

# 三國的大數據時代

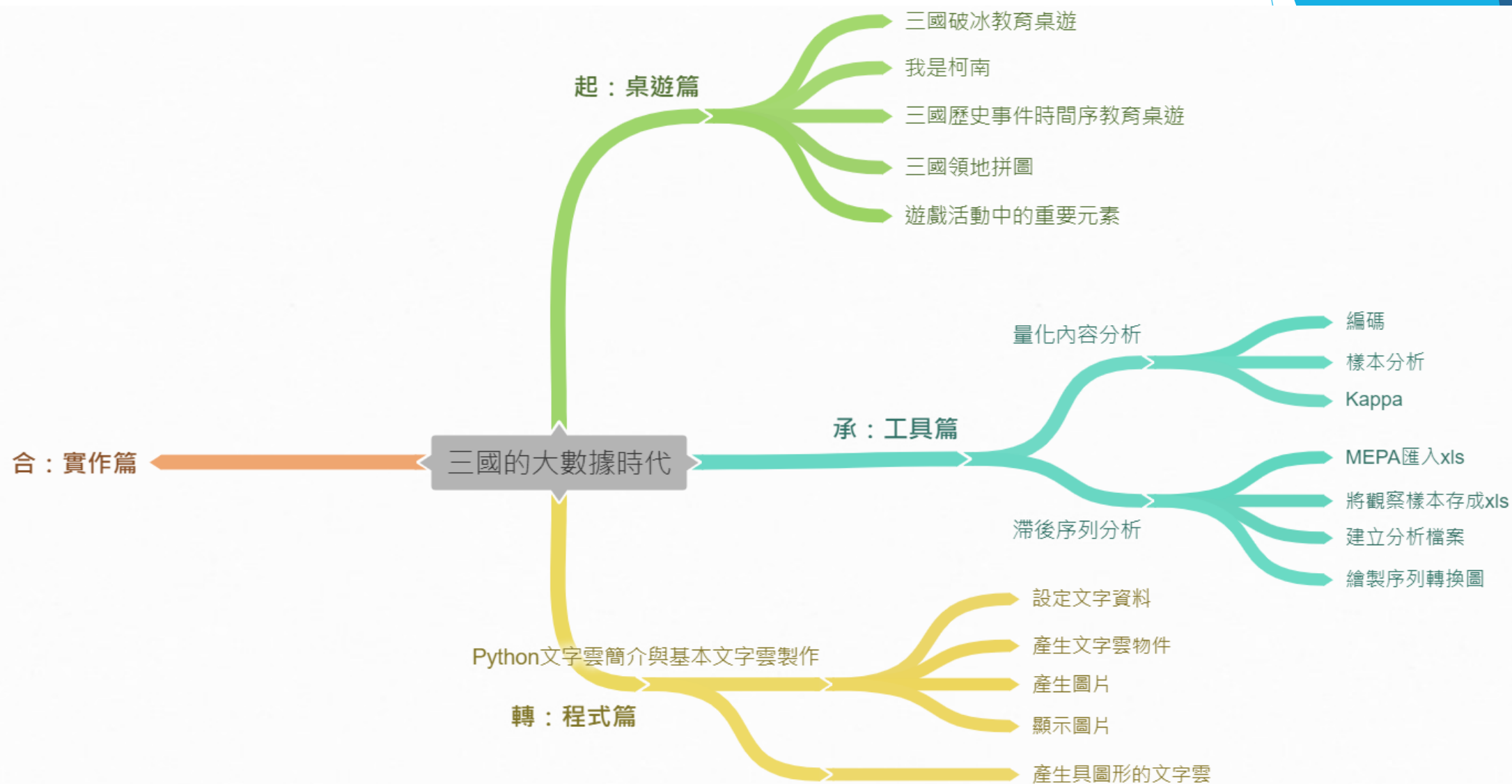


</PLUS>

推動大學(程式設計)教學  
PROMOTING UNIVERSITY COLLEGE PROGRAMS

推動大學程式設計教學計畫。分項六：資料分析領域與學習評量推動團隊（宜蘭大學資工系朱志明老師主編）

# 大綱



# 大綱

1. 起：桌遊篇
2. 承：工具篇
3. 轉：程式篇
4. 合：實作篇

# 起：桌遊篇

1. 三國破冰教育桌遊
2. 我是柯南
3. 三國歷史事件時間序教育桌遊
4. 三國領地拼圖
5. 遊戲活動中的重要元素

# 三國破冰教育桌遊



# 三國破冰教育桌遊

1. 自我介紹。
2. **PK**：出牌核對自己與對方的圖卡是否有相同的圖案，先找到任一相同圖案者（無關大小或方向），即請對方在自己的卡片上簽名（輸的幫贏的簽名）。若皆無相同則繼續找下一位玩家**PK**。
3. 交換：與對方互換圖卡後繼續尋找下一位玩家。
4. 若兩張圖卡完全相同，則為千載難逢的知己，請握手後互相為對方簽名。
5. 時間結束後，計算平均簽名數最多者為勝！

# 三國破冰教育桌遊

策略討論：1分鐘

# 三國破冰教育桌遊

1. 贏的策略?
2. 為何要交換卡片?
3. 在此遊戲中你看(學)到了甚麼?

