cereal factor analysis

September 17, 2021

1 Cereal Data Factor Analysis

Objective: Comprehending the behaviour of customers on taking a particular cereal brands into consideration, based on their attributes. Based on the analysis, trying to understand the underlying similarities between the variables and converting them into factors for further analysis using statistical techniques.

[47]:
```
cereal <- read.csv("C:/Users/amite/Downloads/cereal.csv", header = TRUE)
dim(cereal)
```

1. 235 2. 26

2 Exploratory Data Analysis

2.0.1 Summary of the data

Summarize the data, to see if the data is structured as per the expectations.

[3]:
```
summary(cereal)
```

```
       Cereals      Filling         Natural         Fibre      
 CornFlakes :27   Min.   :1.000   Min.   :1.000   Min.   :1.000  
 Weetabix   :27   1st Qu.:3.000   1st Qu.:3.000   1st Qu.:3.000  
 Vitabrit   :25   Median :4.000   Median :4.000   Median :4.000  
 NutriGrain :24   Mean   :3.881   Mean   :3.783   Mean   :3.528  
 SpecialK   :23   3rd Qu.:4.500   3rd Qu.:4.000   3rd Qu.:4.000  
 RiceBubbles:21   Max.   :5.000   Max.   :5.000   Max.   :5.000  
 (Other)    :88                                                  
     Sweet            Easy            Salt         Satisfying   
 Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :2.000  
 1st Qu.:2.000   1st Qu.:4.000   1st Qu.:1.000   1st Qu.:3.000  
 Median :2.000   Median :5.000   Median :2.000   Median :4.000  
 Mean   :2.506   Mean   :4.532   Mean   :1.991   Mean   :4.004  
 3rd Qu.:3.000   3rd Qu.:5.000   3rd Qu.:3.000   3rd Qu.:5.000  
 Max.   :5.000   Max.   :6.000   Max.   :4.000   Max.   :6.000  

     Energy           Fun             Kids           Soggy      
 Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000  
 1st Qu.:3.000   1st Qu.:2.000   1st Qu.:3.000   1st Qu.:1.000  
```

```
Median :4.000   Median :2.000   Median :4.000   Median :2.000
Mean   :3.643   Mean   :2.617   Mean   :3.843   Mean   :2.255
3rd Qu.:4.000   3rd Qu.:3.000   3rd Qu.:5.000   3rd Qu.:3.000
Max.   :5.000   Max.   :5.000   Max.   :6.000   Max.   :5.000

  Economical        Health          Family        Calories
Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
1st Qu.:3.000   1st Qu.:3.000   1st Qu.:3.000   1st Qu.:2.000
Median :3.000   Median :4.000   Median :4.000   Median :3.000
Mean   :3.217   Mean   :3.809   Mean   :3.877   Mean   :2.702
3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:5.000   3rd Qu.:3.000
Max.   :5.000   Max.   :5.000   Max.   :6.000   Max.   :5.000

    Plain           Crisp          Regular          Sugar
Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
1st Qu.:1.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:1.000
Median :2.000   Median :3.000   Median :3.000   Median :2.000
Mean   :2.268   Mean   :3.204   Mean   :3.072   Mean   :2.145
3rd Qu.:3.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:3.000
Max.   :5.000   Max.   :6.000   Max.   :5.000   Max.   :5.000

    Fruit          Process         Quality          Treat          Boring
Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.00   Min.   :1.00
1st Qu.:1.000   1st Qu.:2.000   1st Qu.:3.000   1st Qu.:2.00   1st Qu.:1.00
Median :1.000   Median :3.000   Median :4.000   Median :3.00   Median :2.00
Mean   :1.694   Mean   :2.936   Mean   :3.694   Mean   :2.63   Mean   :1.83
3rd Qu.:3.000   3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:3.00   3rd Qu.:2.00
Max.   :5.000   Max.   :6.000   Max.   :5.000   Max.   :6.00   Max.   :5.00

  Nutritious
Min.   :1.000
1st Qu.:3.000
Median :4.000
Mean   :3.664
3rd Qu.:4.000
Max.   :5.000
```

It has been observed that there are 235 observations and 26 variables of the 12 different brands of cereals. All the attributes are 'ordinal' and has been marked on the scale between 1-5, but few variables mistakenly entered as 6. Hence, need to be corrected.

[41]:
```
cereal[cereal==6] <- 5
```

Structure of data

[42]:
```
str(cereal)
```

```
'data.frame':   235 obs. of  26 variables:
```

```
 $ Cereals   : Factor w/ 12 levels "AllBran","CMuesli",..: 12 9 9 2 3 8 9 9 8 3
…
 $ Filling   : int  5 1 5 5 4 4 4 4 4 4 …
 $ Natural   : int  5 2 4 5 5 4 4 3 3 3 …
 $ Fibre     : int  5 2 5 5 3 4 3 3 3 3 …
 $ Sweet     : int  1 1 5 3 2 2 2 2 2 2 …
 $ Easy      : num  2 5 5 5 5 5 5 5 5 5 …
 $ Salt      : int  1 2 3 2 2 2 1 1 1 1 …
 $ Satisfying: num  5 5 5 5 5 5 5 5 5 5 …
 $ Energy    : int  4 1 5 5 4 4 5 4 4 4 …
 $ Fun       : int  1 1 5 5 5 5 5 4 4 4 …
 $ Kids      : num  4 5 5 5 5 5 5 5 5 5 …
 $ Soggy     : int  5 3 3 3 1 1 1 1 1 1 …
 $ Economical: int  5 5 3 3 5 5 5 3 3 3 …
 $ Health    : int  5 2 5 5 5 4 5 4 4 4 …
 $ Family    : num  5 5 5 5 3 5 5 5 5 5 …
 $ Calories  : int  1 1 1 1 3 3 3 2 2 2 …
 $ Plain     : int  3 5 1 1 1 1 1 3 3 3 …
 $ Crisp     : num  1 5 5 1 5 5 5 4 4 4 …
 $ Regular   : int  4 1 4 4 3 3 3 4 4 4 …
 $ Sugar     : int  1 2 3 2 1 2 2 1 1 1 …
 $ Fruit     : int  1 1 1 5 1 1 1 1 1 1 …
 $ Process   : num  3 5 2 2 3 3 3 2 2 2 …
 $ Quality   : int  5 2 5 5 5 5 5 4 4 4 …
 $ Treat     : num  1 1 4 5 5 5 5 2 2 2 …
 $ Boring    : int  1 1 1 1 1 1 1 1 1 1 …
 $ Nutritious: int  5 3 5 5 4 4 4 3 3 3 …
```

