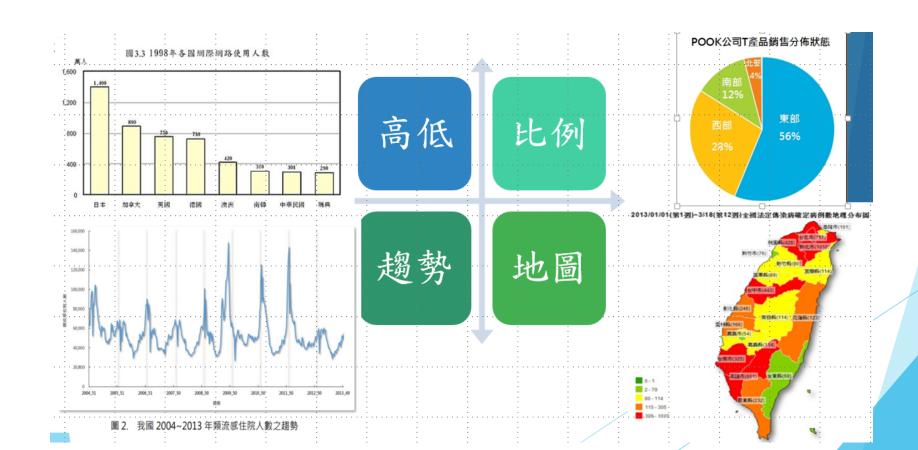
# 統計分析 -SAS EG介紹與繪圖

王雅蕙 東吳大學巨資學院 兼任助理教授 耕莘醫院 醫研中心 助理研究員

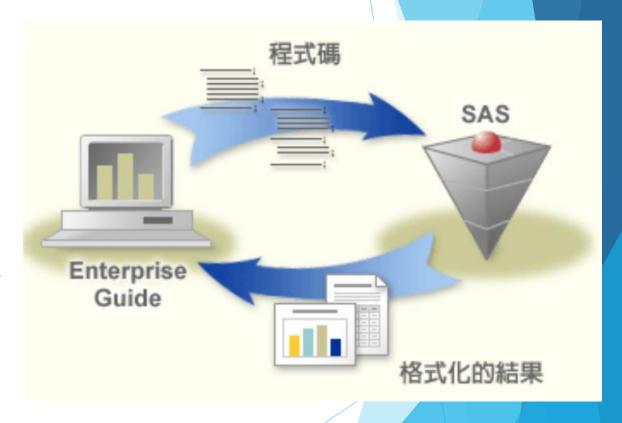
#### Outline

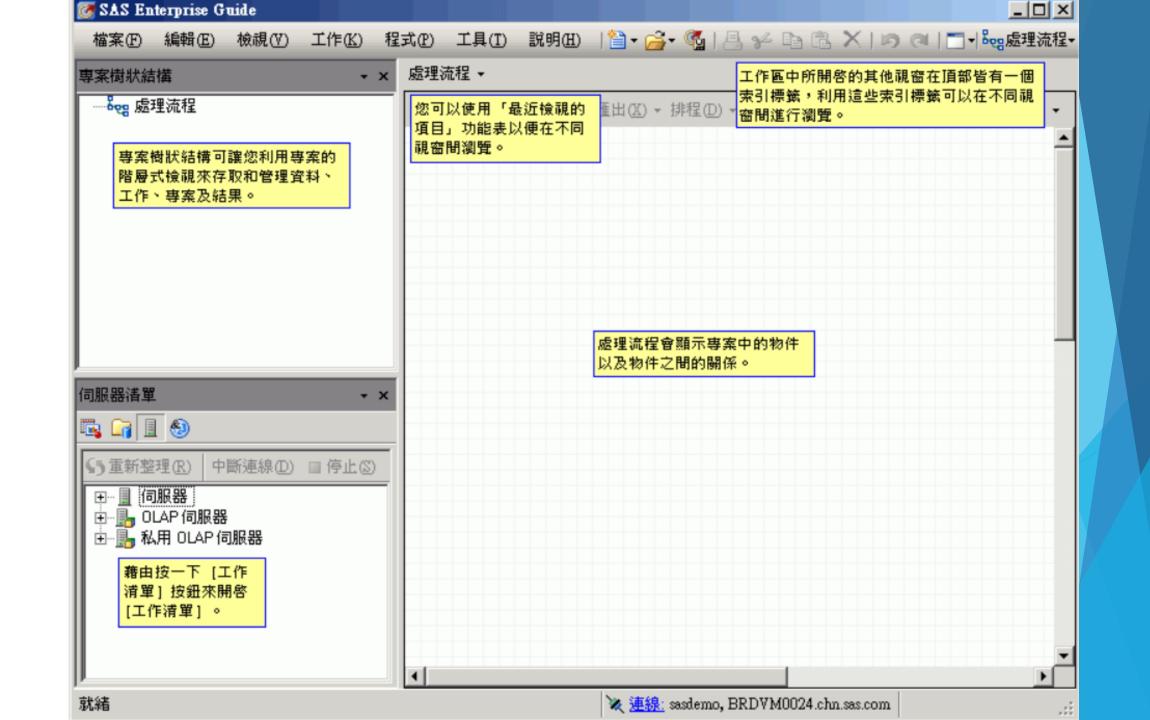
- SAS EG視窗介紹
- 圖表四大類別

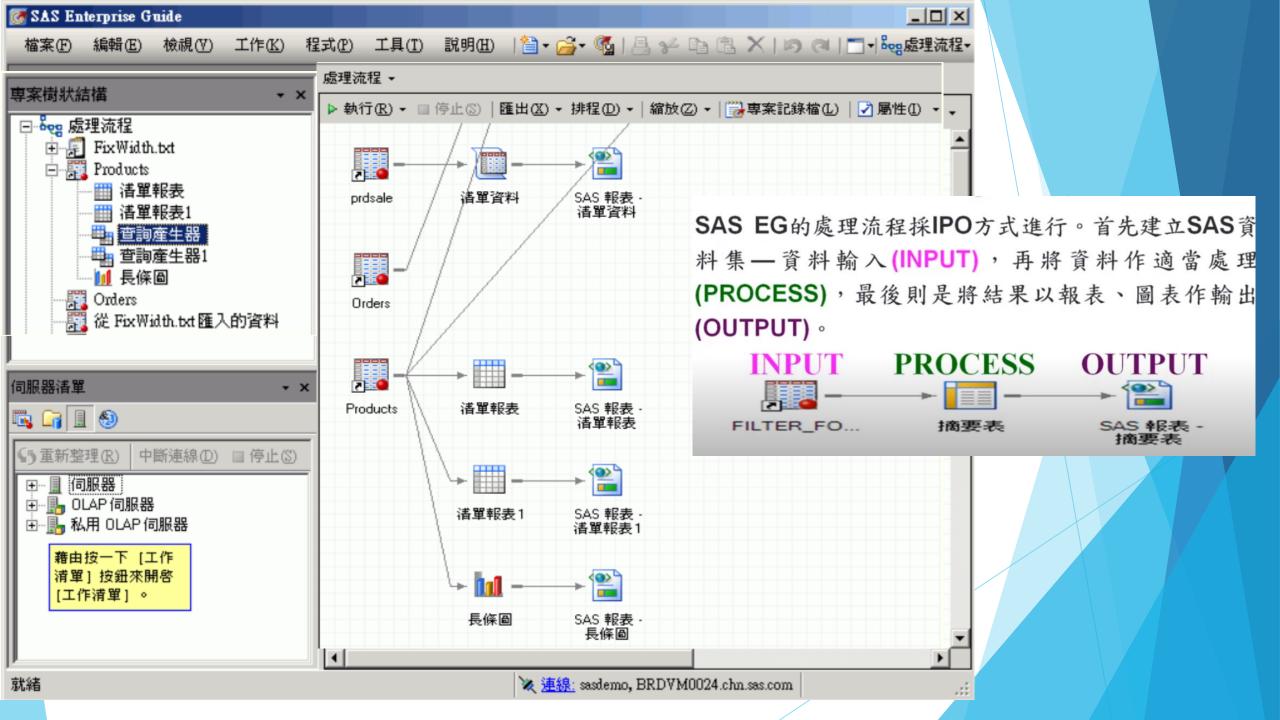


# SAS enterprise guide (SAS EG)

- 透過拖拉點選的方式完成資料處理及 統計分析
- ▶ 同時產生程式碼
- 視覺化的操作介面,資料分析以流程 圖的方式來編輯,降低操作困難,以 方便使用者操作應用程式,與降低統 計分析的門檻
- ▶ 指令碼與自動化
- ▶ 程式碼編輯工具







# SAS EG資料處理

- ▶ 資料的鍵入、匯入...等
- ▶ 資料變項屬性: 類型、標籤、格式...等
- ▶ 資料變項運算:建立新變項、運算、轉置、標準化...等
- ▶ 資料的重新排列: 篩選、排序、排名
- > 資料檔的分割與合併

# SAS EG資料描述

- ▶ 描述統計量:
  - ▶ 平均數、標準差、次數、四分位數...等
- > 統計表格:
  - ▶ 次數分配表、列聯表...等
- ▶ 敘述統計圖:
  - ▶ 直方圖、長條圖、散佈圖、盒形圖、地圖...等
- ▶ 品質管制圖:
  - ▶ p圖、u圖、P-P圖...等

#### SAS EG基礎統計分析

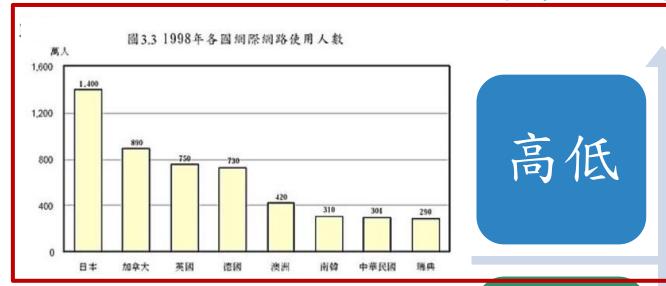
- > 列聯表檢定
  - ▶ 卡方檢定、Fisher's exact test...等
- ▶ T檢定
  - ▶ 獨立樣本t檢定、相依樣本t檢定
- > 變異數分析
  - ▶ 單因子ANOVA、多因子ANOVA...等
- > 相關分析
- > 迴歸分析
  - ▶ 線性迴歸、羅吉斯迴歸...等

#### SAS EG進階統計分析

- > 統計模式
  - ▶ 非線性回歸、廣義線性模式...等
- > 多變量分析
  - ▶ 主成分分析、因素分析、集群分析...等
- > 存活分析
- ▶ 時間序列
- ▶ 資料採礦(text mining)

統計圖表製作

# 四大圖表類型



比例

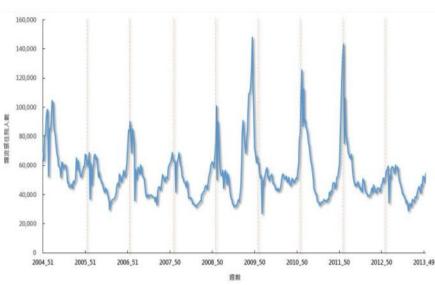
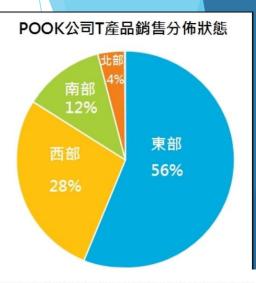


