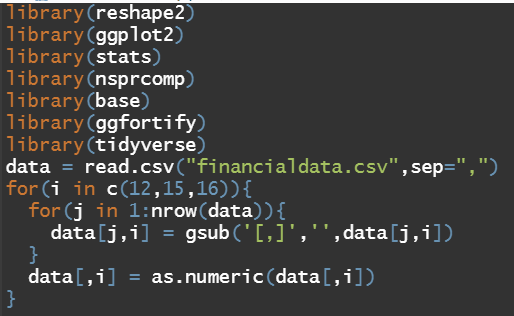
商業分析：SAS / R HW4

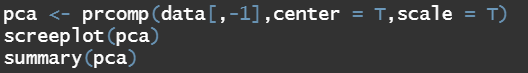
108208004 經濟三 白植允

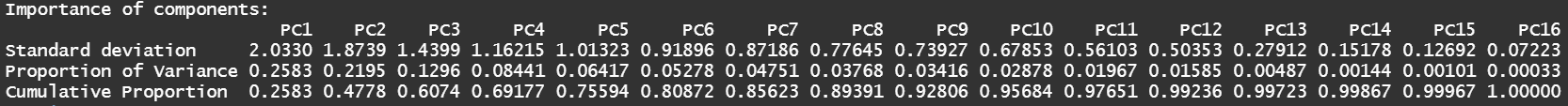
**分析財務報表，了解公司企業特性，以降維的方式，找出有意義的指標，衡量績優公司。**資料financialdata.csv 有163間公司的財務指標。

