

U20 Math 124 - Introduction to (Statistical) Programming with R [Online] Summer 2015

1 Instructor information

Jonathan Homola
Department of Political Science
Office: Seigle 277
E-mail: homola@wustl.edu

2 Course description

This is a 3-credit summer online course to introduce the R Statistical Environment to new users. R is “a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc.” (R Core Team 2015). The goal of this course is to familiarize you with R and the R-Studio IDE, and to provide you with tools to perform some elementary (statistical) programming in R. This will be done with a mix of lecture slides, readings, exercises and a final exam.

This course is not a statistics course, and given the ever increasing number of software packages and advanced statistical routines available in R, it also does not aim at providing a full overview of all the statistical and programming tools that are out there. However, one goal is for you to learn how to do some basic statistical work in R. Therefore, I will expect you to be familiar with basic concepts such as frequency distributions or means, but more complicated terms will be introduced and explained as necessary.

3 Learning objectives

At the conclusion of this course, students will be able to:

- Read data into R, manipulate and analyze it
- Debug, organize, and comment R code
- Understand the R environment for downloading, installing, and using packages

- Do basic programming to write their own functions
- Use loops
- Create standard and customized graphics
- Perform basic statistical operations

4 Prerequisite details

This is an online course. Access to a computer/laptop/Mac, on which new programs can be downloaded and installed is required. Students will be expected to be online multiple times during each week. Students must be comfortable with using technology such as web browsers, the navigation of Blackboard, accessing their system preferences, as well as downloading and installing new software. Previous knowledge of R or any other programming language is not necessary.

5 Course materials

Lectures and Assignments

Course materials and assignments will be posted online each week. There will usually be a number of assigned readings, a short lecture on the week's topics (PDF slides and/or R script), as well as examples of R code and the resulting output.

Readings

The following book will be required and can be purchased from the campus bookstore or Amazon:

de Vries, Andrie and Joris Meys. 2012. R for Dummies (1st edition). Wiley.

A second edition of the book will be published in early July, but for this class we will make use of the first edition. Depending on how quickly we progress, I might also assign a few readings from another book (Braun, W. John and Duncan J. Murdoch. 2008. A First Course in Statistical Programming with R. Cambridge University Press.). However, you do not have to purchase that book, since it is relatively expensive, we will most likely only read one or two chapters from it, and you can access an electronic version for free through the Olin library website.

Moreover, there are a number of high quality monographs on R that are 100% free and available online:

- An Introduction to R
- Using R for Data Analysis and Graphics: J.H. Maindonald

- simpleR - Using R for Introductory Statistics: John Verzani
- The R Guide
- Analysis of Epidemiological Data Using R and Epicalc: Virasakdi Chongsuvivatwong
- Statistics Using R with Biological Examples: Kim Seefeld and Ernst Linder
- An Introduction to R: Software for Statistical Modeling & Computing: Petra Kuhnert and Bill Venables

Software

We will be using the R statistical package (www.r-project.org) and the R-Studio IDE (www.rstudio.com). Both software packages are free of charge. Downloading and installing them will be part of the first section of the class.

6 Grading information and criteria

- Class participation (online discussion): 20%
- Homework assignments: 40%
- Final exam: 40%

A = 90-100%; B = 80-89%; C = 70-79%; D = 60-69%, F = below 60%

Class participation is strongly encouraged. Students will be graded on the basis of participation in online discussions on the Blackboard classroom website. Most commonly, these will be based on coding problems related to the current week's readings/topics or the previous week's assignment.

There will be homework assignments in every week, which will be due in the following week. The assignments will focus on the topics covered in the respective week's readings and lecture. Sometimes, homework will be done in groups together with other students. There will be clear deadlines, and credit will be deducted for late submissions. If for some reason a delay is anticipated, please notify me as early as possible in advance.

There will be one final exam. The exam will be relatively similar to the weekly assignments and will ask you to apply the programming and data management knowledge that you have learned throughout the course. Using the Blackboard discussion section during the final exam will be prohibited.

