Introduction to the People Flow Data

Provided by CSIS, University of Tokyo

HOU Hao
Division of Spatial Information Science
University of Tsukuba
2015.7.30

129th GIS Seminar

CONTENTS

Introduction to the People Flow Data

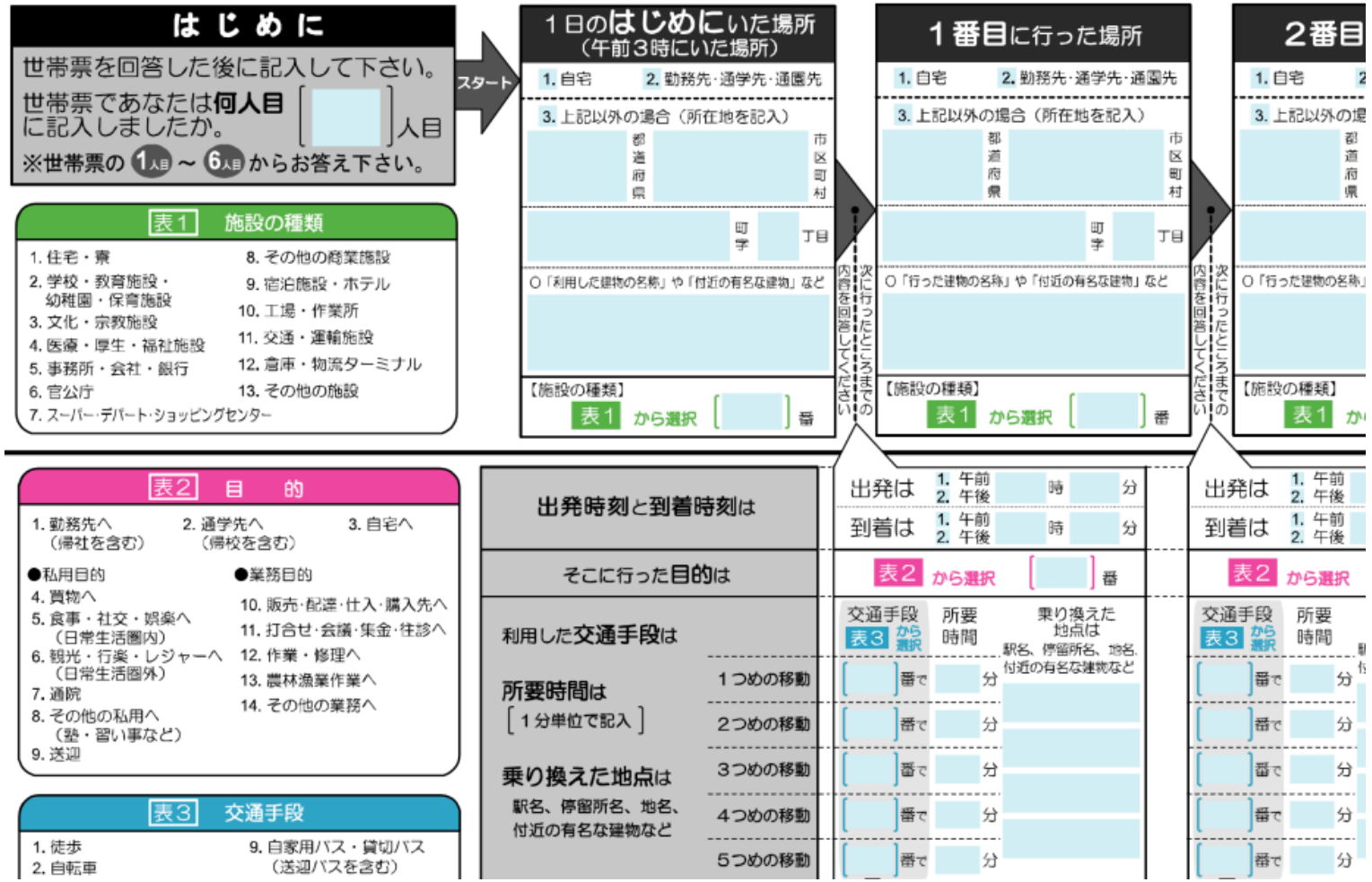
Provided by CSIS, University of Tokyo

Data Introduction

Data Source

Person Trip Survey

- Questionnaires
- Capture movement diaries of one day



(Source from CSIS: http://pflow.csis.u-tokyo.ac.jp/wp-content/uploads/20140724.pflow-ws-tutorial.pdf)

Data Creating Procedure

Qusetionnaire

↓

Geocoding

↓

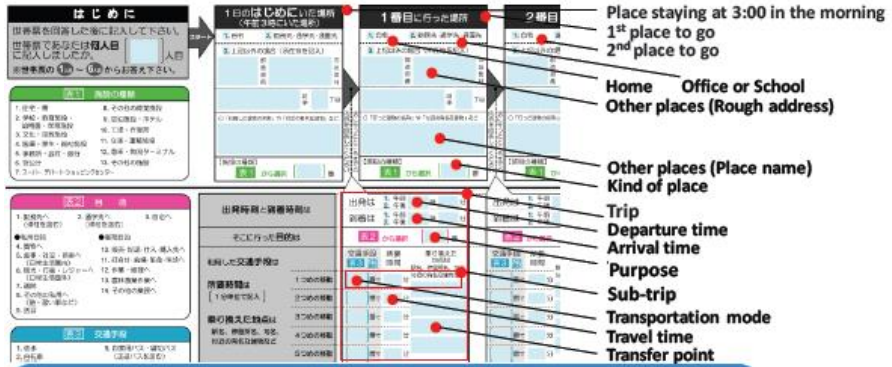
Route choice

↓

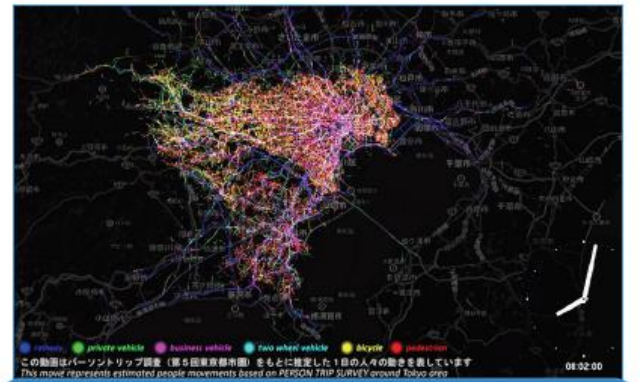
Interpolation

↓

Result by minute



Questionnaire Survey(2008 Tokyo PT)

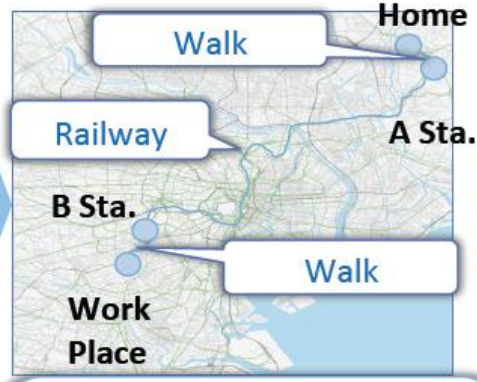


Estimated People Flow Data (Tokyo)

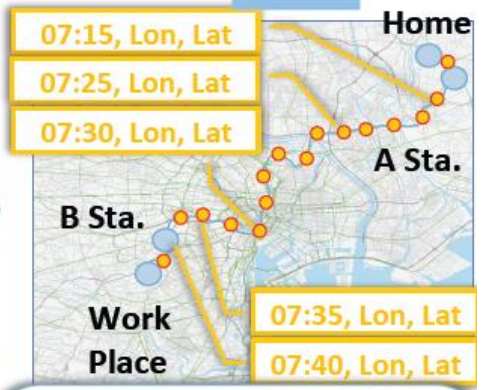
Input

Result

Route Search



Spatiotemporal Interpolation



1) Geocoding OD Position
* Spatial reallocation

2) Route choice along with road/railway topology

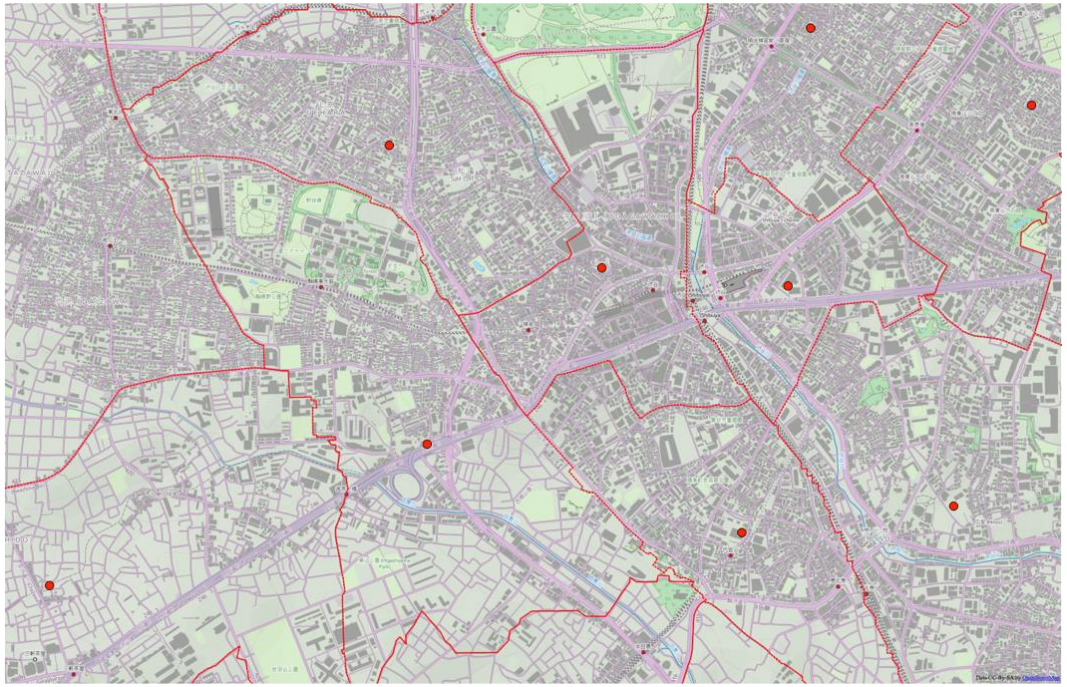
3) Interpolation into 1 minute intervals

(Source from CSIS: http://pflow.csis.u-tokyo.ac.jp/wp-content/uploads/20140724.pflow-ws-tutorial.pdf)

