Utrecht University

June 22, 2020

Exploring Streaming Charts as Proxy Signifiers for Cultural Globalisation

A Distant Exploration with a Critical Data Studies Approach

A Master’s thesis written by **Axel Josh Hommenga**

Supervised by dr. Mirko Tobias Schäfer

Exploring Streaming Charts as Proxy Signifiers for Cultural Globalisation

Student

Axel Josh Hommenga (4088379) - a.j.hommenga@students.uu.nl

Master's programme

New Media & Digital Culture

Supervisor

Dr. Mirko Tobias Schäfer

Second Reader

Dr. René Glas

Wordcount

10583

Abstract

Traditionally, success of popular music artists is measured with economic metrics, but the dominant streaming platform Spotify measure popularity based on its own streaming metrics. This thesis critically explores these metrics and questions how this datafication transforms what we perceive as popular music. I used a list of fifty popular Dutch artists based on the traditional economic metrics, to explore in what way Spotify's metrics transform what we understand as popular music today. In this exploration, informed by cultural analytics, I found a yearly pattern in three years of Spotify´s charts that was not visible before. By building on critical data studies, this research then systematically approached Spotify as a data assemblage, which enabled me to map the data's ontology and its epistemic consequences. In the analysis I found that Spotify´s streaming charts are inconsistent, lack data descriptions and are manipulated by Spotify, making its use problematic.

Keywords

Popular Music Metrics | Critical Data Studies | Data Assemblage | Distant Reading | Cultural Analytics | Proxy Signifiers

Table of Contents

Abstract 2

Introduction 4

Limitations to consider 7

Theory: Cultural Globalisation and Critical Data Studies 9

- Measuring Culture with New Data 9
- A Critical Data Studies Approach 12
- Spotify’s data assemblage 14

Method: Critically Exploring Patterns in Spotify’s Charts 17

- A Distant Exploration of Dutch Cultural Globalisation 17
- Data Retrieval & Visualisation 18
- Critical Data Studies Approach towards Spotify´s Charts 20

Analysis: Popular Music in Spotify’s Terms 22

- Spotify’s Popularity Figures 22
- Ghosts in Spotify’s Data 24
- Popular Dutch Artists in Spotify’s terms 26
- Representativeness of the Charts 30

Conclusion 35

References 38

Appendix 41

- Appendix 1: Spotify’s data assemblage 41
- Appendix 2: The data set 44
- Appendix 3: R code 45
  - R code used for scraping the charts 45
  - R code used for accessing Spotify’s API 47

Introduction

Traditionally, success of popular music artists is determined based on economic metrics like the sales of albums, number of tickets sold and the airplay on the radio, whereby the most popular artists are represented in popular music charts like the Dutch Top 40 or the United States Billboard Charts.[1] However, when Spotify´s streaming platform started catering to a global market of 79 countries with free and premium access to popular music, packaged in playlists instead of albums, the way music is consumed changed. The way popularity is measured thereby transformed too.[2] When a song is consumed, Spotify counts it as a stream and the songs with the most streams will end up in the streaming charts.[3] Such charts become a representation of popular music consumption, according to Ernest Hakanen.[4] The aforementioned traditional charts have started implementing these streams as measures that influence the popularity rankings, yet I argue that a critical analysis of these metrics is lacking.[5]

This thesis explores the publicly available metrics that Spotify uses to describe popular music consumption on its platform. Specifically, it investigates how this datafication produces new power relations in contemporary popular music industries. This is done from the perspective of the platform economy. David Nieborg and Thomas Poell explain that platforms are designed to keep their end-users engaged on the platform.[6] Spotify keeps its users engaged by offering them music consumption, artists biographies, the lyrics to the songs, and an calendar of concerts of their favourite artists.[7] The authors continue that platforms measure these practices, in order to determine which content is successful and which is not.[8] This thesis is particularly interested in the data that Spotify uses to count the consumption of popular music and thereby measure popularity. As Ian Hacking stresses, what data represent is dependent on the metrics used and in what way these are interpreted.[9] The way consumption is measured and represented therefore effects it’s meaning. From my personal Dutch perspective, this research

---

[1] Ernest A. Hakanen, ‘Counting down to Number One: The Evolution of the Meaning of Popular Music Charts’, *Popular Music* 1, no. 17 (1998), https://www.cambridge.org/core/journals/popular-music/article/counting-down-to-number-one-the-evolution-of-the-meaning-of-popular-music-charts/AD2DF9B7B6F5761E029376E2A92EEDAE.
[2] Tiziano Bonini and Alessandro Gandini, ‘“First Week Is Editorial, Second Week Is Algorithmic”: Platform Gatekeepers and the Platformization of Music Curation’, *Social Media + Society*, 2019, https://doi.org/10.1177/2056305119880006.
[3] ‘Spotify Charts’, accessed 14 January 2020, https://spotifycharts.com/regional.
[4] Hakanen, ‘Counting down to Number One: The Evolution of the Meaning of Popular Music Charts’.
[5] The Dutch Top 40 is for instance constantly changing the way their charts are constructed and started taking streaming metrics into account from 2014, see www.top40.nl/geschiedenis.
[6] David B Nieborg and Thomas Poell, ‘The Platformization of Cultural Production: Theorizing the Contingent Cultural Commodity’, *New Media & Society* 20, no. 11 (2018): 4275–92, https://doi.org/10.1177/1461444818769694.
[7] ‘Features – Spotify for Artists’, accessed 19 June 2020, https://artists.spotify.com/features.
[8] Nieborg and Poell, ‘The Platformization of Cultural Production’, 4276.
[9] Ian Hacking, ‘Kinds of People: Moving Targets.’, *Proceedings of Britisch Academy* 151 (2007): 285–318.

therefore answers the following research question: how is the cultural globalisation of the top fifty Dutch popular music artists represented in Spotify's streaming charts and what does it tell us about the metrics themselves?

This study is an exploration of Spotify's charts and the way its streaming metrics are used to represent music consumption on the platform. Therein, I question how Spotify influences which artists are considered successful and how it transforms the meaning of popular music. The goal of this analysis is twofold. First, I explored how Spotify´s datafication of cultural consumption can enrich our understanding of the cultural globalisation of Dutch music. Secondly, it is a critical reflection on the alleged neutrality of the representation of cultural consumption in reductive metrics. Using data to study cultural globalisation is not something new. Cornelius Puschmann and Julian Ausserhofer investigate how accessing behavioural data from social media platforms can enrich social and cultural studies by using application programming interfaces (API).[10] Furthermore, Lev Manovich explores the potential of visual social media data to construct a rich situational awareness.[11] What these studies have in common is their use of data as a representation for cultural consumption. Randolph Kluver and Wayne Fu call these data *proxy signifiers*.[12] The authors explain that cultural research is always challenged with finding proxy signifiers to represent cultural consumption as closely and neutral as possible. These proxy signifiers are never neutral, instead which signifiers are used influences the meaning these are believed to represent.[13]

Therefore, I investigated how the data that Spotify uses to represent music consumption, the streaming metrics, influence what we understand as music consumption today. Specifically, this research is focussed on the most successful Dutch artists according to Spotify, based on the number of streams in the charts. The most successful Dutch popular music artists are already presented in a yearly report, commissioned by the Dutch organisation Buma Cultuur. In this report the global success is determined based on the economic proxy signifiers: copyright earnings of, musical recordings of, and live performances by Dutch artists in foreign countries.[14] In an informal chat with the author Sieb Kroeske, I asked him why the streaming charts are not considered in this report.[15]

---

[10] Cornelius Puschmann and Julian Ausserhofer, 'Social Data APIs', in *The Datafied Society. Studying Culture through Data.*, ed. Mirko Schäfer and Karin van Es (Amsterdam: Amsterdam University Press, 2017), 147–54.
[11] Lev Manovich, 'Cultural Analytics: Analysis and Visualization of Large Cultural Data Sets. A Proposal for Software Studies Initiative', 2007, https://www.mat.ucsb.edu/g.legrady/academic/courses/11w259/cultural_analyticsManovich.pdf.
[12] Randolph Kluver and Wayne Fu, 'Measuring Cultural Globalization in Southeast Asia', 2008, 335–58.
[13] Hacking, 'Kinds of People: Moving Targets.'
[14] Sieb Kroeske, 'Exportwaarde Nederlandse Populaire Muziek', Research Report (Buma Cultuur, 2020).
[15] Via LinkedIn I got in contact with Sieb Kroeske, after which we met in a skype call. I took notes during this informal chat.

He explained that these charts are too vulnerable for manipulation and that streaming does not generate much income for the industry. In my analysis of Spotify's streaming metrics, I discovered different ways how manipulation can be an issue. However, merely considering the economic metrics has its own assumptions and limitations too, especially because platforms like Spotify now offer music consumption for free.[16] Using solely economic metrics does therefore not represent the cultural globalisation of popular music entirely. A critical exploration of other signifiers like streaming metrics is thus needed.

This analysis touches upon in a wider exploration of quantitative data sources in order to map cultural change, which has come to be known as *cultural analytics*.[17] In his introduction of cultural analytics, Manovich explains how distant reading the online dissemination of cultural products can enable to reveal cultural change within cultural periods.[18] He describes how such an analysis can guide a researcher to dive deeper into a specific pattern, by zooming in and out of it. Within this research, I applied an explorative distant reading method to Spotify's streaming charts in order to find patterns in the streaming metrics. When a presumed pattern was discovered I dove deeper into Spotify´s datafication by accessing Spotify´s API, which provided more data signals that described the popular music and its popularity according to Spotify. The results from the aforementioned report by Kroeske, a list of the fifty most popular Dutch artists in 2018, was used as a way to explore how successful they are in Spotify's terms. Thereby, it is not a comparative analysis, but I used this list as guide for the exploration of Spotify´s metrics. The latter is executed by analysing three years of weekly updated Spotify charts from 28 distinct countries and one aggregated *global chart*, in the years 2017 until 2019.

These data were then investigated with a Critical Data Studies approach (CDS). Within CDS I focussed on the criticism towards the emerging role of data and data-driven platforms in shaping the way cultural consumption is datafied, because this process is not as neutral, raw, and valuable as often believed.[19] With this critical lens I approached Spotify as a *data-assemblage*. This enabled me to question how Spotify's streaming data come into existence and what the epistemic consequences for Dutch popular music are.

---

[16] 'Premium vs. Free – Videos – Spotify for Artists', accessed 8 April 2020, https://artists.spotify.com/videos/the-game-plan/premium-vs-free.
[17] Manovich, 'Cultural Analytics: Analysis and Visualization of Large Cultural Data Sets. A Proposal for Software Studies Initiative'.
[18] Ipidem.
[19] Andrew Iliadis and Federica Russo, 'Critical Data Studies: An Introduction', *Big Data & Society* 3, no. 2 (2016): 1–7, https://doi.org/10.1177/2053951716674238; Rob Kitchin and Tracey Lauriault, 'Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work', 2014, https://papers.ssrn.com/abstract=2474112; Andrew Piper, 'There Will Be Numbers', *Journal of Cultural Analytics*, 2016, https://doi.org/10.22148/16.006; Lisa Gitelman, *Raw Data Is an Oxymoron* (Cambridge: The MIT Press, 2013), https://doi.org/0262518287.

Practically, this means that I focussed on two processes that together constituted Spotify's charts and the streaming metrics in it. First, I questioned how the data are interpreted in a *looping* process, which revealed the subjectivity of working with these data. Second, I scrutinized how the data are collected and ultimately how this results in the creation of new norms. This was done by mapping the metrics according to multiple *engines* that contribute to the collection and processing of the charts and its streaming metrics. This approach enabled me to understand how Spotify's streaming metrics "do active work in the world," with the aim of increasing data literacy in the realm of popular music.[20]

This thesis demonstrates how Spotify's standardized streaming metrics can serve as proxies to measure the cultural globalisation of popular music, but it also reveals contingencies, inconsistencies and limitations in Spotify's streaming data that influences its meaning. I discovered that Spotify's charts represent a seasonal rhythm of Dutch popular music, because their music is streamed the most during the summer. This enriches our understanding of the role of these artist in today's globalised popular music cultures. However, I also found that Spotify's streaming metrics are filled with standardised assumptions about what the data represent. Thereby, these assumptions construct new power relations, which arguably transforms global music industries with Spotify in a governing position. This thesis should not be seen as an attempt to describe the cultural globalisation of Dutch popular music industries in its entirety. What is missing, I argue, is a critical exploration of streaming metrics in order to grasp how this contemporary datafication of music consumption changes popular music industries, its consumption, and cultural analysis. That is what this thesis provides. It can guide future research in the changing context of cultural consumption and specifically in the role of datafication in it. Furthermore, it can help actors within cultural industries to rethink their dependency on global data-driven platforms like Spotify.

