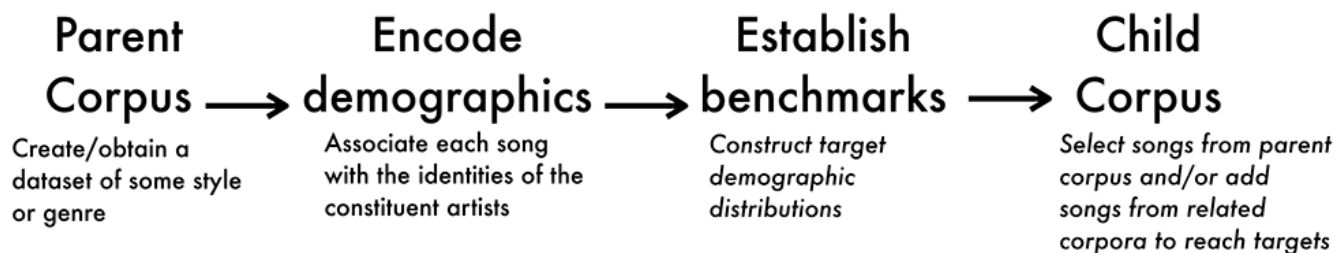A JOURNAL OF THE SOCIETY FOR MUSIC THEORY

MTO 30.1 Examples: Shea et al., Diversity in Music Corpus Studies

(Note: audio, video, and other interactive examples are only available online)
https://mtosmt.org/issues/mto.24.30.1/mto.24.30.1.shea_et_al.html

Example 1. A schematic of the Anti-Discriminatory Alignment System

Parent Corpus → Encode demographics → Establish benchmarks → Child Corpus

Create/obtain a dataset of some style or genre

Associate each song with the identities of the constituent artists

Construct target demographic distributions

Select songs from parent corpus and/or add songs from related corpora to reach targets

Example 2. Sample demographic analyses for top-ranking artists in *Rolling Stone Magazine*'s "500 Greatest Songs of All Time" lists (2004 and 2021 versions) marginalized by race, ethnicity, and or gender

| Artist	Song name	Version	Rank	Year	Nonmale	BIHAP	Primary
Aretha Franklin	Respect	2021	1	1967	1	1	1
Bob Dylan	Like a Rolling Stone	2004	1	1965	0	0	0
Public Enemy	Fight the Power	2021	2	1989	0	1	1
The Rolling Stones	Satisfaction	2004	2	1965	0	0	0

