

Constructing Big Data for Japanese Historical Population: Challenges and Opportunities

Satomi KUROSU (Reitaku University)

and Hao DONG (Princeton University)

(in collaboration with M.Takahashi and A.Nagaoka)

*Mathematical and geographical modelling
for environmental humanities*

Kyoto, October 31, 2016



“It started 50 some years ago...**BDS** was an innovation”*Akira Hayami* (1929-), founder of historical demography in Japan

Outline:

1. A Long Way of Digitization: From Texts to Bits
2. Old and New Approaches of Data Base Construction
3. A New Framework of Analysis: EAP and Marriage
4. Challenges and Opportunities
 - Application of EAP model to other population
 - **Geographical mobility**



July 2015 in Mita, Tokyo

1. Long way to analysis

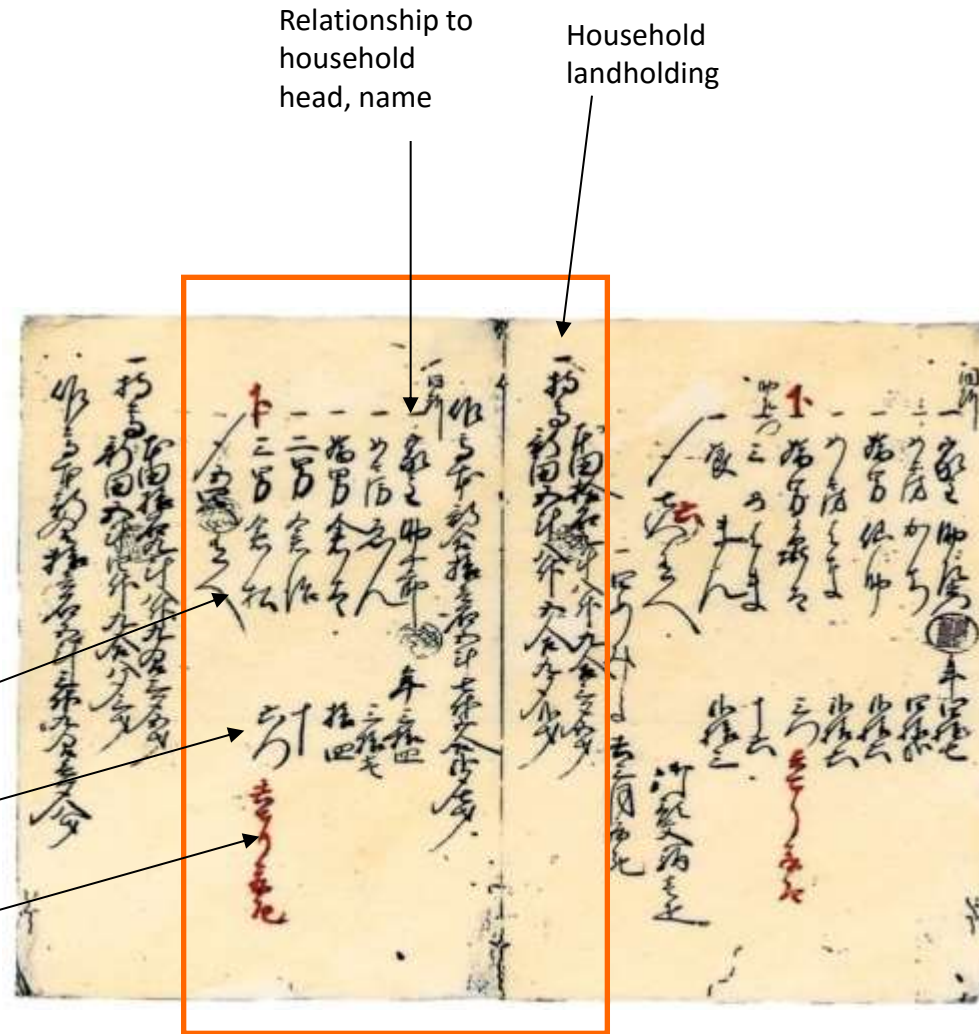
1.1 Original Sources

- ◆ ***Shumon-aratame-cho (SAC)***
Investigation of religious faith 1639-1871--- Christianity was considered dangerous and banned
- ◆ ***Ninbetsu-aratame-cho (NAC):***
Population registration/survey---
Similar to SAC without religious inquiry
- ◆ ***Other***
Kako-cho (Temple registers)
Kainin-kakiage-cho (Pregnancy registers)
Village/regional documents
---numerous!



A scene from religious investigation, *shumon-aratame* (from Siebold, *Japan*)

1.2 Example of Population Registers (NAC)



出典：『江戸時代の東北農村 二本松藩仁井田村』（成松佐恵子 1992）

1.3 BDS

(Basic Data Sheet = time schedule of individual lives and her/his household)

Grandfather =household head "Naoemon"
 Mother "Seki"
 Father "Naosaku"

Transcribing Coding ...and annual information linked manually onto BDS

町	名	性別	戸の主統と稱	年																								計	馬	持高石	順序
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24				
	直右門	男	妻	32	30	4																									
	孝	男	子	33	31	5																									
	也	女	子	34	32	6	2																								
	く	女	子	35	33	7	3																								
	直	男	智	36	34	8	4																								
	作	男	智	37	35	9	5	2																							
	通	男	智	38	36	10	6	3																							
	太	男	孫	39	37	11	7	4																							
	中	男	孫	40	38	12	8	5																							
	喜	男	孫	41	39	13	9	6																							
	兵	男	孫	42	40	14	10	7																							
	七	男	孫	43	41	15	11	8	20																						
	門	男	孫	44	42	16	12	9	21																						
	直右門	男	妻	45	43	17	13	10	12																						
	孝	男	子	46	44	18	14	11																							
	也	女	子	47	45	19	15	12	21																						
	く	女	子	48	46	20	16	13	22																						
	直	男	智	49	47	21	17	14	23																						
	作	男	智	50	48	22	18	15	24																						
	通	男	智	51	49	23	19	16	25																						
	太	男	孫	53	51	25		17																							
	中	男	孫																												
	喜	男	孫																												
	兵	男	孫																												
	七	男	孫																												
	門	男	孫																												
鬼				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				
國分																															

