STAT 425 Final Project

Yulin Song (yulins2), section A3

Mar 17 2021

Summary Statistics & Histograms

```r
# read the table and convert all number
data <- read.csv("Food_Supply_kcal_Data.csv")
n = ncol(data)
data[,2:(n-1)] <- lapply(data[,2:(n-1)], function(data) as.numeric(data))
attach(data)

# getting the pairwise correlation of 25 diet columns and the response "death"
mod <- data[,c(2:26,28)]


#corr_mod <- cor(na.omit(mod))
#pair wise correlation
corr_mod <- cor(mod, use = "pairwise.complete.obs")
corr_mod
```

```
##                           Alcoholic.Beverages Animal.Products   Animal.fats
## Alcoholic.Beverages               1.000000000      0.58829271  0.5623080179
## Animal.Products                   0.588292714      1.00000000  0.7264887768
## Animal.fats                       0.562308018      0.72648878  1.0000000000
## Aquatic.Products..Other           0.050935754      0.01272510 -0.0007014261
## Cereals...Excluding.Beer         -0.550810617     -0.67795107 -0.5255089855
## Eggs                              0.411221328      0.64917791  0.4734109923
## Fish..Seafood                     0.061018015      0.20465453  0.0324930641
## Fruits...Excluding.Wine          -0.028713260     -0.03780897 -0.1359193738
## Meat                              0.487540406      0.84798707  0.4603831210
## Milk...Excluding.Butter           0.404324504      0.78761738  0.4843312945
## Miscellaneous                     0.099345714      0.29921340  0.0351353321
## Offals                            0.133854202      0.31846720  0.0793996121
## Oilcrops                         -0.233094673     -0.22467465 -0.2024715210
## Pulses                           -0.303441961     -0.49170052 -0.3408381122
## Spices                           -0.003244831     -0.12626730 -0.1314870303
## Starchy.Roots                    -0.106954028     -0.42495660 -0.2730718204
## Stimulants                        0.501974436      0.62400619  0.4126803909
## Sugar.Crops                      -0.157943498     -0.13481127 -0.0882848994
## Sugar...Sweeteners                0.210643800      0.40758669  0.2948227063
## Treenuts                          0.043058154      0.27468734  0.2247211661
## Vegetal.Products                 -0.588213318     -0.99999664 -0.7265434879
## Vegetable.Oils                    0.115204133      0.14979360  0.2209216266
## Vegetables                        0.028925541      0.25956366  0.0985897474
## Obesity                           0.294205606      0.55076959  0.4238989123
## Undernourished                   -0.041360497     -0.47412023 -0.3010264388
## Deaths                            0.497379068      0.50321147  0.5394109316
##                           Aquatic.Products..Other Cereals...Excluding.Beer
## Alcoholic.Beverages                  0.0509357535              -0.55081062
## Animal.Products                      0.0127250956              -0.67795107
## Animal.fats                         -0.0007014261              -0.52550899
## Aquatic.Products..Other              1.0000000000              -0.01034739
## Cereals...Excluding.Beer            -0.0103473945               1.00000000
## Eggs                                 0.0620132516              -0.42015432
## Fish..Seafood                        0.1600149392              -0.23618003
```

```
## Fruits...Excluding.Wine           -0.0442440681             -0.34614943
## Meat                               0.0482496064             -0.60778637
## Milk...Excluding.Butter           -0.0947442596             -0.45678766
## Miscellaneous                     -0.0475865980             -0.23896600
## Offals                             0.0329614249             -0.13578856
## Oilcrops                           0.0332393344             -0.03823602
## Pulses                            -0.0718093347              0.18473843
## Spices                            -0.0426828335              0.07997722
## Starchy.Roots                     -0.0504887091             -0.15009761
## Stimulants                        -0.0429660612             -0.42085445
## Sugar.Crops                       -0.0178219220              0.23120125
## Sugar...Sweeteners                 0.0118514573             -0.37530186
## Treenuts                           0.0592021077             -0.20563905
## Vegetal.Products                  -0.0127713478              0.67804820
## Vegetable.Oils                     0.0610381437             -0.31784461
## Vegetables                         0.1633346669             -0.07521957
## Obesity                           -0.1033739669             -0.54421951
## Undernourished                    -0.0977609668              0.22559371
## Deaths                            -0.0697383131             -0.38280445
##                                 Eggs Fish..Seafood Fruits...Excluding.Wine
## Alcoholic.Beverages       0.41122133  0.0610180154            -0.028713260
## Animal.Products           0.64917791  0.2046545312            -0.037808966
## Animal.fats               0.47341099  0.0324930641            -0.135919374
## Aquatic.Products..Other   0.06201325  0.1600149392            -0.044244068
## Cereals...Excluding.Beer -0.42015432 -0.2361800259            -0.346149428
## Eggs                      1.00000000  0.2014833455            -0.070561863
## Fish..Seafood             0.20148335  1.0000000000             0.082014113
## Fruits...Excluding.Wine  -0.07056186  0.0820141128             1.000000000
## Meat                      0.50591468  0.2049711579             0.004722599
## Milk...Excluding.Butter   0.45733073 -0.0807241416            -0.017398746
## Miscellaneous             0.13226105  0.3976036980             0.073627546
## Offals                    0.09397720 -0.0481221340            -0.035662578
## Oilcrops                 -0.28779612  0.3711320077             0.121884829
## Pulses                   -0.40725872 -0.2415950339             0.168120333
## Spices                   -0.02752701  0.1799514148             0.053317338
## Starchy.Roots            -0.42453240 -0.0056015700             0.203684043
## Stimulants                0.41014007  0.1280525110            -0.017417478
## Sugar.Crops              -0.13746325  0.0088074557             0.011530603
## Sugar...Sweeteners        0.42190102  0.0925780738            -0.006778714
## Treenuts                  0.32850378  0.1629732682            -0.032085228
## Vegetal.Products         -0.64925791 -0.2044891578             0.037757281
## Vegetable.Oils            0.25908762 -0.1368861383            -0.082835203
## Vegetables                0.30565413 -0.0348580961             0.033200945
## Obesity                   0.44336961 -0.0001184883             0.115835284
## Undernourished           -0.45994553 -0.2181490601            -0.075341091
## Deaths                    0.45848818 -0.1442571887            -0.035286365
##                                  Meat Milk...Excluding.Butter Miscellaneous
## Alcoholic.Beverages       0.487540406              0.40432450   0.099345714
## Animal.Products           0.847987065              0.78761738   0.299213405
## Animal.fats               0.460383121              0.48433129   0.035135332
## Aquatic.Products..Other   0.048249606             -0.09474426  -0.047586598
## Cereals...Excluding.Beer -0.607786366             -0.45678766  -0.238965997
## Eggs                      0.505914675              0.45733073   0.132261048
## Fish..Seafood             0.204971158             -0.08072414   0.397603698
```

```
## Fruits...Excluding.Wine   0.004722599           -0.01739875   0.073627546
## Meat                      1.000000000            0.44796621   0.357665096
## Milk...Excluding.Butter   0.447966210            1.00000000   0.151817449
## Miscellaneous             0.357665096            0.15181745   1.000000000
## Offals                    0.411185931            0.19198081   0.062835331
## Oilcrops                 -0.105712357           -0.34425962   0.015884876
## Pulses                   -0.437114223           -0.31700356  -0.150857790
## Spices                   -0.159363606           -0.07671262   0.074698983
## Starchy.Roots            -0.307638910           -0.42219086  -0.170606045
## Stimulants                0.482303911            0.56905310   0.363432452
## Sugar.Crops              -0.113136982           -0.11883778  -0.043045313
## Sugar...Sweeteners        0.350844210            0.30130907   0.183636234
## Treenuts                  0.156043883            0.23685002   0.031041056
## Vegetal.Products         -0.848123602           -0.78748198  -0.298660246
## Vegetable.Oils            0.093076080            0.11637463   0.002010436
## Vegetables                0.137093222            0.35324918  -0.027148450
## Obesity                   0.486710004            0.43426407   0.188741560
## Undernourished           -0.330263828           -0.34843534  -0.236391409
## Deaths                    0.336696721            0.44379718  -0.120528277
##                                Offals     Oilcrops      Pulses       Spices
## Alcoholic.Beverages       0.133854202 -0.233094673 -0.30344196 -0.003244831
## Animal.Products           0.318467200 -0.224674647 -0.49170052 -0.126267299
## Animal.fats               0.079399612 -0.202471521 -0.34083811 -0.131487030
## Aquatic.Products..Other   0.032961425  0.033239334 -0.07180933 -0.042682833
## Cereals...Excluding.Beer -0.135788560 -0.038236018  0.18473843  0.079977215
## Eggs                      0.093977203 -0.287796120 -0.40725872 -0.027527013
## Fish..Seafood            -0.048122134  0.371132008 -0.24159503  0.179951415
## Fruits...Excluding.Wine  -0.035662578  0.121884829  0.16812033  0.053317338
## Meat                      0.411185931 -0.105712357 -0.43711422 -0.159363606
## Milk...Excluding.Butter   0.191980814 -0.344259619 -0.31700356 -0.076712616
## Miscellaneous             0.062835331  0.015884876 -0.15085779  0.074698983
## Offals                    1.000000000 -0.051889615 -0.21003823 -0.163953577
## Oilcrops                 -0.051889615  1.000000000  0.04945382  0.006166422
## Pulses                   -0.210038228  0.049453823  1.00000000  0.063522744
## Spices                   -0.163953577  0.006166422  0.06352274  1.000000000
## Starchy.Roots            -0.002398781  0.189489773  0.30893487 -0.006594811
## Stimulants                0.173699535 -0.217002163 -0.38555537 -0.064301451
## Sugar.Crops               0.003055966 -0.026045487 -0.01638684  0.164949723
## Sugar...Sweeteners       -0.083942228 -0.145930500 -0.24115505 -0.002436815
## Treenuts                 -0.037130150 -0.173085425 -0.21748986  0.058165968
## Vegetal.Products         -0.318272781  0.224633111  0.49163677  0.126300348
## Vegetable.Oils           -0.130353762 -0.198111250 -0.18836472 -0.166482069
## Vegetables                0.094536138 -0.220940921 -0.20893925  0.082747892
## Obesity                  -0.031250000 -0.010461795 -0.37429377 -0.129091112
## Undernourished            0.028110877 -0.057616104  0.34779646 -0.280167774
## Deaths                   -0.011837655 -0.307018760 -0.29857700 -0.107808790
##                          Starchy.Roots  Stimulants  Sugar.Crops
## Alcoholic.Beverages       -0.106954028  0.50197444 -0.157943498
## Animal.Products           -0.424956605  0.62400619 -0.134811272
## Animal.fats               -0.273071820  0.41268039 -0.088284899
## Aquatic.Products..Other   -0.050488709 -0.04296606 -0.017821922
## Cereals...Excluding.Beer  -0.150097610 -0.42085445  0.231201248
## Eggs                      -0.424532405  0.41014007 -0.137463255
## Fish..Seafood             -0.005601570  0.12805251  0.008807456
```

```
## Fruits...Excluding.Wine    0.203684043 -0.01741748  0.011530603
## Meat                      -0.307638910  0.48230391 -0.113136982
## Milk...Excluding.Butter   -0.422190862  0.56905310 -0.118837779
## Miscellaneous             -0.170606045  0.36343245 -0.043045313
## Offals                    -0.002398781  0.17369953  0.003055966
## Oilcrops                   0.189489773 -0.21700216 -0.026045487
## Pulses                     0.308934865 -0.38555537 -0.016386839
## Spices                    -0.006594811 -0.06430145  0.164949723
## Starchy.Roots              1.000000000 -0.23628449 -0.022296486
## Stimulants                -0.236284490  1.00000000 -0.102498272
## Sugar.Crops               -0.022296486 -0.10249827  1.000000000
## Sugar...Sweeteners        -0.464920906  0.14243448 -0.194986773
## Treenuts                  -0.185501863  0.19985658 -0.107181462
## Vegetal.Products           0.424788032 -0.62363531  0.134483379
## Vegetable.Oils            -0.183794382  0.03315526 -0.097928038
## Vegetables                -0.293910148  0.26915197  0.057459310
## Obesity                   -0.336096599  0.37361638 -0.217469858
## Undernourished             0.387955279 -0.34776827 -0.004773312
## Deaths                    -0.265054651  0.36518691 -0.139776315
##                          Sugar...Sweeteners    Treenuts Vegetal.Products
## Alcoholic.Beverages             0.210643800  0.04305815      -0.58821332
## Animal.Products                 0.407586690  0.27468734      -0.99999664
## Animal.fats                     0.294822706  0.22472117      -0.72654349
## Aquatic.Products..Other         0.011851457  0.05920211      -0.01277135
## Cereals...Excluding.Beer       -0.375301859 -0.20563905       0.67804820
## Eggs                            0.421901020  0.32850378      -0.64925791
## Fish..Seafood                   0.092578074  0.16297327      -0.20448916
## Fruits...Excluding.Wine        -0.006778714 -0.03208523       0.03775728
## Meat                            0.350844210  0.15604388      -0.84812360
## Milk...Excluding.Butter         0.301309072  0.23685002      -0.78748198
## Miscellaneous                   0.183636234  0.03104106      -0.29866025
## Offals                         -0.083942228 -0.03713015      -0.31827278
## Oilcrops                       -0.145930500 -0.17308543       0.22463311
## Pulses                         -0.241155046 -0.21748986       0.49163677
## Spices                         -0.002436815  0.05816597       0.12630035
## Starchy.Roots                  -0.464920906 -0.18550186       0.42478803
## Stimulants                      0.142434485  0.19985658      -0.62363531
## Sugar.Crops                    -0.194986773 -0.10718146       0.13448338
## Sugar...Sweeteners              1.000000000  0.07073202      -0.40756371
## Treenuts                        0.070732016  1.00000000      -0.27439508
## Vegetal.Products               -0.407563709 -0.27439508       1.00000000
## Vegetable.Oils                  0.174996057  0.25064628      -0.15003766
## Vegetables                     -0.002153968  0.35217648      -0.25908307
## Obesity                         0.603417418  0.26014485      -0.55094670
## Undernourished                 -0.357125029 -0.25690059       0.47420071
## Deaths                          0.302948229  0.24332136      -0.50329697
##                          Vegetable.Oils   Vegetables       Obesity
## Alcoholic.Beverages         0.115204133  0.028925541  0.2942056061
## Animal.Products             0.149793600  0.259563664  0.5507695901
## Animal.fats                 0.220921627  0.098589747  0.4238989123
## Aquatic.Products..Other     0.061038144  0.163334667 -0.1033739669
## Cereals...Excluding.Beer   -0.317844613 -0.075219573 -0.5442195060
## Eggs                        0.259087622  0.305654131  0.4433696139
## Fish..Seafood              -0.136886138 -0.034858096 -0.0001184883
```

