優化通識教育科的數據分析

陳秀騰

2018年6月14日

https://sites.google.com/site/chansautang/

簡報要點:

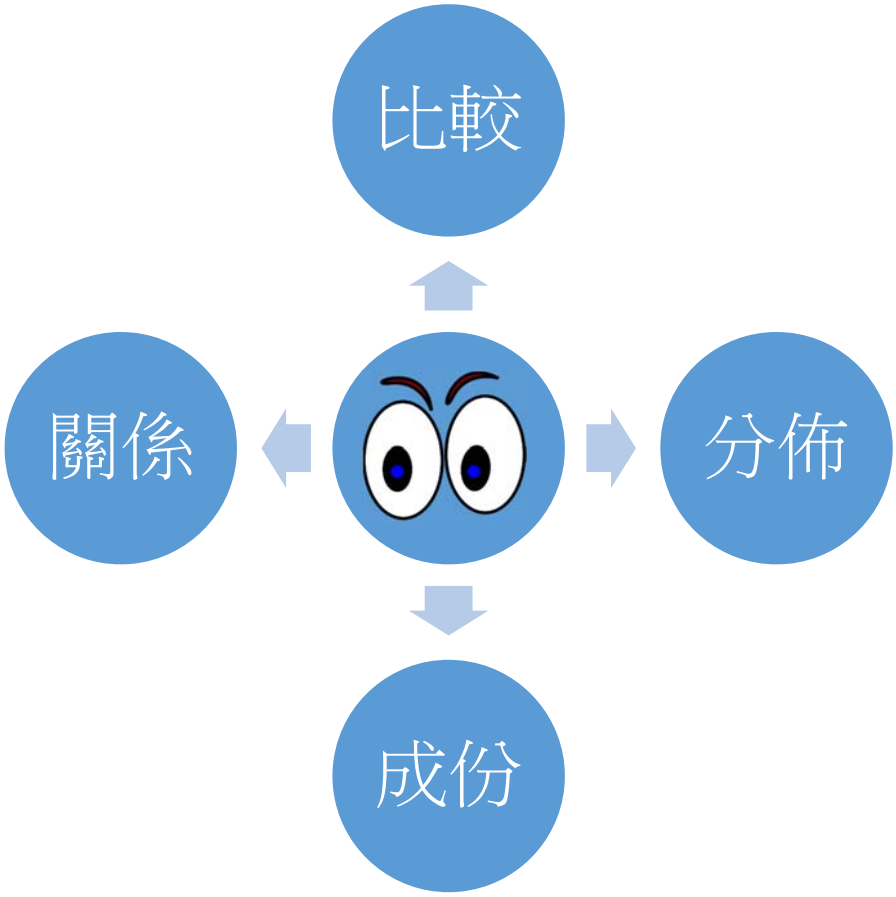
1) 統計圖的選取

2) 收集數據的示例

為何採用 統計圖？

1) 容易引人注意
2) 有助於概括(summarise)大量的數據集合(data set)
3) 把重點放在數據的某一方面(aspect)
4) 表示數據於某一時段的趨勢 (trend)

選取統計圖:


