STAT 331 FINAL PROJECT

```
library("faraway")
protein_train <- read.csv('protein-train.csv')

##EDA

# check for NAs
na.list = sapply(protein_train, function(x){sum(is.na(x))})
na.list[na.list>0]
```

```
## named integer(0)
```

```
#summary of original dataset
head(protein_train)
```

```
##   accuracy   angles carbonylC_carbonylC_medshort carbonylC_carbonylC_medlong
## 1   7.9264 117.9882                            1                           0
## 2   7.7580 113.8061                            1                           0
## 3   7.9628 119.8208                            1                           0
## 4   6.4585 115.7349                            1                           0
## 5   9.9714 126.5727                            1                           0
## 6   8.1091 117.0780                            1                           0
##   carbonylC_carbonylC_long carbonylC_carbonylC_vlong
## 1                        0                         0
## 2                        0                         0
## 3                        0                         1
## 4                        0                         0
## 5                        0                         0
## 6                        0                         0
##   carbonylC_carboxylC_medshort carbonylC_carboxylC_long
## 1                            2                        0
## 2                            0                        0
## 3                            1                        0
## 4                            0                        0
## 5                            1                        0
## 6                            0                        0
##   carbonylC_carboxylC_vlong carbonylC_aliph1HC_medshort
## 1                         0                           0
## 2                         0                           0
## 3                         0                           0
## 4                         0                           1
## 5                         0                           2
## 6                         0                           0
##   carbonylC_aliph1HC_medlong carbonylC_aliph1HC_long carbonylC_aliph1HC_vlong
## 1                          2                       2                        0
## 2                          3                       1                        1
## 3                          2                       1                        3
## 4                          0                       0                        2
```

```
## 5                          2                         2                          3
## 6                          2                         1                          2
##   carbonylC_aliph2HC_short carbonylC_aliph2HC_medshort
## 1                        2                          5
## 2                        1                          3
## 3                        1                          4
## 4                        2                          4
## 5                        0                          4
## 6                        2                          5
##   carbonylC_aliph2HC_medlong carbonylC_aliph2HC_long carbonylC_aliph2HC_vlong
## 1                          5                        6                       14
## 2                          2                        4                       12
## 3                          6                        9                       10
## 4                          3                        5                       10
## 5                          3                        1                        9
## 6                          3                        4                        5
##   carbonylC_aliph3HC_short carbonylC_aliph3HC_medshort
## 1                        0                          3
## 2                        1                          1
## 3                        0                          2
## 4                        0                          1
## 5                        1                          4
## 6                        0                          3
##   carbonylC_aliph3HC_medlong carbonylC_aliph3HC_long carbonylC_aliph3HC_vlong
## 1                          2                        3                       1
## 2                          2                        4                       2
## 3                          3                        2                       3
## 4                          2                        0                       4
## 5                          3                        4                       2
## 6                          0                        3                       4
##   carbonylC_aromaticC_short carbonylC_aromaticC_medshort
## 1                         2                           6
## 2                         0                           5
## 3                         1                           7
## 4                         0                           2
## 5                         0                           2
## 6                         0                           2
##   carbonylC_aromaticC_medlong carbonylC_aromaticC_long
## 1                           7                       11
## 2                           7                        7
## 3                           8                        9
## 4                           6                        7
## 5                           3                        4
## 6                           4                        4
##   carbonylC_aromaticC_vlong carbonylC_scAGN_short carbonylC_scAGN_medshort
## 1                         7                     0                       2
## 2                         8                     0                       2
## 3                        10                     0                       2
## 4                        11                     0                       2
## 5                        11                     0                       2
## 6                        10                     0                       1
##   carbonylC_scAGN_medlong carbonylC_scAGN_long carbonylC_scAGN_vlong
## 1                       0                    0                    1
## 2                       0                    0                    0
```

```
## 3                       0                    1                     0
## 4                       0                    0                     0
## 5                       0                    0                     0
## 6                       1                    0                     0
##   carbonylC_scLysN_short carbonylC_scLysN_medlong carbonylC_scLysN_long
## 1                      0                       1                     0
## 2                      0                       0                     1
## 3                      0                       0                     1
## 4                      0                       0                     0
## 5                      0                       0                     0
## 6                      0                       1                     0
##   carbonylC_scLysN_vlong carbonylC_bbProN_medlong carbonylC_bbProN_long
## 1                      0                        0                     1
## 2                      0                        0                     1
## 3                      0                        0                     1
## 4                      0                        0                     0
## 5                      1                        0                     1
## 6                      0                        1                     0
##   carbonylC_bbProN_vlong carbonylC_hydroxylO_short carbonylC_hydroxylO_medshort
## 1                      0                         3                            0
## 2                      1                         1                            4
## 3                      2                         3                            0
## 4                      1                         2                            2
## 5                      0                         1                            2
## 6                      0                         3                            1
##   carbonylC_hydroxylO_medlong carbonylC_hydroxylO_long
## 1                           1                        3
## 2                           1                        2
## 3                           1                        5
## 4                           0                        1
## 5                           1                        1
## 6                           0                        3
##   carbonylC_hydroxylO_vlong carbonylC_carbonylO_short
## 1                         2                        0
## 2                         3                        1
## 3                         1                        0
## 4                         6                        1
## 5                         6                        0
## 6                         3                        1
##   carbonylC_carbonylO_medshort carbonylC_carbonylO_medlong
## 1                            0                           2
## 2                            0                           1
## 3                            1                           1
## 4                            0                           1
## 5                            1                           1
## 6                            0                           1
##   carbonylC_carbonylO_long carbonylC_carbonylO_vlong carbonylC_carboxylO_short
## 1                        0                         0                        1
## 2                        0                         0                        0
## 3                        0                         2                        1
## 4                        0                         0                        0
## 5                        0                         0                        1
## 6                        0                         0                        0
##   carbonylC_carboxylO_medshort carbonylC_carboxylO_medlong
```

```
## 1                           0                         3
## 2                           0                         0
## 3                           0                         1
## 4                           0                         0
## 5                           0                         1
## 6                           0                         0
##   carbonylC_carboxylO_long carbonylC_carboxylO_vlong carbonylC_sulfur_short
## 1                        0                         0                     0
## 2                        0                         2                     0
## 3                        0                         1                     0
## 4                        0                         1                     0
## 5                        0                         0                     1
## 6                        0                         2                     0
##   carbonylC_sulfur_medshort carbonylC_sulfur_medlong carbonylC_sulfur_long
## 1                         0                        1                    1
## 2                         0                        2                    1
## 3                         0                        1                    0
## 4                         0                        2                    1
## 5                         0                        0                    0
## 6                         0                        1                    0
##   carbonylC_sulfur_vlong carbonylC_bbN_short carbonylC_bbN_medshort
## 1                      1                   0                     2
## 2                      0                   0                     3
## 3                      0                   0                     2
## 4                      1                   0                     2
## 5                      2                   0                     3
## 6                      1                   0                     3
##   carbonylC_bbN_medlong carbonylC_bbN_long carbonylC_bbN_vlong
## 1                     7                  9                  10
## 2                     6                  5                  10
## 3                     6                  9                   9
## 4                     1                  6                   6
## 5                     4                 11                   9
## 6                     4                  5                  10
##   carbonylC_bbCA_medshort carbonylC_bbCA_medlong carbonylC_bbCA_long
## 1                       2                      7                   8
## 2                       4                      3                   8
## 3                       2                      5                   8
## 4                       1                      3                   5
## 5                       3                      6                   8
## 6                       4                      6                   4
##   carbonylC_bbCA_vlong carbonylC_bbC_medshort carbonylC_bbC_medlong
## 1                    8                      3                     3
## 2                    8                      1                     6
## 3                   14                      4                     4
## 4                    5                      1                     1
## 5                    7                      1                     6
## 6                    7                      2                     3
##   carbonylC_bbC_long carbonylC_bbC_vlong carbonylC_bbO_short
## 1                 11                  12                   2
## 2                  7                  10                   1
## 3                  8                  14                   2
## 4                  7                  11                   1
## 5                 10                  11                   1
```

```
## 6                  9                   6                   3
##   carbonylC_bb0_medshort carbonylC_bb0_medlong carbonylC_bb0_long
## 1                      2                    2                 8
## 2                      3                    2                 6
## 3                      2                    3                 7
## 4                      0                    1                 6
## 5                      3                    2                10
## 6                      0                    4                 5
##   carbonylC_bb0_vlong carboxylC_carboxylC_vlong carboxylC_aliph2HC_short
## 1                  17                         0                       0
## 2                  16                         1                       1
## 3                  14                         1                       0
## 4                  16                         0                       0
## 5                  15                         0                       0
## 6                  13                         1                       1
##   carboxylC_aliph2HC_medshort carboxylC_aliph2HC_medlong
## 1                           2                          2
## 2                           1                          0
## 3                           5                          1
## 4                           2                          0
## 5                           2                          2
## 6                           1                          0
##   carboxylC_aliph2HC_long carboxylC_aliph2HC_vlong carboxylC_aliph3HC_medlong
## 1                       4                        5                         1
## 2                       0                        2                         0
## 3                       0                        2                         1
## 4                       0                        1                         0
## 5                       1                        2                         0
## 6                       0                        2                         0
##   carboxylC_aliph3HC_vlong carboxylC_aromaticC_medshort
## 1                        0                            0
## 2                        0                            0
## 3                        0                            2
## 4                        0                            0
## 5                        1                            0
## 6                        0                            0
##   carboxylC_aromaticC_medlong carboxylC_aromaticC_long
## 1                           2                       1
## 2                           0                       0
## 3                           2                       2
## 4                           0                       0
## 5                           0                       0
## 6                           0                       0
##   carboxylC_aromaticC_vlong carboxylC_scAGN_short carboxylC_scAGN_long
## 1                         3                     0                   0
## 2                         0                     0                   0
## 3                         0                     1                   0
## 4                         0                     0                   0
## 5                         0                     1                   0
## 6                         0                     0                   0
##   carboxylC_scAGN_vlong carboxylC_scLysN_long carboxylC_scLysN_vlong
## 1                     0                     0                     0
## 2                     0                     0                     0
## 3                     0                     0                     1
```

```
## 4                     0                     0                     0
## 5                     0                     0                     1
## 6                     0                     0                     0
##   carboxylC_scArgN_vlong carboxylC_hydroxylO_short carboxylC_hydroxylO_medshort
## 1                      0                        1                            0
## 2                      2                        1                            1
## 3                      1                        1                            1
## 4                      0                        1                            0
## 5                      0                        1                            0
## 6                      2                        1                            1
##   carboxylC_hydroxylO_medlong carboxylC_hydroxylO_vlong
## 1                           0                        1
## 2                           0                        0
## 3                           0                        0
## 4                           1                        0
## 5                           1                        1
## 6                           0                        0
##   carboxylC_carbonylO_medshort carboxylC_carbonylO_medlong
## 1                            1                           1
## 2                            0                           0
## 3                            0                           1
## 4                            0                           0
## 5                            1                           0
## 6                            0                           0
##   carboxylC_carbonylO_vlong carboxylC_carboxylO_medlong
## 1                         0                           0
## 2                         0                           0
## 3                         0                           0
## 4                         0                           0
## 5                         0                           0
## 6                         0                           0
##   carboxylC_carboxylO_long carboxylC_carboxylO_vlong carboxylC_sulfur_medshort
## 1                        0                         0                         0
## 2                        1                         3                         0
## 3                        0                         4                         1
## 4                        0                         2                         0
## 5                        0                         2                         1
## 6                        1                         3                         0
##   carboxylC_sulfur_vlong carboxylC_bbN_medshort carboxylC_bbN_medlong
## 1                      1                      1                    2
## 2                      0                      1                    1
## 3                      0                      1                    0
## 4                      0                      1                    0
## 5                      1                      1                    1
## 6                      0                      1                    1
##   carboxylC_bbN_long carboxylC_bbN_vlong carboxylC_bbCA_medshort
## 1                  2                   3                      1
## 2                  1                   1                      2
## 3                  6                   2                      2
## 4                  2                   1                      2
## 5                  2                   3                      1
## 6                  1                   0                      2
##   carboxylC_bbCA_medlong carboxylC_bbCA_long carboxylC_bbCA_vlong
## 1                      1                   2                    4
```

```
## 2                      0                    0                    1
## 3                      2                    2                    1
## 4                      0                    0                    0
## 5                      2                    2                    2
## 6                      0                    0                    1
##   carboxylC_bbC_medshort carboxylC_bbC_medlong carboxylC_bbC_long
## 1                      0                    2                   2
## 2                      1                    1                   1
## 3                      1                    2                   3
## 4                      1                    1                   1
## 5                      0                    2                   2
## 6                      1                    1                   1
##   carboxylC_bbC_vlong carboxylC_bbO_medshort carboxylC_bbO_medlong
## 1                   2                      0                     2
## 2                   1                      0                     2
## 3                   3                      1                     2
## 4                   0                      0                     2
## 5                   2                      0                     2
## 6                   0                      0                     2
##   carboxylC_bbO_long carboxylC_bbO_vlong aliph1HC_aliph1HC_medlong
## 1                  2                   1                         1
## 2                  2                   0                         1
## 3                  2                   4                         1
## 4                  1                   2                         1
## 5                  1                   4                         2
## 6                  1                   1                         1
##   aliph1HC_aliph1HC_long aliph1HC_aliph1HC_vlong aliph1HC_aliph2HC_short
## 1                      0                       1                       0
## 2                      1                       1                       0
## 3                      0                       1                       0
## 4                      1                       1                       0
## 5                      1                       4                       0
## 6                      0                       1                       0
##   aliph1HC_aliph2HC_medshort aliph1HC_aliph2HC_medlong aliph1HC_aliph2HC_long
## 1                          0                        7                      9
## 2                          2                        6                      7
## 3                          0                        9                     12
## 4                          2                        4                      7
## 5                          4                       13                     16
## 6                          0                        5                     13
##   aliph1HC_aliph2HC_vlong aliph1HC_aliph3HC_short aliph1HC_aliph3HC_medshort
## 1                       7                       0                         1
## 2                      15                       1                         2
## 3                       8                       0                         1
## 4                      12                       0                         2
## 5                       9                       0                         3
## 6                      11                       0                         2
##   aliph1HC_aliph3HC_medlong aliph1HC_aliph3HC_long aliph1HC_aliph3HC_vlong
## 1                         1                      3                       1
## 2                         2                      4                       4
## 3                         1                      4                       0
## 4                         4                      1                       5
## 5                         3                      7                       7
## 6                         1                      3                       5
```

```
##   aliph1HC_aromaticC_short aliph1HC_aromaticC_medshort
## 1                        0                           4
## 2                        0                           4
## 3                        0                           3
## 4                        0                           5
## 5                        3                           9
## 6                        0                           3
##   aliph1HC_aromaticC_medlong aliph1HC_aromaticC_long aliph1HC_aromaticC_vlong
## 1                          3                       10                        8
## 2                          6                       10                       11
## 3                          4                       10                        9
## 4                          8                       13                        4
## 5                          8                       10                        8
## 6                          9                       14                       10
##   aliph1HC_scAGN_short aliph1HC_scAGN_medshort aliph1HC_scAGN_medlong
## 1                    1                       0                      1
## 2                    0                       1                      2
## 3                    1                       0                      1
## 4                    0                       0                      1
## 5                    1                       2                      0
## 6                    0                       1                      1
##   aliph1HC_scAGN_long aliph1HC_scAGN_vlong aliph1HC_scArgN_long
## 1                   0                    1                    2
## 2                   0                    1                    2
## 3                   0                    2                    2
## 4                   0                    0                    2
## 5                   2                    2                    2
## 6                   1                    0                    2
##   aliph1HC_bbProN_medlong aliph1HC_bbProN_long aliph1HC_bbProN_vlong
## 1                       0                    2                     0
## 2                       0                    2                     0
## 3                       0                    2                     0
## 4                       0                    1                     2
## 5                       1                    3                     1
## 6                       0                    3                     0
##   aliph1HC_hydroxylO_medshort aliph1HC_hydroxylO_medlong
## 1                           1                          0
## 2                           1                          0
## 3                           1                          1
## 4                           0                          1
## 5                           2                          2
## 6                           1                          0
##   aliph1HC_hydroxylO_long aliph1HC_hydroxylO_vlong aliph1HC_carbonylO_short
## 1                       2                       3                       0
## 2                       3                       3                       0
## 3                       2                       2                       0
## 4                       4                       0                       0
## 5                       3                       3                       2
## 6                       4                       3                       0
##   aliph1HC_carbonylO_medshort aliph1HC_carbonylO_medlong
## 1                           1                          1
## 2                           0                          1
## 3                           1                          0
## 4                           1                          0
```

```
## 5                          1                         0
## 6                          0                         0
##   aliph1HC_carbonylO_long aliph1HC_carbonylO_vlong aliph1HC_carboxylO_vlong
## 1                       1                       0                       1
## 2                       2                       1                       1
## 3                       2                       2                       1
## 4                       1                       0                       0
## 5                       3                       1                       1
## 6                       2                       1                       1
##   aliph1HC_sulfur_short aliph1HC_sulfur_medshort aliph1HC_sulfur_medlong
## 1                     0                       0                      2
## 2                     0                       1                      0
## 3                     0                       0                      1
## 4                     0                       0                      1
## 5                     0                       0                      3
## 6                     0                       0                      0
##   aliph1HC_sulfur_long aliph1HC_sulfur_vlong aliph1HC_bbN_short
## 1                    1                     0                 0
## 2                    1                     2                 0
## 3                    1                     0                 0
## 4                    2                     0                 0
## 5                    1                     1                 1
## 6                    2                     0                 0
##   aliph1HC_bbN_medshort aliph1HC_bbN_medlong aliph1HC_bbN_long
## 1                     0                    6                7
## 2                     1                    7                5
## 3                     1                    7                5
## 4                     0                    9                5
## 5                     0                    8               12
## 6                     0                    8                5
##   aliph1HC_bbN_vlong aliph1HC_bbCA_medshort aliph1HC_bbCA_medlong
## 1                  6                     0                    5
## 2                  7                     0                    6
## 3                  7                     0                    7
## 4                 12                     0                    8
## 5                 12                     1                    7
## 6                 12                     0                    4
##   aliph1HC_bbCA_long aliph1HC_bbCA_vlong aliph1HC_bbC_medshort
## 1                  7                  11                    2
## 2                  6                  12                    1
## 3                  5                  15                    2
## 4                  8                   9                    1
## 5                 15                  18                    2
## 6                 11                  14                    0
##   aliph1HC_bbC_medlong aliph1HC_bbC_long aliph1HC_bbC_vlong aliph1HC_bbO_short
## 1                    5                 3                 11                  0
## 2                    6                 4                  9                  0
## 3                    7                 4                  7                  0
## 4                    7                 9                  7                  0
## 5                    9                 9                 15                  2
## 6                    5                 8                 11                  0
##   aliph1HC_bbO_medshort aliph1HC_bbO_medlong aliph1HC_bbO_long
## 1                     4                   3                5
## 2                     4                   1                3
```

```
## 3                    5                   5                2
## 4                    4                   5                6
## 5                    4                   4                8
## 6                    4                   1                5
##   aliph1HC_bbO_vlong aliph2HC_aliph2HC_short aliph2HC_aliph2HC_medshort
## 1                 7                       1                          8
## 2                13                       1                          6
## 3                 8                       0                         10
## 4                 9                       0                          7
## 5                14                       0                          7
## 6                15                       1                          9
##   aliph2HC_aliph2HC_medlong aliph2HC_aliph2HC_long aliph2HC_aliph2HC_vlong
## 1                        23                     30                      39
## 2                        15                     23                      36
## 3                        22                     23                      34
## 4                        20                     21                      39
## 5                        11                     23                      32
## 6                        22                     24                      42
##   aliph2HC_aliph3HC_vshort aliph2HC_aliph3HC_short aliph2HC_aliph3HC_medshort
## 1                        0                       4                         4
## 2                        1                       1                         4
## 3                        0                       1                         2
## 4                        0                       2                         4
## 5                        1                       1                         9
## 6                        0                       4                         4
##   aliph2HC_aliph3HC_medlong aliph2HC_aliph3HC_long aliph2HC_aliph3HC_vlong
## 1                        11                     14                     13
## 2                        12                     19                     25
## 3                        17                     15                     16
## 4                         7                     18                     21
## 5                        21                     23                     25
## 6                        13                     14                     21
##   aliph2HC_aromaticC_short aliph2HC_aromaticC_medshort
## 1                        3                          22
## 2                        4                          18
## 3                        4                          23
## 4                        2                          17
## 5                        2                          13
## 6                        6                          28
##   aliph2HC_aromaticC_medlong aliph2HC_aromaticC_long aliph2HC_aromaticC_vlong
## 1                         39                     50                     72
## 2                         36                     47                     72
## 3                         36                     50                     60
## 4                         32                     49                     74
## 5                         24                     41                     67
## 6                         31                     44                     63
##   aliph2HC_scAGN_vshort aliph2HC_scAGN_short aliph2HC_scAGN_medshort
## 1                     0                    3                       5
## 2                     0                    1                       2
## 3                     0                    1                       5
## 4                     0                    1                       3
## 5                     0                    1                       3
## 6                     1                    2                       3
##   aliph2HC_scAGN_medlong aliph2HC_scAGN_long aliph2HC_scAGN_vlong
```

```
## 1                      8                   2                  10
## 2                      2                   8                  11
## 3                      7                   6                   8
## 4                      4                   7                   7
## 5                      3                   1                   8
## 6                      4                   2                   5
##   aliph2HC_scLysN_medshort aliph2HC_scLysN_medlong aliph2HC_scLysN_long
## 1                        2                       1                    1
## 2                        0                       2                    0
## 3                        2                       0                    1
## 4                        1                       1                    1
## 5                        0                       2                    0
## 6                        0                       1                    1
##   aliph2HC_scLysN_vlong aliph2HC_scArgN_medshort aliph2HC_scArgN_medlong
## 1                     1                       2                       1
## 2                     2                       2                       1
## 3                     2                       2                       1
## 4                     1                       2                       1
## 5                     2                       2                       0
## 6                     2                       2                       1
##   aliph2HC_scArgN_long aliph2HC_scArgN_vlong aliph2HC_bbProN_medshort
## 1                    1                     5                       0
## 2                    0                     4                       0
## 3                    1                     5                       0
## 4                    0                     4                       1
## 5                    2                     5                       2
## 6                    0                     4                       1
##   aliph2HC_bbProN_medlong aliph2HC_bbProN_long aliph2HC_bbProN_vlong
## 1                       3                    1                     3
## 2                       2                    3                     3
## 3                       3                    2                     3
## 4                       1                    3                     3
## 5                       0                    2                     4
## 6                       2                    8                     6
##   aliph2HC_hydroxylO_short aliph2HC_hydroxylO_medshort
## 1                        1                           5
## 2                        1                           4
## 3                        0                           7
## 4                        1                           4
## 5                        1                           5
## 6                        0                           5
##   aliph2HC_hydroxylO_medlong aliph2HC_hydroxylO_long aliph2HC_hydroxylO_vlong
## 1                          6                      12                      17
## 2                          7                      13                      16
## 3                          6                      12                      18
## 4                          6                      13                      20
## 5                          6                      18                      21
## 6                         10                      11                      25
##   aliph2HC_carbonylO_vshort aliph2HC_carbonylO_short
## 1                         0                        3
## 2                         1                        2
## 3                         0                        2
## 4                         0                        5
## 5                         0                        1
```

```
## 6                          1                          2
##   aliph2HC_carbonylO_medshort aliph2HC_carbonylO_medlong
## 1                           3                          3
## 2                           0                          3
## 3                           3                          5
## 4                           0                          4
## 5                           3                          2
## 6                           3                          3
##   aliph2HC_carbonylO_long aliph2HC_carbonylO_vlong aliph2HC_carboxylO_short
## 1                       6                        9                       2
## 2                       2                        8                       3
## 3                       7                        7                       4
## 4                       2                        8                       2
## 5                       3                        6                       1
## 6                       4                        5                       3
##   aliph2HC_carboxylO_medshort aliph2HC_carboxylO_medlong
## 1                           1                          3
## 2                           1                          0
## 3                           5                          2
## 4                           1                          1
## 5                           3                          3
## 6                           1                          0
##   aliph2HC_carboxylO_long aliph2HC_carboxylO_vlong aliph2HC_sulfur_short
## 1                       8                        9                     2
## 2                       0                        5                     0
## 3                       2                        9                     1
## 4                       1                        3                     2
## 5                       1                        8                     1
## 6                       1                        4                     0
##   aliph2HC_sulfur_medshort aliph2HC_sulfur_medlong aliph2HC_sulfur_long
## 1                        2                       0                    3
## 2                        2                       5                    7
## 3                        2                       0                    3
## 4                        1                       3                    3
## 5                        3                       4                    3
## 6                        1                       2                    8
##   aliph2HC_sulfur_vlong aliph2HC_bbN_short aliph2HC_bbN_medshort
## 1                     7                  1                   19
## 2                     9                  1                   16
## 3                     4                  1                   16
## 4                     9                  1                   15
## 5                     3                  0                   17
## 6                     6                  1                   14
##   aliph2HC_bbN_medlong aliph2HC_bbN_long aliph2HC_bbN_vlong aliph2HC_bbCA_short
## 1                   30                45                 64                    1
## 2                   35                41                 56                    2
## 3                   29                42                 66                    0
## 4                   31                43                 55                    1
## 5                   26                42                 64                    0
## 6                   27                45                 65                    0
##   aliph2HC_bbCA_medshort aliph2HC_bbCA_medlong aliph2HC_bbCA_long
## 1                     15                   44                42
## 2                     10                   39                44
## 3                     15                   43                42
```

```
## 4                     9                   39                44
## 5                    12                   38                37
## 6                    13                   40                47
##   aliph2HC_bbCA_vlong aliph2HC_bbC_short aliph2HC_bbC_medshort
## 1                 61                  1                   20
## 2                 62                  1                   14
## 3                 54                  0                   19
## 4                 56                  0                   16
## 5                 58                  1                   17
## 6                 60                  2                   17
##   aliph2HC_bbC_medlong aliph2HC_bbC_long aliph2HC_bbC_vlong aliph2HC_bbO_vshort
## 1                   32                55                 54                   1
## 2                   36                43                 58                   1
## 3                   31                48                 56                   0
## 4                   26                59                 43                   0
## 5                   33                39                 62                   0
## 6                   31                47                 66                   1
##   aliph2HC_bbO_short aliph2HC_bbO_medshort aliph2HC_bbO_medlong
## 1                 12                   13                  33
## 2                 10                   13                  23
## 3                 10                   12                  32
## 4                  8                    9                  34
## 5                 11                   12                  29
## 6                 10                   15                  24
##   aliph2HC_bbO_long aliph2HC_bbO_vlong aliph3HC_aliph3HC_short
## 1                39                74                       1
## 2                43                49                       2
## 3                38                65                       1
## 4                39                62                       0
## 5                40                64                       2
## 6                47                60                       1
##   aliph3HC_aliph3HC_medshort aliph3HC_aliph3HC_medlong aliph3HC_aliph3HC_long
## 1                          0                         1                      0
## 2                          3                         1                      2
## 3                          0                         1                      1
## 4                          4                         1                      4
## 5                          1                         3                      5
## 6                          0                         5                      1
##   aliph3HC_aliph3HC_vlong aliph3HC_aromaticC_short aliph3HC_aromaticC_medshort
## 1                       4                       2                          4
## 2                       5                       3                          8
## 3                       3                       2                          7
## 4                       3                       2                         10
## 5                       8                      11                         19
## 6                       4                       2                         11
##   aliph3HC_aromaticC_medlong aliph3HC_aromaticC_long aliph3HC_aromaticC_vlong
## 1                         19                      13                      29
## 2                         24                      15                      28
## 3                         20                      11                      27
## 4                         18                      19                      16
## 5                         14                      23                      23
## 6                         22                      26                      22
##   aliph3HC_scAGN_short aliph3HC_scAGN_medshort aliph3HC_scAGN_medlong
## 1                    1                       0                      4
```

```
## 2                    1                      1                      3
## 3                    1                      0                      3
## 4                    0                      1                      0
## 5                    2                      3                      2
## 6                    1                      0                      3
##   aliph3HC_scAGN_long aliph3HC_scAGN_vlong aliph3HC_scLysN_vlong
## 1                   2                    4                     0
## 2                   3                    2                     0
## 3                   3                    2                     0
## 4                   3                    1                     0
## 5                   2                    5                     0
## 6                   1                    3                     0
##   aliph3HC_scArgN_medlong aliph3HC_scArgN_long aliph3HC_scArgN_vlong
## 1                       1                    1                     1
## 2                       1                    1                     1
## 3                       1                    1                     1
## 4                       0                    2                     0
## 5                       1                    1                     1
## 6                       1                    1                     1
##   aliph3HC_bbProN_medlong aliph3HC_bbProN_long aliph3HC_bbProN_vlong
## 1                       0                    3                     1
## 2                       1                    2                     3
## 3                       0                    2                     2
## 4                       1                    1                     1
## 5                       2                    3                     4
## 6                       1                    1                     2
##   aliph3HC_hydroxylO_short aliph3HC_hydroxylO_medshort
## 1                        0                           2
## 2                        0                           1
## 3                        0                           3
## 4                        0                           1
## 5                        2                           3
## 6                        1                           2
##   aliph3HC_hydroxylO_medlong aliph3HC_hydroxylO_long aliph3HC_hydroxylO_vlong
## 1                          2                       3                       6
## 2                          2                       4                       8
## 3                          2                       5                       5
## 4                          2                       4                       4
## 5                          4                       7                       5
## 6                          2                       6                       7
##   aliph3HC_carbonylO_short aliph3HC_carbonylO_medshort
## 1                        2                          1
## 2                        0                          0
## 3                        1                          1
## 4                        1                          1
## 5                        0                          6
## 6                        0                          2
##   aliph3HC_carbonylO_medlong aliph3HC_carbonylO_long aliph3HC_carbonylO_vlong
## 1                          3                       0                       2
## 2                          3                       3                       5
## 3                          2                       2                       2
## 4                          1                       1                       3
## 5                          2                       4                       1
## 6                          1                       2                       4
```

```
