

## **A Matter of Perspective: Taking a Distant View of Music History**

**Andrew Gustar**

### **Introduction**

Investigations into the history of music usually comprise a close examination of individual composers, works, instruments, events, or publications. Musicologists are skilled at unearthing and interrogating a diverse range of sources, analyzing and interpreting the evidence, and collectively building a rich and complex historical narrative. Often overlooked, however, is an alternative methodology which can complement these qualitative studies, and reveal trends and patterns that would otherwise remain invisible. In this article I argue for a quantitative approach which, rather than examining individual entities under the microscope, takes a bird's eye view of whole populations.<sup>1</sup>

Music history is unusually well endowed with population-based data. For centuries, individuals and institutions have collected, listed and catalogued composers, performers, works, concerts, instruments and other elements of the musical world. There are vast numbers of library and publishers' catalogues; biographical dictionaries; composers' complete editions; thematic catalogues; inventories of instruments; and collections of folksongs, hymns, chants, sonatas and other genres. Since the early 20th century the list has expanded to include recording catalogues, record sales charts, and a huge number of other general and specialist lists, collections and databases available in print or online.

Early music is well represented in this universe of data. As well as many primary sources (such as biographical dictionaries,<sup>2</sup> library and publishers' catalogues<sup>3</sup> and repertoire-based publications<sup>4</sup>), there are a growing number of modern datasets focused on early music, covering topics such as printed music and manuscripts,<sup>5</sup> repertoire,<sup>6</sup> concerts<sup>7</sup> and instruments.<sup>8</sup> 'Early music', defined by date, is also covered in many more general sources.

Most of these sources are designed for consulting information about individual items. Quantitative analysis, however, requires data from many or all records in a dataset, the extraction of which is not always straightforward. Handling such data requires some different considerations and techniques from those appropriate to qualitative research, although there are many similarities, in particular the importance of a critical approach to the sources and their interpretation.

At its simplest level, quantitative analysis can indicate the size of populations (such as the number of composers) and the proportions with different characteristics (nationality, period, gender etc). This can be extended to explore geographical distributions, trends and changes over time, or the existence and development of clusters, connections and networks (teacher-pupil relationships, for example). Datasets can be compared to assess their accuracy, representativeness and bias. Snapshots of populations at different times can shed light on the dynamics of survival (of manuscripts or instruments) or of dissemination and reception history.

There are risks in constructing a broad historical narrative by piecing together conclusions from qualitative research. Such studies are, in aggregate, unrepresentative of the broader musical world in that they generally focus on the big names and ignore the many individuals or works about

which little is known. Quantitative techniques can provide hard data to support, refine, or refute hypotheses suggested by qualitative studies. They also give a voice to the multitudes of obscure composers and works which would otherwise remain unknown and ignored.

## **The Population of Early Keyboard Instruments**

As a demonstration, I consider the Boalch-Mould Online (BMO) dataset of early keyboard instruments.<sup>9</sup> I am no expert on keyboard instruments, and apologize for any errors or misunderstandings in what follows: the objective is simply to illustrate some aspects of a statistical approach to studying this sort of data.

The dataset is a snapshot of a subset of a population. The snapshot is based on Boalch and Mould's original published survey, updated by subsequent research (anybody can contribute information to the current database). I will return later to consider the nature of the subset of the population of keyboard instruments represented in BMO.

### ***Gathering and cleaning the data***

Depending on the nature of the source, collecting the data can often be complex and laborious. BMO is relatively straightforward, with a couple of tables that can simply be copied from the website and pasted into a spreadsheet. The first table is a list of instruments (including a reference number, year, type of instrument, maker, present location and compass), and the second table lists makers and their locations. Every instrument also has its own page with further details, which I have not attempted to gather for this exercise, although it would be possible to do so.<sup>10</sup> This would certainly be desirable in a more thorough investigation, in order to take proper account of any unusual characteristics of particular instruments, such as major structural changes or doubts regarding authenticity.

The next step, also usually a lengthy process, is to clean and tidy the data into a useable form. This procedure often highlights problems such as missing information, duplicate records or inconsistent formatting. Tidying, in this case, included:

- use the makers table to give a place of manufacture for each instrument
- splitting the instrument type into a major type ('harpsichord', 'clavichord', etc), and a sub-type ('2M', 'fretted', etc)
- splitting the compass into upper and lower values, and expressing each as a numeric value<sup>11</sup>
- filling in missing or uncertain dates, by taking the mid-point of ranges, or a typical year based on instruments by the same maker.

