# 第 16 天

## 撰寫資料分析報告

我們終於來到最後一天，能夠讀完前面 15 天的內容真的很不簡單！您此刻已經是個程度很不錯的 R 語言使用者，或許您認識的函數、套件還不夠多、語法還不夠熟悉，但那些都會隨著經驗而成長，我們一點都不擔心這個部分。接下來我們希望您可以開始將 R 語言應用到學業或者工作之中，所以在最後一個章節中，我們將說明如何在 RStudio 之中撰寫副檔名為 **.Rmd** 的純文字，並且利用 knitr 套件將這種純文字轉換為資料分析報告。

### **.Rmd** 是什麼？

在說明 **.Rmd** 之前，我們得先提一下 **Markdown** 這個輕量的標示語言（Markup Language），它的目的是使用易於閱讀、撰寫的純文字格式，並能夠選擇性地轉換成以**超文件標示語言（HyperText Markup Language，HTML）**為主的其他檔案格式，其中知名的轉換器為 **Pandoc**，它可以將 **.md** 的純文字檔案轉換為許多的文件格式，包含像是 HTML、Word 文件與 PDF 等。

**.Rmd** 這個副檔名全名為 **R markdown**，撰寫語法跟 **Markdown** 相似度非常高，而 knitr 套件之於 **R markdown** 就像是 **Pandoc** 之於 **Markdown** 一樣，可以轉換成包含 R 語言程式、輸出與圖形的不同文件格式。簡單來說，R 語言使用者可以在 **.Rmd** 純文字檔案中同時寫程式與撰寫說明內容，不需要另外將程式、輸出與圖形另外複製貼上或匯出至文書編輯軟體中。

### 新增一個 R Markdown 檔案

在 RStudio 的選單點選新增一個 R Markdown 文件：

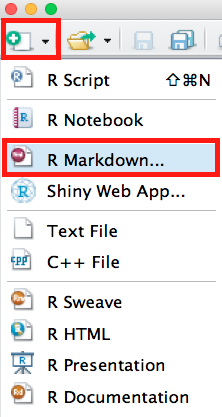


圖 16-1 新增一個 R Markdown 檔案

我們選第一種大類：Document，接著輸入這個文件的標題、作者以及輸出的格式（先選擇 HTML），這些資訊隨時都可以修正：

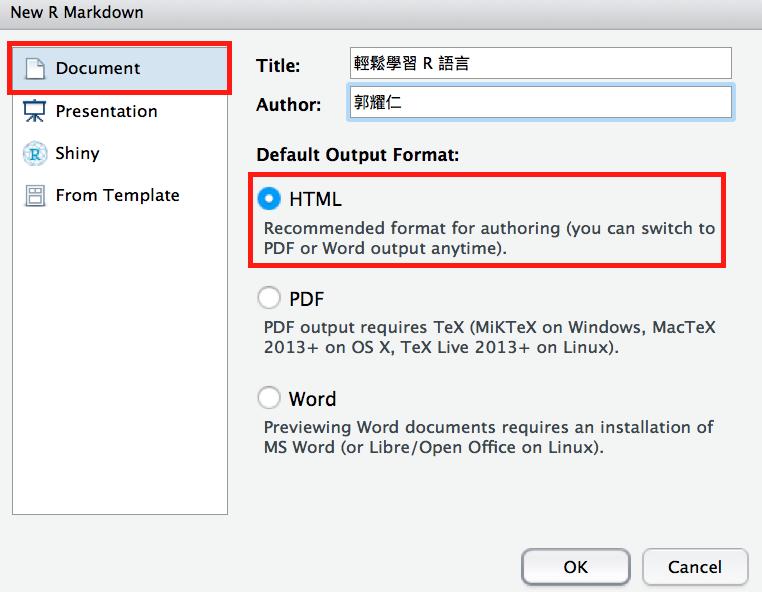


圖 16-2 新增一個 R Markdown 檔案（2）

如果是第一次新增 R Markdown 文件的使用者，RStudio 應該會出現提示訊息安裝 knitr 等相關套件，這時點選同意安裝就可以，新增完畢之後我們就會看到一個有範例說明的 R Markdown 文件：

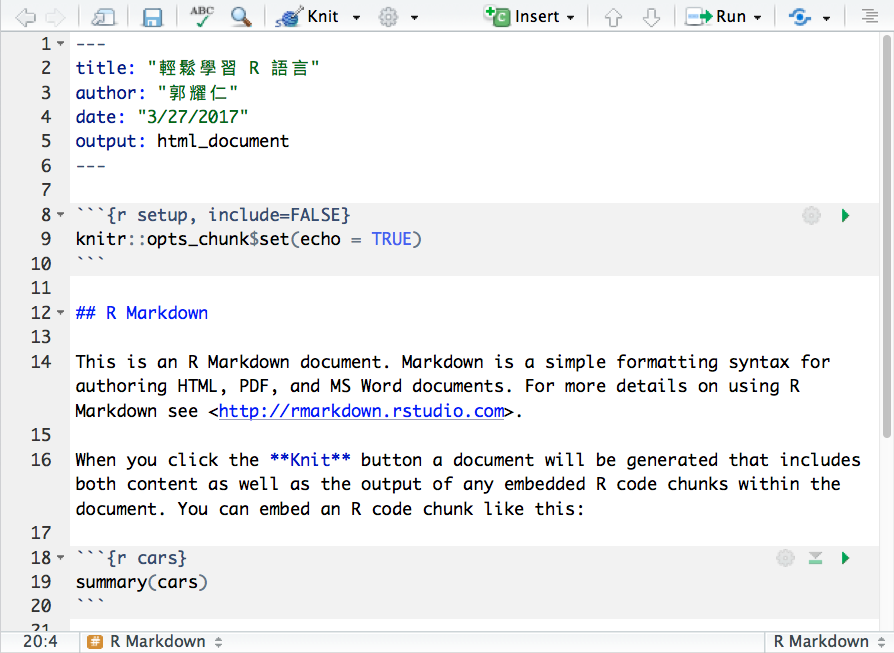


圖 16-3 新增一個 R Markdown 檔案（3）

這時我們可看到在文件上方用 --- 區隔出來的區塊是這個文件的資訊，我們可以任意修改標題、作者、製作日期以及輸出格式，如果希望輸出格式為 PDF 就改為 output: pdf\_document，如果希望輸出格式為 MS Word 就改為 output: word\_document，若指定這兩種格式，必須確定您的電腦有安裝 **TeX** 與 **MS Word**。目前這個檔案已經是一個可以準備輸出的狀態，我們接著點選 RStudio 選單的 **Knit**：

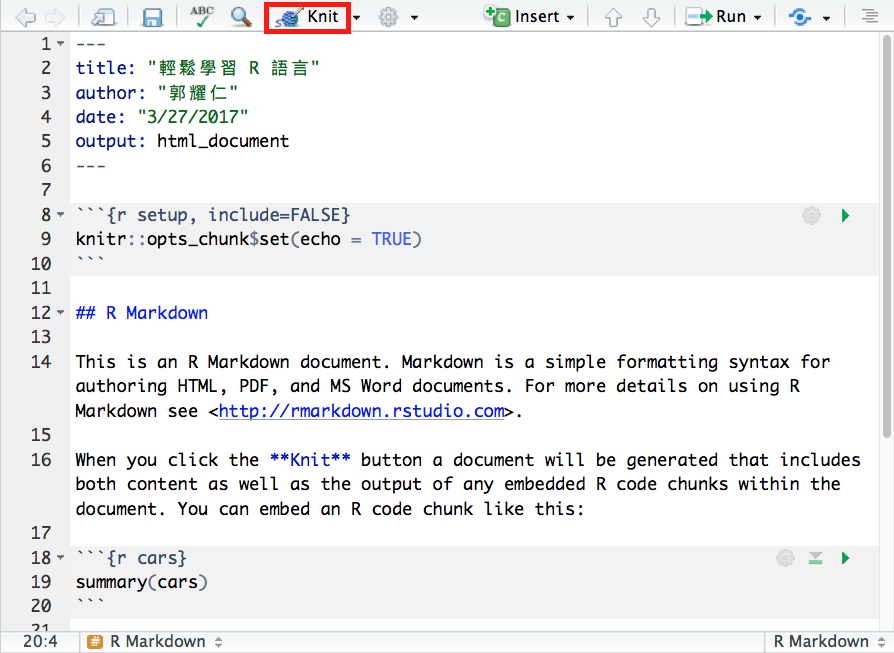


圖 16-4 點選 Knit

這時 RStudio 會要我們先儲存 R Markdown 文件，並且指定一個編碼格式，使用繁體中文我們通常都選擇 **UTF-8** 或者 **BIG5** 編碼：

