# Quantitative Methods of Evaluating Song Lyrics

**Timothy Mitchell**
**May 2020**

## Data Collection → Data Processing → Data Analysis

**Data Collection**

- Web scraping to extract data
- Importing data into statistical software

**Data Processing**

- Preprocessing
- Natural language processing

**Data Analysis**

- Plots and figures
- Statistical tests
- Models and Algorithms

# METRO**LYRICS**

Top Songs    Videos    Features    Artists

# Queen - Bohemian Rhapsody Lyrics

**QUEEN**

print lyrics

edit lyrics

**Popular Right Now**

**Bloody Valentine**
Machine Gun Kelly

**Together**
Sia

**Deen Assalam**
Sabyan Gambus

**Beautiful In White**
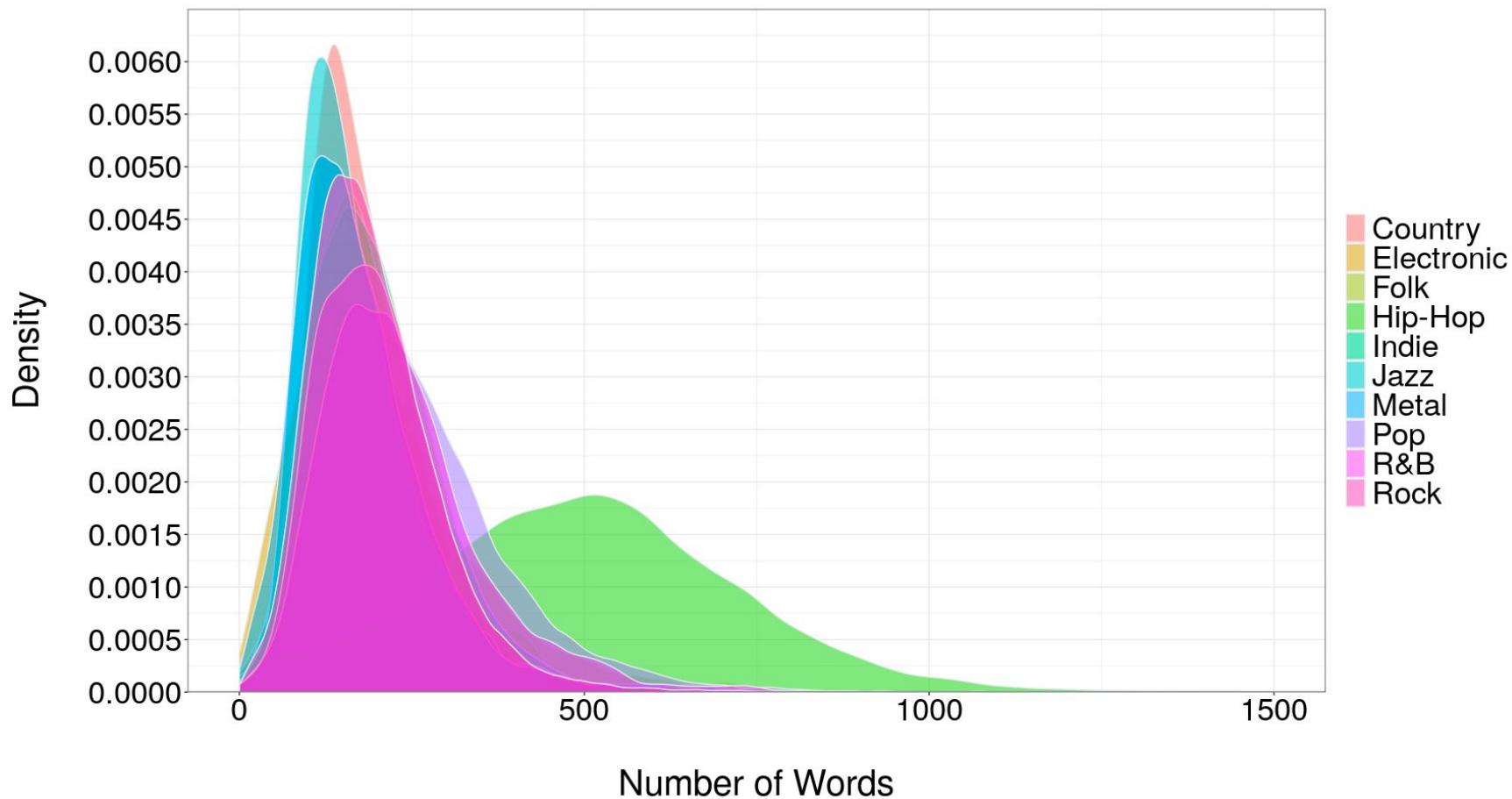Westlife

**In My Feelings**
Drake

Is this the real life, is this just fantasy?
Caught in a landslide no escape from reality
Open your eyes look up to the skies and see
I'm just a poor boy, I need no sympathy

Because I'm easy come, easy go
A little high, little low
Anyway the wind blows
Doesn't really matter to me, to me

Mama, just killed a man
Put a gun against his head
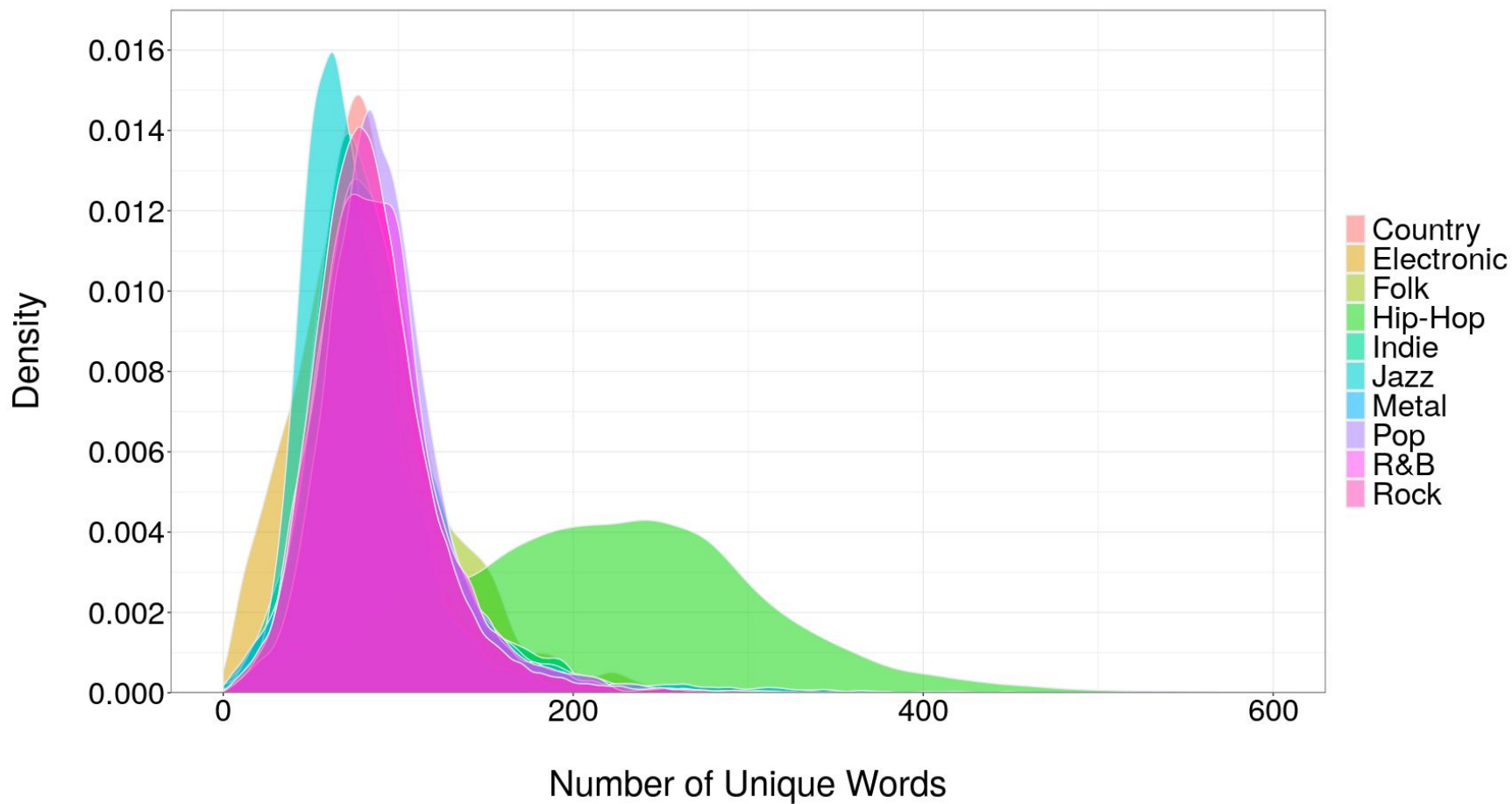Pulled my trigger, now he's dead
Mama, life had just begun