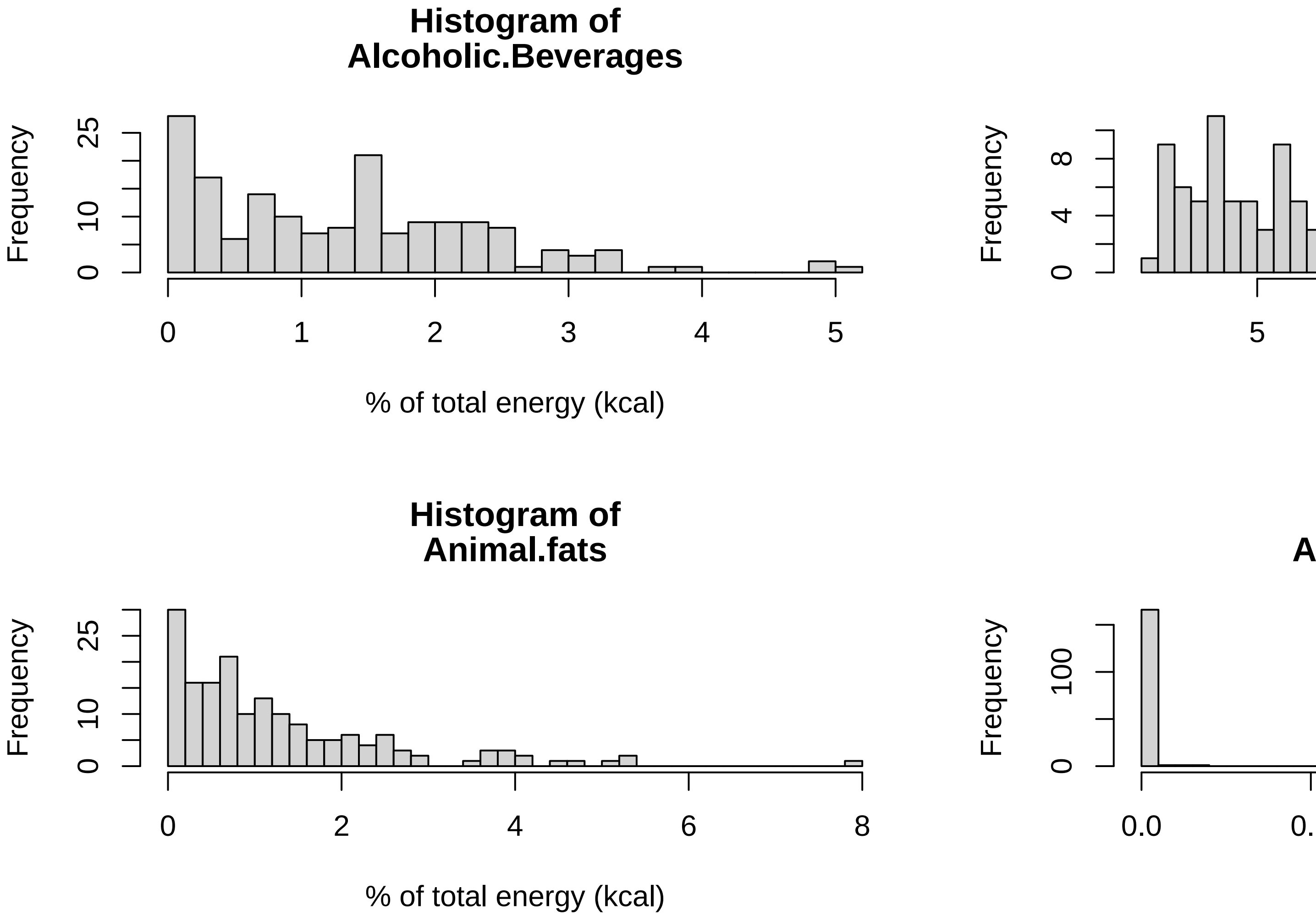
```
## Fruits...Excluding.Wine    -0.082835203  0.033200945  0.1158352841
## Meat                        0.093076080  0.137093222  0.4867100037
## Milk...Excluding.Butter     0.116374628  0.353249179  0.4342640663
## Miscellaneous               0.002010436 -0.027148450  0.1887415605
## Offals                     -0.130353762  0.094536138 -0.0312500003
## Oilcrops                   -0.198111250 -0.220940921 -0.0104617951
## Pulses                     -0.188364724 -0.208939254 -0.3742937691
## Spices                     -0.166482069  0.082747892 -0.1290911115
## Starchy.Roots              -0.183794382 -0.293910148 -0.3360965988
## Stimulants                  0.033155264  0.269151973  0.3736163754
## Sugar.Crops                -0.097928038  0.057459310 -0.2174698575
## Sugar...Sweeteners          0.174996057 -0.002153968  0.6034174175
## Treenuts                    0.250646281  0.352176475  0.2601448517
## Vegetal.Products           -0.150037664 -0.259083072 -0.5509467037
## Vegetable.Oils              1.000000000  0.034939154  0.2921374407
## Vegetables                  0.034939154  1.000000000  0.1662968927
## Obesity                     0.292137441  0.166296893  1.0000000000
## Undernourished             -0.057272681 -0.384962782 -0.5093770923
## Deaths                      0.261394981  0.181358217  0.4830768558
##                           Undernourished      Deaths
## Alcoholic.Beverages        -0.041360497  0.49737907
## Animal.Products            -0.474120232  0.50321147
## Animal.fats                -0.301026439  0.53941093
## Aquatic.Products..Other    -0.097760967 -0.06973831
## Cereals...Excluding.Beer    0.225593711 -0.38280445
## Eggs                       -0.459945525  0.45848818
## Fish..Seafood              -0.218149060 -0.14425719
## Fruits...Excluding.Wine    -0.075341091 -0.03528636
## Meat                       -0.330263828  0.33669672
## Milk...Excluding.Butter    -0.348435336  0.44379718
## Miscellaneous              -0.236391409 -0.12052828
## Offals                      0.028110877 -0.01183765
## Oilcrops                   -0.057616104 -0.30701876
## Pulses                      0.347796462 -0.29857700
## Spices                     -0.280167774 -0.10780879
## Starchy.Roots               0.387955279 -0.26505465
## Stimulants                 -0.347768266  0.36518691
## Sugar.Crops                -0.004773312 -0.13977631
## Sugar...Sweeteners         -0.357125029  0.30294823
## Treenuts                   -0.256900585  0.24332136
## Vegetal.Products            0.474200710 -0.50329697
## Vegetable.Oils             -0.057272681  0.26139498
## Vegetables                 -0.384962782  0.18135822
## Obesity                    -0.509377092  0.48307686
## Undernourished              1.000000000 -0.35217323
## Deaths                     -0.352173227  1.00000000
```

