Hw3

Frank Jiang

3/3/2020

Question 1: a)

```
library(smacof)
```

```
## Warning: package 'smacof' was built under R version 3.6.2
```

```
## Loading required package: plotrix
```

```
##
## Attaching package: 'smacof'
```

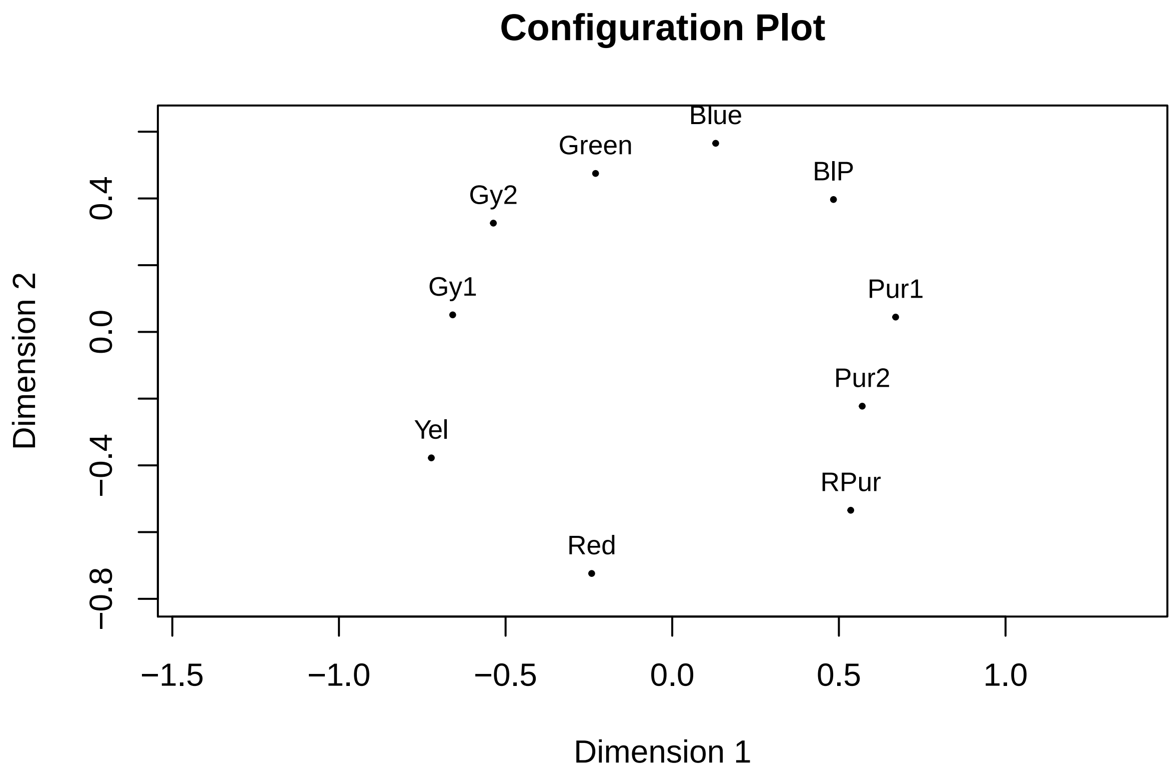
```
## The following object is masked from 'package:base':
##
##     transform
```

```
data(helm)
```

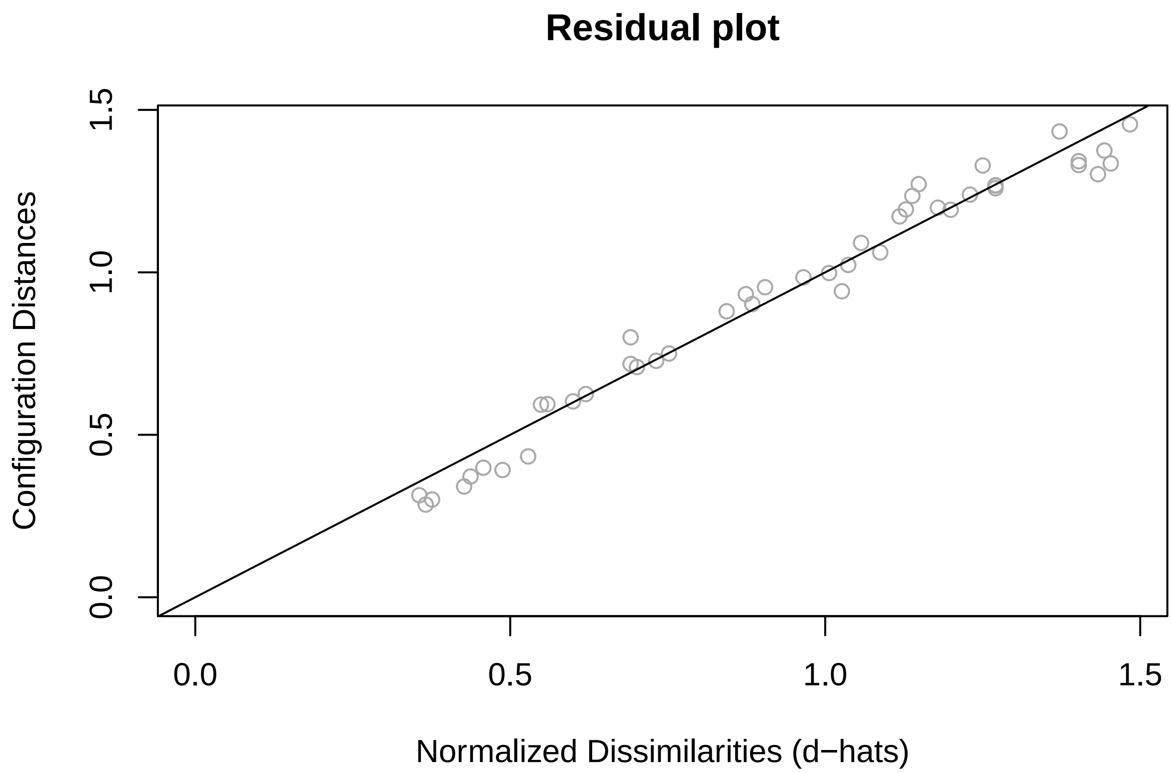
Color wheel is an arrangement of hues aronud a circle, which shows the relationship between primary colors, secondary colors, tertiary colors. For participants with normal color vision, they should percieve different color with great contrast. Thus, if two colors on a color wheel are far from each other, we can expect their perception are with a considerable gap. With our dataset collecting the infomration on people with normal color vision's response to different colors, we can expect the configuration plot to look like the color wheel with a circle shape.

b)

```
N1_dist<- helm$N1
#Ratio MDS
N1.mds.ratio<- mds(N1_dist,type = "ratio")
#configuration plot
plot(N1.mds.ratio)
```




```
#resplot
plot(N1.mds.ratio, plot.type ="resplot")
```



