R語言簡介

王貿

國立臺灣大學行為與資料科學研究中心助理研究員

國立臺灣大學政治學系博士、兼任講師

maowang01@gmail.com

安裝 R 與 RStudio

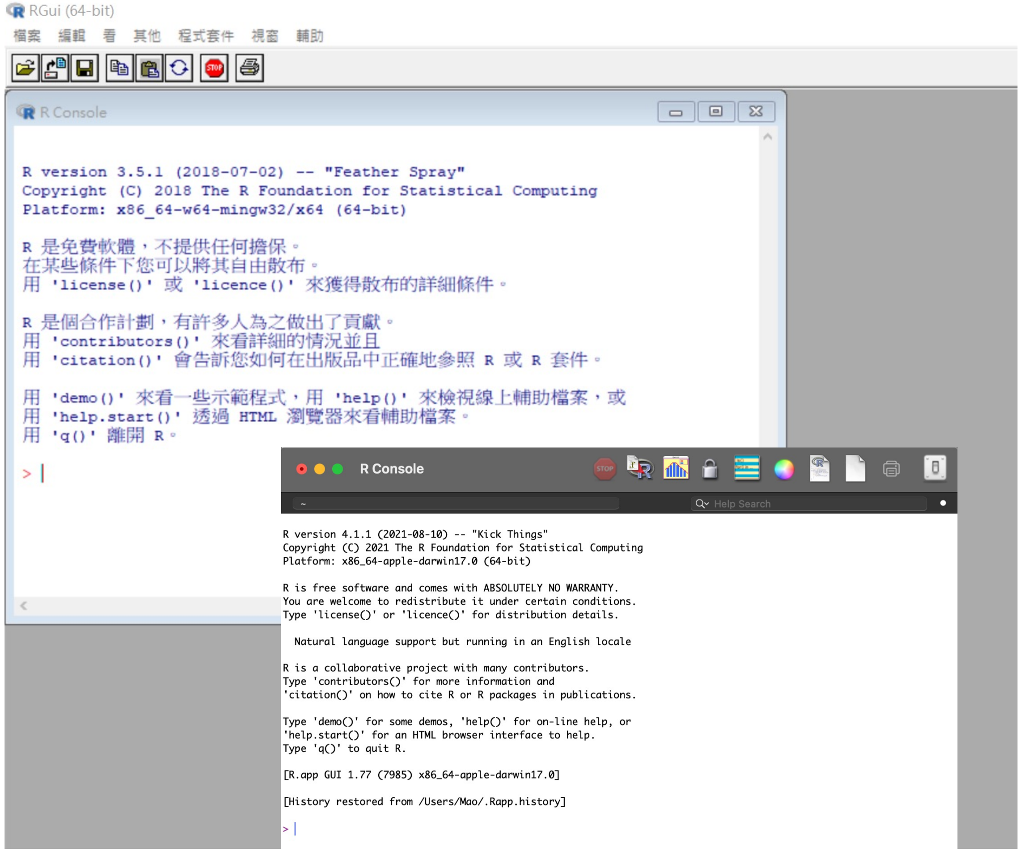
- R
  R for Windows
  (https://cran.csie.ntu.edu.tw/bin/windows/base/)
  R for Mac
  (https://cran.csie.ntu.edu.tw/bin/macosx/)

- **RStudio （請先安裝R）**
  (https://www.rstudio.com/products/rstudio/download/#download)