圖 2. 我國 2004~2013 年類流感住院人數之趨勢

地圖

2-79

115 - 305 306 - 1010



2013/01/01(第1週)~3/18(第12週)全國法定傳染病確定病例數地理分布圖



#### 長條圖

- ▶ 適用於類別資料
  - ▶ 性別、BMI分組
- ▶ 也可適用於數值資料
  - > 例如年齡
  - ▶ 但SAS EG會自動將年齡分組 (例如11-20, 21-30, 31-40...)
- ▶「長條圖精靈」&「長條圖」功能

## 「長條圖精靈」

- ▶ 利用salesperson檔案(SAS EG繪圖範例檔.xlsx中的salesperson工作表)
  - ▶ 25位銷售人員列表,一位銷售人員只在一家商店工作

#### Salesperson:

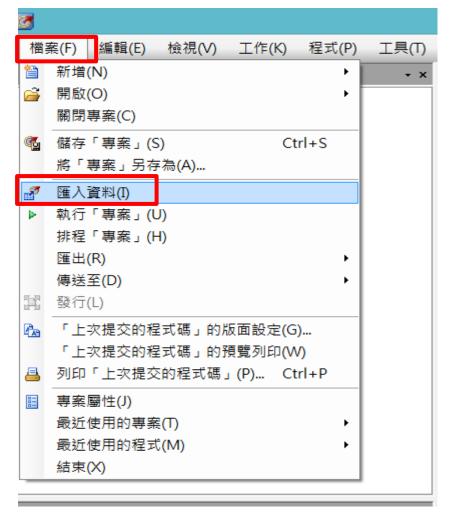
Field Name	Description	Data Type
sales_id	unique Salesperson identifier. ( sales_id is intentionally different than salesperson_id from Sales table / dataset )	Char
name	name of the Salesperson	Char
store_id	store_id where the salesperson works	Numeric
status	employment status of the salesperson:  • A – Active • I – Inactive	Char

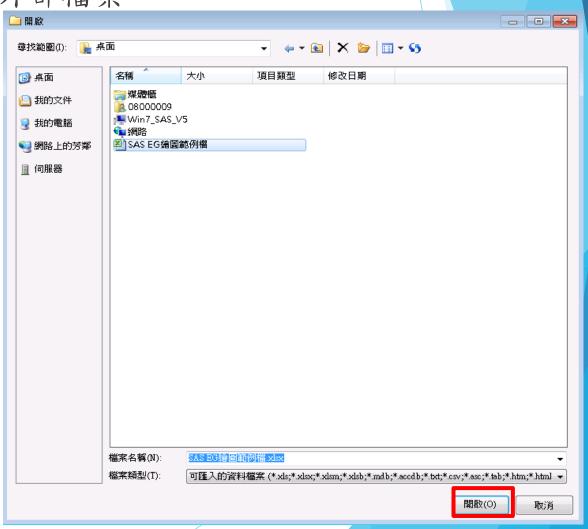
▶ 畫出各分店銷售人員人數分布的長條圖

А	В	C	U
name	sales_id	store_id	status
Jim Goodnight	00001	ST2	Α
Odious Herodias	00002	ST8	Α
John Smith	00003	ST4	Α
MarkBrown	00004	ST3	Α
Susan Jones	00005	ST8	Α
Harry Droogendyk	00006	ST8	Α
Faisal Dosani	00007	ST5	1
Lisa White	00008	ST5	Α
) MaryThompson	00009	ST1	Α
L Don Daniels	00010	ST1	Α
? George Trifunovic	00011	ST7	Α
salesperson	class_birthdate	sales city	y   +

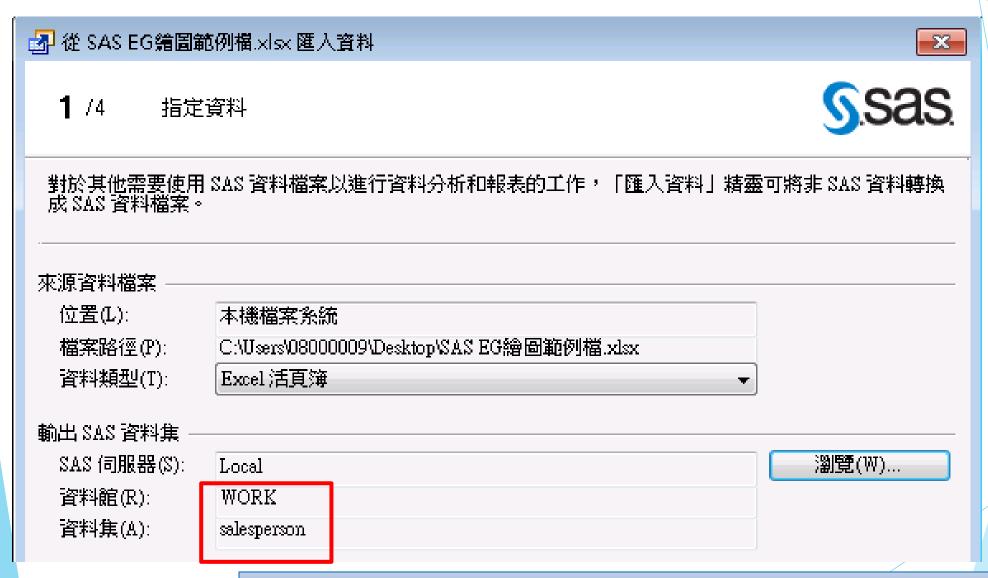
#### 外部資料檔的匯入-Excel檔

▶ 檔案 → 匯入資料 →選取欲匯入的外部檔案





#### 外部資料檔的匯入-Excel檔

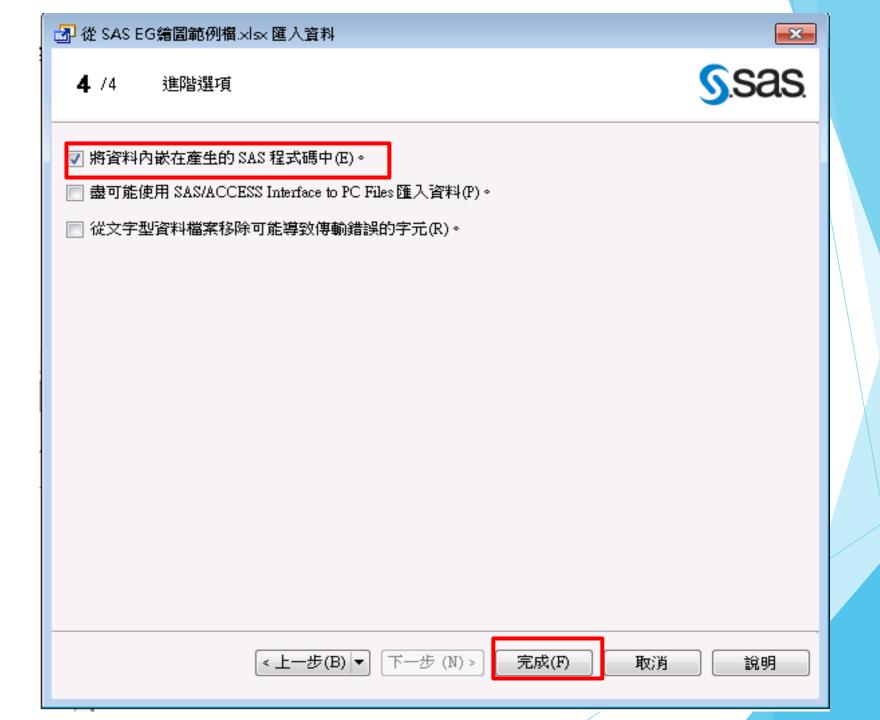


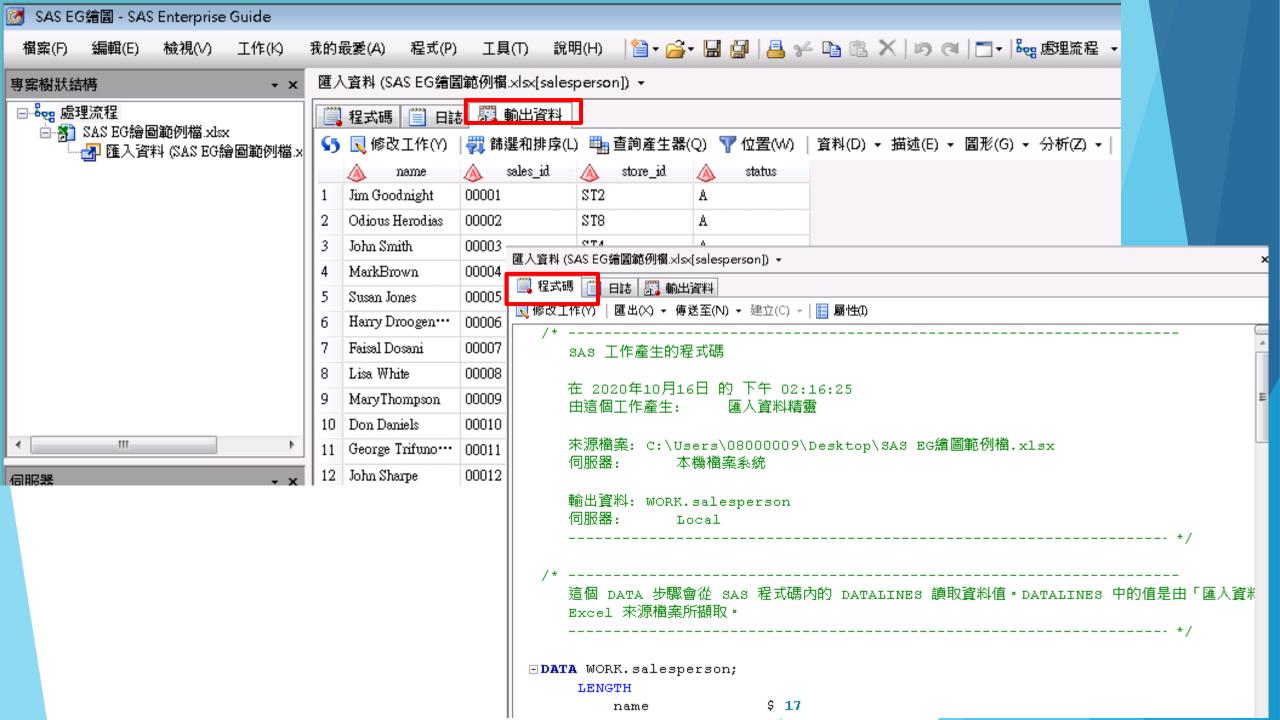
存成SAS work暫存資料夾中的salesperson檔名

選取範圍 ● 使用工作表(W)	河 新風的第一	<b>S.SAS.</b> -列包含欄位名稱(M)
sales city		■以符合 SAS 命名規則(R)。
■ 使用工作表內特定範	<b>圍的儲存格(S)</b>	
左上方儲存格(T): 右下方儲存格(L): ② 必要時展開列範匿		
	重設範圍(A) 图(P)	
使用預先定義的命名範圍		



ales_id的欄份	X					
☑ 包含輸出資料集中的欄位(I)						
名稱(M):	sales_id					
標籤(L):	sales_id					
類型(T):	字串 ▼					
- 來源特性						
原始輸入格式(U):		\$CHAR5.				
輸出特性						
長度(E):		5				
動入格式(P):		\$CHAR5.				
輸出格式(O):		\$CHAR5.				
		確	定 取消			

