

MDS 的簡單介紹

20180614

鍾興潔

Outline

- Multidimensional scaling (MDS)要解決的問題
- Matrix MDS
 - 用飛行距離來做出地圖
 - PCA vs MDS
- Nonmatrix MDS
 - 用各科的成績繪製學生間的關係圖
- (paper)PTT的文字雲

MDS 要解決的問題

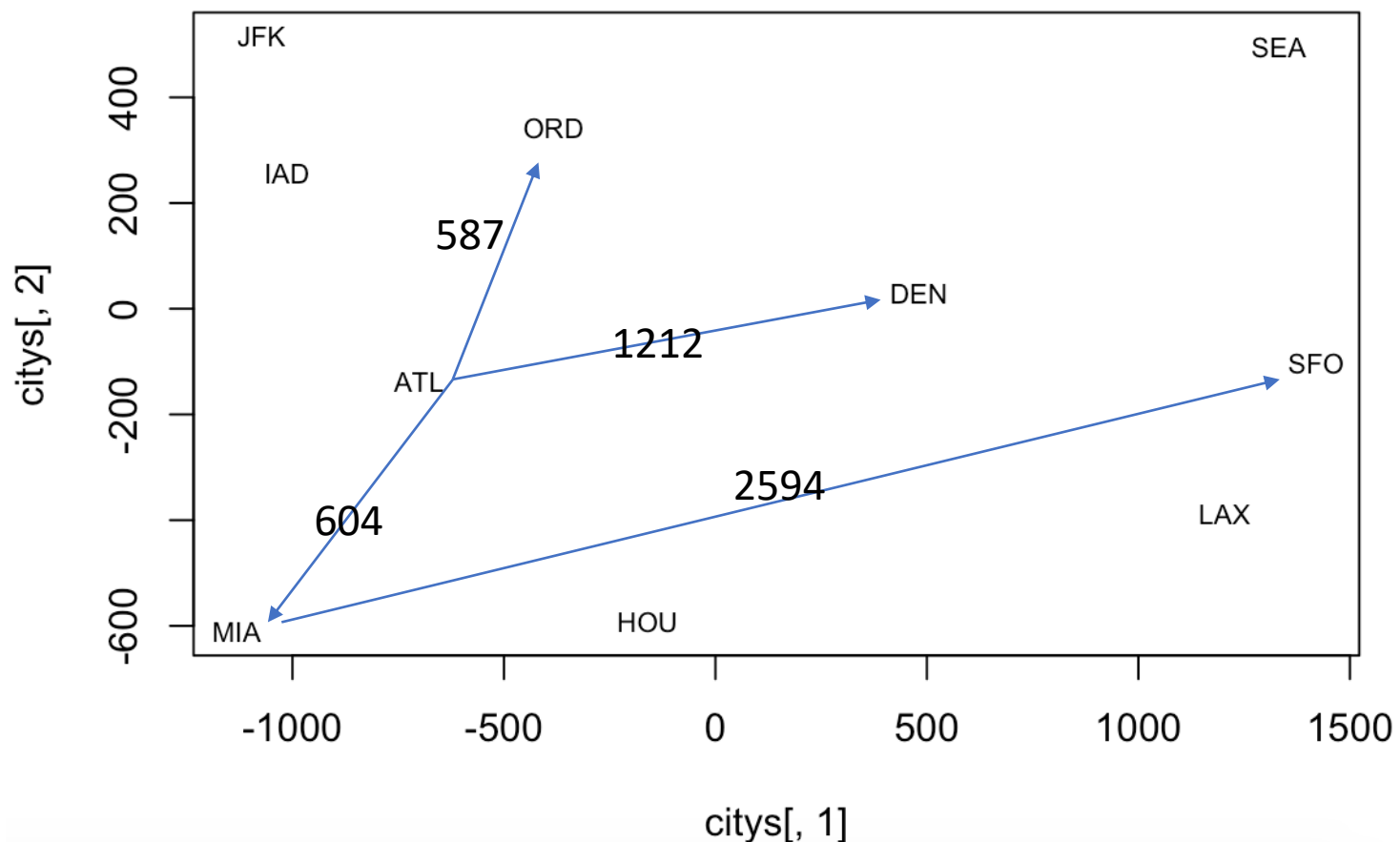
- 當 n 個物件（**object**）中各對物件之間的相似性（或距離）給定時，將這些物件在低維空間中的表示
- 使其盡可能與原先的相似性（或距離），「大體匹配」，使得由降維所引起的任何變形達到最小
- **matrix MDS** : 利用原始相似性（距離）的實際數值為間隔尺度和比率尺度
- **nonmatrix MDS** : 非**matrix** 的 MDS

