MONTGOMERY COLLEGE - Rockville Campus

Business, Science, Math, and Technology Division

Course Syllabus

I. Instructor Information

Abdirisak Mohamed
Email: Abdirisak.Mohamed@MontgomeryCollege.edu
Phone: (240) 672-7529
Office: SC/454C
Mail box: SC/250C
Office Hours:

Tuesday	Thursday

These hours are in the office SC/454C or in the classroom.

II. General Course Information

DATA 101 – INTRO TO DATA SCIENCE (3 Credits)

CRN XXXXXX

Room: SC 350

Course Description

Fundamental coursework on the standards and practices for collecting, organizing, managing, exploring, and using data. Topics include preparation, analysis, and visualization of data and creating analysis tools for larger data sets. PREREQUISITE(S): A grade of C or better in MATH 117/MATH 117A, MATH 217, BSAD 210 or consent of department. Three hours each week.

Format: This course is taught in a lecture/discussion format. Computer exercises may be done in class. Attendance and participation are expected.

Homework: Assignments must be done with the R Swirl package:
http://swirlstats.com/students.html

Projects:

There will be three projects to be submitted as RMarkdown documents. You can work in groups of 2 or 3 but it is not required. However, you are required to be compliant with the academic code of conduct.

(1) The first project will be on data cleaning:

Work on a dataset of your choice. Read the data into a dataframe and clean it using the R package dplyr, i.e., handle missing data, subset, select and filter the data.

(2) The second project is on exploratory data analysis:

Work on a dataset of your choice (e.g., the dataset of project 1). Use the R package ggplot2 to visualize the data and make scatterplots, histograms etc.

(3) The third project is on statistical inference:

. You can work on datasets of your choice. Select a random sample from the data.

Either use bootstrapping to estimate a parameter (e.g., median) of the population or use t-test to do hypothesis testing on the difference of means.

Here are some resources:

https://archive.ics.uci.edu/ml/index.php

https://www.kaggle.com/datasets

https://data.worldbank.org/data-catalog/

https://www.data.gov/

Books:

These books are all free but for some of them you have the option to pay some amount,

An introduction to Data Science by Jeffrey Stanton

The Elements of Data Analytic Style Jeff Leek

Exploratory Data Analysis with R, by Roger Peng

OpenIntro Statistics, by Diez, Barr, and Centinkaya-Rundel

R Programming for Data Science, by Roger Peng

III. Course Outcomes

Upon course completion, a student will be able to:

- Identify and describe the methods and techniques commonly used in data science.
- Demonstrate proficiency with the methods and techniques for obtaining, organizing, exploring, and analyzing data.
- Recognize how data analysis, inferential statistics, modeling, machine learning, and statistical computing can be utilized in an integrated capacity.
- Create and modify customizable tools for data analysis and visualization per the evaluation of characteristics of the data and the nature of the analysis.
- Demonstrate the ability to clean and prepare data for analysis and assemble data from a variety of sources.

IV. Grading

A. Requirements

Homework assignments, two tests, and three projects.
Attendance is required.

B. Course Grade

The course grade is distributed as follows:

Project 1	15%
Project 2	15%
Project 3	15%
Homework	20%
Test 1	15%
Test 2	20%

C. Standards

Letter grades will be determined as follows:

A	90 – 100%
B	80 – 89%
C	70 – 79%
D	60 – 69%
F	< 60%

In the final grade calculation, all partial points are rounded up to whole points.

D. Late Policy

See attendance policy under Section A. Requirements.
No late homework will be accepted.
You will not be given extra time for exams if you are late.

E. Audit Policy

Students auditing the class are expected to attend regularly. Those who do not may be dropped. "Audits" may take exams, but are not required to do so. Audit students do not earn a grade or credit for taking the course and cannot change to become a credit student in the middle of the semester.

V. Classroom Policies

A. Attendance and Withdrawal from Class

Attendance is required for this class, see Section A. Requirements.

It is the responsibility of the student to officially withdraw from a class or classes. If a student neglects to follow the official withdrawal procedure and merely ceases to attend classes, the grade of "F" will be recorded at the end of the semester.

B. Academic Honesty

Students are encouraged to work together on homework and project assignments.

No collaboration is allowed on exams. No cell-phones or any electronic device other than approved software may be used during exams.

No form of academic dishonesty will be tolerated, examples of which can be found in the Student Code of Conduct.

C. Classroom Conduct

Eating is not permitted during class time. Drinks are permitted as long as it is not disruptive to the class.

Cell phones must be turned off or silenced during class time. Anyone receiving or making a call or a text message will be asked to leave the classroom.

Students are expected to adhere to the Student Code of Conduct. If a student persists in disruptive behavior, he or she may be asked to leave the room.

