ACSO 2020: Data Design, Analysis and Representation
Instructor: Dr. Ajay K. Aggarwal
Spring 2024

Email: Ajay.Aggarwal@unt.edu
Office: FRLD 353
Class Hours: T/Th 2-3:20
Office Hours: T/Th 1-2 & by appointment
Class Venue: FRLD 446

Course Description
This course introduces you to basic concepts in data management, statistics, and visualization of data patterns. It shares knowledge about dealing with incomplete and inconsistent information in real-world datasets and discusses techniques to produce reliable insights from structured and unstructured data. Topics covered include data generating processes, data sources, file formats, data management, data wrangling, and data mining.

Prerequisites
Prerequisite(s): DSCI 2710, MATH 1680 or equivalent.

Course Learning Outcomes
At the end of the course, you should be able to do the following:
1. Describe different data sources and file formats.
2. Describe key concepts of data management.
3. Inspect and explore data.
4. Apply data preparation techniques to handle missing values and to subset data.
5. Transform numerical and categorical variables.
6. Create visualization of relationships between two or more variables.
7. Understand spreadsheet modeling.

Required Courseware & Materials
- R can be downloaded at https://cloud.r-project.org
- RStudio a graphical user interface (GUI) for R, can be downloaded at https://www.rstudio.com/products/rstudio/download/#download
- Microsoft Excel, installed in computer labs, is available to download (Office365) at https://aits.unt.edu/support/office365apps
- Canvas (https://unt.instructure.com/login/ldap): The lecture slides, data files, syllabus, and other materials will be posted on Canvas. Please make sure you keep up and check it often.

Course General Guidelines
Brief notes about this course:
This course includes challenging material. To develop a thorough understanding of the course material and
succeed in the course you should attend class consistently, take all assignments seriously, work extensively outside of class, develop good study strategies, and get help from me when you’re struggling. This should aid your understanding of the course material and help you succeed in the course.

It is important to understand the different types of information we use to design a data repository and perform data interpretation. Relationships at among different pieces of information allow us to organize and group data in meaningful and efficient ways.

In data representation, it is key to understand the context of data generating process. Graphs and charts can be easy to understand but may not convey the whole story.

**Guidelines**

You are strongly encouraged to study all the posted material in the course and try and solve problems included in the lecture slides, either individually or in groups. You should replicate all the steps to solve the problems using Excel or R, instead of passively verifying that the provided solutions make sense.

You should work independently or in groups on the homework assignments and case studies. The case studies are intended to assist you in better structuring the learning time you spend on mastering the course material. Exam questions will mostly refer to case analysis and assignments. The best way to prepare for exams is to go over the exercises seen in class.

**Doing the assignments well is essential for success in this course.** The assignments constitute a large portion of your grade in the course. You are encouraged to keep up with the homework and meet the submission deadlines.

**Assessment**

To demonstrate your ability in Data Analysis and Design you need to work problem solving using statistical inference tools. You will be evaluated on a number of homework assignments or problem sets. Rather than being purely numerical, assignments, case studies and exams will require interpretation and thorough analysis of the problems presented showing a conclusion or recommendation whenever required. You should form a work-group for pursuing group projects and develop your collaborating skills. This is an important part of your formation since almost every job or role in industry will require interaction with peers, colleagues and collaborators. You will be provided with varied material to assess four main learning areas:

- **Reflective:** based on videos, articles, books/chapters, or guest testimonials. Students should respond based on a set of either provoking questions or reflections with supporting statements.
- **Case analysis:** a real-world example of the issues or concepts seen in class. Students should analyze problems, and conduct research, test and present potential solutions.
- **Examination:** standard formal tests
- **Discussions:** exchange of ideas on every class meeting will take place. Student-led discussions, electronic or in-class, are always encouraged.

