China Data Institute

chinadatacenter.net
China-data-online.org

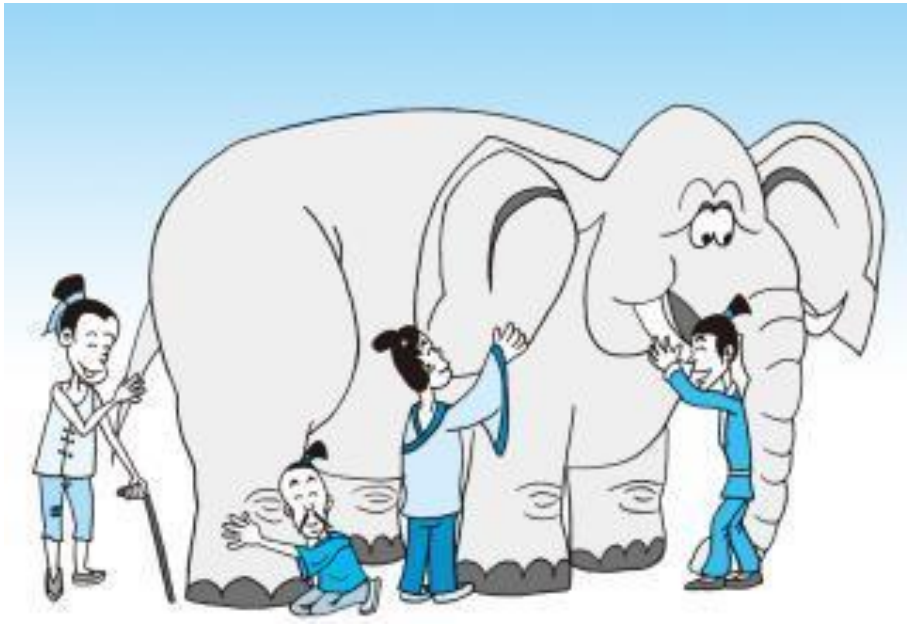
Spatial Data Analysis with China Geo-Explorer

Shuming Bao
China Data Institute

Ecological Approach: A System Theory

Wikipedia: Systems theory. https://en.wikipedia.org/wiki/Systems_theory

"A system is a cohesive conglomeration of **interrelated** and **interdependent** parts that is either natural or man-made. Every system is delineated by its **spatial** and **temporal** boundaries, surrounded and influenced by its **environment**."

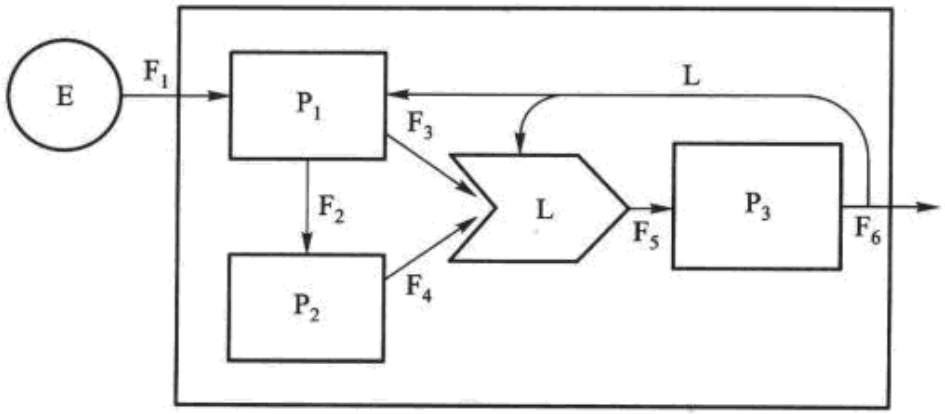


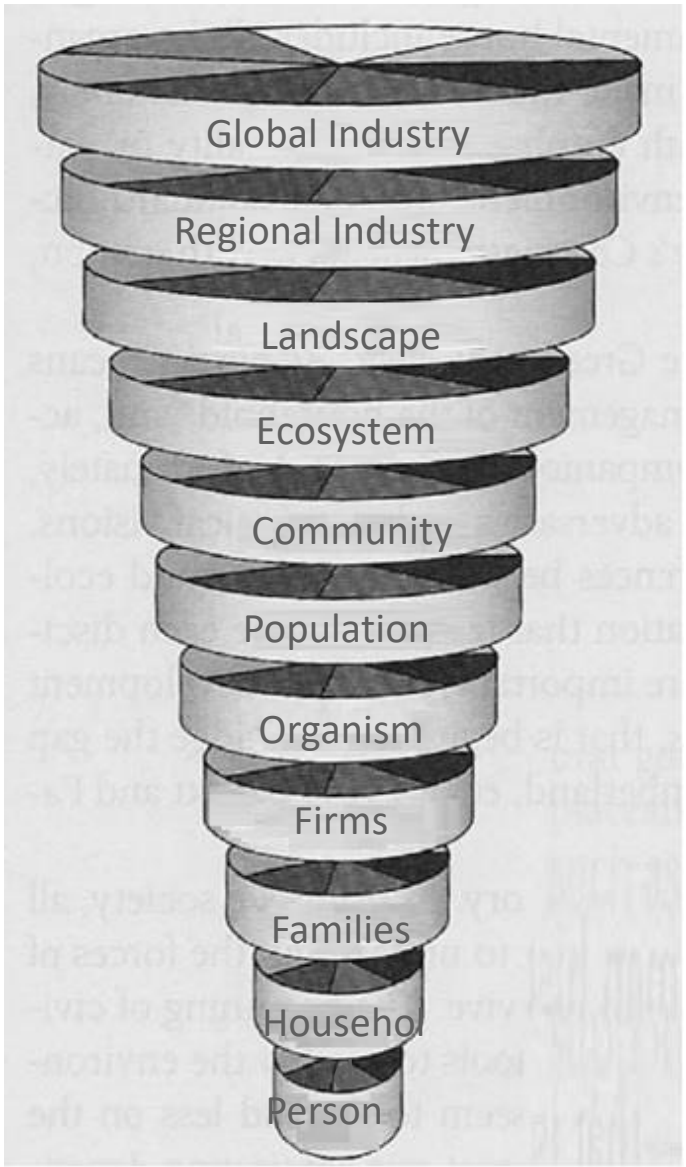
Ecological Models for Social Sciences

Fundamental Assumptions:

- Everything is related
- Degenerative system
- Increasing ecological risks:
  - Environmental risk
  - Health risk
  - Social risk

An graphic model of ecosystem

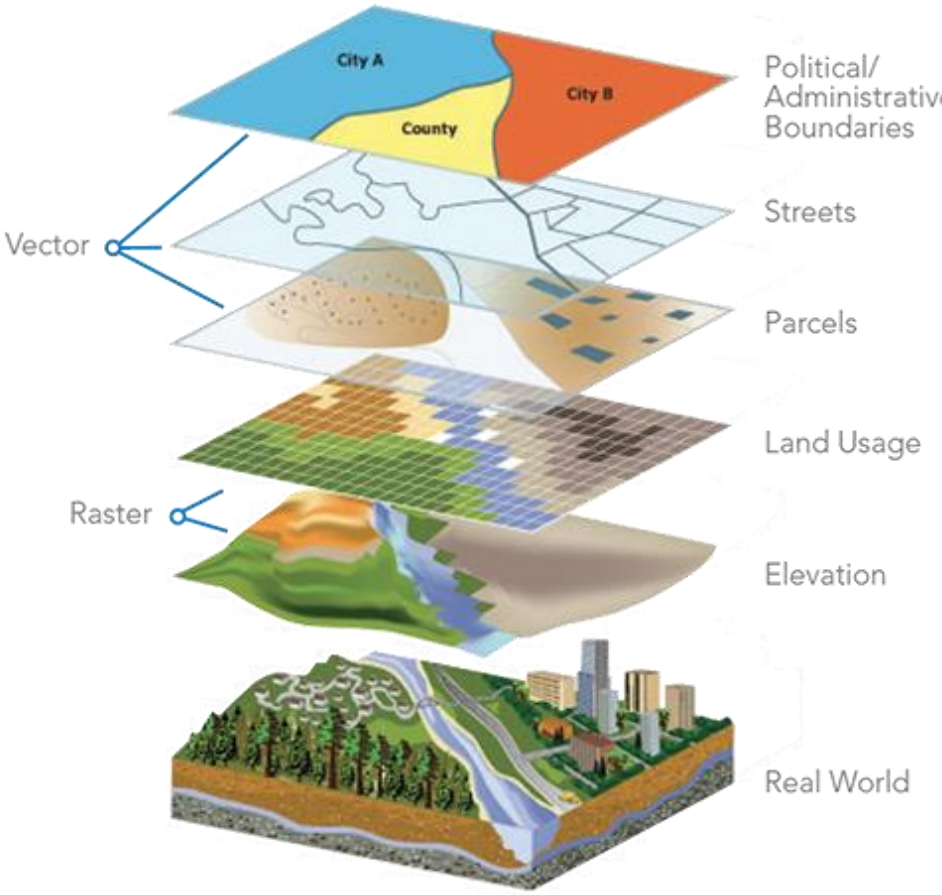




Features of Spatial Data Infrastructure

Spatial Data Platform: A Unique Platform for Multi-sources and Multi-scale Data Integration:

- ☐ Location
- ☐ Time
- ☐ Integration



Topics

- **Principles** for spatial database design
- **Methodologies** for ecological analysis
- **Functions** for spatio-temporal analysis
- **Applications** of China Geo-Explorer
- **Directions** for future work

Principals for Spatial Database Design

Principals:

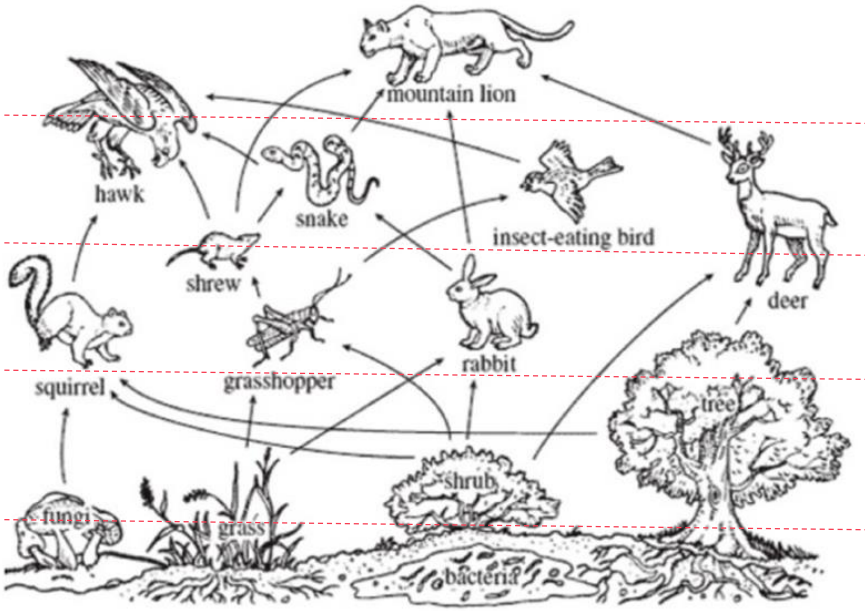
- **Comprehensive** data coverage
- **Comparable** spatio-temporal data structure
- **Compatible** multi-source data structure
- **Consistent** multi-scale data structure

Primary Factors：

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

Methodologies for Ecological Analysis

- **Classification** (spatial and non-spatial)
- **Identification**
  - Patterns
  - Linkages (Inter and intra-regions, Inter and intra-sectors)
  - Changes (spatio-temporal patterns and structures)
- **Modeling**



Tertiary consumers (predator)

Secondary consumers (meat)

Primary consumers (vegetable)

Producers (make own food and do not depend on any other organism for nutrition)

Environment (Detritivores & Decomposers)

Questions for Spatial Data Analysis

- What are the spatial **patterns** of data?
- What are the spatial **structures** of data?
- What are the spatial **trends** (patterns and structures) over the time?
- What are spatial **interactions** between different regions and aspects of the society over the time and space?
- What are the potential **impacts** of proposed actions on the complex systems?

