Data Science for Operational Researchers Using R Online

1. Introduction to R and Posit Cloud

Prof. Jim Duggan,
School of Computer Science
University of Galway.

https://github.com/JimDuggan/explore_or

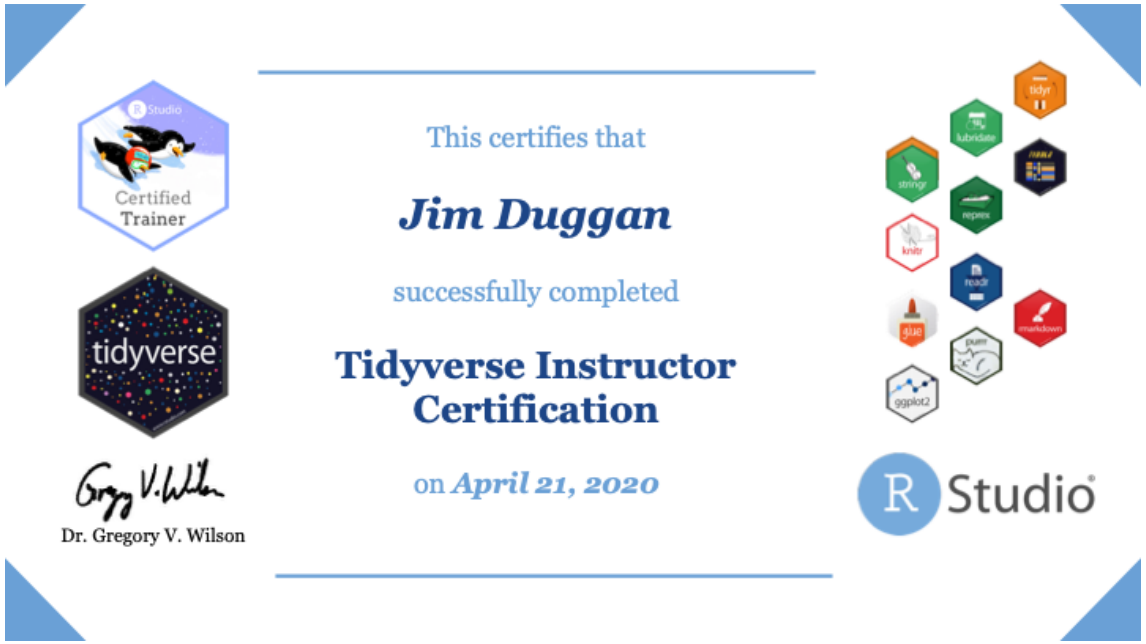
> R is an extremely versatile open source programming language for statistics and data science.
>
> — Norman Matloff (Matloff, 2011)

Motivation

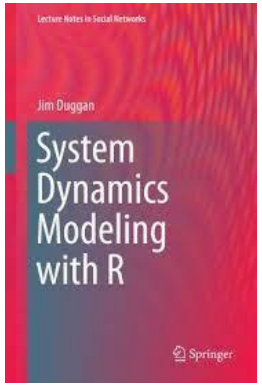
- R is an open source programming language, and a valuable computational tool that can be applied to the field of operations research.
- R provides excellent features such as data representation, data manipulation, and data analysis.
- These features can be integrated with operations research techniques (e.g., simulation, linear programming, and data science) to support an information workflow

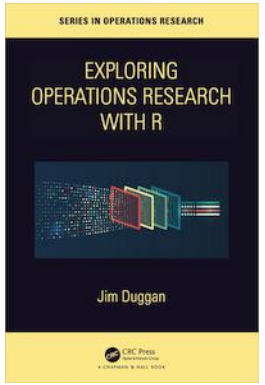
Instructor – Jim Duggan

- Lectures in
  - Programming (R, Python),
  - Modelling & Simulation
- Research interests:
  - System Dynamics
  - Computational Epidemiology
  - Data Science



