ISS4066: Programming for Business Analytics (Undergraduate)

Jaewon Yoo

Assistant Professor
Institute of Service Science
National Tsing Hua University

Fall 2025

Course Duration: Sep. 2025 – Dec. 2025	*Classroom:* TSMC Bldg. R421
Class Time: Monday, 14:20 pm- 17:20 pm	*TA:* Pham Minh Phuong (Chloe)
Office number：(03) 5162543	*Dao Phuong Hoang Diem Uyen (Jessie)*
Instructor email: jaewon.yoo@iss.nthu.edu.tw	*John Loui Santiago*

COURSE DESCRIPTION:

This introductory undergraduate-level course leverages R to equip students with the fundamental skills required for robust business data analysis. The course is designed to address both practical programming skills and the conceptual understanding necessary to apply data science effectively in business contexts.

Specifically, at the end of the semester you should be able to:

- **Data Visualization and Wrangling**: Students will learn to summarize and visualize data, transforming messy data into tidy, analyzable formats.
- **Causality and Regression Analysis**: The course emphasizes evaluating claims about causality and using linear regression to conduct data analysis, crucial for drawing meaningful conclusions in business.
- **Statistical Uncertainty**: Understanding and quantifying uncertainty in data analysis are core components of the curriculum, enabling students to make informed decisions.
- **Professional Tools**: Mastery of professional tools such as R, RStudio, git, and GitHub will be developed, ensuring students are well-prepared for real-world data analytics tasks.

By the end of this course, students will not only be proficient in handling complex business datasets but will also appreciate the importance of meticulous data analysis. The goal is to inspire a passion for data analysis and foster a community among students to deepen their learning and enhance their collaborative skills.

This course welcomes students from any background who are interested in developing foundational data analysis skills applicable to the business sector.

RECOMMENDED TEXTBOOKS:

We will use the following books in this class. Students are expected to complete the assigned reading before joining class, which will assist in understanding the lecture materials and foster in-class discussions. Students can purchase either a hard copy or e-book version of the books.

Be aware that it might take time for the textbooks to be delivered so if you plan to take the course, please make sure to place your order early enough so the books arrive in time.

1. **[MD]** Ismay, Chester and Albert Y. Kim. 2022. Statistical Inference via Data Science: A ModernDive into R and the Tidyverse.
2. **[QSS]** Imai, Kosuke and Nora Webb Willaims. 2022. Quantitative Social Science: An Introduction with Tidyverse, 2022. Princeton University Press.
3. **[IMS]** Mine Cetinkaya-Rundel and Johanna Hardin. 2021. Introduction to Modern Statistics. OpenIntro.
4. **[AAG]** R for Everyone: Advanced Analytics and Graphics, 2nd Edition by Jared P. Lander, O'Reilly Media, 2017.
5. **[VT]** Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics by Nathan Yau, John Wiley & Sons, 2011.

SOFTWARE TOOLS:

Students will use the R statistical programming language to write and run the codes for the assignments and in-class exercises. R is an open-source language that can be used free of charge and can be downloaded from https://www.r-project.org/. For improved productivity and convenience, students will also use RStudio, a widely used integrated development environment (IDE) for R which is available at https://www.rstudio.com/products/rstudio/ for download. If you use a MacBook with the M1 chip/Apple silicon, RStudio might not work on your machine. If it does not, you can also use Visual Studio Code to edit your codes which can be downloaded at https://code.visualstudio.com/.

COURSE EVALUATION:

Grading Policy

Performance in this course is assessed through a blend of practical assignments, participation, and quizzes, designed to gauge your proficiency in understanding of the materials we will cover throughout the semester. Here are how each portion of course contributes to the overall grade:

Category	Percent of Final Grade
Final Project	25%
Assignments	20%
Problem Sets	20%
Quizzes & In-Class Exercises	15%
Attitude/Participation	10%
Attendance	10%

Course Requirements

1. **AI Use Policy*:** Students are permitted to utilize AI tools or Large Language Models (LLMs; e.g., ChatGPT, Claude 2, Bard) responsibly as aids for brainstorming and generating initial drafts for assignments. However, it is imperative that the final submission predominantly reflects the

student's understanding and personal input. To maintain academic integrity, students who choose to use AI tools must adhere to the following guidelines:

a. **Thorough Explanation:** Students must provide a detailed explanation of how the AI tool was used in the completion of the assignment.
b. **Original AI Responses:** Alongside their submission, students are required to submit the original responses generated by the AI tool.
c. **Critical Evaluation and Personal Input:** Students should critically evaluate the AI-generated content and clearly indicate the portions of the work that have been modified or expanded upon with their own insights and understanding.
d. **Proper Attribution:** Proper attribution should be given to the AI-generated content used, clearly indicating the sections that are AI-generated.

