Escalamiento

Tlahuiz Tenorio Giovanni Saúl & Leobardo Enriquez Hernandez

5/5/2024

Veamos cómo se ve la informacion.

```
##        ID          Year_Birth    Education         Marital_Status
##  Min.   :    0   Min.   :1893   Length:2240        Length:2240
##  1st Qu.: 2828   1st Qu.:1959   Class :character   Class :character
##  Median : 5458   Median :1970   Mode  :character   Mode  :character
##  Mean   : 5592   Mean   :1969
##  3rd Qu.: 8428   3rd Qu.:1977
##  Max.   :11191   Max.   :1996
##
##      Income          Kidhome          Teenhome      Dt_Customer
##  Min.   :  1730   Min.   :0.0000   Min.   :0.0000   Length:2240
##  1st Qu.: 35303   1st Qu.:0.0000   1st Qu.:0.0000   Class :character
##  Median : 51382   Median :0.0000   Median :0.0000   Mode  :character
##  Mean   : 52247   Mean   :0.4442   Mean   :0.5062
##  3rd Qu.: 68522   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :666666   Max.   :2.0000   Max.   :2.0000
##  NA's   :24
##     Recency         MntWines         MntFruits     MntMeatProducts
##  Min.   : 0.00   Min.   :   0.00   Min.   :  0.0   Min.   :   0.0
##  1st Qu.:24.00   1st Qu.:  23.75   1st Qu.:  1.0   1st Qu.:  16.0
##  Median :49.00   Median : 173.50   Median :  8.0   Median :  67.0
##  Mean   :49.11   Mean   : 303.94   Mean   : 26.3   Mean   : 166.9
##  3rd Qu.:74.00   3rd Qu.: 504.25   3rd Qu.: 33.0   3rd Qu.: 232.0
##  Max.   :99.00   Max.   :1493.00   Max.   :199.0   Max.   :1725.0
##
##  MntFishProducts  MntSweetProducts  MntGoldProds    NumDealsPurchases
##  Min.   :  0.00   Min.   :  0.00   Min.   :  0.00   Min.   : 0.000
##  1st Qu.:  3.00   1st Qu.:  1.00   1st Qu.:  9.00   1st Qu.: 1.000
##  Median : 12.00   Median :  8.00   Median : 24.00   Median : 2.000
##  Mean   : 37.53   Mean   : 27.06   Mean   : 44.02   Mean   : 2.325
##  3rd Qu.: 50.00   3rd Qu.: 33.00   3rd Qu.: 56.00   3rd Qu.: 3.000
##  Max.   :259.00   Max.   :263.00   Max.   :362.00   Max.   :15.000
##
##  NumWebPurchases  NumCatalogPurchases NumStorePurchases NumWebVisitsMonth
##  Min.   : 0.000   Min.   : 0.000      Min.   : 0.00     Min.   : 0.000
##  1st Qu.: 2.000   1st Qu.: 0.000      1st Qu.: 3.00     1st Qu.: 3.000
##  Median : 4.000   Median : 2.000      Median : 5.00     Median : 6.000
##  Mean   : 4.085   Mean   : 2.662      Mean   : 5.79     Mean   : 5.317
##  3rd Qu.: 6.000   3rd Qu.: 4.000      3rd Qu.: 8.00     3rd Qu.: 7.000
##  Max.   :27.000   Max.   :28.000      Max.   :13.00     Max.   :20.000
##
##   AcceptedCmp3      AcceptedCmp4      AcceptedCmp5      AcceptedCmp1
```

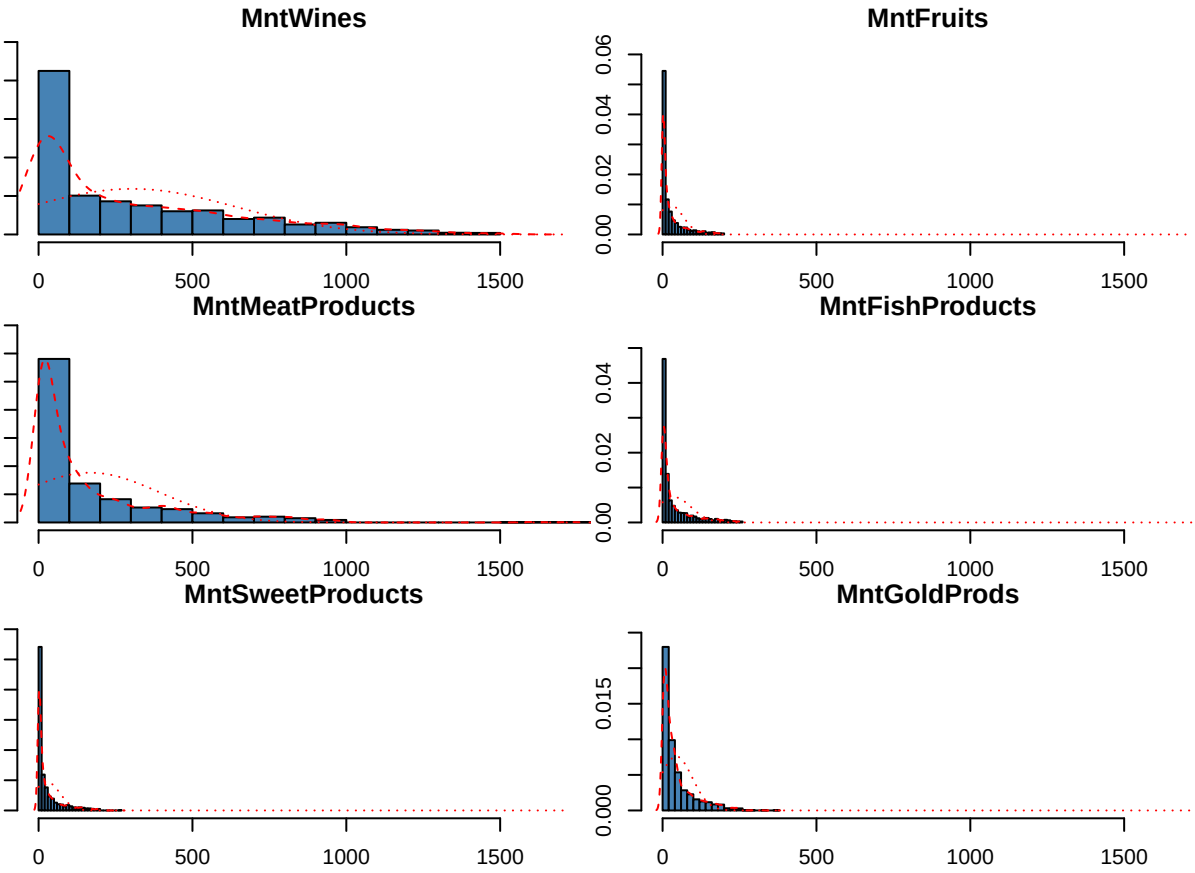
```
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.07277   Mean   :0.07455   Mean   :0.07277   Mean   :0.06429
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##
##   AcceptedCmp2        Complain        Z_CostContact   Z_Revenue
##  Min.   :0.00000   Min.   :0.000000   Min.   :3     Min.   :11
##  1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:3     1st Qu.:11
##  Median :0.00000   Median :0.000000   Median :3     Median :11
##  Mean   :0.01339   Mean   :0.009375   Mean   :3     Mean   :11
##  3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:3     3rd Qu.:11
##  Max.   :1.00000   Max.   :1.000000   Max.   :3     Max.   :11
##
##     Response
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.1491
##  3rd Qu.:0.0000
##  Max.   :1.0000
##
```

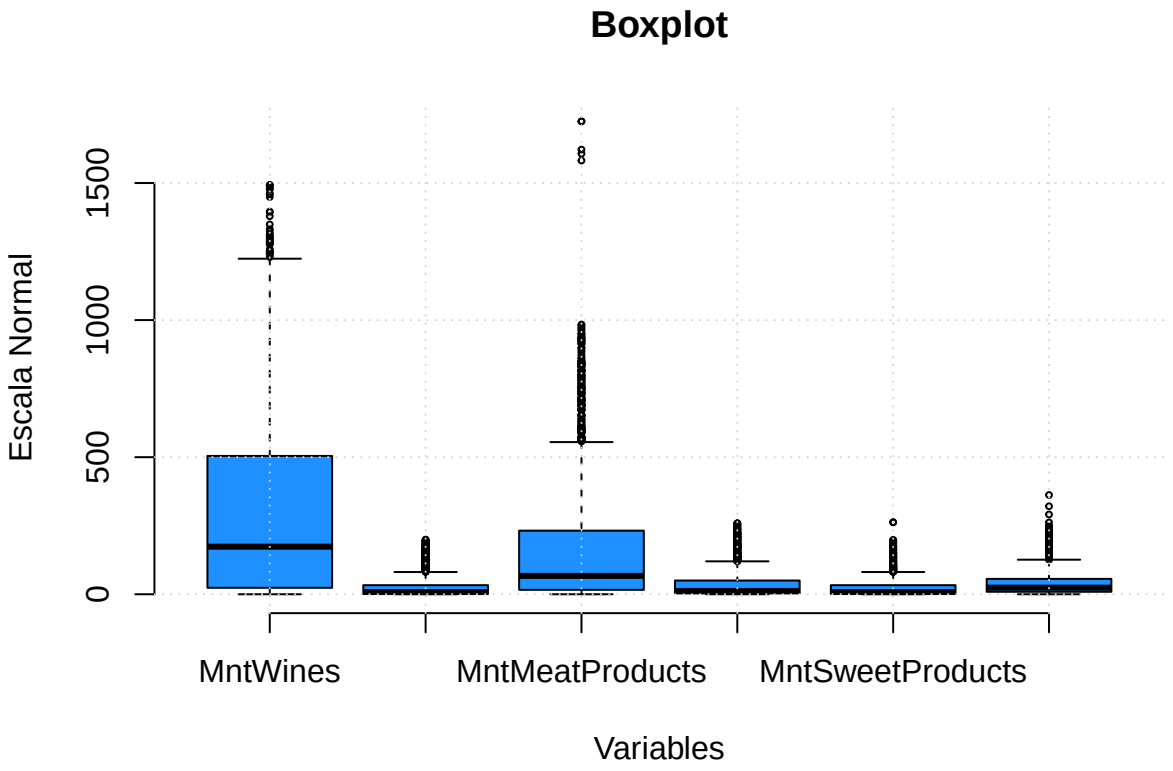
Ya que vimos un resumen rapido de la informacion lo que sigue es tomar las variables que nos serviran de auxiliares y las que nos serviran para el analisis.

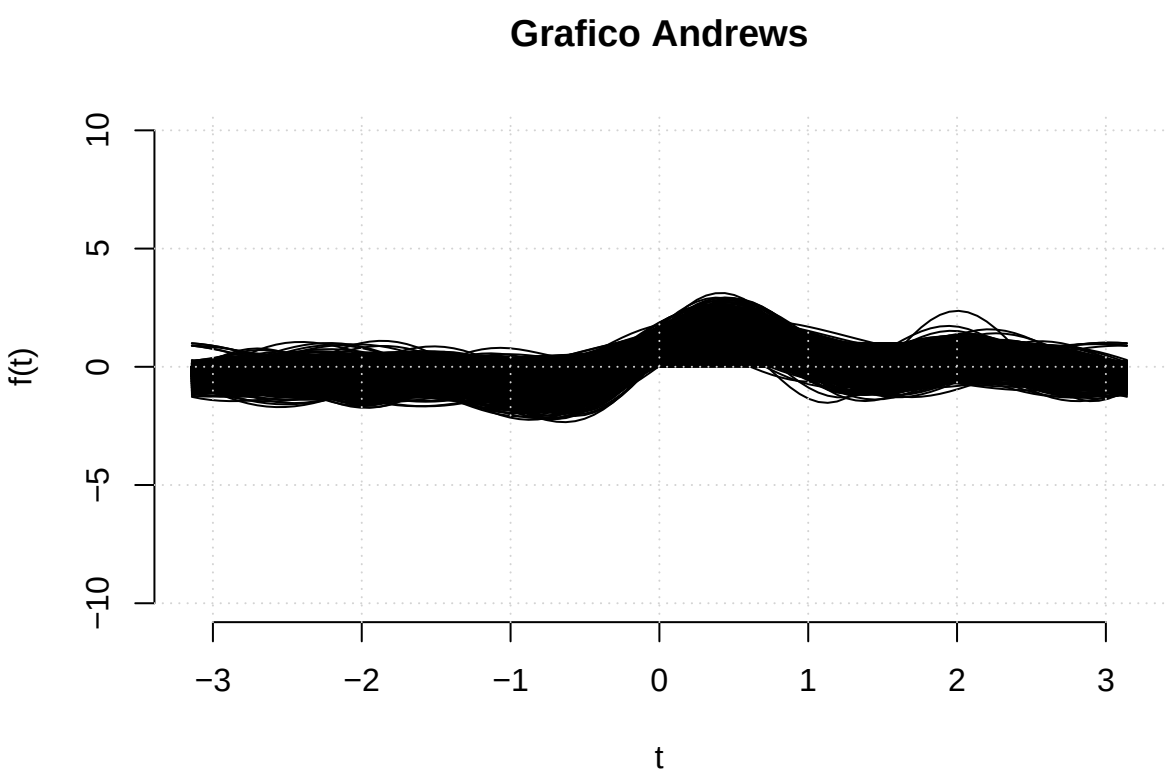
Extraemos los datos.

```
## # A tibble: 6 x 7
##      ID MntWines MntFruits MntMeatProducts MntFishProducts MntSweetProducts
##   <dbl>    <dbl>     <dbl>           <dbl>           <dbl>             <dbl>
## 1  5524      635        88             546             172                88
## 2  2174       11         1               6               2                 1
## 3  4141      426        49             127             111                21
## 4  6182       11         4              20              10                 3
## 5  5324      173        43             118              46                27
## 6  7446      520        42              98               0                42
## # i 1 more variable: MntGoldProds <dbl>
```

Veamos las estadisticas multivariadas con histogramas y box plot.


MntWines
0
500
1000
1500
MntFruits
0.00
0.02
0.04
0.06
0
500
1000
1500
MntMeatProducts
0
500
1000
1500
MntFishProducts
0.00
0.02
0.04
0
500
1000
1500
MntSweetProducts
0
500
1000
1500
MntGoldProds
0.000
0.015
0
500
1000
1500

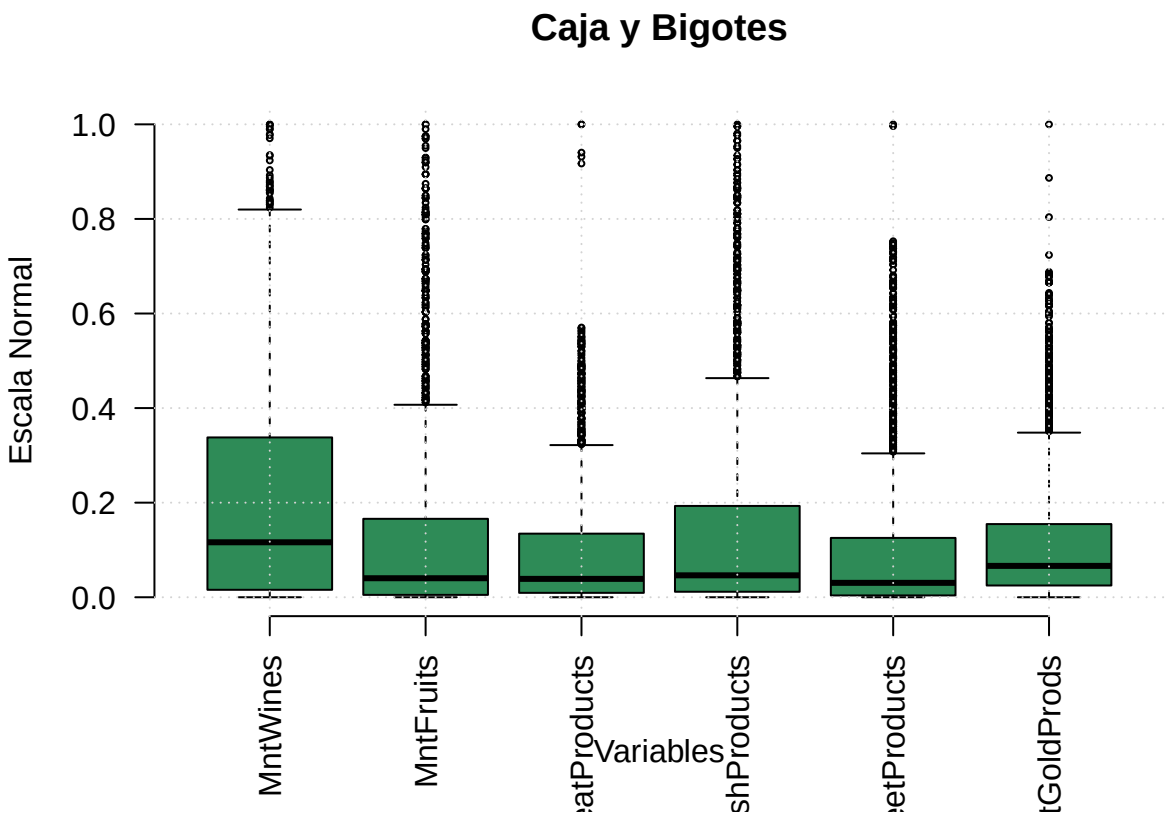

Boxplot
Escala Normal
0
500
1000
1500
MntWines
MntMeatProducts
MntSweetProducts
Variables



El paso siguiente para nuestro analisis es la normalización de nuestros datos, este paso es muy importante ya que lo obtenido en este punto servirá despues para obtener la escala de lickert que se definirá.

MntWines	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds
0.4253182	0.4422111	0.3165217	0.6640927	0.3346008	0.2430939
0.0073677	0.0050251	0.0034783	0.0077220	0.0038023	0.0165746
0.2853315	0.2462312	0.0736232	0.4285714	0.0798479	0.1160221
0.0073677	0.0201005	0.0115942	0.0386100	0.0114068	0.0138122
0.1158741	0.2160804	0.0684058	0.1776062	0.1026616	0.0414365
0.3482920	0.2110553	0.0568116	0.0000000	0.1596958	0.0386740

Con los cambios realizados a los datos veremos cómo cambió esto con unos boxplot, en estos vemos que cambió la escala que teniamos antes y tenemos valores fuera del rango del tercer cuantil, lo cual por la forma en que se ven los puntos nos dice que pueden ser valores extremos.



1 Matriz de datos con escala ordinales

Definicion de rangos para la escala de lickert

0-20 -> 1

21:40 -> 2

40:60 -> 3

61:80 -> 4

81:100 -> 5

Matriz de correlación.

```
R <- cor(datos_new, method = "spearman")
R
```

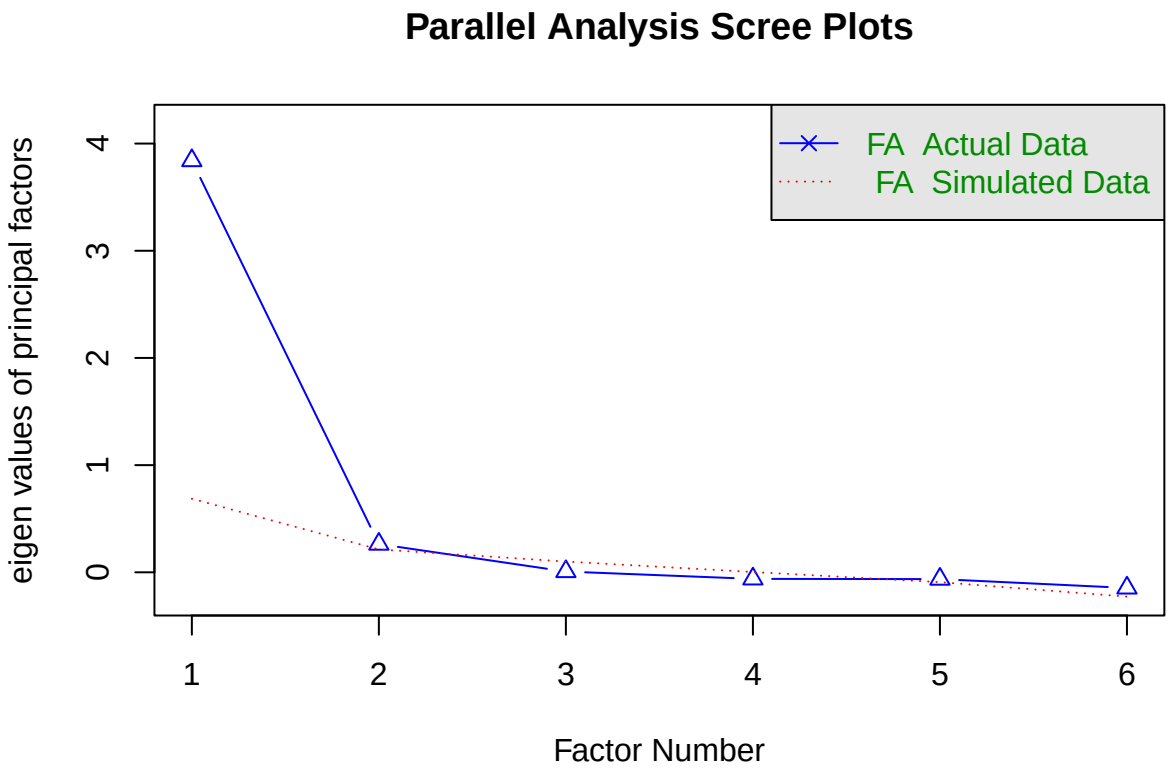
```
##                  MntWines MntFruits MntMeatProducts MntFishProducts
## MntWines         1.0000000 0.5169842       0.8190281       0.5266061
## MntFruits        0.5169842 1.0000000       0.7129579       0.7037176
## MntMeatProducts  0.8190281 0.7129579       1.0000000       0.7245835
## MntFishProducts  0.5266061 0.7037176       0.7245835       1.0000000
## MntSweetProducts 0.5222219 0.6940737       0.7058843       0.7016569
## MntGoldProds     0.5733340 0.5660507       0.6346116       0.5631088
##                  MntSweetProducts MntGoldProds
## MntWines                 0.5222219    0.5733340
## MntFruits                0.6940737    0.5660507
```

```
## MntMeatProducts          0.7058843    0.6346116
## MntFishProducts          0.7016569    0.5631088
## MntSweetProducts         1.0000000    0.5512815
## MntGoldProds             0.5512815    1.0000000
```

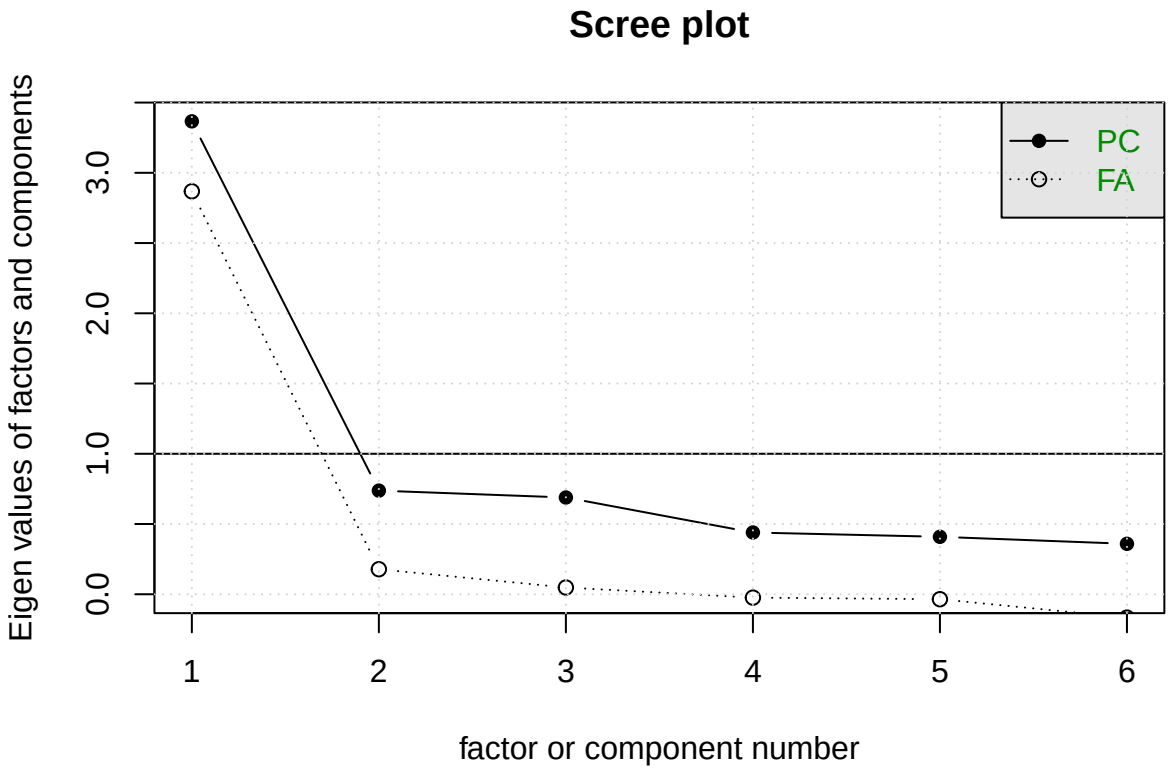
```
## [1] TRUE
```

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = R)
## Overall MSA =  0.86
## MSA for each item =
##         MntWines        MntFruits  MntMeatProducts  MntFishProducts
##             0.77             0.90             0.79             0.90
## MntSweetProducts     MntGoldProds
##             0.91             0.95
```

Del resultados de la prueba de Bartlet tenemos que rechazamos la hipótesis nula, por lo que sí tenemos correlación entre las variables, además con la prueba del KMO obtuvimos un valor de 0.86 lo que significa que tenemos un valor Bueno por lo que sí se puede hacer análisis factorial.




```
## Parallel analysis suggests that the number of factors =  2  and the number of components =  NA
```



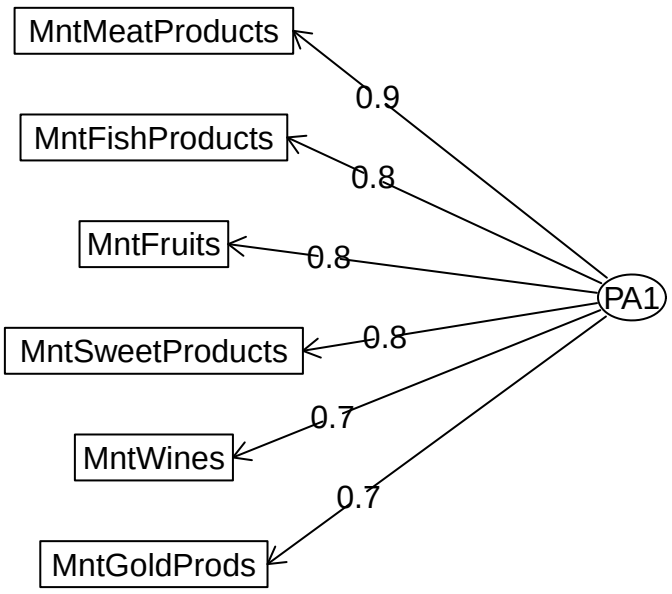
Con el screeplot lo que vemos es que solo tenemos un factor significativo por lo que se usará ese valor para el análisis.

```
## Factor Analysis using method =  pa
## Call: fa(r = R, nfactors = k, rotate = "none", fm = "pa")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                   V  PA1   h2   u2 com
## MntMeatProducts   3 0.93 0.86 0.14   1
## MntFishProducts   4 0.81 0.66 0.34   1
## MntFruits         2 0.80 0.64 0.36   1
## MntSweetProducts 5 0.80 0.63 0.37   1
## MntWines          1 0.74 0.55 0.45   1
## MntGoldProds      6 0.71 0.50 0.50   1
##
##                  PA1
## SS loadings    3.84
## Proportion Var 0.64
##
## Mean item complexity =  1
## Test of the hypothesis that 1 factor is sufficient.
##
## df null model =  15  with the objective function =  4.32 0.3
## df of  the model are 9  and the objective function was  0.46
##  0.3
## The root mean square of the residuals (RMSR) is  0.06
## The df corrected root mean square of the residuals is  0.07
##  0.3
```

