

# POL4835 Applied Data Analysis

Mondays 11:00 - 11:50 AM  
Wednesdays 11:00 AM - 12:50 PM  
Yonhee 402

*Last updated: March 6, 2023*

## Instructor Information

Jeong Hyun Kim, Ph.D.  
Assistant Professor, Department of Political Science & International Studies, Yonsei University  
Office: Yonhee Hall 115  
E-mail: [jhkim1@yonsei.ac.kr](mailto:jhkim1@yonsei.ac.kr)  
Office Hours: Mondays 3-5 PM.  
You can book an appointment here: <https://calendly.com/jhkim1/officehour>

## Course Description

This course is designed to introduce students to the methods and practices by which they can use data to answer questions in political science. The course will focus on helping students to understand the core concepts behind statistical inference and learn how to apply them appropriately to answering substantive problems of interest. In addition, a major component of the course includes learning how to collect, manage, and analyze data using computer software and how to effectively communicate results to others.

Each week will consist of a lecture and in-class activities. Both lectures and in-class activities will be based on R, an open-source software that is widely used for data analysis in social science.

## Course Objectives

By the end of the semester, students should be able to:

- Understand the core concepts in statistical procedures and causal inference.
- Learn how to explore and manage data and conduct statistical analysis.
- Understand core programming concepts in the R language and apply these concepts to complete in-class activities and problem sets.
- Present data and results in a clear and informative manner.
- Independently gather, analyze, interpret, and present your own data.

## Course Materials

There are two textbooks for this course:

- R for Data Science. Hadley Wickham and Garrett Golemund. (R4DS) A free online copy is available here: <https://r4ds.had.co.nz/>
- The Fundamentals of Political Science Research (2nd Edition). 2013. Kellstedt, Paul & Guy Whitten. Cambridge University Press (KW)  
This book is available for purchase at university bookstore. I also have reserved this book in the university's main library for your use.

## Statistical Software

In this course, we use the open-source statistical software R (<http://www.r-project.org>). This package is widely used in various fields of social sciences and statistics. It is also used in many parts of industry, policy, data journalism, and more. R is available for every computing platform, and most importantly, is free. Note that R is much more powerful than other statistical software such as SPSS, STATA and SAS, but it's more difficult to learn. However, it is much more similar in structure to other programming languages and therefore also serves as an introduction to basic programming tasks.

If you find it difficult to install R or R studio, you can use the free [RStudio.cloud](#) service, which lets you run a full instance of R Studio in your web browser. Using this, you can run R without installing anything on your computer.

## Teaching Assistant

Ji Young Park

MA Student, Department of Political Science and International Studies, Yonsei University

E-mail: [tktkclfdl44@naver.com](mailto:tktkclfdl44@naver.com)

## Requirements and Evaluation

- Attendance (10%) Since attendance is crucial to the success in this course, I will track your attendance each class. Attendance starts at 100 points. You are allowed two unexcused absences. After that, missing class without a valid excuse is -3 points. If you are going to miss class due to university or professional events, you must notify TA and me as soon as possible (at least one day prior to the day of your absence). Examples of valid excuses include (but not limited to): serious illness, military training, job interviews, and family emergencies.
- Problem Sets (30%) There will be four problem sets during the semester. Each problem set will count equally towards the final course grade. Collaboration is permitted (and encouraged), but students must write up the code and answers on their own. You will need to turn them in on LearnUs by the specified due date before the beginning of class.

- Midterm (20%) The midterm is scheduled for April 26th during our normal class time. It will take place in person.
- Final (40%) The final exam is scheduled for June 21. It will be in person and you will need to bring your laptop.

## Course Policies

- You are required to complete readings before coming to each class.
- You will frequently make use of computers in this course, during some lecture periods and during every lab. Please be respectful to your instructor and your peers by using your computers only for class-related purposes. Also, please put your phone away before class starts and don't bring it out.
- Because you know the deadline so far in advance, late submissions will result in 0 points. Extensions will be granted only under exceptional circumstances. In each case, I require proof (e.g., a written note from a doctor).
- The best way to communicate with me outside the classroom is via e-mail. I will respond to your emails as soon as I can. For questions related to your attendance, please email your TA and copy me in that email.

## Note on Problem Sets

Problem sets for this course present opportunities for students to discuss questions and collaborate to find a solution together. At the same time, there is a clear distinction between permissible collaboration and unacceptable plagiarism.

Programming necessitates that you reach your own understanding of the problem and discover a path to its solution. During this time, discussions with other people are permitted and encouraged. However, when the time comes to write code that solves the problem, such discussions are no longer appropriate: the code must be your own work.

## Course Schedule (Subject to Change)

- Week 1
  - 3/6: Course Introduction
  - 3/8: Introduction to R, R Studio, and R Markdown. (Reading: R4DS Ch. 20 & Ch. 27)
- Week 2 (3/13 & 3/15)
  - Lecture: Research Design (Reading: KW Ch.2)
  - Lab: Data Management. (Reading: R4DS Ch. 5, Ch. 10-11)
- Week 3 (3/20 & 3/22)

- Lecture: Planning Data Analysis (Reading: KW Ch. 4)
- Lab: Data wrangling in tidyverse. (Reading R4DS Ch. 14-15)
- **Problem Set 1 Assigned** (Due: 3/29)
- Week 4 (3/27 & 3/29)
  - Lecture: Descriptive Statistics (Reading: KW Ch. 5)
  - Lab: Summarizing data.
- Week 5 (4/3 & 4/5)
  - Data Visualization and Exploratory Data Analysis (Reading: R4DS Ch. 3 & Ch. 7)
  - **Problem Set 2 Assigned** (Due: 4/12)
- Week 6 (4/10 & 4/12)
  - Lecture: Probability and Sampling (Reading: KW Ch. 6)
  - Lab: Writing functions and loop (Reading: R4DS Ch. 19 & Ch. 21). Sampling.
- Week 7 (4/17 & 4/19) Confidence Intervals (Reading: KW Ch. 6)
- Week 8
  - 4/24: Midterm Review
  - 4/26: Midterm Exam
- Week 9 (5/1 & 5/3)
  - Hypothesis Testing (Reading KW Ch. 7)
  - **Problem Set 3 Assigned** (Due: 5/10)
- Week 10 (5/8 & 5/10) Causality and Experimental Analysis (Reading: KW Ch. 3; Additional reading TBA)
- Week 11(5/15 & 5/17)
  - Linear Regression I (Reading: KW Ch. 8)
  - **Problem Set 4 Assigned** (Due: 5/24)
- Week 12 (5/22 & 5/24) Linear Regression II (Reading: KW Ch. 9)
- Week 13 (5/29 & 5/31) Basic Text Analysis (Reading: TBA)
- Week 14 (6/5 & 6/7) Case Studies & Review
- Week 15 (6/12 & 6/14) Self-Study
- Week 16 (6/19 & 6/21) Final Exam