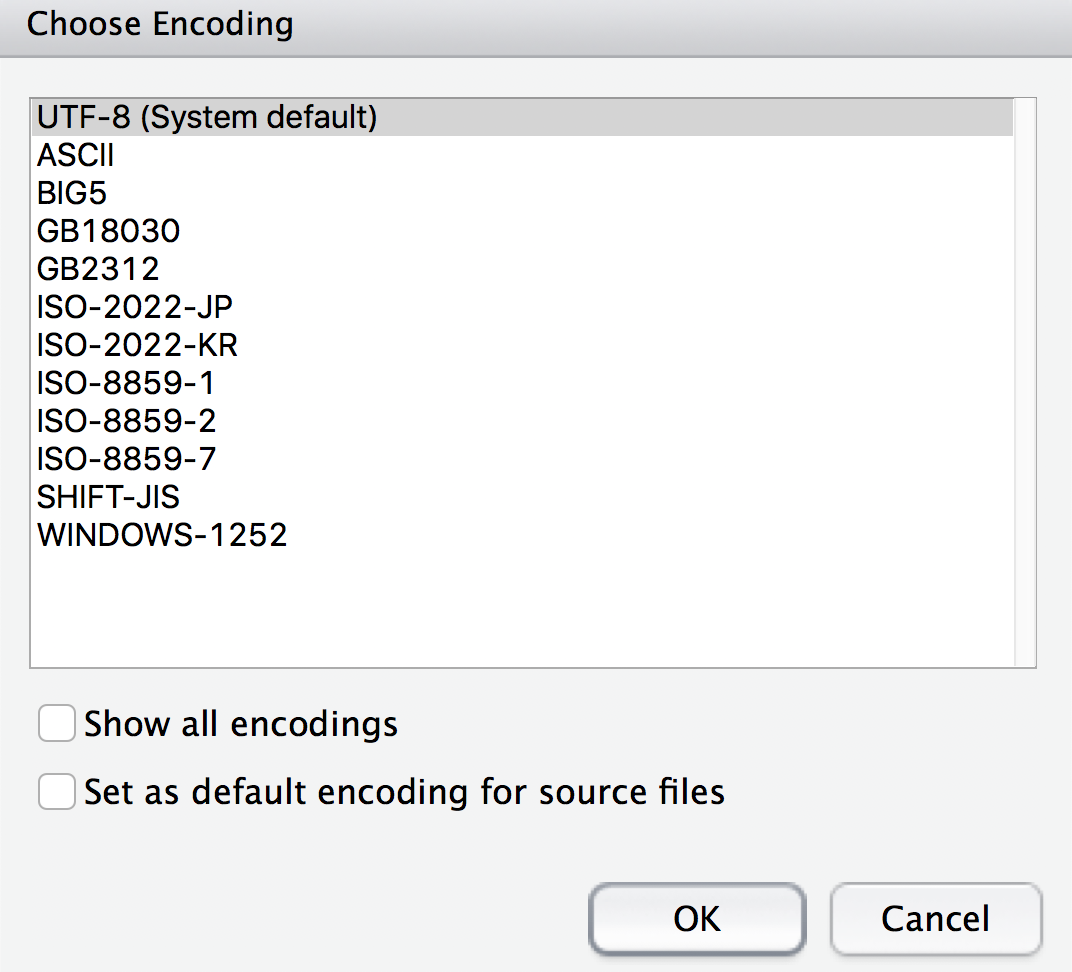


圖 16-5 編碼

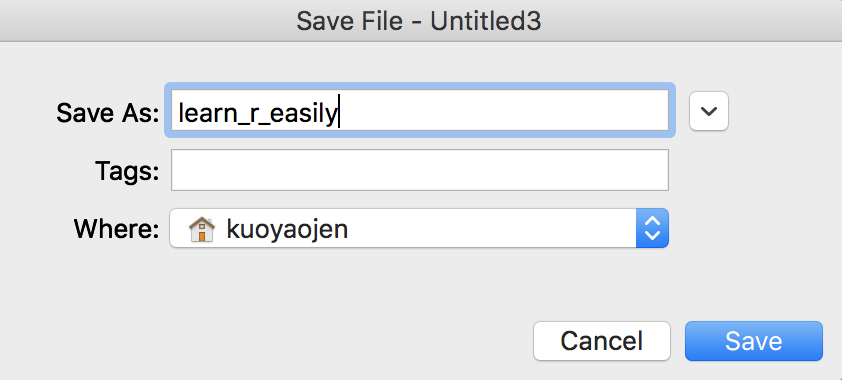


圖 16-6 存檔

我們先選擇將 R Markdown 文件儲存在使用者的根目錄下，存檔完後 knitr 就會在同一個路徑輸出一個 HTML 檔案：

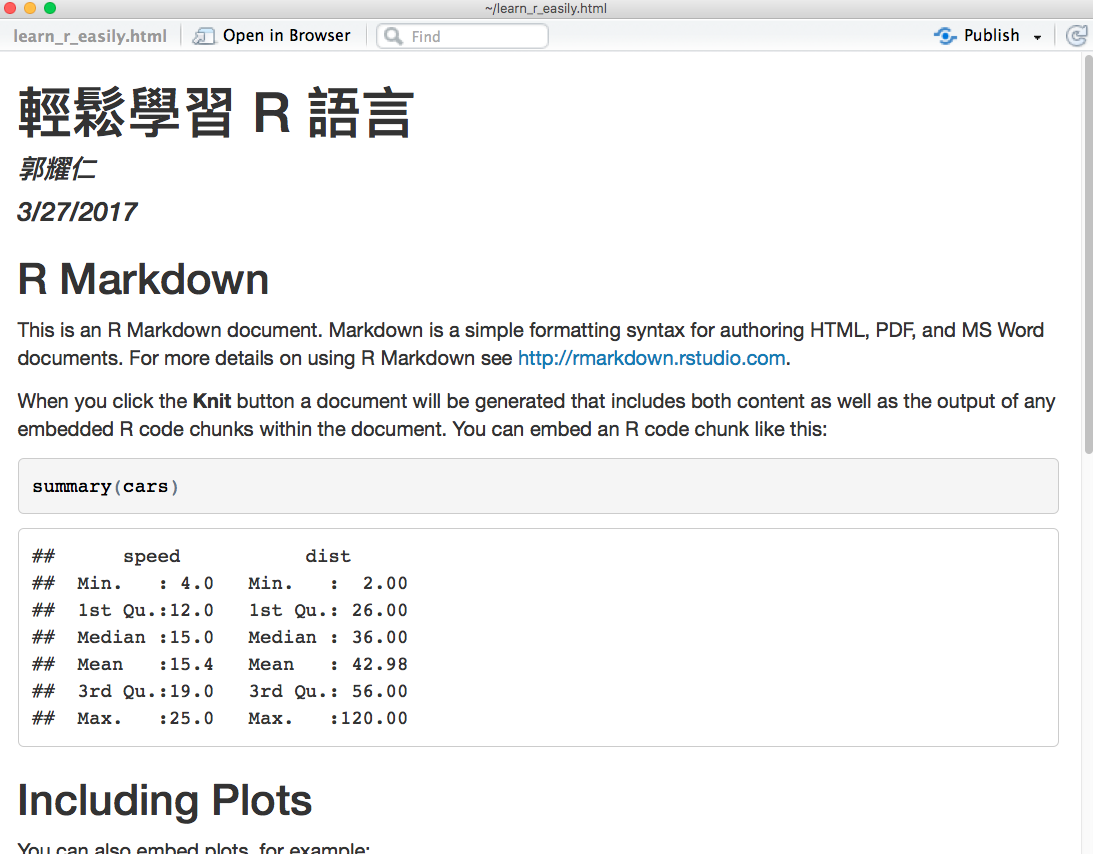


圖 16-7 輸出

裡面的內容是預設的編寫範例，接著我們要撰寫自己的內容。

### 基本的分析文件元素

#### 段落標題

我們的文件中每個段落都會有個標題，我們可以利用一個井字號 # 代表第一層標題，兩個井字號 ## 代表第二層標題，以此類推可以使用到第六層標題。

---  
title: "輕鬆學習 R 語言"  
author: "郭耀仁"  
date: "3/27/2017"  
output: html\_document  
---  
  
