Data Sources and Manipulation

Catalog Description

Examines approaches to locating, acquiring, manipulating, and disseminating data. Imperfection, biases, and other problems in data are examined, and methods for identifying and correcting such problems are introduced. The course covers other topics such as automated collection of large data sets, and extracting, transforming, and reformatting a variety of data and file types.

Extended Course Description

This course will introduce methods and tools for developing application layers that include both front-end and back-end of a web-based system. This course will cover acquiring, installing and running database servers, web servers, modules, and web applications. This course will also cover methods, skills, and processes for developing and maintaining application layers that allow end-users to interact with underlying databases through dynamic web interfaces.

Learning Outcomes

After successfully completing this course you will be able to:
- Identify imperfections, biases, and other problems in data sets;
- Clean up, standardize, and normalize data to prepare for data analysis;
- Extract data from a variety of data types and formats;
- Collect large data sets through scalable, automated means, such as spiders and scrapers;
- Transform data among a variety of formats and standards;
- Explain ethical and equity issues with the collection and use of data.

Donal Heidenb;ad
dheidenb@umd.edu
(pronouns: he/him/his)

Class Meets
Mondays, Wednesdays, & Fridays
10:00am – 10:50am
SYM #0200

Office Hours
TBD
and by appointment

Prerequisites
INST 326 or CMSC 131; INST 327

Course Communication
Announcements relating to this course will be made in the courses ELMS page.
Helpful guidance on writing professional emails (ter.ps/email).

Required Resources

Course website: elms.umd.edu

Textbook: None – Readings will be assigned.
Optional: Python for Everybody (free online) - https://www.py4e.com/book
(Optional Print Version of Above: $10)
Python for Everybody
Paperback: 242 pages
Publisher: CreateSpace Independent Publishing Platform (April 9, 2016)
Language: English
ISBN-10: 1530051126
ISBN-13: 978-1530051120

Campus Policies

It is our shared responsibility to know and abide by the University of Maryland's policies that relate to all courses, which include topics like:

- Academic integrity
- Student and instructor conduct
- Accessibility and accommodations
- Attendance and excused absences
- Grades and appeals
- Copyright and intellectual property

Please visit www.ugst.umd.edu/courserelatedpolicies.html for the Office of Undergraduate Studies' full list of campus-wide policies and follow up with me if you have questions.

Activities, Learning Assessments, & Expectations for Students

Before class you are expected to be prepared by:
- Reading the assigned texts or watching assigned videos
- Performing other activities, as assigned.

During class you will be assigned a variety of activities including, but not limited to:
- Completing "worksheets"comprised of programming exercises
- Participating in discussions
- Writing short reflections
- Performing other activities, as assigned.

In-class activities are graded and there will be a 12 graded activities. The lowest 2 grades will be dropped.

There will be 4 programming assignments. These are to be completed individually.

There will be a mid-term and a final exam. They will be take-home programming exams. They are to be completed individually.

Deadlines are deadlines, but I will accept **late submissions** with penalty for all assignments **EXCEPT** the mid-term and final exams. The penalty for late submission is 1/3 letter grade deduction per 24-hour period (so after 48 hours an A+ effort will results in an A grade; after 72 hours that A+ effort will result in an A- grade). All assignments must be turned in by December 10th in order to receive credit.

Collaboration is working together. Collaboration is not copying and copying is cheating. You may collaborate on in-class Exercises – unless otherwise instructed. You may not collaborate on the Assignments or the Exams. Not collaborating with your group on the team project will have poor results.

Course-Specific Policies

No phones or tablet devices are permitted during our class meetings. If it were feasible, I would ban laptops. Research shows that they present an irresistible distraction and detract from the cooperative learning environment and interfere with learning and active participation. For that reason, the use of phones and tablets will not be permitted during class meetings (except when required for ADS accommodations).

I expect you to make the responsible and respectful decision to refrain from using your cellphone in class. If you have critical communication to attend to, please excuse yourself and return when you are ready. For more information about the science behind the policy watch: http://youtu.be/WwPaw3Fx5Hk

Computers are required for class. Class sessions will involve hands-on activities

which will involve using your computer. The availability of outlets is limited, so you will need to bring your laptop fully charged to each session.

Your attendance in class is expected. Class sessions will involve hands-on activities. You are expected to complete them in class and the activities' files are to be turned in at the end of each session, so that I can identify problem areas.

