



**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**

# Topics

- **Principles** for China database design
- **Data Sources** for China studies
- **Methodologies** for China data processing
- **Functions** for spatio-temporal analysis
- **China Data Lab**

# Principals for Database Design

## Principals:

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

## Primary Factors:

1. Space (boundary)
2. Time
3. Scale (Sub-systems)





# 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	<b>516.9</b>	100	<b>709.9</b>	100	<b>1085.7</b>	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	<b>682.4</b>	100	<b>886.4</b>	100	<b>1303.5</b>	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	<b>3921.6</b>	100	<b>2873.7</b>	100	<b>3279.1</b>	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)

<b>7. 信息传输、计算机服务和软件业</b>	<b>Information Transmission, Computer Services and Software</b>
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
<b>8. 批发和零售业</b>	<b>Wholesale and Retail Trade</b>
批发业	Wholesale
零售业	Retail Trade
<b>9. 住宿和餐饮业</b>	<b>Hotel and Restaurants</b>
住宿业	Hotel
餐饮业	Catering Services
<b>10. 金融业</b>	<b>Financial Intermediation</b>
银行业	Banking
证券业	Securities
保险业	Insurance
其他金融活动	Other Financial Activities
<b>11. 房地产业</b>	<b>Real Estate</b>
租赁和商务服务业	Leasing and Business Services
租赁业	Leasing Services
商务服务业	Business Services
<b>12. 科学研究、技术服务和地质勘查业</b>	<b>Scientific Research, Technical Service and Geologic Prospecting</b>
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
<b>13. 水利、环境和公共设施管理业</b>	<b>Management of Water Conservancy, Environ. and Public Facilities</b>
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
<b>14. 居民服务和其他服务业</b>	<b>Services to Households and Other Services</b>
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

The screenshot displays the homepage of the China Data Online website. The header includes the site's name in English and Chinese, along with a navigation menu. The main content area is organized into several sections, each with a list of links to specific data resources. A red box highlights the 'CHINA SPATIAL DATA' section, which includes links to 'China Geo-Explorer II', 'China Geo-Explorer I', and 'China Map Library'. Another red box highlights the 'CHINA STATISTICS' section, which includes links to 'Monthly Statistics', 'National Statistics', 'Provincial Statistics', 'City Statistics', 'County Statistics', 'Monthly Industrial Data', 'Yearly Industrial Data', 'Statistics on Map', 'Statistical Datasheets', and 'Statistical Charts'. A third red box highlights the 'CENSUS DATA' section, which includes links to 'Census Maps', 'All Census Data', 'Economic Census 2004', 'Industrial Census 1995', 'Census 1982', 'Census 1982 (10%)', 'Census 1990', 'Census 1995 (1%)', 'Province 2000', 'County 2000', 'Census 2005 (1%)', and 'Census Data Search'. Below these sections are 'FREE CHINA MAPS' and 'SAMPLE DATA' sections. On the right side of the page, there is a sidebar with a pie chart titled 'Investment in Fixed Assets for the First Ten Months of 2019' showing the distribution of investment across three industries: Primary (2.23%), Secondary (29.78%), and Tertiary (68.02%). Below the chart is a section titled 'Latest China Statistical News' with several news items and their dates.

Welcome China Data Center, your IP is: 98.224.227.102 Today is: 2019/11/28 EST USA

**ALL CHINA DATA CENTER** China Data Online 中国数据在线

Home | Data Products | Database Demo | Dictionary | Support | Contact | Q&A | Citations | My Account | Logout

### CHINA SPATIAL DATA

- ▶ China Geo-Explorer II
- ▶ China Geo-Explorer I
- ▶ China Map Library

### CHINA STATISTICS

- ▶ Monthly Statistics
- ▶ National Statistics
- ▶ Provincial Statistics
- ▶ City Statistics
- ▶ County Statistics
- ▶ Monthly Industrial Data
- ▶ Yearly Industrial Data
- ▶ Statistics on Map
- ▶ Statistical Datasheets
- ▶ Statistical Charts

### CENSUS DATA

- ▶ Census Maps
- ▶ All Census Data
- ▶ Economic Census 2004
- ▶ Industrial Census 1995
- ▶ Census 1982
- ▶ Census 1982 (10%)
- ▶ Census 1990
- ▶ Census 1995 (1%)
- ▶ Province 2000
- ▶ County 2000
- ▶ Census 2005 (1%)
- ▶ Census Data Search

### FREE CHINA MAPS

2000 Population Census	Pop & Env (1990-1999)	Pop & Env (2000)	Atlas of Industrial Census
------------------------	-----------------------	------------------	----------------------------

### SAMPLE DATA

Major Indicators	Industrial Surveys	Monthly Report	Census Data
------------------	--------------------	----------------	-------------

### Investment in Fixed Assets for the First Ten Months of 2019

Investment in Fixed Asset From Jan to Oct in 2019

Industry	Percentage
Primary industry	2.23%
Secondary industry	29.78%
Tertiary industry	68.02%

### Latest China Statistical News

- Industrial Production Operation in October 2019 (11/15/2019)
- Total Retail Sales of Consumer Goods in October 2019 (11/15/2019)
- Investment in Fixed Assets for the First Ten Months of 2019 (11/15/2019)
- Producer Prices for the Industrial Sector for October 2019 (11/11/2019)
- Consumer Prices for October 2019 (11/11/2019)
- The Gross Imports and Exports Amounts 25.63Trillion Yuan in the First Ten Months... (11/8/2019)
- Industrial Production Operation in September 2019 (10/21/2019)

# 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**

Base Year: ☒ 2000 ☐ 2010 ☐ 2000-2010

Group: ☒ Single group ☐ Multi-groups

Level: ☒ Province ☐ City ☐ County ☐ Town ☐ Any area

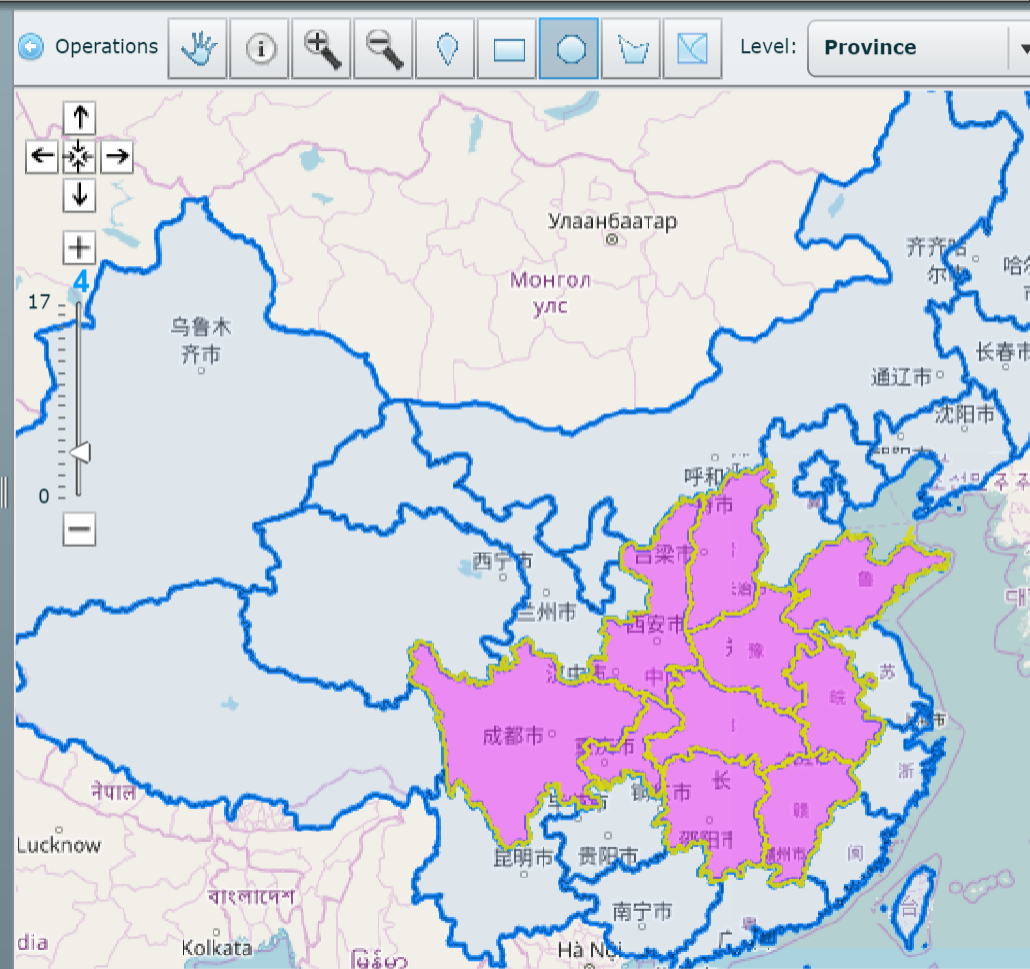
**China (Province)**

Beijing  
Fujian  
Gansu

**Selection**

Anhui  
Chongqing  
Henan  
Hubei  
Hunan  
Jiangxi  
Shaanxi

Selection count: 10



# Administrative Units

Administrative Unit X&Y Location Chart Time Series Establishment Theme Map GIS Map Export China-US Comparison Map Library Welcome, UM Asia Library! 中文 English Log out

Base Year: ☒ 2000 ☐ 2010 ☐ 2000-2010  
 Group: ☒ Single group ☐ Multi-groups  
 Level: ☒ Province ☐ City ☐ County ☐ Town ☐ Any area

Region Customized Report Standard Report Upload your data

HTML PDF CSV Excel RTF ODT

6 index selected

Report type ☐ Summary Report ☒ Compare Report ☐ Original Report

Find:

