SSOAR

Open Access Repository
www.ssoar.info

Historical patterns based on automatically extracted data: the case of classical composers

Borowiecki, Karol J.; O'Hagan, John W.

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:
Borowiecki, K. J., & O'Hagan, J. W. (2012). Historical patterns based on automatically extracted data: the case of classical composers. *Historical Social Research*, *37*(2), 298-314. https://doi.org/10.12759/hsr.37.2012.2.298-314

Nutzungsbedingungen:
Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
https://creativecommons.org/licenses/by/4.0/deed.de

Terms of use:
This document is made available under a CC BY Licence (Attribution). For more Information see:
https://creativecommons.org/licenses/by/4.0

gesis Leibniz-Institut für Sozialwissenschaften

Mitglied der Leibniz-Gemeinschaft

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-372955

Historical Patterns Based on Automatically Extracted Data: The Case of Classical Composers

Karol J. Borowiecki & John W. O'Hagan [*]

Abstract: *»Historische Muster basiert auf automatisch extrahierten Daten: Der Fall klassischer Komponisten«*. The purpose of this paper is to demonstrate the potential for generating interesting aggregate data on certain aspect of the lives of thousands of composers, and indeed other creative groups, from large on-line dictionaries and to be able to do so relatively quickly. A purpose-built java application that automatically extracts and processes information was developed to generate data on the birth location, occupations and importance (using word count methods) of over 12,000 composers over six centuries. Quantitative measures of the relative importance of different types of music and of the different music instruments over the centuries were also generated. Finally quantitative indicators of the importance of different cities over the different centuries in the lives of these composers are constructed. A range of interesting findings emerge in relation to all of these aspects of the lives of composers, which might provide insight and productive lines of enquiry for further work as to why certain composers were so successful in different historical periods.

Keywords: cliometrics, data collection, geographic concentration, creative individual.

1. Introduction

In an earlier article, O'Hagan and Borowiecki (2010) studied the birth location, migration and clustering patterns of over 500 of the most important composers. Arising from this several interesting research papers resulted (see Borowiecki, 2011, 2012). In particular the data allowed the tracking of the movement over time of each composer and thereby provided insights on their work locations for example during war years and whether or not they were in creative clusters when some of their most important works were produced.

[*] Address all communications to: Karol Jan Borowiecki, Trinity College Dublin, Department of Economics, Dublin 2, Ireland, or alternatively Department of Business and Economics, University of Southern Denmark, Campusvej 55, Odense, Denmark;
e-mail: borowiek@tcd.ie.
John W. O'Hagan, Trinity College Dublin, Department of Economics, Dublin 2, Ireland;
e-mail: johagan@tcd.ie.
We wish to thank Antonios Gkogkakis for the design and implementation of the computer application that was used to generate the underlying data set.

However, there are over 15,000 composers listed in *Grove Music Online*[1] and as such their work had to rely on a sample of less than four per cent of these, albeit the sample included almost all of the most important composers listed. A huge amount of painstaking manual data collection was involved in generating this information over four/five months. The question we faced is there any way we could get summary data on all composers via non-manual means, even if these data were nowhere as rich as the manually-collected data. The purpose of this paper is to demonstrate the potential for generating interesting aggregate data on certain aspect of the lives of thousands of composers, and indeed other creative groups, from large on-line dictionaries and to be able to do so relatively quickly. The contribution of this paper perhaps comes particularly to light if one considers the view that "little credit is given to the generation of new data, but these data are the bedrock of [cliometrics]" (Carlos 2010, 106).

Section 2 sets out the methodology applied to obtain the computer-generated summary data. Sections 3, 4 and 5 provide details of the summary information that resulted from this exercise. Section 6 concludes the paper by examining to what uses the data could be applied, both in a general informational sense and in terms of hypotheses to be tested.

2. Methodology

Grove Music Online is one of the leading online resources for music research and contains more than 50,000 signed articles and 30,000 biographies. Each entry appears on one main web page, which consists of several sections such as summary, life, works, bibliography, and sometimes writings. Each section may have subsections, which sometimes contain links to other web pages (for example to certain life periods of an individual). Currently this information is only accessible via the *Grove* online dictionary website. The dictionary provides search forms and enables navigation through the web pages of the results. Obtaining any larger portion of information becomes very time-consuming, whereas extraction of key elements and statistical analysis is practically not feasible without the aid of a computer application.