1. 以PCA或SPCA分析，找出每個主成份能解釋多少變異？大概需要多少個PC來解釋這筆資料？

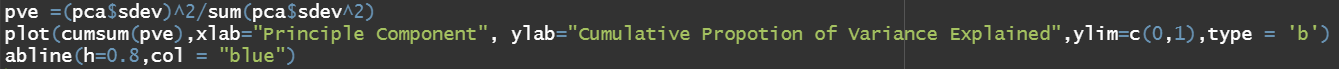


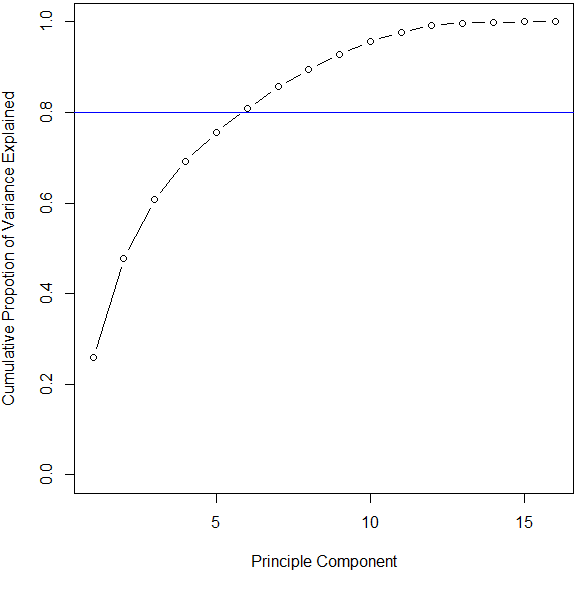
=>將需要的library列出以及將資料整理



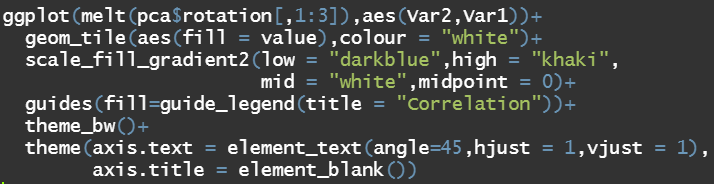
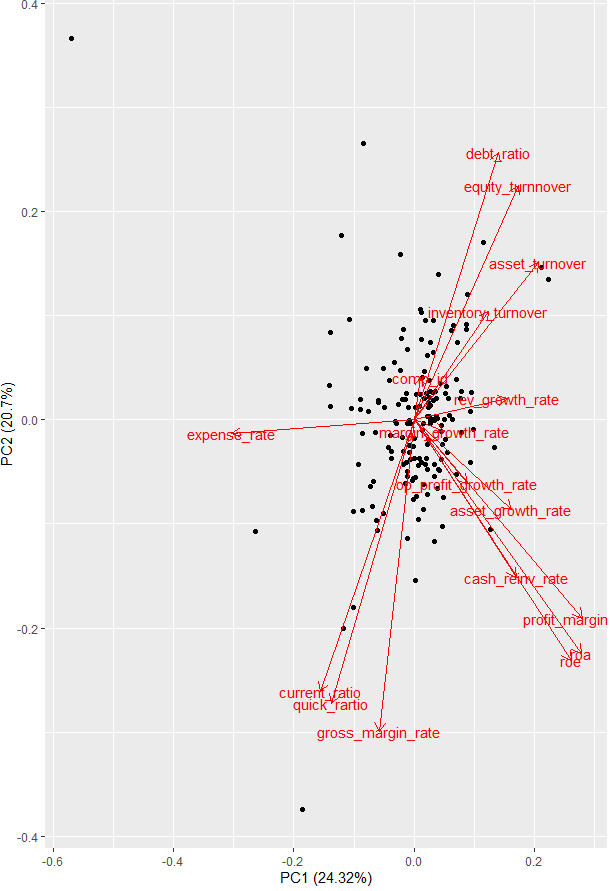
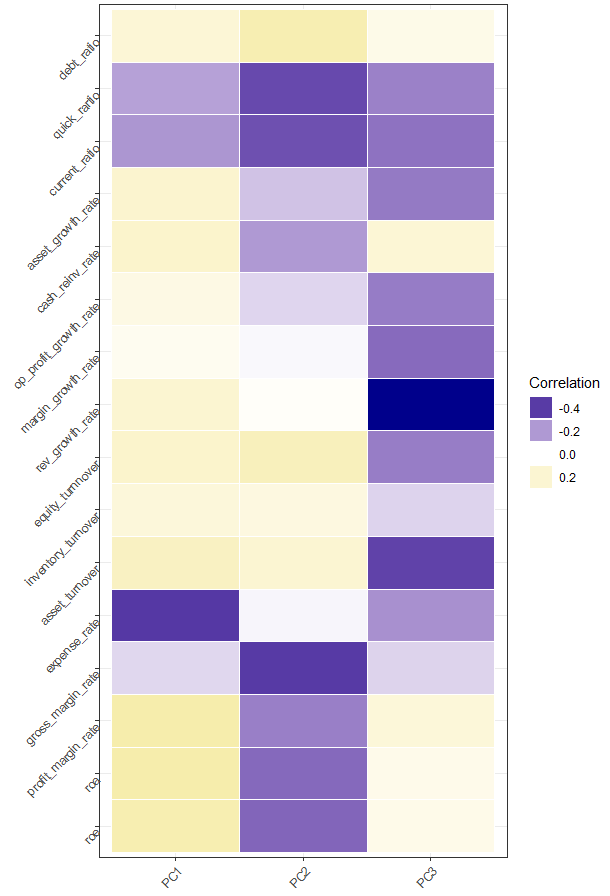
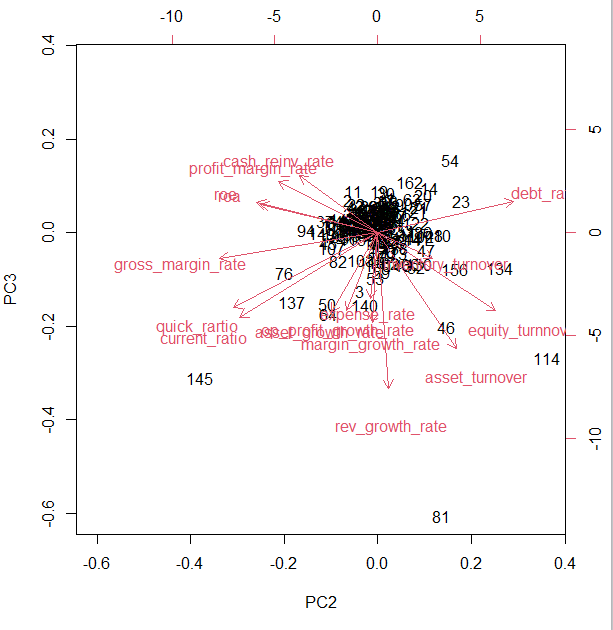
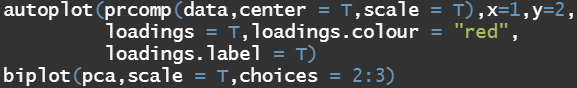