☒ Census 2000  
☒ Economic Census 2008  
☒ Economic Census 2004  
☒ Basic Unit Census 2001  
☐ Industry Census 1995  
☐ Land Use  
☐ Nighttime lights

Operations Level: Province

Theme Layer List

Basic Layer List

Boundary Layer List

temp\_report (43).xls [Compatibility Mode] - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW POWERPivot

Normal Page Break Page Custom Review Layout Views

Zoom 100% Zoom to Selection

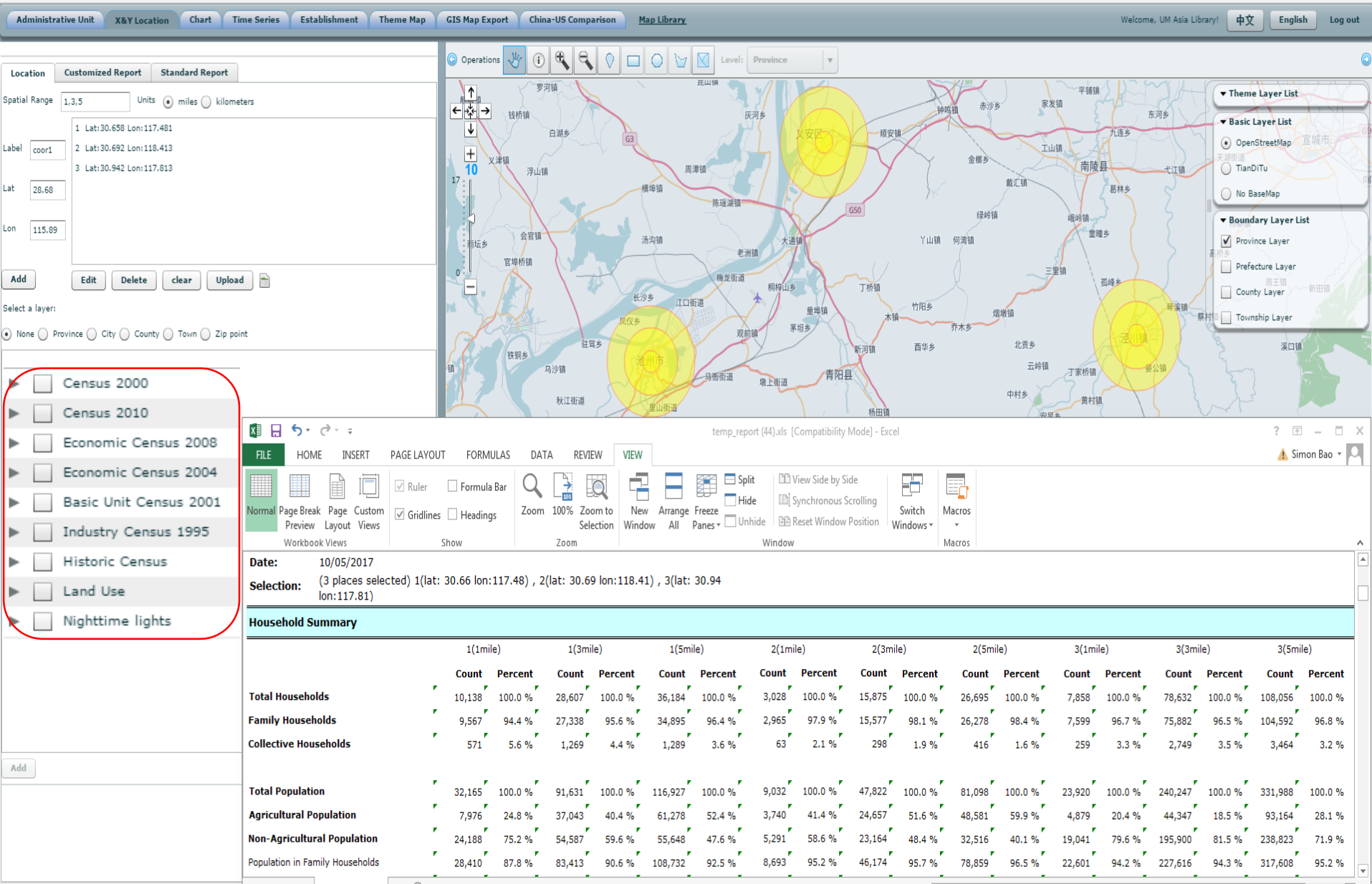
Window

	Anhui	Beijing	Chongqing	Fujian	Gansu	Guangdong	Guangxi
2000 Total Population	58,999,948	13,569,194	30,512,763	34,097,947	25,124,282	85,225,007	43,854,538
2000 Urban Population	15,766,389	10,522,464	10,095,512	14,306,812	6,018,417	47,432,392	12,350,296
2000 Rural Population	43,233,559	3,046,730	20,417,251	19,791,135	19,105,865	37,792,615	31,504,242
2008 Number of All Units	204,959	267,757	139,006	232,466	94,260	617,717	154,443
2004 Number of All Units	230,706	262,847	137,536	226,827	135,138	567,862	194,852
2001 Total Industrial Units	177,879	245,599	89,639	155,771	82,817	400,227	121,763

