

BIOF 439: Data Visualization using R

Course Notes

Abhijit Dasgupta, PhD

Last updated: May 03, 2019

Welcome

This course is an introduction to the statistical programming language [R](#) and various applications. We will cover the entire data analytics pipeline from data ingestion to data wrangling, summarizing, modeling, visualizing and reporting, all using tools found within the R ecosystem.

The version of these notes you are reading now was built on 2019-05-03.

Reproducibility

These notes are written with [bookdown](#), a R package for writing books using [rmarkdown](#). All code in these notes were developed on R version 3.5.3 (2019-03-11), using the same packages pre-installed in your virtual machines. When you're on your own, you will need to install a recent version of R, and also install the corresponding packages, on your computer, for all the code to work. A listing of all the packages used in this course will be available as an appendix.

To build these notes locally, clone or [download](#) the [Github repo](#) hosting these notes, unzip it if necessary, and double-click on `FSI_Book.Rproj`. Assuming you have RStudio installed, this will open this project (more on *RStudio Projects* later). You can then go to the console and enter the following code:

```
bookdown::render_book("index.Rmd") # to build these notes
browseURL("_book/index.html") # to view it
```

Chapter 1

An introduction to BIOF 439

1.0.1 Objectives of this course

- Understand principles of good data visualization
- Know what might make a visualization poor or ineffective
- Get you going using R **for visualization**
 - Various packages
- Creating static and dynamic visualizations using R
- Using the web as a presentation medium

1.0.2 Course resources

- Website <http://www.araastat.com/BIOF439>
- Slack <https://biof439dataviz.slack.com>
- RStudio Cloud <https://rstudio.cloud>

1.1 Homework policies

- Homework assignments will be posted on the website as well as on RStudio Cloud Friday by 9 am
- Homework assignments submissions will be based on a R Markdown file and the corresponding HTML file. These will reside in the RStudio Cloud project for that assignment. We'll be able to see and check them there. **Do not e-mail me anything**
- Homework assignments are due back to me by the following Tuesday at midnight. The instructional team will check the timestamp of the file.
- You may be late on at most 1 homework out of the 6 homeworks that will be assigned.
- Homeworks will count for 50% of your grade.

1.1.0.1 Collaboration

I encourage collaboration and team learning, but homeworks should be one's own work

1.2 Final Project

- A R Markdown document or presentation
- Use your own data
- Use R package(s) to visualize your data sets in at least 3 ways, to show what your data looks like and what your analytic results look like
- Each student will be randomly assigned to 3 peers
 - Critiques based on quality and effectiveness of visualizations
- All final projects will be posted on the website, so we can learn from each other
- I fully expect some of you to blow me away!!

This will count for 20% of your grade

1.3 Class participation

- Ask questions
- Comment on the strengths and weaknesses of visualizations when we work on them

This will count for 30% of your grade

1.4 Exemplar data

I don't work in bioinformatics anymore, or your particular disciplines. So:

1.5 Contact info

1.5.0.1 Email: adasgupta@araastat.com (don't use my NED email)

1.5.0.2 Slack

1.6 Code repository