EEES 4160 Environmental Data Management/
EEES 6160 Advanced Environmental Data Management **Fall 2015**

1. General Information

Instructor:	Song S. Qian	Term	Spring 2015
Email	song.qian@utoledo.edu	Class Day/Time:	MW 2:00-3:15
Office Hours:	MW 1:00-2:00; R 1:00-2:30	Class Location:	BO 1005
Office Phone:	419 530 4230	Credit Hours:	3
Office Location:	BO 3001E		

2. Course/Catalog Description

   A curse in data management for environmental science graduate students and undergraduate seniors, covering advanced data management practices and the use of R for data preparation, evaluation, analysis, visualization, and interpretation. Prerequisite: EEES 2500 or approval of instructor.

3. Course Overview:

   The course introduces techniques and practices in data management, including data preparation, processing, exploratory analysis, visualization and interpretation. The course builds upon students' knowledge in the use of Excel for initial data preparation and processing, and focuses on the use of R, an object oriented programming language, for data visualization, data processing, and exploratory data analysis. The course emphasizes the best practices in data management and prepares students for their subsequent courses and research.

   Materials covered in the class include:

   (a) Introduction – types of data, data analysis as a science, history of data science, relationship to statistics.

   (b) Initial data processing in Excel – importing text data files, manipulating data sheet, basic summary statistics, creating a metadata sheet, exporting data.

   (c) Data management using R – getting started with R and RStudio, data formats in R, data manipulation.

   (d) R programming basics – importing/exporting data from/to Excel and other formats, basic summary statistics, subsetting and merging data files, writing and keeping R script files.

   (e) Data visualization in R – univariate data, bivariate data, multivariate data, categorical data.

   (f) Trellis plots and the concept of conditioning.

   (g) Data evaluation – identifying and documenting potential outliers and other unusual features (using trellis plots), documenting the methods used for recording data values below method detection limits.

   (h) Documenting data analysis using dynamic documents in R – using R markdown document for programming and documentation.

   (i) Class project for graduate students – analyzing a data set from USGS, NOAA, EPA (STORET), or state agencies, and writing a report, including an introduction to the data (why it was collected, where, and how), a summary of main features of the data, plots for visualizing these features, and a tentative interpretation. An R markdown document with detailed R script should be accompanied with the report to document steps used to produce the report.

4. Student Learning Outcomes

   Upon completion of this course, the student will be able to:

(a) Import data from a spreadsheet format to R

(b) Perform necessary data summary and exploratory analysis

(c) Perform quality check on the data

(d) Use advanced graphics for exploratory data analysis

(e) Perform necessary data aggregation and extraction for subsequent analysis

(f) Perform basic programming in R

5. Teaching Strategy

   This course is designed to stimulate student learning through hands-on programming and analysis of real world data in environmental and ecological studies, both as in-class examples and in homework assignments. Students are expected to read all assigned readings before each class and actively participate in-class discussions and demonstration. Lectures include both programming basics and strategies with an emphasis on creative thinking.

6. Prerequisite

   EEES 2500 or approval of instructor.

7. Required Texts and Ancillary Materials

   (a) Main text – J. Stanton, 2013. Introduction to Data Science. An interactive electronic book for iPad and iBook, available free of charge from iTune. Also available as a PDF file.

   (b) R reference text – Petra Kuhnert and Bill Venables, (2006). *An Introduction to R: Software for Statistical Modelling & Computing*, available at the CRAN site, `http://cran.r-project.org/doc/contrib/Kuhnert+Venables-R_Course_Notes.zip`)

   (c) Recommended reading –

   - W. Cleveland, 1993. *Visualizing Data*, Hobart Press, Summit, NJ.

   (d) Software:

   - R – an open source implementation of the S Language, available from CRAN at `http://cran.r-project.org`,
   - RStudio – an open source integrated development environment (IDE) for R.

   (e) Data sets – water quality and water resources data from USGS, climate data from NOAA, environmental monitoring data from EPA and state agencies.

8. Course Expectations

   Class rules

   (a) Students are expected to attend weekly lectures and actively participate in class activities.

   (b) Late work will not be accepted without prior consent from the instructor.

   (c) Students requesting re-grade must make these requests within one week of receiving the graded material. Attach a note explaining the re-grade issue to your homework or exam and submit to instructor. The instructor has the option to re-grade the entire homework or exam.

   (d) Take-home exams must be completed independently.

