

# 應用資料探勘於棒球球種分析

組員：

B033040011 黃柏勳

B033040043 紀儒達

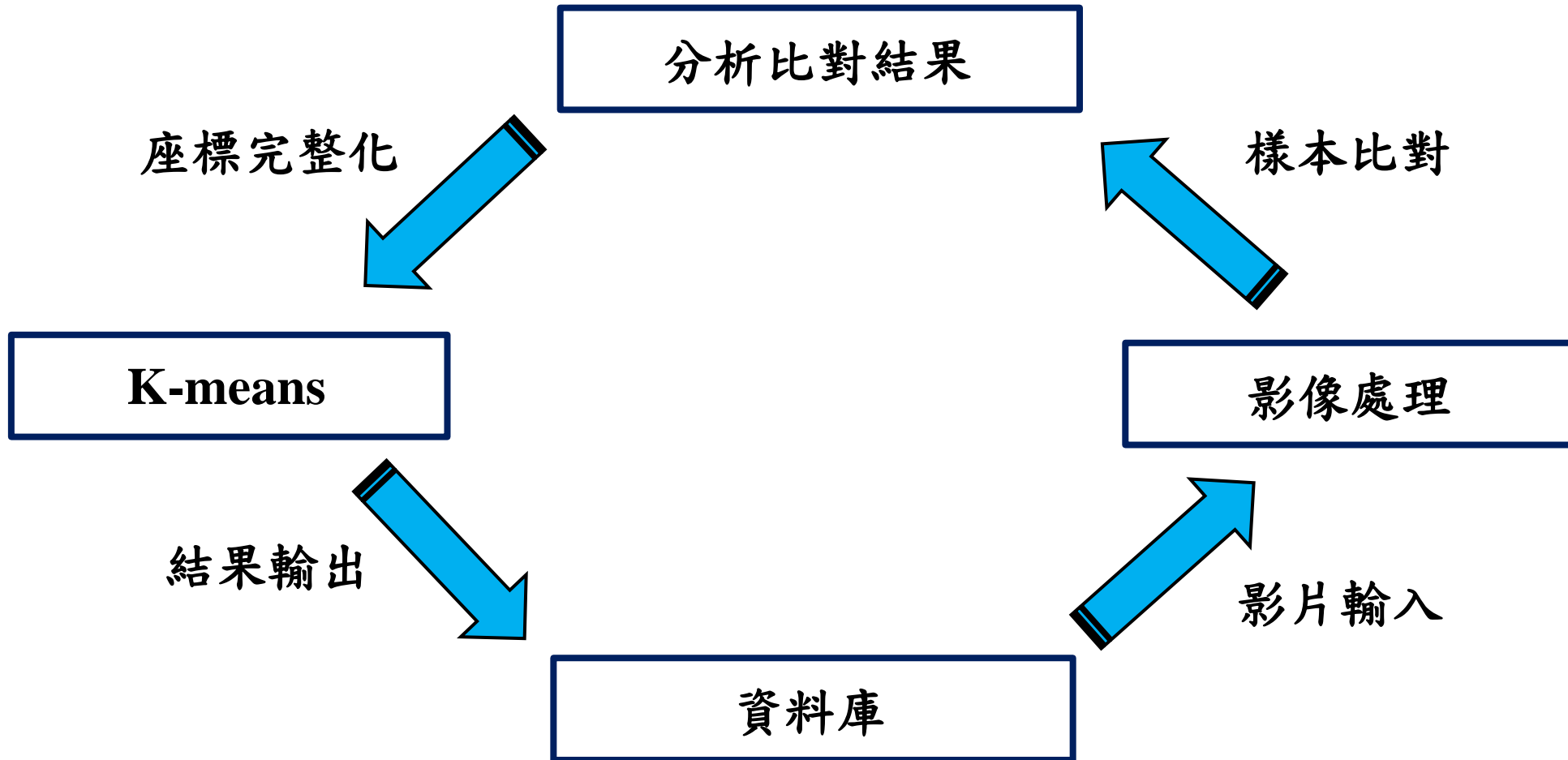
B033040049 曾昱榮

指導教授：張玉盈 教授

# 研究動機

- 對棒球的熱愛
- 棒球技術的演進
- 回饋所學

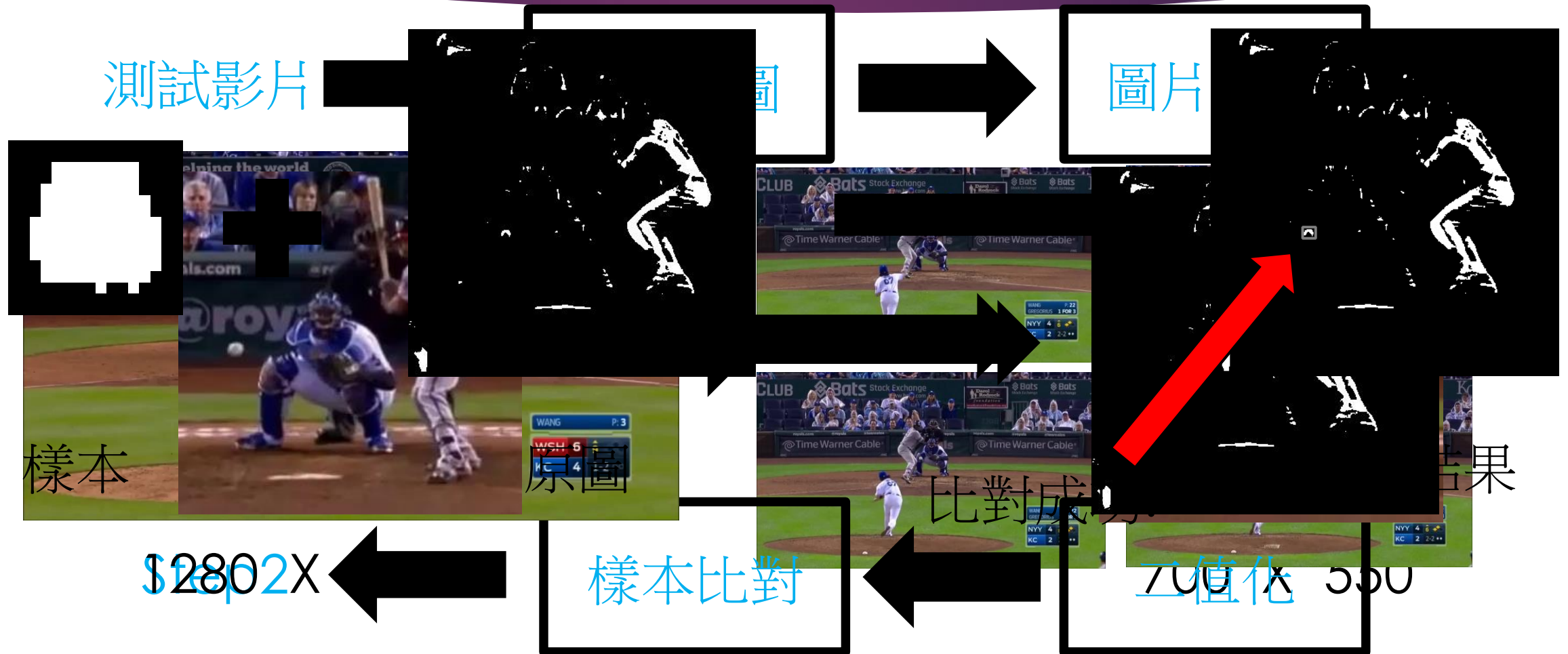
# 流程摘要





# 詳細步驟

# STEP 1 : 影像處理



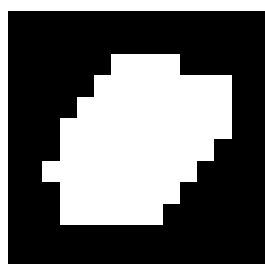
