

# Data Appendix to “Agglomeration: A Dynamic Approach”

W. Walker Hanlon and Antonio Miscio

November 6, 2014

## Contents

<b>1</b>	<b>The British City-Industry Database</b>	<b>2</b>
1.1	Overview of the data source . . . . .	2
1.2	Collection of the British Census . . . . .	2
1.3	Householder’s schedules . . . . .	6
1.4	Geographies . . . . .	7
1.5	Occupation categories . . . . .	9
1.6	Exploring the British City-Industry data . . . . .	16
<b>2</b>	<b>Firm size data from the 1851 Census</b>	<b>17</b>
<b>3</b>	<b>The technology similarity matrix</b>	<b>18</b>
<b>4</b>	<b>Additional material</b>	<b>23</b>

# 1 The British City-Industry Database

## 1.1 Overview of the data source

The British Census of Population is a rich source for information on the long-run development of British cities. C.H. Lee (1979) calls the Census data, “the best single indicator of structural change in the British economy and its component regions in the nineteenth and early twentieth centuries.” But this is also a data source that must be used with some care. The purpose of this document is to provide an overview of the issues a researcher should be aware of when using the Census of Population to understand British cities. It also introduces a set of new datasets which have been constructed for this purpose. For a more detailed review of the British Census of Population data, see Lee (1979) and Higgs (1989).

This project builds on past work, particularly that of Lee (1979), but we differ from previous work in that we focus on cities, rather than the more aggregated units considered by previous researchers. We concentrate on the period from 1851-1911. This is a particularly interesting period because it provides a long and relatively stable window in which reliable data are available to track the development of British cities. It is also one for which a consistent series of city-industry data can be constructed. As noted by Lee (1979), there is a discontinuity in the treatment of distributive trades after 1911 which makes it difficult to generate consistent occupation categories that include both data after 1911 and data prior to 1901. Data from 1841 could potentially be added to our series, but there are some questions about the quality of the data in that year, which represented the Census Office’s first experience with the collection and categorization of detailed occupation data.

We begin with an overview of the data collection procedures used by the British Census. We then deal with the core issues faced by researchers using these data. The first major issue has to do with the geographies used to define different reporting areas, and their evolution over time. We then turn to a discussion of the occupation data provided by the census. These are some of the most interesting and useful data provided by the census, but also some of the most fraught with potential data issues. Finally, we consider some of the unique supplementary data provided by various censuses.

## 1.2 Collection of the British Census

By 1841, the basic approach to data collection that would persist through 1911 had been established by the British Census. The first principle of the collection effort was to capture, at a specific point in time, the exact features of the population. The appointed day for each census from 1841-1911 is given in the table below. For each of these, the census endeavors to capture the location of every person in Britain on the midnight preceding the appointed

day. By 1851, the standard practice was to record the location of each person on a Sunday night near to the first day of April and to collect the recorded information on the following Monday.

Table 1: Dates of the census 1851-1911

Year	Census date	Year	Census date
1841	June 6	1881	April 3
1851	March 30	1891	April 5
1861	April 7	1901	March 30
1871	April 2	1911	April 2

Taking the Census required substantial effort and organization. At the bottom of the organizational pyramid were Enumerators, each of which was responsible for visiting each house in an enumeration district. Supervising the enumerators were Registrars. Each Registrar was responsible for a registration sub-district which they were responsible for dividing into properly sized enumeration districts. The registrars were also responsible for hiring the enumerators. Clear instructions were provided to the registrars on the qualifications required when hiring enumerators<sup>1</sup>:

*The Enumerator, in order fulfill his duties properly, must be a person of intelligence and activity: he must read and write well, and have some knowledge of arithmetic: he must not be infirm, nor of such weak health as may render him unable to undergo the requisite exertion, he should not be younger than eighteen years of age, nor older than sixty five: he must be temperate, orderly and respectable, and be such a person as is likely to conduct himself with strict propriety, and to deserve the good-will of the inhabitants of his District. He should also be well acquainted with the District in which he will be required to act; and it will be an additional recommendation if his occupations have been in any degree of a similar kind.*

Above the Registrars were appointed Superintendent Registrars, each responsible for a registrar's district. Table 2 gives an idea of the number of each of these groups employed in the various years.

The same basic procedure was used in each of these census years. To collect data from individual households, the enumerators dropped off a household schedule several days before the appointed census date. The households were then responsible for filling out this

---

<sup>1</sup>"Instructions to Registrars of Births and Deaths," 13th December, 1850, George Graham, Registrar General, p. 1.

Table 2: Registrars and Enumerators Employed, 1851-1911

<b>Year</b>	<b>Superintendent Registrars</b>	<b>Registrars</b>	<b>Enumerators</b>
1851	624	2,190	30,610
1861	624	2,194	30,329
1871	626	2,195	32,543
1881	630	2,175	34,711
1891	633	2,122	35,507
1901			
1911		2,035	over 35,000

schedule with the relevant information for each individual present on the evening of the Census day. Those who worked at night, such as coal miners, were also included on the households lists, even if they were at work at midnight of the census day. The next day, the enumerator would then visit each of the houses in their enumeration district to pick up the household schedules. Where the schedule was incomplete or had not been filled out, it was the responsibility of the enumerator to interview someone in the household in order to complete the form. Enumeration districts were designed to be sufficiently small that the enumerator could visit each household in one day. Thus, the instructions to the registrars specified that the enumeration district should not exceed 200 houses and should not require the enumerator to travel more than 15 miles. Moreover, an effort was made to keep any single enumeration district from spilling over into multiple geographic areas for which results would be produced.

Separate procedures were used to enumerate travelers, those on ships and barges, and those in public institutions such as schools, jails, or hospitals. Travelers were enumerated based on the inn at which they spent the night or, if traveling at night, the location at which they arrived in the morning. Those on ships and barges were enumerated separately by the Customs service. For public institutions, the registrar was responsible for providing the head of the institution with a larger enumeration schedule and collecting it from them on the day after the census date. As for the homeless, in 1911 we know that the police were responsible for their enumeration. For earlier years it is not clear how they were included, though at least by 1871 we know that “The persons sleeping out of doors, in barns tents, barges, vessels of every kind, men working in mines at night, and people traveling in railway carriages and other conveyances were specially looked after”.

Once the household schedules were collected and checked, the enumerator would then enter the data in his enumeration book. The book, together with all of the householder’s

schedules, were then delivered to the registrar. The registrar would then review the enumerators books, as well as the returns from the public institutions which he has collected, to ensure their accuracy. After certifying their accuracy, the registrar delivered the books (but not the household schedules) to the superintendent registrar, who would also review them, before forwarding them to the census office. At the census office, the information would be reviewed one last time, and then tabulated into the census abstracts from which our data are drawn.