report name

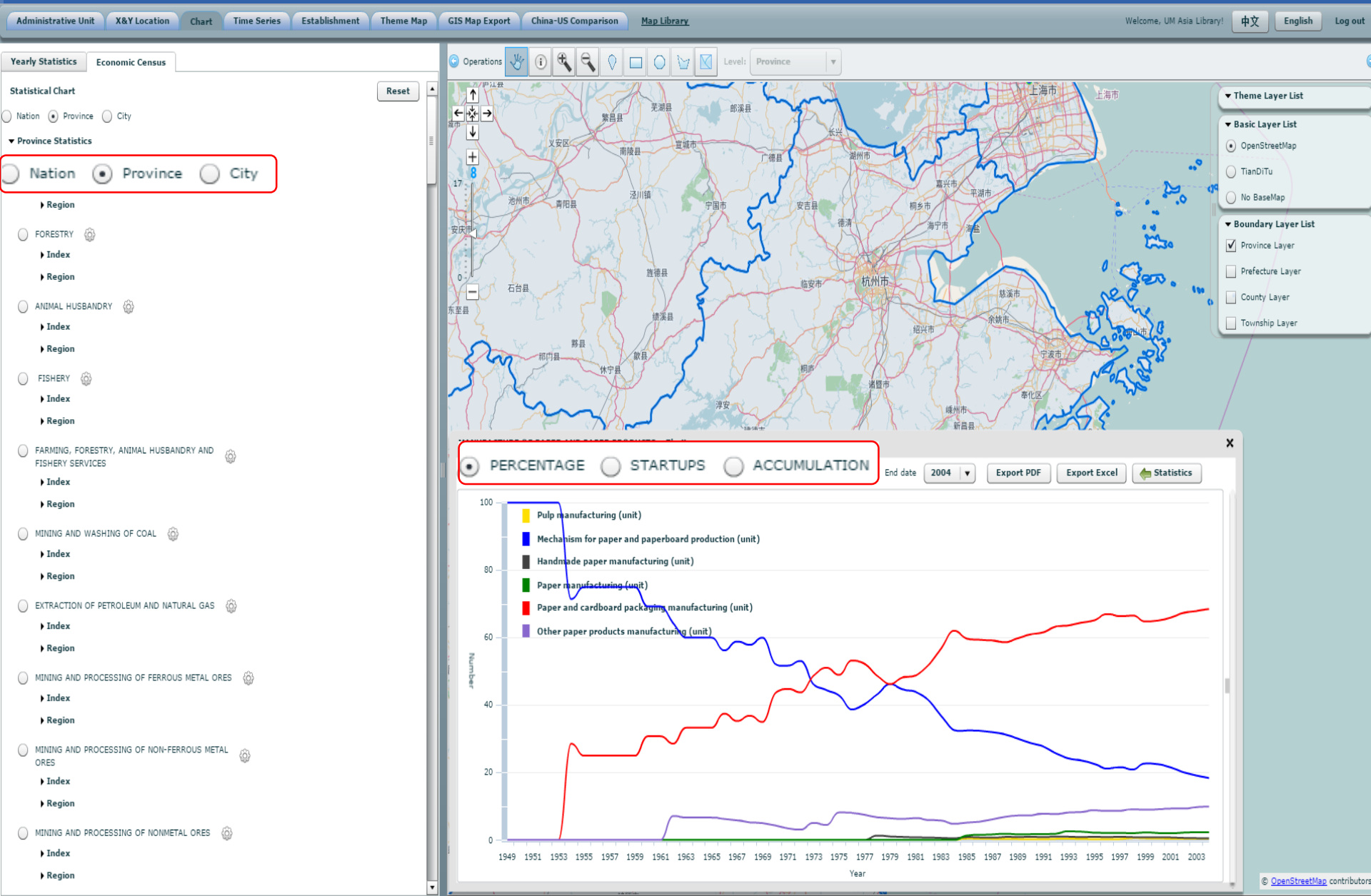
READY

# Location Analysis

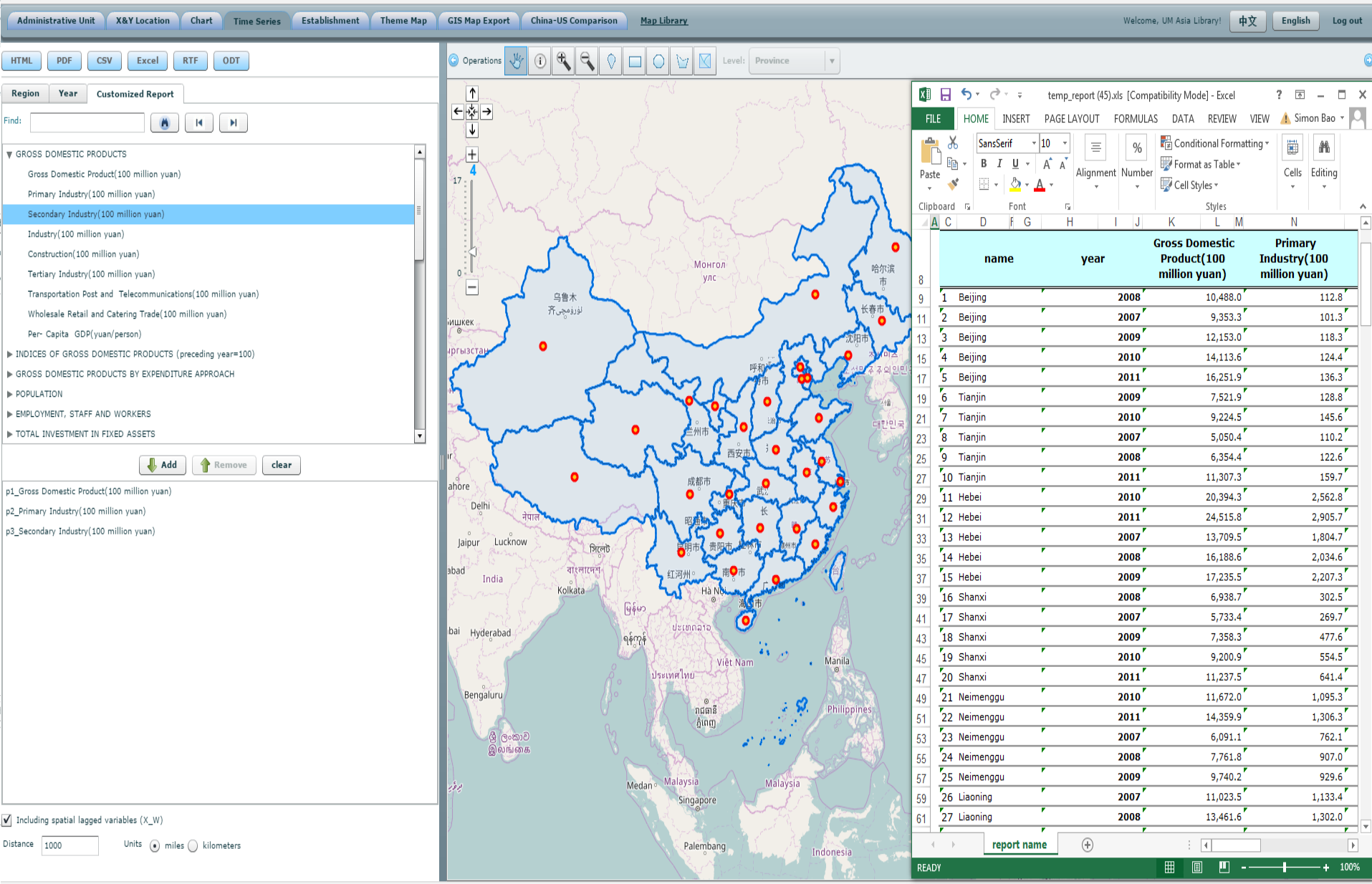




# Charts Analysis (Structural Analysis)

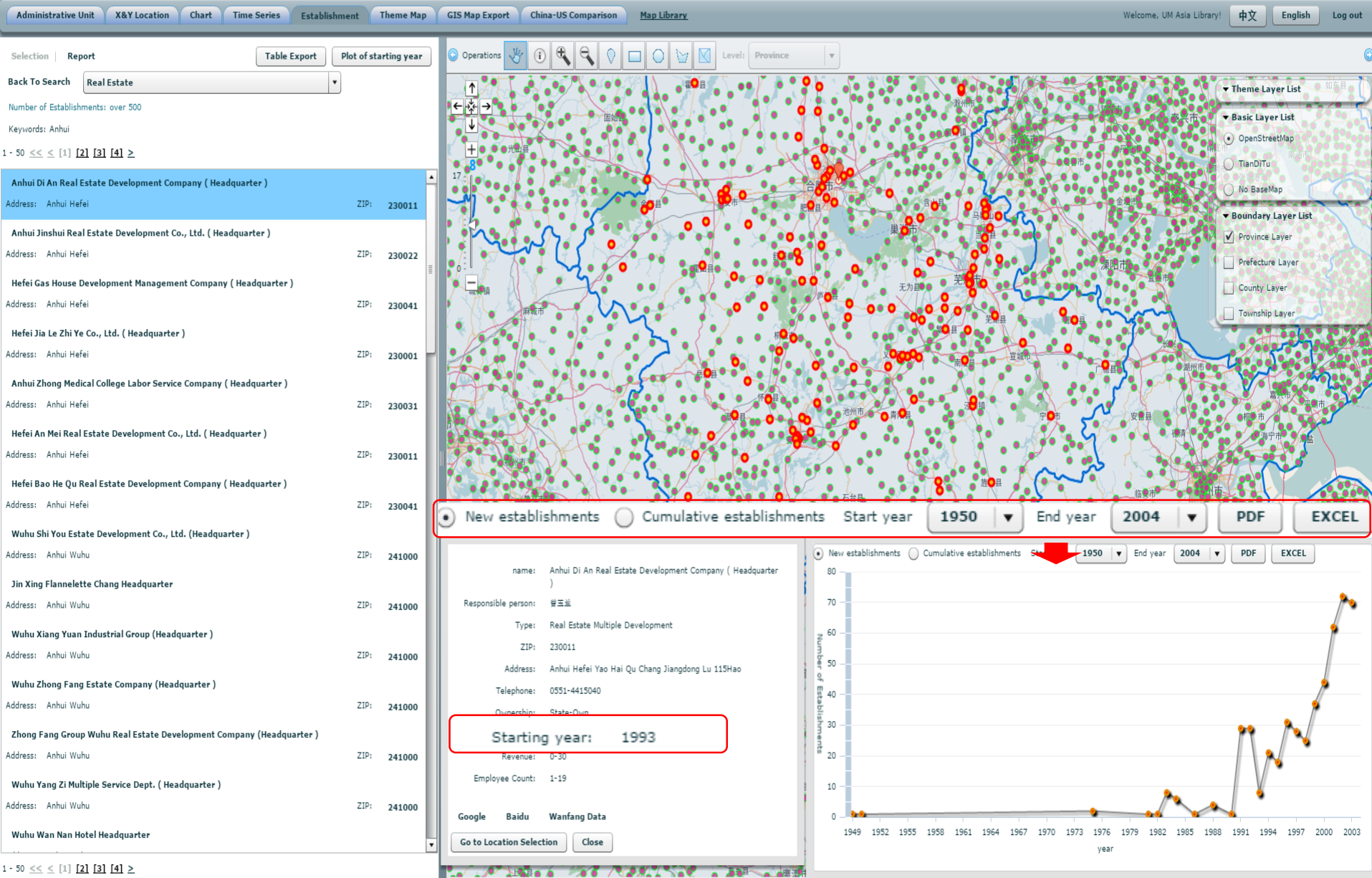


# Time Series Analysis





# Establishment

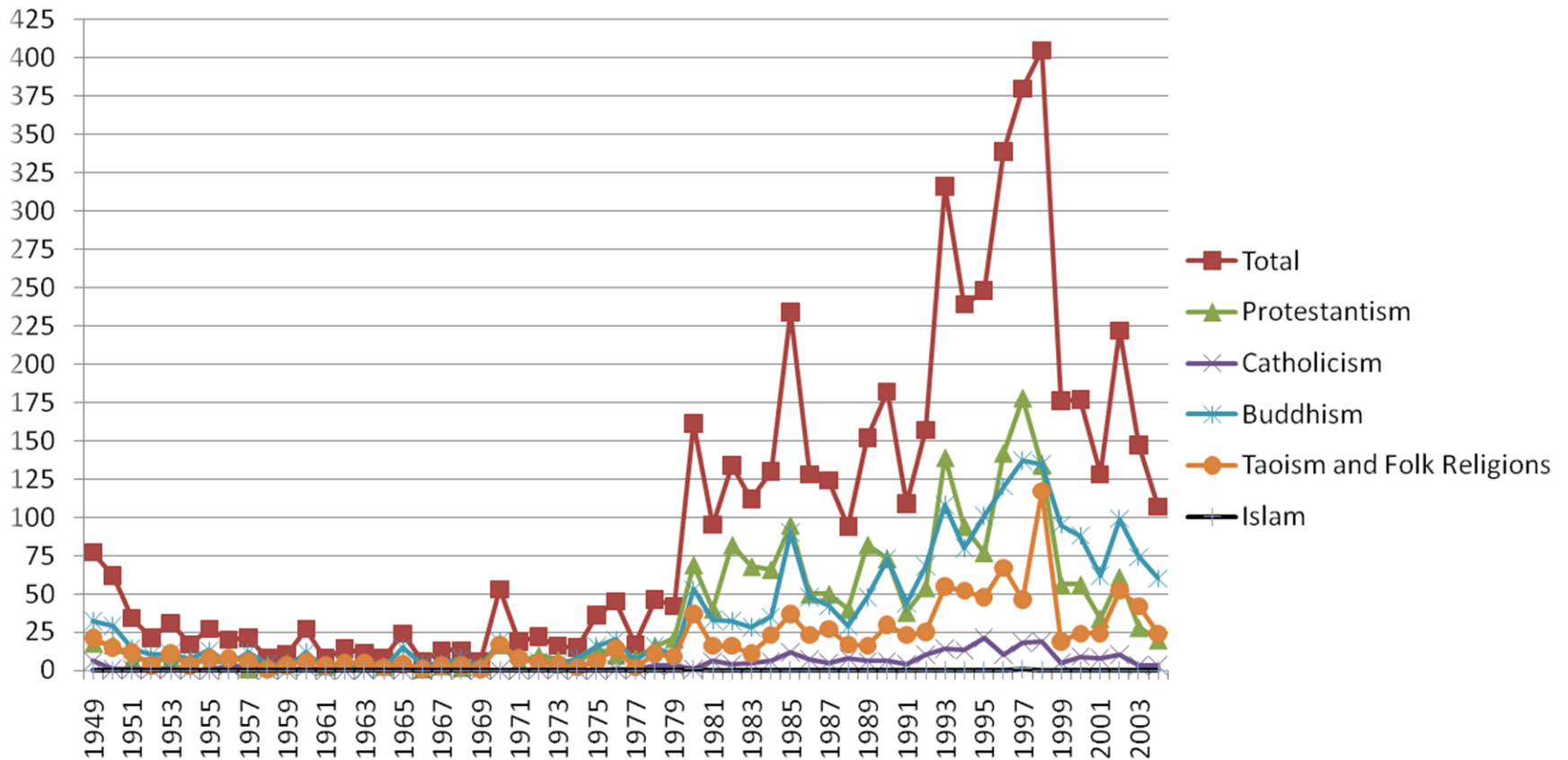


# Panel Data for Spatial Modeling

Time-series Report								
Date:	19/08/2018		Original Variables			Spatially Weighted Variables		
Selection:	(4 selected) zhejiang , tianjin , y		Gross Domestic	Primary Industry	Secondary Industry	Gross Domestic	Primary Industry	Secondary Industry
name	year							
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

# Identification of Possible Policy Impacts

Annual Increment of Religious Organizations in Zhejiang (1949-2004)



# China Data Lab



# 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**

- Research (**reproducible, replicable, generalizable**)
- Teaching (students with different interests and skills)
- Decision support (efficient, effective, and expandable)

# Solution: Cloud Based Platform

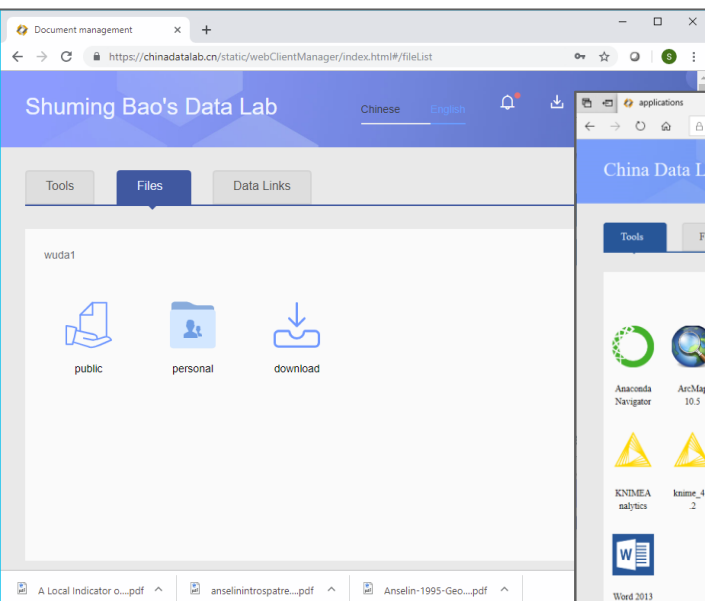
- ❑ **A data center** for China studies based on cloud
- ❑ **A research base** for collaborations on China studies
- ❑ **A development center** for data case studies
- ❑ **A training center** for China studies, including theory, methodology, technology, data and applications for research and teaching