Example 3. Distribution of groups with at least one BIHAP member, with at least one nonmale member, and who yielded a primary status member
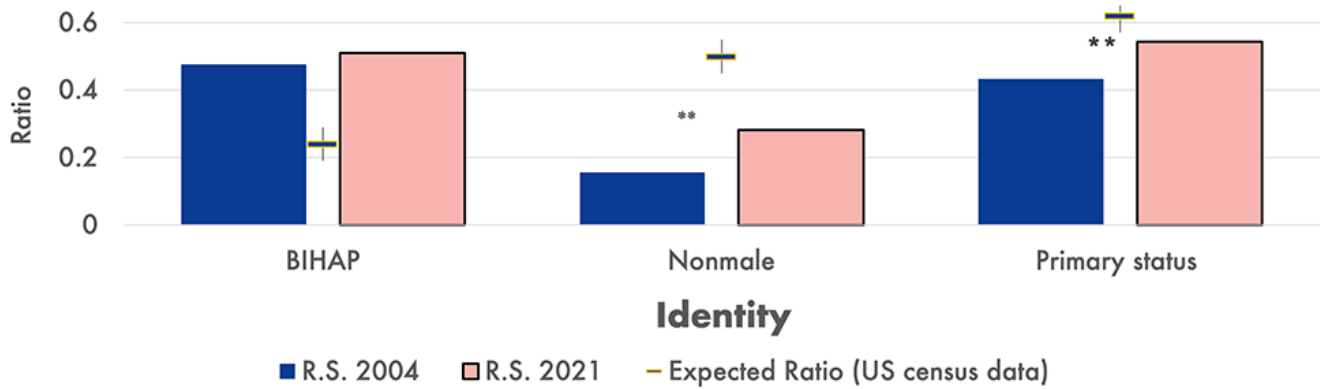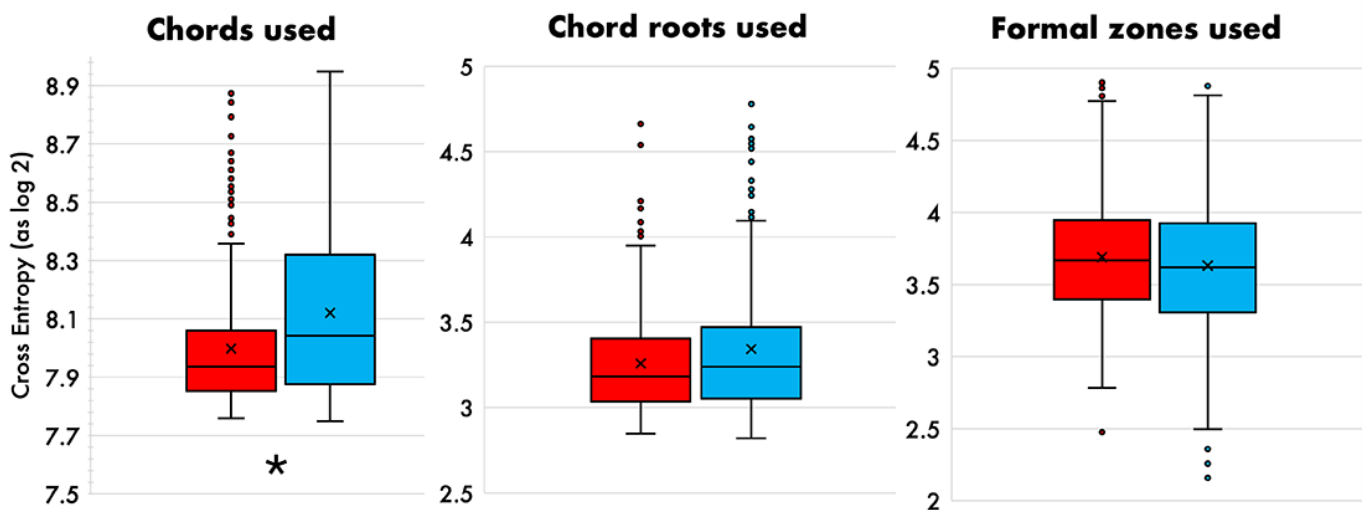

