**Syllabus**

*Course Information*

ECL298 – 014 **Basics of Data Manipulation in R**  2-units, S/U grading

Thursdays 10:30 – 11:50; Winter quarter 2013; 115 Hutchison Hall

*Sponsored by Graduate Studies through the Professors for the Future and Learning Pathways programs*

*Instructor Information*

Primary instructor:

**Anna Steel**

1072 Academic Surge

Office hours: 11-12 Monday or by appointment

[aestephenson@ucdavis.edu](mailto:aestephenson@ucdavis.edu)

Teaching assistants:

**Robert Hijmans**

2001 Wickson Hall, rhijmans@ucdavis.edu

Office hours by appointment

**Matt Whalen**

4345 Storer Hall

Office Hours: T-Th by appointment

mawhalen@ucdavis.edu

*Communication Guidelines*

Please feel free to email Anna any questions; **I will return your email within 48 hours during the week** but reserve the right to **not check my email on the weekends**. In addition, I will be better able to help with R questions if you include a file with some (simulated) example data and code (or a file with the real data if necessary) so I can reproduce the problem. However, we will both **save time if we can arrange a meeting** and review the code together.

*Course Description*

R is powerful, flexible, open-source software for manipulating and analyzing data. It is constantly being improved and expanded by a diverse group of people. Many researchers claim that data analysis with R is the way of the future because of its accessibility (it is free), versatility (advanced tools for almost any field, from finance to GIS and genomics), flexibility (the question is not if you can do it, but how), and reproducibility of analysis (through short scripts). This course will help you to become comfortable with basic data manipulation in R, beginning with installing and setting up the program, learning the basic idiom, and continuing on to manipulating and organizing your data, creating good graphics, and learning how to find more help.

*Learning Objectives*

1. Be able to import, clean, and tidy data in R
2. Be able to calculate descriptive statistics and create illustrative plots from that data
3. Develop/discover a personalized and effective workflow, including file storage, file naming, and script organization
4. Become comfortable finding and using R-help resources

*Course Philosophy*

It takes a while to learn writing R scripts. Just like speaking a second language or learning a sport, you have to *practice* to learn. Personally, I picked up the basics of R scripting through a collaborative project; as we did the analysis I watched a colleague write advanced code, and then applied what I’d learned to much simpler analyses. I spent plenty of time frustrated with on-line help forums, trying to decode the jargon-filled R manuals, and reading through book chapters only to find out it wasn’t what I actually needed to know. It was often maddening, but at the end of the project I discovered that somehow I had become comfortable with the basics of R, and fellow graduate students were seeking my advice. I’m not suggesting that “it was hard for me, so it will be hard for you!” but rather that to truly learn and retain these skills you will need to think through challenging problems and discover solutions. My goal with this course is to provide basic information and structured practice so that you can be more efficient in your learning process.

With this goal in mind I’m using a flipped classroom for this course. Traditionally, class time is when you would listen to a lecture to become familiar with the basic content then you would go home to struggle through assigned problems alone. However, in this course these things will be flipped. I’ll create screen-casts of the lectures so you can familiarize yourself with the basic syntax and commands *at home,* and class will become a period to ask questions and work through problems with feedback and help from your peers and a few more experienced R users (the instructors). This should create a more efficient and effective way to work through the potentially frustrating experience of error messages and failed code.

*Expectations of Students*

Given this approach, the onus of learning is on you as students. Each week will build on the previous week, and to learn these basic skills you must practice them. Therefore I expect the following:

* Watch the screen-cast prior to class
* Attend class
* Turn in your own scripts, even if you completed the problems as a group

As a graduate student myself, I understand that life is busy – mine is too. However, **if you plan to join us you should expect to spend 1-2 hours to prepare for class, 1½ hours in class, and possibly 1 hour to review or finish problem sets. If you cannot commit to 3-5 hours a week, please don’t sign up for this class.** Having said that, I do realize that life happens in unexpected ways. Please contact me with genuine concerns about scheduling, time, or any other issues that arise.

*Assessments*

* Weekly in-class problem sets, uploaded to SmartSite each Friday by 5pm
  + I will use the performance on these problem sets to identify areas where we as a class are struggling, and use that to inform subsequent screen-casts
* Final project, due to SmartSite by Friday March 15th (last week of the quarter)

If fewer than 75% of weekly assignments are submitted on time, you will not receive a passing grade. If the final project does not demonstrate that you can adequately manipulate a dataset you will not receive a passing grade.

*Resources and Textbooks*

* **You MUST have a Laptop** to bring and use during class periods; ideally you will have administrative rights so you can install R and any necessary packages
* No texts are required for this class, although there are useful introductory books available both on line and in hard copy. Links to those electronic books owned by UC Davis will be included on the SmartSite, along with other helpful resources for you to reference (weblinks, command cheatsheets, etc).
* All screen-casts will be posted in the Resources folder on SmartSite, and will be available by noon on the Monday before class.
* Practice problem sets will also be posted on SmartSite to be available for interested students who were unable to enroll in the course due to space or time constraints.

*Preliminary Schedule*

This is a tentative schedule of topics to be covered; the screen-casts will be adjusted to account for how the class is progressing in at any given week. This is first time this course is offered, so I appreciate your flexibility and understanding!

|  |  |  |
| --- | --- | --- |
| **Week** | **Date** | **Screencast Topic** |
| 1 | January 10 | NA - course logistics and basic R structure |
|  |  |  |
| 2 | January 17 | Creating, Importing, Exporting Data |
|  |  |  |
| 3 | January 24 | Calling, Filtering, Describing Data |
|  |  |  |
| 4 | January 31 | Tidying Data (organization, missing values, order…) |
|  |  |  |
| 5 | February 07 | Time formats |
|  |  |  |
| 6 | February 14 | For-Loops vs. 'Vectorization' |
|  |  |  |
| 7 | February 21 | Creating and Customizing Plots |
|  |  |  |
| 8 | February 28 | Basic Statistical Models |
|  |  |  |
| 9 | March 07 | Catch-up - option for additional topics |
|  |  |  |
| 10 | March 14 | Work on Final Project **DUE MARCH 15** |