##   aliph3HC_carboxylO_medshort aliph3HC_carboxylO_medlong
## 1                           1                          0
## 2                           0                          0
## 3                           1                          0
## 4                           0                          0
## 5                           0                          0
## 6                           0                          0
##   aliph3HC_carboxylO_long aliph3HC_carboxylO_vlong aliph3HC_sulfur_short
## 1                       1                       1                     0
## 2                       0                       1                     1
## 3                       1                       2                     0
## 4                       0                       0                     0
## 5                       0                       1                     1
## 6                       0                       1                     0
##   aliph3HC_sulfur_medshort aliph3HC_sulfur_medlong aliph3HC_sulfur_long
## 1                        2                       1                    1
## 2                        4                       0                    2
## 3                        1                       0                    3
## 4                        1                       2                    3
## 5                        1                       3                    3
## 6                        1                       2                    2
##   aliph3HC_sulfur_vlong aliph3HC_bbN_short aliph3HC_bbN_medshort
## 1                     3                  0                    3
## 2                     3                  1                    6
## 3                     1                  0                    2
## 4                     1                  0                    5
## 5                     1                  0                    6
## 6                     2                  0                    5
##   aliph3HC_bbN_medlong aliph3HC_bbN_long aliph3HC_bbN_vlong aliph3HC_bbCA_short
## 1                   10                14                 21                    0
## 2                   13                12                 23                    2
## 3                   10                12                 22                    0
## 4                   12                10                 24                    0
## 5                   11                20                 33                    0
## 6                    9                19                 28                    0
##   aliph3HC_bbCA_medshort aliph3HC_bbCA_medlong aliph3HC_bbCA_long
## 1                      6                    8                15
## 2                      3                   13                17
## 3                      4                    6                22
## 4                      6                   11                17
## 5                      9                   10                22
## 6                      6                   12                17
##   aliph3HC_bbCA_vlong aliph3HC_bbC_short aliph3HC_bbC_medshort
## 1                  21                 0                    5
## 2                  18                 1                    7
## 3                  14                 0                    5
## 4                  13                 0                    7
## 5                  37                 1                    7
## 6                  22                 0                    8
##   aliph3HC_bbC_medlong aliph3HC_bbC_long aliph3HC_bbC_vlong aliph3HC_bbO_short
## 1                    8                17                 21                   2
## 2                    8                17                 21                   4
## 3                    9                12                 22                   3
## 4                   10                16                 12                   4
```

```
## 5                   12                25                33                   5
## 6                    7                20                25                   4
##   aliph3HC_bbO_medshort aliph3HC_bbO_medlong aliph3HC_bbO_long
## 1                     8                    6                 9
## 2                     6                    6                14
## 3                     9                    4                 9
## 4                     8                    6                14
## 5                     9                    8                20
## 6                     8                    4                20
##   aliph3HC_bbO_vlong aromaticC_aromaticC_short aromaticC_aromaticC_medshort
## 1                 29                        5                           22
## 2                 20                        4                           13
## 3                 23                        6                           27
## 4                 18                        2                           13
## 5                 34                        8                           26
## 6                 21                        7                           25
##   aromaticC_aromaticC_medlong aromaticC_aromaticC_long
## 1                          50                       22
## 2                          31                       44
## 3                          43                       28
## 4                          36                       47
## 5                          60                       53
## 6                          36                       39
##   aromaticC_aromaticC_vlong aromaticC_scAGN_short aromaticC_scAGN_medshort
## 1                        23                     5                        5
## 2                        35                     3                        4
## 3                        22                     6                        5
## 4                        33                     0                        2
## 5                        24                     2                        2
## 6                        25                     0                        0
##   aromaticC_scAGN_medlong aromaticC_scAGN_long aromaticC_scAGN_vlong
## 1                      10                    9                     4
## 2                       4                    8                     3
## 3                       9                    9                     6
## 4                       4                    7                     7
## 5                       4                    7                     9
## 6                       2                   10                     4
##   aromaticC_scLysN_medlong aromaticC_scLysN_long aromaticC_scLysN_vlong
## 1                        0                     2                     2
## 2                        0                     2                     3
## 3                        0                     2                     2
## 4                        1                     2                     3
## 5                        0                     2                     2
## 6                        0                     2                     2
##   aromaticC_scArgN_medshort aromaticC_scArgN_medlong aromaticC_scArgN_long
## 1                         0                        0                     0
## 2                         0                        0                     0
## 3                         0                        0                     0
## 4                         0                        0                     0
## 5                         0                        0                     0
## 6                         0                        0                     0
##   aromaticC_scArgN_vlong aromaticC_bbProN_medlong aromaticC_bbProN_long
## 1                      0                        0                    1
## 2                      0                        2                    3
```

```
## 3                     0                       0                     1
## 4                     0                       0                     3
## 5                     0                       0                     2
## 6                     0                       3                     7
##   aromaticC_bbProN_vlong aromaticC_hydroxylO_vshort aromaticC_hydroxylO_short
## 1                     4                          0                         3
## 2                     2                          0                         2
## 3                     4                          0                         3
## 4                     3                          0                         2
## 5                     1                          0                         5
## 6                     4                          0                         2
##   aromaticC_hydroxylO_medshort aromaticC_hydroxylO_medlong
## 1                            5                           6
## 2                            5                           7
## 3                            5                           7
## 4                            2                           8
## 5                            7                           9
## 6                            3                          14
##   aromaticC_hydroxylO_long aromaticC_hydroxylO_vlong aromaticC_carbonylO_short
## 1                       14                        21                        1
## 2                       16                        20                        1
## 3                       12                        20                        1
## 4                       16                        21                        2
## 5                        9                        17                        0
## 6                       10                        20                        2
##   aromaticC_carbonylO_medshort aromaticC_carbonylO_medlong
## 1                            4                           8
## 2                            3                           6
## 3                            3                           8
## 4                            3                           4
## 5                            0                           3
## 6                            2                           4
##   aromaticC_carbonylO_long aromaticC_carbonylO_vlong
## 1                        7                       14
## 2                       10                        4
## 3                        9                       11
## 4                        8                        9
## 5                        5                        4
## 6                        4                        4
##   aromaticC_carboxylO_medshort aromaticC_carboxylO_medlong
## 1                            0                           1
## 2                            0                           0
## 3                            4                           3
## 4                            0                           0
## 5                            0                           0
## 6                            0                           0
##   aromaticC_carboxylO_long aromaticC_carboxylO_vlong aromaticC_sulfur_short
## 1                        4                        6                      1
## 2                        0                        1                      1
## 3                        2                        7                      1
## 4                        0                        0                      5
## 5                        0                        0                      0
## 6                        0                        1                      4
##   aromaticC_sulfur_medshort aromaticC_sulfur_medlong aromaticC_sulfur_long
```

```
## 1                        2                       2                     1
## 2                        5                       8                     8
## 3                        2                       2                     1
## 4                        4                       0                     5
## 5                        2                       1                     5
## 6                        2                       0                     0
##   aromaticC_sulfur_vlong aromaticC_bbN_short aromaticC_bbN_medshort
## 1                      5                    0                      4
## 2                      5                    1                     15
## 3                      4                    0                      5
## 4                      7                    2                      5
## 5                      4                    0                      9
## 6                      6                    0                     12
##   aromaticC_bbN_medlong aromaticC_bbN_long aromaticC_bbN_vlong
## 1                    34                 52                  80
## 2                    28                 50                  77
## 3                    29                 51                  82
## 4                    32                 57                  68
## 5                    24                 55                  73
## 6                    29                 60                  72
##   aromaticC_bbCA_short aromaticC_bbCA_medshort aromaticC_bbCA_medlong
## 1                    2                       8                     33
## 2                    4                      15                     22
## 3                    2                       8                     31
## 4                    2                       7                     35
## 5                    1                      12                     28
## 6                    3                      14                     33
##   aromaticC_bbCA_long aromaticC_bbCA_vlong aromaticC_bbC_short
## 1                  60                  68                   0
## 2                  59                  81                   0
## 3                  62                  71                   0
## 4                  57                  66                   0
## 5                  56                  62                   0
## 6                  64                  75                   0
##   aromaticC_bbC_medshort aromaticC_bbC_medlong aromaticC_bbC_long
## 1                     14                    28                 55
## 2                     13                    30                 51
## 3                     16                    25                 48
## 4                     13                    27                 53
## 5                     12                    24                 47
## 6                     13                    36                 61
##   aromaticC_bbC_vlong aromaticC_bbO_vshort aromaticC_bbO_short
## 1                  80                    0                   4
## 2                  78                    0                   3
## 3                  91                    0                   3
## 4                  62                    0                   3
## 5                  80                    0                   2
## 6                  83                    0                   3
##   aromaticC_bbO_medshort aromaticC_bbO_medlong aromaticC_bbO_long
## 1                     21                    36                 47
## 2                     17                    28                 48
## 3                     18                    35                 48
## 4                     16                    28                 41
## 5                     13                    24                 48
```

```
## 6                     15                    36                 61
##   aromaticC_bbO_vlong scAGN_scAGN_short scAGN_scAGN_medshort
## 1                  71                 0                    1
## 2                  67                 0                    1
## 3                  76                 0                    1
## 4                  65                 1                    0
## 5                  79                 0                    1
## 6                  75                 0                    1
##   scAGN_scAGN_medlong scAGN_scAGN_long scAGN_scAGN_vlong scAGN_scLysN_medshort
## 1                   0                0                 1                     0
## 2                   0                0                 0                     0
## 3                   0                1                 0                     0
## 4                   0                0                 0                     1
## 5                   0                0                 0                     0
## 6                   0                0                 0                     0
##   scAGN_scLysN_medlong scAGN_scLysN_long scAGN_scLysN_vlong
## 1                    0                 1                  0
## 2                    0                 0                  1
## 3                    0                 1                  0
## 4                    0                 0                  0
## 5                    0                 0                  1
## 6                    0                 1                  0
##   scAGN_bbProN_medshort scAGN_bbProN_medlong scAGN_bbProN_long
## 1                     0                    0                 1
## 2                     0                    1                 0
## 3                     0                    0                 1
## 4                     0                    0                 1
## 5                     0                    0                 1
## 6                     1                    0                 0
##   scAGN_bbProN_vlong scAGN_hydroxylO_vshort scAGN_hydroxylO_short
## 1                  1                      1                    1
## 2                  0                      0                    1
## 3                  1                      2                    0
## 4                  0                      0                    1
## 5                  0                      1                    0
## 6                  0                      0                    1
##   scAGN_hydroxylO_medshort scAGN_hydroxylO_medlong scAGN_hydroxylO_long
## 1                        0                       0                   2
## 2                        0                       4                   1
## 3                        0                       0                   2
## 4                        0                       2                   1
## 5                        2                       0                   0
## 6                        1                       2                   1
##   scAGN_hydroxylO_vlong scAGN_carbonylO_vshort scAGN_carbonylO_short
## 1                     1                      0                    0
## 2                     3                      0                    1
## 3                     3                      0                    0
## 4                     1                      0                    1
## 5                     2                      0                    1
## 6                     2                      0                    1
##   scAGN_carbonylO_medshort scAGN_carbonylO_medlong scAGN_carbonylO_long
## 1                        0                       2                   0
## 2                        0                       1                   0
## 3                        0                       2                   1
```

```
## 4                       0                     1                   0
## 5                       0                     1                   0
## 6                       0                     0                   1
##   scAGN_carbonylO_vlong scAGN_carboxylO_vshort scAGN_carboxylO_short
## 1                    0                      0                     1
## 2                    0                      0                     0
## 3                    0                      1                     0
## 4                    0                      0                     0
## 5                    0                      1                     0
## 6                    0                      0                     0
##   scAGN_carboxylO_medshort scAGN_carboxylO_medlong scAGN_carboxylO_long
## 1                        0                      3                    0
## 2                        0                      0                    0
## 3                        1                      0                    1
## 4                        0                      0                    0
## 5                        1                      0                    0
## 6                        0                      0                    0
##   scAGN_carboxylO_vlong scAGN_sulfur_short scAGN_sulfur_medshort
## 1                     0                  0                    1
## 2                     0                  0                    0
## 3                     1                  0                    1
## 4                     0                  0                    1
## 5                     0                  1                    0
## 6                     0                  0                    0
##   scAGN_sulfur_medlong scAGN_sulfur_long scAGN_sulfur_vlong scAGN_bbN_short
## 1                    1                 1                  0              0
## 2                    3                 0                  0              1
## 3                    0                 0                  1              0
## 4                    1                 2                  0              0
## 5                    0                 1                  1              0
## 6                    1                 1                  0              1
##   scAGN_bbN_medshort scAGN_bbN_medlong scAGN_bbN_long scAGN_bbN_vlong
## 1                  3                 4              9               6
## 2                  2                 5              6               4
## 3                  3                 2             10               9
## 4                  0                 2              1              12
## 5                  2                 2             12              11
## 6                  0                 2              6               7
##   scAGN_bbCA_short scAGN_bbCA_medshort scAGN_bbCA_medlong scAGN_bbCA_long
## 1                0                   4                  6               6
## 2                1                   1                  6               7
## 3                0                   4                  3              10
## 4                0                   0                  3               3
## 5                0                   2                  6               5
## 6                1                   3                  1               6
##   scAGN_bbCA_vlong scAGN_bbC_short scAGN_bbC_medshort scAGN_bbC_medlong
## 1                9               1                  2                5
## 2                8               0                  3                4
## 3               10               1                  1                6
## 4               12               0                  0                0
## 5               13               0                  0                5
## 6                8               2                  1                3
##   scAGN_bbC_long scAGN_bbC_vlong scAGN_bbO_vshort scAGN_bbO_short
## 1              6              13                1               2
```

```
## 2              8              5               0              2
## 3              9             13               1              1
## 4              6              9               0              0
## 5             10             14               0              0
## 6              6             11               0              1
##   scAGN_bbO_medshort scAGN_bbO_medlong scAGN_bbO_long scAGN_bbO_vlong
## 1                  1                 1              9               17
## 2                  1                 4              5               10
## 3                  1                 2              9               12
## 4                  0                 1              8                7
## 5                  1                 8              9               10
## 6                  2                 3              4               10
##   scLysN_hydroxylO_long scLysN_hydroxylO_vlong scLysN_carbonylO_vshort
## 1                     0                      0                       0
## 2                     0                      0                       0
## 3                     0                      0                       0
## 4                     0                      2                       1
## 5                     0                      0                       0
## 6                     0                      0                       0
##   scLysN_carbonylO_long scLysN_carbonylO_vlong scLysN_carboxylO_medlong
## 1                     1                      0                       1
## 2                     0                      1                       0
## 3                     1                      0                       0
## 4                     0                      0                       0
## 5                     0                      1                       0
## 6                     0                      0                       0
##   scLysN_carboxylO_long scLysN_carboxylO_vlong scLysN_bbN_medshort
## 1                     1                      0                  0
## 2                     0                      0                  0
## 3                     1                      0                  0
## 4                     0                      0                  2
## 5                     1                      1                  0
## 6                     0                      0                  0
##   scLysN_bbN_medlong scLysN_bbN_long scLysN_bbN_vlong scLysN_bbCA_medshort
## 1                  1               0                3                    0
## 2                  1               0                2                    0
## 3                  1               0                2                    0
## 4                  1               0                2                    1
## 5                  1               0                2                    0
## 6                  0               1                2                    0
##   scLysN_bbCA_medlong scLysN_bbCA_long scLysN_bbCA_vlong scLysN_bbC_medlong
## 1                   1                1                 1                  0
## 2                   1                1                 0                  0
## 3                   1                1                 0                  0
## 4                   1                2                 1                  1
## 5                   0                2                 0                  0
## 6                   0                2                 0                  0
##   scLysN_bbC_long scLysN_bbC_vlong scLysN_bbO_medlong scLysN_bbO_long
## 1               3                1                 2               0
## 2               2                1                 0               2
## 3               3                1                 2               0
## 4               1                2                 0               3
## 5               2                1                 0               2
## 6               2                1                 1               1
```

```
##   scLysN_bbO_vlong scArgN_hydroxylO_short scArgN_hydroxylO_medshort
## 1               2                      2                         1
## 2               3                      2                         1
## 3               2                      2                         1
## 4               1                      2                         1
## 5               2                      2                         1
## 6               1                      2                         1
##   scArgN_hydroxylO_medlong scArgN_hydroxylO_long scArgN_hydroxylO_vlong
## 1                        0                     2                      3
## 2                        0                     2                      3
## 3                        0                     2                      3
## 4                        0                     3                      1
## 5                        0                     2                      2
## 6                        0                     2                      3
##   scArgN_carboxylO_long scArgN_carboxylO_vlong scArgN_bbN_medshort
## 1                     0                      0                  2
## 2                     0                      4                  2
## 3                     0                      2                  2
## 4                     2                      0                  1
## 5                     0                      1                  2
## 6                     0                      4                  2
##   scArgN_bbN_medlong scArgN_bbN_long scArgN_bbN_vlong scArgN_bbCA_medshort
## 1                  3               2                5                    1
## 2                  2               3                3                    1
## 3                  3               2                5                    1
## 4                  3               2                5                    1
## 5                  3               2                5                    1
## 6                  2               3                3                    1
##   scArgN_bbCA_medlong scArgN_bbCA_long scArgN_bbCA_vlong scArgN_bbC_short
## 1                   3                1                 2                1
## 2                   3                0                 3                1
## 3                   3                1                 2                1
## 4                   3                0                 4                1
## 5                   3                1                 2                1
## 6                   3                0                 3                1
##   scArgN_bbC_medshort scArgN_bbC_medlong scArgN_bbC_long scArgN_bbC_vlong
## 1                   1                 2                2                3
## 2                   1                 2                2                0
## 3                   1                 2                2                3
## 4                   1                 1                3                2
## 5                   1                 2                2                2
## 6                   1                 2                2                0
##   scArgN_bbO_vshort scArgN_bbO_short scArgN_bbO_medshort scArgN_bbO_medlong
## 1                 1                1                   1                  1
## 2                 1                1                   1                  1
## 3                 1                1                   1                  1
## 4                 1                1                   1                  1
## 5                 2                0                   1                  1
## 6                 1                1                   1                  1
##   scArgN_bbO_long scArgN_bbO_vlong bbProN_hydroxylO_medshort
## 1               1                3                         0
## 2               1                2                         0
## 3               1                3                         0
## 4               1                3                         0
```

```
## 5               1               4                       0
## 6               1               2                       0
##   bbProN_hydroxylO_medlong bbProN_hydroxylO_long bbProN_hydroxylO_vlong
## 1                        1                    1                      0
## 2                        0                    0                      2
## 3                        0                    2                      0
## 4                        0                    1                      1
## 5                        1                    1                      2
## 6                        1                    0                      2
##   bbProN_carbonylO_medlong bbProN_carbonylO_long bbProN_carbonylO_vlong
## 1                        1                    0                      0
## 2                        0                    0                      2
## 3                        1                    0                      1
## 4                        0                    0                      0
## 5                        1                    0                      0
## 6                        2                    0                      0
##   bbProN_carboxylO_vlong bbProN_sulfur_vlong bbProN_bbN_medshort
## 1                      0                   0                  2
## 2                      0                   0                  1
## 3                      0                   0                  2
## 4                      0                   0                  1
## 5                      1                   0                  1
## 6                      0                   2                  1
##   bbProN_bbN_medlong bbProN_bbN_long bbProN_bbN_vlong bbProN_bbCA_medshort
## 1                  4               2                3                    2
## 2                  2               4                3                    0
## 3                  5               1                4                    2
## 4                  3               4                5                    0
## 5                  3               4                4                    0
## 6                  2               3                7                    0
##   bbProN_bbCA_medlong bbProN_bbCA_long bbProN_bbCA_vlong bbProN_bbC_short
## 1                   4                3                 2                1
## 2                   4                2                 5                0
## 3                   3                4                 2                1
## 4                   4                5                 3                0
## 5                   6                3                 1                0
## 6                   4                2                 7                0
##   bbProN_bbC_medshort bbProN_bbC_medlong bbProN_bbC_long bbProN_bbC_vlong
## 1                   1                  5               2                3
## 2                   1                  3               3                5
## 3                   1                  3               4                3
## 4                   1                  3               6                3
## 5                   1                  5               1                5
## 6                   1                  3               4                2
##   bbProN_bbO_short bbProN_bbO_medshort bbProN_bbO_medlong bbProN_bbO_long
## 1                1                   4                  1               2
## 2                1                   1                  1               4
## 3                0                   3                  2               2
## 4                1                   1                  2               5
## 5                0                   4                  0               4
## 6                0                   2                  2               3
##   bbProN_bbO_vlong hydroxylO_hydroxylO_short hydroxylO_hydroxylO_medshort
## 1                4                         1                           0
## 2                3                         1                           1
```

```
## 3                5                         0                           2
## 4                5                         0                           1
## 5                3                         0                           1
## 6                5                         2                           0
##   hydroxyl0_hydroxyl0_medlong hydroxyl0_hydroxyl0_long
## 1                           0                        1
## 2                           1                        0
## 3                           0                        1
## 4                           2                        2
## 5                           0                        2
## 6                           0                        1
##   hydroxyl0_hydroxyl0_vlong hydroxyl0_carbonyl0_vshort
## 1                         2                          0
## 2                         3                          0
## 3                         3                          0
## 4                         1                          1
## 5                         4                          0
## 6                         5                          2
##   hydroxyl0_carbonyl0_short hydroxyl0_carbonyl0_medshort
## 1                         0                           3
## 2                         2                           1
## 3                         0                           3
## 4                         0                           1
## 5                         1                           1
## 6                         0                           1
##   hydroxyl0_carbonyl0_medlong hydroxyl0_carbonyl0_long
## 1                           2                        1
## 2                           1                        2
## 3                           4                        0
## 4                           1                        1
## 5                           0                        1
## 6                           0                        3
##   hydroxyl0_carbonyl0_vlong hydroxyl0_carboxyl0_vshort
## 1                         0                          1
## 2                         2                          1
## 3                         0                          1
## 4                         4                          1
## 5                         5                          1
## 6                         2                          1
##   hydroxyl0_carboxyl0_short hydroxyl0_carboxyl0_medshort
## 1                         1                           0
## 2                         1                           2
## 3                         1                           1
## 4                         1                           1
## 5                         1                           1
## 6                         1                           2
##   hydroxyl0_carboxyl0_medlong hydroxyl0_carboxyl0_long
## 1                           0                        0
## 2                           0                        1
## 3                           1                        0
## 4                           0                        1
## 5                           0                        1
## 6                           0                        0
##   hydroxyl0_carboxyl0_vlong hydroxyl0_sulfur_short hydroxyl0_sulfur_medshort
```

```
## 1                         2                    0                         0
## 2                         0                    0                         0
## 3                         3                    0                         0
## 4                         1                    0                         0
## 5                         1                    0                         1
## 6                         1                    0                         0
##   hydroxylO_sulfur_medlong hydroxylO_sulfur_long hydroxylO_sulfur_vlong
## 1                        0                     0                      3
## 2                        0                     0                      3
## 3                        0                     1                      0
## 4                        1                     1                      1
## 5                        0                     0                      0
## 6                        0                     0                      0
##   hydroxylO_bbN_short hydroxylO_bbN_medshort hydroxylO_bbN_medlong
## 1                   0                      3                     4
## 2                   3                      5                     2
## 3                   3                      4                     4
## 4                   1                      3                     2
## 5                   0                      8                     2
## 6                   1                      7                     3
##   hydroxylO_bbN_long hydroxylO_bbN_vlong hydroxylO_bbCA_short
## 1                 13                   5                    1
## 2                  8                  14                    3
## 3                 11                   9                    1
## 4                  6                  16                    2
## 5                 10                  16                    2
## 6                 10                  16                    2
##   hydroxylO_bbCA_medshort hydroxylO_bbCA_medlong hydroxylO_bbCA_long
## 1                       3                      6                   4
## 2                       4                      5                   8
## 3                       6                      6                   9
## 4                       1                      5                   7
## 5                       4                      8                   6
## 6                       5                      6                  11
##   hydroxylO_bbCA_vlong hydroxylO_bbC_short hydroxylO_bbC_medshort
## 1                   18                   0                      5
## 2                   13                   2                      4
## 3                   14                   1                      8
## 4                   18                   1                      6
## 5                   22                   2                      2
## 6                   15                   3                      4
##   hydroxylO_bbC_medlong hydroxylO_bbC_long hydroxylO_bbC_vlong
## 1                     5                  8                  12
## 2                     7                  6                  11
## 3                     4                  9                  11
## 4                     5                  5                  12
## 5                     9                 11                  11
## 6                     5                  9                  19
##   hydroxylO_bbO_vshort hydroxylO_bbO_short hydroxylO_bbO_medshort
## 1                    0                   3                     1
## 2                    1                   1                     1
## 3                    0                   3                     2
## 4                    1                   1                     3
## 5                    2                   1                     1
```

```
## 6                    1                    1                        4
##   hydroxylO_bbO_medlong hydroxylO_bbO_long hydroxylO_bbO_vlong
## 1                     5                   8                  16
## 2                     9                   9                   9
## 3                     7                   9                  16
## 4                     9                   6                  10
## 5                     9                  11                  14
## 6                     6                  10                  21
##   carbonylO_carbonylO_medshort carbonylO_carbonylO_medlong
## 1                            0                           1
## 2                            1                           0
## 3                            0                           1
## 4                            0                           1
## 5                            0                           1
## 6                            0                           1
##   carbonylO_carbonylO_long carbonylO_carbonylO_vlong
## 1                        0                         0
## 2                        0                         0
## 3                        0                         1
## 4                        0                         0
## 5                        0                         0
## 6                        0                         0
##   carbonylO_carboxylO_medshort carbonylO_carboxylO_medlong
## 1                            2                           1
## 2                            0                           0
## 3                            1                           0
## 4                            0                           0
## 5                            0                           1
## 6                            0                           0
##   carbonylO_carboxylO_long carbonylO_carboxylO_vlong carbonylO_sulfur_short
## 1                        1                         0                      0
## 2                        0                         0                      0
## 3                        1                         1                      0
## 4                        0                         0                      0
## 5                        0                         0                      1
## 6                        0                         0                      0
##   carbonylO_sulfur_medshort carbonylO_sulfur_medlong carbonylO_sulfur_long
## 1                         0                        2                    0
## 2                         1                        1                    0
## 3                         0                        1                    0
## 4                         1                        0                    2
## 5                         0                        0                    0
## 6                         0                        0                    1
##   carbonylO_sulfur_vlong carbonylO_bbN_short carbonylO_bbN_medshort
## 1                      1                   0                     2
## 2                      2                   0                     1
## 3                      0                   0                     2
## 4                      0                   0                     1
## 5                      3                   0                     3
## 6                      1                   1                     2
##   carbonylO_bbN_medlong carbonylO_bbN_long carbonylO_bbN_vlong
## 1                     5                  6                   9
## 2                     6                  4                   8
## 3                     6                  2                  15
```

```
## 4                    4                  2                   5
## 5                    6                  7                   3
## 6                    1                  7                   5
##   carbonylO_bbCA_short carbonylO_bbCA_medshort carbonylO_bbCA_medlong
## 1                    0                      4                     3
## 2                    1                      1                     3
## 3                    1                      2                     5
## 4                    0                      1                     3
## 5                    0                      3                     7
## 6                    1                      3                     2
##   carbonylO_bbCA_long carbonylO_bbCA_vlong carbonylO_bbC_short
## 1                   5                  12                   0
## 2                   8                   7                   0
## 3                   4                  11                   0
## 4                   3                   6                   0
## 5                   4                  12                   0
## 6                   7                   5                   1
##   carbonylO_bbC_medshort carbonylO_bbC_medlong carbonylO_bbC_long
## 1                      2                    6                  6
## 2                      3                    1                  6
## 3                      2                    6                  4
## 4                      1                    0                  4
## 5                      3                    7                  4
## 6                      0                    6                  2
##   carbonylO_bbC_vlong carbonylO_bbO_short carbonylO_bbO_medshort
## 1                  10                   0                     3
## 2                  11                   1                     1
## 3                  11                   0                     3
## 4                  11                   0                     1
## 5                   9                   2                     2
## 6                  12                   1                     2
##   carbonylO_bbO_medlong carbonylO_bbO_long carbonylO_bbO_vlong
## 1                     5                  6                  9
## 2                     3                  6                 11
## 3                     3                  5                 12
## 4                     0                  5                 12
## 5                     2                 10                 11
## 6                     2                  6                  9
##   carboxylO_carboxylO_medlong carboxylO_carboxylO_long
## 1                           0                       0
## 2                           0                       2
## 3                           0                       1
## 4                           0                       1
## 5                           0                       2
## 6                           0                       2
##   carboxylO_carboxylO_vlong carboxylO_sulfur_medshort carboxylO_sulfur_medlong
## 1                         0                         1                       0
## 2                         2                         0                       0
## 3                         3                         1                       0
## 4                         2                         0                       0
## 5                         0                         0                       1
## 6                         2                         0                       0
##   carboxylO_sulfur_long carboxylO_sulfur_vlong carboxylO_bbN_vshort
## 1                     2                       1                   1
```

```
## 2                     0                     0                     1
## 3                     0                     0                     1
## 4                     0                     0                     1
## 5                     0                     0                     1
## 6                     0                     1                     1
##   carboxylO_bbN_medshort carboxylO_bbN_medlong carboxylO_bbN_long
## 1                      3                     1                  2
## 2                      1                     3                  0
## 3                      1                     4                  5
## 4                      2                     1                  1
## 5                      2                     2                  2
## 6                      1                     3                  0
##   carboxylO_bbN_vlong carboxylO_bbCA_short carboxylO_bbCA_medshort
## 1                   8                    1                       0
## 2                   4                    2                       1
## 3                   9                    1                       2
## 4                   6                    2                       0
## 5                   6                    1                       0
## 6                   4                    2                       1
##   carboxylO_bbCA_medlong carboxylO_bbCA_long carboxylO_bbCA_vlong
## 1                      3                   2                   8
## 2                      1                   1                   3
## 3                      4                   4                   4
## 4                      2                   1                   1
## 5                      4                   5                   3
## 6                      1                   1                   2
##   carboxylO_bbC_short carboxylO_bbC_medshort carboxylO_bbC_medlong
## 1                   0                     1                     1
## 2                   1                     2                     1
## 3                   0                     2                     3
## 4                   1                     1                     2
## 5                   0                     1                     4
## 6                   1                     2                     1
##   carboxylO_bbC_long carboxylO_bbC_vlong carboxylO_bbO_medshort
## 1                  4                   4                     1
## 2                  2                   2                     2
## 3                  6                   6                     2
## 4                  2                   2                     2
## 5                  3                   5                     1
## 6                  2                   1                     2
##   carboxylO_bbO_medlong carboxylO_bbO_long carboxylO_bbO_vlong
## 1                     3                  2                  6
## 2                     3                  2                  2
## 3                     3                  4                  9
## 4                     2                  3                  2
## 5                     2                  6                  3
## 6                     3                  2                  0
##   sulfur_sulfur_vlong sulfur_bbN_short sulfur_bbN_medshort sulfur_bbN_medlong
## 1                   0                0                   0                  5
## 2                   0                0                   3                  6
## 3                   0                0                   0                  4
## 4                   1                0                   0                  1
## 5                   0                0                   0                  6
## 6                   0                0                   1                  4
```

```
##   sulfur_bbN_long sulfur_bbN_vlong sulfur_bbCA_short sulfur_bbCA_medshort
## 1               3                8                 0                    1
## 2               3                6                 0                    4
## 3               3                7                 0                    1
## 4               8                9                 0                    1
## 5               6                6                 0                    2
## 6               4                7                 0                    2
##   sulfur_bbCA_medlong sulfur_bbCA_long sulfur_bbCA_vlong sulfur_bbC_short
## 1                   5                4                 3                0
## 2                   4                5                 6                1
## 3                   3                4                 5                0
## 4                   3                5                 9                0
## 5                   1               11                 5                0
## 6                   2                7                 6                0
##   sulfur_bbC_medshort sulfur_bbC_medlong sulfur_bbC_long sulfur_bbC_vlong
## 1                   1                  4               6                4
## 2                   4                  3               7                2
## 3                   1                  4               4                5
## 4                   2                  2               6                9
## 5                   4                  2               5                7
## 6                   3                  1               5                9
##   sulfur_bbO_short sulfur_bbO_medshort sulfur_bbO_medlong sulfur_bbO_long
## 1                0                   2                  3               6
## 2                2                   1                  2               5
## 3                0                   3                  1               5
## 4                1                   3                  2               5
## 5                3                   1                  2               4
## 6                1                   2                  3               5
##   sulfur_bbO_vlong bbN_bbN_medshort bbN_bbN_medlong bbN_bbN_long bbN_bbN_vlong
## 1                5                6               11           22            18
## 2                9                6               10           17            22
## 3                5                9               14           20            23
## 4                7                7                9           16            19
## 5                4                5               12           20            16
## 6                5                6               13           13            24
##   bbN_bbCA_medshort bbN_bbCA_medlong bbN_bbCA_long bbN_bbCA_vlong
## 1                16               28            32             56
## 2                14               31            22             59
## 3                18               34            35             60
## 4                13               27            25             57
## 5                11               33            31             51
## 6                14               31            27             58
##   bbN_bbC_vshort bbN_bbC_short bbN_bbC_medshort bbN_bbC_medlong bbN_bbC_long
## 1              0            17               10               16           37
## 2              0            18                9               11           42
## 3              0            19               10               19           44
## 4              1            17                9               11           33
## 5              0            13               14               15           38
## 6              0            15               12               13           36
##   bbN_bbC_vlong bbN_bbO_vshort bbN_bbO_short bbN_bbO_medshort bbN_bbO_medlong
## 1            48              3            19               11              10
## 2            37              2            19                7              14
## 3            51              3            21                8              22
## 4            50              0            21                8              10
```

