**Big Data and New Methods in Chinese Social History:**

**Collection, Transcription, Analysis**

Course Syllabus

Matthew Noellert, Cameron Campbell, James Lee

# **Course Description**

This course focuses on the construction of four kinds of big historical data using Chinese sources: social-demographic, social-economic, social-political, and education and employment. We devote three weeks to each data type to discuss 1) the data sources and institutions that produced them, 2) major examples of data acquisition and data transcription, and 3) major scholarly contributions from analyses of these data.

# **Assignments and Learning Outcomes**

Students are expected to familiarize themselves with all four kinds of big historical data, to explore these data quantitatively, and to write about their explorations. The first three writing assignments are based on the CMGPD-LN, the CMGPD-SC, and our ongoing data collection on the Social Origins of Republican Students. The last assignment should derive from our initial work on either land reform materials from Heilongjiang and/or Shanxi, or四清 materials from Shanxi Province.

1. Three 500-word group data exploration write-ups (3 X 15 percent of course grade)
   1. Weeks 4, 7, 10
   2. At least half the write-up must be in prose
   3. Tables and figures count at most 100 words each
2. 5-page original research proposal (45 percent of course grade)
   1. Week 14 – Problematic statement
   2. Week 16 – Topic sentence outline
   3. Week 18 – Final draft due
   4. At least half the proposal must be in prose
   5. Tables and figures count at most 200 words each
3. Class participation (10 percent of course grade)

# Assessment

The grading rubric: All written response assignments must respond to the relevant lectures and/or reading materials, topic, and writing exercise for each respective week. In addition, we will evaluate your written work based on five criteria:

1) **English language -** your ability to write a short narrative with proper word choice and grammar

Below standard: sentences and ideas are incomprehensible.

Meets standard: overall meaning and ideas can be understood, but still contains some minor grammar mistakes.

Above standard: writing is virtually error-free, and ideas are expressed clearly and fluently

2) **Organization -**your ability to organize your narrative using paragraphs with topic sentences and segues

Below standard: no topic sentences, segues, or coherent paragraphs

Meets standard: employs topic sentences and basic segues, but paragraph organization and overall narrative structure is still incomplete

Above standard: constructs a solid, complete narrative structure based on clearly-stated topic sentences, fluent segues, and succinct paragraphs

3) **Evidence -**your ability to provide sufficient evidence to support your analytic thesis

Below standard: simply asserts personal opinion, or attempts to use evidence to support ideas but is ineffective, irrelevant, and/or not convincing

Meets standard: demonstrates meaningful connections between evidence and conclusion

Above standard: uses relevant evidence to convincingly demonstrate that your evidence and therefore your conclusion complement or are superior to previous scholarship

4) **Interpretation -**your ability to write a constructive narrative expressing your own understanding / interpretation of your evidence

Below standard: simply repeats a collection of available ideas

Meets standard: describe your evidence and show how it contradicts or fits prevailing ideas, but without proposing new ideas

Above standard: persuasively show how your evidence extends our understanding of your topic and perhaps even broader topics

5) **Source materials:** your ability to incorporate outside sources to supplement the assigned readings and lectures.

Below standard: does not effectively employ outside sources to support ideas

Meets standard: uses relevant outside sources to support and inform ideas related to the writing theme, but still lacks development of these ideas

Above standard: uses relevant outside sources to effectively develop ideas, convince the reader, and stimulate further discussion

# **Reading Schedule**

Week 1 (October 10) - Introduction and conceptual discussion

梁晨，董浩，李中清。2014.《计量数据库与历史研究》准备投稿至《历史研究》或《中国社会科学》

Campbell, Cameron and James Z. Lee. 2014. *China Multigenerational Panel Dataset (CMGPD) Training Guide*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Wang, Feng. 2008. *Boundaries And Categories: Rising Inequality In Post-Socialist Urban China*. Stanford University Press: 3-45.

## I. Social-Demographic Data and Individual Behavior

Week 2 (October 17) – Data Sources and Institutions：户口册，家谱

Co-taught with Xiao Xing 肖星 and Jiyang激扬

Hao Dong, Cameron Campbell, Satomi Kurosu, Wen-shan Yang, and James Z. Lee, "New Sources for Comparative Social Science: Historical Population Panel Data from East Asia." Manuscript.

Cameron Campbell and James Lee. 2006. ‘State views and local views of population: Linking and comparing genealogies and household registers in Liaoning, 1749-1909.’ History and Computing. 14.1+2:9-29. http://papers.ccpr.ucla.edu/download.php?paper=PWP-CCPR-2004-025

Lee, James Z., and Cameron D. Campbell. China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN), 1749-1909. ICPSR27063-v10. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2014-07-10. <http://doi.org/10.3886/ICPSR27063.v10>

Week 3 (October 25 – Saturday) – From Population Registers to Panel Data

Co-taught with Dong Hao董浩

James Lee, Cameron Campbell, Shuang Chen. 2010. *A User Guide to the China Multi-Generational Panel Data-Liaoning (CMGPD-LN)*. Ann Arbor: Inter‑University Consortium for Political and Social Research Study 27063. <http://www.icpsr.umich.edu/icpsrweb/CMGPD/studies/27063/documentation>

Week 4 (October 31) - From Panel Data to Individual Behavior

LEE, James and WANG Feng. 1999*. One Quarter of Humanity: Malthusian Mythology and Chinese Realities, 1700-2000*. Harvard University Press: 1-23, 42-136.

