A Different History Manifesto

Clare Crowston, Claire Lemercier

A Different History Manifesto: Data Need not Be "Big" or "Clean"

Clare H. Crowston, University of Illinois at Urbana-Champaign
Claire Lemercier, CNRS, Center for the Sociology of Organizations (Sciences Po, Paris)

This text draws on two ongoing conversations. The first conversation, centered on pedagogies in quantitative history (and in particular on good practices of data-set building), recently gave birth to a short introduction to "*Quantitative Methods in the Humanities*," published by Claire Lemercier and Claire Zalc with the University of Virginia Press. The second is a conversation among the three co-authors of an in-progress book manuscript on the history of apprenticeship in 18th and 19th-century France, (written by the two of us and Steven L. Kaplan). This is a conversation about the history of work, in which gender, occupations, and skill figure prominently. It is also a profoundly methodological conversation, due to the fact that we are writing a history that spans more than two centuries and attempts a national scale and the predominantly economic and quantitative approach that has been taken to apprenticeship in recent years.[1] The scale of our project and the need for comparison across time and space call for macro-level analysis; however, our commitment to reconstituting the complex contexts in which apprenticeship was conceived and practiced –including nuclear and extended families, the church, schools, law courts, etc. – require micro-historical approaches. (These dilemmas of a *longue durée* study are similar to methodological problems raised in the context of "global" history).

In the paper that follows, therefore, we attempt to provide general statements about data and method (which may be familiar to our audience at SSHA but perhaps more surprising to the wider historical community) as they are provoked and illustrated by our ongoing research. In so doing, our aim is to show that these principles are not just abstract compromises to the classical dilemma of "quantitative vs. qualitative" history: they offer added value to research on our specific substantive topic. Coming back to the questions listed in the call for papers on "Data and its Discontents", we could sum up our principles as follows:

Data exists; it is as simple as notes taken from historical sources. We tend to call it "data" when we take our notes in a spreadsheet, but it is really any type of notes taken *systematically* from historical sources. Our spreadsheets are verbose and most "digital humanists" would want to "clean" them, but we cherish their dirtiness. It is a sign that the transformation from source to data (notes) happened through our own thought processes (not those of underlings paid to sweep the dirt under the rug) and retained the ambiguities of the sources. To make data better, and to have more colleagues crave it and fewer abhor it, we want to keep it as close as possible to the source, even though it will be dirtier, and more costly to produce in large quantities. But that's fine, because we do not believe data is good only if it is really big. It is good if it is complicated,

[1] See for example, Stephan R. Epstein, "Craft guilds, apprenticeship and technological change in preindustrial Europe," *Journal of Economic History* 583 (1998): 684-702; Patrick Wallis, "Apprenticeship and Training in Premodern England", *Journal of Economic History*, 68, 3, (2008): 832-861; Chris Minns and Patrick Wallis, "Rules and Reality: Quantifying the Practice of Apprenticeship in Early Modern England," *The Economic History Review* 65, 2 (May 2012): 556-579; and David de la Croix, Matthias Doepke, and Joel Mokyr, "Clans, Guilds, and Markets: Apprenticeship Institutions and Growth in the Preindustrial Economy," *The Quarterly Journal of Economics*, 133, 1 (February 2018): 1-70. For a recent study attempting to bridge socio-cultural and economic approaches, see Laura Gowing, "Girls on Forms: Apprenticing Young Women in Seventeenth-Century London," *Journal of British Studies*, 55, 3, (2016): 447-473.

and thus rich in information, but still systematically acquired and noted in a structured way, so that we can simplify it in many different ways if we want to experiment with it. It is good for thinking, even, or especially, when it produces new questions rather than final answers.

This was the manifesto in our title.[2] To flesh out what we mean, this paper will illustrate three main points through cases from our research on apprenticeship in France[3]. First, data can help us say something about the meanings and values of the past – for example on how work performed by different categories of people was valued differently than that of other categories - not just about economic or demographic behaviors. Second, trying to be systematic and to count, if we don't vacuum-clean missing data, can help us notice silences in the sources and thus address important questions, for example on female labor. Third, experimenting with alternative ways to categorize is the only way to learn something new: we strongly oppose attempts to offer tools to "harmonize," "standardize," or "categorize" data that can then putatively be used to respond to any research question. This last point has been at the center of our thinking about who was considered to be an apprentice and how this categorization related to the perceived hierarchy of trades.