# An Integrated Platform for Research and Teaching

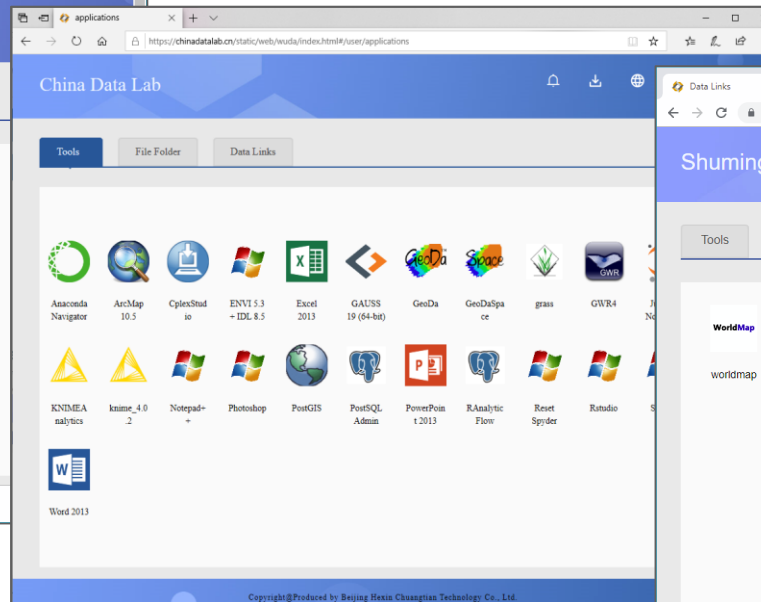
## Main Features:

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

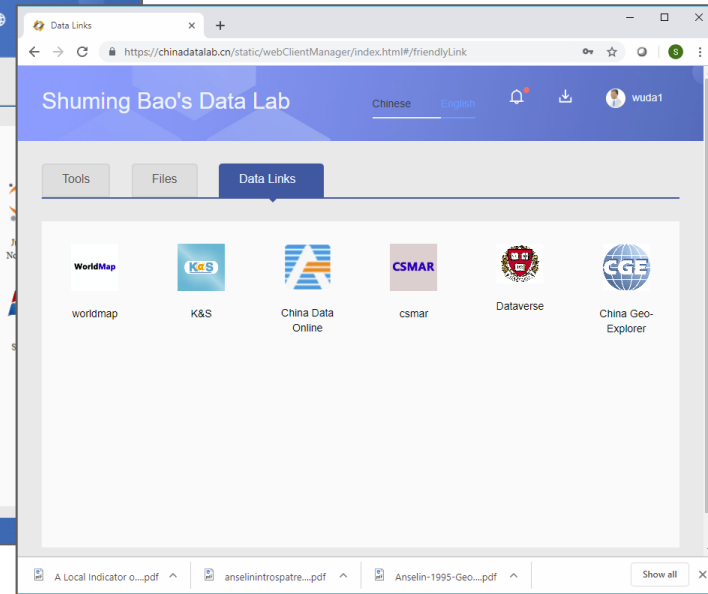
## Data Center



## Tool Center



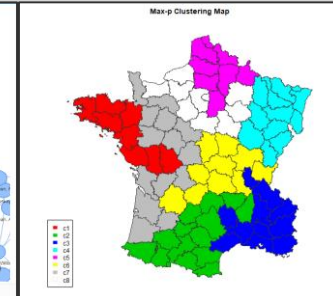
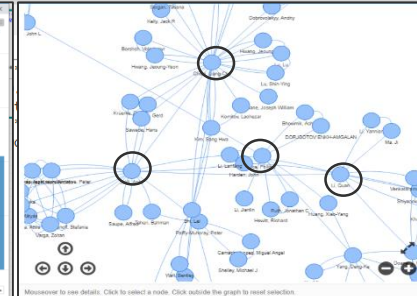
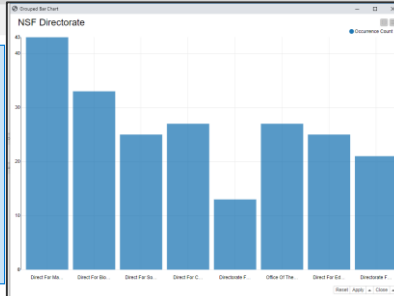
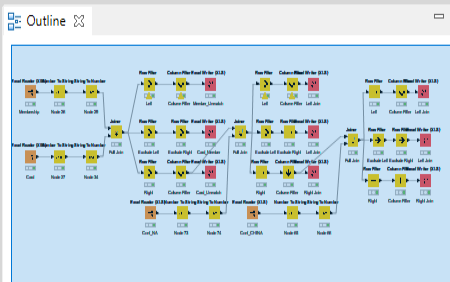
## Sharing Center





The screenshot displays the KNIME Analytics Platform interface. The main workspace shows a complex workflow titled "CES annual report test". The workflow starts with two "Excel Reader (XLS)" nodes, "Membership" and "Card", which feed into "Node 26" and "Node 27" respectively. These nodes then feed into "Node 29" and "Node 34". The workflow continues with various "Joiner", "Row Filter", "Column Filter", and "Excel Writer (XLS)" nodes, including "Full Join", "Left", "Right", "Exclude Left", "Exclude Right", "Card\_Member", "Card\_Unmatch", "Conf\_NA", and "Conf\_CHINA". The workflow ends with "Node 73", "Node 74", "Node 65", and "Node 66".

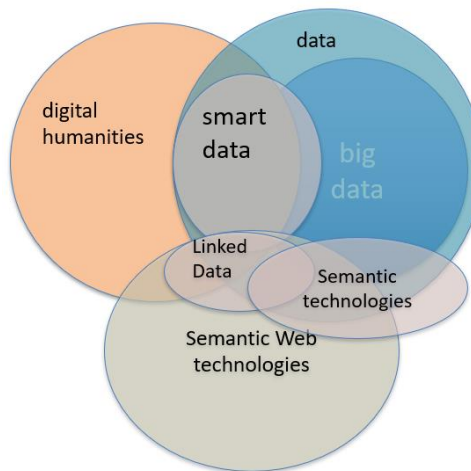
The left sidebar contains the "Node Repository" and "Workflow Coach" panels. The right sidebar shows the "Description" panel for the workflow, titled "CES annual rep...". The bottom section of the interface includes an "Outline" panel, a bar chart titled "NSF Directorate", and a "Max-p Clustering Map".





