Course Syllabus for CRSS 8220  
Advanced Topics in Crop and Soil Sciences  
Data Science and Statistical Programming Applied to Ag.  
UGA Spring Semester 2024

# Course information

## General information

* CRSS 8220 - Advanced Topics in Crop and Soil Sciences - Data Science and Statistical Programming Applied to Agriculture
* Spring Semester 2024
* 3 credit hours

## Meeting times and locations

* **Lectures**: Tuesday at 09:30-11:10
* **Labs**: Thursday at 09:10-11:10
* Location:
  + Athens campus: in person at 1203 Miller Plant Sciences
  + Tifton campus: in person at 601 NESPAL South OR remote
  + Griffin campus: M/F 217 SLC in person, W at 107 SLC in person OR remote

## Prerequisites

STAT 6315 – Statistical Methods for Researchers

## Co-requisites

None.

# Instructor information

## General information

**Dr. Leonardo M. Bastos**, Assistant Professor  
Crop & Soil Sciences Dept.  
4101 Miller Plant Sciences Building, Athens Campus  
University of Georgia  
Email: lmbastos@uga.edu  
URL: leombastos.github.io/bastoslab/

## Office hours

Please make an appointment if you would like a face-to-face meeting the instructor. Otherwise, I am always available by email.

# Course description and details

## Description

This course will expose students to common data analytical workflows in agriculture while utilizing data science principles. For that, students will learn how to develop workflows that include finding and importing data, exploratory data analysis, data wrangling and processing, fitting a model to the data, assessing model quality, extracting model information, and creating publication-ready figures. This type of workflow will be implemented to both designed and observational data commonly found in agricultural sciences, including analysis of variance, regression, and machine learning algorithms. Students will learn how access publicly available data sets for crop, soils, and weather information, and train machine learning models on these data. All the above will be performed while learning and using data science tools for reproducibility like version control, R statistical programming, APIs to publicly available data sets, task automation, and creating online interactive dashboards.

## Course learning outcomes

The general course objective is to provide students with hands-on applied experience in analyzing agricultural data using modern reproducible tools. That involves: - Learning and applying analytical workflows that involve importing data, processing, analyzing, assessing model fit, extracting model information (means and pairwise comparisons, regression coefficients) and producing publication-ready figures for different analysis including ANOVAs and regression. - Conducting analysis of variance workflows for the most commonly used agricultural designed studies (completely randomized design, randomized complete block design, split-plot design) - Conducting linear and non-linear regression workflows. - Learning and applying machine learning concepts (bias-variance trade-off, data split, hyper-parameter optimization, predictive metrics) and algorithms to agricultural observational data (soils, weather, yield). - Doing all the above while learning and using data science tools for reproducibility like version control, statistical programming, APIs to publicly available data sets, task automation, and creating online interactive dashboards.

## Topical Outline

1. Intro to R and RStudio (R script, Rmarkdown, quarto, RStudio Projects)
2. Version control with git and GitHub
3. R APIs to publicly available data (USDA NASS, weather, soil)
4. Data wrangling with dplyr, tidyr, pipe operator
5. Data visualization with ggplot2, gganimate
6. Experimental concepts of experimental unit, randomization, and replication
7. Experimental and treatment designs and ANOVAs (model fit, assumption checking, inference, plot):
8. Completely randomized design (CRD)
9. Randomized complete block design (RCBD)
10. Split-plot
11. Fixed vs. Random effects
12. Repeated measures
13. Automating repetitive tasks through iteration with purrr
14. Linear regression
15. Non-linear regression
16. Regression for finding optimum
17. Machine learning concepts
18. Bias-variance trade-off
19. Data split
20. Hyperparameter optimization
21. Predictive assessment
22. Machine learning models
23. K-means (unsupervised)
24. Conditional inference tree/Random forest (supervised, regression and classification)
25. Dashboards
26. Creating a simple dashboard with shiny apps
27. Publishing a dashboard online

The topical outline is a general plan for the course; deviations announced to the class by the instructor may be necessary.

# Course materials

## Textbook

A textbook is not required. Reading materials will be supplied by instructor and will include benchmark research articles, manuals, and other materials.

## Technology and software requirements

Students will need to have access to:

* A computer (to install software, code along with instructors)
* A second screen (main screen to code along, second screen to watch class if not in person)

If a student does not have access to these resources (personal laptop/desktop and a second screen), please let instructors know to ensure proper accommodations can be made.

## Course website

Important links related to this course:

* [Course website](https://leombastos.github.io/bastoslab/teaching/2023-apa/2023-apa.html)
* [Course GitHub repository](https://github.com/leombastos/2023_AdvPA)
* [Course YouTube channel](https://youtube.com/playlist?list=PLcdv9xoF_6TFzITEruQ1JnP54FdNO4PtI)

# Assessment and Grading

## Grading categories

The grade you receive in this course will be determined from your performance on a mid-term project, a mid-term exam, periodic quizzes, homework assignments and lab reports, a final project, and class participation. These factors will be weighted as follows:

| Activity | Grade |
| --- | --- |
| Mid-term project: Experimental data analysis | 10% |
| Mid-term exam | 10% |
| Homework assignments | 35% |
| In-class quizzes | 15% |
| Final project:Machine learning | 20% |
| Class participation | 10% |

## Written assignment quality

Up to **thirty percent of the grade** on written assignments (mini-project, homework, final project) will be based on quality of communication.

