Teaching R as a GIS: problems, solutions and lessons learned

Robin Lovelace, Rachel Oldroyd
Centre for Spatial Analysis and Policy, University of Leeds

Abstract
R is a fast and flexible programming language and data environment which is widely accepted among the statistical community. It provides an open-source alternative to proprietary software such as SPSS and Stata. Despite R’s acceptance among the quantitative disciplines, it still remains unpopular in the social sciences, especially geography. This unpopularity is somewhat surprising considering the number of advanced spatial packages that have emerged in recent years. Factors contributing to R’s unpopularity may be its Command Line Interface (CLI), the steep learning curve associated with the language or the popularity of conventional Geographical Information Systems (GIS). In an attempt to raise awareness of R’s spatial capabilities and its use as a GIS, a teaching team from the University of Leeds repeatedly delivered a day-long workshop over the period of 18 months. Based on teaching experiences and formal feedback, the workshop materials were incrementally improved based on an understanding that R skills are developed through ‘experiential learning’. A number of measures were introduced to facilitate the learning process and the findings of the teaching activity should be of interest to anyone teaching computer code to beginners. The resulting materials and guidance will be of practical use for others teaching R for spatial applications.

Keywords
GIS, R, Spatial Analysis, Education

Introduction
The features, performance and graphics provided by R’s spatial packages now mean that R can rival market-leading GIS software suites such as QGIS and ArcGIS (Bivand, Pebesma & Gómez-Rubio, 2013). Despite the recent progression of R’s spatial capabilities, the software is still not frequently used by geographers and social scientists. Factors which may contribute to R's unpopularity include; the command line-based interface, the dominance and familiarity of proprietary GIS products and R’s reputation as being difficult to learn. To the majority of computer users, the Command Line Interface (CLI) is an alien concept. The prospect of using a tool which only responds to typed commands is unfamiliar and most social scientists prefer to ‘click and point’ their way to a pretty map. However, once familiarity with the CLI is gained, the benefits of using R compared to other Graphical User Interface (GUI) based programs become clear, these include:
• Ease of reproducibility, including modifying and building on the work of others.
• Speed of undertaking batch processes: a single script can perform the same process on many files with little or no modification.
• Advanced and highly customisable graphical capabilities, empowering the user to visualise data in new ways (e.g. faceted plots and animations).
Highly extensible environment with over 6800 additional libraries (Comprehensive R Archive Network [CRAN], 2014).

In short, R has proved itself as an integrated tool for processing, modelling and mapping geographical data (Ihaka & Gentleman, 1996), but this has yet to translate into widespread uptake by social scientists and geographers. In an attempt to dispel the negative image associated with R, a 1 day workshop was developed and delivered by staff at the University of Leeds. This workshop was delivered through the Geospatial Analysis and Simulation (TALISMAN) project, funded by the UK’s Economic and Social Research Council (ESRC).

Case Study
The aim of the Introductory R for Spatial Analysis (IRSA) workshop was to raise awareness of R’s spatial capabilities by teaching new and emerging methods for processing, analysing and visualising spatial data in R. 6 courses were delivered to approximately 130 participants between the period of November 2013 and March 2015 and informed changes were made to the teaching material and delivery of the workshop throughout this period. These changes were based on formal and informal feedback. Formal feedback was collected from 75 participants during 4 of the courses. The remaining 2 courses were delivered to non-native English speakers as part of international conferences and formal feedback was not collected.

The Teaching Model
Prior to development of the teaching materials, the target audience was identified. It was thought that the majority of attendees would be academics working in a social science related field with some experience of GIS. Measurable objectives were also defined to provide context to the teaching practice and provide a template for the teaching materials. It was hoped that, by the end of the workshop participants would:

- Have an understanding of R’s underlying syntax and structure
- Have an understanding of loading, manipulating, processing and transforming spatial data in R
- Have used R’s graphics packages to create aesthetic maps in R
- Have an awareness of R’s more complex spatial functionality

A teaching model consisting of short lecture style talks alongside independent work was adopted for the workshop. It was thought that this approach would facilitate experiential learning, ‘learning by doing’, and would allow participants to work at a pace which was comfortable for them. This would also provide the opportunity and time for individuals to explore concepts which were not familiar to them or which were particularly interesting. It was decided that an existing tutorial named ‘A short Introduction to R’ by Harris (2012) would be used in the first half of the workshop and a more spatially-inclined tutorial would be used in the second half to introduce more complex functionality. The latter tutorial was named ‘Creating Maps in R’ (www.github.com/rlovelave/Creating-Maps-in-R) and was purposefully written for use in this workshop.

In order to provide a quantitative measure of course effectiveness, feedback forms were collected after each workshop. The feedback form consisted of a
number of Likert scale questions. The results were summarised whereby each category was given an average score out of 100 based upon participant feedback. The categories were; the content was interesting; the course materials were useful; speakers knowledge; clarity of explanations; difficulty of the course; prior knowledge assumed; speed of presentation; time for questions. This quantitative reflection process provided an objective way of highlighting which parts of the course were successful and which were not.