```
## 5            45              1            18              11              17
## 6            44              4            16              10              15
##   bbN_bbO_long bbN_bbO_vlong bbCA_bbCA_medshort bbCA_bbCA_medlong
## 1           35            42                  1                24
## 2           39            40                  4                15
## 3           38            42                  3                24
## 4           41            35                  2                16
## 5           36            36                  1                21
## 6           34            44                  2                18
##   bbCA_bbCA_long bbCA_bbCA_vlong bbCA_bbC_medshort bbCA_bbC_medlong
## 1             22              21                20               26
## 2             23              23                21               22
## 3             23              28                21               30
## 4             27              17                17               23
## 5             25              17                20               25
## 6             25              23                20               26
##   bbCA_bbC_long bbCA_bbC_vlong bbCA_bbO_vshort bbCA_bbO_short bbCA_bbO_medshort
## 1            33             66               0              6                21
## 2            30             56               0              5                23
## 3            37             69               1              7                19
## 4            31             62               0              3                22
## 5            38             57               0              7                17
## 6            31             62               0              9                16
##   bbCA_bbO_medlong bbCA_bbO_long bbCA_bbO_vlong bbC_bbC_medshort
## 1               18            36              64                5
## 2               13            31              62                5
## 3               23            42              58                7
## 4               17            31              54                5
## 5               22            41              54                6
## 6               16            39              58                8
##   bbC_bbC_medlong bbC_bbC_long bbC_bbC_vlong bbC_bbO_short bbC_bbO_medshort
## 1              20           19            18             3               13
## 2              19           15            24             3               14
## 3              22           16            29             4               17
## 4              17           16            25             1               12
## 5              17           22            17             6               12
## 6              13           19            25             7                8
##   bbC_bbO_medlong bbC_bbO_long bbC_bbO_vlong bbO_bbO_vshort bbO_bbO_short
## 1              30           36            50              0             1
## 2              20           40            43              0             2
## 3              31           39            54              0             2
## 4              32           32            44              0             1
## 5              26           35            50              0             2
## 6              26           43            46              0             3
##   bbO_bbO_medshort bbO_bbO_medlong bbO_bbO_long bbO_bbO_vlong
## 1                5              15           24            25
## 2                5              10           21            27
## 3                7              15           21            26
## 4                4              15           22            17
## 5               11               8           17            34
## 6                5              16           17            25
```

```
summary(protein_train)
```

```
##     accuracy          angles      carbonylC_carbonylC_medshort
```

```
##  Min.   : 2.371   Min.   :113.8   Min.   :0.000
##  1st Qu.: 6.372   1st Qu.:122.0   1st Qu.:1.000
##  Median : 7.704   Median :124.9   Median :1.000
##  Mean   : 7.488   Mean   :124.8   Mean   :1.143
##  3rd Qu.: 8.683   3rd Qu.:127.7   3rd Qu.:1.000
##  Max.   :10.482   Max.   :139.6   Max.   :4.000
##  carbonylC_carbonylC_medlong carbonylC_carbonylC_long carbonylC_carbonylC_vlong
##  Min.   :0.0000              Min.   :0.0000           Min.   :0.0000
##  1st Qu.:0.0000              1st Qu.:0.0000           1st Qu.:0.0000
##  Median :0.0000              Median :0.0000           Median :0.0000
##  Mean   :0.2354              Mean   :0.2076           Mean   :0.2729
##  3rd Qu.:0.0000              3rd Qu.:0.0000           3rd Qu.:0.0000
##  Max.   :3.0000              Max.   :2.0000           Max.   :3.0000
##  carbonylC_carboxylC_medshort carbonylC_carboxylC_long
##  Min.   :0.0000               Min.   :0.0000
##  1st Qu.:0.0000               1st Qu.:0.0000
##  Median :0.0000               Median :0.0000
##  Mean   :0.2256               Mean   :0.1131
##  3rd Qu.:0.0000               3rd Qu.:0.0000
##  Max.   :3.0000               Max.   :3.0000
##  carbonylC_carboxylC_vlong carbonylC_aliph1HC_medshort
##  Min.   :0.0000            Min.   :0.0000
##  1st Qu.:0.0000            1st Qu.:0.0000
##  Median :0.0000            Median :0.0000
##  Mean   :0.2472            Mean   :0.4275
##  3rd Qu.:0.0000            3rd Qu.:1.0000
##  Max.   :4.0000            Max.   :4.0000
##  carbonylC_aliph1HC_medlong carbonylC_aliph1HC_long carbonylC_aliph1HC_vlong
##  Min.   :0.000              Min.   :0.000           Min.   :0.000
##  1st Qu.:1.000              1st Qu.:1.000           1st Qu.:1.000
##  Median :1.000              Median :2.000           Median :1.000
##  Mean   :1.572              Mean   :2.012           Mean   :1.467
##  3rd Qu.:2.000              3rd Qu.:3.000           3rd Qu.:2.000
##  Max.   :6.000              Max.   :7.000           Max.   :6.000
##  carbonylC_aliph2HC_short carbonylC_aliph2HC_medshort
##  Min.   :0.000            Min.   : 0.000
##  1st Qu.:1.000            1st Qu.: 2.000
##  Median :1.000            Median : 3.000
##  Mean   :1.264            Mean   : 3.581
##  3rd Qu.:2.000            3rd Qu.: 4.000
##  Max.   :6.000            Max.   :10.000
##  carbonylC_aliph2HC_medlong carbonylC_aliph2HC_long carbonylC_aliph2HC_vlong
##  Min.   : 1.000             Min.   : 0.000          Min.   : 3.000
##  1st Qu.: 3.000             1st Qu.: 3.000          1st Qu.: 7.000
##  Median : 4.000             Median : 4.000          Median : 9.000
##  Mean   : 4.116             Mean   : 4.591          Mean   : 9.428
##  3rd Qu.: 5.000             3rd Qu.: 6.000          3rd Qu.:11.000
##  Max.   :13.000             Max.   :13.000          Max.   :20.000
##  carbonylC_aliph3HC_short carbonylC_aliph3HC_medshort
##  Min.   :0.0000           Min.   :0.000
##  1st Qu.:0.0000           1st Qu.:1.000
##  Median :0.0000           Median :2.000
##  Mean   :0.5149           Mean   :1.837
##  3rd Qu.:1.0000           3rd Qu.:3.000
```

```
##  Max.   :4.0000           Max.   :7.000
##  carbonylC_aliph3HC_medlong carbonylC_aliph3HC_long carbonylC_aliph3HC_vlong
##  Min.   :0.000              Min.   :0.000           Min.   :0.000
##  1st Qu.:1.000              1st Qu.:1.000           1st Qu.:2.000
##  Median :2.000              Median :2.000           Median :3.000
##  Mean   :2.285              Mean   :2.518           Mean   :3.117
##  3rd Qu.:3.000              3rd Qu.:3.000           3rd Qu.:4.000
##  Max.   :8.000              Max.   :8.000           Max.   :9.000
##  carbonylC_aromaticC_short carbonylC_aromaticC_medshort
##  Min.   :0.0000            Min.   : 0.000
##  1st Qu.:0.0000            1st Qu.: 2.000
##  Median :0.0000            Median : 3.000
##  Mean   :0.3325            Mean   : 3.506
##  3rd Qu.:1.0000            3rd Qu.: 4.000
##  Max.   :6.0000            Max.   :12.000
##  carbonylC_aromaticC_medlong carbonylC_aromaticC_long carbonylC_aromaticC_vlong
##  Min.   : 0.000              Min.   : 2.00            Min.   : 2.000
##  1st Qu.: 4.000              1st Qu.: 6.00            1st Qu.: 6.000
##  Median : 6.000              Median : 7.00            Median : 8.000
##  Mean   : 5.756              Mean   : 7.56            Mean   : 7.805
##  3rd Qu.: 7.000              3rd Qu.: 9.00            3rd Qu.: 9.000
##  Max.   :15.000              Max.   :17.00            Max.   :19.000
##  carbonylC_scAGN_short carbonylC_scAGN_medshort carbonylC_scAGN_medlong
##  Min.   :0.0000        Min.   :0.000            Min.   :0.0000
##  1st Qu.:0.0000        1st Qu.:1.000            1st Qu.:0.0000
##  Median :1.0000        Median :1.000            Median :0.0000
##  Mean   :0.7282        Mean   :1.244            Mean   :0.7996
##  3rd Qu.:1.0000        3rd Qu.:2.000            3rd Qu.:1.0000
##  Max.   :4.0000        Max.   :5.000            Max.   :7.0000
##  carbonylC_scAGN_long carbonylC_scAGN_vlong carbonylC_scLysN_short
##  Min.   :0.0000       Min.   :0.0000        Min.   :0.0000
##  1st Qu.:0.0000       1st Qu.:0.0000        1st Qu.:0.0000
##  Median :0.0000       Median :0.0000        Median :0.0000
##  Mean   :0.4404       Mean   :0.5663        Mean   :0.1151
##  3rd Qu.:1.0000       3rd Qu.:1.0000        3rd Qu.:0.0000
##  Max.   :5.0000       Max.   :5.0000        Max.   :1.0000
##  carbonylC_scLysN_medlong carbonylC_scLysN_long carbonylC_scLysN_vlong
##  Min.   :0.0000           Min.   :0.0000        Min.   :0.000
##  1st Qu.:0.0000           1st Qu.:0.0000        1st Qu.:0.000
##  Median :0.0000           Median :0.0000        Median :0.000
##  Mean   :0.1603           Mean   :0.2852        Mean   :0.242
##  3rd Qu.:0.0000           3rd Qu.:1.0000        3rd Qu.:0.000
##  Max.   :1.0000           Max.   :1.0000        Max.   :1.000
##  carbonylC_bbProN_medlong carbonylC_bbProN_long carbonylC_bbProN_vlong
##  Min.   :0.0000           Min.   :0.0000        Min.   :0.0000
##  1st Qu.:0.0000           1st Qu.:0.0000        1st Qu.:0.0000
##  Median :0.0000           Median :0.0000        Median :0.0000
##  Mean   :0.2641           Mean   :0.3792        Mean   :0.4563
##  3rd Qu.:0.0000           3rd Qu.:1.0000        3rd Qu.:1.0000
##  Max.   :3.0000           Max.   :3.0000        Max.   :4.0000
##  carbonylC_hydroxylO_short carbonylC_hydroxylO_medshort
##  Min.   :0.00              Min.   :0.00
##  1st Qu.:1.00              1st Qu.:1.00
##  Median :2.00              Median :2.00
```

```
##  Mean   :1.89              Mean   :1.82
##  3rd Qu.:2.00              3rd Qu.:2.00
##  Max.   :5.00              Max.   :5.00
##  carbonylC_hydroxylO_medlong carbonylC_hydroxylO_long carbonylC_hydroxylO_vlong
##  Min.   :0.000               Min.   :0.000           Min.   :0.000
##  1st Qu.:0.000               1st Qu.:2.000           1st Qu.:2.000
##  Median :1.000               Median :3.000           Median :3.000
##  Mean   :1.238               Mean   :2.773           Mean   :2.939
##  3rd Qu.:2.000               3rd Qu.:3.000           3rd Qu.:4.000
##  Max.   :5.000               Max.   :8.000           Max.   :9.000
##  carbonylC_carbonylO_short carbonylC_carbonylO_medshort
##  Min.   :0.00              Min.   :0.0000
##  1st Qu.:0.00              1st Qu.:0.0000
##  Median :1.00              Median :1.0000
##  Mean   :0.63              Mean   :0.6202
##  3rd Qu.:1.00              3rd Qu.:1.0000
##  Max.   :4.00              Max.   :4.0000
##  carbonylC_carbonylO_medlong carbonylC_carbonylO_long carbonylC_carbonylO_vlong
##  Min.   :0.000               Min.   :0.0000          Min.   :0.0000
##  1st Qu.:1.000               1st Qu.:0.0000          1st Qu.:0.0000
##  Median :1.000               Median :0.0000          Median :0.0000
##  Mean   :1.524               Mean   :0.4502          Mean   :0.5247
##  3rd Qu.:2.000               3rd Qu.:1.0000          3rd Qu.:1.0000
##  Max.   :6.000               Max.   :5.0000          Max.   :5.0000
##  carbonylC_carboxylO_short carbonylC_carboxylO_medshort
##  Min.   :0.0000            Min.   :0.0000
##  1st Qu.:0.0000            1st Qu.:0.0000
##  Median :0.0000            Median :0.0000
##  Mean   :0.1788            Mean   :0.1912
##  3rd Qu.:0.0000            3rd Qu.:0.0000
##  Max.   :2.0000            Max.   :3.0000
##  carbonylC_carboxylO_medlong carbonylC_carboxylO_long carbonylC_carboxylO_vlong
##  Min.   :0.0000              Min.   :0.0000          Min.   :0.0000
##  1st Qu.:0.0000              1st Qu.:0.0000          1st Qu.:0.0000
##  Median :0.0000              Median :0.0000          Median :1.0000
##  Mean   :0.3094              Mean   :0.1942          Mean   :0.9435
##  3rd Qu.:1.0000              3rd Qu.:0.0000          3rd Qu.:2.0000
##  Max.   :3.0000              Max.   :4.0000          Max.   :6.0000
##  carbonylC_sulfur_short carbonylC_sulfur_medshort carbonylC_sulfur_medlong
##  Min.   :0.0000         Min.   :0.0000           Min.   :0.0000
##  1st Qu.:0.0000         1st Qu.:0.0000           1st Qu.:0.0000
##  Median :0.0000         Median :0.0000           Median :0.0000
##  Mean   :0.1552         Mean   :0.5103           Mean   :0.4568
##  3rd Qu.:0.0000         3rd Qu.:1.0000           3rd Qu.:1.0000
##  Max.   :3.0000         Max.   :4.0000           Max.   :3.0000
##  carbonylC_sulfur_long carbonylC_sulfur_vlong carbonylC_bbN_short
##  Min.   :0.0000        Min.   :0.0000         Min.   :0.0000
##  1st Qu.:0.0000        1st Qu.:0.0000         1st Qu.:0.0000
##  Median :0.0000        Median :1.0000         Median :0.0000
##  Mean   :0.3834        Mean   :0.6809         Mean   :0.1377
##  3rd Qu.:1.0000        3rd Qu.:1.0000         3rd Qu.:0.0000
##  Max.   :3.0000        Max.   :4.0000         Max.   :3.0000
##  carbonylC_bbN_medshort carbonylC_bbN_medlong carbonylC_bbN_long
##  Min.   :1.00           Min.   : 1.00         Min.   : 2.000
```

```
##  1st Qu.:1.00           1st Qu.: 3.00         1st Qu.: 8.000
##  Median :1.00           Median : 4.00         Median : 9.000
##  Mean   :1.77           Mean   : 4.53         Mean   : 9.474
##  3rd Qu.:2.00           3rd Qu.: 6.00         3rd Qu.:11.000
##  Max.   :7.00           Max.   :13.00         Max.   :19.000
##  carbonylC_bbN_vlong carbonylC_bbCA_medshort carbonylC_bbCA_medlong
##  Min.   : 2.0        Min.   :1.000           Min.   : 0.000
##  1st Qu.: 7.0        1st Qu.:2.000           1st Qu.: 3.000
##  Median : 9.0        Median :3.000           Median : 5.000
##  Mean   : 9.2        Mean   :2.739           Mean   : 4.862
##  3rd Qu.:11.0        3rd Qu.:3.000           3rd Qu.: 6.000
##  Max.   :22.0        Max.   :8.000           Max.   :14.000
##  carbonylC_bbCA_long carbonylC_bbCA_vlong carbonylC_bbC_medshort
##  Min.   : 1.00       Min.   : 3.000       Min.   : 1.000
##  1st Qu.: 6.00       1st Qu.: 8.000       1st Qu.: 2.000
##  Median : 7.00       Median :10.000       Median : 3.000
##  Mean   : 7.39       Mean   : 9.791       Mean   : 2.783
##  3rd Qu.: 9.00       3rd Qu.:11.000       3rd Qu.: 4.000
##  Max.   :16.00       Max.   :20.000       Max.   :10.000
##  carbonylC_bbC_medlong carbonylC_bbC_long carbonylC_bbC_vlong
##  Min.   : 0.000        Min.   : 2.000     Min.   : 4.00
##  1st Qu.: 2.000        1st Qu.: 6.000     1st Qu.:11.00
##  Median : 3.000        Median : 8.000     Median :13.00
##  Mean   : 3.691        Mean   : 7.721     Mean   :12.63
##  3rd Qu.: 5.000        3rd Qu.: 9.000     3rd Qu.:14.00
##  Max.   :11.000        Max.   :17.000     Max.   :23.00
##  carbonylC_bbO_short carbonylC_bbO_medshort carbonylC_bbO_medlong
##  Min.   :0.000       Min.   :0.000          Min.   : 0.000
##  1st Qu.:1.000       1st Qu.:1.000          1st Qu.: 2.000
##  Median :2.000       Median :2.000          Median : 3.000
##  Mean   :2.018       Mean   :2.009          Mean   : 3.006
##  3rd Qu.:2.000       3rd Qu.:3.000          3rd Qu.: 4.000
##  Max.   :6.000       Max.   :8.000          Max.   :11.000
##  carbonylC_bbO_long carbonylC_bbO_vlong carboxylC_carboxylC_vlong
##  Min.   : 1.000     Min.   : 5.00       Min.   :0.0000
##  1st Qu.: 5.000     1st Qu.:13.00       1st Qu.:0.0000
##  Median : 7.000     Median :15.00       Median :0.0000
##  Mean   : 6.792     Mean   :15.18       Mean   :0.4157
##  3rd Qu.: 8.000     3rd Qu.:17.00       3rd Qu.:1.0000
##  Max.   :15.000     Max.   :26.00       Max.   :1.0000
##  carboxylC_aliph2HC_short carboxylC_aliph2HC_medshort
##  Min.   :0.0000           Min.   :1.000
##  1st Qu.:0.0000           1st Qu.:2.000
##  Median :0.0000           Median :2.000
##  Mean   :0.2585           Mean   :2.064
##  3rd Qu.:1.0000           3rd Qu.:2.000
##  Max.   :3.0000           Max.   :6.000
##  carboxylC_aliph2HC_medlong carboxylC_aliph2HC_long carboxylC_aliph2HC_vlong
##  Min.   :0.0000             Min.   :0.0000         Min.   : 0.000
##  1st Qu.:0.0000             1st Qu.:0.0000         1st Qu.: 2.000
##  Median :0.0000             Median :0.0000         Median : 2.000
##  Mean   :0.5457             Mean   :0.8684         Mean   : 2.715
##  3rd Qu.:1.0000             3rd Qu.:1.0000         3rd Qu.: 3.000
##  Max.   :5.0000             Max.   :8.0000         Max.   :12.000
```

```
##  carboxylC_aliph3HC_medlong carboxylC_aliph3HC_vlong
##  Min.   :0.0000             Min.   :0.000
##  1st Qu.:0.0000             1st Qu.:0.000
##  Median :0.0000             Median :0.000
##  Mean   :0.1146             Mean   :0.332
##  3rd Qu.:0.0000             3rd Qu.:1.000
##  Max.   :3.0000             Max.   :3.000
##  carboxylC_aromaticC_medshort carboxylC_aromaticC_medlong
##  Min.   :0.0000               Min.   :0.0000
##  1st Qu.:0.0000               1st Qu.:0.0000
##  Median :0.0000               Median :0.0000
##  Mean   :0.2323               Mean   :0.5673
##  3rd Qu.:0.0000               3rd Qu.:0.0000
##  Max.   :6.0000               Max.   :9.0000
##  carboxylC_aromaticC_long carboxylC_aromaticC_vlong carboxylC_scAGN_short
##  Min.   :0.0000           Min.   :0.0000            Min.   :0.0000
##  1st Qu.:0.0000           1st Qu.:0.0000            1st Qu.:0.0000
##  Median :0.0000           Median :0.0000            Median :0.0000
##  Mean   :0.5899           Mean   :0.6321            Mean   :0.1701
##  3rd Qu.:0.0000           3rd Qu.:1.0000            3rd Qu.:0.0000
##  Max.   :7.0000           Max.   :6.0000            Max.   :2.0000
##  carboxylC_scAGN_long carboxylC_scAGN_vlong carboxylC_scLysN_long
##  Min.   :0.0000       Min.   :0.0000        Min.   :0.0000
##  1st Qu.:0.0000       1st Qu.:0.0000        1st Qu.:0.0000
##  Median :0.0000       Median :0.0000        Median :0.0000
##  Mean   :0.1249       Mean   :0.2734        Mean   :0.3104
##  3rd Qu.:0.0000       3rd Qu.:0.0000        3rd Qu.:1.0000
##  Max.   :3.0000       Max.   :3.0000        Max.   :1.0000
##  carboxylC_scLysN_vlong carboxylC_scArgN_vlong carboxylC_hydroxylO_short
##  Min.   :0.0000         Min.   :0.000          Min.   :0.00
##  1st Qu.:0.0000         1st Qu.:0.000          1st Qu.:1.00
##  Median :0.0000         Median :0.000          Median :1.00
##  Mean   :0.3099         Mean   :0.741          Mean   :1.07
##  3rd Qu.:1.0000         3rd Qu.:2.000          3rd Qu.:1.00
##  Max.   :1.0000         Max.   :4.000          Max.   :3.00
##  carboxylC_hydroxylO_medshort carboxylC_hydroxylO_medlong
##  Min.   :0.0000               Min.   :0.0000
##  1st Qu.:0.0000               1st Qu.:0.0000
##  Median :0.0000               Median :1.0000
##  Mean   :0.3952               Mean   :0.6521
##  3rd Qu.:1.0000               3rd Qu.:1.0000
##  Max.   :3.0000               Max.   :3.0000
##  carboxylC_hydroxylO_vlong carboxylC_carbonylO_medshort
##  Min.   :0.0000            Min.   :0.0000
##  1st Qu.:0.0000            1st Qu.:0.0000
##  Median :0.0000            Median :0.0000
##  Mean   :0.2724            Mean   :0.1095
##  3rd Qu.:0.0000            3rd Qu.:0.0000
##  Max.   :3.0000            Max.   :2.0000
##  carboxylC_carbonylO_medlong carboxylC_carbonylO_vlong
##  Min.   :0.0000              Min.   :0.0000
##  1st Qu.:0.0000              1st Qu.:0.0000
##  Median :0.0000              Median :0.0000
##  Mean   :0.2004              Mean   :0.1876
```

```
##  3rd Qu.:0.0000              3rd Qu.:0.0000
##  Max.   :3.0000              Max.   :3.0000
##  carboxylC_carboxylO_medlong carboxylC_carboxylO_long carboxylC_carboxylO_vlong
##  Min.   :0.0000              Min.   :0.0000           Min.   :0.000
##  1st Qu.:0.0000              1st Qu.:0.0000           1st Qu.:2.000
##  Median :0.0000              Median :0.0000           Median :2.000
##  Mean   :0.1244              Mean   :0.4486           Mean   :2.133
##  3rd Qu.:0.0000              3rd Qu.:1.0000           3rd Qu.:3.000
##  Max.   :4.0000              Max.   :4.0000           Max.   :4.000
##  carboxylC_sulfur_medshort carboxylC_sulfur_vlong carboxylC_bbN_medshort
##  Min.   :0.0000            Min.   :0.00           Min.   :1.000
##  1st Qu.:0.0000            1st Qu.:0.00           1st Qu.:1.000
##  Median :0.0000            Median :0.00           Median :1.000
##  Mean   :0.1269            Mean   :0.24           Mean   :1.079
##  3rd Qu.:0.0000            3rd Qu.:0.00           3rd Qu.:1.000
##  Max.   :1.0000            Max.   :2.00           Max.   :3.000
##  carboxylC_bbN_medlong carboxylC_bbN_long carboxylC_bbN_vlong
##  Min.   :0.0000        Min.   :1.000      Min.   :0
##  1st Qu.:0.0000        1st Qu.:1.000      1st Qu.:2
##  Median :0.0000        Median :2.000      Median :3
##  Mean   :0.5036        Mean   :2.106      Mean   :3
##  3rd Qu.:1.0000        3rd Qu.:3.000      3rd Qu.:4
##  Max.   :5.0000        Max.   :6.000      Max.   :8
##  carboxylC_bbCA_medshort carboxylC_bbCA_medlong carboxylC_bbCA_long
##  Min.   :1.000           Min.   :0.0000         Min.   :0.0000
##  1st Qu.:1.000           1st Qu.:0.0000         1st Qu.:0.0000
##  Median :1.000           Median :1.0000         Median :1.0000
##  Mean   :1.483           Mean   :0.8006         Mean   :0.8623
##  3rd Qu.:2.000           3rd Qu.:1.0000         3rd Qu.:1.0000
##  Max.   :5.000           Max.   :5.0000         Max.   :6.0000
##  carboxylC_bbCA_vlong carboxylC_bbC_medshort carboxylC_bbC_medlong
##  Min.   :0.0          Min.   :0.0000         Min.   :1.000
##  1st Qu.:1.0          1st Qu.:0.0000         1st Qu.:1.000
##  Median :2.0          Median :0.0000         Median :2.000
##  Mean   :2.2          Mean   :0.4265         Mean   :1.732
##  3rd Qu.:3.0          3rd Qu.:1.0000         3rd Qu.:2.000
##  Max.   :8.0          Max.   :4.0000         Max.   :5.000
##  carboxylC_bbC_long carboxylC_bbC_vlong carboxylC_bbO_medshort
##  Min.   :0.000      Min.   :0.000       Min.   :0.0000
##  1st Qu.:1.000      1st Qu.:0.000       1st Qu.:0.0000
##  Median :1.000      Median :1.000       Median :0.0000
##  Mean   :1.364      Mean   :1.344       Mean   :0.2734
##  3rd Qu.:2.000      3rd Qu.:2.000       3rd Qu.:0.0000
##  Max.   :7.000      Max.   :8.000       Max.   :3.0000
##  carboxylC_bbO_medlong carboxylC_bbO_long carboxylC_bbO_vlong
##  Min.   :0.000         Min.   :0.000      Min.   :0.000
##  1st Qu.:1.000         1st Qu.:1.000      1st Qu.:1.000
##  Median :2.000         Median :1.000      Median :2.000
##  Mean   :1.777         Mean   :1.212      Mean   :2.002
##  3rd Qu.:2.000         3rd Qu.:2.000      3rd Qu.:3.000
##  Max.   :6.000         Max.   :6.000      Max.   :9.000
##  aliph1HC_aliph1HC_medlong aliph1HC_aliph1HC_long aliph1HC_aliph1HC_vlong
##  Min.   :0.000             Min.   :0.0000         Min.   :1.000
##  1st Qu.:1.000             1st Qu.:0.0000         1st Qu.:1.000
```

```
##  Median :1.000             Median :0.0000         Median :2.000
##  Mean   :1.002             Mean   :0.2826         Mean   :1.996
##  3rd Qu.:1.000             3rd Qu.:1.0000         3rd Qu.:2.000
##  Max.   :3.000             Max.   :4.0000         Max.   :6.000
##  aliph1HC_aliph2HC_short aliph1HC_aliph2HC_medshort aliph1HC_aliph2HC_medlong
##  Min.   :0.0000          Min.   :0.000              Min.   : 2.000
##  1st Qu.:0.0000          1st Qu.:1.000              1st Qu.: 5.000
##  Median :0.0000          Median :1.000              Median : 7.000
##  Mean   :0.3011          Mean   :1.632              Mean   : 6.803
##  3rd Qu.:1.0000          3rd Qu.:3.000              3rd Qu.: 8.000
##  Max.   :2.0000          Max.   :7.000              Max.   :16.000
##  aliph1HC_aliph2HC_long aliph1HC_aliph2HC_vlong aliph1HC_aliph3HC_short
##  Min.   : 4.000         Min.   : 4.00           Min.   :0.0000
##  1st Qu.: 8.000         1st Qu.: 8.00           1st Qu.:0.0000
##  Median : 9.000         Median :10.00           Median :0.0000
##  Mean   : 9.657         Mean   :10.44           Mean   :0.1922
##  3rd Qu.:11.000         3rd Qu.:12.00           3rd Qu.:0.0000
##  Max.   :20.000         Max.   :23.00           Max.   :3.0000
##  aliph1HC_aliph3HC_medshort aliph1HC_aliph3HC_medlong aliph1HC_aliph3HC_long
##  Min.   :0.000              Min.   :0.000             Min.   : 0.000
##  1st Qu.:1.000              1st Qu.:1.000             1st Qu.: 2.000
##  Median :2.000              Median :2.000             Median : 3.000
##  Mean   :1.673              Mean   :2.077             Mean   : 3.199
##  3rd Qu.:2.000              3rd Qu.:3.000             3rd Qu.: 4.000
##  Max.   :7.000              Max.   :7.000             Max.   :10.000
##  aliph1HC_aliph3HC_vlong aliph1HC_aromaticC_short aliph1HC_aromaticC_medshort
##  Min.   : 0.000          Min.   :0.0000           Min.   : 1.00
##  1st Qu.: 2.000          1st Qu.:0.0000           1st Qu.: 3.00
##  Median : 3.000          Median :0.0000           Median : 5.00
##  Mean   : 3.267          Mean   :0.9424           Mean   : 4.89
##  3rd Qu.: 4.000          3rd Qu.:2.0000           3rd Qu.: 6.00
##  Max.   :10.000          Max.   :5.0000           Max.   :13.00
##  aliph1HC_aromaticC_medlong aliph1HC_aromaticC_long aliph1HC_aromaticC_vlong
##  Min.   : 1.000             Min.   : 3.0           Min.   : 3.000
##  1st Qu.: 6.000             1st Qu.: 8.0           1st Qu.: 7.000
##  Median : 7.000             Median :10.0           Median : 8.000
##  Mean   : 7.337             Mean   :10.2           Mean   : 8.612
##  3rd Qu.: 8.000             3rd Qu.:12.0           3rd Qu.:10.000
##  Max.   :15.000             Max.   :21.0           Max.   :18.000
##  aliph1HC_scAGN_short aliph1HC_scAGN_medshort aliph1HC_scAGN_medlong
##  Min.   :0.0000       Min.   :0.000           Min.   :0.000
##  1st Qu.:0.0000       1st Qu.:0.000           1st Qu.:0.000
##  Median :0.0000       Median :1.000           Median :1.000
##  Mean   :0.2585       Mean   :1.038           Mean   :1.053
##  3rd Qu.:0.0000       3rd Qu.:2.000           3rd Qu.:2.000
##  Max.   :3.0000       Max.   :4.000           Max.   :5.000
##  aliph1HC_scAGN_long aliph1HC_scAGN_vlong aliph1HC_scArgN_long
##  Min.   :0.0000      Min.   :0.0000       Min.   :0.000
##  1st Qu.:0.0000      1st Qu.:0.0000       1st Qu.:2.000
##  Median :1.0000      Median :1.0000       Median :2.000
##  Mean   :0.8566      Mean   :0.9902       Mean   :1.977
##  3rd Qu.:1.0000      3rd Qu.:2.0000       3rd Qu.:2.000
##  Max.   :6.0000      Max.   :5.0000       Max.   :3.000
##  aliph1HC_bbProN_medlong aliph1HC_bbProN_long aliph1HC_bbProN_vlong
```

```
##  Min.   :0.0000          Min.   :0.000        Min.   :0.0000
##  1st Qu.:0.0000          1st Qu.:1.000        1st Qu.:0.0000
##  Median :0.0000          Median :2.000        Median :1.0000
##  Mean   :0.3947          Mean   :1.436        Mean   :0.6737
##  3rd Qu.:1.0000          3rd Qu.:2.000        3rd Qu.:1.0000
##  Max.   :2.0000          Max.   :4.000        Max.   :5.0000
##  aliph1HC_hydroxylO_medshort aliph1HC_hydroxylO_medlong aliph1HC_hydroxylO_long
##  Min.   :0.000               Min.   :0.0000             Min.   :0.000
##  1st Qu.:0.000               1st Qu.:0.0000             1st Qu.:2.000
##  Median :1.000               Median :0.0000             Median :2.000
##  Mean   :0.667               Mean   :0.5308             Mean   :2.451
##  3rd Qu.:1.000               3rd Qu.:1.0000             3rd Qu.:3.000
##  Max.   :3.000               Max.   :3.0000             Max.   :8.000
##  aliph1HC_hydroxylO_vlong aliph1HC_carbonylO_short aliph1HC_carbonylO_medshort
##  Min.   :0.000            Min.   :0.0000           Min.   :0.000
##  1st Qu.:2.000            1st Qu.:0.0000           1st Qu.:0.000
##  Median :3.000            Median :0.0000           Median :0.000
##  Mean   :2.819            Mean   :0.1408           Mean   :0.573
##  3rd Qu.:3.000            3rd Qu.:0.0000           3rd Qu.:1.000
##  Max.   :7.000            Max.   :2.0000           Max.   :3.000
##  aliph1HC_carbonylO_medlong aliph1HC_carbonylO_long aliph1HC_carbonylO_vlong
##  Min.   :0.0000             Min.   :0.000           Min.   :0.000
##  1st Qu.:0.0000             1st Qu.:1.000           1st Qu.:1.000
##  Median :1.0000             Median :1.000           Median :1.000
##  Mean   :0.8967             Mean   :1.434           Mean   :1.466
##  3rd Qu.:1.0000             3rd Qu.:2.000           3rd Qu.:2.000
##  Max.   :5.0000             Max.   :6.000           Max.   :6.000
##  aliph1HC_carboxylO_vlong aliph1HC_sulfur_short aliph1HC_sulfur_medshort
##  Min.   :0.0000           Min.   :0.0000       Min.   :0.0000
##  1st Qu.:0.0000           1st Qu.:0.0000       1st Qu.:0.0000
##  Median :1.0000           Median :0.0000       Median :1.0000
##  Mean   :0.8561           Mean   :0.1028       Mean   :0.6043
##  3rd Qu.:1.0000           3rd Qu.:0.0000       3rd Qu.:1.0000
##  Max.   :3.0000           Max.   :2.0000       Max.   :4.0000
##  aliph1HC_sulfur_medlong aliph1HC_sulfur_long aliph1HC_sulfur_vlong
##  Min.   :0.000           Min.   :0.000        Min.   :0.0000
##  1st Qu.:0.000           1st Qu.:1.000        1st Qu.:0.0000
##  Median :1.000           Median :1.000        Median :0.0000
##  Mean   :1.061           Mean   :1.149        Mean   :0.5997
##  3rd Qu.:2.000           3rd Qu.:2.000        3rd Qu.:1.0000
##  Max.   :5.000           Max.   :5.000        Max.   :4.0000
##  aliph1HC_bbN_short aliph1HC_bbN_medshort aliph1HC_bbN_medlong
##  Min.   :0.0000     Min.   :0.000         Min.   : 1.000
##  1st Qu.:0.0000     1st Qu.:0.000         1st Qu.: 5.000
##  Median :0.0000     Median :1.000         Median : 6.000
##  Mean   :0.1804     Mean   :1.332         Mean   : 6.382
##  3rd Qu.:0.0000     3rd Qu.:2.000         3rd Qu.: 7.000
##  Max.   :2.0000     Max.   :7.000         Max.   :13.000
##  aliph1HC_bbN_long aliph1HC_bbN_vlong aliph1HC_bbCA_medshort
##  Min.   : 4.00     Min.   : 3.000     Min.   :0.0000
##  1st Qu.: 6.00     1st Qu.: 7.000     1st Qu.:0.0000
##  Median : 8.00     Median : 8.000     Median :1.0000
##  Mean   : 7.86     Mean   : 8.439     Mean   :0.9548
##  3rd Qu.: 9.00     3rd Qu.:10.000     3rd Qu.:1.0000
```

```
##  Max.   :17.00     Max.   :16.000     Max.   :5.0000
##  aliph1HC_bbCA_medlong aliph1HC_bbCA_long aliph1HC_bbCA_vlong
##  Min.   : 2.000        Min.   : 2.00      Min.   : 5.00
##  1st Qu.: 5.000        1st Qu.: 6.00      1st Qu.:11.00
##  Median : 6.000        Median : 8.00      Median :12.00
##  Mean   : 6.386        Mean   : 7.94      Mean   :12.46
##  3rd Qu.: 8.000        3rd Qu.: 9.00      3rd Qu.:14.00
##  Max.   :13.000        Max.   :17.00      Max.   :21.00
##  aliph1HC_bbC_medshort aliph1HC_bbC_medlong aliph1HC_bbC_long
##  Min.   :0.000         Min.   : 3.000       Min.   : 1.000
##  1st Qu.:1.000         1st Qu.: 6.000       1st Qu.: 5.000
##  Median :1.000         Median : 7.000       Median : 6.000
##  Mean   :1.705         Mean   : 7.076       Mean   : 6.089
##  3rd Qu.:2.000         3rd Qu.: 8.000       3rd Qu.: 7.000
##  Max.   :6.000         Max.   :16.000       Max.   :14.000
##  aliph1HC_bbC_vlong aliph1HC_bbO_short aliph1HC_bbO_medshort
##  Min.   : 3.000     Min.   :0.0000     Min.   :2.00
##  1st Qu.: 8.000     1st Qu.:0.0000     1st Qu.:3.00
##  Median :10.000     Median :0.0000     Median :4.00
##  Mean   : 9.658     Mean   :0.5262     Mean   :4.22
##  3rd Qu.:11.000     3rd Qu.:1.0000     3rd Qu.:5.00
##  Max.   :19.000     Max.   :4.0000     Max.   :9.00
##  aliph1HC_bbO_medlong aliph1HC_bbO_long aliph1HC_bbO_vlong
##  Min.   : 0.000       Min.   : 1.000    Min.   : 3.000
##  1st Qu.: 2.000       1st Qu.: 4.000    1st Qu.: 8.000
##  Median : 3.000       Median : 5.000    Median :10.000
##  Mean   : 3.528       Mean   : 5.654    Mean   : 9.713
##  3rd Qu.: 5.000       3rd Qu.: 7.000    3rd Qu.:11.000
##  Max.   :10.000       Max.   :13.000    Max.   :19.000
##  aliph2HC_aliph2HC_short aliph2HC_aliph2HC_medshort aliph2HC_aliph2HC_medlong
##  Min.   :0.0000          Min.   : 2.000             Min.   : 5.00
##  1st Qu.:0.0000          1st Qu.: 6.000             1st Qu.:11.00
##  Median :0.0000          Median : 8.000             Median :13.00
##  Mean   :0.5668          Mean   : 7.718             Mean   :13.73
##  3rd Qu.:1.0000          3rd Qu.: 9.000             3rd Qu.:16.00
##  Max.   :5.0000          Max.   :17.000             Max.   :32.00
##  aliph2HC_aliph2HC_long aliph2HC_aliph2HC_vlong aliph2HC_aliph3HC_vshort
##  Min.   : 9.00          Min.   :17.00           Min.   :0.0000
##  1st Qu.:17.00          1st Qu.:30.00           1st Qu.:0.0000
##  Median :20.00          Median :32.00           Median :0.0000
##  Mean   :20.16          Mean   :32.63           Mean   :0.2359
##  3rd Qu.:23.00          3rd Qu.:36.00           3rd Qu.:0.0000
##  Max.   :35.00          Max.   :49.00           Max.   :2.0000
##  aliph2HC_aliph3HC_short aliph2HC_aliph3HC_medshort aliph2HC_aliph3HC_medlong
##  Min.   :0.00            Min.   : 0.00              Min.   : 2.00
##  1st Qu.:1.00            1st Qu.: 5.00              1st Qu.:10.00
##  Median :2.00            Median : 6.00              Median :12.00
##  Mean   :1.94            Mean   : 6.41              Mean   :11.87
##  3rd Qu.:3.00            3rd Qu.: 8.00              3rd Qu.:14.00
##  Max.   :7.00            Max.   :16.00              Max.   :25.00
##  aliph2HC_aliph3HC_long aliph2HC_aliph3HC_vlong aliph2HC_aromaticC_short
##  Min.   : 6.00          Min.   : 7.00           Min.   : 0.000
##  1st Qu.:13.00          1st Qu.:16.00           1st Qu.: 2.000
##  Median :16.00          Median :19.00           Median : 3.000
```

```
##  Mean   :16.18          Mean   :19.02          Mean   : 3.622
##  3rd Qu.:19.00          3rd Qu.:22.00          3rd Qu.: 5.000
##  Max.   :35.00          Max.   :34.00          Max.   :15.000
##  aliph2HC_aromaticC_medshort aliph2HC_aromaticC_medlong aliph2HC_aromaticC_long
##  Min.   : 4.00               Min.   :15.00              Min.   :28.00
##  1st Qu.:12.00               1st Qu.:27.00              1st Qu.:40.00
##  Median :15.00               Median :30.00              Median :44.00
##  Mean   :15.24               Mean   :31.29              Mean   :44.73
##  3rd Qu.:18.00               3rd Qu.:35.00              3rd Qu.:49.00
##  Max.   :40.00               Max.   :62.00              Max.   :71.00
##  aliph2HC_aromaticC_vlong aliph2HC_scAGN_vshort aliph2HC_scAGN_short
##  Min.   :42.0             Min.   :0.0000        Min.   : 0.000
##  1st Qu.:59.0             1st Qu.:0.0000        1st Qu.: 1.000
##  Median :65.0             Median :0.0000        Median : 2.000
##  Mean   :64.4             Mean   :0.1182        Mean   : 1.648
##  3rd Qu.:70.0             3rd Qu.:0.0000        3rd Qu.: 2.000
##  Max.   :90.0             Max.   :3.0000        Max.   :11.000
##  aliph2HC_scAGN_medshort aliph2HC_scAGN_medlong aliph2HC_scAGN_long
##  Min.   :0.000           Min.   : 0.000         Min.   : 0.000
##  1st Qu.:2.000           1st Qu.: 3.000         1st Qu.: 3.000
##  Median :3.000           Median : 4.000         Median : 4.000
##  Mean   :2.843           Mean   : 4.138         Mean   : 4.618
##  3rd Qu.:4.000           3rd Qu.: 5.000         3rd Qu.: 6.000
##  Max.   :9.000           Max.   :11.000         Max.   :14.000
##  aliph2HC_scAGN_vlong aliph2HC_scLysN_medshort aliph2HC_scLysN_medlong
##  Min.   : 1.000       Min.   :0.00             Min.   :0.0000
##  1st Qu.: 6.000       1st Qu.:0.00             1st Qu.:0.0000
##  Median : 8.000       Median :0.00             Median :1.0000
##  Mean   : 8.101       Mean   :0.26             Mean   :0.7477
##  3rd Qu.:10.000       3rd Qu.:0.00             3rd Qu.:1.0000
##  Max.   :18.000       Max.   :3.00             Max.   :4.0000
##  aliph2HC_scLysN_long aliph2HC_scLysN_vlong aliph2HC_scArgN_medshort
##  Min.   :0.000        Min.   :0.00          Min.   :1.000
##  1st Qu.:1.000        1st Qu.:1.00          1st Qu.:2.000
##  Median :1.000        Median :2.00          Median :2.000
##  Mean   :1.566        Mean   :2.11          Mean   :1.884
##  3rd Qu.:2.000        3rd Qu.:3.00          3rd Qu.:2.000
##  Max.   :7.000        Max.   :7.00          Max.   :4.000
##  aliph2HC_scArgN_medlong aliph2HC_scArgN_long aliph2HC_scArgN_vlong
##  Min.   :0.0000          Min.   :0.000        Min.   :2.000
##  1st Qu.:1.0000          1st Qu.:1.000        1st Qu.:4.000
##  Median :1.0000          Median :1.000        Median :5.000
##  Mean   :0.8469          Mean   :1.135        Mean   :4.641
##  3rd Qu.:1.0000          3rd Qu.:2.000        3rd Qu.:5.000
##  Max.   :4.0000          Max.   :5.000        Max.   :7.000
##  aliph2HC_bbProN_medshort aliph2HC_bbProN_medlong aliph2HC_bbProN_long
##  Min.   :0.0000           Min.   :0.00            Min.   :0.000
##  1st Qu.:0.0000           1st Qu.:1.00            1st Qu.:1.000
##  Median :1.0000           Median :1.00            Median :2.000
##  Mean   :0.7323           Mean   :1.47            Mean   :1.907
##  3rd Qu.:1.0000           3rd Qu.:2.00            3rd Qu.:3.000
##  Max.   :4.0000           Max.   :5.00            Max.   :9.000
##  aliph2HC_bbProN_vlong aliph2HC_hydroxylO_short aliph2HC_hydroxylO_medshort
##  Min.   :1.000         Min.   :0.000            Min.   : 2.000
```

```
##  1st Qu.:1.000         1st Qu.:0.000           1st Qu.: 5.000
##  Median :2.000         Median :1.000           Median : 6.000
##  Mean   :2.408         Mean   :1.028           Mean   : 5.867
##  3rd Qu.:3.000         3rd Qu.:2.000           3rd Qu.: 7.000
##  Max.   :9.000         Max.   :7.000           Max.   :14.000
##  aliph2HC_hydroxylO_medlong aliph2HC_hydroxylO_long aliph2HC_hydroxylO_vlong
##  Min.   : 1.000             Min.   : 5.00           Min.   : 8.00
##  1st Qu.: 5.000             1st Qu.:12.00           1st Qu.:14.00
##  Median : 6.000             Median :13.00           Median :17.00
##  Mean   : 6.764             Mean   :13.42           Mean   :16.61
##  3rd Qu.: 8.000             3rd Qu.:15.00           3rd Qu.:19.00
##  Max.   :17.000             Max.   :22.00           Max.   :28.00
##  aliph2HC_carbonylO_vshort aliph2HC_carbonylO_short aliph2HC_carbonylO_medshort
##  Min.   :0.0000            Min.   :0.00            Min.   :0.000
##  1st Qu.:0.0000            1st Qu.:1.00            1st Qu.:1.000
##  Median :0.0000            Median :2.00            Median :2.000
##  Mean   :0.2456            Mean   :1.98            Mean   :2.213
##  3rd Qu.:0.0000            3rd Qu.:3.00            3rd Qu.:3.000
##  Max.   :3.0000            Max.   :7.00            Max.   :9.000
##  aliph2HC_carbonylO_medlong aliph2HC_carbonylO_long aliph2HC_carbonylO_vlong
##  Min.   : 0.000             Min.   : 0.000          Min.   : 1.00
##  1st Qu.: 2.000             1st Qu.: 3.000          1st Qu.: 6.00
##  Median : 3.000             Median : 5.000          Median : 8.00
##  Mean   : 3.253             Mean   : 5.025          Mean   : 7.88
##  3rd Qu.: 4.000             3rd Qu.: 6.000          3rd Qu.:10.00
##  Max.   :11.000             Max.   :14.000          Max.   :20.00
##  aliph2HC_carboxylO_short aliph2HC_carboxylO_medshort
##  Min.   :1.000            Min.   :1.000
##  1st Qu.:1.000            1st Qu.:1.000
##  Median :2.000            Median :2.000
##  Mean   :1.879            Mean   :2.188
##  3rd Qu.:3.000            3rd Qu.:3.000
##  Max.   :8.000            Max.   :9.000
##  aliph2HC_carboxylO_medlong aliph2HC_carboxylO_long aliph2HC_carboxylO_vlong
##  Min.   :0.000              Min.   : 0.000          Min.   : 1.000
##  1st Qu.:0.000              1st Qu.: 1.000          1st Qu.: 5.000
##  Median :1.000              Median : 1.000          Median : 7.000
##  Mean   :1.458              Mean   : 2.008          Mean   : 7.194
##  3rd Qu.:2.000              3rd Qu.: 3.000          3rd Qu.: 9.000
##  Max.   :8.000              Max.   :16.000          Max.   :19.000
##  aliph2HC_sulfur_short aliph2HC_sulfur_medshort aliph2HC_sulfur_medlong
##  Min.   :0.0000        Min.   :0.000           Min.   :0.000
##  1st Qu.:0.0000        1st Qu.:1.000           1st Qu.:1.000
##  Median :1.0000        Median :2.000           Median :2.000
##  Mean   :0.6608        Mean   :1.666           Mean   :2.317
##  3rd Qu.:1.0000        3rd Qu.:3.000           3rd Qu.:3.000
##  Max.   :4.0000        Max.   :7.000           Max.   :8.000
##  aliph2HC_sulfur_long aliph2HC_sulfur_vlong aliph2HC_bbN_short
##  Min.   : 0.000       Min.   : 0.000        Min.   :0.0000
##  1st Qu.: 2.000       1st Qu.: 3.000        1st Qu.:0.0000
##  Median : 3.000       Median : 4.000        Median :0.0000
##  Mean   : 3.056       Mean   : 3.808        Mean   :0.5663
##  3rd Qu.: 4.000       3rd Qu.: 5.000        3rd Qu.:1.0000
##  Max.   :10.000       Max.   :11.000        Max.   :5.0000
```

```
##  aliph2HC_bbN_medshort aliph2HC_bbN_medlong aliph2HC_bbN_long
##  Min.   : 7.00         Min.   :16.00        Min.   :25
##  1st Qu.:12.00         1st Qu.:24.00        1st Qu.:37
##  Median :14.00         Median :27.00        Median :41
##  Mean   :14.58         Mean   :27.54        Mean   :41
##  3rd Qu.:16.00         3rd Qu.:31.00        3rd Qu.:45
##  Max.   :26.00         Max.   :46.00        Max.   :65
##  aliph2HC_bbN_vlong aliph2HC_bbCA_short aliph2HC_bbCA_medshort
##  Min.   :42.00      Min.   :0.0000      Min.   : 4.00
##  1st Qu.:56.00      1st Qu.:0.0000      1st Qu.: 9.00
##  Median :61.00      Median :1.0000      Median :10.00
##  Mean   :61.27      Mean   :0.7862      Mean   :10.72
##  3rd Qu.:66.00      3rd Qu.:1.0000      3rd Qu.:12.00
##  Max.   :89.00      Max.   :6.0000      Max.   :24.00
##  aliph2HC_bbCA_medlong aliph2HC_bbCA_long aliph2HC_bbCA_vlong
##  Min.   :25.00         Min.   :24.00      Min.   :35.00
##  1st Qu.:34.00         1st Qu.:36.00      1st Qu.:48.00
##  Median :37.00         Median :41.00      Median :53.00
##  Mean   :37.24         Mean   :40.92      Mean   :53.84
##  3rd Qu.:40.00         3rd Qu.:45.75      3rd Qu.:59.00
##  Max.   :53.00         Max.   :68.00      Max.   :78.00
##  aliph2HC_bbC_short aliph2HC_bbC_medshort aliph2HC_bbC_medlong
##  Min.   :0.000      Min.   : 7.00         Min.   :15.00
##  1st Qu.:1.000      1st Qu.:12.00         1st Qu.:29.00
##  Median :1.000      Median :14.00         Median :32.00
##  Mean   :1.381      Mean   :14.41         Mean   :32.08
##  3rd Qu.:2.000      3rd Qu.:16.00         3rd Qu.:35.00
##  Max.   :6.000      Max.   :26.00         Max.   :56.00
##  aliph2HC_bbC_long aliph2HC_bbC_vlong aliph2HC_bbO_vshort aliph2HC_bbO_short
##  Min.   :26.00     Min.   :33.00      Min.   :0.0         Min.   : 2.000
##  1st Qu.:37.00     1st Qu.:50.00      1st Qu.:0.0         1st Qu.: 7.000
##  Median :42.00     Median :55.00      Median :0.0         Median : 8.000
##  Mean   :42.24     Mean   :55.21      Mean   :0.5         Mean   : 8.364
##  3rd Qu.:47.00     3rd Qu.:60.00      3rd Qu.:1.0         3rd Qu.:10.000
##  Max.   :63.00     Max.   :86.00      Max.   :4.0         Max.   :16.000
##  aliph2HC_bbO_medshort aliph2HC_bbO_medlong aliph2HC_bbO_long
##  Min.   : 6.00         Min.   :14.00        Min.   :23.00
##  1st Qu.:12.00         1st Qu.:24.00        1st Qu.:36.00
##  Median :14.00         Median :27.00        Median :40.00
##  Mean   :14.25         Mean   :27.28        Mean   :40.08
##  3rd Qu.:16.00         3rd Qu.:30.00        3rd Qu.:44.00
##  Max.   :28.00         Max.   :47.00        Max.   :66.00
##  aliph2HC_bbO_vlong aliph3HC_aliph3HC_short aliph3HC_aliph3HC_medshort
##  Min.   :36.00      Min.   :0.000           Min.   :0.000
##  1st Qu.:51.00      1st Qu.:0.000           1st Qu.:1.000
##  Median :56.00      Median :1.000           Median :1.000
##  Mean   :56.36      Mean   :1.155           Mean   :1.362
##  3rd Qu.:62.00      3rd Qu.:2.000           3rd Qu.:2.000
##  Max.   :81.00      Max.   :5.000           Max.   :6.000
##  aliph3HC_aliph3HC_medlong aliph3HC_aliph3HC_long aliph3HC_aliph3HC_vlong
##  Min.   :0.000             Min.   :0.000          Min.   : 0.000
##  1st Qu.:1.000             1st Qu.:1.000          1st Qu.: 3.000
##  Median :2.000             Median :2.000          Median : 4.000
##  Mean   :1.728             Mean   :2.504          Mean   : 3.982
```

```
##  3rd Qu.:2.000             3rd Qu.:3.000          3rd Qu.: 5.000
##  Max.   :6.000             Max.   :9.000          Max.   :10.000
##  aliph3HC_aromaticC_short aliph3HC_aromaticC_medshort
##  Min.   : 0.000           Min.   : 3.00
##  1st Qu.: 3.000           1st Qu.: 8.00
##  Median : 5.000           Median :10.00
##  Mean   : 5.368           Mean   :10.94
##  3rd Qu.: 8.000           3rd Qu.:13.00
##  Max.   :16.000           Max.   :24.00
##  aliph3HC_aromaticC_medlong aliph3HC_aromaticC_long aliph3HC_aromaticC_vlong
##  Min.   : 7.00              Min.   : 4.00           Min.   : 9.00
##  1st Qu.:15.00              1st Qu.:17.00           1st Qu.:18.00
##  Median :17.00              Median :21.00           Median :21.00
##  Mean   :17.17              Mean   :20.22           Mean   :21.58
##  3rd Qu.:20.00              3rd Qu.:24.00           3rd Qu.:25.00
##  Max.   :35.00              Max.   :35.00           Max.   :38.00
##  aliph3HC_scAGN_short aliph3HC_scAGN_medshort aliph3HC_scAGN_medlong
##  Min.   :0.000        Min.   :0.000           Min.   :0.000
##  1st Qu.:0.000        1st Qu.:1.000           1st Qu.:1.000
##  Median :1.000        Median :1.000           Median :2.000
##  Mean   :1.058        Mean   :1.458           Mean   :2.026
##  3rd Qu.:2.000        3rd Qu.:2.000           3rd Qu.:3.000
##  Max.   :6.000        Max.   :6.000           Max.   :8.000
##  aliph3HC_scAGN_long aliph3HC_scAGN_vlong aliph3HC_scLysN_vlong
##  Min.   :0.000       Min.   :0.000        Min.   :0.0000
##  1st Qu.:1.000       1st Qu.:2.000        1st Qu.:0.0000
##  Median :2.000       Median :2.000        Median :0.0000
##  Mean   :2.187       Mean   :2.637        Mean   :0.1978
##  3rd Qu.:3.000       3rd Qu.:3.000        3rd Qu.:0.0000
##  Max.   :8.000       Max.   :9.000        Max.   :2.0000
##  aliph3HC_scArgN_medlong aliph3HC_scArgN_long aliph3HC_scArgN_vlong
##  Min.   :0.0000          Min.   :0.000        Min.   :0.0000
##  1st Qu.:0.0000          1st Qu.:1.000        1st Qu.:0.0000
##  Median :1.0000          Median :1.000        Median :1.0000
##  Mean   :0.6295          Mean   :1.355        Mean   :0.6742
##  3rd Qu.:1.0000          3rd Qu.:2.000        3rd Qu.:1.0000
##  Max.   :5.0000          Max.   :5.000        Max.   :5.0000
##  aliph3HC_bbProN_medlong aliph3HC_bbProN_long aliph3HC_bbProN_vlong
##  Min.   :0.000           Min.   :0.000        Min.   :0.000
##  1st Qu.:1.000           1st Qu.:1.000        1st Qu.:1.000
##  Median :1.000           Median :1.000        Median :1.000
##  Mean   :1.058           Mean   :1.546        Mean   :1.594
##  3rd Qu.:1.000           3rd Qu.:2.000        3rd Qu.:2.000
##  Max.   :4.000           Max.   :6.000        Max.   :6.000
##  aliph3HC_hydroxylO_short aliph3HC_hydroxylO_medshort
##  Min.   :0.0000           Min.   :0.000
##  1st Qu.:0.0000           1st Qu.:1.000
##  Median :0.0000           Median :2.000
##  Mean   :0.6675           Mean   :1.661
##  3rd Qu.:1.0000           3rd Qu.:2.000
##  Max.   :5.0000           Max.   :6.000
##  aliph3HC_hydroxylO_medlong aliph3HC_hydroxylO_long aliph3HC_hydroxylO_vlong
##  Min.   :0.000              Min.   : 1.000          Min.   : 1.000
##  1st Qu.:2.000              1st Qu.: 3.000          1st Qu.: 4.000
```

```
##  Median :2.000              Median : 4.000          Median : 5.000
##  Mean   :2.626              Mean   : 4.182          Mean   : 5.129
##  3rd Qu.:3.000              3rd Qu.: 5.000          3rd Qu.: 6.000
##  Max.   :8.000              Max.   :10.000          Max.   :12.000
##  aliph3HC_carbonylO_short aliph3HC_carbonylO_medshort
##  Min.   :0.0000           Min.   :0.000
##  1st Qu.:0.0000           1st Qu.:1.000
##  Median :0.0000           Median :1.000
##  Mean   :0.6413           Mean   :1.298
##  3rd Qu.:1.0000           3rd Qu.:2.000
##  Max.   :4.0000           Max.   :6.000
##  aliph3HC_carbonylO_medlong aliph3HC_carbonylO_long aliph3HC_carbonylO_vlong
##  Min.   :0.000              Min.   :0.000           Min.   : 0.000
##  1st Qu.:1.000              1st Qu.:1.000           1st Qu.: 2.000
##  Median :2.000              Median :2.000           Median : 3.000
##  Mean   :2.114              Mean   :2.262           Mean   : 2.919
##  3rd Qu.:3.000              3rd Qu.:3.000           3rd Qu.: 4.000
##  Max.   :8.000              Max.   :8.000           Max.   :10.000
##  aliph3HC_carboxylO_medshort aliph3HC_carboxylO_medlong aliph3HC_carboxylO_long
##  Min.   :0.000                Min.   :0.0000            Min.   :0.000
##  1st Qu.:0.000                1st Qu.:0.0000            1st Qu.:0.000
##  Median :0.000                Median :0.0000            Median :0.000
##  Mean   :0.147                Mean   :0.1958            Mean   :0.221
##  3rd Qu.:0.000                3rd Qu.:0.0000            3rd Qu.:0.000
##  Max.   :4.000                Max.   :5.0000            Max.   :4.000
##  aliph3HC_carboxylO_vlong aliph3HC_sulfur_short aliph3HC_sulfur_medshort
##  Min.   :0.000            Min.   :0.0000       Min.   :0.00
##  1st Qu.:1.000            1st Qu.:0.0000       1st Qu.:1.00
##  Median :1.000            Median :0.0000       Median :2.00
##  Mean   :1.313            Mean   :0.5339       Mean   :1.82
##  3rd Qu.:2.000            3rd Qu.:1.0000       3rd Qu.:2.00
##  Max.   :5.000            Max.   :3.0000       Max.   :6.00
##  aliph3HC_sulfur_medlong aliph3HC_sulfur_long aliph3HC_sulfur_vlong
##  Min.   :0.000           Min.   :0.000        Min.   :0.000
##  1st Qu.:1.000           1st Qu.:1.000        1st Qu.:1.000
##  Median :1.000           Median :1.000        Median :2.000
##  Mean   :1.266           Mean   :1.491        Mean   :1.811
##  3rd Qu.:2.000           3rd Qu.:2.000        3rd Qu.:3.000
##  Max.   :5.000           Max.   :6.000        Max.   :8.000
##  aliph3HC_bbN_short aliph3HC_bbN_medshort aliph3HC_bbN_medlong
##  Min.   :0.0000     Min.   : 0.000        Min.   : 3.00
##  1st Qu.:0.0000     1st Qu.: 3.000        1st Qu.: 8.00
##  Median :1.0000     Median : 4.000        Median :10.00
##  Mean   :0.8232     Mean   : 4.439        Mean   :10.35
##  3rd Qu.:1.0000     3rd Qu.: 6.000        3rd Qu.:12.00
##  Max.   :5.0000     Max.   :11.000        Max.   :22.00
##  aliph3HC_bbN_long aliph3HC_bbN_vlong aliph3HC_bbCA_short
##  Min.   : 7.00     Min.   :15.00      Min.   :0.0000
##  1st Qu.:14.00     1st Qu.:22.00      1st Qu.:0.0000
##  Median :16.00     Median :24.00      Median :1.0000
##  Mean   :16.35     Mean   :24.35      Mean   :0.6362
##  3rd Qu.:19.00     3rd Qu.:27.00      3rd Qu.:1.0000
##  Max.   :29.00     Max.   :40.00      Max.   :3.0000
##  aliph3HC_bbCA_medshort aliph3HC_bbCA_medlong aliph3HC_bbCA_long
```

```
