Multivariate exploration of the questionnaire and typology of the surveyed people

The results are provided by the
EnQuireR package

November 4, 2009

Contents

List of Figures

1 Quick overview of the questionnaire

The analysis was performed on 300 individuals described by 18 variables:

- breakfast (breakfast , no breakfast)
- afternoon.tea (afternoon tea , no afternoon tea)
- evening (evening , no evening)
- after.lunch (after lunch , no after lunch)
- after.dinner (after dinner , no after dinner)
- anytime (anytime , no anytime)
- home (home , no home)
- work (no work , work)
- tea.house (no tea house , tea house)
- friends (friends , no friends)
- restaurant (no restaurant , restaurant)
- pub (no pub , pub)
- tea.type (black , green , perfumed)
- how (lemon , milk , other , pure)
- sugar (no sugar , sugar)
- shape (bulk , tea bag , tea bag+bulk)
- location.of.purchase (large-scale retail stores , large-scale retail stores+specialized sho specialized shop)
- price (downmarket-p , famous-brand-p , supermarket-p , unknown-p , up-scale-p , variable-p)

Moreover, the dataset contained 0% of missing values.

2 Multivariate exploration of the questionnaire

2.1 Graphical representations of the questionnaire

The following results are obtained by performing a Multiple Correspondence Analysis (MCA) on the previous 18 variables. This method provides two important graphical displays, a representation of the individuals (surveyed people) and a representation of the categories (answers given by the surveyed people). The first two main axes of variability explain 17.99% of the information contained in the dataset (9.88% for the first factorial axis and 8.1% for the second one). In some cases the analyst may want to introduce supplementary quantitative variables.

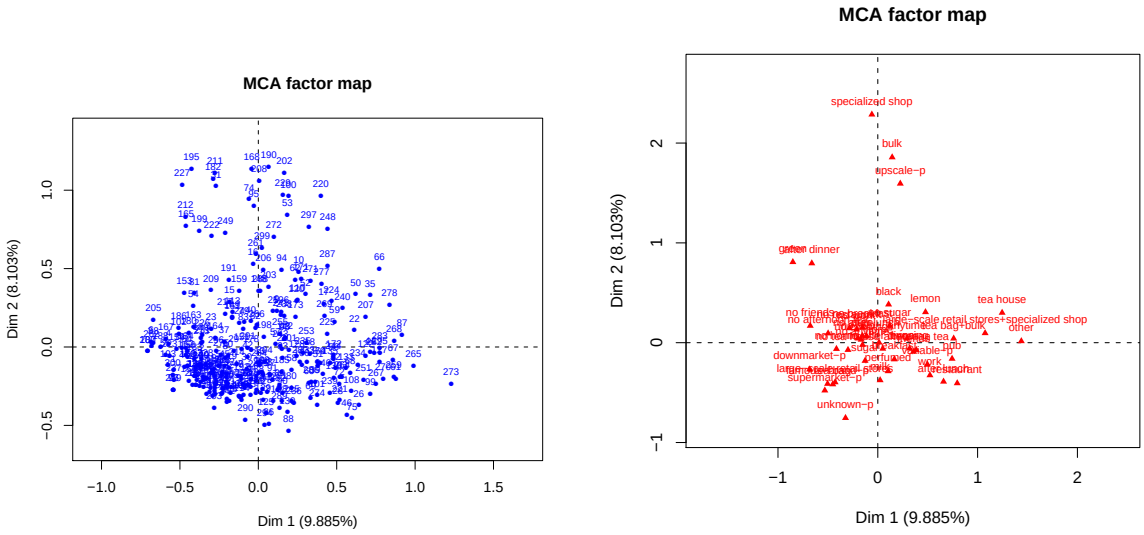

Figure 1: Representations of the individuals and of the categories on axes 1 and 2

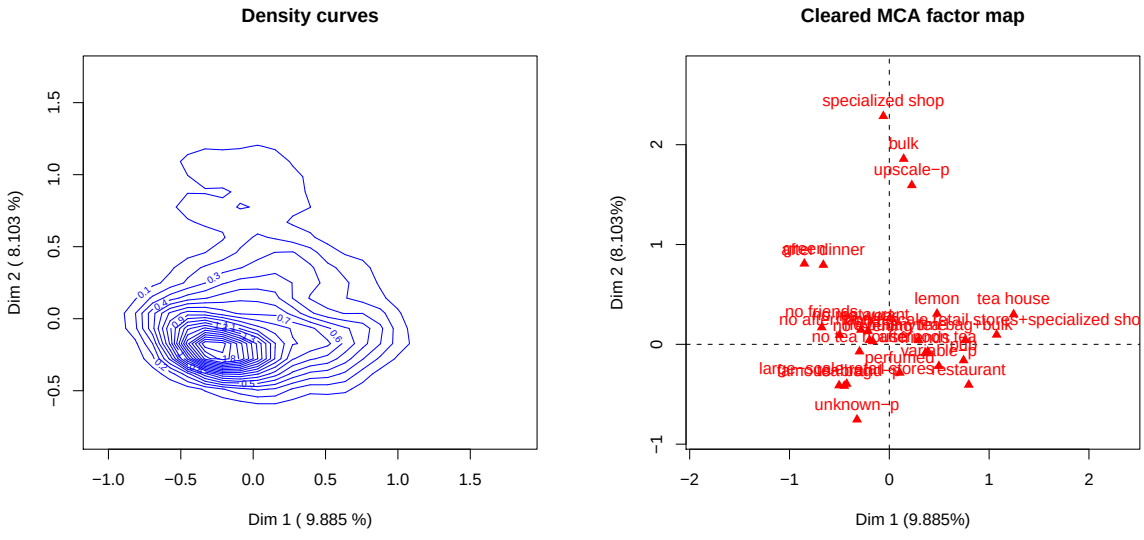

Figure 2: Representation of the individuals using density curves and enhanced representation of the categories

