Data Science for Urban Scientists
Fall 2025

Instructor

Wookjae Yang
Assistant Professor of Urban Policy and Administration
Email/office: wookjae.yang@inu.ac.kr / Building 28, Room 413

Course Description

Outline/Purpose:

This course introduces students to data science techniques for urban spatial and big data analysis. In this class, we will focus on three parts: urban big data, statistical analysis, and machine learning. This course will teach students how to harness the power of big data by mastering the way they are collected, organized, and analyzed to support better decision-making in an urban planning context. Students will learn the basic tools needed to manipulate large datasets derived from various open-data platforms, from data collection to storage and approaches to analysis. In addition, students will learn how to use various data analytic tools R to clean, wrangle, analyze, and visualize data. The course will also expose students to statistical programming with R and introduce them to basic machine learning and deep learning techniques.

Course Learning Outcomes:

The course aims to develop students' fundamental knowledge and skills in urban data science using R programming. After successful completion of this course, students will be able to:

(1) Understand and learn how to process urban spatial and big data
(2) Describe important theories and concepts in the analysis and modeling of urban spatial big data
(3) Understand the role of AI and machine learning in urban data analytics
(4) Critically review real-world applications of urban data science
(5) The ability to deal with R programming is not a prerequisite for this class. Beyond the content covered in the class, the instructor recommends that students put more effort into becoming familiar with R coding. Students do not need to memorize all the R coding syntax, but they should understand and memorize the key R coding syntax that is repeatedly used in the class.

Class Delivery Method: PPT lecture slides + Lab practice documents
Time: Monday, Wednesday (12:00 – 13:30)
Classroom: Building 28, Room 406
Credit: 3 credits
Language: English
Grading:

- This class follows absolute evaluation.
- A+: 95 and above, A0: 90 – 94, B+: 85 – 89, B0: 80 – 84, C+: 75 – 79, C0: 70 – 74, D+: 65 – 69, D0: 60 – 64, F: Below 60

	Exam	Attendance	Assignment
Percentage of grade evaluation	40%	20%	40%

- Exam (40%): Three or four students will be grouped to work on a term project. For the mid-term exam, students will present their project proposal. In the final exam, students will present final team projects that are going to be prepared throughout the semester. There is no written tests in this class. Please stay tuned, assignment guidelines will be uploaded to the LMS
- Attendance (20%): Out of a maximum of 20 points (학칙시행세칙 제 56 조 제 2 항) → For general subjects (3 credits), 1/3 point will be deducted for every 1 hour of absence → 1 point will be deducted for 3 hours of absence. Students who miss more than one-third of the actual class hours or engage in academic dishonesty will not be granted course credits, regardless of their exam scores or other grades (학생시행세칙 제 56 조 제 3 항).
- Assignment (40%): Students will submit weekly lab assignments. Every assignment is due before every Thursday class (due date will be indicated in the LMS) and should be submitted on LMS under assignments. Labs are intended to provide you with hands-on experience in land use planning and practices. Lab documents will be provided at the start of each lab session and students will work on the exercises in-class with help of the instructor. Extra credits will be granted when submitting extra lab sessions

Course schedule:

Week 01. Introduction to Urban Data Science

- Basics of urban data science
- What urban data science deals with – urban issues
- *Lab #1: Basics of R programming*
  - Installing R and RStudio
  - The R console
  - R scripts and comments
  - R studio IDE (integrated development environment)

Week 02. R Coding Basics

- Libraries in R
- R coding basics
- Data types and structure in R
- *Lab #2: R coding basics practice*
  - *dplyr* package in R

Week 03. Urban Big Data and spatial data

- Definition of Urban Big Data
- Big data in urban planning
- Examples of Urban data science
- Good project topics in urban analytics using R
- *Lab #3: Urban data processing*
  - Importing/exporting data
  - Spatial data, coordinate reference systems and transforming CRS
  - Mapping Data

Week 04. Urban Data Visualization

- Creating graphs (Univariate/ Bivariate/ Multivariate graphs)
- Spatial data mapping
- Statistical models visualization (bar chart, scatter plot, bubble chart, box plot, heatmap, and etc)
- *Lab #4: Data visualization in R*
    - Reference: Kabacoff, R. (2024). Modern Data Visualization with R

Week 05. Data cleaning

- Data cleaning definition
- Importance of data cleaning
- Data cleaning techniques
- *Lab #5. Data cleaning with R*
    - Dealing with missing data
    - Strings and recoding variables
    - References: Edwin de Jonge and Markvander Loo (2013). An introduction to data cleaning with R

Week 06. No classes on this week (* Chu-Seok holiday)

Week 07. Data wrangling

- Data wrangling definition
- Importance of data wrangling
- Data wrangling process
- *Lab #6. Data wrangling with R*
    - Subsetting data
    - Data classes
    - Reshaping data (wide <-> long)
    - Spatial data wrangling
    - References: Boehmke, B. C., (2016). Data Wrangling with R. In Use R

Week 08. Mid-term

- Term project proposal presentation
- Submit a project proposal as per the instructions in the file
- The proposal will be documented in one or two pages, including the title, background, literature gap, objective, data collection plan, methods, results, and what to expect from the results

Week 09. Exploratory data analysis (EDA)

- What is EDA?
- Types of EDA: non-graphical vs graphical / univariate vs multivariate
- Summary statistics: central tendency, dispersion, outliers
- Correlation analysis: identifying relationships between variables using correlation matrices and scatterplots
- Visualization tools: histograms, boxplots, pair plots, correlation heatmaps
- *Lab #7. EDA with R*

Week 10. Inferential Statistics

- Fisher's exact test, Chi-square test, t-test, Analysis of variance test
- Correlation test

- *Lab #8. Inferential statistics with R*

Week 11. Statistical Analysis 1

- Linear Regression
- Logistic Regression
- Poisson/Negative binomial Regression
- Multinomial Regression
- *Lab #9. Statistical analysis1 with R*

Week 12. Statistical Analysis 2

- Linear Regression
- Logistic Regression
- Poisson/Negative binomial Regression
- Multinomial Regression
- *Lab #10. Statistical analysis2 with R*

Week 13. Spatial Analysis

- Spatial Error Model (SEM)
- Spatial Lag Model (SLM)
- Geographically Weighted Regression (GWR)
- Model comparison and interpretation of diagnostics

Week 14. Basic machine learning

- What is machine learning? Supervised vs. unsupervised learning
- Key concepts: features, labels, training/testing split, overfitting
- Overview of algorithms:
    - Classification: k-Nearest Neighbors (kNN), Decision Trees
    - Regression: Linear regression, Random Forest regression
    - Clustering**:** k-Means clustering (optional, if time permits)
- Model evaluation: accuracy, confusion matrix, RMSE

Week 15. Advanced machine learning (AI, deep learning)

- Neural networks
- Deep learning

Week 16. Term project presentation