

24TH FEBRUARY 2019

The Harkive Project: popular music reception, digital technologies, and data analysis

Craig Hamilton

Craig Hamilton is a Research Fellow in the School of Media at Birmingham City University. His research explores contemporary popular music reception practices and the role of digital, data, and internet technologies on the business and cultural environments of music consumption. This research is built around the development of [The Harkive Project](#), an online, crowd-sourced method of generating data from music consumers about their everyday relationships with music and technology. Craig is also the Co-Managing Editor of [Riffs: Experimental Research on Popular Music](#).

Abstract

Through an analysis of how respondents to [The Harkive Project](#) describe their use of vinyl records, this article will demonstrate and reflect upon the development of an experimental methodological approach derived from the fields of digital humanities and cultural analytics, and show how this was applied to my 'home' discipline of popular music studies. Before proceeding to my analysis, I first describe the context and rationale for taking this approach. In reflecting on this approach I discuss how it enabled me to explore how data-derived knowledge creation works through practice within contemporary popular music culture, highlighting some of the issues raised by data-related technologies and techniques in both popular music culture and in arts and humanities research. My hope is that work in this area may help popular music studies begin to account for the technologies and practices that have so changed the field. Towards that aim, and in consideration of Sandvig and Hargittai's recent work

highlighting the importance of ‘benchwork’, my article links to code, sample data, and instructional blog posts that may enable scholars to replicate and/or build upon my work.

Introduction

This paper was delivered at the 2018 [Listening Experience Database \(LED\)](#) project conference at The Open University, Milton Keynes, on 7 March 2018. The conference took place just a few days after I completed my PhD at Birmingham City University and was in fact my first outing as a newly minted doctor. I am delighted, then, to have been invited to revisit that talk for this volume of LED proceedings.

During my [AHRC-Midland3Cities](#)-funded research project I looked at popular music reception, with a particular focus on the digital, data, and internet technologies that have over the last 20 years helped bring about such huge changes in my field of study. In this article I want to provide an overview of my research, and in particular describe how working through a number of methodological issues ultimately led me towards ideas and practices that may be more broadly understood as ‘data science’. Ultimately, and just as I did in my thesis, I want to argue that a greater practical understanding of and critical engagement with the digital, data, and internet technologies is possible, both for popular music scholars such as myself, and also for the millions of people who engage with popular music in their everyday lives. Along the way I hope to demonstrate that my work represents a small step towards that.

After briefly describing [Harkive](#), the project that underpinned my research, I will demonstrate how I arrived at using automated data collection and computational analysis techniques in my work. By walking through an example of the type of analysis I undertook, I want to highlight some of the potential benefits and problems with such an approach. I will then reflect on some of those potential benefits and problems, and suggest some next steps. My hope is that the work undertaken during my project may provide a springboard for future work, and in particular for the creation of new tools, platforms, and research projects that may enable consumers and scholars alike to develop useful and productive epistemic responses to the role of digital, data, and internet technologies in popular music. To begin, however, I should briefly explain what The Harkive Project is.

The Harkive Project

Harkive is an online, crowd-sourced project that runs on a single day in July of each year. It invites people to provide detail and reflection on their experiences with music across the course of a single day. Since the project first ran in 2013 it has gathered over 10,000 individual entries. The reflections and detail contained within the Harkive dataset come from posts made to social media platforms, from participants who have emailed the project directly, or from people who have completed an online form. The shortest entry in the database contains only two words, the longest almost 4,000.

Taken as a whole the Harkive dataset contains descriptions and reflections on many different forms of music reception. By music reception, I am taking Keith Negus' definition of 'how people receive, interpret and use music as a cultural form while engaging in specific social activities.'

[footnote] [1] Keith Negus, *Popular Music in Theory: An Introduction* (Middletown: Wesleyan University Press, 1997), p. 4. [/footnote]

This definition encompasses how we engage with music in our everyday lives much more usefully than perhaps the embodied act of 'listening', or more commercially focused ideas of 'consumption' could. This is because music is not just something we hear, or buy; it is also something we talk about, think about, or otherwise use in a variety of ways.

The texts collected by Harkive represent snapshots of different individuals' engagement with music, and each (to varying degrees) contains descriptions of respondents' use of various technologies, or the locations and everyday situations involved when engaged in music reception activity. Some stories are inspired by memories, others detail engagement with technologies, and many show how deeply experiences with music and technology are woven into the rhythms and routines of everyday life.

Before proceeding to how I analysed those texts, I should like to provide a little more context for my research. This context is crucial, because it was engaging with the broader issues of debate around contemporary popular music reception, and in particular the role of digital, data, and internet technologies within that, that ultimately informed the methodological approach I took.

Research context

Over the last two decades digital, data, and internet technologies have emerged as important and influential factors in how popular music is produced, distributed and

consumed. These technologies, allied to practices of data collection and computational analysis, now play a significant role both in how audiences engage with music, and how those audiences are understood. A key point here is that popular music audiences are now highly individualised, and defined according to a growing number of new categorical variables. At the same time, however, audiences are also understood through the large-scale agglomeration of data points. An example here would be streaming services such as Spotify, which provide access to music through an interface that facilitates the capture, analysis and use of data about the daily activity of millions of people. We may also wish to consider social media platforms, where people discuss and share music, or search engines and online retailers, all of which gather data about users and derive forms of knowledge from that which is then deployed according to what Hartmann et al. call Data-Derived Business Models (DDBMs)

[footnote] [2] Philipp Max Hartmann, Mohamed Zaki, Niels Feldmann and Andy Neely, 'Big data for big business? A taxonomy of data-driven business models used by start-up firms. A Taxonomy of Data-Driven Business Models Used by Start-Up Firms', University of Cambridge, available at https://cambridgeservicealliance.eng.cam.ac.uk/news/March2014Paper, accessed 8 February 2019. [/footnote]

– that is, models relying on data as a key resource.

