I used the following code to recode my qualitative responses into quantitative numbers.

data(package = "ltm")

library(tidyverse)

df<- Environment

x<-df %>% mutate_all(~ str_replace(., "^$", NA_character_)) %>% mutate_all(.funs = ~ as.integer(recode(.x = .,"not very concerned"=0,"slightly concerned"=1, "very concerned"=2)))

library("writexl"

write_xlsx(df,"D:\\df.xlsx")

ddf<-df %>% mutate_all(~ str_replace(.,"^$", NA_character_)) %>%mutate_all(.funs = ~ as.integer(recode(.x = .,"not very concerned"=0,"slightly concerned"=1, "very concerned"=2)))

First i find the correlation matrix of the data. I scan the matrix for any correlation below .3 or above.9. Most variables seem close enough to .3

raqMatrix<-cor(ddf) round(raqMatrix, 2)

```
             LeadPetrol RiverSea RadioWaste
LeadPetrol         1.00     0.41       0.29
RiverSea           0.41     1.00       0.42
RadioWaste         0.29     0.42       1.00
AirPollution       0.48     0.59       0.54
Chemicals          0.31     0.42       0.64
Nuclear            0.29     0.34       0.54
             AirPollution Chemicals Nuclear
LeadPetrol           0.48      0.31    0.29
RiverSea             0.59      0.42    0.34
RadioWaste           0.54      0.64    0.54
AirPollution         1.00      0.51    0.41
Chemicals            0.51      1.00    0.52
Nuclear              0.41      0.52    1.00
> 
```

I then do a Barlett test. For factor analysis to work we need some relationships between variables and if the R-matrix were an identity matrix then all correlation coefficients would be zero. Therefore, we want this test to be significant (i.e., have a significance value less than .05). A significant test tells us that the R-matrix is not an identity matrix; therefore, there are some relationships between the variables we hope to include in the analysis. The result is significant.

```
> cortest.bartlett(raqMatrix)
$chisq
[1] 206.5277

$p.value
[1] 9.989891e-36

$df
[1] 15
```

Then i do a KMO test. F or these data the overall value is .83, so we should be confident that the sample size and the data are adequate for factor analysis.KMO can be calculated for multiple and individual variables. The value of KMO should be above the bare minimum of .5 for all variables (and preferably higher) as well as overall. The KMO values for individual variables are produced by the kmo() function too. For these data all values are well above .5, which is good news. If you find any variables with values below .5 then you should consider excluding them from the analysis (or run the analysis with and without that variable and note the difference). Removal of a variable affects the KMO statistics, so if you do remove a variable be sure to rerun the kmo() function on the new data.

Multicollinearity can be detected by looking at the determinant of the R-matrix. We then check the determinant. This value is greater than the necessary value of 0.00001. As such, our determinant does not seem problematic.

DSUR.noof::kmo(raqData)

```
Error in kmo(raqData) : could not f
> DSUR.noof::kmo(raqData)
$overall
[1] 0.8357451

$report
[1] "The KMO test yields a degree c
```

Type '/' for commands

```
$individual
  LeadPetrol    RiverSea  RadioWaste AirPollution   Chemicals
   0.8527141   0.8447277   0.8110176    0.8203403   0.8348433
     Nuclear
   0.8733315
```

```
> det(cor(raqData))
[1] 0.1167639
> |
```

h2 is the communalities (which are sometimes called h2). These communalities are all equal to 1 because we have extracted 23 items, the same as the number of variables: we've explained all of the variance in every variable. When we extract fewer factors (or components) we'll have lower communalities. Next to the communality column is the uniqueness column, labelled u2. This is the amount of unique variance for each variable, and it's 1 minus the communality; because all of the communalities are 1, all of the uniquenesses are 0.8

The next thing to look at after the factor loading matrix is the eigenvalues.