I. Meanings from Data: The Value of Female Apprentices

We do not want our book to employ quantitative methods for the economic aspects of apprenticeship and qualitative methods for its cultural aspects – especially as, in practice, sources and situations generally mix the economic and the cultural. Like other historians, notably those working in the methods of Italian micro-history[4], we strive to read cultural meaning and value through quantitative data we have created from our reading of historical sources.

To take one example, we studied evidence on early termination of apprenticeship in the eighteenth and nineteenth centuries. One of the most common types of disputes involving apprenticeship had to do with apprentices leaving before the end of the contract (based on evidence from litigation, this was the most frequent complaint brought to the courts in mid-eighteenth-century Paris and Lyon and in mid-19th century Paris). From the nineteenth century onward, economists, and commentators generally, claimed that apprentices would be most likely to leave after roughly two thirds of the time, at a moment when they would have learned enough to be able to be employed as waged workers[5]. Since masters and mistresses invested time and resources at the beginning of the contract and would not be remunerated by the apprentices'

[2] The title alluded to David Armitage et Jo Guldi, *The History Manifesto*, Cambridge, Cambridge University Press, 2014, which, in our view, only revived the "quantitative vs. qualitative history" divide and the lack of attention of some first-generation quantitative historians to the construction of data.

[3] We give more details on most of these cases in Clare H. Crowston, Steven L. Kaplan, and Claire Lemercier, (2018), « Les apprentissages parisiens aux XVIIIe et XIXe siècles », *Annales HSS*, vol. 73, n° 4, published in a special issue about the renewals in quantitative history (still forthcoming, in spite of the 2018 date, and to be published in English in 2020).

[4] One of our main references here is the use of prices (in the context of sales of land) to reveal values and social norms in Giovanni Levi, *Inheriting power: the story of an exorcist*, Chicago, University of Chicago Press, 1988.

[5] Patrick Wallis, "Apprenticeship and Training in Premodern England", *Journal of Economic History*, 68-3, 2008, p. 832-861 discussed this in the context of twentieth-century economics, but this "two-stage model" of apprenticeship was already ubiquitous in texts by nineteenth-century commentators, for example Édouard Ducpétiaux, *De la condition physique et morale des jeunes ouvriers et des moyens de l'améliorer. Tome 2*, Bruxelles, Méline, Cans et cie, 1843, p. 400.

unpaid labor at the end, courts should redress this loss by granting damages. However, our results show that disputed departures did not occur at the moment identified by economists as the most propitious for an apprentice to depart. This is equally true of mid-19th century Paris, based on our study of labor court decisions, as it was of Old Regime Lyon – where detailed guild registers show that about 20% of apprentices from the late 1680s through the 1760s left early, but with no clear clustering of departures at any one point of the contract[6].

If apprentices did not choose to leave at the supposedly most economically profitable moment, nor did their masters and mistresses seem to ask for higher damages if departures happened when two thirds or three quarters of the duration had passed. And judges equally perversely failed to follow the economists' presumptions in their decisions. (This analysis is based on a regression involving judgments from several hundred labor court cases in mid-19th century Paris that tried to disentangle the correlation between timing and damages from other correlations) So we have learned a little something, however indirect, about visions of the phases of the contract by actors who have not produced manifestoes, diaries, memoirs, or other written sources. The little something is interesting because it contradicts what almost everybody who had access to print wrote and what economic historians have long maintained about apprenticeship.