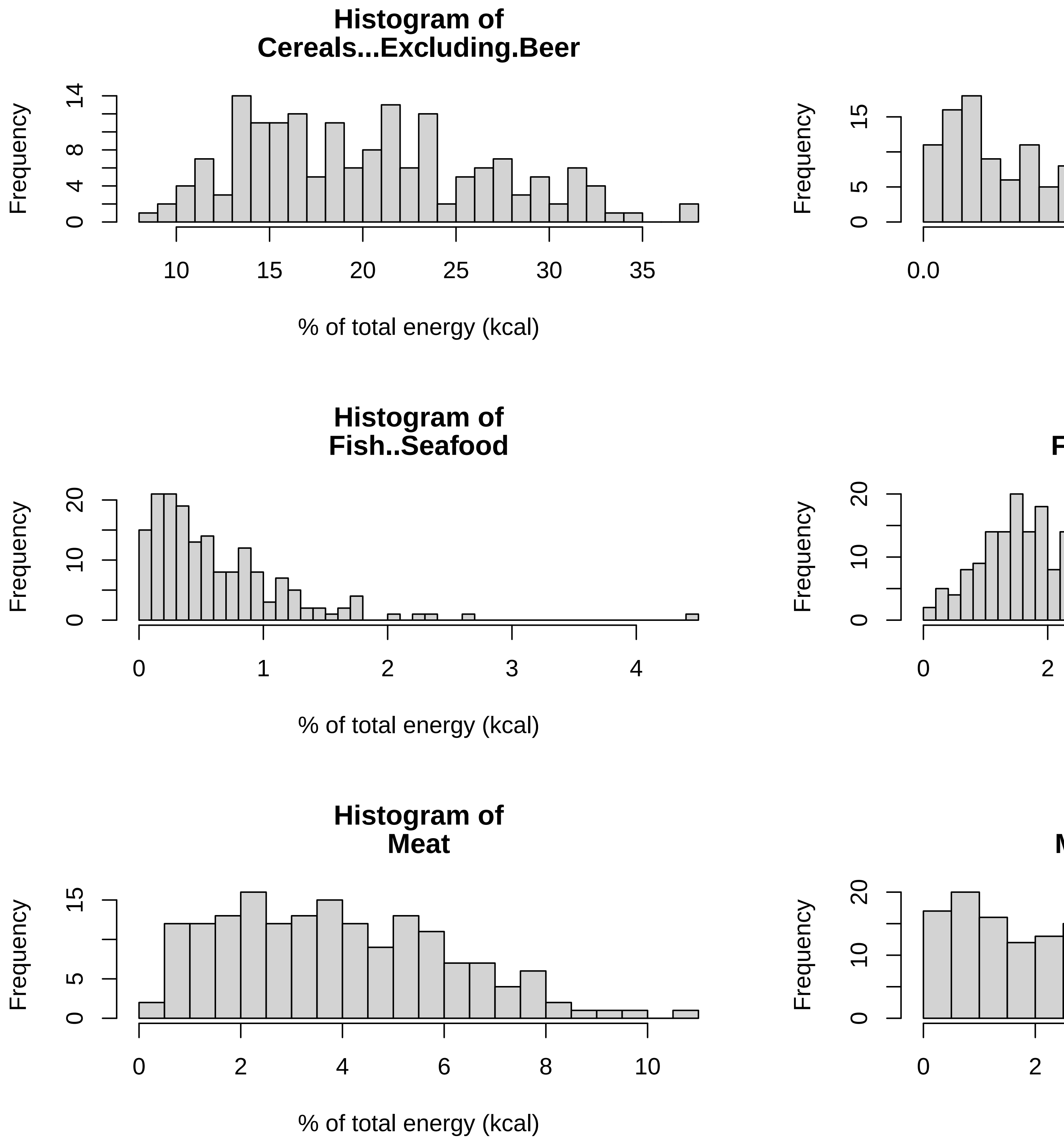
```
# summary statistics of the 26 columns (Figure 1 in report)
sum_mod <- summary(mod)
sum_mod
```

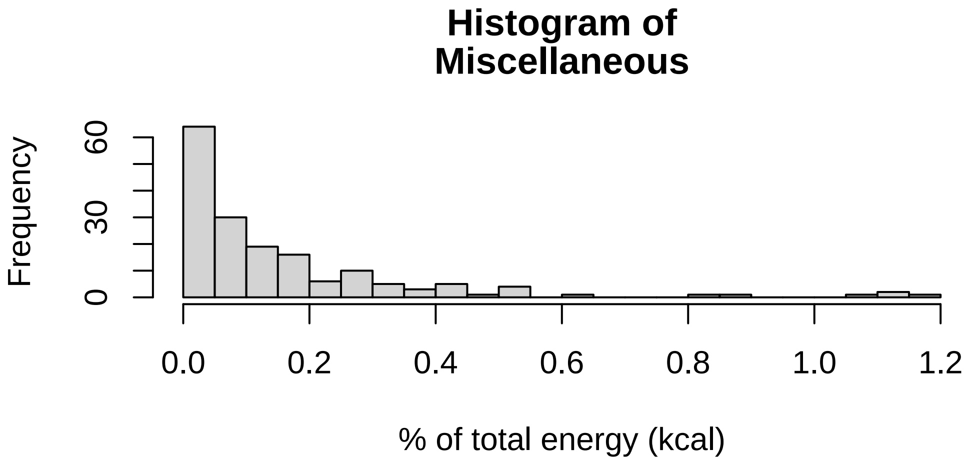
```
##  Alcoholic.Beverages Animal.Products   Animal.fats     Aquatic.Products..Other
##  Min.   :0.0000      Min.   : 1.624   Min.   :0.0000   Min.   :0.000000
##  1st Qu.:0.3613      1st Qu.: 5.083   1st Qu.:0.3428   1st Qu.:0.000000
```

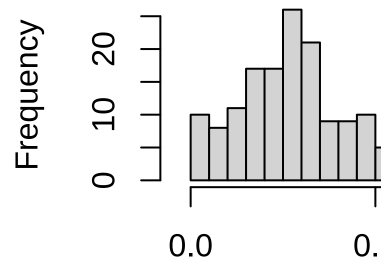
```
##  Median :1.2446      Median : 9.034   Median :0.8775   Median :0.000000
##  Mean   :1.3252      Mean   : 9.295   Mean   :1.2674   Mean   :0.002786
##  3rd Qu.:2.0280      3rd Qu.:13.175   3rd Qu.:1.7632   3rd Qu.:0.000000
##  Max.   :5.1574      Max.   :22.291   Max.   :7.8007   Max.   :0.400700
##
##  Cereals...Excluding.Beer      Eggs        Fish..Seafood
##  Min.   : 8.957           Min.   :0.0188   Min.   :0.0000
##  1st Qu.:15.306           1st Qu.:0.1410   1st Qu.:0.2402
##  Median :19.620           Median :0.4037   Median :0.4783
##  Mean   :20.365           Mean   :0.4285   Mean   :0.6315
##  3rd Qu.:24.841           3rd Qu.:0.6330   3rd Qu.:0.8697
##  Max.   :37.526           Max.   :1.4461   Max.   :4.4183
##
##  Fruits...Excluding.Wine      Meat        Milk...Excluding.Butter
##  Min.   :0.1471          Min.   : 0.298   Min.   :0.1169
##  1st Qu.:1.2245          1st Qu.: 2.081   1st Qu.:1.1078
##  Median :1.6948          Median : 3.687   Median :2.7198
##  Mean   :2.0120          Mean   : 3.896   Mean   :2.9245
##  3rd Qu.:2.3707          3rd Qu.: 5.278   3rd Qu.:4.3196
##  Max.   :8.8540          Max.   :10.567   Max.   :9.9441
##
##  Miscellaneous         Offals           Oilcrops          Pulses
##  Min.   :0.00000   Min.   :0.00000   Min.   : 0.0179   Min.   :0.0000
##  1st Qu.:0.02470   1st Qu.:0.07825   1st Qu.: 0.2993   1st Qu.:0.2967
##  Median :0.08805   Median :0.11825   Median : 0.6363   Median :0.7084
##  Mean   :0.15933   Mean   :0.14122   Mean   : 1.1035   Mean   :1.1089
##  3rd Qu.:0.19173   3rd Qu.:0.17663   3rd Qu.: 1.1902   3rd Qu.:1.5472
##  Max.   :1.18220   Max.   :0.80150   Max.   :10.4822   Max.   :7.5638
##
##      Spices        Starchy.Roots      Stimulants       Sugar.Crops
##  Min.   :0.00000   Min.   : 0.2938   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.03635   1st Qu.: 1.1123   1st Qu.:0.07765   1st Qu.:0.00000
##  Median :0.08590   Median : 1.5449   Median :0.20675   Median :0.00000
##  Mean   :0.18320   Mean   : 3.0839   Mean   :0.30537   Mean   :0.01788
##  3rd Qu.:0.22798   3rd Qu.: 2.9245   3rd Qu.:0.42080   3rd Qu.:0.00000
##  Max.   :1.22020   Max.   :19.6759   Max.   :2.00900   Max.   :0.59300
##
##  Sugar...Sweeteners    Treenuts       Vegetal.Products Vegetable.Oils
##  Min.   :0.6786     Min.   :0.00000   Min.   :27.71    Min.   : 0.9325
##  1st Qu.:3.4222     1st Qu.:0.04662   1st Qu.:36.83    1st Qu.: 3.1263
##  Median :4.6784     Median :0.17400   Median :40.97    Median : 4.6607
##  Mean   :4.8212     Mean   :0.26162   Mean   :40.71    Mean   : 4.8724
##  3rd Qu.:6.3458     3rd Qu.:0.38958   3rd Qu.:44.94    3rd Qu.: 6.4279
##  Max.   :9.5492     Max.   :1.42100   Max.   :48.39    Max.   :10.3839
##
##    Vegetables        Obesity      Undernourished      Deaths
##  Min.   :0.0957   Min.   : 2.10   Min.   : 2.50   Min.   :0.000000
##  1st Qu.:0.6026   1st Qu.: 8.50   1st Qu.: 5.70   1st Qu.:0.002013
##  Median :1.0031   Median :21.20   Median : 9.90   Median :0.011998
##  Mean   :1.0863   Mean   :18.71   Mean   :14.46   Mean   :0.039370
##  3rd Qu.:1.3670   3rd Qu.:25.70   3rd Qu.:18.95   3rd Qu.:0.069503
##  Max.   :3.3524   Max.   :45.60   Max.   :59.60   Max.   :0.185428
##                   NA's   :3       NA's   :51      NA's   :6
```

