A Demographic Sampling Model and Database for Addressing Racial, Ethnic, and Gender Bias in Popular-music Empirical Research

NICHOLAS J. SHEA [1]
Arizona State University

ABSTRACT: This report summarizes the development and application of a demographic encoding model designed to assist researchers in aligning dataset diversity with real-world diversity in popular-music corpus studies. Drawing on sampling strategies in machine-learning research and encoding procedures in health sciences and the humanities, the model and its associated open-access data provides researchers with a tool to generate more inclusive databases along the parameters of race, ethnicity, and gender. The model itself attempts to reconcile the intersectional boundaries of personal identity with the binarity required by statistical encoding and analysis. Importantly, it facilitates a mindful approach through conditional parameters; for example, by minimizing the risk of tokenizing minoritized artists in multi-member ensembles by considering said artist's agency and demographic proportion within the group. Applying the model to artist samples from various popular-music corpora affirms the underrepresentation of non-white and non-male artists in related research. In response, the report outlines how a researcher might utilize intentional demographic sampling when developing future corpus-based popular-music studies.

Submitted 2021 August 29; accepted 2021 September 15.
Published 2023 August 10; https://doi.org/10.18061/emr.v17i1.8531

KEYWORDS: *popular music, race, ethnicity, gender, sampling, corpus studies*

A core tenet of empirical research is that a robust sample offers a better approximation of tendencies within the broader target population. Despite this, datasets developed for music research do not always reflect real-world population diversity regarding an artist or composer's racial, ethnic, and/or gender identity. Corpora of Western European art music (i.e., "classical" music) demonstrate this with an overt demographic bias toward a small collection of composers who are white and male (e.g., Devaney et al., 2015; Neuwirth et al., 2018). Popular-music corpora the *Rolling Stone 200* (RS200; de Clercq & Temperley, 2011) and *McGill Billboard Hot 100* (MBB; Burgoyne, Wild, & Fujinaga, 2011) meanwhile appear to sidestep this issue by featuring songs by a comparatively diverse contingent of artists. However, data from this study challenge this assumption: applying the demographic model to artist lists from the *RS200*, *MBB*, and a more robust independent sample from *ultimate-guitar.com* (UG; Shea, 2020) demonstrates a continued need for researcher intervention to avoid amplifying real-world biases against non-white and non-male popular-music artists in corpus studies.

Demographic biases can manifest in various ways in music. For example, Kinney (2018) reports that urban secondary schools in the United States are less likely to attract students to music elective offerings than suburban schools. Urban districts historically enroll a greater proportion of non-white students and are also underfunded due to lower support from property taxes (Reschovsky, 2016). As such, urban schools, and resultingly non-white students, are less likely to receive access to quality music education in the United States. Conversely, in music research and pedagogy, resources such as corpora or textbooks can marginalize certain demographic groups through canonization. That is, when a resource chooses to include one work over another, these resources implicitly signal the works as more important. Palfy and Gilson (2018) and Ewell (2020) address this issue of canonization explicitly in their surveys of music theory textbooks. These authors argue that, even though some textbooks do include a handful of works by non-white composers, the disproportional underrepresentation of Black and other non-white composers indicates that works by white composers are still those most worth studying.

This data report presents a demographic model and an accompanying database to address canonization biases in future popular-music corpus work. It specifically draws on sampling strategies in machine-learning research and interdisciplinary encoding procedures to foster more inclusive corpora building in future work along the parameters of race, ethnicity, and gender.

ARTIST DEMOGRAPHIC DATABASE

Data Sources

This report's dataset consists of demographic information for popular-music artists (*n* = 1,438) featured in the *RS200*, *MBB*, and *UG* song databases. *RS200* artists (*n* = 121) derive from the song list featured on Trevor de Clercq's personal website.[2] *MBB* artist (*n* = 417) names were gathered from the dataset index on The McGill Billboard Project website.[3] Artist names from the *UG* sample (*n* = 1,132) were parsed from the top-rated "pop" and "rock" songs encoded in the GuitarPro file format (*n* = 5,393 songs) featured in Shea (2019).[4] Demographic data were gathered from a variety of public sources, including Wikipedia.org, nndb.com, artist websites, and published artist interviews in print and online magazines.

