

成為初級資料分析師 I R 程式設計與資料科學應用

資料視覺化

郭耀仁

The simple graph has brought more information to the data analyst's mind than any other device.

John Tukey

大綱

- 認識範例資料集: gapminder
- ggplot2 基礎
- ggplot2 技巧
- 使用 plotly 做一個 gapminder replica

認識範例資料集： gapminder

資料視覺化經典案例

- [Napoleon's Russian campaign by Charles Minard](https://en.wikipedia.org/wiki/Charles_Joseph_Minard#/media/File:Minard.png)
(https://en.wikipedia.org/wiki/Charles_Joseph_Minard#/media/File:Minard.png).
- [Gapminder by Hans Rosling](https://youtu.be/jbkSRLYSojo) (<https://youtu.be/jbkSRLYSojo>).

關於 Gapminder, Hans Rosling 與 Factfulness

- [Gapminder \(https://www.gapminder.org/\)](https://www.gapminder.org/).
- [Hans Rosling \(https://en.wikipedia.org/wiki/Hans_Rosling\)](https://en.wikipedia.org/wiki/Hans_Rosling).
- [Factfulness \(https://www.youtube.com/watch?v=5uooRe07mYM\)](https://www.youtube.com/watch?v=5uooRe07mYM).

獲得 gapminder 範例資料集

- 安裝 gapminder 套件
- 載入 gapminder 套件

安裝 gapminder 套件

- 透過 RStudio 的 Packages 功能頁籤
- 透過 `install.packages()` 函數

```
install.packages("gapminder")
```


載入 gapminder 套件

- 透過 RStudio 的 Packages 功能頁籤
- 透過 `library()` 函數

```
library("gapminder")
```

隨堂練習： gapminder 範例資料集有 m 列， n 欄， 請問 m =?, n =?

```
In [1]: library("gapminder")
```

```
In [3]: ans
```

'gapminder 範例資料集有 1704 列， 6 欄'

隨堂練習：請顯示 gapminder 範例資料集前三列與後三列

In [5]:

```
first_three
```

country	continent	year	lifeExp	pop	gdpPercap
Afghanistan	Asia	1952	28.801	8425333	779.4453
Afghanistan	Asia	1957	30.332	9240934	820.8530
Afghanistan	Asia	1962	31.997	10267083	853.1007

In [6]:

```
last_three
```

country	continent	year	lifeExp	pop	gdpPercap
Zimbabwe	Africa	1997	46.809	11404948	792.4500
Zimbabwe	Africa	2002	39.989	11926563	672.0386
Zimbabwe	Africa	2007	43.487	12311143	469.7093

隨堂練習：gapminder 範例資料集有幾個國家？ 幾個洲別？

In [8]:

```
ans
```

```
`gapminder` 範例資料集有 142 個國家、5 個洲別'
```

隨堂練習：gapminder 範例資料集有哪些年份？

In [10]:

```
ans
```

```
1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 2002 2007
```

其他檢視 `gapminder` 範例資料集的函數： `summary()`
與 `str()`

```
In [ ]: summary(gapminder)  
        str(gapminder)
```

ggplot2 基礎

什麼是 ggplot2?

ggplot2 is a system for declaratively creating graphics, based on The Grammar of Graphics. You provide the data, tell ggplot2 how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details.

gg as in Grammer of Graphics

安裝 ggplot2 套件

- 透過 RStudio 的 Packages 功能頁籤
- 透過 `install.packages()` 函數

```
install.packages("ggplot2")
```

載入 ggplot2 套件

- 透過 RStudio 的 Packages 功能頁籤
- 透過 `library()` 函數

```
library("ggplot2")
```

基礎視覺化圖形

- 觀察資料相關性的散佈圖 (Scatter Plot)
- 觀察類別型資料的長條圖 (Bar Plot)
- 觀察資料散佈的直方圖 (Histogram)
- 觀察數值變化趨勢的線圖 (Line Plot)
- 觀察不同類別資料散佈的盒鬚圖 (Boxplot)

如何創建一個 ggplot2 圖形

- 使用 `ggplot()` 函數做資料映射
- 使用 `geom_()` 函數調整圖形種類
- 使用 `+` 連結不同的函數，堆疊圖層

```
In [ ]: library("gapminder") # data  
library("ggplot2")          # plotting  
library("dplyr")             # data manipulations
```

觀察資料相關性的散佈圖 (Scatter Plot)

使用 `ggplot(aes(x, y)) + geom_point()`

```
In [ ]: gapminder %>%  
  ggplot(aes(x = gdpPercap, y = lifeExp)) +  
  geom_point()
```

觀察類別型資料的長條圖（Bar Plot）：長條高度為觀測值個數

使用 `ggplot(aes(x)) + geom_bar()`

```
In [ ]: gapminder %>%  
  filter(year == 2007) %>%  
  ggplot(aes(x = continent)) +  
    geom_bar()
```