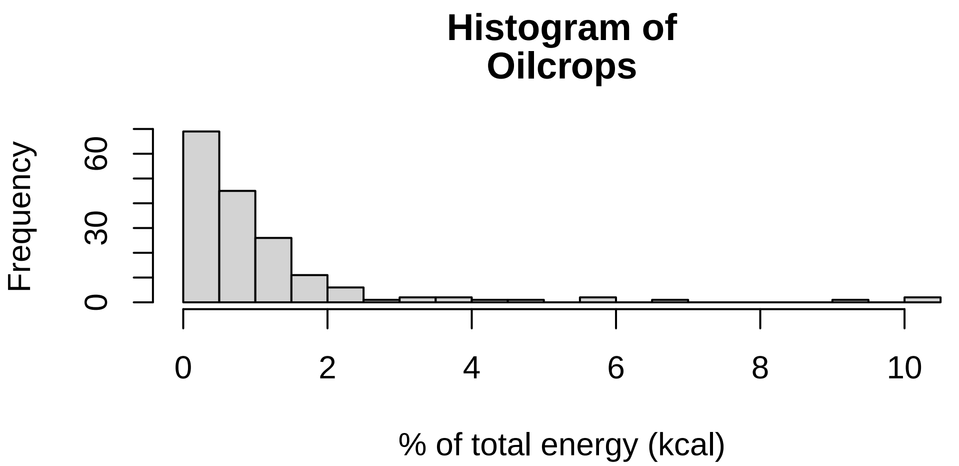
```r
# histogram of the 26 columns (Figure 2 in report)
n_breaks = 35
for (i in 1:23) {
hist(mod[,i],main = c("Histogram of",colnames(mod[i])),
     xlab = "% of total energy (kcal)", breaks = n_breaks)
}
```

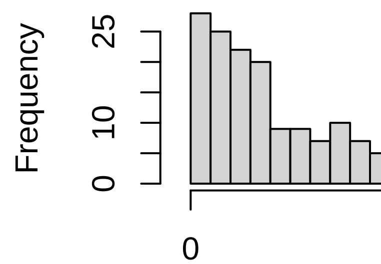


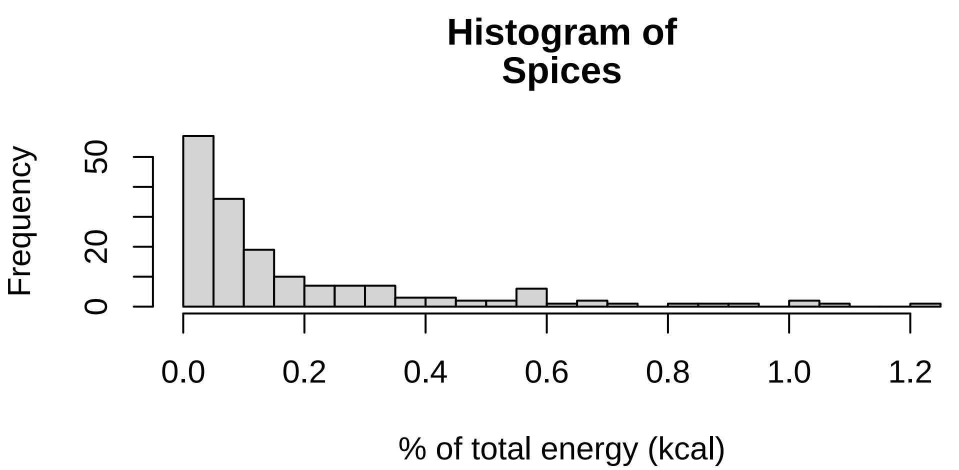

Histogram of
Cereals...Excluding.Beer
Frequency
% of total energy (kcal)
Histogram of
Fish..Seafood
Frequency
% of total energy (kcal)
Histogram of
Meat
Frequency
% of total energy (kcal)
Frequency
Frequency
Frequency

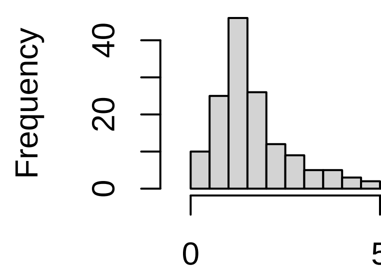

Histogram of
Miscellaneous
Frequency
0
30
60
0.0
0.2
0.4
0.6
0.8
1.0
1.2
% of total energy (kcal)

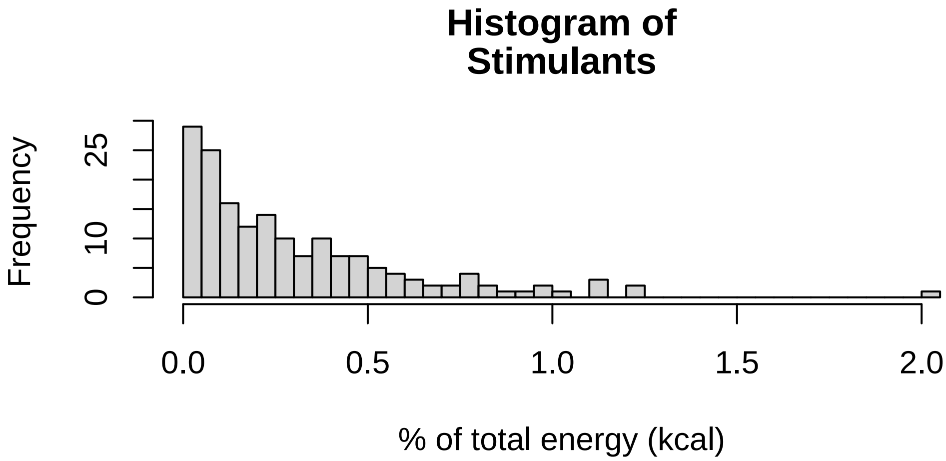

Frequency
0
10
20
0.0

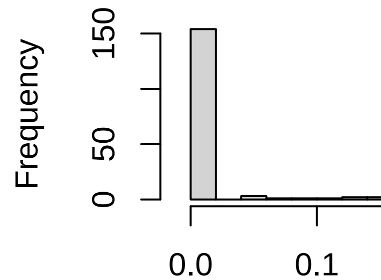

Histogram of
Oilcrops
Frequency
0
30
60
0
2
4
6
8
10
% of total energy (kcal)

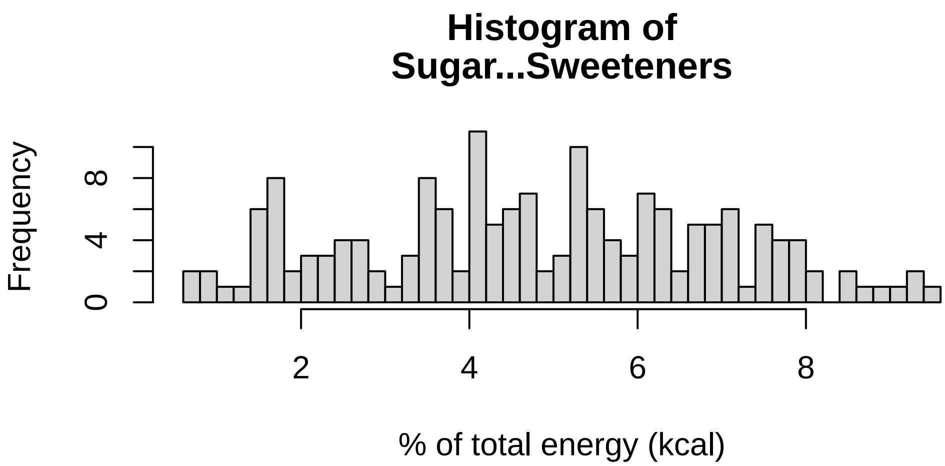

Frequency
0
10
25
0

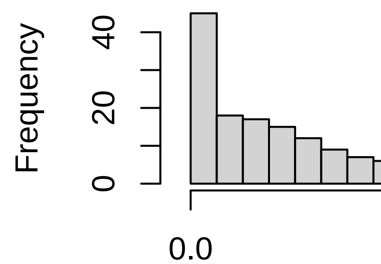

Histogram of
Spices
Frequency
0
20
50
0.0
0.2
0.4
0.6
0.8
1.0
1.2
% of total energy (kcal)

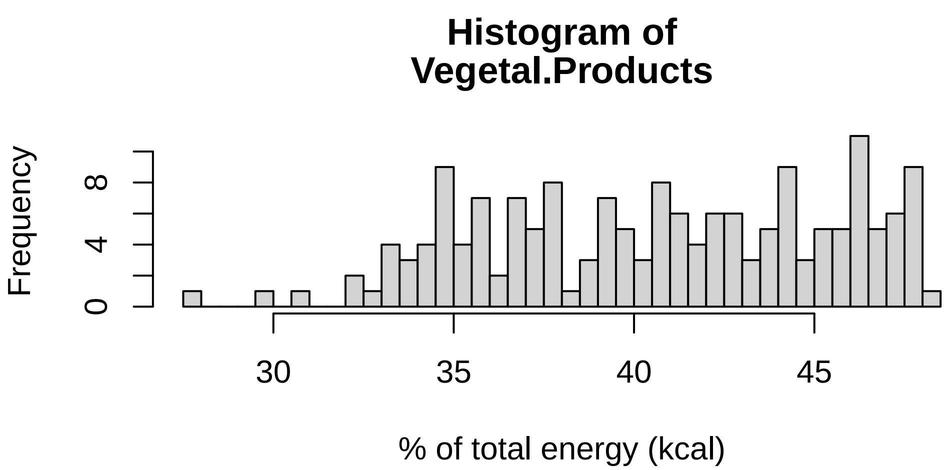

Frequency
0
20
40
0

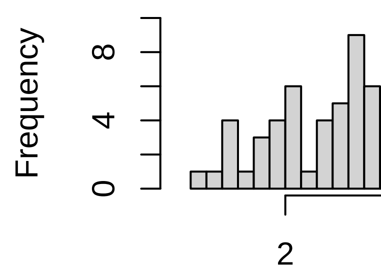

Histogram of
Stimulants
Frequency
0
10
25
0.0
0.5
1.0
1.5
2.0
% of total energy (kcal)

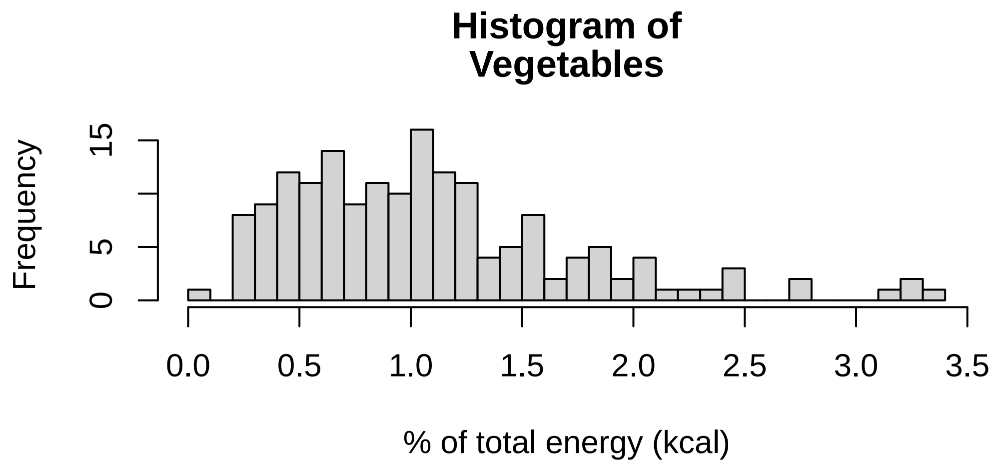

Frequency
0
50
150
0.0
0.1


Histogram of
Sugar...Sweeteners
Frequency
0
4
8
2
4
6
8
% of total energy (kcal)

