

DSCI 521: Data Analysis and Interpretation

Course Syllabus

Credits: 3 hour lecture [3 credits]

Winter 2019

Short Title: Analysis and Interpretation

General Information

Course Instructor: Jake Ryland Williams

Instructor Contact Information: Jake.Williams@drexel.edu

Office Hours, Location: Tuesdays, 5–6 PM; Rush Building, Room 417

Student Learning Information

Course Description

Introduces methods for data analysis and their quantitative foundations in application to pre-processed data. Covers reproducibility and interpretation for project life cycle activities, including data exploration, hypothesis generation and testing, pattern recognition, and task automation. Provides experience with analysis methods for data science from a variety of quantitative disciplines. Concludes with an open-ended term project focused on the application of data exploration and analysis methods with interpretation via statistical, algorithmic, and mathematical reasoning.

College/Department: College of Computing & Informatics

Repeat Status: Not repeatable for credit

Restrictions: None

Prerequisites: None

Course Purpose Within a Program of Study

This course provides an application-oriented introduction to quantitative reasoning for data science. Hands-on experience is provided with a number analytic techniques and evaluation methods in the context of their quantitative foundations. Via implementation, an understanding for method function, applicability, setup, and interpretation is provided under mathematical frameworks.

This course is a core course in the Data Science Master's program.

Statement of Expected Learning

The course objectives are to:

- observe and explore a variety of quantitative methods for data analysis;
- understand methods' evaluation techniques to interpret their output;
- implement and evaluating methods to gain technical experience with data; and
- reproducibly execute an analytic project and represent/communicate its results faithfully.

As learning outcomes, students completing this course should be able to take data and a task and identify methods for analysis, reproducibly implement them by using widely-available tools, reason quantitatively to interpret their results, and conduct evaluations for the faithful communication of output.

Course Materials

Required and Recommended Texts, Readings, and Resources

Note: all text readings are supplemental to the course lecture notes and will be assigned on a weekly basis. Python Data Science Handbook is available free of charge on Github:

- <https://jakevdp.github.io/PythonDataScienceHandbook/>

and The Python Data Science Handbook and Data Science from Scratch are available to Drexel students through the [University Libraries](#). Specific text information is as follows:

- Data Science from Scratch (DSFS). ISBN: 978-1491901427, O'Riley, April 2015
- Python Data Science Handbook (PDSH). ISBN: 978-1491912058, O'Riley, November 2016
- The Data Science Handbook (TDSH). ISBN:9781119092940, John Wiley & Sons, 2017

Required and Supplemental Materials and Technologies

Note: instructions and discussion of the following materials and technologies are provided in Chapter 0 of the course lecture notes. Students are expected to have the following by the start of the first week:

- A Github account: <https://www.github.com>
- A command line environment with Python (version 3) installed
- The Jupyter notebooks interactive development environment

Lecture Notes

The primary course materials consist of a collection of interactive Jupyter notebooks, which may be found on the course blackboard website and on the following private Github repository:

- <http://github.com/jakerylandwilliams/DSCI521/>

For access to the Github repository, all students are required to sign up for an account and post user names on the course discussion board. An invite to the Github repository will follow.

The course lecture notes are broken down into the following topics, which roughly correspond to a week of content each. More information on course scheduling can be found in that section, below.

- Chapter 0: System configuration, programming, and numeracy
- Chapter 1: Processing numeric data
- Chapter 2: Feature engineering and language processing
- Chapter 3: Exploratory data analysis and visualization
- Chapter 4: Networks, metrics, and applications

Chapter 5: Data transformation and analysis of functions
Chapter 6: Counting and probability for quantitative frameworks
Chapter 7: Approaching regression and classification problems
Chapter 8: Model pipelines and comparisons

Assignments, Assessments, and Evaluations

Graded Assignments and Learning Activities

Homework: Structured, individual assignments will be distributed according to four topic areas:

1. Text analysis and feature engineering
2. Network and exploratory data analysis, including visualization
3. Probabilistic modeling and prediction
4. Machine learning and regression

These assignments will be composed in a modular fashion, with each module/problem worth about 35–45 points apiece. The total assignment value requirement for the term is 400 *target* points and there will be about 550 *possible* points available across all modules. Each module can be completed and submitted separately. This means there will be roughly 12–16 modules, total.

Project: One open-ended group assignment will have two phases:

1. Data Set Identification, Motivational Exploration, and Proposed Methods Implementation
2. Methods Implementation, Evaluation, and Interpretation, with Documentation and Dissemination

Grading Matrix

Students will not receive letter grades for individual assignments. Grades are calculated as:

Project:	30% (10% Proposal, 20% Implementation)
Homework:	70% (out of 400 target points, cumulative from 550 possible)
<hr/>	
Total:	100%

Grade Scale

The following scale will be used to convert points to letter grades:

<i>Points</i>	<i>Grade</i>	<i>Points</i>	<i>Grade</i>	<i>Points</i>	<i>Grade</i>
97-100	A+	82-86.99	B	70-71.99	C-
92-96.99	A	80-81.99	B-	67-69.99	D+
90-91.99	A-	77-79.99	C+	60-66.99	D

87-89.99	B+	72-76.99	C	0-59.99	F
----------	----	----------	---	---------	---

Note that the instructor may revise this conversion if/when necessary.