Spelling, grammar, punctuation, and clarity of writing are evidence of written communication quality.

## Class participation

Active class participation is important for you to achieve the learning goals of the class. To receive maximum credit for class participation you must

* attend every class period (lecture or laboratory)
* arrive on time and remain for the entire class period
* you are actively engaged and attentive throughout the class period
* participate in the class discussion and ask and answer questions

## Grading scale

Final grades will be assigned as follows:

| Letter | Grade |
| --- | --- |
| A | 93 and above |
| A- | 90-92 |
| B+ | 87-89 |
| B | 83-86 |
| B- | 80-82 |
| C+ | 77-79 |
| C | 73-76 |
| C- | 70-72 |
| D+ | 67-69 |
| D | 63-66 |
| D- | 60-62 |
| F | 59 and below |

## Extra credit opportunities

Extra credit opportunities may be made available during projects, homework assignments, and exams, at the discretion of the instructor.

# Course statements and policies

## Academic honesty

UGA Student Honor Code: “I will be academically honest in all of my academic work and will not tolerate academic dishonesty of others.” A Culture of Honesty, the University’s policy and procedures for handling cases of suspected dishonesty, can be found at www.uga.edu/ovpi.

For this course, all lab reports, projects, and other assignments can be discussed with your classmates but any work you turn in must be your own.

Students can work together through coding exercises, but direct copying and pasting from a colleague will be considered plagiarism.

If using code from an online source, it is ok to copy and paste IF proper credit is given (e.g., showing the website source from where the code was obtained).

Unless explicitly stated, artificial intelligence-based technologies, such as ChatGPT, must not be used to generate responses for student assignments.

## Attendance policy

Students are expected to attend every class period.

Students on the **Athens campus** must attend class in-person. If a special circumstance arise (illness, travel, etc.), student absence or remote attendance **must be informed to instructors prior to that class period**.

Students on the **Tifton and Griffin campuses** may attend class in-person on their campuses or remote using the zoom link information. Student absence **must be informed to instructors prior to that class period**.

Per Board of Regents policy, I reserve the right to drop students from the class roll who **miss more than 5 class periods unexcused**. Such students will be given a WF grade.

## Disclaimer

The course syllabus is a general plan for the course; deviations announced to the class by the instructor may be necessary.

## Make-up procedures

* There will be no make-ups for missed quizzes. Any missed quiz will be recorded as a zero
* Exams can be made up only with a note from a doctor or if you can document extenuating circumstances. Any unexcused missed exam will be recorded as a zero
* Homework assignments will be accepted up to one week beyond the due date. The penalty for submitting a late assignment is one letter grade
* Homework assignments may be submitted late without penalty in case of illness, extenuating circumstances, or if prior arrangements are made with the instructors. All late assignments are due within a week of the original due date or within a week of when a student returns from an illness

## Mental Health and Wellness Resources

* If you or someone you know needs assistance, you are encouraged to contact Student Care and Outreach in the Division of Student Affairs at 706-542-7774 or visit https://sco.uga.edu/. They will help you navigate any difficult circumstances you may be facing by connecting you with the appropriate resources or services.
* UGA has several resources for a student seeking mental health services (https://www.uhs.uga.edu/bewelluga/bewelluga) or crisis support (https://www.uhs.uga.edu/info/emergencies).
* If you need help managing stress anxiety, relationships, etc., please visit BeWellUGA (https://www.uhs.uga.edu/bewelluga/bewelluga) for a list of FREE workshops, classes, mentoring, and health coaching led by licensed clinicians and health educators in the University Health Center.
* Additional resources can be accessed through the UGA App.

## Disability statement

If you plan to request accommodations for a disability, please register with the Disability Resource Center. They can be reached by visiting Clark Howell Hall, calling 706-542-8719 (voice) or 706-542-8778 (TTY), or by visiting https://sitedrc.uga.edu

# Resources

Below there are some resources for students to further your knowledge in topics ranging from using quarto files, vector and raster manipulation in R, data visualization, and geostatistics.

* [quarto guide](https://quarto.org/docs/guide/)
* [Geocomputation with R book](https://geocompr.robinlovelace.net/index.html)
* [Introduction to Spatial Data Programming with R book](http://132.72.155.230:3838/r/index.html)
* [ggplot2 cheatsheet](https://github.com/rstudio/cheatsheets/blob/main/data-visualization-2.1.pdf)
* [dplyr data transformation cheatsheet](https://github.com/rstudio/cheatsheets/blob/main/data-transformation.pdf)
* [tidyr data tidying cheatsheet](https://github.com/rstudio/cheatsheets/blob/main/tidyr.pdf)
* [sf cheatsheet](https://github.com/rstudio/cheatsheets/blob/main/sf.pdf)
* [stars documentation](https://r-spatial.github.io/stars/)
* [Colorblind-friendly palettes in R](https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html)
* [Violin/density plot vs. boxplots](https://blog.bioturing.com/2018/05/16/5-reasons-you-should-use-a-violin-graph/)
* [File naming conventions](https://speakerdeck.com/jennybc/how-to-name-files?slide=1)
* [Workflow maintenance](https://speakerdeck.com/jennybc/zen-and-the-art-of-workflow-maintenance)