For popular music scholars these are intriguing developments, but studying these new environments is difficult. This is because the systems of data collection and analysis that facilitate them are technologically complex, subject to rapid change, and are often hidden behind commercial and legal firewalls.

[footnote] [3] Mike Ananny, 'Toward an Ethics of Algorithms Convening, Observation, Probability, and Timeliness,' *Science, Technology & Human Values*, 2015, 0162243915606523. [/footnote]

At the same time, however, the use of online technologies by many people during the course of their everyday lives is providing scholars with new opportunities and methods for undertaking research in the humanities. This in turn is leading to questions about the role of the researcher, and – in the case of popular music studies – how we as scholars may take into account the new technologies and practices that have so changed the field. Of particular interest to me are automated recommender systems, the manner in which digital interfaces foreground (or not) content to audiences, and the new ways in which audiences are conceived of and organised.

Given the growing importance of streaming services, social media platforms, search engines, and so on, within the field, popular music scholars wishing to understand contemporary conditions of music reception are faced with the problem of exploring

the consequences of systems that they presently lack sufficient access to, or else the technical knowledge and skills required to fully understand.

[footnote] [4] Danah Boyd and Kate Crawford, 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon,' *Information, Communication & Society* 15, no. 5 (2012), pp. 662–79. [/footnote]

Although in the case of streaming services, for example, recent work by Hagen,

[footnote] [5] Anja Nylund Hagen, 'The Playlist Experience: Personal Playlists in Music Streaming Services,' *Popular Music and Society* 38, no. 5 (2015), pp. 625–645. [/footnote]

Nowak,

[footnote] [6] Raphaël Nowak, 'Investigating the Interactions between Individuals and Music Technologies within Contemporary Modes of Music Consumption,' *First Monday* 19, no. 10 (2014). [/footnote]

Prey,

[footnote] [7] Robert Prey, 'Henri Lefebvre and the Production of Music Streaming Spaces,' *Sociologica* 9, no. 3 (2015), pp. 1–22. [/footnote]

and Webster et al.

[footnote] [8] Jack Webster et al., 'Towards a Theoretical Approach for Analysing Music Recommender Systems as Sociotechnical Cultural Intermediaries,' in *Proceedings of the 8th ACM Conference on Web Science* (ACM, 2016), pp. 137–145. [/footnote]

has made progress in helping to develop our understanding, a wider question for popular music scholars is how such an inquiry into these recently emerged systems may be usefully undertaken. Through The Harkive Project, and via the methods I have developed, my intention has been to see if I could make a contribution to that wider question by discovering a little more about the new conditions of everyday music reception.

Towards a method

In terms of the practicalities of my doctoral work, the size, variety and complexity of the data I had collected through Harkive immediately presented two interrelated methodological and intellectual problems:

- How could I derive useful information and insight from the large collection of unstructured texts I had gathered?

- How could I do that in such a way that said something useful and interesting about contemporary popular music reception?

In attempting to deal with those questions, I decided that a potentially productive route would be to subject the Harkive data to similar computational analysis processes to those used by commercial organisations seeking to understand the way people engage with popular music. What I mean here is the idea that activities such as listening through streaming services, Googling your new favourite artist, or discussing music with friends on social media – in other words, activities that would fall under the definition of ‘music reception’ – now often take place within environments where, as Bernhard Rieder describes, data capture, analysis and output are integrated.

[footnote] [9] Bernhard Rieder, ‘Big Data and the Paradox of Diversity’. *Digital Culture & Society*, 2(2), pp. 39-54. [/footnote]

Because of this, music reception activities now create data points that can be aggregated and analysed in order to produce a form of knowledge that in turn informs interface design, or the foregrounding of certain content, which ultimately impacts upon the experiences we may have. We can consider here, for example, automated recommendation and curated playlists offered by streaming services, or product recommendations positioned within the interfaces of online retailers and social media platforms. Through analysing the Harkive texts in a similar manner, and alongside discovering what respondents were saying about their music reception, I wanted to explore the processes involved with attempts to represent complex elements of individual real-world experience through the medium of data and computational analysis.

As attractive as this proposed direction for my research was, however, it revealed an immediate problem related to my own research skills. I attempted to define this problem early on in my research journey, in this quote taken from my research notebook:

I am a reasonably tech-savvy media scholar, but I am not a data scientist, or a coder. Yet I am building a PhD research project that hinges on my ability to make sense of my data through computational techniques.

In short, I had backed myself into a practical, methodological and intellectual cul-de-sac.

The process of finding my way out of that cul-de-sac started with a period of trying to get to grips with the technologies and practices involved. Through an intense period of reading, attending conferences, listening to ‘data science’ podcasts, and several months of self-directed learning in the R programming language, I slowly got to grips with the practicalities of how data and computational techniques are used in commercial settings and academic research. A key realisation, and indeed a consideration for the purposes of the remainder of this article also, is that the texts gathered by Harkive were in a digital format and were thus reducible to data points that could be counted, analysed, and otherwise computationally processed at scale. Stripping away the unique, individual detail contained within each response, the ‘raw material’ I had to work with could be understood as data. My research project was now, concurrently, a philosophical exploration of how data-derived ‘knowledge’ is created and a data science ‘problem’ is to be solved.