# 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
- ...

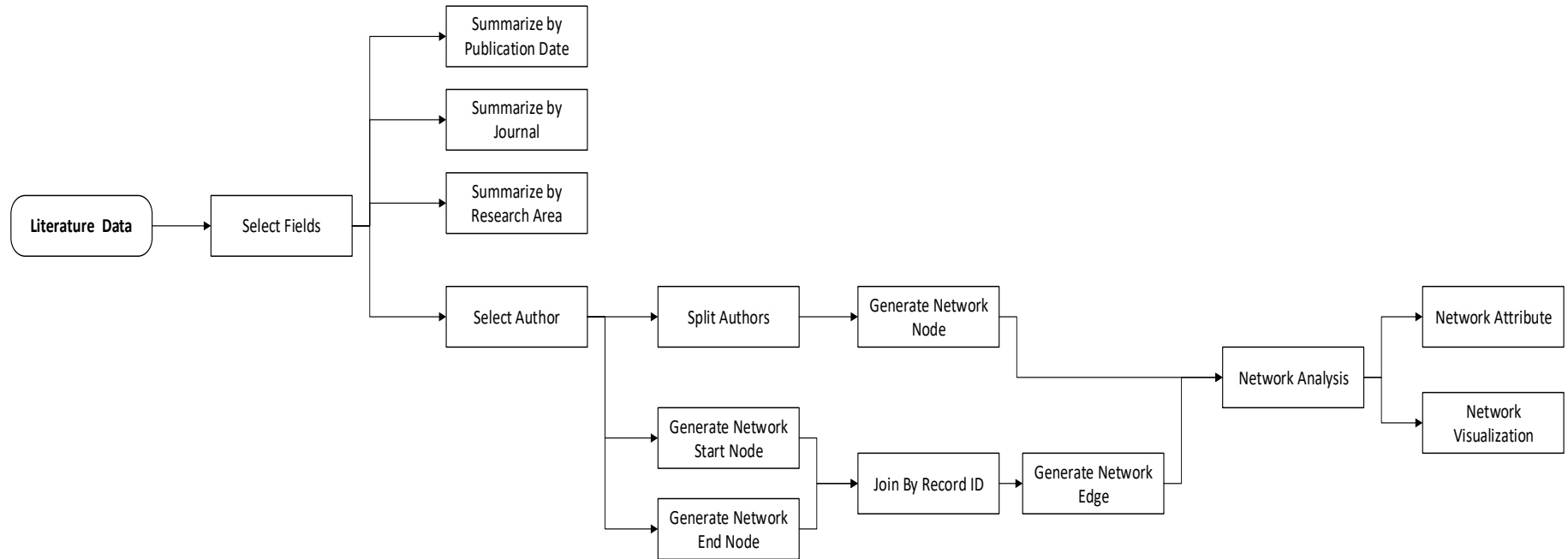
## ■ Awarded Grants

- Title
- PI
- Co-PI
- Email Address
- Institution
- NSF Organization
- Start Date
- Expiration Date
- Awarded Amount
- NSF Directorate

# Data Input

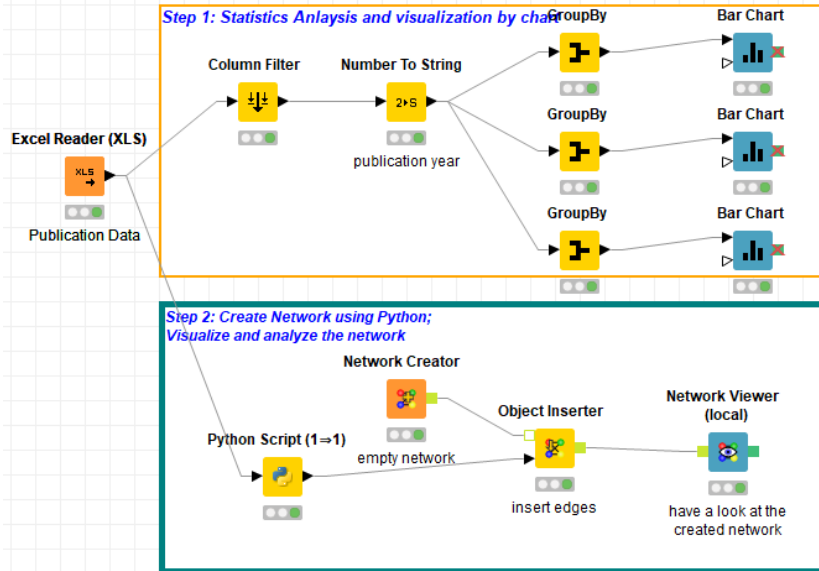
<b>File Name</b>	<b>Format</b>	<b>Data Sources</b>
publication.xls	.xls	Web of Science
patent.xls	.xls	ProQuest
grant.xls	.xls	NSF website

# The Flowchart for Data Analysis

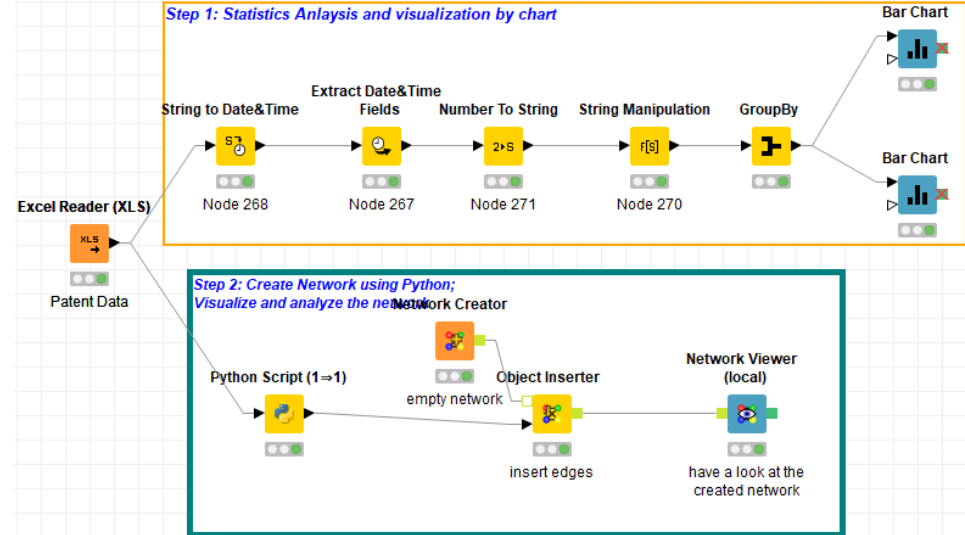


# Knime Workflow for Literature Analysis

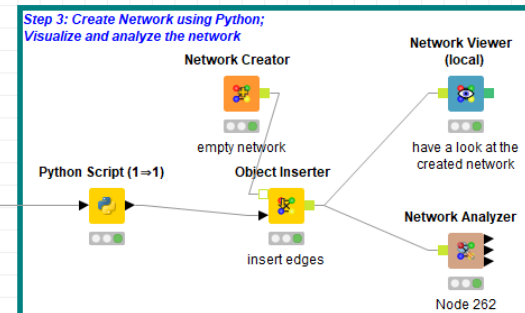
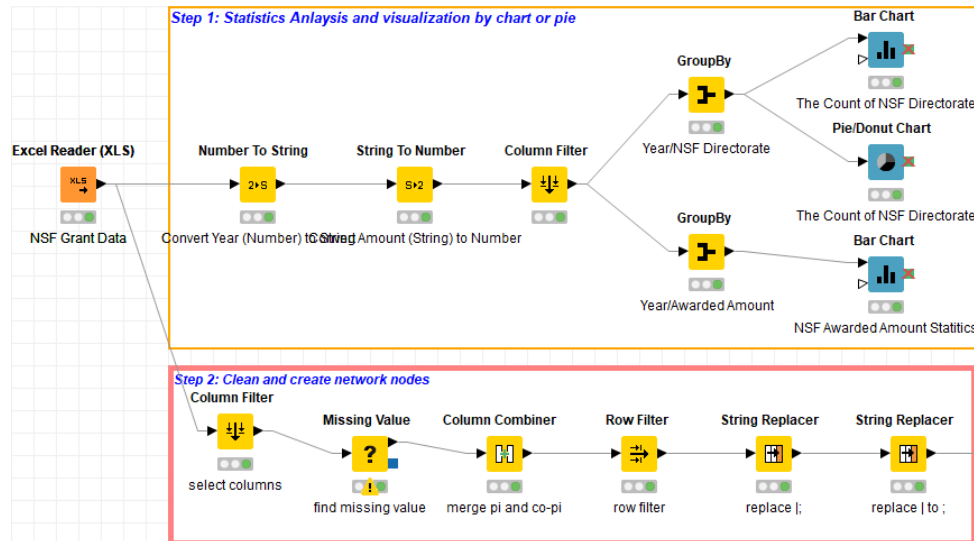
## Publication Analysis



