Spatial Social Science

Workflow-Based Practices Series I

Workbook for Quantitative Methods and Socio-Economic Applications in GIS

Fahui Wang | Shuming Bao

Lingbo Liu

Spatial Data Lab
CGA @ Harvard University
Louisiana State University
Wuhan University

Spatial Social Science

Workflow-Based Practices Series I

Spatial Data Lab
CGA @ Harvard University
Louisiana State University
Wuhan University

Contents

- About this project
- Benefit from this project
- About KNIME workflow
- Prototype of the work
- Future work
- How to join the project?
- Website and contacts

Spatial Science

Workflow-Based Practices Series I

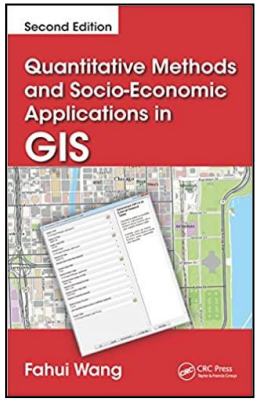
Workbook for Quantitative Methods and Socio-Economic Applications in GIS

1 About this project

Book and Author

Quantitative Methods and Socio-Economic Applications in GIS

By Fahui Wang, Louisiana State University



A bestseller, the book integrates GIS and quantitative methods and demonstrates them in various policy-relevant socio-economic applications with step-by-step instructions and datasets. The book illustrates the range of computational methods and covers common tasks and major issues encountered in a spatial environment. It provides a platform for learning technical skills and quantitative methods in the context of addressing real-world problems, giving you instant access to the tools to resolve major socio-economic issues.

Fahui Wang is a Professor and former Chair in the Department of Geography and Anthropology, Louisiana State University. He earned the B.S. in geography from Peking University, and M.A. in economics and PhD in city and regional planning from the Ohio State University. His research has evolved the broad theme of spatially-integrated social sciences, public policy and planning in Geographic Information Systems. He is among the top 1% most-cited researchers in Geography in the world.

Table of Contents

- **GIS AND BASIC SPATIAL ANALYSIS TASKS**

1. Getting Started with ArcGIS
2. Measuring Distance and Time
3. Spatial Smoothing and Spatial Interpolation

- **BASIC QUANTITATIVE METHODS AND APPLICATIONS**

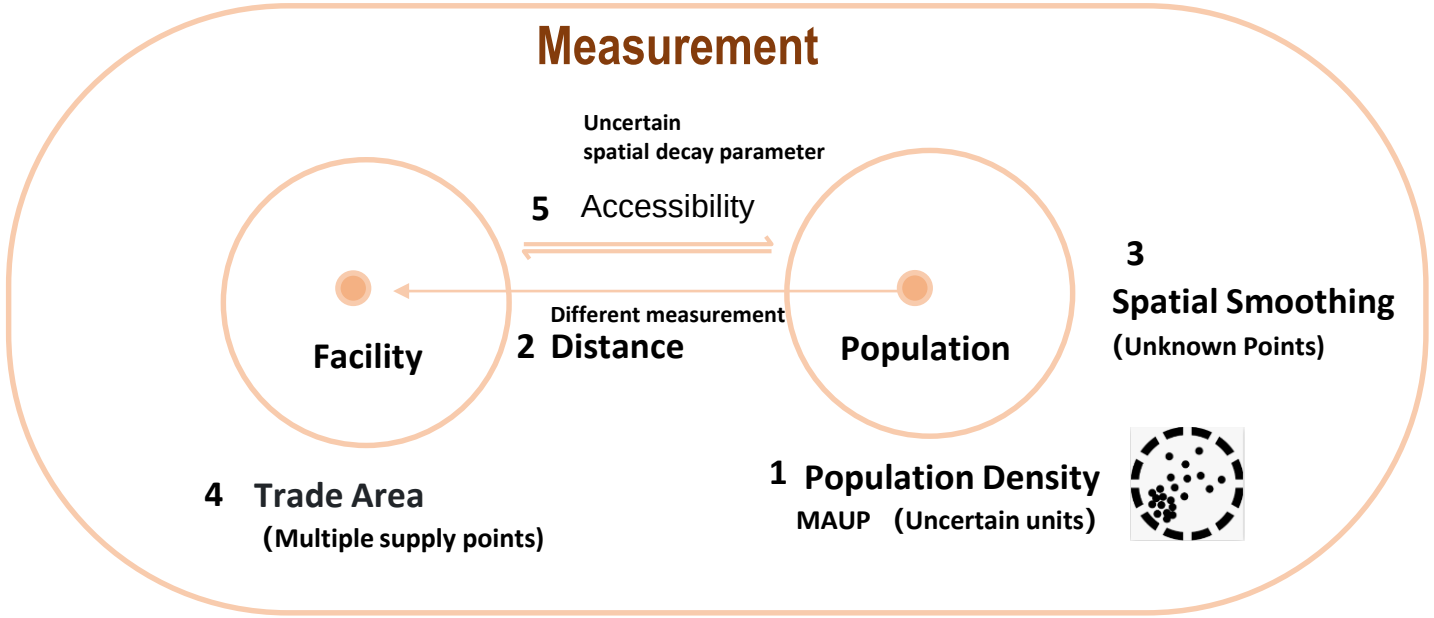
4. GIS-Based Trade Area Analysis and Application in Business Geography
5. GIS-Based Measures of Spatial Accessibility and Application in Examining Health Care Access
6. Function Fittings by Regressions and Application in Analyzing Urban Density Patterns
7. Principal Components, Factor and Cluster Analyses, and Application in Social Area Analysis
8. Spatial Statistics and Applications

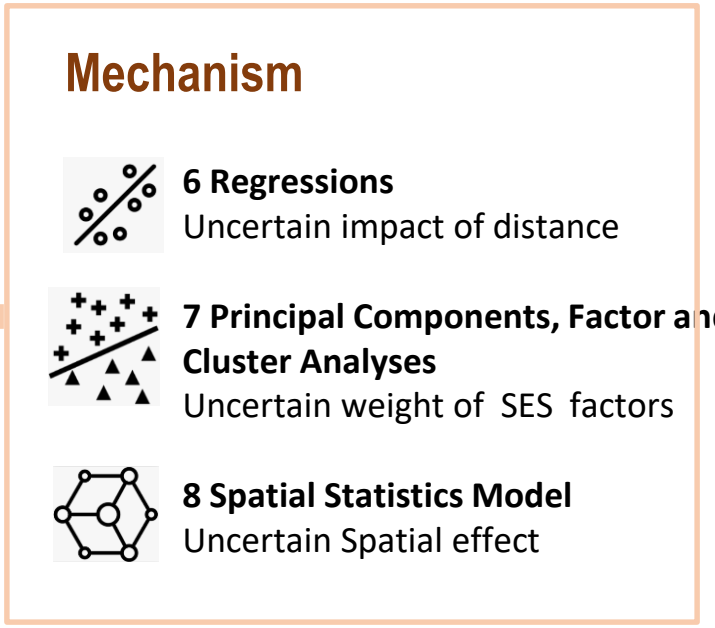
- **ADVANCED QUANTITATIVE METHODS AND APPLICATIONS**

9. Regionalization Methods and Application in Analysis of Cancer Data
10. System of Linear Equations and Application of Garin-Lowry Model in Simulating Urban Population and Employment Patterns
11. Linear Programming and Applications in Examining Wasteful Commuting and Allocating Health Care Providers
12. Monte Carlo Method and Its Application in Urban Traffic Simulation

Chapter Structure

Popular Methods in Social Spatial Science, Uncertain Geographic Context Problem





Optimization / Simulation

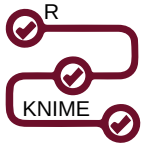
9. Regionalization Methods and Application in Analysis of Cancer Data （MAUP）
10. System of Linear Equations in Simulating Urban Population and Employment Patterns
11. Linear Programming in Examining Wasteful Commuting and Allocating Health Care Providers
12. Monte Carlo Method in Urban Traffic Simulation

Goal and Objectives of the workbook project

GOAL: to promote **REPLICABLE**, **REPRODUCIBLE** and **EXPANDABLE** research

OBJECTIVES:

DEVELOP a workbook for “Quantitative Methods and Socio-Economic Applications in GIS”, including test data sets, workflow, report, and presentation.



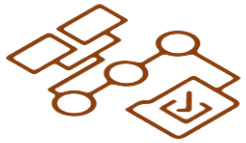
REPRODUCE the cases in the book based on workflow for efficient data analysis

HELP students from any fields learn and **MASTER** quantitative methods, technology and applications for spatial analysis in social sciences

APPLY and **EXPAND** those methods to other studies, **BUILD** knowledge map and seek research gap

What you can benefit from this project

LEARN GIS, workflow tools and quantitative methods for spatial analysis for socioeconomic studies.



How to **USE KNIME** and build workflows for data analysis?

How to **REPRODUCE CASE** studies for spatial data analysis with KNIME ?

How to use the existing workflows for other data or **RESEARCH QUESTIONS**?