Using a number of automated collection methods,

[footnote] [10] For a detailed breakdown of this process, see the instructional overview provided on The Harkive project website: <http://harkive.org/datcolzap> . [footnote]

Harkive data was subsequently gathered into a single database organised according to the principles of tidy data,

[footnote] [11] Hadley Wickham, ‘Tidy Data,’ *Journal of Statistical Software* 59, no. 10 (2014), pp. 1–23. [footnote]

which made it ready for computational analysis at the point it was collected. The dataset contained text-based submissions and quantitative survey responses, along with metadata gathered during the collection processes – including time/date stamps, and detail on which platforms each text was collected from – and additional variables generated through the use of a series of unsupervised machine learning algorithms, including topic modelling, which is described in more detail below. This meant that the data about music reception activities available to me could be understood and analysed in a number of different ways, ranging from the close readings of texts more usually associated with humanities research, through to the clustering, visualisation and analysis of abstractions generated through computational/algorithmic processes that rendered the original texts as data. The method also allowed for analyses that combined these approaches.

All of which is to say that what I ended up with was an experimental approach that conceived of Harkive as both a ‘space’ in which people could reflect upon their engagement with music, and simultaneously a ‘place’ able to replicate many of the commercial practices related to data collection and processing. Through this, I sought to critically engage with the growing role of data-related technologies associated with

music reception, whilst simultaneously exploring the use of those techniques in popular music studies research. An intriguing sub-question that emerged from this methodological development was to attempt to discover what patterns, correlations and other potentially interesting insights could be derived through the use of techniques that are largely guided by their own internal mathematical logics, and less by the human researcher. In the next sections I will walk through one of the ways I explored this question.

LDA topic modelling – overview

One unsupervised algorithmic technique I employed was topic modelling, which David Blei

[footnote] [12] David M. Blei, 'Topic Modeling and Digital Humanities,' *Journal of Digital Humanities* 2, no. 1 (2012), pp. 8–11. [/footnote]

defines as a process that:

provides a suite of algorithms to discover hidden thematic structure in large collections of texts. The results of topic modelling algorithms can be used to summarize, visualize, explore, and theorize about a corpus.

[footnote] [13] Blei's article appears online in the *Journal of Digital Humanities*, and as such no page numbers are attributed to his quotes. The full article is available here: <http://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei/> [/footnote]

Topics can better be understood as recurring data points (in this case, words) across a larger dataset (a corpus of text documents). The model, meanwhile, is a mathematical representation of the extent to which each individual entry in a dataset – that is the documents within a given corpus – contains data points – that is topics/words. Applying this to the case of my own research and the Harkive texts, the immediate potential in a process of this kind was that the larger a collection of documents is, the more difficult and labour intensive it becomes to manually explore, encode and reveal common themes within it. Beyond simply saving time and effort, however, a further potential advantage lay in the possibility that there may be themes or topics within the dataset that were not immediately apparent. In other words, these may be hidden or

otherwise made 'latent' by the complexity and scale of the corpus. The broad rationale behind topic modelling is that latent themes in text corpora may be revealed by mathematical processing.

I employed the most commonly used approach in automated topic modelling, Latent Dirichlet Allocation (LDA), which was originally developed in 2003 by Blei and colleagues at Berkley. Blei has since argued that automated modelling of this kind can help humanities scholars 'build a statistical lens that encodes...specific knowledge, theories, and assumptions about texts'.

[footnote] [14] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, 'Latent Dirichlet Allocation,' *Journal of Machine Learning Research* 3, no. Jan (2003), pp. 993–1022. [/footnote]

Blei reveals that LDA in particular can be understood as being based on two assumptions:

- There is a finite number of patterns of words, or groups of terms, that occur together within a corpus.
- Each document within a corpus exhibits these to a varying degree.

Although the documents and words are observable to a manual reading, the thematic structure – in other words, the topics – may be hidden, and will be more likely to be hidden as the scale and complexity of a corpus increases.

LDA infers these hidden structures based on what can be computationally processed (that is, by counting the frequency with which unique words occur across documents), and represents these in terms of the probabilistic likelihood that a document belongs to a given topic. It was this process and rationale that was applied to the Harkive dataset from the years 2013–2016 inclusive.

LDA Topic modelling and Harkive texts

Using the topicmodels package with the R software, the LDA process described above revealed the following words associated with five topics, organised here according to frequency, with the top 20 words associated with each topic shown in Figure 1 below. The LDA process of course had no 'knowledge' about the contents or meaning of the document corpus and had processed it purely according to the statistical probability of words appearing in documents. Nevertheless the initial results were interesting, and I was able to interpret the topics as being relatively distinct. The appearance, for

instance, of words such as morning, home, office, train, commute, and so on, were interpreted as being related to ideas of ‘Time and Place’. Another topic containing words such as love, play, sing, hear, summer and favourite was interpreted as being more closely associated with ‘Emotions and Experience’. There were other topics related to ‘Formats and Methods’, another more generally about ‘Listening’, and finally one that appeared to relate to The Harkive Project itself, with words such as post, tweet, project, blog, and write.

	Time & Place	Contributing to Harkive	Listening	Formats & Methods	Emotions & Experience
1	morn	time	play	album	song
2	home	post	sound	radio	love
3	nowplay	tweet	record	spotify	play
4	ipod	video	track	listen	time
5	offic	head	tri	playlist	vinyl
6	tune	youtub	day	track	boy
7	shuffl	hour	album	car	live
8	walk	look	peopl	bit	sing
9	thank	project	actual	iphon	hear
10	soundtrack	write	friend	mix	band
11	headphon	juli	hear	stream	start
12	train	stori	time	start	summer
13	pop	watch	world	bbc	sun
14	commut	night	found	get	version
15	bus	blog	feel	itun	favourit
16	whilst	stuff	heard	phone	beauti
17	drive	check	lot	week	girl
18	hot	bit	decid	even	litti
19	singl	final	band	enjoy	blue
20	danc	late	piec	soundcloud	make

Figure 1: Top 20 frequently occurring words in each LDA topic. Colour-coded according to research interpretation, with words interpreted as being most closely aligned to the interpreted topics colour-coded.