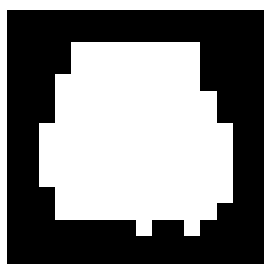
# STEP 2 : 分析比對結果(1/2)

結果成功!!!

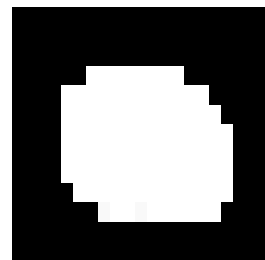
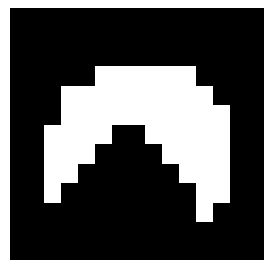
樣本1

樣本2

結果錯誤!!!



結果錯誤!!!



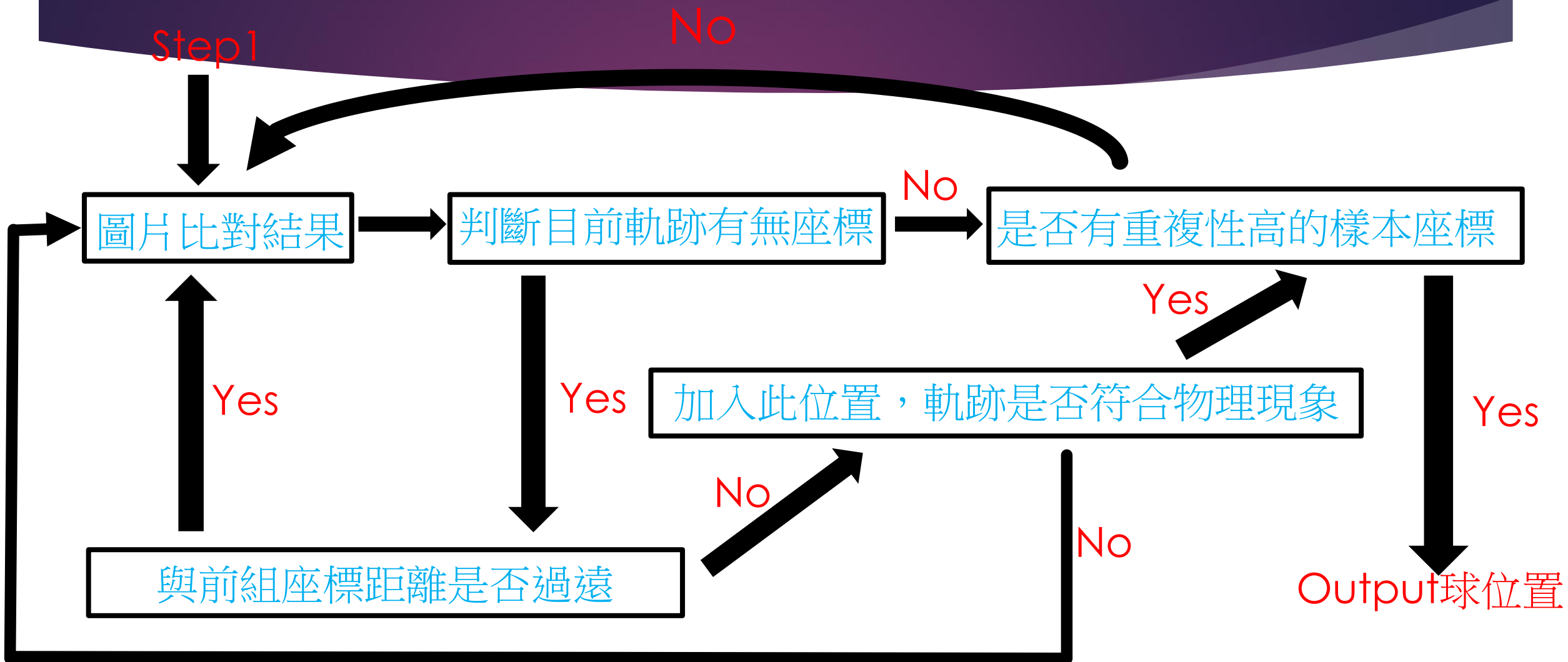
結果成功!!!