2.1 Creating a correlation matrix and correlation plot

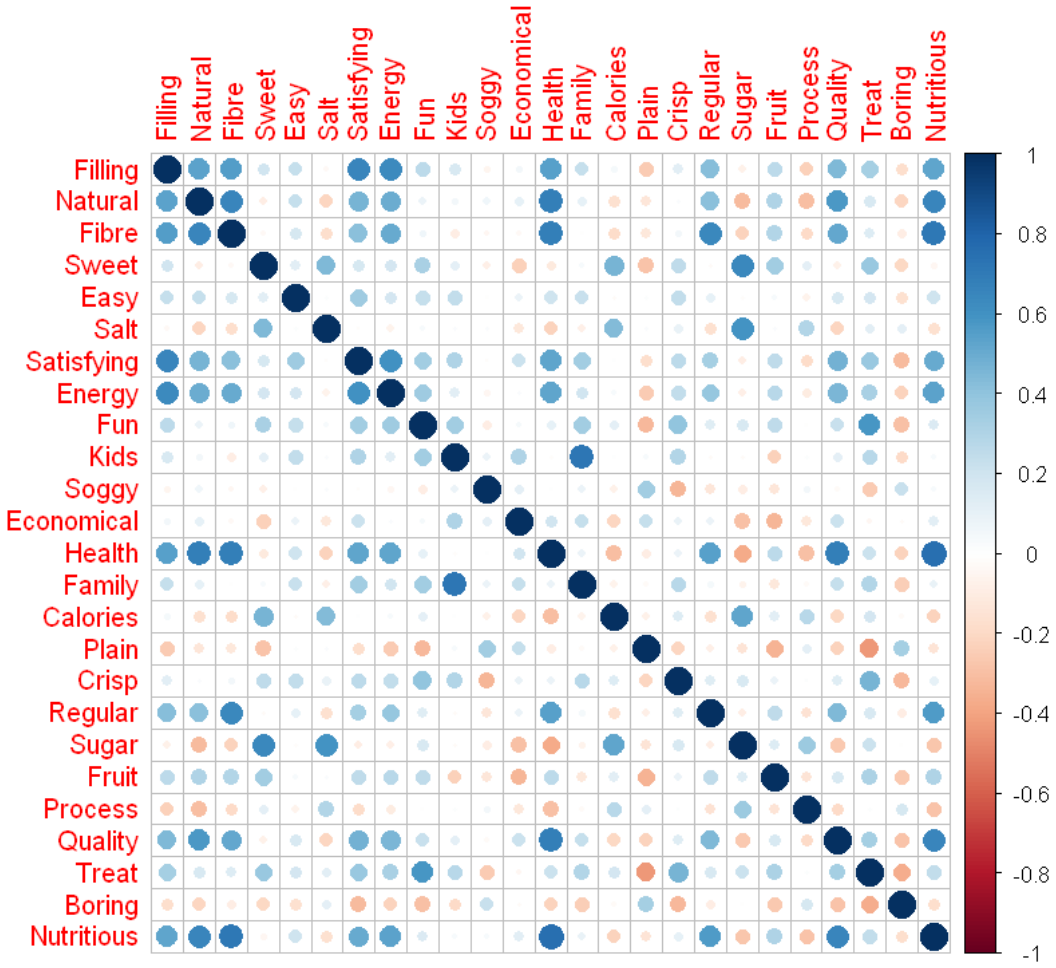
[6]:
```
cereal1 <- cereal[,-1]
cor.cereal <- cor(cereal1)
round(cor.cereal,2)
```

	Filling	Natural	Fibre	Sweet	Easy	Salt	Satisfying	Energy	Fun	Kids	...	Plain
Filling	1.00	0.54	0.55	0.19	0.24	-0.04	0.65	0.64	0.27	0.16	...	-0.25
Natural	0.54	1.00	0.65	-0.09	0.23	-0.22	0.47	0.49	0.08	0.06	...	-0.14
Fibre	0.55	0.65	1.00	-0.04	0.18	-0.17	0.41	0.50	0.06	-0.09	...	-0.12
Sweet	0.19	-0.09	-0.04	1.00	0.12	0.44	0.18	0.18	0.33	0.12	...	-0.29
Easy	0.24	0.23	0.18	0.12	1.00	0.01	0.35	0.18	0.24	0.24	...	0.02
Salt	-0.04	-0.22	-0.17	0.44	0.01	1.00	-0.01	-0.07	0.03	0.02	...	0.02
Satisfying	0.65	0.47	0.41	0.18	0.35	-0.01	1.00	0.60	0.35	0.30	...	-0.18
Energy	0.64	0.49	0.50	0.18	0.18	-0.07	0.60	1.00	0.35	0.13	...	-0.26
Fun	0.27	0.08	0.06	0.33	0.24	0.03	0.35	0.35	1.00	0.34	...	-0.32
Kids	0.16	0.06	-0.09	0.12	0.24	0.02	0.30	0.13	0.34	1.00	...	0.03
Soggy	-0.06	0.07	-0.04	-0.08	-0.01	0.02	-0.01	-0.05	-0.10	0.09	...	0.35
Economical	0.05	0.10	-0.03	-0.24	0.09	-0.13	0.21	0.03	0.04	0.30	...	0.23
Health	0.55	0.69	0.68	-0.12	0.20	-0.23	0.52	0.52	0.10	-0.01	...	-0.10
Family	0.23	0.11	-0.01	0.04	0.23	-0.09	0.35	0.19	0.35	0.72	...	-0.03
Calories	0.05	-0.16	-0.19	0.47	-0.02	0.44	0.01	0.03	0.11	0.01	...	-0.08
Plain	-0.25	-0.14	-0.12	-0.29	0.02	0.02	-0.18	-0.26	-0.32	0.03	...	1.00
Crisp	0.13	0.02	0.05	0.26	0.24	0.10	0.26	0.25	0.40	0.29	...	-0.21
Regular	0.42	0.42	0.65	-0.03	0.11	-0.16	0.33	0.39	0.14	-0.03	...	-0.08
Sugar	-0.08	-0.32	-0.23	0.65	-0.02	0.59	-0.09	-0.09	0.17	-0.02	...	-0.15
Fruit	0.26	0.30	0.29	0.35	0.04	0.03	0.25	0.27	0.25	-0.23	...	-0.34
Process	-0.23	-0.31	-0.19	0.11	-0.07	0.30	-0.19	-0.10	-0.01	0.01	...	0.12
Quality	0.44	0.58	0.51	-0.08	0.16	-0.22	0.47	0.46	0.22	0.11	...	-0.23
Treat	0.34	0.17	0.14	0.37	0.18	0.12	0.37	0.32	0.58	0.28	...	-0.43
Boring	-0.18	-0.22	-0.10	-0.20	-0.17	0.11	-0.32	-0.22	-0.30	-0.20	...	0.33
Nutritious	0.53	0.65	0.71	-0.05	0.20	-0.16	0.50	0.54	0.16	0.03	...	-0.14

If the relationship is very strong, the value of the output to be closer to a positive one. If the relationship is inverse, then the value corresponds to a negative value, basically implying that the two quantities are inversely proportional. And, anything closer to zero, implies that the quantities are not related at all.