But what is perhaps more important is that, for our mid-19th century sample, the best predictor of damages granted is damages requested. And damages requested varied a great deal, all other things being equal, depending on the gender of the apprentice (not, or not significantly, according to the gender of the master or mistress). Our analysis shows that the mistress of a girl would ask for 95 francs less than the master of a boy, in a context where the median demand was 200 francs. At first sight, this result provides useful but hardly earth-shattering confirmation of the well-known fact that female labor was valued less than male labor. Upon reflection, however, it raises interesting questions in the specific context of damages for an uncompleted apprenticeship. In particular this finding obliges us to consider the range of possible reasons why masters and mistresses asked for lower damages in these cases. For example, the loss of unpaid work might be less valued, because girls would receive lower wages anyway (this would be an economic calculation by the master or mistress); or the master or mistress might know that the court would not grant as much as for a boy, whatever the demand (especially as the judges were all male), and decide on a more "reasonable" claim. A third possibility is that the cost of the first years of apprenticeship was reckoned to be lower for girls, because it was more legitimate to ask them to perform "services" for the master or mistress' household. Thus, even lacking occupational skills, a girl could contribute valuable labor to the household through gendered domestic activities, such as cleaning, cooking and childcare. This indeed is what our qualitative analysis of a group of detailed cases involving complaints about domestic chores demanded of female apprentices suggests. Thus, our regression makes it clear that girls' labor was "valued" less than boys' labor and demands of us further thought and reflection on how and why this was the case.

II. Results from missing data

The issue of the "value" of apprentices' labor was a case where we had comparable, systematic data on girls and boys, because cases involving apprentices of the two genders were discussed in courts. But of course, the history of female labor is often a story of missing data, and

[6] Ruben Schalk, Patrick Wallis, Clare Crowston and Claire Lemercier, "Failure or flexibility? Exits from apprenticeship training in pre-modern Europe", *Journal of Interdisciplinary History*, 48-2, 2017, p. 131-158.

our research is no exception[7]. We surmise, however, that the systematic gathering and quantitative analysis of data can be used to advance source criticism and make better sense of missing data. What was the share of girls among apprentices in Paris and what did they learn? Data gathered by other people offered us seemingly solid answers. The French national archives have created a dataset of metadata on all notarized contracts for 1761[8]. At that time, the notion of "apprenticeship" was supposed to apply only to contracts established in the context of a guild, and guilds mandated that the contracts be notarized. So, if we defined the category on the basis of these legal norms, we found 1,404 new apprentices in 1761; 11% were girls. One century later, the Paris Chamber of Commerce surveyed all industrial and artisanal firms: clerks went door to door to fill in forms and the Chamber published a one-thousand page book that was among the first to use cross-tabulations; in many ways, this enterprise resembles twentieth-century official surveys.[9] Entrepreneurs were asked about their apprentices; the category was not defined in the questionnaire but seemed taken for granted. The Chamber found 19,700 apprentices in 1860; 28% were girls.

Should we conclude that the abolition of the guild system, which had excluded most women from guild membership, had rendered apprenticeship more accessible for girls? As historians of female labor, we knew from experience that a little bit of source criticism was necessary before jumping to conclusions. The more general point here might seem obvious, but is still sometimes ignored by students and even colleagues: source criticism also applies to "data," especially when it was gathered by other people, and especially when it looks very clean. It happens that seamstresses, the largest female guild in Paris, mostly used one notary for their apprenticeship contracts; there were so many contracts that he stored them separately, and those contracts have apparently been lost for 1761. Based on my previous research on seamstresses,[10] I was able to estimate a number of new contracts for that year. Including seamstresses increases the share of girls among new apprentices from ca. 11% to ca. 30%.

Did, then, the share of girls slightly *decrease* from 1761 to 1860, and should we blame the abolition of female guilds? Not exactly. We already knew from a pioneering chapter by Joan Scott that the survey by the Chamber of Commerce should not be taken at face value, especially as regards female labor.[11] But Scott had used external and internal criticism of the *text* of the

---

[7] On research by "cunning historians" who have written the history of female labor from often missing data, see e.g. the recent survey by Raffaella Sarti, Anna Bellavitis and Manuela Martini, "Introduction", *in* Raffaella Sarti, Anna Bellavitis and Manuela Martini (eds.), *What is Work? Gender at the Crossroads of Home, Family, and Business from the Early Modern Era to the Present*, New York, Berghahn Books, 2018, p. 1-84.

