

Lab 05

Data Visualization with ggplot2

內建資料: diamonds

```
library(ggplot2)  
diamonds
```

carat	cut	color	clarity	depth	table	price	x	y	z
0.23	Ideal	E	SI2	61.5	55	326	3.95	3.98	2.43
0.21	Premium	E	SI1	59.8	61	326	3.89	3.84	2.31
0.23	Good	E	VS1	56.9	65	327	4.05	4.07	2.31
0.29	Premium	I	VS2	62.4	58	334	4.20	4.23	2.63
0.31	Good	J	SI2	63.3	58	335	4.34	4.35	2.75
0.24	Very Good	J	VVS2	62.8	57	336	3.94	3.96	2.48

內建資料: diamonds

```
library(ggplot2)
```

```
diamonds
```

```
library(dplyr)
```

```
diam <- diamonds %>% sample_n(1500)
```

carat

y

z

0.23

3.98

2.43

0.21

3.84

2.31

0.23

Good

E

VS1

56.9

65

327

4.05

4.07

2.31

0.29

Premium

I

VS2

62.4

58

334

4.20

4.23

2.63

0.31

Good

J

SI2

63.3

58

335

4.34

4.35

2.75

0.24

Very Good

J

VVS2

62.8

57

336

3.94

3.96

2.48

Template

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

data

+ before new line

```
ggplot(data = diam) +  
  geom_point(mapping = aes(x = carat, y = price))
```

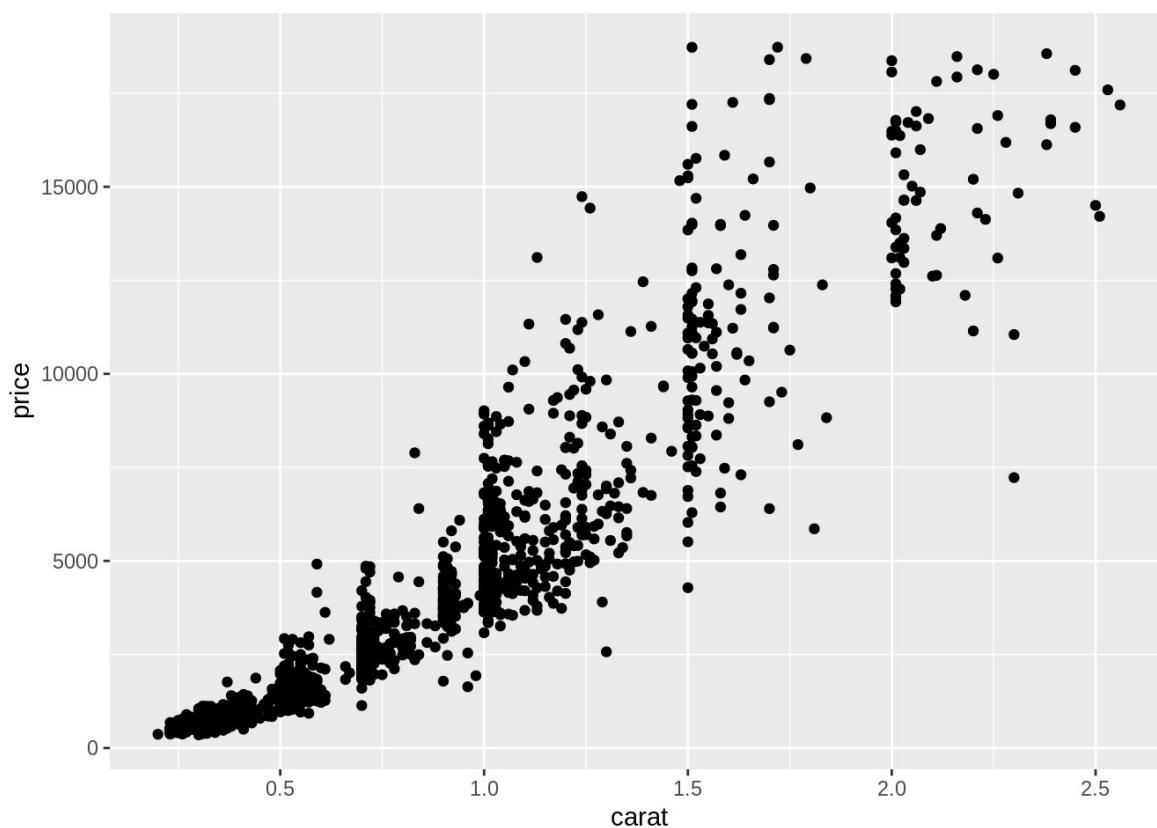
type of layer

aes()

x variable

y variable

散布圖 (scatter plot)



```
ggplot(data = diam) +  
  geom_point(mapping = aes(x = carat, y = price))
```

