QTM 150: Introduction to Statistical Computing I (R)

Instructor
Zachary Binney, PhD
Assistant Professor, Quantitative Theory and Methods
Office: Candler 218
Email: zbinney@emory.edu
Office hours: Tuesdays 4:00-5:00pm, Wednesdays 8:00am-9:00am, Thursdays 1:00-2:00pm (all times U.S. ET)

Time/Location
Friday 9:40-10:40am (Sec. 1)
Friday 11:20am-12:20pm (Sec. 2)
Friday 1:00-2:00pm (Sec. 3)
Zoom Address

Prerequisites
None! All are welcome.

What Will You Learn?
R. We’ll provide an introduction to the free R statistical programming language with a focus on using it to analyze and present data. We’ll cover how to import data, manipulate and reorganize it, and create beautiful and informative graphs and tables to present the data and your insights to others.

Contents
Course Description and Learning Outcomes ....................................................................................... 2
Course Materials ................................................................................................................................... 3
Computing Needs ................................................................................................................................. 3
Course Structure and Online Sessions ................................................................................................. 4
Assignments and Grading ..................................................................................................................... 5
Questions and Communication ........................................................................................................... 7
Learning During a Pandemic ................................................................................................................... 9
Diversity ................................................................................................................................................ 10
Access and Disability Resources ......................................................................................................... 10
Class Session Recording ....................................................................................................................... 11
Academic Integrity ................................................................................................................................. 12
Course Schedule (Fall 2020) ................................................................................................................. 13
Course Description and Learning Outcomes

Course Description
This course provides an introduction to the programming language R for analyzing data. The material and class format are targeted towards turning you into proficient data analysts able to actively implement the methods that will be covered in class. This will require you to practice the material both during and outside of class.

Learning Outcomes
By the end of the course, students should (1) be fluent in using R and Rstudio for data analysis, (2) be able to deal with complex real data and seamlessly transform that data into the most convenient form for analysis or reporting, (3) use graphics to explore and understand data, and (4) gain familiarity with basic data collection, storage, and manipulation.

Summary: you’ll be able to take raw data and make cool stuff like this on your own with your laptop!
Course Materials

Primary Course Resources

The primary materials that contain the information you will be responsible for on assignments and quizzes will be the Tutorials and videos on Canvas, custom-made for this class from various resources.

Textbook

*R for Data Science* (R4DS) by Hadley Wickham and Garrett Grolemund.

Cost: free! It’s available online here: [https://r4ds.had.co.nz/](https://r4ds.had.co.nz/)

You don’t have to have this – it’s not required – but a.) it’s free and b.) I strongly recommend you review the sections I point out that go with each class tutorial, especially if you have no prior programming experience. The reinforcement will really help, and the book is quite accessible (even fun!).

Other Materials for Learning R

There is a *universe* of free R materials out there – free online textbooks, written textbooks, websites, online courses (e.g. Coursera), interactive tutorials, and presentations. This is one of the great things about the R (#rstats on Twitter) community. On Canvas I have compiled a *sampling* of other materials you might find useful, along with a brief opinion on them if I have reviewed them. *These are purely optional – even if you wanted to review them all they would end up being pretty repetitive*, so please don’t feel compelled.

If you stumble across any materials I haven’t included here that you find useful, please let me know!

Computing Needs

*It is strongly recommended you have a laptop or desktop (not a tablet) for this course.* You will get the most out of this class if you can install R and RStudio onto your own machine as you’ll be able to use it for data analysis far into the future! R and RStudio are available for Windows, Mac, and Linux, but unfortunately not for tablets or Chromebooks (iOS, Android, or Chrome OS). If you don’t have a laptop or desktop, don’t fear! We’re an accessible class. You can use R and RStudio in the cloud (i.e. through a web browser). **Email me for more details if you need the Cloud option.**
Course Structure and Online Sessions

Weekly Expectations

This course is an online, primarily asynchronous course. What this means is, essentially, that you may work through the materials at your own pace and at a time of your choosing each week.

