

全國大專校院數位人文大數據學生競賽

德州撲克行為模式數據分析 以及牌組關聯性分析

台科大工管系大四：廖宣雅
台科大工管系大四：吳睿得
台科大工管系大四：廖祐萱





職業玩家，不可能一直都是靠運氣來贏得勝利
試著靠統計、資料分析冷靜地來做決策
順便贏得豐沛的賭金！

資料來源：

德州撲克，贏的人每局打牌狀況 (朋友以美金100購買)

```
PokerStars Hand #188892583550: Hold'em No Limit ($0.10/$0.25 USD) - 2018/07/19 23:55:25 ET
Table 'Aase' 6-max Seat #5 is the button
Seat 1: JamesCrossyy ($29.87 in chips)
Seat 2: ChikoKz740 ($26.70 in chips)
Seat 3: Donkey84Kong ($26.53 in chips)
Seat 4: takerunner ($7.90 in chips)
Seat 5: Tálas ($25.23 in chips)
Seat 6: INeedJP ($25.19 in chips)
INeedJP: posts small blind $0.10
JamesCrossyy: posts big blind $0.25
*** HOLE CARDS ***
ChikoKz740: folds
Donkey84Kong: folds
takerunner: folds
Tálas: folds
INeedJP: raises $0.50 to $0.75
JamesCrossyy: calls $0.50
*** FLOP *** [2h Jc 8c]
INeedJP: bets $0.70
JamesCrossyy: calls $0.70
*** TURN *** [2h Jc 8c] [5h]
INeedJP: checks
JamesCrossyy: bets $1.83
INeedJP: folds
Uncalled bet ($1.83) returned to JamesCrossyy
JamesCrossyy collected $2.77 from pot
JamesCrossyy: doesn't show hand
*** SUMMARY ***
Total pot $2.90 | Rake $0.13
Board [2h Jc 8c 5h]
Seat 1: JamesCrossyy (big blind) collected ($2.77)
Seat 2: ChikoKz740 folded before Flop (didn't bet)
Seat 3: Donkey84Kong folded before Flop (didn't bet)
Seat 4: takerunner folded before Flop (didn't bet)
Seat 5: Tálas (button) folded before Flop (didn't bet)
Seat 6: INeedJP (small blind) folded on the Turn
```



問題：

玩牌無法理性分析，想要科學化、
更可靠的依據來決策打牌
(不想輸錢啊！)