用飛行距離來做出地圖

- 下表是美國十個城市之間的飛行距離，我們想在平面座標上據此標出這**10**城市之間的相對位置，使之盡可能接近表中的距離資料

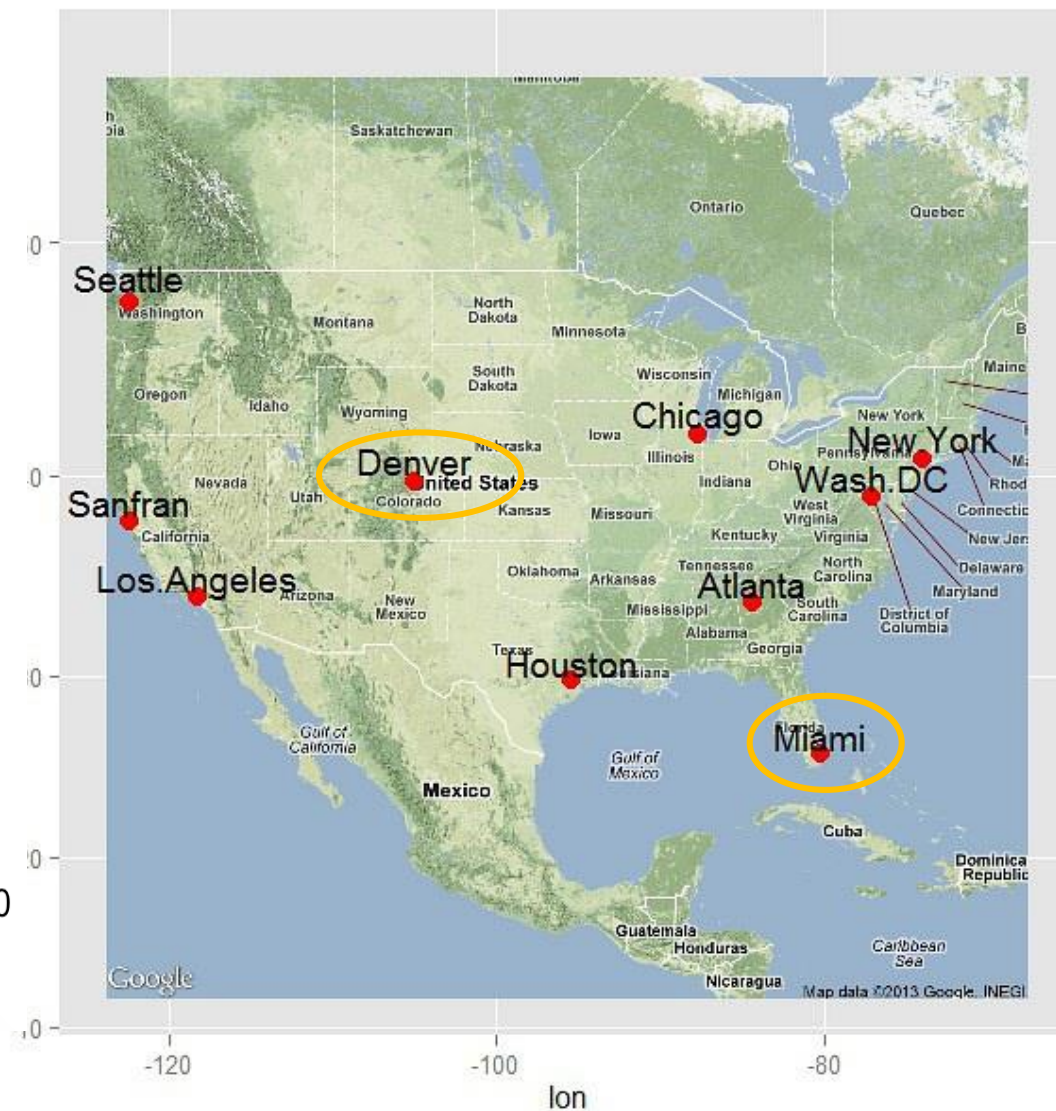
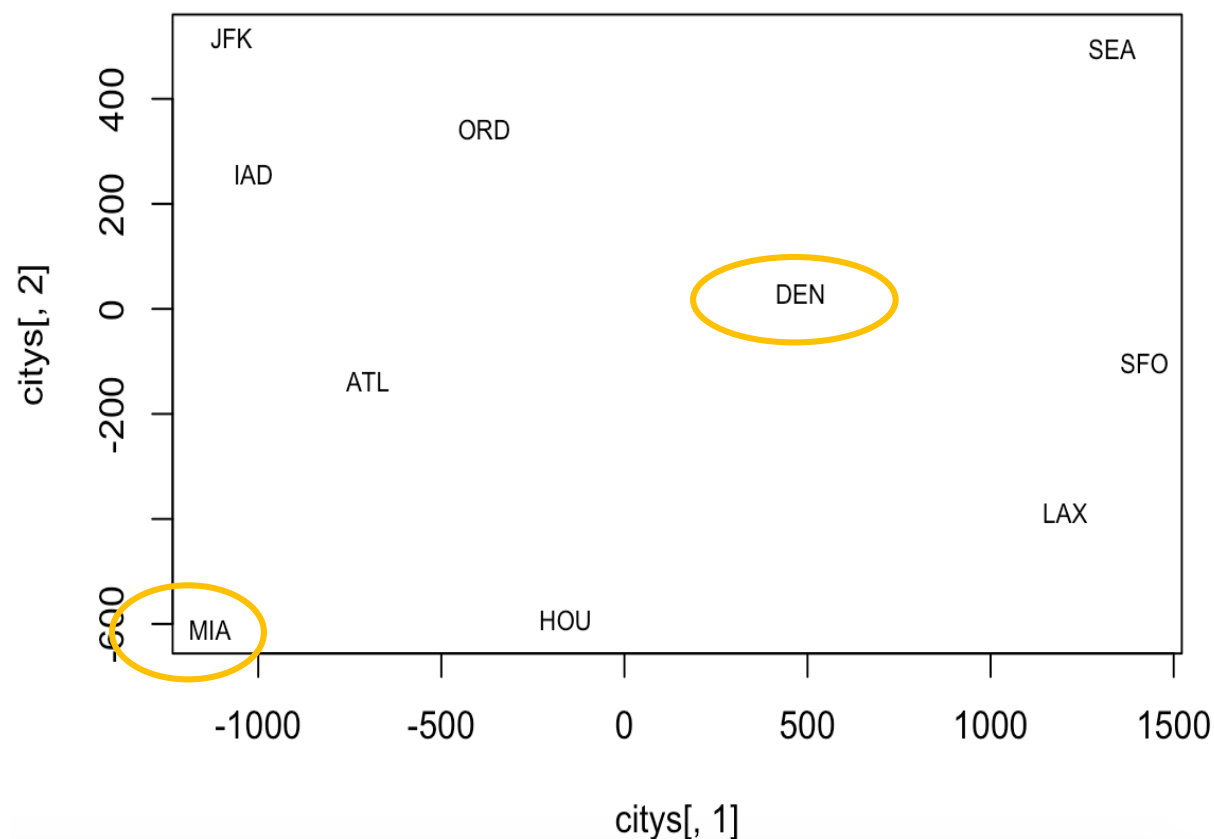
	A	B	C	D	E	F	G	H	I	J	K
1		ATL	ORD	DEN	HOU	LAX	MIA	JFK	SFO	SEA	IAD
2	ATL	0	587	1212	701	1936	604	748	2139	2182	543
3	ORD	587	0	920	940	1745	1188	713	1858	1737	597
4	DEN	1212	920	0	879	831	1726	1631	949	1021	1494
5	HOU	701	940	879	0	1374	968	1420	1645	1891	1220
6	LAX	1936	1745	831	1374	0	2339	2451	347	959	2300
7	MIA	604	1188	1726	968	2339	0	1092	2594	2734	923
8	JFK	748	713	1631	1420	2451	1092	0	2571	2408	205
9	SFO	2139	1858	949	1645	347	2594	2571	0	678	2442
10	SEA	2182	1737	1021	1891	959	2734	2408	678	0	2329
11	IAD	543	597	1494	1220	2300	923	205	2442	2329	0

