Introduction into cross-sectional data and databases

Siegfried Gruber

Laboratory of Historical Demography (MPIDR)

научной школы «Новые информационные технологии в исторических исследованиях: от источника к методу»

Jekaterinburg, Russia, 8 – 10 September 2014
Outline

• Sources

• Uses

• Challenges

• Databases
Sources

- Census and census-like sources
- 1 point in time
- Population characteristics
- Population changes only partly available
- Many comparative sources available
Types of sources

- Kind of source (painting, sculpture, writing …)
- Private or public (state or church)
- Original source/edition/image/compilation/database
- Amount of data
- Population composition or change
“A census of population may be defined as the total process of collecting, compiling, evaluating, analyzing and publishing or otherwise disseminating demographic, economic and social data pertaining, at a specified time, to all persons in a country or delimited part of a country” (United Nations 1998, p. 3)
Characteristics of a census

• National legal authority
• Defined enumeration area
• Complete coverage
• Simultaneous enumeration
• Individual enumeration
• Periodic enumeration
• Publication of results
• Dissemination of results
(Goyer 1986)
Census lists

- Census taking has become a general practice in 19th century Europe
- USA: required by constitution
- In the course of time additional information added
- Now register-based censuses replacing “real” census
Census lists

- Early population counting already in ancient times (Egypt, Babylonia, China, Palestine, Rome)
- 1st modern census? Nouvelle France/Canada 1665, Iceland 1703
- Population characteristics
<table>
<thead>
<tr>
<th>Name</th>
<th>Relationship</th>
<th>Age</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Icha Bonamianski</td>
<td>Husband</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chaya Zona Fgo</td>
<td>Wife/Spouse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idel Syn Fk</td>
<td>Daughter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michael Karochnicki</td>
<td>Father</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dobra Zona Fgo</td>
<td>Daughter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abram Falkowski</td>
<td>Brother</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chaya Zona Fgo</td>
<td>Sister/Spouse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abram Minnowski Litpl</td>
<td>Nephew</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pinchas Falkowski</td>
<td>Son</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pesha Falkowska Talpa</td>
<td>Nephew</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tova Zora Fgo</td>
<td>Daughter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kmzka Falkowska Talpa</td>
<td>Nephew</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Juda Sztoldmawski</td>
<td>Sister/Spouse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shava Zona Fgo</td>
<td>Daughter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abram Syn Fk</td>
<td>Brother</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noms et prénoms des habitants</td>
<td>Indication numérique des individus</td>
<td>Profession</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------------</td>
<td>------------</td>
<td></td>
</tr>
</tbody>
</table>
| Meisen Mathias                | 1                                 | 46
| Meisen Marie Josephine        | 1                                 | 43
| Meisen Mathias                | 1                                 | 14
| Meisen Adelaide               | 1                                 | 12
| Meisen Pierre                 | 1                                 | 10
| Meisen Elisabeth              | 1                                 | 7
| Meisen Dominique              | 1                                 | 3
| Meisen Josephine              | 1                                 | 2
| Meisen Félise                 | 1                                 | 1
| Wissen Mathias                | 1                                 | 64
| Wissen Jeanne                 | 1                                 | 54
| Wissen Jeanne Marie           | 1                                 | 44
| Wissen Marie agnes            | 1                                 | 21
| Rock guillaume                | 1                                 | 20
| Rammelhorst Marie             | 1                                 | 123
| Dauenzahl                     | 1                                 | 87

Census 1812 Dép. de la Roer, Borcette/Burtscheid
<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Age</th>
<th>Occupation</th>
<th>Residence</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>Male</td>
<td>30</td>
<td>Laborer</td>
<td>Vienna</td>
<td>Husband</td>
</tr>
<tr>
<td>Jane</td>
<td>Female</td>
<td>25</td>
<td>Teacher</td>
<td>Bratislava</td>
<td>Wife</td>
</tr>
</tbody>
</table>