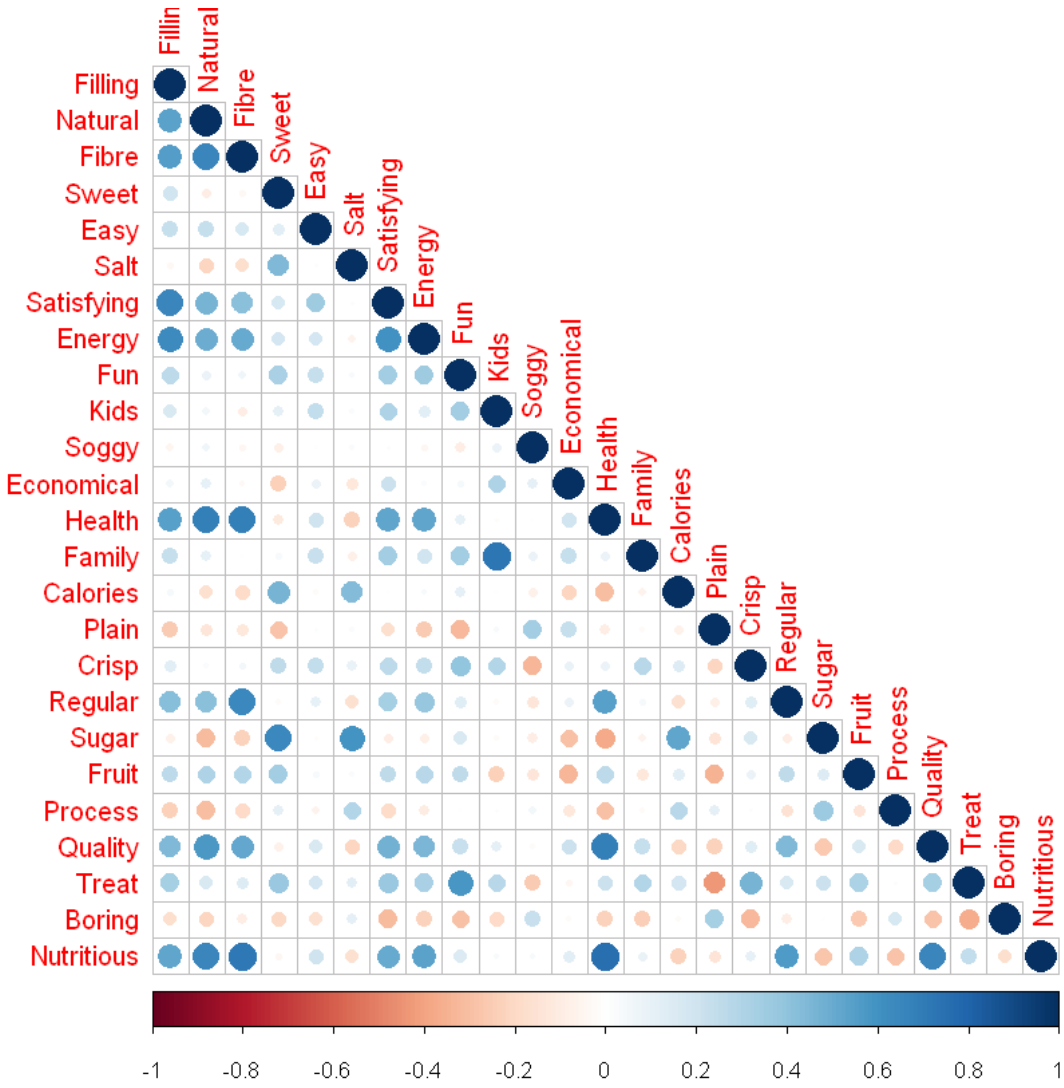
[40]:
```
library(corrplot)
```

[39]:
```
corrplot(cor.cereal, method = "circle",bg = "white")
```

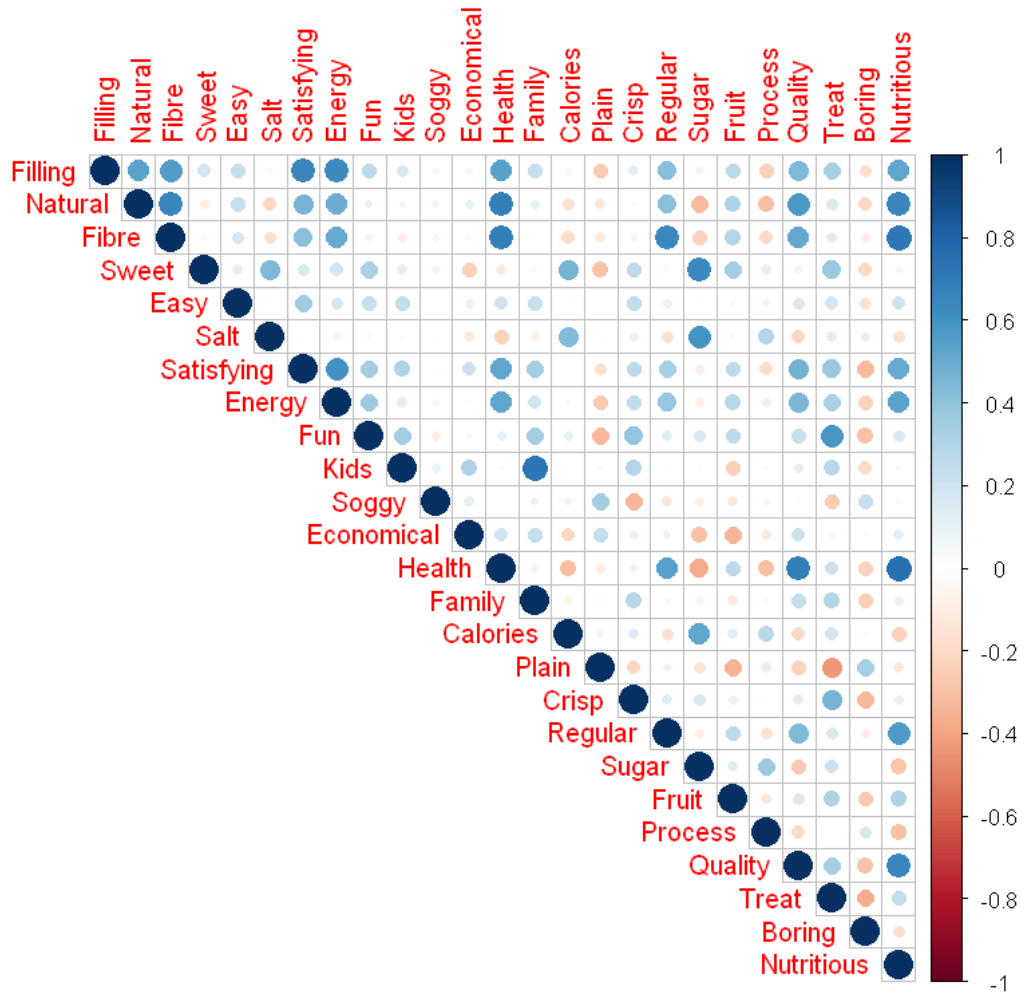


Anything that is in blue is highly correlated, anything that's in red or closer to red, is inversely correlated and if it's white it basically means it's closer to zero(no strong relationship).

