable (the term most associated with each artist) into 13 broader genre categories to make comparisons easier and renamed to genre



- The dataset is unevenly distributed across genres, so certain correlations may represent certain genres more strongly than others
  - Rock is the most represented
  - Soul is the least represented

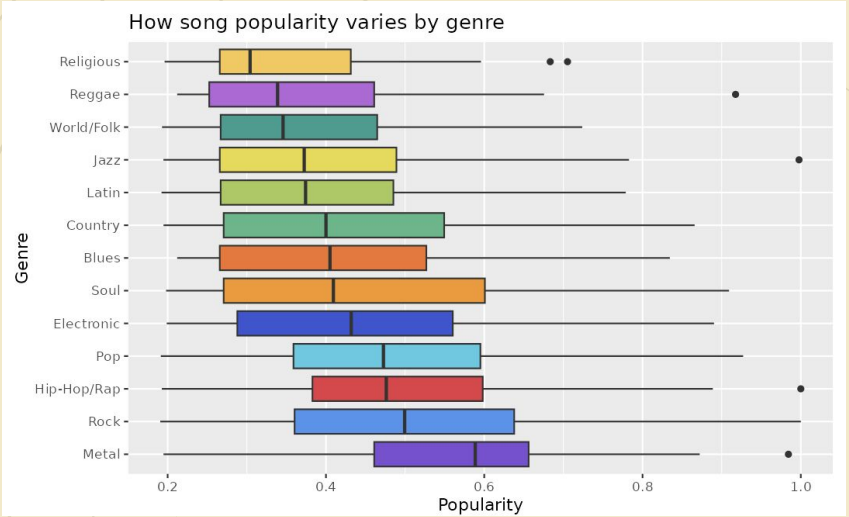
Average Song Popularity Over Time By Genre

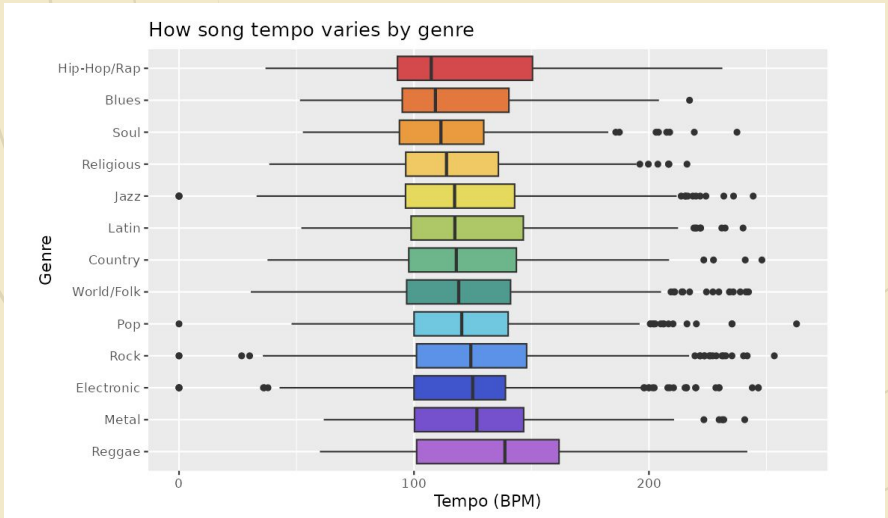
- ★ Interested to see how genre affects song popularity
- ★ Song popularity fluctuates over time, and Religious music shows the most fluctuations
- ★ Shows how song popularity varies across genres and release years, revealing whether popularity patterns change over time



How does song popularity and song tempo vary by genre?

- ★ Popularity varies more by genre than tempo
- ★ Median tempos are fairly similar, mostly around 110-130 BPM
- ★ The distributions overlap substantially, so genre alone does not clearly distinguish a song's popularity or tempo
- ★ Outliers: many genres also have unusually fast songs