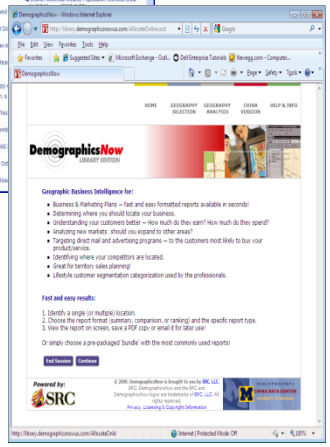
Functions for Spatio-temporal Data Analysis

- ❑ Report Analysis
- ❑ Chart Analysis
- ❑ Spatial Analysis
- ❑ Statistical Analysis
- ❑ Map Analysis

China Geo-Explorer

An Integration of Spatial Data and Analysis for China Studies

Statistics



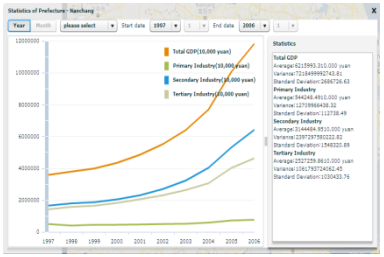
Census

GIS

Data

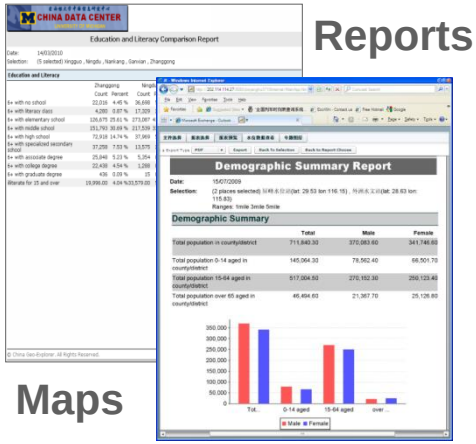
Output

Charts

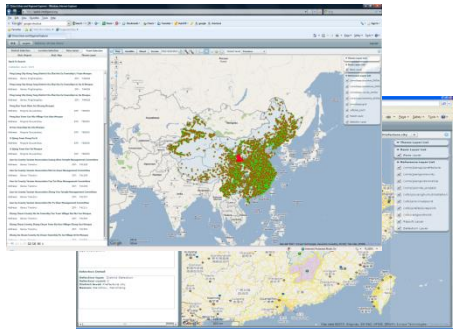


Tables

Reports



Maps



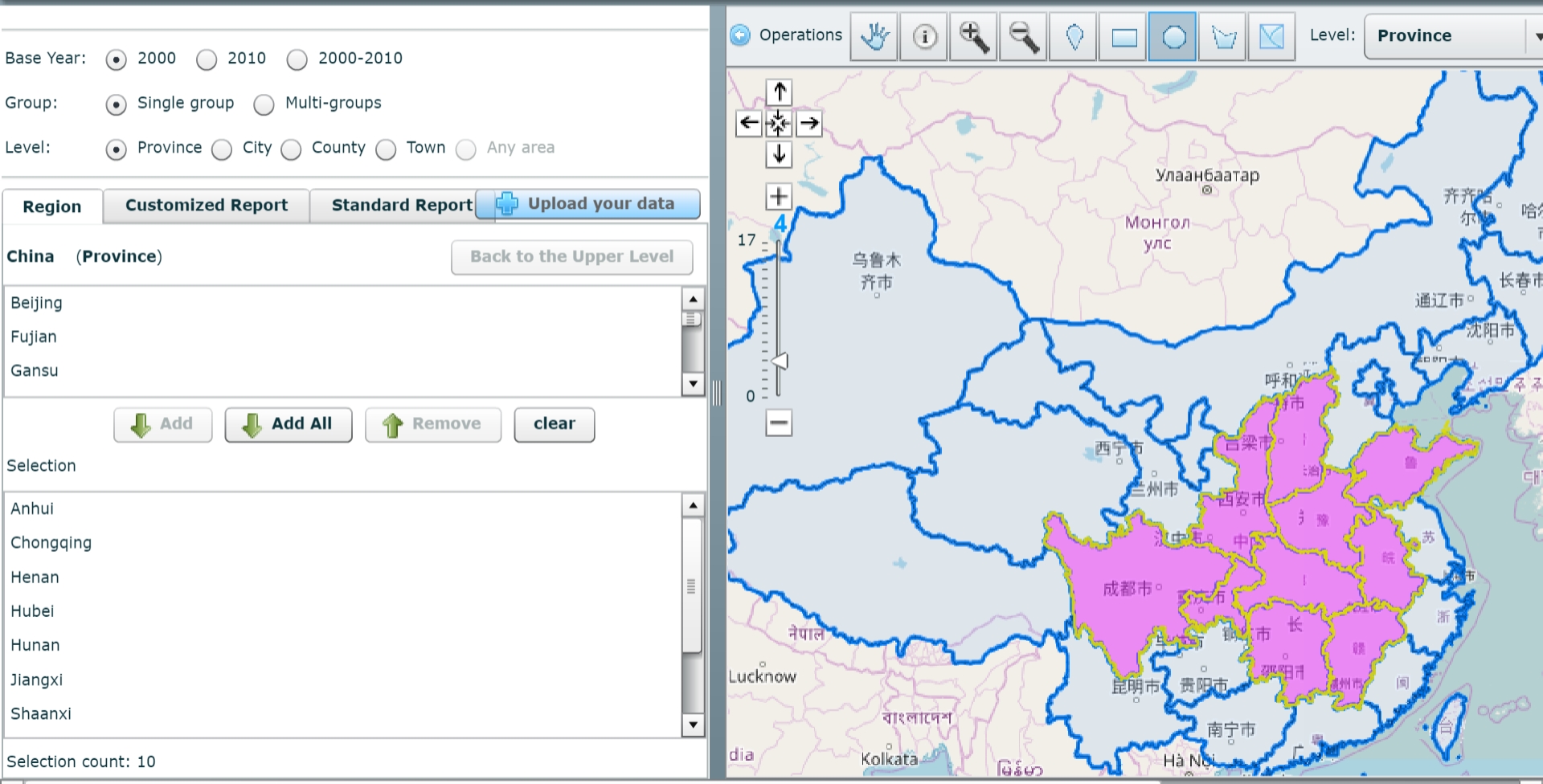
Data: China

- **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

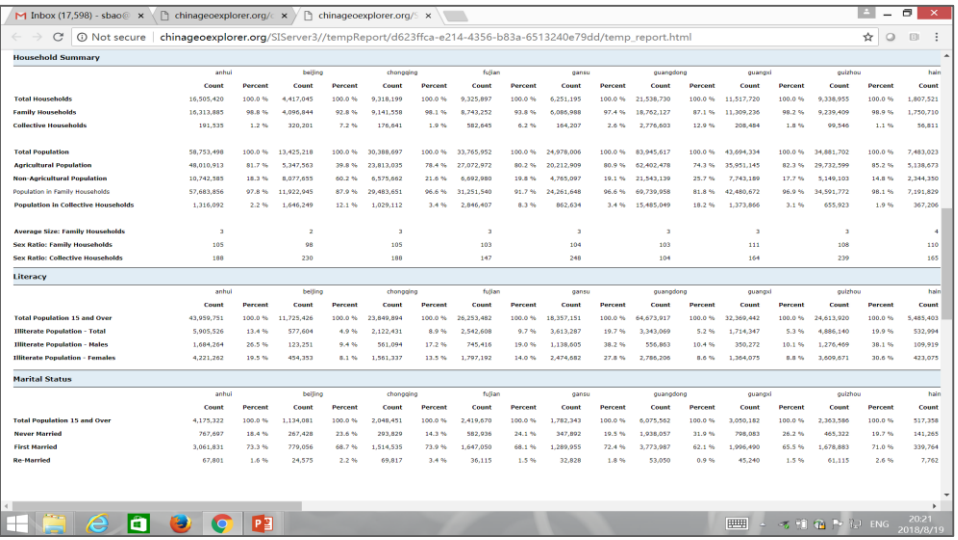
China Geo-Explorer

Administrative Units – Location – Chart – Time Series – Establishment – GIS Map Export – China-US Comparison – Map Library