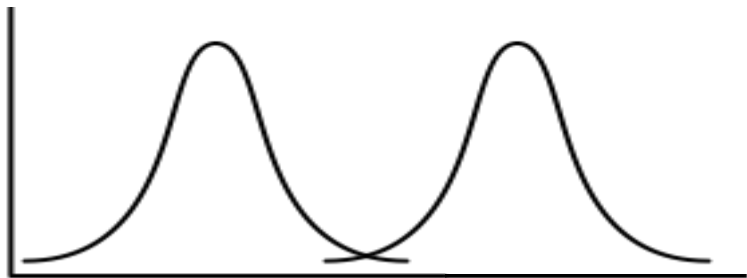Total Words per Song, By Genre (Scaled)

Unique Words per Song, By Genre (Scaled)

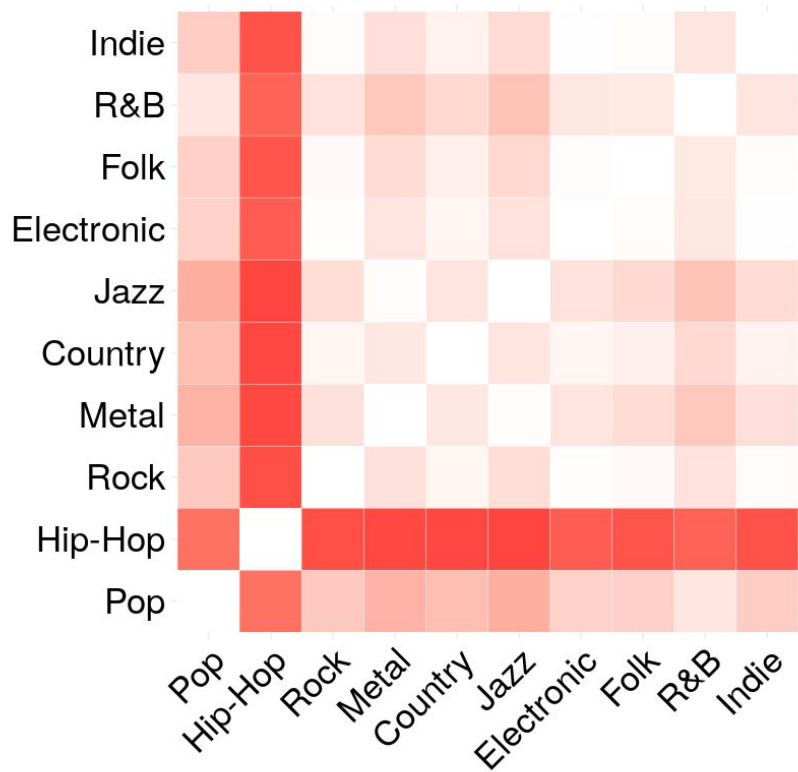**Cliff's