比較
關係
分佈
成份

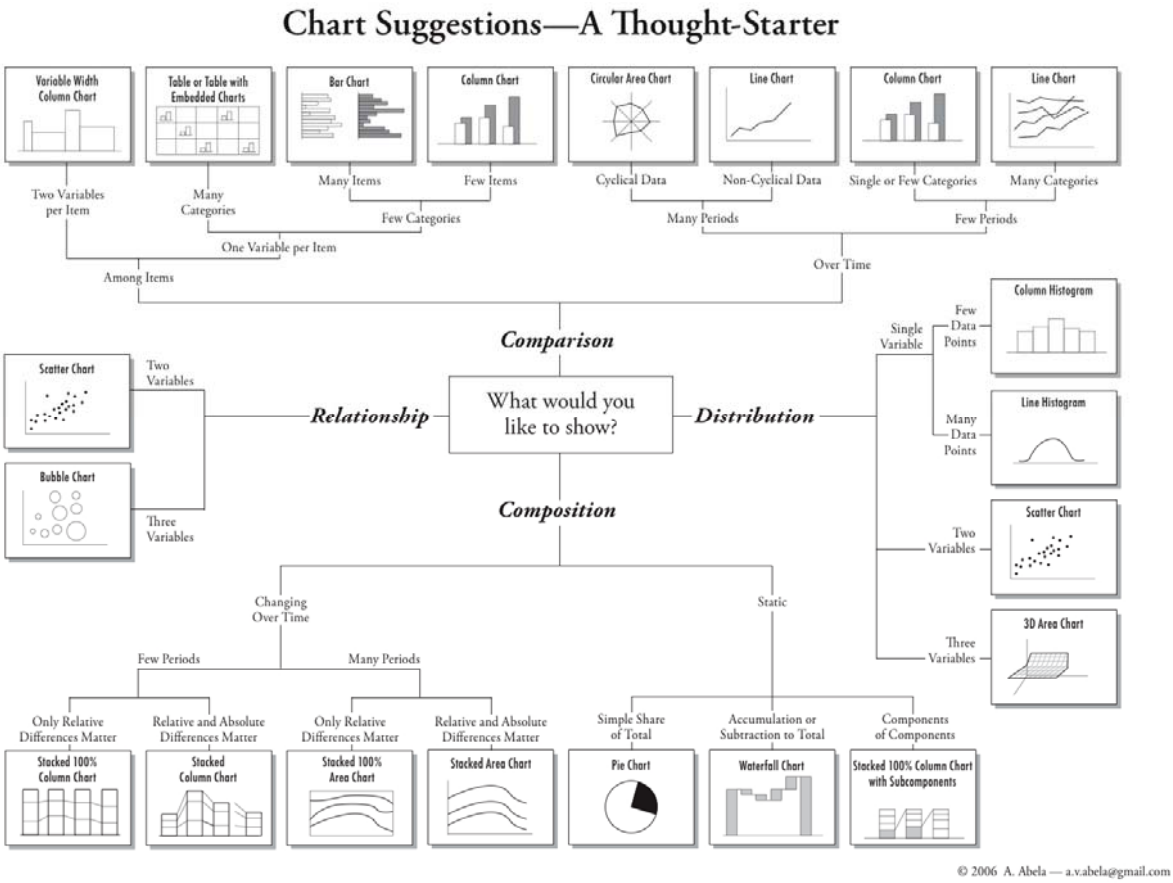
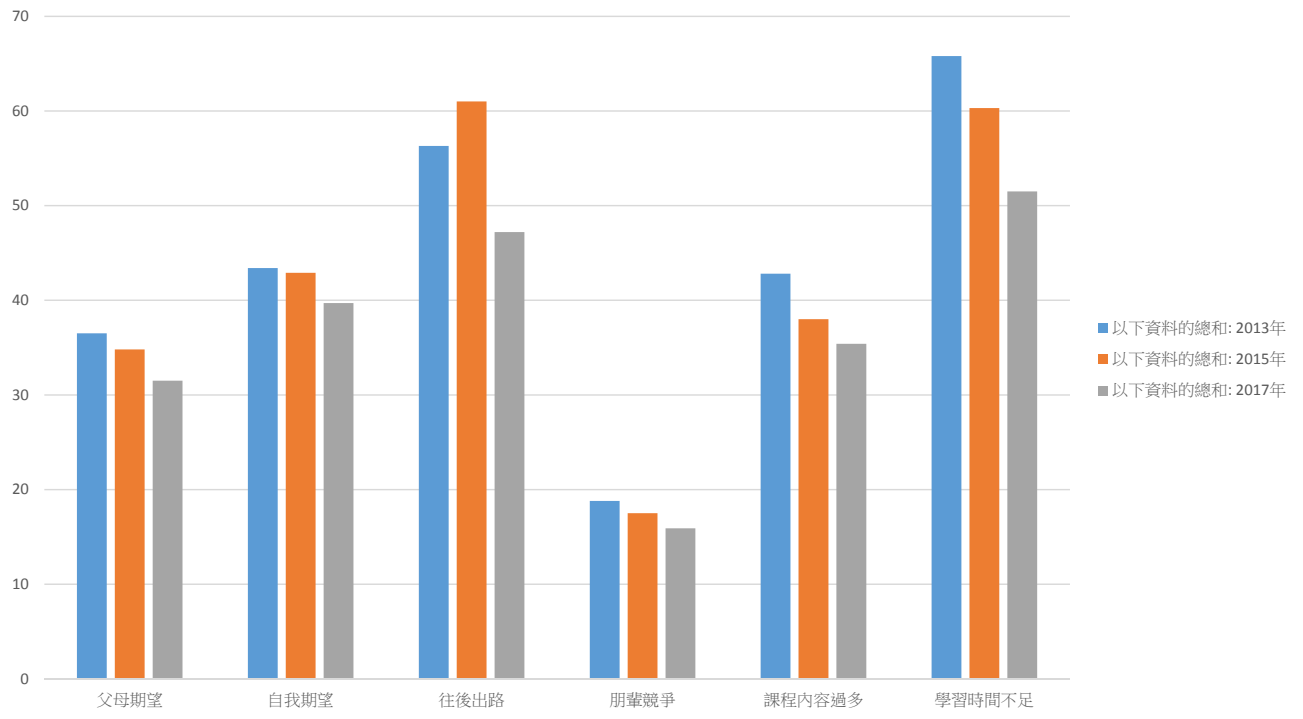


