

```
In [2]: import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
import seaborn as sns
import scipy.cluster.hierarchy as sch
from sklearn.cluster import AgglomerativeClustering
```

```
In [7]: crime = pd.read_csv('crime_data.csv')
crime
```

```
Out[7]:
```

	Unnamed: 0	Murder	Assault	UrbanPop	Rape
0	Alabama	13.2	236	58	21.2
1	Alaska	10.0	263	48	44.5
2	Arizona	8.1	294	80	31.0
3	Arkansas	8.8	190	50	19.5
4	California	9.0	276	91	40.6
5	Colorado	7.9	204	78	38.7
6	Connecticut	3.3	110	77	11.1
7	Delaware	5.9	238	72	15.8
8	Florida	15.4	335	80	31.9
9	Georgia	17.4	211	60	25.8
10	Hawaii	5.3	46	83	20.2
11	Idaho	2.6	120	54	14.2
12	Illinois	10.4	249	83	24.0
13	Indiana	7.2	113	65	21.0
14	Iowa	2.2	56	57	11.3
15	Kansas	6.0	115	66	18.0
16	Kentucky	9.7	109	52	16.3
17	Louisiana	15.4	249	66	22.2
18	Maine	2.1	83	51	7.8
19	Maryland	11.3	300	67	27.8
20	Massachusetts	4.4	149	85	16.3
21	Michigan	12.1	255	74	35.1
22	Minnesota	2.7	72	66	14.9
23	Mississippi	16.1	259	44	17.1
24	Missouri	9.0	178	70	28.2
25	Montana	6.0	109	53	16.4
26	Nebraska	4.3	102	62	16.5
27	Nevada	12.2	252	81	46.0
28	New Hampshire	2.1	57	56	9.5
29	New Jersey	7.4	159	89	18.8
30	New Mexico	11.4	285	70	32.1
31	New York	11.1	254	86	26.1
32	North Carolina	13.0	337	45	16.1
33	North Dakota	0.8	45	44	7.3

	Unnamed: 0	Murder	Assault	UrbanPop	Rape
34	Ohio	7.3	120	75	21.4
35	Oklahoma	6.6	151	68	20.0
36	Oregon	4.9	159	67	29.3
37	Pennsylvania	6.3	106	72	14.9
38	Rhode Island	3.4	174	87	8.3
39	South Carolina	14.4	279	48	22.5
40	South Dakota	3.8	86	45	12.8
41	Tennessee	13.2	188	59	26.9
42	Texas	12.7	201	80	25.5
43	Utah	3.2	120	80	22.9
44	Vermont	2.2	48	32	11.2
45	Virginia	8.5	156	63	20.7
46	Washington	4.0	145	73	26.2
47	West Virginia	5.7	81	39	9.3
48	Wisconsin	2.6	53	66	10.8
49	Wyoming	6.8	161	60	15.6

```
In [8]: #Normalized data fuction
def norm_func(i):
    x=(i-i.min())/(i.max()-i.min())
    return(x)
```

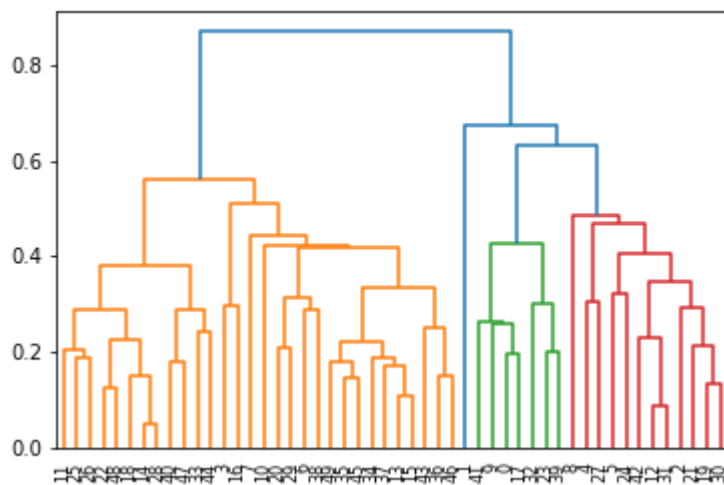
```
In [9]: df_norm=norm_func(crime.iloc[:,1:])
df_norm
```

```
Out[9]:
```

