China Data Institute

chinadatacenter.net
China-data-online.org

Introduction to China Data Online and China Data Lab

Shuming Bao
China Data Institute

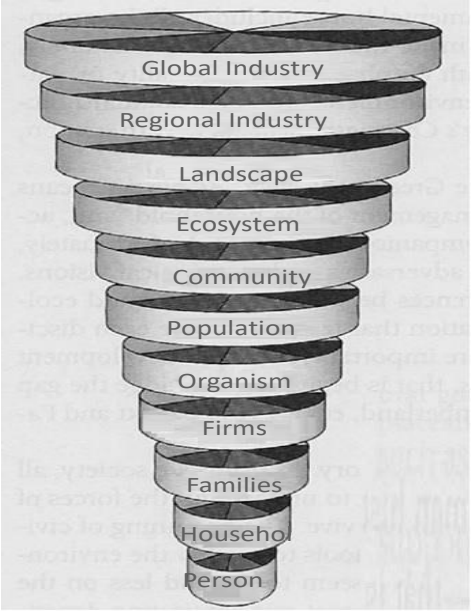
Most Challenges for China Data Studies

- **Availability**
- **Accessibility**
- **Comparability**

Principals for Spatial Data Design

Principals:

- **Comprehensive** data coverage
- **Comparable** spatio-temporal data structure
- **Compatible** multi-source data structure
- **Consistent** multi-scale data structure



The Principles of China Data Design

- Authority
- Completion
- Uniqueness

China Data Sources

- **Government Statistics**
  - Provincial Statistics (1949 -)
  - City Statistics (1996 -)
  - County Statistics (1997 -)
- **Population Census**
  - Census 1953
  - Census 1964
  - Census 1982
  - Census 1990
  - Census 2000/2010 (province, city, county, township, GRID)
- **Economic Census**
  - Industrial Census 1995 (province, city, county, ZIP)
  - Basic Unit Census 2001 (province, city, county, ZIP)
  - Economic Census 2004/2008 (province, city, county, ZIP)
- **Establishments** (more than 7 millions companies and organizations)
- **Geography and Environment**
  - Land Use data
  - Night-Time data

- **Base Maps**
  - 2000
  - 2010
  - 2000-2010

Population Census Data with GIS Maps

> 2,000 demographic variables in population Census

Geographical Levels:

Country
|
Province
|
Prefecture
|
County
|
Township
|
1 sq km Grid

Census data variables:

- ❑ General Information
- ❑ Nationalities
- ❑ Age Structure
- ❑ Household Structure
- ❑ Education
- ❑ Fertility
- ❑ Deaths
- ❑ Marriage
- ❑ Migration
- ❑ Housing Status
- ❑ Industries and Occupations

Economic Census Data

> Cover about 900 economic sectors with more than 7 million business units

Industries

- 852 industries

Products

- 3 primary products

Ownerships

- 23 different industries

Revenue

- 15 revenue ranges

Employment

- 10 employment ranges

Employment

1
1-19
20-49
50-99
100-499
500-999
1000-4999
5000-29999
30000-49999
50000+

Revenue

(in 10,000 Yuan)
0-30
30---50
50---100
100--300
300--500
500--1000
1000-3000
3000-5000
5000-10000
10000-30000
30000-50000
50000-100000
100000-150000
150000-200000
200000 and over

Business Unit Classification

2004, 2008 & 2013 Business and Private Units 单位数与个体经营户数

	2004		2008		2013	
	(in 10,000)	(%)	(in 10,000)	(%)	(in 10,000)	(%)
1. Legal Unit 法人单位	**516.9**	100	**709.9**	100	**1085.7**	100
Enterprise 企业法人	325	62.9	495.9	69.9	820.8	75.6
Government 机关、事业法人	90	17.4	95.9	13.5	103.7	9.6
Non-profit org 社会团体法人	101.9	19.7	118.1	16.6	161.1	14.8
2. Economic Unit 产业活动单位	**682.4**	100	**886.4**	100	**1303.5**	100
Manufacture 第二产业	167.5	24.6	230	25.9	287.5	22.1
Service 第三产业	514.9	75.4	656.4	74.1	1015.9	77.9
3. Private 个体经营户	**3921.6**	100	**2873.7**	100	**3279.1**	100
Manufacture 第二产业	588.7	15	253.8	8.8	188.3	5.7
Service 第三产业	3332.9	85	2619.9	91.2	3090.8	94.3

Industrial Classification (18 Categories)

行业分类	Industrial Classification
1. 农、林、牧、渔业	1. Farming, Forestry, Animal Husbandry and Fishery
2. 采矿业	2. Mining and Quarrying
3. 制造业	3. Manufacturing
4. 电力、燃气及水的生产和供应业	4. Production and Distribution of Electric Power, Gas and Water
5. 建筑业	5. Construction
6. 交通运输、仓储和邮政业	6. Transport, Storage and Post
7. 信息传输、计算机服务和软件业	7. Information Transmission, Computer Services and Software
8. 批发和零售业	8. Wholesale and Retail Trade
9. 住宿和餐饮业	9. Hotel and Restaurants
10. 金融业	10. Financial Intermediation
11. 房地产业	11. Real Estate
12. 科学研究、技术服务和地质勘查业	12. Scientific Research, Technical Service and Geologic Prospecting
13. 水利、环境和公共设施管理业	13. Management of Water Conservancy, Environ. and Public Facilities
14. 居民服务和其他服务业	14. Services to Households and Other Services
15. 教育	15. Education
16. 卫生、社会保障和社会福利业	16. Health, Social Security and Social Welfare
17. 文化、体育和娱乐业	17. Culture, Sports and Entertainment
18. 公共管理和社会组织	18. Public Management and Social Organization

Industrial Classification: (2 digits)

7. 信息传输、计算机服务和软件业	**Information Transmission, Computer Services and Software**
电信和其他信息传输服务业	Telecommunications and Other Information Transfer Services
计算机服务业	Computer Services
软件业	Software Industry
8. 批发和零售业	**Wholesale and Retail Trade**
批发业	Wholesale
零售业	Retail Trade
9. 住宿和餐饮业	**Hotel and Restaurants**
住宿业	Hotel
餐饮业	Catering Services
10. 金融业	**Financial Intermediation**
银行业	Banking
证券业	Securities
保险业	Insurance
其他金融活动	Other Financial Activities
11. 房地产业	**Real Estate**
租赁和商务服务业	Leasing and Business Services
租赁业	Leasing Services
商务服务业	Business Services
12. 科学研究、技术服务和地质勘查业	**Scientific Research, Technical Service and Geologic Prospecting**
研究与试验发展	Research and Experimental Development
专业技术服务业	Professional Skill Services
科技交流和推广服务业	Services of Scientific and Technological Exchange and Popularization
地质勘查业	Geological Prospecting
13. 水利、环境和公共设施管理业	**Management of Water Conservancy, Environ. and Public Facilities**
水利管理业	Management of Water Conservancy
环境管理业	Management of Environment
公共设施管理业	Management of Public Facilities
14. 居民服务和其他服务业	**Services to Households and Other Services**
居民服务业	Resident Services
其他服务业	Other Services

Industrial Classification: (4 digits)

7. Information Transmission, Computer Services and Software
信息传输、计算机服务和软件业

7.1. 电信和其他信息传输服务业 Telecommunications and Other Information Transfer Services	7.1.1 固定电信服务	Fixed telecommunications services
	7.1.2 移动电信服务	Mobile telecommunications services
	7.1.3 其他电信服务	Other telecommunication services
	7.1.4 互联网信息服务	Internet Information Services
	7.1.5 有线广播电视传输服务	Cable television transmission services
	7.1.6 无线广播电视传输服务	Radio and TV transmission services
	7.1.7 卫星传输服务	Satellite transmission services
7.2. 计算机服务业 Computer Services	7.2.1 计算机系统服务	Computer system services
	7.2.2 数据处理	Data processing
	7.2.3 计算机维修	Computer maintenance
	7.2.4 其他计算机服务	Other computer services
7.3. 软件业 Software Industry	7.3.1 基础软件服务	Based software services
	7.3.2 应用软件服务	Application software services
	7.3.3 其他软件服务	Other software services

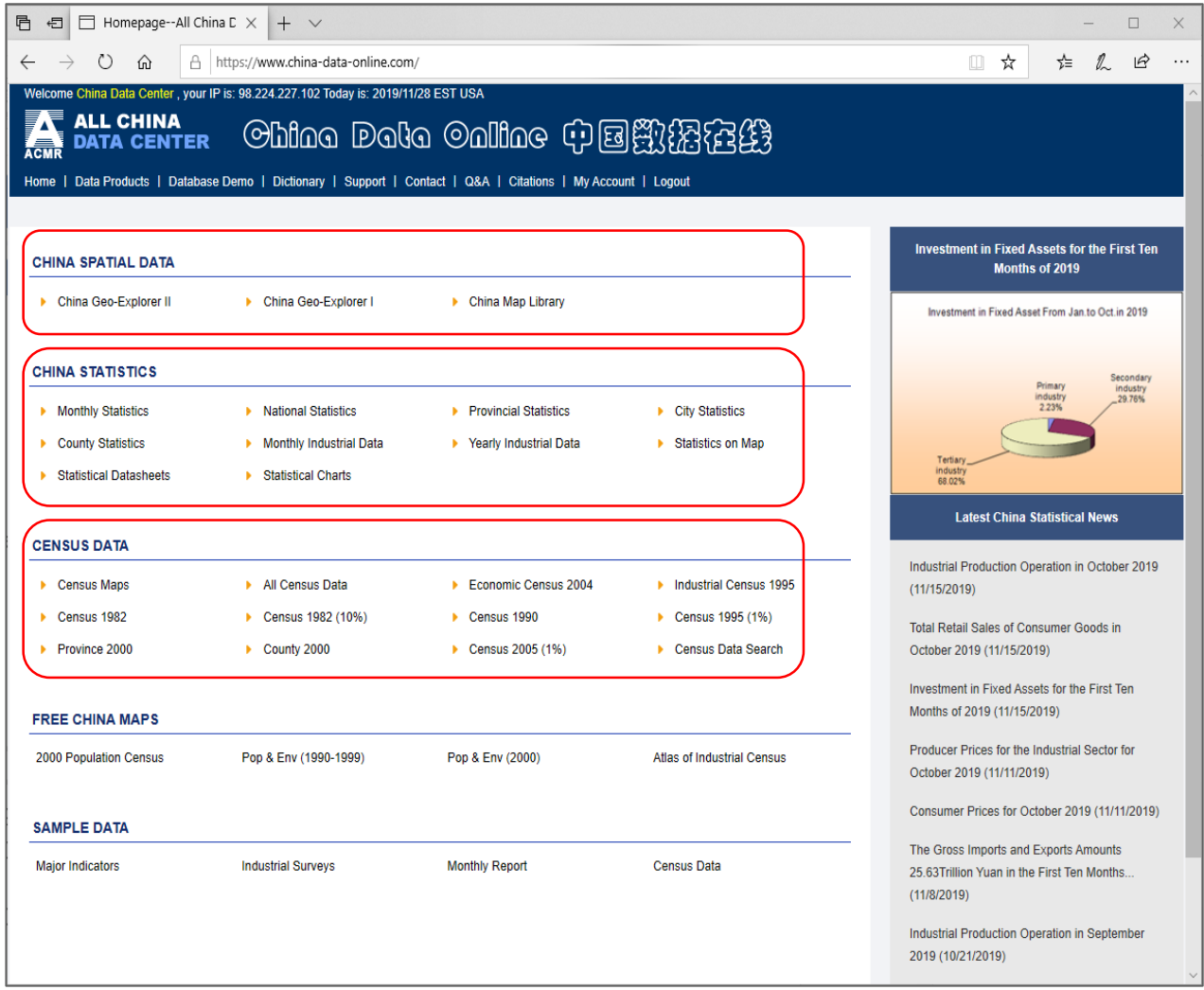
China Data Online: http://china-data-online.com

Statistical Database:

- Monthly Statistics
- National Statistics
- Provincial Statistics
- City Statistics
- County Statistics
- Monthly Industrial Data
- Yearly Industrial Data
- Statistics on Map
- Statistical Yearbooks

Census Database:

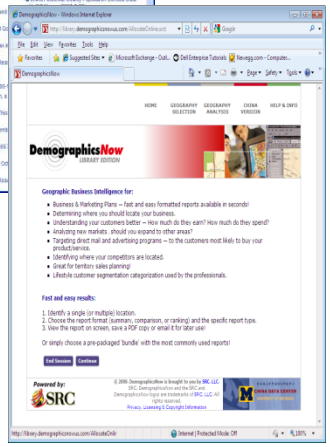
- Population Census 1982
- Population Census 1990
- Population Survey 1995, 2005
- Province Census 2000
- County Census 2000
- Economic Census 2004