- **https://bids.github.io/2019-01-17-bids/ (安裝介紹影片在最下方)**

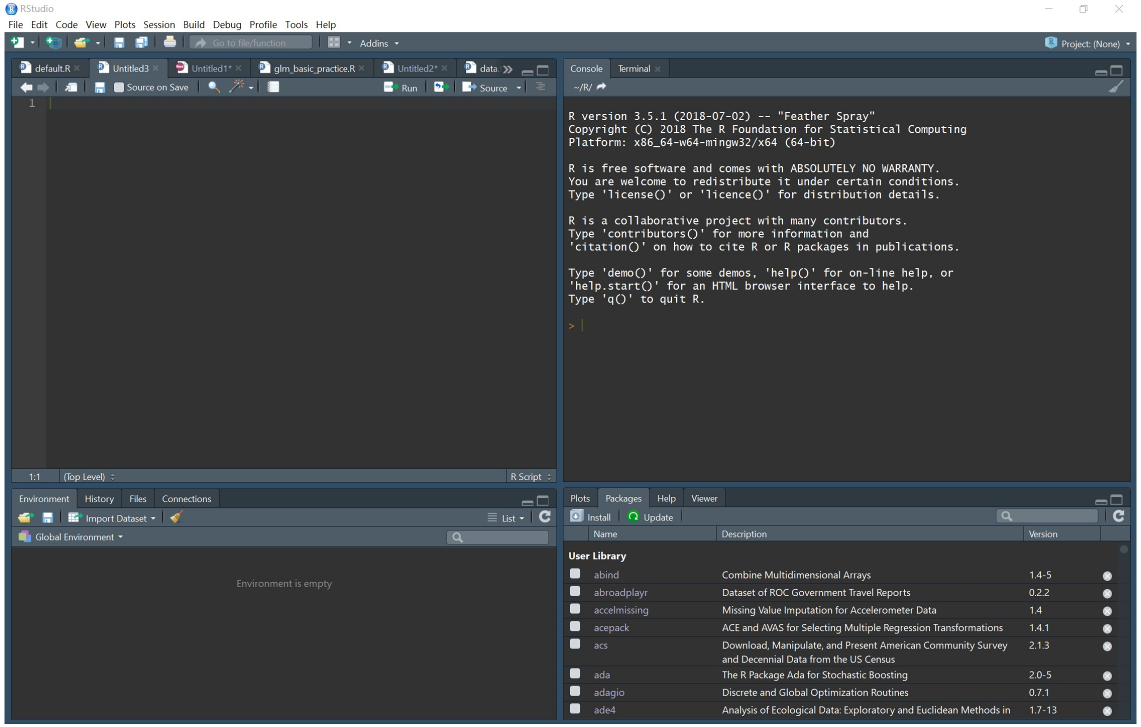
R

RStudio (IDE)



```
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R 是免費軟體，不提供任何擔保。
在某些條件下您可以將其自由散布。
用 'license()' 或 'licence()' 來獲得散布的詳細條件。

R 是個合作計劃，有許多人為之做出了貢獻。
用 'contributors()' 來看詳細的情況並且
用 'citation()' 會告訴您如何在出版品中正確地參照 R 或 R 套件。

用 'demo()' 來看一些示範程式，用 'help()' 來檢視線上輔助檔案，或
用 'help.start()' 透過 HTML 瀏覽器來看輔助檔案。
用 'q()' 離開 R。

> |
```

```
R version 4.1.1 (2021-08-10) -- "Kick Things"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (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.

  Natural language support but running in an English locale

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.

[R.app GUI 1.77 (7985) x86_64-apple-darwin17.0]

[History restored from /Users/Mao/.Rapp.history]

> |
```



```
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (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.

> |
```

	Name	Description	Version
User Library			
□	abind	Combine Multidimensional Arrays	1.4-5
□	abroadplayr	Dataset of ROC Government Travel Reports	0.2.2
□	accelmissing	Missing Value Imputation for Accelerometer Data	1.4
□	acepack	ACE and AVAS for Selecting Multiple Regression Transformations	1.4.1
□	acs	Download, Manipulate, and Present American Community Survey and Decennial Data from the US Census	2.1.3
□	ada	The R Package Ada for Stochastic Boosting	2.0-5
□	adagio	Discrete and Global Optimization Routines	0.7.1
□	ade4	Analysis of Ecological Data: Exploratory and Euclidean Methods in	1.7-13

