QTM 530: COMPUTING I

Contact Hours: Two 1.25-hour or one 2.5-hour session weekly, plus 4.5 hours of regular, out-of-class work required as preparation for in-class work
Credit Hours: 3
Prerequisite: None

Instructor:	Xxx
Semester	Fall 20xx
Meeting Time and Place:	xxx
Office:	xxx
Office Hours:	xxx
Email/Contact:	xxx
Course Website:	xxx
TA:	xxx

COURSE OBJECTIVES

This course provides as a hands-on introduction to computing techniques related to data analysis. By the end of the course, students should be able to (1) retrieve, organize, and manipulate both structured and unstructured data formats from various sources including web, (2) gain fundamentals of computational backgrounds including advanced topics such as sampling and simulation methods, (3) use proper visualization to explore data for better interpretation, (4) be fluent in two programming languages, R and Python, for statistical and scientific computing, and (5) learn software engineering for professional management of user-custom packages. R Markdown and Jupyter Notebook will be used as the primary IDEs for programming exercises. Cloud computing using Amazon Web Services and Google Cloud Platform will be adapted for scalable computing.

CLASS REQUIREMENTS

Grades will be based on

- Homework assignments (25%)
- Quizzes (25%)
- 1 cumulative exam (25%)
- 1 data analysis project (25%)

HOMEWORK ASSIGNMENTS (25%)

Assignments will be given weekly, and you have approximately one week to complete each one. The homework will consist of data analysis. Each assignment will have its due date indicated and should be submitted through Canvas. Late submissions will be accepted but with penalty (25% grade reduction per day). No homework submitted four days after the due date will be graded. To accommodate unexpected circumstances, your lowest homework grade will be automatically dropped at the end of the semester. Working together on the homework assignments is encouraged, but you must submit your own work individually. It is also highly recommended that you make your solo effort on all the problems before consulting others.

QUIZZES (25%)

In addition to the weekly homework assignments, weekly quizzes will be administered in class to evaluate your understanding of the course materials on regular basis. Quizzes are to be completed individually. **No collaboration of any kind is allowed on the exams**. Similar to homework assignments, your lowest quiz grade will be dropped automatically at the end of the semester to accommodate unexpected circumstances.

EXAMS (25%)

One cumulative exam will be proctored in-class. **No collaboration of any kind is allowed on the exams**.

DATA ANALYSIS PROJECT (25%)

The final term project is a group-based work. Each group will consist of 4 to 5 students. The details of the project will be discussed after the Fall break.

HONOR CODE

All students enrolled at Emory are expected to abide by the Emory College Honor Code. Any type of academic misconduct is not allowed which includes 1) receiving or giving information about the content or conduct of an ex- amination knowing that the release of such information is not allowed and 2) plagiarizing, whether intentionally or unintentionally, in any assignment. For the activities that are considered to be academically dishonest, refer to the Honor Code: http://catalog.college.emory.edu/academic/policies-regulations/honor-code.html.

DISABILITY ACCOMMODATIONS

If you are seeking classroom accommodations or academic adjustments under the Americans with Disabilities Act, you are required to register with Office of Accessibility Services (http://accessibility.emory.edu). To receive academic accommodations for this class, please obtain the relevant letter and meet with me at the beginning of the semester. Students are expected to give two weeks' notice of the need for accommodations.

REQUIRED TEXTBOOK

- *R for Data Science: Visualize, Model, Transform, Tidy, and Import Data* by Hadley Wickham and Garrett Grolemund.

SUPPLEMENTAL TEXTBOOK

- *A First Course in Statistical Programming with R, 2nd edition*, by W. John Braun and Duncan J. Murdoch.
- *A Handbook of Statistical Analyses using R, 3rd edition,* by T. Hothorn and B. Everitt

Part I: Data Organization, Manipulation & Acquisition

Week 1: Statistical Computing

- R Programming
- Dynamic documentation: R Markdown (https://rmarkdown.rstudio.com)
- Version control: GitHub (https://github.com)
- Reading: Chapters 2 & 27, Wickham and Grolemund
- Assignment: HW1

Week 2: Data Organization

- Data types and structures
- Tables: tidy data, relational data
- Packages: `tidyr`, `forcats`, `lubridate`, `stringr`
- Reading: Chapters 12 & 13, Wickham and Grolemund
- Assignment: HW2

Week 3: Data Manipulation

- Search: `filter()`, `select()`
- Transformation: `arrange()`, `mutate()`
- Packages: `dplyr`
- Reading: Chapter 5, Wickham and Grolemund
- Assignment: HW3

Week 4: Data Acquisition

- Data collection: web scrapping
- Data aggregation into the standard file format: `CSV`, `JSON`
- Packages: `rvest`
- Reading: rvest: easy web scraping with R, Wickham
  https://blog.rstudio.com/2014/11/24/rvest-easy-web-scraping-with-r/
- Assignment: HW4

Part II: Fundamentals

Week 5: Programming Design

- Control structures: conditions, loops
- Data structures: vectors, matrices
- Functions: parameters, return types
- Reading: Chapters 19 – 21, Wickham and Grolemund
- Assignment: HW5

Week 6: Mathematical Foundation

- Probability
- Calculus
- Linear algebra
- Reading: TBA

- Assignment: HW6 and Group Data Analysis Project Distributed (GDAP)

Week 7: Sampling

- Sampling methods
- Random vs. conditional sampling
- Sampling with and without replacement
- Reading: https://en.wikipedia.org/wiki/Sampling_(statistics)
- Assignment: HW7

Week 8: Simulation

- Random number generation
- Density functions: uniform, gaussian, probabilistic
- Linear vs. non-linear models
- Reading: Chapter 20, Peng (https://bookdown.org/rdpeng/rprogdatascience)
- Assignment: HW8 and GDAP Outline

Part III: Data Presentation

Week 9: Data Visualization

- Plotting: statter, chi, bubble
- Kernel density estimation: geometric, topological
- Packages: `ggplot2`, `plotly`, `maps`
- Reading: Chapters 3 & 7, Wickham and Grolemund
- Assignment: HW9

Week 10: Interactive Web

- Data analytics
- Text mining
- Packages: `tidytext`, `shiny`
- Reading: Text Mining with R, Silge and Robinson (tidytextmining.com)
- Assignment: HW10

Part IV: Advanced Computing Techniques

Week 11: Scientific Computing

- Python programming
- Numerical computing
- Packages: `numpy`, `panda`, `jupyter`
- Reading: https://numpy.org/devdocs/
- Assignment: HW11 and GDAP draft

Week 12: Advanced Programming

- Data structures: list, set, dictionary
- Recursive calls
- Functional programming
- Reading: https://docs.python.org/3/howto/functional.html
- Assignment: HW12

Week 13: Software Engineering

- Debugging
- Source code control
- Interactive development and code review
- Reading: TBA
- Assignment: HW13

Week 14: Performance and Optimization

- Cloud computing
- Parallelism: multithreaded applications and concurrency
- Numerical computing in Graphics Processing Units (GPUs)
- Reading: TBA
- Assignment: HW14 and GDAP due