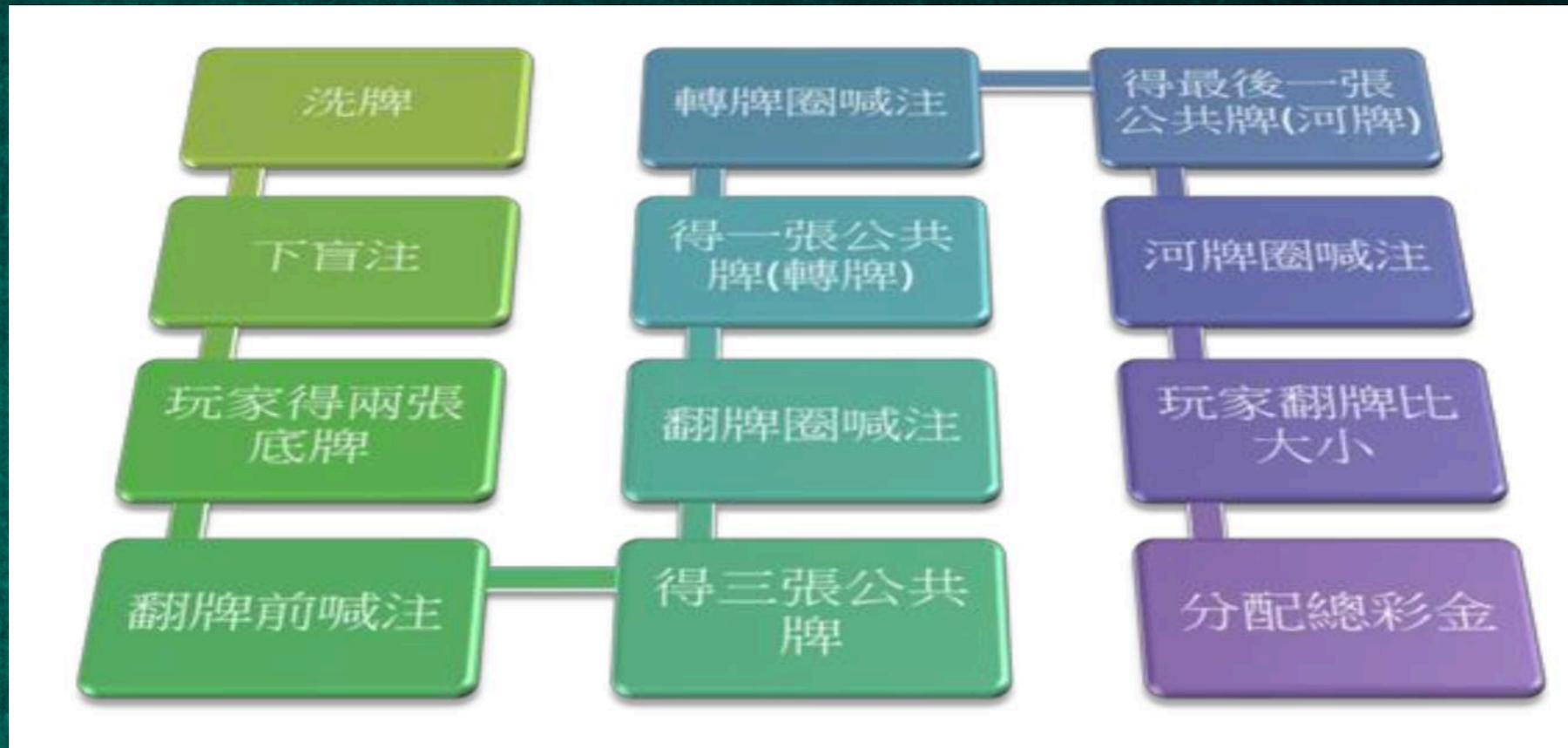
目標：

在玩牌的時候，看對手做的動作
以及其本身個性，可以推測出其
牌組強度

>>這次專案以翻牌圈為觀察目標

(設想假設，每個人他有固定的出牌習慣，好牌時會怎麼打/壞牌怎麼打)

遊戲流程：



大小：



專業術語：

下注 (Bet)

一回合下注中，首次押下的籌碼。玩家下注後，對手下更高金額的注稱為「加注」

跟注 (Call)

一位玩家下注或加注後，另一位玩家於底池投入相同額度的籌碼，稱為「跟注」。

加注 (Raise)

在對手下注之後下更高金額的注。

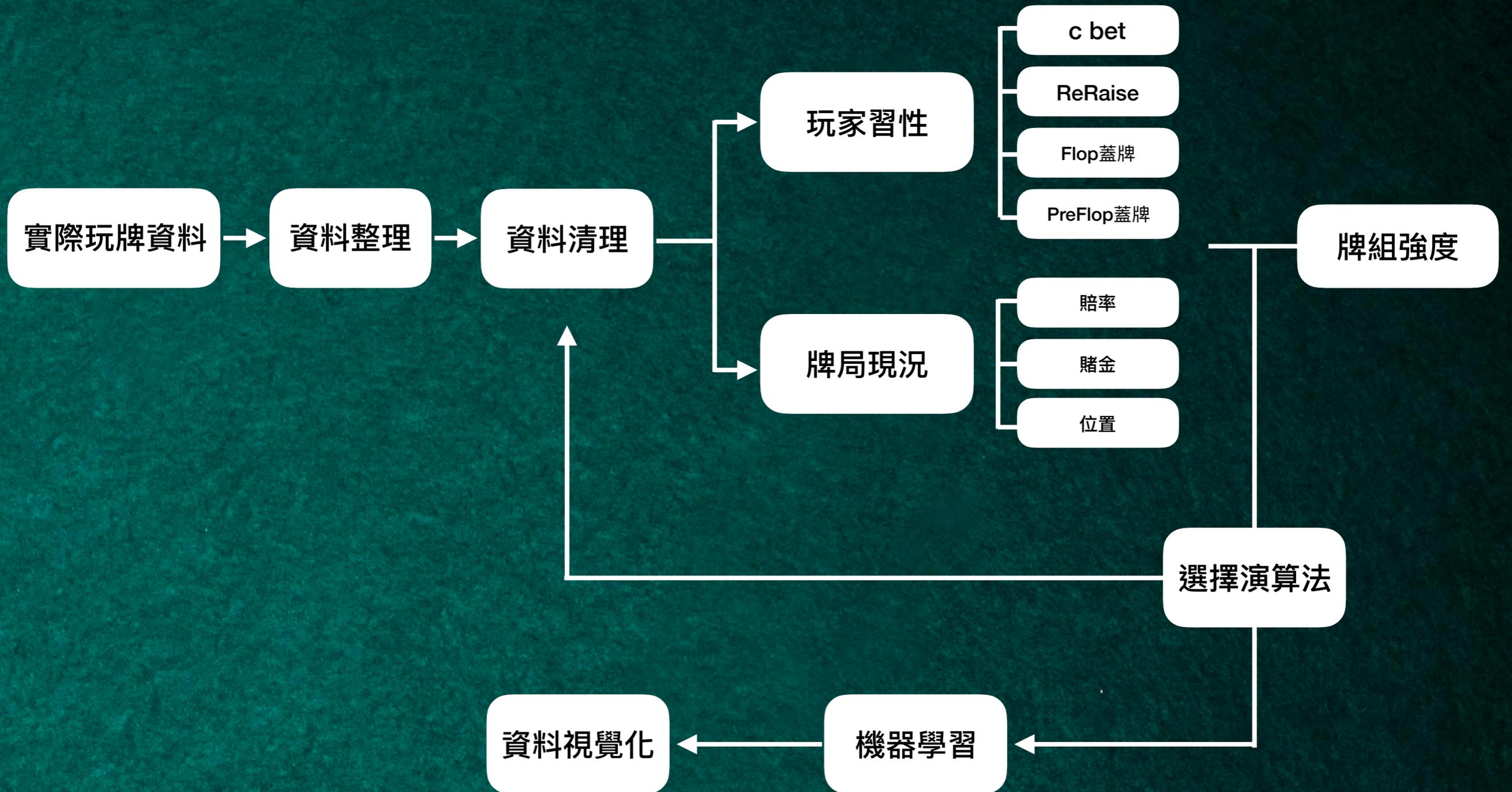
過牌 (Check)

