GEOL 5303: Computer Applications in the Geosciences Syllabus
Spring 2017, CRN 27752
Lecture & Lab: F 9:30-12:30 Prospect Hall 300

Instructor
Deana Pennington
Department of Geological Sciences
GEO 305C
ddpennington@utep.edu
Office Hours: During class or by appointment

Required Text: Hey, Tony, Stewart Tansley, and Kristin Tolle (2009). The fourth paradigm: data-intensive scientific discovery. Microsoft Research Publication. Redmond, WA.
Paperback available from Amazon.com
Kindle Edition $0.99
Free PDF download here:
https://blogs.microsoft.com/next/2011/08/11/the-future-of-technology-for-just-0-99/#sm.0001xf8cyjcnsdj010dl3uh5tjapp

Several journal articles will also be assigned, available through UTEP's library.

General Information
The ever-evolving technology landscape has generated a plethora of new data types, data acquisition methodologies, and data analytic approaches. The pervasiveness of readily available data of all kinds is fundamentally changing science, which traditionally has often lacked sufficient data. At the same time, voluminous data does not ensure the right data for a given scientific purpose; nor does it mean that existing relevant data can be found or easily manipulated. Indeed, a scientist hardly knows where to start -- yet 21st scientists must learn to navigate these waters. This course will be a first attempt at preparing earth and environmental science students to embrace emerging technologies for "big data" and the data management skills required to use them.

Class will be taught in one, 3-hour session each Friday in the new Socio-Environmental and Geospatial Analysis (SEGA) lab in Prospect Hall 300. Each class will be comprised of diverse activities, including a combination of short lecture, hands-on tutorials, remote presentations, group work, and full group discussion.

Prerequisite: A willingness to try new things, determination to work through the inevitable technical glitches, and enthusiasm for integrating a rich new set of ideas into your knowledge.

Course Objectives
- Develop a basic understanding of the core concepts of data-intensive science.
- Foster ability to solve complex geoscience problems using advanced techniques
- Exposure to a breadth of data management, analysis and visualization approaches
- Develop knowledge of existing and emerging tools, techniques, and methods
- Develop a vision of future geoscience technical platforms currently under development

Academic dishonesty
A student's submission of work for academic credit indicates that the work is the student's own. Any outside assistance should be acknowledged. While cooperation during class is

encouraged, assignments must be constructed and written by each individual student unless a collaborative approach is specifically assigned.

Makeup policy
Due dates are firm. 10% will be deducted for each day an assignment is late. Assignments will not be accepted more than one week late. Reading summaries are due at the start of class and are intended to facilitate discussion. No reading summaries will be accepted late.

Students with Disabilities
Students with disabilities are referred to the Center for Accommodations and Support Services (CASS; http://sa.utep.edu/cass/) who will work with the instructor to identify appropriate adaptations to better foster a positive teaching and learning experience.

Drop Deadline
The deadline to drop this class is March 30. No requests for a withdrawal will be approved by the College of Science after that date.

Grading

Class participation	10%
Glossary	10%
Reflective writing (2)	20%
Technical activities (4)	40%
Group presentations	20%

Class participation
The class participation grade will be based on a combination of attendance, assigned tutorials, any preparatory assignment you are given to complete before class, plus having a question or comment for guest presenters. Active participation will make the class much more interesting and engaging for everyone. Please come to class prepared to contribute your thoughts and ideas. This is especially true when we have a guest presenter.

Rubric: Students will receive 1 point for each class they attend, tutorial they complete, and question they ask guest presenters.

Glossary
Students will collaboratively develop a glossary of data science terms. Each student must submit a minimum of ten terms and their definitions. Gain extra credit by submitting more terms, or being particularly helpful in defining terms.

Rubric: Students will receive 1 point for each full contribution they make to the glossary. A full contribution is a word or phrase plus its definition. Adding a word without a definition or defining a term someone else contributed will earn ½ point.

