NC STATE UNIVERSITY

ST 558 Course Syllabus

Data Science for Statisticians

Table of Contents

NC STATE UNIVERSITY

INSTRUCTOR INFORMATION

Name	Office Phone	Email	Office Location
Justin Post	919-515-0637	jbpost2@ncsu.edu	SAS Hall 5272

Virtual Office Hours

My zoom office hours are Wednesdays, 10-12pm EST and Saturdays, 11:30-12:30pm EST.
If these times don't work for you, please let me know and we can try to find a time to meet individually!

The teaching assistant (TA), David Elsheimer, will have zoom office hours on Monday/Wednesday from 6-8pm EST or by appointment.

Preferred Method of Communication

The general discussion board is the preferred method of communication for content questions and email for everything else.

Response Time

You should receive a response within one business day at the latest (generally it should be shorter than that during business hours).

Weekly Update

I'll send a weekly reminder email to help everyone keep on track and communicate any common problems seen across assignments.

COURSE INFORMATION

Course Website: https://wolfware.ncsu.edu/courses/my-wolfware/
Course Credit Hours: 3

Meeting Time and Tool Used

No in person meetings or synchronous activities (other than optional zoom office hours)

NC STATE UNIVERSITY

Prerequisites/Corequisites

ST 555 (or equivalent), ST 511/ST 513/ST 517 (or equivalent)

COURSE OVERVIEW

Catalog Description

Methods for reading, manipulating, and combining data sources including databases. Custom functions, visualizations, and summaries. Common analyses done by data scientists. Methods for communicating results including dashboards. Regular access to a computer for homework and class exercises is required.

Structure

The course is completely **asynchronous**, which means that students have no real-time class meeting requirements. Instead students will watch videos, complete and upload homework assignments/projects, and participate in discussion board posts. Students should set aside sufficient time in their schedules to complete these materials.

To obtain course help there are a number of options:

- General Discussion Board - This should be used for any question you feel comfortable asking and having others view. The TA, other students, and I will answer questions on this board. This will be the fastest way to receive a response! **You should refrain from posting large portions of your programs in questions or responses. Stick to the most relevant line or lines of code.** Repeated violations of this policy will be considered an academic integrity violation.
- E-mail - If there is a question that you don't feel comfortable asking the whole class you can use e-mail. The TA and I will be checking daily (during the regular work week).
- Zoom Office Hour Sessions - These sessions can be used to share screens and have multiple users. You can do text chat, voice, and video. They are great for a class like this!

There are four broad units of the course:

- R as an Environment for Data Science
- Data, Collaboration, and Efficiency
- Statistical/MachineLearning
- Scaling & Sharing Analyses

These sections each have their material broken up into subtopics that generally span a week. Dates are provided and hopefully this will help everyone stay on track.

Each subtopic will usually have videos to watch, optional readings, and an assignment to put what you've learned into practice. The notes and code used for each video are available as well.

NC STATE UNIVERSITY

LEARNING OUTCOMES

Course Learning Outcomes (COs): At the end of this course students will be able to

- explain the steps and purpose of programs (CO 1)
- efficiently read in, combine, and manipulate data (CO 2)
- utilize help and other resources to customize programs (CO 3)
- write programs using good programming practices (CO 4)
- explore data and perform common analyses (CO 5)
- create reports, web pages, and dashboards to display and communicate results (CO 6)

COURSE MATERIALS

Required Textbook and/or Software

Textbooks: None - we'll use some free online texts and articles

Software: Students in this course will use the R statistical software and the R Studio IDE. This software is open source, works on all major platforms, and is free to anyone. It is widely used in statistics and data science (behind python in data science but still quite popular).

TECHNOLOGY REQUIREMENTS

Hardware

NC State's Online and Distance Education provides technology requirements and recommendations for computer hardware.

Software

> Moodle and Wolfware
  - Moodle Accessibility Statement
  - Moodle Privacy Policy
  - NCSU Privacy Policy

> Adobe Reader (for reading PDF files)
  - Accessibility Statement
  - Adobe Privacy Policy

> Zoom:
  - Zoom Accessibility Statement
  - Zoom Privacy Policy

> G Suite
  - Accessibility Statement
  - Privacy Policy

> Office 365
  - Accessibility Statement

- Privacy Policy

> R Statistical Software and R Studio for programming
> Miktex distribution for creation of PDF files
> Docker for creating containers
> Github for version control and collaboration. We recommend the use of the github rather than NCSU's github.
> Headsets with microphone (optional for office hours)

Computer Skills and Digital Information Literacy

The prerequisite coursework requires that students have completed a programming course. This implies that students are expected to have a strong understanding of how computers function and the logic required to instruct them.