如果前面的玩家皆未下注，玩家可以選擇「過牌」，也就是保留權利先不下注，讓下一位玩家行動。過牌的玩家可以在同一回合下注中稍後再跟注或加注。

蓋牌 (Fold)

放棄手中的牌，退出牌局，稱為「蓋牌」。蓋牌的玩家放棄贏得底池的權利。

架構流程： f (牌局現況 + 玩家習性) = 牌組強度



資料的選取：

有打到最終圈，贏最多場遊戲的前十位玩家
測試量：660mb

```
*** HOLE CARDS ***
Ladybird1367: folds
Syed32: calls $0.25
DyxSpirit: raises $0.75 to $1
kelele45: folds
zatoplus: folds
StingnWo: folds
Syed32: calls $0.75
*** FLOP *** [8d 8c 7h]
Syed32: checks
DyxSpirit: checks
*** TURN *** [8d 8c 7h] [6h]
Syed32: checks
DyxSpirit: bets $1
Syed32: calls $1
*** RIVER *** [8d 8c 7h 6h] [4c]
Syed32: checks
DyxSpirit: checks
*** SHOW DOWN ***
Syed32: shows [Th Jd] (a pair of Eights)
DyxSpirit: shows [Qc 4d] (two pair, Eights and Fours)
DyxSpirit collected $4.15 from pot
*** SUMMARY ***
Total pot $4.35 | Rake $0.20
Board [8d 8c 7h 6h 4c]
Seat 1: StingnWo (big blind) folded before Flop
Seat 2: Ladybird1367 folded before Flop (didn't bet)
Seat 3: Syed32 showed [Th Jd] and lost with a pair of Eights
Seat 4: DyxSpirit showed [Qc 4d] and won ($4.15) with two pair, Eights and Fours
Seat 5: kelele45 (button) folded before Flop (didn't bet)
Seat 6: zatoplus (small blind) folded before Flop
```

一、玩家習性：

激進型

這種類型的玩家玩的手牌種類很廣，即使獲勝率不夠的情況下，也容易 call 或 Raise，自身有牌時 Raise 的機率很高，同時也伴隨著高比例的 Bluff。

保守型

玩家玩的手牌較緊，但由於相對少的情況下，可預測性就變的較高。當你加注於他時，他們總是十分謹慎，棄牌率高。相對地，如果是他們主動加注，擁有大牌之比例便相當高。

高手型

結合上述，取各型菁華，不時調整自己的模式。根據對手改變，讓大家捉摸不定，沉著冷靜，觀察力強。

一、玩家習性：

>>收集data，計算代表玩家的屬性指標

PreFlop蓋牌率, Flop蓋牌率, Flop中ReRaise率, C-Bet in Flop率

- PreFlop蓋牌率

在preflop階段fold次數

有玩preflop

- C-Bet in Flop率

在 Preflop 是加注者，
在 flop 又持續下注。

- Flop蓋牌率

在flop階段fold次數

有玩flop

- Flop中ReRaise率

當是下注方時，被再次
加注的機率。(中大牌機
率低，太常進行此行為，
代表詐唬機率高。)

二、牌局現況：

依照我們的data，看玩家那局所做的動作及賠率

- 公牌強度
- 翻牌前底池
- 動作
- 投注金額
- 前一局投注金額
- 給別人賠率
- 別人給的賠率
- Preflop aggression
- 座位

三、找出3+2時的牌組強度



HS.flop : 由holdem套件算出的牌組強度(手牌+公牌)