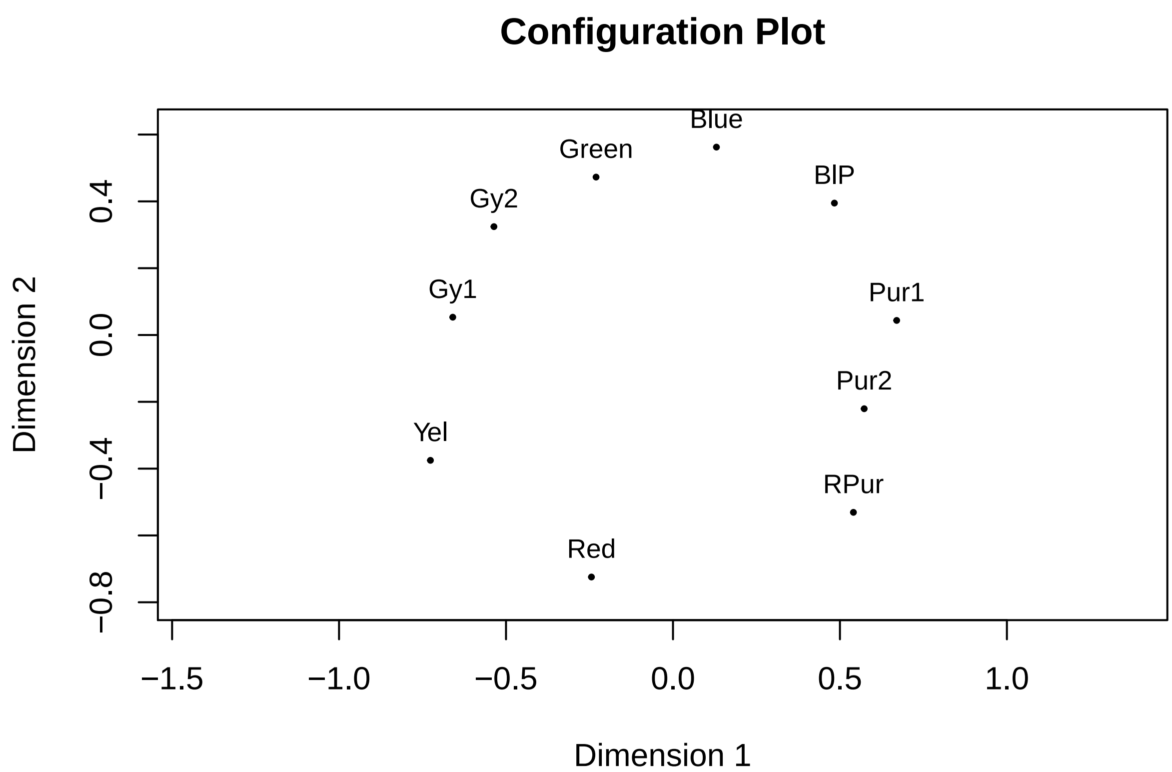

```r
#stress
N1.mds.ratio$stress
```

```
## [1] 0.06117711
```

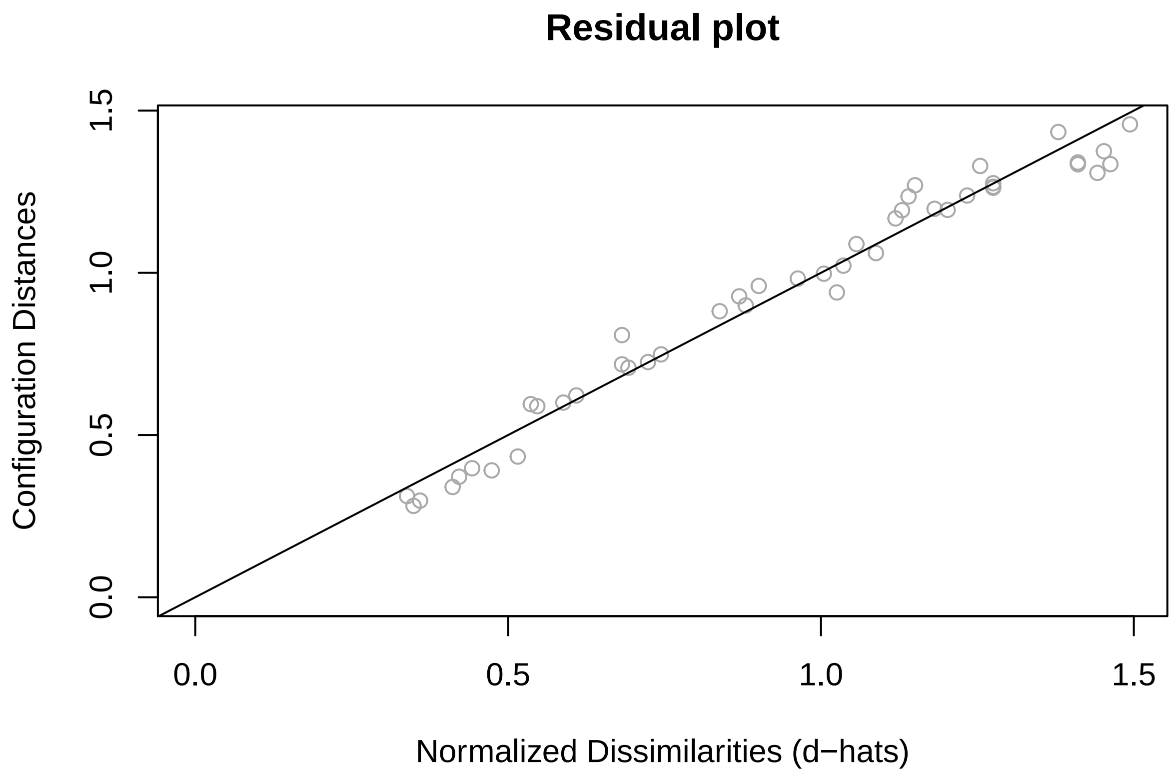
```r
#stress per point
summary(N1.mds.ratio)
```

```
##
## Configurations:
##            D1      D2
## RPur   0.5356 -0.5345
## Red   -0.2417 -0.7240
## Yel   -0.7228 -0.3775
## Gy1   -0.6587  0.0511
## Gy2   -0.5367  0.3260
## Green -0.2301  0.4749
## Blue   0.1303  0.5653
## BlP    0.4838  0.3968
## Pur1   0.6702  0.0445
## Pur2   0.5701 -0.2225
##
##
## Stress per point (in %):
##  RPur   Red   Yel   Gy1   Gy2 Green  Blue   BlP  Pur1  Pur2
##  9.26 10.94 14.32 11.61  8.49  4.14 12.72 11.63  5.48 11.41
```

```
#Interval MDS
N1.mds.interval<- mds(N1_dist,type = "interval")
#configuration plot
plot(N1.mds.interval)
```




```
#resplot
plot(N1.mds.interval, plot.type = "resplot")
```