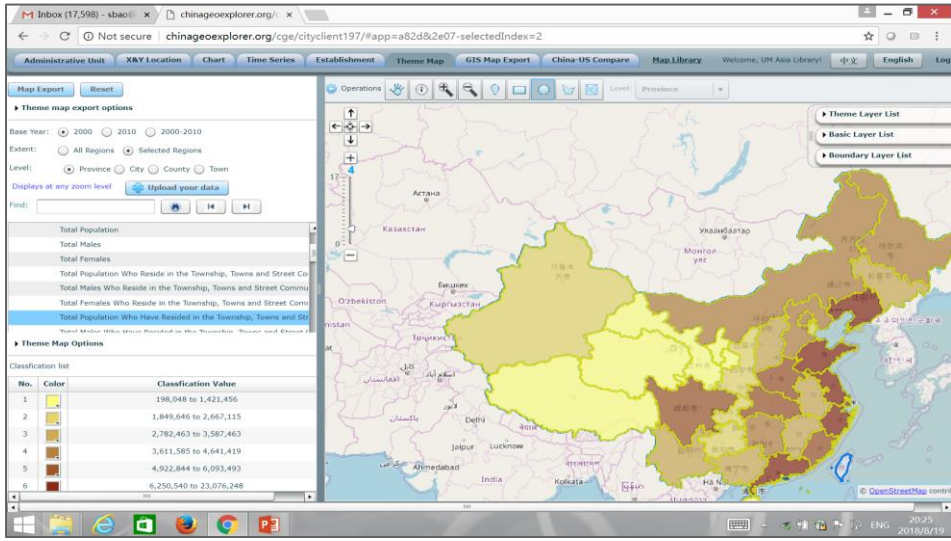


Tests of Spatial Patterns

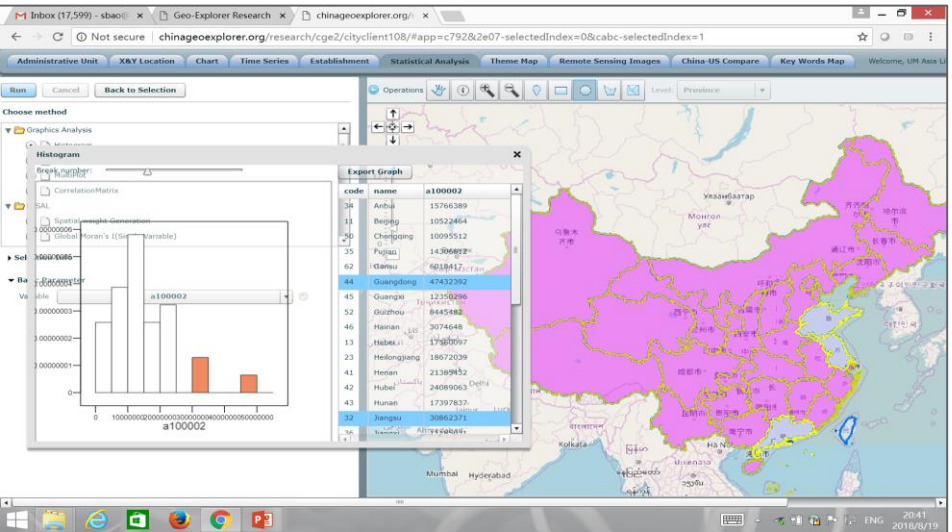
Reports



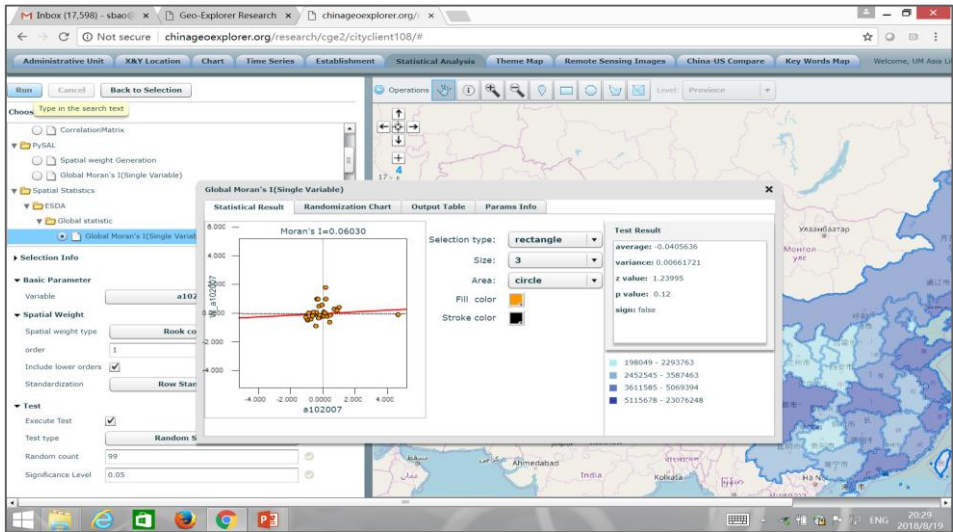
Maps



Charts: Histogram

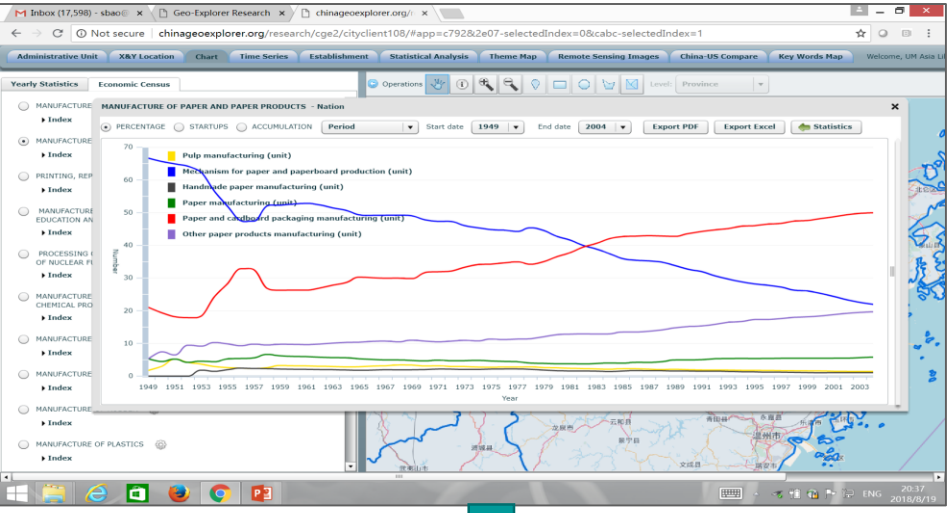


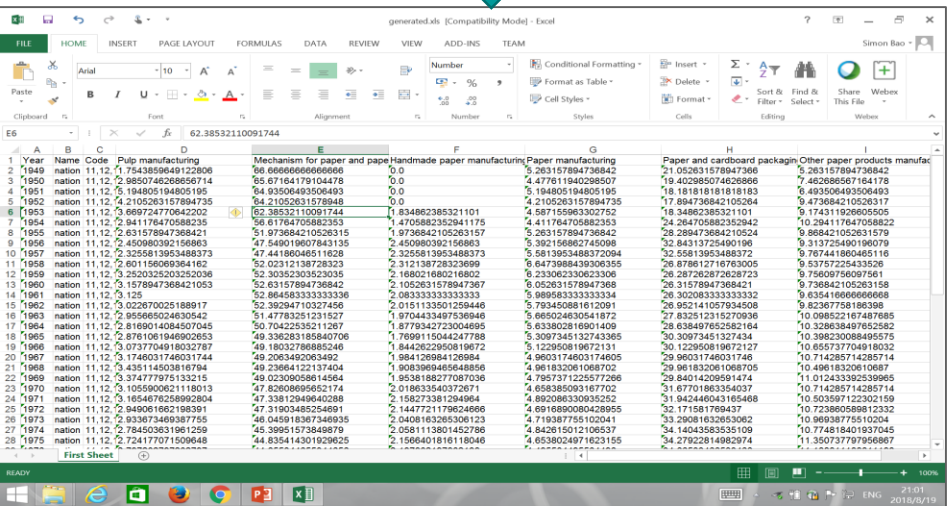
Morans' I: Spatial Autocorrelation



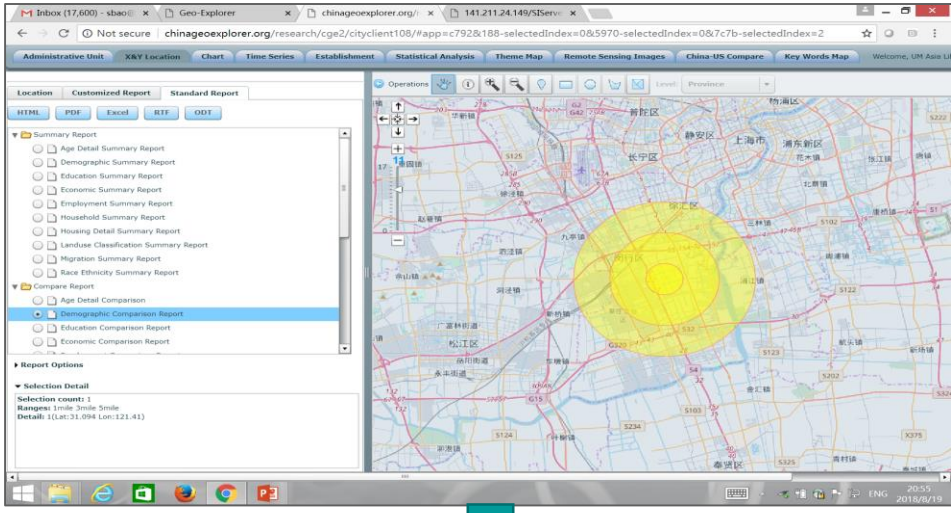
Tests of Spatial Structures

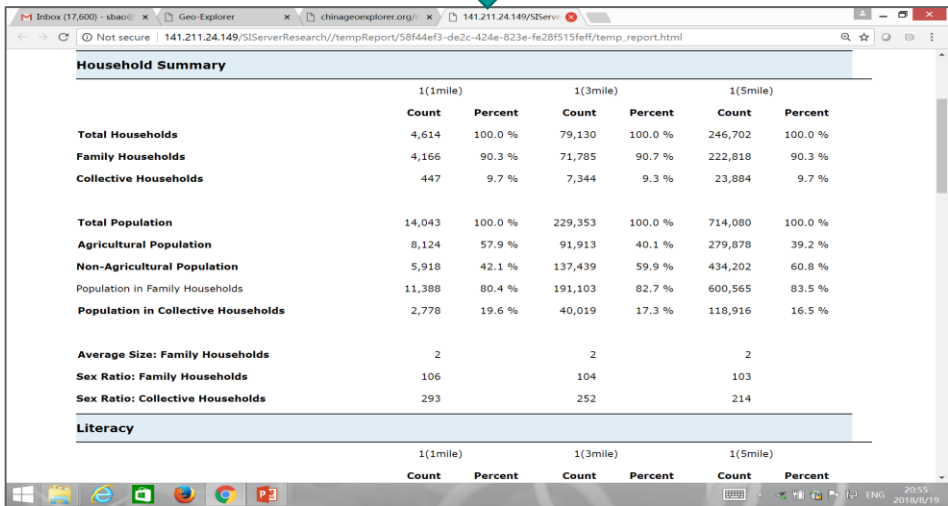
Industrial Structures





Spatial Structures





Household Summary

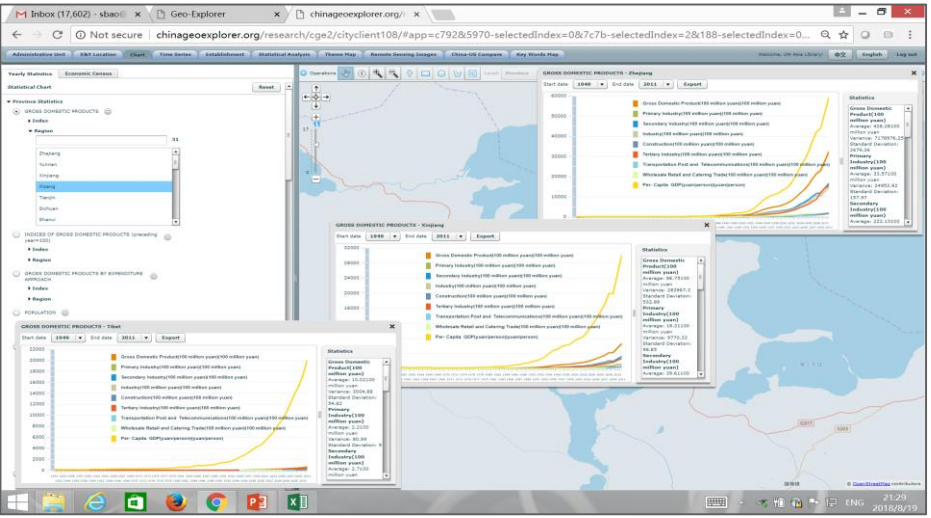
	1(1mile)		1(3mile)		1(5mile)	
	Count	Percent	Count	Percent	Count	Percent
Total Households	4,614	100.0 %	79,130	100.0 %	246,702	100.0 %
Family Households	4,166	90.3 %	71,785	90.7 %	222,818	90.3 %
Collective Households	447	9.7 %	7,344	9.3 %	23,884	9.7 %
Total Population	14,043	100.0 %	229,353	100.0 %	714,080	100.0 %
Agricultural Population	8,124	57.9 %	91,913	40.1 %	279,878	39.2 %
Non-Agricultural Population	5,918	42.1 %	137,439	59.9 %	434,202	60.8 %
Population in Family Households	11,388	80.4 %	191,103	82.7 %	600,565	83.5 %
Population in Collective Households	2,778	19.6 %	40,019	17.3 %	118,916	16.5 %
Average Size: Family Households	2		2		2	
Sex Ratio: Family Households	106		104		103	
Sex Ratio: Collective Households	293		252		214	

Literacy

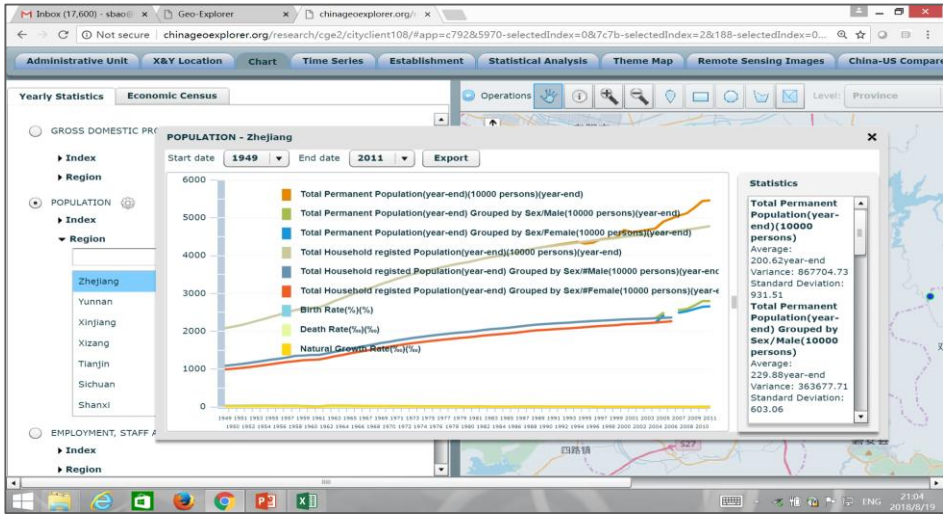
	1(1mile)		1(3mile)		1(5mile)	
	Count	Percent	Count	Percent	Count	Percent

