Data Science Training Webinars
Spatiotemporal Simulation

LINEAR PROGRAMMING AND APPLICATIONS IN EXAMINING WASTEFUL COMMUTING AND ALLOCATING HEALTHCARE PROVIDERS

Lingjie Hu

Wuhan University

July 2, 2022

Center for Geographic Analysis

Harvard University

Geo-computation Center for Social Sciences

Wuhan University

China Data Institute

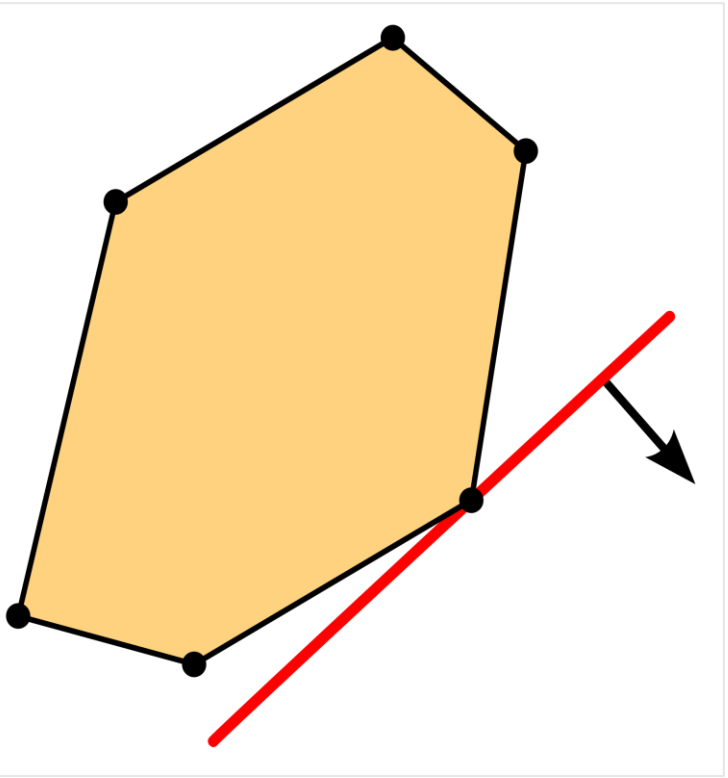
Future Data Lab

Contents

Linear Programming	线性规划
Examining Wasteful Commuting	浪费性通勤测算
Integer Programming	整数规划
Location-Allocation Problems	区位优化问题
P-median Model	p中位模型
Case study	案例分析
Flow Chart	流程图
Workflow Implementation	KNIME工作流
The Outputs from the Workflow	KNIME工作流输出
Step by Step Workflow Execution	KNIME逐步执行

Introduction

Linear programming is an important optimization technique in socioeconomic analysis and planning.



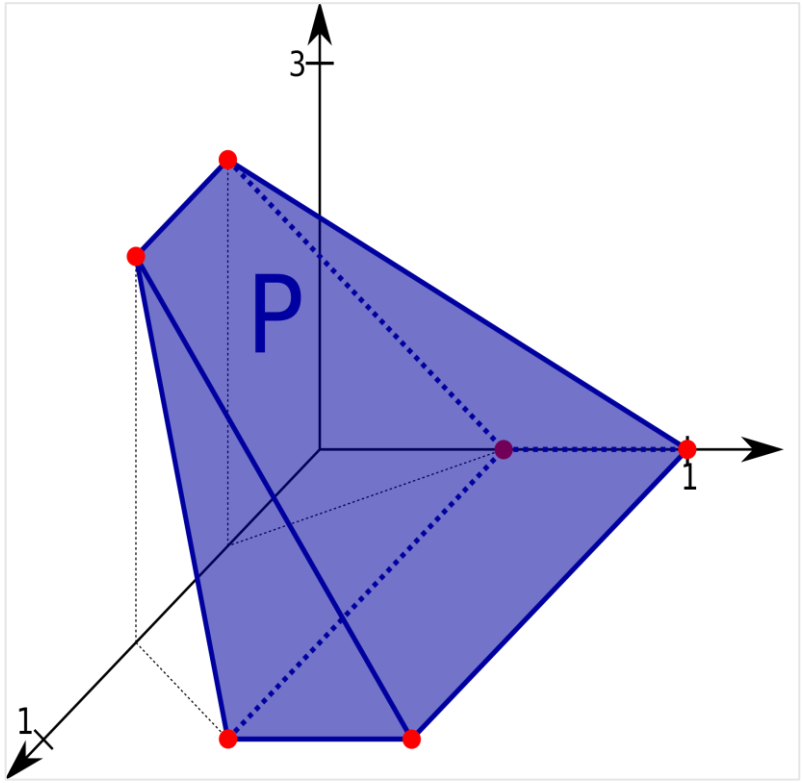


Image Source https://en.wikipedia.org/wiki/Linear_programming

It seeks to maximize or minimize an objective function subject to a set of constraints. Linear programming methods are often used to **examine the issue of wasteful commuting** and **solve facility location and allocation problem**.

Linear Programming

A **linear programming(LP)** function is in its standard form if it seeks to maximize the objective function

$$z = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$$

subject to,

$$a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n \leq b_1$$

$$a_{21} x_1 + a_{22} x_2 + \cdots + a_{2n} x_n \leq b_2$$

$$\vdots$$

$$a_{m1} x_1 + a_{m2} x_2 + \cdots + a_{mn} x_n \leq b_m$$

where,

$$x_i \geq 0, b_i \geq 0$$

Wasteful Commuting

Concept

The issue of **wasteful commuting** was first raised by Hamilton from the perspective of urban planning. Assuming that residents can freely swap houses, the planning problem is to minimize total commuting given the locations of houses and jobs.

Methods

- Hamilton monocentric city model
- **White's linear programming based method**

Development

- Excess commuting
- Relationship between commuting and urban land use
- Explanation of intra-city commuting differences

Examining Wasteful Commuting

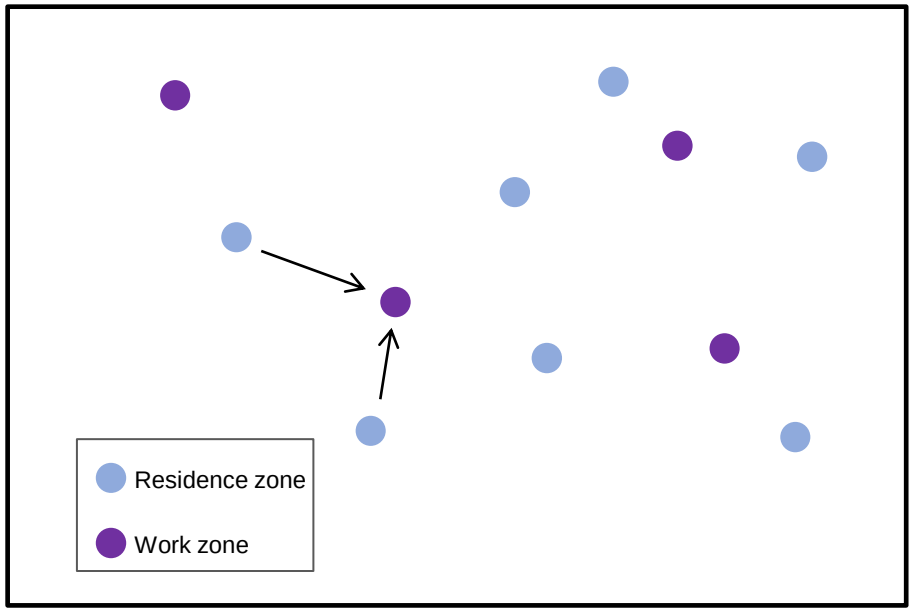
Given the number of resident workers P_i *at* *i*(i = 1, 2, …, n), and the number of jobs at E_j at j (j = 1, 2, …, m), the minimum commute is the solution to the following linear programming problem:

Minimize: $\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}x_{ij}$

Subject to: $\sum_{j=1}^{m} x_{ij} \le P_i$ for all $i(= 1,2,\cdots,n)$

$\sum_{i=1}^{n} x_{ij} \le E_i$ for all $j(= 1,2,\cdots,m)$

$x_{ij} > 0$ for all $i(= 1,2,\cdots,n)$ and all $j(= 1,2,\cdots,m)$



Where c_{ij} is the commute distance (time) from residential location *i* to job site *j*, and x_{ij} is the number of commuters on that route.

Integer Programming

If some of the decision variables in a linear programming problem are restricted to only integer values, the problem is referred to as an integer programming problem (ILP). Similar to the LP standard form, a linear programming problem is written as

Maximize: $\sum_{j=1}^{n} c_j x_j$

Subject to: $\sum_{j=1}^{n} a_{ij} x_j \leq b_i$ for all $i(= 1,2,\cdots,m)$ and

integers $x_j \geq 0$ for all $j(= 1,2,\cdots,n)$.

Solving the ILP requires special approaches such as the cutting planes method or the branch and bound method, and the latter is more popular.

Location-Allocation Problems

Location-allocation is to find a set of facilities that will best serve demand from surrounding areas.