The eigenvalues associated with each factor represent the variance explained by that particular linear component. R calls these SS loadings (sums of squared loadings), because they are the sum of the squared loadings.

```
ix
               PC1   PC2   PC3   PC4   PC5   PC6 h2       u2 com
LeadPetrol    0.60  0.59  0.50 -0.16  0.08  0.09  1 -2.2e-16 3.2
RiverSea      0.72  0.36 -0.42  0.31  0.25  0.12  1  5.6e-16 3.1
RadioWaste    0.79 -0.35 -0.08 -0.21 -0.16  0.41  1  8.9e-16 2.3
AirPollution  0.81  0.24 -0.19 -0.06 -0.42 -0.27  1 -6.7e-16 2.1
Chemicals     0.78 -0.33 -0.05 -0.31  0.34 -0.26  1  5.6e-16 2.4
Nuclear       0.70 -0.38  0.37  0.47 -0.04 -0.07  1  5.6e-16 3.0

                       PC1  PC2  PC3  PC4  PC5  PC6
SS loadings           3.26 0.91 0.61 0.49 0.39 0.34
Proportion Var        0.54 0.15 0.10 0.08 0.06 0.06
Cumulative Var        0.54 0.70 0.80 0.88 0.94 1.00
Proportion Explained  0.54 0.15 0.10 0.08 0.06 0.06
Cumulative Proportion 0.54 0.70 0.80 0.88 0.94 1.00
```

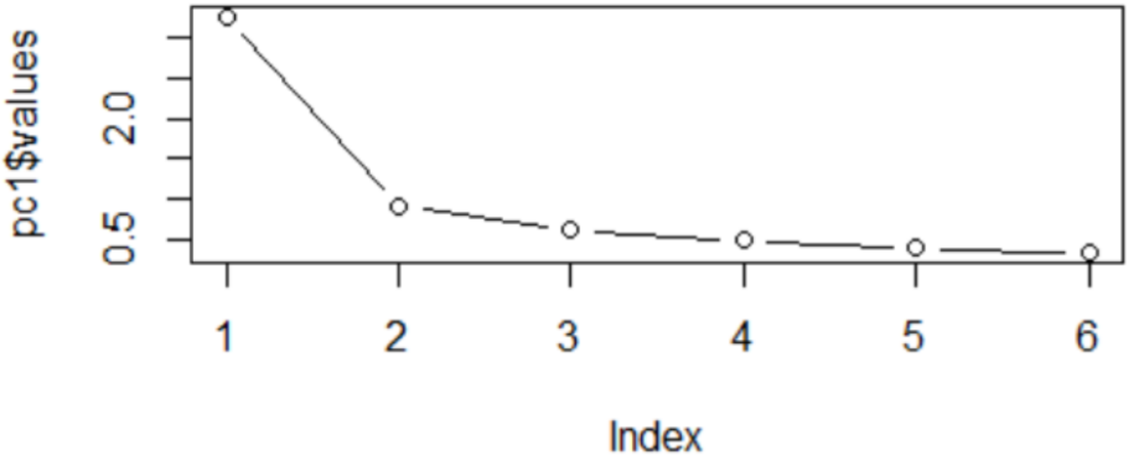
R also displays the eigenvalues in terms of the proportion of variance explained. Factor 1 explains 3.26 units of variance out of a possible 6(the number of factors) so as a proportion this is 3.26/6= .54 ; this is the value that R reports. We can convert these proportions to percentages by multiplying by 100; so, factor 1 explains 54% of the total variance.

Using Kaiser's criterian we extract only 1 factor. By Jolliffe's criterion we retain 2 factors. Scree plot suggests 1 factor.