1. `ggplot()`:「初始化」圖片
2. `geom_*()`: 為圖片疊上新圖層

```
ggplot(data = diam) +  
  geom_point(mapping = aes(x = carat, y = price))
```

散布圖 (scatter plot)

Mappings

Aesthetic Mappings

視覺屬性

資料

點的 **x 軸位置**



carat
(鑽石重量)

點的 **y 軸位置**



price
(鑽石價錢)

點的 **顏色**



clarity
(鑽石淨度)

變項

```
ggplot(diam) + geom_point(aes(x = carat, y = price, color = clarity))
```

視覺屬性

Your Turn

Aesthetic Mappings

視覺屬性

變項

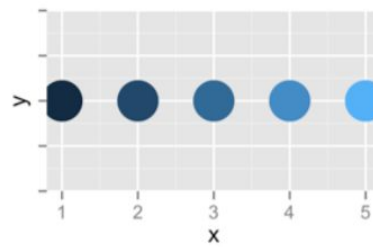
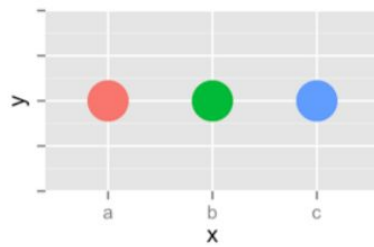
```
ggplot(diam) + geom_point(aes(x = carat, y = price, color = clarity))  
ggplot(diam) + geom_point(aes(x = carat, y = price, size = clarity))  
ggplot(diam) + geom_point(aes(x = carat, y = price, shape = clarity))  
ggplot(diam) + geom_point(aes(x = carat, y = price, alpha = clarity))
```

1. 比較 color, size, shape, alpha 這幾種視覺屬性的效果
2. 將視覺元素用在 **連續性** (e.g. carat) 和 **類別性** (e.g. clarity) 變項會發生什麼事？

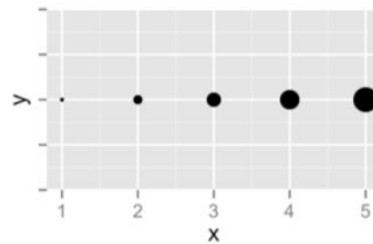
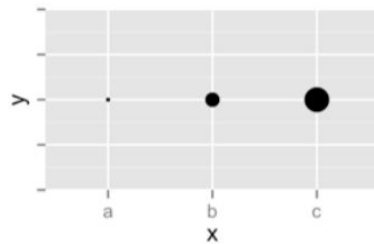
Discrete

