

DRAFT: JSGS 843 Data Science for Health Analytics and Decision Support

| | UNIVERSITY OF REGINA CAMPUS | UNIVERSITY OF SASKATCHEWAN CAMPUS |
|-------------------------|---|-----------------------------------|
| INSTRUCTOR: | Dr. Gavin Simpson | |
| PHONE: | 1 306 337 8863 (Office) 1 306 201 4791 (Cell) | |
| E-MAIL: | gavin.simpson@uregina.ca | |
| STUDENT HOURS: | TBC | |
| OFFICE LOCATION: | Working from home | |
| TERM: | Fall 2020 | |
| ROOM: | Online | |
| DATE AND TIME: | Sep 2nd – Dec 8th 2020 | |

CALENDAR DESCRIPTION

In this course you will learn the key data literacy and data analysis skills required to investigate complex data sets to answer pressing health care questions and effectively communicate your results to peers. Taking a broad-but-shallow approach, the course follows the stages of the CRISP-DM data life cycle; you will learn to import and filter data using databases; how to prepare data for analysis; to choose appropriate data visualizations; to perform exploratory data analysis to understand the properties of data; to use modern machine learning techniques to analyze data, and; how best to present your findings and use them to inform evidence-based decision making. No prior programming or coding experience is required to succeed on the course, just a willingness to learn new ways to use a computer to process and analyze data.

ATTRIBUTES OF JSGS GRADUATES

1. Management, Governance, and Leadership: Ability to inspire support for a vision or course of action and successfully direct the teams, processes, and changes required to accomplish it.
2. Communication and Social Skills: Ability to communicate effectively and build enduring, trust-based interpersonal, professional relationships.
3. Systems Thinking and Creative Analysis: Ability to identify key issues and problems, analyze them systematically, and reach sound, innovative conclusions.

4. Public Policy and Community Engagement: Ability to understand how organizational and public policies are formulated, their impact on public policy and management and how to influence their development.
5. Continuous Evaluation and Improvement: Commitment to on-going evaluation for continuous organizational and personal improvement.
6. Policy Knowledge: Ability to analyze and contribute content to at least one applied policy field.

COURSE CONTENT AND APPROACH

The course will be taught as a combination of recorded lectures and computer-based lab classes, with weekly online meetings where we will discuss the content for that week and allow for Q&A, clarification etc.

COURSE OUTLINE AND ASSIGNMENTS

The course will cover the following topics (the order of topics may change)

- What is data science?
- Data types and storage
- Data wrangling
 - Data import and processing
- SQL and working with databases
- Data visualization
 - The theory of visualization
 - The grammar of graphics with ggplot
 - Dashboards
- Descriptive statistics
 - Simple data summaries such as the mean, variance, range and how to compute these from large data
- Machine learning basics
 - Overview of machine learning
 - Tree-based models
 - Text mining and analysis
- Dynamic reporting
 - Literate programming and why
 - RMarkdown

- Version control
 - Why do we need version control?
 - Using git for version control with GitHub
- Working from the command line
 - Navigating the file system

The entire course will be available from the start of classes (September 2nd) and you are free to work at your own pace. There are no internal deadlines within the course; all assessed elements are to be submitted by Dec 8th. The weekly online student hours sessions will be an opportunity for students to ask any questions about the course, assignments, etc.

DESCRIPTION OF ASSIGNMENTS

Lab assignments will test your understanding of the course content and application of practical data analysis skills. Building upon analysis methods developed in the computer-based lab classes, you will be tasked with analyzing a similar data set and answering a set of questions related to the analysis. This requires you to demonstrate the practical R skills required to investigate and analyze a data set, prepare a data visualization, or short summary report, produce a dashboard, etc.

COMPUTER AND SOFTWARE

You will need a desktop or laptop computer to participate in the course. The recorded videos can be accessed using YouTube on a variety of devices, but you will need a computer and a version of RStudio software installed locally or via Rstudio.cloud. Other software may need to be installed too to complete all the assignments and activities. Full details will be provided to you ahead of the course start date so you can prepare your computing environment before classes start.

