IE 5390
Special Topics Industrial Engineering

"Data Visualization for Decision Making"

Fall 2020

Instructor: Sergio Luna, Ph.D.

Course Web Address: https://blackboardlearn.utep.edu/ultra/courses/_116349_1/cl/outline
Course Schedule: Wednesday (6:30 pm – 8:00 pm MST)
Contact Information: salunafong@utep.edu
Virtual Office Hours: Tuesday 11:00am - 1:00 p.m. MST or by appointment
Virtual session URL: Zoom Link to Weekly meeting

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 mostly on R statistical programming language, the presented concepts and techniques are general enough that 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, among other type of visualization charts. Additionally, students will get hands-on-experience with basic data pre-processing techniques including summaries, data cleaning and transformation, introduction to interactive visualizations, and fundamentals of machine learning.

No (proficient) 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.
    o The student will develop critical thinking skills by addressing 5W questions when developing visualizations. What and So-What will be further explored.

- Understand the conventional approach to manipulate and pre-process data while manipulating large datasets. Transformed data will then be analyzed to generate insights and provide stakeholders with data-driven recommendations.
    o The students will be able to load, manipulate, transform, and clean data using R and RStudio.

- Use R statistical language to generate data visualizations.
    o The student will be able to create different plots (Bar and Scatterplots, Time series, Interactive Charts, Histograms, etc.) by creating R scripts.

- Develop public speaking and project management experience by creating a multi-chart static poster or interactive dashboard while disseminating the key insights to an audience.

- o Students will develop 2 major projects that will 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 Monday to Sunday. 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.
  - o Course Web Address: https://blackboardlearn.utep.edu/ultra/courses/_116349_1/cl/outline
  - o Virtual session URL: Zoom Link to Weekly meeting

- Virtual office hours will be held Tuesday 11:00 a.m. - 1:00 p.m. MST or by appointment. Please email me at 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 make it easier for me to manage assignment files when downloading to my computer.

- Please submit the following files:

  - o Working **R file** ("assignment1_LastName_FirstName.R") - # Commented
  - o A **PDF file** describing the process and steps taken while manipulating data
  - o The **dataset (raw data)** used to run the R submitted file
  - o 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	Aug 24th – 30th, 2020	Orientation & Introduction and Course Overview From Data Visualization to Data Analytics	
Week 02	Aug 31st – Sep 6th, 2020	Introduction to R	
Week 03	Sep 7th – 13th, 2020	Data Manipulation 1	**HW #1** due **9/13** – Data Manipulation 1
Week 04	Sep 14th – 20th, 2020	Data Manipulation 2	**HW #2** due **9/20** – Data Manipulation 2
Week 05	Sep 21st – 27th, 2020	Basic Graphs	
Week 06	Sep 28th – Oct 4th, 2020	Graphs with Enhancement	**HW #3** due **10/4** - Basic Graphs & Enhancement
Week 07	Oct 5th – 11th, 2020	Heat Map	Release Mid-term exam
Week 08	Oct 12th – 18th, 2020	Spatial Data	**HW #4** due **10/18** - Heatmap & Spatial Data
Week 09	**Oct 19th – 25th, 2020**	**Mid-term Project Break**	**Mid-Term Project due: 10/25**
Week 10	Oct 26th – Nov 1st, 2020	Categorical Data	
Week 11	Nov 2th – 8th, 2020	Time Series	
Week 12	Nov 9th – 15th, 2020	Multivariate Data	**HW #5** due **11/15** - Categorical, Time Series & Multivariate Define Final Project
Week 13	Nov 16th – 22nd, 2020	Interactive Visualization	
Week 14	Nov 23rd – 29th, 2020	Network Graphs	**HW #6** due **11/29** - Interactive Visualization and Network Graphs
Week 15	**Nov 30th – Dec 3rd, 2020**	**Final Project**	**Final Project due: 12/5**

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.

- **Exams:** 1 Mid-term exam will be assigned.

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

  Grading rubric:
  - o Context: Clearly defined topic that answers a specific question.
  - o 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.
  - o Importance: the extent to which the visualization addresses problems facilitates decision making.
  - o Relevancy: Visualization contains no color, symbolism, or text that is irrelevant to the question the visualization seeks to answer.
  - o Grammar of graphics: Applies appropriate graphic variable types for the data type and scale.
  - o Aesthetic Design: Meticulous care given to colors, shape, size, background, annotation and overall design.
  - o Key insights communication. Clearly communicate the purpose, objective, analysis and insights and recommendations to the audience.

- **Project(s):** 1 Final Project will be assigned.
  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:
  - o All of the criteria evaluated at during the mid-term project, plus:
  - o 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.
  - o Layout: design multi-chart static poster or interactive dashboard.
  - o 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	35%
Mid-term project	30%
Final Project	35%

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 programing 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