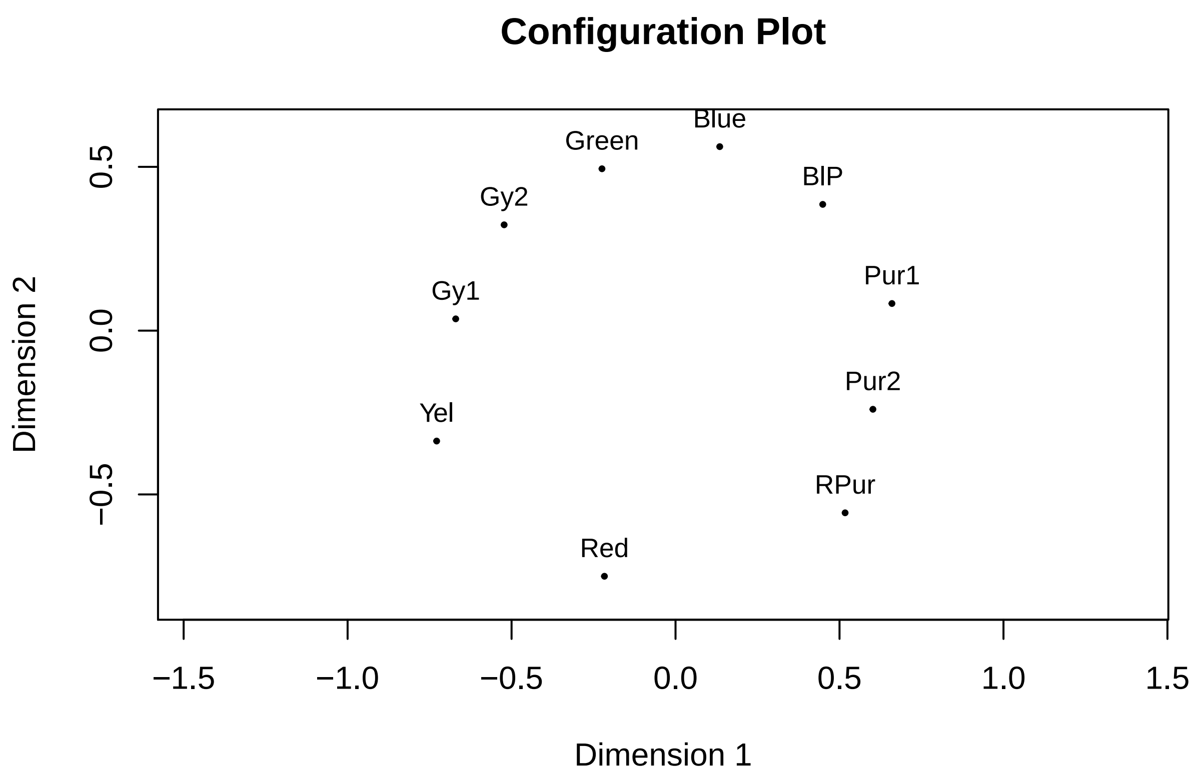

```
#stress
N1.mds.interval$stress
```

```
## [1] 0.06061565
```

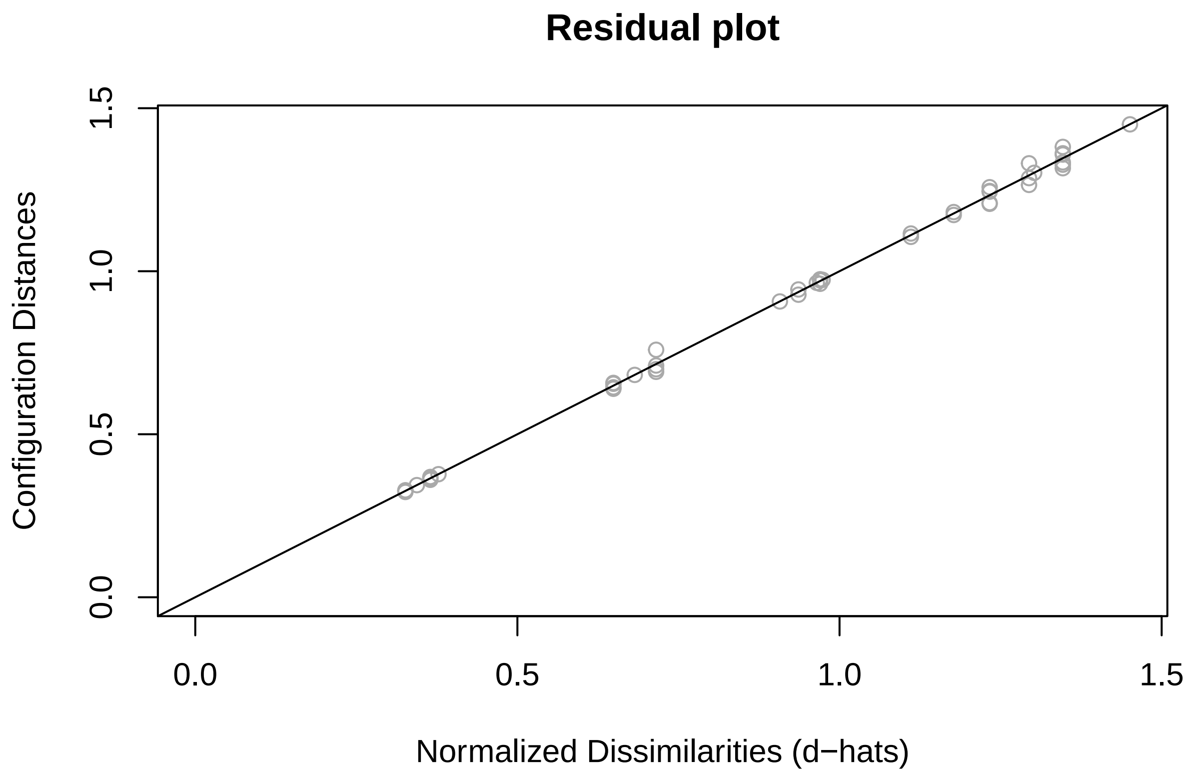
```
#stress per point
summary(N1.mds.interval)
```

```
##
## Configurations:
##            D1      D2
## RPur   0.5404 -0.5309
## Red   -0.2442 -0.7245
## Yel   -0.7263 -0.3752
## Gy1   -0.6594  0.0533
## Gy2   -0.5361  0.3243
## Green -0.2303  0.4727
## Blue   0.1301  0.5624
## BlP    0.4834  0.3948
## Pur1   0.6699  0.0436
## Pur2   0.5726 -0.2207
##
##
## Stress per point (in %):
##  RPur   Red   Yel   Gy1   Gy2 Green  Blue   BlP  Pur1  Pur2
## 10.89 14.32 14.22 10.73  7.70  3.20 11.80 11.24  4.66 11.25
```

```r
#Ordinal MDS
N1.mds.ordinal<- mds(N1_dist,type ="ordinal")
#configuration plot
plot(N1.mds.ordinal)
```




```r
#resplot
plot(N1.mds.ordinal, plot.type = "resplot")
```




```r
#stress
N1.mds.ordinal$stress
```

```
## [1] 0.01554193
```

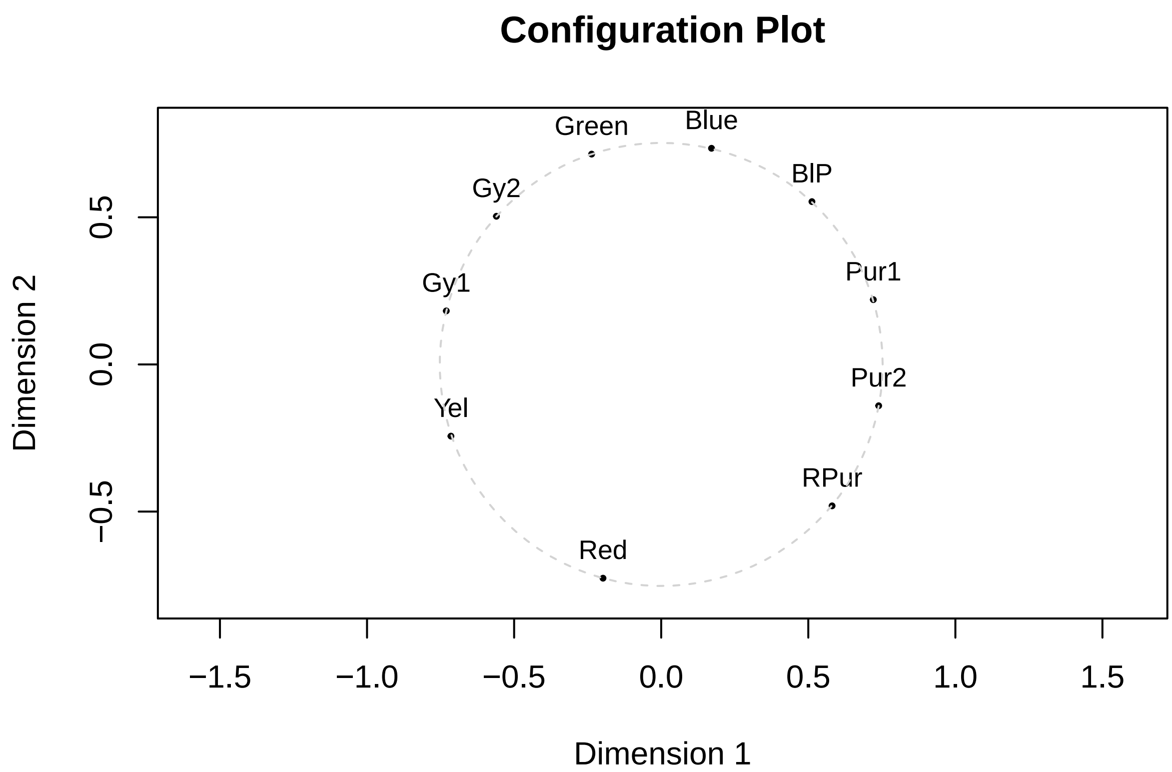
```r
#stress per point
summary(N1.mds.ordinal)
```

```
##
## Configurations:
##            D1      D2
## RPur   0.5171 -0.5562
## Red   -0.2168 -0.7498
## Yel   -0.7285 -0.3370
## Gy1   -0.6703  0.0359
## Gy2   -0.5227  0.3231
## Green -0.2244  0.4942
## Blue   0.1347  0.5617
## BlP    0.4489  0.3854
## Pur1   0.6599  0.0828
## Pur2   0.6019 -0.2401
##
##
## Stress per point (in %):
##  RPur   Red   Yel   Gy1   Gy2 Green  Blue   BlP  Pur1  Pur2
## 16.46 17.29 14.02  8.97 10.98  1.60  3.45 10.53 12.56  4.13
```

