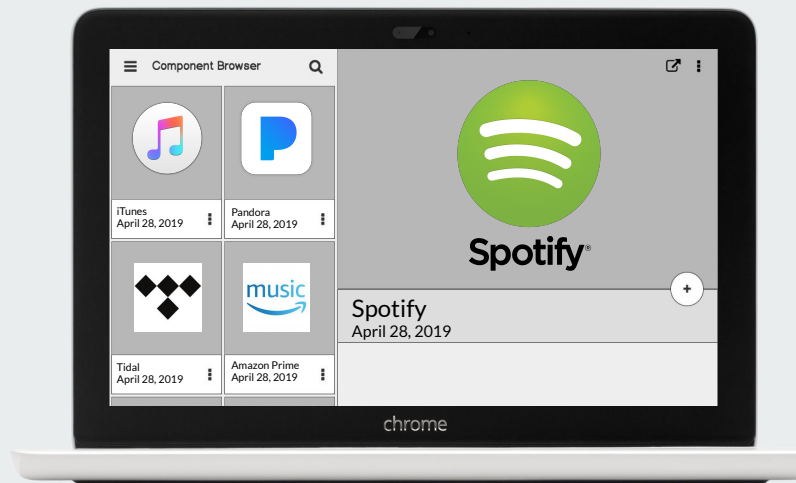# Journey to a Song's True Value

Natalie Austin, Victoria Cardenas, Yilin Chen, Zehui Jin, Alan Koolik

Data

- We used a 10,000 song subset of the Million Song Dataset.
- Once cleaned, we were left with around 4,300 songs
- Each song identified by 21 unique features

Artist Features:

Name, Hotness,
Location, Grouping
(low, medium, or high
hotness)

Song:

Name, Year, Genre,
Decade, Hotness

Musicality:

Duration, Key, Loudness,
Mode, Mode Confidence
Interval, Start of Fade
Out, End of Fade In,
Tempo, Tempo Group,
Time Signature, Time
Signature C.I

Song vs. Artist

- Distribution of song hotness is different among artists of various levels of popularity
  - Higher Artist.Hotness → Higher song hotness

- Indicates Artist.Hotness is an important feature for Song.hotness

Low. Medium. High.
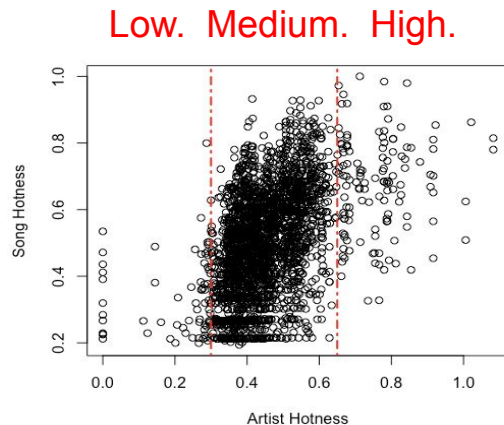
Song Hotness Distrubution by Artist Hotness

Low: Artist.Hotness<0.3
Medium: 0.2<=Artist Hotness<0.65
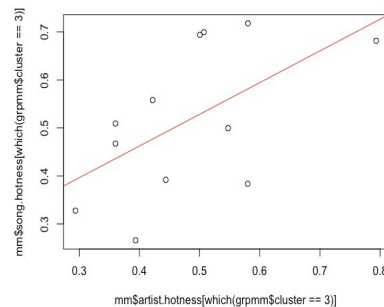High: Artist.Hotness>=0.65
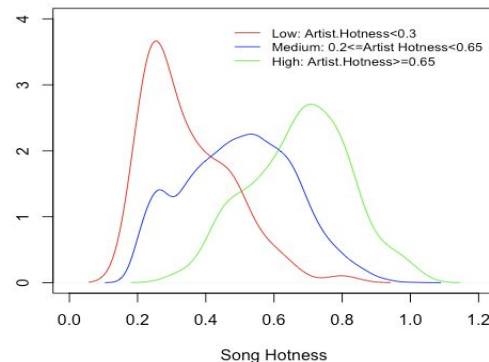
Song Hotness

