

## 95-868 Applied Exploratory Data Analysis, Summer 2024

### Course Logistics

Class: Tue, Thu, 7:00 pm – 8:20 pm (Zoom)

Instructor: Raja Sooriamurthi, PhD

Office: HBH 3017

Email: [raja@cmu.edu](mailto:raja@cmu.edu)

Teaching Assistant:

Jose Eduardo Oros Chavarria ([joroscha](mailto:joroscha))

*This course is an introduction to the principles and practice of effective data exploration and visualization in the framework of reproducible research.*

**Learning Objectives:** On the successful completion of the course, students will be able to:

1. How to use R to perform basic data tasks such as filtering, aggregating, and organizing data sets, and for production of graphics.
2. Understand the Fundamentals of Data and Reproducible Data Analysis.
3. Create Statistical Graphics and effective visualizations.

### Course Components:

- i. **Lectures.** EDA is an exciting and rapidly evolving field. *To fully engage in classroom discussions, you are expected to attend all class sessions and come prepared for each class.* Class participation contributes towards the final grade assessment. There will be in-class assignments and occasionally unannounced short quizzes at the beginning of class. Students who have an unexcused absence or tardiness will not be able to make up these assignments and quizzes. Unexcused absences can reflect upon your grade. In the event of a situation requiring you to be absent (e.g., sickness, job interview) please contact the professor in *advance*.

Please note that this is a summer course and will proceed at a rapid pace spread over just 6 weeks. *Missing even one or two classes may cause you to fall considerably behind the rest of the class.*

- ii. **Assignments.** Homework problems provide you with the opportunity to learn, practice, and test your knowledge and understanding of the material. You will be graded on your ability to think critically about both the data and application at hand, code graphics properly, incorporate relevant statistical methods, and write concisely about your analyses. We will give you adequate time to work on the problems, and the graders will work hard to return your homework in a timely manner. Unfortunately, this means that late homework will not be accepted.

**Code.** All code should be written in R and Quarto. Students should follow the popular [tidyverse style guide](#). If a student's submitted code does not adhere to this style guides, students will lose up to 10% credit on that assignment. If you wish to use a different (but well-defined) style guide, please talk to the instructor.

**Homework Format.** Homeworks should have the student's name and Andrew ID at the very top/beginning. Questions should be answered in order. All answers should be clearly marked and labeled. Answers should be written in the context of the problem when applicable. Proper spelling and grammar should always be used – this means using complete sentences, proper punctuation, etc. Deviating from this format may result in your assignment not being graded. You are encouraged to discuss homework problems and collaborate with

classmates. **However, the work you submit must be your own.** This means, in particular, that each student must independently write up each problem, including all code and written responses. Instances of identical, nearly identical, or copied homework will be considered cheating and plagiarism. The use of material from previous semesters of this course or from any other source to solve homework and exam problems is regarded as unauthorized assistance and therefore as a violation of the Carnegie Mellon University code of academic integrity.

## Assessment

Mastery of the course content will be demonstrated via a combination of homework and a final project:

Component	Weight
Homework (4-5)	60
Project (report + presentation)	40

## Tentative Course Schedule

Please check the detailed week by week course schedule for slides, handouts etc. All course content will be available from Canvas. The actual topics we discuss, and their depth will depend on the class's interest and pace. Given the broad nature of the topics in the course, there is no single book that discusses the course content. The lectures will be self-contained for the needed background for the assignments. For additional background, all of the following books are *excellent* resources:

- [Storytelling with Data: A Data Visualization Guide for Business Professionals](#) - Cole Nussbaumer Knaflic
- [Data Visualization: A practical introduction](#) - Kieran Healy
- [Fundamentals of Data Visualization](#) - Claus Wilke
- [R for Data Science](#) - Hadley Wickham and Garrett Grolemund
- [ggplot2: Elegant Graphics for Data Analysis \(3rd edition\)](#) - Hadley Wickham, Danielle Navarro, and Thomas Lin Pedersen.

Cole Knaflic's book is technology agnostic and is an excellent introduction to storytelling with visualization. It is available through the CMU library. The rest of the books discuss visualization in the context of R. The R community is very generous and supportive in terms of sharing their knowledge and expertise. All of these excellent texts are available for free on the Web.

## Class Policies

*Academic Integrity:* Unless explicitly stated otherwise, *all work needs to be individually done.* While it is fine to discuss general ideas, all submitted work must be your own. Sharing of work with another student or using the work of another's when completing your own will result in a grade of zero. Any case of suspected cheating will be brought to the Dean's attention. If you referred to external sources or consulted with others be sure to clearly indicate so. Be sure to familiarize yourself with the University policies on academic integrity <http://www.cmu.edu/policies/student-and-student-life/academic-integrity.html>.

*For Students with Learning Disabilities:* If you wish to request an accommodation due to a documented disability, please inform your instructor and contact the Office of Disability Resources <http://www.cmu.edu/disability-resources>

### ***Take care of yourself***

Do your best to maintain a healthy lifestyle by eating well, exercising, avoiding drugs and alcohol, getting enough sleep, and taking time to relax. Despite what you might hear, using your time to take care of yourself will actually help you achieve your academic goals more than spending too much time studying.

All of us benefit from support and guidance during times of struggle. There are many helpful resources available on campus. An important part of the college experience is learning how to ask for help. Take the time to learn about all that's available and take advantage of it. Ask for support sooner rather than later – this always helps.

If you or anyone you know experiences any academic stress, difficult life events, or difficult feelings like anxiety or depression, we strongly encourage you to seek support. Consider reaching out to a friend, faculty or family member you trust for assistance connecting to the support that can help. Counseling and Psychological Services (CaPS) is here for you: call 412-268-2922 and visit their website at <http://www.cmu.edu/counseling>. Over 25% of students reach out to CaPS some time during their time at CMU. <http://www.cmu.edu/teaching/designteach/design/syllabus/syllabussupport.html>

### ***Every individual must be treated with respect.***

We are diverse in many ways, and this diversity is fundamental to building and maintaining an equitable and inclusive campus community. Diversity can refer to multiple ways that we identify ourselves, including but not limited to race, color, national origin, language, sex, disability, age, sexual orientation, gender identity, religion, creed, ancestry, belief, veteran status, or genetic information. Each of these diverse identities, along with many others not mentioned here, shape the perspectives our students, faculty, and staff bring to our campus. We, at CMU, will work to promote diversity, equity and inclusion not only because diversity fuels excellence and innovation, but because we want to pursue justice. We acknowledge our imperfections while we also fully commit to the work, inside and outside of our classrooms, of building and sustaining a campus community that increasingly embraces these core values. Each of us is responsible for creating a safer, more inclusive environment.

Unfortunately, incidents of bias or discrimination do occur, whether intentional or unintentional. They contribute to creating an unwelcoming environment for individuals and groups at the university. Therefore, the university encourages anyone who experiences or observes unfair or hostile treatment on the basis of identity to speak out for justice and support, within the moment of the incident or after the incident has passed. Anyone can share these experiences using the following resources:

- Center for Student Diversity and Inclusion: [csti@andrew.cmu.edu](mailto:csti@andrew.cmu.edu), (412) 268-2150
- Report-It online anonymous reporting platform: [www.reportit.net](http://www.reportit.net) username: tartans password: plaid

All reports will be documented and deliberated to determine if there should be any following actions. Regardless of incident type, the university will use all shared experiences to transform our campus climate to be more equitable and just.

***Let's have a fun and productive course!***