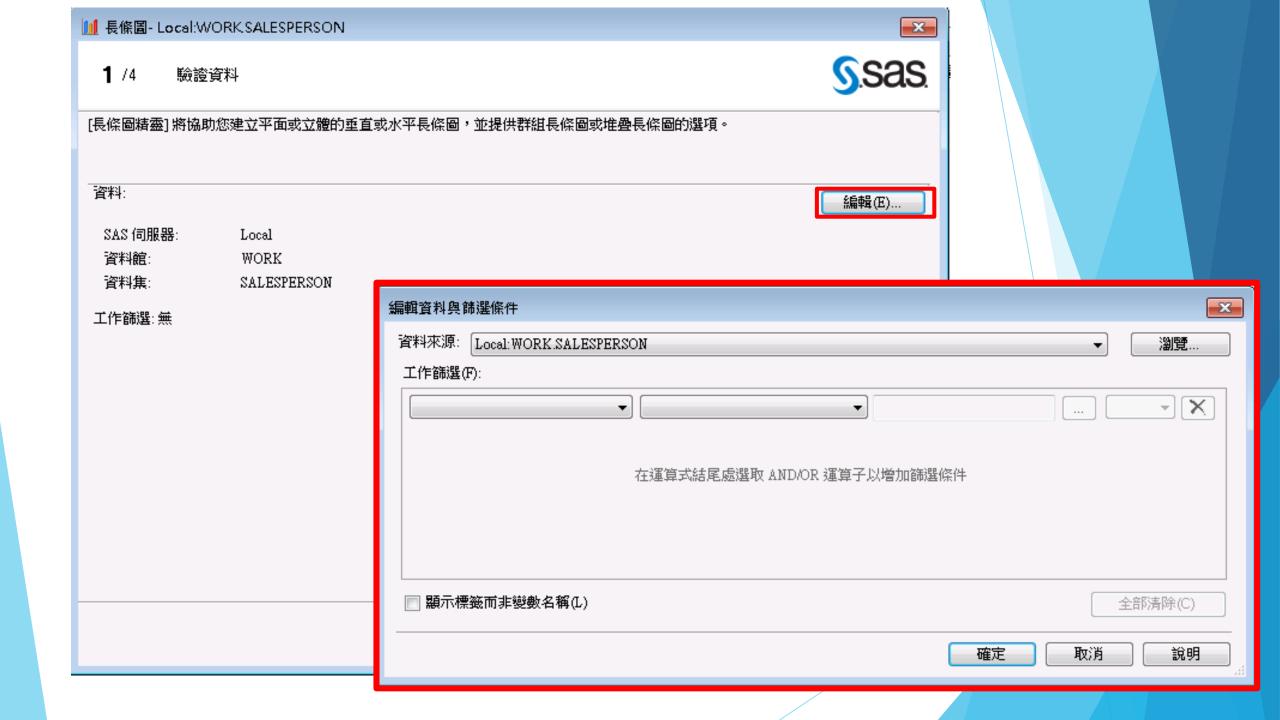


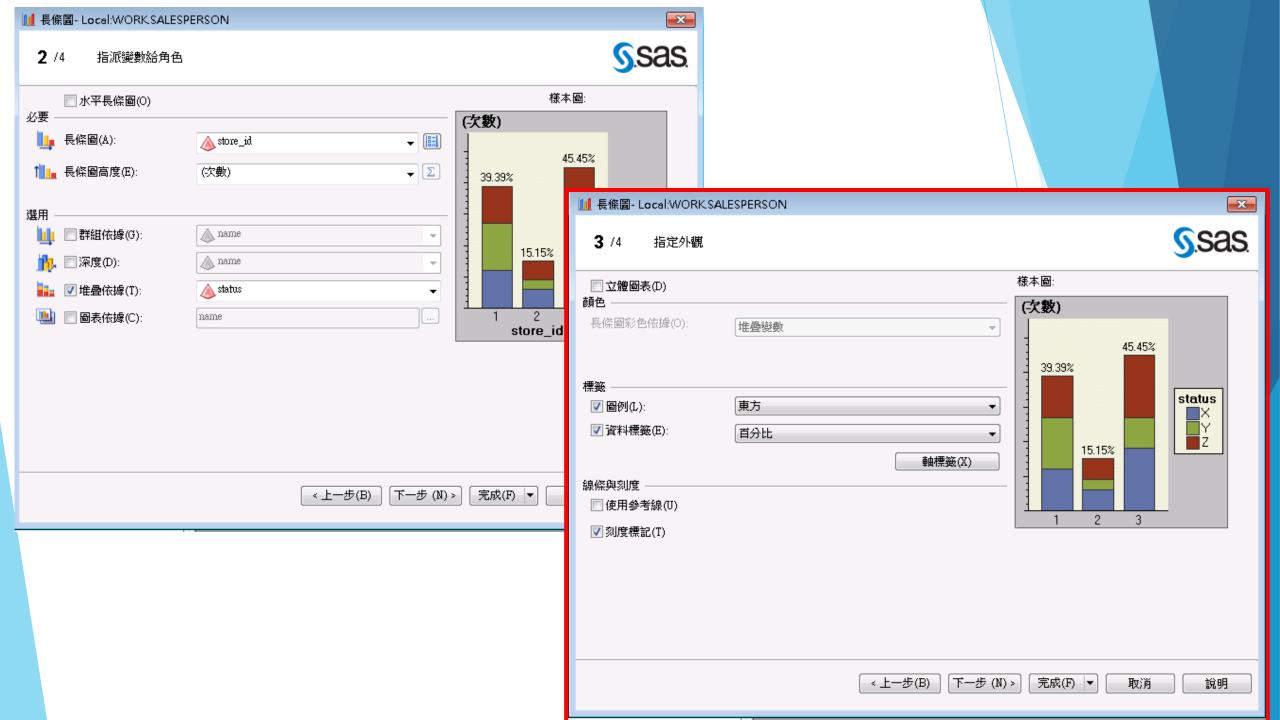


## 「長條圖精靈」

- ▶ 利用salesperson檔案
- ▶ 畫出各分店銷售人員人數分布的長條圖
- ▶ 根據status變項堆疊

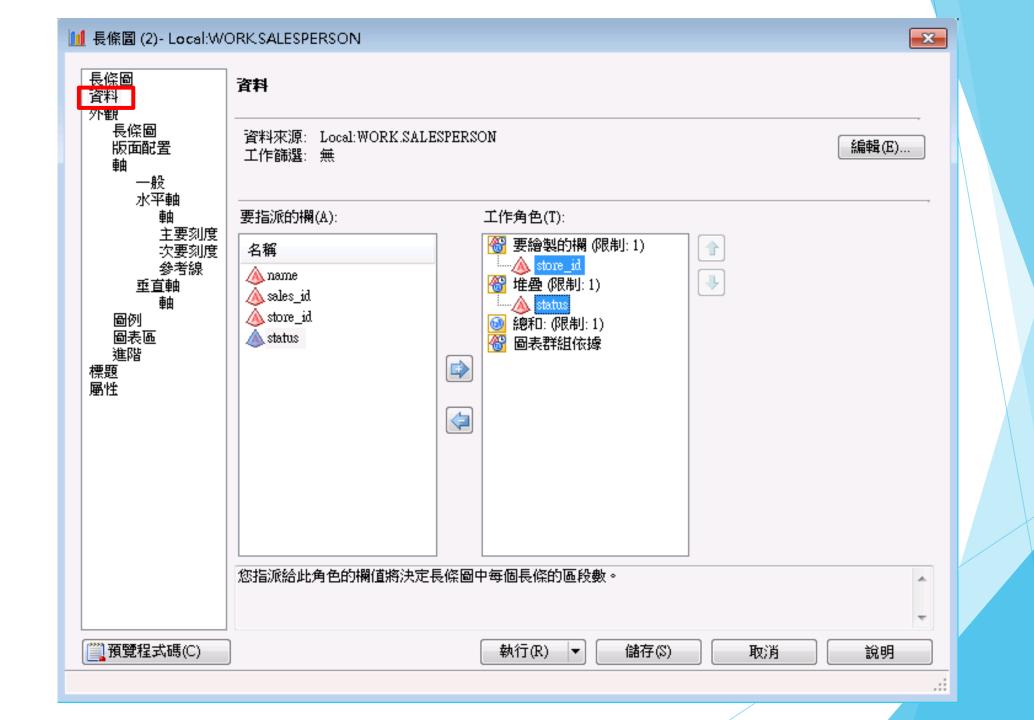


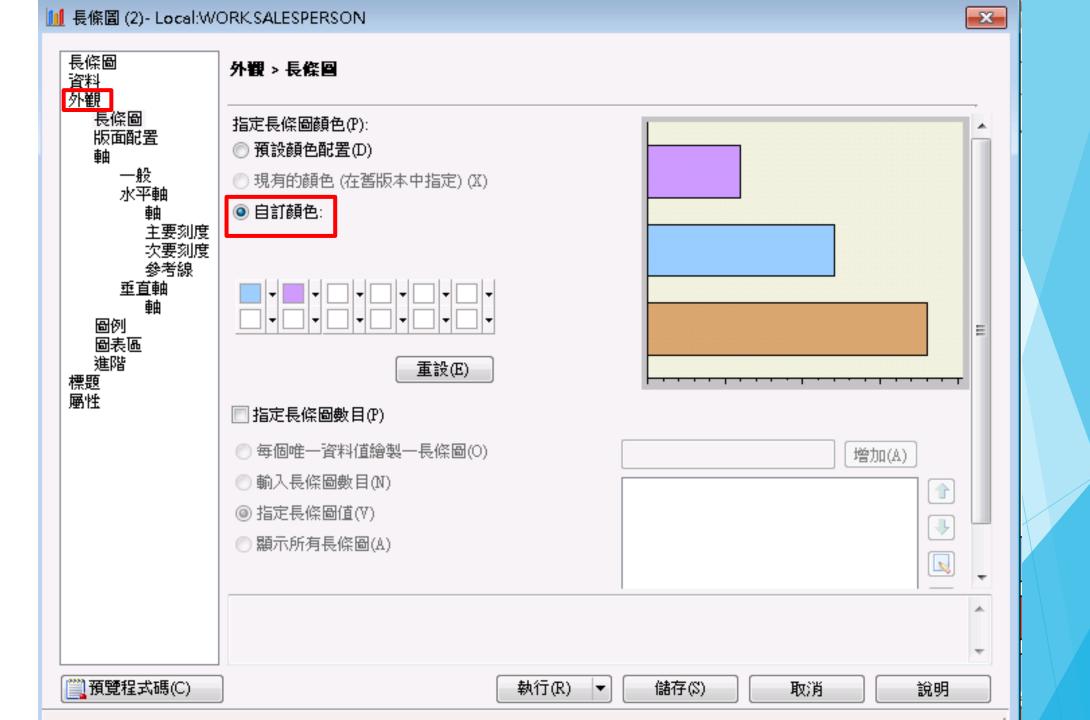


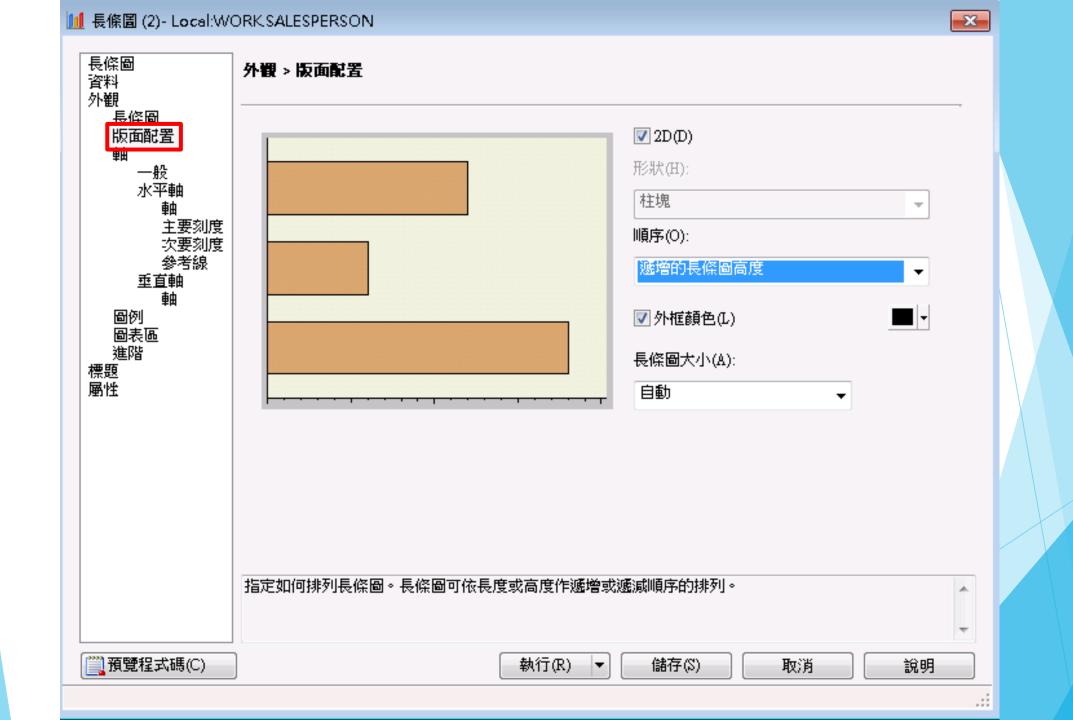




