ISOM33390 Business Programming in R – Fall 2018

Instructor	Jia Jia
Contact	Email: justinjia@ust.hk Office: LSK 5045 **Begin subject: [ISOM3390]... <--- Note!**
Office Hours	By appt.
Course Schedule and Classroom	Lecture: Mon. & Wed. 9:00 – 10:20 (LSK1005) Lab: Thu. 9:00 – 9:50 (LSKG005)
Course Website	Accessible from Canvas

1. Course Overview

In the era of Internet of Things (IoT) and big data, in order to fuel their decision making, firms need to analyze massive amounts of data for idiosyncratic problems in more efficient ways. Business applications and analytics relying only on basic spreadsheets and prepacked software tools are no longer adequate, and implementing sophisticated algorithms with custom programs starts to prevail. In addition, businesses are spending more time capturing data from various sources and curating the data before applying advanced analytic techniques. As such, mastering a programming language that helps in accomplishing all these tasks is crucial for students aiming for data applications or business analytics jobs.

With its extensive data visualization capabilities and continuously growing libraries, R is widely considered the broadest analytical platform in the field of business analytics. This course can prepare you with R programming skills for putting analytics and modeling techniques into use by working with cases of emerging business applications, including Web scraping, text analytics, social network analysis, etc.

2. Course Goals and Objectives

At the end of this course, you will be able to:

- Understand generic programming language concepts in R
- Know how to obtain data from a variety of sources and tidy data for downstream analysis tasks with R
- Understand the basic principles of constructing data graphics and be familiarized with the plotting systems and visualization features in R
- Understand how to write R scripts and use various R packages for business applications
- Use R Markdown to write reports that includes R code and the code's automatically-generated output

3. Prerequisites

ISOM3230 Business Applications Programming & ISOM3360 Data Mining

4. Lecture Notes, Textbook, and Readings

For most classes I will hand out lecture notes, which will outline the primary material for the class. Other readings (posted to Canvas or distributed in class) are intended to supplement the material we

learn in class. They give alternative perspectives and additional details about the topics we cover.

5. Course Website

All materials for the course will be posted on Canvas, including the syllabus, readings, slides used in class, and homework assignments. Please check the course website frequently for updates.

6. Grading

The grade breakdown is as follows:

- Lab participation: 10%
- Homework (4): 40%
- Midterm quiz: 25%
- Final project: 25% (10% out of 25% will be based on your team-mates' assessment of your contribution)

7. Important Notes on Labs

This is primarily a lecture-based course. But follow-up labs and assignments ensure that you will get a hands-on experience with concepts learned from lectures. Therefore, student participation is an essential part of the learning process. During the lab session, I will expect you to be entirely devoted to the class by following the instructions and completing the exercises.

8. Homework Assignments, Midterm Exam, and the Final Project

The **homework assignments** are designed for you to explore specific topics in a structured way. There will be a total of **4 individual** homework assignments (using Canvas website), each comprising conceptual questions to be answered and hands-on tasks. Assignments will be graded and returned promptly. **The due date of each assignment** will be announced upon its release on Canvas.

You may work together on the homework assignments, but all of the material that is turned in for grading must be produced individually. For example, you may form study groups and work out homework solutions together and then sharing what you've learned, but it would not be permissible for someone to prepare an answer set and then for others to copy those answers and submit it as their own work. Turning in copied files is specifically prohibited; you **must** individually write (type) any material that is submitted for grading.

Late policy: turn in your assignment early if there is any uncertainty about your ability to turn it in on the due date. Assignments up to 24 hours late will have their grade reduced by 25%; assignments up to one week late will have their grade reduced by 50%. After one week, late assignments will receive no credit.

An in-class **midterm quiz** is to be scheduled on **October 22**. Makeup quizzes will be offered only in cases of documented health or family emergencies or for official, university-sanctioned activities. Advanced notification of missing an examination is required. Any uncoordinated absence from an exam will result in a score of 0 for the exam.

No final will be administered. Instead, you will be assigned to small groups to work on a final project. You will select project topics provided by the professors. (Multiple groups can take on the same project.) Each group will cooperate on writing code, documenting it, writing a report, and making a presentation on the

project. One component of your final project grade will be based on your team-mates' assessment of your contribution to the project.