2.2 Highlights on the two principal axes of variability

2.2.1 Characterization of the first factorial axis

The most meaningful variables characterizing the first factorial axis are:

- `location.of.purchase`
- `tea.house`
- `shape`
- `friends`
- `restaurant`
- `afternoon.tea`
- `price`
- `pub`
- `work`
- `how`
- `tea.type`
- `after.lunch`
- `evening`
- `anytime`
- `after.dinner`
- `breakfast`
- `sugar`

The most meaningful categories characterizing the positive side of the first axis are:

- `tea house`
  - Contribution: 11.24
  - V-Test: 10.54
  - Frequency in the population: 19.33 %
- `large-scale retail stores+specialized shop`
  - Contribution: 11.26
  - V-Test: 11.02
  - Frequency in the population: 26 %
- `friends`
  - Contribution: 3.16
  - V-Test: 8.53
  - Frequency in the population: 65.33 %
- `restaurant`
  - Contribution: 6.25
  - V-Test: 8.23

  - Frequency in the population: 26.33 %

- `afternoon tea`
  - Contribution: 3.14
  - V-Test: 7.58
  - Frequency in the population: 56.33 %
- `tea bag+bulk`
  - Contribution: 6.79
  - V-Test: 8.88
  - Frequency in the population: 31.33 %
- `pub`
  - Contribution: 4.36
  - V-Test: 6.63
  - Frequency in the population: 21 %
- `work`
  - Contribution: 2.97
  - V-Test: 5.77
  - Frequency in the population: 29 %
- `variable-p`
  - Contribution: 3.45
  - V-Test: 6.63
  - Frequency in the population: 37.33 %
- `after lunch`
  - Contribution: 2.38
  - V-Test: 4.72
  - Frequency in the population: 14.67 %

The most meaningful categories characterizing the negative side of the first axis are:

- `no tea house`
  - Contribution: 2.69
  - V-Test: -10.54
  - Frequency in the population: 80.67 %
- `no friends`
  - Contribution: 5.95
  - V-Test: -8.53
  - Frequency in the population: 34.67 %
- `large-scale retail stores`
  - Contribution: 4.38
  - V-Test: -9.86
  - Frequency in the population: 64 %
- `no restaurant`

  - Contribution: 2.23
  - V-Test: -8.23
  - Frequency in the population: 73.67 %

- `tea bag`
  - Contribution: 4.32
  - V-Test: -8.92
  - Frequency in the population: 56.67 %

- `no afternoon tea`
  - Contribution: 4.05
  - V-Test: -7.58
  - Frequency in the population: 43.67 %

- `no pub`
  - Contribution: 1.16
  - V-Test: -6.63
  - Frequency in the population: 79 %

- `no work`
  - Contribution: 1.21
  - V-Test: -5.77
  - Frequency in the population: 71 %

- `green`
  - Contribution: 2.98
  - V-Test: -5.17
  - Frequency in the population: 11 %

- `no after lunch`
  - Contribution: 0.41
  - V-Test: -4.72
  - Frequency in the population: 85.33 %

2.2.2 Characterization on the second factorial axis

The most meaningful variables characterizing the second factorial axis are:

- `location.of.purchase`
- `price`
- `shape`
- `tea.type`
- `restaurant`
- `after.dinner`
- `work`
- `sugar`
- `how`

- after.lunch
- breakfast
- tea.house
- friends

The most meaningful categories characterizing the positive side of the second axis are:

- specialized shop
  - Contribution: 23.89
  - V-Test: 13.18
  - Frequency in the population: 10 %
- upscale-p
  - Contribution: 20.51
  - V-Test: 12.77
  - Frequency in the population: 17.67 %
- bulk
  - Contribution: 18.92
  - V-Test: 11.86
  - Frequency in the population: 12 %
- green
  - Contribution: 3.28
  - V-Test: 4.91
  - Frequency in the population: 11 %
- no restaurant
  - Contribution: 0.71
  - V-Test: 4.19
  - Frequency in the population: 73.67 %
- after dinner
  - Contribution: 2.03
  - V-Test: 3.77
  - Frequency in the population: 7 %
- no work
  - Contribution: 0.57
  - V-Test: 3.6
  - Frequency in the population: 71 %
- no sugar
  - Contribution: 0.68
  - V-Test: 3.03
  - Frequency in the population: 51.67 %
- no after lunch

  - Contribution: 0.17
  - V-Test: 2.79
  - Frequency in the population: 85.33 %
- `no breakfast`
  - Contribution: 0.56
  - V-Test: 2.76
  - Frequency in the population: 52 %

The most meaningful categories characterizing the negative side of the second axis are:

- `large-scale retail stores`
  - Contribution: 4.59
  - V-Test: -9.13
  - Frequency in the population: 64 %
- `tea bag`
  - Contribution: 4.49
  - V-Test: -8.24
  - Frequency in the population: 56.67 %
- `large-scale retail stores+specialized shop`
  - Contribution: 0.11
  - V-Test: 0.98
  - Frequency in the population: 26 %
- `tea bag+bulk`
  - Contribution: 0.03
  - V-Test: 0.49
  - Frequency in the population: 31.33 %
- `perfumed`
  - Contribution: 2.39
  - V-Test: -6.62
  - Frequency in the population: 64.33 %
- `restaurant`
  - Contribution: 1.98
  - V-Test: -4.19
  - Frequency in the population: 26.33 %
- `unknown-p`
  - Contribution: 1.04
  - V-Test: -2.66
  - Frequency in the population: 4 %
- `no after dinner`
  - Contribution: 0.15

  - V-Test: -3.77
  - Frequency in the population: 93 %
- `famous-brand-p`
  - Contribution: 2.46
  - V-Test: -4.85
  - Frequency in the population: 31.67 %
- `work`
  - Contribution: 1.41
  - V-Test: -3.6
  - Frequency in the population: 29 %

3 Typology on the individuals

3.1 Choice of the number of clusters

The ascendant hierarchical clustering (AHC) lead to a partition made of 3 clusters. Those clusters are displayed in the following representations: a graphical representation of the individuals according to the cluster they belong to, a representation of the center of gravity of each group enhanced by a confidence ellipse, a representation of the individuals according to the cluster they belong to by the use of density curves.

