**廖昱瑋 311551094 HW3、4**

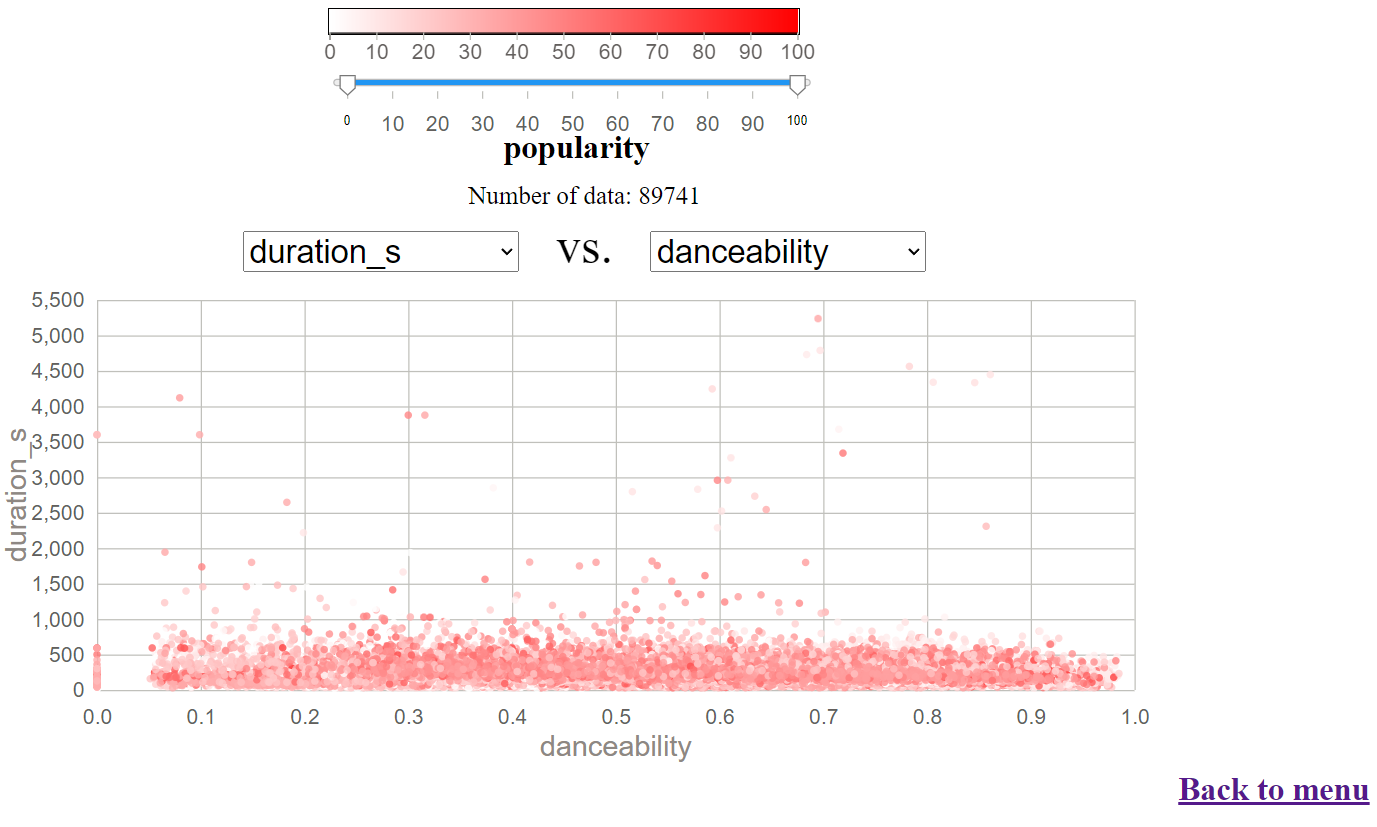
* **目標**

觀察不同popularity的track有哪些不同特性分布，以分析怎樣的歌較受市場歡迎，未來創作新歌曲時，可參考此資料迅速掌握市場口味。我以Scatter Plot及Collapsible Tree兩張圖呈現資料的視覺化。

