

## STAT 5730 Introduction to R for data science

lecture: M 12:10pm – 2:00pm in [Arps Hall 012](#)  
 instructor: Vincent Q. Vu ([vqv@stat.osu.edu](mailto:vqv@stat.osu.edu))  
 office: Cockins Hall 428b  
 office hours: W 1:00pm – 2:50pm, or by appointment  
 web: Class schedule, assignments, and course announcements will be posted on Carmen ([carmen.osu.edu](http://carmen.osu.edu))  
 prerequisites: STAT 1350, 1450, or 1550 or equivalent or permission of instructor

## I Overview

R is a freely available statistical computing environment and programming language. It has become a dominant workhorse for modern statistical research and data analysis, and is being widely adopted in industrial data analytics as well. The primary goal of the course is to teach students how to use R for data analysis: both (1) efficient use of the R computing environment and (2) effective programming in the R language.

There are formal prerequisites for the course. This is a *statistics* course, so the examples and applications demonstrated in the class will be oriented towards data analysis and statistical endeavors. Basic numeracy and familiarity with statistics is expected for motivation and perspective. No programming experience is required.

## 2 Course materials & computing

### • Required reading

- (R4DS) Golemund and Wickham (2016): *R for Data Science*. O'Reilly. ISBN: 9781491910382. (web: [r4ds.had.co.nz](http://r4ds.had.co.nz)). The web version of the book is can be accessed freely from any web browser. Electronic access to the print version of the book is available at <https://www.safaribooksonline.com/library/view/-/9781491910382/?ar>.

**Note that the web and print versions have different chapter numbering.**

- (HoPR) Golemund (2014): *Hands-On Programming with R*. O'Reilly. ISBN: 9781449359089. Electronic access to the book is available at <https://www.safaribooksonline.com/library/view/-/9781449359089/?ar>
- Print copies of both books can be purchased directly from [oreilly.com](http://oreilly.com). After visiting one of the above links, if you sign-up for a O'Reilly account with your OSU email address and install the appropriate app to your iOS or Android device (<https://www.oreilly.com/online-learning/apps.html>), you should be able to download the books for offline access.

- **Software**

- R ([www.r-project.org](http://www.r-project.org))
- RStudio ([www.rstudio.com](http://www.rstudio.com))

You are expected to be able to access working installations of **current versions** of the required software. RStudio Server login access will be provided to students registered in the course by the Ohio Supercomputer Center (OSC, [www.osc.edu](http://www.osc.edu)). This will allow you to access R via the RStudio IDE from any web browser. Alternatively, you can also install R and RStudio on your personal computer by downloading the software from the links above.

### 3 Tentative schedule

The following is a tentative schedule of topics. Reading for R4DS refers to the numbering of the web version of the book. We may deviate from this schedule, so pay attention to announcements.

Week	Topic	Due	Reading	Extra Reading
1	Introduction to R, RStudio and R Markdown		R4DS 1.4–1.6, 4, 27.1–27.4.3	HoPR 1–2
2	Visualization in R with ggplot2		R4DS 3, 28.2	R4DS 28
3	<b>No class</b>	HW1		
4	Data types and representation	HW2	HoPR 1, 3–5	
5	Data frames and data manipulation	HW3	HoPR 3–4, R4DS 5.1–5.5, 10	
6	Data summarization	HW4	R4DS 5.6–5.7	
7	Data import and workflows	HW5	R4DS 6, 8, 11.1–11.2, 11.6	R4DS 7
8	<b>Exam</b>			
9	Tidy data and relational data		R4DS 12–13	
10	<b>Spring break</b>			
11	Functions and conditional execution	HW6, Project proposal	R4DS 19, HoPR 6–7	
12	Strings and factors	HW7	R4DS 14–15	
13	Iteration and functional programming	HW8	R4DS 21, HoPR 9	
14	Dates and times	HW9	R4DS 16	
15	Working with many models and list-columns	HW10	R4DS 25	
16		Project presentation		

## 4 Coursework & grading

There will be homework, a midterm exam, and a class project. Grading will be based on the following components:

- 50% Homework (10 assignments, lowest score dropped)
- 25% Exam (in-class)
- 25% Project

Learning to compute and program requires practice. So homeworks will be assigned on a weekly basis. These will mainly consist of exercises designed to reinforce the concepts covered in class during the previous week. Late homework will not be accepted, however you will be allowed to drop one homework score from your grade. I recommend completing all of the homeworks, even if you plan to drop one.

There will be one in-class exam. It will be open book/internet access, but absolutely no communicating with other humans will be allowed. The format of the exam will consist of conceptual questions as well as some computing and programming problems involving data.

Students will work in small groups on a final project consisting of producing an interactive Shiny app. I will provide a list of topics. Each group will cooperate in the design, development, and making a presentation on the project.

## 5 Rules and policies

### 5.1 Attendance

It is the responsibility of students to make up for any material covered in class during any absence.

### 5.2 Academic misconduct

It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term “academic misconduct” includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487). For additional information, see the Code of Student Conduct <http://studentlife.osu.edu/csc/>.

### 5.3 Disability services

The University strives to make all learning experiences as accessible as possible. If you anticipate or experience academic barriers based on your disability (including mental health, chronic or temporary medical conditions), please let me know immediately so that we can privately discuss options. To establish reasonable accommodations, I may request that you register with Student Life Disability Services. After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion. SLDS contact information: [slds@osu.edu](mailto:slds@osu.edu); 614-292-3307; [slds.osu.edu](http://slds.osu.edu); 098 Baker Hall, 113 W. 12th Avenue.