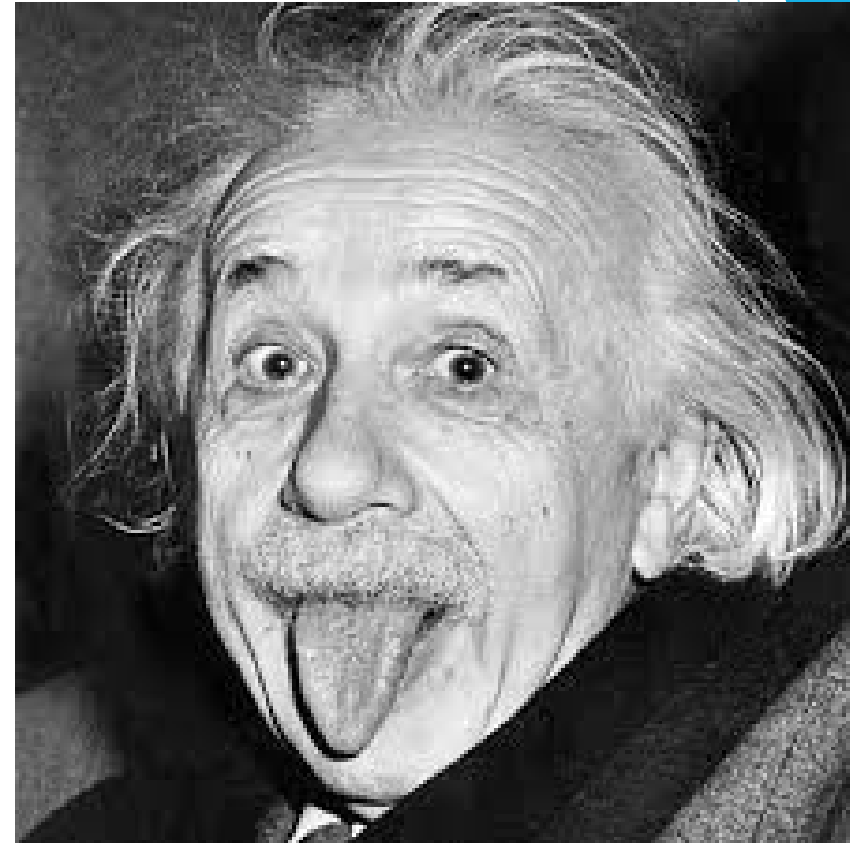


# 三國破冰教育桌遊

1. 認識夥伴
2. 享受互動樂趣
3. 建立圖案記憶
4. 激發情意與注意力

# 三國桌遊篇

- ▶ Play is the highest form of research.
- ▶ 探究的最高形式就是 - 遊戲。



Albert Einstein (亞伯特·愛因斯坦)

# 三國桌遊篇

- ▶ 遊戲是因著人們內在的目的，並且在自為的意義上達到自由的生命活動。



德國哲學家，康德1724-1804

# 三國桌遊篇：我是柯南

Q：相片的組成元素？

A：時間、地點、目的...

# 三國桌遊篇：我是柯南

4 個關鍵字+一張照片





美國、大學、一群人、第二次世界大戰

# 三國桌遊篇：我是柯南

1945年，賓州大學



美國、零件、諾貝爾獎、第二次世界大戰後



# 三國桌遊篇：我是柯南

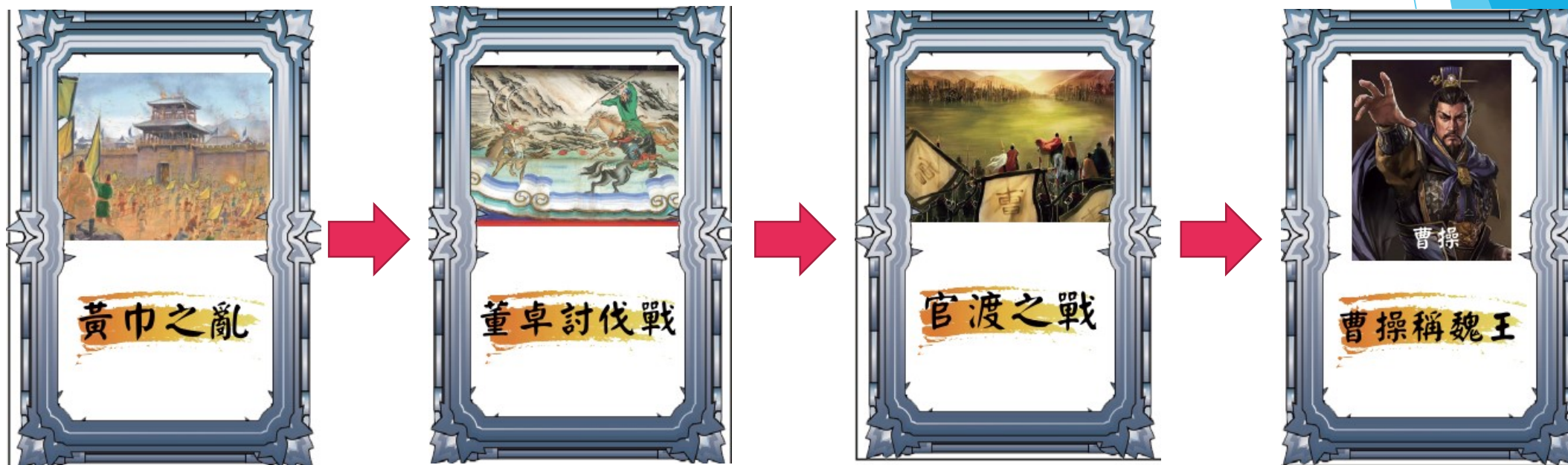
1947年，電晶體

# 三國桌遊篇：我是柯南



# 三國桌遊篇：三國歷史事件時間序教育桌遊

- ▶ 玩法：將30張的三國歷史事件卡牌，按照時間發生順序排列。



# 三國桌遊篇：三國領地拼圖





# 三國桌遊篇：遊戲活動中的重要元素

1. 自由度
2. 控制感
  - $2 + 3 = ?$  VS  $( ? ) = 5$
3. 不確定性
4. 新奇感
5. 成就感



# 承：工具篇

- ▶ 量化內容分析(Quantitative Content Analysis, QCA)
- ▶ 滯後序列分析(Lab Sequential Analysis, LSA)

# 工具篇：QCA及LSA分析流程



# 工具篇：量化內容分析

- ▶ 量化內容分析法(Quantitative Content Analysis, QCA)：將定性(質性)資料轉化為定量(量化)資料的一種分析方法，也稱為資訊分析或文獻分析。
- ▶ 所謂的內容(Content)，指的是資料的型式，如：文字、聲音或影片等具有研究價值的內容，均可做為分析之資料。