These procedures remained fairly constant across the 1851-1911 period, with some minor changes. For example, in 1911 the procedure was changed to eliminate the copying of households schedules into enumerator's books. Instead, the central census office tabulated results directly from the household returns. Also, in 1911, a specially designed set of maps were produced to improve the accuracy with with the limits of different areas were defined.

The Census office was keenly aware that the cooperation of the public played a crucial role in determining their success. The Census Office used newspapers and schools to inform the public of the importance of filling out their census returns. Thus, in 1871, "And circulars were addressed to the editors of newspapers, explaining the objects and uses of the Census, in order that, if they thought proper, they might impart information on the subject to their readers, and thus aid in securing complete and correct returns."<sup>2</sup> By 1901, other means were also being used, such as holding special lessons in schools on the importance of the census.

The Act authorizing the Census made compliance by households a legal obligation. The penalty for failing to comply was up to five pounds over all of the census years, a sum that would have been substantial for most families, even in 1911. In the event that households refused to provide information, the enumerators were instructed to,

*In case of refusal to make a written return, or to answer the questions which the Enumerator is authorized to put, he may remind persons so refusing of the penalty to which they are liable for each offense; and may show them the extracts from the Act printed with these instructions. He may also warn any person whom he suspects of giving willfully false information, of the penalty imposed for that offense. If any person should still refuse, he must take a note of the name and residence of such person, and report the refusal to the Registrar as soon as possible. (Instructions to Enumerators, 1851 Census)*

The available evidence suggests that refusal to provide information was limited. For example, the 1861 reports that,

---

<sup>2</sup>"Mode of Taking the Census", Census Report of 1871, Appendix B, p. 160.

*...not a single instance occurred in which it was found necessary to adopt proceedings under the penal clauses of the Act of Parliament. Several of the enumerators who had acted in the same capacity ten years before remarked that a more intelligent appreciation of the objects and uses of the inquiry, combined with the utmost willingness to furnish the returns, was evinced by the poorer population; a result which may be fairly ascribed to the co-operation of the educated and influential classes of society, particularly of the clergy of all denominations and of the public writers in the press. (General Report, Census of 1861, p. 3)*

### 1.3 Householder's schedules

The householder's schedules remained fairly similar over the 1851-1911 period, though there was some expansion in the material covered. This evolution can be seen in the 1851 and 1911 schedules presented in Figures 2-5 at the end of this document. Each schedule included on the front page some detailed instructions as well as an example of how to fill out the form. Particularly important for our purposes are the instruction related to filling out the occupation field. For example, in 1851, manufacturing workers are instructed that:

*In the case of WORKERS IN MINES OR MANUFACTURES, and generally in the constructive ARTS, the particular branch of work and the material, are always to be distinctly expressed if they are not implied in the names, as in Coal-miner, Brass-founder, Wool-carder, Silk-throwster. Where the trade is much sub-divided, both trade and branch are to be returned thus – "Watchmaker – Finisher;" "Printer – Compositor." A person following MORE THAN ONE DISTINCT TRADE may insert his occupations in the order of importance.*

It is clear from these instructions that, at least in this category, the intent was to capture something closer to what we would today think of as industry rather than occupation. In addition, masters, journeyman and apprentices were instructed to be identified as such, with masters listing the number of persons employed.

The schedules also include clear instruction on the treatment of women and children. For example, the 1851 instructions state:

*The occupations of women who are regularly employed from home, or at home, in any but domestic duties, to be distinctly recorded. So also children and young persons. Against the names of children above five years of age, if daily attending school, or receiving regular tuition under a master or governess at home, write "Scholar," and in the latter case add "at home".*

## 1.4 Geographies

We turn now to addressing the key issues that the researcher must grapple with when using the Census of Population data. This study uses a geographic definition that expands over time as the cities we study grew. This has the advantage of including only urban area within the geographic definition. The trade off is that choices about whether to expand the definition over time will potentially affect observed city growth. As a result, variable geographies are likely to be more attractive when one is interested in studying changes in the composition of a city while controlling for overall city growth, as is done in this study.

To construct variable geographic definitions of cities, we rely on the geographies for which occupation data are reported, since one of the main uses of the variable-geography data is in analyzing the industrial composition of cities. Table 3 describes the geographic units for which occupation data are reported for cities in each year.

Table 3: Geographic definitions for which city occupation data are reported

<b>Year</b>	<b>Geographic definition</b>	<b>Details of reported city occupation data</b>
1851	Principal Towns	Based on either municipal or parliamentary limits
1861	Principal Towns	Same as in 1851
1871	Principal Towns	Same as in 1851
1881	Urban Sanitary Districts	Only USDs with population over 50,000
1891	Urban Sanitary Districts	Only USDs with population over 50,000
1901	County or Municipal Boroughs	
1911	County or Municipal Boroughs	

There is evidence that the changing geographic categorizations had a modest impact on the data. One way to check this is to take advantage of the fact that in most of the Censuses, the census department has constructed populations in a city in the previous census year, but using the census geographies for the current year. These can be compared to the population reported in the previous census in order to observe the amount of population growth due to the change in the city's geographic boundaries.

To make this point concrete, consider an example. In 1911, the geographic boundaries of the city of Stockport were expanded to include urban areas on the city edge. Specifically, the 1911 census reports that,

*By the Stockport (Extension) Order, 1901, which came into operation on the 9th November, 1901, [after the 1901 census], Stockport County Borough was extended to include Reddish Urban District (and Civil Parish)...*

For comparison purposes, in 1911 the census department then went back to the 1901 census and constructed a population for Stockport in 1901 that included the Reddish Urban District (as well as other areas added at the same time). We can then compare the population in Stockport reported in the 1901 census (for the 1901 boundaries) to the population reported for 1901 in the 1911 census (for the 1911 boundaries), to get an idea of how much the geographic shifts influenced the population of the city.

Table 4 describes, for each of the years for which these back-casting populations were constructed, the average of the absolute value of the change in population due to the shift in city geographies in each year. We can see that the average change in city population due to shifting city boundaries is around 4-5% in any given year. This value does not jump in the 1871-1881 period, when the census shifts from reporting values for Principal Towns to reporting values for Urban Sanitary Districts, nor does it jump in the 1891-1901 period, when the shift is from Urban Sanitary Districts to Municipal or County Boroughs. This suggests that these definitional changes are not substantially affecting the results. The geographies move in discrete bursts. Thus, in any given year most cities show no geographic shifts at all, while a few cities show discrete shifts that often substantially increase the population. Once a city has experienced a substantial shift that expands the city population, there is generally no shift in the preceding decade or two.

