Course Syllabus

Site:	Rose-Hulman Institute of Technology	Printed by:	Eric Reyes
Course:	2021F MA386	Date:	Monday, December 7, 2020, 12:50 PM
Book:	Course Syllabus		

Description

The syllabus contains information about the structure of the course, including the unique grading approach taken in the class. It is important to familiarize yourself with this document and reference it often.

Table of contents

MA386 Statistical Programming

Course Description

Across all disciplines, data is being integrated into several aspects of decision making. Whether for statistical analysis or data science more broadly, skills such as data wrangling, data visualization, and computational analysis are valuable. When working closely with raw data, point-and-click programs can be lacking. Whether for reproducibility or flexibility, statistical computing languages that allow you to script the work (such as R, Python, or Julia) are often necessary. We will introduce the basic structure of these languages (primarily referencing R) and describe their use in the statistical analysis pipeline moving from obtaining data to address a question through communicating the results of your investigation. We will discuss general programming concepts related to popular statistical tools such as simulation, bootstrapping, and modeling. As with all statistical courses, there will be an emphasis placed on statistical literacy (interpretation and clear communication of statistical methods, results, and concepts) and statistical reasoning (defining the need for data to address questions, modeling variability in a process, and choosing appropriate methodology to address a question of interest).

Course Learning Objectives

Data is everywhere, and tools are needed to turn data into information. This course discusses the use of scripting in a statistical computing language to aid in this process. At the end of this course, students will be able to perform the following tasks:

(A) **Associate** a computational task with the appropriate function(s) or package(s) (suites of related functions) in a statistical computing language.

(B) Given a script, **describe** the computational task being performed.

(C) Given a computational task from the workflow of a statistical project, **construct** a script to complete the task.

(D) Given a research objective, **integrate** multiple computational tasks to provide a data-driven conclusion.

(E) **Communicate** the solution to a computational task, identifying and describing key steps/chunks within the solution.

(F) **Express the value** of scripting a solution to a computational task.

(G) **Identify resources** that generalize the material covered in class in order to learn new tools for solving a novel computational task.

Course Prerequisites

MA223 (Engineering Statistics I) or MA382 (Introduction to Statistics with Probability) *and a* previous programming course (any language). In particular, students should be familiar with basic statistical inference (confidence intervals and p-values), the bootstrap procedure, descriptive statistics (histograms, boxplots, means, medians, etc.), and essential elements of programming (for-loops, conditional statements, functions, etc.).

Contact Information

Eric Reyes, Ph.D.
Statistician
Associate Professor | Department of Mathematics

Office: DL-110, D (inside Engineering Management Suite)
Phone: 812.877.8287
Email: reyesem@rose-hulman.edu

Contacting the Instructor:
I do not have regular office hours dedicated to this class. If you have questions, feel free to email or message me on Teams. I keep my Outlook calendar and Teams status up to date. Alternatively, you may always email me to schedule an appointment.

I will respond to email within 1 business day; I am not available on the weekends, as this is reserved for my family.

Instructor Biography

After graduating from Rose-Hulman Institute of Technology in 2006 with a degree in Mathematics and Economics, I attended graduate school at North Carolina State University where I earned my Ph.D. in Statistics under the direction of Dr. Dennis Boos and Dr. Len Stefanski. My primary interest is biostatistics - the application of statistical methodology to medical research. As a former participant in the NHLBI Integrated Biostatistical Trainee Program for CVD Research, I spent five years as an intern at the Duke Clinical Research Institute serving as a statistical consultant under the direction of Karen Pieper.

My research interests include methods for variable selection (the process of discerning which variables are useful for predicting a response) and statistics education (how to teach my discipline in a way that gives the best student learning experience).

I was hired into the math department to teach primarily statistics. I believe it is important for every student on campus to be statistically literate. And, I believe that the ability to script an analysis is one of the most useful skills for an engineer or scientist in today's world (and is vital for anyone interested in data science or statistics). In addition to teaching statistics, I am the faculty adviser for the InterVarsity Christian Fellowship chapter on campus.

Mode of Delivery

This course is a hybrid, flipped class. That is, some sessions will be held face-to-face, while other engagement will take place online. Lecture content will be delivered remotely. An in-class meeting will be held once each week for additional learning activities. The bulk of your time in this course will be spent on completing course assignments.

In these types of courses, it can feel like you are "teaching yourself." The activities are meant to structure your learning. In addition, the best way to learn programming is to actually do it (often through trial and error and always through struggling with development). The course has been structured to provide ample time for working through these struggles, while having the instructor as a resource for completing these assignments.

Textbook

The course will primarily make use of **R for Data Science** (by Grolemund and Wickham). This text is freely available online and can be accessed using the above link or from the link on the main Moodle course page.

Lectures will focus on broad concepts common to all statistical computing languages. You may choose to use any statistical computing language during the course; however, the instructor and additional resources provided will support/illustrate R (hence the choice of textbook). For those wishing to use Python, we recommend the **Python Data Science Handbook**; for those wishing to use Julia, we recommend **A Comprehensive Tutorial to Learn Data Science with Julia from Scratch**; and, for those wishing to use SAS, there are few free resources, but the instructor could recommend some texts if interested. Whichever language you choose to use, you will get the most out of this course if you adopt a single language throughout the term. While in practice, analysts commonly move between multiple tools to complete a task, the pace of the course is best addressed by focusing on the new concepts in a familiar language.

Software

- R: This freely available statistical software package is used in academia and industry. While it focuses on statistical analysis, there are a suite of packages making it useful for every step in the statistical analysis pipeline.
- RStudio: This is an integrated development environment (IDE) for R that nicely integrates some additional features we will be exploring in the course. While the use of the IDE is not required, it greatly simplifies some of the operations discussed with regard to reproducible research.

Instructions for Installing the Course Software are made available in the Getting Started document on the main Course Moodle page. If you choose to use a different statistical computing language, you are responsible for determining appropriate software for completing the tasks in the class. The instructor may not be able to provide support for other software.

Grading Procedures

Statistics is a unique discipline in that it exists solely to aid decision-makers in other fields. In this role, a proper statistical analysis has the potential to facilitate real change as it informs researchers. However, a wonderful analysis is useless if the data is constructed inappropriately or the program written to construct the analysis does not function as intended. That is, a substandard program can have significant consequences. This course is designed to reflect this practice and to encourage you to perform at your best throughout the term. As a benefit, when you leave this course, you will have developed a portfolio of high-quality examples to showcase your proficiency with managing data and writing programs as part of a statistical analysis pipeline. This can go a lot further with potential employers than a line item on your resume stating you took a specific course.

In order to help you achieve the objectives of the course, I will be implementing "specifications grading." That is, instead of taking a weighted average of points earned on a series of assignments throughout the term, you will earn grades based on the requirements you choose to complete (described below). Each assignment is graded pass/fail (meaning no partial credit is awarded). In order to help you achieve a passing score, very clear expectations will be provided on every assignment. You have complete control over your grade in the course; the instructor will provide you with the necessary tools and feedback to help you achieve your desired grade.

Grade **Requirements**
A • Successfully complete 2 *Programming Projects* • Successfully complete 9 *Portfolio Problems* • Successfully complete 9 *Homework Assignments*
B • Successfully complete 1 *Programming Project* • Successfully complete 9 *Portfolio Problems* • Successfully complete 9 *Homework Assignments*
C • Successfully complete 8 *Portfolio Problems* • Successfully complete 9 *Homework Assignments*
D • Successfully complete 4 *Portfolio Problems* • Successfully complete 5 *Homework Assignments*

Plus grades will not be awarded.

Description of Course Assignments

The course is divided into ten modules. The first five modules cover essential topics in scripting an analysis, and the final five modules cover intermediate and advanced topics related to the statistical analysis pipeline. In addition to lectures, reading assignments, and class activities, each module will have several opportunities to assess your progress toward the course learning objectives. Each assessment item is described below.