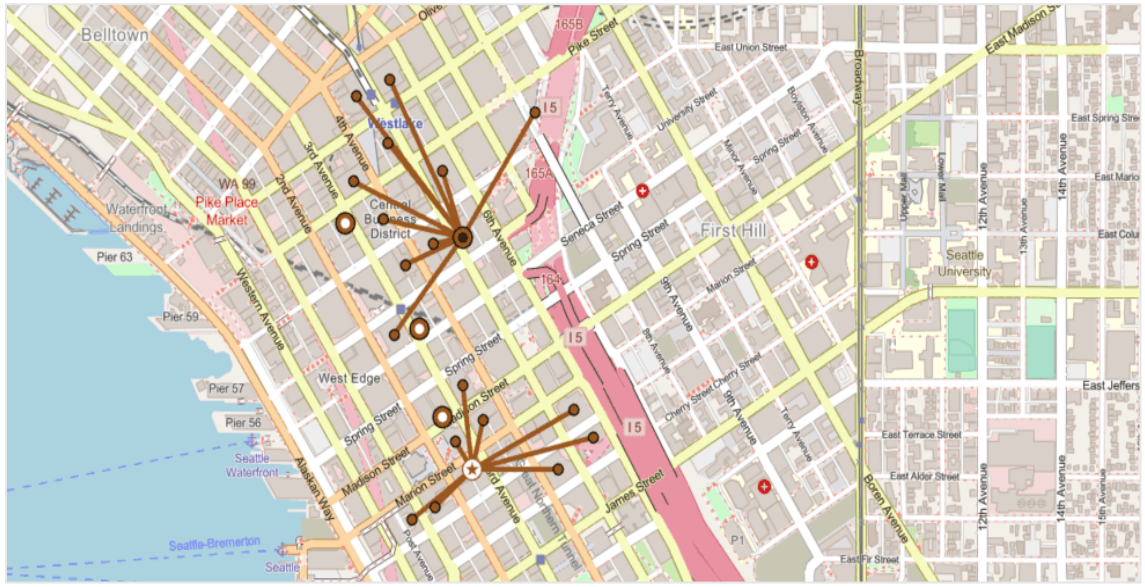


Image Source https://developers.arcgis.com/documentation/mapping-apis-and-services/routing/location-allocation/

- **Healthcare providers**
- Fire stations
- …

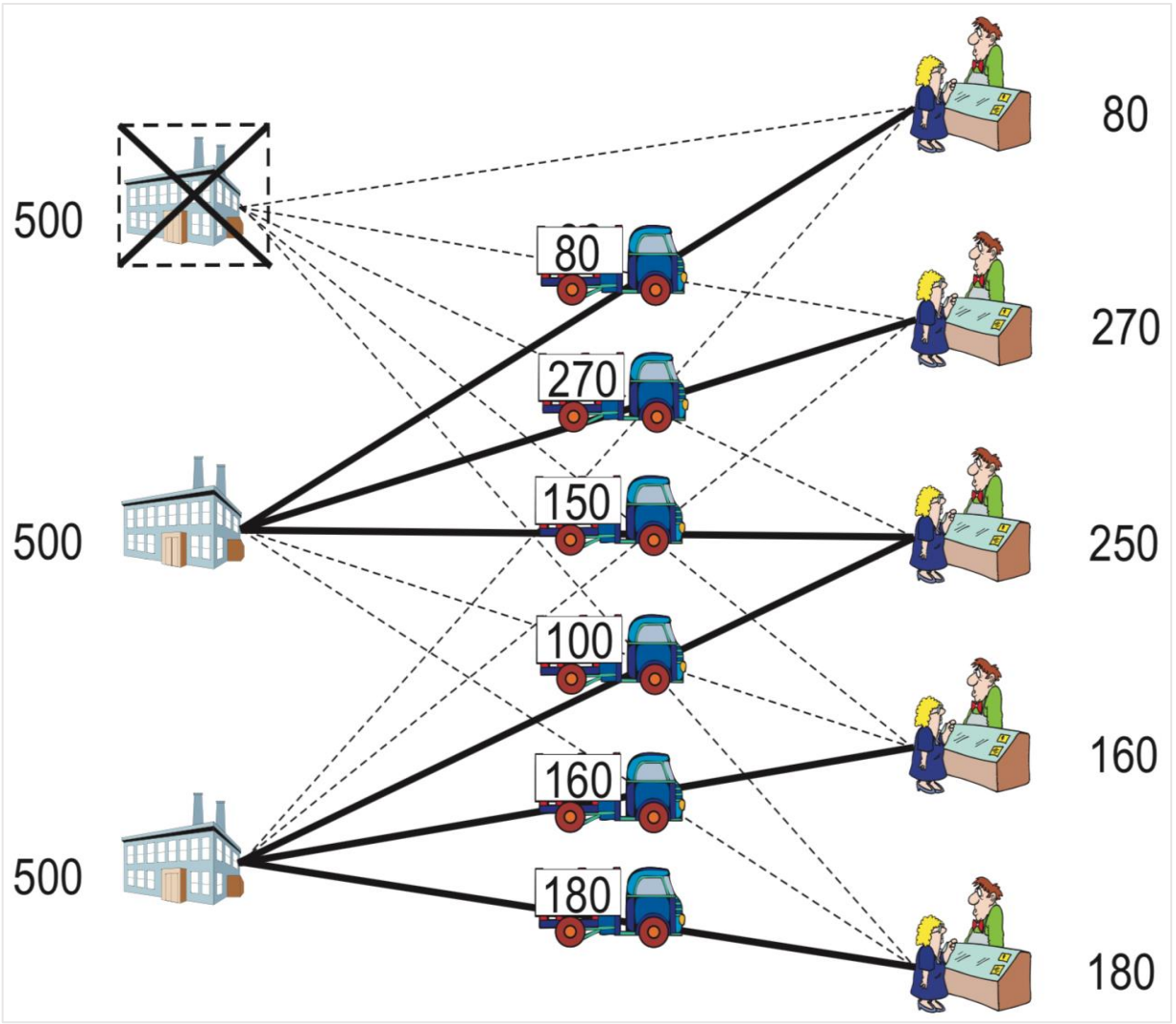


Image Source https://scipbook.readthedocs.io/en/latest/flp.html

Location-Allocation Problems

Three classic location-allocation problems can be used to illustrate the formulation of ILP problems.

Location-Allocation Model	Objective	Constraints	Problem type in ArcGIS location-allocation analysis
p-median problem	Minimize total distance (time)	Locate p facilities; all demands are covered	Minimize Impedance
location set covering problem (LSCP)	Minimize the number of facilities	All demands are covered	Minimize Facilities
maximum covering location problem (MCLP)	Maximize coverage	Locate p facilities; demand is covered if it is within a specified distance (time) of a facility	Maximize Capacitated Coverage (by setting an impedance cutoff)

P-median Model

The p-median model formulation is

Minimize: $Z = \sum\sum a_i d_{ij} x_{ij}$

Subject to: $x_{ij} \leq x_{jj} \quad \forall i, j, i \neq j$ (each demand assignment is restricted to what has been located)

$\sum_{j=1}^{m} x_{ij} = 1 \quad \forall i$ (each demand must be assigned to a facility)

$\sum_{j=1}^{m} x_{jj} = p \quad \forall j$ (exactly p facilities are located)

$x_{ij} = 1,0 \quad \forall i, j$ (a demand area is either assigned or not to a facility)

Where :

i indexes demand areas $(i = 1, 2, \ldots, n)$

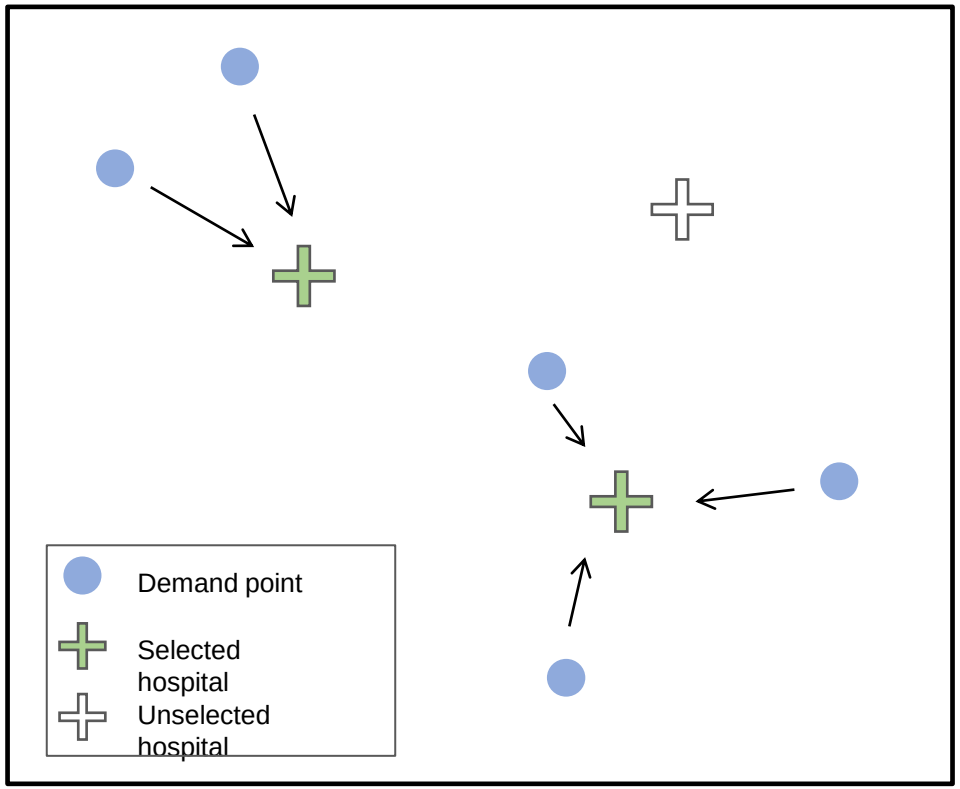
j indexes candidate facility sites $(j = 1, 2, \ldots, m)$

p is the number of facilities to be located

a_i is the amount of demand at area i

d_{ij} is the distance or time between demand i and facility j

x_{ij} is 1 if demand i is assigned to facility j or 0 otherwise.