用MDS得出降維後的相對座標

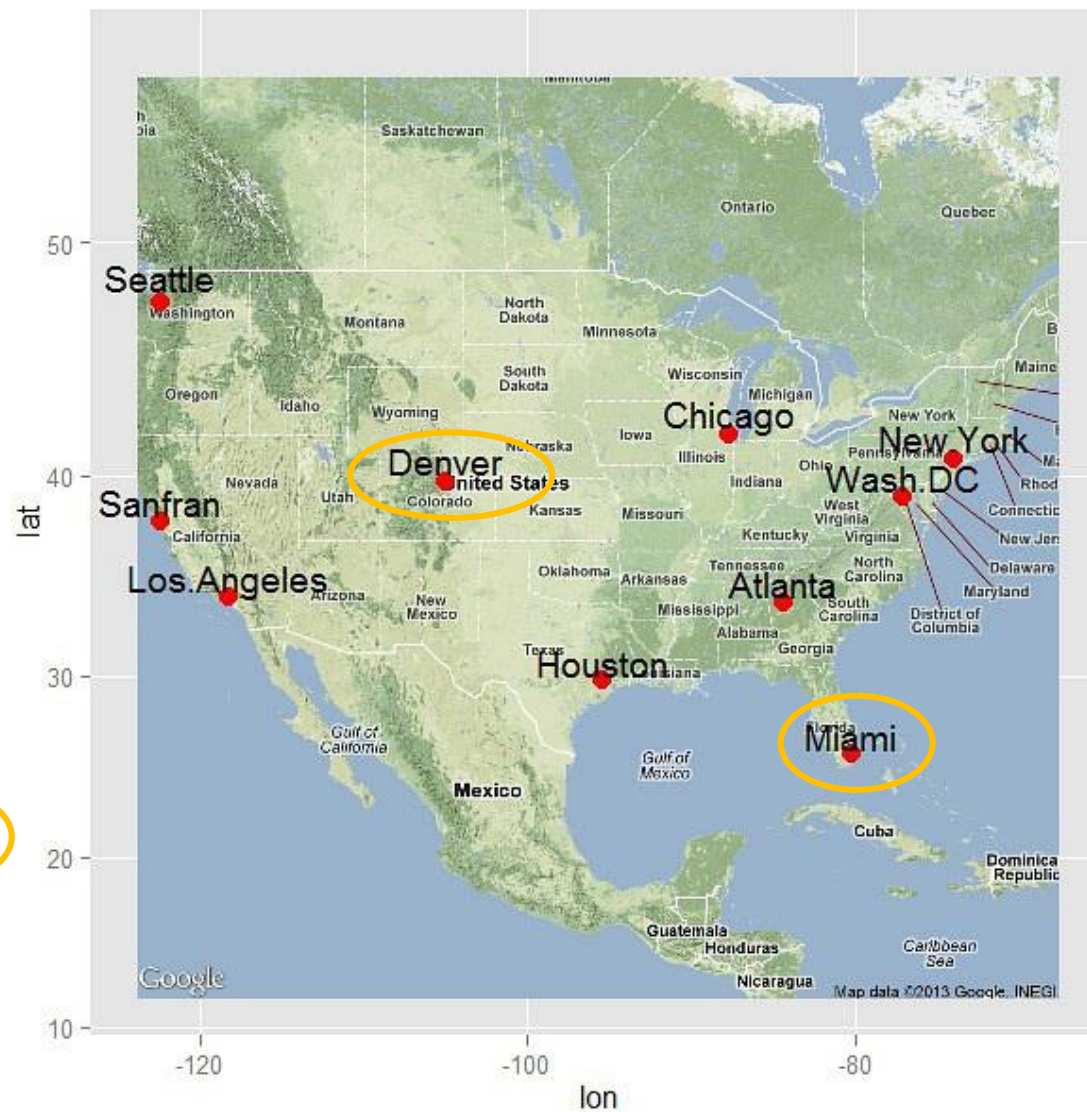
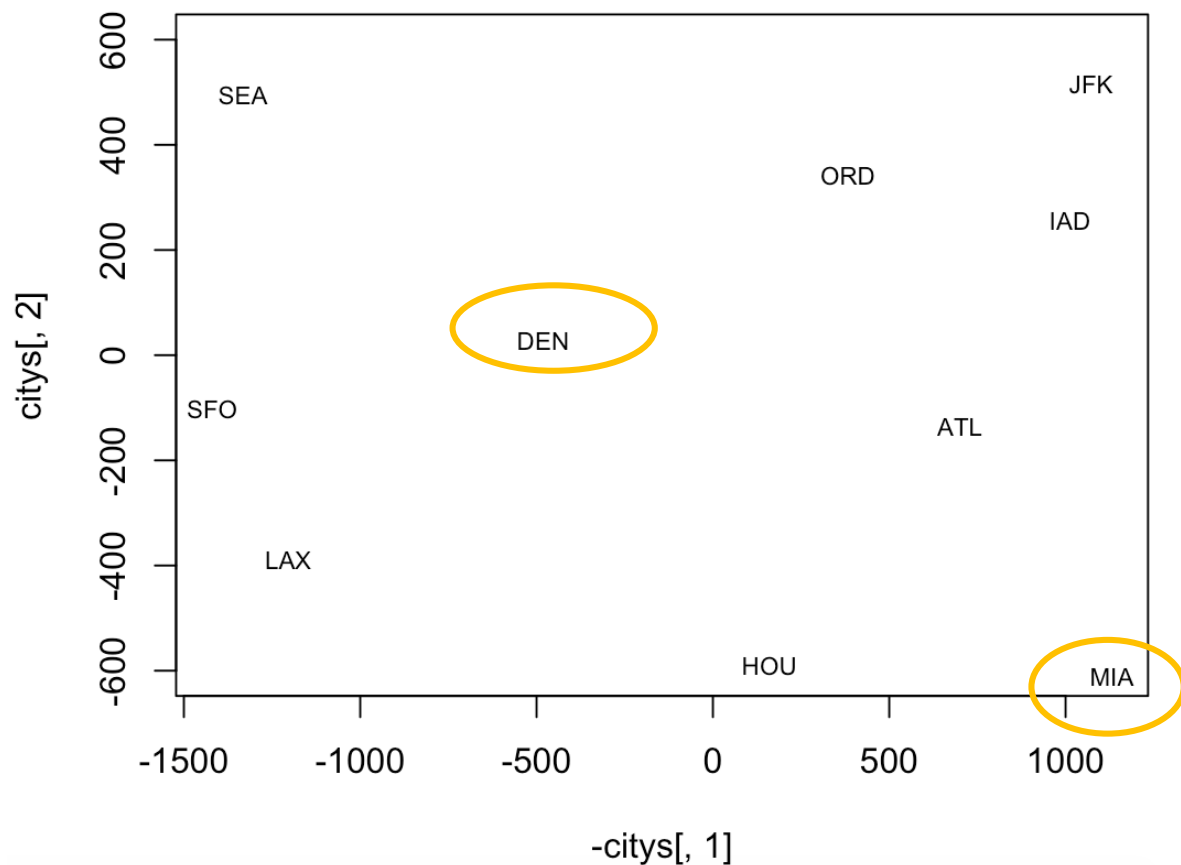