delta** (**δ**): measures overlap between 2 distributions
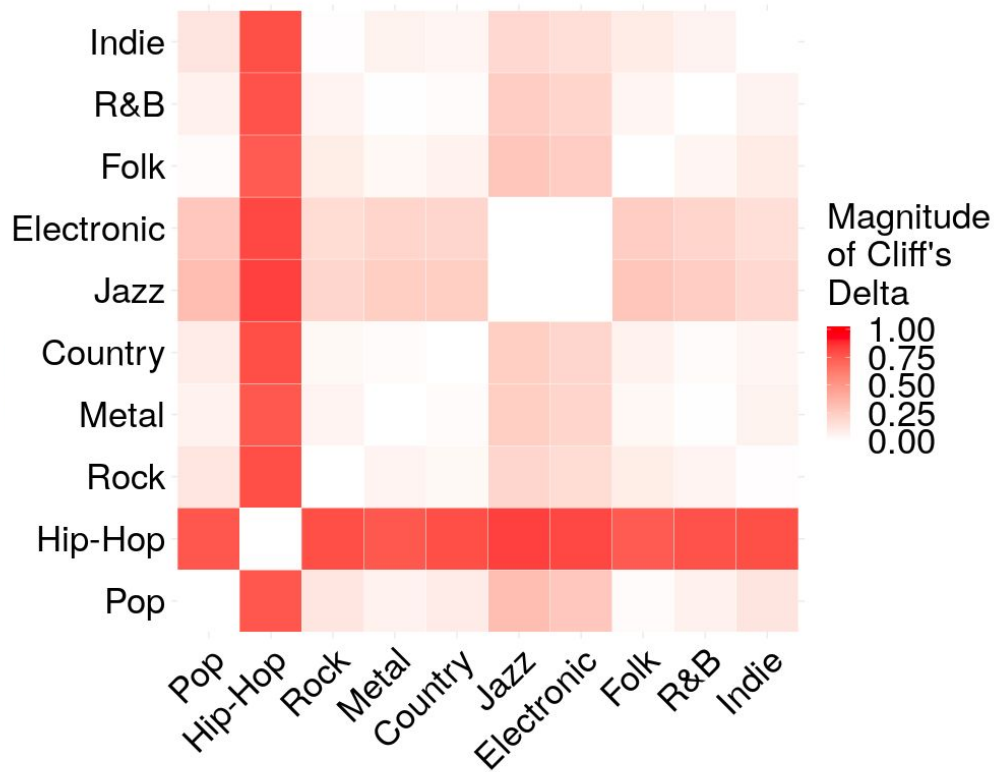
**Less overlap:**                    **More overlap:**