Continuous

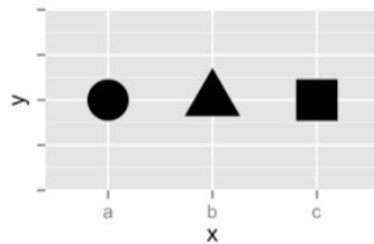
Color



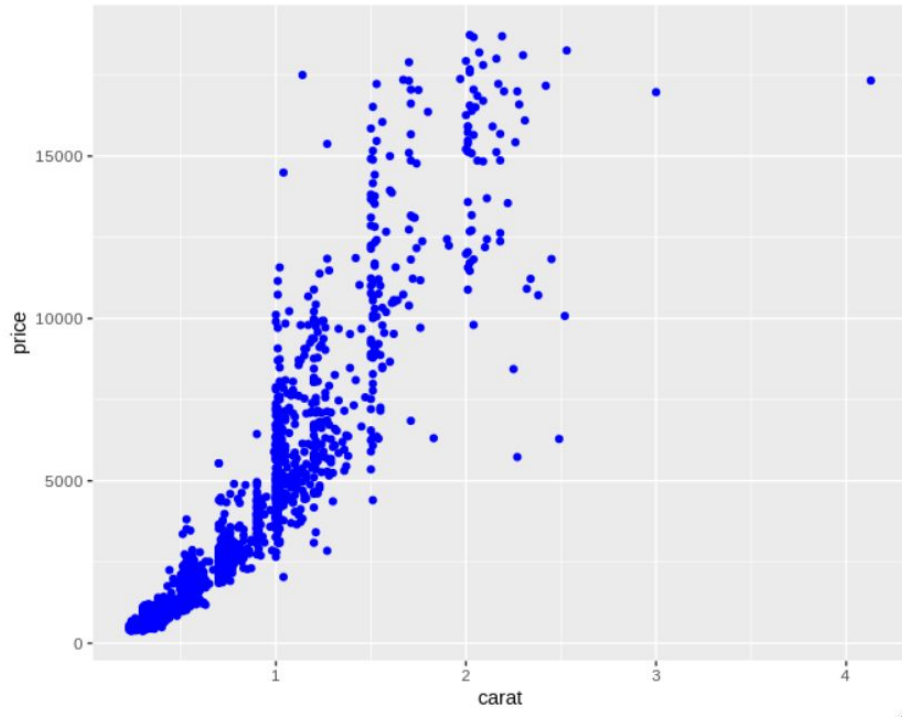
Size



Shape



Pop Quiz

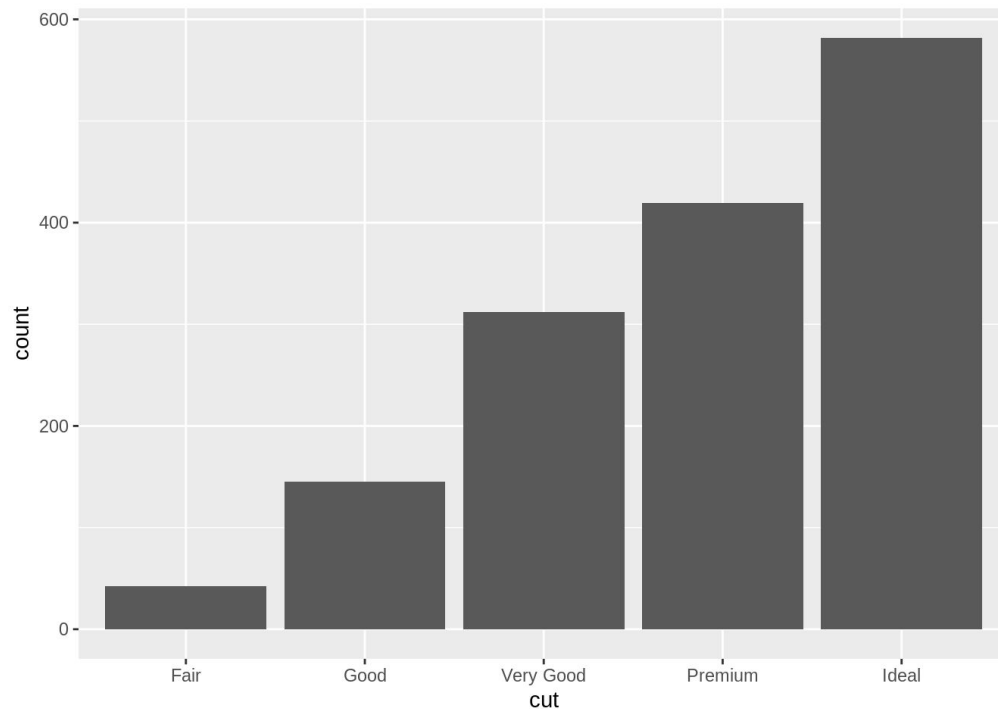


A `ggplot(diam) + geom_point(aes(x = carat, y = price, color = "blue"))`

B `ggplot(diam) + geom_point(aes(x = carat, y = price), color = "blue")`

Statistical Transformations

```
ggplot(diam) + geom_bar(aes(x = cut))
```



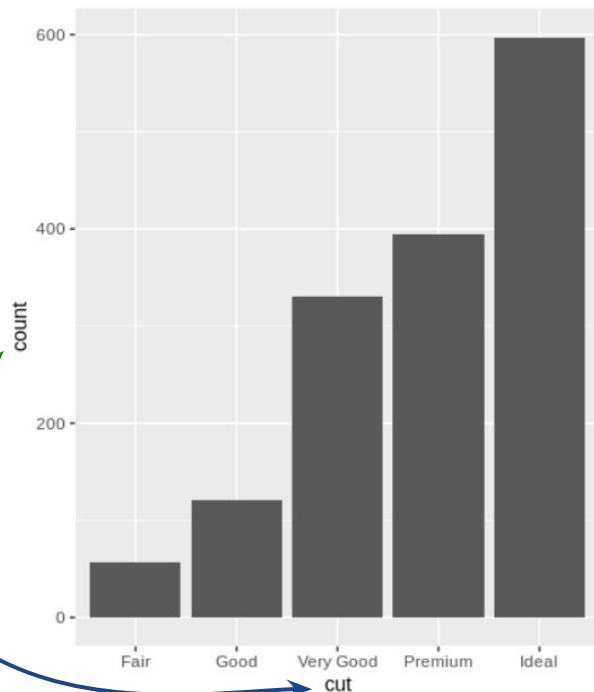
```
ggplot(diam) +  
  geom_bar(aes(x = cut), stat = "count")
```

stat_count()



