DSCI 101 Foundations of Data Science I, 4 Credits

University of Oregon, Data Science Program

Course Description

This course utilizes a quantitative approach to explore fundamental concepts in data science.

Students will develop key skills in programming and statistical inference as they interact with real-world data sets across a variety of domains. Ethical ramifications of data collection, data-driven decision making, and privacy will be explored.

This course is designed to be accessible to students without prior experience in computer

programming or statistics.

Prerequisites: None - This course is designed for entry-level students from any major. It is designed specifically for students who have not previously taken statistics or computer science courses.

Major Requirement Required for all data science students

Format Two 1.5 hours classroom lectures / demos / discussions, one 1.0 hour lab.

Lecture: recorded lectures will be recorded and posted within a few days of class. Demos and slides will be posted on canvas before the lecture.

Lab: Quizzes will be given in labs, and you must take the quiz at your assigned lab time. Please attend your assigned lab.

Course Materials

Textbook: Computational and Inferential Thinking: The Foundations of Data Science, a free online textbook that includes Jupyter notebooks and public data sets for its examples.

Software: Python 3 with data science modules from an Anaconda installation such as Numpy and Matplotlib, as well as the datascience and otter-grader modules. *With the Chrome or Firefox browser, you can use the UO Talapas super-computer and avoid the installing any software on your computer; it's also the only way you could use a Chromebook or similar machine with the course material.*

Coaching Model: Instructors will play a role similar to a coach for athletes. We'll give you some information and feedback, but your active, engaged participation will lead to genuine learning. As with a sport, practice, practice, practice and your understanding will improve.

Learning Outcomes

Upon successful completion of this course each student should be able to:

1. Discover**,** organize, analyze, and visualize data using the Python programming language
2. Identify potential errors in data collection and analysis and discuss their consequences
3. Apply concepts of statistical inference including sampling and simulation to create and test models
4. Develop and test null and alternative hypotheses to answer domain-specific questions

(questions in domains such as science, business, economics, political science, journalism, athletics …)

5. Outline ethical ramifications of data collection, data-driven decision making, data sharing, and privacy

Assessment

- **Homework** will be assigned weekly and unless otherwise noted will be due Friday at midnight the week following after the HW is posted (the first HW is due later, see schedule below). Assignments help you apply the material discussed in class. Homework lets you evaluate your current level of understanding of course topics. Show all work necessary to reach a solution. Partial credit will be awarded for the proper procedure regardless of the final answer.

- **Labs sections** will meet on a weekly basis. Lab assignments are due the Friday after the lab took place, except otherwise noted (the first lab assignment is due later, see schedule below). If you need extra help, you can come to additional lab sections. If you pass all of the graders and correctly upload your assignment on time, you will receive 100%. No late submissions are accepted.

- **Project:** Additionally, one course project will be completed during the term. No late submissions are accepted.

The final course grade will be determined using these weighted components: subject to revision

- Homework 14% (7 x 2% each)
- Lab Assignments 28% (7 x 4% each)
- Course Project1 14% (1 x 14%)
- Quizzes 24 % (4 x 6%)
- Final Quiz 20% (1 x 20%)

The course will be graded on the following scale

A+ = 96.67-100%	A = 93.34-96.66%	A- = 90.0-93.33%
B+ = 86.67-89.99%	B = 83.34-86.66%	B- = 80.0-83.33%
C+ = 76.67-79.99%	C = 73.34-76.66%	C- = 70.0-73.33%
D+ = 66.67-69.99%	D = 63.34-66.66%	D- = 60.0-63.33

F = 0-59.99%

Student Expectations

Attendance Attend both classes and labs regularly. Actively engage in classroom discussions.

Readings Come to class having read the corresponding reading material. This will not only facilitate better assimilation of the lecture material but encourage more thorough and thoughtful discussions during the class.

LP= Prof Ponisio, DW= Prof Wilkins

Course Outline *Subject to adjustments*

Week	Date	Topic	Reading	Week's Lab / Quiz (lab assignments due Friday of the week the lab took place at midnight)	Week's Hm (HW due Monday week AFTER it is assigned at midnight)

1	Jan 5	Introduction (LP)	1.1, 1.2, 1.3	Lab1, due Friday Jan 15 (exception to usual due dates)	Hm 1 due Jan 18 (exception to usual due dates)
	Jan 7	Cause and Effect vs Association (LP)	2, 2.1 – 2.5		
2	Jan 12	Data Types, Conversions, arrays (LP)	3 – 3.3, 4 – 4.3	Lab 2 (due Jan 15)	Hm 2 (due Jan 19) (exception to usual due dates)
	Jan 14	Building Tables (LP)	3.4, 5		
3	Jan 19	Building Tables cont (DW)	6.1, 6.2	Lab 3 (due Jan 22), **Quiz 1**	Hm 3 (due Jan 29)
	Jan 21	Histograms, Functions (DW)	6.3, 6.4		
4	Jan 26	Group (DW)	7, 7.1	Lab 4 (due Jan 29)	Hm 4 (due Feb 5)
	Jan 28	Join, Draw Maps (DW)	7.2, 7.3		
5	Feb 2	Conditionals and iteration (LP)	8, 8.1	Lab 5 (due Feb 5), **Quiz 2**	Hm 5 (due Feb 12)
	Feb 4	Chance, probability, sampling (LP)	8.2, 8.3		
6	Feb 9	Models, Distributions (LP)	8.4	Lab 6 (due Feb 12)	Hm 6 (due Feb 19)
	Feb 11	Decisions, Uncertainty (LP)	8.5		
7	Feb 16	A/B Testing (DW)	12.1	Lab 7 (due Feb 26**), Quiz 3**	Hm 7 (due 26)
	Feb 18	Causality (DW) Benford's Law, Ethics	12.3		
8	Feb 23	Wikipedia search for "Benford's Law" (DW)	*	finish Lab 7 (due Feb 26**)** Start project in lab	
	Feb 25	"Ethics and Data Science" by Loukides et al Free on Amazon (40 pages total, readable) (DW)	*		
9		"Ethics and Data Science" by Loukides et al Free on Amazon (40 pages total, readable) (DW)	*	Work on project in lab, **Quiz 4**	

		Bee communities in agriculture (special topics, LP)	*	
10		Review / Special Topics / Student Project demo		Review in lab
		Review / Special Topics / Student Project demo		Project Due
Final	Mar 18	Final Quiz (Thur 18 Mar 12:30 – 14:30)		**Final Quiz 5**

Collaboration

Collaboration is an excellent way to **learn** and **share insights** (but not share the actual code for the homework or assignments due), so share ideas, concepts, insights. It's widely used in the "real world" both in business and research. But note that **copying and pasting code does not contribute to learning**.

We encourage you to ask questions and discuss questions (but not during quizzes).

Do: collaborate and identify the people you have worked together with when you submit assignments.
Don't: copy and paste code. Both those who copy and those who provide the copy will face consequences (see Academic Misconduct below).
Don't: post solutions on slack.