LSCP Model

The location set covering problem formulation is

Minimize: $Z = \sum_{j=1}^{m} x_j$

Subject to: $\sum_{j=1}^{N_i} x_j \geq 1 \quad \forall i$ (a demand area must be within the critical distance or time of at least one open facility site)

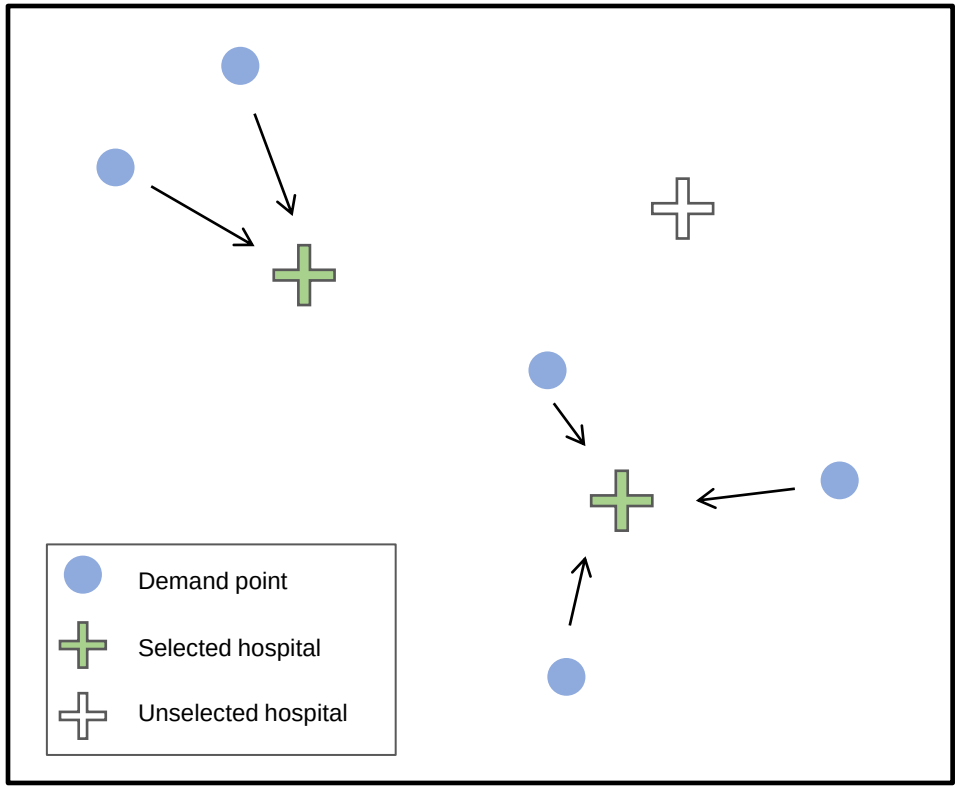
$x_j = 1,0 \quad \forall j$ (a candidate facility is either open or closed)

Where :

N_i is the set of facilities where the distance or time between demand i and facility j is less than the critical distance or time s, i.e.,$d_{ij} \leq d_0$

x_j is 1 if a facility is open at candidate site j or 0 otherwise

i, j, m and n are the same as in the above p-median model formulation.



MCLP Model

The maximum covering location problem formulation is

Minimize: $Z = \sum_{i=1}^{m} a_i y_i$

Subject to: $\sum_{j=1}^{N_i} x_j + y_i \geq 1 \quad \forall i$ (a demand area must be within the critical distance or time of at least one open facility site or it is not covered)

$\sum_{j=1}^{m} x_j = p$ (exactly p facilities are located)

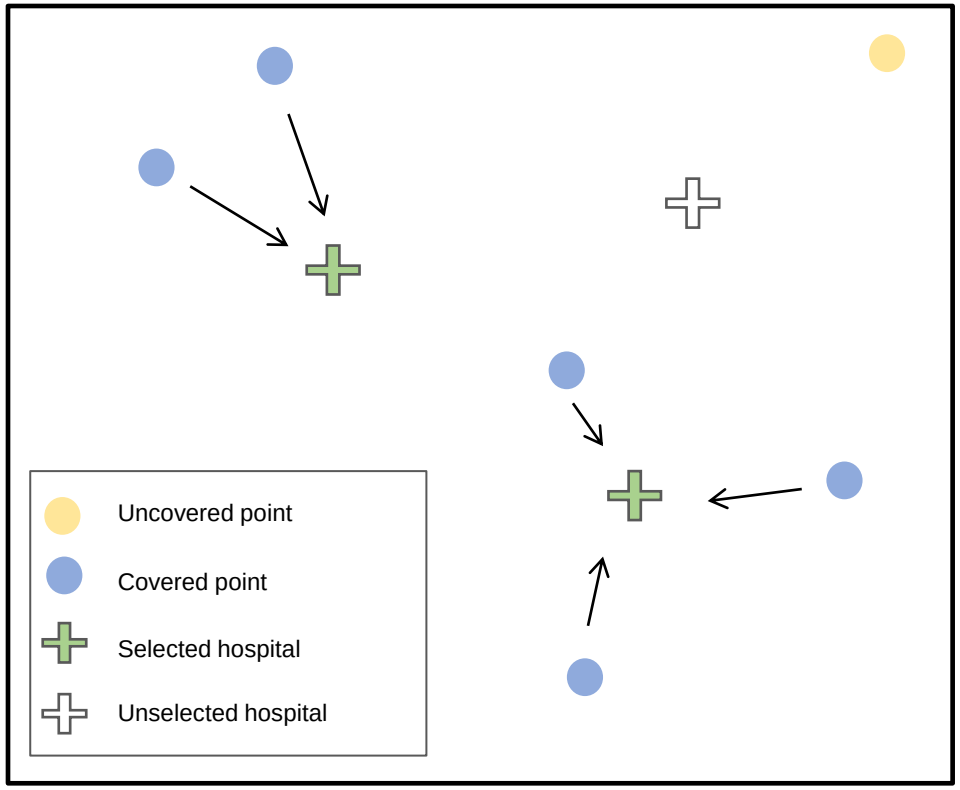
$x_j = 1,0 \quad \forall j$ (a candidate facility is either open or closed)

$y_i = 1,0 \quad \forall i$ (a demand area is either not covered or covered)

Where :