為什麼要學程式語言？

- 簡化繁瑣重複的工作
- make your life easier

為什麼程式語言要學 R？

- 免費！
- 功能強大（各種分析都可以辦到，繪圖尤其強大）。
- 易學（對人文社會科學背景者較容易）。
- 友善的學習社群。

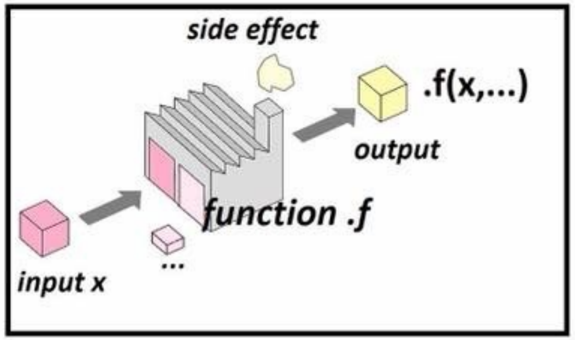
R 語言

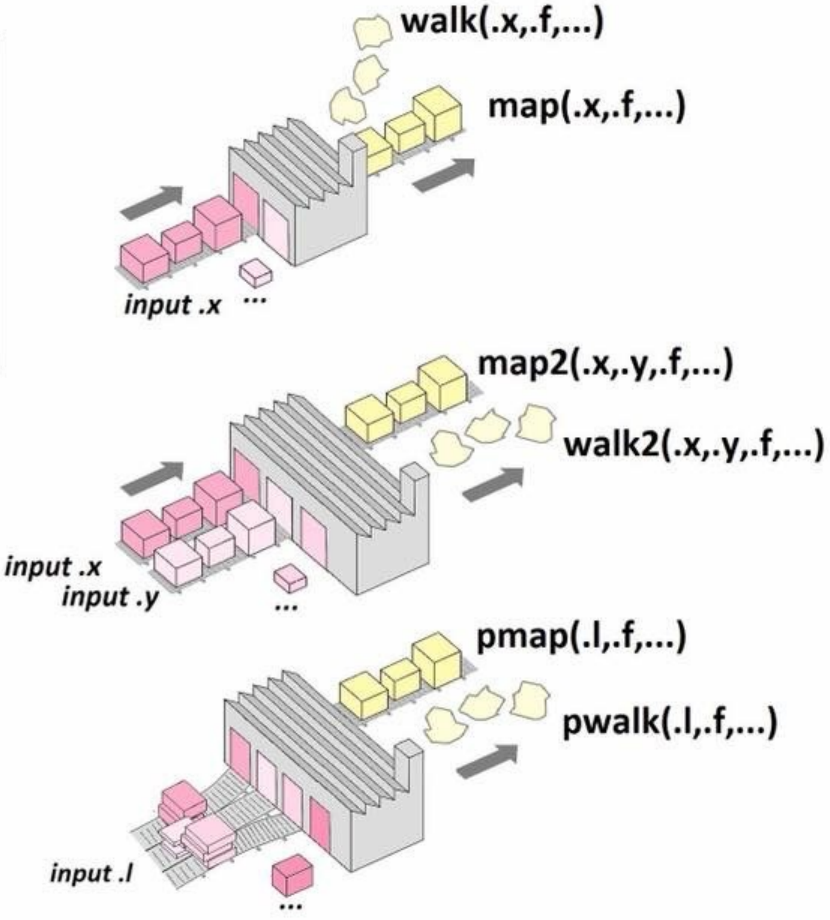
- Ross Ihaka

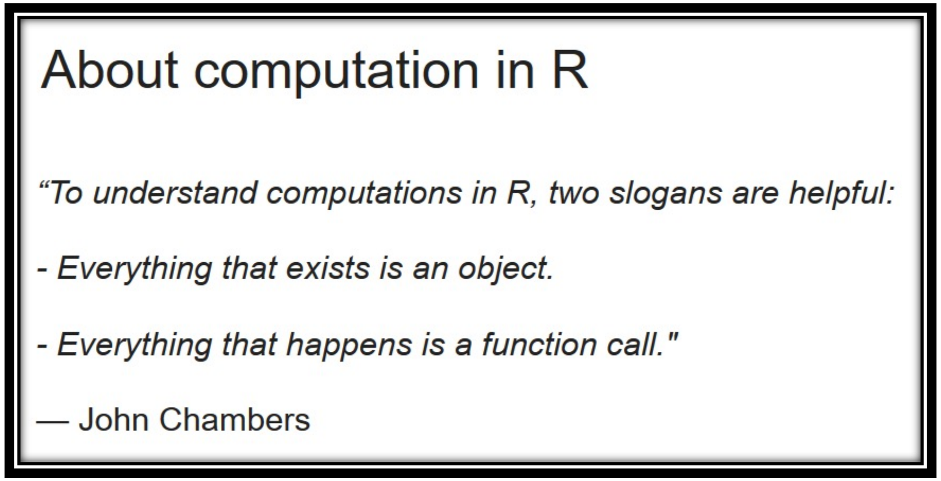
- Robert C. Gentleman

R 語言的基本要素

- 物件（Object）
- 函式（Function）







About computation in R

“To understand computations in R, two slogans are helpful:

- Everything that exists is an object.

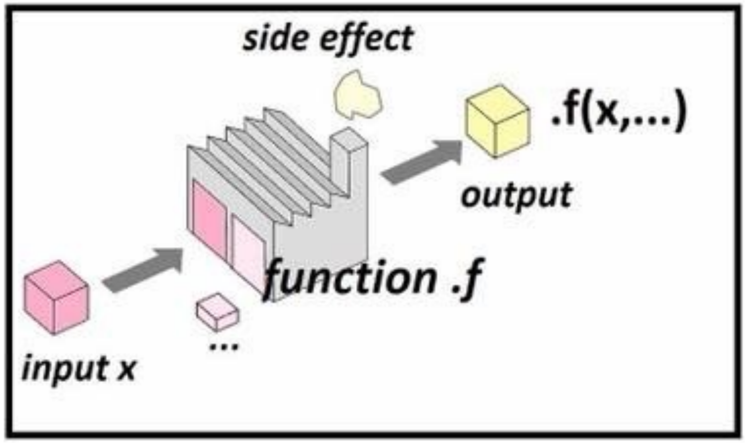
- Everything that happens is a function call."

— John Chambers

R 語言範例

- print(x,

  digits = getOption("digits"), ...)