** Remember, attempting to cheat the system by heavily relying on AI-generated content without substantial personal input is ultimately the student's loss, as it undermines the learning process and personal growth.

2. **Academic Honesty and Plagiarism*:** All work submitted for academic evaluation must be the student's own. The penalty for violation of academic integrity will result in a zero for that assignment for the first time. Subsequent violation(s) will result in a failing grade for the course. Plagiarism will also not be tolerated.

   Academic dishonesty comprises of, but is not limited to, the following:

   - Cheating: Copying from other students' quizzes and assignments or allowing others to copy from one's own.
   - Plagiarism: Using other people's original work without giving appropriate credit or acknowledgment to the authors or sources.
   - Self-plagiarism: Submitting a piece of work in more than one course without the explicit permission of the instructors involved.
   - Misrepresentation of authorship: Submitting work as one's own, which has been prepared by or purchased from others.

   Students will be asked to upload their submission materials to Turnitin.com, an online plagiarism checker, to ensure academic integrity. Read more about online submission on http://learning.site.nthu.edu.tw/p/412-1319-7120.php?Lang=en.

3. **Attendance:** All students are expected to attend every class. Please bring your own hard copy of the course materials, which will be distributed by the instructor before class. If you have any urgent reason to miss a class, you are still responsible for the materials covered during the class and are expected to complete the required work. Attendance will be taken on a regular basis and will count towards your participation score (10%). Class missing will cause about 1% loss of final grade. Students who miss a class should inform me or the TA prior to the class via email or phone call.

4. **Final Project*:** This is an open-ended data analysis project on a topic of your interest. Whether you are fascinated by voting trends in presidential elections, racial disparities in policing, NBA salary distributions, or the box-office economics of the Marvel Cinematic Universe, the core steps should involve: 1) develop a clear research question; 2) locate and prepare data that can address that question; 3) apply the analytical techniques covered in the course to answer the question. You should be able to present your findings in a format suitable for a general audience.

The aim is to produce a polished, portfolio-ready project that demonstrates the skills you've acquired to prospective employers. Your deliverable will be a publicly accessible article or webpage that includes: a concise overview of the research question and dataset; a visualization that directly addresses the question; a table or figure presenting a regression model and a plain-language explanation of the results; and a brief paragraph discussing the study's limitations, threats to inference, and how the analysis could be improved (for instance, by collecting better data). Finally, the piece must link to a GitHub repository containing all code used to load, clean, analyze, and visualize the data and to generate the article.

The data collection and cleaning must be meaningful -- it's not sufficient to simply use a pre-cleaned data from an *R* package. Self-collected data is allowed and even encouraged, though beware you may need to undergo the Institutional Review Board (IRB) review process to be able to use the data for your final project.

Milestone	**Course Week**
Creating a GitHub Repository	4
Data and Proposal	6
First Visualization	9
First Analysis	11
Final Report	15

5. **Assignments:** There will be a total of FIVE assignments throughout the semester. The assignments will be relevant to our class topic and most likely will require you to run and report a statistical method we will cover in class. For instance, after we cover ordinary least squares (OLS) estimation, students will be asked to run a dataset (to be provided by the instructor) and report the results in a simple article format. Students will be required to write a brief report using R Markdown/Quarto for the assignment. Each assignment is worth 4% of the final grade. Assignments are to be submitted on the due date. *Late submissions will be penalized 1 percentage point of the assignment's weight per day* (e.g., for an assignment worth 4% of the course grade, turning it in 2 days late caps the attainable score at 2%).

6. **Problem Sets:** Throughout the semester, you will be required to complete problem sets that make up 20% of your final grade. These problem sets will consist of a mix of exercises taken from the recommended textbooks and are designed to reinforce your understanding of key concepts covered in class.

7. **Quizzes & in-class exercises:** To keep the learning momentum going, students will be asked to complete multiple quizzes about the course materials throughout the semester. In order to foster learning, students will also be given in-class exercises, some of which will be graded & will count towards the final letter grade.

8. **Attitude/participation:** In class, the most important thing for the students is to stay active and engaged about the topic being discussed. Positive contributions to class discussions will increase your score towards attitude. When we discuss a topic in class, effective discussions are only possible if everyone is well prepared. Please, be prepared to open and engage in discussions with your opinions and thoughts.