Limitations to consider

Nevertheless, there are multiple limitations which should be considered when reading this thesis. First, by focussing on the charts alone the data set is already limited. The same metrics could reveal different patterns outside of Spotify´s Top 200 lists. This threshold of a maximum of two hundreds songs did however enable to understand how Spotify is changing the way we understand popularity and global success. A second limitation was

[20] Kitchin and Lauriault, 'Towards Critical Data Studies'.

caused by technicalities, which I describe in detail in the method section. Because of it, I used weekly updated instead of daily updated charts. In a comparison of the two, I found how there is more variation in the daily charts than the weekly ones. Still, the inquiry in the weekly charts enabled to find how there is a cultural rhythm within a year, which Buma Cultuur´s (BC) representation of popular Dutch artist did not reveal yet. Thus, I was able to demonstrate how it matters which metrics are used to study the cultural globalisation of Dutch popular music. This demonstration should be understood empirically. Thereby, I leave the discussion of what 'Dutch' popular music is out of scope. This analysis shows how publicly available popular music data can be used to enrich existing studies like BC's yearly report. Thereby, I am aware that, by only focussing on BC's fifty most successful artists, I cannot say much about the export value of Dutch popular music on Spotify. Another selection could result in different outcomes. What this analysis does show, is that it matters which metrics are used to determine global successful artists, because the streaming charts show different results than BC's report.

Furthermore, this research project takes an explorative approach towards Spotify as a data-driven system. Therein, I make use of publicly available information and data. Because of this, the results should not be understood as a holistic explanation of Spotify's data. This means that some elements and apparatuses that influence the meaning of the data remained hidden to me. However, it is precisely the goal of this analysis to lay bare what is visible and what remains hidden, because this can teach us that the data should not be taken for granted. Instead there will always be hidden processes that influence the way public data come into existence. This analysis contributes to the CDS aim to develop data literacy in order to tackle precisely this problem, by becoming aware of visible and hidden elements and apparatuses. It should be understood as an empirically contribution, to which many more should and will follow.

Theory: Cultural Globalisation and Critical Data Studies

In this section I describe how *proxy signifiers* can be used to measure cultural change by introducing *cultural analytics*. From a *critical data studies* perspective, I then explain why *proxy signifiers* that have been used to measure cultural globalisation should be reconsidered. Finally, I explain how approaching Spotify as a *data assemblage* can reveal the way that Spotify's streaming metrics do work in the world. Together these steps provide a framework to answer the main research question: how is the cultural globalisation of the top fifty Dutch popular music artists represented in Spotify's streaming charts and what does it tell us about the metrics themselves?

Measuring Culture with New Data

This thesis joins an exploration of new online data sources for cultural analyses. The use of data can enable new arguments, as Lev Manovich describes.[21] In his introduction of what he calls *cultural analytics* (CA), he explains that the contemporary prevalence of quantitative datafications can be very promising for research. With CA, online data are used to map cultural change from day to day, which enables to discover patterns that otherwise remain invisible. Thereby, Manovich opposes CA to the way museums represent cultural change as a static picture of the world with a start and an end date.[22] Such static representations do not take the cultural change during these periods into account, he says.[23] Manovich explains that these cultural changes can be found by exploring online data that describe cultural practices. By doing so, he takes a distant approach to the actual cultural practices, because he finds patterns in the datafication of cultural content and its consumption instead of the cultural practice itself. Franco Moretti denotes that distant reading is "a condition of knowledge [that] allows you to focus on units that are much smaller or much larger than the text."[24] Thereby, enabling a researcher to look beyond the canon. The use of data is not only promising but a necessity in some cases too. Christopher Bail observes how social and cultural studies are dealing with declining survey rates, which encourages researchers to find new sources to continue their analysis.[25] However, what can these new data tell us?

---

[21] Manovich, 'Cultural Analytics: Analysis and Visualization of Large Cultural Data Sets. A Proposal for Software Studies Initiative'.
[22] Ipidem.
[23] Ipidem.
[24] Franco Moretti, 'Conjectures on World Literature', *New Left Review* 1 (2000): 57.
[25] Christopher A. Bail, 'The Cultural Environment: Measuring Culture with Big Data', *Theory and Society* 43, no. 3–4 (2014): 466, https://doi.org/10.1007/s11186-014-9216-5.

While online data sources can be promising, they are often messy and can be inaccurate.[26] According to Bail "the greatest challenge […] is to develop new techniques to measure the unspoken or implicit meanings."[27] He critiques the reductive character of data and questions whether data can represent cultural practices in their full complexity. Nevertheless, cultural research has always been challenged to find signifiers that represent cultural consumption as closely and neutral as possible.[28] Randolph Kluver and Wayne Fu explain how data can be used as *proxy signifiers* for cultural practices.[29] In their study of the cultural globalisation of the Southeast Asian popular movie industry, they used the number of imported and exported movies as proxy signifiers. Therein, I understand cultural globalisation as "the cultural exchange between countries," in the words of Achterberg et al., because I focus on proxy signifiers that represent which countries Dutch songs are exchanged to.[30] The authors used popular music charts from several countries as proxy signifiers for the cultural globalisation of popular music. However, Marco Caselli observes that measuring cultural globalisation like this is often met with inconsistent, diverse and incomplete data sets, because countries use different measurement protocols and data formats, or the data sets are simply not updated.[31] This changes when globally operating platforms are used as a data source.

Major platforms like Spotify operate on a global market and standardize the data that are used to measure the cultural consumption of its users.[32] In 2008 the streaming platform Spotify launched, which since then grew out to be one of the most dominant channels for popular music consumption worldwide, with more than 270 million monthly active users today.[33] On Spotify a consumption is measured when a song is streamed for at least thirty seconds and the most streamed songs are presented in Top 200 charts for every country.[34] Thereby, the platform "set[s] global, rather than local, standards," because the datafication of popular music consumption is standardized for all 79 countries

[26] Kenneth Cukier and Viktor Mayer-Schoenberger, 'The Rise of Big Data: How It's Changing the Way We Think about the World', *Foreign Affairs* 92, no. 3 (2013): 29.
[27] Bail, 'The Cultural Environment', 467.
[28] Kluver and Fu, 'Measuring Cultural Globalization in Southeast Asia'.
[29] Ipidem.
[30] Peter Achterberg et al., 'A Cultural Globalization of Popular Music? American, Dutch, French, and German Popular Music Charts (1965 to 2006)', *American Behavioral Scientist* 55, no. 11 (2011): 590–91, https://doi.org/10.1177/0002764211398081.
[31] Marco Caselli, 'Nation States, Cities, and People: Alternative Ways to Measure Globalization', *SAGE Open* 3, no. 4 (2013): 2158244013508417, https://doi.org/10.1177/2158244013508417.
[32] Anne Helmond, 'The Platformization of the Web: Making Web Data Platform Ready', *Social Media + Society* 1, no. 2 (2015): 1–11, https://doi.org/10.1177/2056305115603080.
[33] Horacio Gutierrez, 'Annual Report 2019 - Spotify Technology S.A.', Annual Report (Washington, D.C.: Spotify Technology S.A., 2019).
[34] 'Stats – FAQ – Spotify for Artists', accessed 22 January 2020, https://artists.spotify.com/faq/stats.

where the platform operates.[35] This makes cultural analysis presumably easier, because the charts can all be collected from the same data source.

However, I argue that the number of streams represent substantially different behaviour than the number of album sales. Traditionally, charts are based on economic metrics like the money earned with album sales, whereby the most popular artists are represented in popular music charts like the Dutch Top 40 or the United States Billboard Charts.[36] Today the same charts are taking streams into account too, but can these metrics be combined?[37] One can purchase an album or an individual track, yet this does not tell us anything about its consumption. On the contrary, a stream is counted when a Spotify user actually streamed a song. Besides this difference, Luis Aguiar and Joel Waldfogel stress that streaming influences the more traditional metrics, because music consumption is increasingly enjoyed on streaming platforms and less by e.g. listening to an album on a CD.[38] Because of this shift towards streaming platforms, it makes sense to consider the metrics that platforms provide as proxies. Nevertheless, these numbers represent distinct modes of music consumption.

Popular music is not bundled in the same way on platforms as in CD´s. In the context of the news industry Poell and Nieborg describe how "platform integration has led to large-scale content unbundling."[39] While news items used to be solely consumed in the context of a specific printed newspaper, platforms now digitally recommend specific items to individual users.[40] A single item thereby becomes 'unbundled' from its original context. While they mainly focussed on the news industry, I think the same process can be observed in the popular music industry. Tiziano Bonini and Alessandro Gandini namely found that most Spotify users discover and consume popular music through playlists.[41] Thereby, the songs are unbundled from the traditional albums, and re-bundled into lists with songs from other artists.[42] Eriksson et al. describe that these lists are curated to fit specific "taste-profiles."[43] According to Tiziano Bonini and Alessandro Gandini this curational process is constantly informed by multiple data-signals and, as

---

[35] Nieborg and Poell, 'The Platformization of Cultural Production', 4285.
[36] Hakanen, 'Counting down to Number One: The Evolution of the Meaning of Popular Music Charts'.
[37] Stichting Nederlandse Top 40, 'Geschiedenis Nederlandse Top 40', Top40.nl, accessed 21 January 2020, https://www.top40.nl/geschiedenis; 'Billboard Finalizes Changes to How Streams Are Weighted for Billboard Hot 100 & Billboard 200', Billboard, 1 May 2018, http://www.billboard.com/articles/news/8427967/billboard-changes-streaming-weighting-hot-100-billboard-200.
[38] Luis Aguiar and Joel Waldfogel, 'As Streaming Reaches Flood Stage, Does It Stimulate or Depress Music Sales?', *International Journal of Industrial Organization* 57 (2018): 278–307, https://doi.org/10.1016/j.ijindorg.2017.06.004.
[39] Nieborg and Poell, 'The Platformization of Cultural Production', 4287.
[40] Ipidem.
[41] Bonini and Gandini, '"First Week Is Editorial, Second Week Is Algorithmic"'.
[42] Maria Eriksson et al., *Spotify Teardown. Inside the Black Box of Streaming Music* (London: The MIT Press, 2019), 116.
[43] Ipidem.

Eriksson et al. demonstrate, some data-signals are privileged over others.[44] What the streaming metrics represent is thus influenced by the way music is consumed, bundled and datafied.

A critical perspective towards this datafication, clarifies how Spotify´s metrics are manipulated by the platform. David Berry explains that datafication is a "simplification and standardisation of the external world so that it can be stored and manipulated within code."[45] This is a manipulation of the data that begins with the way Spotify uses data to determine which songs will end up in the playlists and subsequently have the highest chance of getting in the charts. For example: more than half a million Spotify users are subscribed to the hip-hop playlist "Woordenschat."[46] Getting your songs in these playlists therefore equals reaching a wide audience, subsequently enhancing the chance of getting in the charts. In interviews with multiple Dutch artists, the newspaper *de Volksrant* describes how Dutch artists are changing their music in order to be selected for the playlists by Spotify´s processes.[47] This illustrates the perceived power of Spotify in determining the success of artists. In other words, Spotify is not only representing music consumption in streaming charts, but also an actor that influences the charts with data-driven playlists.

