Reviews in Digital Humanities • Vol. 2, No. 3

Review: SongData

Kayleigh Voss¹

¹University of Texas at Austin

Published on: Mar 08, 2021

DOI: 10.21428/3e88f64f.31cc3cab

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Project

SongData

Project Team

[Jada Watson](#), Principal Investigator, University of Ottawa

[André Vellino](#), Collaborator, University of Ottawa

Project URL

<https://songdata.ca/>

Project Reviewers

[Kayleigh Voss](#), University of Texas at Austin

Project Overview**Jada Watson**

What can we learn about a music genre through its discography? How can information *about* music help us better understand a genre's history and evolution? How can information about music's consumption be used to discover an industry's behavior and practices? How can biographic data about the artists responsible for musical works within a discography lead to new discoveries about a genre's origins, community, and culture?

Launched publicly in September 2018, SongData (which stands for Studies of Networks and Genres through Data) is interested in investigating the potential uses of discographic and biographical data to learn more about how popular music genres form, develop, and evolve over time. Adopting methods of Big Data research in the humanities and social sciences and influenced by the concept of prosopography or collective biography, SongData is developing an approach for collecting and organizing music industry data to engage in deep examinations of the genre's culture and the industry's structure and practices. We are seeking to explore what information *about* songs and artists can tell us about connections between them and the broader socio-cultural and institutional frameworks that govern genres.

There are currently two active projects under SongData, both funded by grants from the Social Sciences and Humanities Research Council of Canada. These projects emerged out of an interest in learning more about growing inequalities within the country music industry, made public by a 2015 interview by a radio consultant who

advocated for a practice of programming songs by women at 13-15% of a station's playlist. In addition to the gender inequality that was becoming increasingly more apparent on weekly popularity charts and radio playlists, the overwhelming and historic whiteness of the industry could no longer be ignored—especially given the diverse roots of the genre. For these projects, we are curating a database of all singles that charted on the *Billboard* Hot Country Songs chart from its inception in 1944 to 2017. The longest-running country singles chart, Hot Country Songs, has tabulated popularity through a variety of formats and methods (from jukebox records to modern streaming) and is comprised of 19,844 charting singles over a 70-year period. These projects are described on the [Projects](#) pages, but the critical analysis of the data is presented in the articles, reports, the [Keepers of the Flame blog](#), and featured in various interviews or podcasts archived on the [Press](#) page.

Each single in this dataset is being augmented with discographic information about the recorded song (songwriter, producer, label, album, etc.) and biographical details about the broader network of individuals involved in the creative process, from composition to recorded production. The underlying technology for the prosopography is a spreadsheet. The historical Hot Country Songs data was acquired directly from Billboard's Research Services Office in .csv tabular format, which includes the name of the artist, featured artist(s), and song title, along with the charting history for each single. We are enriching the data with discographic metadata that will enable us to make new discoveries about the community of artists that have contributed to the development of country music culture. Discographic data can be scraped from a variety of sources, but it is often incomplete, inaccurate, or missing altogether. Since the data needed is already available in catalogues published by chart expert Joel Whitburn, much of the discographic augmentation has been done manually. The discographic augmentation process generates the list of artists, songwriters, and producers, which we are augmenting with biographical details. These details were already available in the Whitburn catalogues for lead performers, but we are relying on artist records from [MusicBrainz](#) for the prosopography of songwriters and producers. We downloaded all of the artist records in their database, and my research assistant (Fatima Sajadi) wrote a Python script to sort the data by genre and generate a .csv file of records for artists affiliated with country music. Because the original dataset from *Billboard* is under a license agreement, we cannot just republish the chart data. However, once completed, the prosopography and discography will be made available via the SongData website for those interested in using the data to conduct their own research. A dataset of the Hot Country Songs chart from 1987-2017

has been completed, and we have been using [RapidMiner](#) to analyze the data. The first study to emerge from this data was published in [Popular Music & Society](#) and an article about the impact of the changing chart methodology is soon to be published in a special issue on Popular Music Curation in the *Journal of Popular Music History*.

In addition to these funded projects, SongData has been exploring issues of equity, diversity, and inclusion on the weekly and year-end airplay reports for country format radio in the USA and in Canada. This work, outlined under the [RadioData](#) page on the website, has been conducted in consultation with the advocacy group WOMAN Nashville, included in CMT's EqualPlay campaign, and cited as a source in a [brief](#) submitted to the US Federal Communications Commission in response to the National Association of Broadcasters' proposal for further deregulation of radio. It has also been cited in the Grammy Recording Academy's [Report on Inclusion and Diversity](#) in the music industry. While the American and Canadian country music industries have been the primary focus of SongData's activities, future work will expand to other genre formats, including Top 40, rock, R&B, and adult contemporary.

