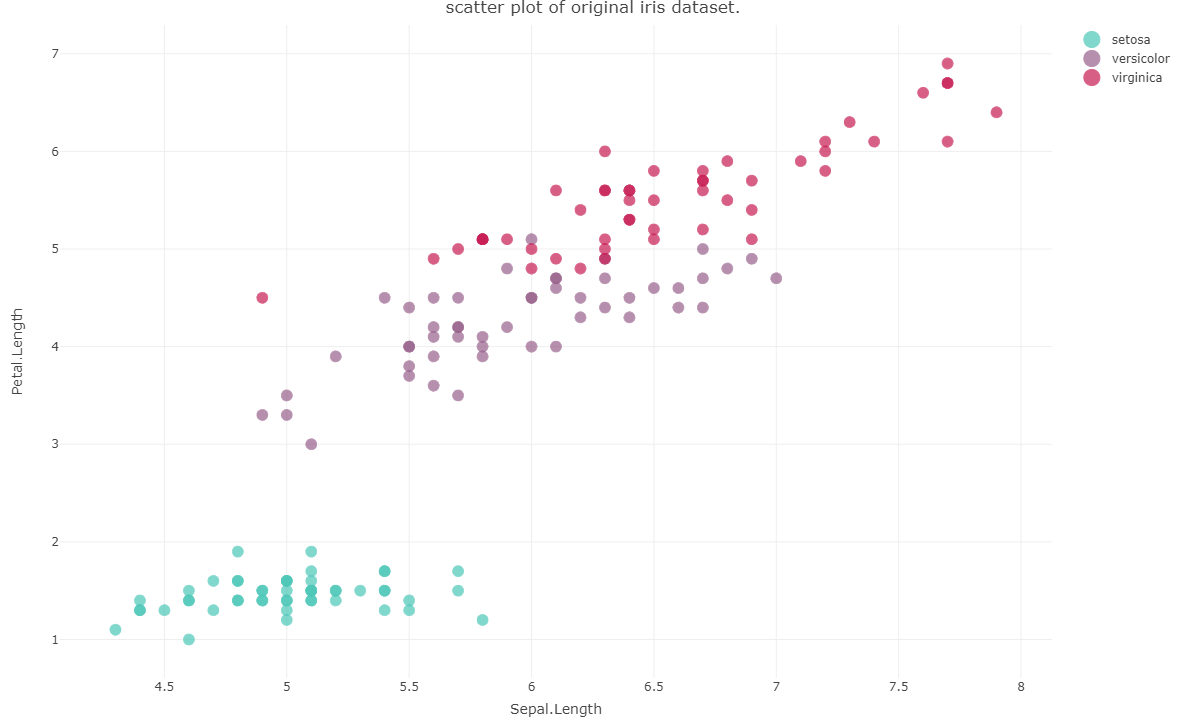
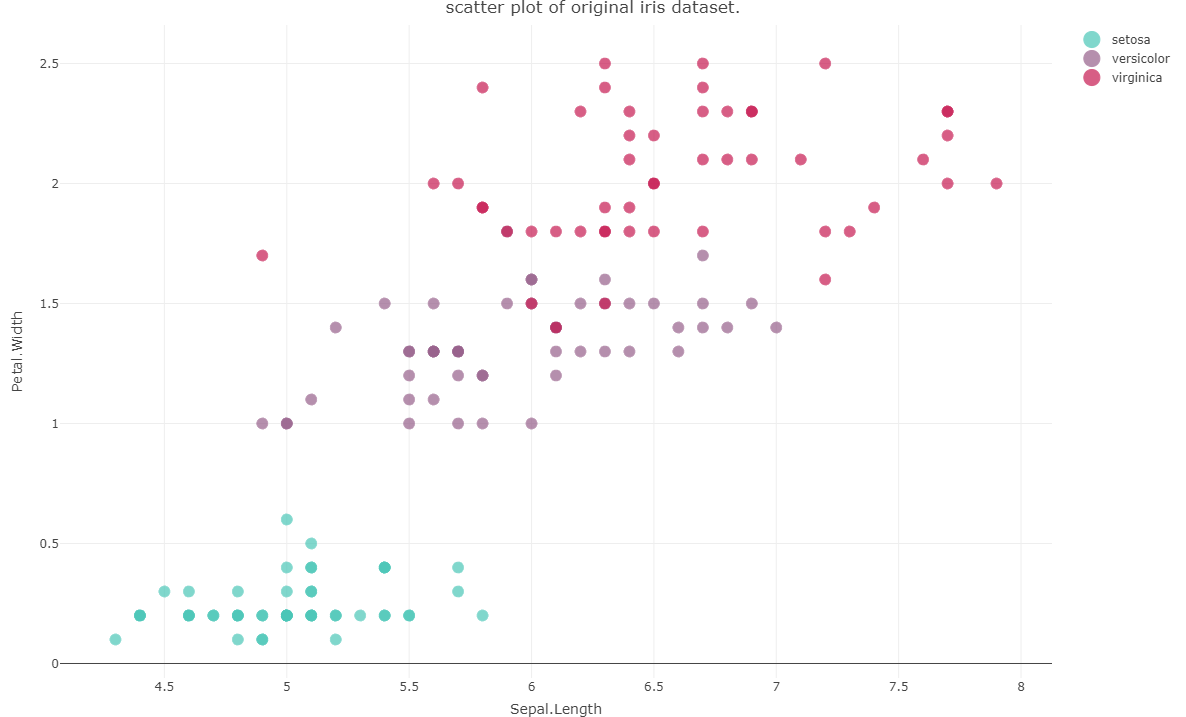