Table 4: Average absolute percentage change in city population due to shifting geographic boundaries

<b>1861-1871</b>	<b>1871-1881</b>	<b>1881-1891</b>	<b>1891-1901</b>	<b>1901-1911</b>
6.12%	4.89%	2.96%	4.47%	4.00%

This table shows the average change in city population due to shifts in the geographic extent of cities for each of the years from 1861-1911 for 25 of the cities included in the analysis database. So, for example, the first column of the table shows the average over the absolute value of the shift in population in 1861 that would have been generated by a shift from the 1861 city boundaries to the 1871 city boundaries, given the distribution of population in 1861. The remaining columns conduct the same exercise for later years.

There are a few small naming issues in the data that one should be aware of. One is Kingston-upon-Hull, which is often listed as just Hull. Another is Portsmouth, which in the early records is often listed as Portsea Island. Another issue arises with Plymouth, which is available in the Census starting in 1851 but is not included in our analysis database. The issue here is that in 1851 and 1861, the Plymouth entries include Devonport. Starting in 1871, Devonport was not included in the Plymouth entries, but separate occupation data

for Devonport was also not listed, making it impossible to construct a consistent series for Plymouth. It is also worth noting that in the analysis data, Manchester always includes Salford, an adjoining town, since these essentially comprise a single urban area and, in some years, data are only provided for the two areas together.

## 1.5 Occupation categories

For those interested in the role played by economic forces in the development of British cities, the occupation data reported in the Census can provide a powerful tool, though one that must be handled with care. The first thing that should be understood about these data is that thinking about them as reflecting what we think of today as occupations is misleading. In fact, they are generally much closer to what we would think of as industries, with only a few categories that appear closer to a true occupation, a point made in the 1911 Census:

*...the complexity of modern industry is such that especially in dealing with manufacturing processes it is impossible for the most part to classify [occupations] by the nature of the work done by the individual. The headings covering the processes of production and manufacture will be found, therefore, to refer as a rule to the article made or material worked rather than the process carried on by the individual worker. (Instructions to Clerks Employed in Classifying Occupations, Census of 1911, p. 1)*

This feature is made explicit in the instructions printed on every householders schedule in each census year, which put emphasis on the fact that the occupation category should include the “*branch of work*” and the material worked, particularly for manufacturing workers. The instructions printed on each householder’s schedule also made it clear that occupation information should be provide for women and children, as well as the men, including those who worked from home. Of the instructions on the schedule, roughly half were devoted to the occupation column.

The instructions to the Enumerators emphasized the importance placed on the occupational categories, which were the most difficult area in which to obtain accurate results. The 1851 instructions state,

*He [the Enumerator] should pay particular attention to the column headed “Rank, Profession, or Occupation,” and take care that what is inserted under that head is in strict conformity with the instructions applicable to that column.*

To emphasize this point, and provide further details, two additional pages of instructions and examples were attached to the Enumerator’s instructions providing further guidance on the completion of the occupational category.

In 1911 the household schedule changed slightly, with the inclusion of both an “Occupation” and an “Industry” entry. The motivation for this change appears to be to encourage more explicitly the households to list the industry in which they were employed. These entries were used together in constructing the occupational tabulations in the published census abstracts. No separate tabulations of industry and occupation were published until 1921.

One issue faced in the data collection phase is that many people may hold more than one occupation. The approach taken by the Census in dealing with this issue was to leave the judgment up to the workers themselves. Thus,

The enumerators were instructed to this effect, that “a person following more than one distinct trade may insert his occupations in the *order of their importance;*” and in the classification the first occupation was generally taken. (Census of 1861, General Report, p. 30)

Other than a limited listing of the second occupations of farmers, the census thus does not provide information on the secondary employment of workers. In studying the 1881 Census returns, Woollard (1999) suggests that, “Most multiple occupations, however, cause no real problem as the occupations are similar or would be classified in the same general order. (The exceptions seem usually to concern publicans of any description, shoemakers, grocers, and drapers.)”

Having collected the household schedules, the Census office was faced with a bewildering array of occupation names. In order to deal with this, the Census office produced dictionaries covering the names and instructing the clerks how to classify each. These dictionaries were updated every few decades to account for changes in nomenclature and the emergence of new industries. For example, in 1881, the dictionary used in the preceding decades (which contained 7,000 occupational names) was deemed to be out of date. To build a new dictionary, the Census office,

*...sent out circulars to leading manufacturers, asking for information as to the designations used in their branches of industry, and the information thus collected we supplemented by searches through trade directories, and especially by a preliminary examination of the enumeration books from the chief industrial centers. By these means we eventually collected together between eleven and twelve thousand different occupations having each its name.*

Using these dictionaries, the clerks of the central Census office then aggregated the occupations under a set number of headings. Table 5 shows the number of headings used for the town-level data in each of the Census years (even more detailed headings were sometimes used for the national data or for London). One thing to notice in Table 5 is that the categories are fluctuating over time. A consequence of this is that obtaining consistent

series will require matching different categories over time and then aggregating to a smaller number of distinct occupational groups. Table 5 also highlights the fact that how the Census reported occupations changed in 1901 to include substantially less detail.

Table 5: Count of reported occupational categories in town data for each census year

<b>1851</b>	<b>1861</b>	<b>1871</b>	<b>1881</b>	<b>1891</b>	<b>1901</b>	<b>1911</b>
375	479	436	402	349	157	264

To deal with these changing occupational categories we collapse the occupations into 26 consistent private-sector industry categories. No consistent crosswalk is available for matching industry categories over the 1851-1911 period, so it fell to us to construct one. In many cases this was straightforward as the categories were unchanged across years. Where categories did change, our matching was based on our understanding of the occupation. We also used the information provided by the census that classified each occupation into particular orders and sub-orders as well as the work of Lee (1979). Occupations that fell into the same order or sub-order were likely to be related. Starting in 1891, the census began publishing crosswalks relating each occupation in that year to the category in the previous census, and these are used for the years after 1871. Below, we describe the subcategories included in each of the 26 industries in the British City-Industry database.

There are a number of additional issues to consider when dealing with the occupational data. The first is that, prior to 1881, some retired persons were known to have listed their former occupation in the occupation category and were therefor tabulated under their former occupation. The Census of 1881 (p. 28) describes the issue,

Figure 1: Subcategories included in each of the 26 industry categories

**Chemicals and Drugs**

Chemical manufacture  
 Drugs and Medical Instruments  
 Dye, paint manufacture

**Instrument and jewelry makers**

Gold and silver products  
 Instrument manufacture  
 Jewelry  
 Musical instrument maker  
 Watch makers

**Metal and machinery makers**

Anchor and chain manufacture  
 Arms manufacture Other than guns  
 Blacksmiths and whitesmiths  
 Brass working  
 Button maker  
 Copper working  
 Cutler  
 Engine and machine makers  
 File maker  
 Gun manufacture  
 Iron manufacture  
 Ironmonger  
 Lead working  
 Millwright  
 Nail manufacture  
 Other implements  
 Other metalworkers  
 Patternmakers, draughtsman, etc.  
 Pin, needle makers  
 Tin working  
 Tool makers  
 Wheelwright  
 Wire maker  
 Zinc working