##  Min.   : 1.000         Min.   : 2.00         Min.   : 7.00
##  1st Qu.: 5.000         1st Qu.: 8.00         1st Qu.:15.00
##  Median : 6.000         Median :10.00         Median :17.00
##  Mean   : 6.187         Mean   :10.22         Mean   :17.49
##  3rd Qu.: 7.000         3rd Qu.:12.00         3rd Qu.:19.00
##  Max.   :14.000         Max.   :24.00         Max.   :30.00
##  aliph3HC_bbCA_vlong aliph3HC_bbC_short aliph3HC_bbC_medshort
##  Min.   : 8.0        Min.   :0.0000     Min.   : 0.000
##  1st Qu.:18.0        1st Qu.:0.0000     1st Qu.: 4.000
##  Median :21.0        Median :1.0000     Median : 6.000
##  Mean   :20.9        Mean   :0.8736     Mean   : 5.793
##  3rd Qu.:24.0        3rd Qu.:1.7500     3rd Qu.: 7.000
##  Max.   :37.0        Max.   :5.0000     Max.   :14.000
##  aliph3HC_bbC_medlong aliph3HC_bbC_long aliph3HC_bbC_vlong aliph3HC_bbO_short
##  Min.   : 4.00        Min.   : 6.0      Min.   : 9.00      Min.   :0.000
##  1st Qu.: 9.00        1st Qu.:14.0      1st Qu.:19.00      1st Qu.:2.000
##  Median :11.00        Median :17.0      Median :21.00      Median :3.000
##  Mean   :10.71        Mean   :16.9      Mean   :21.57      Mean   :3.121
##  3rd Qu.:12.00        3rd Qu.:20.0      3rd Qu.:24.00      3rd Qu.:4.000
##  Max.   :21.00        Max.   :30.0      Max.   :35.00      Max.   :8.000
##  aliph3HC_bbO_medshort aliph3HC_bbO_medlong aliph3HC_bbO_long
##  Min.   : 2.000        Min.   : 0.00        Min.   : 5.00
##  1st Qu.: 6.000        1st Qu.: 5.00        1st Qu.:13.00
##  Median : 8.000        Median : 6.00        Median :16.00
##  Mean   : 7.884        Mean   : 6.75        Mean   :15.56
##  3rd Qu.: 9.000        3rd Qu.: 9.00        3rd Qu.:18.00
##  Max.   :16.000        Max.   :16.00        Max.   :27.00
##  aliph3HC_bbO_vlong aromaticC_aromaticC_short aromaticC_aromaticC_medshort
##  Min.   :11.00      Min.   : 0.000            Min.   : 0.00
##  1st Qu.:20.00      1st Qu.: 4.000            1st Qu.:17.00
##  Median :22.00      Median : 6.000            Median :22.00
##  Mean   :22.99      Mean   : 5.714            Mean   :21.31
##  3rd Qu.:26.00      3rd Qu.: 7.000            3rd Qu.:26.00
##  Max.   :41.00      Max.   :16.000            Max.   :41.00
##  aromaticC_aromaticC_medlong aromaticC_aromaticC_long aromaticC_aromaticC_vlong
##  Min.   : 4.00               Min.   :10.00            Min.   : 8.00
##  1st Qu.:31.00               1st Qu.:30.00            1st Qu.:28.00
##  Median :38.00               Median :36.00            Median :34.00
##  Mean   :37.86               Mean   :37.16            Mean   :34.29
##  3rd Qu.:44.00               3rd Qu.:46.00            3rd Qu.:40.00
##  Max.   :64.00               Max.   :70.00            Max.   :59.00
##  aromaticC_scAGN_short aromaticC_scAGN_medshort aromaticC_scAGN_medlong
##  Min.   :0.000         Min.   : 0.000           Min.   : 0.000
##  1st Qu.:0.000         1st Qu.: 2.000           1st Qu.: 4.000
##  Median :2.000         Median : 3.000           Median : 6.000
##  Mean   :1.803         Mean   : 3.428           Mean   : 5.622
##  3rd Qu.:3.000         3rd Qu.: 5.000           3rd Qu.: 8.000
##  Max.   :9.000         Max.   :10.000           Max.   :15.000
##  aromaticC_scAGN_long aromaticC_scAGN_vlong aromaticC_scLysN_medlong
##  Min.   : 0.000       Min.   : 0.000        Min.   :0.0000
##  1st Qu.: 6.000       1st Qu.: 5.000        1st Qu.:0.0000
##  Median : 7.000       Median : 7.000        Median :0.0000
##  Mean   : 7.128       Mean   : 7.186        Mean   :0.6269
##  3rd Qu.: 8.000       3rd Qu.: 9.000        3rd Qu.:1.0000
```

```
##  Max.   :16.000       Max.   :23.000        Max.   :4.0000
##  aromaticC_scLysN_long aromaticC_scLysN_vlong aromaticC_scArgN_medshort
##  Min.   :0.000         Min.   :0.00           Min.   :0.0000
##  1st Qu.:1.000         1st Qu.:2.00           1st Qu.:0.0000
##  Median :2.000         Median :2.00           Median :0.0000
##  Mean   :1.703         Mean   :2.22           Mean   :0.1459
##  3rd Qu.:2.000         3rd Qu.:3.00           3rd Qu.:0.0000
##  Max.   :6.000         Max.   :7.00           Max.   :4.0000
##  aromaticC_scArgN_medlong aromaticC_scArgN_long aromaticC_scArgN_vlong
##  Min.   :0.0000           Min.   : 0.0000       Min.   : 0.0000
##  1st Qu.:0.0000           1st Qu.: 0.0000       1st Qu.: 0.0000
##  Median :0.0000           Median : 0.0000       Median : 0.0000
##  Mean   :0.4239           Mean   : 0.3489       Mean   : 0.9039
##  3rd Qu.:0.0000           3rd Qu.: 0.0000       3rd Qu.: 1.0000
##  Max.   :9.0000           Max.   :11.0000       Max.   :13.0000
##  aromaticC_bbProN_medlong aromaticC_bbProN_long aromaticC_bbProN_vlong
##  Min.   :0.000            Min.   :0.000         Min.   :0.000
##  1st Qu.:0.000            1st Qu.:1.000         1st Qu.:2.000
##  Median :0.000            Median :2.000         Median :2.000
##  Mean   :0.518            Mean   :1.857         Mean   :2.544
##  3rd Qu.:1.000            3rd Qu.:2.000         3rd Qu.:4.000
##  Max.   :6.000            Max.   :8.000         Max.   :9.000
##  aromaticC_hydroxylO_vshort aromaticC_hydroxylO_short
##  Min.   :0.0000             Min.   : 0.000
##  1st Qu.:0.0000             1st Qu.: 2.000
##  Median :0.0000             Median : 3.000
##  Mean   :0.1213             Mean   : 2.837
##  3rd Qu.:0.0000             3rd Qu.: 4.000
##  Max.   :2.0000             Max.   :11.000
##  aromaticC_hydroxylO_medshort aromaticC_hydroxylO_medlong
##  Min.   : 0.000               Min.   : 0.000
##  1st Qu.: 4.000               1st Qu.: 6.000
##  Median : 5.000               Median : 7.000
##  Mean   : 5.114               Mean   : 7.585
##  3rd Qu.: 7.000               3rd Qu.: 9.000
##  Max.   :14.000               Max.   :16.000
##  aromaticC_hydroxylO_long aromaticC_hydroxylO_vlong aromaticC_carbonylO_short
##  Min.   : 2.00            Min.   : 2.00             Min.   :0.000
##  1st Qu.:10.00            1st Qu.:16.00             1st Qu.:1.000
##  Median :13.00            Median :18.00             Median :2.000
##  Mean   :12.17            Mean   :18.41             Mean   :1.481
##  3rd Qu.:15.00            3rd Qu.:21.00             3rd Qu.:2.000
##  Max.   :25.00            Max.   :34.00             Max.   :8.000
##  aromaticC_carbonylO_medshort aromaticC_carbonylO_medlong
##  Min.   : 0.00                Min.   : 0.000
##  1st Qu.: 2.00                1st Qu.: 3.000
##  Median : 3.00                Median : 4.000
##  Mean   : 2.74                Mean   : 4.559
##  3rd Qu.: 4.00                3rd Qu.: 6.000
##  Max.   :10.00                Max.   :14.000
##  aromaticC_carbonylO_long aromaticC_carbonylO_vlong
##  Min.   : 1.00            Min.   : 0.000
##  1st Qu.: 6.00            1st Qu.: 5.000
##  Median : 7.00            Median : 7.000
```

```
##  Mean   : 6.92            Mean   : 7.671
##  3rd Qu.: 8.00            3rd Qu.: 9.000
##  Max.   :15.00            Max.   :21.000
##  aromaticC_carboxylO_medshort aromaticC_carboxylO_medlong
##  Min.   :0.0000               Min.   : 0.0000
##  1st Qu.:0.0000               1st Qu.: 0.0000
##  Median :0.0000               Median : 0.0000
##  Mean   :0.4522               Mean   : 0.8772
##  3rd Qu.:0.0000               3rd Qu.: 0.0000
##  Max.   :7.0000               Max.   :11.0000
##  aromaticC_carboxylO_long aromaticC_carboxylO_vlong aromaticC_sulfur_short
##  Min.   : 0.000           Min.   : 0.000            Min.   :0.000
##  1st Qu.: 0.000           1st Qu.: 0.000            1st Qu.:0.000
##  Median : 0.000           Median : 1.000            Median :1.000
##  Mean   : 1.134           Mean   : 1.543            Mean   :1.349
##  3rd Qu.: 1.750           3rd Qu.: 2.000            3rd Qu.:2.000
##  Max.   :14.000           Max.   :18.000            Max.   :8.000
##  aromaticC_sulfur_medshort aromaticC_sulfur_medlong aromaticC_sulfur_long
##  Min.   :0.0               Min.   :0.000            Min.   : 0.000
##  1st Qu.:2.0               1st Qu.:1.000            1st Qu.: 1.000
##  Median :2.0               Median :2.000            Median : 2.000
##  Mean   :2.4               Mean   :1.739            Mean   : 2.513
##  3rd Qu.:3.0               3rd Qu.:2.000            3rd Qu.: 4.000
##  Max.   :9.0               Max.   :8.000            Max.   :10.000
##  aromaticC_sulfur_vlong aromaticC_bbN_short aromaticC_bbN_medshort
##  Min.   : 1.000         Min.   : 0.0000     Min.   : 1.00
##  1st Qu.: 3.000         1st Qu.: 0.0000     1st Qu.: 6.00
##  Median : 5.000         Median : 0.0000     Median : 9.00
##  Mean   : 4.708         Mean   : 0.9373     Mean   : 8.62
##  3rd Qu.: 6.000         3rd Qu.: 2.0000     3rd Qu.:11.00
##  Max.   :13.000         Max.   :10.0000     Max.   :19.00
##  aromaticC_bbN_medlong aromaticC_bbN_long aromaticC_bbN_vlong
##  Min.   :15.00         Min.   :32.00      Min.   :49.00
##  1st Qu.:23.00         1st Qu.:46.00      1st Qu.:64.00
##  Median :26.00         Median :51.00      Median :70.00
##  Mean   :26.06         Mean   :50.59      Mean   :69.88
##  3rd Qu.:29.00         3rd Qu.:55.00      3rd Qu.:75.00
##  Max.   :42.00         Max.   :70.00      Max.   :93.00
##  aromaticC_bbCA_short aromaticC_bbCA_medshort aromaticC_bbCA_medlong
##  Min.   :0.000        Min.   : 3.0           Min.   :14.0
##  1st Qu.:1.000        1st Qu.: 9.0           1st Qu.:25.0
##  Median :2.000        Median :11.0           Median :28.0
##  Mean   :2.106        Mean   :11.1           Mean   :28.8
##  3rd Qu.:3.000        3rd Qu.:13.0           3rd Qu.:32.0
##  Max.   :9.000        Max.   :26.0           Max.   :46.0
##  aromaticC_bbCA_long aromaticC_bbCA_vlong aromaticC_bbC_short
##  Min.   :33.00       Min.   :51.00        Min.   :0.0000
##  1st Qu.:46.00       1st Qu.:66.00        1st Qu.:0.0000
##  Median :50.00       Median :70.00        Median :0.0000
##  Mean   :50.02       Mean   :69.41        Mean   :0.5164
##  3rd Qu.:54.00       3rd Qu.:74.00        3rd Qu.:1.0000
##  Max.   :71.00       Max.   :98.00        Max.   :6.0000
##  aromaticC_bbC_medshort aromaticC_bbC_medlong aromaticC_bbC_long
##  Min.   : 7.00          Min.   :14.00         Min.   :30.00
```

```
##  1st Qu.:11.00          1st Qu.:20.00         1st Qu.:44.00
##  Median :13.00          Median :24.00         Median :48.00
##  Mean   :13.92          Mean   :24.51         Mean   :48.41
##  3rd Qu.:16.00          3rd Qu.:28.00         3rd Qu.:53.00
##  Max.   :28.00          Max.   :42.00         Max.   :72.00
##  aromaticC_bbC_vlong aromaticC_bbO_vshort aromaticC_bbO_short
##  Min.   :48.00       Min.   :0.0000       Min.   : 0.000
##  1st Qu.:68.00       1st Qu.:0.0000       1st Qu.: 2.000
##  Median :72.00       Median :0.0000       Median : 3.000
##  Mean   :72.51       Mean   :0.2045       Mean   : 3.581
##  3rd Qu.:78.00       3rd Qu.:0.0000       3rd Qu.: 5.000
##  Max.   :95.00       Max.   :3.0000       Max.   :16.000
##  aromaticC_bbO_medshort aromaticC_bbO_medlong aromaticC_bbO_long
##  Min.   : 7.00          Min.   :16.0          Min.   :27.00
##  1st Qu.:13.00          1st Qu.:25.0          1st Qu.:42.00
##  Median :15.00          Median :29.0          Median :46.00
##  Mean   :15.12          Mean   :28.6          Mean   :46.44
##  3rd Qu.:17.00          3rd Qu.:32.0          3rd Qu.:50.00
##  Max.   :29.00          Max.   :50.0          Max.   :71.00
##  aromaticC_bbO_vlong scAGN_scAGN_short scAGN_scAGN_medshort scAGN_scAGN_medlong
##  Min.   :47.00       Min.   :0.0000    Min.   :0.0000       Min.   :0.0000
##  1st Qu.:62.00       1st Qu.:0.0000    1st Qu.:0.0000       1st Qu.:0.0000
##  Median :68.00       Median :0.0000    Median :0.0000       Median :0.0000
##  Mean   :68.34       Mean   :0.4111    Mean   :0.5005       Mean   :0.3931
##  3rd Qu.:74.00       3rd Qu.:1.0000    3rd Qu.:1.0000       3rd Qu.:1.0000
##  Max.   :91.00       Max.   :3.0000    Max.   :4.0000       Max.   :3.0000
##  scAGN_scAGN_long scAGN_scAGN_vlong scAGN_scLysN_medshort scAGN_scLysN_medlong
##  Min.   :0.0000   Min.   :0.0000    Min.   :0.0000        Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000    1st Qu.:0.0000        1st Qu.:0.0000
##  Median :0.0000   Median :0.0000    Median :0.0000        Median :0.0000
##  Mean   :0.2641   Mean   :0.2713    Mean   :0.1506        Mean   :0.1033
##  3rd Qu.:0.0000   3rd Qu.:0.0000    3rd Qu.:0.0000        3rd Qu.:0.0000
##  Max.   :3.0000   Max.   :3.0000    Max.   :1.0000        Max.   :1.0000
##  scAGN_scLysN_long scAGN_scLysN_vlong scAGN_bbProN_medshort
##  Min.   :0.0000    Min.   :0.0000     Min.   :0.0000
##  1st Qu.:0.0000    1st Qu.:0.0000     1st Qu.:0.0000
##  Median :0.0000    Median :0.0000     Median :0.0000
##  Mean   :0.2117    Mean   :0.2631     Mean   :0.1079
##  3rd Qu.:0.0000    3rd Qu.:1.0000     3rd Qu.:0.0000
##  Max.   :1.0000    Max.   :2.0000     Max.   :2.0000
##  scAGN_bbProN_medlong scAGN_bbProN_long scAGN_bbProN_vlong
##  Min.   :0.000        Min.   :0.0000    Min.   :0.0000
##  1st Qu.:0.000        1st Qu.:0.0000    1st Qu.:0.0000
##  Median :0.000        Median :0.0000    Median :0.0000
##  Mean   :0.259        Mean   :0.3094    Mean   :0.4625
##  3rd Qu.:0.000        3rd Qu.:1.0000    3rd Qu.:1.0000
##  Max.   :3.000        Max.   :3.0000    Max.   :4.0000
##  scAGN_hydroxylO_vshort scAGN_hydroxylO_short scAGN_hydroxylO_medshort
##  Min.   :0.0000         Min.   :0.000         Min.   :0.0000
##  1st Qu.:0.0000         1st Qu.:1.000         1st Qu.:0.0000
##  Median :0.0000         Median :1.000         Median :1.0000
##  Mean   :0.3715         Mean   :1.147         Mean   :0.8798
##  3rd Qu.:1.0000         3rd Qu.:2.000         3rd Qu.:1.0000
##  Max.   :3.0000         Max.   :4.000         Max.   :5.0000
```

```
##  scAGN_hydroxylO_medlong scAGN_hydroxylO_long scAGN_hydroxylO_vlong
##  Min.   :0.000           Min.   :0.000        Min.   :0.000
##  1st Qu.:1.000           1st Qu.:1.000        1st Qu.:1.000
##  Median :1.000           Median :2.000        Median :1.000
##  Mean   :1.518           Mean   :1.745        Mean   :1.582
##  3rd Qu.:2.000           3rd Qu.:2.000        3rd Qu.:2.000
##  Max.   :6.000           Max.   :7.000        Max.   :6.000
##  scAGN_carbonylO_vshort scAGN_carbonylO_short scAGN_carbonylO_medshort
##  Min.   :0.0000         Min.   :0.0000        Min.   :0.0000
##  1st Qu.:0.0000         1st Qu.:0.0000        1st Qu.:0.0000
##  Median :0.0000         Median :1.0000        Median :0.0000
##  Mean   :0.1259         Mean   :0.5478        Mean   :0.6192
##  3rd Qu.:0.0000         3rd Qu.:1.0000        3rd Qu.:1.0000
##  Max.   :2.0000         Max.   :3.0000        Max.   :5.0000
##  scAGN_carbonylO_medlong scAGN_carbonylO_long scAGN_carbonylO_vlong
##  Min.   :0.00            Min.   :0.0000       Min.   :0.0000
##  1st Qu.:1.00            1st Qu.:0.0000       1st Qu.:0.0000
##  Median :1.00            Median :0.0000       Median :0.0000
##  Mean   :1.39            Mean   :0.5699       Mean   :0.5092
##  3rd Qu.:2.00            3rd Qu.:1.0000       3rd Qu.:1.0000
##  Max.   :5.00            Max.   :6.0000       Max.   :5.0000
##  scAGN_carboxylO_vshort scAGN_carboxylO_short scAGN_carboxylO_medshort
##  Min.   :0.0000         Min.   :0.0000        Min.   :0.0000
##  1st Qu.:0.0000         1st Qu.:0.0000        1st Qu.:0.0000
##  Median :0.0000         Median :0.0000        Median :0.0000
##  Mean   :0.1305         Mean   :0.1285        Mean   :0.2384
##  3rd Qu.:0.0000         3rd Qu.:0.0000        3rd Qu.:0.0000
##  Max.   :2.0000         Max.   :3.0000        Max.   :2.0000
##  scAGN_carboxylO_medlong scAGN_carboxylO_long scAGN_carboxylO_vlong
##  Min.   :0.0000          Min.   :0.0000       Min.   :0.0000
##  1st Qu.:0.0000          1st Qu.:0.0000       1st Qu.:0.0000
##  Median :0.0000          Median :0.0000       Median :0.0000
##  Mean   :0.1511          Mean   :0.3011       Mean   :0.4851
##  3rd Qu.:0.0000          3rd Qu.:0.0000       3rd Qu.:1.0000
##  Max.   :4.0000          Max.   :4.0000       Max.   :4.0000
##  scAGN_sulfur_short scAGN_sulfur_medshort scAGN_sulfur_medlong
##  Min.   :0.0000     Min.   :0.0000        Min.   :0.0000
##  1st Qu.:0.0000     1st Qu.:0.0000        1st Qu.:0.0000
##  Median :0.0000     Median :0.0000        Median :0.0000
##  Mean   :0.2975     Mean   :0.3356        Mean   :0.4681
##  3rd Qu.:1.0000     3rd Qu.:1.0000        3rd Qu.:1.0000
##  Max.   :4.0000     Max.   :3.0000        Max.   :3.0000
##  scAGN_sulfur_long scAGN_sulfur_vlong scAGN_bbN_short  scAGN_bbN_medshort
##  Min.   :0.0000    Min.   :0.0000     Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.0000    1st Qu.:0.0000     1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.0000    Median :1.0000     Median :0.0000   Median :1.000
##  Mean   :0.4974    Mean   :0.6377     Mean   :0.1459   Mean   :1.378
##  3rd Qu.:1.0000    3rd Qu.:1.0000     3rd Qu.:0.0000   3rd Qu.:2.000
##  Max.   :4.0000    Max.   :4.0000     Max.   :3.0000   Max.   :8.000
##  scAGN_bbN_medlong scAGN_bbN_long   scAGN_bbN_vlong  scAGN_bbCA_short
##  Min.   : 0.000    Min.   : 0.000   Min.   : 3.000   Min.   :0.000
##  1st Qu.: 3.000    1st Qu.: 5.000   1st Qu.: 7.000   1st Qu.:0.000
##  Median : 4.000    Median : 7.000   Median : 9.000   Median :0.000
##  Mean   : 4.396    Mean   : 7.242   Mean   : 9.366   Mean   :0.353
```

```
##  3rd Qu.: 6.000    3rd Qu.: 9.000   3rd Qu.:11.000   3rd Qu.:1.000
##  Max.   :12.000    Max.   :18.000   Max.   :20.000   Max.   :4.000
##  scAGN_bbCA_medshort scAGN_bbCA_medlong scAGN_bbCA_long  scAGN_bbCA_vlong
##  Min.   :0.000       Min.   : 0.000    Min.   : 1.000   Min.   : 3
##  1st Qu.:1.000       1st Qu.: 3.000    1st Qu.: 5.000   1st Qu.: 8
##  Median :2.000       Median : 4.000    Median : 6.000   Median :10
##  Mean   :2.201       Mean   : 4.367    Mean   : 6.377   Mean   :10
##  3rd Qu.:3.000       3rd Qu.: 6.000    3rd Qu.: 8.000   3rd Qu.:12
##  Max.   :8.000       Max.   :13.000    Max.   :15.000   Max.   :22
##  scAGN_bbC_short  scAGN_bbC_medshort scAGN_bbC_medlong scAGN_bbC_long
##  Min.   :0.0000   Min.   :0.000      Min.   : 0.000    Min.   : 1.00
##  1st Qu.:0.0000   1st Qu.:1.000      1st Qu.: 2.000    1st Qu.: 6.00
##  Median :0.0000   Median :1.000      Median : 3.000    Median : 8.00
##  Mean   :0.6485   Mean   :1.673      Mean   : 3.644    Mean   : 7.88
##  3rd Qu.:1.0000   3rd Qu.:2.000      3rd Qu.: 5.000    3rd Qu.:10.00
##  Max.   :6.0000   Max.   :8.000      Max.   :12.000    Max.   :17.00
##  scAGN_bbC_vlong scAGN_bbO_vshort scAGN_bbO_short scAGN_bbO_medshort
##  Min.   : 3.00   Min.   :0.0000   Min.   :0.000   Min.   :0.000
##  1st Qu.: 8.00   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.000
##  Median :10.00   Median :0.0000   Median :1.000   Median :1.000
##  Mean   :10.53   Mean   :0.4358   Mean   :1.094   Mean   :1.347
##  3rd Qu.:13.00   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:2.000
##  Max.   :26.00   Max.   :4.0000   Max.   :5.000   Max.   :8.000
##  scAGN_bbO_medlong scAGN_bbO_long  scAGN_bbO_vlong scLysN_hydroxylO_long
##  Min.   : 0.000    Min.   : 0.00   Min.   : 3.00   Min.   :0.0000
##  1st Qu.: 2.000    1st Qu.: 6.00   1st Qu.:10.00   1st Qu.:0.0000
##  Median : 3.000    Median : 8.00   Median :12.00   Median :0.0000
##  Mean   : 3.488    Mean   : 7.62   Mean   :12.34   Mean   :0.1809
##  3rd Qu.: 5.000    3rd Qu.: 9.00   3rd Qu.:14.00   3rd Qu.:0.0000
##  Max.   :12.000    Max.   :17.00   Max.   :25.00   Max.   :3.0000
##  scLysN_hydroxylO_vlong scLysN_carbonylO_vshort scLysN_carbonylO_long
##  Min.   :0.0000         Min.   :0.00000         Min.   :0.0000
##  1st Qu.:0.0000         1st Qu.:0.00000         1st Qu.:0.0000
##  Median :0.0000         Median :0.00000         Median :0.0000
##  Mean   :0.3299         Mean   :0.09918         Mean   :0.1968
##  3rd Qu.:1.0000         3rd Qu.:0.00000         3rd Qu.:0.0000
##  Max.   :3.0000         Max.   :1.00000         Max.   :1.0000
##  scLysN_carbonylO_vlong scLysN_carboxylO_medlong scLysN_carboxylO_long
##  Min.   :0.0000         Min.   :0.0000           Min.   :0.0000
##  1st Qu.:0.0000         1st Qu.:0.0000           1st Qu.:0.0000
##  Median :0.0000         Median :0.0000           Median :0.0000
##  Mean   :0.2431         Mean   :0.1912           Mean   :0.5026
##  3rd Qu.:0.0000         3rd Qu.:0.0000           3rd Qu.:1.0000
##  Max.   :2.0000         Max.   :2.0000           Max.   :2.0000
##  scLysN_carboxylO_vlong scLysN_bbN_medshort scLysN_bbN_medlong scLysN_bbN_long
##  Min.   :0.000          Min.   :0.0000      Min.   :0.0000     Min.   :0.0000
##  1st Qu.:0.000          1st Qu.:0.0000      1st Qu.:0.0000     1st Qu.:0.0000
##  Median :0.000          Median :0.0000      Median :1.0000     Median :1.0000
##  Mean   :0.518          Mean   :0.1686      Mean   :0.8078     Mean   :0.7364
##  3rd Qu.:1.000          3rd Qu.:0.0000      3rd Qu.:1.0000     3rd Qu.:1.0000
##  Max.   :2.000          Max.   :3.0000      Max.   :4.0000     Max.   :4.0000
##  scLysN_bbN_vlong scLysN_bbCA_medshort scLysN_bbCA_medlong scLysN_bbCA_long
##  Min.   :0.00     Min.   :0.0000       Min.   :0.0000      Min.   :0.000
##  1st Qu.:1.00     1st Qu.:0.0000       1st Qu.:0.0000      1st Qu.:1.000
```

```
##  Median :2.00     Median :0.0000       Median :0.0000      Median :2.000
##  Mean   :1.95     Mean   :0.1048       Mean   :0.4902      Mean   :1.797
##  3rd Qu.:2.00     3rd Qu.:0.0000       3rd Qu.:1.0000      3rd Qu.:2.000
##  Max.   :6.00     Max.   :2.0000       Max.   :4.0000      Max.   :5.000
##  scLysN_bbCA_vlong scLysN_bbC_medlong scLysN_bbC_long scLysN_bbC_vlong
##  Min.   :0.0000    Min.   :0.0000     Min.   :0.000   Min.   :0.000
##  1st Qu.:0.0000    1st Qu.:0.0000     1st Qu.:2.000   1st Qu.:1.000
##  Median :1.0000    Median :0.0000     Median :2.000   Median :2.000
##  Mean   :0.8916    Mean   :0.2924     Mean   :2.053   Mean   :1.936
##  3rd Qu.:1.0000    3rd Qu.:0.0000     3rd Qu.:3.000   3rd Qu.:2.000
##  Max.   :5.0000    Max.   :4.0000     Max.   :5.000   Max.   :6.000
##  scLysN_bbO_medlong scLysN_bbO_long scLysN_bbO_vlong scArgN_hydroxylO_short
##  Min.   :0.00       Min.   :0.000   Min.   :0.000    Min.   :0.000
##  1st Qu.:0.00       1st Qu.:1.000   1st Qu.:2.000    1st Qu.:2.000
##  Median :1.00       Median :1.000   Median :2.000    Median :2.000
##  Mean   :0.76       Mean   :1.357   Mean   :2.141    Mean   :1.783
##  3rd Qu.:1.00       3rd Qu.:2.000   3rd Qu.:3.000    3rd Qu.:2.000
##  Max.   :4.00       Max.   :5.000   Max.   :6.000    Max.   :3.000
##  scArgN_hydroxylO_medshort scArgN_hydroxylO_medlong scArgN_hydroxylO_long
##  Min.   :0.000             Min.   :0.000            Min.   :0.000
##  1st Qu.:1.000             1st Qu.:0.000            1st Qu.:2.000
##  Median :1.000             Median :0.000            Median :2.000
##  Mean   :1.053             Mean   :0.334            Mean   :2.109
##  3rd Qu.:1.000             3rd Qu.:0.000            3rd Qu.:2.000
##  Max.   :5.000             Max.   :6.000            Max.   :5.000
##  scArgN_hydroxylO_vlong scArgN_carboxylO_long scArgN_carboxylO_vlong
##  Min.   :1.000          Min.   :0.000         Min.   :0.000
##  1st Qu.:2.000          1st Qu.:0.000         1st Qu.:1.000
##  Median :2.000          Median :0.000         Median :1.000
##  Mean   :2.354          Mean   :0.185         Mean   :1.725
##  3rd Qu.:3.000          3rd Qu.:0.000         3rd Qu.:3.000
##  Max.   :4.000          Max.   :6.000         Max.   :8.000
##  scArgN_bbN_medshort scArgN_bbN_medlong scArgN_bbN_long scArgN_bbN_vlong
##  Min.   :1.000       Min.   :2.000      Min.   :1.000   Min.   :2.000
##  1st Qu.:2.000       1st Qu.:2.000      1st Qu.:2.000   1st Qu.:5.000
##  Median :2.000       Median :3.000      Median :2.000   Median :5.000
##  Mean   :2.059       Mean   :2.749      Mean   :2.105   Mean   :4.675
##  3rd Qu.:2.000       3rd Qu.:3.000      3rd Qu.:2.000   3rd Qu.:5.000
##  Max.   :3.000       Max.   :4.000      Max.   :5.000   Max.   :6.000
##  scArgN_bbCA_medshort scArgN_bbCA_medlong scArgN_bbCA_long scArgN_bbCA_vlong
##  Min.   :1.000        Min.   :2.000       Min.   :0.0000   Min.   :1.000
##  1st Qu.:1.000        1st Qu.:3.000       1st Qu.:1.0000   1st Qu.:2.000
##  Median :1.000        Median :3.000       Median :1.0000   Median :3.000
##  Mean   :1.109        Mean   :2.893       Mean   :0.9538   Mean   :2.687
##  3rd Qu.:1.000        3rd Qu.:3.000       3rd Qu.:1.0000   3rd Qu.:3.000
##  Max.   :2.000        Max.   :4.000       Max.   :4.0000   Max.   :5.000
##  scArgN_bbC_short scArgN_bbC_medshort scArgN_bbC_medlong scArgN_bbC_long
##  Min.   :1.000    Min.   :0.0000      Min.   :0.000      Min.   :1.000
##  1st Qu.:1.000    1st Qu.:1.0000      1st Qu.:2.000      1st Qu.:2.000
##  Median :1.000    Median :1.0000      Median :2.000      Median :2.000
##  Mean   :1.118    Mean   :0.8818      Mean   :1.842      Mean   :2.163
##  3rd Qu.:1.000    3rd Qu.:1.0000      3rd Qu.:2.000      3rd Qu.:2.000
##  Max.   :2.000    Max.   :1.0000      Max.   :3.000      Max.   :4.000
##  scArgN_bbC_vlong scArgN_bbO_vshort scArgN_bbO_short scArgN_bbO_medshort
```

```
##  Min.   :0.000    Min.   :0.000     Min.   :0.0000   Min.   :0.0000
##  1st Qu.:1.000    1st Qu.:1.000     1st Qu.:0.0000   1st Qu.:1.0000
##  Median :2.000    Median :1.000     Median :1.0000   Median :1.0000
##  Mean   :1.798    Mean   :1.379     Mean   :0.6213   Mean   :0.9383
##  3rd Qu.:2.000    3rd Qu.:2.000     3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :6.000    Max.   :2.000     Max.   :2.0000   Max.   :2.0000
##  scArgN_bbO_medlong scArgN_bbO_long scArgN_bbO_vlong bbProN_hydroxylO_medshort
##  Min.   :0.000      Min.   :0.000   Min.   :1.000    Min.   :0.0000
##  1st Qu.:1.000      1st Qu.:1.000   1st Qu.:3.000    1st Qu.:0.0000
##  Median :1.000      Median :1.000   Median :3.000    Median :0.0000
##  Mean   :1.058      Mean   :1.104   Mean   :3.184    Mean   :0.1331
##  3rd Qu.:1.000      3rd Qu.:1.000   3rd Qu.:4.000    3rd Qu.:0.0000
##  Max.   :2.000      Max.   :3.000   Max.   :4.000    Max.   :2.0000
##  bbProN_hydroxylO_medlong bbProN_hydroxylO_long bbProN_hydroxylO_vlong
##  Min.   :0.0000           Min.   :0.0000        Min.   :0.000
##  1st Qu.:0.0000           1st Qu.:0.0000        1st Qu.:1.000
##  Median :0.0000           Median :0.0000        Median :1.000
##  Mean   :0.2605           Mean   :0.4281        Mean   :1.149
##  3rd Qu.:1.0000           3rd Qu.:1.0000        3rd Qu.:2.000
##  Max.   :2.0000           Max.   :2.0000        Max.   :4.000
##  bbProN_carbonylO_medlong bbProN_carbonylO_long bbProN_carbonylO_vlong
##  Min.   :0.0000           Min.   :0.0000        Min.   :0.0000
##  1st Qu.:0.0000           1st Qu.:0.0000        1st Qu.:0.0000
##  Median :0.0000           Median :0.0000        Median :0.0000
##  Mean   :0.2492           Mean   :0.3417        Mean   :0.5159
##  3rd Qu.:0.0000           3rd Qu.:1.0000        3rd Qu.:1.0000
##  Max.   :2.0000           Max.   :3.0000        Max.   :4.0000
##  bbProN_carboxylO_vlong bbProN_sulfur_vlong bbProN_bbN_medshort
##  Min.   :0.0000         Min.   :0.0000      Min.   :0.000
##  1st Qu.:0.0000         1st Qu.:0.0000      1st Qu.:1.000
##  Median :0.0000         Median :0.0000      Median :1.000
##  Mean   :0.1326         Mean   :0.1136      Mean   :1.245
##  3rd Qu.:0.0000         3rd Qu.:0.0000      3rd Qu.:2.000
##  Max.   :3.0000         Max.   :2.0000      Max.   :5.000
##  bbProN_bbN_medlong bbProN_bbN_long bbProN_bbN_vlong bbProN_bbCA_medshort
##  Min.   :0.000      Min.   :0.000   Min.   :0.0      Min.   :0.0000
##  1st Qu.:1.000      1st Qu.:2.000   1st Qu.:2.0      1st Qu.:0.0000
##  Median :2.000      Median :3.000   Median :3.0      Median :1.0000
##  Mean   :2.276      Mean   :2.686   Mean   :3.1      Mean   :0.8083
##  3rd Qu.:3.000      3rd Qu.:3.000   3rd Qu.:4.0      3rd Qu.:1.0000
##  Max.   :7.000      Max.   :9.000   Max.   :9.0      Max.   :3.0000
##  bbProN_bbCA_medlong bbProN_bbCA_long bbProN_bbCA_vlong bbProN_bbC_short
##  Min.   :0.000       Min.   :1.000    Min.   :0.000     Min.   :0.0000
##  1st Qu.:2.000       1st Qu.:2.000    1st Qu.:1.000     1st Qu.:0.0000
##  Median :3.000       Median :3.000    Median :2.000     Median :0.0000
##  Mean   :2.947       Mean   :2.887    Mean   :1.803     Mean   :0.5067
##  3rd Qu.:4.000       3rd Qu.:4.000    3rd Qu.:3.000     3rd Qu.:1.0000
##  Max.   :8.000       Max.   :8.000    Max.   :7.000     Max.   :2.0000
##  bbProN_bbC_medshort bbProN_bbC_medlong bbProN_bbC_long bbProN_bbC_vlong
##  Min.   :0.0000      Min.   :1.000      Min.   :0.000   Min.   :1.000
##  1st Qu.:0.0000      1st Qu.:3.000      1st Qu.:1.000   1st Qu.:2.000
##  Median :1.0000      Median :3.000      Median :2.000   Median :3.000
##  Mean   :0.6012      Mean   :3.285      Mean   :2.069   Mean   :3.038
##  3rd Qu.:1.0000      3rd Qu.:4.000      3rd Qu.:3.000   3rd Qu.:4.000
```

```
##  Max.   :3.0000      Max.   :6.000      Max.   :8.000   Max.   :9.000
##  bbProN_bbO_short bbProN_bbO_medshort bbProN_bbO_medlong bbProN_bbO_long
##  Min.   :0.0000   Min.   :0.00        Min.   :0.000      Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:1.00        1st Qu.:1.000      1st Qu.:3.000
##  Median :0.0000   Median :2.00        Median :1.000      Median :3.000
##  Mean   :0.4404   Mean   :2.06        Mean   :1.248      Mean   :3.502
##  3rd Qu.:1.0000   3rd Qu.:3.00        3rd Qu.:2.000      3rd Qu.:4.000
##  Max.   :2.0000   Max.   :5.00        Max.   :5.000      Max.   :8.000
##  bbProN_bbO_vlong hydroxylO_hydroxylO_short hydroxylO_hydroxylO_medshort
##  Min.   :0.000    Min.   :0.0000            Min.   :0.0000
##  1st Qu.:1.000    1st Qu.:0.0000            1st Qu.:0.0000
##  Median :2.000    Median :0.0000            Median :1.0000
##  Mean   :2.138    Mean   :0.4121            Mean   :0.6932
##  3rd Qu.:3.000    3rd Qu.:1.0000            3rd Qu.:1.0000
##  Max.   :8.000    Max.   :2.0000            Max.   :3.0000
##  hydroxylO_hydroxylO_medlong hydroxylO_hydroxylO_long hydroxylO_hydroxylO_vlong
##  Min.   :0.0000              Min.   :0.000            Min.   :0.000
##  1st Qu.:0.0000              1st Qu.:1.000            1st Qu.:2.000
##  Median :1.0000              Median :1.000            Median :2.000
##  Mean   :0.5874              Mean   :1.175            Mean   :2.461
##  3rd Qu.:1.0000              3rd Qu.:2.000            3rd Qu.:3.000
##  Max.   :4.0000              Max.   :5.000            Max.   :7.000
##  hydroxylO_carbonylO_vshort hydroxylO_carbonylO_short
##  Min.   :0.0000             Min.   :0.0000
##  1st Qu.:0.0000             1st Qu.:0.0000
##  Median :0.0000             Median :1.0000
##  Mean   :0.4728             Mean   :0.8314
##  3rd Qu.:1.0000             3rd Qu.:1.0000
##  Max.   :3.0000             Max.   :4.0000
##  hydroxylO_carbonylO_medshort hydroxylO_carbonylO_medlong
##  Min.   :0.000                Min.   :0.000
##  1st Qu.:0.000                1st Qu.:0.000
##  Median :1.000                Median :1.000
##  Mean   :1.203                Mean   :1.631
##  3rd Qu.:2.000                3rd Qu.:3.000
##  Max.   :5.000                Max.   :9.000
##  hydroxylO_carbonylO_long hydroxylO_carbonylO_vlong hydroxylO_carboxylO_vshort
##  Min.   :0.000            Min.   :0.000             Min.   :0.0000
##  1st Qu.:1.000            1st Qu.:1.000             1st Qu.:1.0000
##  Median :2.000            Median :2.000             Median :1.0000
##  Mean   :1.795            Mean   :1.709             Mean   :0.9928
##  3rd Qu.:3.000            3rd Qu.:2.000             3rd Qu.:1.0000
##  Max.   :6.000            Max.   :8.000             Max.   :3.0000
##  hydroxylO_carboxylO_short hydroxylO_carboxylO_medshort
##  Min.   :0.000             Min.   :0.000
##  1st Qu.:1.000             1st Qu.:1.000
##  Median :1.000             Median :1.000
##  Mean   :1.092             Mean   :1.331
##  3rd Qu.:1.000             3rd Qu.:2.000
##  Max.   :4.000             Max.   :5.000
##  hydroxylO_carboxylO_medlong hydroxylO_carboxylO_long hydroxylO_carboxylO_vlong
##  Min.   :0.0000              Min.   :0.0000           Min.   :0.0000
##  1st Qu.:0.0000              1st Qu.:0.0000           1st Qu.:0.0000
##  Median :0.0000              Median :1.0000           Median :1.0000
```

```
##  Mean   :0.2811              Mean   :0.7251           Mean   :0.8628
##  3rd Qu.:0.0000              3rd Qu.:1.0000           3rd Qu.:1.0000
##  Max.   :3.0000              Max.   :5.0000           Max.   :6.0000
##  hydroxylO_sulfur_short hydroxylO_sulfur_medshort hydroxylO_sulfur_medlong
##  Min.   :0.0000         Min.   :0.000             Min.   :0.0000
##  1st Qu.:0.0000         1st Qu.:0.000             1st Qu.:0.0000
##  Median :0.0000         Median :0.000             Median :0.0000
##  Mean   :0.1418         Mean   :0.221             Mean   :0.3052
##  3rd Qu.:0.0000         3rd Qu.:0.000             3rd Qu.:1.0000
##  Max.   :2.0000         Max.   :3.000             Max.   :3.0000
##  hydroxylO_sulfur_long hydroxylO_sulfur_vlong hydroxylO_bbN_short
##  Min.   :0.0000        Min.   :0.0000         Min.   :0.0000
##  1st Qu.:0.0000        1st Qu.:0.0000         1st Qu.:0.0000
##  Median :0.0000        Median :1.0000         Median :1.0000
##  Mean   :0.4604        Mean   :0.7765         Mean   :0.8052
##  3rd Qu.:1.0000        3rd Qu.:1.0000         3rd Qu.:1.0000
##  Max.   :4.0000        Max.   :4.0000         Max.   :5.0000
##  hydroxylO_bbN_medshort hydroxylO_bbN_medlong hydroxylO_bbN_long
##  Min.   : 2.000         Min.   : 0.000        Min.   : 3.000
##  1st Qu.: 4.000         1st Qu.: 3.000        1st Qu.: 8.000
##  Median : 5.000         Median : 4.000        Median : 9.000
##  Mean   : 4.786         Mean   : 4.154        Mean   : 9.603
##  3rd Qu.: 6.000         3rd Qu.: 5.000        3rd Qu.:11.000
##  Max.   :11.000         Max.   :17.000        Max.   :19.000
##  hydroxylO_bbN_vlong hydroxylO_bbCA_short hydroxylO_bbCA_medshort
##  Min.   : 4.00       Min.   :0.000        Min.   : 1.000
##  1st Qu.:10.00       1st Qu.:1.000        1st Qu.: 3.000
##  Median :12.00       Median :2.000        Median : 4.000
##  Mean   :12.58       Mean   :1.693        Mean   : 3.788
##  3rd Qu.:15.00       3rd Qu.:2.000        3rd Qu.: 5.000
##  Max.   :25.00       Max.   :6.000        Max.   :10.000
##  hydroxylO_bbCA_medlong hydroxylO_bbCA_long hydroxylO_bbCA_vlong
##  Min.   : 1.000         Min.   : 0.00       Min.   : 6.00
##  1st Qu.: 5.000         1st Qu.: 6.00       1st Qu.:13.00
##  Median : 6.000         Median : 8.00       Median :16.00
##  Mean   : 6.311         Mean   : 7.51       Mean   :15.54
##  3rd Qu.: 8.000         3rd Qu.: 9.00       3rd Qu.:18.00
##  Max.   :15.000         Max.   :17.00       Max.   :25.00
##  hydroxylO_bbC_short hydroxylO_bbC_medshort hydroxylO_bbC_medlong
##  Min.   :0.000       Min.   : 1.000         Min.   : 1.000
##  1st Qu.:0.000       1st Qu.: 3.000         1st Qu.: 4.250
##  Median :1.000       Median : 4.000         Median : 6.000
##  Mean   :1.084       Mean   : 4.447         Mean   : 6.063
##  3rd Qu.:2.000       3rd Qu.: 5.000         3rd Qu.: 7.000
##  Max.   :6.000       Max.   :12.000         Max.   :13.000
##  hydroxylO_bbC_long hydroxylO_bbC_vlong hydroxylO_bbO_vshort
##  Min.   : 3.000     Min.   : 3.00       Min.   :0.0000
##  1st Qu.: 7.000     1st Qu.:10.00       1st Qu.:0.0000
##  Median : 9.000     Median :12.00       Median :1.0000
##  Mean   : 9.094     Mean   :11.98       Mean   :0.6346
##  3rd Qu.:11.000     3rd Qu.:14.00       3rd Qu.:1.0000
##  Max.   :19.000     Max.   :24.00       Max.   :5.0000
##  hydroxylO_bbO_short hydroxylO_bbO_medshort hydroxylO_bbO_medlong
##  Min.   :0.000       Min.   :0.000          Min.   : 2.000
```

```
##  1st Qu.:1.000       1st Qu.:1.000          1st Qu.: 6.000
##  Median :1.000       Median :2.000          Median : 7.000
##  Mean   :1.412       Mean   :2.521          Mean   : 6.994
##  3rd Qu.:2.000       3rd Qu.:3.000          3rd Qu.: 8.000
##  Max.   :6.000       Max.   :9.000          Max.   :15.000
##  hydroxylO_bbO_long hydroxylO_bbO_vlong carbonylO_carbonylO_medshort
##  Min.   : 3.00      Min.   : 6.0        Min.   :0.0000
##  1st Qu.: 8.00      1st Qu.:12.0        1st Qu.:0.0000
##  Median :10.00      Median :14.0        Median :0.0000
##  Mean   :10.24      Mean   :13.9        Mean   :0.4861
##  3rd Qu.:12.00      3rd Qu.:16.0        3rd Qu.:1.0000
##  Max.   :21.00      Max.   :23.0        Max.   :3.0000
##  carbonylO_carbonylO_medlong carbonylO_carbonylO_long carbonylO_carbonylO_vlong
##  Min.   :0.0000              Min.   :0.0000           Min.   :0.0000
##  1st Qu.:0.0000              1st Qu.:0.0000           1st Qu.:0.0000
##  Median :1.0000              Median :0.0000           Median :0.0000
##  Mean   :0.6742              Mean   :0.3546           Mean   :0.2395
##  3rd Qu.:1.0000              3rd Qu.:1.0000           3rd Qu.:0.0000
##  Max.   :4.0000              Max.   :3.0000           Max.   :3.0000
##  carbonylO_carboxylO_medshort carbonylO_carboxylO_medlong
##  Min.   :0.000                Min.   :0.0000
##  1st Qu.:0.000                1st Qu.:0.0000
##  Median :0.000                Median :0.0000
##  Mean   :0.258                Mean   :0.2395
##  3rd Qu.:0.000                3rd Qu.:0.0000
##  Max.   :4.000                Max.   :3.0000
##  carbonylO_carboxylO_long carbonylO_carboxylO_vlong carbonylO_sulfur_short
##  Min.   :0.0000           Min.   :0.0000            Min.   :0.0000
##  1st Qu.:0.0000           1st Qu.:0.0000            1st Qu.:0.0000
##  Median :0.0000           Median :0.0000            Median :0.0000
##  Mean   :0.2857           Mean   :0.3798            Mean   :0.2292
##  3rd Qu.:0.0000           3rd Qu.:1.0000            3rd Qu.:0.0000
##  Max.   :4.0000           Max.   :3.0000            Max.   :3.0000
##  carbonylO_sulfur_medshort carbonylO_sulfur_medlong carbonylO_sulfur_long
##  Min.   :0.0000            Min.   :0.0000           Min.   :0.0000
##  1st Qu.:0.0000            1st Qu.:0.0000           1st Qu.:0.0000
##  Median :0.0000            Median :0.0000           Median :0.0000
##  Mean   :0.3201            Mean   :0.4193           Mean   :0.4563
##  3rd Qu.:1.0000            3rd Qu.:1.0000           3rd Qu.:1.0000
##  Max.   :4.0000            Max.   :3.0000           Max.   :4.0000
##  carbonylO_sulfur_vlong carbonylO_bbN_short carbonylO_bbN_medshort
##  Min.   :0.0000         Min.   :0.0000      Min.   :0.0000
##  1st Qu.:0.0000         1st Qu.:0.0000      1st Qu.:0.0000
##  Median :1.0000         Median :0.0000      Median :1.0000
##  Mean   :0.6973         Mean   :0.1984      Mean   :0.8597
##  3rd Qu.:1.0000         3rd Qu.:0.0000      3rd Qu.:1.0000
##  Max.   :4.0000         Max.   :4.0000      Max.   :6.0000
##  carbonylO_bbN_medlong carbonylO_bbN_long carbonylO_bbN_vlong
##  Min.   : 0.000        Min.   : 0.000     Min.   : 1.00
##  1st Qu.: 3.000        1st Qu.: 5.000     1st Qu.: 7.00
##  Median : 4.000        Median : 7.000     Median : 9.00
##  Mean   : 4.145        Mean   : 6.861     Mean   : 8.92
##  3rd Qu.: 5.000        3rd Qu.: 9.000     3rd Qu.:11.00
##  Max.   :11.000        Max.   :19.000     Max.   :20.00
```

```
##  carbonylO_bbCA_short carbonylO_bbCA_medshort carbonylO_bbCA_medlong
##  Min.   :0.0000       Min.   :0.000           Min.   : 0.000
##  1st Qu.:0.0000       1st Qu.:1.000           1st Qu.: 2.000
##  Median :0.0000       Median :2.000           Median : 4.000
##  Mean   :0.4949       Mean   :1.736           Mean   : 3.626
##  3rd Qu.:1.0000       3rd Qu.:2.000           3rd Qu.: 5.000
##  Max.   :4.0000       Max.   :7.000           Max.   :12.000
##  carbonylO_bbCA_long carbonylO_bbCA_vlong carbonylO_bbC_short
##  Min.   : 0.000      Min.   : 3.00        Min.   :0.0000
##  1st Qu.: 5.000      1st Qu.: 8.00        1st Qu.:0.0000
##  Median : 6.000      Median :10.00        Median :0.0000
##  Mean   : 6.342      Mean   :10.33        Mean   :0.3006
##  3rd Qu.: 8.000      3rd Qu.:12.00        3rd Qu.:0.0000
##  Max.   :17.000      Max.   :22.00        Max.   :5.0000
##  carbonylO_bbC_medshort carbonylO_bbC_medlong carbonylO_bbC_long
##  Min.   :0.000          Min.   : 0.000        Min.   : 0.000
##  1st Qu.:1.000          1st Qu.: 2.000        1st Qu.: 5.000
##  Median :2.000          Median : 4.000        Median : 6.000
##  Mean   :1.786          Mean   : 3.754        Mean   : 6.247
##  3rd Qu.:2.000          3rd Qu.: 5.000        3rd Qu.: 8.000
##  Max.   :8.000          Max.   :12.000        Max.   :15.000
##  carbonylO_bbC_vlong carbonylO_bbO_short carbonylO_bbO_medshort
##  Min.   : 1.00       Min.   :0.0000      Min.   : 0.00
##  1st Qu.: 8.00       1st Qu.:0.0000      1st Qu.: 1.00
##  Median :10.00       Median :1.0000      Median : 2.00
##  Mean   :10.33       Mean   :0.9486      Mean   : 2.31
##  3rd Qu.:12.00       3rd Qu.:1.0000      3rd Qu.: 3.00
##  Max.   :23.00       Max.   :5.0000      Max.   :10.00
##  carbonylO_bbO_medlong carbonylO_bbO_long carbonylO_bbO_vlong
##  Min.   : 0.000        Min.   : 0.000     Min.   : 3.0
##  1st Qu.: 2.000        1st Qu.: 4.000     1st Qu.:10.0
##  Median : 3.000        Median : 6.000     Median :12.0
##  Mean   : 3.108        Mean   : 6.038     Mean   :12.2
##  3rd Qu.: 4.000        3rd Qu.: 8.000     3rd Qu.:14.0
##  Max.   :11.000        Max.   :15.000     Max.   :25.0
##  carboxylO_carboxylO_medlong carboxylO_carboxylO_long carboxylO_carboxylO_vlong
##  Min.   :0.00000             Min.   :0.00             Min.   :0.0000
##  1st Qu.:0.00000             1st Qu.:1.00             1st Qu.:0.0000
##  Median :0.00000             Median :2.00             Median :1.0000
##  Mean   :0.09301             Mean   :1.62             Mean   :0.7831
##  3rd Qu.:0.00000             3rd Qu.:2.00             3rd Qu.:1.0000
##  Max.   :4.00000             Max.   :4.00             Max.   :4.0000
##  carboxylO_sulfur_medshort carboxylO_sulfur_medlong carboxylO_sulfur_long
##  Min.   :0.0000            Min.   :0.0000           Min.   :0.0000
##  1st Qu.:0.0000            1st Qu.:0.0000           1st Qu.:0.0000
##  Median :0.0000            Median :0.0000           Median :0.0000
##  Mean   :0.1542            Mean   :0.1968           Mean   :0.2091
##  3rd Qu.:0.0000            3rd Qu.:0.0000           3rd Qu.:0.0000
##  Max.   :3.0000            Max.   :3.0000           Max.   :3.0000
##  carboxylO_sulfur_vlong carboxylO_bbN_vshort carboxylO_bbN_medshort
##  Min.   :0.000          Min.   :1.000        Min.   :1.000
##  1st Qu.:0.000          1st Qu.:1.000        1st Qu.:1.000
##  Median :0.000          Median :1.000        Median :1.000
##  Mean   :0.426          Mean   :1.043        Mean   :1.275
```

```
##  3rd Qu.:1.000          3rd Qu.:1.000        3rd Qu.:1.000
##  Max.   :4.000          Max.   :4.000        Max.   :6.000
##  carboxylO_bbN_medlong carboxylO_bbN_long carboxylO_bbN_vlong
##  Min.   :1.000         Min.   : 0.000     Min.   : 2.000
##  1st Qu.:1.000         1st Qu.: 1.000     1st Qu.: 4.000
##  Median :2.000         Median : 2.000     Median : 6.000
##  Mean   :1.959         Mean   : 2.362     Mean   : 5.894
##  3rd Qu.:3.000         3rd Qu.: 4.000     3rd Qu.: 7.000
##  Max.   :8.000         Max.   :11.000     Max.   :14.000
##  carboxylO_bbCA_short carboxylO_bbCA_medshort carboxylO_bbCA_medlong
##  Min.   :1.000        Min.   :0.0000          Min.   :1.000
##  1st Qu.:1.000        1st Qu.:0.0000          1st Qu.:2.000
##  Median :1.000        Median :0.0000          Median :2.000
##  Mean   :1.406        Mean   :0.5606          Mean   :2.272
##  3rd Qu.:2.000        3rd Qu.:1.0000          3rd Qu.:3.000
##  Max.   :5.000        Max.   :6.0000          Max.   :9.000
##  carboxylO_bbCA_long carboxylO_bbCA_vlong carboxylO_bbC_short
##  Min.   : 0.000      Min.   : 0.000       Min.   :0.0000
##  1st Qu.: 1.000      1st Qu.: 2.000       1st Qu.:0.0000
##  Median : 2.000      Median : 4.000       Median :0.0000
##  Mean   : 2.689      Mean   : 3.891       Mean   :0.3422
##  3rd Qu.: 4.000      3rd Qu.: 5.000       3rd Qu.:1.0000
##  Max.   :10.000      Max.   :13.000       Max.   :4.0000
##  carboxylO_bbC_medshort carboxylO_bbC_medlong carboxylO_bbC_long
##  Min.   :1.000          Min.   : 0.00         Min.   :0.000
##  1st Qu.:1.000          1st Qu.: 1.00         1st Qu.:1.000
##  Median :1.000          Median : 2.00         Median :2.000
##  Mean   :1.525          Mean   : 1.88         Mean   :2.541
##  3rd Qu.:2.000          3rd Qu.: 2.00         3rd Qu.:3.000
##  Max.   :7.000          Max.   :10.00         Max.   :9.000
##  carboxylO_bbC_vlong carboxylO_bbO_medshort carboxylO_bbO_medlong
##  Min.   : 1.000      Min.   :1.000          Min.   :0.000
##  1st Qu.: 3.000      1st Qu.:1.000          1st Qu.:1.000
##  Median : 4.000      Median :2.000          Median :2.000
##  Mean   : 4.342      Mean   :1.761          Mean   :2.035
##  3rd Qu.: 6.000      3rd Qu.:2.000          3rd Qu.:3.000
##  Max.   :15.000      Max.   :7.000          Max.   :8.000
##  carboxylO_bbO_long carboxylO_bbO_vlong sulfur_sulfur_vlong sulfur_bbN_short
##  Min.   : 0.00      Min.   : 0.000      Min.   :0.0000      Min.   :0.0000
##  1st Qu.: 1.00      1st Qu.: 2.000      1st Qu.:0.0000      1st Qu.:0.0000
##  Median : 2.00      Median : 4.000      Median :0.0000      Median :0.0000
##  Mean   : 2.68      Mean   : 3.957      Mean   :0.1228      Mean   :0.3422
##  3rd Qu.: 4.00      3rd Qu.: 5.000      3rd Qu.:0.0000      3rd Qu.:1.0000
##  Max.   :10.00      Max.   :16.000      Max.   :2.0000      Max.   :4.0000
##  sulfur_bbN_medshort sulfur_bbN_medlong sulfur_bbN_long sulfur_bbN_vlong
##  Min.   :0.000       Min.   : 0.000     Min.   : 1.00   Min.   : 1.000
##  1st Qu.:1.000       1st Qu.: 3.000     1st Qu.: 4.00   1st Qu.: 4.000
##  Median :1.000       Median : 5.000     Median : 5.00   Median : 5.000
##  Mean   :1.637       Mean   : 4.644     Mean   : 5.33   Mean   : 5.574
##  3rd Qu.:2.000       3rd Qu.: 6.000     3rd Qu.: 7.00   3rd Qu.: 7.000
##  Max.   :8.000       Max.   :12.000     Max.   :13.00   Max.   :13.000
##  sulfur_bbCA_short sulfur_bbCA_medshort sulfur_bbCA_medlong sulfur_bbCA_long
##  Min.   :0.0000    Min.   :0.000        Min.   : 0.000      Min.   : 1.000
##  1st Qu.:0.0000    1st Qu.:1.000        1st Qu.: 2.000      1st Qu.: 5.000
```

```
##  Median :0.0000    Median :2.000        Median : 3.000      Median : 6.000
##  Mean   :0.4866    Mean   :2.179        Mean   : 3.351      Mean   : 6.187
##  3rd Qu.:1.0000    3rd Qu.:3.000        3rd Qu.: 4.000      3rd Qu.: 7.000
##  Max.   :4.0000    Max.   :7.000        Max.   :13.000      Max.   :14.000
##  sulfur_bbCA_vlong sulfur_bbC_short sulfur_bbC_medshort sulfur_bbC_medlong
##  Min.   : 0.000    Min.   :0.0000   Min.   :0.00        Min.   : 0.000
##  1st Qu.: 3.000    1st Qu.:0.0000   1st Qu.:2.00        1st Qu.: 3.000
##  Median : 4.000    Median :0.0000   Median :3.00        Median : 4.000
##  Mean   : 4.461    Mean   :0.2718   Mean   :3.04        Mean   : 3.737
##  3rd Qu.: 6.000    3rd Qu.:0.0000   3rd Qu.:4.00        3rd Qu.: 5.000
##  Max.   :13.000    Max.   :3.0000   Max.   :9.00        Max.   :11.000
##  sulfur_bbC_long  sulfur_bbC_vlong sulfur_bbO_short sulfur_bbO_medshort
##  Min.   : 1.000   Min.   : 1.000   Min.   :0.000    Min.   :0.000
##  1st Qu.: 4.000   1st Qu.: 4.000   1st Qu.:1.000    1st Qu.:1.000
##  Median : 5.000   Median : 5.000   Median :2.000    Median :2.000
##  Mean   : 5.451   Mean   : 5.387   Mean   :1.835    Mean   :2.054
##  3rd Qu.: 7.000   3rd Qu.: 7.000   3rd Qu.:2.000    3rd Qu.:3.000
##  Max.   :14.000   Max.   :15.000   Max.   :6.000    Max.   :8.000
##  sulfur_bbO_medlong sulfur_bbO_long  sulfur_bbO_vlong bbN_bbN_medshort
##  Min.   :0.000      Min.   : 0.000   Min.   : 0.000   Min.   : 2.000
##  1st Qu.:1.000      1st Qu.: 4.000   1st Qu.: 4.000   1st Qu.: 6.000
##  Median :2.000      Median : 5.000   Median : 5.000   Median : 7.000
##  Mean   :2.383      Mean   : 5.232   Mean   : 5.295   Mean   : 6.855
##  3rd Qu.:3.000      3rd Qu.: 6.000   3rd Qu.: 7.000   3rd Qu.: 8.000
##  Max.   :9.000      Max.   :12.000   Max.   :16.000   Max.   :14.000
##  bbN_bbN_medlong  bbN_bbN_long   bbN_bbN_vlong   bbN_bbCA_medshort
##  Min.   : 4.00   Min.   :11.00   Min.   :11.00   Min.   : 6.00
##  1st Qu.:10.00   1st Qu.:16.00   1st Qu.:18.00   1st Qu.:13.00
##  Median :12.00   Median :18.00   Median :20.00   Median :15.00
##  Mean   :11.89   Mean   :18.28   Mean   :20.67   Mean   :15.08
##  3rd Qu.:14.00   3rd Qu.:20.00   3rd Qu.:23.00   3rd Qu.:17.00
##  Max.   :23.00   Max.   :30.00   Max.   :36.00   Max.   :26.00
##  bbN_bbCA_medlong bbN_bbCA_long   bbN_bbCA_vlong  bbN_bbC_vshort
##  Min.   :18.00    Min.   :15.00   Min.   :39.00   Min.   :0.0000
##  1st Qu.:27.00    1st Qu.:25.00   1st Qu.:53.00   1st Qu.:0.0000
##  Median :29.00    Median :28.00   Median :58.00   Median :0.0000
##  Mean   :29.82    Mean   :29.07   Mean   :57.87   Mean   :0.3279
##  3rd Qu.:32.00    3rd Qu.:33.00   3rd Qu.:62.00   3rd Qu.:1.0000
##  Max.   :51.00    Max.   :54.00   Max.   :84.00   Max.   :4.0000
##  bbN_bbC_short   bbN_bbC_medshort bbN_bbC_medlong  bbN_bbC_long
##  Min.   : 9.00   Min.   : 2.00    Min.   : 4.00   Min.   :22.0
##  1st Qu.:14.00   1st Qu.: 9.00    1st Qu.:13.00   1st Qu.:34.0
##  Median :16.00   Median :11.00    Median :16.00   Median :37.0
##  Mean   :15.59   Mean   :10.89    Mean   :16.34   Mean   :37.5
##  3rd Qu.:17.00   3rd Qu.:12.00    3rd Qu.:19.00   3rd Qu.:41.0
##  Max.   :23.00   Max.   :20.00    Max.   :36.00   Max.   :58.0
##  bbN_bbC_vlong   bbN_bbO_vshort  bbN_bbO_short   bbN_bbO_medshort
##  Min.   :31.00   Min.   :0.000   Min.   : 9.00   Min.   : 3.00
##  1st Qu.:42.00   1st Qu.:1.000   1st Qu.:16.00   1st Qu.: 9.00
##  Median :47.00   Median :2.000   Median :18.00   Median :11.00
##  Mean   :47.37   Mean   :2.162   Mean   :17.57   Mean   :10.65
##  3rd Qu.:52.00   3rd Qu.:3.000   3rd Qu.:19.00   3rd Qu.:13.00
##  Max.   :79.00   Max.   :7.000   Max.   :26.00   Max.   :24.00
##  bbN_bbO_medlong  bbN_bbO_long   bbN_bbO_vlong   bbCA_bbCA_medshort
```

```
##  Min.   : 4.00   Min.   :21.00   Min.   :25.00   Min.   :0.000
##  1st Qu.:13.00   1st Qu.:34.00   1st Qu.:37.00   1st Qu.:1.000
##  Median :16.00   Median :37.00   Median :41.00   Median :2.000
##  Mean   :15.86   Mean   :37.12   Mean   :41.95   Mean   :2.258
##  3rd Qu.:18.00   3rd Qu.:40.00   3rd Qu.:47.00   3rd Qu.:3.000
##  Max.   :30.00   Max.   :62.00   Max.   :68.00   Max.   :7.000
##  bbCA_bbCA_medlong bbCA_bbCA_long bbCA_bbCA_vlong bbCA_bbC_medshort
##  Min.   :11.00     Min.   :15.0   Min.   :10.00   Min.   :10.00
##  1st Qu.:17.00     1st Qu.:21.0   1st Qu.:18.00   1st Qu.:17.00
##  Median :19.00     Median :24.0   Median :20.00   Median :18.00
##  Mean   :19.56     Mean   :24.2   Mean   :20.35   Mean   :18.47
##  3rd Qu.:22.00     3rd Qu.:27.0   3rd Qu.:23.00   3rd Qu.:20.00
##  Max.   :32.00     Max.   :40.0   Max.   :34.00   Max.   :30.00
##  bbCA_bbC_medlong bbCA_bbC_long   bbCA_bbC_vlong  bbCA_bbO_vshort
##  Min.   :14.00    Min.   :20.00   Min.   :42.00   Min.   :0.0000
##  1st Qu.:25.00    1st Qu.:31.00   1st Qu.:55.00   1st Qu.:0.0000
##  Median :28.00    Median :34.50   Median :58.00   Median :0.0000
##  Mean   :27.92    Mean   :34.81   Mean   :59.17   Mean   :0.1213
##  3rd Qu.:31.00    3rd Qu.:39.00   3rd Qu.:63.00   3rd Qu.:0.0000
##  Max.   :49.00    Max.   :54.00   Max.   :83.00   Max.   :2.0000
##  bbCA_bbO_short   bbCA_bbO_medshort bbCA_bbO_medlong bbCA_bbO_long
##  Min.   : 0.000   Min.   :10.00     Min.   : 9.00    Min.   :19.00
##  1st Qu.: 4.000   1st Qu.:17.00     1st Qu.:16.25    1st Qu.:33.00
##  Median : 5.000   Median :19.00     Median :20.00    Median :36.00
##  Mean   : 5.585   Mean   :19.31     Mean   :20.01    Mean   :36.91
##  3rd Qu.: 7.000   3rd Qu.:21.00     3rd Qu.:23.00    3rd Qu.:40.00
##  Max.   :12.000   Max.   :30.00     Max.   :39.00    Max.   :61.00
##  bbCA_bbO_vlong bbC_bbC_medshort bbC_bbC_medlong  bbC_bbC_long
##  Min.   :35.0   Min.   : 0.000   Min.   :10.00   Min.   : 9.00
##  1st Qu.:52.0   1st Qu.: 4.000   1st Qu.:17.00   1st Qu.:17.00
##  Median :57.0   Median : 6.000   Median :18.00   Median :19.00
##  Mean   :57.3   Mean   : 5.758   Mean   :18.46   Mean   :19.18
##  3rd Qu.:62.0   3rd Qu.: 7.000   3rd Qu.:20.00   3rd Qu.:21.00
##  Max.   :82.0   Max.   :14.000   Max.   :30.00   Max.   :33.00
##  bbC_bbC_vlong   bbC_bbO_short    bbC_bbO_medshort bbC_bbO_medlong
##  Min.   :10.00   Min.   : 0.000   Min.   : 3.00    Min.   :15.00
##  1st Qu.:18.00   1st Qu.: 3.000   1st Qu.:10.00    1st Qu.:26.00
##  Median :21.00   Median : 4.000   Median :12.00    Median :29.00
##  Mean   :20.81   Mean   : 3.743   Mean   :12.42    Mean   :29.28
##  3rd Qu.:23.00   3rd Qu.: 5.000   3rd Qu.:15.00    3rd Qu.:32.00
##  Max.   :37.00   Max.   :12.000   Max.   :24.00    Max.   :46.00
##   bbC_bbO_long   bbC_bbO_vlong  bbO_bbO_vshort   bbO_bbO_short
##  Min.   :23.00   Min.   :27.0   Min.   :0.0000   Min.   :0.000
##  1st Qu.:35.00   1st Qu.:41.0   1st Qu.:0.0000   1st Qu.:2.000
##  Median :38.00   Median :46.0   Median :0.0000   Median :3.000
##  Mean   :38.61   Mean   :46.6   Mean   :0.1161   Mean   :2.715
##  3rd Qu.:42.00   3rd Qu.:51.0   3rd Qu.:0.0000   3rd Qu.:4.000
##  Max.   :63.00   Max.   :76.0   Max.   :3.0000   Max.   :9.000
##  bbO_bbO_medshort bbO_bbO_medlong  bbO_bbO_long   bbO_bbO_vlong
##  Min.   : 1.000   Min.   : 3.00   Min.   :11.00   Min.   :15.00
##  1st Qu.: 5.000   1st Qu.:10.00   1st Qu.:19.00   1st Qu.:22.00
##  Median : 6.000   Median :12.00   Median :21.00   Median :25.00
##  Mean   : 6.168   Mean   :12.17   Mean   :21.71   Mean   :24.82
##  3rd Qu.: 7.000   3rd Qu.:14.75   3rd Qu.:24.00   3rd Qu.:27.00
```

```
##  Max.   :14.000   Max.   :24.00   Max.   :36.00   Max.   :41.00
```

```
# Scatter plots
par ( mfrow = c(3 ,3))
plot ( protein_train$angles , protein_train$accuracy , ylab ="Accuracy", xlab ="Angles")
plot ( protein_train$carbonylC_aliph2HC_short , protein_train$accuracy , ylab ="Accuracy", xlab =" carbo
plot ( protein_train$carbonylC_bbCA_medlong , protein_train$accuracy , ylab ="Accuracy", xlab =" carbony
plot ( protein_train$scAGN_bbC_medshort , protein_train$accuracy , ylab ="Accuracy", xlab =" scAGN_bbC_m
plot ( protein_train$scArgN_carboxylO_long , protein_train$accuracy , ylab ="Accuracy", xlab =" scArgN_c
plot ( protein_train$bbC_bbO_vlong , protein_train$accuracy , ylab ="Accuracy", xlab =" bbC_bbO_vlong")
plot ( protein_train$aliph2HC_aliph3HC_vshort , protein_train$accuracy , ylab ="Accuracy", xlab =" aliph



mfull <- lm(accuracy~., data = protein_train)
#mfull$coefficients

summary(mfull)
```

```
##
## Call:
## lm(formula = accuracy ~ ., data = protein_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.65667 -0.21862 -0.00315  0.22672  1.28079
##
## Coefficients: (2 not defined because of singularities)
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -4.127e+00  3.101e+00  -1.331 0.183545
