

# Interactively exploring supply and demand in the UK independent music scene

Matt McVicar, Cédric Mesnage, Jeffrey Lijffijt, and Tijl De Bie

Intelligent Systems Lab, University of Bristol, UK

`mattjamesmcvicar@gmail.com`, `cedric.mesnage@bristol.ac.uk`,  
`jefrey.lijffijt@bristol.ac.uk`, `tijl.debie@gmail.com`

**Abstract.** We present an exploratory data mining tool useful for finding patterns in the geographic distribution of independent UK-based music artists. Our system is interactive, highly intuitive, and entirely browser-based, meaning it can be used without any additional software installations from any device. The target audiences are artists, other music professionals, and the general public. Potential uses of our software include highlighting discrepancies in supply and demand of specific music genres in different parts of the country, and identifying at a glance which areas have the highest densities of independent music artists.

## 1 Motivation

Supply and demand are important micro-economic properties of products, affecting the price consumers are willing to pay for an item or service. With respect to music, the ‘supply’ of a particular music genre<sup>1</sup> can be thought of as the number of artists actively recording and performing music in this style. ‘Demand’ on the other hand refers to a population’s desire to spend resources (time, money) purchasing and listening to a given music genre, as well as attending live concerts in this style.

Quantification of the supply and demand of music is not straightforward; until very recently only the demand for the *most popular* genres was known, through charts produced by companies such as Billboard<sup>2</sup> or The Official Charts Company<sup>3</sup>. Our aim is to provide estimates of supply and demand at much finer granularities by using simple statistical techniques combined with the massive amount of data available on the web. In particular, music artist platforms such as SoundCloud, MixCloud and YouTube<sup>4</sup> present us with a unique chance to study the supply of particular music genres outside the mainstream, whilst fan activity on (micro)blogging sites such as Twitter and Facebook give us a window into the demand for particular musical styles.

---

<sup>1</sup> ‘style’ and ‘genre’ are considered to be synonyms in this paper, examples of which include *Classical*, *Rock* etc.

<sup>2</sup> <http://www.billboard.com/>

<sup>3</sup> <http://www.officialcharts.com/home/>

<sup>4</sup> <https://soundcloud.com/>, <https://www.mixcloud.com/>, <https://www.youtube.com>

We are particularly interested in lesser-known artists as they may represent fresh talent or emerging music styles. Musicians such as these may never reach commercial breakthrough if they are not identified and linked with their fans, and our tool can help record labels identify growing genres and thus artists at an early stage in their career, leading to greater profit potential. In contrast to existing work on geo-located twitter music [1], [2], our system pairs artists active on the web with their fans, and is also concerned exclusively with non-mainstream music. Furthermore, we provide an intuitive system for browsing the patterns seen in the geographical distribution of these niche music artists.

The remainder of this demo paper is organised as follows. In Section 2 we outline our data collection methods. Section 3 gives an overview of our system, and we conclude in Section 4.

## 2 Data Collection and Processing

We collected data about 54,192 UK independent artists from Reverbnation.com<sup>5</sup>. In particular, we collected artist name, genre(s), location and links to their Twitter, Soundcloud and/or Youtube accounts. Using Twitter handles of the artists (21,616 artists volunteered this information) we periodically gathered tweets using this handle. Currently, our MongoDB database is storing 10,005,640 tweets, tweeted from 1,844,278 different accounts. 181,042 of these tweets we have found to be geolocalised and can therefore be used to assess the demand for a genre in a particular location.

To localise our data, we obtained the location from ReverbNation (free text) and Twitter (curated by Twitter) and semi-automatically binned locations into the 83 counties of England, 26 districts of Northern Ireland, 32 unitary authorities of Scotland and the 32 principal areas of Wales. We found that many artists listed their location simply as ‘London’, which is not itself a well-defined area. We therefore collected all London boroughs (Camden, Tower Hamlets etc.) into one region called simply *London*.

