STT - Mini Project

Code Output File

Aditya Jain

2022-10-28

Contents

Question 1

(a)

The affinity matrix was created as shown below for 15 nodes. The below matrix has been created for just 1 simulation but each and every simulation would have it's own affinity matrix. The affinity matrix is of 15 columns and 15 rows which will give the information of which node is connected to which node. For ex- if on index (1,2) and (2,1) there is a "1" in affinity matrix, then that means there is a connection between nodes 1 and 2.

(b)

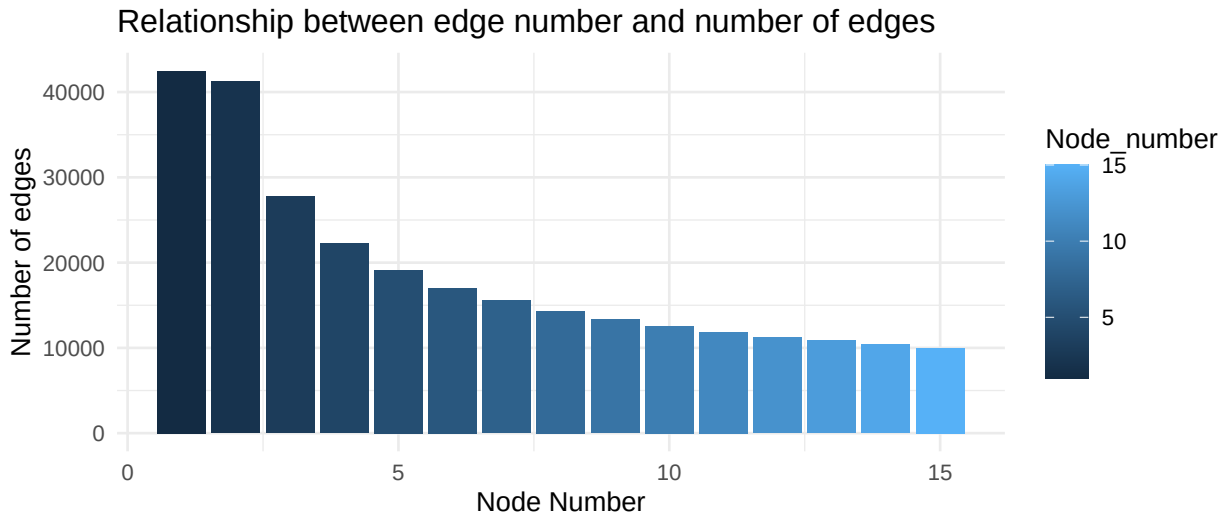
The edge vector contains the information of how many edges are associated with that particular node. So, for ex- if edge[1] has the value of 6 then that means Node number 1 is associated with or connected with 6 edges/nodes.

Therefore, for this question's sub-part 10,000 simulations was done and for each simulation the edge vector was stored, which was subsequently added to the previous simulation's edge vector to evaluate that how many edges has been connected with each node from Node 1-15 for total of 10,000 simulations.

Observations-

The number of edges each node has was not equal. When we plotted the distribution plot i.e. histogram, between number of edges connected to it and node number, it was observed that the relationship is an exponential distribution curve. This is shown below.

It can be seen that maximum edges was connected to Initial Nodes i.e. Node 1,2,3 and 4. The total edges connected with node 1 and node 2 are approximately same and subsequently after that the number of edges decreases with the increase in the Node number.

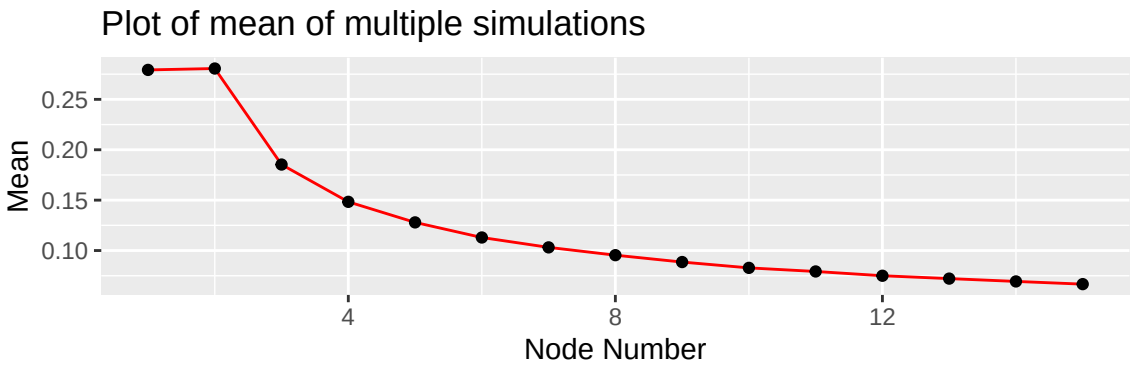


(c)

```
## [1] 5.944221
```

```
## [1] 47.96571
```

It can be seen that initial 1/3rd of the total nodes, i.e. for 15 nodes it is 4.95 (approx. 5), is associated with 50% of the total edges formed and the mean value is approximately 6.



For multiple simulations we evaluated the mean for each node and then plotted the distribution using the affinity matrix. So, for each node and for all simulations total number of connections were calculated and divided by the total count which finally gave the mean value for each node. It was found that the mean value is also distributed exponentially and there is a decrease in it's value as the number of nodes is increased.

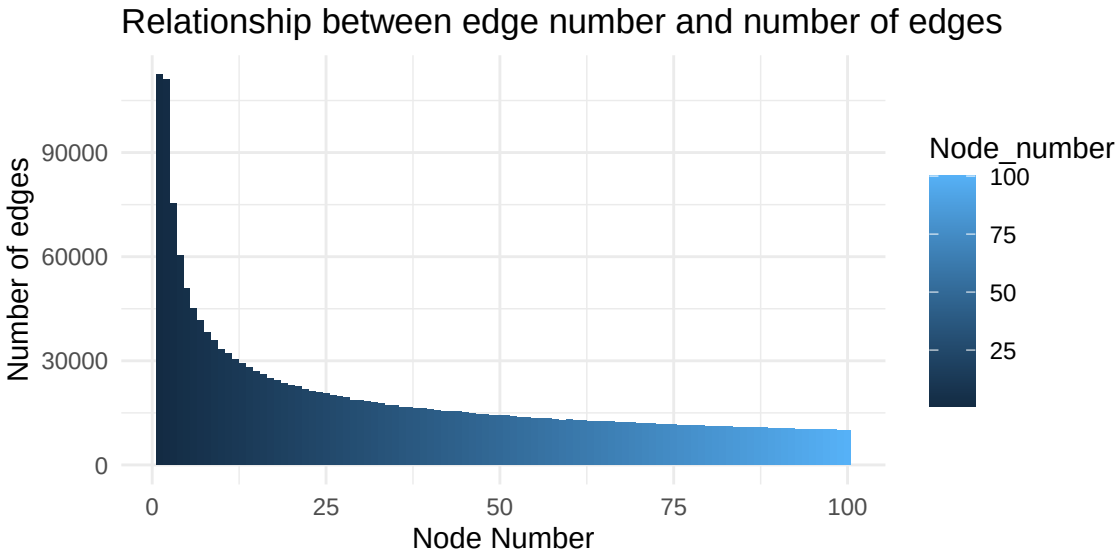
(d)