## angles                       -7.486e-03  5.394e-03  -1.388 0.165448
## carbonylC_carbonylC_medshort -3.164e-02  8.218e-02  -0.385 0.700271
## carbonylC_carbonylC_medlong  -6.772e-02  8.068e-02  -0.839 0.401409
## carbonylC_carbonylC_long     -9.423e-02  6.551e-02  -1.438 0.150583
## carbonylC_carbonylC_vlong    -8.465e-02  4.073e-02  -2.078 0.037892 *
## carbonylC_carboxylC_medshort -8.688e-02  5.896e-02  -1.474 0.140842
## carbonylC_carboxylC_long     -4.307e-02  5.655e-02  -0.762 0.446417
## carbonylC_carboxylC_vlong    -4.575e-02  4.270e-02  -1.071 0.284210
## carbonylC_aliph1HC_medshort  -9.763e-02  4.808e-02  -2.031 0.042482 *
## carbonylC_aliph1HC_medlong   -3.142e-02  4.183e-02  -0.751 0.452745
## carbonylC_aliph1HC_long      -2.419e-02  3.436e-02  -0.704 0.481607
## carbonylC_aliph1HC_vlong     -4.229e-03  2.297e-02  -0.184 0.853969
## carbonylC_aliph2HC_short     -8.577e-02  3.910e-02  -2.193 0.028456 *
## carbonylC_aliph2HC_medshort  -9.476e-02  2.855e-02  -3.320 0.000927 ***
## carbonylC_aliph2HC_medlong   -4.127e-02  2.194e-02  -1.881 0.060235 .
## carbonylC_aliph2HC_long      -3.177e-02  1.568e-02  -2.025 0.043027 *
## carbonylC_aliph2HC_vlong     -1.291e-02  9.204e-03  -1.403 0.160856
## carbonylC_aliph3HC_short     -1.465e-02  4.345e-02  -0.337 0.736059
## carbonylC_aliph3HC_medshort   4.959e-02  3.863e-02   1.284 0.199477
## carbonylC_aliph3HC_medlong    6.276e-02  3.415e-02   1.838 0.066303 .
## carbonylC_aliph3HC_long       3.587e-02  2.687e-02   1.335 0.182239
## carbonylC_aliph3HC_vlong      1.809e-02  1.656e-02   1.092 0.274981
## carbonylC_aromaticC_short     7.909e-02  4.509e-02   1.754 0.079670 .
## carbonylC_aromaticC_medshort -1.434e-02  2.553e-02  -0.562 0.574495
## carbonylC_aromaticC_medlong  -7.470e-03  2.086e-02  -0.358 0.720372
```

```
## carbonylC_aromaticC_long     -1.591e-02  1.428e-02  -1.114 0.265418
## carbonylC_aromaticC_vlong    -8.890e-03  1.069e-02  -0.832 0.405745
## carbonylC_scAGN_short        -9.469e-02  6.793e-02  -1.394 0.163576
## carbonylC_scAGN_medshort     -8.404e-02  6.627e-02  -1.268 0.205031
## carbonylC_scAGN_medlong      -8.147e-02  6.137e-02  -1.328 0.184548
## carbonylC_scAGN_long         -4.041e-02  4.874e-02  -0.829 0.407267
## carbonylC_scAGN_vlong        -1.737e-02  2.923e-02  -0.594 0.552439
## carbonylC_scLysN_short        7.321e-02  7.389e-02   0.991 0.321937
## carbonylC_scLysN_medlong      5.359e-02  6.233e-02   0.860 0.390054
## carbonylC_scLysN_long        -2.001e-02  6.259e-02  -0.320 0.749266
## carbonylC_scLysN_vlong        9.811e-02  5.352e-02   1.833 0.067040 .
## carbonylC_bbProN_medlong     -8.236e-02  5.140e-02  -1.602 0.109350
## carbonylC_bbProN_long        -4.392e-02  4.364e-02  -1.006 0.314503
## carbonylC_bbProN_vlong       -8.427e-02  2.876e-02  -2.930 0.003451 **
## carbonylC_hydroxylO_short    -1.310e-01  5.483e-02  -2.388 0.017070 *
## carbonylC_hydroxylO_medshort -1.337e-01  4.669e-02  -2.864 0.004256 **
## carbonylC_hydroxylO_medlong  -6.135e-02  3.940e-02  -1.557 0.119721
## carbonylC_hydroxylO_long     -2.872e-02  3.026e-02  -0.949 0.342794
## carbonylC_hydroxylO_vlong    -5.743e-02  1.845e-02  -3.113 0.001896 **
## carbonylC_carbonylO_short     1.720e-01  5.678e-02   3.029 0.002500 **
## carbonylC_carbonylO_medshort  1.215e-01  5.525e-02   2.199 0.028089 *
## carbonylC_carbonylO_medlong   5.772e-02  5.474e-02   1.054 0.291875
## carbonylC_carbonylO_long      8.859e-02  4.507e-02   1.966 0.049554 *
## carbonylC_carbonylO_vlong     1.013e-01  2.949e-02   3.434 0.000614 ***
## carbonylC_carboxylO_short     3.968e-01  8.490e-02   4.674 3.27e-06 ***
## carbonylC_carboxylO_medshort  2.767e-01  7.547e-02   3.667 0.000256 ***
## carbonylC_carboxylO_medlong   2.192e-01  6.413e-02   3.418 0.000650 ***
## carbonylC_carboxylO_long      1.966e-01  5.277e-02   3.725 0.000204 ***
## carbonylC_carboxylO_vlong     8.787e-02  2.914e-02   3.015 0.002617 **
## carbonylC_sulfur_short       -1.003e-01  9.182e-02  -1.092 0.274933
## carbonylC_sulfur_medshort     2.448e-02  7.880e-02   0.311 0.756145
## carbonylC_sulfur_medlong     -3.022e-02  6.370e-02  -0.474 0.635328
## carbonylC_sulfur_long        -6.866e-02  4.897e-02  -1.402 0.161173
## carbonylC_sulfur_vlong       -4.713e-02  3.072e-02  -1.534 0.125202
## carbonylC_bbN_short           1.793e-02  4.988e-02   0.360 0.719265
## carbonylC_bbN_medshort       -5.284e-03  3.250e-02  -0.163 0.870875
## carbonylC_bbN_medlong        -1.643e-02  2.308e-02  -0.712 0.476785
## carbonylC_bbN_long           -5.604e-03  1.676e-02  -0.334 0.738076
## carbonylC_bbN_vlong          -1.357e-02  1.005e-02  -1.350 0.177242
## carbonylC_bbCA_medshort      -5.385e-02  2.808e-02  -1.918 0.055329 .
## carbonylC_bbCA_medlong       -4.537e-02  2.312e-02  -1.962 0.049969 *
## carbonylC_bbCA_long          -2.846e-02  1.679e-02  -1.695 0.090378 .
## carbonylC_bbCA_vlong         -6.187e-03  1.005e-02  -0.616 0.538181
## carbonylC_bbC_medshort       -5.407e-02  2.994e-02  -1.806 0.071191 .
## carbonylC_bbC_medlong        -7.858e-02  2.312e-02  -3.399 0.000698 ***
## carbonylC_bbC_long           -6.291e-02  1.561e-02  -4.030 5.91e-05 ***
## carbonylC_bbC_vlong          -4.576e-02  9.640e-03  -4.747 2.30e-06 ***
## carbonylC_bbO_short          -4.308e-02  3.853e-02  -1.118 0.263649
## carbonylC_bbO_medshort       -4.243e-03  2.940e-02  -0.144 0.885264
## carbonylC_bbO_medlong        -6.652e-03  2.176e-02  -0.306 0.759931
## carbonylC_bbO_long            5.428e-03  1.511e-02   0.359 0.719496
## carbonylC_bbO_vlong           1.861e-03  8.902e-03   0.209 0.834410
## carboxylC_carboxylC_vlong    -1.294e-01  6.180e-02  -2.095 0.036413 *
## carboxylC_aliph2HC_short      3.164e-02  8.186e-02   0.386 0.699211
```

```
## carboxylC_aliph2HC_medshort  -2.198e-02  6.012e-02  -0.366 0.714722
## carboxylC_aliph2HC_medlong    7.831e-04  4.600e-02   0.017 0.986420
## carboxylC_aliph2HC_long       9.652e-03  2.990e-02   0.323 0.746875
## carboxylC_aliph2HC_vlong      3.495e-02  1.610e-02   2.172 0.030075 *
## carboxylC_aliph3HC_medlong   -2.432e-02  6.080e-02  -0.400 0.689194
## carboxylC_aliph3HC_vlong     -7.174e-02  3.286e-02  -2.183 0.029205 *
## carboxylC_aromaticC_medshort  4.654e-02  4.376e-02   1.064 0.287680
## carboxylC_aromaticC_medlong   7.415e-02  4.250e-02   1.745 0.081315 .
## carboxylC_aromaticC_long      4.225e-02  3.481e-02   1.214 0.225017
## carboxylC_aromaticC_vlong     2.188e-02  2.262e-02   0.967 0.333623
## carboxylC_scAGN_short         1.587e-02  7.229e-02   0.220 0.826259
## carboxylC_scAGN_long          5.976e-02  5.213e-02   1.146 0.251892
## carboxylC_scAGN_vlong         9.746e-04  3.704e-02   0.026 0.979014
## carboxylC_scLysN_long         7.755e-02  8.028e-02   0.966 0.334251
## carboxylC_scLysN_vlong       -1.289e-03  6.926e-02  -0.019 0.985151
## carboxylC_scArgN_vlong       -3.528e-02  4.712e-02  -0.749 0.454160
## carboxylC_hydroxylO_short    -2.543e-01  1.572e-01  -1.618 0.105926
## carboxylC_hydroxylO_medshort  4.410e-02  1.134e-01   0.389 0.697325
## carboxylC_hydroxylO_medlong   3.438e-02  8.423e-02   0.408 0.683186
## carboxylC_hydroxylO_vlong    -2.231e-02  3.254e-02  -0.686 0.493062
## carboxylC_carbonylO_medshort -9.404e-02  6.446e-02  -1.459 0.144836
## carboxylC_carbonylO_medlong  -6.848e-02  5.985e-02  -1.144 0.252779
## carboxylC_carbonylO_vlong     1.415e-02  3.608e-02   0.392 0.695053
## carboxylC_carboxylO_medlong  -3.309e-02  7.567e-02  -0.437 0.661999
## carboxylC_carboxylO_long     -5.343e-02  6.549e-02  -0.816 0.414699
## carboxylC_carboxylO_vlong     2.937e-02  4.675e-02   0.628 0.530046
## carboxylC_sulfur_medshort     4.047e-03  5.794e-02   0.070 0.944333
## carboxylC_sulfur_vlong        1.789e-02  3.464e-02   0.517 0.605525
## carboxylC_bbN_medshort        1.409e-01  6.862e-02   2.053 0.040301 *
## carboxylC_bbN_medlong         7.043e-02  4.921e-02   1.431 0.152633
## carboxylC_bbN_long            7.278e-02  3.386e-02   2.149 0.031802 *
## carboxylC_bbN_vlong           1.648e-02  2.002e-02   0.823 0.410696
## carboxylC_bbCA_medshort      -7.352e-02  6.998e-02  -1.051 0.293633
## carboxylC_bbCA_medlong       -6.510e-02  5.557e-02  -1.172 0.241572
## carboxylC_bbCA_long          -4.446e-03  3.801e-02  -0.117 0.906903
## carboxylC_bbCA_vlong         -2.349e-02  2.223e-02  -1.057 0.290684
## carboxylC_bbC_medshort       -6.773e-02  7.275e-02  -0.931 0.352040
## carboxylC_bbC_medlong        -1.211e-02  5.330e-02  -0.227 0.820331
## carboxylC_bbC_long           -5.444e-02  3.857e-02  -1.411 0.158344
## carboxylC_bbC_vlong          -5.685e-02  2.248e-02  -2.529 0.011550 *
## carboxylC_bbO_medshort       -6.618e-02  5.351e-02  -1.237 0.216396
## carboxylC_bbO_medlong        -4.674e-03  4.538e-02  -0.103 0.917976
## carboxylC_bbO_long            2.206e-03  3.374e-02   0.065 0.947885
## carboxylC_bbO_vlong          -2.087e-02  1.827e-02  -1.142 0.253559
## aliph1HC_aliph1HC_medlong     2.201e-02  6.691e-02   0.329 0.742205
## aliph1HC_aliph1HC_long        7.096e-02  4.972e-02   1.427 0.153769
## aliph1HC_aliph1HC_vlong       2.862e-02  3.404e-02   0.841 0.400662
## aliph1HC_aliph2HC_short      -1.394e-02  4.815e-02  -0.290 0.772235
## aliph1HC_aliph2HC_medshort    1.251e-02  2.381e-02   0.526 0.599270
## aliph1HC_aliph2HC_medlong     3.284e-02  1.673e-02   1.963 0.049825 *
## aliph1HC_aliph2HC_long        2.494e-02  1.266e-02   1.969 0.049165 *
## aliph1HC_aliph2HC_vlong      -6.217e-03  9.207e-03  -0.675 0.499636
## aliph1HC_aliph3HC_short       1.101e-01  5.230e-02   2.105 0.035517 *
## aliph1HC_aliph3HC_medshort    4.284e-02  3.858e-02   1.110 0.267093
```

```
## aliph1HC_aliph3HC_medlong    -4.081e-03  3.153e-02  -0.129 0.897050
## aliph1HC_aliph3HC_long        7.543e-04  2.408e-02   0.031 0.975014
## aliph1HC_aliph3HC_vlong      -1.081e-02  1.598e-02  -0.677 0.498617
## aliph1HC_aromaticC_short      8.005e-02  3.043e-02   2.631 0.008619 **
## aliph1HC_aromaticC_medshort   5.710e-02  1.832e-02   3.117 0.001865 **
## aliph1HC_aromaticC_medlong    7.032e-02  1.550e-02   4.537 6.25e-06 ***
## aliph1HC_aromaticC_long       6.525e-02  1.263e-02   5.168 2.75e-07 ***
## aliph1HC_aromaticC_vlong      6.584e-02  1.116e-02   5.897 4.74e-09 ***
## aliph1HC_scAGN_short          9.725e-02  5.297e-02   1.836 0.066608 .
## aliph1HC_scAGN_medshort       1.278e-01  4.213e-02   3.034 0.002463 **
## aliph1HC_scAGN_medlong        9.251e-02  3.605e-02   2.566 0.010399 *
## aliph1HC_scAGN_long           7.078e-02  2.965e-02   2.387 0.017115 *
## aliph1HC_scAGN_vlong          2.625e-02  1.939e-02   1.354 0.175956
## aliph1HC_scArgN_long          1.559e-01  2.043e-01   0.763 0.445376
## aliph1HC_bbProN_medlong       3.027e-01  7.583e-02   3.992 6.92e-05 ***
## aliph1HC_bbProN_long          1.239e-01  5.461e-02   2.269 0.023460 *
## aliph1HC_bbProN_vlong         7.053e-02  3.601e-02   1.959 0.050389 .
## aliph1HC_hydroxylO_medshort   1.050e-01  3.913e-02   2.684 0.007367 **
## aliph1HC_hydroxylO_medlong    1.170e-01  3.614e-02   3.238 0.001233 **
## aliph1HC_hydroxylO_long       4.901e-02  2.667e-02   1.838 0.066338 .
## aliph1HC_hydroxylO_vlong      4.951e-02  1.988e-02   2.490 0.012898 *
## aliph1HC_carbonylO_short     -2.918e-02  5.641e-02  -0.517 0.605002
## aliph1HC_carbonylO_medshort   4.818e-05  4.447e-02   0.001 0.999136
## aliph1HC_carbonylO_medlong   -1.737e-02  3.660e-02  -0.475 0.635051
## aliph1HC_carbonylO_long      -2.008e-02  2.967e-02  -0.677 0.498666
## aliph1HC_carbonylO_vlong     -7.343e-03  1.887e-02  -0.389 0.697200
## aliph1HC_carboxylO_vlong     -1.582e-02  3.479e-02  -0.455 0.649387
## aliph1HC_sulfur_short         1.862e-02  7.013e-02   0.265 0.790716
## aliph1HC_sulfur_medshort     -1.812e-01  4.828e-02  -3.752 0.000183 ***
## aliph1HC_sulfur_medlong      -1.208e-01  3.800e-02  -3.178 0.001520 **
## aliph1HC_sulfur_long         -7.492e-02  3.099e-02  -2.417 0.015776 *
## aliph1HC_sulfur_vlong        -2.977e-02  2.213e-02  -1.345 0.178836
## aliph1HC_bbN_short           -5.762e-02  5.341e-02  -1.079 0.280865
## aliph1HC_bbN_medshort        -5.269e-02  3.272e-02  -1.610 0.107540
## aliph1HC_bbN_medlong         -1.581e-02  2.596e-02  -0.609 0.542555
## aliph1HC_bbN_long            -8.049e-03  1.907e-02  -0.422 0.673005
## aliph1HC_bbN_vlong           -2.811e-02  1.228e-02  -2.289 0.022234 *
## aliph1HC_bbCA_medshort        7.361e-02  3.799e-02   1.937 0.052936 .
## aliph1HC_bbCA_medlong         1.033e-02  2.839e-02   0.364 0.716067
## aliph1HC_bbCA_long            2.138e-02  2.070e-02   1.033 0.301845
## aliph1HC_bbCA_vlong           1.030e-02  1.250e-02   0.824 0.410132
## aliph1HC_bbC_medshort         1.511e-02  3.747e-02   0.403 0.686858
## aliph1HC_bbC_medlong         -3.837e-02  2.893e-02  -1.326 0.185019
## aliph1HC_bbC_long            -3.458e-02  2.138e-02  -1.618 0.106014
## aliph1HC_bbC_vlong           -1.988e-02  1.244e-02  -1.598 0.110214
## aliph1HC_bbO_short           -4.104e-02  4.255e-02  -0.965 0.334932
## aliph1HC_bbO_medshort        -6.077e-02  3.160e-02  -1.923 0.054700 .
## aliph1HC_bbO_medlong         -1.064e-01  2.407e-02  -4.420 1.07e-05 ***
## aliph1HC_bbO_long            -7.306e-02  1.713e-02  -4.264 2.16e-05 ***
## aliph1HC_bbO_vlong           -9.127e-03  1.027e-02  -0.889 0.374239
## aliph2HC_aliph2HC_short      -2.009e-02  2.284e-02  -0.879 0.379304
## aliph2HC_aliph2HC_medshort   -2.186e-02  1.202e-02  -1.818 0.069252 .
## aliph2HC_aliph2HC_medlong    -9.785e-03  8.567e-03  -1.142 0.253612
## aliph2HC_aliph2HC_long       -1.417e-02  6.345e-03  -2.233 0.025742 *
```

```
## aliph2HC_aliph2HC_vlong      -4.133e-03  4.699e-03  -0.880 0.379291
## aliph2HC_aliph3HC_vshort      2.875e-02  4.758e-02   0.604 0.545858
## aliph2HC_aliph3HC_short       4.021e-03  1.946e-02   0.207 0.836329
## aliph2HC_aliph3HC_medshort   -2.192e-02  1.304e-02  -1.681 0.093069 .
## aliph2HC_aliph3HC_medlong    -2.496e-02  1.022e-02  -2.441 0.014783 *
## aliph2HC_aliph3HC_long       -2.917e-02  8.070e-03  -3.615 0.000313 ***
## aliph2HC_aliph3HC_vlong       1.237e-02  6.030e-03   2.051 0.040488 *
## aliph2HC_aromaticC_short      2.094e-02  1.102e-02   1.899 0.057799 .
## aliph2HC_aromaticC_medshort   2.509e-02  7.108e-03   3.529 0.000432 ***
## aliph2HC_aromaticC_medlong    6.087e-03  5.248e-03   1.160 0.246302
## aliph2HC_aromaticC_long       1.048e-02  4.370e-03   2.399 0.016588 *
## aliph2HC_aromaticC_vlong      1.653e-02  3.431e-03   4.819 1.62e-06 ***
## aliph2HC_scAGN_vshort         1.377e-01  4.947e-02   2.783 0.005471 **
## aliph2HC_scAGN_short          9.155e-02  2.491e-02   3.676 0.000247 ***
## aliph2HC_scAGN_medshort       7.051e-02  1.898e-02   3.716 0.000212 ***
## aliph2HC_scAGN_medlong        5.295e-02  1.504e-02   3.520 0.000446 ***
## aliph2HC_scAGN_long           2.204e-02  1.080e-02   2.041 0.041458 *
## aliph2HC_scAGN_vlong          1.572e-02  7.607e-03   2.066 0.038988 *
## aliph2HC_scLysN_medshort     -7.747e-02  4.325e-02  -1.791 0.073516 .
## aliph2HC_scLysN_medlong      -2.575e-02  3.303e-02  -0.779 0.435877
## aliph2HC_scLysN_long          2.932e-03  2.802e-02   0.105 0.916655
## aliph2HC_scLysN_vlong         1.154e-02  2.725e-02   0.424 0.671943
## aliph2HC_scArgN_medshort     -3.535e-01  3.878e-01  -0.912 0.362170
## aliph2HC_scArgN_medlong      -2.035e-01  1.136e-01  -1.791 0.073538 .
## aliph2HC_scArgN_long         -1.879e-03  5.340e-02  -0.035 0.971936
## aliph2HC_scArgN_vlong         1.091e-01  3.707e-02   2.944 0.003302 **
## aliph2HC_bbProN_medshort      1.305e-02  3.438e-02   0.380 0.704246
## aliph2HC_bbProN_medlong       8.523e-03  2.455e-02   0.347 0.728538
## aliph2HC_bbProN_long          1.474e-02  1.936e-02   0.761 0.446746
## aliph2HC_bbProN_vlong         2.265e-02  1.358e-02   1.668 0.095527 .
## aliph2HC_hydroxylO_short      5.950e-02  1.821e-02   3.267 0.001117 **
## aliph2HC_hydroxylO_medshort   2.456e-02  1.326e-02   1.852 0.064250 .
## aliph2HC_hydroxylO_medlong    3.081e-02  1.125e-02   2.739 0.006251 **
## aliph2HC_hydroxylO_long       1.997e-02  8.282e-03   2.411 0.016067 *
## aliph2HC_hydroxylO_vlong      3.261e-03  6.185e-03   0.527 0.598088
## aliph2HC_carbonylO_vshort     9.282e-02  4.501e-02   2.062 0.039382 *
## aliph2HC_carbonylO_short      4.320e-02  2.493e-02   1.733 0.083366 .
## aliph2HC_carbonylO_medshort   8.581e-03  1.874e-02   0.458 0.647111
## aliph2HC_carbonylO_medlong    2.265e-02  1.587e-02   1.427 0.153802
## aliph2HC_carbonylO_long       1.511e-02  1.109e-02   1.362 0.173319
## aliph2HC_carbonylO_vlong      1.987e-02  7.535e-03   2.638 0.008451 **
## aliph2HC_carboxylO_short      1.375e-03  3.971e-02   0.035 0.972375
## aliph2HC_carboxylO_medshort  -1.714e-03  3.161e-02  -0.054 0.956760
## aliph2HC_carboxylO_medlong   -1.256e-02  2.654e-02  -0.474 0.635935
## aliph2HC_carboxylO_long      -2.354e-02  1.817e-02  -1.295 0.195510
## aliph2HC_carboxylO_vlong     -2.701e-02  9.789e-03  -2.759 0.005876 **
## aliph2HC_sulfur_short         1.837e-02  2.898e-02   0.634 0.526166
## aliph2HC_sulfur_medshort     -3.633e-02  2.014e-02  -1.804 0.071485 .
## aliph2HC_sulfur_medlong      -9.203e-03  1.630e-02  -0.565 0.572359
## aliph2HC_sulfur_long         -1.985e-03  1.312e-02  -0.151 0.879774
## aliph2HC_sulfur_vlong        -7.273e-03  9.829e-03  -0.740 0.459485
## aliph2HC_bbN_short           -2.712e-02  2.385e-02  -1.137 0.255674
## aliph2HC_bbN_medshort        -3.819e-02  1.225e-02  -3.117 0.001871 **
## aliph2HC_bbN_medlong          1.039e-02  9.242e-03   1.124 0.261308
```

```
## aliph2HC_bbN_long             7.665e-03  6.936e-03   1.105 0.269303
## aliph2HC_bbN_vlong            5.735e-03  4.284e-03   1.339 0.180911
## aliph2HC_bbCA_short           1.450e-02  3.087e-02   0.470 0.638700
## aliph2HC_bbCA_medshort        2.653e-02  1.590e-02   1.668 0.095536 .
## aliph2HC_bbCA_medlong         5.157e-03  1.117e-02   0.462 0.644358
## aliph2HC_bbCA_long            2.474e-03  7.516e-03   0.329 0.742059
## aliph2HC_bbCA_vlong          -3.036e-03  4.737e-03  -0.641 0.521611
## aliph2HC_bbC_short            1.074e-02  3.079e-02   0.349 0.727254
## aliph2HC_bbC_medshort         6.607e-03  1.611e-02   0.410 0.681857
## aliph2HC_bbC_medlong         -5.320e-03  1.170e-02  -0.455 0.649517
## aliph2HC_bbC_long             1.378e-02  8.169e-03   1.687 0.091853 .
## aliph2HC_bbC_vlong            7.686e-03  4.915e-03   1.564 0.118088
## aliph2HC_bbO_vshort           6.219e-02  3.938e-02   1.579 0.114555
## aliph2HC_bbO_short            1.748e-03  1.657e-02   0.105 0.916000
## aliph2HC_bbO_medshort         1.227e-03  1.160e-02   0.106 0.915799
## aliph2HC_bbO_medlong         -8.969e-03  8.574e-03  -1.046 0.295720
## aliph2HC_bbO_long            -6.025e-03  5.763e-03  -1.046 0.295971
## aliph2HC_bbO_vlong           -6.957e-03  3.627e-03  -1.918 0.055328 .
## aliph3HC_aliph3HC_short      -4.559e-02  3.829e-02  -1.191 0.233983
## aliph3HC_aliph3HC_medshort   -3.365e-02  3.360e-02  -1.001 0.316809
## aliph3HC_aliph3HC_medlong    -6.046e-04  2.683e-02  -0.023 0.982022
## aliph3HC_aliph3HC_long       -4.342e-03  2.101e-02  -0.207 0.836290
## aliph3HC_aliph3HC_vlong       1.768e-02  1.353e-02   1.307 0.191541
## aliph3HC_aromaticC_short      9.313e-03  1.199e-02   0.777 0.437476
## aliph3HC_aromaticC_medshort   7.799e-03  9.244e-03   0.844 0.399018
## aliph3HC_aromaticC_medlong   -1.285e-02  7.323e-03  -1.755 0.079528 .
## aliph3HC_aromaticC_long      -7.750e-03  5.937e-03  -1.305 0.192002
## aliph3HC_aromaticC_vlong     -1.229e-02  4.778e-03  -2.571 0.010246 *
## aliph3HC_scAGN_short         -8.393e-02  3.010e-02  -2.788 0.005381 **
## aliph3HC_scAGN_medshort      -3.880e-02  2.868e-02  -1.353 0.176271
## aliph3HC_scAGN_medlong       -5.747e-02  2.542e-02  -2.261 0.023954 *
## aliph3HC_scAGN_long          -4.502e-02  1.987e-02  -2.266 0.023644 *
## aliph3HC_scAGN_vlong         -3.015e-02  1.300e-02  -2.319 0.020560 *
## aliph3HC_scLysN_vlong         5.213e-02  3.797e-02   1.373 0.170074
## aliph3HC_scArgN_medlong      -7.720e-02  1.058e-01  -0.730 0.465730
## aliph3HC_scArgN_long          6.657e-02  1.015e-01   0.656 0.512006
## aliph3HC_scArgN_vlong         1.980e-01  7.581e-02   2.611 0.009136 **
## aliph3HC_bbProN_medlong      -7.433e-02  3.331e-02  -2.232 0.025814 *
## aliph3HC_bbProN_long         -3.455e-02  2.679e-02  -1.290 0.197325
## aliph3HC_bbProN_vlong         2.095e-02  2.017e-02   1.038 0.299260
## aliph3HC_hydroxylO_short     -9.024e-02  2.912e-02  -3.099 0.001986 **
## aliph3HC_hydroxylO_medshort  -3.936e-02  2.375e-02  -1.658 0.097619 .
## aliph3HC_hydroxylO_medlong   -4.472e-04  1.954e-02  -0.023 0.981742
## aliph3HC_hydroxylO_long       8.807e-03  1.533e-02   0.574 0.565753
## aliph3HC_hydroxylO_vlong     -8.820e-04  1.119e-02  -0.079 0.937180
## aliph3HC_carbonylO_short      4.600e-03  3.211e-02   0.143 0.886104
## aliph3HC_carbonylO_medshort  -1.139e-05  2.939e-02   0.000 0.999691
## aliph3HC_carbonylO_medlong    3.596e-03  2.537e-02   0.142 0.887310
## aliph3HC_carbonylO_long       1.754e-04  2.014e-02   0.009 0.993052
## aliph3HC_carbonylO_vlong     -9.195e-05  1.288e-02  -0.007 0.994306
## aliph3HC_carboxylO_medshort  -7.467e-02  4.514e-02  -1.654 0.098366 .
## aliph3HC_carboxylO_medlong   -5.415e-02  3.750e-02  -1.444 0.148988
## aliph3HC_carboxylO_long      -4.571e-02  3.675e-02  -1.244 0.213805
## aliph3HC_carboxylO_vlong     -1.350e-02  2.074e-02  -0.651 0.515169
```

```
## aliph3HC_sulfur_short         9.287e-02  3.484e-02   2.666 0.007779 **
## aliph3HC_sulfur_medshort      1.318e-01  2.661e-02   4.955 8.22e-07 ***
## aliph3HC_sulfur_medlong       1.028e-01  2.378e-02   4.321 1.67e-05 ***
## aliph3HC_sulfur_long          5.059e-02  1.993e-02   2.539 0.011243 *
## aliph3HC_sulfur_vlong         3.395e-02  1.555e-02   2.183 0.029236 *
## aliph3HC_bbN_short            2.620e-02  2.952e-02   0.887 0.375048
## aliph3HC_bbN_medshort        -1.730e-02  2.006e-02  -0.862 0.388659
## aliph3HC_bbN_medlong         -8.430e-03  1.559e-02  -0.541 0.588728
## aliph3HC_bbN_long             2.938e-03  1.156e-02   0.254 0.799347
## aliph3HC_bbN_vlong            3.015e-03  7.030e-03   0.429 0.668079
## aliph3HC_bbCA_short           6.450e-02  3.543e-02   1.820 0.068968 .
## aliph3HC_bbCA_medshort        2.378e-02  2.440e-02   0.975 0.329806
## aliph3HC_bbCA_medlong         4.754e-02  1.816e-02   2.617 0.008978 **
## aliph3HC_bbCA_long            3.110e-02  1.279e-02   2.432 0.015166 *
## aliph3HC_bbCA_vlong           1.715e-02  7.641e-03   2.244 0.024992 *
## aliph3HC_bbC_short            3.644e-02  3.033e-02   1.202 0.229743
## aliph3HC_bbC_medshort         2.457e-02  2.308e-02   1.064 0.287367
## aliph3HC_bbC_medlong          6.966e-03  1.786e-02   0.390 0.696611
## aliph3HC_bbC_long            -9.864e-04  1.202e-02  -0.082 0.934633
## aliph3HC_bbC_vlong            1.548e-02  7.554e-03   2.050 0.040594 *
## aliph3HC_bbO_short           -3.134e-04  2.086e-02  -0.015 0.988017
## aliph3HC_bbO_medshort         1.899e-02  1.718e-02   1.105 0.269279
## aliph3HC_bbO_medlong          1.133e-02  1.285e-02   0.882 0.377721
## aliph3HC_bbO_long             5.736e-03  9.715e-03   0.590 0.555019
## aliph3HC_bbO_vlong            5.926e-05  6.208e-03   0.010 0.992386
## aromaticC_aromaticC_short    -3.079e-03  9.783e-03  -0.315 0.753025
## aromaticC_aromaticC_medshort -8.645e-03  5.761e-03  -1.501 0.133714
## aromaticC_aromaticC_medlong  -1.215e-04  3.894e-03  -0.031 0.975108
## aromaticC_aromaticC_long     -2.639e-03  3.598e-03  -0.734 0.463363
## aromaticC_aromaticC_vlong    -9.593e-04  3.739e-03  -0.257 0.797531
## aromaticC_scAGN_short        -1.069e-02  2.269e-02  -0.471 0.637827
## aromaticC_scAGN_medshort     -2.275e-02  1.555e-02  -1.463 0.143801
## aromaticC_scAGN_medlong      -1.728e-02  1.207e-02  -1.432 0.152288
## aromaticC_scAGN_long          9.038e-03  1.079e-02   0.838 0.402265
## aromaticC_scAGN_vlong         3.874e-03  7.035e-03   0.551 0.582007
## aromaticC_scLysN_medlong     -3.710e-02  3.651e-02  -1.016 0.309746
## aromaticC_scLysN_long        -2.389e-02  3.014e-02  -0.793 0.428083
## aromaticC_scLysN_vlong       -6.348e-02  2.642e-02  -2.402 0.016431 *
## aromaticC_scArgN_medshort     3.176e-01  1.564e-01   2.030 0.042522 *
## aromaticC_scArgN_medlong     -1.012e-01  6.921e-02  -1.462 0.143868
## aromaticC_scArgN_long        -9.231e-02  5.284e-02  -1.747 0.080878 .
## aromaticC_scArgN_vlong       -5.260e-02  3.116e-02  -1.688 0.091690 .
## aromaticC_bbProN_medlong     -8.852e-02  2.882e-02  -3.072 0.002174 **
## aromaticC_bbProN_long        -7.657e-03  2.057e-02  -0.372 0.709813
## aromaticC_bbProN_vlong       -2.808e-02  1.581e-02  -1.775 0.076060 .
## aromaticC_hydroxylO_vshort   -1.130e-01  4.735e-02  -2.387 0.017138 *
## aromaticC_hydroxylO_short    -3.296e-02  2.028e-02  -1.625 0.104375
## aromaticC_hydroxylO_medshort -3.850e-02  1.442e-02  -2.671 0.007670 **
## aromaticC_hydroxylO_medlong  -3.852e-02  1.017e-02  -3.787 0.000160 ***
## aromaticC_hydroxylO_long     -2.340e-02  7.662e-03  -3.054 0.002305 **
## aromaticC_hydroxylO_vlong    -8.792e-03  6.194e-03  -1.419 0.156060
## aromaticC_carbonylO_short    -5.955e-03  2.386e-02  -0.250 0.802938
## aromaticC_carbonylO_medshort  4.862e-03  1.601e-02   0.304 0.761394
## aromaticC_carbonylO_medlong   2.727e-02  1.281e-02   2.129 0.033430 *
```

```
## aromaticC_carbonylO_long      1.856e-04  1.122e-02   0.017 0.986802
## aromaticC_carbonylO_vlong     1.103e-02  7.519e-03   1.466 0.142805
## aromaticC_carboxylO_medshort  1.120e-02  2.603e-02   0.430 0.667021
## aromaticC_carboxylO_medlong  -2.715e-02  2.322e-02  -1.169 0.242513
## aromaticC_carboxylO_long     -1.672e-02  2.181e-02  -0.767 0.443404
## aromaticC_carboxylO_vlong     9.091e-03  1.529e-02   0.595 0.552194
## aromaticC_sulfur_short       -5.659e-02  1.947e-02  -2.906 0.003719 **
## aromaticC_sulfur_medshort    -9.800e-03  1.447e-02  -0.678 0.498198
## aromaticC_sulfur_medlong      2.858e-02  1.515e-02   1.887 0.059441 .
## aromaticC_sulfur_long         3.707e-02  1.281e-02   2.893 0.003881 **
## aromaticC_sulfur_vlong        1.325e-02  1.057e-02   1.254 0.210067
## aromaticC_bbN_short          -1.017e-02  2.103e-02  -0.484 0.628620
## aromaticC_bbN_medshort       -1.835e-02  1.237e-02  -1.484 0.138163
## aromaticC_bbN_medlong        -1.314e-02  9.435e-03  -1.392 0.164084
## aromaticC_bbN_long            6.164e-03  6.842e-03   0.901 0.367805
## aromaticC_bbN_vlong           8.217e-03  4.753e-03   1.729 0.084104 .
## aromaticC_bbCA_short         -3.216e-04  2.650e-02  -0.012 0.990319
## aromaticC_bbCA_medshort      -8.366e-03  1.565e-02  -0.535 0.592993
## aromaticC_bbCA_medlong       -7.959e-03  1.150e-02  -0.692 0.488915
## aromaticC_bbCA_long          -8.701e-03  7.993e-03  -1.089 0.276555
## aromaticC_bbCA_vlong          4.557e-03  4.838e-03   0.942 0.346354
## aromaticC_bbC_short           1.345e-01  3.708e-02   3.628 0.000297 ***
## aromaticC_bbC_medshort        3.256e-02  1.746e-02   1.865 0.062419 .
## aromaticC_bbC_medlong         3.147e-02  1.306e-02   2.410 0.016117 *
## aromaticC_bbC_long            2.000e-02  8.710e-03   2.296 0.021840 *
## aromaticC_bbC_vlong           5.132e-03  5.323e-03   0.964 0.335179
## aromaticC_bbO_vshort         -1.654e-01  5.720e-02  -2.891 0.003902 **
## aromaticC_bbO_short          -2.011e-02  1.733e-02  -1.160 0.246262
## aromaticC_bbO_medshort       -7.973e-03  1.097e-02  -0.727 0.467620
## aromaticC_bbO_medlong         1.287e-03  8.011e-03   0.161 0.872409
## aromaticC_bbO_long            2.522e-04  5.739e-03   0.044 0.964960
## aromaticC_bbO_vlong           9.423e-03  3.923e-03   2.402 0.016456 *
## scAGN_scAGN_short             1.197e-01  6.197e-02   1.931 0.053700 .
## scAGN_scAGN_medshort          1.289e-01  6.337e-02   2.034 0.042164 *
## scAGN_scAGN_medlong           1.214e-01  6.000e-02   2.023 0.043331 *
## scAGN_scAGN_long              9.336e-02  5.039e-02   1.853 0.064171 .
## scAGN_scAGN_vlong             6.382e-03  3.583e-02   0.178 0.858647
## scAGN_scLysN_medshort        -4.709e-04  5.580e-02  -0.008 0.993268
## scAGN_scLysN_medlong         -3.680e-02  5.829e-02  -0.631 0.527959
## scAGN_scLysN_long             5.547e-04  5.289e-02   0.010 0.991634
## scAGN_scLysN_vlong           -4.471e-03  4.396e-02  -0.102 0.919004
## scAGN_bbProN_medshort        -1.876e-01  6.063e-02  -3.095 0.002013 **
## scAGN_bbProN_medlong         -6.979e-02  4.793e-02  -1.456 0.145627
## scAGN_bbProN_long            -3.615e-02  3.620e-02  -0.998 0.318245
## scAGN_bbProN_vlong           -5.018e-03  2.505e-02  -0.200 0.841261
## scAGN_hydroxylO_vshort        2.208e-02  5.017e-02   0.440 0.659934
## scAGN_hydroxylO_short        -1.345e-02  4.029e-02  -0.334 0.738542
## scAGN_hydroxylO_medshort     -1.036e-02  3.437e-02  -0.301 0.763087
## scAGN_hydroxylO_medlong      -1.844e-03  2.813e-02  -0.066 0.947732
## scAGN_hydroxylO_long         -1.872e-02  2.218e-02  -0.844 0.398738
## scAGN_hydroxylO_vlong        -6.939e-03  1.606e-02  -0.432 0.665796
## scAGN_carbonylO_vshort       -1.321e-01  8.699e-02  -1.518 0.129176
## scAGN_carbonylO_short        -1.220e-01  7.760e-02  -1.572 0.116115
## scAGN_carbonylO_medshort     -9.289e-02  6.566e-02  -1.415 0.157384
```

```
## scAGN_carbonylO_medlong      -7.779e-05  5.529e-02  -0.001 0.998878
## scAGN_carbonylO_long          6.504e-03  4.282e-02   0.152 0.879294
## scAGN_carbonylO_vlong        -6.924e-03  2.677e-02  -0.259 0.795951
## scAGN_carboxylO_vshort       -3.096e-01  8.796e-02  -3.520 0.000448 ***
## scAGN_carboxylO_short        -2.064e-01  7.044e-02  -2.930 0.003455 **
## scAGN_carboxylO_medshort     -1.450e-01  5.590e-02  -2.594 0.009584 **
## scAGN_carboxylO_medlong      -1.561e-01  5.096e-02  -3.063 0.002240 **
## scAGN_carboxylO_long         -1.134e-01  3.928e-02  -2.888 0.003947 **
## scAGN_carboxylO_vlong        -8.591e-02  2.536e-02  -3.388 0.000725 ***
## scAGN_sulfur_short           -6.279e-02  6.128e-02  -1.025 0.305790
## scAGN_sulfur_medshort        -4.655e-02  5.233e-02  -0.890 0.373807
## scAGN_sulfur_medlong          1.425e-02  4.355e-02   0.327 0.743529
## scAGN_sulfur_long             1.157e-03  3.461e-02   0.033 0.973346
## scAGN_sulfur_vlong           -1.727e-02  2.448e-02  -0.705 0.480680
## scAGN_bbN_short               4.430e-02  4.580e-02   0.967 0.333655
## scAGN_bbN_medshort            4.388e-02  2.407e-02   1.823 0.068549 .
## scAGN_bbN_medlong             3.889e-02  1.864e-02   2.087 0.037128 *
## scAGN_bbN_long                5.221e-02  1.329e-02   3.929 8.97e-05 ***
## scAGN_bbN_vlong               1.983e-02  8.170e-03   2.427 0.015379 *
## scAGN_bbCA_short             -2.404e-02  4.215e-02  -0.570 0.568604
## scAGN_bbCA_medshort           4.875e-03  2.926e-02   0.167 0.867715
## scAGN_bbCA_medlong            2.435e-02  2.196e-02   1.108 0.267877
## scAGN_bbCA_long               2.202e-03  1.556e-02   0.141 0.887502
## scAGN_bbCA_vlong             -8.257e-04  8.663e-03  -0.095 0.924076
## scAGN_bbC_short              -2.666e-02  4.150e-02  -0.642 0.520714
## scAGN_bbC_medshort           -3.551e-02  3.111e-02  -1.141 0.253887
## scAGN_bbC_medlong            -1.016e-02  2.045e-02  -0.497 0.619434
## scAGN_bbC_long                6.764e-03  1.459e-02   0.464 0.643079
## scAGN_bbC_vlong              -1.181e-02  8.630e-03  -1.369 0.171244
## scAGN_bbO_vshort              1.125e-01  4.810e-02   2.339 0.019482 *
## scAGN_bbO_short               1.110e-01  3.806e-02   2.915 0.003616 **
## scAGN_bbO_medshort            5.095e-02  2.601e-02   1.959 0.050330 .
## scAGN_bbO_medlong             4.317e-02  1.789e-02   2.414 0.015937 *
## scAGN_bbO_long                3.103e-02  1.242e-02   2.499 0.012590 *
## scAGN_bbO_vlong               5.611e-03  8.262e-03   0.679 0.497191
## scLysN_hydroxylO_long         5.334e-05  4.670e-02   0.001 0.999089
## scLysN_hydroxylO_vlong       -3.846e-02  3.249e-02  -1.184 0.236668
## scLysN_carbonylO_vshort      -7.124e-02  7.437e-02  -0.958 0.338335
## scLysN_carbonylO_long        -2.773e-02  4.310e-02  -0.643 0.520126
## scLysN_carbonylO_vlong       -6.734e-04  4.013e-02  -0.017 0.986616
## scLysN_carboxylO_medlong     -1.604e-02  6.358e-02  -0.252 0.800905
## scLysN_carboxylO_long         8.014e-02  5.290e-02   1.515 0.130006
## scLysN_carboxylO_vlong        3.346e-02  5.042e-02   0.664 0.507096
## scLysN_bbN_medshort           1.837e-01  9.048e-02   2.030 0.042526 *
## scLysN_bbN_medlong            7.543e-02  6.404e-02   1.178 0.239074
## scLysN_bbN_long              -1.396e-02  4.888e-02  -0.286 0.775304
## scLysN_bbN_vlong             -2.788e-02  2.790e-02  -0.999 0.317907
## scLysN_bbCA_medshort         -2.075e-01  1.032e-01  -2.011 0.044533 *
## scLysN_bbCA_medlong          -2.228e-01  6.392e-02  -3.486 0.000507 ***
## scLysN_bbCA_long             -1.608e-01  5.019e-02  -3.205 0.001385 **
## scLysN_bbCA_vlong            -5.408e-02  3.109e-02  -1.740 0.082171 .
## scLysN_bbC_medlong            9.711e-02  5.410e-02   1.795 0.072920 .
## scLysN_bbC_long               1.346e-01  4.501e-02   2.990 0.002840 **
## scLysN_bbC_vlong             -2.009e-02  3.178e-02  -0.632 0.527423
```

```
## scLysN_bbO_medlong           -4.968e-03  4.151e-02  -0.120 0.904751
## scLysN_bbO_long              -4.755e-02  3.521e-02  -1.351 0.177021
## scLysN_bbO_vlong             -4.583e-02  2.619e-02  -1.750 0.080350 .
## scArgN_hydroxylO_short        2.477e-02  8.991e-02   0.276 0.782965
## scArgN_hydroxylO_medshort    -1.733e-02  1.198e-01  -0.145 0.885013
## scArgN_hydroxylO_medlong      3.129e-02  1.273e-01   0.246 0.805917
## scArgN_hydroxylO_long         6.283e-02  6.594e-02   0.953 0.340831
## scArgN_hydroxylO_vlong        1.660e-02  5.070e-02   0.327 0.743450
## scArgN_carboxylO_long        -3.990e-02  4.761e-02  -0.838 0.402173
## scArgN_carboxylO_vlong        4.746e-02  3.119e-02   1.522 0.128301
## scArgN_bbN_medshort           2.216e-01  2.416e-01   0.917 0.359191
## scArgN_bbN_medlong            5.754e-01  2.493e-01   2.308 0.021171 *
## scArgN_bbN_long               1.977e-01  1.603e-01   1.233 0.217828
## scArgN_bbN_vlong              3.541e-02  7.532e-02   0.470 0.638349
## scArgN_bbCA_medshort          4.255e+00  9.106e-01   4.672 3.29e-06 ***
## scArgN_bbCA_medlong           3.493e+00  9.223e-01   3.788 0.000159 ***
## scArgN_bbCA_long              3.927e-01  1.657e-01   2.369 0.017964 *
## scArgN_bbCA_vlong            -2.260e-01  6.410e-02  -3.526 0.000436 ***
## scArgN_bbC_short             -1.857e+00  4.892e-01  -3.796 0.000154 ***
## scArgN_bbC_medshort                  NA         NA      NA       NA
## scArgN_bbC_medlong           -1.989e+00  4.266e-01  -4.661 3.47e-06 ***
## scArgN_bbC_long              -1.410e+00  3.950e-01  -3.569 0.000371 ***
## scArgN_bbC_vlong              2.674e-01  7.330e-02   3.648 0.000275 ***
## scArgN_bbO_vshort             5.500e-02  1.857e-01   0.296 0.767197
## scArgN_bbO_short                     NA         NA      NA       NA
## scArgN_bbO_medshort           1.385e+00  3.257e-01   4.252 2.28e-05 ***
## scArgN_bbO_medlong            1.140e+00  2.507e-01   4.545 6.01e-06 ***
## scArgN_bbO_long              -5.032e-01  1.534e-01  -3.280 0.001066 **
## scArgN_bbO_vlong             -3.777e-01  1.132e-01  -3.335 0.000878 ***
## bbProN_hydroxylO_medshort     7.788e-02  4.497e-02   1.732 0.083545 .
## bbProN_hydroxylO_medlong      7.072e-02  4.012e-02   1.763 0.078151 .
## bbProN_hydroxylO_long         9.063e-02  3.554e-02   2.550 0.010888 *
## bbProN_hydroxylO_vlong        8.311e-02  3.091e-02   2.689 0.007259 **
## bbProN_carbonylO_medlong      4.067e-02  3.907e-02   1.041 0.298026
## bbProN_carbonylO_long         8.144e-03  3.410e-02   0.239 0.811308
## bbProN_carbonylO_vlong        3.392e-02  2.329e-02   1.456 0.145572
## bbProN_carboxylO_vlong        3.548e-02  3.719e-02   0.954 0.340282
## bbProN_sulfur_vlong          -4.371e-02  4.077e-02  -1.072 0.283843
## bbProN_bbN_medshort           7.333e-02  4.522e-02   1.621 0.105168
## bbProN_bbN_medlong            4.079e-02  3.506e-02   1.164 0.244808
## bbProN_bbN_long               4.448e-02  2.739e-02   1.624 0.104658
## bbProN_bbN_vlong              4.586e-03  1.586e-02   0.289 0.772505
## bbProN_bbCA_medshort          1.136e-01  5.465e-02   2.080 0.037754 *
## bbProN_bbCA_medlong           4.986e-02  4.235e-02   1.177 0.239342
## bbProN_bbCA_long              4.180e-02  2.870e-02   1.457 0.145454
## bbProN_bbCA_vlong             9.702e-03  1.799e-02   0.539 0.589741
## bbProN_bbC_short             -9.722e-02  6.374e-02  -1.525 0.127450
## bbProN_bbC_medshort          -7.123e-02  5.773e-02  -1.234 0.217440
## bbProN_bbC_medlong           -4.058e-02  4.334e-02  -0.936 0.349305
## bbProN_bbC_long              -1.360e-02  3.109e-02  -0.437 0.661883
## bbProN_bbC_vlong             -2.712e-03  1.838e-02  -0.148 0.882721
## bbProN_bbO_short              4.777e-03  4.605e-02   0.104 0.917386
## bbProN_bbO_medshort           1.113e-02  4.219e-02   0.264 0.791945
## bbProN_bbO_medlong            1.447e-02  3.245e-02   0.446 0.655714
```