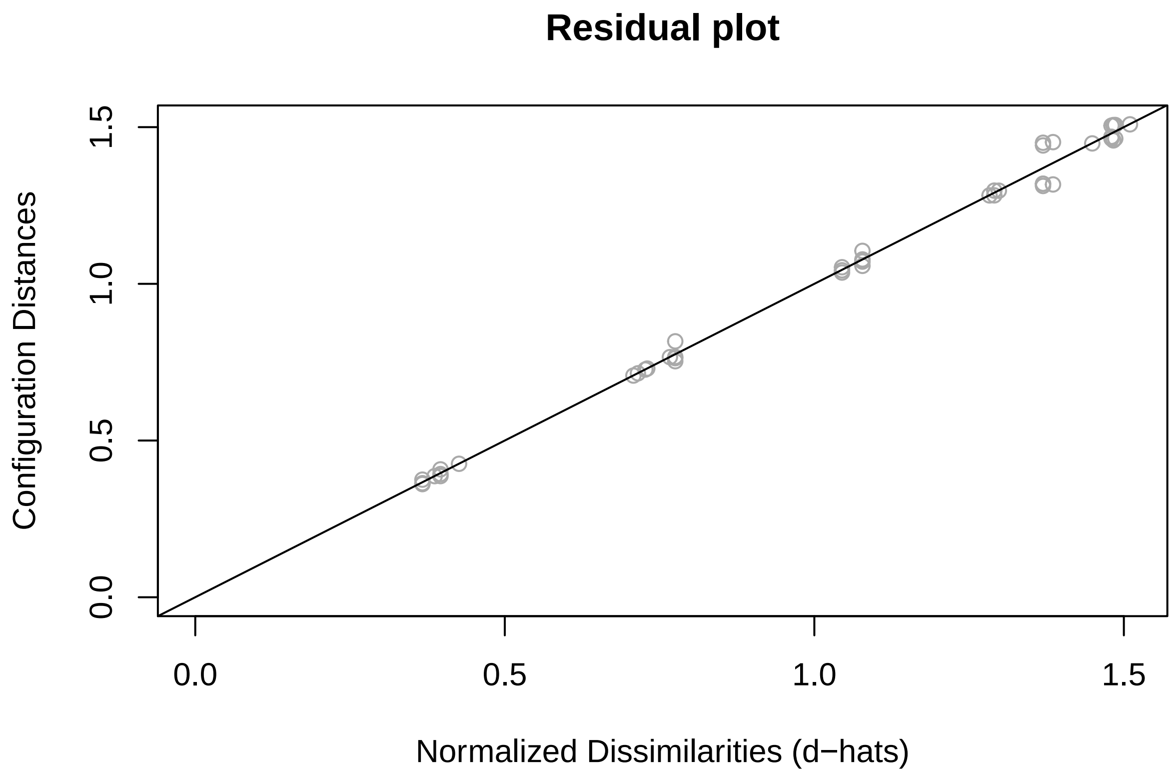
According to the MDS examination, we can tell that the configuration appear to comform to the color wheel hypothesis.

c)

```
#Spherically-constrained ordinal analysis
N1.Sphere.ordinal<- smacofSphere(N1_dist, type = "ordinal")
#configuration plot
plot(N1.Sphere.ordinal)
```




```
#resplot
plot(N1.Sphere.ordinal, plot.type = "resplot")
```




```r
#stress
N1.Sphere.ordinal$stress
```

```
## [1] 0.02874312
```

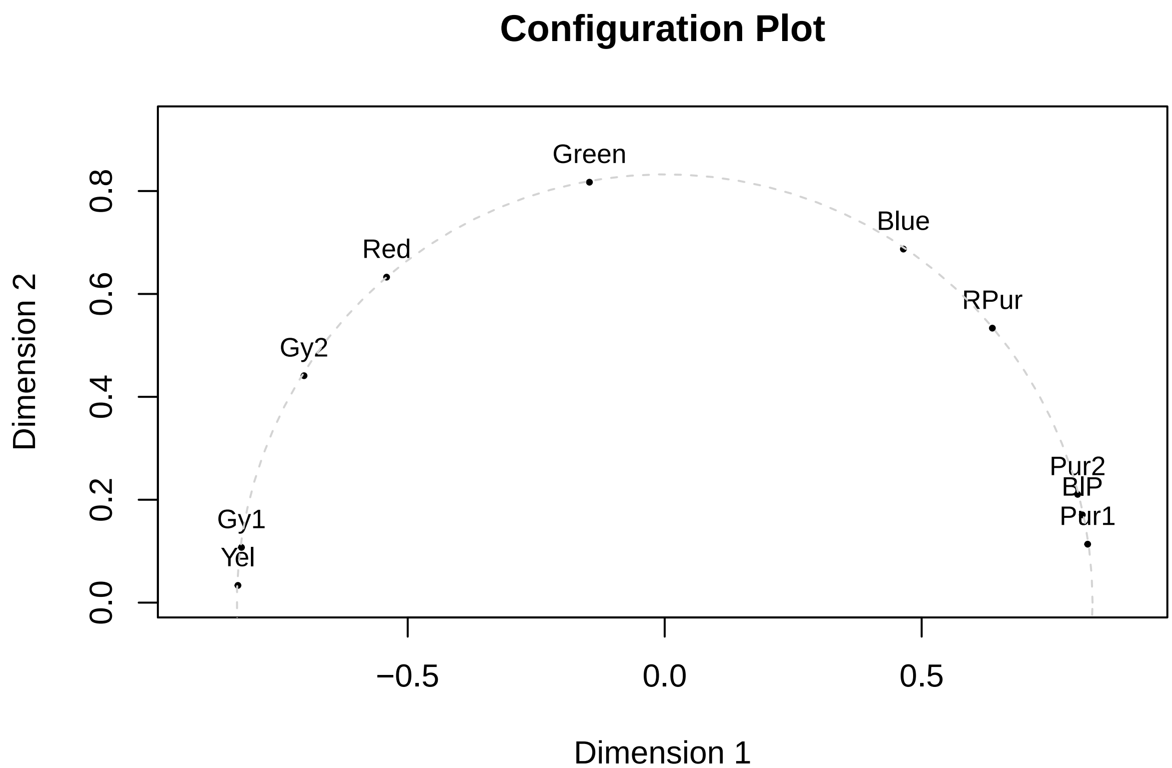
```r
#stress per point
summary(N1.Sphere.ordinal)
```

```
##
## Configurations:
##            D1      D2
## RPur   0.5809 -0.4805
## Red   -0.1978 -0.7264
## Yel   -0.7147 -0.2436
## Gy1   -0.7306  0.1819
## Gy2   -0.5602  0.5035
## Green -0.2365  0.7147
## Blue   0.1710  0.7345
## BlP    0.5124  0.5532
## Pur1   0.7211  0.2201
## Pur2   0.7393 -0.1404
```