**Oil, soap, and rubber products**

Oil colourman  
 Soap maker  
 Rubber and oil products  
 Other bone, etc. products  
 Tallow, candles

**Dress**

Glover  
 Haberdasher  
 Hatter (Hat Maker)  
 Hosiery Manufacture  
 Milliner  
 Rag gatherer  
 Seamstress, shirtmaker  
 Shoe Maker  
 Straw hat, plait manufacture  
 Tailor  
 Umbrella, etc. manufacture  
 Other misc. dress workers

**Earthenware and bricks**

Brickmaking  
 Earthenware dealers  
 Earthenware manufacture  
 Glass manufacture  
 Other misc. earthenware products

**Leather and hair goods**

Saddle makers  
 Leather goods  
 Hair goods  
 Tanner, fellmonger  
 Quill goods  
 Fur workers

**Textiles**

Carpet and rug manufacture  
 Cotton textile finishing  
 Cotton textile manufacture  
 Draper (wool)  
 Flax and linen  
 Fustian manufacture  
 Hemp, jute, sackcloth  
 Other cotton and flax workers  
 Other wool workers  
 Rope makers  
 Silk manufacture  
 Thread  
 Wool stapler  
 Wool textile manufacture  
 Worsted and stuff manufacture

Figure 2: Subcategories included in each of the 26 industry categories (cont.)

**Paper and publishing**

Bookbinder  
Lithographs, pictures, photographers  
Other paper products  
Paper manufacture  
Paper stainer  
Printer  
Publisher  
Stationary

**Wood products & furniture**

Carvings  
Cooper  
Misc. wood, bark products  
Misc. wood manufactures  
Sawyer  
Timber  
Wood furniture  
Woodman

**Water and gas service**

Gasworks service  
Waterworks service

**Food processing and sales**

Baker, confectioner  
Butcher  
Cheesemonger  
Corn miller  
Corn, vegetable merchant  
Greengrocer  
Grocer  
Milk seller  
Other food products  
Other meat and fish dealer  
Salt maker  
Sugar refining

**Shipbuilding**

Shipbuilding

**Vehicles**

Coachmakers (including railway)  
Cycle and motorcars

**Construction**

Bricklayer  
Builders (construction)  
Carpenter  
Mason  
Misc. construction  
Painters and plumbers  
Plasterers and paperhangers  
Slater

**Mining**

Coal merchant  
Coal mining  
Iron miner  
Other clay, stone products  
Other coal and coke workers  
Other miners  
Stone workers

**Spirituos drinks**

Maltster  
Brewer

**Tobacco**

Tobacconist

Figure 3: Subcategories included in each of the 26 industry categories (cont.)

**Professionals**

Architect  
Art, theater, music related  
Doctors  
Engineers (civil)  
Law  
Musicians  
Other medial workers  
Scientists  
Surveyors  
Writers  
Dental

**Merchants, agents, accountants, etc.**

Accountant  
Auctioneer  
Broker, agent, factor, etc.  
Merchant  
Misc. commercial occupations

**Retail shopkeepers and salesmen**

Hawker, peddler  
Pawnbroker  
Salesman  
Shopkeeper, general

**Messengers, porters, warehousing**

Messengers and porters  
Other messages (telegraph, etc.)  
Warehousing

**Railway transport**

Railway transport

**Road transport**

Carter or hauler  
Coach, bus operator  
Toll collectors  
Tramways

**Sea and canal transport**

Ship transport  
Canal transport

**General service**

College, club, lodge service  
Domestic Service  
Hairdresser  
Innkeepers and Beersellers  
Laundry  
Lodging House Keepers  
Medical institution service  
Other Boarding and Lodging  
Other drinks  
Wine merchant

*As regards persons “retired” from any business, we found ourselves in some doubt [as to how to include them]. In the Census of 1871 such persons had been considered as following the business from which they had really retired, and were abstracted accordingly. To depart from this former practice would, of course, interfere in some measure with the ready comparison of the returns for 1881 with those of 1871. But, on the other hand, it was known that a very inconsiderable proportion of persons who had retired from business made mention of their former occupation in their schedules, and that, consequently, if such persons were included, the return made by us under any occupation would be neither of persons actually so occupied, nor yet of those together with those who had retired from the trade, a large proportion of the latter being omitted. We found by careful examination of the enumeration books for an entire county, including a large town, that, had we included the “retired,” as was done in 1871, the persons returned by us under any heading would on an average have been about 2 per cent more than they are actually. On the whole, seeing that the difference was so small, we thought it best altogether to omit those who had retired from business...*

The practice of omitting retirees who listed an occupation from the occupation returns begun in 1881 was then carried on in the latter years. Later we will discuss some recommendations for how to address this issue when using the data.

A similar issue exists with workers who were unemployed. In general, those who were temporarily unemployed were included under their usual occupation. It goes without saying that the Census data are not a useful tool for considering temporary unemployment patterns. In a similar vein, those in asylums, jails, and hospitals were commonly classed under their previous occupation prior to 1881. Lee (1979) conducted an analysis of the share of the population listed as employed at various years and concluded that the main break in practice was between 1871 and 1881, with a drop of about 5% in the activity rate. However, it is unclear how much of this might have been due to cyclical economic factors.

A second issue of concern is the treatment of makers vs. dealers. This is a tricky issue because, while in some cases firms specialize as either makers or dealers, others, particularly small firms, undertake both activities at the same time. Prior to 1901, both makers and dealers working in the same industry were often lumped together, though in cases where many specialized dealers existed we see individual dealer categories. For example, the 1851 census separately identifies “Coal, Merchant or Dealer,” but also combined some makers and dealers in smaller categories, such as “Other Workers, Dealers in Hemp.” Starting in 1901, the Census made an effort to separate workers and dealers, but with limited success. In general, it seem to us that the best course of action is simply to include workers and dealers in the same industry together in a single consistent category.

A third issue in using the occupational data reported in the published Census Abstracts is that the age categories reported change over time. In 1851 and 1861, detailed occupation data for principal towns are reported in two age categories, “Under 20 years” and “20 years and up.” In 1871, only the “20 years and up” category appears for the town-level data. In 1881 and 1891, occupation is reported only for all ages together, and in 1901-1911 it is reported only for those over 10, which amounts to essentially reporting all workers.

The key issue in constructing a consistent series is therefore 1871, where a substantial number of workers under 20 would not be included in the tabulations. Note that a consistent series of occupation counts for workers over 20 can be constructed for 1851-1871 without making any adjustment, and when running analysis over just those years that may be the preferred approach. To construct a consistent database over a longer period requires that we make an adjustment to the 1871 population counts. To do so, we take advantage of the fact that occupations counts by age are available for 1871 at the national level. Under the assumption that the fraction of workers under 20 in a particular industry is fairly constant across locations, we can use the national share of workers under 20 to adjust the city occupation data to reflect all workers.

