MATH 3430 Computing with Data

Course Description: Computational methods for collecting, manipulating, exploring, and graphing data. Basic principles of exploratory data analysis and statistical graphical methods. Methods for downloading and organizing large data sets. All of the computing methods will be illustrated using a high-level language such as R or Python.

Prerequisites: CS 2010 and C or better in MATH 1310 or MATH 1340 and MATH 1350.

Instructor Information:
Name:
E-mail:
Class time:
Class meeting place:
Office hours:

Course Aim: The general plan of the course is to introduce the student to computing methods for collecting, manipulating, exploring, and graphing data. The basic principles of exploratory data analysis and statistical graphics will be presented together with programming commands for implementing these statistical methods. The student will be introduced to the challenges of downloading and organizing data from the Internet. Once a statistical data analysis is presented, the student will learn software tools for writing reports that combine statistical output and textual material.

Learning Objectives:

- (EDA) Understand and apply basic software commands for graphing and summarizing a single collection of numerical data
- (EDA) Understand and apply software commands for comparing collections of numerical data by graphs and summaries
- (EDA) Understand and apply software commands for graphing and summarizing relationships between numerical data
- (EDA) Understand and apply commands for tabulating and graphing tabular data, to understand distributions and relationships
- (Data Collection) Understand and apply methods for importing data tables into the statistical software program
- (Data Manipulation) Understand the basic data types represented in a statistics software program
- (Data Manipulation) Understand and apply software commands for manipulating data including selecting specific cases or variables, sorting, creating new variables, or summarizing groups by different measures
- (Large Datasets) Understand and apply basic strategies for working with large datasets
- (Large Datasets) Understand and implement basic tasks with databases
- (Graphing) Understand and be able to implement basic principles in creating graphs

- (Graphing) Understand and be able to apply software commands for creating basic graphs for one or two numerical or categorical variables
- (Presentation) Understand software tools for blending text and statistical content and be able to write a statistical report using these tools

Texts:

- (Required) Wickham and Grolemund (2016), R for Data Science, O'Reilly
  Online version of text: http://r4ds.had.co.nz/
- Albert, J. and Rizzo, M. (2012), *R by Example*, Springer.
  (Chapters are downloadable through SpringerLink.)

Software:

The course will be using R and RStudio software and it is expected that the student has a laptop/desktop where both R and RStudio can be installed. Many of the assignments will be submitted as R Markdown documents created in RStudio.

Course Schedule (tentative)

Week 1 Introduction to data science, R Markdown
Week 2 R basics (variables, vectors, data frames)
Week 3 Introduction to graphing using ggplot2
Week 4 ggplot2 work (continued)
Week 5 Manipulations with data frames using dplyr
Week 6 Working with dplyr and other data wrangling packages
Week 7 Exploring data for single variable
Week 8 Exploring data – covariation between multiple variables
Week 9 Reading in data from files
Week 10 Working with character data, regular expressions
Week 11 Character data (continued)
Week 12 Mining twitter data
Week 13 Working with mapping data
Week 14 Final project
Week 15 Final project

Assessment:

Student performance will be evaluated by (1) assignments, (2) work in the classroom labs and (3) work on the mini projects.

Each in-class lab and mini-project will be graded both on the correctness of the data programming and on the correct interpretation of the computational work in the context of the application. Since the mini-project will sometimes require the students to find their own data, part of the grading will be based on the suitability of the chosen dataset to address the particular computational data task.

Grading:

Standard 90-100 A, 80-89 B, etc. Your final grade is expected to be weighed as follows:

Class assignments	50%
Mini-projects and final project	50%

Pedagogical Efforts for Learning:

The data science material is learned by working on activities and course assignments. Much of the class time will be devoted to activity-style learning. If you need help, it is strongly encouraged that you ask questions to the instructor or fellow students. At the present time, the Learning Commons is not equipped to handled questions on data science material.

University Policies and Department Mediation:

Disability Policy: In accordance with the University policy, if the student has a documented disability and requires accommodations to obtain equal access in this course, he or she should contact the instructor at the beginning of the semester and make this need known. Students with disabilities must verify their eligibility through the Office of Disability Services for Students, 413 South Hall, 419-372-8495. (http://www.bgsu.edu/offices/sa/disability/ (Links to an external site.))

Student veteran-friendly campus: BGSU educators recognize student veterans' rights when entering and exiting the university system. If you are a student veteran, please communicate with your instructor so reasonable accommodations can be made for absence when drilling or being called to active duty (See http://www.bgsu.edu/veteran/ (Links to an external site.) for more information).

Codes of Conduct and Academic Honesty Policy: The instructor and students in this course will adhere to the University's general Codes of Conduct defined in the *BGSU Student Handbook*. Specifically, the Code of Academic Conduct (Academic Honesty Policy) requires that students do not cheat, fabricate, plagiarize or facilitate academic dishonesty. Students who passively engage in cheating (i.e. allowing others to cheat off of them) may receive the same consequences as the person copying.

Departmental Mediator: Dr. Kit Chan, kchan@bgsu.edu

See http://www.bgsu.edu/departments/math/mediator (Links to an external site.).