CSCI 556 Scientific Data Analysis and Visualization

COURSE SYLLABUS: FALL, 2017

Instructor: Mehmet Malcok, Ph.D.

Office Location: Virtual Office Online (Web based course)

Office Hours: Email can be sent any weekday and usually gets a response within 24 hours

Office Phone: Not Available, use online contact via email

TAMUC Email Address: Mehmet.Malcok@tamuc.edu

COURSE INFORMATION

Materials – Textbooks, Readings, Supplementary Readings:

Textbooks Required:

1. Practical Data Science With R

By Nina Zumel and John Mount,

Publisher: Manning, ISBN 13: 978-1-617291-56-2

Optional Reference Books:

1. The Art of R Programming
   By Norman Mattloff

A free version is available at: http://heather.cs.ucdavis.edu/~matloff/132/NSPpart.pdf

2. An Introduction to Statistical Learning, with Applications in R,

   By Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani.

   A free version of this book is available at:

   http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Fourth%20Printing.p

Course Description:

Big scientific data sets are growing exponentially both in size and complexity. Extracting meaningful information from these data requires not only programming skills, but also understanding the analysis work-flows, mathematical models and visualization tools that help to condense large amounts of information into a comprehensible story. We will introduce standard statistical data analysis and modeling methods such as correlation functions, linear regression, clustering, pattern extraction, classification, data mining, as well as Monte Carlo methods which are commonly used in creating simulations in the computational sciences. Different analysis and visualization packages popular in scientific modeling, analysis, and visualization will be introduced.

Objectives:

This course is for graduate students to study statistical and machine learning techniques used in data analysis.

Student Learning Outcomes:

After completing this class, the successful student should have attained the following skills:

1. **The student will gain detailed knowledge about the goal and techniques of the data analysis and visualization process.**

2. **The student will understand the steps in characterizing and understanding data.**

3. **The student will be able to build effective predictive models.**

4. **The student will be able to build models that rely on memorizing training data.**

5. **The student will be able to build models that have an explicit additive structure.**

6. **The student will be able to build models for data that has no labeled training data available: Unsupervised learning**

7. **The student will be able to use software applications for data analysis.**

COURSE REQUIREMENTS

Instructional / Methods / Activities Assessments

The text for the class will be covered in this class using the text author's slides modified, animated and highlighted by the course instructor. By the end of the semester, the students should have read the assigned chapter sections, studied the presentations, taken quizzes, participated in discussions and have spent some time to be certain they have a general understanding of the reference material in the appendices. The table of contents of the text is basically an outline for the course.

All quiz activity, required discussion inputs, and any other assignments are provided to the student as each class unit is opened and the week progresses.

Grading

Discussions: **10%**

Homework Assignments   25%

Quizzes: **20%**

Midterm Exam: **20%**

Final exam: **25%**

Letter grades will be assigned according to the following scale, rounded to nearest tenth of a point according to Excel worksheet process:

A = at least 90.0% of total

B = at least 80.0% but less than 90.0%

C = at least 70.0% but less than 80.0%

D = at least 60.0% but less than 70.0%

F = less than 60.0% of the total points

For Example: 89.94 rounds to 89.9 and would award you a B grade

      But: 89.95 rounds to 90.0 and would award you an A grade.

TECHNOLOGY REQUIREMENTS

• To fully participate in online courses you will need to use a current Flash enabled internet browser. For PC and Mac users the suggested browser is Mozilla Firefox.

• You will need regular access to a computer with a broadband Internet connection. The minimum computer requirements are:

o 512 MB of RAM, 1 GB or more preferred
o Broadband connection required courses are heavily video intensive
o Video display capable of high-color 16-bit display 1024 x 768 or higher resolution

• You must have a:

o Sound card, which is usually integrated into your desktop or laptop computer
o Speakers or headphones.
o *For courses utilizing video-conferencing tools and/or an online proctoring solution, a webcam and microphone are required.

• Both versions of Java (32 bit and 64 bit) must be installed and up to date on your machine. At a minimum Java 7, update 51, is required to support the learning management system. The most current version of Java can be downloaded at: JAVA web site http://www.java.com/en/download/manual.jsp

• Current anti-virus software must be installed and kept up to date.

