香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

# Releasing research data: Our experiences with the China Multigenerational Panel Databases (CMGPD)

**Cameron Campbell**
**HKUST**

**On behalf of the Lee-Campbell Research Group**

***Columbia University***
***29 March** 2016*

# Outline

- Why should we think about data sharing/data release?
- Three projects
  - CMGPD-Liaoning
  - CMGPD-Shuangcheng
  - Jinshenlu 缙绅录
- What we learned
  - From our experience
  - From the experience of others

# Data sharing/data release: Why?

- Sharing/release of research data initially emerged as a norm in physical and life sciences.
  - To facilitate replication.
- Funding agencies supporting research in social science and other areas now routinely expect/demand plans for data preservation, sharing/release.
- Journals increasingly expect data and code to be made available in an online appendix or elsewhere.

# Additional reasons

- If you care about citations, releasing your data may help with your counts.
  - Hopefully, others who use your data will cite your data, your documentation, and the papers you wrote with it.
- People in other fields may find uses for your data that you never anticipated.
  - Possibly incorporating it into aggregatations.
  - And if they cite you, more sweet citations.
- Releasing your data and your code will increase trust in your results.

Google

Scholar

Edit

Export

Cameron Campbell

# China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN), 1749–1909 [Computer file]

Authors JZ Lee, C Campbell

Publication date 2010

Pages 10-01

Publisher Ann Arbor, MI: Inter-university Consortium for

Total citations Cited by 16


Scholar articles

China multi-generational panel dataset, Liaoning (CMGPD-LN), 1749-1909 *
JZ Lee, CD Campbell - 2010
Cited by 14 - Related articles

China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN), 1749-1909 [Computer file]. ICPSR27063-v2 *
JZ Lee, CD Campbell - Ann Arbor, MI: Inter-university Consortium for Political ..., 2010
Cited by 3 - Related articles

China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN), 1749-1909. ICPSR27063-v10 *
JZ Lee, CD Campbell - Ann Arbor, MI: Inter-university Consortium for Political ..., 2014
Cited by 2 - Related articles - All 3 versions

China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN), 1749–1909 [Computer file]
JZ Lee, C Campbell - 2010
Cited by 2 - Related articles

China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN) *
JZ Lee, CD Campbell
Related articles

# CMGPD-LN

- China Multigenerational Panel Dataset-Liaoning
  - http://www.icpsr.umich.edu/icpsrweb/DSDR/studies/27063
- James Lee began collection and entry in the early 1982.
  - Microfilmed materials from the Liaoning Provincial Archives
- I joined in 1987, and helped organize and analyze the data.
  - dBase III+ and dBase IV!
- The dataset eventually grew to 1.5 million records describing 260,000 people between 1749 and 1909.
- We released the data at the Interuniversity Consortium for Political and Social Research in 2010.
  - Preparation of the CMGPD-SC and documentation for public release via ICPSR DSDR was supported by United States Department of Health and Human Services National Institutes of Health Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) 1R01HD070985-01 "Multi-generational Demographic and Landholding Data: CMGPD-SC Public Release.
- Cited in 75 publications, including our own.

# Format and organization

- 1.5 million triennial observations of 260,000 people from 1749-1909
  - 1,051 paternal descent groups identified through record linkage
  - 698 communities
- Longitudinal
  - Resemble triennial censuses
  - Individuals and households are listed in the same order in successive registers, and can be linked
- Generational depth
  - Paternal lines linked at least 7 generations

# Producing the CMGPD-LN (I)

- The registers from which the CMGPD-LN was produced are held in the Liaoning Provincial Archives
- They have been microfilmed and archived by FamilySearch (formerly the Genealogical Society of Utah) https://familysearch.org/
- The images we use are scanned from microfilm.
- We provide these images to coders in China.
- Coders in China transcribe contents to Excel spreadsheets
  - Copy previous spreadsheet over and update based on contents of new register
  - Link each record in the new register to that individual's record in the previous register, based on record number.

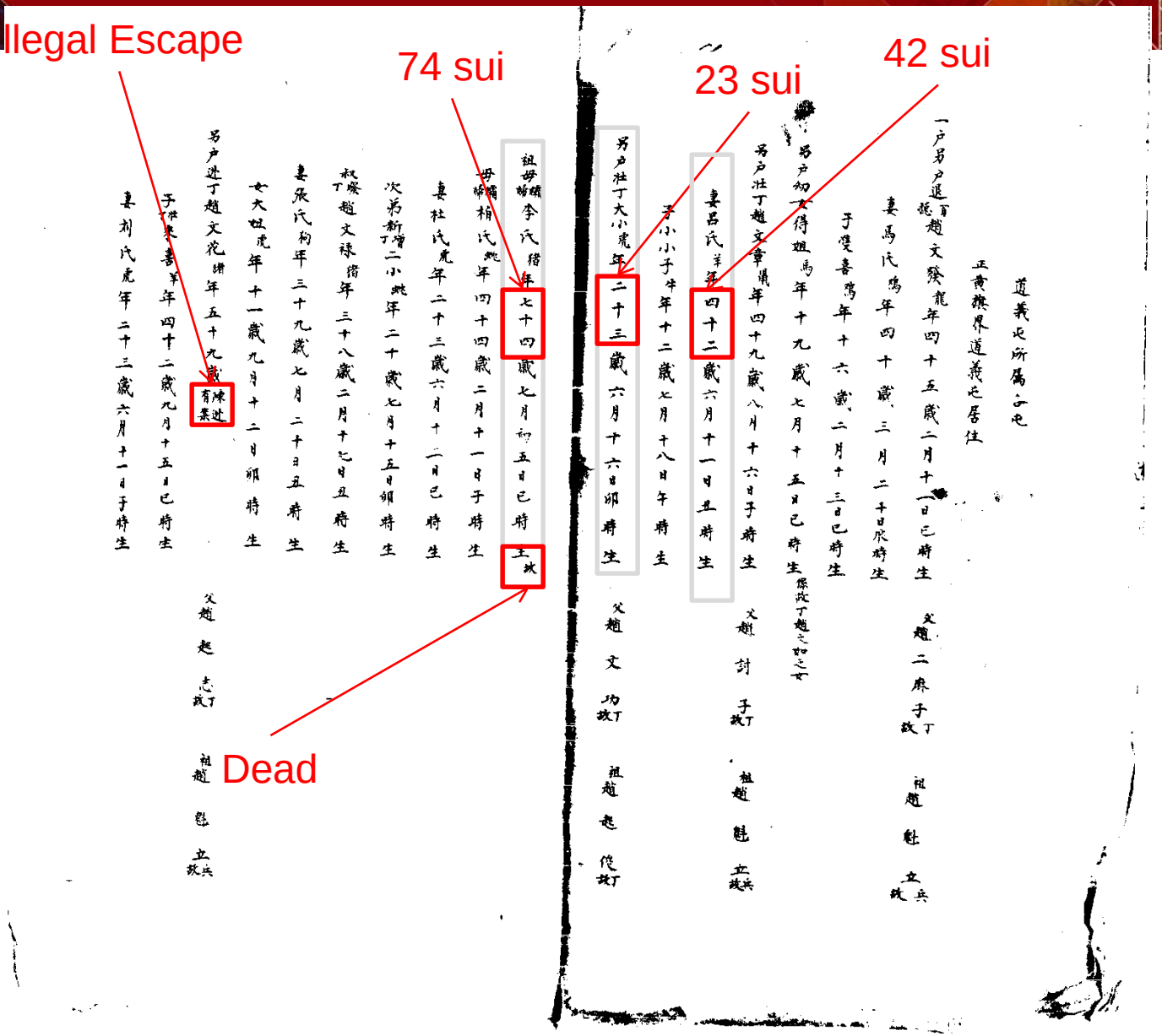
# Producing the CMGPD-LN (II)

- STATA programs import the contents of the spreadsheets and perform error-checking
  - Programs check for inconsistencies across registers
- Reports sent to coders for cleaning
  - Original registers coded ‘as is’: if there is an inconsistency in the original register, it is coded that way, and dealt with later by the software.
- STATA programs link kin and generate variables for analysis
  - Link records of the same individual in successive registers, and assign a common ID (PERSON_ID).
  - Link children to parents based on RELATIONSHIP
  - Reconstruct pedigrees