Apart from these general issues, there are some more specific issues that apply to particular occupational categories. The agricultural occupations are particularly fraught because the ambiguity involved in deciding at what point a child or wife working in a farm home was listed as an agricultural worker. As a general rule we exclude agricultural occupations from our data.

## 1.6 Exploring the British City-Industry data

The British city-industry data series were constructed using the tabulations from the Census Abstracts published by the Census Office for each census year. Scanned version of these published volumes have been made available by the U.K. Data Archive at the University of Essex through the [histpop.org](http://histpop.org) website. The data, coming from hundreds of pages of scanned PDFs, were then digitized by hand using a double-entry procedure to reduce error rates.

Figure 4 in the appendix provides time-series plots of the national employment patterns for the 26 analysis industries across the 31 cities for which the data are available. There are a number of patterns visible in these graphs that suggest that these occupational matches are allowing us to construct consistent and reliable indicators of the evolution of industry employment over time. In general the plots show reasonably smooth growth paths over time. In addition, the pattern of volatility also seems reasonable; capital goods industries such as Construction, Earthenware & Bricks, and Shipbuilding, show the greatest volatility over time, while industries such as General Services and Professionals show extremely smooth series.

The variation in growth rates across the different industries fits the historical record well. Established industries such as Textiles, Leather & Hair Goods, and Clothing, Shoes, etc., show consistent but slow growth. In contrast, Second Industrial Revolution industries, such as Chemicals & Drugs, Instruments, Oil & Soap, and Vehicles, show clear accelerations in the later years of the series. The Metal & Machinery industry, which includes a mix of first and second Industrial Revolution industries, falls somewhere in between.

Another comforting pattern is that series that we would expect to move together often do so. One example is the construction-related industries – Construction, Earthenware & Bricks, and Wood & Furniture – all of which experience slow growth between 1881-1891, rapid growth from 1891-1901, and slow growth again from 1901-1911. Another pair of related industries that moves together is Shipbuilding and Sea & Canal Transport, both of which grow rapidly from 1851-1861, decelerate from 1861-1871, and then resume modest growth.

## **2 Firm size data from the 1851 Census**

The 1851 census was particularly detailed, both in the amount of information collected and the detail of information reported. One piece of information included in this year is the distribution of workers across different firm sizes. These data were collected from the household schedules, where individuals listing themselves as “Masters” (i.e., business owners) were instructed to list the number of workers that they employ. Unfortunately, many masters did not specify the number of men employed and it is difficult to tell apart those who worked for themselves without employees from those who simply failed to state the number of workers they employed. As a result, the returns are incomplete. Also, since failing to state the number of workers may correlated with the size of the firm, it is not clear that working with only the entries where the number of workers were reported can allow us to accurately observe the firm size distribution for firms with at least one employee.

Despite these issues, the firm size data can still be useful for comparing across industries. If we are willing to assume that, conditional on firm size, the propensity to report did not vary substantially across industries, then the firm size data can be used to differentiate industries characterized by larger or smaller firms. The firm size data are reported at the county level. This geographic unit is too large to directly link to cities, particularly in counties with many cities such as Lancashire or Yorkshire. In general, we treat the firm size data as a national characteristic for comparing across industries, rather than as a regional characteristic.

### 3 The technology similarity matrix

Here we briefly describe how the technology similarity matrix was constructed. The technology similarity matrix is based on patent data, unlike the data we have discussed previously. Patent data, while not perfect, is widely used as a means for tracking technological progress. It has also been used to track flows of information across firms and industries (e.g., Jaffe *et al.* (1993)).

There are two building blocks needed to generate the technological similarity matrix. The first is a mapping from patents to industries. This comes from work by Nuvolari & Tartari (2011) in which they hand matched each British patent before 1841 to one industry category. The Nuvolari & Tartari (NT) industry categories are then matched to categories from the British city-industry database. In most cases it is possible to generate a direct match between the NT industries and the industries in the British city-industry database, but there are a couple of exceptions. The exceptions are:

- The NT industry “Food” contains patents related to both the “Food” and the “Spirituuous drinks” industries from the city-industry database
- The NT industry “Chemicals” contains patents related to both the “Chemicals & drugs” and the “Oils, soaps etc.” industries in the city-industry database
- The NT industry “Construction” contains is primarily related to the “Construction” industry in the city-industry database, but contains a few patents that should be in the “Earthenware, bricks, etc.” industry

For these exceptions, I conducted a second review of the patents using the patent titles, in order to classify the patents into the proper industry category from the British city-industry database. The resulting matches are described in Table 7 below.

Once each patent has been matched to one industry, the second necessary ingredient in constructing the technological similarity matrix is a mapping of connections between each patent. This is done using the names of the inventors attached to each patent. These are available from the Cradle of Inventions Database, which has been used in previous studies such as Brunt *et al.* (2012) and Hanlon (2012). As a first step, the inventors names are cleaned and standardized. Next, a set of known patent agents are dropped from the database. This helps deal with the fact that occasionally patent agents may be listed as the inventor when they file a patent for the true inventor, particularly if the true inventor is foreign. For each pair of industries, we then calculate the number of inventors that file at least one patent in each. The technological similarity matrix between any two industries  $i$  and  $j$  is then,

$$TECHSIM_{ij} = \frac{CommonInventors_{ij}}{Inventors_i + Inventors_j}.$$

The matrix generated by this exercise is shown in Table 8. Looking closely at this matrix, we observe patterns that tend to make sense. For example, consider the first column, Agriculture. This industry tends to have relatively little relatedness to other industries, and the most related industry to Agriculture is Food, which seems reasonable. Another industry with little relationship to other sectors is Mining, which shows a substantial level of relatedness only to Construction. The industries showing the broadly high levels of relatedness are Chemicals & Drugs and Metal & Manufactures. Thus, our results indicate that these are the most central industries in technological space.

Consider a couple of other examples. Shipbuilding shows moderate levels of technological similarity to other industries, with the strongest relationships being to Metals & Manufactures and Vehicles. Shipbuilding shows very low levels technological similarity to Earthenware/bricks, followed by Agriculture and Dress.

Textiles, another important industry in Britain during this period, shows strong technological similarity to Metals & Manufactures, followed by Paper and Chemicals & Drugs. The latter two make a lot of sense, given that chemicals such as bleaches and dyes are an important part of textile finishing, as well as Paper-making.

Finally, consider the Wood & Furniture industry. This industry is most closely related to Construction, followed by Metals & Manufactures, Shipbuilding, Paper, and Chemicals. This industry has very little technological similarity to Mining.

As these examples illustrate, it appears that our approach is delivering a measure of technological similarity that seems reasonable given our knowledge of these industries.