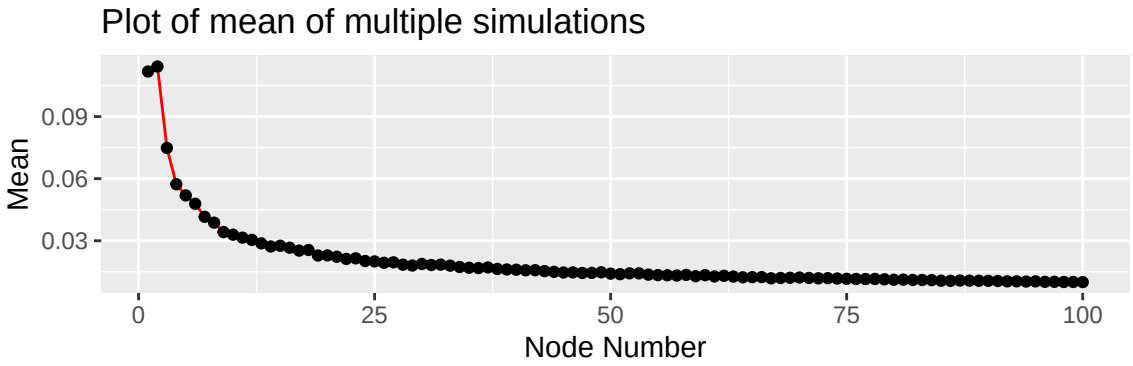
```
## [1] 34.31144
```

```
## [1] 49.43379
```

In this, it was found that approx. 1/4 of total nodes were associated with 50% of the total edges formed and the mean value is approximately 34.

In this sub-part of the question, the number of nodes were increased to 100 and again the same distribution graph was plotted. This plot made the picture clearer that the plot is an exponential distribution. This can be visualized below.





It was found that the mean value is also distributed exponentially and there is a sharp decrease in it's value as the number of nodes is increased.

Question 2

Some basic information gathered from the internet about the data and the variables -

- The nndb data is a flattened version of the USDA National Nutrition Database
- Each record is a food composition data per 100 grams for all major source of food in USA

The columns are mostly self-explanatory. The nutrient columns end with the units, so:

- Nutrient_g is in grams
- Nutrient_mg is in milligrams
- Nutrient_mcg is in micro grams
- Nutrient_USRDA is in percentage of US Recommended Daily Allows (e.g. 0.50 is 50%)

```
## [1] "8618" "45"
```

```
##  [1] "ID"               "FoodGroup"        "ShortDescrip"     "Descrip"
##  [5] "CommonName"       "MfgName"          "ScientificName"   "Energy_kcal"
##  [9] "Protein_g"        "Fat_g"            "Carb_g"           "Sugar_g"
## [13] "Fiber_g"          "VitA_mcg"         "VitB6_mg"         "VitB12_mcg"
## [17] "VitC_mg"          "VitE_mg"          "Folate_mcg"       "Niacin_mg"
## [21] "Riboflavin_mg"    "Thiamin_mg"       "Calcium_mg"       "Copper_mcg"
## [25] "Iron_mg"          "Magnesium_mg"     "Manganese_mg"     "Phosphorus_mg"
## [29] "Selenium_mcg"     "Zinc_mg"          "VitA_USRDA"       "VitB6_USRDA"
## [33] "VitB12_USRDA"     "VitC_USRDA"       "VitE_USRDA"       "Folate_USRDA"
## [37] "Niacin_USRDA"     "Riboflavin_USRDA" "Thiamin_USRDA"    "Calcium_USRDA"
## [41] "Copper_USRDA"     "Magnesium_USRDA"  "Phosphorus_USRDA" "Selenium_USRDA"
## [45] "Zinc_USRDA"
```

