

COMP 499 Introduction to Data Analytics

2019 Course Outline

Greg Butler

Data Science Research Centre

and

Centre for Structural and Functional Genomics

and

Computer Science and Software Engineering

Concordia University, Montreal, Canada

`gregb@cs.concordia.ca`

COMP 499 — Course Summary

Course Web Site:

<http://users.encs.concordia.ca/~gregb/home/comp499-s2019.html>

Instructor: Greg Butler, EV-3.219, gregb@encs

<http://users.encs.concordia.ca/~gregb>

Lectures: Tuesdays & Thursdays 13:15 – 15:45 H-562

Labs: Tuesdays & Thursdays 10:15 – 12:15 H-847

Labs: Tuesdays & Thursdays 16:15 – 18:15 H-847

Lectures and Labs are Mandatory.

Office Hours: By appointment in EV 3.219

- Ask questions at lectures!

Recommended Books

- ▶ *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*, by Wes McKinney, O'Reilly Media, 2012.
- ▶ *Data Crunching: Solve Everyday Problems using Java, Python and More*, by Greg Wilson, The Pragmatic Bookshelf, 2005. This book is out of print, but can be found in the Library.
- ▶ *Data Science from Scratch: First Principles with Python*, by Joel Grus, O'Reilly, 2015.
- ▶ *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*, by Hadley Wickham, Garrett Golemund, O'Reilly Media, 2016.
- ▶ Material on available data resources and data analysis problems from a range of disciplines will be provided.

Evaluation

Attendance	10%
Assignments \times 2	15%
Project — Individual	25%
Midterm Exam	15%
Final Exam	35%
Total	100%

You must pass both project and exam components of the course.

Assignments

Submit to eas electronic submission system as Jupyter notebooks.

Marked in Labs

Assignment 1 — Exploratory Data Analysis

Perform Lab 5 using Python to explore a dataset

Assignment 2 — Data Wrangling

Perform Lab 6-7 data wrangling task in Python for Movie dataset as Jupyter notebook

Project

Select a dataset — analyse it

What are the questions?

Data Wrangling

Exploratory Data Analysis

Model Construction

Story Telling

Project Report

Do project as a Jupyter notebook.

Prepare presentation for class to tell your story.

Submit pdf version of report to eas electronic submission system as Jupyter notebooks.

Examinations

Midterm

60 minute examination

true-false, multiple choice, short answer

focus on terminology and basic technical matters

Final

three-hour examination

true-false, multiple choice, short answer

focus on process of data analytics: data wrangling, exploratory data analysis, modeling, story telling, validation (maybe)

COMP 499 is a slot course!

Second time COMP 499 is being taught

Be prepared to be flexible.

Course schedule and content is subject to change.

My organization may not be the best.

Accelerated summer schedule!

COMP 499 moves at twice the normal pace.

Expect to spend 20+ hours/week on COMP 499.

Be proactive in watching videos, reading, and doing labs.

Labs are essential — this is a DOING subject!

My organization may not be the best.