Supply (demand) for each genre and region in our maps was then estimated simply as number of artists (tweets about artists) divided by population of region. These were generally on different scales, making a direct comparison challenging. Instead, we employed the following statistical approach which aims to quantify how balanced the supply and demand is given a background balance given by the UK as a whole.

Let  $a_i$  and  $t_i$  be the number of artists and tweets for region  $i$ . Furthermore, suppose that  $a$  and  $t$  are the number of artists in this genre for the UK in its entirety. Considering  $a_i$  and  $t_i$  as samples from independent Poisson random variables, we then regard the respective rate parameters  $\lambda_{a,i}$  and  $\lambda_{t,i}$  as proxy measures for supply and demand for a given genre in region  $i$ . The degree to which region  $i$ 's music market is out of equilibrium can then be quantified by comparing  $\log(\lambda_{a,i}/\lambda_{t,i})$  with  $\log(\lambda_a/\lambda_t)$ ; the latter is considered the equilibrium.

<sup>5</sup> <http://www.reverbnation.com/>

Estimating this log ratio parameter in a maximum likelihood setting can be conducted following Wald’s approach [3], the Maximum Likelihood Estimate for  $\log(\lambda_{a,i}/\lambda_{t,i})$  is  $\mu_i = \log((a_i + 0.5)/(t_i + 0.5))$ , with an estimated variance of  $\sigma_i^2 = \frac{1}{a_i+0.5} + \frac{1}{t_i+0.5}$ . Due to the large population size of the UK in its entirety, the ‘prior’ disequilibrium  $\log(\lambda_a/\lambda_t)$  can be estimated accurately as  $\mu = \log(a/t)$  with negligible variance. As a measure of the disequilibrium, we thus used the  $z$ -score of  $\mu_i$ :  $z_i = \frac{\mu_i - \mu}{\sigma_i}$ .

### 3 Interface

Our system is live for users to experiment with<sup>6</sup>, while a screenshot is provided in Figure 1. Users interact with our system in the following way. First, they select a genre from a dropdown menu. This then automatically loads a new set of maps for the supply, demand and net data. Each region of the UK is colored by intensity on a green scale for supply, purple for demand and diverging green-purple for net divergence from the equilibrium.

Mouse-hovering over a particular region then shows the relevant data in a clear and concise way using the recently-developed D3 Javascript library (data-driven documents, <http://d3js.org/>). By exploring these maps, we hope that users will be able to delve deep into the data in an intuitive and fun way, exploring many gigabytes of data with no technical expertise required.

The final map (showing net supply vs. demand) may be of particular interest, since users can discover neighbouring regions in which there is a lack of supply and excess demand (or vice-versa). This indicates that there is an opportunity for *arbitrage* in the market for this genre: artists may wish to book live concerts or promote their work in a neighbouring county if there is a higher demand for their style than in their own locality.