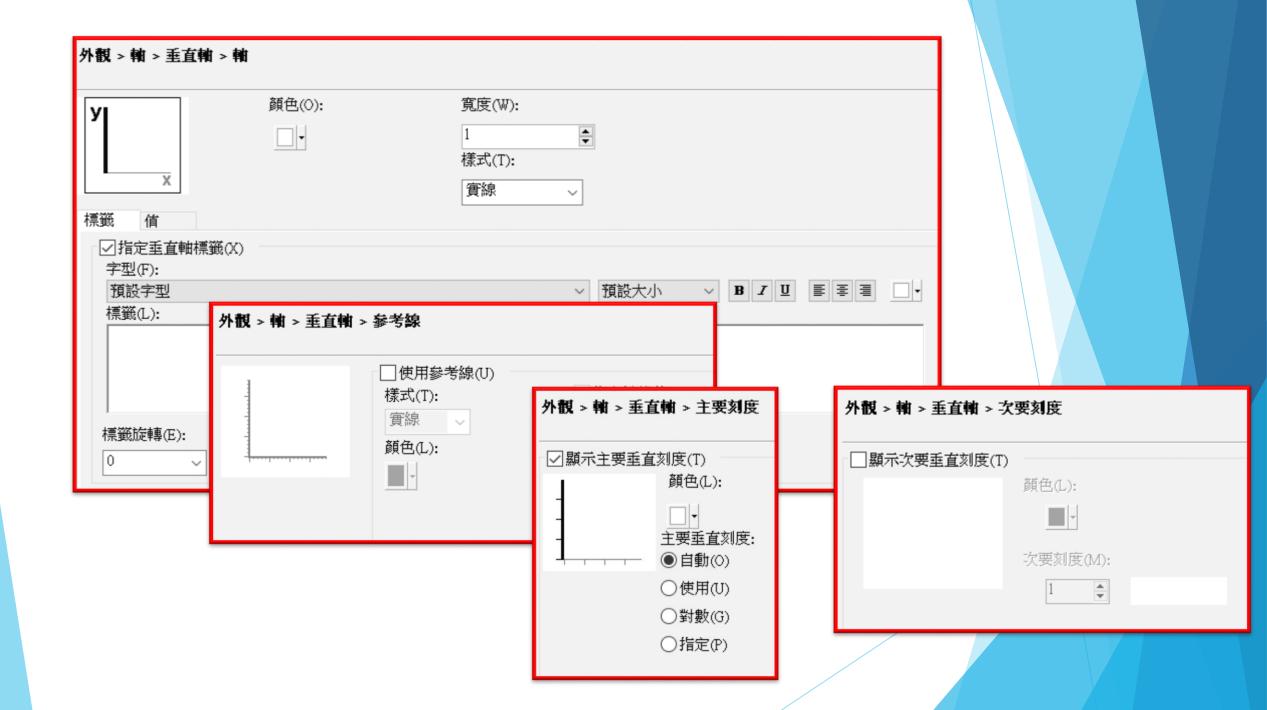




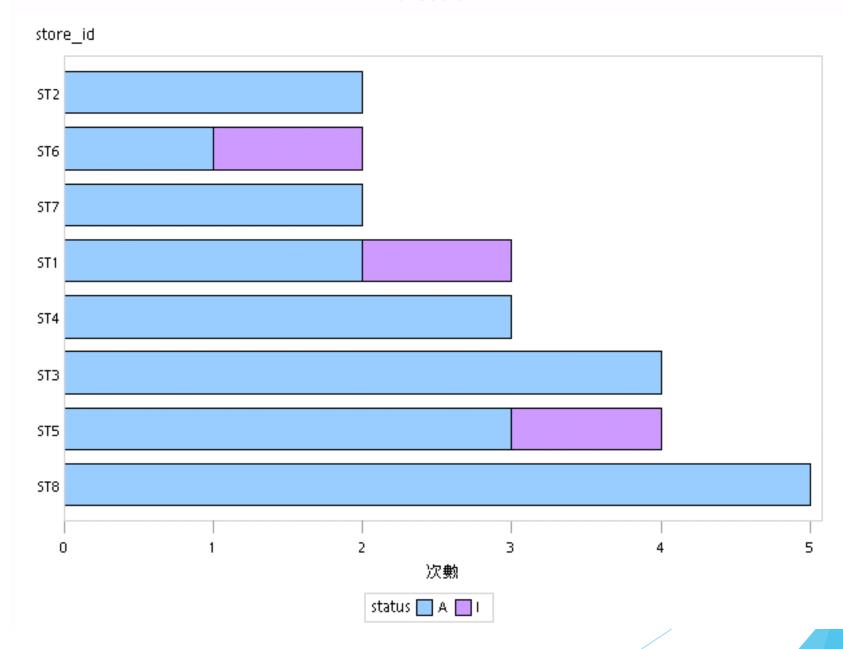




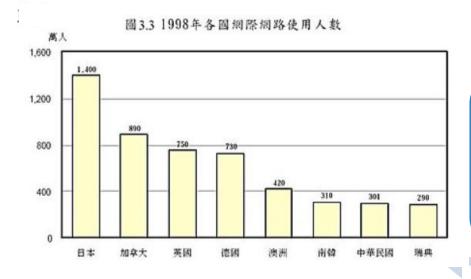




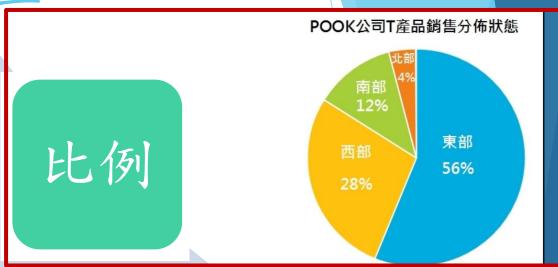




# 四大圖表類型



高低



115 - 305 306 - 1010

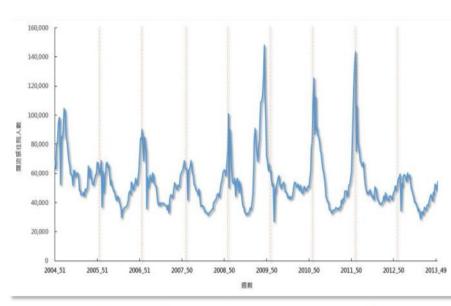
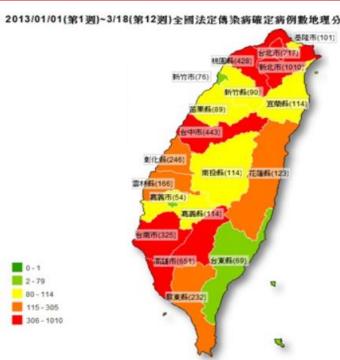


