

**Institute of Service Science, College of Technology Management,  
National Tsing Hua University**

**Course Syllabus  
Programming for Business Analytics (PBA)**

*Course Duration:* Feb 2023 – Jul 2023  
*Class Time:* Tuesday, 14:20 pm- 17:20 pm  
*Classroom:* TSMC Bldg. R430

*Instructor:* Jaewon Yoo ([jaewon.yoo@iss.nthu.edu.tw](mailto:jaewon.yoo@iss.nthu.edu.tw))  
*TA:* Daniel Gonzalez ([danielvinicioglo410@outlook.com](mailto:danielvinicioglo410@outlook.com))

**COURSE DESCRIPTION:**

This course introduces programming for business analytics using R and aims to cover a wide array of topics that are needed to utilize analytical tools to solve business problems. As an introductory course for business analytics, students will spend a substantial amount of time learning how to conduct exploratory data analysis by visualizing and examining data to detect interesting (model-free) patterns. Throughout the course, we will discuss topics that are related to visualization, descriptive statistics, basic inferential statistics (estimation and hypothesis testing, e.g., t-test and chi-squares), and regression analysis.

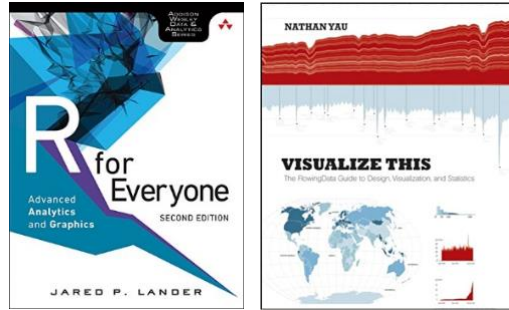
Students will also learn 1) how to navigate the R ecosystem (i.e., RStudio, CRAN, R Markdown), 2) import and preprocess data (“wrangling” needed to clean and prepare the data for analysis), 3) perform data analyses, and 4) produce reports to communicate the results and make business decisions based on empirical evidence. The emphasis is on familiarizing students to work proficiently with the statistical tool and preparing students to “wrestle” with real-world data in a quest to extract business insights. The door is open to students with any background who are willing to learn the fundamentals of empirical analysis for business.

**COURSE GOALS:**

- Learn the fundamentals to operate/navigate the R statistical programming language.
- Be familiar with data analytics basics.
- Understand how to identify and avoid common mistakes in metrics, visualizations, and interpretations.
- Be prepared for advanced business analytics (BA) topics.

**REQUIRED TEXTBOOK:**

1. **[TX1] R for Everyone: Advanced Analytics and Graphics, 2nd Edition** by Jared P. Lander, O'Reilly Media, 2017.
2. **[TX2] Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics** by Nathan Yau, John Wiley & Sons, 2011.



Students can choose to purchase either the hard copy or the e-book for both textbooks 1 and 2, but please be prepared to have the textbooks ready before the beginning of the semester. Students can order the hard copy textbooks directly at <https://bookdepository.com/> (provides free shipping worldwide):

- <https://www.bookdepository.com/R-for-Everyone-Jared-Lander/9780134546926> for **TX1** and
- <https://www.bookdepository.com/Visualize-This-Nathan-Yau/9780470944882> for **TX2**.

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.

### 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

Assignments	45%
In-class quizzes & Exercises	30%
Attitude/Participation	10%
Attendance	15%
<hr/>	
Total	100%

#### Course Requirements

1. **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>.