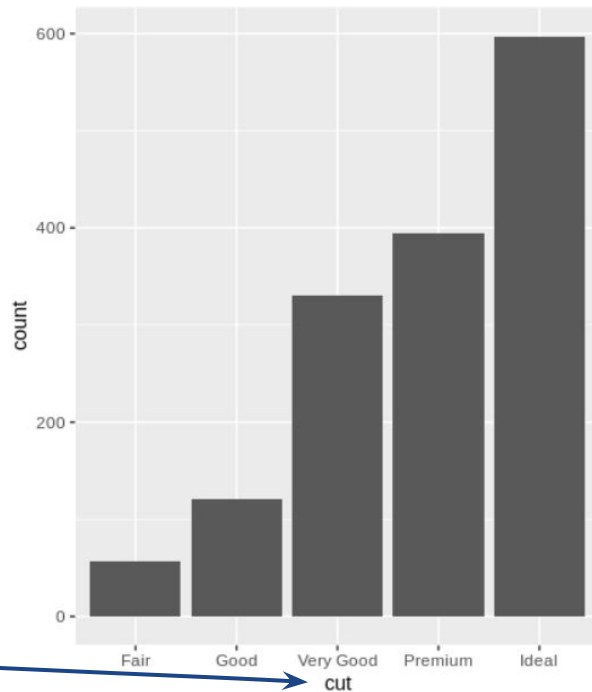
carat	cut	color	clarity	depth	table	price	x	y	z
0.23	Ideal	E	SI2	61.5	55	326	3.95	3.98	2.43
0.21	Premium	E	SI1	59.8	61	326	3.89	3.84	2.31
0.23	Good	E	VS1	56.9	65	327	4.05	4.07	2.31
0.29	Premium	I	VS2	62.4	58	334	4.20	4.23	2.63
0.31	Good	J	SI2	63.3	58	335	4.34	4.35	2.75
...