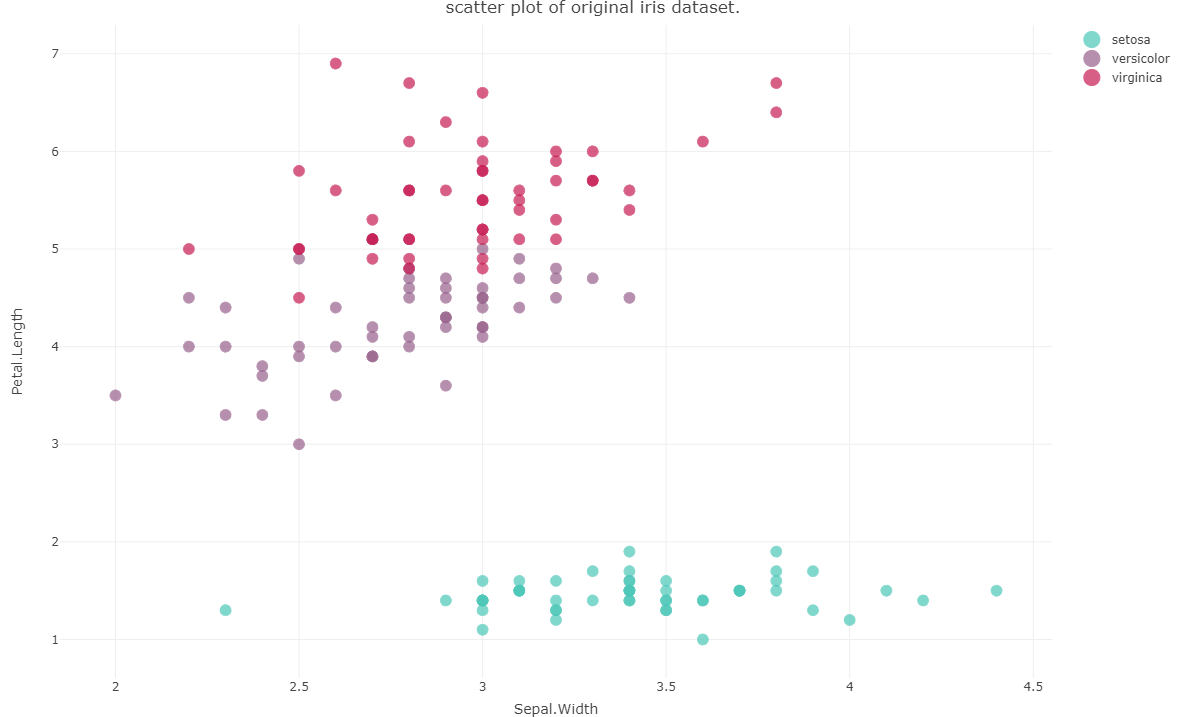
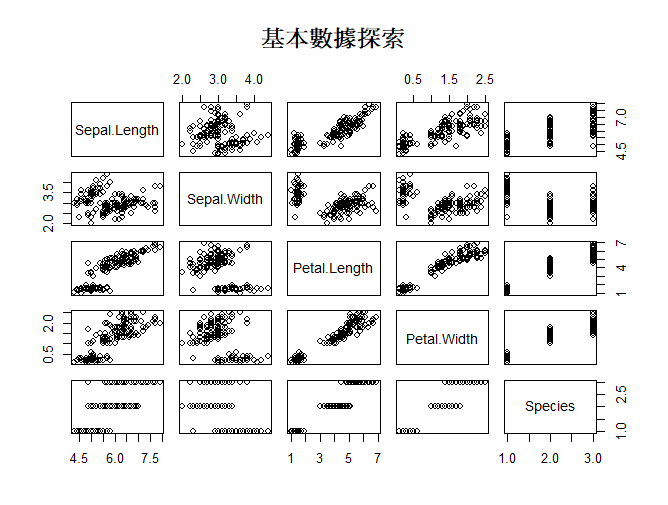
1. **基本數據探索（EDA）**：
2. 
3. 