觀察類別型資料的長條圖（Bar Plot）：長條高度為摘要數值

使用 `ggplot(aes(x, y)) + geom_bar(stat = "identity")`

```
In [ ]: gapminder %>%  
  filter(year == 2007) %>%  
  mutate(pop_numeric = as.numeric(pop)) %>%  
  group_by(continent) %>%  
  summarise(ttl_pop = sum(pop_numeric)) %>%  
  ggplot(aes(x = continent, y = ttl_pop)) +  
    geom_bar(stat = "identity")
```


觀察資料散佈的直方圖 (Histogram)

使用 `ggplot(aes(x)) + geom_histogram()`

```
In [ ]: gapminder %>%  
  ggplot(aes(x = gdpPercap)) +  
    geom_histogram(bins = 40)
```

觀察數值變化趨勢的線圖 (Line Plot)

使用 `ggplot(aes(x, y)) + geom_line()`

```
In [ ]: gapminder %>%  
  filter(country %in% c("Taiwan", "Japan", "China")) %>%  
  ggplot(aes(x = year, y = gdpPercap, color = country)) +  
    geom_line()
```

觀察不同類別資料散佈的盒鬚圖 (Boxplot)

使用 `ggplot(aes(x, y)) + geom_boxplot()`

```
In [ ]: gapminder %>%  
  ggplot(aes(x = continent, y = gdpPercap, color = continent)) +  
  geom_boxplot()
```

ggplot2 技巧

常用的 ggplot2 技巧

- 加入圖標題與軸標籤
- 加入註釋
- 加入中文字（macOS 使用者會遭遇的問題）
- 調整座標軸
- 在一個畫布上繪製多個子圖形

加入圖標題與軸標籤

使用 `ggtitle()` + `xlab()` + `ylab()`

```
In [ ]: gapminder %>%  
  ggplot(aes(x = gdpPercap, y = lifeExp)) +  
  geom_point() +  
  ggtitle("Wealth vs. Health") +  
  xlab("GDP Per Capita") +  
  ylab("Life Expectancy")
```

加入註釋

使用 `geom_text()`

```
In [ ]: n_obs <- gapminder %>%  
  group_by(continent) %>%  
  summarise(nrows = n())  
  
n_obs %>%  
  ggplot(aes(x = continent, y = nrows)) +  
  geom_bar(stat = "identity") +  
  geom_text(aes(label = n_obs$nrows, y = n_obs$nrows), vjust = -1)
```

加入中文字 (macOS 使用者會遭遇的問題)

使用 `theme(text = element_text(family = FONTS_SUPPORT_TC))`

```
In [ ]: gapminder %>%  
  ggplot(aes(x = gdpPercap, y = lifeExp)) +  
    geom_point() +  
    ggtitle("財富與健康") +  
    xlab("人均 GDP") +  
    ylab("預期壽命") +  
    theme(text = element_text(family = "Heiti TC Light"))
```


調整座標軸

- 使用 `scale_x_continuous()` 與 `scale_y_continuous()` 調整座標軸上下界與量尺

```
In [ ]: gapminder %>%  
  ggplot(aes(x = gdpPercap, y = lifeExp)) +  
  geom_point() +  
  scale_x_continuous(limits = c(0, 50000))
```

```
In [ ]: gapminder %>%  
  ggplot(aes(x = gdpPercap, y = lifeExp)) +  
  geom_point() +  
  scale_x_continuous(trans = "log10")
```

在一個畫布上繪製多個子圖形

使用 `facet_wrap(vars(CATEGORICAL_COLUMN))`

```
In [ ]: gapminder %>%  
  ggplot(aes(x = gdpPercap, y = lifeExp, color = continent)) +  
    geom_point() +  
    facet_wrap(vars(continent))
```

使用 plotly 做一個 gapminder replica

關於 **plotly** 套件

幫助 R 語言使用者不需要額外去學習 JavaScript 就能夠建立出互動性、具備 D3.js 及 WebGL 特性的圖表

安裝 plotly 套件

- 透過 RStudio 的 Packages 功能頁籤
- 透過 `install.pacakges()` 函數

```
install.pacakges("plotly")
```

載入 plotly 套件

- 透過 RStudio 的 Packages 功能頁籤
- 透過 `library()` 函數

```
library("plotly")
```

```
In [ ]: library("plotly")
library("gapminder")
radius <- sqrt((gapminder$pop)/pi)

p <- gapminder %>%
  plot_ly(
    x = ~gdpPercap,
    y = ~lifeExp,
    size = ~pop,
    color = ~continent,
    frame = ~year,
    text = ~country,
    hoverinfo = "text",
    type = 'scatter',
    mode = 'markers',
    sizes = c(min(radius), max(radius))
  ) %>%
  layout(
    xaxis = list(
      type = "log"
    )
  )
```

In []:

p

延伸閱讀

- <https://ggplot2.tidyverse.org> (<https://ggplot2.tidyverse.org>).
- <https://plot.ly/r/> (<https://plot.ly/r/>).