```
Chapter 4 - Clustering Models

Part 2 - Hierarchial methods

Setting up for clustering analysis
```

In [11]:
```
import numpy as np
import pandas as pd

import matplotlib.pyplot as plt
from pylab import rcParams
import seaborn as sb

import sklearn
import sklearn.metrics as sm
```

In [12]:
```
from sklearn.cluster import AgglomerativeClustering

import scipy
from scipy.cluster.hierarchy import dendrogram, linkage
from scipy.cluster.hierarchy import fcluster
from scipy.cluster.hierarchy import cophenet
from scipy.spatial.distance import pdist
```

In [13]:
```
np.set_printoptions(precision=4, suppress=True)
plt.figure(figsize=(10,3))
%matplotlib inline
plt.style.use('seaborn-whitegrid')
```

In [14]:
```
address = 'C:/Users/danal/Desktop/ExerciseFiles/Data/mtcars.csv'

cars = pd.read_csv(address)
cars.columns = ['car_names', 'mpg', 'cyl', 'disp', 'hp', 'drat', 'wt', 'qsec', 'v

X = cars[['mpg', 'disp', 'hp', 'wt']].values
y = cars.iloc[:,(9)].values
```

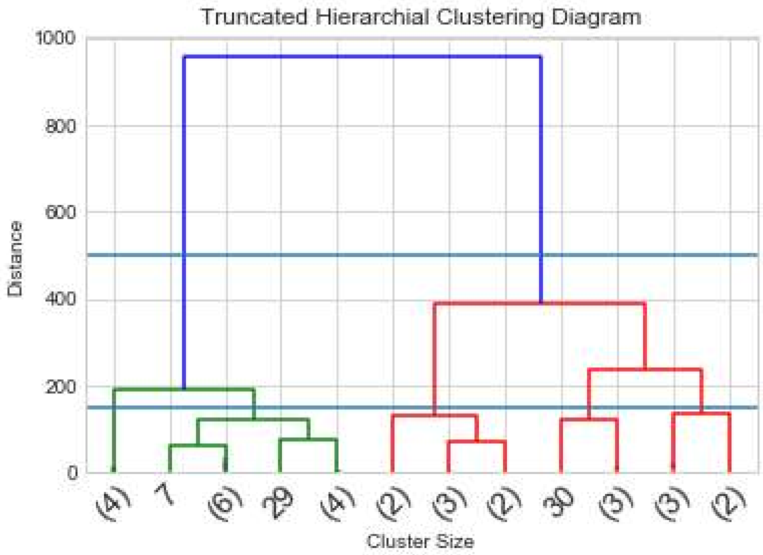
Using scipy to generate dendrograms

In [15]:
```
z = linkage(X, 'ward')
```

In [16]:
```
dendrogram(z, truncate_mode='lastp', p=12, leaf_rotation=45., leaf_font_size=15,

plt.title('Truncated Hierarchial Clustering Diagram')
plt.xlabel('Cluster Size')
plt.ylabel('Distance')

plt.axhline(y=500)
plt.axhline(y=150)
plt.show()
```



Generating hierarchial clusters

In [17]:
```
k=2

Hclustering = AgglomerativeClustering(n_clusters=k, affinity='euclidean', linkage
Hclustering.fit(X)

sm.accuracy_score(y, Hclustering.labels_)
```

Out[17]:
```
0.78125
```

In [18]:

```
k=2

Hclustering = AgglomerativeClustering(n_clusters=k, affinity='manhattan', linkage
Hclustering.fit(X)

sm.accuracy_score(y, Hclustering.labels_)
```

Out[18]:

```
0.71875
```

In [19]:

```
k=2

Hclustering = AgglomerativeClustering(n_clusters=k, affinity='euclidean', linkage
Hclustering.fit(X)

sm.accuracy_score(y, Hclustering.labels_)
```

Out[19]:

```
0.78125
```