To explore these initial results further it was also possible to visualise them alongside other variables, including the results of other unsupervised analyses. In Figure 2, for instance, we can see how topic allocation plays out according to the various channels by which texts were gathered, suggesting that texts gathered from Facebook, email and Flickr contained a higher proportion of texts allocated to the ‘Emotions and Experience’ topic by LDA processing.

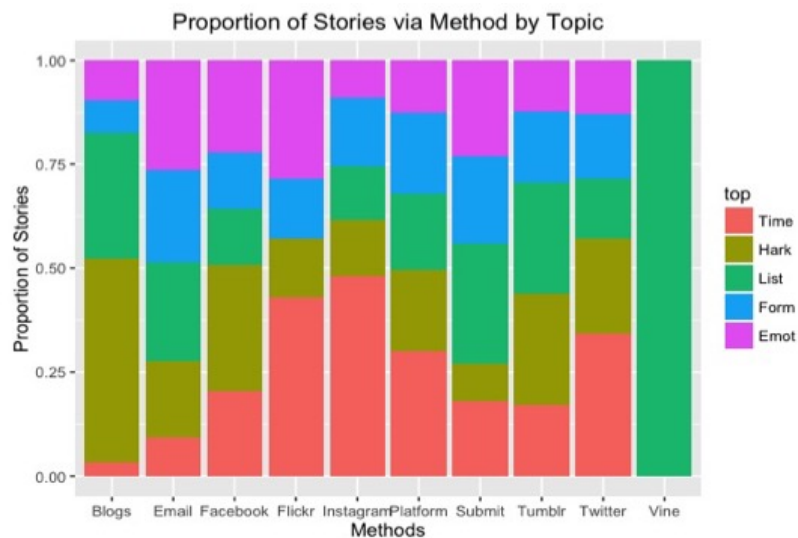


Figure 2: Proportion of texts gathered via each automated collection method according to LDA topic allocation

Figure 3 visualises the LDA results alongside another unsupervised algorithmic analysis, sentiment analysis, and suggested a high concentration of texts in the ‘Time and Place’ topic clustered around a neutral sentiment, while ‘Emotions and Experience’ texts appeared to exhibit a wider spread of positive and negative scores. Figure 4, meanwhile, visualises some frequent words revealed by the LDA process in terms of trends across the lifetime of the project. The rise in mentions of Spotify and vinyl since 2013, along with the downward trend in mentions of iTunes and downloads, appears to tally somewhat with what appears to be happening in popular music consumption more widely.

[footnote] [15] See, for instance, IFPI reports on the revenues derived from recorded music: <http://www.ifpi.org/news/IFPI-GLOBAL-MUSIC-REPORT-2018> [footnote]

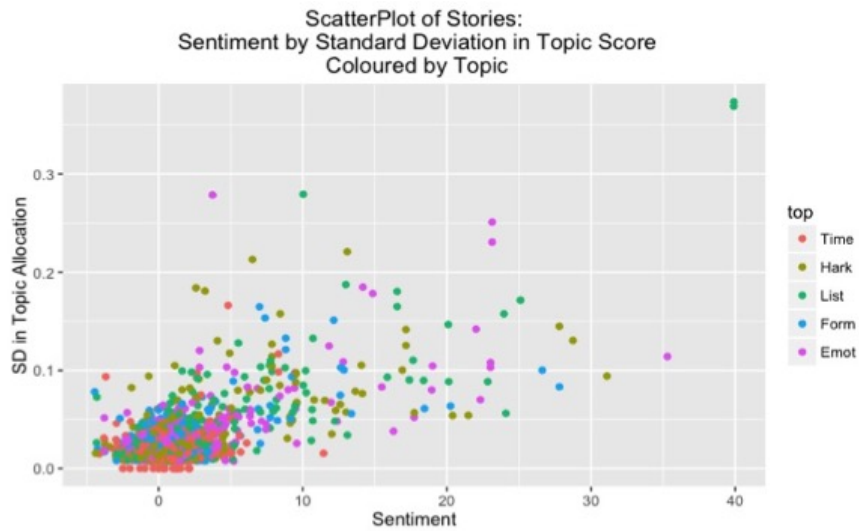


Figure 3: Scatterplot demonstrating relationship between Sentiment Analysis scores and Standard Deviation in Topic Allocation, coloured according to LDA Topic Allocation.