- 參數（Arguments）

object an object for which a summary is desired.

digits minimal number of significant digits

- R語言的程式碼是**有區分大小寫**（case sensitive）

R 語言參數說明

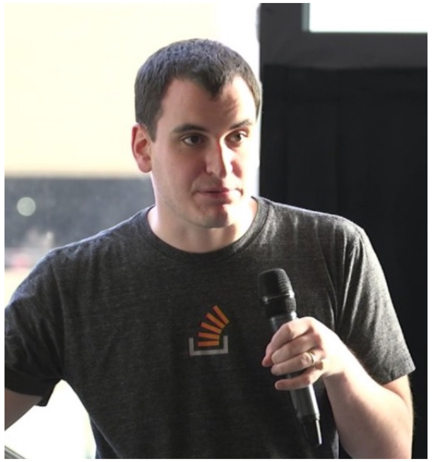
- 必要參數
- x a numeric vector, matrix or data frame
- 預設參數
- y NULL (default) or a vector, matrix or data frame with compatible dimensions to x. The default is equivalent to y = x (but more efficient).

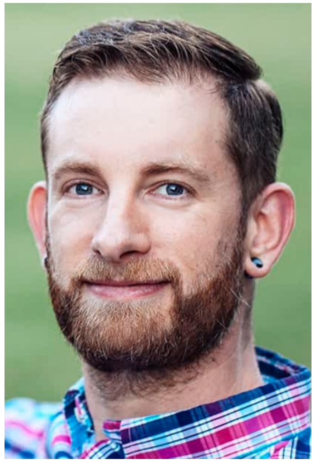
- 依參數位置（position）
- 依參數名稱

```
cor(  x,
      y = NULL,
      use = "everything",
      method = c("pearson", "kendall", "spearman"))
```

R 語言開發生態

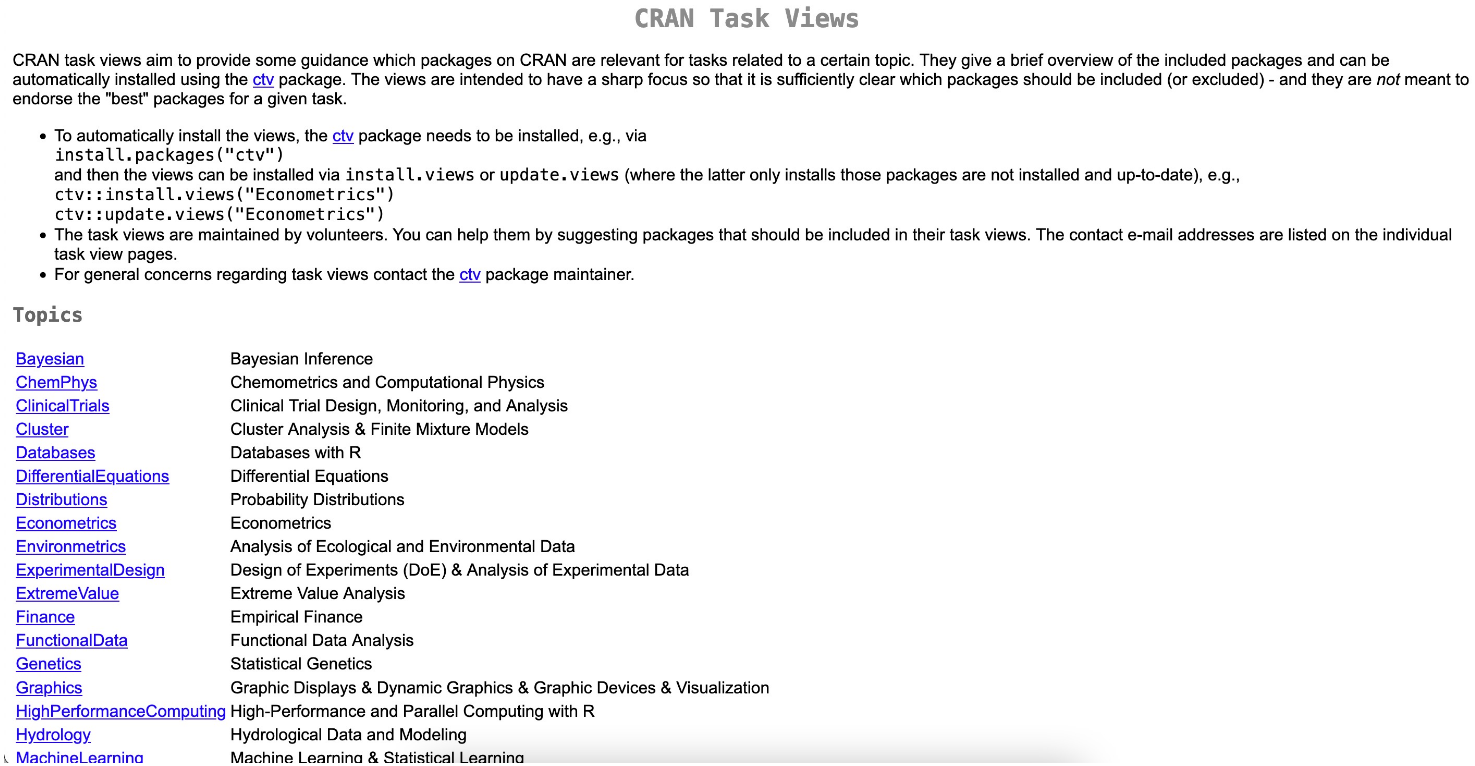
- R Development Core Team (Base R)
- 其他套件（ package ）製作者