# Contents of original data

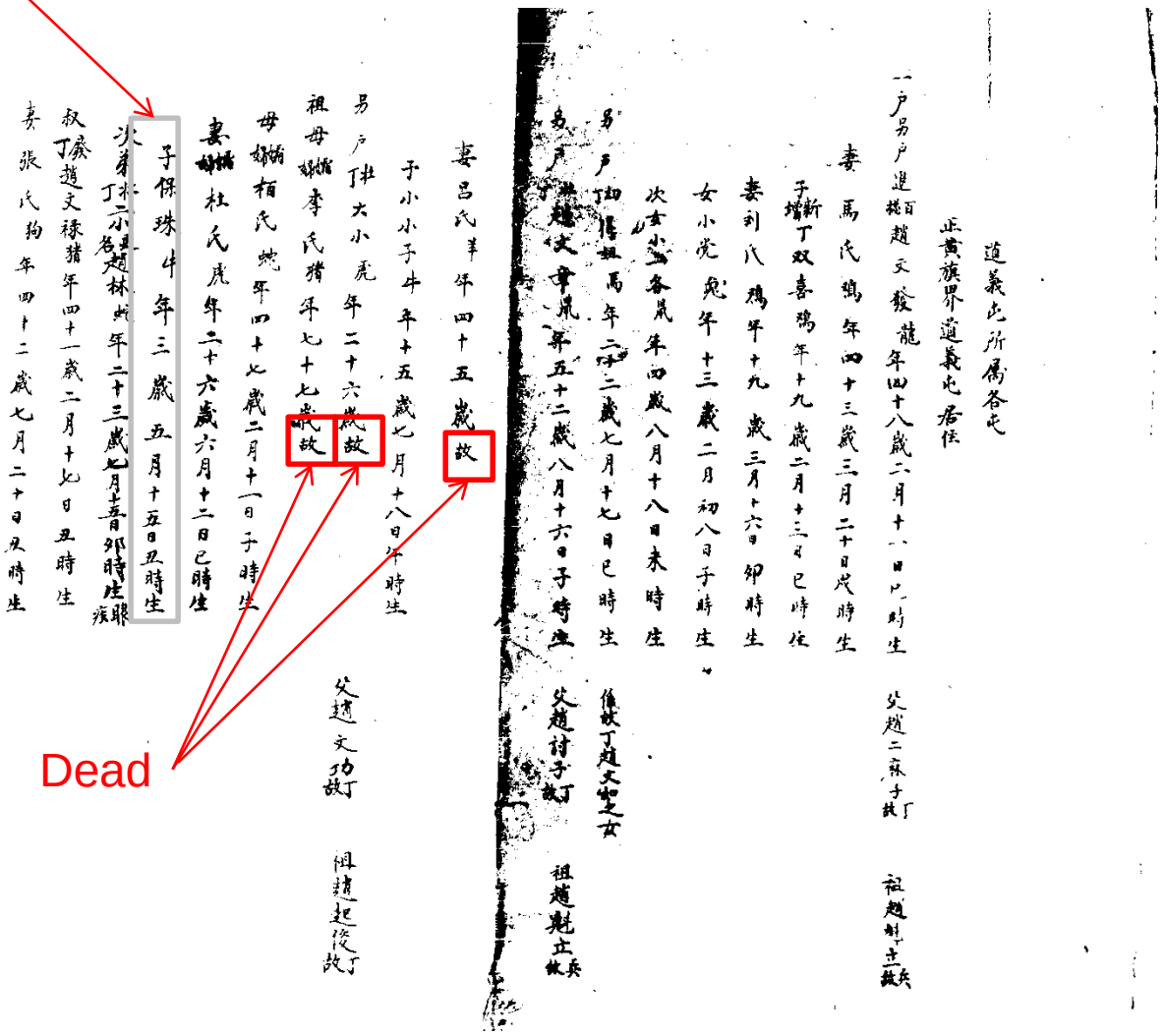
- Relationship
- Official position, title or status (for adult males)
- Name
  - Married women only had a maiden name.
- Zodiac year of birth
- Age in *sui*
- Birth date
- Annotations of events (mostly exits) that have occurred since the last register
  - Death, illegal escape, legal out-migration, out-marriage, out-remarriage
- Household heads had the names of their father and grandfather recorded