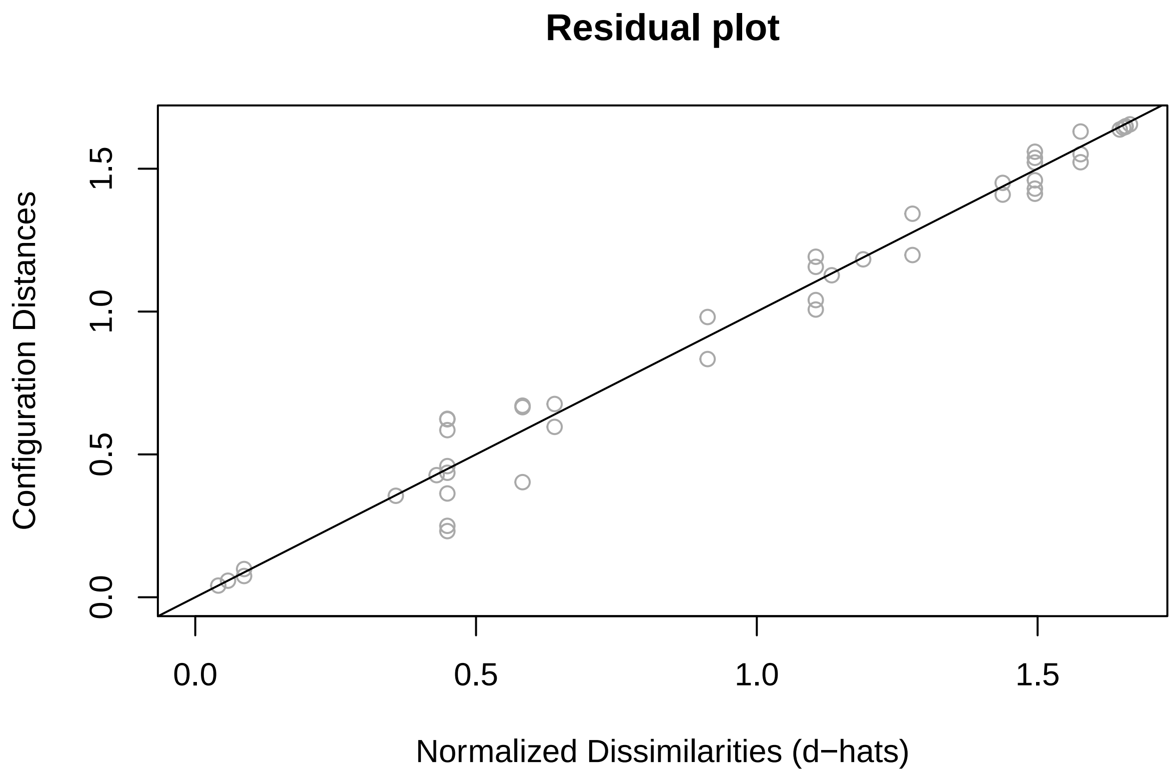
The data comform really well to the hypothesis as the configuration plot forms a color-circle. There does not exist a large loss of fitted data comparing to allowing the configuration to be free. Therefore, we can conclude that the color circle hypothesis is true. The spherical fit differs with the ordinal fit at the small amount of loss fitted at the end of the residual plot. When the normalized dissmilaries approaching 1.5, the fitted data seems to be off the fitted line comparing to the ordinal analysis.

d)

```r
#Spherically-constrained ordinal analysis with CD4
CD4_dist<- helm$CD4
CD4.Sphere.ordinal<- smacofSphere(CD4_dist, type = "ordinal")
#configuration plot
plot(CD4.Sphere.ordinal)
```




```r
#resplot
plot(CD4.Sphere.ordinal, plot.type = "resplot")
```




```r
#stress
CD4.Sphere.ordinal$stress
```

```
## [1] 0.0813597
```

```r
#stress per point
summary(CD4.Sphere.ordinal)
```

```
##
## Configurations:
##            D1     D2
## RPur   0.6376 0.5335
## Red   -0.5411 0.6326
## Yel   -0.8303 0.0335
## Gy1   -0.8234 0.1074
## Gy2   -0.7017 0.4411
## Green -0.1464 0.8172
## Blue   0.4645 0.6872
## BlP    0.8125 0.1707
## Pur1   0.8230 0.1136
## Pur2   0.8035 0.2105
```

The figure agrees with the color wheel hypothesis, as the configuration graph does not form the shape of circle since the participant has abnormal color vision. As we can observe in the graph, there is not a large enough amount of dissimilarity between different colors such as grey, yellow; The configuration graph of CD4 form a larger sphere which is different from the full circle of N1. In conclusion, grey1, and yellow; navy and purple seems to be similar for the participant.

e) Yes, I believe that using MDS to analyze color comparison data can lead to a development of a test for color-deficient vision. First of all, MDS provide a analysis on the color comparison data that are collected from the tester. If the configuration plot does not form to be a circle, then there might be a color-deficient according to the color circle hypothesis.Then, we can further analyze what color are similar to the participant by observing the dissimilarity between two different colors.

Question 2:

a)

```
library(readxl)
library(vcd)
```

```
## Loading required package: grid
```

```
VIJ_data<- read_excel("vIJ.xlsx")
#data cleaning
VIJ_collab<- c("Avoidant","Secure","Resistant")
VIJ_rowlab<- c("GE","UK","NL","SW","IS","JP","CH","US")
VIJ_tab<- as.table(matrix(
                 c(48,  77,  11,
                   16,  54,   2,
                   66, 169,  16,
                   11,  38,   2,
                    8,  76,  34,
                    5,  65,  26,
                    9,  18,   9,
                  260, 797, 173),
                  byrow=T, ncol=3))
rownames(VIJ_tab)<- VIJ_rowlab
colnames(VIJ_tab)<- VIJ_collab
#pearson chi-square analysis
chisq.test(VIJ_tab)
```

```
## Warning in chisq.test(VIJ_tab): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  VIJ_tab
## X-squared = 102.42, df = 14, p-value = 1.626e-15
```

```
chisq.test(VIJ_tab)$res
```

```
## Warning in chisq.test(VIJ_tab): Chi-squared approximation may be incorrect
```