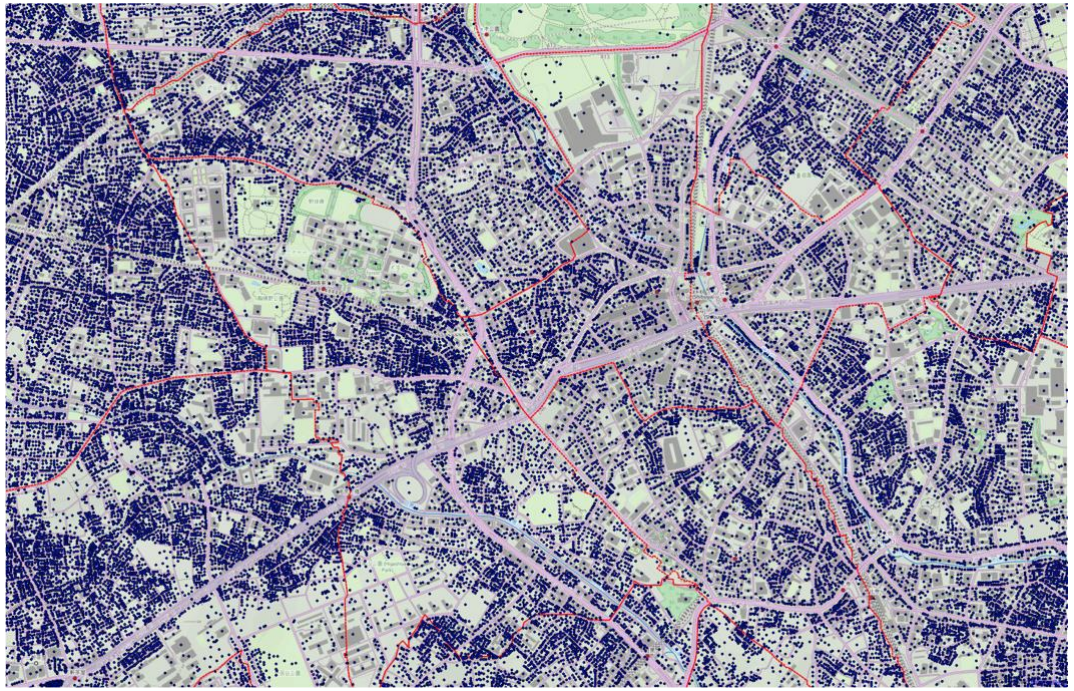
Spatial Distribution Data

Transition from OD-zone to building-based

Old Version



New Version



(Source from CSIS: http://pflow.csis.u-tokyo.ac.jp/wp-content/uploads/20140724.pflow-ws-tutorial.pdf)

Available Datasets (Spatial Distribution Version)

- Tokyo 1998, 2008
- Kyoto/Osaka/Kobe 2000
- Nagoya 2001, 2011
- Central Shizuoka 2001
- Northern Kyushu 2005

Available Datasets (OD Zone Version, Japan)

- Tokyo 1988
- Central Hokkaido 2006
- Okinawa 2006
- Yamaguchi/Kofu 2003
- Sendai 2002
- Matsuyama 2007
- Toyama/Takaoka 1999
- Kochi 1997
- Nagano 2001
- Kanazawa 2007
- Shizuoka 2001
- Miyazaki 2001
- Koriyama 2006
- Asahigawa 2002
- Okayama 1994
- Akita 2005

Available Datasets (OD Zone Version, International)

- Manila(Philippines), 1996
- Jakarta(Indonesia), 2000
- Hanoi(Vietnam), 2004
- Dhaka(Bangladesh), 2009

Data Structure

Personal attributes

- Person ID
- Gender
- Age
- Home address
- Occupation
- Magnification Factor

Trip attributes

- Trip ID
- Sub Trip ID
- Purpose
- Transportation mode

Field ID	Field Name	Description
1	PID	Unique person ID
2	TNO	Trip number
3	SNO	Sub trip number
4	LON	Longitude position
5	LAT	Latitude position
6	GENDER	Gender
7	AGE	Age group
8	ZCODE	Current location by zone code
9	OCCUP	Person occupation
10	PURPOSE	Purpose to trip
11	MAGFAC	Adjustment Factor
12	MAGFAC2	Adjustment Factor
13	TCODE	Mode of transportation

Data Structure (Tokyo 2008)

Age

Code	Value	Code	Value
1	Age between 0 - 5	10	Age between 45 - 50
2	Age between 5 - 10	11	Age between 50 - 55
3	Age between 10 - 15	12	Age between 55 - 60
4	Age between 15 - 20	13	Age between 60 - 65
5	Age between 20 - 25	14	Age between 65 - 70
6	Age between 25 - 30	15	Age between 70 - 75
7	Age between 30 - 35	16	Age between 75 - 80
8	Age between 35 - 40	17	Age between 80 - 85
9	Age between 40 - 45	18	Age above 85

Gender

Code	Value
1	Male
2	Female
9	Unknown

(Source from SIS, University of Tsukuba: http://land.geo.tsukuba.ac.jp/persontrips/explanation.htm)

Data Structure (Tokyo 2008)

Occupation

Code	Value	Code	Value
1	Agricultural/Forestry/Fishery	9	Manager
2	Labor/Factory (Blue Collar)	10	Other Occupation
3	Sales	11	Elementary and Junior-high Student
4	Service	12	High School Student
5	Transport Service	13	College and University Student
6	Security Service	14	House-wife
7	Office Worker	15	No-occupation
8	Professional	16	Others (Not Categorized)
		99	Unknown

(Source from SIS, University of Tsukuba: http://land.geo.tsukuba.ac.jp/persontrips/explanation.htm)

Data Structure (Tokyo 2008)

Transportation mode

Code	Value	Code	Value
1	Walk	9	Private bus
2	Bicycle	10	Public bus
3	Motor-bicycle	11	Monorail transit
4	Motor-bike	12	Train/subway
5	Taxi	13	Ship
6	Passenger vehicles	14	Aircraft
7	Mini car	15	Others
8	Freight vehicle	97	No movement
		99	Unknown

(Source from SIS, University of Tsukuba: http://land.geo.tsukuba.ac.jp/persontrips/explanation.htm)

Data Structure (Tokyo 2008)

Trip Purpose

Code	Value	Code	Value
1	To-From Office	9	To Send/Pick Up Activity
2	To-From School	10	For Selling and Buying
3	To Home	11	For Appointment
4	For Shopping	12	To/For Work (Fixing and Repairing)
5	For Short Recreation	13	To Agri./Forestry/Fishery Work
6	For Sight Seeing and Leisure	14	Other Business Purpose
7	For Medical Treatment	99	Others
8	For Attending Class		

(Source from SIS, University of Tsukuba: http://land.geo.tsukuba.ac.jp/persontrips/explanation.htm)

Data Structure

Attention on the difference of attributes code among datasets

Dhaka

Code	Purpose
1	To Home
2	To Work
3	To School
4	Shopping
5	Socializing
6	Recreation
7	Religious
8	Medical
9	Employment
10	Goods
11	Others
99	Unknown

Jakarta

Code	Purpose
1	To Work
2	To School
3	For Shopping
4	Business
5	Private
6	To Home
7	Others
9	Unknown
99	Unknown

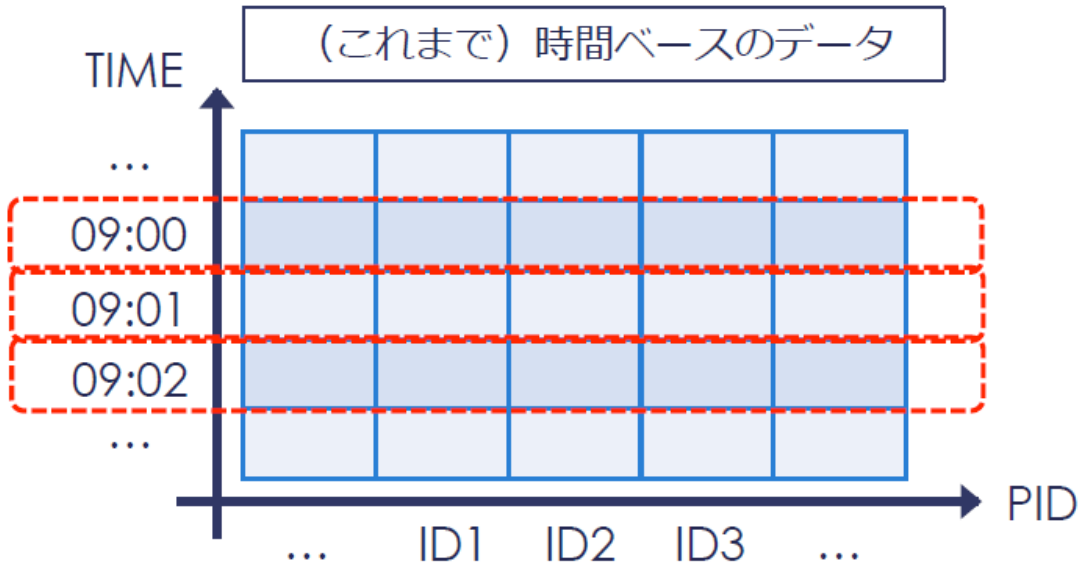
Tokyo 2008

Code	Value
1	To-From Office
2	To-From School
3	To Home
4	For Shopping
5	For Short Recreation
6	For Sight Seeing and Leisure
7	For Medical Treatment
8	For Attending Class
9	To Send/Pick Up Activity
10	For Selling and Buying
11	For Appointment
12	To/For Work (Fixing and Repairing)
13	To Agri./Forestry/Fishery Work
14	Other Business Purpose
99	Others

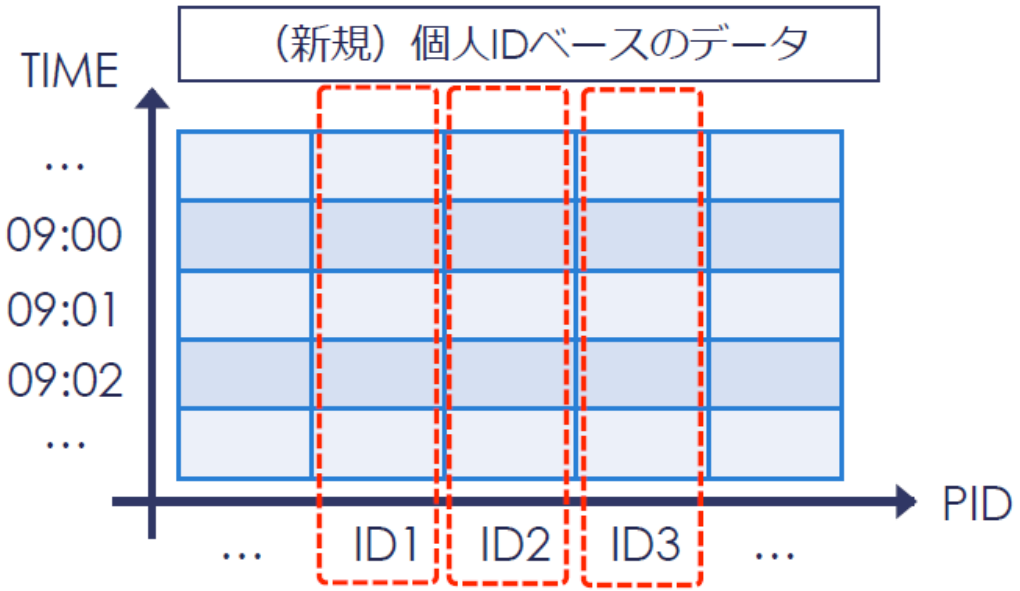
(Source from SIS, University of Tsukuba: http://land.geo.tsukuba.ac.jp/persontrips.asia/explanation.htm)