File Format and Licenses

Data are presented in a CSV format. A searchable and downloadable version is available on Google Sheets.[5] No licenses are required to access the data.

DATA GENERATION AND COLLECTION

Demographic data are generated by following interdisciplinary encoding practices, including those in health sciences (Hauck et al., 2011) and survey reports conducted by the USC Annenberg Inclusion Initiative (Smith et al., 2020a, 2020b) and the Institute for Composer Diversity ("Composer Diversity Database," 2021). The following section summarizes the encoding procedure as it corresponds to evaluating the parameters of gender, race, ethnicity, and primary status.

Encoding Procedures

Broadly, the encoding procedure involves searching online resources for demographic information about popular-music artists and encoding this information as a series of dichotomous variables.[6] These variables are strictly operationalized to avoid making problematic assumptions about an artist's identity. For example, encoders are not permitted to encode variables based on visual evidence alone. Encoders also cannot encode certain variables, such as ethnicity, unless they are made explicit by the found source. To mitigate risk of artist mis-categorization, encoders are provided with a set of guidelines and a training sample before beginning data entry. The encoders then meet with the database developer to discuss their findings, clarify ambiguities, and cross-validate their results.

The author encoded demographic data for the *RS200* and *MBB* corpora. A team of four undergraduate and graduate students encoded the *UG* sample. To ensure accuracy, the author then hired an independent reviewer to verify all demographic data as it is included in this report.

Gender, *race*, and *ethnicity* are encoded under the condition that at least one member of the ensemble meets the described demographic criteria. The *primary status* condition, established by Shea (2019, p. 94), considers a minoritized artist's agency and identity in proportion with other ensemble members to avoid tokenism. That is, *primary status* and its related *gender* and *BIPOC* (i.e., "Black, Indigenous, and people of color," Garcia, 2020) conditions consider the demographic makeup of the entire ensemble.

GENDER

Artist gender is defined via two parameters, represented in columns. The *non-male* column parameter indicates the identity of artists whose gender corresponds to those historically assigned at birth as "male" or “female.”[7] These labels are also implemented for the gender identity of transgender artists. A “1” is assigned to the *non-male* column if any artist within the ensemble identifies as non-male, while a “0” is assigned for artists who identify as male. The *non-cis* variable is meanwhile reserved for any artist within the ensemble whose gender identity does not align with those that have historically been assigned at birth.[8] A “1” in the *non-cis* column indicates an artist identifies as non-cisgender (e.g., transgender, non-binary, etc.) and a “0” indicates the reverse. If an artist’s gender identity under the *non-cis* variable cannot be ascertained, the encoder enters a “0” for this column and defers to the pronouns used in the online source to categorize the artist accordingly under the *non-male* column.[9]

RACE AND ETHNICITY

Race and ethnicity are encoded as separate column parameters. Conditions for these parameters frequently overlap, but their distinction in the dataset reflects the broad ways in which non-white persons have been subject to marginalization. Artist race is treated as the socially determined distinction between human groups based on “perceived common physical characteristics” that are inherent from birth (Cornell & Hartman, 1998, p. 24). Race is primarily externally imposed on marginalized persons, such as by white Europeans as they enslaved African peoples (Cornell & Hartman, 1998, p. 24). Artist ethnicity is meanwhile defined as “a sense of common ancestry based on cultural attachments, past linguist heritage, religious affiliations, claimed kinship, or some physical trait” (p. 19).[10] Hispanic or Latino persons are treated by the United States Census as members of an ethnic category, while white, Black/African American, American Indian or Alaska Native, Asian, and Native Hawaiian or other Pacific Islander are all racial categories.[11]

When categorizing artists, encoders adhere to the following guidelines: encoders do not 1) encode race or ethnicity based on visual evidence alone, 2) encode race or ethnicity unless it is made explicit in the online resource, and 3) distinguish artists who are multiracial or multiethnic (e.g., Beyoncé) from those who are not. A “1” in the *race* or *ethnicity* column indicates an artist is non-white via the outlined parameters.