A Critical Data Studies Approach

Representing cultural practices with proxy signifiers is never a neutral process. Data are often perceived to be raw, open and transparent, but as Lisa Gitelman denotes, data never are.[48] Based on this critical perspective I believe that it is important to understand how the meaning of the data is constructed. Therefore, I build upon the emerging field of *critical data studies* (CDS). This is a research and thinking approach "that applies critical social theory to data to explore the ways in which they [data] are never simply neutral, objective, independent, raw representations of the world, but are situated, contingent, relational, contextual, and do active work in the world."[49] In their introduction of CDS, Andrew Iliadis and Frederica Russo elaborate how it is concerned with "the identification

---

[44] Bonini and Gandini, '"First Week Is Editorial, Second Week Is Algorithmic"'; Eriksson et al., *Spotify Teardown. Inside the Black Box of Streaming Music,* 101–2.
[45] D. Berry, *The Philosophy of Software: Code and Mediation in the Digital Age* (New York: Palgrave Macmillan, 2011), https://doi.org/10.1057/9780230306479.
[46] 'Woordenschat', Spotify, accessed 17 June 2020, https://open.spotify.com/playlist/37i9dQZF1DX19xRtMyA5LM.
[47] Haro Kraak, 'Hoe streamingdiensten als Spotify de muziek veranderen', *de Volkskrant*, 19 November 2017, sec. Cultuur & Media, https://www.volkskrant.nl/gs-b10594f9.
[48] Gitelman, *Raw Data Is an Oxymoron*.
[49] Kitchin and Lauriault, 'Towards Critical Data Studies'.

of social data problems, the design of critical frameworks for addressing social data problems, and the application of social solutions to increase data literacy."[50] I join their critical exploration by studying how Spotify's streaming metrics "do active work in the world," with the aim of increasing data literacy in the realm of popular music.[51]

In the analysis I therefore critically analyse the *representativeness* of the data. As I described above, a stream represents different behaviour than the purchase of an entire album. Achterberg et al. use charts as proxy signifiers for cultural globalisation yet pay no attention to the way the charts are constructed. While their study was finished just before streaming platforms became dominant, other factors like downloads also influence the charts, as Aguiar and Waldfogel denote.[52] This does not mean that these new data sources cannot be valuable. In his critical reflection on the use of cultural analytics (CA), Andrew Piper is positive about the exploration of public data for cultural analyses, but describes how CA often lacks a critical reflection on its evidence.[53] He explains that the data's *representativeness* should be considered, meaning that a researcher should always describe what the data do and do not represent and what the epistemic consequences of this are. In Buma Cultuur´s (BC) report, Sieb Kroeske notes that only economic metrics were used to determine which fifty Dutch artists are most successful globally. However, a decent explanation of the consequences is missing. As described in the introduction, Kroeske told me that the streaming metrics are too vulnerable for manipulation, making them questionable as proxy signifiers. Nevertheless, I argue that streaming metrics should be explored, because platforms like Spotify now offer music consumption for free.[54] When the music of a Dutch artist would spread globally, but is mostly consumed by users with a free account, this artist will not be represented in BC's report. Hence, the report´s representativeness is limited, which should always be considered when working with data.

Nevertheless, the fear of manipulation is relevant. As mentioned above, Spotify manipulates the charts by recommending specific songs above others. Furthermore, the way data from such public data sources can be accessed is a manipulation of the data too. API´s provide targeted access to the platform´s data, which has the connotation of transparency, openness and neutrality, as Bail observes, but he explains that it is in fact

[50] Iliadis and Russo, 'Critical Data Studies', 5.
[51] Kitchin and Lauriault, 'Towards Critical Data Studies'.
[52] Aguiar and Waldfogel, 'As Streaming Reaches Flood Stage, Does It Stimulate or Depress Music Sales?'
[53] Piper, 'There Will Be Numbers'.
[54] 'Premium vs. Free – Videos – Spotify for Artists'.

uncertain how the data are generated and which data the API hides.[55] Cornelius Puschmann and Julian Ausserhofer also argue that API´s govern how and which data can be accessed.[56] Therefore, critically reflecting on the way the data can be accessed is essential in order to grasp the data´s meanings. In the following part I therefore explain how studying Spotify's data assemblage enabled me to critically reflect on the way these processes influence the data's meanings.

Spotify's data assemblage

Kitchin and Lauriault explain how a data-driven system like Spotify can be studied with a critical data lens, by studying its *data assemblage*.[57] They define a data assemblage as a "complex socio-technical system, composed of many apparatuses and elements that are thoroughly entwined, whose central concern is the production of a data."[58] Critically engaging with elements and apparatuses of an assemblage can reveal the work that a data-driven system does.[59] I understand a data assemblage as a structural way to map different human and non-human actors that influence the meaning of the data which the entire system produces. Therefore, I use Kitchin and Lauriault's lists of possible elements and apparatuses as a framework to start mapping them. When this mapping is done, a researcher can reflect on the way those elements and apparatuses influence the data's meanings. From that perspective, I approached Spotify as a data assemblage of which I critically studied the streaming data's ontology, or in other words the way these data came into existence, and how this influences its meaning as proxy signifiers for the cultural globalisation of Dutch popular music.

Kitchin and Lauriault explain that a data assemblage is built on two processes introduced by Ian Hacking. The first is what Hacking calls "looping effect."[60] It entails how "data are classified and organised" and in what way this effects the data´s meaning.[61] This is an ongoing looping process as data move through different systems, are interpreted by different institutions and individuals and used for multiple purposes. Critically analysing how the data move through this loop and in what way that effects its alleged meaning can reveal the work that Spotify's data assemblage does.[62] Hacking's second

[55] Bail, 'The Cultural Environment'.
[56] Puschmann and Ausserhofer, 'Social Data APIs'.
[57] Kitchin and Lauriault, 'Towards Critical Data Studies'.
[58] Ipidem.
[59] Kitchin and Lauriault, 'Towards Critical Data Studies'.
[60] Hacking, 'Kinds of People: Moving Targets.', 286.
[61] Kitchin and Lauriault, 'Towards Critical Data Studies'.
[62] Ipidem.

process is what he calls "engines of discovery, which have side-effects […] for these are engines for making up people."[63] He stresses how we keep "making up people" by making up categories which are based on the data available at that time.[64] Hacking explains that as time passes, we come up with more data to measure and quantify human lives, hence these categories are not static but "moving targets."[65] The engines of this process can be studied by describing what is counted, how it is quantified and how it sets new norms. Spotify counts music consumption with streaming metrics and uses these to quantify popularity. Spotify thus *makes up* a category of popular artists based on the processes in its data assemblage.

It is not possible to lay bare every influencing factor in a data assemblage. As an example, Spotify manipulates how the playlists are curated, which is an ongoing process of change. Thereby, I understand these data-driven playlists as elements of what the authors call the *practices* apparatus, because Spotify uses the data in their curational processes.[66] Even when solely analysing how this element influences the meaning of the streaming data, its inner workings keep changing as moving targets and it is not directly visible how. The latter is closely related to what the authors call the *governmentalities and legalities* apparatus, wherein I classified Spotify´s interface and the terms of service as elements.[67] These elements are what Poell, Nieborg and van Dijck call "governing instruments" that govern interactions with the platform's features with black box mechanisms.[68] In this black box the inner workings remain hidden to me, because I do not have access to Spotify's programming division and I did not talk to Spotify's legal department to find out why they made certain decisions in their terms of service. While some of these processes can be revealed, others will always remain hidden.

Furthermore, apparatuses like *subjectivities and communities* emphasize how this critical process is subjective in itself. Not only the subjective decisions by elements like data producers and curators influence its meaning, but the subjective decisions by third party users too. As I critically explored Spotify's data as proxy signifiers for cultural globalisation, I am constructing meaning informed by my personal experiences too. Studying Spotify's data assemblage is therefore not an objective process, but more so a subjective process of empirically exploring apparatuses and elements and how these

---

[63] Hacking, 'Kinds of People: Moving Targets.'
[64] Ipidem.
[65] Idem, 312.
[66] Kitchin and Lauriault, 'Towards Critical Data Studies'.
[67] Thomas Poell, David Nieborg, and José van Dijck, 'Platformisation', *Internet Policy Review* 8, no. 4 (2019): 8, https://policyreview.info/concepts/platformisation.
[68] Ipidem.

could influence its meaning. While this makes working with a data assemblage somewhat problematic, it also emphasizes the necessity. Because Spotify shares their data to be used by anyone, we should acknowledge how little we actually know about its ontology. Thereby, critically engaging with some elements and apparatuses, and demonstrating the subjectivity of the data, makes us more aware what the data can and cannot represent as proxy signifiers. In the following section, I describe how I did this with an explorative distant reading method with a critical data studies approach.

Method: Critically Exploring Patterns in Spotify's Charts

In this section I describe the practical steps that were taken to answer the main research question: how is the cultural globalisation of the top fifty Dutch popular music artists represented in Spotify's streaming charts and what does it tell us about the metrics themselves? This starts with a description of the explorative distant reading method that enabled me to find patterns in the cultural globalisation of Dutch popular music with Spotify's charts as proxy signifiers. Thereby, I question how *popular* a selection of Dutch popular artists is. Then I explain how the data is collected by accessing Spotify's API and web-scraping Spotify's charts. The section concludes by describing how I explored elements and apparatuses of Spotify´s data assemblage in order to reveal its epistemic consequences. With this method I could critically explore how Spotify's charts influence the meaning of popular music.

A Distant Exploration of Dutch Cultural Globalisation

The yearly Buma Cultuur (BC) report about the export value of Dutch popular music does not reveal how cultural change of Dutch popular music is happening within a year. The report merely informs us about the fifty most successful Dutch artists in one year. On the contrary, Spotify's charts are updated daily and weekly. In order to find the presumed cultural change within this year, this research is inspired by the explorative approach of cultural analytics. Thereby, this analysis takes a distant approach to the consumption of popular music by studying Spotify´s charts as proxy signifiers. This distance can reveal hidden patterns, as Manovich elaborates, that would not be visible when closely analysing every song in detail.[69] In this way this analysis aims to reveal hidden patterns of the cultural globalisation of Dutch popular music artists.

In order to move beyond the songs and discover patterns, the data is collected for a longer time span. In his work, Moretti investigates datafications of literary history over a time span of multiple decades. Thereby, he applies distant reading to literary texts in order uncover cultural change in for instance the world bank policies.[70] Where Moretti is not necessarily concerned with computed datafications of cultural practices, Manovich

[69] Manovich, 'Cultural Analytics: Analysis and Visualization of Large Cultural Data Sets. A Proposal for Software Studies Initiative'.
[70] Franco Moretti, 'Graphs, Maps, Trees. Abstract Models for Literary History.', *New Left Review* 24 (2003); Franco Moretti and Dominique Pestre, 'Bankspeak: The Language of World Bank Reports', *New Left Review* 92 (2015).

emphasizes the analytical value of using big online data sets.[71] Thereby, Manovich aims to discover patterns by analysing online cultural practices from a macro perspective. One example of his work is a project called *Selfiecity*.[72] For this project he used quantifications of online self-portraits in order to discover similarities and differences between the way people make these images in distinct cities across the world. In a similar manner I use Spotify's datafication of music consumption to discover patterns on this global platform. While he selected images randomly, this analysis specifically selected three years of Spotify charts like Moretti selected yearly world bank policies. This enables me to explore yearly patterns and thereby enrich BC's yearly report. In the analysis I build on the BC report that presents the fifty most successful Dutch artists in 2018. Hence, I collected Spotify's Top 200 charts between 2017 until 2019. This timespan starts in 2017, because this is the first year Spotify shares its data and ends in 2019. Thereby, I chose to use three full years, as 2020 was only halfway when this project was executed. Only investigating one year would be too limited, as opposed to a yearly pattern. In the next part I describe how the data were collected and the way the data were explored with data visualisations.

Data Retrieval & Visualisation

Spotify's charts were retrieved using computational means. The charts can be accessed on a public website called *Spotify charts*.[73] Spotify is available in 79 countries, which Spotify calls markets.[74] The consumption on Spotify is represented in separate charts for all of these countries. Besides, Spotify has one aggregated *global chart* that contains the most popular two hundred songs on the entire platform. With RStudio, a user-friendly programming interface of the statistical programming language R, I moved the charts from the website into an Excel file.[75] Thereby, I built upon readymade R packages that enable me to collect the data. The *Rvest* package was the most important one, because this enabled me to retrieve the charts from the website.[76] The collected data are the names of the artists, the names of the songs, the number of streams generated, and the ranks on the charts. The initial goal was to collect the daily updated charts of all the 79 countries in order to find a global pattern on the entire platform. Yet, Excel limits the number of

[71] Manovich, 'Cultural Analytics: Analysis and Visualization of Large Cultural Data Sets. A Proposal for Software Studies Initiative'.
[72] 'Selfiecity', selfiecity, accessed 17 December 2019, http://selfiecity.net/.
[73] 'Spotify Charts'.
[74] 'Spotify — Company Info', Spotify, accessed 26 October 2019, https://newsroom.spotify.com/company-info/.
[75] RStudio Team, *RStudio* (Boston, MA: RStudio Inc., 2019), https://rstudio.com/; R Core Team, *R: A Language and Environment for Statistical Computing.* (Vienna: R Foundation for Statistical Computing, 2013), http://www.R-project.org/.
[76] In the appendix you can find the specific code used.

rows to 1,048,576 in order to function correctly, which was not enough for the latter goal.[77] After I collected one year of daily charts for only ten countries the Excel file was full. Therefore, the data retrieval code was altered to collect weekly charts instead. This enabled me to collect twenty-eight local charts and the global chart for the three-year timeframe. While this is still limited, this list of countries provides empirical evidence for the countries that the BC report specifies as generating the most economic value for the Dutch music industry.[78] When the twenty-nine charts were collected, which took about thirty minutes per chart, I randomly took samples and checked these with the original data source. You can find the data set and its specifications in appendix 2. After this check, the data was ready for exploration.