In order to overcome this constraint we developed a purpose-built java application that automatically extracts and processes information, similar to that used in other contexts for example by newspapers. The purpose of the underlying application is a first attempt to automate information extraction from *Grove Music Online*. The application conducts a search for all the composers stored in *Grove Music Online*. For all the results (composer entries) it obtains the related

[1] This multivolume dictionary is "a critically organized repository of historically significant information" (Grove 2011 Preface).

stored web pages, in order to extract their content. The information acquired is then processed to extract predefined elements (for example composer's full name). Further processing was then carried out to provide statistical data such as word occurrence of predefined terms and word count in different sections of the results.

The processing done on the set of web pages comprising each composer's life was threefold. First, key elements such as the full name, place of birth, place of death, birth date, death date, nationality and list of occupations were extracted. Second, a word count for all the sections – that is life, works, bibliography and writings – in the result pages was calculated. This calculation takes into account the fact that a section may consists of several web pages. Third, each entry was scanned to count the occurrence of predefined terms in several categories, such as, for example, geographic locations, music instruments or types of works. All the search results are stored in order to allow further investigation of the data.

The entries in the dictionary had to be corrected in two cases.[2] Besides, several adjustments had to be conducted due to erroneous coding of the entries.[3] In 171 cases the information was not coded appropriately in the dictionary and had to be adjusted. For example, some dates of birth were erroneously stored under birth place. Such mistakes can be detected and manually corrected in the database. Next, the data set contained double entries if a composer was listed in both the *Grove Music Online* and the *New Grove Dictionary of Opera*. As a result, 1,331 double entries had to be filtered out. Furthermore, 231 individuals who were identified by the Grove search application as composers, have no such mention in their occupation (and are only described as 'songwriter', 'hymn writer' etc.). Those artists are usually modern musicians and have been dropped from further analysis with the motivation to include only classical composers. Finally, we dropped 66 entries that are for whole families of composers, rather than for an individual.[4]

Following the above we were left with 14,087 composer entries. For all these we are able to calculate the length of each entry and count the occurrence of predefined terms. However when it comes to key elements (such as for example place of birth) information is sometimes missing or incomplete, especially for earlier time periods. Therefore, in order to enable a meaningful statistical analysis we have to constrain the sample to individuals born in or after the 15th century. As a result we were left with 12,201 composers born in or after

---

[2] David L. Downing birth year is 1822 (and not 1922) and Carl Wolfsohn died in 1907 (not 1807).

[3] Incorrect coding is not visible to the reader of the dictionary and can be detected only in the *html*-code. It can however lead to incomplete results when it is searched for certain keywords.

[4] Note, that composers listed in the family entry would usually have separate entries.

the 15th century for whom the birth century and birth country is known. In all cases the occupation list is available while the birth place and death place is known for only 8,728 of those composers.

3. Birth Location and Occupational Profile

Birth Locations

Tables 1 and 2 provide a direct comparison between the birth locations of the 500+ most important composers looked at in O'Hagan and Borowiecki (2010) and the 12,000+ examined in this study. The table covers the 15th to the 20th centuries.

Table 1 provides the basic data that can be compared with Table 1 in O'Hagan and Borowiecki (2010). The broad picture there is confirmed when the much larger sample is examined. These are the very strong positions of Italy in the 15th to 17th centuries and Germany from the 15th to the 19th centuries, and the rise of the US in the 19th and 20th centuries. In the later centuries though there was nothing like the concentration of earlier centuries: for example, the US accounted for 18 per cent of all composers in the 20th century, whereas in the 15th to 18th centuries the Germanic countries accounted for around 25 per cent of the total in each century. Looked at differently, the 'other' categories (Eastern Europe, Rest of Europe, Rest of World) accounted for just 20 per cent of the total in the 18th century but 50 per cent of the total in the 20th century.

