

主題: 第 3 章圖形文法繪圖-ggplot2 簡介

作者: 李明昌

日期: 2025-07-26

目錄

3.1 認識 ggplot2 套件	2
ggplot2 雙軸 (Y1,Y2) 繪圖	14
ggplot2-新增數學式-以迴歸分析模型為例	14
弦圖 (Chord diagram) ——	14
3.2 熟悉開放資料集於 ggplot2 應用	15
資料集願望清單	19
參考資料	19
謝謝您的聆聽, Q & A	20

大綱:

3. 圖形文法繪圖

- 3.1 認識 ggplot2 套件
- 3.2 熟悉開放資料集於 ggplot2 應用

R Markdown 技巧:

- `include = FALSE` 隱藏程式碼和標記的輸出, 但會在背景中執行 R 程式碼。預設值 `include = FALSE`。

- `echo = FALSE` 隱藏程式碼，但會在背景中執行 R 程式碼並顯示（繪圖）結果。預設值 `echo = TRUE` 表示呈現程式碼與結果。
- Rmd 檔案名稱與路徑使用英文字母，不建議使用中文字型。
- Rmd 檔案轉換為 Word 檔案參考: <https://bookdown.org/yihui/rmarkdown/word-document.html>
- Rmd 檔案中文字型 YouTube: <https://youtu.be/6Wc75BH02iE>

3.1 認識 ggplot2 套件

ggplot2 是一個強大的視覺化套件，可以用來建立美觀並呈現直觀式與有效的圖表。

ggplot2 的三大元素：

- **data=**：指定視覺化的資料集。
- **aes()**：將資料中的變數對應到繪圖的美學屬性 (aesthetic properties)，即是指定 X、Y 等對應變數。
- **geom_**：指定幾何物件的類型，例：點、線、長條圖，例：`geom_point`、`geom_line`、`geom_smooth`、`geom_bar`。
- ggplot2 參考: <https://r-graph-gallery.com/>

```
# 載入套件
library(ggplot2)
```

```
# ?diamonds
```

```
# 資料結構
```

```
str(diamonds) # 53940*10
```

```
## tibble [53,940 x 10] (S3: tbl_df/tbl/data.frame)
## $ carat   : num [1:53940] 0.23 0.21 0.23 0.29 0.31 0.24 0.24 0.26 0.22 0.23 ...
## $ cut     : Ord.factor w/ 5 levels "Fair"<"Good"<...: 5 4 2 4 2 3 3 3 1 3 ...
## $ color   : Ord.factor w/ 7 levels "D"<"E"<"F"<"G"<...: 2 2 2 6 7 7 6 5 2 5 ...
## $ clarity: Ord.factor w/ 8 levels "I1"<"SI2"<"SI1"<...: 2 3 5 4 2 6 7 3 4 5 ...
## $ depth   : num [1:53940] 61.5 59.8 56.9 62.4 63.3 62.8 62.3 61.9 65.1 59.4 ...
## $ table   : num [1:53940] 55 61 65 58 58 57 57 55 61 61 ...
## $ price   : int [1:53940] 326 326 327 334 335 336 336 337 337 338 ...
## $ x       : num [1:53940] 3.95 3.89 4.05 4.2 4.34 3.94 3.95 4.07 3.87 4 ...
## $ y       : num [1:53940] 3.98 3.84 4.07 4.23 4.35 3.96 3.98 4.11 3.78 4.05 ...
## $ z       : num [1:53940] 2.43 2.31 2.31 2.63 2.75 2.48 2.47 2.53 2.49 2.39 ...
```

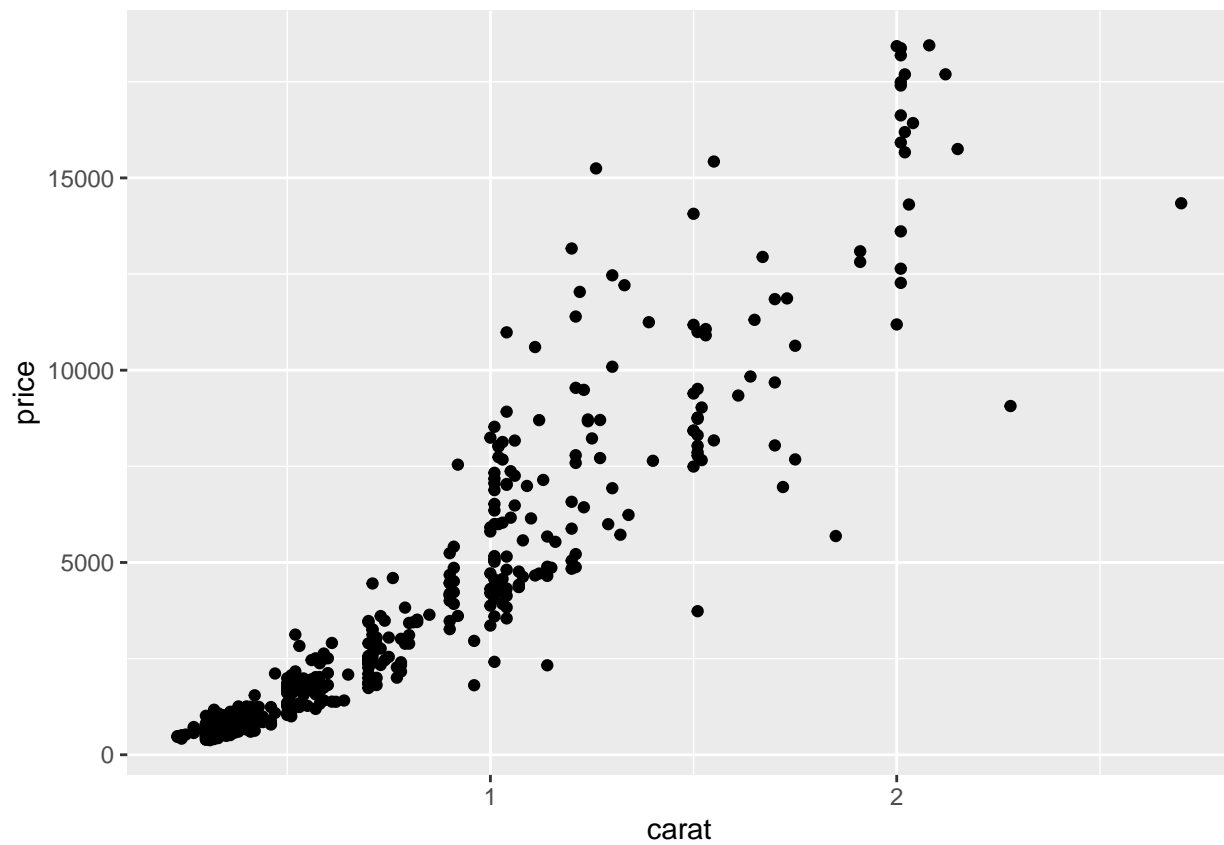
```
# 設定亂數種子，使用大於或等於 1 整數
set.seed(168)

# 使用 sample 函數自所有列數隨機抽取 500 筆資料
dsmall <- diamonds[sample(nrow(diamonds), 500),]

# 顯示結果
dsmall # 500*10
```

```
## # A tibble: 500 x 10
##   carat cut      color clarity depth table price      x      y      z
##   <dbl> <ord>    <ord> <ord>    <dbl> <dbl> <int> <dbl> <dbl> <dbl>
## 1  0.42 Premium  G     VS1     62.4   59   984  4.8   4.78  2.99
## 2  1.01 Ideal   I     I1      61.6   56  2416  6.45  6.43  3.97
## 3  1.09 Very Good F     VS2     61.4   54  6990  6.66  6.7   4.1
## 4  0.31 Good    D     SI1     63.9   54   732  4.3   4.27  2.74
## 5  1.01 Ideal   G     VS1     59.1   55  7059  6.6   6.53  3.88
## 6  0.3  Premium  G     VS1     62.5   59   787  4.32  4.25  2.68
## 7  0.58 Ideal   J     VS2     60.9   57  1324  5.38  5.4   3.28
## 8  0.51 Ideal   I     SI1     61.7   54  1000  5.18  5.19  3.2
## 9  1.5  Good    H     SI1     61     59  9394  7.31  7.39  4.48
## 10 0.4  Ideal   G     IF      61.1   56  1257  4.76  4.79  2.92
## # i 490 more rows
```