cut	count
Fair	1610
Good	4906
Very Good	12082
Premium	13791
Ideal	21551



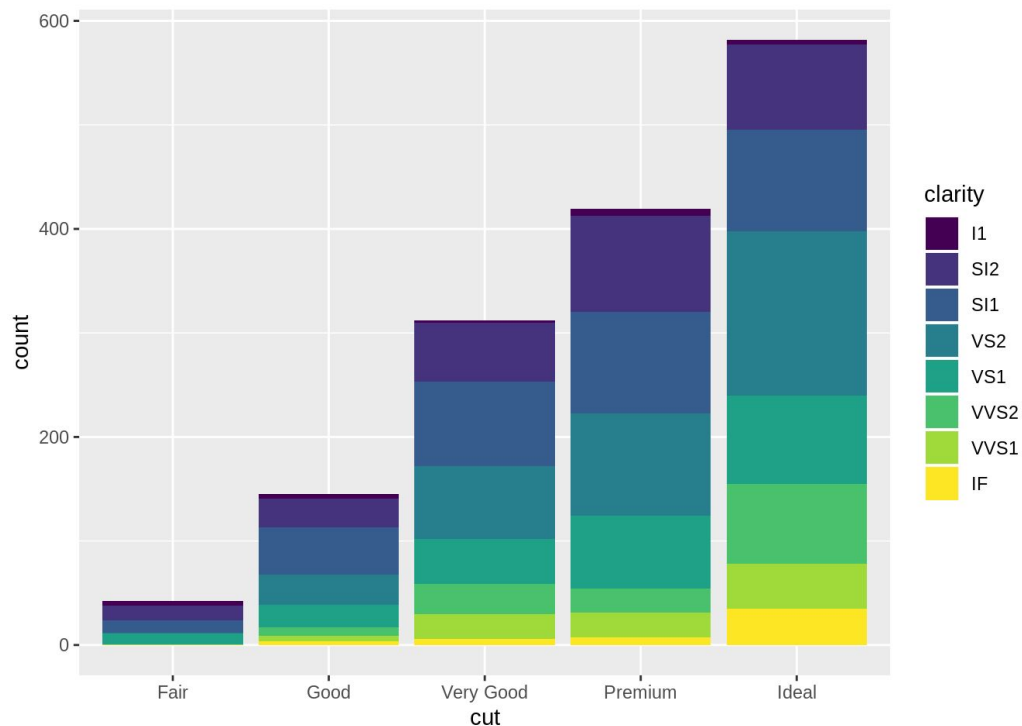
如果只有「整理好」的資料呢？

cut	count
Fair	1610
Good	4906
Very Good	12082
Premium	13791
Ideal	21551



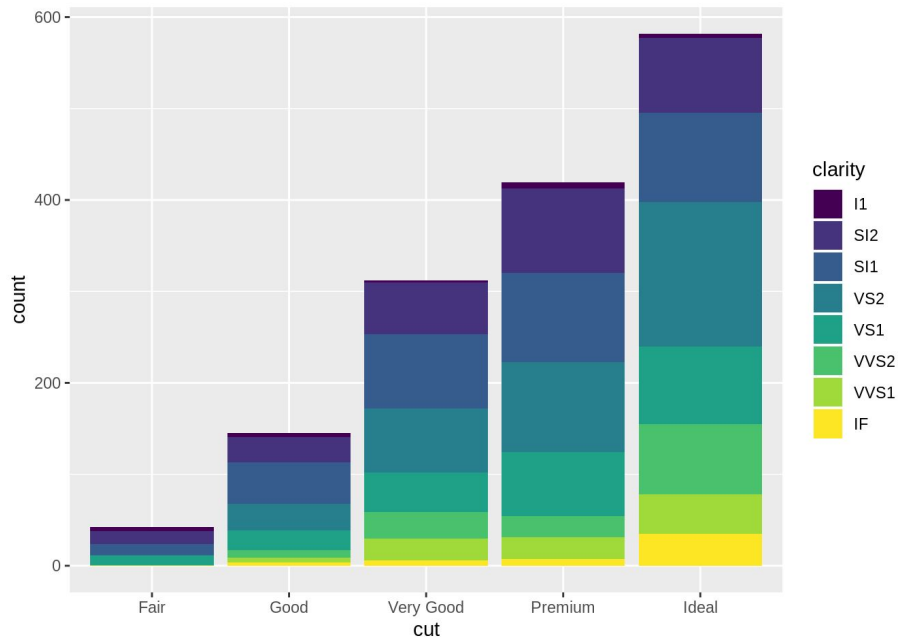
Position Adjustments

```
ggplot(diam) +  
  geom_bar(aes(x = cut, fill = clarity))
```



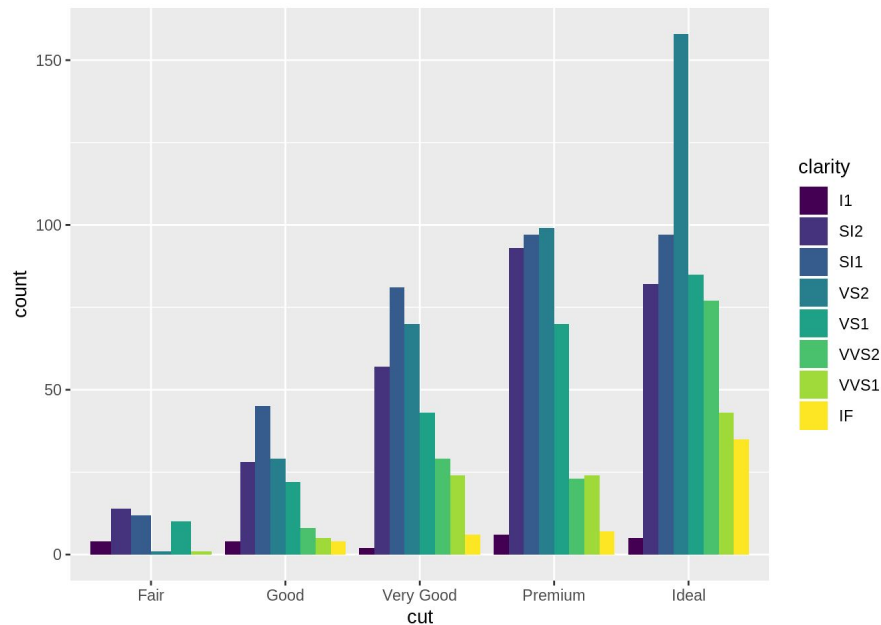
geom_bar 預設

```
ggplot(diam) +  
  geom_bar(aes(x = cut, fill = clarity),  
            position = "stack")
```



