

Day 1: Introduction to R Statistical Analysis Software

Qingyin Cai

Department of Applied Economics
University of Minnesota

Summer 2025

WebR Status



Outline

- Introduction
 - Course Review
 - Icebreaker
- Motivation
 - What you can do with R?
 - How do we use R in course in APEC?
 - Basic knowledge about R and Rstudio

Course overview

Intro

- This one-week course is a boot camp designed to introduce you to R statistical software. My goal is to build a strong foundation for your Ph.D.-level econometrics courses and future research.
- **By the end of this week, you will be able to:**
 - to create and manipulate the base-R object data.
 - to do data manipulation with `data.table` package.
 - to do data visualization with `ggplot2` package.
 - to conduct regression analysis with `lm()` and make a publish-ready regression table with `modelsummary()` package.
 - to write Monte code for Carlo simulations using `for` loop function.

Course overview

Style

- We will meet each day from 1:00 PM to 4:00 PM, with office hours immediately following.
- Each lecture is divided into three sessions, with each session consisting of a 50-minute lecture and a 10-minute break.
- We will have in-class exercises at the end of each topic, and after-class exercises (optional) to practice!

Course overview

References

No textbook is required. Below are recommended resources:

Recommended Reading

- [R for Data Science](#)
- [ggplot2 Book](#)
- [Intro to data.table](#)
- [Intro to Econometrics with R](#)

Essential Cheatsheets

- [Base R](#)
- [data.table](#)
- [ggplot2](#)
- [R Markdown](#)

About myself

- **Qingyin Cai**
 - From China
 - Fifth-year Ph.D. in Applied Economics
 - Area of interests: Food and Agriculture Economics, Consumer Economics, and Environment Economics.

- **Introduce yourself**
 - What's your name?
 - What's your program?
 - Where are you from?
 - What brings you to UMN?

Motivation to learn R

What is R?

- R is a powerful programming language for a wide range of tasks:
 - Data Manipulation: cleaning, reshaping, merging datasets, API.
 - Various Analysis: descriptive analysis, regression, GIS, spatial analysis, machine learning.
 - Data visualization.
- It's a great tool to communicate your results with others (documentation, papers, slides, books, etc.).

Comparison of R, Stata, and Python

Criteria	R	Stata	Python
Primary Use	Statistical analysis, visualization, research	Economics/social science research; valued for tested results	General-purpose; machine learning, web scraping, automation
Cost	Open-source	Commercial license	Open-source
Data Visualization	Excellent	Less flexible or aesthetically pleasing	Very powerful, but can be verbose
Ecosystem	Large academic community; many packages on CRAN	Strong in economics but smaller user base	Huge, diverse community
Handling Big Data	Base R is memory-bound; packages like <code>data.table</code> / <code>arrow</code> improve performance	Memory-bound	Excellent

AI and Learning R

- AI can help, but it cannot replace understanding.
 - Tools like ChatGPT or Copilot can generate R code, but you need to know if the code is **correct and appropriate**.
 - You'll understand why a method works, not just how to run it.
- Academic integrity & skill development.
 - Employers and researchers may expect you to adapt and debug code yourself.
- Long-term benefit.
 - Once you know R, AI becomes a more powerful assistant — you can ask better questions and spot mistakes.

R in the APEC curriculum

- We use R extensively in the Econometrics series (APEC 8211-8214).
 - To conduct regression analysis (OLS, IV, FE, etc.).
 - To conduct Monte Carlo simulations.
 - e.g., To understand the difference in variance inference techniques.
- Don't worry if you are new to this!
 - Basic knowledge is enough to start.
 - If you would like to learn R programming further, I recommend that you take Programming for Econometrics (APEC8221) and Big Data Methods in Economics (APEC8222).

Rstudio

What is it?

- You can use  app to write and run R codes, but it has a terrible graphic user interface.
- **Rstudio** is an Integrated Development Environment. It provides a user-friendly interface to write and run R code, view plots, and manage files.
- You must install R (the engine) before you can use RStudio!
- R studio looks like this:

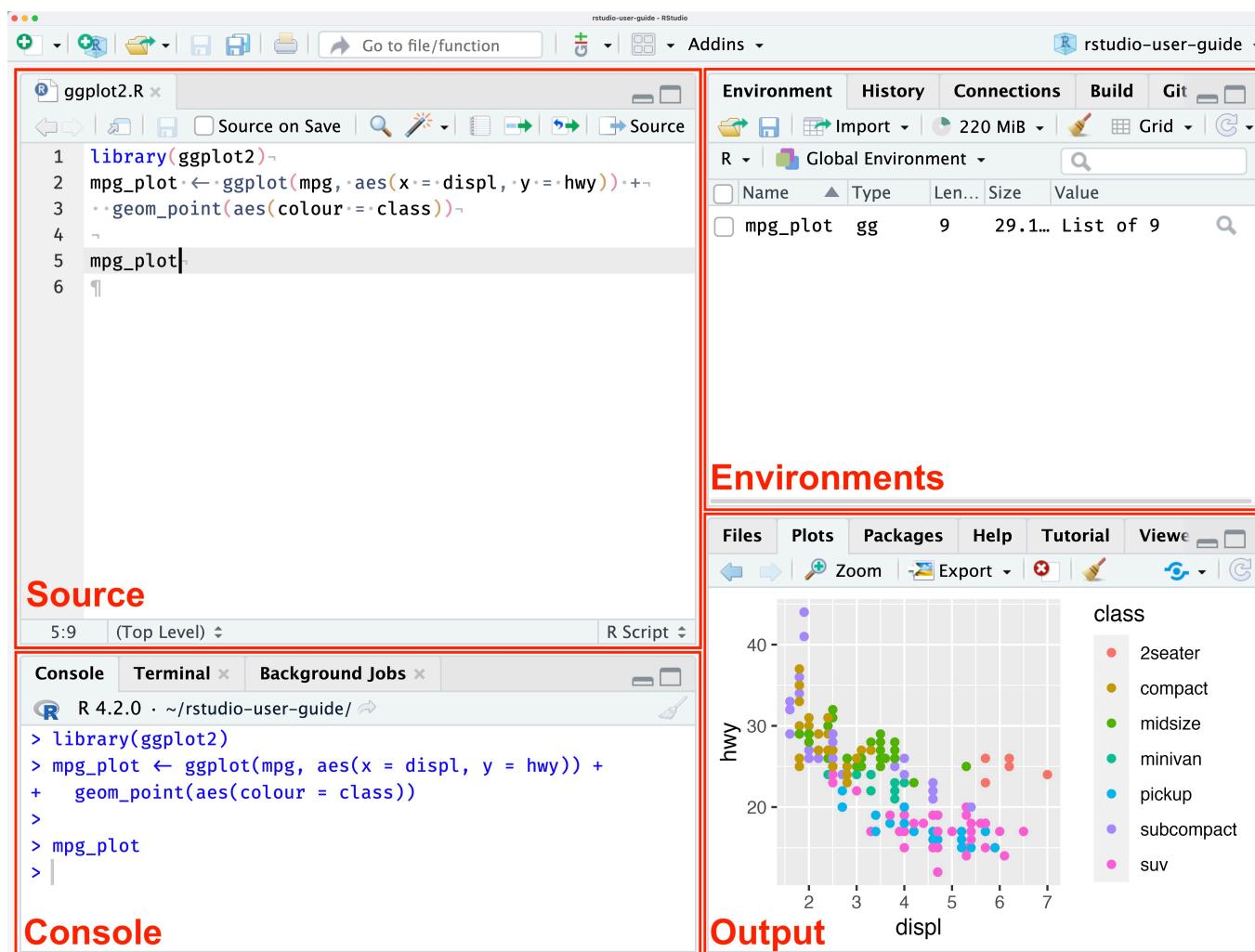


Image Source: Rstudio User Guide

Rstudio

Create New R code File

- To create new R script file, click the **+** button on the top-left corner of the Rstudio, or hit **Ctrl + Shift + N** (**Cmd + Shift + N** on mac).
- To save the file, click the floppy disk icon , or **Ctrl + S** (**Cmd + S** on macOS).