Each week I will distribute some written interactive tutorial notes. Here is how I recommend you consume these:

1. The written tutorials are your primary resource; please review them and work through the interactive exercises until you feel comfortable with the material.
2. Strongly Recommended: Practice all included code (that is, re-write and run the code yourself in your own script or RMarkdown document separate from the tutorial) on your own or in small groups.
3. Optional: I also recommend you read the relevant sections of R for Data Science that correspond with each tutorial, especially if you’re confused about anything.

Once you feel comfortable with the material, take the weekly open-book quiz on Canvas and complete that week’s portion of the homework assignment. That’s it!

Class Sessions

Notice I said the class is primarily asynchronous. We will meet for 50-minute “synchronous” sessions via Zoom during our scheduled “class time” about 5 times during the semester – once for each major topic area. These dates are indicated on the schedule below. Attendance is optional but encouraged. Why attend?

1. The chance to ask questions and discuss anything you found unclear!
2. Bonus discussions and live-coding of extra things I think are fun that are relevant to the current lesson.
3. Community with me and your fellow students.
Assignments and Grading

Homework (30% of total grade, 6% each split across 5 assignments)

Homework will be assigned regularly throughout the semester, with 1 assignment corresponding roughly to 1 major topic. Occasionally, a homework “assignment” will be split into 2-3 parts, but it will be submitted graded together as 1 assignment. Homework should be generated into HTML via R Markdown and submitted on Canvas. All work should be shown including R code, R output, and written answers to questions.

You are encouraged to talk about homework with others in the class but you have to submit your own work.

All homework is due by 6 pm U.S. ET on its due date (and not 11:59 pm, as you may be used to from other classes). This is to encourage you to not work late into the night – sleep is important! You may get a 1-week, no questions asked extension on 2 of these 5 assignments – I will track these automatically over the semester. You do not need to email me a request.

Weekly Quizzes (40% of total grade, 12 quizzes at 4% each, 2 lowest grades dropped)

Each week there will be a brief Canvas quiz on the material from the tutorials. This is an open-book, open-note, open-Internet quiz. I invite you to review your tutorials, watch my videos again, or Google the topic, but do not search for the wording of the particular question (for example, sites listing specific exam questions are off-limits). You also may not interact with another live person, for example chatting with a friend or tutor, for these quizzes. Anything else is fair game. You may get a 1-week, no questions asked extension on 3 of these 12 quizzes – I will track these automatically over the course of the semester. You do not need to email me a request.

Project (30%, split into a project pitch (10%) and finished product (20%)):

The culmination of this class is a statistical computing project designed to provide an overview of data analysis through descriptive statistics. Students will be required to identify an interesting dataset and analyze it on their own, providing summary tables and graphics that demonstrate data-based insights and basic R mastery. You will also
be asked to create a brief (<5 minute) video presentation for your classmates and me. These may be completed alone or in groups of up to 4. More detailed instructions will be provided later in the semester. I am limited in my ability to grant extensions on the final project because of grading deadlines.

Below is the grading scale for this class. Grades are rounded to the nearest point (89.50% = 90%, 89.49% = 89%) at the end of the semester. I maintain the right to arbitrarily curve the class at my discretion, but it will only be done to *boost* grades (i.e. make each cutoff easier to reach), never the reverse. This class is not designed to be mean enough to merit a curve, though.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>105%–</td>
<td>92%–</td>
<td>89%–</td>
<td>86%–</td>
<td>79%–</td>
<td>69%–</td>
</tr>
<tr>
<td>93%</td>
<td>90%</td>
<td>87%</td>
<td>83%</td>
<td>76%–</td>
<td>67%</td>
</tr>
<tr>
<td>89%</td>
<td>86%–</td>
<td>82%–</td>
<td>79%–</td>
<td>72%–</td>
<td>66%–</td>
</tr>
<tr>
<td>90%</td>
<td>87%</td>
<td>80%</td>
<td>77%</td>
<td>73%</td>
<td>67%</td>
</tr>
<tr>
<td>86%–</td>
<td>83%</td>
<td>77%</td>
<td>73%</td>
<td>70%</td>
<td>63%</td>
</tr>
<tr>
<td>80%</td>
<td>77%</td>
<td>73%</td>
<td>70%</td>
<td>67%</td>
<td>60%</td>
</tr>
<tr>
<td>79%–</td>
<td>76%–</td>
<td>72%–</td>
<td>69%–</td>
<td>62%–</td>
<td>59%–</td>
</tr>
<tr>
<td>77%</td>
<td>73%</td>
<td>70%</td>
<td>67%</td>
<td>62%–</td>
<td>0%</td>
</tr>
</tbody>
</table>