PRIMARY STATUS

The last column parameter*, primary status*, aims to disrupt demographic tokenism within the dataset. This parameter considers the demographic makeup of the entire ensemble. As an example, D.H. Peligro is the drummer for Dead Kennedys and is the only Black member of an otherwise all-white group. Guitarist and singer-songwriter Tracy Chapman is also a Black musician who leads backing bands whose demographic makeup varies depending on the performance, but whose members are often white. While Peligro and Chapman’s identities as Black artists are underrepresented — within their ensembles and across the sampled popular-music corpora — Chapman arguably holds increased agency as the title artist, lead singer, and primary songwriter of her group. She is both the public face and has a large degree of creative control compared to Peligro. Equating the Dead Kennedys as demographically analogous to Tracy Chapman therefore runs the risk of tokenizing Peligro’s Blackness as currency for a somewhat flimsy measure of diversification. Which is to say, it is inappropriate to categorize the Dead Kennedys as a primarily diverse group just because one member is Black.

A group is considered primarily diverse under the *primary status* column if 1) more than half of its members are minoritized by race, ethnicity, or gender, or 2) if the public-facing or title member of the ensemble is of minoritized demographic status. The former measures primacy by proportion, while the latter does so by artist agency. However, ambiguous cases where proportion and agency are at odds inevitably arise. Encoders use their own discretion to make judgements as supported by the evidence available to them and offer notes in the *comments* column for clarification.

A final consideration regarding *primary status* is selectivity during sampling. In a study on the field of music theory’s white racial frame, Ewell (2020) criticizes the Society for Music Theory’s Committee on Race and Ethnicity for splitting its focus amongst many types of diversity, which he argues marginalizes efforts to foster racial diversity (para. 6.3). Similarly, sampling artists via the primary status parameter as it stands runs the risk of over-fitting the data on categories such as gender at the sacrifice of race or ethnicity. Put another way, the primary status category acts as a broader measure of different types of diversity, but currently does not allow researchers to parse for primary-status artists along the lines of race or gender in isolation. In response to this potentially problematic lack of nuance, two additional parameters are implemented. Following Strmic-Pawl et al. (2018) and Smith et al. (2020a), *race* and *ethnicity* are synthesized in coordination with *primary status* to generate the *BIPOC* column. The *gender* column is similarly generated from the *non-male, non-cis*, and *primary status* parameters.[12] The *BIPOC* and *gender* columns therefore indicate primary-status artists who are so by measures of race and ethnicity or gender, respectively. Table 1 describes each sampling condition summarized in this section.

Table 1. Description of encoded demographic parameters and example artists/ensembles.

Column	Marker	Description	Example artist/ensemble
Non-male	0	All-male ensemble	Led Zeppelin
	1	Any other distribution of member identity	The Cranberries; Dolores O'Riordan (female)
Non-cis	0	All cisgender members	The Beatles
	1	Any other distribution of member gender identity	Fever Ray; Karin Dreijer (they/them)
Race	0	All-white ensemble	The Who
	1	Any other distribution of member identity	Fine Young Cannibals; Roland Gift (Black)
Ethnicity	0	All-white ensemble	Johnny Cash
	1	Any other distribution of member identity	Hombres G; all members (Spanish)
Primary status	0	Non-white/male member assumes a secondary role and constitutes less than half of the membership of the ensemble OR there are no non-white/male members.	Coldplay
	1	Non-white/male member is a founding member, forward-facing member of the ensemble, and/or composes material OR more than half of public-facing members are non-white/male.	Little Mix; all members (female), Leigh-Anne Pinnock (Black)

OTHER INFORMATION AND COMMENTS

Other identifying and potentially marginalizing characteristics such as sexuality, disability, level of education, and age are not summarized as column parameters and therefore are not used in the current sampling procedure. However, encoders frequently noted these relevant characteristics when observed. Database users can view related data by artist under the *comments* column. Finally, the author assumes all responsibility for any mis-categorization of an artist. The author also recognizes that one's gender identity can shift. As such, database users can submit requests to change or update an artist's demographic characteristics via Google Forms.[13] All requests will be reviewed for accuracy and implemented as soon as possible.