圖 2. 我國 2004~2013 年類流感住院人數之趨勢

地圖

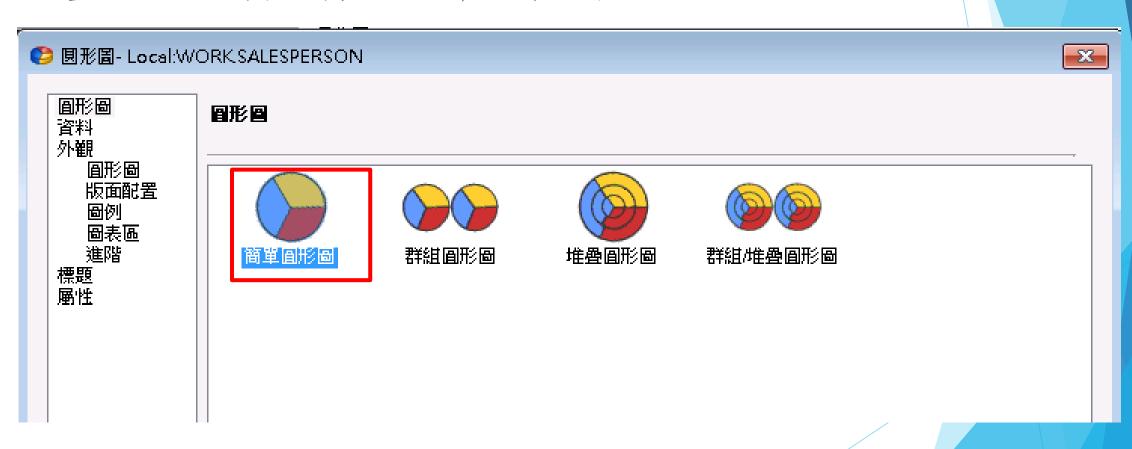


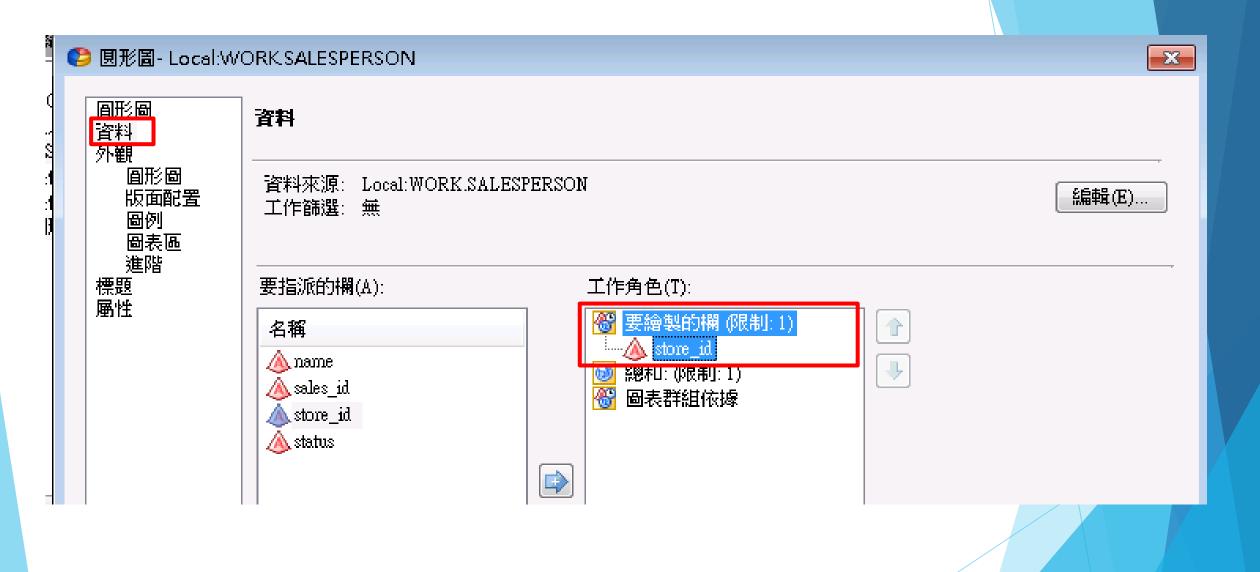
#### 圓形圖(圓餅圖)

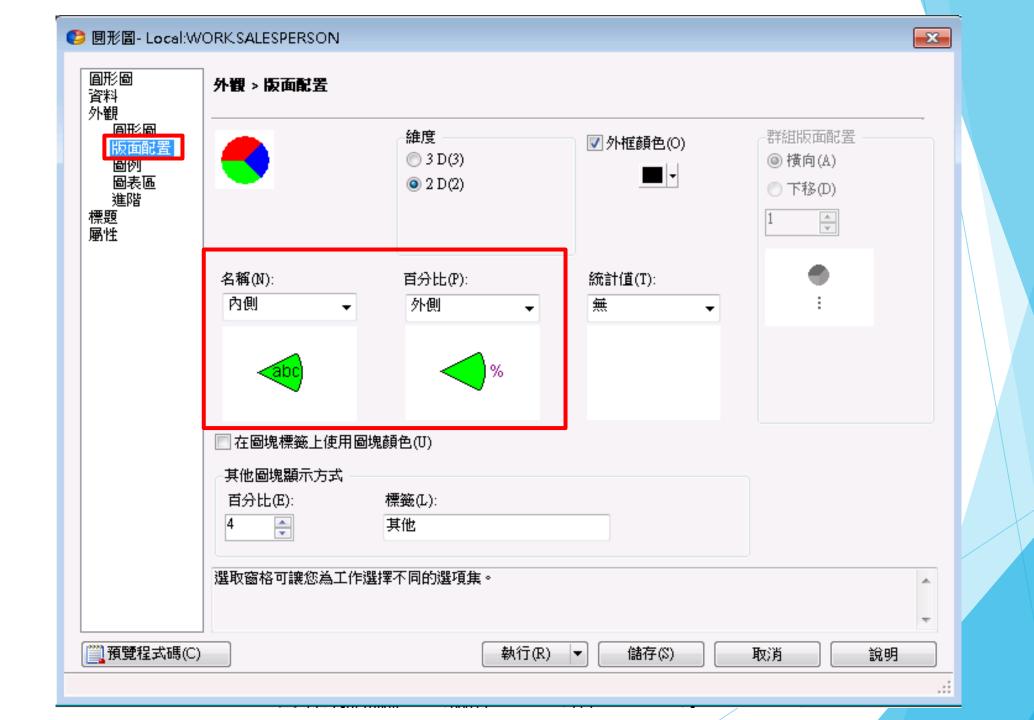
- ▶ 適用於類別資料
- ▶ 也可適用於數值資料,但不建議使用
  - ▶ 會自動分組
- ▶「圓形圖精靈」&「圓形圖」功能

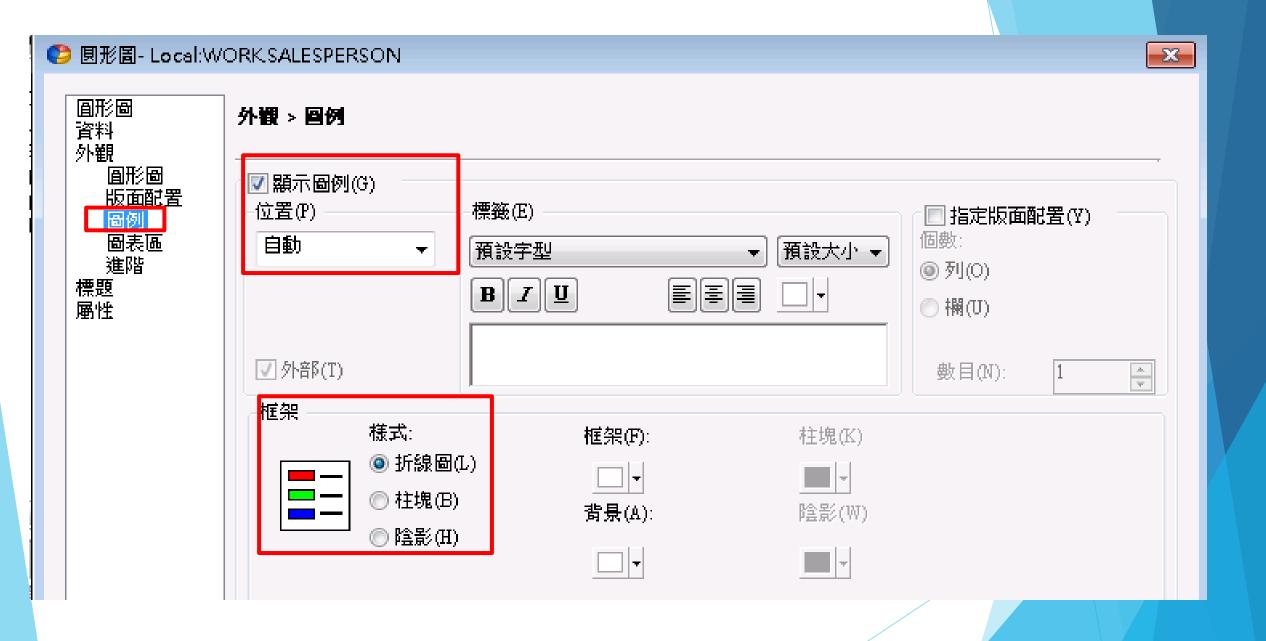
# 「圓形圖」功能

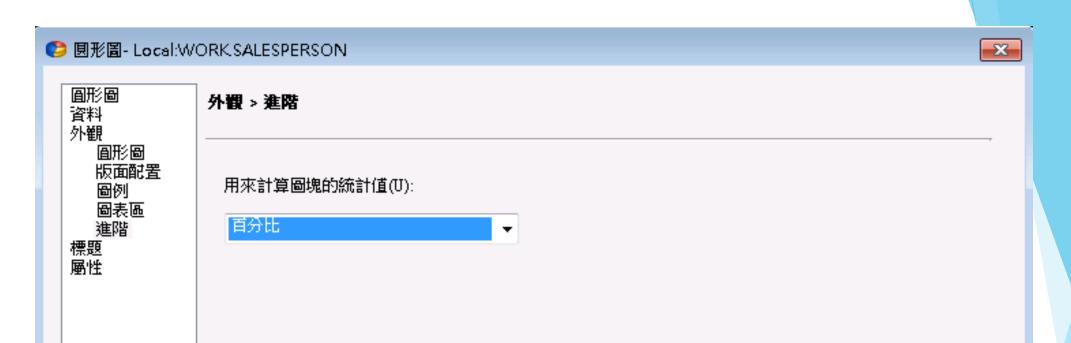
▶ 畫出各分店銷售人員人數分布的圓形圖

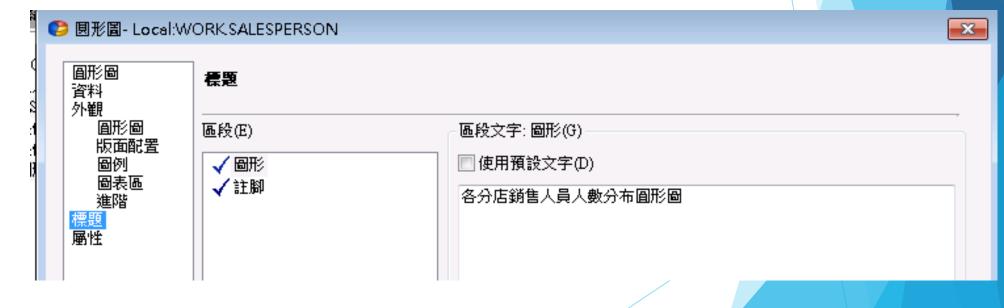




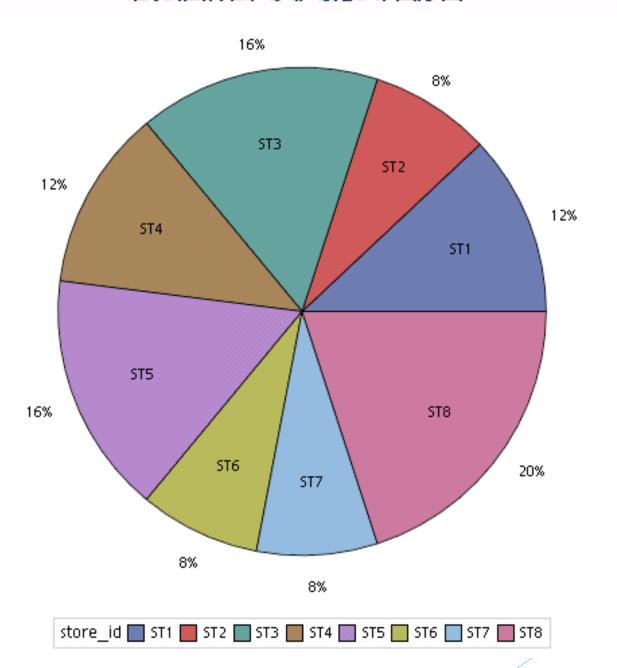








#### 各分店銷售人員人數分布圓形圖



## 直方圖

- ▶ 常用於數值型變項
- > 資料的分布狀態
  - ▶ 例如是否為常態分佈
- ▶「描述」→「分配分析」
- ▶ 請利用class\_birthdate檔案,畫出weight變項的直方圖
  - ▶ 19名國中生基本資料

	Α	В	С	D	Е	F
1	Name	Sex	Age	Height	Weight	Birthdate
2	Alfred	М	14	69	112.5	10/26/2004
3	Alice	F	13	56.5	84	11/16/2005
4	Barbara	F	13	65.3	98	01/15/2005
5	Carol	F	14	62.8	102.5	07/04/2004
6	Henry	M	14	63.5	102.5	12/01/2004
7	James	M	12	57.3	83	06/15/2006