樣本4

樣本3

# STEP 2：分析比對結果(2/2)

Step 1



No

No

是否有重複性高的樣本座標

Yes

Yes

Yes

加入此位置，軌跡是否符合物理現象

Yes

No

No

與前組座標距離是否過遠

Output球位置

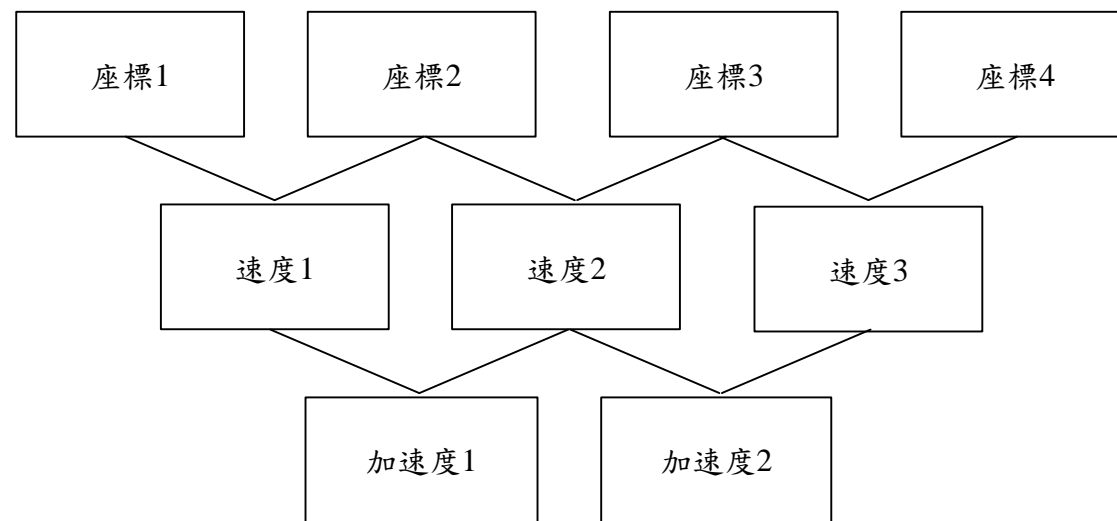
# STEP3 : 座標完整化

等加速度運動之公式  $\Delta X = vt + \frac{at^2}{2}$

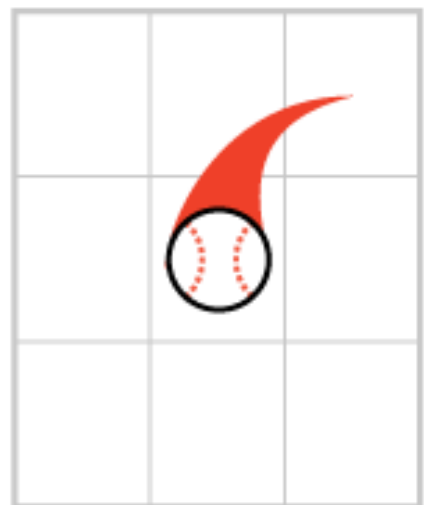
205	163		
220	163		
235	165		
Null	Null	→	251 168
267	177		
Null	Null	→	282 182
Null	Null	→	299 192
316	220		
333	213		
351	226		