- Within the data set, their are 6 non numerical variables consisting of categorical and nominal data
- One Energy column, the total energy in kilo calorie per 100 gm of food
- Five basic nutrients like protein, fat, carb. etc in grams
- Vitamins and other minerals in mili and micro grams
- The rest of the columns contain values limited per day by US government
- The data majorly contains numeric values from column energy_kcal to zinc_usrda
- Nominal and categorical data are stored as Factor, with levels equal to number of unique data points
- Id is stored as int data type
- Common, Manufacturing and Scientific names are the only fields with missing or null values

```
##                 Energy_kcal Protein_g Fat_g Carb_g Sugar_g Fiber_g VitA_mcg
## Energy_kcal             1.0       0.1   0.8    0.5     0.3     0.2      0.0
## Protein_g               0.1       1.0   0.1   -0.3    -0.3    -0.1      0.0
## Fat_g                   0.8       0.1   1.0   -0.1     0.0     0.0      0.0
## Carb_g                  0.5      -0.3  -0.1    1.0     0.6     0.5      0.0
## Sugar_g                 0.3      -0.3   0.0    0.6     1.0     0.1      0.0
## Fiber_g                 0.2      -0.1   0.0    0.5     0.1     1.0      0.0
## VitA_mcg                0.0       0.0   0.0    0.0     0.0     0.0      1.0
## VitB6_mg                0.1       0.2   0.0    0.2     0.1     0.2      0.1
## VitB12_mcg              0.0       0.2   0.0   -0.1    -0.1    -0.1      0.6
## VitC_mg                 0.0      -0.1  -0.1    0.1     0.1     0.1      0.1
## VitE_mg                 0.3       0.0   0.3    0.1     0.1     0.2      0.0
## Folate_mcg              0.1       0.0  -0.1    0.3     0.2     0.2      0.1
```

```
## Niacin_mg               0.2       0.4   0.0    0.2     0.0     0.1      0.2
## Riboflavin_mg           0.2       0.2   0.0    0.2     0.1     0.2      0.3
## Thiamin_mg              0.2       0.1   0.0    0.3     0.1     0.2      0.1
## Calcium_mg              0.1       0.0   0.0    0.2     0.1     0.2      0.0
## Copper_mg               0.1       0.2   0.0    0.1     0.0     0.2      0.6
## Iron_mg                 0.2       0.1   0.0    0.4     0.1     0.4      0.1
## Magnesium_mg            0.3       0.2   0.1    0.3     0.0     0.5      0.0
## Manganese_mg            0.0       0.0   0.0    0.1     0.0     0.1      0.4
## Phosphorus_mg           0.2       0.4   0.1    0.1     0.0     0.2      0.1
## Selenium_mg             0.1       0.4   0.0   -0.1    -0.1     0.0      0.0
## Zinc_mg                 0.1       0.4   0.0    0.0     0.0     0.1      0.1
## VitA_USRDA              0.0       0.0   0.0    0.0     0.0     0.0      1.0
## VitB6_USRDA             0.1       0.2   0.0    0.2     0.1     0.2      0.1
## VitB12_USRDA            0.0       0.2   0.0   -0.1    -0.1    -0.1      0.6
## VitC_USRDA              0.0      -0.1  -0.1    0.1     0.1     0.1      0.1
## VitE_USRDA              0.3       0.0   0.3    0.1     0.1     0.2      0.0
## Folate_USRDA            0.1       0.0  -0.1    0.3     0.2     0.2      0.1
## Niacin_USRDA            0.2       0.4   0.0    0.2     0.0     0.1      0.2
## Riboflavin_USRDA        0.2       0.2   0.0    0.2     0.1     0.2      0.3
## Thiamin_USRDA           0.2       0.1   0.0    0.3     0.1     0.2      0.1
## Calcium_USRDA           0.1       0.0   0.0    0.2     0.1     0.2      0.0
## Copper_USRDA            0.1       0.2   0.0    0.1     0.0     0.2      0.6
## Magnesium_USRDA         0.3       0.2   0.1    0.3     0.0     0.5      0.0
## Phosphorus_USRDA        0.2       0.4   0.1    0.1     0.0     0.2      0.1
## Selenium_USRDA          0.1       0.4   0.0   -0.1    -0.1     0.0      0.0
## Zinc_USRDA              0.1       0.4   0.0    0.0     0.0     0.1      0.1
##                  VitB6_mg VitB12_mcg VitC_mg VitE_mg Folate_mcg Niacin_mg
## Energy_kcal           0.1        0.0     0.0     0.3        0.1       0.2
## Protein_g             0.2        0.2    -0.1     0.0        0.0       0.4
## Fat_g                 0.0        0.0    -0.1     0.3       -0.1       0.0
## Carb_g                0.2       -0.1     0.1     0.1        0.3       0.2
## Sugar_g               0.1       -0.1     0.1     0.1        0.2       0.0
## Fiber_g               0.2       -0.1     0.1     0.2        0.2       0.1
## VitA_mcg              0.1        0.6     0.1     0.0        0.1       0.2
## VitB6_mg              1.0        0.3     0.3     0.3        0.6       0.7
## VitB12_mcg            0.3        1.0     0.0     0.1        0.2       0.3
## VitC_mg               0.3        0.0     1.0     0.1        0.1       0.2
## VitE_mg               0.3        0.1     0.1     1.0        0.2       0.2
## Folate_mcg            0.6        0.2     0.1     0.2        1.0       0.5
## Niacin_mg             0.7        0.3     0.2     0.2        0.5       1.0
## Riboflavin_mg         0.6        0.4     0.2     0.2        0.6       0.7
## Thiamin_mg            0.4        0.1     0.1     0.1        0.5       0.6
## Calcium_mg            0.2        0.0     0.1     0.1        0.1       0.1
## Copper_mcg            0.1        0.6     0.0     0.1        0.1       0.1
## Iron_mg               0.5        0.2     0.1     0.2        0.5       0.5
## Magnesium_mg          0.3        0.0     0.1     0.2        0.2       0.3
## Manganese_mg          0.1        0.2     0.0     0.1        0.1       0.1
## Phosphorus_mg         0.2        0.1     0.0     0.1        0.1       0.3
## Selenium_mg           0.1        0.2     0.0     0.0        0.0       0.2
## Zinc_mg               0.4        0.3     0.0     0.2        0.3       0.5
## VitA_USRDA            0.1        0.6     0.1     0.0        0.1       0.2
## VitB6_USRDA           1.0        0.3     0.3     0.3        0.6       0.7
## VitB12_USRDA          0.3        1.0     0.0     0.1        0.2       0.3
## VitC_USRDA            0.3        0.0     1.0     0.1        0.1       0.2
```

```
## VitE_USRDA            0.3        0.1     0.1     1.0        0.2       0.2
## Folate_USRDA          0.6        0.2     0.1     0.2        1.0       0.5
## Niacin_USRDA          0.7        0.3     0.2     0.2        0.5       1.0
## Riboflavin_USRDA      0.6        0.4     0.2     0.2        0.6       0.7
## Thiamin_USRDA         0.4        0.1     0.1     0.1        0.5       0.6
## Calcium_USRDA         0.2        0.0     0.1     0.1        0.1       0.1
## Copper_USRDA          0.1        0.6     0.0     0.1        0.1       0.1
## Magnesium_USRDA       0.3        0.0     0.1     0.2        0.2       0.3
## Phosphorus_USRDA      0.2        0.1     0.0     0.1        0.1       0.3
## Selenium_USRDA        0.1        0.2     0.0     0.0        0.0       0.2
## Zinc_USRDA            0.4        0.3     0.0     0.2        0.3       0.5
##                  Riboflavin_mg Thiamin_mg Calcium_mg Copper_mcg Iron_mg
## Energy_kcal                0.2        0.2        0.1        0.1     0.2
## Protein_g                  0.2        0.1        0.0        0.2     0.1
## Fat_g                      0.0        0.0        0.0        0.0     0.0
## Carb_g                     0.2        0.3        0.2        0.1     0.4
## Sugar_g                    0.1        0.1        0.1        0.0     0.1
## Fiber_g                    0.2        0.2        0.2        0.2     0.4
## VitA_mcg                   0.3        0.1        0.0        0.6     0.1
## VitB6_mg                   0.6        0.4        0.2        0.1     0.5
## VitB12_mcg                 0.4        0.1        0.0        0.6     0.2
## VitC_mg                    0.2        0.1        0.1        0.0     0.1
## VitE_mg                    0.2        0.1        0.1        0.1     0.2
## Folate_mcg                 0.6        0.5        0.1        0.1     0.5
## Niacin_mg                  0.7        0.6        0.1        0.1     0.5
## Riboflavin_mg              1.0        0.6        0.2        0.3     0.5
## Thiamin_mg                 0.6        1.0        0.1        0.1     0.4
## Calcium_mg                 0.2        0.1        1.0        0.1     0.3
## Copper_mcg                 0.3        0.1        0.1        1.0     0.2
## Iron_mg                    0.5        0.4        0.3        0.2     1.0
## Magnesium_mg               0.2        0.2        0.3        0.3     0.4
## Manganese_mg               0.1        0.0        0.1        0.2     0.1
## Phosphorus_mg              0.2        0.2        0.6        0.2     0.2
## Selenium_mcg               0.1        0.1        0.0        0.1     0.1
## Zinc_mg                    0.4        0.2        0.1        0.3     0.4
## VitA_USRDA                 0.3        0.1        0.0        0.6     0.1
## VitB6_USRDA                0.6        0.4        0.2        0.1     0.5
## VitB12_USRDA               0.4        0.1        0.0        0.6     0.2
## VitC_USRDA                 0.2        0.1        0.1        0.0     0.1
## VitE_USRDA                 0.2        0.1        0.1        0.1     0.2
## Folate_USRDA               0.6        0.5        0.1        0.1     0.5
## Niacin_USRDA               0.7        0.6        0.1        0.1     0.5
## Riboflavin_USRDA           1.0        0.6        0.2        0.3     0.5
## Thiamin_USRDA              0.6        1.0        0.1        0.1     0.4
## Calcium_USRDA              0.2        0.1        1.0        0.1     0.3
## Copper_USRDA               0.3        0.1        0.1        1.0     0.2
## Magnesium_USRDA            0.2        0.2        0.3        0.3     0.4
## Phosphorus_USRDA           0.2        0.2        0.6        0.2     0.2
## Selenium_USRDA             0.1        0.1        0.0        0.1     0.1
## Zinc_USRDA                 0.4        0.2        0.1        0.3     0.4
##                  Magnesium_mg Manganese_mg Phosphorus_mg Selenium_mcg Zinc_mg
## Energy_kcal               0.3          0.0           0.2          0.1     0.1
## Protein_g                 0.2          0.0           0.4          0.4     0.4
## Fat_g                     0.1          0.0           0.1          0.0     0.0
```

