CSSS/POLS 510 MLE Lab

Lab 1. R Review + Intro to `RMarkdown` and `Overleaf`

Ramses Llobet

September 26, 2024

Disclaimer

The current version of the lab materials is adapted from those drafted by previous TAs for this course.

About me

- Welcome to the first lab section of POLS / CS&SS 510!
- I am Ramses, a fifth year Ph.D. student in **Political Science**.
  - Research interests: political economy and applied statistics.
- I am from Barcelona, Spain.
  - Please **do not** hesitate to stop me if you don't hear or understand me properly.
- **Do not** hesitate to ask questions. No question is silly. :)

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.rmd`, `lab1_key.rmd`, and `lab1_data.csv`
  - I will also provide a compressed *ZIP* file with all materials.

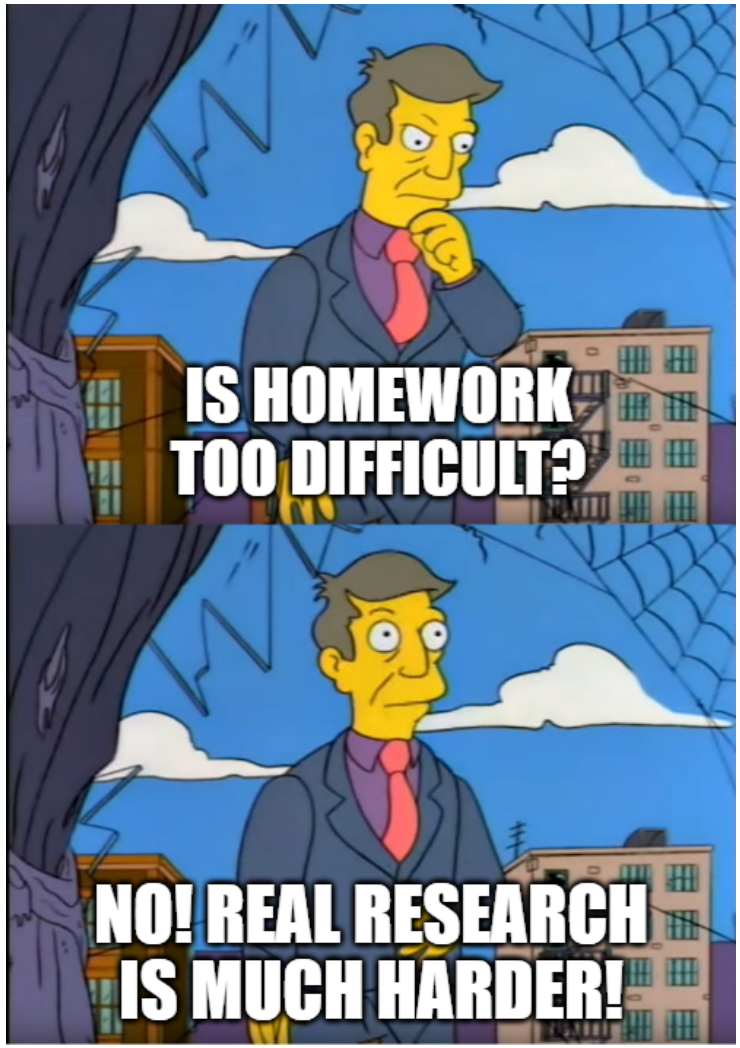
2. **Office Hours**: after labs or by appointment: *rllobet@uw.edu*.

- Trouble-shooting, questions about the lecture and assignments, etc.
- Please email or text me in Slack me with time and a short comment of the topic you want to discuss.
- Zoom link: (**meeting room**)

Logistics (Cont.)

3. **Homework**: 5 due every 2 weeks or so

- Must be typed up (not hand-written).
- Using LaTeX in Overleaf or R Studio with R Markdown.
- We will use two of **Chris's packages** extensively:
  - *simcf*
  - *tile*



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 programming "by hand" using *optim*, extract parameters of interest, and interpret these in probabilities.
   - Compute **predictions** and use simulation to find the confidence intervals of $\widehat{\pi}_i$ across counterfactuals values of covariates 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.
   - **Visualize** computations and quantities of interests.

`R` setup

- How to install `R` and `R-studio`.
  - R-4.4.1 for Windows
  - R-4.4.1 for macOS
- R-studio can be downloaded from posit's repository.
- To render **PDF** files in markdown, you will need to install the package `tinytex`.
  - At the end of today's lab review script file, you will find the functions to install it.

