

DSO-545: Statistical Computing and Data Visualization**Spring 2024, 16278 R****Location: JKP104, M, W 12:30 P.M. – 1:50 P.M.**

Instructor: Dr. Arif Ansari
Office: BRI 401 R (Main Campus) and Zoom Office hours
Office Hours: 5:00-5:45 p.m. on Class days and by Zoom Appointment
Phone: (213) 999 3554
Email: aansari@marshall.usc.edu

Emergency Contact number: 213-740-0172 or 213-999-3554E-mail: aansari@marshall.usc.edu**I. Course Description**

Statistical computing and data visualization are increasingly important and necessary aspects of a data analyst job. Whether they are dealing with small, big, structured or unstructured data, data analysts are expected to be able to access data from various sources, apply the latest statistical methodologies, and communicate their findings to others in novel visualizations.

In this course, students will learn how to make sense of data, and not the specifics of modeling. The course starts with statistical computing, and students will gain experience with a programming language called **R** (<http://cran.r-project.org>), a GNU-license statistical package. They will learn the practice of data cleaning, reshaping of data, basic tabulations, and aggregations in order to be able to produce high quality visualizations.

Then, the course proceeds with graphics that are critical elements of modern data analysis and presentation. From initial exploration of a data set to the final presentation of results to the end user, statistical graphics play a very vital role in shaping our understanding of our data. Through proper use of graphics, we can make critical discoveries, and communicate them clearly. Conversely, poor use or misuse of graphics can seriously mislead (by accident or design).

II. Course Learning Objectives

In this course, the students will gain explicit experience with programming language concepts such as variables, assignments, flow control, functions, parameters, data structures, input and output, error handling, debugging, and so on. In addition, students will learn how to write a computer program in **R** to create a graph from scratch and manipulate its different attributes (axis, title, orientation, color, etc.), identify appropriate data visualization techniques given particular requirements imposed by the data, analyze and criticize examples of visualizations from newspapers, scientific papers, business reports, and journals. Create visualizations that help organizations to take advantage of opportunities.

1. Describe Data from business point of view.
2. Learn Best Business Practices for Selecting Appropriate Visualization & Reporting for Business Needs
3. Apply Statistical Analytics and Visualization for Business problems.
4. Introduce Statistical Analysis R software and Tableau based Visualization Analytics.
5. Use appropriate visualization for analyzing the data – for achievable strategies and solutions that allow organizations to take advantage of opportunities that create value for its stakeholders.
6. Develop quantitative material to support written arguments – Prescriptive solutions that can create short-term and long-term value for organizations.
7. Apply R programming in an integrated approach to understanding and analyzing business problems, which can be complex, messy, unstructured, and beyond formulaic analysis.

III. Prerequisites

I expect basic “operational” knowledge from an introductory stats course. Prior R knowledge is not required, but you will rather learn the basics of R in this course

IV. Software

This is a hand-on course and it is computationally intensive. We will primarily be using **R** (<http://cran.r-project.org>), a GNU-license statistical package. We will also be introducing **Tableau** for interactive graphics.

There are many reasons for us to focus on R:

- The cost: While commercial distributions exist, open-source R is free.
- The rich features: R has an estimated user community of 2 million, which includes thousands of contributors from different domains expanding the language’s capabilities through new libraries.
- The quality: R libraries are enhanced by domain experts and field-tested by the large user base including other experts with real datasets in real analysis scenarios.
- The learning resources: Thanks to the active user community, plenty of tutorials and sample code are readily available.

Rstudio is a recommended interface for the R software. It is also free, and it runs on Windows, Mac, and Linux operating systems. <http://www.rstudio.org>

Students are expected to bring their laptops to class during all class sessions.

V. Methods

This class is taught through a combination of lectures, computer labs, hands on computing tasks in homework and exams. The assignments integrate multiple computational topics in the content of a modern data problem. Students gain hand on experience with statistical concepts flowing from contextual problem solving with data, and they make their own discoveries by posing and answering questions rather than solely fitting models or using “this week’s lecture’s methodology” as a computing exercise.

VI. Suggested Books

Notes will be posted on blackboard, and the following is a list of suggested books:

Wickham, H and Golemund, G (2017) R for Data Science (Note: 2nd Edition is available but more on Modeling than Visualization)

Yau, N. (2011). Visualize this: The flowing data guide to design, visualization, and statistics. Indianapolis: Wiley Publishing.