Table 6: Industry firm size data from 1851 Census of Population

Industry	Median			Median		
	Median firm size	worker's firm size	Average firm size	Median firm size*	worker's firm size*	Average firm size*
Chemicals and drugs	1	7	5	2	10	6
Services	1	1	2	1	2	3
Dress	1	5	4	2	6	6
Merchants, agents, accountants	2	9	6	3	10	7
Shopkeepers and salesmen	1	3	3	3	4	5
Road transport	1	3	3	3	5	5
Sea and canal transport	3	20	8	5	20	10
Engineers and surveyors	7	20	11	7	20	11
Vehicles	4	10	7	5	10	9
Shipbuilding	3	40	11	5	40	14
Building	2	7	5	3	8	7
Food processing	1	2	3	2	3	4
Oils, soaps, etc	1	5	4	2	6	5
Leather, hair, etc	2	8	5	3	9	7
Drinks	2	8	6	3	9	7
Tobacco	1	20	7	5	20	11
Wood and furniture	2	6	5	3	7	6
Textiles	3	150	17	5	150	22
Paper, etc.	2	20	8	4	20	11
Mining	3	50	13	4	50	15
Earthenware, bricks, etc	4	100	16	6	100	19
Instruments and jewelry	1	7	5	3	9	7
Metal and machinery manuf.	1	20	6	2	20	8

\* Values in columns 3-6 are calculated dropping entries by masters reporting zero employees or not reporting the number of employees.

Table 7: Industry category matches and patent counts, 1821-1841

<b>British city-industry database industry</b>	<b>Nuvolari-Tartari industry</b>	<b>Patent count</b>	<b>British city-industry database industry</b>	<b>Nuvolari-Tartari industry</b>	<b>Patent count</b>
Agriculture	Agriculture	108	Metals & manufactures	Metallurgy	259
Chemicals/Drugs	Chemicals	292	"	Manufacturing	200
"	Medicines	61	"	Hardware	277
Clothing	Clothing	81	"	Engines	662
Construction	Construction	191	Oils, soap etc.	Oils, soap*	62
Earthenware/bricks & glass	Pottery	75	Paper/publishing	Paper	204
"	Glass	32	Shipbuilding	Ships	280
Food	Bricks*	10	Spirituous drinks	Drinks*	63
Instruments	Food	181	Textiles	Textiles	711
Leather/hair goods	Instruments	155	Vehicles	Carriages	273
Mining	Leather	63	Wood/furniture	Furniture	224
	Mining	27	Water/gas service	Water, gas*	71

\* indicates categories that were not in the original NT database. Patents in the "Military" industry in the NT database are not used in constructing the technological similarity matrix.

Table 8: Industry technological similarity matrix

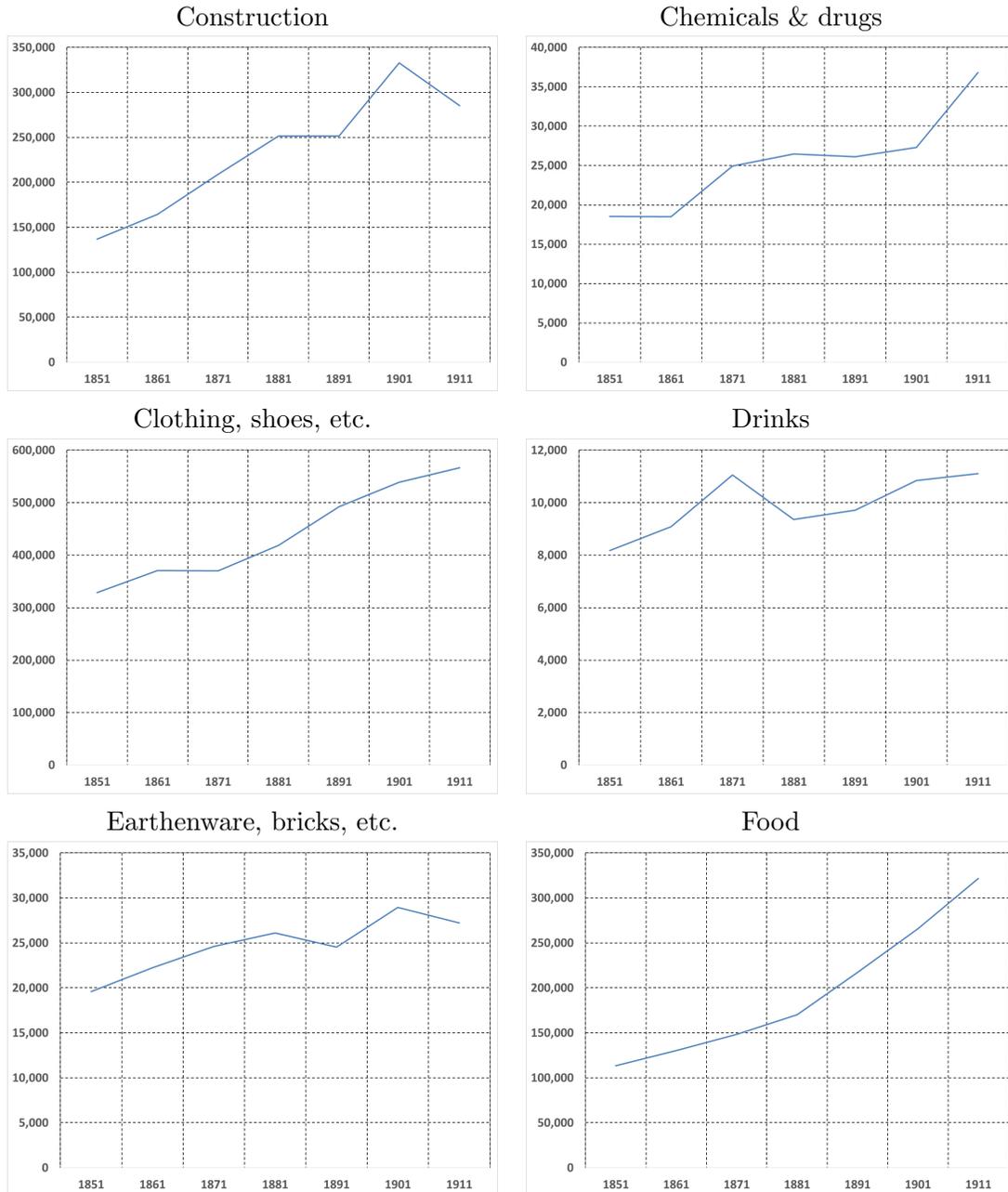
	Agriculture		Construction	Chemicals/ drugs	Dress	Earthenware/ bricks	Food	Instruments/ jewelry	Leather goods	Metal/ manufactures	Mining	Oil, soap etc.	Paper, etc	Shipbuilding	Spiritous drinks	Textiles	Vehicles	Water & gas service	Wood/ furniture		
Agriculture		0.008	0.003	0.006	0.016	0.021	0.021	0.022	0.013	0.011	0.016	0.006	0.008	0.010	0.000	0.007	0.000	0.016	0.009	0.013	0.011
Construction	0.008		0.019	0.008	0.019	0.018	0.016	0.024	0.018	0.037	0.021	0.000	0.003	0.028	0.005	0.015	0.020	0.018	0.020	0.018	0.029
Chemicals/ drugs	0.003	0.019		0.011	0.018	0.034	0.034	0.012	0.008	0.041	0.000	0.020	0.022	0.029	0.008	0.023	0.023	0.020	0.020	0.023	0.023
Dress	0.006	0.008	0.011		0.006	0.005	0.005	0.005	0.024	0.009	0.011	0.008	0.000	0.010	0.000	0.022	0.007	0.000	0.004	0.004	0.004
Earthenware/ bricks	0.016	0.019	0.018	0.006		0.021	0.021	0.022	0.013	0.011	0.016	0.006	0.008	0.006	0.007	0.010	0.009	0.013	0.013	0.011	0.011
Food	0.021	0.016	0.034	0.005	0.021			0.015	0.010	0.023	0.006	0.020	0.016	0.027	0.042	0.016	0.016	0.020	0.016	0.018	0.018
Instruments/ jewelry	0.005	0.024	0.012	0.005	0.022	0.015	0.015		0.005	0.021	0.007	0.011	0.007	0.017	0.000	0.005	0.028	0.016	0.020	0.006	0.006
Leather goods	0.000	0.018	0.008	0.024	0.013	0.010	0.010	0.005		0.007	0.000	0.019	0.005	0.014	0.000	0.005	0.003	0.000	0.017	0.017	0.017
Metal/ manufactures	0.013	0.037	0.041	0.009	0.011	0.023	0.023	0.021	0.007		0.009	0.008	0.025	0.048	0.011	0.041	0.040	0.014	0.027	0.027	0.027
Mining	0.000	0.021	0.000	0.011	0.016	0.006	0.006	0.007	0.000	0.009		0.000	0.000	0.016	0.000	0.004	0.000	0.000	0.000	0.000	0.000
Oil, soap etc.	0.007	0.000	0.020	0.008	0.006	0.020	0.020	0.011	0.019	0.008	0.000		0.009	0.011	0.010	0.005	0.003	0.000	0.004	0.004	0.004
Paper, etc	0.016	0.003	0.022	0.000	0.008	0.016	0.016	0.007	0.005	0.025	0.000	0.009		0.023	0.000	0.030	0.023	0.009	0.023	0.023	0.023
Shipbuilding	0.009	0.028	0.029	0.010	0.006	0.027	0.027	0.017	0.014	0.048	0.016	0.011	0.023		0.011	0.014	0.035	0.018	0.025	0.025	0.025
Spiritous drinks	0.007	0.005	0.008	0.000	0.007	0.042	0.042	0.000	0.000	0.011	0.000	0.010	0.000	0.011		0.004	0.003	0.000	0.009	0.009	0.009
Textiles	0.012	0.015	0.023	0.022	0.010	0.016	0.016	0.005	0.005	0.041	0.004	0.005	0.030	0.014	0.004		0.017	0.007	0.018	0.018	0.018
Vehicles	0.006	0.020	0.023	0.007	0.009	0.016	0.016	0.028	0.003	0.040	0.000	0.003	0.023	0.035	0.003	0.017		0.021	0.019	0.019	0.019
Water & gas service	0.007	0.018	0.020	0.000	0.013	0.020	0.020	0.016	0.000	0.014	0.000	0.000	0.009	0.018	0.000	0.007	0.021		0.004	0.004	0.004
Wood/ furniture	0.004	0.029	0.023	0.004	0.011	0.018	0.018	0.006	0.017	0.027	0.000	0.004	0.023	0.025	0.009	0.018	0.019	0.004	0.004	0.004	0.004