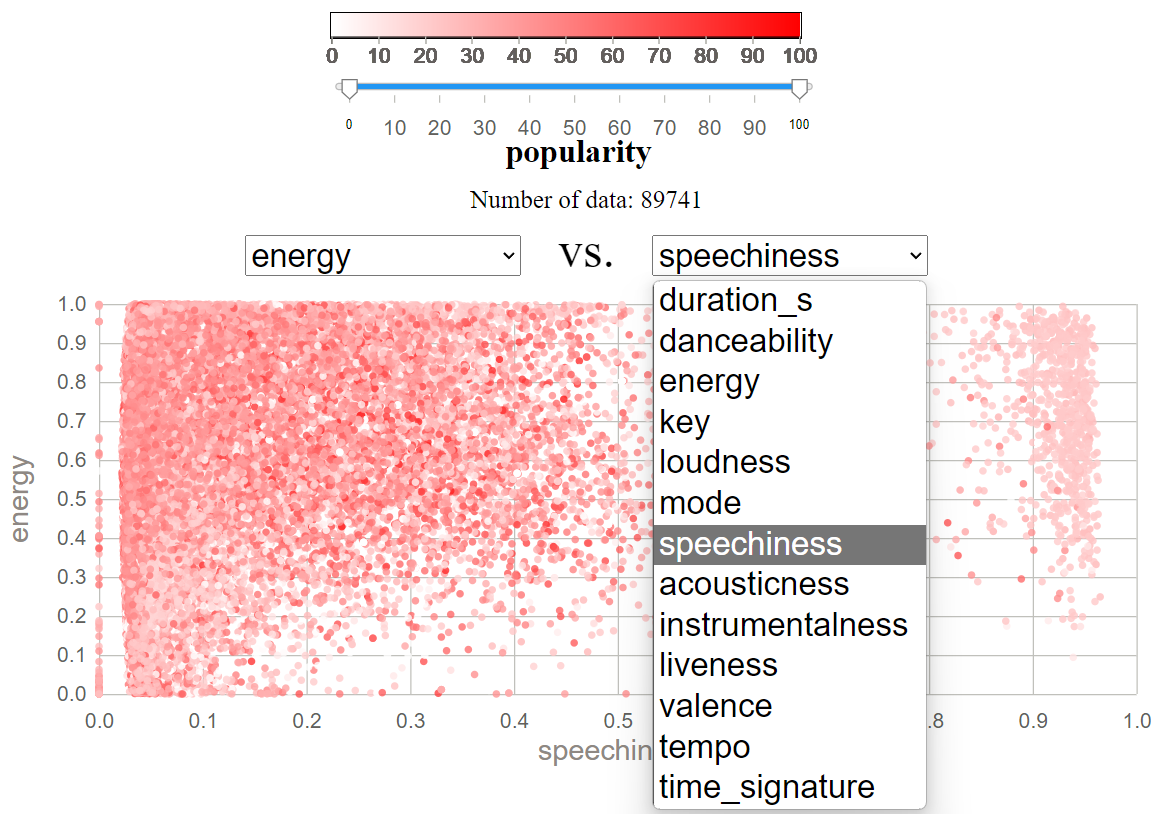
* **資料預處理**

Dataset中共有114,000筆tracks，但其中有很多是重複的，所以先過濾track\_id相同的資料，取獨有的89,741筆資料。

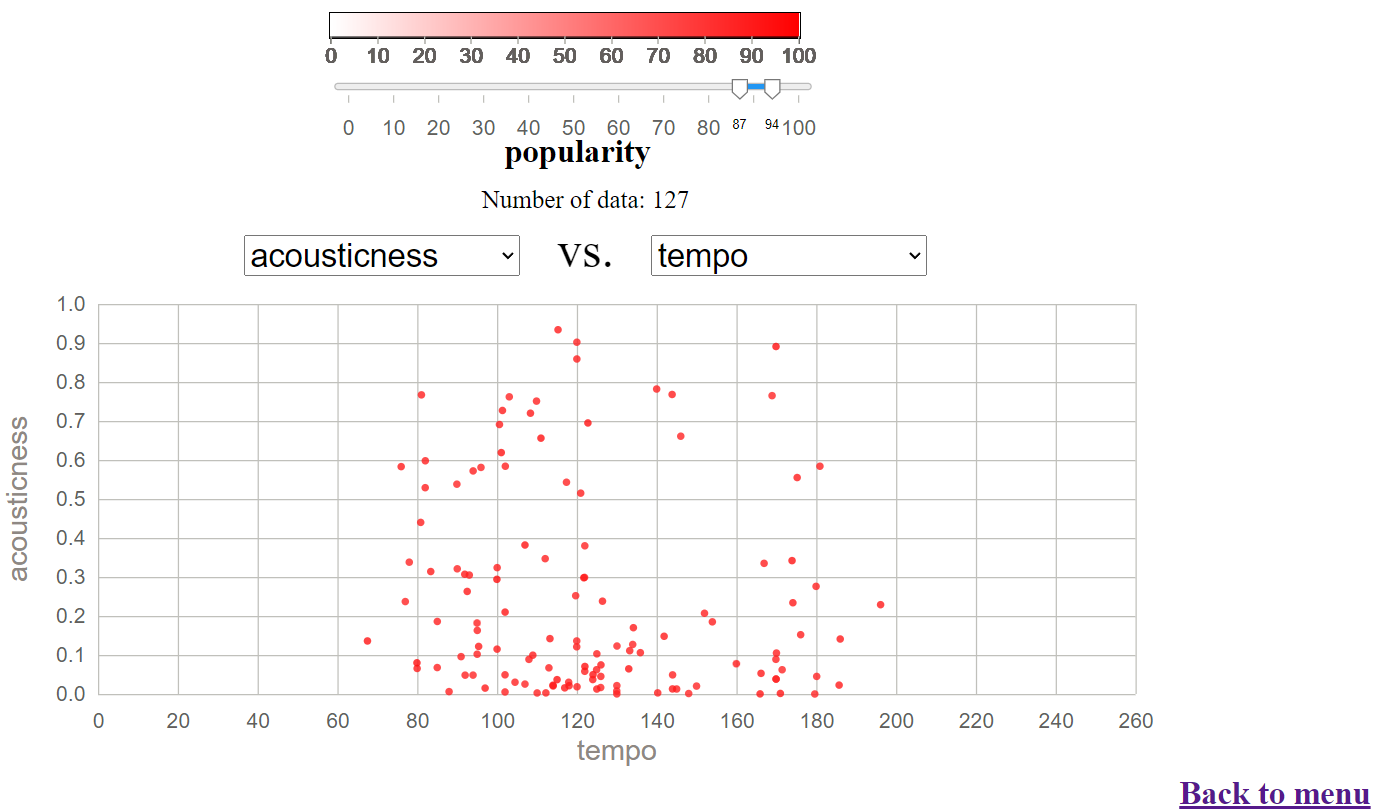