**Extra Credit (up to 5%)**

There are three sources of extra credit I invite you to avail yourselves of:

1. Answering your classmate’s questions on the Canvas Discussion Board ([more details under Questions](#)). To encourage learning from your peers as well as me, each time you are the first to provide a complete and
correct answer to a course-relevant, non-repeated question, you will get 1% extra credit, up to 2% total (2 total responses). “Complete,” “correct,” “course-relevant,”, and “non-repeated” are all solely determined by the instructor, but you will be informed each time a response counts for extra credit!

2. I invite you to make a meme, a TikTok video, or any other social media output (Tweetstorm, Instagram whatever-that-entails) about something you learned in this course, a link to which is submitted on Canvas by the project due date. This is worth 2% extra credit.

3. So we can all get to know each other, before the second week of class I’d like you to make a brief Flipgrid video introducing yourself, giving us any interesting fact about you of your choice, and telling us why you’re here to learn R. More details will be provided on Canvas. This is worth 1% extra credit.

Questions and Communication

Communication is very important in online classes to foster community and maintain good relations between students and instructors.

First stop – the Internet!

If you have a question about the material try searching around a bit for an answer first on Google or coding help sites such as Stack Exchange (or the statistics version, Stats Exchange). It’s good practice. I do it all the time. All programmers do. The signal that you’ve made it as an R programmer is not that you know how to do everything, but that you know the words and terms to use to search for how to do things you don’t know (or have just forgotten).

If you are struggling to find the answer after about 5 minutes, though, please don’t continue struggling. Ask for help!

Asking for Help

There are 3 options for asking questions in this class, listed in order of preference:

1. Canvas Discussions Board: This is our preferred communication option. It is where all questions about course material and assignments should go. In short, anything that could conceivably be useful to another student should go here. This reduces duplicated efforts, encourages interacting
with your fellow students, and puts everyone on a level playing field. It’s win-win-win!

Please also check the Discussions board to see if someone else had your question and it’s already been answered. That way you get an answer faster.

I understand some students may be shy, but please know: there are no stupid questions, and I have a zero tolerance policy for any bullying or mockery on the boards. We’re all here to learn together and help each other.

2. **Office Hours:** You can always ask questions about or for extra help with the course material in office hours, though I might ask you to consider making a Discussion Board post with what we talked about if it’ll be helpful for others.

Office hours are also a great option if you have questions about material that extends beyond this course, want to talk about careers, mentoring, academic advice (e.g. you’re curious about Quantitative Theory and Methods), or simply want to introduce yourself so I know you better later for things like letters of recommendation or research opportunities.

I didn’t often take advantage of office hours when I was an undergraduate because I didn’t really understand what they were for and slightly regretted it later on.

3. **Email:** The third option. If it’s a question that will be useful to other students, I will ask you to put it on the board instead. Issues appropriate for email include questions specific to your or your group’s assignment or project, or other personal matters that need to not be broadcast to the rest of the class (e.g. accommodation requests).

For emails and Discussion board posts, please allow 24 hours for a response from me, or 48 hours on weekends (your fellow students may be quicker, but don’t take what they say as 100% correct until I sign off on the response). Please plan accordingly.
Learning During a Pandemic

(Text modified from Prof. Andrew Heiss at Georgia State.)

Life is rough right now. None of us is really doing well. We’re all just hanging on, myself included. And there’s a lot of uncertainty about what will happen this fall in particular.

