

資料前處理：

#讀入資料集並去除第一欄和資料分析無關的 comp_id

#檢查各欄位資料型態發現 op_profit_growth_rate, current_ratio 和 quick_rartio 為 chr 型態，使用 as.numeric()轉為 num 型態，同時使用 gsub()將千分位逗號去除避免轉換錯誤

#最終檢查完成皆為 num 型態

Code:

```
library(tidyverse)
```

```
findata <- read.csv("financialdata.csv")
```

```
findata <- findata[,-1]
```

```
attach(findata)
```

```
str(findata)
```

```
> str(findata)
```

```
'data.frame': 162 obs. of 16 variables:
 $ roe                : num  3.06 23.56 25.68 -3.41 10.9 ...
 $ roa                : num  2.21 17.84 14.29 -0.72 7.69 ...
 $ profit_margin_rate : num  4.4 39.45 16.82 3.86 13.99 ...
 $ gross_margin_rate  : num  18.1 50.6 37 16.9 34.3 ...
 $ expense_rate       : num  14.8 11 20.1 13 20.3 ...
 $ asset_turnover     : num  0.38 0.5 0.86 0.66 0.61 0.38 0.82 0.77 0.72 0.81 ...
 $ inventory_turnover : num  6.93 7.88 2.54 8.87 3.99 5.16 5.64 4.66 6.79 3.73 ...
 $ equity_turnnover   : num  0.69 0.67 1.59 1.55 0.89 0.5 1.43 1.45 0.88 4.16 ...
 $ rev_growth_rate    : num  0.96 3.11 41.75 12.73 13.07 ...
 $ margin_growth_rate : num  -10.93 4.21 116.49 13.25 35.83 ...
 $ op_profit_growth_rate: chr  "6.05" "2.01" "1,708.73" "22.41" ...
 $ cash_reinv_rate    : num  4.66 11.06 5.31 0.56 5.91 ...
 $ asset_growth_rate  : num  1.93 5.59 24.33 17.54 29.6 ...
 $ current_ratio      : chr  "158.03" "238.97" "187.85" "184.99" ...
 $ quick_rartio       : chr  "118.71" "215.17" "110.85" "158.19" ...
 $ debt_ratio         : num  45.7 23.6 44.2 51.4 30.1 ...
```

```
findata$op_profit_growth_rate <- as.numeric(gsub(",", "",
```

```
unlist(op_profit_growth_rate)))
```

```
class(findata$op_profit_growth_rate)#gsub 可替換符號
```

```
> class(findata$op_profit_growth_rate)#gsub可替換符號
```

```
[1] "numeric"
```

```
findata$current_ratio <- as.numeric(gsub(",", "", unlist(current_ratio)))
```

```
findata$quick_rartio <- as.numeric(gsub(",", "", unlist(quick_rartio)))
```

```
str(findata)
```

```
'data.frame': 162 obs. of 16 variables:
 $ roe : num 3.06 23.56 25.68 -3.41 10.9 ...
 $ roa : num 2.21 17.84 14.29 -0.72 7.69 ...
 $ profit_margin_rate : num 4.4 39.45 16.82 3.86 13.99 ...
 $ gross_margin_rate : num 18.1 50.6 37 16.9 34.3 ...
 $ expense_rate : num 14.8 11 20.1 13 20.3 ...
 $ asset_turnover : num 0.38 0.5 0.86 0.66 0.61 0.38 0.82 0.77 0.72 0.81 ...
 $ inventory_turnover : num 6.93 7.88 2.54 8.87 3.99 5.16 5.64 4.66 6.79 3.73 ...
 $ equity_turnnover : num 0.69 0.67 1.59 1.55 0.89 0.5 1.43 1.45 0.88 4.16 ...
 $ rev_growth_rate : num 0.96 3.11 41.75 12.73 13.07 ...
 $ margin_growth_rate : num -10.93 4.21 116.49 13.25 35.83 ...
 $ op_profit_growth_rate: num 6.05 2.01 1708.73 22.41 79.26 ...
 $ cash_reinv_rate : num 4.66 11.06 5.31 0.56 5.91 ...
 $ asset_growth_rate : num 1.93 5.59 24.33 17.54 29.6 ...
 $ current_ratio : num 158 239 188 185 229 ...
 $ quick_rartio : num 119 215 111 158 173 ...
 $ debt_ratio : num 45.7 23.6 44.2 51.4 30.1 ...
```

1.