BS.flop : 由holdem套件算出的公牌強度

str_draw : 買順，從公牌組成順子的可能性

lu_draw : 買花，從公牌組成同花的可能性

rank.flop : 牌組等級，0 高牌 1 一對 2 兩對 3三條 4
順子 5 同花 6葫蘆 7同花順 8皇家同花順

skl.rank : 由David Sklansky 提出的起手牌推薦表所
對應的等級



四、（玩家動作＋屬性→牌組強度）分析

（A玩家動作＋屬性→牌組強度）data 分析

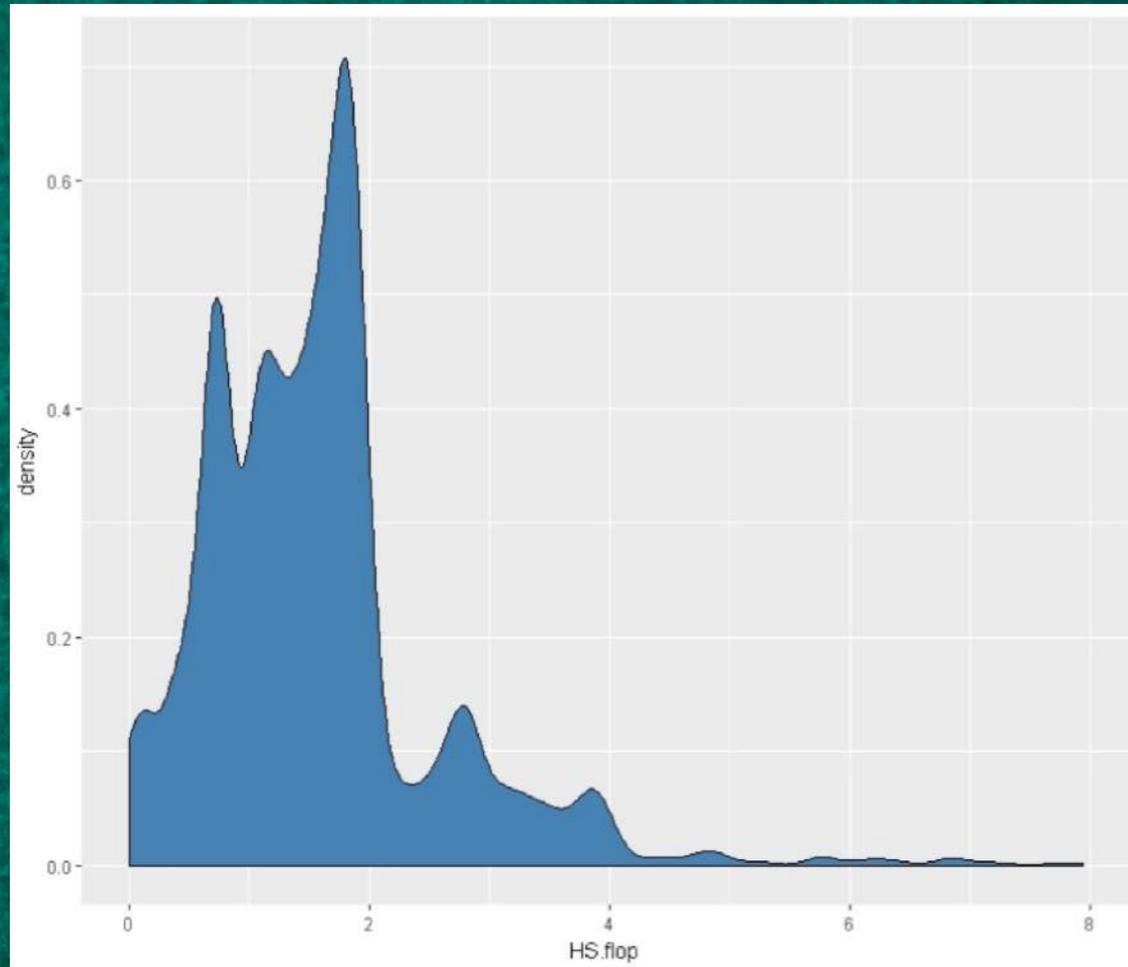
（B玩家動作＋屬性→牌組強度）data 分析

（C玩家動作＋屬性→牌組強度）data 分析

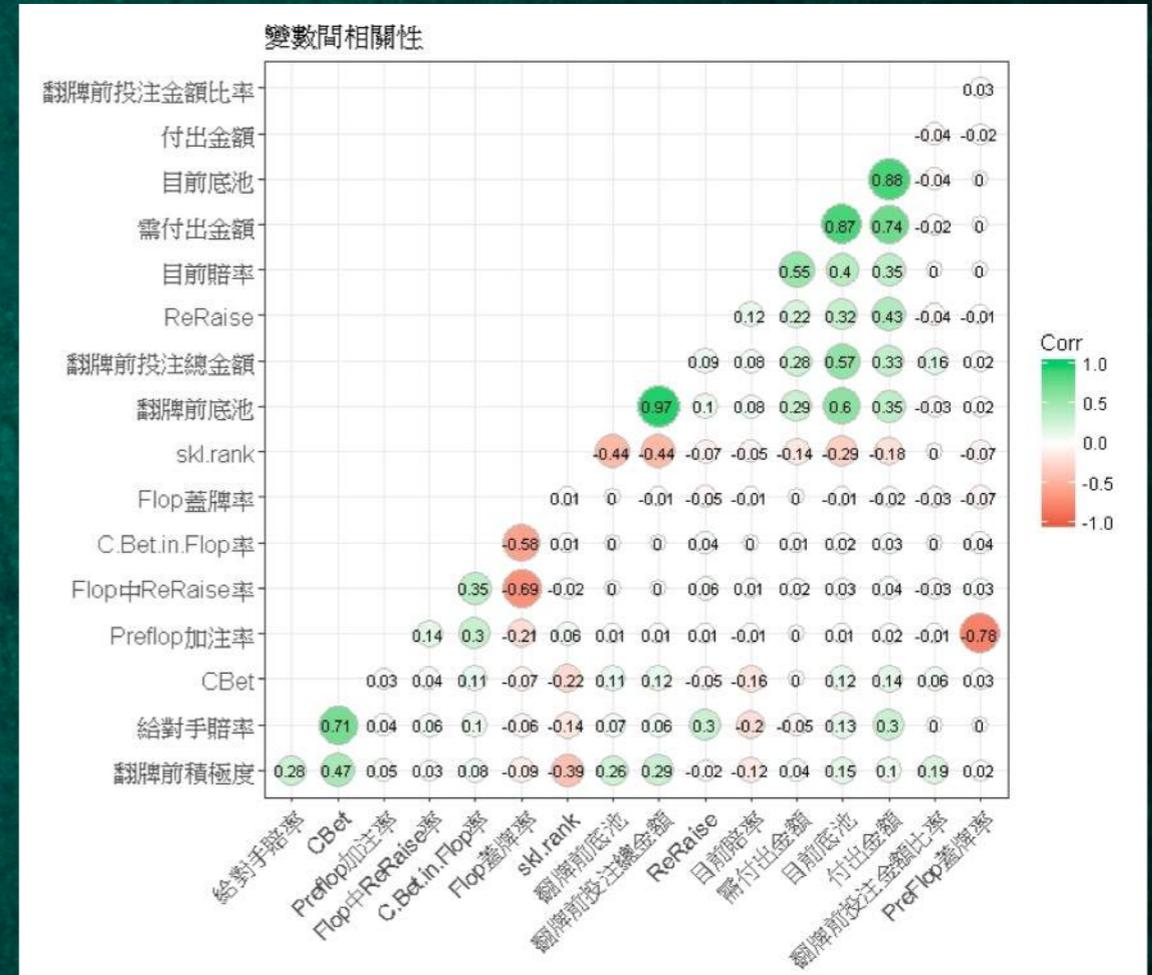
·
·

五、以結果論，我們找出各類玩家動作與牌組強度的關係、規則

結果：

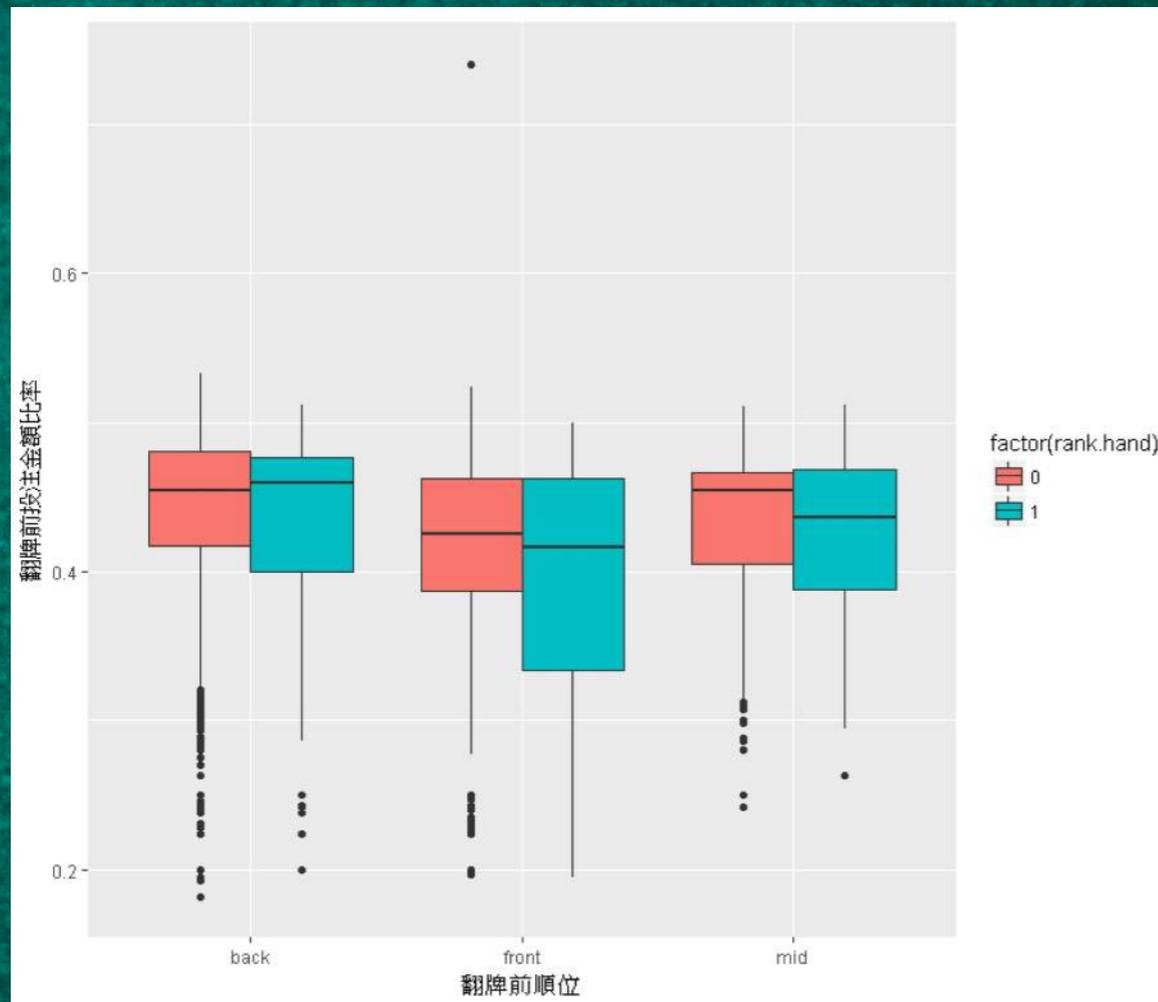


手牌強度密度圖

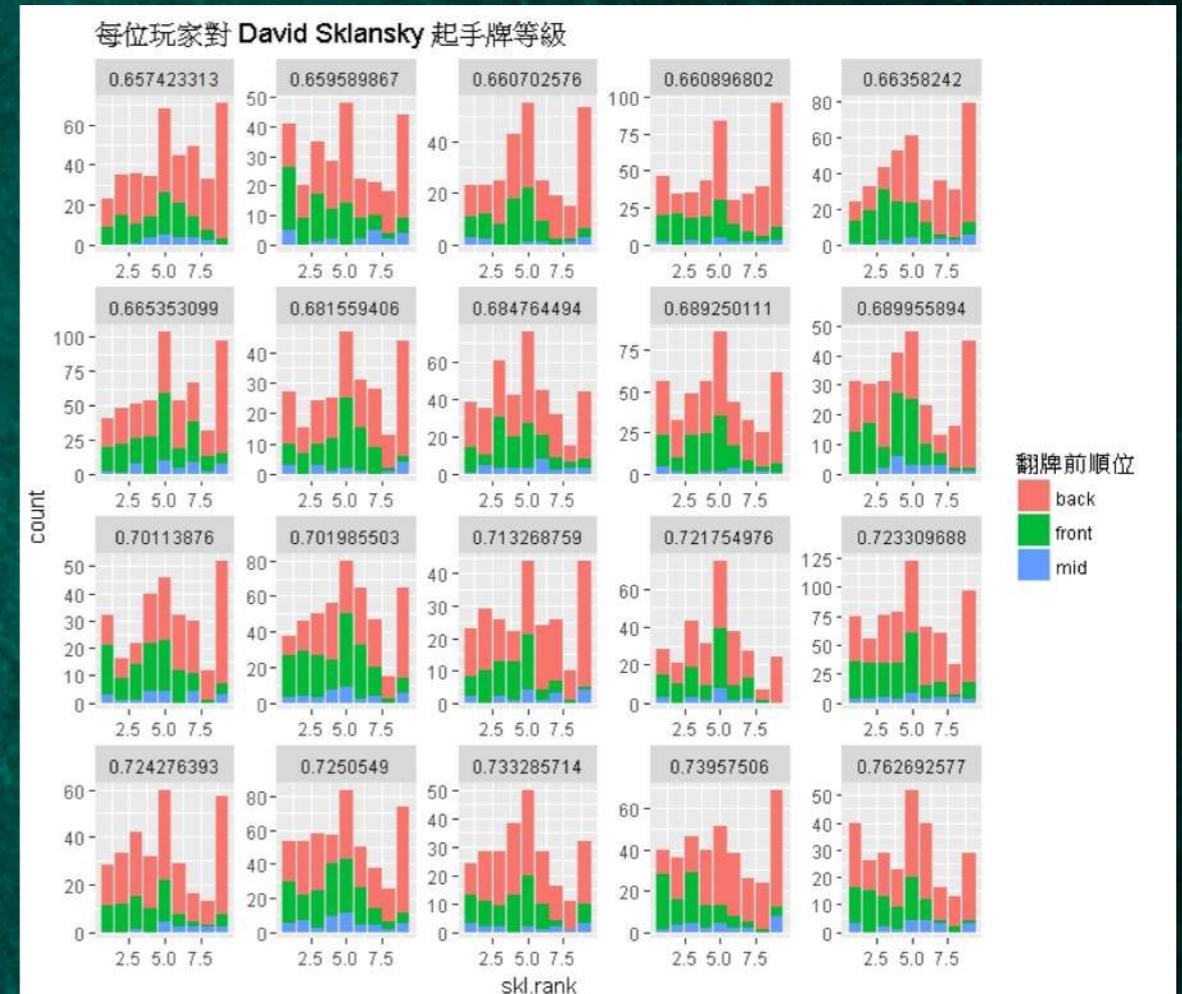


各變數關聯性分析圖

結果：

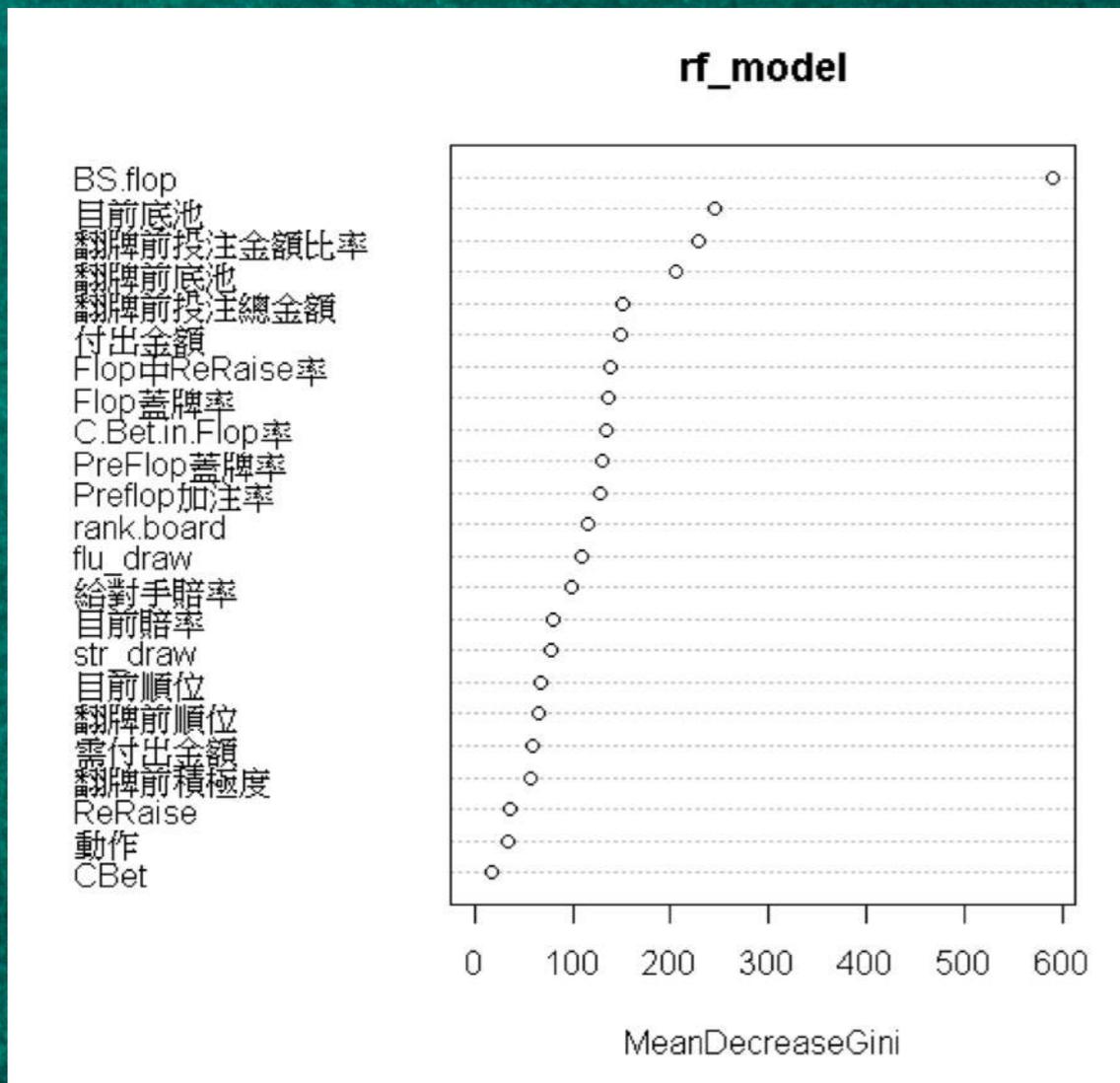


順位與起手牌對翻牌前
投注金額比例盒狀圖

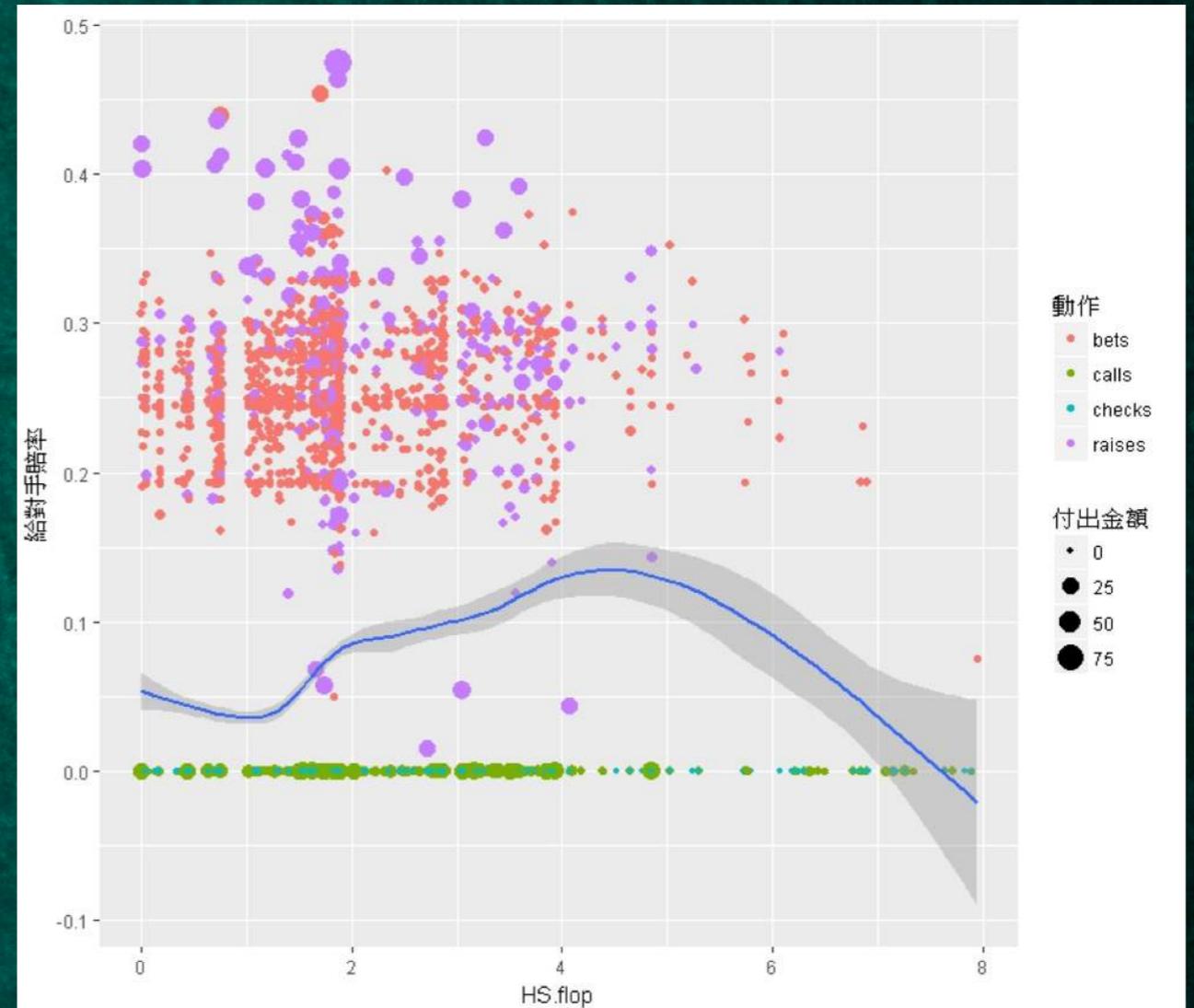


每位玩家依照蓋牌率對David
Sklansky起手牌等級長條圖