Difference in Number of Words by Genre

Difference in Number of Unique Words by Genre

Magnitude of Cliff's Delta
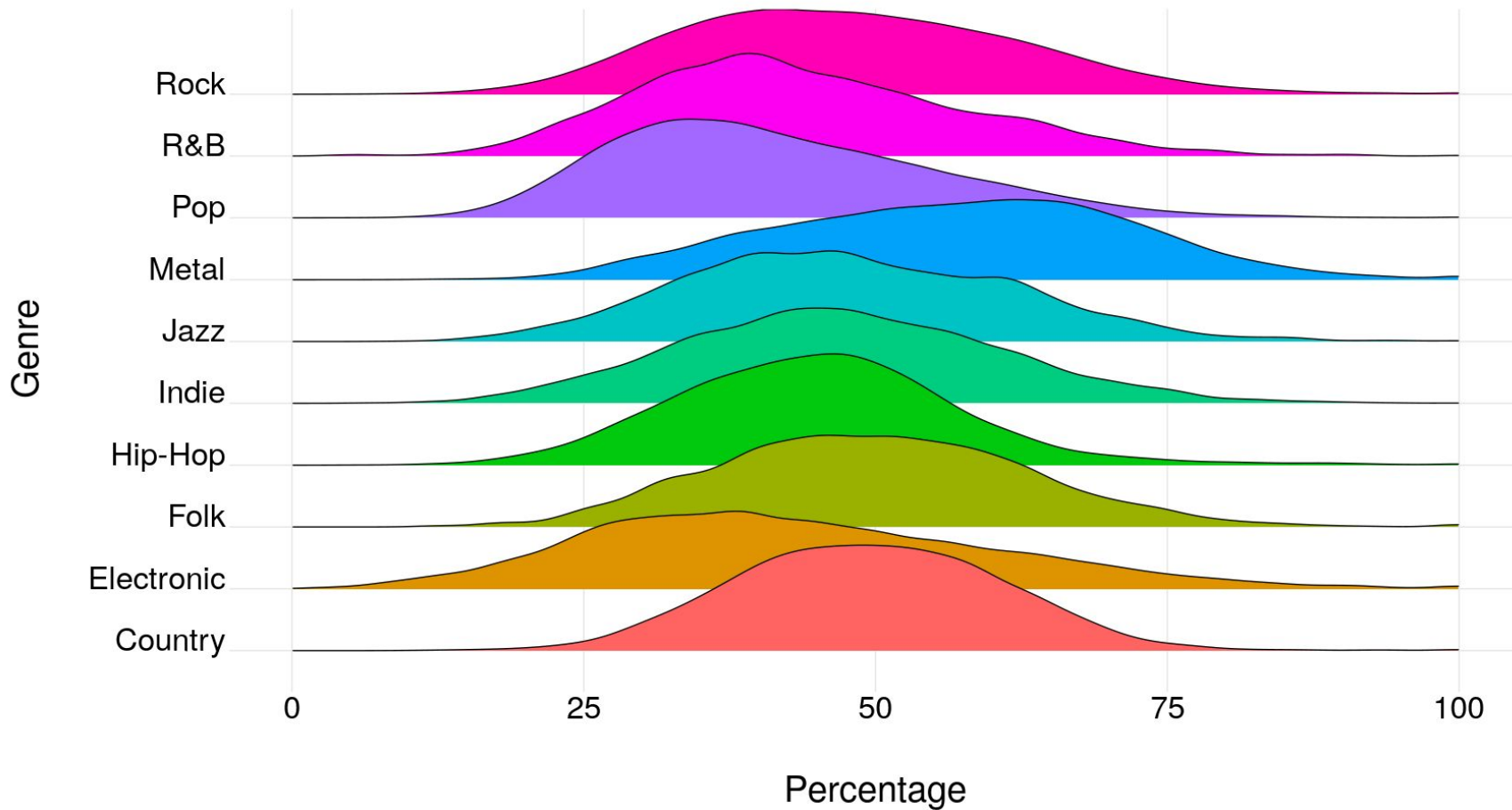1.00
0.75
0.50
0.25
0.00

The **Kruskal–Wallis analysis of variance test** indicated a significant effect of **genre** on **total words**, $p < 2.2 \times 10^{-16}$, and $\varepsilon^2 = 0.347$, a relatively strong effect size