**Course Grading**

<table>
<thead>
<tr>
<th>Grade Type</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midterm</td>
<td>20%</td>
</tr>
<tr>
<td>Final exam / Project</td>
<td>25%</td>
</tr>
<tr>
<td>Assignments (Individual &amp; Group)</td>
<td>40%</td>
</tr>
<tr>
<td>Attendance &amp; Class Participation</td>
<td>15%</td>
</tr>
</tbody>
</table>

While total course grade of 90, 80, 70, and 60 will assure you of receiving grades of “A”, “B”, “C”, and “D” respectively, your instructor curves grades in your favor based upon the overall class performance. Extra credit will be awarded for participation in activities announced by the instructor and for sharing innovative ideas that add value to the course material. All extra credits will be used in the computation of your final grade.
• **Assignments:** These are available in CONNECT. Individual Homework assignments should be attempted alone. Group assignments should be attempted groups of up to 3. Due dates for all assignments are provided in the detailed schedule in this syllabus.

• **Final Exam / Project:** You may be asked to work on a group final project to substitute for the final examination. Details will be discussed during class. Suggestions for topic and data selection and project presentation will be provided.

**University and School Policies and my course learning mindset**

- **DIVERSITY & INCLUSION:** As members of the UNT community, we have all made a commitment to be part of an institution that respects and values the identities of the students and employees with whom we interact. UNT does not tolerate identity-based discrimination, harassment, and retaliation so we will work as a class to collaborate in ways that encourage inclusivity. Although disagreements and debates are encouraged, personal attacks are unacceptable. Together, we can ensure a safe and welcoming classroom for all.

- **CAREGIVER RESPONSIBILITIES:** I have great respect for students who are balancing their pursuit of education with the responsibilities of caring for children or other family members. If you run into challenges that require you to miss a class, or if your caregiving responsibilities are interfering with your ability to engage in learning, please contact me. There may be some instances of flexibility we can offer to support your learning.

- **STAY INFORMED:** Access regularly the course material posted on Canvas. If you feel that you are struggling with the material, please contact me.

- **COMPLAINTS:** I value the many perspectives students bring to our classroom. Please work with me to create a classroom culture of open communication, mutual respect, and inclusion. All discussions should be respectful and civil. If you ever feel like this is not the case, please stop by my office and let me know. We are all learning together. If you wish to register a complaint, you should first discuss your complaint with me. If you wish to carry it further, contact Dr. Dianne Gravley (the program director) and then other instances in New College, but only after first discussing it with your instructor.

- **EXAMS:** Even though exams are stressing, you are perfectly capable of solve them successfully. If you are facing extenuating circumstances, please reach out to me as soon as possible to receive support. I normally require written requests to fulfill UNT regulations. We can then discuss alternative arrangements.

- **LEARNING SUPPORT:** I'm here for you. My aim is to facilitate your learning process. Please do not hesitate to ask questions to the instructor (Dr. Ajay). I will respond to your questions as quickly as possible via email. Sometimes, similar questions may be raised by several students, in which case I will send the entire class email in Canvas.

- **WITHDRAWING / DROPPING THE COURSE:** Keep in mind that if you are considering to withdraw the class, make sure you received all the possible support before. If you still feel necessary to do so, please check the academic calendar for properly withdraw before the scheduled last drop day. If you stop attending class, you should execute the drop procedure since failure to do so will result in a grade of “F” which cannot be changed.

- **ACADEMIC INTEGRITY:** This course adheres to the UNT policy on academic integrity. The policy can be found at [https://vpaa.unt.edu/fs/resources/academic/integrity](https://vpaa.unt.edu/fs/resources/academic/integrity). Remember that if you engage in any form of academic dishonesty, you will receive a failing grade on the test or assignment, or a failing grade in the course. In addition, the case may be reported to the UNT Dean of Students/Academic Integrity Office, which maintains a database of related violations. Students are expected to read [https://policy.unt.edu/policy/06-003](https://policy.unt.edu/policy/06-003) UNT’s Student Standards of Academic Integrity with defines academic dishonesty and sets out the consequences of unethical behavior.