## 基本功  
  
### 開發環境與語法  
  
### 變數類型與資料結構  
  
### 迴圈與流程控制  
  
## 應用  
  
### 讀取資料  
  
### 探索性資料分析

輸出結果如下：

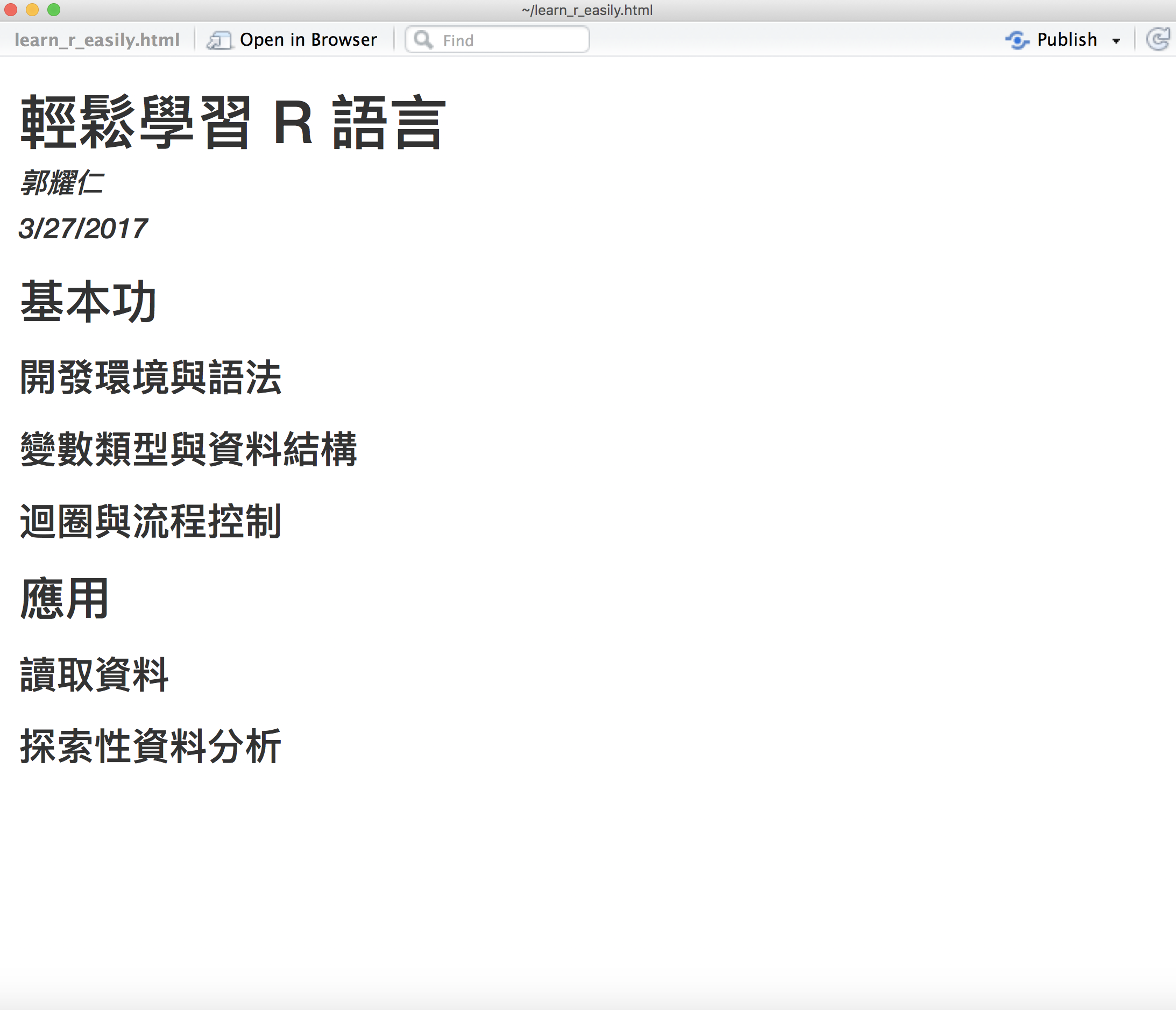


圖 16-8 段落標題

#### 段落內文

接著開始撰寫內文，直接輸入文字即為段落內文，還可以使用 \*\*粗體\*\* 或者 \*斜體\* 強調內文中的重點：

## 基本功  
  
紮實的基本功包括對於\*\*變數類型\*\*與\*資料結構\*的掌握程度，以及懂得使用\*\*迴圈\*\*與\*流程控制\*撰寫邏輯清楚而簡潔的程式。  
  
### 開發環境與語法  
  
### 變數類型與資料結構  
  
### 迴圈與流程控制  
  
## 應用  
  
一但掌握了基本功，就需要\*\*大量\*\*的\*應用場景\*來磨練這些技巧。  
  
### 讀取資料  
  
### 探索性資料分析

輸出結果如下：

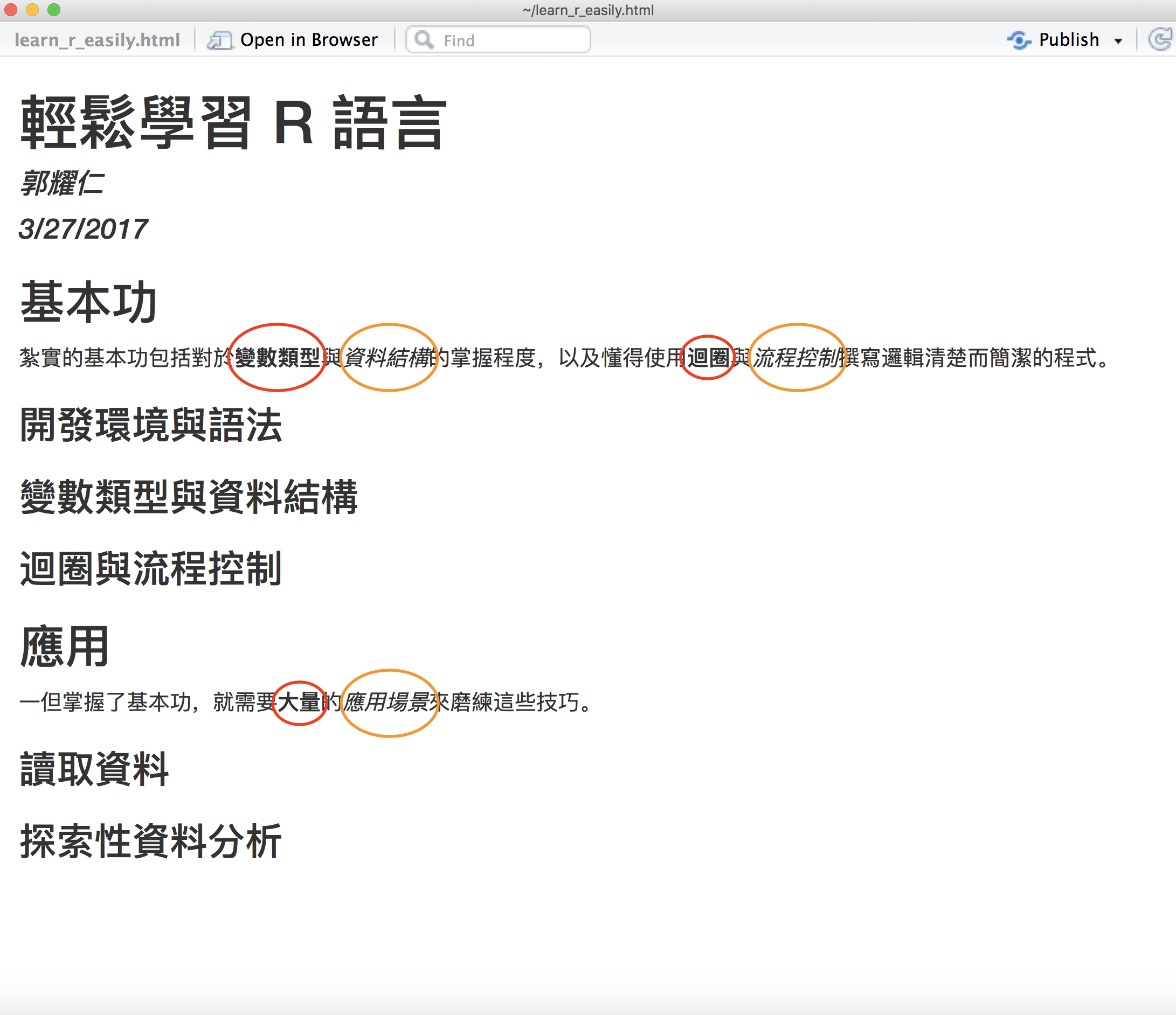


圖 16-9 段落內文

#### 清單

接著我們為文件加入清單，標點形式的清單可以利用 - 或者 \* 建立，並以 Tab 鍵（或四個空格）來建立巢狀結構的標點形式清單；編號形式的清單則直接輸入數字：

## 基本功  
  
紮實的基本功包括對於\*\*變數類型\*\*與\*資料結構\*的掌握程度，以及懂得使用\*\*迴圈\*\*與\*流程控制\*撰寫邏輯清楚而簡潔的程式。  
  
### 開發環境與語法  
  
1. 安裝 R  
2. 安裝 RStudio  
  
### 變數類型與資料結構  
  
- 變數類型：  
 - 數值  
 - numeric  
 - integer  
 - complex  
 - 邏輯值  
 - 文字  
  
- 資料結構：  
 - 一維  
 - 向量  
 - 因素向量  
 - 二維  
 - 矩陣  
 - 資料框  
 - 多維  
 - 陣列  
 - 清單

輸出結果如下：

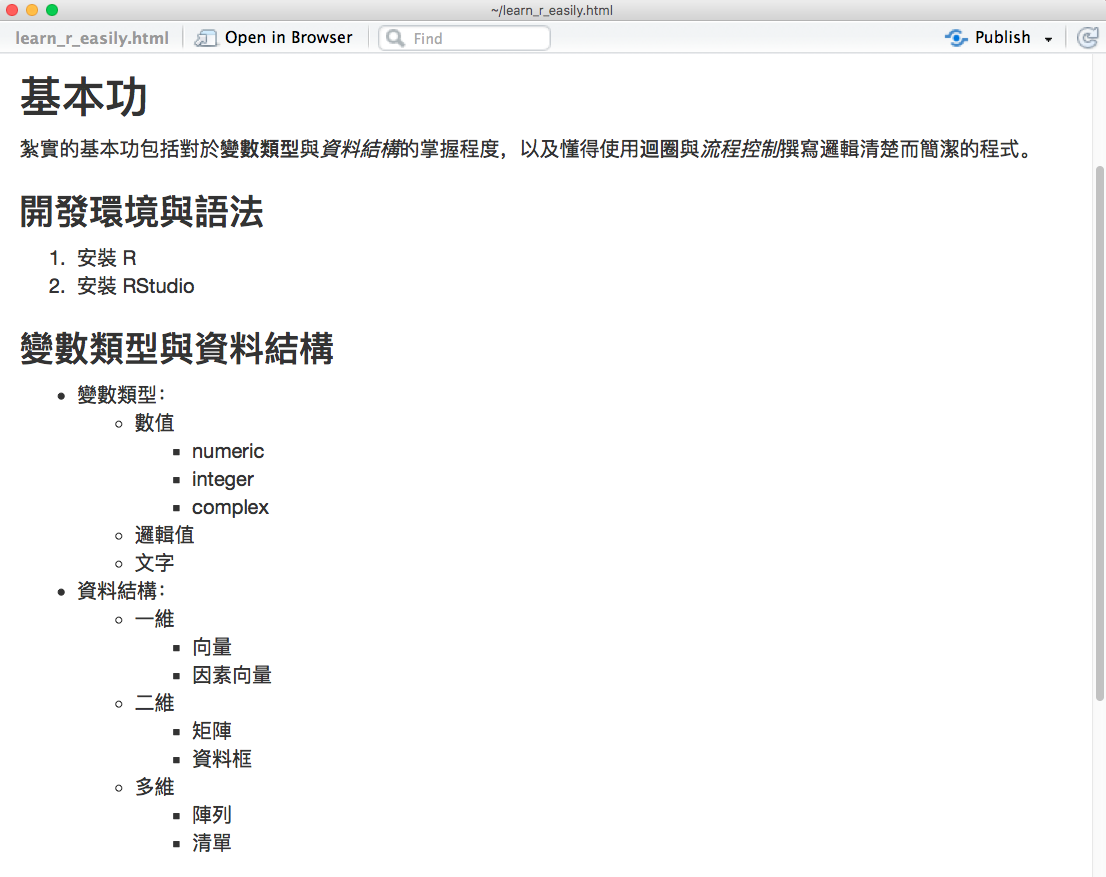


圖 16-10 清單

#### 程式區塊

資料分析文件中很常需要加入程式區塊（Code Chunks），我們使用 `` 將行內程式（Inline Code）包起來，或者使用 ``` 將程式區塊包起來，` 這個符號叫做 tilt，您可以在鍵盤最左上角的按鍵找到它，它在 tab 的上方，數字 1 的左方。