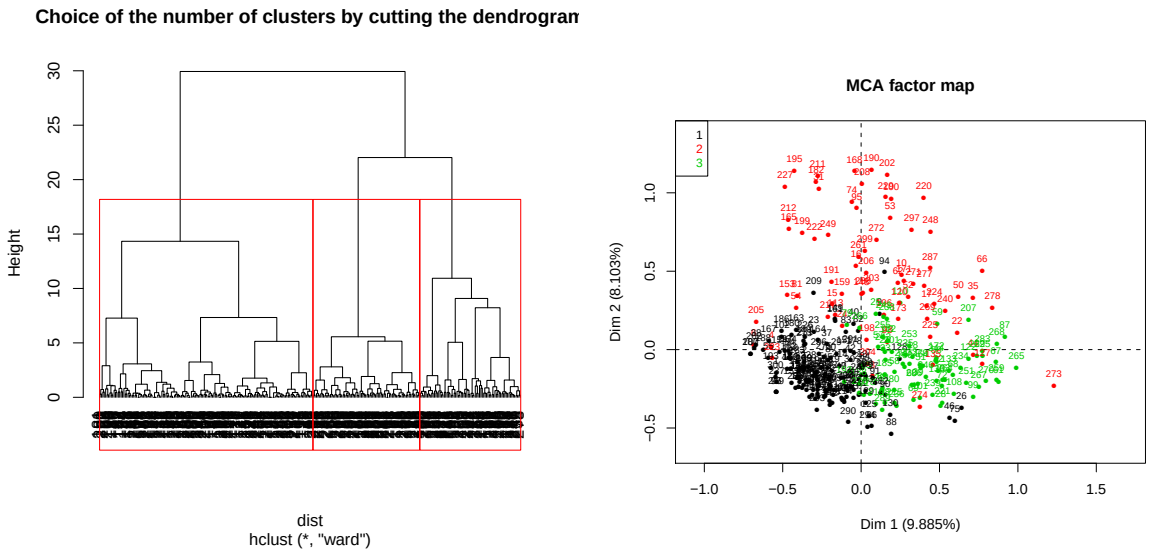

Figure 3: Number of clusters chosen by the analyst; representation of the individuals according to their cluster

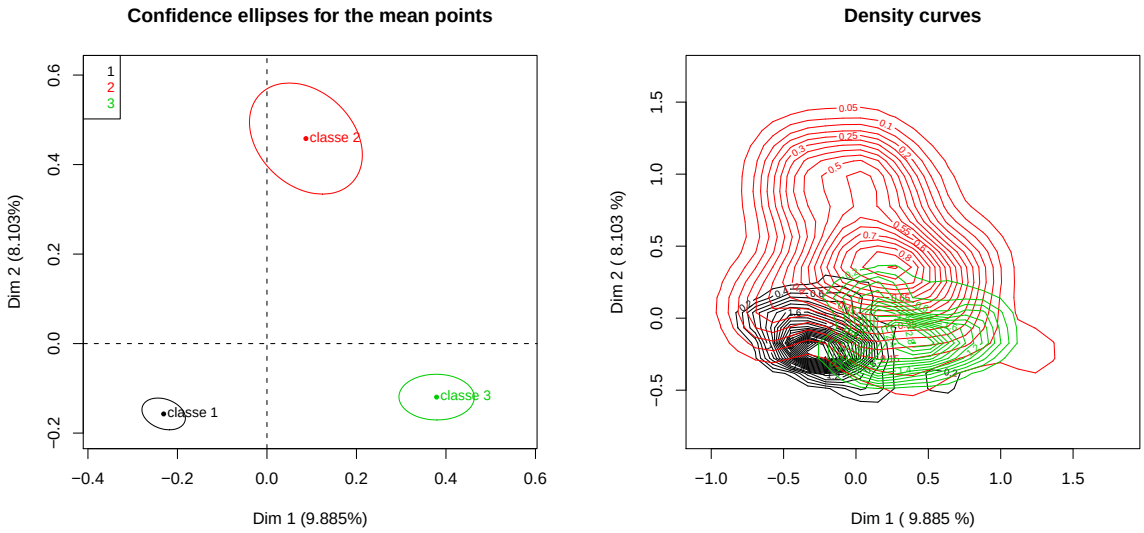

Figure 4: Centers of gravity with confidence ellipses; representation of the individuals according to their cluster with density curves