i, j, m and n are the same as in the above p-median model formulation

N_i and x_j are the same as in the above LSCP model formulation



Case Study 11A

Title Measuring Wasteful Commuting in Columbus, Ohio

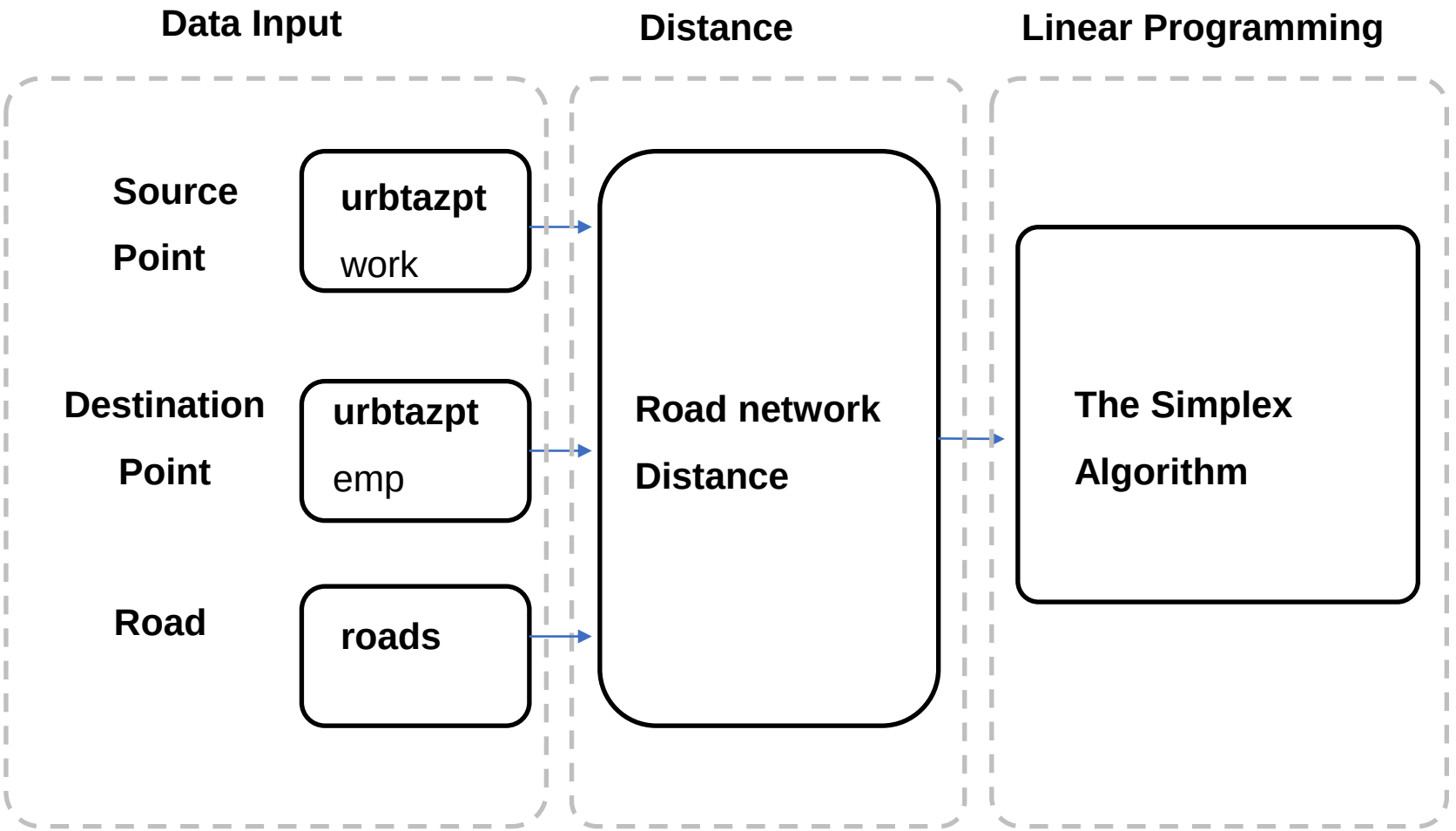
Objectives Minimize wasteful commuting

Methodology The Simplex Algorithm

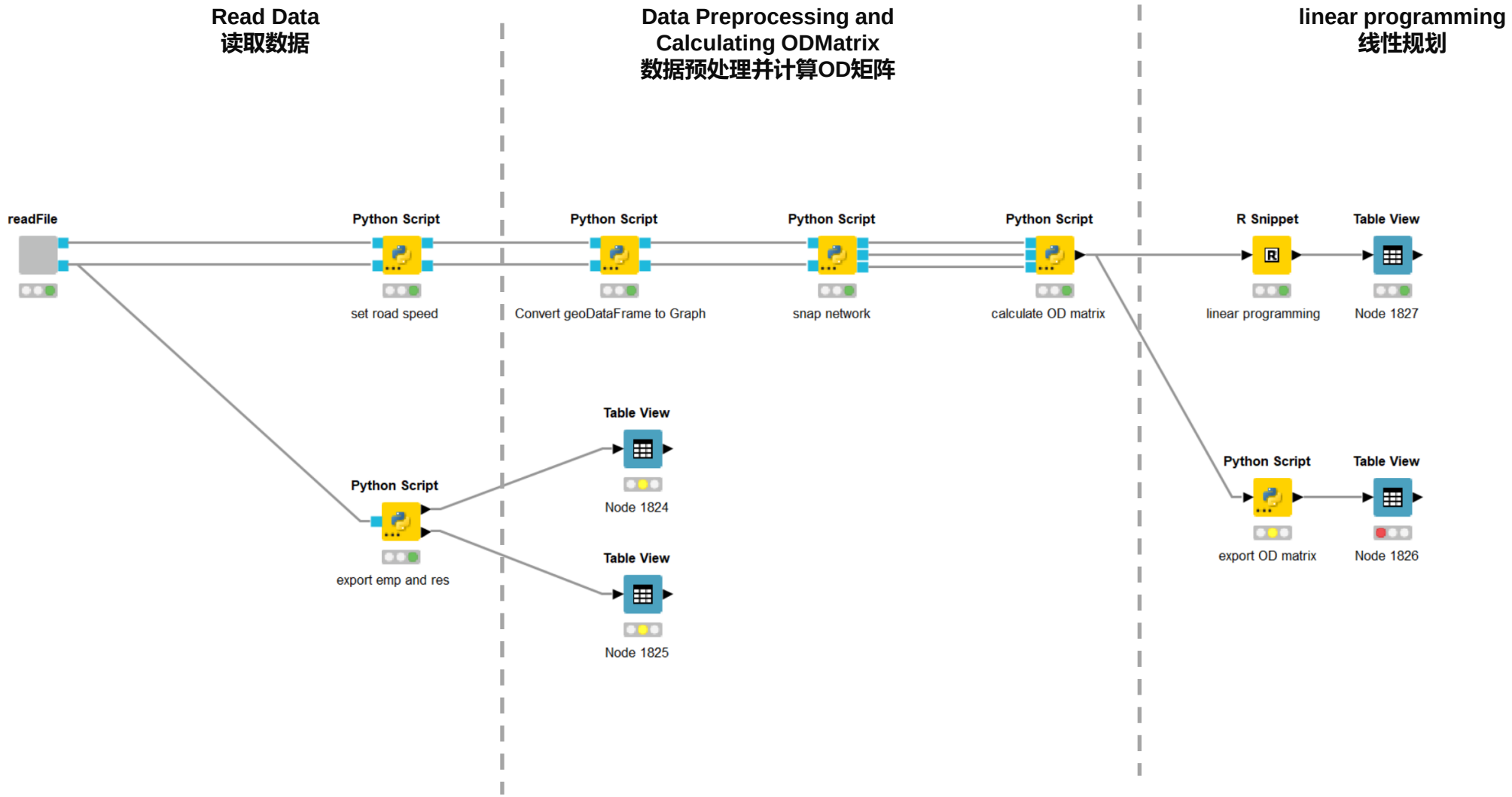
Data

Shapefile	Type	File Detail	Key Fields	Key Variable	Data Source
urbtazpt	Point	traffic analysis centroids	**work;emp**	Resident and employment Population in each census tract	TIGER files 1990 US census ; 1990 CTPP Urban Element data sets
urbtaz	Polygon	traffic analysis zone			TIGER files 1990 US census ; 1990 CTPP Urban Element data sets
roads	Polyline	the road network			TIGER files 1990 US census

Flowchart for Case Study



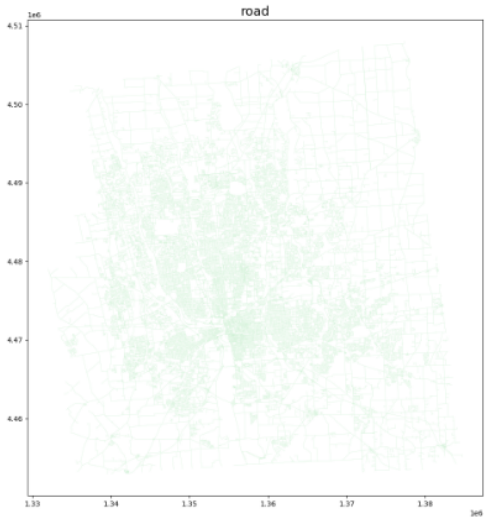
The Workflow Implementation



Step by Step Workflow Execution

01 Read Data | 读取数据

roadnet



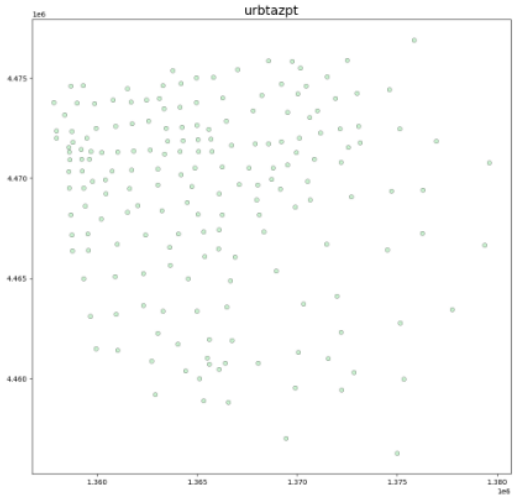
Upload road zip

Select file roads.zip

data sample

road.shp (zip)

demand and supply point



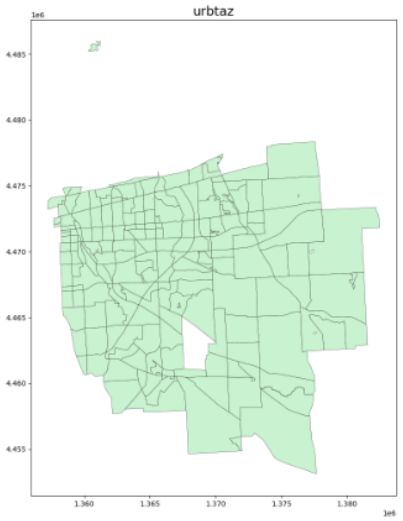
Upload urbtazpt zip

Select file urbtazpt.zip

data sample

urbtazpt.shp (zip)

demand and supply zone



Upload urbtaz zip

Select file urbtaz.zip

data sample

urbtaz.shp (zip)

Step by Step Workflow Execution

02 Data Preprocessing and Calculating ODMatrix | 数据预处理并计算OD矩阵

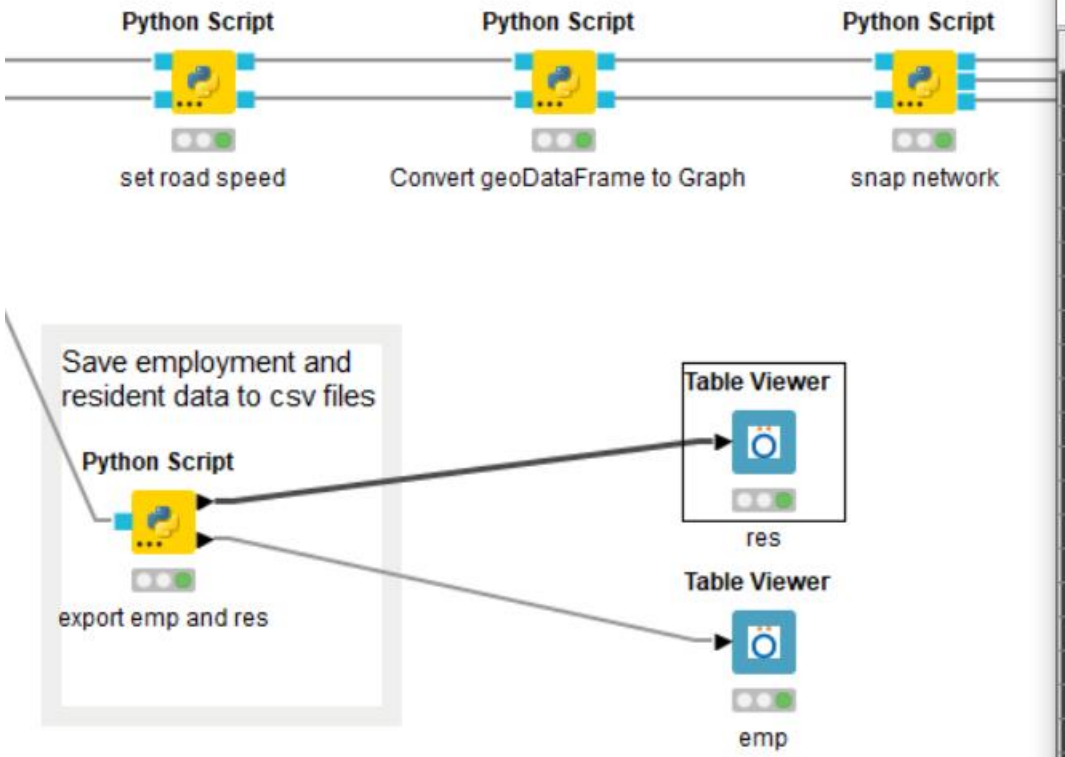


Table View - 0:1781 - Table Viewer (res)

File Hilite Navigation View Output

Row ID	OBJECTID	AREA	TAZ	POPU	WORK	EMP
Row0	1	20,237,489...	4504	2142	1215	18
Row1	2	8,971,957.101	4511	402	181	10
Row2	3	7,180,775.844	4412	2227	993	1070
Row3	4	6,625,105.568	6308	169	99	520
Row4	5	5,336,216.463	6310	870	475	12
Row5	6	4,595,197.213	6324	1231	648	23
Row6	7	5,070,328.297	6311	246	129	66
Row7	8	6,238,362.552	5313	115	35	44
Row8	9	3,562,001.316	5303	3990	1913	187
Row9	10	3,771,856.553	5311	2493	1421	128
Row10	11	4,644,963.869	4320	96	69	42
Row11	12	5,476,080.052	4401	80	41	0
Row12	13	1,448,665.641	4329	1182	587	155
Row13	14	1,878,883.037	4401A	43	24	87
Row14	15	16,264,089.73	4403	877	492	13
Row15	16	1,520,755.781	4406	239	75	69
Row16	17	4,884,677.763	4404	342	131	124
Row17	18	12,202,707...	6407	352	216	661
Row18	19	10,083,036...	6408	497	298	0
Row19	20	19,503,674...	4506	1439	754	17
Row20	21	54,251.62	4405	97	64	305
Row21	22	4,411,089.689	2417	257	114	315
Row22	23	4,322,570.273	2409	177	95	56