Summary Statistics

The following section summarizes demographic trends in the *RS200* and *MBB* popular-music corpora. The model is also applied to the *UG* artist sample to compare how these trends might reflect in an independent and more robust collection of artists (n = 962 unique artists).[14] *RS200* and *MBB* artists are those whose songs were respectively selected for encoding by measures of critical acclaim and commercial success, while artists included in the *UG* sample are based on online musician ratings of song transcriptions. The *UG* sample therefore provides an alternative measure of artist popularity.

Figures 1 and 2 respectively model the distribution of artist identity under the *gender* and *BIPOC* parameters in each individual corpus sample. Table 2 provides summary statistics for each parameter across the combined corpora. As shown, all three samples largely prioritize white and male artists, even under the less restrictive conditions of *non-male*, *non-cis*, *race*, and *ethnicity*. The Appendix includes additional summary figures, including by song count for the *RS200* and *MBB* corpora.

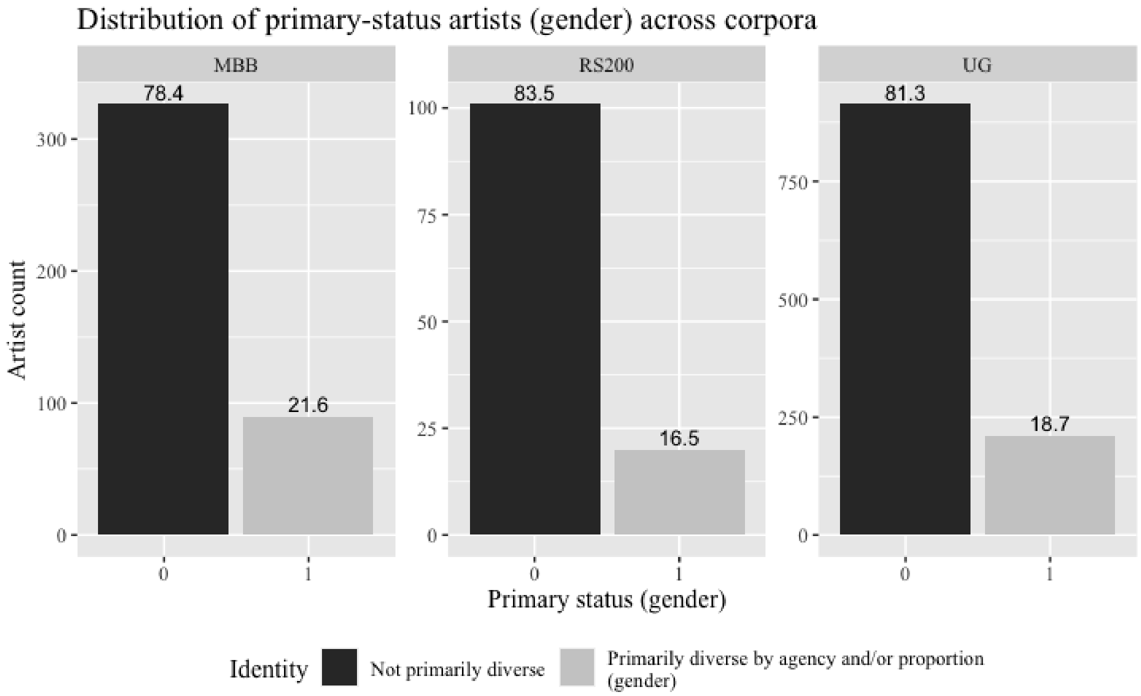

Figure 1. Distribution of primary-status artists (gender) across corpora artist samples.