Tests of Spatial Trends

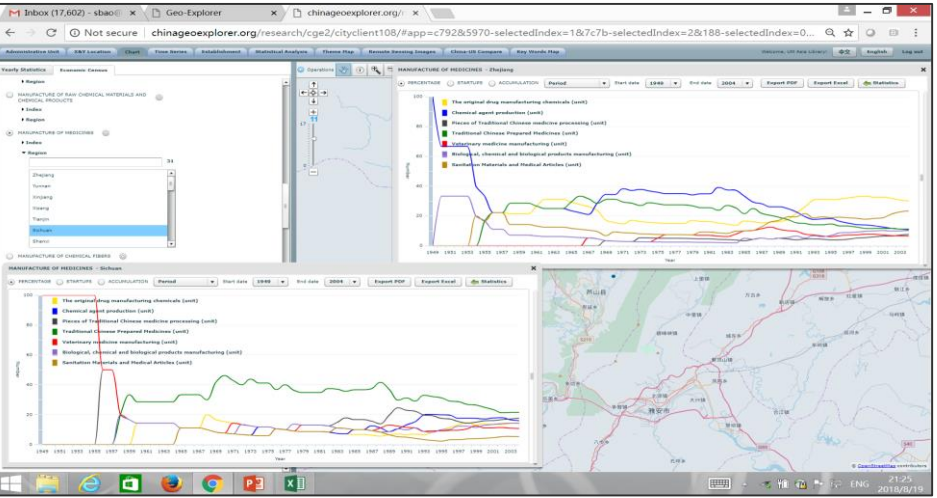
GDP: Zhejiang/Xinjiang/Tibet



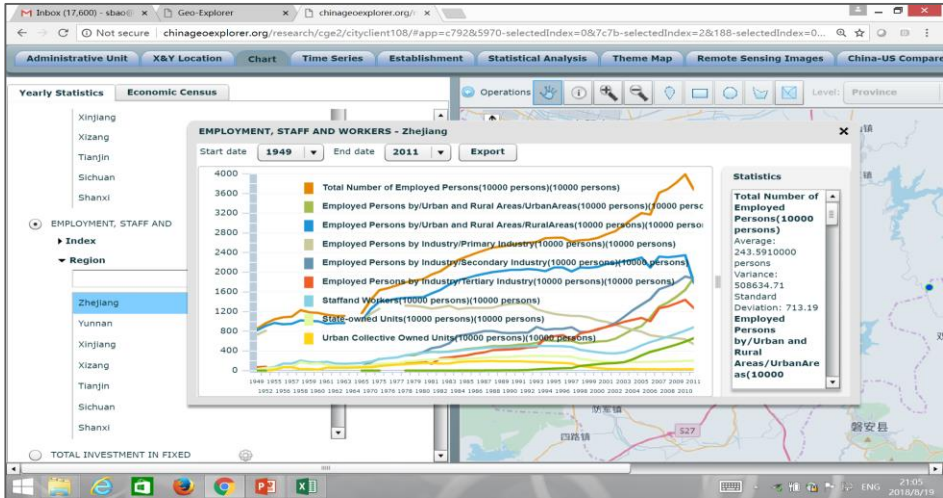
Population



Manufacture of Medicine: Zhejiang & Sichuan

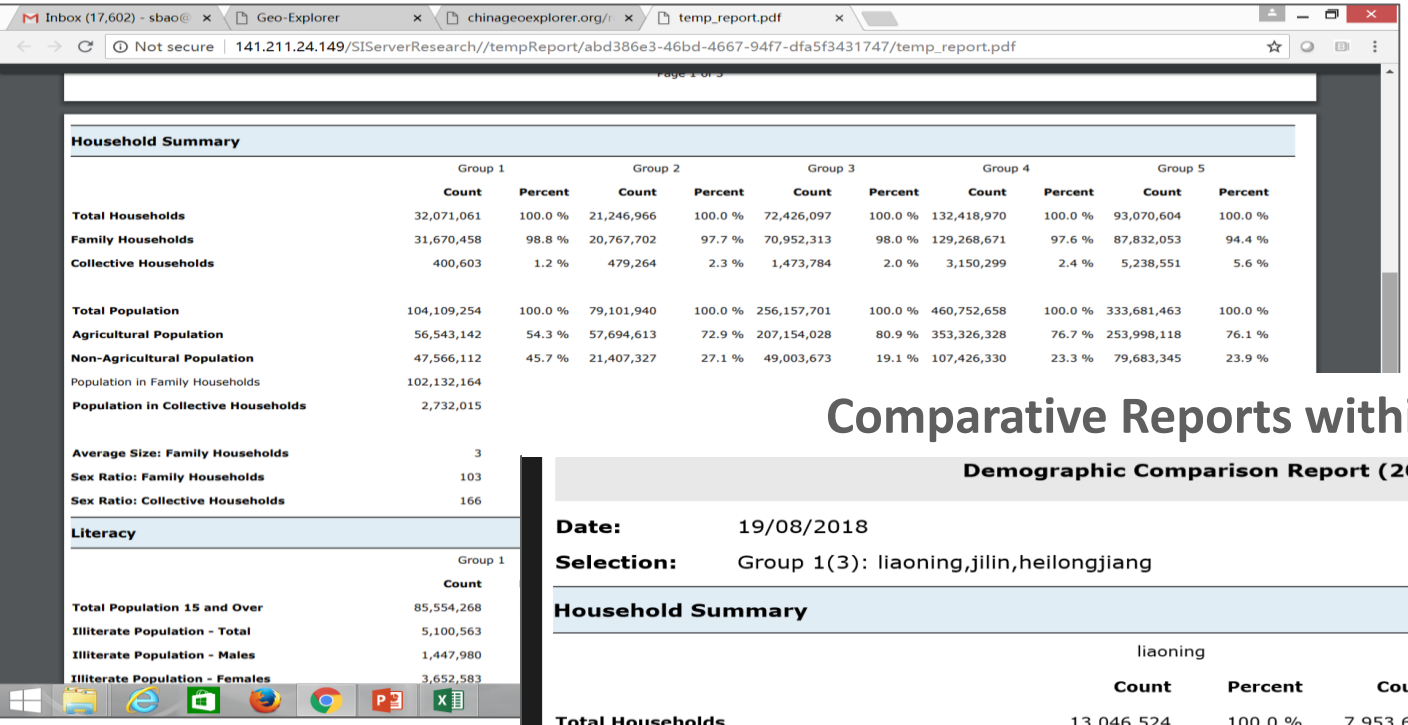


Employment



Tests of Inter and Intra-Regional Similarities

Comparative Reports between Groups



Household Summary

	Group 1		Group 2		Group 3		Group 4		Group 5	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Total Households	32,071,061	100.0 %	21,246,966	100.0 %	72,426,097	100.0 %	132,418,970	100.0 %	93,070,604	100.0 %
Family Households	31,670,458	98.8 %	20,767,702	97.7 %	70,952,313	98.0 %	129,268,671	97.6 %	87,832,053	94.4 %
Collective Households	400,603	1.2 %	479,264	2.3 %	1,473,784	2.0 %	3,150,299	2.4 %	5,238,551	5.6 %
Total Population	104,109,254	100.0 %	79,101,940	100.0 %	256,157,701	100.0 %	460,752,658	100.0 %	333,681,463	100.0 %
Agricultural Population	56,543,142	54.3 %	57,694,613	72.9 %	207,154,028	80.9 %	353,326,328	76.7 %	253,998,118	76.1 %
Non-Agricultural Population	47,566,112	45.7 %	21,407,327	27.1 %	49,003,673	19.1 %	107,426,330	23.3 %	79,683,345	23.9 %
Population in Family Households	102,132,164									
Population in Collective Households	2,732,015									
Average Size: Family Households	3									
Sex Ratio: Family Households	103									
Sex Ratio: Collective Households	166									

Literacy

	Group 1
	Count
Total Population 15 and Over	85,554,268
Illiterate Population - Total	5,100,563
Illiterate Population - Males	1,447,980
Illiterate Population - Females	3,652,583

Comparative Reports within Groups

Demographic Comparison Report (2000)

Date: 19/08/2018

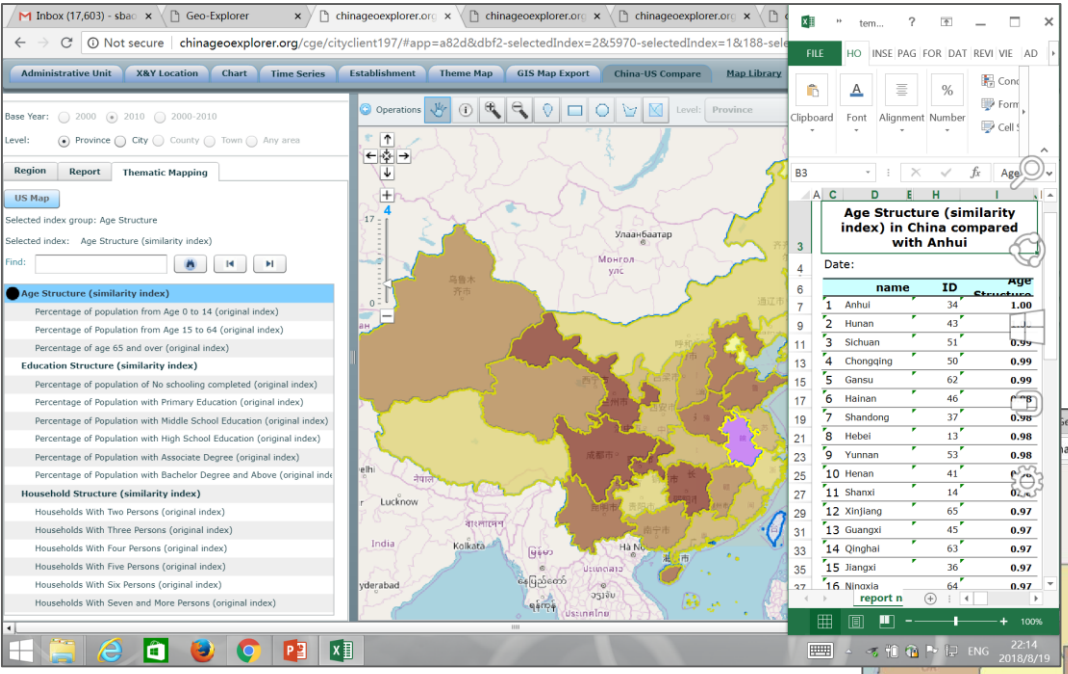
Selection: Group 1(3): liaoning,jilin,heilongjiang

Household Summary

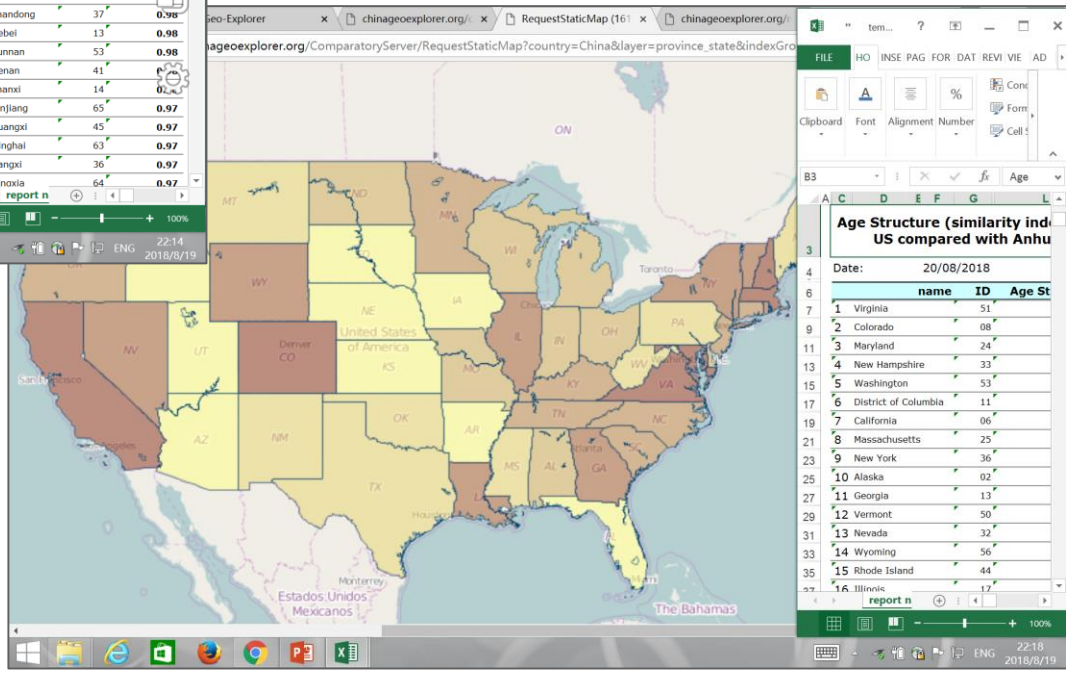
	liaoning		jilin		heilongjiang	
	Count	Percent	Count	Percent	Count	Percent
Total Households	13,046,524	100.0 %	7,953,615	100.0 %	11,070,922	100.0 %
Family Households	12,866,262	98.6 %	7,848,446	98.7 %	10,955,750	99.0 %
Collective Households	180,262	1.4 %	105,169	1.3 %	115,172	1.0 %
Total Population	41,553,243	100.0 %	26,591,484	100.0 %	35,964,527	100.0 %
Agricultural Population	22,462,770	54.1 %	15,076,240	56.7 %	19,004,132	52.8 %
Non-Agricultural Population	19,090,473	45.9 %	11,515,244	43.3 %	16,960,395	47.2 %
Population in Family Households	40,597,750	97.1 %	26,076,948	97.3 %	35,457,466	97.8 %
Population in Collective Households	1,226,662	2.9 %	725,243	2.7 %	780,110	2.2 %
Average Size: Family Households	3		3		3	
Sex Ratio: Family Households	102		103		103	
Sex Ratio: Collective Households	171		168		159	

Tests of Regional Similaries and Dissimilarities

Similarity Index between Anhui and Other Provinces of China



Similarity Index between Anhui and States of the US



Tests of Industrial Linkages by Space and Time

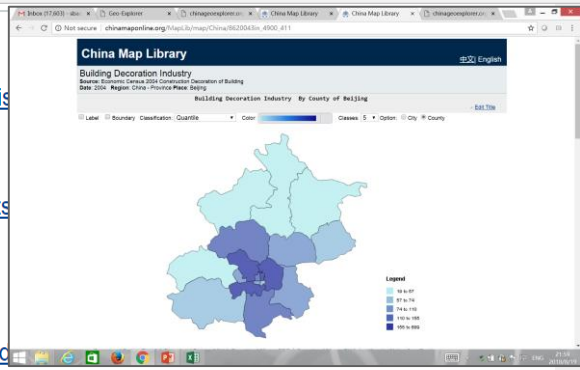
Inbox (17,603) - sbao | Geo-Explorer | chinageoexplorer.org | China Map Library | China Map Library | chinageoexplorer.org

Not secure | chinamaponline.org/MapLib/map/China/8620043in_4900_411

Rank of Related Industries 2004 By Spatial Relevance

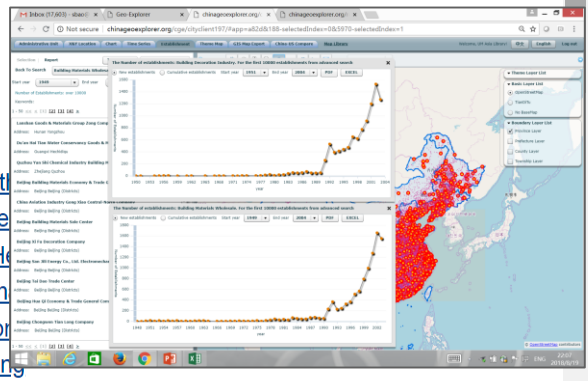
1 Other machinery and equipment rental
2 Building Materials Wholesale
3 Other cleaning services
4 Construction and installation industry
5 Cars, motorbikes Maintenance
6 Other food wholesale
7 Other interior decoration materials retail
8 Automotive retail
9 Flower retail
10 Cosmetics and health supplies wholesale
11 Management of urban greening
12 Paint retail
13 Real estate development and management
14 Hardware retail
15 Retail of Auto Parts and Accessories
16 Semiconductor manufacturing dis
17 Repair of household appliances
18 Fruit and vegetables wholesale
19 Cars, motorbikes and spare parts
20 Laundry and dyeing Services
21 Building cleaning services
22 Retail
23 Railways, roads, tunnels and bric
24 Metal and metal ore wholesale
25 Taxi passenger
26 Kitchen, bathroom appliances and daily groceries wholesale
27 Textiles, knitwear and raw materials wholesale
28 Western medicine wholesale
29 Other cultures and office machinery manufacturing
30 Furniture retail

Export All



Rank of Related Industries 2004 By Temporal Relevance (Startups)

1 Manufacture of metal doors and windows
2 Industrial Automation System device manufacturing
3 Property management
4 Project management services
5 Paint manufacturing
6 Retail of Auto Parts and Accessories
7 Electronic components and parts manufacturing
8 Supply with General Instruments and other equipment manufacturing
9 Transport agency services
10 Environmental Pollution Prevention Equipment
16 Wood furniture manufacturing
17 Other activities not specified in t
18 Processing crude oil and petrole
19 Manufacture of Nutritious and H
20 Power electronic components m
21 Distribution switch control equipr
22 Textile and apparel manufacturing
23 Refrigeration and air-conditioning equipment manufacturing
24 Hardware retail
25 Other real estate activities