# Daoyi 1816



# Daoyi 1819

New arrival

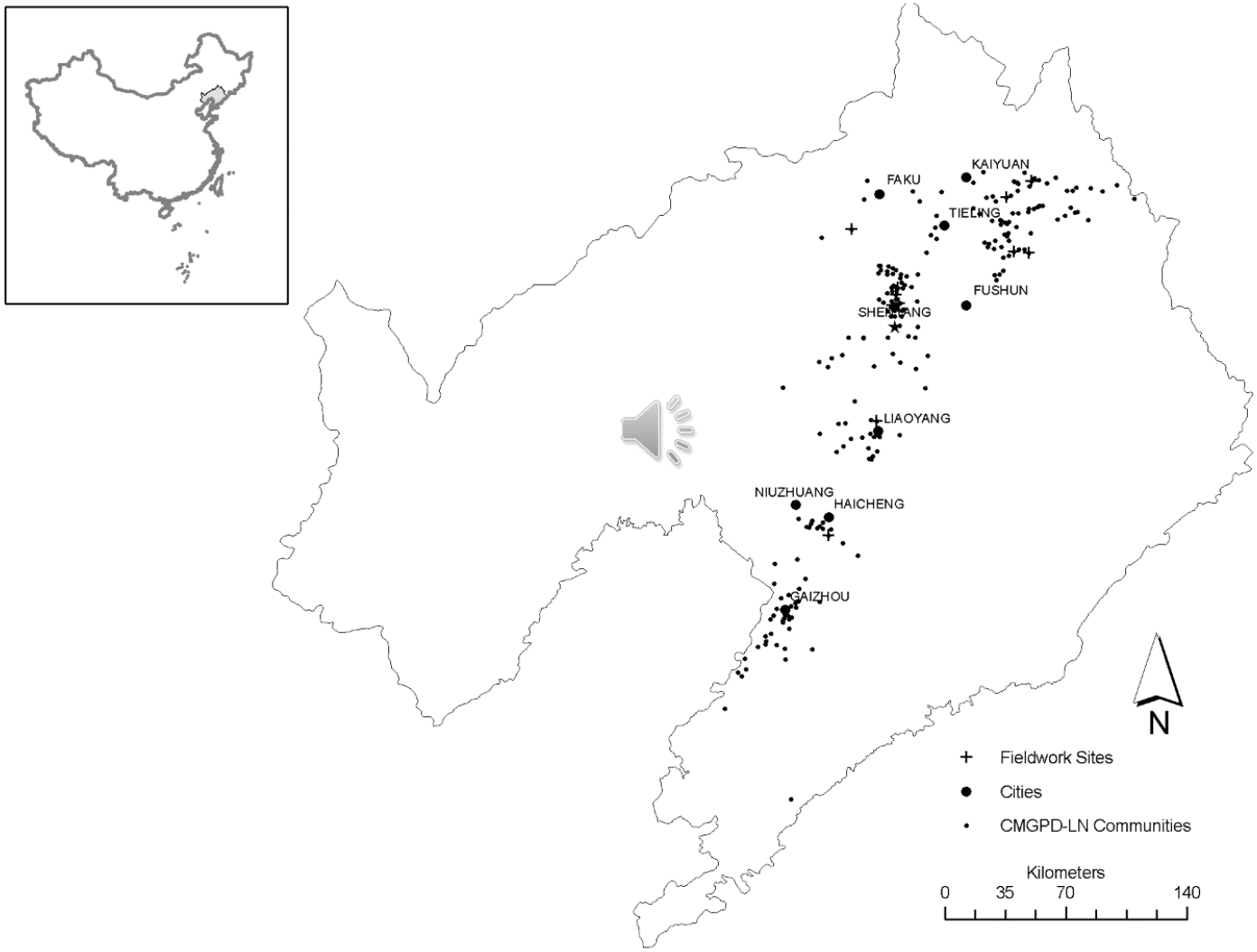


# Daoyi 1816 and 1819

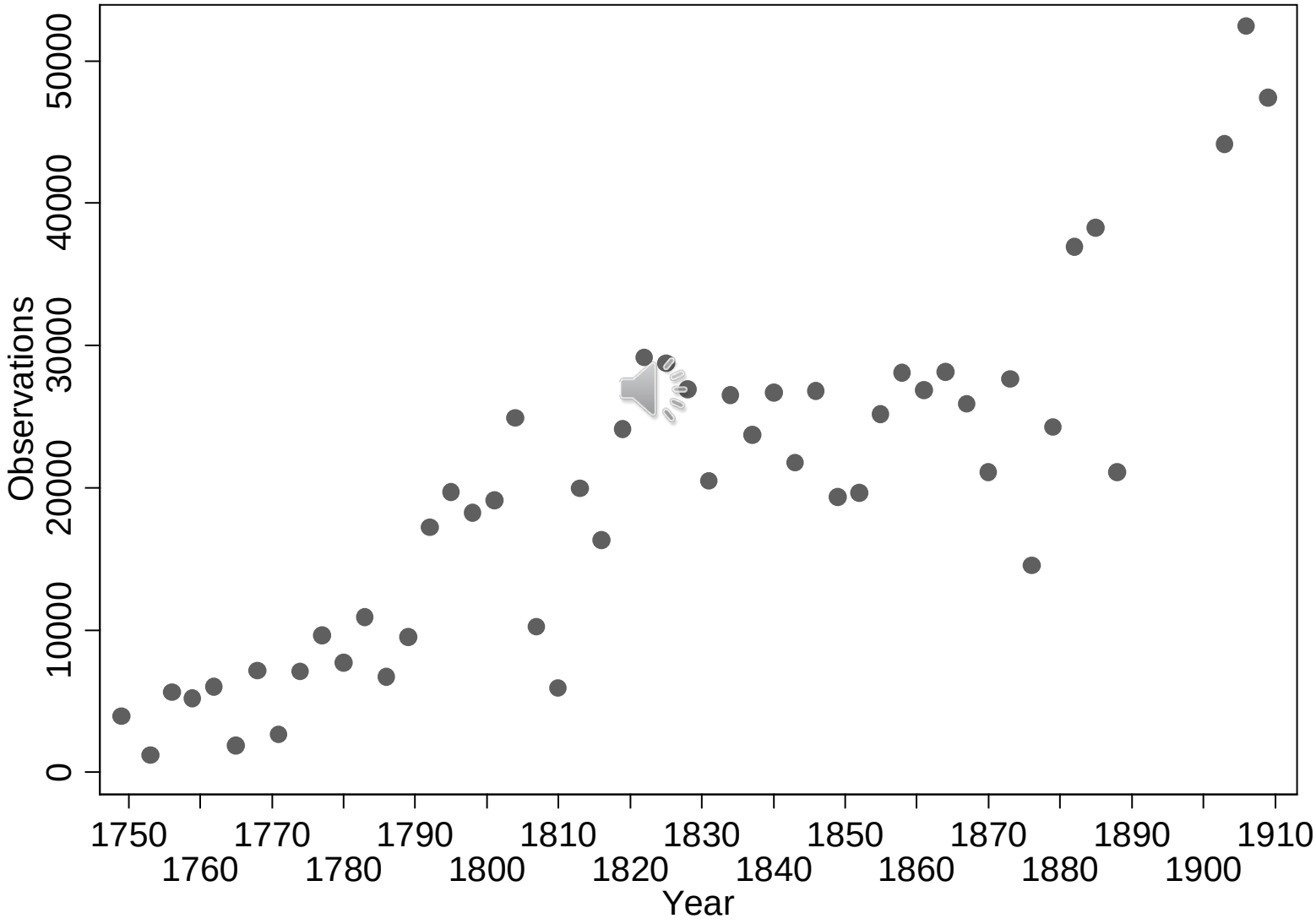
| | A | B | C | D | E | F | G | H | I | J | L | M | N | O | P | Q | R | S | T | U | AB | AC | AE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | case | date | interpol | clan | hid | yihu | rhhead | ws | address | modage | sex | gen | rank | occu | marital | name | vital1 | vital2 | birthyr | age | lcase | ldate | id |
| 2 | 1 | 816 | 0 | 1 | 1 | 0 | e | | 101 | 45 | 2 | 3 | 2 | 11 | 4 | zhao wenfa | 0 | 0 | 5 | 45 | 4 | 813 | 2374 |
| 3 | 2 | 816 | 0 | 1 | 1 | 0 | w | | 101 | 40 | 1 | 3 | 0 | 0 | 4 | ma | 0 | 0 | 10 | 40 | 5 | 813 | 5347 |
| 4 | 3 | 816 | 0 | 1 | 1 | 0 | 1s | | 101 | 16 | 2 | 4 | 0 | 0 | 2 | shuangxi | 0 | 0 | 10 | 16 | 6 | 813 | 5642 |
| 5 | 4 | 816 | 0 | 1 | 2 | 0 | e | | 101 | 19 | 1 | 3 | 0 | 0 | 2 | dejie | 0 | 0 | 7 | 19 | 2 | 813 | 5641 |
| 6 | 5 | 816 | 0 | 1 | 3 | 0 | e | | 101 | 49 | 2 | 3 | 0 | 1 | 1 | zhao wenzhang | 0 | 0 | 1 | 49 | 7 | 813 | 5 |
| 7 | 6 | 816 | 0 | 1 | 3 | 0 | w | | 101 | 42 | 1 | 3 | 0 | 0 | 1 | lu | 0 | 0 | 8 | 42 | 8 | 813 | 4008 |
| 8 | 7 | 816 | 0 | 1 | 3 | 0 | 1s | | 101 | 12 | 2 | 4 | 0 | 0 | 2 | xiaoxiao | 0 | 0 | 8 | 12 | 9 | 813 | 6103 |
| 9 | 8 | 816 | 0 | 1 | 4 | 0 | e | | 101 | 23 | 2 | 3 | 0 | 1 | 1 | daixiao | 0 | 0 | 3 | 23 | 10 | 813 | 4616 |
| 10 | 9 | 816 | 0 | 1 | 4 | 0 | fm | | 101 | 74 | 1 | 1 | 0 | 0 | 0 | li | 1 | 0 | 12 | 74 | 11 | 813 | 27 |
| 11 | 10 | 816 | 0 | 1 | 4 | 0 | m | | 101 | 44 | 1 | 2 | 0 | 0 | 3 | bai | 0 | 0 | 6 | 44 | 12 | 813 | 4615 |
| 12 | 11 | 816 | 0 | 1 | 4 | 0 | w | | 101 | 23 | 1 | 3 | 0 | 0 | 1 | du | 7 | 0 | 3 | 23 | 0 | 0 | 6981 |
| 13 | 12 | 816 | 0 | 1 | 4 | 0 | 2yb | | 101 | 20 | 2 | 3 | 0 | 3 | 2 | erxiao | 0 | 0 | 6 | 20 | 13 | 813 | 5643 |
| 14 | 13 | 816 | 0 | 1 | 4 | 0 | f2yb | | 101 | 38 | 2 | 2 | 0 | 6 | 1 | zhao wenlu | 0 | 0 | 12 | 38 | 14 | 813 | 2378 |
| 15 | 14 | 816 | 0 | 1 | 4 | 0 | f2ybw | | 101 | 39 | 1 | 2 | 0 | 0 | 1 | zhang | 0 | 0 | 11 | 39 | 15 | 813 | 5010 |
| 16 | 15 | 816 | 0 | 1 | 4 | 0 | f2yb1d | | 101 | 11 | 1 | 3 | 0 | 0 | 2 | daniu | 0 | 0 | 3 | 11 | 16 | 813 | 6104 |
| 17 | 16 | 816 | 0 | 1 | 5 | 0 | e | | 101 | 65 | 2 | 3 | 0 | 13 | 3 | zhao wenhua | 4 | 0 | 12 | 59 | 17 | 813 | 20 |
| 18 | 17 | 816 | 0 | 1 | 5 | 0 | 1s | | 101 | 42 | 2 | 4 | 0 | 1 | 4 | laixi | 0 | 0 | 8 | 42 | 18 | 813 | 2377 |
| 19 | 18 | 816 | 0 | 1 | 5 | 0 | 1sw | | 101 | 23 | 1 | 4 | 0 | 0 | 4 | liu | 0 | 0 | 3 | 23 | 19 | 813 | 6105 |

| | A | B | C | D | E | F | G | H | I | J | L | M | N | O | P | Q | R | S | T | U | V | W | X | AB | AC | AE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | case | date | interpol | clan | hid | yihu | rhhead | ws | address | modage | sex | gen | rank | occu | marital | name | vital1 | vital2 | birthyr | age | month | day | hour | lcase | ldate | id |
| 2 | 1 | 819 | 0 | 1 | 1 | 0 | e | | 101 | 48 | 2 | 3 | 2 | 11 | 4 | zhao wenfa | 0 | 0 | 5 | 48 | 2 | 11 | 6 | 1 | 816 | 2374 |
| 3 | 2 | 819 | 0 | 1 | 1 | 0 | w | | 101 | 43 | 1 | 3 | 0 | 0 | 4 | ma | 0 | 0 | 10 | 43 | 3 | 20 | 11 | 2 | 816 | 5347 |
| 4 | 3 | 819 | 0 | 1 | 1 | 0 | 1s | | 101 | 19 | 2 | 4 | 0 | 3 | 1 | shuangxi | 0 | 0 | 10 | 19 | 2 | 13 | 6 | 3 | 816 | 5642 |
| 5 | 4 | 819 | 0 | 1 | 1 | 0 | 1sw | | 101 | 19 | 1 | 4 | 0 | 0 | 1 | liu | 7 | 0 | 10 | 19 | 3 | 16 | 4 | 0 | 0 | 7203 |
| 6 | 5 | 819 | 0 | 1 | 1 | 0 | 1d | | 101 | 13 | 1 | 4 | 0 | 0 | 2 | xiaodong | 6 | 0 | 4 | 13 | 2 | 8 | 1 | 0 | 0 | 7204 |
| 7 | 6 | 819 | 0 | 1 | 1 | 0 | 2d | | 101 | 4 | 1 | 4 | 0 | 0 | 2 | xiaoergo | 6 | 0 | 1 | 4 | 8 | 18 | 8 | 0 | 0 | 7205 |
| 8 | 7 | 819 | 0 | 1 | 2 | 0 | e | | 101 | 22 | 1 | 3 | 0 | 50 | 2 | dejie | 0 | 0 | 7 | 22 | 7 | 17 | 6 | 4 | 816 | 5641 |
| 9 | 8 | 819 | 0 | 1 | 3 | 0 | e | | 101 | 52 | 2 | 3 | 0 | 1 | 3 | zhao wenzhang | 0 | 0 | 1 | 52 | 8 | 16 | 1 | 5 | 816 | 5 |
| 10 | 9 | 819 | 0 | 1 | 3 | 0 | w | | 101 | 45 | 1 | 3 | 0 | 0 | 0 | lu | 1 | 0 | 8 | 45 | 0 | 0 | 0 | 6 | 816 | 4008 |
| 11 | 10 | 819 | 0 | 1 | 3 | 0 | 1s | | 101 | 15 | 2 | 4 | 0 | 0 | 2 | xiaoxiao | 0 | 0 | 2 | 15 | 7 | 18 | 7 | 7 | 816 | 6103 |
| 12 | 11 | 819 | 0 | 1 | 4 | 0 | e | | 101 | 26 | 2 | 3 | 0 | 1 | 0 | daixiao | 1 | 0 | 3 | 26 | 0 | 0 | 0 | 8 | 816 | 4616 |
| 13 | 12 | 819 | 0 | 1 | 4 | 0 | fm | | 101 | 77 | 1 | 1 | 0 | 0 | 0 | li | 1 | 0 | 12 | 77 | 0 | 0 | 0 | 9 | 816 | 27 |
| 14 | 13 | 819 | 0 | 1 | 4 | 0 | m | | 101 | 47 | 1 | 2 | 0 | 0 | 3 | bai | 0 | 0 | 6 | 47 | 2 | 11 | 1 | 10 | 816 | 4615 |
| 15 | 14 | 819 | 0 | 1 | 4 | 0 | w | | 101 | 26 | 1 | 3 | 0 | 0 | 3 | du | 0 | 0 | 3 | 26 | 6 | 12 | 6 | 11 | 816 | 6981 |
| 16 | 15 | 819 | 0 | 1 | 4 | 0 | 1s | | 101 | 3 | 2 | 4 | 0 | 0 | 2 | baozhu | 6 | 0 | 2 | 3 | 5 | 15 | 2 | 0 | 0 | 7206 |
| 17 | 16 | 819 | 0 | 1 | 4 | 0 | 2yb | | 101 | 23 | 2 | 3 | 0 | 1 | 2 | zhao lin | 0 | 0 | 6 | 23 | 7 | 15 | 4 | 12 | 816 | 5643 |
| 18 | 17 | 819 | 0 | 1 | 4 | 0 | f2yb | | 101 | 41 | 2 | 2 | 0 | 6 | 1 | zhao wenlu | 0 | 0 | 12 | 41 | 2 | 17 | 2 | 13 | 816 | 2378 |
| 19 | 18 | 819 | 0 | 1 | 4 | 0 | f2ybw | | 101 | 42 | 1 | 2 | 0 | 0 | 1 | zhang | 0 | 0 | 11 | 42 | 7 | 20 | 2 | 14 | 816 | 5010 |