Such tidying might introduce approximations (such as estimated dates), and ignore some information (such as taking the lowest pitch irrespective of whether it is part of a short octave). Judgements need to be made, such as what to do if a maker worked in two or more centres. If done in a neutral way, this process can improve the statistical weight of a dataset, although care needs to be taken when drawing conclusions.

It is often desirable to augment the data with information from other sources. For example, I assigned latitude and longitude coordinates so that makers' and owners' locations could be plotted on a map. There are online systems that will assist with this process, although it is important to check and manually correct, as far as possible, any missing or suspicious results.

## Exploring the Data

The simplest exploration of any data is to count it. There are 2,374 distinct records in BMO. Of these, 1,329 instruments (56%) have complete information on date, maker, ownership, location and compass: the rest are missing at least one of these items.<sup>12</sup> A year (or estimated year) exists for over 97% of records, the maker is known for 95%, the current owner is recorded for 75%, and the compass for 75%.

More interesting is to count the data in different categories. Table 1, for example, shows the number of instruments by type and quarter-century of manufacture, showing how the Virginal gave way to the Harpsichord and Spinet, with the Clavichord living on after the other types had largely fallen out of fashion:

	Harpsichord	Clavichord	Spinet	Virginal	Other	Total
15C Q4	-	-	-	1	1	2
16C Q1	3	-	-	2	-	5
16C Q2	10	6	-	21	2	39
16C Q3	17	2	1	48	-	68
16C Q4	28	2	5	31	9	75
17C Q1	63	-	1	58	10	132
17C Q2	85	-	3	42	6	136
17C Q3	57	5	5	38	3	108
17C Q4	88	10	64	30	2	194
18C Q1	56	16	69	10	7	158
18C Q2	85	63	84	3	7	242
18C Q3	224	114	126	11	10	485
18C Q4	179	208	52	5	6	450
19C+	28	164	15	13	-	220
<i>Unknown</i>	23	10	16	7	4	60
<b>Total</b>	<b>946</b>	<b>600</b>	<b>441</b>	<b>320</b>	<b>67</b>	<b>2,374</b>

Table 1: Entries in BMO database by type and period

Whilst tables can be useful, trends and patterns are often clearer when presented graphically. It is also often more useful to show proportions rather than absolute numbers. Figure 1 shows countries' changing share of the harpsichord market between 1550 and 1800, with Italy's dominant position being temporarily supplanted by Flemish instruments during the early 17th century, followed by the development of French production from 1650, a brief boom in German manufacture at the turn of the 18th century, followed by British dominance by the end of the period.<sup>13</sup> Undated instruments have been ignored. Although the number of instruments in BMO with missing dates is quite small, this has the potential to distort the apparent trend if, for example, the chance of dates being missing varies significantly by age or country.

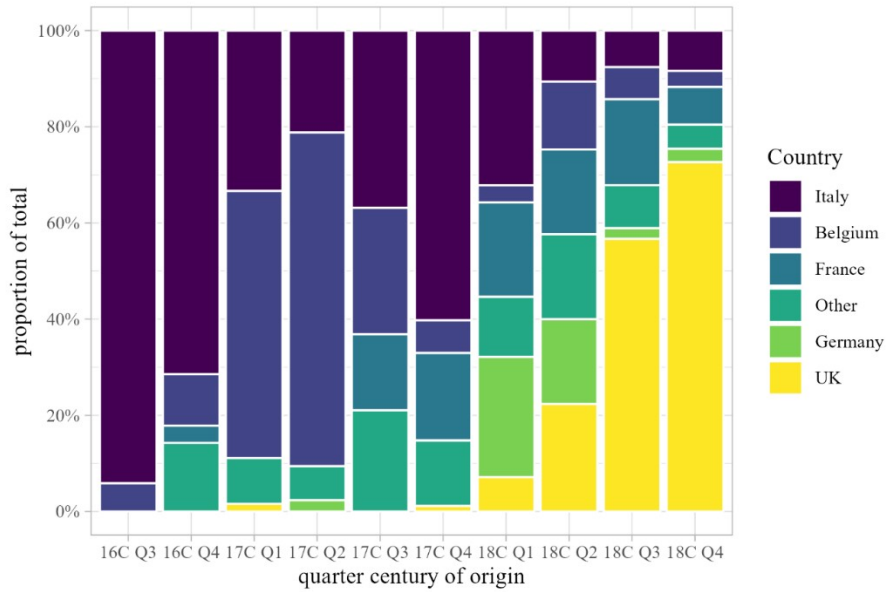


Figure 1 Harpsichords in BMO by country and quarter century of manufacture

The same data can be illustrated with maps, such as the series in Figure 2, showing the relative harpsichord production of different locations over the half-centuries from 1500 to 1800.