# 工具篇：量化內容分析

## ▶ 三國演義之溫酒斬華雄

<https://www.youtube.com/watch?v=20BWVZLD1uU&t=24s>

編碼：A. 驚慌，B. 平靜，C. 高興，D. 鄙視，E. 其他

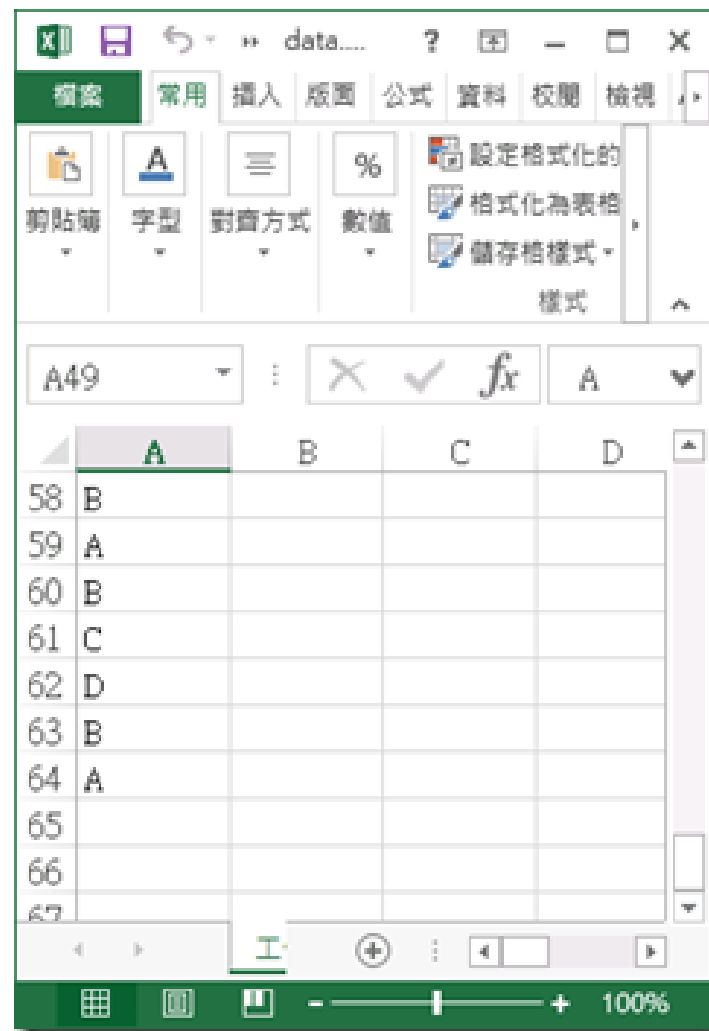
# 工具篇：滯後序列分析

► 使用【多重事件記錄分析】MEPA (Multiple Episode Protocol Analysis)。

1. 將觀察樣本存成xls
2. MEPA匯入xls
3. 建立分析檔案
4. 繪製序列轉換圖

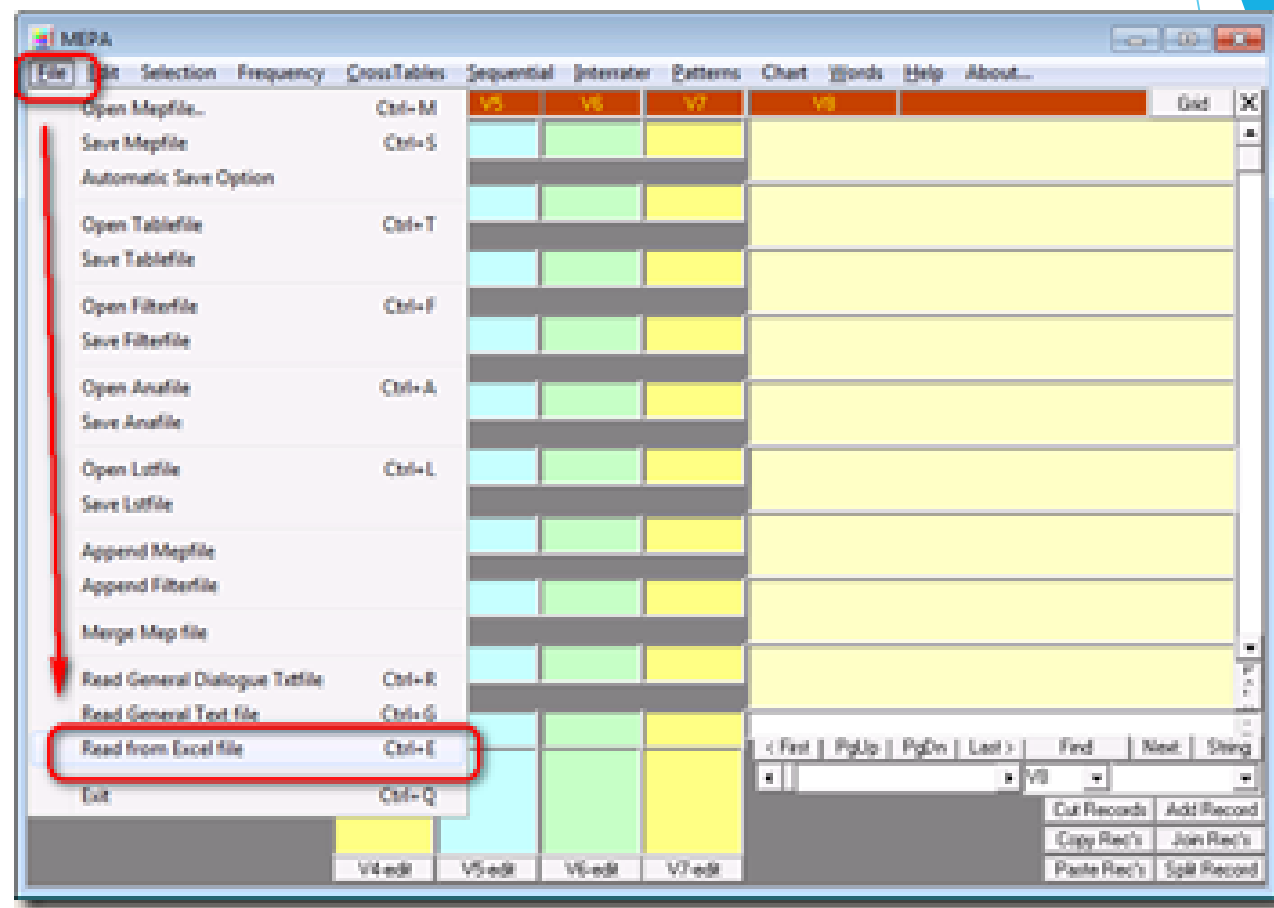
# 工具篇：滯後序列分析

## 1. 