Tufte, E. (2001). The Visual Display of Quantitative Information. Connecticut: Graphics Press LLC. Wickham, H.

(2016). ggplot2: Elegant Graphics for Data Analysis.

Chang, W. (2013). R Graphics Cookbook.

Sleeper, R. (2018). Practical Tableau: 100 Tips, Tutorials, and Strategies. Optional Book

VII. Structure

Each topic in the course is motivated by a data problem, and will be supported by relevant data.

VIII. Assessment

- Homework
To do well in this course you will need to spend 4-5 hours a week (outside of class!), and the homework are designed to encourage you do that. For each homework assignment, you will need to revise the week's work, as well synthesize some new information, from the help pages or the web.
- Participation will answering questions on theory, concepts and Business Practices on blackboard.
- There will be a Midterm and Final Exam in this class. You can use the notes provided on blackboard.

Assessment	% of Grade
Homework	30%
Participation	20%
Midterm	25%
Final Exam	25%

IX. Schedule

Assessment	Date	Place
Midterm	TBD	JKP 104
Final	FINAL EXAM – May 3rd – 11:00 – 1:00 p.m.	TBD

X. Topics

- Data Science Tools: R, R Markdown, Rstudio
- Exploratory data analysis: one variable, two variables, multivariate, linear regression, deep learning framework
- Visualizing data: ggplot2
- Spatial graphics: gmap
- Time series: lubridate (if time allows)
- Web scraping: xml
- Elements of Tableau for data visualization and dashboarding

OPEN EXPRESSION AND RESPECT FOR ALL

An important goal of the educational experience at USC Marshall is to be exposed to and discuss diverse, thought-provoking, and sometimes controversial ideas that challenge one's beliefs. In this course we will support the values articulated in the USC Marshall "[Open Expression Statement](#)."

Open Expression and Respect For All

An important goal of the educational experience at USC Marshall is to be exposed to and discuss diverse, thought-provoking, and sometimes controversial ideas that challenge one's beliefs. In this course we will support the values articulated in the USC Marshall "[Open Expression Statement](#)."

Schedule of class – This is Approximate Time Line

Date		Topic	Reading from textbooks	Reading from Class notes	Due/Other
Week of 1/08/24	1	Introduction to R	R for Data Science Chapter 1, 2 and 4	Dr. Ansari Notes	
Week of 1/15/24	2	Why Visualize? ggplot	R for Data Science Chapter 3	Dr. Ansari Notes	Turn in your Lab 1
Week of 1/22/24	3	ggplot - Continued	R for Data Science Chapter 3	Dr. Ansari Notes	Turn in your Lab 2
Week of 1/29/24	4	dplyr Package	R for Data Science Chapter 5	Dr. Ansari Notes	
Week of 2/5/24	5	dplyr Package - Continued Statistics – t-test & Regression Model in R	R for Data Science Chapter 5	Dr. Ansari Notes	Turn in your Lab 3
Week of 2/12/24	6	Statistics – t-test & Regression Model in R Factor Variable	R for Data Science Chapter 15	Dr. Ansari Notes	
Week of 2/19/24	7	Factor Variable - Continued	R for Data Science Chapter 15	Dr. Ansari Notes	Turn in your Lab 4
Week of 2/26/24	8	Review	Sample Midterm	Dr. Ansari Notes	
Week of 3/04/24	9	Midterm / Time Series Data – Spatial Analysis	R for Data Science Chapter 16	Dr. Ansari Notes	
Week of 3/18/24	10	Time Series Data – Spatial Analysis Lubridate Package	R for Data Science Chapter 16	Dr. Ansari Notes	Turn in your Lab 5
Week of 3/25/24	11	Text Data – Stringr Package	R for Data Science Chapter 14	Dr. Ansari Notes	
Week of 4/01/24	12	Data Mining in R and Tableau Introduction	Class Notes	Dr. Ansari Notes	Turn in your Lab 6
Week of 4/08/24	13	Additional Topic in R (deep learning)	Class Notes	Dr. Ansari Notes	Turn in your Lab 7
Week of 4/15/24	14	Interactive Web	R for Data Science Chapter 29		
Week of 4/22/24	15	Review		Dr. Ansari Notes	Turn in your Lab 8
4/29/24	16	STUDY DAY			
5/03/24	16	FINAL EXAM – May 3rd – 11:00 – 1:00 p.m.			