COURSE TEXTBOOK

There are two course texts. The first course text is: Freeman and Ross (2019) *Programming Skills for Data Science*. The book is currently CAD\$23.99 for the Kindle edition, which you can read on your desktop/laptop or tablet if you have one – you don't need a Kindle to read it. [[Amazon Link](#)]. The paperback version is currently CAD\$58.74 [[Amazon Link](#)]. For option purchase options, see the Google Books [[page](#)] for the book.

The second course text is: Kelleher & Tierney (2018) *Data Science*. The MIT Press Essential Knowledge Series. The Kindle edition is currently CAD\$13.19 [[Amazon Link](#)] while the paperback version is CAD\$21.22 [[Amazon Link](#)]. For option purchase options, see the Google Books [[page](#)] for the book.

Throughout the course, we'll make extensive use of a range of open access resources

EVALUATION

The course will be assessed via a combination of:

- multiple choice and short answer questions/quizzes completed online through URCourses
- lab assignments involving analysis of a representative & relevant data set and answering related questions arising
- preparation of a short report, data visualization, dashboard, etc. that builds upon the skills learned in each module

The aim of these assignments is to test your understanding of the principles and processes of data science as well as the practical data analysis and interpretive skills required to use and analyze public health data.

Additional informal quizzes will be used to evaluate student progress; feedback for these will be provided so you are strongly encouraged to complete them, but they will not be formally assessed as part of your final mark for the course.

Grades will be assigned on the following basis: one mark is worth one mark towards your total mark for the course and therefore your final % will be calculated as *marks achieved / total marks available*.

LEARNING REMOTELY

The University of Regina has prepared a resource page with lots of useful information for preparing to study remotely. While this is especially relevant for students taking classes that are switching from face-to-face to remote delivery, there is lots of useful information relevant to students taking this course. [[link](#)]

LATE ASSIGNMENTS

Assignments will be accepted through URCourses, up to the end of the course (8th December, 2020). Any assignments that are submitted beyond the end of the course will be awarded a grade of 0. No extensions will be provided as final marks for the course need to be determined and entered by 8th December, 2020. It is envisaged that you will organize yourself to complete the assignments at your own pace throughout the period the course is running. As such there are no deadlines for individual assignments. This allows you the greatest flexibility to study at your own pace.

STUDENTS WITH SPECIAL NEEDS

University of Regina (U of R): Students in this course who, because of a disability, may have a need for accommodations are encouraged to discuss this need with the instructor and to contact the Coordinator of Special Needs Services at (306) 585-4631.

U OF S: Students in this course who, because of a disability, may have a need for accommodations are encouraged to discuss this need with the instructor and to contact Disability Services for Students (DSS) at 966-7273.

STUDENTS EXPERIENCING STRESS

University of Regina (U of R): Students in this course who are experiencing stress can seek assistance from the University of Regina Counseling Services. For more information, please see the attached document, visit this website: <http://www.uregina.ca/student/counselling/contact.html>, or call (306) 585-4491 between 8:30 a.m. to 4:30 p.m. Saskatchewan time Monday to Friday.

ACADEMIC INTEGRITY AND CONDUCT

U of R: Ensuring that you understand and follow the principles of academic integrity and conduct as laid out by the University of Regina (available at <http://www.uregina.ca/gradstudies/grad-calendar/policy-univ.html>) is vital to your success in graduate school. Ensuring that your work is your own and reflects both your own ideas and those of others incorporated in your work is important: ensuring that you acknowledge the ideas, words, and phrases of others that you use is a vital part of the scholarly endeavour. If you have any questions at all about academic integrity in general or about specific issues, contact your course instructor to discuss your questions.

U OF S: Understanding and following the principles of academic integrity and conduct as laid out in the University of Saskatchewan's Guidelines for Academic Conduct is vital to your success in graduate school (available at www.usask.ca/university_secretary/council/reports_forms/reports/guide_conduct.php). Ensuring that your work is your own and reflects both your own ideas and those of others incorporated in your work is important: ensuring that you acknowledge the ideas, words, and phrases of others that you use is a vital part of the scholarly endeavour. If you have any questions at all about academic integrity in general or about specific issues, contact any faculty member and we can discuss your questions.