ENG 21:59 2018/8/19

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

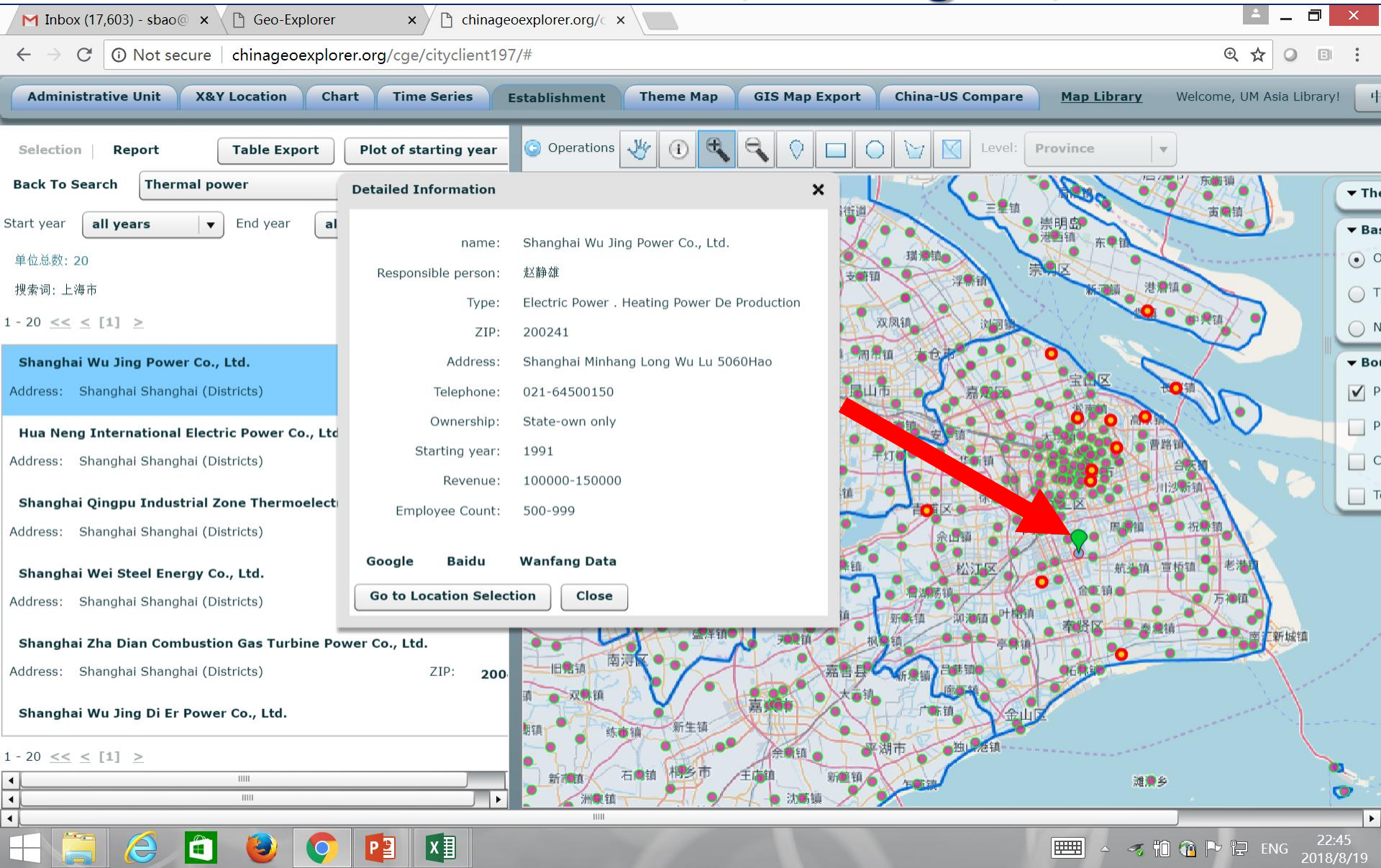
Selection: (4 selected) zhejiang , tianjin , y

Original Variables | **Spatially Weighted Variables**

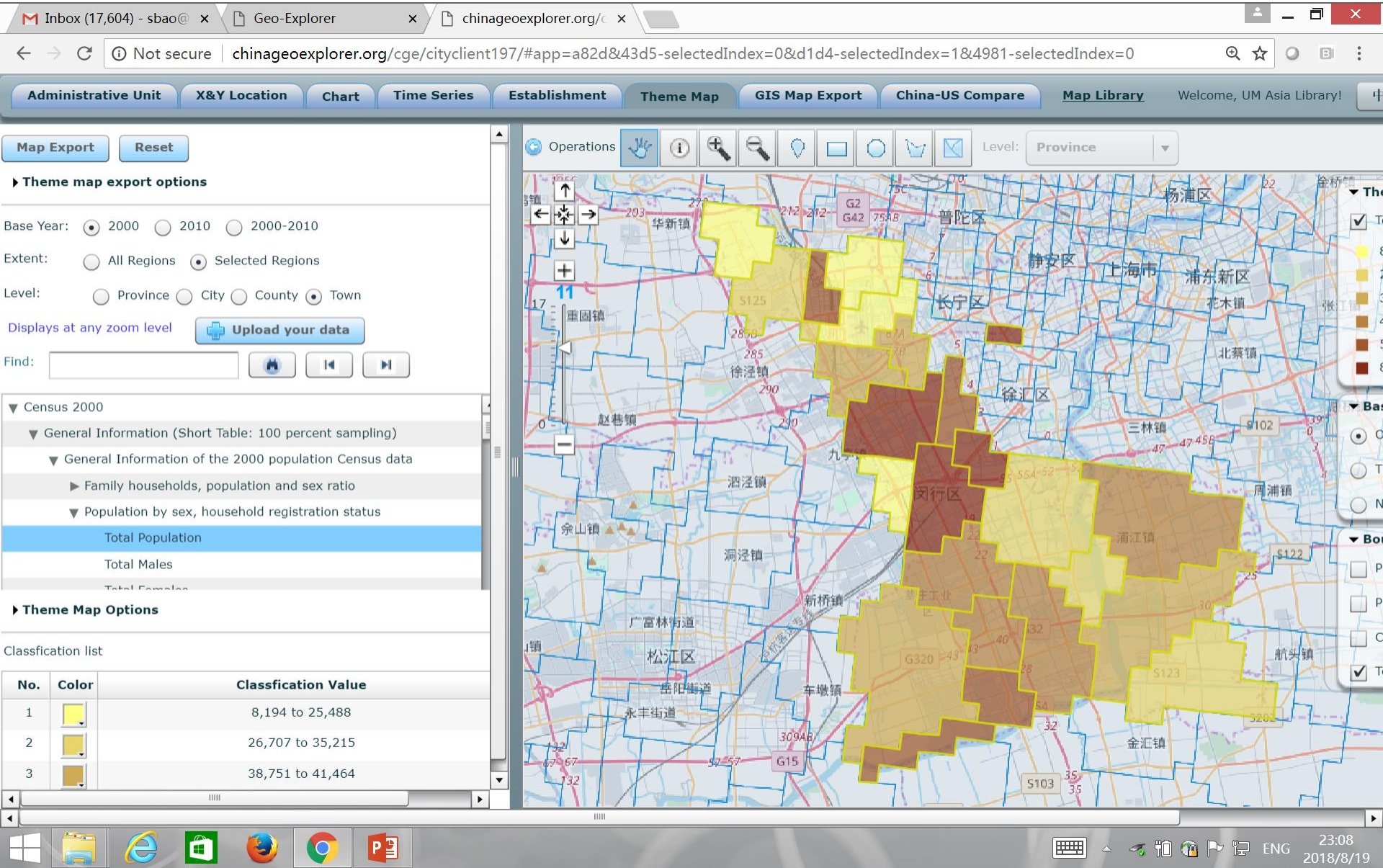
	name	year	Gross Domestic	Primary Industry	Secondary Industry	Gross Domestic	Primary Industry	Secondary Industry
1	tianjin	2009	7,521.9	128.8	3,987.8	15,159.8	1,324.7	7,640.2
2	tianjin	2010	9,224.5	145.6	4,840.2	18,081.4	1,526.9	9,226.2
3	tianjin	2006	4,344.3	118.2	2,488.3	9,570.7	918.0	4,939.7
4	tianjin	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

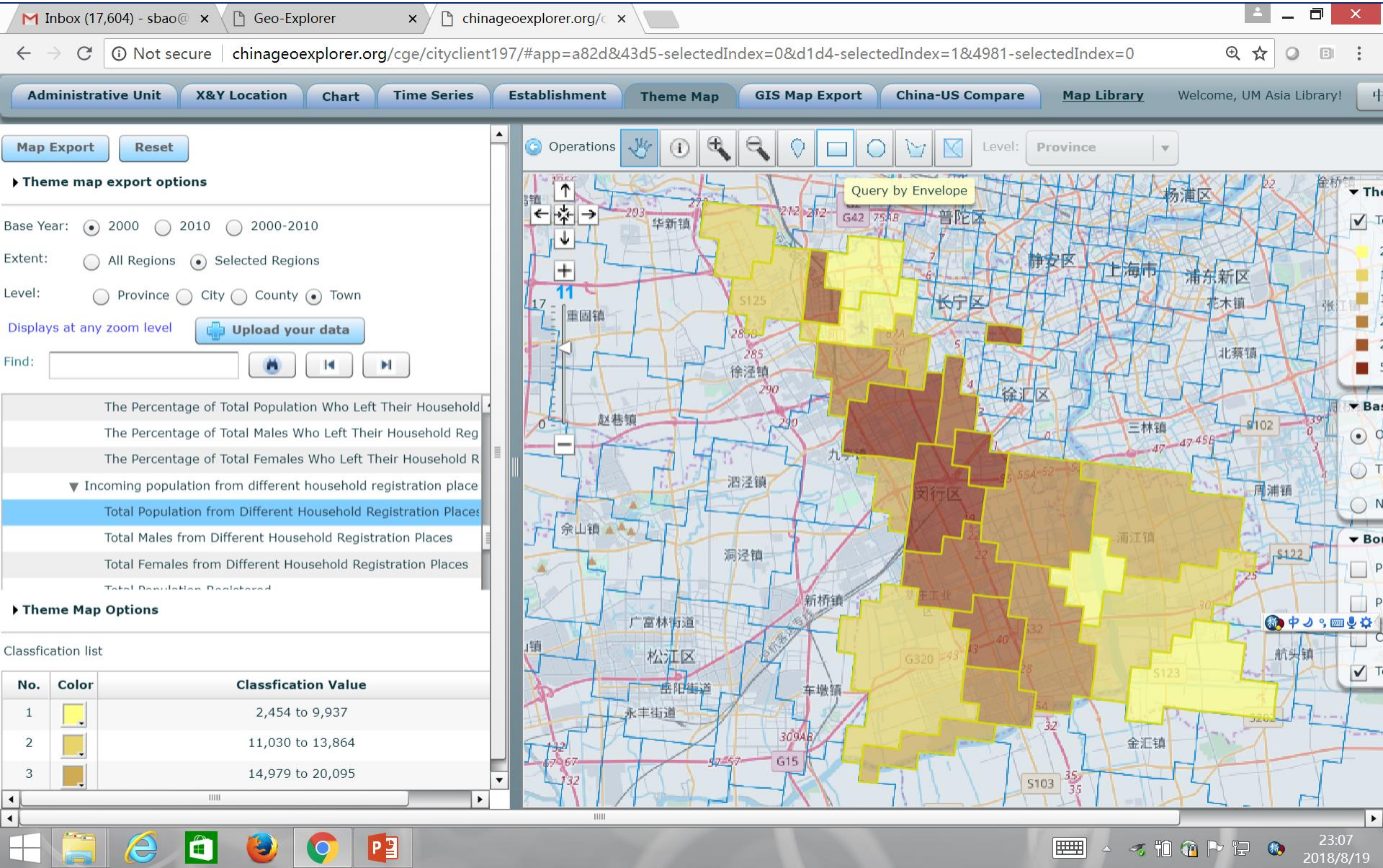
Case Study: Risk Assessment of Wujing Thermal Power Plant, Shanghai, China