• Run a browser check through the Pearson LearningStudio Technical Requirements website. Browser Check http://help.ecollege.com/LS_Tech_Req_WebHelp/en-us/#LS_Technical_Requirements.htm#Browset

Running the browser check will ensure your internet browser is supported.

Pop-ups are allowed.

JavaScript is enabled.

Cookies are enabled.

• You will need some additional free software (plug-ins) for enhanced web browsing. Ensure that you download the free versions of the following software:

o Adobe Reader https://get.adobe.com/reader/
o Adobe Flash Player *(version 17 or later)* https://get.adobe.com/flashplayer/
o Adobe Shockwave Player https://get.adobe.com/shockwave/
o Apple Quick Time http://www.apple.com/quicktime/download/

• At a minimum, you must have Microsoft Office 2013, 2010, 2007 or Open Office. Microsoft Office is the standard office productivity software utilized by faculty, students, and staff. Microsoft Word is the standard word processing software, Microsoft Excel is the standard spreadsheet software, and Microsoft PowerPoint is the standard presentation software. Copying and pasting, along with attaching/uploading documents for assignment submission, will also be required. If you do not have Microsoft Office, you can check with the bookstore to see if they have any student copies.

• For additional information about system requirements, please see: System Requirements for LearningStudio https://secure.ecollege.com/tamuc/index.learn?action=technical

ACCESS AND NAVIGATION

Pearson LearningStudio (eCollege) Access and Log in Information

This course will be facilitated using Pearson LearningStudio, the learning management system used by Texas A&M University-Commerce. To get started with the course, go to

myLeo and from the top menu ribbon select eCollege. Then on the upper left side of the screen click on the My Courses tab. http://www.tamuc.edu/myleo.aspx

You will need your campus-wide ID (CWID) and password to log into the course. If you do not know your CWID or have forgotten your password, contact the Center for IT Excellence (CITE) at 903.468.6000 or helpdesk@tamuc.edu.

Note: It is strongly recommended you perform a "Browser Test" prior to the start of your course. To launch a browser test login to Pearson LearningStudio, click on the My Courses tab, and then select the Browser Test link under Support Services.

Pearson LearningStudio Student Technical Support

Texas A&M University-Commerce provides students technical support for the use of Pearson LearningStudio.

Technical assistance is available 24/7 (24 hours, 7 days a week).

If you experience LearningStudio (eCollege) technical problems, contact the LearningStudio helpdesk at 1-866-656-5511 (toll free) or visit Pearson 24/7 Customer Support Site http://247support.custhelp.com/

The student help desk may be reached in the following ways:

Chat Support: Click on *'Live Support'* on the tool bar within your course to chat with a Pearson LearningStudio Representative.

Phone: 1-866-656-5511 (Toll Free) to speak with Pearson LearningStudio Technical Support Representative.

Accessing Help from within Your Course: Click on the *'Tech Support'* icon on the upper left side of the screen inside the course. Then you will be able to get assistance via online chat or by phone.

Note: Personal computer and internet connection problems do not excuse the requirement to complete all course work in a timely and satisfactory manner. Each student needs to have a backup method to deal with these inevitable problems. These methods might include the availability of a backup PC at home or work, the temporary use of a computer at a friend's home, the local library, office service companies, Starbucks, a TAMUC campus open computer lab, etc.

Policy for Reporting Problems with Pearson LearningStudio

Should students encounter Pearson LearningStudio based problems while submitting assignments/discussions/comments/exams, the following procedure must be followed:

1. Students must report the problem to the help desk. You may reach the helpdesk at 1-866-656-5511.
2. Students must file their problem with the helpdesk and obtain a helpdesk ticket number
3. Once a helpdesk ticket number is in your possession, students should email me to advise me of the problem and provide me with the helpdesk ticket number.

4. I will call the helpdesk to confirm your problem and follow up with you

PLEASE NOTE: Your personal computer and internet access problems are not a legitimate excuses for filing a ticket with the Pearson LearningStudio Help Desk. Only Pearson LearningStudio based problems are legitimate reasons to contact the Help Desk.