Table 2 provides a direct comparison of these patterns when looking at the top 500+ composers and the full sample of 12,000+ composers. A value of 1.00 in this table indicates that a country's share of each was the same, whereas a value above 1.00 indicates that its share of top composers was relatively higher than that of all listed composers and the opposite for a value below 1.00. The lowest value possible, 0.00, simply indicates that the country had no one in the top 500+ in that century. This then allows us to address the questions whether the distribution by birth location of the 12,000+ composers is significantly different to that of the 500+ composers. Furthermore, it can be observed in which centuries and countries the biggest differences apply.

There are very marked differences in the distribution by birth location when the two groups are compared. One comparison of note is that between Britain and Germany. In the case of Germany from the 17th to the 20th centuries its share in the top 500+ significantly exceeds its share in the top 12,000+ (the same applied to France in every century bar the 18th) whereas the exact opposite applies in the case of Britain. This might suggest a country-specific bias in *Grove*. The top 500+ were chosen on the basis of many musical sources (see Murray, 2003) from many different countries whereas the *Grove* is a British publication. This does not explain the whole picture though, as values of well

below 1.00 are also evident for the Netherlands, although this could be a small numbers issue.

The other interesting comparison is that by century. As seen in Table 1, there is a dramatically different distribution for the 12,000+ composers compared to that for the 500+ composers, which might be expected. The biggest differences are for the 16th and 20th centuries; for the former little information is probably available, except for the better-known composers. The opposite would apply in the 20th century where so much information appears to be available for a huge range of composers, only a small few of whom would rank in the top 500+ over the centuries.

Occupations

Table 3 examines the occupations of the 12,000+ composers and provides information whether the primary occupation was listed as composer. Furthermore, it can be analyzed how many of them had other occupations and if so how many other occupations? Finally, an interesting comparison between different centuries and different countries can be conducted.

As may be seen in the first column of the table, for the vast majority their main employment was composing. However, there are some interesting variations that are difficult to understand. The first is the dip in the 18th century, where only 58 per cent are listed as having composing as their primary occupation. The second is the rise in this figure to 88 per cent in the 20th century.

The picture that emerges in relation to other occupations is one that is well known in relation to creative people like composers and visual artists (see Benhamou, 2011). A large proportion of composers over the centuries relied on income from occupations other than composing. This varied from 74 per cent in the 19th century, to 45 per cent or more in the 15th and 20th centuries. Indeed in the 19th centuries over 30 per cent of all composers had three or more occupations. These are high figures especially when one considers that it is only the most successful composers that would be included in *Grove* and hence the group most likely to receive the lion's share of their income from composing. In relation to the 20th century what is surprising is that while 88 per cent had composing as their primary occupation, almost half of the total still had another occupation.

4. Importance of Composers, Types of Music, Instruments Used

Word Counts per Composer

Table 4 provides some information that throws light on the importance of the various composers by century and country; the word count measures the num-

ber of words in the main description (i.e. life section) of each biographical entry.[5] The method is crude of course and is taking word count as an indicator of importance but yet is informative (see Kelly and O'Hagan 2007; O'Hagan and Kelly 2005 and O'Hagan and Hellmanzik 2008).

As can be seen there is a marked variation in word count per composer by century and by country. As one might expect, there is the lowest word count per composer in the 20th century: because information is available on so many composers it is likely that entries could be included for many less important composers than in the earlier centuries. This could also explain the high figure for the 15th century; information might only have been available for the top composers and as such one would expect a much higher average entry per composer.

The variation by country is harder to interpret. What is measured in the table is word count per composer for each country and each century. To measure the importance of each country we would need also to combine this information with that in Table 1. Comparing the average for each country to that for all countries the following emerges.

The prominence of France and Germany in the 17th to the 20th century is even more marked than indicated in Table 1. The opposite applies in the case of the Netherlands and US. In the case of Britain the word count is also above the average for these centuries but this again may reflect a country-specific bias given that the source of this information is British. This then would be a double bias; namely a disproportionate number of composers listed for Britain (compared to the top 500+ ranking based on multiple, international sources) and a higher word count per composer even though many of these might be considered less important.