Format of Data

Time-based (old version)



PID-based (new version)



(Source from CSIS: http://pflow.csis.u-tokyo.ac.jp/wp-content/uploads/20140724.pflow-ws-tutorial.pdf)

Format of Data

Time-based (old version)

1	1	1	2008/10/1 0:00	139.7701	35.69787	1	4	123	4	99	83	0	97
2	1	1	2008/10/1 0:00	139.7759	35.69773	2	14	123	14	99	37	0	97
3	1	1	2008/10/1 0:00	139.7665	35.68351	1	11	103	8	99	62	0	97
4	1	1	2008/10/1 0:00	139.7814	35.69694	1	7	124	15	99	64	0	97
5	1	1	2008/10/1 0:00	139.7407	35.68188	1	5	110	7	99	83	0	97
6	1	1	2008/10/1 0:00	139.7398	35.68198	1	12	110	9	99	33	0	97
7	1	1	2008/10/1 0:00	139.7413	35.69229	2	8	112	4	99	45	0	97
8	1	1	2008/10/1 0:00	139.7271	35.69609	2	9	112	4	99	55	0	97
9	1	1	2008/10/1 0:00	139.746	35.69014	1	17	111	15	99	50	0	97
10	1	1	2008/10/1 0:00	139.7781	35.69381	2	6	124	15	99	47	0	97

PID-based (new version)

1	1	1	2008/10/1 0:00	139.7701	35.69787	1	4	123	4	99	83	0	97
1	1	1	2008/10/1 0:01	139.7701	35.69787	1	4	123	4	99	83	0	97
1	1	1	2008/10/1 0:02	139.7701	35.69787	1	4	123	4	99	83	0	97
1	1	1	2008/10/1 0:03	139.7701	35.69787	1	4	123	4	99	83	0	97
1	1	1	2008/10/1 0:04	139.7701	35.69787	1	4	123	4	99	83	0	97
1	1	1	2008/10/1 0:05	139.7701	35.69787	1	4	123	4	99	83	0	97
1	1	1	2008/10/1 0:06	139.7701	35.69787	1	4	123	4	99	83	0	97
1	1	1	2008/10/1 0:07	139.7701	35.69787	1	4	123	4	99	83	0	97
1	1	1	2008/10/1 0:08	139.7701	35.69787	1	4	123	4	99	83	0	97
1	1	1	2008/10/1 0:09	139.7701	35.69787	1	4	123	4	99	83	0	97

Introduction to the People Flow Data

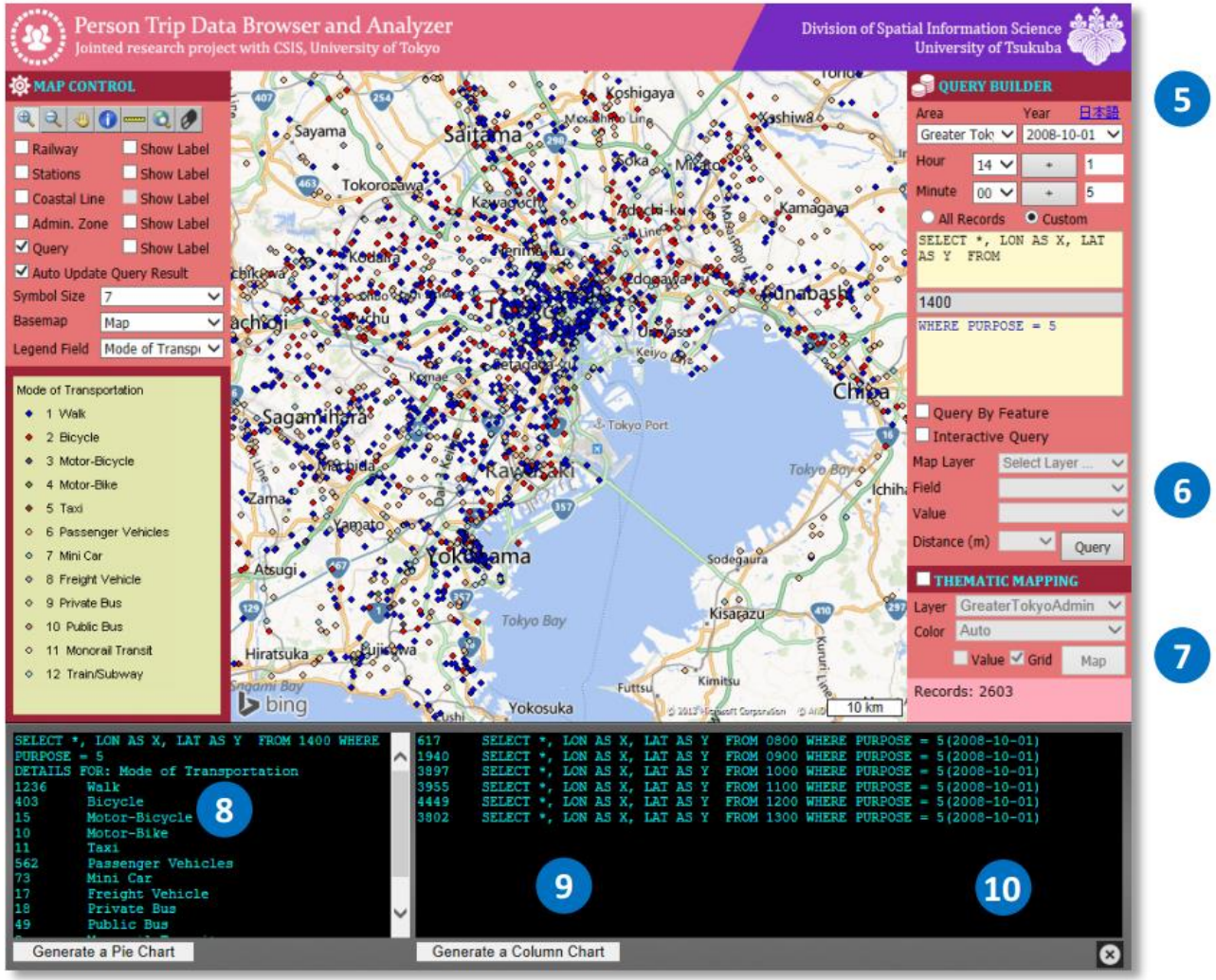
Provided by CSIS, University of Tokyo

How to use time-based people flow data

Visulization of time-based data

(Source from CSIS: https://www.youtube.com/watch?v=pR1g2S6OGUo)

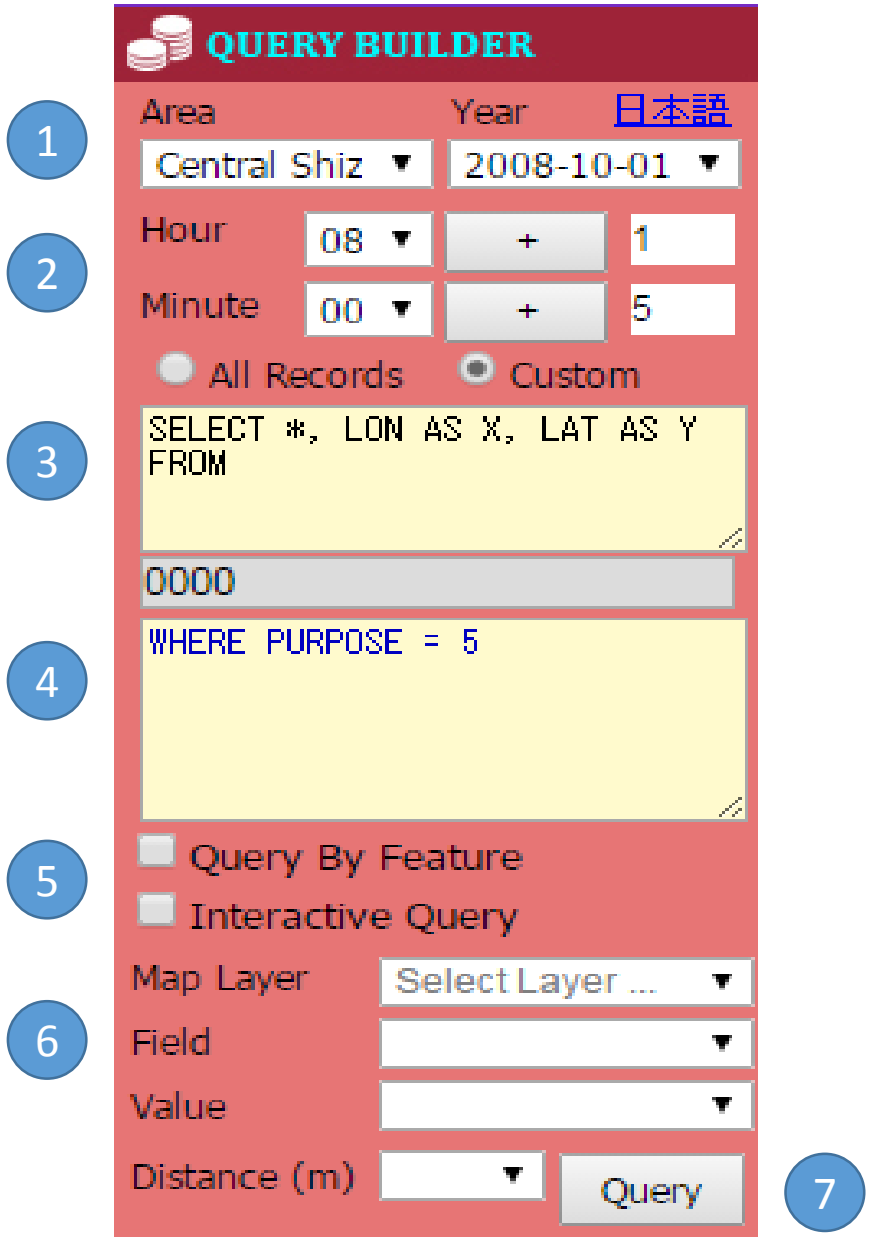
Person Trip Data Browser and Analyzer



1. Map controls (Zoom In/Out, Pan, Get Attribute Information, etc.)
2. Map layers control (Layer On-Off, Label On-Off)
3. Map symbol size
4. Show map legend and select attribute field for generating sum of each category
5. Query builder Spatial
6. Query by Feature/ Interactive Spatial Query
7. Thematic Mapping
8. Result of sum of each category in user defined attribute field
9. Result of each query string
10. Result of spatial query

Query Builder

1. Area and Year selection
2. Time selection
3. Column fields selection
4. Record selection
5. Feature Query
6. Interactive Query
7. Start Query