You strongly are encouraged to check for your internet browser compatibility **BEFORE** the course begins and take the Pearson LearningStudio tutorial offered for students who may require some extra assistance in navigating the Pearson LearningStudio platform.

myLeo Support

Your myLeo email address is required to send and receive all student correspondence. Please email helpdesk@tamuc.edu or call us at 903-468-6000 with any questions about setting up your myLeo email account. You may also access information at myLeo. https://leo.tamuc.edu

Learner Support

The One Stop Shop was created to serve you by providing as many resources as possible in one location. http://www.tamuc.edu/admissions/onestopshop/

The Academic Success Center provides academic resources to help you achieve academic success. http://www.tamuc.edu/campusLife/campusServices/academicSuccessCenter/

FREE MobilE APPS

The Courses apps for phones have been adapted to support the tasks students can easily complete on a smaller device. Due to the smaller screen size course content is not presented.

The Courses app is free of charge. The mobile Courses Apps are designed and adapted for different devices.

PEARSON COURSES	**App Title:**	**iPhone** – Pearson LearningStudio Courses for iPhone **Android** – LearningStudio Courses - Phone
	Operating System:	**iPhone** - OS 6 and above **Android** – Jelly Bean, Kitkat, and Lollipop OS
	iPhone App URL:	https://itunes.apple.com/us/app/pearson-learningstudio-courses/id977280011?mt=8
	Android App URL:	https://play.google.com/store/apps/details?id=com.pearson.lsphone

Once downloaded, search for Texas A&M University-Commerce, and it should appear on the list. Then you will need to sign into the myLeo Mobile portal.

The Courses App for Android and iPhone contain the following feature set:

• View titles/code/Instructor of all Courses enrolled in online

• View and respond to all discussions in individual Courses

• View Instructor Announcements in individual Courses

• View Graded items, Grades and comments in individual Courses

• Grade to Date

• View Events (assignments) and Calendar in individual Courses

• View Activity Feed for all courses

• View course filters on activities

• View link to Privacy Policy

• Ability to Sign out

• Send Feedback

LearningStudio Notifications

Students can be alerted to course activities via text on their mobile phones or up to two email addresses.

Based on their preferences, students can automatically receive a push notification with every new: course announcement, threaded discussion post, grade, and/or assignment without having to login to the course. Enrolled students will automatically receive email notifications for announcements and can opt out of this feature. To receive text notifications, students must opt in.

To begin setting up notifications, go into your course in LearningStudio and click on the bell-shaped Notifications icon on the main menu ribbon.

By default the student's university email address will appear. This cannot be changed in LearningStudio. Additional email addresses may be added by clicking the Add button. After all of the other selections are completed be sure to click the Save and Finish button

Access and Navigation of the Course Web Site

Access to class material and Uploading to an eCollege Dropbox are described below. However, if you have not taken a TAMUC online course previously, it is required that students take the online student tutorial for eCollege before you begin participation.

1. You can access eCollege either directly or through your myLeo account.

To connect to eCollege directly, set your browser to go to http://online.tamuc.org/ Enter your campus ID in the User ID box on left side of the screen and your password in the Password box below the User Id box (it's the same id and password that you use when logging in to your myLeo account). Click on Go to class. As long as eCollege itself is up, you can connect to eCollege this way even when myLeo is down or the university's home page is inaccessible. Or your can access eCollege through myLeo, log on to your myLeo account and click on eCollege in the bar in the upper right of the screen under the myLeo logo (there may be an additional link available in the menu on the left side of your screen). You can bypass the university's home page by going tohttp://online.tamu-commerce.edu

2. Once you are in eCollege click on the MyCourses tab (next to Home) in the upper left corner of the screen between the university logo and the date.

3. Scroll down to the bottom of the MyCourses screen and click on the class you want from the list of online and web-enhanced courses you may be taking this semester. In this case click on CSCI556 Scientific Data Analysis & Visualization.

At this point you can access course material and course tools. You should first read any announcements that are posted, and then enter the class Unit. The Course Home is the only area that is open during any time the entire 16 weeks of the Fall Semester.