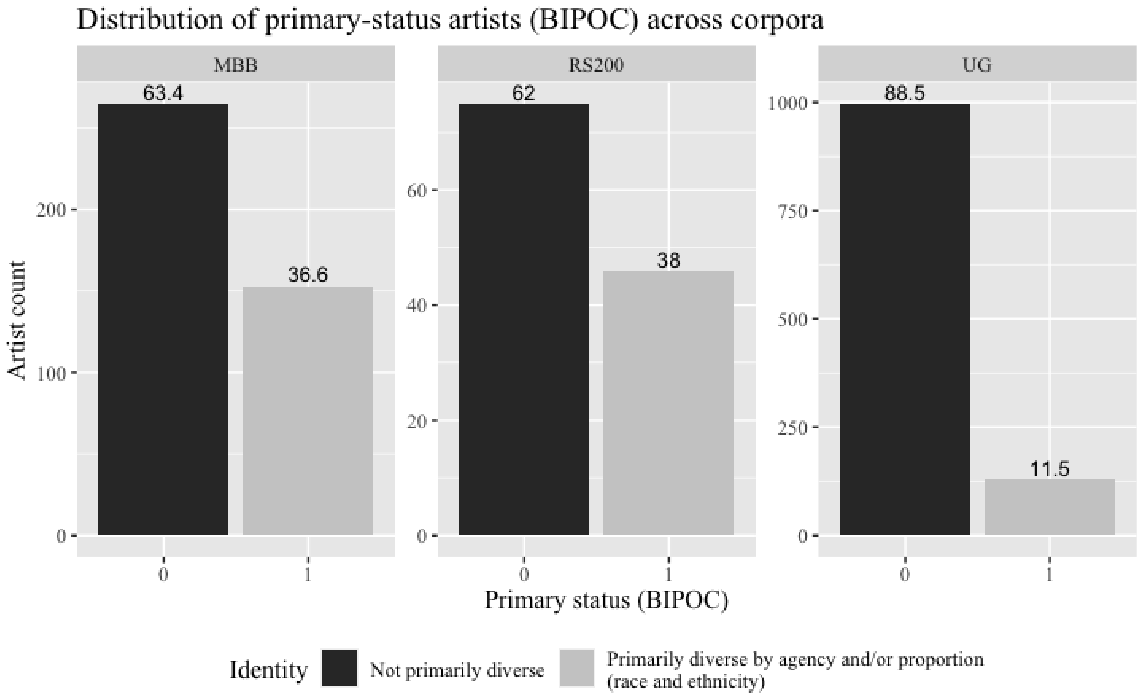

Figure 2. Distribution of primary-status artists (BIPOC) across corpora artist samples.

Table 2. Proportional distribution of demographic parameters in *McGill Billboard Hot 100*, *Rolling Stone 200*, and *ultimate-guitar.com* artist samples.

	At least one member				Agency and proportion			
corpora	non-male	non-cis	race	ethnicity	primary	BIPOC	gender	*n*
MBB	.237	.002	.372	.089	.452	.370	.217	414
RS200	.182	.000	.397	.116	.438	.380	.165	121
UG	.226	.007	.120	.118	.253	.115	.187	1126

POTENTIAL APPLICATIONS

The purpose of this data report is not to determine a universal benchmark for artist diversity in corpus studies. As shown by Smith et al. (2020b), artist demographics can vary widely across historical period and genre (p. 20), meaning researchers will need to determine which parameters best suit their representational needs. However, once these needs are assessed, a few lines of R code can be used to generate a suitable sample.

