

**Spring 2023**

**Instructor: Sergio Luna, Ph.D.**

Course Web Address: [Link to Blackboard shell](#)

Course Schedule: Monday (6:00 pm – 8:50 pm MST)

Contact Information: [salunafong@utep.edu](mailto:salunafong@utep.edu)

Live session: [Zoom session](#)

Office Hours: Thursday 3:00pm - 4:30 p.m. MST or by appointment

Office session URL: [Zoom session](#)

Teaching Assistant Hours: Ms. Karen Gonzalez

Office Hours: Tuesday 4:30pm - 6:30 pm MST or by appointment

Email: [kdgonzalez@miners.utep.edu](mailto:kdgonzalez@miners.utep.edu)

## COURSE DESCRIPTION

This course provides a formal introduction to the process of data manipulation, analysis, and visualization to drive the strategic decision-making process. It leverages such methods with the corresponding problem-solving and story-telling skills to complement data visualization. Although the course is based mainly on R statistical programming language, the presented concepts and techniques are general enough that they can be extended to other statistical or data analysis tools. The course covers basic plotting techniques, including bar and scatter plots, time series, heatmaps, geographic maps, networks graphs, and other visualization charts. Additionally, students will get hands-on experience with primary data pre-processing techniques, including summaries, data cleaning and transformation, introduction to interactive visualizations, and fundamentals of machine learning.

No (professional) programming experience is required.

## STUDENT LEARNING OUTCOMES

After successful completion of this course, students will be able to:

- Understand and implement an analytical framework for conducting data exploration and analysis. Also, the student will be able to apply data manipulation techniques and create high quality plots to understand the phenomena of interest.
  - The student will develop critical thinking skills by addressing 5W questions when creating visualizations. What and So-What will be further explored.
- Understand the conventional approach to manipulating and pre-processing data large datasets. Data will be transformed to be analyzed to generate insights and provide stakeholders with data-driven recommendations.
  - The students will be able to load, manipulate, transform, and clean data using R and RStudio.
- Use R statistical language to generate data visualizations.
  - The student will be able to create different plots (Bar and Scatterplots, Time series, Interactive Charts, Histograms, etc.) by creating coding scripts.

- Develop public speaking and project management experience by creating a multi-chart static poster or interactive dashboard while disseminating critical insights to an audience. Students will develop two significant projects that need to be managed in terms of data extraction, research scope, analysis, insights generated, insights presentation, and suggested outcomes. Text documents and oral presentation skills are being practiced and evaluated.
  - Students will develop two significant projects that need to be managed in terms of data extraction, research scope, analysis, insights generated, insights presentation, and suggested outcomes. Text documents and oral presentation skills are being practiced and evaluated.

## COURSE FORMAT AND STRUCTURE

- Our weeks will run from Tuesday to Monday. I will post information (online activities, discussion starters, etc.) in Blackboard for the upcoming week by Sunday evening, so that when you log in on Monday, you can begin the new week.
  - Course Web Address: [Link to Blackboard shell](#)
  - Virtual session URL: [Zoom session](#)
- Virtual office hours will be held Thursday 2:00 p.m. - 3:30 p.m. MST or by appointment. Please email me at [salunafong@utep.edu](mailto:salunafong@utep.edu) to schedule a meeting.
- Assignments are due by 11:59 p.m. MST on the due date listed in the course schedule. Late submissions will be accepted, up 24 hours after the due date. However, a 50% of the maximum will be deducted from the assignment score. **No assignments will be accepted after 24 hours from the due date.**
- An assignment file should be appended by the respective assignment number, your Last Name and First Name, such as "assignment1\_LastName\_FirstName.pdf". This makes it easier for me to manage assignment files when downloading to my computer.
- Please submit the following files:
  - Working **R file** ("assignment1\_LastName\_FirstName.R") - # Commented
  - A **PDF file** describing the process and steps taken while manipulating data
  - The **dataset (raw data)** used to run the R submitted file
  - NOTE: Points will be deducted if any of the above items is missing with your submission.

## TENTATIVE COURSE SCHEDULE

The schedule below is subject to change. If for any reason I am required to make any amendments, I will be informing you via Blackboard, email and MS Teams.

Assignments are due by 11:59 p.m. MST on the day specified in the "Assignment" column seen below.