You most likely know people who have lost their jobs, have tested positive for COVID-19, have been hospitalized, or perhaps have even died. You may have altered family care responsibilities and/or you are likely facing uncertain job prospects. If it helps, I graduated straight into the Great Recession in 2008 and came through it OK (though many people had a much rougher time). Seriously, I got hired at a consulting firm 2 months before Lehman Brothers collapsed. So you’re not doomed.

I’m fully committed to making sure that you learn everything you were hoping to learn from this class! I will make whatever accommodations I can to help you finish your exercises, do well on your projects, and learn and understand the class material. Under ordinary conditions, I am flexible and lenient with grading and course expectations when students face difficult challenges. Under pandemic conditions, that flexibility and leniency is intensified.

If you tell me you’re having trouble, I will not judge you or think less of you. I hope you’ll extend me the same grace.

You never owe me personal information about your health (mental or physical). You are always welcome to talk to me about things that you’re going through, though. If I can’t help you, I usually know somebody who can – such as Counseling and Career Services.

If you need extra help, or if you need more time with something, or if you feel like you’re behind or not understanding everything, do not suffer in silence! Talk to me. I will work with you. I promise.

Please sign up for a time to meet with me during office hours. I’m also available through e-mail.

I want you to learn lots of things from this class, but I primarily want you to stay healthy, balanced, and grounded during this *gestures wildly at everything*. 
Diversity

Diversity and Inclusivity

Oxford College of Emory University’s ideals of inclusivity require that we foster an environment where people of diverse backgrounds, identities, abilities, and ideologies are affirmed, respected, and seen as a source of strength; where we strive to learn together, and ultimately thrive communally. If we at all fail to support these ideals, then we encourage discussion towards improvement, and we hope that this statement affirms your right to seek those discussions via dialogue with faculty, staff, and your peers.

I try my best to create a welcoming and inclusive environment for all students regardless of race, gender and gender identity, religion, immigration status, nationality, parental status, or age. If I ever fall short of that, I encourage you to approach me personally after class, during office hours, or via email and help me understand how I can do a better job of that. There are also college-level resources to express concerns if you would prefer not to approach me personally.

Kids in the Classroom

If you’re a parent and you ever have a childcare issue and need to bring your child to class, I completely understand and welcome their presence. Fair warning, though: I retain the right to make funny faces at them.

COVID-19 and Other Long-Term Health Issues

If you are placed under quarantine for COVID-19 or are going through any longer-term health or personal issues that may affect your success in the class, please let me know. In addition – or alternatively, if you’re more comfortable – I invite you to coordinate with the Advising Support Center (ASC) oxacadsvc@emory.edu. They are very helpful.

Access and Disability Resources

As the instructor of this course I endeavor to provide an inclusive learning environment. I want every student to succeed. The Office of Accessibility Services (OAS) works with students who have disabilities to provide reasonable accommodations. In order to receive consideration for reasonable accommodations, please contact the OAS and complete the registration
process. Faculty may not legally provide you with accommodations until an accommodation letter has been processed and discussed with them; accommodations do not start until this point and are not retroactive, so I ask you to let me know about accommodations ASAP so I can best help you.

Students registered with OAS who receive a letter outlining specific academic accommodations are invited to immediately coordinate a meeting with their professors to discuss a protocol to implement accommodations that will (or may) be needed over the course of the semester. This meeting should occur as early in the term as possible. Contact OAS for more information at (770) 784-4690 or oas_oxford@emory.edu.

**Class Session Recording**

Our class sessions on Zoom / our in-person class sessions will all be audio visually recorded for students in the class to refer back to the information, and for enrolled students who are unable to attend live.

Lectures and other classroom presentations presented through video conferencing and other materials posted on Canvas are for the sole purpose of educating the students enrolled in the course. The release of such information (including but not limited to directly sharing, screen capturing, or recording content) is strictly prohibited, unless the instructor states otherwise. Doing so without the permission of the instructor will be considered an Honor Code violation and may also be a violation of other state and federal laws, such as the Copyright Act.

