![A close up of a sign

Description generated with very high confidence]()

BIOF 339: Practical R

Number of Credits: 2

Fall 2020 Term A

Syllabus

Instructor: Abhijit Dasgupta, PhD

Contact Information:

E-mail: adasgupta+biof339@araastat.com

Phone:

Virtual Office Hours:

Preferred Method of Communication: Via Slack. Sign up via [this link](file:///Users/abhijit/Downloads/(https:/join.slack.com/t/biof339/shared_invite/zt-hczp2mg1-Yh0yqms52wAA8H445jkBUg)

Course Information:

Prerequisites, if any: None

Course Description: The goal of this course is to introduce R as an analysis platform and tool for data science rather than a programming language. Throughout the course, emphasis will be placed on example-driven learning. Topics to be covered include: installation of R and R packages; command line R; R data types; loading data in R; manipulating data; exploring data through visualization; statistical tests; correcting for multiple comparisons; building models; generating publication-quality graphics; creating reports using R Markdown. No prior programming experience is required.

Course Website (Canvas):

Course materials will be posted at 10 am on Monday of each week. Homework will be submitted on Canvas by Sunday 11:59 pm at the end of each week.

Learning Materials:

Required and Recommended Texts: There are no required texts for this class

Required Journal Articles: There are no required journal articles for this class

Course Goals

When you complete the course successfully, you will be able to:

* Run R and RStudio, making use of inherent R features
* Find and make use of the extensive packages (R add-ons) available for analyzing biological and other forms of data
* Load, manipulate, and combine data to make it amenable to further analyses
* Visualize data with extensive graphics capabilities of R (including ggplot)
* Use R to run statistical models and hypothesis tests and report results conforming to standards expected in scientific journals
* Write reports using the powerful rmarkdown package and its derivatives

Structure of the Course

| **Week** | **Topic** |
| --- | --- |
| Week 1 | Introduction to R: Working environment and data structures |
| Week 2 | Using packages to enhance data ingestion, munging, and reporting |
| Week 3 | Data visualization for exploration and reporting |
| Week 4 | Statistical analyses using R |
| Week 5 | Statistical learning using R |
| Week 6 | Designing and analyzing experiments, with a sprinkling of bioinformatics |
| Week 7 | Reproducible documents for analytic reporting |

The Learning Process

My philosophy is to get students up and running with R for data analysis as quickly as possible. As such, this course is opinionated, in that I make certain choices of what parts of R to teach to make things most accessible and useful. The course will be a mixture of didactic lessons, interactive tutorials and exercises, culminating in a final project that brings different aspects of the course together into a single document, either a HTML report or presentation, mimicking a typical workflow in data science of going from data acquisition to reporting. We will learn some of the details or nuances of R as we go along when they come up, but the general approach will be to begin with the end in mind.

Students can be successful in this course through following the teaching materials, participating in discussions on Slack, and practice. R is a language in the same way that French or Japanese is a language (you’re just talking to a computer), and so the only way to retain the knowledge gained in this class is to **use it**. The exercises and tutorials are meant to get you used to using R for different purposes, so please do them diligently.

This course should take around 4-6 hours of time weekly, depending on the week.

Important Dates:

Drop deadline

Audit and withdrawal deadline

Holidays

Communication:

This class will communicate primarily via [Slack](http://www.slack.com). Please join the BIOF339 Slack channel using [this link](https://join.slack.com/t/biof339/shared_invite/zt-hczp2mg1-Yh0yqms52wAA8H445jkBUg).

You will see a channel #fall2020-a. Please join this channel. Please use Slack for broadcasting messages, answering questions and the like. When you ask a question, please ask it under the #general or #fall202-a channels, so others can learn as well. I should respond within 24 hours.

The Canvas Discussion forum will be used for guided class discussions.

Etiquette:

The most important thing is to be polite, considerate and empathetic in all communications and discussions. There are different levels of knowledge about R in this class, and so some questions may appear trivial to some but are essential for others. Be kind, and if you can help a classmate, do so with grace and civility. The class learns best if we all help and support each other.