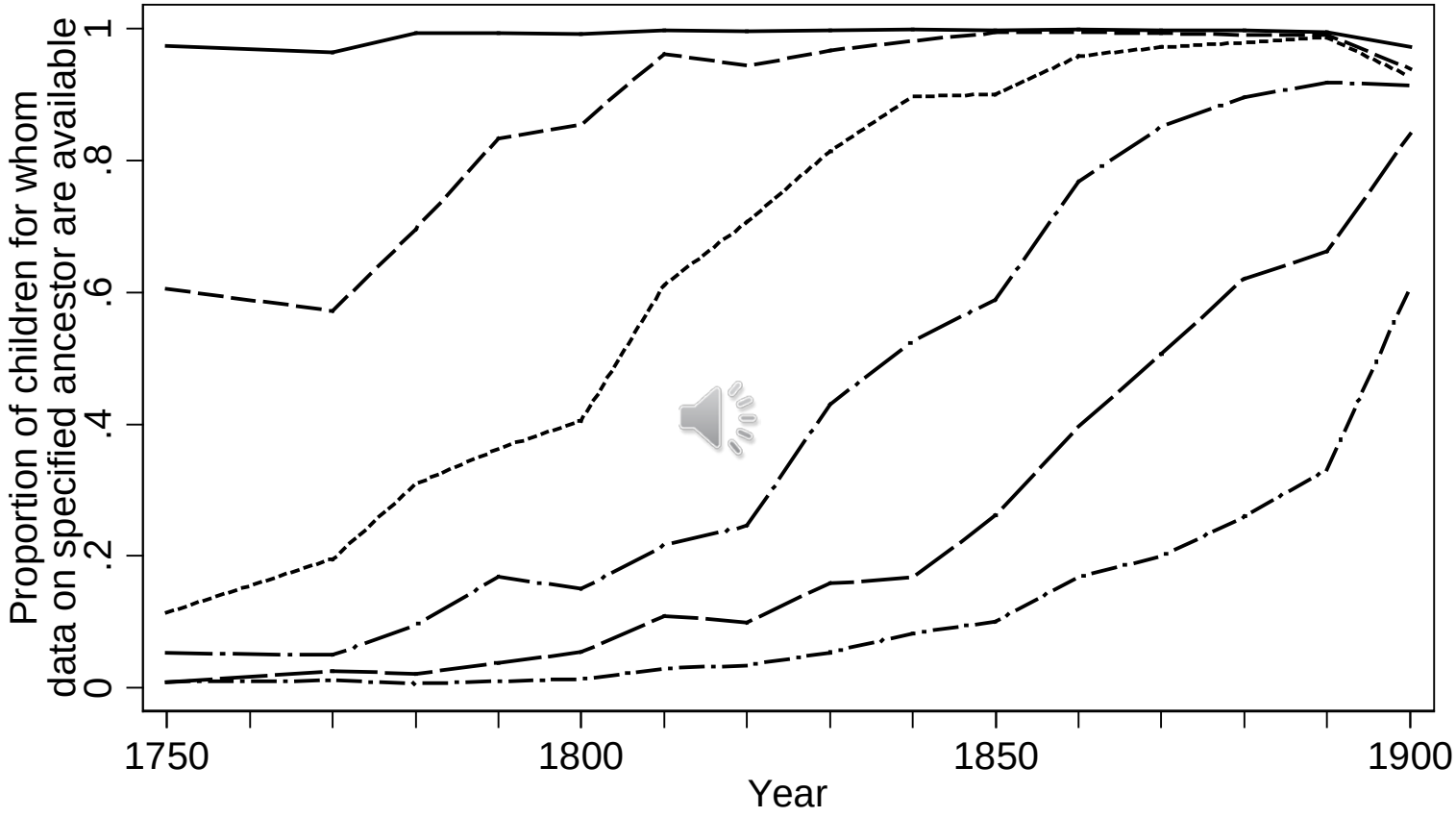
# Geographic coverage of the CMGPD-LN



# Observations by Year in the CMGPD-LN



# Multi-generational Linkage in the CMGPD-LN




[redacted]

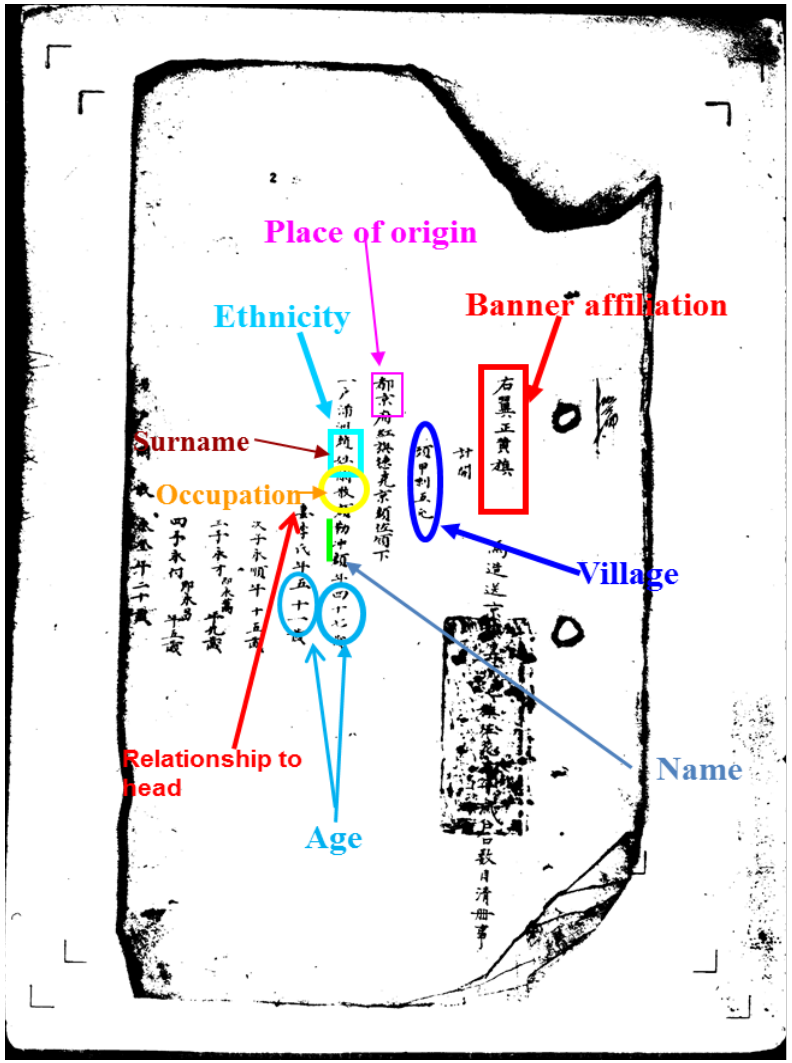
# China Multi-Generational Panel Dataset - Shuangcheng (CMGPD-SC)

