



Course Syllabus

# Data Sources and Manipulation

**INST 447**  
Spring 2018

## Learning Outcomes

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.

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 Heidenblad

[dheidenb@umd.edu](mailto:dheidenb@umd.edu)

(pronouns: he/him/his)

## Class Meets

0101 - Monday,  
Wednesday, & Fridays  
09:00 - 09:50  
KEY #0103

0102 - Tuesdays &  
Thursdays  
12:30pm - 1:45pm  
TYD #0102

## Office Hours

HBK #4121F  
TBD  
and by appointment

## Teaching Assistants

TBD

## Prerequisites

INST326 or CMSC131;  
and INST327.

## Course Communication

Time-sensitive information will go out via ELMS announcement. "I did not get an ELMS notification" **is not** a valid excuse.

Please contact me through email to discuss questions, absences, or accommodations.

Please consider this guidance on writing professional emails

## Required Resources

### Website:

[elms.umd.edu](http://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

### Course videos:

These pre-recorded lectures provide the core “content” of the course. They are available through ELMS.

### Software:

There will be a Virtual Computer Lab (VCL) environment set-up with all of the required software. However, you will need to install at least those software listed under “Minimum,” “Complete” lists all of the software we will use.

VCL - <https://ischoolvcl.umd.edu/>

### Minimum

- Python 3.6 - <https://www.python.org/>
- IDE or text editor of your preference (PyCharm, Atom, Notepad++, etc)

### Complete

- Anaconda - <https://www.anaconda.com/>
- OpenRefine - <http://openrefine.org/>

## **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](http://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

**This is a “flipped” class.** This means that you are expected to watch the lectures and read materials posted to Canvas *before* each class session. There will also be a short quiz on the materials due *before* each class. These quizzes are designed to help me to understand areas where you are having difficulty.

**Why a flipped class?** Research indicates that flipped courses are the most effective learning environment. Studies also show that students strongly believe that a traditional lecture-based courses are most effective despite all of the evidence demonstrating the opposite. Flipped courses have significantly lower course evaluation scores because of this strongly held belief.

**Before class** you are expected to be prepared by:

- Reading the assigned texts
- Watching the assigned videos
- Completing the quiz
- 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 graded activity every class period.

**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 20% deduction per 24-hour period (so after 48 hours is its a 40% deduction).

The **Mid-Term** and **Final Exams** will be take home.

**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**.

## 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.

The lowest 2 quizzes are dropped and 2 absences are excused without question. However, I do appreciate notice before you have to miss class for an excusable reason and those prearranged absences do not count against those "freebies."

**Tardiness is distracting.** Take consideration of your classmates and be on time. If you are going to be more than 5 minutes late, then consider taking advantage of your 2 excused absences. If you are more than 10 minutes late, then I will expect an explanation. I retain the right to count your unexcused tardiness against your "freebies."

## 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](https://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**

**Note:** Canvas does not always calculate grades properly. I saw about a +/-5% difference. If you are concerned about your grade, please see me.

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.

Late work will not be accepted for course credit so please plan to have it submitted well before the scheduled deadline. 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.

Learning Assessments	#	Points Each	Category Weight
<b>Quizzes:</b> pre-class quizzes submitted on ELMS	14	5	5%
<b>Reflections:</b> in-class reflections	14	5	5%
<b>Worksheets:</b> in-class programming exercises	28	10	15%
<b>Assignments:</b> out-of-class programming assignments	6	25	25%
<b>Exams:</b> Mid-Term & Final	2	100	25%
<b>Group Project (GP):</b> hand-on demonstrations	4	Varies	25%

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 as making the cut (89.99  $\neq$  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

#	Planned Topic	Assignment	Deadline (11:59pm)
1	Introduction & Overview		
2	Command Line Intro and Python Review		
3	Data Sources & Storage; Problems, Issues, & Bias	Assignment 1	2/11
4	Metadata Standards & Extraction		
5	Intro to XML; Parsing XML	Assignment 2	2/25
6	Querying & Transforming XML		
7	Intro to JSON	Assignment 3	3/11
8	More JSON	Mid Term	3/16
SB	Spring Break		
9	Web Scrapings & APIs	Project Proposal	4/1
10	More APIs	Assignment 4	4/8
11	Cleaning Data		
12	More Cleaning	Assignment 5	4/22
13	Even More Cleaning		
14	Advanced Topics	Assignment 6	5/7
15	Project Presentations	Project Report	
FW	Finals Week	Final Exam	5/17

**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.