

# GEOG0030: Geocomputation

Justin van Dijk

Last modified: 2022-11-25



# Contents

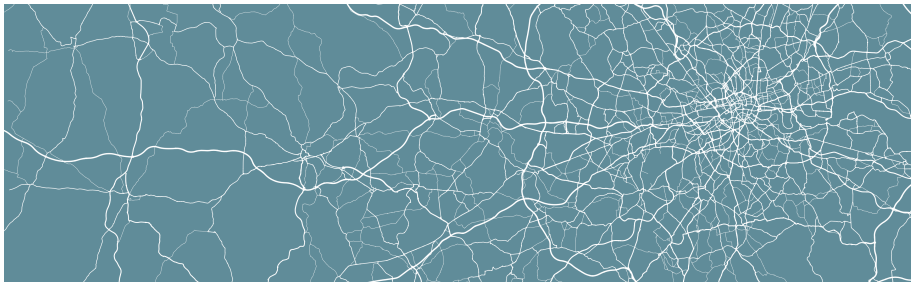
<b>Module Overview</b>	<b>5</b>
<b>Module Introduction</b>	<b>7</b>
Welcome . . . . .	7
Moodle . . . . .	7
Module overview . . . . .	8
Troubleshooting . . . . .	8
Acknowledgements . . . . .	9
<b>Foundational Concepts</b>	<b>11</b>
<b>1 Geocomputation: An Introduction</b>	<b>13</b>
1.1 Reading list . . . . .	13
1.2 Getting started . . . . .	14
1.3 Software . . . . .	14
1.4 Before you leave . . . . .	21
<b>Additional Resources</b>	<b>23</b>
<b>2 Data Sources</b>	<b>25</b>
2.1 Open Data . . . . .	25
2.2 Safeguarded Data . . . . .	26



# Module Overview



# Module Introduction



## Welcome

Welcome to **Geocomputation**. This module will introduce you both to the principles of spatial analysis as well as provide you with a comprehensive introduction to the use of programming. Over the next ten weeks, you will learn about the theory, methods and tools of spatial analysis through relevant case studies. We will start by using QGIS before moving to the R programming language. You will learn how to find, manage and clean spatial, demographic and socioeconomic datasets, and then analyse them using core spatial and statistical analysis techniques.

## Moodle

Moodle is the central point of contact for GEOG0030 and it is where all important information will be communicated such as key module and assessment information. This workbook contains links to all reading material as well as the content of all computer tutorials

## Module overview

The topics covered over the next ten weeks are:

Week	Section	Topic
1	Foundational Concepts	Geocomputation: An Introduction
2	Foundational Concepts	GIScience and GIS software
3	Foundational Concepts	Cartography and Visualisation
4	Foundational Concepts	Programming for Data Analysis
5	Foundational Concepts	Programming for Spatial Analysis
	<b>Reading week</b>	<b>Reading week</b>
6	Core Spatial Analysis	Analysing Spatial Patterns I: Geometric Operations and Spatial Queries
7	Core Spatial Analysis	Analysing Spatial Patterns II: Spatial Autocorrelation
8	Core Spatial Analysis	Analysing Spatial Patterns III: Point Pattern Analysis
9	Advanced Spatial Analysis	Rasters, Zonal Statistics and Interpolation
10	Advanced Spatial Analysis	Transport Network Analysis

## Troubleshooting

Spatial analysis can yield fascinating insights into geographical relationships, albeit at times it can be challenging, particularly when we combine this with learning how to program at the same time. You will most likely encounter many error messages, experience software crashes, and spend hours to identify bugs in your code. However, the rewards of learning how to programmatically solve complex spatial problems will be very much worth it in the end.

If you need specific assistance with this course please:

- Ask a question at the end of a lecture or during the computer practical.
- Attend the Department's **Coding Therapy sessions** that are run on a weekly basis.
- Check the Moodle assessment tab for queries relating to this module's assessment.



If after pursuing all these avenues you still need help, you can book into our office hours. You can use an office hour to discuss a geographical concept in relation to the material, assessment or for any personal matters relevant to the completion of the module.

## Acknowledgements

This year's workbook is updated and compiled using:

- The GEOG0030: Geocomputation 2021-2021 workbook as created and compiled by Dr Jo Wilkin.
- The GEOG0030: Geocomputation 2021-2022 workbook.