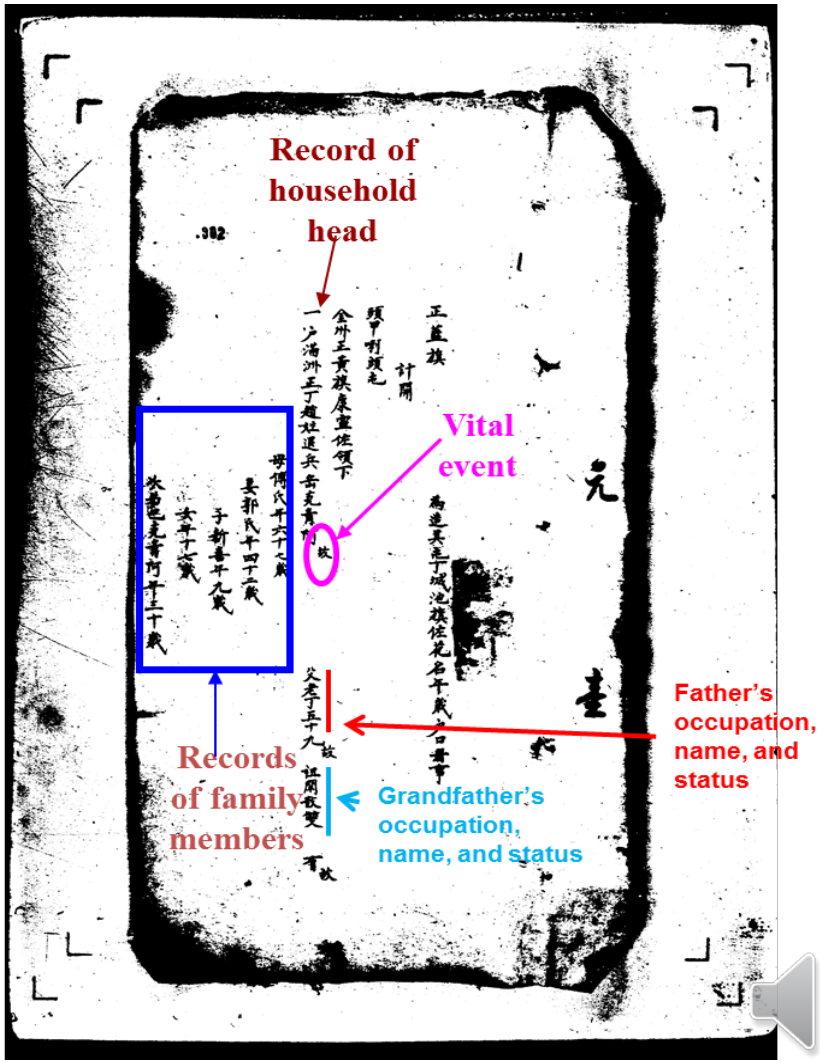
- Longitudinal individual, household, and community information on the demographic and socioeconomic characteristics of a migrant population living in Shuangcheng (1866 to 1912)
- Three categories of bannermen: metropolitan (*jingqi*), rural (*tunding*), and floating (*fuding*)
  - *Jingqi* (metropolitan) were rusticated bannermen who resettled from Beijing in the first half of the 19th century.
  - *Tunding* (rural) were farmers relocated from elsewhere in northeast China at roughly the same time.
  - *Fuding* (floating) were unofficial migrants, mostly from elsewhere in northeast China.
- Publicly released at ICPSR: http://www.icpsr.umich.edu/icpsrweb/DSDR/studies/35292

# Sources for the CMGPD-SC

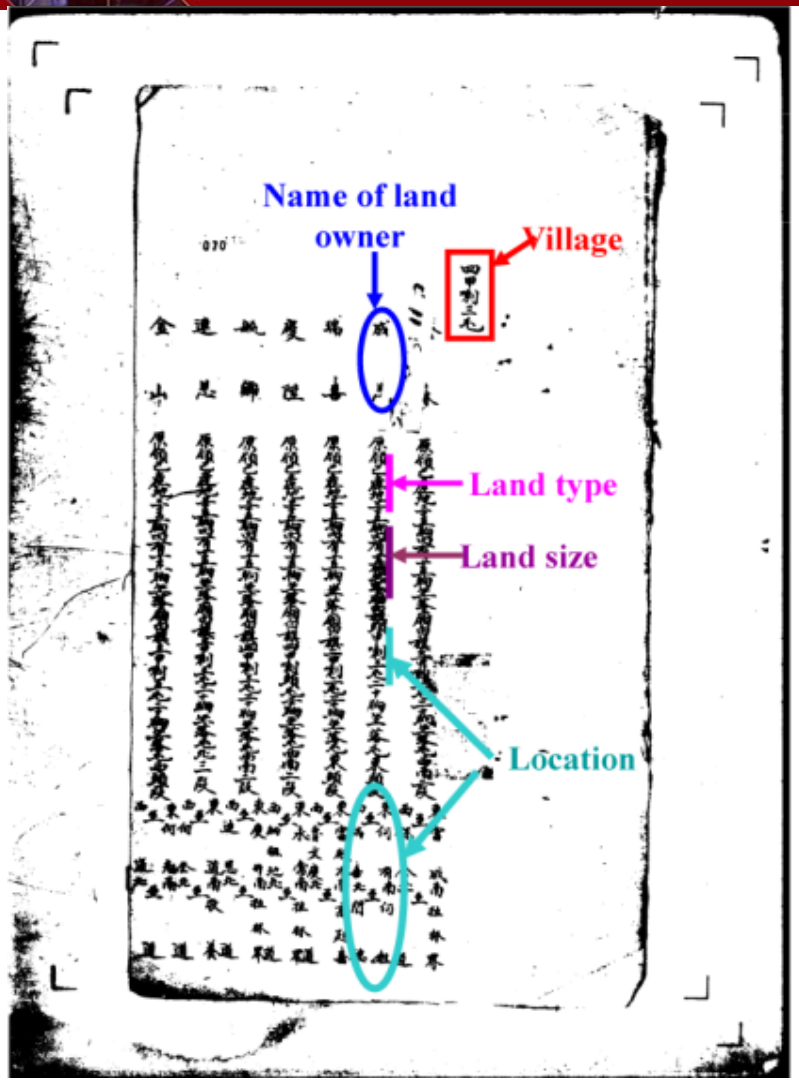
- Qing population registers (260 coded and linked registers from 1866 to 1913):
  1. Metropolitan population (jingqi): annual
  2. Rural population (tunding): annual
  3. Floating population (fuding): triennial
- 1,346,829 observations for 108,100 linked individuals
- Land registers from 1870, 1876, 1882, 1887, 1889, and 1906: over 19,000 cultivated plots

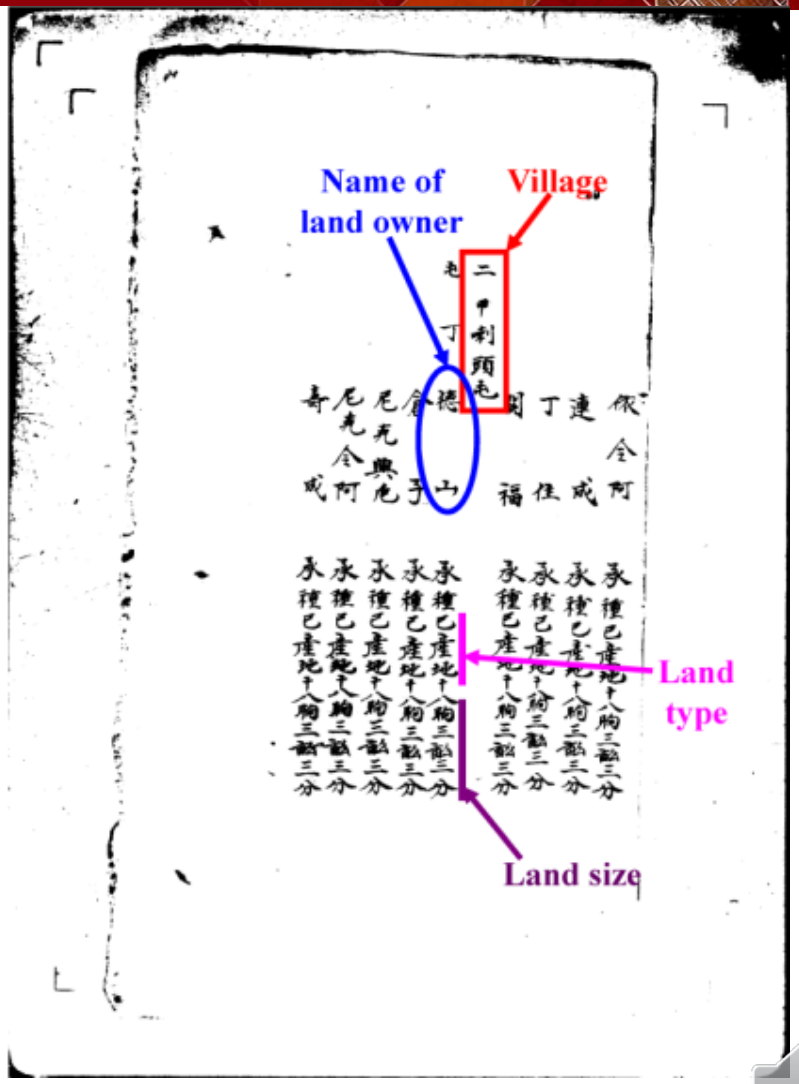
# Sample household register pages from Shuangcheng