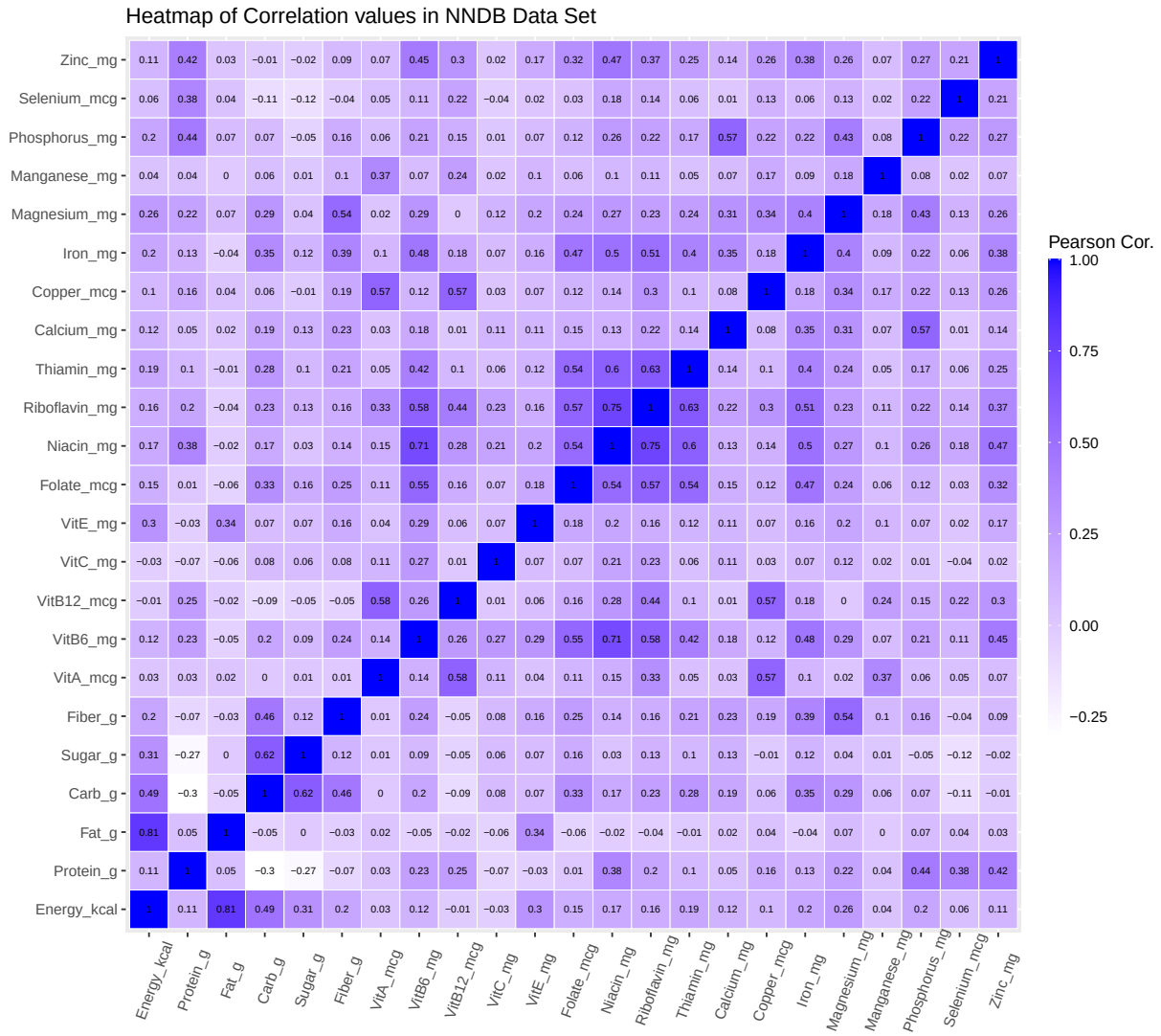
```
## Carb_g                  0.3         0.1          0.1         -0.1     0.0
## Sugar_g                 0.0         0.0          0.0         -0.1     0.0
## Fiber_g                 0.5         0.1          0.2          0.0     0.1
## VitA_mcg                0.0         0.4          0.1          0.0     0.1
## VitB6_mg                0.3         0.1          0.2          0.1     0.4
## VitB12_mcg              0.0         0.2          0.1          0.2     0.3
## VitC_mg                 0.1         0.0          0.0          0.0     0.0
## VitE_mg                 0.2         0.1          0.1          0.0     0.2
## Folate_mcg              0.2         0.1          0.1          0.0     0.3
## Niacin_mg               0.3         0.1          0.3          0.2     0.5
## Riboflavin_mg           0.2         0.1          0.2          0.1     0.4
## Thiamin_mg              0.2         0.0          0.2          0.1     0.2
## Calcium_mg              0.3         0.1          0.6          0.0     0.1
## Copper_mcg              0.3         0.2          0.2          0.1     0.3
## Iron_mg                 0.4         0.1          0.2          0.1     0.4
## Magnesium_mg            1.0         0.2          0.4          0.1     0.3
## Manganese_mg            0.2         1.0          0.1          0.0     0.1
## Phosphorus_mg           0.4         0.1          1.0          0.2     0.3
## Selenium_mcg            0.1         0.0          0.2          1.0     0.2
## Zinc_mg                 0.3         0.1          0.3          0.2     1.0
## VitA_USRDA              0.0         0.4          0.1          0.0     0.1
## VitB6_USRDA             0.3         0.1          0.2          0.1     0.4
## VitB12_USRDA            0.0         0.2          0.1          0.2     0.3
## VitC_USRDA              0.1         0.0          0.0          0.0     0.0
## VitE_USRDA              0.2         0.1          0.1          0.0     0.2
## Folate_USRDA            0.2         0.1          0.1          0.0     0.3
## Niacin_USRDA            0.3         0.1          0.3          0.2     0.5
## Riboflavin_USRDA        0.2         0.1          0.2          0.1     0.4
## Thiamin_USRDA           0.2         0.0          0.2          0.1     0.2
## Calcium_USRDA           0.3         0.1          0.6          0.0     0.1
## Copper_USRDA            0.3         0.2          0.2          0.1     0.3
## Magnesium_USRDA         1.0         0.2          0.4          0.1     0.3
## Phosphorus_USRDA        0.4         0.1          1.0          0.2     0.3
## Selenium_USRDA          0.1         0.0          0.2          1.0     0.2
## Zinc_USRDA              0.3         0.1          0.3          0.2     1.0
##                  VitA_USRDA VitB6_USRDA VitB12_USRDA VitC_USRDA VitE_USRDA
## Energy_kcal             0.0         0.1          0.0        0.0        0.3
## Protein_g               0.0         0.2          0.2       -0.1        0.0
## Fat_g                   0.0         0.0          0.0       -0.1        0.3
## Carb_g                  0.0         0.2         -0.1        0.1        0.1
## Sugar_g                 0.0         0.1         -0.1        0.1        0.1
## Fiber_g                 0.0         0.2         -0.1        0.1        0.2
## VitA_mcg                1.0         0.1          0.6        0.1        0.0
## VitB6_mg                0.1         1.0          0.3        0.3        0.3
## VitB12_mcg              0.6         0.3          1.0        0.0        0.1
## VitC_mg                 0.1         0.3          0.0        1.0        0.1
## VitE_mg                 0.0         0.3          0.1        0.1        1.0
## Folate_mcg              0.1         0.6          0.2        0.1        0.2
## Niacin_mg               0.2         0.7          0.3        0.2        0.2
## Riboflavin_mg           0.3         0.6          0.4        0.2        0.2
## Thiamin_mg              0.1         0.4          0.1        0.1        0.1
## Calcium_mg              0.0         0.2          0.0        0.1        0.1
## Copper_mcg              0.6         0.1          0.6        0.0        0.1
## Iron_mg                 0.1         0.5          0.2        0.1        0.2
```

```
## Magnesium_mg            0.0         0.3          0.0         0.1         0.2
## Manganese_mg            0.4         0.1          0.2         0.0         0.1
## Phosphorus_mg           0.1         0.2          0.1         0.0         0.1
## Selenium_mg             0.0         0.1          0.2         0.0         0.0
## Zinc_mg                 0.1         0.4          0.3         0.0         0.2
## VitA_USRDA              1.0         0.1          0.6         0.1         0.0
## VitB6_USRDA             0.1         1.0          0.3         0.3         0.3
## VitB12_USRDA            0.6         0.3          1.0         0.0         0.1
## VitC_USRDA              0.1         0.3          0.0         1.0         0.1
## VitE_USRDA              0.0         0.3          0.1         0.1         1.0
## Folate_USRDA            0.1         0.6          0.2         0.1         0.2
## Niacin_USRDA            0.2         0.7          0.3         0.2         0.2
## Riboflavin_USRDA        0.3         0.6          0.4         0.2         0.2
## Thiamin_USRDA           0.1         0.4          0.1         0.1         0.1
## Calcium_USRDA           0.0         0.2          0.0         0.1         0.1
## Copper_USRDA            0.6         0.1          0.6         0.0         0.1
## Magnesium_USRDA         0.0         0.3          0.0         0.1         0.2
## Phosphorus_USRDA        0.1         0.2          0.1         0.0         0.1
## Selenium_USRDA          0.0         0.1          0.2         0.0         0.0
## Zinc_USRDA              0.1         0.4          0.3         0.0         0.2
##                  Folate_USRDA Niacin_USRDA Riboflavin_USRDA Thiamin_USRDA
## Energy_kcal               0.1          0.2              0.2           0.2
## Protein_g                 0.0          0.4              0.2           0.1
## Fat_g                    -0.1          0.0              0.0           0.0
## Carb_g                    0.3          0.2              0.2           0.3
## Sugar_g                   0.2          0.0              0.1           0.1
## Fiber_g                   0.2          0.1              0.2           0.2
## VitA_mcg                  0.1          0.2              0.3           0.1
## VitB6_mg                  0.6          0.7              0.6           0.4
## VitB12_mcg                0.2          0.3              0.4           0.1
## VitC_mg                   0.1          0.2              0.2           0.1
## VitE_mg                   0.2          0.2              0.2           0.1
## Folate_mcg                1.0          0.5              0.6           0.5
## Niacin_mg                 0.5          1.0              0.7           0.6
## Riboflavin_mg             0.6          0.7              1.0           0.6
## Thiamin_mg                0.5          0.6              0.6           1.0
## Calcium_mg                0.1          0.1              0.2           0.1
## Copper_mcg                0.1          0.1              0.3           0.1
## Iron_mg                   0.5          0.5              0.5           0.4
## Magnesium_mg              0.2          0.3              0.2           0.2
## Manganese_mg              0.1          0.1              0.1           0.0
## Phosphorus_mg             0.1          0.3              0.2           0.2
## Selenium_mg               0.0          0.2              0.1           0.1
## Zinc_mg                   0.3          0.5              0.4           0.2
## VitA_USRDA                0.1          0.2              0.3           0.1
## VitB6_USRDA               0.6          0.7              0.6           0.4
## VitB12_USRDA              0.2          0.3              0.4           0.1
## VitC_USRDA                0.1          0.2              0.2           0.1
## VitE_USRDA                0.2          0.2              0.2           0.1
## Folate_USRDA              1.0          0.5              0.6           0.5
## Niacin_USRDA              0.5          1.0              0.7           0.6
## Riboflavin_USRDA          0.6          0.7              1.0           0.6
## Thiamin_USRDA             0.5          0.6              0.6           1.0
## Calcium_USRDA             0.1          0.1              0.2           0.1
```

```
## Copper_USRDA              0.1          0.1              0.3           0.1
## Magnesium_USRDA           0.2          0.3              0.2           0.2
## Phosphorus_USRDA          0.1          0.3              0.2           0.2
## Selenium_USRDA            0.0          0.2              0.1           0.1
## Zinc_USRDA                0.3          0.5              0.4           0.2
##                  Calcium_USRDA Copper_USRDA Magnesium_USRDA Phosphorus_USRDA
## Energy_kcal                0.1          0.1              0.3               0.2
## Protein_g                  0.0          0.2              0.2               0.4
## Fat_g                      0.0          0.0              0.1               0.1
## Carb_g                     0.2          0.1              0.3               0.1
## Sugar_g                    0.1          0.0              0.0               0.0
## Fiber_g                    0.2          0.2              0.5               0.2
## VitA_mcg                   0.0          0.6              0.0               0.1
## VitB6_mg                   0.2          0.1              0.3               0.2
## VitB12_mcg                 0.0          0.6              0.0               0.1
## VitC_mg                    0.1          0.0              0.1               0.0
## VitE_mg                    0.1          0.1              0.2               0.1
## Folate_mcg                 0.1          0.1              0.2               0.1
## Niacin_mg                  0.1          0.1              0.3               0.3
## Riboflavin_mg              0.2          0.3              0.2               0.2
## Thiamin_mg                 0.1          0.1              0.2               0.2
## Calcium_mg                 1.0          0.1              0.3               0.6
## Copper_mcg                 0.1          1.0              0.3               0.2
## Iron_mg                    0.3          0.2              0.4               0.2
## Magnesium_mg               0.3          0.3              1.0               0.4
## Manganese_mg               0.1          0.2              0.2               0.1
## Phosphorus_mg              0.6          0.2              0.4               1.0
## Selenium_mcg               0.0          0.1              0.1               0.2
## Zinc_mg                    0.1          0.3              0.3               0.3
## VitA_USRDA                 0.0          0.6              0.0               0.1
## VitB6_USRDA                0.2          0.1              0.3               0.2
## VitB12_USRDA               0.0          0.6              0.0               0.1
## VitC_USRDA                 0.1          0.0              0.1               0.0
## VitE_USRDA                 0.1          0.1              0.2               0.1
## Folate_USRDA               0.1          0.1              0.2               0.1
## Niacin_USRDA               0.1          0.1              0.3               0.3
## Riboflavin_USRDA           0.2          0.3              0.2               0.2
## Thiamin_USRDA              0.1          0.1              0.2               0.2
## Calcium_USRDA              1.0          0.1              0.3               0.6
## Copper_USRDA               0.1          1.0              0.3               0.2
## Magnesium_USRDA            0.3          0.3              1.0               0.4
## Phosphorus_USRDA           0.6          0.2              0.4               1.0
## Selenium_USRDA             0.0          0.1              0.1               0.2
## Zinc_USRDA                 0.1          0.3              0.3               0.3
##                  Selenium_USRDA Zinc_USRDA
## Energy_kcal                 0.1        0.1
## Protein_g                   0.4        0.4
## Fat_g                       0.0        0.0
## Carb_g                     -0.1        0.0
## Sugar_g                    -0.1        0.0
## Fiber_g                     0.0        0.1
## VitA_mcg                    0.0        0.1
## VitB6_mg                    0.1        0.4
## VitB12_mcg                  0.2        0.3
```