Statement on Academic Conduct and Support Systems

Academic Integrity:

The University of Southern California is a learning community committed to developing successful scholars and researchers dedicated to the pursuit of knowledge and the dissemination of ideas. Academic misconduct, which includes any act of dishonesty in the production or submission of academic work, compromises the integrity of the person who commits the act and can impugn the perceived integrity of the entire university community. It stands in opposition to the university's mission to research, educate, and contribute productively to our community and the world.

All students are expected to submit assignments that represent their own original work, and that have been prepared specifically for the course or section for which they have been submitted. You may not submit work written by others or "recycle" work prepared for other courses without obtaining written permission from the instructor(s).

Other violations of academic integrity include, but are not limited to, cheating, plagiarism, fabrication (e.g., falsifying data), collusion, knowingly assisting others in acts of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

The impact of academic dishonesty is far-reaching and is considered a serious offense against the university. All incidences of academic misconduct will be reported to the Office of Academic Integrity and could result in outcomes such as failure on the assignment, failure in the course, suspension, or even expulsion from the university.

For more information about academic integrity see [the student handbook](#) or the [Office of Academic Integrity's website](#), and university policies on [Research and Scholarship Misconduct](#).

Please ask your instructor if you are unsure what constitutes unauthorized assistance on an exam or assignment, or what information requires citation and/or attribution.

The use of AI Generators in the course:

Not Permitted in Exam and Homework Submission, can use to generate and compare code.

In this course, I encourage you to use artificial intelligence (AI)-powered programs to help you with writing the Code for comparison with your written Code. Since the objective of this course is to learn R language you will not be allowed to use AI to write your lab code but you can use AI to compare your written code with AI generated code. To adhere to our university values, you must cite any AI-generated material (e.g., text, images, etc.) included or referenced in your work and provide the prompts used to generate the content. Using an AI tool to generate content without proper attribution will be treated as plagiarism and reported to the Office of Academic Integrity. Please review the instructions in each assignment for more details on how and when to use AI Generators for your submissions.

Students and Disability Accommodations:

USC welcomes students with disabilities into all of the University's educational programs. [The Office of Student Accessibility Services](#) (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion.

This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at osas.usc.edu. You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

Support Systems:

[Counseling and Mental Health](#) - (213) 740-9355 – 24/7 on call

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

[988 Suicide and Crisis Lifeline](#) - 988 for both calls and text messages – 24/7 on call

The 988 Suicide and Crisis Lifeline (formerly known as the National Suicide Prevention Lifeline) provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week, across the United States. The Lifeline is comprised of a national network of over 200 local crisis centers, combining custom local care and resources with national standards and best practices. The new, shorter phone number makes it easier for people to remember and access mental health crisis services (though the previous 1 (800) 273-8255 number will continue to function indefinitely) and represents a continued commitment to those in crisis.

[Relationship and Sexual Violence Prevention Services \(RSVP\)](#) - (213) 740-9355(WELL) – 24/7 on call

Free and confidential therapy services, workshops, and training for situations related to gender- and power-based harm (including sexual assault, intimate partner violence, and stalking).

[Office for Equity, Equal Opportunity, and Title IX \(EEO-TIX\)](#) - (213) 740-5086

Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

[Reporting Incidents of Bias or Harassment](#) - (213) 740-5086 or (213) 821-8298

Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

[The Office of Student Accessibility Services \(OSAS\)](#) - (213) 740-0776

OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.

[USC Campus Support and Intervention](#) - (213) 740-0411

Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

[Diversity, Equity and Inclusion](#) - (213) 740-2101

Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

[USC Emergency](#) - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call

Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

[USC Department of Public Safety](#) - UPC: (213) 740-6000, HSC: (323) 442-1200 – 24/7 on call

Non-emergency assistance or information.

[Office of the Ombuds](#) - (213) 821-9556 (UPC) / (323-442-0382 (HSC)

A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

[Occupational Therapy Faculty Practice](#) - (323) 442-2850 or otfp@med.usc.edu

Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.