China Geo-Explorer

An Integration of Spatial Data and Analysis for China Studies

Statistics

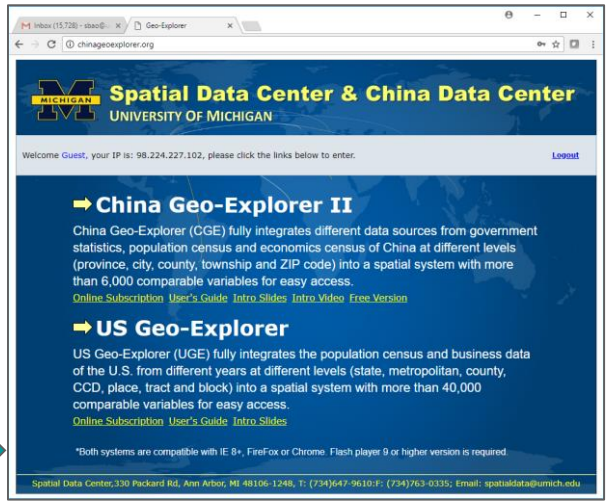


Census

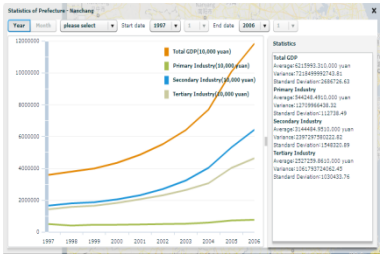
GIS

Data

Output

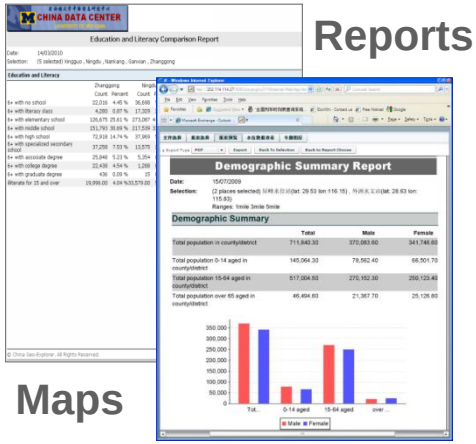


Charts

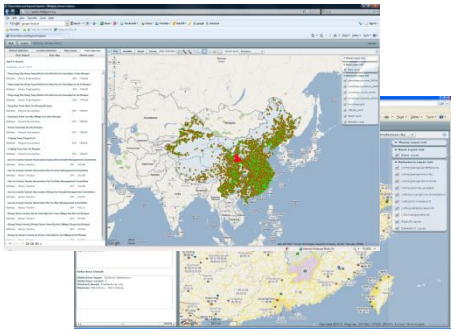


Tables

Reports



Maps



Unique Features of China Spatial Data

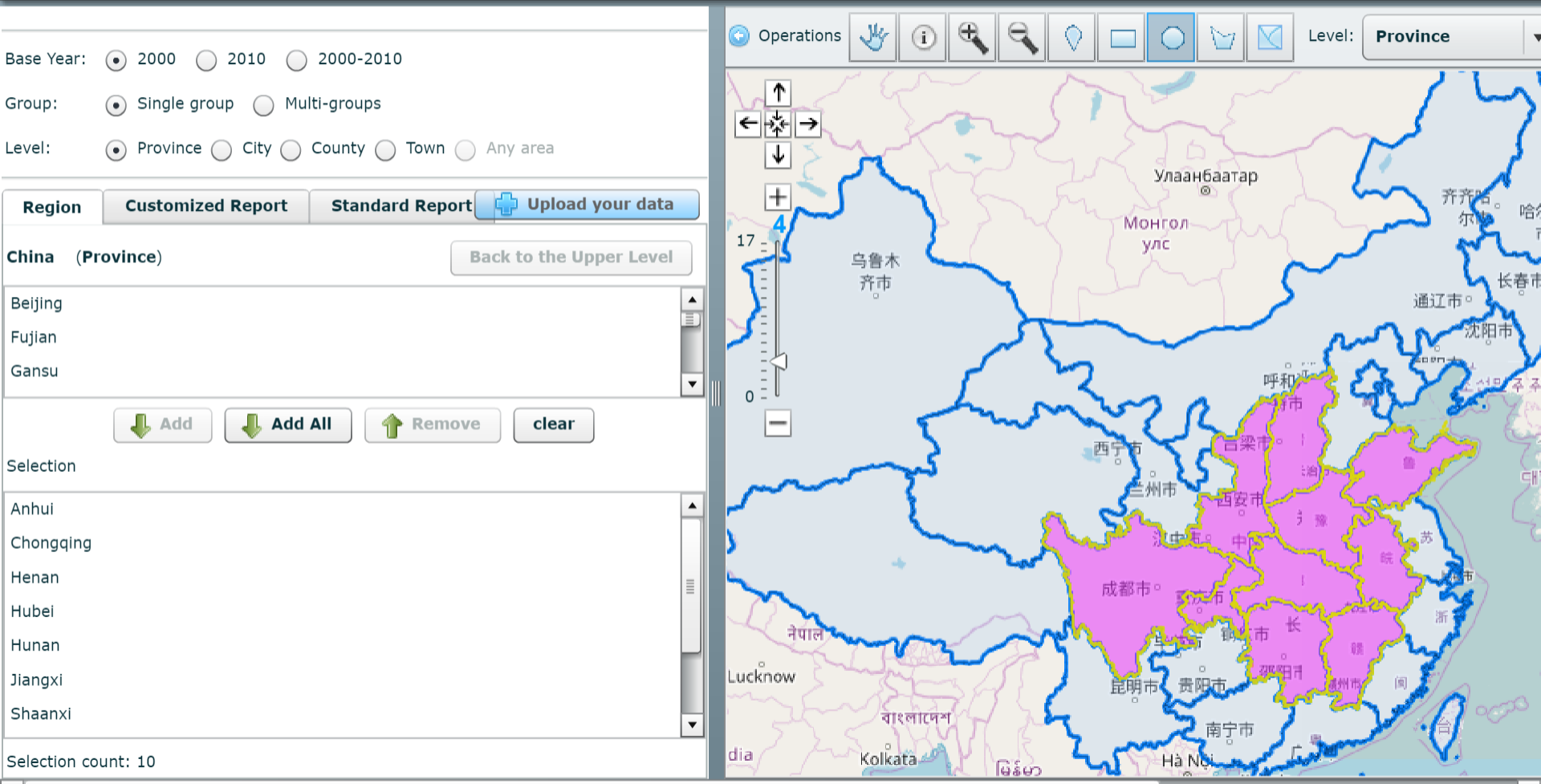
- The mostly complete collection in China's history
- Detailed data for nation, province, city, county, district and township
- Complete coverage for all provinces, cities, counties, and townships
- All data are comparable across time and region with the adjusted base map (2000, 2010)
- Most data in CGE are unique and not available in official publications
  - Population census data were compiled from the source data directly
  - Economic census data were compiled from the establishment data aggregated at province, city, county and ZIP level
  - All data have been integrated with GIS maps

Primary Functions

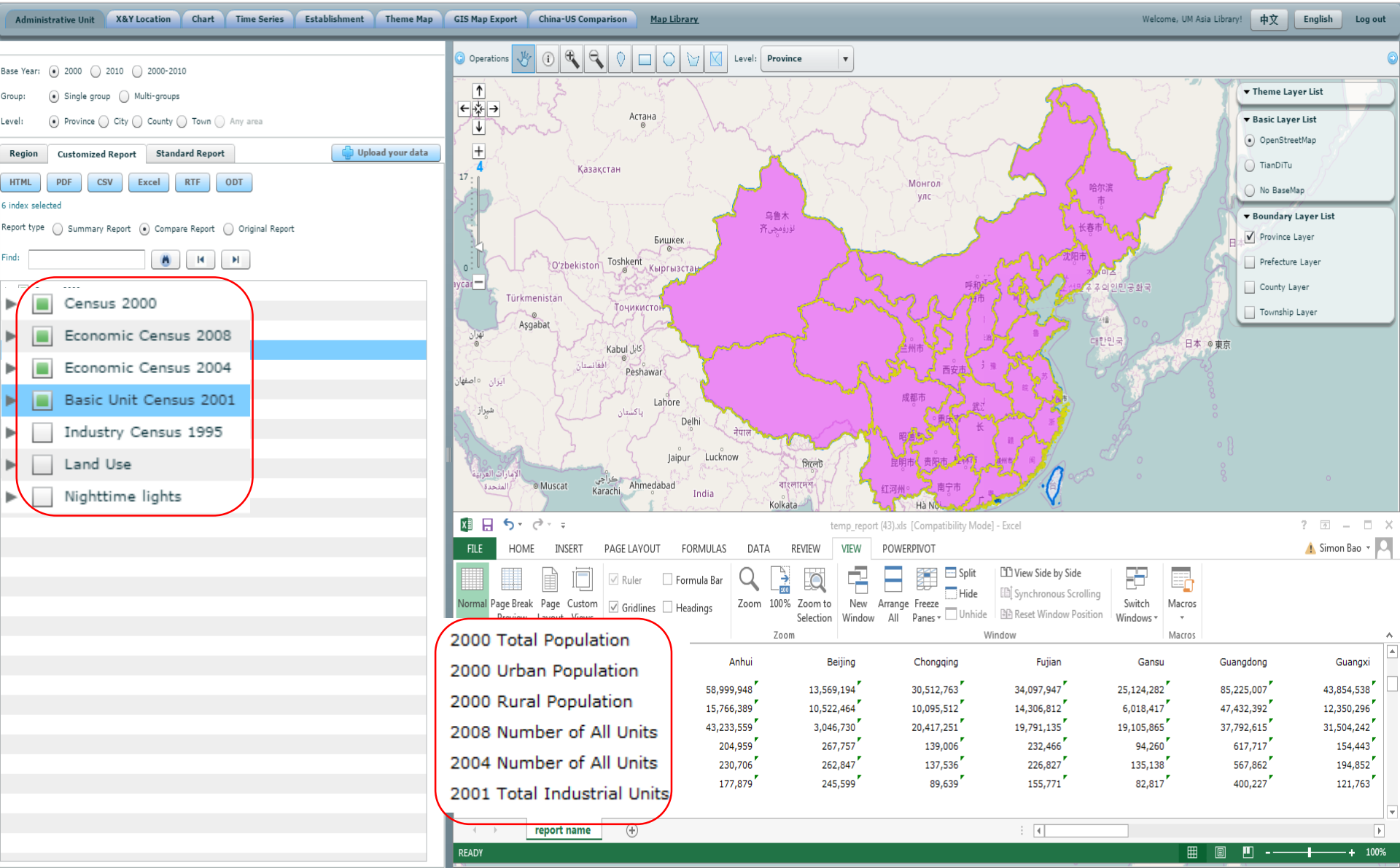
- **Data Selection**
  - By administrative units (province, city, county, township)
  - By groups
  - By location (X&Y) and spatial range (km or miles)
  - By time-series statistics (province, city and county)
  - By establishments (province, city, county and ZIP)
- **Reporting**
  - Summary report
  - Comparison report
  - Original data report
- **Export**
  - Data tables (Excel)
  - Reports (Excel, Word, PDF)
  - GIS maps (Shape)
  - Maps (PDF)
- **Map Library with Metadata**
  - Pre-defined maps
  - Easy links between maps of different spatial levels
  - Easy links to related industries

China Geo-Explorer

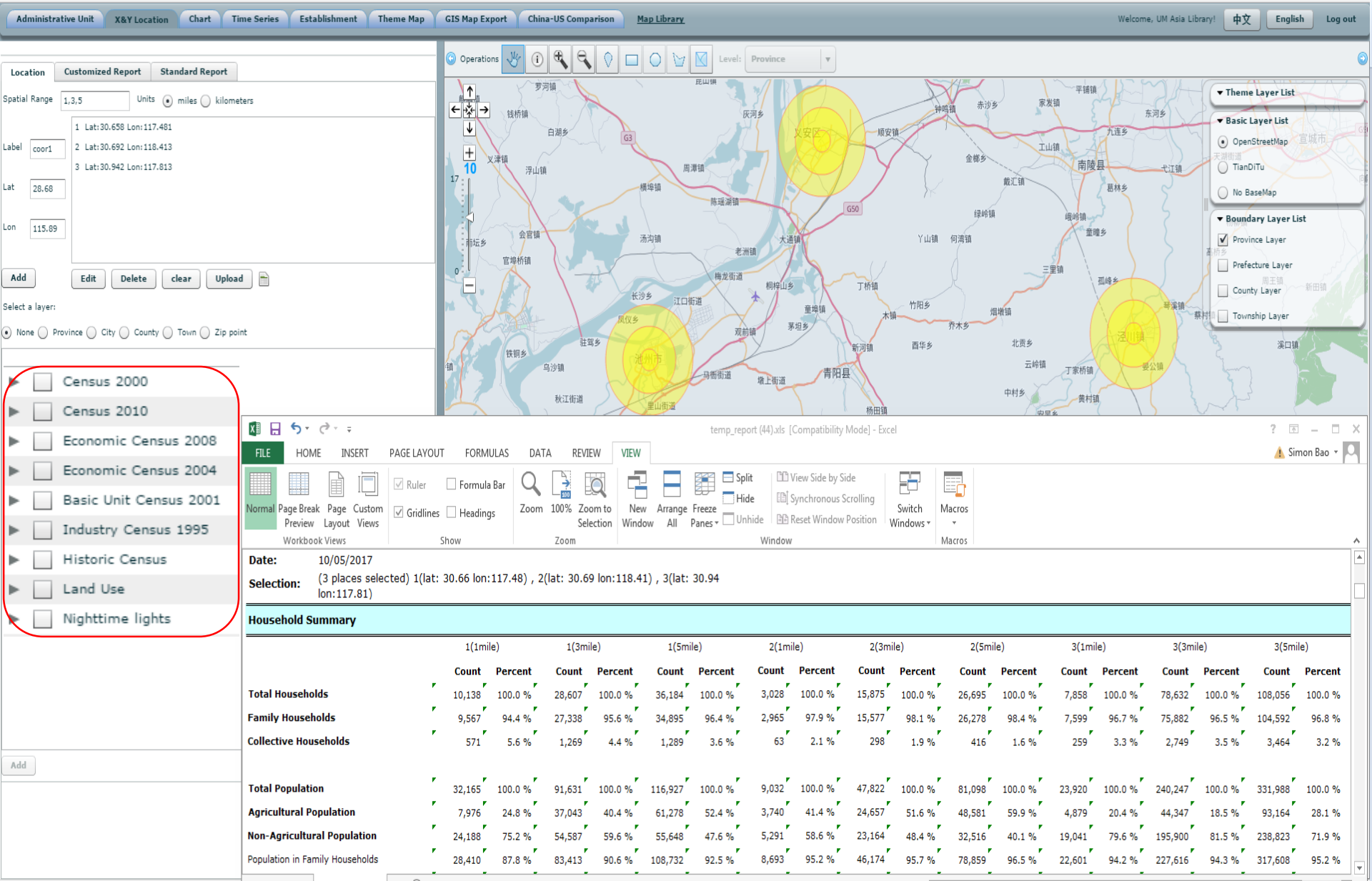
Administrative Units – Location – Chart – Time Series – Establishment – GIS Map Export – Map Library