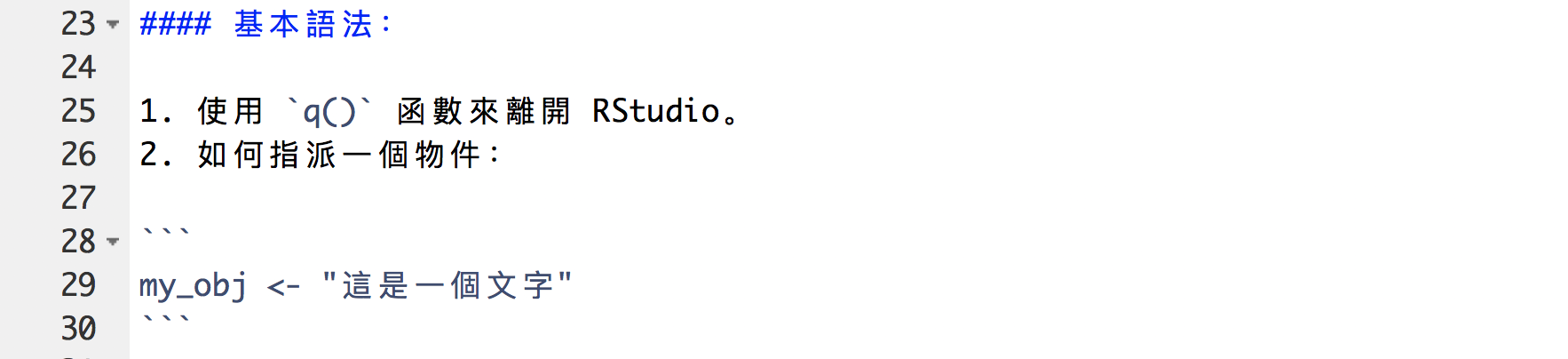


圖 16-11 程式區塊

輸出結果如下：

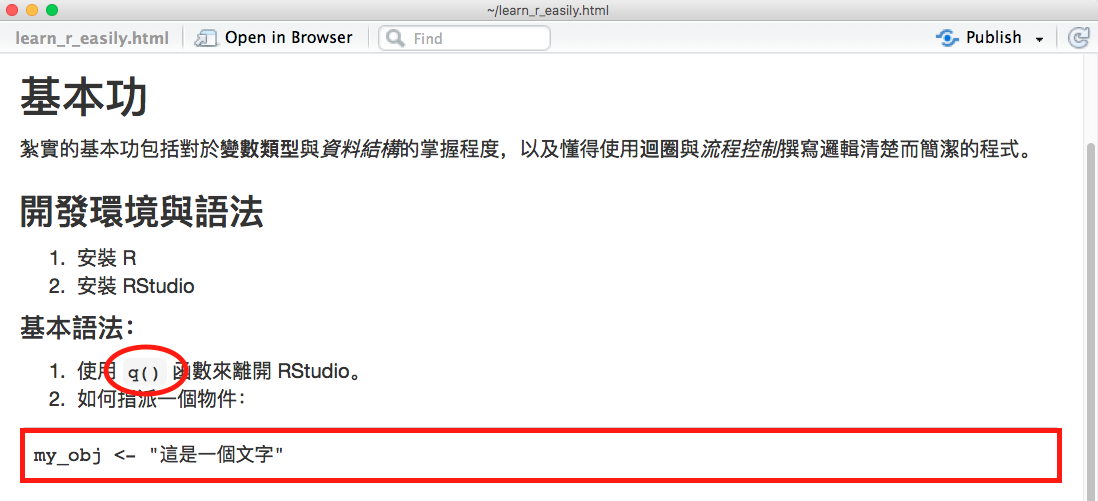


圖 16-12 程式區塊（2）

如果在程式區塊旁邊加註 {r}，在輸出的時候這段程式就會被 R 語言執行，並且將輸出結果或圖形直接顯示在程式區塊的下方，如此一來就可以達到不需仰賴其他文書編輯軟體的目的：