R 能做什麼事？

- 統計分析
- 文字探勘
- 社會網絡分析
- 空間分析
- 網路爬蟲
- ……



CRAN Task Views

CRAN task views aim to provide some guidance which packages on CRAN are relevant for tasks related to a certain topic. They give a brief overview of the included packages and can be automatically installed using the ctv package. The views are intended to have a sharp focus so that it is sufficiently clear which packages should be included (or excluded) - and they are *not* meant to endorse the "best" packages for a given task.

- To automatically install the views, the ctv package needs to be installed, e.g., via
  `install.packages("ctv")`
  and then the views can be installed via `install.views` or `update.views` (where the latter only installs those packages are not installed and up-to-date), e.g.,
  ```
  ctv::install.views("Econometrics")
  ctv::update.views("Econometrics")
  ```
- The task views are maintained by volunteers. You can help them by suggesting packages that should be included in their task views. The contact e-mail addresses are listed on the individual task view pages.
- For general concerns regarding task views contact the ctv package maintainer.

Topics

Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
Databases	Databases with R
DifferentialEquations	Differential Equations
Distributions	Probability Distributions
Econometrics	Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data
ExtremeValue	Extreme Value Analysis
Finance	Empirical Finance
FunctionalData	Functional Data Analysis
Genetics	Statistical Genetics
Graphics	Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
HighPerformanceComputing	High-Performance and Parallel Computing with R
Hydrology	Hydrological Data and Modeling
MachineLearning	Machine Learning & Statistical Learning

https://cran.r-project.org/web/views/

為什麼有人覺得 R 不好學？

- **難處1：要記各種函式 (function)**
- 解答：沒有人真的能全部記住，重點是該函式的說明文件（documentation）清楚嗎？

- **難處2：入門的門檻不低，要學的套件（package）太多！**
- 解答：理解基本的資料結構，可以降低後續學習的門檻；學習具有同樣設計邏輯的套件（如tidyverse）。

怎麼問問題？

- **Rep**roducible **ex**ample (**reprex**) https://github.com/tidyverse/reprex

- 讓別人能最小化的**重現**你的問題，才能夠幫你處理問題。
  - A minimal dataset, necessary to reproduce the error
  - The minimal runnable code necessary to reproduce the error, which can be run on the given dataset.

https://stackoverflow.com/questions/5963269/how-to-make-a-great-r-reproducible-example

Coding style

- https://style.tidyverse.org/
- 讓你自己及合作者更容易讀你的code。

其他學習資源

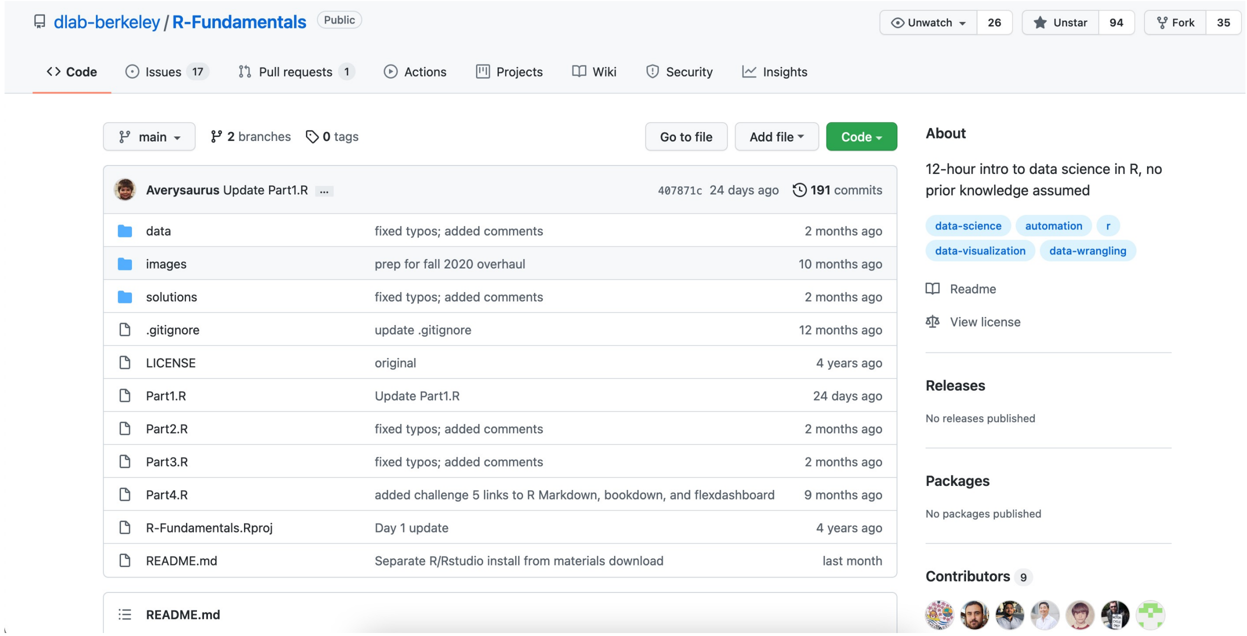
- DataCamp
- Coursera
- edX
- Cheat sheet
  https://www.rstudio.com/resources/cheatsheets/