One such application could be to address the relative underrepresentation of Black artists in rock music. Johnson (2018, p. 38) and Redd (1985, p. 41) both argue that Black artists were essentially sequestered from measures of mainstream commercial success when the Billboard charts implemented the “rhythm and blues” genre label that distinguishes works by Black artists from those by white artists under the “rock” genre label. Redd specifically argues that the racial motivations for these labels are clear given that rock music and rhythm and blues music are functionally equivalent. Given the propensity for music theory studies, including the *RS200*, to prefer the umbrella term “rock” to encompass a wide variety of genres, there is an obvious concern that Black artists are unduly overlooked in existing resources. Similarly, female Black musicians such as Sister Rosetta Tharpe and Memphis Minnie were seminal in establishing the stylistic norms of rock music (Jackson, 1995; Lewis, 2018), but are subsequently underrepresented in current corpora. The accompanying R code takes these observations into consideration. [15] Specifically, it outlines how to create an artist sample (*n* = 200) that prioritizes BIPOC artists using R packages from the tidyverse (Whickham et al., 2019) and “splitstackshape” (Mahto, 2019, p. 26) when applied to the combined corpora artist database. This hypothetical sample has the following distribution of artist parameters: 50% of *primary-status* artists by race or ethnicity (*n* = 100), half of whom are non-male (*n* = 50), added to a random sample (*n* = 100) of other artists.

ACKNOWLEDGEMENTS

This article has been copy edited and layout edited by Jonathan Tang.

NOTES

[1] Correspondence can be addressed to: Nicholas J. Shea, ASU School of Music Dance and Theatre, 50 E Gammage Pkwy, Tempe, AZ 85281, njshea@asu.edu

[2] File “billboard-2.0-index.csv” retrieved from https://ddmal.music.mcgill.ca/research/The_McGill_Billboard_Project_(Chord_Analysis_Dataset)/

[3] http://rockcorpus.midside.com/overview/rs200.txt

[4] Shea (2019, pp. 93–95) outlines the procedure for parsing artists from *ultimate-guitar.com*.

[5] The spreadsheet can be accessed and downloaded as a csv file at https://docs.google.com/spreadsheets/d/1WA1aYrGgqlO96dpj6Gv0XpCBczWg7V4cQpdiR6zG67w/edit?usp=sharing.

[6] The Institute for Composer Diversity has living composers' self-report demographic information. Because researchers normally do not have the ability to interview popular-music artists, it is necessary to look to online sources, following the precedent set by Smith et al. (2020b).

[7] Column names are italicized to distinguish encoded variables from the real-world categories they represent.

[8] Non-cis identities can include but are not limited to intersex, non-binary, third gender, transgender, two spirit, transgender man, and transgender woman. These categories match those found on the Institute for Composer Diversity Database (https://www.composerdiversity.com/composer-diversity-database).

[9] Again, this follows the procedure used by Smith et al. (2018), who clarify "gender was assigned by scrutinizing online information, industry databases, pronoun use, and online interviews" (p. 22).

[10] See "Race" in the *Stanford Encyclopedia of Philosophy* (May 2020).

[11] The utility of US Census categories is subject to discussion, as they are frequently employed in studies on the social construct of race (Ifekwunigwe et al., 2017), but are also argued to further discriminate in political practice (Nobles, 2000; Strmic-Pawl et al., 2018).

[12] The column names *BIPOC* and *gender* reflect an attempt to disrupt the potential implication of whiteness and maleness as default as conveyed through the previous column labels *nonmale*, *non-cis*, *race*, and *ethnicity*. Following the above discussion of US Census categories, monolithic terms for personal identities can be used to both harm and empower. It is also where the principles of intersectionality (Krenshaw, 1989; Oluo, 2019, pp. 70–82) appear most incompatible with the binarity required in empirical analysis. When wrestling with this issue I am reminded of an axiom of David Huron's: "Reduction is a method, not a belief."

[13] Requests can be submitted at the following form: https://forms.gle/hKfEtwZgRAVdZuzHA.

[14] Each sample shares mutual artists: RS200 and MBB (n = 22), RS200 and UG (n = 31), MBB and UG (n = 97), mutual artists across all samples (n = 40). Overlapping artists are indicated under the *corpora* column of the linked dataset.