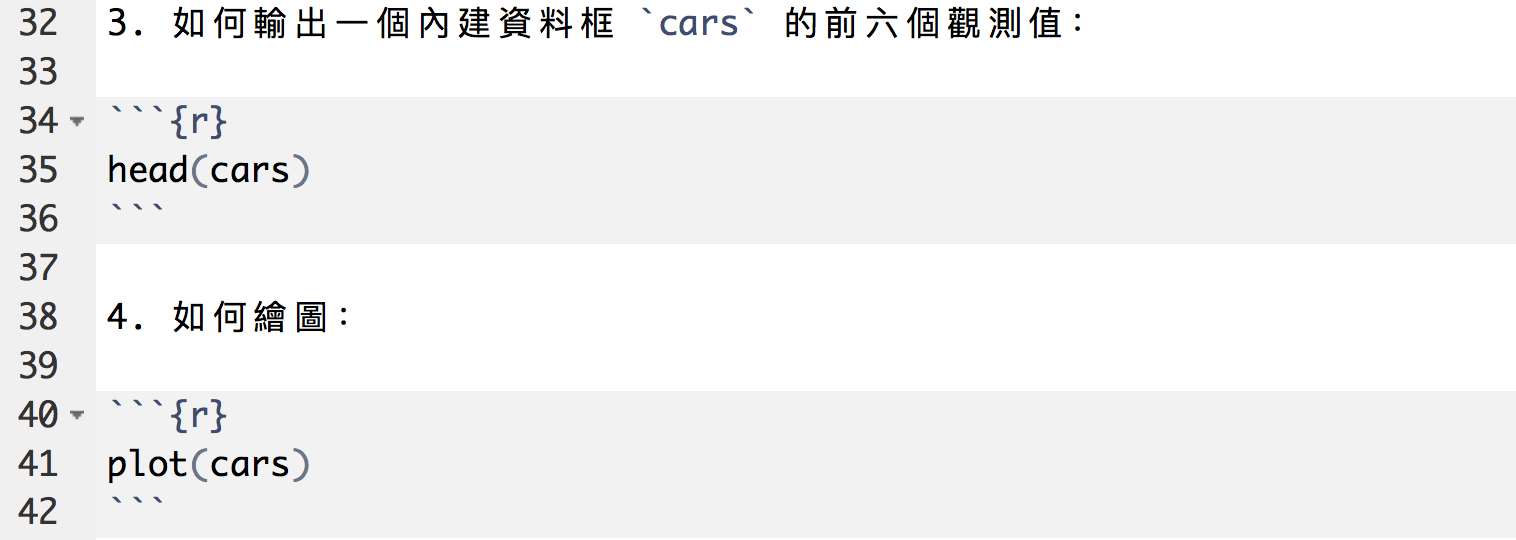


圖 16-13 程式區塊（3）

輸出結果如下：



圖 16-14 程式區塊（4）

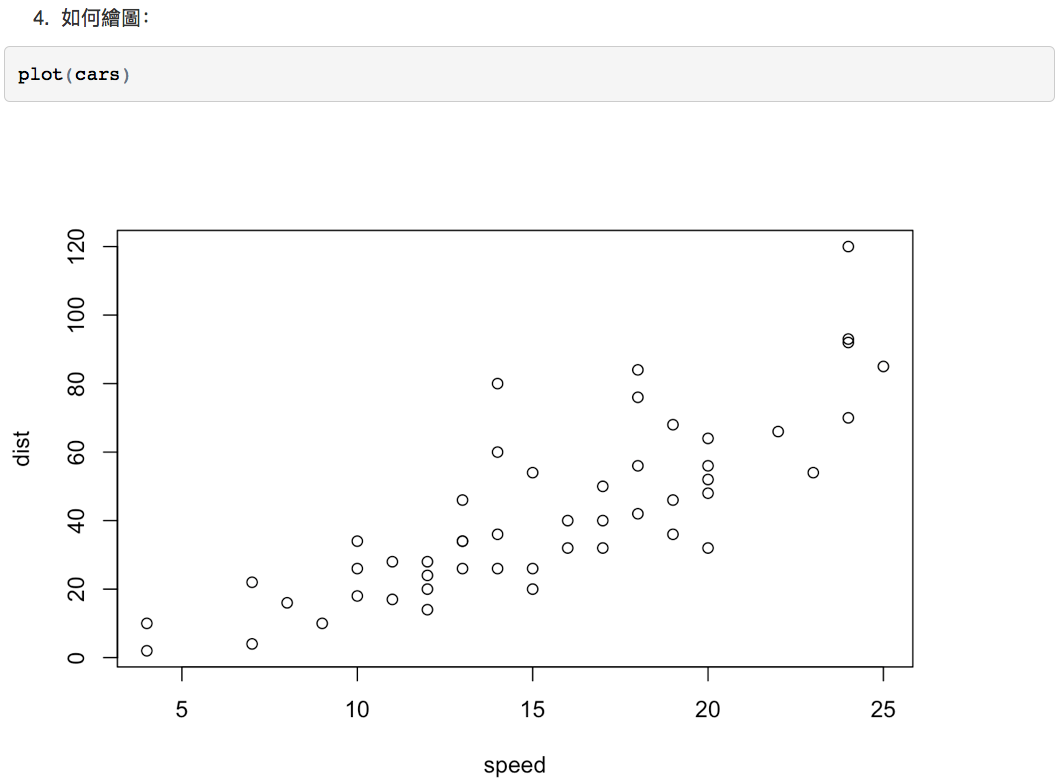


圖 16-15 程式區塊（5）

在 {r} 大括號中還可以加入更多的參數設定：

|  |  |
| --- | --- |
| 參數 | 設定內容 |
| echo | 是否要將程式顯示在文件中（預設為 TRUE） |
| results | 是否要將執行結果顯示在文件中（預設為 markup），可以另外選擇 asis、hold 或 hide |
| warning | 是否要將程式執行的警告訊息顯示在文件中（預設為 TRUE） |
| error | 是否允許有錯誤訊息的程式（預設為 FALSE） |
| message | 是否要將程式執行的回傳訊息顯示在文件中（預設為 TRUE） |

調整 echo 參數為 FALSE，只會將程式執行後的結果輸出在文件中，但是不會輸出程式碼：



圖 16-16 程式區塊（6）

輸出結果如下：

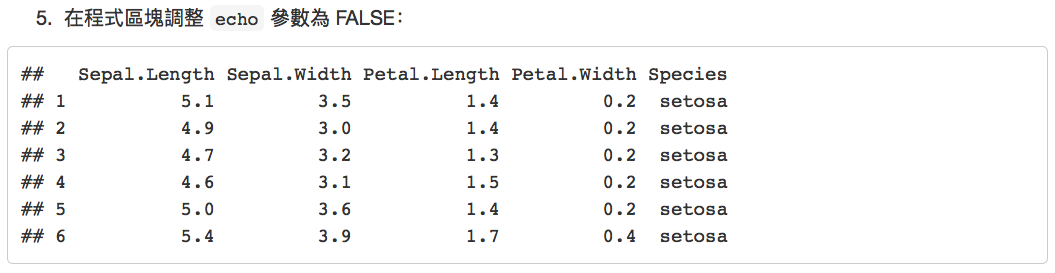