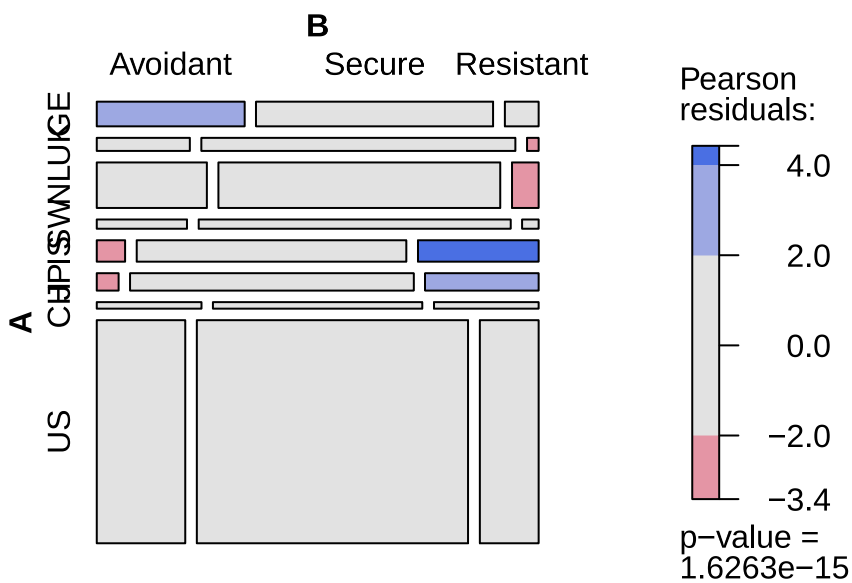
```
##       Avoidant      Secure   Resistant
## GE  3.55079811 -1.21589094 -1.77276290
## UK  0.17777600  1.04962238 -2.50646212
## NL  1.73140181  0.45297175 -3.14137746
## SW  0.04838129  0.83997642 -1.88896752
## IS -3.41086511 -0.08329749  4.42709017
## JP -3.41044839  0.32602380  3.53542279
## CH  0.48720395 -1.11796643  1.82750988
## US -0.08981494 -0.09932660  0.32804652
```

```
chisq.test(VIJ_tab)$std
```

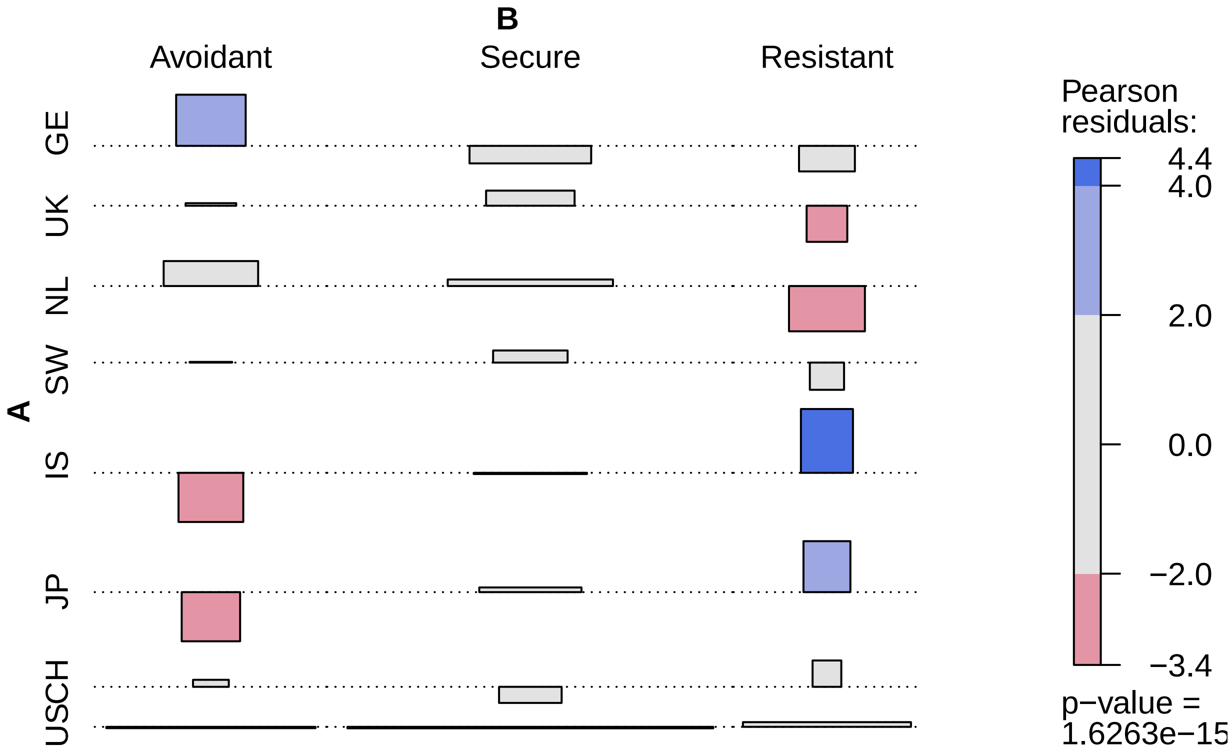
```
## Warning in chisq.test(VIJ_tab): Chi-squared approximation may be incorrect
```

```
##        Avoidant      Secure   Resistant
## GE  4.14562194 -2.13004224 -1.97725971
## UK  0.20406447  1.80782854 -2.74855715
## NL  2.08721244  0.81935062 -3.61774758
## SW  0.05523408  1.43888685 -2.06017216
## IS -3.96305592 -0.14522035  4.91397988
## JP -3.93949063  0.56507719  3.90138939
## CH  0.55407314 -1.90772135  1.98547939
## US -0.16377980 -0.27177358  0.57147412
```

```
mosaic(VIJ_tab,shade=TRUE)
```




```
assoc(VIJ_tab,shade=TRUE)
```



According to the pearson chi-square test, our p-value is 1.626*e^-15 which is lower than the signifance level(0.05), therefore, I would say that the data is not indepedent. There is a association between different country and the characteristics of the infants. It makes sense in a way that in different countries, culture might include infant's attitude to caregivers, also the training that caregiver recieves might be different in each country.

b)

```
library(FactoMineR)
```

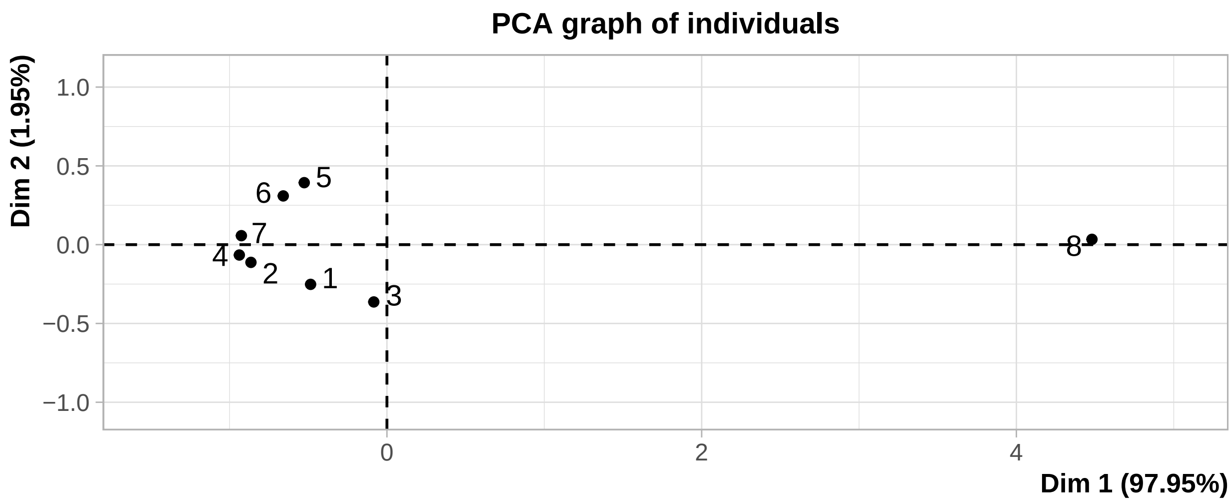
```
## Warning: package 'FactoMineR' was built under R version 3.6.2
```