9. Tentative Schedule of Lecture Topics

The following table shows the planned list of topics that we plan to cover. Please note that this schedule is tentative and is subject to adjust as the semester progresses.

<table>
<tr><th>Week</th><th>Date</th><th>Lecture Topic</th><th>Remarks</th></tr>
<tr><td>1</td><td>Sep. 3</td><td>Topic 1: Course Introduction

• Course mechanics;
• Overview and history of R;
• Overview of the planned topics
• R Markdown</td><td></td></tr>
<tr><td></td><td>Sep. 5</td><td rowspan="2">Topic 2: Data Types in R

• Vectors, lists, factors, matrices and arrays, data frames, etc.;
• Subsetting and assignment</td><td></td></tr>
<tr><td>2</td><td>Sep. 10</td><td></td></tr>
<tr><td></td><td>Sep. 12</td><td>Topic 3: Flow Controls

• Conditionals, loops, etc.</td><td></td></tr>
<tr><td>3</td><td>Sep. 17</td><td rowspan="3">Topic 4: Functions, Environments, and Scoping rules

• Functions as objects;
• Arguments (inputs) and return values (outputs);
• Named arguments and defaults;
• Environments;
• Scoping rules</td><td></td></tr>
<tr><td></td><td>Sep. 19</td><td></td></tr>
<tr><td>4</td><td>Sep. 24</td><td></td></tr>
<tr><td></td><td>Sep. 26</td><td rowspan="2">Topic 5: Loop functions and the split-apply-combine paradigm

• Abstractions of the split/apply/combine pattern
• Tools for split/apply/combine in base R</td><td></td></tr>
<tr><td>5</td><td>Oct. 1</td><td>Public holiday, no class</td></tr>
<tr><td></td><td>Oct. 3</td><td rowspan="2">Topic 6: Data Wrangling: A Tidy Approach

• tibbles: A modern take on data frames;
• %>%: A forward-pipe operator;</td><td></td></tr>
<tr><td>6</td><td>Oct. 8</td><td></td></tr>
<tr><td></td><td>Oct. 10</td><td></td><td></td></tr>
<tr><td>7</td><td>Oct. 15</td><td></td><td></td></tr>
</table>

	Oct. 17	• tidyr for tidying data. • dplyr for data manipulation	Public holiday, no class
8	Oct. 22	**Midterm Quiz (in-class)**	
	Oct. 24		
9	Oct. 29	Topic 7: R Base Plotting System • plot(): A generic plotting function; • The painter model and primitive graphics elements; • Other high-level graphics functions; • Multi-panel Layouts	
	Oct. 31	Topic 8: ggplot Plotting System • Grammar of graphics; • Data, geoms, and aesthetic mapping; • Optional features of ggplot();	
10	Nov. 5		
	Nov. 7	Topic 9: String Basics • Characters, strings, text data; • Base R functions for finding matches, splitting strings, and substituting according to patterns; • Regular expressions	
11	Nov. 12		
	Nov. 14	Topic 10: Web Scraping with R • Html and CSS basics • rvest for extracting tagged data • Dynamic Web Scraping	
12	Nov. 19		
	Nov. 21		
13	Nov. 26	Topic 11: Text analytics • Text analytics with tidy text format; • Sentiment analysis; • Tokenizing and relationships between words (N-grams and correlations)	
	Nov. 28		

10. Tentative Lab Schedule

Week	Date	Lab Topic	
1	Sep. 6	R and RStudio installation and R markdown syntax	
2	Sep. 13	Exploring data frames	
3	Sep. 20	Using control structures	
4	Sep. 27	Writing and calling functions	
5	Oct. 4	Practicing advanced looping and split-apply-combine manipulation	
6	Oct. 11	Data Wrangling with dplyr and tidyr I	
7	Oct. 18	Data Wrangling with dplyr and tidyr II	
8	Oct. 25	**No Lab**	Midterm Exam
9	Nov. 1	Plotting with R base graphics system	
10	Nov. 8	Plotting with ggplot2	
11	Nov. 15	Splitting and querying with regexes	
12	Nov. 22	Web scraping	
13	Nov. 29	Text Analytics	