

The first principle I propose is that our *Mission*, as users and creators of software for data analysis, is to enable the best and most thorough exploration of data possible. That means that users of the software must be able to ask meaningful questions about their applications, quickly and flexibly.

— John Chambers ([Chambers, 2008](#))

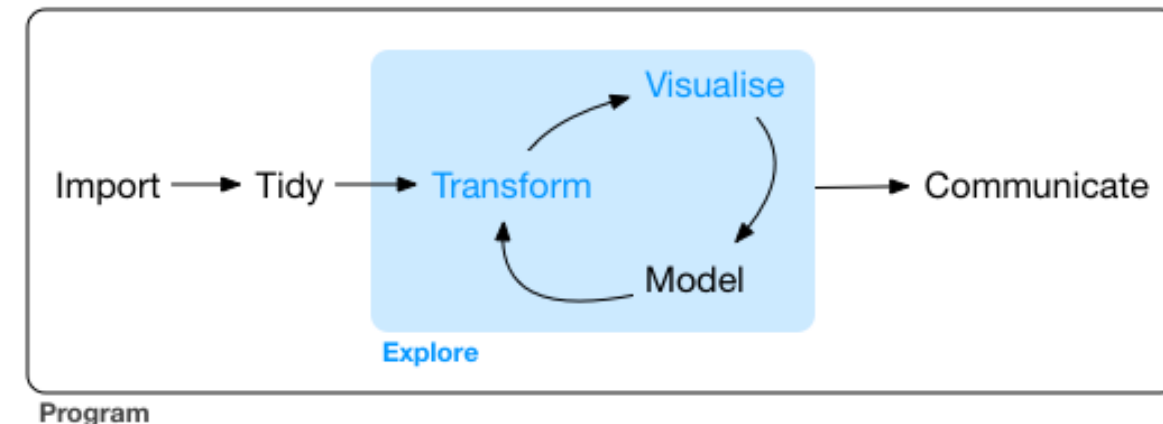
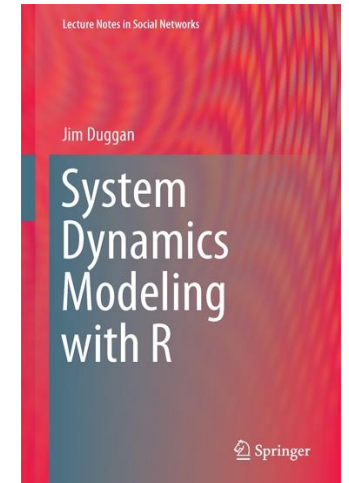
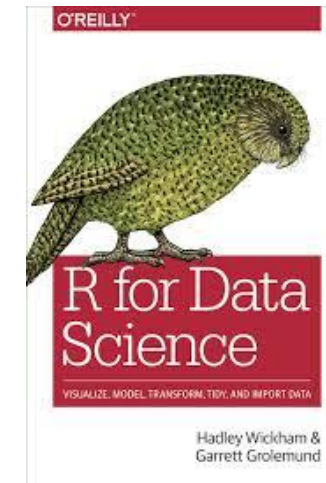
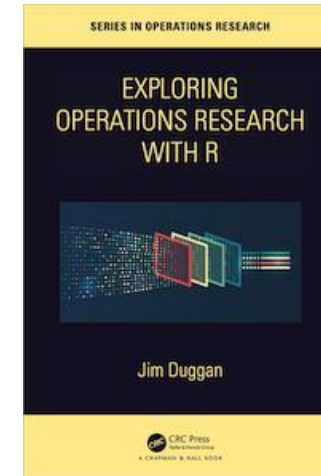
Data Science for Operational Researchers using R

01 - Introduction

https://github.com/JimDuggan/explore_or

Goal

- Improve your personal productivity for generating rapid and insightful results from large data sets.
- Learn about the role of R in the overall data science life process.
- Appreciate the power of R and the 'tidyverse' set of packages for increasing productivity