Importance of Different Types of Music

Table 5 provides information in relation to the types of music and instruments which figured most frequently in relation to the various composers. We concentrate in this section on the former.

By way of explanation, each biographical entry was scanned for a set of predefined terms (e.g. 'symphony'). Those terms have been counted and assorted into a group of music types, consisting of concert works (symphony, sinfonietta, symphonic suite, symphonic, tone poem, rhapsody, overture, oratorio, waltz), chamber works (chamber, sonata, quartet, art song, cantata, scherzo, motet), theatre works (ballet, opera, incidental music, zarzuela, operetta, li-

[5] We also compiled the word count for each composer in the works, bibliography, and writings sections and found a very high correlation between these and the main biographical entry. The average word count for the latter was 516.6, and 257.0, 120.4 and 10.6 for the other three.

bretto), church works (mass, church cantata, requiem, oratorio) or march works (march).

These are certainly broad categories but are perhaps a useful first step in establishing the type of information that can be made available. In relation to each category we calculated the word count per composer and also per 1,000 words; to check for any biases resulting from say some important composers having very long entries and hence having several references say in the case of Bach to organ music. These two different measures are presented alongside each other in Table 5. Some interesting findings emerge.

- The count per thousand words for the church music category in the 15th century was 1.44 (the top for any category for that century) but its rank declined thereafter, dropping to a figure of only 0.33 in the 20th century.
- In contrast, the word count for the theatre category rose from a low of 0.05 in the 15th century to a high of 3.40 in the 19th century, dropping back to 2.38 in the 20th century.
- A similar but less dramatic story applied to the chamber music category, although even in the 15th century it had already a word count of 1.11. This figure had risen to 2.38 in the 20th century putting it as the top category in that century.
- The word count for the march category has been remarkably consistent over the centuries: for example 0.65 in the 15th century and 0.59 in the 20th century.
- While the word count for the concert category was extremely low in the 15th and 16th centuries, it did not vary dramatically over the following centuries and was never ranked higher than third out of the five categories.

Music Instruments

The categorisation of music instruments into different groups is even more problematic than that for music types, but further subdivision is possible should the broad categories presented in Table 5 prove of interest.

The group of music instruments were divided into the following: violin family instruments (violin, viola, cello, double bass, viol), lute/guitar family (classical guitar, lute, mandolin, harp), keyboard family (clavichord, piano, harpsichord, organ), woodwind family (bassoon, clarinet, English horn, flute, oboe, piccolo, saxophone, recorder), brass family (French horn, trombone, trumpet, tuba) and percussion instruments (snare drum, tenor drum, bass drum, timpani, tambourine, cymbals, gong, triangle, vibraphone, xylophone, marimba). As with music types the word counts per composer and per 1,000 words are presented in Table 5, with only the more meaningful latter measure included in the table.

A number of interesting findings again emerge in relation to families of music instruments.

- The keyboard family has held its prominence over the centuries, especially in the 19th and 20th centuries. For example the word count per thousand words for this family of instruments was 0.66 in the 15th century compared to the next ranked (lute/guitar) of 0.32: the figure had risen to 3.30 in the 20th century, compared to the next ranked (violin) of 1.18.
- The violin family ranked a low third in the 15th and 16th centuries, but first in the 17th century and a strong second thereafter.
- The word count for the lute/guitar family has been fairly stable over the centuries but its word count value declined markedly relative to that for the keyboard and violin families from the 17th century on and by the 19th century was ranked way behind these two.
- The word count for the woodwind family has also been well behind those for the keyboard and violin families and its ranking (third or fourth) has interchanged with that for the lute/guitar families over the different centuries.
- The brass and percussion families have been consistently ranked fifth and sixth and their word count scores were much lower in recent centuries than even the third and fourth ranked categories.

5. Important Cities

City Citations

Table 6 provides information on the number of times a city was mentioned per thousand words in the biographical/life section of composers. What this then gives us perhaps is a snapshot of the changing importance of cities over the centuries in the lives of composers. A fairly clear picture emerges.