## References

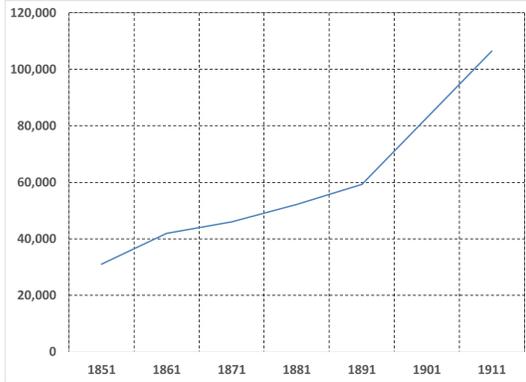
- Brunt, Liam, Lerner, Josh, & Nicholas, Tom. 2012. Inducement Prizes and Innovation. *Journal of Industrial Economics*, **60**(4), 657–696.
- Hanlon, W. Walker. 2012 (May). *Necessity is the Mother of Invention: Input Supplies and Directed Technical Change*. Working Paper.
- Higgs, Edward. 1989. *Making Sense of the Census*. London: Her Majesty's Stationary Office.
- Jaffe, Adam B., Trajtenberg, Manuel, & Henderson, Rebecca. 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics*, **108**(3), 577–598.
- Lee, C.H. 1979. *British Regional Employment Statistics 1841-1971*. Cambridge, UK: Cambridge University Press.
- Nuvolari, Alessandro, & Tartari, Valentina. 2011. Bennet Woodcroft and the Value of English Patents, 1617-1841. *Explorations in Economic History*, **48**, 97–115.
- Woollard, Matthew. 1999. *The Classification of Occupations in the 1881 Census of England and Wales*. Mimeo.

## 4 Additional material

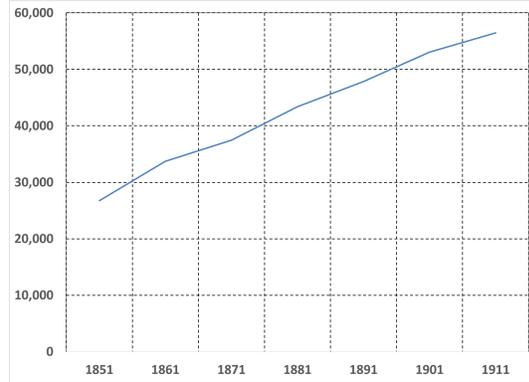
Figure 4: Employment by industry group across all analysis cities, 1851-1911



Instruments & jewelry



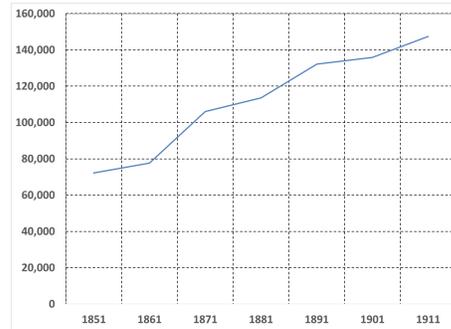
Leather goods, etc.