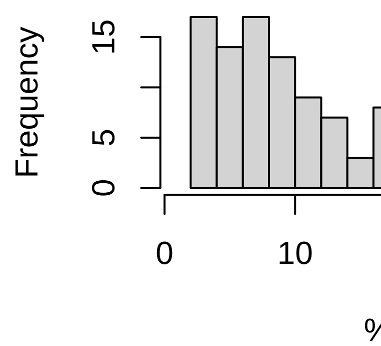

Frequency
0
20
40
0.0

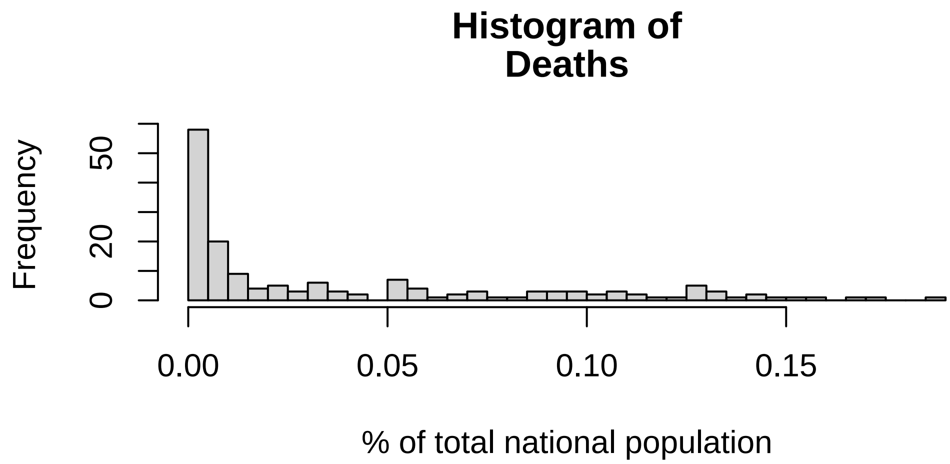

Histogram of
Vegetal.Products
Frequency
0
4
8
30
35
40
45
% of total energy (kcal)


Frequency
0
4
8
2



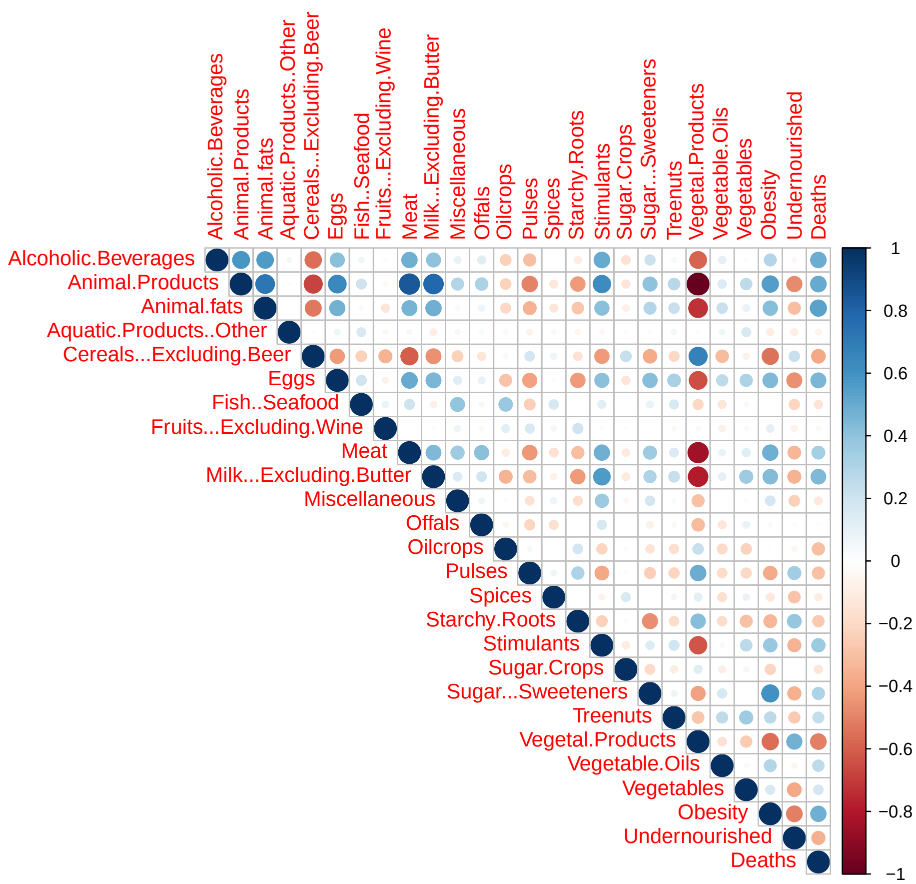

```
for (i in 24:26) {
hist(mod[,i],main = c("Histogram of",colnames(mod[i])),
     xlab = "% of total national population", breaks = n_breaks)
}
```





Correlation Plot and High correlation entries

```
# pairwise correlation Plot (Figure 3 in report)
library(corrplot)
corrplot_mod <- corrplot(corr_mod, type = 'upper')
```



```r
# which column & rows in corr_mod where absolute value of correlation happens to be >= 0.75

## indices of the above criteria happens
## (counted in each column from top to bottom and then across columns from left to right)

(highcorr_indices <- which(abs(corr_mod) >= 0.75 & corr_mod < 1))
```

```
##  [1]  35  36  47 210 229 236 255 522 529 530
```

```r
(columns_length <- length(corr_mod[1,]))
```

```
## [1] 26
```

```r
## "which rows" is calculated by:
## ind mod col_length
(rows_indices <- highcorr_indices %% columns_length)
```

```
##  [1]  9 10 21  2 21  2 21  2  9 10
```

```r
## "which columns" is calculated by:
(cols_indices <- 1 + ((highcorr_indices - rows_indices) / columns_length))
```

```
##  [1]  2  2  2  9  9 10 10 21 21 21
```

```r
# correlation between Meat and Animal Products
corr_mod[9,2]
```

```
## [1] 0.8479871
```

```r
# correlation between Vegetal Products and Animal Products
corr_mod[21,2]
```

```
## [1] -0.9999966
```

```r
# correlation between Vegetal Products and Meat
corr_mod[21,9]
```

```
## [1] -0.8481236
```

```r
# in corr_plot, (9,2) and (2,9) corresponds to Meat&Animal Products; (high positive correlation)
# (21,2) and (2,21) corresponds to Vegetal Products&Animal Products; (high negative correlation)
# (21,9) and (9,21) corresponds to Vegetal Products & Meat. (high positive correlation)
# These three pairs have pairwise high correlation
```

Multiple Linear Regression

```r
# multiple linear regression using all predictors
############## We will use full model firstly to predict, name (Full Model) in report.

#too much missing of Undernourished, not appropriate to drop all of the
#observations that just missing of Undernourished
#51/170 missing proportion is about 30% which is too high, we can just drop it
mod$Undernourished <- NULL

#then we can remove missing values of observations
mod2 <- na.omit(mod)
#now we have 163 data only 7 removed from original data, if we just remove
#all missing values, we only get 112 observations, we would lost information from over 50 observations
#which is not appropriate, so drop Undernourished first is a better choice
dim(mod2)
```

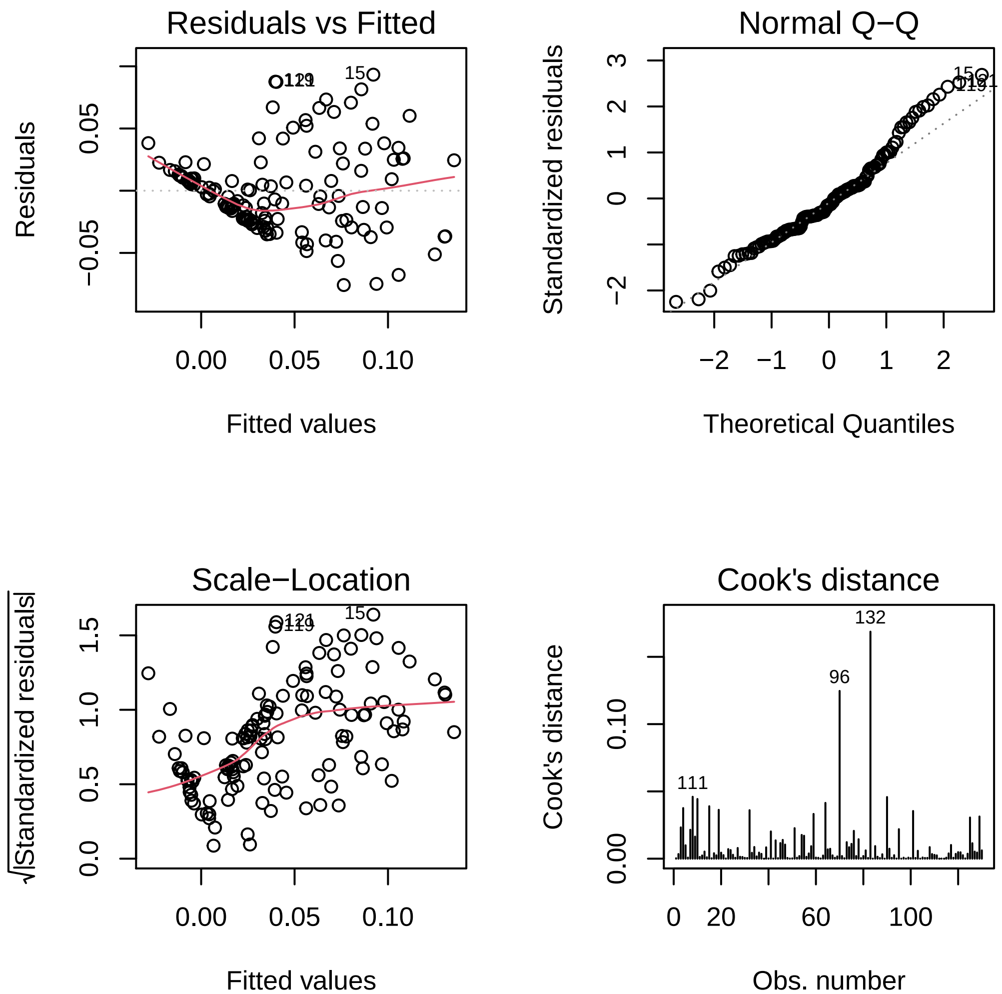
```
## [1] 163  25
```

```r
set.seed(1)

id <- sample(1:nrow(mod2), 0.8*nrow(mod2))
traindata <- mod2[id, ]
testdata <- mod2[-id,]



full_model <- lm(Deaths ~ ., data = traindata)
summary(full_model)
```

```
##
## Call:
## lm(formula = Deaths ~ ., data = traindata)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.075835 -0.023405 -0.004554  0.015912  0.093274
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -4.955e+03  2.914e+03  -1.700   0.0921 .
## Alcoholic.Beverages         4.964e+01  2.918e+01   1.701   0.0919 .
## Animal.Products             4.970e+01  2.914e+01   1.705   0.0911 .
## Animal.fats                 4.940e+01  2.915e+01   1.695   0.0931 .
## Aquatic.Products..Other     4.852e+01  2.920e+01   1.662   0.0995 .
## Cereals...Excluding.Beer    4.962e+01  2.918e+01   1.701   0.0920 .
## Eggs                        4.941e+01  2.915e+01   1.695   0.0930 .
## Fish..Seafood               4.939e+01  2.915e+01   1.694   0.0932 .
## Fruits...Excluding.Wine     4.962e+01  2.918e+01   1.701   0.0920 .
## Meat                        4.940e+01  2.915e+01   1.695   0.0931 .
## Milk...Excluding.Butter     4.940e+01  2.915e+01   1.695   0.0931 .
## Miscellaneous               4.956e+01  2.918e+01   1.699   0.0923 .
## Offals                      4.940e+01  2.915e+01   1.695   0.0931 .
## Oilcrops                    4.962e+01  2.918e+01   1.701   0.0920 .
```

```
## Pulses                     4.962e+01  2.918e+01   1.701   0.0920 .
## Spices                     4.962e+01  2.918e+01   1.701   0.0920 .
## Starchy.Roots              4.962e+01  2.918e+01   1.701   0.0920 .
## Stimulants                 4.964e+01  2.918e+01   1.701   0.0919 .
## Sugar.Crops                4.961e+01  2.918e+01   1.700   0.0921 .
## Sugar...Sweeteners         4.962e+01  2.918e+01   1.701   0.0920 .
## Treenuts                   4.966e+01  2.918e+01   1.702   0.0918 .
## Vegetal.Products           4.947e+01  2.911e+01   1.700   0.0922 .
## Vegetable.Oils             4.962e+01  2.918e+01   1.701   0.0920 .
## Vegetables                 4.962e+01  2.918e+01   1.700   0.0920 .
## Obesity                    8.780e-04  6.079e-04   1.444   0.1517
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03737 on 105 degrees of freedom
## Multiple R-squared:  0.5607, Adjusted R-squared:  0.4603
## F-statistic: 5.584 on 24 and 105 DF,  p-value: 2.43e-10
```

```
#too much insignificant predictors and model diagnostics:

par(mfrow = c(2,2))
plot(full_model,1:4)
```

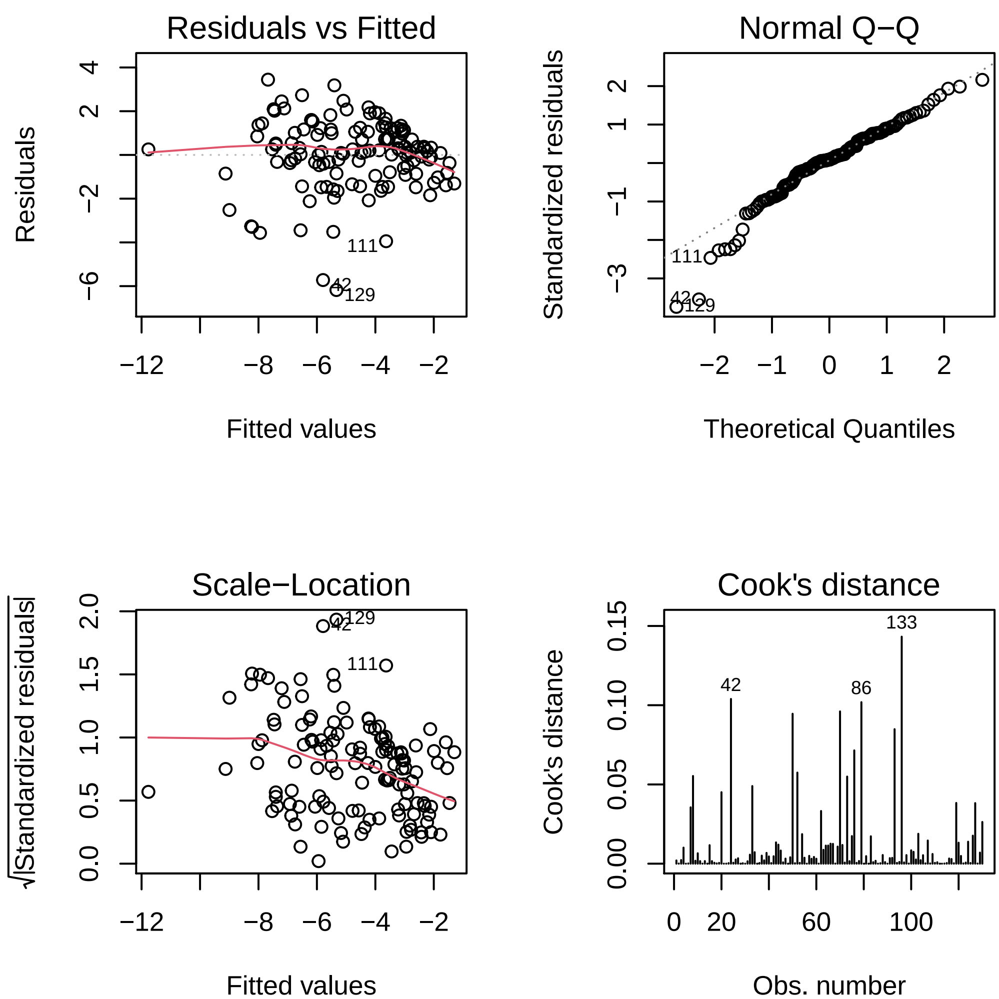



```
par(mfrow = c(1,1))

#Check model assumptions:

#1) independent assumption: the points are randomly distributed around the zero mean line, it is true.

#2) linearity assumption: the residuals plot shows there is no special curve, the linearity asumption i

#3) constant variance assumption: the residuals plot also shows  the spread of residuals changes across


#4) normality assumption: the normal qq plot shows that some outliers far from  the line at the two tai


#use transformation
```

```r
#add 0.00001 to avod log(0)

log_full_model <- lm( log(Deaths + 0.00001)  ~ ., data = traindata)
summary(log_full_model)
```

```
##
## Call:
## lm(formula = log(Deaths + 1e-05) ~ ., data = traindata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.1799 -0.8408  0.1472  1.0719  3.4416
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1.618e+05  1.383e+05  -1.170    0.245
## Alcoholic.Beverages       1.623e+03  1.385e+03   1.172    0.244
## Animal.Products           1.599e+03  1.383e+03   1.156    0.250
## Animal.fats               1.637e+03  1.383e+03   1.183    0.239
## Aquatic.Products..Other   1.566e+03  1.385e+03   1.130    0.261
## Cereals...Excluding.Beer  1.622e+03  1.385e+03   1.172    0.244
## Eggs                      1.638e+03  1.383e+03   1.185    0.239
## Fish..Seafood             1.636e+03  1.383e+03   1.183    0.240
## Fruits...Excluding.Wine   1.623e+03  1.385e+03   1.172    0.244
## Meat                      1.637e+03  1.383e+03   1.183    0.239
## Milk...Excluding.Butter   1.637e+03  1.383e+03   1.184    0.239
## Miscellaneous             1.622e+03  1.384e+03   1.172    0.244
## Offals                    1.639e+03  1.383e+03   1.185    0.239
## Oilcrops                  1.622e+03  1.385e+03   1.172    0.244
## Pulses                    1.622e+03  1.385e+03   1.172    0.244
## Spices                    1.622e+03  1.385e+03   1.171    0.244
## Starchy.Roots             1.622e+03  1.385e+03   1.172    0.244
## Stimulants                1.623e+03  1.385e+03   1.172    0.244
## Sugar.Crops               1.621e+03  1.385e+03   1.171    0.244
## Sugar...Sweeteners        1.623e+03  1.385e+03   1.172    0.244
## Treenuts                  1.624e+03  1.385e+03   1.173    0.243
## Vegetal.Products          1.613e+03  1.381e+03   1.168    0.245
## Vegetable.Oils            1.623e+03  1.385e+03   1.172    0.244
## Vegetables                1.622e+03  1.385e+03   1.172    0.244
## Obesity                   3.504e-02  2.885e-02   1.215    0.227
##
## Residual standard error: 1.773 on 105 degrees of freedom
## Multiple R-squared:  0.6047, Adjusted R-squared:  0.5143
## F-statistic: 6.692 on 24 and 105 DF,  p-value: 2.134e-12
```

```r
#too much insignificant predictors and model diagnostics:

par(mfrow = c(2,2))
plot(log_full_model,1:4)
```




```
par(mfrow = c(1,1))

#assumptions much better, no serious issues now, but we need perform model selections

# Before that, do prediction

# but before that, we will try to predict

pred_1 <- predict(log_full_model, testdata)
pred_1 <- exp(pred_1) -  1e-05
mse_1 <- mean((pred_1 - testdata$Deaths)^2)
mse_1
```

```
## [1] 0.0417249
```

```
aic <- step(log_full_model, trace = 0)
bic <-  step(log_full_model, trace = 0, k = log(nrow(traindata)))
summary(aic)
```