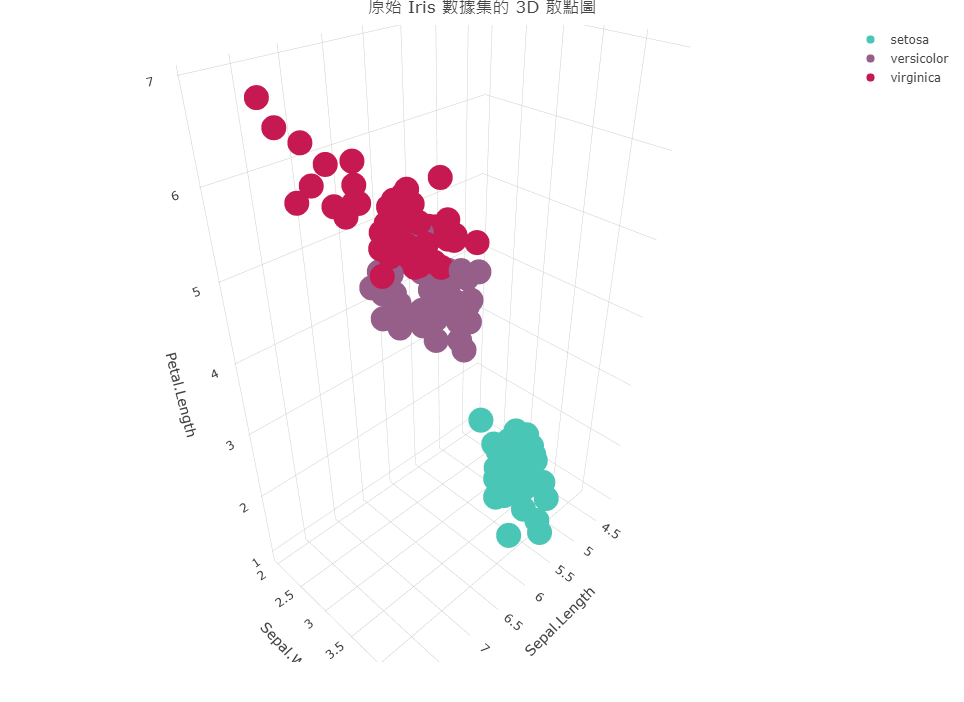
我們可先結合基本數據探索圖對2d plot圖做分析

**第1個2D散點圖（萼片長度與花瓣長度比）**：這個圖會顯示出萼片長度與花瓣長度之間的關係。從一般的觀察來看，這兩個特徵之間應該有比較強的正相關，即花瓣越長的鳶尾花，其萼片也應越長。

**第2個2D散點圖（萼片長度與花瓣寬度比）**：這張圖展示了萼片長度與花瓣寬度之間的關係，這個關係可能不如萼片長度與花瓣長度之間的關係那麼緊密。

**第3個2D散點圖（萼片寬度與花瓣長度比）**：萼片寬度與花瓣長度的關係可能會揭示特定種類鳶尾花的特點，例如某些種類的花瓣長度可能不受萼片寬度的太多影響。

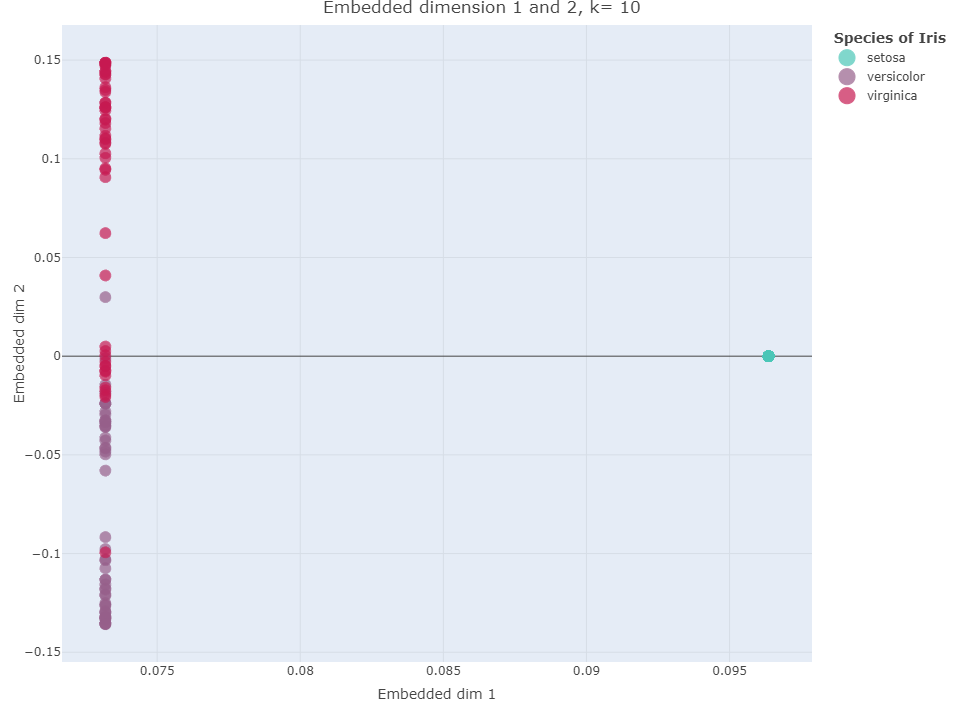
從2d散點圖中，可觀察出三種花的品種，setosa品種的特徵幾乎與另兩種花朵分離，而versicolor和virginica之間則較相近。

