ISS4066: Programming for Business Analytics (Undergraduate)

Jaewon Yoo

Assistant Professor
Institute of Service Science
National Tsing Hua University

Fall 2026

Course Duration: Sep. 2026 – Dec. 2026

Class Time: Monday, 14:20 pm- 17:20 pm

Office number：(03) 5162543

Instructor email: jaewon.yoo@iss.nthu.edu.tw

Classroom: TSMC Bldg. R421

TA: Anh Tran (anhtranpt@iss.nthu.edu.tw), Daffa Onielda (daffaonielda@gapp.nthu.edu.tw)

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. In practice this covers filtering, summarizing, grouping, joining, and reshaping data, and building clear exploratory graphics such as histograms, boxplots, and layered plots.

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. It distinguishes randomized experiments from observational studies and builds from simple to multiple linear regression, including categorical predictors and model fit.

Statistical Uncertainty: Understanding and quantifying uncertainty in data analysis are core components of the curriculum, enabling students to make informed decisions. This includes sampling distributions and the central limit theorem, the bootstrap and confidence intervals, hypothesis testing (including permutation tests and power analysis), and inference for regression.

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. You will also write your own functions in R and produce reproducible reports with R Markdown and Quarto.

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. Even as agentic AI now writes much of the code, the essential skill has shifted from writing it to directing and verifying it: without the underlying building blocks, you cannot tell when an analysis is wrong, and you end up steered by the tool rather than steering it. The course thus builds programming and visualization foundations first, then turns to statistical inference for business analytics.

This course welcomes students from any background who are interested in developing foundational data analysis skills applicable to the business sector. A graduate-level section (ISS5066) is also offered, with higher expectations for independent, research-grade work. Students may not receive credit for both sections.

READING AND CLASS PREPARATION:
Complete the Required reading in the weekly schedule before class and try the examples in R. These freely accessible readings introduce the concepts we will use in lectures and discussions. Arrive ready to work through examples and ask questions. In class, we will explain and revisit R fundamentals through real data analysis.

MD, IMS, and provided course materials support required preparation. Further reading offers additional explanation and depth; purchasing the commercial textbooks is not required. Follow the course website for direct links and the specified sections.

1. **[MD]** Ismay, Chester and Albert Y. Kim. 2022. Statistical Inference via Data Science: A ModernDive into R and the Tidyverse. Free online text.
2. **[IMS]** Cetinkaya-Rundel, Mine and Johanna Hardin. 2021. Introduction to Modern Statistics, first edition. OpenIntro. Free online text.
3. **[QSS]** Imai, Kosuke and Nora Webb Williams. 2022. Quantitative Social Science: An Introduction in tidyverse. Princeton University Press. Optional.
4. **[AAG]** Lander, Jared P. 2017. R for Everyone: Advanced Analytics and Graphics, 2nd ed. Addison-Wesley Professional. Optional.
5. **[VT]** Yau, Nathan. 2011. Visualize This: The FlowingData Guide to Design, Visualization, and Statistics. John Wiley & Sons. Optional.

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. You can also use Visual Studio Code to edit your code, 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 in-class exercises, 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	40%
Assignments	20%
Problem Sets	15%
Quizzes & In-Class Exercises	5%
Attitude/Participation	10%
Attendance	10%

Course Requirements

1. **AI Use Policy*:** Students are permitted to use AI tools, including large language models and agentic coding assistants (e.g., ChatGPT, Claude, Gemini), for assignments and the final project. You are fully responsible for everything you submit. The final project is designed so that a high-quality result requires genuine understanding; AI accelerates competent work but does not substitute for it. 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. In line with NTHU's Implementation Guidelines on AI Ethics Statements and Learning Assessment (September 2026): if AI use is identified but not disclosed as required, the work may be re-evaluated or assigned no credit; in-class exercises are completed without generative AI; when the instructor uses AI to produce teaching materials, this is indicated in the materials; and by enrolling in this course, students are considered to have agreed to this statement.

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*:** In lieu of in-class examinations (typically two midterms and a final), the final research project is the primary summative assessment of this course. It is completed in teams and includes a final in-class presentation with Q&A; a grading rubric will be provided. 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	Due
Creating a GitHub Repository	Sun Oct 11
Data and Proposal	Sun Oct 25
First Visualization	Sun Nov 15
First Analysis	Sun Dec 6
Final Report	Sun Dec 20
Showcase and Q&A	Mon Dec 21

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 three problem sets that make up 15% 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. **In-class exercises: To keep the learning momentum going, students will complete five short in-class exercises (1% each) during class, without generative AI, and upload them to eeclass the same day. Each exercise ends with a verification step: check what your code actually did before you trust the output. There are no take-home quizzes.**

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	**Preparation and reading**	In-class deliverable
#1 (Sep 7)	Course Introduction and R Programming Basics: • Course aims, organization, and expectations • Working in R and RStudio: objects, functions, and packages • Reproducible reports with Quarto	**Required reading:** MD §§1.1–1.4. Install R and RStudio before class. **Further (optional):** AAG §§4.3, 5, 6 VT Ch. 1–2.	Self-introduction on the eeclass discussion board by Fri Sep 11 Swirl tutorials 1–3 by Week 2