The datasets used in this workbook contain:

- Crime data obtained from data.police.uk (Open Government Licence)
- National Statistics data © Crown copyright and database right [2015] (Open Government Licence)
- Ordnance Survey data © Crown copyright and database right [2015]
- Public Health England © Crown copyright 2021



# Foundational Concepts



# Chapter 1

## Geocomputation: An Introduction

This week's lecture provided you with a thorough introduction on Geocomputation, outlining how and why it is different to a traditional GIScience course. We set the scene for the remainder of the module and explained how the foundational concepts that you will learn in the first half of term sit within the overall module. This week we start easy by setting up our work environment and set up the software that we will need over the coming weeks.

### 1.1 Reading list

#### Essential readings

- Brundson, C. and Comber, A. 2020. Opening practice: Supporting reproducibility and critical spatial data science. *Journal of Geographical Systems* 23: 477–496. [Link]
- Longley, P. *et al.* 2015. Geographic Information Science & Systems, **Chapter 1: Geographic Information: Science, Systems, and Society**. [Link]
- Singleton, A. and Arribas-Bel, D. 2019. Geographic Data Science. *Geographical Analysis*. [Link]

#### Suggested readings

- Miller, H. and Goodchild, M. 2015. Data-driven geography. *GeoJournal* 80: 449–461. [Link]

- Goodchild, M. 2009. Geographic information systems and science: Today and tomorrow. *Annals of GIS* 15(1): 3-9. [Link]

## 1.2 Getting started

Over the next few weeks, we will be taking a closer look at many of the foundational concepts that will ultimately enable you to confidently and competently analyse spatial data using both programming and GIS software. You will further learn how to plan, structure and conduct your own spatial analysis using programming – whilst making decisions on how to best present your work, which is a crucial aspect of any type of investigation but of particular relevance to your dissertation.

To help with this, we highly recommend that you try to stay organised with your work, including taking notes and making yourself a coding handbook. We would also suggest to list the different datasets you come across - and importantly, the scales and different projections you use them at - more on this over the next weeks. Finally, you should also make notes about the different spatial analysis techniques you come across, including the different properties they assess and parameters they require to run.

## 1.3 Software

This course primarily uses the R programming language, although we start by using QGIS in the next two weeks to give you a basic foundation in the principles of spatial analysis.

**Note** Please follow the instructions below to install both R and QGIS onto your own personal computer. If you cannot install the software on your personal computer or you are not planning to bring your own laptop to the computer practicals, please refer to the UCL Desktop and RStudio Server section below. Please make sure that you have access to a working installation of QGIS and R (including relevant packages) **before** the first hands-on practical session next week.

### 1.3.1 QGIS Installation

QGIS is an open-source graphic user interface GIS with many community developed add-on packages (or plugins) that provide additional functionality to the software. You can download and install QGIS on your personal machine by going to the QGIS website: [Link].

**Note** We recommend installing the **Long Term Release** (*QGIS 3.22 LTR*) as this version should be the most stable version. For Windows users: the QGIS installation may be a little slow.

After installation, start QGIS to see if the installation was successful and no errors are shown after start up.

### 1.3.2 R and RStudio Installation

R is both a programming language and software environment - in the form of RStudio- originally designed for statistical computing and graphics. R's great strength is that it is open-source, can be used on any computer operating system, and is free for anyone to use and contribute to. Because of this, it is rapidly becoming the statistical language of choice for many academics and has a very large user community with people constantly contributing new packages to carry out all manner of statistical, graphical, and importantly for us, geographical tasks.

Installing R takes a few relatively simple steps involving two programmes. First there is the R programme itself. Follow these steps to get it installed on your computer:

1. Navigate in your browser to your nearest CRAN mirror: [\[Link\]](#)
2. If you use a Windows computer, click on *Download R for Windows*. Then click on *base*. Download and install **R 4.2.x for Windows**. If you use a Mac computer, click on *Download R for macOS* and download and install **R-4.2.x.pkg**

That is it! You now have installed the latest version of R on your own machine. However, to make working with R a little bit easier we also need to install something called an Integrated Development Environment (IDE). We will use RStudio:

1. Navigate to the official webpage of RStudio: [\[Link\]](#)
2. Download and install RStudio Desktop on your computer (**free version!**)

After this, start RStudio to see if the installation was successful and no errors are shown after start up.