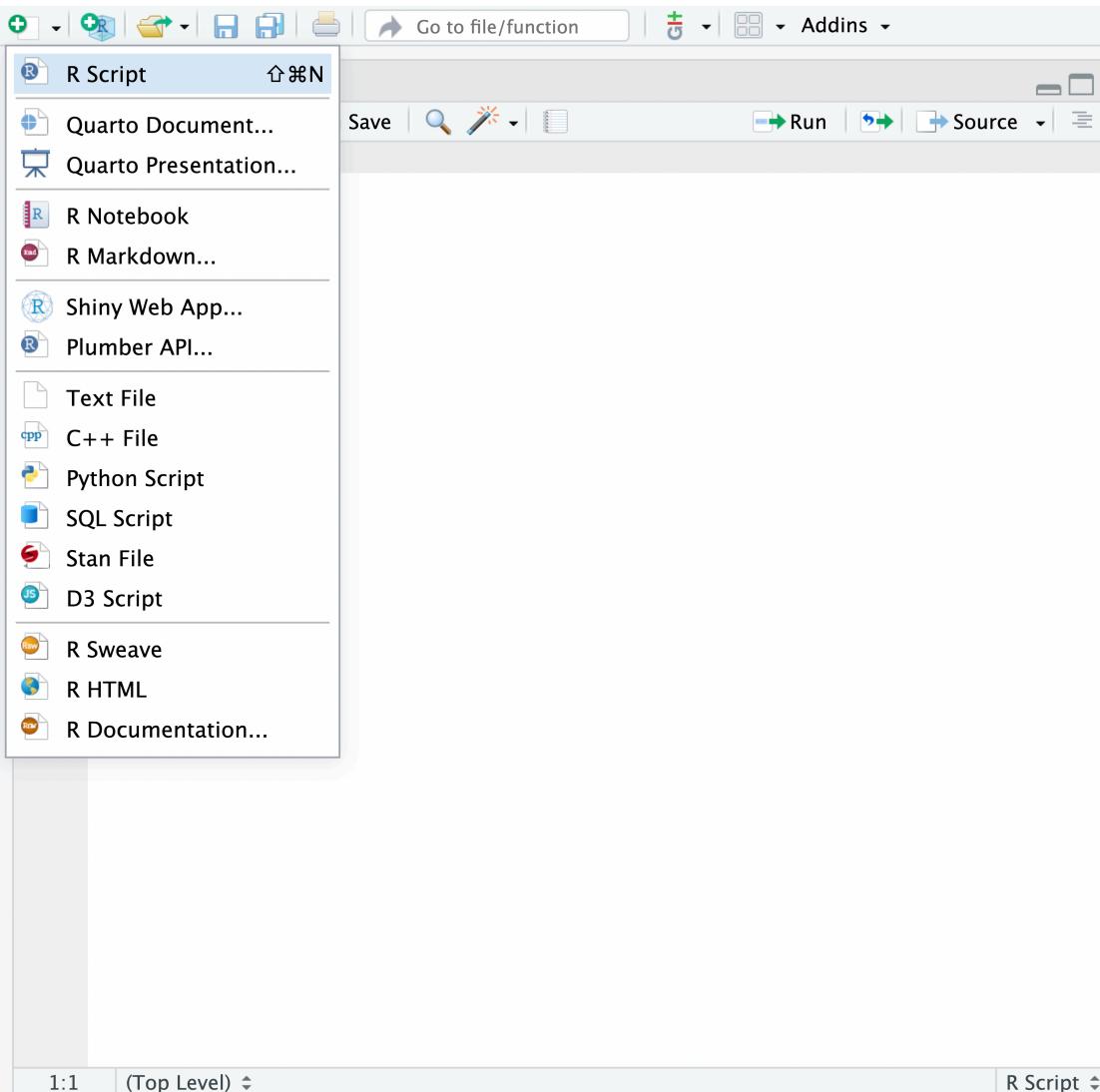
