DSCI 325: Management of Structured Data
Spring 2026

Instructor: Chris Malone
Office: Gildemeister 141
Email: cmalone@winona.edu
Website: www.statsclass.org

Text: There is no text required for this course. The following references may be useful.
- Modern Data Science with R by Baumer, B., Kaplan, D., Horton, N.
  Link: https://mdsr-book.github.io/mdsr2e/index.html
- Python Data Science Handbook by VanderPlas, J.
  Link: https://jakevdp.github.io/PythonDataScienceHandbook/

Course Description: This course will give students an overview of the issues related to the management of structured data. Topics to be covered in this course include: data warehousing, data integrity and quality, data cleansing, basic programming concepts, the construction of simple algorithms, and the appropriate descriptive and graphical summaries of data. Commonly used software packages for the analysis and management of data will be emphasized.

Prerequisites: DSC 210 Data Science

Learning Outcomes: A student who has successfully completed this course will be able to:
- demonstrate an advanced aptitude for cleaning, preparing, and combining data from multiple sources.
- develop and apply a data management process in multiple technologies.
- demonstrate the apply window functions to solve appropriate problems.
- conduct a data science project which will require the formulation of the problem.

Assessments:
>   Tasks (about 50% - 66% of grade)
>   You will be asked to complete several small tasks to demonstrate your mastery of concepts discussed during a particular class period. In addition, students will be required to complete more substantial tasks which will involve the use of several concepts / methods. Work submitted past the deadline will be assessed a 10% penalty for being late and will not be accepted solutions are posted.
>
>   Exams / Projects (About 33% - 50% of grade)
>   There will be a combination of exams and/or projects in this course. These exams / projects will be substantial in nature and will be worth between 33% and 50% of your grade. These assessments will test your ability to adapt your data management skills to new situations and/or extend your knowledge of methods presented in class. Exams will likely be a combination of in-class and out-of-class.

Grades:

Your grade will be determined by your performance on tasks and exams / projects. The tasks will be worth slightly more than the exams / projects. I do no weighting, so a point is worth a point in this class.

Your final grade will be determined using the following percentages.

Your Percentage	Grade
90% of greater	A
80% - 89.9%	B
70% – 79.9%	C
60% - 69.9%	D
Less than 60%	F

Extras:

- I encourage you to stay organized with the course content – there will be a combination of note taking, handouts, on-the-fly programming, and impromptu discussions.
- Attendance is mandatory. If you miss class, it is your responsibility to get the material and get yourself caught up.
- If necessary, I reserve the right to make policy changes for this course as the semester progresses.

Topics Covered – Course Outline:

1. Review of tidy data and data verbs
   a. Tidy data and the golden rule
   b. Select, filter, mutate, group aggregation
   c. Joins, unions, reshaping
2. Introduction to regular expressions
   a. Basic regular expression syntax
   b. Filtering and transforming textual data using regular expressions.
3. Introduction to window function
   a. Simple and partitioned windows
   b. Applications: column statistics, standardizing, smoothing
   c. Adding order and order statistics (ranks based off order)
   d. Applications: Overall rank, group rank, filtering by rank
4. Introduction to SQL
   a. Basics of SELECT, FILTER ,and MUTATE
   b. Data verbs related to FILTER: NULL, IN, GLOB, LIKE, NOT
   c. Transforming string and dates
   d. GROUPBY and SUMMARIZE
   e. Pivot tables.vi. Subqueries
      Notes: SQL Lite Studio Making databases for notes/assignments

5. Advanced Data Verbs in SQL (3 weeks)
   a. Set-like operations on similar tables: UNION, INTERSECTION, and DIFFERENCE
   b. Various table joins: FULL OUTER, LEFT, RIGHT, INNER
   c. Joining 3+ tables
   d. Simultaneous JOIN + GROUPBY + SUMMARIZE
   e. Joins and filters: WHERE VS HAVING
   f. Joins and aggregation: Window functions
6. Window functions in SQL
   a. Basic summaries and ranking using PARTITION BY or ORDER BY
   b. Defining windows with the WINDOW clause
   c. Defining frames by ROW or RANGE
   d. Applications: Standardizing, ranking, filter via rank, smoothers
7. Introduction to Data Management in Python
   a. SELECT, FILTER, MUTATE
   b. GROUPBY and SUMMARIZE
   c. Joins, unions, and reshaping
   d. Window functions
8. Python Data APIs
   a. Basic concepts for APIs
   b. Connecting to an API/authentication
   c. Examples of Python interfaces to APIs