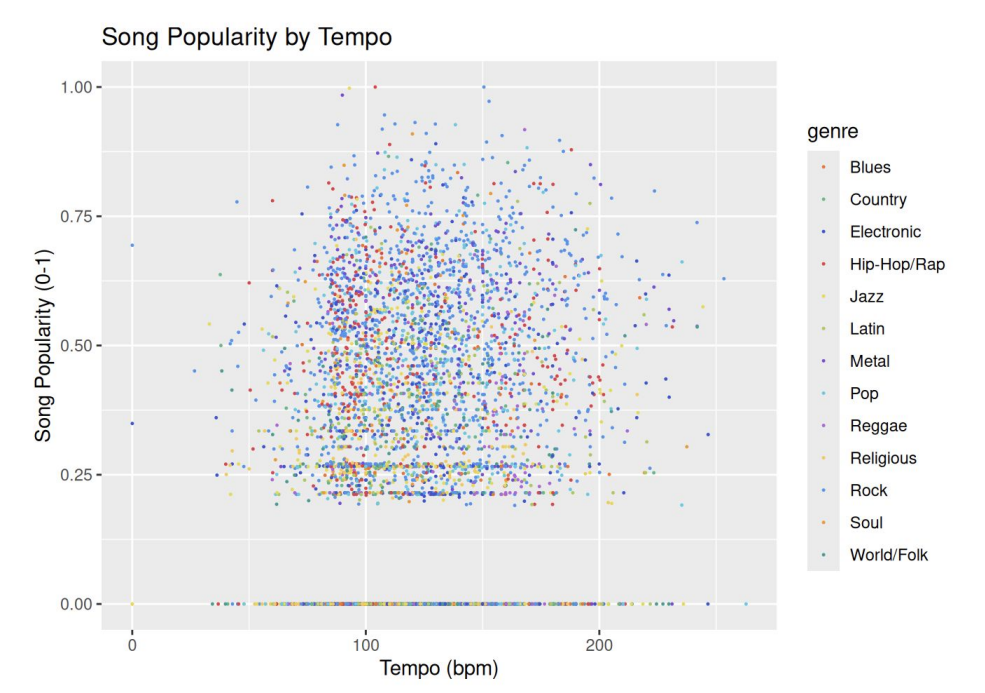




Song Popularity vs. Tempo

	r.squared	adj.r.squared
	<dbl>	<dbl>
1	0.00536	0.00513

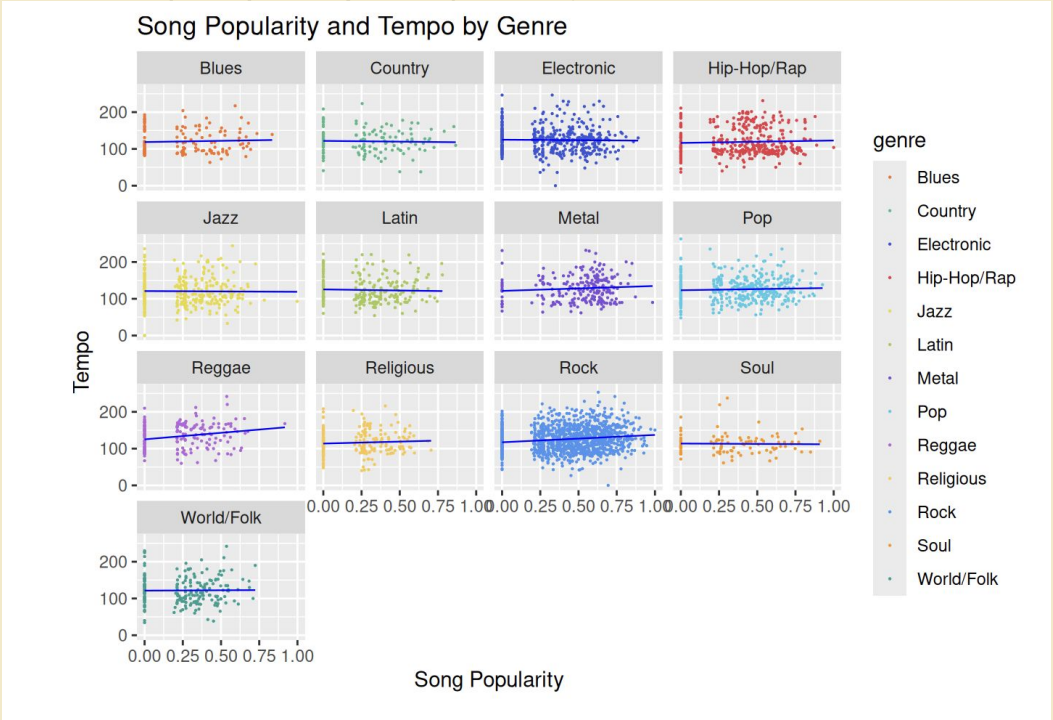
- ★ No clear relationship between song tempo and popularity
- ★ Most songs clustered between 90-160 BPM, popularity ranging in this region from ~0.2 - 0.8.
- ★ R^2 = 0.5% → tempo alone is not a strong predictor of popularity