First data exploration write-up due.

## II. Education-Employment Data and Social Mobility

Week 5 (November 7) – Data Sources and Institutions: 登科录，学生登记卡, 縉紳錄, 问卷

梁晨，张浩，李兰，阮丹青，康文林，李中清. 2013. 《无声的革命：北京大学, 苏州大学学生社会来源研究, 1949-2002》. 北京三联出版社，1-57.

Cameron Campbell. 2013. ‘Spatial, Temporal, and Social Network Influences on Officials' Careers during the Qing: Creation and Analysis of a National Database from the Jin Shen Lu:’ A GRF Application to the Hong Kong Research Grant Council.

James Lee, Hongbo Wang, Liang Chen. 2012. ‘Social Origins of University Students in Republican China’. A GRF Application to the Hong Kong Research Grant Council.

Week 6 (November 17 – Monday) – From Student Cards to Familial Origins

Co-taught with Professor Liang Chen梁晨

梁晨，张浩，李兰，阮丹青，康文林，李中清. 2013. 《无声的革命：北京大学, 苏州大学学生社会来源研究, 1949-2002》. 北京三联出版社， 58-194， 278-280.

Week 7 (November 21) – From Familial Origins to Social Mobility

梁晨，张浩，李兰，阮丹青，康文林，李中清. 2013. 《无声的革命：北京大学, 苏州大学学生社会来源研究, 1949-2002》. 北京三联出版社， 195-277，281-300.

王洪波, 李兰，李中清. 2013《苏州大学校友调查问卷》.

Second data exploration write-up due.

## III. Social-Economic Data and Wealth Distribution

Week 8 (November 29) – Data Sources and Institutions: 地亩册，鱼鳞图册

Co-taught with Prof. James Z. Lee, Xiao Xing肖星 and Jiyang激扬

韦庆元。1961.《明代黄册制度》,北京：中华书局。

[栾成显](http://book.douban.com/search/%E6%A0%BE%E6%88%90%E6%98%BE)。 1988. 《明代黄册研究》北京：中国社会科学出版社

Lee, James Z., Shuang Chen, Cameron D. Campbell, and Hongbo Wang. China Multi-Generational Panel Dataset, Shuangcheng (CMGPD-SC), 1866-1914. ICPSR35292-v1. Ann Arbor: Inter-university Consortium for Political and Social Research [distributor], 2014-07-28. <http://doi.org/10.3886/ICPSR35292.v1>

Week 9 (December 5) – From Land Registers to Linked Datasets

Wang, Hongbo, Shuang Chen, Hao Dong, Matt Noellert, Cameron Campbell, and

James Z Lee. 2013. *China Multi-Generational Panel Dataset, Shuangcheng*

*(CMGPD-SC) 1866-1914. User Guide.* Ann Arbor, MI: Inter-university Consortium

for Political and Social Research.

Week 10 (December 12) – From Linked Datasets to the Distribution of Wealth

Chen,Shuang. 2009. “Where Urban Migrants Met Rural Settlers: State Categories, Social Boundaires, and Wealth Stratification in Northeast China, 1815-1913”. PhD Dissertation, History Department. Ann Arbor, MI: University of Michigan, Chapter 8, “Wealth Stratification: State Categories and Individual Agency”.

Third data exploration write-up due.

## IV. Social-Political Data and Social Stratification

Week 11 (December 19) – Data Sources: 阶级划分表，被分地人登记表, 四清成份登记表

NOELLERT, Matthew. 2014. “New Perspectives on Communist Land Reform: Evidence from Northeast China, 1946-1948.” HKUST Humanities PhD Dissertation, Chapter 2, “New Data and New Perspectives”.

Week 12 (December 26) – From Revolution Registers to Big Data

NOELLERT, Matthew. 2014. “New Perspectives on Communist Land Reform: Evidence from Northeast China, 1946-1948.” HKUST Humanities PhD Dissertation, Part One, “Patterns”.

Week 14 (January 11) – From Big Data to Social Stratification

Co-taught with Prof. Cameron Campbell

NOELLERT, Matthew. 2014. “New Perspectives on Communist Land Reform: Evidence from Northeast China, 1946-1948.” HKUST Humanities PhD Dissertation, Part Two, “Practices”.

Research Proposal Problematic due

Week 16 (January 23): Topic sentence outline due

Week 17 (January 30): Final research proposal due