Chart Suggestions—A Thought-Starter
Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart
Two Variables per Item
Many Categories
Many Items
Few Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories
Few Categories
Many Periods
Few Periods
One Variable per Item
Among Items
Over Time
Comparison
Scatter Chart
Two Variables
Relationship
What would you like to show?
Distribution
Bubble Chart
Three Variables
Composition
Single Variable
Few Data Points
Many Data Points
Column Histogram
Line Histogram
Two Variables
Scatter Chart
Three Variables
3D Area Chart
Changing Over Time
Static
Few Periods
Many Periods
Only Relative Differences Matter
Relative and Absolute Differences Matter
Only Relative Differences Matter
Relative and Absolute Differences Matter
Simple Share of Total
Accumulation or Subtraction to Total
Components of Components
Stacked 100% Column Chart
Stacked Column Chart
Stacked 100% Area Chart
Stacked Area Chart
Pie Chart
Waterfall Chart
Stacked 100% Column Chart with Subcomponents
© 2006 A. Abela — a.v.abela@gmail.com

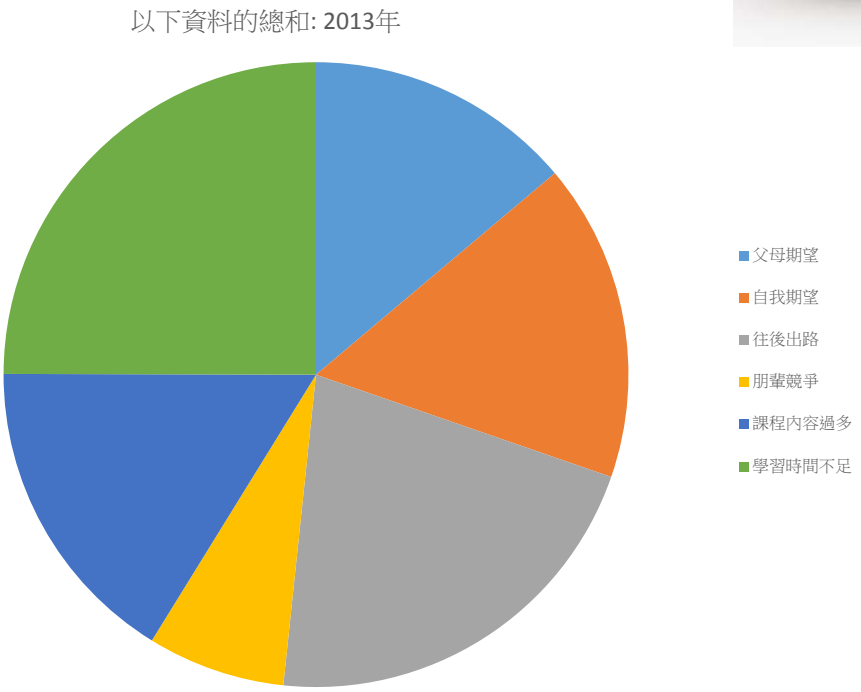
選取統計圖:



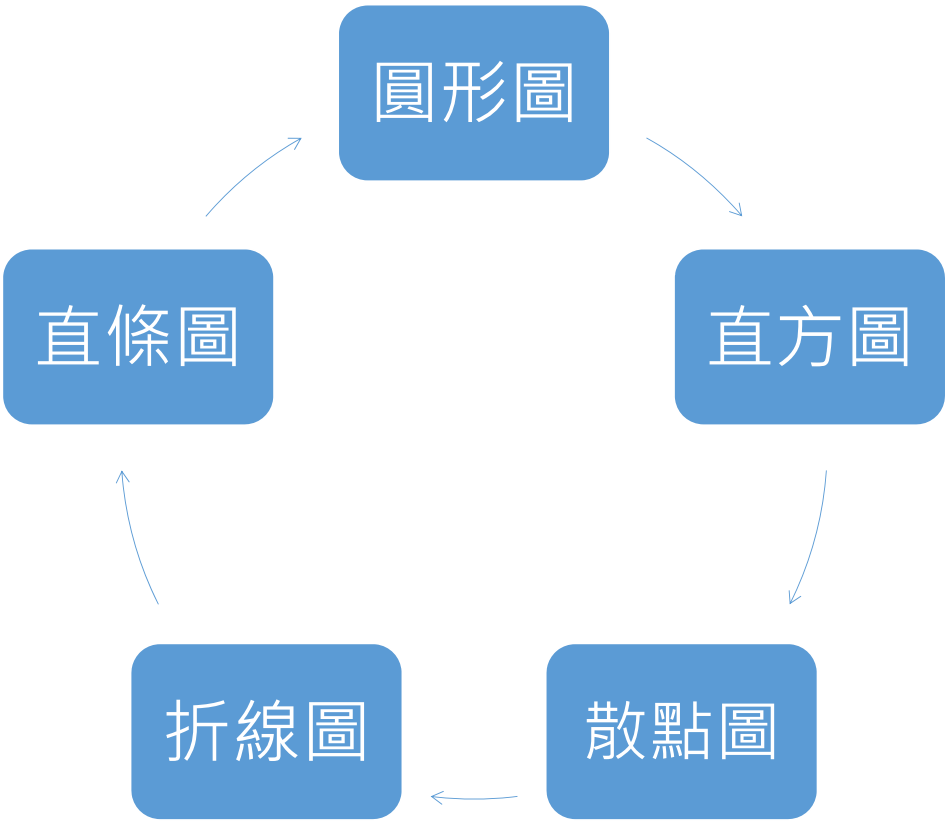
資料來自2018年香港中學文憑考試通識教育試卷一第2題資料回應題

選取統計圖:

選取統計圖:

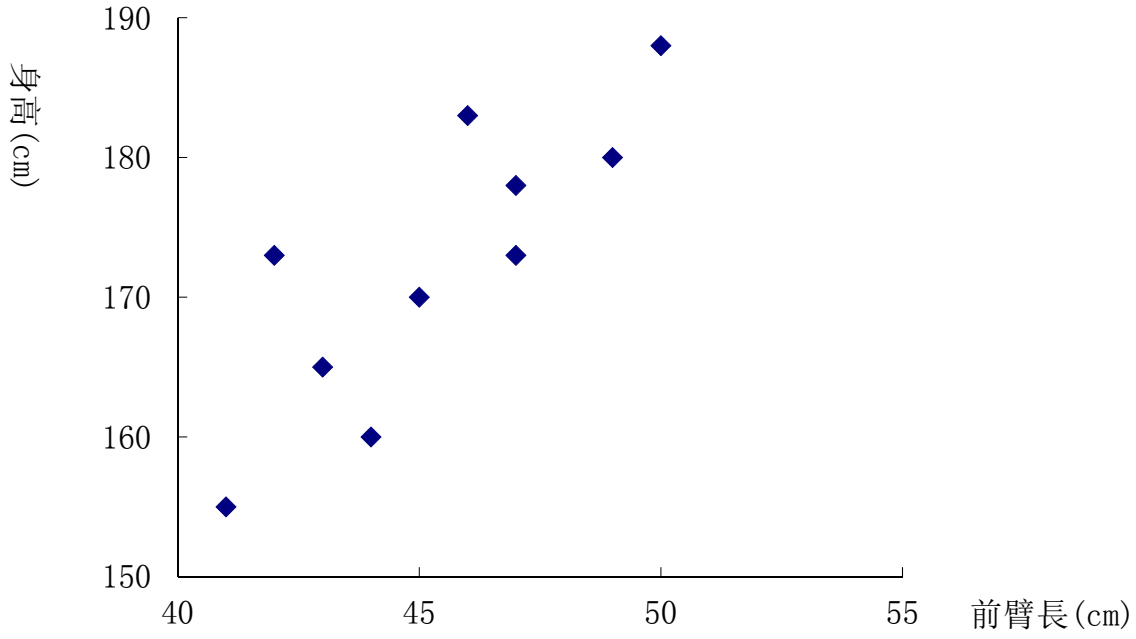

以下資料的總和: 2013年
父母期望
自我期望
往後出路
朋輩競爭
課程內容過多
學習時間不足

統計圖


圓形圖
直方圖
散點圖
折線圖
直條圖

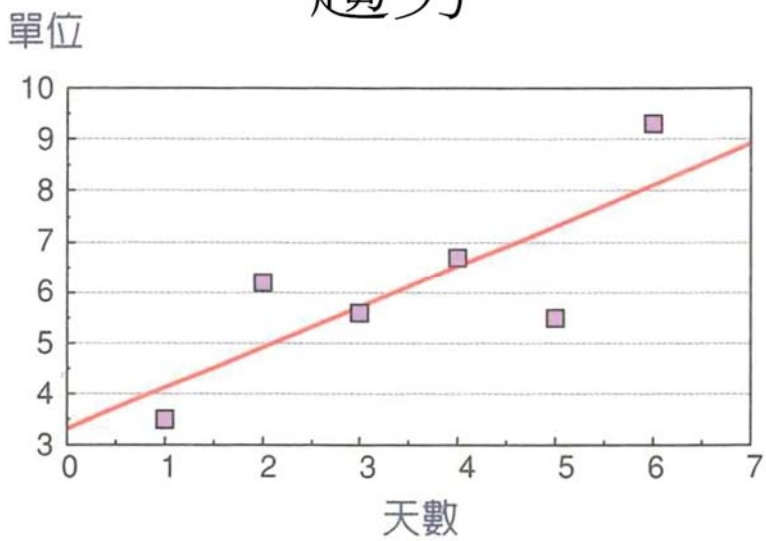
散點圖(scatter diagram)

1. 以點的密集程度和趨勢來表示兩種現象的相關關係
2. 繪製要點：橫軸代表一個變量，縱軸代表另一個變量
3. 縱橫軸的坐標起點不一定要從 0 開始
4. 反映兩事物間的相關關係，主要用於相關回歸分析