Logistics - R

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.

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

- For R:
  - Introductory:
    - *Hands-On Programming with R* (Grolemund 2014).
    - R cheat sheets.
  - Intermediate:
    - *R for Data Science* (Grolemund and Wickham 2023, 2nd edition).
    - *Data Visualization: A Practical Introduction* (Healy 2018).
    - *Graphical Data Analysis with R* (Unwin 2015).
  - Advanced:
    - *Advanced R* (Wickham 2019).
- For MLE:
  - *Maximum Likelihood for Social Science* (Ward and Ahlquist, 2018).
    - Book Materials code, data, etc.
  - *Unifying Political Methodology* (King 1998).

`R` review

- R basics.
- Data wrangling with `dplyr`.
- Install ‘tinytext“ for RMarkdown.

Let's open `RStudio` and **review_scrip.R**.

R-Markdown

- Save the following Cheat Sheet for RMarkdown.
- If any of you is looking for an general introduction for RMarkdown, I suggest you to check Chapter 27 from Wickham and Grolemund (2017) - **R for Data Science**.
- If you want a more comprehensive guide, then check Xie et al. (2021) - **R Markdown: The Definitive Guide**.
- Another, more applied, resource is Xie et al. (2022) - **R Markdown Cookbook**.

R-Markdown

- `RMarkdown` is a document format that allows you to integrate `R` **code** and **output** into a single document.
- Besides `R` code and output, it can also include **text**, **images**, and other **multimedia elements**, allowing for rich and informative documents.
- *Pandoc* is a free and open-source **document converter** that can convert documents from one markup language to another.
  - In the context of `Rmarkdown`, pandoc is the underlying document converter (sfotware) that converts the R-markdown file into a final output format, such as **HTML**, **PDF**, or **Word**.

R-Markdown

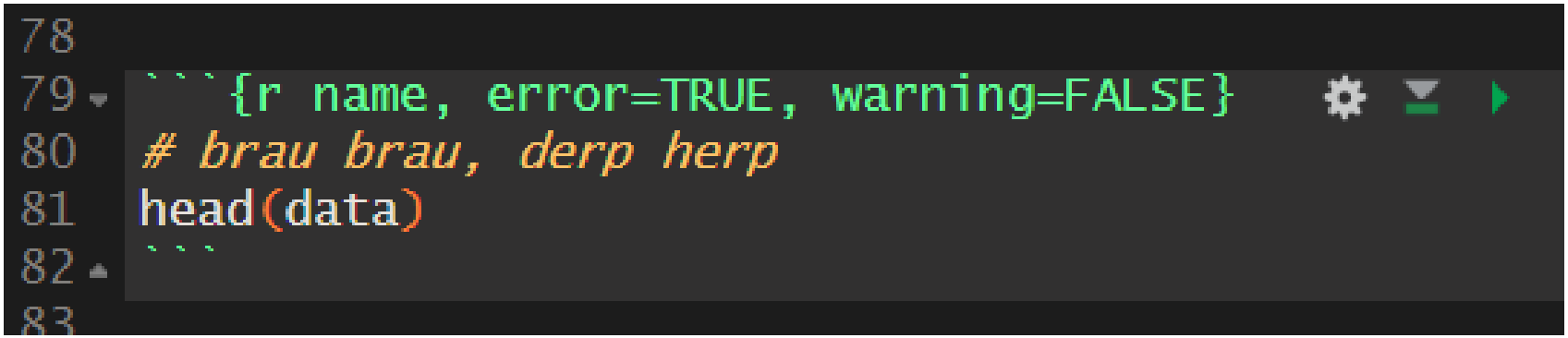
- The output format of the final document can be customized using options in the **YAML header** or external templates.

```yaml
---
title: "Lab 1 - Intro to RMarkdown"
author: "Your name"
date: \today
output:
  pdf_document:
    latex_engine: pdflatex
fontsize: 12pt
editor_options:
  chunk_output_type: console
---
```

- The YAML header in RMarkdown is a block of configuration settings at the beginning of the document enclosed by three hyphens (---).
- It is used to specify document metadata and other settings such as the document title, author, output format, and more.