- K-means cluster purely by songs attribute
- Examine Artist.hotness vs Song.hotness in each cluster
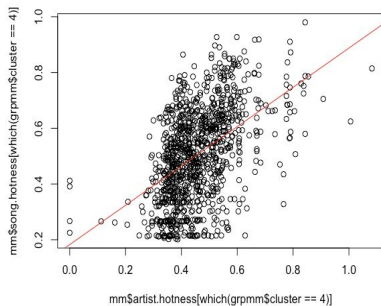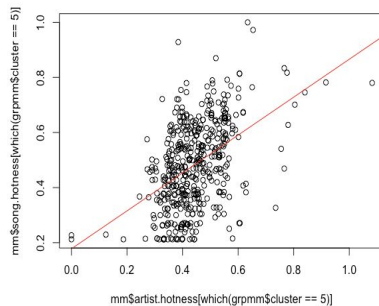- Regression of Song.hotness ~ Artist.hotness, p-value < 2e-16
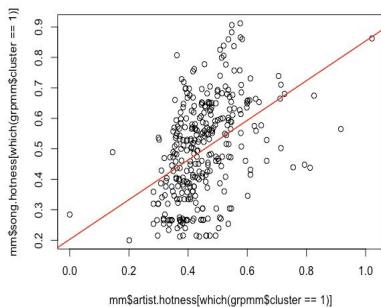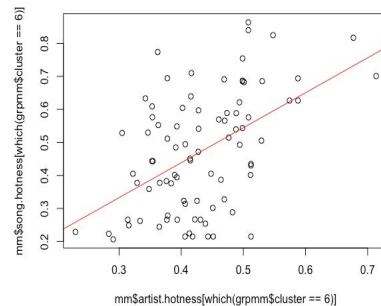

$\rho_1 = 0.4639$
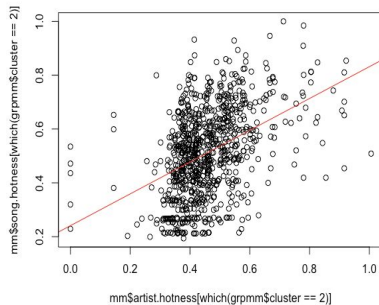
$\rho_2 = 0.4274$

$\rho_3 = 0.5690$

$\rho_4 = 0.4861$

$\rho_5 = 0.4622$

$\rho_6 = 0.4957$

Conclusion 1:

With an adjusted r-squared value of 0.22, the popularity of an artist is a relatively good indicator of the popularity of a song. The more popular an artist is, the more likely his/her song will be a hit.

Song.hotness vs Song's attributes

- Split data into low, medium, high classes by artist.hotness
- Apply PCA in the medium group to see if there was a clear hierarchy in Song.hotness.
- First component: duration, …
- Second component: mode, end_of_fade_in, …

PCA
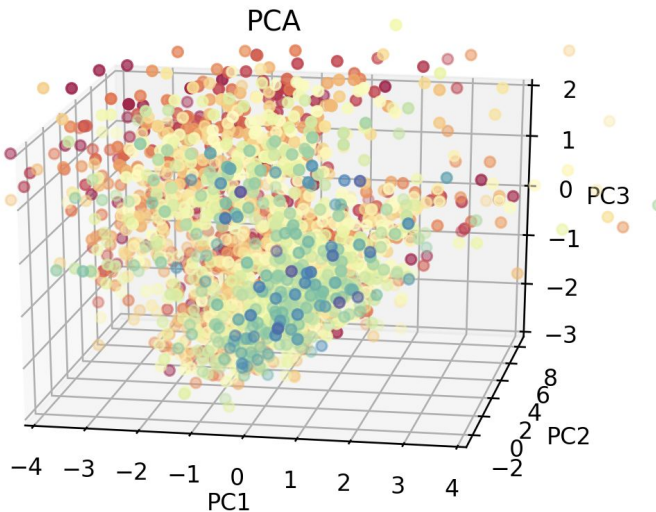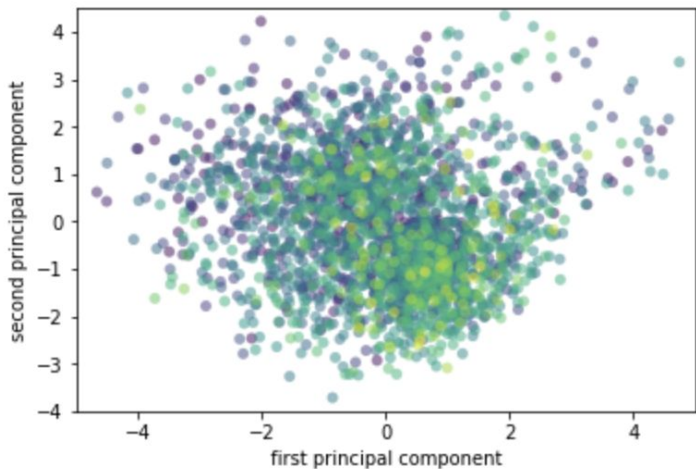
Independent of Artist, Song vs Features