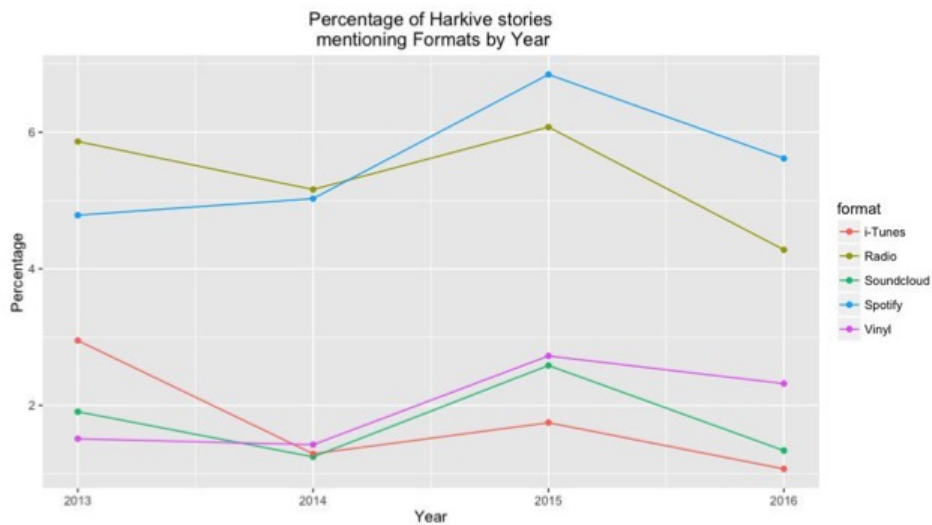


Figure 4: Percentage of texts containing keywords across each of the years 2013–2016, inclusive.

Although these initial results and exploratory visualisations are potentially useful and may lead to further questions, it was still necessary to delve a little deeper, both in terms of finding interesting things about music reception, but also in terms of exploring how the process of analysing data in this way produces forms of knowledge. An example of a potentially useful route towards this can be seen if we go back to the LDA topic allocation outlined in Figure 1 and look specifically at words relating to the ‘Formats and Methods’ topic. We may notice, for instance, that the word *vinyl* appears instead amongst the ‘Emotions and Experience’ topic, and that words related to mobile-based digital listening formats and technologies, such as *ipod*, *shuffle* and *headphones*, appear in the ‘Time and Place’ topic. An interesting initial observation is that these allocations seems to perhaps fit with some more widely held notions of those two modes of listening as having distinct characteristics: vinyl is often associated with a more ‘warm’, ‘authentic’ and ‘real’ experience of listening; digital listening by comparison is seen as cold, distracted, and functional. This may suggest that the texts gathered by Harkive perhaps contain descriptions that repeat similar ideas, and moreover that the LDA process has been efficient in revealing this. But is that indeed the case?

	Time	Hark	List	Form	Emot
1	0.18867925	0.20754717	0.20754717	0.18867925	0.20754717
2	0.18867925	0.18867925	0.18867925	0.22641509	0.20754717
3	0.21428571	0.17857143	0.23214286	0.17857143	0.19642857
4	0.18518519	0.2037037	0.18518519	0.24074074	0.18518519
5	0.18867925	0.20754717	0.18867925	0.18867925	0.22641509
6	0.19230769	0.19230769	0.23076923	0.19230769	0.19230769
7	0.20754717	0.18867925	0.18867925	0.18867925	0.22641509
8	0.19642857	0.21428571	0.19642857	0.19642857	0.19642857
9	0.17857143	0.17857143	0.23214286	0.19642857	0.21428571
10	0.22222222	0.22222222	0.18518519	0.18518519	0.18518519

Figure 5: LDA Topic Allocation scores for the first 10 documents in the corpus. Each row produces 5 values, each associated with a topic, that are divisions of a total score of 1.

In fact, a closer look at the numbers behind such an observation suggests that the differences between documents, and thus their alignment with discrete topics, were perhaps less stark than the corpus-wide overview suggested. The topic modelling process is based on the assumption that documents within a corpus exhibit relationships to *all* topics in varying degrees, and the figures in Figure 6 show those relationships in terms of a total score that is equal to 1. We may note that there are some very marginal differences between topic allocations, and so corpus level observations – in this case about digital and analogue listening methods – need to be augmented by a further, close reading before any claims could be reasonably made. Do the assumptions we carry about the technologies of vinyl records and the iPod, for instance, that appear to have been foregrounded and validated by the results of the

computational reading facilitated by the LDA process, carry over into a close, manual reading of the texts concerned? In other words, how far does Blei's 'statistical lens' take us in terms of theorising about a corpus? By looking more closely at texts containing mentions of vinyl, we are able to explore this further.

Case study: music reception and vinyl

Isolating from the corpus those texts containing the word vinyl returned $n=139$ entries, which represented 1.83% of the total corpus. From the following visualisations we can immediately observe two things. Figure 7 reveals that not only are stories mentioning vinyl present in all topics, but that a similar number ($n=40$) appear in the 'Time and Place' topic to those appearing in the 'Emotions and Experience' topic ($n=33$). Almost immediately these results appear to challenge the conclusions suggested by the corpus-wide analysis discussed in the previous section. We can see that vinyl is discussed in many other topics than the one it was allocated to by the LDA process, and which has been interpreted as related to ideas of the emotional and experiential. To what extent, then, do texts containing the word vinyl allocated to particular topics exhibit the characteristics our interpretation has assigned to those topics? A closer reading of the 139 texts containing the word vinyl in terms of their allocation to different LDA topics reveals some interesting results.

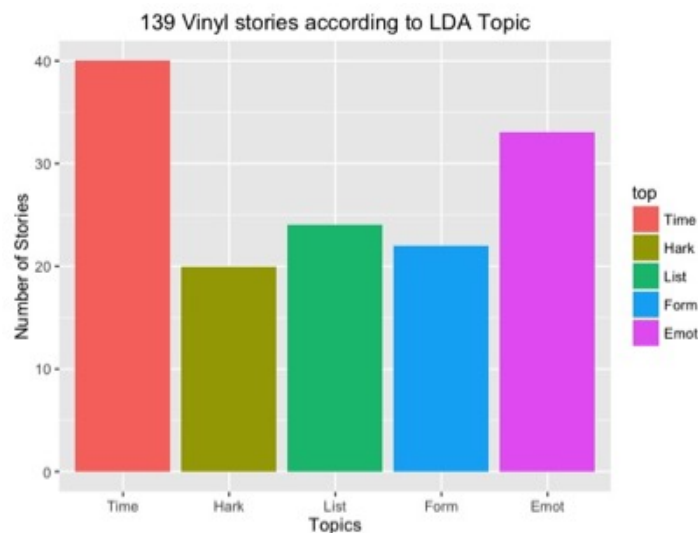


Figure 6: LDA topic allocation of texts containing the word vinyl.

