

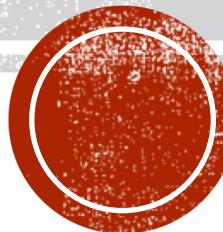
DATA SCIENCE WITH PYTHON

CS 498 (SPECIAL TOPICS IN COMPUTER SCIENCE)

Dr. Malak Abdullah

Computer Science Department

Jordan University of Science and Technology



WELCOME

To CS 498 (Special Topics in Computer Science) offered by Computer Science Department!

The course focuses on the analysis of messy, real life data to perform predictions using statistical and machine learning methods

COURSE INSTRUCTOR

❖ Dr. Malak Abdullah

❖ Ph.D. in Computer Science from UNC-Charlotte, NC, USA 2018

❖ Area of interests:

- ❖ Programming languages (mainly: C++, Java, Python, R)
- ❖ Data Science
- ❖ Natural Language Processing
- ❖ Machine Learning
- ❖ Deep Learning
- ❖ Sentiment and Emotion Detection
- ❖ Social Network Analysis

❖ Coached JUST teams for ACM programming competitions (2011-2014)

❖ Old website: <https://webpages.uncc.edu/~mabdull5>

❖ Office hours: To be announced later (TBAL)

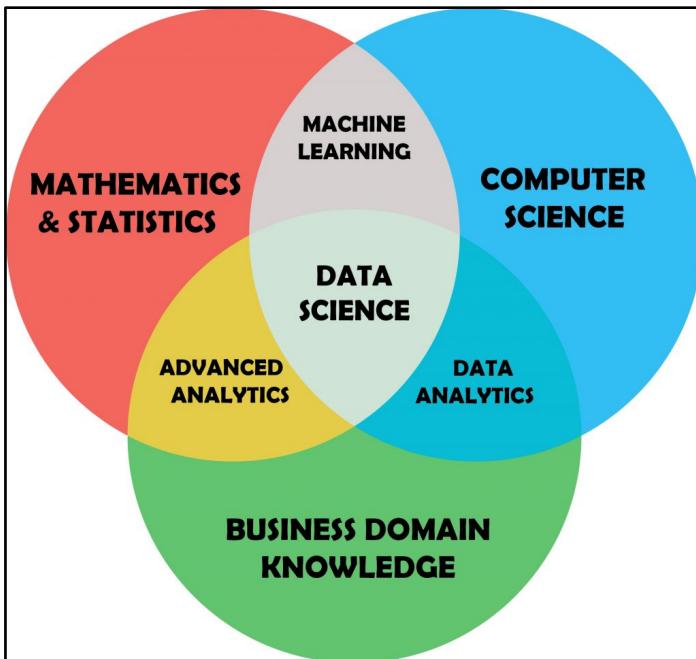
❖ We will have 2 TAs (to be announced later (TBAL))



WHAT IS DATA SCIENCE?

❖ **Data science** is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured

From Wikipedia, the free encyclopedia



Google best jobs in the us

All Images Videos News Maps More Settings Tools

About 2,070,000,000 results (0.63 seconds)

The 50 Best Jobs in America — And How Much They Pay

- Data Scientist.
- Devops Engineer.
- Marketing Manager.
- Occupational Therapist.
- HR Manager.
- Electrical Engineer.
- Mobile Developer.
- Product Manager.

More items...

The 50 Best Jobs in America | Money - Time
time.com/money/5114734/the-50-best-jobs-in-america-and-how-much-they-pay/

[About this result](#) [Feedback](#)

