Ex7_template_section2

October 25, 2019

1 Import dependencies

```
In [1]: import numpy as np
        from scipy.cluster.hierarchy import linkage, dendrogram
        from collections import defaultdict
        import matplotlib.pyplot as plt
        from mpl_toolkits import mplot3d
```

2 Utility

```
In [2]: ### Kudos to http://www.nxn.se/valent/extract-cluster-elements-by-color-in-python

        class Clusters(dict):
            def _repr_html_(self):
                html = '<table style="border: 0;">'
                for c in self:
                    hx = rgb2hex(colorConverter.to_rgb(c))
                    html += '<tr style="border: 0;">' \
                    '<td style="background-color: {0}; ' \
                               'border: 0;">' \
                    '<code style="background-color: {0};">'.format(hx)
                    html += c + '</code></td>'
                    html += '<td style="border: 0"><code>'
                    html += repr(self[c]) + '</code>'
                    html += '</td></tr>'

                html += '</table>'

                return html

        def get_cluster_classes(den, label='ivl'):
            cluster_idxs = defaultdict(list)
            for c, pi in zip(den['color_list'], den['icoord']):
                if c == 'grey':
                    continue
                for leg in pi[1:3]:
```

```
            i = (leg - 5.0) / 10.0
            if abs(i - int(i)) < 1e-10:
                cluster_idxs[c].append(int(i))

    cluster_classes = Clusters()
    for c, l in cluster_idxs.items():
        i_l = [den[label][i] for i in l]
        cluster_classes[c] = i_l

    return cluster_classes
```

3 Load the dataset 'cities_coordinates.txt'

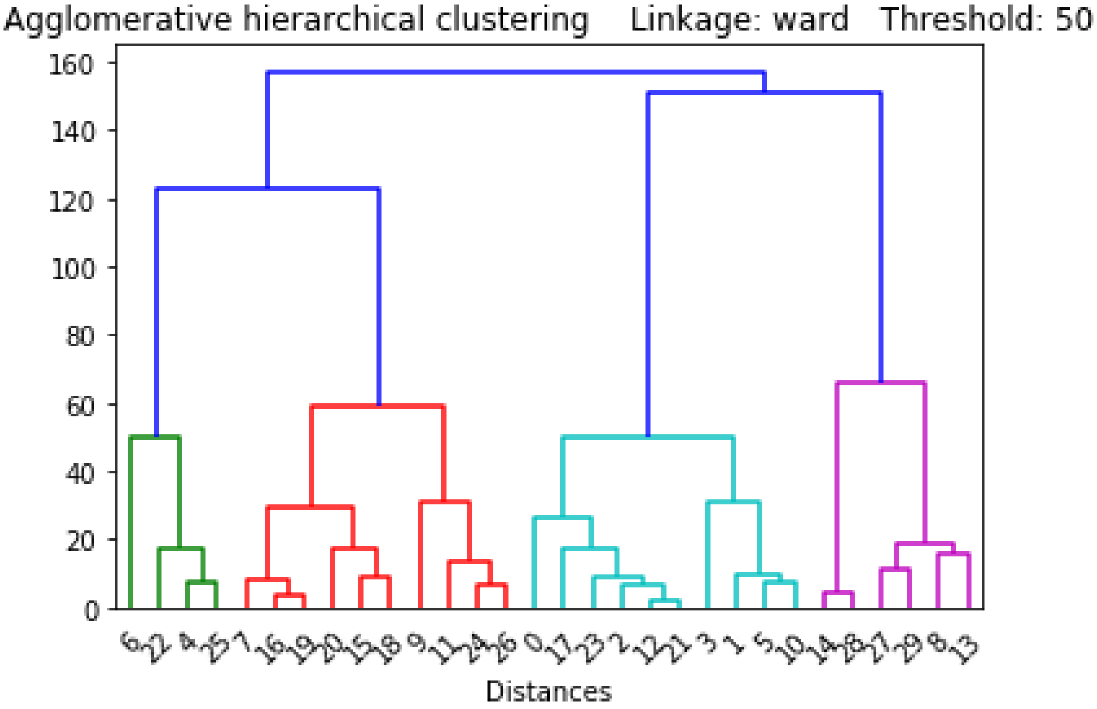
```
In [25]: ### TO DO ###
         X = np.loadtxt('cities_coordinates.txt', delimiter=',', usecols=(0, 1, 2))
         label = np.loadtxt('cities_coordinates.txt', dtype=np.str , delimiter=',', usecols=(3
```

