

Science

[Home](#) [Syllabus](#) [Schedule](#) [Project](#) [Fame](#) [Resources](#)

Syllabus

Welcome to introduction to data science (COMP 5360 / MATH 4100)! The major goals of this course are to learn how to use tools for acquiring, cleaning, analyzing, exploring, and visualizing data; making data-driven inferences and decisions; and effectively communicating results. These will be accomplished through course activities on the following data science topics:

- Introduction to data analysis tools in Python
- Descriptive statistics
- Data structures with Pandas
- Introductory hypothesis testing and statistical inference
- Web scraping and data acquisition via APIs
- Linear regression
- Classification methods, including logistic regression, k-nearest neighbors, decision trees, support vector machines, and neural networks
- Data visualization
- Clustering methods
- Dimensionality reduction, including principle component analysis
- Network analysis
- Rating, ranking, and elections
- Cleaning and reformatting messy datasets using regular expressions or dedicated tools such as open refine
- Natural language processing
- Ethics of big data

A major component of this course will be learning how to use python-based programming tools to apply these methods to real-life datasets.

Learning Outcomes

At the end of the course, a student should be able to:

- Acquire data through web-scraping and data APIs
- Clean and reshape messy datasets

- Use exploratory tools such as clustering and visualization tools to analyze data
- Perform linear regression analysis
- Use methods such as logistic regression, nearest neighbors, decision trees, support vector machines, and neural networks to build a classifier
- Apply dimensionality reduction tools such as principle component analysis
- Perform basic analysis of network data
- Evaluate outcomes and make decisions based on data
- Effectively communicate results

Course Logistics

Prerequisites

Completed at least one of the following:

- MATH 1170 - Calculus for Biologists I (4)
- MATH 1210 - Calculus I (4)
- MATH 1250 - Calculus for AP Students I (4)
- MATH 1310 - Engineering Calculus I (4)
- MATH 1311 - Accelerated Engineering Calculus I (4)

Recommended Prerequisites/Corequisites: Some programming experience with Python or a similar language, as demonstrated by the ability to write short programs incorporating variables, lists and strings, loop structures, and data file input and output. More advanced mathematics, such as linear algebra or introductory statistics, is also recommended.

If in doubt, ask one of the instructors. **You should also own a notebook computer that you can bring to class.**

Textbooks

There is no required textbook for the class. However, students may find it useful to consult the following textbooks for reference.

Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython

Read for free on Campus

2nd edition

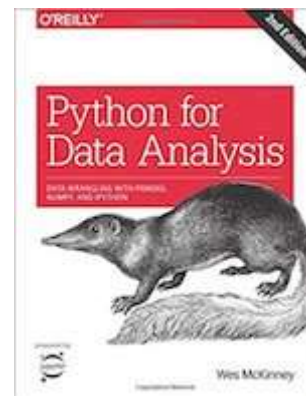
Wes McKinney

O'Reilly Media (2017)

Get complete instructions for manipulating, processing, cleaning, and crunching datasets in Python. Updated for Python 3.6, the second edition of this hands-on guide is packed with practical case studies that show you

how to solve a broad set of data analysis problems effectively. You'll learn the latest versions of pandas, NumPy, IPython, and Jupyter in the process.

Written by Wes McKinney, the creator of the Python pandas project, this book is a practical, modern introduction to data science tools in Python. It's ideal for analysts new to Python and for Python programmers new to data science and scientific computing. Data files and related material are available on GitHub.



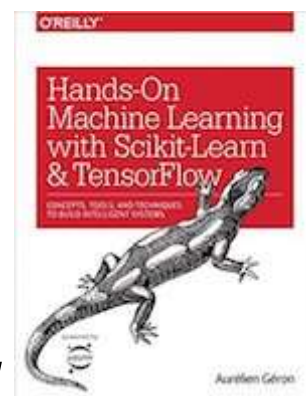
Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems

[Read for free on Campus](#)

Aurélien Géron

O'Reilly Media (2017)

Through a series of recent breakthroughs, deep learning has boosted the entire field of machine learning. Now, even programmers who know close to nothing about this technology can use simple, efficient tools to implement programs capable of learning from data. This practical book shows you how.



By using concrete examples, minimal theory, and two production-ready Python frameworks—scikit-learn and TensorFlow—author Aurélien Géron helps you gain an intuitive understanding of the concepts and tools for building intelligent systems. You'll learn a range of techniques, starting with simple linear regression and progressing to deep neural networks. With exercises in each chapter to help you apply what you've learned, all you need is programming experience to get started.

Other Relevant Books

Data Science from Scratch: First Principles with Python, Joel Grus

O'Reilly Media (2015)

[Code](#)

Doing Data Science: Straight Talk from the Frontline

Cathy O'Neil, Rachel Schutt

O'Reilly Media (2013)

Learning the Pandas Library: Python Tools for Data Munging, Analysis, and Visualization

Matt Harrison

CreateSpace Independent Publishing Platform (2016)

Mining of Massive Datasets

Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman

2nd ed., Cambridge University Press (2014)

[Download from book website](#)

Data Mining: Concepts and Techniques

Jiawei Han, Micheline Kamber, and Jian Pei
3rd ed., Morgan Kaufmann (2011)

An Introduction to Statistical Learning

Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani
Springer Texts in Statistics (2015)
[Read for free](#)

Deep Learning with Python

Francois Chollet
Manning Publications Co. (2018)

Course Components

Lectures

The class meets twice a week for lectures and joint class activities. The [weekly schedule](#) is posted on the course website.

Lectures contain both theoretical knowledge and technical components to give you the skills to successfully complete the homework assignments and projects. Lectures will often consist of a short presentation and live coding, followed by time to complete some exercises with the help of the instructors. Lecture topics and times are announced on the schedule. **You are expected to bring your own computer with the necessary software installed to all labs.**

Class activities are designed to help you master the relevant materials, to work on your homework in groups, and to help you start your project.

Group Activities

Most lectures will be accompanied by a group activity, where you, for example, solve a programming problem together. To incentivize participation, you now must submit 5 discussions of group activities held during lecture. There will be 10-15 of these activities, you can pick which ones you want to write up and submit. This constitutes 5% of your grade, so it's important to do.

Reading Assignments

The course schedule includes required weekly readings – you are free to study ahead, but the schedule ensures that you are prepared for the activities in class and the homework. The goal of the reading assignments is to familiarize yourself with new terminology and definitions, to learn new statistics and programming skills, and to determine which part of the subject needs more attention. The homework

assignments will contain questions about the mandatory readings. When answering these, please be brief and to the point!

Homework

Neither mathematics or computer science are spectator sports; mastery of either subject requires a significant amount of practice! Homework assignments provide an opportunity to practice your programming skills, think about analytical concepts in a new way, and to test your understanding of the material. The homework assignments are also designed to prepare you for your course project. You should view the homework as an opportunity to learn, and not to “earn points”. The homework will be graded holistically to reflect this objective.

Homework assignments are due on Fridays, 11:59 pm MST/MDT, unless stated otherwise. **For due dates see [weekly schedule](#) and submit via the [canvas](#).**

The assignments are published on [GitHub](#).

Homework submissions will be handled through [canvas](#). Submit a zip file that includes all files needed to execute the homework.

Homework Rules & Hints

A couple of **important rules** to make our lives easier:

- See our [collaboration policy](#) to see what is permissible and what is academic misconduct and to learn how to quote your sources.
- We recommend you use version control while you are working on your homework using a private repository. GitHub offers [free private repositories for students](#), [BitBucket](#) also provides free private repositories. Every time you finish a chunk of work, or when you are done for the day, push your changes to a repository. This will avoid data loss, and you always will be able to recover what you already pushed. **Make sure that your work is NOT PUBLICLY ACCESSIBLE.**
- We will grade your work based on the Python, Jupyter, and library version used in lectures. Make sure that your code is compatible with these versions.

Course Project

At the core of this course is a course project. The goal of the project is to analyze a topic of your choosing and present your findings. You will acquire and clean the data; use tools from the class to explore, describe, and analyze the data; and evaluate the results to make predictions. The path to a good project will involve mistakes and wrong turns. It is important to recognize that these missteps are invaluable on the path to a great project, but will require a significant amount of time. **It is therefore imperative that you begin your project early!** The project has an intermediate milestone that will allow you to get feedback and to iterate. In your project, you will work closely with classmates in 2-3 person teams. You can find more information on the [project page](#).

Course Policies

Assessment Procedure

Your final grade will be determined by your performance on the various aspects of the class:

- Homework: 45%, assessed on your individual submissions. **We will drop the lowest homework grade.**
- Project: 50%, assessed on meeting the project criteria and your peer assessment. The 50% is split between the two milestones and the proposal. 5% are assigned to the proposal, 15% are assigned to your first milestone, 30% to your final submission.
- Group Activities: 5%, 1% for each activity.

We will evaluate your work holistically beyond mechanical correctness and focus on the overall quality of the work.

In addition, the instructors will select the top projects through a review and voting process. These projects will be featured on the website and the project teammates will receive bonus points. In addition, the teammates will be awarded chocolate!

The scale for assigning letter grades is as follows (based on a total of 100 points). This scale might be curved based on overall class performance while ensuring fairness to all.

A 100-93 A- 93-90 B+ 90-87 B 87-83 B- 83-80 C+ 80-77 C 77-73 C- 73-70 D+ 70-67 D 67-63 D- 63-60 E 60-0

Project Group Peer Assessment

In the professional world, three important features affect your productivity and success: your own effort, the effort of people you depend on, and the way you work together. For this reason we have chosen a team-based approach that values all three of those features. After each team-based project you will provide an assessment of the contributions of the members of your team, including yourself. Your scores on the projects are adjusted up or down depending on the following factors:

- Your teammates' view of your contributions to the team
- The accuracy of your own assessment of your contributions
- The accuracy of your assessment of each of your teammates' contributions

Your teammates' assessment of your contributions and the accuracy of your self-assessment will be considered as part of your overall course evaluation.

Collaboration, Cheating and Plagiarism Policy

You are welcome to discuss the course's ideas, material, and homework with others in order to better understand it, but **the work you turn in must be your own** (or for the project, yours and your teammates'). For example, you must write your own code, design your own visualizations, and critically evaluate the results in your own words. You may not submit the same or similar work to this course that you have submitted or will submit to another. Nor may you provide or make available solutions to homeworks to individuals who take or may take this course in the future.

As a simple rule: **every time you copy code into your project or homework, you are likely to violate this policy**. The exception is code provided in lectures: you are free to copy that as a starting point.

In **homeworks** you must not use libraries or code provided on the internet except when explicitly permitted in the instructions.

In your **project**, you may use **limited** parts of code found online, provided its license allows you to re-use it. You are free to use general purpose frameworks or libraries (e.g., Node.js, Bootstrap, JQuery, etc.) You may not use plotting libraries such as plot.ly.

If you plan to use source code that was not written by you (except general purpose frameworks or libraries), please obtain prior approval from the teaching staff in writing. You must acknowledge any source code that was not written by you using a proper citation (author, year, title, time accessed, URL) directly in your source code (comment or header) and provide a link to the source. You can also acknowledge sources in a README.txt file if you used whole classes or libraries. You also must include these references clearly visible on your project website.

For a general rule of thumb, at least 80% of the code should be written by you and your teammates (not counting general purpose frameworks or libraries). When in doubt, please consult the teaching staff.

We will use both manual and automatic methods to check your submissions for plagiarism and will also check against online sources and submissions from previous years. For details on the policy, please refer to the [School of Computing Cheating Policy](#).

For this particular class, each plagiarism instance will lead to an academic misconduct sanction and a zero grade on the assignment; two sanctions will lead to a failing grade in this course, two such infractions will lead to a ban on all CS programs.

Please read carefully the School of Computing (SoC) policies and guidelines. For undergraduate students, please read the undergraduate handbook <https://handbook.cs.utah.edu/>. For graduate students, please read the graduate handbook: <http://www.cs.utah.edu/graduate/resources/>

Make sure to familiarize yourself with the above policies!

College of Engineering Guidelines: Academic Calendar, Policies

Please review the college of engineering guidelines, which you can find [here](#). These guidelines contain important dates regarding adding, dropping, and withdrawing from classes as well as the College Policy regarding repeating courses.

Missed Activities and Assignment Deadlines

All submissions related to projects must be turned in on time. Homeworks are subject to the late day policy stated below. We understand, however, that certain factors may occasionally interfere with your ability to participate or to hand in work on time. If that factor is an extenuating circumstance such as a medical condition, we will ask you to provide documentation directly issued by a doctor's office or the University, and we will try to work out an agreeable solution with you (and your team).

Deadline extensions (of both homework assignments and project submissions) may be arranged due to a documented medical emergency. A documented medical emergency is defined as a verifiable document from a doctor's office. Getting a COVID-19 test by itself is not a medical emergency. Once approved, the student will get a 5-day extension (including weekends) for an assignment. Such a document should be (preferably) provided 1 day before the deadline. Documented emergencies other than medical emergencies (e.g. loss of power) will be dealt with on a case-by-case basis. Please email the instructor with documented proof of the emergency (e.g. from the power company). Such a document does not automatically guarantee approval.

Late Policy

You can turn in your homework assignments up to two days late, however, for each day that an assignment is turned in late we will deduct 10% off the total possible points. That is, one-day late is 10% off, two-days is 20% off. So, if your assignment is two days late, the max number of points (out of 10) that you can receive is 8. By permission of the instructor in extenuating circumstances, you may use more than two late days, however, the 10% rule per day will still apply.

It is important to note that the **late policy does not apply to submissions related to projects**, which must be turned in on time.

Homework Drop Policy

We will automatically remove the homework with the lowest score from calculating your grade.

Regrading Policy

It is very important to us that all assignments are properly graded. If you believe there is an error in your assignment grading, please submit a regrading request that includes an explanation via **Piazza within 7 days** of receiving the grade. No regrade requests will be accepted orally, and no regrade requests will be accepted more than 7 days after you receive the grade for the assignment.

Online etiquette

Students and the teaching staff (instructors and TAs) are expected to create a respectful online learning environment. All online interactions (including but not limited to emails, Piazza, Canvas, Zoom) are expected

to follow common rules for good online etiquette:

- Be respectful and be professional.
- Be aware of strong language, all caps, and exclamation points.
- Be careful with humor and sarcasm.
- Do not post or share (even privately) inappropriate material.
- Disrespectful or inappropriate online communications will be deleted from online platforms (e.g., Piazza and Canvas). Severe cases may be referred to the appropriate committee or office within the University for possible disciplinary actions.

Additional Accommodations

If students need to seek an ADA accommodation due to a disability, please contact the Center for Disability and Access (CDA). CDA will work with us to determine what, if any, ADA accommodations are reasonable and appropriate. While this class is taught in person, class materials will be made available online to accommodate the instructional needs of students who are quarantined or self-isolated due to COVID-19, or who have ADA accommodations.

Recommendations during the COVID-19 pandemic

This class follows the latest recommendation regarding the COVID-19 pandemic:
<https://coronavirus.utah.edu/>.

Student Wellness

Personal concerns such as stress, anxiety, relationship difficulties, depression, cross-cultural differences, etc., can interfere with a student's ability to succeed and thrive at the University of Utah. For helpful resources contact the Center for Student Wellness at www.wellness.utah.edu or 801-581-7776.

Support Student Wellbeing During COVID-19

A list of COVID-19 specific course accommodations:

- This class does not have a "credit or no credit" option. Please contact the instructor if you have any questions.
- The class lectures are in-person. However, lectures will be streamed via Zoom and Youtube. And course materials will be made available online.

Rates of burnout, anxiety, depression, isolation, and loneliness have noticeably increased during the pandemic. If you need help, reach out for campus mental health resources, including counseling, training, and other support. See: <https://studentaffairs.utah.edu/mental-health-resources/index.php>

Consider participating in a Mental Health First Aid or other wellness-themed training provided by our Center for Student Wellness and sharing these opportunities with your peers, teaching assistants and department colleagues: <https://studentaffairs.utah.edu/mental-health-first-aid.php> <https://wellness.utah.edu/workshops-training/>

The Americans with Disabilities Act

The University of Utah seeks to provide equal access to its programs, services, and activities for people with disabilities. If you will need accommodations in this class, reasonable prior notice needs to be given to the Center for Disability Services, 162 Olpin Union Building, (801) 581-5020. CDS will work with you and the instructor to make arrangements for accommodations. All written information in this course can be made available in an alternative format with prior notification to the Center for Disability Services.