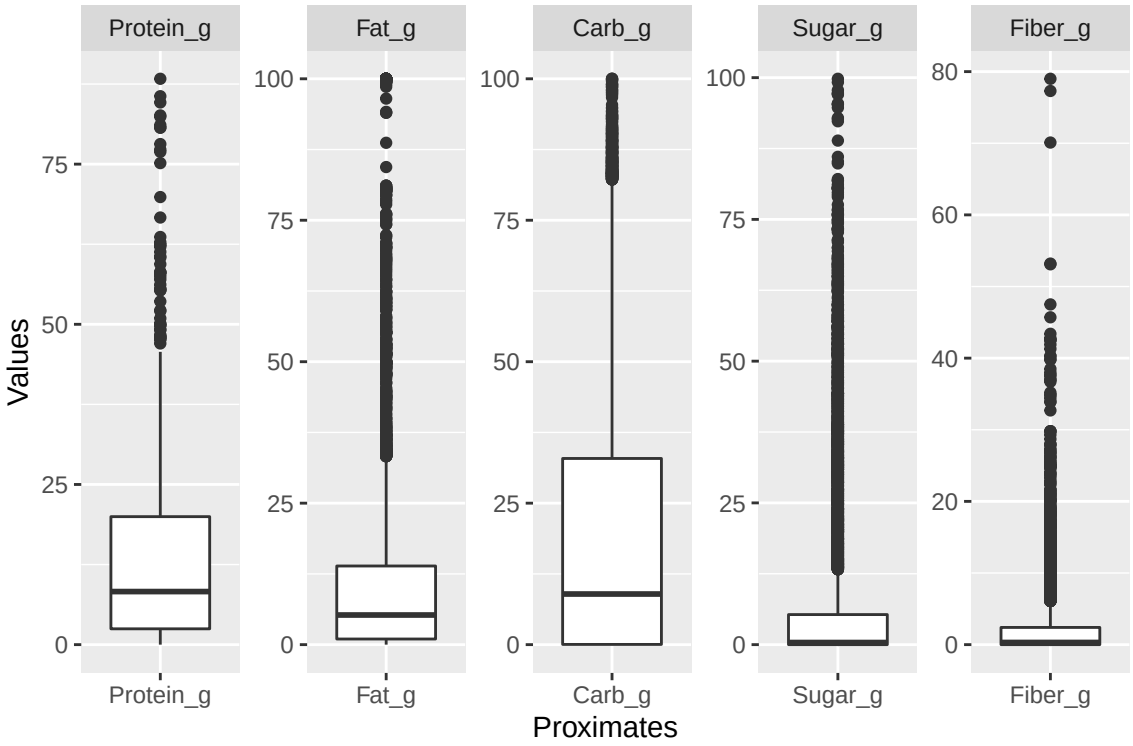
```
## VitC_mg                  0.0        0.0
## VitE_mg                  0.0        0.2
## Folate_mcg               0.0        0.3
## Niacin_mg                0.2        0.5
## Riboflavin_mg            0.1        0.4
## Thiamin_mg               0.1        0.2
## Calcium_mg               0.0        0.1
## Copper_mcg               0.1        0.3
## Iron_mg                  0.1        0.4
## Magnesium_mg             0.1        0.3
## Manganese_mg             0.0        0.1
## Phosphorus_mg            0.2        0.3
## Selenium_mcg             1.0        0.2
## Zinc_mg                  0.2        1.0
## VitA_USRDA               0.0        0.1
## VitB6_USRDA              0.1        0.4
## VitB12_USRDA             0.2        0.3
## VitC_USRDA               0.0        0.0
## VitE_USRDA               0.0        0.2
## Folate_USRDA             0.0        0.3
## Niacin_USRDA             0.2        0.5
## Riboflavin_USRDA         0.1        0.4
## Thiamin_USRDA            0.1        0.2
## Calcium_USRDA            0.0        0.1
## Copper_USRDA             0.1        0.3
## Magnesium_USRDA          0.1        0.3
## Phosphorus_USRDA         0.2        0.3
## Selenium_USRDA           1.0        0.2
## Zinc_USRDA               0.2        1.0
```