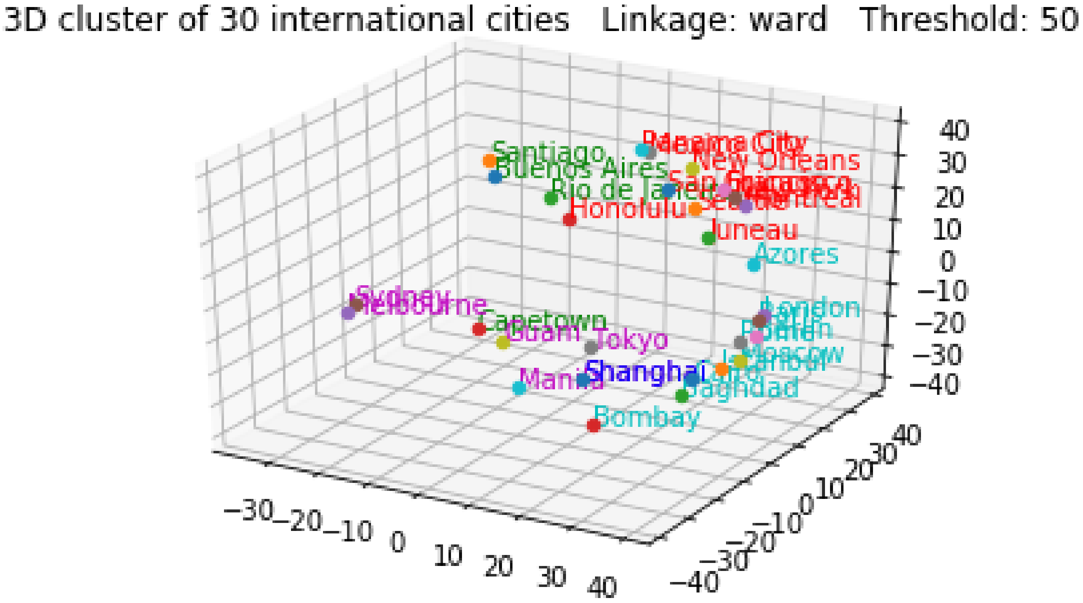
4 Hierarchical clustering and dendrograms