	<b>Murder</b>	<b>Assault</b>	<b>UrbanPop</b>	<b>Rape</b>
<b>0</b>	0.746988	0.654110	0.440678	0.359173
<b>1</b>	0.554217	0.746575	0.271186	0.961240
<b>2</b>	0.439759	0.852740	0.813559	0.612403
<b>3</b>	0.481928	0.496575	0.305085	0.315245
<b>4</b>	0.493976	0.791096	1.000000	0.860465
<b>5</b>	0.427711	0.544521	0.779661	0.811370
<b>6</b>	0.150602	0.222603	0.762712	0.098191
<b>7</b>	0.307229	0.660959	0.677966	0.219638
<b>8</b>	0.879518	0.993151	0.813559	0.635659
<b>9</b>	1.000000	0.568493	0.474576	0.478036
<b>10</b>	0.271084	0.003425	0.864407	0.333333
<b>11</b>	0.108434	0.256849	0.372881	0.178295
<b>12</b>	0.578313	0.698630	0.864407	0.431525
<b>13</b>	0.385542	0.232877	0.559322	0.354005
<b>14</b>	0.084337	0.037671	0.423729	0.103359
<b>15</b>	0.313253	0.239726	0.576271	0.276486
<b>16</b>	0.536145	0.219178	0.338983	0.232558
<b>17</b>	0.879518	0.698630	0.576271	0.385013
<b>18</b>	0.078313	0.130137	0.322034	0.012920
<b>19</b>	0.632530	0.873288	0.593220	0.529716
<b>20</b>	0.216867	0.356164	0.898305	0.232558
<b>21</b>	0.680723	0.719178	0.711864	0.718346
<b>22</b>	0.114458	0.092466	0.576271	0.196382
<b>23</b>	0.921687	0.732877	0.203390	0.253230
<b>24</b>	0.493976	0.455479	0.644068	0.540052
<b>25</b>	0.313253	0.219178	0.355932	0.235142
<b>26</b>	0.210843	0.195205	0.508475	0.237726
<b>27</b>	0.686747	0.708904	0.830508	1.000000
<b>28</b>	0.078313	0.041096	0.406780	0.056848
<b>29</b>	0.397590	0.390411	0.966102	0.297158
<b>30</b>	0.638554	0.821918	0.644068	0.640827
<b>31</b>	0.620482	0.715753	0.915254	0.485788
<b>32</b>	0.734940	1.000000	0.220339	0.227390
<b>33</b>	0.000000	0.000000	0.203390	0.000000

	Murder	Assault	UrbanPop	Rape
34	0.391566	0.256849	0.728814	0.364341
35	0.349398	0.363014	0.610169	0.328165
36	0.246988	0.390411	0.593220	0.568475
37	0.331325	0.208904	0.677966	0.196382
38	0.156627	0.441781	0.932203	0.025840
39	0.819277	0.801370	0.271186	0.392765
40	0.180723	0.140411	0.220339	0.142119
41	0.746988	0.489726	0.457627	0.506460
42	0.716867	0.534247	0.813559	0.470284
43	0.144578	0.256849	0.813559	0.403101
44	0.084337	0.010274	0.000000	0.100775
45	0.463855	0.380137	0.525424	0.346253
46	0.192771	0.342466	0.694915	0.488372
47	0.295181	0.123288	0.118644	0.051680
48	0.108434	0.027397	0.576271	0.090439
49	0.361446	0.397260	0.474576	0.214470

```
In [7]: dendrogram =sch.dendrogram(sch.linkage(df_norm,method ='average'))
```



```
In [3]: # Create Clusters (y)
hclusters=AgglomerativeClustering(n_clusters=5,affinity='euclidean',linkage='ward')
hclusters
```

```
Out[3]: AgglomerativeClustering(n_clusters=5)
```

```
In [10]: y=pd.DataFrame(hclusters.fit_predict(df_norm),columns=['clustersid'])  
y['clustersid'].value_counts()
```

```
Out[10]: 0      13  
         2      12  
         1       9  
         4       9  
         3       7  
         Name: clustersid, dtype: int64
```

```
In [11]: # Adding clusters to dataset
crime['clustersid']=hclusters.labels_
crime
```

```
Out[11]:
```

	Unnamed: 0	Murder	Assault	UrbanPop	Rape	clustersid
0	Alabama	13.2	236	58	21.2	3
1	Alaska	10.0	263	48	44.5	0
2	Arizona	8.1	294	80	31.0	0
3	Arkansas	8.8	190	50	19.5	4
4	California	9.0	276	91	40.6	0
5	Colorado	7.9	204	78	38.7	0
6	Connecticut	3.3	110	77	11.1	1
7	Delaware	5.9	238	72	15.8	1
8	Florida	15.4	335	80	31.9	0
9	Georgia	17.4	211	60	25.8	3
10	Hawaii	5.3	46	83	20.2	1
11	Idaho	2.6	120	54	14.2	2
12	Illinois	10.4	249	83	24.0	0
13	Indiana	7.2	113	65	21.0	4
14	Iowa	2.2	56	57	11.3	2
15	Kansas	6.0	115	66	18.0	4
16	Kentucky	9.7	109	52	16.3	4
17	Louisiana	15.4	249	66	22.2	3
18	Maine	2.1	83	51	7.8	2
19	Maryland	11.3	300	67	27.8	0
20	Massachusetts	4.4	149	85	16.3	1
21	Michigan	12.1	255	74	35.1	0
22	Minnesota	2.7	72	66	14.9	2
23	Mississippi	16.1	259	44	17.1	3
24	Missouri	9.0	178	70	28.2	0
25	Montana	6.0	109	53	16.4	2
26	Nebraska	4.3	102	62	16.5	2
27	Nevada	12.2	252	81	46.0	0
28	New Hampshire	2.1	57	56	9.5	2
29	New Jersey	7.4	159	89	18.8	1
30	New Mexico	11.4	285	70	32.1	0
31	New York	11.1	254	86	26.1	0
32	North Carolina	13.0	337	45	16.1	3