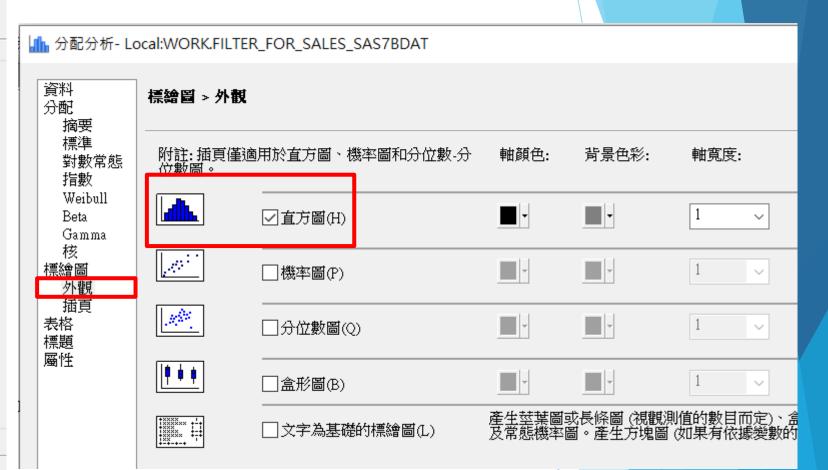
#### 匯入資料 (SAS EG繪圖範例檔:xlsx[class\_birthdate]) ▼

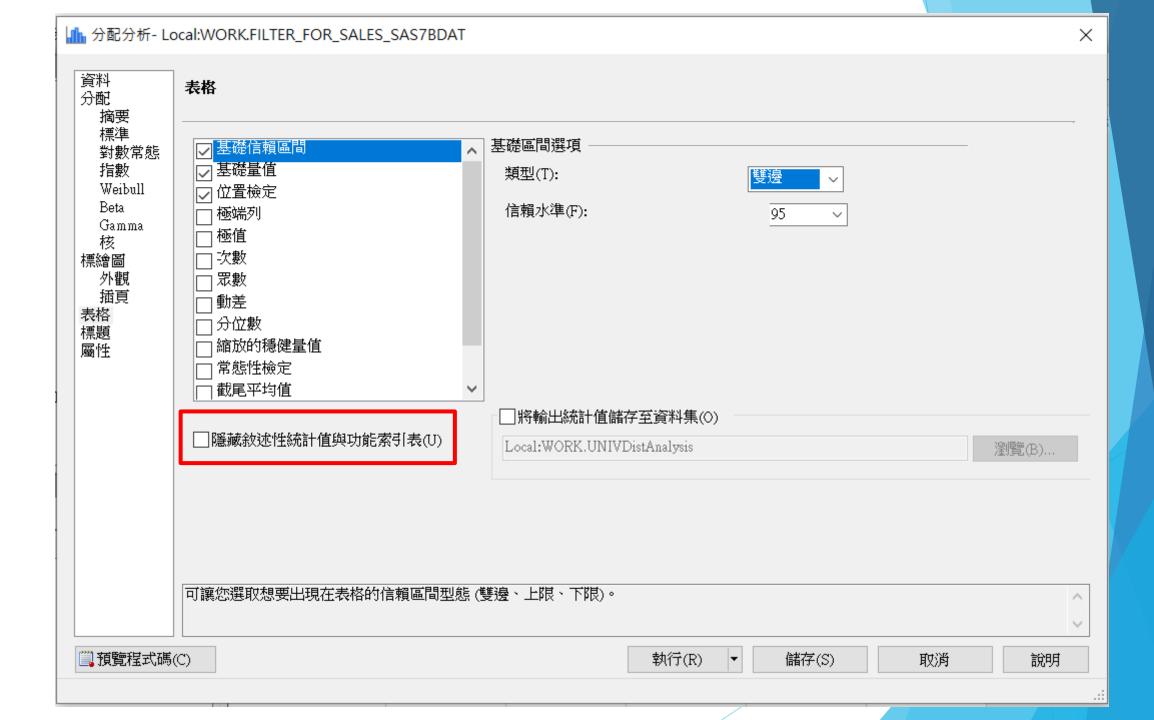
2222	,程式碼 📋 日試	調 輸出資料				
<b>\$</b> 5	凤 修改工作(Y)	🎳 篩選和排序(L)	) 🏪 查詢產生器(	Q) 🍸 位置(W)	資料(D) ▼	描述(E) - 圖形(G) - 分析(Z) -   匯出(X) -
	<u></u> Name	🔈 Sex	Age	Height	₩eig	■ 清單資料(L)
1	Alfred	М	14	69		
2	Alice	F	13	56.5		∑ 摘要統計(S)
3	Barbara	F	13	65.3		
4	Carol	F	14	62.8		
5	Henry	М	14	63.5		│
6	James	M	12	57.3		
7	Jane	F	12	59.8		—
8	Janet	F	15	62.5		分配分析(D)
9	Jeffrey	М	13	62.5		■ 単因子次數(O)
10	John	М	12	59		表格分析(A)
11	Joyce	F	11	51.3		20.2 21172 C.UC



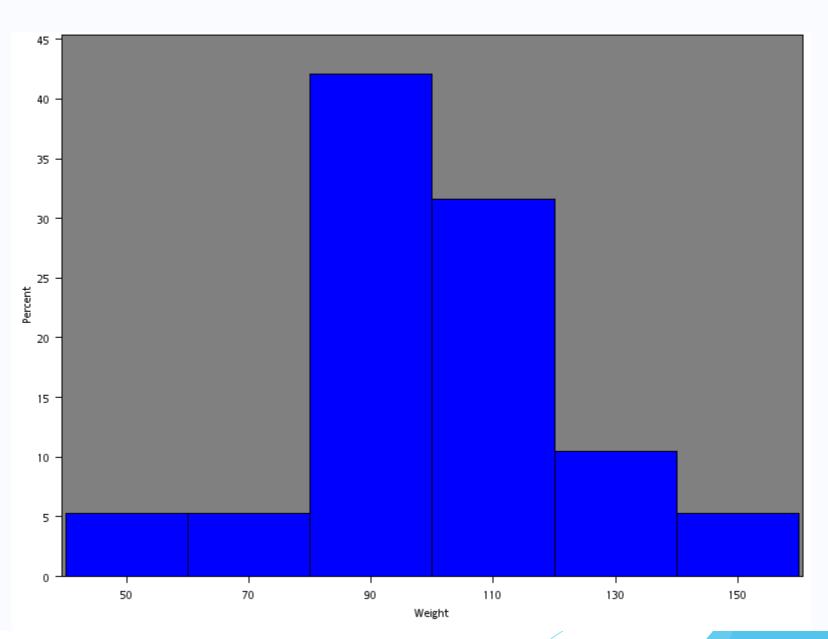
#### ▲ 分配分析- Local:WORK.FILTER\_FOR\_SA

資料 分配 > 摘要 分配 摘要 標準 可用的分配 對數常態 指數 ■常態(N) Weibull Beta ■對數常態(L) Gamma ■指數(E) 標繪圖 外觀 Weibull(W) 插頁 表格 Beta(B) 標題 屬性 Gamma(G) □核(K) 圖形樣式 ● 傳統(T) ODS(D)