# 計算加速度

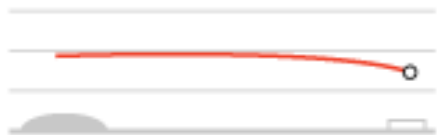


# 軌跡曲線量化



縱向軌跡加速度小

橫向軌跡加速度小

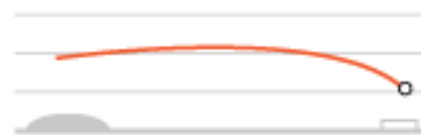


Fastball



縱向軌跡加速度大

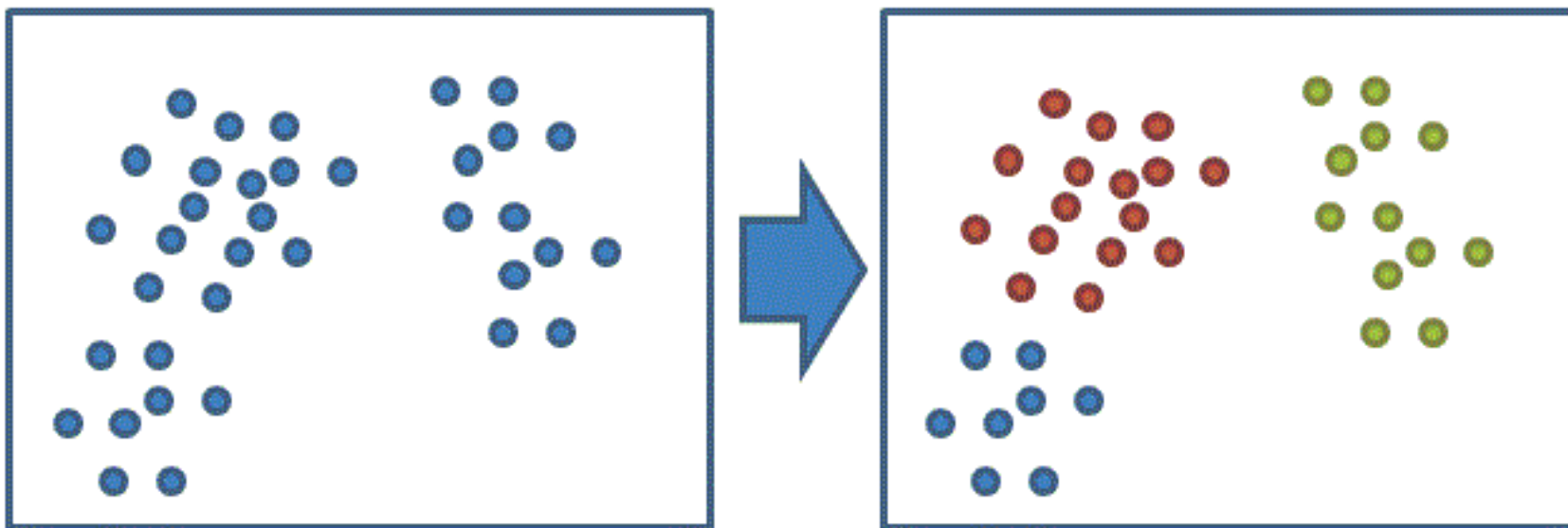
橫向軌跡加速度大



Slider

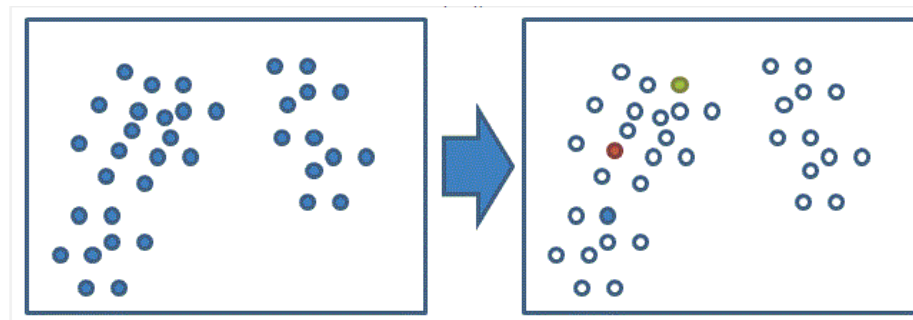
# K-MEANS演算法介紹

- ▶ 分群演算法
- ▶ K代表分群的群數

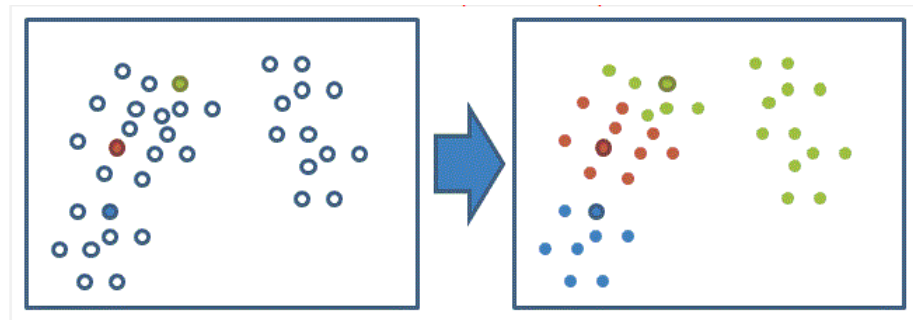