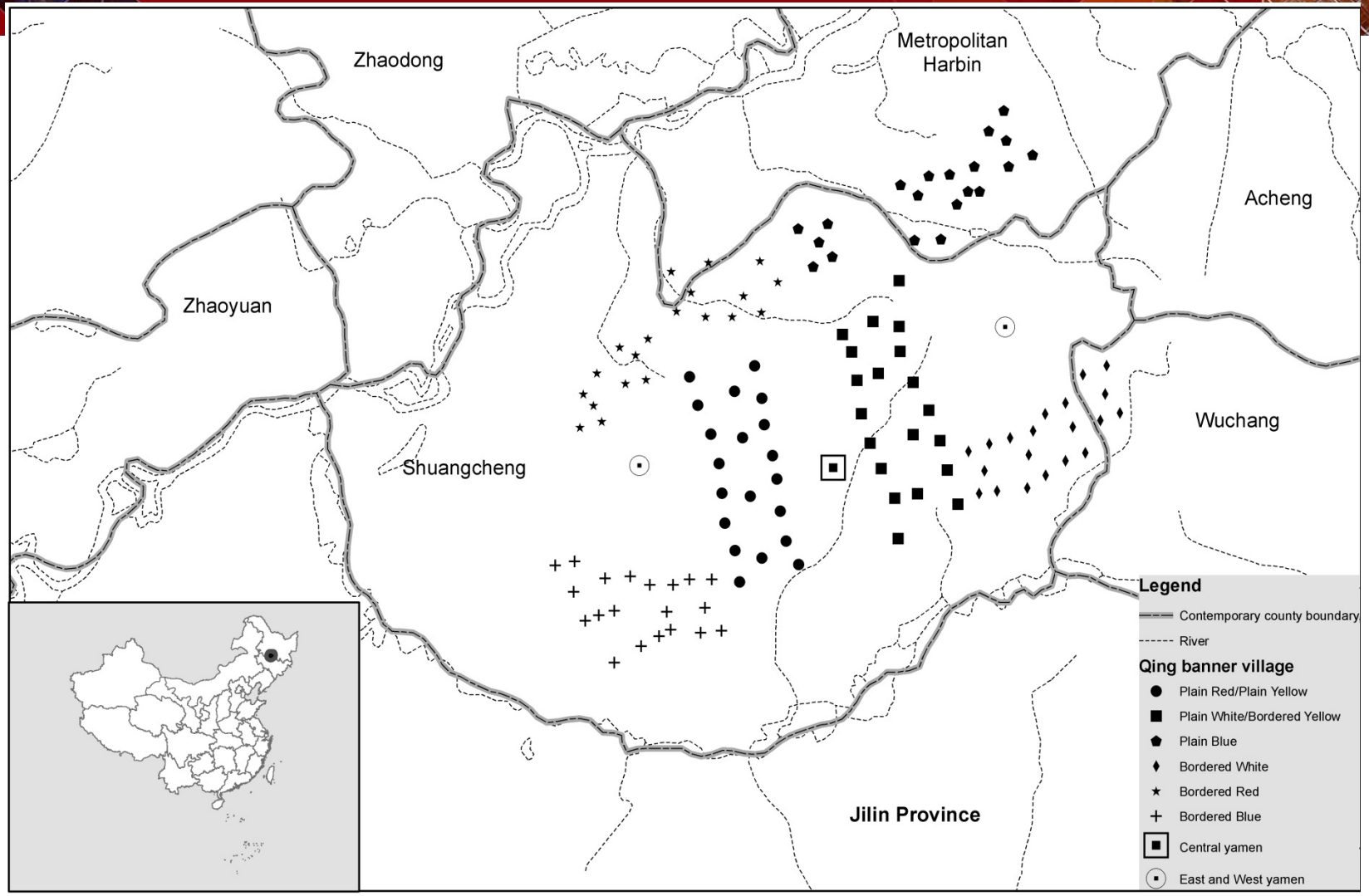



# Sample landholding register pages from Shuangcheng



# Shuangcheng map



# New project: *jinshenlu* (缙绅录)

- The core of the database is the Qing *jinshenlu* (缙绅录), which listed officials on a quarterly basis.
- We are currently working with a collection published by Tsinghua in 2008 which includes 209 out of 230 surviving editions held in the university library.
- Another 120 editions are held at the National Library, Peking University Library, other libraries including international libraries.
- Editions record 13000-15000 officials, from the Six Minstries all the way down to county administrations.
- Our current data include far more civil than military officials and as a result some civil officials who started their career in the military and then transitioned may only be captured from the time they were already well advanced in their career.

# 《清代缙绅录数据库》文献来源

- 以清代缙（搢）绅录文献为基础
- 清华大学图书馆现存230余种，其中209种已结集出版，即《清代缙绅录集成》（大象出版社2008年版）
- 国家图书馆、北京大学图书馆等十九个图书馆，另有120余种未出版
- 海外图书馆也藏有一定数量的缙绅录
- 目前录入的缙绅录数据以清华大学图书馆出版的《清代缙绅录集成》为基础

# Distinctive features of the *jinshenlu*

- As far as we know, the *jinshenlu* is the only complete record of an entire pre-twentieth century national bureaucracy, certainly of this scale, that is also easy to transcribe into a database.
- It covers the period from the mid 18th century to the beginning of the 20th.
- While the original data are cross-sectional, we can produce longitudinal records of individuals by nominative linkage.
- Initial experiments indicate that nominative linkage to such other sources as the 提名录 is also straightforward.

# 清代缙绅录数据库的研究价值和潜力

- 据我们所知，清代缙绅录是世界范围内唯一的长时段的国家职官手册，尤其适合录入成为数据库
  - 其他国家公务人员也有个人档案或履历记载，但是不容易做成统一格式的数据库
- 长时段：清华大学版清代缙绅录集成涵盖乾隆到宣统
- 横截面：按季度有特定时间点的完整记录，涵盖绝大部分官员和吏员及其隶属机构
- 个人数据的可追踪性：可以通过个人信息的连接追踪其仕途的迁转
- 与其他数据库的连接潜力：比如进士题名录、乡试硃卷、清代官员履历档案

# Contents

- Surname, given name and appellation.
  - Most Manchus and other non-Han only have a given name.
- Province and county of origin (for Han)
- Examination degree
- Banner affiliation, ethnicity
- Government office and position
- Bureaucratic rank (pinji 品级) 1-9
  - 1 was highest, and 9 the lowest.
  - Salaries of the highest rank were 6 times that of the lowest rank.
  - Year of appointment for current position
  - Sometimes appointment authority (Ministry of Personnel)
- Details on the posting, including ratings for difficulty and sensitivity.

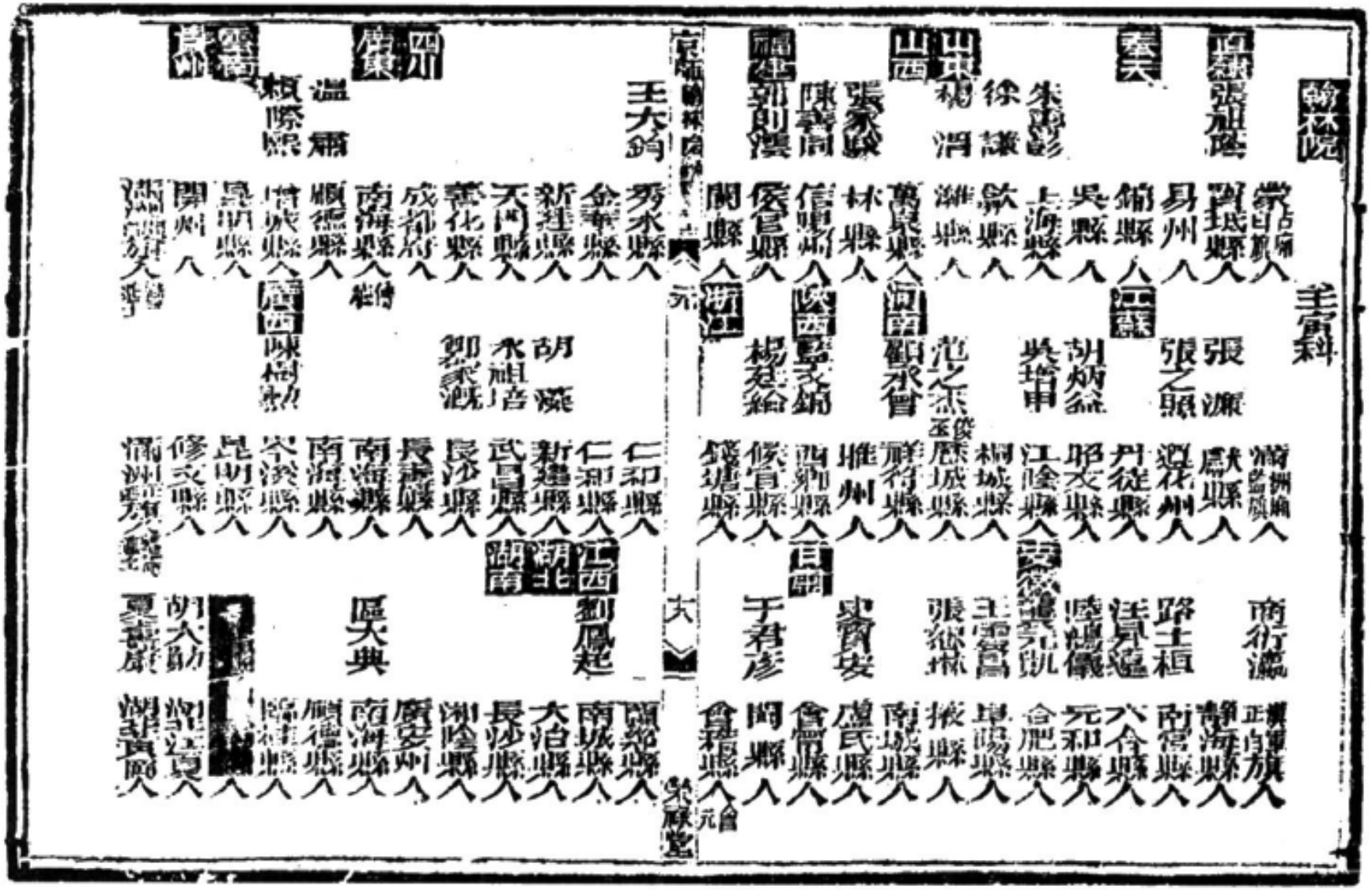
# 缙绅录主要内容

- 姓名
  - 民人： 姓名、字号
  - 旗人： 大部分只有名
- 来源省、来源府厅州县等
- 科名
  - 部分通过与CBDB中的清代进士题名录数据连接进行补充
- 八旗旗分
- 满洲、蒙古、汉军
- 机构，官职
- 品级
  - 只有一小部分官职名称中包括了品级，但大部分情况中我们根据《清会典》中的清代官员官职品级来确定官职的相应品级。
- 铨选方式、铨选年代
- 其他内容包括：加级、地域要紧、冲繁疲难程度等