[36]:
```
corrplot(cor.cereal, type = "lower",bg = "white")
```



[38]:
```
corrplot(cor.cereal, type = "upper",bg = "white")
```



2.2 Testing the assumptions to do factor analysis

KMO test for Measuirng Sampling Accuracy

KMO test for measuring sampling accuracy(MSA): Kaiser-Meyer-Olkin test gives us the measures to check whether the data is suitable for data analysis or not. Observing each variable, check the model that is present, these are the two adequacy checkpoints which are present. While observing the variable we are trying to estimate the propertion of the varience that is present between all the obsereved variables. Proportion of varience is basically confirming whether the varience is very high or the varience is very low, high isn't preferable.

[15]:
```
library(psych)
cereal.KMO <- KMO(cor.cereal)
```

```
cereal.KMO
```

```
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = cor.cereal)
Overall MSA =  0.85
MSA for each item =
   Filling    Natural      Fibre      Sweet       Easy       Salt Satisfying
      0.89       0.90       0.88       0.78       0.83       0.82       0.91
    Energy        Fun       Kids      Soggy Economical     Health     Family
      0.91       0.85       0.67       0.63       0.73       0.92       0.73
  Calories      Plain      Crisp    Regular      Sugar      Fruit    Process
      0.86       0.82       0.83       0.87       0.78       0.77       0.80
   Quality      Treat     Boring Nutritious
      0.91       0.88       0.87       0.92
```

According to Kaiser himself a KMO in the .80's is meritorious!. Thus, our sample is large enough for factor analysis or principal component analysis.

[16]:
```
cereal.barlett <- cortest.bartlett(cor.cereal, nrow(cereal))
cereal.barlett
```

$chisq 2877.73856777721

$p.value 0

$df 300

Bartlett Test of Spehericity - The p-value is <.001, thus the null hypothesis is rejected (The null hypothesis is that the corelation matrix is an identity matrix i.e. there is no scope for dimentionality reduction.). Thus, the dimensionality reduction is a possibility using Factor analysis or principal component analysis.

2.3 Running PCA to identify the number of factors

[17]:
```
cereal.pca <- princomp(cereal1,scores = TRUE, cor = TRUE)
summary(cereal.pca)
```

```
Importance of components:
                          Comp.1    Comp.2     Comp.3     Comp.4     Comp.5
Standard deviation     2.5515645 1.9473508 1.57931249 1.29699429 1.04196617
Proportion of Variance 0.2604193 0.1516870 0.09976912 0.06728777 0.04342774
Cumulative Proportion  0.2604193 0.4121063 0.51187538 0.57916315 0.62259089
                           Comp.6     Comp.7     Comp.8     Comp.9    Comp.10
Standard deviation     0.97215571 0.92371682 0.88941254 0.85594266 0.83528812
Proportion of Variance 0.03780347 0.03413011 0.03164219 0.02930551 0.02790825
Cumulative Proportion  0.66039436 0.69452447 0.72616666 0.75547217 0.78338042
                          Comp.11    Comp.12    Comp.13    Comp.14   Comp.15
Standard deviation     0.80508009 0.74210795 0.72900837 0.69819277 0.6456121
Proportion of Variance 0.02592616 0.02202897 0.02125813 0.01949893 0.0166726
Cumulative Proportion  0.80930658 0.83133555 0.85259368 0.87209260 0.8887652
                          Comp.16    Comp.17    Comp.18    Comp.19    Comp.20
```

```
Standard deviation     0.62203549 0.60340597 0.60072703 0.55329582 0.52496345
Proportion of Variance 0.01547713 0.01456395 0.01443492 0.01224545 0.01102346
Cumulative Proportion  0.90424233 0.91880628 0.93324119 0.94548664 0.95651011
                          Comp.21     Comp.22     Comp.23     Comp.24
Standard deviation     0.51267066 0.492791255 0.467311607 0.445682156
Proportion of Variance 0.01051325 0.009713729 0.008735206 0.007945303
Cumulative Proportion  0.96702336 0.976737086 0.985472292 0.993417595
                           Comp.25
Standard deviation     0.405660102
Proportion of Variance 0.006582405
Cumulative Proportion  1.000000000
```

Interpretations from the above values,

According to KMO statistics concept, all principal component values which has an Eigen value of greater than 1, can be taken into consideration for the dimension reduction technique.

