1. 1. Food set

This dataset contains information on 25 different countries and the types of foods they consume

summary(food)

Country RedMeat WhiteMeat Eggs Milk Fish Cereals

Albania : 1 Min. : 4.400 Min. : 1.400 Min. :0.500 Min. : 4.90 Min. : 0.200 Min. :18.60

Austria : 1 1st Qu.: 7.800 1st Qu.: 4.900 1st Qu.:2.700 1st Qu.:11.10 1st Qu.: 2.100 1st Qu.:24.30

Belgium : 1 Median : 9.500 Median : 7.800 Median :2.900 Median :17.60 Median : 3.400 Median :28.00

Bulgaria : 1 Mean : 9.828 Mean : 7.896 Mean :2.936 Mean :17.11 Mean : 4.284 Mean :32.25

Czechoslovakia: 1 3rd Qu.:10.600 3rd Qu.:10.800 3rd Qu.:3.700 3rd Qu.:23.30 3rd Qu.: 5.800 3rd Qu.:40.10

Denmark : 1 Max. :18.000 Max. :14.000 Max. :4.700 Max. :33.70 Max. :14.200 Max. :56.70

(Other) :19

Starch Nuts Fr.Veg

Min. :0.600 Min. :0.700 Min. :1.400

1st Qu.:3.100 1st Qu.:1.500 1st Qu.:2.900

Median :4.700 Median :2.400 Median :3.800

Mean :4.276 Mean :3.072 Mean :4.136

3rd Qu.:5.700 3rd Qu.:4.700 3rd Qu.:4.900

Max. :6.500 Max. :7.800 Max. :7.900

* 1. Unexpstates

This dataset contains unemployment data by state.

summary(raw)

AL AK AZ AR CA CO CT

Min. : 3.300 Min. : 5.900 Min. : 3.700 Min. : 4.000 Min. : 4.700 Min. :2.600 Min. :2.100

1st Qu.: 4.575 1st Qu.: 6.900 1st Qu.: 4.800 1st Qu.: 5.100 1st Qu.: 5.775 1st Qu.:4.400 1st Qu.:4.300

Median : 6.300 Median : 7.600 Median : 5.800 Median : 6.300 Median : 6.900 Median :5.500 Median :5.100

Mean : 6.645 Mean : 8.033 Mean : 6.121 Mean : 6.474 Mean : 7.204 Mean :5.384 Mean :5.182

3rd Qu.: 7.850 3rd Qu.: 9.300 3rd Qu.: 6.725 3rd Qu.: 7.600 3rd Qu.: 8.200 3rd Qu.:6.125 3rd Qu.:6.025

Max. :14.300 Max. :11.500 Max. :11.600 Max. :10.100 Max. :12.600 Max. :8.800 Max. :9.400

DE FL GA HI ID IL IN

Min. :2.800 Min. : 3.300 Min. : 3.300 Min. :2.300 Min. :2.700 Min. : 4.200 Min. : 2.600

1st Qu.:3.600 1st Qu.: 5.000 1st Qu.: 4.700 1st Qu.:3.600 1st Qu.:5.100 1st Qu.: 5.400 1st Qu.: 4.600

Median :4.600 Median : 5.900 Median : 5.400 Median :4.900 Median :5.500 Median : 6.400 Median : 5.400

Mean :5.186 Mean : 6.227 Mean : 5.685 Mean :4.736 Mean :5.923 Mean : 6.876 Mean : 6.021

3rd Qu.:6.900 3rd Qu.: 7.200 3rd Qu.: 6.300 3rd Qu.:5.800 3rd Qu.:7.200 3rd Qu.: 8.025 3rd Qu.: 6.825

Max. :9.300 Max. :12.300 Max. :10.500 Max. :9.900 Max. :9.600 Max. :12.900 Max. :12.700

IA KS KY LA ME MD MA

Min. :2.500 Min. :3.000 Min. : 4.10 Min. : 3.60 Min. :3.100 Min. :3.400 Min. : 2.600

1st Qu.:3.700 1st Qu.:4.100 1st Qu.: 5.50 1st Qu.: 5.60 1st Qu.:4.600 1st Qu.:4.200 1st Qu.: 4.100

Median :4.350 Median :4.600 Median : 6.05 Median : 6.70 Median :5.400 Median :4.800 Median : 5.300

Mean :4.760 Mean :4.661 Mean : 6.82 Mean : 7.21 Mean :5.794 Mean :5.184 Mean : 5.632