Homework Assignment

There will be one homework assignment within each module (for a total of 10 assignments). A homework assignment consists primarily of short programming exercises which are straightforward extensions of examples discussed in the lecture material or in-class activities implementing a step in a computational task. These exercises are auto-graded to provide immediate feedback. A homework assignment can additionally consist of a class discussion or additional task related to the application of the course concepts. Be sure to review the **Criteria for Successfully Completing a Homework Assignment**.

Portfolio Problem

In practice, a computational task involves integrating several smaller steps in the presence of a rich context. Further, it is required to provide some minimal amount of communication involving the proposed solution. Portfolio problems provide an opportunity to hone such skills. There will be one Portfolio Problem assigned in each module (for a total of 10 assignments). Be sure to review the **Criteria for Successfully Completing a Portfolio Problem**.

Programming Project

Moving from data to insights requires linking several computational tasks; further, communication of both the solution and the insights is paramount for collaborating with others. Programming projects are a recognition of mastery of course content. There will be two such projects in the course; one project will be due around midterm, and the other will be due at the end of the term. Be sure to review the **Criteria for Successfully Completing a Programming Project**.

Tokens

I recognize that life happens. We each have our individual priorities, and these may not align with the course schedule and requirements at all times. In those moments, the pass/fail nature of the course may be overwhelming. Here are two tips for being successful in this course:

1. Begin assignments early in order to provide an opportunity to ask intelligent questions of the instructor when needed. *It is unlikely that you will be successful in this course if you begin assignments the night before they are due.*
2. Change your mindset about obtaining grades. Success in a specifications-graded course requires that you master content. That is, "knowing a little about a lot of things" will not equate to success. Passing the course does not require engaging with every module of the course, while obtaining a C does.

Even with the best intentions, we may struggle to begin assignments early. In these cases, you make use of tokens. Tokens create a type of barter economy in the class. You can trade your tokens for leniency on the course policies. Each student will be given 1 token at the beginning of the term. You will have an opportunity to earn an additional token for each of two optional assignments in the class; so, you have the opportunity to have at most 3 tokens in the course. The value of a token is described below.

Commodity **Price**
Automatic "successful completion" of a *Homework Assignment* 1 Token
Automatic "successful completion" of a *Portfolio Problem* 2 Tokens
Revision of a *Portfolio Problem* 1 Token
Minor revision of a *Programming Project* 1 Token
Major revision of a *Programming Project* 2 Tokens

Access to Grades

All assignments will be submitted electronically through an associated Moodle dropbox. Feedback will be placed in the same dropbox, and all grades can be viewed in the Gradebook on Moodle.

Late Work

Late assignments will not be accepted. Due to the nature of the course, incomplete assignments will most certainly not meet the requirements for a successful completion. If you will not be able to meet a deadline, you should contact the instructor as soon as possible via email to discuss your situation.

Attendance Policy

The majority of this course will be completed remotely. The face-to-face sessions will provide a chance to practice skills and build on the lecture content. While you are encouraged to attend, there is no penalty for missing a face-to-face session. *If you are feeling ill, you **should** remain home.* This is both for your benefit as well as others in the class.

Technology and Communication Policy

As outlined in the Student Handbook, Rose-Hulman students are expected to be responsible adults and to behave at all times with honor and integrity. In this class, this includes the following:

- Images should be associated with the content of the assignment and appropriate for a general audience (I should be able to show my children). The display of objects or images which marginalize any group will constitute harassment of others who see these objects or images; this includes sexually suggestive objects or images.
- All communication (whether electronic or in-person), including messages, email, and forum posts, need to be respectful. You should consider not only what was *intended* by the sender with the message but also what may be *heard* by the recipient.
- If you are a recipient of or a witness to any type of harassment or bullying, either online or in person, please contact your instructor immediately.