MDS地圖和實際地圖的比較

左右好像顛倒了!?



MDS地圖完成了



A practical algorithm of MDS

- 1. start with distances d_{ij}
- 2. define $A = -\frac{1}{2}d_{ij}^2$
- 3. put $B = (a_{ij} - a_{i\cdot} - a_{\cdot j} + a_{\cdot\cdot})$
- 4. find the eigenvalues $\lambda_1, \dots, \lambda_p$ and the associated eigenvectors $\gamma_1, \dots, \gamma_p$ where the eigenvectors are normalised
- 5. Choose an appropriate number of dimensions p (ideally $p = 2$)
- 6. The coordinates of the n points in the Euclidean space are given by $x_{ij} = \gamma_{ij} \lambda_j^{0.5}$ for $i = 1, \dots, n$ and $j = 1, \dots, p$.

PCA and MDS

Definition 17.1

A $(n \times n)$ distance matrix $\mathcal{D} = (d_{ij})$ is Euclidean if for some points $x_1, \dots, x_n \in \mathbb{R}^p$; $d_{ij}^2 = (x_i - x_j)^\top (x_i - x_j)$.

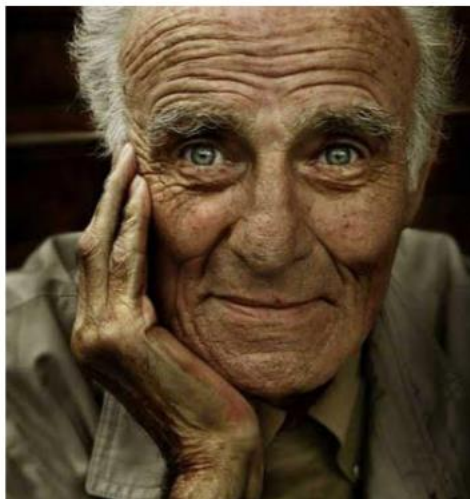
The following result tells us whether a distance matrix is Euclidean or not.

Theorem 17.1

Define $\mathcal{A} = (a_{ij})$, $a_{ij} = -\frac{1}{2}d_{ij}^2$ and $\mathcal{B} = \mathcal{H}\mathcal{A}\mathcal{H}$ where \mathcal{H} is the centering matrix. \mathcal{D} is Euclidean if and only if \mathcal{B} is positive semidefinite. If \mathcal{D} is the distance matrix of a data matrix \mathcal{X} , then $\mathcal{B} = \mathcal{H}\mathcal{X}\mathcal{X}^\top\mathcal{H}$. \mathcal{B} is called the inner product matrix.

