Math 3013X1 / 5883X1
Computing with Data - Fall 2017
Course Outline (September 6 version)

Instructor: Hugh Chipman, Hayward House, Room 110, hugh.chipman@gmail.com
Office Hours: Mon 2:30 - 3:30, Thu 2:00 - 3:00, or by appointment, or by chance
Lectures: T/Th 8:30-10:00, HSH 147
Textbook: "R for Data Science" by Garrett Grolemund and Hadley Wickham available free online at http://r4ds.had.co.nz/ Hard copy can be purchased online (see webpage). Additional online resources will be made available as the term progresses (see Acorn)

Marking Scheme
Tests (2): 25%
Assignments (5): 25%
Project: 25%
Exam: 25%

5883 Marking Scheme:
The independent study (below) will be worth 10 points. All other elements have the same values as in the 3013 marking scheme. Course mark = 100*(your mark out of 110)/110.

Assignments: There will be 5 assignments during the term (dates below). Solutions to all questions will be posted after the due date. Late assignments will not be accepted, except for special circumstances.

Tests: There will be two tests, given in class (dates below). Tests will be open book, open computer.

Project: There will be a take-home project assigned. Students will be expected to apply the methods learned in the course to the analysis of a substantial data set.

Exam: The exam a regular 3-hour exam scheduled by the registrar's office. It will be open book, open computer.

Expectations for 5883:
The graduate offering of this course will be held with 3013. Additional expectations are:

1. Some assignment, test and exam questions may be different from 3013;
2. Independent study of a more advanced topic. Present to the class a summary of a set of assigned readings and demonstration of related methods in R.

It is anticipated that graduate students will meet with the instructor outside of class to discuss assigned readings (research papers). Note that this presentation is in addition to the project.

Tentative course schedule:

Tuesday Sep 19	Assignment 1 due
Tuesday Oct 3	Assignment 2 due
Thursday Oct 12	Test 1
Tuesday Oct 24	Assignment 3 due
Thursday Nov 2	Assignment 4 due
Thursday Nov 9	Test 2
Tuesday Nov 21	5883 presentations
Tuesday Nov 28	Assignment 5 due
Tuesday Dec 5	Project due

Computer Use:
The course will make extensive use of the (free) R software environment for statistical computing and graphics, and the RStudio add-on interface to R (www.rstudio.com). Students are encouraged to install R + RStudio on their laptop before classes begin. Copies of R may be downloaded from http://www.r-project.org/ . R and RStudio run on Linux, Windows and MacOS. See also https://hub.acadiau.ca/TDClient/KB/ArticleDet?ID=309

Acorn:
All course notes and other information will be posted on Acorn.

Note: It is your responsibility to be aware of course requirements and deadlines, and to be present for tests and exams. If you are unable to attend a lecture, it is your responsibility both to make up any material covered in your class, as well as find out any announcements made during the lecture missed. It is a courtesy to your instructor to advise him when you must be absent from lectures. Missing tests or assignment deadlines because of illness or personal emergencies should be documented by signed excuses from physicians or counsellors.

Although collaboration in solving assignments is acceptable, the copying of assignments is inappropriate and will be considered cheating. Submitting someone else's work as your own, copying on tests, and other forms of cheating will be dealt with under the clause concerning Academic Integrity in the Acadia Academic Calendar.

Students with Disabilities
If you are a student with a documented disability who anticipates needing accommodations, please inform Kathy O'Rourke (902-585-1823) in Accessible Learning Services disability.access@acadiau.ca.

PREREQUISITES: Math 2223 or 2243 with a minimum of C-.

Course description:
Students will learn to implement a substantial statistical analysis project in a programming environment. This will include efficient storage, retrieval and manipulation of data; exploration, visualization, modelling, statistical inference and prediction; coding to support reproducible research, including writing algorithms and conducting simulations; scaling up computation for handling big data.