## Patent Analysis

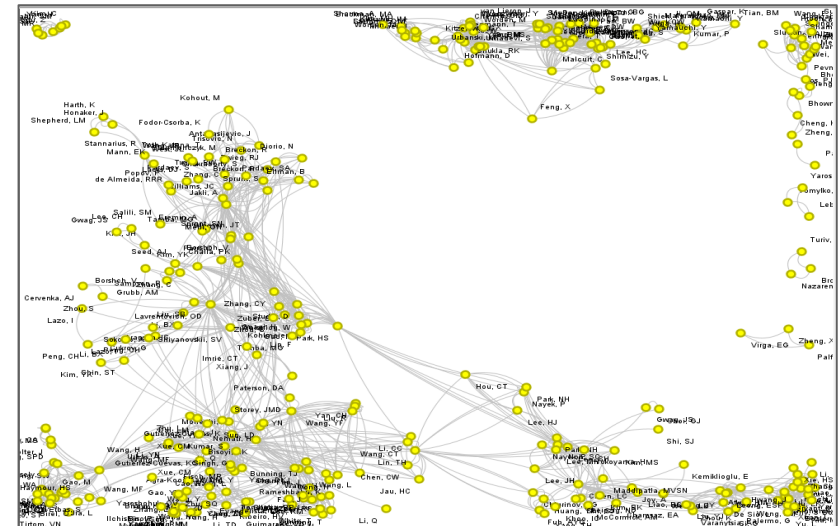
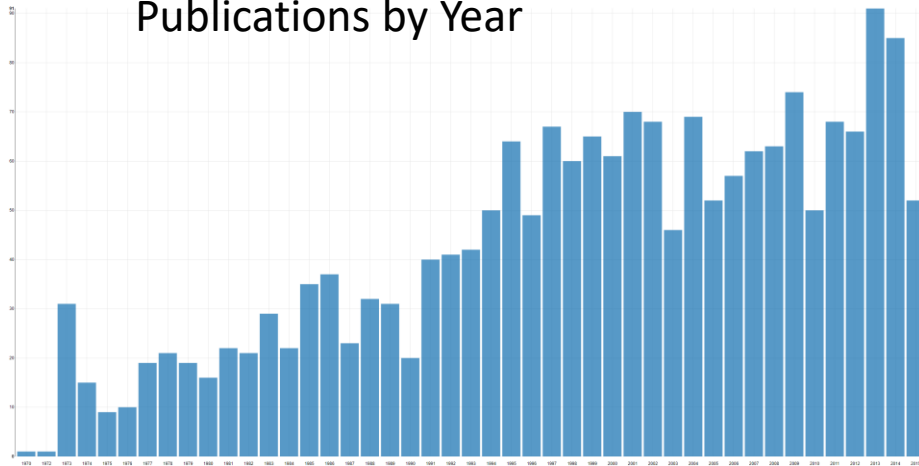


## Grant Analysis



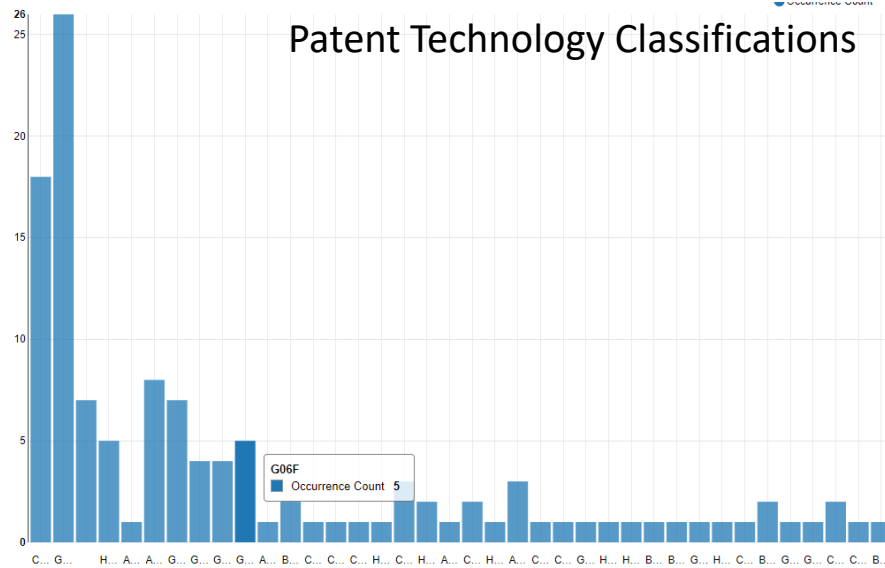
# Results from Publication Analysis

Publications by Year

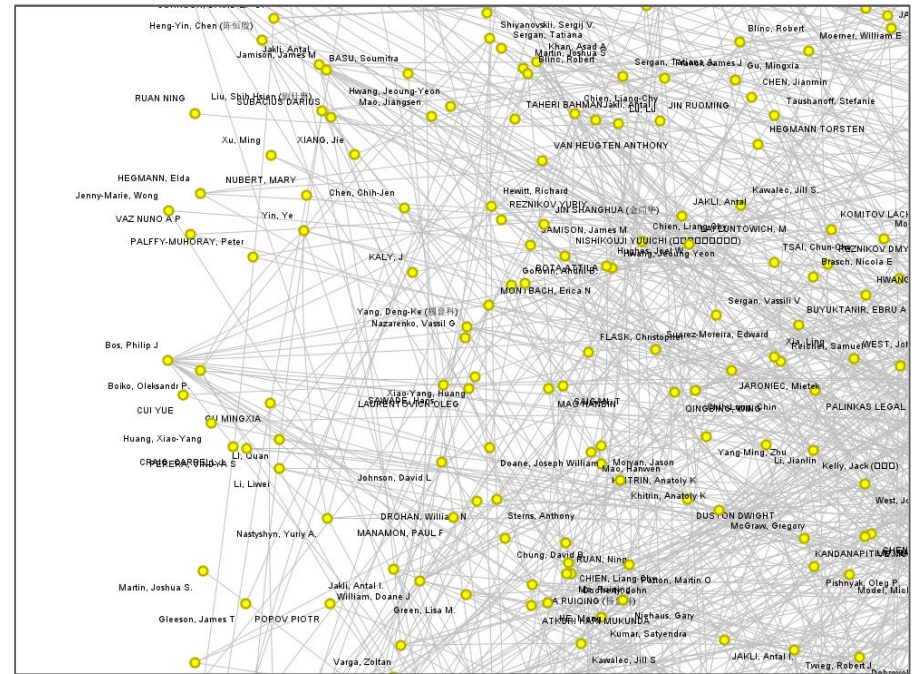


[S] Object id	[D] Node d...	[D] Node d...	[D] In degree	[D] In degr...	[D] Out de...	[D] Out de...	[D] Closen...	[D] Node w...	[D] Avg. n...	[D] Clusteri...	[D] Hub score	[D] Authori...	[D] Baryce...
Feng, X	5	0.326	5	0.326	5	0.326	0.483	5	1	1	0.069	0.069	0.008
Wang, F	5	0.326	5	0.326	5	0.326	0.621	5	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

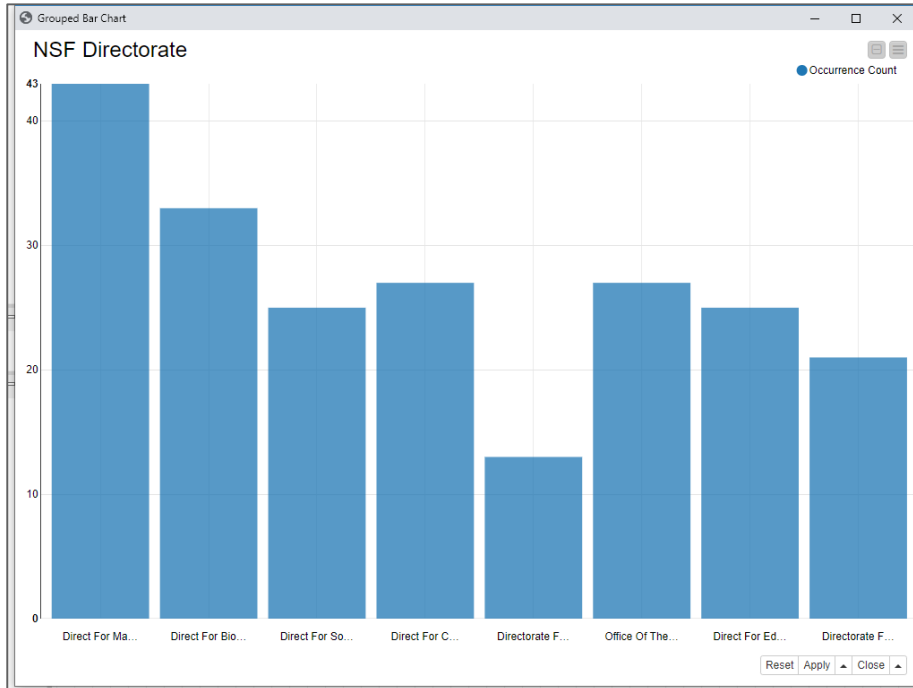


S	Object id	D Node d...	D Node d...	D In degree	D In degr...	D Out de...	D Out de...	D Node w...	D Avg. n...
	Chao-Chiun, Liang	9	0.771	9	0.771	9	0.771	9	1
	SHOKOUHIMEH...	5	0.428	5	0.428	5	0.428	5	1
	Nemati, Hossein ...	9	0.771	9	0.771	9	0.771	9	1
	Hwang, Jeoung ...	3	0.257	3	0.257	3	0.257	3	1
	QIAN 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