Often in homework assignments students will be asked a few questions that require outside of class material. This implies that students are expected to be able to search through online help forums, vignettes, and the like to distill and extract relevant information to solve problems.

NETIQUETTE

Netiquette is the term used to describe the special set of rules for online communication.

Students should be aware that their behavior impacts other people, even online. I hope that we will all strive to develop a positive and supportive environment and will be courteous to fellow students and your instructor. Due to the nature of the online environment, there are some things to remember when taking an online course and engaging with others.

Tips for Success:

> **Do**: Follow the same standards of behavior that you subscribe to offline. Keep in mind that all online communication is documented and therefore permanent.
> **Don't**: Flame others in discussion forums. Flaming is the act of responding in a highly critical, sarcastic, or ridiculing manner – especially if done on a personal level. Remember that these discussions are meant for constructive exchanges and learning!

> **Do**: Ensure you are using forums to get help when you are struggling! It is ok to reach out to help after giving a good faith effort.
> **Don't**: Go for long periods of time without communicating to your instructors or classmates. It is important to stay a part of the online community!

> **Do**: Remember to read over your posts before selecting "Submit."
> **Don't**: Use slang, poor grammar, and other informal language in discussion forums or email messages to instructors or classmates (if possible).

Additional resources

> DELTA's Netiquette or Best Practices for Teaching Online
> Netiquette – Ethics in Computing

NC STATE UNIVERSITY

GRADING

Grading Policy

It is the student's responsibility to be aware of their grades in the course and the appropriate level of work required. Your final grade in this course will depend on the following (subject to change with notification):

Item	Portion of Grade
Video questions	**0%**
Blog Posts	**8% in total**
Homework (Generally programming assignments with discussion questions included)	**17% in total (lowest score dropped)**
Projects (3)	**15% each** **3rd project acts as the final for the course**
Exams (2)	**15% each**

Videos questions: The videos will have questions throughout to keep you active in your learning and to give you a check of how well you are grasping the material presented.. These are not a part of your formal grade.

Blog Posts: There are four short blog posts and a required test/course reflection discussion forum. These are all graded for completeness (that is, if you've responded appropriately to all the prompts you will get full credit).

Homework: There will be many small homework assignments to get you programming a lot. The best way to learn a programming language is to do it! There will also be some explanation questions included. These will each have equal weight. The lowest (%-wise) homework will be dropped. As such, no late work will be accepted so please get your submissions in on time. Generally, each part of each question will be graded on a 3 point scale:

- 3 points for completely correct
- 2 points for a minor error
- 1 point for a major error or a number of minor errors
- 0 points for multiple major/minor errors

Projects: There will be 3 larger projects - some of these will require collaboration with others (data science is often a collaboration!). Details will follow as we get closer to the projects being assigned. Grading guides/rubrics will be provided with the assignments. No late work will be accepted for these.

NC STATE UNIVERSITY

The "final exam" for the course will be the third project.

Exams: All exams are **closed book and closed notes**. The exams do not require a proctor, but you are not allowed to use any material during the exam and you cannot discuss the exams with other students. The exams will cover some programming concepts in R, but I will try to refrain from testing you explicitly on minute details of syntax (unless it is something deemed important enough to know off the top of your head). In addition to R, data science concepts, algorithm ideas, and other topics from the class are fair game.

Students who are unable to take an exam during the designated exam window (see schedule) for a legitimate unavoidable reason may be given the opportunity for a make-up exam or be given a reweighting of their scores. The exams are limited to 120 minutes. Many of the questions will be multiple choice, fill in the blank, or true/false. Other questions will be in essay form. These are generally graded on a 0/1 or 0/1/2/3 basis similar to the homework grading.

Feedback Schedule: Some of the learning activities/assessments (video questions and parts of the exams) are automatically graded. For the videos this feedback should be immediate. For the exams, feedback should be available after the exam timeframe closes. You should double check your wrong answers to make sure there aren't any formatting issues causing a correct answer to be marked as incorrect.

The assignments, projects, and some exam questions, however, are manually graded. My goal as an instructor is to provide feedback and a grade within 5 business days of when you turn these assessments in. However, sometimes it might take a little longer (especially for projects).