Spatial Science

Workflow-Based Practices Series I

Workbook for Quantitative Methods and Socio-Economic Applications in GIS

2 About KNMIE workflow

Open Source Platform KNIME

Open for Innovation
KNIME

Hub Blog Forum Events Careers Contact **Download**

SOFTWARE / SOLUTIONS / LEARNING / PARTNERS / COMMUNITY / ABOUT

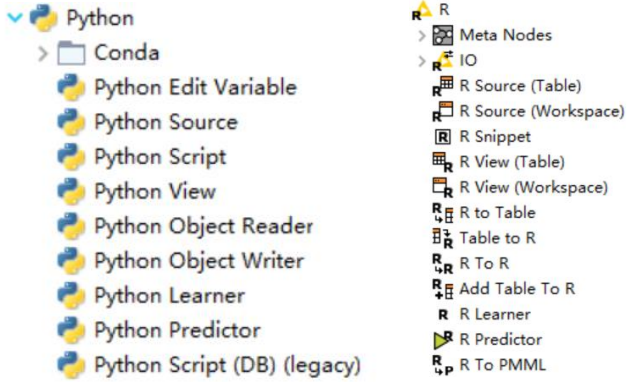
www.knime.com

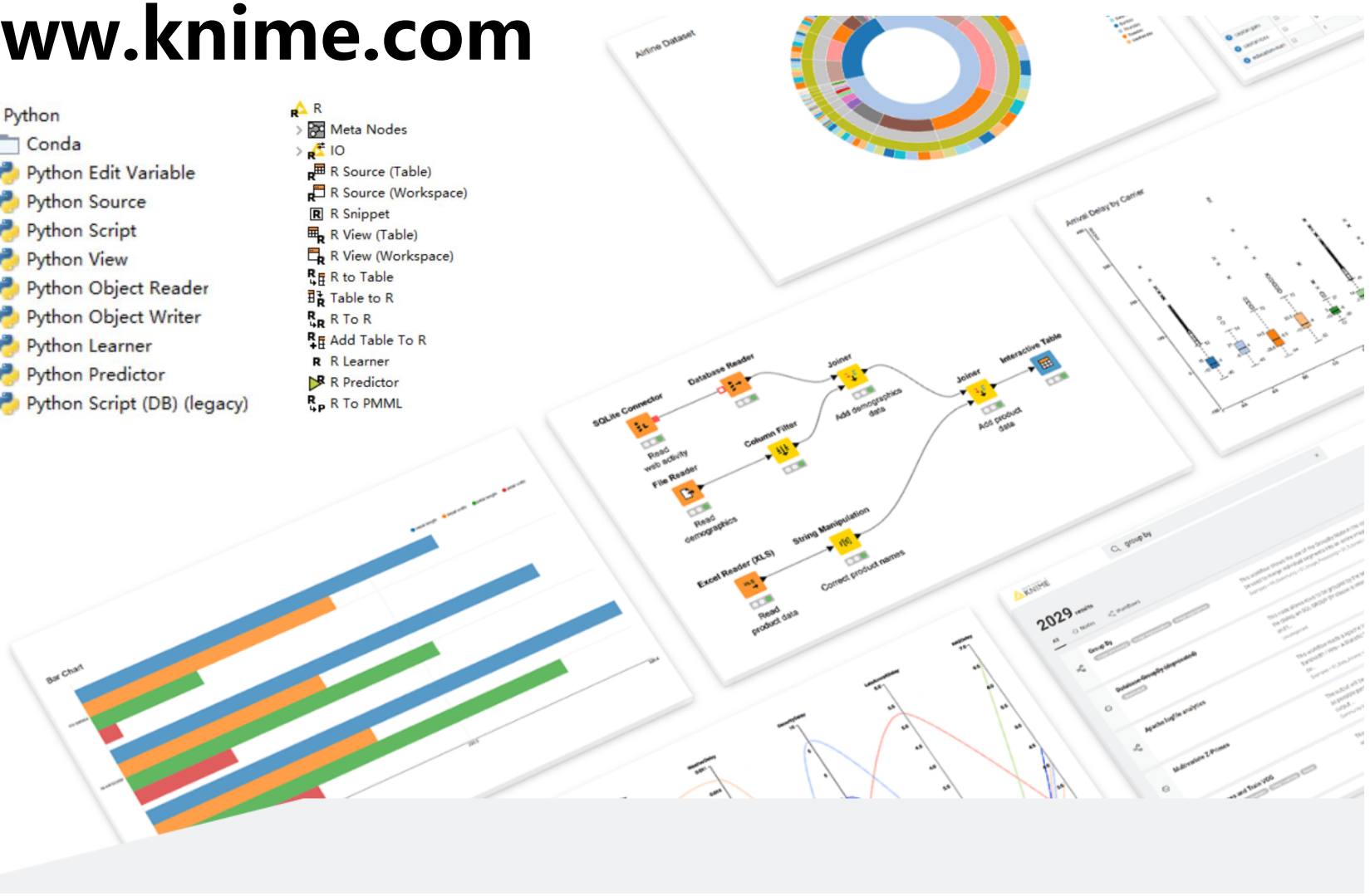
End to End Data Science

At KNIME, we build software to create and productionize data science using one easy and intuitive environment, enabling every stakeholder in the data science process to focus on what they do best.

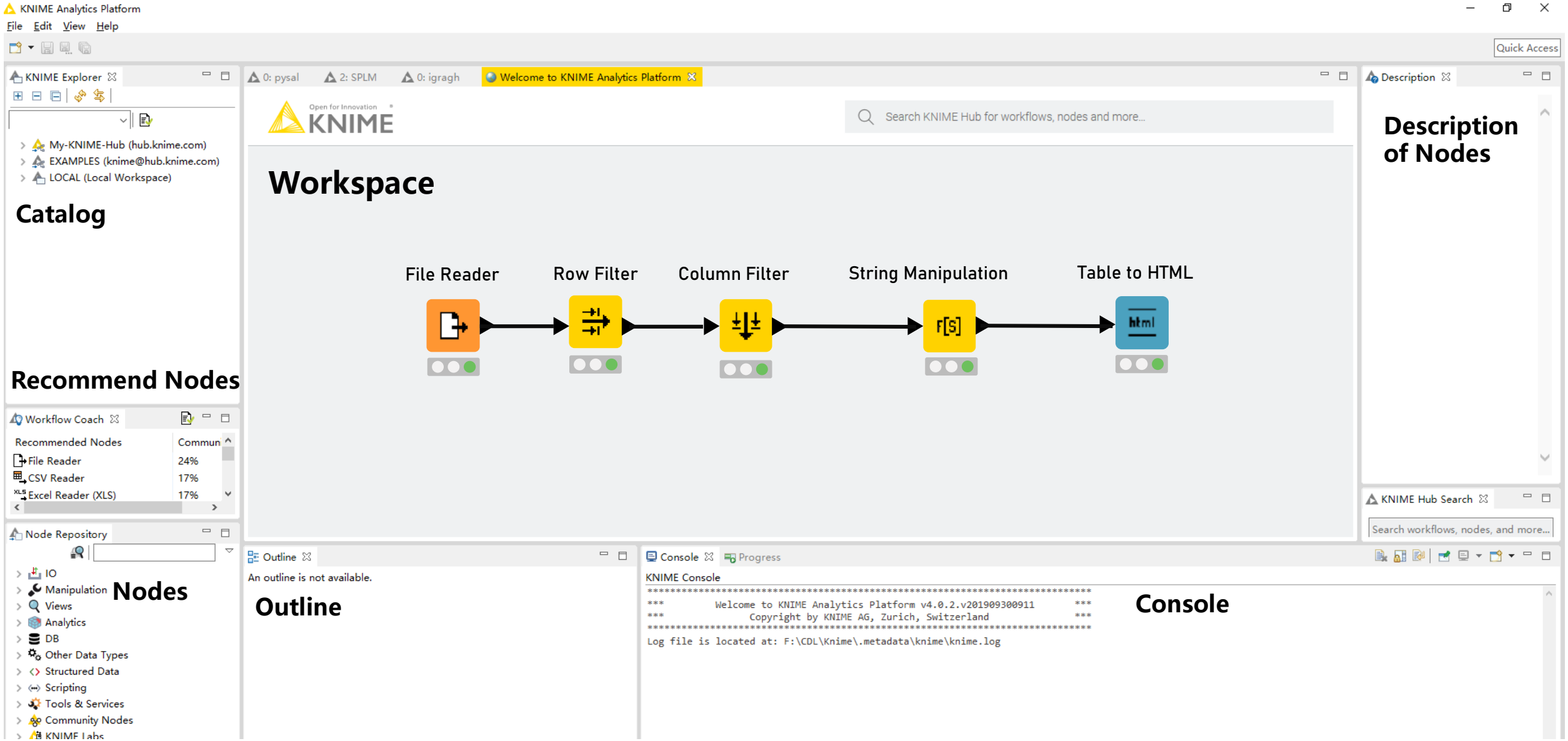
2000+ Tools | Python | R | Java

KNIME Software | KNIME Open Source Philosophy





Interface and Workspace



Built-in Nodes

Input/output

- IO
  - Read
  - Write
  - Connectors
  - File Folder Utility
  - Other
  - File Handling (legacy)
  - Cache
- Read
  - Excel Reader
  - File Reader
  - ARFF Reader
  - CSV Reader
  - Line Reader
  - Table Reader
  - PMML Reader
  - Fixed Width File Reader
  - Model Reader
  - Read Excel Sheet Names
  - Read Images
  - Simple File Reader
  - Explorer Browser

Manipulation

- Manipulation
  - Column
  - Row
  - Table
  - PMML
- Column
  - Binning
  - Convert & Replace
    - Category To Number
    - Category To Number (Apply)
    - Cell Replacer
    - Column Auto Type Cast
    - Column Rename
    - Column Rename (Regex)
    - Constant Value Column
    - Math Formula
    - Math Formula (Multi Column)
    - Number To Category (Apply)
    - Number To String
    - String To Number

Visualization

- JavaScript
  - Generic JavaScript View
  - Bar Chart
  - Box Plot
  - Conditional Box Plot
  - Decision Tree View
  - Heatmap
  - Histogram
  - Lift Chart
  - Line Plot
  - Parallel Coordinates Plot
  - Pie/Donut Chart
  - ROC Curve
  - Scatter Plot
  - Stacked Area Chart
  - Sunburst Chart

Modelling

- Mining
  - Bayes
  - Clustering
  - Rule Induction
  - Neural Network
  - Decision Tree
  - Decision Tree Ensemble
  - Misc Classifiers
  - Ensemble Learning
  - Item Sets / Association Rules
  - Linear Discriminant Analysis
  - Linear/Polynomial Regression
  - Logistic Regression
  - MDS
  - PCA
  - PMML