某地 20歲 男青年身高和前臂長的關係

趨勢

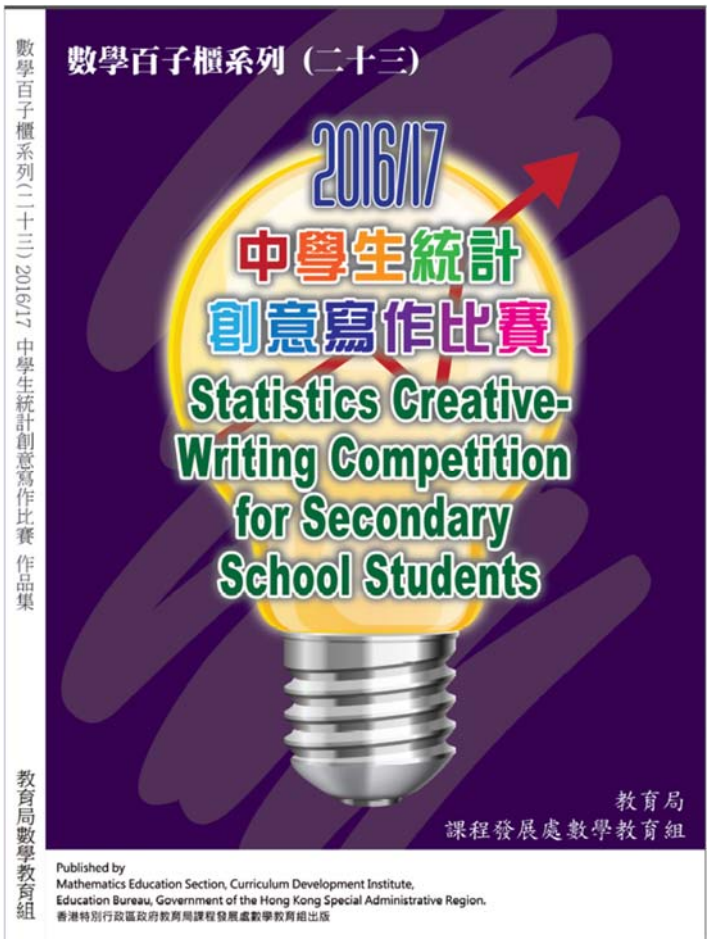


趨勢分析技巧
把數據過度簡化

注意事項

1. 統計圖類型
   - 離散資料: 直條圖
   - 連續資料: 直線圖 或 直方圖
   - 構成比資料: 圓形圖
   - 雙變量資料: 散點圖
2. 合適的標題。標題寫在圖的下方，要能夠概括圖的內容
3. 直條圖、直線圖、和直方圖的縱、橫坐標上要有刻度和單位，刻度要均勻等距。縱橫周長度之比為 5：7 較合適
4. 直條圖與直方圖縱坐標要求從 0 開始。
5. 比較不同事物時用不同的線條和顏色來表示，並附上圖例

收集數據示例



收集數據示例

收集數據示例

收集數據示例

Table 1: Distribution of age group in peak and non-peak hours

Timeslot	Peak hours*			
Total number	77			
Age group+	children	teens	working	elderly
Number of people using the seats	9	2	42	24

收集數據示例

Timeslot	Non-peak hours			
Total number	77			
Age group+	children	teens	working	elderly
Number of people using the seats	20	3	17	37

*Peak hours refers to the period of 7:00a.m. to 9:00 a.m. and the period of 6:00p.m. to 7:30p.m. during Mondays to Fridays.

+ Children are those below 13. Teens are those with the age between 13 and 19 years old. Working are the people between 20 and 59 years old. Elderly are those with the age above 60 years old.

收集數據示例

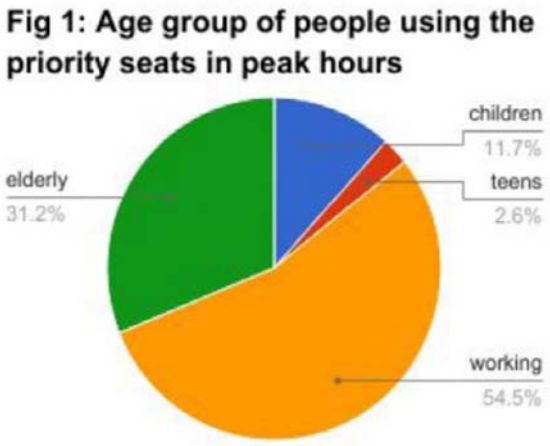
Table 2: Frequency of yielding seats to those in need

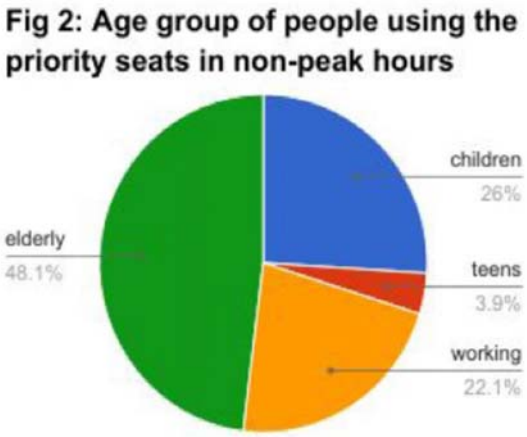
	0 times	1 times	2 or more times
Peak hours	56	17	4
Non-peak hours	68	4	0

