Lab_Hierarchical clustering

KUMAR GAURAV 20122065

```
import pandas as pd
```

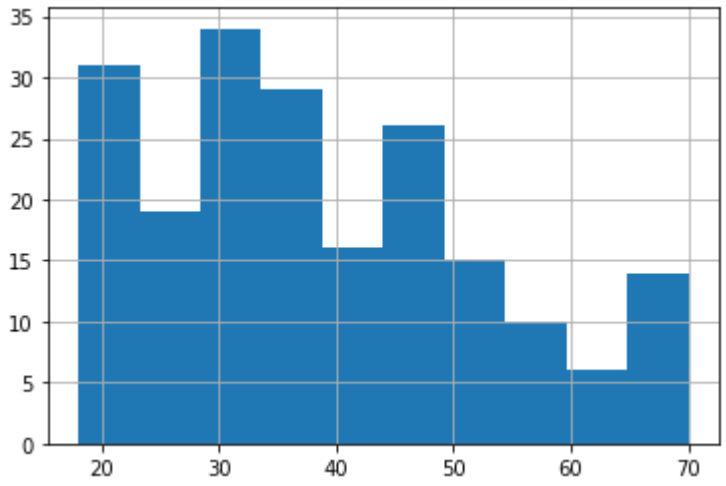
```
df = pd.read_csv('Mall_Customers.csv')
```

```
df.head()
```

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

```
df['Age'].hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7efc2c001710>
```



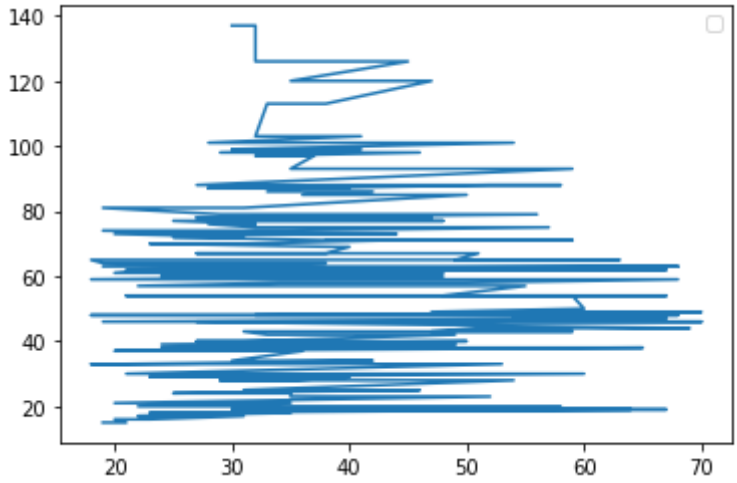

```
df['Annual Income (k$)'].hist(color = 'r')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7efc2b9a7090>
```


```
import matplotlib.pyplot as plt
```

```
plt.plot(df['Age'],df['Annual Income (k$)'])
plt.legend()
plt.show()
```

```
No handles with labels found to put in legend.
```




```
data = df.iloc[:,3:5]
data
```

	Annual Income (k$)	Spending Score (1-100)
0	15	39
1	15	81
2	16	6
3	16	77
4	17	40
...	...	...
195	120	79
196	126	28
197	126	74
198	137	18
199	137	83

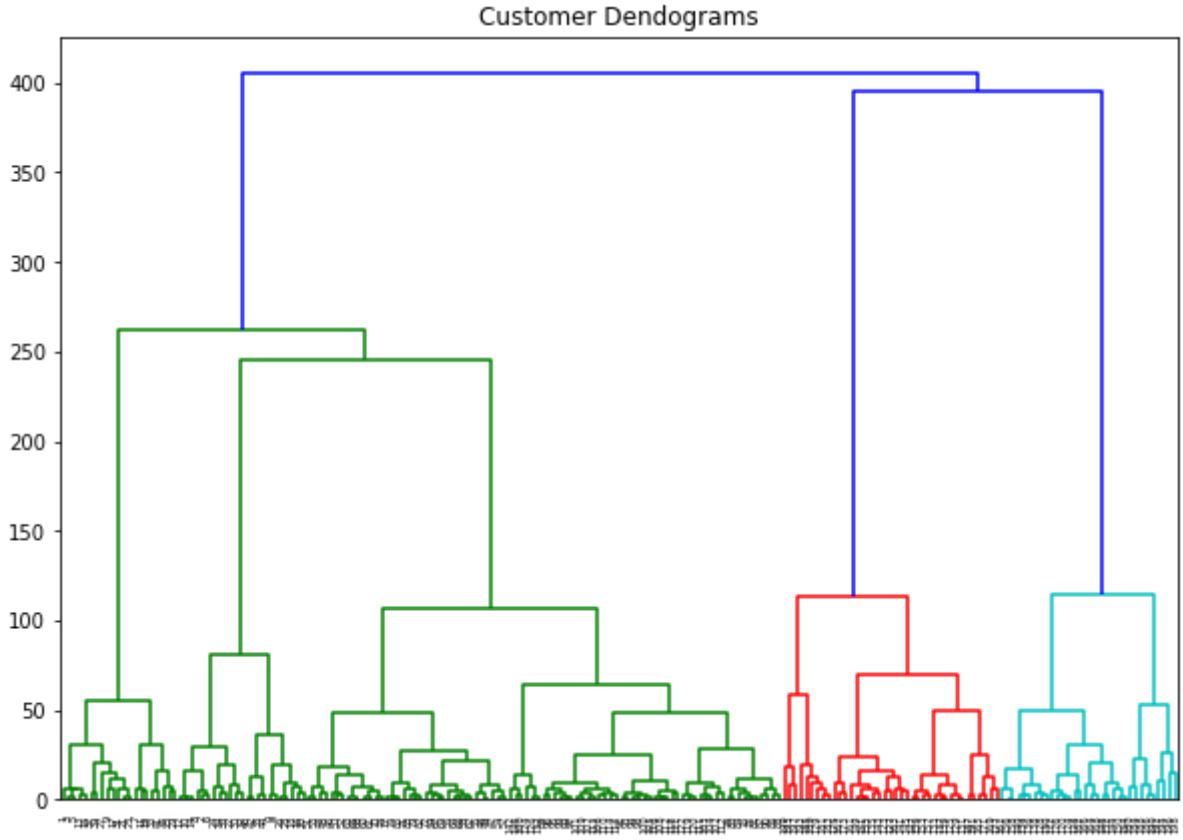
200 rows × 2 columns

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 2 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Annual Income (k$)      200 non-null    int64
 1   Spending Score (1-100)  200 non-null    int64
dtypes: int64(2)
memory usage: 3.2 KB
```

```
import scipy.cluster.hierarchy as shc


plt.figure(figsize=(10,7))
plt.title("Customer Dendograms")
dend = shc.dendrogram(shc.linkage(data, method='ward'))
```




```
from sklearn.cluster import AgglomerativeClustering


cluster = AgglomerativeClustering(n_clusters=5, affinity= 'euclidean',linkage = 'ward')
cluster.fit_predict(data)
```

```
array([4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3,
       4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 1,
       4, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 2, 1, 2, 0, 2, 0, 2,
       1, 2, 0, 2, 0, 2, 0, 2, 0, 2, 1, 2, 0, 2, 1, 2, 0, 2, 0, 2, 0, 2,
       0, 2, 0, 2, 0, 2, 1, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2,
```

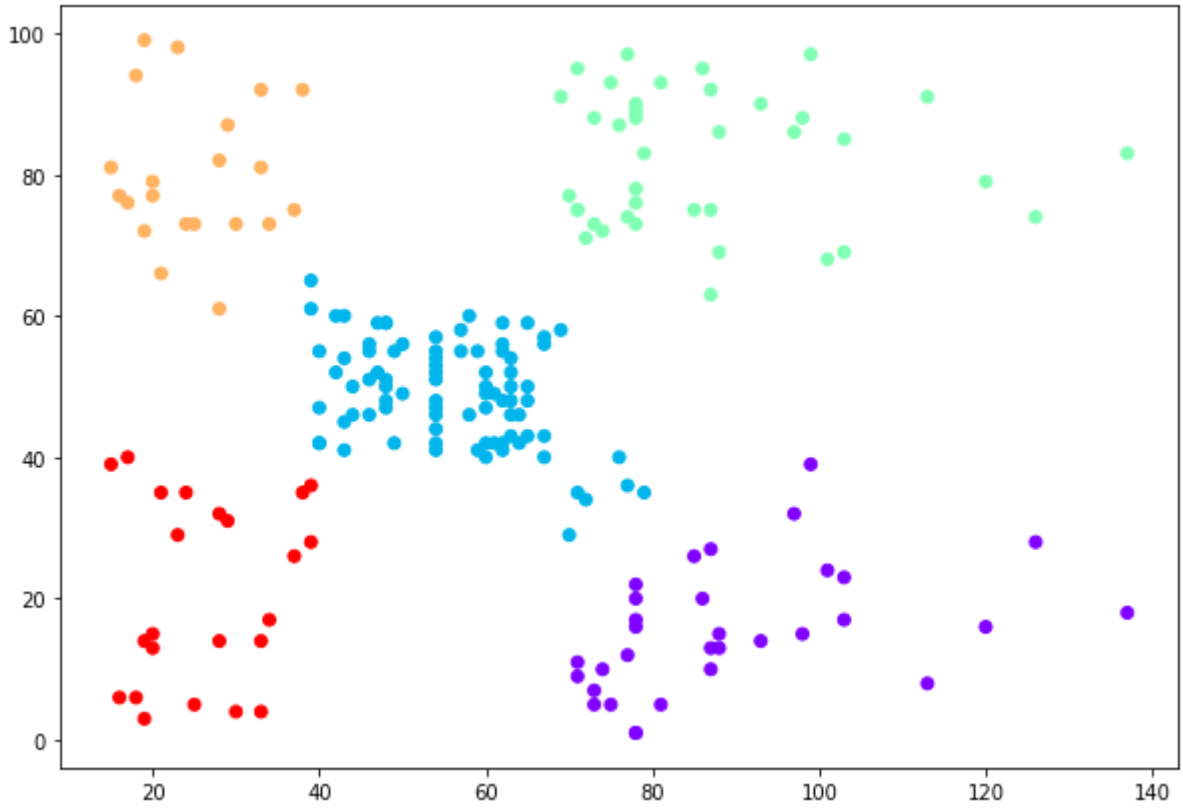
```
       0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2,
       0, 2])
```

```
#You can see the cluster labels from all of your data points. Since we had five clusters,
```

```
data.head()
```

	Annual Income (k$)	Spending Score (1-100)
0	15	39
1	15	81
2	16	6
3	16	77
4	17	40

```
plt.figure(figsize=(10, 7))
plt.scatter(data.iloc[:,0], data.iloc[:,1], c=cluster.labels_, cmap='rainbow')
plt.show()
```




```
#You can see the data points in the form of five clusters.
#The data points in the bottom right belong to the customers
#with high salaries but low spending. These are the customers
#that spend their money carefully. Similarly, the customers
#at top right (green data points), these are the customers
#with high salaries and high spending
```

✓ 0s completed at 4:26 PM