Step by Step Workflow Execution

03 linear programming | 线性规划

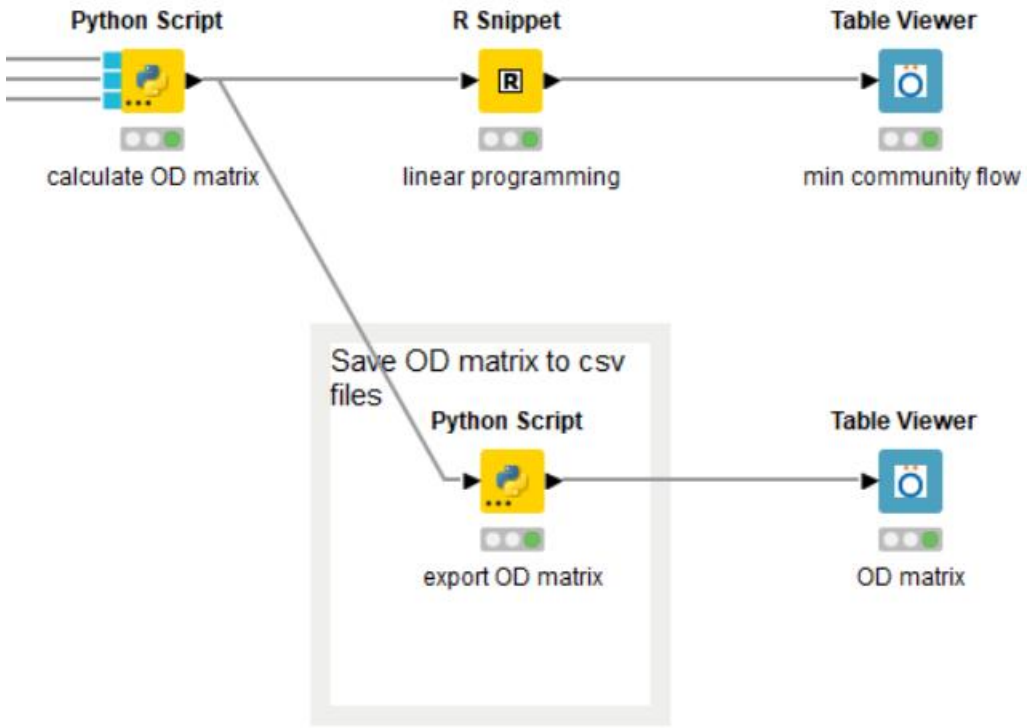


Table View - 3:1840 - Table Viewer

File Hilite Navigation View Output

Row ID	from	to	dist
Row0	175	1	20.334
Row1	175	2	22.742
Row2	175	3	21.882
Row3	175	4	21.791
Row4	175	5	18.965
Row5	175	6	19.491
Row6	175	7	21.406
Row7	175	8	26.193
Row8	175	9	19.35
Row9	175	10	25.235
Row10	175	11	29.044
Row11	175	12	21.253
Row12	175	13	4.033
Row13	175	14	2.575
Row14	175	15	7.735
Row15	175	16	2.917
Row16	175	17	4.933
Row17	175	18	7.253
Row18	175	19	11.794
Row19	175	20	10.135
Row20	175	21	12.272

Step by Step Workflow Execution

Results | 结果

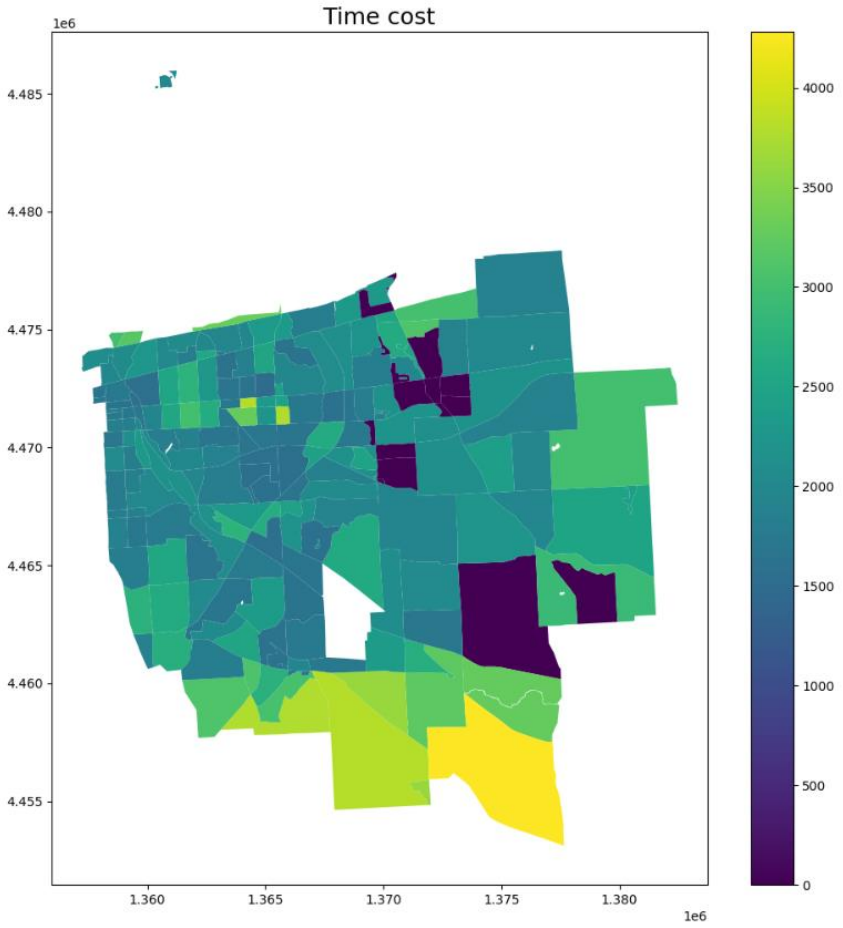


Table View - 0:1784 - Table Viewer (min community flow)

File Hilite Navigation View Output

Row ID	Oid	Did	Time	Flow
1	1	1	0	18
2	1	2	5.142	0
3	1	3	9.234	0
4	1	4	33.744	0
5	1	5	32.65	0
6	1	6	32.784	0
7	1	7	35.231	0
8	1	8	31.658	0
9	1	9	25.279	0
10	1	10	23.008	0
11	1	11	24.909	0
12	1	12	20.641	0
13	1	13	17.816	0
14	1	14	18.341	0
15	1	15	18.898	0
16	1	16	15.763	0
17	1	17	7.45	0
18	1	18	15.686	0
19	1	19	35.12	0
20	1	20	12.013	0
21	1	21	7.202	0
22	1	22	28.437	0
23	1	23	33.268	0
24	1	24	28.222	0

Analysis

Explanation

The output table contain origin and destination tracts, and the travel time and number of commuters on the route. The minimum commute time is **3.53 minutes** per resident worker. The actual mean commute time was **20.40 minutes** in the study area in 1990, implying an astonishing **82.7%** wasteful commuting.