- Paris stands out as the most important by far in this measure: it was ranked fifth in the 15th and 16th centuries but first in the 17th to 20th centuries inclusive. In terms of actual word count the values range from 0.38 in the 15th century, to a high of 2.31 in the 18th century, but still at 2.01 in the 19th century.
- London was the next highest ranked: eight in the 15th century, seventh in the 16th, fifth in the 17th and in second in the 18th and 19th centuries, and fourth in the 20th century. Its word count value as with Paris reached its peak in the 18th century, at 1.56, dropping to 0.62 by the 20th century.
- Vienna also had high word counts, especially when one takes into account the fact that it has always been a much smaller city than either London or Paris. Its word count value peaked at 1.27 in the 18th century, it was ranked above New York in the 19th century and was still ranked seventh the 20th century.
- Rome topped the rankings in the 15th century, dropped to second in the 16th and 17th centuries, to eight in the 18th century, ninth in the 19th century but back at sixth in the 20th century.

- Berlin appears in the top ten ranking only in the 18th century, and was ranked third most important city in both the 19th and 20th centuries, reaching a word count high of 1.21 in the 19th century.
- New York entered the rankings in the 19th century (at fifth place) and stood at second place in the 20th century. Its highest word count score was 0.89, but this was in the 20th century when the word count score for all cities dropped to significantly lower levels than in other centuries.
- Finally as can be seen in Table 6 several other cities figured prominently in one or two centuries but were not ranked in the top ten in most centuries; for example in earlier centuries, Leipzig, Nuremburg, and Venice, and in more recent decades Moscow and Prague.

Birth and Death Locations

Table 7 is another attempt to measure the importance of cities, in terms of number of composers who were born and/or died in each of the cities listed. The six cities looked at here are: Berlin, London, Paris, New York, Rome and Vienna as these have been predominant locations across most studied centuries. Again a number of key findings emerge.

- The results support the main finding of Table 6, namely the prominence of Paris. Even in terms of number of composers born it dominates for most of the time covered. London had more composers born in it than Paris in the 19th century, but only just and this could result from the country-bias referred to already.
- The difference between number of births and deaths can give some insight in terms of the importance of a city also, as if the number of deaths greatly exceeds the number of births this would suggest significant migration to that city in the period in question. Again Paris stands out in this regard, with the number of deaths of composers in the city greatly exceeding the number of births in most centuries.
- This is true for some periods also for London and New York, confirming also their major importance as centres for composers.
- Berlin and Vienna are again prominent in terms of both births and the excess of deaths over births. Indeed, if one was to combine these two Germanic cities (which together would still be a good bit smaller than London, Paris or New York) their importance becomes even more significant.

6. Concluding Comments

The main purpose of this paper was to illustrate how one can obtain and exploit automatically data, produced in a fraction of the time taken to compile some of these data manually even from on-line sources and with avoidance of man-made mistakes, which could prove useful to historic study. Even on the basis of

what has been produced from this short project we feel that useful findings have been generated. Some could argue of course that the broad picture, outlined in the paper, based on the automatically-extracted data, are well known already. There are two responses to this. First is this true? Second even if the broad findings are well known in a general sense this paper provides specific quantitative evidence, based on a huge number of composers, and based on very explicit methodology.

The paper adds considerable knowledge to the findings of O'Hagan and Borowiecki (2010) in the sense not only that it covers 12,000+ composers (as opposed to 500+ there) but also in terms of identifying key cities, occupational profiles, types of work and music instruments. This is despite the fact their paper involved many more months of data collection than the work for the current article. The key advantage of the O'Hagan and Borowiecki (2010) study though is that for its construction they also compiled detailed data on the year to year work locations of the 500+ composers, which enabled very useful further work to be undertaken.[6]