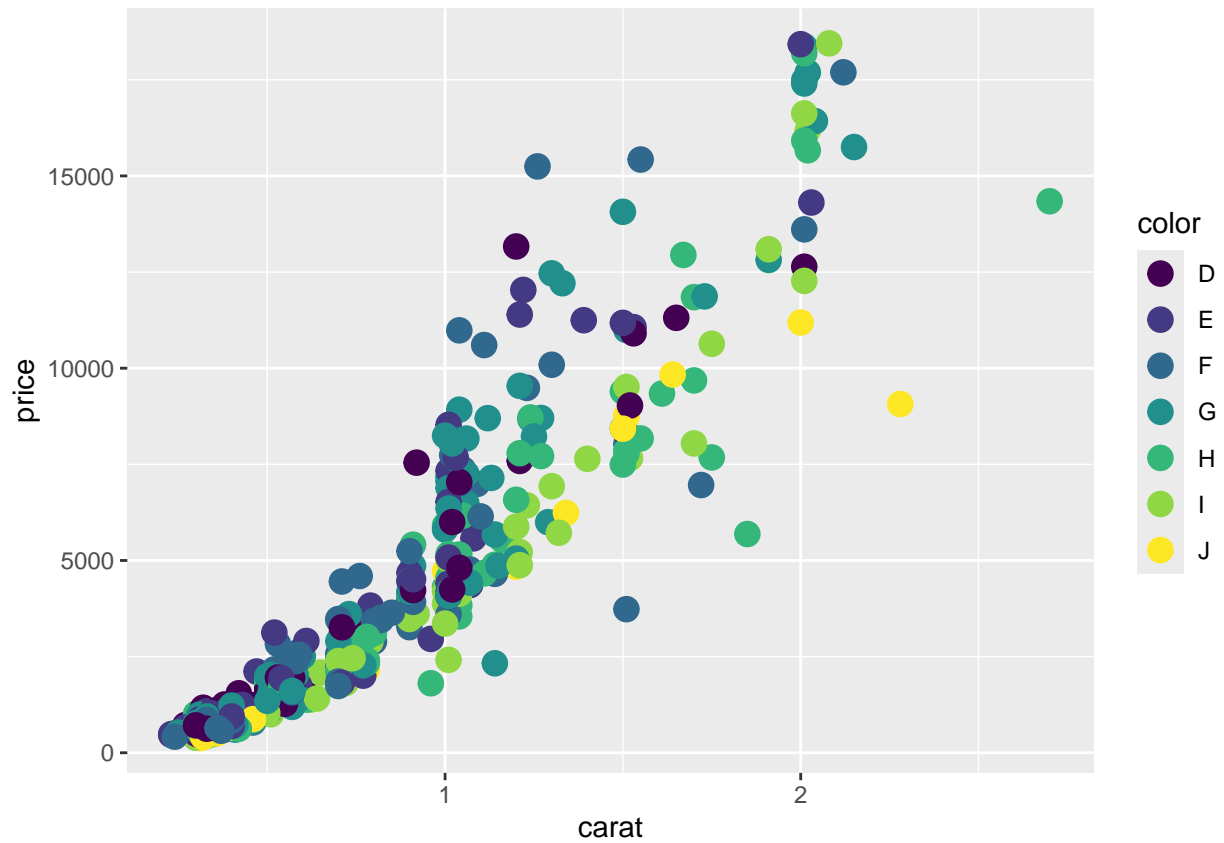
```
# ggplot: 散佈圖
p <- ggplot(data=dsmall, mapping=aes(carat, price)) +
  geom_point()
p
```



```
# ggplot: 群組散佈圖
```