R 語言基礎

- R語言基礎：簡介
- R語言基礎：資料篩選與整理
- R語言基礎：資料探索與分析
- R語言基礎：程式設計基礎

- R語言進階：資料整理

The Social Science Data Lab at UC Berkeley



dlab-berkeley / R-Fundamentals

2 branches 0 tags

Averysaurus Update Part1.R ... 407871c 24 days ago 191 commits

data	fixed typos; added comments	2 months ago
images	prep for fall 2020 overhaul	10 months ago
solutions	fixed typos; added comments	2 months ago
.gitignore	update .gitignore	12 months ago
LICENSE	original	4 years ago
Part1.R	Update Part1.R	24 days ago
Part2.R	fixed typos; added comments	2 months ago
Part3.R	fixed typos; added comments	2 months ago
Part4.R	added challenge 5 links to R Markdown, bookdown, and flexdashboard	9 months ago
R-Fundamentals.Rproj	Day 1 update	4 years ago
README.md	Separate R/Rstudio install from materials download	last month

README.md

About

12-hour intro to data science in R, no prior knowledge assumed

data-science automation r data-visualization data-wrangling

Readme

View license

Releases

No releases published

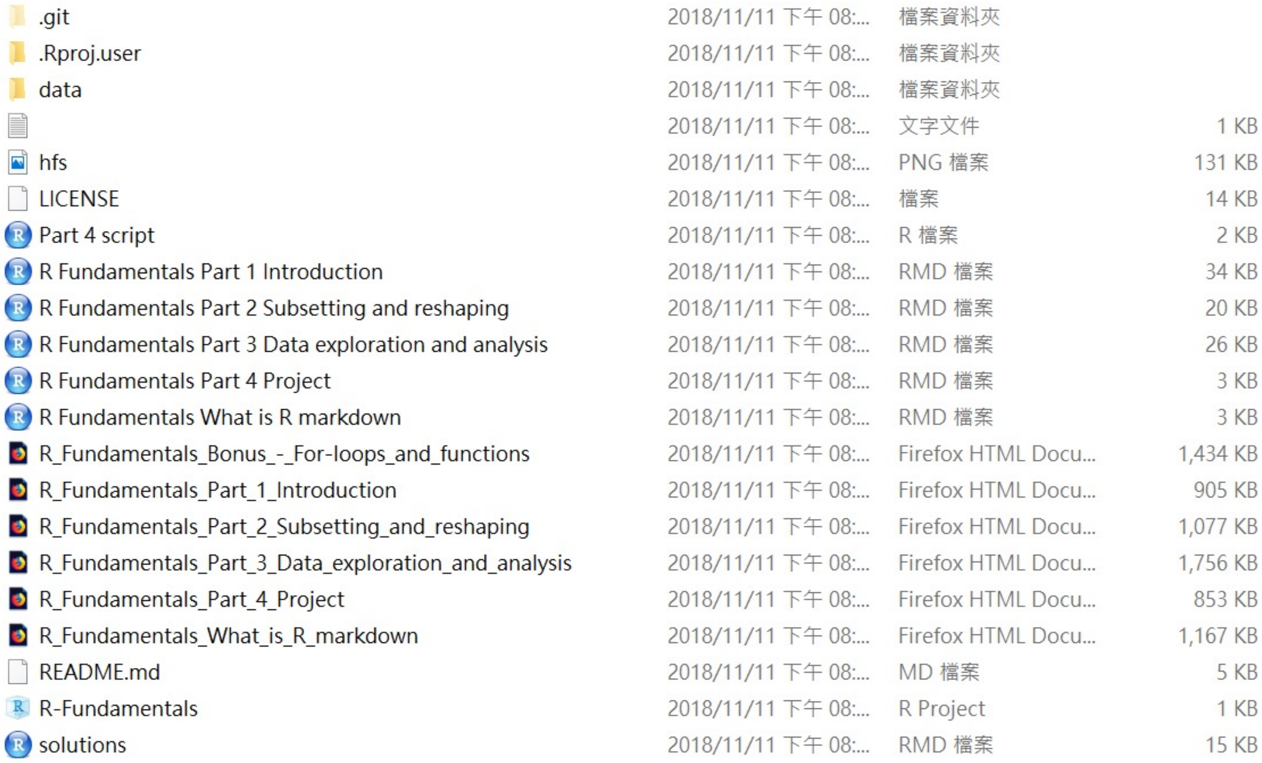
Packages

No packages published

Contributors 9

實際操作

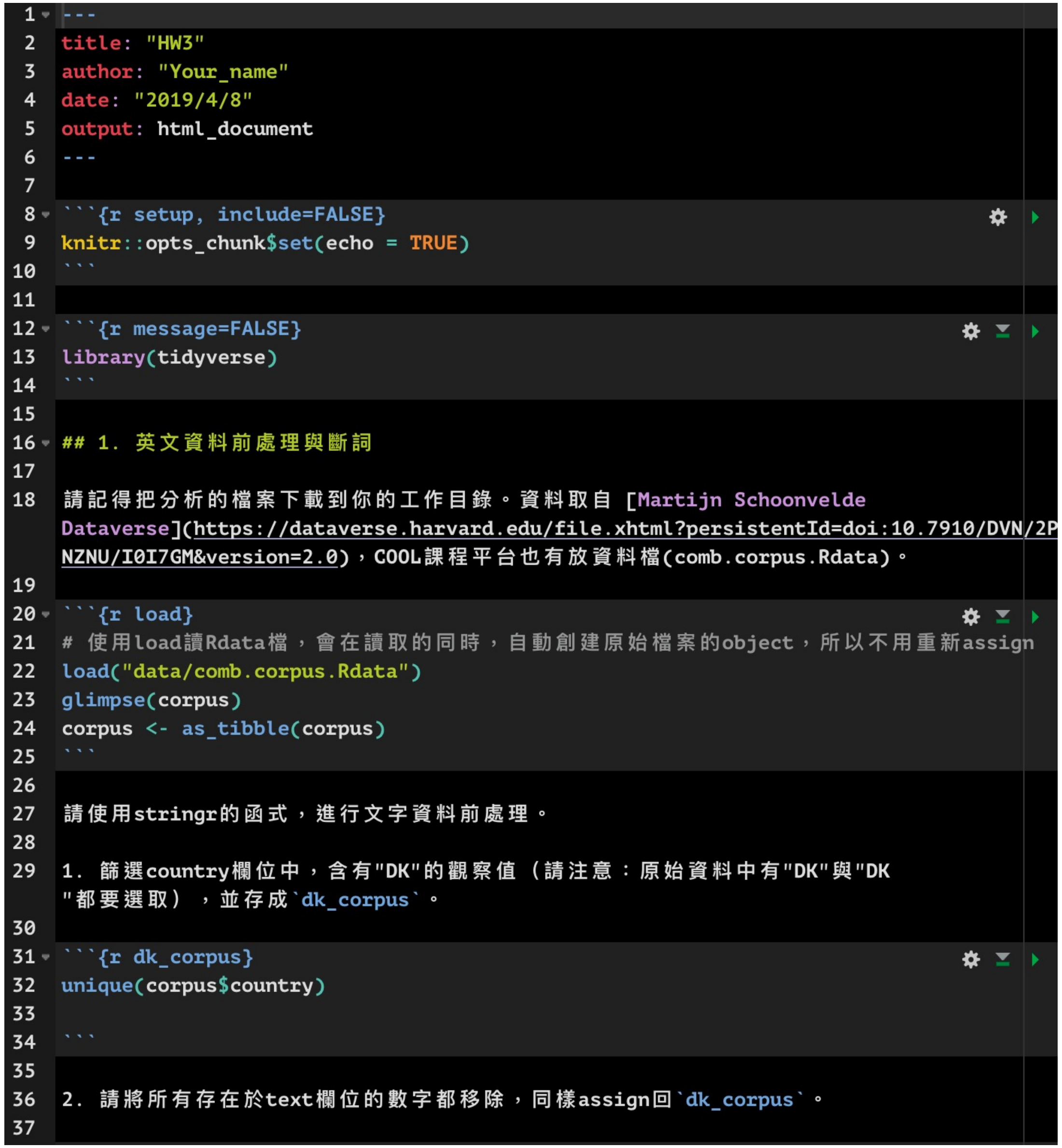
- R Markdown files



.git	2018/11/11 下午 08:...	檔案資料夾	