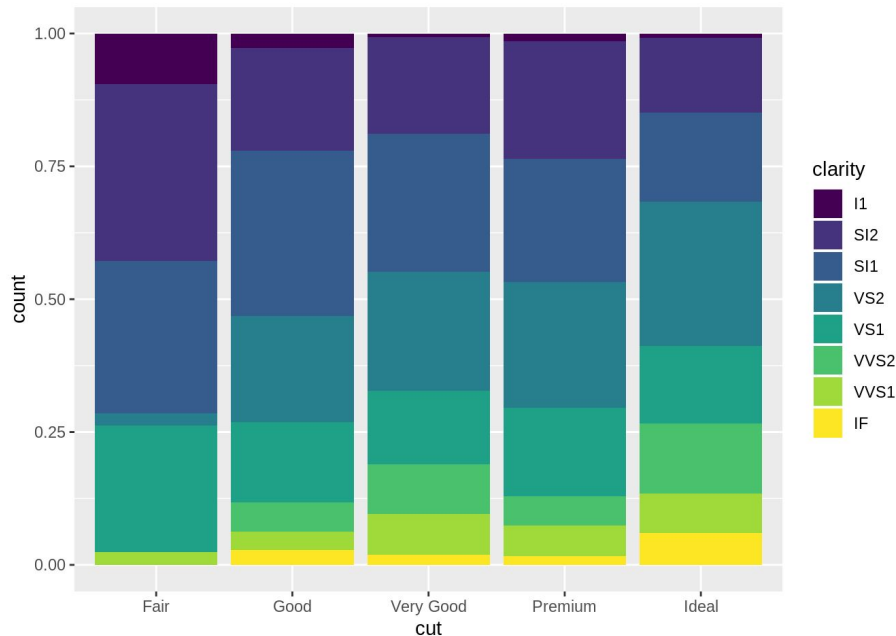
比較各 clarity 類別的數量

```
ggplot(diam) +  
  geom_bar(aes(x = cut, fill = clarity),  
            position = "dodge")
```



比較各 clarity 類別的比例

```
ggplot(diam) +  
  geom_bar(aes(x = cut, fill = clarity),  
           position = "fill")
```