Future work

- Temporal trends of intraurban commuting
- Commuting Variability by Wage Groups

Case Study 11B

Title Allocating Healthcare Providers in Baton Rouge, Louisiana

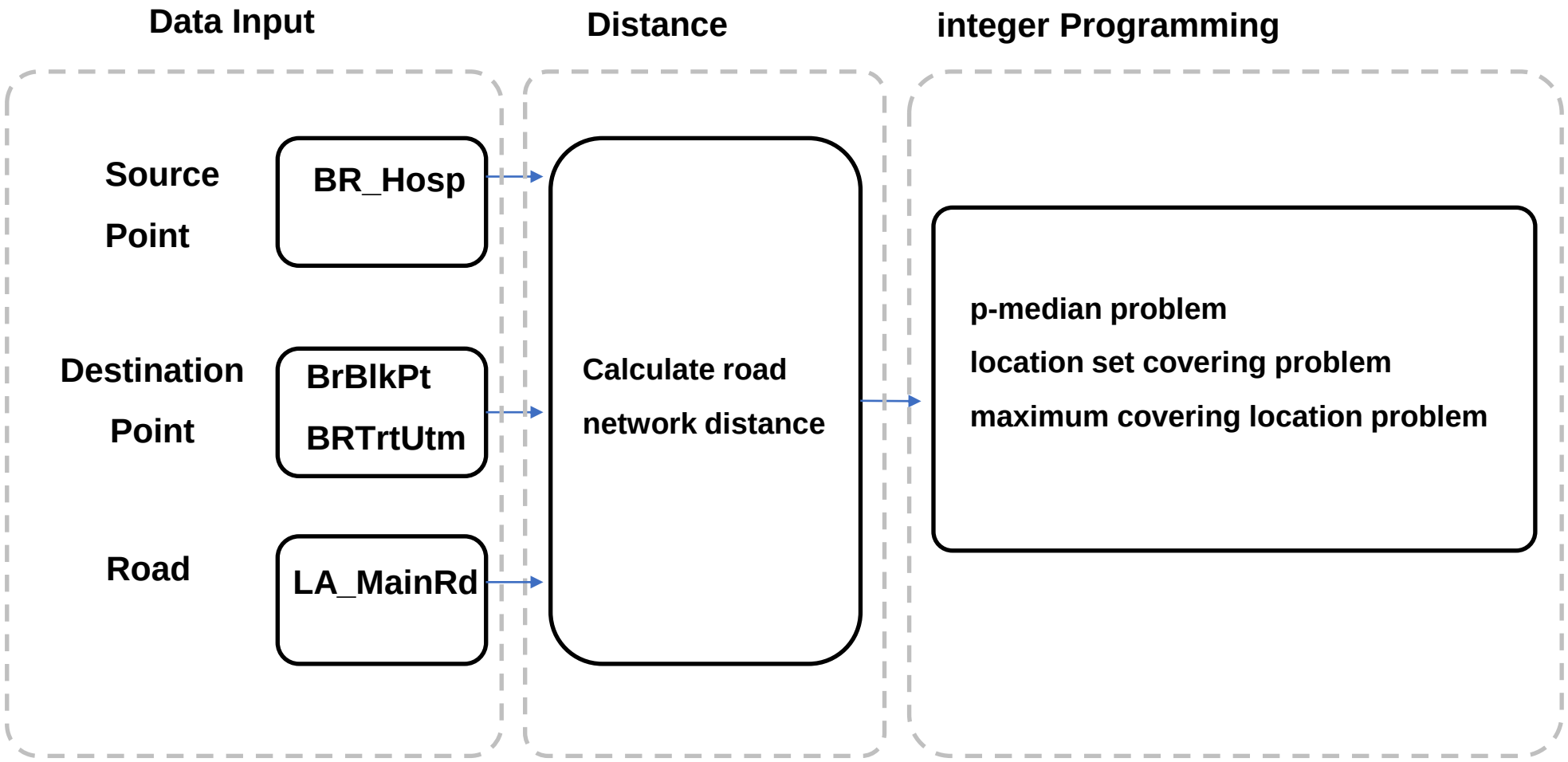
Objectives Location-Allocation Problems

Methodology p-median model;LSCP model;MCLP model

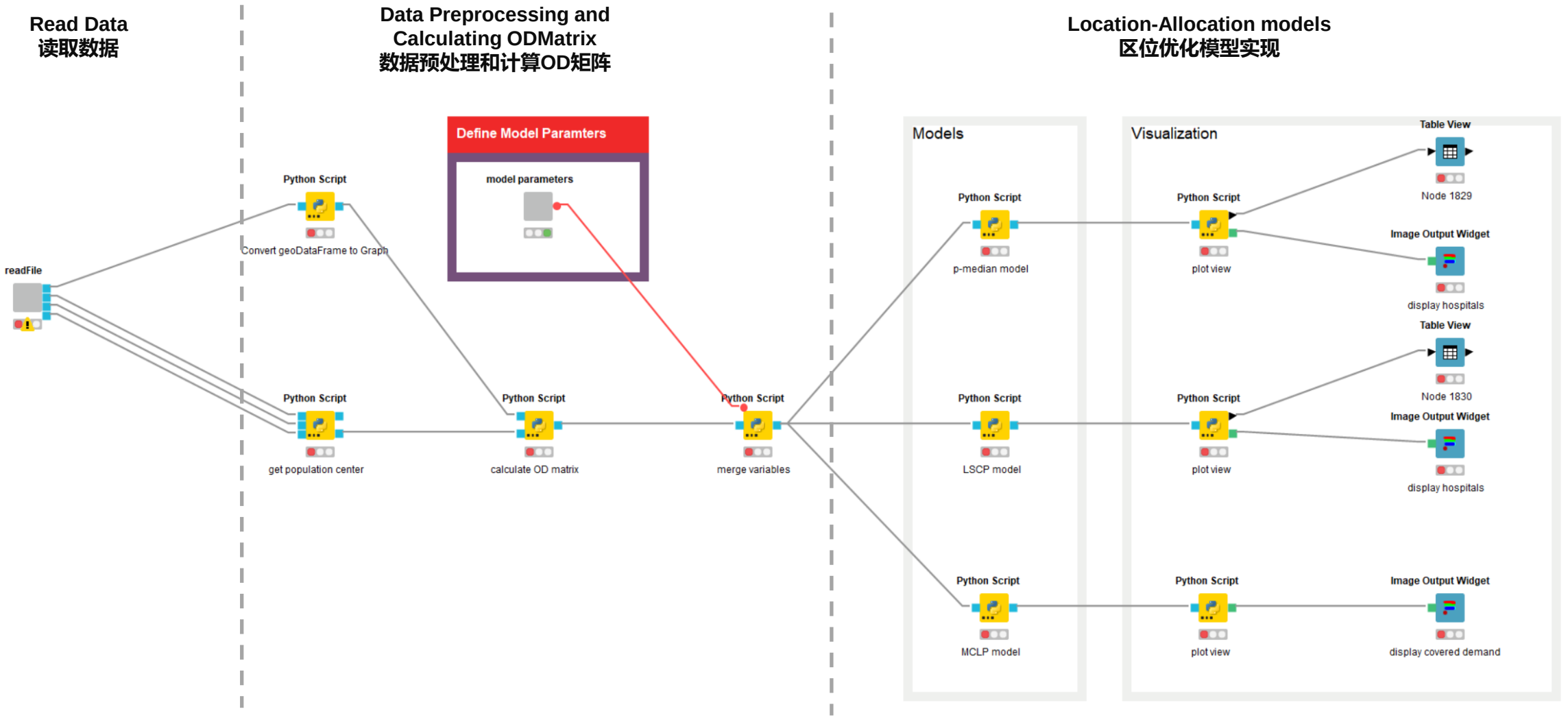
Data

Shapefile	Type	File Detail	Key Fields	Key Variable	Data Source
BR_Hosp	Point	26 hospitals			Case Study 1
BrBlkPt	Point	census block centroids	POP10	Population in each census tract	Case Study 1
BRTrtUtm	Polygon	census tracts			Case Study 3B
LA_MainRd	Polyline	the road network			Case Study 2

Flowchart for Case Study



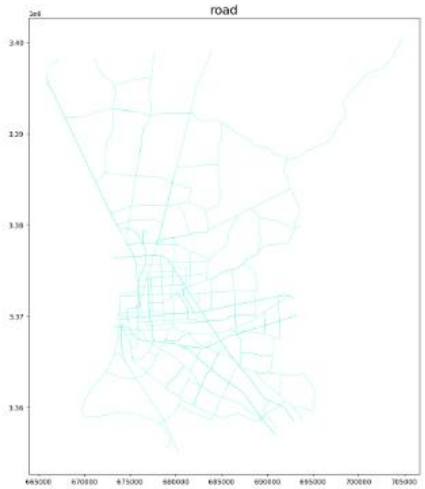
The Workflow Implementation



Step by Step Workflow Execution

01 Read Data | 读取数据

roadnet



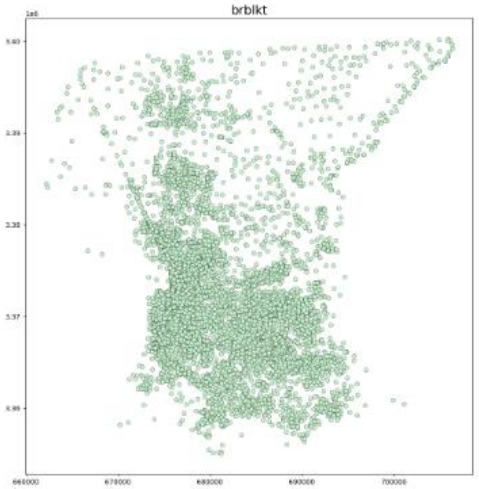
data sample

LA_MainRd.shp (zip)

Upload LA_MainRd zip

Select file LA_MainRd.zip

brblkt



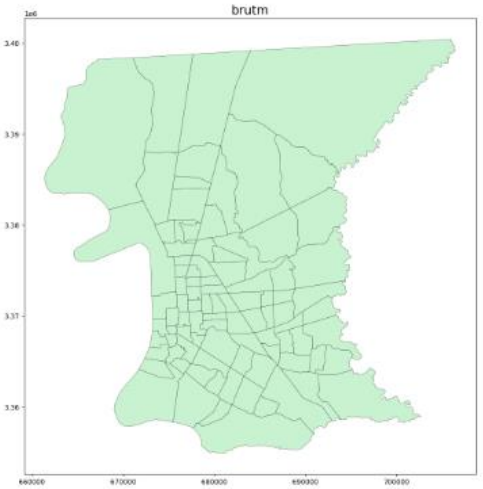
data sample

BRBlkPt.shp (zip)

Upload BRBlkPt zip

Select file BRBlkPt.zip

brutm



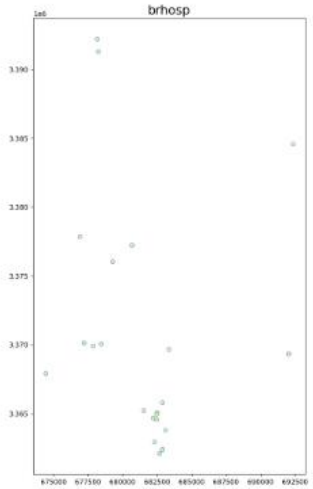
data sample

BRTrtUtm.shp (zip)

Upload BRTrtUtm zip

Select file BRTrtUtm.zip

brhosp



data sample

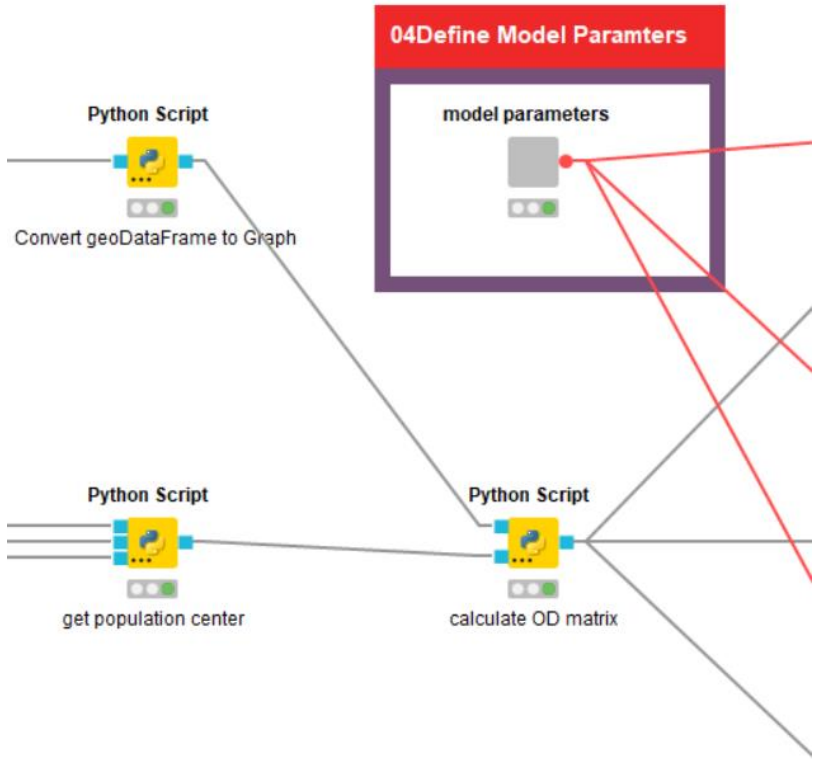
BR_Hosp.shp (zip)