* **Scatter Plot**



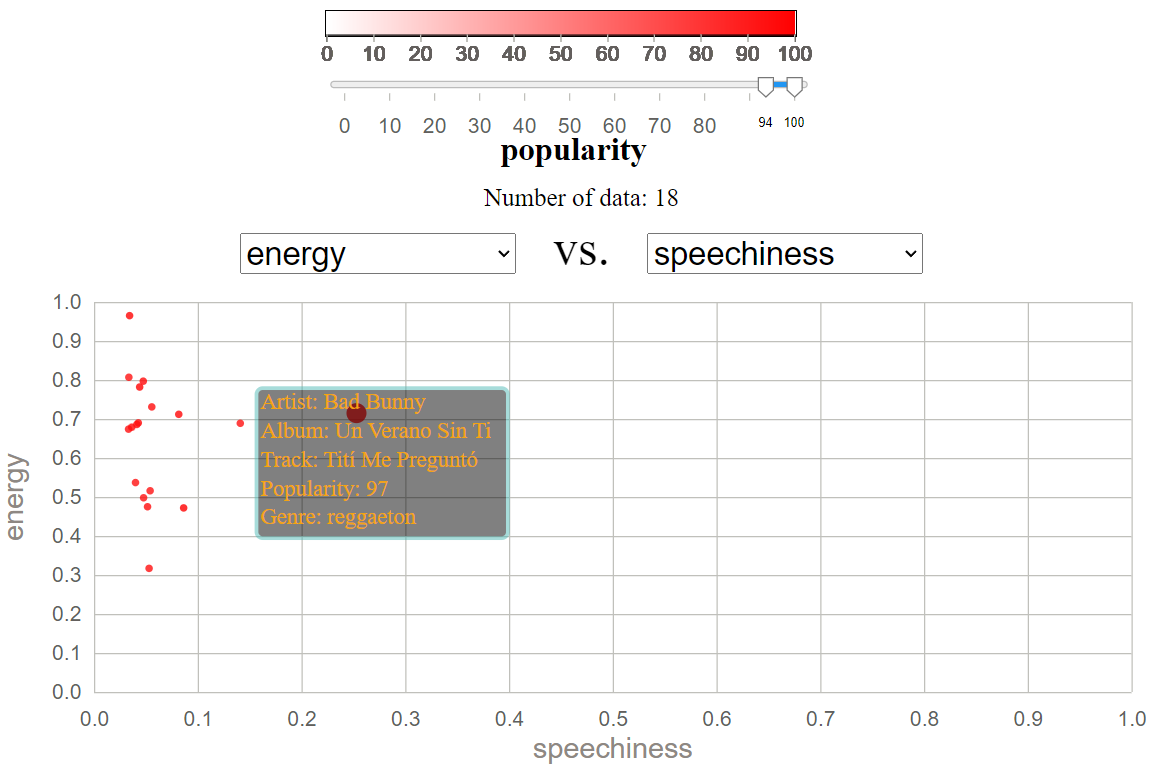
Scatter plot中每一個點代表一個track，以不同顏色深度的紅點代表該track的popularity，點越紅該track越有名。