```
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy
##                                                  PA1
## Correlation of (regression) scores with factors   0.97
## Multiple R square of scores with factors          0.94
## Minimum correlation of possible factor scores     0.87
```

```
##
## Loadings:
##                  PA1
## MntWines         0.743
## MntFruits        0.801
## MntMeatProducts  0.929
## MntFishProducts  0.809
## MntSweetProducts 0.795
## MntGoldProds     0.705
##
##                  PA1
## SS loadings    3.841
## Proportion Var 0.640
```

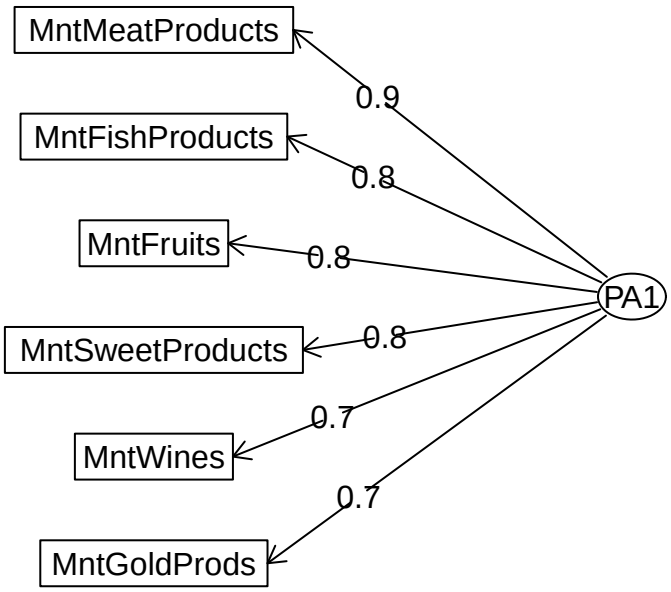
Factor Analysis




```
## Factor Analysis using method =  pa
## Call: fa(r = R, nfactors = k, rotate = "none", fm = "pa")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                  V  PA1   h2   u2 com
## MntMeatProducts  3 0.93 0.86 0.14   1
## MntFishProducts  4 0.81 0.66 0.34   1
```

```
## MntFruits        2 0.80 0.64 0.36   1
## MntSweetProducts 5 0.80 0.63 0.37   1
## MntWines         1 0.74 0.55 0.45   1
## MntGoldProds     6 0.71 0.50 0.50   1
##
##                 PA1
## SS loadings    3.84
## Proportion Var 0.64
##
## Mean item complexity =  1
## Test of the hypothesis that 1 factor is sufficient.
##
## df null model =  15  with the objective function =  4.32 0.3
## df of  the model are 9  and the objective function was  0.46
##  0.3
## The root mean square of the residuals (RMSR) is  0.06
## The df corrected root mean square of the residuals is  0.07
##  0.3
## Fit based upon off diagonal values = 0.99
## Measures of factor score adequacy
##                                                  PA1
## Correlation of (regression) scores with factors   0.97
## Multiple R square of scores with factors          0.94
## Minimum correlation of possible factor scores     0.87
##
## Loadings:
##                  PA1
## MntWines         0.743
## MntFruits        0.801
## MntMeatProducts  0.929
## MntFishProducts  0.809
## MntSweetProducts 0.795
## MntGoldProds     0.705
##
##                  PA1
## SS loadings    3.841
## Proportion Var 0.640
```

Factor Analysis



De los ajustes realizados con el factor que se escogió vemos que si una persona compra Productos como GoldProds, Wines, además de alimentos como FishProducts entre otros es un consumidor que tiene un ingreso mas alto con respecto a los que no compran este tipo de alimentos, por otro lado si tenemos que una persona solo compra cosas como MeatProducts Fruits sin tomar productos como Fish y Wines entonces es un consumidor que solo compra lo necesario o la canasta básica.