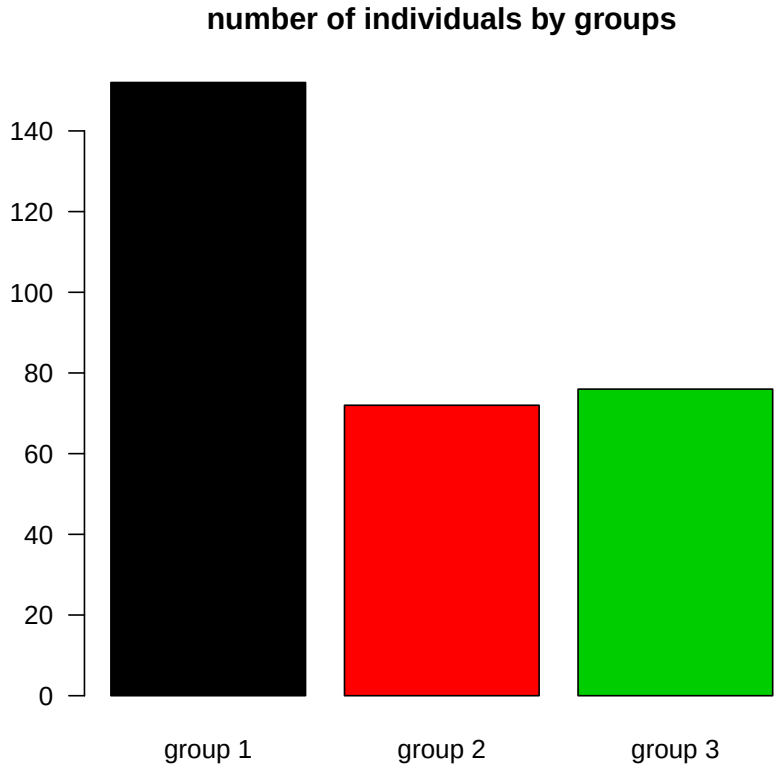

Figure 5: Number of individuals per cluster

3.2 Simultaneous comparison of the clusters with respect with the most relevant variables

3.2.1 Number of individuals by cluster for the variable price

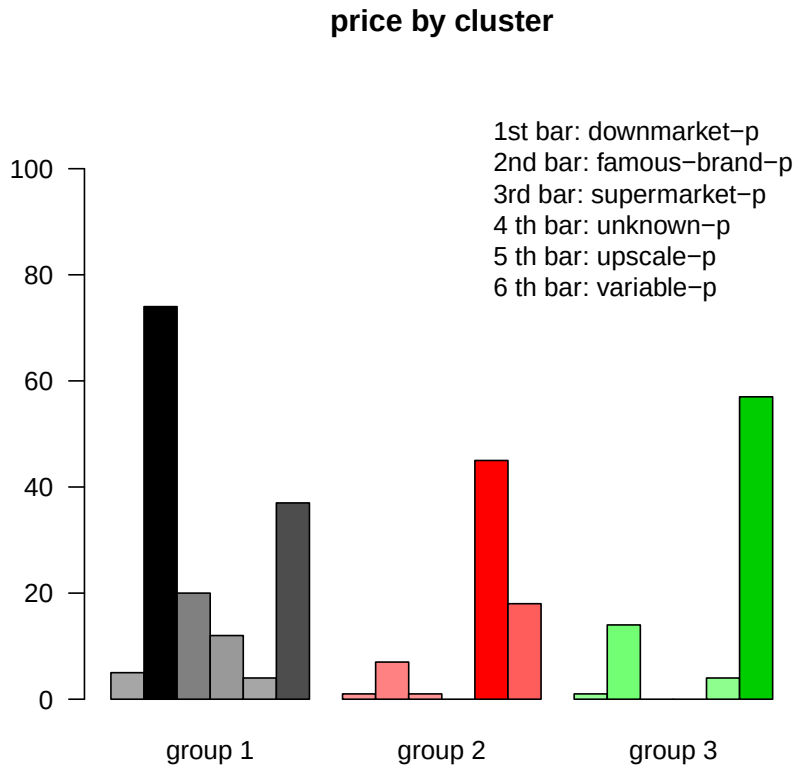

Figure 6: Variable price

3.2.2 Number of individuals by cluster for the variable location.of.purchase

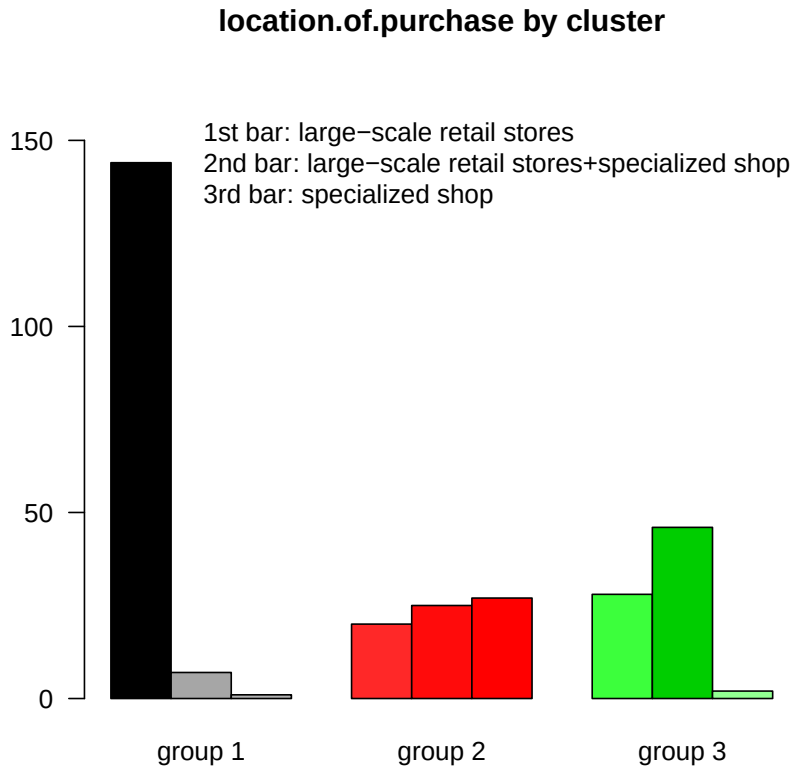

Figure 7: Variable location.of.purchase

3.2.3 Number of individuals by cluster for the variable shape