3rd Qu.:5.525 3rd Qu.:5.100 3rd Qu.: 8.25 3rd Qu.: 8.10 3rd Qu.:6.900 3rd Qu.:6.200 3rd Qu.: 6.800

Max. :8.600 Max. :7.200 Max. :12.00 Max. :12.80 Max. :9.000 Max. :8.400 Max. :11.100

MI MN MS MO MT NE NV

Min. : 3.300 Min. :2.500 Min. : 4.900 Min. : 2.800 Min. :3.200 Min. :2.200 Min. : 3.800

1st Qu.: 6.550 1st Qu.:4.000 1st Qu.: 6.300 1st Qu.: 4.800 1st Qu.:5.000 1st Qu.:2.775 1st Qu.: 4.800

Median : 7.400 Median :4.700 Median : 7.100 Median : 5.600 Median :5.600 Median :3.200 Median : 5.700

Mean : 8.194 Mean :4.938 Mean : 7.822 Mean : 5.867 Mean :5.765 Mean :3.512 Mean : 6.394

3rd Qu.: 9.500 3rd Qu.:5.600 3rd Qu.: 8.925 3rd Qu.: 6.400 3rd Qu.:6.500 3rd Qu.:4.100 3rd Qu.: 7.300

Max. :16.800 Max. :9.100 Max. :13.500 Max. :10.600 Max. :8.800 Max. :6.700 Max. :14.400

NH NJ NM NY NC ND OH

Min. :2.100 Min. : 3.600 Min. : 3.400 Min. : 4.000 Min. : 3.100 Min. :2.600 Min. : 3.800

1st Qu.:3.300 1st Qu.: 4.600 1st Qu.: 5.800 1st Qu.: 5.200 1st Qu.: 4.500 1st Qu.:3.300 1st Qu.: 5.400

Median :4.100 Median : 5.900 Median : 6.800 Median : 6.450 Median : 5.300 Median :4.000 Median : 6.100

Mean :4.411 Mean : 6.203 Mean : 6.832 Mean : 6.583 Mean : 5.618 Mean :4.099 Mean : 6.810

3rd Qu.:5.400 3rd Qu.: 7.400 3rd Qu.: 7.900 3rd Qu.: 7.700 3rd Qu.: 6.400 3rd Qu.:4.700 3rd Qu.: 7.925

Max. :7.600 Max. :10.700 Max. :10.000 Max. :10.300 Max. :11.200 Max. :6.800 Max. :13.900

OK OR PA RI SC SD TN

Min. :2.800 Min. : 4.700 Min. : 4.000 Min. : 2.900 Min. : 3.200 Min. :2.500 Min. : 3.900

1st Qu.:4.100 1st Qu.: 5.600 1st Qu.: 5.000 1st Qu.: 4.800 1st Qu.: 5.000 1st Qu.:3.200 1st Qu.: 5.300

Median :5.100 Median : 6.600 Median : 5.900 Median : 5.550 Median : 6.150 Median :3.600 Median : 5.900

Mean :5.287 Mean : 7.172 Mean : 6.517 Mean : 6.237 Mean : 6.349 Mean :3.768 Mean : 6.542

3rd Qu.:6.300 3rd Qu.: 8.500 3rd Qu.: 7.700 3rd Qu.: 7.700 3rd Qu.: 6.900 3rd Qu.:4.400 3rd Qu.: 7.425

Max. :9.200 Max. :12.100 Max. :12.900 Max. :12.700 Max. :12.500 Max. :6.000 Max. :12.800

TX UT VT VA WA WV WI

Min. :4.200 Min. : 2.500 Min. :2.40 Min. :2.20 Min. : 4.400 Min. : 3.900 Min. : 3.000

1st Qu.:5.100 1st Qu.: 3.700 1st Qu.:3.70 1st Qu.:3.80 1st Qu.: 5.600 1st Qu.: 5.975 1st Qu.: 4.300

Median :6.100 Median : 4.800 Median :4.50 Median :4.50 Median : 6.700 Median : 7.800 Median : 4.900

Mean :6.156 Mean : 4.991 Mean :4.84 Mean :4.62 Mean : 6.982 Mean : 8.422 Mean : 5.446

3rd Qu.:7.000 3rd Qu.: 5.900 3rd Qu.:5.90 3rd Qu.:5.40 3rd Qu.: 8.100 3rd Qu.:10.500 3rd Qu.: 6.400

Max. :9.300 Max. :10.000 Max. :8.80 Max. :7.80 Max. :12.200 Max. :18.100 Max. :11.500