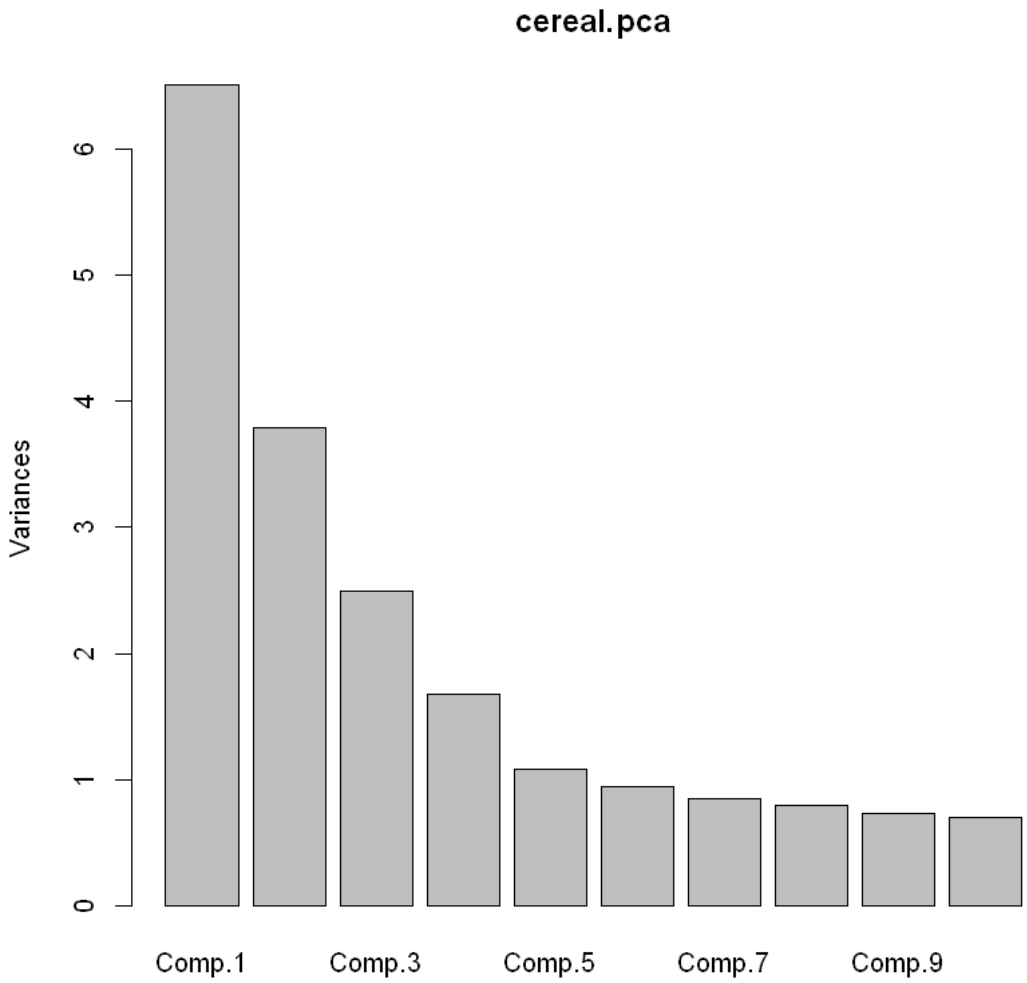
Based on that criteria, in our case we will take first 5 principal components into consideration.

From the output values, Comp 1 explains 26% of variation of the whole data set, likewise Comp 2 explains 15%, Comp 3 explains 10%, Comp 4 explians 7% and Comp 5 explains 4% of variation of the entire data set.

Therefore, it implies PCA1 to PCA5 itself are able to explain 62% of the data.

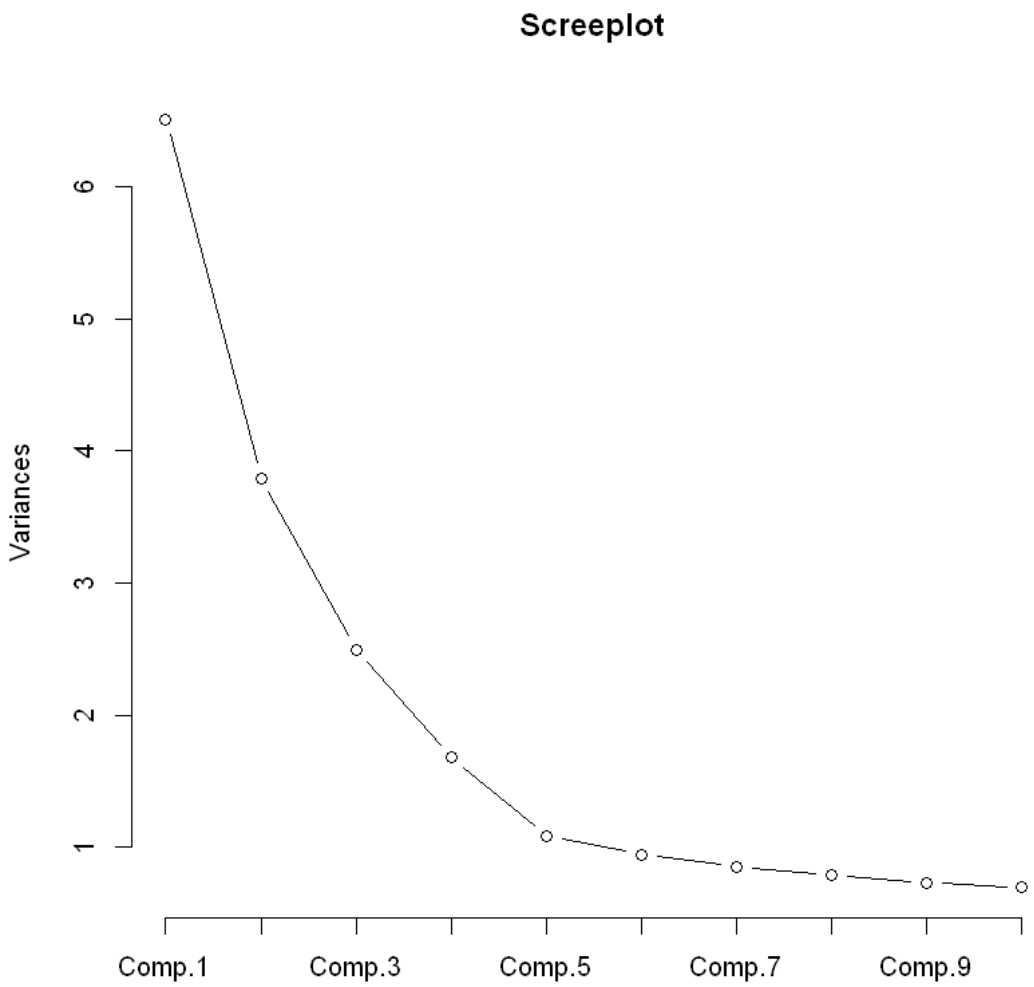
Plot of principal components with variance A PCA plot converts the correlation among all of the cells into 2D graph. Values that are highly correlated cluster together.

[19]:
```
plot(cereal.pca,col = "grey")
```



After PC5 the bar flattens and there is no further steep.

[20]:
```
screeplot(cereal.pca,type = "lines",main = "Screeplot")
```



In the above Screeplot , after the PC5 which is above Eigen value(Variances) 1,the "Elbow" shape becomes a line and flattens. The analysis suggests that the number of factors = 5.