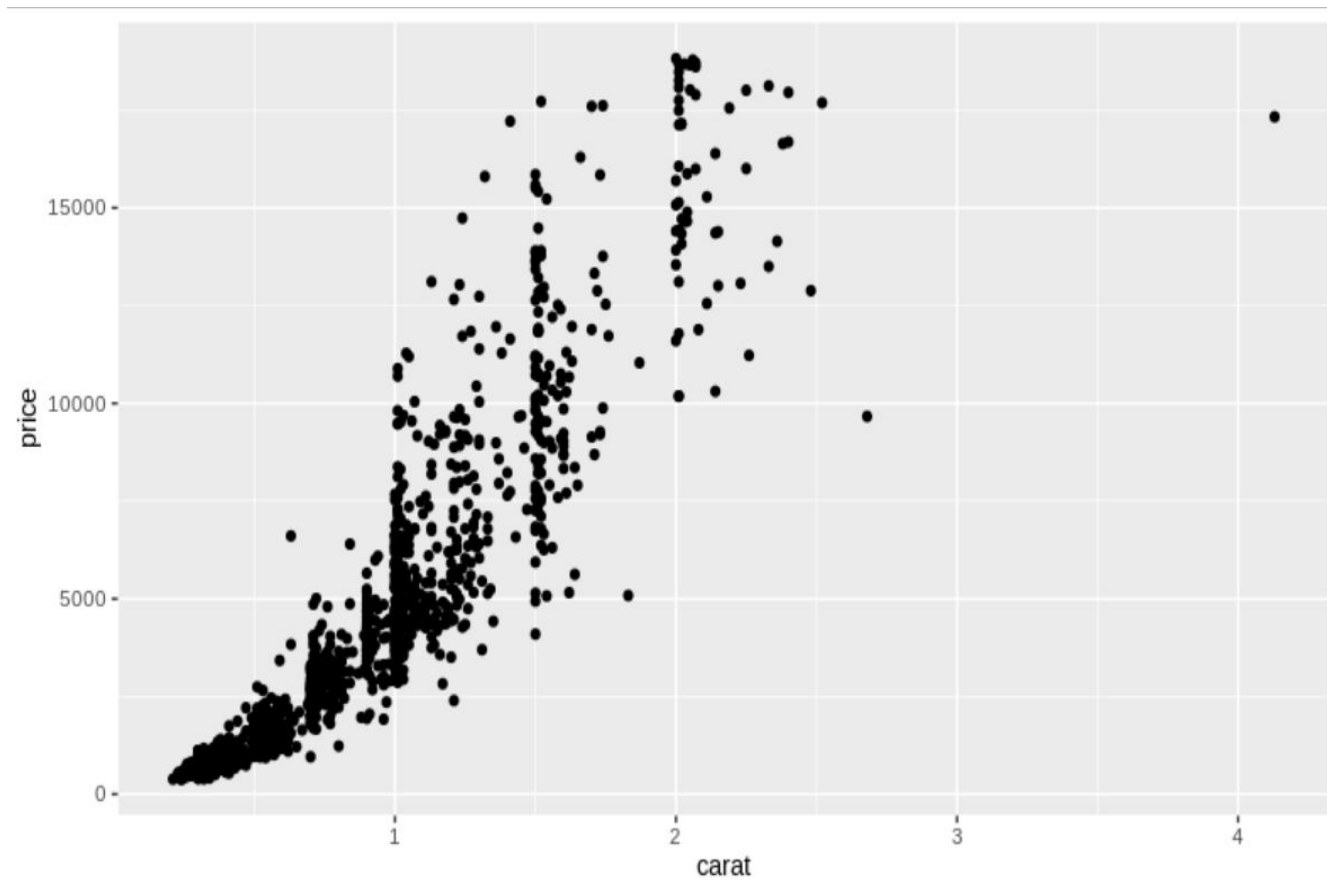


Template

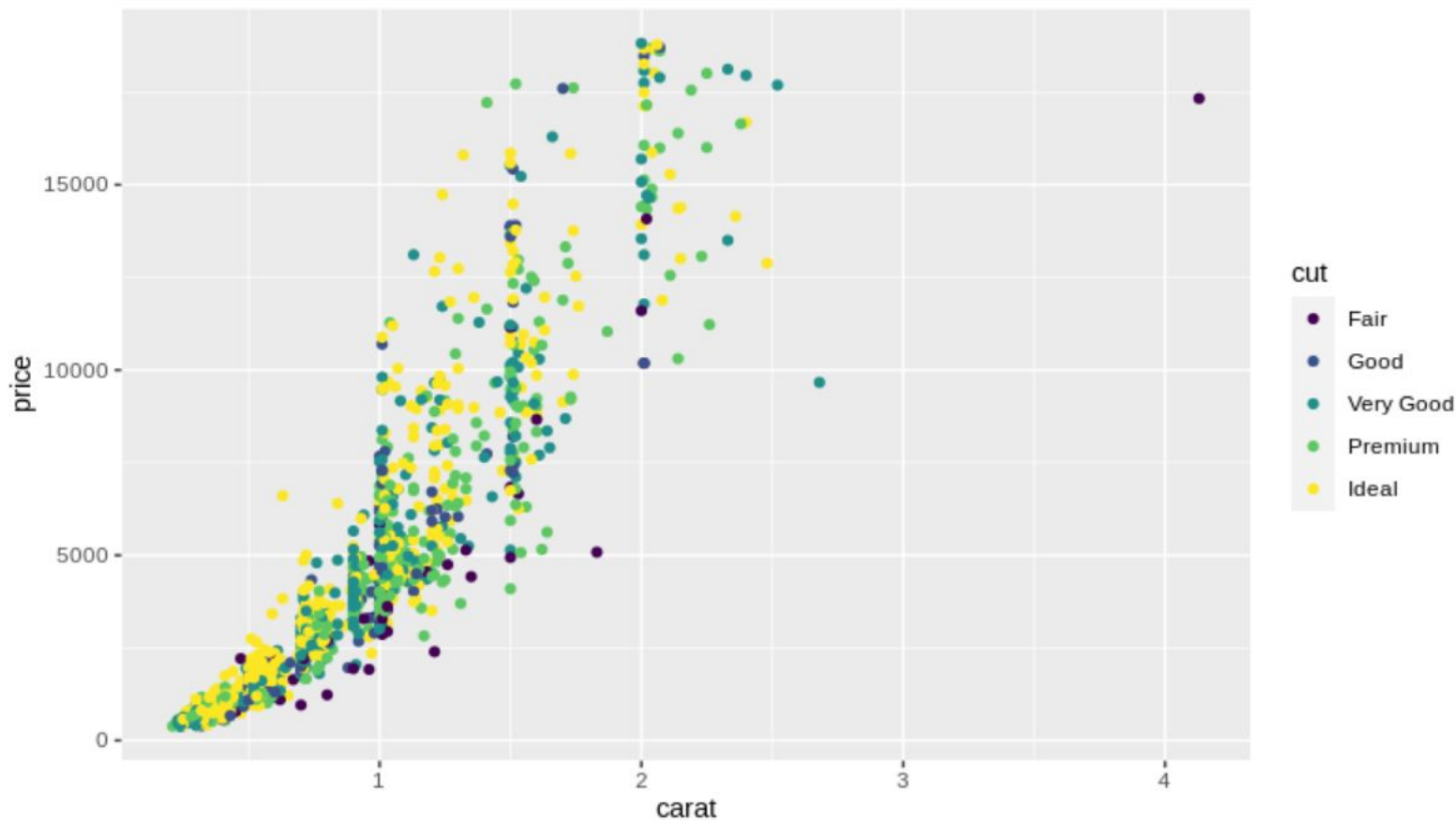
```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>),  
    stat = <STAT>,  
    position = <POSITION>  
  )
```