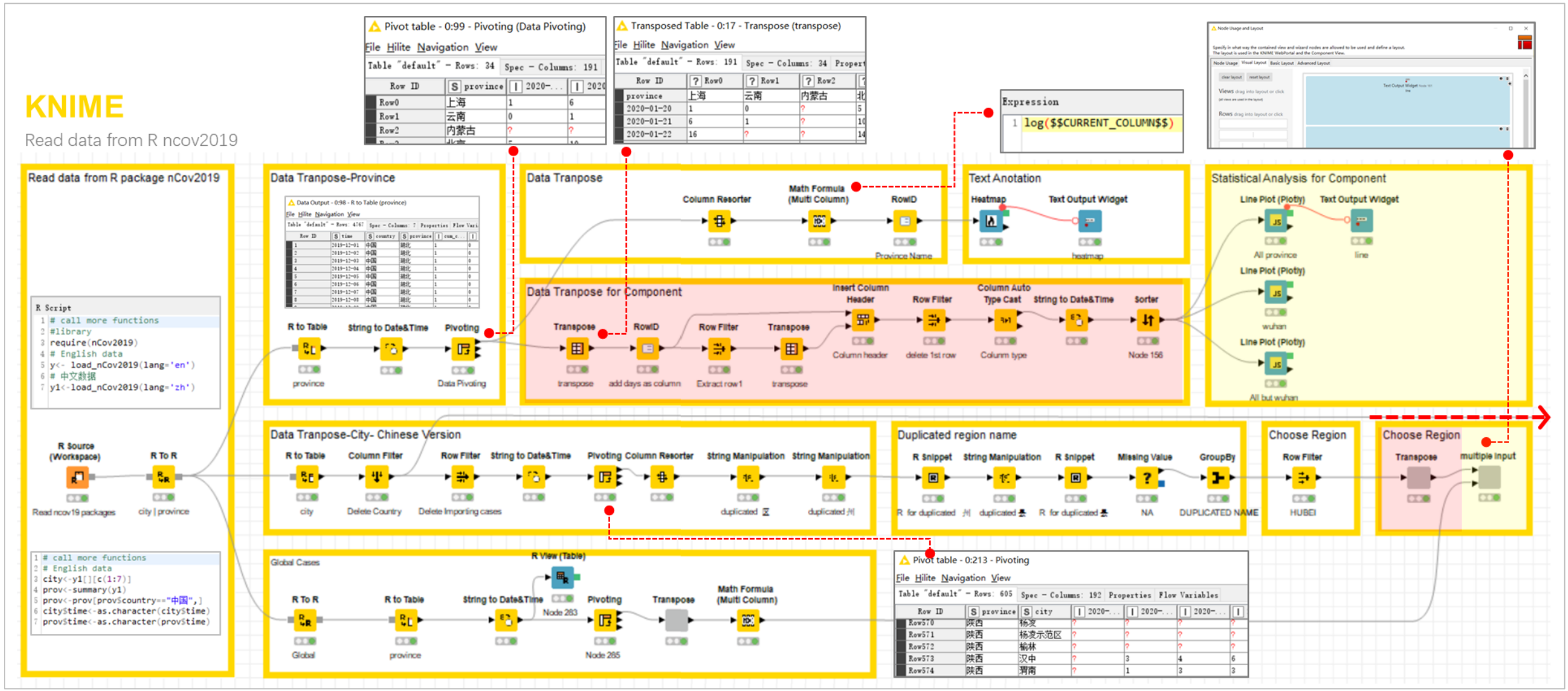
Advance +Plugin

- Community Nodes
  - ESRI Shapefiles
- KNIME Labs
  - Amazon Web Services
  - JavaScript Views (Labs)
  - Open Street Map
  - Column Expressions
  - Joiner (Labs)
  - Row Filter (Labs)
  - Row Splitter (Labs)
  - Variable Expressions
- Workflow Control
- Workflow Abstraction
- Reporting
- Chemistry

Scripting Nodes

- Java
  - Java Snippet
  - Java Snippet (simple)
  - Java Snippet Row Filter
  - Java Snippet Row Splitter

Example Works



Spatial Science

Workflow-Based Practices Series I

Workbook for Quantitative Methods and Socio-Economic Applications in GIS

3 Prototype Work- Chapter 1

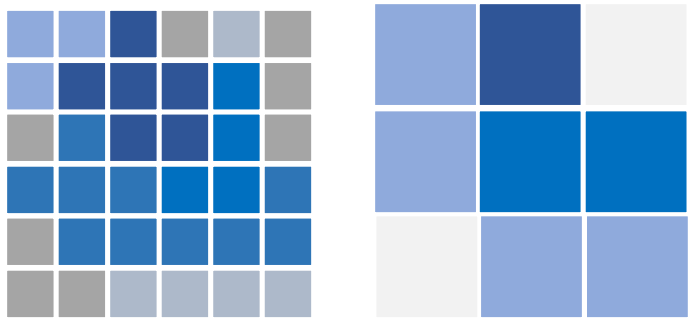
Contents of Chapter I

- ☐ Background
  - GIS | Coordinates system
- ☐ Main Concept
  - Features of Spatial Data
  - Join of Spatial Data
- ☐ Tool of Spatial Analysis
  - Coordinates System and Projection
  - Spatial Query
  - Spatial Query
  - Spatial Join
  - Attribute Join
  - Map Overlay
- ☐ Case Study
  - **Spatial Exploratory Analysis of Urban Center and Population**

Title & Objectives

Title **Mapping Population Density**

Objectives **Understanding the spatial feature of population distribution**
Comparing the result of different measurement of population Density

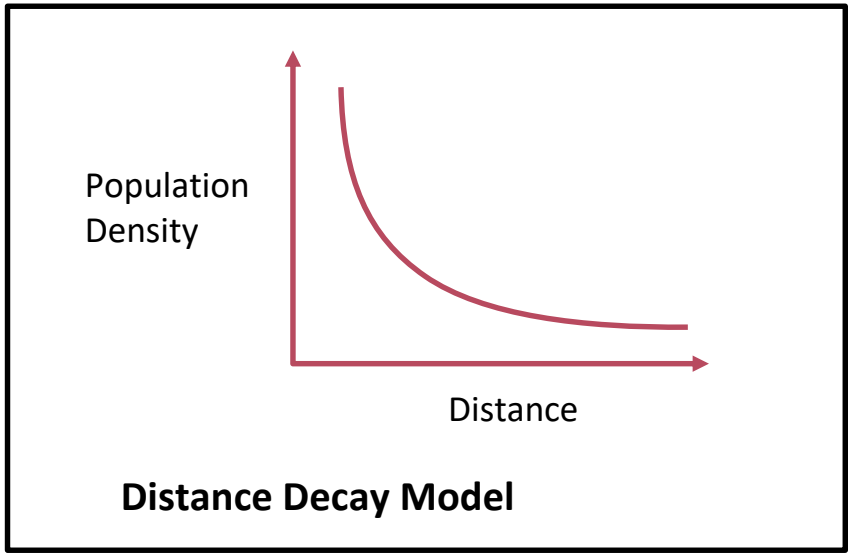


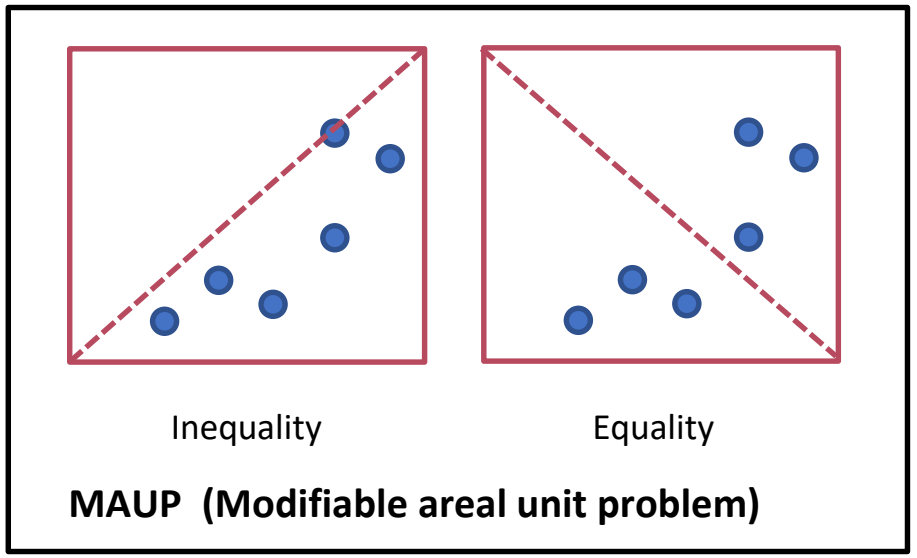
Methodology

Question

Distance Decay Model (Alonso Model)

MAUP (Modifiable areal unit problem)





Methods

Multiple Buffer Ring of Population Density

Different Scale of Geographic Units

Data

Data Source

US Census Data 2010

https://github.com/UrbanGISer/Quantitative-Methods-and-Socioeconomic-Applications-in-GIS

Administrative boundary of Louisianan
County-level administrative boundary of Louisianan
Tract-level administrative boundary of Louisianan (Population)
Center of East Baton Rough Parish
Block-level Census Data of East Baton Rough Parish (Population)

Variables & Definition

Distance to BR center :

Euclidean distance between the centroid of geographic units to down town center of EBRP