[21]:
```
loadings(cereal.pca)
```

```
Loadings:
          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
Filling    0.293                0.176  0.107         0.232  0.406
Natural    0.294 -0.132         0.101  0.139                       0.210
Fibre      0.287 -0.123  0.210  0.138 -0.155        -0.148  0.119
Sweet              0.398  0.117  0.143  0.155  0.114                -0.178
Easy       0.136        -0.171  0.121         0.667 -0.478         0.163
Salt               0.280         0.373 -0.127  0.109  0.210 -0.220
```

```
Satisfying  0.292        -0.108  0.152  0.101  0.124  0.182  0.110
Energy      0.285                0.131        -0.128  0.112  0.297  0.104
Fun         0.161  0.270 -0.162 -0.113        -0.175 -0.230        -0.353
Kids               0.129 -0.498               -0.103         0.139
Soggy             -0.142 -0.113  0.446  0.479 -0.173 -0.173 -0.311 -0.103
Economical        -0.147 -0.366        -0.237         0.458 -0.349 -0.162
Health      0.318 -0.161                                      -0.193
Family      0.124        -0.460         0.137 -0.199 -0.120  0.172
Calories           0.324  0.110  0.216                0.280  0.101  0.259
Plain      -0.129 -0.208 -0.158  0.374 -0.143  0.232 -0.120
Crisp       0.121  0.252 -0.171 -0.185 -0.402  0.186
Regular     0.243         0.142        -0.381        -0.142        -0.359
Sugar              0.384  0.142  0.201                       -0.114 -0.113
Fruit       0.154  0.148  0.342 -0.111  0.282        -0.185 -0.189 -0.178
Process    -0.134  0.155         0.263 -0.339 -0.409 -0.364         0.480
Quality     0.295                              -0.225        -0.307  0.206
Treat       0.190  0.302        -0.150        -0.190        -0.155
Boring     -0.162 -0.152         0.334 -0.158 -0.153         0.361 -0.418
Nutritious  0.316 -0.116  0.102  0.114                       -0.167
            Comp.10 Comp.11 Comp.12 Comp.13 Comp.14 Comp.15 Comp.16 Comp.17
Filling                      0.109           0.108   0.111           0.278
Natural                     -0.281   0.110          -0.272  -0.379   0.167
Fibre       -0.177                           -0.139          -0.213   0.197
Sweet       -0.130           0.227           -0.214  -0.132   0.294   0.306
Easy         0.248   0.230   0.114   0.188
Salt                 0.352  -0.158  -0.282   0.306   0.206  -0.379  -0.218
Satisfying   0.127  -0.124   0.232  -0.108   0.348           0.170  -0.137
Energy       0.231  -0.283   0.117  -0.347  -0.160   0.107          -0.157
Fun          0.390  -0.167           0.163           0.312  -0.344  -0.172
Kids        -0.298   0.116  -0.108                  -0.132  -0.189
Soggy                -0.121          -0.171  -0.348
Economical   0.223           0.261   0.205          -0.356  -0.132   0.102
Health                                                      0.124   0.109
Family      -0.369                                   0.141           0.144  -0.243
Calories             -0.310  -0.343   0.520  -0.233                  -0.201
Plain       -0.173  -0.458  -0.179           0.261   0.378   0.205   0.143
Crisp                -0.261  -0.358  -0.465  -0.247  -0.227   0.106
Regular     -0.345  -0.103   0.262   0.243  -0.104                  -0.117
Sugar       -0.190   0.148   0.218          -0.141           0.118
Fruit                -0.302                   0.462  -0.427          -0.166
Process      0.131  -0.132   0.238           0.173  -0.227           0.137
Quality              0.225           0.159  -0.109   0.102   0.425  -0.362
Treat        0.198   0.124  -0.300   0.179   0.191   0.181   0.163   0.512
Boring       0.359   0.227  -0.299                  -0.330   0.218
Nutritious           0.105  -0.156                                  -0.125
            Comp.18 Comp.19 Comp.20 Comp.21 Comp.22 Comp.23 Comp.24 Comp.25
Filling      0.188   0.268           0.416   0.124   0.413   0.230
Natural     -0.223   0.437          -0.335                          -0.315
```

```
Fibre                    0.339   0.183  -0.277  -0.503   0.189   0.344
Sweet      -0.381  -0.169   0.120   0.101   0.227  -0.213          -0.331
Easy                     -0.191
Salt                     -0.121   0.120   0.159  -0.131
Satisfying  0.313          0.412  -0.465          -0.150
Energy     -0.314         -0.519  -0.141  -0.112  -0.115           0.116
Fun        -0.182   0.135   0.324   0.105                  -0.102
Kids       -0.203  -0.356          -0.221   0.245   0.248           0.394
Soggy       0.419
Economical                -0.117   0.160  -0.178
Health                             0.192   0.226          -0.767   0.270
Family              0.219  -0.143   0.309  -0.267  -0.272  -0.186  -0.248
Calories    0.121  -0.203                                  -0.160
Plain      -0.283   0.135                                   0.125
Crisp       0.210   0.174   0.126           0.108
Regular     0.283          -0.270  -0.178   0.305                  -0.188
Sugar               0.395          -0.252  -0.421   0.268  -0.136   0.338
Fruit                    -0.127   0.144                   0.160   0.217
Process                            0.102                           -0.118
Quality    -0.160   0.160                   0.261           0.380   0.136
Treat       0.204          -0.316  -0.211  -0.151  -0.149
Boring     -0.101
Nutritious          -0.461                  -0.461   0.453          -0.332

               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
SS loadings      1.00   1.00   1.00   1.00   1.00   1.00   1.00   1.00   1.00
Proportion Var   0.04   0.04   0.04   0.04   0.04   0.04   0.04   0.04   0.04
Cumulative Var   0.04   0.08   0.12   0.16   0.20   0.24   0.28   0.32   0.36
               Comp.10 Comp.11 Comp.12 Comp.13 Comp.14 Comp.15 Comp.16 Comp.17
SS loadings       1.00    1.00    1.00    1.00    1.00    1.00    1.00    1.00
Proportion Var    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.04
Cumulative Var    0.40    0.44    0.48    0.52    0.56    0.60    0.64    0.68
               Comp.18 Comp.19 Comp.20 Comp.21 Comp.22 Comp.23 Comp.24 Comp.25
SS loadings       1.00    1.00    1.00    1.00    1.00    1.00    1.00    1.00
Proportion Var    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.04
Cumulative Var    0.72    0.76    0.80    0.84    0.88    0.92    0.96    1.00
```