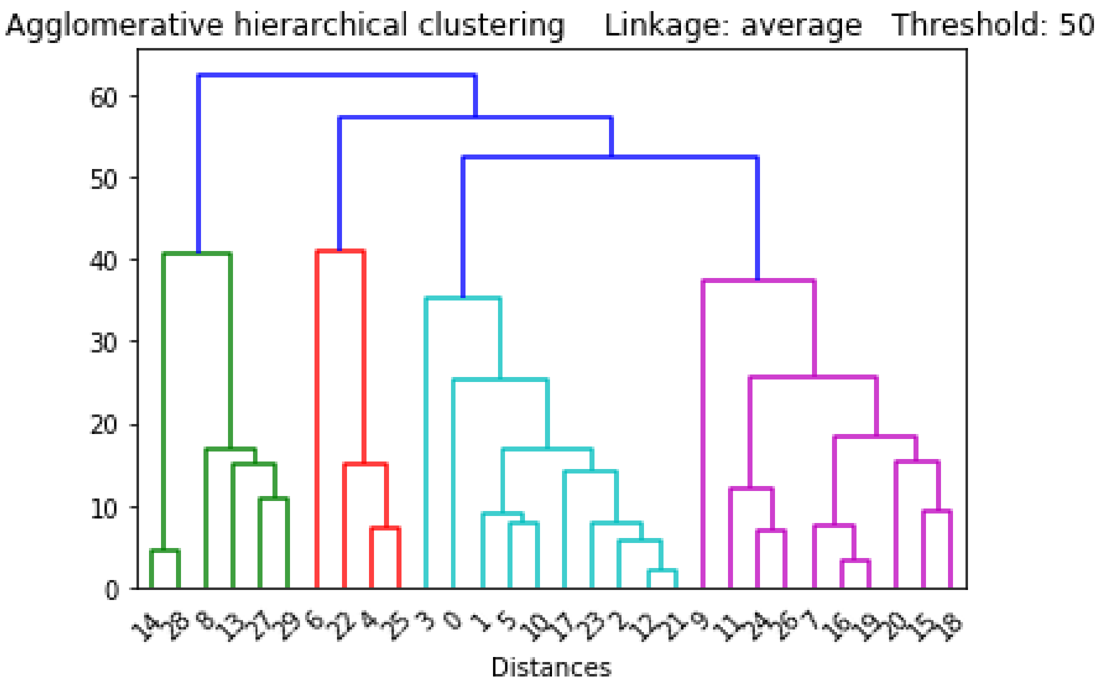
```
In [43]: color_threshold = 50

         for l, linkage_type in enumerate(['ward', 'average', 'complete', 'single']):
             # Do hierarchichal clustering:
             clusters_hierarchy = linkage(X, linkage_type)
             plt.figure(l+1)
             plt.title("Agglomerative hierarchical clustering    Linkage: " + str(linkage_type
             plt.xlabel("Distances") # Depending on the orientation of your dendrogram, it cou
             # Extract and display the dendrogram:
             cities_dendrogram = dendrogram(clusters_hierarchy)
             # Get the colors of every city:
             colored_clusters = get_cluster_classes(cities_dendrogram)
             fig = plt.figure(l+5)
             ax = plt.axes(projection='3d')
             ax.set_title("3D cluster of 30 international cities   Linkage: " + str(linkage_ty
             # Display every city on a scatter plot, with the color of its cluster:
             for color, cluster_cities in colored_clusters.items():
                                                             #        print(color, cl
                 # Find the index in y of every city in cluster_cities:
                 cluster_cities_idx = cluster_cities
                 for city_idx in cluster_cities_idx:
                     city_idx = int(city_idx)
                     # Extract the 3D coordinates of the city, and its name:
                     city_point = X[city_idx,:]
                     city_name = label[city_idx]
                     # Add the city point to the scatterplot, color this point with the color
                     ax.scatter3D(city_point[0], city_point[1], city_point[2], cmap=color)
                     # Annotate the scatterplot with the name of the city, colored as well:
```