## Tentative Schedule

Module	Date	Topic(s)	Assignment
Week 01	Jan 17 <sup>th</sup> – 22 <sup>nd</sup> , 2022	Orientation & Introduction and Course Overview From Data Visualization to Data Analytics	
Week 02	Jan 23 <sup>rd</sup> – 29 <sup>th</sup> , 2022	Introduction to R	
Week 03	Jan 30 <sup>st</sup> – Feb 5 <sup>th</sup> , 2022	Data Manipulation 1	<b>HW #1 due 2/5 – Data Manipulation 1</b>
Week 04	Feb 6 <sup>th</sup> – 12 <sup>th</sup> , 2022	Data Manipulation 2	<b>HW #2 due 2/12 – Data Manipulation 2</b>
Week 05	Feb 13 <sup>th</sup> – 19 <sup>th</sup> , 2022	Basic Graphs	
Week 06	Feb 20 <sup>nd</sup> – 26 <sup>th</sup> , 2022	Graphs with Enhancement	<b>HW #3 due 2/26 - Basic Graphs &amp; Enhancement</b>
Week 07	Feb 27 <sup>th</sup> – March 5 <sup>th</sup> , 2022	Heat Map	<b>Release Mid-term exam</b>
Week 08	March 6 <sup>th</sup> – 12 <sup>th</sup> , 2022	Spatial Data	<b>HW #4 due 3/12 - Heatmap &amp; Spatial Data</b>
<b>Week 09</b>	<b>March 13<sup>th</sup> – 19<sup>th</sup>, 2022</b>	<b>Spring Break Mid-term Project Break</b>	<b>Mid-Term Project due: 3/19</b>
Week 10	March 20 <sup>nd</sup> – March 26 <sup>th</sup> , 2022	Categorical Data	
Week 11	March 27 <sup>th</sup> – April 2 <sup>nd</sup> , 2022	Time Series	<b>HW #5 due 4/2 - Categorical, Time Series &amp; Multivariate</b> Define Final Project
Week 12	April 3 <sup>rd</sup> – 9 <sup>th</sup> , 2022	Interactive Visualization	
Week 13	April 10 <sup>th</sup> – 16 <sup>th</sup> , 2022	Network Graphs	<b>HW #6 due 5/16 - Interactive Visualization and Network Graphs</b>
Week 14	April 17 <sup>th</sup> – April 23 <sup>rd</sup> , 2022	CRISP-DM Presentation	
<b>Week 15</b>	<b>April 24<sup>th</sup> – April 30<sup>th</sup>, 2022</b>	<b>Final Project</b>	<b>Final Project due: 4/30 @ 11:59pm MST</b>

## COURSE MATERIALS

- **Textbook(s): No textbook is required for this course.** Students will be presented with slides, and online tutorials. In case students are interested, the following books are recommended:

[R Graphics Cookbook Practical Recipes for Visualizing Data](#)

[Storytelling with Data: A Data Visualization Guide for Business Professionals](#)

## COURSE REQUIREMENTS

- **Attendance:** Attendance is not mandatory but recommended.
- **Assignments:** A total of 6 homework assignments will be assigned throughout the semester. Data visualization exercise is 60%, and CRISP-DM is 40%.
- **Exams:** 1 Mid-term exam will be assigned. 60% Visualizations and CRISP-DM is 40%.

Students will work individually with assigned project and deliver a "data story" report.

Grading rubric:

- Context: Clearly defined topic that answers a specific question.
  - Data Cleanliness: The extent to which the data has been manipulated to remove potential bias injection to the resulting visualizations. Identifying incomplete, incorrect, inaccurate or irrelevant parts of the data.
  - Importance: the extent to which the visualization addresses problems facilitate decision making.
  - Relevancy: Visualization contains no color, symbolism, or text that is irrelevant to the question the visualization seeks to answer.
  - Grammar of graphics: Applies appropriate graphic variable types for the data type and scale.
  - Aesthetic Design: Meticulous care given to colors, shape, size, background, annotation and overall design.
  - Key insights communication. Clearly communicate the purpose, objective, analysis and insights and recommendations to the audience.
- **Project(s):** 1 Final Project will be assigned. 60% Visualizations and CRISP-DM is 40%.

Students will work individually on their chosen project and create a poster and "data story" report.

The final project aims to capture the knowledge and skills learned throughout the semester. It evaluates the students' approach when creating visualizations that support the decision-making process.

Grading rubric:

- All the criteria evaluated at during the mid-term project, plus:
- Evidence: Data comes from multiple highly reputable sources; data is supplemented with references to relevant scholarship; methodology behind how the data was collected is explained.
- Layout: design multi-chart static poster or interactive dashboard.
- Key insights communication. Clearly communicate the purpose, objective, analysis and insights and recommendations to the audience.

## GRADING PROCEDURES

Grades will be based on the following weights:

Homework assignments	40%
Mid-term project	30%
Final project	30%

Final grading rubric will be as followed:

A	90 - 100
B	80-89
C	70 - 79
D	60 – 69
F	< 60

## TECHNOLOGY REQUIREMENTS

Required Software:

- Microsoft Word
- Microsoft Excel
- Microsoft PowerPoint
- R Project for Statistical Computing (R programming language) - [Download R language](#)
- RStudio - [Download RStudio](#)
- Gephi - [Download Gephi software](#)

Download and install R and RStudio (Brief introduction will be provided during Orientation and Introduction to R session. If you want to practice in advance: please watch this YouTube video if you have never used R before: [Install R and RStudio](#).

## LEARNING ACCOMMODATIONS

The Center for Accommodations and Support Services (CASS) aspires to provide students with disabilities, accommodations and support services to help them pursue their academic, graduation, and career goals. For more information concerning services for students with disabilities, please contact the Center for Accommodations and Support Services at <https://www.utep.edu/student-affairs/cass/>

## INCLUSIVITY

### **Name and Pronoun Usage**

As this course includes group work and class discussion, it is vitally important for us to create an educational environment of inclusion and mutual respect. This includes the ability for all students to have their chosen gender pronoun(s) and chosen name affirmed. If the class roster does not align with your name and/or pronouns, please inform the instructor of the necessary changes.

You are expected to treat your instructor and all other participants in the course with courtesy and respect. Disrespectful conduct and harassing statements will not be tolerated and may result in disciplinary actions