Administrative Units

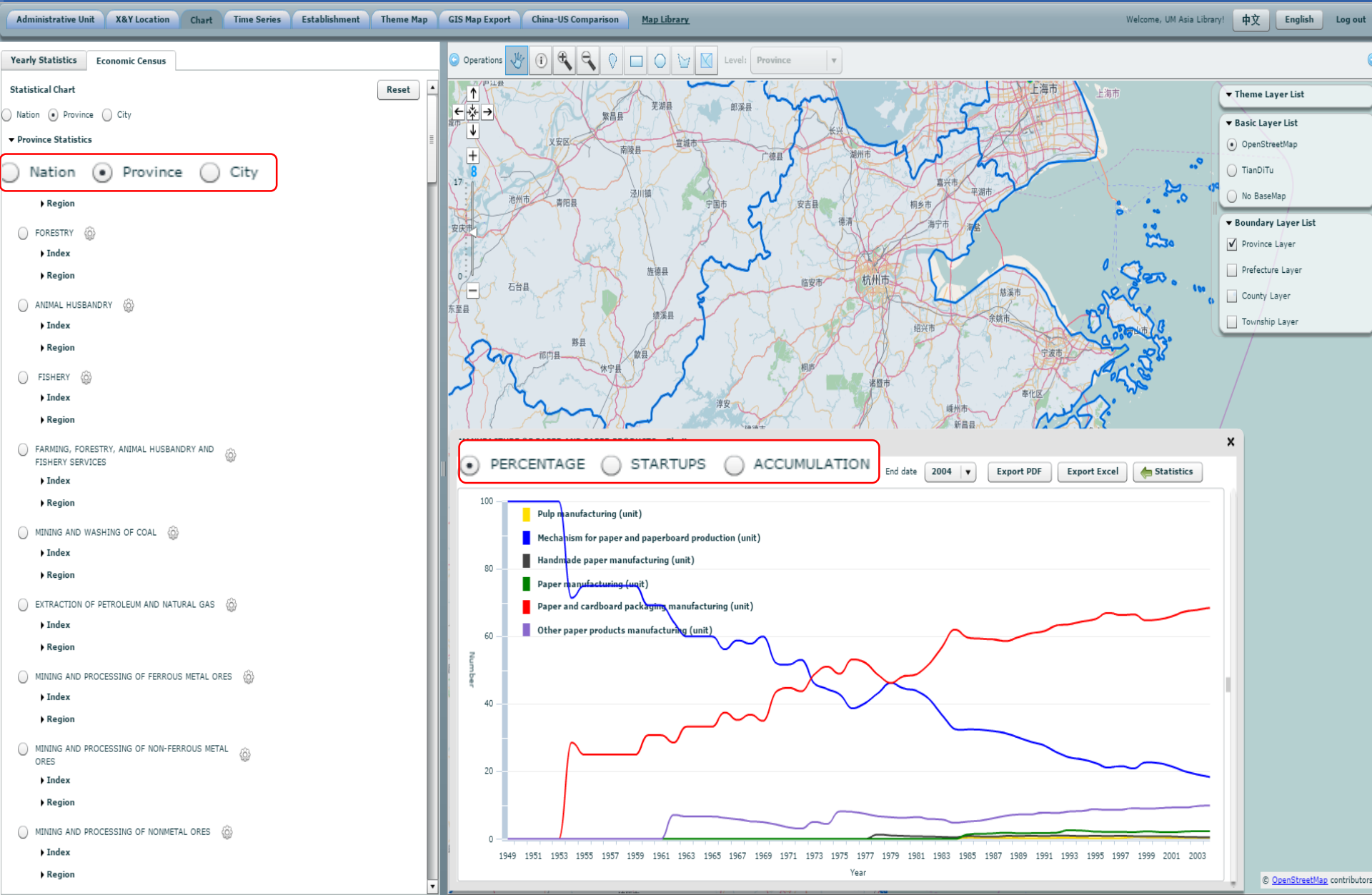


Location Analysis

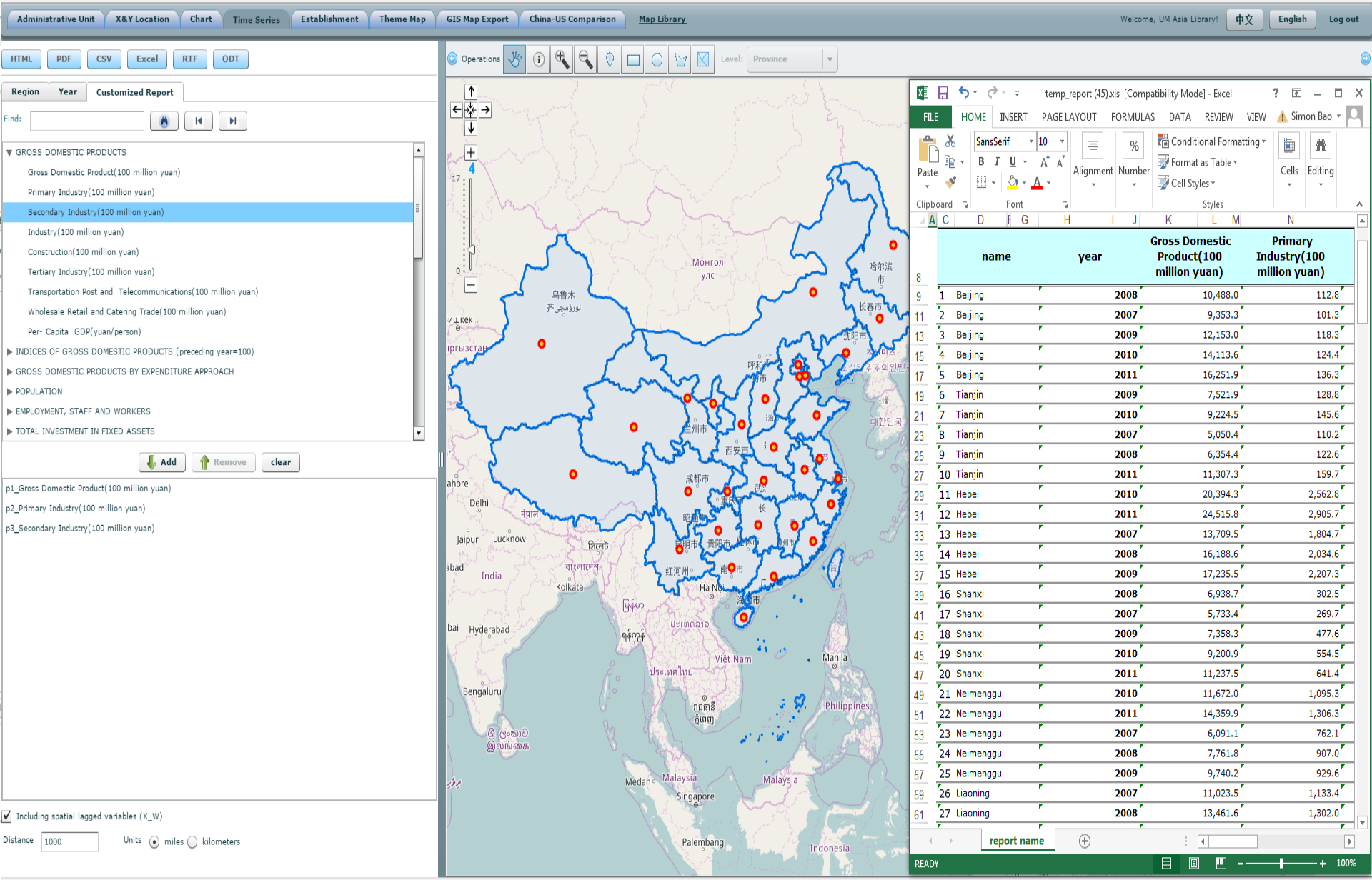


	1(1mile)		1(3mile)		1(5mile)		2(1mile)		2(3mile)		2(5mile)		3(1mile)		3(3mile)		3(5mile)	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Total Households	10,138	100.0 %	28,607	100.0 %	36,184	100.0 %	3,028	100.0 %	15,875	100.0 %	26,695	100.0 %	7,858	100.0 %	78,632	100.0 %	108,056	100.0 %
Family Households	9,567	94.4 %	27,338	95.6 %	34,895	96.4 %	2,965	97.9 %	15,577	98.1 %	26,278	98.4 %	7,599	96.7 %	75,882	96.5 %	104,592	96.8 %
Collective Households	571	5.6 %	1,269	4.4 %	1,289	3.6 %	63	2.1 %	298	1.9 %	416	1.6 %	259	3.3 %	2,749	3.5 %	3,464	3.2 %
Total Population	32,165	100.0 %	91,631	100.0 %	116,927	100.0 %	9,032	100.0 %	47,822	100.0 %	81,098	100.0 %	23,920	100.0 %	240,247	100.0 %	331,988	100.0 %
Agricultural Population	7,976	24.8 %	37,043	40.4 %	61,278	52.4 %	3,740	41.4 %	24,657	51.6 %	48,581	59.9 %	4,879	20.4 %	44,347	18.5 %	93,164	28.1 %
Non-Agricultural Population	24,188	75.2 %	54,587	59.6 %	55,648	47.6 %	5,291	58.6 %	23,164	48.4 %	32,516	40.1 %	19,041	79.6 %	195,900	81.5 %	238,823	71.9 %
Population in Family Households	28,410	87.8 %	83,413	90.6 %	108,732	92.5 %	8,693	95.2 %	46,174	95.7 %	78,859	96.5 %	22,601	94.2 %	227,616	94.3 %	317,608	95.2 %

Charts Analysis (Structural Analysis)

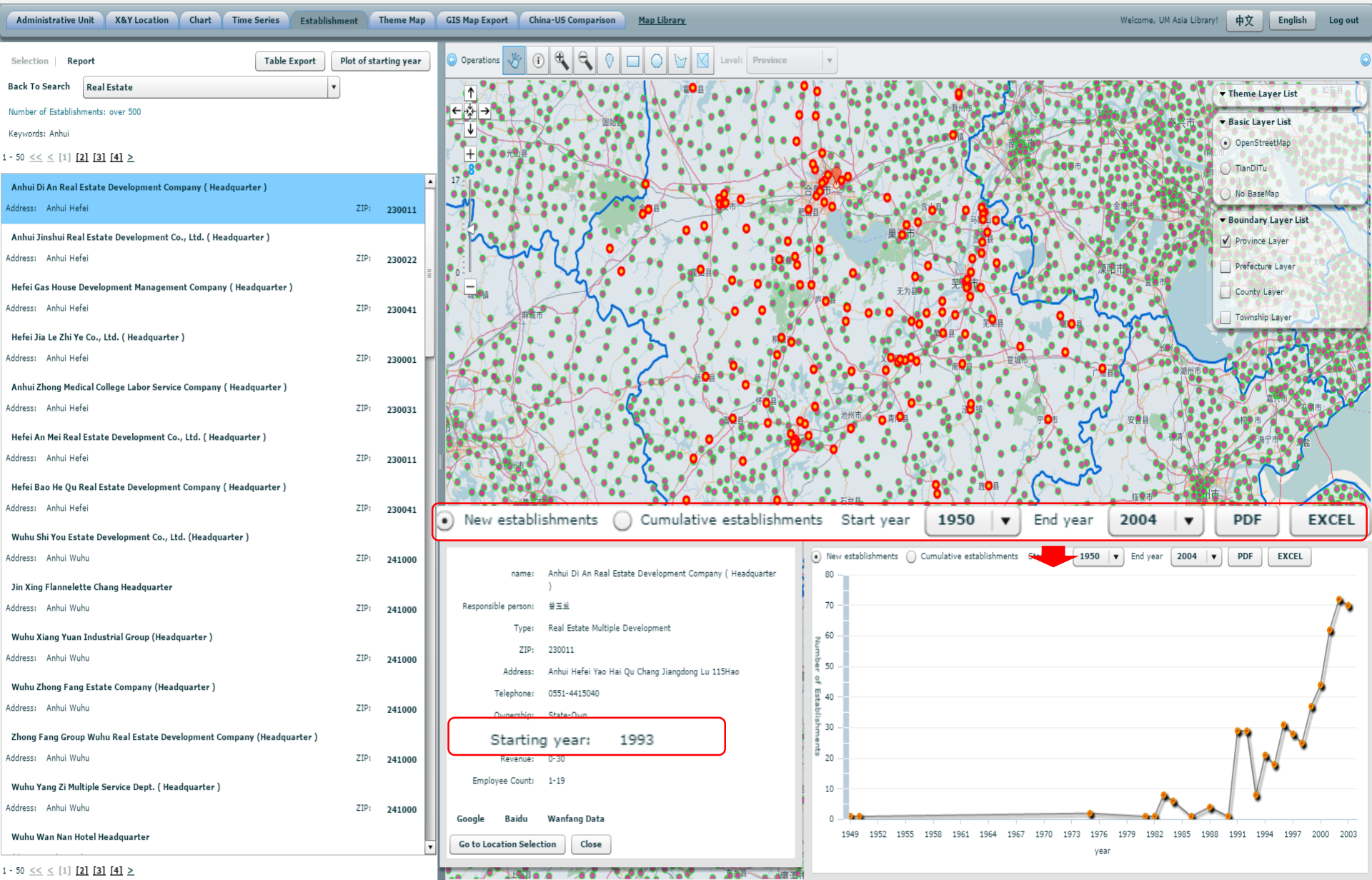


Time Series Analysis



	name	year	Gross Domestic Product(100 million yuan)	Primary Industry(100 million yuan)
1	Beijing	2008	10,488.0	112.8
2	Beijing	2007	9,353.3	101.3
3	Beijing	2009	12,153.0	118.3
4	Beijing	2010	14,113.6	124.4
5	Beijing	2011	16,251.9	136.3
6	Tianjin	2009	7,521.9	128.8
7	Tianjin	2010	9,224.5	145.6
8	Tianjin	2007	5,050.4	110.2
9	Tianjin	2008	6,354.4	122.6
10	Tianjin	2011	11,307.3	159.7
11	Hebei	2010	20,394.3	2,562.8
12	Hebei	2011	24,515.8	2,905.7
13	Hebei	2007	13,709.5	1,804.7
14	Hebei	2008	16,188.6	2,034.6
15	Hebei	2009	17,235.5	2,207.3
16	Shanxi	2008	6,938.7	302.5
17	Shanxi	2007	5,733.4	269.7
18	Shanxi	2009	7,358.3	477.6
19	Shanxi	2010	9,200.9	554.5
20	Shanxi	2011	11,237.5	641.4
21	Neimenggu	2010	11,672.0	1,095.3
22	Neimenggu	2011	14,359.9	1,306.3
23	Neimenggu	2007	6,091.1	762.1
24	Neimenggu	2008	7,761.8	907.0
25	Neimenggu	2009	9,740.2	929.6
26	Liaoning	2007	11,023.5	1,133.4
27	Liaoning	2008	13,461.6	1,302.0