If you miss class, it is your responsibility to make the effort to find out what you missed and to make up any in-class work.

Tardiness is distracting. Take consideration of your classmates and be on time.

Get Some Help!

You are expected to take personal responsibility for you own learning. This includes acknowledging when your performance does not match your goals and doing something about it. Everyone can benefit from some expert guidance on time management, note taking, and exam preparation, so I encourage you to consider visiting http://ter.ps/learn and schedule an appointment with an academic coach. Sharpen your communication skills (and improve your grade) by visiting http://ter.ps/writing and schedule an appointment with the campus Writing Center. Finally, if you just need someone to talk to, visit http://www.counseling.umd.edu.

Everything is free because you have already paid for it, and **everyone needs help**… all you have to do is ask for it.

Names/Pronouns and Self Identifications

The University of Maryland recognizes the importance of a diverse student body, and we are committed to fostering equitable classroom environments. I invite you, if you wish, to tell us how you want to be referred to both in terms of your name and your pronouns (he/him, she/her, they/them, etc.). The pronouns someone indicates are not necessarily indicative of their gender identity. Visit trans.umd.edu to learn more.

Additionally, how you identify in terms of your gender, race, class, sexuality, religion, and dis/ability, among all aspects of your identity, is your choice whether to disclose (e.g., should it come up in classroom conversation about our experiences and perspectives) and should be self-identified, not presumed or imposed. I will do my best to address and refer to all students accordingly, and I ask you to do the same for all of your fellow Terps.

Grades

Grades are not given, but earned. Your grade is determined by your performance on the learning assessments in the course and is assigned individually (not curved). If earning a particular grade is important to you, please speak with me at the beginning of the semester so that I can offer some helpful suggestions for achieving your goal.

All assessment scores will be posted on the course ELMS page. If you would like to review any of your grades (including the exams), or have questions about how something was scored, please email me to schedule a time for us to meet in my office.

I am happy to discuss any of your grades with you, and if I have made a mistake I will immediately correct it. Any formal grade disputes must be submitted in writing and within one week of receiving the grade.

Class Activities	20%
• 12 in class exercises (drop 2)	
Homework	20%
• 4 programming assignments	
Group Project	20%
• Project proposal (4%)	
• Project Status Update (1%)	
• Project presentation (4%)	
• Project report (11%)	
Exams	40%
• Midterm (20%)	
• Final (20%)	

Final letter grades are assigned based on the percentage of total assessment points earned. To be fair to everyone I have to establish clear standards and apply them consistently, so please understand that being close to a cutoff is not the same this as making the cut (89.99 ≠ 90.00). It would be unethical to make exceptions for some and not others.

Final Grade Cutoffs									
+	97.00%	+	87.00%	+	77.00%	+	67.00%		
A	93.00%	B	83.00%	C	73.00%	D	63.00%	F	<60.0%
-	90.00%	-	80.00%	-	70.00%	-	60.00%		

Course Schedule

Week Number#	Week Of	Topic & Readings	Worksheet	Due
1	8/27	Introduction & Overview	• WKS01	
2	9/3	Data Sources & Identifying Problems, Issues, & Bias	• WKS02	
3	9/10	Intro to Visualization	• WKS03	Project Proposal
4	9/17	Metadata Standards & Extraction	• WKS04	
5	9/24	Intro to XML; Parsing XML; Validating & Transforming XML	• WKS05	Assignment 1
6	10/1	Databases	• WKS06	
7	10/8	Intro to JSON	• WKS07	Assignment 2
8	10/15	Thinking Through Solutions		Mid Term
9	10/22	Web Scrapings & APIs	• WKS09	
10	10/29	Regular Expressions	• WKS10	
11	11/5	More Regular Expressions	• WKS11	Assignment 3
12	11/12	Other Tools	• WKS12	
13	11/19	Thanksgiving Break		
14	11/26	Translating these Skills into the Workplace	• WKS14	Assignment 4
15	12/3	Project Presentations		Project Report
16	12/10	Review		
FW		Finals Week		Final Exam

Note: This is a tentative schedule, and subject to change as necessary – monitor the course ELMS page for current deadlines. In the unlikely event of a prolonged university closing, or an extended absence from the university, adjustments to the course schedule, deadlines, and assignments will be made based on the duration of the closing and the specific dates missed.