

CPSC380: INTRO TO DATA SCIENCE

COURSE INTRODUCTION

Dongsheng Che



Course Information

- Course Number: CPSC380
- Course Title: Introduction to Data Science
- Term: Spring 2023
- Time and Location: MWF 11-11:50AM, SciTech352

Office Hours

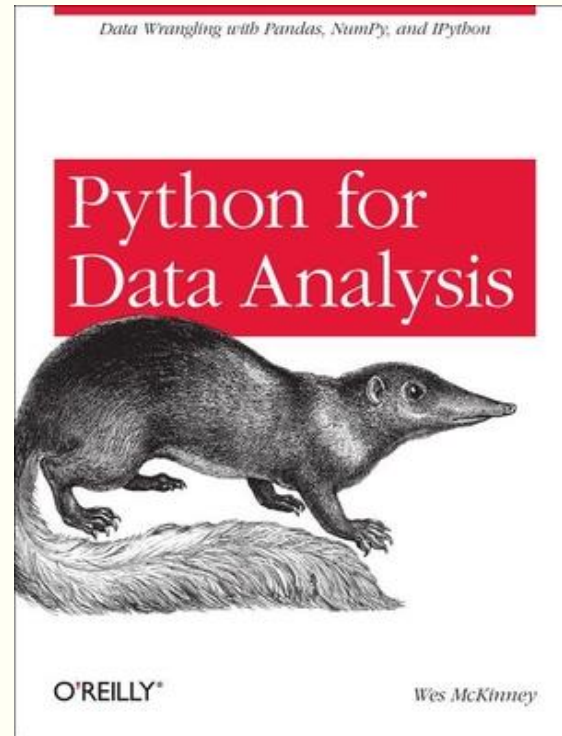
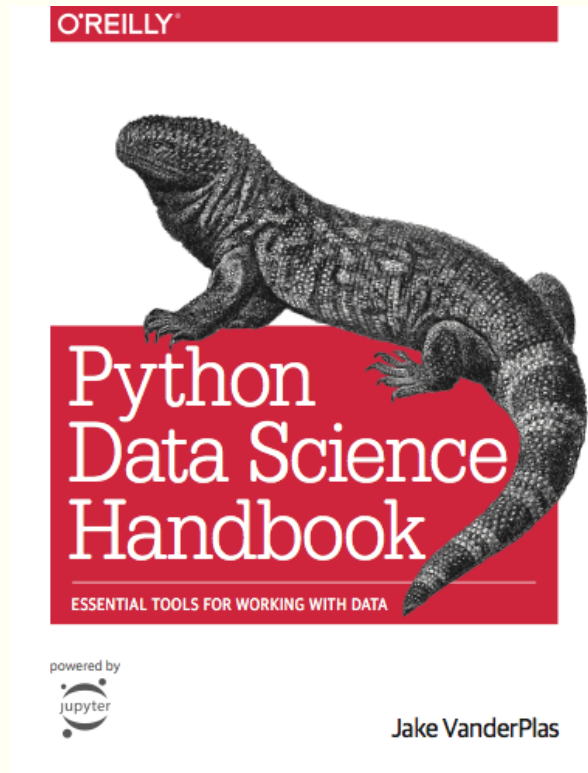
- Instructor: Dr. Dongsheng Che
- Email: dche@esu.edu
- Office Hours: MWF 9:20-11:00AM

Course Prerequisites

- One of the following programming courses:
 - CPSC120: Introduction to Computer Programming for Science and Engineering (Python)
 - CPSC230: Programming Principles and Practice (C++)
 - MATH318: Exploratory Data Analysis with R (R)
- One of the following statistics courses:
 - MATH 110 - GN: General Statistics
 - MATH 311 - Statistics I

Textbooks

(<https://www.py4e.com/book>)



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

<https://www.programmer-books.com/wp-content/uploads/2019/04/Python-for-Data-Analysis-2nd-Edition.pdf>

<https://github.com/wesm/pydata-book>

Other Online Resources

- Python (<https://www.python.org>)
- Numpy (<https://numpy.org/>)
- Python Data Analysis Library (<https://pandas.pydata.org/>)
- Matplotlib (<https://matplotlib.org/>)

Tentative Schedule

- Week1-3: Crash course on python
- Week4-5: Numpy
- Week6-9: Pandas
- Week10-11: Visualization
- Week12: Scikit-Learn
- Week13: Time Series
- Week14: Data Science Applications
- Week 15: Project presentations

Grading

- Seven programming assignments (70 points)
 - One program on python
 - Two programs on numpy
 - Two programs on pandas
 - Two programs on visualization
- Term project (30 points)
 - Pick a dataset on your own (more detail information will be provided)
 - Do exploratory data analysis
 - Write a project report
 - Present your project