WY

Min. :2.300

1st Qu.:3.900

Median :4.900

Mean :5.019

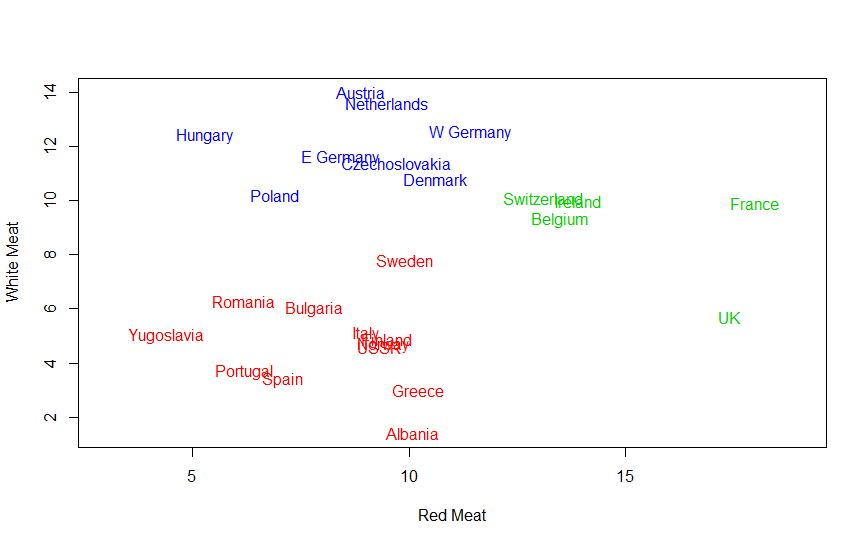
3rd Qu.:6.100

Max. :9.100

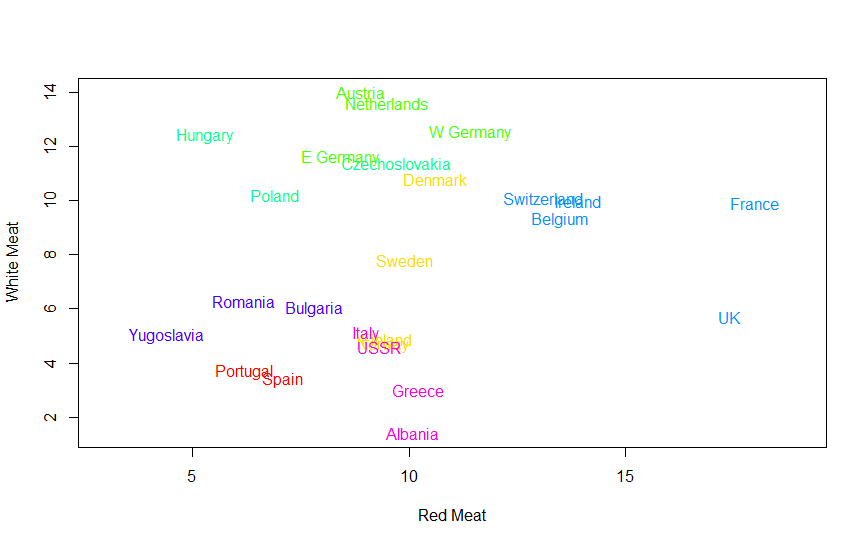


Example 1:

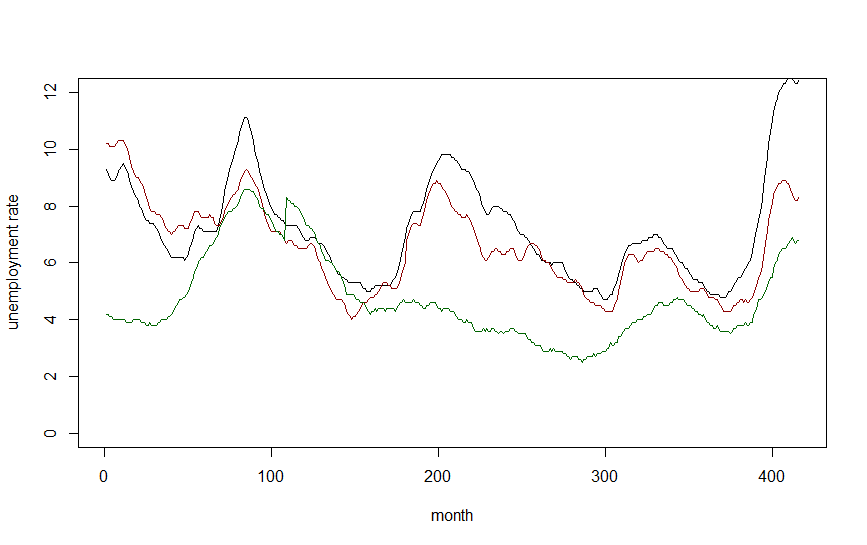
The following plot is designed to look at which countries have similar dietary preferences by using kmeans. This specific plot uses three clusters to classify countries.