<https://r4ds.had.co.nz>

Instructor –Jim Duggan



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



<https://education.rstudio.com/trainers/people/duggan+jim/>

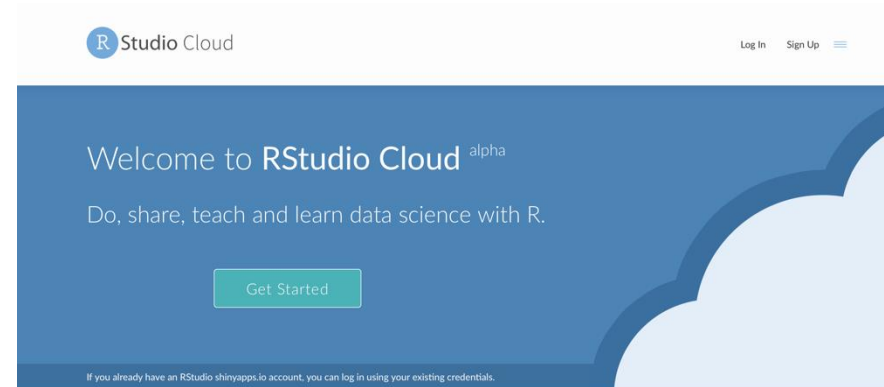
https://twitter.com/_jimduggan

Course Overview

Topic	Description
	<u>Session 1 (half-day)</u>
1	Introduction to R and Posit Cloud
2	Atomic Vectors
3	Functions, Lists and Functionals
4	Data Frames and Tibble
5	ggplot2
	<u>Session 2 (half-day)</u>
6	Data transformation with dplyr
7	Relational Data with dplyr
8	Processing data with purrr
9	Exploratory Data Analysis – Examples

The R Project for Statistical Computing

- R's *mission* is to enable the best and most thorough exploration of data possible (Chambers 2008).
- It is a dialect of the S language, developed at Bell Laboratories
- ACM noted that S “*will forever alter the way people analyze, visualize, and manipulate data*”



```
1 # We use this for processing the answer
2 # In programming, we "stand on the shoulders of giants"
3 library(stringi)
4
5 # This gets the input from the user.
6 # The result is stored in a variable
7 # Variables are important in programming!
8 name <- readline(prompt="Enter a name: ")
9
10 # We call a specially designed function to get the answer
11 # In R, we call functions all the time
12 # A function is a "mini-program"
13 ans <- stri_reverse(name)
14
15 # After all this work, we output the result
16 cat("The reverse of ", name, "is ==>", ans)
17
```

First Steps: Posit Cloud – Create Your Account

 posit Cloud

Friction free data science

Posit Cloud (formerly RStudio Cloud) lets you access Posit's powerful set of data science tools right in your browser – no installation or complex configuration required.

GET STARTED

ALREADY A USER? LOG IN

If you already have a shinyapps.io account, you can [log in](#) using your existing credentials.

New Project ▾

 New RStudio Project

 New Jupyter Project

 New Project from Git Repository

New Project from Git Repository



URL of your Git Repository 

`https://github.com/JimDuggan/Data-Science-for-OR`

OK

View the R IDE

The screenshot displays the Posit Cloud R IDE interface. The top navigation bar shows the user's workspace, "Data-Science-for-OR", and the R version, 4.3.0. The left sidebar contains links to "Your Workspace", "CT1100 Workspace", "New Space", "Learn", "Guide", "What's New", "Primers", "Cheat Sheets", "Help", "Current System Status", "Posit Community", "Info", "Plans & Pricing", and "Terms and Conditions".

