DATA 301 / DATA 501 / COSC 301 - Introduction to Data Analytics
Winter 2017 Term 2

Instructor:	**Dr. Ramon Lawrence**
Class Schedule:	2:00 p.m. – 3:30 p.m. Tuesday/Thursday
Location:	ART 366
Lab time/location:	**L01** - 3:30 p.m. to 5:30 p.m. Tuesdays in SCI 234
	L02 - 5:30 p.m. to 7:30 p.m. Tuesdays in SCI 234
	L03 - 5:30 p.m. to 7:30 p.m. Wednesdays in SCI 234
	L04 - 3:30 p.m. to 5:30 p.m. Wednesdays in SCI 234
	L05 - 3:30 p.m. to 5:30 p.m. Thursdays in SCI 234
	L06 - 5:30 p.m. to 7:30 p.m. Thursdays in SCI 234
Office Hours:	9 a.m. to 10 a.m. Tuesday/Thursday or by appointment
Office Location:	ASC 349
Phone:	807-9390
E-mail:	ramon.lawrence@ubc.ca (preferred contact method)
Course URL:	https://people.ok.ubc.ca/rlawrenc/teaching/301/

Course Description

Official Calendar: Techniques for computation, analysis, and visualization of data using software. Manipulation of small and large data sets. Automation using scripting. Real-world applications from life sciences, physical sciences, economics, engineering, or psychology. No prior computing background is required. Credit will be granted for only one of COSC 301, DATA 301 or DATA 501. [3-2-0] **Prerequisite:** Third-year standing.

Specific description: This course provides an introduction to data analytics to train students with practical industrial techniques for data manipulation, analysis, reporting, and visualization.
This is not an introduction to programming using Python. Programming techniques will be taught to automate data analysis. Introduction to programming courses are COSC 111 or COSC 123. Prior computing experience is not required, but is helpful including COSC 122 or COSC 111.

Marking and Evaluation

Clickers	**5 %**	
Assignments	**30 %**	(weekly assignments)
Two Midterm Exams	**30 %**	(in class, 15% each)
Final Exam	**35 %**	(cumulative, three hours)

Graduate Student Evaluation:

Clickers	**5 %**	
Assignments	**20 %**	(weekly assignments)
Project	**10 %**	
Two Midterm Exams	**30 %**	(in class, 15% each)
Final Exam	**35 %**	(cumulative, three hours)

- **Graduate students are responsible for a substantial data analytics project. Optional or bonus assignment questions for undergraduates will often be required for grad students.**
- **A student must receive a combined grade of at least 50% on the exams (midterms and final) to pass the course. Otherwise, the student will be assigned a maximum overall grade of 45.**

Textbook and Reference Material:

- ***A clicker is required.*** All notes are available online. A textbook is ***not required***.

Expectations

- Attend **all** classes and prepare before attending class.
- Read the notes **before** the lecture.
- Learn the material in the course by completing all assignments.
- Enjoy attending class and feel free to participate according to your own personalities. Feel free to ask questions by raising your hand or speaking out at appropriate times.
- Please actively participate in class discussions, questions, and problem solving exercises.
- **I want all students to pass the course, receive a good grade, and feel the course was beneficial.**

Homework Expectation

For this course, it is expected that you will spend ***at least six hours per week in out-of-class preparation***.

Grievances and Complaints Procedures

A student who has a complaint related to this course should follow the procedures summarized below.
• The student should attempt to resolve the matter with the instructor first. Students may talk first to someone other than the instructor if they do not feel, for whatever reason, that they can directly approach the instructor.
• If the complaint is not resolved to the student's satisfaction, the student should go to the departmental chair John Braun at SCI 388, 807-8032.

Your Responsibilities

Your responsibilities to this class and to your education as a whole include attendance and participation. You have a responsibility to help create a classroom environment where all may learn. At the most basic level, this means you will respect the other members of the class and the instructor and treat them with the courtesy you hope to receive in return. Inappropriate classroom behavior may include: disruption of the classroom atmosphere, engaging in non-class activities, talking on a cell-phone, inappropriate use of profanity in classroom discussion, use of abusive or disrespectful language toward the instructor, a student in the class, or about other individuals or groups.

Academic Integrity

The academic enterprise is founded on honesty, civility, and integrity. As members of this enterprise, all students are expected to know, understand, and follow the codes of conduct regarding academic integrity. At the most basic level, this means submitting only original work done by you and acknowledging all sources of information or ideas and attributing them to others as required. This also means you should not cheat, copy, or mislead others about what is your work. Violations of academic integrity (i.e., misconduct) lead to the breakdown of the academic enterprise, and therefore serious consequences arise and harsh sanctions are imposed. For example, incidences of plagiarism or cheating may result in a mark of zero on the assignment or exam and more serious consequences may apply if the matter is referred to the President's Advisory Committee on Student Discipline. Careful records are kept in order to monitor and prevent recurrences. **If you have any questions about how academic integrity applies to this course, please consult with your professor.**

Disability Services

If you require disability-related accommodations to meet the course objectives, please contact the Disability Resource Centre in UNC 227. More information is at: http://students.ok.ubc.ca/drc.

Equity, Human Rights, Discrimination and Harassment

UBC does not condone discrimination or harassment in classrooms, living or work environments on campus. For information about UBC's policies related to equity go to: http://equity.ok.ubc.ca/.