Available Datasets

Japan

Tokyo 1988/1998/2008

Northern Kyoshu 2005

Central Shizuoka 2001

International

Manila 1996

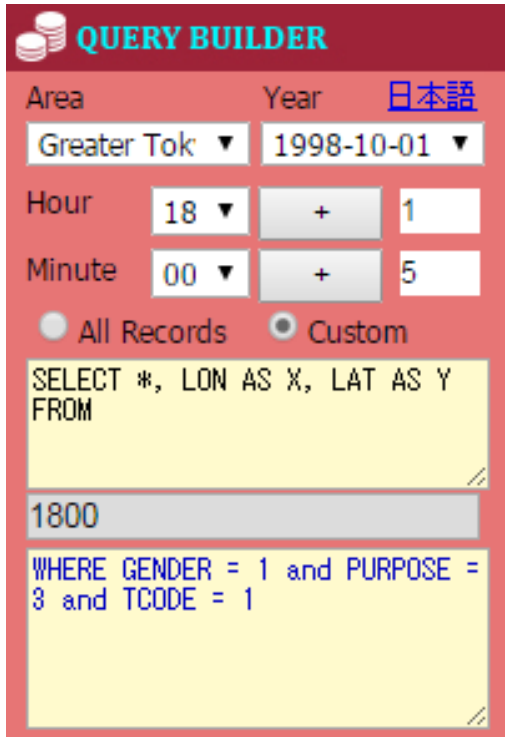
Jakarta 2000

Hanoi 2004

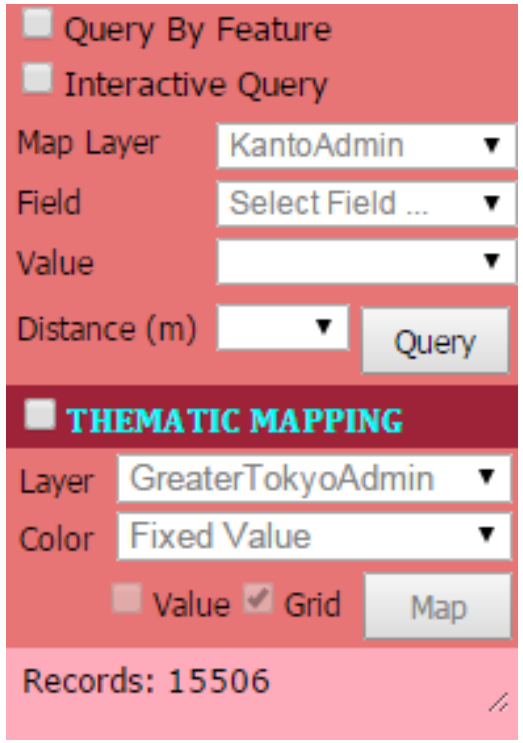
Dhaka 2009

Sample Query

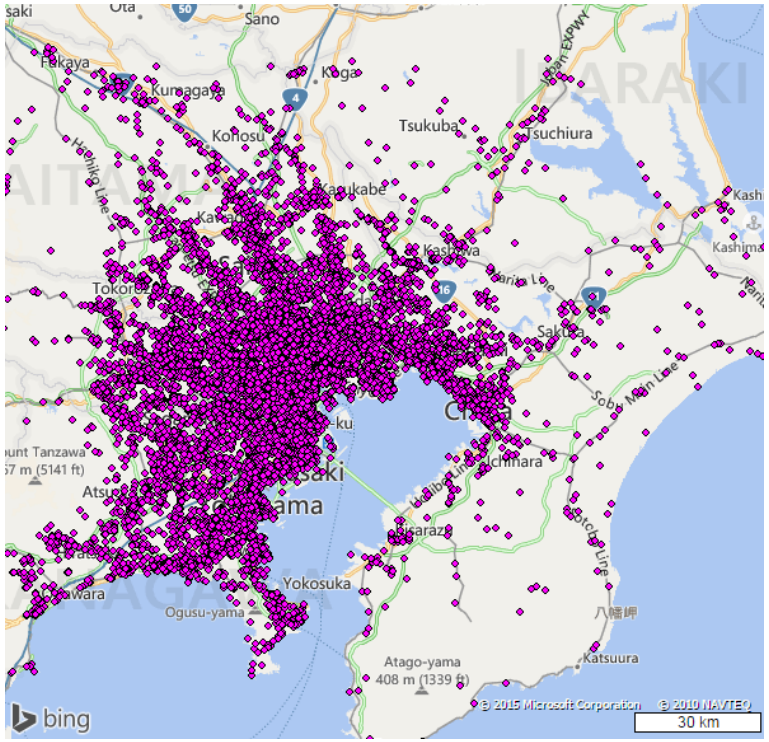
Set the query sentence



(Male & To Home & Walk)



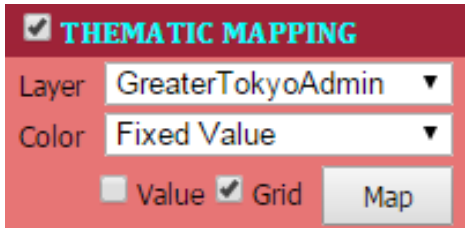
(Query)

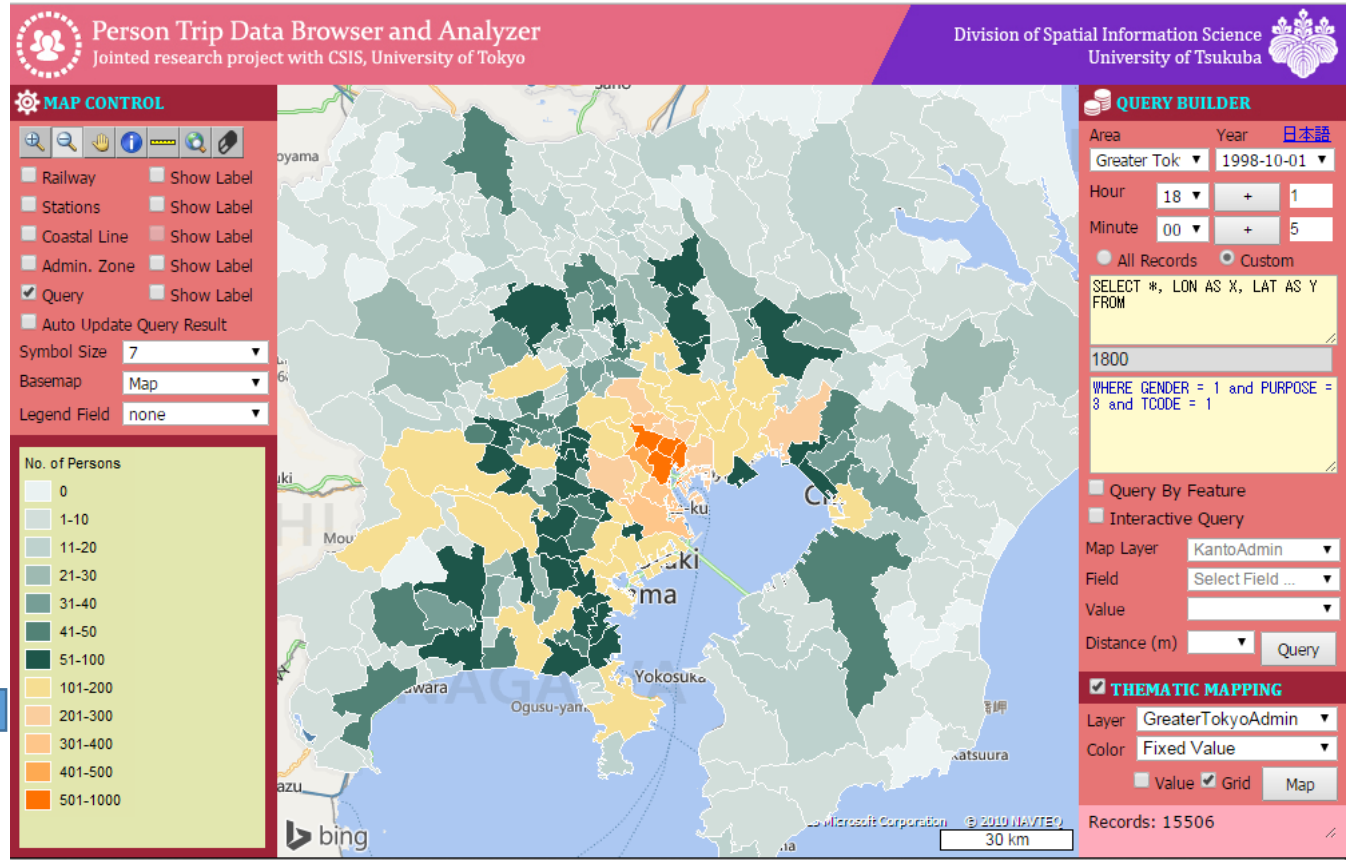


(Result)

Sample Query

Visulization (Thematic Mapping)

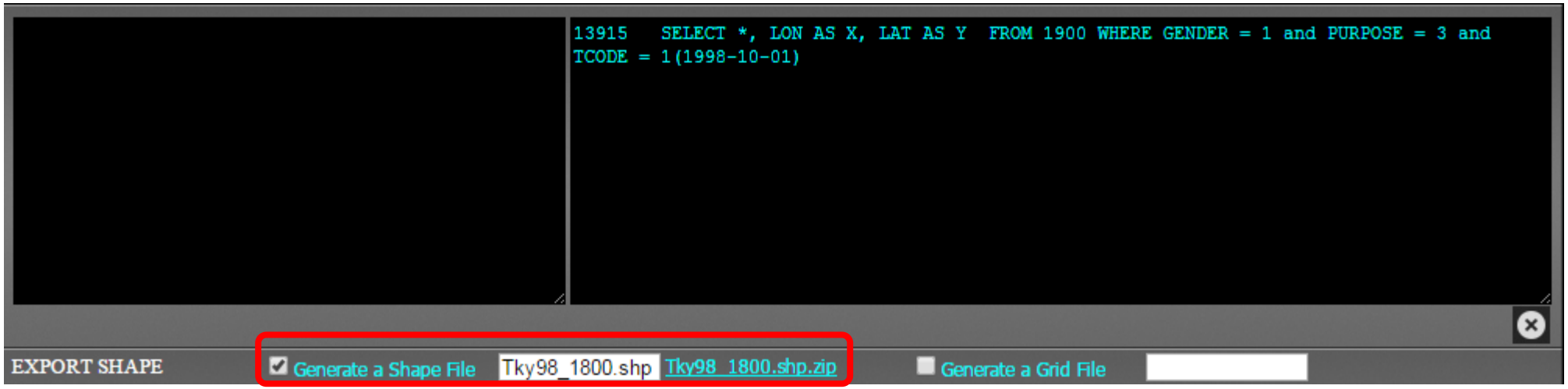


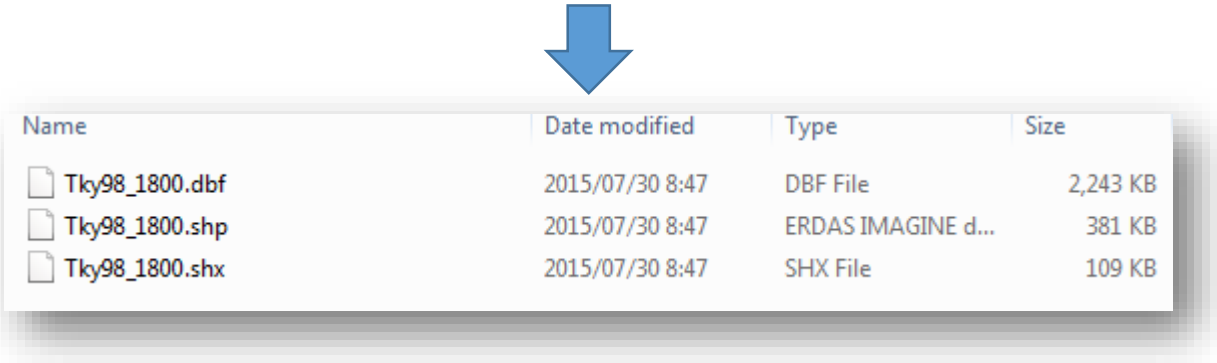



```
THEMATIC MAPPING (Summary By Administration Units)(1998-10-01)
-----------------------------------------------------------------
SELECT *, LON AS X, LAT AS Y  FROM 1800 WHERE GENDER = 1 and PURPOSE = 3 and TCODE = 1
-----------------------------------------------------------------
2       いすみ市
3       一宮町
136     中央区
0       九十九里町
28      佐倉市
50      八千代市
4       八街市
0       勝浦市
```