```
## bbProN_bbO_long               4.669e-02  2.348e-02   1.989 0.046968 *
## bbProN_bbO_vlong              5.126e-02  1.573e-02   3.259 0.001147 **
## hydroxylO_hydroxylO_short     5.604e-03  4.246e-02   0.132 0.895009
## hydroxylO_hydroxylO_medshort -1.290e-02  3.150e-02  -0.410 0.682116
## hydroxylO_hydroxylO_medlong  -1.048e-03  3.147e-02  -0.033 0.973447
## hydroxylO_hydroxylO_long     -1.340e-02  2.323e-02  -0.577 0.564093
## hydroxylO_hydroxylO_vlong     4.393e-03  1.602e-02   0.274 0.783921
## hydroxylO_carbonylO_vshort    7.664e-02  5.354e-02   1.432 0.152528
## hydroxylO_carbonylO_short     7.278e-02  4.246e-02   1.714 0.086756 .
## hydroxylO_carbonylO_medshort  7.403e-02  3.466e-02   2.136 0.032893 *
## hydroxylO_carbonylO_medlong   4.961e-02  2.892e-02   1.715 0.086543 .
## hydroxylO_carbonylO_long      2.923e-02  2.313e-02   1.264 0.206516
## hydroxylO_carbonylO_vlong     9.626e-03  1.577e-02   0.611 0.541613
## hydroxylO_carboxylO_vshort    9.855e-02  1.196e-01   0.824 0.410079
## hydroxylO_carboxylO_short     3.420e-02  8.630e-02   0.396 0.691944
## hydroxylO_carboxylO_medshort -4.606e-02  6.916e-02  -0.666 0.505556
## hydroxylO_carboxylO_medlong  -6.227e-02  5.202e-02  -1.197 0.231480
## hydroxylO_carboxylO_long     -3.341e-02  3.292e-02  -1.015 0.310420
## hydroxylO_carboxylO_vlong     8.489e-03  2.185e-02   0.388 0.697751
## hydroxylO_sulfur_short        1.248e-01  4.242e-02   2.941 0.003335 **
## hydroxylO_sulfur_medshort     1.156e-01  3.282e-02   3.521 0.000445 ***
## hydroxylO_sulfur_medlong      5.765e-02  2.874e-02   2.006 0.045051 *
## hydroxylO_sulfur_long         3.961e-02  2.305e-02   1.719 0.085887 .
## hydroxylO_sulfur_vlong       -4.742e-03  1.765e-02  -0.269 0.788285
## hydroxylO_bbN_short           3.382e-02  2.623e-02   1.289 0.197465
## hydroxylO_bbN_medshort        2.057e-02  2.071e-02   0.993 0.320719
## hydroxylO_bbN_medlong         3.363e-02  1.652e-02   2.036 0.041977 *
## hydroxylO_bbN_long            2.260e-02  1.255e-02   1.801 0.071907 .
## hydroxylO_bbN_vlong           1.659e-02  7.815e-03   2.123 0.033961 *
## hydroxylO_bbCA_short         -8.476e-02  3.066e-02  -2.764 0.005787 **
## hydroxylO_bbCA_medshort      -5.940e-02  2.519e-02  -2.358 0.018523 *
## hydroxylO_bbCA_medlong       -2.942e-02  2.050e-02  -1.435 0.151460
## hydroxylO_bbCA_long          -5.213e-03  1.387e-02  -0.376 0.707138
## hydroxylO_bbCA_vlong         -5.428e-03  8.227e-03  -0.660 0.509462
## hydroxylO_bbC_short          -7.148e-03  3.811e-02  -0.188 0.851234
## hydroxylO_bbC_medshort        2.005e-02  2.922e-02   0.686 0.492853
## hydroxylO_bbC_medlong         1.216e-02  2.271e-02   0.535 0.592521
## hydroxylO_bbC_long           -7.612e-03  1.545e-02  -0.493 0.622196
## hydroxylO_bbC_vlong          -1.301e-02  9.471e-03  -1.374 0.169791
## hydroxylO_bbO_vshort          6.185e-02  3.898e-02   1.587 0.112795
## hydroxylO_bbO_short           2.605e-02  2.747e-02   0.948 0.343080
## hydroxylO_bbO_medshort        2.405e-02  2.106e-02   1.142 0.253670
## hydroxylO_bbO_medlong         2.310e-02  1.674e-02   1.380 0.167836
## hydroxylO_bbO_long            2.904e-02  1.192e-02   2.437 0.014954 *
## hydroxylO_bbO_vlong           1.133e-02  7.730e-03   1.466 0.142794
## carbonylO_carbonylO_medshort -2.771e-02  4.989e-02  -0.555 0.578669
## carbonylO_carbonylO_medlong  -1.078e-02  5.220e-02  -0.207 0.836407
## carbonylO_carbonylO_long     -3.102e-03  5.216e-02  -0.059 0.952593
## carbonylO_carbonylO_vlong    -2.745e-02  3.954e-02  -0.694 0.487607
## carbonylO_carboxylO_medshort -2.495e-02  5.034e-02  -0.495 0.620339
## carbonylO_carboxylO_medlong  -1.326e-02  4.766e-02  -0.278 0.780946
## carbonylO_carboxylO_long     -5.633e-02  4.006e-02  -1.406 0.159852
## carbonylO_carboxylO_vlong    -4.974e-02  2.590e-02  -1.920 0.055069 .
## carbonylO_sulfur_short        2.083e-02  6.474e-02   0.322 0.747708
```

```
## carbonylO_sulfur_medshort     5.425e-02  5.587e-02   0.971 0.331722
## carbonylO_sulfur_medlong      7.580e-02  4.708e-02   1.610 0.107662
## carbonylO_sulfur_long         1.359e-02  3.612e-02   0.376 0.706758
## carbonylO_sulfur_vlong        6.142e-02  2.524e-02   2.434 0.015073 *
## carbonylO_bbN_short           4.855e-02  3.912e-02   1.241 0.214816
## carbonylO_bbN_medshort        2.216e-02  2.618e-02   0.846 0.397472
## carbonylO_bbN_medlong        -8.055e-05  1.832e-02  -0.004 0.996492
## carbonylO_bbN_long            1.518e-02  1.363e-02   1.113 0.265831
## carbonylO_bbN_vlong           1.467e-02  8.777e-03   1.672 0.094806 .
## carbonylO_bbCA_short          1.718e-03  4.269e-02   0.040 0.967899
## carbonylO_bbCA_medshort       4.053e-04  3.231e-02   0.013 0.989995
## carbonylO_bbCA_medlong       -5.376e-03  2.259e-02  -0.238 0.811906
## carbonylO_bbCA_long          -1.103e-02  1.575e-02  -0.700 0.483987
## carbonylO_bbCA_vlong         -4.136e-03  9.322e-03  -0.444 0.657364
## carbonylO_bbC_short          -8.656e-03  4.136e-02  -0.209 0.834237
## carbonylO_bbC_medshort        4.886e-03  2.907e-02   0.168 0.866547
## carbonylO_bbC_medlong         3.075e-02  2.193e-02   1.402 0.161127
## carbonylO_bbC_long            1.856e-02  1.531e-02   1.212 0.225828
## carbonylO_bbC_vlong           8.858e-03  8.989e-03   0.985 0.324569
## carbonylO_bbO_short           3.751e-02  2.985e-02   1.257 0.209118
## carbonylO_bbO_medshort       -1.149e-02  2.371e-02  -0.484 0.628135
## carbonylO_bbO_medlong        -3.823e-03  1.848e-02  -0.207 0.836095
## carbonylO_bbO_long            1.536e-02  1.283e-02   1.197 0.231621
## carbonylO_bbO_vlong           6.138e-03  7.897e-03   0.777 0.437200
## carboxylO_carboxylO_medlong  -8.007e-03  6.290e-02  -0.127 0.898717
## carboxylO_carboxylO_long     -4.441e-02  5.606e-02  -0.792 0.428447
## carboxylO_carboxylO_vlong    -2.255e-02  3.716e-02  -0.607 0.544067
## carboxylO_sulfur_medshort     4.685e-02  4.130e-02   1.134 0.256878
## carboxylO_sulfur_medlong      8.328e-02  3.681e-02   2.263 0.023830 *
## carboxylO_sulfur_long         1.917e-02  3.155e-02   0.608 0.543457
## carboxylO_sulfur_vlong        1.800e-02  2.293e-02   0.785 0.432478
## carboxylO_bbN_vshort          1.849e-01  1.081e-01   1.711 0.087387 .
## carboxylO_bbN_medshort       -1.468e-01  4.512e-02  -3.254 0.001167 **
## carboxylO_bbN_medlong        -6.186e-02  3.283e-02  -1.884 0.059765 .
## carboxylO_bbN_long           -3.904e-03  2.243e-02  -0.174 0.861839
## carboxylO_bbN_vlong           1.835e-02  1.356e-02   1.354 0.176114
## carboxylO_bbCA_short          1.073e-01  8.025e-02   1.337 0.181315
## carboxylO_bbCA_medshort       1.296e-01  5.390e-02   2.404 0.016357 *
## carboxylO_bbCA_medlong        8.296e-02  3.910e-02   2.122 0.034026 *
## carboxylO_bbCA_long           6.112e-02  2.656e-02   2.301 0.021526 *
## carboxylO_bbCA_vlong          2.451e-02  1.454e-02   1.686 0.092036 .
## carboxylO_bbC_short          -6.581e-02  7.332e-02  -0.897 0.369640
## carboxylO_bbC_medshort        6.631e-03  5.064e-02   0.131 0.895829
## carboxylO_bbC_medlong        -7.590e-03  3.812e-02  -0.199 0.842195
## carboxylO_bbC_long            8.327e-03  2.790e-02   0.299 0.765369
## carboxylO_bbC_vlong           2.113e-02  1.552e-02   1.362 0.173564
## carboxylO_bbO_medshort        2.859e-02  3.407e-02   0.839 0.401495
## carboxylO_bbO_medlong        -1.642e-02  3.011e-02  -0.545 0.585635
## carboxylO_bbO_long            7.248e-03  2.077e-02   0.349 0.727156
## carboxylO_bbO_vlong           1.353e-02  1.271e-02   1.064 0.287439
## sulfur_sulfur_vlong           1.180e-01  3.789e-02   3.115 0.001879 **
## sulfur_bbN_short             -4.344e-02  4.167e-02  -1.042 0.297400
## sulfur_bbN_medshort          -6.464e-02  3.154e-02  -2.050 0.040597 *
## sulfur_bbN_medlong           -5.904e-02  2.442e-02  -2.417 0.015769 *
```

```
## sulfur_bbN_long             -2.186e-02  1.857e-02  -1.177 0.239265
## sulfur_bbN_vlong            -1.359e-02  1.184e-02  -1.148 0.251085
## sulfur_bbCA_short           -3.163e-02  5.805e-02  -0.545 0.585998
## sulfur_bbCA_medshort         3.897e-02  4.162e-02   0.936 0.349347
## sulfur_bbCA_medlong          6.761e-02  3.173e-02   2.131 0.033306 *
## sulfur_bbCA_long             3.951e-02  2.227e-02   1.774 0.076260 .
## sulfur_bbCA_vlong            2.210e-02  1.368e-02   1.616 0.106434
## sulfur_bbC_short             1.148e-01  5.601e-02   2.051 0.040512 *
## sulfur_bbC_medshort          7.773e-02  3.901e-02   1.992 0.046532 *
## sulfur_bbC_medlong           2.389e-03  3.169e-02   0.075 0.939927
## sulfur_bbC_long              1.518e-02  2.257e-02   0.672 0.501393
## sulfur_bbC_vlong            -9.172e-03  1.357e-02  -0.676 0.499099
## sulfur_bbO_short            -1.030e-01  3.500e-02  -2.944 0.003301 **
## sulfur_bbO_medshort         -6.275e-02  2.853e-02  -2.199 0.028030 *
## sulfur_bbO_medlong          -1.828e-02  2.190e-02  -0.834 0.404222
## sulfur_bbO_long             -4.082e-02  1.616e-02  -2.526 0.011670 *
## sulfur_bbO_vlong            -3.374e-02  1.107e-02  -3.049 0.002346 **
## bbN_bbN_medshort            -4.329e-02  2.133e-02  -2.030 0.042573 *
## bbN_bbN_medlong             -1.718e-02  1.703e-02  -1.009 0.313268
## bbN_bbN_long                -9.018e-03  1.293e-02  -0.697 0.485690
## bbN_bbN_vlong               -3.904e-03  7.482e-03  -0.522 0.601862
## bbN_bbCA_medshort            1.806e-02  1.785e-02   1.012 0.311883
## bbN_bbCA_medlong            -3.094e-03  1.363e-02  -0.227 0.820535
## bbN_bbCA_long                9.200e-03  9.429e-03   0.976 0.329384
## bbN_bbCA_vlong               1.290e-02  5.366e-03   2.404 0.016376 *
## bbN_bbC_vshort              -1.072e-01  3.692e-02  -2.903 0.003762 **
## bbN_bbC_short               -7.747e-02  2.636e-02  -2.938 0.003360 **
## bbN_bbC_medshort            -3.479e-02  2.058e-02  -1.691 0.091124 .
## bbN_bbC_medlong              9.680e-03  1.379e-02   0.702 0.482674
## bbN_bbC_long                 6.734e-03  9.565e-03   0.704 0.481537
## bbN_bbC_vlong               -7.194e-03  5.266e-03  -1.366 0.172128
## bbN_bbO_vshort               3.143e-02  2.816e-02   1.116 0.264553
## bbN_bbO_short                2.477e-02  2.236e-02   1.108 0.268112
## bbN_bbO_medshort            -3.046e-03  1.678e-02  -0.182 0.855988
## bbN_bbO_medlong              8.190e-03  1.234e-02   0.664 0.507079
## bbN_bbO_long                 1.234e-03  8.515e-03   0.145 0.884778
## bbN_bbO_vlong                9.494e-03  4.848e-03   1.958 0.050428 .
## bbCA_bbCA_medshort          -5.026e-02  2.888e-02  -1.741 0.081999 .
## bbCA_bbCA_medlong           -6.480e-02  2.021e-02  -3.206 0.001382 **
## bbCA_bbCA_long              -2.295e-02  1.404e-02  -1.635 0.102314
## bbCA_bbCA_vlong             -3.287e-02  8.362e-03  -3.931 8.91e-05 ***
## bbCA_bbC_medshort           -8.588e-03  1.753e-02  -0.490 0.624212
## bbCA_bbC_medlong            -2.798e-02  1.369e-02  -2.044 0.041114 *
## bbCA_bbC_long               -2.820e-02  9.970e-03  -2.828 0.004756 **
## bbCA_bbC_vlong              -1.979e-02  5.657e-03  -3.497 0.000486 ***
## bbCA_bbO_vshort             -3.262e-01  7.816e-02  -4.173 3.21e-05 ***
## bbCA_bbO_short               1.671e-02  2.463e-02   0.679 0.497490
## bbCA_bbO_medshort            9.930e-03  1.756e-02   0.565 0.571839
## bbCA_bbO_medlong             2.064e-02  1.302e-02   1.585 0.113299
## bbCA_bbO_long                1.021e-02  8.844e-03   1.154 0.248674
## bbCA_bbO_vlong               1.068e-03  4.994e-03   0.214 0.830709
## bbC_bbC_medshort             2.484e-03  2.603e-02   0.095 0.924006
## bbC_bbC_medlong              5.871e-03  2.010e-02   0.292 0.770295
## bbC_bbC_long                 2.498e-02  1.462e-02   1.709 0.087748 .
```

```
## bbC_bbC_vlong                 1.637e-02  7.993e-03   2.048 0.040791 *
## bbC_bbO_short                -3.426e-02  2.546e-02  -1.346 0.178550
## bbC_bbO_medshort             -3.634e-03  1.824e-02  -0.199 0.842141
## bbC_bbO_medlong              -1.258e-02  1.320e-02  -0.953 0.340795
## bbC_bbO_long                 -3.449e-03  9.268e-03  -0.372 0.709846
## bbC_bbO_vlong                -7.528e-03  5.471e-03  -1.376 0.169100
## bbO_bbO_vshort                1.194e-01  5.685e-02   2.100 0.035945 *
## bbO_bbO_short                -5.165e-02  2.760e-02  -1.871 0.061535 .
## bbO_bbO_medshort             -1.925e-02  2.047e-02  -0.941 0.347099
## bbO_bbO_medlong              -1.473e-02  1.484e-02  -0.992 0.321172
## bbO_bbO_long                 -1.013e-02  1.071e-02  -0.946 0.344532
## bbO_bbO_vlong                -2.199e-03  6.826e-03  -0.322 0.747425
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4161 on 1262 degrees of freedom
## Multiple R-squared:  0.9483, Adjusted R-squared:  0.9203
## F-statistic: 33.89 on 683 and 1262 DF,  p-value: < 2.2e-16
```

```
# remove the 2 variables which were perfectly correlated together and create a new dataframe
protein = subset(protein_train, select = -c(scArgN_bbC_medshort,scArgN_bbO_short) )

# correlation with accuracy

cor.list = sapply(protein[-1], function(x){cor(protein$accuracy,x)})

cor.tab  = data.frame(var = names(cor.list), r=cor.list,row.names = NULL)

cor.tab2 = subset(cor.tab, abs(r)>0.05)

vars = cor.tab2$var

vars = as.character(vars)
protein1 = protein[ c("accuracy",as.character(vars))]

# new linear model based on the new dataframe
mshort<- lm(accuracy ~., data=protein1)
summary(mshort)
```

```
##
## Call:
## lm(formula = accuracy ~ ., data = protein1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.58435 -0.25936 -0.00239  0.25409  1.79997
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -1.982e+00  2.875e+00  -0.689 0.490640
## angles                       -7.546e-03  5.325e-03  -1.417 0.156686
## carbonylC_carbonylC_medshort  4.325e-02  6.759e-02   0.640 0.522332
## carbonylC_carbonylC_medlong  -2.600e-02  7.115e-02  -0.365 0.714862
## carbonylC_carbonylC_long     -7.542e-02  6.349e-02  -1.188 0.235025
## carbonylC_carbonylC_vlong    -4.877e-02  4.130e-02  -1.181 0.237918
```

```
## carbonylC_carboxylC_medshort -9.643e-02  6.019e-02  -1.602 0.109381
## carbonylC_carboxylC_long      2.232e-02  5.381e-02   0.415 0.678345
## carbonylC_carboxylC_vlong    -1.963e-02  3.801e-02  -0.517 0.605563
## carbonylC_aliph1HC_medshort  -9.157e-02  3.949e-02  -2.319 0.020528 *
## carbonylC_aliph1HC_medlong   -1.950e-02  3.114e-02  -0.626 0.531125
## carbonylC_aliph1HC_long      -1.683e-02  2.274e-02  -0.740 0.459544
## carbonylC_aliph2HC_short     -1.955e-02  2.964e-02  -0.660 0.509637
## carbonylC_aliph2HC_medshort  -4.878e-02  2.101e-02  -2.322 0.020393 *
## carbonylC_aliph2HC_medlong   -1.842e-02  1.910e-02  -0.964 0.334988
## carbonylC_aliph2HC_long      -1.716e-02  1.410e-02  -1.217 0.223740
## carbonylC_aliph2HC_vlong     -3.495e-03  8.419e-03  -0.415 0.678156
## carbonylC_aliph3HC_short     -1.633e-02  4.024e-02  -0.406 0.684989
## carbonylC_aliph3HC_medshort   4.508e-02  3.317e-02   1.359 0.174308
## carbonylC_aliph3HC_medlong    4.288e-02  2.583e-02   1.660 0.097076 .
## carbonylC_aliph3HC_long       6.239e-03  1.836e-02   0.340 0.734105
## carbonylC_aromaticC_short     8.838e-02  4.103e-02   2.154 0.031381 *
## carbonylC_aromaticC_medshort -4.753e-03  1.853e-02  -0.257 0.797589
## carbonylC_aromaticC_long     -2.766e-03  1.083e-02  -0.255 0.798554
## carbonylC_aromaticC_vlong    -5.086e-03  9.352e-03  -0.544 0.586644
## carbonylC_scAGN_short        -4.891e-02  2.571e-02  -1.903 0.057297 .
## carbonylC_scAGN_medlong       2.118e-02  2.780e-02   0.762 0.446273
## carbonylC_scAGN_long          1.694e-02  3.519e-02   0.481 0.630440
## carbonylC_scAGN_vlong         7.988e-03  2.589e-02   0.308 0.757765
## carbonylC_scLysN_medlong      2.101e-02  6.022e-02   0.349 0.727165
## carbonylC_scLysN_long        -4.176e-02  5.695e-02  -0.733 0.463446
## carbonylC_scLysN_vlong       -1.353e-03  5.166e-02  -0.026 0.979108
## carbonylC_bbProN_medlong     -9.694e-02  3.327e-02  -2.914 0.003625 **
## carbonylC_hydroxylO_short    -8.982e-02  3.895e-02  -2.306 0.021256 *
## carbonylC_hydroxylO_medshort -8.692e-02  3.232e-02  -2.689 0.007244 **
## carbonylC_hydroxylO_medlong  -3.161e-02  2.703e-02  -1.169 0.242575
## carbonylC_hydroxylO_long      3.123e-02  1.744e-02   1.791 0.073535 .
## carbonylC_carbonylO_medshort -6.772e-03  3.488e-02  -0.194 0.846075
## carbonylC_carbonylO_medlong  -6.091e-02  4.374e-02  -1.393 0.163980
## carbonylC_carbonylO_long      7.142e-03  4.115e-02   0.174 0.862251
## carbonylC_carbonylO_vlong     4.694e-02  2.820e-02   1.665 0.096167 .
## carbonylC_carboxylO_short     2.274e-01  6.623e-02   3.434 0.000611 ***
## carbonylC_carboxylO_medshort  1.161e-01  5.740e-02   2.023 0.043290 *
## carbonylC_carboxylO_medlong   4.867e-02  4.797e-02   1.015 0.310488
## carbonylC_carboxylO_vlong     3.150e-02  2.270e-02   1.388 0.165441
## carbonylC_sulfur_short       -1.497e-01  4.763e-02  -3.142 0.001711 **
## carbonylC_sulfur_medlong     -4.009e-02  3.208e-02  -1.250 0.211636
## carbonylC_sulfur_long        -8.873e-02  3.712e-02  -2.390 0.016974 *
## carbonylC_sulfur_vlong       -7.259e-02  2.752e-02  -2.637 0.008442 **
## carbonylC_bbN_medshort        9.989e-03  1.826e-02   0.547 0.584355
## carbonylC_bbN_long            1.412e-02  7.867e-03   1.794 0.072971 .
## carbonylC_bbCA_medshort       6.917e-03  1.670e-02   0.414 0.678787
## carbonylC_bbCA_long          -3.470e-03  9.889e-03  -0.351 0.725735
## carbonylC_bbCA_vlong          3.790e-03  8.349e-03   0.454 0.649910
## carbonylC_bbC_medshort       -7.717e-02  2.050e-02  -3.765 0.000173 ***
## carbonylC_bbC_medlong        -9.271e-02  1.655e-02  -5.603 2.52e-08 ***
## carbonylC_bbC_long           -6.842e-02  1.351e-02  -5.064 4.64e-07 ***
## carbonylC_bbC_vlong          -5.053e-02  9.281e-03  -5.445 6.09e-08 ***
## carbonylC_bbO_medlong        -5.490e-03  1.281e-02  -0.429 0.668273
## carbonylC_bbO_long            3.289e-03  1.192e-02   0.276 0.782622
```

```
## carbonylC_bbO_vlong           8.196e-04  8.312e-03   0.099 0.921465
## carboxylC_carboxylC_vlong    -1.353e-01  5.924e-02  -2.283 0.022550 *
## carboxylC_aliph2HC_medshort  -1.065e-02  2.477e-02  -0.430 0.667220
## carboxylC_aliph2HC_vlong      2.115e-02  1.280e-02   1.652 0.098791 .
## carboxylC_aliph3HC_medlong   -6.470e-02  5.685e-02  -1.138 0.255282
## carboxylC_aliph3HC_vlong     -6.824e-02  3.309e-02  -2.062 0.039375 *
## carboxylC_aromaticC_long      1.326e-02  1.723e-02   0.770 0.441573
## carboxylC_scAGN_short        -1.261e-02  6.580e-02  -0.192 0.848092
## carboxylC_scAGN_vlong        -2.358e-02  3.014e-02  -0.782 0.434197
## carboxylC_scLysN_long         5.276e-02  7.806e-02   0.676 0.499223
## carboxylC_scLysN_vlong        9.842e-02  6.840e-02   1.439 0.150407
## carboxylC_scArgN_vlong        1.430e-02  4.233e-02   0.338 0.735577
## carboxylC_hydroxylO_medlong   4.470e-02  4.163e-02   1.074 0.283098
## carboxylC_carbonylO_medshort -2.885e-02  6.563e-02  -0.440 0.660290
## carboxylC_carbonylO_medlong  -8.079e-03  6.045e-02  -0.134 0.893694
## carboxylC_carboxylO_medlong  -5.465e-02  6.360e-02  -0.859 0.390278
## carboxylC_carboxylO_long     -4.202e-03  4.258e-02  -0.099 0.921402
## carboxylC_sulfur_medshort    -2.832e-03  5.719e-02  -0.050 0.960521
## carboxylC_sulfur_vlong       -2.704e-04  3.507e-02  -0.008 0.993850
## carboxylC_bbN_medshort        7.693e-02  5.506e-02   1.397 0.162596
## carboxylC_bbN_vlong           7.799e-03  1.524e-02   0.512 0.608893
## carboxylC_bbCA_medlong       -6.020e-03  3.563e-02  -0.169 0.865853
## carboxylC_bbCA_long           5.557e-03  3.082e-02   0.180 0.856951
## carboxylC_bbCA_vlong         -3.616e-02  2.048e-02  -1.766 0.077622 .
## carboxylC_bbC_medlong         7.897e-02  3.583e-02   2.204 0.027693 *
## carboxylC_bbC_long           -1.254e-03  3.101e-02  -0.040 0.967756
## carboxylC_bbC_vlong          -2.739e-02  1.996e-02  -1.372 0.170279
## carboxylC_bbO_medshort       -9.164e-02  4.930e-02  -1.859 0.063270 .
## carboxylC_bbO_medlong         8.095e-03  3.742e-02   0.216 0.828771
## carboxylC_bbO_long            4.039e-03  2.413e-02   0.167 0.867123
## aliph1HC_aliph1HC_medlong     3.609e-02  6.253e-02   0.577 0.563903
## aliph1HC_aliph1HC_long        1.196e-01  4.731e-02   2.529 0.011556 *
## aliph1HC_aliph1HC_vlong       6.460e-02  3.358e-02   1.924 0.054574 .
## aliph1HC_aliph2HC_medshort    2.118e-02  2.128e-02   0.995 0.319812
## aliph1HC_aliph2HC_medlong     3.951e-02  1.582e-02   2.497 0.012649 *
## aliph1HC_aliph2HC_long        2.838e-02  1.197e-02   2.371 0.017867 *
## aliph1HC_aliph2HC_vlong       1.529e-03  8.958e-03   0.171 0.864499
## aliph1HC_aliph3HC_medlong    -5.126e-02  2.176e-02  -2.355 0.018648 *
## aliph1HC_aliph3HC_long       -2.000e-02  2.150e-02  -0.930 0.352396
## aliph1HC_aliph3HC_vlong      -1.710e-02  1.512e-02  -1.131 0.258440
## aliph1HC_aromaticC_short      1.302e-01  2.894e-02   4.500 7.33e-06 ***
## aliph1HC_aromaticC_medshort   6.576e-02  1.721e-02   3.821 0.000139 ***
## aliph1HC_aromaticC_medlong    8.086e-02  1.478e-02   5.471 5.27e-08 ***
## aliph1HC_aromaticC_long       7.651e-02  1.215e-02   6.300 3.95e-10 ***
## aliph1HC_aromaticC_vlong      6.697e-02  1.078e-02   6.212 6.83e-10 ***
## aliph1HC_scAGN_short         -2.191e-02  4.126e-02  -0.531 0.595459
## aliph1HC_scAGN_medshort       5.018e-02  2.613e-02   1.920 0.055022 .
## aliph1HC_scAGN_long          -1.914e-02  1.872e-02  -1.022 0.306733
## aliph1HC_scAGN_vlong         -2.770e-02  1.598e-02  -1.733 0.083323 .
## aliph1HC_scArgN_long          5.487e-03  1.955e-01   0.028 0.977618
## aliph1HC_bbProN_medlong       3.883e-01  7.305e-02   5.316 1.23e-07 ***
## aliph1HC_bbProN_long          1.503e-01  5.273e-02   2.850 0.004438 **
## aliph1HC_bbProN_vlong         6.623e-02  3.565e-02   1.858 0.063374 .
## aliph1HC_hydroxylO_medshort   5.225e-02  3.579e-02   1.460 0.144499
```

```
## aliph1HC_hydroxylO_medlong    6.471e-02  3.256e-02   1.987 0.047076 *
## aliph1HC_hydroxylO_long      -7.705e-03  2.267e-02  -0.340 0.733961
## aliph1HC_carbonylO_medshort   2.787e-02  3.305e-02   0.843 0.399152
## aliph1HC_carbonylO_medlong   -6.184e-04  2.940e-02  -0.021 0.983223
## aliph1HC_carbonylO_long      -1.256e-03  2.547e-02  -0.049 0.960685
## aliph1HC_carbonylO_vlong     -1.688e-02  1.675e-02  -1.008 0.313589
## aliph1HC_carboxylO_vlong      1.936e-02  3.494e-02   0.554 0.579609
## aliph1HC_sulfur_medshort     -5.345e-02  3.166e-02  -1.689 0.091501 .
## aliph1HC_sulfur_medlong       5.378e-03  2.532e-02   0.212 0.831781
## aliph1HC_sulfur_long          1.595e-02  2.261e-02   0.706 0.480547
## aliph1HC_bbN_short           -1.303e-01  5.200e-02  -2.506 0.012318 *
## aliph1HC_bbN_medshort        -9.397e-02  3.033e-02  -3.099 0.001982 **
## aliph1HC_bbN_medlong         -4.137e-02  2.379e-02  -1.739 0.082259 .
## aliph1HC_bbN_long            -2.170e-02  1.763e-02  -1.231 0.218627
## aliph1HC_bbN_vlong           -3.796e-02  1.184e-02  -3.207 0.001373 **
## aliph1HC_bbCA_medshort        2.886e-02  3.611e-02   0.799 0.424205
## aliph1HC_bbCA_medlong        -3.428e-02  2.592e-02  -1.323 0.186141
## aliph1HC_bbCA_long           -8.891e-03  1.929e-02  -0.461 0.644970
## aliph1HC_bbCA_vlong          -2.702e-04  1.203e-02  -0.022 0.982079
## aliph1HC_bbC_medshort         4.870e-02  2.244e-02   2.171 0.030109 *
## aliph1HC_bbC_long            -1.507e-02  1.257e-02  -1.198 0.230977
## aliph1HC_bbC_vlong           -6.020e-03  1.004e-02  -0.600 0.548908
## aliph1HC_bbO_short           -4.364e-04  3.541e-02  -0.012 0.990169
## aliph1HC_bbO_medshort         1.549e-02  1.978e-02   0.783 0.433810
## aliph1HC_bbO_long            -2.849e-02  1.175e-02  -2.425 0.015436 *
## aliph1HC_bbO_vlong            1.533e-02  8.766e-03   1.749 0.080453 .
## aliph2HC_aliph2HC_medlong    -7.419e-03  6.558e-03  -1.131 0.258128
## aliph2HC_aliph2HC_long       -7.774e-03  5.140e-03  -1.512 0.130673
## aliph2HC_aliph3HC_vshort      3.218e-02  4.644e-02   0.693 0.488464
## aliph2HC_aliph3HC_short      -9.909e-03  1.826e-02  -0.543 0.587437
## aliph2HC_aliph3HC_medshort   -2.277e-02  1.237e-02  -1.841 0.065800 .
## aliph2HC_aliph3HC_medlong    -2.780e-02  1.011e-02  -2.749 0.006054 **
## aliph2HC_aliph3HC_long       -2.474e-02  8.038e-03  -3.078 0.002126 **
## aliph2HC_aliph3HC_vlong       1.543e-02  6.068e-03   2.542 0.011120 *
## aliph2HC_aromaticC_short      1.185e-02  9.782e-03   1.212 0.225865
## aliph2HC_aromaticC_medshort   2.764e-02  6.444e-03   4.289 1.92e-05 ***
## aliph2HC_aromaticC_medlong    1.006e-02  4.680e-03   2.150 0.031724 *
## aliph2HC_aromaticC_vlong      1.299e-02  3.238e-03   4.010 6.37e-05 ***
## aliph2HC_scAGN_vshort         9.807e-02  4.906e-02   1.999 0.045811 *
## aliph2HC_scAGN_short          4.449e-02  2.288e-02   1.944 0.052057 .
## aliph2HC_scAGN_medshort       3.158e-02  1.777e-02   1.777 0.075809 .
## aliph2HC_scAGN_medlong        2.816e-02  1.415e-02   1.990 0.046733 *
## aliph2HC_scAGN_long           4.994e-03  1.053e-02   0.474 0.635276
## aliph2HC_scAGN_vlong          1.572e-02  7.555e-03   2.081 0.037609 *
## aliph2HC_scLysN_medshort     -6.723e-02  3.671e-02  -1.831 0.067273 .
## aliph2HC_scLysN_long          2.473e-02  2.184e-02   1.132 0.257616
## aliph2HC_scLysN_vlong         6.094e-02  2.446e-02   2.492 0.012826 *
## aliph2HC_scArgN_medshort     -6.510e-01  3.694e-01  -1.762 0.078244 .
## aliph2HC_scArgN_medlong      -1.630e-01  1.071e-01  -1.523 0.128093
## aliph2HC_scArgN_long         -4.488e-02  4.844e-02  -0.927 0.354267
## aliph2HC_scArgN_vlong         4.152e-02  3.437e-02   1.208 0.227140
## aliph2HC_bbProN_medshort     -1.908e-02  3.096e-02  -0.616 0.537671
## aliph2HC_bbProN_medlong      -1.591e-02  2.071e-02  -0.768 0.442553
## aliph2HC_bbProN_long         -6.295e-03  1.593e-02  -0.395 0.692799
```

```
## aliph2HC_hydroxylO_medshort   1.958e-02  1.131e-02   1.732 0.083521 .
## aliph2HC_hydroxylO_medlong    3.538e-02  9.509e-03   3.720 0.000207 ***
## aliph2HC_hydroxylO_long       1.718e-02  7.109e-03   2.417 0.015760 *
## aliph2HC_carbonylO_vshort     8.739e-02  3.963e-02   2.205 0.027594 *
## aliph2HC_carbonylO_medshort  -1.854e-02  1.292e-02  -1.435 0.151388
## aliph2HC_carbonylO_medlong   -3.694e-05  1.217e-02  -0.003 0.997578
## aliph2HC_carbonylO_long      -2.234e-03  9.385e-03  -0.238 0.811871
## aliph2HC_carboxylO_medshort  -1.169e-03  1.879e-02  -0.062 0.950398
## aliph2HC_carboxylO_medlong   -6.314e-03  1.623e-02  -0.389 0.697295
## aliph2HC_carboxylO_long      -1.084e-02  1.187e-02  -0.913 0.361525
## aliph2HC_sulfur_vlong        -3.021e-03  9.273e-03  -0.326 0.744645
## aliph2HC_bbN_medshort        -4.432e-02  8.888e-03  -4.986 6.91e-07 ***
## aliph2HC_bbN_medlong          1.266e-02  5.527e-03   2.292 0.022070 *
## aliph2HC_bbN_vlong            2.024e-03  3.167e-03   0.639 0.522757
## aliph2HC_bbCA_short           5.610e-03  2.904e-02   0.193 0.846838
## aliph2HC_bbCA_medshort        1.873e-02  1.378e-02   1.359 0.174472
## aliph2HC_bbCA_medlong         2.950e-03  9.723e-03   0.303 0.761597
## aliph2HC_bbCA_long            3.719e-03  6.455e-03   0.576 0.564618
## aliph2HC_bbCA_vlong           2.525e-03  4.138e-03   0.610 0.541815
## aliph2HC_bbC_short            8.065e-03  2.729e-02   0.296 0.767616
## aliph2HC_bbC_medshort         9.788e-03  1.231e-02   0.795 0.426863
## aliph2HC_bbC_medlong         -6.939e-03  8.573e-03  -0.809 0.418441
## aliph2HC_bbC_long             1.120e-02  6.140e-03   1.825 0.068261 .
## aliph2HC_bbC_vlong            3.454e-03  4.198e-03   0.823 0.410785
## aliph2HC_bbO_vshort           2.481e-02  3.643e-02   0.681 0.495936
## aliph2HC_bbO_short            9.362e-04  1.186e-02   0.079 0.937109
## aliph2HC_bbO_vlong           -4.796e-03  3.135e-03  -1.530 0.126275
## aliph3HC_aliph3HC_medlong     2.268e-02  2.021e-02   1.122 0.261902
## aliph3HC_aliph3HC_long       -7.469e-04  1.901e-02  -0.039 0.968660
## aliph3HC_aliph3HC_vlong       2.036e-02  1.334e-02   1.527 0.127073
## aliph3HC_aromaticC_short      3.593e-03  1.180e-02   0.305 0.760744
## aliph3HC_aromaticC_medshort   1.389e-04  9.068e-03   0.015 0.987783
## aliph3HC_aromaticC_medlong   -1.920e-02  7.122e-03  -2.697 0.007085 **
## aliph3HC_aromaticC_long      -1.422e-02  5.734e-03  -2.481 0.013228 *
## aliph3HC_aromaticC_vlong     -1.227e-02  4.682e-03  -2.620 0.008875 **
## aliph3HC_scAGN_short         -4.486e-02  2.589e-02  -1.733 0.083335 .
## aliph3HC_scAGN_medshort       1.817e-03  2.192e-02   0.083 0.933938
## aliph3HC_scAGN_medlong       -9.201e-03  1.674e-02  -0.550 0.582527
## aliph3HC_scAGN_vlong         -8.975e-03  1.099e-02  -0.817 0.414339
## aliph3HC_scArgN_medlong       4.767e-02  1.035e-01   0.461 0.645194
## aliph3HC_scArgN_long          1.451e-01  1.002e-01   1.448 0.147811
## aliph3HC_scArgN_vlong         1.408e-01  7.307e-02   1.927 0.054115 .
## aliph3HC_bbProN_medlong      -3.723e-02  3.261e-02  -1.142 0.253776
## aliph3HC_bbProN_long         -1.506e-02  2.686e-02  -0.561 0.574971
## aliph3HC_bbProN_vlong         1.181e-02  2.040e-02   0.579 0.562687
## aliph3HC_hydroxylO_medshort  -7.520e-03  1.856e-02  -0.405 0.685361
## aliph3HC_hydroxylO_medlong    2.847e-02  1.701e-02   1.674 0.094384 .
## aliph3HC_hydroxylO_long       4.245e-02  1.369e-02   3.100 0.001971 **
## aliph3HC_hydroxylO_vlong      1.788e-02  1.038e-02   1.723 0.085155 .
## aliph3HC_carbonylO_short     -2.738e-03  2.968e-02  -0.092 0.926521
## aliph3HC_carbonylO_medshort  -7.117e-03  2.707e-02  -0.263 0.792661
## aliph3HC_carbonylO_medlong   -3.376e-03  2.359e-02  -0.143 0.886235
## aliph3HC_carbonylO_long      -1.079e-02  1.830e-02  -0.590 0.555365
## aliph3HC_carbonylO_vlong     -9.041e-03  1.202e-02  -0.752 0.452128
```

```
## aliph3HC_carboxylO_long      -7.230e-02  3.584e-02  -2.018 0.043815 *
## aliph3HC_carboxylO_vlong     -3.464e-02  2.055e-02  -1.686 0.092069 .
## aliph3HC_sulfur_short         6.669e-02  2.684e-02   2.484 0.013092 *
## aliph3HC_sulfur_medshort      5.770e-02  1.707e-02   3.380 0.000744 ***
## aliph3HC_sulfur_long          3.730e-03  1.450e-02   0.257 0.797088
## aliph3HC_bbN_short            8.664e-02  2.229e-02   3.887 0.000106 ***
## aliph3HC_bbN_medlong          9.364e-03  9.638e-03   0.971 0.331463
## aliph3HC_bbN_long             9.803e-03  8.930e-03   1.098 0.272489
## aliph3HC_bbN_vlong            2.662e-03  6.107e-03   0.436 0.662982
## aliph3HC_bbCA_short           7.078e-02  3.391e-02   2.087 0.037030 *
## aliph3HC_bbCA_medshort        1.516e-02  2.162e-02   0.701 0.483258
## aliph3HC_bbCA_medlong         4.078e-02  1.628e-02   2.505 0.012351 *
## aliph3HC_bbCA_long            2.616e-02  1.201e-02   2.178 0.029552 *
## aliph3HC_bbCA_vlong           2.057e-02  7.476e-03   2.751 0.006020 **
## aliph3HC_bbC_short            3.570e-02  2.946e-02   1.212 0.225775
## aliph3HC_bbC_medshort         3.260e-02  2.116e-02   1.541 0.123590
## aliph3HC_bbC_medlong         -6.402e-03  1.549e-02  -0.413 0.679498
## aliph3HC_bbC_long            -1.599e-02  1.063e-02  -1.504 0.132678
## aliph3HC_bbC_vlong            1.102e-02  7.336e-03   1.502 0.133256
## aliph3HC_bbO_short           -2.840e-02  1.799e-02  -1.578 0.114713
## aliph3HC_bbO_medshort         1.739e-04  1.223e-02   0.014 0.988652
## aliph3HC_bbO_long            -1.249e-03  7.286e-03  -0.171 0.863862
## aliph3HC_bbO_vlong            1.783e-03  5.623e-03   0.317 0.751234
## aromaticC_aromaticC_medlong  -5.065e-03  3.726e-03  -1.359 0.174290
## aromaticC_aromaticC_long     -3.457e-03  3.540e-03  -0.977 0.328950
## aromaticC_aromaticC_vlong    -7.847e-04  3.661e-03  -0.214 0.830282
## aromaticC_scAGN_short        -1.108e-02  2.104e-02  -0.526 0.598736
## aromaticC_scAGN_medshort     -2.026e-02  1.300e-02  -1.559 0.119184
## aromaticC_scAGN_medlong      -2.392e-02  1.024e-02  -2.336 0.019630 *
## aromaticC_scAGN_long          3.630e-03  9.320e-03   0.389 0.696995
## aromaticC_scAGN_vlong         4.473e-03  6.909e-03   0.647 0.517481
## aromaticC_scLysN_medlong     -4.905e-02  3.437e-02  -1.427 0.153714
## aromaticC_scLysN_long         1.444e-02  2.867e-02   0.504 0.614655
## aromaticC_scLysN_vlong       -6.485e-02  2.567e-02  -2.527 0.011627 *
## aromaticC_scArgN_medshort     5.630e-01  1.439e-01   3.912 9.57e-05 ***
## aromaticC_scArgN_medlong     -1.254e-01  5.412e-02  -2.316 0.020673 *
## aromaticC_scArgN_long        -1.640e-01  4.442e-02  -3.693 0.000230 ***
## aromaticC_bbProN_medlong     -5.092e-02  2.621e-02  -1.943 0.052266 .
## aromaticC_bbProN_long         1.333e-02  1.920e-02   0.694 0.487563
## aromaticC_hydroxylO_short    -2.734e-02  1.794e-02  -1.524 0.127702
## aromaticC_hydroxylO_medshort -4.403e-02  1.316e-02  -3.347 0.000838 ***
## aromaticC_hydroxylO_medlong  -4.848e-02  9.524e-03  -5.091 4.04e-07 ***
## aromaticC_hydroxylO_long     -2.465e-02  7.271e-03  -3.391 0.000716 ***
## aromaticC_carbonylO_short     1.378e-02  2.215e-02   0.622 0.533831
## aromaticC_carbonylO_medlong   1.777e-02  1.026e-02   1.731 0.083603 .
## aromaticC_carbonylO_long      8.448e-04  9.168e-03   0.092 0.926592
## aromaticC_carbonylO_vlong     6.371e-03  7.116e-03   0.895 0.370743
## aromaticC_carboxylO_medshort  1.153e-02  1.797e-02   0.642 0.521153
## aromaticC_sulfur_short       -3.867e-02  1.860e-02  -2.079 0.037753 *
## aromaticC_sulfur_medshort    -8.349e-03  1.423e-02  -0.587 0.557505
## aromaticC_sulfur_medlong      2.710e-02  1.413e-02   1.918 0.055364 .
## aromaticC_sulfur_long         4.583e-02  1.265e-02   3.624 0.000301 ***
## aromaticC_sulfur_vlong        1.959e-02  1.032e-02   1.897 0.057993 .
## aromaticC_bbN_short           1.093e-02  1.713e-02   0.638 0.523671
```

```
## aromaticC_bbN_medshort       -4.814e-03  7.533e-03  -0.639 0.522854
## aromaticC_bbN_long            7.899e-03  5.111e-03   1.545 0.122445
## aromaticC_bbN_vlong           7.870e-03  4.287e-03   1.836 0.066568 .
## aromaticC_bbCA_short          1.948e-02  2.143e-02   0.909 0.363545
## aromaticC_bbCA_medlong       -1.815e-03  6.721e-03  -0.270 0.787129
## aromaticC_bbCA_long           1.580e-03  5.848e-03   0.270 0.787049
## aromaticC_bbCA_vlong          9.277e-03  4.259e-03   2.178 0.029546 *
## aromaticC_bbC_medlong        -3.694e-03  7.756e-03  -0.476 0.633924
## aromaticC_bbC_long           -4.529e-04  6.460e-03  -0.070 0.944114
## aromaticC_bbC_vlong          -4.980e-03  4.613e-03  -1.080 0.280541
## aromaticC_bbO_vshort         -1.169e-01  4.933e-02  -2.370 0.017943 *
## aromaticC_bbO_short           1.314e-02  1.183e-02   1.111 0.266765
## aromaticC_bbO_medshort        2.144e-02  8.173e-03   2.623 0.008811 **
## aromaticC_bbO_medlong         1.573e-02  6.300e-03   2.497 0.012630 *
## aromaticC_bbO_long            5.500e-03  5.130e-03   1.072 0.283852
## aromaticC_bbO_vlong           1.256e-02  3.766e-03   3.334 0.000876 ***
## scAGN_scAGN_medshort          1.377e-02  3.675e-02   0.375 0.707970
## scAGN_scAGN_medlong           7.309e-03  4.363e-02   0.168 0.866976
## scAGN_scAGN_long              4.638e-02  4.315e-02   1.075 0.282594
## scAGN_scAGN_vlong            -1.926e-02  3.533e-02  -0.545 0.585668
## scAGN_scLysN_medshort         5.018e-02  5.321e-02   0.943 0.345831
## scAGN_scLysN_medlong          4.458e-03  5.225e-02   0.085 0.932024
## scAGN_scLysN_long            -1.376e-02  4.193e-02  -0.328 0.742791
## scAGN_bbProN_medlong         -2.084e-02  3.407e-02  -0.612 0.540833
## scAGN_bbProN_long             8.116e-03  2.814e-02   0.288 0.773086
## scAGN_bbProN_vlong           -1.457e-02  2.269e-02  -0.642 0.520904
## scAGN_hydroxylO_vshort        8.556e-02  4.551e-02   1.880 0.060287 .
## scAGN_hydroxylO_short         1.854e-02  3.397e-02   0.546 0.585335
## scAGN_hydroxylO_medshort      2.597e-02  2.676e-02   0.971 0.331953
## scAGN_hydroxylO_medlong       2.459e-02  1.964e-02   1.252 0.210881
## scAGN_carbonylO_medshort     -2.466e-02  3.090e-02  -0.798 0.424959
## scAGN_carbonylO_medlong       4.563e-02  3.367e-02   1.355 0.175612
## scAGN_carbonylO_long          3.399e-02  3.340e-02   1.017 0.309102
## scAGN_carbonylO_vlong         1.682e-02  2.456e-02   0.685 0.493595
## scAGN_carboxylO_vshort       -1.332e-01  6.170e-02  -2.159 0.030990 *
## scAGN_carboxylO_medshort      8.036e-03  4.296e-02   0.187 0.851633
## scAGN_carboxylO_medlong       1.982e-02  3.985e-02   0.497 0.618989
## scAGN_sulfur_short            6.536e-03  3.746e-02   0.174 0.861508
## scAGN_sulfur_medlong          1.526e-02  2.781e-02   0.549 0.583141
## scAGN_sulfur_long            -3.056e-03  2.880e-02  -0.106 0.915522
## scAGN_sulfur_vlong           -3.448e-03  2.336e-02  -0.148 0.882656
## scAGN_bbN_medshort            1.768e-02  1.552e-02   1.140 0.254653
## scAGN_bbN_medlong             2.734e-02  1.292e-02   2.116 0.034493 *
## scAGN_bbN_long                4.587e-02  1.121e-02   4.091 4.53e-05 ***
## scAGN_bbN_vlong               8.330e-03  7.628e-03   1.092 0.275030
## scAGN_bbCA_short             -5.261e-02  2.870e-02  -1.833 0.066940 .
## scAGN_bbCA_medlong            3.301e-02  1.246e-02   2.650 0.008144 **
## scAGN_bbCA_long               6.441e-04  1.222e-02   0.053 0.957977
## scAGN_bbCA_vlong             -1.008e-03  8.068e-03  -0.125 0.900547
## scAGN_bbC_short               2.448e-02  2.250e-02   1.088 0.276731
## scAGN_bbC_medlong             4.145e-04  1.282e-02   0.032 0.974208
## scAGN_bbC_long                1.304e-02  1.216e-02   1.073 0.283585
## scAGN_bbC_vlong              -7.109e-03  7.995e-03  -0.889 0.374097
## scAGN_bbO_short               3.597e-02  1.896e-02   1.897 0.058016 .
```