Example 4. Song counts per half decade a) for BIHAP groups in the 2004 and 2021 lists, and b) for gender groupings in the 2004 and 2021 lists
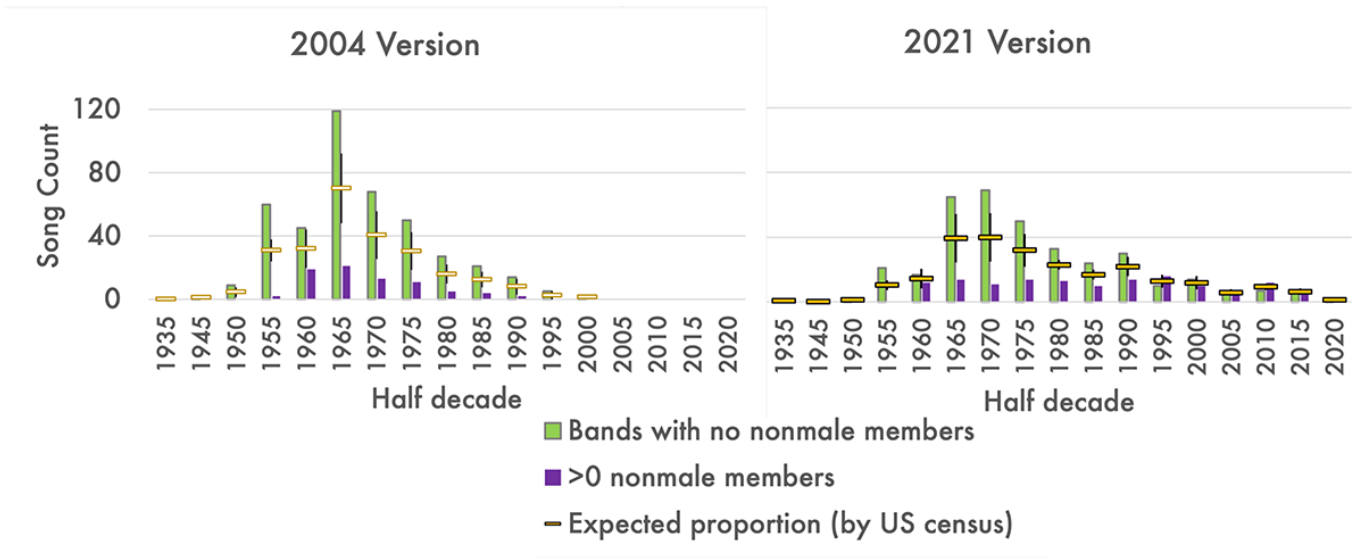
Example 5. The ratio of groups with at least one BIHAP and nonmale member; $n=500$ for each full corpus, $n=47$ for post-2004 RS data
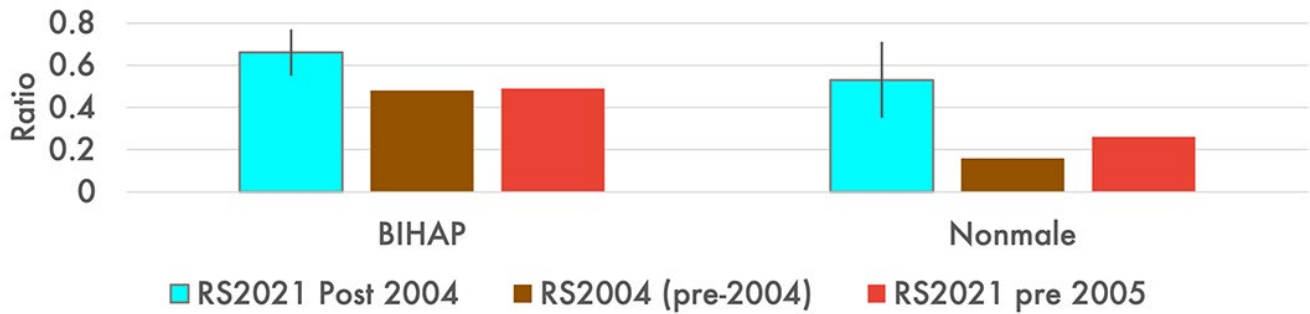
Example 6. The cross-entropy distribution between individual songs and the larger corpus in the all-white-male/not-all-white-male divisions of the McGill Billboard corpus
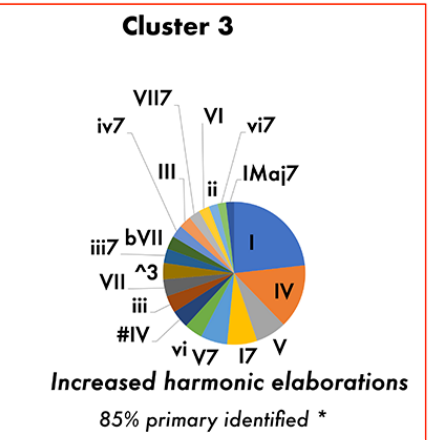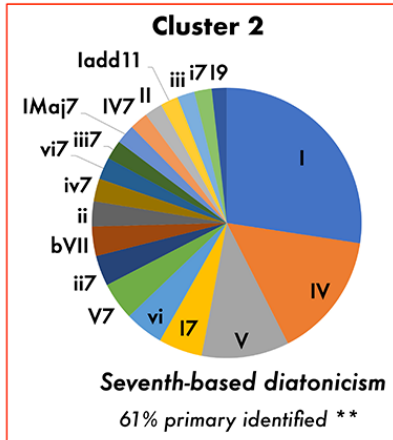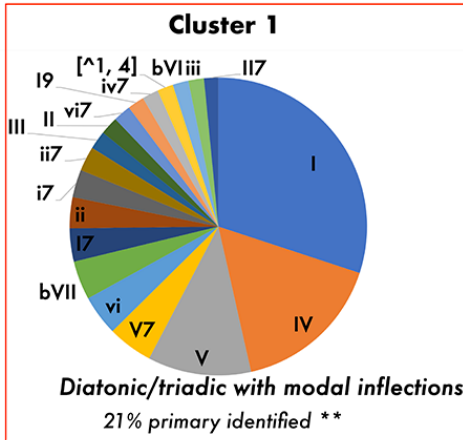
Example 7. The results of a *k*-means clustering on the chord frequencies in each song within the McGill-Billboard corpus, such that *k* = 3

Three harmonic clusters in the McGill-Billboard Corpus

cluster comprises songs by significantly more artists *without* primary designations

clusters comprise songs by significantly more artists *with* primary designations

Cluster 1
Diatonic/triadic with modal inflections
21% primary identified **

Cluster 2
Seventh-based diatonicism
61% primary identified **

Cluster 3
Increased harmonic elaborations
85% primary identified *

Example 8. Sample demographic annotations from each genre

Artist	Song	Year	Genre	Nonmale	Non-cis	Race & Eth.	Primary
Heavy D & the Boyz	Got Me Waiting	1997	Hip-hop	0	0	1	1
Shania Twain	(If You're Not In It For Love) I'm Outta Here!	1996	Country	1	0	0	1
Amorphis	Black Winter Day	1994	Metal	0	0	0	0
Spice Girls	Wannabe	1996	Pop	1	0	1	1