TENTATIVE WEEKLY SCHEDULE:

Week	Topics	Prepare before class	In-class deliverable
#1	Introduction to PBA: • Welcome and motivation: why learn programming & business analytics? • Overview of course details * Read AAG Chap 1 and 2 after class.	Read: Course syllabus	**Self-introduction** on Canvas general discussion board

#2	The nuts and bolts of R programming:
	• Review of programming structures
	• Different types of programming errors
	• Let's take a touR!
	• Operators for R programming
	• What can we do with R?
	
	Recap of Basic Statistics:
	• Probability and inference
	• Random variables (r.v.s)
	• Probability distributions
	• Expectations? Spread?
		Read:
		MD Ch 1
		AAG Ch 4.3, 5, and 6
		VT Ch 1 and 2
		
		Install *R* & *R Studio*
			Complete Swirl R Tutorials
			`1: Basic Building Blocks`
			`2: Workspace and Files`
			`3: Sequences of Numbers`

#3	Data Types and Structures: • Numeric, character/string, logical, dates & times • Vectors, Matrices, and arrays • Lists and data frames Data Visualization for Exploratory Analysis I: • Building plots by layers • Histograms and boxplots • Grouped data * In-class group exercises	Read: MD Ch 2 AAG Ch 3, 4, 7.1, 9, & 10 AAG Ch 27 and 28 VT Ch 3 and 4	Read a handout for '*Reporting Tools (R Markdown and Quarto)*'
#4	User-Defined Functions in R: • Creating & nesting fn. • Local vs. global var. • Scoping rules in R • Creating binary operators Data Visualization II • Principles of analytic graphics * In-class group exercises	Read: AAG Ch 7.2 and 11 VT Ch 5 and 6	**AS1** (The Nuts and Bolts of *R* Programming for Business Analytics) Complete group formation survey
#5	Data Wrangling: • Operating on rows • Operating on columns • Operating on groups • Creating barplots	Read: AAG Ch 14, 15, and 16 VT Ch 7 MD Ch 3	

#6	Causality: • What is it? • Randomized experiments • Calculating effects • Observational studies	Read: QSS Ch 2.1-2.5	
#7	Relationships and Importing/Tidying Data: • Z-Scores and standardization • Correlation • Pivoting data longer • Joining datasets	Read: AAG Ch 11 MD Ch 4 QSS Ch 3.5-3.6	**AS2** (**Visualization & Causality**: Air Pollution and Job Training Program)
#8	Prediction: • Predicting election outcomes • Loops • Evaluating the predictions • Time-series plot * Review: causal effect estimation for AS2	Read: AAG 4.7 and 12 VT Chap 8 and 9	
#9	Regression: • Prediction • Modeling with a line • Linear regression in R • Model fit	Read: MD Ch 5 QSS Ch 4.1 and 4.2.1-4.2.4	
#10	More on Regression: • Multiple regression • Categorical independent variables * In-class group exercises	Read: MD Ch 6.1-6.2 QSS Ch 4.2.6-4.3.2	**AS3** (**Regression & Prediction**: Ethnic Bias and Voter Turnout)
#11	Sampling and Sampling Distribution: • Sampling exercise and framework • Random var. and probability distributions • Normal variables and the central limit theorem	Read: MD Ch 7	
#12	Bootstrap and Confidence Intervals: • Resampling from our samples • Bootstrap CIs for a difference in means • Bootstrap CIs for a difference in ATEs • Computing and interpreting CIs	Read: AAG Chap 18 and 19 MD Ch 8 IMS Ch 12	
#13	Hypothesis Testing: • Hypothesis testing using *[infer]* • Two-sample tests • Two-sample permutation tests • Issues with hypothesis testing • Power analysis	Read: MD Ch 9 IMS Ch 11	**AS4** (**Bootstrapping**: Rags to Riches TV & Economic Mobility)
#14	Individual meeting by appointment. *(no in-class meeting)*		

#15	Models of Uncertainty: • Using the normal for inference • CI for experiments • Inference for linear regression	Read: IMS Ch 13	**AS5** (**Hypothesis Testing**)
#16	Inference for Linear Regression: • Uncertainty estimation for regression • Presenting OLS regressions • Wrapping up the class	Read: QSS Ch 7.3	

Notes: The course schedule is subject to change if necessary (**AAG**, **VT**, **MD**, **QSS**, and **IMS**, refer to 'Recommended Textbooks' in Page 1; **AS**: assignment). **Assignments are due at the beginning of the class on the submission date.**