It is also possible to build on the findings of this work and consider many hypotheses to be tested later, perhaps involving further data collection. It could be asked was the availability of certain instruments in particular historic periods a key factor in the success of some composers, for example Liszt and the piano, Rodrigo and the guitar, Brahms and large orchestras, etc. It might also be asked whether or not migration to clusters of composers was more essential in some periods than others; for example if church music was the main focus in earlier centuries this would have implications for where people worked, their sources of funding and also the talents of some composers relative to others. The questions asked in O'Hagan and Borowiecki (2010) in relation to cities very prominent in the classical musical world could be asked with even more force, as the dominance of five/six cities over the centuries is even more marked using the data here than in their paper. Why was Paris such a dominant centre for composers and visual artists but not for scientists for example? And why was it so dominant in the classical musical world? In a similar vein it could be asked why was a small city like Vienna such a major centre for musical activity? The findings also throw light in relation to current research on the work patterns of creative people. It is not a recent phenomenon that composers had to rely on several occupations for their income and much has been written on the implications of this in a modern context (see Benhamou, 2011).

What has been conducted here of course could be applied also to other large electronically-available information sources relating say to physicists, chemists, visual artists, literary artists, philosophers, soft-ware designers, architects and (and even economists and historians!). What this would allow is a broad

[6] See Borowiecki, 2011a and 2012.

brush overview of the different creative occupations in relation to birth locations, prominent cities, types of activity, etc, and in this context in attempting to explain the different patterns might throw light on the patterns observed for each individual activity. Such new methods could mark an important if small step forward perhaps in quantitative historical enquiry.

Appendix

Table 1: Total Number (share per century) of Births of all Composers (N=12,201)

Century of birth	It	Low	Fr	Ger	Brit	Ru	Sp	EE	RoE	US	RoW	Total	Share by century	Share of 500+ composers by century
15th	35 (0.16)	37 (0.17)	30 (0.13)	56 (0.25)	40 (0.18)	0 (0.00)	17 (0.08)	1 (0.00)	7 (0-03)	0 (0.00)	0 (0.00)	223	(0.02)	(0.08)
16th	449 (0.40)	110 (0.10)	79 (0.07)	240 (0.21)	118 (0.10)	0 (0.00)	64 (0.06)	28 (0.02)	36 (0.03)	0 (0.00)	4 (0.00)	1128	(0.09)	(0.021)
17th	409 (0.33)	53 (0.04)	157 (0.13)	318 (0.26)	123 (0.10)	0 (0.00)	69 (0.06)	36 (0.03)	52 (0.04)	1 (0.00)	11 (0.01)	1229	(0.10)	(0.18)
18th	364 (0.19)	61 (0.03)	263 (0.14)	522 (027)	199 (0.10)	25 (0.01)	72 (0.04)	200 (0.10)	147 (0.008)	50 (0.03)	29 (0.02)	1932	(0.16)	(0.19)
19th	235 (0.08)	121 (0.04)	257 (0.08)	465 (0.15)	321 (0.10)	134 (0.04)	131 (0.04)	569 (0.18)	278 (0.09)	338 (0.11)	261 (0.08)	3110	(0.25)	(0.29)
20th	133 (0.03)	159 (0.03)	173 (0.04)	383 (0.08)	394 (0.09)	170 (0.04)	87 (0.02)	917 (0.20)	476 (0.10)	776 (0.17)	911 (0.20)	4579	(0.38)	(0.05)
1900-49	113 (0.03)	124 (0.03)	148 (0.04)	317 (0.08)	319 (0.08)	140 (0.04)	65 (0.02)	785 (0.21)	379 (0.10)	663 (0.18)	726 (0.19)	3779	(0.31)	(0.05)
Total	1625	541	959	1984	1195	329	440	1751	996	1165	1216	12201		

Note: The last composer presented in O'Hagan and Borowiecki (2010) was born in 1911. To facilitate interpretation, an additional category for composers born 1900-1949 has been added. It = Italy; Low = Low Countries; Fr = France; Ger = Germanic Countries; Brit = British Isles; Ru = Russia; Sp = Spain; EE = Eastern Europe; RoE = Rest of Europe; US = United States; RoW = Rest of World.