Name
 Sex
 Relationship to household head

Horse Landholding

Events and notations

1.4 Data entry using *Shumon-data-entry program*

BDS

操作 入力方法 表示 世帯 世帯員 作成 抽出 検索 置換

史料年 嘉永 5年(1852年)壬子 家番号 1:忠吉

	1	1	1	1	1	1	1	1	1	1	1	1	1	
	-001	-002	-003	-004	-005	-006	-007	-008	-009	-010	-011	-012	-013	-014
	忠吉	たか	忠吉	半重	かん	ふり	忠右衛門	忠太郎	けん	甚吉	三治郎	宗助	ゆう	きゑ
	家主	忠吉女房	忠吉嫡男	忠吉二男	忠吉母	たか養女	ふり入夫	忠右衛門嫡男	忠右衛門二女	忠右衛門三男	忠吉弟	忠吉弟	下女	下女
嘉永 5年(1852年)壬子	*26	26	5	3	65	29	29	10	5	2B	15	13	14E	16E
嘉永 6年(1853年)癸丑	*27	27	6											
嘉永 7年(1854年)甲寅	*28	28	7											
安政 2年(1855年)乙卯	*29	29	8											
安政 3年(1856年)丙辰	*30	30	9											
安政 4年(1857年)丁巳	*31	31	10											
安政 5年(1858年)戊午	*32	32	11											
安政 6年(1859年)己未	*33	33	12											
安政 7年(1860年)庚申	*34	34	13											
万延 2年(1861年)辛酉	*35													
文久 2年(1862年)壬戌	*36													
文久 3年(1863年)癸亥	*37													
文久 4年(1864年)甲子	*38													
元治 2年(1865年)乙丑	*39													
慶応 2年(1866年)丙寅														
慶応 3年(1867年)丁卯														
慶応 4年(1868年)戊辰														
明治 2年(1869年)己巳														
明治 3年(1870年)庚午														

新世帯員
世帯員別家
世帯員追加
世帯員消去
番号入力
続柄入力
名前入力
年齢入力
地文直置
追跡

番号 続柄 名前 地文

非表示

Digitalized BDS---
household/individual information
for the duration of observation

陸奥国安藝郡郡山上町

操作 入力方法 表示 世帯 世帯員 作成 抽出 検索 置換

史料年 嘉永 5年(1852年)壬子 家番号 1:忠吉

村役 持高

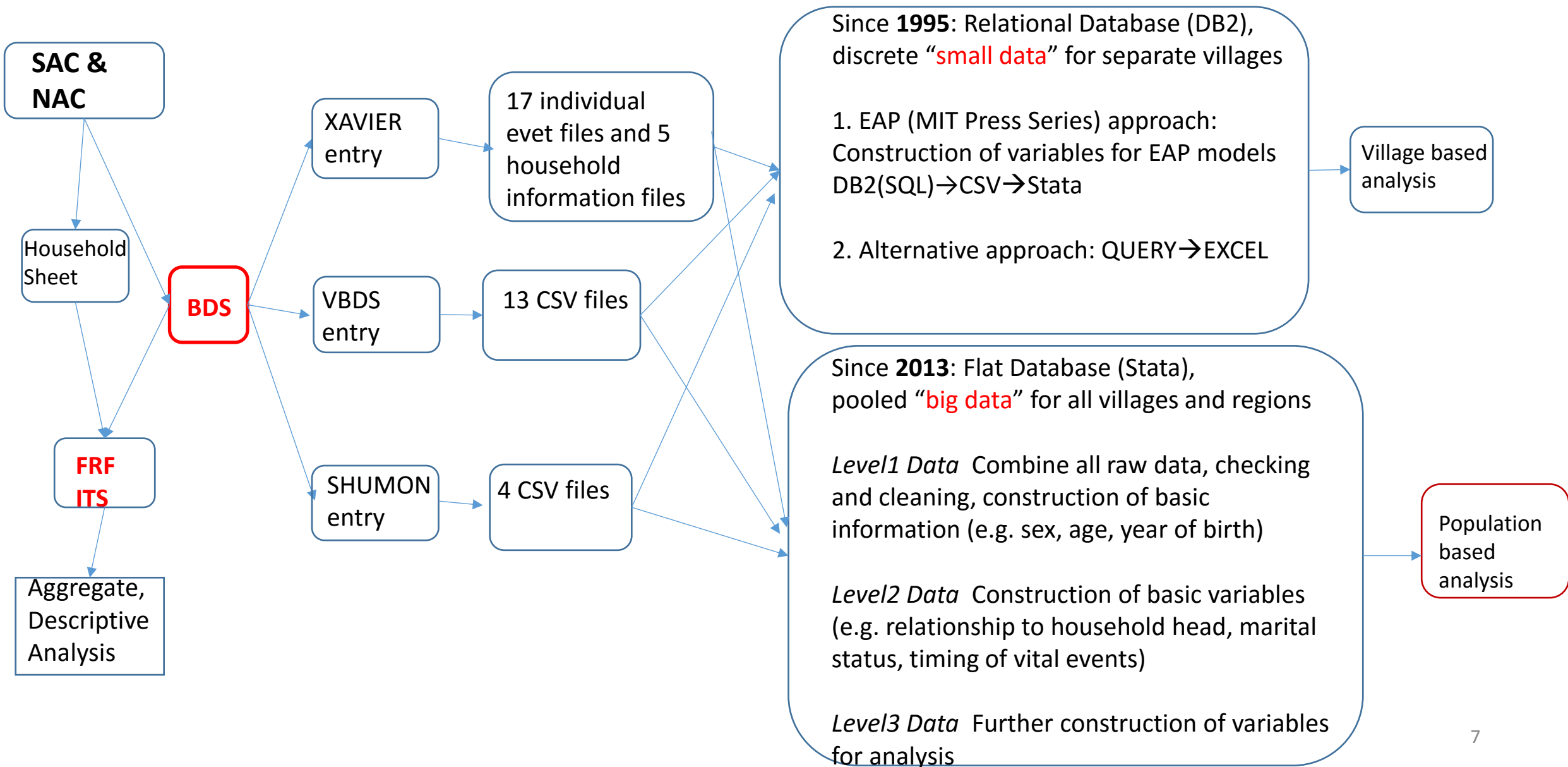
	個人番号	続柄	名前	年齢	地文
1	1-001	家主	忠吉	26	
2	1-002	忠吉女房	たか	26	
3	1-003	忠吉嫡男	忠吉	5	
4	1-004	忠吉二男	半重	3	
5	1-005	忠吉母	かん	65	
6	1-006	たか養女	ふり	29	
7	1-007	ふり入夫	忠右衛門	29	
8	1-008	忠右衛門嫡男	忠太郎	10	
9	1-009	忠右衛門二女	けん	5	
10	1-010	忠右衛門三男	甚吉	2	去6月、出生
11	1-011	忠吉弟	三治郎	15	
12	1-012	忠吉弟	宗助	13	
13	1-013	下女	ゆう	14	西より、越後国蒲原郡新潟古町与喜蔵、
14	1-014	下女	きゑ	16	当下町九蔵、娘ヨリ 質物