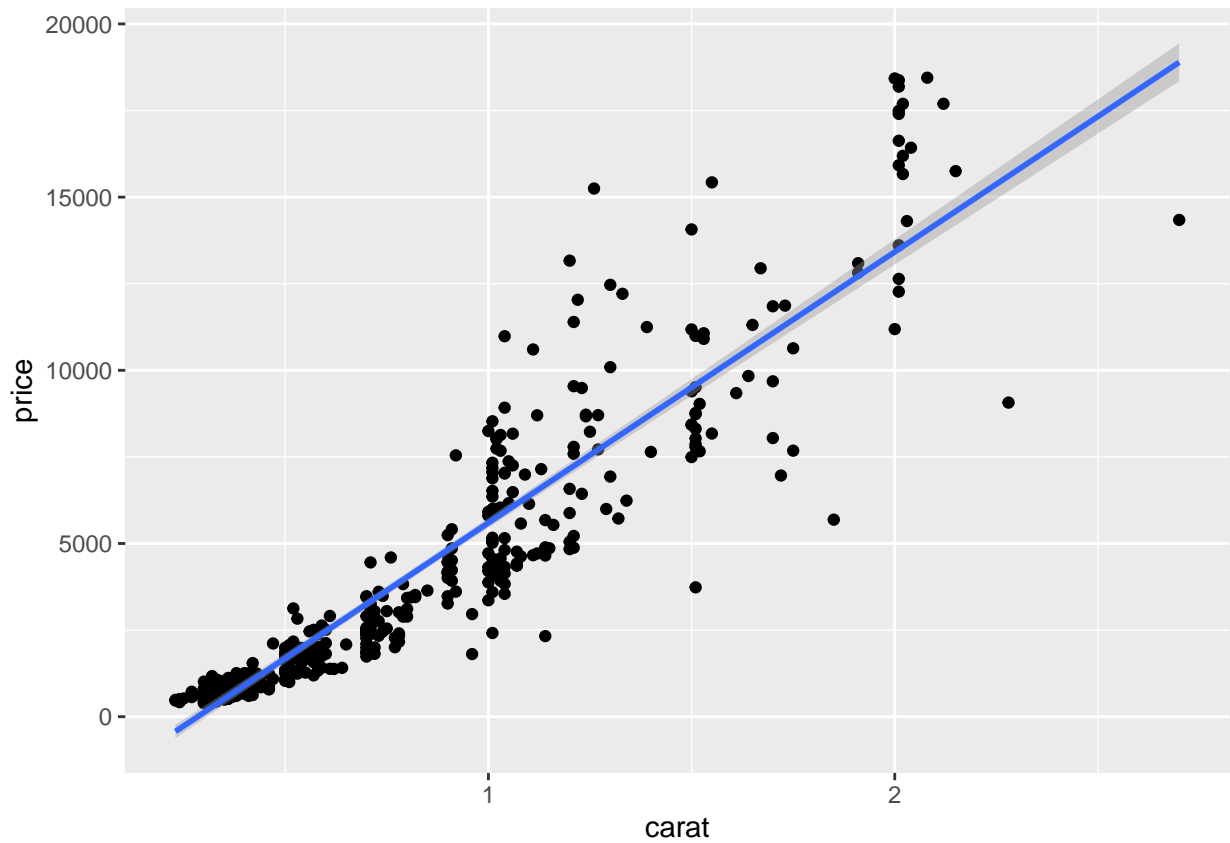
```
p <- ggplot(data=dsmall, mapping=aes(carat, price, color=color)) + # 等號左側 color 表示群組參數，右  
  geom_point(size=4)
```

```
p
```



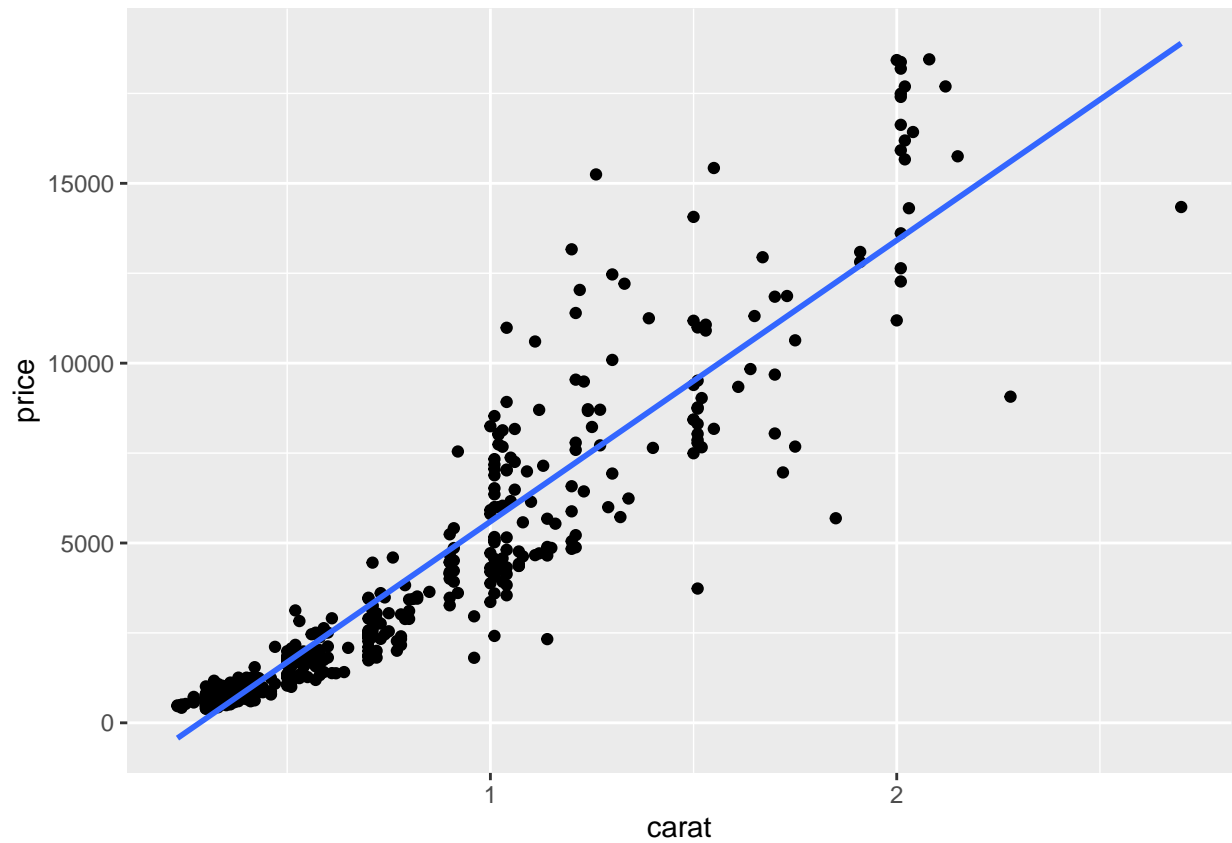
```
# ggplot: 散點圖 + 線性迴歸 +95% 信賴區間
p <- ggplot(dsmall, aes(carat, price)) +
  geom_point() +
  geom_smooth(method="lm")
p
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



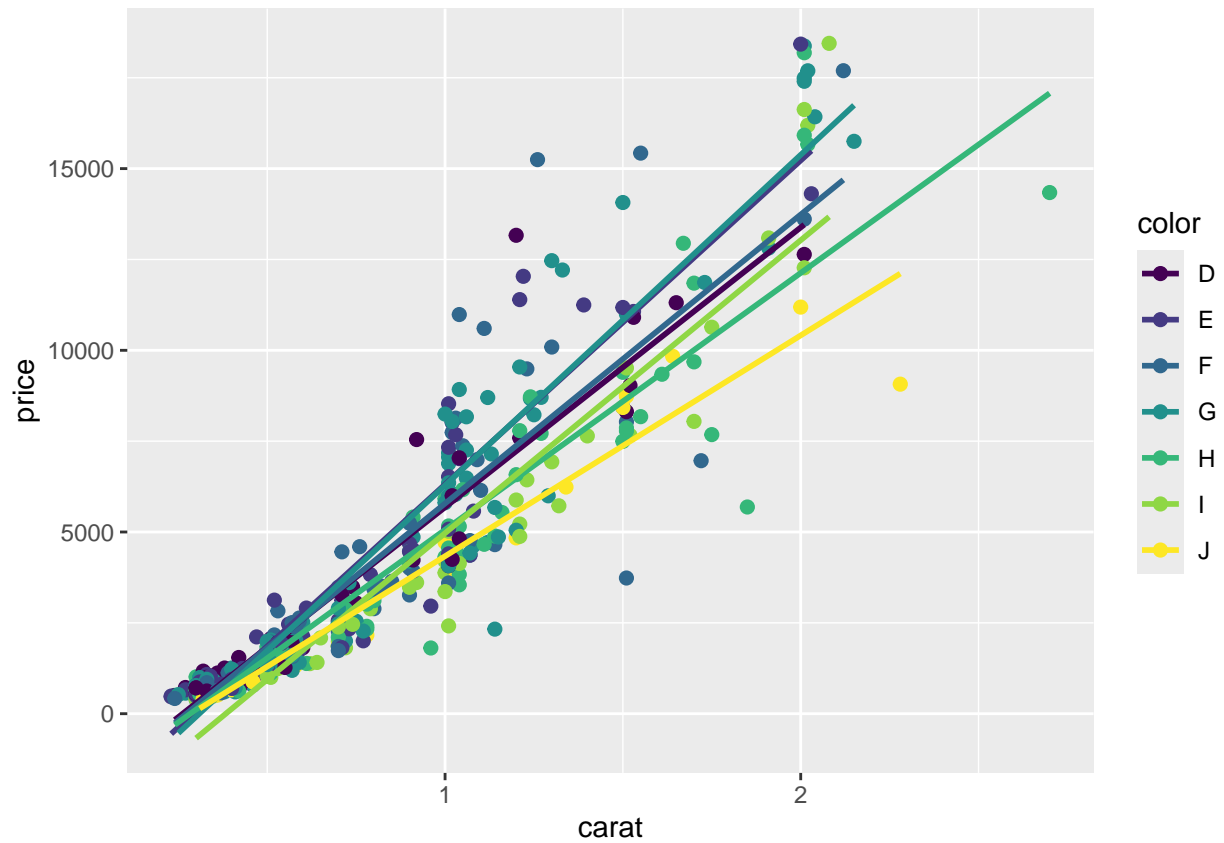
```
# ggplot: 散點圖 + 線性迴歸 + 不顯示信賴區間
ggplot(dsmall, aes(carat, price)) +
  geom_point() +
  geom_smooth(method="lm", se=FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