Academic Misconduct Policy

Expectations for upholding academic integrity, and the importance of academic integrity within the institution, as well as the department policy on academic integrity, can be located in the last chapter of the course syllabus (Institute Policies). In this section, we simply outline the penalty for academic misconduct. If a student commits academic misconduct on an assignment, the following action will be taken by the instructor:

- A letter will be sent to the student, the Head of the Department of Mathematics, and the Dean of Students outlining the incident and the penalty applied.
- The student's grade will be capped at a D in the course, regardless of the assignment on which the misconduct occurs. This policy is to ensure that academic misconduct on a *Portfolio Problem* is not more severe than misconduct on a *Programming Project*.

This policy is consistent the Mathematics Department's Academic Integrity Policy. To ensure you are not found in violation of the standards of academic integrity, you are encouraged to read the **Student Handbook**.

Guidelines for Appropriately Working Together (Avoiding Misconduct)

If you work with someone on a *Portfolio Problem* or *Programming Project*, you must cite the name of the collaborator(s) within the document and explicitly discuss their contribution to the work. Similarly, if you use an outside resource (something other than the provided class materials), you should cite your source. These are the basic guidelines. Here are some additional tips which can help draw the distinction between collaboration and academic misconduct:

- If you find an existing post helpful on stackExchange (or similar forum) explaining a topic or troubleshooting an error in your code, it need not be cited. If you create a post on such a forum and receive a direct reply, it should be cited and the contribution discussed in your work. If the existing post provides a solution to the specific task you are performing, then it should be cited.
- It is okay to *discuss* solutions with one another, but you should not be looking at someone else's work when you write up your own solution.
- It is helpful to avoid writing down solutions in the presence of other people - instead, outline the steps and stop there.
- While it is common to write things on a whiteboard, consider instead chatting and each person taking their own notes on the conversation; that way, you are always working on your own notes and write-up.
- *Never* look at someone else's solution; instead, if you are asking them a question, talk about it together.
- Instead of asking "how did you do this problem?" ask questions about specific steps: "what was the big idea?" "what functions did you rely on in order to filter the dataset?"

If you are in doubt, work alone. No assignment is worth jeopardizing your career at Rose-Hulman or your reputation.

Attendance Policy

The success (or failure) of Rose-Hulman's coronavirus-containment measures depends on each one of us. As part of this effort, your grade in the course is not directly linked to your attendance in class. I encourage you to shift from thinking of "being in class" as your goal and instead "engaging with course content." Short version: if you are experiencing any symptoms, believe you were exposed to COVID-19, or have been asked to quarantine or isolate, please do not come to class. In class activities are meant to help explain course content; but, no new information is shared during those times. The other course meeting is just a time to engage with the professor virtually if you have questions on the assignments.

Here are a few additional things to keep in mind regarding your attendance in class.

- Be vigilant and closely monitor your health; be honest about any worrying symptoms.
- Even if you or others will be inconvenienced, even if you worry about what you'll miss in class, *stay home* and contact your professors *via email* if you have any symptoms.
- It is more important than ever to start assignments early; don't allow yourself to fall behind.
- Let go of perfectionist expectations of yourself and others; be open-minded about course delivery, virtual office hours, remote gatherings with classmates, etc.

Expectations for Professionalism in a Group

Whether working in industry or academia, engineers, scientists, and analysts will be required to collaborate with others at some point. To maximize learning of some concepts in class, there may be times when you are encourage or required to work within a group. Group members will have different levels of mastery of the course content and different skill sets. As a result, a healthy group does not necessarily mean "equal participation" of group members. However, it does recognize that every group member is "equally valuable." Every member is responsible for the work submitted by the group; therefore, every member is expected to act professionally at all times - which includes active participation in meetings and aiding in the construction of final documents. At a minimum, professionalism includes:

- No one is done until everyone is done.
- You have the right to ask anyone in your group for help.

- You have the duty to assist, to the best of your abilities, anyone in your group who asks for help before turning to the instructor.
  - Helping peers means explaining thinking, not giving answers or doing the work for them.
  - Provide justification when you make a statement to another member of the group.
- If the group has a question, it is appropriate to ask the instructor.
  - The group has a question when everyone shares the question.
  - The group has a question when an individual is uncertain after consulting with other members of the group and receiving their responses.