追記 馬二匹

14 人内 男 8 人
女 6 人

Digitalized SAC/NAC ---annual
information by household

2 Old and New Approaches of Data Base Construction: From Village-Based to Population Based Analysis



Separate CSV Datasheets, Relational SQL Database, and Flat STATA Dataset

- Separate CSV Datasheets
 - Each file contains information of a specific aspect and/or for analytical units of a specific level
 - For example, each village in Shumon data include separate csv datasheets for household time-variant information, individual time-variant information, vital events, etc., respectively.
- Relational SQL Database
 - A database includes all separate datasheets linked by common primary identifiers
 - Low efficiency and high demand in data management (e.g. variable construction, data restrictions, etc.).
 - Data management and analysis are separate
- Flat STATA Dataset
 - A unified data file integrates complete information contained by separate datasheets
 - Straightforward and relatively efficient for data management
 - Data management and analysis are in one flow.

Steps of database construction using DB2 & STATA

- 1) Coding (categorizing) events using Excel
- 2) Importing files (containing time-variant & time-constant info. at individual, household & community levels, and also events) to DB2
- 3) Construction of 'tables' using DB2 (with commands written in SQL), that contain variables to measure risks of events, individual & household characteristics, and kinship relationships
- 4) Exporting 'tables' to STATA
- 5) Merging tables using STATA, and construction of additional variables with STATA

- Name
- Takasago_Aiyamachi_jibun.xlsx
- Takasago_Aiyamachi_households.xlsx
- Minamisugita_vil_tv.xlsx
- Minamisugita_persons.xlsx
- Yukinobu_vil_tv.xlsx
- Yukinobu_persons.xlsx
- Yukinobu_jibun.xlsx
- Yukinobu_households.xlsx
- Yanbe_vil_tv.xlsx
- Yanbe_persons.xlsx

Nishijyo_households

Home Insert Page Layout Formulas Data Review View

Calibri (Body) 14

fx vil_name

A	B	C	D	E	F	G	H	I	J
vil_name	year	h_order	hhid	murayaku	land	etc	total	male	female
1	西条	1773	10001	1	百姓代	48592合	馬1頭	11	7
2	西条	1773	10002	2	0合	権兵衛家		4	0
3	西条	1773	10003	3	0合		5	2	0
4	西条	1773	10004	4	0合			2	0
5	西条	1773	10005	5	0合			5	1
6	西条	1773	10006	6	33435合			7	0
7	西条	1773	10007	7	8284合			4	0
8	西条	1773	10008	8	0合			5	0
9	西条	1773	10009	9	2364合			1	0
10	西条	1773	10010	10	0合			4	0
11	西条	1773	10011	11	0合			5	2
12	西条	1773	10012	12	百姓代	6744合		4	0
13	西条	1773	10013	13	0合			1	0
14	西条	1773	10014	14	0合			2	0
15	西条	1773	10015	15	0合			4	0
16	西条	1773	10016	16	0合			1	0
17	西条	1773	10017	17	0合			1	0
18	西条	1773	10018	18	0合			1	0

households +

Nishijyo_persons

Home Insert Page Layout Formulas Data Review View

Normal Page Layout Custom Views

fx vil_name

A	B	C	D	E	F	G	H	I	J	K	L	M	N
vil_name	year	hhid	p_order	ego	sex	shuha	danna	ref	rel	name	age	presence	
1	西条	1773	1	1 1-001	男子	禪宗	心岩院		戸主	権兵衛	38	UCHI	
2	西条	1773	1	2 1-002	女子	禪宗	心岩院		女房	女房	23	UCHI	
3	西条	1773	1	3 1-003	男子	禪宗	心岩院		父	浄確	70	UCHI	
4	西条	1773	1	4 1-005	女子	禪宗	心岩院		母	ふう	57	UCHI	
5	西条	1773	1	5 1-006	男子	禪宗	心岩院		弟	伊予八	27	UCHI	
6	西条	1773	1	6 1-007	男子	禪宗	心岩院		弟	儀助	21	UCHI	
7	西条	1773	1	7 1-008	女子	禪宗	心岩院		妹	きの	15	UCHI	
8	西条	1773	1	8 2-005	男子	禪宗	心岩院		下男	又助	44	UCHI	
9	西条	1773	1	9 86-004	男子	禪宗	心岩院		下男	免蔵	29	UCHI	
10	西条	1773	1	10 1-009	男子	禪宗	心岩院		下男	丹七	27	UCHI	
11	西条	1773	1	11 1-012	女子	禪宗	心岩院		下女	はる	20	UCHI	
12	西条	1773	1	12 1-501	女子	禪宗	心岩院		妹	ふさ	36	SOTO	
13	西条	1773	1	13 1-502	男子	禪宗	心岩院		弟	傳左衛門	33	SOTO	
14	西条	1773	1	14 1-503	男子	禪宗	心岩院		弟	租桂	18	SOTO	
15	西条	1773	1	15 1-004	男子	禪宗	心岩院			岩吉	0	SHU	
16	西条	1773	1	16 1-013	男子	禪宗	心岩院			谷助	0	SHU	
17	西条	1773	1										
18	西条	1773	1										

Nishijyo_jibun

Home Insert Page Layout Formulas Data Review View

fx vil_name

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
vil_name	year	hhid	ego	e_order	jyear	jmonth	jdate	date	place	house	dir	etc	event	typ	
1	西条	1773	1 1-002	1	1764	0	0	9年以前	安八郡里	李兵衛娘			縁付来	#T	
2	西条	1773	1 1-006	51	1773	99	0						別家	#F	
3	西条	1773	1 2-005	1	1773	0	0			源八伴			一年季	#F	
4	西条	1773	1 2-005	52	1773	99	0						奉公引	#F	
5	西条	1773	1 86-004	1	1773	0	0		当村	銀七伴			一年季	#F	
6	西条	1773	1 86-004	52	1773	99	0						奉公引	#T	
7	西条	1773	1 1-009	1	1773	0	0		楡又村	与平次伴			一年季	#T	
8	西条	1773	1 1-009	52	1773	99	0						奉公引	#F	
9	西条	1773	1 1-012	1	1773	0	0		里村	小助娘			奉公	#T	
10	西条	1773	1 1-012	52	1773	99	0						奉公引	#F	
11	西条	1773	1 1-501	1	1771	0	0	2年以前	楡又村	半左衛門	へ		縁付	#T	
12	西条	1773	1 1-502	1	1757	0	0	16年以前	安八郡大	伊藤作之丞殿			養子二	#T	
13	西条	1773	1 1-503	1	1773	0	0			尾州稲葉宿			弟子	#T	
14	西条	1773	1 1-004	51	1773	99	0						出生	#F	
15	西条	1773	1 1-013	51	1773	99	0						一年季	#F	
16	西条	1773	2 2-002	1	1765	0	0	1765年	安八郡中村				縁付来	#T	
17	西条	1773	2 2-005	1	1773	0	0		安八郡水	左右衛門娘			一年季	#T	
18	西条	1773	2 2-005	52	1773	99	0		当村	権兵衛		通い	縁付来	#T	
19	西条	1773	2 2-005	52	1773	99	0		当村				奉公引	#F	