Grading Scale

This course uses this grading scale:

Low	Letter	High
97 ≤	A+	≤ 100
93 ≤	A	< 97
90 ≤	A-	< 93
87 ≤	B+	< 90
83 ≤	B	< 87
80 ≤	B-	< 83
77 ≤	C+	< 80
73 ≤	C	< 77
70 ≤	C-	< 73
67 ≤	D+	< 70

63 ≤	D	< 67
60 ≤	D-	< 63
0 ≤	F	< 60

Note: This course may not be taken as an **S/U graded course**.

Students who wish to **audit** the course with satisfactory status must register officially for the course and will be required to obtain an 80% or greater in total on the homework assignments (lowest dropped) and complete at least 2 of the 3 projects with an 80% or better on average to receive credit.

COURSE SCHEDULE

A full detailed schedule is available in the Resources & Information (Start Here!) section.

Please note: course schedule is subject to change.

COURSE POLICIES

Late Assignments

The lowest (%-wise) homework will be dropped. As such, no late work will be accepted so please get your submissions in on time. Projects must be turned in on time. If some issue arises in which you are unable to complete the project prior to the due date, please contact Dr. Post immediately to work out other arrangements.

Incomplete Grades

Incomplete (IN) grades are given only as specified in university regulations.

Attendance and Participation

There are generally no attendance or participation requirements. Attendance is really done through your watching of videos and participation is done through your discussion board posts and assignments.
NC State's Attendance Policy: *https://policies.ncsu.edu/regulation/reg-02-20-03-attendance-regulations/*
Withdrawl Process: *https://studentservices.ncsu.edu/your-classes/withdrawal/process/*

General

Be nice! Also, some things in this course are going to be hard - technology doesn't always work. Be prepared to troubleshoot, check stackoverflow and other online help resources to get things working. **Many homework assignments will include some new aspects that will require you to problem solve (just like real life!).**

NC STATE UNIVERSITY

Course Assistance

As mentioned previously, to obtain course help there are a number of options:

- General Discussion Board - This should be used for any question you feel comfortable asking and having others view. The TA, other students, and I will answer questions on this board. This will be the fastest way to receive a response! **You should refrain from posting large portions of your programs in questions or responses. Stick to the most relevant line or lines of code.**
- E-mail - If there is a question that you don't feel comfortable asking the whole class you can use e-mail. The TA and I will be checking daily (during the regular work week).
- Zoom Office Hour Sessions - These sessions can be used to share screens and have multiple users. You can do text chat, voice, and video. They are great for a class like this!

If you have technical difficulties, you can email the helpdesk (help@ncsu.edu, 919-515-4357). For specific questions about wolfware, moodle, or other instructional tools, you can contact LearnTech (learntech@ncsu.edu, 919-513-7094)

UNIVERSITY POLICIES

Academic Integrity and Honesty

Students are required to comply with the university policy on academic integrity found in the Code of Student Conduct. Therefore, students are required to uphold the university pledge of honor and exercise honesty in completing any assignment.

Please refer to the Academic Integrity web page for a detailed explanation of the University's policies on academic integrity and some of the common understandings related to those policies.

Students may be required to disclose personally identifiable information to other students in the course, via electronic tools like email or web-postings, where relevant to the course. Examples include online discussions of class topics and posting of student coursework. All students are expected to respect the privacy of each other by not sharing or using such information outside the course.

Students are responsible for reviewing the NC State University PRR's which pertains to their course rights and responsibilities:

> Equal Opportunity and Non-Discrimination Policy Statement and additional references
> Code of Student Conduct
> Grades and Grade Point Average
> Credit-Only Courses
> Audits

Students with Disabilities

Reasonable accommodations will be made for students with verifiable disabilities. In order to take advantage of available accommodations, students must register with the Disability Resource Office at Holmes Hall, Suite 304,Campus Box 7509, 919-515-7653 . For more information on NC State's policy on working with students with disabilities, please see the Academic Accommodations for Students with Disabilities Regulation (REG02.20.01)

Trans-Inclusive Statement

In an effort to affirm and respect the identities of transgender students in the classroom and beyond, please contact me if you wish to be referred to using a name and/or pronouns other than what is listed in the student directory.