Students who participate with their camera engaged or utilize a profile image are agreeing to have their video or image recorded. If you are unwilling to consent to have your profile or video image recorded, be sure to keep your camera off and do not use a profile image.

Likewise, students who un-mute during class and participate orally are agreeing to have their voices recorded. If you are not willing to consent to have your voice recorded during class, you will need to keep your mute button activated and communicate exclusively using the "chat" feature, which allows students to type questions and comments live.

Please read the [Rules of Zoom Engagement](#) for further advice on participating in our Zoom class sessions.
Academic Integrity

Upon every individual who is a part of Emory University falls the responsibility for maintaining in the life of Emory a standard of unimpeachable honor in all academic work. The Honor Code of Emory College is based on the fundamental assumption that every loyal person of the University not only will conduct his or her own life according to the dictates of the highest honor, but will also refuse to tolerate in others action which would sully the good name of the institution. Academic misconduct is an offense generally defined as any action or inaction which is offensive to the integrity and honesty of the members of the academic community. The typical sanction for a violation of the Emory Honor Code is an F in the course. Any suspected case of academic misconduct will be referred to the Emory Honor Council.
# Course Schedule (Fall 2020)

This is a tentative outline of the schedule and is subject to change.

<table>
<thead>
<tr>
<th>Date</th>
<th>Topics</th>
<th>Tutorials to be Completed by Indicated Date</th>
<th>Assignments Due on Indicated Date</th>
<th>Live Class Meeting (Optional)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-Aug</td>
<td>Intro to QTM150 Installing R and RStudio</td>
<td>0.1</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>28-Aug</td>
<td>Intro to R, RStudio, and R Markdown Create your first visualization!</td>
<td>0.2</td>
<td>Quiz 1</td>
<td></td>
</tr>
<tr>
<td>4-Sep</td>
<td>Opening and Exploring Datasets Data Visualization I</td>
<td>1.1</td>
<td>Quiz 2</td>
<td></td>
</tr>
<tr>
<td>11-Sep</td>
<td>Data Visualization II</td>
<td>1.2</td>
<td>Quiz 3 HW 1</td>
<td></td>
</tr>
<tr>
<td>18-Sep</td>
<td>Data Visualization III</td>
<td>1.3</td>
<td>Quiz 4</td>
<td></td>
</tr>
<tr>
<td>25-Sep</td>
<td>Data Visualization III Catch-up Week</td>
<td>1.3</td>
<td>HW 2</td>
<td>X</td>
</tr>
<tr>
<td>2-Oct</td>
<td>Importing and Exporting Data Exploratory Data Analysis</td>
<td>2.1</td>
<td>Quiz 5</td>
<td></td>
</tr>
<tr>
<td>9-Oct</td>
<td>Data Transformation I</td>
<td>3.1</td>
<td>Quiz 6</td>
<td></td>
</tr>
<tr>
<td>16-Oct</td>
<td>Data Transformation II</td>
<td>3.2</td>
<td>Quiz 7</td>
<td>X</td>
</tr>
<tr>
<td>23-Oct</td>
<td>Data Types I</td>
<td>4.1</td>
<td>Quiz 8 HW3</td>
<td></td>
</tr>
<tr>
<td>30-Oct</td>
<td>Data Types II</td>
<td>4.2</td>
<td>Quiz 9</td>
<td>X</td>
</tr>
<tr>
<td>6-Nov</td>
<td>If-Else Statements For-Loops</td>
<td>5.1</td>
<td>Quiz 10 HW 4</td>
<td></td>
</tr>
<tr>
<td>13-Nov</td>
<td>Writing Your Own Functions Sampling and Simulations</td>
<td>5.2</td>
<td>Quiz 11</td>
<td></td>
</tr>
<tr>
<td>20-Nov</td>
<td>Analytical Pitfalls (Denominators, Confounding)</td>
<td>6.1</td>
<td>Quiz 12 HW 5</td>
<td>X</td>
</tr>
<tr>
<td>9-Dec</td>
<td></td>
<td></td>
<td>Project!</td>
<td></td>
</tr>
</tbody>
</table>