```
import warnings
warnings.filterwarnings('ignore')

import pandas as pd
import numpy as np
from plotnine import *

from sklearn.preprocessing import StandardScaler

from sklearn.cluster import AgglomerativeClustering


from sklearn.cluster import KMeans
from sklearn.mixture import GaussianMixture

from sklearn.metrics import silhouette_score

import scipy.cluster.hierarchy as sch
from matplotlib import pyplot as plt

%matplotlib inline
```

```
# wide data

tests_wide = pd.read_csv("https://raw.githubusercontent.com/cmparlettpelleriti/CPSC392ParlettPelleriti/master/Data/testperform.csv")

tests_wide.head()
```

	zero	one	two	three	four	id
0	64.27	65.950884	69.865513	70.071844	71.557456	1
1	65.42	68.334638	71.774954	71.488519	72.414197	2
2	71.58	75.002741	77.169056	80.038548	80.307990	3
3	51.71	52.610899	53.005482	51.800135	53.829034	4
4	73.18	74.575287	76.507547	76.102415	75.441424	5

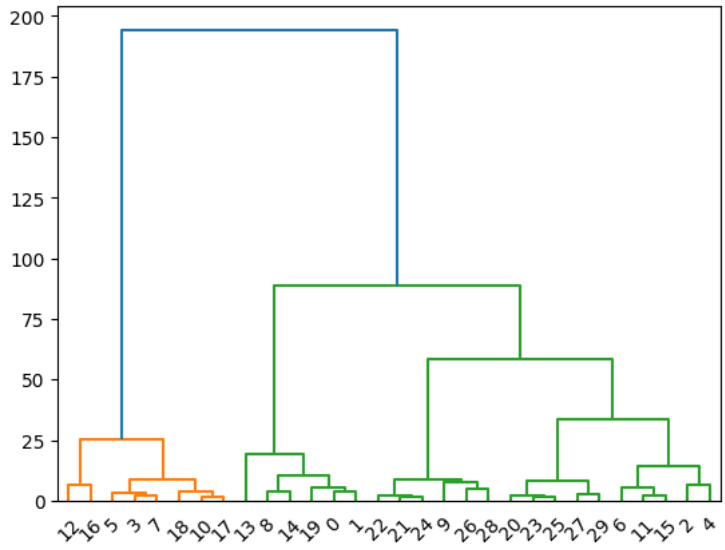
```
features = ["zero", "one", "two", "three", "four"]

X = tests_wide[features]

hac = AgglomerativeClustering(affinity = "euclidean",
                              linkage = "ward")
```

```
hac.fit(X)

dendro = sch.dendrogram(sch.linkage(X, method='ward'))
```




```
hac = AgglomerativeClustering(n_clusters = 3,
                              affinity = "euclidean",
                              linkage = "ward")
hac.fit(X)
```

```
AgglomerativeClustering(affinity='euclidean', n_clusters=3)
```

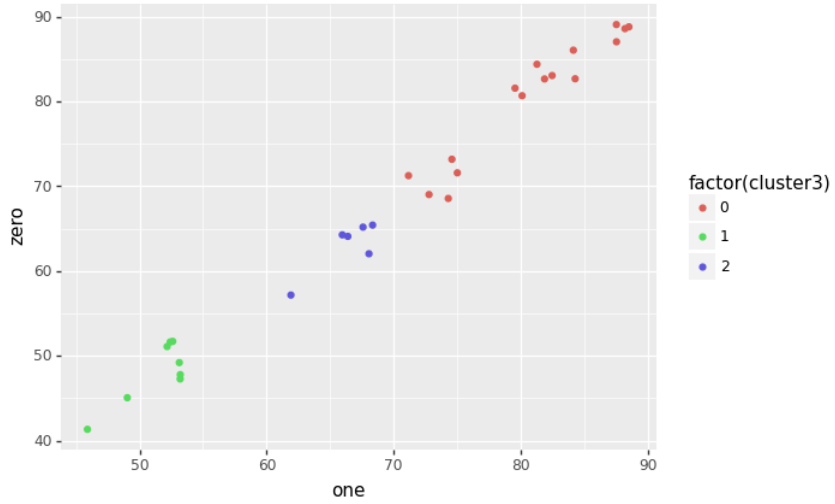
```
membership = hac.labels_
membership
```

```
array([2, 2, 0, 1, 0, 1, 0, 1, 2, 0, 1, 0, 1, 2, 2, 0, 1, 1, 1, 2, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0])
```

```
silhouette_score(X,membership)
```

```
0.5275892679012398
```

```
tests_wide["cluster3"] = membership
(ggplot(tests_wide, aes(x = "one", y = "zero")) + geom_point(aes(color = "factor(cluster3)")))
```




```
<ggplot: (8771688317454)>
```


✓ 1s completed at 00:35