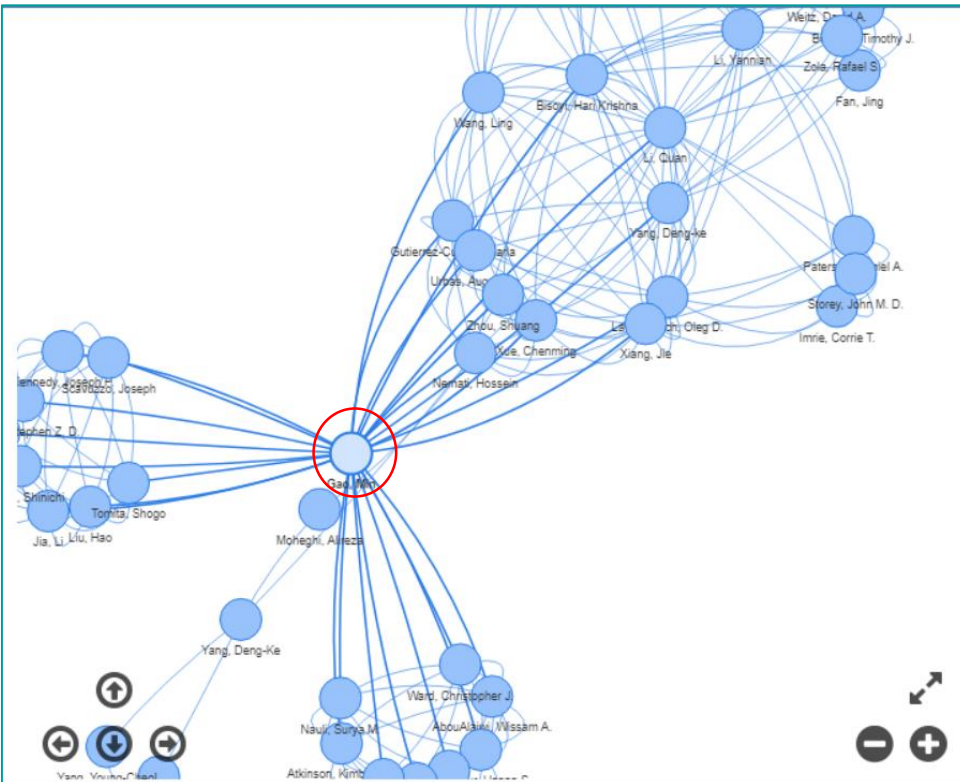
# Results from Grant Analysis



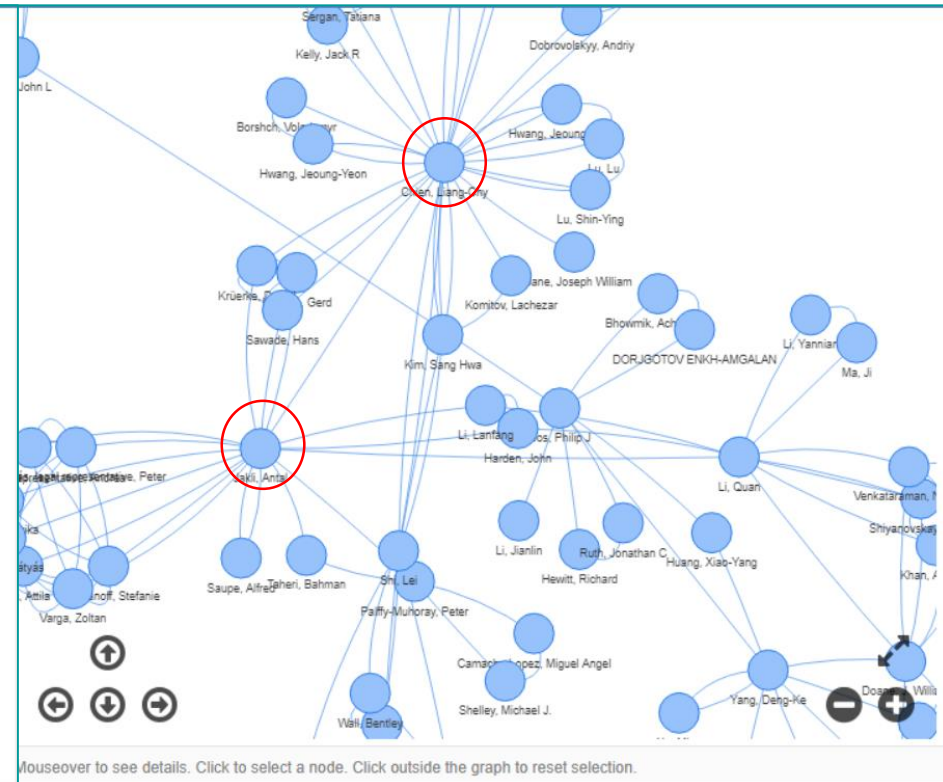
S Object id	D Node d...	D Node degree %	D In degree	D In degree %	D Out degree	D Out degree %	D Node weigh...	D Avg. n...
Genevieve Davis	1	0.787	1	0.787	1	0.787	1	1
Gerassimos Pe...	1	0.787	1	0.787	1	0.787	1	1
Xiaoyu Zheng	2	1.575	2	1.575	2	1.575	2	1
Noah FriedkinE...	1	0.787	1	0.787	1	0.787	1	1
William 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.



Journal of Corporate Finance

Volume 58, October 2019, Pages 92-111



### Going public in China: Reverse mergers versus IPOs ☆

Charles M.C. Lee <sup>a</sup>, Yuanyu Qu <sup>b</sup>, Tao Shen <sup>c</sup>

[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.

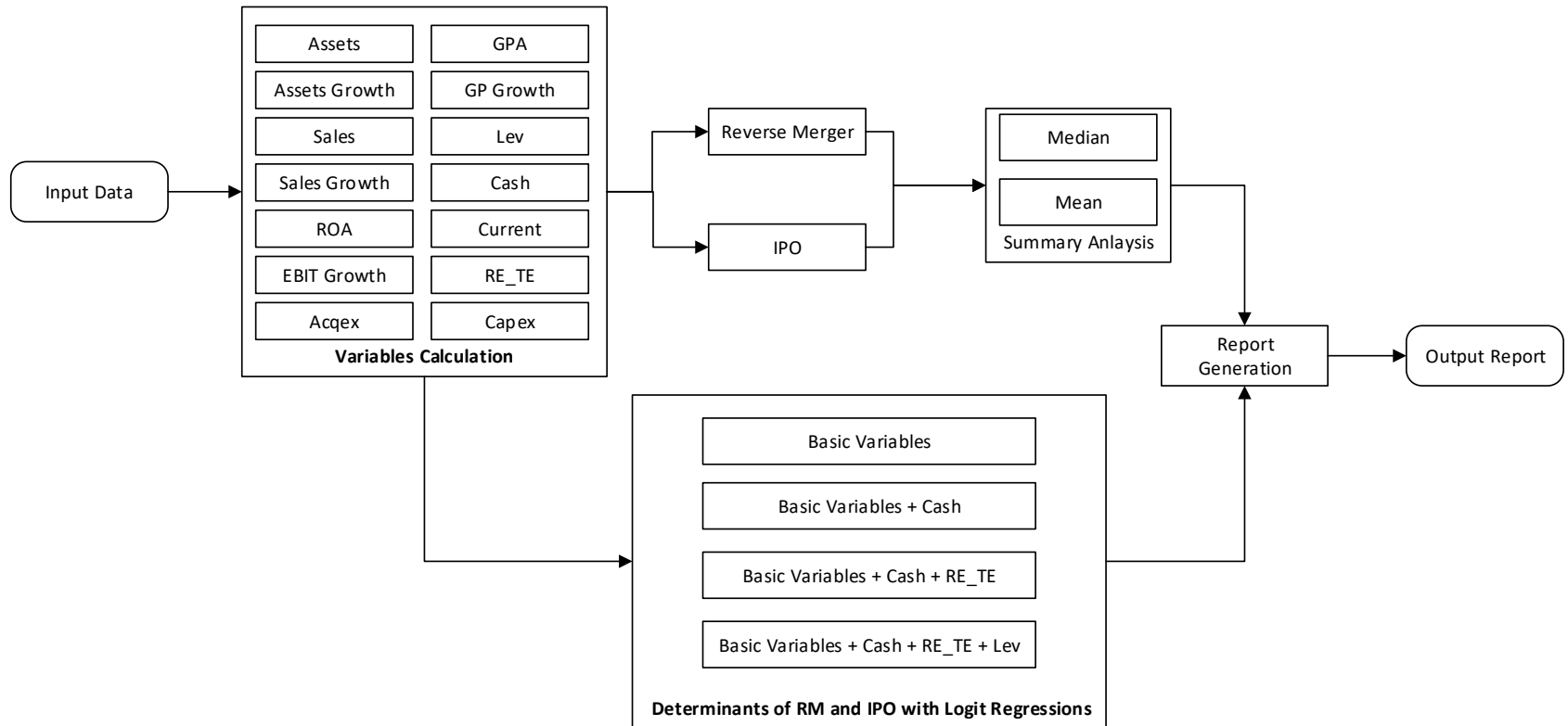
#### Abstract

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.