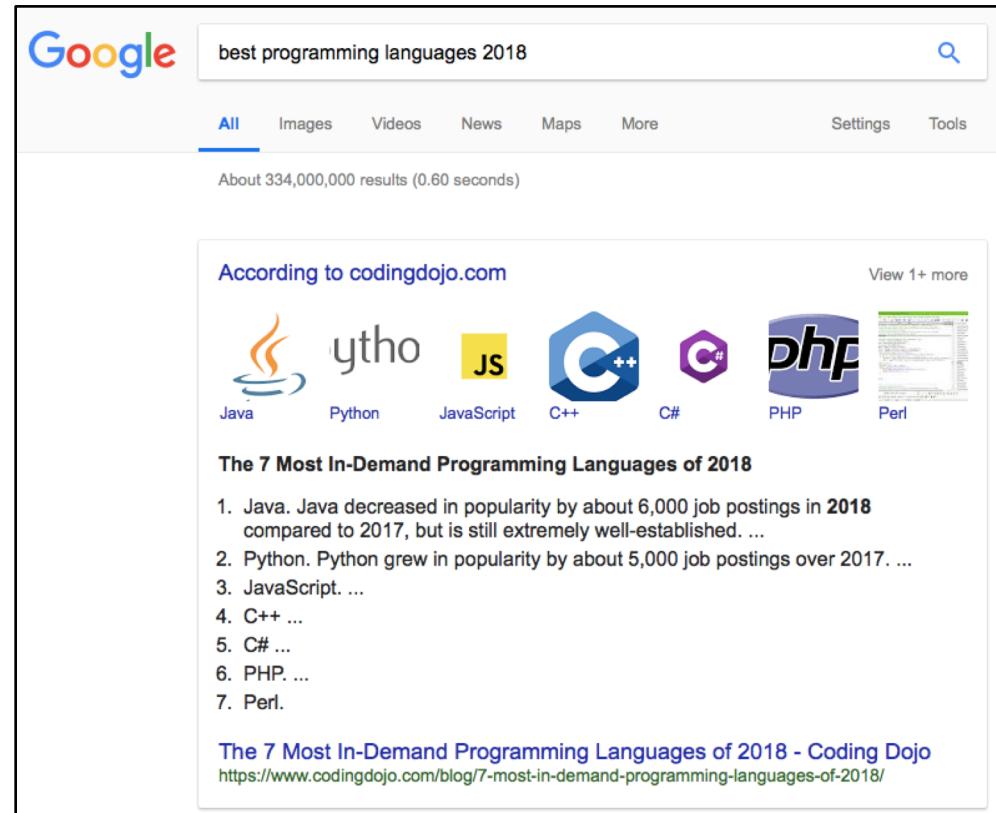
BEST JOBS
U.S. News & WORLD REPORT
RANKINGS
money.usnews.com

PREREQUISITES

- ❖ This course is intended for learners who have a basic knowledge of programming in any language (Java, C++, python, etc.).
- ❖ I will assume you already know how to:
 - ❖ Create and assign variables
 - ❖ Write programs with loops
 - ❖ Write programs with conditions
 - ❖ Create and use functions (methods)
- ❖ Knowing that most of you are unfamiliar with python, we have the first two weeks dedicated to getting you up to speed with basic programming in python.

COURSE OVERVIEW

- ❖ This course will introduce you to the field of data science!
- ❖ You will learn:
 - ❖ Data Analysis: includes import data, explore it, analyze it, learn from it, and visualize it.
 - ❖ Machine Learning
 - ❖ Natural Language Processing
- ❖ Tools including:
 - ❖ python
 - ❖ jupyter notebooks
 - ❖ pandas
 - ❖ numpy
 - ❖ matplotlib
 - ❖ scikit learn
 - ❖ nltk
 - ❖ And many other tools



LEARNING OBJECTIVE

- ❖ Basic process of data science
- ❖ Python and Jupyter notebooks
- ❖ An applied understanding of how to manipulate and analyze (text) datasets
- ❖ Basic statistical analysis and machine learning methods
- ❖ How to effectively visualize results

By the end of the course, you should be able to find a dataset, formulate a research question, use the tools and techniques of this course to explore the answer to that question, and share your findings.

COURSE OUTLINE

- ❖ Week 1 and 2: Python Basics
- ❖ Week 3: Visualizing data
- ❖ Week 4 and 5: Mathematics (Linear Algebra, Statistics, Probability)
- ❖ Week 6: Getting and Working with Data for our Project
- ❖ Week 7 and 8: Natural Language Processing
- ❖ Week 9 and 10: Machine Learning
- ❖ Week 11 and 12: Neural Network
- ❖ Week 13 and 14: DO DATA SCIENCE!!

GRADING OUT OF 100

❖ Quizzes	15
❖ Assignments and lab	25
❖ First Exam	10
❖ Second Exam	10
❖ Final exam	40 (Project 20 and Exam 20)

- ❖ For the final project, you will get a dataset (Twitter dataset) and learn how to explore it and detect emotions and sentiments by using machine learning and deep learning methods!
- ❖ Please hold your questions for the final project until later!

