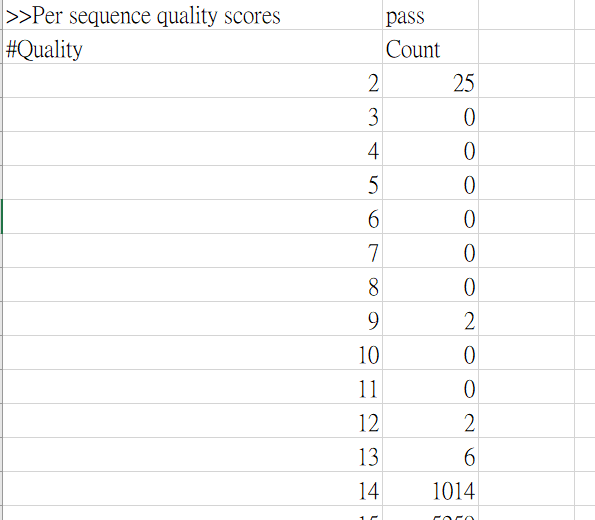
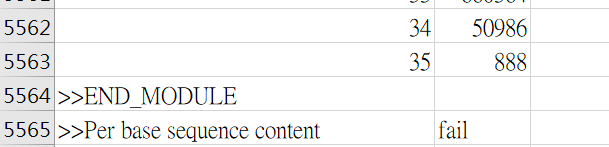
Count Q30 (data preprocessing practice for pandas and Python不過你要用其他方式解決也可以啦XD)

1. 檔案名稱叫做fastqc\_data.txt
2. 從這個檔案的>>Per sequence quality scores這行開始會看到下面這種格式



最後尾端結束在這裡



1. 目標:算出(#Quality>=30的Count)/(Count總加總)然後輸出這個結果到csv檔
2. 注意:檔案欄位數量,欄位名稱
3. Bonus:使用者提供複數個zip檔在同一資料夾下,你直接用code解好所有檔案壓縮之後算出值並輸出結果到同一個csv檔裡

這是我之前寫過的一支幫實驗室助理批次處理資料的小程式,因為在同一個資料夾下會跑出一堆zip檔,然後每支zip檔下都會有fastqc\_data.txt這個檔案,要把所有zip人肉解壓縮,開fastqc\_data.txt來算出Q30,然後統整到一張表實在是太花時間(助理手邊的是100多個,我提供給你兩個玩玩),現在可以用這支code解決,我覺得拿來練習還不錯,所以你可以寫一遍XD 上面題目沒bonus的是簡化版,有bonus的就是我當初實際寫的東西的版本,希望你玩的開心,難度太低的話再和我說XD

(算這個是做啥星期五再和你說)

quime