- All the USRDA attributes are redundant as they are perfectly correlated, hence removing these features for further analysis



The correlation heat map indicates positive correlation between variables

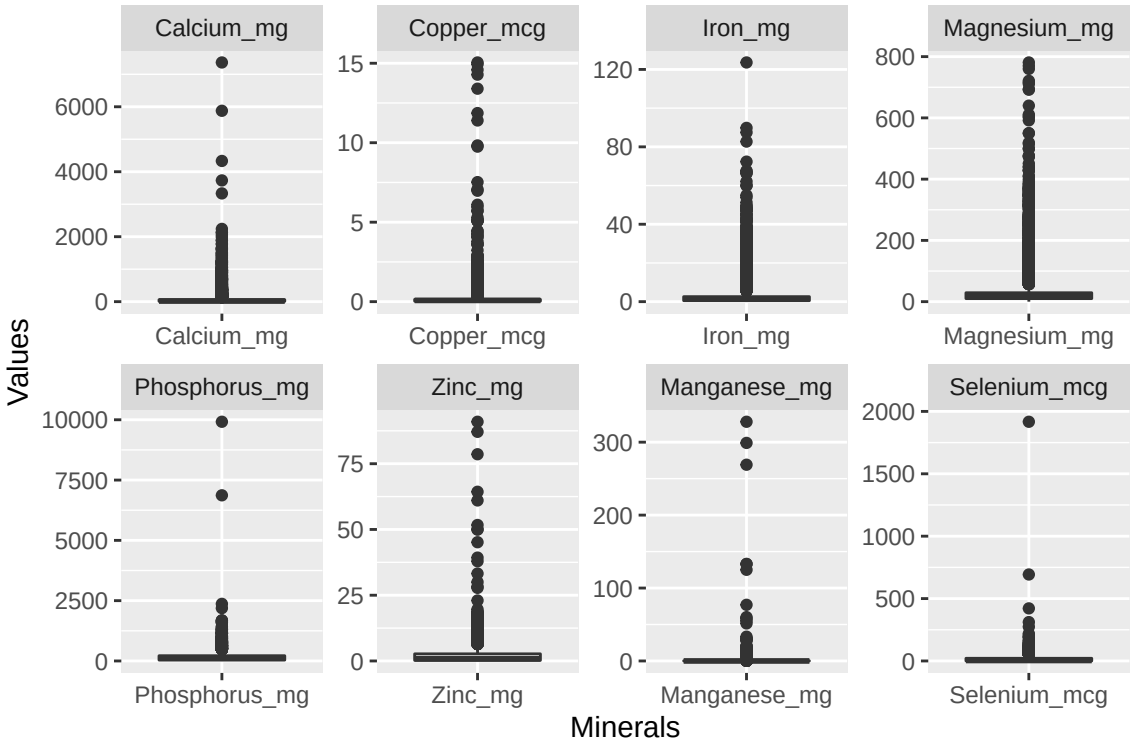
```
## No id variables; using all as measure variables
```



The above box plots shows us the distribution of five main nutrients in food.