Table 3: Distribution of yielding of seats to those in need

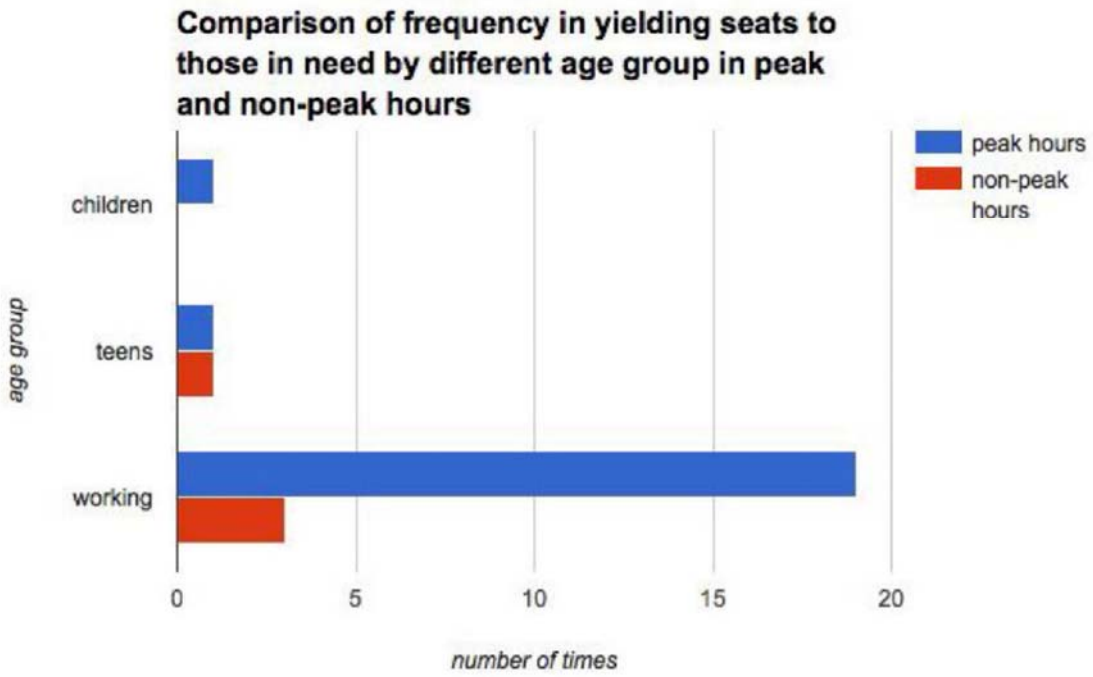
	Children	Teens	Working
Peak hours (times)	1	1	19
Non-peak hours (times)	0	1	3

收集數據示例





收集數據示例



收集數據示例

Fig 4: Frequency of yielding seats in peak hours

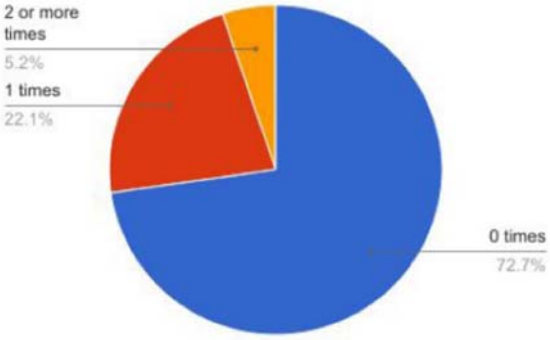
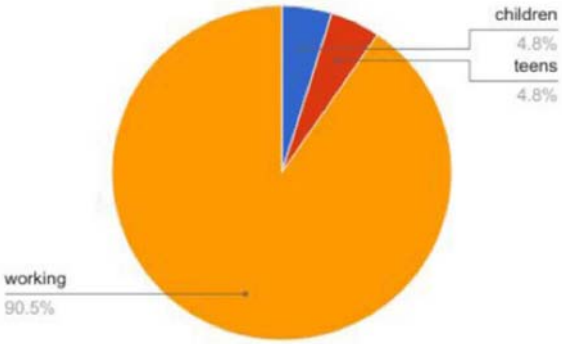


Fig 5: Distribution of yielding of seats to those in need by different age groups in peak hours



收集數據示例

Table 2

Frequency of yielding seats to those in need	0 times	1 times	2 or more times	Total number
Peak hours	56	17	4	77
Non-peak hours	68	4	0	77


感謝!