- When considering medium artist.hotness, a simple regression comes up with many of the same significant features as PCA
  - Duration, loudness, and genre (metal, pop, hip-hop, punk, r&b, rock)
- Adjusted R-squared is 0.14

Song Hotness vs Duration

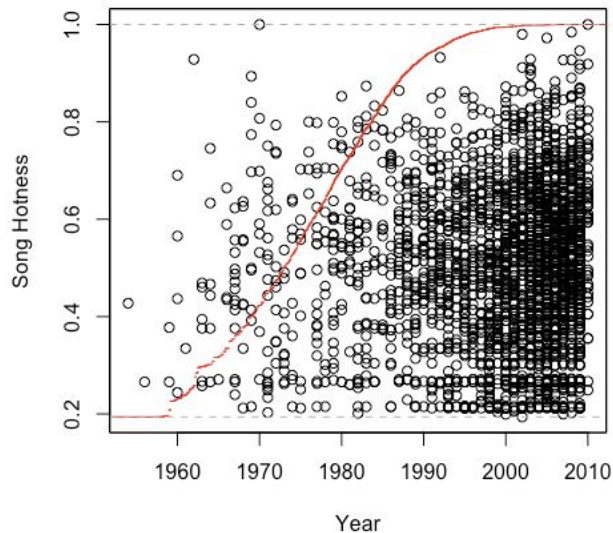**Song Hotness vs Loudness**

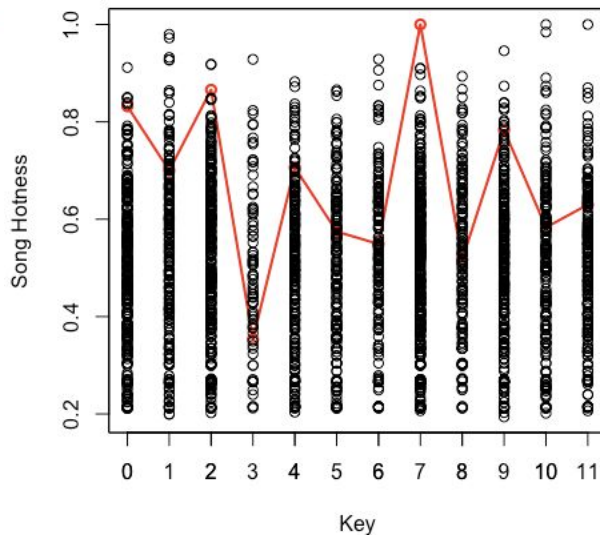**Song Hotness by Genre**
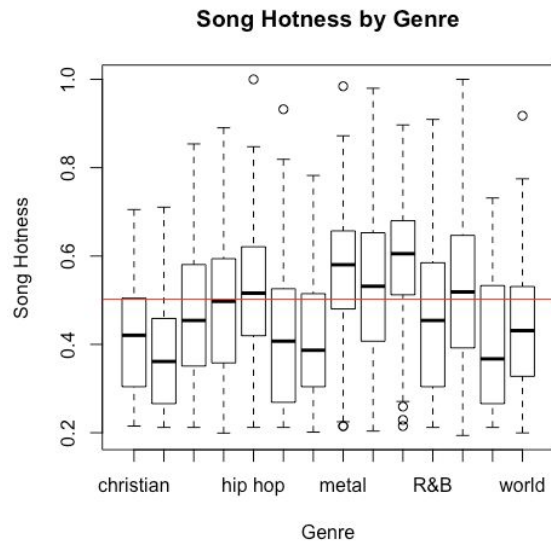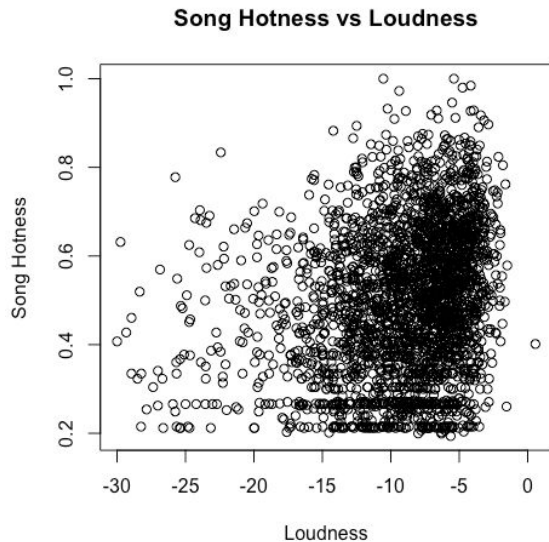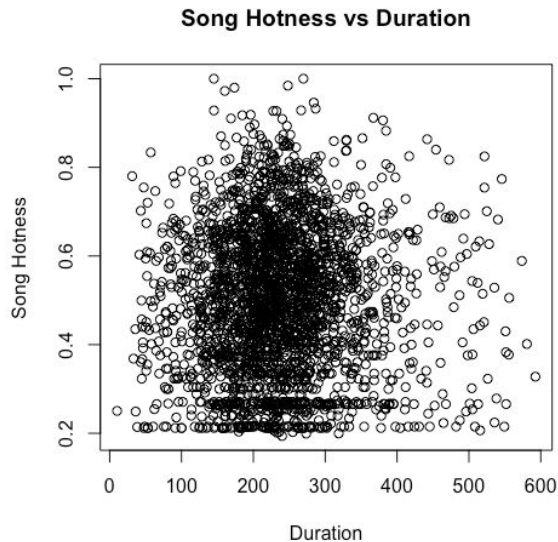
Conclusion 2:
Certain song features are important components that contribute to a song's "hotness", but they do not fully explain what makes a song popular.
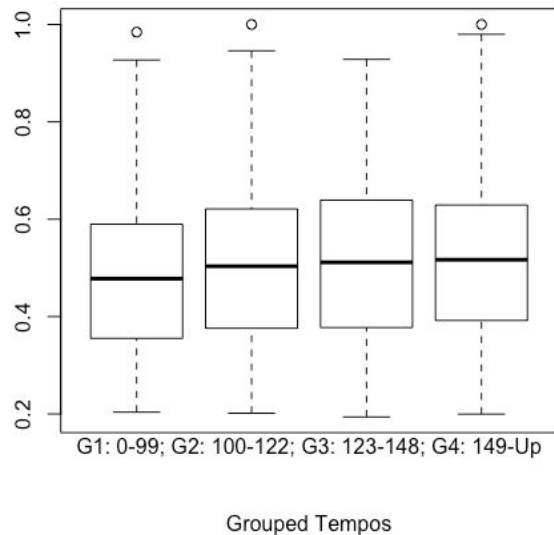
Other Things to Look At

Song Hotness by Year

Song Hotness by Key

Song Hotness by Grouped Tempos

G1: 0-99; G2: 100-122; G3: 123-148; G4: 149-Up

Limitations

- Data available does not provide a complete representation of a song
  - No data on the lyrics
    - Topics of lyrics might affect the popularity of a song
    - Catchy lyrics add value; regardless if they're good or bad
  - No information on the performing artist's label/manager
    - Record labels, in the past more so than now, could push a song
  - No data on the marketing effort for each song
    - Publicity campaigns can vary widely for each song

Major Takeaways

In trying to make a song popular...

- Boost an artist's profile
  - Social media presence, collaborations with better known artists

- Choose a song that's...
  - Not over 400 seconds
  - Of greater than -10 dB in loudness
  - Of genre punk, metal, pop

- And attach it to a notable artist