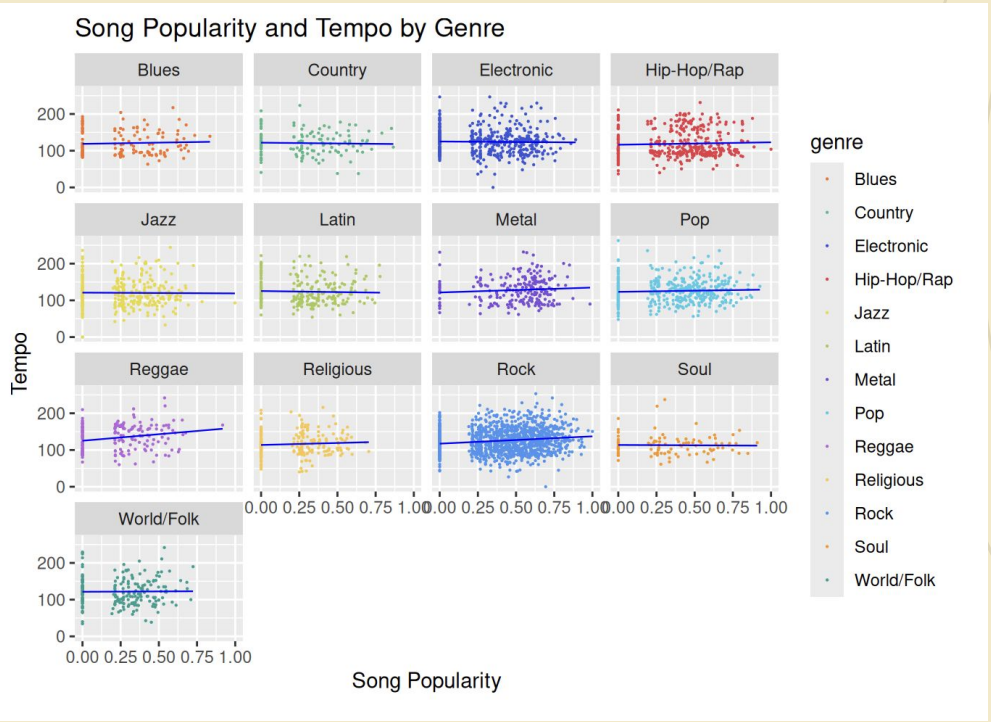
By Genre

- Tempo and song popularity have weak relationships across genres, as shown by flat regression lines.

- Reggae, Metal, & Rock have positive relationships - showing that faster songs are somewhat associated with popularity in these genres.



Linear Regression By Genre



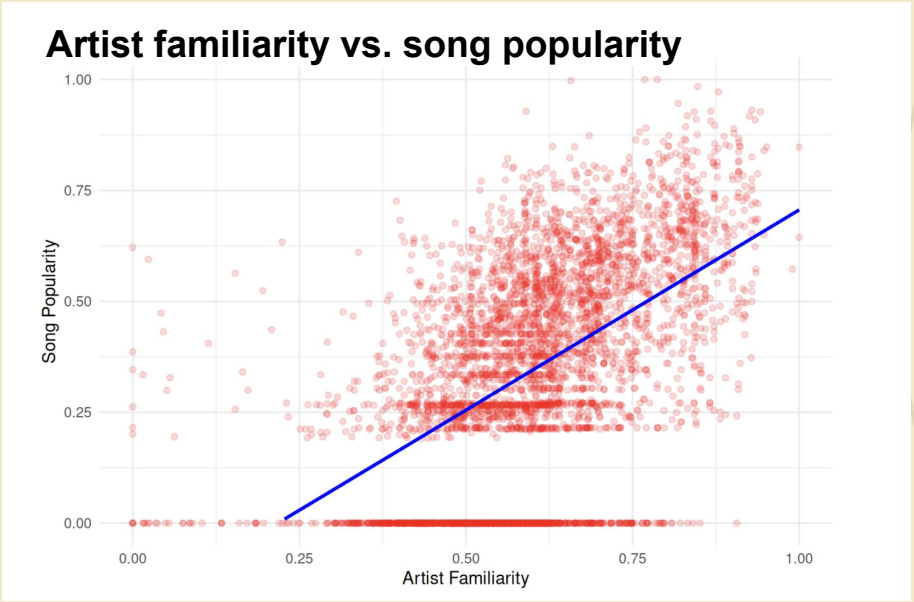
Tempo on popularity

```
+ glance(tempo_on_pop)
+
# A tibble: 1 × 12
  r.squared adj.r.squared
      <dbl>         <dbl>
1   0.00536       0.00513
```

