

An Applied Geographic Information Systems and Science Course in R

Andrew MacLachlan¹ and Adam Dennett¹

¹ The Bartlett Centre for Advanced Spatial Analysis, University College London

DOI: [10.21105/jose.00141](https://doi.org/10.21105/jose.00141)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 05 May 2020

Published: 13 April 2022

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Updates

The content discussed within this paper refers to the module created for the academic year 2020-2021: https://andrewmaclachlan.github.io/CASA0005repo_20202021/. At the conclusion of each year content is copied to a new repository ending with the academic year it was taught, whilst the content applicable to the current academic year remains on the primary repository: <https://github.com/andrewmaclachlan/CASA0005repo>. This allows the authors and external users to track the development of the content.

Summary

This paper presents a masters level geographic information systems and science course (CASA0005) developed by the Bartlett Centre for Advanced Spatial Analysis (CASA), University College London (UCL). CASA0005 is a compulsory module for both MSc programs within the department, running annually in the autumn term with between 100 and 150 enrolled students. During the summer of 2019 the module was transformed from static practicals with some legacy content related to commercial software to an online interactive book primarily for the open source R data science programming language written using the Bookdown package ([Xie, 2016](#)).

Resource story

The geographic information systems and science course (CASA0005) is a core module on the CASA MSc and MRes programmes. Originally conceived primarily using ESRI's ArcMap with some Quantum GIS (QGIS), in recent years, consecutive module leaders and module staff, namely the authors of this paper, have transitioned and developed the content to R, with some QGIS. This reflects feedback from alumni, academia and industry that value programmatic and importantly reproducible GIS analysis. The course is representative of the expertise of the module staff with point pattern originating from Adam Dennett, raster analysis from Andrew MacLachlan and geographical regression a collaboration between the two.

Statement of need

The buzzword 'Big data' and associated roles of data scientists have become common additions to academia, industry and government organisations seeking to optimise workflows and base decisions on factual evidence ([Donoho, 2017](#); [Li et al., 2016](#)). In response,

the demand for data scientists has significantly grown, with two major recruitment websites LinkedIn and Glassdoor reporting a respective 56% and 44% year on year increase in advertised roles along with classification as the most promising career based on salary, growth, promotion opportunities and satisfaction within the USA in 2019 (Glassdoor, 2020; Pattabiraman, 2019). In a similar theme data from the UK's Universities and Colleges Admissions Service (UCAS) shows that the number of postgraduate level data science courses within the UK has grown from 104 across 51 institutions to 195 across 85 institutions between the academic years commencing in 2019 and 2020 (The Universities and Colleges Admissions Service, 2020). However, for the ability to use data to its full capability the spatial aspect is an essential component, yet the majority of current academic resources provide limited coverage of advanced, applied and reproducible geospatial methodologies using unfiltered and real world data examples (Arribas-Bel, 2019)

Whilst materials such as the R for Data Science book provide a foundational and essential resource for learning the R data science programming language, consideration of spatial data is extremely limited (Grolemund & Wickham, 2017). The welcome addition of Geocomputation with R specifically addressed this resource gap, but often used basic examples that are unrepresentative of the complex spatial data wrangling, transformation and manipulation skills required by master's level graduates (Lovelace et al., 2019). Consequently, this paper presents an academic teaching resource using applied real world government (e.g. Cosh, 2020) and academic examples (e.g. Guha et al., 2018) manipulating complex and large datasets in creating reproducible and actionable outputs, being made possible through delivery as a taught master's level module.

Learning objectives

Learning objectives are provided at the start of each individual practical that combine to achieve the overall module learning objectives.