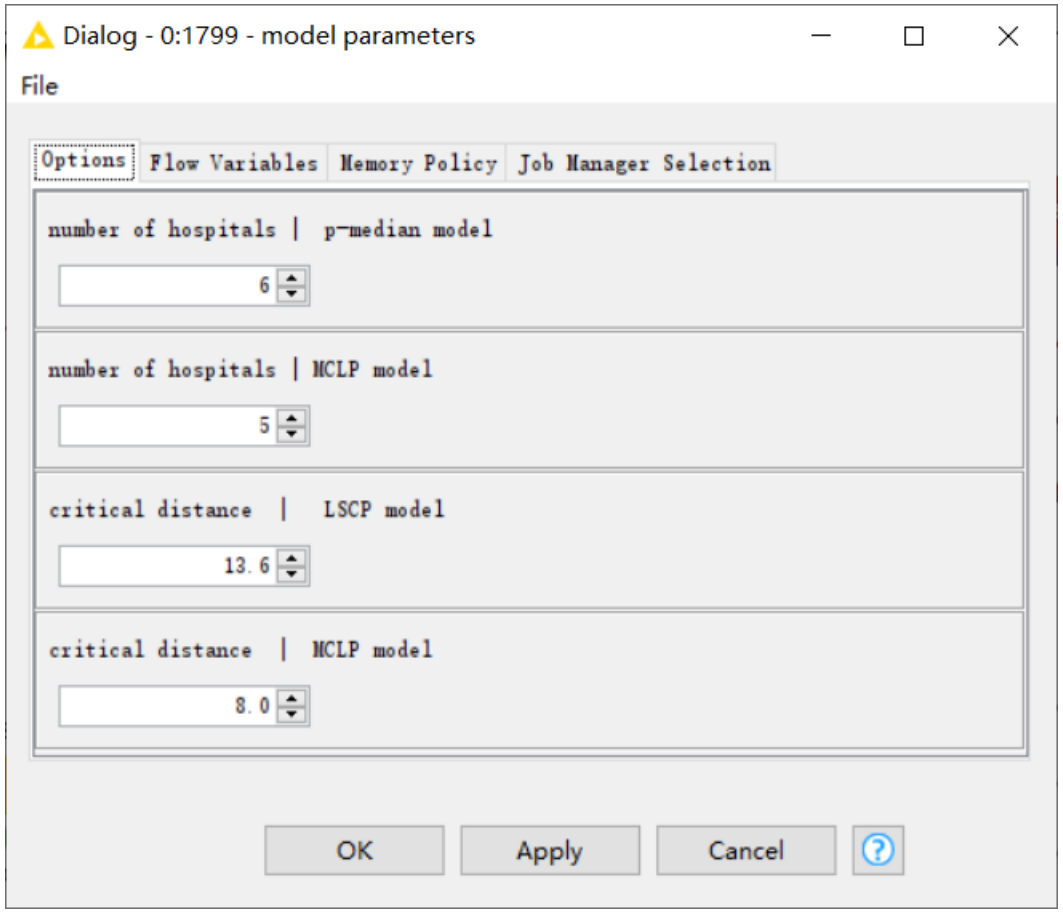
Upload BR_Hosp zip

Select file BR_Hosp.zip

Step by Step Workflow Execution

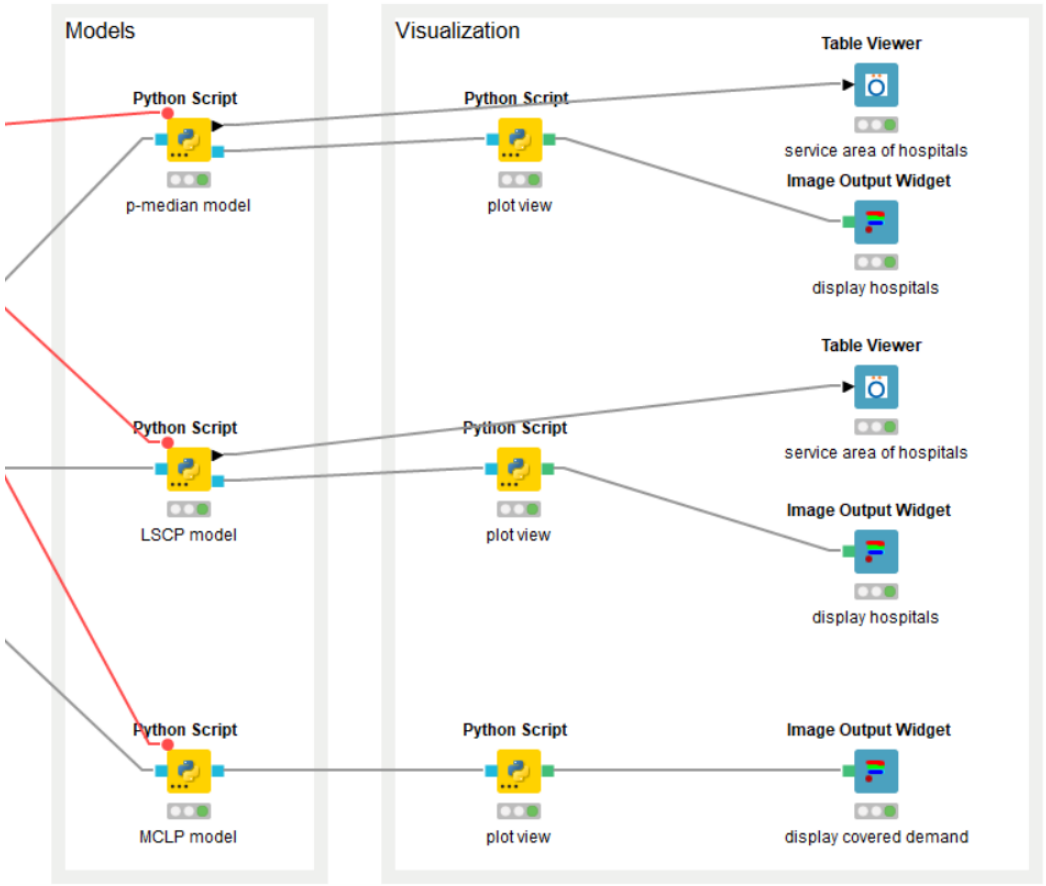
02 Data Preprocessing and Calculating ODMatrix | 数据预处理和计算OD矩阵





Step by Step Workflow Execution

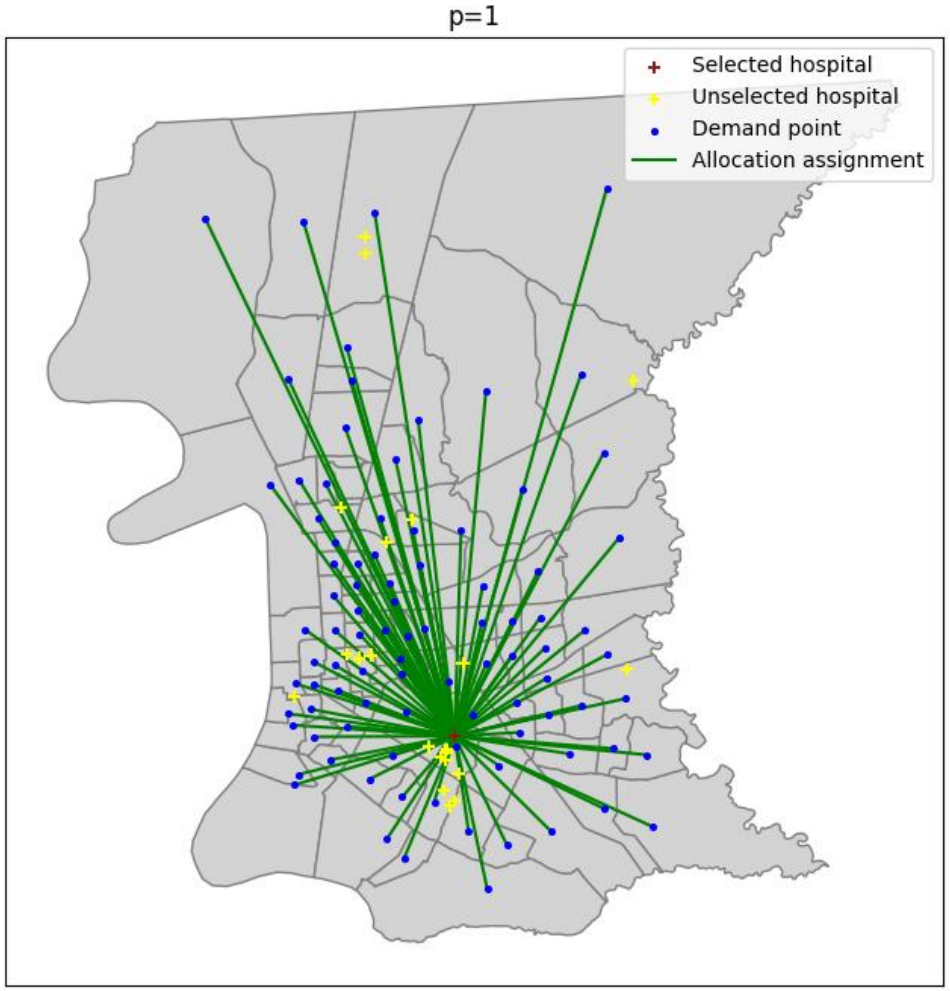
03 Location-Allocation models | 区位优化模型实现




```
import pulp
import pandas as pd

BRTrtPt = input_object_1
brhosp = input_object_2
dist_df = input_table_1

num_trt=len(BRTrtPt)
num_hosp=len(brhosp)
problem=pulp.LpProblem("MyProblem",sense=pulp.LpMinimize)

def getIndex(s):
    li=s.split()
    if len(li)==2 :
        return int(li[-1])
    else :
        return int(li[-2]),int(li[-1])

X=['X'+' '+str(i)+' '+str(j) for i in range(num_trt) for j in range(num_hosp)]
Y=['Y'+' '+str(j) for j in range(num_hosp) ]
varX=pulp.LpVariable.dicts("分配",X,cat="Binary")
varY=pulp.LpVariable.dicts("是否分配",Y,cat="Binary")

object_list=[]
for it in X:
    i,j=getIndex(it)
    cost = dist_df[(dist_df['from']== i) & (dist_df['to']== j)]['cost'].values[0]
    object_list.append(BRTrtPt['DP0010001'][i]*cost*varX[it])
problem+=pulp.lpSum(object_list)

problem+=pulp.lpSum([varY[it] for it in Y])==5
for i in range(num_trt):
    problem+=pulp.lpSum([varX['X'+' '+str(i)+' '+str(j)] for j in range(num_hosp) ])==1
for i in range(num_trt):
    for j in range(num_hosp):
        problem+=(varX['X'+' '+str(i)+' '+str(j)]-varY['Y'+' '+str(j)]<=0)

problem.solve()
print(pulp.LpStatus[problem.status])

area = pd.DataFrame(columns=['OID','name','num_tract','total_population'])
brhosp['choosen']=0
for it in Y:
```