Reflective Writing (2 assignments)
Students will be asked to reflect on their perceptions of the role of emerging technologies in their field at the beginning and end of the class, and write a one page, single spaced reaction. Reflective writing is expected to demonstrate thoughtful consideration of all of the course content. “Thoughtful” implies “full of thought.” These must show that the student is thinking about the material deeply.

Technical activities (4)
There will be four activities assigned, that require students to explore particular technologies more deeply than possible in class. These will require students to take the material presented in class and delve into how those technologies are being applied within the student's field of interest. This will include doing a literature review combined with a survey of tools and techniques, along with a technical assessment of the usefulness and usability of available tools.

Ratings: E = Excellent, VG = Very Good, G = Good, F = Fair, P = Poor, M = Missing

ITEM	WORTH	RATING	EARNED
Introduction to the topic(s). Summarize the information provided in class, in assigned reading, and encountered on the web. Do not list specific projects or tools in this section; rather, describe general approaches.	10		
Cite each reading assignment at least once, and include in a list of references.	10		
Create a specific list of projects, tools, or methods relevant to the topic(s), and describe who, what, where, when, why, and how for each. Who is involved, both as creator and/or intended user? What is it about? Where is the work occurring - related institutions and/or field sites? When is the work occurring – currently, in the past, or in the future? Why is the work occurring – what problems are they trying to address? How does it work, if you can easily figure this out. Your list should include items mentioned during the lecture, but must also include at least five you discovered on your own, related to your field.	50		
Reflect on potential relevance to your work. How could you make use of these methods and tools?	20		
Grammar and writing style.	10		
Vocabulary. List the vocabulary words you contributed to. Used for separate vocabulary grade.	0		
Total			

Group presentations (1 assignment)
Group presentations will revolve around assessing the current and potential future states of technology in different geoscience disciplines. Students will be assigned to a group that works on disciplines related to each student's primary interests.

Ratings: E = Excellent, VG = Very Good, G = Good, F = Fair, P = Poor, M = Missing

ITEM	WORTH	RATING	EARNED
Content is well organized and well designed. The presentation is logical and flows from title, introduction, body, and conclusion. Contextual information is early, and details come later.			
Slides contain minimal text, supporting the presentation primarily with visuals and bulleted short phrases			
Preparedness: Presenters were well practiced and professional.			
The presenters were clear, articulate and concise. Their language was appropriate for the audience.			
Presenters engaged audience, with questions and discussion			
Total			

TENTATIVE SCHEDULE (subject to change)

DATE	TOPIC	READING	ASSIGNMENT DUE
20-Jan	Introduction to data-intensive science		
27-Jan	Version Control & GitHUB (Dr. Pennington out of town – Perry Houser will facilitate)	Foreword, Jim Gray on eScience; Geoscience 2020	Reflective writing 1: Data science in your field
3-Feb	Data collection	Pt 1: Instrumenting the earth	
10-Feb	Data management		DC & DM report
17-Feb	Data integration	Pt 1: Gray's Laws: Database-Centric Computing; Pt 3: Platform for all that I know	
24-Feb	Data analysis	Peters et al. (2014); Bui (2016)	
3-Mar	Geology Spring Colloquim -- No class: Work on data integration, mining & machine learning activity		DI, DM & ML report
10-Mar	Data processing	Pt 3: Impact of Workflow Tools	
17-Mar	Spring Break; No class: Work on scientific workflow activity		SW report
24-Mar	Visualization & visual analytics	Pt 3: Visualization	
31-Mar	Holiday No class: Visualization activity		VIS report
7-Apr	Work in groups	Pt 4: From Web 2.0 to Global Database	
14-Apr	Spring study day No class: work in groups		
21-Apr	Presentations		Group presentation
28-Apr	Dissemination & Sharing; Curation & Archiving	Pt 4: All Aboard: Toward a Machine Friendly Scholarly Communication System	
5-May	Dead day		
	Final reflective writing		