After having taken this module, students should be able to:

- Develop a working knowledge of QGIS and R to support the application of GI Science techniques
- Visualise geographic information through producing appropriate maps to high cartographic standards
- Carry out spatial data management tasks (joining attribute to geometry data, cleaning data, converting between file formats and spatial reference systems)
- Interpret data and apply relevant spatial analyses (e.g. auto correlation/hot spot analysis, areal interpolation, point in polygon/buffer analysis, spatial statistical analysis) to answer a variety of spatial problems
- Explain and evaluate common issues with geographic data such as representation and uncertainty
- Apply and critique (spatial) statistical analysis techniques to infer relationships between spatial phenomena
- Experience the diversity of the global spatial data landscape and evaluate the relative drawbacks and merits of different spatial datasets

Content

The course can be accessed through the GitHub pages website <https://andrewmaclachlan.github.io/CASA0005repo/>. The content is equally divided into two main parts:

- Part 1 entitled GIS tools (practicals 1-5) focuses on demonstrating the main spatial software, toolsets and skills that the students will require throughout the module, MSc programme and in future spatial data science careers. These sessions involve spatial data loading and manipulation to produce descriptive statistics alongside cartographically sound mapped outputs in a reproducible manner.
- Part 2 entitled GIS analysis (practicals 6-9) builds upon the foundations established during the previous 5 sessions that form part 1, with the focus moving to analysing spatial data in response to proposed research questions, demonstrating types of analysis expected for the assessment of the course. Methods taught in this section include: spatial pattern analysis (Ester et al., 1996), remotely sensed raster data processing (Artis & Carnahan, 1982), bootstrapped, linear and geographically weighted regression (Brunsdon et al., 1996; Efron, 1988). The final practical concludes both part 2 and the course through exploring: advanced programming techniques, interactive mapping, further regression methods and cross validation.

Whilst practicals build upon concepts taught in previous weeks each could be used as a stand-alone tutorial.

Instructional design

The module runs over 11 weeks, composed of two 5-week teaching periods separated by a University wide reading week. It forms 15 of the 120 credits required for the completion of the MSc with an expected total student workload of 150 hours. The course assumes no prior knowledge and attracts a wide variety of students from across the University. It is designed to introduce key weekly theoretical concepts within an interactive 1-hour lecture followed by an associated practical from the Bookdown workbook 24 hours later. During this gap between the lecture and practical session students are encouraged to independently work through the self-guided material, with the practical sessions being used to solve problems and discuss previously taught concepts, often in a collaborative manner, following the flipped classroom pedagogical approach (Tucker, 2012). These practical sessions are run by two academic staff members who are supported by between four and five postgraduate teaching assistants. Throughout all delivered sessions students are encouraged to post questions (anonymously if needed) to a Slack channel monitored by the module academic team and engage with any ensuing in-class or virtual discussion.

Students transition from basic geo-spatial tools in part 1 to analysis focused practicals in part 2 that demonstrate methodological approaches used to solve recent policy and academic challenges. The shift from subject-based to problem-based learning in part 2 of the course is designed to develop independence and critical thinking. Whilst the ‘problem-first’ rule of problem-based learning is adhered to, the exact format of the part 2 practicals differs slightly from full traditional definition, as the material guides students to a potential answer (Kwan, 2009). This is in order to showcase methodological approaches available whilst stimulating ideas to solve similar research questions in preparation for their final enquiry-based learning assessment task. Within the assessment task students are required to develop and answer a pertinent or topical geo-spatial question/hypothesis/issue in a logical, scientific and reproducible manner (Kahn & O’Rourke, 2005). The progression through the module in this manner follows UCL’s current connected curriculum framework (also UCL’s educational framework) that aims to deliver a research-based education. Specifically, the enquiry based learning (that includes problem-based learning) aspects of this module align with expanding student research activity (point 2) and connecting academic learning to a workplace environment (point 4) (Fung, 2017). The applied skills acquired from this course in the first term of the MSc set a foundation for the remaining two terms and future careers in spatial data science.

Experience of use

The practical materials have been developed over several years with regular updates and revisions added before each annual delivery of the course to capture emerging or current trends in spatial analysis and responding to student feedback from previous years. Having the complete course materials in one searchable online and interactive book has been welcomed by the majority of students. In addition, owing to the logical Bookdown project format any issues with the taught concepts or code raised by the students can be easily and quickly resolved and the online book updated, improving student satisfaction.