# Contents

- Surname, given name and appellation.
  - Most Manchus and other non-Han only have a given name.
- Province and county of origin (for Han)
- Examination degree
- Banner affiliation, ethnicity
- Government office and position
- Bureaucratic rank (pinji 品级) 1-9
  - 1 was highest, and 9 the lowest.
  - Salaries of the highest rank were 6 times that of the lowest rank.
  - Year of appointment for current position
  - Sometimes appointment authority (Ministry of Personnel)
- Details on the posting, including ratings for difficulty and sensitivity.

# 《清代缙绅录集成》实例



# 已录入数据格式

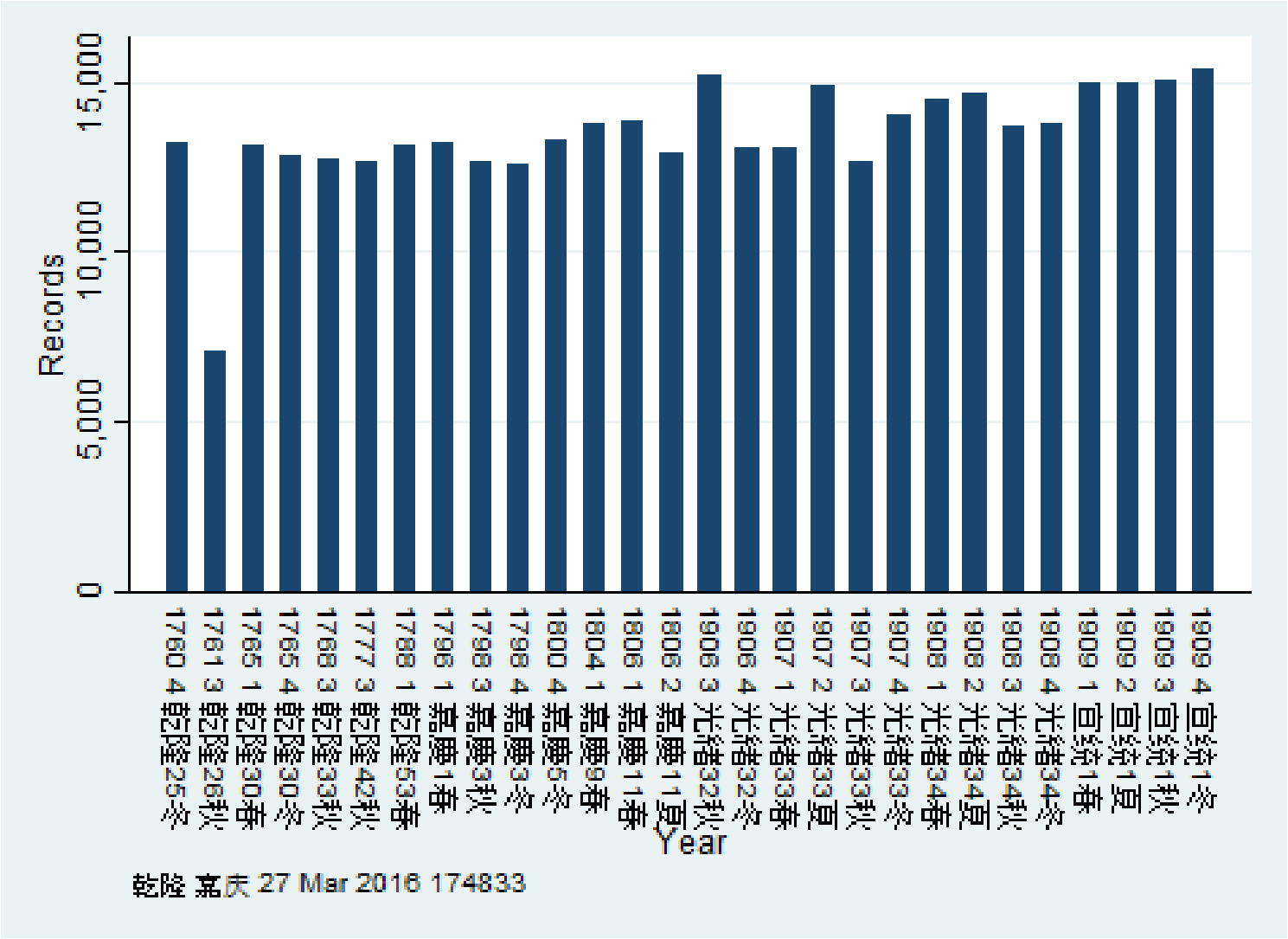
| | L | N | O | S | U | V | Y | Z | AA | AD | AE | AF | AG | AH | AI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | diqu | yamen | erji_jigou | guanzhi | guanxian_pinan_chengdu | jiaji | xing | ming | zihao | laiyuan_she | laiyuan_xia | minzu | qifen | keju_1 | keju_nian_ |
| 61 | 京師 | 內閣衙門 | 學士 | 內閣學士兼禮部侍郎 | | 3 | | ？保住 | | | | 滿洲 | 正藍旗 | | |
| 62 | 京師 | 內閣衙門 | 學士 | 內閣學士兼禮部侍郎 | | 3 | | 保成 | | | | 滿洲 | 鑲紅旗 | | |
| 63 | 京師 | 內閣衙門 | 學士 | 內閣學士兼禮部侍郎管理太常寺事務 | | 3 | | 紮郎阿 | | | | 滿洲 | 正紅旗 | | |
| 64 | 京師 | 內閣衙門 | 學士 | 內閣學士兼禮部侍郎文淵閣直閣事 | | 3 | | 阿肅 | 雨齋 | | | 滿洲 | 廂白旗 | | 甲戌 |
| 65 | 京師 | 內閣衙門 | 學士 | 內閣學士兼禮部侍郎公中佐領 | | 3 | | 依蘭泰 | ？芳 | | | 滿洲 | 廂紅旗 | 進士 | |
| 66 | 京師 | 內閣衙門 | 學士 | 內閣學士兼禮部侍郎 | | 3 | 管 | 幹珍 | 松? | 江蘇 | 陽? | | | | 丙戌 |
| 67 | 京師 | 內閣衙門 | 學士 | 內閣學士兼禮部侍郎 | | 3 | 張 | 若? | ？？ | 安徽 | 桐城 | | | 監生 | |
| 68 | 京師 | 內閣衙門 | 學士 | 內閣學士兼禮部侍郎管理？部事務 | | 3 | 鄒 | 奕孝 | | 江蘇 | 金匱 | | | ？？ | 丁醜 |
| 69 | 京師 | 內閣衙門 | 學士 | 內閣學士兼禮部侍郎文淵閣直閣事 | | 3 | 胡 | 高望 | 豫堂 | 浙江 | 仁和 | | | | 辛巳 |
| 70 | 京師 | 內閣衙門 | 侍讀學士 | 內閣侍讀學士 | | 3 | | 明祿 | | | | 滿洲 | 正紅旗 | | |
| 71 | 京師 | 內閣衙門 | 侍讀學士 | 內閣侍讀學士 | | 3 | | 善寶 | | | | 滿洲 | 廂黃旗 | | |
| 72 | 京師 | 內閣衙門 | 侍讀學士 | 內閣侍讀學士 | | 3 | | 富崑 | | | | 滿洲 | 廂藍旗 | | |
| 73 | 京師 | 內閣衙門 | 侍讀學士 | 內閣侍讀學士 | | 3 | | 武爾？ | | | | 滿洲 | 廂紅旗 | | |
| 74 | 京師 | 內閣衙門 | 侍讀學士 | 內閣侍讀學士 | | 3 | | 明泰 | | | | 蒙古 | 廂白旗 | | |
| 75 | 京師 | 內閣衙門 | 侍讀學士 | 內閣侍讀學士 | | 3 | | 五靈阿 | | | | 蒙古 | 正白旗 | | |
| 76 | 京師 | 內閣衙門 | 侍讀學士 | 內閣侍讀學士 | | 3 | 沈 | 鹹熙 | ？陔 | 浙江 | 歸安 | | | | 庚辰 |
| 77 | 京師 | 內閣衙門 | 侍讀學士 | 內閣侍讀學士 | | 3 | 趙 | 文興 | | | | 漢軍 | 正黃旗 | | |
| 78 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 1 | | 傅森 | 春圃 | | | 滿洲 | 鑲黃旗 | 監生 | |
| 79 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 2 | | 富勒赫 | | | | 滿洲 | 正黃旗 | | |
| 80 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 1 | | 保明 | | | | 滿洲 | 正白旗 | | |
| 81 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 1 | | 紮拉芬 | | | | 滿洲 | 廂黃旗 | | |
| 82 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 1 | | 德寧 | | | | 滿洲 | 鑲紅旗 | 生員 | |
| 83 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 1 | | 雙慶 | | | | 滿洲 | 正紅旗 | | |
| 84 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 1 | | 揆交 | | | | 滿洲 | 正白旗 | 生員 | |
| 85 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 3 | | 官亮 | | | | 滿洲 | 正藍旗 | 生員 | |
| 86 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 4 | | 那淇 | 竹軒 | | | 滿洲 | 鑲紅旗 | | |
| 87 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 1 | | 富克進 | | | | 滿洲 | 廂藍旗 | 生員 | |
| 88 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 1 | | 色布星額 | | | | 蒙古 | 鑲紅旗 | | |
| 89 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 4 | | 色克通額 | | | | 滿洲 | 廂黃旗 | 生員 | |
| 90 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 1 | 董 | 斌 | | | | 漢軍 | 正黃旗 | 舉人 | |
| 91 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 2 | 張 | 紹孟 | | | | 滿洲 | 正黃旗 | 生員 | |
| 92 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 1 | 方 | 大川 | | 安徽 | 歙縣 | | | | 辛卯 |
| 93 | 京師 | 內閣衙門 | 侍讀 | 侍讀 | | 2 | | | | 江蘇 | 吳縣 | | | | 乙未 |
| 94 | 京師 | 內閣衙門 | 典籍廳 | 典籍 | | 1 | | 明秀 | | | | 滿洲 | 廂黃旗 | | |
| 95 | 京師 | 內閣衙門 | 典籍廳 | 典籍 | | 2 | | 薩炳阿 | | | | 滿洲 | 正黃旗 | 生員 | |