To view the class material during all other Units, click on the Unit topic and view all topics or elements in the section during the scheduled open time. To make an assignment upload during an open Unit, follow these directions:

1. On the course page, click on Dropbox in the Tools menu bar at the top of the screen, then on the Dropbox page, click onSubmit Assignment.

3. In the Basket drop-down menu, click on the dropbox you want (Program 1, Discussion, Assignment1,etc.). The dropbox selected will appear in the Basket box.

4. Where it says "Attachments: Add/Remove" under the edit window, click on Add/Remove.

5. In the Add and Remove Attachments window, either type the name (including path) of the file you wish to upload in the dialog box to the right of File to Attach: , or click on the Browse button to search for the file you wish to upload. The name of the file selected will appear in the small dialog box to the left of the browse button.

6. If this is the only (or last) file you wish to upload, click on the Attach File button underneath the Browse button. If this is not the last file to be uploaded at this time, click on the Save & Attach Another File button instead of the Attach File button and browse for the next file. The name(s) of the selected file(s) will be listed in the Submit Assignment window underneath Attachments: Add/Remove. If you remember that you wanted to upload an additional file, you can click on the Submit assignment button again and browse for another file. A message to accompany the file(s) can be entered into the edit window.

7. Click on the Submit Assignment button to initiate the file transfer. The dropbox window will now show the basket name under Outbox with a document icon for each file you uploaded. You can hover the mouse over a document icon to see the name of the file uploaded. If you're not sure it's the correct document, you can click on it to download a copy of the file you uploaded.

8. Click on the key icon in the lower left corner of the screen to exit from the eCollege.

COMMUNICATION AND SUPPORT

Interaction with Instructor and Other Students Statement:

I respond to email questions from the university email. No other email address will be allowed besides my TAMUC email address from your TAMUC email address.

You are encouraged to study the material with your fellow students, but do not work together on assignments of any kind submitted and testing periods. The allowable activity must fall within the academic integrity procedures outlined by the university.

COURSE AND UNIVERSITY PROCEDURES/POLICIES

Course Specific Procedures:

Late Policy:

Quizzes, Homework Assignments and Discussions must be accomplished in the Unit specified by the date given and submitted according to the e-college online process. There will be 10% reduction each day after the due date.

Syllabus Change Policy

The syllabus is a guide. Circumstances and events, such as student progress, may make it necessary for the instructor to modify the syllabus during the semester. Any changes made to the syllabus will be announced in advance.

Make-up Policy:

Make-up will be permitted only in the case of having and providing me a formal Institution excuse.

Drops:

If you find that you cannot complete the course, please don't forget to drop the course as early as you can. If you simply disappear and stop attending, your grade will be based on only the points you have actually earned.

University Specific Procedures

Student Conduct

All students enrolled at the University shall follow the tenets of common decency and acceptable behavior conducive to a positive learning environment. The Code of Student Conduct is described in detail in the Student Guidebook.

http://www.tamuc.edu/admissions/registrar/documents/studentGuidebook.pdf

Students should also consult the Rules of Netiquette for more information regarding how to interact with students in an online forum: Netiquette http://www.albion.com/netiquette/corerules.html

TAMUC Attendance:

Since this is an online learning course, each student is expected to regularly login to the course website at TAMUC eCollege. You will be required to login regularly each week to view a unit presentation, to take a quiz or to submit a discussion. It is expect that you will need to spend approximately 3-6 hours a week watching the presentations, completing quizzes and submitting discussion requirements. This will bring your class time roughly equivalent to the weekly class attendance required for an on-campus class.

You must watch for changed announcements at eCollege or in emails to know when something has been updated, **Your activity on e-College web site is monitored and you will get instructor emails if you become deficient in your attendance or assignments**.

