

02.522: Urban Data & Methods I:  
Computational Urban Analysis

*Ate Poorthuis*



# Contents

<b>1</b>	<b>Syllabus</b>	<b>5</b>
1.1	Class Schedule . . . . .	5
1.2	Course Description . . . . .	5
1.3	Format . . . . .	5
1.4	Expectations . . . . .	6
1.5	Assessment . . . . .	6
1.6	Deadlines . . . . .	7
1.7	Software . . . . .	7
1.8	Detailed Outline . . . . .	7



# Chapter 1

## Syllabus

### 1.1 Class Schedule

**January Term 2020**

**Tue 12.30-2pm** Think Thank 7 (1.409)

**Thu 11.00-12.30pm** Think Thank 17 (2.202)

Office Hours: By appointment

### 1.2 Course Description

This course focuses on quantitative and computational approaches to urban analytics and data science. It exposes students to new ways of collecting large datasets (“big data”) and innovative methods of analysing such datasets. It draws on both more conventional methods such as (spatial) statistics, as well as how to appropriately use methods from data science and machine learning within an urban context.

### 1.3 Format

The course is structured around two 1.5 hour classes per week that integrate lecture, discussion and in-class activities and exercises in an interactive manner. The class is further structured around 6 blocks (each lasting two weeks).

## 1.4 Expectations

Students are expected to be present and actively participate in each class, as well as on the class online forum (Slack). Before coming to class, you will have read the assigned readings and you are coming to class prepared to participate in discussion and exercises.

You are also expected to produce your own work, whether individually or in groups. Do not copy work from the internet or other published sources without proper citations. This is plagiarism and if a student is found to be doing so, he or she will be subject to disciplinary measures including potentially failing the course.

## 1.5 Assessment

There will be a variety of assessments throughout the semester. Emphasis is on your performance overall, with relatively low weight placed on individual items. Continued participation throughout the semester will enable you to do well in this course.

Assessment Items	Percentage	Period
Class Participation	15%	Throughout term
Assignments (six)	45%	Throughout term
Final Project	40%	Week 14

### 1.5.1 Assignments

Each block consists of a series of exercises that culminate in an assignment/report that will be due before the start of the next block (Monday 23.59).

### 1.5.2 Final Project

For the final project, you will develop conduct a computational analysis of a topic and dataset of your own choice. This is an opportunity to explore (an aspect of) your final term project in more detail – or take a deep-dive in a topic or dataset that you’re interested in. You can choose to use already existing datasets, or collect your own but you must use one or more computational methods to help answer your research question. In Week 6, you will hand in a research proposal of a maximum of 1000 words. It should discuss your motivation for the project; its objectives and research question; data requirements (does the data already exist? where do you get it from? does it need a lot of

post-processing); and the methods you plan on utilizing. The remainder of the term is spent on collecting and analyzing data. A prototype of your analysis is due at the end of Week 13 and the final version at the end of Week 14. You are required to hand in both the code, a written paper and (where appropriate) visualizations as part of your Github repository. You can do this in a single form factor through one or more RMarkdown documents.

## 1.6 Deadlines

Deadlines are as noted in the course syllabus or on the specific assignment. If something is due on a specific date, you have until midnight on that day to submit the assignment.

## 1.7 Software

We will use R, RStudio and a series of packages (most of them compatible with the Tidyverse). All software used in the course is open-source and freely available.

## 1.8 Detailed Outline

### 1.8.1 Block 1: Analysis of HDB Resale Prices I

1. Reproducible Science & Project Management (+ re-view of tidyverse basics)
  - Reading:
    - Excuse me, do you have a moment to talk about version control? **(done)**
    - Happy Git with R (Chapter 1-15) **(done)**
    - Reproducible research and quantitative geography **(done)**
    - Pro Git (optional)
2. Exploratory Data Analysis (Univariate Statistics & Visualization)
  - Reading:
    - Happy Git with R (Chapter 20-23) **(done)**
    - Data Visualization Chapter 3
    - Burt et al., *Elementary Statistics for Geographers*, (Chapter 3)

### 1.8.2 Block 2: Analysis of HDB Resale Prices II

#### 3. Sampling, Bootstrapping & Confidence Intervals

- Reading:
  - Burt et al., *Elementary Statistics for Geographers*, (Chapter 6)
  - Modern Dive Chapter 8
  - Modern Dive Chapter 9

#### 4. Correlation & Linear Regression

- Reading:
  - Burt et al., *Elementary Statistics for Geographers*, (Chapter 4)
  - Modern Dive Chapter 6

### 1.8.3 Block 3: Geodemographics of SG Neighborhoods I

#### 5. Dimension Reduction (MDS/PCA)

- Reading:
  - Gatrell, Anthony C. “Multidimensional Scaling.” In *Quantitative Geography*. Routledge & Kegan Paul, 1981.
  - Principal Component Analysis, Explained Visually
  - Wylly’s notes on PCA

#### 6. Dimension Reduction (T-SNE)

- Reading:
  - Van der Maaten’s Google Tech Talk

### 7. Break

### 1.8.4 Block 4: Geodemographic of SG Neighborhoods II

#### 8. Clustering (non-spatial)

- Reading:
  - Fortunato, F. Community Detection in Graphs. Specifically section 4a-c.
  - Jain, A. Data clustering: 50 years beyond K-means
  - Ester et al. A Density-Based Algorithm for Discovering Clusters

#### 9. Clustering (spatial, including refresher on spatial data structures)

- Reading:
  - Guo, D. Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP)
  - Poorthuis, A. How to draw a neighborhood



### 1.8.5 Block 5: Spatial Statistics

10. Spatial Statistics I (Spatial Autocorrelation)
  - Reading:
    - TBA
11. Spatial Statistics II (Spatial Regression Modeling)
  - Reading:
    - TBA

### 1.8.6 Block 6: Classification

12. Classification (logistic regression & SVM)
  - Reading:
    - TBA
13. Classification (overflow, possible excursion to deep learning)
  - Reading:
    - TBA
14. Final Project Studio