Sample Query

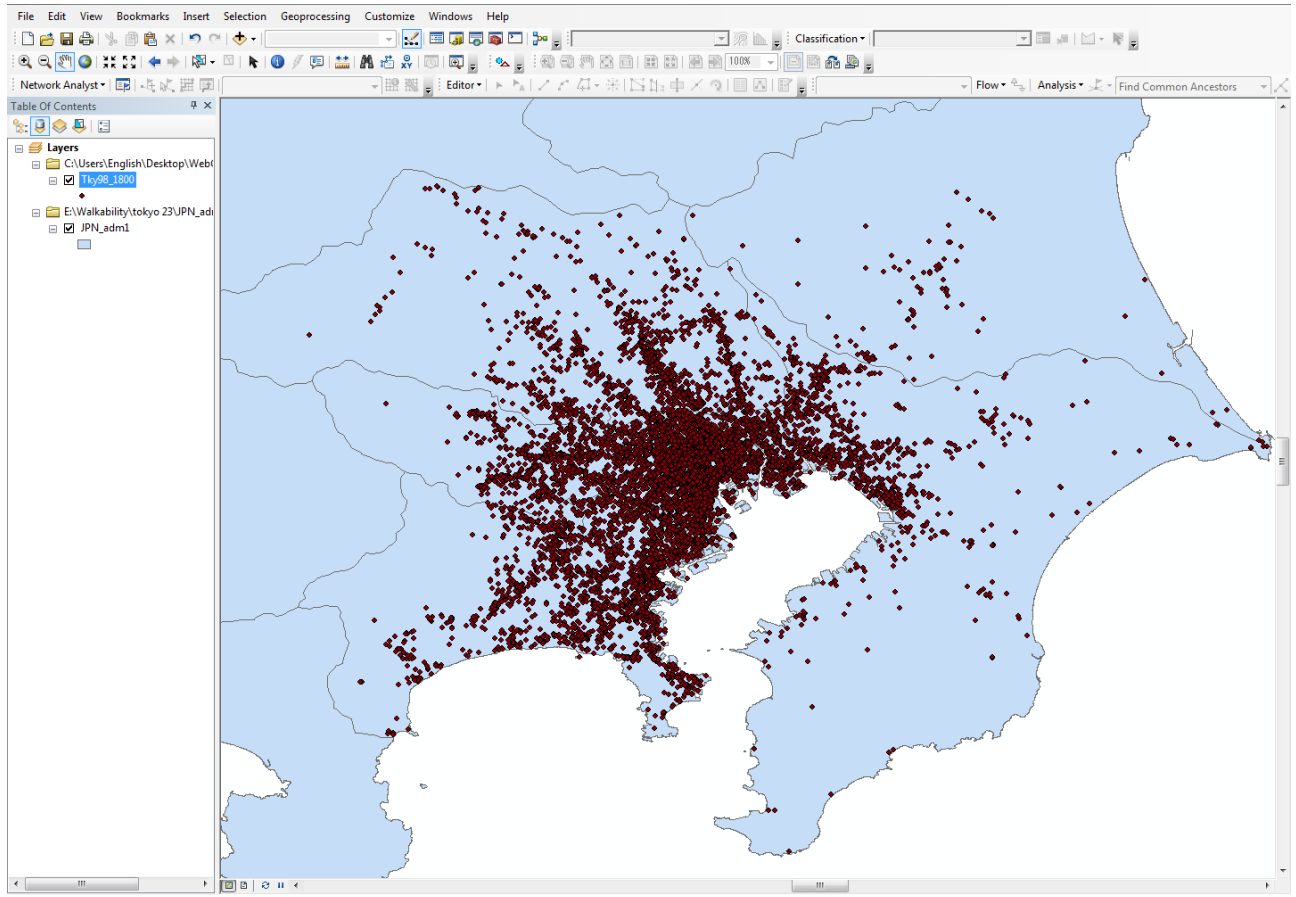
Generate a shapefile





Sample Query

Generated data in ArcGIS



Table

Tky98_1800

FID	Shape	X	Y	PID	TN	SN	TTIME	GENDER	AGE	ZCOD	OCC	PUR	TCODE
0	Point	139.579405	35.705396	37	4	1	1800	1	7	41100	8	3	1
1	Point	139.786401	35.697784	92	4	1	1800	1	5	41100	8	3	1
2	Point	139.731506	35.673319	151	6	1	1800	1	9	41100	9	3	1
3	Point	139.768066	35.675877	173	4	1	1800	1	11	41100	8	3	1
4	Point	139.942299	35.720059	238	6	3	1800	1	13	41100	9	3	1
5	Point	139.803501	35.668962	266	4	1	1800	1	8	41100	9	3	1
6	Point	140.122346	35.733927	268	4	1	1800	1	3	41100	13	3	1
7	Point	139.9616	35.728825	271	4	3	1800	1	6	41100	9	3	1
8	Point	139.946299	35.714236	311	4	3	1800	1	7	41100	6	3	1
9	Point	139.949898	35.724009	317	4	3	1800	1	4	41100	5	3	1
10	Point	139.943199	35.713926	357	4	3	1800	1	14	41100	9	3	1
11	Point	139.942276	35.717081	381	8	3	1800	1	14	41100	8	3	1
12	Point	139.747282	35.639364	511	4	1	1800	1	9	41101	9	3	1
13	Point	139.947273	35.727825	729	4	3	1800	1	9	41101	6	3	1
14	Point	140.00242	35.695713	769	4	1	1800	1	3	41101	13	3	1
15	Point	139.947569	35.715369	802	4	3	1800	1	11	41101	4	3	1
16	Point	140.165211	35.663181	829	10	1	1800	1	9	41101	9	3	1
17	Point	139.694133	35.693991	834	4	1	1800	1	7	41101	9	3	1
18	Point	139.729182	35.66087	1029	4	1	1800	1	5	41102	9	3	1
19	Point	139.645001	35.679308	1035	4	1	1800	1	5	41102	9	3	1
20	Point	140.033257	35.666266	1063	4	1	1800	1	4	41102	9	3	1
21	Point	139.934187	35.720726	1091	4	4	1800	1	3	41102	12	3	1
22	Point	139.793152	35.657797	1100	4	1	1800	1	6	41102	9	3	1
23	Point	139.92575	35.722209	1144	4	3	1800	1	4	41102	13	3	1
24	Point	139.92575	35.722209	1164	4	3	1800	1	8	41102	4	3	1
25	Point	139.896679	35.69217	1210	4	1	1800	1	3	41102	13	3	1
26	Point	139.928641	35.723651	1216	4	3	1800	1	14	41102	10	3	1
27	Point	139.929566	35.727404	1247	4	3	1800	1	6	41103	10	3	1
28	Point	139.925911	35.727384	1256	10	5	1800	1	11	41103	10	3	1
29	Point	139.929299	35.725896	1311	4	3	1800	1	13	41103	10	3	1
30	Point	139.927563	35.721017	1345	4	4	1800	1	16	41103	9	3	1
31	Point	139.927563	35.721017	1358	4	3	1800	1	9	41103	9	3	1
32	Point	139.921905	35.726906	1406	4	3	1800	1	4	41103	3	3	1
33	Point	139.738472	35.596385	1501	4	1	1800	1	4	41103	9	3	1
34	Point	139.908115	35.728431	1509	4	3	1800	1	6	41103	9	3	1

(0 out of 13915 Selected)

Tky98_1800

Introduction to the People Flow Data

Provided by CSIS, University of Tokyo

How to use ID-based people flow data

ID-based Data Transformation

Original

PID	time	Lon	Lat	transport
123	00:00	Lon0	Lat0	Stay
123	00:01	Lon0	Lat0	Stay
123	…	…	…	…
123	08:10	Lon0	Lat0	Stay
123	08:10	Lon0	Lat0	Move
123	08:11	Lon1	Lat1	Move
123	…	…	…	…
123	08:30	LonN	LatN	Move
123	08:30	LonN	LatN	Stay
123	08:31	LonN	LatN	Stay
123	…	…	…	…

Transformed

PID	Start_time	End_time	Lon	Lat	transport
123	00:00	08:10	Lon0	Lat0	Stay
123	08:10	08:11	Lon0	Lat0	Move
123	08:11	08:12	Lon1	Lat1	Move
123	…	…	…	…	…
123	08:30	08:30	LonN	LatN	Move

(Source from CSIS: http://pflow.csis.u-tokyo.ac.jp/wp-content/uploads/20140724.pflow-ws-tutorial.pdf)

Java Environment

http://java.com/en/download/uninstallapplet.jsp

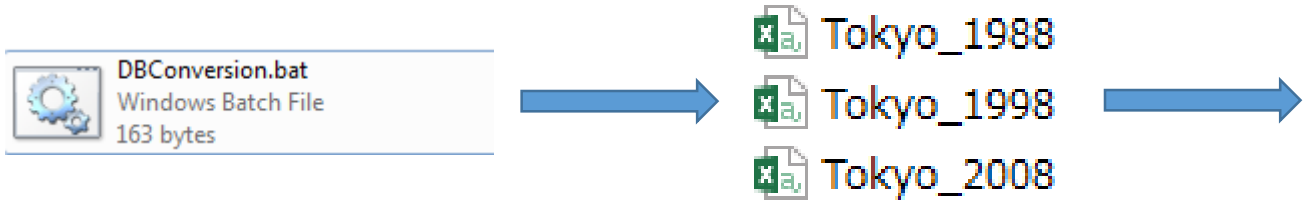
Check version higher than 1.7

http://www.java.com/en/download/ie_manual.jsp

Data Preperation (Compress)