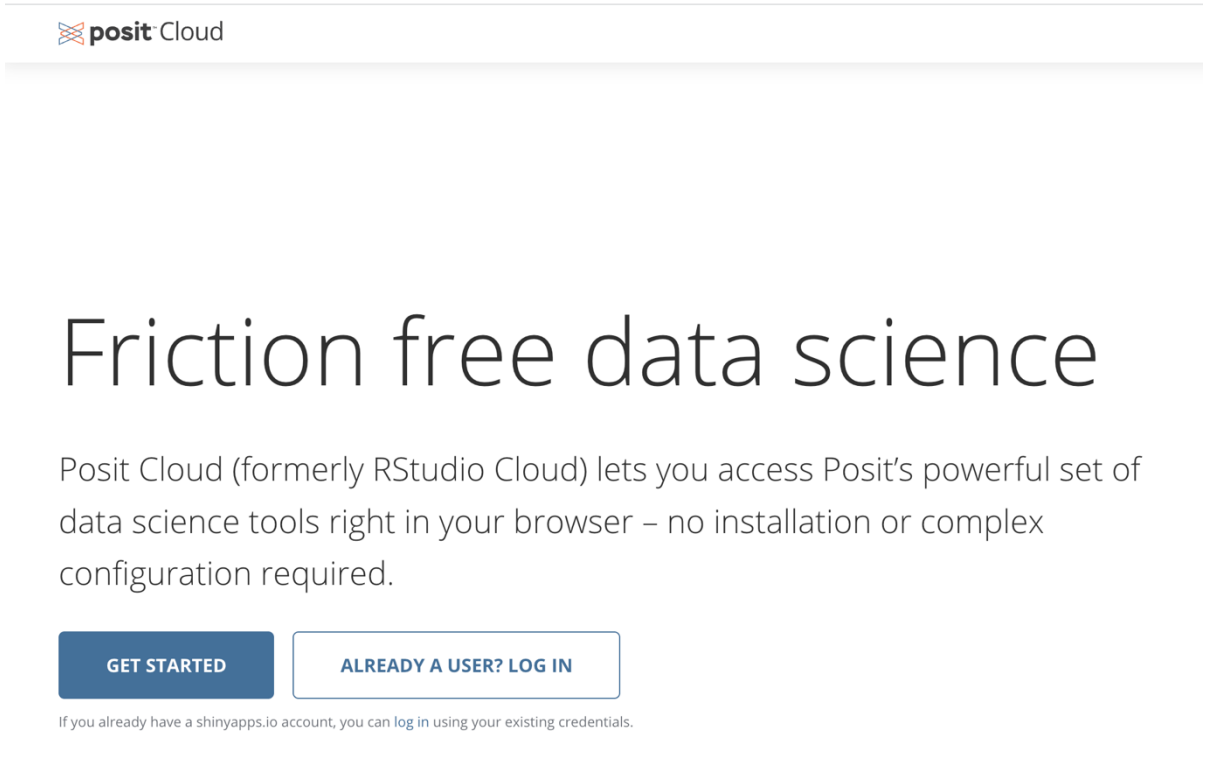
https://github.com/JimDuggan

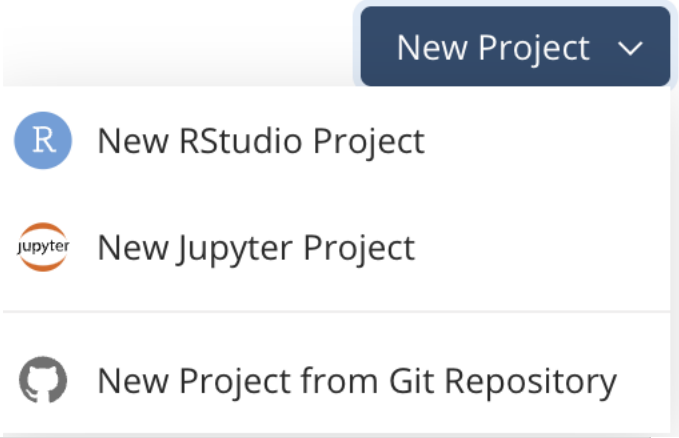


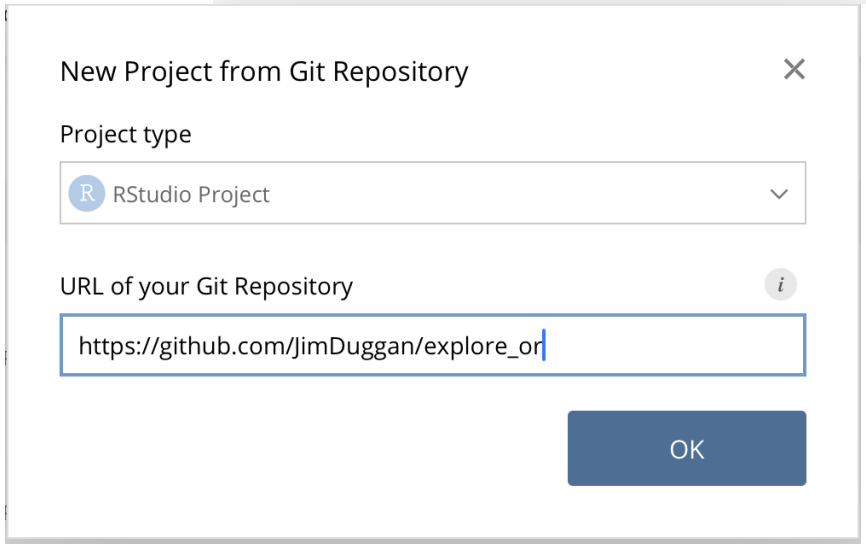


First Steps: Posit Cloud – Create Your Account

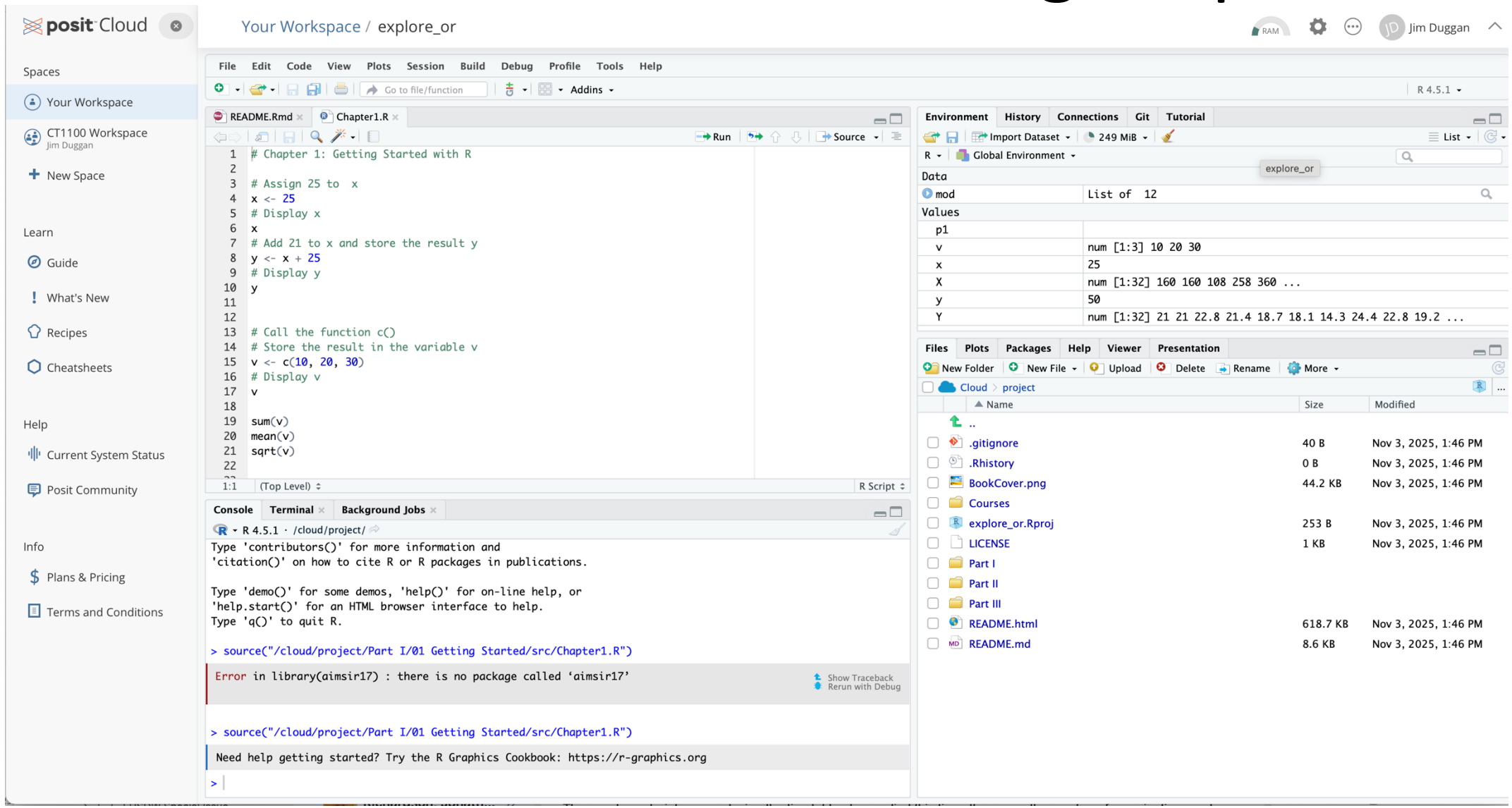
https://posit.cloud/



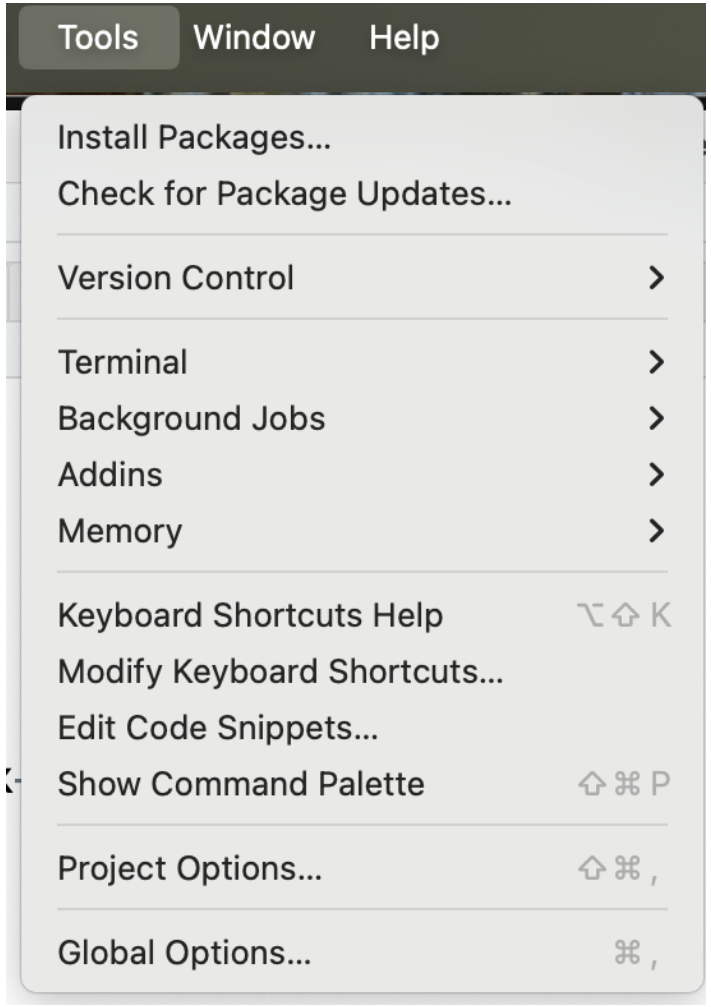


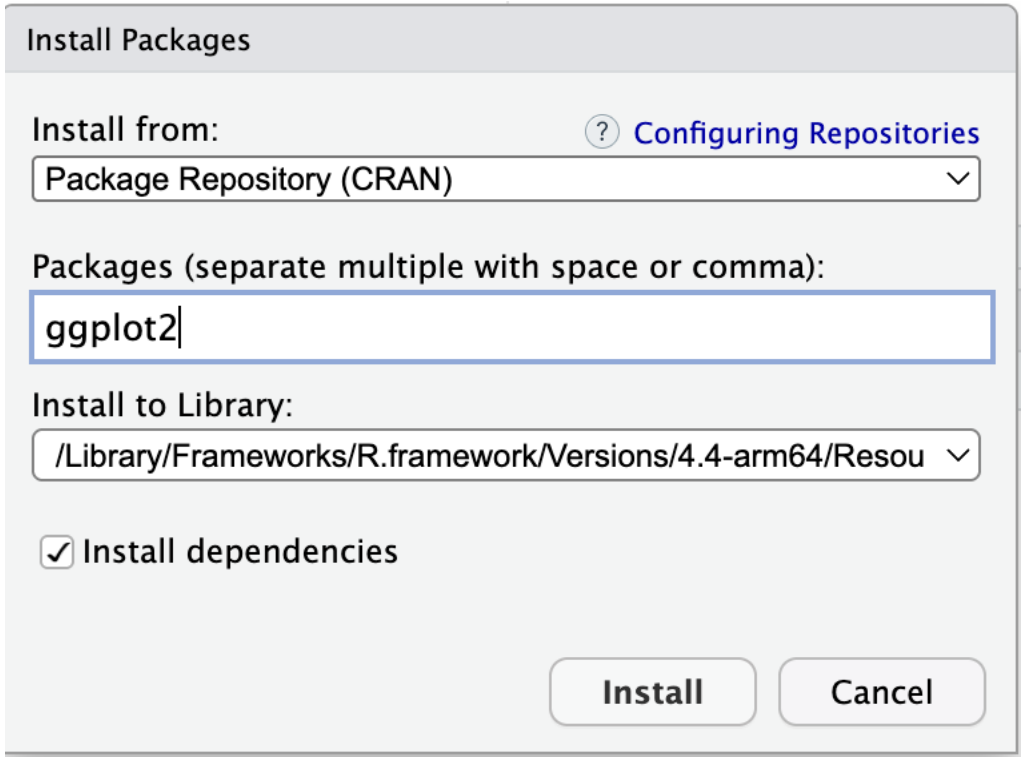