Basic Needs Security

Any student who faces challenges securing their food or housing or has other severe adverse experiences and believes this may affect their performance in the course is encouraged to notify the professor if you are comfortable in doing so. Alternatively, you can contact the Division of Academic and Student Affairs to learn more about the Pack Essentials program https://dasa.ncsu.edu/pack-essentials/

Moodle Student Services Block

On the right hand side of the Moodle page there is a student services block that has many useful links for obtaining help. There are academic support services (tutorial center, distance ed help), student support services (counseling, disability resources, etc), technology support services (help with online learning), and the general NCSU help desk linked for your convenience!

COURSE EVALUATIONS

ClassEval is the end-of-semester survey for students to evaluate instruction of all university classes. The current survey is administered online and includes 12 closed-ended questions and 3 open-ended questions. Deans, department heads, and instructors may add a limited number of their own questions to these 15 common-core questions.

Each semester students' responses are compiled into a ClassEval report for every instructor and class. Instructors use the evaluations to improve instruction and include them in their promotion and tenure dossiers, while department heads use them in annual reviews. The reports are included in instructors' personnel files and are considered confidential.

Online class evaluations will be available for students to complete during the last two weeks of the semester for full semester courses and the last week of shorter sessions. Students will receive an email directing them to a website to complete class evaluations. These become unavailable at 8am on the first day of finals.

> Contact ClassEval Help Desk: classeval@ncsu.edu

NC STATE UNIVERSITY

> ClassEval website
> More information about ClassEval

REFERENCES

Below is a list of references for the material in the course

Books:

- Wickham, H. and Grolemund, G. (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media
- Wickham, H. (2015). *Advanced R*. Chapman & Hall/CRC the R series
- R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing
- Wickham, H. (2015). *R packages*. O'Reilly Media
- Bryan, J., the STAT 545 TAs, Hester, J. (2020). *Happy Git and GitHub for the useR.*
- James, G., Witten, D., Hastie, T, and Tibshirani, R. (2017). *An Introduction to Statistical Learning.* Spring Texts in Statistics
- Hastie, T., Tibshirani, R., and Friedman, J. (2017). *The Elements of Statistical Learning.* Springer Series in Statistics

Websites:

- https://medium.com/odscjournal/data-scientists-versus-statisticians-8ea146b7a47f
- https://www.springboard.com/blog/machine-learning-engineer-vs-data-scientist/
- https://www.simplilearn.com/data-science-vs-data-analytics-vs-machine-learning-article
- https://mixpanel.com/blog/2016/03/30/this-is-the-difference-between-statistics-and-data-science/
- https://www.smashingmagazine.com/2014/08/build-blog-jekyll-github-pages/
- https://guides.github.com/activities/hello-world/
- https://cran.r-project.org/
- https://support.rstudio.com/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages
- https://www.r-bloggers.com/consistent-naming-conventions-in-r/
- https://google.github.io/styleguide/Rguide.html
- https://datacarpentry.org/R-ecology-lesson/05-r-and-databases.html
- https://db.rstudio.com/dplyr/
- https://www.w3schools.com/sql/
- https://api.census.gov/data.html
- https://cran.r-project.org/web/packages/censusapi/vignettes/getting-started.html
- https://www.programmableweb.com/apis/directory
- https://www.programmableweb.com/news/how-to-access-any-restful-api-using-r-language/how-to/2017/07/21
- https://gitforwindows.org/
- https://osxdaily.com/2014/02/12/install-command-line-tools-mac-os-x/
- http://www.sthda.com/english/wiki/descriptive-statistics-and-graphics#descriptive-statistics-for-a-single-group
- http://www.parallelr.com/r-with-parallel-computing/

Back to TOC

NC STATE UNIVERSITY

- https://www.r-bloggers.com/scheduling-r-markdown-reports-via-email/
- https://tensorflow.rstudio.com/guide/tfestimators/estimator_basics/
- https://tensorflow.rstudio.com/guide/keras/
- https://www.asimovinstitute.org/neural-network-zoo/?utm_content=buffer09a5e&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer
- https://pathmind.com/wiki/neural-network
- https://medium.com/tmobile-tech/r-can-api-c184951a24a3
- https://medium.com/tmobile-tech/using-docker-to-deploy-an-r-plumber-api-863ccf91516d
- https://docker-curriculum.com/
- https://colinfay.me/docker-r-reproducibility/

SYLLABUS MODIFICATION STATEMENT

Some modifications may be needed to the course as we go along. If the syllabus is modified in a substantial way, students will be contacted via an announcement.