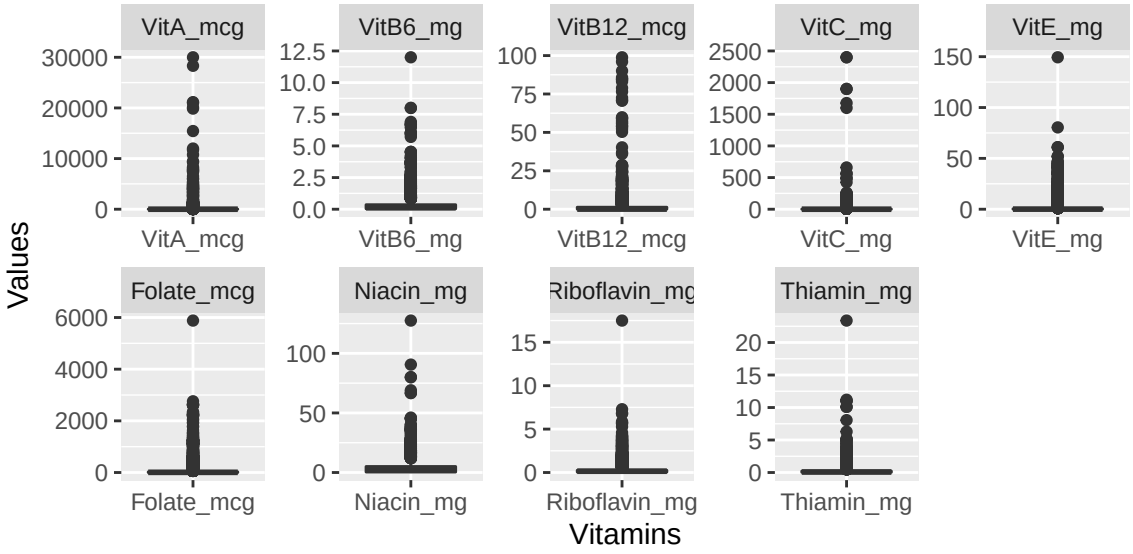
- We can clearly identify large numbers of outliers in every field, especially in sugar
- All nutrient category contains few products which are way above the US daily accepted limit

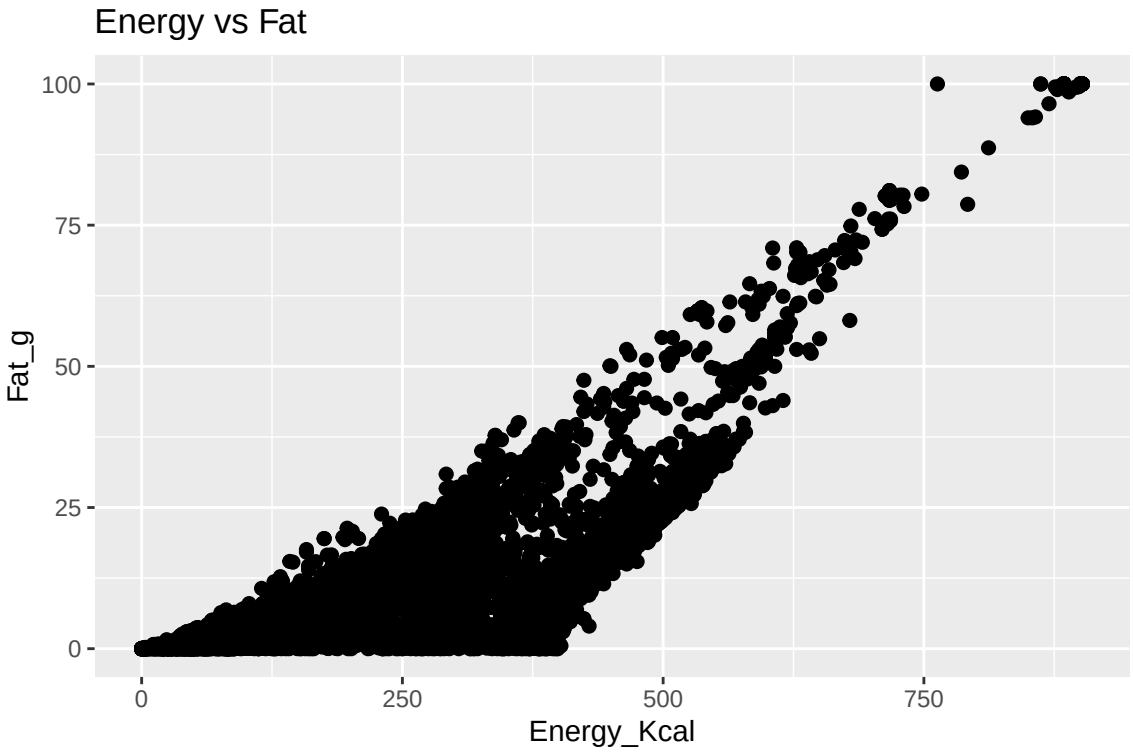
```
## No id variables; using all as measure variables
```



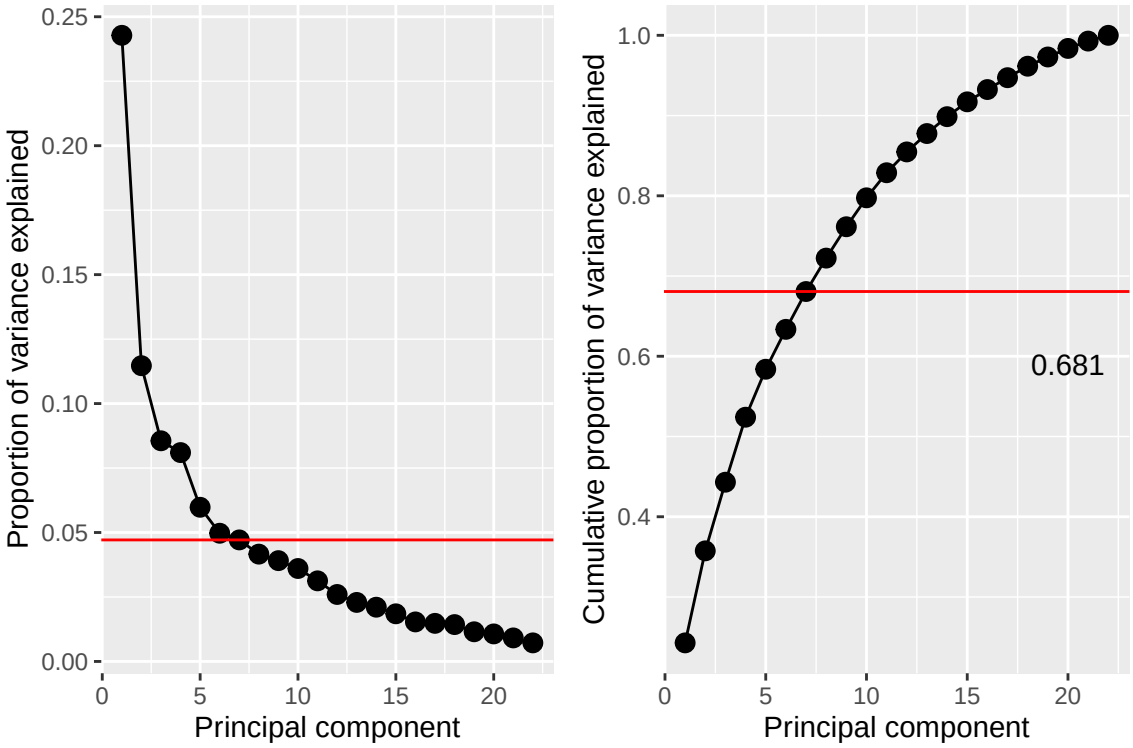
Minerals like phosphorous and selenium are present in huge amount in products, Way above accepted values by the US department.

```
## No id variables; using all as measure variables
```