將觀察樣本存成 xls



# 工具篇：滯後序列分析

## 2. MEPA匯入xls



# 工具篇：滯後序列分析

## 2. MEPA匯入xls

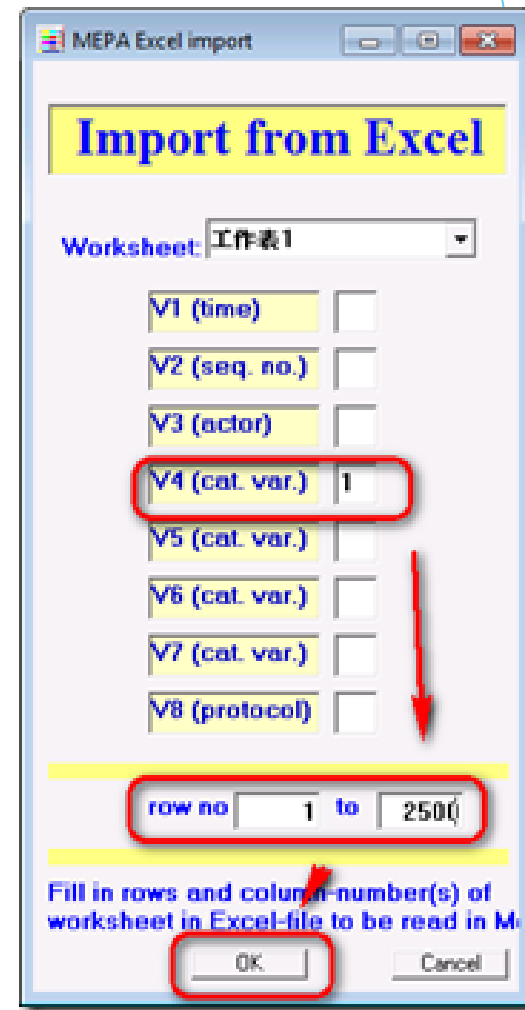
開啟MEPA之後，開啟左上角的File，然後選擇「Read from Excel file」。



# 工具篇：滯後序列分析

## 2. MEPA匯入xls

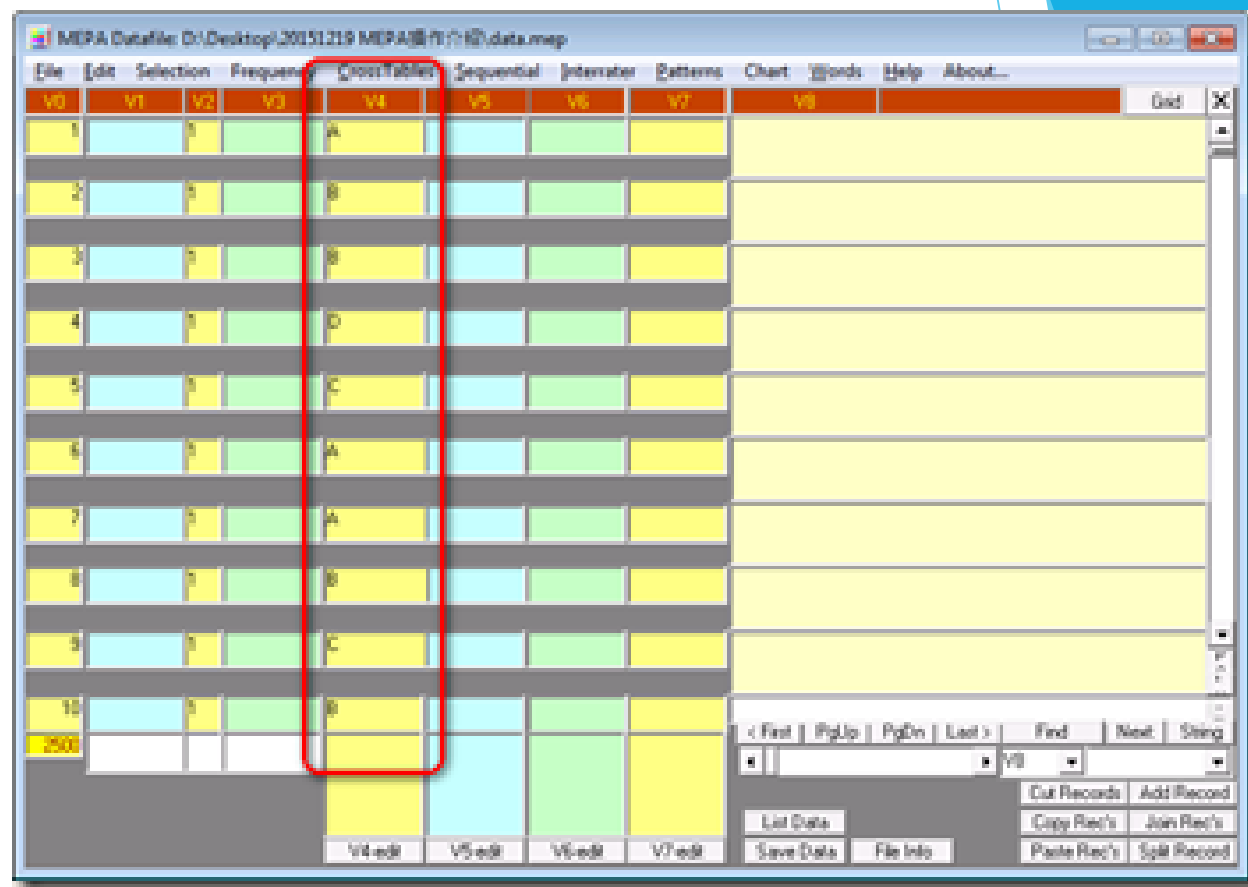
- 選擇要分析的xls檔案，例如這個data.xls。
- 接著會出現Import from Excel對話視窗。我們要設定兩個地方：
- V4 (cat. var.)是編碼樣本，我們輸入1，表示這是Excel的第1欄。
- row no 1 to 2500，這個2500是指有多少筆資料需要分析。如果你的Excel資料的列數算到5000，那就是有5000筆，請輸入5000。