The main console area shows the R version 4.3.0 (2023-04-21) -- "Already Tomorrow" and the copyright notice for the R Foundation for Statistical Computing. It also displays the license information and the R Foundation's disclaimer. The console output shows the R version and the license information.

```
R version 4.3.0 (2023-04-21) -- "Already Tomorrow"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

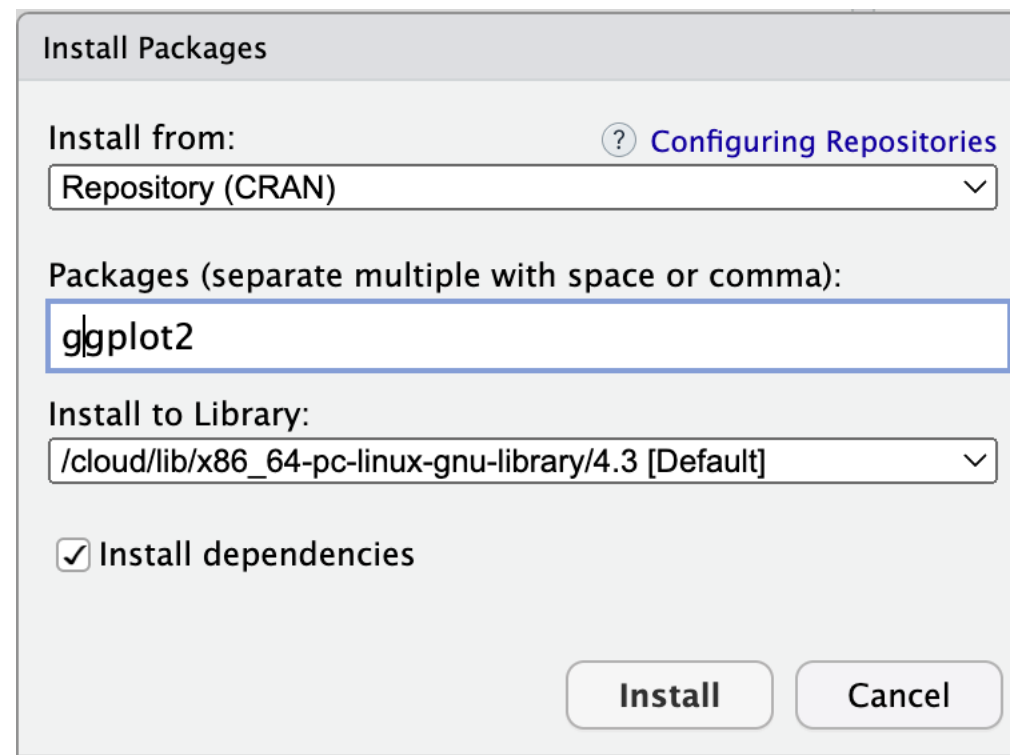
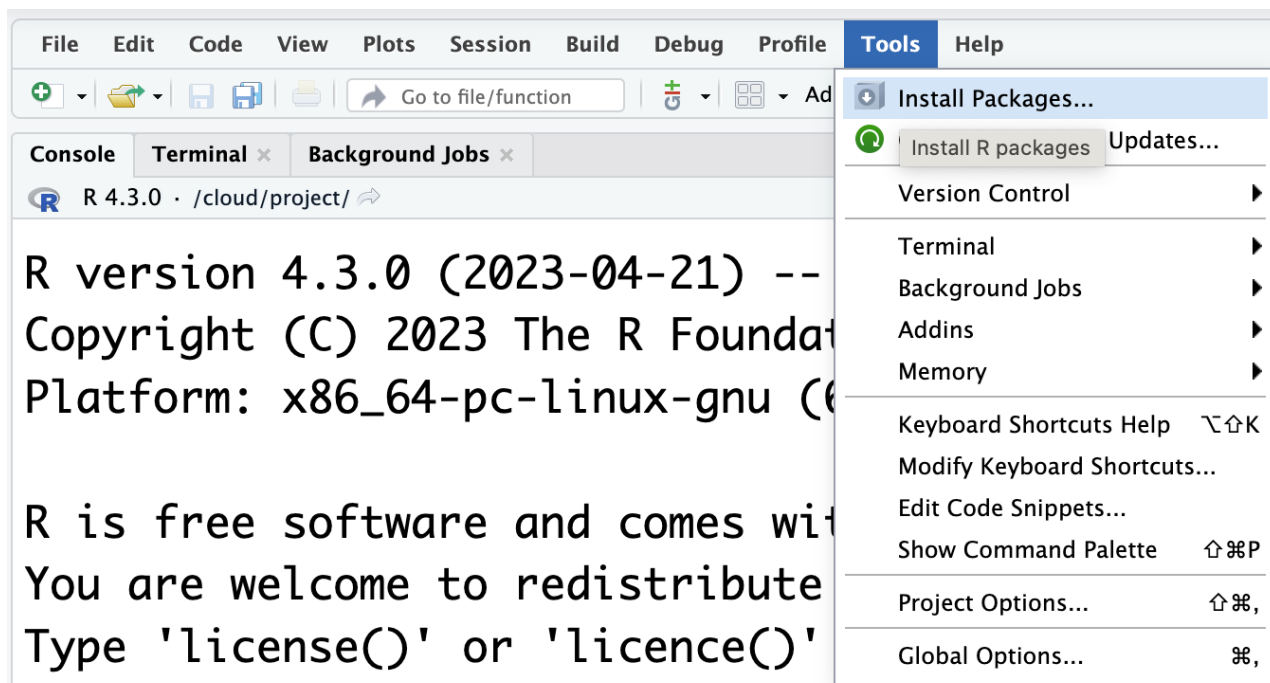
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> WorldPhones
      N.Amer Europe Asia S.Amer Oceania Africa Mid.Amer
1951  45939  21574  2876   1815   1646    89     555
1956  60423  29990  4708   2568   2366   1411    733
1957  64721  32510  5230   2695   2526   1546    773
1958  68484  35218  6662   2845   2691   1663    836
1959  71799  37598  6856   3000   2868   1769    911
1960  76036  40341  8220   3145   3054   1905   1008
1961  79831  43173  9053   3338   3224   2005   1076
> ?WorldPhones
> |
```