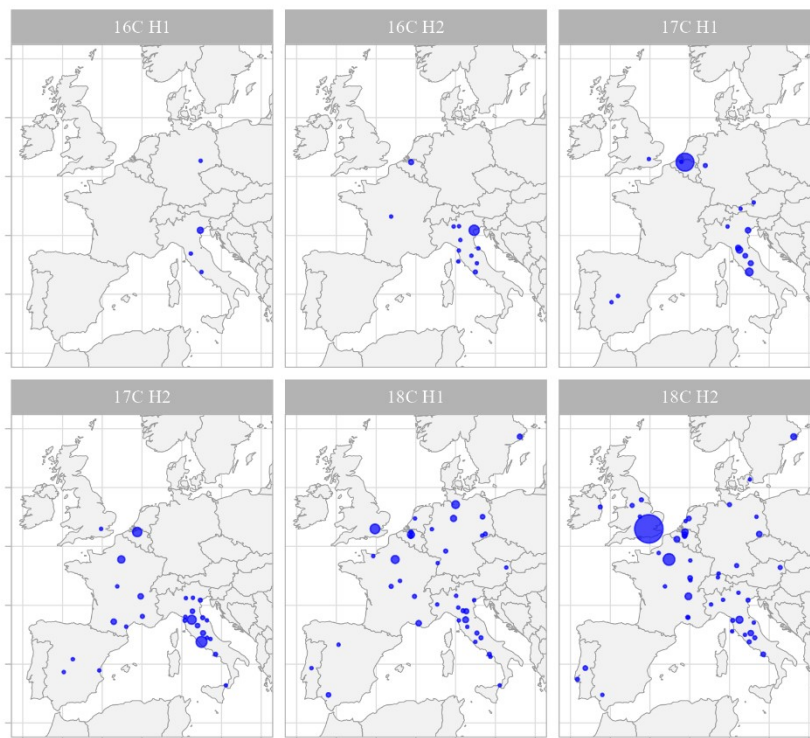


Figure 2 Origins of harpsichords in BMO by half-century of manufacture

About 60% of makers listed in BMO only have one instrument to their name, whereas a handful of families of makers are credited with 50 or more surviving instruments. Figure 3 shows the proportion of maker families against the number of instruments they produced on a *log-log* plot, where both axes increase multiplicatively rather than additively. The points fall roughly on a downward sloping straight line, indicating that makers' productivity follows a 'power law'

distribution, which is a typical consequence of market forces based on fashion and herd behaviour. The implication is that people bought Kirkman or Ruckers instruments largely because other people did so.<sup>14</sup>

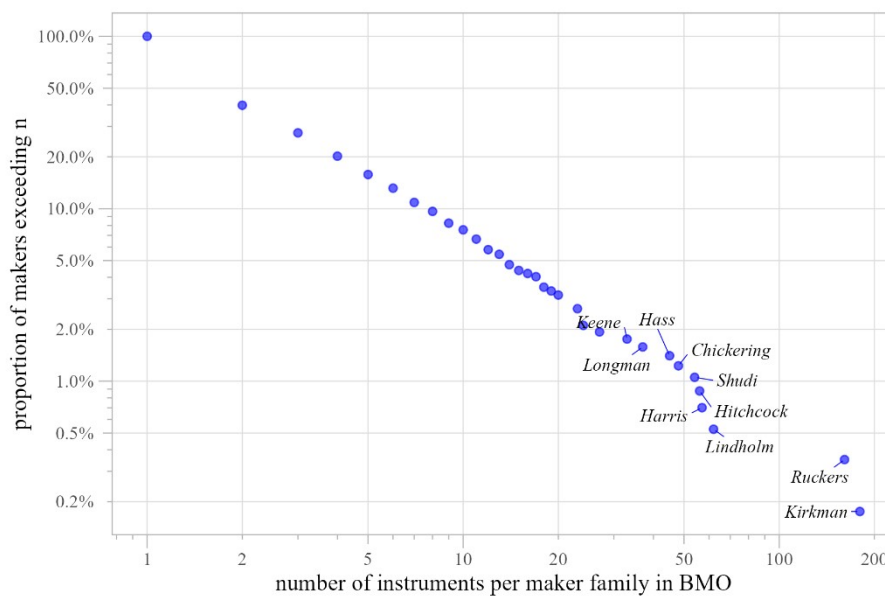


Figure 3 Number of instruments in BMO by maker family (note logarithmic scales)