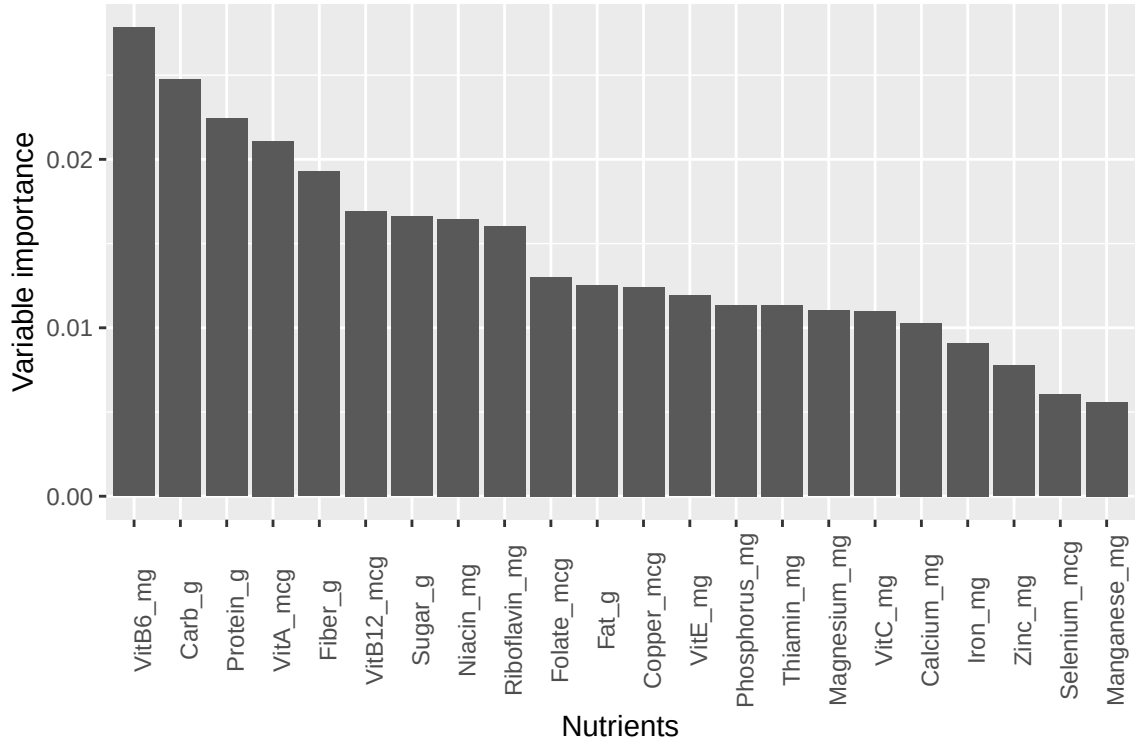


The scatter plot shows positive correlation between energy and fat. As visualized, as fat increases the energy Kcal also increases.

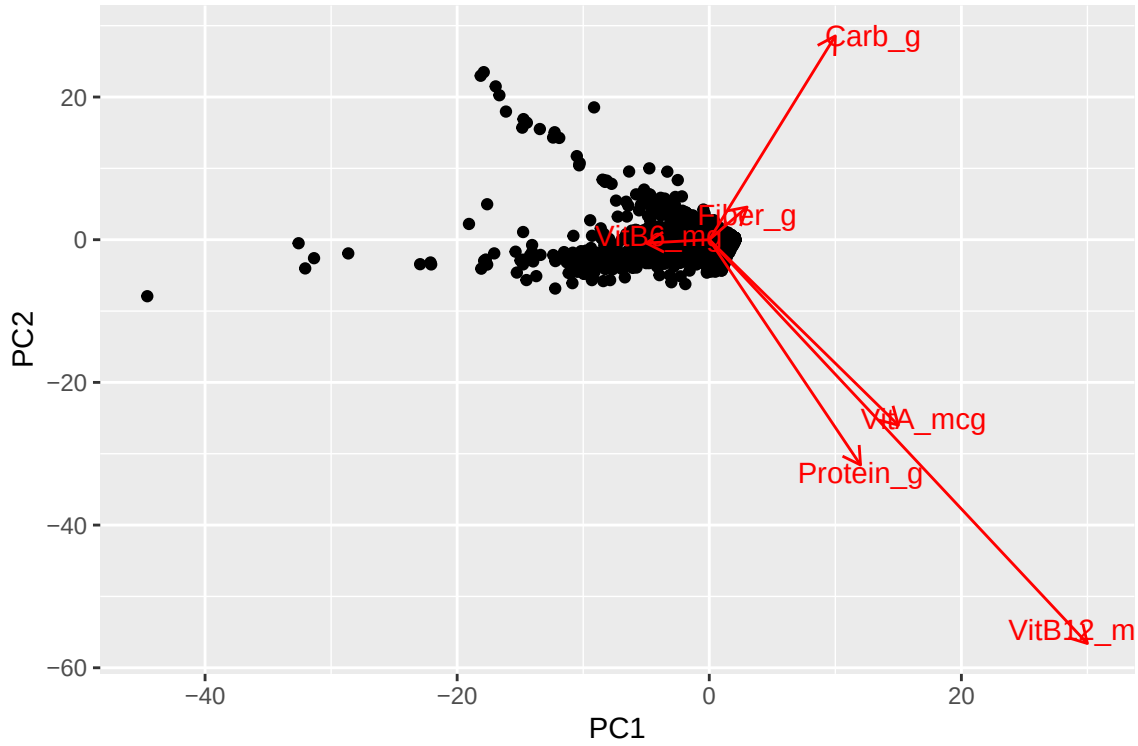


Principal component analysis helps to extract features in our data set which explains high proportion of variance. By applying PCA on our data, we are able to explain 70 percent of variance with the help of only

7 components.



The bar chart shows contribution of nutrients in the top 7 principal component.



The biplot shows 6 most significant nutrients impacting the first two principal components.

- It is observed that vitamins, protein, carb and fibers are the best indicators of variance in the data