

CS 350/550 Computational Tools and Techniques for Data Analysis

Section 1 - Spring 2008 M, W, F 2:45 – 3:35 p.m. RC 153

Last Updated: March 30, 2008

Description: Introduction to the representation, visualization, and modeling of large data sets. Data analysis using standard high level software tools. Topics include data filtering, clustering, classification, and data mining. A familiarity with Excel is assumed plus some exposure to programming in languages such as C/C++ or Java. Knowledge of basic statistics is useful but not required. Material covered will be relevant to applications areas in science and engineering. Four credit hours.

Prerequisites: None.

Instructor: Dr. Ronald F. Taylor, RC 340, 775-5122, ronald.taylor(at)wright.edu, 6:30 – 7:30 p.m. on Monday and Wednesday also 2:00 – 3:00 p.m. on Tuesday and Thursday (other times by appointment).

Required Textbook: Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining, Glenn J. Myatt, John Wiley & Sons, 2007, ISBN-13: 978-0-470-07471-8.

Recommended Textbook: Advanced Excel for Scientific Data Analysis, Robert de Levie, Oxford University Press, 2004, ISBN-13 978-0-19-517089-4.

Course Home Page and WebCT: <http://www.cs.wright.edu/people/faculty/rtaylor/cs350> available by the start of second week of class. We will also be using WebCT for posting of grades and submittal of some assignments or portions of assignments. Students should familiarize themselves with accessing WebCT: <http://wisdom.wright.edu/>. Students are also responsible for accessing the Course Home Page or WebCT for printing copies of resource materials. Some handouts will be given in class.

Software: Using numerical programs for data analysis is an important part of this course. Assignments will include the use of Excel, MATLAB, and possibly some scripting languages. In lectures we will use Microsoft Office Excel 2007 software and students should consider upgrading home computers to the latest version if necessary. Many times numerical work can be done on a scientific or programmable calculator. MATLAB is very useful, and you may want to consider purchasing the Student Edition if you have a PC that can support it. It is expected that students will spend a minimum of 2 hours per week working in a computer lab or equivalent environment enhancing their data analysis skills and working on assignments for this course.

Computers and Computing Accounts: You must be able to access the Web and have a WSU Student Login to Wings, e-mail, and WebCT. Check your WSU e-mail on a regular basis for any course announcements from the instructor. Get familiar with the use of the PCs in Russ Center 152C to access software such as Excel and MATLAB if you do not have it on your own PC. Needed computing topics be covered in class and handouts or web citations given as appropriate. Check the University computing information at <http://www.wright.edu/cats/help/guides/students/index.html> as well as that for the College of Engineering and Computer Science at: <http://www.cs.wright.edu/help/services.shtml>

Grading Policy: Mid-term exam and quizzes – 35% . One comprehensive final – 40%. Homework/Project assignments – 25%. Quizzes may be in class or take-home: points included with mid-term score. Students registered at the graduate level (i.e. CS 550) will be required to complete extra problems, programs and/or special projects as part of the Homework/Project component of this course. Expect about six major Homework/Project assignments. Some problems assigned will be considered "practice" and may not be graded. In general, one week will be given to prepare these assignments. Smaller homework problems/investigations may be due the next class period. Student presentations of projects may be included in the course. Grade Based on Course Average:
A: 100-90, **B:** less than 90-80, **C:** less than 80-70, **D:** less than 70-60, **F:** less than 60-0.

Class Policies: No late or early exams unless verifiable emergency. No make-up quizzes: quizzes may be unannounced. Attendance at lecture is not a component of your grade. However, students are expected to attend all lectures and to participate in class discussion. Attendance may be taken in the course to better get to know students. In cases of infrequent attendance, lower homework and exam grades will inevitably result since a significant portion of lecture material is not covered in the text. All Homework/Project assignments are due at the start of class and/or in WebCT on the date and time specified. Grades on late assignments will be reduced by 10%. Submittals more than one day late will not be graded - "zero" grade assigned. Exceptions to the above policies may be made unusual circumstances when documentation is provided in writing -- otherwise expect strict enforcement of the policies. All work submitted must be your own unless group assignments are explicitly made by the instructor; sharing of program code or copying problem solutions/codes from any source will result in at least a homework grade of "zero" for all involved and possibly a grade of "F" for the course. University procedures for plagiarism will be strictly followed. Sharing ideas and general mathematical and computer skills with others outside of class is encouraged. Students are expected to read, understand and follow the University Academic Integrity Policy at: <http://www.wright.edu/students/judicial/integrity.html>

Schedule: Topics will vary. At times the textbook readings are supplemental to the topics presented in lecture. We do expect to cover most of the Myatt text and several key chapters of de Levie. A copy of the de Levie text is useful, but not required to follow the lectures and complete the assignments. "Notes" refer to "lecture notes" and other recourses presented during class. **Exams dates and times** are firm.

Week	Topic/Tests etc.	Readings/Reference
1	Overview of fundamental problems related to data acquisition and analysis	Myatt Chapter 1, 2
2	Storage, manipulation, and visualization of using spreadsheets	Myatt Chapter 3 de Levie Chapter 1, Appendix
3	Use of spreadsheets continued.	Myatt Chapter 4 and Notes
4	Data normalization, missing or corrupted data values, filtering. Introduction to scripting languages.	Notes and Myatt Chapter 5, 6
5	Introduction to tools for modeling and simulation (MATLAB etc). (Mid-Term Exam – Friday May 2nd, full class period)	Notes
6	Modeling and simulation tools continued. Excel and MATLAB applications.	Notes and Myatt Chapter 7
7	Introduction to visualization. Excel applications.	Notes
8	Visualization continued and overview of pattern recognition.	Notes
9	Introduction to data mining and supporting tools. Excel applications.	Myatt Chapter 8, 9
10	Special topics and review.	de Levie Chapter 2
Finals	Comprehensive Final – Wednesday June 11th 3:15-5:15 p.m.	