#2 (Sep 14)	Data Visualization: • Exploring college education and poverty in Midwest counties • Building plots layer by layer with ggplot2: geoms, labels, scales, and mapping versus setting aesthetics • Comparing distributions with histograms, density plots, boxplots, and facets • Plotting grouped data and trends over time with Gapminder	**Required reading:** MD §§1.2–1.4 (review) MD §§2.1–2.7. **Further (optional):** AAG §7.2 AAG Ch. 3–5 VT Ch. 3–4.	In-class exercise E1 (gapminder) AS1 released Due: Group formation survey, Sun Sep 20
#3 (Sep 21)	Data Wrangling with dplyr: • Operating on rows, columns, and groups • Bar plots	**Required reading:** MD §§3.1–3.6 MD §3.8 MD §2.8. **Further (optional):** AAG §§12.1–12.9 VT Ch. 7.	In-class exercise E2 GitHub setup guide posted
#4 (Sep 28)	No class (Teachers' Day): • Use the week for Assignment 1, the R basics reference slides, and swirl; TAs hold office hours as usual	No class meeting; no new assigned reading.	
#5 (Oct 5)	Causality: • What is a causal effect? • Randomized experiments; calculating effects • Observational studies	**Required reading:** IMS §§2.2–2.3 Course notes: potential outcomes, ATEs, and group means (posted before class). **Further (optional):** QSS §§2.1–2.5.	In-class exercise E3 20-minute debrief of E3 AS2 released Due: AS1 Due: Milestone 1: GitHub repository, Sun Oct 11
#6 (Oct 12)	Relationships, Importing, and Tidying Data: • Z-scores and correlation • Writing reusable functions for standardization • Pivoting longer; joining datasets	**Required reading:** MD §§4.1–4.2 MD §§3.7.1–3.7.3 MD §5.1.1 IMS §13.2.2. **Further (optional):** AAG §§6.1, 15.2–15.3, 18.2 QSS §§3.5–3.6.	Optional, Oct 12–23: book a slot via Google Calendar to discuss your research question before the proposal Due: Milestone 2: data and proposal, Sun Oct 25

#7 (Oct 19)	Prediction and Iteration: • Predicting election outcomes • Loops; evaluating predictions; time-series plots	**Required reading:** Course notes and R examples: prediction and iteration (posted before class) review MD §2.4. **Further (optional):** QSS §4.1 AAG Ch. 10 VT Ch. 4.	
#8 (Oct 26)	No class (Retrocession Day observed)	No class meeting; no new assigned reading.	
#9 (Nov 2)	Regression and Model Fit: • Modeling with a line; linear regression in R • Model fit	**Required reading:** MD §5.1 MD §5.3.2 IMS §7.2.5. **Further (optional):** QSS §§4.2.1–4.2.4, 4.2.6 IMS §§7.1–7.3.	AS3 and Problem Set 1 (IMS Ch. 7) released Course withdrawal period Nov 2–20 Due: AS2
#10 (Nov 9)	More on Regression: • Multiple regression • Categorical predictors	**Required reading:** MD §5.2 MD §§6.1–6.2. **Further (optional):** QSS §§4.2.6–4.3.2 MD §6.3.	In-class exercise E4 (NYC condominiums) Problem Set 2 (IMS Ch. 8) released Due: Problem Set 1, Sun Nov 15 Due: Milestone 3: first visualization, Sun Nov 15
#11 (Nov 16)	Sampling and Sampling Distributions: • Sampling framework • Random variables; the normal distribution and the central limit theorem	**Required reading:** MD §§7.1–7.3 MD §7.5. **Further (optional):** IMS §§13.1–13.3.	Due: AS3 Due: Problem Set 2, Sun Nov 22
#12 (Nov 23)	The Bootstrap and Confidence Intervals: • Resampling; bootstrap CIs for means and ATEs • Computing and interpreting CIs	**Required reading:** MD §§8.1–8.5 IMS §20.2. **Further (optional):** AAG §21.4 MD §8.6.	AS4 released

#13 (Nov 30)	Hypothesis Testing: • Hypothesis tests with infer; two-sample and permutation tests • Issues with hypothesis testing; power	**Required reading:** MD §§9.1–9.4 IMS §20.1 IMS §14.4. **Further (optional):** MD §§9.5–9.6.	In-class exercise E5 Problem Set 3 (IMS Ch. 11–12) released Team consultations Nov 30–Dec 4: book a slot via Google Calendar Due: Milestone 4: first analysis, Sun Dec 6
#14 (Dec 7)	Models of Uncertainty: • Normal approximation and standard errors for confidence intervals and hypothesis tests • Comparing groups: confidence intervals and two-sample tests for experiments	**Required reading:** IMS §§13.1–13.3 IMS §13.6 IMS §20.4. **Further (optional):** IMS §§13.4–13.5.	AS5 released Due: AS4 Due: Problem Set 3, Sun Dec 13
#15 (Dec 14)	Inference for Regression: • Uncertainty for regression coefficients • Presenting OLS regressions	**Required reading:** MD §§10.1–10.3 IMS §25.1. **Further (optional):** QSS §7.3 MD §10.4.	
#16 (Dec 21)	Project Showcase, Q&A, and Peer Evaluation: • Teams present at stations; the instructor and TAs rotate for Q&A • Peer evaluation	Review the project instructions and evaluation rubric on eeclass, and prepare your presentation and responses to questions.	Due: Final report, Sun Dec 20 Due: AS5, Sun Dec 27

Notes: The course schedule is subject to change if necessary. MD, IMS, AAG, QSS, and VT refer to the readings listed above; AS: assignment; E: in-class exercise. Required reading is preparation before class; Further reading is optional. Assignments are due at the beginning of class on the due date; Sunday deadlines are 23:59. Sep 28 and Oct 26 are university holidays with no class meeting.