Establishment



Panel Data for Spatial Modeling

Inbox (17,603) - sbao@ | Geo-Explorer | chinageoexplorer.org/c | chinageoexplorer.org/r | 141.211.24.149/SIServe

Not secure | 141.211.24.149/SIServerResearch//tempReport/597336bd-e3a1-4a8e-8835-65619db5d2e9/temp_report.html

Time-series Report

Date: 19/08/2018

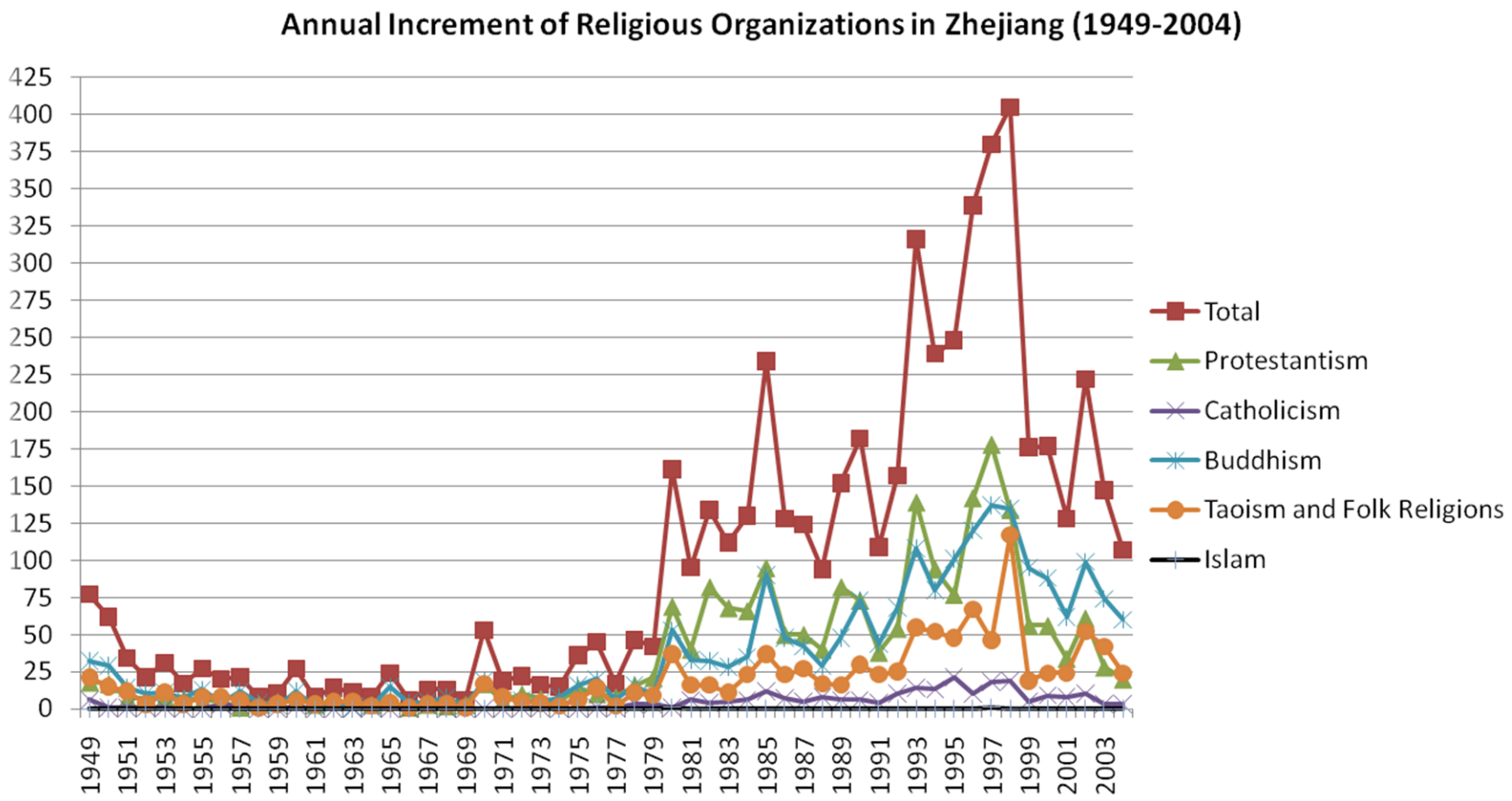
Selection: (4 selected) zhejiang , tianjin , y

Original Variables | **Spatially Weighted Variables**

	name	year	Gross Domestic	Primary Industry	Secondary Industry	Gross Domestic	Primary Industry	Secondary Industry
1	tianjin	2009	7,521.9	128.8	3,987.8	15,159.8	1,324.7	7,640.2
2	tianjin	2010	9,224.5	145.6	4,840.2	18,081.4	1,526.9	9,226.2
3	tianjin	2006	4,344.3	118.2	2,488.3	9,570.7	918.0	4,939.7
4	tianjin	2007	5,050.4	110.2	2,892.5	11,455.6	1,050.3	5,902.5
5	tianjin	2008	6,354.4	122.6	3,821.1	13,626.3	1,241.1	7,117.9
6	tianjin	2011	11,307.3	159.7	5,928.3	21,441.7	1,762.1	10,970.9
7	zhejiang	2010	27,722.3	1,360.6	14,297.9	17,809.7	1,551.2	8,940.6
8	zhejiang	2011	32,318.8	1,583.0	16,555.6	21,163.0	1,798.8	10,653.4
9	zhejiang	2006	15,742.5	925.1	8,509.6	9,421.9	946.0	4,764.4
10	zhejiang	2007	18,780.4	986.0	10,148.5	11,259.4	1,083.4	5,669.9
11	zhejiang	2008	21,486.9	1,095.4	11,580.3	13,348.6	1,287.2	6,812.6
12	zhejiang	2009	22,990.3	1,163.1	11,908.5	14,943.3	1,347.8	7,377.5
13	yunnan	2010	7,224.2	1,108.4	3,223.5	12,753.8	1,372.5	6,304.8
14	yunnan	2011	8,893.1	1,411.0	3,780.3	15,398.3	1,639.1	7,734.8
15	yunnan	2006	3,981.3	749.8	1,712.6	6,481.9	835.0	3,042.2
16	yunnan	2007	4,741.3	837.4	2,051.1	7,801.2	985.9	3,683.3
17	yunnan	2008	5,700.1	1,020.9	2,451.1	9,276.7	1,187.1	4,471.8
18	yunnan	2009	6,169.8	1,067.6	2,582.5	10,583.4	1,190.0	5,030.8
19	xinjiang	2008	4,203.4	691.1	2,086.7	678.7	83.0	322.6
20	xinjiang	2006	3,045.3	527.8	1,459.3	465.3	60.3	205.6
21	xinjiang	2007	3,523.2	628.7	1,647.5	562.9	69.2	258.1
22	xinjiang	2009	4,277.1	759.7	1,929.6	761.3	85.6	356.0
23	xinjiang	2010	5,437.5	1,078.6	2,592.2	929.0	101.8	454.3
24	xinjiang	2011	6,610.1	1,139.0	3,225.9	1,138.1	114.8	592.0

ENG 22:29 2018/8/19

Identification of Possible Policy Impacts



chinadatalab.net

China Data Lab

Participated by

China Data Institute 中国数据研究所 (chinadatacenter.net)
Center for Geographical Analysis, Harvard University 哈佛大学地理分析中心
All China Marketing Research, Ltd. 华通人信息技术有限公司
China Consortium for Finance and Economics Education 中国财经教育资源共享联盟 (knowledgeatshare.cn)
Business School, East China University of Science and Technology 华东理工大学商学院 (bs.ecust.edu.cn/bsweb2016en)
Geo-computation Center for Social Sciences, Wuhan University 武汉大学社会地理计算联合中心(www.lmars.whu.edu.cn/gcss/index.php/en)

Challenges for Data Research and Teaching

- **Data Sharing**
  - Licensed data
  - Restricted data
  - Sensitive data
  - Large size data
  - Research data generated from different projects
- **Tool Sharing**
  - Licensed and free tools
  - Integrated environment for tools for data
  - Maintenance and updates
- **Research Results Sharing （Differentiated training and teaching)**
  - Research (**reproducible, replicable, generalizable**)
  - Teaching (students with different interests and skills)
  - Decision support (efficient, effective, and expandable)

Solution: Cloud Based Platform

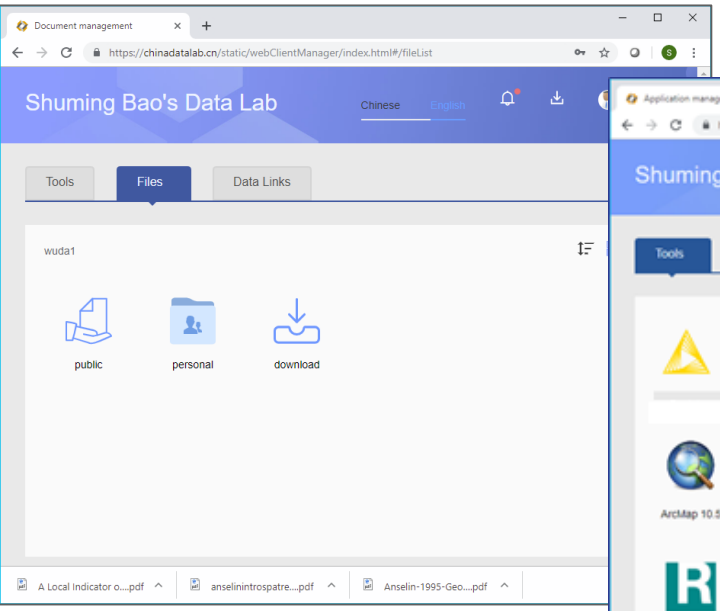
- **A data center** for China studies based on cloud
- **A research base** for collaborations on China studies
- **A development center** for data case studies
- **A training center** for China studies, including theory, methodology, technology, data and applications for research and teaching

An Integrated Platform for Research and Teaching

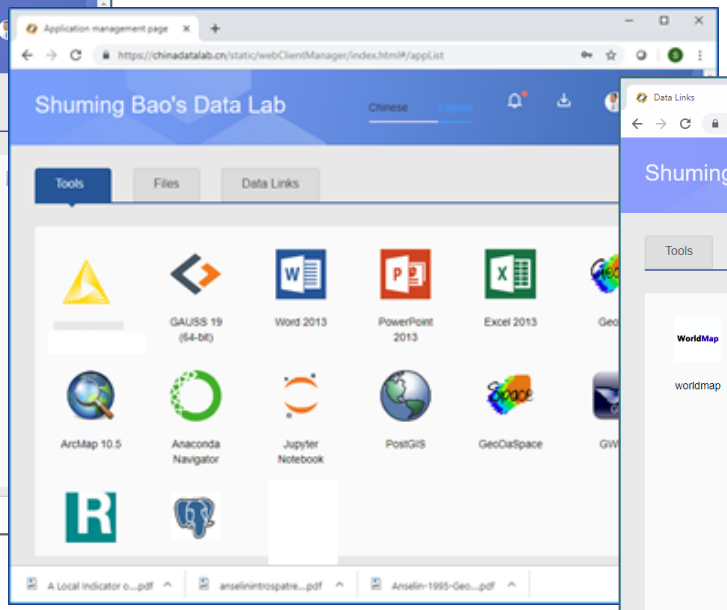
Main Features:

- ❑ Data available only on the cloud
- ❑ Tools available on the cloud
- ❑ All computation are on the cloud
- ❑ No maintenance required for end users