IDE Available for running scripts



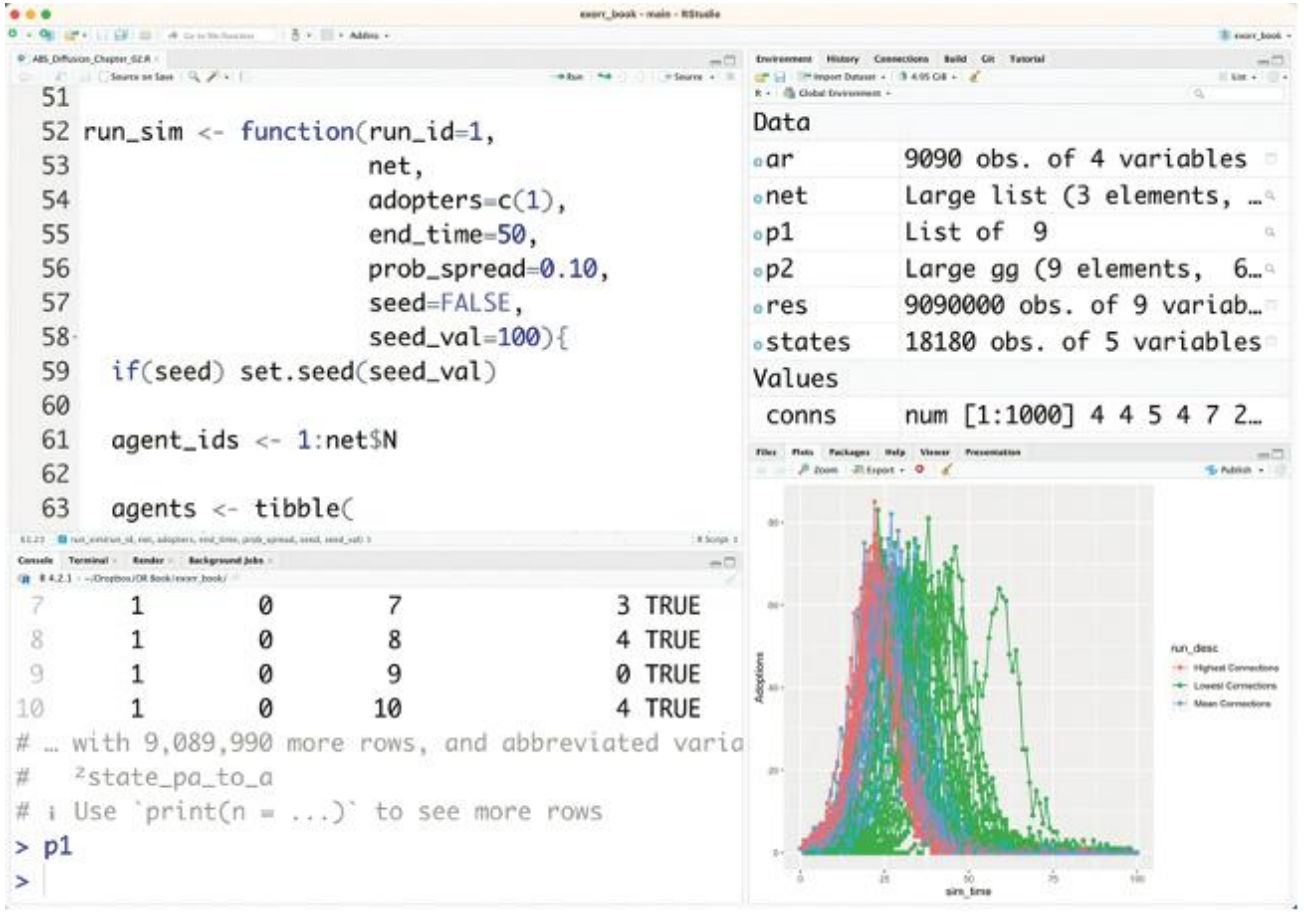
Install Packages





Posit cloud IDE

- A source code editor, where R scripts can be created, edited, and saved.
- The R console, which allows you to create and explore variables, and observe the immediate impact of passing an instruction to R.
- Files and plots, where we can see the full file system for your written code (folders, sub-folders, files), and also view any generated plots.
- The global environment which shows what variables you have created, and therefore which ones can be explored and used for data processing.



Some tips…

1. The "Source" button, which appears on the editor panel, will execute the currently active R script from start to finish.
2. The question mark is used to provide help. For example ?sample will inform you of how the function sample() works.
3. The double question mark provides access to further resources. For example, if you type ??ggplot2 at the console, you can browse vignettes relating to plotting with ggplot2.
4. The menu option "Session>Clear Workspace" will clear all variables from your global environment. This is useful before you run a script, just to ensure that no current variable could disrupt your data processing workflow.
5. The menu option "Tools>Install Packages. . . " will allow you to access CRAN and install any of the available libraries.