### 1.3.3 UCL Desktop and RStudio Server

As an alternative to installing QGIS and R with RStudio onto your personal device, there are some other options. Firstly, both programmes are available through Desktop@UCL Anywhere as well as all UCL computers on campus. In

case of R, there is also an RStudio server version available which you can access through your web browser: [\[Link\]](#)

You should be able to log in with your normal UCL username and password. After logging in, you should see the RStudio interface appear.

**Note** If it is the first time you log on to RStudio server you may only see the RStudio interface appear once you have clicked on the *start a new session* button. More importantly: if you are not on campus, RStudio server will only work with an active Virtual Private Network (VPN) connection that links your personal computer into UCL's network. Details on setting up a VPN connection can be found in UCL's VPN connection guides: [\[Link\]](#)

### 1.3.4 R package installation

Now we have installed or have access to QGIS and R, we need to customise R. Many useful R function come in packages, these are free libraries of code written and made available by other by R users. This includes packages specifically developed for data cleaning, data wrangling, visualisation, mapping, and spatial analysis. To save us some time, we will install all R packages that we will need over the next ten weeks in one go. Without going into detail on the RStudio (Server) interface, copy and paste the following code into the **console**. You can execute the code by hitting **Enter**. This may take a while.

```
# install all packages that we need
install.packages(c("tidyverse", "sf", "tmap", "osmdata", "RColorBrewer", "janitor",
                  "spdep", "dbscan", "raster", "spatstat", "gstat", "dodgr"))
```

Once you have installed the packages, we need to check whether we can in fact load them into our R session. Copy and paste the following code into the **console**, and executed by hitting **Enter** again.

```
# load all packages
library(tidyverse)
```

```
## -- Attaching packages -----
## v ggplot2 3.4.0      v purrr  0.3.5
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```



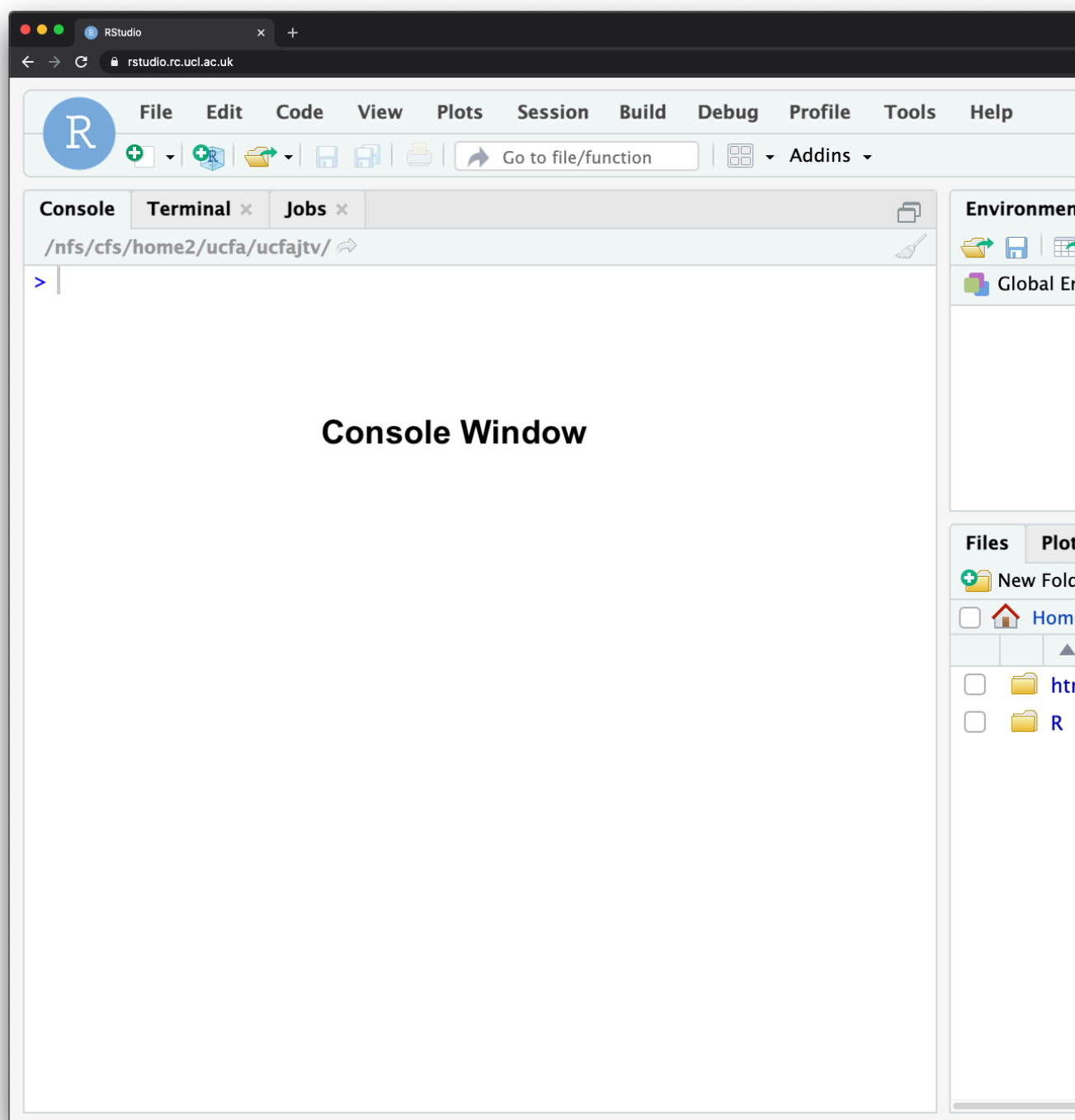


Figure 1.1: The RStudio Server interface.