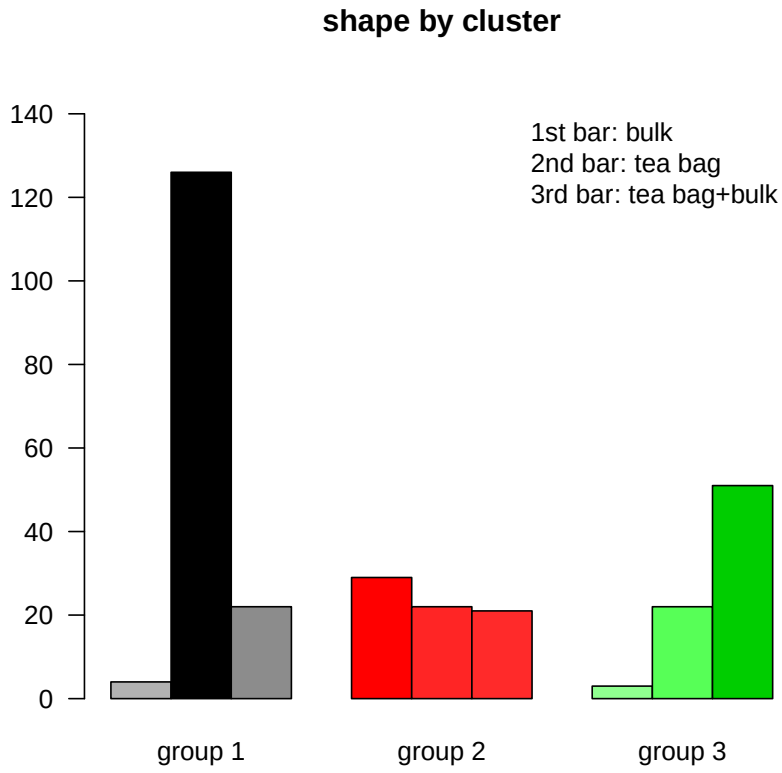

Figure 8: Variable shape

3.2.4 Number of individuals by cluster for the variable after.lunch

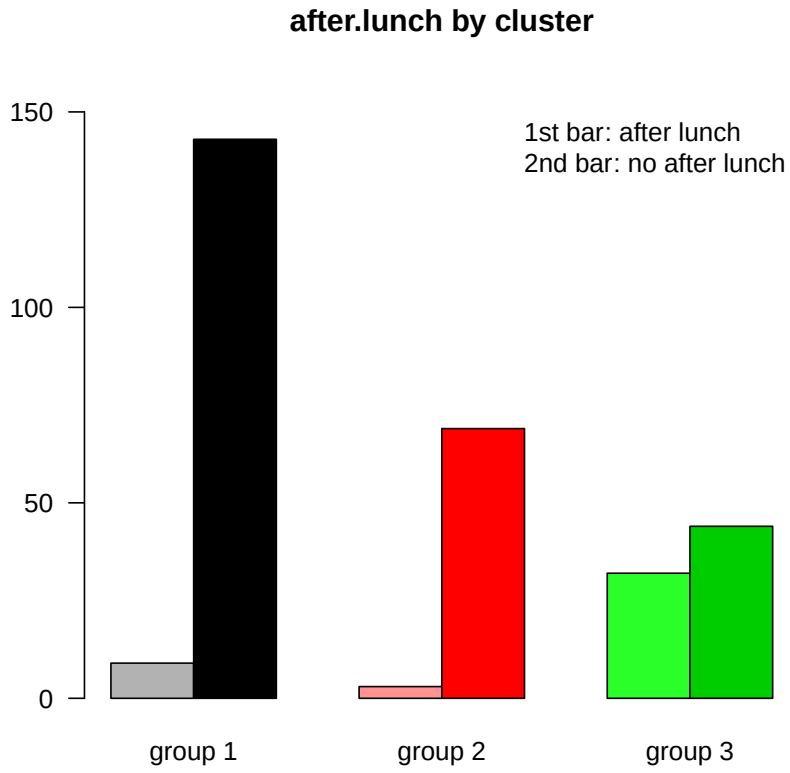

Figure 9: Variable after.lunch

3.2.5 Number of individuals by cluster for the variable after.dinner

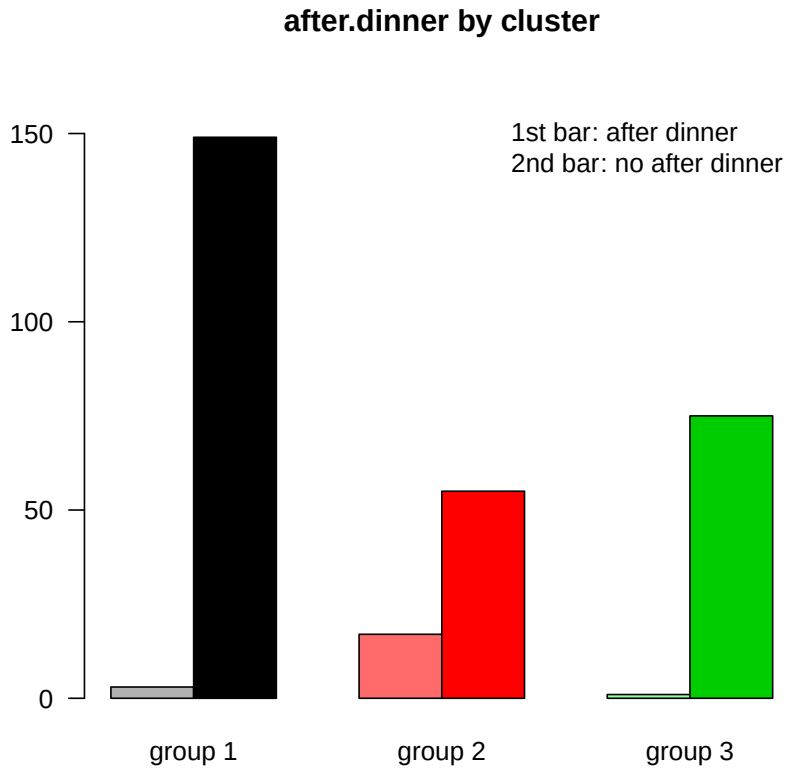

Figure 10: Variable after.dinner

3.2.6 Number of individuals by cluster for the variable tea.type

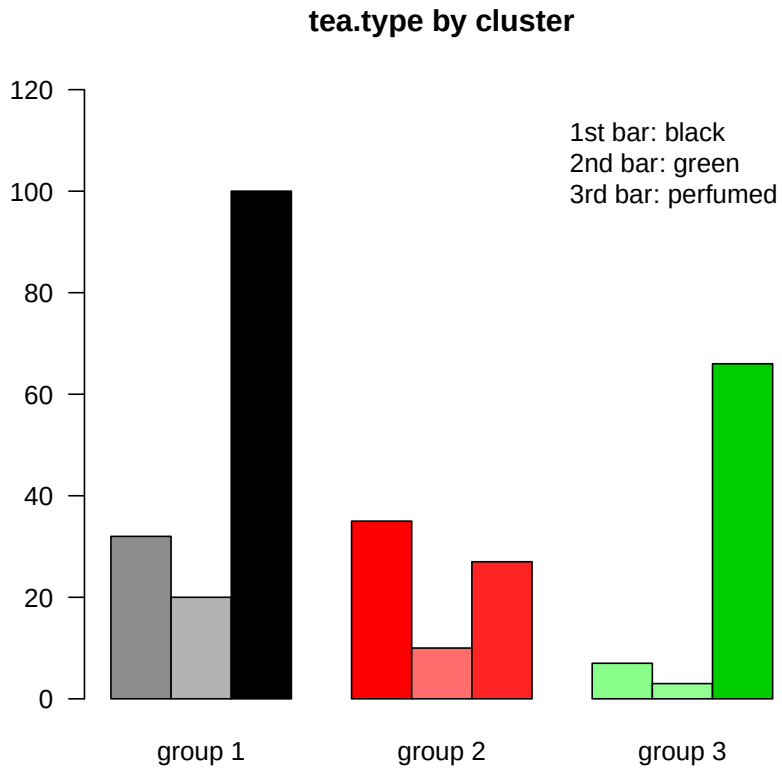

Figure 11: Variable tea.type

3.2.7 Number of individuals by cluster for the variable tea.house