You are expected to resolve typical issues of group dynamics; the instructor can help with this as needed. If you feel a member of your group is not acting in a professional manner to the extend that you do not feel comfortable with that individual's name attached to the work submitted, please contact the instructor in writing (via email) immediately. If, after investigating, the instructor determines that a student has a pattern of unprofessionalism (negligence) within a group, the following action will be taken by the instructor:

- A letter will be sent to the Head of the Department of Mathematics outlining the incident.
- The student's final course grade will be reduced by 1 letter grade (A to B, C to D, etc.).

In a remote learning environment, the key to success is regular engagement with course material, and that is accomplished through remaining disciplined with your schedule. Here are the biggest lies you can tell yourself about this course:

1. *I can get everything done over the weekend*. The truth is that you might get the work done, but the work will most likely not be of sufficient quality to earn a successful completion of a Portfolio Problem. If you wait until the last minute to begin assignments, you will most likely find yourself saying "I understood it when I read it, but my code is not working now." Coding requires iteration; you need to save time to revise your initial work.
2. *I just need to find an example*. We learn to program by writing programs, not copying others' programs. That is not to say we never look on stackoverflow or in a text for an example. We should rely heavily on past accomplishments and the vast resources at our disposal. But, if your solution to every problem is to Google "how do I ..." then you will never develop the skills to solve the problems on your own. Work through the guided notes and attempt the problems; learn the building blocks so that you can assemble your own solutions.

At a minimum, you will be expected, *each week*, to put 2-3 hours into reading and guided notes, an additional hour on learning activities in class, 1-2 hours on homework assignments, and an additional 2-4 hours working on the portfolio problem; for those that intend to complete the course projects, you should also budget an additional 1-2 hours each week on these projects. This includes time spent in class. Each student has different commitments during the term. Most of you have a full course load; some are working; others are involved in groups across campus; some may be caring for family members. And, let's face it, this term looks a bit different than the Rose-Hulman you remember or expected, and we may experience some significant disruptions! Therefore, the schedule that works best for you will vary, but I can offer you some advice.

Below is a rough schedule of the course activities and due dates.

	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Weekend**
Week 0				Introduction to Course	In-Class Activity	
Week 1		Workshop Hours		Module 1 Assignments Due	In-Class Activity	
Week 2		Workshop Hours		Module 2 Assignments Due	In-Class Activity	
Week 3		Workshop Hours		Module 3 Assignments Due	In-Class Activity	
Week 4		Workshop Hours		Module 4 Assignments Due	In-Class Activity	
Week 5		Workshop Hours		*Fall Break*	*Fall Break*	*Fall Break*
Week 6	Module 5 Assignments Due Project 1 Due	In-Class Activity			Workshop Hours	
Week 7	Module 6 Assignments Due	In-Class Activity			Workshop Hours	
Week 8	Module 7 Assignments Due	In-Class Activity			Workshop Hours	
Week 9	Module 8 Assignments Due	In-Class Activity			Workshop Hours	
Week 10	Module 9 Assignments Due Project 2 Due	Workshop Hours			Module 10 Assignments Due	

Based on this, here are my recommendations:

- Complete approximately half of the reading and guided notes *prior* to the In-Class Activity. The In-Class Activities will be examples, often worked in pairs or as a group, of the material. The more exposure you have to the material in advance, the more you will get out of these activities.
- Complete the remainder of the reading and guided notes on the same evening as the In-Class Activity.
- Begin the module assignments the day following the In-Class Activity.
- Use the Workshop Hours as a chance to ask questions of the instructor regarding your assignments. Do your best to finish up the assignments that same day.
- Work on the Course Projects *regularly* throughout the term; these assignments are *not* meant to be completed in a single week.

There is definitely a rhythm to the course (albeit Fall Break messes with that a bit). It helps to develop a consistent routine during the first module and continue that through the term.

Students with Accessibility Needs

Rose-Hulman is committed to working with students who have special needs or disabilities. Such students may be eligible to receive accommodations that provide equal access to learning, the living and learning environment, and college activities. Visit the **Accessibility Services website** for more information. Requests for academic accommodations must be documented with and approved by the Accessibility Services office before they can be implemented in this course.