[8] It was one of the so-called "bases ARNO" and has now become part of the general online finding aid of the National Archives (https://www.siv.archives-nationales.culture.gouv.fr), but thanks to Gilles Postel-Vinay, we have been able to use a tabular version.

[9] *Statistique de l'industrie à Paris résultant de l'enquête faite par la Chambre de commerce pour l'année 1860*, Paris, Chambre de commerce, 1864.

[10] Clare Haru Crowston, *Fabricating Women : The Seamstresses of Old Regime France, 1675-1791* (Duke University Press, 2001).

[11] Joan W. Scott, « Statistical Representations of Work: The Chamber of Commerce's Statistique de l'Industrie à Paris, 1847-48 », *in* Steven L. Kaplan & Cynthia J. Koepp (eds.), *Work in France: representations, meaning, organization, and practice*, Ithaca, Cornell University Press, 1986, p. 335-363, discussing *Statistique de l'industrie à Paris résultant de l'enquête faite par la Chambre de Commerce pour les années 1847-1848*, Paris, Guillaumin, 1851. The two surveys used the same methods, with small changes in the questionnaire and the presentation of results.

survey to point at its biases. We drew on her insights, but added data criticism based on more data gathering and data analysis. The Chamber of Commerce in fact published two surveys, one on the year 1848, the other on the year 1860. In the survey on 1848, gender was not systematically used as a category to present numbers of apprentices, but a few calculations allowed us to deduce that the Chamber found one thousand *fewer* female apprentices in 1860 than in 1848, even though the administrative limits of Paris had been expanded and the general population was higher. Comparing results at the scale of trades was not straightforward either, since the Chamber had changed some of its categories, but it led us to conclude that was had happened was most likely that in 1860, the Chamber did not register hundreds of apprentice linen-drapers (*lingères*), washerwomen and seamstresses. Those were female guilds in the eighteenth century; in the nineteenth century, these apprentices would have worked with female entrepreneurs, often subcontractors.

Our best guess as to the share of girls among Parisian apprentices is thus the one that we could reconstruct for 1848: roughly one third, similar to 1761 – a finding that we had not anticipated. We could then move on to discuss the fact that their occupations had become a bit more diverse by the mid-19th century.

III. Experiments in categorizing occupation

As this last point suggests, our aim is not merely to explore differences between situations explicitly considered as "apprenticeship" by the actors, such as in the court cases discussed above. We are also interested in the boundaries of apprenticeship: why some situations were labeled to be "apprenticeships" and others were not, even though they were in many ways similar. We are also very much concerned with perceived hierarchies of skill among youths labeled as "apprentices". In workshops on the history of apprenticeship, we have regularly heard colleagues casually interpreting prices paid to masters by families, or the planned durations of apprenticeships, as being the obvious consequence of the fact that some trades were more "skilled" than others. Of course, you would need more time, and pay more, to become a silversmith than a washerwoman. If one is willing to study conceptions of "skill" among contemporaries rather than imposing one's own prejudices, and to hypothesize that contemporaries might not have all agreed on these conceptions, data on apprenticeship offers interesting ways to think about lived categories, and hierarchies, of occupations.

Our approach to these issues has mostly been to keep the terminology as written in our source and then to aggregate categories based on one specific analytic question after another, using specific criteria for each question. Our general tenet in addressing problems of categorization is that time spent refining our questions and exploring data, including exceptions, weird data, data that seems difficult to categorize is time well spent. As with missing and inconclusive data, weird data often helps to refine questions. In terms of substantive results, this is, we think, more efficient than having colleagues spend a lot of time establishing general reference lists purporting to make comparisons easier between occupations in, say, twelfth-century Iceland and twentieth-century Nigeria, and other colleagues trying to interpret the exact phrasing of their sources with reference to such lists. We do not object at all to distant comparisons; but we do say that two occupations are equivalent if we decide to consider them as such, for specific substantive reasons – they are not, in themselves, to be considered similar because they would be manual, textile, or, even worse, unskilled.