As it is also possible to deliver each practical as a standalone tutorial parts have been included in workshop events and summer schools such as the CASA doctoral summer school for advanced spatial modelling.

References

- Arribas-Bel, D. (2019). A course on Geographic Data Science. *Journal of Open Source Education*, 2(14), 42. <https://doi.org/10.21105/jose.00042>
- Artis, D. A., & Carnahan, W. H. (1982). Survey of emissivity variability in thermography of urban areas. *Remote Sensing of Environment*, 12(4), 313–329. [https://doi.org/10.1016/0034-4257\(82\)90043-8](https://doi.org/10.1016/0034-4257(82)90043-8)
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4), 281–298. <https://doi.org/10.1016/B978-008044910-4.00447-8>
- Cosh, G. (2020). *GLA Housing and Land Short-term and holiday letting in London*.
- Donoho, D. (2017). 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, 26(4), 745–766. <https://doi.org/10.1080/10618600.2017.1384734>
- Efron, B. (1988). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics* (Vol. 2, pp. 569–593). Springer. <http://projecteuclid.org/euclid.aop/1176996548>
- Ester, M., Kriegel, H.-P., Sander, J., & Xiaowe, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd*, 96, 226–231. <https://doi.org/10.1016/B978-044452701-1.00067-3>
- Fung, D. (2017). *UCL Connected Curriculum* (2nd ed.). UCL Press.
- Glassdoor. (2020). *50 Best Jobs in America for 2019*. https://www.glassdoor.co.uk/List/Best-Jobs-in-America-LST_KQ0,20.htm
- Grolemund, Garrett., & Wickham, Hadley. (2017). *R for Data Science*. O'Reilly.
- Guha, S., Govil, H., Dey, A., & Gill, N. (2018). Analytical study of land surface temperature with NDVI and NDBI using Landsat 8 OLI and TIRS data in Florence and Naples city, Italy. *European Journal of Remote Sensing*, 51(1), 667–678. <https://doi.org/10.1080/22797254.2018.1474494>
- Kahn, P., & O'Rourke, K. (2005). Understanding enquiry-based learning. *Handbook of Enquiry & Problem Based Learning*, 1–12. <https://doi.org/10.1002/9781118944707.ch10>
- Kwan, A. (2009). Problem-based Learning. In M. Tight, K. H. Mok, J. Huisman, & C. Morphew (Eds.), *The routledge international handbook of higher education*. Taylor & Francis.

- Li, S., Dragicevic, S., Castro, F. A., Sester, M., Winter, S., Coltekin, A., Pettit, C., Jiang, B., Haworth, J., Stein, A., & Cheng, T. (2016). Geospatial big data handling theory and methods: A review and research challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 119–133. <https://doi.org/10.1016/j.isprsjprs.2015.10.012>
- Lovelace, Robin., Nowosad, Jakub., & Muenchow, Jannes. (2019). *Geocomputation with R*. CRC Press. <https://doi.org/10.1201/9780203730058>
- Pattabiraman, K. (2019). *LinkedIn's Most Promising Jobs of 2019*. <https://blog.linkedin.com/2019/january/10/linkedins-most-promising-jobs-of-2019>
- The Universities and Colleges Admissions Service. (2020). *UCAS Postgraduate courses*. <https://digital.ucas.com/coursedisplay/results/providers?studyYear=2019&destination=Postgraduate&postcodeDistanceSystem=imperial&pageNumber=1&sort=MostRelevant&searchTerm=data%20science>
- Tucker, B. (2012). The flipped classroom: Online instruction at home frees class time for learning. *Education Next*, 12, p. 82–83. ISBN: 15399664
- Xie, Y. (2016). *Bookdown: Authoring books and technical documents with R markdown*. CRC Press.