x、y軸有下拉式選單可以分別選擇track的不同屬性，以觀察各popularity在不同屬性的分布。

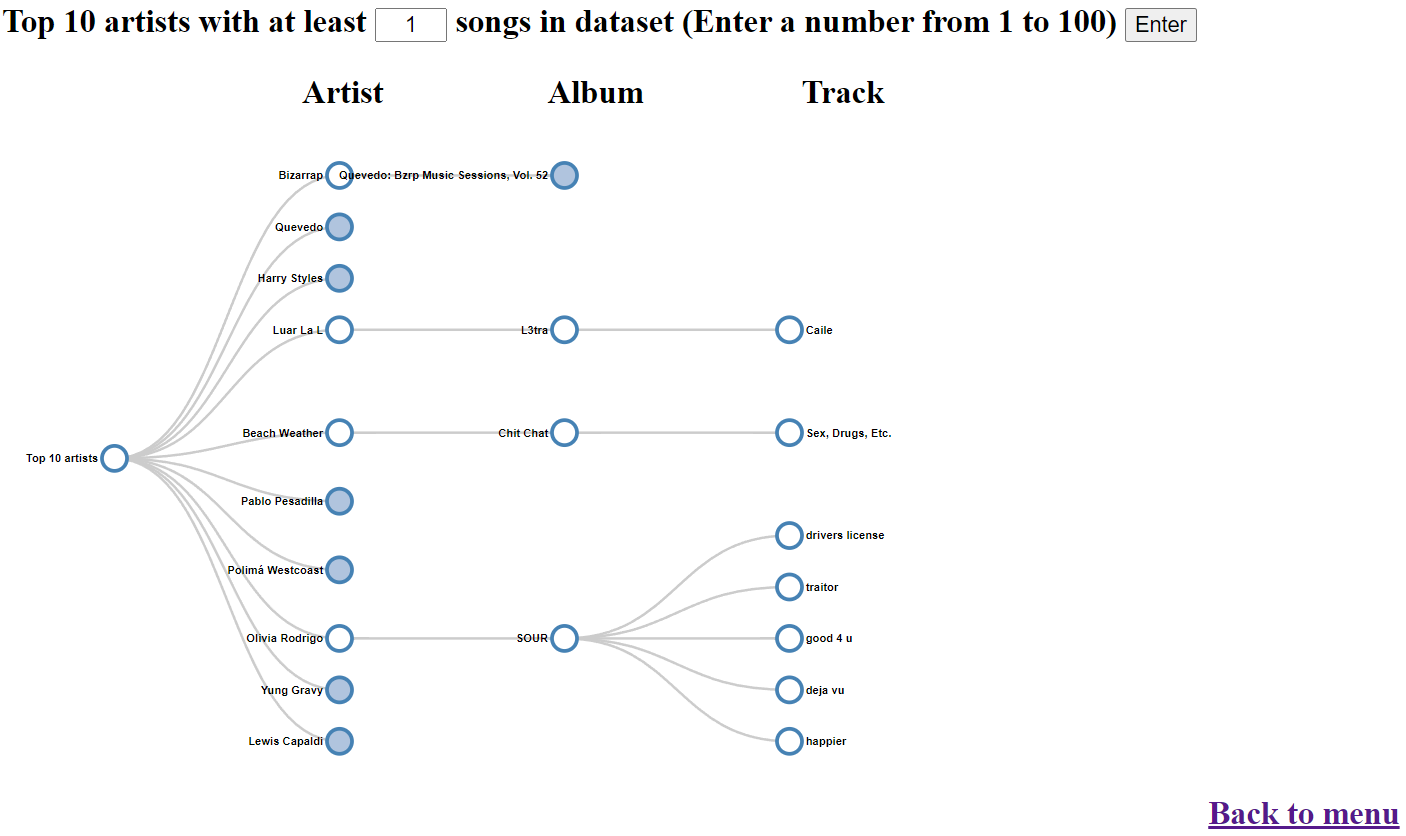


由於資料量大，圖中的點會大量重疊，不利我們觀察，所以上方我製作一個可拖曳的slider，可讓圖中只顯示特定範圍popularity的資料點。

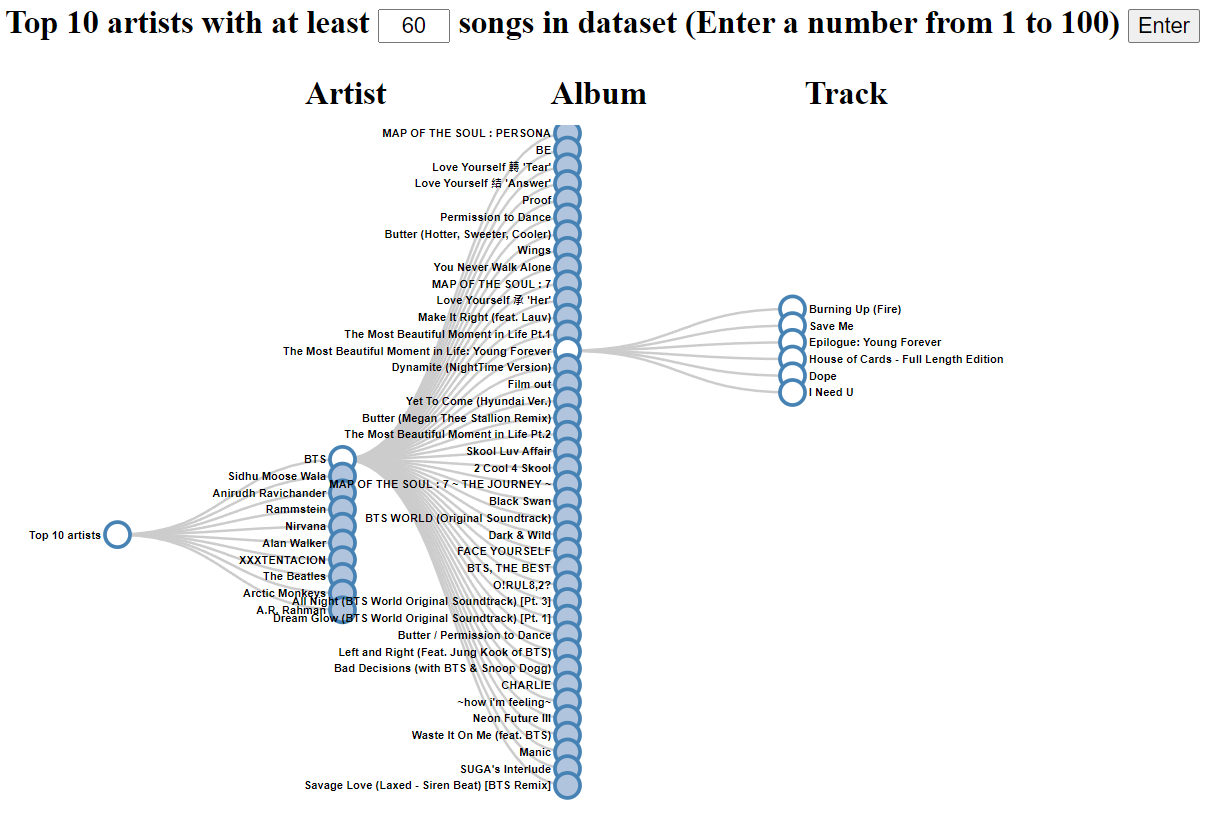


將游標放在點上，會出現該資料點的資訊。由此圖可觀察出，較受歡迎的track energy值屬於中高，speechiness則很低。