# K-MEANS演算法流程(1/2)

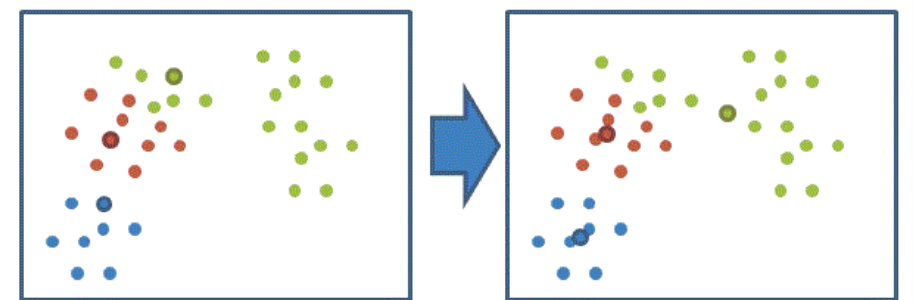
STEP 1 : 隨機選取資料組中的k筆資料當作初始群中心



STEP 2 : 計算每個資料對應到最短距離的群中心



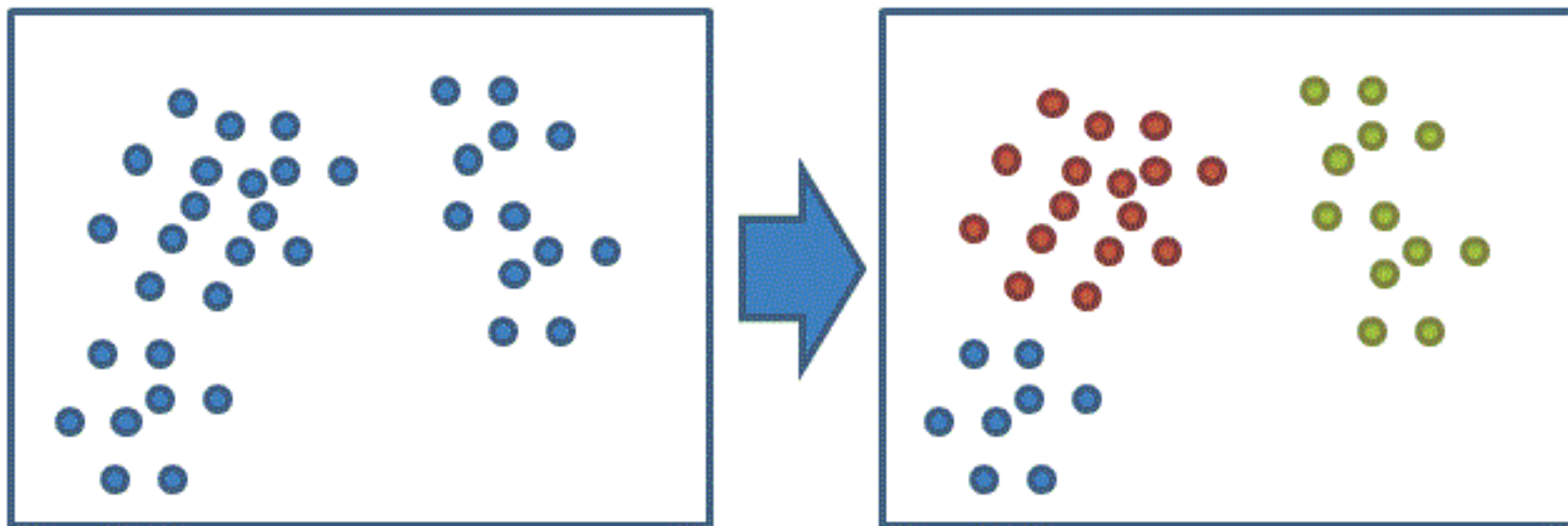
STEP 3 : 利用目前的分類重新計算群中心



# K-MEANS演算法流程

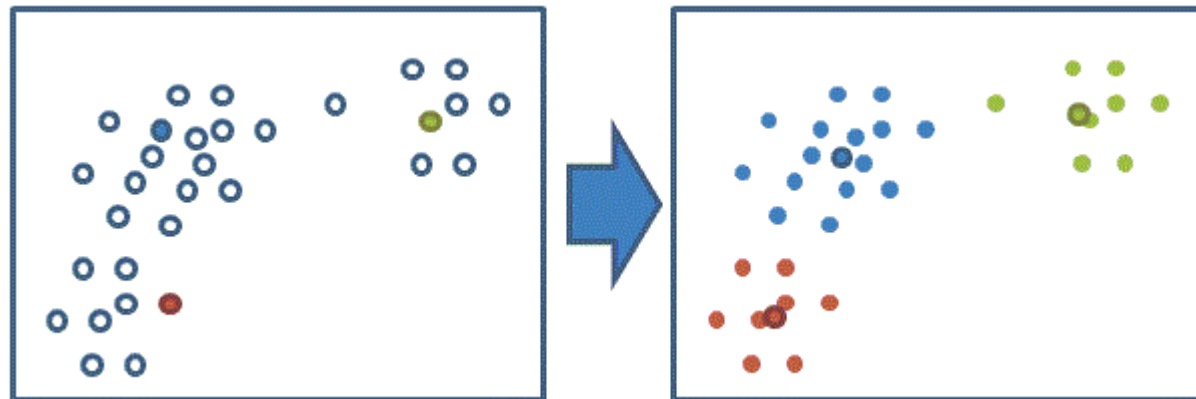
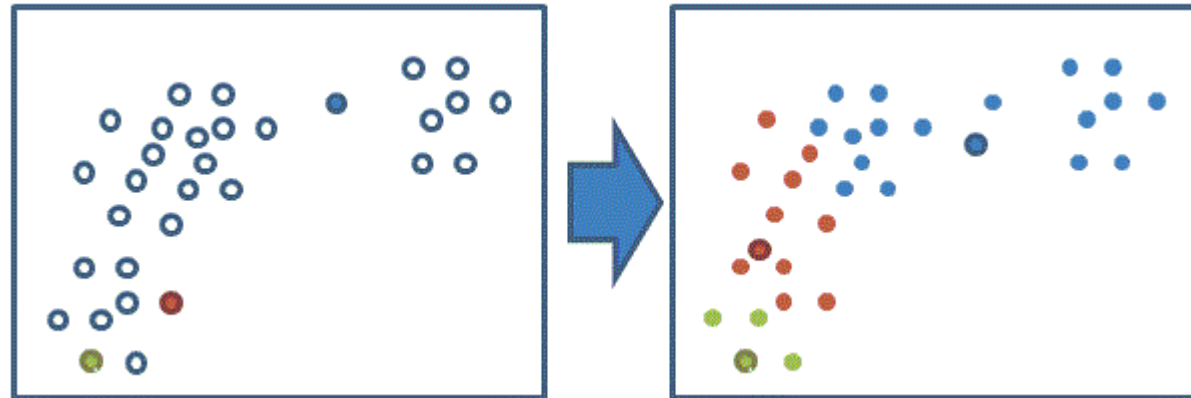
(2/2)