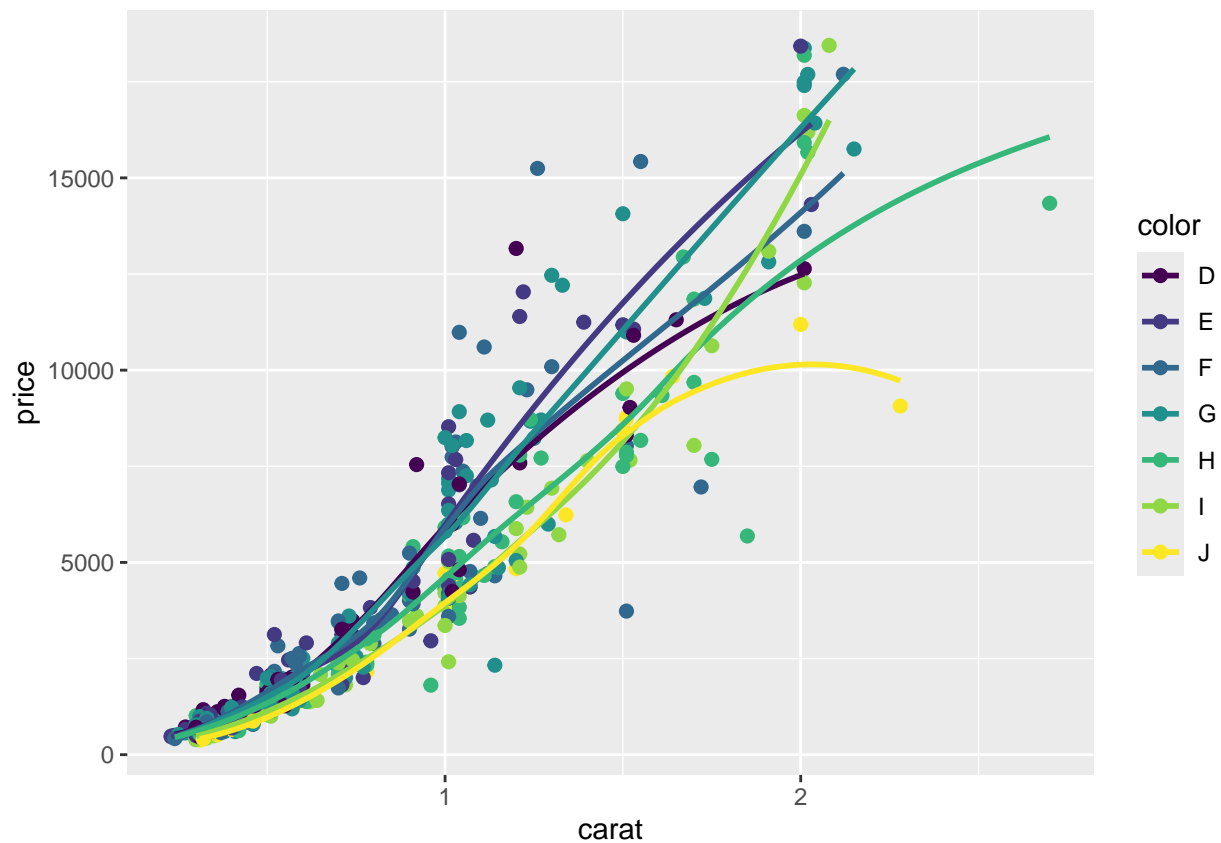
```
# ggplot: 散點圖 + 群組 lm
ggplot(dsmall, aes(carat, price, group=color)) +
  geom_point(aes(color=color), size=2) +
  geom_smooth(aes(color=color), method="lm", se=FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



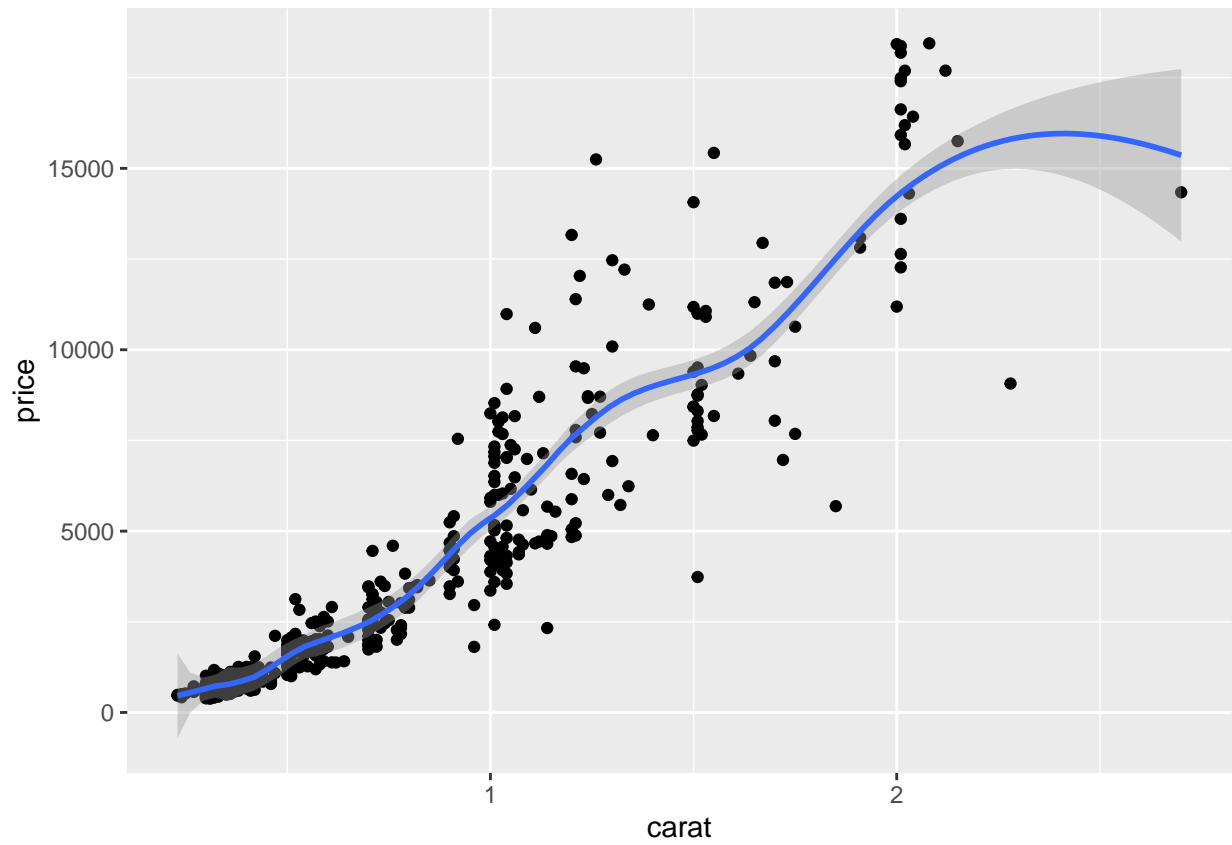
```
# ggplot: 散點圖 + 群組 loess
ggplot(dsmall, aes(carat, price, group=color)) +
  geom_point(aes(color=color), size=2) +
  geom_smooth(aes(color=color), method="loess", se=FALSE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