Policies:

Academic Policies

This course adheres to all FAES policies described in the academic catalog and student handbook, including the Academic Integrity policy listed on page 11 of the academic catalog and student handbook. Be certain that you are knowledgeable about all of the policies listed in this syllabus, in the academic catalog and student handbook, and on the FAES website. As a student in this program, you are bound by those policies.

Copyright

All course materials are the property of FAES and are to be used for the student’s individual academic purpose only. Any dissemination, copying, reproducing, modification, displaying, or transmitting of any course material for any other purpose is prohibited, will be considered misconduct, and may be cause for disciplinary action. In addition, encouraging academic dishonesty by distributing information about course materials or assignments which would give an unfair advantage to others may violate the FAES Academic Integrity policy. Course materials may not be exchanged or distributed for commercial purposes, for compensation, or for any purpose other than use by students enrolled in the course. Distributions of course materials may be subject to disciplinary action.

Guidelines for Disability Accommodations

FAES is committed to providing reasonable and appropriate accommodations to students with disabilities. Students with documented disabilities should contact Dr. Mindy Maris, Assistant Dean of Academic Programs.

Dropping the Course

Students are responsible for understanding FAES policies, procedures, and deadlines regarding dropping or withdrawing from the course or switching to audit status.

Harassment

FAES adheres to the NIH’s harassment policies, which can be found at the following link:

<https://hr.nih.gov/working-nih/civil/statement-workplace-harassment>

Faculty and students in FAES courses are responsible for being familiar with the NIH’s harassment policies and adhering to them.

Attendance

It is in your best interest to use, utilize, question and understand all the instructional material provided, and to submit questions and homework in a timely manner. Since this course is completely asynchronous, there is no attendance required at particular times.

Participation

Participation will be judged through the assigned discussions as well as through activity on Slack.

Assignment Submission

Assignment submission is through Canvas. Each submission will consist of a R Markdown file and the corresponding HTML file. **Both are required.** Just submitting the R Markdown doesn’t let us see the results easily, and just submitting the HTML doesn’t let us evaluate your code. If you have trouble knitting the R Markdown to HTML, let me know and I can help. If it’s really impossible and you’re tearing your hair out, reach out to me at least by Saturday so I can see if (a) I can help, or (b) I can see if reasonable accommodation can be made. The latter will be a rarity, generally.

Due Dates

Homework is assigned at 10am each Monday and is due by 11:59pm the following Sunday.

Late Submission Policies

No late submissions of homework or discussion are allowed. However, for homework, I will only use the top 4 scores for your grade, so you will have the option of not submitting or doing poorly on 2 of them.

Step-by-Step Guidelines for Submitting Assignments:

The guidelines for submitting assignments will be posted as a screencast during the first week of class.

Expectations for instructor’s feedback on assignments:

We will get your assignment grades and feedback to you within a week of submission.

Major Assignments:

Grades will be based on the following requirements:

1. Homeworks for each week are due Sunday at 11:59pm (50%)
   * No late homeworks
   * We’ll have 6 homeworks, I’ll score the top 4 for grade
2. Final project: A R Markdown report/presentation demonstrating an end-to-end data analysis in R using your own data, from data ingestion to munging to analyses and graphics, with a brief introduction and conclusion (30%)
3. Class participation (20%): Discussion topics in weeks 2-6

Final project

* Create a R Markdown document or presentation
* Use your own data, or data available on the web (legally)
* Show me that you can
  + import data into R
  + manipulate (munge) the data
  + perform some analysis on the data
  + create a visualization
  + create a report in R Markdown
* 5 minute lightning talks that can be recorded using Quicktime or [Screencastify](https://www.screencastify.com/)

Weekly Schedule:

| **Week** | **Topic** |
| --- | --- |
| Week 1 | Introduction to R: Working environmnent and data structures |
| Week 2 | Using packages to enhance data ingestion, munging, and reporting |
| Week 3 | Data visualization for exploration and reporting |
| Week 4 | Statistical analyses using R |
| Week 5 | Statistical learning using R |
| Week 6 | Designing and analyzing experiments, with a sprinkling of bioinformatics |
| Week 7 | Reproducible documents for analytic reporting |