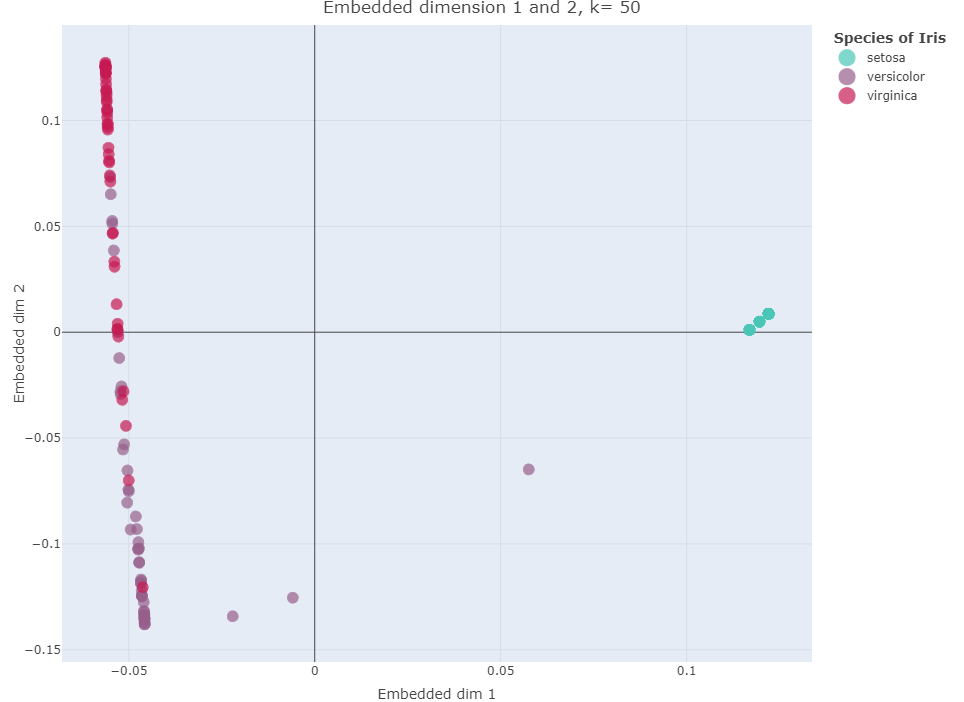


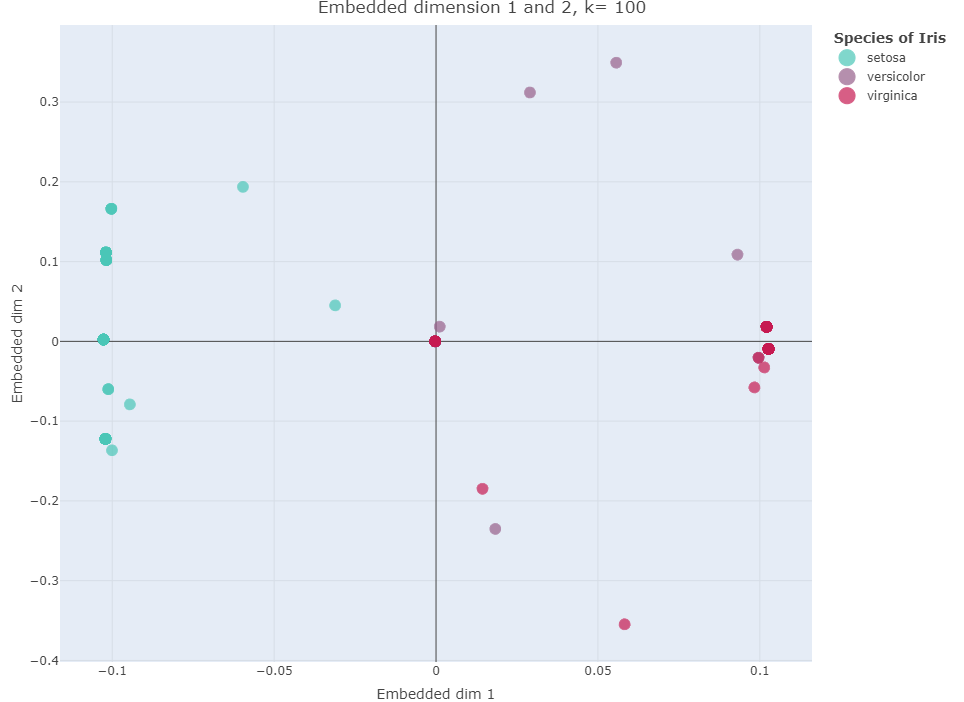
**原始數據集的3D散點圖**：

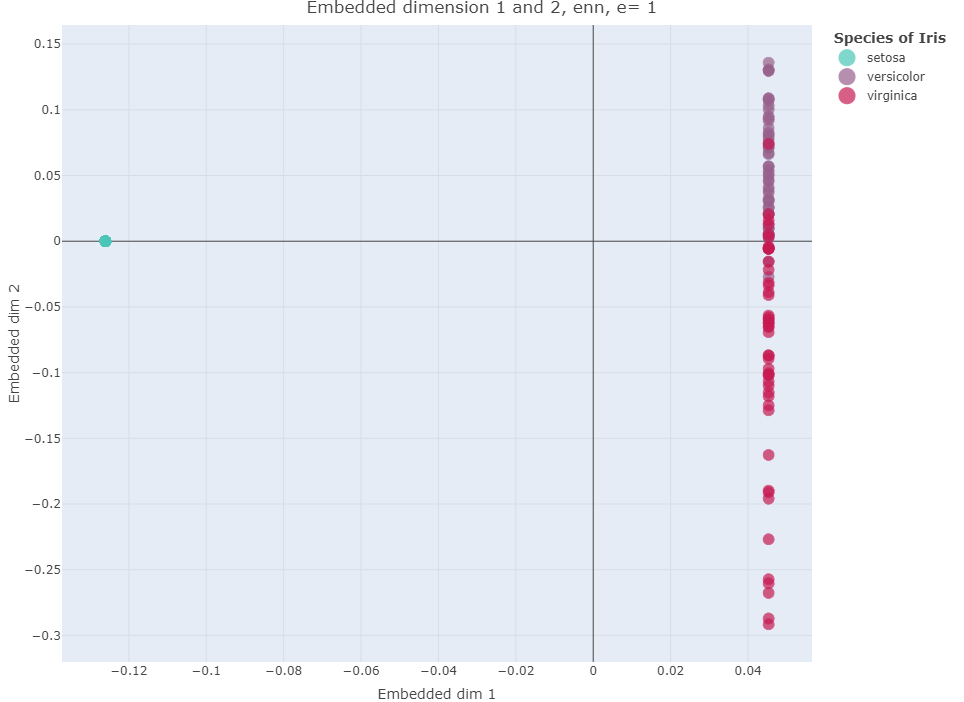
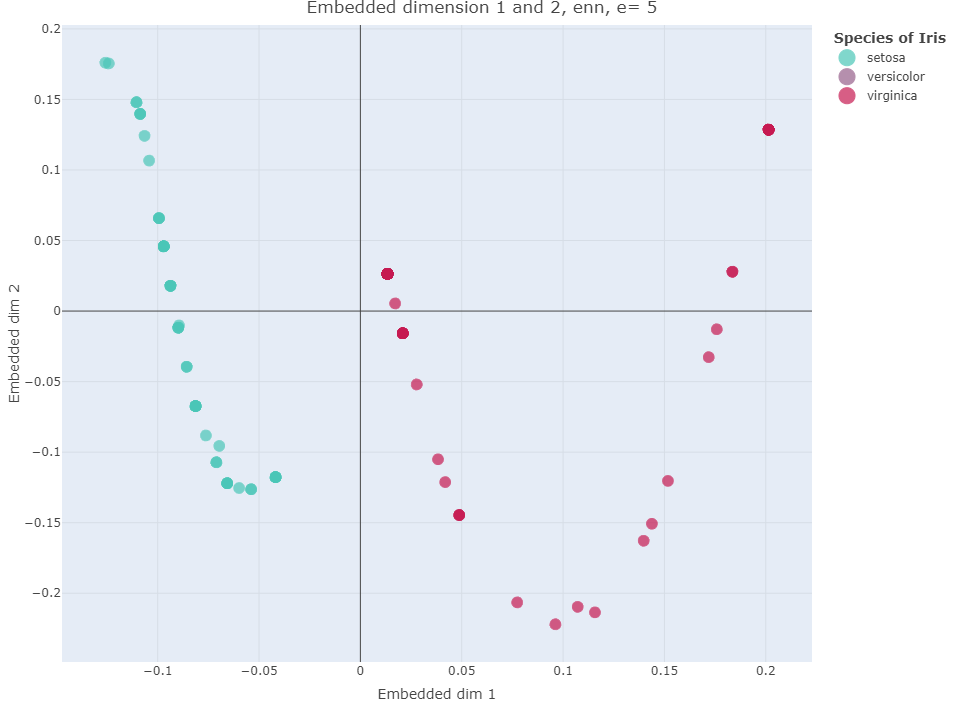
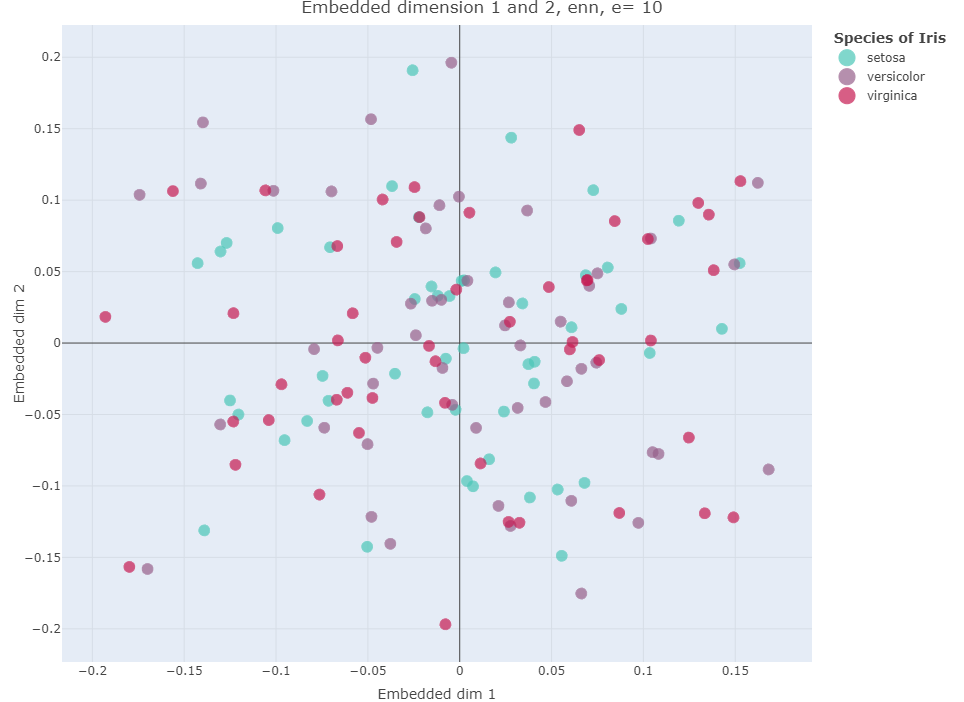
* + 第三張圖是一個3D散點圖，展示了使用前三個變量的數據分佈。也可觀察出setosa品種的特徵幾乎與另兩種花朵分離，而versicolor和virginica之間則較相近。

**對比分析：**

1. **KNN方法的對比**：





* + 圖像顯示了**k=10**, **k=50**, 和 **k=100**的KNN方法的結果。
  + 在**k=10**時，setosa品種幾乎完全分離，而versicolor和virginica之間有重疊。這表明較小的**k**值強調了數據的局部結構。
  + 隨著**k**值增加到**k=50**和**k=100**，setosa品種仍然清晰分離，但versicolor和virginica開始在空間中分散，顯示出數據的全局結構更為明顯。
  + 
  + 
  + 