Of the ($n=40$) vinyl stories contained within the 'Time and Place' topic, 20% ($n=8$) contain only information about what the respondent had played. These were texts where the respondent mentions only the name of an artist or record they were listening to at the time, and have mentioned that they were using vinyl, but have not provided any additional context. Of the remaining 32 texts, 69% ($n=22$) explicitly

reference listening whilst working, or being in domestic spaces, and in the examples below we can see vinyl records being used in mundane, everyday situations, including ironing school uniforms and hanging out washing. Considering first of all the topic allocation of 'Time and Place', the activities accompanying vinyl listening here are emblematic of what we may expect, yet appear to challenge the conclusion suggested by the corpus-level overview:

Quick coffee before next set of jobs / chores. Ornette Coleman – Twins, vinyl (#295)

[footnote] [16] All quotes and extracts from Harkive stories presented in this and subsequent chapters will be identified only by their unique story number allocated by the collection process outlined above. [/footnote]

Decided the Marvin's too uptempo for the heat. Playing 'One On One' by Bob James & Earl Klugh on vinyl LP, hanging out the washing (#7040)

Working from home today so will mainly be listening to vinyl (#5764)

Ironing school uniforms 'Toussaint' by Allen Toussaint on vinyl and what a cracker this is (#2377)

First up for @harkive 2014, Beirut, The Rip Tide on vinyl while my daughter has her breakfast (#6056)

#harkive Whilst kids getting ready for bed, had bits of Armand Van Helden 'Killing Puritans' and Faith No More 'The Real Thing' from iTunes, then a bit of 'American Supreme' by Suicide on vinyl to bring them right down...(#1125)

In the 'Emotions and Experience' topic, meanwhile, we perhaps begin to see support for the conclusion suggested by the corpus-level analysis. Although, as in the 'Time and Place' topic, some texts (33% (n=11)) contain only information about what has been played and provide no further context, of the remaining 22 texts 45% (n=10) make specific reference to the aesthetics and physicality of the 'spinning' vinyl record as an object. These features appear to play a central role in its descriptions of its use, as evidenced by the examples below:

I really want to get this Bosconi Stallions set on vinyl as it comes in a box that looks like this [photo attached] (#3180)

It's my Vertigo copy of Autobahn, I love this sleeve almost as much as the record (#330)

On to vinyl (must remember to take it off at the end, can't risk warping!) (#7643)

I listen to music in all formats, CD, Cassette, MP3 but my favourite is still vinyl. I just love the ceremony of playing a record (#4159)

I'm spending Harkive Day at home with a pile of records to listen to...I love rooting through boxes of old records looking for the next addition to my..collection; it's a great way of discovering new old music and I've gained loads of favourite artists this way (#5126)

Looking forward to getting home and spinning some new charity shop vinyl purchases for pleasure and #harkive (#3111)

A further 27% (n=6) of the texts made explicit reference to vinyl in terms of memory ('I'd forgotten how good this sounded!'). In the first example below, a longer-form text about playing a record, the respondent recalls going to a Bruce Springsteen show. Meanwhile, 50% (n=11) of the texts in this topic also discussed listening to vinyl in other experiential terms, referencing physical and emotional reactions to the music they heard on vinyl:

I remember being lent Born To Run, The River, and Greetings From Asbury Park by our very musically wise youth leader with a 'you MUST listen to these', and I did, and loved them too. But for some reason I stopped listening to Springsteen at all a few years after that, my vinyl got put away in a cupboard. (#2155)

Mrs R out again so loud vinyl time. This one really shakes the walls (#1364)

Darn That Dream' from 'Undercurrent' by Bill Evans & Jim Hall on vinyl. Music that doesn't break a sweat whilst the rest of us do (#7039)

Little Richard & His Band Pt 2 on 7' #vinyl (London 1957). Imagine hearing Tutti Frutti for the 1st time. (#71)

In the 'Formats and Methods' topic cluster, which we may recall contained almost all the other common methods through which people listen (that is, radio, Spotify, iTunes, and so on), we find 22 texts, 9% (n=2) containing only information about what has been played and provide no further context. Of the remaining 20 texts, however, 75% (n=15) discuss vinyl within the context of other technologies, services and formats. Here we can see evidence of the type of fractured and heterogeneous listening suggested by Nowak,

[footnote] [17] Raphaël Nowak, 'Investigating the Interactions between Individuals and Music Technologies within Contemporary Modes of Music Consumption,' *First Monday* 19, no. 10 (2014). [/footnote]

where meaning is derived not so much from the properties, affordances or perceptions of a particular format, but rather from the 'circuit of practices' Maguadda

[footnote] [18] Paolo Magaudda, 'When Materiality 'bites Back': Digital Music Consumption Practices in the Age of Dematerialization,' *Journal of Consumer Culture* 11, no. 1 (2011), pp. 15–36. [/footnote]

describes that together combine to constitute cultural practices around the reception of music. Vinyl listening is variously informed by the use of streaming services, or else discussed in terms of the download codes that now accompany new vinyl releases:

I guess I make most of my musical discoveries in the evenings and then do further investigation/research during the day via Spotify (before deciding to either download the album from eMusic or buy on CD/vinyl) (#5822)

Finally streaming 'Electric' by Pet Shop Boys after ten annoying minutes of having to download Pandora on a new phone and resetting my password. If it's good I'm pre ordering the vinyl. 11:34 Well that was fantastic. Aside from the ads (#162)

This morning was something from Wolves In The Throne Room, I got their new album on LP from my local indie vinyl place last week and the handy digital download means the songs live in all my devices immediately (#5560)

Taken together this brief analysis of the small number of texts containing the word vinyl demonstrates that this particular format can be seen as being described by respondents in many other ways than the initial corpus-wide computational analysis

and subsequent clustering and visualisation suggested. As well as evidence of vinyl providing the types of visceral, aesthetic and emotional experiences that Barthamski and Woodward

[footnote] [19] Dominik Bartmanski and Ian Woodward, 'Vinyl: The Analogue Record in the Digital Age' (London: Bloomsbury, 2014). [/footnote]

suggest are defining factors of vinyl use, we can also see it being used as a background accompaniment to more mundane, everyday activities – something perhaps more commonly associated with digital technologies.

[footnote] [20] See, for example, Jonathan Sterne, 'The mp3 as cultural artifact', *New media & Society*, 8(5), 2006, pp. 825–842. [/footnote]

We also see vinyl being used as part of more complex listener practices that make use of different formats and technologies as part of individualised cultural practices. An interesting thing to observe also is the apparent absence of any qualifying statements regarding the 'better' or 'more authentic' qualities of vinyl listening in comparison to other formats and technologies. In the main we see the use of vinyl records as being narrated by respondents as part of a complex, wider landscape of music reception in a variety of everyday contexts – including in commuting, working, social and other situations. These everyday situations are what Felski calls the 'mundane activities that frame our forays into more esoteric or exotic worlds',

[footnote] [21] Rita Felski, 'The invention of everyday life'. *New formations*, (39), pp. 13–31. [/footnote]

and here – recalling again Maguagga – we can see that vinyl is but one way the use of available technologies provide the means through which these forays occur. Perhaps more pertinently, this section has also shown that computational analyses, if used uncritically and unreflexively, could be used to generate results that may help reinforce or otherwise appear to support assumptions we may have – in this case, around the manner in which vinyl is used – when in actuality the detail beneath the abstractions reveals that in individual cases those assumptions are problematic. A key reflexive observation enabled by this analysis is that computational processes have both numerous benefits and limitations when used in humanities research and with cultural texts. The responsibility for ascertaining the difference between these two extremes falls to the critical researcher, who is required to make careful decisions at each stage of the research process, and particularly when it comes to interpreting the results of computational analyses.

Reflection and discussion

In reflecting upon the issues and questions that have informed the development of my method, I must consider first of all how as a researcher I initially lacked the technical skills required to collect, prepare and analyse data in the manner I had identified as being of potential use. As such my project became as much about how to conceive of new methods for studying the reception of music as it did about studying music reception.

The approach I arrived at drew upon similar methods to those used in the commercial environments of popular music, and to methods associated with the computational turn in humanities research.

[footnote] [22] David M. Berry, 'The Computational Turn: Thinking about the Digital Humanities,' *Culture Machine* 12, no. 0 (2011), p. 2. [/footnote]

Data collection processes and computational techniques have been shown in the analysis above to be in equal parts technically efficient, potentially useful in question formation, but also inherently reductive and in a manner which often prevents them from capturing and accurately reflecting complex cultural practices. In particular, text-based, qualitative data is a difficult form of data to process using computational methods and can lead to results that are problematic. I must recognise also, however, the extent to which the different modes of analysis afforded by my chosen method have enabled me to arrive at different forms of insight (and further questions) that may not have arisen through methods usually associated with the humanities alone. In other words, the observations derived from such an approach speak as much to the approach itself as they do to their application.

I still do not consider myself a coder, or a data scientist, and have attempted to utilise some complex mathematical processes from an under-privileged viewpoint. This is perhaps representative of a wider problem in the humanities when it comes to work of this kind, where as scholars we are attracted to the affordances of large datasets and computational techniques through their increasing availability and falling barriers to entry, but are simultaneously ill equipped to adequately explain and explore those methods of analysis.

By publishing regular notes on The Harkive Project website, or through the creation of interfaces such as the recently created [data explorer](#) – a resource providing both interactive access to the Harkive data and analysis and also raw code and instructional videos that enable replication of my work – I am attempting to reveal and reflect upon what Sandvig and Hargittai call the messy benchwork

[footnote] [23] Christian Sandvig and Eszter Hargittai, 'How to Think about Digital Research,' *Digital Research Confidential: The Secrets of Studying Behavior Online* (Cambridge, MA: MIT Press, 2015), p. 1. [/footnote]

involved when attempting to put such techniques to use. They argue that – apart from in ethnographic work – there is very little notion of ‘bench science’ in the humanities and social sciences, but that there should be. Their point is that the ‘workaday’ practices of our research processes need to be highlighted, particularly in areas of work that look at digital media and the internet, because these are producing the ‘new methods, new opportunities, and new challenges for understanding human behavior and society.’ As the authors state, the desired outcome is a space where ‘researchers can reveal the messy details of what they are actually doing, aiming towards mutual reflection, creativity, and learning that advances the state of the art’.

[footnote] [24] Christian Sandvig and Eszter Hargittai, ‘How to Think about Digital Research,’ *Digital Research Confidential: The Secrets of Studying Behavior Online* (Cambridge, MA: MIT Press, 2015), p. 5. [/footnote]

It is my hope that my work around The Harkive Project may make a small contribution in this regard.