```
## scAGN_bbO_medlong             2.683e-02  1.212e-02   2.213 0.027064 *
## scAGN_bbO_long                2.109e-02  1.041e-02   2.025 0.043020 *
## scAGN_bbO_vlong              -1.206e-03  7.878e-03  -0.153 0.878390
## scLysN_carbonylO_vshort       6.742e-04  6.499e-02   0.010 0.991725
## scLysN_carbonylO_long        -6.064e-02  4.313e-02  -1.406 0.159940
## scLysN_carbonylO_vlong       -5.499e-02  3.975e-02  -1.383 0.166751
## scLysN_carboxylO_medlong     -5.497e-02  6.347e-02  -0.866 0.386537
## scLysN_carboxylO_long         6.865e-02  5.207e-02   1.319 0.187539
## scLysN_carboxylO_vlong        1.988e-03  4.912e-02   0.040 0.967728
## scLysN_bbN_medlong            3.409e-03  4.692e-02   0.073 0.942097
## scLysN_bbN_long              -2.793e-02  4.092e-02  -0.682 0.495115
## scLysN_bbN_vlong             -6.278e-02  2.694e-02  -2.330 0.019925 *
## scLysN_bbCA_medlong          -1.090e-01  4.820e-02  -2.260 0.023950 *
## scLysN_bbCA_long             -7.244e-02  4.255e-02  -1.703 0.088869 .
## scLysN_bbCA_vlong            -2.700e-02  2.967e-02  -0.910 0.362891
## scLysN_bbC_long               1.247e-01  3.200e-02   3.898 0.000102 ***
## scLysN_bbC_vlong             -2.278e-02  2.795e-02  -0.815 0.415062
## scLysN_bbO_medlong           -9.258e-04  3.587e-02  -0.026 0.979414
## scLysN_bbO_long              -4.885e-02  2.952e-02  -1.655 0.098221 .
## scLysN_bbO_vlong             -5.076e-02  2.564e-02  -1.980 0.047908 *
## scArgN_hydroxylO_medshort     3.159e-02  7.797e-02   0.405 0.685423
## scArgN_hydroxylO_medlong      2.184e-02  6.895e-02   0.317 0.751523
## scArgN_hydroxylO_vlong        1.566e-02  4.322e-02   0.362 0.717100
## scArgN_carboxylO_long         7.495e-03  4.158e-02   0.180 0.856960
## scArgN_bbN_medshort           2.212e-01  2.354e-01   0.940 0.347470
## scArgN_bbN_medlong            7.011e-01  2.421e-01   2.896 0.003835 **
## scArgN_bbN_long               2.909e-01  1.562e-01   1.863 0.062662 .
## scArgN_bbN_vlong              1.697e-02  7.056e-02   0.241 0.809930
## scArgN_bbCA_medshort          4.021e+00  8.542e-01   4.707 2.76e-06 ***
## scArgN_bbCA_medlong           2.903e+00  8.646e-01   3.358 0.000806 ***
## scArgN_bbCA_long              3.192e-01  1.540e-01   2.072 0.038397 *
## scArgN_bbCA_vlong            -1.562e-01  6.017e-02  -2.595 0.009544 **
## scArgN_bbC_short             -2.297e+00  4.731e-01  -4.854 1.34e-06 ***
## scArgN_bbC_medlong           -1.838e+00  4.133e-01  -4.447 9.39e-06 ***
## scArgN_bbC_long              -1.383e+00  3.811e-01  -3.630 0.000293 ***
## scArgN_bbC_vlong              2.023e-01  6.908e-02   2.928 0.003467 **
## scArgN_bbO_vshort             2.819e-01  1.766e-01   1.596 0.110721
## scArgN_bbO_medshort           1.310e+00  3.110e-01   4.213 2.68e-05 ***
## scArgN_bbO_medlong            1.026e+00  2.444e-01   4.197 2.87e-05 ***
## scArgN_bbO_long              -4.773e-01  1.464e-01  -3.260 0.001140 **
## scArgN_bbO_vlong             -3.842e-01  1.099e-01  -3.494 0.000489 ***
## bbProN_hydroxylO_medshort     1.279e-02  3.976e-02   0.322 0.747816
## bbProN_hydroxylO_medlong     -1.092e-02  2.996e-02  -0.365 0.715454
## bbProN_carbonylO_medlong      8.108e-02  3.262e-02   2.486 0.013040 *
## bbProN_carbonylO_vlong        1.185e-02  2.011e-02   0.589 0.555988
## bbProN_carboxylO_vlong        7.755e-02  3.758e-02   2.064 0.039226 *
## bbProN_sulfur_vlong          -2.905e-02  4.165e-02  -0.698 0.485576
## bbProN_bbN_medlong            1.118e-02  1.440e-02   0.777 0.437457
## bbProN_bbN_vlong             -1.854e-02  1.169e-02  -1.586 0.113006
## bbProN_bbCA_medshort          8.499e-02  3.375e-02   2.518 0.011900 *
## bbProN_bbCA_medlong           3.637e-02  2.260e-02   1.609 0.107813
## bbProN_bbCA_vlong             2.291e-04  1.214e-02   0.019 0.984943
## bbProN_bbC_vlong             -5.721e-03  1.202e-02  -0.476 0.634162
## bbProN_bbO_short             -2.751e-02  3.414e-02  -0.806 0.420482
```

```
## bbProN_bbO_medshort          -1.264e-02  2.906e-02  -0.435 0.663780
## bbProN_bbO_long               3.030e-02  1.567e-02   1.934 0.053334 .
## bbProN_bbO_vlong              3.499e-02  1.281e-02   2.731 0.006393 **
## hydroxylO_hydroxylO_short     2.022e-02  4.127e-02   0.490 0.624185
## hydroxylO_hydroxylO_medshort  1.054e-02  2.936e-02   0.359 0.719810
## hydroxylO_hydroxylO_medlong   2.108e-02  2.663e-02   0.792 0.428633
## hydroxylO_hydroxylO_vlong     1.238e-02  1.425e-02   0.869 0.385179
## hydroxylO_carbonylO_medshort  2.729e-02  1.886e-02   1.447 0.148157
## hydroxylO_carbonylO_medlong   2.283e-02  1.752e-02   1.303 0.192691
## hydroxylO_carboxylO_medshort -7.069e-02  3.444e-02  -2.052 0.040308 *
## hydroxylO_carboxylO_long     -3.113e-02  2.790e-02  -1.116 0.264794
## hydroxylO_carboxylO_vlong    -7.170e-03  2.141e-02  -0.335 0.737719
## hydroxylO_bbN_short           3.924e-02  1.986e-02   1.975 0.048405 *
## hydroxylO_bbN_long            6.671e-03  7.744e-03   0.861 0.389130
## hydroxylO_bbN_vlong           1.005e-02  6.390e-03   1.573 0.116043
## hydroxylO_bbCA_medshort      -8.770e-03  1.305e-02  -0.672 0.501608
## hydroxylO_bbCA_long           1.244e-03  7.609e-03   0.163 0.870151
## hydroxylO_bbC_short          -1.830e-02  2.441e-02  -0.750 0.453590
## hydroxylO_bbC_medshort        2.983e-02  1.570e-02   1.900 0.057639 .
## hydroxylO_bbC_medlong         3.758e-02  1.192e-02   3.153 0.001647 **
## hydroxylO_bbC_long            6.534e-03  8.450e-03   0.773 0.439497
## hydroxylO_bbO_vshort          5.231e-02  3.071e-02   1.704 0.088654 .
## hydroxylO_bbO_medshort       -2.639e-02  1.358e-02  -1.944 0.052123 .
## hydroxylO_bbO_medlong        -1.969e-02  1.040e-02  -1.892 0.058627 .
## hydroxylO_bbO_vlong           2.643e-04  5.929e-03   0.045 0.964453
## carbonylO_carbonylO_medshort  3.022e-02  4.899e-02   0.617 0.537436
## carbonylO_carbonylO_medlong   6.986e-02  5.269e-02   1.326 0.185053
## carbonylO_carbonylO_long      7.032e-02  5.254e-02   1.338 0.180993
## carbonylO_carbonylO_vlong     1.828e-02  3.981e-02   0.459 0.646285
## carbonylO_carboxylO_medshort -1.004e-02  4.983e-02  -0.201 0.840366
## carbonylO_carboxylO_medlong  -1.211e-02  4.472e-02  -0.271 0.786579
## carbonylO_carboxylO_long     -1.556e-02  3.386e-02  -0.460 0.645904
## carbonylO_sulfur_short        6.964e-02  4.723e-02   1.474 0.140599
## carbonylO_sulfur_medshort     9.225e-02  3.613e-02   2.553 0.010771 *
## carbonylO_sulfur_medlong      8.578e-02  3.379e-02   2.539 0.011237 *
## carbonylO_sulfur_long         3.901e-02  3.114e-02   1.253 0.210543
## carbonylO_sulfur_vlong        7.409e-02  2.397e-02   3.092 0.002030 **
## carbonylO_bbN_long            9.841e-03  8.129e-03   1.211 0.226202
## carbonylO_bbN_vlong           1.127e-02  7.081e-03   1.591 0.111724
## carbonylO_bbCA_long          -9.234e-03  9.074e-03  -1.018 0.309040
## carbonylO_bbCA_vlong         -7.705e-04  7.489e-03  -0.103 0.918074
## carbonylO_bbC_medshort        4.337e-03  1.430e-02   0.303 0.761666
## carbonylO_bbC_long           -4.713e-03  9.300e-03  -0.507 0.612423
## carbonylO_bbC_vlong          -9.110e-04  7.359e-03  -0.124 0.901493
## carbonylO_bbO_medshort       -7.965e-03  1.260e-02  -0.632 0.527367
## carbonylO_bbO_long            2.886e-02  9.200e-03   3.137 0.001740 **
## carbonylO_bbO_vlong           6.576e-03  6.866e-03   0.958 0.338341
## carboxylO_carboxylO_medlong  -4.811e-02  5.363e-02  -0.897 0.369786
## carboxylO_carboxylO_long      3.094e-02  3.691e-02   0.838 0.402021
## carboxylO_carboxylO_vlong     1.620e-03  2.760e-02   0.059 0.953222
## carboxylO_sulfur_medshort     4.971e-02  4.054e-02   1.226 0.220318
## carboxylO_sulfur_medlong      6.837e-02  3.642e-02   1.877 0.060704 .
## carboxylO_sulfur_long         1.881e-03  3.129e-02   0.060 0.952069
## carboxylO_sulfur_vlong        4.478e-02  2.342e-02   1.912 0.056102 .
```

```
## carboxylO_bbN_vshort          1.724e-01  8.694e-02   1.983 0.047507 *
## carboxylO_bbN_medshort       -1.105e-01  3.491e-02  -3.165 0.001584 **
## carboxylO_bbN_long            1.100e-02  1.673e-02   0.657 0.511107
## carboxylO_bbN_vlong           2.684e-02  1.132e-02   2.371 0.017878 *
## carboxylO_bbCA_medshort       5.866e-02  3.523e-02   1.665 0.096152 .
## carboxylO_bbCA_medlong        2.407e-02  3.374e-02   0.713 0.475659
## carboxylO_bbCA_long           1.545e-02  2.491e-02   0.621 0.535022
## carboxylO_bbCA_vlong          1.559e-02  1.442e-02   1.081 0.279660
## carboxylO_bbC_medshort       -1.383e-02  3.634e-02  -0.381 0.703517
## carboxylO_bbC_medlong        -1.512e-02  3.168e-02  -0.477 0.633142
## carboxylO_bbC_long            4.330e-03  2.592e-02   0.167 0.867378
## carboxylO_bbC_vlong           2.607e-02  1.490e-02   1.750 0.080354 .
## carboxylO_bbO_medshort        1.014e-02  3.152e-02   0.322 0.747684
## carboxylO_bbO_medlong        -1.799e-02  2.774e-02  -0.649 0.516570
## carboxylO_bbO_long           -5.002e-03  1.832e-02  -0.273 0.784912
## carboxylO_bbO_vlong          -1.173e-03  1.121e-02  -0.105 0.916690
## sulfur_sulfur_vlong           8.128e-02  3.862e-02   2.105 0.035505 *
## sulfur_bbN_short              8.684e-03  3.958e-02   0.219 0.826374
## sulfur_bbN_medshort           1.051e-02  2.816e-02   0.373 0.709097
## sulfur_bbN_medlong           -2.773e-02  1.993e-02  -1.392 0.164228
## sulfur_bbN_long               9.385e-03  1.360e-02   0.690 0.490284
## sulfur_bbN_vlong             -5.408e-03  9.942e-03  -0.544 0.586527
## sulfur_bbCA_short            -4.916e-02  4.620e-02  -1.064 0.287499
## sulfur_bbCA_medshort          2.986e-02  3.058e-02   0.977 0.328905
## sulfur_bbCA_medlong           2.822e-02  2.168e-02   1.302 0.193205
## sulfur_bbCA_long              1.258e-02  1.417e-02   0.888 0.374683
## sulfur_bbC_medlong           -4.727e-02  1.279e-02  -3.696 0.000228 ***
## sulfur_bbC_vlong             -2.722e-02  8.904e-03  -3.057 0.002279 **
## sulfur_bbO_medlong            1.721e-02  1.262e-02   1.364 0.172733
## sulfur_bbO_long              -2.219e-02  1.084e-02  -2.047 0.040795 *
## sulfur_bbO_vlong             -2.445e-02  9.365e-03  -2.611 0.009126 **
## bbN_bbN_medshort             -3.597e-02  1.600e-02  -2.248 0.024736 *
## bbN_bbN_medlong              -1.062e-02  9.843e-03  -1.079 0.280823
## bbN_bbN_vlong                -9.367e-03  5.466e-03  -1.714 0.086763 .
## bbN_bbCA_medshort            -1.494e-02  1.460e-02  -1.023 0.306390
## bbN_bbCA_medlong             -4.234e-02  9.794e-03  -4.323 1.64e-05 ***
## bbN_bbCA_long                -2.017e-02  5.871e-03  -3.436 0.000608 ***
## bbN_bbC_vshort               -3.211e-02  2.965e-02  -1.083 0.279070
## bbN_bbC_short                -1.925e-02  1.327e-02  -1.451 0.147049
## bbN_bbC_medlong               2.018e-02  8.469e-03   2.383 0.017287 *
## bbN_bbC_long                  1.897e-02  5.736e-03   3.306 0.000968 ***
## bbN_bbO_vshort               -1.109e-02  2.560e-02  -0.433 0.664934
## bbN_bbO_short                -1.934e-02  1.904e-02  -1.016 0.309804
## bbN_bbO_medshort             -3.694e-02  1.502e-02  -2.460 0.014000 *
## bbN_bbO_medlong              -1.512e-02  1.075e-02  -1.407 0.159499
## bbN_bbO_long                 -1.579e-02  6.914e-03  -2.284 0.022543 *
## bbN_bbO_vlong                 8.625e-06  4.190e-03   0.002 0.998358
## bbCA_bbCA_medshort            1.599e-03  1.934e-02   0.083 0.934105
## bbCA_bbCA_long                2.348e-02  6.328e-03   3.711 0.000214 ***
## bbCA_bbC_medshort             7.677e-03  1.438e-02   0.534 0.593588
## bbCA_bbC_medlong             -8.794e-03  8.042e-03  -1.093 0.274391
## bbCA_bbC_vlong               -1.104e-02  3.874e-03  -2.851 0.004424 **
## bbCA_bbO_vshort              -3.559e-01  7.625e-02  -4.667 3.33e-06 ***
## bbCA_bbO_short                2.164e-03  1.463e-02   0.148 0.882458
```

```
## bbCA_bbO_medlong              8.190e-05  5.749e-03   0.014 0.988635
## bbC_bbC_medshort             -4.402e-02  2.471e-02  -1.781 0.075120 .
## bbC_bbC_medlong              -3.190e-02  1.851e-02  -1.723 0.085126 .
## bbC_bbC_long                 -4.163e-03  1.328e-02  -0.313 0.754007
## bbC_bbC_vlong                -1.087e-03  7.367e-03  -0.148 0.882667
## bbC_bbO_short                -2.569e-02  2.052e-02  -1.252 0.210727
## bbC_bbO_medshort              3.065e-04  1.267e-02   0.024 0.980709
## bbC_bbO_medlong              -4.688e-03  8.866e-03  -0.529 0.597053
## bbC_bbO_long                  5.142e-03  6.007e-03   0.856 0.392199
## bbO_bbO_vshort                1.451e-01  5.098e-02   2.845 0.004500 **
## bbO_bbO_short                -3.026e-02  1.634e-02  -1.852 0.064202 .
## bbO_bbO_medlong              -3.801e-04  8.468e-03  -0.045 0.964201
## bbO_bbO_long                 -8.544e-03  7.293e-03  -1.172 0.241584
## bbO_bbO_vlong                -4.752e-03  5.739e-03  -0.828 0.407754
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4604 on 1440 degrees of freedom
## Multiple R-squared:  0.9278, Adjusted R-squared:  0.9024
## F-statistic: 36.63 on 505 and 1440 DF,  p-value: < 2.2e-16
```

```
max(vif(mshort))
```

```
## [1] 662.2596
```

```
# while loop to go through all the vifs, and only keep those variables with a vif < 10
cutoff <- 10
flag <- TRUE
while(flag){
  fit <- lm(accuracy ~ ., data=protein1)
  vfit <- vif(fit)
  if(max(vfit) > cutoff){
    protein1 <- subset(protein1, select = -get(names(vfit)[which.max(vfit)]))
  } else {
    flag <- FALSE
   }
}

#print(fit)
#print(vfit)


##MODEL SELECTION
set.seed(20680907)
library(MASS)

#start with one train/validation split
N <- nrow(protein1)
trainInd <- sample(1:N, round(N*0.8), replace=F)
trainSet <- protein1[trainInd,]
validSet <- protein1[-trainInd,]

# Full model and empty model with just intercept
full <- lm(accuracy ~ ., data = trainSet)
empty <- lm(accuracy ~ 1, data =trainSet)
```

```r
# Stepwise forward with BIC
#m1<-stepAIC(object = empty, scope = list(upper = full, lower = empty), direction = "forward", k = log(

#save(m1, file= "m1.rda")
load("m1.rda")
summary(m1)
```

```
##
## Call:
## lm(formula = accuracy ~ aliph1HC_aliph2HC_long + scLysN_bbC_vlong +
##     aliph2HC_bbN_medshort + aliph1HC_aromaticC_medshort + carbonylC_aromaticC_short +
##     aromaticC_hydroxylO_medlong + aliph1HC_aromaticC_vlong +
##     bbC_bbO_short + aliph1HC_aliph2HC_medlong + sulfur_bbC_vlong +
##     aromaticC_hydroxylO_long + sulfur_bbC_medlong + aliph1HC_aromaticC_long +
##     carboxylC_bbN_vlong + aliph1HC_bbC_long + scAGN_bbN_long +
##     aliph1HC_aliph1HC_vlong + bbN_bbCA_medlong + aromaticC_sulfur_long +
##     aliph3HC_bbN_short + aliph1HC_aromaticC_medlong + aliph2HC_aromaticC_medshort +
##     aliph1HC_bbO_long + aliph1HC_aromaticC_short + aliph2HC_aromaticC_vlong +
##     carboxylO_bbN_long + aliph3HC_bbC_vlong + bbN_bbCA_long +
##     scArgN_bbO_medlong + carboxylC_scLysN_vlong + carbonylC_aromaticC_long +
##     aliph1HC_scArgN_long + aliph2HC_bbN_medlong + aromaticC_aromaticC_vlong +
##     aliph3HC_hydroxylO_long + scAGN_carbonylO_medshort + carbonylC_bbC_medlong +
##     scAGN_bbN_medlong + aliph1HC_bbC_medshort + bbN_bbN_medshort +
##     carbonylC_bbProN_medlong + aromaticC_hydroxylO_medshort +
##     aromaticC_bbO_vshort + aliph1HC_aliph3HC_medlong + aliph1HC_aliph1HC_long +
##     carboxylO_bbO_vlong + hydroxylO_carbonylO_medlong + carboxylO_bbO_long +
##     bbN_bbO_vlong + aliph1HC_bbProN_medlong + bbO_bbO_long +
##     aliph1HC_bbN_medshort + hydroxylO_bbC_medlong + aromaticC_bbO_medlong +
##     scLysN_carboxylO_long + bbN_bbC_vshort + carboxylC_carboxylC_vlong +
##     bbCA_bbC_vlong + bbO_bbO_vshort + bbN_bbN_medlong + bbProN_carboxylO_vlong +
##     aliph1HC_bbN_vlong + carbonylC_scAGN_long + bbC_bbO_medshort +
##     aliph2HC_carboxylO_medlong + carbonylC_sulfur_short + scArgN_carboxylO_long +
##     aromaticC_bbN_short + carbonylC_aliph2HC_short + bbCA_bbO_vshort +
##     carbonylC_carboxylC_long + aliph1HC_bbO_vlong + aliph2HC_scArgN_vlong +
##     aromaticC_bbO_vlong + carbonylO_sulfur_medshort + carboxylO_sulfur_vlong +
##     aliph2HC_aliph3HC_vlong + aliph1HC_scAGN_long + aliph3HC_bbCA_short +
##     bbProN_bbCA_medshort, data = trainSet)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.93817 -0.32832  0.00956  0.32344  1.83944
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    2.429261   0.417619   5.817 7.34e-09 ***
## aliph1HC_aliph2HC_long         0.043373   0.008094   5.358 9.73e-08 ***
## scLysN_bbC_vlong              -0.108361   0.018656  -5.808 7.71e-09 ***
## aliph2HC_bbN_medshort         -0.047947   0.006378  -7.517 9.66e-14 ***
## aliph1HC_aromaticC_medshort    0.111254   0.011749   9.469  < 2e-16 ***
## carbonylC_aromaticC_short      0.157372   0.031089   5.062 4.67e-07 ***
## aromaticC_hydroxylO_medlong   -0.050777   0.007349  -6.909 7.22e-12 ***
## aliph1HC_aromaticC_vlong       0.087069   0.008601  10.123  < 2e-16 ***
## bbC_bbO_short                 -0.076757   0.010714  -7.164 1.23e-12 ***
```

```
## aliph1HC_aliph2HC_medlong     0.036946   0.009141   4.042 5.57e-05 ***
## sulfur_bbC_vlong             -0.039380   0.007390  -5.329 1.14e-07 ***
## aromaticC_hydroxylO_long     -0.039516   0.005257  -7.517 9.67e-14 ***
## sulfur_bbC_medlong           -0.057085   0.010735  -5.318 1.21e-07 ***
## aliph1HC_aromaticC_long       0.082276   0.008214  10.017  < 2e-16 ***
## carboxylC_bbN_vlong           0.016195   0.013338   1.214 0.224867
## aliph1HC_bbC_long            -0.014981   0.009485  -1.579 0.114445
## scAGN_bbN_long                0.057561   0.006421   8.964  < 2e-16 ***
## aliph1HC_aliph1HC_vlong       0.122734   0.025266   4.858 1.31e-06 ***
## bbN_bbCA_medlong             -0.038316   0.005031  -7.616 4.63e-14 ***
## aromaticC_sulfur_long         0.063347   0.010931   5.795 8.32e-09 ***
## aliph3HC_bbN_short            0.108719   0.020437   5.320 1.20e-07 ***
## aliph1HC_aromaticC_medlong    0.079468   0.009307   8.539  < 2e-16 ***
## aliph2HC_aromaticC_medshort   0.028099   0.004533   6.199 7.38e-10 ***
## aliph1HC_bbO_long            -0.042737   0.010685  -4.000 6.66e-05 ***
## aliph1HC_aromaticC_short      0.124749   0.022292   5.596 2.61e-08 ***
## aliph2HC_aromaticC_vlong      0.018180   0.002710   6.709 2.79e-11 ***
## carboxylO_bbN_long            0.020765   0.012329   1.684 0.092357 .
## aliph3HC_bbC_vlong            0.024201   0.004607   5.253 1.71e-07 ***
## bbN_bbCA_long                -0.019315   0.004233  -4.563 5.45e-06 ***
## scArgN_bbO_medlong            0.410671   0.086244   4.762 2.11e-06 ***
## carboxylC_scLysN_vlong        0.115932   0.042965   2.698 0.007048 **
## carbonylC_aromaticC_long      0.037363   0.007614   4.907 1.03e-06 ***
## aliph1HC_scArgN_long          0.751719   0.092540   8.123 9.50e-16 ***
## aliph2HC_bbN_medlong          0.025140   0.003903   6.441 1.60e-10 ***
## aromaticC_aromaticC_vlong    -0.012045   0.002539  -4.745 2.29e-06 ***
## aliph3HC_hydroxylO_long       0.048201   0.011295   4.267 2.10e-05 ***
## scAGN_carbonylO_medshort     -0.086765   0.017806  -4.873 1.22e-06 ***
## carbonylC_bbC_medlong        -0.040960   0.009391  -4.361 1.38e-05 ***
## scAGN_bbN_medlong             0.035249   0.007995   4.409 1.11e-05 ***
## aliph1HC_bbC_medshort         0.062420   0.016372   3.813 0.000143 ***
## bbN_bbN_medshort             -0.092115   0.011234  -8.200 5.19e-16 ***
## carbonylC_bbProN_medlong     -0.137199   0.031004  -4.425 1.03e-05 ***
## aromaticC_hydroxylO_medshort -0.049315   0.010309  -4.784 1.89e-06 ***
## aromaticC_bbO_vshort         -0.248490   0.039814  -6.241 5.66e-10 ***
## aliph1HC_aliph3HC_medlong    -0.077008   0.015992  -4.815 1.62e-06 ***
## aliph1HC_aliph1HC_long        0.240857   0.036889   6.529 9.07e-11 ***
## carboxylO_bbO_vlong           0.029186   0.009230   3.162 0.001598 **
## hydroxylO_carbonylO_medlong   0.057454   0.012450   4.615 4.28e-06 ***
## carboxylO_bbO_long            0.046036   0.011332   4.063 5.11e-05 ***
## bbN_bbO_vlong                 0.013521   0.003031   4.461 8.78e-06 ***
## aliph1HC_bbProN_medlong       0.270172   0.052267   5.169 2.68e-07 ***
## bbO_bbO_long                 -0.022575   0.005244  -4.305 1.78e-05 ***
## aliph1HC_bbN_medshort        -0.065046   0.017494  -3.718 0.000208 ***
## hydroxylO_bbC_medlong         0.038199   0.008149   4.688 3.02e-06 ***
## aromaticC_bbO_medlong         0.013381   0.004398   3.043 0.002387 **
## scLysN_carboxylO_long         0.150475   0.031321   4.804 1.71e-06 ***
## bbN_bbC_vshort               -0.067528   0.027020  -2.499 0.012556 *
## carboxylC_carboxylC_vlong    -0.168787   0.039246  -4.301 1.81e-05 ***
## bbCA_bbC_vlong               -0.014537   0.003335  -4.359 1.40e-05 ***
## bbO_bbO_vshort                0.159233   0.048278   3.298 0.000996 ***
## bbN_bbN_medlong              -0.024165   0.007943  -3.042 0.002391 **
## bbProN_carboxylO_vlong        0.128668   0.039441   3.262 0.001131 **
## aliph1HC_bbN_vlong           -0.029483   0.008453  -3.488 0.000501 ***
```

```
## carbonylC_scAGN_long          0.084848   0.021078   4.025 5.98e-05 ***
## bbC_bbO_medshort             -0.023355   0.005323  -4.387 1.23e-05 ***
## aliph2HC_carboxylO_medlong    0.051074   0.012777   3.997 6.73e-05 ***
## carbonylC_sulfur_short       -0.133928   0.038736  -3.457 0.000561 ***
## scArgN_carboxylO_long        -0.117905   0.024640  -4.785 1.88e-06 ***
## aromaticC_bbN_short           0.035444   0.012515   2.832 0.004687 **
## carbonylC_aliph2HC_short      0.065691   0.021558   3.047 0.002351 **
## bbCA_bbO_vshort              -0.333337   0.073821  -4.515 6.82e-06 ***
## carbonylC_carboxylC_long      0.161798   0.047109   3.435 0.000610 ***
## aliph1HC_bbO_vlong            0.028694   0.007831   3.664 0.000257 ***
## aliph2HC_scArgN_vlong         0.074301   0.024382   3.047 0.002350 **
## aromaticC_bbO_vlong           0.008995   0.002627   3.424 0.000635 ***
## carbonylO_sulfur_medshort     0.090116   0.026583   3.390 0.000718 ***
## carboxylO_sulfur_vlong        0.067090   0.022569   2.973 0.003000 **
## aliph2HC_aliph3HC_vlong       0.015608   0.004879   3.199 0.001408 **
## aliph1HC_scAGN_long          -0.049525   0.017219  -2.876 0.004082 **
## aliph3HC_bbCA_short           0.067754   0.025361   2.672 0.007633 **
## bbProN_bbCA_medshort          0.065279   0.024596   2.654 0.008039 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5409 on 1476 degrees of freedom
## Multiple R-squared:  0.8725, Adjusted R-squared:  0.8656
## F-statistic: 126.3 on 80 and 1476 DF,  p-value: < 2.2e-16
```

```
BIC(m1)
```

```
## [1] 3024.507
```

```
pred1 <- predict(m1, newdata = validSet)
sqrt(mean((validSet$accuracy - pred1)^2)) # RMSE on validation
```

```
## [1] 0.6188801
```

```
sqrt(mean(m1$residuals^2)) # RMSE on train
```

```
## [1] 0.526641
```

```
# forward stepwise again, with a larger L0 penalty (e.g., twice the usual BIC penalty)
#m2 <- stepAIC(object = empty, scope = list(upper = full, lower = empty), direction = "forward", k = 2*

#save(m2, file= "m2.rda")
load("m2.rda")
summary(m2)
```

```
##
## Call:
## lm(formula = accuracy ~ aliph1HC_aliph2HC_long + scLysN_bbC_vlong +
##     aliph2HC_bbN_medshort + aliph1HC_aromaticC_medshort + carbonylC_aromaticC_short +
##     aromaticC_hydroxylO_medlong + aliph1HC_aromaticC_vlong +
##     bbC_bbO_short + aliph1HC_aliph2HC_medlong + sulfur_bbC_vlong +
##     aromaticC_hydroxylO_long + sulfur_bbC_medlong + aliph1HC_aromaticC_long +
##     carboxylC_bbN_vlong + aliph1HC_bbC_long + scAGN_bbN_long +
##     aliph1HC_aliph1HC_vlong + bbN_bbCA_medlong + aromaticC_sulfur_long +
##     aliph3HC_bbN_short + aliph1HC_aromaticC_medlong + aliph2HC_aromaticC_medshort +
##     aliph1HC_bbO_long + aliph1HC_aromaticC_short + aliph2HC_aromaticC_vlong +
##     carboxylO_bbN_long + aliph3HC_bbC_vlong + bbN_bbCA_long +
```

```
##     scArgN_bbO_medlong + carboxylC_scLysN_vlong + carbonylC_aromaticC_long +
##     aliph1HC_scArgN_long + aliph2HC_bbN_medlong + aromaticC_aromaticC_vlong +
##     aliph3HC_hydroxylO_long + scAGN_carbonylO_medshort + carbonylC_bbC_medlong +
##     scAGN_bbN_medlong, data = trainSet)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4759 -0.3778  0.0149  0.3848  2.2892
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   2.533948   0.405186   6.254 5.20e-10 ***
## aliph1HC_aliph2HC_long        0.061357   0.008862   6.924 6.48e-12 ***
## scLysN_bbC_vlong             -0.159296   0.018309  -8.700  < 2e-16 ***
## aliph2HC_bbN_medshort        -0.040121   0.006850  -5.857 5.77e-09 ***
## aliph1HC_aromaticC_medshort   0.135201   0.012517  10.801  < 2e-16 ***
## carbonylC_aromaticC_short     0.226118   0.034135   6.624 4.83e-11 ***
## aromaticC_hydroxylO_medlong -0.072865   0.007921  -9.199  < 2e-16 ***
## aliph1HC_aromaticC_vlong      0.125104   0.008724  14.340  < 2e-16 ***
## bbC_bbO_short                -0.094872   0.010529  -9.011  < 2e-16 ***
## aliph1HC_aliph2HC_medlong     0.044980   0.009316   4.828 1.52e-06 ***
## sulfur_bbC_vlong             -0.057484   0.007966  -7.216 8.41e-13 ***
## aromaticC_hydroxylO_long     -0.042631   0.005551  -7.680 2.83e-14 ***
## sulfur_bbC_medlong           -0.079939   0.011605  -6.888 8.25e-12 ***
## aliph1HC_aromaticC_long       0.106279   0.008603  12.353  < 2e-16 ***
## carboxylC_bbN_vlong           0.071152   0.013733   5.181 2.50e-07 ***
## aliph1HC_bbC_long            -0.012990   0.009515  -1.365 0.172375
## scAGN_bbN_long                0.058209   0.007090   8.209 4.70e-16 ***
## aliph1HC_aliph1HC_vlong       0.099798   0.026411   3.779 0.000164 ***
## bbN_bbCA_medlong             -0.046007   0.004835  -9.515  < 2e-16 ***
## aromaticC_sulfur_long         0.081134   0.011479   7.068 2.39e-12 ***
## aliph3HC_bbN_short            0.145819   0.020467   7.125 1.61e-12 ***
## aliph1HC_aromaticC_medlong    0.085360   0.009944   8.584  < 2e-16 ***
## aliph2HC_aromaticC_medshort   0.032314   0.004654   6.944 5.64e-12 ***
## aliph1HC_bbO_long            -0.069758   0.010606  -6.577 6.57e-11 ***
## aliph1HC_aromaticC_short      0.165391   0.022317   7.411 2.07e-13 ***
## aliph2HC_aromaticC_vlong      0.017642   0.002665   6.619 4.99e-11 ***
## carboxylO_bbN_long            0.055419   0.011255   4.924 9.40e-07 ***
## aliph3HC_bbC_vlong            0.024377   0.004883   4.992 6.66e-07 ***
## bbN_bbCA_long                -0.020572   0.003617  -5.688 1.54e-08 ***
## scArgN_bbO_medlong            0.457185   0.083272   5.490 4.70e-08 ***
## carboxylC_scLysN_vlong        0.213552   0.042742   4.996 6.52e-07 ***
## carbonylC_aromaticC_long      0.046141   0.008369   5.513 4.14e-08 ***
## aliph1HC_scArgN_long          0.508017   0.099074   5.128 3.31e-07 ***
## aliph2HC_bbN_medlong          0.024744   0.004275   5.788 8.62e-09 ***
## aromaticC_aromaticC_vlong    -0.012437   0.002596  -4.790 1.83e-06 ***
## aliph3HC_hydroxylO_long       0.053663   0.012085   4.440 9.63e-06 ***
## scAGN_carbonylO_medshort     -0.083377   0.020267  -4.114 4.10e-05 ***
## carbonylC_bbC_medlong        -0.050393   0.010389  -4.851 1.36e-06 ***
## scAGN_bbN_medlong             0.036422   0.008852   4.115 4.09e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6292 on 1518 degrees of freedom
```

```
## Multiple R-squared:  0.8226, Adjusted R-squared:  0.8181
## F-statistic: 185.2 on 38 and 1518 DF,  p-value: < 2.2e-16
```

```
BIC(m2)
```

```
## [1] 3230.409
```

```
pred2 <- predict(m2, newdata = validSet)
sqrt(mean((validSet$accuracy - pred2)^2)) # RMSE on validation
```

```
## [1] 0.6570954
```

```
sqrt(mean(m2$residuals^2)) # RMSE on train
```

```
## [1] 0.6212791
```

```
#Forward-Backward
#m3 <- stepAIC(object = empty, scope = list(upper = full, lower = empty), direction = "both", k = log(n

#save(m3, file= "m3.rda")
load("m3.rda")
summary(m3)
```

```
##
## Call:
## lm(formula = accuracy ~ aliph1HC_aliph2HC_long + scLysN_bbC_vlong +
##     aliph2HC_bbN_medshort + aliph1HC_aromaticC_medshort + carbonylC_aromaticC_short +
##     aromaticC_hydroxylO_medlong + aliph1HC_aromaticC_vlong +
##     bbC_bbO_short + aliph1HC_aliph2HC_medlong + sulfur_bbC_vlong +
##     aromaticC_hydroxylO_long + sulfur_bbC_medlong + aliph1HC_aromaticC_long +
##     scAGN_bbN_long + aliph1HC_aliph1HC_vlong + bbN_bbCA_medlong +
##     aromaticC_sulfur_long + aliph3HC_bbN_short + aliph1HC_aromaticC_medlong +
##     aliph2HC_aromaticC_medshort + aliph1HC_bbO_long + aliph1HC_aromaticC_short +
##     aliph2HC_aromaticC_vlong + aliph3HC_bbC_vlong + bbN_bbCA_long +
##     scArgN_bbO_medlong + carboxylC_scLysN_vlong + carbonylC_aromaticC_long +
##     aliph1HC_scArgN_long + aliph2HC_bbN_medlong + aromaticC_aromaticC_vlong +
##     aliph3HC_hydroxylO_long + scAGN_carbonylO_medshort + carbonylC_bbC_medlong +
##     scAGN_bbN_medlong + aliph1HC_bbC_medshort + bbN_bbN_medshort +
##     carbonylC_bbProN_medlong + aromaticC_hydroxylO_medshort +
##     aromaticC_bbO_vshort + aliph1HC_aliph3HC_medlong + aliph1HC_aliph1HC_long +
##     carboxylO_bbO_vlong + hydroxylO_carbonylO_medlong + carboxylO_bbO_long +
##     bbN_bbO_vlong + aliph1HC_bbProN_medlong + bbO_bbO_long +
##     aliph1HC_bbN_medshort + hydroxylO_bbC_medlong + scLysN_carboxylO_long +
##     bbN_bbC_vshort + carboxylC_carboxylC_vlong + bbCA_bbC_vlong +
##     bbN_bbN_medlong + aliph2HC_carboxylO_medlong + carbonylC_scAGN_long +
##     bbO_bbO_short + carbonylC_sulfur_short + bbCA_bbO_vshort +
##     carbonylO_sulfur_medshort + scArgN_carboxylO_long + bbProN_carboxylO_vlong +
##     aliph2HC_scArgN_vlong + carbonylC_carboxylC_long + carbonylC_aliph2HC_short +
##     aromaticC_bbCA_vlong + aliph2HC_aliph3HC_long + aromaticC_bbN_short +
##     aliph1HC_bbN_vlong + aliph1HC_bbO_vlong + aliph2HC_bbN_vlong +
##     aliph1HC_scAGN_long, data = trainSet)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2761 -0.3437  0.0111  0.3393  2.0508
##
## Coefficients:
```

```
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    2.402801   0.417709   5.752 1.07e-08 ***