```
library(sf)
```

```
## Linking to GEOS 3.10.2, GDAL 3.4.2, PROJ 8.2.1; sf_use_s2() is TRUE
```

```
library(tmap)  
library(osmdata)
```

```
## Data (c) OpenStreetMap contributors, ODbL 1.0. https://www.openstreetmap.org/copyright
```

```
library(RColorBrewer)  
library(janitor)
```

```
##  
## Attaching package: 'janitor'  
##  
## The following objects are masked from 'package:stats':  
##  
##      chisq.test, fisher.test
```

```
library(spdep)
```

```
## Loading required package: sp  
## Loading required package: spData  
## To access larger datasets in this package, install the spDataLarge  
## package with: `install.packages('spDataLarge',  
## repos='https://nowosad.github.io/drat/', type='source')`
```

```
library(dbscan)
```

```
##  
## Attaching package: 'dbscan'  
##  
## The following object is masked from 'package:stats':  
##  
##      as.dendrogram
```

```
library(raster)
```

```
##  
## Attaching package: 'raster'  
##  
## The following object is masked from 'package:dplyr':  
##  
##      select
```

```
library(spatstat)
```

```
## Loading required package: spatstat.data
## Loading required package: spatstat.geom
## spatstat.geom 3.0-3
##
## Attaching package: 'spatstat.geom'
##
## The following objects are masked from 'package:raster':
##
##   area, rotate, shift
##
## Loading required package: spatstat.random
## spatstat.random 3.0-1
## Loading required package: spatstat.explore
## Loading required package: nlme
##
## Attaching package: 'nlme'
##
## The following object is masked from 'package:raster':
##
##   getData
##
## The following object is masked from 'package:dplyr':
##
##   collapse
##
## spatstat.explore 3.0-5
## Loading required package: spatstat.model
## Loading required package: rpart
## spatstat.model 3.0-2
## Loading required package: spatstat.linnet
## spatstat.linnet 3.0-3
##
## spatstat 3.0-2
## For an introduction to spatstat, type 'beginner'
```

```
library(gstat)
```

```
##
## Attaching package: 'gstat'
##
## The following object is masked from 'package:spatstat.explore':
##
##   idw
```