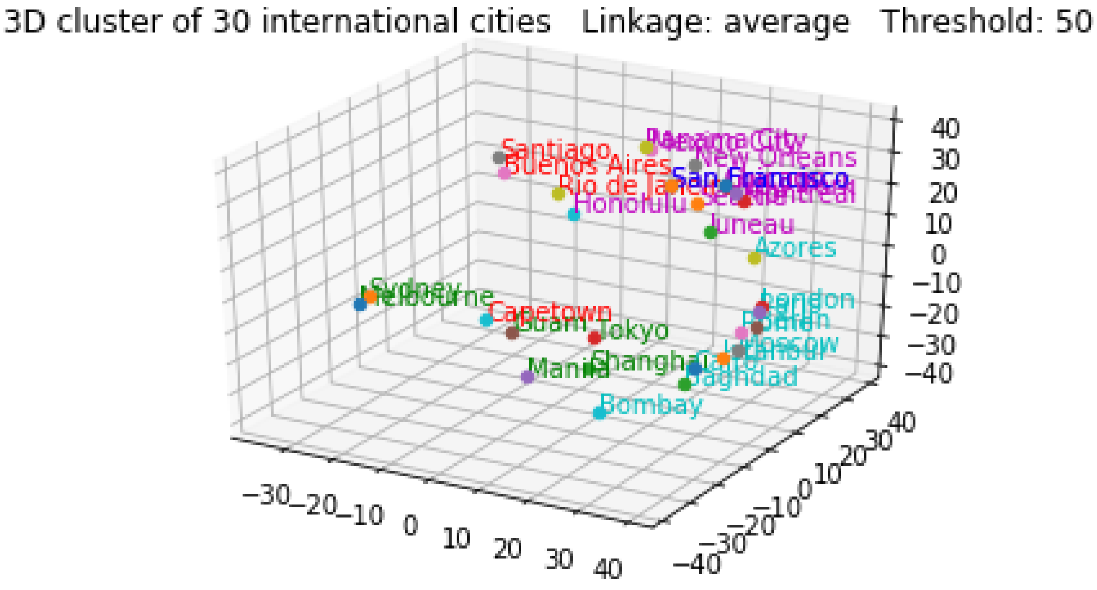
```
            ax.text(city_point[0], city_point[1], city_point[2],city_name, color=colo
plt.show()
```