## aliph1HC_aliph2HC_long         0.053661   0.008389   6.396 2.13e-10 ***
## scLysN_bbC_vlong              -0.115602   0.018421  -6.275 4.57e-10 ***
## aliph2HC_bbN_medshort         -0.044526   0.006330  -7.034 3.06e-12 ***
## aliph1HC_aromaticC_medshort    0.120614   0.011679  10.328  < 2e-16 ***
## carbonylC_aromaticC_short      0.164500   0.031156   5.280 1.48e-07 ***
## aromaticC_hydroxylO_medlong   -0.050416   0.007417  -6.797 1.54e-11 ***
## aliph1HC_aromaticC_vlong       0.085732   0.008692   9.863  < 2e-16 ***
## bbC_bbO_short                 -0.038111   0.010620  -3.588 0.000343 ***
## aliph1HC_aliph2HC_medlong      0.039269   0.009289   4.228 2.51e-05 ***
## sulfur_bbC_vlong              -0.038917   0.007443  -5.228 1.95e-07 ***
## aromaticC_hydroxylO_long      -0.043191   0.005506  -7.845 8.25e-15 ***
## sulfur_bbC_medlong            -0.056187   0.010713  -5.245 1.79e-07 ***
## aliph1HC_aromaticC_long        0.090470   0.008257  10.957  < 2e-16 ***
## scAGN_bbN_long                 0.055807   0.006501   8.584  < 2e-16 ***
## aliph1HC_aliph1HC_vlong        0.139707   0.025505   5.478 5.06e-08 ***
## bbN_bbCA_medlong              -0.042739   0.004858  -8.797  < 2e-16 ***
## aromaticC_sulfur_long          0.053975   0.010570   5.106 3.71e-07 ***
## aliph3HC_bbN_short             0.123730   0.019416   6.373 2.47e-10 ***
## aliph1HC_aromaticC_medlong     0.080223   0.009245   8.677  < 2e-16 ***
## aliph2HC_aromaticC_medshort    0.031150   0.004527   6.880 8.78e-12 ***
## aliph1HC_bbO_long             -0.046213   0.009937  -4.651 3.60e-06 ***
## aliph1HC_aromaticC_short       0.145989   0.021832   6.687 3.22e-11 ***
## aliph2HC_aromaticC_vlong       0.018809   0.002728   6.895 7.97e-12 ***
## aliph3HC_bbC_vlong             0.023026   0.004647   4.955 8.06e-07 ***
## bbN_bbCA_long                 -0.021065   0.004194  -5.023 5.71e-07 ***
## scArgN_bbO_medlong             0.386727   0.086058   4.494 7.54e-06 ***
## carboxylC_scLysN_vlong         0.125173   0.042393   2.953 0.003200 **
## carbonylC_aromaticC_long       0.033047   0.007658   4.315 1.70e-05 ***
## aliph1HC_scArgN_long           0.711104   0.092549   7.684 2.80e-14 ***
## aliph2HC_bbN_medlong           0.027738   0.003888   7.134 1.51e-12 ***
## aromaticC_aromaticC_vlong     -0.011897   0.002560  -4.647 3.66e-06 ***
## aliph3HC_hydroxylO_long        0.051772   0.011270   4.594 4.72e-06 ***
## scAGN_carbonylO_medshort      -0.087427   0.017901  -4.884 1.15e-06 ***
## carbonylC_bbC_medlong         -0.043490   0.009506  -4.575 5.16e-06 ***
## scAGN_bbN_medlong              0.037415   0.008054   4.645 3.70e-06 ***
## aliph1HC_bbC_medshort          0.067511   0.016486   4.095 4.45e-05 ***
## bbN_bbN_medshort              -0.090507   0.011203  -8.079 1.34e-15 ***
## carbonylC_bbProN_medlong      -0.145503   0.031240  -4.658 3.49e-06 ***
## aromaticC_hydroxylO_medshort -0.052516   0.010601  -4.954 8.10e-07 ***
## aromaticC_bbO_vshort          -0.218962   0.039282  -5.574 2.95e-08 ***
## aliph1HC_aliph3HC_medlong     -0.082600   0.016104  -5.129 3.29e-07 ***
## aliph1HC_aliph1HC_long         0.256634   0.037597   6.826 1.27e-11 ***
## carboxylO_bbO_vlong            0.033366   0.008768   3.806 0.000147 ***
## hydroxylO_carbonylO_medlong    0.057452   0.012363   4.647 3.66e-06 ***
## carboxylO_bbO_long             0.050933   0.010769   4.729 2.47e-06 ***
## bbN_bbO_vlong                  0.013412   0.003072   4.365 1.36e-05 ***
## aliph1HC_bbProN_medlong        0.258476   0.049309   5.242 1.82e-07 ***
## bbO_bbO_long                  -0.019297   0.005257  -3.671 0.000250 ***
## aliph1HC_bbN_medshort         -0.057962   0.017265  -3.357 0.000807 ***
## hydroxylO_bbC_medlong          0.034662   0.008172   4.242 2.36e-05 ***
## scLysN_carboxylO_long          0.174563   0.029458   5.926 3.86e-09 ***
## bbN_bbC_vshort                -0.082895   0.027162  -3.052 0.002315 **
```

```
## carboxylC_carboxylC_vlong    -0.189042   0.038258  -4.941 8.65e-07 ***
## bbCA_bbC_vlong               -0.014965   0.003382  -4.424 1.04e-05 ***
## bbN_bbN_medlong              -0.024361   0.007850  -3.103 0.001950 **
## aliph2HC_carboxylO_medlong    0.049173   0.012693   3.874 0.000112 ***
## carbonylC_scAGN_long          0.075970   0.021179   3.587 0.000345 ***
## bbO_bbO_short                -0.048840   0.012353  -3.954 8.06e-05 ***
## carbonylC_sulfur_short       -0.161486   0.038997  -4.141 3.65e-05 ***
## bbCA_bbO_vshort              -0.332933   0.073167  -4.550 5.79e-06 ***
## carbonylO_sulfur_medshort     0.073984   0.026895   2.751 0.006017 **
## scArgN_carboxylO_long        -0.112380   0.024689  -4.552 5.75e-06 ***
## bbProN_carboxylO_vlong        0.109278   0.039562   2.762 0.005812 **
## aliph2HC_scArgN_vlong         0.078742   0.024440   3.222 0.001301 **
## carbonylC_carboxylC_long      0.172284   0.047386   3.636 0.000287 ***
## carbonylC_aliph2HC_short      0.058028   0.021877   2.653 0.008075 **
## aromaticC_bbCA_vlong          0.011177   0.002860   3.908 9.71e-05 ***
## aliph2HC_aliph3HC_long       -0.014768   0.005134  -2.876 0.004081 **
## aromaticC_bbN_short           0.045346   0.012352   3.671 0.000250 ***
## aliph1HC_bbN_vlong           -0.027647   0.008392  -3.294 0.001010 **
## aliph1HC_bbO_vlong            0.026360   0.007491   3.519 0.000446 ***
## aliph2HC_bbN_vlong            0.008121   0.002847   2.852 0.004402 **
## aliph1HC_scAGN_long          -0.048337   0.017317  -2.791 0.005317 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5464 on 1483 degrees of freedom
## Multiple R-squared:  0.8693, Adjusted R-squared:  0.8629
## F-statistic: 135.1 on 73 and 1483 DF,  p-value: < 2.2e-16
```

```
BIC(m3)
```

```
## [1] 3011.712
```

```
pred3 <- predict(m3, newdata = validSet)
sqrt(mean((validSet$accuracy - pred3)^2)) # RMSE on validation
```

```
## [1] 0.6229044
```

```
sqrt(mean(m3$residuals^2)) # RMSE on train
```

```
## [1] 0.5332196
```

```
#Forward-Backward with larger penalty
#m4 <- stepAIC(object = empty, scope = list(upper = full, lower = empty), direction = "both", k = 2*log

#save(m4, file= "m4.rda")
load("m4.rda")
summary(m4)
```

```
##
## Call:
## lm(formula = accuracy ~ aliph1HC_aliph2HC_long + scLysN_bbC_vlong +
##     aliph2HC_bbN_medshort + aliph1HC_aromaticC_medshort + carbonylC_aromaticC_short +
##     aromaticC_hydroxylO_medlong + aliph1HC_aromaticC_vlong +
##     bbC_bbO_short + aliph1HC_aliph2HC_medlong + sulfur_bbC_vlong +
##     aromaticC_hydroxylO_long + sulfur_bbC_medlong + aliph1HC_aromaticC_long +
##     carboxylC_bbN_vlong + scAGN_bbN_long + bbN_bbCA_medlong +
##     aromaticC_sulfur_long + aliph3HC_bbN_short + aliph1HC_aromaticC_medlong +
```

```
##     aliph2HC_aromaticC_medshort + aliph1HC_bbO_long + aliph1HC_aromaticC_short +
##     aliph2HC_aromaticC_vlong + carboxylO_bbN_long + aliph3HC_bbC_vlong +
##     bbN_bbCA_long + scArgN_bbO_medlong + carboxylC_scLysN_vlong +
##     carbonylC_aromaticC_long + aliph2HC_bbN_medlong + aliph1HC_scArgN_long +
##     aromaticC_aromaticC_vlong + aliph3HC_hydroxylO_long + scAGN_carbonylO_medshort +
##     aliph1HC_bbC_medshort + carbonylO_sulfur_medshort + aliph1HC_hydroxylO_medlong,
##     data = trainSet)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6091 -0.3586  0.0160  0.3728  2.4598
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   2.994476   0.411300   7.281 5.31e-13 ***
## aliph1HC_aliph2HC_long        0.055644   0.008922   6.237 5.78e-10 ***
## scLysN_bbC_vlong             -0.158642   0.018065  -8.782  < 2e-16 ***
## aliph2HC_bbN_medshort        -0.039336   0.006780  -5.801 7.99e-09 ***
## aliph1HC_aromaticC_medshort   0.140426   0.012474  11.257  < 2e-16 ***
## carbonylC_aromaticC_short     0.251020   0.033869   7.411 2.07e-13 ***
## aromaticC_hydroxylO_medlong -0.071905   0.007926  -9.072  < 2e-16 ***
## aliph1HC_aromaticC_vlong      0.119993   0.008778  13.670  < 2e-16 ***
## bbC_bbO_short                -0.095490   0.010528  -9.070  < 2e-16 ***
## aliph1HC_aliph2HC_medlong     0.037053   0.009456   3.919 9.30e-05 ***
## sulfur_bbC_vlong             -0.065944   0.007993  -8.250 3.38e-16 ***
## aromaticC_hydroxylO_long     -0.037846   0.005555  -6.813 1.37e-11 ***
## sulfur_bbC_medlong           -0.089190   0.011525  -7.739 1.82e-14 ***
## aliph1HC_aromaticC_long       0.102868   0.008509  12.089  < 2e-16 ***
## carboxylC_bbN_vlong           0.073427   0.013566   5.412 7.21e-08 ***
## scAGN_bbN_long                0.043683   0.006956   6.280 4.42e-10 ***
## bbN_bbCA_medlong             -0.048627   0.004657 -10.441  < 2e-16 ***
## aromaticC_sulfur_long         0.087388   0.011305   7.730 1.94e-14 ***
## aliph3HC_bbN_short            0.127198   0.020611   6.171 8.66e-10 ***
## aliph1HC_aromaticC_medlong    0.076740   0.010229   7.502 1.06e-13 ***
## aliph2HC_aromaticC_medshort   0.031533   0.004642   6.793 1.57e-11 ***
## aliph1HC_bbO_long            -0.083415   0.009892  -8.432  < 2e-16 ***
## aliph1HC_aromaticC_short      0.173312   0.022184   7.812 1.04e-14 ***
## aliph2HC_aromaticC_vlong      0.015827   0.002660   5.950 3.33e-09 ***
## carboxylO_bbN_long            0.063181   0.011102   5.691 1.51e-08 ***
## aliph3HC_bbC_vlong            0.025193   0.004858   5.186 2.44e-07 ***
## bbN_bbCA_long                -0.021310   0.003520  -6.054 1.77e-09 ***
## scArgN_bbO_medlong            0.453995   0.083734   5.422 6.85e-08 ***
## carboxylC_scLysN_vlong        0.203646   0.042675   4.772 2.00e-06 ***
## carbonylC_aromaticC_long      0.048709   0.008346   5.836 6.52e-09 ***
## aliph2HC_bbN_medlong          0.023573   0.004249   5.547 3.42e-08 ***
## aliph1HC_scArgN_long          0.531617   0.098370   5.404 7.54e-08 ***
## aromaticC_aromaticC_vlong    -0.013671   0.002600  -5.259 1.66e-07 ***
## aliph3HC_hydroxylO_long       0.056027   0.011702   4.788 1.85e-06 ***
## scAGN_carbonylO_medshort     -0.088372   0.020193  -4.376 1.29e-05 ***
## aliph1HC_bbC_medshort         0.073710   0.016885   4.365 1.35e-05 ***
## carbonylO_sulfur_medshort     0.123712   0.029901   4.137 3.71e-05 ***
## aliph1HC_hydroxylO_medlong    0.114368   0.029238   3.912 9.57e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.6284 on 1519 degrees of freedom
## Multiple R-squared:  0.8229, Adjusted R-squared:  0.8186
## F-statistic: 190.8 on 37 and 1519 DF,  p-value: < 2.2e-16
```

```
BIC(m4)
```

```
## [1] 3220.17
```

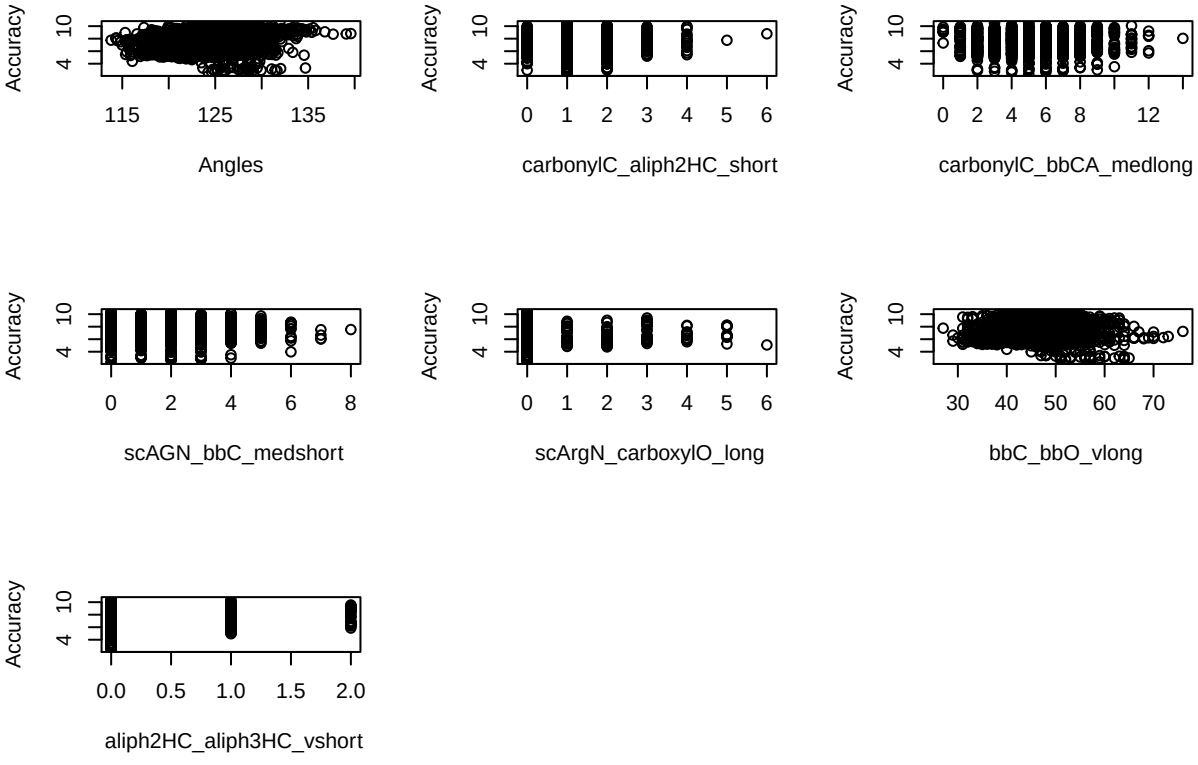
```
pred4 <- predict(m4, newdata = validSet)
sqrt(mean((validSet$accuracy - pred4)^2)) # RMSE on validation
```

```
## [1] 0.682042
```

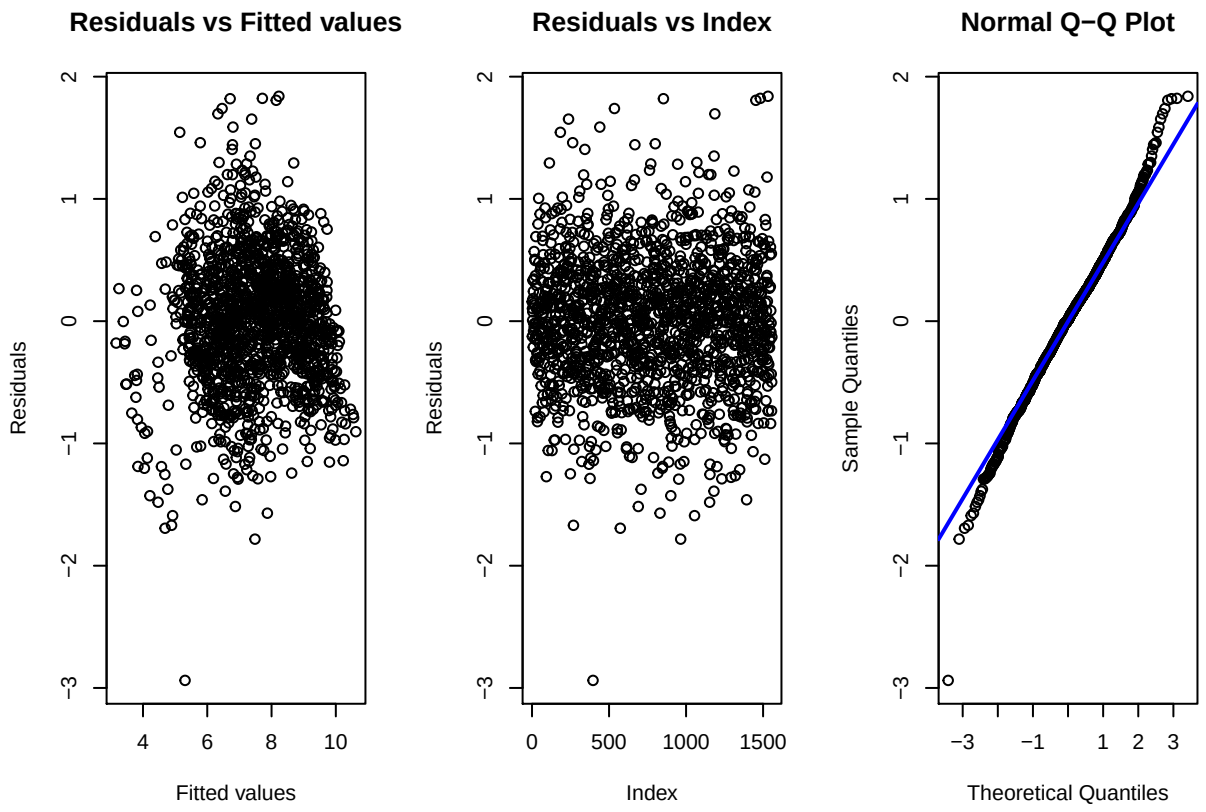
```
sqrt(mean(m4$residuals^2)) # RMSE on train
```

```
## [1] 0.6207029
```

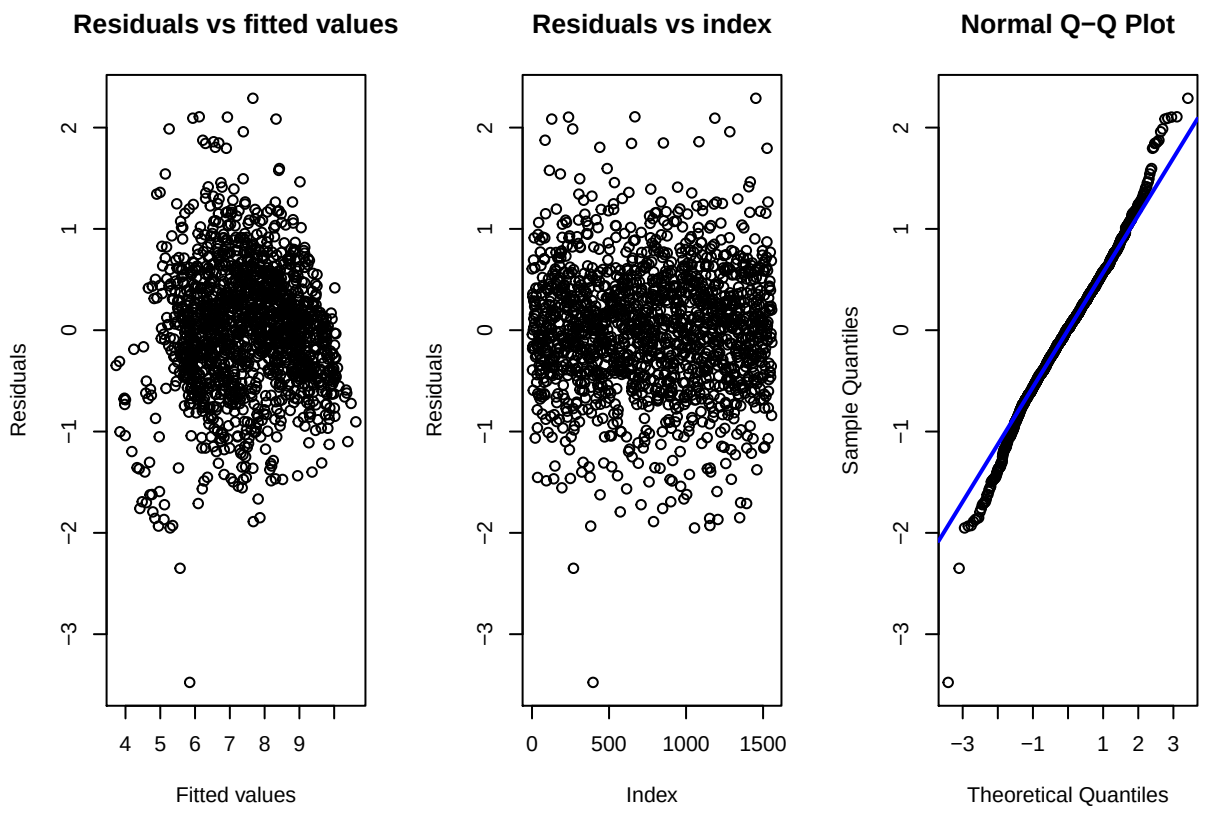
```
#Plots of Models to Check Model Assumptions

#m1
par(mfrow=c(1,3))
```



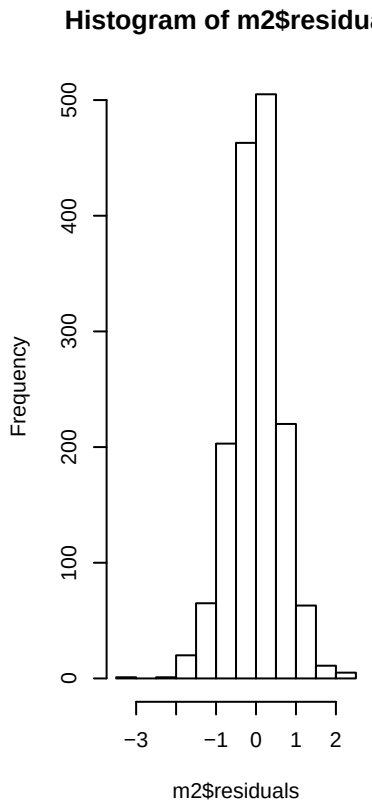

```
plot(m1$fitted.values,m1$residuals, xlab="Fitted values",
ylab="Residuals", main="Residuals vs Fitted values")
plot(1:nrow(trainSet),m1$residuals, xlab="Index",
ylab="Residuals", main="Residuals vs Index")
qqnorm(m1$residuals)
qqline(m1$residuals, col="blue", lwd=2)
```



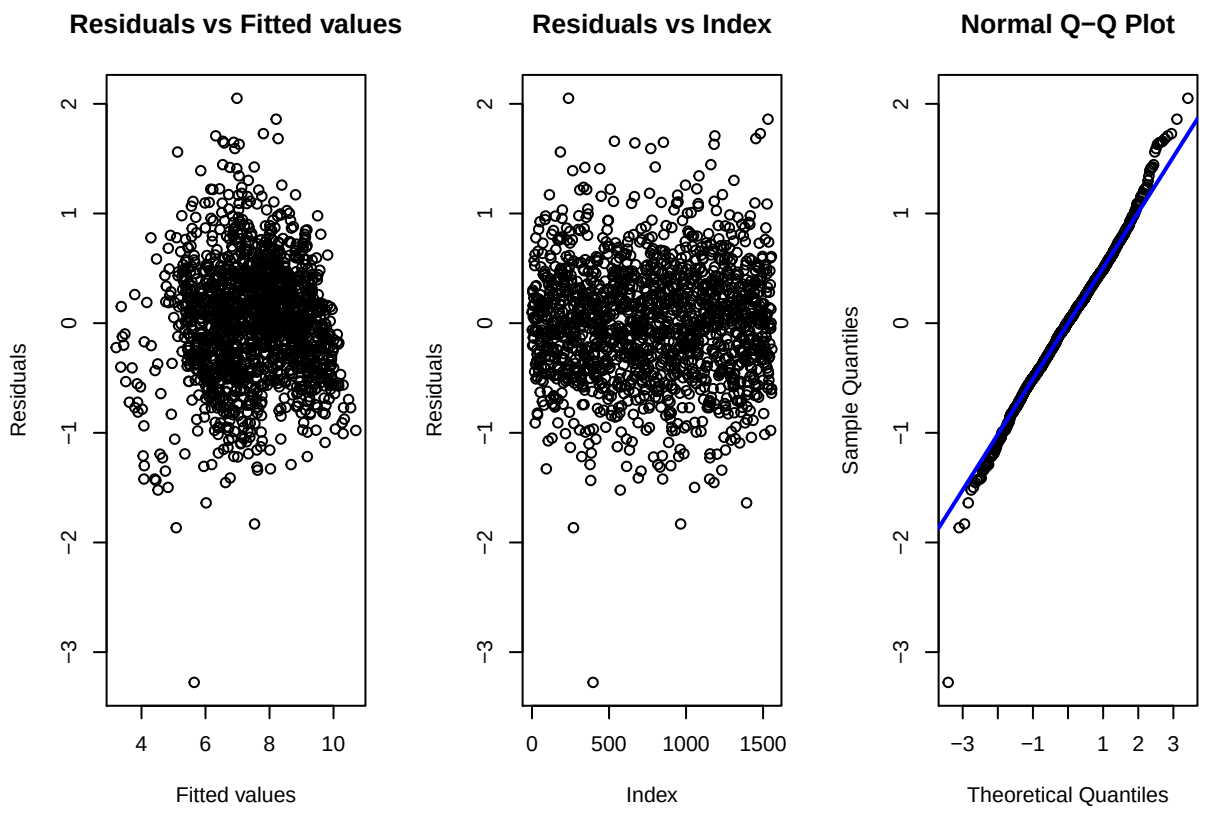

```
#m2
plot(m2$fitted.values,m2$residuals, xlab="Fitted values",
ylab="Residuals", main="Residuals vs fitted values")
plot(1:nrow(trainSet),m2$residuals, xlab="Index",
ylab="Residuals", main="Residuals vs index")
qqnorm(m2$residuals)
qqline(m2$residuals, col="blue", lwd=2)
```



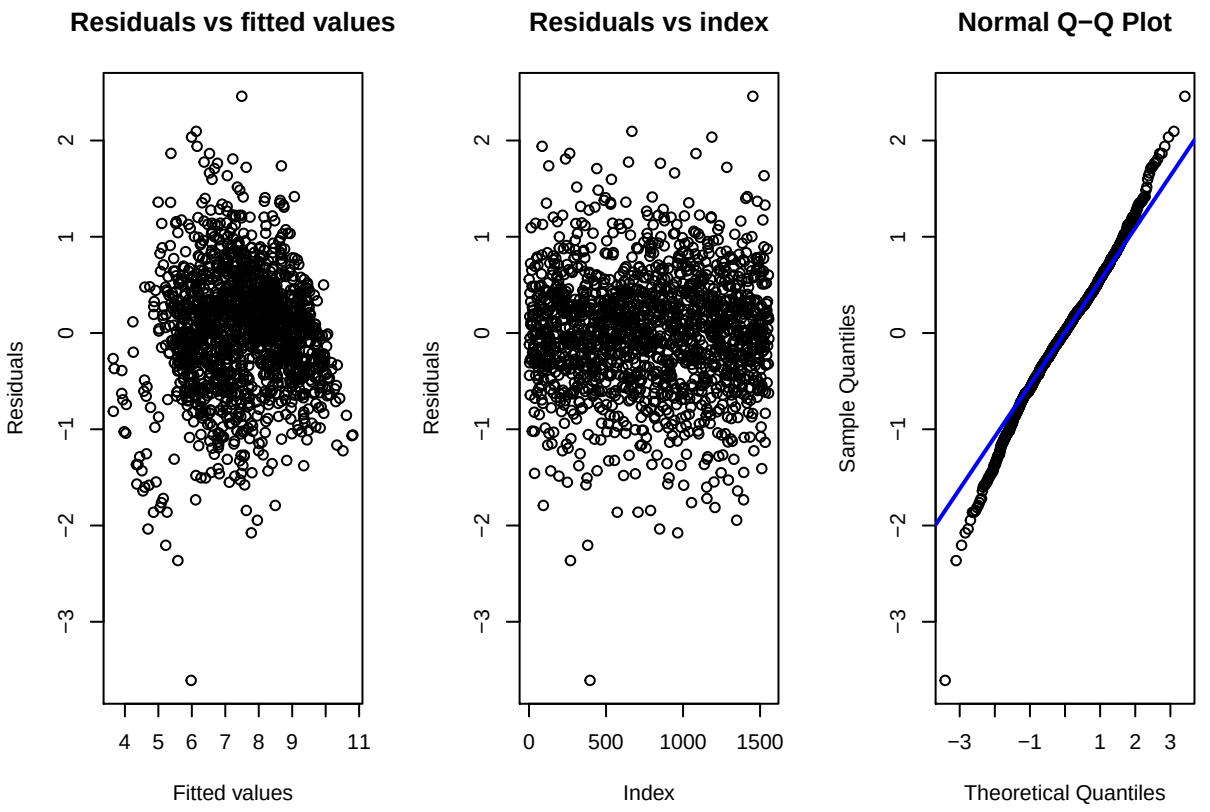

```
hist ( m2$residuals )

#m3
par(mfrow=c(1,3))
```




```
plot(m3$fitted.values,m3$residuals, xlab="Fitted values",
ylab="Residuals", main="Residuals vs Fitted values")
plot(1:nrow(trainSet),m3$residuals, xlab="Index",
ylab="Residuals", main="Residuals vs Index")
qqnorm(m3$residuals)
qqline(m3$residuals, col="blue", lwd=2)
```




```
#m4
plot(m4$fitted.values,m4$residuals, xlab="Fitted values",
ylab="Residuals", main="Residuals vs fitted values")
plot(1:nrow(trainSet),m4$residuals, xlab="Index",
ylab="Residuals", main="Residuals vs index")
qqnorm(m4$residuals)
qqline(m4$residuals, col="blue", lwd=2)
```




```
hist ( m4$residuals )

# K fold cross validation to choose model selection method
K <- 5
validSetSplits <- sample((1:N)%%K + 1)
RMSE1 <- c()
RMSE2 <- c()
RMSE3 <- c()
RMSE4 <-c()
for (k in 1:K) {
  validSet <- protein1[validSetSplits==k,]
  trainSet <- protein1[validSetSplits!=k,]

  full <- lm(accuracy ~ ., data = trainSet)
  empty <- lm(accuracy ~ 1, data = trainSet)

  load("m1.rda")
  pred1 <- predict(m1, newdata = validSet)
  RMSE1[k] <- sqrt(mean((validSet$accuracy - pred1)^2))

  load("m2.rda")
  pred2 <- predict(m2, newdata = validSet)
  RMSE2[k] <- sqrt(mean((validSet$accuracy - pred2)^2))

  load("m3.rda")
  pred3 <- predict(m3, newdata = validSet)
  RMSE3[k] <- sqrt(mean((validSet$accuracy - pred3)^2))
```

```
  load("m4.rda")
  pred4 <- predict(m4, newdata = validSet)
  RMSE4[k] <- sqrt(mean((validSet$accuracy - pred4)^2))

}


RMSE1
```

```
## [1] 0.5562111 0.5407755 0.5622529 0.5534633 0.5178010
```

```
RMSE2
```

```
## [1] 0.6539304 0.6346173 0.6130136 0.6214550 0.6191332
```

```
RMSE3
```

```
## [1] 0.5706701 0.5446036 0.5599246 0.5598758 0.5253923
```

```
RMSE4
```

```
## [1] 0.6541111 0.6276039 0.6212410 0.6393845 0.6242970
```

```
mean(RMSE1)
```

```
## [1] 0.5461008
```

```
mean(RMSE2)
```

```
## [1] 0.6284299
```

```
mean(RMSE3)
```

```
## [1] 0.5520933
```

```
mean(RMSE4)
```

```
## [1] 0.6333275
```

```
# turns out m1 is indeed the better procedure among these 4 based on CV prediction error
# if we decide on procedure m1, we can apply procedure m1 to all observations
# to get a final model for future predictions
# e.g.,
full <- lm(accuracy ~ ., data = protein1)
empty <- lm(accuracy ~ 1, data = protein1)

#mfinal <- stepAIC(object = empty, scope = list(upper = full, lower = empty), direction = "forward", k =


#save(mfinal, file= "mfinal.rda")
load("mfinal.rda")

summary(mfinal)
```

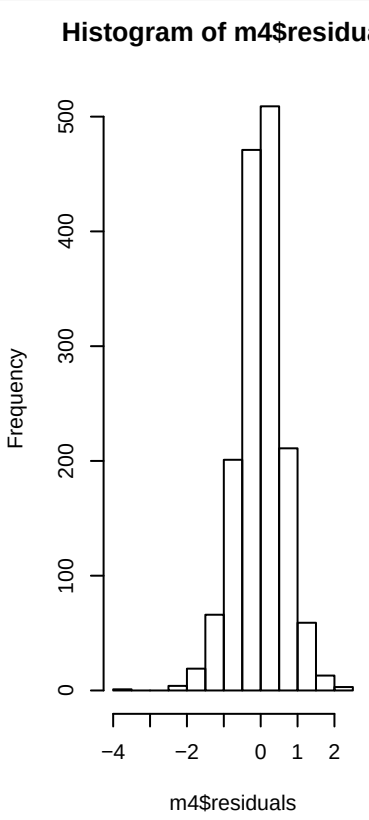
```
##
## Call:
## lm(formula = accuracy ~ aliph1HC_aliph2HC_long + scLysN_bbC_vlong +
##     aliph2HC_bbN_medshort + aliph1HC_aliph2HC_medlong + bbC_bbC_medshort +
##     aromaticC_sulfur_short + aromaticC_hydroxylO_medlong + aliph1HC_aromaticC_medshort +
##     carbonylC_aromaticC_short + aliph1HC_aromaticC_vlong + bbN_bbCA_medlong +
##     aliph1HC_aromaticC_long + aliph1HC_aliph1HC_vlong + aliph1HC_bbN_vlong +
```

```
##     scAGN_bbN_long + carboxylO_bbC_vlong + aromaticC_hydroxylO_long +
##     bbN_bbN_medlong + sulfur_bbC_medlong + aliph2HC_aromaticC_vlong +
##     aliph3HC_bbC_vlong + aliph1HC_bbO_long + aliph1HC_aromaticC_medlong +
##     aliph3HC_bbN_short + aliph2HC_scArgN_vlong + aliph3HC_aromaticC_long +
##     aromaticC_bbCA_vlong + bbO_bbO_short + carbonylC_aromaticC_long +
##     carboxylC_scLysN_vlong + sulfur_bbC_vlong + aromaticC_sulfur_long +
##     aliph2HC_bbN_medlong + scAGN_carbonylO_medshort + bbO_bbO_long +
##     aliph1HC_aliph1HC_long + aliph1HC_scArgN_long + aliph1HC_aromaticC_short +
##     carboxylC_aromaticC_long + aromaticC_hydroxylO_medshort +
##     carbonylC_bbC_medlong + hydroxylO_carbonylO_medlong + aliph3HC_hydroxylO_long +
##     bbN_bbN_medshort + carboxylC_bbC_medlong + aliph2HC_aromaticC_medshort +
##     scArgN_bbO_medlong + carboxylC_carboxylC_vlong + scArgN_carboxylO_long +
##     carbonylC_bbProN_medlong + scAGN_bbCA_medlong + aliph3HC_aromaticC_vlong +
##     carboxylO_bbN_vlong + scLysN_carboxylO_long + bbCA_bbO_vshort +
##     sulfur_bbCA_short + carbonylC_sulfur_short + bbCA_bbC_vlong +
##     aliph2HC_bbN_vlong + aliph1HC_bbO_vlong + bbProN_carboxylO_vlong +
##     aromaticC_scAGN_vlong + carbonylO_sulfur_medshort + aliph1HC_bbC_medshort +
##     aromaticC_bbO_vlong + hydroxylO_bbC_medlong + aromaticC_hydroxylO_short +
##     carboxylC_bbO_medlong + bbProN_bbCA_medshort + carbonylC_scAGN_long +
##     bbN_bbCA_long + aromaticC_bbO_vshort + carbonylC_carboxylO_short +
##     carbonylC_bbCA_medshort + aliph1HC_aliph3HC_medlong + scAGN_carboxylO_vshort +
##     sulfur_sulfur_vlong + aliph1HC_bbProN_medlong + aliph1HC_bbN_medshort +
##     aromaticC_aromaticC_vlong + aliph2HC_aliph3HC_vlong + aliph3HC_bbCA_short +
##     aromaticC_bbO_medlong + aliph3HC_carboxylO_long, data = protein1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.87372 -0.33177 -0.00649  0.33619  1.90183
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   1.735463   0.376227   4.613 4.24e-06 ***
## aliph1HC_aliph2HC_long        0.046797   0.007420   6.307 3.55e-10 ***
## scLysN_bbC_vlong             -0.109515   0.016506  -6.635 4.24e-11 ***
## aliph2HC_bbN_medshort        -0.037766   0.005822  -6.486 1.13e-10 ***
## aliph1HC_aliph2HC_medlong     0.034923   0.008321   4.197 2.83e-05 ***
## bbC_bbC_medshort             -0.034217   0.008762  -3.905 9.75e-05 ***
## aromaticC_sulfur_short       -0.031489   0.014993  -2.100 0.035846 *
## aromaticC_hydroxylO_medlong  -0.058413   0.006726  -8.684  < 2e-16 ***
## aliph1HC_aromaticC_medshort   0.093504   0.010569   8.847  < 2e-16 ***
## carbonylC_aromaticC_short     0.169676   0.028389   5.977 2.72e-09 ***
## aliph1HC_aromaticC_vlong      0.091808   0.007989  11.492  < 2e-16 ***
## bbN_bbCA_medlong             -0.037791   0.004278  -8.835  < 2e-16 ***
## aliph1HC_aromaticC_long       0.088988   0.007854  11.331  < 2e-16 ***
## aliph1HC_aliph1HC_vlong       0.138224   0.023551   5.869 5.18e-09 ***
## aliph1HC_bbN_vlong           -0.039990   0.007419  -5.390 7.94e-08 ***
## scAGN_bbN_long                0.040275   0.005879   6.850 9.98e-12 ***
## carboxylO_bbC_vlong           0.039622   0.008203   4.830 1.48e-06 ***
## aromaticC_hydroxylO_long     -0.038628   0.004948  -7.807 9.72e-15 ***
## bbN_bbN_medlong              -0.029963   0.007139  -4.197 2.83e-05 ***
## sulfur_bbC_medlong           -0.060841   0.009725  -6.256 4.88e-10 ***
## aliph2HC_aromaticC_vlong      0.016871   0.002573   6.556 7.16e-11 ***
## aliph3HC_bbC_vlong            0.020388   0.004135   4.930 8.95e-07 ***
## aliph1HC_bbO_long            -0.049413   0.008885  -5.561 3.07e-08 ***
```

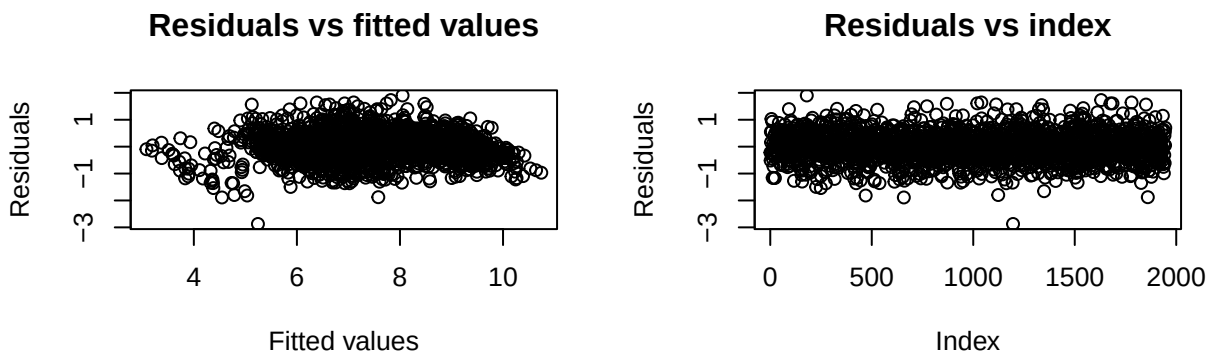
```
## aliph1HC_aromaticC_medlong    0.073905   0.008724   8.472  < 2e-16 ***
## aliph3HC_bbN_short            0.103115   0.018679   5.520 3.86e-08 ***
## aliph2HC_scArgN_vlong         0.111865   0.021842   5.121 3.35e-07 ***
## aliph3HC_aromaticC_long      -0.016094   0.003753  -4.289 1.89e-05 ***
## aromaticC_bbCA_vlong          0.014861   0.002710   5.485 4.71e-08 ***
## bbO_bbO_short                -0.059256   0.011574  -5.120 3.38e-07 ***
## carbonylC_aromaticC_long      0.027979   0.006908   4.050 5.32e-05 ***
## carboxylC_scLysN_vlong        0.133323   0.039466   3.378 0.000745 ***
## sulfur_bbC_vlong             -0.036977   0.006901  -5.358 9.45e-08 ***
## aromaticC_sulfur_long         0.048626   0.009815   4.954 7.91e-07 ***
## aliph2HC_bbN_medlong          0.022535   0.003533   6.378 2.26e-10 ***
## scAGN_carbonylO_medshort     -0.080476   0.016565  -4.858 1.28e-06 ***
## bbO_bbO_long                 -0.022536   0.004645  -4.852 1.32e-06 ***
## aliph1HC_aliph1HC_long        0.177699   0.032920   5.398 7.61e-08 ***
## aliph1HC_scArgN_long          0.747241   0.084052   8.890  < 2e-16 ***
## aliph1HC_aromaticC_short      0.131340   0.021014   6.250 5.07e-10 ***
## carboxylC_aromaticC_long      0.050266   0.014397   3.491 0.000492 ***
## aromaticC_hydroxylO_medshort -0.054078   0.009673  -5.591 2.59e-08 ***
## carbonylC_bbC_medlong        -0.039571   0.008639  -4.581 4.94e-06 ***
## hydroxylO_carbonylO_medlong   0.039897   0.010942   3.646 0.000273 ***
## aliph3HC_hydroxylO_long       0.047843   0.010096   4.739 2.31e-06 ***
## bbN_bbN_medshort             -0.085114   0.010383  -8.198 4.51e-16 ***
## carboxylC_bbC_medlong         0.055137   0.018188   3.031 0.002467 **
## aliph2HC_aromaticC_medshort   0.023891   0.004459   5.358 9.45e-08 ***
## scArgN_bbO_medlong            0.376225   0.076831   4.897 1.06e-06 ***
## carboxylC_carboxylC_vlong    -0.201210   0.036259  -5.549 3.28e-08 ***
## scArgN_carboxylO_long        -0.141090   0.021915  -6.438 1.54e-10 ***
## carbonylC_bbProN_medlong     -0.162327   0.027701  -5.860 5.46e-09 ***
## scAGN_bbCA_medlong            0.032091   0.008092   3.966 7.59e-05 ***
## aliph3HC_aromaticC_vlong     -0.016470   0.003700  -4.451 9.05e-06 ***
## carboxylO_bbN_vlong           0.033386   0.008189   4.077 4.75e-05 ***
## scLysN_carboxylO_long         0.135168   0.026724   5.058 4.66e-07 ***
## bbCA_bbO_vshort              -0.357844   0.066206  -5.405 7.32e-08 ***
## sulfur_bbCA_short            -0.076015   0.023236  -3.271 0.001089 **
## carbonylC_sulfur_short       -0.122569   0.034872  -3.515 0.000451 ***
## bbCA_bbC_vlong               -0.013418   0.002998  -4.475 8.10e-06 ***
## aliph2HC_bbN_vlong            0.010102   0.002476   4.080 4.70e-05 ***
## aliph1HC_bbO_vlong            0.025221   0.006720   3.753 0.000180 ***
## bbProN_carboxylO_vlong        0.108459   0.036233   2.993 0.002796 **
## aromaticC_scAGN_vlong         0.016279   0.005167   3.150 0.001657 **
## carbonylO_sulfur_medshort     0.067980   0.023919   2.842 0.004530 **
## aliph1HC_bbC_medshort         0.054798   0.014515   3.775 0.000165 ***
## aromaticC_bbO_vlong           0.009477   0.002503   3.786 0.000158 ***
## hydroxylO_bbC_medlong         0.031438   0.007389   4.255 2.20e-05 ***
## aromaticC_hydroxylO_short    -0.021857   0.012057  -1.813 0.070026 .
## carboxylC_bbO_medlong         0.074944   0.018545   4.041 5.53e-05 ***
## bbProN_bbCA_medshort          0.051779   0.022096   2.343 0.019216 *
## carbonylC_scAGN_long          0.055106   0.019164   2.876 0.004079 **
## bbN_bbCA_long                -0.014252   0.003733  -3.818 0.000139 ***
## aromaticC_bbO_vshort         -0.172256   0.036952  -4.662 3.36e-06 ***
## carbonylC_carboxylO_short     0.209093   0.043206   4.839 1.41e-06 ***
## carbonylC_bbCA_medshort       0.046890   0.012774   3.671 0.000249 ***
## aliph1HC_aliph3HC_medlong    -0.057652   0.014674  -3.929 8.84e-05 ***
## scAGN_carboxylO_vshort       -0.161624   0.048607  -3.325 0.000901 ***
```

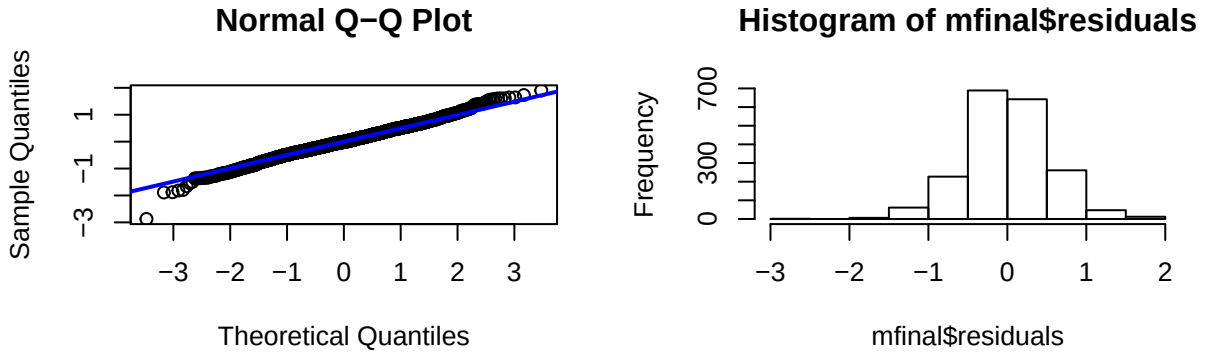
```
## sulfur_sulfur_vlong           0.107797   0.038941   2.768 0.005693 **
## aliph1HC_bbProN_medlong       0.253891   0.046763   5.429 6.40e-08 ***
## aliph1HC_bbN_medshort        -0.062378   0.015463  -4.034 5.71e-05 ***
## aromaticC_aromaticC_vlong    -0.009547   0.002331  -4.096 4.38e-05 ***
## aliph2HC_aliph3HC_vlong       0.018129   0.004401   4.119 3.98e-05 ***
## aliph3HC_bbCA_short           0.079538   0.022277   3.570 0.000365 ***
## aromaticC_bbO_medlong         0.011696   0.003969   2.947 0.003248 **
## aliph3HC_carboxylO_long      -0.076702   0.027012  -2.840 0.004566 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5429 on 1861 degrees of freedom
## Multiple R-squared:  0.8702, Adjusted R-squared:  0.8643
## F-statistic: 148.5 on 84 and 1861 DF,  p-value: < 2.2e-16
```

```
#Model Assumption Check on mfinal
par(mfrow=c(2,2))
```

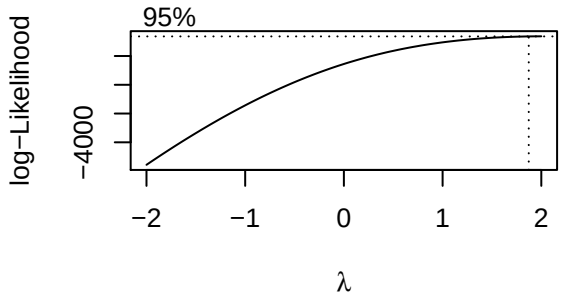



```
plot(mfinal$fitted.values,mfinal$residuals, xlab="Fitted values",
ylab="Residuals", main="Residuals vs fitted values")
plot(1:nrow(protein1),mfinal$residuals, xlab="Index",
ylab="Residuals", main="Residuals vs index")
qqnorm(mfinal$residuals)
qqline(mfinal$residuals, col="blue", lwd=2)
hist ( mfinal$residuals)
```



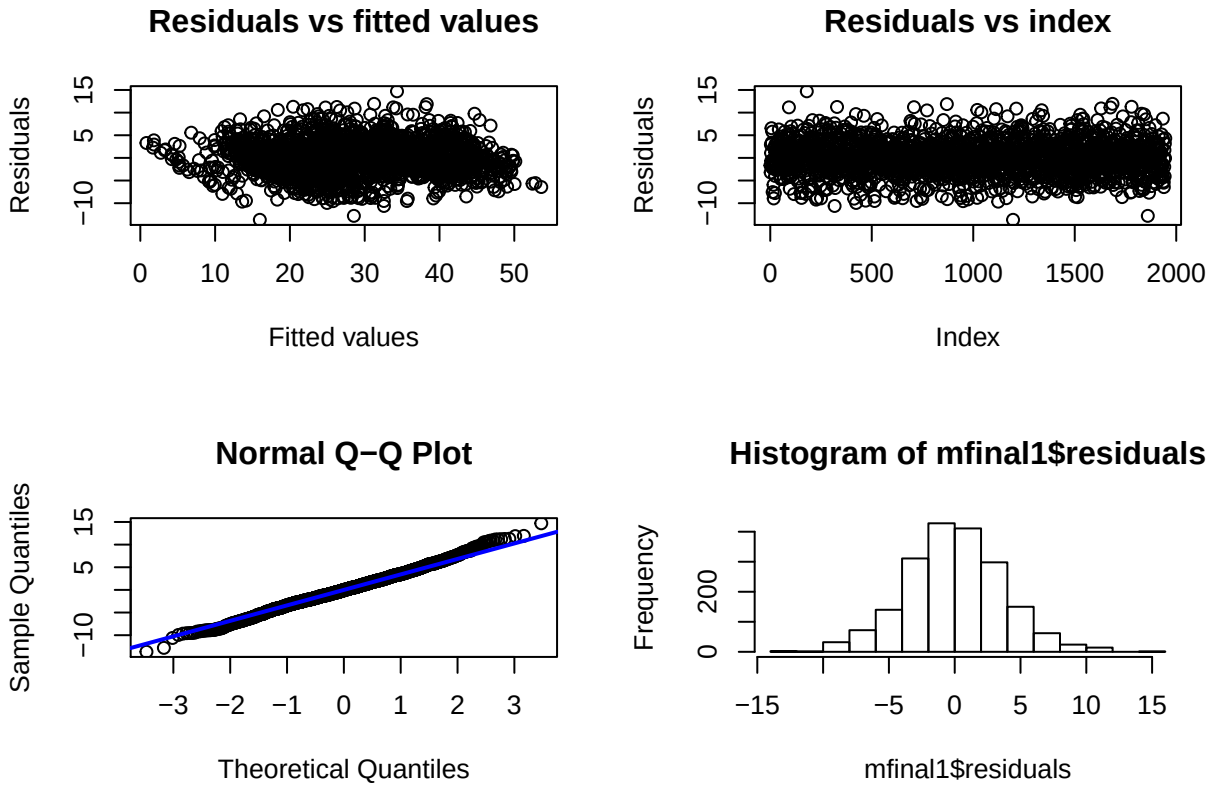



```
#Box-Cox Transformation
colz <- names(mfinal$coefficients)
colz <- colz[2:length(colz)]
mfinal.ind <- match(colz, colnames(protein1))
mfinal.ind <- append(1, mfinal.ind)

bc <- boxcox(mfinal)
lambda <- bc$x[which.max(bc$y)]
mfinal1 <- lm( (accuracy^lambda - 1)/lambda ~., data=protein1[,mfinal.ind])

#Model Assumption Check on mfinal1
par(mfrow=c(2,2))
```




```
plot(mfinal1$fitted.values,mfinal1$residuals, xlab="Fitted values",
ylab="Residuals", main="Residuals vs fitted values")
plot(1:nrow(protein1),mfinal1$residuals, xlab="Index",
ylab="Residuals", main="Residuals vs index")
qqnorm(mfinal1$residuals)
qqline(mfinal1$residuals, col="blue", lwd=2)
```

```r
hist ( mfinal1$residuals)
```




```r
summary(mfinal1)
```

```
##
## Call:
## lm(formula = (accuracy^lambda - 1)/lambda ~ ., data = protein1[,
##     mfinal.ind])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.6683  -2.3080  -0.0632   2.3044  14.7033
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -10.55805    2.61046  -4.045 5.46e-05 ***
## aliph1HC_aliph2HC_long          0.39777    0.05149   7.726 1.80e-14 ***
## scLysN_bbC_vlong               -0.77263    0.11453  -6.746 2.02e-11 ***
## aliph2HC_bbN_medshort          -0.26347    0.04040  -6.522 8.93e-11 ***
## aliph1HC_aliph2HC_medlong       0.36403    0.05773   6.306 3.58e-10 ***
## bbC_bbC_medshort               -0.29534    0.06079  -4.858 1.29e-06 ***
## aromaticC_sulfur_short         -0.23550    0.10403  -2.264 0.023703 *
## aromaticC_hydroxylO_medlong    -0.34871    0.04667  -7.472 1.21e-13 ***
## aliph1HC_aromaticC_medshort     0.58187    0.07333   7.935 3.61e-15 ***
## carbonylC_aromaticC_short       1.27262    0.19698   6.461 1.33e-10 ***
## aliph1HC_aromaticC_vlong        0.61366    0.05543  11.070  < 2e-16 ***
## bbN_bbCA_medlong               -0.26859    0.02968  -9.050  < 2e-16 ***
## aliph1HC_aromaticC_long         0.63528    0.05449  11.658  < 2e-16 ***
## aliph1HC_aliph1HC_vlong         1.06365    0.16341   6.509 9.70e-11 ***
```

```
## aliph1HC_bbN_vlong            -0.26904    0.05148  -5.226 1.93e-07 ***
## scAGN_bbN_long                0.31741    0.04079   7.781 1.18e-14 ***
## carboxylO_bbC_vlong           0.30787    0.05692   5.409 7.17e-08 ***
## aromaticC_hydroxylO_long     -0.28274    0.03433  -8.235 3.33e-16 ***
## bbN_bbN_medlong              -0.24233    0.04953  -4.893 1.08e-06 ***
## sulfur_bbC_medlong           -0.30002    0.06748  -4.446 9.25e-06 ***
## aliph2HC_aromaticC_vlong      0.11350    0.01786   6.357 2.59e-10 ***
## aliph3HC_bbC_vlong            0.15585    0.02869   5.432 6.32e-08 ***
## aliph1HC_bbO_long            -0.28266    0.06165  -4.585 4.84e-06 ***
## aliph1HC_aromaticC_medlong    0.52340    0.06053   8.647  < 2e-16 ***
## aliph3HC_bbN_short            0.75269    0.12960   5.808 7.43e-09 ***
## aliph2HC_scArgN_vlong         0.44922    0.15155   2.964 0.003074 **
## aliph3HC_aromaticC_long      -0.08057    0.02604  -3.094 0.002003 **
## aromaticC_bbCA_vlong          0.06695    0.01880   3.561 0.000378 ***
## bbO_bbO_short                -0.27893    0.08031  -3.473 0.000526 ***
## carbonylC_aromaticC_long      0.14859    0.04793   3.100 0.001963 **
## carboxylC_scLysN_vlong        1.04321    0.27383   3.810 0.000144 ***
## sulfur_bbC_vlong             -0.28836    0.04788  -6.022 2.07e-09 ***
## aromaticC_sulfur_long         0.30625    0.06810   4.497 7.31e-06 ***
## aliph2HC_bbN_medlong          0.15647    0.02452   6.382 2.20e-10 ***
## scAGN_carbonylO_medshort     -0.56403    0.11494  -4.907 1.00e-06 ***
## bbO_bbO_long                 -0.11131    0.03223  -3.454 0.000565 ***
## aliph1HC_aliph1HC_long        1.35104    0.22842   5.915 3.94e-09 ***
## aliph1HC_scArgN_long          4.77619    0.58320   8.190 4.81e-16 ***
## aliph1HC_aromaticC_short      0.92454    0.14580   6.341 2.86e-10 ***
## carboxylC_aromaticC_long      0.30161    0.09989   3.019 0.002568 **
## aromaticC_hydroxylO_medshort -0.31168    0.06711  -4.644 3.65e-06 ***
## carbonylC_bbC_medlong        -0.28573    0.05994  -4.767 2.02e-06 ***
## hydroxylO_carbonylO_medlong   0.31490    0.07592   4.148 3.51e-05 ***
## aliph3HC_hydroxylO_long       0.34289    0.07005   4.895 1.07e-06 ***
## bbN_bbN_medshort             -0.64549    0.07204  -8.960  < 2e-16 ***
## carboxylC_bbC_medlong         0.33229    0.12620   2.633 0.008533 **
## aliph2HC_aromaticC_medshort   0.13385    0.03094   4.326 1.60e-05 ***
## scArgN_bbO_medlong            2.40861    0.53309   4.518 6.63e-06 ***
## carboxylC_carboxylC_vlong    -1.18886    0.25158  -4.726 2.47e-06 ***
## scArgN_carboxylO_long        -0.93417    0.15206  -6.144 9.85e-10 ***
## carbonylC_bbProN_medlong     -1.07902    0.19220  -5.614 2.27e-08 ***
## scAGN_bbCA_medlong            0.21304    0.05615   3.794 0.000153 ***
## aliph3HC_aromaticC_vlong     -0.09019    0.02567  -3.513 0.000454 ***
## carboxylO_bbN_vlong           0.23526    0.05682   4.141 3.62e-05 ***
## scLysN_carboxylO_long         0.92755    0.18542   5.002 6.20e-07 ***
## bbCA_bbO_vshort              -2.47416    0.45937  -5.386 8.12e-08 ***
## sulfur_bbCA_short            -0.32822    0.16122  -2.036 0.041912 *
## carbonylC_sulfur_short       -0.92907    0.24196  -3.840 0.000127 ***
## bbCA_bbC_vlong               -0.05839    0.02080  -2.807 0.005058 **
## aliph2HC_bbN_vlong            0.06481    0.01718   3.772 0.000167 ***
## aliph1HC_bbO_vlong            0.19583    0.04663   4.200 2.80e-05 ***
## bbProN_carboxylO_vlong        0.69230    0.25140   2.754 0.005949 **
## aromaticC_scAGN_vlong         0.13731    0.03585   3.830 0.000133 ***
## carbonylO_sulfur_medshort     0.50595    0.16596   3.049 0.002331 **
## aliph1HC_bbC_medshort         0.38097    0.10071   3.783 0.000160 ***
## aromaticC_bbO_vlong           0.06255    0.01737   3.602 0.000324 ***
## hydroxylO_bbC_medlong         0.19092    0.05127   3.724 0.000202 ***
## aromaticC_hydroxylO_short    -0.18893    0.08366  -2.258 0.024042 *
```

```
## carboxylC_bbO_medlong          0.59767    0.12867   4.645 3.64e-06 ***
## bbProN_bbCA_medshort           0.24416    0.15332   1.593 0.111427
## carbonylC_scAGN_long           0.29987    0.13297   2.255 0.024237 *
## bbN_bbCA_long                 -0.09656    0.02590  -3.728 0.000198 ***
## aromaticC_bbO_vshort          -0.98885    0.25639  -3.857 0.000119 ***
## carbonylC_carboxylO_short      1.40676    0.29979   4.693 2.90e-06 ***
## carbonylC_bbCA_medshort        0.31115    0.08863   3.511 0.000458 ***
## aliph1HC_aliph3HC_medlong     -0.43459    0.10181  -4.269 2.07e-05 ***
## scAGN_carboxylO_vshort        -0.81688    0.33726  -2.422 0.015527 *
## sulfur_sulfur_vlong            0.58701    0.27019   2.173 0.029939 *
## aliph1HC_bbProN_medlong        1.51069    0.32447   4.656 3.45e-06 ***
## aliph1HC_bbN_medshort         -0.48920    0.10729  -4.559 5.46e-06 ***
## aromaticC_aromaticC_vlong     -0.08123    0.01617  -5.023 5.58e-07 ***
## aliph2HC_aliph3HC_vlong        0.15787    0.03054   5.169 2.60e-07 ***
## aliph3HC_bbCA_short            0.29993    0.15457   1.940 0.052478 .
## aromaticC_bbO_medlong          0.05235    0.02754   1.901 0.057457 .
## aliph3HC_carboxylO_long       -0.41075    0.18742  -2.192 0.028531 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.767 on 1861 degrees of freedom
## Multiple R-squared:  0.879,  Adjusted R-squared:  0.8736
## F-statistic:   161 on 84 and 1861 DF,  p-value: < 2.2e-16
```

```
#Prediction of mfinal
protein_test <- read.csv('protein-test.csv')

pred <- predict(mfinal1, data=protein_test)
writeLines(as.character(pred), "mypreds.txt")
```