Respect for Diversity

It is our intent that students from all diverse backgrounds and perspectives be well-served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that the students bring to this class be viewed as a resource, strength and benefit. It is our intent to present materials and activities that are respectful of diversity: gender identity, sexuality, disability, age, socioeconomic status, ethnicity, race, nationality, religion, and culture. Your suggestions are encouraged and appreciated. Please let us know ways to improve the effectiveness of the course for you personally, or for other students or student groups.

Student Name and Personal Pronoun

Class rosters are provided to the instructor with the student's legal name as well as "Preferred first name" (if previously entered by you in the Student Profile section of your CIS account). While CIS refers to this as merely a preference, I will honor you by referring to you with the name and pronoun that feels best for you in class, on papers, exams, group projects, etc. Please advise me of any name or pronoun changes (and please update CIS) so I can help create a learning environment in which you, your name, and your pronoun will be respected.

Addressing Sexual Misconduct

Title IX makes it clear that violence and harassment based on sex and gender (which includes sexual orientation and gender identity/expression) is a civil rights offense subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories such as race, national origin, color, religion, age, status as a person with a disability, veteran's status or genetic information. If you or someone you know has been harassed or assaulted, you are encouraged to report it to the Title IX Coordinator in the Office of Equal Opportunity and Affirmative Action, 135 Park Building, 801-581-8365, or the Office of the Dean of Students, 270 Union Building, 801-581-7066. For support and confidential consultation, contact the Center for Student Wellness, 426 SSB, 801-581-7776. To report to the police, contact the Department of Public Safety, 801-585-2677(COPS).

Veterans Center

If you are a student veteran, the U of Utah has a Veterans Support Center (Links to an external site.) located in Room 161 in the Olpin Union Building. Hours: M-F 8-5 pm. Please visit their website for more information about what support they offer, a list of ongoing events, and links to outside resources. Please also let the instructors know if you need any additional support in this class for any reason.

Learners of English as an additional/second language

If you are an English language learner, please be aware of several resources on campus that will support you with your language and writing development. These resources include the Writing Center (Links to an external site.), the Writing Program (Links to an external site.), and the English Language Institute (Links to an external site.). Please let the instructor know if there is any additional support you would like to discuss for this class.

Undocumented Student Support Statement

Immigration is a complex phenomenon with broad impact—those who are directly affected by it, as well as those who are indirectly affected by their relationships with family members, friends, and loved ones. If your immigration status presents obstacles to engaging in specific activities or fulfilling specific course criteria, confidential arrangements may be requested from the Dream Center. Arrangements with the Dream Center will not jeopardize your student status, your financial aid, or any other part of your residence. The Dream Center offers a wide range of resources to support undocumented students (with and without DACA) as well as students from mixed-status families. To learn more, please contact the Dream Center at 801.213.3697 or visit <https://dream.utah.edu/>

Safety

The University of Utah values the safety of all campus community members. To report suspicious activity or to request a courtesy escort, call campus police at 801-585-COPS (801-585-2677). You will receive important emergency alerts and safety messages regarding campus safety via text message. For more information regarding safety and to view available training resources, including helpful videos, visit <https://safeu.utah.edu/>.

Credits

This class occasionally uses material developed for Harvard's **CS 109**, taught by Hanspeter Pfister, Joe Blitzstein, Rhaul Dave, and Verena Kayning. We have drawn on materials and examples found online and give credit by linking to the original source. You can find these credits mainly by direct links to the sources from the slides (e.g., hyperlinked from images). Please contact us if you find materials where the credit is missing or that you would rather have removed.

User Notice for Copyrighted Materials on Course Websites

This course website, and all original content provided as part of this course is licensed under the [creative commons cc by license](#). Other content such as text, images, graphics, audio and video clips, (collectively, the “Content”), are protected by copyright law. In some cases, the copyright is owned by third parties, and we are making the third-party content available to you under the fair use doctrine. Fair use permits only certain limited uses of the Content. You may use this Content only for your personal, noncommercial educational and scholarly use.

Note

The information provided here is meant to serve as an outline and guide for our course. Please note that the instructor may modify it with reasonable notice to you. The instructors may also modify the course schedule to accommodate the needs of our class. Any changes will be announced during lectures and/or posted on Canvas under Announcements.

Introduction to Data Science

[Canvas](#)

[Piazza](#)

 [datascience-course](#) © The University of Utah, 2017-2022