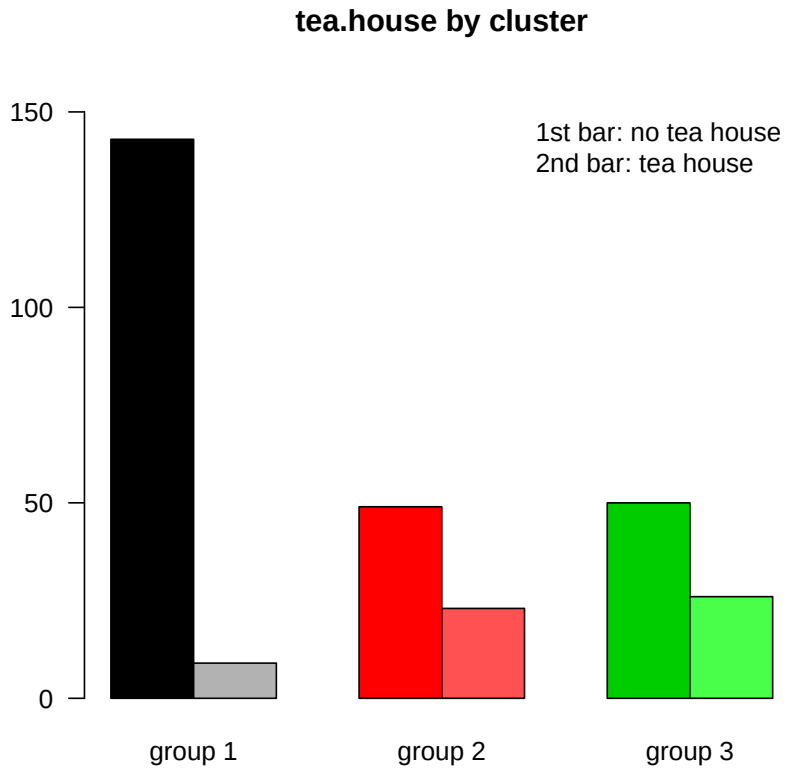

Figure 12: Variable tea.house

3.2.8 Number of individuals by cluster for the variable friends

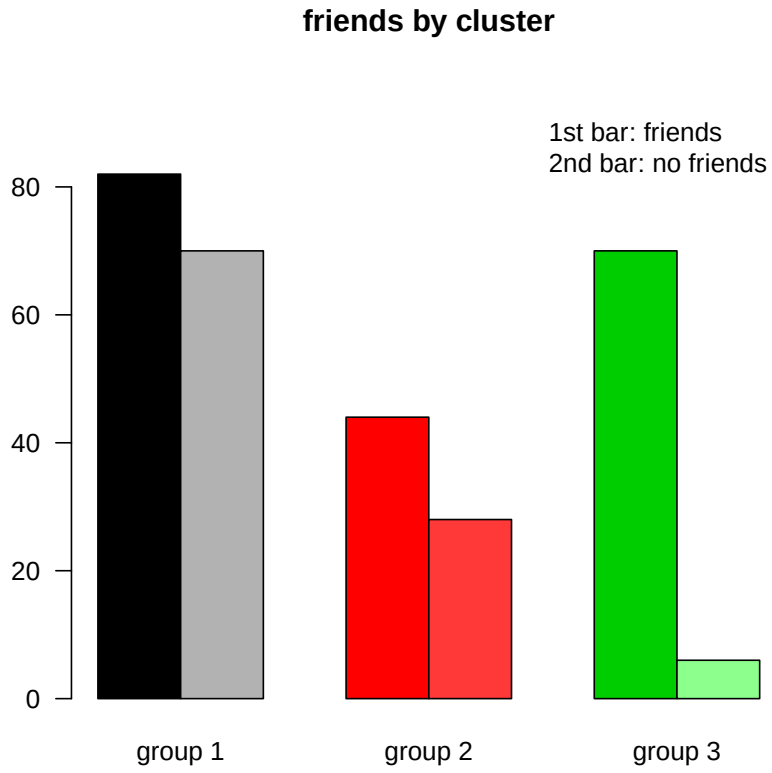

Figure 13: Variable friends

3.2.9 Number of individuals by cluster for the variable how

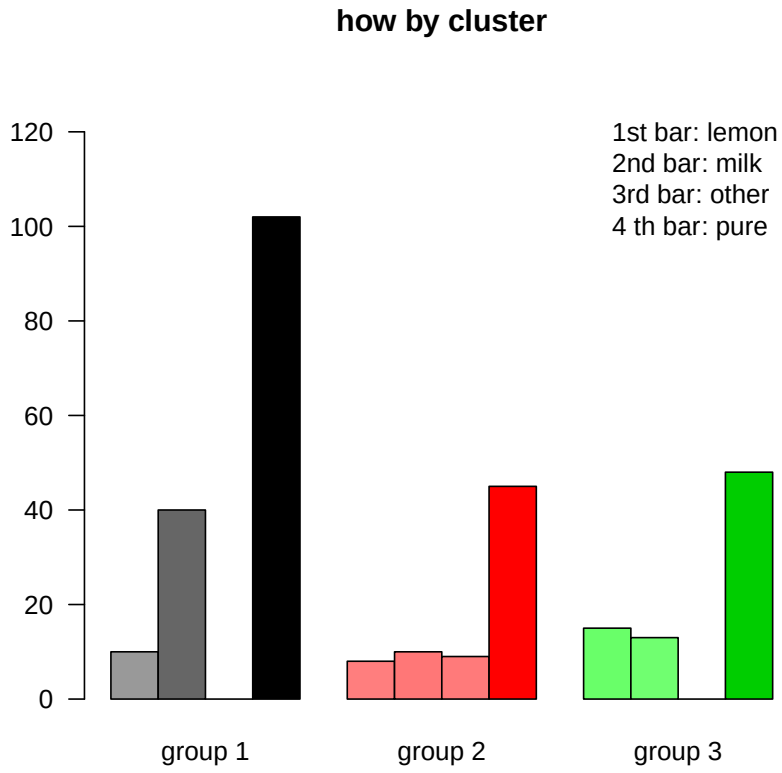

Figure 14: Variable how

3.2.10 Number of individuals by cluster for the variable pub

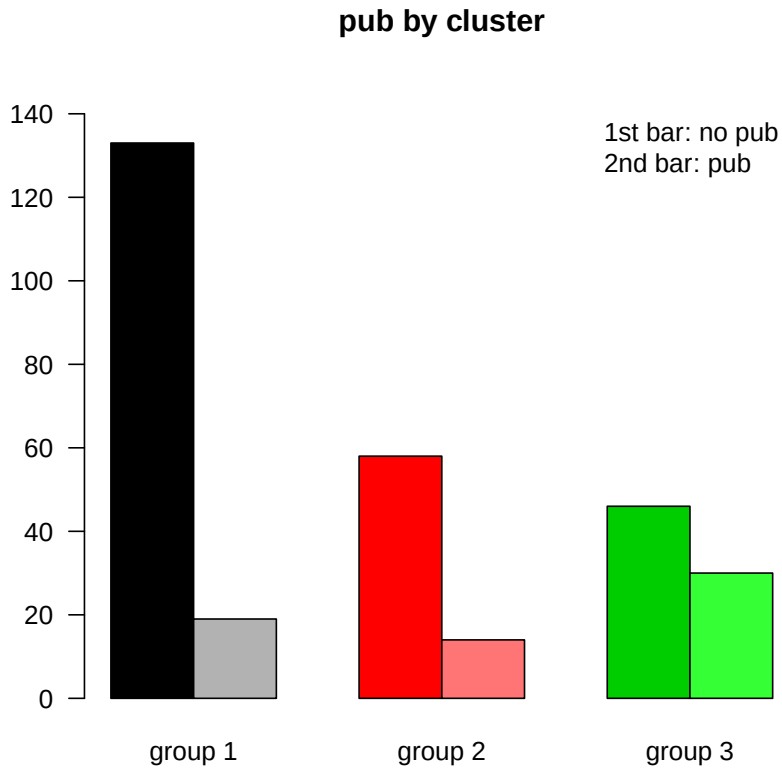

Figure 15: Variable pub

3.3 Automatic description of each cluster

The cluster 1 (152 individuals) includes the individuals possessing the following categories:

- `location.of.purchase=large-scale retail stores`

  64 % of the individuals possess this category in the global population versus 94.74 % in the cluster 1 .

  Moreover, 75 % of the individuals possessing this category belong to the cluster 1 .

- `shape=tea bag`

  56.67 % of the individuals possess this category in the global population versus 82.89 % in the cluster 1 .

  Moreover, 74.12 % of the individuals possessing this category belong to the cluster 1 .