Conclusion

The analysis in this chapter will, I hope, have helped raise new questions for the reader on a general level about how we may approach the complexity of contemporary music reception and – specifically – how we may begin to critically engage with the data-derived technologies and practices that now play a key role in those music reception activities. My aim has been to provide both an argument for and perhaps even a suggested route towards a more practical engagement with data, digital, and internet technologies. The intention has been to suggest new means by which scholars may be prompted to think about the efficacy of data systems when they are applied to popular music, to challenge numerous assumptions around what a data point can and does represent, and to consider ways in which we may formulate new ways of working that are able to critically engage with digital, internet, and data technologies through practical work that attempts to understand their operations, benefits, and consequences.

In terms of the practicalities of method that may underpin such an approach, in reflecting upon the process of developing and performing my research, I can offer in the first instance that the process of learning involved with attempting work of this kind is hard – particularly if the researcher is approaching it from scratch, as I did – but, equally, that it is possible to arrive at a point where such work can be undertaken. I have shown also that when approaching data-related research projects, thinking about

collection, organisation and analyses as one interrelated process is extremely useful, and probably essential.

More specifically, unsupervised machine-learning algorithms such as topic modelling can efficiently help reveal trends and patterns within a text-based corpus. However, although the results on the surface appear unequivocal, a closer examination of the numbers behind any claims related to those results reveals problems, and further questions. However, exploring both the results and the process nevertheless points towards potentially fruitful lines of enquiry that can be facilitated by further computational techniques, such as the automated extraction of specific elements based on results, keywords, and other variables, which is particularly useful with large datasets. Such processes, as I have shown, can simultaneously assist with question formation and – on a practical level – with the foregrounding of texts in large corpora. As I hope to have shown, however, it was only when considered alongside manual, close readings of texts – facilitated in part by automated process – that the results of computational processing were fully understood. As such computational techniques should be used to augment, rather than replace, close readings of texts.

As I continue to develop my own analyses into popular music reception, learning from one coding error after another through experimenting with practice and method, the limitations and affordances of working in this way are slowly revealed in the process of writing my notes, in the detail of the painfully slow, step-by-step benchwork. It is my hope that other scholars may find the resources I have made available on [The Harkive Project website](#), and through interfaces such as the [2013–17 Data Explorer](#), of use. They may also find that work of this kind can be replicated, built upon, and may lead to the creation of new questions, collaborations, and projects as together we endeavour to understand the role of digital, data, and internet technologies in the field of popular music studies.

Select bibliography

Ananny, Mike. 'Toward an Ethics of Algorithms Convening, Observation, Probability, and Timeliness,' *Science, Technology & Human Values*, 2015, 0162243915606523.

Berry, David M. 'The Computational Turn: Thinking about the Digital Humanities,' *Culture Machine* 12, no. 0, 2011, p. 2.

Blei, David M. 'Topic Modeling and Digital Humanities,' *Journal of Digital Humanities* 2, no. 1, 2012, pp. 8–11.

Boyd, Danah and Crawford, Kate. 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon,' *Information, Communication & Society* 15, no. 5, 2012, pp. 662–679.

Felski, Rita. 'The invention of everyday life'. *New Formations* (39), 1992, pp. 13–31.

Hamilton, Craig. 'The Harkive Data Explorer v1.0', Web application / Research Database / Scholarly Resource, <http://www.harkive.org/data1317>, accessed 12 February 2019.

Negus, Keith. *Popular Music in Theory: An Introduction*. Middletown: Wesleyan University Press, 1997.

Nowak, Raphaël. 'Investigating the Interactions between Individuals and Music Technologies within Contemporary Modes of Music Consumption,' *First Monday* 19, no. 10, 2014.

Prey, Robert. 'Henri Lefebvre and the Production of Music Streaming Spaces,' *Sociologica* 9, no. 3, 2015, pp. 1–22.

Sandvig, Christian and Hargittai, Eszter. 'How to Think about Digital Research,' *Digital Research Confidential: The Secrets of Studying Behavior Online*, 2015, p. 1.

Webster, Jack, Gibbins, Nicholas, Halford, Susan and Hrcs, Brian J. 'Towards a Theoretical Approach for Analysing Music Recommender Systems as Sociotechnical Cultural Intermediaries,' in *Proceedings of the 8th ACM Conference on Web Science*, pp. 137–145. ACM, 2016.

Wickham, Hadley. 'Tidy Data,' *Journal of Statistical Software* 59, no. 10, 2014, pp. 1–23.

Please leave a comment or share via social media by clicking on the icons.

View the [online publication details](#)

[footnote] [25] The Experience of Listening to Music: Methodologies, Identities, Histories has been Open Access funded by the Arts & Humanities Research Council (AHRC), grant AH/J013986/1, <https://ledbooks.org/proceedings2019> The collection has been peer reviewed, edited by David Rowland and Helen Barlow, and subsequently prepared for online publication by the Knowledge Media Institute (KMi) of The Open University. Published by: The Open University, Walton Hall, Milton Keynes MK7 6AA. Copyright © The Open University. First published: July 2019. ISBN: 9781473028647. PDFs displaying some of the content from the online collection is available from https://ledbooks.org/proceedings2019/getting-the-best-from-this-website/ You can experience the online publication as it was originally designed at https://ledbooks.org/proceedings2019 View copyright information relating to the publication here: https://ledbooks.org/proceedings2019/2019/02/24/harkive-project/

<https://ledbooks.org/proceedings2019/copyright> [footnote]

Copyright © The Open University