# 清代缙绅录集成数据库录入进度

| 公元年 | 春 | 夏 | 秋 | 冬 | 公元年 | 春 | 夏 | 秋 | 冬 | 公元年 | 春 | 夏 | 秋 | 冬 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1726 | | | | | 1840 | | | | | 1881 | | | | |
| 1748 | | | | | 1842 | | | | | 1882 | | | | |
| 1760 | | | | | 1845 | | | | | 1884 | | | | |
| 1761 | | | | | 1846 | | | | | 1885 | | | | |
| 1765 | | | | | 1847 | | | | | 1886 | | | | |
| 1766 | | | | | 1848 | | | | | 1887 | | | | |
| 1768 | | | | | 1849 | | | | | 1888 | | | | |
| 1777 | | | | | 1851 | | | | | 1889 | | | | |
| 1788 | | | | | 1852 | | | | | 1890 | | | | |
| 1796 | | | | | 1853 | | | | | 1892 | | | | |
| 1797 | | | | | 1854 | | | | | 1893 | | | | |
| 1798 | | | | | 1856 | | | | | 1894 | | | | |
| 1800 | | | | | 1857 | | | | | 1895 | | | | |
| 1804 | | | | | 1858 | | | | | 1896 | | | | |
| 1806 | | | | | 1859 | | | | | 1897 | | | | |
| 1812 | | | | | 1860 | | | | | 1898 | | | | |
| 1814 | | | | | 1865 | | | | | 1899 | | | | |
| 1816 | | | | | 1866 | | | | | 1900 | | | | |
| 1817 | | | | | 1867 | | | | | 1901 | | | | |
| 1820 | | | | | 1869 | | | | | 1902 | | | | |
| 1824 | | | | | 1870 | | | | | 1903 | | | | |
| 1826 | | | | | 1871 | | | | | 1904 | | | | |
| 1827 | | | | | 1872 | | | | | 1905 | | | | |
| 1830 | | | | | 1873 | | | | | 1906 | | | | |
| 1833 | | | | | 1874 | | | | | 1907 | | | | |
| 1834 | | | | | 1875 | | | | | 1908 | | | | |
| 1836 | | | | | 1876 | | | | | 1909 | | | | |
| 1837 | | | | | 1877 | | | | | 1910 | | | | |
| 1838 | | | | | 1878 | | | | | 1911 | | | | |
| 1839 | | | | | 1879 | | | | | 1912 | | | | |

已录入　　录入中　　存在但未录入

# 本报告分析数据


Records
0
5,000
10,000
15,000
1760 4 乾隆25冬
1761 3 乾隆26秋
1765 1 乾隆30春
1765 4 乾隆30冬
1768 3 乾隆33秋
1777 3 乾隆42秋
1788 1 乾隆53春
1798 1 嘉庆1春
1798 3 嘉庆3秋
1798 4 嘉庆3冬
1800 4 嘉庆5冬
1804 1 嘉庆9春
1806 1 嘉庆11春
1806 2 嘉庆11夏
1906 3 光绪32秋
1906 4 光绪32冬
1907 1 光绪33春
1907 2 光绪33夏
1907 3 光绪33秋
1907 4 光绪33冬
1908 1 光绪34春
1908 2 光绪34夏
1908 3 光绪34秋
1908 4 光绪34冬
1909 1 宣统1春
1909 2 宣统1夏
1909 3 宣统1秋
1909 4 宣统1冬
Year
乾隆 嘉庆 27 Mar 2016 174833

# From the beginning

- For data entry, *use paid labor.*
  - Ensure an explicit understanding that you control the data.
  - Conducting data entry ‘on the cheap’ by having students or collaborators do it for academic credit or publications may lead to trouble later.
- Data entry should seek to transcribe the raw data as faithfully as possible.
  - Don’t ask data entry personnel to make judgment calls, or ‘correct’ obviously wrong data.
  - Data cleaning should come later, and should be a separate stage.
  - Distinguish between a raw file and a ‘clean’ work file
- When constructing datasets, be wary of using custom or proprietary software.
  - As much as possible, stay with open standards, and off-the-shelf software platforms.
  - Generally, keep it simple.
- If you share data with collaborators or students, make sure you have clear, preferably written understandings about what they can do with the data.
- When acquiring data, play by the rules. Papers using data acquired in an irregular fashion, for example via personal connections, may be unpublishable.
- Secure your data!

# Document everything

- Pay attention to field names in databases
  - Avoid cryptic field names. You won't remember what they are 6 months later.
- Preserve and document code
  - If you hire a programmer, make sure that they are documenting their code. If they leave, and you have thousands of lines of undocumented code, it may be easier for a new programmer to start from scratch than try to 'fix' existing, undocumented code.
- Document decisions made during data cleaning and preparation of a 'clean' work file.
  - Archive email and minutes of meetings.

# Venues for releasing your data

- Posting your data on your personal website is only a short-term solution
- Many universities, including HKUST, have online repositories.
  - http://repository.ust.hk/ir/
- Discipline-specific repositories
  - For example, ICPSR for social science https://www.icpsr.umich.edu/icpsrweb/landing.jsp
- Online appendices at journals.


Home ▾ Scholar Profiles

Search Publications: campbell All Fields Find Advanced ☑ Retain my current filters

## Narrow Search

☐ Has open access documents

REMOVE FILTERS

✖ Format: Dataset

FORMAT ▲

Dataset ✔

AUTHOR ▲

Campbell, Cameron Dougall 2
Lee, James 2
Campbell, Cameron D 1
Dong, Hao 1
Lee, James Z 1

SUBJECT ▲

Household Register 1
Korea 1
Longitudinal historical population data 1
Tansung 1

JOURNAL ▲

China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN), 1749-1909 1
DataSpace@HKUST 1

Showing **1** - **2** of **2** (0.03 seconds)

Export **0** items to: ◉ Email ○ CSV ○ Refworks Go

Sort by Relevance

☐ Longitudinal Links to Construct the Korean Multi-Generational Panel Dataset – Tansung (KMGPD-TS) from the Tansung Household Registers
Author(s): Campbell, Cameron Dougall ; Lee, James ; Dong, Hao
Source: DataSpace@HKUST,
Dataset, 2015

☐ China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN), 1749-1909
Author(s): Lee, James Z. ; Campbell, Cameron D.
Source: China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN), 1749-1909, Ann Arbor, MI: Inter-university Consortium for Political and Social Research, 2014-07-10, ICPSR27063-v10
Dataset, 2014

# Preparing for release

- If you construct databases by transcribing the contents of images, you should be able to release the database you construct, but may not be free to share the original images.
- If your dataset includes information on living individuals and was not constructed from publicly available sources, you must consider confidentiality and privacy.
  - De-identification can be difficult. It is extremely difficult to anonymize data on individuals.
- If privacy and confidentiality are an issue, you should disseminate your data through a repository that has relevant experience.
  - ICPSR, for example, can handle the release sensitive data, and work with institutions to ensure protocols are followed.

# Outreach

- Simply posting your data on the web may not attract users.
- You will need to provide documentation, especially if the data are complex.
- For the CMGPD, we conducted two-week workshops at Shanghai Jiaotong University in summer 2011, 2012, 2013 and 2014, as well as a workshop at UCLA in January 2016.
- We have also made presentations in other venues.


中国清代辽宁多代人口数据库暨社会人口学研究暑期学校(第二期)
上海交通大学中国东北历史与社会研究中心
2012.7.19