1. **ENN方法的對比**：
   * 對於**e=1**, **e=5**, 和 **e=10**的ENN方法的結果，隨著**e**值的增加，數據點趨於在空間內形成更連續的區域。
   * 在**e=1**時，可以看到三個品種都相對集中在空間的不同區域，但versicolor和virginica品種之間的區分不夠明顯。
   * 當**e**增加到**e=5**，品種之間的分離開始變得清晰，特別是對setosa品種。
   * 在**e=10**時，散點圖顯示分離度並沒有提高，反而變得較為混亂，這表明對於ENN方法，過大的**e**值可能會導致過度強調全局結構，從而使得品種間的界限模糊。
   * 一張含有 文字, 螢幕擷取畫面, 繪圖, 行 的圖片

     自動產生的描述
   * 一張含有 文字, 螢幕擷取畫面, 繪圖, 行 的圖片

     自動產生的描述
   * 一張含有 文字, 螢幕擷取畫面, 繪圖, 圖表 的圖片

     自動產生的描述
   * 一張含有 文字, 螢幕擷取畫面, 繪圖, 圖表 的圖片

     自動產生的描述
   * 一張含有 文字, 螢幕擷取畫面, 圖表, 繪圖 的圖片

     自動產生的描述
   * 一張含有 螢幕擷取畫面, 文字, 圖表, 繪圖 的圖片

     自動產生的描述
2. **Proportion方法的對比**：

**Proportion=0.05** - 在這張圖中，Setosa類別（藍色點）與其他兩個類別清晰地分開。Versicolor和Virginica之間有較大的重疊。這顯示了一個非常疏的鄰域圖，導致這兩個類別難以用線性方法分開。

**Proportion=0.1** - 類似於Proportion=0.05的結果，但是這次Versicolor和Virginica之間的重疊似乎更加密集，這可能意味著鄰域圖變得稍微密集一些。

**Proportion=0.3** - 在此圖中，三個品種的分布都開始彼此分離，尤其是Versicolor和Virginica品種，它們在嵌入空間中表現出更清晰的群聚邊界。

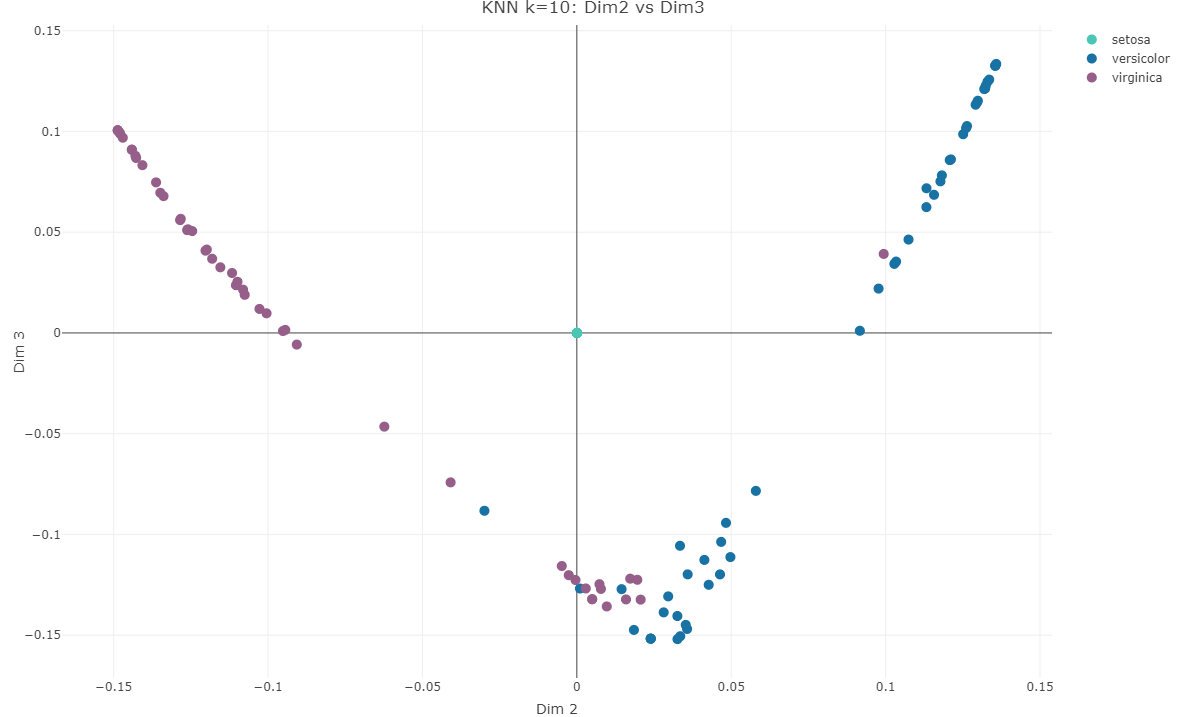
**Proportion=0.5** - 這張圖中三個品種的點分佈更加分散。Setosa品種仍然清晰地與其他兩個品種分離，而Versicolor和Virginica品種之間的界限也更為明確。