MDS 要解決的問題

- Sometimes, there is no strict metric on original points
- Example: How much do you like the portraits?
(1: Not at all, 10: Very much)

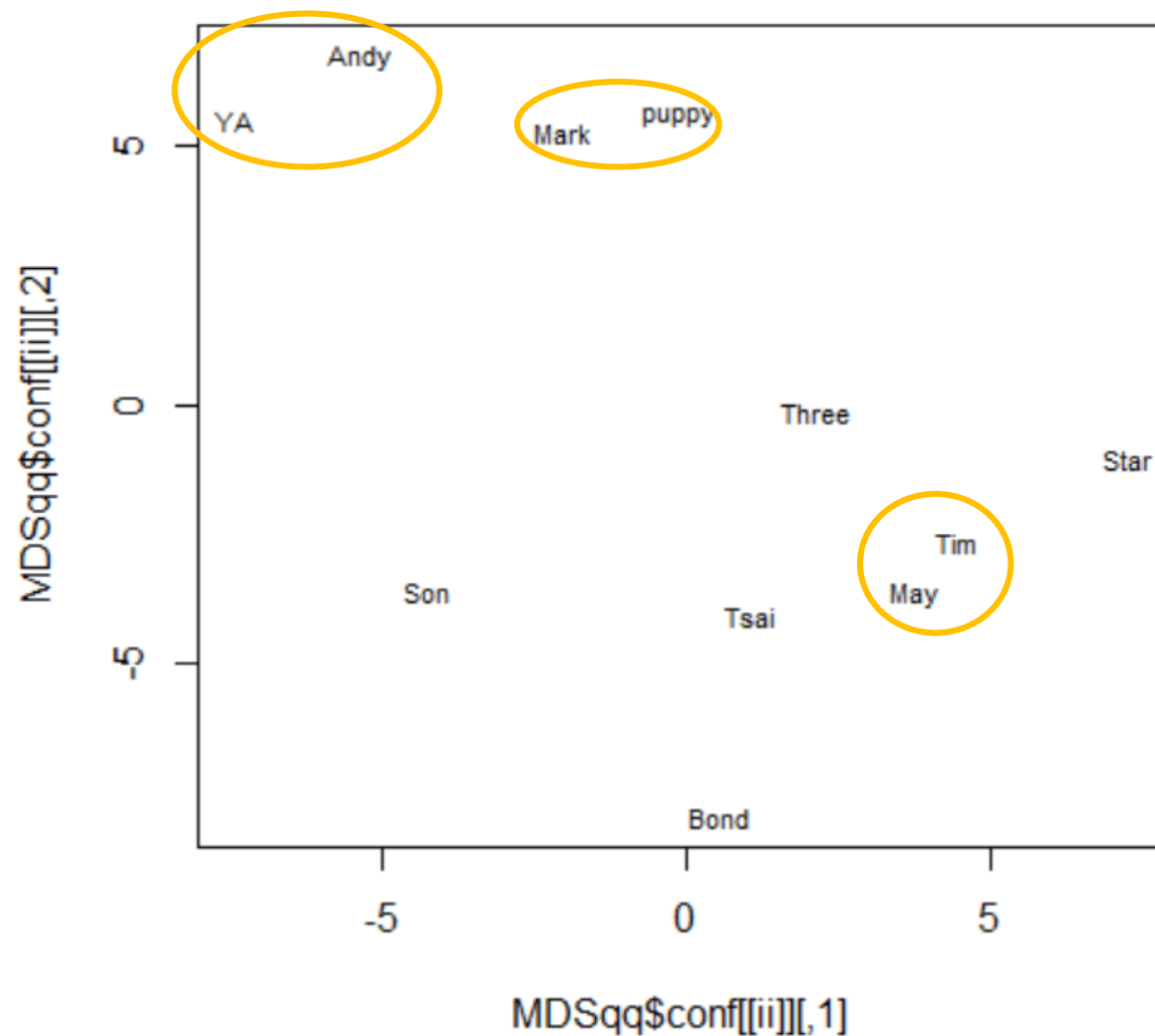


繪製學生間的綜合成績圖

- 下表是某校**12**位同學的各科表現(**GPA**)，我們想在平面上據此標出這**12**位同學間的相對位置，使之盡可能接近**GPA**的表現

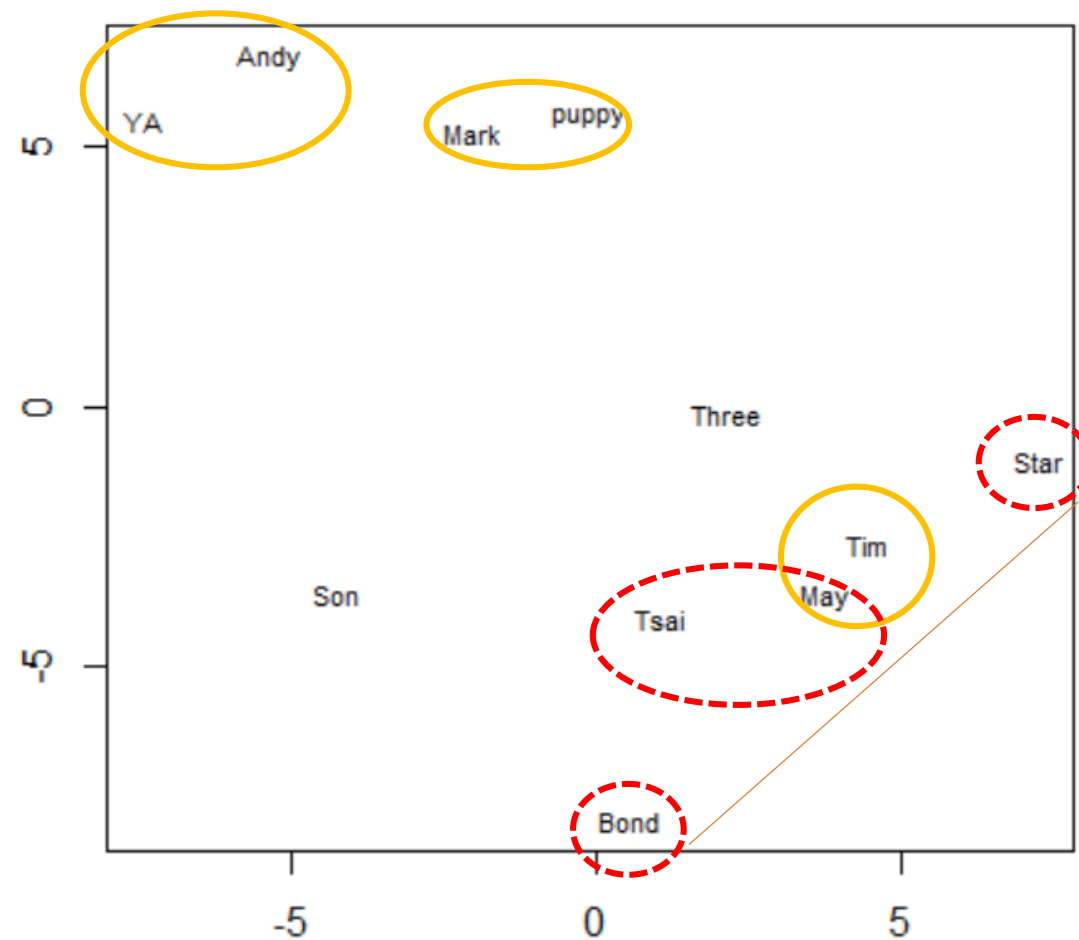
	APLM	ML	TS	NN	GLM	ASP	MS
Mark	9	8	9	6	8	8	5
Andy	10	9	10	9	8	10	8
YA	10	9	10	8	10	8	10
Three	5	5	5	5	5	7	4
Son	8	4	3	6	4	9	9
Tim	1	5	4	5	5	7	5
May	1	4	5	6	4	8	6
Tsai	2	5	5	3	5	7	8
Bond	3	2	3	2	2	9	8
Star	1	2	5	7	6	6	3
Puppy	7	9	10	5	7	7	6