Course Schedule

The course's schedule follows the lecture notes at roughly one week per chapter with the expectation that students will configure systems and review or work through the numeracy and processing fundamentals in **Chapter 0**. Some chapters may extend over multiple weeks, depending on the section's pace. **Week 10** and the regularly scheduled **final exam** period (the same day of week and time of the regularly scheduled class meetings, but during exam week) are reserved for final project presentations. Please observe the following (tentative) schedule; weekly activities may change depending on the pace of class.

- Week 0:
 - DSFS: Chapter 1
 - TDSH: Chapters 1–2
 - Homework: configure system
- Week 1:
 - DSFS: Chapters 2 & 4
 - Project: team formation; begin Phase 1
 - Homework: begin Assignment Group 1
- Week 2:
 - DSFS: Chapters 10 (pgs. 121–133) & 20 (pgs. 239–244)
 - TDSH: Chapter 16
 - Project: continue Phase 1
 - Homework: continue Assignment Group 1
- Week 3:
 - DSFS: Chapters 3 & 5
 - TDSH: Chapters 5.1–5.12
 - PDSH: Chapters 4.01–4.02, 4.04–4.06, 4.08–4.09
 - Project: continue Phase 1
 - Homework: begin Assn. Group 2 (Thursday); complete Assn. Group 1 (Friday)
- Week 4:
 - DSFS: Chapter 21 (pgs. 255–256, 264–266)
 - Project: begin Phase 2 (Thursday); complete Phase 1 (Friday)
 - Homework: continue Assignment Group 2
- Week 5:
 - DSFS: Chapter 8
 - Project: continue Phase 2
 - Homework: begin Assn. Group 3 (Thursday); complete Assn. Group 2 (Friday)
- Week 6:
 - DSFS: Chapter 6
 - TDSH: Chapters 6, 18.0–18.3, 18.5–18.7, 19.0–19.2
 - Project: continue Phase 2
 - Homework: continue Assignment Group 3

- Week 7:
 - DSFS: Chapters 14, 15 (pgs. 171–183), & 16 (pgs. 189–192, 196–200)
 - TDSH: Chapters 8.1–8.4, 8.5.5, 10.6, 10.7 (through 10.7.2), & 11.0–11.6
 - PDSH: 5.06 & 5.09
 - Project: continue Phase 2
 - Homework: begin Assn. Group 4 (Thursday); complete Assn. Group 3 (Friday)
- Week 8:
 - DSFS: Chapters 11, 13, 17 (pgs. 201–203 & 211) & 19
 - TDSH: Chapters 8.5.1–8.5.4, 8.5.7
 - PDSH: 5.01–5.04, 5.05, 5.07–5.08 & 5.11
 - Project: continue Phase 2
 - Homework: continue Assignment Group 4
- Week 9:
 - DSFS: Chapter 18
 - TDSH: Chapters 8.5.8 & 24.1–24.5
 - Project: continue Phase 2
 - Homework: complete Assignment Group 4
- Week 10:
 - DSFS: Chapters 12, & 22 (as needed for projects)
 - Project: complete Phase 2; presentations
- Week 11: (final exam week)
 - Project: complete Phase 2; presentations

Academic Policies

This course follows university, college, and department policies, including but not limited to:

- Academic Honesty: http://www.drexel.edu/provost/policies/academic_dishonesty.asp
- Student Life Honesty Policy from Judicial Affairs: <http://www.drexel.edu/provost/policies/academic-integrity>
- Students with Disability
Statement: <http://drexel.edu/oed/disabilityResources/faculty/SyllabusStatement/>
- Course Drop Policy: http://www.drexel.edu/provost/policies/course_drop.asp
- Department Academic Integrity
Policy: <http://drexel.edu/cs/academics/undergrad/policies/academic-integrity/>
- Drexel Student Learning Priorities: <http://drexel.edu/provost/assessment/outcomes/dslp/>
- Office of Disability Resources: http://www.drexel.edu/ods/student_reg.html

The instructor(s) may, at his/her/their discretion, change any part of the course before or during the term, including assignments, grade breakdowns, due dates, and schedule. Such changes will be communicated to students via the course web site. This web site should be checked regularly and frequently for such changes and announcements.

Students [requesting accommodations](#) due to a disability at Drexel University need to request a current Accommodations Verification Letter (AVL) in the [ClockWork database](#) before accommodations can be made. These requests are received by Disability Resources (DR), who then issues the AVL to the appropriate contacts. For additional information, visit the DR website at drexel.edu/oed/disabilityResources/overview/, or contact DR for more information by phone at 215.895.1401, or by email at disability@drexel.edu.

