

MSc Urban Spatial Science

Data Science for Spatial Systems Course Outline



Huanfa Chen



Welcome Everyone

- In this course you're going to learn how to:
 - Applied Machine Learning
 - Casual Inference
 - Applications in spatial/urban domains
- In the following Weeks you'll be doing:
 - Python Programming | Data Handling | Data Analysis | Machine learning

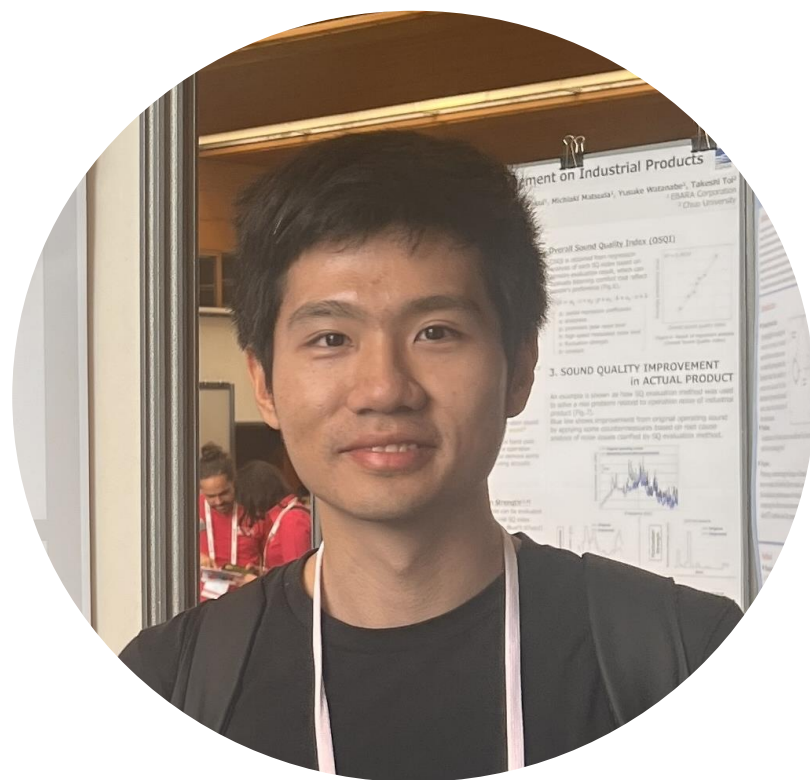
The Teaching Team



Huanfa Chen



Esra Suel



Xiang Fang



Zhixuan Liu



Xinglei Wang



Shitian Zhang

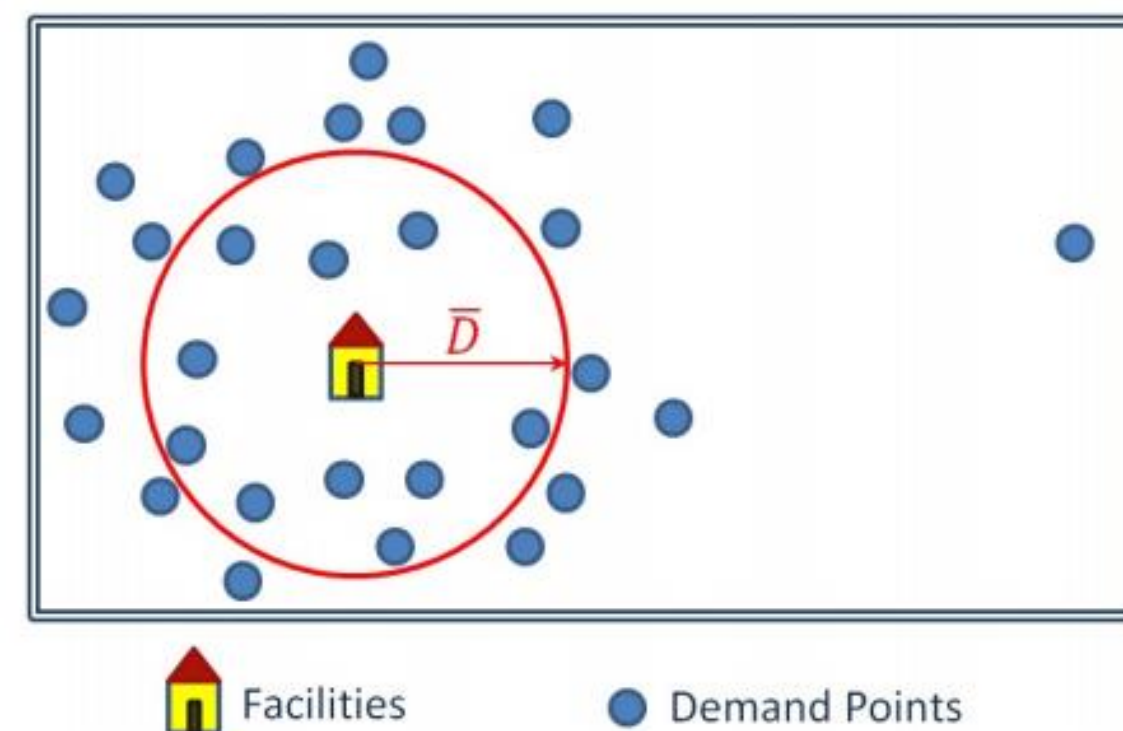
The Teaching Team

Associate Professor in Spatial Data Science

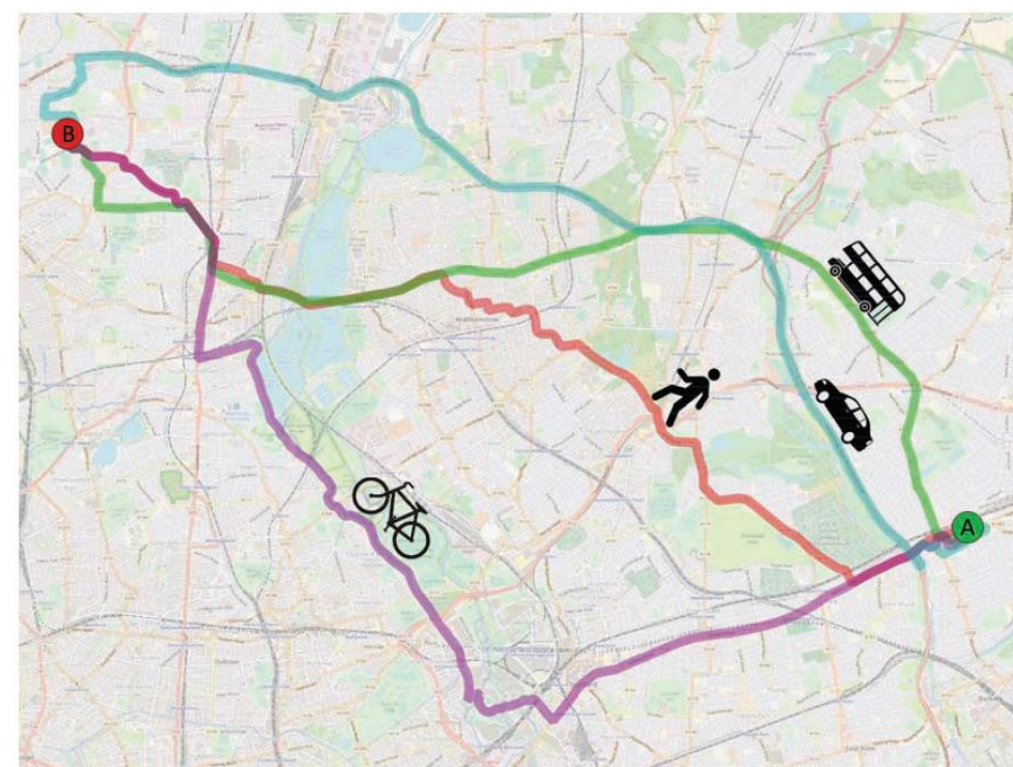
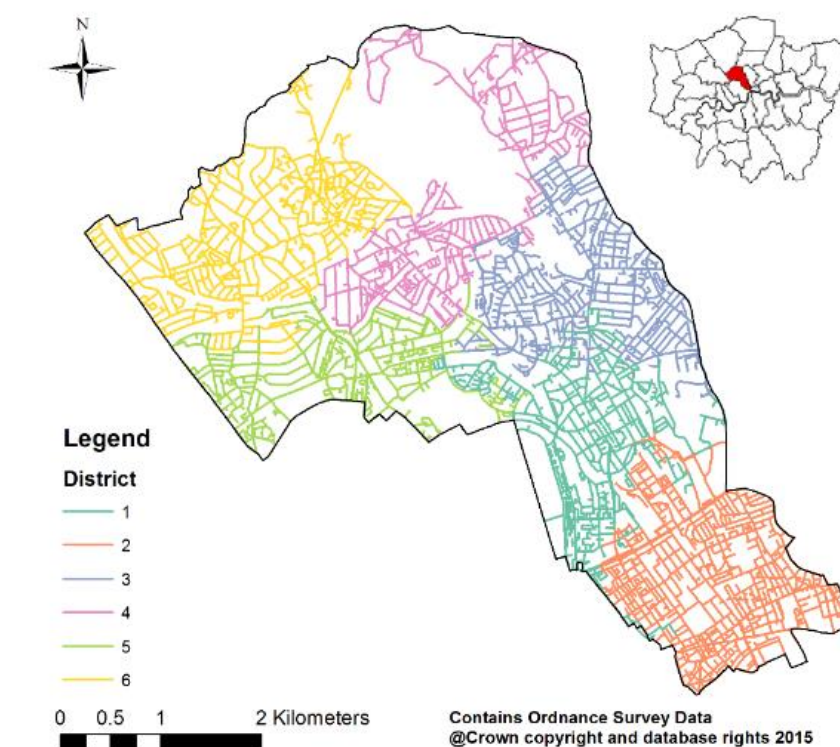


Huanfa Chen

Facility location selection



Police districting problem (spatial clustering)



Machine learning for transport

- Travel mode prediction (how people travel)
- Travel purpose estimation (why people travel)

The Teaching Team



Esra Suel

Associate Professor in City Modelling

Background: Transportation (PhD) & Urban Health (Postdoc)

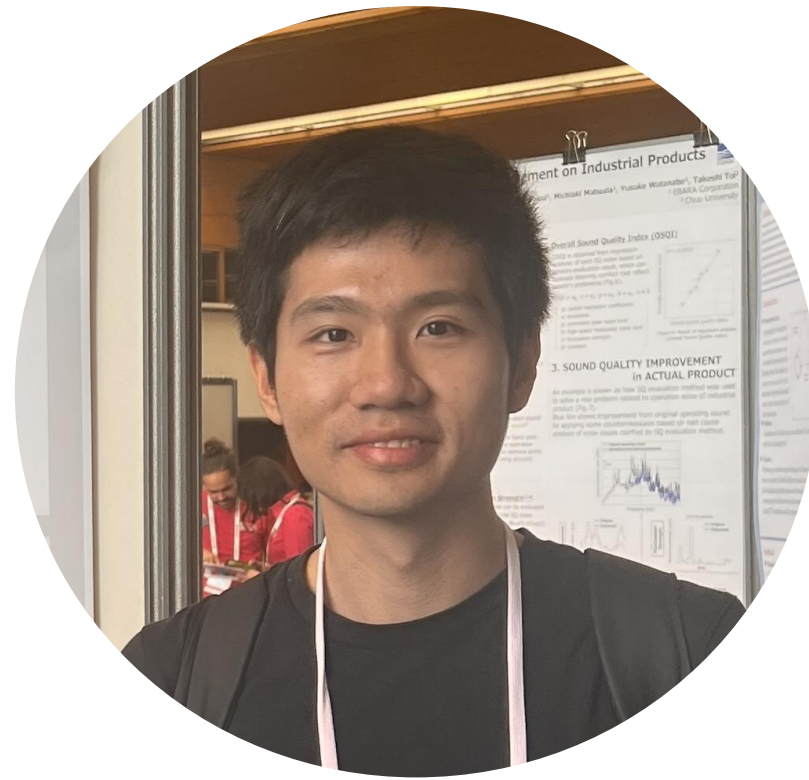
CASA Arup City Modelling Lab with Adam Dennett and Gerry Casey

Research themes: mobility, energy and transport, housing, inequalities.

Relevant active projects:

- Exposure metrics from street images using supervised deep learning algorithms for investigating associations with health outcomes
- Using street and satellite images for urban change detection using unsupervised learning

The Teaching Team



Xiang Fang

I am a PhD candidate in the Acoustics and Soundscape group at the UCL Institute for Environmental Design and Engineering at the Bartlett.

Background: Architecture (undergraduate and master's degrees)

Research interests:

- Soundscape prediction and modelling
- Human perception of sound and its interaction with urban space
- Acoustic metamaterials for soundscape design

Contact: xiang.fang.22@ucl.ac.uk

The Teaching Team



Zhixuan (Willow) Liu

PhD candidate in CASA

Doctoral researcher in Department for transport UK (Dft)

Research Interest:

- Economic impacts of transport infrastructure
- Agglomeration impacts
- Land Use and Transport Integrated model
- Quantitative spatial economic model

Contact: zhixuan.liu21@ucl.ac.uk

The Teaching Team



Xinglei Wang

 [xlwang233.github.io](https://github.com/xlwang233)

 xinglei.wang.21@ucl.ac.uk

PhD student at SpaceTimeLab at Dept of Civil, Environmental and Geomatic Engineering