This exploration starts by visualizing the data. According to Manovich this step enables to turn data into insightful patterns.[79] For this analysis, I used the data visualization tool Tableau that enabled me to visually plot the chart data in order to discover patterns throughout the three years.[80] Manovich explains that a researcher should be able to zoom in on a specific pattern and see the songs in detail, in order to create a rich situational awareness.[81] While Manovich is mostly concerned with formal visual characteristics, this analysis applies his explorative approach to Spotify's further datafication of the consumption of popular music on the platform. The level of detail that this analysis focusses on does not describe the songs themselves, but solely meta information about the song like the publication date and collaborating artists. Access towards more detail about the songs can be obtained by accessing Spotify's Application Programming Interface (API), as described in the previous section. Spotify's API was accessed with R code too, which I attached in appendix 3. For this purpose, I used the *SpotifyR* package that enabled me to retrieve structured data from the API.[82] This enabled access to the publication dates of the songs, collaborating artists, and the countries where the songs are available. Because of this detailed access, I considered to widen the definition of Dutch popular music artists towards all artists in the charts that Spotify labelled as Dutch. However, when exploring the available data, I found that this label did

[77] 'Excel Specifications and Limits', accessed 5 June 2020, https://support.office.com/en-us/article/excel-specifications-and-limits-1672b34d-7043-467e-8e27-269d656771c3.
[78] Kroeske, 'Exportwaarde Nederlandse Populaire Muziek'.
[79] Manovich, 'Cultural Analytics: Analysis and Visualization of Large Cultural Data Sets. A Proposal for Software Studies Initiative'.
[80] *Tableau: Business Intelligence and Analytics Software* (Seattle, WA: Tableau Software), accessed 18 January 2020, https://www.tableau.com/.
[81] Manovich, 'Cultural Analytics: Analysis and Visualization of Large Cultural Data Sets. A Proposal for Software Studies Initiative'.
[82] 'The Comprehensive R Archive Network', accessed 18 January 2020, https://cran.r-project.org/.

not exist in this public data set. Therefore, I focussed on the fifty artists from the BC report.

Critical Data Studies Approach towards Spotify´s Charts

In order to reveal the epistemic consequences of Spotify's datafication of popular music consumption, this analysis explores elements and apparatuses of its data assemblage. As a first step I therefore used Kitchin and Lauriault's list of apparatuses and elements in order to find which ones might influence how Spotify's charts and its streaming metrics come into existence. I did so, by examining how Spotify describes the meaning of its data on its *Spotify for Developers* environment and on its online *Spotify Community*.[83] In appendix 1 you can find the results of this exploration. The elements that I found were then used as a guide to dive deeper into the data and find empirical evidence of its epistemic consequences. This was used to critically reflect on the patterns that I found in the distant exploration of Spotify's charts. When I found a pattern, I used Hacking's two interpretative processes, the *looping effect* and the *engines*. This enabled me to grasp how the patterns become meaningful in Spotify's data assemblage. Practically, this means that that I critically examined elements that the data looped through before I could access it. In terms of the engines, I questioned how these data were counted, what these are intended to represent, and how this creates norms.

Furthermore, I applied a critical reflection towards the work that my analysis does. Cultural analytics (CA) has been criticized because data visualisations are used as representations for human behaviour, without a critical reflection towards the way the data relate to what these are used to represent.[84] From this perspective, Lindsey Caplan writes that Manovich' method is not informed by theory, which makes it unsuitable for academic research according to her.[85] I recognize this limitation, but with critical data studies as theoretical framework I argue that I am well equipped for a proper analysis. Furthermore, the explorative approach of CA enabled to go beyond what I already know and find unexpected patterns, as I demonstrate in the next section. Andrew Piper acknowledges this constraint too and explains that a CA researcher should be reflexive towards the used data and the explorative decisions and question the *representativeness*

[83] 'Home | Spotify for Developers', accessed 18 January 2020, https://developer.spotify.com/; 'The Spotify Community', accessed 7 June 2020, https://community.spotify.com/.
[84] Lindsay Caplan, 'Method without Methodology: Data and the Digital Humanities' 72, no. E-flux journal (2016), Worker01.e-flux.com/pdf/article_9006656.pdf.
[85] Ipidem.

of the analysis.[86] For that reason, I made notes of the steps that I took when collecting, processing and visualising the data. Subsequently, I questioned what this does and does not represent in the same manner as I critically explore the Spotify's charts. I aim to be as transparent about my decisions as possible, in order to invite further academic debate about the datafication of popular music or any other cultural practice.

Concluding, this section described how an explorative distant reading method with a critical data approach enabled to explore Spotify's charts and the work these do. Furthermore, it explained how an exploration of three years of Spotify´s charts can enable to discover patterns that would otherwise remain invisible. In the following section the research findings are presented.

[86] Piper, 'There Will Be Numbers'.

Analysis: Popular Music in Spotify's Terms

In this section I present how the cultural globalisation of the top fifty Dutch popular music artists is represented in Spotify's streaming charts and what it tells us about the metrics themselves. First, I demonstrate how Spotify´s metrics lack decent descriptions. Then, I share empirical evidence of the incompleteness of the data set. This is followed by a critical description of a Dutch pattern of cultural globalisation. Finally, this analysis shows how the representativeness of Spotify's charts should be questioned, because its standardized data are not as obvious and meaningful as these might seem.

Spotify's Popularity Figures

Spotify's data are not as open as these might seem. In order to get access some of the data with the API, a developers account is needed. When making the latter account, the user has to accept the "Spotify Developer Terms of Service" which serve as a governing instrument.[87] In the *terms of service* (ToS) Spotify describes "[t]he Spotify platform is provided on an 'as is' and 'as available' basis," which implies that the platform is obtaining a power relation with the users of the API functionality.[88] Namely, because the data are available only in the format that Spotify decided on and the data set can change any time. This governmentality apparatus of Spotify´s data assemblage makes the external users of these data dependent on Spotify's hidden decisions. Furthermore, in the ToS the company stresses that "Spotify expressly disclaims any warranty […] that the results are accurate or reliable or consistent with your expectations."[89] Thereby, I argue that Spotify acknowledges the interpretability and subjectivity of the meaning of all data that the platform provides, yet it also distances itself from any interpretations made by third parties. Therefore, the data are subject to a looping effect, whereby its meaning is subject to a changing and hidden classification and organisation of the data by Spotify. Furthermore, this effect continues when researchers or other third parties use these data, because every individual and organisation interprets the data from their perspective. Therein, they choose which objects of focus to prioritise in their analysis. Spotify thus leaves the interpretation of the streaming metrics open, which is problematic because data are never neutral, as Gitelman stressed.[90]

---

[87] 'Spotify Developer Terms of Service | Spotify for Developers', accessed 4 May 2020, https://developer.spotify.com/terms/.
[88] Ipidem.
[89] Ipidem.
[90] Gitelman, *Raw Data Is an Oxymoron*.

Once the ToS are accepted, the account is available for use. However, before I could access the API, a client ID has to be set up. This unique ID provides access to Spotify's API and the data behind it.[91] This can be done by filling out a form describing what you intend to use the API for. There is no explanation for the reason the company needs this information, yet when not provided the API cannot be accessed, which makes this form a governing instrument too. There are a few open fields and one list of checkboxes with the question: "what are you building?"[92] I selected "I don't know," assuming that this would reveal all data instead of highlighting specific data to match a specific purpose like Spotify features in a "TV" or "Mobile App".[93] The selection triggered a notification describing that only non-commercial use of the API is allowed. However, I found no description, note or other indications of the impact this selection would have on further use of the API. This makes the use of the data very dependent on Spotify's black box governance, while the data's ontology remains hidden.

The same counts for the way the charts come into existence. Spotify describes its *Spotify charts* website as "the best place to see the latest Spotify Chart figures."[94]

> *These figures are generated using a formula that protects against any artificial inflation of chart positions. Note: Due to this formula, you might notice the data here differs from other reported stream numbers we share.*

Thus, one of the stages of the interpretative looping process of Spotify's streaming metrics is a *formula* that influences the chart positions. Interestingly, there is no description of this formula besides the fact that it protects against artificial inflation. Imagine you are listening to one song over and over again, simply because it is your favourite. Every time you listen to it for more than thirty seconds a stream is counted according to Spotify.[95] However, as Eriksson et al. have shown this can also be automated without Spotify finding out, besides some artists asked their fans to play their music on repeat while sleeping.[96] This automatic classification if a stream is real or artificially generated could classify an actual person actively listening to a song in the same way as a person sleeping, or an algorithm pressing play. It therefore matters what kind of parameters the formula considers when spotting artificial manipulation.

[91] 'Web API | Spotify for Developers', accessed 4 May 2020, https://developer.spotify.com/documentation/web-api/.
[92] 'My Dashboard | Spotify for Developers', accessed 4 May 2020, https://developer.spotify.com/dashboard/.
[93] Ipidem.
[94] 'Stats – FAQ – Spotify for Artists'.
[95] Ipidem.
[96] Eriksson et al., *Spotify Teardown. Inside the Black Box of Streaming Music*, 97–98.

The same logic can be applied to the way a consumption is counted by Spotify. When mapping the engines of Spotify's data assemblage, I noticed how Spotify is setting a norm for what a consumption means. A stream is counted when a song "is streamed for over 30 seconds," thus implying that this is a valuable threshold without explaining why.[97] This threshold is namely used to determine the charts' positions and as a way to measure if the artists should get paid by Spotify.[98] Because of this, artists are stimulated to grasp the audience attention in the first thirty seconds. However, what happens if a track grasps the audience attention for the first thirty seconds, but the end-users get bored and skip the track after these crucial seconds? Such a stream represents different behaviour than a stream where the listener is enjoying every second of the song.

Furthermore, not all streams are the same. Spotify is reporting about the possibility for artists to *window* their music to specific end-users, enabling to target either end-users with a premium paid account or with a free advertisement-based account.[99] According to Spotify these two categories are important when determining what an artist gets paid.[100] Thereby, all the revenue from premium users is shared among the artists whose songs were streamed by premium users and the same for free accounts. In Spotify's shareholders letter the company reports more free users than premium users, but more income from premium users. Hence, a stream by a premium user is economically more valuable for an artist than a free user. Nevertheless, the presentation of streams in the charts suggests that all of these streams are the same, whilst these are not.

Ghosts in Spotify's Data

Spotify's charts are not as complete as they might seem at first hand. Spotify's presentation of three years of daily and weekly updated charts implies that the company knows how the consumption on its platform develops from day to day. Furthermore, the fact that authoritative organisations like the Dutch Top 40 and the Billboard charts are using their streaming data emphasizes that Spotify's metrics have become a significant measure in what we understand as popular music today. However, when collecting the charts, I ran into multiple computational errors. A close examination of the returned error revealed why it occurred. Apparently, some of the days and weeks were missing. Why

[97] 'Stats – FAQ – Spotify for Artists', accessed 30 March 2020, https://artists.spotify.com/faq/stats.
[98] 'Premium vs. Free – Videos – Spotify for Artists'.
[99] Tim Ingham, 'Windowing on Spotify Is Finally Happening - but Is It Worth It for the Labels?', *Music Business Worldwide*, 13 April 2018, https://www.musicbusinessworldwide.com/windowing-on-spotify-is-finally-happening-but-is-it-worth-it-for-the-labels/.
[100] 'Premium vs. Free – Videos – Spotify for Artists'.

was this data missing? On the original data source, the Spotify charts website, I did not find any explanation why the data were missing. It was simply a missing day or week in the drop-down menu on Spotify's website. It could mean that Spotify was not working at that moment, meaning that no music was consumed on Spotify at all. However, it can also mean that something went wrong in the tracking. Either way, it demonstrates that Spotify's data set is not as complete as it might seem. This impacts the popularity rankings and how the data are made meaningful. Whatever the reason for these missing data-signals is, it emphasizes how fragile and contingent the streaming metrics are, which makes its use as proxy signifiers for phenomena like cultural globalisation problematic.

The charts themselves became even more exemplary for the contingency of streaming data. When I was taking samples of the data set, to check if the values were consistent with the original data source, I found that some artist names and song names were missing in the data. These names were removed by Spotify, leaving some rankings empty. Again, the reason behind it remains hidden in Spotify's black box. In figure 1 on the next page, I visualised the stream count of these removed artists over the course of the analysed three years. This reveals how most streams by these removed artists were generated in the second half of 2017, with another small peak in December 2017. The count of streams in this first peak equals about one percent of all the streams in the charts in the same period, thus a significant part of global consumption on Spotify is now unknown. After 2017 the number of streams generated by removed artists seems to be stabilised, yet present. Interestingly, this count only started in the second half of 2017 and was non-existent before. This suggests that a significant change has been made in one of the elements of Spotify's data assemblage, yet it remains hidden which one and why.