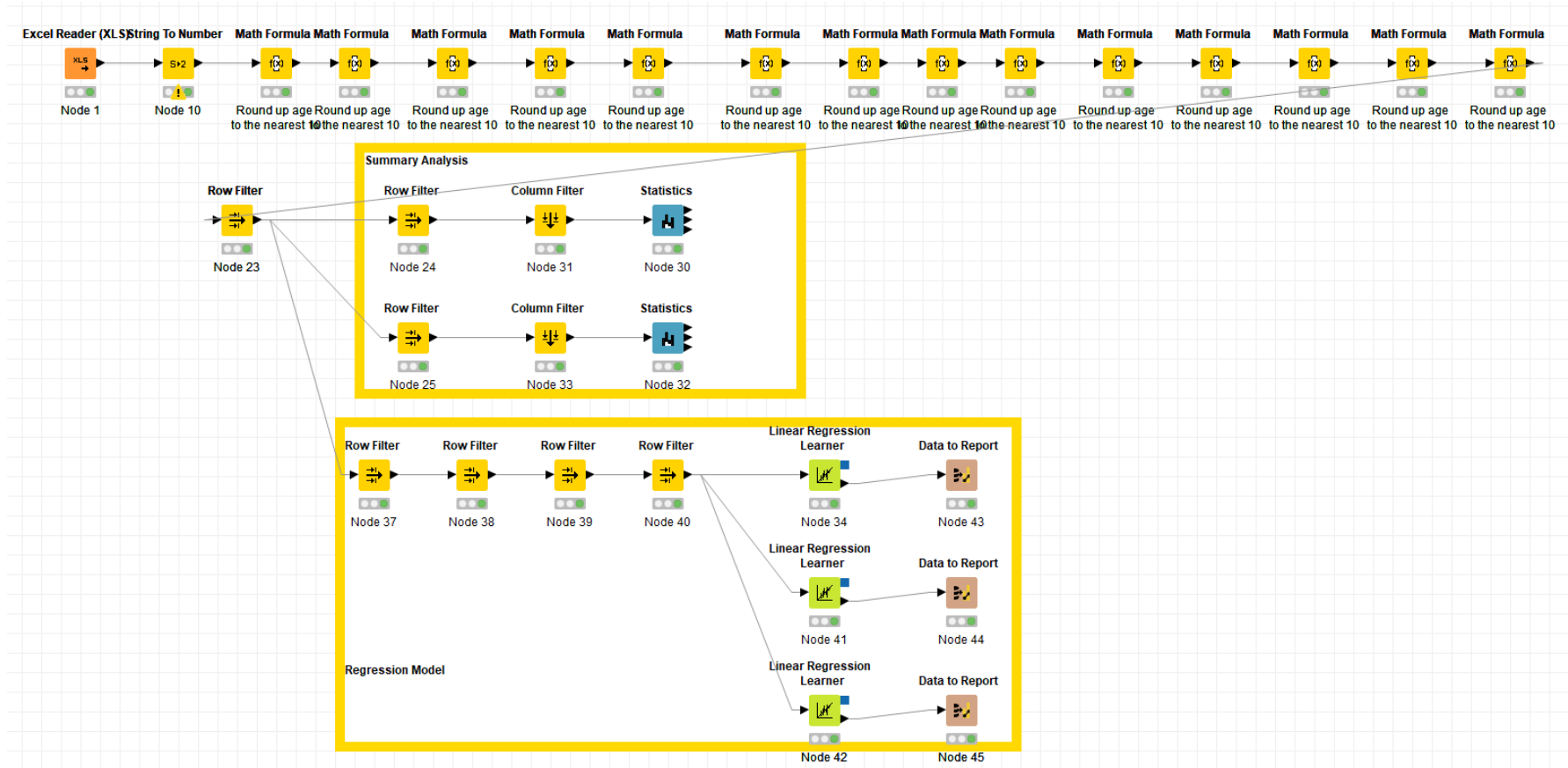
# Data Sources

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 <a href="http://www.cninfo.com.cn">www.cninfo.com.cn</a> , 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	S Column	D Min	D Max	D Mean	D Std. deviation	D Variance	D Overall ...	I	I	I	I	I	I	I	Histogram
GPA	GPA	-0.007	0.638	0.162	0.141	0.02	17.024	5	0	0	0	0	0	0	
ROA	ROA	-0.157	0.42	0.085	0.099	0.01	9.094	3	0	0	0	0	0	0	
GP Growth	GP Growth	-627....	56,679....	1,260.835	7,203.23	51,886,524....	132,387.667	5	0	0	0	0	0	0	
assets	assets	309,7....	359,80....	18,987....	43,036,002,985....	1,852,097,5....	2,088,583,6....	0	0	0	0	0	0	0	
assets gr...	assets gr...	-0.299	376.438	4.996	36.275	1,315.889	549.613	0	0	0	0	0	0	0	
sales	sales	12,65....	57,482....	5,055,5....	8,239,117,930.447	67,883,064....	530,827,55....	5	0	0	0	0	0	0	
sales growth	sales growth	-0.918	87.484	2.938	12.784	163.419	323.195	0	0	0	0	0	0	0	
EBIT	EBIT	-476....	10,705....	805,421....	1,450,585,376.419	2,104,197,9....	86,180,152....	3	0	0	0	0	0	0	
lev	lev	0.045	0.956	0.492	0.235	0.055	54.115	0	0	0	0	0	0	0	

## Results from Logit Regressions

S Variable	D Coeff.	D Std. Err.	D t-value	D 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

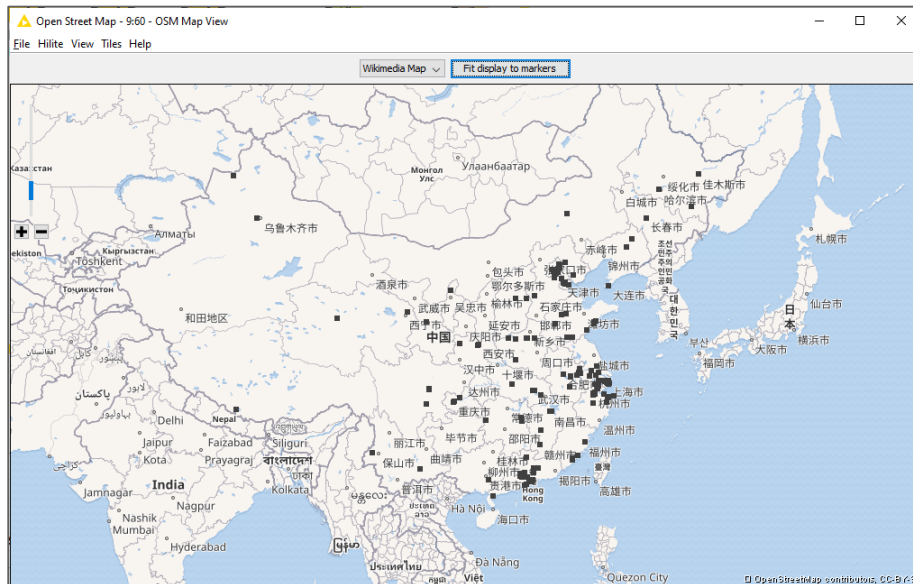
S Variable	D Coeff.	D Std. Err.	D t-value	D P> t
GPA	-0.164	0.117	-1.403	0.161
ROA	0.67	0.19	3.519	0
assets	0	0	0.868	0.385
assets growth	0	0.001	0.056	0.955
sales	-0	0	-0.176	0.86
sales growth	0.012	0.002	4.927	0
EBIT	-0	0	-0.767	0.443
lev	0.129	0.052	2.477	0.013
current	0.014	0.087	0.159	0.873
re_te	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

保留三位小数

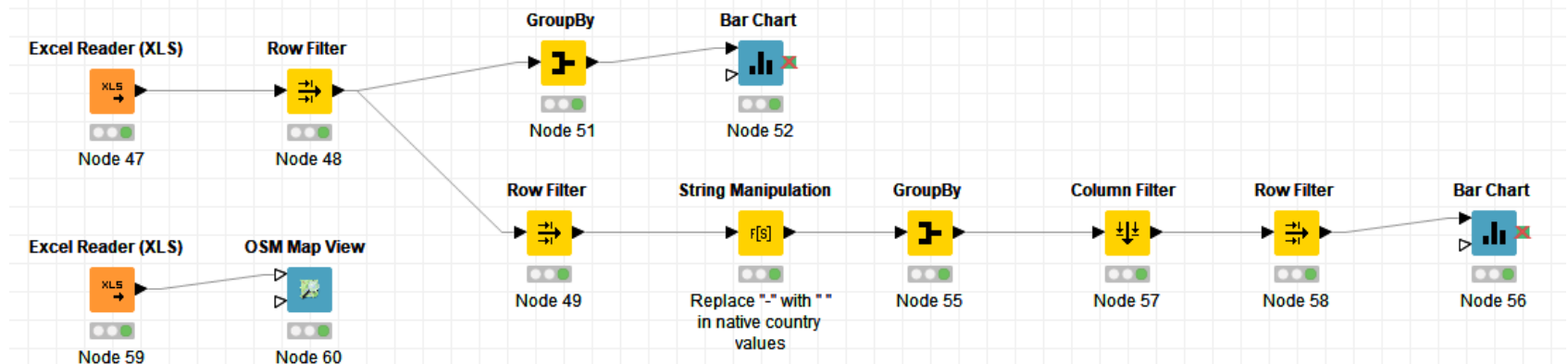
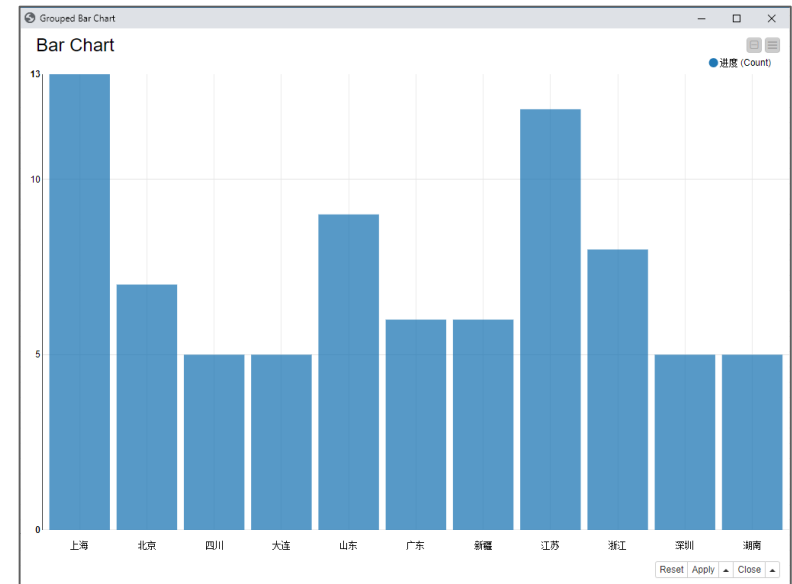


# Results from the Expanded Analysis

## ❑ Spatial Distribution of RM Firms



## ❑ Number of RM Firms by Province





## ❑ CDL Platform for Workflow Data Analysis



# China Data Lab for Research and Teaching