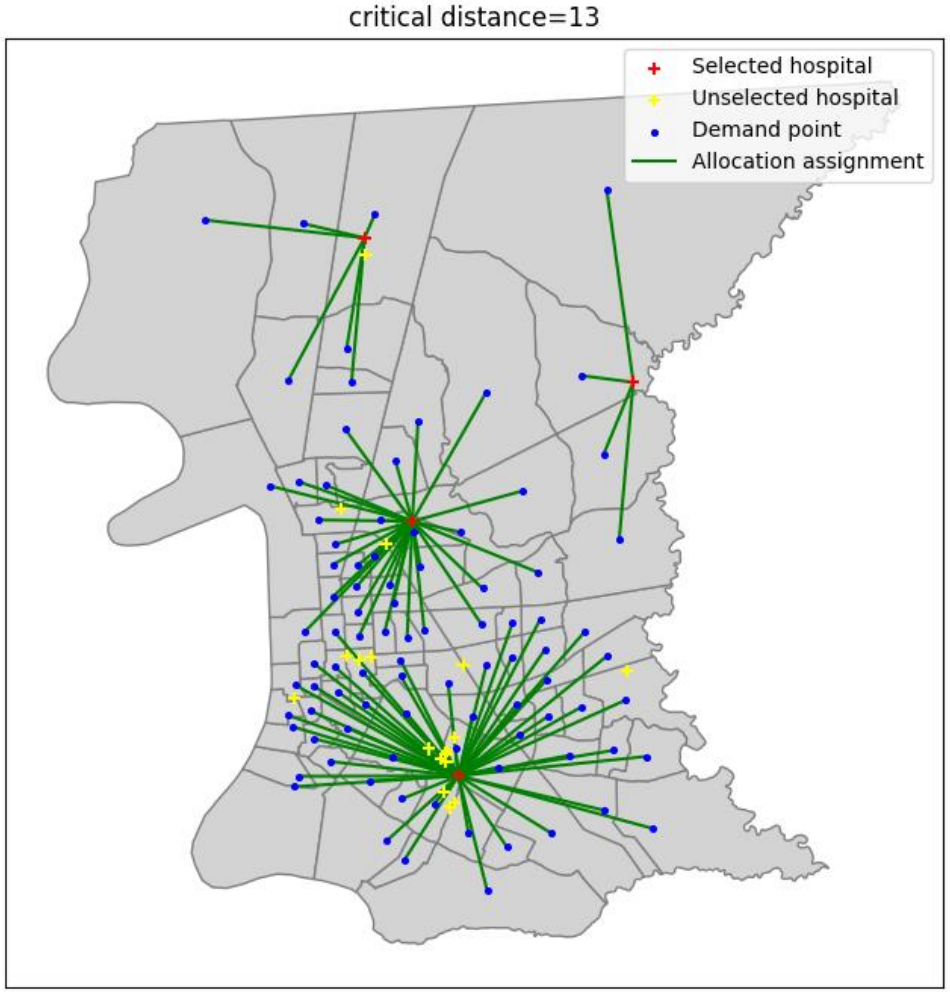
Step by Step Workflow Execution

Results | p-median model



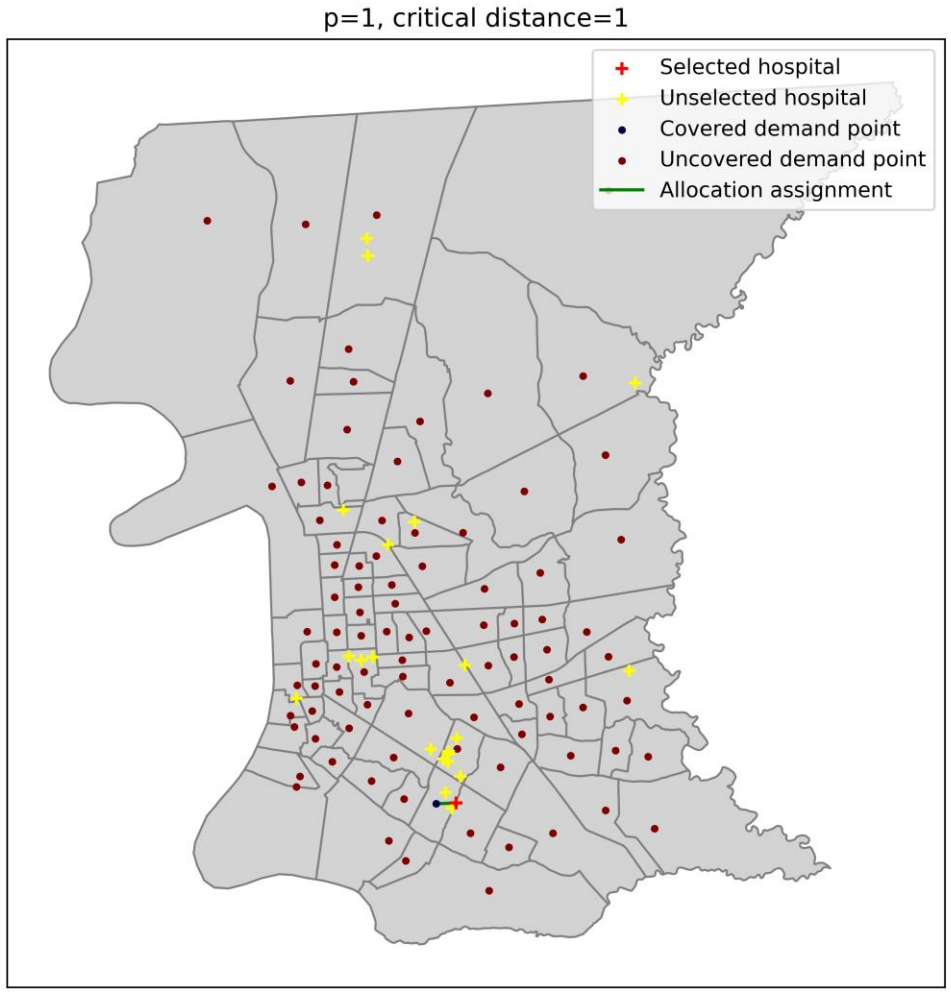
Step by Step Workflow Execution

Results | LSCP model



Step by Step Workflow Execution

Results | MCLP model



Analysis

Explanation

The output gifs show which facility a tract is assigned to. The green line shows the allocation assignment between a demand point (census tract) and a chosen hospital (clinic). We can see the optimal assignment results under different parameter settings.

Future work

- Two-Step Optimization for Spatial Accessibility Improvement
- A two-step approach to planning new facilities towards equal accessibility

Benefit of simulation

- Identify sensitive factors of important influence
- obtain the optimal results in different situations

Discussion

- The location-allocation models are just a set of examples of LP applications in location analysis.
- These models are simple and easy to implement and can be used in allocating healthcare providers, logistics centers, firefighting facilities.
- With the increase of the size of the location problems, it is difficult for the precise algorithm to obtain the optimal solution, which promotes the application of the inexact algorithm represented by the heuristic algorithm.
- The service capacity and construction cost of the hospital are not considered.
- The challenge of uncertain factors ,such as dynamic demand.

References

Fahui Wang, 2015. Quantitative Methods and Socio-Economic Applications in GIS. https://www.routledge.com/Quantitative-Methods-and-Socio-Economic-Applications-in-GIS/Wang/p/book/9781138843622.

Luo J, Tian L, Luo L, et al. Two-Step Optimization for Spatial Accessibility Improvement: A Case Study of Health Care Planning in Rural China[J]. 2017.

Li X, Wang F, Yi H. A two-step approach to planning new facilities towards equal accessibility[J]. Environment and Planning B: Urban Analytics and City Science, 2017, 44(6): 994-1011.

Hu Y, Wang F, Wilmot C G. Commuting variability by wage groups in Baton Rouge, 1990–2010[J]. Papers in Applied Geography, 2017, 3(1): 14-29.

Hu Y, Wang F. Temporal trends of intraurban commuting in Baton Rouge, 1990–2010[J]. Annals of the American Association of Geographers, 2016, 106(2): 470-479.

THANKS

Spatial Data Lab http://spatialdatalab.org

Contact Spatialdatalab@lists.fas.harvard.edu

lingjie@whu.edu.cn

Center for Geographic Analysis

Harvard University

Geo-computation Center for Social Sciences

Wuhan University

China Data Institute

Future Data Lab