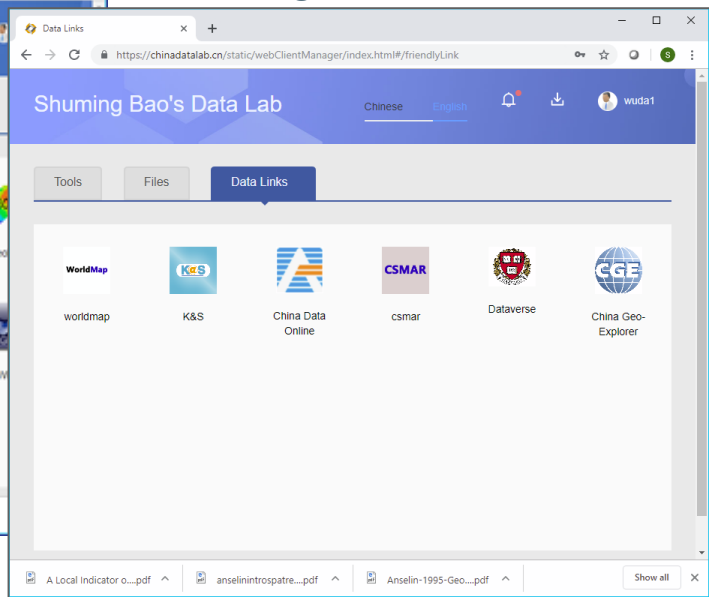
Data Center



Tool Center



Sharing Center



China Data Lab

http://chinadatalab.cn

GAUSS

GeoDa

GWR4

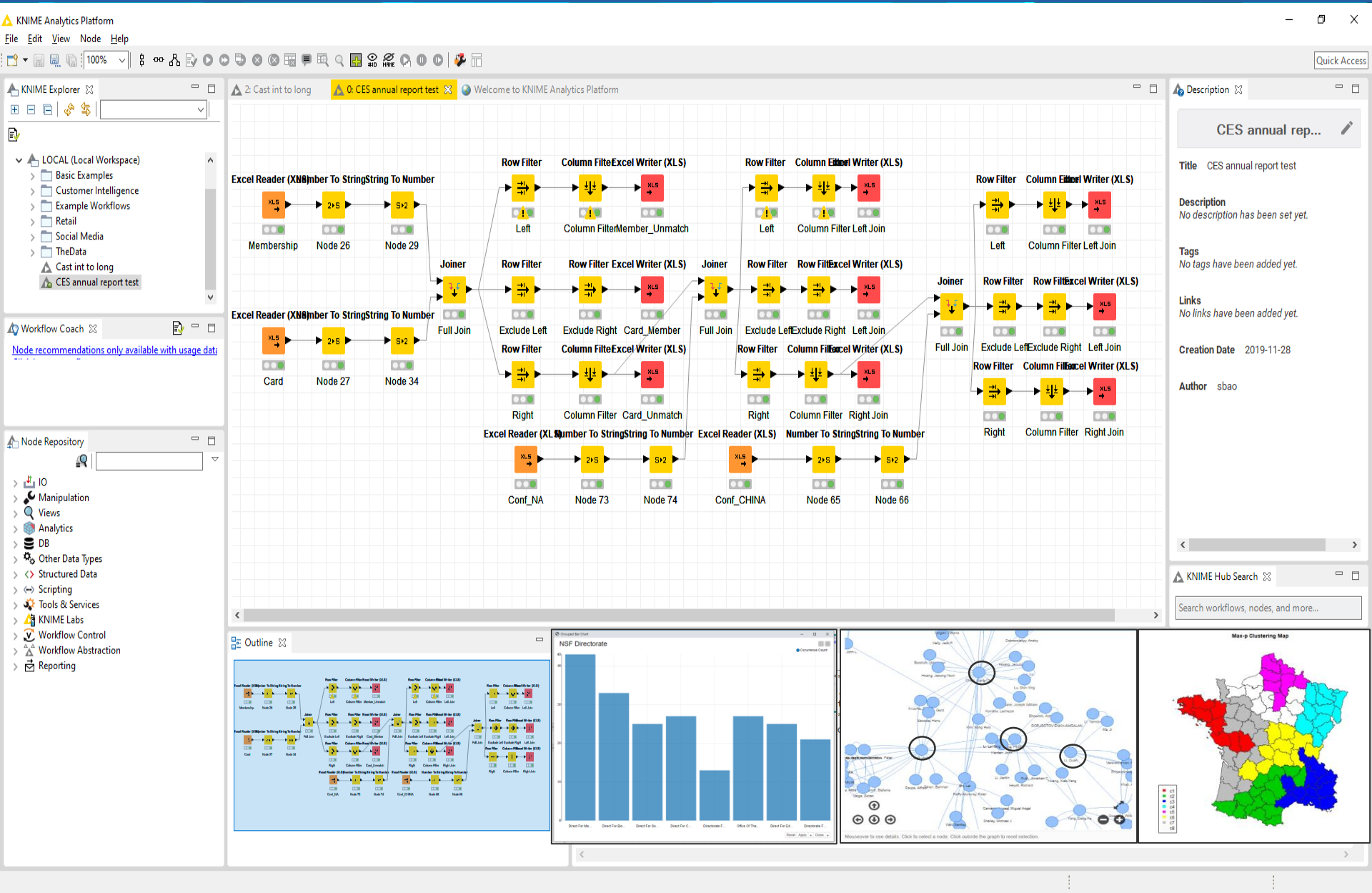
ArcMap

Jupyter

RAnalyticFlow

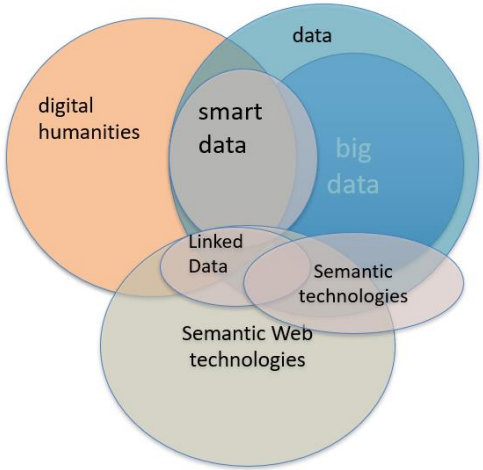
KNIMEAnalytics

Workflow Based Data Analysis



Case Study I: Literature Analysis with KNIME

- **Goal**: develop and demonstrate a network framework of the historical Innovation and Invention at the Liquid Crystal Institute, Kent State University (PI: Marcia Lei Zeng, et al.)



References:

- Li, H., Zeng, M., Zhang, Y., Ye, X., & Hu, T. (2017). Tackling Innovation Networks with Smart Data: A Case Study of the Liquid Crystal Institute at Kent State University. In DH.
- Zeng, M. L., Zhang, Y., Li, H., & Polyakov, S. (2015). Exploring Smart Data Approaches to the History of Innovation and Invention at Liquid Crystal Institute at Kent State University. In Digital Libraries: Providing Quality Information: The 17th International Conference on Asia-Pacific Digital Libraries, ICADL 2015, Seoul, Korea, December 9-12, 2015. Proceedings (Vol. 9469, p. 346). Springer.

Objectives

- Replicate data analysis procedures using previous scientific literature data based on workflow;
- Expanded data analysis based on publication, patent, and NSF grant data;
- Applications of workflow for research and teaching related to network analysis based on publication, patent, grant data, as well as other data.

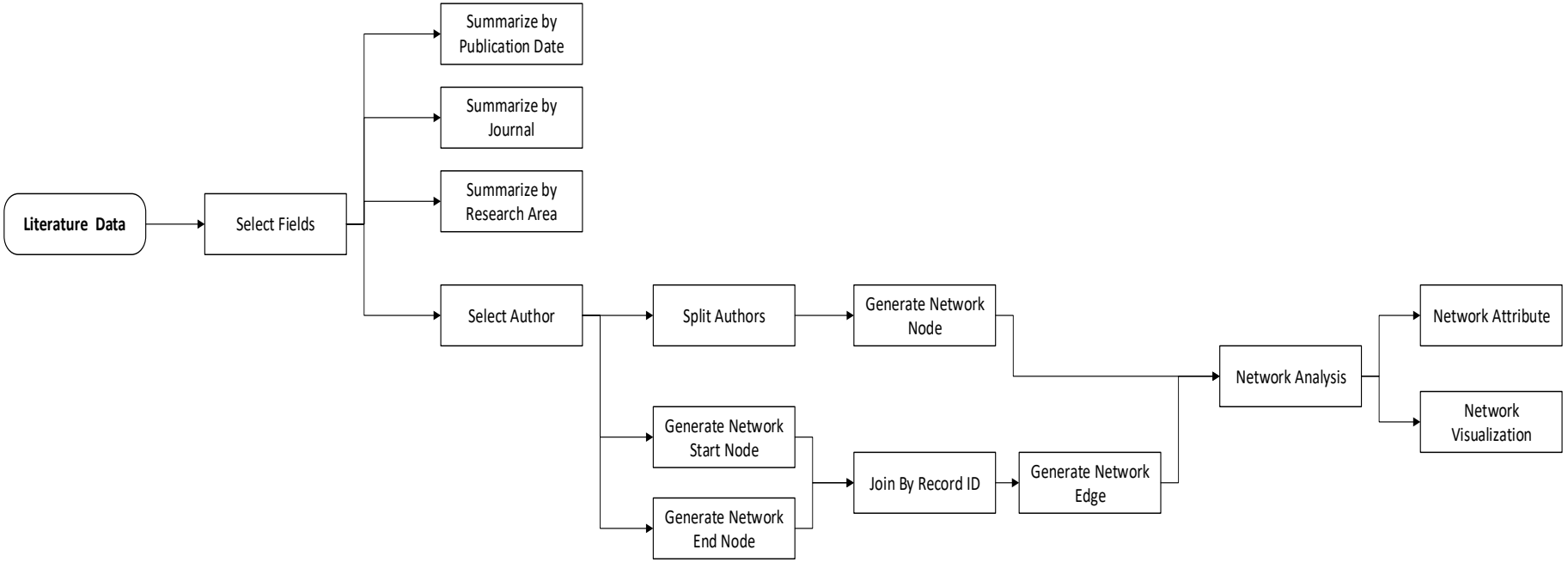
Data Sources

- **Publication Data**
  - Title
  - Author
  - Affiliation
  - Key words
  - Abstract
  - Publication Date
  - Journal
  - Volume
  - Issue
  - …
- **Patent Data**
  - Title
  - Inventor
  - Inventor Location
  - Publication Date
  - Assignee
  - Assignee Location
  - CPC
  - IPC
  - USPC
  - Abstract
  - …
- **Awarded Grants**
  - Title
  - PI
  - Co-PI
  - Email Address
  - Institution
  - NSF Organization
  - Start Date
  - Expiration Date
  - Awarded Amount
  - NSF Directorate

Data Input

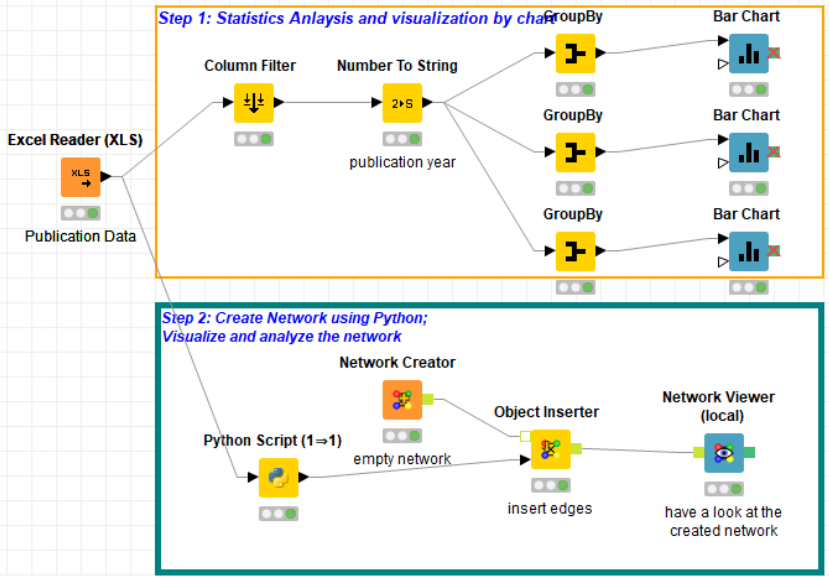
File Name	Format	Data Sources
publication.xls	.xls	Web of Science
patent.xls	.xls	ProQuest
grant.xls	.xls	NSF website