Similarly, I found a significant effect of **genre** on **unique words**, $p < 2.2 \times 10^{-16}$, and $\varepsilon^2 = 0.390$, a relatively strong effect size
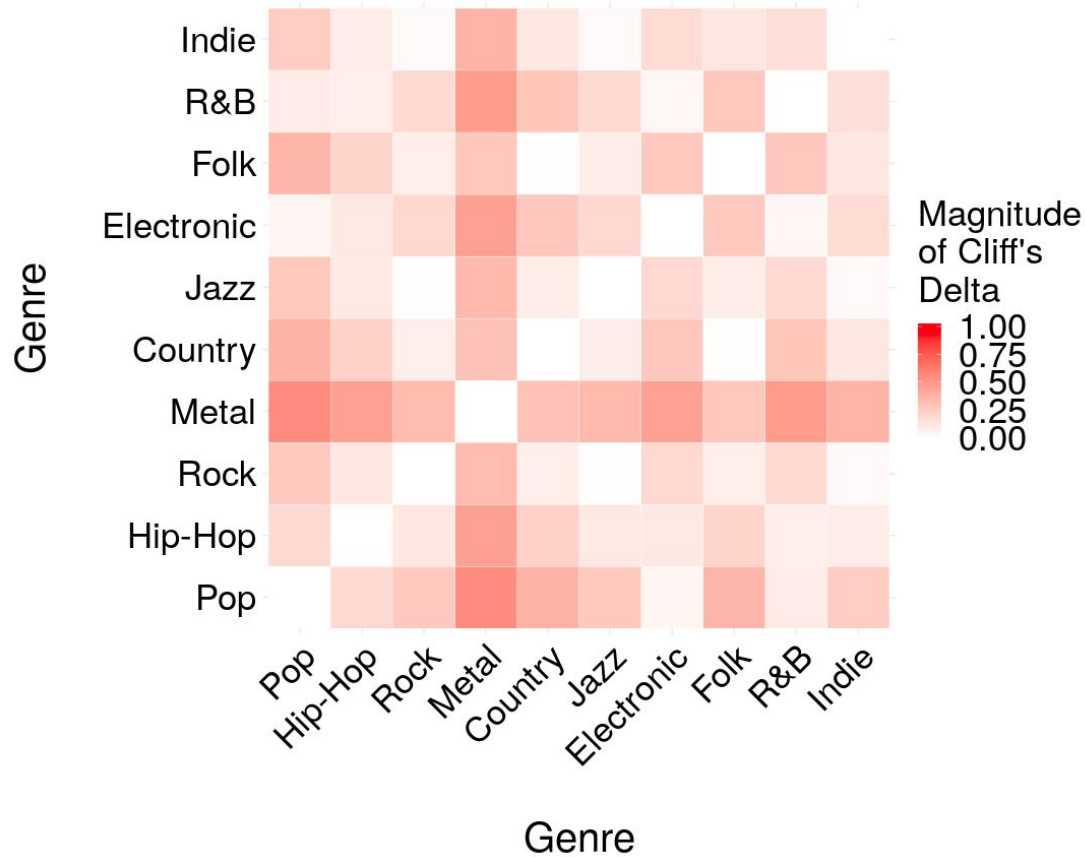
Results of the **pairwise Wilcoxon rank-sum test** confirmed a significant difference in the mean number of total and unique words between Hip-Hop and other genres

# Percentage of Words That Are Unique, By Genre

Difference in Percentage of Words that are Unique by Genre

|  | Type of Consonance | Example |
|---|---|---|
| 1 | **Doublet** | Donald Duck |
| 2 | **Triplet** | Big Brown Bear |
| 3 | **Quartet** | Happy Handsome Healthy Horse |
| 4 | **2 singlets** | Lots of Love |
| 5 | **3 singlets** | Why not Waste a Weekend |
| 6 | **4 singlets** | Back when Burning Man was Blacks in Birmingham |
| 7 | **Doublet with 1 singlet** | Bed Bath and Beyond |
| 8 | **Doublet with 2 singlets** | Stood Still and Stopped the Sound |
| 9 | **Triplet with 1 singlet** | Betty Botter Bought some Butter |
| 10 | **Complex** (≥ 5 of a letter) | Peter Piper Picked a Peck of Pickled Peppers |