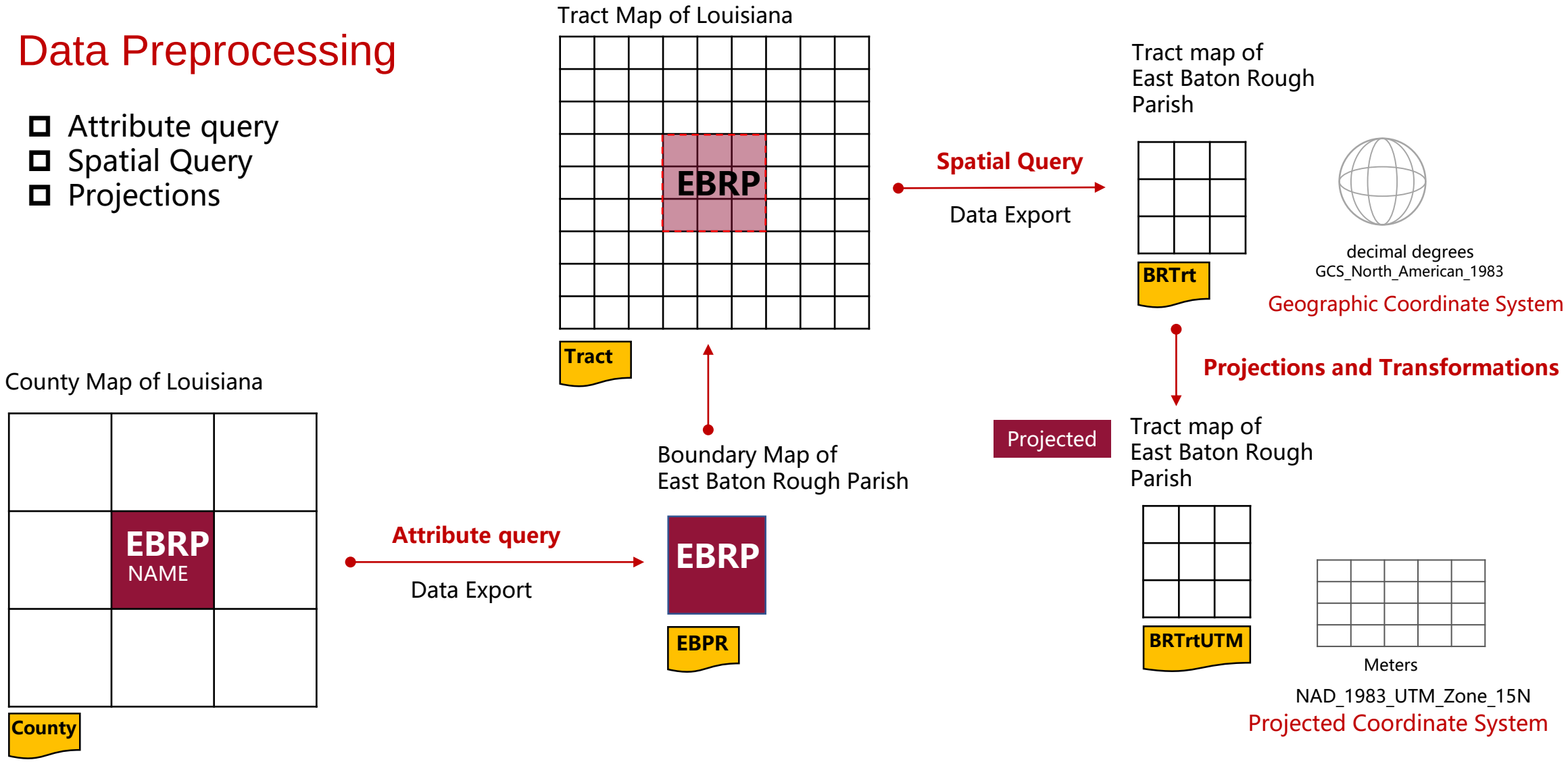
Population density of Rings：

Average density of total population by area of buffer rings

Flow chart I- Data Preprocessing

Data Preprocessing

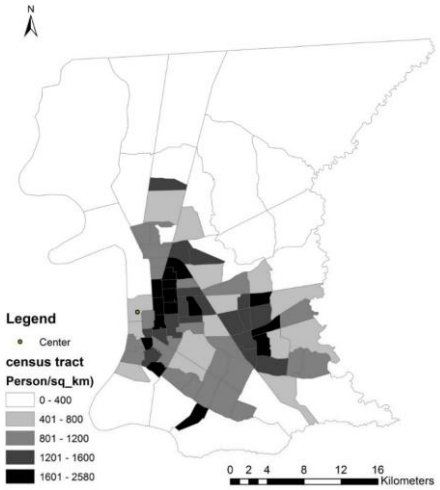
- Attribute query
- Spatial Query
- Projections



Flow chart II- Visualization

Visualization

- Calculate Geometry
- Field Calculator
- Mapping



Projected

Tract map of
East Baton Rough Parish

BRTrtUTM

Calculate Geometry

Square
Meters

Area

BRTrtUTM

Attribute Table

Area

Field Calculator

[DP0010001]/([Area]/1000 000)