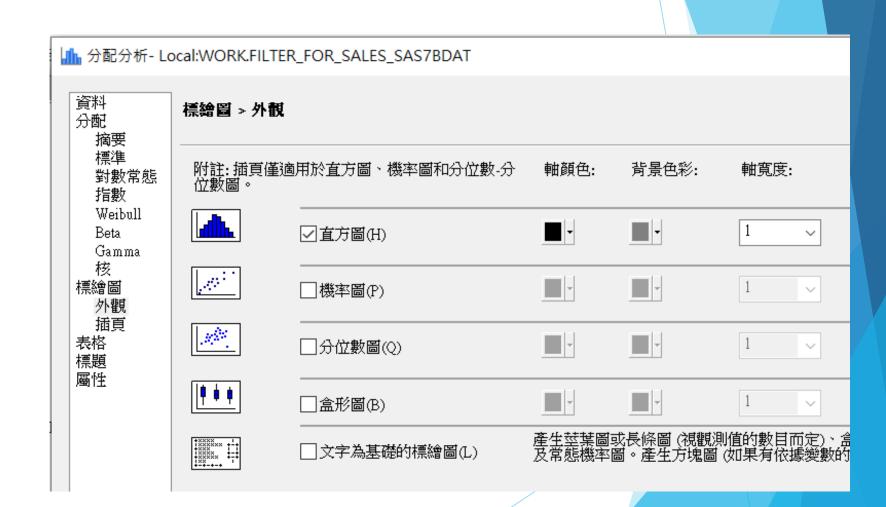


# 下列項目的分配分析: Weight UNIVARIATE程序

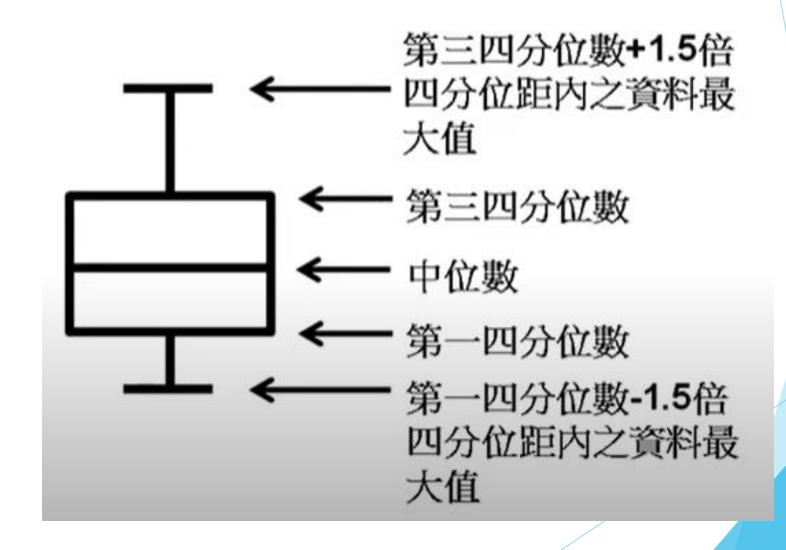


# 「描述」一「分配分析」

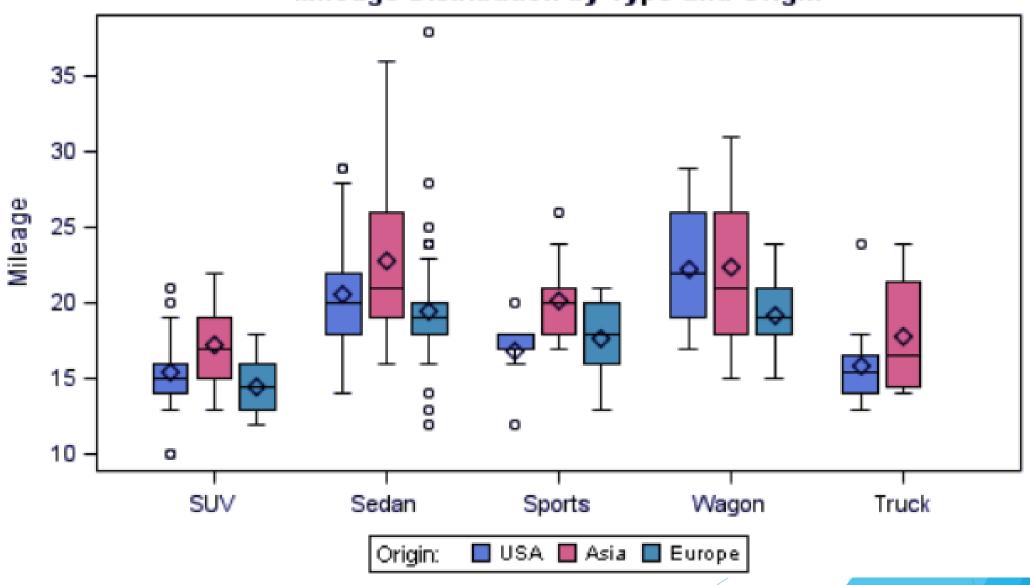
- 直方圖
- > 機率圖
- > 分位數圖
- ▶ 盒形圖



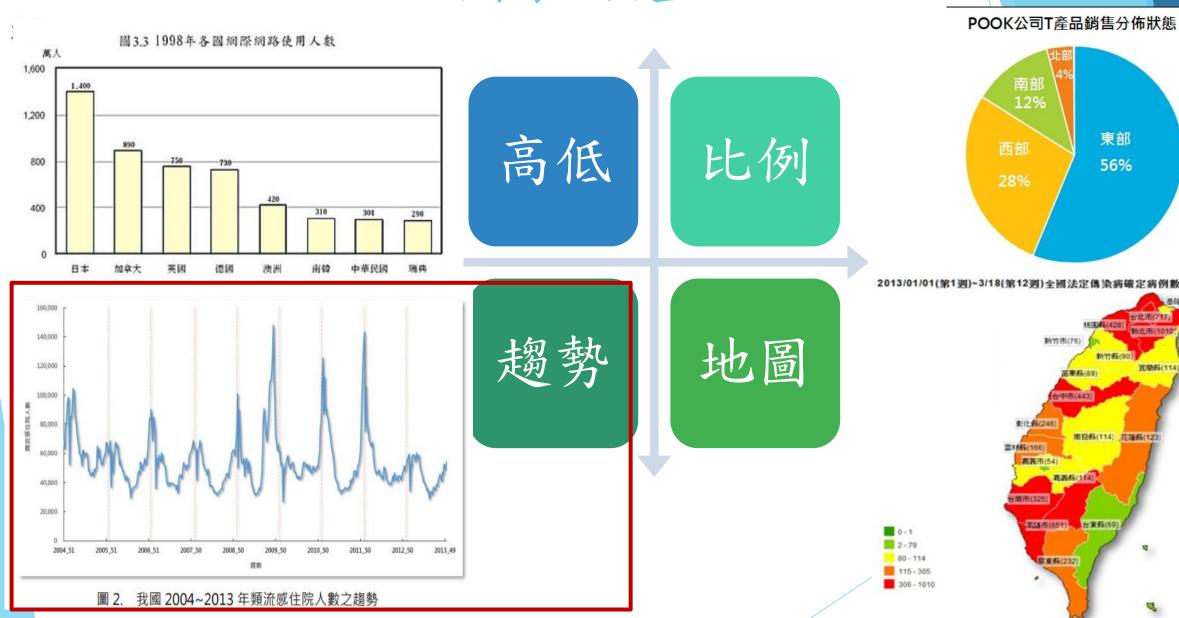
## 盒形圖



### Mileage Distribution by Type and Origin



# 四大圖表類型



### 折線圖

- ▶ 通常會跟觀察時間有關
- > 可以觀察時間趨勢
- > 可視為時間序列圖的一種
- ▶請使用SAS EG繪圖範例檔.xlsx中的"每周肺炎及流感死亡人數"工作表,畫出每周流感及肺炎死亡人數趨勢變化

匯入資料 (SAS EG繪圖範例檔以sҳ[每周肺炎及流感死亡人數]) ▼

2223	程式碼 📋 日誌	調 輸出資料	
<b>\$</b> 5	凤 修改工作(Y) 🍴	🤁 篩選和排序(L)	➡ 查詢產生器(Q) ▼
	▲ 死亡年週	▲ 年齢層	肺炎及流感 ⑩ 死亡人數
1	201926	0-49歳	9.00
2	201927	0-49歳	11.00
3	201928	0-49歳	9.00
4	201929	0-49歳	5.00
5	201930	0-49歳	11.00

匯入	匯入資料 (SAS EG繪圖範例檔以sҳ[每周肺炎及流感死亡人數]) ▼ ×							×	
2222	程式碼 📋 日説	調 輸出資料							
_	■ 修改工作(Y)	- │ <mark>₹</mark> 7 篩選和排序(L)	■ 查詢產生器(Q)	₹ 位置(W)	│ 資料(D) ·	▼ 描述(E)	▼ 🖺	形(G) • 分析(Z) •	F
	🔌 死亡年週	▲ 年齢層	肺炎及流感 1937 死亡人數					長條圖精靈(H) 長條圖(B)	
1	201926	0-49歲	9.00			L	-		
2	201927	0-49歳	11.00			- 1		圓形圖精靈(E)	
3	201928	0-49歳	9.00			- 0		<b>圓形圖(Р)</b>	
4	201929	0-49歳	5.00				~	折線圖精靈(W)	
5	201930	0-49歳	11.00				~	折線圖(L)	
6	201931	0-49歳	10.00						
7	201932	0-49歳	8.00					散佈圖(S) 数体图(5)た00	
8	201933	0-49歳	9.00					散佈圖矩陣(×)	
9	201934	0-49歳	8.00			10		區域圖(A)	
10	201935	0-49歳	9.00					條線圖(I)	
11	201936	0-49歳	11.00				80	氣泊圖(U)	
12	201937	0-49歳	15.00					環形圖(D)	
13	201938	0-49歳	10.00				ď	等高線圖(N)	
14	201939	0-49歳	8.00				<u> P</u>	盒形圖(O)	
15	201940	0-49歳	13.00				<b>*</b>	雷達圖(R)	
16	201941	0-49歳	14.00			4		曲面圖(F)	
17	201942	0-49歳	16.00				H	方 <b>塊圖</b> (T)	
18	201943	0-49歳	15.00					地圖圖表(M)	
19	201944	0-49歲	12.00				- 24	開啟 ODS 圖形設計程式	
20	201945	0-49歳	10.00					顯示 ODS 統計圖(C)	·(~)
21	201946	N-49歳	8 00			1	<u></u>	# 現 / I O D 3 # 7 语   回   ( C )	

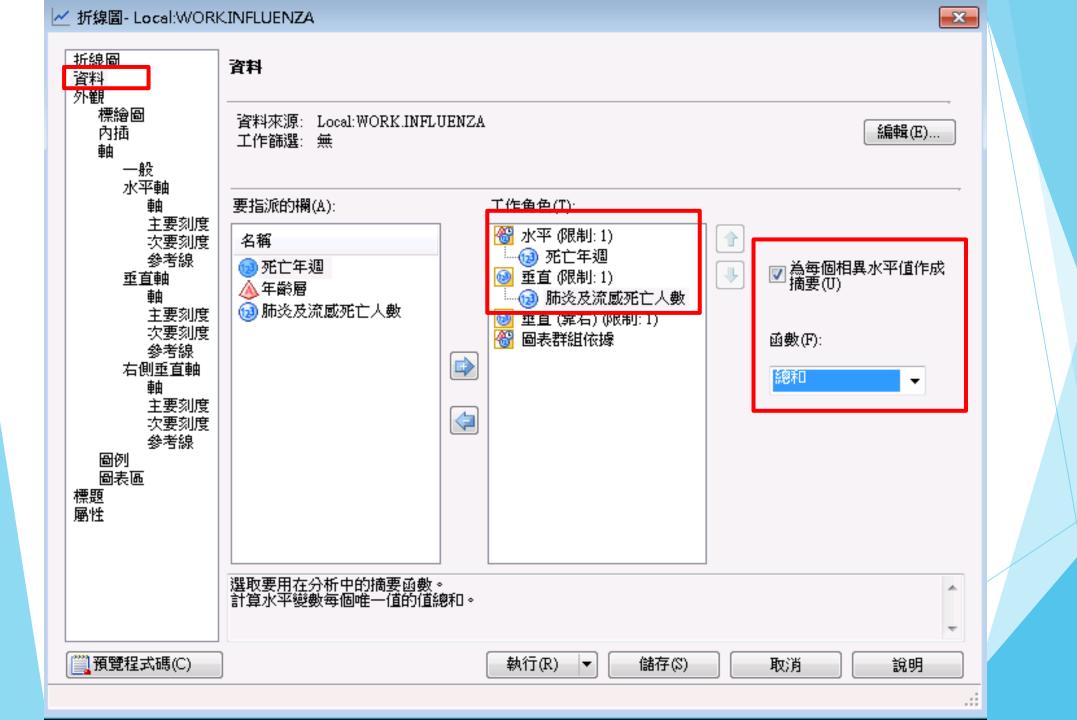
○ 預覽程式碼(C)

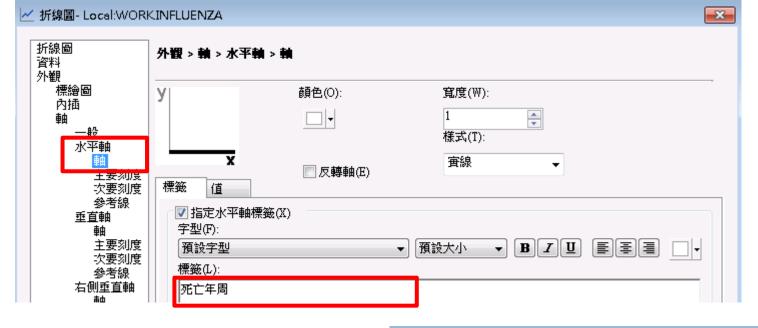
執行(R)

儲存(3)

取消

說明





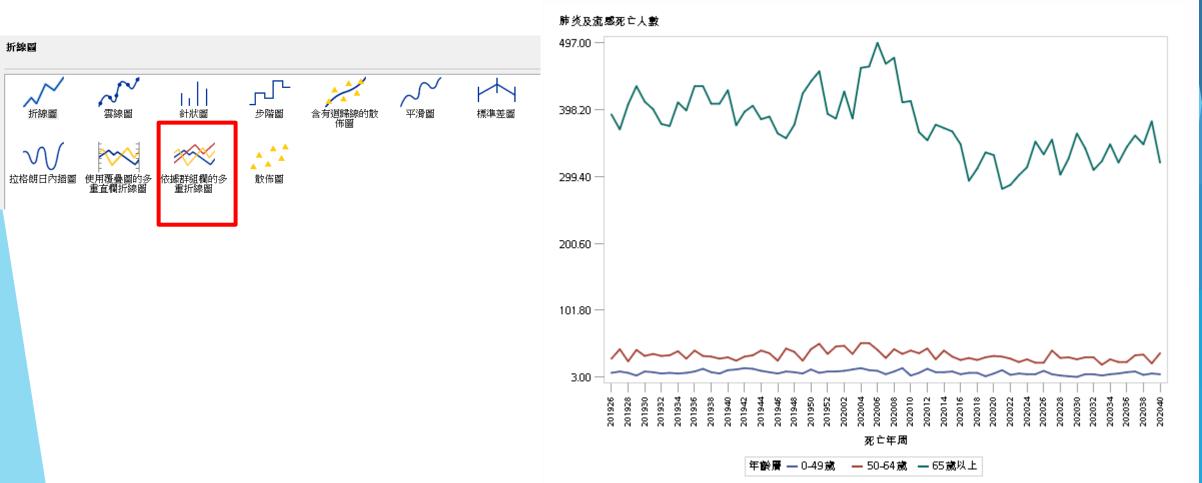




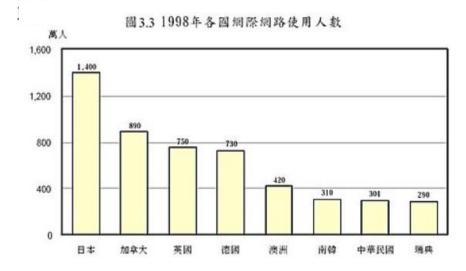
### 課堂練習

試著想想看,若是想要畫出"不同年齡層之每周流感及肺炎死亡人數趨勢變化",要如何繪圖?