Optional

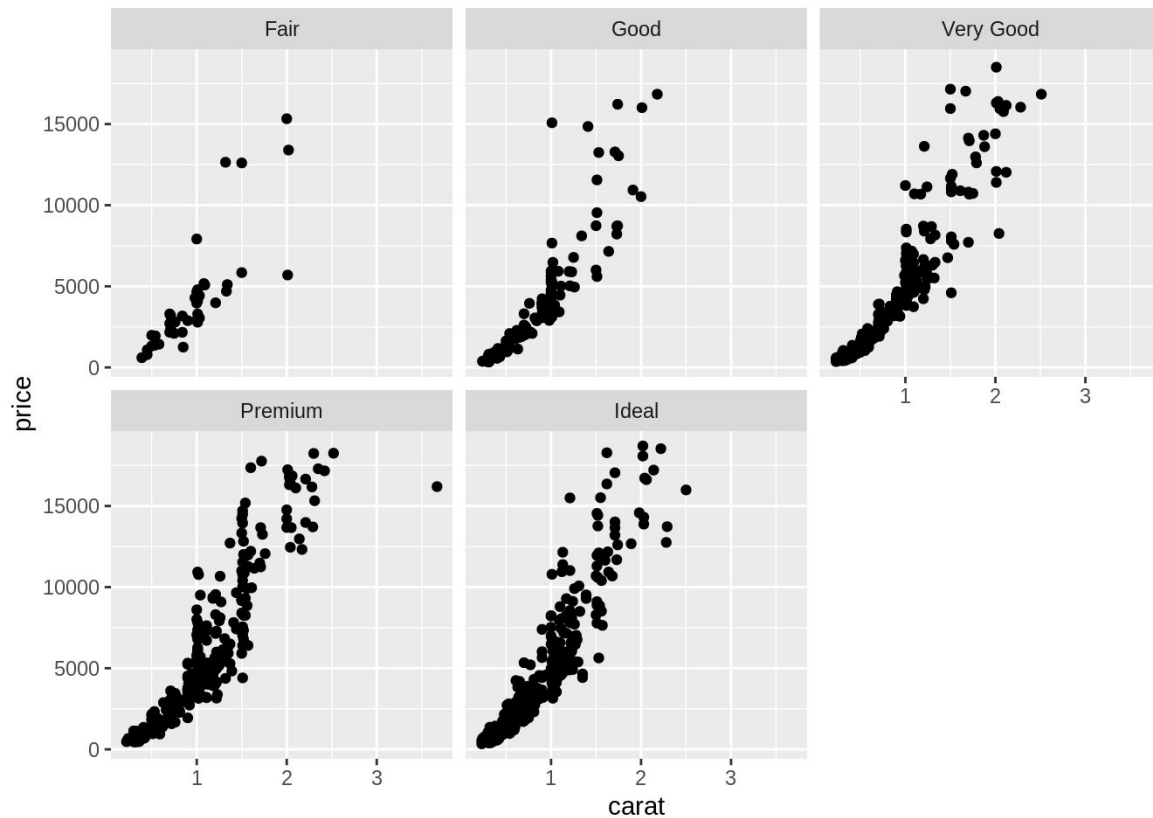
Facet



各種 cut 之下的散布圖



各種 cut 之下的散布圖



Template

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>),  
    stat = <STAT>,  
    position = <POSITION>  
  ) +  
  <FACET_FUNCTION>
```

Geoms

Geom Functions

Reference

geom_point

散布圖

geom_boxplot

箱形圖 (盒狀圖)

geom_histogram

直方圖

geom_bar

長條圖

...

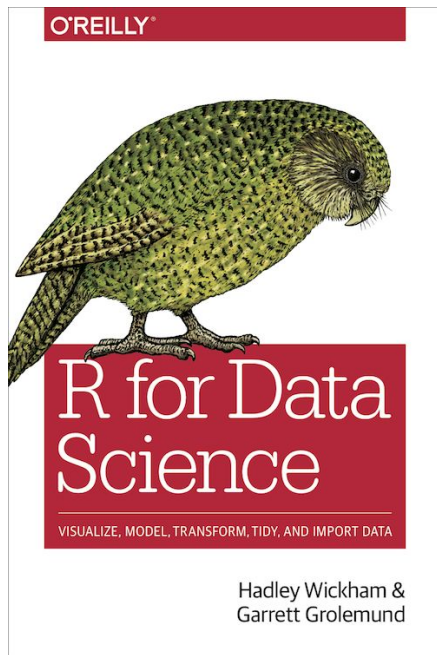
...

geom_sf

地圖

Cheatsheet

ggplot2 學習資源



Ch. 3 [Data visualisation](#)



ggplot2, 3rd edition