Figure 4 shows the lowest notes of harpsichords from the five main manufacturing countries. In Italy, C (i.e. two octaves below middle C) has always been most common, with a handful of early examples of GG and, later, FF. 17<sup>th</sup> century Flemish harpsichords were evenly spread between C, FF and GG (interestingly, the Italians focused entirely on C harpsichords while the Flemish were at their peak). From 1680 onwards the French, Germans, and then the British largely ignored C, with FF becoming the standard. The British later even experimented with CC, although this did not catch on. Note that rebuilds (as in the 18th-century French *ravalement*, where compasses were expanded but the original maker date retained), need to be checked when assessing this data.

According to BMO, 22% of recorded harpsichords are currently in the US, followed by 21% in the UK and 15% in Germany. However, BMO is an English language source, with British and American editors, and (as is often the case) there might well be some over-representation in favour of the 'home' language or countries. It can be instructive to compare other sources - ideally from other countries or in different languages - to ascertain the nature of such bias, although in practice the available options are often limited.

Another database of early keyboard instruments is at early-keyboard.com (EK), the website of Tony Chinnery, a British instrument maker based in Tuscany, which lists 1,119 instruments.<sup>15</sup> The UK is home to 23% of the harpsichords listed in EK, with 19% in Germany, and just 18% in the US, suggesting a different regional bias compared with BMO.

Comparing sources also reveals something about the larger population of early keyboard instruments. If, say, only 80% of instruments in EK can be found in BMO, then, if we simplistically assume that these are random samples from the same population,<sup>16</sup> we can argue that BMO's 2,374 instruments only represent about 80% of the population, which would therefore be close to 3,000.