Backgrounds: Geomatics Engineering (Undergraduate); Geographical Information Engineering (Master's); Geoinformatics (PhD)

Research interests: human mobility, urban analytics, GeoAI

Projects:

- London high street footfall time series analysis
- Human mobility prediction using large language models
- Urban space representation learning from POIs
- Enhanced location embeddings for user next location prediction

The Teaching Team



Shitian Zhang

PhD candidate in CASA

Background:

Urban Planning (Undergraduates); Space Syntax (Postgraduates)

Research Interests: Spatial Optimisation

Projects:

- Spatial layout simulation and optimisation for healthcare spaces
- Web-based decision support system for wind farm layout optimisation

Contact: shitian.zhang.22@ucl.ac.uk

Moodle and Slack

- Moodle enroll key for CASA0006: **CASA2023**
- Slack: Join the Slack space at casa-students-2023-24.slack.com and the channel of **#0006-data-science-spatial-systems**.
- Please log into Slack using your UCL email address.

Week by week

What's going to happen when

- 10-week module
 - 15 credits course
- 1.5-hour lectures + 1.5-hour workshops
 - Workshop: quiz + Python notebook
- Individual Assessment
 - Reproducible Python notebooks
- Keep checking Moodle and Slack
 - Heavy forum use to help you all work together

Lectures & Workshops

Where are we going to be

Venue: IOE - Bedford Way (20) W3.01

Time: 9:00 – 12:00, Wednesday

Please inform CASA-teaching@ucl.ac.uk if you can't attend a lecture or workshop.

Check the Moodle for more details.

1 Introduction to Module

2 Supervised Machine Learning

3 Tree-based Methods

4 Analysis Workflow

5 Artificial Neural Networks

6 Panel Regression

7 Dimensionality Reduction

8 Spatial Clustering

9 Difference in Difference

10 Regression Discontinuity

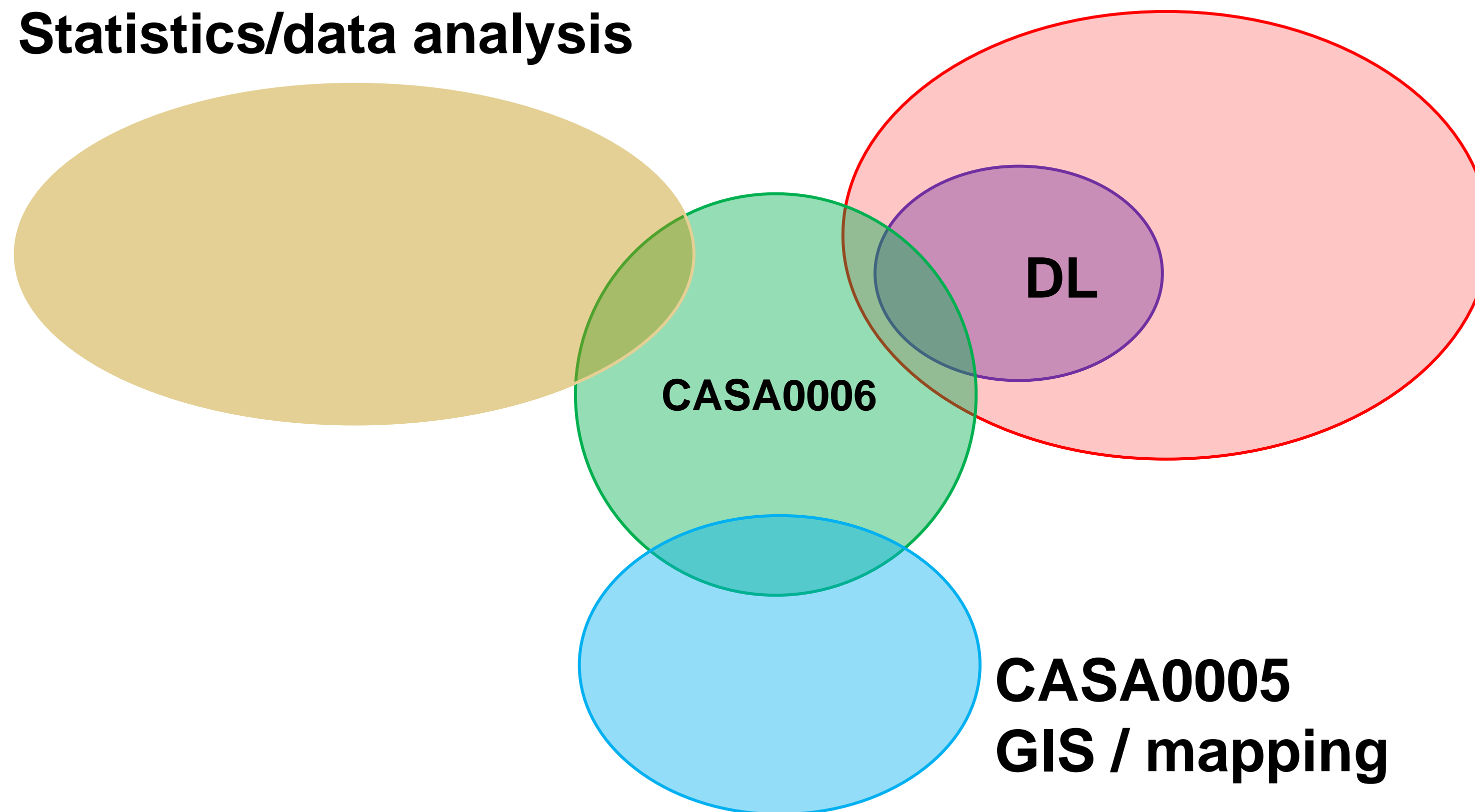
Why are we teaching this?