圖 16-17 程式區塊（7）

調整 results 參數為 "hide"，只會輸出程式碼在文件中但不會顯示執行後的結果：

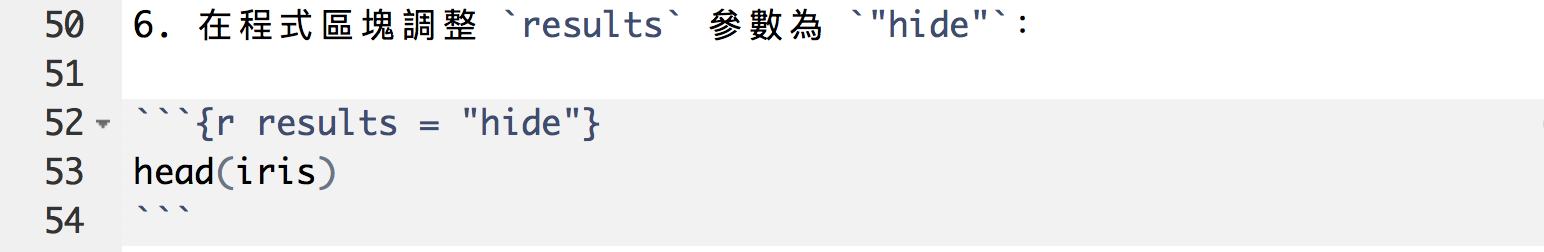


圖 16-18 程式區塊（8）

輸出結果如下：

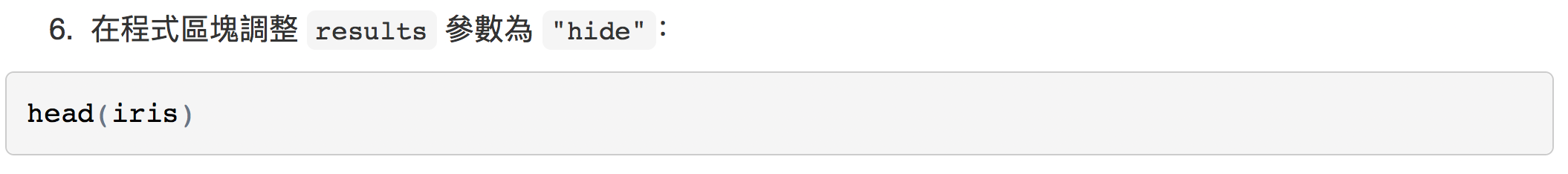


圖 16-19 程式區塊（9）

調整 warning 參數為 FALSE，會在文件中隱藏警告訊息，如果沒有調整 warning 參數輸出的文件會帶有警告訊息：

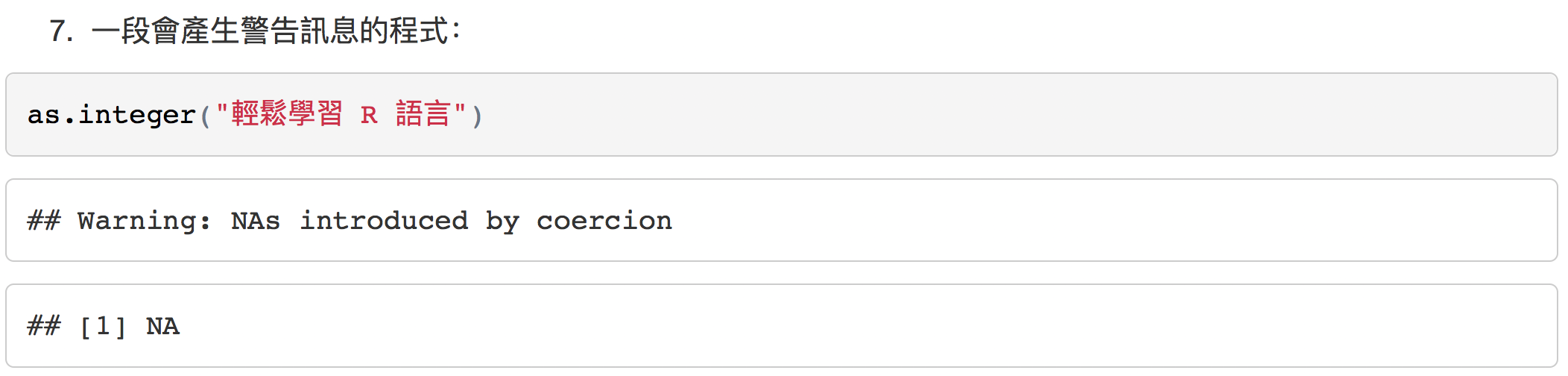


圖 16-20 程式區塊（10）

調整 warning 參數之後的輸出結果：

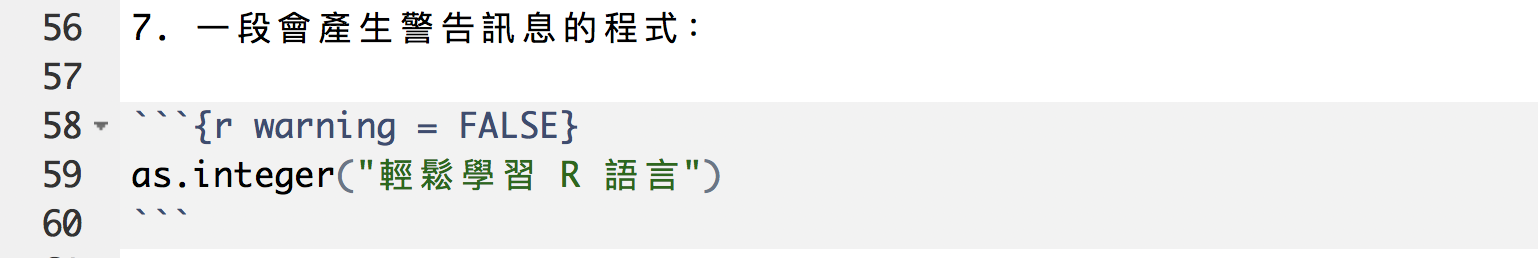


圖 16-21 程式區塊（11）



圖 16-22 程式區塊（12）

預設 R Markdown 文件中如果有錯誤的 R 程式，就會無法輸出，如果程式區塊的 error 參數調整為 TRUE，會允許輸出程式執行錯誤的文件，如果沒有調整 error 參數將會不能輸出：