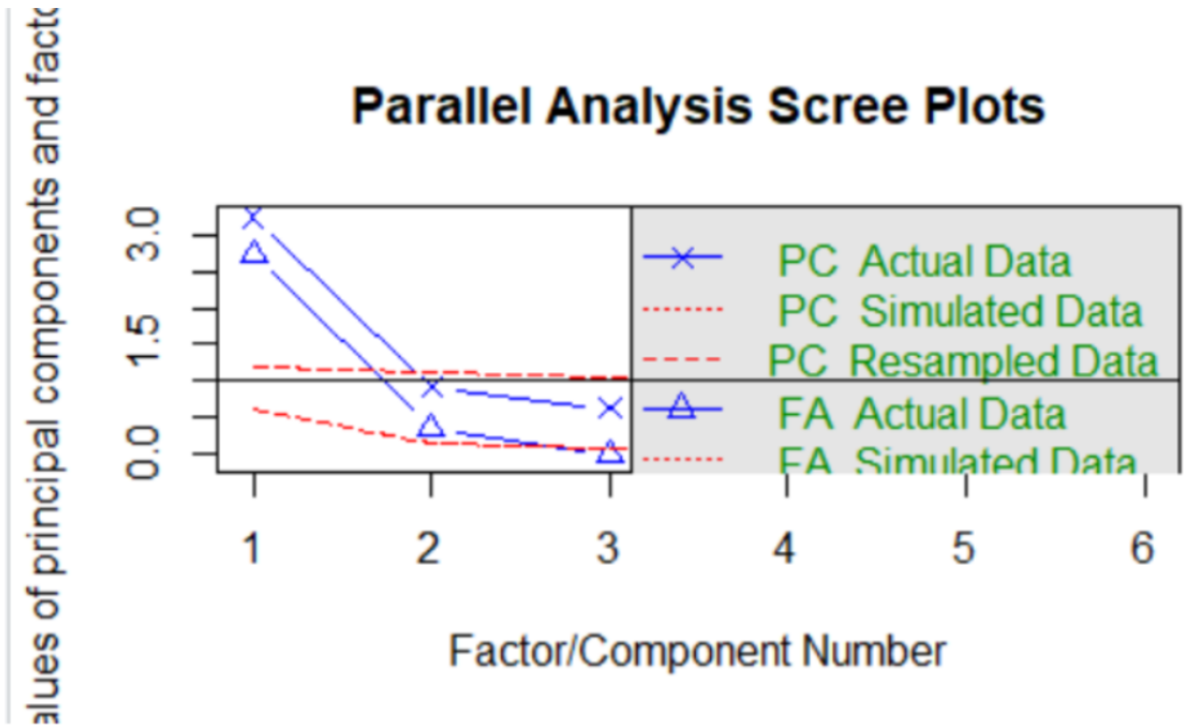
```
Fit based upon off diagonal values = 1
> pc1$values
[1] 3.2621004 0.9138358 0.6093809 0.4859021 0.3866825 0.3420984
> |
```

R also displays the eigenvalues in terms of the proportion of variance explained. Factor 1 explains 3.26 units of variance out of a possible 6(the number of factors) so as a proportion this is 3.26/6= .54 ; this is the value that R reports. We can convert these proportions to percentages by multiplying by 100; so, factor 1 explains 54% of the total variance.

Using Kaiser's criterian we extract only 1 factor. By Jolliffe's criterion we retain 2 factors. Scree plot suggests 1 factor.



Parallel analysis suggests that the number of factors = 2 and the number of components = 1



I will extract 2 factors.

pc2 <- principal(raqData, nfactors = 1, rotate = "none")

```
1X
               PC1   PC2   h2   u2 com
LeadPetrol    0.60  0.59 0.71 0.29 2.0
RiverSea      0.72  0.36 0.65 0.35 1.5
RadioWaste    0.79 -0.35 0.75 0.25 1.4
AirPollution  0.81  0.24 0.71 0.29 1.2
Chemicals     0.78 -0.33 0.72 0.28 1.3
Nuclear       0.70 -0.38 0.64 0.36 1.5

                        PC1  PC2
SS loadings            3.26 0.91
Proportion Var         0.54 0.15
Cumulative Var         0.54 0.70
Proportion Explained   0.78 0.22
Cumulative Proportion  0.78 1.00

Mean item complexity =  1.5
```

The communalities (the h2 column) and uniquenesses (the u2 column) are changed. The communality is the proportion of common variance within a variable . Principal components analysis works on the initial assumption that all variance is common; therefore, before extraction the communalities are all 1. In effect, all of the variance associated with a variable is assumed to be common variance. Once factors have been extracted, we have a better idea of how much variance is, in reality, common. The communalities in the output reflect this common variance. So, for example, we can say that 43% of the variance associated with question 1 is common, or shared, variance.

Another way to look at these communalities is in terms of the proportion of variance explained by the underlying factors. Before extraction, there were as many factors as there are variables, so all variance is explained by the factors and communalities are all 1. However, after extraction some of the factors are discarded and so some information is lost. The retained factors cannot explain all of the variance present in the data, but they can explain some. The amount of variance in each variable that can be explained by the retained factors is represented by the communalities after extraction

Fit based upon off diagonal values = 0.95 Values over 0.95 are often considered indicators of good fit, and as our value is 0.96, this indicates that four factors are sufficient.

```
                         PC1  PC2
SS loadings           3.26 0.91
Proportion Var        0.54 0.15
Cumulative Var        0.54 0.70
Proportion Explained  0.78 0.22
Cumulative Proportion 0.78 1.00

Mean item complexity =  1.5
Test of the hypothesis that 2 components are sufficient.

The root mean square of the residuals (RMSR) is  0.1
 with the empirical chi square  90.28  with prob <  1.1e-18

Fit based upon off diagonal values = 0.95
```

I did an orthogonal rotation using varimax on 2 factors. I get the following results. Item 3 5 6 load on factor 1 and item 124 load on factor 2.

1X

	item	RC1	RC2	h2	u2	com
RadioWaste	3	0.83		0.75	0.25	1.2
Chemicals	5	0.81		0.72	0.28	1.2
Nuclear	6	0.78		0.64	0.36	1.1
LeadPetrol	1		0.84	0.71	0.29	1.0
RiverSea	2	0.31	0.74	0.65	0.35	1.3
AirPollution	4	0.46	0.71	0.71	0.29	1.7

Using oblimin oblique rotation i get the following results. Now the variables load more highly and we can also see correlation between the factors. If the constructs were independent then we would expect oblique rotation to provide an identical solution to an orthogonal rotation and the component correlation matrix should be an identity matrix

```
ix
              item  TC1   TC2   h2   u2 com
RadioWaste       3  0.85        0.75 0.25 1.0
Chemicals        5  0.82        0.72 0.28 1.0
Nuclear          6  0.82        0.64 0.36 1.0
LeadPetrol       1        0.90  0.71 0.29 1.1
RiverSea         2        0.73  0.65 0.35 1.1
AirPollution     4  0.32  0.64  0.71 0.29 1.5
```

```
 With component correlations of
     TC1  TC2
TC1 1.00 0.48
TC2 0.48 1.00
```

When an oblique rotation is conducted the factor matrix is split into two matrices: the pattern matrix and the structure matrix . For orthogonal rotation these matrices are the same. The pattern matrix contains the **factor loadings** and is comparable to the factor matrix that we interpreted for the orthogonal rotation. The structure matrix **takes into account the relationship between factors** (in fact it is a product of the pattern matrix and the matrix containing the correlation coefficients between factors). **Most researchers interpret the pattern matrix, because it is usually simpler; however, there are situations in which values in the pattern matrix are suppressed because of relationships between the factors. Therefore, the structure matrix is a useful double-check.**

The following is the structure matrix. Several variables load highly on both factors. This might be expected as some constructs might be highly correlated. If this is justified or not should be checked with our theory.

```
> pc4$loadings %*% pc4$Phi
                  TC1       TC2
LeadPetrol   0.2884369 0.8306511
RiverSea     0.4904084 0.7950838
RadioWaste   0.8666310 0.4347094
AirPollution 0.6280087 0.7942613
Chemicals    0.8471178 0.4434516
Nuclear      0.7997503 0.3447617
```

The following are the factor scores for participants

```
Fit based upon off diagonal values = 0.95
> pc5$scores
              TC1          TC2
  [1,]  0.74335140  0.740744231
  [2,]  0.74335140  0.740744231
  [3,]  0.74335140  0.740744231
  [4,]  0.74335140  0.740744231
  [5,]  0.74335140  0.740744231
  [6,]  0.74335140  0.740744231
  [7,]  0.74335140  0.740744231
  [8,]  0.74335140  0.740744231
  [9,]  0.74335140  0.740744231
 [10,]  0.74335140  0.740744231
 [11,]  0.74335140  0.740744231
 [12,]  0.74335140  0.740744231
 [13,]  0.74335140  0.740744231
 [14,]  0.74335140  0.740744231
```