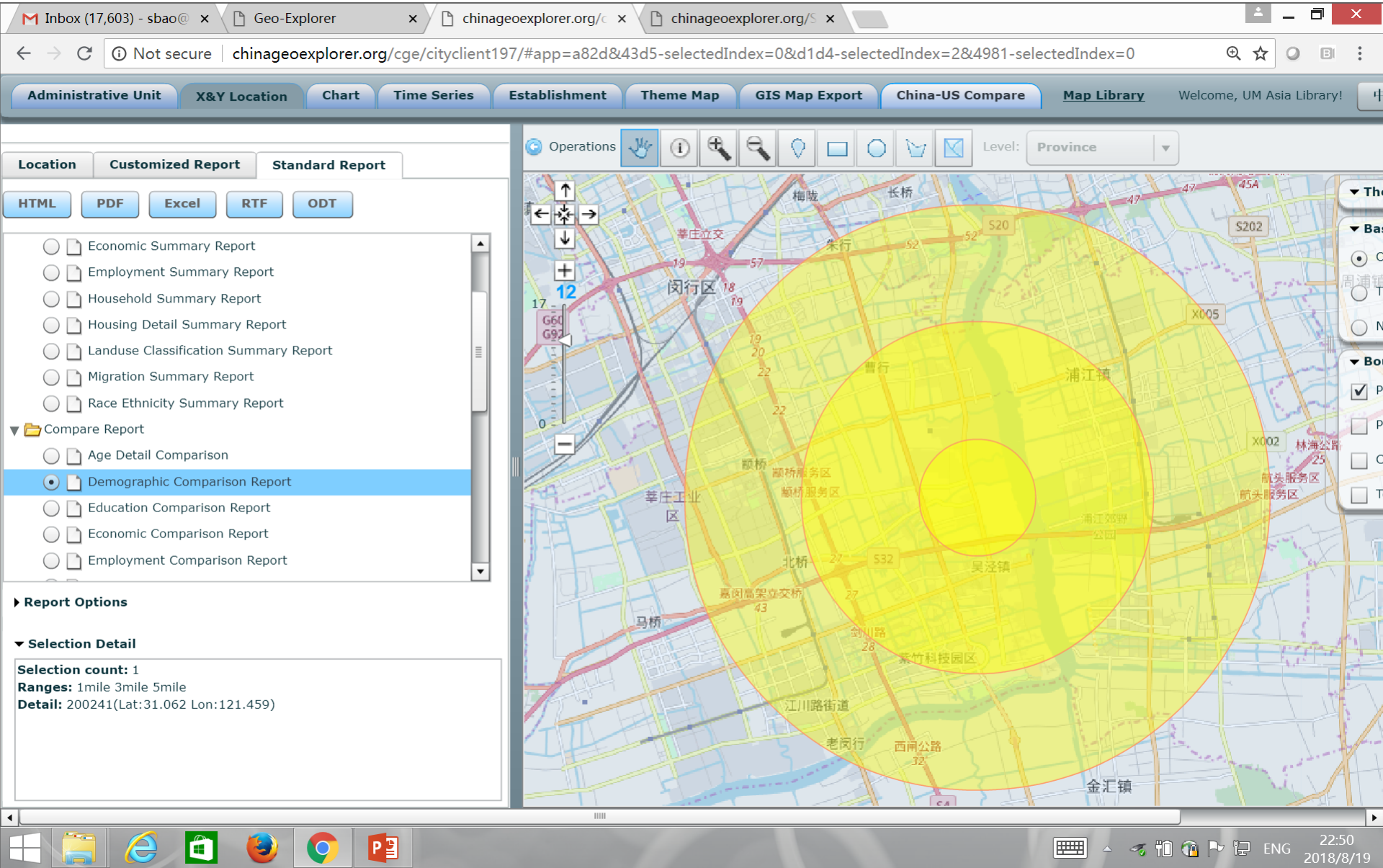
Population by Township of Minhang



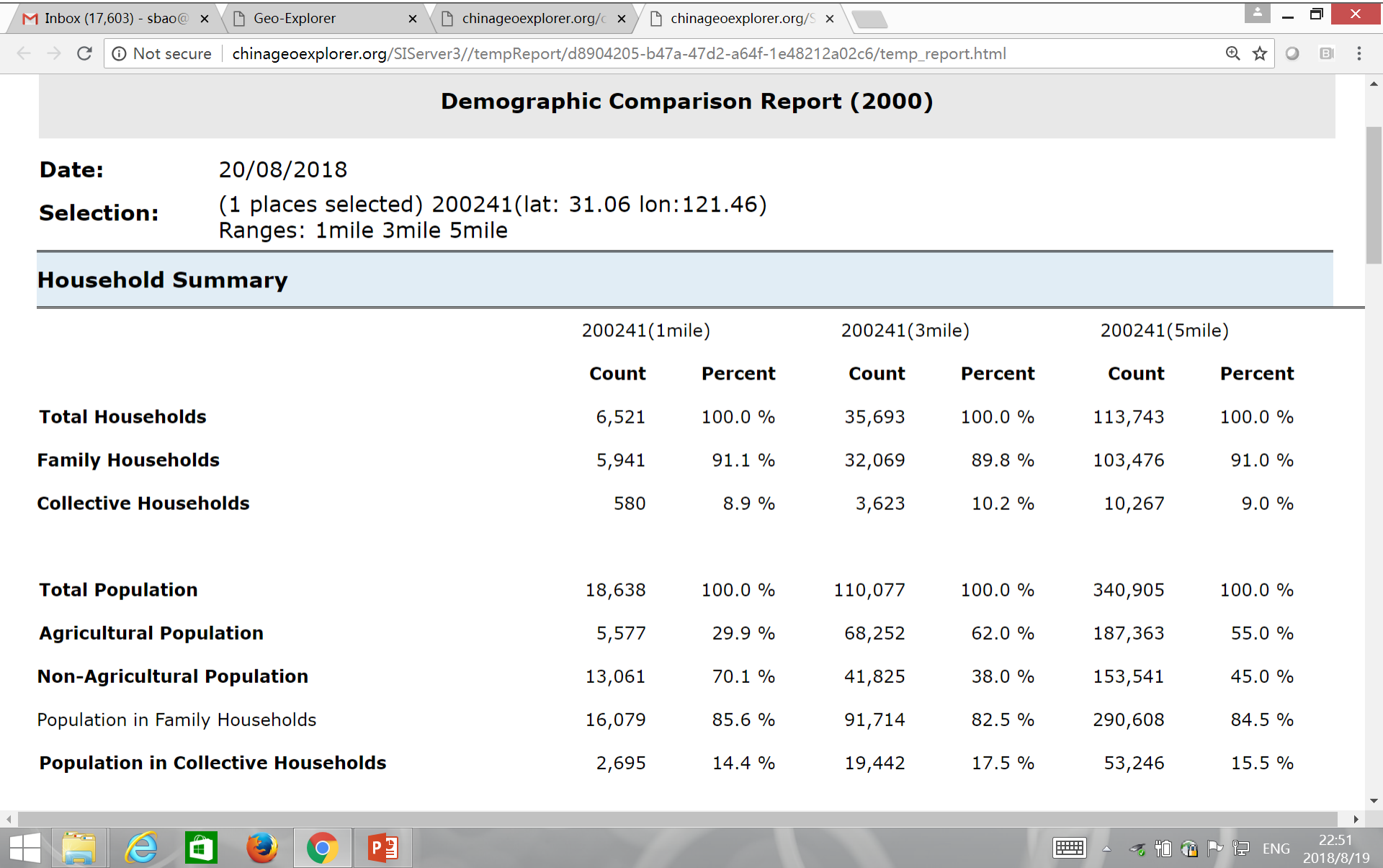
Migration from Other Places



Location Analysis



Demographic Report by Spatial Ranges



Demographic Comparison Report (2000)

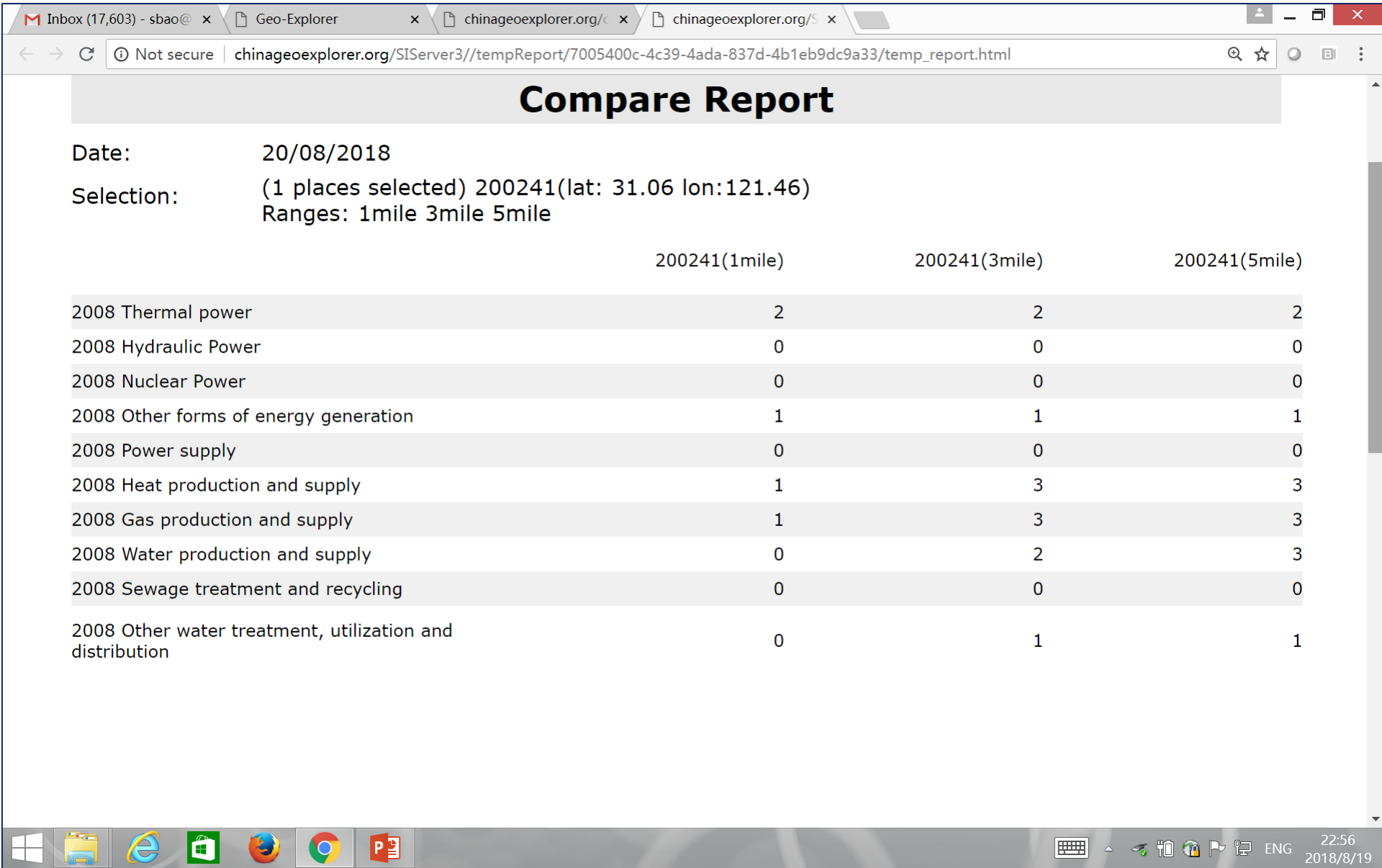
Date: 20/08/2018

Selection: (1 places selected) 200241(lat: 31.06 lon:121.46)
Ranges: 1mile 3mile 5mile

Household Summary

	200241(1mile)		200241(3mile)		200241(5mile)	
	Count	**Percent**	**Count**	**Percent**	**Count**	**Percent**
Total Households	6,521	100.0 %	35,693	100.0 %	113,743	100.0 %
Family Households	5,941	91.1 %	32,069	89.8 %	103,476	91.0 %
Collective Households	580	8.9 %	3,623	10.2 %	10,267	9.0 %
Total Population	18,638	100.0 %	110,077	100.0 %	340,905	100.0 %
Agricultural Population	5,577	29.9 %	68,252	62.0 %	187,363	55.0 %
Non-Agricultural Population	13,061	70.1 %	41,825	38.0 %	153,541	45.0 %
Population in Family Households	16,079	85.6 %	91,714	82.5 %	290,608	84.5 %
Population in Collective Households	2,695	14.4 %	19,442	17.5 %	53,246	15.5 %