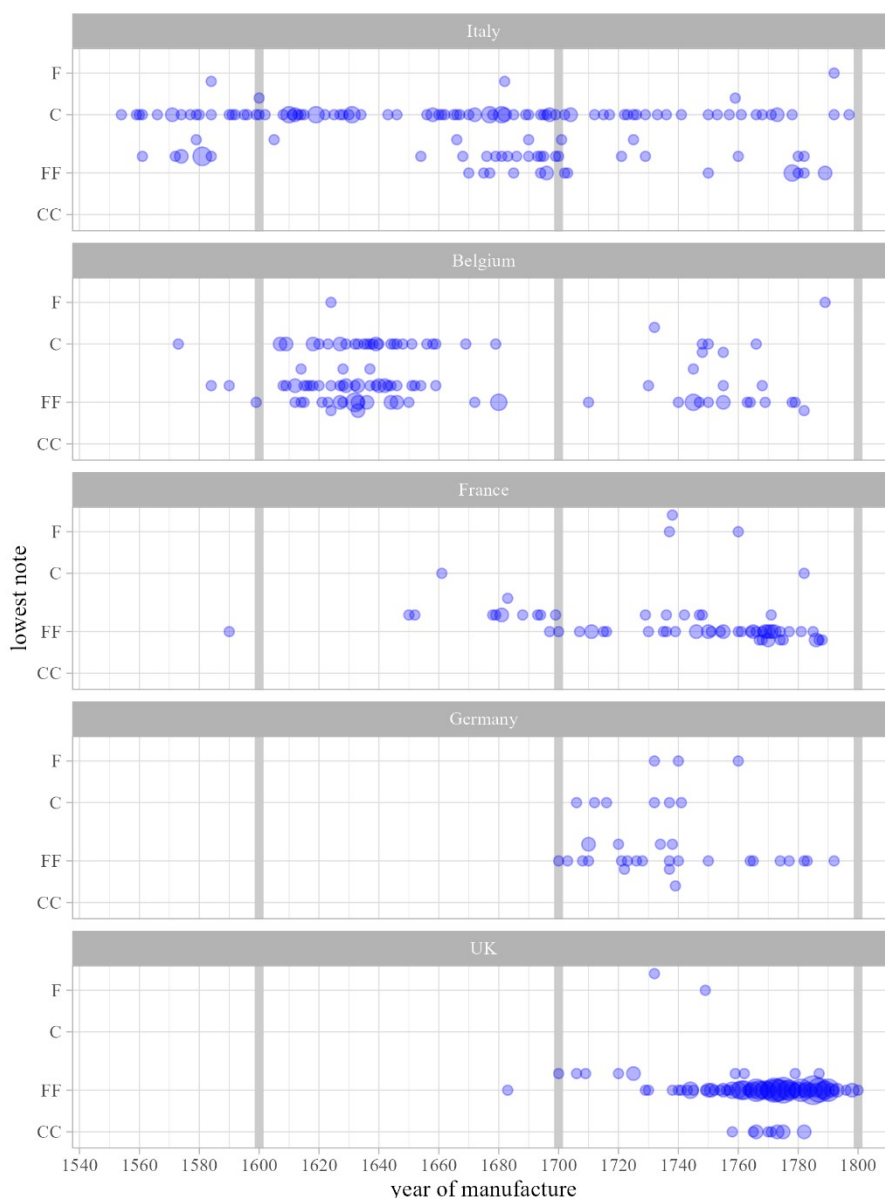


Figure 4 Lowest notes of harpsichords in BMO for main manufacturing countries, by date

Unfortunately, matching records between datasets can be particularly difficult and time-consuming, as sources contain varying information and present it differently. EK and BMO express instrument descriptions, locations, makers' names and owners in different ways, so each of these variables requires a separate exercise to compare the layout, formatting, language and any typographical errors between the two sources. This requires considerable manual intervention and can never be 100% accurate, especially with large datasets.

Having attempted to standardise and deduplicate the data on five key fields (type, year, maker, owner and location), I found that just 42% of instruments in EK matched an entry in BMO on all five fields, with another 37% matching four, and 16% matching three. Many four-field and some three-field matches are probably correct, since some information is missing from one or both datasets, so probably at least 80% of EK instruments can be found in BMO. There are definitely some entries in EK that are not in BMO, but between these and the five-field matches there are many question marks, some of which cannot be confirmed either way because of missing data.



## Discussion

This brief analysis of BMO has aimed to demonstrate the power of a quantitative approach to the many datasets relevant to the history of music, and to illustrate some of the considerations and difficulties that arise. Although this example has covered early keyboard instruments, similar techniques could be used to study populations of composers, works, publications, concerts, or any topic for which data is available. Any dataset offers many potential avenues of investigation, and there are multiple analytical techniques that can be employed. There is no shortage of sources with which a dataset can be combined or compared, and there are limitless ways of presenting the results.

Like any research methodology, the success of this sort of analysis is dependent on the skill, experience and judgement of the researcher, and on the nature and quality of the data. The process is dependent upon, and revealing of, the biases inherent in the datasets. We have seen some indication that the contents of BMO may be skewed by its geographical and linguistic context, and similar biases are common across many such datasets. Other biases might result from aesthetic judgements (such as guides to recommended recordings), the availability of information (more is written about composers about whom most is already known), or differential survival rates (prestigious instruments in wealthy households were less likely to end up as firewood than basic ones from modest homes). It is also generally the case that the available data is skewed towards the regions where music has historically been most avidly collected, catalogued, studied and written about.

Quantitative techniques are perhaps most useful when employed alongside qualitative methods. They can identify and quantify patterns and trends, and are useful for testing hypotheses, but they rarely help resolve the *why* and *how* questions that are best approached qualitatively. Some historical musicologists, of course, will not be comfortable with using statistical techniques, nor with the technological wrangling that is sometimes involved in extracting useable data from complex sources. Help should not be far away, however: plenty of statisticians and computer scientists will be interested in music almost as much as they are excited by the prospect of new sources of data to play with!

Although the focused spotlight will remain an essential component of the historical musicologist's toolbox, hopefully I have demonstrated that things can sometimes be seen more clearly from a high vantage point with a powerful floodlight.

*Andrew Gustar's first degree was in mathematics, and he is now a Visiting Fellow with the Open University, having been awarded a PhD for his thesis, Statistics in Historical Musicology. His website [www.musichistorystats.com](http://www.musichistorystats.com) contains further discussion and examples of quantitative techniques in music history.*

## Notes

<sup>1</sup> The title of this article is inspired by Franco Moretti, *Distant Reading* (London, 2013), who argues for a quantitative and computational approach to the study of literature, in contrast to traditional 'close reading'.