The Flowchart for Data Analysis

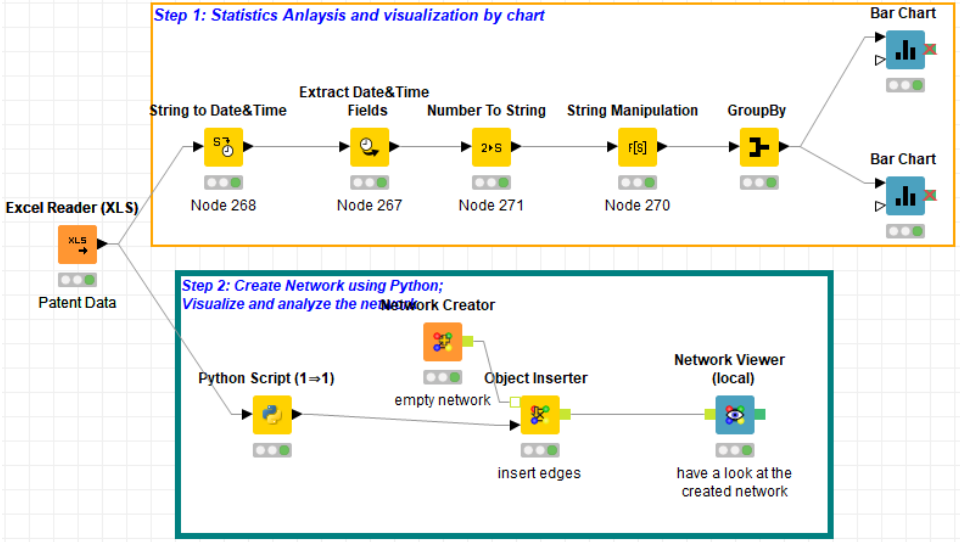


Knime Workflow for Literature Analysis

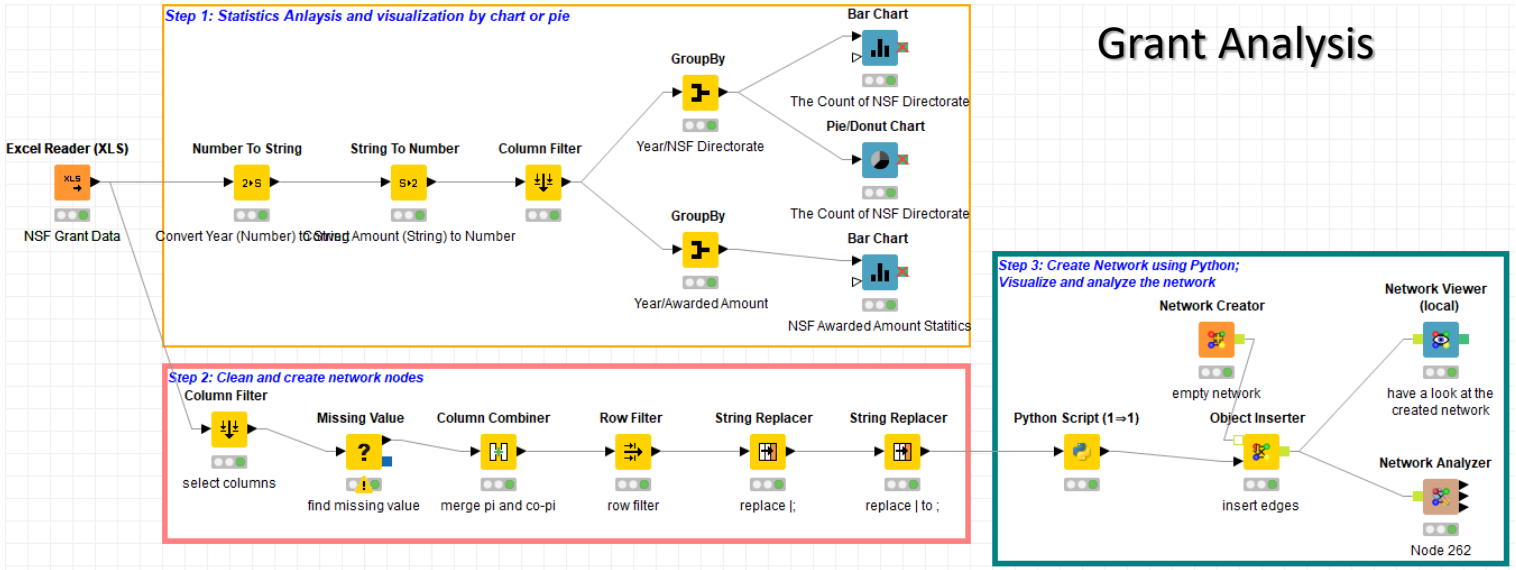
Publication Analysis



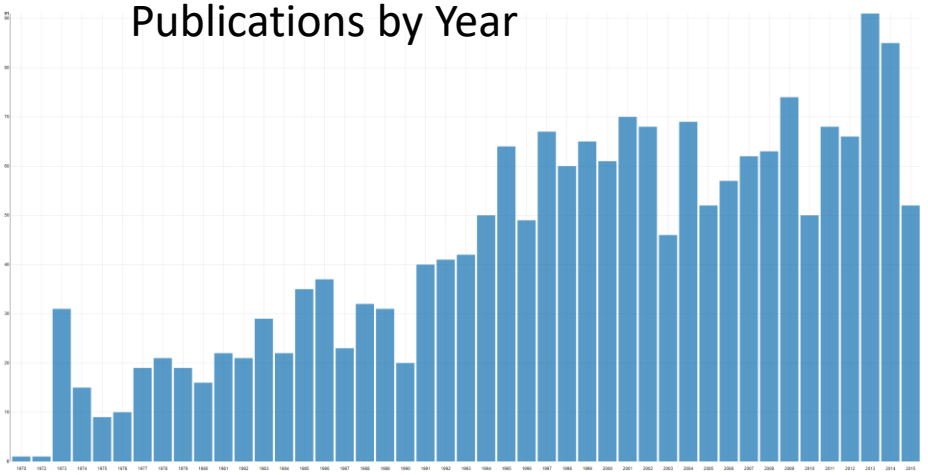
Patent Analysis

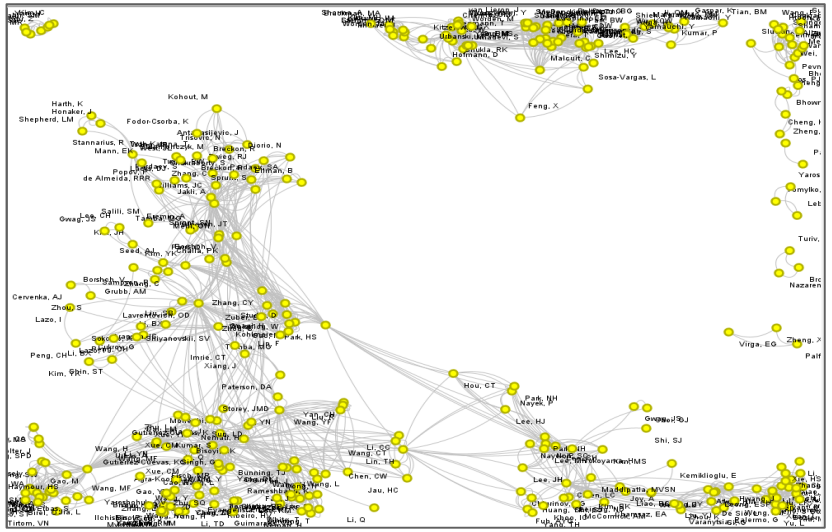


Grant Analysis



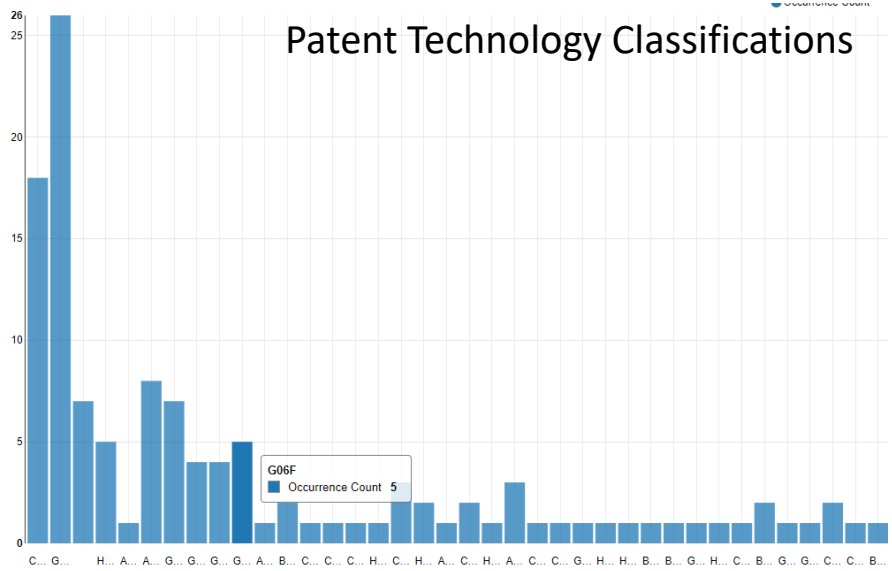
Results from Publication Analysis



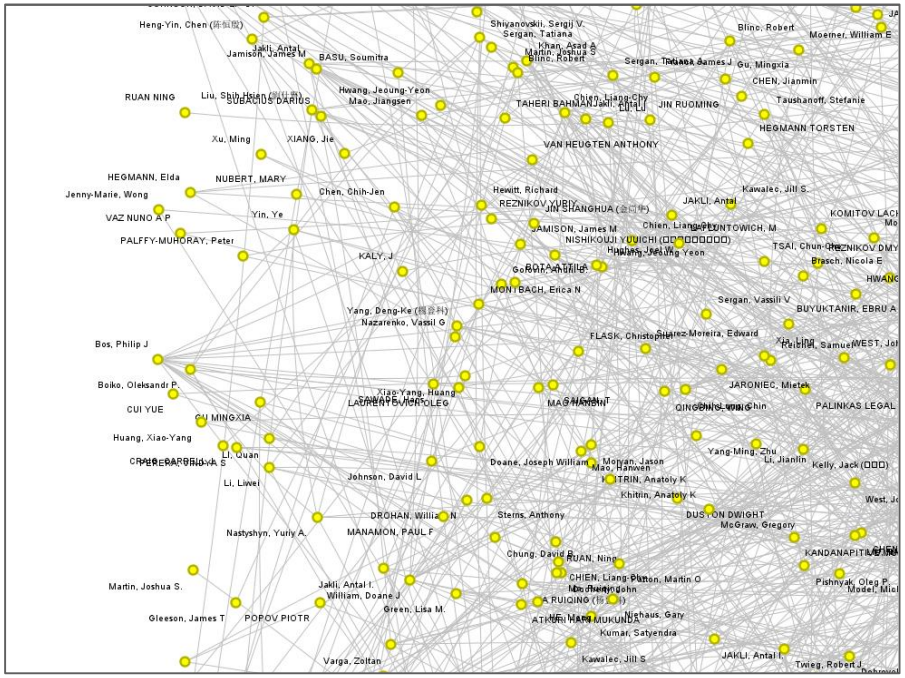


Object id	Node d...	Node d...	In degree	In degr...	Out de...	Out de...	Closen...	Node w...	Avg. n...	Clusteri...	Hub score	Authori...	Baryce...
Feng, X	5	0.326	5	0.326	5	0.326	0.483	5	1	1	0.069	0.069	0.008
Wang, F	5	0.326	5	0.326	5	0.326	0.621	5	1	1	0	0	0.04
Sharma, A	22	1.433	22	1.433	22	1.433	0.589	22	1	0.706	0.801	0.801	0.009
Yao, WH	8	0.521	8	0.521	8	0.521	0.292	8	1	1	0	0	0.001
Wang, MF	5	0.326	5	0.326	5	0.326	0.32	5	1	0.9	0	0	0.001
Lu, W	9	0.586	9	0.586	9	0.586	0.49	9	1	1	0.091	0.091	0.007
Wang, H	4	0.261	4	0.261	4	0.261	0.32	4	1	1	0	0	0.001
Antanasijev...	7	0.456	7	0.456	7	0.456	0.287	7	1	1	0	0	0.001
Reich, R	8	0.521	8	0.521	8	0.521	0.923	8	1	1	0	0	0.1
Lu, L	8	0.521	8	0.521	8	0.521	0.917	8	1	1	0	0	0.1
Lebovka, N	2	0.13	2	0.13	2	0.13	1	2	1	1	0	0	0.5
Ma, J	21	1.368	21	1.368	21	1.368	0.258	21	1	0.367	0	0	0.001
Malgras, V	7	0.456	7	0.456	7	0.456	0.375	7	1	1	0.01	0.01	0.006
Park, HS	23	1.498	23	1.498	23	1.498	0.405	23	1	0.526	0	0	0.002
Beltrano, G	18	1.173	18	1.173	18	1.173	0.517	18	1	1	0.776	0.776	0.008
Kohlmeier, A	15	0.977	15	0.977	15	0.977	0.363	15	1	1	0	0	0.002
Sampson, P	5	0.326	5	0.326	5	0.326	0.327	5	1	1	0	0	0.001
Umadevi, S	8	0.521	8	0.521	8	0.521	0.497	8	1	0.571	0.079	0.079	0.008
Moheghi, A	2	0.13	2	0.13	2	0.13	0.286	2	1	1	0	0	0.001

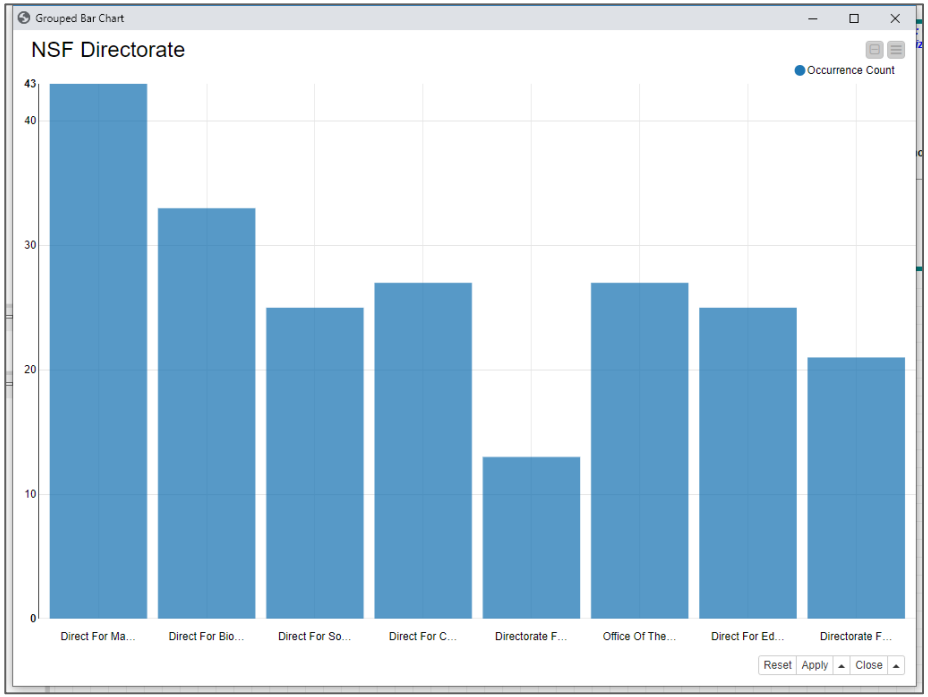
Results from Patent Analysis

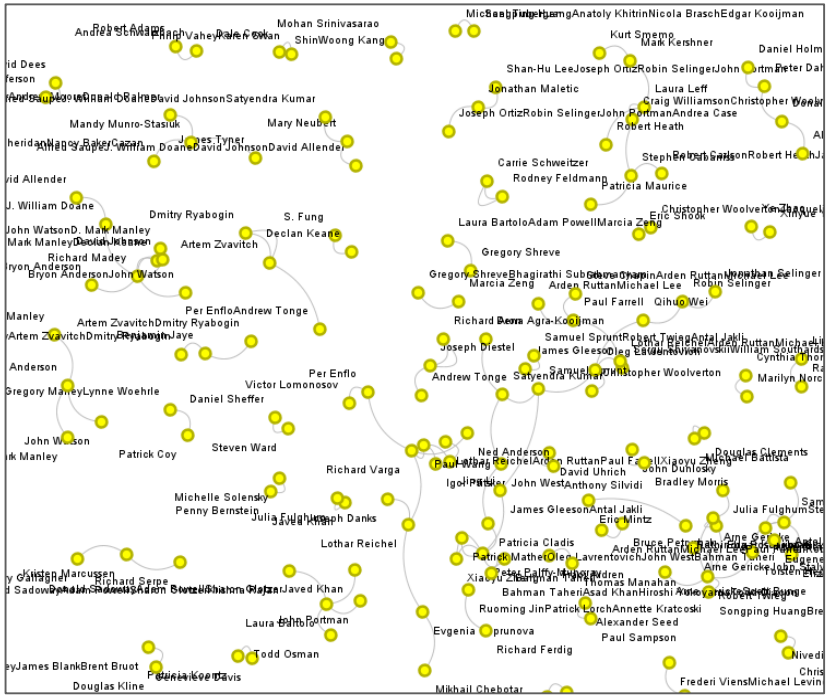


Object id	Node d...	Node d...	In degree	In degr...	Out de...	Out de...	Node w...	Avg. n...
Chao-Chiun, Liang	9	0.771	9	0.771	9	0.771	9	1
SHOKOUHIMEH...	5	0.428	5	0.428	5	0.428	5	1
Nemati, Hossein ...	9	0.771	9	0.771	9	0.771	9	1
Hwang, Jeoung ...	3	0.257	3	0.257	3	0.257	3	1
QIAN LIANGQI (...	2	0.171	2	0.171	2	0.171	2	1
Soehnlen, Eric S...	8	0.686	8	0.686	8	0.686	8	1
Chen, Cheng	4	0.343	4	0.343	4	0.343	4	1
Osher, Lawrence	9	0.771	9	0.771	9	0.771	9	1
Nastyshyn, Yuri...	6	0.514	6	0.514	6	0.514	6	1
Bhowmik, Achint...	2	0.171	2	0.171	2	0.171	2	1
Gleeson, James T	1	0.086	1	0.086	1	0.086	1	1
Dobrovolskyy, A...	3	0.257	3	0.257	3	0.257	3	1
GLEESON, Jame...	3	0.257	3	0.257	3	0.257	3	1
Kelly, Jack (□□□)	4	0.343	4	0.343	4	0.343	4	1
Li, Liwei	5	0.428	5	0.428	5	0.428	5	1
Palffy-Muhoray,...	7	0.6	7	0.6	7	0.6	7	1
Tsai, Chen Chu (...	9	0.771	9	0.771	9	0.771	9	1



