HANYANG UNIVERSITY

SOC3035: DATA ANALYSIS

Sam Hyun Yoo

Fall 2022

E-mail: samyoo@hanyang.ac.kr
Office: 408 Social Sciences
Office Hours: Wed 4~5pm or by appointment
Department of Sociology
Class Hours: Wed 1~4 pm
Class Room: 303 Social Sciences

E-mail: samyoo@hanyang.ac.kr
Office: 408 Social Sciences
Office Hours: Wed 4~5pm or by appointment
Department of Sociology
Class Hours: Wed 1~4 pm
Class Room: 303 Social Sciences

Course Overview

Course description

This course is an introduction to the field of data analysis and visualization in sociology. It will guide you how to manage, analyze, and visualize quantitative data from diverse topics in sociology. This course is designed to learn the basic knowledge and usages of statistical software, such as Excel and R, in data wrangling, data analysis, and visualization. Basic computer skills are essential to successfully complete this course, though intermediate or advanced computer skills are preferred.

Required prerequisite: CUL2004 *Invitation to Sociology* & SOC2002 *Social Statistics* or equivalent; Students are required to have taken both CUL2004 and SOC2002 prior to enrolling this course. Otherwise, please contact the instructor before registering.

Course objectives

Upon successful completion of this course, students will

- Understand how to manage, analyze, and interpret quantitative data.
- Have practical knowledge and senses of how to present quantitative data through effective graphs and figures.
- Gain on hands-on skills and experience with *R* and *RStudio* (especially *tidyverse* packages).

Textbooks & Materials

This course uses several textbooks, online resources, and lecture slides. As R is an open-source software, most of textbooks and manuals are freely available online. Required textbooks and additional resources are listed as below.

Required textbooks

The first three books [1, 2, 3] are about data analysis and visualization using R while the next two books [4, 5] explains fundamental knowledge on data visualization. The last one [6] explains how to install git and connect it with Rstudio.

1. Wickham, Hadley & Garrett Grolemund. 2017. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Sebastopol, California: O'Reilly Media. (http://r4ds.had.co.nz/)
2. Healy, Kieran. 2019. *Data Visualization: A Practical Introduction*. Princeton: Princeton University Press. (http://socviz.co/)
3. Ismay, Chester & Albert Y. Kim. 2016. *Statistical Inference via Data Science: A Modern Dive into R and the Tidyverse*. CRC Press. (http://moderndive.com).
4. Wilke, Claus O. 2016. Fundamentals of Data Visualization. O'Reilly Media. (https://serialmentor.com/dataviz/).
5. Camores, Jorge. 2016. *Data at Work: Best Practices for Creating Effective Charts and Information Graphics in Microsoft Excel*. New Riders.
6. Bryan, Jennifer. 2016. *Happy Git and GitHub for the UseR*.(http://happygitwithr.com)

Additional books and resources

7. Grolemund, Garrett. 2022. *Hands-on programming with R: Write your own functions and simulations*. O'Reilly Media, Inc. (https://rstudio-education.github.io/hopr/)
8. Chang, Winston. 2018. *R Graphics Cookbook: Practical Recipes for Visualizing Data*. O'Reilly Media. (https://r-graphics.org/)
9. Xie, Y., J. J. Allaire, Garrett Grolemund, 2022. R markdown: The definitive guide. Chapman and Hall/CRC.(https://bookdown.org/yihui/rmarkdown/)
10. Data Carpentry. *Data Carpentry R for data analysis for Ecology* translated in Korean (http://aispiration.com/R-ecology-lesson/#fn:r-ecology) This is a translated version of R basics for those who are more comfortable in Korean than in English
11. Robinson, David. *Datacamp Course: Introduction to the Tidyverse*. (https://www.datacamp.com/courses/introduction-to-the-tidyverse) This is an interactive website for R programing and data science. I am checking the possibility of classroom registration, which allows you to use the website free for six months.

Lecture slides

The instructor also provides and uses lecture slides and additional materials as needed. Lecture slides will be posted on the *HY-ON* course website 3~5 days before each lecture. Lecture slides are based on the textbooks and other materials and may include potential errors and typos. Please let me know if you find any.

Please review the required readings and slides in advance. As this course involves in computing and statistical thinking, students who are not familiar with these areas may have difficulty in following course contents. I strongly recommend you study and practice in advance.

Laptop & Software

Laptop is required for this course. Please bring your own laptop (*MacOS* or *Windows*) to this course. This course includes computational practices using *R* and *RStudio*, as well as *MSExcel*. *R* and *RStudio* are free, open-source software programs and available online. It is also necessary to download, install, and update *R* packages. There are many PCs in computer labs across campus available to students, but installing new software and packages on the PCs is often restricted by denying administrative privileges to regular users. Therefore, it is essential for students to bring their own laptop to every class meeting.

Details on how to download and install *R* and *RStudio* will be provided in the first week of the semester – this information is also available in the textbook listed above and online. Please note that *R* and *RStudio* can be installed on any operating system including *MacOS* and *Linux*. However, I primarily use *Windows*. Although I have been learning *MacOS* recently, I may not be able to offer practical support for *MacOS* and *Linux* users.

Course Requirements & Grading

Demanding course schedule

This course includes computer-based exercises, which require statistical knowledge and *R* programming language. It is imperative and crucial that you stay focused, organized, and diligent in keeping up with the materials, assignments, and exams.

If you have never used *R*, *tidyverse*, or *ggplot*2 before, you may need to spend more time on learning it. Please check and review one of the core textbooks listed above, as well as course materials. The one written by *Wickham and Grolemund* (2017) is especially useful for learning *R* packages (*tidyverse* and *ggplot*2) required for this course. Please note that *Camores* (2016) is about data visualization using Excel and not related to *R* and *RStudio*. There are many useful websites on Data Science using R as well. Please check online first; most of your questions can be found.

If you are not confident at social statistics and not familiar with computing skills, you may have difficulty following the main contents of this course. Please find colleagues who can help you and work together. Otherwise, I recommend you take other stat courses or review relevant books first and take this course next time.

Required readings, lecture slides, & lectures

Lecture slides and course contents are based on the core textbooks and other resources. Again, most textbooks on *R* and *Rstudio* are freely available online and provide enough explanation with helpful examples. Although lecture slides are designed to help you understand major concepts and how to handle data using *R* and *RStudio*, those include limited information only. Therefore, please read the textbooks first and review slides. Then, it will be much easier to follow lectures in a classroom. Please remember that students are responsible for reading assigned chapters each week.

Course assessment & grading

Your grade will be based on course attendance (10%), class participation (10%), in-class & home assignments (50%), a final presentation and report (30%). Possible grades are as follow: A+, A, B+, B, C+, C, D+, D, & F.

Table 1: Course requirements and grading

Type	Content	Points
Course attendance	One point deduction for each absence	10
Class participation	Qualitative evaluaton (average 7/10)	10
Assignments	In-class & home assignments (3~5) including mid-term	50
Final presentation & report	Submit a report after a presentation	30
Total		100

Note: This is are subject to change. Any change will be notified in class.

Course attendance

Course attendance is required. Learning data analysis and visualization requires you to gradually but constantly build your knowledge and skills. Therefore, it is difficult to catch up once you get behind. Any unexcused absence will be penalized with one point deduction. Arriving late or leaving early without permission will count as 1/2 unexcused absence.

Absences are only 'excused' if they are one of the types on the university's list of acceptable excuses, such as injury, illness, death in family, etc. Students seeking for excused absence should contact the administrative office of your home department/college first and submit the Excused Absence Request Form. Once the requested form is approved, contact your instructor with a photo copy of the approved document. Please check the details with the administrative office.

Note that according to the university policy, students who do not attend 2/3 of the class or more in a single semester will not be able to receive the credit and will be graded "F" for the course.

Class participation

You are expected to actively engage in learning activities in this course. This include extra efforts in learning and involvement in both face-to-face and online activities, such as discussions, debates, group work, and helping other students. To encourage and promote academic atmosphere, your attitudes, efforts, collaboration, and preparation are also taken into account. Please do not assume that you will get 10/10 points for class participation; average score is usually around 7~8. Criteria for class participation includes but not limited to the following:

- comes to class prepared
- respect instructor and other students
- listens attentively and take notes
- contributes to conversation and discussion
- provides constructive feedback
- shows interest in and respect for others' views and work;
- participates actively in small groups.
- present or submit an excellent work

Assignments

This course will have several in-class and homework assignments throughout the semester. These assignments aim to help you understand statistical concepts, data analysis and visualization in

R. Each assignment will be explained in class and also will be posted on *HY-ON* (a LMS website for this course). Of course, your assignments should be submitted via *HY-ON* (a course website, LMS) unless instructed to do otherwise. Please note that late assignment will not be accepted.

Depending on circumstances, there might be a mid-term assignment, which requires more work than a regular one – isn't it better than a mid-term exam? Details will be discussed and provided in class.

Final presentation & report

To promote a sense of data analysis and visualization in a real-world context, I will ask students to acquire quantitative data of interest and plan, carry out and report on a simple project involving techniques discussed in this course. Depending on class size, final report can be carried out as an individual or group project.

Your research topics and data can be discussed in class. However, quantitative data should be relevant to the sub-fields of sociology. Findings from the projects will be presented in a class a couple of weeks before the end of semester. Final report is due a week after the presentation. Details will be discussed in a class.

Presentation: Each of you will give a presentation based on your project to the class as if at an academic conference. You will speak for 8 to 12 minutes, depending on class size. Please be prepare for the presentation with tidy slides. Your slides should include research purpose, backgrounds, data & methods, and results – mostly figures.

Final report: Based on your research, you should submit a final report via *HY-ON* course site. Your final report should be a reproducible version of a quantitative research. For your paper, you will:

- write using RMarkdown (or `knitr`) with interspersed R code and text,
- begin with raw data collection,
- include all steps of data cleaning and analysis,
- present figures rather than tables,
- use BibTEXfor citations,
- automatically generate PDF, html, and ($^{eee}w_{www_w}$!) Word versions,
- do your work in a GitHub repository,
- submit a link to that repo.

Papers are due the last day of exam week, Week 16, Dec 21 before midnight.

Student Academic Conference

Department of Sociology at *Hanyang University* holds the *Student Academic Conference*(SAC) at the end of each semester in which undergraduate students conducting independent and group works for credit can present their research findings as from of oral presentations.

The SAC will provide an invaluable opportunity for you to share your findings with other students and faculty members.It will help you deepen your understanding of core themes of sociology and your research. Participating in such an event will be helpful for your professional development with valuable experience. Scholarships may be also awarded for selected participants.

Course Policy

Academic integrity & dishonesty

Academic honesty is expected of all students in all examinations, papers, laboratory work, academic transactions and records. The possible sanctions include, but are not limited to, appropriate grade penalties, and course failure.

1. Students need to attend more than two thirds (2/3) of the total number of classes in order to take the final exam.
2. When the act of cheating is discovered, received credit will be cancelled in accordance with the school rules or internal regulations.

Course guide for disabled students

Major contents of teaching support services for students with disabilities.

1. visual impairments: prior class registration support, help to move, writing (note takers)-typing help, exemption from English classes, extension of test time, enlarged textbook.
2. physical disability: prior class registration support, help to move, writing (note takers)-typing help, extension of test time (upper body disorder).
3. hearing impairments: prior class registration support, writing (note takers)-typing help, exemption from English classes (except for balance disorder).
4. developmental and intellectual disabilities: prior class registration support, writing (note takers)-typing help.
5. Other disabled students: prior class registration support, writing (note takers)-typing help.
6. Other matters: Test time can be extended according to the professor's judgment regardless of type of disability.

Students who take this course can get appropriate level of support service including the support listed above depending on the students' individual characteristics and needs through consultation with professors and the support center for students with disabilities. If you have any questions concerning support service for students with disabilities, you can contact professor or support center for students with disabilities

Support Center for Students with Disabilities (Seoul) 02-2220-0776, (ERICA) 031-400-4502

Policy on missed deadlines

It is your responsibility to be aware of timelines and deadlines in this course. Always make sure to give yourself plenty of time to complete your assignments given your work and personal schedules.

- Extensions for assignments and exams will not be allowed.
- Last minute work or personal conflicts are not justification for missed assignments.
- Having computer/internet problems or limited access to an alternate computer will not be acceptable justifications for missing a deadline.
- Please contact your instructor if you have any questions on these issues.

Mutual respect, professionalism, & courtesy.

You are expected to treat your instructor and all other students in the course with courtesy and respect. The topics in social sciences, especially in sociology, are pretty diverse from class, gender,

politics, and religion, to social inequality and irregularities. You are encouraged to actively engage in debates and discussion, but your comments should be factual, constructive, and free from harassing statements, based upon facts and documentation. I ask for your cooperation in keeping an atmosphere of mutual respect in the classroom.

At the same time, it is disruptive and inappropriate if you do surf the web, text messages, or focus on other work during during class meeting. I may exercise the right to ask students to leave if they engage in one of these manners.

Course Schedule

Week 1, Sep 07 : Course Overview

Readings: Wickham (Ch 1), Healy (Ch 2), Ismay & Kim (Ch 1) (optional) Bryan (Ch 1-16)

Assignment:

- Install *R* and *RStudio* on your machine and submit proof
- (optional) Explore the Introduciton to Tidyverse at DataCamp website

Week 2, Sep 14 : Introduction to R and RStudio

Readings: Wickham (Ch 1), Healy (Ch 2), Ismay & Kim (Ch 1)

Week 3, Sep 21 : R Basic I

Readings: Wickham (Ch 1)

Week 4, Sep 28 : R Basic II

Readings: Wickham (Ch 1)

Week 5, Oct 05 : Data Visualization

Readings: Camores (Ch 1-5), Healy (Ch 1)

Assignment:

- Find bad and good examples of data visualization from the sources (magazines, news articles, reports, research papers, etc.) that published since Jan 1, 2022

Week 6, Oct 12 : Visualizing Data with R I

Readings: Wickham (Ch 2-3), Healy (Ch 3-4), Ismay & Kim (Ch 2)

Assignment:

- Visualizing data with built-in data; Otherwise getting certificate from DataCamp (TBD)

Week 7, Oct 19 : Data Import in R

Readings: Wickham (Ch 11), Ismay & Kim (Ch 4)

Week 8, Oct 26 : Tidy Data I

Readings: Wickham (Ch 5), Healy (Ch 5), Ismay & Kim (Ch 3)

Assignment:

- Research proposal with quantitative data; provide detailed explanation for data and methods

Week 9, Nov 02 : Tidy Data II

Readings: Wickham (Ch 9-16), Ismay & Kim (Ch 4)

Week 10, Nov 09 : Visualizing Data with R II

Readings: Wickham (Ch 2-3, 7), Wilke (Ch 6-10)

Assignment:

- Visualizing data with diverse approaches

Week 11, Nov 16 : Descriptive Statistics

Readings: Wickham (Ch 7)

Week 12, Nov 23 : Statistical Analysis: Comparing Means

Readings: Wickham (Ch 7)

Week 13, Nov 30 : Statistical Aanalysis: Association & Correlation

Readings: Wickham (Ch 7), Ismay & Kim (Ch 5)

Week 14, Dec 07 : Statistical Analysis: Regression

Readings: Ismay & Kim (Ch 5-6)

Assignment:

- Presentation

Week 15, Dec 14 : R Applications

Readings: Wickham (Ch 26-30)

Assignment:

- Presentation cont. (if necessary)
- Final report due (TBD)

Week 16, Dec 21 : Summary and Q&A

Readings: additional resources

Assignment:

- (optional) Explore R Markdown, Bookdown, Blogdown, Shiny apps, and GIS etc.

Note: This course schedule is subject to change. Any change will be notified in class.