**Problems and Solutions**

A well-defined problem associated with developing training workshops, is pitching the material at the appropriate level. It was decided that the workshop would be advertised as entry level, suitable for beginners who had little or no experience with R. In reality, a range of participants with varying levels of experience registered for the event. These ranged from proficient R users with limited experience of spatial data, to participants who regularly used spatial data but had no experience of R and very infrequently, participants who had limited experience of both R and spatial data. The extreme range of knowledge introduced a number of problems. Not only were the materials required to cover R’s basic syntax and capabilities, they also had to cover general programming techniques and basic geographical concepts whilst maintaining the interest of the higher-level users.

The majority of the problems experienced over the 18 month period stemmed from the range of participant knowledge. It soon became clear that completing two separate tutorials was not conducive to effective learning. The participants felt pressured into completing the tutorial within the limit of the one day period; ‘I acted more like a typewriter’, ‘I was typing code without understanding what I was doing’. To address this problem, the ‘Creating Maps in R’ tutorial was extended and adapted to form an entry level tutorial on which the IRSA course would be solely based. Although this meant that less material was covered, this approach allowed attendees more time to comprehend the code and attain a greater level of understanding. This was reflected in the feedback as the average score for ‘course difficulty’ rose from 93 to 97 and the score for ‘prior knowledge assumed’ rose from 89 to 96 out of 100, where a higher score reflected higher suitability.

Reducing the amount of material introduced another problem. Participants who were more proficient in R skipped over the basics covered at the beginning of the tutorial and finished earlier than expected. In an attempt to preserve the introductory level of the ‘Creating Maps in R tutorial’ but also maintain the attention of these participants, a number of optional extension exercises were introduced. These were added to the end of each section allowing higher-level users to further explore R’s spatial functionality. A conscious effort was made to clearly signpost these exercises as optional, as it was found that other participants felt pressured into completing all sections if not. Over the duration of the 18 months the score for the ‘Content was interesting’ and ‘Materials were useful’ categories increased from 84 to 89 and 85 to 91 respectively. The feedback also reflected upon the success of the optional exercises; ‘Although I started the course with considerable R experience, and the course had to go
through the basics. The pace was just enough and there were enough extra bits of information to keep me interested’, ‘I think the course strikes a good balance between being accessible for an R novice and interesting for more experiences R users’.

Originally it was assumed that participants would have some familiarity with geographical datasets, however this was not true of all the attendees. A number of attendees were unfamiliar with the shapefile format used in the tutorial and struggled to understand the differentiation between geometric data, attribute data and the associated R ‘slots’. In order to address this problem, it was decided that a traditional GIS would be used to supplement the teaching. QGIS provided a quick and easy way to visualise the geometry and attribute data of a shapefile for participants who were not familiar with geographical formats. This initial understanding proved pivotal to the comprehension of the concepts that followed and was extremely successful.

The aim of the IRSA workshop was to raise awareness of the spatial capabilities of R and to dispel the negative image associated with the language. The majority of participants had not used R before and it was therefore important to create a good first impression. The benefits of using Integrated Development Environments (IDE) for beginners are well defined (Debuse & Lawley, 2012), therefore the R Studio IDE was used to support the learning process. The auto-completion function in R studio was particularly useful as it meant that the complexities of the language did not impair learning. The help window allowed for quick and easy access to documentation and the workspace pane creates a list of active data, values and functions which helps the debugging process. It is thought that these devices all helped to ease the steep learning curve associated with the language.

Lessons Learned
In conclusion, the IRSA course has been extremely successful and continues to be in high demand. The materials have been improved incrementally over an 18 month period and a number of lessons have been learned. When communicating computer code to beginners it is important to provide a supportive learning environment and introduce devices to ease the learning process. The use of an IDE is one example, provided it is not over complicated (Reis & Cartwright, 2004). Quality always rules quantity in terms of materials and it is best to cater for the needs of the less experienced users as optional tasks can always be provided. Having a high number of demonstrators is essential when teaching an entry-level course and participants appreciate individual interaction; ‘Explanations given by the tutors in practical sessions was excellent’. A basic understanding of geographical data formats was essential to the success of the course and it should not be assumed that participants will be familiar with these formats. Traditional techniques have proved successful in providing familiarisation with the datasets. Lastly, it is important to mention the appropriateness of the course. This workshop was aimed at academics who had some experience of geographic data. In reality, a range of partipants attended. A small minority of attendees had used neither R or a GIS and the IRSA course was unsuitable for their needs. Training in traditional GIS techniques may have been more suitable. Defining the target audience and clearly signposting any prerequisites is essential in providing
participant satisfaction.

**References**