DEADLINES

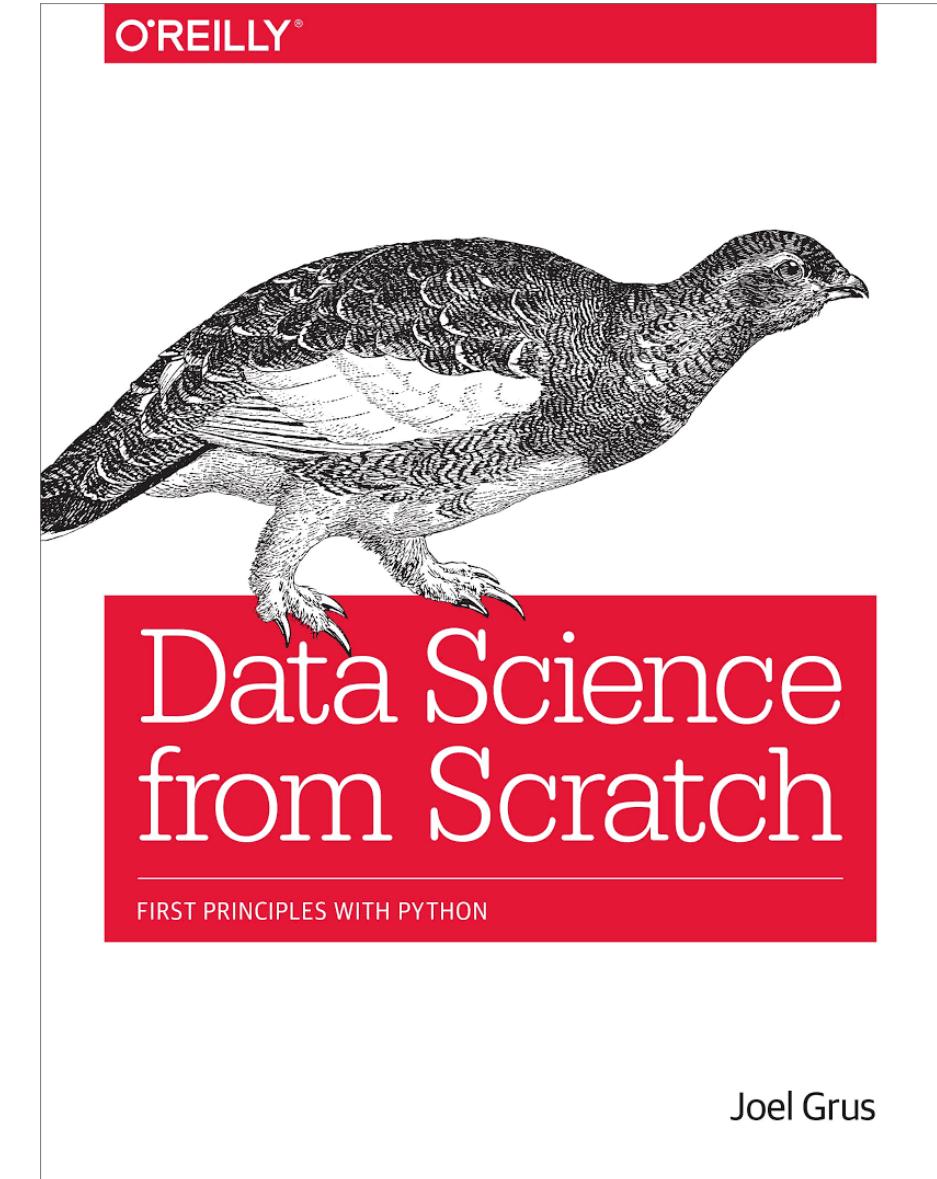
- ❖ Each Assignment will have a deadline. Every day late will deduct %20 of the grade.
- ❖ Copying code will get -5 for the first time (which means zero for the assignment and -5 from your exams). For the second time you will have to drop the course or you will get F for this course!
- ❖ No excuses for not submitting your assignments on time!
- ❖ There are no retake or makeup for the quizzes or the assignments.

COMMUNICATION

- ❖ Will be announced later!

TEXT BOOK AND REFERENCES

- ❖ Data Science from Scratch (Author Joel Grus)
- ❖ Reference: We will use different online videos and courses
 - ❖ To be announced later



ATTENDANCE

- ❖ Attendance is very important for the course. In accordance with university policy, students missing more than 10% of total classes are subject to failure. Penalties may be assessed without regard to the students performance. Attendance will be recorded at the beginning or end of each class.

NEXT CLASS

- ❖ Download Python and Anaconda
- ❖ Setup your Jupyter Notebooks

WELCOME ONCE AGAIN!

❖ See you next class!!