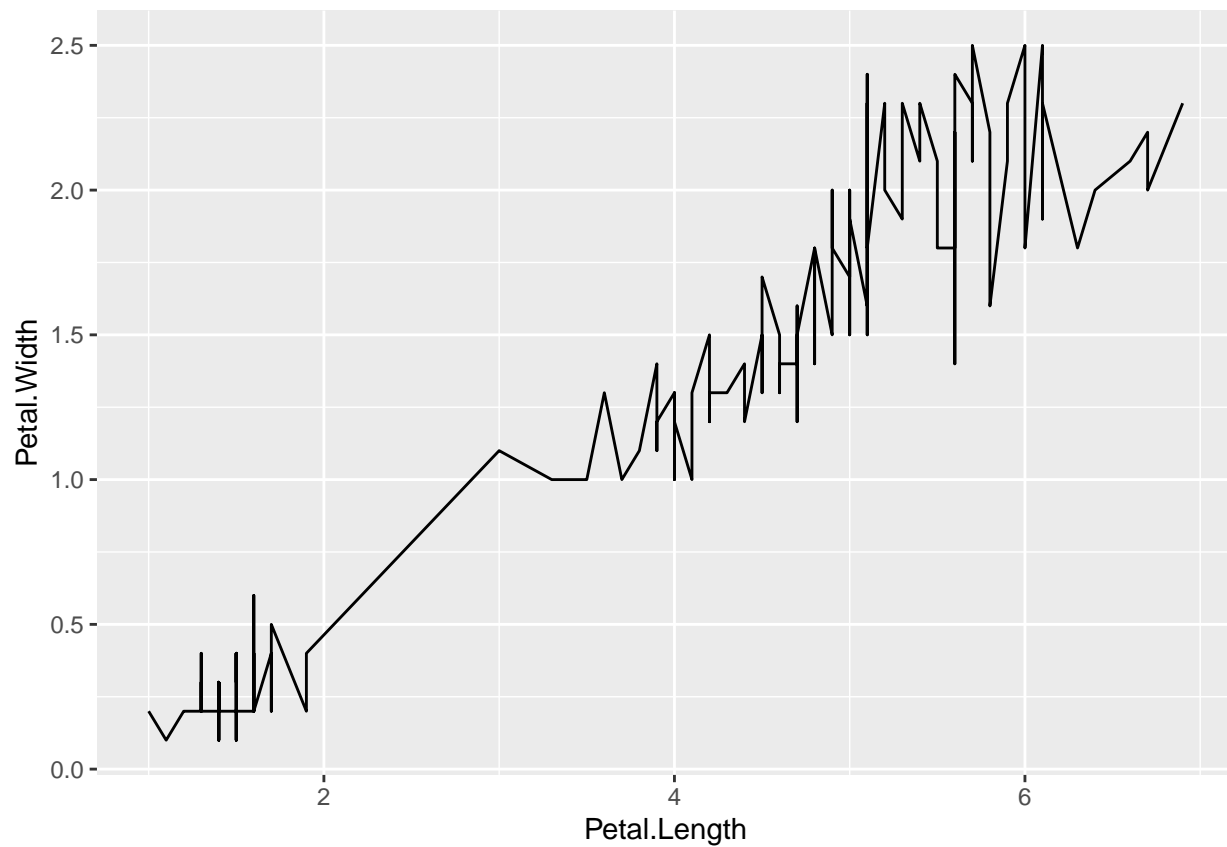



```
# ggplot: 散點圖 +span  
ggplot(dsmall, aes(carat, price)) +  
  geom_point() +  
  geom_smooth(span = 0.3)
```