# 工具篇：滯後序列分析

## 2. MEPA匯入xls

- 設定完成之後按下OK。
- 主視窗往下移動，你會發現V4已經匯入了來自Excel的資料。

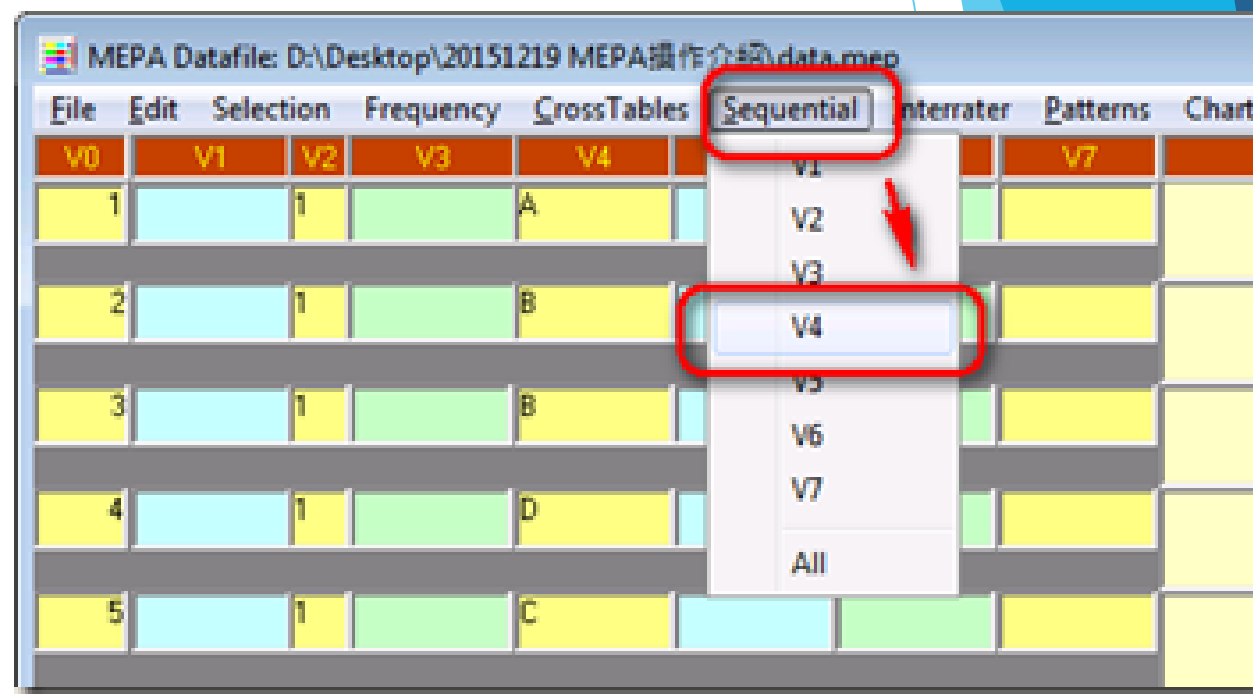


# 工具篇：滯後序列分析

使用【多重事件記錄分析】MEPA  
(Multiple Episode Protocol Analysis)。

## 3. 建立分析檔案

接著按下上面「Sequential」，  
選擇要分析的「V4」。



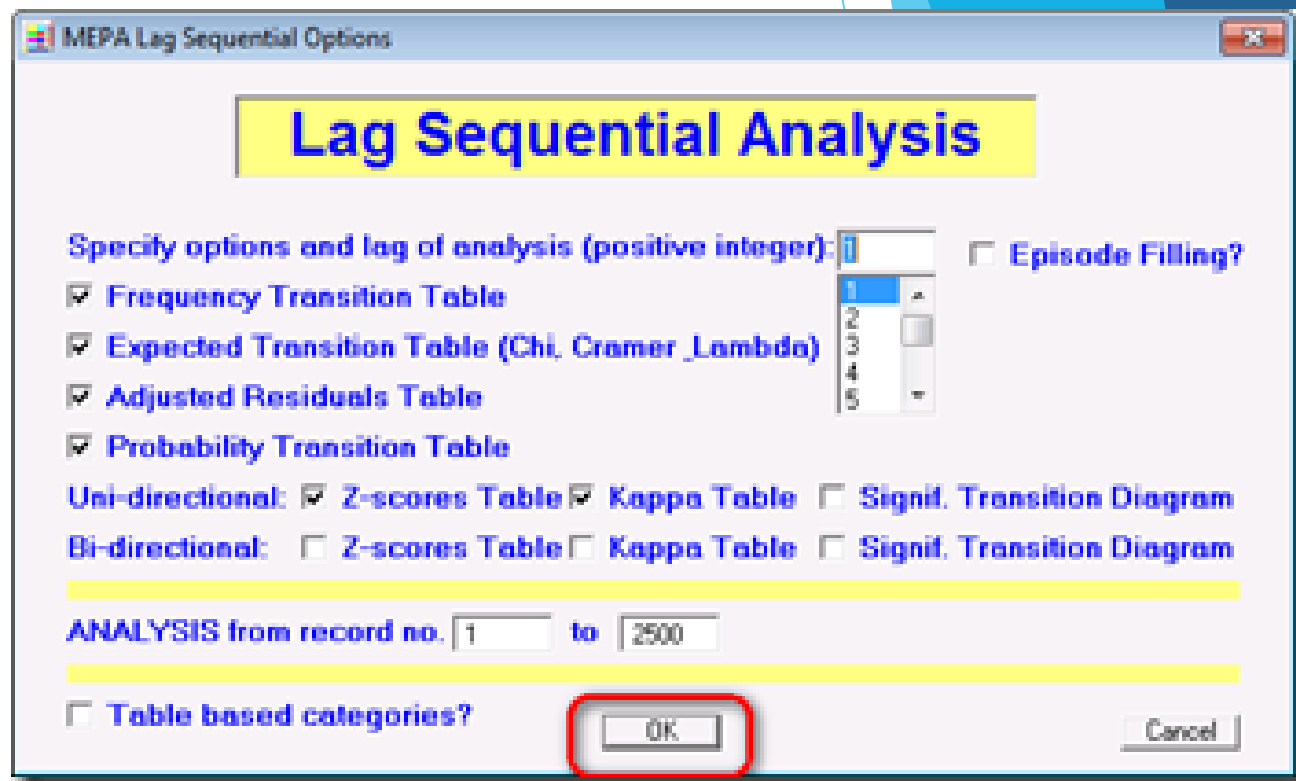


# 工具篇：滯後序列分析

使用【多重事件記錄分析】MEPA  
(Multiple Episode Protocol Analysis)。

## 3. 建立分析檔案

這時候會出現Lag Sequential Analysis滯後序列分析輸入視窗，請按下「OK」。很多參數設定還要再研究，但至少我確定那個lag of analysis是1沒錯。