用各科的成績繪製學生間的關係圖



平均值和關係圖的比較

	APLM	ML	TS	NN	GLM	ASP	MS	AVERAGE	DIFF
Mark	9	8	9	6	8	8	5	7.571428571	0.285714
Puppy	7	9	10	5	7	7	6	7.285714286	1.857143
Andy	10	9	10	9	8	10	8	9.142857143	0.142857
YA	10	9	10	8	10	8	10	9.285714286	4.142857
Three	5	5	5	5	5	7	4	5.142857143	1
Son	8	4	3	6	4	9	9	6.142857143	1.571429
Tim	1	5	4	5	5	7	5	4.571428571	0.285714
May	1	4	5	6	4	8	6	4.857142857	0.142857
Tsai	2	5	5	3	5	7	8	5	0.857143
Bond	3	2	3	2	2	9	8	4.142857143	0.142857
Star	1	2	5	7	6	6	3	4.285714286	4.285714

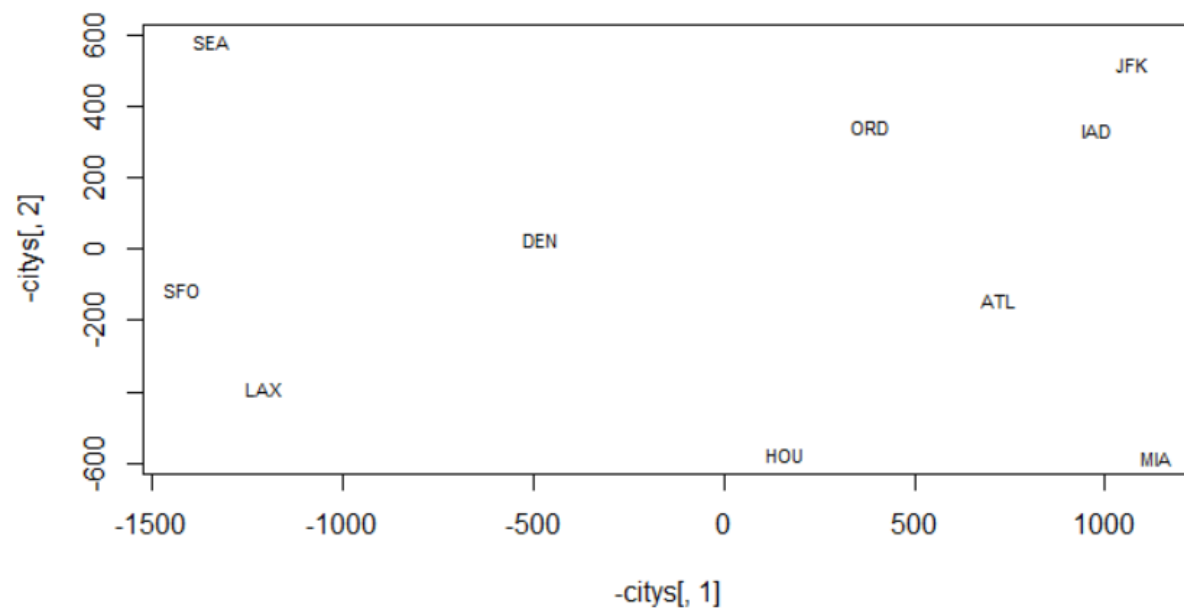
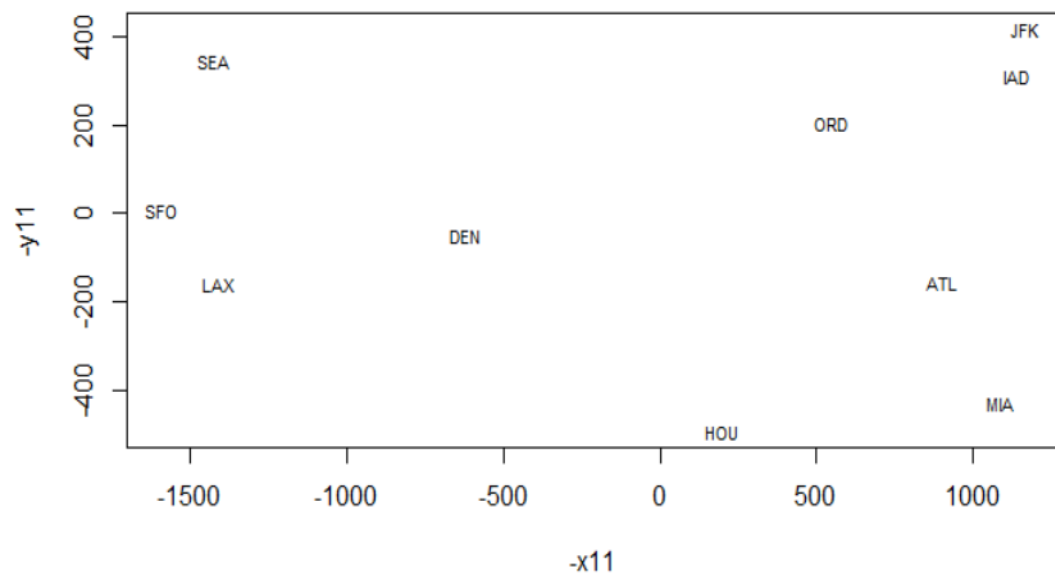