TECHNICAL SUPPORT

For Blackboard support go to USC ITS Services Website or call USC ITS at 213.740.555

For Zoom Support or call (888) 799-9666 ext. 2

Appendix I. MARSHALL GRADUATE PROGRAMS LEARNING GOALS

How DSO 545 Contributes to Marshall Graduate Program Learning Goals

Marshall Graduate Program Learning Goals	DSO 548 Objectives that support this goal	Assessment Method*
Learning Goal #1: Develop Personal Strengths. Our graduates will develop a global and entrepreneurial mindset, lead with integrity, purpose and ethical perspective, and draw value from diversity and inclusion.		
1.1 Possess personal integrity and a commitment to an organization’s purpose and core values.	5	Class Participation, Quizzes, Exams
1.2 Expand awareness with a global and entrepreneurial mindset, drawing value from diversity and inclusion.	1	Class Participation, Quizzes, Exams
1.3 Exhibit awareness of ethical dimensions and professional standards in decision making.	4	Class Participation, Quizzes, Exams
Learning Goal #2: Gain Knowledge and Skills. Our graduates will develop a deep understanding of the key functions of business enterprises and will be able to identify and take advantage of opportunities in a complex, uncertain and dynamic business environment using critical and analytical thinking skills.		
2.1 Gain knowledge of the key functions of business enterprises.	1, 2, 3, 4	Exams, Lab
2.2 Acquire advanced skills to understand and analyze significant business opportunities, which can be complex, uncertain and dynamic.	3, 4, 5, 6	Exams, Lab
2.3 Use critical and analytical thinking to identify viable options that can create short-term and long-term value for organizations and their stakeholders.	4	Exams, Lab
Learning Goal #3: Motivate and Build High Performing Teams. Our graduates will achieve results by fostering collaboration, communication and adaptability on individual, team, and organization levels.		
3.1 Motivate and work with colleagues, partners, and other stakeholders to achieve organizational purposes.	5,6	Participation, Lab
3.2 Help build and sustain high-performing teams by infusing teams with a variety of perspectives, talents, and skills and aligning individual success with team success and with overall organizational success.		Participation, Lab
3.3 Foster collaboration, communication and adaptability in helping organizations excel in a changing business landscape.	2, 5, 6	Participation, Lab

Appendix II.

SAMPLE PEER EVALUATION FORM – please note, we don’t have group project for this class.

Please identify your team and team members for the ____ Project(s) that you worked on. Then rate all your team members, including yourself, based on the **contributions** of each team member for the selected assignment according to the criteria listed below. On a scale of 0 – 2 with 0 indicating does not meet expectations, 1 meets expectations and 2 exceeds expectations, rate each person on each of the five criteria. Lastly, add up the points for each person with the maximum number of points for each person being 10. In the box below, describe the exact contributions of each team member, including yourself.

Team Members/ Assessment Criteria of Team Contributions	Team Member 1	Team Member 2	Team Member 3	Yourself
1. Role Performance				
2. Assists Team Members				
3. Listening and Discussing				
4. Research and Information Sharing				
5. Time Management				
Total				

Contribution details:

Appendix III.

Sample Course Participation Rubric

The following rubric provides a guide as to how course participation will be assessed.

Active Participation	Moderate Participation	Low Participation
Exhibits evidence of having completed all assignments and activities according to guidelines that were assigned	Attempts to participate and has completed most assignments and activities	Exhibits lack of preparation and non-completion of required assignments
Initiates discussion and supports points using page-	Supports points during discussion but uses	Rarely initiates discussion and is not

specific references to readings or specific reference points in film/videos	general references to readings and other materials	able to reference required readings or other materials
Furtheres the discussion and builds on the ideas of others; comments and questions reflect having thought deeply about the material	Furtheres the discussion and builds on the ideas of others; general or limited references to course materials	Comments do not further the discussion and do not exhibit careful reflection on the material

Appendix IV

SAMPLE MIDPOINT COURSE EVALUATION QUESTIONS

In order to continuously improve the effectiveness of our class, could you please take a few moments to answer the following questions:

1. How well do the course objectives support your general business knowledge and personal career goals?
2. What have you liked about this course so far?
3. Do you have any suggestions for improving the course experience?