a. Boundaries of Apprenticeship

In the eighteenth century, our research has demonstrated the use of the terms "apprenticeship" and "apprentice" for situations clearly deviating from the model of training within the confines of the guild system with the goal of becoming a guild master. Once again, female labor provides one of the best demonstrations of the conceptual blurriness around what is and was taken as a well-established and taken for granted norm. In the thousands of contracts indexed by archivists for eighteenth-century Paris (all the notarized contracts for 1751 and 1761, plus a handful of smaller thematic datasets), I found 94 training contracts for girls with male guild masters, among which 38 girls were described as "apprentices". This was forbidden by guild statutes, but the contracts were still notarized. An even more dramatic appropriation of the term occurred in a number of charitable institutions in late seventeenth-century Paris in which female artisans taught vocational skills to poor female "apprentices" in large classrooms. This training, explicitly labeled as "apprenticeship", also included instruction religious devotion and reading and writing[12]. By contrast, male youths receiving professional training to become merchants, notaries and other skilled occupations were described as "clerks", rather than as apprentices.

After the abolition of the guilds in the eighteenth century, there was still a legal definition of apprenticeship, but it was no longer linked to guilds. As apprenticeship no longer conveyed legal privileges vis-à-vis guild monopolies, the boundary between apprentices and young workers became even more blurred. For example, when asked to count apprentices, administrators gave answers that were clearly based on different definitions of apprenticeship from region to region.[13] Therefore, the fact that we do find apprentices described as such in some census lists and not in others probably signals different readings of situations by the masters and mistresses or by the census clerks as well as empirically different organizations of labor. Census lists provide an invaluable source that places apprentices in the context of a household, giving information on their parents and siblings, or on their master and mistress, the master or mistress's family and the rest of the co-resident workforce. For the nineteenth century, they are also the main source of names and ages of apprentices, information that allows us to track them in other sources and say something about their trajectory.

Therefore, the persons labelled "apprentices" in census lists are apprentices for us, because their master or mistress or a clerk described them as such (just as the eighteenth-century charity school girl was), and we compare girls and boys, apprentices in different cities and occupations, those who live with their master or mistress or with their parents, in order to better understand how apprenticeship worked in practice. But we also paid attention to, and gathered systematic data on other teenagers in the same city (in the censuses that we studied, mostly around 1850, few apprentices were younger than 13 or older than 18). This allowed us to derive the proportion of apprentices among teenagers, but also to systematically compare teenagers listed as "apprentices" with those ascribed an artisanal, industrial, or commercial occupation, without an explicit status or labeled as "workers", "boys", "girls", etc.

[12] Clare H. Crowston, « L'apprentissage hors des corporations. Les formations professionnelles alternatives à Paris sous l'Ancien Régime », *Annales. Histoire, Sciences Sociales*, 60-2, 2005, p. 409-441.

[13] See for example the answers of prefects in the 1880s in Archives nationales, F12 4831. The census lists that we discuss here have been digitized (as images) and are available on the website of department archives.

For example, in the Caen census of 1856 it is likely that the "baker boys" (*garçons boulangers*) were in the same social and legal situation as the "apprentice bakers" elsewhere, as we found no "apprentice baker" in Caen. Not noticing such differences in terminology would have led us to misinterpret overall proportions of apprentices in each city. Moreover, systematically comparing ways to describe teenagers across occupations – ie those labeled "boys"/"girls," "apprentices," "workers," "clerks" (*clerc* or *employé*), "assistants" (*commis*), etc. or just with the name of the occupation – provides us with interesting hierarchies of occupations, not entirely consistent across places. As in the eighteenth century, a young lawyer would always, everywhere, be a clerk, not an apprentice.[14] In Caen in 1856, a young lacemaker – there were many – would be a lacemaker, not an apprentice or a worker; on the contrary, the two last statuses were mentioned for seamstresses. More generally, girls were rarely called "apprentices" in Caen in 1856, much less often than in Lyon in 1851, for example, in spite of the fact that female guilds had thrived in Caen before the Revolution.