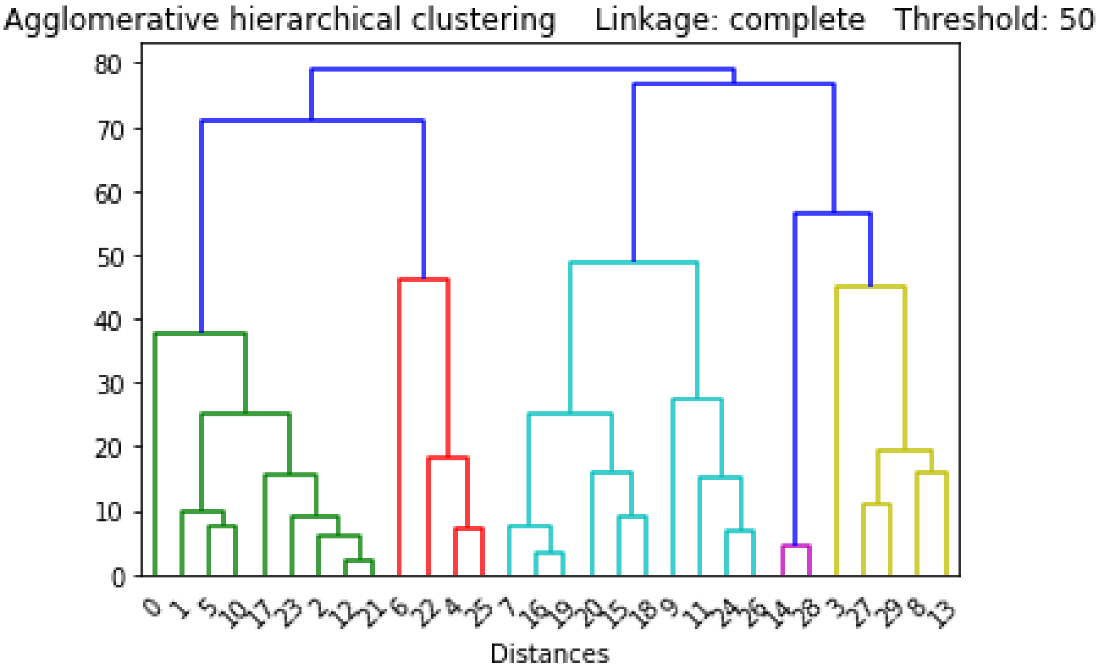


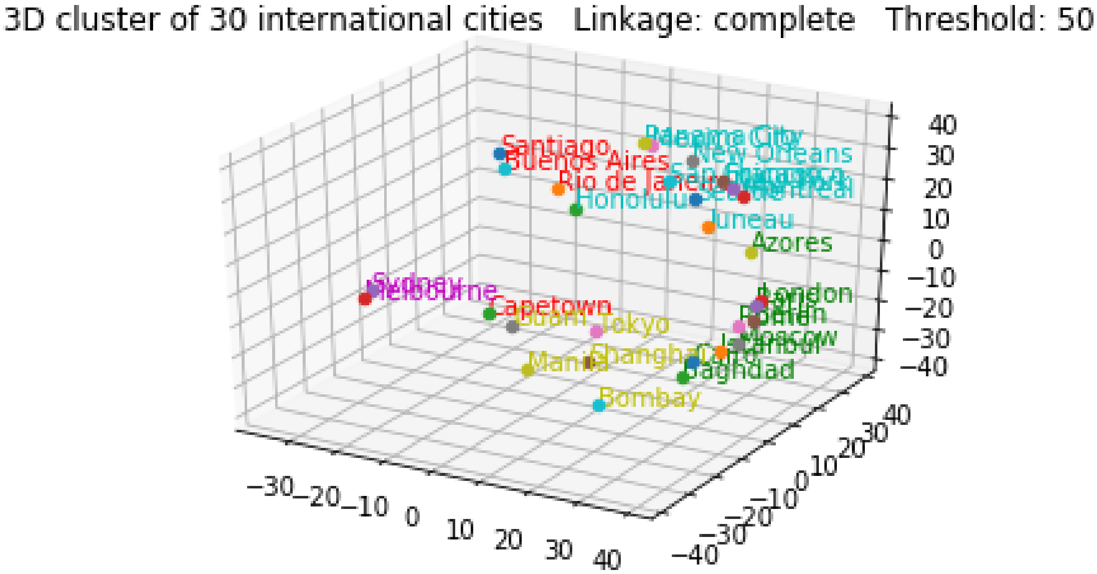


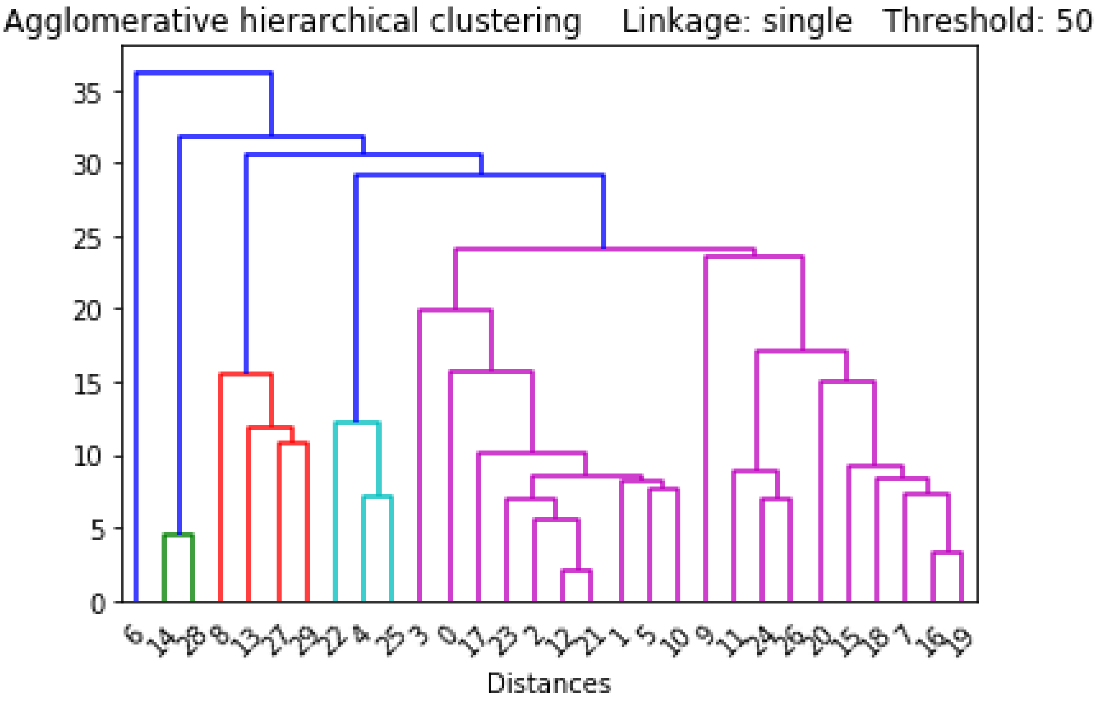
Agglomerative hierarchical clustering Linkage: average Threshold: 50

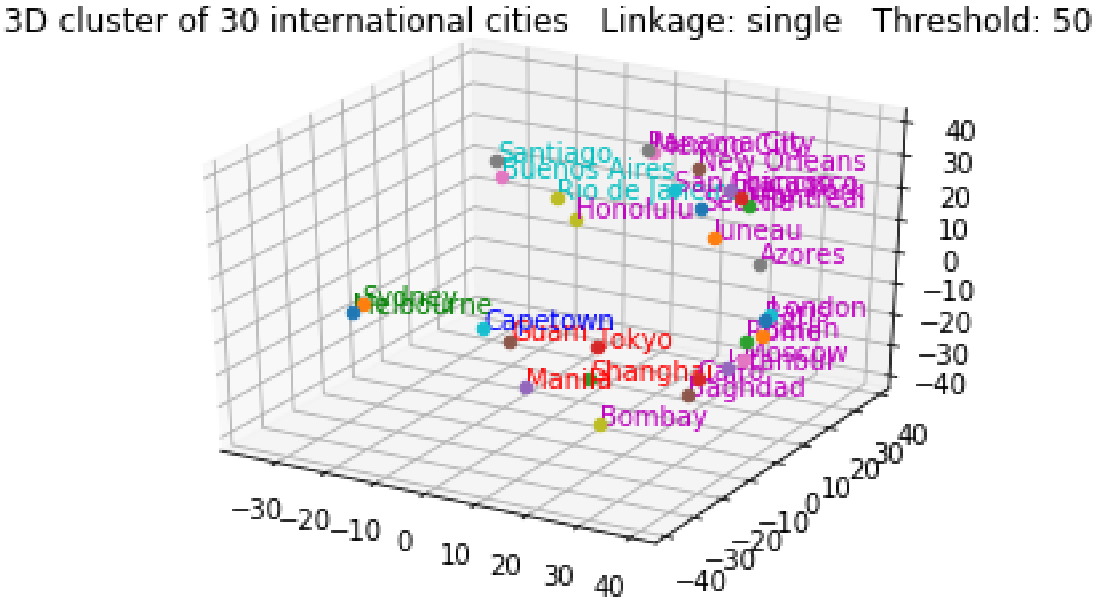


3D cluster of 30 international cities Linkage: average Threshold: 50


Agglomerative hierarchical clustering   Linkage: complete   Threshold: 50


3D cluster of 30 international cities   Linkage: complete   Threshold: 50


Agglomerative hierarchical clustering   Linkage: single   Threshold: 50
35
30
25
20
15
10
5
0
6
14
28
8
13
27
29
22
4
25
3
0
17
23
2
12
21
1
5
10
9
11
24
26
20
15
18
7
16
19
Distances


3D cluster of 30 international cities   Linkage: single   Threshold: 50
Panama City
Mexico City
New Orleans
Santiago
Buenos Aires
Rio de Janeiro
San Francisco
Chicago
Seattle
Montreal
Honolulu
Juneau
Azores
Sydney
Melbourne
Capetown
Guam
Tokyo
London
Paris
Berlin
Rome
Moscow
Istanbul
Cairo
Shanghai
Manila
Baghdad
Bombay
40
30
20
10
0
−10
−20
−30
−40
−30
−20
−10
0
10
20
30
40
40
30
20
10
0
−10
−20
−30
−40

6