Size	Kind
89 KB	Micros... (.xlsx)
65 KB	Micros... (.xlsx)
12 KB	Micros... (.xslm)
2.8 MB	Micros... (.xlsx)
11 KB	Micros... (.xlsx)
640 KB	Micros... (.xlsx)
72 KB	Micros... (.xlsx)
108 KB	Micros... (.xlsx)
11 KB	Micros... (.xlsx)
4.4 MB	Micros... (.xlsx)
661 KB	Micros... (.xlsx)
497 KB	Micros... (.xlsx)
10 KB	Micros... (.xlsx)
206 KB	Micros... (.xlsx)
10 KB	Micros... (.xlsx)
1.4 MB	Micros... (.xlsx)
210 KB	Micros... (.xlsx)
167 KB	Micros... (.xlsx)
11 KB	Micros... (.xlsx)
913 KB	Micros... (.xlsx)
911 KB	Micros... (.xlsx)
145 KB	Micros... (.xlsx)
10 KB	Micros... (.xlsx)
1 MB	Micros... (.xlsx)
75 KB	Micros... (.xlsx)
167 KB	Micros... (.xlsx)
10 KB	Micros... (.xlsx)
881 KB	Micros... (.xlsx)
73 KB	Micros... (.xlsx)
108 KB	Micros... (.xlsx)
11 KB	Micros... (.xlsx)
3.3 MB	Micros... (.xlsx)
420 KB	Micros... (.xlsx)
294 KB	Micros... (.xlsx)
10 KB	Micros... (.xlsx)
304 KB	Micros... (.xlsx)
155 KB	Micros... (.xlsx)
4 KB	Micros... (.xlsx)
11 KB	Micros... (.xlsx)
1.3 MB	Micros... (.xlsx)
118 KB	Micros... (.xlsx)
195 KB	Micros... (.xlsx)
12 KB	Micros... (.xlsx)
4.8 MB	Micros... (.xlsx)
1.1 MB	Micros... (.xlsx)
706 KB	Micros... (.xlsx)
1.5 MB	Micros... (.xlsx)
265 KB	Micros... (.xlsx)
11 KB	Micros... (.xlsx)
11 KB	Micros... (.xlsx)
220 KB	Micros... (.xlsx)
7.5 MB	Micros... (.xlsx)
2.4 MB	Micros... (.xlsx)

Review

```
Command _rc
1 do "/Us...
2 use "/U...
3 brow
```

Results

```
-----+----- (R)
/-----+-----
/-----+----- 14.1 Copyright 1985-2015 StataCorp LP
Statistics/Data Analysis StataCorp
4905 Lakeway Drive
MP - Parallel Edition College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

Single-user 6-core Stata perpetual license:
Serial number: 501406266721
Licensed to: Population and Family History Project
Reitaku University

Notes:
1. Unicode is supported; see help unicode_advice.
2. More than 2 billion observations are allowed; see help obs_advice.
3. Maximum number of variables is set to 5000; see help set_maxvar.
4. New update available; type
```

```
. do "/Users/hdong/Dropbox/Shared
. clear all
. set more off
. capture log close
. program drop _all
. cd "/Users/hdong/Dropbox/Shared
/Users/hdong/Dropbox/Shared folder
. log using "Calculations for the
name: <unnamed>
log: /Users/hdong/Dropbox
log type: smcl
opened on: 21 Jul 2016, 18:20:59
```

```
Command
```

vil_name[2]	vil_name	year	hhid	p_order	ego	ref
98010	西条	1821	6	7		
98011	西条	1821	6	8		
98012	西条	1821	6	9		
98013	西条	1821	6	10		
98014	西条	1821	7	1		
98015	西条	1821	75	1		
98016	西条	1821	75	2		
98017	西条	1821	75	3		
98018	西条	1821	75	4		
98019	西条	1821	75	5		
98020	西条	1821	75	6		
98021	西条	1821	75	7		
98022	西条	1821	75	8	75-009	東本願寺 西福寺
98023	西条	1821	75	9	75-501	東本願寺 西福寺
98024	西条	1821	0	1	64-004	源之助
98025	西条	1821	0	2	64-005	源之助
98026	西条	1821	0	4	8-005	留吉
98027	西条	1821	0	5	8-004	留吉
98028	西条	1821	0	6	11-007	
98029	西条	1821	0	7	11-006	はつ
98030	西条	1821	0	8	11-010	はつ
98031	西条	1821	0	9	11-008	
98032	西条	1821	0	10	10-009	彦七