Missing an Exam

Only students who miss the final exam for a reason that corresponds to the University of British Columbia Okanagan's policy on excused absences from examinations will be permitted to take the final exam at a later time. A make-up exam may have a question format different from the regular exam. **There will be no make-up midterm exams.** If the reason for absence is satisfactory, the student's final exam will be worth more of the final grade. Further information on Academic Concession can be found under Policies and Regulation in the Okanagan Academic Calendar http://www.calendar.ubc.ca/okanagan/index.cfm?tree=3,48,0,0.

Course Objectives

***Course Format*:** Interactive classes consisting of topic introduction, understanding evaluation using clickers, and concept mastery with in-class exercises. Practical skills and applications of topics are covered in computer labs as well as practice using data analytics systems and software.

Learning Outcomes:

- Ability to manipulate, extract, convert, and integrate data from different sources.
- Ability to perform advanced Excel analysis including what-if scenarios, pivot tables, and VBA scripting.
- Ability to use relational databases including creating tables and querying using SQL.
- Ability to use scripting programs to automate repetitive and large tasks and improve efficiency.

Course Objectives:

- Understand data representation formats and techniques and how to use them.
- Experience using a wide-range of data analytics tools including Excel, SQL databases, GIS, and visualization and reporting software.
- Develop a computational thinking approach to problem solving and use programs to solve data tasks.

Graduate Student Objectives:

- In addition to the skills learned in the cross-listed DATA 301 course, DATA 501 will extend the student's skills in conducting independent research studies with data analytical techniques. These skills may also be applied to research in the student's thesis work.
- The project will involve background literature review, data analysis of a real-world data set, and public presentation of the results.

Project Description

A data analysis project is selected by the student. The project must manipulate a real-world data set, perform analysis with at least two of the tools/techniques in the course (Excel, SQL, Python, R), and be summarized in a report and public presentation. Projects where the data analyzed can have real-world impacts by changing policy or behavior are especially encouraged.

The project is worth 10% of the overall grade. The project has several deliverables with the following percentages:

- Proposal (10%) – Topic discussion, selection of data set, background research
- Initial Results (10%) – Mid-semester checkpoint on progress and initial analysis
- Presentation (10%) – Presentation of results
- Final report (70%) – Overall report and data analysis and recommendations

A student that receives less than 60% on the project will fail the course and be awarded a maximum overall mark of 45.

Course Outline

The course has a substantial amount of material to be covered in a short time. This requires the student make a strong effort to prepare before class so that the material can be practiced in class. Below is an outline of the topics. The professor is not bound to these topics and timelines as they only serve as a general reference.

Date	Topics Covered and Description
January 4 (TH)	**First day of classes. Introduction to course. What is data analytics?**
January 9 (T)	Data Representation: Data and metadata ; file formats and encoding ; text/binary files
January 11 (TH)	Excel I: Introduction to Excel – formulas, formatting, aggregate functions
January 16 (T)	Excel II: Data Analysis in Excel – sorting, filtering, charts, what-if scenarios, pivot tables
January 18 (TH)	Excel III: Excel scripting – macros, VBA
January 23 (T)	Databases I: Introduction to Relational Databases ; Creating a Database
January 25 (TH)	Databases II: Querying using SQL
January 30 (T)	Databases III: Advanced Querying using SQL
February 1 (TH)	Linux/Unix Command Line Introduction
February 6 (T)	Python I: Introduction to Python
February 8 (TH)	**Midterm Exam #1**
Feb. 13 & 15	**No classes during mid-term break**
February 20 (T)	Python II: Decisions and Loops
February 22 (TH)	Python III: Reading and Writing Files
February 27 (T)	Python IV: Data Analysis with Python
March 1 (TH)	Python V: Data Analysis with Map-Reduce
March 6 (T)	Data Analysis with R – Brief statistics intro/review
March 8 (TH)	Data Analysis with R (cont.)
March 13 (T)	**Midterm Exam #2**
March 15 (TH)	Data Analysis with R (cont.)
March 20 (T)	GIS
March 22 (TH)	Data Visualization I: Reporting
March 27 (T)	Data Visualization II: Tableau
March 29 (TH)	Data Analytics applications: Google Analytics, Scientific Data Sets, Open Data, Web Data Services, Data Integration
April 3 (T)	**Demonstrations of data analytics projects**
April 5 (TH)	**Review for final exam**

Laboratory times: The laboratory time will be spent performing assignments and practice questions using data analytics software.

Week	Dates	Topics Covered and Description
1	January 1 - 5	**No labs during first week of class**
2	January 8 – 12	Lab 1: Excel – Analyzing and Reporting Data
3	January 15 – 19	Lab 2: Excel – Pivot tables and What-If Scenarios
4	January 22 – 26	Lab 3: Excel – Macros and VBA Scripting
5	Jan. 29 – Feb. 2	Lab 4: Creating and Querying a Relational Database in Microsoft Access
6	February 5 – 9	Lab 5: Introduction to Python
7	February 12 - 16	**No labs during mid-term break**
8	February 19 – 23	Lab 5: Introduction to Python (cont.)
9	Feb. 26 – Mar. 2	Lab 6: Python Data Analysis
10	March 5 – 9	Lab 7: Python Databases and Map-Reduce
11	March 12 – 16	Lab 8: Introduction to R
12	March 19 – 23	Lab 9: GIS using Google Maps
13	March 26 – 30	Lab 10: Visualization with Tableau
14	April 2 - 6	**No new labs. Work on bonus labs.**