Business of Power Industries by Spatial Ranges



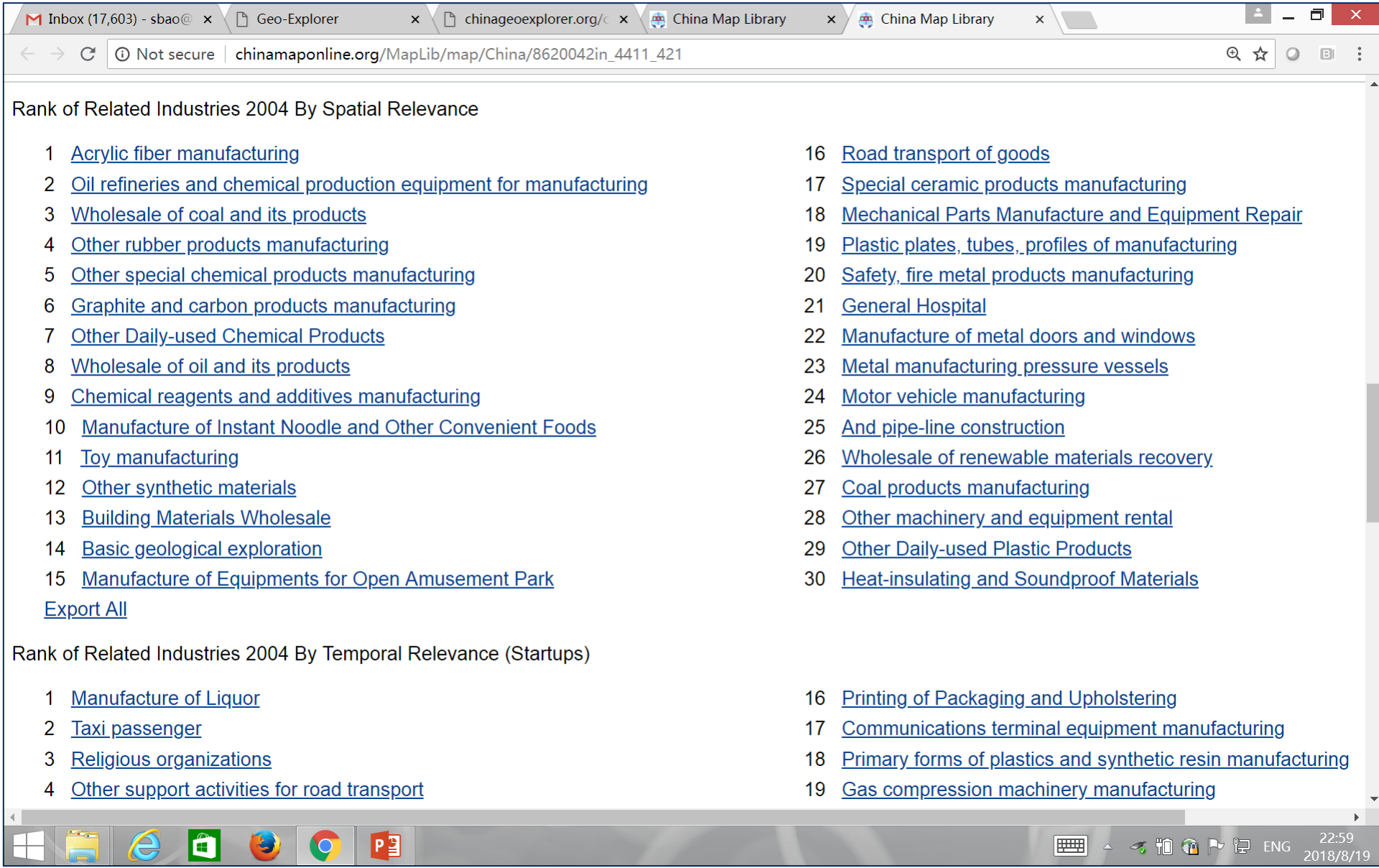
Compare Report

Date: 20/08/2018

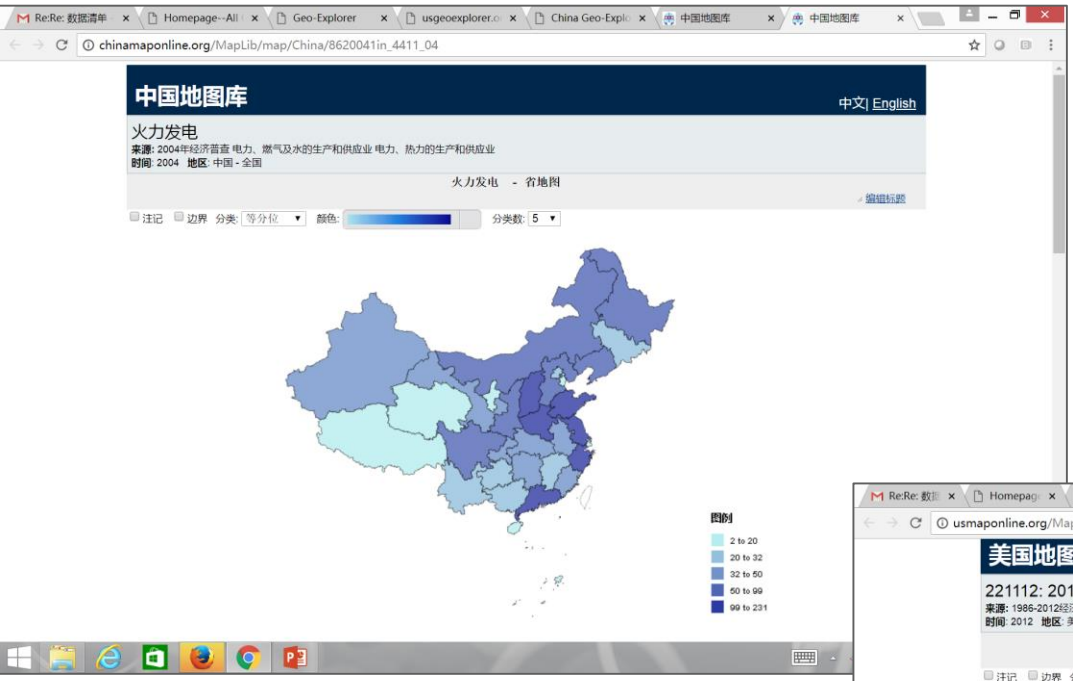
Selection: (1 places selected) 200241(lat: 31.06 lon:121.46)
Ranges: 1mile 3mile 5mile

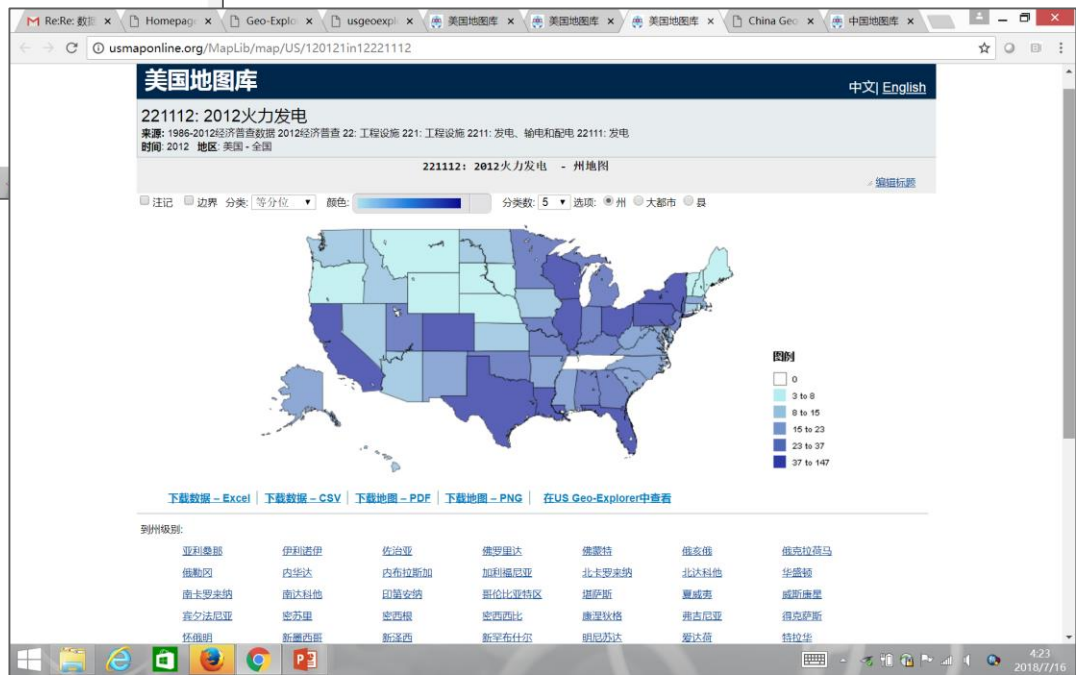
	200241(1mile)	200241(3mile)	200241(5mile)
2008 Thermal power	2	2	2
2008 Hydraulic Power	0	0	0
2008 Nuclear Power	0	0	0
2008 Other forms of energy generation	1	1	1
2008 Power supply	0	0	0
2008 Heat production and supply	1	3	3
2008 Gas production and supply	1	3	3
2008 Water production and supply	0	2	3
2008 Sewage treatment and recycling	0	0	0
2008 Other water treatment, utilization and distribution	0	1	1

Correlations between Thermal Power and Other Power Plants by Space and Time



Spatial Distribution of Thermal Power Plants in the US and China





Future Directions: Fragmented Data Stitching

Big Data Sources:

Mobile facilities
- Cell phones
- Mobile tablets

APPs
- QQ, WeChat, Facebook, Twitter, Tumblr etc.

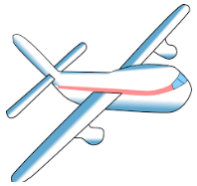
Transportation network
- Bus
- Train
- Bicycle

Traditional Business systems:
- Data produced by Public Agencies
- Data produced by businesses

Internet of Things:
- Data from sensors
- Data from online shopping

Features of Big Data:

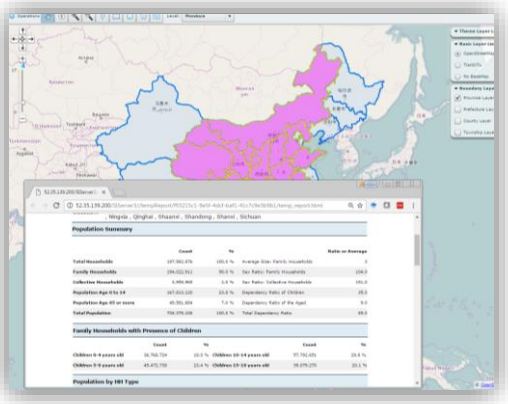
- **Multi-sources**
- **Space and time**
- **Classifications**
- **Complex structure**