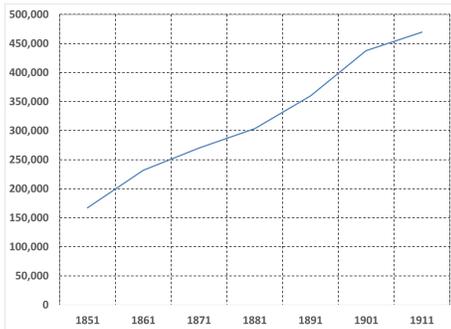
Merchants, agents, etc.



Messengers, porters, communication



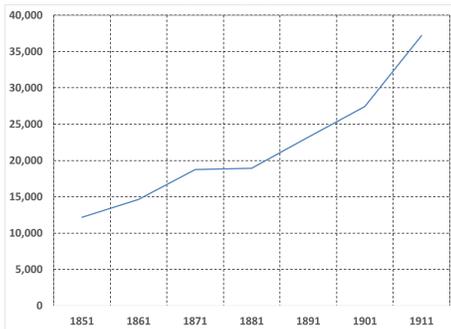
Metal & Machinery



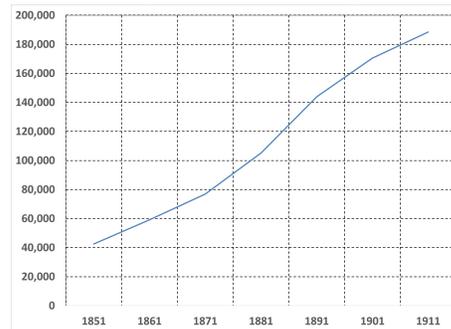
Mining related



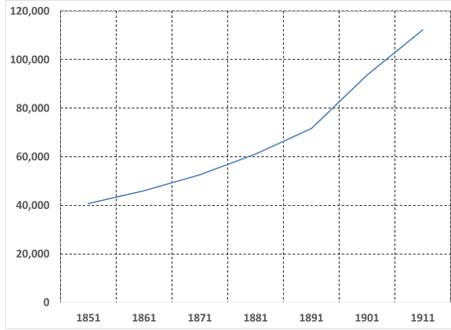
Oils, soaps, etc.



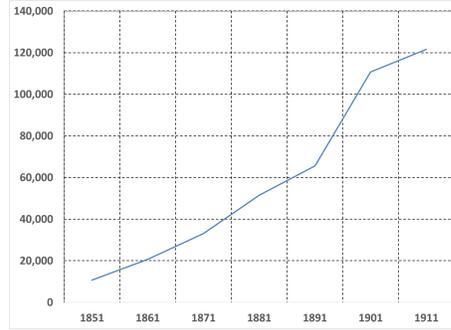
Paper & Publishing



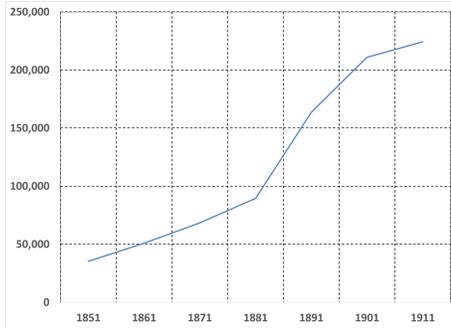
Professionals



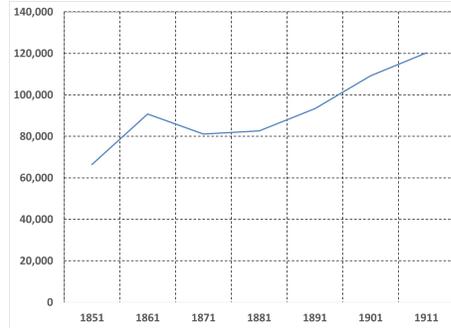
Railway transport



Road transport



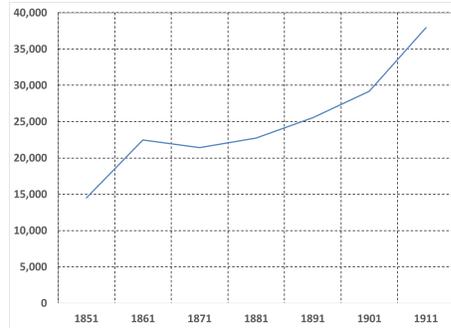
Sea & Canal Transport



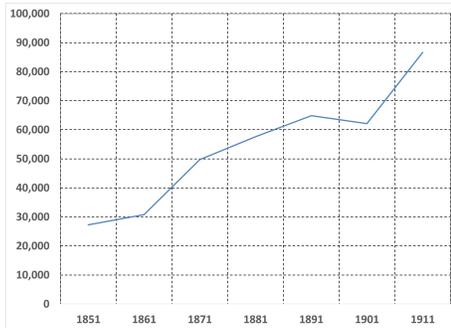
General services



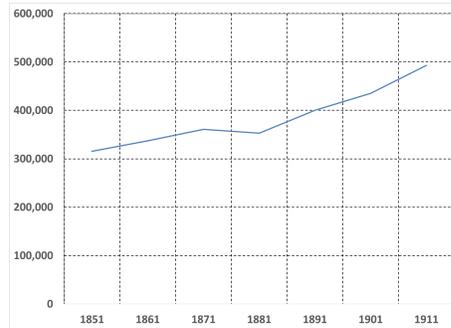
Shipbuilding



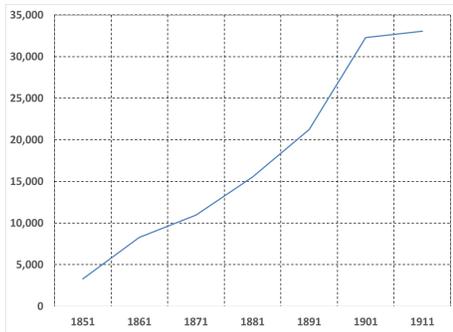
Shopkeepers



Textiles



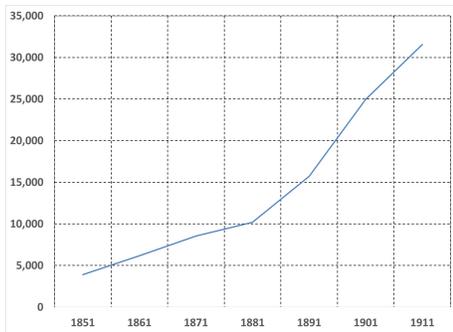
Tobacco



Vehicles



Water & Gas Utilities



Wood & Furniture