Population Density

BRTrtUTM

Attribute Table

PopuDen

Mapping

Choropleth/symbology

Density
Map

Population Density

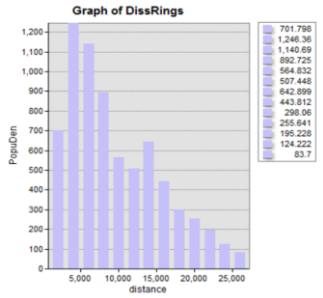
BRTrtUTM

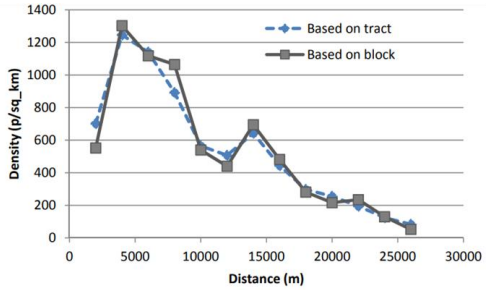
Flow chart III – Spatial Analysis

Alonso Model

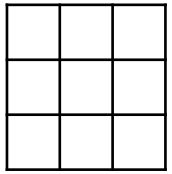
Spatial Overlay

- Multiple Ring Buffer
- Intersect
- Dissolve



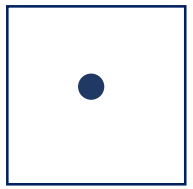


Tract map of
East Baton Rough Parish



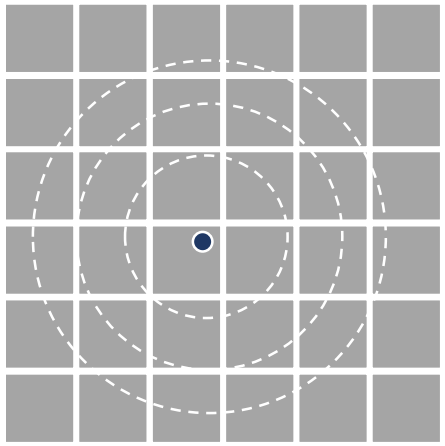
BRTrtUTM

Tract map of
East Baton Rough Parish



BRCenter

Multiple Ring Buffer



Rings

Intersect

TrtRings

Dissolve

Attribute Inheriting

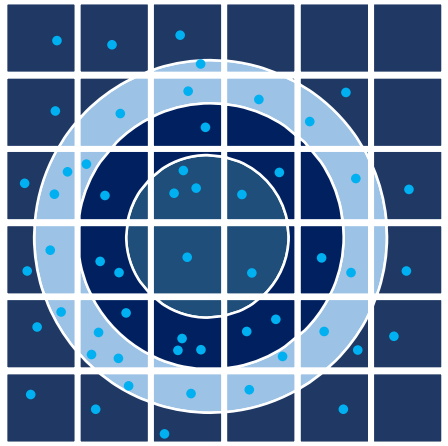
Density
Population

Distance

Tract

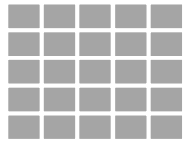
Ring

DissRings



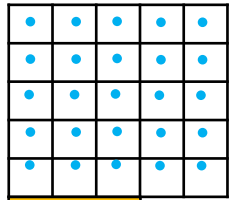
RingBlkP

Spatial Join



BRBlkUtm

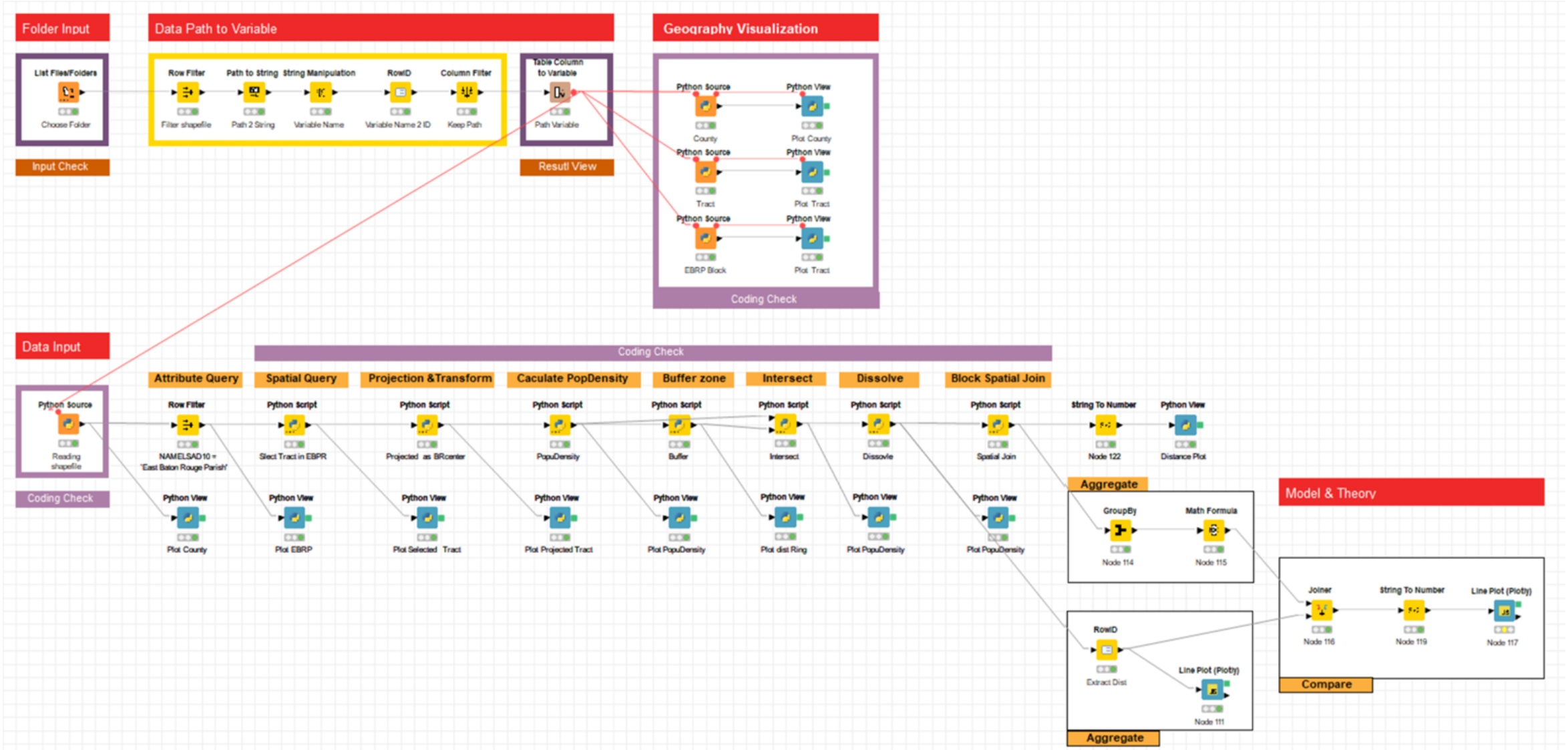
Feature to Point



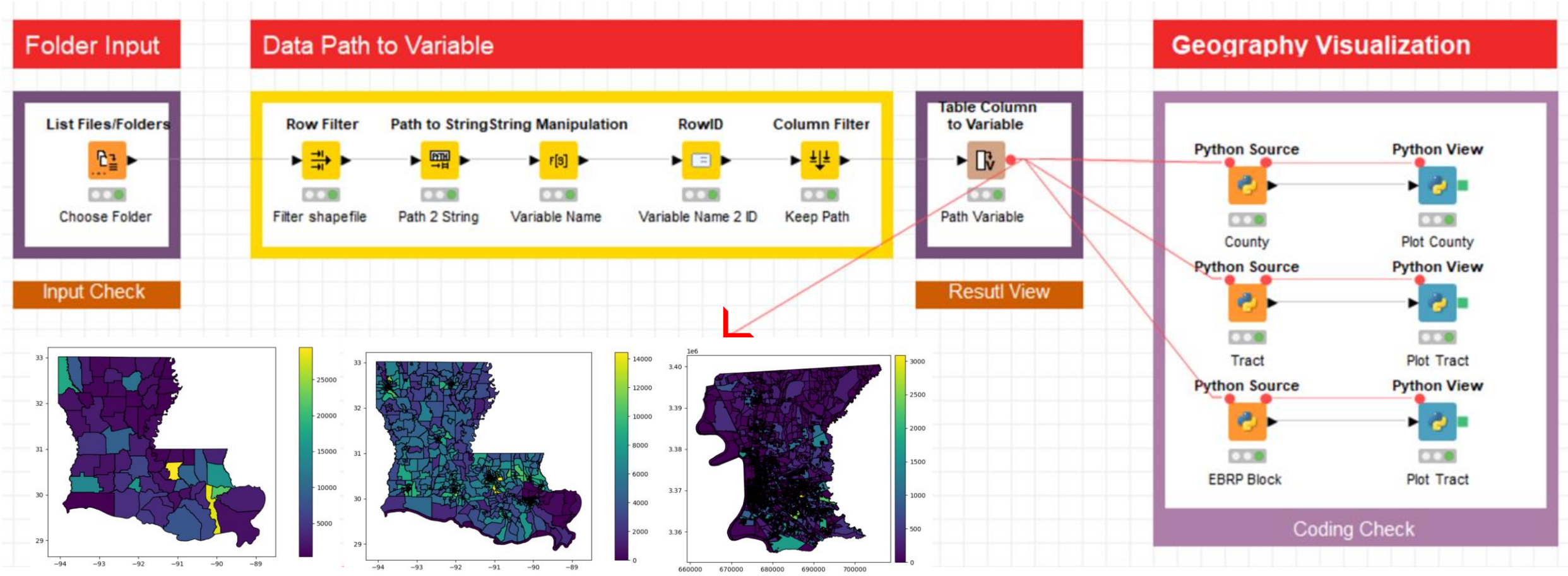
BRBlkPt

MAUP modifiable areal unit problem

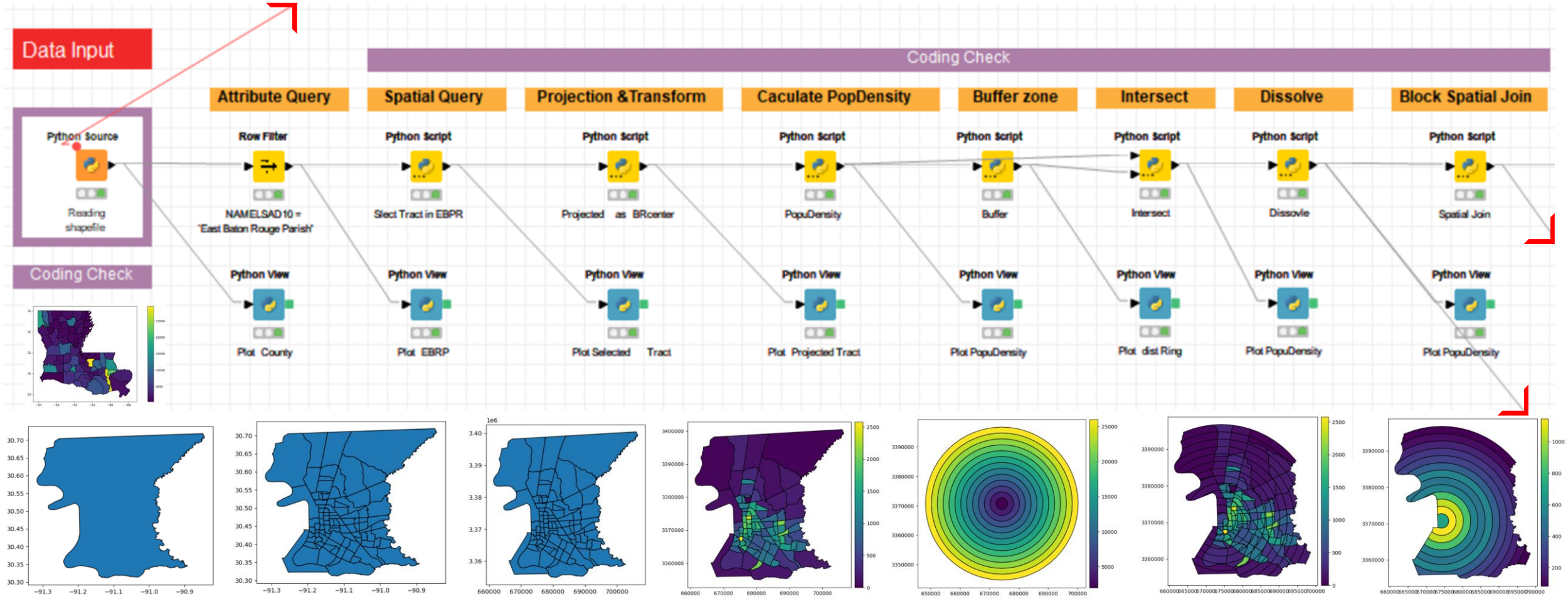
KNIME Workflow Implementation



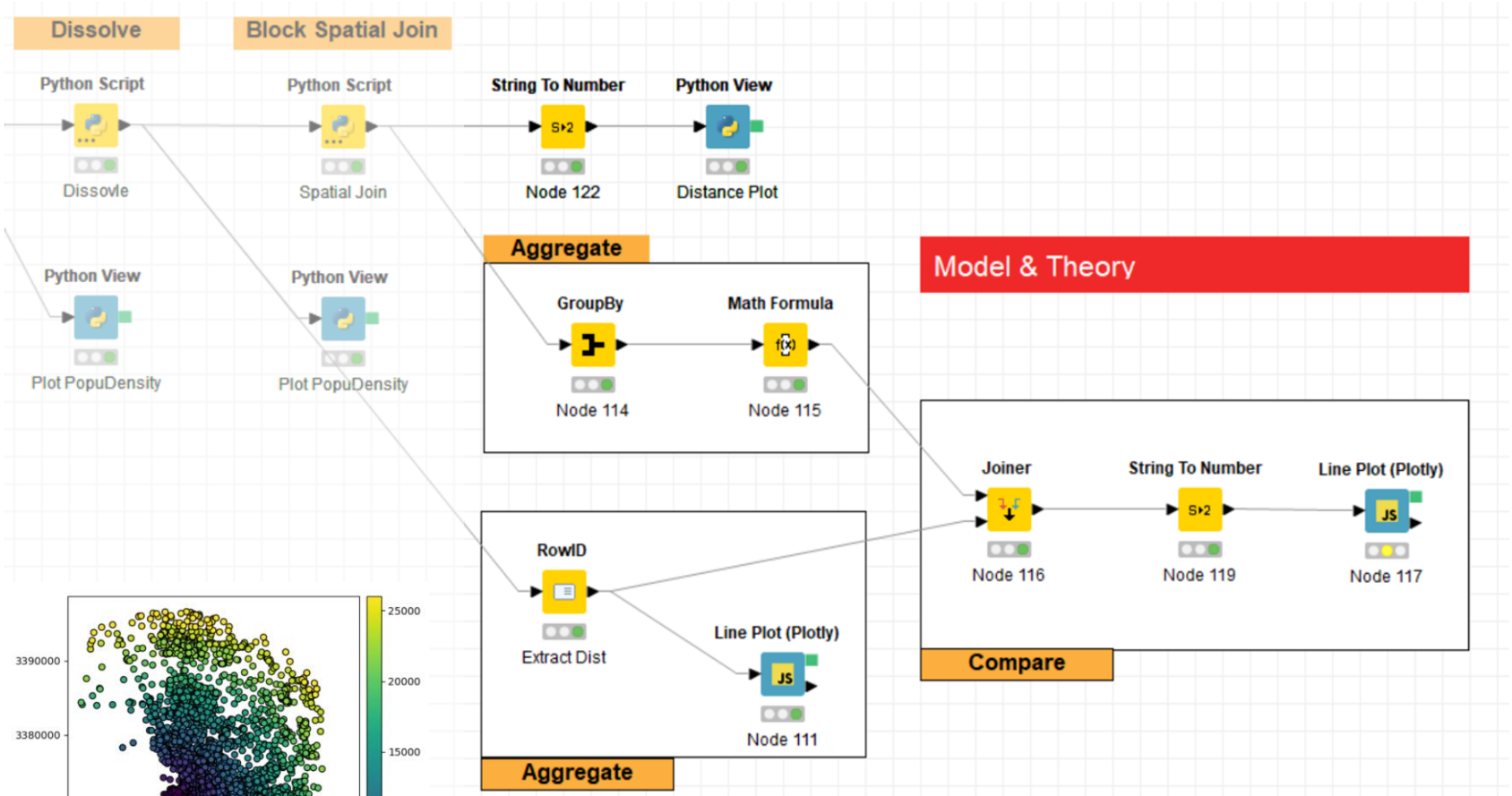
Step by Step Workflow Execution I

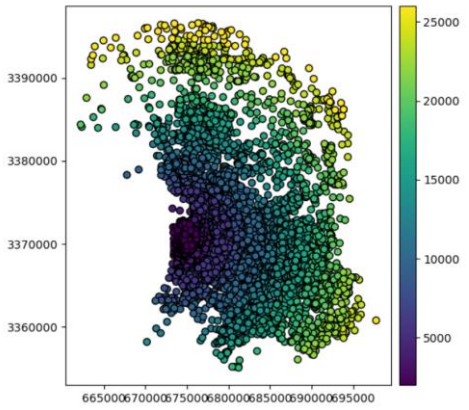


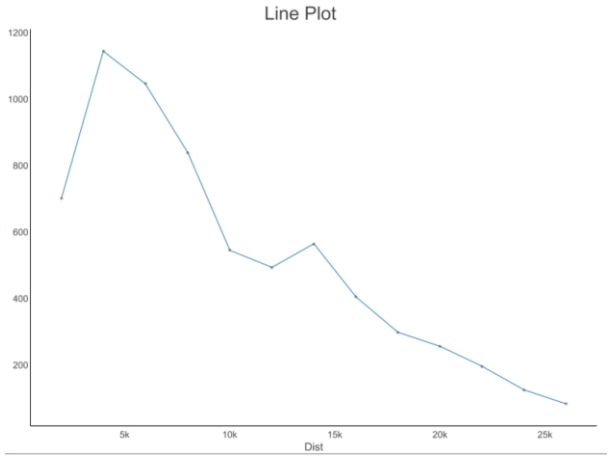
Step by Step Workflow Execution II

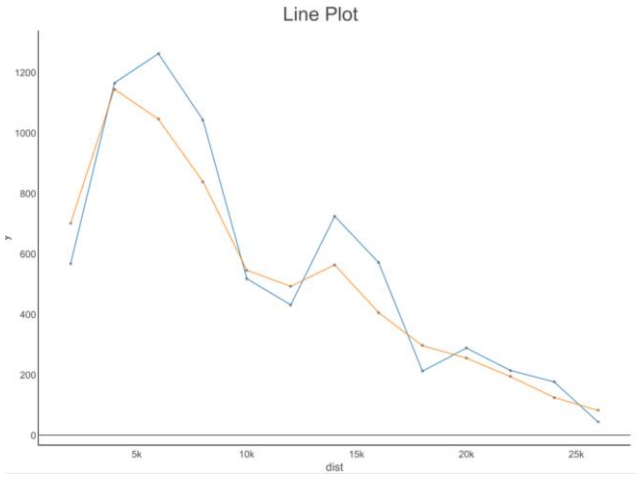


Step by Step Workflow Execution III

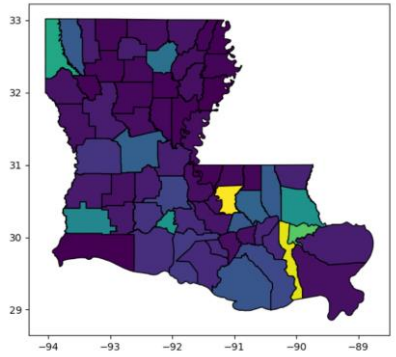