參考資料

- <https://blog.csdn.net/baimafujinji/article/details/79407478>
- <https://stat.ethz.ch/education/semesters/ss2012/ams/slides/v4.1.pdf>
- Textbook chapter 17
- <https://www.quora.com/Whats-the-difference-between-MDS-and-PCA>
- <http://sun.csim.scu.edu.tw/~jjhuang/Data/RM/20.ppt>

APPENDIX

手動MDS 套件MDS



20.3 多元尺度的模型適合度評估

進行 MDS 分析時通常以壓力係數(stress)作為衡量標準，根據Kruskal (1964)的解釋，不同的壓力係數水準，有其代表的配適程度。

壓力係數	配適程度
0.200	Poor (不好)
0.100	Fair (還可以)
0.050	Good (好)
0.025	Excellent (非常好)
0.000	Perfect (完全配合)

壓力係數之計算，是以在知覺圖中成對事物之距離(d_{ij})與其平均距離(\hat{d}_{ij})之差來計算。

20.3 多元尺度的模型適合度評估

多元尺度適合度評估是計算 R^2 並判定 R^2 是否合適， R^2 之計算公式如下：

$$R^2 = 1 - \frac{sse}{sst}$$

sse：殘差平方和(sum of squared error)

sst：整體平方和(sum of squared total)

R^2 的意涵是直接解釋最佳尺度資料(optimally scaled data)的變異數中，可由多元尺度法解釋的部份。因此多元尺度分析也可以以 R^2 作為同時參考的標準，當 R^2 愈大時(即越接近1)，表示配合性愈好，通常 R^2 在 0.9 以上即視為非常好的配適性。

20.4 多元尺度法範例

本案共有250位受測者，透過電腦處理首先將6個不同職業之受測者對上述afm、pm、acm、opf、ops、ks_co、ks_ks、ift、idt、ab、bli、cpi、cri等十三個影響線上社群行為意圖的變數之平均值求出如表20—2所示。茲利用此一平均值矩陣，將資料輸入電腦中，執行SPSS多元尺度分析軟體，即可計算afm~cpi各個線上社群項目在X軸與Y軸二元空間之座標如表20—3所示。

此座標位置可以在X軸與Y軸之空間中，點上座標位置點，並由位置點再將X軸與Y軸之原點與各變數之座標位置點相連結即得到各變數在二元空間之向量，請參看圖20—1所示。

20.4 多元尺度法範例

表 20 - 2 受測者對各種構面之平均評分

群組 變數名稱	學生1	公務員2	員工3	老闆4	自由業5	其它行業6
從屬動機(afm)	2.9148	2.1667	2.7080	3.0708	2.6875	4.0500
權力動機(pm)	4.2955	3.8704	4.4447	4.7972	5.4375	5.9000
成就動機(acm)	5.5000	5.3889	5.5841	5.8962	6.3750	6.0000
支持(opf)	4.3977	4.4074	4.1947	4.7170	4.5625	5.1000
網站效能(ops)	4.6250	4.6481	4.4115	5.0094	4.2500	6.2000
能力(ab)	5.1439	5.0251	5.2476	5.4845	6.0417	5.4000
與公司分享(ks_co)	3.7574	3.6909	3.5986	4.0378	3.7921	4.9993
與會員分享(ks_ks)	4.3258	4.3831	4.3362	4.7984	5.2917	5.4000
資訊基礎信任(ift)	5.1761	5.0556	5.0774	5.2689	5.6563	6.1000
認同基礎信任(idt)	4.8125	4.3981	4.6040	5.2594	5.5625	6.0000
品牌忠誠意圖(bli)	5.2501	5.1114	5.4306	6.0943	6.0417	5.8667
社群參與意圖(cpi)	4.8750	4.3704	4.5487	4.8774	4.8750	5.4000
社群推薦意圖(cri)	5.5455	5.4630	5.8274	6.0755	5.8750	6.0000