```
library(dodgr)
```

You will see some information printed to your console but as long as you do not get a message that is similar to **Error: package or namespace load failed for <packagename>** or **Error: package '*<packagename>*' could not be loaded** all should be fine.

**Note** Even if you have used R or RStudio Server before and already installed some of the packages in the above list, do re-install all packages to make sure you have the latest versions. Legacy installations that have not been updated may lead to problems when going through the tutorials.

### 1.3.5 A note on ArcGIS

ArcGIS Pro (previously ArcMap) is the main commercial GIS software that you may have already used - or seen/heard about through other modules or even job adverts. We do not use ArcGIS Pro in our Practicals for several reasons:

- Computing requirements for ArcGIS Pro are substantial and it **only** operates on the Windows Operating System. For Mac users, using ArcGIS Pro (and ArcMap) would require using either a Virtual Machine or running a copy of Windows OS on a separate partition of your hard drive.
- It is **proprietary** software, which means you need a license to use the software. For those of us in education, the University covers the cost of this license, but when you leave, you will need to pay for a personal license (around £100 for non-commercial use) to continue using the software and repeat any analysis you have used the software for.
- Whilst ArcPro can use pure Python (and even R) as a programming language within it through scripts and notebooks, it primarily relies on its own **ArcPy** and **ArcGIS API for Python** packages to run the in-built tools and analytical functions. To use these packages, you still need a license which makes it difficult to share your code with others *if* they do not have their own ArcGIS license.

Recent developments in the ArcPro software, however, does make it an attractive tool for spatial data science and quantitative geography - it has cross-user functionality, from data analysts who like to use a tool called Notebooks for their code development, to those focused more on cartography and visualisation with in-built bridges to Adobe's Creative Suite. We therefore do not want to put you off looking into ArcGIS in the future, but for this course, we want to ensure the reproducibility of your work.

**Note** This also means that the analysis you will be doing for your coursework assignment must be completed in R and QGIS. Specific guidance on the coursework assignment and permitted software will be made available at the end of Reading Week.

## 1.4 Before you leave

You should now be all ready to go with the computer practicals the coming week. That is it for this week!



## Additional Resources





## Chapter 2

# Data Sources

Below you will find a list of resources that you might want to explore when sourcing data for your coursework assignment or your dissertation. This is by no means an exhaustive list, but simply contains some suggestions of websites that you may want to use.

**Note** You are **not limited** to using these datasets for your coursework assignment or your dissertation.

### 2.1 Open Data

The following websites contain Open Data or link to Open Data from several respectable data providers:

- AirBnB Data
- Bike Docking Data (ready for R)
- Camden Air Action
- Consumer Data Research Centre
- DEFRA
- DIVA-GIS
- Edina (e.g. OS mastermap)
- EU Tourism Data
- Eurostat
- Geofabrik (OSM data)
- Global Weather Data
- Google Dataset Search
- Johns Hopkins COVID19 Data (ready for R)
- King's College Data on Air Pollution
- London Data Store

- London Tube PM2.5 Levels
- NASA EARTHDATA
- NASA SocioEconomic Data and Applications Center (SEDAC)
- NHS Data (ready for R)
- nomis Official Census and Labour Market Statistics
- Office for National Statistics Geoportal
- Office for National Statistics
- Open Topography
- Tesco Store Data (London)
- TfL Cycling Data
- TfL Open Data
- Tidy Tuesday Data (not exclusively spatial data)
- Uber Travel Time Data
- UK COVID19 Data
- UK Data Service
- US Census Data
- US City Open Data Census
- USGS Earth Explorer
- WorldPop GitHub
- WorldPop

Some other websites that could be helpful:

- Awesome Public Datasets; general collection of datasets, although not limited to spatial data.
- Free GIS data; long list with lots of GIS datasets on many different topics and covering many different areas.

## 2.2 Safeguarded Data

Undergraduate students can also apply for a **Safeguarded** dataset held by the Consumer Data Research Centre. There is a process to access these **Safeguarded** datasets, which is detailed on the CDRC website. Please be aware that it normally takes 4-5 weeks for your application to be processed.

As part of the process, you will need to say in your application why you want that specific dataset and what you are planning to do with it. You will also need to have at least thought about the ethical implications of using that data and provide this with your data application (alongside your standard ethics application).

Some of the datasets held by the CDRC that you can apply for are:

- Bicycle Sharing System Docking Station Observations

- CDRC Modelled Ethnicity Proportions - LSOA Geography
- FCA Financial Lives Survey
- Local Data Company - SmartStreetSensor Footfall Data – Research Aggregated data
- NHS Hospital Admission Rates by Ethnic Group and other Characteristics
- Speedchecker Broadband Internet Speed Tests

**Note** Given that the application can take several weeks, the Safeguarded CDRC datasets may be useful for your dissertation work but probably not for the GEOG0030 coursework assignment. However, any of the CDRC datasets that are marked as **Open Data** do not require this application process and you can download these datasets directly after registering on the website.