data management_Shumon

```
gen notdone = 1
save `notdone', replace

foreach i of local vil_name{
    capture confirm file "/Users/hdong/Dropbox/Shumon relation coding check/'i' for rel
    if _rc != 0 {
        use Shumon_persons, clear
        keep if vil_name == "'i'"
        keep vil_name
        keep if _n == 1
        gen notdone = 1
        save `one', replace
        use `notdone', clear
        append using `one'
        display "'i'"
        save `notdone', replace
    }
    else {
        tempfile 'i'
        import excel using "/Users/hdong/Dropbox/Shumon relation coding check/'i' for r
        gen relcode = relcode_c
        replace relcode = relcode_m if relcode_m != ""
        keep rel relcode
        save `one', replace
        use `codebook', clear
        append using `one'
        drop if rel == ""
        bysort rel: keep if _n == 1
        display "'i'"
        save `codebook', replace
    }
}

use `codebook', clear
codebook rel relcode
export excel using "/Users/hdong/Dropbox/Shumon relation coding check/codebook.xlsx", f

use Shumon_persons, clear

merge m:1 rel using `codebook', nogenerate

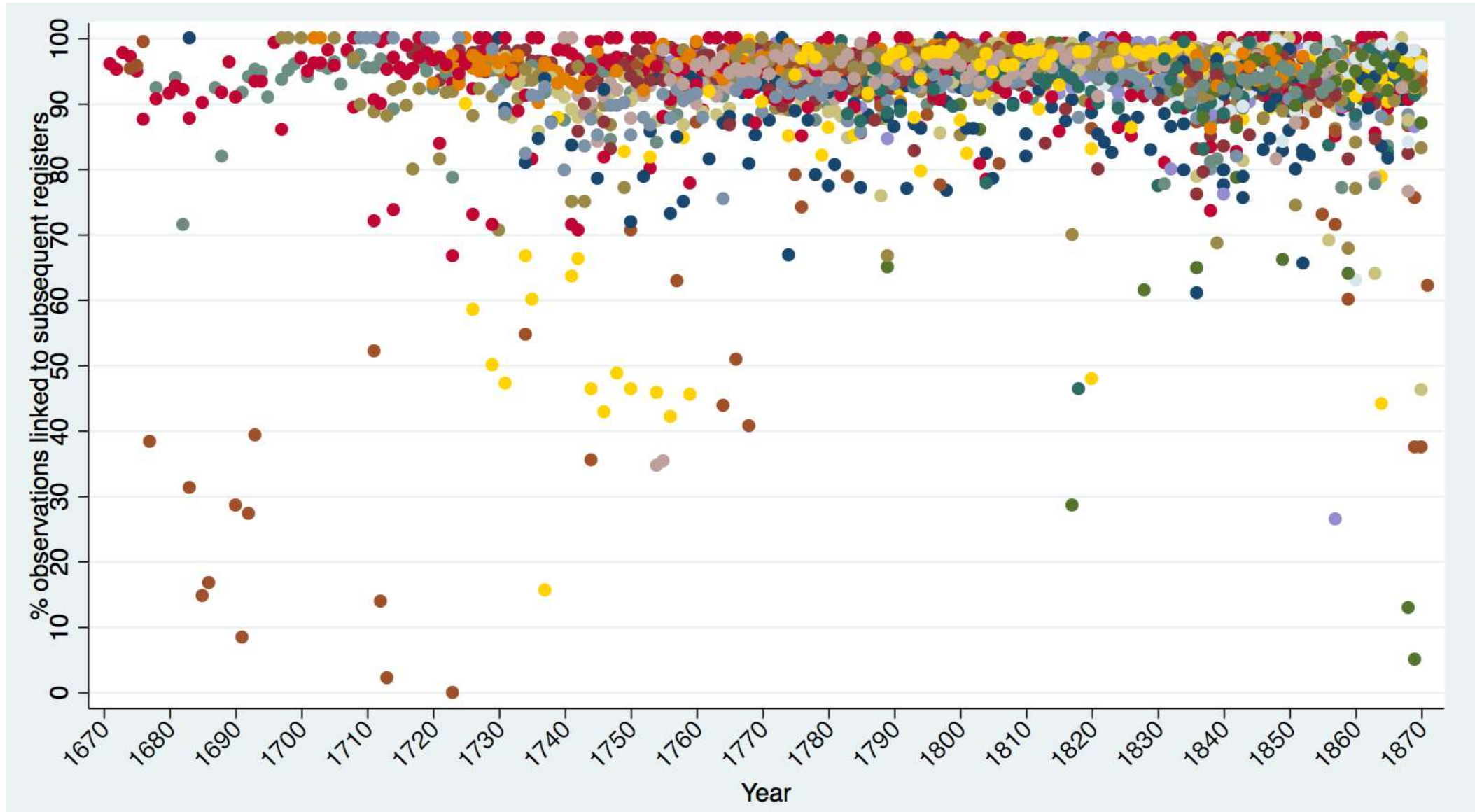
ren relcode relcode_c
replace relcode_c = "e" if rel == ""
replace relcode_c = "r" if rel == "" & (age == 0 | presence == "S0T0" | presence == "SH
```

Variables

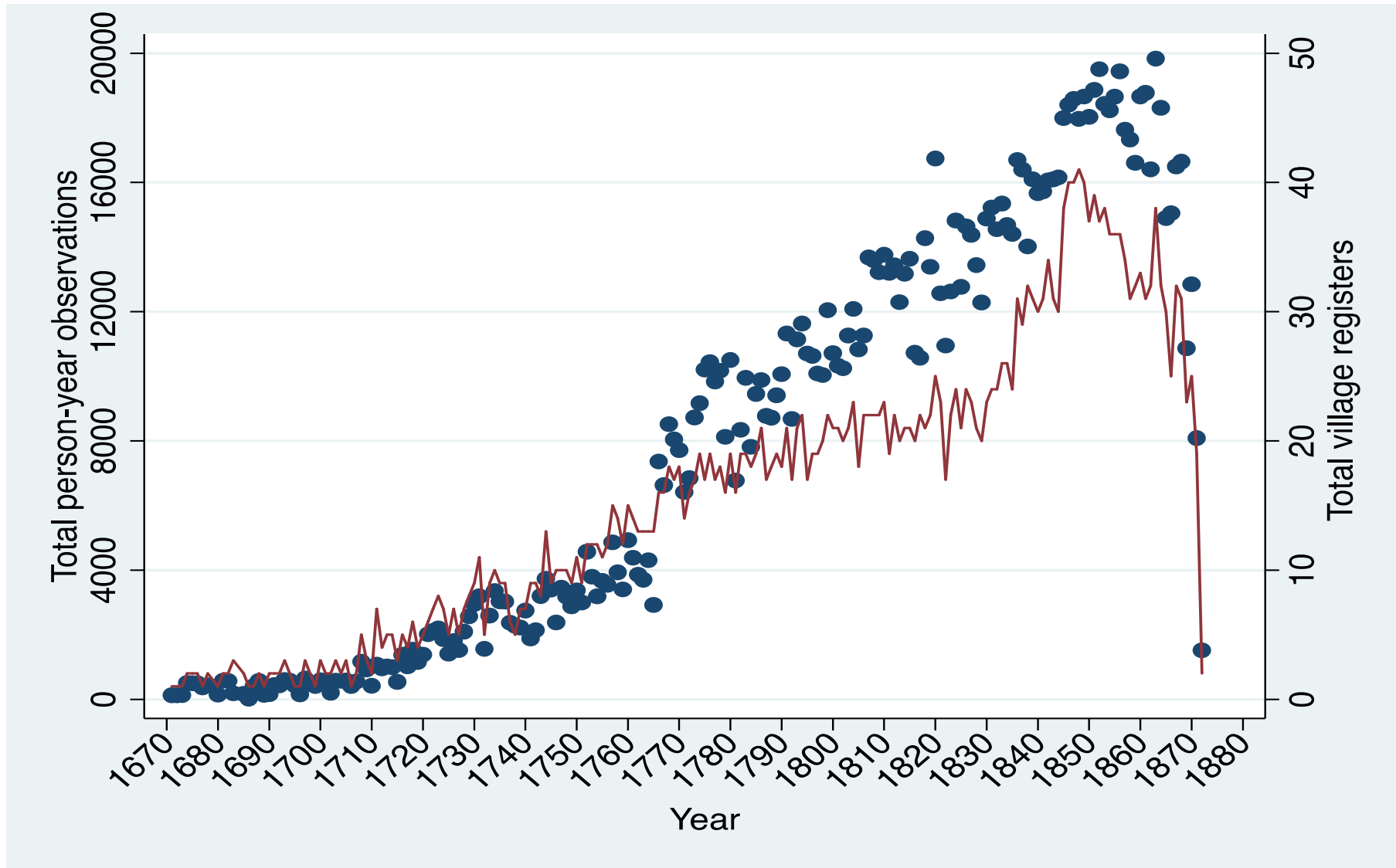
Name	Label
vil_name	vil_name
year	year
hhid	hhid
p_order	p_order
ego	ego
sex_old	sex
sex	sex2
shuha	shuha
danna	danna
ref	ref
rel	rel
name	name
age	age
presence	presence
person_id	group(vil...
household_id	group(vil...
present	