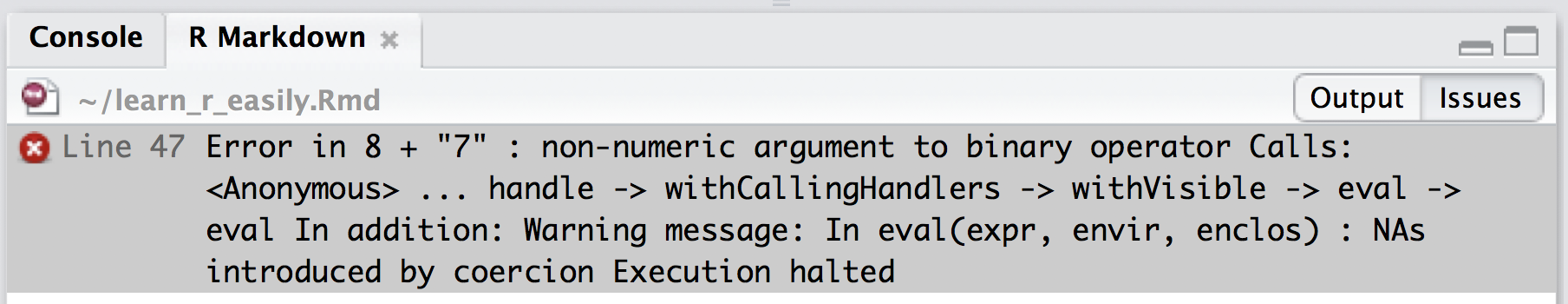


圖 16-23 程式區塊（13）

調整 error 參數之後的輸出結果：

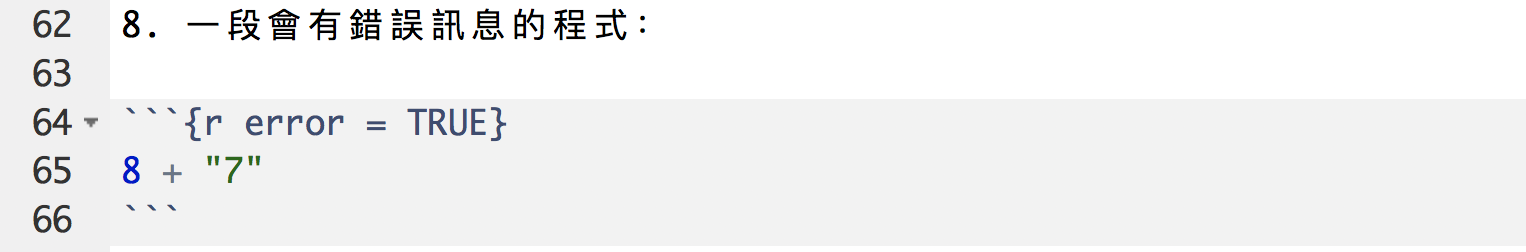


圖 16-24 程式區塊（14）

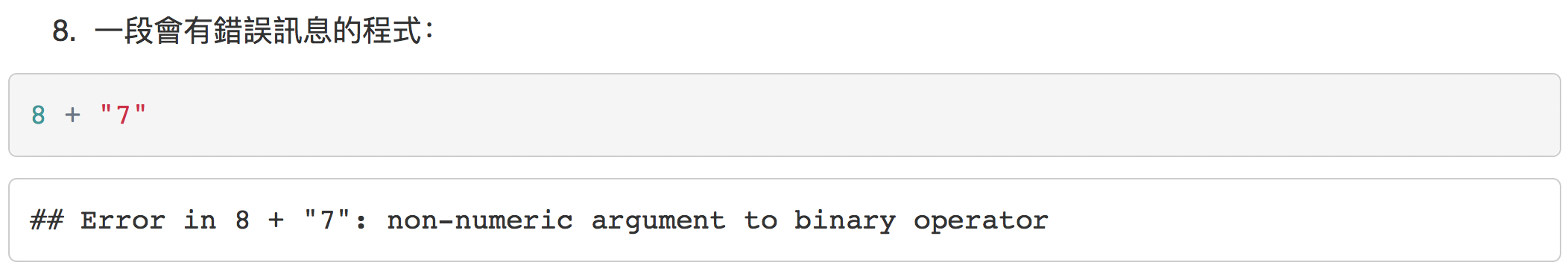


圖 16-25 程式區塊（15）

調整 message 參數為 FALSE，會在文件中隱藏訊息，如果沒有調整 message 參數輸出的文件會帶有回傳訊息：

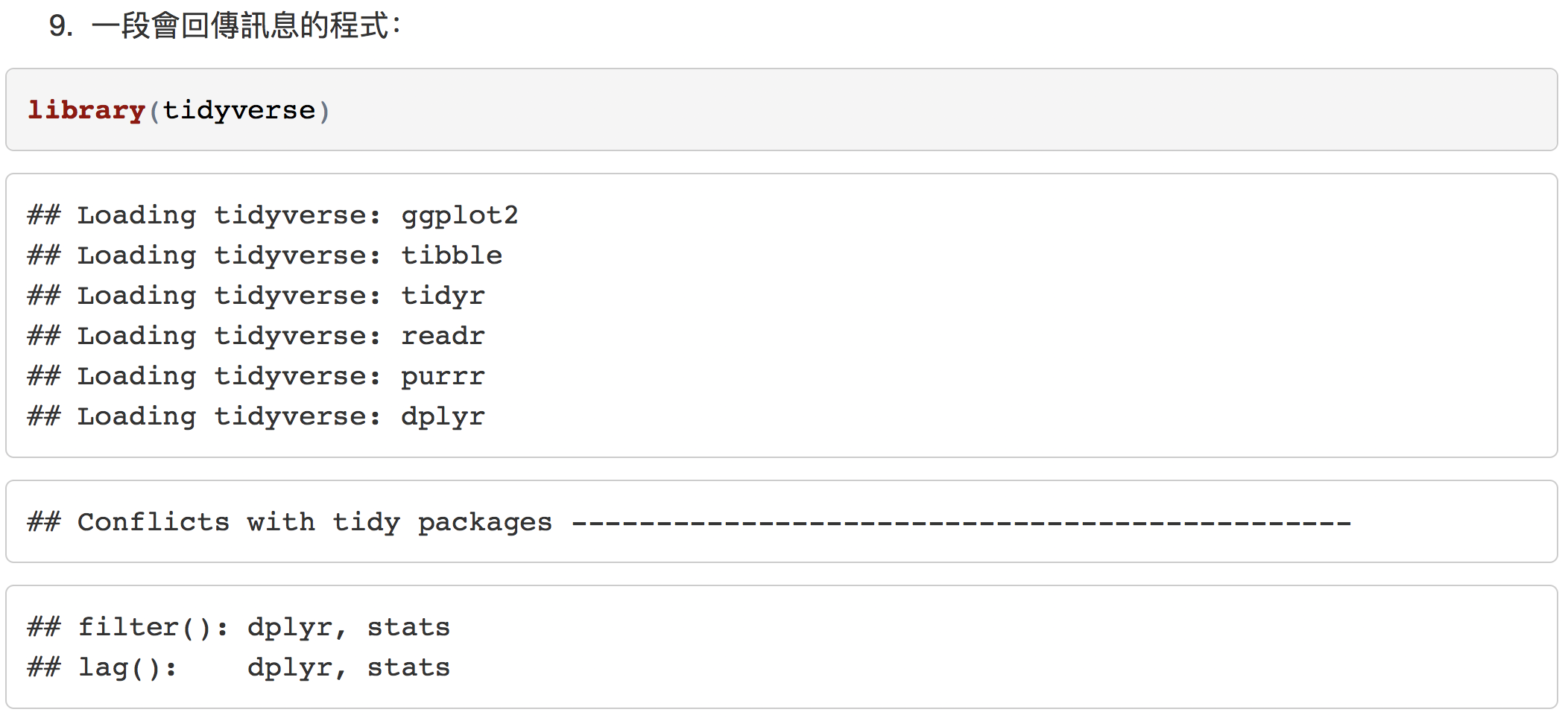


圖 16-26 程式區塊（16）

調整 message 參數之後的輸出結果：

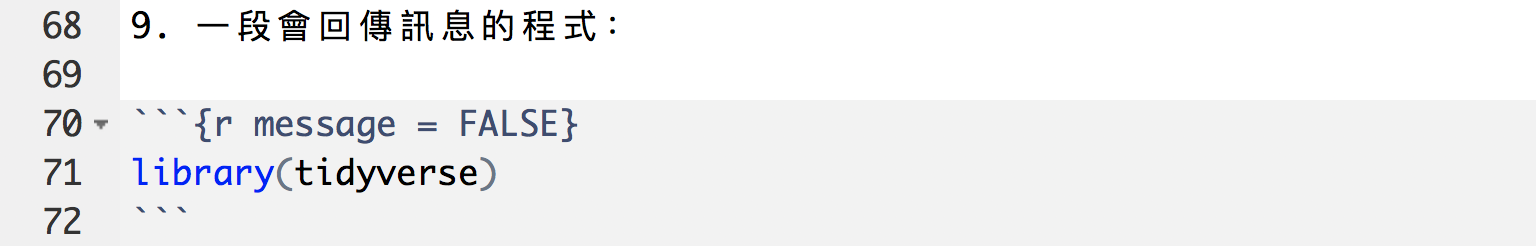


圖 16-27 程式區塊（17）

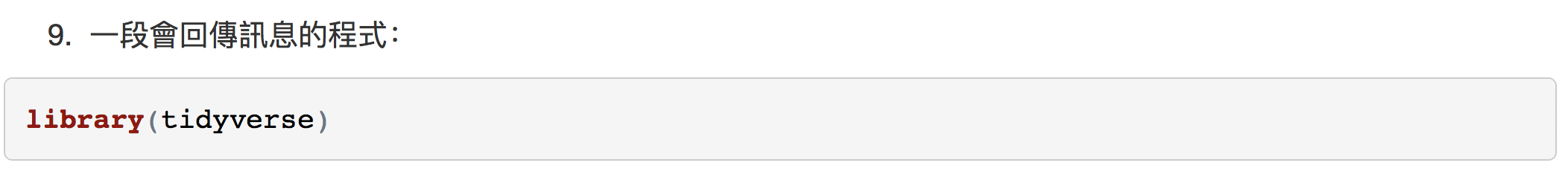


圖 16-28 程式區塊（18）

#### 連結

在文件中加入連結使用 [連結顯示名稱](連結網址 URL)：

### 開發環境與語法  
  
1. 安裝 R：[CRAN](https://cran.r-project.org/)  
2. 安裝 RStudio：[RStudio](https://www.rstudio.com/products/rstudio/download/)

輸出結果如下：



圖 16-29 連結

#### 圖片

在文件中加入圖片使用 ![圖片顯示名稱](圖片連結位址)，連結位址可以指定一個本機的圖片，也可以指定一個網址，假設我們在 雲端儲存了一張圖片 Rlogo.png：

### 開發環境與語法  
  
1. 安裝 R：[CRAN](https://cran.r-project.org/)  
2. 安裝 RStudio：[RStudio](https://www.rstudio.com/products/rstudio/download/)  
  
![R\_logo](https://storage.googleapis.com/learn-r-the-easy-way.appspot.com/screenshots\_ch16/Rlogo.png)

輸出結果如下：



圖 16-30 圖片

#### 引用

在文件中加入引用的段落使用 >：