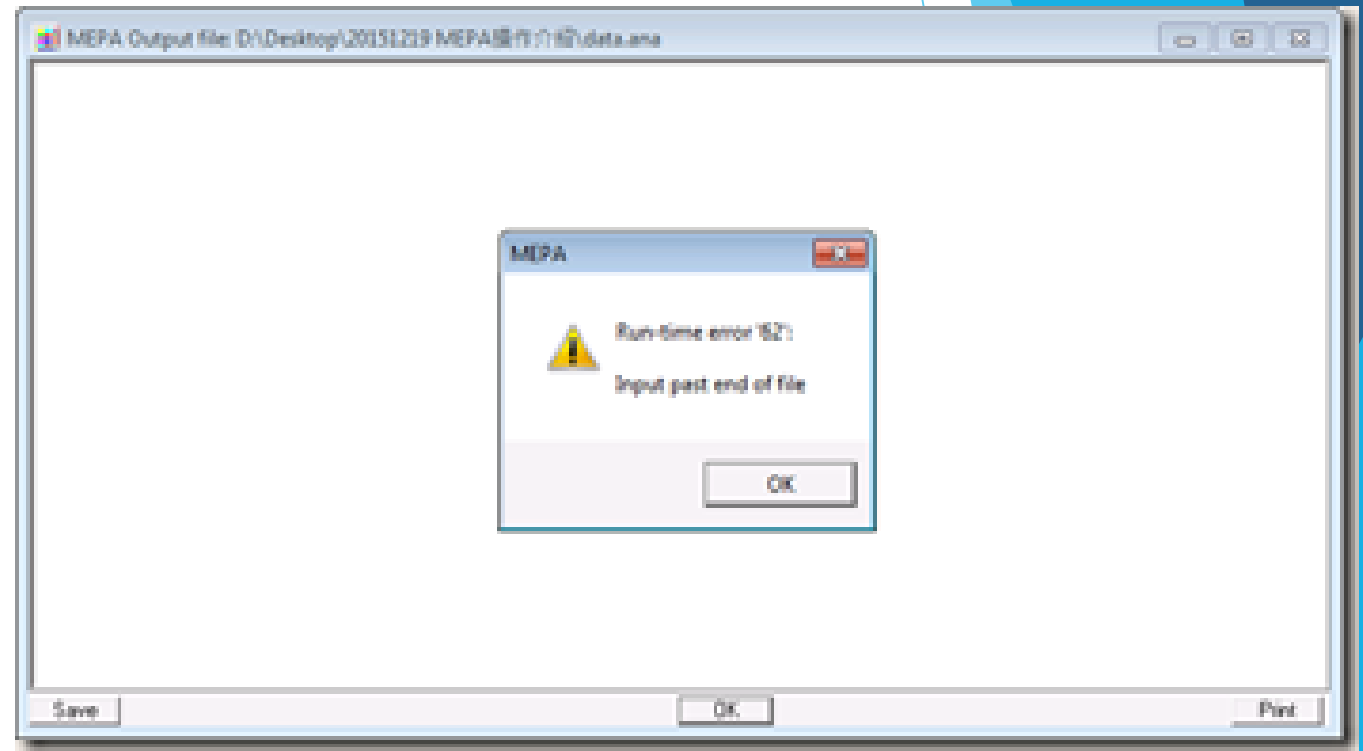


# 工具篇：滯後序列分析

使用【多重事件記錄分析】MEPA  
(Multiple Episode Protocol Analysis)。

## 3. 建立分析檔案

這時候MEPA會顯示錯誤，這是正常現象。

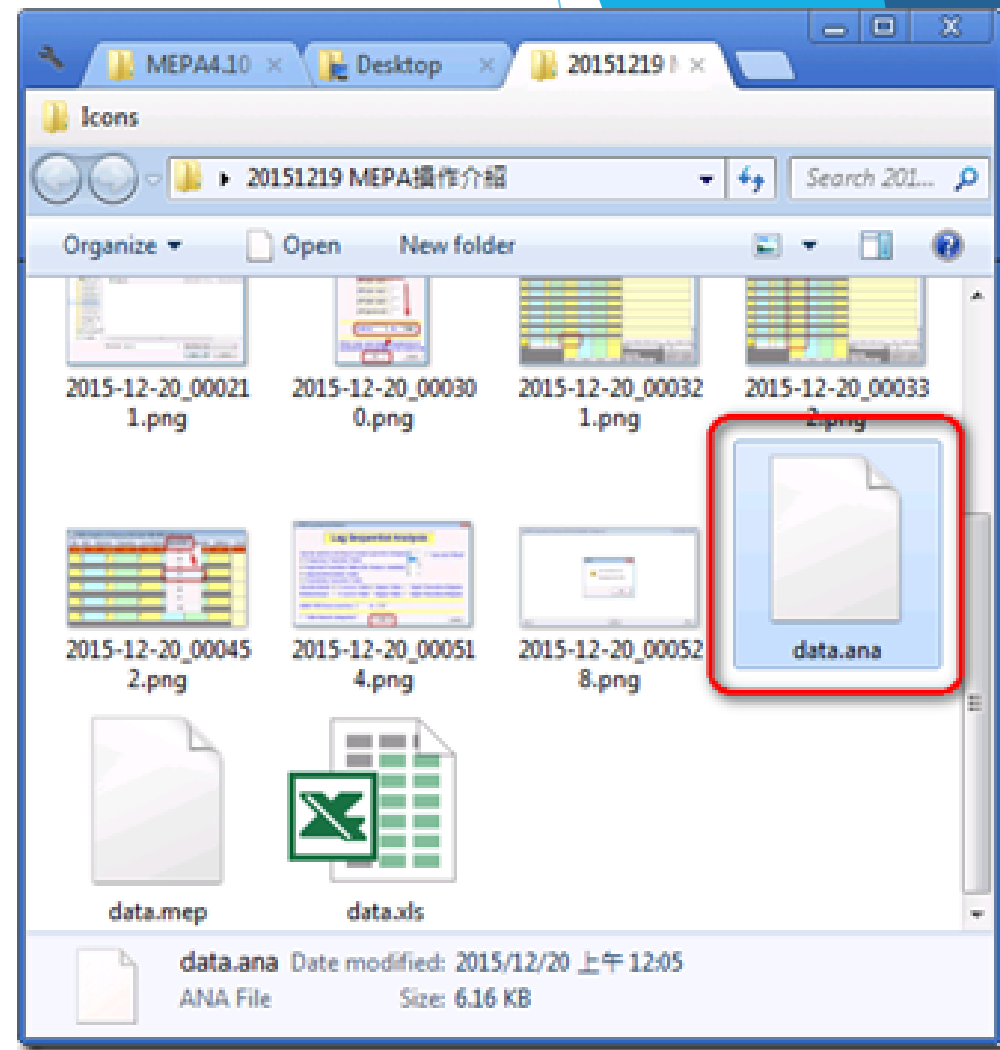


# 工具篇：滯後序列分析

使用【多重事件記錄分析】MEPA (Multiple Episode Protocol Analysis)。

## 3. 建立分析檔案

接著開啟你輸入檔案的資料夾位置，如果是「data.xls」的話，旁邊應該可以找到新增的「data.ana」。請使用記事本軟體，例如Notepad++ Portable來開啟。



# 工具篇：滯後序列分析

使用【多重事件記錄分析】MEPA  
(Multiple Episode Protocol Analysis)。

## 3. 建立分析檔案

畫面往下拉到  
「Adjusted  
Residuals Table」，  
這就是我們的答案。

檔案(F) 編輯(E) 格式(O) 檢視(V) 說明								
Adjusted Residuals Table (Z-scores)				Bakeman & Quera (1995)			Row at Lag: -1	
	A1	A2	B3	B5	C3	C4	D1	D4
A1	9.49	0.89	0.07	-1.43	-1.17	-0.83	-1.96	-1.62
A2	0.10	6.58	-0.45	-0.45	-0.37	-0.26	-0.87	-0.79
B3	-0.99	-0.43	9.31	-0.22	5.66	-0.12	-0.41	-0.37
B5	-0.99	-0.43	-0.22	9.31	-0.18	-0.12	-0.41	-0.37
C3	-0.81	-0.35	-0.18	-0.18	6.96	9.95	-0.34	-0.30
C4	-0.57	-0.25	-0.12	-0.12	-0.10	-0.07	-0.24	-0.21
D1	-1.98	-0.86	-0.43	2.16	-0.35	-0.25	10.01	0.75
D4	-1.15	-0.77	-0.39	-0.39	-0.32	-0.22	0.74	10.80
E1	-0.25	-0.50	-0.25	-0.25	-0.21	-0.15	-0.48	-0.44