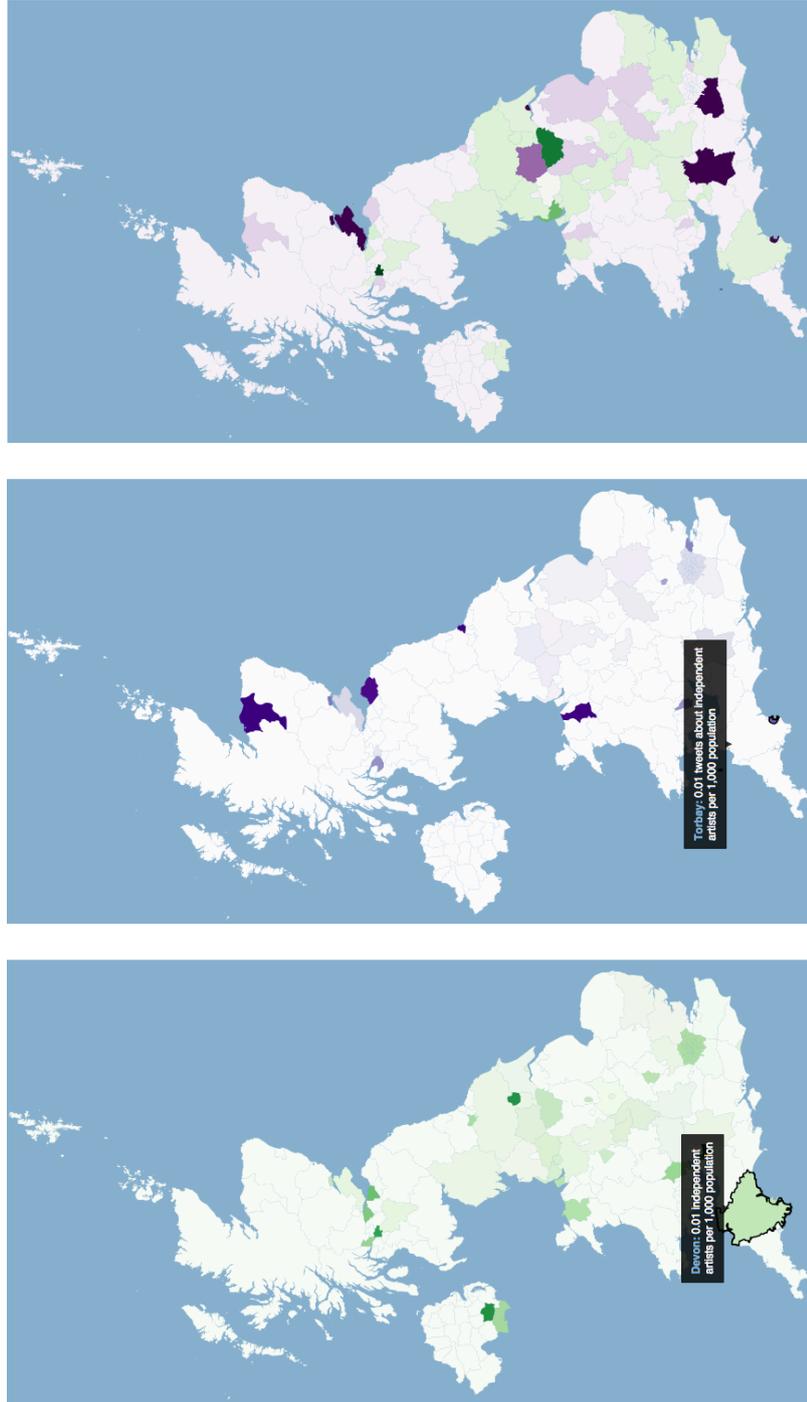
### 4 Conclusions

In this demo, we introduced an exploratory data analysis tool for identifying geographic trends in the independent UK music scene. Our system is web-based, requires no installation other than a modern browser, and is highly intuitive to use. We are optimistic that this tool can be used by music fans and professionals alike, to navigate trends in the distribution of music genres through the UK.

### References

1. A. Bellogín, A. P. de Vries, and J. He. Artist popularity: Do web and social music services agree? In *Proc. of ICWSM*, pages 673–676, 2013.
2. D. Hauger, M. Schedl, A. Košir, and M. Tkalčič. The million musical tweets dataset: What can we learn from microblogs. In *Proc. of ISMIR*, pages 189–194, 2013.
3. R. Price and D. Bonett. Estimating the ratio of two poisson rates. *Computational Statistics & Data Analysis*, 34(3):345–356, 2000.

<sup>6</sup> available at: <http://tinyurl.com/qybeoke>, password “ecmlpkdd\_review”



**Fig. 1.** Supply, demand, and disequilibria of/for Dubstep music in the UK. Left: supply per region, measured by number of independent Dubstep artists divided by population. Center: demand, measured by density of tweets about independent Dubstep artists. Right: net supply vs. demand, which shows that the neighbouring regions of Devon and Torbay have excess demand and supply respectively.