Table 2: Share of Top in All Composers

Century of birth	It	Low	Fr	Ger	Brit	Ru	Sp	EE	RoE	US	RoW
15th	1.06	2.15	1.77	0.66	0.40		0.00	0.00			
16th	0.85	1.18	1.37	0.081	1.84		1.36	0.39	0.00		
17th	1.30	0.26	1.22	1.25	0.44		0.00	0.76	0.26	0.00	0.00
18th	1.16	0.33	1.07	1.70	0.40	0.00	0.28	1.01	0.14	0.00	0.00
19th	1.18	0.35	2.82	1.37	0.60	3.18	0.49	0.52	0.46	0.82	0.16
20th	1.38	1.15	4.23	3.83	0.93	2.15	2.11	0.00	0.38	1.18	0.00

Note: See Table 1.

Table 3: Occupational Profile by Century (N=12,334)

	Importance of occupation as 'composer'			Number of occupations			
Century of birth	Primary	Secondary	Total	1 Occupation	2	3+	Total
15th	0.70	0.30	1.00	0.54	0.35	0.12	1.00
16th	0.77	0.23	1.00	0.38	0.43	0.19	1.00
17th	0.70	0.30	1.00	0.31	0.48	0.21	1.00
18th	0.58	0.42	1.00	0.28	0.47	0.25	1.00
19th	0.64	0.36	1.00	0.26	0.43	0.31	1.00
20th	0.88	0.12	1.00	0.53	0.33	0.14	1.00

Note: Each biography in Grove (2011) contains a list of occupations. If the occupation list begins with 'composer', or equivalent, the individual is marked as 'Primary' composer, or 'Secondary' otherwise. The number of occupations listed is measured in the later part of the table.

Table 4: Importance of Composers. Word Count in the Life Section (N=14,087)

Century of birth	It	Low	Fr	Ger	Brit	Ru	Sp	EE	RoE	US	RoW	Average
15th	725.6	1994.1	1200.0	498.6	505.0		576.4	381.0	630.0			813.8
16th	542.8	600.4	441.9	444.9	877.1		527.8	377.3	334.6		504.0	516.8
17th	580.4	270.0	652.1	547.6	585.3		351.5	424.7	320.8	282.0	410.4	442.5
18th	608.9	344.9	470.3	519.3	555.2	457.0	331.6	388.8	395.9	260.4	360.4	426.6
19th	498.3	324.3	787.2	606.2	543.6	916.8	408.5	591.0	465.3	444.7	330.5	537.9
20th	498.5	312.7	446.1	381.4	450.2	348.7	266.6	288.9	304.9	392.1	293.1	362.1
Average	575.8	641.1	666.3	499.6	586.1	574.2	410.4	408.6	408.6	344.8	379.7	516.6

Note: The word count measures the number of words in the life section of each biographical entry.

Table 5: Importance of Types of Works and Instruments.
Word Count per Thousand Words in the Life Section (N=14,087)

Century of birth	Type of work		Type of instrument	
	Count (per thousand words)		Count (per thousand words)	
15th	church	1.44	keyboard	0.66
	chamber	1.11	guitar	0.32
	march	0.65	violin	0.03
	theater	0.05	woodwind	0.01
	concert	0.00	brass	0.01
			percussion	0.00
16th	chamber	1.02	keyboard	0.76
	march	0.64	guitar	0.63
	church	0.54	violin	0.37
	theater	0.18	woodwind	0.06
	concert	0.07	brass	0.05
			percussion	0.01
17th	theater	1.45	violin	1.58
	chamber	1.30	keyboard	1.36
	march	0.91	guitar	0.51
	church	0.70	woodwind	0.35
	concert	0.48	brass	0.14
			percussion	0.03
18th	theater	2.41	keyboard	3.07
	chamber	1.61	violin	2.61
	march	1.00	woodwind	0.93
	concert	0.73	guitar	0.90
	church	0.69	brass	0.05
			percussion	0.03
19th	theater	3.40	keyboard	5.27
	chamber	2.14	violin	1.74
	concert	1.61	guitar	0.44
	march	1.01	woodwind	0.28
	church	0.43	brass	0.06
			percussion	0.03
20th	chamber	2.38	keyboard	3.30
	theater	1.98	violin	1.18
	concert	1.40	woodwind	0.69
	march	0.59	guitar	0.42
	church	0.33	brass	0.16
			percussion	0.07

Note: The word count measures the occurrence of predefined terms grouped into the above categories per thousand words in the life section.