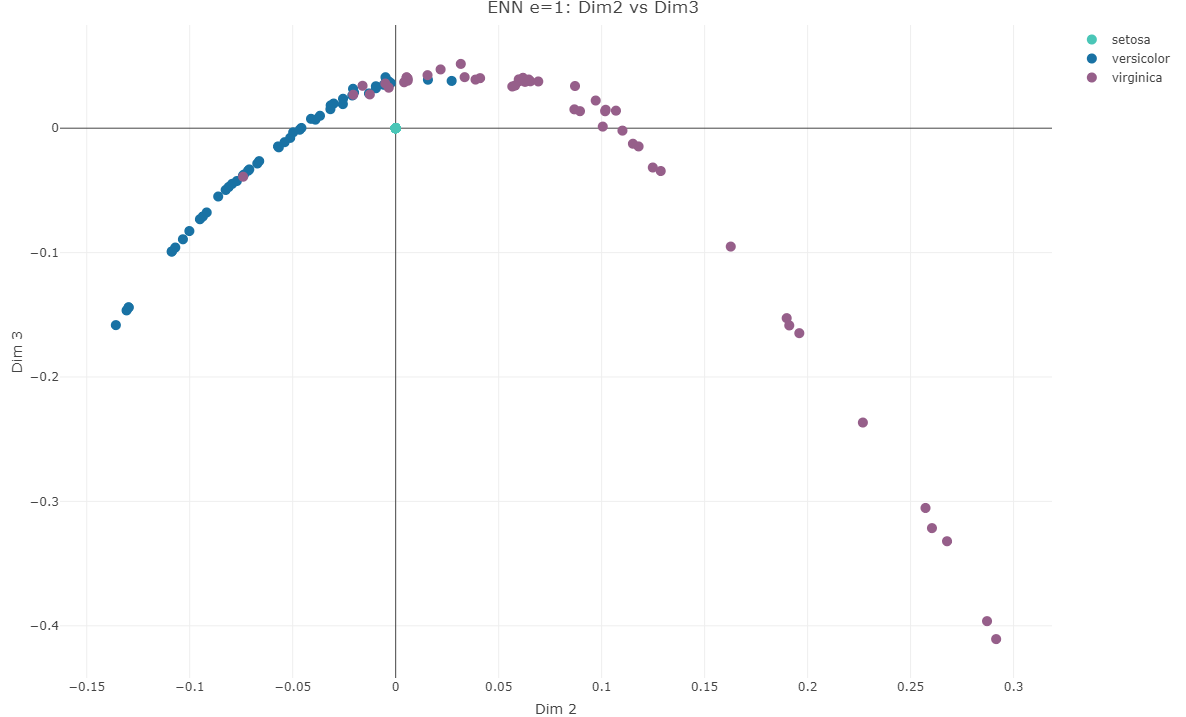
**Proportion=0.7** - 隨著鄰域圖變得更加密集，三個品種的點似乎開始重疊在一起，特別是Versicolor和Virginica，顯示出更多的重疊區域。

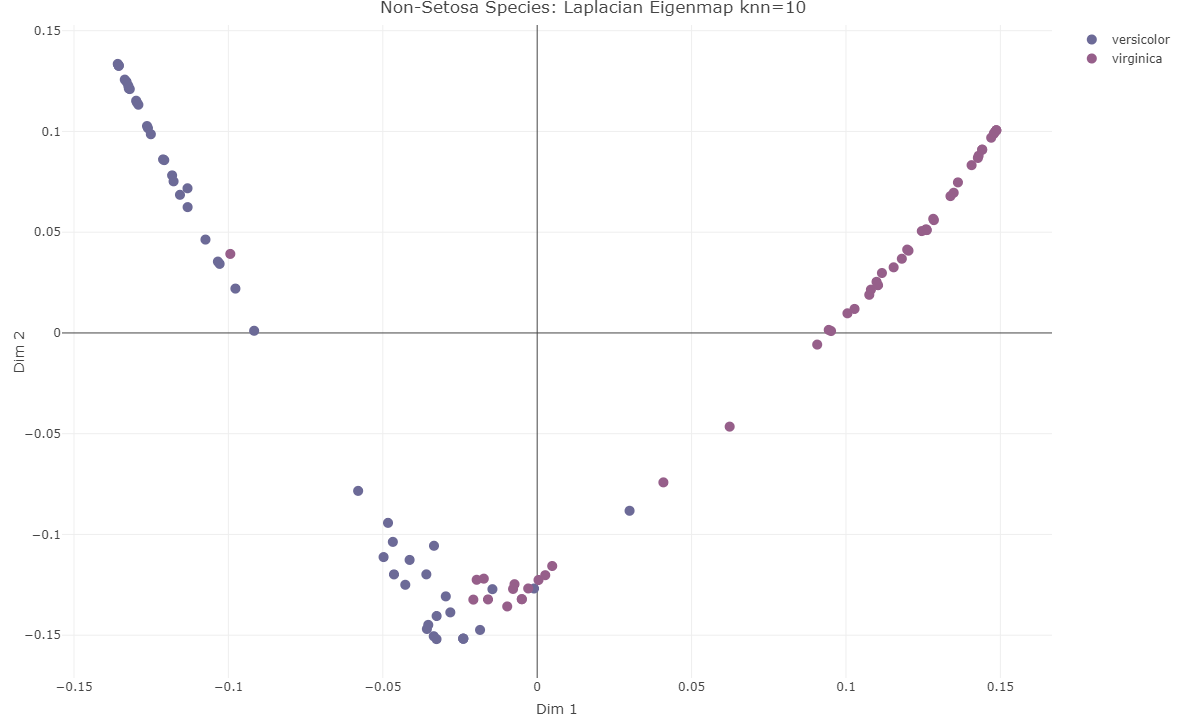
**Proportion=1** - 所有點都彼此連接，造成了一個完全密集的鄰域圖。這導致品種間界限的進一步模糊，特別是Versicolor和Virginica之間。