R and Functions

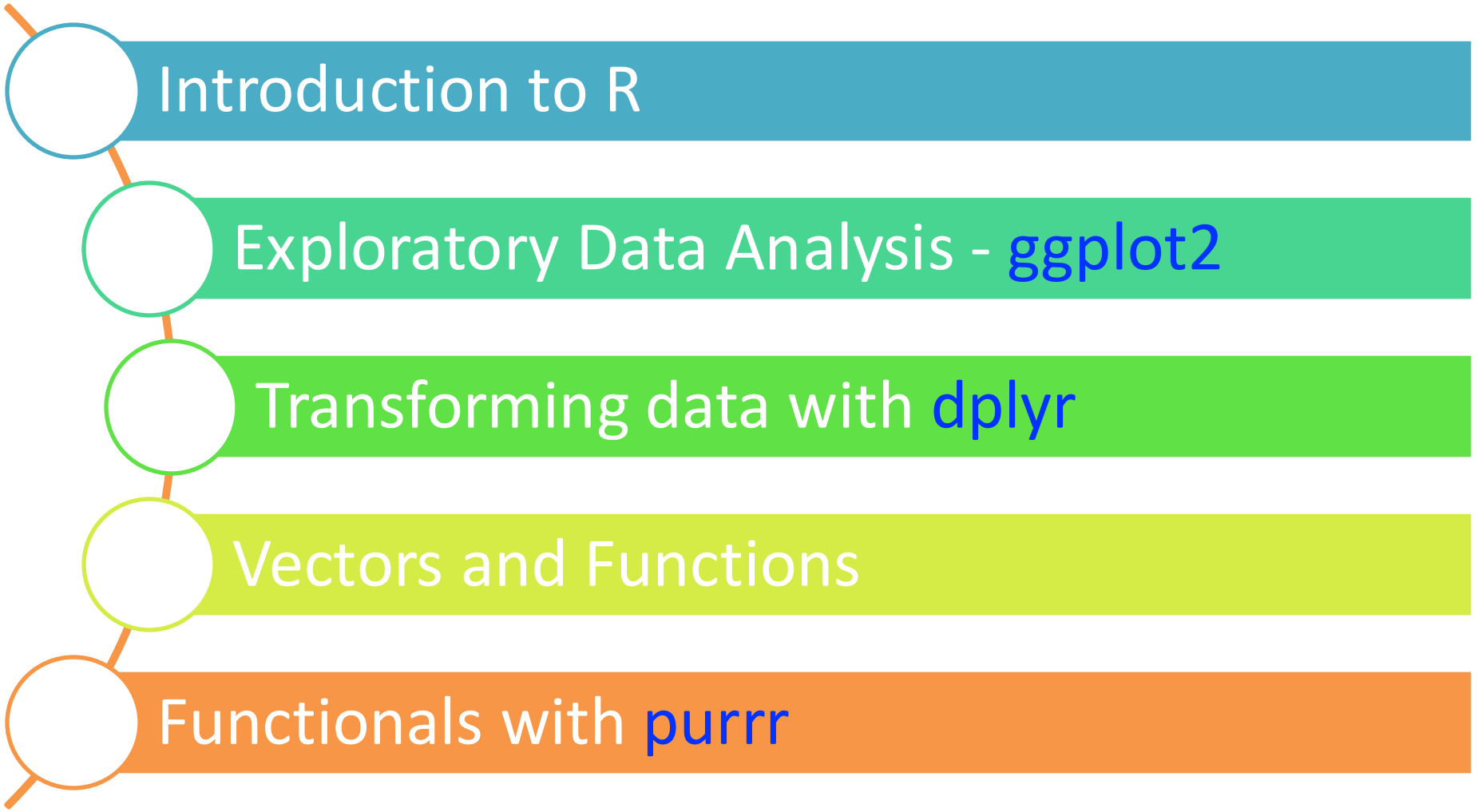
- Functions are central to programming in R.
- There are two kinds of functions: (1) those available within base R, and other libraries; and (2) functions written by programmers for their own use.
- (Note the term base R is used to describe the core version of R).
- All functions take a set of arguments as input, and then return a result.

```
# Call the function c()
# Store the result in the variable v
v <- c(10, 20, 30)
# Display v
```

```
v
#> [1] 10 20 30
```

```
sum(v)
#> [1] 60
mean(v)
#> [1] 20
sqrt(v)
#> [1] 3.162 4.472 5.477
```

Day 1 Overview



Day 2 Overview