Proportion of observations linked to subsequent registers

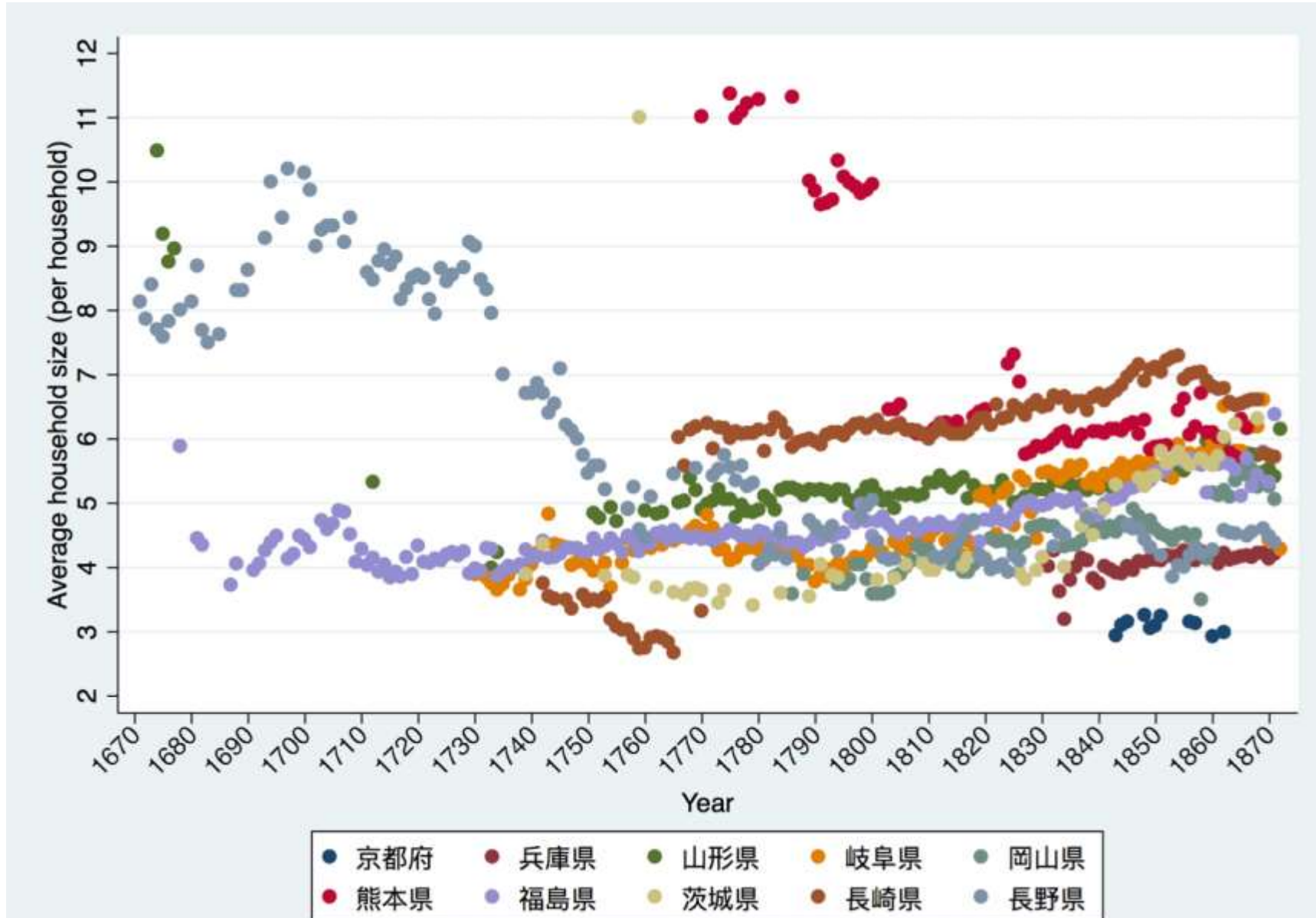
100 Thousand Lives in 200 Years



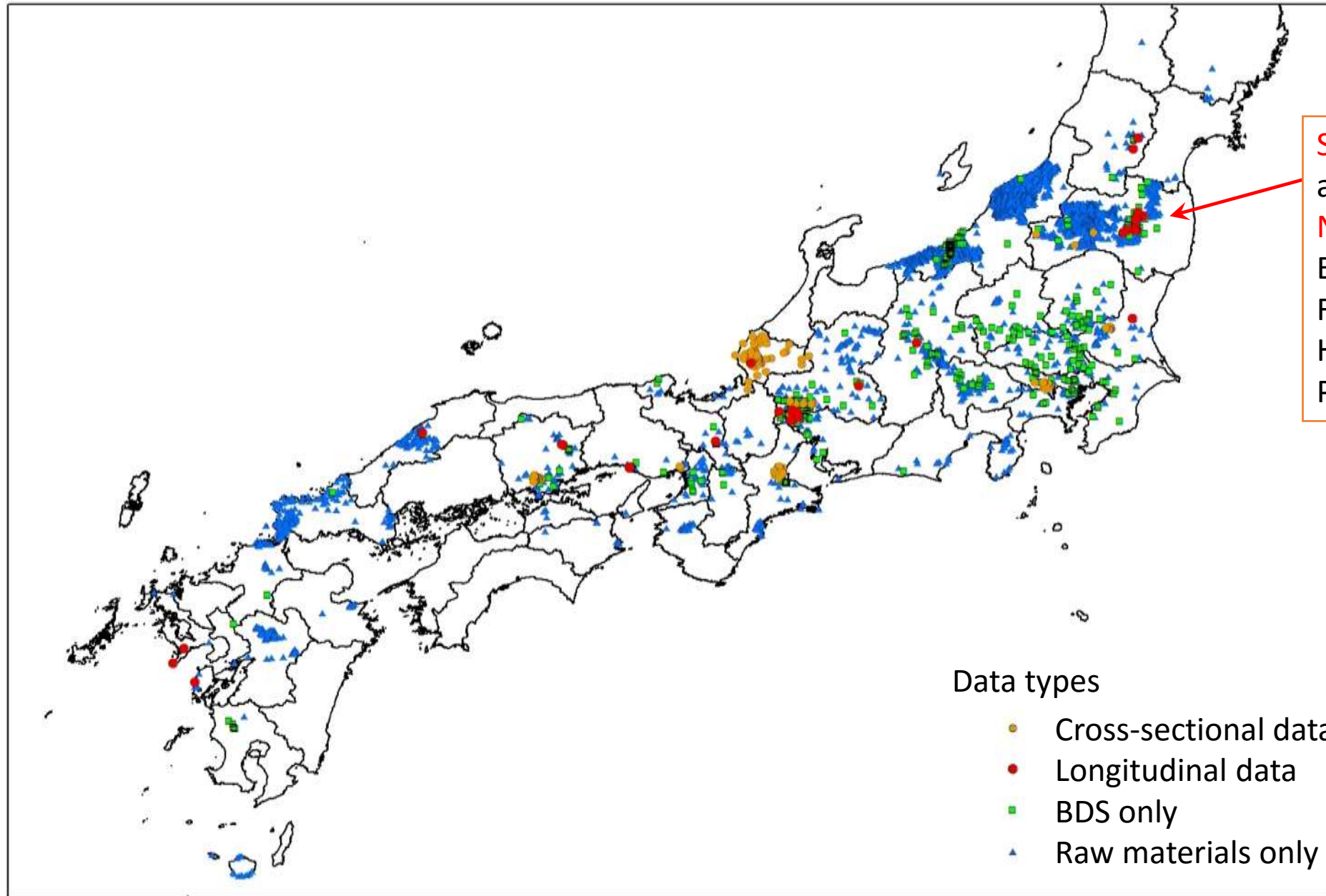
Total person-year observations for 52 villages, 1670-1870



Average size of household by current prefecture: 1670-1870