STEP 4 : 重複step 2,3直到收斂 (群心中移動距離趨近於零)



# 初始値設定(1/2)

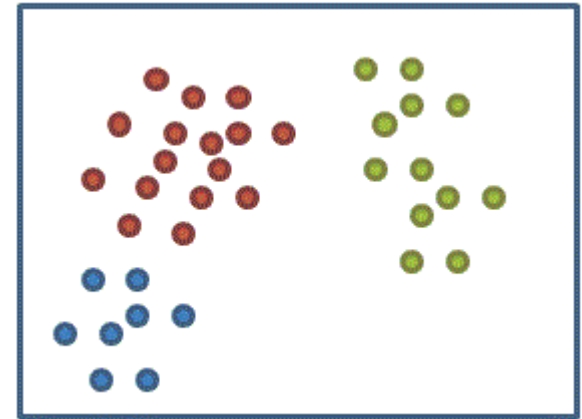
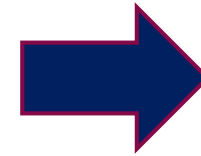
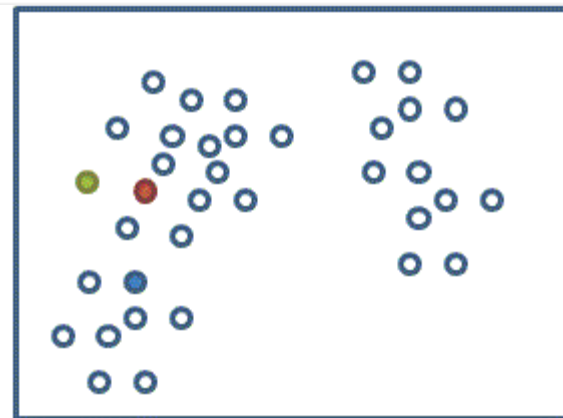
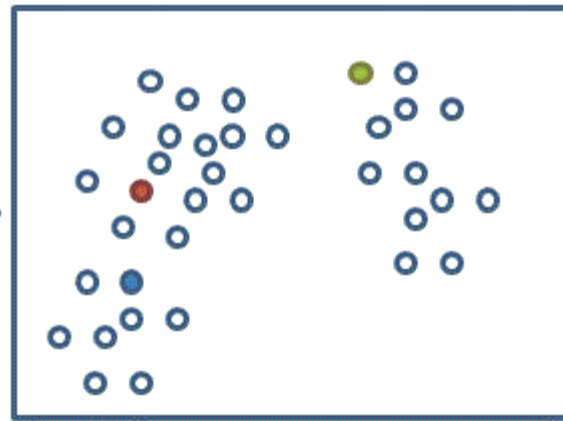
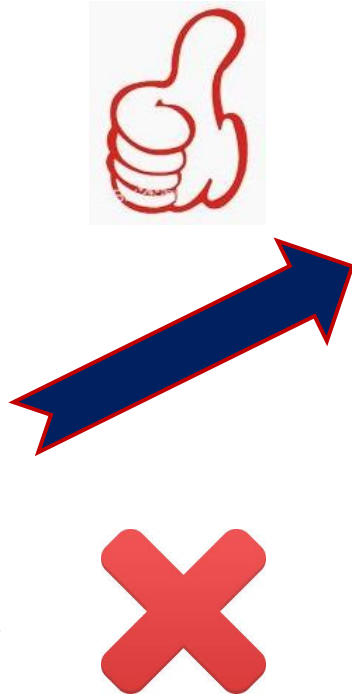
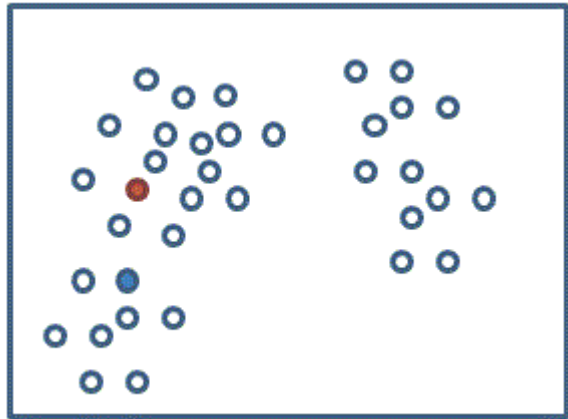
**CASE:**