Broken barriers bounded by the bomb beat

Buildings are broken, basically I'm bombarding

Casually create catastrophes, casualties

Canceling cats got their canopies collapsing
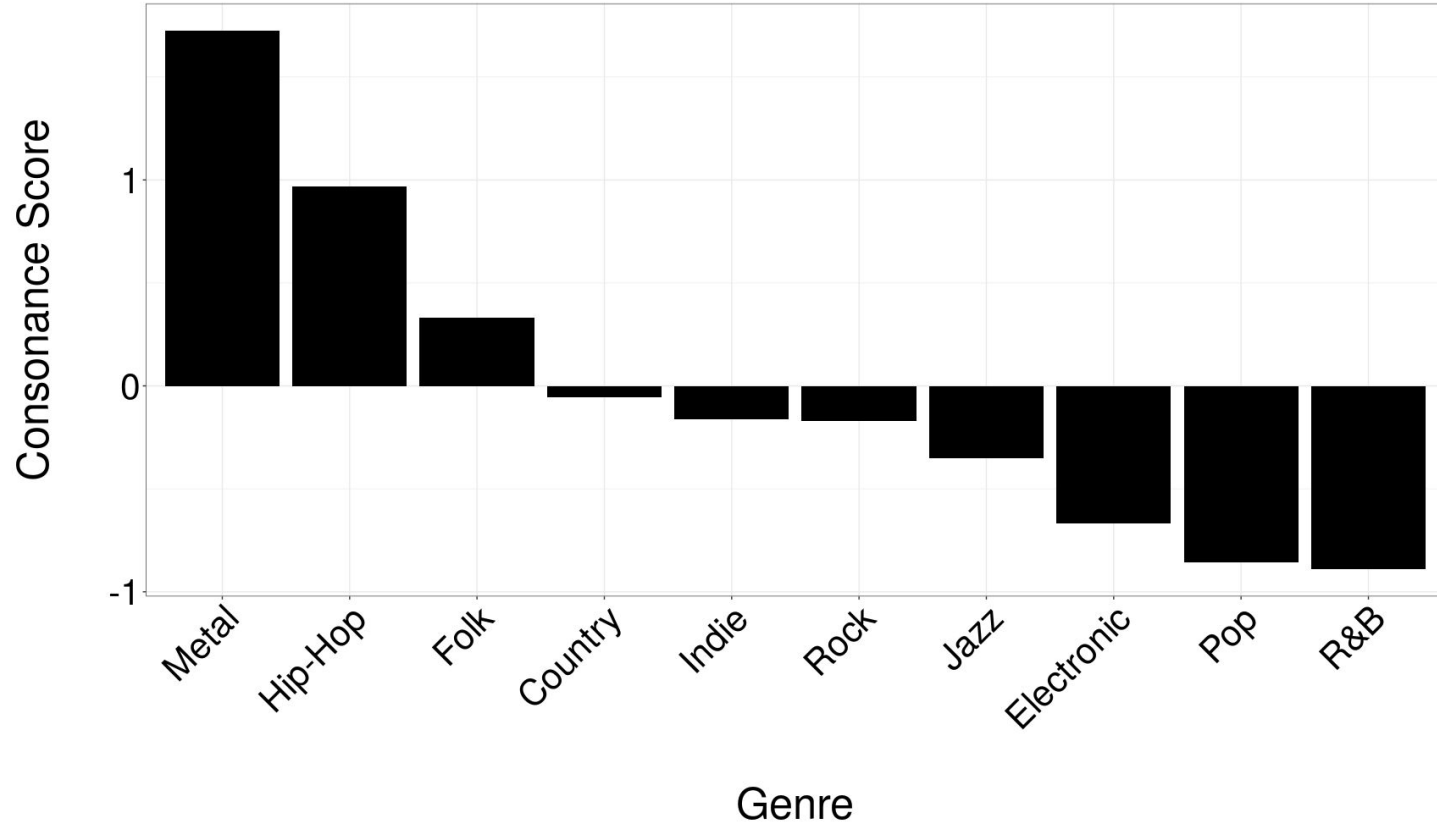
} *Alphabet Aerobics*, Blackalicious

A alligators all around

B bursting balloons

C catching colds

D doing dishes

} *Alligators All Around*, Carole King

Consonance Score, By Genre
Consonance Score was Calculated by Adding Together
Scaled Measures of 10 Different Kinds of Consonance

- clawed fingers forcing
- man made monster morbid
- sedative substances

} *Experimental Stage*, Benediction

- hearing echoes from your heart
- recalling retreating returning retrieving
- a leader a learner a lawful beginner
  a lodger of lunacy so lucid
- rumbling thunder rolling

} *Rebirth*, Angra

- wind whipped foams warlords
- braves brightly burn berserks
- memories invade my mind
- brave friends battle their flashing blades

} *North Sea Storm*, Amon Amarth

# Questions?

Email: tmitchell8@luc.edu

# Appendix

1. Tokenize lyrics
2. Extract the initial letter of each word
3. Scan lyrics for clusters of words with > 1 appearance of a consonant and < 7 words between successive consonants
4. Bypass stop words and duplicate words during the search, to ignore repetitive motifs like "dance, dance, dance"
5. Analyze frequencies for 10 types of consonance
6. Weight frequencies by number of words per song
7. Pool frequencies for the letters b, d, f, h, l, m, n, p, r, v
8. Calculate row sums of the centered and scaled data matrix to obtain a **consonance score** for each song