結果：



資料內各變數重要度



翻牌時玩家動作對手牌強度之關係圖

結果：

```

                Sample size: 5129
    Frequency of class labels: 1298, 2874, 513, 342, 46, 22, 23, 10, 1
                Number of trees: 1000
    Forest terminal node size: 5
    Average no. of terminal nodes: 806.777
No. of variables tried at each split: 5
                Total no. of variables: 23
    Resampling used to grow trees: swr
    Resample size used to grow trees: 5129
                Analysis: RF-C
                Family: class
                Splitting rule: gini *random*
    Number of random split points: 10
                Normalized brier score: 57.5
                AUC: 83.94
                Error rate: 0.39, 0.64, 0.17, 0.53, 0.82, 0.98, 1, 1, 1, 1

Confusion matrix:

      predicted
observed  0  1  2  3  4  5  6  7  8  class.error
0  469  826  0  3  0  0  0  0  0  0.6387
1  348 2375 135 16 0  0  0  0  0  0.1736
2   6  245 243 19 0  0  0  0  0  0.5263
3   9  211  61 61 0  0  0  0  0  0.8216
4   3   38  0  4  1  0  0  0  0  0.9783
5   1   20  0  0  1  0  0  0  0  1.0000
6   0    7 12  4  0  0  0  0  0  1.0000
7   0    4  6  0  0  0  0  0  0  1.0000
8   0    1  0  0  0  0  0  0  0  1.0000

Overall error rate: 38.6%
```

訓練資料內預測結果

結果：

```
Sample size of test (predict) data: 2198
      Number of grow trees: 1000