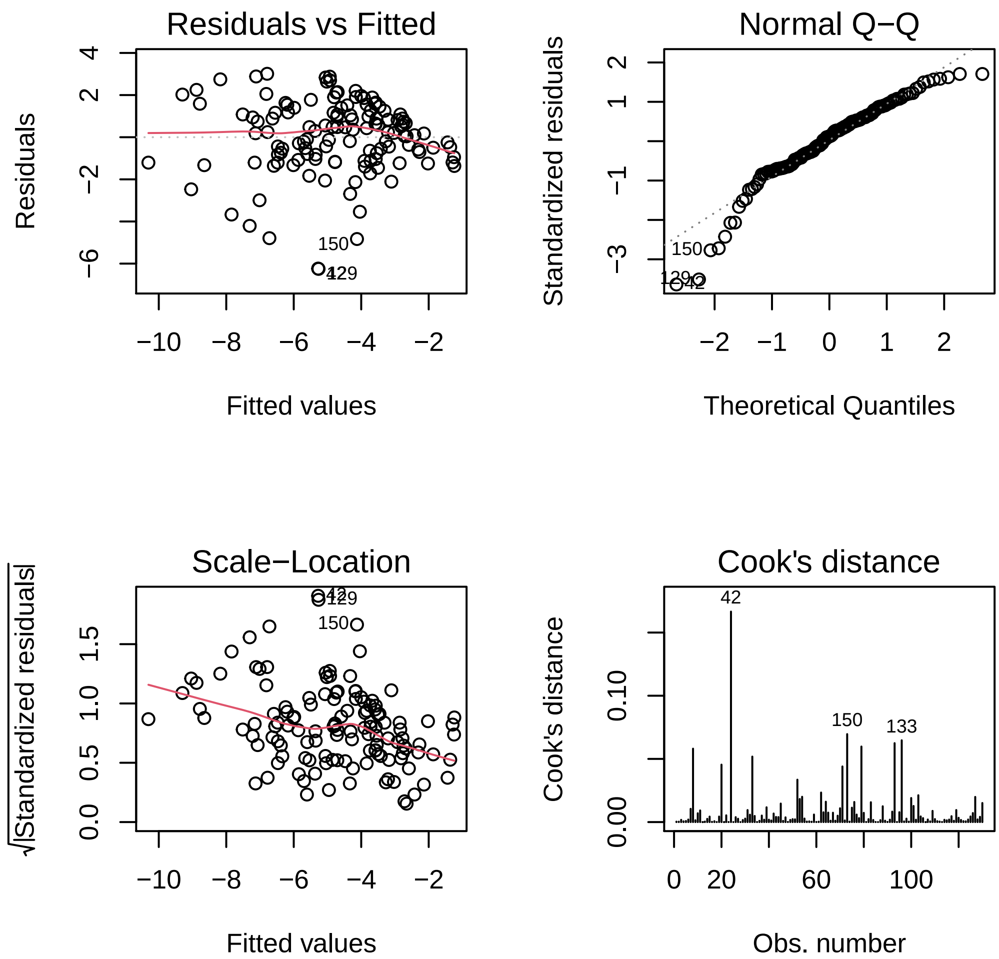
```
##
## Call:
## lm(formula = log(Deaths + 1e-05) ~ Alcoholic.Beverages + Animal.fats +
##     Cereals...Excluding.Beer + Eggs + Fish..Seafood + Fruits...Excluding.Wine +
##     Meat + Milk...Excluding.Butter + Miscellaneous + Offals +
##     Oilcrops + Pulses + Spices + Starchy.Roots + Stimulants +
##     Sugar.Crops + Sugar...Sweeteners + Treenuts + Vegetable.Oils +
##     Vegetables, data = traindata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0705 -0.7834  0.1523  1.0665  3.3911
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1167.12     699.53  -1.668   0.0981 .
## Alcoholic.Beverages         23.72      13.99   1.695   0.0929 .
## Animal.fats                 23.31      14.00   1.664   0.0989 .
## Cereals...Excluding.Beer    23.21      13.99   1.659   0.0999 .
## Eggs                        24.70      13.96   1.769   0.0797 .
## Fish..Seafood               22.10      14.02   1.577   0.1177
## Fruits...Excluding.Wine     23.38      13.98   1.672   0.0974 .
## Meat                        22.90      13.98   1.638   0.1043
## Milk...Excluding.Butter     23.44      13.98   1.676   0.0966 .
## Miscellaneous               22.95      14.19   1.617   0.1088
## Offals                      24.80      14.30   1.734   0.0858 .
## Oilcrops                    22.94      14.01   1.637   0.1045
## Pulses                      23.09      13.99   1.651   0.1017
## Spices                      22.58      14.03   1.609   0.1105
## Starchy.Roots               23.16      13.99   1.656   0.1007
## Stimulants                  23.78      14.04   1.694   0.0931 .
## Sugar.Crops                 20.75      13.88   1.495   0.1379
## Sugar...Sweeteners          23.39      13.99   1.672   0.0975 .
## Treenuts                    25.15      14.10   1.784   0.0772 .
## Vegetable.Oils              23.29      13.98   1.665   0.0987 .
## Vegetables                  23.13      14.05   1.646   0.1027
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.779 on 109 degrees of freedom
## Multiple R-squared:  0.5869, Adjusted R-squared:  0.5112
## F-statistic: 7.744 on 20 and 109 DF,  p-value: 2.811e-13
```

```r
summary(bic)
```

```
##
## Call:
## lm(formula = log(Deaths + 1e-05) ~ Alcoholic.Beverages + Animal.fats +
##     Cereals...Excluding.Beer + Eggs + Fruits...Excluding.Wine +
##     Milk...Excluding.Butter + Starchy.Roots + Sugar...Sweeteners +
##     Treenuts + Vegetable.Oils, data = traindata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -6.2486 -1.0354  0.2173  1.1617  3.0101
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -23.07670    2.96162  -7.792 2.75e-12 ***
## Alcoholic.Beverages       0.86379    0.21582   4.002 0.000109 ***
## Animal.fats               0.43298    0.18144   2.386 0.018597 *
## Cereals...Excluding.Beer   0.34107    0.06811   5.008 1.93e-06 ***
## Eggs                      1.58544    0.70088   2.262 0.025509 *
## Fruits...Excluding.Wine   0.48749    0.16481   2.958 0.003737 **
## Milk...Excluding.Butter   0.68307    0.12397   5.510 2.11e-07 ***
## Starchy.Roots             0.28967    0.08536   3.394 0.000937 ***
## Sugar...Sweeteners        0.48974    0.11293   4.337 3.05e-05 ***
## Treenuts                  1.78027    0.65536   2.716 0.007582 **
## Vegetable.Oils            0.45738    0.09192   4.976 2.21e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.828 on 119 degrees of freedom
## Multiple R-squared:  0.5237, Adjusted R-squared:  0.4837
## F-statistic: 13.09 on 10 and 119 DF,  p-value: 3.343e-15
```

```r
#it seems BIC selected model is better, all predictors are significant, the R-squared value
#is not lower too much compared with full model

#model diagnostics:

par(mfrow = c(2,2))
plot(bic,1:4)
```




```r
par(mfrow = c(1,1))

#Check model assumptions:

#1) independent assumption: the points are randomly distributed around the zero mean line, it is true.

#2) linearity assumption: the residuals plot shows there is no special curve, the linearity asumption i

#3) constant variance assumption: the residuals plot also shows  the spread of residuals does not chang


#4) normality assumption: the normal qq plot shows that only some outliers far from  the line at the tw
```

```r
#then we need to check unusual points

n = dim(traindata)[1]

outliers <- which(abs(rstudent(bic)) > 2.5)
outliers
```

```
##  42  25 150 129
##  24  33  73  93
```

```r
highleverages <- which(hatvalues(bic) >  2 * mean(hatvalues(bic)  ))
highleverages
```

```
##  43  71   2  96  28  31 132 133 138  77
##   3  22  43  70  71  76  83  96 129 130
```

```r
stronginfluences <- which(cooks.distance(bic) > 4 / (nrow(traindata) - length(coef(bic))))
stronginfluences
```

```
## 111  33  42  25  28 150  86 129 133
##   8  20  24  33  71  73  79  93  96
```

```r
#remove them

ids <- unique(c(outliers ,highleverages ,stronginfluences ))


bic2 <- lm(formula = log(Deaths + 1e-05) ~ Alcoholic.Beverages + Animal.fats +
    Cereals...Excluding.Beer + Eggs + Fruits...Excluding.Wine +
    Milk...Excluding.Butter + Starchy.Roots + Sugar...Sweeteners +
    Treenuts + Vegetable.Oils, data = traindata[-ids, ])

summary(bic2)
```

```
##
## Call:
## lm(formula = log(Deaths + 1e-05) ~ Alcoholic.Beverages + Animal.fats +
##     Cereals...Excluding.Beer + Eggs + Fruits...Excluding.Wine +
##     Milk...Excluding.Butter + Starchy.Roots + Sugar...Sweeteners +
##     Treenuts + Vegetable.Oils, data = traindata[-ids, ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1513 -0.7460  0.0383  1.0084  2.6328
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -19.20332    2.62498  -7.316 6.00e-11 ***
## Alcoholic.Beverages        0.88061    0.17773   4.955 2.88e-06 ***
## Animal.fats                0.40080    0.16333   2.454 0.015827 *
## Cereals...Excluding.Beer   0.26963    0.06059   4.450 2.20e-05 ***
## Eggs                       1.76568    0.60970   2.896 0.004625 **
## Fruits...Excluding.Wine    0.62893    0.17164   3.664 0.000396 ***
## Milk...Excluding.Butter    0.42046    0.11224   3.746 0.000298 ***
## Starchy.Roots              0.11226    0.07317   1.534 0.128080
## Sugar...Sweeteners         0.43211    0.09611   4.496 1.83e-05 ***
## Treenuts                   1.92729    0.60922   3.164 0.002054 **
## Vegetable.Oils             0.28011    0.07884   3.553 0.000578 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.333 on 102 degrees of freedom
## Multiple R-squared:  0.6628, Adjusted R-squared:  0.6297
## F-statistic: 20.05 on 10 and 102 DF,  p-value: < 2.2e-16
```

```
#model diagnostics check, this potentially can be used as final best model
library(lmtest)
dwtest(bic2)
```

```
##
##  Durbin-Watson test
##
## data:  bic2
## DW = 1.9415, p-value = 0.3696
## alternative hypothesis: true autocorrelation is greater than 0
```

```
bptest(bic2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  bic2
## BP = 16.401, df = 10, p-value = 0.08872
```

```
shapiro.test(residuals(bic2))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(bic2)
## W = 0.97833, p-value = 0.06324
```

```
#both p values are larger than 0.05, so the residauls have constant variance and to be normality,
#so linear model is valid


#predictions


pred_2 <- predict(bic2, testdata)
pred_2 <- exp(pred_2) -  1e-05
mse_2 <- mean((pred_2 - testdata$Deaths)^2)
mse_2
```