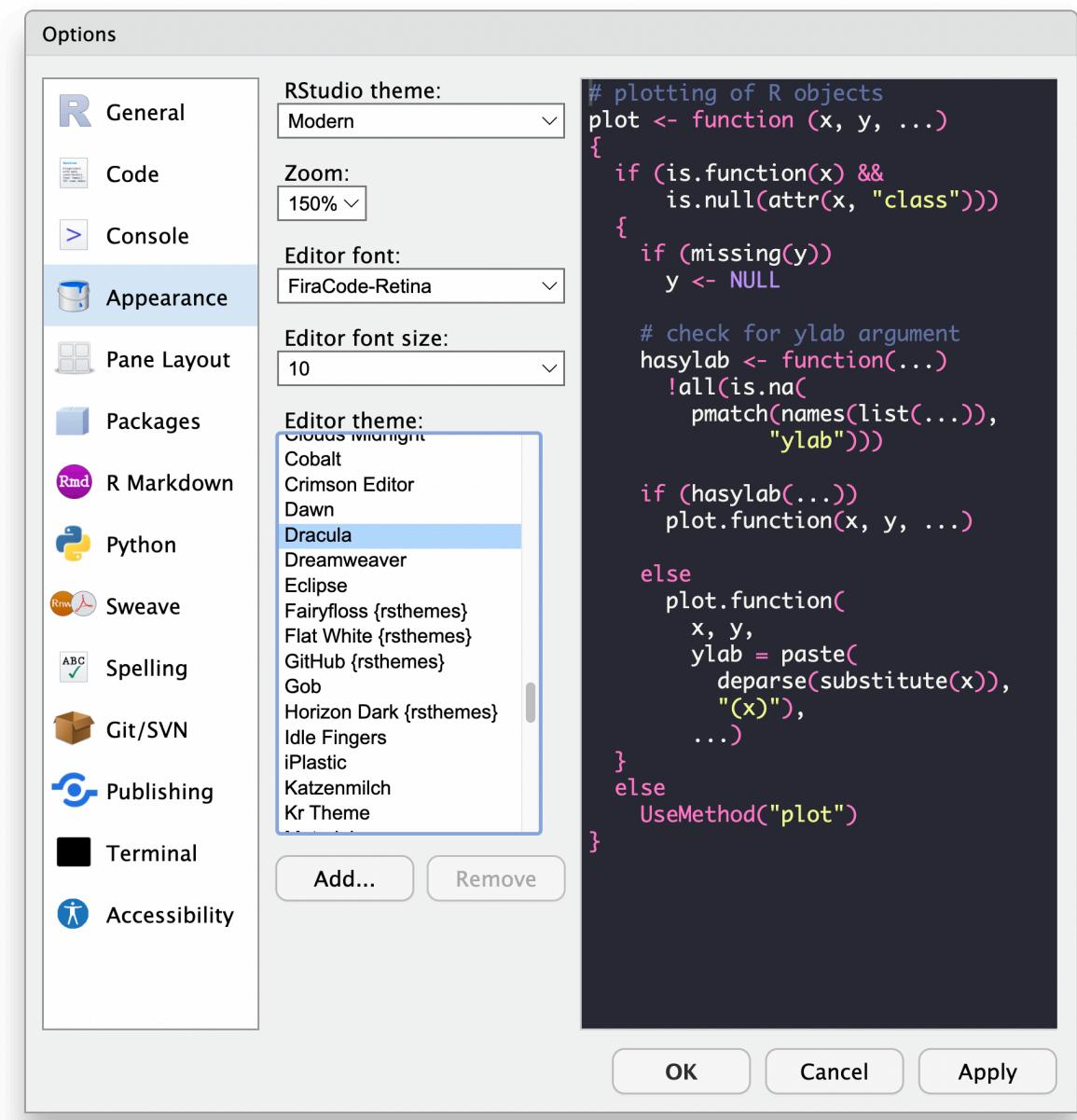


