

# CSSS/POLS 510 MLE Lab

## 1. Logistics and R Refresher

Inhwan Ko

Oct 1, 2021

# About TA

## ① Research interests

- Political economy of clean energy transition ([interactive map](#))
- [Climate policy](#), [negotiation](#), and [climate clubs](#)
- Text-as-data (in [Korean](#) and English), time-series analysis, mixed methods

## ② Current projects

- Cross-country case studies on renewable energy (land-use) conflicts (causes and consequences)
- Determinants of “the quality of net-zero emissions”
- Climate clubs as inter-national AND inter-firm organizations

# Disclaimer

The current version of the lab materials is adapted from those drafted by our two previous TAs for this course, Brian Leung and Kenya Amano.

# Logistics

- 1 **Lab Sessions:** Fridays, 3:30 - 5:20pm via Zoom ([link](#))
  - Reviews lecture materials using examples; Q & A for assignments and lectures
  - Materials will be available on the [course website](#)
  - Always look for these files: lab1\_slide.pdf, lab1\_practice.rmd, lab1\_key.rmd, and lab1\_data.csv
- 2 **Office Hours:** Thursdays, 1:00 - 3:00 pm, or by appointment
  - Trouble-shooting, questions about the lecture and assignments, etc.
  - Please make an appointment for in-person meeting
  - Or simply walk-in using Zoom ([link](#))

# Logistics (Cont.)

- ③ **Homework:** 5-6 due every 2 weeks or so
  - Must be typed up (not hand-written)
  - Ideally, done using R or R Studio with write up in  $\text{\LaTeX}$
  - Using R Studio with R Markdown is an easy way to do this (Will work on this next week)
  - We will use two of Chris's packages extensively: `simcf` and `tile`
- ④ Please make sure you name the assignment files properly as follow:
  - Assignments to me: `CSSS510HW1_InhwanKo.pdf`
  - Assignments to Chris: Make sure to include all your co-authors' names in the file name

# Logistics - Goals

- 1 When this course is over, you should be able to do the following (and much more):
  - ▶ Identify the proper distribution and model for your data (logistic, ordered, multinomial, count)
  - ▶ Run the model using both the glm function and “by hand” using optim, extract parameters of interest, and interpret these in probabilities
  - ▶ Compute predicted probabilities and use simulation to find the confidence intervals of  $\hat{\pi}_i$  across counterfactuals values of covariates  $\mathbf{x}_i$
  - ▶ Use cross-validation to assess the predictive accuracy of several models and also compare these models across a variety of in-sample goodness of fit tests
  - ▶ Use one of several algorithms to impute missing data

- 1 **The stuff in R:** For the homework assignments and project you will need to feel comfortable
  - ▶ importing (and exporting) data sets
  - ▶ tidying and transforming data
  - ▶ analyzing data (conceptual part of the course)
  - ▶ generating plots of your data and results
  - ▶ writing basic functions and loops for repeated procedures

- ② I have to read lots of your code. Please be considerate when writing code and submitting assignments.
- ▶ Do not print unnecessary code and output. Learn how to use `results = "hide"` and `echo = TRUE` in R Markdown.
  - ▶ Name well
    - ★ functions vs. all other objects
    - ★ readability is about consistency (`dot.naming`, `CamelCaseNaming`, `pothole_naming`)
    - ★ short, clear, consistent – help future you (and present me)



# Logistics - R

- ② I have to read lots of your code. Please be considerate when writing code and submitting assignments.
  - ▶ Specify arguments fully, e.g.

```
rbinom(n = 1000, size = 30, prob = 0.49) # GOOD!
```

```
rbinom(1000, 30, 0.49) # LESS GOOD!
```

- See the Google R styleguide for an example

# Logistics - R Useful resources

- R
  - ▶ *R for Data Science* (Golemund and Wickham 2016)
  - ▶ *Quantitative Social Science : An Introduction* (Imai 2017)
  - ▶ DataCamp: <https://www.datacamp.com>
  - ▶ R cheat sheets: <https://rstudio.com/resources/cheatsheets/>
- R Markdown
  - ▶ *R Markdown: The Definitive Guide* (Xie, Allaire, and Golemund 2019)

# Logistics - R Useful resources

- Data visualization
  - ▶ *Data Visualization: A Practical Introduction* (Healy 2018)
  - ▶ *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures* (Wilke 2019)
- Others
  - ▶ Stack Overflow: <https://stackoverflow.com>
  - ▶ TidyTuesday Project: <https://github.com/rfordatascience/tidytuesday>

# Logistics - Social Sciences & Computing

- 1 There are best practices for computing in the social sciences. You should aim for transparency and replicability in your work in general, and clarity and consistency in your code.
  - ▶ Best Practices (Wilson et al. 2014)
  - ▶ Good Enough (Wilson et al. 2017)

# R refresher

## 1 Overview

- R is a language and environment for statistical computing and graphics
  - ▶ *Object-oriented* style of programming
  - ▶ System-supplied or user-defined functionality as *functions*
  - ▶ Extended via *packages*
- RStudio is an integrated development environment for R, which includes:
  - ▶ a console to run R code
  - ▶ an editor to write code and text
  - ▶ tools for plotting, history, debugging and workspace management

# R refresher

## 2 Data Types

- ▶ character, numeric (integer or double), logical, complex
- ▶ data can also be missing

## 3 Data Structures

### • Matrices vs. data frames

- ▶ Matrices can only contain one **homogenous** type of vectors
- ▶ Data frames can contain **heterogeneous** types of vectors, and thus are more flexible

## 3 Data Structure - Summary

|    | Homogeneous   | Heterogeneous |
|----|---------------|---------------|
| 1d | Atomic vector | List          |
| 2d | Matrix        | Data frame    |
| nd | Array         |               |

- For much more see [here](#) or [here](#)

# R refresher

## ④ R as calculator

- ▶ Standard mathematical operators (e.g. + - \* / ^ etc.)
- ▶ Functions (e.g., mean()) take arguments (inputs)
- ▶ Logical operators (e.g. ==, >, <, >=, <=, !=) return TRUE FALSE or NA



# R refresher

## 4 R as calculator

```
1 + 7  
  
sqrt(2*2)  
  
factorial(5)  
  
(1 + 7)>= 4  
  
mean(c(1,7))
```

- 5 Create objects with assignment operator `<-`
  - ▶ Don't use `=` for assignment (even though it works)

# R refresher

Let's open RStudio and Lab 1 practice code