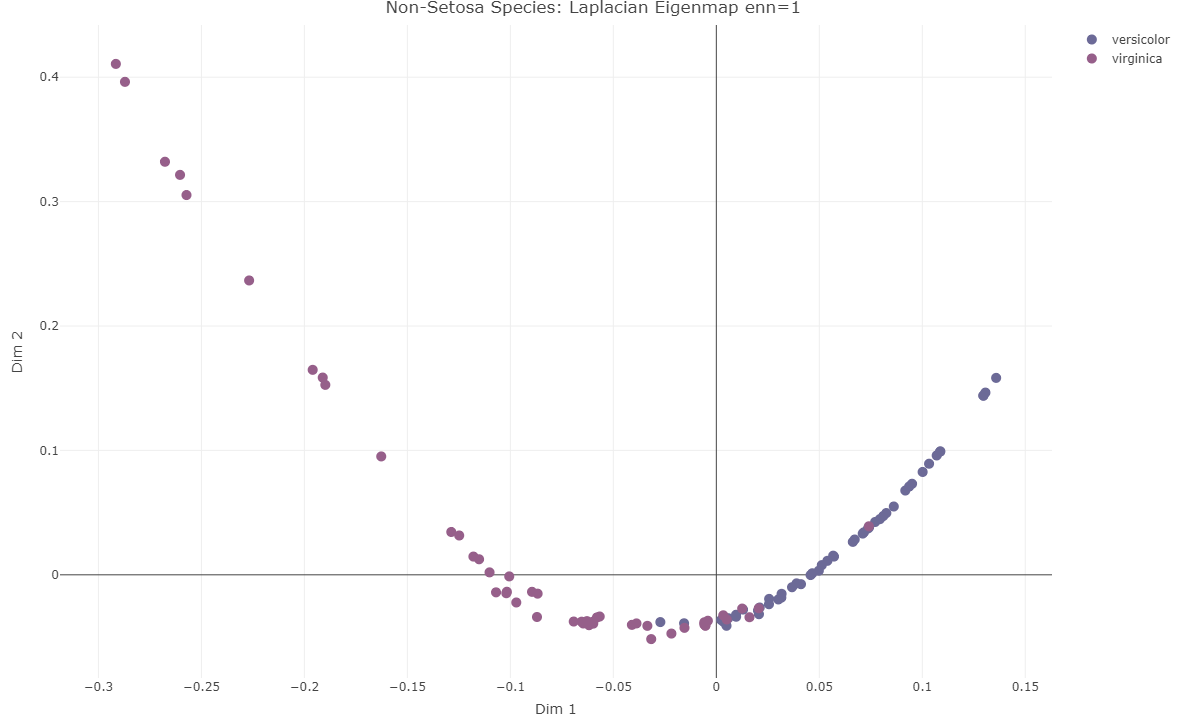
在選擇最佳proportion值時，目標是在保持品種間清晰分界的同時，最大程度地保持品種內部的結構。根據這些圖，一個中間值的proportion（例如0.3或0.5）似乎提供了最佳的品種分離和結構保持。

使用Proportion方法構造的鄰域圖揭示了數據點在空間中相對均勻分散的情況，與KNN和ENN方法相比，品種之間的區分不如ENN方法清晰。

**4.延伸分析**:由上面的分析可看出從k=10和e=1的這2張圖可以看到，降維後的第一個維度，即dim 1的數值，僅僅起到了將setosa的數據點與其它兩個種類的數據點分離的作用，所以我們針對這兩個情況再做延伸分析







1. **KNN k=10 (Dim2 vs Dim3)**：
   * 在**k=10**的KNN方法下，第二和第三維度顯示了不同品種之間的明顯分離。特別是，Setosa在這兩個維度上與其他兩種品種分開，而Versicolor和Virginica則在Dim3上有較好的分離，但在Dim2上有重疊。
2. **ENN e=1 (Dim2 vs Dim3)**：
   * 對於**e=1**的ENN方法，在第二和第三維度上同樣能看出Setosa的清晰分離。然而，Versicolor和Virginica品種之間的分離在這兩個維度上並不明顯。這可能表明ENN方法在這些參數設定下對這兩個品種的分離效果有限。
3. **KNN k=10 (Non-Setosa Species)**：
   * 從僅包含Versicolor和Virginica品種的數據集進行的KNN降維分析中，我們可以觀察到這兩個品種在空間中的分佈具有一定程度的分離，尤其是在Dim1。這表明，移除Setosa品種可以幫助我們更清晰地了解剩餘品種的結構差異。
4. **ENN e=1 (Non-Setosa Species)**：
   * ENN方法同樣顯示，在僅包含Versicolor和Virginica品種的數據集中，這兩個品種在降維後的空間中顯示出一定程度的分離。這表明ENN方法能夠揭示這兩個相似品種之間潛在的結構差異。

**評述：**

* 在選擇合適的方法時，需要根據數據的具體結構和分析的目標進行評估。對於鳶尾花數據集，ENN方法在適中的**e(e=5)**值下似乎提供了良好的品種分離，能顯示出3d圖的特徵。
* KNN方法在小的**k**值下突顯了局部結構，但隨著**k**值增大，數據的全局結構更為明顯，但可能會導致不同品種間的邊界變得模糊。
* Proportion方法提供了一種不同的視角，它可能更適合揭示數據的全局分佈，而非細節的局部分類結構。