Results from Grant Analysis

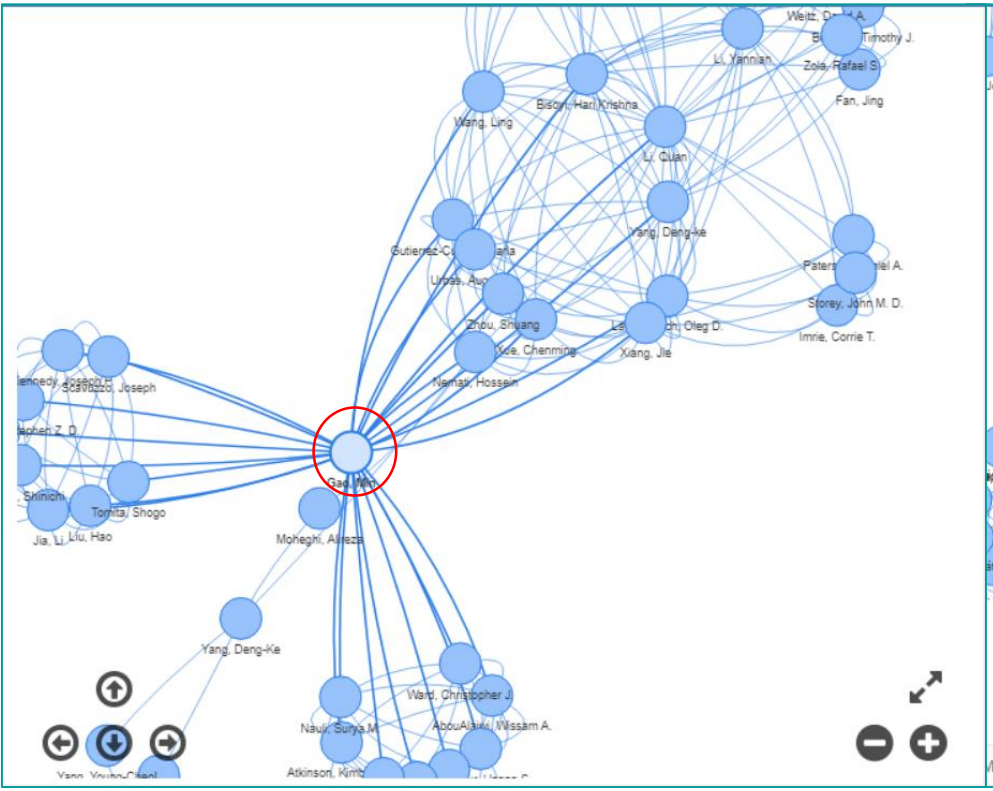




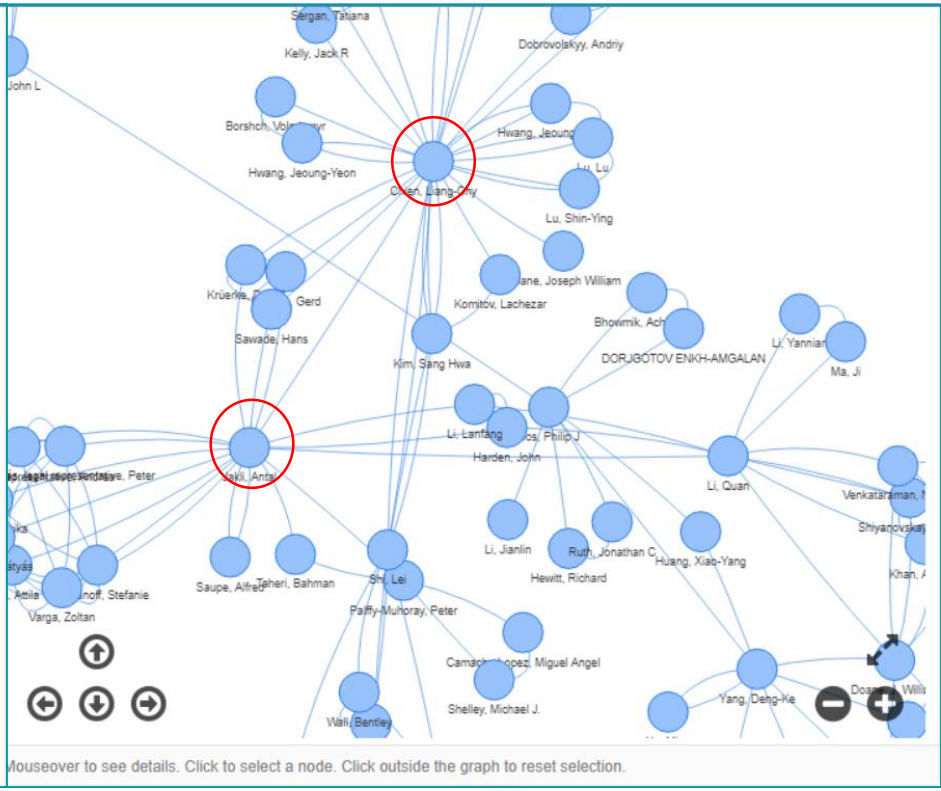
Object id	Node d...	Node degree %	In degree	In degree %	Out degree	Out degree %	Node weigh...	Avg. n...
Genevieve Davis	1	0.787	1	0.787	1	0.787	1	1
Gerassimos Pe...	1	0.787	1	0.787	1	0.787	1	1
Xiaoyu Zheng	2	1.575	2	1.575	2	1.575	2	1
Noah FriedkinE...	1	0.787	1	0.787	1	0.787	1	1
William Kalkhoff	1	0.787	1	0.787	1	0.787	1	1
Robin Selinger	2	1.575	2	1.575	2	1.575	2	1
Paul Farrell	3	2.362	3	2.362	3	2.362	3	1
Joseph OrtizD...	1	0.787	1	0.787	1	0.787	1	1
Ben FinneyMar...	1	0.787	1	0.787	1	0.787	1	1

Network Analysis of Scholars

Authors of Publications



Participants of Grants



Case Study II: Financial Analysis with Firm Data

Going public in China: Reverse mergers versus IPOs

Aim: This study examines the decision to go public in China through an initial public offering (IPO) versus a reverse merger (RM) transaction.

ELSEVIER

Journal of Corporate Finance
Volume 58, October 2019, Pages 92-111

Going public in China: Reverse mergers versus IPOs ☆

Charles M.C. Lee [a], Yuanyu Qu [b], Tao Shen [c]

Show more

https://doi.org/10.1016/j.jcorpfin.2019.04.003

Get rights and content

Highlights

- We study firms' choice to go public through reverse mergers (RMs) versus initial public offerings (IPOs) in China.
- Pre-listing RM firms are larger, more profitable, and less politically-connected than pre-listing IPO firms.
- RM firms also have superior post-listing performance, both in terms of operations and stock returns.
- These results are in sharp contrast to the evidence on RMs from developed countries.

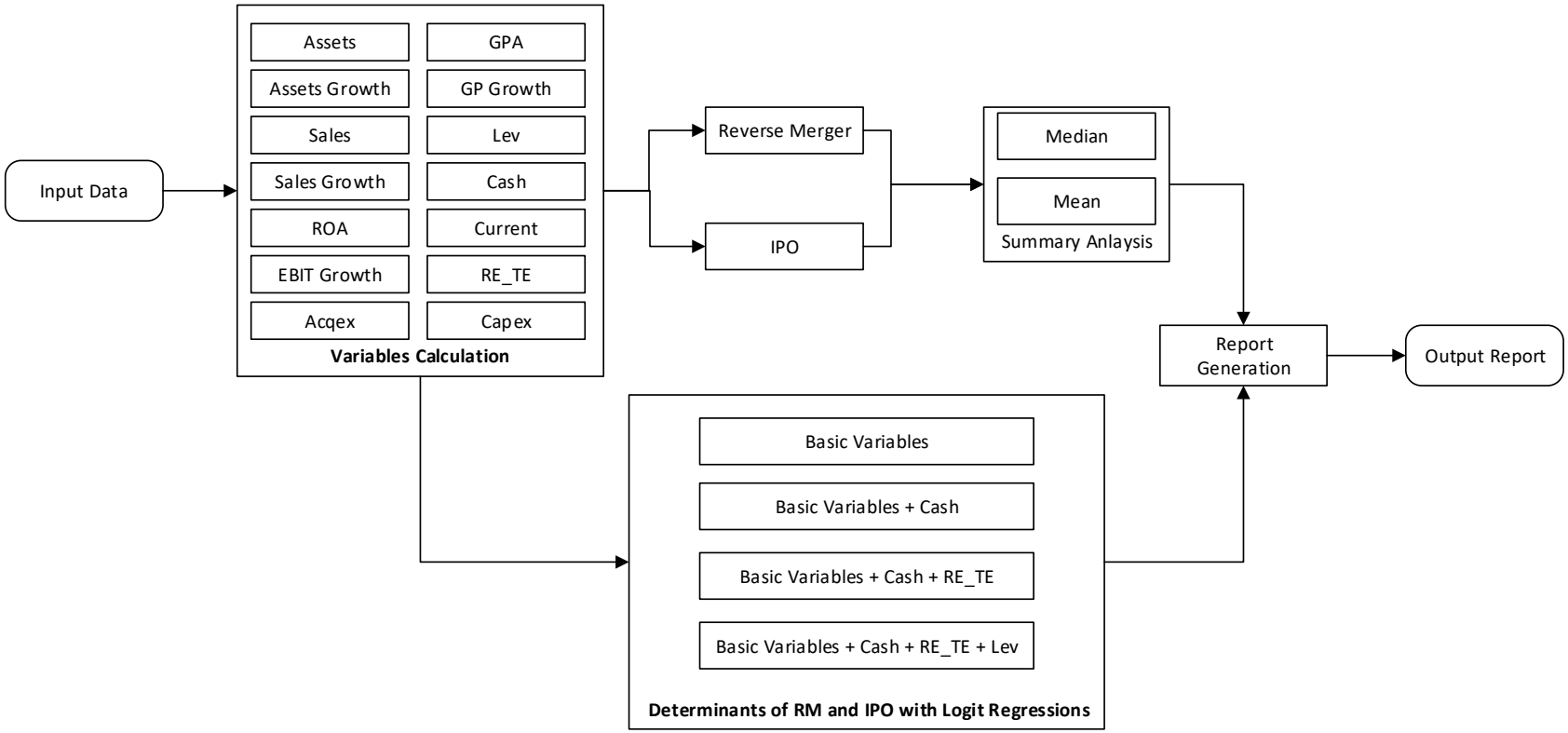
Abstract

We study firms that go public through reverse mergers (RMs) versus initial public offerings (IPOs) in China. Using a manually assembled data set, we show that pre-listing RM firms are larger, more profitable, and less politically connected than pre-listing IPO firms. Chinese RM firms also have superior post-listing performance, in terms of both operations and stock returns, compared to IPOs matched on industry and size. Unlike IPOs, RM firms do not underperform the market in the long run. These results are in sharp contrast to the evidence on RMs from developed countries. We trace these differences to China's stringent and potentially biased IPO policies, which appear to preclude even high-quality firms from accessing public markets.