<sup>2</sup> Early biographical dictionaries include Johann Mattheson, *Grundlage einer Ehren-pforte, woran der tüchtigsten Capellmeister, Componisten, Musikgelehrten, Tonkünstler &c. Leben, Wercke, Verdienste &c. erscheinen sollen* (Berlin, 1910 [1740]) and Ernst Ludwig Gerber, *Historisch-biographisches Lexicon der Tonkünstler* (Leipzig, 1790). Even sources such as Charles Burney, *A General History of Music, from the Earliest Ages to the Present Period* (New York and London, 1935 [1789]) have useful indexes of names that can be analysed.

<sup>3</sup> Such as the Portuguese Royal Library catalogue: Paulo Craesbeeck, *Primeira parte do index da livraria de musica do muyto alto* (Lisbon, 1649; Breitkopf & Co., 1762–65). *Catalogo delle sinfonie, partite, overture, soli, duetti, trii, quattri e concerti per il violin, flauto traverso, cembalo ed altri stromenti, che si trovano in manuscritto nella Officina musica di Giovanni Gottlob Breitkopf in Lipsia* (Leipzig, 1762-65).

<sup>4</sup> One such is Leone Allacci, *Drammaturgia* (Venice, 1755). There are also early thematic catalogues such as William Barton, *The Book of Psalms in metre* (London, 1644).

<sup>5</sup> Examples include the Digital Image Archive of Medieval Music, DIAMM (<https://www.diamm.ac.uk/>), or Répertoire International des Sources Musicales, RISM (<https://rism.info/>). Also of interest are lists of lists, such as Robert Eitner, *Bibliographie der Musik-Sammelwerke des XVI. und XVII. Jabrunders* (Berlin, 1877).

<sup>6</sup> Composer complete editions would fall into this category. Other repertoire-based datasets include the Motet Cycles Database (<http://motetcycles.ch/>); Cantus (<https://cantusdatabase.org/>); and Gareth Curtis and Andrew Wathey, 'Fifteenth-Century English Liturgical Music: A List of the Surviving Repertory', *Royal Musical Association Research Chronicle*, xxvii (1994), pp.1-69.

<sup>7</sup> Examples include Michael Tilmouth, 'Calendar of References to Music in Newspapers Published in London and the Provinces (1660–1719)', *Royal Musical Association Research Chronicle*, i (1961) and ii (1962); and Simon McVeigh, *Calendar of London Concerts 1750-1800* (1994) (<https://research.gold.ac.uk/id/eprint/10342/>).

<sup>8</sup> As well as the catalogues of museums and collections, there are more general surveys such as Donald Boalch, rev Charles Mould, *Makers of the Harpsichord and Clavichord 1440-1840* (London, 3/1995) or Frederick Crane, *Extant Medieval Musical Instruments: A Provisional Catalogue by Types* (Iowa City, 1972).

<sup>9</sup> BMO (<https://www.boalch.org>) is the online successor to Boalch and Mould (1995).

<sup>10</sup> The makers and instruments tables can be accessed from the website's 'Reference' menu. A better table of all instruments (including the compass information) can be obtained by using the 'Search' facility to search for the letter 'a' (or any other common character) across all fields.

<sup>11</sup> I used the standard MIDI pitch codes, expressed in semitones starting at CCCC = 0 (so middle C is 60).

<sup>12</sup> Two records (BMO numbers 267 and 1931) are completely blank. On closer inspection, the first is marked as a duplicate, and the other has a note saying that the 'Tannenberg harpsichord' does not exist. Another (2400) may be a dummy entry, as the maker is given as 'testmaker'.

<sup>13</sup> The chart legend attributes Flemish harpsichords to 'Belgium', a country which did not exist until the 19th century. The historical volatility of European borders means that it is often convenient to use modern countries for consistency.

<sup>14</sup> I discuss power laws in music history at more length in Andrew Gustar, 'Fame, Obscurity and Power Laws in Music History', *Empirical Musicology Review*, xiv/3-4 (2020), pp.186–215.

<sup>15</sup> See <https://www.early-keyboard.com/archive/results.php>.

<sup>16</sup> In practice, we cannot assume that EK and BMO are *independent* samples, as one may draw on the other, and both will have used many of the same original sources, such as museum catalogues.