2. **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 (15%). 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.
3. **Assignments:** There will be a total of FOUR 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 two- to three-page long report about the given topics unless otherwise noted and submit it along with the R code script used for the assignment. Each assignment is worth 11.25% of the final grade. Assignments are to be submitted on the due date. *Late submission will incur a 1% loss to the grade each day.*
4. **In-class quizzes & exercises:** At the beginning of each class, students will be given a quiz about the course materials we would have discussed in the previous class. In order to foster learning, students will also be given in-class exercises, some of which will be graded & will count towards the letter grades.
5. **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	Class introduction and background: Basic concepts, terminologies, and history of R. An overview of the R ecosystem.  * Read <a href="#">TX1 Chap 1 and 2</a> after class.		
#2	The nuts and bolts of R programming: Basics of Programming & Intro. to R Operators, variables, and data types (e.g., levels of measurement, vector, matrix, array).	Read: <a href="#">TX1 Chap 4.3, 5, and 6</a> <a href="#">TX2 Chap 1 and 2</a>  Install <i>R</i> & <i>R Studio</i>	
#3	Syntax basics & Intro. to R cont.: Lists and data frames. Factor, table, string, date, and time.  In-class exercises.	Read: <a href="#">TX1 Chap 3, 4, 9, and 10</a> <a href="#">TX1 Chap 27 and 28</a> <a href="#">TX2 Chap 3</a>  Handout on Reporting Tools (R Markdown and Quarto)	
#4	Control structures & Loops in R: User-defined functions, loops (i.e., apply functions) conditional expressions (e.g., if/else, for, while).  Review Quiz #1 together!	Read: <a href="#">TX1 Chap 11</a> <a href="#">TX2 Chap 6</a>  Watch: Loops & -apply func. (50 min)	In-class quiz #1
#5	Data transformation I: Data munging using dplyr (e.g., selecting, appending, sorting, sampling, filtering, pivoting)  Installing and loading packages.	Read: <a href="#">TX1 Chap 14 and 15</a> <a href="#">TX2 Chap 7</a>  Watch: Data munging operations I (80 min)	<b>AS1</b> (Syntax operation)
#6	Data transformation II: Merging, binning, reshaping, date operations, and string parsing.	Read: <a href="#">TX1 Chap 16</a>  Watch: Data munging operations II (110 min)	

#7	Visualization I: Exploratory graphs, plotting systems, and types of plots.  Review Quiz #2 together!	Read: <a href="#">TX1 Chap 7.1</a> <a href="#">TX2 Chap 4</a>	In-class quiz #2
#8	Visualization II: Principles of analytic graphs and the application of ggplot2 package.	Read: <a href="#">TX1 Chap 7.2</a> <a href="#">TX2 Chap 5</a>	
#9	Visualization III: In-class exercises and Q&A session for AS2.	Read: <a href="#">AS2</a>	
#10	Statistics review and hypothesis testing: Review of basic statistics, T-test, and ANOVA.	Read: <a href="#">TX1 Chap 18 and 19</a>  Watch: Hypothesis testing using ANOVA (30 min)	<b>AS2</b> (Visualization)
#11	Estimation and dirty data problems I: Simple linear regression. Multiple linear regression.	Read: <a href="#">TX1 4.7 and 12</a> <a href="#">TX2 Chap 8 and 9</a>  Watch: Simple linear reg. (30 min) Multiple linear reg. (40 min)	
#12	Estimation and dirty data problems II: Model fits	Read: Handout on model fit	
#13	Estimation and dirty data problems III: Handling missing data (MCAR, MAR, MNAR).  Review Quiz #3 together!	Read: <a href="#">TX1 Chap 25.3</a>	In-class quiz #3
#14	Statistical methods for exploratory analysis I: Hierarchical clustering. K-Means clustering.	Read: <a href="#">TX1 Chap 25.1</a>	<b>AS3</b> (Statistical analysis & hypothesis testing)
#15	Individual group meetings (appointment-based)  (no class)	Watch: k-means clustering (40 min)	

#16	Statistical methods for exploratory analysis II: Business applications of clustering algorithms.  In-class exercises.	Read: TX1 Chap 25	Take-home quiz #4
#17	Individual meeting/Q&A Session for AS4.  End-of-Semester Lunch!	Read: <u>AS4</u>	
#18	Submit Assignment 4: EDA and Clustering  (no class)		<b>AS4</b> (Clustering)

*Notes:* The course schedule is subject to change if necessary (**TX1:** textbook 1, **TX2:** textbook 2, **AS:** assignment). **Assignments are due at the beginning of the class on the submission date.**