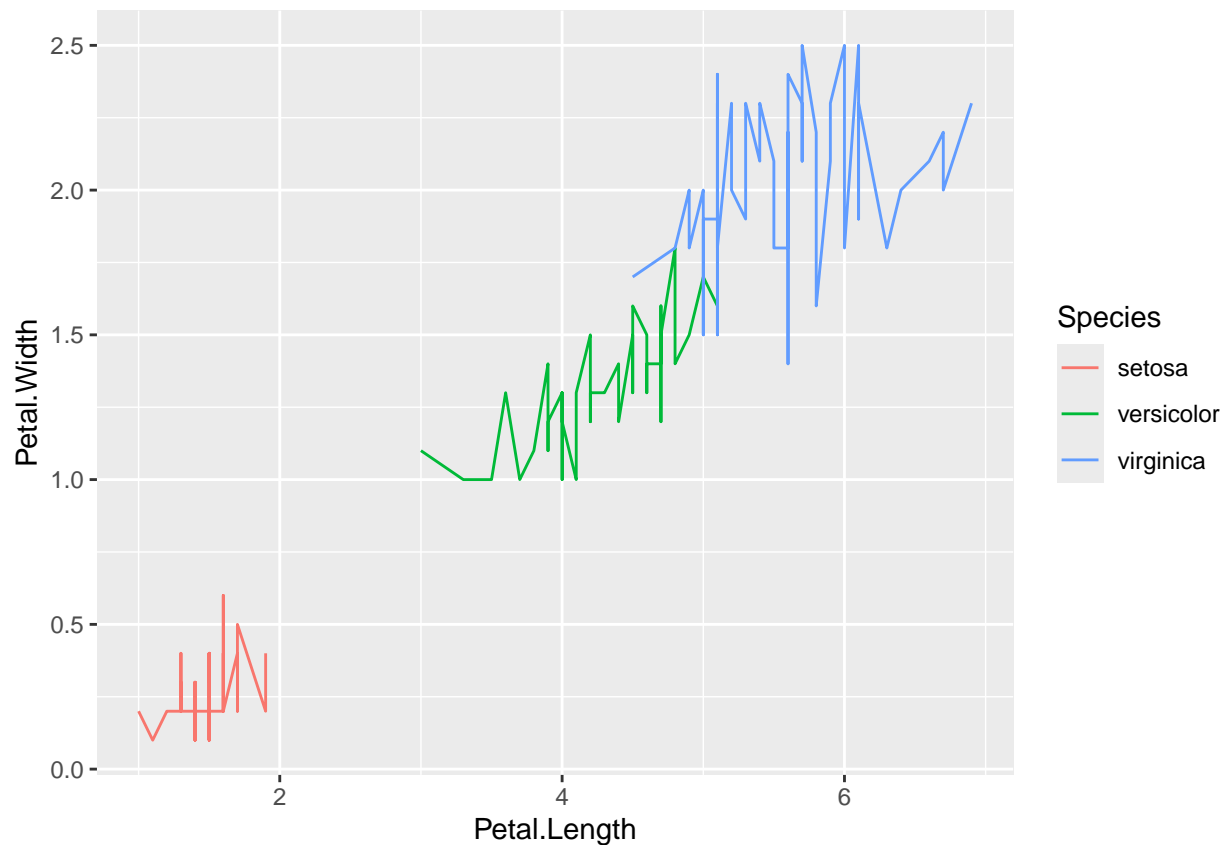
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



```
# ggplot: 線圖  
ggplot(iris, aes(Petal.Length, Petal.Width)) +  
  geom_line()
```



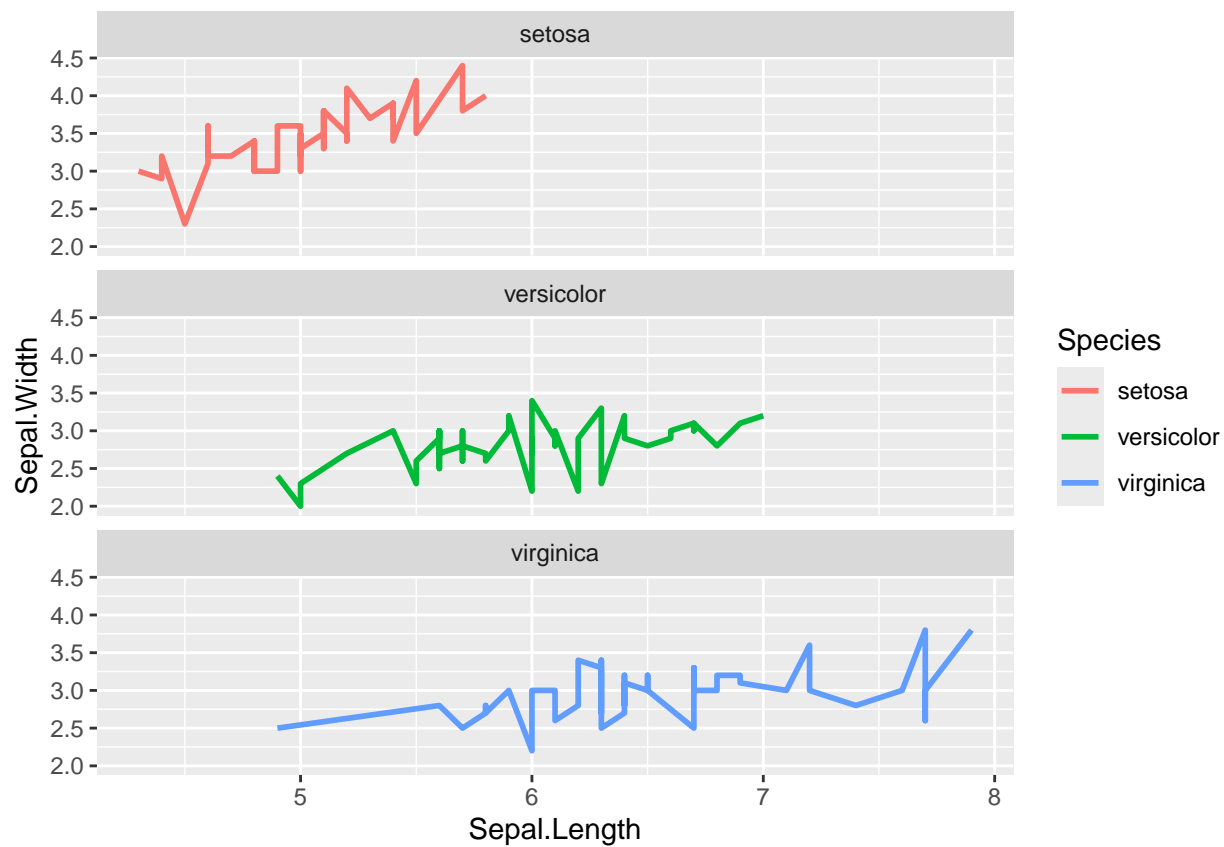
```
# ggplot: 線圖 + 群組 + 顏色
ggplot(iris, aes(Petal.Length, Petal.Width, group=Species, color=Species)) + # group: 群組
  geom_line()
```