■ Data Collection at Population and Family History Project, Reitaku U.



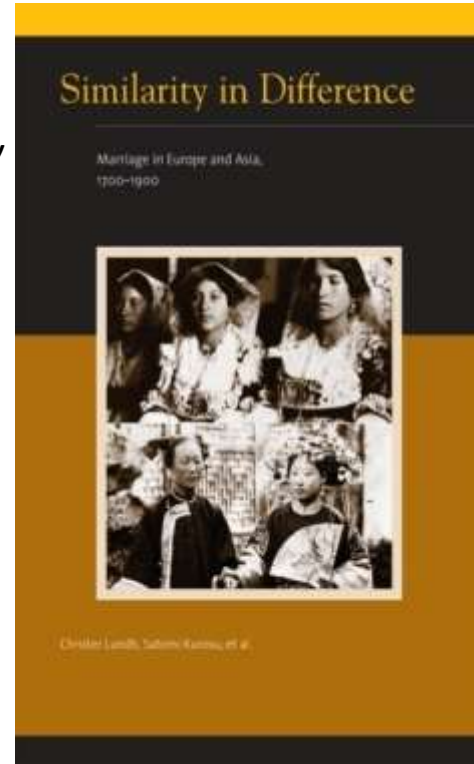
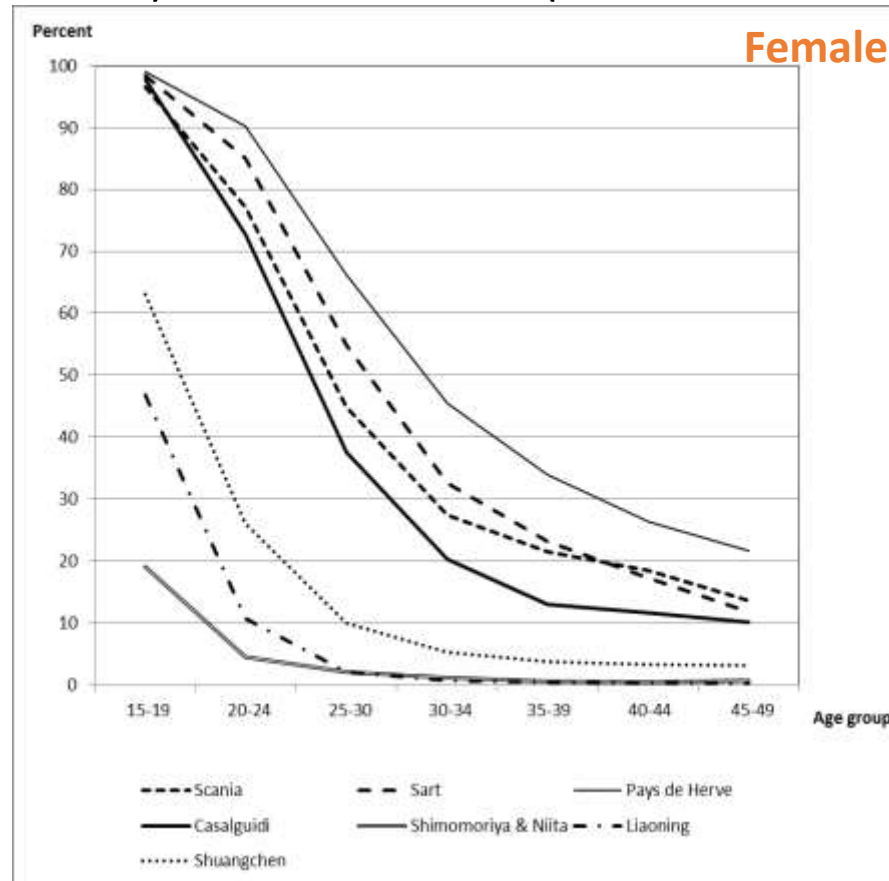
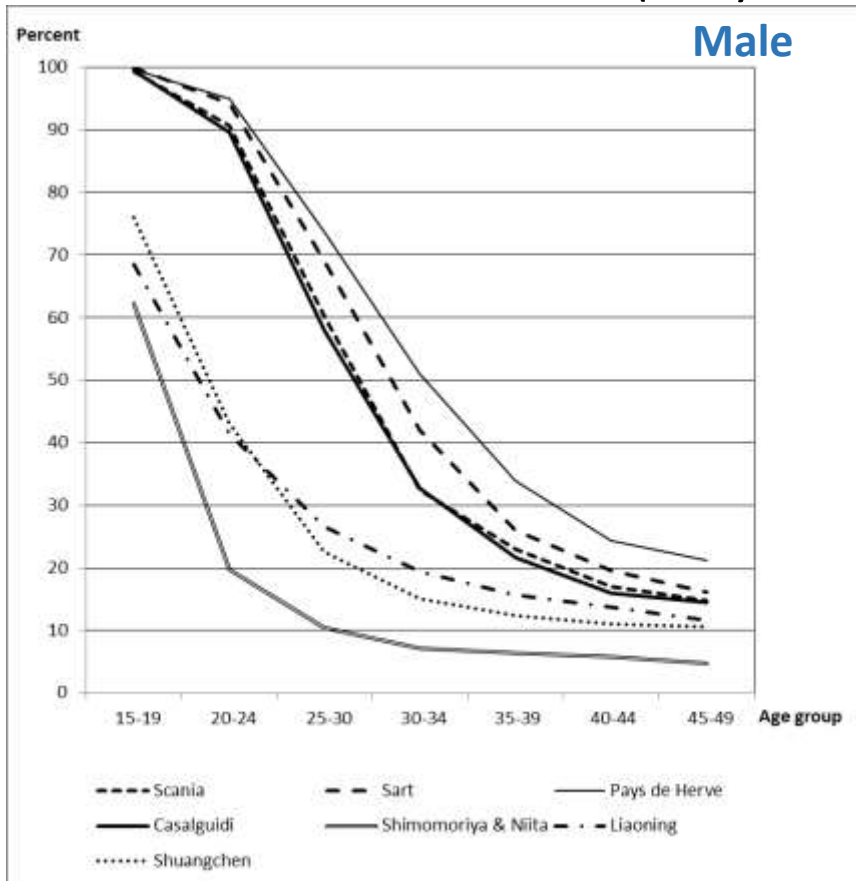
Shimomoriya
and
Niita used for
EAP (Eurasian
Family and
History
Project)