Average no. of grow terminal nodes: 806.777
      Total no. of grow variables: 23
      Resampling used to grow trees: swr
      Resample size used to grow trees: 5129
      Analysis: RF-C
      Family: class
      Test set Normalized brier score: 56.71
      Test set AUC: 83.91
      Test set error rate: 0.38, 0.66, 0.17, 0.48, 0.85, 0.95, 1, 1, 1, NA
```

Confusion matrix:

	predicted									
observed	0	1	2	3	4	5	6	7	8	class.error
0	187	365	0	0	0	0	0	0	0	0.6612
1	128	1053	78	4	0	0	0	0	0	0.1663
2	0	95	110	5	0	0	0	0	0	0.4762
3	7	84	21	19	0	0	0	0	0	0.8550
4	1	16	0	2	1	0	0	0	0	0.9500
5	0	6	0	0	0	0	0	0	0	1.0000
6	0	3	9	0	0	0	0	0	0	1.0000
7	0	2	2	0	0	0	0	0	0	1.0000
8	0	0	0	0	0	0	0	0	0	NaN

overall error rate: 37.67%

驗證資料預測結果

創新的價值：

我們嘗試建構一個模型，探討德州撲克玩家習性、其動作與當下牌局狀況，進行分析並預測。

包括當你手牌強度為4(三條)以上結果基本上是穩贏的以及順位偏後者，投注金額比率也較多，也利用David Sklansky 提出的起手牌推薦表所對應等級為基礎做出的手牌強度變數與局面狀況的各種負相關，並且深入探討此項變數，同時也運用手邊data的詳細資訊加上三總運算演算法來完成最佳預測結果。

創新的價值：

我們目前程式只寫到第二個階段（翻牌面）的數據分析與預測，但其實下幾個階段也是依樣畫葫蘆而已，預測的準確度也超過了50%，一般而言玩家在每個階段給的賠率很少超過一個33%，所以依照我們測試的結果的期望值是正的！

我們未來打算加上更多的樣本數做分析，再把轉牌跟河牌的程式完成，之後做成人工AI也就是機器人，可以與玩家真實對戰，做出GTO賽局理論最佳化的決定！

困難與挑戰：

- 資料格式雜亂
- 真正能用的資料受到侷限
- 樣本分佈不均
- 事實上還是有很多變數要考量

參考輔助：

- 德州撲克相關論文

<https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1006&context=statsp>

- 專業玩家朋友
- 比賽影片