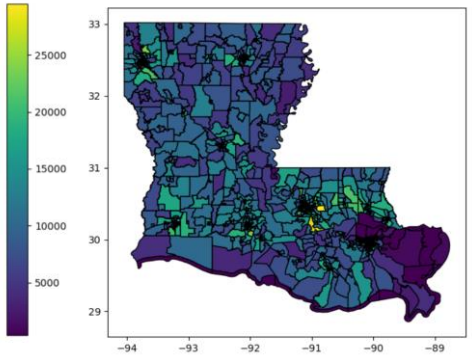




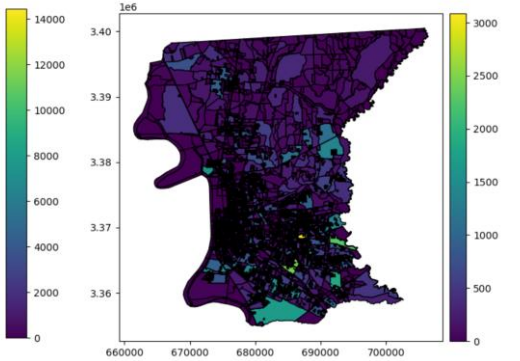
The Outputs from the Workflow



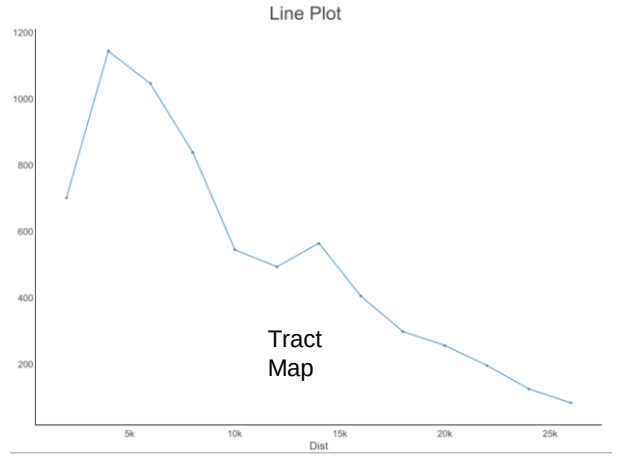
County map



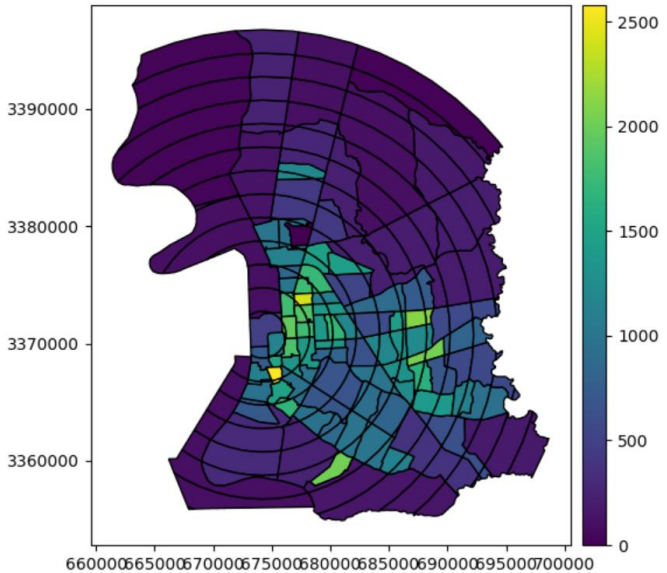
Tract map



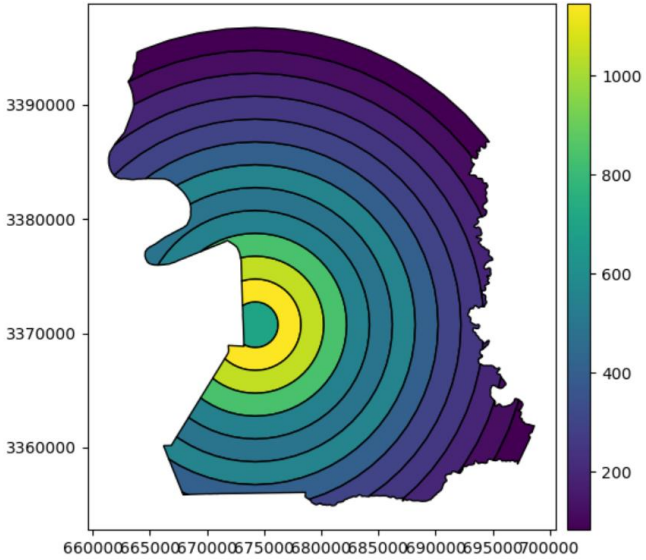
Block map of East Baton Rouge Parish



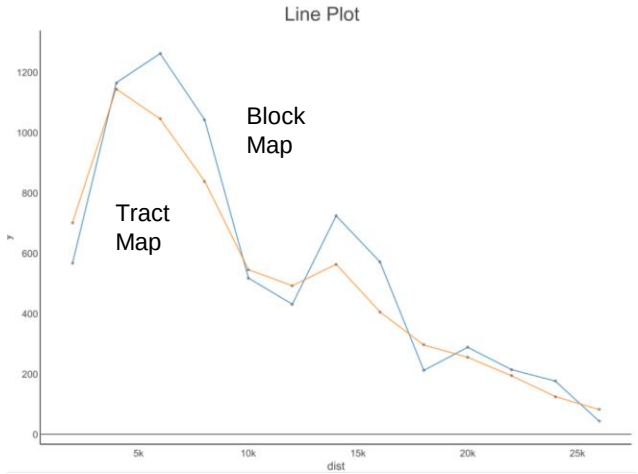
Distance VS Population Density



Multiple Buffer Rings

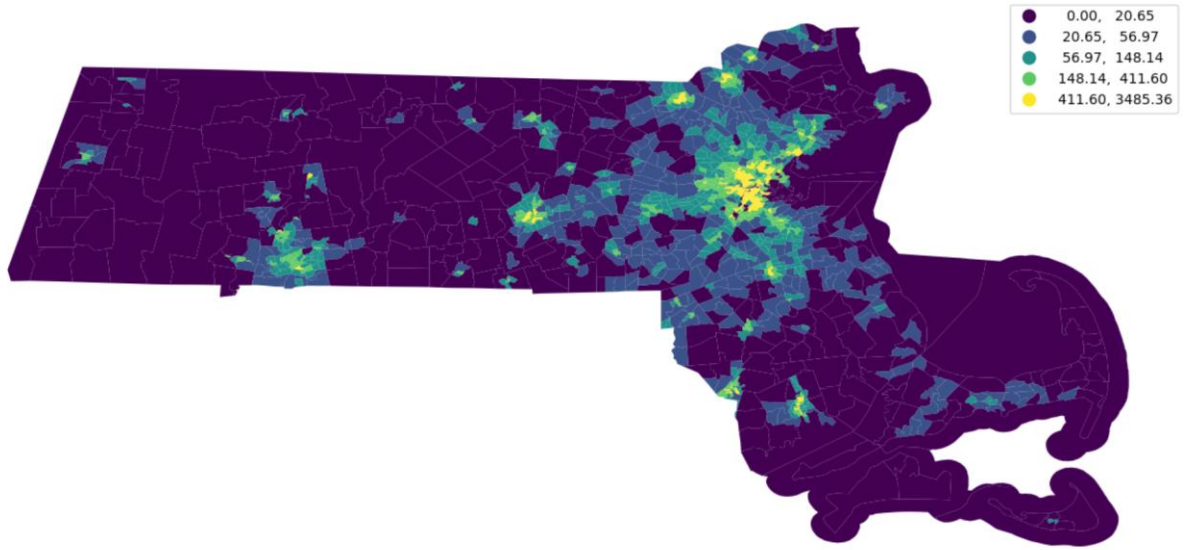


Dissolve



Distance VS Population Density

Discussion & Expansion



Model

Classic Model of Distance vs Population Density

Population Density vs Urban Sprawl

Urban Centrality Model- Monocenter vs Multiple centers

MAUP

How to alleviate the potential influence?

What is difference between 2015 and 2021?

More relevant issue

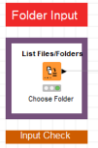
Expansion

Different area, Mechanism

Step by Step Instructions for Running the Workflow

Step 1: **Download data** from Google Drive data folder
Step 2: **Download workflow** from Google Drive workflow folder
Step 3: **Open KNIME** from local PC or China Data Lab Cloud Platform
Step 4: **Import KNIME workflow** file (Chapter1.knwf)
Step 5: **Configure "List files/Folders"** for each table and figure
Step 6: Click **Run** function from the top menu
Step 7: **Display the outputs**:

- **Geographical Visualization :** LA County map; LA Tract; Block of EBRP