[22]:
```
print(cereal.pca$loadings, cutoff = 0.2)
```

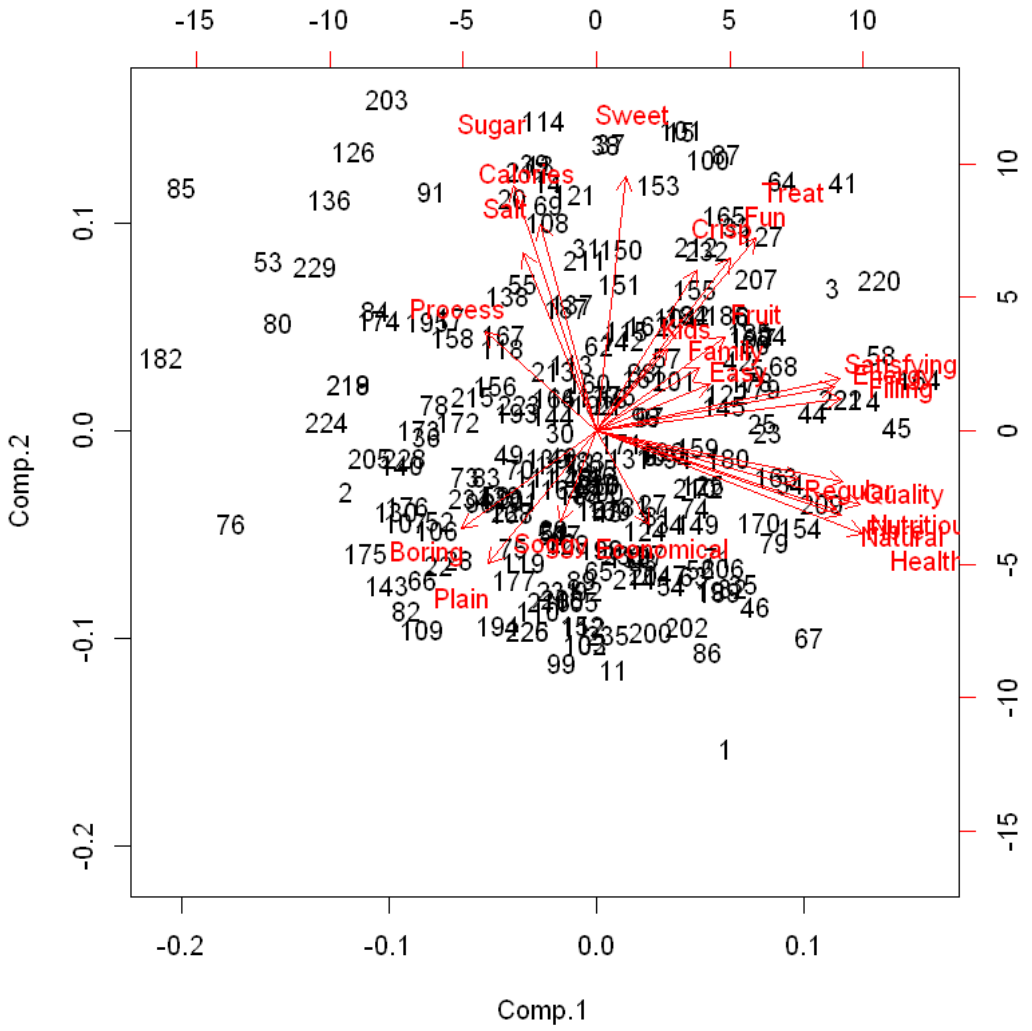
```
Loadings:
           Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
Filling     0.293                                  0.232  0.406
Natural     0.294                                                  0.210
Fibre       0.287         0.210
Sweet               0.398
Easy                                        0.667 -0.478
Salt                0.280         0.373                0.210 -0.220
```

```
Satisfying  0.292
Energy      0.285                                                 0.297
Fun                 0.270                             -0.230          -0.353
Kids                      -0.498
Soggy                             0.446  0.479                  -0.311
Economical                -0.366         -0.237          0.458 -0.349
Health      0.318
Family                    -0.460
Calories            0.324          0.216                 0.280          0.259
Plain              -0.208          0.374          0.232
Crisp               0.252                -0.402
Regular     0.243                        -0.381                         -0.359
Sugar               0.384          0.201
Fruit                       0.342          0.282
Process                           0.263 -0.339 -0.409 -0.364          0.480
Quality     0.295                               -0.225         -0.307  0.206
Treat               0.302
Boring                            0.334                          0.361 -0.418
Nutritious  0.316
            Comp.10 Comp.11 Comp.12 Comp.13 Comp.14 Comp.15 Comp.16 Comp.17
Filling                                                                0.278
Natural                     -0.281                  -0.272  -0.379
Fibre                                                       -0.213
Sweet                        0.227          -0.214           0.294   0.306
Easy         0.248   0.230
Salt                 0.352          -0.282   0.306   0.206  -0.379  -0.218
Satisfying                   0.232           0.348
Energy       0.231  -0.283          -0.347
Fun          0.390                                  0.312  -0.344
Kids        -0.298
Soggy                                       -0.348
Economical   0.223           0.261   0.205          -0.356
Health
Family      -0.369                                                     -0.243
Calories            -0.310  -0.343   0.520  -0.233                     -0.201
Plain               -0.458                  0.261   0.378   0.205
Crisp               -0.261  -0.358  -0.465  -0.247  -0.227
Regular     -0.345           0.262   0.243
Sugar                        0.218
Fruit               -0.302                  0.462  -0.427
Process                      0.238                  -0.227
Quality              0.225                                  0.425  -0.362
Treat                       -0.300                                  0.512
Boring       0.359   0.227  -0.299                  -0.330   0.218
Nutritious
            Comp.18 Comp.19 Comp.20 Comp.21 Comp.22 Comp.23 Comp.24 Comp.25
Filling              0.268           0.416           0.413   0.230
Natural     -0.223   0.437          -0.335                          -0.315
```

```
Fibre                    0.339          -0.277  -0.503           0.344
Sweet       -0.381                      0.227  -0.213          -0.331
Easy
Salt
Satisfying   0.313          0.412  -0.465
Energy      -0.314         -0.519
Fun                        0.324
Kids        -0.203  -0.356         -0.221   0.245   0.248           0.394
Soggy        0.419
Economical
Health                                      0.226          -0.767   0.270
Family               0.219           0.309  -0.267  -0.272          -0.248
Calories            -0.203
Plain       -0.283
Crisp        0.210
Regular      0.283         -0.270           0.305
Sugar                0.395         -0.252  -0.421   0.268           0.338
Fruit                                                               0.217
Process
Quality                                     0.261           0.380
Treat        0.204         -0.316  -0.211
Boring
Nutritious          -0.461                 -0.461   0.453          -0.332

               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
SS loadings      1.00   1.00   1.00   1.00   1.00   1.00   1.00   1.00   1.00
Proportion Var   0.04   0.04   0.04   0.04   0.04   0.04   0.04   0.04   0.04
Cumulative Var   0.04   0.08   0.12   0.16   0.20   0.24   0.28   0.32   0.36
               Comp.10 Comp.11 Comp.12 Comp.13 Comp.14 Comp.15 Comp.16 Comp.17
SS loadings       1.00    1.00    1.00    1.00    1.00    1.00    1.00    1.00
Proportion Var    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.04
Cumulative Var    0.40    0.44    0.48    0.52    0.56    0.60    0.64    0.68
               Comp.18 Comp.19 Comp.20 Comp.21 Comp.22 Comp.23 Comp.24 Comp.25
SS loadings       1.00    1.00    1.00    1.00    1.00    1.00    1.00    1.00
Proportion Var    0.04    0.04    0.04    0.04    0.04    0.04    0.04    0.04
Cumulative Var    0.72    0.76    0.80    0.84    0.88    0.92    0.96    1.00
```

[23]:
```
biplot(cereal.pca, choices = c(1,2))
```



From the loadings, we are able to interpret that Filling,Natural,Fibre,Satisfying,Energy,Health,Regular,Quality, Nutritious are the variables that are added and gives the weightage to Comp 1. Similarly, we can identify the variables that adds weightage in the Comps 2,3,4,5. As a result, we have 4 components determning the variability of our data.

