**STAT 413/613 Data Science**

**Summer 2021**

**J Dickens, PhD**

* **Time**: Mondays and Wednesdays 5:30 PM to 8:40 PM
* **Instructor**: Dr. J Dickens
* **Email : jdickens@american.edu**
* **Office**: DMTI 208C
* **Virtual Office Hours (via Zoom)**: Thursdays 11am – 1pm, Fridays 7pm – 9pm , Sundays 7pm – 9pm
* **Virtual Class Sessions** (Mondays and Wednesdays 5:30pm to 8:40pm)

**Overview of Topics and Course Objectives**

This course builds on the R/Tidyverse programming skills developed in STAT-412/612 for the collection, organization, analysis, interpretation, and presentation of data. Topics include version control, web scraping, date manipulation, vectorized operations, web application development with R Shiny, big data manipulation, R’s statistical modeling functions, and dimensionality reduction and clustering for data exploration.

**Learning Objectives:**

* Gain expertise in the git version control software.
* Collaborate and share code using [GitHub](https://github.com/)
* Scrape data from the web, either with or without using an API.
* Implement vectorized operations in R using functional programming techniques.
* Build and deploy data science web applications using R Shiny.
* Review elementary statistics and implement basic statistical procedures in R.
* Wrangle and analyze large datasets using data.table.
* Understand the ethical questions associated with data collection and analysis.

**Materials**

* **Required**: A computer with [R](https://cran.r-project.org/), [RStudio](https://rstudio.com/products/rstudio/download/), and [git](https://git-scm.com/downloads) installed.
* **Books**: All course material is freely available online.
  + R for Data Science by Wickham and Grolemund: <http://r4ds.had.co.nz/>
  + Advanced R by Wickham: <http://adv-r.had.co.nz/>
  + Mastering Shiny by Wickham: <https://mastering-shiny.org/>
  + Git for Scientists by McBain: <https://milesmcbain.github.io/git_4_sci/>
  + OpenIntro Statistics by Diez, Cetinkaya-Rundel, and Barr: <https://www.openintro.org/stat/textbook.php?stat_book=os>
  + blogdown: Creating Websites with R Markdown by Xie, Thomas, and Hill: <https://bookdown.org/yihui/blogdown/>

**Topics/Objectives coverage schedule (modifications and adjustments may occur)**

**Week 1 / Vectors, Functions, Iteration**

**Week 2 / Git and Github (Branching, Forking)**

**Week 3 / Statistics Linear Models (Linear Regression, Many Models)**

**Week 4 / Shiny Apps**

**Week 5 / Data Table Web Scraping (with and without APIs)**

**Week 6 / Introduction to Python Additional Topic(s) (TBD)**

**Class Structure**

We will hold virtual classes via **Blackboard Collaborate** during the scheduled class times. Class sessions will be a blend of lecture, class discussion, exercises, and labs. There is no expectation to share video during class if you do not want to, unless you have a really cool Zoom background.

To foster collaboration, at the beginning of the semester I will divide students into teams of about 3 to 4 students. You can discuss homework issues, coding problems, etc with your teammates. The forum for team communication will be through GitHub.

If you cannot attend lectures, I will expect you to participate in other ways. For example, you can show up to office hours, you can post questions on GitHub, you can answer other students’ questions on GitHub, or you can post links relating to data science (e.g. cool blog posts, exciting new R packages, etc).

During exercises and labs, I will split the class up into breakout sessions of 2-4 students. One student should share their screen and the other students can discuss and type suggestions while the group solves the exercise together. Students should alternate on who is the screen sharer. One group will be chosen to present their results

**Graded Work**

**Assignments**

* There will be approximately 5-6 formal homework assignments throughout the summer session in addition to a take home midterm and a take home final exam. Anticipate the Midterm at the end of week 3 and the Final Exam at the end of week 6.
* Homework assignments instructions for projects as well as exams will be posted on Blackboard and GitHub.
* To account for life circumstances that pop up, I will also drop the lowest homework assignment grade.

**Group Projects**

All students will prepare a final project using the tools learned in the class. This project will be completed in groups of 2-3 students. Work with me to get your project topic approved.

Your project will involve creating a Shiny app to interactively analyze a real-world dataset. As a part of the project, you will present this Shiny app to the rest of the class during a virtual presentation during final exam week.

Those taking this course for graduate credit (STAT 613) will also be required to submit a Graduate project. The Graduate Project can be a written report detailing a statistical analysis concerning a topic or a project that features an R programming application not studied in class. You will run a statistical inference procedure, verify the assumptions of your statistical procedure, and write up concluding remarks.

**Grading**

**Students will be subject to the grading process outlined below, depending on their classification as graduate or undergraduate**

**Undergraduate Students:**

|  |  |
| --- | --- |
| **Assignment** | **Grade Proportion** |
| Midterm Exam | **.15** |
| Final Exam | **.15** |
| Homework | **.45** |
| Classwork/Labs | **.10** |
| Shiny App Project | **.15** |

**Graduate Students:**

|  |  |
| --- | --- |
| **Assignment** | **Grade Proportion** |
| Midterm Exam | **.15** |
| Final Exam | **.15** |
| Homework | **.30** |
| Classwork/Labs | **.10** |
| Shiny App Project | **.15** |
| Graduate Project | **.15** |

You may receive assistance from other students in the class as well as myself, and / or the TA designated for this class., but your submissions must be composed of your own thoughts, coding, and explanations. I expect you to get ideas from online resources such as stackoverflow or github when you get stuck. Please cite your source when you do so and be specific about what you have added to it. You should be able to redo the code “cold” when you do this. Failure to do so is a violation of AU’s Academic Integrity Code.