Run the Java program

【**Windows**】⇒ DBConversion.bat
【**Mac**】⇒ DBConversion.sh



PID	TID	STID	Start_time	End_time	Longitude	Latitude
1	1	1	10/1/2008 0:00	10/1/2008 10:00	139.7701	35.69787
1	2	1	10/1/2008 10:00	10/1/2008 10:01	139.7701	35.69814
1	2	1	10/1/2008 10:01	10/1/2008 10:02	139.7694	35.69853
1	2	1	10/1/2008 10:02	10/1/2008 10:03	139.7687	35.69896
1	2	1	10/1/2008 10:03	10/1/2008 10:04	139.7679	35.69933
1	2	1	10/1/2008 10:04	10/1/2008 10:05	139.7672	35.69967
1	2	1	10/1/2008 10:05	10/1/2008 10:06	139.7663	35.69993
1	2	1	10/1/2008 10:06	10/1/2008 10:07	139.7655	35.70016
1	2	1	10/1/2008 10:07	10/1/2008 10:08	139.7647	35.7004
1	2	1	10/1/2008 10:08	10/1/2008 10:08	139.7638	35.70064
1	2	2	10/1/2008 10:08	10/1/2008 10:09	139.7646	35.6999
1	2	2	10/1/2008 10:09	10/1/2008 10:10	139.7652	35.6997
1	2	2	10/1/2008 10:10	10/1/2008 10:11	139.7657	35.6995
1	2	2	10/1/2008 10:11	10/1/2008 10:12	139.7663	35.69929
1	2	2	10/1/2008 10:12	10/1/2008 10:13	139.7669	35.69909

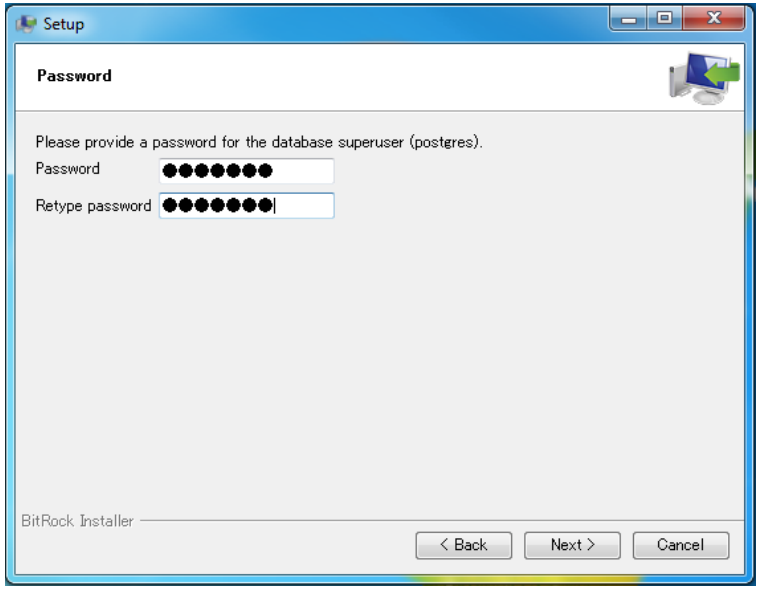
Create new csv file **(DB_Import_File.csv)**

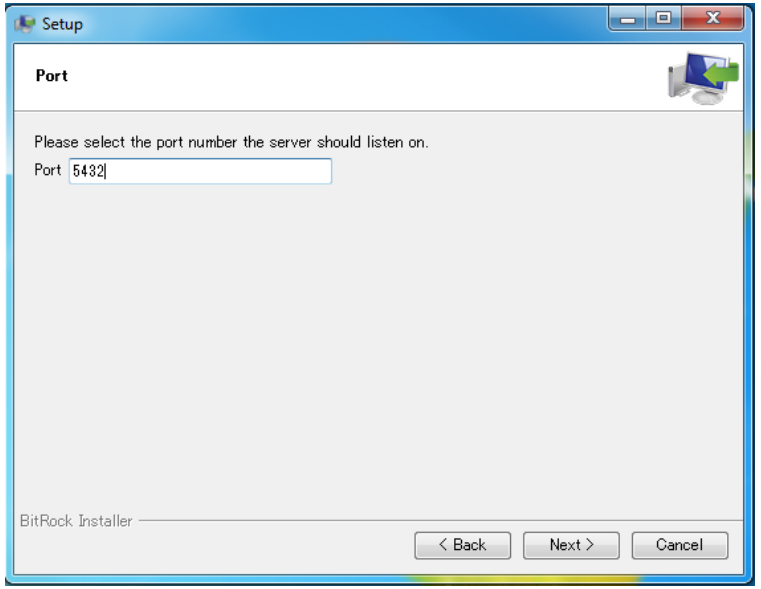
600,000 ID needs **4 ~ 5** hours for compression (differ in each computer)

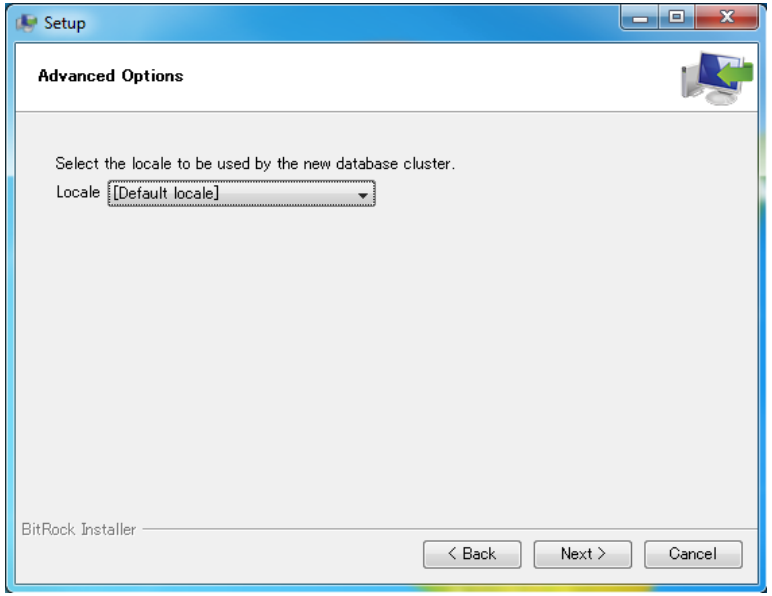
Take Tokyo 1988 dataset as an example, data size decreased from **68.1 GB** to **4.65 GB**

