

# Comment on Data in Labor History session

SSHA 2019 in Chicago

Session 49

Peter B. Meyer

U.S Bureau of Labor Statistics

(Views and findings are those of the author, not the Bureau)

## Presenters:

[A Different History Manifesto: Data Need Not Be “Big” or “Clean”](#) • Clare H. Crowston

[Occupational Data and Classification: An Old and Recurring Dilemma for Labor and Social Historians](#) • Margo Anderson

[Measuring Skill: Gender, Race, and Occupations in the Early-20th Century U.S. Census](#) • Cristina Groeger

[The Persistence and Change of Classification in Official Statistics: The Case of Forced Labour.](#) • Theresa Wobbe

# Crowston: Data Not “Big” or “Clean”

- Systematic collection of history —> data
- Analysts “clean” data, which can include standardizing and overcategorizing; losing or adding key ambiguities
  - Especially for female apprentices
  - and others not treated as central
- Meanings and values are in the evidence/data
- Multiple categorization approaches help
  
- Here many sources: Guild records, Chamber of Commerce surveys, labor courts, contracts not approved by guilds any longer
  - Courts: 20% of Parisian apprentices left contracts early

# Apprenticeship data — small? dirty?

- Authors found many apprenticeship contracts which weren't known before
- 136 trades here, and ~11 variables
- Drew out — invented — a grouping into four classes of these, relating sectors of work to guild/occupation to conditions of work
- Synthesis => Small datasets, that include invention and categorization by specialists

# Margo Anderson on occupation classification

- Anderson's work is fundamental and unique
- In US census, occupation and info was recorded as open-ended text and classified by specialists
  - Some of whom I met
- Problem: categories appear, disappear, or are residuals (“operative, not specified”, “not elsewhere classified”)
- Worse: Specialists often do not have much info to classify occ (and ind); Census chief Alba Edwards said to use relational information (sex, race, age, nativity, location...)
- So it would be circular/unreliable to use occupation to make detailed inference about these categories, and social class
- 1900, 1910: change in counting women as employed with occupation (Bose)
- 1940: “too many” women were in skilled trades so Census reclassified

# Margo Anderson on occupation classification

- Census procedures are relevant & necessary to understand occupation — it's metadata!
- But sometimes hidden
- Related problems: changing concepts of labor force, unemployment, retirement, slaves, Indians, housework, military, the institutionalized, volunteers, apprentices, persons at leisure, persons in new industry
- Relatedly: changing race in the data
- She noted in earlier work absence of advisory committees to Census
- New Census format: Fewer questions. Two sexes.

# Groeger on social class and occupation data in 1880s Boston

- Occupation doesn't perfectly predict "class" — e.g. teachers, musicians, merchant/dealers
- High social class in Boston: professionals, traders, financiers, and manufacturers
- 1880 Census data didn't distinguish "confectioner" staff from owner of establishment
- Groeger finds Census manuscripts reveal more: about house, neighbors, and servants if any
- Many examples

# Possible approaches to ambiguous “social class”

- Unbundle the elements: track wealth, neighborhood, education, etc. separately
- Combine elements into an index. (“She .75 upper class”)
- Distinguish elites by type: financial, educational, power, wealthy, honorific, artistic upper classes, which overlap
- Construct social networks with enriched relations: “married”, “hired”, “was witness for”, “trusted”, “advocated for”
  - Then construct the class you want in context
- (And be bold ; there is no perfect way anyway)

# Wobbe on forced labor

- “Forced labor” isn’t always classified as labor
- 1920s ILO categorized “forced labour” legally, but not in data
  - Earlier, U.S. slaves had skills, tasks, but were not recorded as having occupations
  - Issues: work for wages / for others / for market
- 20 million people now
- Wobbe’s view; it should be categorized as labor
- Is on ILO project to establish definitions and practices
- Concept: Duress in hiring, work conditions, or leaving



# Synthetic thoughts and the future

- Use multiple sources!
- Use enriched databases, e.g. with text, not just rows on a spreadsheet
  - Wikis can help — If each observation is rich, then make a wiki page out of it and work out how to classify it by seeing and summarizing other cases. This addresses a deep problem of historical quantitative research.
- Data science can help, if wise
  - Know the data; learn what predicts what; and understand it
  - If all blacksmiths in the data are male, but we know there were female blacksmiths, report this and adapt as well as possible
  - Match to other data ; benchmark