Table 6: Importance of Cities. Word Count per Thousand Words in the Life Section (N=14,087)

15th		16th		17th		18th		19th		20th	
Name	Count	Name	Count	Name	Count	Name	Count	Name	Count	Name	Count
Rome	0.694	Venice	1.163	Paris	1.108	Paris	2.31	Paris	2.071	Paris	1.065
Nuremberg	0.584	Rome	1.047	Rome	0.979	London	1.56	London	1.318	New York	0.822
Florence	0.472	Naples	0.745	Bologna	0.868	Vienna	1.267	Berlin	1.214	Berlin	0.654
Venice	0.434	Milan	0.551	Venice	0.832	Naples	0.879	Vienna	1.033	London	0.620
Strasbourg	0.383	Bologna	0.443	London	0.748	Berlin	0.726	New York	0.792	Moscow	0.455
Paris	0.375	Paris	0.442	Naples	0.561	Munich	0.511	Leipzig	0.732	Rome	0.442
Vienna	0.309	London	0.342	Vienna	0.497	Prague	0.495	Warsaw	0.654	Vienna	0.440
London	0.294	Florence	0.33	Hamburg	0.403	Rome	0.462	Prague	0.596	Prague	0.412
Leipzig	0.236	Nuremberg	0.317	Leipzig	0.372	Leipzig	0.452	Rome	0.482	Bucharest	0.295
Geneva	0.209	Madrid	0.312	Dresden	0.341	Venice	0.431	Moscow	0.422	Budapest	0.247

Note: the word count measures the occurrence of city names per thousand words in the main description.

Table 7: Number of Births and Deaths in Important Cities per Century (N=8,728)

	Paris		London		Vienna		Rome		Berlin		New York	
Century of birth	Births	Deaths	Births	Deaths	Births	Deaths	Births	Deaths	Births	Deaths	Births	Deaths
15th	2	4	2	2	0	3	0	9	0	0	0	0
16th	16	26	11	43	2	14	24	57	2	4	0	0
17th	64	96	28	67	8	44	52	60	2	5	0	0
18th	84	249	77	141	65	119	33	24	22	62	0	4
19th	117	227	119	168	73	83	20	37	43	71	23	117

Note: 'Births'/'Deaths' measure the number of composer births/deaths in a given city and century. The birth place is available for 12,333 individuals and the death place for 9,280. The data used here is the intersection of both, i.e. 8,728 observations. 20th century has been omitted as for most composers the death place is not known.

References

Benhamou, Francoise. 2011. Artists' labour market. In *A Handbook of Cultural Economics*, ed. Ruth Towse, 2nd edition. Cheltenham: Edward Elgar.

Borowiecki, Karol J. 2011. Geographic clustering and productivity: An instrumental variable approach for classical composers. *Trinity Economics Papers No. 0611.*

Borowiecki, Karol J. 2012. Are composers different? Historical evidence on conflict-induced migration (1816-1997). *European Review in Economic History*, forthcoming.

Carlos, Ann. 2010. Reflection on reflections: review essay on reflections on the cliometric revolution: conversations with economic historians. *Cliometrica* 4: 97-111.

Grove Music Online. 2011. Oxford Music Online, <http://www.oxfordmusiconline .com> (accessed on 5-8 March 2011).

O'Hagan, John, and Kelly Elish. 2007. Geographic clustering of economic activity: The case of prominent western visual artists. *Journal of Cultural Economics* 31: 109-28.

O'Hagan John, and Karol J. Borowiecki. 2010. Birth Location, Migration and Clustering of Important Composers: Historical Patterns. *Historical Methods* 43: 81-90.

O'Hagan John, and C. Hellmanzik. 2008. Clustering and migration of important visual artists: Broad historical evidence. *Historical Methods* 41: 121-36.

O'Hagan, John, and E. Kelly. 2005. Identifying the most important artists in a historical context: Methods used and initial results. *Historical Methods* 38: 118-25.