Project Review

Carl Teegerstrom and Kayleigh Voss

SongData, a project led by Jada Watson with collaborator André Vellino at the University of Ottawa, applies big data and collective biography methods to the history of the *Billboard* Hot Country Songs chart between 1944 and 2017. SongData delves deeply into these charts, past the numbers, casting a critical eye on this history. Working with the advocacy group WOMAN Nashville to advance efforts to combat discriminatory radio regulations and programming, SongData pays particular focus to gender discrimination. As any country radio listener may notice, artists who are women get less air time, which in turn influences sales and awards. A simple way to quantify—and potentially correct—this imbalance is by looking at the numbers as SongData does: who charts, when, and how often. In addition to these efforts to combat discrimination, SongData also combines discographic and biographical data to reveal networks between artists, songs, producers, eras, and trends in this 74-year-long period, visualizing and quantifying country music history as it has never been done before.

SongData has licensed the *Billboard* Hot Country Songs charts from *Billboard* directly. Their team then manually appends discographic data, which includes details like the

album, label, producer, songwriter, etc., sourced from other publications. The discographic data is then augmented with biographical data about the people involved in making each song. SongData pulls together these various, already existing datasets for the first time with tools like RapidMiner and Gephi as well as Python. Due to the nature of the licensing agreement with *Billboard*, the charts cannot be reproduced online, but the additional layers of discographic and biographic data will be made available once they've been completed.

A central mission of the project—to illuminate networks in country music—won't be possible until the data is made visible and accessible, a goal that is being addressed by the project team. For now, users can read overviews and analyses of SongData's two current projects, "Country Music's 'Geo-cultural' Origins" and "Gender Representation in Country Music," both of which are funded by the Social Sciences and Humanities Research Council of Canada. These two projects tackle some of the most pressing issues in country music—gender and racial inequality—by tracing and uncovering country music's diverse history. Contrary to country music's white, male image, women and Black, brown, and Indigenous people have played an outsized role in the genre's history; these projects serve to amplify their previously silenced voices. While the site is searchable, future improvements might address improved search functionalities and linking—only bibliographic citations are provided in search results rather than full-text prose from the publications, and results appear one per page, which slows down the process unnecessarily.

Part of Watson's mission is to reclaim the importance of women who, due to discriminatory radio programming, don't see their music and its influence on the genre reflected in the Hot Country Songs chart. To that end, the site's blog, "Keepers of the Flame," focuses on both what is and isn't included in the SongData database. So far the blog features 26 posts such as, "Is There Gender Equity on Satellite Radio? Let's take a look!" and "20 Years of Country Music Association Awards Nominations." In these analytical posts, SongData offers a narrative, quasi-journalistic examination of the dataset and—importantly—its silences. The style of writing is accessible: never too-academic, occasionally personal, but serious, validating, and compelling. This is where SongData shows its potential both to rewrite country music history and to provide quantitative evidence that can be leveraged for change in the industry. The project may consider broadening its footprint by providing a moderated community contribution section where users can respond to, interact with, contribute, and query the data themselves.

Recognizing that keeping a contemporary digital project current is a constant struggle, the period from 2017 to 2020 that is not accounted for in the dataset might serve many researchers' needs. For most large-scale analysis, a 3-year gap might be considered negligible. But these three years in particular have seen undeniably historically important moments for country music that SongData should account for if Watson wants to paint a full picture of the industry that includes its issues and growing pains alongside its accomplishments. Especially relevant might be the controversy surrounding Lil Nas X's breakthrough hit "Old Town Road," which debuted on the *Billboard* Hot Country Songs chart in March of 2019 before *Billboard* reclassified it and had it removed, sparking conversations about racism in country music: why is country music so white, and how does *Billboard* help continue this practice?

Kacey Musgraves, politically outspoken both in her music and her public appearances, won Album of the Year at the Grammys in 2019, though her performance on *Billboard*'s Hot Country Songs chart wouldn't have foretold it. Why doesn't *Billboard* account for the popular success of women in country music—especially those who buck conventions? In response to widespread Black Lives Matter protests in 2020, several high-charting country music acts altered their names to distance themselves from the language and symbolism of the former Confederacy: Lady Antebellum has chosen to go by Lady A, appropriated from a Black singer-songwriter; The Dixie Chicks, widely reviled for their public denouncement of George W. Bush, are now known as The Chicks. Will these changes affect these groups' popularity and radio playtime? Do country music listeners want this kind of change? Does the industry? Watson undoubtedly has these questions in mind; adding additional researchers to the project either directly as collaborators or as community contributors might help get some of these and other questions answered in future iterations.

Once complete, SongData will prove to be an invaluable resource for both close and distant reading of specific genres of music (Watson indicates she wants to expand the project to include other genres besides country). *Billboard*'s data takes on new and rich meaning once paired with historical information, and the data, once disclosed, should make good on its promise to illuminate patterns and connections through the history of country music.