=>每個pc能解釋的變異如上圖「Proportion of Variance」所示





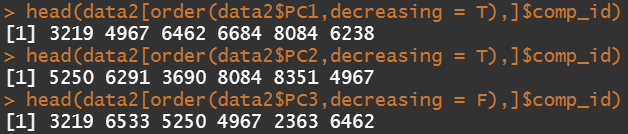
=>大約需6個才能解釋80%的變異

1. 找出前三個主成份分別重點變數為何並解釋。

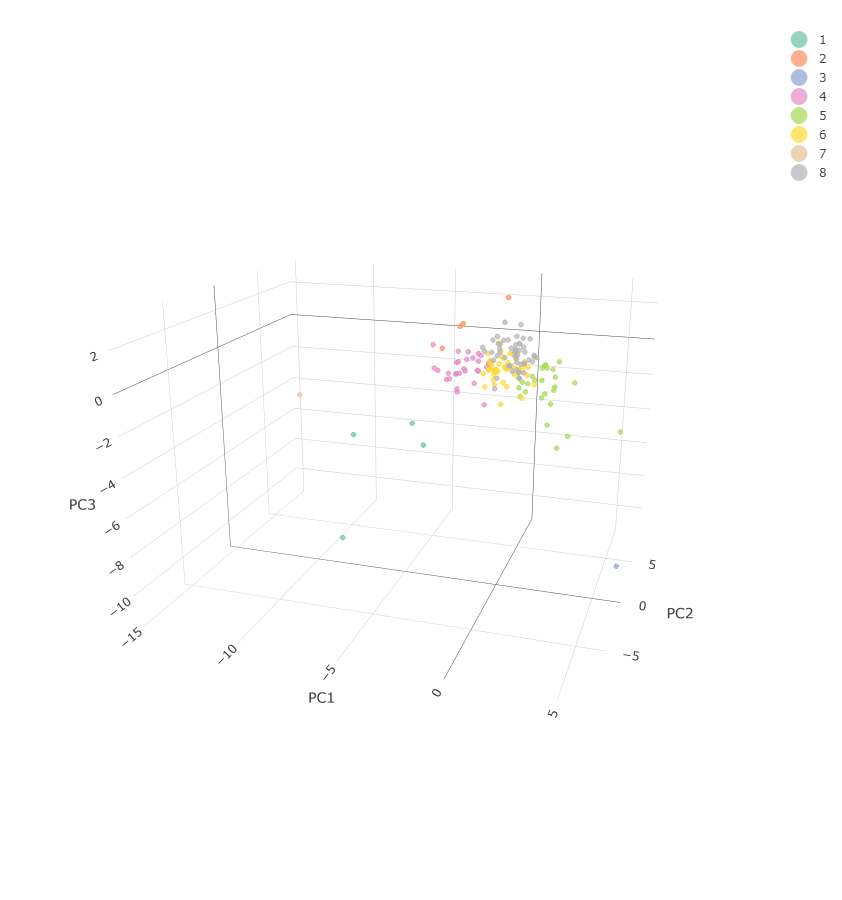
* PC1的重點變數是「營業費用率, 營業利益率, 股東權益報酬率, 資產報酬率」，可用來關注這家公司對股東權益的影響，用上述指標能看出一家公司獲利的能力。
* PC2的重點變數是「營業毛利率, 速動比率, 流動比例，負債比率,淨值週轉率」，可用來觀察一家公司的償債能力，以他的毛利率與其他債務負擔的比率來看這家公司的風險高低。
* PC3的重點變數是「營收成長率，資產週轉率」，這些變數可用來觀測一家公司的資產管理以及營收狀況，看出這家公司運用資產的效率。
* PC1數字越大，代表公司獲利能力越好，越值得投資。PC2數字越大， 代表公司倒債風險越高，需要注意。PC3數字越小，資產投資創造的營收效率越好，越值得投資。

1. 找出適合投資的公司。





=>3219的獲利能力最高且資產運用效率很好，而5250雖然目前營收指數不是前六名，但未來目前資金利用效率也不錯，且倒債風險很低，是值得持續關注的公司，4967的營運狀況也良好，雖然獲利能力可能低於3219，但是是屬於風險較低的公司，其他榜上的公司可以依各項指標做投資選擇。



=>可以到這裡看3D圖:<https://rpubs.com/Dragalia2001/897858>

附錄:程式碼

library(reshape2)

library(ggplot2)

library(stats)

library(nsprcomp)

library(base)

library(ggfortify)

library(tidyverse)

data = read.csv("financialdata.csv",sep=",")

for(i in c(12,15,16)){

for(j in 1:nrow(data)){

data[j,i] = gsub('[,]','',data[j,i])

}

data[,i] = as.numeric(data[,i])

}

pca <- prcomp(data[,-1],center = T,scale = T)

screeplot(pca)

summary(pca)

pve =(pca$sdev)^2/sum(pca$sdev^2)

plot(cumsum(pve),xlab="Principle Component", ylab="Cumulative Propotion of Variance Explained",ylim=c(0,1),type = 'b')

abline(h=0.8,col = "blue")

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

geom\_tile(aes(fill = value),colour = "white")+

scale\_fill\_gradient2(low = "darkblue",high = "khaki",

mid = "white",midpoint = 0)+

guides(fill=guide\_legend(title = "Correlation"))+

theme\_bw()+

theme(axis.text = element\_text(angle=45,hjust = 1,vjust = 1),

axis.title = element\_blank())

autoplot(prcomp(data,center = T,scale = T),x=1,y=2,

loadings = T,loadings.colour = "red",

loadings.label = T)

biplot(pca,scale = T,choices = 2:3)

data2 <- cbind(data,pca$x)

head(data2[order(data2$PC1,decreasing = T),]$comp\_id)

head(data2[order(data2$PC2,decreasing = T),]$comp\_id)

head(data2[order(data2$PC3,decreasing = F),]$comp\_id)

k <- kmeans(data2[,18:20],centers = 8)

group <- as.factor(k$cluster)

data3 <- cbind(data2[,c(1,18:20)],group)

library(plotly)

plot\_ly(data3, x = ~PC1, y = ~PC2, z = ~PC3,color = group, size=20,text = ~paste("Comp\_id:",comp\_id)) %>%

add\_markers() %>%

layout(scene = list(xaxis = list(title = 'PC1'),

yaxis = list(title = 'PC2'),

zaxis = list(title = 'PC3')))