Use PostgreSQL to Manage Data

PostgreSQL（http://www.postgresql.org/）

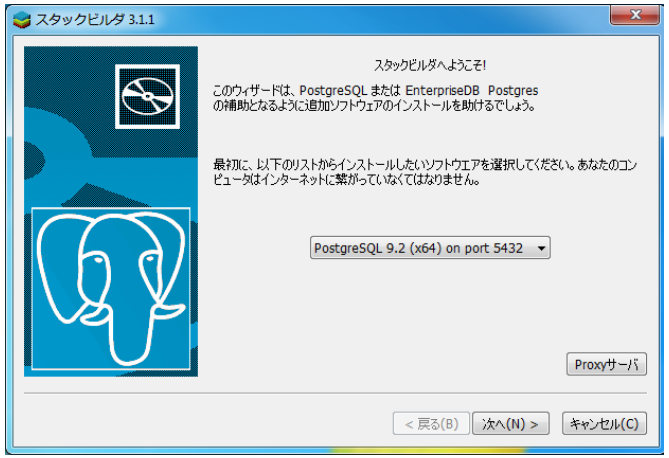




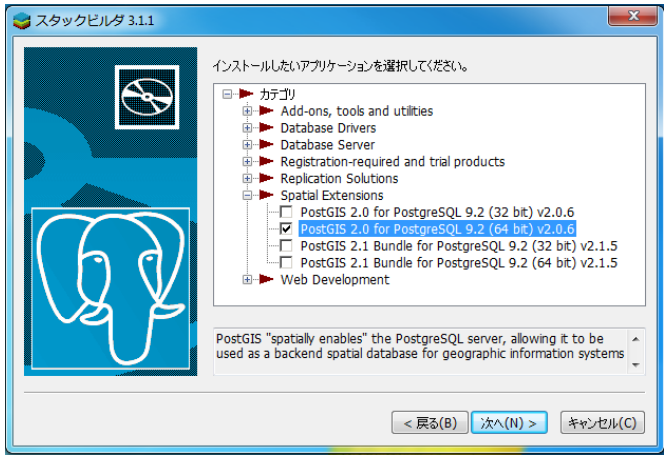


PostGIS（http://postgis.net/ ）

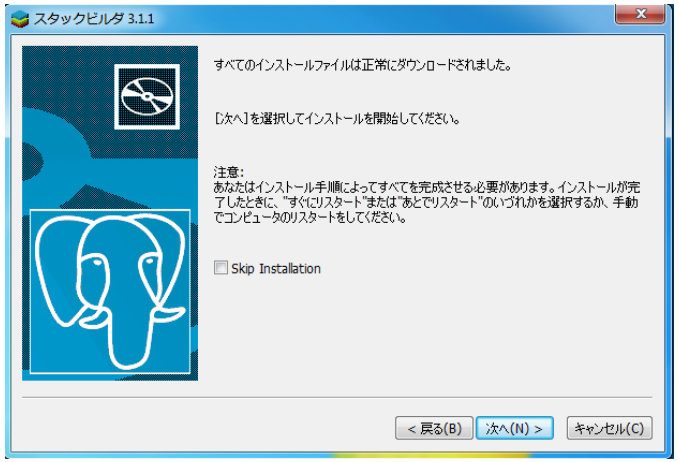
1



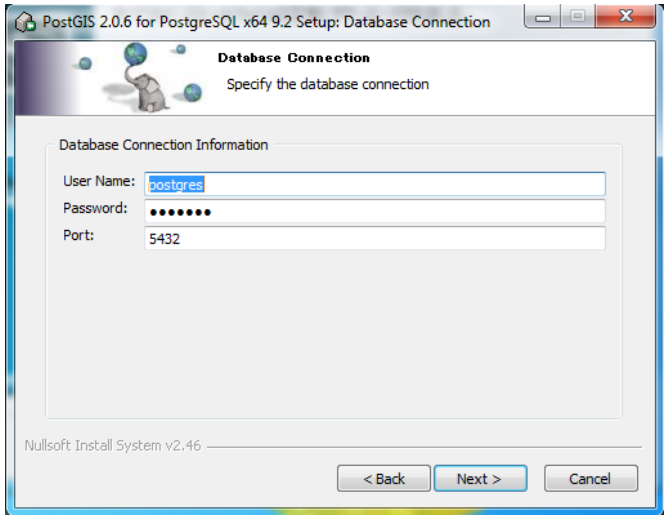
2



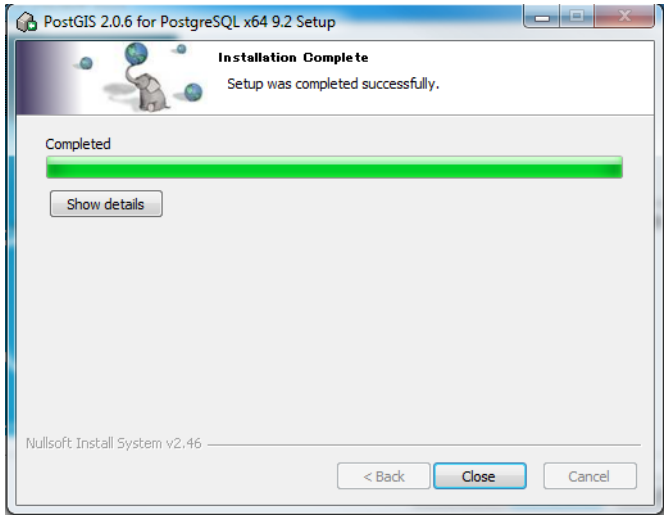
3



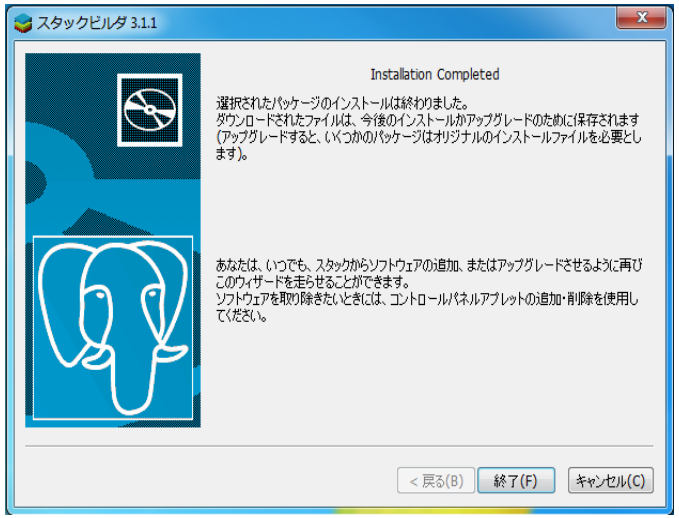
4



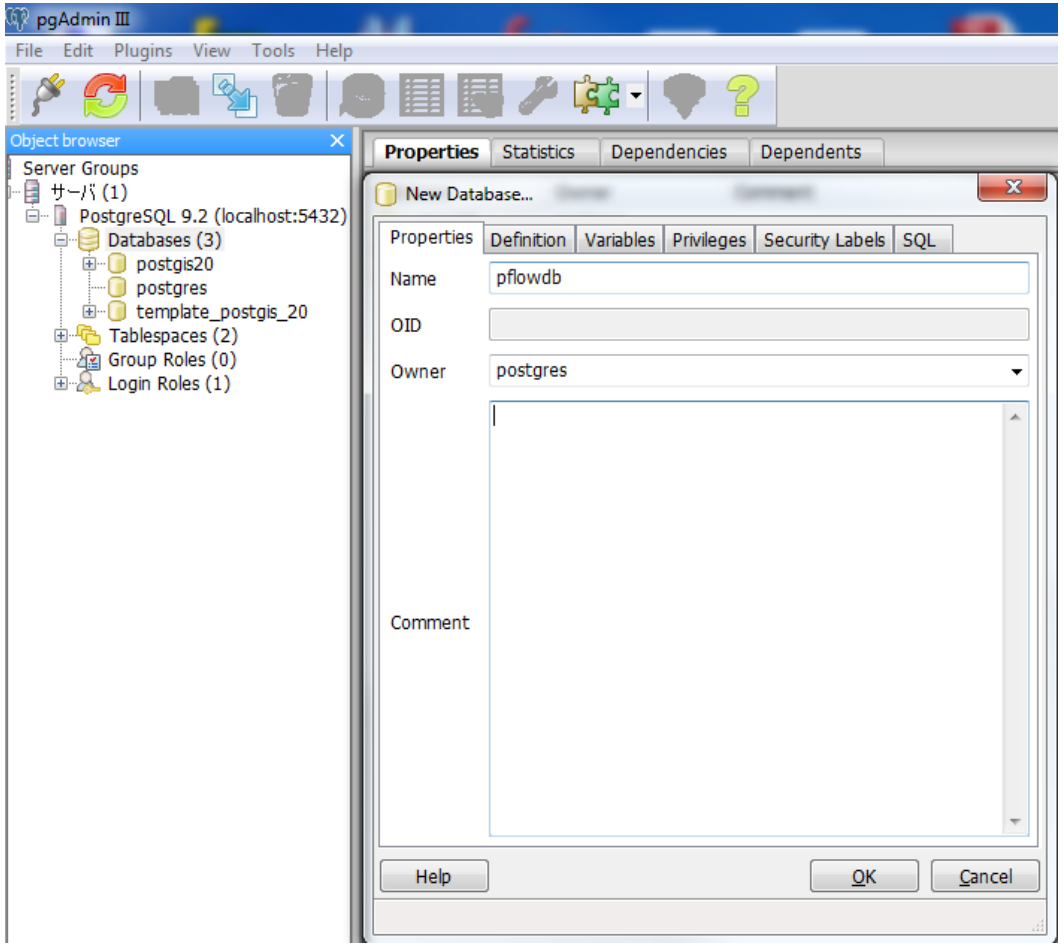
5

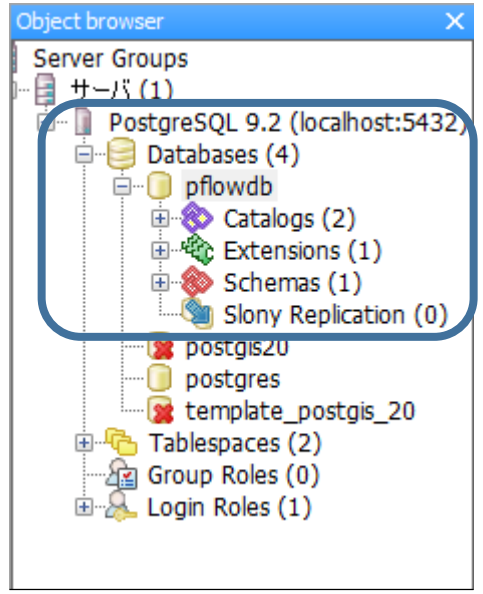


6



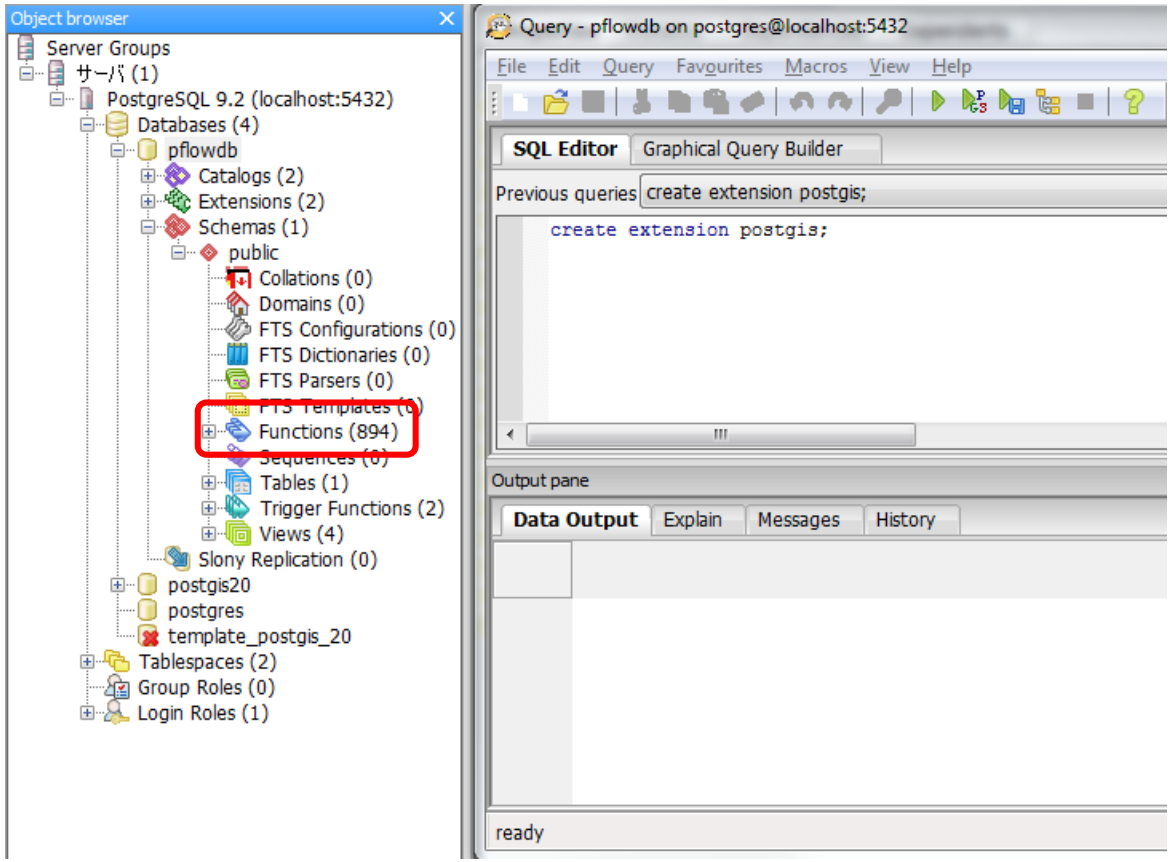
Build Database



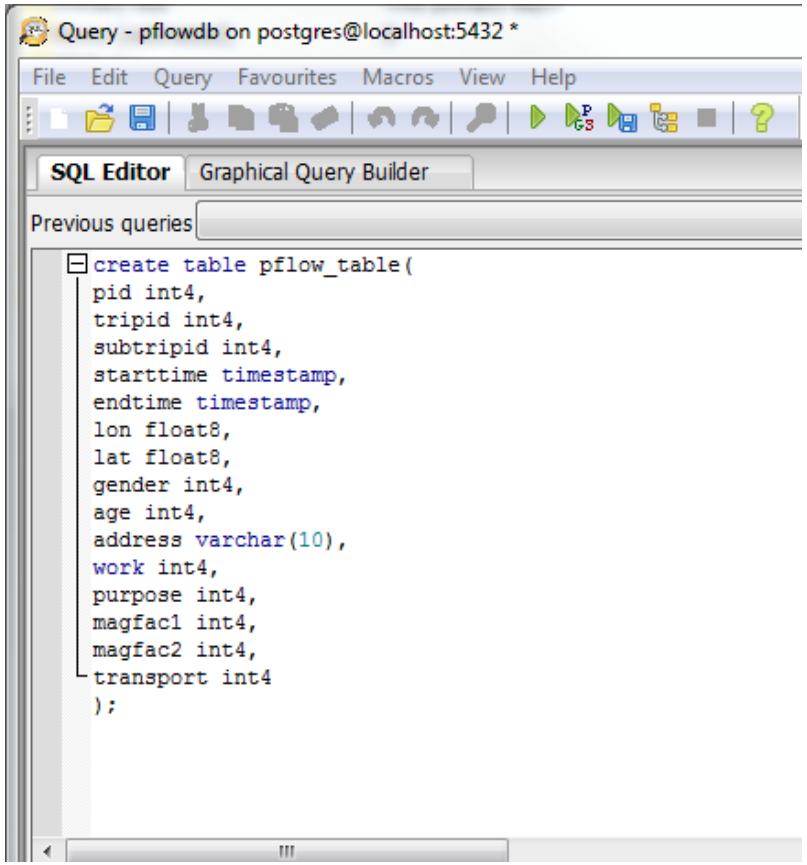


Import Data into PostgreSQL

Add PostGIS Extension

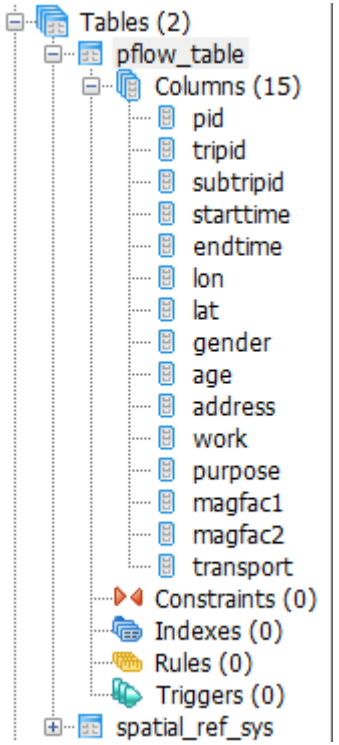


Create Vacant Table



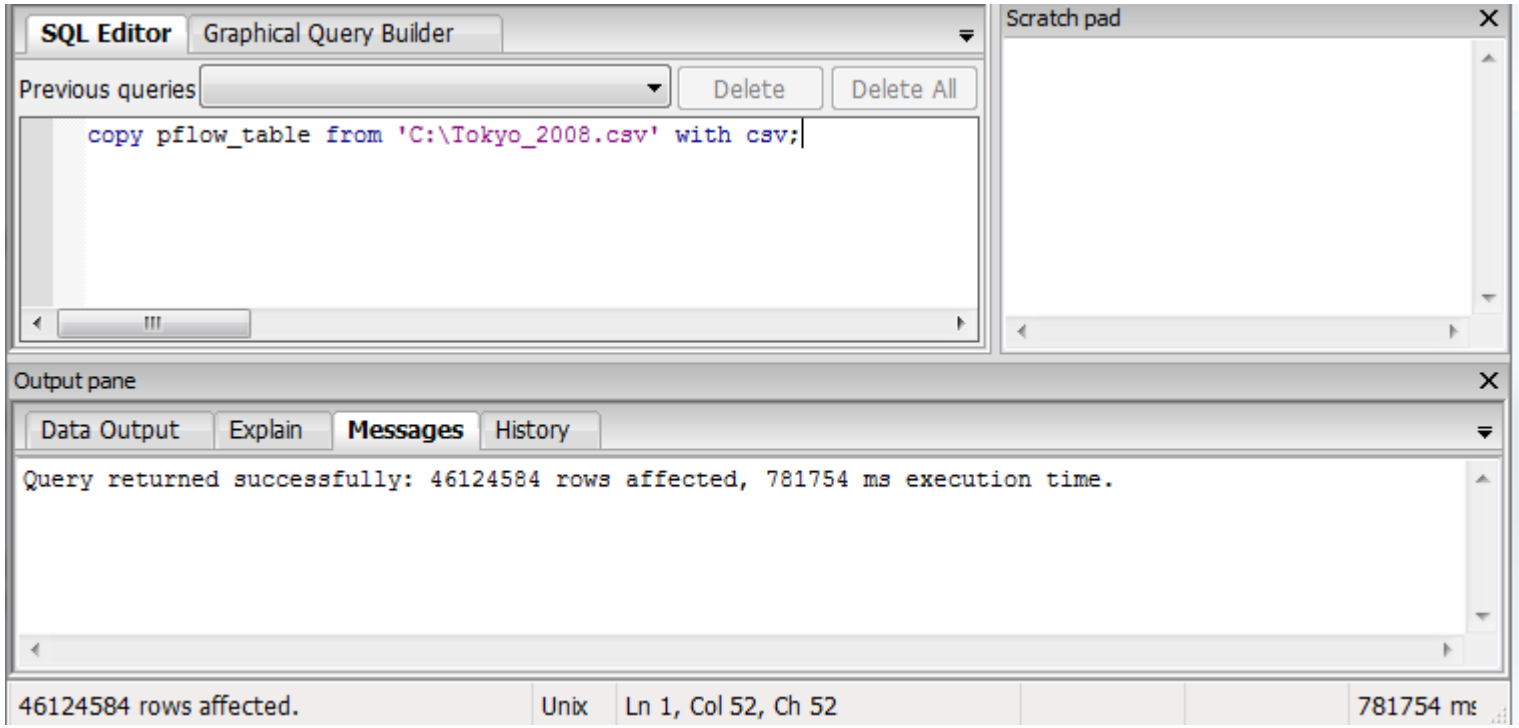
Import Data into PostgreSQL

New pflow table



Fill in table

copy pflow_table from '**C:\Tokyo_2008.csv**' with csv;



Sample Query

Rocord all the moving people at 10:00 a.m.

SQL Editor | Graphical Query Builder

Previous queries | Delete | Delete All

```
select pid,starttime,purpose,transport,lon,lat
from pflow_table
where starttime <= '2008-10-01 10:00:00'
and '2008-10-01 10:00:00' < endtime
and transport<>97
order by pid;
```

Output pane

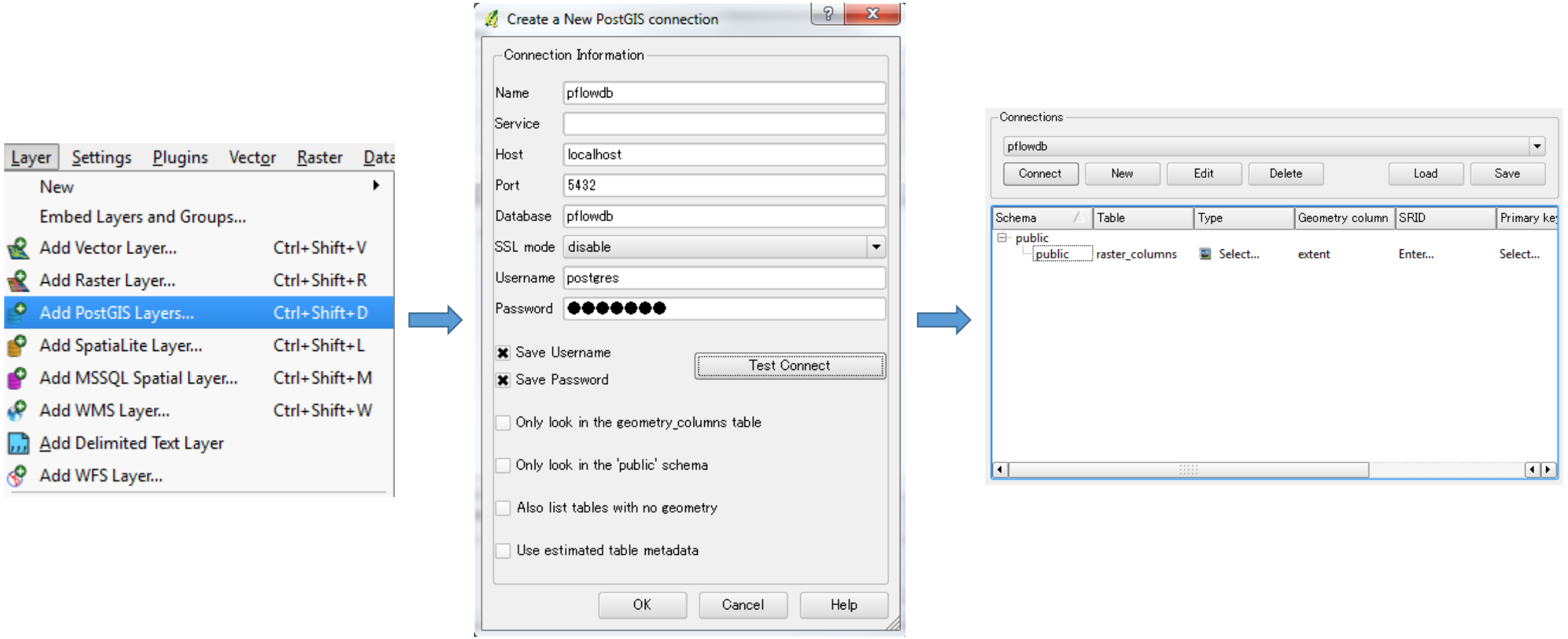
Data Output | Explain | Messages | History

	pid integer	starttime timestamp without time zone	purpose integer	transport integer	lon double precision	lat double precision
1	1	2008-10-01 10:00:00	11	1	139.770139	35.698142
2	27	2008-10-01 10:00:00	1	1	139.74033	35.683243
3	36	2008-10-01 10:00:00	1	11	139.789733	35.638964
4	43	2008-10-01 10:00:00	3	1	139.746404	35.693567
5	62	2008-10-01 10:00:00	3	2	139.753903	35.701716
6	76	2008-10-01 10:00:00	7	1	139.734643	35.689084
7	77	2008-10-01 10:00:00	7	1	139.742442	35.6904
8	88	2008-10-01 10:00:00	8	1	139.745716	35.698583
9	94	2008-10-01 10:00:00	11	1	139.692921	35.68755
10	108	2008-10-01 10:00:00	1	1	139.741855	35.688084
11	133	2008-10-01 10:00:00	1	12	139.69695	35.701463
12	157	2008-10-01 10:00:00	3	1	139.745329	35.693958
13	174	2008-10-01 10:00:00	1	1	139.790088	35.706524
14	177	2008-10-01 10:00:00	5	12	139.672303	35.743617
15	199	2008-10-01 10:00:00	1	1	139.772402	35.695609
16	204	2008-10-01 10:00:00	3	2	139.771502	35.692371
17	223	2008-10-01 10:00:00	14	1	139.737636	35.691529
18	229	2008-10-01 10:00:00	1	4	139.74038	35.688784
19	233	2008-10-01 10:00:00	1	1	139.752341	35.699833
20	254	2008-10-01 10:00:00	14	1	139.77179	35.689493
21	256	2008-10-01 10:00:00	3	1	139.783777	35.689051
22	259	2008-10-01 10:00:00	4	6	139.708542	35.66106
23	263	2008-10-01 10:00:00	8	2	139.74018	35.686576
24	279	2008-10-01 10:00:00	4	12	139.76025	35.692165

OK. | Unix | Ln 6, Col 14, Ch 176 | 36991 rows. | 39247 ms

Connect Database From QGIS

Layer -> Add PostGIS Layers…



Extract Data by Query Layer

Set the query sentence