20.4 多元尺度法範例

表 20 - 3 十三項研究變數在知覺屬性空間之座標

變數名稱	X	Y
afm	3.5396	0.1184
pm	0.2326	0.5421
acm	-1.6183	0.0915
opf	0.6120	-0.1907
ops	0.2241	-0.7540
ab	-0.9574	0.3336
ks_co	1.6906	-0.1602
ks_ks	0.2417	0.2744
ift	-0.8423	-0.1490
idt	-0.3355	0.2418
bli	-1.3418	0.0652
cpi	0.1616	-0.1425
cri	-1.6066	-0.2709

20.4 多元尺度法範例

此外，以六個職業別1~6為計算單元，也一樣可以得到職業別1~6在X軸與Y軸二元空間之相對位置，這些相對位置若能與變數之相對位置比較，即可以看各種職業別比較重視哪些變數，而較不重視哪些變數。

Krusal壓力係數為0.04025， R^2 的值為0.99658，顯示空間構形與影響線上社群行為意圖點間距離資料配合相當好。

20.4 多元尺度法範例

表 20 - 4 不同職業別在知覺屬性空間之座標

變數名稱	X軸	Y軸
學生1	0.7598	0.3858
公務員2	1.4851	0.4257
員工3	0.9793	-0.0370
老闆4	-0.3408	-0.1417
自由業5	-0.5221	-1.2497
其它行業6	-2.3613	0.6168

20.4 多元尺度法範例

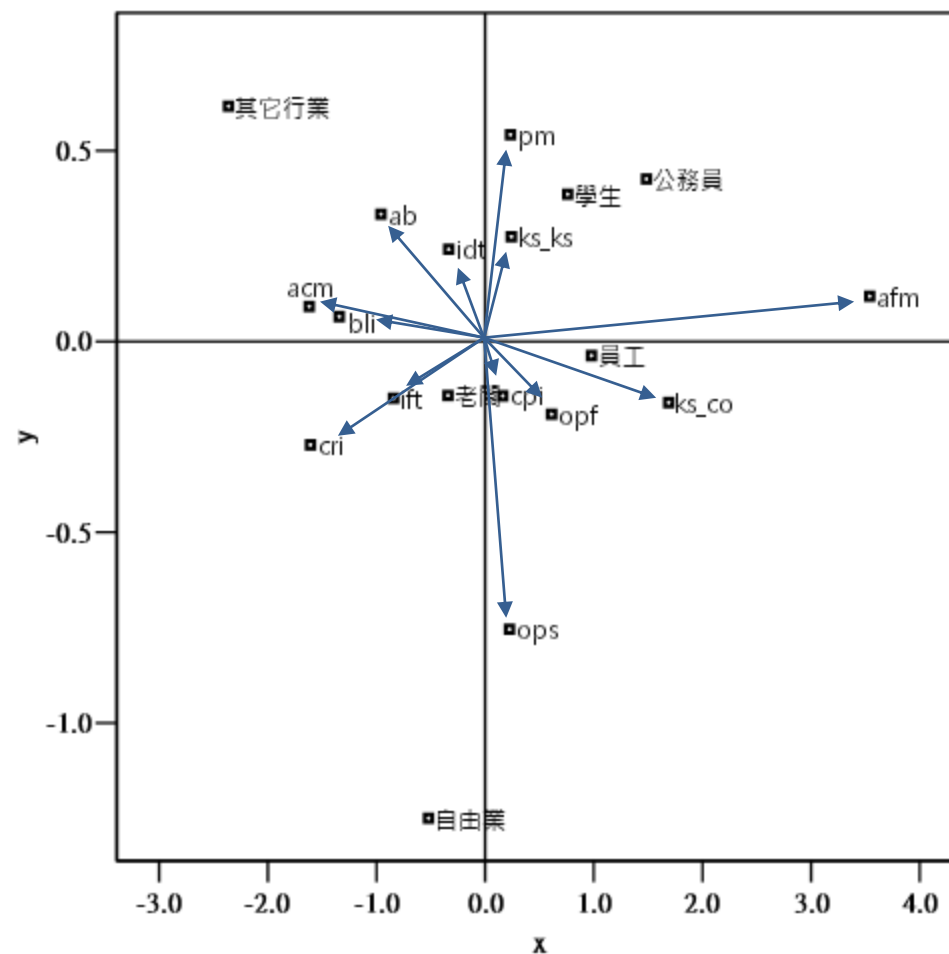


圖 20 - 3 不同職業別在不同構面之知覺空間圖

20.4 多元尺度法範例

表 20 - 5 不同職業別影響線上行為意圖

群組 變數名稱	學生 1	公務員 2	員工 3	老闆 4	自由業 5	其它 6
afm		◎	◎			
pm	◎	◎				
acm						◎
opf			◎	◎		
ops					◎	
ab						◎
ks_co			◎			
ks_ks	◎	◎				
ift				◎		
idt	◎					◎
bli						◎
cpi				◎		
cri					◎	

20.4 多元尺度法範例

由圖20—3之知覺圖與表20—5之示意圖，我們可以得知，
對於影響線上社群行為意圖過程中，

學生最重視權力動機、與會員分享及認同基礎；

公務員最重視從屬動機、權力動機以及與會員分享；

企業員工最重視從屬動機、支持以及與公司分享；

老闆最重視支持、資訊基礎與社群參與；

自由業最重視網站效能與社群推薦；