What is crucial to note here is that non-consistent descriptions of teenagers at work in censuses – and other documents – are not, for us, an annoyance, something that should disappear thanks to proper "data cleaning." We do not relegate the messy notes to paper in order to achieve a "clean" spreadsheet. On the contrary, a complicated spreadsheet including exceptions and multiple re-categorizations is our primary tool. Then calculations can tell us something about the entangled questions of the definition of apprenticeship, on the one hand, and the occupational hierarchy, on the other. We know from discursive sources that apprenticeship was associated with trades that were deemed skilled, but were not considered as professions, like lawyers and notaries; from sources such as the census lists, we confirm this general association, but we can draw precise boundaries, show how these boundaries differed according to time and place, how tight or blurred they were, etc.

b. "Skills" and the division of labor

The same principles apply to the issue of categorizing occupations. Sometimes, a handful of weird cases stimulate ideas for a more general categorization. Not incidentally, some of the eighteenth-century female "apprenticeship" contracts described above labeled the trade to be learnt as a specialty, or a specific task, not by using the name of a guild. For example, the female apprentice of a gilder would learn to gild tobacco boxes.[15] This hints at a division of labor in workshops that was hidden by the names of guilds. One century later, the Chamber of Commerce, in its surveys, praised the economic efficiency of the division of labor and listed as distinct trades activities that had been mere specialties inside eighteenth-century guilds, such as "polishing for jewelers" (a mostly female trade). But it still counted these workers and apprentices as part of the overall "industry" of their masters. It is only incidentally that the survey mentioned that workers with many different specialties in fact still worked in each workshop. For example, the accompanying text on locksmith workshops listed 25 (all male) different names of occupations for their workers; but their respective numbers and wages were not investigated. So

[14] Sven Steffens, « Le nom de l'apprenti : une analyse du vocabulaire socioprofessionnel », *Revue belge de philologie et d'histoire*, 79-2, 2001, p. 591-617 offered a first foray into these differences in vocabulary – but without any quantification.

[15] Archives nationales, MC/ET/IV/549, 12 June 1747.

we cannot know, for example, how many apprentices in the workshops of locksmiths learned mechanics generally and went on to use this set of skills in a different trade.

Going back and forth between the weird elements of our eighteenth and nineteenth-century sources thus provides a warning against the prevalent idea that apprentices learned all of their master's or mistress's trade and just that; however, it does not allow us to directly measure the transfer of skills. But we can give quantitative answers to other aspects of the general question "what does the division of labor do to the transferability of skills?" We can categorize eighteenth-century guilds as female, officially mixed, male with only male apprentices, and male with some female apprentices, or according to the uniformity or lack thereof or the description of skills imparted in notarized contracts[16]. Of course this is not direct information about what happened in workshops, but is also related to how individual masters and mistresses had recourse to notaries and the extent to which their occupational identities were encompassed by the guild to which they belonged. These ancillary issues must be taken into account when interpreting numbers. Nonetheless, paying close attention to trade descriptions given in these contracts provides some insight into an otherwise quite elusive question.

Likewise, the lists of workers' occupations in the Chamber of Commerce survey offer us scales of occupations, from those found in just one trade to the most ubiquitous. We thus find mechanics in 43 different trades in the survey for 1860, lathe hands (*tourneurs*) in 41 trades; and, among female occupations, polishers in 21 trades, burnishers (*brunisseuses*), who also treated the surface of metals, in 19. This is not to be taken as a direct and complete measurement of transferability either. It is likely, for example, that those who answered the questionnaire, the clerks, and the authors at the Chamber themselves, especially in the second survey, were less interested in the inner workings of workshops led by women than those led by men. Yet there are few if any other sources that give this level of detail on hundreds of trades.