R-Markdown

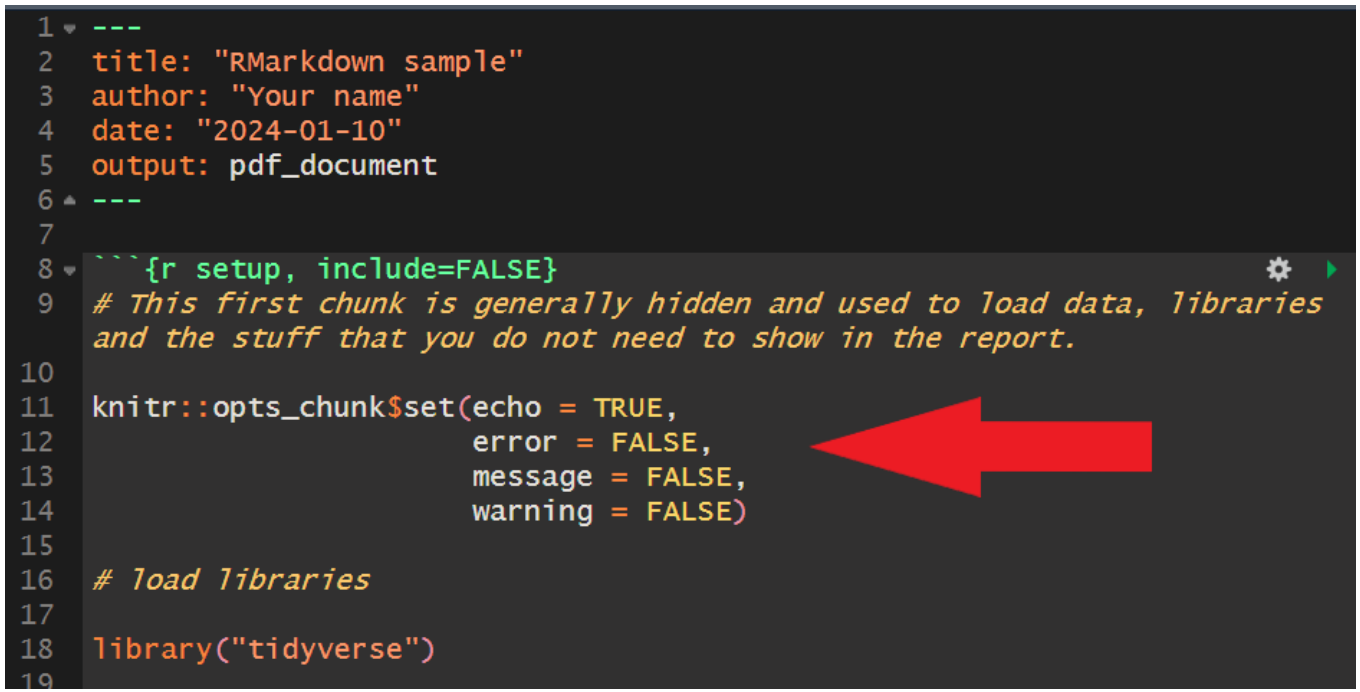
- **Code chunks** are sections of `R` code that can be executed and embedded within an `RMarkdown` document.



- Code chunks can be inserted using the syntax `{r}` and closed with “‘.
  - Short cut in Windows: `Ctrl + Alt + I`
  - Short cut in macOS: `Cmd + Option + I`
- Code chunks can be customized with various **chunk options**.

R-Markdown

- **Note**: set the function `knitr::opts_chunk$set()` with any general setting without repeating it in every code chunk.
- Recommendation chunck options for Homework

