STA 6233: Advanced R Programming – Spring 2021
STA 6233 R Programming for Data Science

Instructor: Matthew J. Martinez, Ph.D.   **Email**: matthew.martinez@utsa.edu
Website: https://www.linkedin.com/in/mattdemography/
Course Site: https://github.com/mattdemography/STA_6233_Spring2021
Github: https://github.com/mattdemography
Twitter: https://twitter.com/mattdemography
Office Location: By Appointment (Online)   **Office Phone**: 726.666.0385 (Text/Call)

Office Hours/Availability: Thursdays 5:00 – 6:00pm online Zoom Meeting. I will make every effort to respond to texts within 24hours and e-mails as soon as I can. Admittedly you will receive a much faster response if you text me. I will not respond over the weekend. Do not text/call after 9:00pm CST on any day.

Class Location/Meeting Times
Online; Wednesdays 5:00 – 7:45pm

Description of Course

This course is designed to introduce students to the statistical program R for data analysis and manipulation. Topics include preprocessing/manipulating/combining datasets, summarizing and visualizing data techniques, writing functions, object oriented programming, data simulation and resampling methods, and interfacing R with other programming languages such as SQL, Python, C++, and Hadoop. Techniques for efficient programming will be stressed. The concept of high-performance computing (multi-core/parallel-processing) is also demonstrated.

This course is designed to move forward your utilization of R in a practical environment. Although you may find the applications discussed in this course beneficial for research purposes, the main aim of this course will be to equip you to use R as one would at a daily job in a Fortune 500 company. In this sense, the course materials and assignments will focus more on practical application than theoretical understanding.

By completing this course, you will have the opportunity to:
1. Develop high-end graphical representations of data
2. Find ways to simplify your code for readability and performance
3. Learn how to interface R with SQL, Python, and Github
4. Use web and document scraping for innovation and curation of data
5. Learn how to present your highly technical work to non-technical audiences in a clear, concise, and brief manner

Required Materials

All required readings will be available on Blackboard Learn. You will have to incur costs of printing posters, but this is minimal compared to textbooks. Additionally you will have to have access to a laptop to use during the course.

Course Policies

Attendance
Students are expected to attend and actively participate in the classroom discussions. Failure to attend will significantly decrease your success in the class. You are responsible for the material presented in class regardless of whether you are in attendance. If you show up at the end of the semester and I don't recognize you, chances are you haven't attended a significant portion of the course.

Classroom Etiquette
I consider the classroom to be a safe, productive environment. As such, we will maintain a harassment-free classroom. I encourage you to share your viewpoints in an appropriate manner and expect that you will respect the viewpoints of others. Part of classroom etiquette is remaining engaged throughout course time. Please refrain from side conversations as they are distracting to others around you and to me. I hear your conversations and it throws off my train of thought usually resulting in longer class periods. I will call you out for repeated offenses and it will be awkward – because you are adults and should act as such.

Projects and Presentations
Assignments are due on the day indicated in the syllabus before the start of class at 5:30PM via Learn. **No late work will be accepted.** Additionally, assignments *will not* be accepted via email. *We will discuss all assignments in much more detail in class.

Teams
Given the size of this course, you all will be assigned to a team. This team provides troubleshooting support of each other's code. They are your first line of defense for trouble with code such as reviewing code to find where potential errors exist. If team members cannot find a solution within a reasonable time, then you may e-mail/text me directly concerning this. Note that I will not continue to troubleshoot your code. As a burgeoning data scientist this will largely be your and your team's responsibility.

Grades
This is a graduate level course. I assume you are in this course to gain valuable information. I also assume you will put forth maximum effort and work during and outside of class. Failure to complete projects and presentations may result in an incomplete or failing grade.

University Policies
Please visit http://www.utsa.edu/syllabus for University policies including those related to accommodations and academic dishonesty.

PROJECTS

This course will have two main projects and presentations of those projects along with an evaluation of your peers' work.

Document/Web-scrapping
This project focuses on the creation of a unique dataset using data that are mined from documents and/or web sources. You will be responsible for coming up with a research question that you desire to answer using data that are previously unavailable. This task is designed to sharpen your skills in creative problem solving and ability to distill copious information into usable products using R packages centered on text manipulation, data frame management, feature creation, and regular expressions.

You will create a poster style presentation that will be showcased in a course symposium open to the department and the public. You will also evaluate your peer's work providing constructive criticism and encouragement.

Interactive Project
For this project you will have to design an interactive presentation using data to answer a different research question. This interactive project will either be a web document utilizing graphs and interactives from ggplot, Plotly, Google Vis, or d3 or you will create and deploy a Shiny Application. This task will allow you to use R packages centered on graphical representations of data including learning how to write and compile R Markdown files. Additionally, this project sharpens your ability to distill complex data to everyday audiences in a manner they find engaging and understandable.

You will publish your results online via Github and will be responsible for providing feedback to two of your classmates on their slide deck, use of graphics and interactivity, and your own ability to understand the main takeaway from their project.

Summary of Assignments
Poster Presentation
Evaluation of Peer's Posters
Interactive Project
Evaluation of Peer's Interactive Projects
Classroom Participation

Grading Scale
A+ = Successful Completion of All Projects
Incomplete/No-Credit = Unsatisfactory Work

Tentative Class Schedule

Tentative Course Schedule of Assignments

Date(s)	Material to be Covered	Readings/Assignments/Exercises/Etc.
Week of January 19th	• Introduction to Course and Introduction to R and R Studio	
Week of January 25th	• Data Manipulation: plyr, dplyr, Data Table	
Week of February 1st	• Data Readability and Performance: Functions, Loops, and Package Creation	Idea for Document/Web Mining Due
Week of February 8th	• Document and Web Mining	
Week of February 15th	• NO CLASS – ERCOT MESSED UP	
Week of February 22nd	• R Connections: Using and Learning Regular Expressions	
Week of March 1st	• Making Presentation Worthy Graphs: ggplot	
Week of March 8th	• **Spring Break – No Class**	
Week of March 15th	• **Course Symposium (Online)**	**Document/Web Mining Project Due**
Week of March 22nd	• Github • R Markdown	**Peer Evaluation of Document/Web Mining Projects Due**
Week of March 29th	• Making Interactive Presentations: Plotly and Google Vis	
Week of April 5th	• Shinny Apps Part 1	Idea for Interactive Project Due
Week of April 12th	• Shinny Apps Part 2	
Week of April 19th	• R Connections: SQL Coding Examples & Hadoop	
Week of April 26th	• **Working Week – No Class**	**Interactive Project Due April 30th @ 5:30pm**

Week of May 3rd	**All Peer Evaluations Due** Friday, May 8th 5:45pm