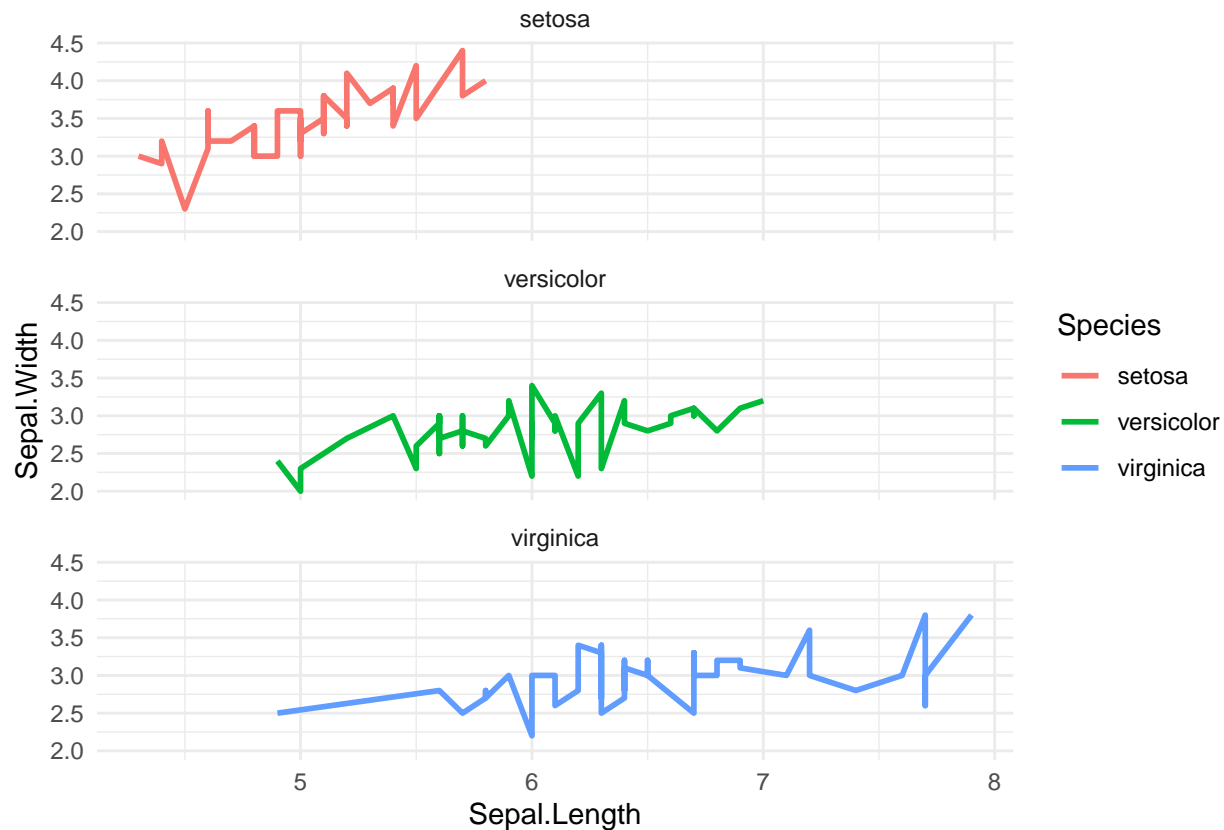
```
# ggplot: 線圖 + 分面畫圖
```

```
ggplot(iris, aes(Sepal.Length, Sepal.Width)) +  
  geom_line(aes(color=Species), size=1) +  
  facet_wrap(~Species, ncol=1)
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## i Please use `linewidth` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was  
## generated.
```



```
# ggplot: 線圖 + 分面
ggplot(iris, aes(Sepal.Length, Sepal.Width)) +
  geom_line(aes(color=Species), size=1) +
  facet_wrap(~Species, ncol=1) +
  theme_minimal()
```



```
# theme_ 設定繪圖風格，例：theme_gray、theme_bw、theme_minimal、theme_classic，參考`?theme_classic`
```

ggplot2 雙軸 (Y1,Y2) 繪圖

參考: <http://rwepa.blogspot.com/2017/09/ggplot2-2y-packagelist.html>

ggplot2-新增數學式-以迴歸分析模型為例

參考: <http://rwepa.blogspot.com/2020/09/ggplot2-equation.html>

弦圖 (Chord diagram) ——

參考: <http://rwepa.blogspot.com/2019/10/chord-diagram.html>

3.2 熟悉開放資料集於 ggplot2 應用

ChatGPT: 模擬 10 萬筆台積電公司製程不良率資料, 資料至少包括年月日時間, location, 批號, 生產量, 使用 dplyr, ggplot2 套件進行資料預處理分析, 視覺化分析.

```
# 載入套件
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##      filter, lag

## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
# 設定資料集大小
datasize <- 100000

# 亂數種子
set.seed(123)

# 建立模擬製程資料
df <- data.frame(
  id = 1:datasize,
  year = sample(2018:2023, size = datasize, replace = TRUE),
  month = sample(1:12, size = datasize, replace = TRUE),
  day = sample(1:31, size = datasize, replace = TRUE),
  location = sample(c(" 台中", " 高雄", " 新竹"), size = datasize, replace = TRUE),
  batch_number = paste0("Batch", sample(1:10, size = datasize, replace = TRUE)),
  production_volume = runif(datasize, min = 0.001, max = 100),
  defect_rate = rnorm(datasize, mean = 0.05, sd = 0.01)
)

# 轉換為 tbl_df
```

```
df <- tibble::as_tibble(df)
```

```
# 類別
```

```
class(df) # "tbl_df" "tbl" "data.frame"
```

```
## [1] "tbl_df"      "tbl"        "data.frame"
```

```
df
```

```
## # A tibble: 100,000 x 8
```

```
##      id  year month   day location batch_number production_volume defect_rate
##    <int> <int> <int> <int> <chr>      <chr>              <dbl>         <dbl>
##  1     1    2020     5    21  台中      Batch3              98.3          0.0584
##  2     2    2023     9     6  台中      Batch2               8.41          0.0409
##  3     3    2020     2    10  新竹      Batch6              41.5          0.0612
##  4     4    2019     6     3  高雄      Batch8              16.1          0.0511
##  5     5    2019     3    19  新竹      Batch7              69.6          0.0509
##  6     6    2023    11     9  台中      Batch10             15.7          0.0433
##  7     7    2020     4     5  台中      Batch2              68.0          0.0371
##  8     8    2022     5    19  新竹      Batch2              70.4          0.0426
##  9     9    2021     6    10  台中      Batch1              95.2          0.0225
## 10    10    2023     3     2  台中      Batch1              74.5          0.0475
## # i 99,990 more rows
```

```
# 新增製程日期變數
```

```
df <- df %>%
```

```
  mutate(process_date = as.Date(paste0(year, "/", month, "/", day)))
```

```
# 計算每日平均不良率
```

```
df_summary <- df %>%
```

```
  group_by(process_date) %>%
```

```
  summarise(AvgDefectRate = mean(defect_rate))
```

```
# 刪除 NA 值
```

```
df_summary <- df_summary %>%
```

```
  na.omit()
```

```
# 資料結構
```

```
str(df_summary)
```



```
## tibble [2,191 x 2] (S3: tbl_df/tbl/data.frame)
## $ process_date : Date[1:2191], format: "2018-01-01" "2018-01-02" ...
## $ AvgDefectRate: num [1:2191] 0.0503 0.0492 0.0503 0.0503 0.0514 ...
## - attr(*, "na.action")= 'omit' Named int 2192
## ..- attr(*, "names")= chr "2192"
```