E2	-0.57	-0.25	-0.12	-0.12	-0.10	-0.07	-0.24	-0.21
E3	-1.82	-1.19	-0.59	-0.59	-0.48	-0.34	-1.14	-1.03
E4	-0.57	-0.25	-0.12	-0.12	-0.10	-0.07	-0.24	-0.21
E6	-1.67	-1.70	-0.85	-0.85	-0.69	-0.49	-1.62	-1.47
F1	-0.99	-0.43	-0.22	-0.22	-0.18	-0.12	-0.41	-0.37
F2	-1.77	-0.77	-0.39	-0.39	-0.32	-0.22	0.74	-0.67
F3	-0.99	-0.43	-0.22	-0.22	-0.18	-0.12	-0.41	-0.37
F4	-0.81	-0.35	-0.18	-0.18	-0.14	-0.10	-0.34	-0.30
F5	-2.63	-0.10	-0.57	-0.57	-0.47	-0.33	-1.10	-0.99
I	-0.57	-0.25	-0.12	-0.12	-0.10	-0.07	-0.24	-0.21

# 工具篇：滯後序列分析

使用【多重事件記錄分析】MEPA  
(Multiple Episode Protocol Analysis)。

## 4.繪製序列轉換圖

調整殘差表(Z-scores)如下：

	A	B	C	D
A	15.70	3.17	-10.82	-9.77
B	-2.99	-15.15	24.14	2.88
C	-2.99	3.10	-8.20	6.74
D	-9.63	9.78	-6.24	-0.27

# 工具篇：滯後序列分析

使用【多重事件記錄分析】MEPA  
(Multiple Episode Protocol Analysis)。

## 4.繪製序列轉換圖

其中有大於1.96的序列轉換表示有顯著，我以紅字標示，列出結果如下：

A > A: 15.70

A > B: 3.17

B > C: 24.14

B > D: 2.88

C > B: 3.10

C > D: 6.74

D > B: 9.78

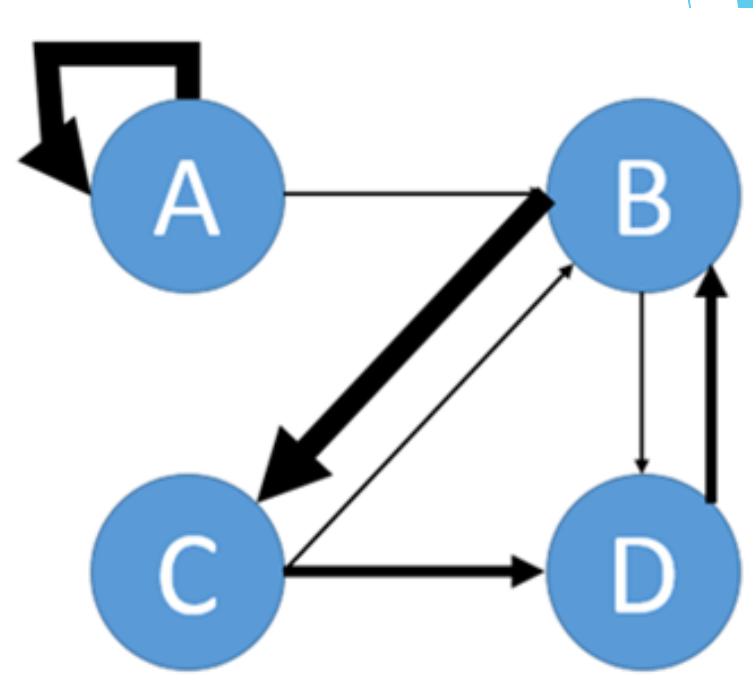
	A	B	C	D
A	15.70	3.17	-10.82	-9.77
B	-2.99	-15.15	24.14	2.88
C	-2.99	3.10	-8.20	6.74
D	-9.63	9.78	-6.24	-0.27