   Course work expectations

   (a) Data project is an important part of the class. Students are encouraged to select their own project topics in the first two weeks of the semester.

(b) Examples are an important part of the class. But any given example is unlikely to be intuitive to all. Students are encouraged to submit/recommend examples.

(c) The objective of data management and processing is to facilitate subsequent statistical analysis. Students are expected to learn common data formats for statistical modeling in R.

(d) Students are encouraged to study together in order to better understand the materials. Students must complete individual assignments independently. Exams (including take-home exams) must be completed independently.

9. Grading Scheme:

   - Individual homework (40% for EEES 6160 and 60% for EEES 4160) – 10 homework assignments (using Excel and R for data manipulation and visualization)
   - Midterm and Final exams (30% and 40% for EEES 6160 and 4160, respectively)
   - Course project (30% for EEES 6160 only)

   Each category will be graded in percentage and the final score is the weighted average of the three categories (also in percentage).

   Course grades will be curved based on the final score. But in general, A (>90%), A- (85-90%), B (75-85%), C (50-75%), F (<50%).

   Students who do not attend class or stop attending at some point throughout the semester will be given a final grade of F which will impact your overall grade point average. To formally withdraw from this or any other course you need to contact the Registrars Office.

10. Communication Guidelines

    As your instructor, I am here to help, and will do my best to respond to mail within 24 to 48 hours. Students are expected to check their UT email account frequently for important course information. In addition, if you are having difficulty in the course or trouble understanding any aspect of it, please let me know.

    A question-answer session will be held during Mondays lecture. Questions submitted via Email will be addressed on the following Monday.

11. University Policies

    - Policy Statement on Non-Discrimination on the basis of Disability (ADA)

      The University is an equal opportunity educational institution. Please read The Universitys Policy Statement on Nondiscrimination on the Basis of Disability Americans with Disability Act Compliance.

    - Academic Accommodations

      The University of Toledo is committed to providing equal access to education for all students. If you have a documented disability or you believe you have a disability and would like information regarding academic accommodations/adjustments in this course please contact the Student Disability Services Office.

12. Academic Policies

    As a student in my course and enrolled at The University of Toledo you should be familiar with the policies that govern the institutions academic processes, for example, Academic Dishonesty, Enrollment Status, and Grades and Grading. Please read Graduate Academic Policies.

    Students are expected to attend every class meeting of courses in which they are registered. Please read the Missed Class Policy.

13. Course Schedule

Date	Topic	Readings	Contents/Activities	Assignment
Week 1 8/24,26	Overview Introduction to R, RStudio, and R Markdown	JS Chs 1-2 RStudio introduction Data objects, dates	Using R and RStudio Writing using R Markdown	HW1: K&V Lab1
Week 2 8/31,9/2	R Basics Import data Prepare data in Excel	JS Ch 3 K&V *R Objects*	Reading data Writing R data Loading & saving	HW 2: K&V Lab2
Week 3 9/9	Graphics in R K&V: Graphics: An Introduction	JS Chs 4-6 bivariate and multivariate data	Display distribution scatter plots scater plot matrix	
Week 4 9/14,16	Examples Finnish Lakes Neuse River chla		data presentation Factor object	HW3: K&V Lab 3
Week 5 9/21,23	Data manipulation	JS chs 4-6, K&V pp 97-108	Subscripting Aggregation Merging	HW 4: K&V Lab 4
Week 6 9/28,30	Data manipulation examples, functions Review	JS ch 7	More on aggregation	
Week 7 10/7	Midterm	JS Chs 1-7 K&V pp1-108	(take home)	
Week 8 10/12,14	Class Projects Requirements		Character objects Maps	HW 5: project proposal
Week 9 10/19,21	Data Manipulation Reshaping data		package `reshape2`	
Week 10 10/26,28	Projects Programming: `for` loop		Data, objectives approaches	Proposal presentations
Week 11 11/2,4	Examples USGS-EUSE data FL Nutrient data			HW 6: merging
Week 12 11/9	Data in the age of Internet Trellis plots	JS Chs 10-13		HW 7: subsetting
Week 13 11/16,18	Advanced graphics Trellis plots	K&V Handout	EDA	HW 8: K&V Lab 9
Week 14 11/23	Class project Conditional plots			HW 9: Conditional plots
Week 15 11/30,12/2	Programming Graphics		Functions Math symbols	HW10: using `reshape2`
Week 16 12/7, 9	Student Presentations			
12/16	Final Exam			

Last modified: Wednesday 19th August, 2015, 12:48