

# MGS 3701 / MKT 3950: Data Mining

## Software / Rmarkdown

---

Chungil Chae



2022/01/01 (updated: 2022-02-17)



# Required Software in Class

R (<https://cran.r-project.org/>)

- Rstudio (<https://www.rstudio.com/products/rstudio/download/>)
  - tidyverse (r package)
  - psych (r package)
  - ggplot2 (r package)
  - tidytex (r package)
- Python (<https://www.python.org/downloads/>)
  - numpy (python package)
  - pandas (python package)
  - dash (python package)
- Visual Studio Code
  - (<https://code.visualstudio.com/download>)
- Tableau (Student licence:
  - <https://www.tableau.com/academic/students>)

In this course, I am not going to teach how to install programs. Program installation and fixing error is your responsibility.



# Environment

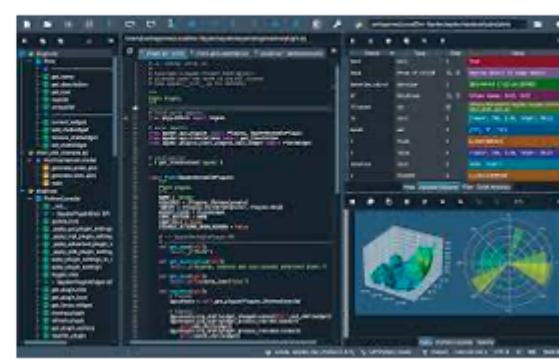
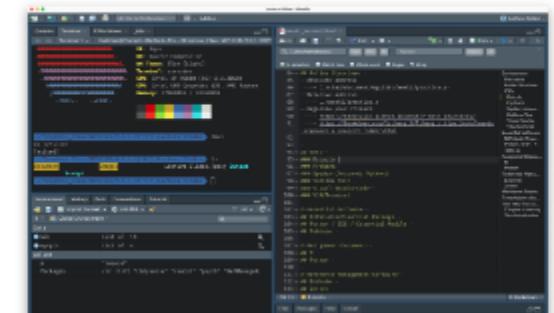
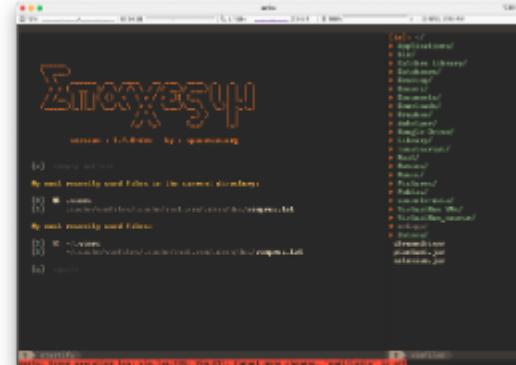
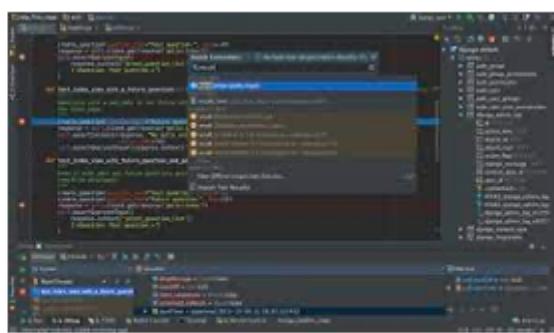
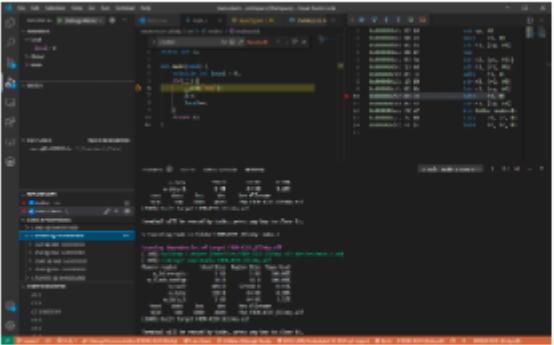
## File name

- No white space
  - using hyphen(-) or underbar(\_)
- File name
  - ID\_Chinese Name in Eng\_title.extension
  - 1111111\_ChungilChae\_lab1\_practice.rmd

## Folder Structure

- Absolute address
  - C:/chad/document/mgs3101/week1/practice.r
- Relative address
  - ../week1/practice.r
- Organize your folders
  - [https://kdestasio.github.io/post/r\\_best\\_practices/](https://kdestasio.github.io/post/r_best_practices/)
  - [https://bookdown.org/lyzhang10/lzhang\\_r\\_tips\\_book/how-to-organize-a-project-folder.html](https://bookdown.org/lyzhang10/lzhang_r_tips_book/how-to-organize-a-project-folder.html)

# IDEs





# Essential Software

## R/Rstudio/Essential Package

RStudio is an Integrated Development Environment for R, a programming language for statistical computing and graphics. It is available in two formats: RStudio Desktop is a regular desktop application while RStudio Server runs on a remote server and allows accessing RStudio using a web browser.

## Python / IDE / Essential Module

Python is an interpreted high-level general-purpose programming language. Its design philosophy emphasizes code readability with its use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects

## Tableau

Tableau Software is an American interactive data visualization software company focused on business intelligence.



# Markdown Basics

Markdown is a lightweight markup language for creating formatted text using a plain-text editor. John Gruber and Aaron Swartz created Markdown in 2004 as a markup language that is appealing to human readers in its source code form. Markdown is widely used in blogging, instant messaging, online forums, collaborative software, documentation pages, and readme files. The initial description of Markdown contained ambiguities and raised unanswered questions. To correct these problems, later implementations introduced subtle differences from the original version as well as syntax extensions.

<https://markdownlivepreview.com/>



## R markdown cheatsheet

(<https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>)

## Jupyter notebook cheatsheet

([https://www.edureka.co/blog/wp-content/uploads/2018/10/Jupyter\\_Notebook\\_CheatSheet\\_Edureka.pdf](https://www.edureka.co/blog/wp-content/uploads/2018/10/Jupyter_Notebook_CheatSheet_Edureka.pdf))



# Demonstration

- How to create project in R
- How to install r packages
- How to create a document in
- how to install python module
- Practice basics of markdown



# Chapter Video

<https://www.youtube.com/watch?v=DNS7i2m4sB0&list=TLPOQMTcwMjIwMjLxhVmSmWxNQ&index=2>

R Markdown with RStudio





# References

Shmueli, G., P. C. Bruce, P. Gedeck, and N. R. Patel (2019). *Data mining for business analytics: concepts, techniques and applications in Python*. John Wiley & Sons.

Shmueli, G., P. C. Bruce, I. Yahav, N. R. Patel, and K. C. Lichtendahl Jr (2017). *Data mining for business analytics: concepts, techniques, and applications in R*. John Wiley & Sons.



# Assignment

## Introducing your self

- To do
  - 1 page self introduction(single space)
  - Add your picture (your face should be clearly shown)
- Requirement
  - Picture of your face
  - Your student ID
  - Your name (Chinese name in eng and Eng name)
  - Your introduction
- Format
  - id\_name\_assignment1.pdf
    - 111111\_Chungil\_Chae\_assignment1.pdf
  - PDF
- Submission
  - Class LMS
- Due: Feb 20 (Sun), 11:58pm(Bejing Time)