Census 1869 Austria-Hungary
<table>
<thead>
<tr>
<th>№</th>
<th>Фамилија</th>
<th>Препознатак</th>
<th>Род</th>
<th>Возраст</th>
<th>Статус</th>
<th>Епифеномен</th>
<th>Грчко поношење</th>
<th>Национална раса</th>
<th>Чао</th>
<th>Области</th>
<th>Место</th>
<th>Зошто се државно поношење добио</th>
<th>Примета</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Милушиновић</td>
<td>46</td>
<td>23</td>
<td>18</td>
<td>жена</td>
<td>простак</td>
<td>прво поношење</td>
<td>серб</td>
<td>државни</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Стошић</td>
<td>22</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Добровчаковић</td>
<td>38</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Пантомичевић</td>
<td>48</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>№</td>
<td>Фамилия</td>
<td>Имя</td>
<td>Отчество</td>
<td>Возраст</td>
<td>Пол</td>
<td>Род</td>
<td>Место рождения</td>
<td>Вероятность</td>
<td>Отметка</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>--------------</td>
<td>-------------</td>
<td>----------</td>
<td>---------</td>
<td>-----</td>
<td>-----</td>
<td>----------------</td>
<td>-------------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Алексеев</td>
<td>Алексеев</td>
<td></td>
<td>46</td>
<td>М</td>
<td>Род</td>
<td>Воронежской губ. Орловской</td>
<td>При дозна</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Алексеев</td>
<td>Алексеев</td>
<td></td>
<td>47</td>
<td>М</td>
<td>Род</td>
<td>Воронежской губ. Орловской</td>
<td>При дозна</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Алексеев</td>
<td>Алексеев</td>
<td></td>
<td>48</td>
<td>М</td>
<td>Род</td>
<td>Воронежской губ. Орловской</td>
<td>При дозна</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Алексеев</td>
<td>Алексеев</td>
<td></td>
<td>49</td>
<td>М</td>
<td>Род</td>
<td>Воронежской губ. Орловской</td>
<td>При дозна</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Алексеев</td>
<td>Алексеев</td>
<td></td>
<td>50</td>
<td>М</td>
<td>Род</td>
<td>Воронежской губ. Орловской</td>
<td>При дозна</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Алексеев</td>
<td>Алексеев</td>
<td></td>
<td>51</td>
<td>М</td>
<td>Род</td>
<td>Воронежской губ. Орловской</td>
<td>При дозна</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Алексеев</td>
<td>Алексеев</td>
<td></td>
<td>52</td>
<td>М</td>
<td>Род</td>
<td>Воронежской губ. Орловской</td>
<td>При дозна</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Алексеев</td>
<td>Алексеев</td>
<td></td>
<td>53</td>
<td>М</td>
<td>Род</td>
<td>Воронежской губ. Орловской</td>
<td>При дозна</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Алексеев</td>
<td>Алексеев</td>
<td></td>
<td>54</td>
<td>М</td>
<td>Род</td>
<td>Воронежской губ. Орловской</td>
<td>При дозна</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Volkstzählung am 1. Dezember 1900.
Grossherzogthum Mecklenburg-Schwerin.

Zählkarte für in der Haushaltung Anwesende.
Zählkarte Nr. 1, zur Haushaltungsliste Nr. 18, des Zählbezirks Nr. 18.