3.1 A New Framework of Analysis: EAP and Marriage

➤ Eurasian Population and Family History Project (EAP):

- A study of marriage in preindustrial Europe and Asia that goes beyond the Malthusian *East–West dichotomy* to find variation within regions and commonality across regions.
- Comparable individual panel data for complete rural communities from population registers in **Sweden, Belgium, Italy, Japan and China 1700-1900** (1,000,000 person years or observations)

➤ %Never-married → **East** (early and universal) and the **West** (late and selective)



*Similarity in Difference:
Marriage in Europe and
Asia, 1700-1900*
Lundh, Kurosu, et al
(The MIT Press 2014)

3.2 EAP and Marriage : Similarities

- The EAP applied identical or nearly identical models relating marriage outcomes to characteristics of the community, the household and the individual
- Found more **similarities**: importance of *power* (location within family and other status hierarchies) and *property* (location within SES/wealth hierarchies) in structuring inequality
 - strong positive association between SES and timing of marriage in all *male* populations
 - household relationship was important in Asian context; however, parental and other family contextual influences were apparent in Europe as well--- e.g. sibling effect in first marriage, child effect in remarriage
- Differences within Europe and within Asia are as important as differences between the two regions

3.3 EAP and Japan: differences

- A) In **Europe**, marriage was closely linked to **reproduction**, in **Asia** it was a **promise of future reproduction**
 - >What is marriage? Not just for reproduction and personal preference/taste but for household labor, to secure family continuity and old age care; important for survival
- B) In Asia, marriage was not a formation of a new household: new member was incorporated into his/her parental household
- C) In Japan, **stem family system** was ideal and practiced---only one from one generation stays and the rest leaves → two types of marriage
 - Virilocal** (wife marry into live with husband's family) vs. **Uxorilocal** (husband marry into live with wife's family)
 - >SES important not only for men but also for **inheriting women** (i.e. uxorilocal marriage for women)
 - >presence of sibling was important for marriage: However, **sib size** mattered in Europe but **order** mattered in Asia (being first vs. rest → difference in life course patterns)
- D) In Japan, **remarriage** was the most frequent → but just like Europe, higher SES was associated with lower likelihood of remarriage (opposite in China)

4. Challenges and Opportunities

- **Time to accumulate more detailed comparisons:** Apply the EAP approach/methods to new data from other villages and towns
 1. Regional variations (Northeast, Central, Southwest) to be examined with micro data
 2. Non-rural setting (e.g. **Koriyama town**) or non-agricultural setting (e.g. fishing vil.)
- **Comparison within Asia and/or East-Asia** (e.g. joint vs. stem family)
- **Linking past to present:** Implication to contemporary demography
- **Link to contextual information** (climate change, famine, policy...)
- **Relationship among demographic factors** (mortality, reproduction, marriage, remarriage, migration)
- **Geographical mobility:** Use of detailed migration records
 - ➔ **How to model?! Time trend x life course x geographical mobility**
(local time, household time, individual time)

4.2 Geographical mobility

% Distribution of Reasons of **Out-migration** by Sex:
Shimomoriya and Niita 1716-1870

Reason	Male	Female
Marriage/adoption	10.4	45.1
Service	65.1	34.9
Absconding	16.8	12.6
Change of legal domicile	5.9	5.9
Other	1.8	1.5
(Number of events)	(1,598)	(1,131)

Tsuya NO, Kurosu S. Social class and migration in two northeastern Japanese villages, 1716– 1870. *History of the Family*. 2013;18:434–455.

Linking past to present...

Koriyama: Now and then (around 1800)



十返舎一九『方言修行金草鞋（むだしゅぎょうかねのわらじ）』（文化10(1813)年～天保5(1834)年刊)