Aware of these caveats, we used the indicator of the degree of division of labor, or diversity of skills, as one variable among others to describe each trade listed by the Chamber of Commerce. Other variables, such as the share of very small workshops, the proportion of female workers, or of apprentices in the workforce, were more directly copied or calculated from the survey. From a dozen such variables describing the workforce generally, we derived an automatic classification of 136 trades into four contrasted classes.[17] Then we checked the correlations

[16] For example, in eighteenth-century Dijon (Archives départementales de la Côte-d'Or), some masters promised to teach « la profession et l'art d'orfévrerie et bijouterie » (e.g. $4^{E}12/10$, 4 December 1743), others « la profession d'orfévrerie » (e.g., with the same notary, in the same year, $4^{E}12/10$, 22 November 1743); some mistresses promised to teach to or « blanchir et travailler en linges » (e.g. $4^{E}12/1900$, 18 August 1776), others the « profession de blanchisseuse » (e.g. $4^{E}2/2198$, 13 January 1859), others « le métier de blanchisseuse et à coudre le linge » (e.g. $4^{E}12/15$, 8 October 1747). We have not finished inputting this data and cannot therefore give definite numbers, but we consider it important not only to notice these variations but also to try to quantify them.

[17] More specifically: we first used principal component analysis on 11 variables that described the workshops (with only one related to apprenticeship: the share of apprentices in the workforce). We then performed cluster analysis on the first five dimensions of the PCA (more specifically, a variant of hierarchical ascending classification that uses k-means based on class paragons to stabilize the results – the hcpc procedure in the R-package FactoMineR: François Husson, Sébastien Lê and Jérôme Pagès, *Exploratory Multivariate Analysis by Example Using R*, Boca Raton, CRC Press, 2010). The algorithm suggested a solution in four classes. In plain English: the classes were grouped? by an algorithm so that trades in the same class would be as similar as possible to each other, and trades in different classes as different as possible from one another, based on our 11 primary variables. This analysis produced four classes, into which each of the trades was assigned. We then checked whether the trades in each of the classes were

between these classes of trades and variables that more specifically described the conditions of the apprenticeships. Our analysis showed that the four classes also differed in terms of conditions of apprenticeship, thus seemingly confirming the empirical evidence for the validity of the classification. For example, contracts were longer in one class than in the others; apprentices were more often only paid in kind in one of the classes. There is some correlation between these classes derived from data on the workforce and classical groupings, e.g. food vs. fashion, but not much, and there is almost no correlation with the fact that the trade had had a guild in the eighteenth century or not. We have thus taken data gathered, with many biases, by the Chamber of Commerce, we have processed it, on the basis of a critical scrutiny of the source, to build categories of trades that are absent from the source but, in our view, make sense as a description of the workforce. This analysis has produced new knowledge on apprenticeship – and possibly on Parisian industry generally. Yet we emphasize that a different research method could process the data differently and produce different, albeit equally valid, classes of trades.

*

In this presentation, we have given several short examples, at the expense of proving precise details on any of them, because it is important for us that the apprenticeship project relies on many different datasets, not just one. Comparing datasets, over time, space, and types of sources, helped us to better spot missing data, make sense of weird cases, and generally perform source criticism. It's not that the errors cancel themselves out, but the biases become more visible and we can make conscious decisions about what to do about them.

The consequence is also that we do not have one really big database, and that each is indeed quite small. For example, our census data is based on a sample of streets for Lyon and Caen, and we cannot dream of studying more than four or five places for one or two dates based on censuses. Well, we could, but we would need to divide labor much further and we have decided against it. We have inputted most of our data ourselves; for some of the sources that were easy to copy, especially censuses, we have used students, but we followed their work closely so as to incite them not to "clean" the data themselves, and we then categorized data in multiple ways, not just in order to produce results, but to grasp what was in the source as if we had read it (and we did so while striving to pay them a living wage for their labor). Colleagues often think that they will hate data if they input it themselves, because it will be so boring, and they will have to leave apart, or take as separate notes, everything that is really interesting in sources – stories, exceptions, strange language. We think that we can make them crave data if we consider that constructing data is just a slightly different way to take notes. This implies that we, and they, have to learn to sample for rich data, rather than hiring armies of data workers.

statistically associated or not to a series of other variables such as economic sectors defined by the Chamber of Commerce (food, clothing, etc.), the existence of a guild before 1791, etc. and to variables related to the conditions of apprenticeship (duration, etc.).