The environment panel on the right shows the loaded packages, including "package:datasets". The "Data" section displays the "WorldPhones" dataset, showing the number of telephones in various regions of the world (in thousands) for the years 1951, 1956, 1957, 1958, 1959, 1960, and 1961. The "Values" section shows the data for the "WorldPhones" dataset, with columns for "ability.cov", "airmiles", "AirPassengers", and "airquality".

The "The World's Telephones" documentation panel on the right provides a description of the dataset, its usage, and its source. The description states: "The number of telephones in various regions of the world (in thousands)." The usage section shows the "WorldPhones" dataset. The format section states: "A matrix with 7 rows and 8 columns. The columns of the matrix give the figures for a given region, and the rows the figures for a year. The regions are: North America, Europe, Asia, South America, Oceania, Africa, Central America. The years are: 1951, 1956, 1957, 1958, 1959, 1960, 1961." The source section states: "AT&T (1961) The World's Telephones." The references section states: "McNeil, D. R. (1977) Interactive Data Analysis. New York: Wiley." The examples section shows the command "Run examples".

(5) Install packages...



```
> install.packages("ggplot2")
Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
(as 'lib' is unspecified)
also installing the dependencies 'colorspace', 'utf8', 'farver', 'labeling', 'munsell', 'R6', 'RColorBrewer', 'viridisLite',
'fansi', 'magrittr', 'pillar', 'pkgconfig', 'cli', 'glue', 'gtable', 'isoband', 'lifecycle', 'rlang', 'scales', 'tibble', 'v
ctrs', 'withr'
```


(7) Packages required

Package	Purpose
ggplot2	Produce graphics for data
dplyr	Analysis of data held in tibbles/data frames
aimsir17	2017 Weather data for Ireland
purrr	To iterate over data structures
tidyr	To tidy rectangular data, and to nest data sets