	Unnamed: 0	Murder	Assault	UrbanPop	Rape	clustersid
33	North Dakota	0.8	45	44	7.3	2
34	Ohio	7.3	120	75	21.4	4
35	Oklahoma	6.6	151	68	20.0	4
36	Oregon	4.9	159	67	29.3	1
37	Pennsylvania	6.3	106	72	14.9	4
38	Rhode Island	3.4	174	87	8.3	1
39	South Carolina	14.4	279	48	22.5	3
40	South Dakota	3.8	86	45	12.8	2
41	Tennessee	13.2	188	59	26.9	3
42	Texas	12.7	201	80	25.5	0
43	Utah	3.2	120	80	22.9	1
44	Vermont	2.2	48	32	11.2	2
45	Virginia	8.5	156	63	20.7	4
46	Washington	4.0	145	73	26.2	1
47	West Virginia	5.7	81	39	9.3	2
48	Wisconsin	2.6	53	66	10.8	2
49	Wyoming	6.8	161	60	15.6	4

In [12]: `crime.groupby('clustersid').agg(['mean']).reset_index()`

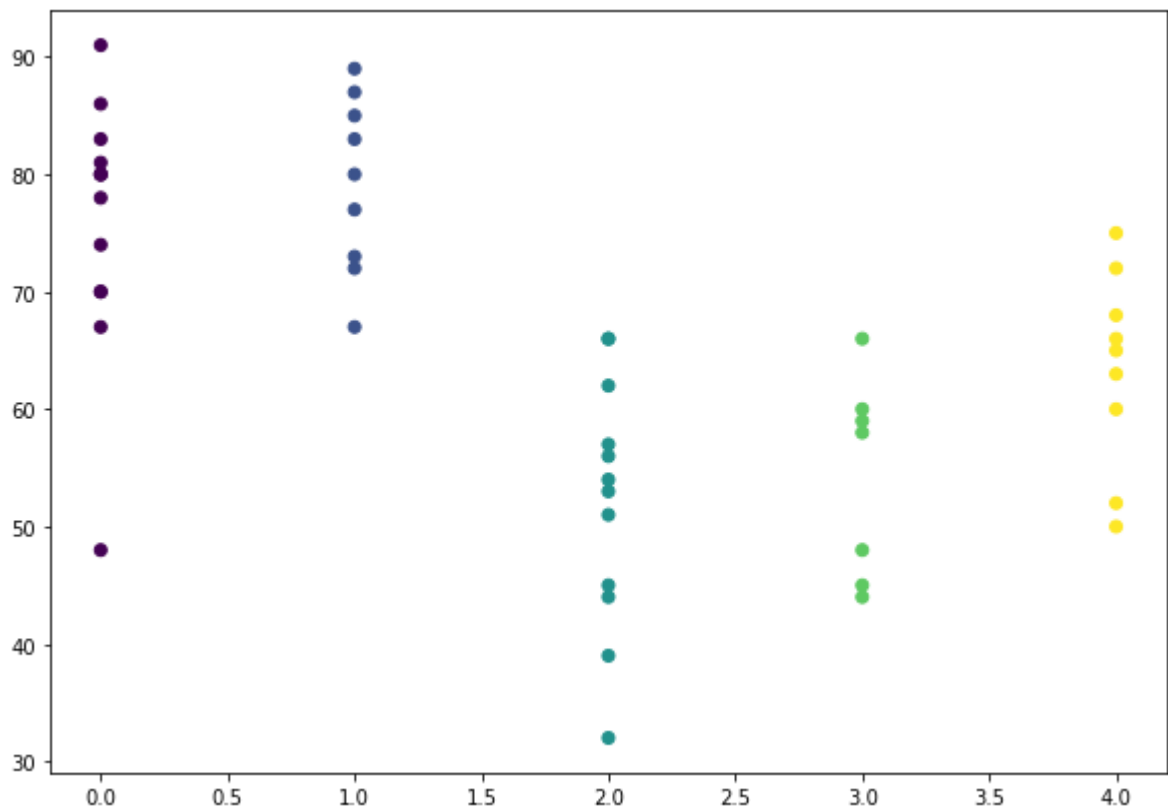
Out[12]:

	clustersid	Murder mean	Assault mean	UrbanPop mean	Rape mean
0	0	10.815385	257.384615	76.000000	33.192308
1	1	4.644444	144.444444	79.222222	18.766667
2	2	3.091667	76.000000	52.083333	11.833333
3	3	14.671429	251.285714	54.285714	21.685714
4	4	7.466667	135.666667	63.444444	18.600000



```
In [13]: # Plot Clusters
plt.figure(figsize=(10, 7))
plt.scatter(crime['clustersid'], crime['UrbanPop'], c=hclusters.labels_)
```

Out[13]: <matplotlib.collections.PathCollection at 0x268a878dfa0>



In [ ]: