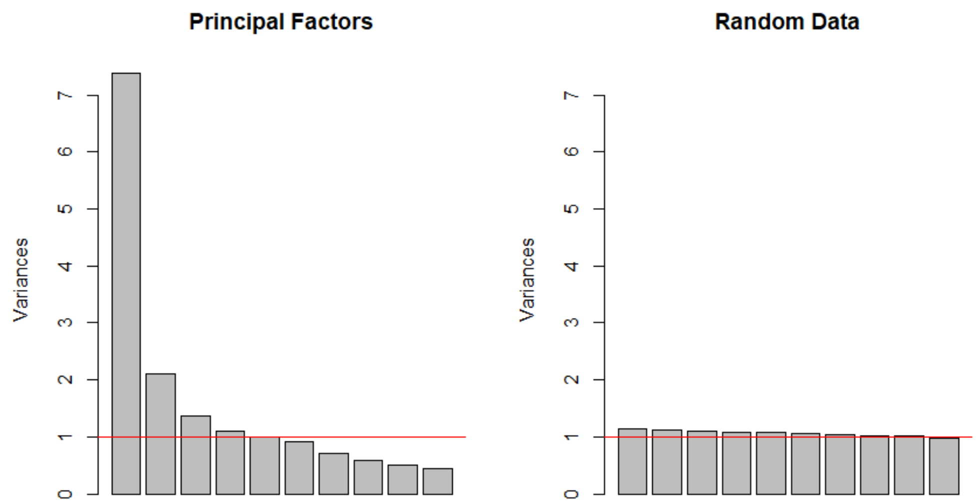
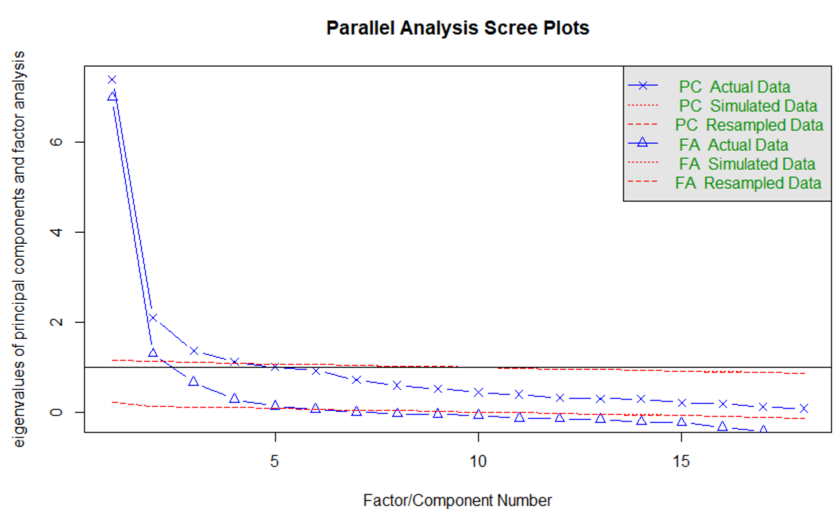
The first Eigen vectors and Eigen values were calculated from a correlation matrix using the prcomp function to try to determine how many principal factors should be used in the principal factor analysis. The first four principal factors account for 66% of all the variance captured by a total of eighteen factors, as can be seen in the Figure below.

```
Importance of components:
                          PC1    PC2     PC3     PC4     PC5     PC6     PC7     PC8     PC9    PC10    PC11    PC12
Standard deviation     2.7157 1.4484 1.16508 1.05256 0.99956 0.96126 0.84349 0.77094 0.71638 0.66091 0.62692 0.55870
Proportion of Variance 0.4097 0.1166 0.07541 0.06155 0.05551 0.05133 0.03953 0.03302 0.02851 0.02427 0.02183 0.01734
Cumulative Proportion  0.4097 0.5263 0.60170 0.66325 0.71875 0.77009 0.80961 0.84263 0.87114 0.89541 0.91725 0.93459
                         PC13    PC14    PC15    PC16    PC17    PC18
Standard deviation     0.5500 0.53464 0.45911 0.42101 0.34358 0.28814
Proportion of Variance 0.0168 0.01588 0.01171 0.00985 0.00656 0.00461
Cumulative Proportion  0.9514 0.96727 0.97898 0.98883 0.99539 1.00000
```

Thirty random matrices were generated and the average for each one of their corresponding elements was calculated. The variance explained by each one of the columns of said matrix was plotted on the graph to the right on the figure below. A red line was drawn at variance equals one, allowing a comparison between the variance explained by randomly generated data, or noise, and the variance explained by each of the principal factors calculated, on the left. This figure indicates that four, maybe five, principal factors should explain more variance than just noise would and, therefore, can be used in the analysis.



A parallel analysis using the fa.parallel function was conducted, resulting in the figure below. The “PC Actual Data” line is the one corresponding to the variance explained by the principal factors previously calculated and the “PC Simulated Data” is the equivalent to the randomly generated data. Not only can the parallel analysis be used to confirm what the figure above indicated, as far as how many principal factors to use in the principal factor analysis, but also gives us insight regarding variance explained by each principal factor calculated by common factor analysis, which will be introduced as an alternative to principal factor analysis later on.



The principal factors were re-calculated using the principal function from the psych package with, in the middle, and without, on the left, the VARIMAX rotation. As can be seen in the figure below the loadings in the middle are much easier to interpret than the ones on the left as they are not spread over as many components. A common factor analysis was calculated using the factanal function, producing the factors seen on the right, which have even more multi factor loadings than the principal factors calculated without VARIMAX. Considering that the purpose of this analysis is driving understanding from the principal factors and their loadings, a principal factor analysis using VARIMAX seems to be the best option.

```
Loadings:
                             PC1    PC2    PC3    PC4
Kidhome                    -0.660
logp1_Income                0.785
logp1_MntFishProducts       0.761
logp1_MntFruits             0.765
logp1_MntGoldProds          0.699
logp1_MntMeatProducts       0.931
logp1_MntSweetProducts      0.755
logp1_MntWines              0.868
logp1_NumCatalogPurchases   0.892
logp1_NumStorePurchases     0.817
logp1_NumWebPurchases       0.718  0.406
logp1_NumWebVisitsMonth    -0.583  0.432  0.434
Teenhome                           0.797
logp1_NumDealsPurchases            0.763  0.450
Year_Birth                        -0.438  0.607
Recency                                          0.688
Tot_Accepted                                    -0.683
Complain

                 PC1   PC2   PC3   PC4
SS loadings    7.375 2.098 1.357 1.108
Proportion Var 0.410 0.117 0.075 0.062
Cumulative Var 0.410 0.526 0.602 0.663
```

```
Loadings:
                             RC1    RC2    RC3    RC4
Kidhome                    -0.550
logp1_Income                0.682
logp1_MntFishProducts       0.791
logp1_MntFruits             0.793
logp1_MntGoldProds          0.749
logp1_MntMeatProducts       0.915
logp1_MntSweetProducts      0.786
logp1_MntWines              0.816
logp1_NumCatalogPurchases   0.844
logp1_NumStorePurchases     0.803
logp1_NumWebPurchases       0.761
logp1_NumDealsPurchases            0.857
logp1_NumWebVisitsMonth    -0.457  0.716
Year_Birth                               -0.765
Teenhome                           0.417  0.729
Recency                                         -0.677
Tot_Accepted                                     0.732
Complain

                 RC1   RC2   RC3   RC4
SS loadings    6.939 1.957 1.805 1.237
Proportion Var 0.386 0.109 0.100 0.069
Cumulative Var 0.386 0.494 0.595 0.663
```

```
Loadings:
                          Factor1 Factor2 Factor3 Factor4
logp1_Income               0.693
logp1_MntMeatProducts      0.704   0.621
logp1_MntWines             0.896
logp1_NumCatalogPurchases  0.674   0.547
logp1_NumStorePurchases    0.656   0.460
logp1_NumWebPurchases      0.628   0.410           0.406
logp1_MntFishProducts              0.752
logp1_MntFruits                    0.750
logp1_MntGoldProds         0.404   0.566
logp1_MntSweetProducts             0.746
Teenhome                                   0.928
logp1_NumDealsPurchases                            0.646
logp1_NumWebVisitsMonth                            0.720
Year_Birth                                -0.414
Kidhome                   -0.444
Recency
Complain
Tot_Accepted

               Factor1 Factor2 Factor3 Factor4
SS loadings      4.032   3.568   1.409   1.403
Proportion Var   0.224   0.198   0.078   0.078
Cumulative Var   0.224   0.422   0.500   0.578
```

Correlation matrix and principal factor analysis seem to indicate that the more kids at home a customer has the less purchases they are likely to make, especially meat and wine products. These customers are also much less likely to make purchases either at the store or from a catalog, preferring to visit the website and make deal purchases. It can also be inferred that customers who visit the website are usually looking for deals whether they have kids or not. On the other hand, even though customers who buy at the store and from the catalog might also buy online it is clear that the higher the income of a customer, the less likely they are to check the website and the more likely they are to buy at the store or from a catalog. It stems from this analysis that it could be beneficial to have mostly deals on the website as well as a section dedicated to families with kids.