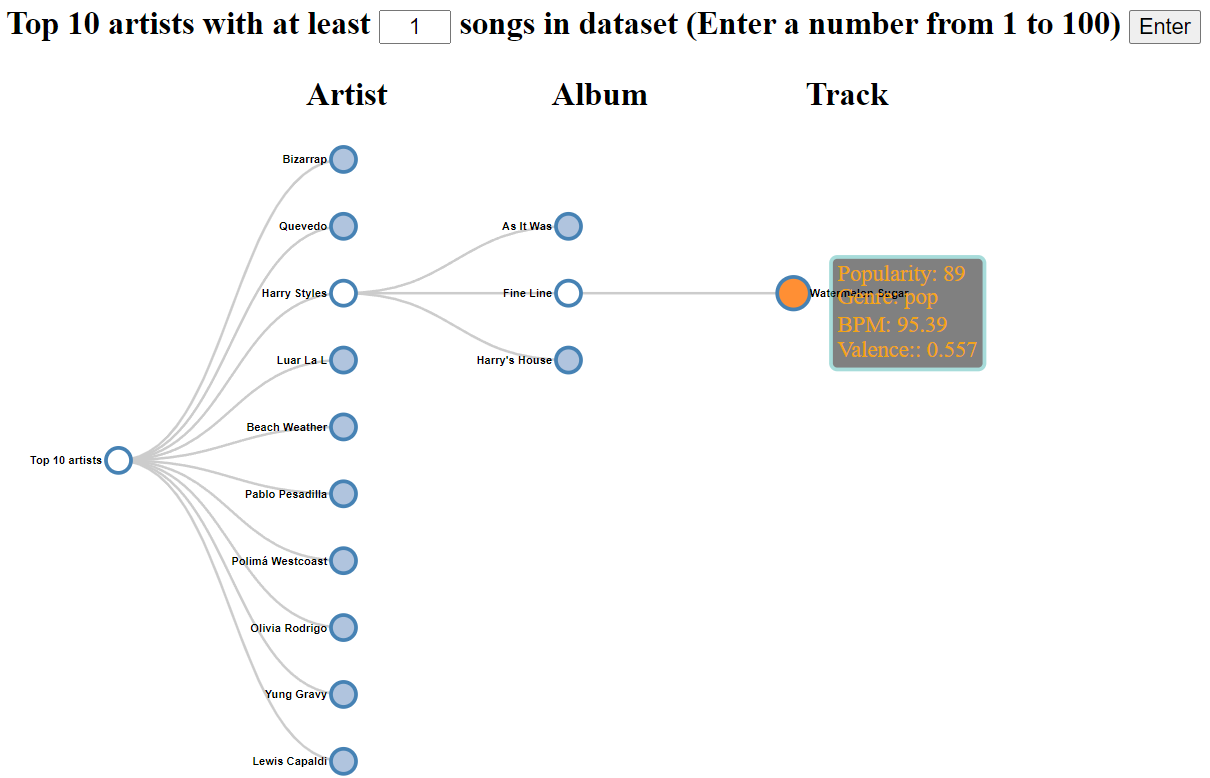
* **Collapsible Tree**



找出top 10 artists並以collapsible tree呈現他的所有作品，計算top 10 artists時，我先將每個artist的所有tracks的popularity取平均，如果一個track有多個創作者，所有其創作者都會使用到該track popularity來他們各自計算平均，再從各artist的平均popularity中取最高的前10名。除了tree中最尾端的node，每個node都可以透過點擊展開或收合，藍色實心的node為還可以展開的node，藍框空心的node為已經展開的node。



從collapsible tree的第一張圖中可以觀察出top 10 artists的創作量都很少，這並符合直覺，在這種情況計算出的top 10 artists可能只是某一首track曾經爆紅，而不是該artist本人很有名，所以我在圖的上方多加了可讓使用者輸入參數的地方，限制呈現的top 10 artists在dataset中至少要有幾個tracks。方塊中允許輸入數字1~100，並設有防呆裝置，若輸入不在這個範圍會無效。



將游標放在tree中最尾端的node，圖中會顯示該track的資訊。