2.3.1 Factor Analysis

Factor analysis without rotation

[24]:
```
factanal(cereal1,factors = 5,rotation = "none")
```

```
Call:
factanal(x = cereal1, factors = 5, rotation = "none")
```

```
Uniquenesses:
   Filling    Natural      Fibre      Sweet       Easy       Salt Satisfying
     0.286      0.389      0.312      0.360      0.847      0.517      0.369
    Energy        Fun       Kids      Soggy Economical     Health     Family
     0.431      0.525      0.239      0.775      0.709      0.214      0.352
  Calories      Plain      Crisp    Regular      Sugar      Fruit    Process
     0.578      0.549      0.648      0.550      0.199      0.562      0.764
   Quality      Treat     Boring Nutritious
     0.392      0.391      0.671      0.241

Loadings:
           Factor1 Factor2 Factor3 Factor4 Factor5
Filling     0.719   0.243           0.143  -0.338
Natural     0.763  -0.116
Fibre       0.757  -0.113  -0.283   0.137
Sweet               0.757  -0.244
Easy        0.289   0.188   0.168
Salt       -0.258   0.495  -0.215   0.341
Satisfying  0.690   0.295   0.132          -0.205
Energy      0.682   0.237                  -0.204
Fun         0.287   0.533   0.191  -0.266
Kids        0.164   0.365   0.762   0.135
Soggy              -0.177   0.143   0.413
Economical  0.183  -0.172   0.440   0.143   0.117
Health      0.854  -0.177  -0.112           0.110
Family      0.260   0.310   0.694
Calories   -0.237   0.536  -0.206   0.131  -0.134
Plain      -0.232  -0.352   0.187   0.480
Crisp       0.197   0.450   0.172  -0.233   0.164
Regular     0.605          -0.195           0.196
Sugar      -0.356   0.716  -0.325   0.185   0.147
Fruit       0.324   0.257  -0.447  -0.247
Process    -0.336   0.227           0.198   0.176
Quality     0.746                           0.213
Treat       0.368   0.589          -0.328   0.124
Boring     -0.314  -0.283  -0.114   0.370
Nutritious  0.834          -0.142           0.170

                Factor1 Factor2 Factor3 Factor4 Factor5
SS loadings       6.049   3.368   2.062   1.165   0.486
Proportion Var    0.242   0.135   0.082   0.047   0.019
Cumulative Var    0.242   0.377   0.459   0.506   0.525

Test of the hypothesis that 5 factors are sufficient.
The chi square statistic is 318.9 on 185 degrees of freedom.
The p-value is 3.48e-09
```

From the above, we observe that the variables “Easy”, “Soggy”, “Economical”, “Crisp”, “Process” and “Boring” have high ’Uniquenesses - means it does not correlate with other variables and can be used as it is. We observe that the loadings are not very clear, hence it has to be rotated to increase the very clear interpretability. It is done to maximize the high loadings and minimize low loadings.

2.3.2 Factor loadings with Varimax rotation

[25]:
```
factanal(cereal1,factors = 5,rotation = "varimax")
```

```
Call:
factanal(x = cereal1, factors = 5, rotation = "varimax")

Uniquenesses:
   Filling    Natural      Fibre      Sweet       Easy       Salt Satisfying
     0.286      0.389      0.312      0.360      0.847      0.517      0.369
    Energy        Fun       Kids      Soggy Economical     Health     Family
     0.431      0.525      0.239      0.775      0.709      0.214      0.352
  Calories      Plain      Crisp    Regular      Sugar      Fruit    Process
     0.578      0.549      0.648      0.550      0.199      0.562      0.764
   Quality      Treat     Boring Nutritious
     0.392      0.391      0.671      0.241

Loadings:
           Factor1 Factor2 Factor3 Factor4 Factor5
Filling     0.646           0.147   0.186   0.486
Natural     0.731  -0.214                   0.155
Fibre       0.816
Sweet               0.696   0.354           0.164
Easy        0.228                   0.292
Salt                0.686
Satisfying  0.571           0.202   0.369   0.355
Energy      0.610           0.228   0.161   0.343
Fun         0.126   0.155   0.545   0.364
Kids                                0.868
Soggy                      -0.452   0.138
Economical         -0.257  -0.190   0.410  -0.104
Health      0.839  -0.271
Family                     0.133   0.789
Calories   -0.157   0.591   0.122           0.179
Plain      -0.115          -0.635   0.100  -0.147
Crisp               0.148   0.465   0.318
Regular     0.658
Sugar      -0.177   0.854   0.171
Fruit       0.339   0.159   0.433  -0.294   0.155
Process    -0.218   0.380  -0.105          -0.181
Quality     0.682  -0.222   0.221   0.190
```

```
Treat       0.234   0.208   0.659   0.278
Boring     -0.150          -0.513  -0.193
Nutritious  0.849  -0.154

               Factor1 Factor2 Factor3 Factor4 Factor5
SS loadings      5.044   2.583   2.440   2.353   0.711
Proportion Var   0.202   0.103   0.098   0.094   0.028
Cumulative Var   0.202   0.305   0.403   0.497   0.525

Test of the hypothesis that 5 factors are sufficient.
The chi square statistic is 318.9 on 185 degrees of freedom.
The p-value is 3.48e-09
```

After the rotation we observe that the 5 factors are able to explain about 52% of the data. We infer that based on the latent relation between the variables they are grouped into as "Factors". The factor loadings explains the weightage of each variable has in the Factors. We can also see that the some of the variables are present in more than one factor, which is known as the 'Communality' effect. As a result, the factor scores for all the 235 observations are known,from which prediction or classification can be done.

The variable that depend highly on factor -

Factor 1 -'State of well being', Filling,Natural,Fibre,Satisfying,Energy,Health,Regular,Quality,Nutritious

Factor 2 - 'Nutritional Information', Sweet,Salt, Calories, Sugar, Process,

Factor 3 - Fun, Soggy, Crisp, Fruit, Treat, Boring,Plain - 'Experience'

Factor 4 - Kids, Economical,Family - 'Family'

Factor 5 - Filling, Satisfying, Energy - 'Supporting variables'