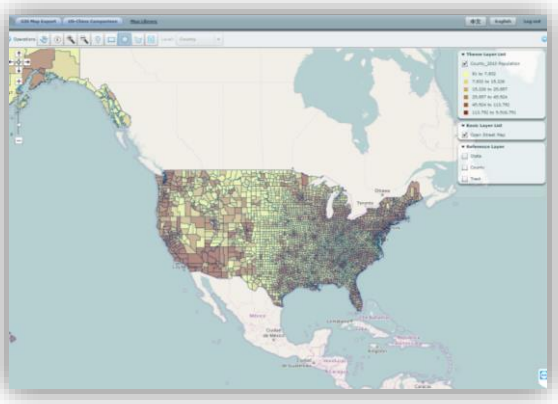
amazon

Google

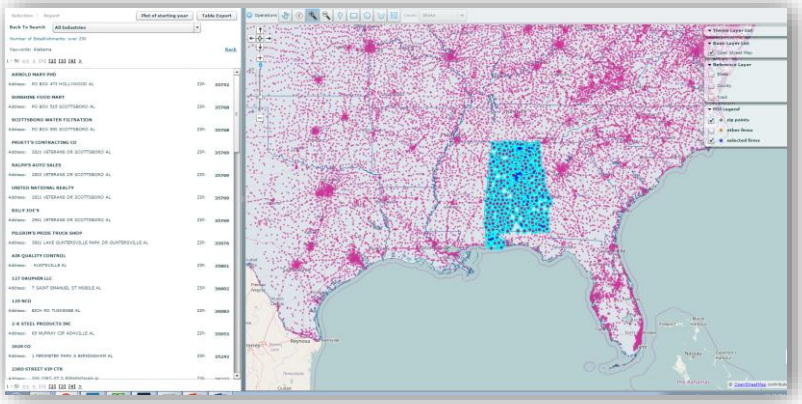
Geo-Explorers: Integrated Data Analysis



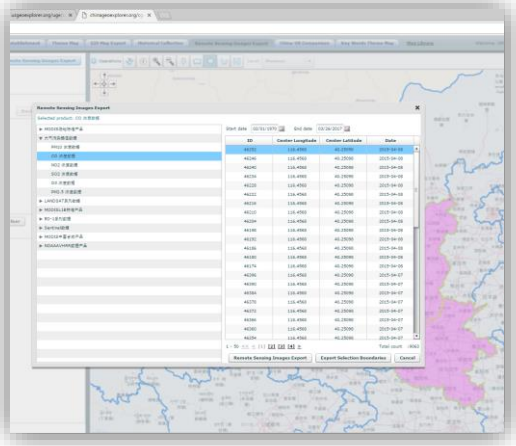
Reporting



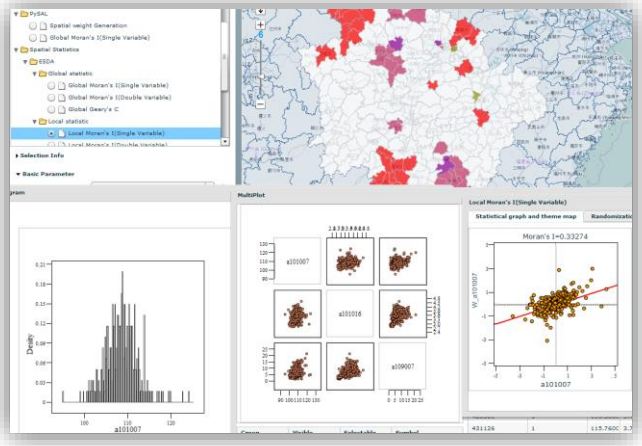
Thematic Maps



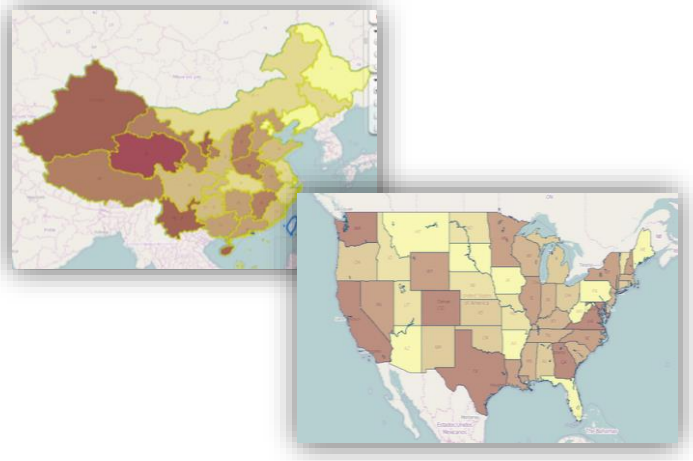
Establishment search



Remote Sensing Data



Exploratory Spatial Data Analysis

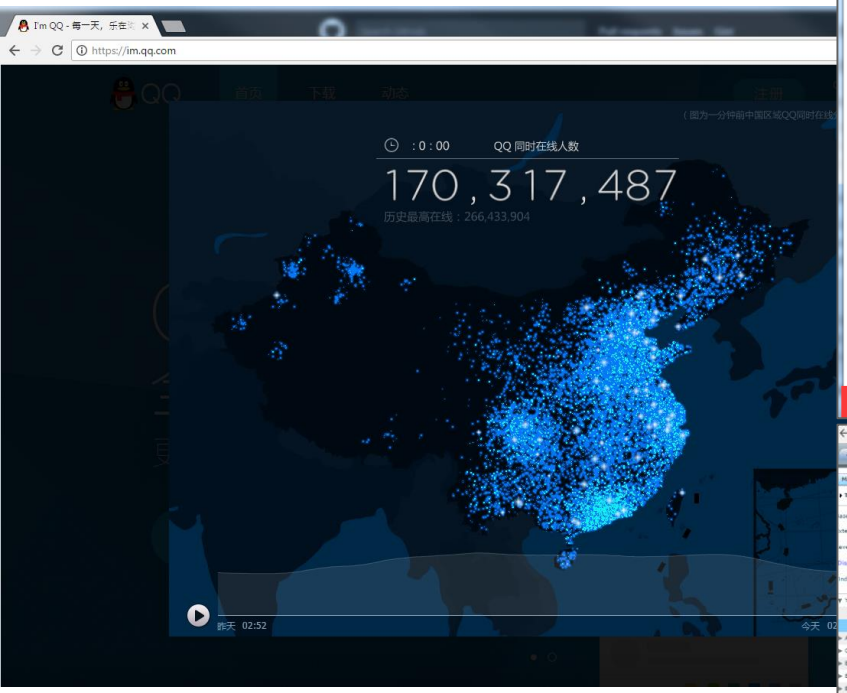


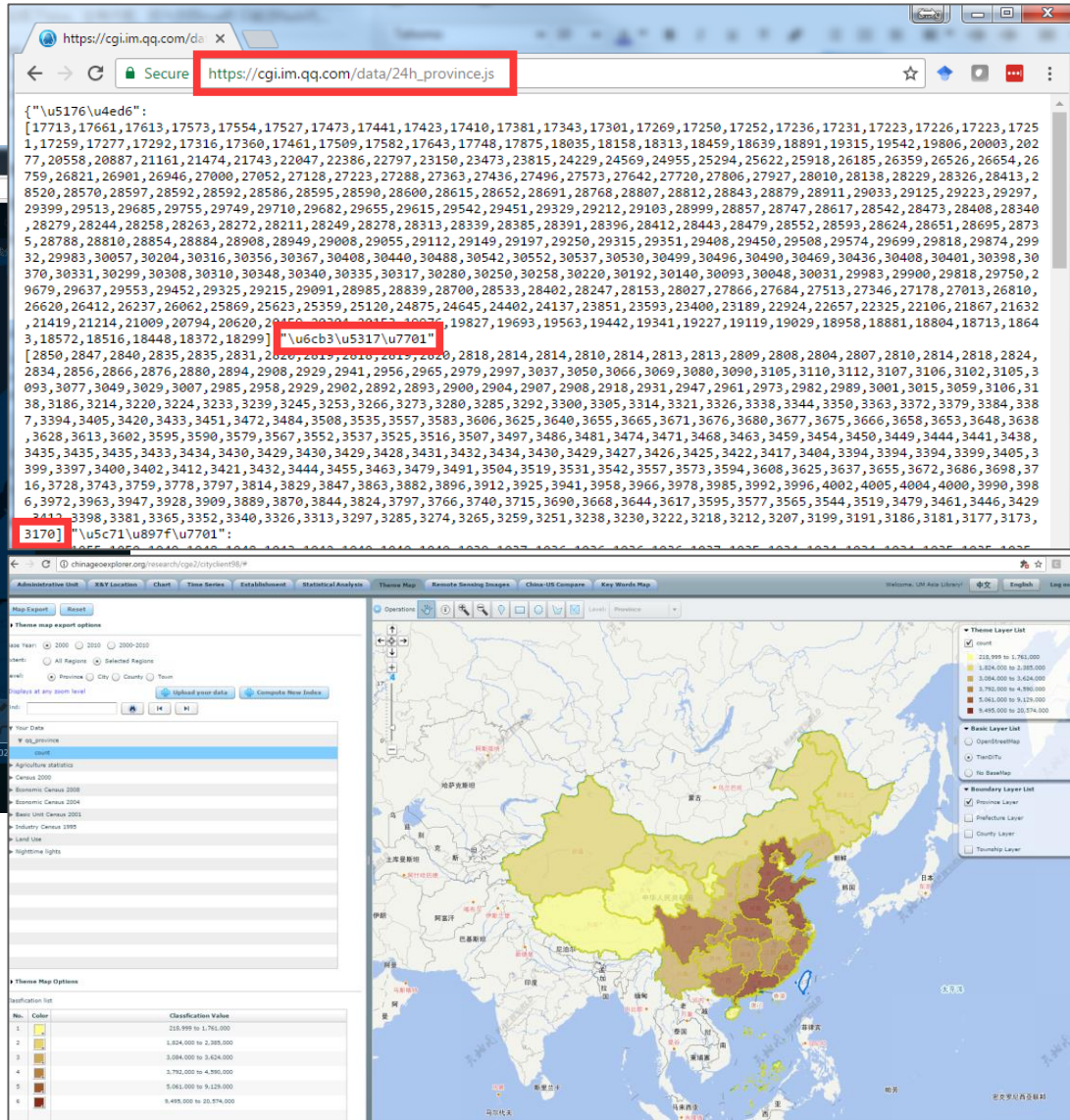
China-US Comparative analysis

Identifying Spatial Population with QQ Data

https://cgi.im.qq.com/data/24h_province.js

https://im.qq.com/





Identifying Population Activities with Twitter Data

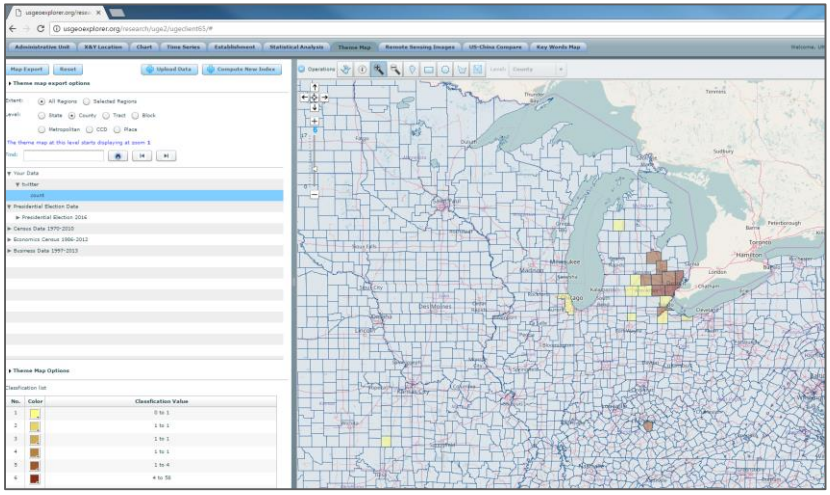
- Step 1: API Register
- Step 2: Search Twitter Data through Tweepy
- Step 3: Transform Twitter Data into GeoJson
- Step 4: Upload to GeoExplorer
- Step 5: Spatial analysis and visualization

https://dev.twitter.com/overview/api/tweets


```
"place":
{
    "attributes":{},
     "bounding_box":
    {
        "coordinates":
        [[
                [-77.119759,38.791645],
                [-76.909393,38.791645],
                [-76.909393,38.995548],
                [-77.119759,38.995548]
        ]],
        "type":"Polygon"
    }
```

```
"coordinates":
{
    "coordinates":
    [
        -75.14310264,
        40.05701649
    ],
    "type":"Point"
}
```

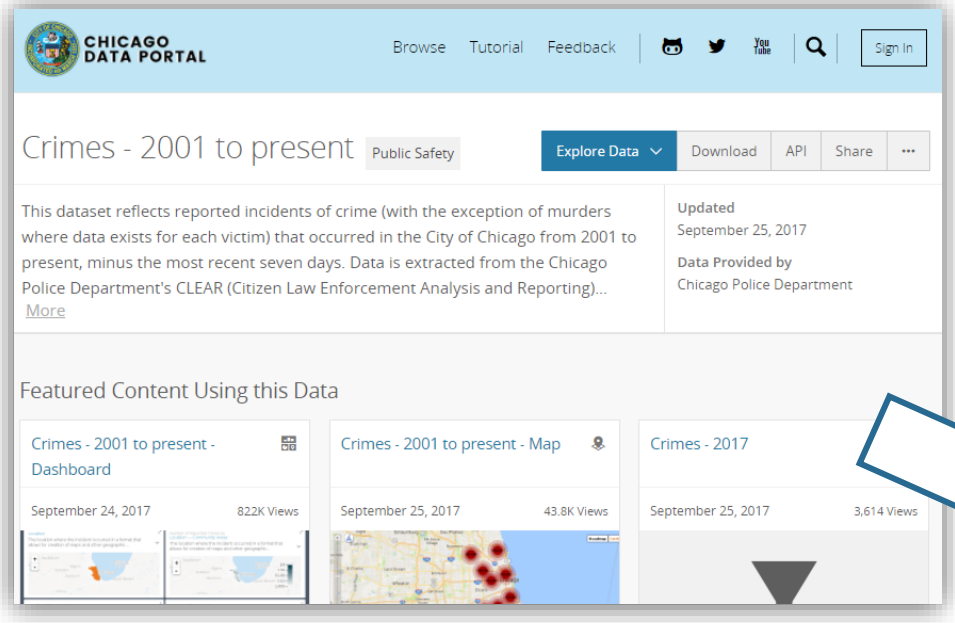
```
"location":"San Francisco, CA",
```

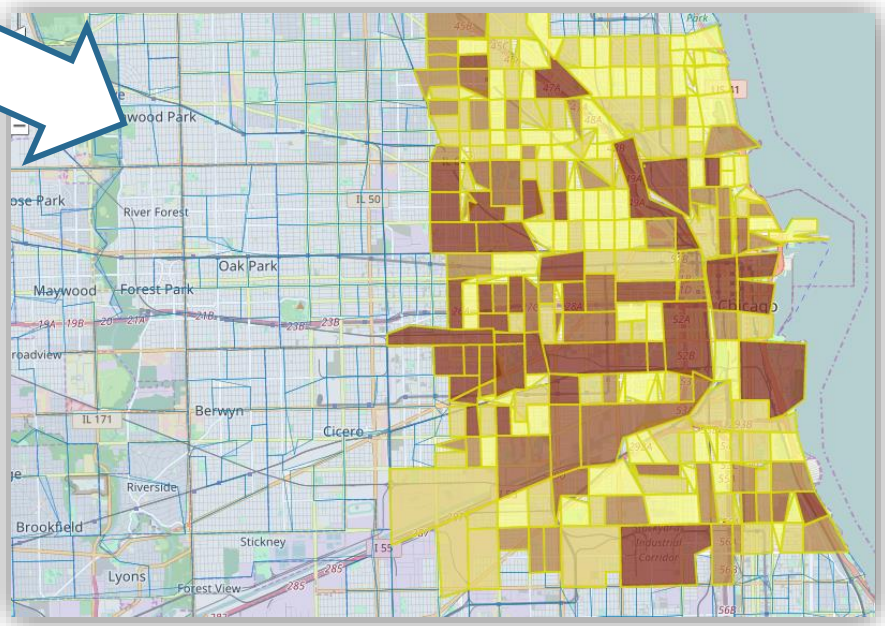
Identifying Spatial Crime with Web Crawling

Open Government Data

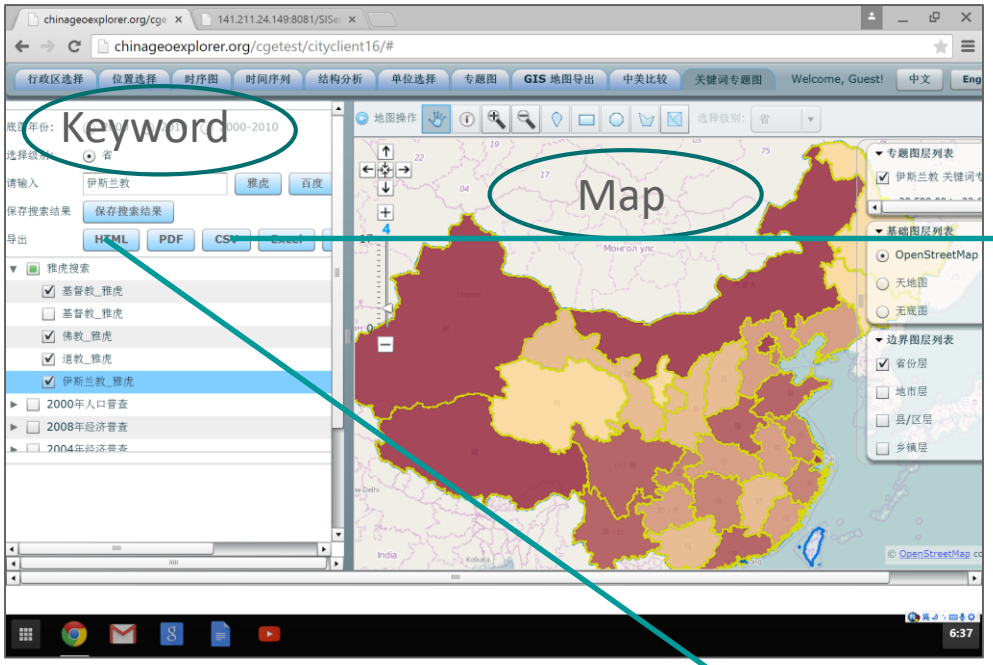
https://data.cityofchicago.org/

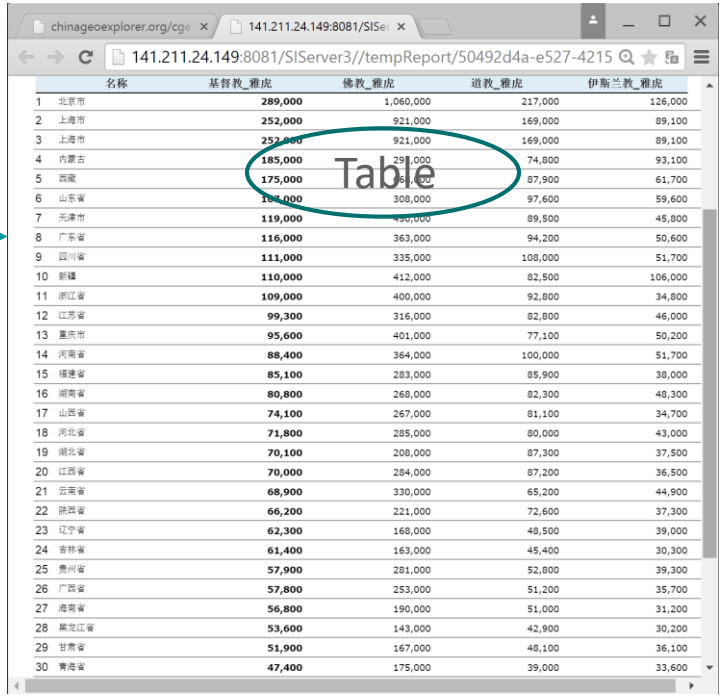


Crime Map by Tract



Big Data Integration – Web Search





Results from keyword search:

- Non-structural data -> Structural data
- Non-spatial data -> Spatial data
- Text data -> Numeric data
- Unlimited keywords -> Unlimited data
- Complimentary data to baseline data

http://china-data-online.com

Contact

China Data Institute

office@chinadatacenter.net

(734)680-3943