For more information about the attendance policy please visit the Attendance webpage and Procedure 13.99.99.R0.01.

http://www.tamuc.edu/admissions/registrar/generalInformation/attendance.aspx

http://www.tamuc.edu/aboutUs/policiesProceduresStandardsStatements/rulesProcedures/13students/academic/13.99.99.R0.01.pdf

Academic Integrity:

University, college, and department policies against academic dishonesty will be enforced. Students found guilty academic dishonesty are subject to receiving an "F". Academic dishonesty includes, but is not limited to: cheating, plagiarizing, fabricating of information or citations, facilitating acts of academic dishonesty by others, having unauthorized possession of previous examinations or assignments, submitting work done by another person, tampering with the academic work of other students, and working with anyone during a quiz or exam.

For more details and the definition of academic dishonesty see the following procedures:

Undergraduate Academic Dishonesty 13.99.99.R0.03

http://www.tamuc.edu/aboutUs/policiesProceduresStandardsStatements/rulesProcedures/13students/undergraduates/13.99.99.R0.03UndergraduateAcademicDishonesty.pdf

Graduate Student Academic Dishonesty 13.99.99.R0.10

ADA Statement

Students with Disabilities

The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities. If you have a disability requiring an accommodation, please contact:

Office of Student Disability Resources and Services

Texas A&M University-Commerce

Gee Library- Room 132

Phone (903) 886-5150 or (903) 886-5835

Fax (903) 468-8148

Email: Rebecca.Tuerk@tamuc.edu

Website: Office of Student Disability Resources and Services

http://www.tamuc.edu/campusLife/campusServices/studentDisabilityResourcesAndServices/

Nondiscrimination Notice
Texas A&M University-Commerce will comply in the classroom, and in online courses, with all federal and state laws prohibiting discrimination and related retaliation on the basis of race, color, religion, sex, national origin, disability, age, genetic information or veteran status. Further, an environment free from discrimination on the basis of sexual orientation, gender identity, or gender expression will be maintained.

Campus Concealed Carry Statement

Texas Senate Bill - 11 (Government Code 411.2031, et al.) authorizes the carrying of a concealed handgun in Texas A&M University-Commerce buildings only by persons who have been issued and are in possession of a Texas License to Carry a Handgun. Qualified law enforcement officers or those who are otherwise authorized to carry a concealed handgun in the State of Texas are also permitted to do so. Pursuant to Penal Code (PC)

46.035 and A&M-Commerce Rule 34.06.02.R1, license holders may not carry a concealed handgun in restricted locations. For a list of locations, please refer to (http://www.tamuc.edu/aboutUs/policiesProceduresStandardsStatements/rulesProcedures/34SafetyOfEmployeesAndStudents/34.06.02.R1.pdf) and/or consult your event organizer). Pursuant to PC 46.035, the open carrying of handguns is prohibited on all A&M-Commerce campuses. Report violations to the University Police Department at 903-886-5868 or 9-1-1.

COURSE OUTLINE/CONTENT:

Week1 08/28/2017-09/03/2017	Introduction to Data Science and R Programming
Week2 09/04/2017-09/10/2017	Basic Statistics with R and Loading Data Into R
Week3 09/11/2017-09/17/2017	Exploring Data
Week4 09/18/2017-09/24/2017	Managing Data
Week5 09/25/2017-10/01/2017	Linear Regression
Week6 10/02/2017-10/08/2017	Logistic Regression
Week7 10/09/6/2017-10/15/2017	K-Nearest Neighbors Classifier.
Week8 10/16/2017-10/22/2017	Hierarchical Clustering Midterm Exam
Week9 10/23/2017-10/29/2017	Visualization of Clusters
Week10 10/30/2017-11/05/2017	K-Means
Week11 11/06/2017-11/12/2017	Regression and Classification Trees
Week12 11/13/2017-11/19/2017	Bagging And Random Forests
Week13 11/20/2017-11/26/2017	Generalized Additive Models

Week14 11/27/2017-12/03/2017	Support Vector Machines
Week15 12/04/2017-12/10/2017	Kernel Methods
Final Exam Week 12/11/2017-12/15/2017	Final Exam

Disclaimer:

This syllabus, schedule, policies, procedures, and assignments are subject to change in the event of extenuating circumstances. It is the student's responsibility to read and abide by all of the instructor's emails and all course announcements which may provide updated versions of this information.