.Rproj.user	2018/11/11 下午 08:...	檔案資料夾	
data	2018/11/11 下午 08:...	檔案資料夾	
	2018/11/11 下午 08:...	文字文件	1 KB
hfs	2018/11/11 下午 08:...	PNG 檔案	131 KB
LICENSE	2018/11/11 下午 08:...	檔案	14 KB
Part 4 script	2018/11/11 下午 08:...	R 檔案	2 KB
R Fundamentals Part 1 Introduction	2018/11/11 下午 08:...	RMD 檔案	34 KB
R Fundamentals Part 2 Subsetting and reshaping	2018/11/11 下午 08:...	RMD 檔案	20 KB
R Fundamentals Part 3 Data exploration and analysis	2018/11/11 下午 08:...	RMD 檔案	26 KB
R Fundamentals Part 4 Project	2018/11/11 下午 08:...	RMD 檔案	3 KB
R Fundamentals What is R markdown	2018/11/11 下午 08:...	RMD 檔案	3 KB
R_Fundamentals_Bonus_-_For-loops_and_functions	2018/11/11 下午 08:...	Firefox HTML Docu...	1,434 KB
R_Fundamentals_Part_1_Introduction	2018/11/11 下午 08:...	Firefox HTML Docu...	905 KB
R_Fundamentals_Part_2_Subsetting_and_reshaping	2018/11/11 下午 08:...	Firefox HTML Docu...	1,077 KB
R_Fundamentals_Part_3_Data_exploration_and_analysis	2018/11/11 下午 08:...	Firefox HTML Docu...	1,756 KB
R_Fundamentals_Part_4_Project	2018/11/11 下午 08:...	Firefox HTML Docu...	853 KB
R_Fundamentals_What_is_R_markdown	2018/11/11 下午 08:...	Firefox HTML Docu...	1,167 KB
README.md	2018/11/11 下午 08:...	MD 檔案	5 KB
R-Fundamentals	2018/11/11 下午 08:...	R Project	1 KB