資料摘要

```
summary(df_summary)
```

```
## process_date AvgDefectRate
## Min. :2018-01-01 Min. :0.04381
## 1st Qu.:2019-07-02 1st Qu.:0.04895
## Median :2020-12-31 Median :0.04996
## Mean :2020-12-31 Mean :0.04995
## 3rd Qu.:2022-07-01 3rd Qu.:0.05098
## Max. :2023-12-31 Max. :0.05534
```

前 6 筆資料

```
head(df_summary)
```

```
## # A tibble: 6 x 2
## process_date AvgDefectRate
## <date> <dbl>
## 1 2018-01-01 0.0503
## 2 2018-01-02 0.0492
## 3 2018-01-03 0.0503
## 4 2018-01-04 0.0503
## 5 2018-01-05 0.0514
## 6 2018-01-06 0.0474
```

後 6 筆資料

```
tail(df_summary)
```

```
## # A tibble: 6 x 2
## process_date AvgDefectRate
## <date> <dbl>
## 1 2023-12-26 0.0521
## 2 2023-12-27 0.0499
## 3 2023-12-28 0.0516
```

```
## 4 2023-12-29      0.0485
## 5 2023-12-30      0.0507
## 6 2023-12-31      0.0512
```

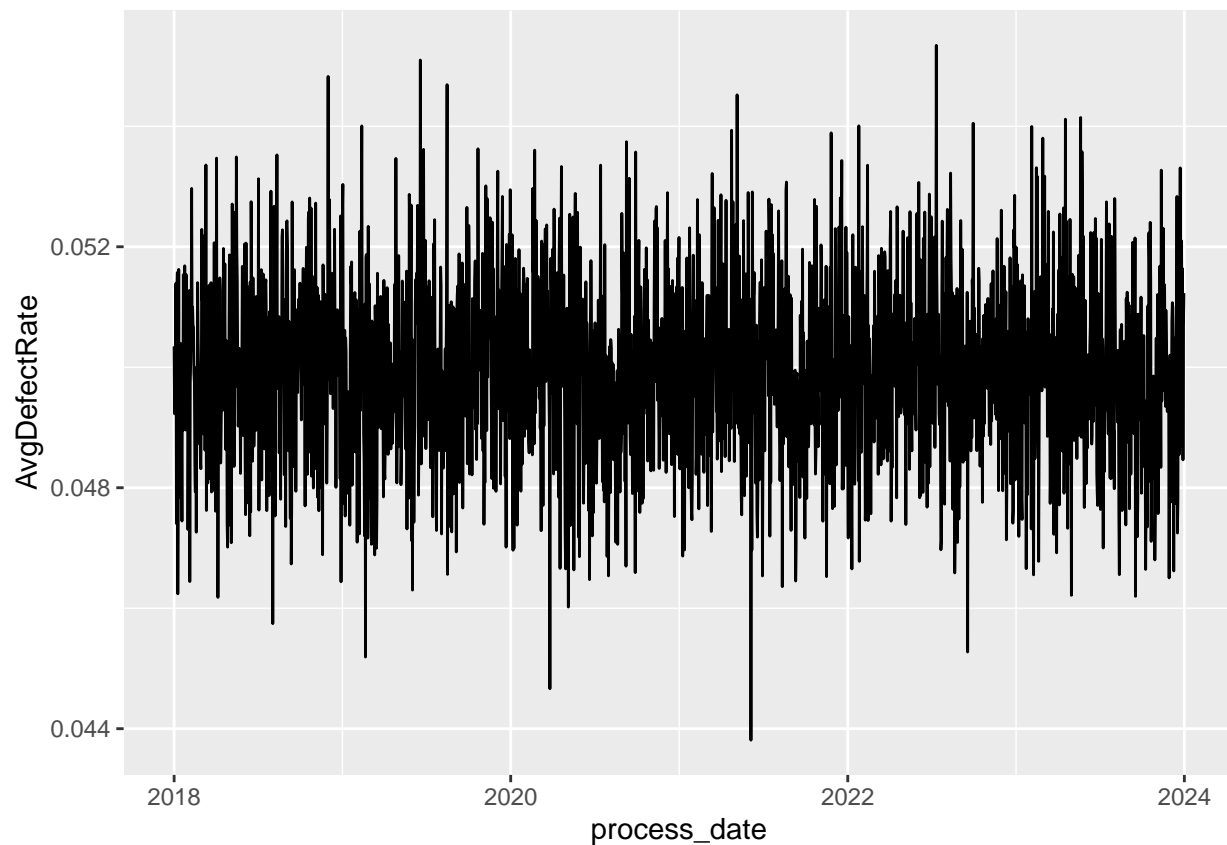
```
# 顯示資料
```

```
df_summary
```

```
## # A tibble: 2,191 x 2
##   process_date AvgDefectRate
##   <date>         <dbl>
## 1 2018-01-01      0.0503
## 2 2018-01-02      0.0492
## 3 2018-01-03      0.0503
## 4 2018-01-04      0.0503
## 5 2018-01-05      0.0514
## 6 2018-01-06      0.0474
## 7 2018-01-07      0.0512
## 8 2018-01-08      0.0516
## 9 2018-01-09      0.0462
## 10 2018-01-10     0.0482
## # i 2,181 more rows
```

```
# 使用 ggplot2 資料視覺化
```

```
ggplot(df_summary, aes(x = process_date, y = AvgDefectRate)) +
  geom_line()
```



```
# 發現什麼樣式 (patterns) --> NONE  
# Try: 優化視覺化結果, 加入平均線
```

```
# Try: facet_wrap_year
```

```
# Try: facet_wrap_month
```

```
# Try: facet_wrap_location
```

資料集願望清單

參考: https://github.com/rwepa/r_data_scientist/blob/main/r_training_advanced_ppt_codes_2025/r_data_science_advncanced_02_dplyr.R#L165

參考資料

- RWEPA <http://rwepa.blogspot.com/>

- iPAS-R-tutorial https://github.com/rwepa/ipas_bda/blob/main/ipas-r-program.R
- R 入門資料分析與視覺化應用教學 (付費) <https://mastertalks.tw/products/r?ref=MCLEE>
- R 商業預測與應用 (付費) <https://mastertalks.tw/products/r-2?ref=MCLEE>

謝謝您的聆聽, Q & A