Tempo on popularity by genre

```
+ genre_reg
# A tibble: 13 × 13
# Groups:   genre [13]
   genre      r.squared a
   <chr>          <dbl>
 1 Blues        0.00227
 2 Country      0.000873
 3 Electron…    0.000373
 4 Hip-Hop/…    0.00172
 5 Jazz         0.000136
 6 Latin        0.00122
 7 Metal        0.00671
 8 Pop          0.00221
 9 Reggae       0.0472
10 Religious    0.00463
11 Rock         0.0207
12 Soul         0.000326
13 World/Fo…    0.000104
```

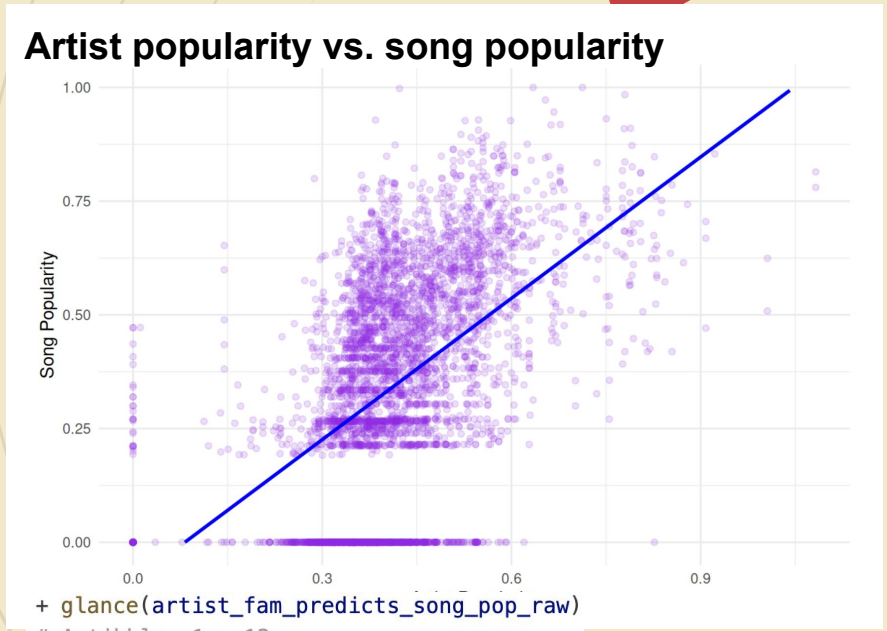
Influence of artist on song popularity




```
+ glance(artist_pop_predicts_song_pop_raw)
# A tibble: 1 × 12
```

r.squared	adj.r.squared
<dbl>	<dbl>
0.0553	0.0552

	r.squared	adj.r.squared
	<dbl>	<dbl>
1	0.227	0.226




```
+ glance(artist_fam_predicts_song_pop_raw)
# A tibble: 1 × 12
```

	r.squared	adj.r.squared
	<dbl>	<dbl>
1	0.0688	0.0687

r.squared	adj.r.squared
<dbl>	<dbl>
0.250	0.250

What did we learn?

- Originally predicted that faster tempos would be more popular, but we found out there's not a strong correlation between tempo and song popularity.
  - In some genres (Rock and Metal), higher tempo songs are more popular
- **Artist fame** explains the variability in song popularity more than tempo
- Which genres have the highest tempos (surprising that Hip Hop and Rap have the slowest median tempos)
- How song popularity across genres **varies over time** (Religious music fluctuating the most was interesting)

Critiques

- ★ Shifted our focus away from tempo to other variables after calculating the R value because tempo barely affected song popularity
- ★ Our genre groupings were somewhat arbitrary and each group had a different amount of songs
- ★ Controlling for more variables could have helped us isolate what really makes a song grow in popularity
- ★ Erroneous values in the dataset → had to filter many songs out as a result.

Thank You!