7 Course outline

- Week 1: Introduction – print/read syllabus; obtain course materials; become familiar with necessary technology/Blackboard; introductions discussion
Downloading and installing R, and R-Studio; basic syntax (comments, naming conventions etc.); quitting R and saving R objects
- Week 2: Data Analysis – data types and data structures; basic operations; data import and export; univariate data; bivariate data; regression analysis
- Week 3: Plotting – basic plotting commands; plotting categorical data; two-dimensional plotting commands; setting up the graphics window; combining different graphs; important ways to export your graphs
- Week 4: Basic programming – for, if, and while loops; apply
- Week 5: R Packages – finding, downloading, and installing packages
- Week 6: Advance programming – functions; plyr; debugging
- Week 7: Statistical programming – simulation; sampling; computational linear algebra; linear programming

8 Calendar of Assignments

Week	Assignment	Deadline
Week 1	Blackboard: Introductions and general discussion Weekly Assignment: Programming homework	Sun, July 5, 6pm Sun, July 5, 6pm
Week 2	Blackboard: General discussion Weekly Assignment: Programming homework	Sun, July 12, 6pm Sun, July 12, 6pm
Week 3	Blackboard: General discussion Weekly Assignment: Programming homework	Sun, July 19, 6pm Sun, July 19, 6pm
Week 4	Blackboard: General discussion Weekly Assignment: Programming homework	Sun, July 26, 6pm Sun, July 26, 6pm
Week 5	Blackboard: General discussion Weekly Assignment: R Package homework	Sun, August 2, 6pm Sun, August 2, 6pm
Week 6	Blackboard: Group discussion – Discuss last week’s packages Weekly Assignment: Programming homework	Sun, August 9, 6pm Sun, August 9, 6pm
Week 7	Blackboard: General discussion Weekly Assignment: Final Exam	Fri, August 14, 6pm Fri, August 14, 6pm

9 Technology

Technical support

This is a fully online, technology-based course. Because computers are not perfect, plan on having technical issues at least once during the term. While this can cause some incredibly

frustrating moments, the overall benefits of the technology do outweigh any issues that may arise. Just be ready to contact technical support in the event of difficulties. Send an email to blackboardhelp@wustl.edu with any questions/problems concerning Blackboard. (This is technical support for any problems you experience within the Blackboard classroom only, not other external software). For additional Blackboard resources and assistance, visit the Blackboard Answers & Support page on the University College Website.

Technological issues related to R and R-Studio will be dealt with in class. Solutions will either be suggested by me or discussed on the course's Blackboard website. Students are responsible for resolving any other technical issues with their personal computer/laptop/Mac such as outdated software, permissions, etc.

“Netiquette” statement on internet communication

- Remember your audience. If you would not say it in a face-to-face classroom, do not include it in the online discussions either. Consider what you write, as it is a permanent record and can be retrieved easily. Use courtesy and common sense in all your electronic communications.
- Write in complete sentences and check spelling before you post anything in class.
- DON'T TYPE IN ALL CAPS. This is hard to read and is considered “shouting”.
- Respect the opinions of others and be sensitive to the diverse nature of people in the class. Keep in mind that although you cannot “see” your classmates, you can show respect for individual differences. Diversity issues may include the following and others: race, ethnicity, religion, disabilities, gender, sexual orientation, age, social class, marital status, urban vs. rural dwellers, etc.
- No profanity will be allowed. This includes writing in punctuation. For example, `##%$#!` is considered profanity and is not permitted. Also, language expressed in inappropriate acronyms is not acceptable.

Expectations for student attendance

You are expected to have an active presence within Blackboard (our online classroom) and that is one of the ways in which I will monitor your attendance in the course. I am able to see when last you logged into Blackboard and how much time you have spent in various areas of the classroom. The class is not self-paced because we have specific due dates for all assignments, but you will work on the course at your own convenience within those parameters. Lack of participation in discussions or not answering email messages sent to you will be considered lack of attendance and will result in a loss of points.

10 UC policies

Students are bound by the University College policy on academic integrity in all aspects of this course. All references to ideas and texts other than the students' own must be so indicated through appropriate footnotes, whether the source is a book, an online site, the professor, etc. All students are responsible for following the rules outlined in the document regarding the university academic integrity policy: ucollege.wustl.edu/faculty/academic_integrity.

11 Helpful websites

- An Introduction to R
- R For Dummies Cheat Sheet
- An R Reference Card
- R Biostat Resources at Vanderbilt