[15] Sample code can be accessed at https://github.com/njshea/EMR-demo/blob/9393543469edb66e1e77c4c3bc4bdc782ef69c58/R-sample-code.txt

REFERENCES

Burgoyne, J. A., Wild, J., & Fujinaga, I. (2011). An expert ground-truth set for audio chord recognition and music analysis. In A. Klapuri & C. Leider (Eds.), *Proceedings of the 12th International Society for Music Information Retrieval Conference* (pp. 633-638). Canada: ISMIR. https://doi.org/10.5281/zenodo.1417547

de Clercq, T., & Temperley, D. (2011). A corpus analysis of rock harmony. *Popular Music*, *30*(1), 47–70. https://doi.org/10.1017/S026114301000067X

Institute for Composer Diversity. (n.d.) *Composer Diversity Database*. Retrieved August 27, 2021, from https://www.composerdiversity.com/composer-diversity-database

Cornell, S., & Hartmann, D. (2006). *Ethnicity and race* (2nd ed.). Thousand Oaks, CA: Pine Forge Press.

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, *1*(8), 139–167.

Devaney, J., Arthur, C., Condit-Schultz, N., & Nisula, K. (2015). Theme and variation encodings with roman numerals (TAVERN): A new dataset for symbolic music analysis. In M. Müller & F. Wiering (Eds.), *Proceedings of the 16th International Society for Music Information Retrieval Conference* (pp. 728–734). Canada: ISMIR. https://doi.org/10.5281/zenodo.1417497

Ewell, P. A. (2020). Music theory and the white racial frame. *Music Theory Online*, *26*(2). https://doi.org/10.30535/mto.26.2.4

Garcia, S. E. (2020, June 17). Where did BIPOC come from? *The New York Times*. Retrieved from https://www.nytimes.com/article/what-is-bipoc.html

Hauck, F. R., Tanabe, K. O., & Moon, R. Y. (2011). Racial and ethnic disparities in infant mortality. *Seminars in Perinatology*, *35*(4), 209–220. https://doi.org/10.1053/j.semperi.2011.02.018

Ifekwunigwe, J. O., Wagner, J. K., Yu, J.-H., Harrell, T. M., Bamshad, M. J., & Royal, C. D. (2017). A qualitative analysis of how anthropologists interpret the race construct. *American Anthropologist*, *119*(3), 422–434. https://doi.org/10.1111/aman.12890

James, M., & Burgos, A. (2020). Race. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020 edition). Stanford, CA: Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/sum2020/entries/race/

Jackson, J. M. (1995). The changing nature of gospel music: A Southern case study. *African American Review*, *29*(2), 185–200. https://doi.org/10.2307/3042290

Johnson, T. (2018). *Analyzing genre in post-millennial popular music*. Doctoral dissertation, City University of New York, NY: The Graduate Center. Retrieved from https://academicworks.cuny.edu/gc_etds/2884

Kinney, D. W. (2018). Selected nonmusic predictors of urban students' decisions to enroll and persist in middle and high school music ensemble electives. *Journal of Research in Music Education*, *67*(1), 23–44. https://doi.org/10.1177/0022429418809972

Mahto, A. (2019). Package 'splitstackshape' (Version 1.4.8). Retrieved from https://cran.r-project.org/web/packages/splitstackshape/splitstackshape.pdf

Neuwirth, M., Harasim, D., Moss, F. C., & Rohrmeier, M. (2018). The annotated Beethoven corpus (ABC): A dataset of harmonic analyses of all Beethoven string quartets. *Frontiers in Digital Humanities*, *5*(1). https://doi.org/10.3389/fdigh.2018.00016

Nobles, M. (2000). *Shades of citizenship: Race and the census in modern politics*. Redwood City, CA: Stanford University Press. https://doi.org/10.1515/9780804780131

Oluo, I. (2019). *So you want to talk about race*. New York, NY: Seal Press.

Palfy, C., & Gilson, E. (2018). The hidden curriculum in the music theory classroom. *The Journal of Music Theory Pedagogy*, 1–32.