<table>
<thead>
<tr>
<th>Angaben</th>
<th>Wert</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Vor- und Familienname</td>
<td>Anna Koch</td>
</tr>
<tr>
<td>2. Geschlecht</td>
<td>männlich</td>
</tr>
<tr>
<td>3. Familienstand</td>
<td>verheiratet</td>
</tr>
<tr>
<td>5. Geburtsort</td>
<td>Plau</td>
</tr>
<tr>
<td>6. Beruf, Stand, Gewerbe, Geschäfts oder Tätigkeitszweig</td>
<td>Haushalt, Hausfrau</td>
</tr>
<tr>
<td>7. a. Gerichts-Ort, in welcher der Wohnung (Vorstadt)</td>
<td>Plau</td>
</tr>
<tr>
<td>8. Religionsbezeichnung</td>
<td>Evangelisch</td>
</tr>
<tr>
<td>9. Muttersprache</td>
<td>Deutsch</td>
</tr>
<tr>
<td>10. Geburtsort</td>
<td>Plau</td>
</tr>
<tr>
<td>11. Besondere Angaben</td>
<td></td>
</tr>
<tr>
<td>Family Number</td>
<td>Name</td>
</tr>
<tr>
<td>---------------</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>Durn</td>
</tr>
<tr>
<td>2</td>
<td>Visha</td>
</tr>
<tr>
<td>3</td>
<td>Shisha</td>
</tr>
<tr>
<td>4</td>
<td>Info</td>
</tr>
<tr>
<td>5</td>
<td>Dasha</td>
</tr>
<tr>
<td>6</td>
<td>Alija</td>
</tr>
<tr>
<td>7</td>
<td>Fatih</td>
</tr>
<tr>
<td>8</td>
<td>Bajram</td>
</tr>
<tr>
<td>9</td>
<td>Branim</td>
</tr>
</tbody>
</table>
Registration systems

A population registration system can be defined as “an individualized data system, that is, a mechanism of continuous recording, and/or of coordinated linkage, of selected information pertaining to each member of the resident population of a country in such a way to provide the possibility of determining up-to-date information concerning the size and characteristics of that population at selected time intervals” (United Nations 1969)
Population registers

Some started already in the 18th century, in Belgium since 1846, Netherlands since 1850

Base inventory of population and its characteristics plus all changes that follow
<table>
<thead>
<tr>
<th>Woningen.</th>
<th>Wijk. Nummer</th>
<th>NAAM</th>
<th>VOORNAMEN</th>
<th>BEROEP of MIDDEN VAN bestaan</th>
<th>GOEDHUIST.</th>
<th>OSGBREWED.</th>
<th>GEBRIJZE MIED.</th>
<th>WEDUWEMALE</th>
<th>WEDUWEL</th>
<th>LAATSTE WONEN</th>
<th>DAGTEKENING VAN PLAATS, DAG, MAAND, JAAR VAN GEBORTE.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 153</td>
<td></td>
<td>MELHAGH</td>
<td>CHRISTIANE</td>
<td>Schraaff, P. C.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Geboren in de Gemeente, 4 Jun 1853</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CAROLINE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Overleden in de Gemeente, 6 Dec 1855</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MEHAGGE</td>
<td>FRANZESIAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Verbonden op 22 Augustus 1850</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MEDEGER</td>
<td>MARTINUS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gebrand op 3 Maart 1856</td>
</tr>
</tbody>
</table>

**Dutch population register 1826-1848 Leiden**
Status Barochia Olpeh.  
Sub initium Anni 1750.

Nomina, Numerus, Rites & Condicio  
Familiorum, Patrum, & Matris familii,  
liberorum, familiarum, assillarum,  
nee non incolarum, viduorum, viduarum.  
juvenum & virginum.

A. zu v. Vigilius  
28.

Peters  
27.

Peters  
27.

Peters  
27.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.

Hotman  
26.
## Register-type family book, Diocese of Rottenburg

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Name</th>
<th>Birth</th>
<th>Confirmation</th>
<th>Baptism</th>
<th>Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sabina</td>
<td>26 Oct 1834</td>
<td></td>
<td>14 Aug 1836</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Maria</td>
<td>14 Oct 1836</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Josefine</td>
<td>15 Apr 1838</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Franz</td>
<td>16 Feb 1840</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Karl</td>
<td>8 Apr 1842</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Albert</td>
<td>24 Mar 1843</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Other sources

Tax lists:
- Head tax lists
- Soul revision lists

Other church listings:
- Lists of communicants
- Lists of confessions
<table>
<thead>
<tr>
<th>Alter</th>
<th>männl.</th>
<th>weibl.</th>
<th>männl.</th>
<th>weibl.</th>
<th>männl.</th>
<th>weibl.</th>
</tr>
</thead>
<tbody>
<tr>
<td>30.</td>
<td></td>
<td></td>
<td>2.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>60.</td>
<td>59.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60.</td>
<td>30.</td>
<td>19.</td>
<td>17.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34.</td>
<td>43.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Amount and quality of sources