#因為 PCA 分析結果比較難解釋，所以我使用 spca 分析，k 取 16(findata 有 16 個變數)*4(我取期待非 0 係數為 4)=64

#由 summary()中的 Proportion of Variance 可知道每個主成份能解釋多少變異，例如 PC1 可以解釋約 11.54%(0.1154)的變異

#由 summary()中的 Cumulative Proportion 可知達到 PC9 後可解釋 80%以上的變異，因此可知大概需要取 9 個 PC 解釋這筆資料

Code:

```
library(nsprcomp)
```

```
spca <- nscumcomp(findata, k=64, nneg=T, scale=T)
```

```
summary(spca)
```

```
> summary(spca)
Importance of components:
      PC1    PC2    PC3    PC4    PC5    PC6    PC7
Standard deviation 0.9067 0.8903 0.8744 0.8502 0.78070 0.74310 0.73882
Proportion of Variance 0.1154 0.1113 0.1073 0.1015 0.08556 0.07752 0.07663
Cumulative Proportion 0.1154 0.2267 0.3340 0.4355 0.52109 0.59861 0.67524
      PC8    PC9    PC10    PC11    PC12    PC13    PC14
Standard deviation 0.71378 0.6731 0.62675 0.56572 0.44299 0.43555 0.34539
Proportion of Variance 0.07152 0.0636 0.05515 0.04493 0.02755 0.02663 0.01675
Cumulative Proportion 0.74676 0.8104 0.86550 0.91043 0.93798 0.96461 0.98136
      PC15    PC16
Standard deviation 0.26380 0.25135
Proportion of Variance 0.00977 0.00887
Cumulative Proportion 0.99113 1.00000
```

2.

#由下圖，顏色越深表示該變數對該主成份越重要，所以第一主成份(PC1)的重點變數為 rev_growth_rate, debt_ratio, current_ratio, margin_growth_rate, asset_turnover；第二主成份(PC2)的重點變數為 inventory_turnover, quick_rartio, asset_growth_rate, cash_reinv_rate；第三主成份(PC3)的重要變數為 roe, current_ratio, expense_rate

(變數順序大致依照顏色深->淺列出)

Code:

```
library(reshape2)
```

```
ggplot(melt(spca$rotation[,1:3]), aes(Var2, Var1)) +
```

```
  geom_tile(aes(fill = value), colour = "white") +
```

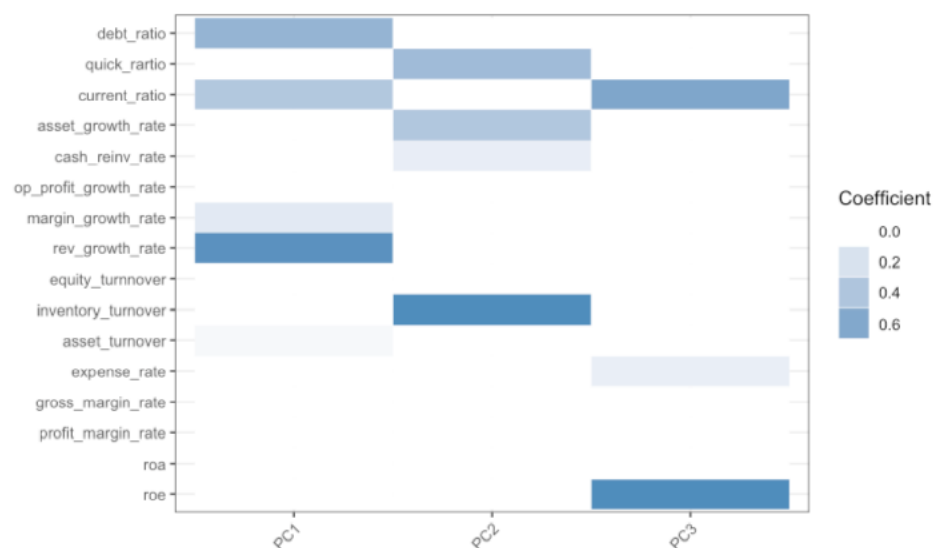
```
  scale_fill_gradient2(low = "white", high = "steelblue") +
```

```
  guides(fill=guide_legend(title="Coefficient")) +
```

```
  theme_bw() +
```

```
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1),
```

```
        axis.title = element_blank())
```



3.

承 2.第一主成份(PC1)的圖和得出的重要變數，又這些變數皆呈正相關，所以尋找適合投資的公司時可以檢查該公司有沒有高的 rev_growth_rate, debt_ratio, current_ratio, margin_growth_rate, asset_turnover