- **STUDENTS WITH DISABILITIES:** The New College complies with the Americans with Disabilities Act (ADA) in making reasonable accommodations for qualified students with disability. If you have an established disability you should register with the Office for Disability Accommodation (ODA) and receive further instructions. Please see contact me as soon as possible if you have any questions.

- **DEADLINES:** Dates of drop deadlines, final exams, etc., are published in the university catalog and the schedule of classes. Please be sure you keep informed about these dates.
• INCOMPLETE GRADE (I): The grade of "I" is not given except for rare and very unusual emergencies, as per university guidelines. An “I” grade cannot be used to substitute your poor performance in class. Do not let that happen and contact me as soon as possible!
• CAMPUS CLOSING: In the event of an official campus closing, please check your UNT e-mail for instructions on how to turn in assignments, how the due dates are modified, etc.

Schedule and weekly learning goals

The description and timelines contained in the syllabus are subject to change at Professor’s discretion based on the group learning’s pace or other events that may impact the current schedule.

It is meant to be a guide and several items are subject to change. Exams may be moved in time & will be announced. It is STRONGLY recommended that you adhere to the schedule outlined below. This way, you should progress at a reasonable, sustainable pace. You will also be able to keep up with all duedates.

Schedule and Weekly Learning Goals

<table>
<thead>
<tr>
<th>Main Topic</th>
<th>Dates</th>
<th>Sub Topics</th>
<th>Assignments, Cases, Exams, &amp; Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Management &amp; Wrangling</td>
<td>1/16 – 1/18</td>
<td>Introductions &amp; Course Syllabus</td>
<td>Read the syllabus</td>
</tr>
<tr>
<td></td>
<td>1/23 – 1/25</td>
<td><strong>L.0</strong> Introductory concepts: Data Management &amp; Wrangling</td>
<td>Assignment 1</td>
</tr>
<tr>
<td></td>
<td>1/30 – 2/1</td>
<td><strong>L.1</strong> Data Management Basics, Integration &amp; Techniques</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2/6 - 2/15</td>
<td><strong>L.2</strong> Relational Database Basics &amp; Database Management System</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2/20 – 2/22</td>
<td><strong>L.3</strong> Data Retrieval: Introduction to Relation Database Concepts (Access / SQL)</td>
<td>Assignment 2</td>
</tr>
<tr>
<td></td>
<td>2/27 – 3/5</td>
<td><strong>L.4</strong> Data Inspection: Counting and Sorting Excel functions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3/7</td>
<td>Demonstration of R &amp; R Studio Basics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3/19 – 3/21</td>
<td><strong>L.5</strong> Data Preparation: Handling missing values &amp; Sub setting</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>L.6</strong> Transforming Numerical Data: Binning and Mathematical Transformation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3/26 – 3/28</td>
<td><strong>L.7</strong> Transforming Categorical Data</td>
<td>Assignment 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>MIDTERM</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3/26 – 3/28</td>
<td><strong>L.8</strong> Decision Model Components: Deterministic &amp; Stochastic</td>
<td>Assignment 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>L.9</strong> Spreadsheet Engineering: Influence Diagrams</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>L.10</strong> What-If Analysis Data table, Goal Seek, and Scenario Manager</td>
<td>Assignment 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>L.11</strong> Avoiding &amp; Detecting Errors in Spreadsheet Modeling</td>
<td>Assignment 7</td>
</tr>
</tbody>
</table>
| Data Visualization | 4/2 – 4/11 | **L.12** Data Visualization  
Visualize A Single Variables | Assignment 8 |
|-------------------|------------|-------------------------------------------------|--------------|
| 4/16 - 4/30       | **L.13** Pivot-Tables Introduction  
Simple Queries & Charts  
Work on Project Data Connections | Assignment 9 |
| 5/2 – 5/9         | Practice for Final / Work on Final Project  
Practice for Final / Work on Final Project | Final Exam / Project |