Worse in East and Southeast Europe:

• Vital registration starts later
• Census taking in Russian and Ottoman Empire starts later
• Quality of data worse (less education and worse administration)

• Many sources no longer available because of destruction
Microdata

- All known characteristics for each individual person
- Researchers can custom-design their own tabulations
- Multivariate analyses are possible
- Small groups can be investigated
Source criticizm

• Who counts?
• Who might want to avoid being counted?
• Mobile populations
• Information about ethnic or religious minorities
Challenges

- Incomplete data
- Errors in data
- Ambiguous and fuzzy data
- Missing data
- Finding appropriate data might be challenging
- Transcription of data is time consuming
How to analyze events or life courses?

- Use other sources
- Link consecutive censuses
- Develop synthetic measures
How to analyze events?

- Births: child-woman-ratio, own-child-method, children ever born
- Marriages: Singulate mean age at marriage
- Migration: if information about place of birth, year of arrival
Challenges in analyzing household structures

- Analysis of persons and not families or households
- Control for age
- Control for demographic conditions
- Use simple measures
- Pay as much attention to denominators as to numerators

(Ruggles 2012)
Tobler’s first law of geography: “everything is related to everything else, but near things are more related than distant things” (Tobler 1970: 236)

- Spatial aspects still not enough included in historical research
- better possibilities due to available GIS-software
Data bases

Censuses:
- IPUMS-USA
- NAPP
- Mosaic
- Viennese Database on European Family History

Church books:
- Demographic Database Umeå
IPUMS-USA: usa.ipums.org/usa/index.shtml
IPUMS-USA

- Minnesota Population Center
- Started in 1989
- More than 50 samples of 15 censuses since 1850
- Samples of surveys since 2000
North Atlantic Population Project

Census microdata from Canada, Great Britain, Germany, Iceland, Norway, Sweden, and the United States from 1801 to 1910. The project’s goals are to:

- Harmonize data, including many complete count datasets
- Link individuals between census years for longitudinal analysis
- Disseminate the data absolutely freely

The North Atlantic Population Project gratefully acknowledges the collaboration and contributions of

THE CHURCH OF JESUS CHRIST OF LATTER-DAY SAINTS

Funding provided by:

National Institute of Child Health and Human Development
National Science Foundation

Copyright © Minnesota Population Center, University of Minnesota
North Atlantic Population Project

- Minnesota Population Center
- Since 2001
- 25 samples of 1801-1910 from 7 countries
Recovering Surviving Census Records to Reconstruct Population, Economic, and Cultural History

The Mosaic project identifies, gathers, harmonizes, and distributes surviving historical census microdata for Europe and beyond. It aims at building a comprehensive and detailed resource for the historic study of populations by historians, demographers, economists, and other researchers.

The Mosaic project builds on historical integrated census microdata projects such as the Integrated Public Use Microdata Series (IPUMS) and the North Atlantic Population Project (NAPP). Records are distributed in the same integrated format, so that historical comparisons can be made across time and space.

Data for the Mosaic project is provided by its partners, which include an international set of institutions in Europe and beyond.

The Mosaic project began with a conference in May 2011 (Events) and is coordinated at the Max Planck Institute for Demographic Research.

For more details click here.
• Until now coordinated by the Max Planck Institute for Demographic Research in Rostock
• Started in 2011
• 25 samples with 200,000 persons
NAPP and Mosaic in Europe
Georeferenced Mosaic data
Mosaic data already available for research
Common features of IPUMS, NAPP, and Mosaic

- Data is for free for scientific research, but registration is necessary
- Proper citation is required
- Standardized variables
- Occupations are coded into HISCO (occhisco)
Downloading and setting up data

IPUMS-International: https://international.ipums.org/international/

IPUMS USA: https://usa.ipums.org/usa/

NAPP: https://www.nappdata.org/napp/

Mosaic: http://www.censusmosaic.org
Comments or Questions?

www.demogr.mpg.de