Emergency Information

To receive email or text messages regarding emergency situations that may impact campus and, possibly, the delivery of classes, register for RAVE alerts and/or follow @Rose-HulmanAlert on Twitter. Any announcements about the Institute's ability to offer classes will be shared on Rose-Hulman's website.

Student Handbook

This course adheres to all policies described in the **Student Handbook**. A few key sections are briefly outlined below. In brief, Rose-Hulman expects its students to be responsible adults and to behave at all times with honor and integrity. All students are expected to abide by this code and to aid in its enforcement by reporting violations of it.<\p>

Dropping the Course

You are responsible for understanding the university's policies and procedures regarding withdrawing from courses found in the current catalog. You should be aware of the current deadlines according to the **Rose-Hulman Academic Calendar.** More information for Drops and Adds can be found on the **Registrar's site here**.

Academic Integrity

Academic integrity is an integral part of the Rose-Hulman community. It is important that all members of our community learn to properly acknowledge the important contributions of others in our respective fields, both within Rose-Hulman and external to Rose-Hulman. Understanding how to work in collaboration with others and how to incorporate their work into your own, and then appropriately acknowledging them, demonstrates your intellectual maturity and a high degree of professionalism. Academic integrity refers to maintaining a high standard of honesty in academic conduct. All students and faculty are encouraged and required to show academic integrity at all times. On the other hand, academic misconduct is a failure of academic integrity. Specifically, academic misconduct is cheating, plagiarism, or interfering with the academic progress of other students.

The **Academic Rules and Procedures document** provides extensive rules and procedures for academic and other misconduct. The Mathematics Department **follows these rules seriously**. The minimum penalty for such misconduct is for the instructor to award zero credit for whatever test, exam, project or quiz on which the misconduct occurs, even if it results in a lowered or failing grade. A report of the misconduct will be sent to the Dean of Students and the Mathematics Department Head. Faculty members may exact a higher penalty, up to and including failure in the course if they feel the misconduct warrants such action. Students may appeal the sanctions to the rules and discipline committee, per the cited web page.

Plagiarism is a serious offense, and students are expected to adhere to the Rose-Hulman policy on plagiarism and cheating. Some individuals might say that they did not understand what plagiarism was when they took credit for someone else's ideas, but ignorance is not an excuse for lack of academic integrity. It is each student's responsibility to know the Rose-Hulman policy on academic honesty, including plagiarism, cheating, dishonest conduct, and collusion. This not only includes misrepresenting others' work as your own, but also summarizing, paraphrasing, use of any other material in your work, and incorrect or incomplete citations and references. Using the same work for multiple courses is also dishonest. If you have any questions concerning rules, procedures, or about academic honesty, plagiarism, cheating, dishonest conduct or collusion, please speak with your instructor.

Course Information Copyrights

You have accessed this document through a system located at Rose-Hulman Institute of Technology using your private and personal authentication information. By doing so, you affirm that you assent to the provisions of the United States Copyright Act, Title 17 of the U.S. Code. Course materials available through this system may be protected by copyright law. This material is only for the use of students enrolled in the specific course(s). Protected materials on this site may not be further disseminated by the user to any other persons.

For further information please refer to **Rose-Hulman Library information on Copyright**.

Diversity Statement

Rose-Hulman is commited to being an **inclusive community** in which the multiplicty of values, beliefs, intellectual viewpoints, and cultural perspectives enrich learning and inform scholarship.

Online Access Requirements

Rose-Hulman welcomes students from around the world and encourages faculty, staff, and students to travel around the world. However, geopolitical conditions and compliance with export law and regulations prevent us from delivering certain kinds of educational experiences and/or using certain kinds of Institute technologies in some locations. For example, students in locations with limited access to the internet in general, or with restricted access to portions of the internet, or which are embargoed by the U.S. Directorate of Defense Trade, may not be able to successfully complete Rose-Hulman courses.

Disclaimer

The instructor reserves the right to modify the course content, schedule, topics, policies, etc. outlined in this syllabus.