# 工具篇：滯後序列分析

使用【多重事件記錄分析】MEPA  
(Multiple Episode Protocol Analysis)。

## 4.繪製序列轉換圖

根據上述序列，我們將各個行為畫成圓形節點，序列轉移顯著程度則是用粗細不同的線條來表示，越粗表示序列轉移越顯著(Z-score越高)，沒有連線表示沒有顯著序列轉移。



# 轉：程式篇

## ► 文字雲

1. 文字雲是一個視覺化文字資料相當有用的工具。
2. 它從一段的文字中找出字的出現頻率，然後依照出現頻率的多寡，調整單字在文字雲中的大小。文字雲的假設是，越重要的字出現頻率會越高，所以應該用越大的字形來顯示。



# 文字雲簡介與基本文字雲製作

- ▶ 相關模組：
- ▶ 需要產生中文 ( 日文 ) 的文字雲，則需還要額外使用 Jieba(PyKakasi) 套件。

```
!pip install pykakasi
```

```
from PIL import Image  
import matplotlib.pyplot as plt  
from matplotlib.font_manager import FontProperties  
from wordcloud import WordCloud, STOPWORDS  
import cv2, requests, jieba, pykakasi
```

# 文字雲簡介與基本文字雲製作

```
BasePath = '/content/drive/MyDrive/Python/'
imgName1 = BasePath + 'imagesDB/heart.png'
imgName2 = BasePath + 'imagesDB/heartRV.png'
image1    = BasePath + 'faceBase/Lenna_01.png'
txtName1  = BasePath + 'supData/Walden.txt'   # 湖濱散記
txtName2  = BasePath + 'supData/三國志.txt'
txtName3  = BasePath + 'supData/方丈記.txt'
resImg1   = BasePath + "results/result1.png"
resImg2   = BasePath + "results/result2.png"
fntName1  = BasePath + 'supData/楷書體.ttf'
fntName2  = BasePath + 'supData/隸書體.ttf'
fntName3  = BasePath + 'supData/日文體.ttc'
font1 = FontProperties(fname = fntName2)
```

# 文字雲簡介與基本文字雲製作

- ▶ 產生文字雲：
- ▶ 文字雲的產生問題可以拆分成如下四個步驟：
  1. 設定文字資料
  2. 產生文字雲物件
  3. 產生圖片
  4. 顯示圖片

# 文字雲簡介與基本文字雲製作

# 1. 設定文字資料

```
wds = """It becomes necessary for one people to dissolve the \
political bands which have connected them with another, and to \
assume among the powers of the earth, the separate and equal \
station to which the laws of nature and of nature's God entitle \
them, a decent respect to the opinions of mankind requires that \
they should declare the causes which impel them to the separation."""
```

# 2. 產生文字雲物件

```
wc = WordCloud(background_color="pink", max_words=800,
               stopwords=set(STOPWORDS), max_font_size=40, random_state=42)
wc.generate(wds)    # 3. 產生圖片
plt.axis(False)    # 4. 顯示圖片
_=plt.imshow(wc)
```

# 文字雲簡介與基本文字雲製作

# 1. 設定文字資料

```
wds = """It becomes necessary for one people to dissolve the \
political bands which have connected them with another, and to \
assume among the powers of the earth, the separate and equal \
station to which the laws of nature and of nature's God entitle \
them, a decent respect to the opinions of mankind requires that \
they should declare the causes which impel them to the separation."""
```

# 2. 產生文字雲物件

```
wc = WordCloud(background_color="pink", max_words=800,
               stopwords=set(STOPWORDS), max_font_size=40, random_state=42)
```

wc.generate(wds) # 3. 產生圖片

plt.axis(False) # 4. 顯示圖片

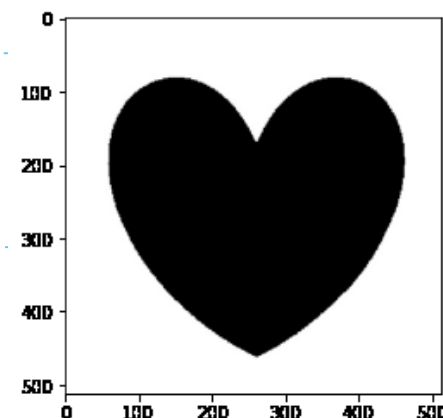
\_ = plt.imshow(wc)



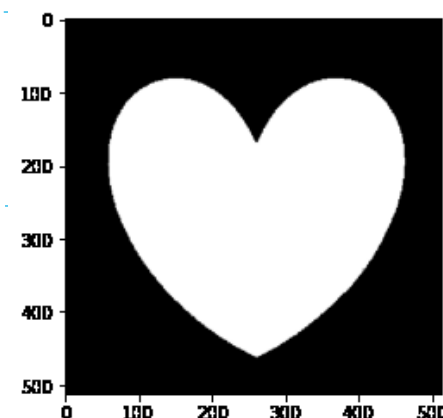
# 特定形狀的文字雲

## ▶ 產生具圖形的文字雲

```
img = cv2.imread(imgName1)
RGB_img = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
_ = plt.imshow(RGB_img)
```



```
img = cv2.imread(imgName2)
RGB_img = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
_ = plt.imshow(RGB_img)
```



# 特定形狀的文字雲

```
# 讀取文字資料
wds=open(txtName1, "r").read()
img = cv2.imread(imgName1)
#####
wc = WordCloud(background_color="white", max_words=800,
               stopwords=set(STOPWORDS), max_font_size=60, mask=img)
wc.generate(wds)
plt.axis(False)
_=plt.imshow(wc)
```



# 合：實作篇

- ▶ 關羽溫酒斬華雄

(<https://www.youtube.com/watch?v=20BWVZLD1uU&t=24s>)

- ▶ 編碼：A.生氣、B.緊張、C.高興、D.懷疑、E.驚訝、  
F.驕傲、G.沮喪、H.其他