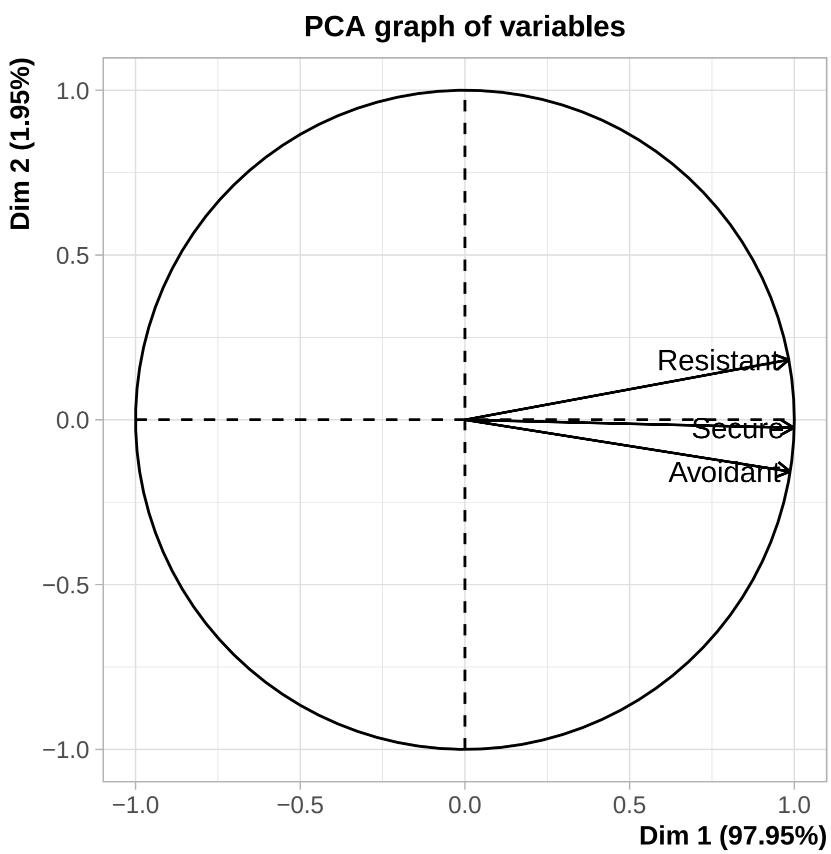
```
#component score
VIJ.PCA<- PCA(VIJ_data[,-1], graph =F)
summary(VIJ.PCA)
```

```
##
## Call:
## PCA(X = VIJ_data[, -1], graph = F)
##
##
## Eigenvalues
##                        Dim.1   Dim.2   Dim.3
## Variance               2.939   0.058   0.003
## % of var.             97.951   1.948   0.102
## Cumulative % of var.  97.951  99.898 100.000
##
## Individuals
##                 Dist     Dim.1     ctr    cos2     Dim.2     ctr    cos2     Dim.3
```

```
## 1          |  0.558 | -0.485  1.001  0.756 | -0.252 13.573  0.204 |  0.112
## 2          |  0.874 | -0.864  3.177  0.978 | -0.112  2.705  0.017 | -0.062
## 3          |  0.377 | -0.084  0.030  0.049 | -0.364 28.275  0.932 | -0.052
## 4          |  0.941 | -0.938  3.743  0.993 | -0.066  0.926  0.005 | -0.040
## 5          |  0.657 | -0.526  1.175  0.641 |  0.393 33.111  0.359 |  0.008
## 6          |  0.728 | -0.659  1.847  0.819 |  0.309 20.469  0.180 | -0.022
## 7          |  0.928 | -0.925  3.640  0.993 |  0.057  0.696  0.004 |  0.056
## 8          |  4.480 |  4.480 85.388  1.000 |  0.034  0.246  0.000 |  0.000
##               ctr   cos2
## 1          51.322  0.040 |
## 2          15.817  0.005 |
## 3          11.210  0.019 |
## 4           6.495  0.002 |
## 5           0.264  0.000 |
## 6           1.976  0.001 |
## 7          12.916  0.004 |
## 8           0.000  0.000 |
##
## Variables
##               Dim.1    ctr   cos2    Dim.2    ctr   cos2    Dim.3    ctr
## Avoidant  |  0.987 33.166  0.975 | -0.157 42.206  0.025 |  0.027 24.628
## Secure    |  0.999 33.943  0.997 | -0.024  0.994  0.001 | -0.045 65.063
## Resistant |  0.983 32.891  0.966 |  0.182 56.800  0.033 |  0.018 10.309
##              cos2
## Avoidant   0.001 |
## Secure     0.002 |
## Resistant  0.000 |
```

```r
#PCA graph
PCA(VIJ_data[,-1])
```


PCA graph of individuals
Dim 2 (1.95%)
Dim 1 (97.95%)
1.0
0.5
0.0
−0.5
−1.0
0
2
4
1
2
3
4
5
6
7
8




```
## **Results for the Principal Component Analysis (PCA)**
## The analysis was performed on 8 individuals, described by 3 variables
## *The results are available in the following objects:
##
##    name               description
## 1  "$eig"             "eigenvalues"
## 2  "$var"             "results for the variables"
## 3  "$var$coord"       "coord. for the variables"
## 4  "$var$cor"         "correlations variables - dimensions"
## 5  "$var$cos2"        "cos2 for the variables"
## 6  "$var$contrib"     "contributions of the variables"
## 7  "$ind"             "results for the individuals"
## 8  "$ind$coord"       "coord. for the individuals"
## 9  "$ind$cos2"        "cos2 for the individuals"
## 10 "$ind$contrib"     "contributions of the individuals"
## 11 "$call"            "summary statistics"
## 12 "$call$centre"     "mean of the variables"
## 13 "$call$ecart.type" "standard error of the variables"
## 14 "$call$row.w"      "weights for the individuals"
## 15 "$call$col.w"      "weights for the variables"
```

```
#eigenvalues
VIJ.PCA$eig
```

```
##         eigenvalue percentage of variance
## comp 1 2.938523204             97.9507735
## comp 2 0.058427324              1.9475775
```

```
## comp 3 0.003049472                0.1016491
##        cumulative percentage of variance
## comp 1                             97.95077
## comp 2                             99.89835
## comp 3                            100.00000
```

Base on the PCA we run on this dataset, we can conclude that there is probably only dimension of data that matters since 97% of the variance can be explained or heavily lie on the first dimension. Therefore, there is no really point of performing PCA on a dataset that only has one dimension of data. However, we could apply Correspondence analysis in this situation. As the data itself is a cross-tabular data that are categorical or ratio-scale.

c)

```r
library(ca)
```

```
## Warning: package 'ca' was built under R version 3.6.2
```

```r
library(factoextra)
```

