

DATA 100 Course Syllabus

DATA 100 001 - Introduction to Data Science in Python (3)

Alternate Title: Introduction to Scientific Computing in Python (3)

Course Description

DATA 100 (3) Introduction to Data Science in Python (3)

Fundamentals of data science with an emphasis on computational thinking, testing, debugging, and working with data sets. Real-world applications from all disciplines including the sciences, engineering, humanities, sociology, economics, political sciences, psychology, and others. No prior computing background is required.

Prerequisite: None

Learning Outcomes

Upon successful completion of this course, students will be able to:

1. construct workflows to load, process, clean, and analyze data.
2. develop the ability to "think computationally" using programming principles.
3. appraise the quality of data and assess its limitations in answering questions.
4. identify and use different data types to solve data science problems.
5. practice the creation of loops, conditionals, and functions to analyze data.
6. understand the role of testing to write sustainable code.
7. understand the role of version control to create reproducible reports and sustainable analyses.
8. apply the skills and techniques in this course to generate reproducible reports.

Consider removing these learning outcomes:

- understand the role of data in our society and explain how to use data to answer questions.
 - Comment: Too much detail, and too much to expect in a first year course
- examine the ethics of conducting data science analyses responsibly.
 - Comment: Important but too ambitious for a first year course?
- design and develop strategies for solving data science problems.
 - Comment: Adjust this learning outcome so it is appropriate for a 1st year course.

Assessment

Item	Weight	Frequency
Learning Logs	10%	Weekly
Labs	30%	Weekly
Project	30%	Weekly
Tests	30%	Bi-weekly
Final Exam	0%	Exam Period

N.B. I would like to petition for this course not to have a final exam, as it does not make sense pedagogically for this type of course. I would prefer to distribute the Final Exam weighting to Labs, and the project as shown above.

Passing Criteria

All students must satisfy ALL conditions to pass the course:

1. Pass the Labs with an average grade of at least 50%, with no more than 4 missed labs.
2. Pass the Tests with an average grade of at least 50%.
3. Pass the Project with a grade of at least 40%.
4. Pass the Course overall with a grade of at least 50%.

If a student does not satisfy the appropriate requirements, the student will be assigned the **lower** of their earned course grade or, a maximum overall grade of 45 in the course.

Textbook

Portions of the following (open source) textbooks will be assigned as reading:

- [Python Data Science Handbook](#), by Jake VanderPlas
- [Python for Data Analysis](#), by Wes McKinney

Eventually, an open textbook will be developed using open resources.

Schedule

Wk	Starting	Topics	Project	Lab	Learning Logs	Tests
1	Week 1	Introduction to Data Science and your Computer			LL 1	
2	Week 2	Navigating the Command Line		L1	LL 2	Test 1
3	Week 3	Introduction to Jupyter Notebook	PM1	L2	LL 3	
4	Week 4	Version Control with Git		L3	LL 4	Test 2
5	Week 5	Introduction to computation in Python with numpy	PM2	L4	LL 5	
6	Week 6	Loading and working with data		L5	LL 6	Test 3
7	Week 7	Data Types: Lists, dictionaries, sets, tuples	PM3	L6		
8	Week 8	Control flow: conditionals		L7	LL 7	Test 4
9	Week 9	Loops and Functions	PM4	L8	LL 8	
10	Week 10	Data analysis with scipy		L9	LL 9	Test 5
11	Week 11	Data visualization	PM5	L10	LL 10	
12	Week 12	Testing and the git flow			LL 11	
13	Week 13	Reproducible data analyses	PM6		LL 12	