**Generally, I do not accept late assignments**. To accommodate extraordinary circumstances and illness, I will drop the lowest assignment grade. You should not use this as an excuse to skip an assignment, but rather to handle a situation where you are unable to submit an assignment in a timely manner due to an emergency.

**Grading Scale**

| **Grade** | **Lower** | **Upper** |
| --- | --- | --- |
|  |  |  |
| A | 95 | 100 |
| A- | 89 | 94 |
| B+ | 85 | 88 |
| B | 80 | 84 |
|  |  |  |
| C+ | 75 | 79 |
| C | 70 | 74 |
| C- | 65 | 69 |
| D | 60 | 64 |
| F | 0 | 59 |
|  |  |  |

**Academic Integrity and Group Work**

* All students should familiarize themselves with American University’s Academic Code of Integrity: <https://www.american.edu/academics/integrity/code.cfm>.
* You may receive assistance from other students in the class and me, but your submissions must be composed of your own thoughts, coding and words.
* I expect you to get ideas from online resources such as Stackoverflow or GitHub when you get stuck (this is what real programmers and data scientists do). Please cite your source **using urls** when you do so.
* You should be able to explain your work on assignments and the project and your rationale. **Based on your explanation (or lack thereof), I may modify your grade.**
* I wish to emphasize that looking/using past homework solutions is a violation of the Academic Code of Integrity (even if it is “just” for confirmation purposes).
* It is a violation of the Academic Code of Integrity if you obtain past homework solutions from students who took the course previously.
* It is a violation of the Academic Code of Integrity if you obtain past homework solutions from students who are currently taking the course.
* There are a bunch of small mistakes in my old homework solutions. It is easy for me to see if you copied these mistakes and have been using my solutions. It is usually not enough to just change variable names to avoid detection.
* All solutions that I provide are under my copyright. These solutions are for personal use only and may not be distributed to anyone else. Giving these solutions to others, including other students or posting them on the internet, is a violation of my copyright and a violation of the student code of conduct.

**Assistance/Support**

Before receiving any assistance on a specific homework problem, please make sure that you have read through the class materials, and that you have made a fair attempt at the problem.

1. **If you are having ANY trouble with the class, please come see me about it as soon as possible. Do not wait until it is too late.**
2. Always feel welcome to talk to me during my office hours (no appointment necessary). These hours are for you to ask questions. Office hours can be busy so please come prepared with specific questions.
3. I also set aside a few hours each week to meet with students outside of office hours. Just send me an email to ask for a private meeting.
4. Use your peers! Feel free to work with your classmates on assignments. Just make sure that you write down the solutions in your own words.
5. You are also encouraged to ask me questions online via email.
6. Additional support services are available on campus that may assist you in successfully completing the course requirements. Details provided by each support service’s office are provided below.
   * The Academic Support and Access Center (x3360, MGC 243) supports the academic development and educational goals of all AU students while also providing support to students with disabilities. They offer workshops on topics of interest to all students such as time management, note taking, critical thinking, memory skills, and test taking. Additional support includes free private and group tutoring in many subjects, supplemental instruction, The Math Lab and The Writing Lab.
   * Students wishing to receive accommodations for a disability, are to bring their documentation directly to the Academic Support and Access Center (ASAC-<http://www.american.edu/ocl/asac/index.cfm>), in MGC 243, x3360. ASAC, in turn, will notify me of the accommodation required. Keep in mind that accommodations can only begin when I have been notified. This means that students should take care of this at the start of the semester, **before** the work for which they require accommodation is due.
   * The Counseling Center (x3500, MGC 214) is here to help students make the most of their university experience, both personally and academically. We offer individual and group counseling, urgent care, self-help resources, referrals to private care, as well as programming to help you gain the skills and insight needed to overcome adversity and thrive while you are in college. Contact the Counseling Center to make an appointment in person or by telephone, or visit the Counseling Center page on the AU website for additional information.
   * Center for Diversity & Inclusion (X3651, MGC 201) is dedicated to enhancing LGBTQ, Multicultural, First Generation, and Women’s experiences on campus and to advance AU’s commitment to respecting & valuing diversity by serving as a resource and liaison to students, staff, and faculty on issues of equity through education, outreach, and advocacy.
   * OASIS: The Office of Advocacy Services for Interpersonal and Sexual Violence (X7070) provides free and confidential advocacy services for anyone in the campus community who is impacted by sexual violence (sexual assault, dating or domestic violence, and stalking).

**A Few Additional Notes**

* I expect you to be courteous to me and your fellow classmates both inside and outside of the classroom. This generally just involves a bit of common sense. Cell phones need to be silenced and put away during class. Laptops should be out during class time for use only on class activities. Please save texting, typing/sending emails, checking Facebook, etc. for outside of class time. Any correspondence pertaining to the course needs to be handled in a respectful manner.
* A grade of incomplete will only be given under extreme circumstances and will not be granted to any student who is failing.
* In the event of an emergency, please refer to the AU Web site (<http://www.american.edu/emergency>) and the AU information line at (202) 885-1100 for general university-wide information. In the event that class is canceled for ANY reason I will communicate with you via email and Blackboard to let you know what work, reading, etc. you will be responsible for.