#### 每周流感及肺炎死亡人数趨勢變化--年齡分層

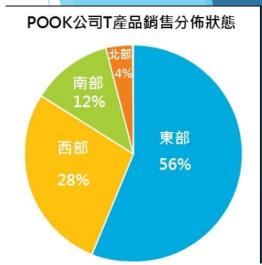


# 四大圖表類型



高低

比例



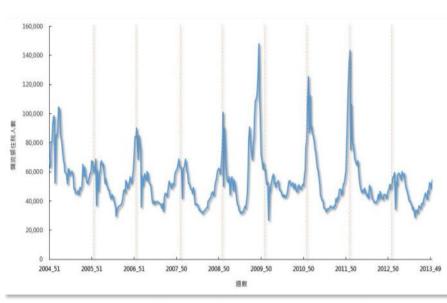
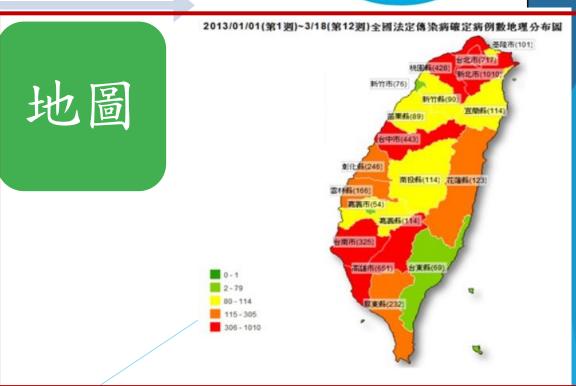


圖 2. 我國 2004~2013 年類流感住院人數之趨勢

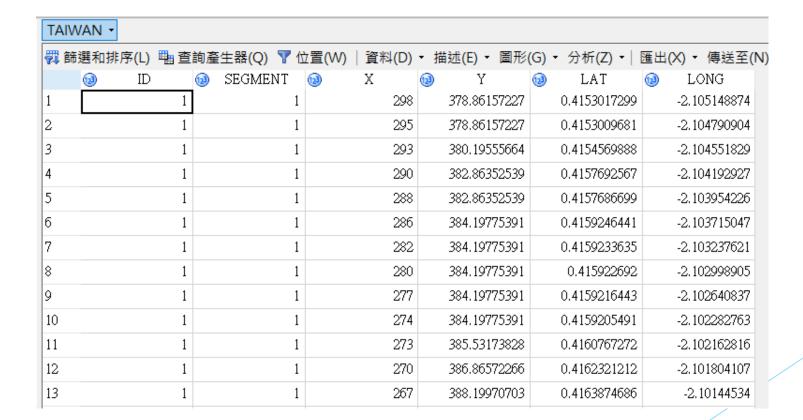
趨勢

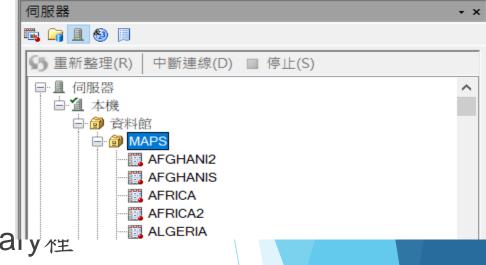


地圖資料

### 地圖

- > SAS有全世界地圖圖資
- ▶ 以SAS資料集形式 (.sas7bdat) 儲存MAPS的Library程







#### 

8	篩選和排序(L) 🖷 查	፤詢產生器(Q) 🍸 位置(	(W)  資料(D) ·	· 描述(E) · 圖形	(G) ▼ 分析(Z) ▼	匯出(※) ▼ 傳送至(1
	₀ ID		Х	⊚ Y	⊚ LAT	LONG     ■
1	1	1	298	378.86157227	0.4153017299	-2.105148874
2	1	1	295	37	81	-2.104790904
3	1	1	293	38 終 3	<b>建度</b> ※	-2.104551829
4	1	1	290	38 ""	567	-2.104192927
5	1	1	288	382.86352539	0.4157686699	-2.103954226
6	1	1	286	384.19775391	0.4159246441	-2.103715047
7	1	1	282	384.19775391	0.4159233635	-2.103237621
8	1	1	280	384.19775391	0.415922692	-2.102998905
9	1	1	277	384.19775391	0.4159216443	-2.102640837
10	1	1	274	384.19775391	0.4159205491	-2.102282763
1	. 1	1	273	385.53173828	0.4160767272	-2.102162816
12	1	1	270	386,86572266	0.4162321212	-2.101804107
13	1	1	267	388.19970703	0.4163874686	-2.10144534
14	1	1	264	388.19970703	0.4163862141	-2.101087198
13	1	1	261	388.19970703	0.4163849112	-2.100729056
16	1	1	258	386.86572266	0.4162270082	-2.100371629
17	' 1	1	255	386.86572266	0.4162256103	-2.100013517
18	1	1	253	385.53173828	0.4160680994	-2.09977553
19	1	1	251	386.86572266	0.4162236731	-2.099536024
20	1	1	248	389.53369141	0.4165352704	-2.09917631
2	. 1	1	249	392.20166016	0.416848884	-2.099294111
22	1	1	252	394.86987305	0.4171635117	-2.099650785
23	1	1	255	398.87182617	0.4176346157	-2.100006819

#### TAIWAN2 -

41	篩選和排序(L) 🖷	查詢產生器(Q) 🍸	位置(W)   資料([	O) ▼ 描述(E) ▼ 圖刑
	→ _MAP_GEOM ETRY_	○ COUNTRY	⊚ ID	
1	id 1 of maps.tai…	281	1	CHANGHUA
2	id 2 of maps.tai…	281	2	CHIAI
3	id 3 of maps.tai…	281	3	CHILUNG
4	id 4 of maps.tai…	281	4	HSINCHU
5	id 5 of maps.tai…	281	5	HUALIEN
б	id δ of maps.tai…	281	б	ILAN
7	id 7 of maps.tai…	281	7	KAOHSIUNG
8	id 8 of maps.tai…	281	8	KAOHSIUNG
9	id 9 of maps.tai…	281	9	MIAOLI
10	id 10 of r	划 66 18 1	7夕经 <sup>[0</sup>	NANTOU
11	id 11 of r 百寸	制的縣市	1石件  1	PENGHU
12	id 12 of maps.ta···	281	12	PINGTUNG
13	id 13 of maps.ta…	281	13	TAICHUNG CI…
14	id 14 of maps.ta…	281	14	TAICHUNG
15	id 15 of maps.ta…	281	15	TAINAN
16	id 1б of maps.ta…	281	16	TAINAN CITY
17	id 17 of maps.ta…	281	17	TAIPEI
18	id 18 of maps.ta···	281	18	TAIPEI CITY
19	id 19 of maps.ta…	281	19	TAITUNG
20	id 20 of maps.ta…	281	20	TAOYUAN
21	id 21 of maps.ta···	281	21	YUNLIN

## 練習檔案

- ► SAS EG繪圖範例檔.xlsx中的"city"工作表ID變項需和地圖檔的ID相同
- n\_hosp
  - ▶ 醫療機構數
- n\_perhosp
  - ▶ 平均每一醫療機構服務人數
- n\_physician
  - > 執業醫師數
- staff\_tenthousand
  - ▶ 每萬人口執業醫事人員

city	y <del>*</del>					
<b>41</b>						匯出(X) ▼ 傳送至
		n_hosp	n_perhosp		⊚ staff_tenthou…	•
1	新北市	3200	1239.63	5849	78.62	17
2	臺北市	3450	783.28	9909	181.93	18
3	桃園市	1523	1351.5	4084	111.78	20
4	臺中市	3342	813.83	6851	128.37	13
5	臺南市	1875	1004.95	3770	121.01	15
б	高雄市	2906	956.29	6490	130.86	7
7	宜蘭縣	337	1361.36	736	114.26	б
8	新竹縣	360	1493.42	537	69.81	4
9	苗栗縣	391	1450.47	670	77.32	9
10	彰化縣	1062	1216.08	2282	100.2	1
11	南投縣	434	1185.06	748	85.98	10
12	雲林縣	508	1388.5	1041	87.39	21
13	嘉義縣	271	1936.47	803	95.14	2
14	屏東縣	655	1294.53	1265	99.02	12
15	臺東縣	161	1394.22	319	100.77	19
16	花蓮縣	278	1199.25	852	147.99	5
17	澎湖縣	89	1143.35	136	72.33	11
18	基隆市	297	1256.15	775	104.24	3
19	新竹市	438	986.27	937	128.64	4
20	嘉義市	409	662.31	966	218.84	2
21	金門縣	50	2554.46	76	33.9	
22	連江縣	5	2501.2	21	64.77	
23	高雄市	2906	956.29	6490	130.86	8
24	臺中市	3342	813.83	6851	128.37	14
25	臺南市	1875	1004.95	3770	121.01	16

~	ı	T		•
•	ı	•	y	

<b></b>			位置(M) │ 資料(D)	) ▼ 描述(F) ▼	  形(G) ▼ 分析(Z) ▼   匯出(X) ▼ 傳送至(I
<b>是</b> 春	oity_name	回 n_hosp			長條圖精靈(H)
1	新北市	3200	1239.63	5	
2	臺北市	3450	783.28	9	
3	桃園市	1523	1351.5	41 😝	(-)
4	臺中市	3342	813.83	6 ~	
5	臺南市	1875	1004.95	3' 📈	
б	高雄市	2906	956.29	6	
7	宜蘭縣	337	1361.36	(4)	
8	新竹縣	360	1493.42	la l	區域圖(A)
9	苗栗縣	391	1450.47	lesi.	[ 條線圖(I)
10	彰化縣	1062	1216.08	2: 80	空 氣泡圖(U)
11	南投縣	434	1185.06	•	環形圖(D)
12	雲林縣	508	1388.5	1	
13	嘉義縣	271	1936.47	[ <u>0</u> 00	` '
14	屏東縣	655	1294.53	1 🖈	
15	臺東縣	161	1394.22	<b>₩</b>	
1б	花蓮縣	278	1199.25	#	, , ,
17	澎湖縣	89	1143.35		
18	基隆市	297	1256.15	4	
19	新竹市	438	986.27	4	顯示 ODS 統計圖(C)
20	嘉義市	409	662.31	966	218.84 2