## 基本功  
  
> R, at its heart, is a functional programming (FP) language.  
  
By Hadley Wickham

輸出結果如下：

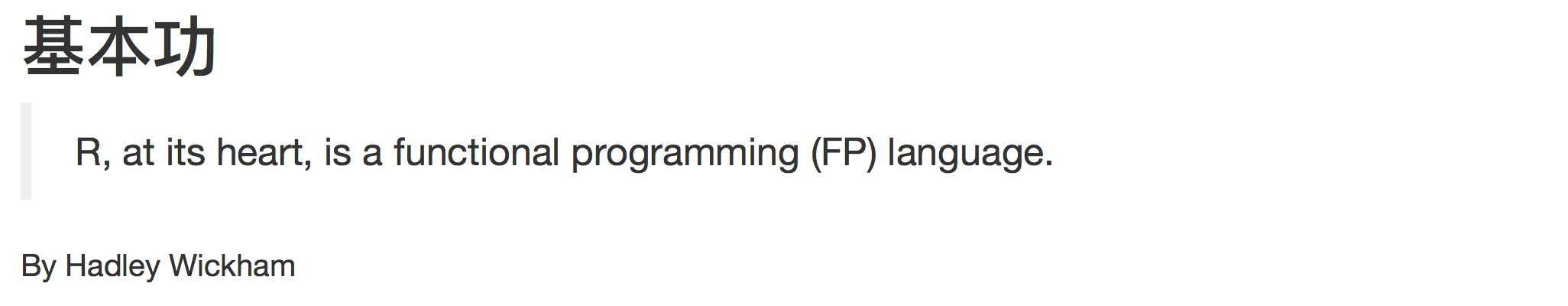


圖 16-31 引用

#### 表格

在文件中加入簡單的表格：

### 讀取資料  
  
|資料格式|函數|套件|  
|--------|----|----|  
|結構化文字|`read.table()`|`utils`|  
|非結構化文字|`readLines()`|`base`|  
|Excel 試算表|`read\_excel()`|`readxl`|  
|JSON|`fromJSON()`|`jsonlite`|

輸出結果如下：



圖 16-32 表格

#### 本章節練習的完整 R Markdwon 文件

完整的 R Markdown 文件可以在這個連結下載：

[learn\_r\_easily.Rmd](https://storage.googleapis.com/learn-r-the-easy-way.appspot.com/data_ch16/learn_r_easily.Rmd)

### 小結

好啦！第十六天的內容就是這麼多，現在您已經懂得使用 R Markdown 文件來撰寫資料分析報告，趕快在下一個專案開始使用 R 語言進行實戰吧！

### 練習題

###### 開始使用 R Markdown 撰寫您工作或學習的資料分析專案報告。

### 延伸資訊

* [Dynamic Documents with R and knitr](https://www.amazon.com/dp/1498716962/ref=cm_sw_su_dp)
* [knitr](https://yihui.name/knitr/)
* [Markdown - 維基百科](https://zh.wikipedia.org/wiki/Markdown)
* [Pandoc](http://pandoc.org/)
* [Package knitr](https://cran.r-project.org/web/packages/knitr/knitr.pdf)
* [R Markdown](http://rmarkdown.rstudio.com)