The following graph is the same except it has been created with a model of 7 clusters instead of 3:

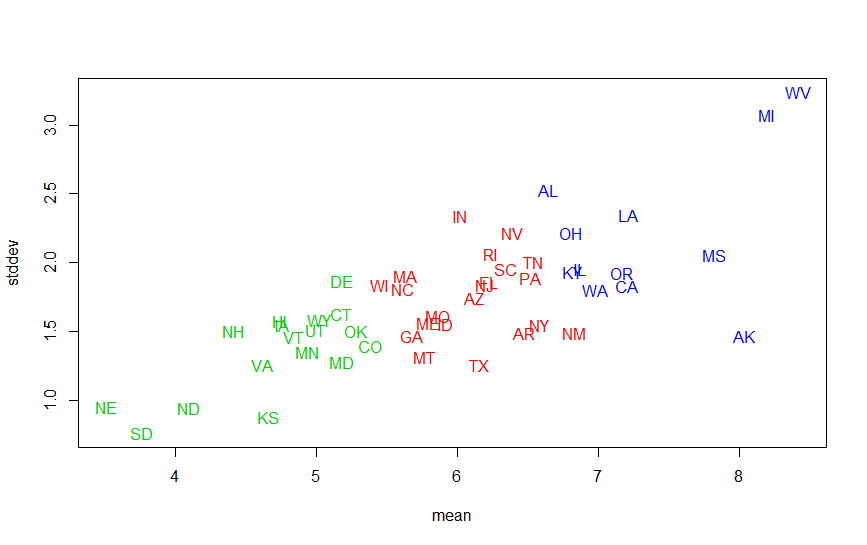


Example 2:



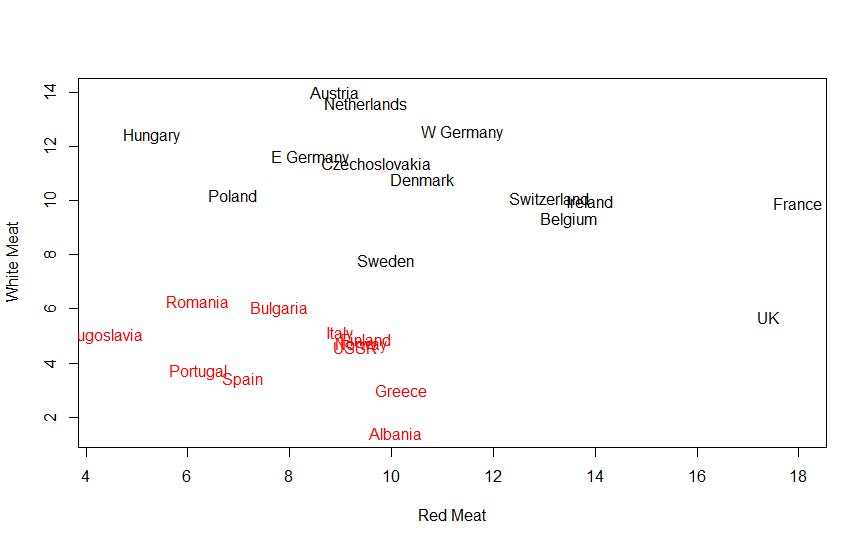
Unemployment rate in tree states to show trends. After Identifying state unemployment rates move differently. After this the script generate several different groupings of clusters.

The script then looks to cluster elements by the standard deviation and the average unemployment. This model creates a clustering as follows:



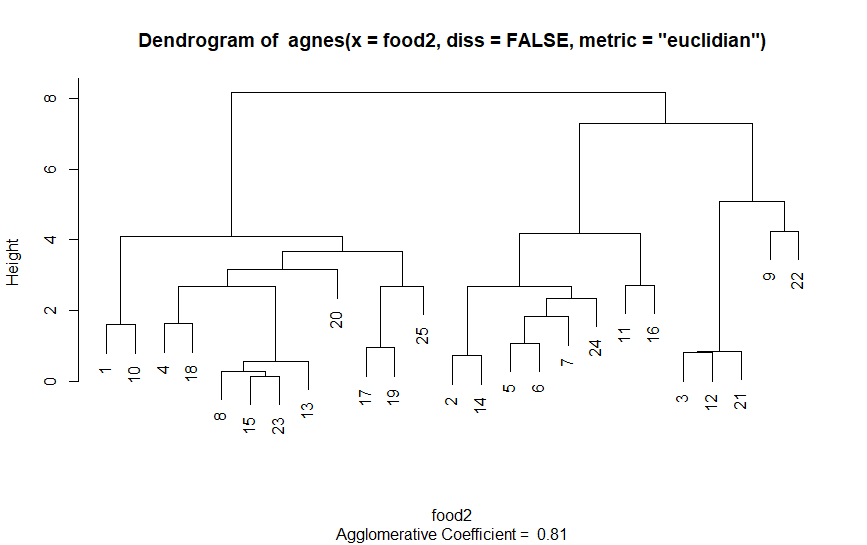
Example 3:

This example is again dealing with the protein data set. This model creates a binary model where a country is either a heavy consumer of red or white meat.



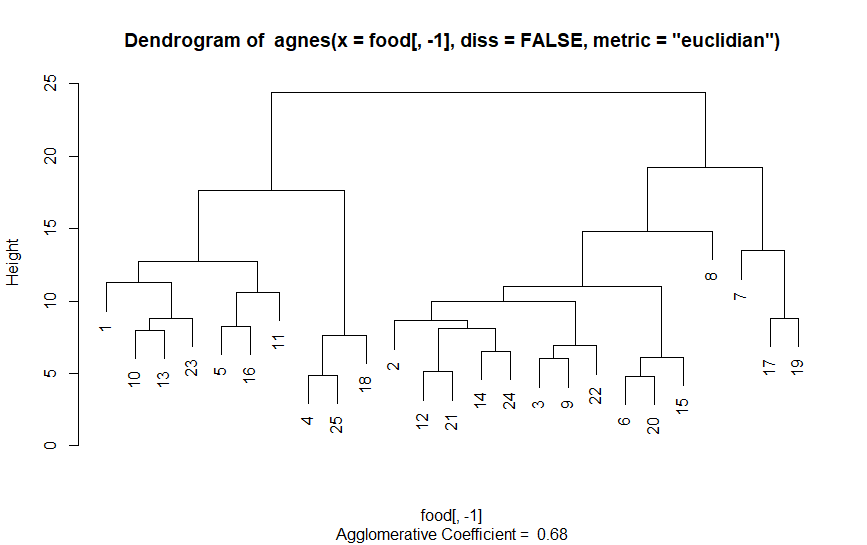
Example 4:

This example uses hierarchy clustering in order to determine clusters which are more likely to prefer red or white meat:

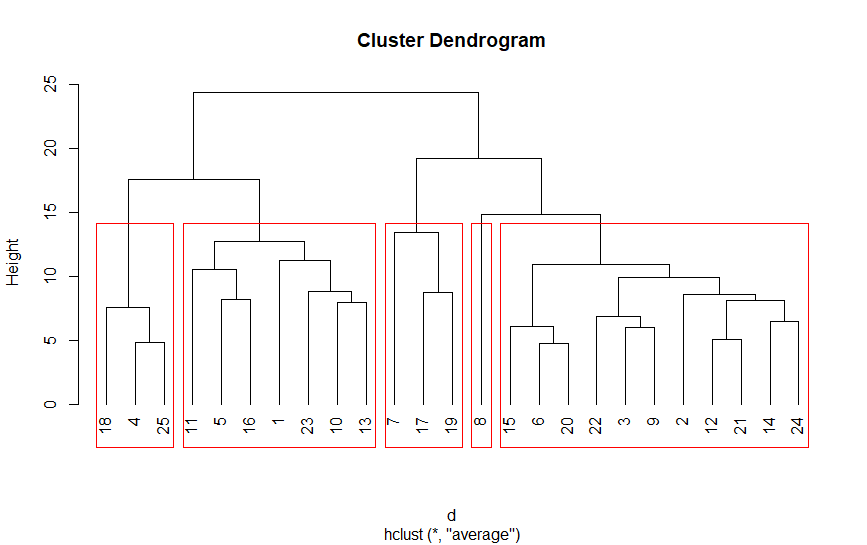


The above demonstrates the hierarchy structure when just red and white meat are considered as parameters

The below graph illustrates the hierarchy created when all foods are taken into account instead of just red and white meats:



After finding an optimal clustering level and the hierarchy structure the clustering results in the following graph:



Example 5:

This final example takes the unemployment data set and runs it through a hierarchy clustering model in order to cluster states into unemployment buckets.

