Problem 1

```
=== Cluster statistics ===
              mpg            displacement               horsepower                weight                 acceleration
             mean       var          mean          var         mean         var         mean            var          mean        var
cluster
0        27.365414  41.976309   131.934211  2828.083391   84.300061  369.143491  2459.511278  182632.099872    16.298120   5.718298
1        13.889062   3.359085   358.093750  2138.213294  167.046875  756.521577  4398.593750   74312.340278    13.025000   3.591429
2        17.510294   8.829892   278.985294  2882.492318  124.470588  713.088674  3624.838235   37775.809263    15.105882  10.556980

=== Origin statistics ===
             mpg            displacement               horsepower                 weight                 acceleration
            mean       var          mean          var         mean          var         mean            var          mean       var
origin
1       20.083534  40.997026   245.901606  9702.612255  118.814769  1569.532304  3361.931727  631695.128385    15.033735  7.568615
2       27.891429  45.211230   109.142857   509.950311   81.241983   410.659789  2423.300000  240142.328986    16.787143  9.276209
3       30.450633  37.088685   102.708861   535.465433   79.835443   317.523856  2221.227848  102718.485881    16.172152  3.821779

=== Crosstab cluster vs origin ===
origin     1   2   3
cluster
0        120  67  79
1         64   0   0
2         65   3   0
```

The results of the analyses show that there is a degree of clarity in the relationship between cluster assignments and vehicle origin labelling, but not a perfect correspondence:
Cluster 1 exclusively grouped American cars, characterized by the lowest mean MPG (13.89) and the highest mean displacement, horsepower, and weight. This clearly identifies a segment of typical American heavy-duty, high-consumption vehicles.
Cluster 2 also predominantly consisted of American cars (65 out of 68), exhibiting traits similar to Cluster 1 (e.g., low MPG of 17.51, high weight) but less extreme.
Cluster 0 was more diverse, containing all Japanese (79) and most European (67) cars, alongside a substantial number of American cars (120). This cluster represented vehicles with higher average MPG (27.37) and lower average weight and displacement, typical of more fuel-efficient models across all origins.
In summary, the hierarchical clustering successfully distinguished distinct groups of vehicles, particularly isolating segments of American cars based on their physical and performance characteristics. While not a perfect one-to-one mapping for all origins due to the mixed nature of one cluster, a clear relationship between cluster assignment and vehicle origin is evident, especially in identifying less fuel-efficient, heavier American vehicles.

Problem 2

```
Shape of data: (506, 13)
k=2  Silhouette=0.3601
k=3  Silhouette=0.2575
k=4  Silhouette=0.2658
k=5  Silhouette=0.2878
k=6  Silhouette=0.2625

  Best k by silhouette: 2

=== Cluster mean (scaled features) ===
          crim     zn  indus   chas    nox     rm    age    dis    rad    tax  ptratio      b  lstat
cluster
0       -0.390  0.262 -0.620  0.003 -0.585  0.243 -0.435  0.457 -0.584 -0.631   -0.286  0.326 -0.446
1        0.725 -0.488  1.153 -0.005  1.087 -0.452  0.809 -0.850  1.085  1.174    0.531 -0.607  0.830

=== Centroid coordinates (scaled) ===
    crim     zn  indus   chas    nox     rm    age    dis    rad    tax  ptratio      b  lstat
0 -0.390  0.262 -0.620  0.003 -0.585  0.243 -0.435  0.457 -0.584 -0.631   -0.286  0.326 -0.446
1  0.725 -0.488  1.153 -0.005  1.087 -0.452  0.809 -0.850  1.085  1.174    0.531 -0.607  0.830
```

Among k = 2 ⋯ 6, k = 2 yields the highest Silhouette score = 0.3601, clearly outperforming the other choices (> 0.07 margin). Therefore, k = 2 is selected as the optimal number of clusters.

Cluster	Key mean shifts (relative to 0)	Interpretation
Cluster 0	crim ↓ indus ↓ nox ↓ tax ↓ rad ↓ , zn ↑ rm ↑ dis ↑ b ↑	Low-crime, low-industry, cleaner air, larger rooms, farther from employment centres – neighbourhoods with generally higher living quality.
Cluster 1	crim ↑ indus ↑ nox ↑ tax ↑ rad ↑ , zn ↓ rm ↓ dis ↓ b ↓	High-crime, high-industry, more pollution, higher taxes, smaller rooms, close to main roads – areas with comparatively lower residential desirability.

The printed Cluster Mean and Centroid Coordinates are identical for all 13 features (up to three decimals). This is expected, because after convergence, a k-means centroid is the arithmetic mean of all points assigned to that cluster.

Summary: Silhouette analysis indicates that k = 2 provides the best clustering structure for the Boston Housing data. Consistent cluster means and centroid coordinates confirm that the k-means algorithm has converged properly and that the centroids truly represent the central tendency of their respective clusters.

Problem 3

```
使用 load_wine() 载入数据
数据维度: (178, 13)

Homogeneity Score   : 0.8788
Completeness Score  : 0.8730
```

Metric interpretation
Homogeneity measures cluster purity: a score of 1 means each cluster contains samples from only one true class.
Completeness measures class completeness: a score of 1 means all samples of a given class are assigned to the same cluster.

Result analysis
Both scores are above 0.87, indicating that K-Means with k = 3 recovers the underlying wine classes very well. Homogeneity is slightly higher than completeness, suggesting that clusters are highly pure, while a few classes are still split across clusters to a minor extent.
Overall, the clustering closely reproduces the true class structure, with only a small number of mis-assignments or class splits.