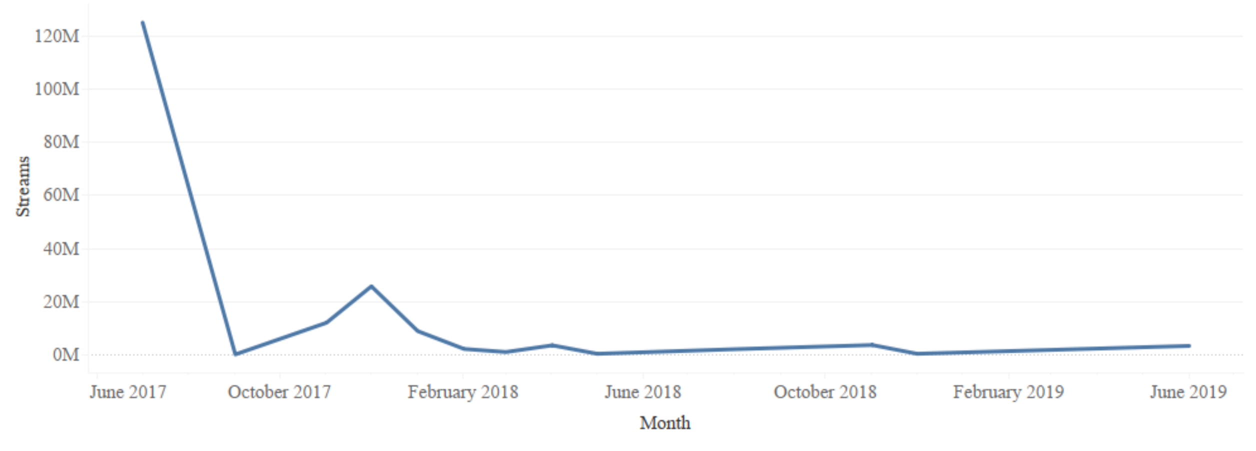

Figure 1: The number of streams for artists that have been removed from Spotify plotted over the course of three years.

In order to get a song in the global charts, it has to be streamed for about two and a half million times in one week. Because of this threshold it is not surprising that not many artists get their songs in these charts. In relation to charts from small countries (in terms of inhabitants count) like the Scandinavian countries, the Baltic states, and the Netherlands, there are between fifteen and forty four percent less artists in the global charts.[101] Therefore, it is even more interesting to see how more than forty two percent of streams by removed artists are generated by songs in those charts. In Spotify's *annual report*, they state how one of the risks for the company's existence is the fact that the music owners can decide to pull their music from the platform.[102] Therefore, I argue that there is reason to believe that this gap in the data represents artists that removed their music from the platform, as was the case for Taylor Swift in 2014 for instance.[103] Yet, we cannot be sure to which artist these numbers belong, which transforms these numbers into ghosts. This demonstrates that Spotify is dependent on external parties within their data assemblage too.

Popular Dutch Artists in Spotify's terms

With Spotify´s charts, Spotify sets norms for what popularity on the platform means. Out of the fifty Dutch artists that the analysis started with, forty-seven can be found on

[101] There are charts, like Brazil, the US, and Japan, that have less distinct artists in the charts than the global chart. Yet, I argue that the global chart is the clearest example of the argument, because it represents the most popular artists on the entire platform.
[102] Gutierrez, 'Annual Report 2019 - Spotify Technology S.A.', 11.
[103] Charlotte Alter, 'Taylor Swift Just Removed Her Music From Spotify', *Time*, 3 November 2014, https://time.com/3554438/taylor-swift-spotify/.

Spotify. Of those, only seventeen were found in the charts and of those seventeen only six Dutch artists made it to the global charts. In Spotify's terms it is therefore suggested that only six of the Dutch artists are successful globally. Here it should be noted that with this, Spotify creates a norm for what global success means. The fact that Spotify presents the most popular music in lists of two hundred songs is already accompanied with the assumption that this amount is a good threshold. Thus, a song and its artists are popular only when getting enough streams to get a position in this list. This decision in the charts looping process influences what the charts can represent. Thereby, Spotify's threshold is different from traditional charts like the Dutch Top 40, with forty songs, and Buma Cultuur's top fifty. As Spotify´s list is much larger, it is surprising that only seventeen of the selected artists are in these charts. Especially, because other research has revealed that this characteristic of Spotify's charts could even reveal songs in the more diverse *long tail*.[104]

Getting a song in the charts does not equal infinite popularity for the Dutch artists on Spotify. When plotting the stream count over the course of the three selected years, a seasonal pattern became visible (see figure 2 on the next page). The Dutch artists that obtained ranking in Spotify's charts were mostly popular on the platform during the summer period, with the most streams in august. There is one deviation in the beginning of 2017, where the Dutch artists generated more streams to get in the charts than the summer period in that same year. Nevertheless, I argue that this uncovers a summer pattern of Dutch popular music on the global platform. This becomes more apparent when I compared it with the median number of streams of all artists in the charts. This is plotted in figure 3 and does not show similar high peaks specifically during the summer months. This suggests that Spotify's users do not necessarily listen to music more during the summer months. It therefore appears to be a specific patterns for the fifty Dutch artists.

[104] Michelangelo Harris et al., 'Analyzing the Spotify Top 200 Through a Point Process Lens', *ArXiv*, 2019, 200, https://doi.org/1910.01445.

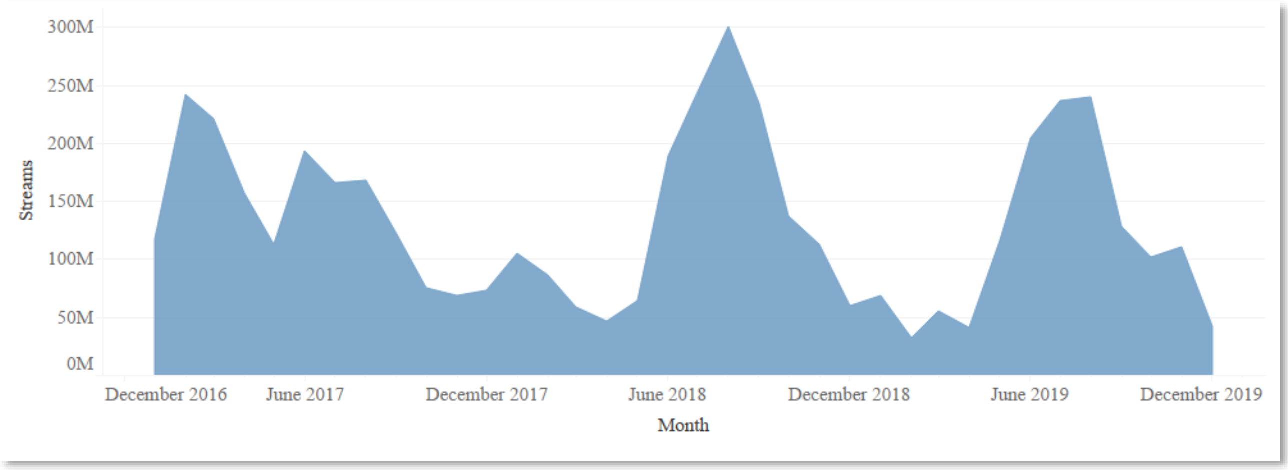

Figure 2: The number of streams by the selected Dutch artists in Spotify's charts. These are plotted over the course of three years in the chart above.

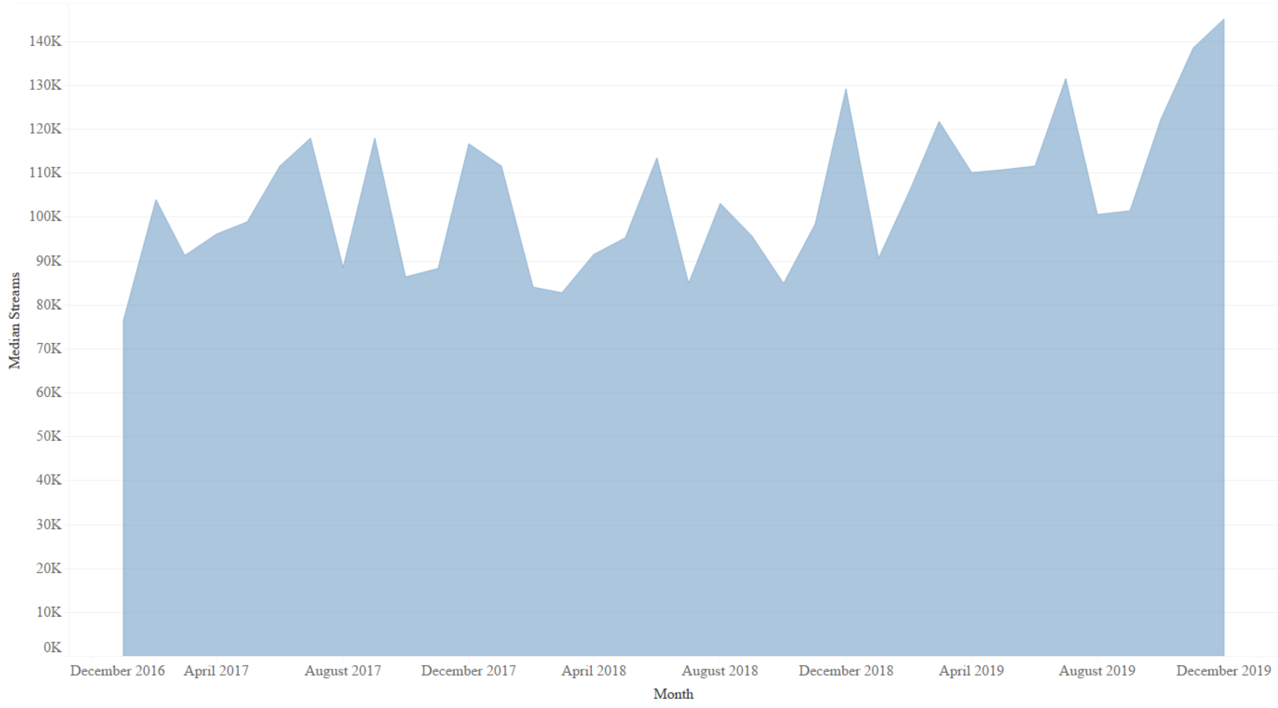

Figure 3: The median number of streams in all the charts by all the artists. These are plotted over the course of three years.

When diving deeper into Spotify´s datafication of specific Dutch songs, the contingency of Dutch artists to Spotify´s platform becomes visible. One of the Dutch artists that is most popular, based on the number of streams and charts he got his songs in, is Martin Garrix. This Dutch DJ had a total of nineteen songs in numerous Spotify charts between 2017 and 2019. In a previous study of Spotify´s Top 200 charts, Michelangelo Harris et al. argue how popularity is not only determined by getting songs in the charts, but more so by how long a song remains there.[105] In this analysis I observed how, for Garrix's songs, the latter becomes apparent in Spotify's charts. In figure 4, I visualised the number of streams for his most popular songs. This shows how the songs are getting in the charts one after the other. Furthermore, only when one song is fading out of the charts a new song enters. I argue that this is an example of the platform's "unbundling" effect, because Garrix is popular one song at the time instead of one bundled album at the time.[106] This seems to be the result of a strategy, because these popular songs were all released shortly before they entered the charts. The same counts for songs of the other Dutch artist. With the only exceptions being the *We Like to Party* by the Vengaboys (1998) and *Drowning - Avicii Radio Edit* by Armin van Buuren. All the other selected Dutch songs were released the same year as they were in the charts. This demonstrates how weekly updated charts can reveal a pattern that is not visible with aggregated yearly numbers alone. Hence, it is crucial to understand what the charts can and cannot represent, therefore the next part is dedicated to the chart´s *representativeness*.

[105] Harris et al., 200.
[106] Nieborg and Poell, 'The Platformization of Cultural Production', 4287.

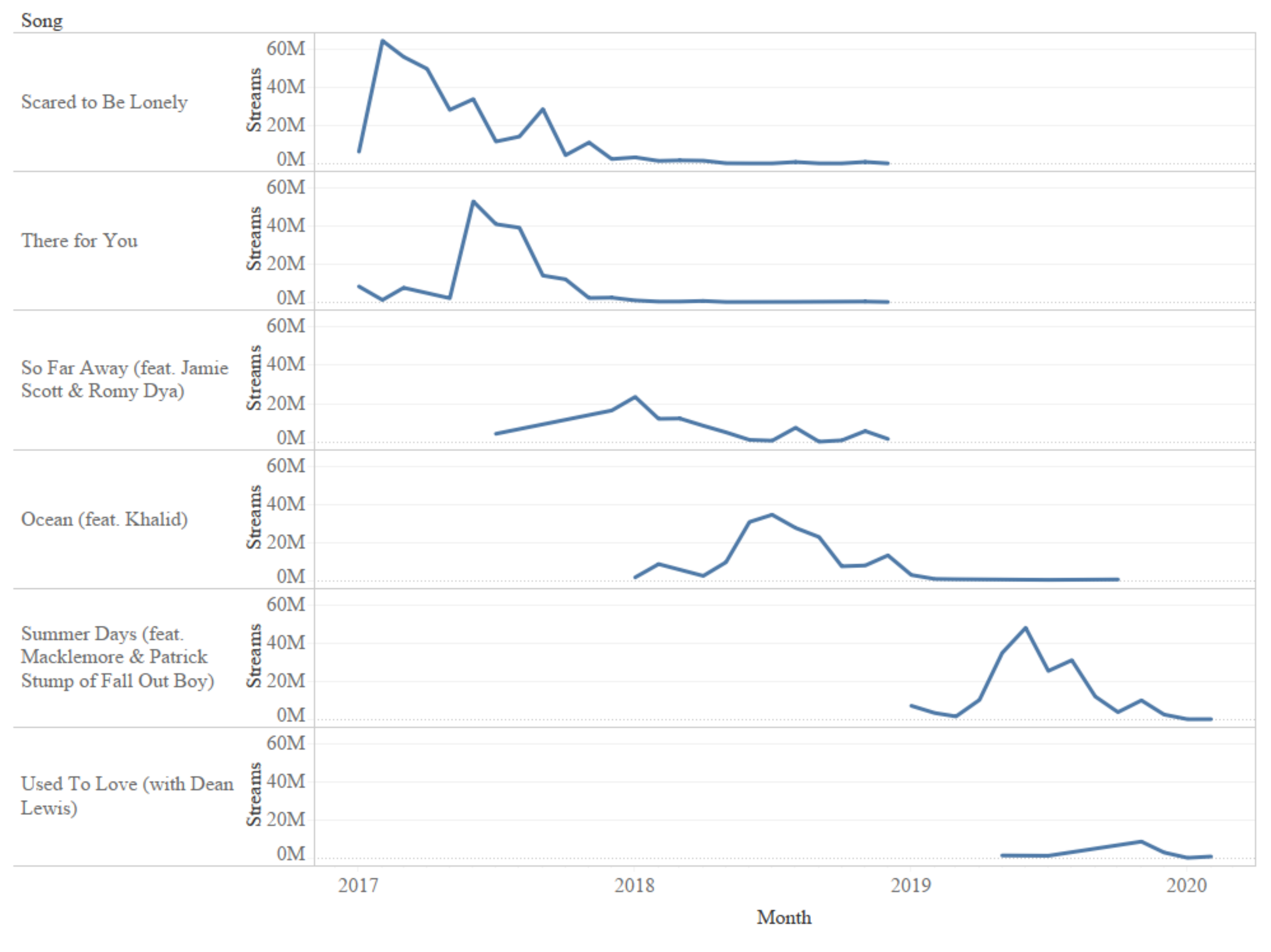

Figure 4: The number of streams for Martin Garrix' most popular songs plotted over the course of three years.

Representativeness of the Charts

Spotify uses standardizing engines in its data assemblage. This is problematic, because it makes it seem as if what these numbers represent is similar for all 79 countries where Spotify is active. As demonstrated above, Spotify is setting new norms for popularity by using specific metrics in its own way. Furthermore, the metrics used to measure and represent popularity are the same for all the countries. In other words, Spotify´s streaming charts are standardizing numbers to measure popularity globally. As mentioned before, this is done “so that it can be stored and manipulated within code.”[107] Yet, as Andrew Piper explains, the degree of representativeness is often overestimated in cultural

[107] Berry, *The Philosophy of Software*.

analytics research.[108] This means that the extent to which the streaming data can represent real-life events like a music consumption is more limited than often presumed.

This issue can be illustrated by visualizing the median number of streams a song needs in order to get in the distinct charts at all, as you can see in figure 5 on the next page. Here the bars represent the number of streams a song needs in order to get in the charts. The chart with the biggest threshold is the US chart.[109] Here, a song needs to be streamed for almost one million times in a week to get in the chart. For countries like Japan, Belgium and the three Baltic states this threshold is lower than sixteen thousand. It is not surprising that the latter numbers are lower than the US, because less people live there. However, when comparing the relationship between the number of inhabitants and the median number of streams, an interesting misbalance becomes apparent. Looking at the number of inhabitants in Japan and the US in 2018, based on a public *Worldometers* data set, Japan has about thirty-seven percent of the total number of inhabitants in the US. However, making the same comparison with the median number of streams shows that Japan’s count is only three percent of the number of streams for the US charts. On the other hand, the same comparison with the US and Germany seems more in balance. There, Germany has twenty-five percent of the number of inhabitants of the US and about twenty-seven percent of the stream count. Note that the number of streams does not directly relate to individual users, nevertheless this comparison demonstrates how the US dominates Spotify’s streaming charts.

[108] Piper, ‘There Will Be Numbers’.
[109] Here I left out the global chart, where the threshold is two and a half times bigger than the US chart. This enables be to explore how Spotify’s popularity norm differs around the world.

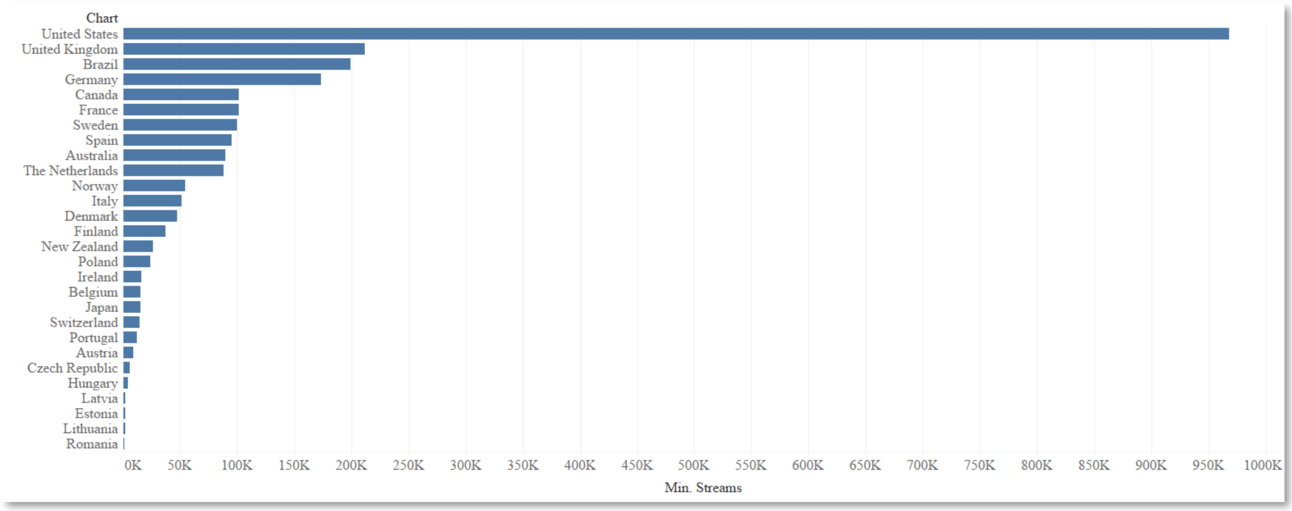

Figure 5: The number of streams a song needs to generate in order to get into the charts plotted per country.

When an artist gets in the US charts alone, this equals forty percent of the streams needed to get in the global charts. Getting in some of the smaller charts like Italy, The Netherlands, Poland or the Scandinavian countries is much further removed from the same number of streams. Yet, as discussed above this is not always well balanced with the number of people living there. Spotify describes its global chart as the place to see "the most played tracks" on its entire platform, which makes it seem as a neutral representation of the most successful songs globally.[110] However, the step to get in these global charts is much lower for an artist that got in the US charts than any of the other ones. From a strategic point of view an artist can decide to focus on the bigger local markets, like the US, Germany and the UK, in order to get in the global charts to be considered popular globally. Opposite to this, an artist that operates in the more diverse long tail could find their audience across the world in many smaller markets. This is afforded by the platform´s personalised recommendation algorithms, as e.g. Poell et al. and Anderson describe as an effect of the platform economy.[111] This kind of music crosses more borders, which can be seen as a measure of its degree of cultural globalisation, as Achterberg et al. denote.[112] Nevertheless, such artists have a lower chance of getting in the charts, because Spotify does not measure global success based on borders crossed but on the total number of streams. When the charts are used as proxy signifiers for the

---

[110] 'Spotify Charts'.
[111] Poell, Nieborg, and Dijck, 'Platformisation'; Chris Anderson, *The Long Tail: Why The Future of Business Is Selling Less of More*. (New York: Hyperion, 2006).
[112] Achterberg et al., 'A Cultural Globalization of Popular Music? American, Dutch, French, and German Popular Music Charts (1965 to 2006)'.

cultural globalisation of music, it's representativeness should therefore not simply be taken for granted.

Furthermore, the way Spotify's charts come into existence changes throughout time. First of all, Spotify's charts were available before 2017, the first hit of the Spotify charts website on the *WayBack Machine* web archive is on July 29th, 2012.[113] Today however, Spotify only makes the chart's numbers accessible from 2017. What has changed for Spotify to decide to start with new charts? Perhaps, the thirty second threshold for a stream count was different before. The Dutch Top 40 has also made multiple changes to the way their charts are constructed, and they share a description of their changes on a dedicated webpage.[114] BC offers an annotation of the measuring method too in their yearly report.[115] Spotify however, offers no annotation of changes made to their charts. Thereby, the company leaves the data open for interpretation. This effects the looping process, because experts, external systems and organisations, artists, and frankly anyone with an internet connection is free to give meaning to the data in the way that best suits them. In order to understand what these data can represent these external parties are however dependent on the way these data are represented by Spotify, which is not a static definition either.

The latter is exemplified in the charts data, which became visible when diving into collaborating artists. So far, I have referred to the global success of individual artists and their songs, but many songs in the charts are the result of collaborations. The *Used To Love* song from Martin Garrix is for instance a collaboration with Dean Lewis and *Rise* from Nicky Romero is featuring the artist Matluck. However, in the Spotify charts all of these songs are connected to only one artist. The collaborating artist are mentioned in the same data-signal as the name of the song. Thereby, Spotify thus constructs a certain hierarchy where one artist is presented as the song´s artist, while the collaborators are mere collaborators and not particularly referred to as 'artists.' However, the API data does not make the same distinction and refers to all the artist as 'artists.'[116] It is not sure why Spotify does this, yet it emphasized the dependency of popular music to Spotify's datafication. Furthermore, this exploration of the three years of charts enabled to find more evidence of this contingency of Spotify's data. In total I found seven different ways

[113] 'SpotifyCharts.Com --- Music Charts for Spotify!', 29 July 2012, http://web.archive.org/web/20120729055442/http://www.spotifycharts.com/.
[114] Top 40, 'Geschiedenis Nederlandse Top 40'.
[115] Kroeske, 'Exportwaarde Nederlandse Populaire Muziek'.
[116] 'Web API | Spotify for Developers'.

in which a reference to the collaborating artists is made in the charts. Namely: "(With [artist])", "feat. [artist]", "[feat. [artist]]", "(& [artist])", "; feat. [artist]", "([artist] & [artist])" and "(feat. [artist])". The words used are somewhat the same, but the structure of the data is different, because different symbols are used to distance it from the song name. Maybe these different notations have different values connected to it. For instance: "With" can mean a duet and "; feat." can refer to a smaller role of the collaborating artist as is suggested in an online popular music forum.[117] However, Spotify leaves all of this open for anyone's imagination. This degree of inconsistency and lack of explanation makes it hard to grasp what the streaming metrics actually mean, making its use questionable.

[117] 'What's Difference Between (W/ & or Feat./)?, Posted by Elayblooze - Rate Your Music', accessed 12 June 2020, https://rateyourmusic.com/board_message/message_id_is_3107221.