Example 9. An outline of the ADAS process as applied to four popular music genres

Genre	Parent corpora	Supplementary corpora	Percent marginalized Pre-ADAS	Population (estimated from US Census data)
Metal	Pasbani 2017, Podoshen 2017, Hartmann 2018 Loudwire Staff 2020	Billboard "Hot 100", MTV's *Headbangers Ball*, The Official Finnish Charts, Loadedradio 2021, Välimäki 2017	nonmale: 6.7% BIHAP: 28.4%	nonmale: 51.1% BIHAP: 27.7%
Hip Hop	All songs that reached #1 on Billboard's "Hot Rap Songs" chart	Other top-ranking songs from Billboard's "Hot Rap Songs"	nonmale: 22.7% BIHAP: 98.2%	
Country	Top 20 on "Hot Country Songs" from the Billboard Year-End charts	Blessen 2021, Fawale 2020, SongData 2020	nonmale: 20.5% BIHAP: 1.5%	
Pop	Billboard's "Top Songs of the 90s," using songs as classified as "pop" by iTunes	None	nonmale: 60.4% BIHAP: 42.9%	

Example 10. Distribution of prechorus sections in each genre overall; 95% confidence intervals were calculated using binomial distributions; asterisks indicate a difference from the expected average at $p < .05$ (*) and $p < .01$ (**)
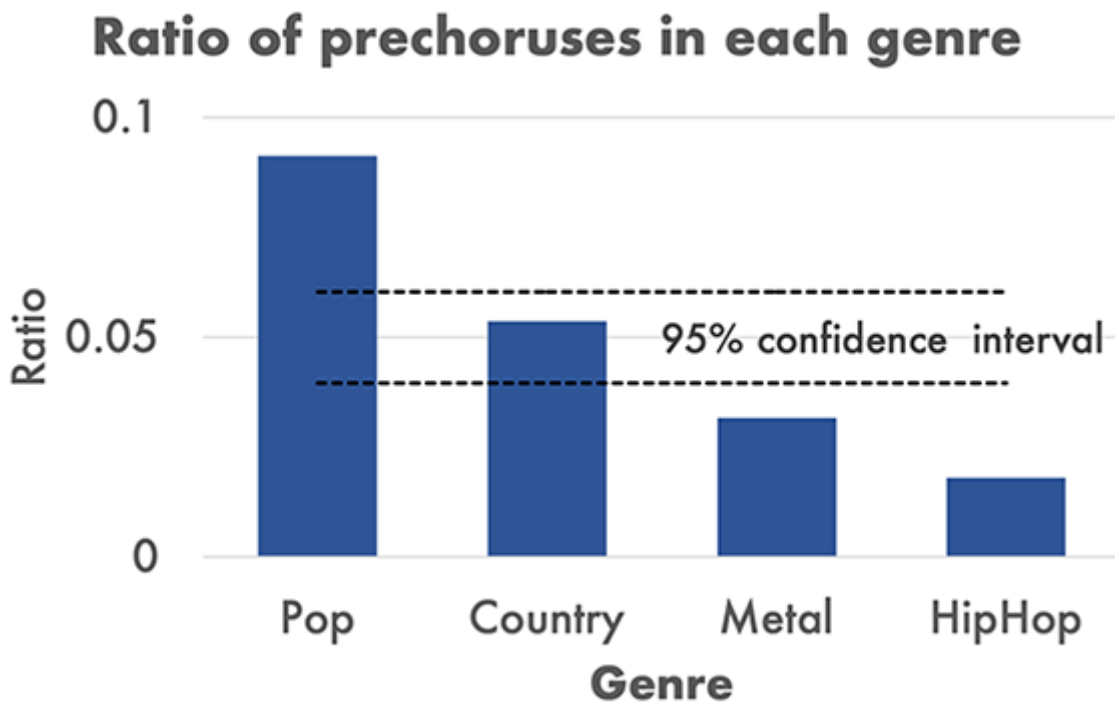
Example 11. Distribution of prechorus sections divided by gender; 95% confidence intervals were calculated using binomial distributions; asterisks indicate a difference from the expected average at $p < .05$ (*) and $p < .01$ (**)

Average number of prechoruses per song