- `price=famous-brand-p`

  31.67 % of the individuals possess this category in the global population versus 48.68 % in the cluster 1 .

  Moreover, 77.89 % of the individuals possessing this category belong to the cluster 1 .

- `tea.house=no tea house`

  80.67 % of the individuals possess this category in the global population versus 94.08 % in the cluster 1 .

  Moreover, 59.09 % of the individuals possessing this category belong to the cluster 1 .

- `price=supermarket-p`

  7 % of the individuals possess this category in the global population versus 13.16 % in the cluster 1 .

  Moreover, 95.24 % of the individuals possessing this category belong to the cluster 1 .

- `after.lunch=no after lunch`

  85.33 % of the individuals possess this category in the global population versus 94.08 % in the cluster 1 .

  Moreover, 55.86 % of the individuals possessing this category belong to the cluster 1 .

- `friends=no friends`

  34.67 % of the individuals possess this category in the global population versus 46.05 % in the cluster 1 .

  Moreover, 67.31 % of the individuals possessing this category belong to the cluster 1 .

- `pub=no pub`

  79 % of the individuals possess this category in the global population versus 87.5 % in the cluster 1 .

  Moreover, 56.12 % of the individuals possessing this category belong to the cluster 1 .

- `price=unknown-p`

  4 % of the individuals possess this category in the global population versus 7.89 % in the cluster 1 .

  Moreover, 100 % of the individuals possessing this category belong to the cluster 1 .

- `after.dinner=no after dinner`

  93 % of the individuals possess this category in the global population versus 98.03 % in the cluster 1 .

  Moreover, 53.41 % of the individuals possessing this category belong to the cluster 1 .

The cluster 2 (72 individuals) includes the individuals possessing the following categories:

- `price=upscale-p`

  17.67 % of the individuals possess this category in the global population versus 62.5 % in the cluster 2 .

  Moreover, 84.91 % of the individuals possessing this category belong to the cluster 2 .

- `location.of.purchase=specialized shop`

  10 % of the individuals possess this category in the global population versus 37.5 % in the cluster 2 .

  Moreover, 90 % of the individuals possessing this category belong to the cluster 2 .

- `shape=bulk`

  12 % of the individuals possess this category in the global population versus 40.28 % in the cluster 2 .

  Moreover, 80.56 % of the individuals possessing this category belong to the cluster 2 .

- `after.dinner=after dinner`

  7 % of the individuals possess this category in the global population versus 23.61 % in the cluster 2 .

  Moreover, 80.95 % of the individuals possessing this category belong to the cluster 2 .

- `tea.type=black`

  24.67 % of the individuals possess this category in the global population versus 48.61 % in the cluster 2 .

  Moreover, 47.3 % of the individuals possessing this category belong to the cluster 2 .

- `how=other`

  3 % of the individuals possess this category in the global population versus 12.5 % in the cluster 2 .

  Moreover, 100 % of the individuals possessing this category belong to the cluster 2 .

- `after.lunch=no after lunch`

  85.33 % of the individuals possess this category in the global population versus 95.83 % in the cluster 2 .

  Moreover, 26.95 % of the individuals possessing this category belong to the cluster 2 .

- `tea.house=tea house`

  19.33 % of the individuals possess this category in the global population versus 31.94 % in the cluster 2 .

  Moreover, 39.66 % of the individuals possessing this category belong to the cluster 2 .

- `sugar=no sugar`

  51.67 % of the individuals possess this category in the global population versus 63.89 % in the cluster 2 .

  Moreover, 29.68 % of the individuals possessing this category belong to the cluster 2 .

The cluster 3 (76 individuals) includes the individuals possessing the following categories:

- `price=variable-p`

  37.33 % of the individuals possess this category in the global population versus 75 % in the cluster 3 .

  Moreover, 50.89 % of the individuals possessing this category belong to the cluster 3 .

- `location.of.purchase=large-scale retail stores+specialized shop`

  26 % of the individuals possess this category in the global population versus 60.53 % in the cluster 3 .

  Moreover, 58.97 % of the individuals possessing this category belong to the cluster 3 .

- `shape=tea bag+bulk`

  31.33 % of the individuals possess this category in the global population versus 67.11 % in the cluster 3 .

  Moreover, 54.26 % of the individuals possessing this category belong to the cluster 3 .

- `after.lunch=after lunch`

  14.67 % of the individuals possess this category in the global population versus 42.11 % in the cluster 3 .

  Moreover, 72.73 % of the individuals possessing this category belong to the cluster 3 .

- `friends=friends`

  65.33 % of the individuals possess this category in the global population versus 92.11 % in the cluster 3 .

  Moreover, 35.71 % of the individuals possessing this category belong to the cluster 3 .

- `tea.type=perfumed`

  64.33 % of the individuals possess this category in the global population versus 86.84 % in the cluster 3 .

  Moreover, 34.2 % of the individuals possessing this category belong to the cluster 3 .

- `pub=pub`

  21 % of the individuals possess this category in the global population versus 39.47 % in the cluster 3 .

  Moreover, 47.62 % of the individuals possessing this category belong to the cluster 3 .

- `tea.house=tea house`

  19.33 % of the individuals possess this category in the global population versus 34.21 % in the cluster 3 .

  Moreover, 44.83 % of the individuals possessing this category belong to the cluster 3 .

- `restaurant=restaurant`

  26.33 % of the individuals possess this category in the global population versus 42.11 % in the cluster 3 .

  Moreover, 40.51 % of the individuals possessing this category belong to the cluster 3 .

- `work=work`

  29 % of the individuals possess this category in the global population versus 42.11 % in the cluster 3 .

  Moreover, 36.78 % of the individuals possessing this category belong to the cluster 3 .