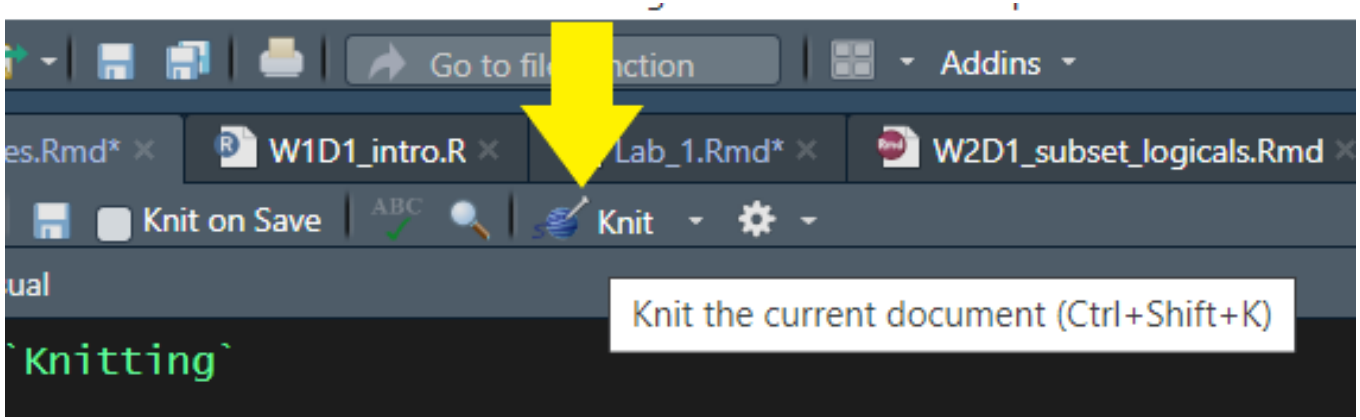


R-Markdown

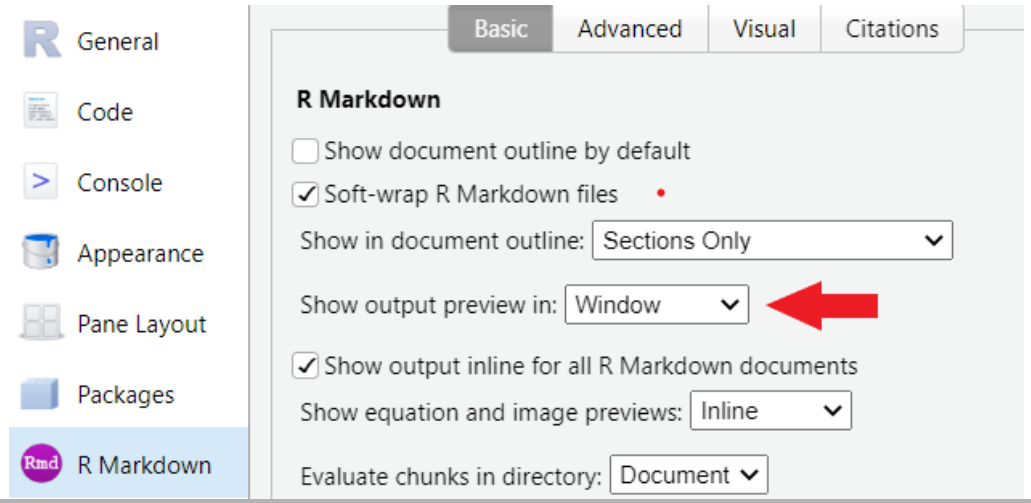
- In `RMarkdown`, **rendering** a document means converting the source `RMarkdown` file into its final output format (using pandoc).
- To render a document, we need to `Knit`, knitting is the process of taking the RMarkdown file and converting it into a single, cohesive document that can be rendered into different formats (HTML, PDF, etc).
  - To produce **PDF file**, you need `TeX` files.
- Easy way: Install the `tinytex` package: `install.packages("tinytex")`. Then run `tinytex::install_tinytex()`.

Knitting

- To knit:



- Auxiliary window for output preview:



R-Markdown

- Live demonstration and practice exercise:
  - Open the file `RMarkdown_sample.rmd`
  - **Optional** homework: `lab1_practice.rmd`

Intro to LaTeXwith `Overleaf`

- Alternatively, we have `Overleaf`: https://www.overleaf.com/
  - An online LaTeXeditor
    - Integrated PDF preview panel
    - Quality of life features: auto-complete commands, auto-close brackets, keyboard shortcuts, etc.
    - Numerous templates: journal articles, books, CVs, slides, posters, etc.
    - Easy collaboration (But not free)
    - Integrated with `Zotero` and `Mendeley` for bibliography management
    - Integrated with `Git` for version control

Intro to LaTeXwith `Overleaf`

- Before we dive in, useful resources
  - The Not So Short Introduction to LaTeX 2_ε (Oetiker et al., 2023)
    - Learn LaTeXin 280 pages / minutes
  - 'Overleaf' documentation
    - Contains intro to basic LaTeX, `Overleaf`, and many practical guides
  - TeXat StackExchange
  - General: Mathematics and Tables and TikZ
  - Beamer Theme: here
  - Bibliography: natbib, doi2bib, text2bib
  - Other: here

Intro to LaTeXwith `Overleaf`

- For an introductory tutorial series, I recommend Dr.Trefor Bazett's YouTube channel.
- Some useful templates:
  - Thesis: here
  - Homework: my sample.
  - Working paper: Kenya's sample, and Chris's sample
  - Academic journal: here
  - Presentation slides (Beamer): here and here
  - Poster: here
  - CV: here and here
  - Graphs, trees, diagrams (TikZ): here and here

F I N