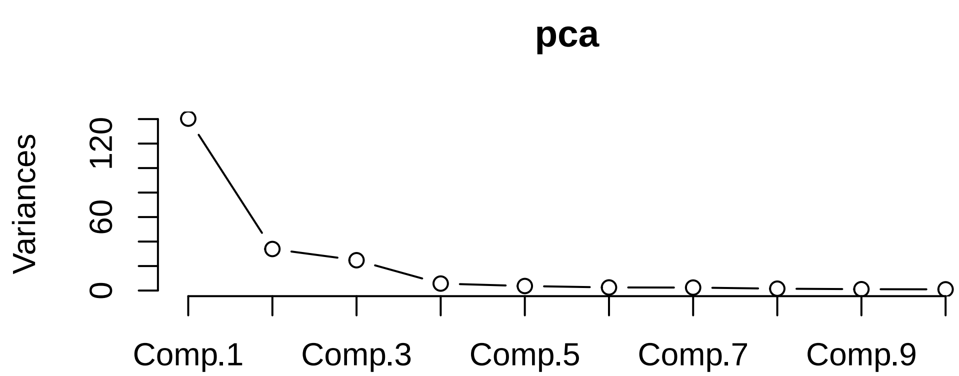
```
## [1] 0.006352532
```

```
# BIC2 Model: Adjusted R^2 0.6297; Predicted MSE: 0.00635
```

```r
#use PCA

XX <- mod2[,-ncol(mod2)]

pca <- princomp(XX)

summary(pca) # first 4 PCS already explain  0.93623305 that 93.6% percent of variance in original data
```

```
## Importance of components:
##                           Comp.1    Comp.2    Comp.3     Comp.4     Comp.5
## Standard deviation     11.8468335 5.8243133 4.9765493 2.38682863 1.92433711
## Proportion of Variance  0.6418012 0.1551263 0.1132537 0.02605184 0.01693395
## Cumulative Proportion   0.6418012 0.7969275 0.9101812 0.93623305 0.95316700
##                           Comp.6    Comp.7      Comp.8      Comp.9     Comp.10
## Standard deviation     1.59517671 1.5646738 1.243739010 1.055400087 1.001686752
## Proportion of Variance 0.01163626 0.0111955 0.007073827 0.005093663 0.004588384
## Cumulative Proportion  0.96480327 0.9759988 0.983072594 0.988166257 0.992754641
##                           Comp.11     Comp.12     Comp.13     Comp.14
## Standard deviation     0.803919533 0.643843847 0.519905018 0.258951995
## Proportion of Variance 0.002955433 0.001895645 0.001236073 0.000306644
## Cumulative Proportion  0.995710074 0.997605719 0.998841792 0.999148436
##                            Comp.15      Comp.16      Comp.17      Comp.18
## Standard deviation     0.2394509377 0.2252708738 0.2106224529 1.387777e-01
## Proportion of Variance 0.0002621978 0.0002320631 0.0002028641 8.807146e-05
## Cumulative Proportion  0.9994106338 0.9996426968 0.9998455610 9.999336e-01
##                            Comp.19      Comp.20      Comp.21      Comp.22
## Standard deviation     9.476482e-02 6.722777e-02 3.055554e-02 7.797110e-03
## Proportion of Variance 4.106674e-05 2.066776e-05 4.269489e-06 2.780118e-07
## Cumulative Proportion  9.999747e-01 9.999954e-01 9.999996e-01 9.999999e-01
##                            Comp.23      Comp.24
## Standard deviation     4.325434e-03 2.384993e-05
## Proportion of Variance 8.555694e-08 2.601178e-12
## Cumulative Proportion  1.000000e+00 1.000000e+00
```

```r
plot(pca, type = "l")
```



```r
#use the first 4 principle compoents is appropriate
#as it is an elbow point that after this point, the decrease is very slow

#interpret which variables contribute to the 4 PCS, the larger
#the absolute coefficients, the more important in the corresponding PC
#for example, Obesity  , Cereals...Excluding.Beer ,Animal.Products ,Vegetal.Products  are most #importa
round(pca$loadings[,1:4],2)
```

```
##                          Comp.1 Comp.2 Comp.3 Comp.4
## Alcoholic.Beverages        0.05   0.07   0.03   0.01
## Animal.Products            0.34   0.28   0.40   0.18
## Animal.fats                0.07   0.06   0.07  -0.01
## Aquatic.Products..Other    0.00   0.00   0.00   0.00
## Cereals...Excluding.Beer  -0.42  -0.61   0.40   0.33
## Eggs                       0.02   0.00   0.02  -0.01
## Fish..Seafood              0.00   0.03   0.01   0.00
## Fruits...Excluding.Wine    0.02   0.04  -0.12  -0.04
## Meat                       0.14   0.11   0.12   0.15
## Milk...Excluding.Butter    0.11   0.07   0.18   0.04
## Miscellaneous              0.00   0.00   0.00   0.00
## Offals                     0.00   0.01   0.01   0.01
## Oilcrops                  -0.02   0.01  -0.08   0.05
## Pulses                    -0.04   0.01  -0.08  -0.05
## Spices                     0.00   0.00   0.00   0.00
## Starchy.Roots             -0.12   0.30  -0.59   0.47
## Stimulants                 0.01   0.01   0.01   0.02
## Sugar.Crops                0.00   0.00   0.00   0.00
## Sugar...Sweeteners         0.11  -0.08   0.05  -0.28
## Treenuts                   0.01   0.00   0.00  -0.01
## Vegetal.Products          -0.34  -0.28  -0.40  -0.18
## Vegetable.Oils             0.06  -0.03  -0.06  -0.68
## Vegetables                 0.01  -0.01   0.03   0.01
## Obesity                    0.73  -0.59  -0.30   0.17
```

```r
#then use the first 4 PCs  to build regression model
newmod <- data.frame(Deaths = mod2$Deaths, pca$scores[,1:4])
traindata_pca <-  newmod [id, ]
testdata_pca <- newmod[-id,]


pca_model <- lm(Deaths ~ ., data = traindata_pca)
summary(pca_model)
```

```
##
## Call:
## lm(formula = Deaths ~ ., data = traindata_pca)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.084150 -0.025959 -0.003519  0.015586  0.111611
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  0.0417346  0.0037254  11.203  < 2e-16 ***
## Comp.1       0.0023764  0.0003169   7.500 1.04e-11 ***
## Comp.2       0.0001938  0.0006525   0.297   0.7669
## Comp.3       0.0013665  0.0007889   1.732   0.0857 .
## Comp.4      -0.0015489  0.0016362  -0.947   0.3456
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04244 on 125 degrees of freedom
## Multiple R-squared:  0.3254, Adjusted R-squared:  0.3038
## F-statistic: 15.07 on 4 and 125 DF,  p-value: 4.404e-10
```

```
pred <- predict(pca_model, testdata_pca)

mse_pca <- (mean((pred - testdata_pca$Deaths)^2))
mse_pca
```

```
## [1] 0.001167658
```

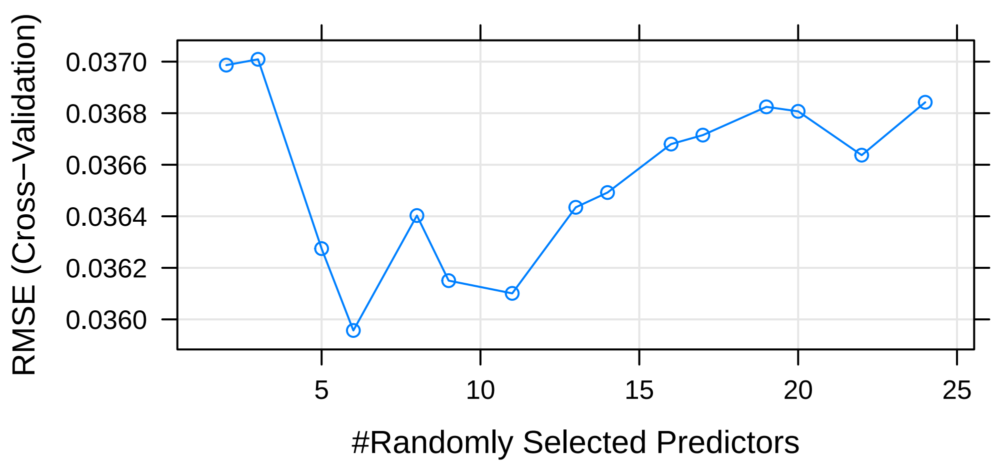
```
#MSE is low
```

```
#random forest model

library(caret)
set.seed(1)


#first use 5-folds cross validation to find the best model

ctrl <- trainControl(method = "cv",
                     number = 5)

#as random forest is not like  linear model, it does not require normality and etc, so
#we can just use deaths as response

fit <- train(Deaths ~ ., data = traindata,
             method = "rf",
             tuneLength = 15,
             trControl = ctrl)

fit
```

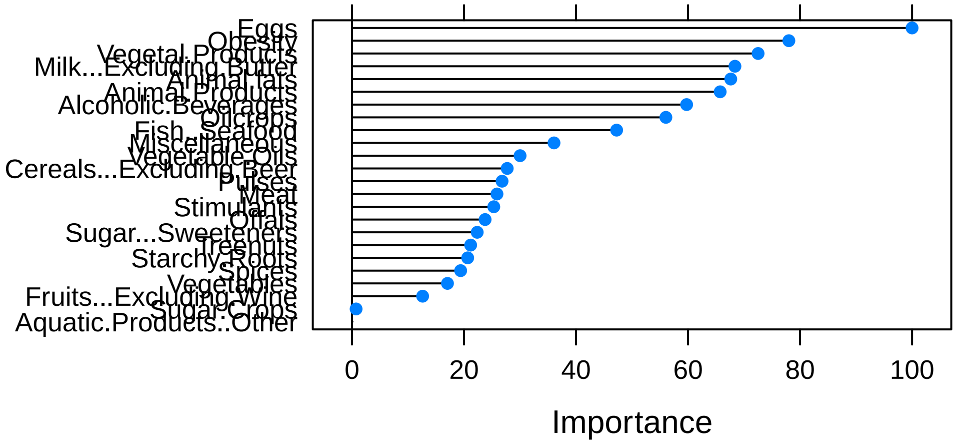
```
## Random Forest
##
## 130 samples
##  24 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 103, 106, 104, 104, 103
## Resampling results across tuning parameters:
##
##   mtry  RMSE        Rsquared   MAE
```

```
##     2     0.03698656  0.4949551  0.02802159
##     3     0.03700902  0.4845998  0.02806513
##     5     0.03627461  0.5112348  0.02733289
##     6     0.03595704  0.5152842  0.02729757
##     8     0.03640289  0.5068570  0.02736163
##     9     0.03615023  0.5105689  0.02715860
##    11     0.03610102  0.5118531  0.02706086
##    13     0.03643484  0.5013919  0.02729049
##    14     0.03649212  0.4970947  0.02710573
##    16     0.03668016  0.4899405  0.02721713
##    17     0.03671491  0.4907500  0.02710269
##    19     0.03682450  0.4901316  0.02734995
##    20     0.03680716  0.4886602  0.02719106
##    22     0.03663716  0.4929768  0.02715344
##    24     0.03684231  0.4880076  0.02706455
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was mtry = 6.
```

```r
plot(fit)
```




```r
#the best parameter mtry is 6, RMSE is lowest


#check variables important plot to find which predictors affect the deaths seriously,
#the top ones are most important ones
plot(varImp(fit))
```




```
#predictions


pred <- predict(fit, testdata)

mse2 <- (mean((pred - testdata$Deaths)^2))
mse2
```

```
## [1] 0.0009464952
```

```
#MSE is much lower, so the model performs much better than linear model
```