Image Source: Rstudio User Guide

Rstudio

Change the theme (Optional)

- You can change the appearance of Rstudio by going to **Tools -> Global Options -> Appearance -> Editor theme** and select your favorite theme.



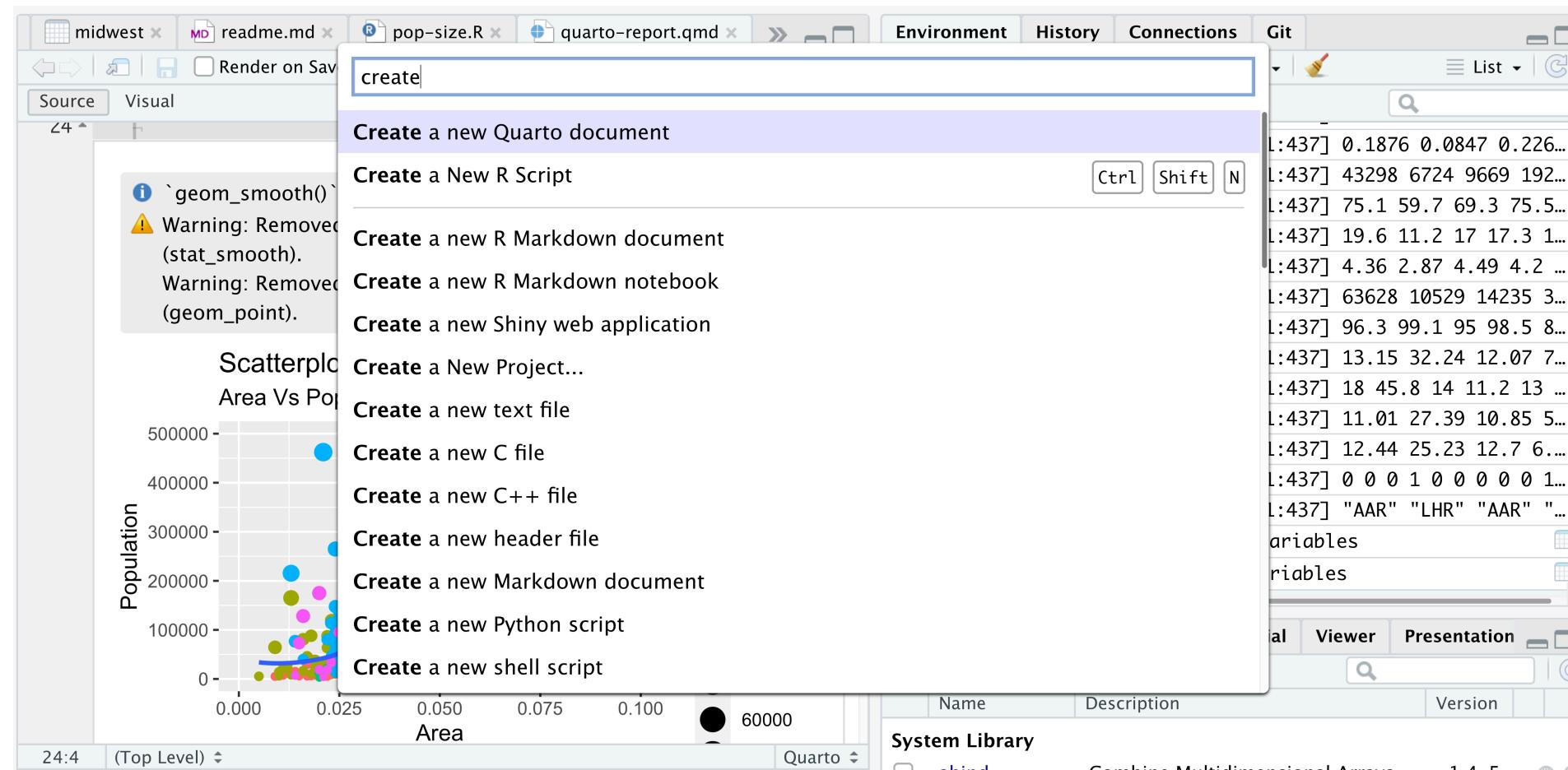
Multiple panes (Optional)

- You can have multiple code panes in Rstudio.
- To create a new pane, go to [Tools](#) -> [Global Options](#) -> [Pane Layout](#) -> [Add Column](#).
- In the same window, you can also change the layout of the panes.

Rstudio

Command Palette (Optional)

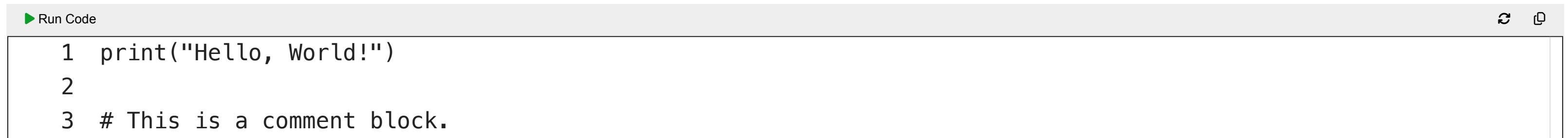
- Recent R-studio has a new feature called “Command Palette.”
- Hit **Ctrl + Shift + P** (**Cmd + Shift + P** on macOS) on your keyboard, or go to **Tools -> Show Command Palette**.
- From here, you can search for and do almost anything: create new files, open projects, etc.



Source: Rstudio User Guide

Rstudio: Running Code

Let's write some codes.



A screenshot of the RStudio code editor. The code pane contains the following R code:

```
1 print("Hello, World!")
2
3 # This is a comment block.
```

The code editor has a toolbar at the top with icons for Run Code (green triangle), Undo (left arrow), Redo (right arrow), and Close (square). The code is syntax-highlighted, with "print" in blue and the comment line starting with "#".

R code

- Any thing you write in the source (or console) pane is regarded as R code.
- To run (execute) the code, select the code line, and click the “Run” bottom, or use the shortcut key: **Ctrl + Enter** (**Cmd + Enter** on macOS).

Comment block

- Any line starting with a **#** is a comment. R will ignore it. Use comments to leave notes for yourself and others!

Summary

- You are now familiar with the basics of RStudio. As long as you know how to create, save, and run a script, you are ready for the next lecture.
- For more details, see the [official RStudio IDE Cheatsheet](#).
- While RStudio is the most popular tool, you can also run R in other editors like [Visual Studio Code](#) to run R. Nevertheless, Rstudio is a great starting point to get familiar with R.