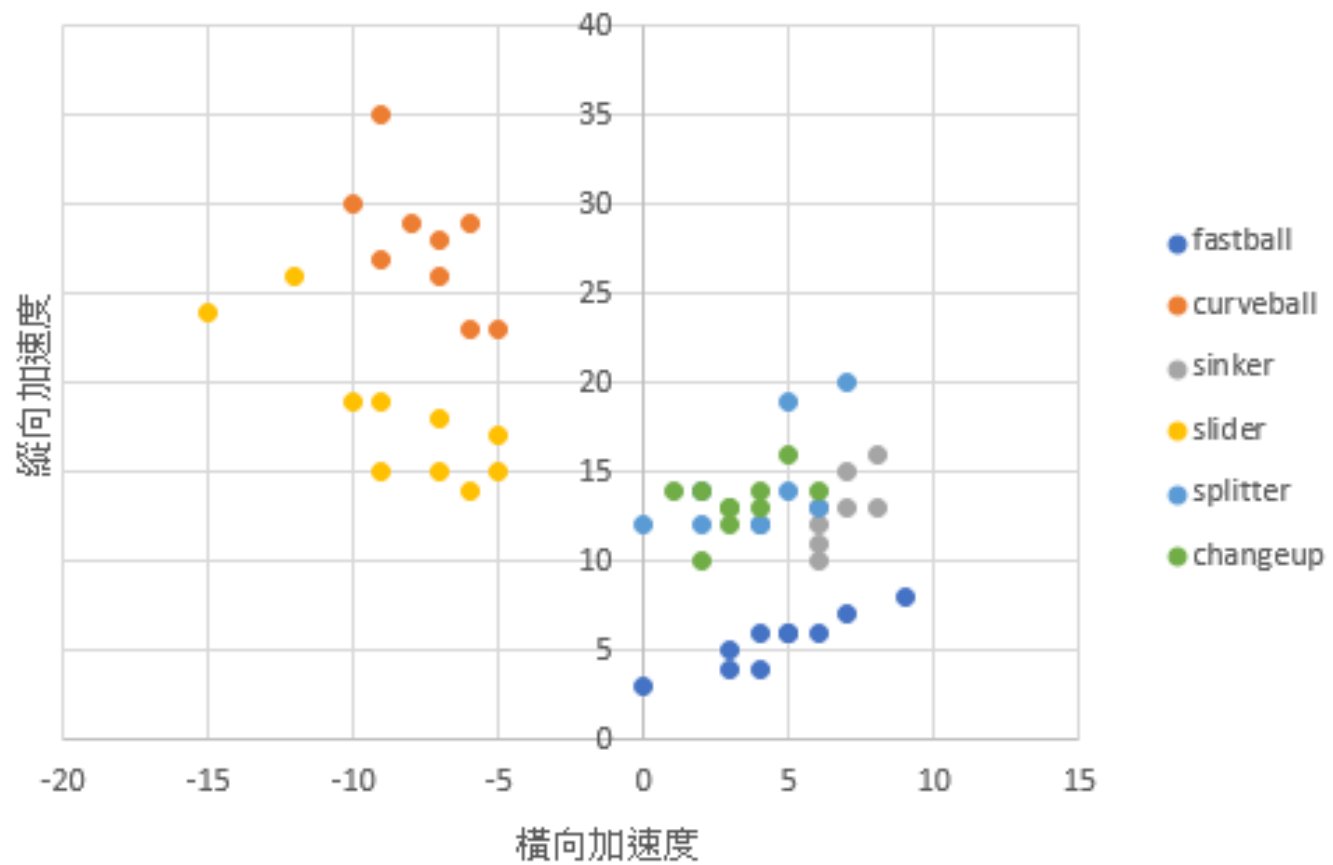
# 初始值設定(2/2)

解決方法：計算每個隨機初始點兩兩之距離，距離越大  
選中機率越高

EX:



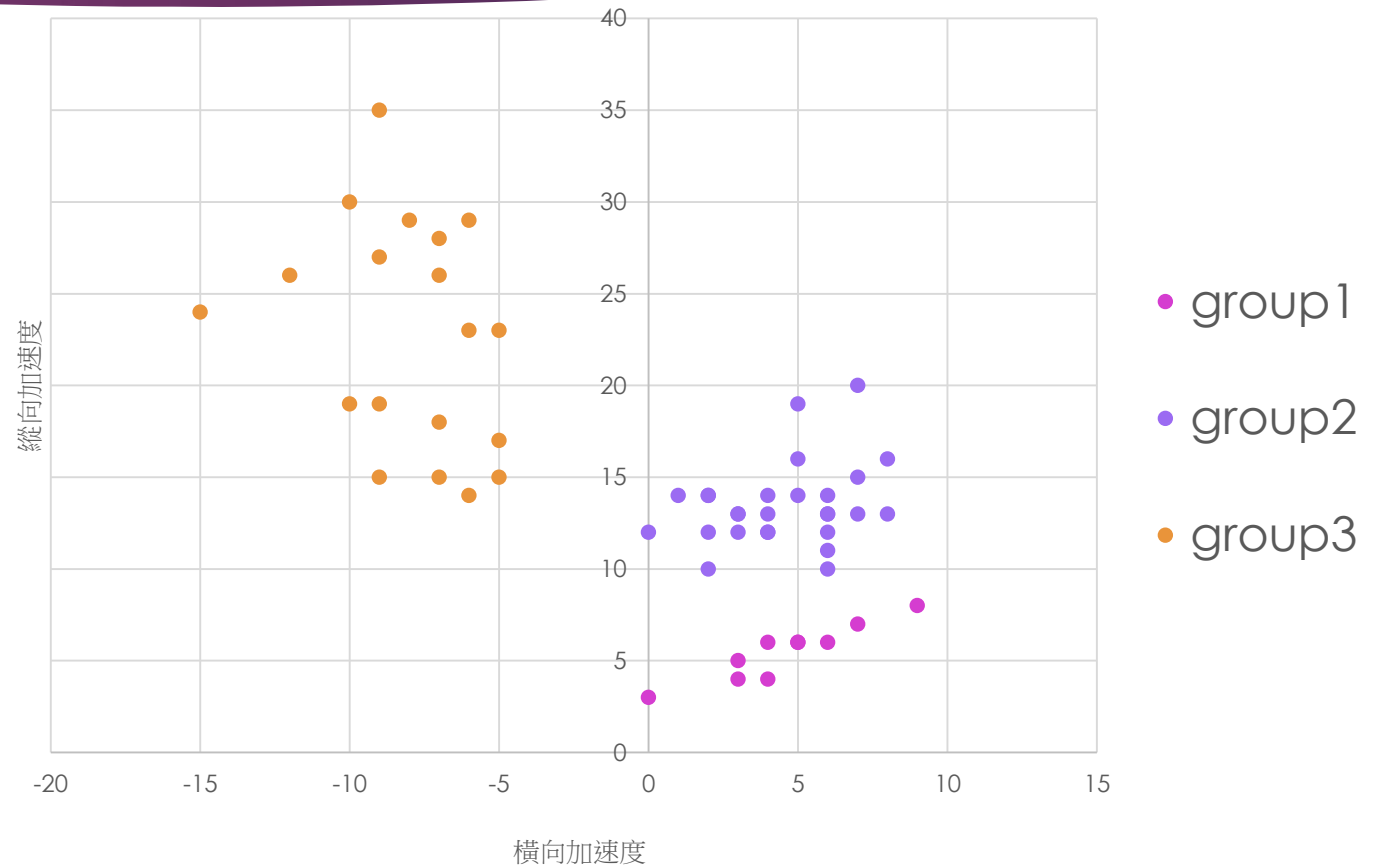
# 大量球種數據





# K-MEANS(K=3)分群

- ▶ 可分出不同軌跡特性變化球
- ▶ Group1:Fastball
- ▶ Group2:Sinkers,Splitter,Changeup
- ▶ Group3:Curveball,Slider

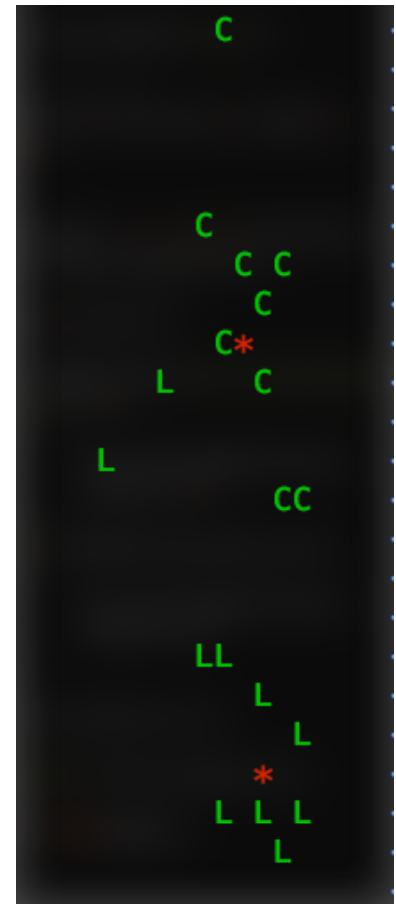


# GROUP2分群

- ▶ Group2: Sinker, Splitter, Changeup
- ▶ 如何區分?
- ▶ Ans : 加入球速做為辨別依據
- ▶ 球速依據MLB 2016 球種資料庫統計作為評判標準
- ▶ Sinker: 91.3MPH
- ▶ Splitter: 84.5MPH
- ▶ Changeup: 83.6MPH

# GROUP3再分群

- ▶ Group3:Curveball,Slider
- ▶ 如何區分?
- ▶ Ans: 針對Group3做K-means(K=2)分群



# 球種辨識流程圖