CASA0007

CASA0013

Statistics/data analysis

Machine Learning



- Building on techniques from CASA0007/CASA0013/CASA0005
- Extracting knowledge and insights from spatial data

Pre-requisites

- Working knowledge of Python, Jupyter, Python notebooks, Markdown, Docker, Command line in Linux/Unix based OS, github, file formats
- A general understanding of statistics, linear regression
- A general understanding of GIS files, projection, UK census units
- No previous knowledge with machine learning is required

Course philosophy

- This is a practical and hands-on course.
- We will cover basic concepts and background theory relating to ML and causal inference, but not in great mathematical depth or rigor.
- Mainly focus on implementing and running ML algorithms with examples in a spatial context, discussing the implications
- Using the variables and algorithms in the right way

An aerial photograph of an airport tarmac, showing several aircraft parked at gates, ground service equipment, and various airport infrastructure. The image is overlaid with a semi-transparent dark grey layer containing the title text.

DSSS Assessment

Technical Analysis and Visualisation Report

Individual Assessment

Module Assessment

- 100% mark from coursework
 - 100% - From Analysis / Python Notebook
- Individual Assessment
 - Working by yourself
- Max 2,000 word report with in-built analysis in form of Python Notebook
 - Deadline: 24 April 2024 at 5pm

An aerial photograph of an airport tarmac and runways, showing several aircraft parked at gates and on the tarmac. The image is darkened to serve as a background for the text.

Examples of Work

What we want you to strive to create

About this document

This is the assignment for the UCL CASA module Data Science for Spatial Systems (CASA0006)

This is a self contained Jupyter notebook with structure introduction, discussion and embeded code and bibliography.

Code is hosted [here](#)

Data is availble [here](#)

Multi-level segmentation of electricity consumption data:

the case of the Netherlands

This paper explores data from Dutch electricity providers and establishes provider-level traits based on clustering of cities and streets they supply to.

It highlights the varying customer and area profiles that each serves and serves to identify potential priorities

| [1.0 Intro](#) | [2.0 Lit. review](#) | [3.0 Method](#) | [4.0 Data](#) | [5.0 Explore](#) | [6.0 Analysis](#) | [7.0 Discussion and conclusion](#) | [Bibliography](#) | [Appendix](#) |

https://github.com/antoniosfiala/Electricity_clustering

7.0 | Discussion and conclusion

In conclusion, this paper has carried clustering of electricity usage data across the Netherlands with the aim of answering the following research questions with each addressed below:

- RQ1 | What segments of energy use can be observed in the Dutch market?
- RQ2 | Does the data suggest a relationship between the supplier and the segment it supplies?
- RQ3 | How do clusters vary across geographical scales?

RQ3: Some variation can be observed across the levels, with clustering at city level yielding three clusters whereas street and postcode level analysis suggests four is the appropriate number.

RQ1: Along the measures available, categories were observed that were chiefly defined by variation in smart-meter and low-tariff use that appear to correspond to the supplier they fall under.

RQ2: Building on RQ1 it appears that there may be a relationship between the supplier and the extent to which smart meters are used.

The implication of this is that both at institutional, governmental level and supplier level, there may be room to improve targetting to transition the remaining areas to smart metering.

https://github.com/antoniosfiala/Electricity_clustering

General Recap

- Going to be challenging but worthwhile
 - . You'll learn real-world analytical skills and how to pull components together to make data shine.
- Learn how to work on a data science project
 - . Skills that will be really useful in the coming years
- Present your work
 - . Tell a story with your data