Conclusion

This thesis started with the question of how the cultural globalisation of the top fifty Dutch popular music artists is represented in Spotify's streaming charts and what it tells us about the metrics themselves. I demonstrated that Spotify´s streaming charts influence the way the popularity of Dutch artists is determined. Aguiar and Waldfogel already stressed that streaming platforms influence traditional modes of music consumption, like the purchase of an album.[118] Traditional charts that use these traditional metrics to measure the most popular artists, like Buma Cultuur (BC), are therefore influenced by streaming platforms too. As more people consume music on Spotify´s platform every year, Spotify´s metrics cannot be ignored any longer.[119] As opposed to BC's static list of the fifty most successful Dutch artists, this analysis discovered a yearly pattern. This pattern shows that the Dutch artists mainly provide soundtracks to Spotify's global audience during the summer. In Spotify's charts, the Dutch artists are successful one song at the time, while traditional charts measure popularity based on the success of entire albums.[120] Still, organisations like the Dutch Top 40 increasingly use the streaming metrics to determine their charts. However, as I have argued, these metrics represent a different cultural practice than purchasing an album. Frankly, combining these metrics to represent the same thing is wrong, because as Ian Hacking stresses, the representation changes as the proxy signifiers change too.[121] For this reason, it is crucial to grasp what the streaming charts can and cannot represent.

While forty-eight of BC´s list of successful Dutch artists can be found on Spotify, only a few managed to get in the streaming charts. Seventeen of BC´s top fifty artists are well represented in Spotify's charts, implying that Spotify *agrees* that these artists are successful, most of the Dutch artists did not make it to the charts. Therefore, the artists that Spotify considers successful differ from BC's definition of success. With this analysis, I contribute to the goal of cultural analytics to explore online data sources to create a rich situational awareness of cultural practices. In this case, Spotify's weekly updated charts enabled to find patterns that were not visible before. Nevertheless, these patterns only represent seventeen out of the initial fifty selected Dutch artists. The representativeness of these charts for all fifty artists, let alone Dutch popular music in

[118] Aguiar and Waldfogel, 'As Streaming Reaches Flood Stage, Does It Stimulate or Depress Music Sales?'
[119] 'Spotify Reports Fourth Quarter and Full-Year 2019 Earnings', 5 February 2020, https://newsroom.spotify.com/2020-02-05/spotify-reports-fourth-quarter-and-full-year-2019-earnings/.
[120] Aguiar and Waldfogel, 'As Streaming Reaches Flood Stage, Does It Stimulate or Depress Music Sales?'
[121] Hacking, 'Kinds of People: Moving Targets.'

general, is therefore limited. For the artists that are represented in Spotify's streaming charts, the meaning of their proxy signifiers remains opaque.

The analysis demonstrated that Spotify's streaming charts are incomplete, contingent to, and manipulated by its data assemblage. Spotify presents its streaming metrics as if these are neutral, but as I have shown Spotify's datafication e.g. favours US listening behaviour over others. Besides, the platform offers an extremely limited explanation of what the data mean, as if Spotify's data speak for themselves. In critical data studies (CDS), this is observed as a common shortcoming of publicly available data sources, as Gitelman, Kitchin and Lauriault, and Cukier and Mayer-Schoenberger denote.[122] Because of it, any organisation or individual trying to find meaning in Spotify's streaming charts is subject to its changes in an ongoing looping process. Every stage of this process effects the meaning of the data, as Hacking stresses.[123] The metrics themselves are not static either. The charts are ever changing, transforming its streaming metrics to be "moving targets."[124] The biggest problem is not that these are moving, but that the changes are not acknowledged and remain hidden in Spotify's black box. Without a critical understanding of the streaming metrics this could lead to false conclusions, which is especially problematic as more third parties, like artists and institutions such as the Dutch Top 40, become reliant on Spotify´s metrics. One of the aims of CDS is to increase data literacy for anyone working with data, as Kitching and Lauriault explain.[125] This thesis contributes to this goal for media and culture scholars. When they want to use Spotify's metrics as proxy signifiers for cultural phenomenon, this thesis can be used as a guide to read Spotify's data.

Concluding, I understand Spotify's charts as a representation of the growing dominance of platforms over the way popular music cultures develop globally. Especially, because its users can also read the artists biography, find the lyrics to the songs, and find the future concerts of their favourite artists on Spotify.[126] Therefore, not getting your songs in a popular data-driven playlist, because Spotify´s datafication thinks your songs cannot be successful, could lead to less concert tickets sold. Hence, how Spotify measures and defines success matters. As Spotify and its streaming metrics are becoming more important for global and Dutch popular music industries, I encourage

---

[122] Gitelman, *Raw Data Is an Oxymoron*; Kitchin and Lauriault, 'Towards Critical Data Studies'; Cukier and Mayer-Schoenberger, 'The Rise of Big Data'.
[123] Hacking, 'Kinds of People: Moving Targets.', 286.
[124] Idem, 312.
[125] Kitchin and Lauriault, 'Towards Critical Data Studies'.
[126] 'Features – Spotify for Artists'.

future research to continue their critical engagements with such data-driven platforms. Finally, I encourage public policy makers and actors operating in the cultural industries to rethink their dependency on these platforms and question: what standards is the platform's datafication constructing, could my work be considered successful in its data assemblage, and most importantly, do I want to adjust my work to its standards? Spotify's growing dominance as a place to discover and consume music should not be underestimated, therefore the question remains if you can be a globally successful Dutch artist without it.

References

Achterberg, Peter, Johan Heilbron, Dick Houtman, and Stef Aupers. 'A Cultural Globalization of Popular Music? American, Dutch, French, and German Popular Music Charts (1965 to 2006)'. *American Behavioral Scientist* 55, no. 11 (2011): 589–608. https://doi.org/10.1177/0002764211398081.

Aguiar, Luis, and Joel Waldfogel. 'As Streaming Reaches Flood Stage, Does It Stimulate or Depress Music Sales?' *International Journal of Industrial Organization* 57 (2018): 278–307. https://doi.org/10.1016/j.ijindorg.2017.06.004.

Alter, Charlotte. 'Taylor Swift Just Removed Her Music From Spotify'. *Time*, 3 November 2014. https://time.com/3554438/taylor-swift-spotify/.

Anderson, Chris. *The Long Tail: Why The Future of Business Is Selling Less of More.* New York: Hyperion, 2006.

Bail, Christopher A. 'The Cultural Environment: Measuring Culture with Big Data'. *Theory and Society* 43, no. 3–4 (2014): 465–82. https://doi.org/10.1007/s11186-014-9216-5.

Berry, D. *The Philosophy of Software: Code and Mediation in the Digital Age*. New York: Palgrave Macmillan, 2011. https://doi.org/10.1057/9780230306479.

Billboard. 'Billboard Finalizes Changes to How Streams Are Weighted for Billboard Hot 100 & Billboard 200', 1 May 2018. http://www.billboard.com/articles/news/8427967/billboard-changes-streaming-weighting-hot-100-billboard-200.

Bonini, Tiziano, and Alessandro Gandini. '"First Week Is Editorial, Second Week Is Algorithmic": Platform Gatekeepers and the Platformization of Music Curation'. *Social Media + Society*, 2019. https://doi.org/10.1177/2056305119880006.

Caplan, Lindsay. 'Method without Methodology: Data and the Digital Humanities' 72, no. E-flux journal (2016). Worker01.e-flux.com/pdf/article_9006656.pdf.

Caselli, Marco. 'Nation States, Cities, and People: Alternative Ways to Measure Globalization'. *SAGE Open* 3, no. 4 (2013): 2158244013508417. https://doi.org/10.1177/2158244013508417.

Cukier, Kenneth, and Viktor Mayer-Schoenberger. 'The Rise of Big Data: How It's Changing the Way We Think about the World'. *Foreign Affairs* 92, no. 3 (2013): 28–40.

Eriksson, Maria, Rasmus Fleischer, Anna Johansson, Pelle Snickars, and Patrick Vonderau. *Spotify Teardown. Inside the Black Box of Streaming Music*. London: The MIT Press, 2019.

'Excel Specifications and Limits'. Accessed 5 June 2020. https://support.office.com/en-us/article/excel-specifications-and-limits-1672b34d-7043-467e-8e27-269d656771c3.

'Features – Spotify for Artists'. Accessed 19 June 2020. https://artists.spotify.com/features.

Gitelman, Lisa. *Raw Data Is an Oxymoron*. Cambridge: The MIT Press, 2013. https://doi.org/0262518287.

Gutierrez, Horacio. 'Annual Report 2019 - Spotify Technology S.A.' Annual Report. Washington, D.C.: Spotify Technology S.A., 2019.

Hacking, Ian. 'Kinds of People: Moving Targets.' *Proceedings of Britisch Academy* 151 (2007): 285–318.

Hakanen, Ernest A. 'Counting down to Number One: The Evolution of the Meaning of Popular Music Charts'. *Popular Music* 1, no. 17 (1998).

https://www.cambridge.org/core/journals/popular-music/article/counting-down-to-number-one-the-evolution-of-the-meaning-of-popular-music-charts/AD2DF9B7B6F5761E029376E2A92EEDAE.
Harris, Michelangelo, Brian Liu, Cean Park, Ravi Ramireddy, Gloria Ren, Max Ren, Shangdi Yu, Andrew Daw, and Jamol Pender. 'Analyzing the Spotify Top 200 Through a Point Process Lens'. *ArXiv*, 2019. https://doi.org/1910.01445.
Helmond, Anne. 'The Platformization of the Web: Making Web Data Platform Ready'. *Social Media + Society* 1, no. 2 (2015): 1–11. https://doi.org/10.1177/2056305115603080.
'Home | Spotify for Developers'. Accessed 18 January 2020. https://developer.spotify.com/.
Iliadis, Andrew, and Federica Russo. 'Critical Data Studies: An Introduction'. *Big Data & Society* 3, no. 2 (2016): 1–7. https://doi.org/10.1177/2053951716674238.
Ingham, Tim. 'Windowing on Spotify Is Finally Happening - but Is It Worth It for the Labels?' *Music Business Worldwide*, 13 April 2018. https://www.musicbusinessworldwide.com/windowing-on-spotify-is-finally-happening-but-is-it-worth-it-for-the-labels/.
Kitchin, Rob, and Tracey Lauriault. 'Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work', 2014. https://papers.ssrn.com/abstract=2474112.
Kluver, Randolph, and Wayne Fu. 'Measuring Cultural Globalization in Southeast Asia', 2008, 335–58.
Kraak, Haro. 'Hoe streamingdiensten als Spotify de muziek veranderen'. *de Volkskrant*, 19 November 2017, sec. Cultuur & Media. https://www.volkskrant.nl/gs-b10594f9.
Kroeske, Sieb. 'Exportwaarde Nederlandse Populaire Muziek'. Research Report. Buma Cultuur, 2020.
Manovich, Lev. 'Cultural Analytics: Analysis and Visualization of Large Cultural Data Sets. A Proposal for Software Studies Initiative', 2007. https://www.mat.ucsb.edu/g.legrady/academic/courses/11w259/cultural_analyticsManovich.pdf.
Moretti, Franco. 'Conjectures on World Literature'. *New Left Review* 1 (2000): 54–68.
Moretti, Franco. 'Graphs, Maps, Trees. Abstract Models for Literary History.' *New Left Review* 24 (2003).
Moretti, Franco, and Dominique Pestre. 'Bankspeak: The Language of World Bank Reports'. *New Left Review* 92 (2015).
'My Dashboard | Spotify for Developers'. Accessed 4 May 2020. https://developer.spotify.com/dashboard/.
Nieborg, David B, and Thomas Poell. 'The Platformization of Cultural Production: Theorizing the Contingent Cultural Commodity'. *New Media & Society* 20, no. 11 (2018): 4275–92. https://doi.org/10.1177/1461444818769694.
Piper, Andrew. 'There Will Be Numbers'. *Journal of Cultural Analytics*, 2016. https://doi.org/10.22148/16.006.
Poell, Thomas, David Nieborg, and José van Dijck. 'Platformisation'. *Internet Policy Review* 8, no. 4 (2019). https://policyreview.info/concepts/platformisation.
'Premium vs. Free – Videos – Spotify for Artists'. Accessed 8 April 2020. https://artists.spotify.com/videos/the-game-plan/premium-vs-free.
Puschmann, Cornelius, and Julian Ausserhofer. 'Social Data APIs'. In *The Datafied Society. Studying Culture through Data.*, edited by Mirko Schäfer and Karin van Es, 147–54. Amsterdam: Amsterdam University Press, 2017.