solutions	2018/11/11 下午 08:...	RMD 檔案	15 KB

作業範例

- Introduction to R https://www.datacamp.com/courses/free-introduction-to-r
- Intermediate R https://www.datacamp.com/courses/intermediate-r

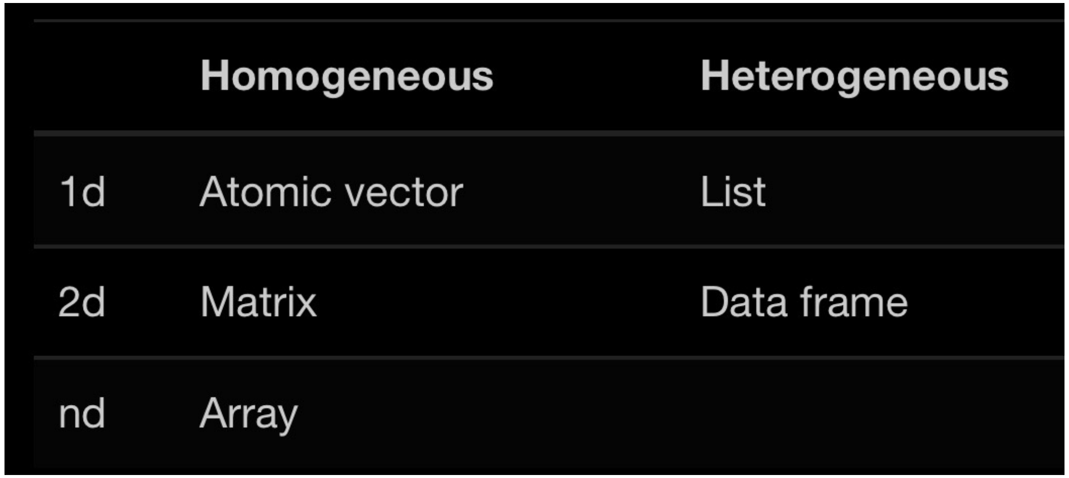


~~~
---
title: "HW3"
author: "Your_name"
date: "2019/4/8"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

```{r message=FALSE}
library(tidyverse)
```

## 1. 英文資料前處理與斷詞

請記得把分析的檔案下載到你的工作目錄。資料取自 [Martijn Schoonvelde Dataverse](https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/2PNZNU/I0I7GM&version=2.0)，COOL課程平台也有放資料檔(comb.corpus.Rdata)。

```{r load}
# 使用load讀Rdata檔，會在讀取的同時，自動創建原始檔案的object，所以不用重新assign
load("data/comb.corpus.Rdata")
glimpse(corpus)
corpus <- as_tibble(corpus)
```

請使用stringr的函式，進行文字資料前處理。

1. 篩選country欄位中，含有"DK"的觀察值（請注意：原始資料中有"DK"與"DK"都要選取），並存成`dk_corpus`。

```{r dk_corpus}
unique(corpus$country)

```

2. 請將所有存在於text欄位的數字都移除，同樣assign回`dk_corpus`。

~~~

資料類型

- 數值（ numeric ）
- 字串（ character, string ）
- 邏輯判斷（ logical ）

儲存格式



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

什麼是文字探勘？

- Github repository
- https://github.com/aleszu/textanalysis-shiny

- Shiny App實做 https://storybench.shinyapps.io/textanalysis/