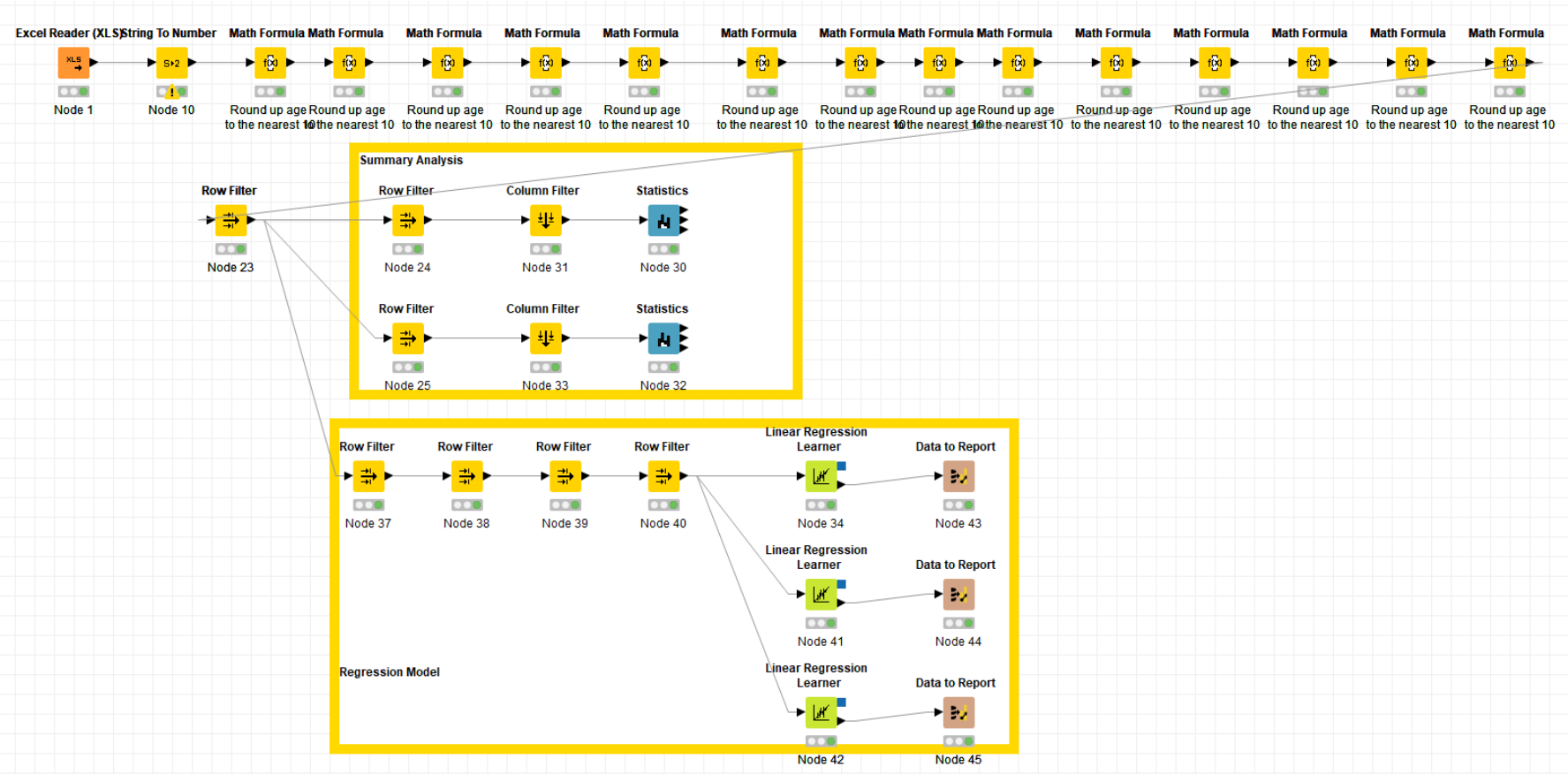
Data Sources

Name	Format	Description
IPO.xlsx	xlsx	The financial and stock returns data of listed firms are from the China Stock Market and Accounting Research (CSMAR) Database
RM.xlsx	xlsx	The data is from the iFinD database provided by Tong Hua Shun (THS), a major financial data service company in China
Firm.xlsx	xlsx	The financial information on each RM proposal from www.cninfo.com.cn, a CSRC-authorized website that archives documents and filings for listed firms

Flow Chart for Data Analysis



KNIME Workflow for Data Analysis



Results from Data Analysis

Summary Statistics of RM and IPO

Row ID	Column	Min	Max	Mean	Std. deviation	Variance	Overall ...	Histogram
GPA	GPA	-0.007	0.638	0.162	0.141	0.02	17.024	-0 0.6
ROA	ROA	-0.157	0.42	0.085	0.099	0.01	9.094	-0.2 0.4
GP Growth	GP Growth	-627....	56,679....	1,260.835	7,203.23	51,886,524....	132,387.667	-627 56,679
assets	assets	309,7...	359,80...	18,987,...	43,036,002,985....	1,852,097,5...	2,088,583,6...	3E8 4E11
assets gr...	assets gr...	-0.299	376.438	4.996	36.275	1,315.889	549.613	-0 376
sales	sales	12,65...	57,482,...	5,055,5...	8,239,117,930.447	67,883,064,...	530,827,55...	1E7 6E10
sales growth	sales gro...	-0.918	87.484	2.938	12.784	163.419	323.195	-1 87
EBIT	EBIT	-476,...	10,705,...	805,421...	1,450,585,376.419	2,104,197,9...	86,180,152,...	-5E8 1E10
lev	lev	0.045	0.956	0.492	0.235	0.055	54.115	0 1

Results from Logit Regressions

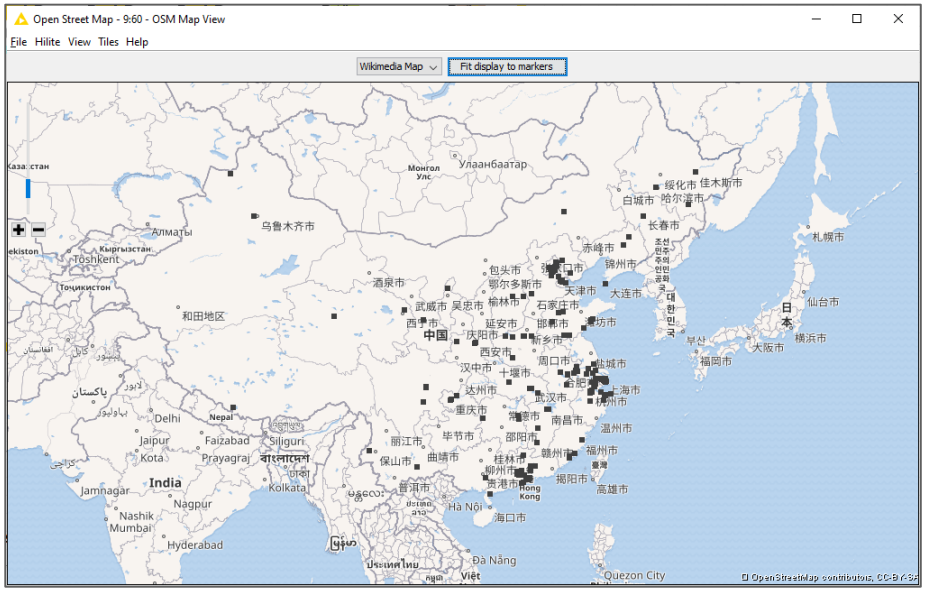
Variable	Coeff.	Std. Err.	t-value	P>|t|
GPA	-0.197	0.117	-1.687	0.092
ROA	0.624	0.19	3.287	0.001
assets	0	0	1.458	0.145
assets growth	-0	0.001	-0.078	0.938
sales	0	0	0.12	0.905
sales growth	0.012	0.002	5.151	0
EBIT	-0	0	-0.792	0.428
current	-0.029	0.085	-0.336	0.737
re_te	0.018	0.005	3.762	0
acqex	-0	0	-0.604	0.546
capex	0	0	2.998	0.003
net payout	-0	0	-1.294	0.196
Intercept	0.08	0.02	4.044	0

Variable	Coeff.	Std. Err.	t-value	P>|t|
GPA	-0.164	0.117	-1.403	0.161
ROA	0.67	0.19	3.519	0
assets	0	0	0.868	0.385
assets growth	0	0.001	0.056	0.955
sales	-0	0	-0.176	0.86
sales growth	0.012	0.002	4.927	0
EBIT	-0	0	-0.767	0.443
lev	0.129	0.052	2.477	0.013
current	0.014	0.087	0.159	0.873
re_te	0.018	0.005	3.707	0
acqex	-0	0	-0.445	0.657
capex	0	0	2.677	0.008
net payout	-0	0	-0.635	0.526
AO	0	0	0.186	0.852
Intercept	0.011	0.034	0.31	0.757

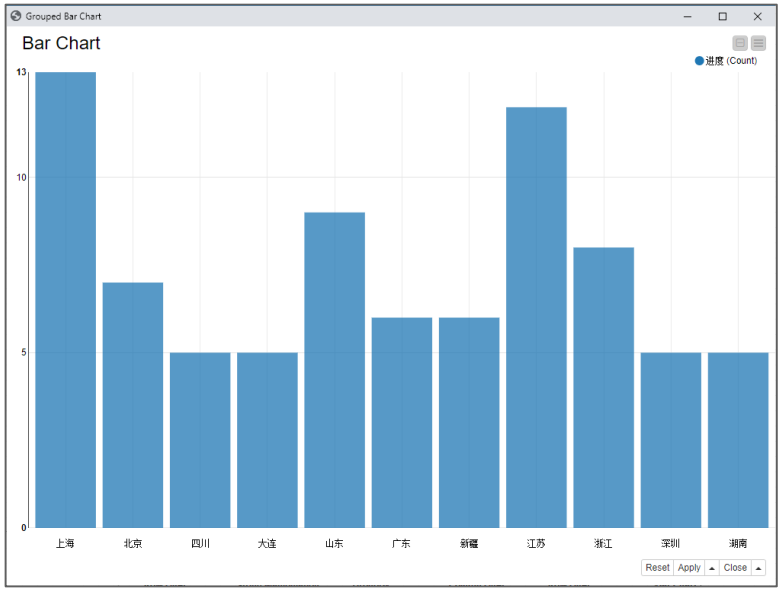
*0:保留三位小数

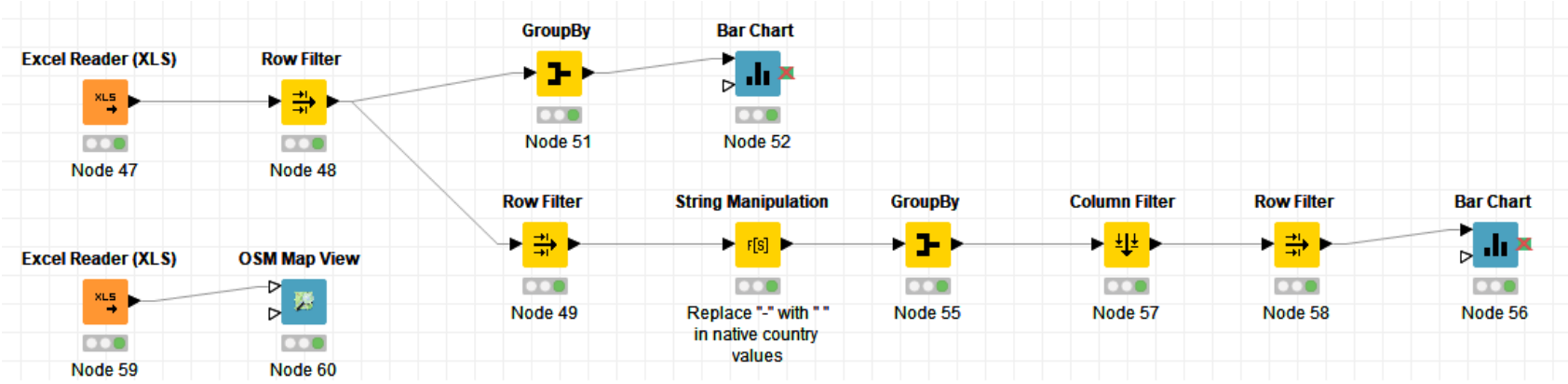
Results from the Expanded Analysis

- **Spatial Distribution of RM Firms**



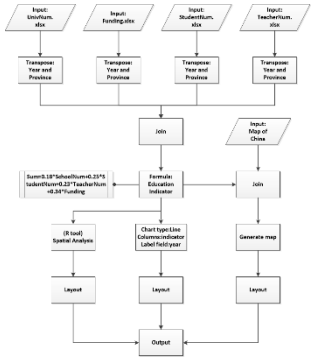
- **Number of RM Firms by Province**





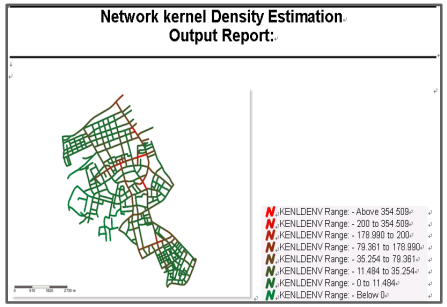
Replicable, Reproducible and Expandable

❑CDL Platform for Workflow Data Analysis

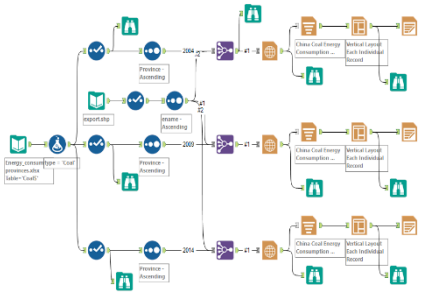


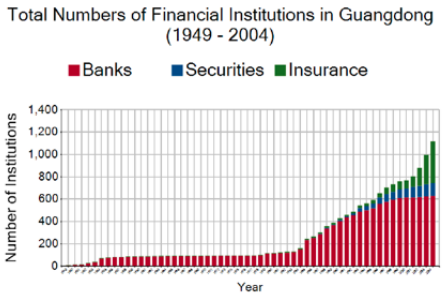
Environment

Education

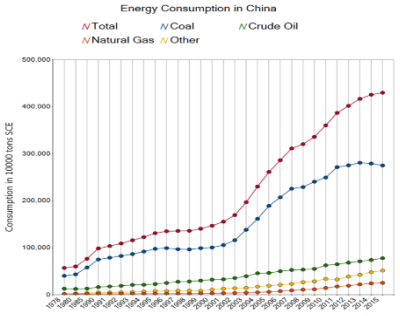


Transportation

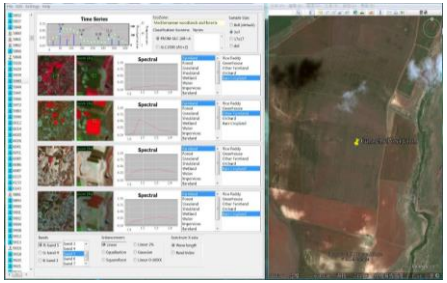




Economics

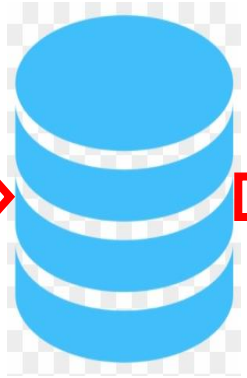


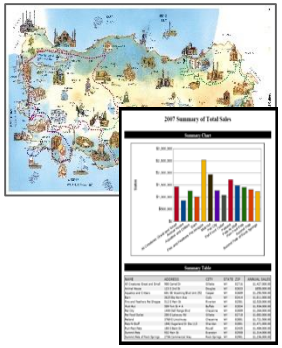
Energy

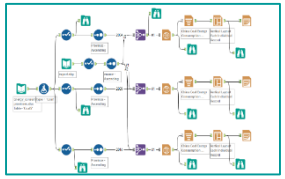


Land Use

China Data Lab for Research and Teaching







TRAINING

Workshop

Learn

Teaching

Knowledge

Skills

Web Sites

China Data Lab

http://chinadatalab.net

China Data Online

http://china-data-online.com

China Data Lab on the Cloud

http://chinadatalab.cn

office@chinadatacenter.net