In addition to course requirements and objectives that are in this syllabus, Montgomery College has information on its web site (see link below) to assist you in having a successful experience both inside and outside of the classroom. It is important that you read and understand this information. The link below provides information and other resources to areas that pertain to Student Success such as: Student Behavior (Student Code of Conduct); Student e-mail, College Tobacco Free Policy; Course Withdrawal and Refund Information; Resources for Military Service Members, Veterans and Dependents; how to access information on delayed openings and closings; how to register for Montgomery College's Alert System and how closings and delays can impact your classes.

http://cms.montgomerycollege.edu/mcsyllabus/

Finally, any student who may need accommodations due to a disability, please contact the Disability Support Services office on your campus as soon as possible (R-CB122; G-SA172; or TP/SS-ST233). If you have an accommodation letter from DSS, please contact me to discuss arrangements for your accommodations. Any student who may need assistance in the event of an emergency evacuation must identify to the Disability Support Services Office; guidelines for emergency evacuations are at:
https://cms.montgomerycollege.edu/EDU/Plain2.aspx?id=4162

If you have any questions please bring them to your professor. As rules and regulations change they will be updated and you will be able to access them through the link. If any student would like a written copy of these policies and procedures, the professor would be happy to provide them. By registering for this class and staying in this class, you are indicating that you acknowledge and accept these policies.

.

D. Cancellation of Classes

If class is cancelled, read the corresponding sections as indicated in the attached schedule. Any quiz or test scheduled for class that is cancelled will be given when the class next meets.

- If the College opens or closes at a time when more than 50% of a class period will be missed, that class will be cancelled for the day.

- If less than 50% of a class will be missed, that class will meet for the remaining portion of its regularly scheduled time.
- For the most up-to-date information regarding College openings, closings, or emergencies, all students, faculty, and staff are encouraged to sign up for email and text alerts via the Montgomery County MC ALERT at https://alert.montgomerycountymd.gov/index.php?CCheck=1

Important Dates:
JAN-30-2019– Refund-6% Drop Deadline
FEB-13-2019– No Grade,Change Audit/Credi Deadline
APR-15-2019– W Grade Drop Deadline

E. Tentative course schedule (subject to change)

Class #	*Date*	*Material*	*Due Dates*		
1	*Tuesday, January 29, 2019*	*Introduction and Course Overview*			
2	*Thursday, January 31, 2019*	*Data Science Tools and Processes R, RStudio, Git/GitHub, Rmarkdown, SWIRL*			
3	*Tuesday, February 5, 2019*	*R Programming: Vectors, Factors, Lists, Dataframes Missing values*			
4	*Thursday, February 7, 2019*	*Reproducibility with Rmarkdown, Control Structures*	*Assignment on Tools 2/7/2019*		
5	*Tuesday, February 12, 2019*	*Vectorized Operations, Functions, Subsetting*	*Swirl Assignment on R Programming 2/12/2019*		
6	*Thursday, February 14, 2019*	*Installing packages, Tidyverse, dplyr*			
7	*Tuesday, February 19, 2019*	*Reshaping, Merging, Melt, Cast*			
8	*Thursday, February 21, 2019*	*Group_by, Reading Data, Dates, regualr expressions*			

9	*Tuesday, February 26, 2019*	*Relational Database: SQL Projection, Aggregation.*	*Swirl Assignment on Data Cleaning 2/26/2019*		
10	*Thursday, February 28, 2019*	*Presentation Project 1*	*Project 1 2/28/2019*		
11	*Tuesday, March 5, 2019*	*Exploratory Data analysis: Principles*	*Test 1*		
12	*Thursday, March 7, 2019*	*Base Plot,*	*Assignment on SQL : DataCamp 3/5/2019*		
13	*Tuesday, March 12, 2019*	*Spring Break*			
14	*Thursday, March 14, 2019*	*Spring Break*			
	Tuesday, March 19, 2019	*Probability Distributions and Sampling in R*	*Swirl Assignment on EDA 3/19/2019*		
15	*Thursday, March 21, 2019*	*Getting Data In and Out of R: Reading files, Twitter API*			
16	*Tuesday, March 26, 2019*	*EDA example: ggplot2, Simpson's paradox*			
17	*Thursday, March 28, 2019*	*Statistical Inference: Hypothesis Testing in R*			
18	*Tuesday, April 2, 2019*	*Frequentist vs. Bayesian, Kolmogorov's axioms of Probability*			
19	*Thursday, April 4, 2019*	*Presentation Project 2*	*Project 2 4/4/2019*		
20	*Tuesday, April 9, 2019*	*Simulations in R:*			
21	*Thursday, April 11, 2019*	*Binomial, Poisson, and Normal Distributions*			
22	*Tuesday, April 16, 2019*	*Central Limit Theorem, Confidence Intervals*	*Swirl Assignment for Inference 4/16/2019*		
23	*Thursday, April 18, 2019*	*Causal Inference*			
24	*Tuesday, April 23, 2019*	*P-values, Power*	*Test 2 4/23/2019*		
25	*Thursday, April 25, 2019*	*Multiple Testing, adjusted P-value*			

26	*Tuesday, April 30, 2019*	*Simulation in R: Bootsrapping*			
27	*Thursday, May 2, 2019*	*Ethics in Data Science*			
28	*Tuesday, May 7, 2019*	*Machine Learning: Linear Regression*	*Project 3*		
29	*Thursday, May 9, 2019*	*Bayesian approach to Linear Regression*			

The professor reserves the right to make changes to this syllabus.
All changes will be made in writing.

Last updated April 1, 2019