- **Data Manipulation :** County; EBRP boundary ;Tract of EBRP; Projected EBRP tract; EBRP tract by Population Density; Multiple buffer zone of BR center ; Intersected map of Tract and Buffer Rings ; Dissolved rings by population density
- **Statistics** : Line plot with distance vs population density



All you need is
to Choose folder containing such files

List of Documents for the Case Study

Data sets	BRBlkUtm.shp; BRcenter.shp; County.shp; State.shp; Tract.shp
Workflow file	Chapter1.knwf
PPT presentation	Chapter1.PPT
Output files	Chapter1_Output (Folder)
Case Study Report file	Chapter1-report.doc
Computing Environment	Knime 4.3.2; Python 3.x Pandas; geopandas; matplotlib; numpy ; shapely

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.

A. STEWART FOTHERINGHAM (1981) SPATIAL STRUCTURE AND DISTANCE-DECAY PARAMETERS, Annals of the Association of American Geographers, 71:3, 425-436, DOI: 10.1111/j.1467-8306.1981.tb01367.x

Fotheringham AS, Wong DWS. The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environment and Planning A: Economy and Space*. 1991;23(7):1025-1044. doi:10.1068/a231025

YingchengLi , XingjianLiu . How did urban polycentricity and dispersion affect economic productivity? A case study of 306 Chinese cities. Landscape and Urban Planning,2018. https://doi.org/10.1016/j.landurbplan.2018.01.007