Reschovsky, A. (2016). *The future of U.S. public school revenue from property tax*. Cambridge, MA: University of Madison-Wisconsin Lincoln Institute of Land Policy. Retrieved from https://www.lincolninst.edu/sites/default/files/pubfiles/future-us-public-school-revenue-policy-brief_0.pdf

Redd, L. (1985). Rock! It's still rhythm and blues. *The Black Perspective in Music*, *13*(1), 31–47. https://doi.org/10.2307/1214792

Shea, N. J. (2020). *Ecological models of musical structure in pop-rock, 1950–2019*. Doctoral dissertation, Ohio State University, OH. Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=osu158755665247824

Smith, S. L., Choueiti, M., & Pieper, K. (2020a). *Inequality in 1,300 popular films: Examining portrayals of gender, race/ethnicity, LGBTQ & disability from 2007 to 2019*. Los Angeles, CA: USC Annenberg Inclusion Initiative. Retrieved from https://assets.uscannenberg.org/docs/aii-inequality_1300_popular_films_09-08-2020.pdf

Smith, S. L., Pieper, K., Clark, H., Case, A., & Choueiti, M. (2020b). *Inclusion in the recording studio?* Los Angeles, CA: USC Annenberg Inclusion Initiative. Retrieved from https://assets.uscannenberg.org/docs/aii-inclusion-recording-studio-20200117.pdf

Strmic-Pawl, H. V., Jackson, B. A., & Garner, S. (2018). Race counts: Racial and ethnic data on the U.S. census and the implications for tracking inequality. *Sociology of Race and Ethnicity*, *4*(1), 1–13. https://doi.org/10.1177/2332649217742869

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

APPENDIX

Table 3. Summary of unique artist demographic categories in the *McGill Billboard Hot 100* artist sample (*n* = 417 artists).

non-male	non-cis	race	ethnicity	primary	BIPOC	gender	*n*
0	0	0	0	0	0	0	193
0	0	1	0	1	1	0	83
1	0	1	0	1	1	1	53
1	0	0	0	1	0	1	33
0	0	0	1	0	0	0	17
0	0	1	1	1	1	0	9
1	0	0	0	0	0	0	7
0	0	1	0	0	0	0	6
0	0	0	1	1	1	0	4
1	0	0	1	1	1	1	3
0	0	1	1	0	0	0	2
0	0	0	0	1	0	0	1
0	1	0	0	0	0	0	1
1	0	0	1	0	0	0	1
1	0	1	1	1	1	1	1

Table 4. Summary of unique artist demographic categories in the *Rolling Stone 200* artist sample (*n* = 121 artists).

non-male	non-cis	race	ethnicity	primary	BIPOC	gender	*n*
0	0	0	0	0	0	0	53
0	0	1	0	1	1	0	32
1	0	1	0	1	1	1	13
0	0	0	1	0	0	0	11
1	0	0	0	1	0	1	6
0	0	0	0	1	0	0	1

0	0	1	0	0	0	0	1
0	0	1	1	0	0	0	1
1	0	0	0	0	0	0	1
1	0	0	1	0	0	0	1
1	0	1	1	1	1	1	1

Table 5. Summary of unique artist demographic categories in the *ultimate-guitar.com* sample (*n* = 1127 artists).

non-male	non-cis	race	ethnicity	primary	BIPOC	gender	*n*
0	0	0	0	0	0	0	737
1	0	0	0	1	0	1	151
0	0	1	1	1	1	0	48
1	0	0	0	0	0	0	40
0	0	0	1	0	0	0	31
1	0	1	1	1	1	1	22
0	0	1	0	0	0	0	20
0	0	1	0	1	1	0	18
1	0	0	1	1	1	1	16
1	0	1	0	1	1	1	16
0	0	0	1	1	1	0	8
0	0	1	1	0	0	0	7
1	1	0	0	1	0	1	4
1	0	1	0	0	0	0	3
0	1	0	0	0	0	0	1
0	1	0	0	1	0	1	1
1	0	0	1	0	0	0	1
1	1	0	0	0	0	0	1
1	1	1	0	1	1	1	1

Table 6. Summary of artist demographic categories by song count in the *McGill Billboard Hot 100* and *Rolling Stone 200* corpora.

	At least one member				Agency and proportion			
corpora	non-male	non-cis	race	ethnicity	primary	BIPOC	gender	*n*
MBB	165	1	234	69	298	233	153	734
RS200	26	0	73	30	80	71	24	200