R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing, 2013. http://www.R-project.org/.
RStudio Team. *RStudio*. Boston, MA: RStudio Inc., 2019. https://rstudio.com/.
selfiecity. 'Selfiecity'. Accessed 17 December 2019. http://selfiecity.net/.
Spotify. 'Spotify — Company Info'. Accessed 26 October 2019. https://newsroom.spotify.com/company-info/.
'Spotify Charts'. Accessed 14 January 2020. https://spotifycharts.com/regional.
'Spotify Developer Terms of Service | Spotify for Developers'. Accessed 4 May 2020. https://developer.spotify.com/terms/.
'Spotify Reports Fourth Quarter and Full-Year 2019 Earnings', 5 February 2020. https://newsroom.spotify.com/2020-02-05/spotify-reports-fourth-quarter-and-full-year-2019-earnings/.
'SpotifyCharts.Com --- Music Charts for Spotify!', 29 July 2012. http://web.archive.org/web/20120729055442/http://www.spotifycharts.com/.
'Stats – FAQ – Spotify for Artists'. Accessed 22 January 2020. https://artists.spotify.com/faq/stats.
'Stats – FAQ – Spotify for Artists'. Accessed 30 March 2020. https://artists.spotify.com/faq/stats.
Tableau: Business Intelligence and Analytics Software. Seattle, WA: Tableau Software. Accessed 18 January 2020. https://www.tableau.com/.
'The Comprehensive R Archive Network'. Accessed 18 January 2020. https://cran.r-project.org/.
'The Spotify Community'. Accessed 7 June 2020. https://community.spotify.com/.
Top 40, Stichting Nederlandse. 'Geschiedenis Nederlandse Top 40'. Top40.nl. Accessed 21 January 2020. https://www.top40.nl/geschiedenis.
'Web API | Spotify for Developers'. Accessed 4 May 2020. https://developer.spotify.com/documentation/web-api/.
'What's Difference Between (W/ & or Feat./)?, Posted by Elayblooze - Rate Your Music'. Accessed 12 June 2020. https://rateyourmusic.com/board_message/message_id_is_3107221.
Spotify. 'Woordenschat'. Accessed 17 June 2020. https://open.spotify.com/playlist/37i9dQZF1DX19xRtMyA5LM.

Appendix

Appendix 1: Spotify's data assemblage

Apparatus	**Elements**	**Spotify**
Systems of thought	Modes of thinking, theories, models, ideologies etc.	Spotify aims to be a world leader in audio content. Claiming to have a two-sided market strategy serving content consumers on the one side and content producers on the other side. Within the latter side they company states it aims to enable more artists to make a living from the content they share on Spotify.
Forms of knowledge	Websites, word of mouth, chat forums, research text etc.	The company shares what the data mean that they collect and process on multiple public websites, in their Spotify.for.Developers environment, on their publicly facilitated community forum and e.g. in their shareholders letters. However, it is not shared how the algorithms work and reverse engineering is prohibited.
Finance	Business models, investments, profit etc.	The company is partly funded by Swedish subsidies. The company has financial stakeholders in Europe and in the U.S. They report the income in two segments: premium subscribers & ad-supported users. While most users have an ad-supported account, most revenue comes from premium subscribers.
Political economy	Policy, public and political opinion, tax regimes etc.	The company operates across national boundaries. Therefore, they must deal with different local policies. However, in order to grow as quickly as possible, they tend to standardize their products as much as possible to fit in as much local policies as possible.
Governmentalities and legalities	Data standards, file formats, system requirements, protocols,	Spotify has data standards for the audio files that can be uploaded. These are controlled by third-party gatekeepers. The company also has regulations for

	regulations, ethical considerations, etc.	the kind of content it allows. Furthermore, it must deal with copyrights which vary across the world. The company made their own data standard for measuring user growth (Monthly Active Users) which they claim is a good measure out of their experience. Furthermore, they measure a consumption as a stream.
Materialities and infrastructures	Paper/pens, computers, digital devices, sensors, scanners, databases, networks, servers, buildings, etc.	The platform can be accessed on smartphones, tablets, personal computers, smart board computers in cars, Smart TV's, etcetera. Which are all tracked with Spotify's meta-data tracking. The data can be accessed by third parties through the API and in one of their online insights' environments.
Practices	Techniques, ways of doing, learned behaviours, scientific conventions, etc.	The company collects user behavior data and meta-data. Spotify uses data it collects and processes to construct data-driven playlists by algo-torial curators and to server personalized advertisements. They analyze audio files with A.I. technologies. The company shares some of the data on their websites and through their API.
Organizations and institutions	Government agencies, universities, conferences, clubs and societies, committees and boards, communities of practice, etc.	Much of their infrastructure is built upon open-source code made by many third parties, yet they do not share how their own code functions. Also, third party aggregators decide which content gets on the platform. Part of the company is owned by major U.S. labels and Chinese Streaming provider Tencent. Which implies some influence of those parties in the way Spotify functions.

Subjectivities and communities	Of data producers, experts, curators, managers, analysts, scientists, politicians, users, citizens, etc.	The algorithms are written by human coders, mostly from the Stockholm office. The company claims this team consists of programmers from around the world. The figures are protected by undisclosed algorithms, so the decisions made in the processing, hence their meaning is uncertain to outsiders.
Places	Labs, offices, field sites, data centers, server farms, business parks, etc., and their agglomerations	The data is stored on Google Cloud Servers according to their stakeholder letters. The company has major offices in Stockholm, London and New York and other smaller offices throughout the world.
Marketplace	For data, its derivatives (e.g., text, tables, graphs, maps), analysts, analytic software, interpretations, etc.	The data are transformed to personalized playlists which are valuable to end-users in ad-supported and premium accounts. The user data are sold to advertisers and presented for instance in advertisements and Spotify.Me promotions which highlight the data richness Spotify can offer to advertisers. Audience insights can be analyzed by artists and other content creators when it's about the consumption of their cultural products. The company shares data insights in their Spotify.for.Artists environment.

Appendix 2: The data set

The data was collected in an Excel file. I made the file available for download in a public Google Drive folder that can be accessed by typing the following URL in your web browser:

https://drive.google.com/drive/folders/1OOIuvwZrueMI6KgJkq9ys6MB2RwPY9kc?usp=sharing

This file consists of multiple tabs, which are listed below.

1. **Artists**. This is a list of the top fifty artists from the Buma Cultuur report with the following data fields:
    a. Artist: the name of the artist.
    b. Spotify ID: the unique identifier of this artist.
2. **All songs**: This is a list of all the charts that were collected with the following data fields:
    a. Rank: position of the song in the charts.
    b. Song: the name of the song.
    c. Featuring artists: the name of the collaborating artists.
    d. Artist: the name of the artist.
    e. Streams: the number of streams the song generated.
    f. Date: the date the song was in the charts.
    g. Chart: the specific chart the song was in.
3. **Dutch Songs:** This is a list of all the Dutch songs with the following data fields:
    a. Song: the name of the song.
    b. Release data: the day the song was released according to Spotify.

Appendix 3: R code

R code used for scraping the charts

```
#the readymade packages
library(rvest)
library(tidyverse)
library(magrittr)
library(scales)
library(knitr)
library(lubridate)

#describing which are the weeks we want to scrape
start <- "2016/12/30"
end1 <- "2020/01/05"

#This is where the charts are scraped from
url <- "https://spotifycharts.com/regional/global/weekly/"

start <-as.Date(start)
end <- as.Date(end)
start <-start + (6 - wday(start))
end <- end +(5-wday(end))

week_start <- seq(start, (end-7), by = "week")
week_end <- seq((start + 7) , end, by = "week")

week_def <- paste(week_start,week_end, sep="--")
week_def

unite.url<- function(x){
  full_url <- paste0(url, x)
  full_url
}
#Run the function
search_url <- unite.url(week_def)
search_url

SpotifyScrape <- function(x){
  page <- x
  rank <- page %>%
    read_html() %>%
    html_nodes('.chart-table-position') %>%
    html_text() %>%
    as.data.frame()
  track <- page %>%
    read_html() %>%
    html_nodes('strong') %>%
    html_text() %>%
    as.data.frame()
  artist <- page %>%
```

```
    read_html() %>%
    html_nodes('.chart-table-track span') %>%
    html_text() %>%
    as.data.frame()
  streams <- page %>%
    read_html() %>%
    html_nodes('td.chart-table-streams') %>%
    html_text() %>%
    as.data.frame()
  dates <- page %>%
    read_html() %>%
    html_nodes('.responsive-select~ .responsive-select+ .responsive-select .responsive-
select-value') %>%
    html_text() %>%
    as.data.frame()
  charts <- page %>%
    read_html() %>%
    html_nodes('.responsive-select .responsive-select-value') %>%
    .[[1]] %>%
    html_text() %>%
    as.data.frame()


  #combine, name, and make it a tibble
  chart <- cbind(rank, track, artist, streams, dates, charts)
  names(chart) <- c("Rank", "Track", "Artist", "Streams", "Date", "Chart")
  chart <- as.tibble(chart)

  return(chart)
}
spotify_data <- map_df(search_url, SpotifyScrape)

#cleaning the data set
spotify_data %<>%
  mutate( Artist = gsub("^by ", "", Artist),
          Streams = gsub(",", "", Streams),
          Streams = as.numeric(Streams),
          Date = as.Date(spotify_data$Date,"%m/%d/%Y"),
          Track = as.character(Track)
          ) -> spotify_data

head(spotify_data)
#write data in a local csv file
setwd("[file location on computer]")
write.csv2(spotify_data, "SpotifyChartsWeekly_[country].csv")
```

R code used for accessing Spotify's API

```
library(spotifyr)
library(dplyr)
library(tidyverse)

#authorisation that can be requested on the Spotify for developers website
Sys.setenv(SPOTIFY_CLIENT_ID = "[Private ID]")
Sys.setenv(SPOTIFY_CLIENT_SECRET = "[Private Code]")

access_token <- get_spotify_access_token()

#declaring list of artist id's I want to check. I found these ID's by manually searching for the 50 artist on Spotify.
artists_ids <-
c("4D75GcNG95ebPtNvoNVXhz","2o5jDhtHVPhrJdv3cEQ99Z","1l2ekx5skC4gJH8dj
ERwh1","60d24wfXkVzDSfLS6hyCjZ","18T0SpXY06tqdiyBTYTIgi","0SfsnGyD8Fp
IN4U4WCkBZ5","47z7ZrgFoBvVpCnElCE3Zh","1V3VTM7VspiQjcmRhC010n","6c
EuCEZu7PAE9ZSzLLc2oQ","5nki7yRhxgM509M5ADlN1p","7kABWMhjA5GIl9PB
EasBPt","20gsENnposVs2I4rQ5kvrf","6xQvQwZQQuq9R3TdPNbcR8","2rTo8KIkBT
FjQS7VvaKYQ4","53cQZtWDwDJwVCNZlfJ6Qk","58wDbO29TUyzGTz7gZqOue","
4x7gxsrTH3gThvSKZPPwaQ","4jGpKAmwvU263l0tUh4xKU","6hJtgCB3L5cnJSND
7sp6GU","2UgphhGSlC9QWgaZWUOCkl","7zpN81tVvPwlHcJSkSCyRa","0cwmNvc
lzPd8mQnoHuIksj","2ohlvFf9PBsDELdRstPtlP","7jp5R1aY7kmwBYX3jIu9yk","5jm3
x1qIibWdKSEMw2G011","29IzCNDrWGPFXMmtCY3N83","7dc6hUwyuIhrZdh80ea
CEE","3hJv5p2HwekJysNB2NDnEC","4sQNUQjOYj9rV2sdfJ8laS","33aAnmwlO8v4
0zpiBWhqTc","6VkPP9ugQTWDbCpj0EwU81","5QySqc6yAFDx9m7fedFZmC","7e1
BNCygl2Gf7CX8LrByPv","12SPNXi0aDpFt0rMVbmLrr","5ChF3i92IPZHduM7jN3d
pg","3hciB8bkQeECakkTWmrPEH","0c103ZyWDycpfVxR0lNrjm","37PL04N8kBW
WW69jdrMMWT","3gRJFoOKgAkHI7EZTWKXRy","6BrvowZBreEkXzJQMpL174"
,"5nri3hyKmKBGAfvjBi0mK0","5Aw0IGM5JS3FuTgtRsDWGA","1PFc84GHTYPL0
G12JHh9Mc","1EAFXic0Cfiwpe7nSuTrGL","0e1whBY5xv3rF7Zj5cxLD9","5WdqBA
QhGFCrZvBKXiPIu7","0jNDKefhfSbLR9sFvcPLHo")

#retrieving artists information
artist_info <- get_artists(artists_ids, authorization = get_spotify_access_token(),
        include_meta_info = FALSE)

#setting artist information in a dataframe
artists_df <- data.frame(
  artist_info$id,
  artist_info$name,
  artist_info$popularity,
  artist_info$followers.total)

artists_df

#write data in a local csv file
setwd("[file location]")
write.csv2(artists_df, "artist info.csv")
```

Thank you for reading!