Spatial Science

Workflow-Based Practices Series I

Workbook for Quantitative Methods and Socio-Economic Applications in GIS

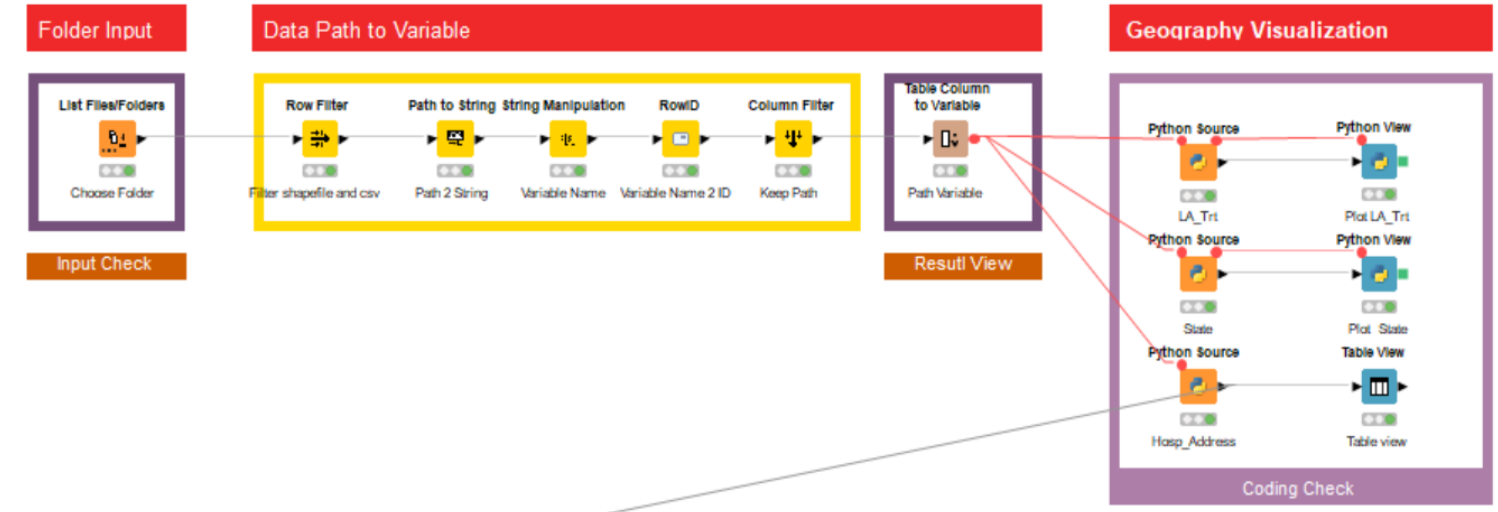
4 Future Work

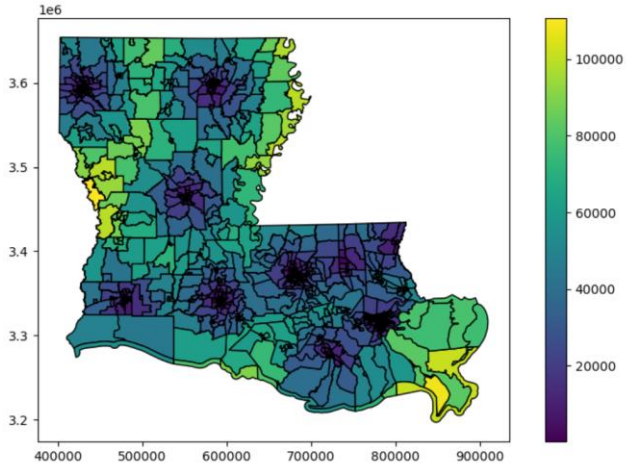
Workflow Transplanting

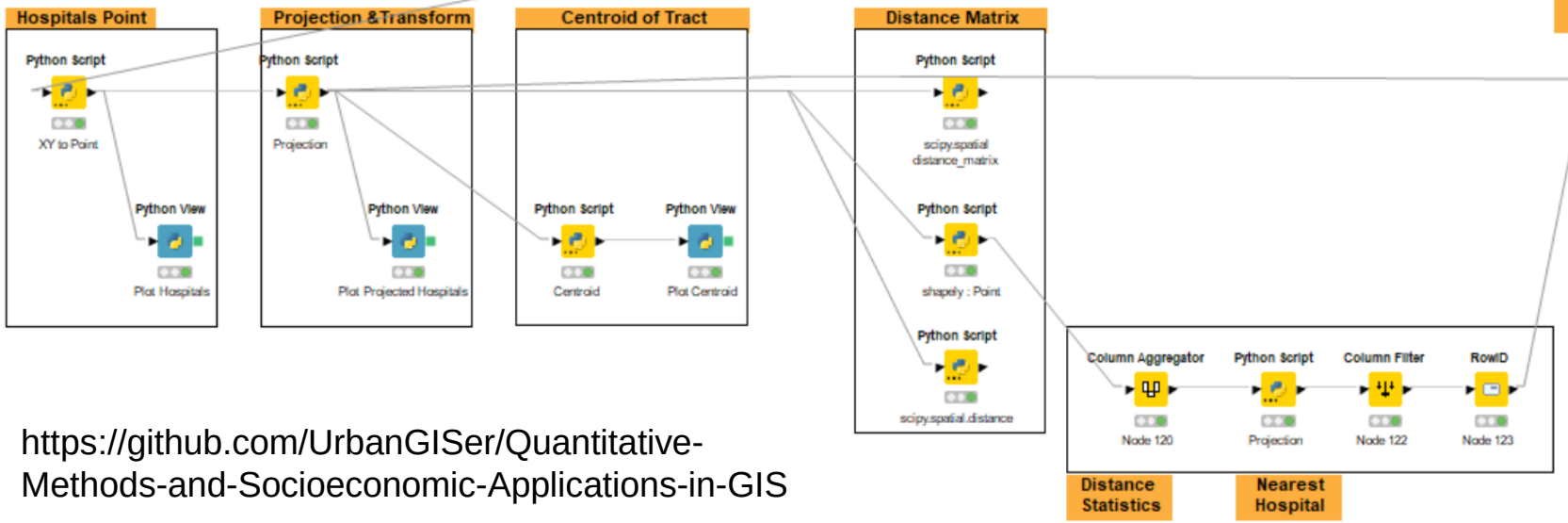
Adjustable | Replicable | GIS Automation

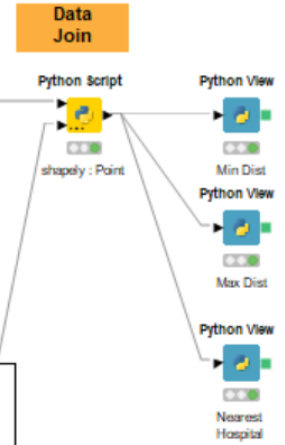
Adjustable Parameters	Buffering zone parameters
Transplantable Automation Functions	Project and transformation Buffer zone Intersect Dissolve Spatial join
KNIME node	Attribute query Data Aggregate Attribute Join Line plot

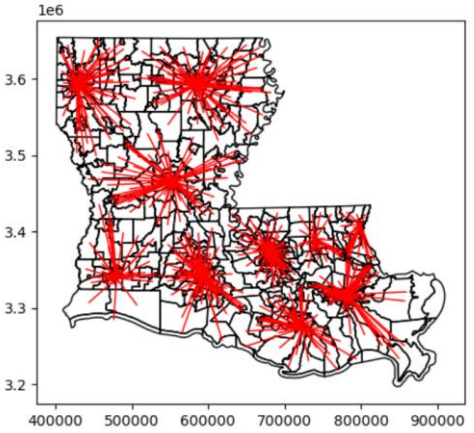
Workflows for All Chapters





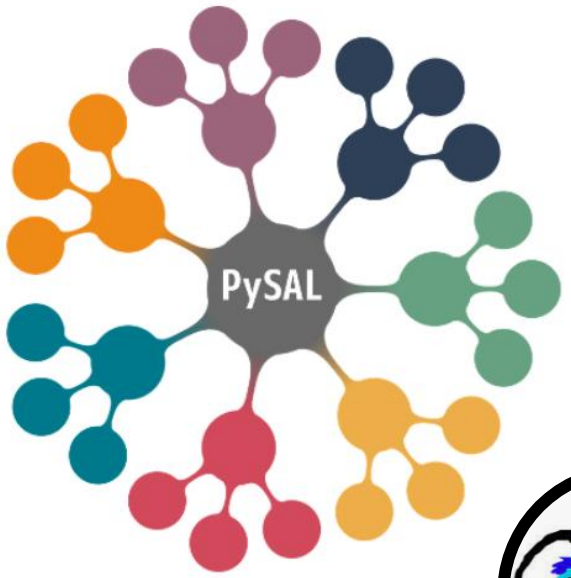






https://github.com/UrbanGISer/Quantitative-Methods-and-Socioeconomic-Applications-in-GIS

Automation GIS



GeoPandas

plotly

The Python Spatial Analysis Library
for open source, cross platform
Geospatial Data Science

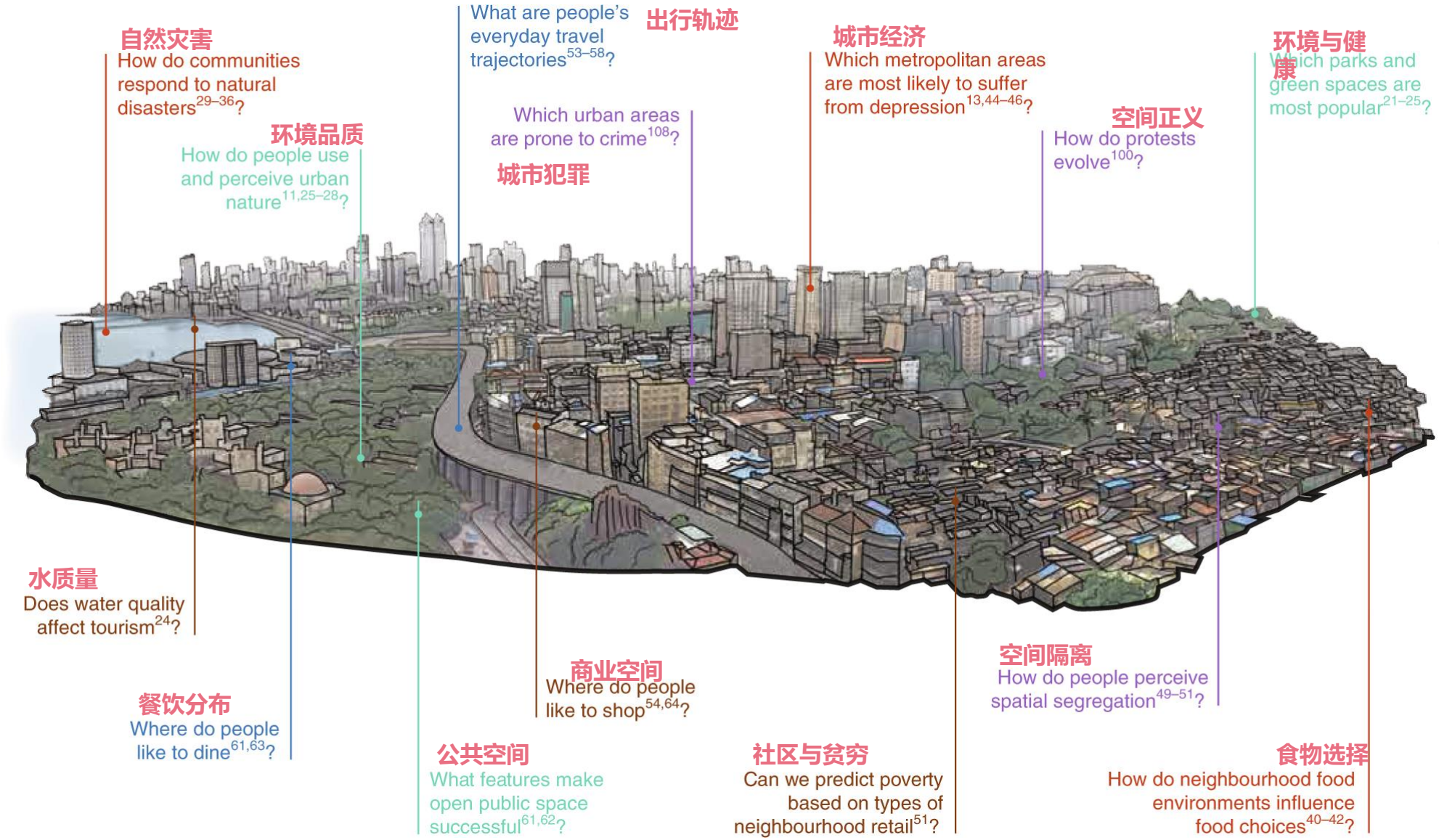
PYGeoDA

matplotlib

SciPy

NetworkX
Network Analysis in Python

Model Integration & Workbench



How to Join the Project?

☐ Requirements:

- Self-motivation on spatial data science and spatial social sciences
- Commitment of 3-6 hours per week on the project
- Participation in regular online meetings on the project
- Timing reports on the project by following the schedule

☐ Tasks:

- Reading the materials for the project
- Preparing the data for case studies
- Testing (or developing) workflows for case studies
- Writing the case study report
- Drafting the PPT for project presentation

☐ Application:

- Send a copy of CV with photo to spatialdatalab@lists.fas.harvard.edu
- Join an online interview

Website and Contact

Project website:

http://spatialdatalab.org

Contact:

spatialdatalab@lists.fas.harvard.edu

THANKS

Spatial Social Science

Workflow-Based Practices Series I

Workbook for Quantitative Methods and Socio-Economic Applications in GIS

Fahui Wang | Shuming Bao

Lingbo Liu

Spatial Data Lab
CGA @ Harvard University
Louisiana State University
Wuhan University