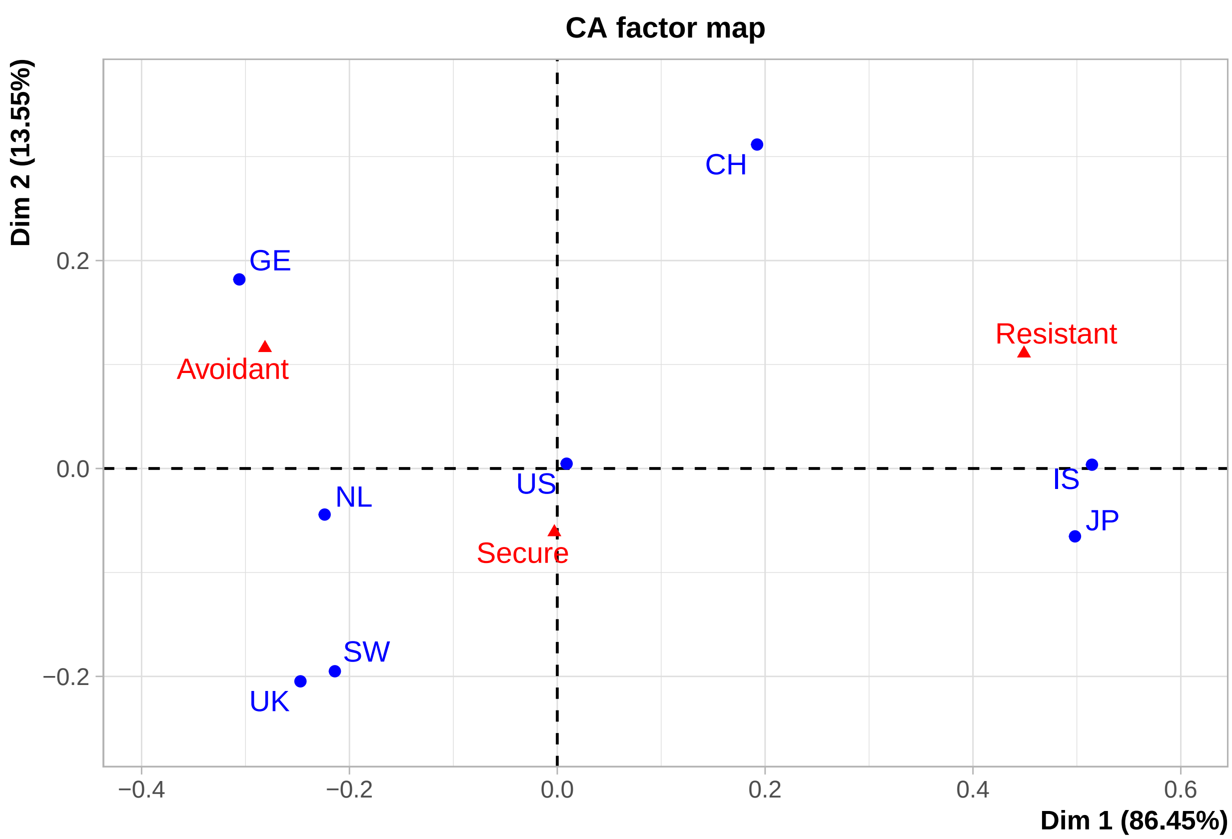
```
## Warning: package 'factoextra' was built under R version 3.6.2

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.6.2

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```r
VIJ_CA<- CA(VIJ_tab, ncp=5,graph=TRUE)
```



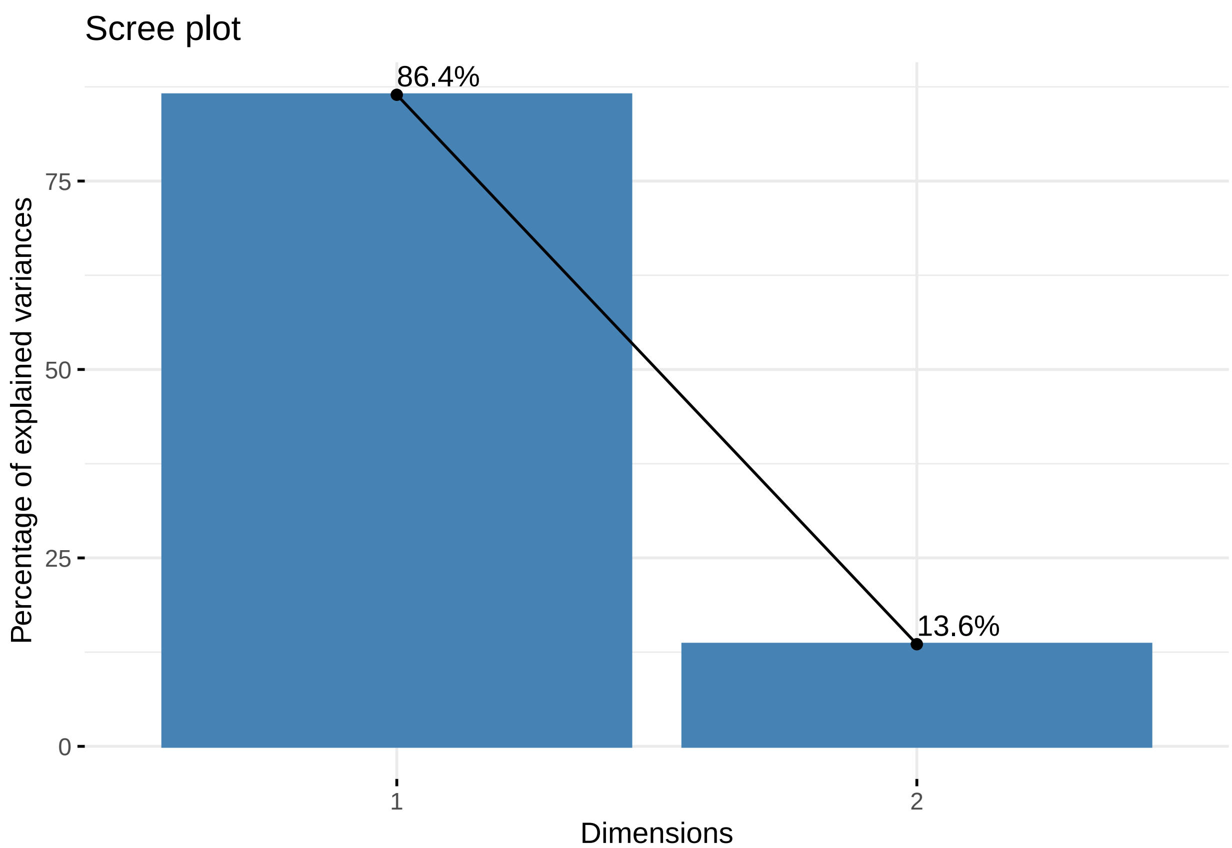
```r
VIJ_CA
```

```
## **Results of the Correspondence Analysis (CA)**
## The row variable has  8  categories; the column variable has 3 categories
## The chi square of independence between the two variables is equal to 102.4215 (p-value =  1.626285e-
## *The results are available in the following objects:
##
##    name              description
## 1  "$eig"            "eigenvalues"
## 2  "$col"            "results for the columns"
## 3  "$col$coord"      "coord. for the columns"
## 4  "$col$cos2"       "cos2 for the columns"
## 5  "$col$contrib"    "contributions of the columns"
## 6  "$row"            "results for the rows"
## 7  "$row$coord"      "coord. for the rows"
## 8  "$row$cos2"       "cos2 for the rows"
## 9  "$row$contrib"    "contributions of the rows"
## 10 "$call"           "summary called parameters"
## 11 "$call$marge.col" "weights of the columns"
## 12 "$call$marge.row" "weights of the rows"
```

```r
summary(VIJ_CA)
```

```
##
## Call:
## CA(X = VIJ_tab, ncp = 5, graph = TRUE)
##
## The chi square of independence between the two variables is equal to 102.4215 (p-value =  1.626285e-
##
## Eigenvalues
##                        Dim.1   Dim.2
## Variance               0.044   0.007
## % of var.             86.446  13.554
## Cumulative % of var.  86.446 100.000
##
## Rows
##             Iner*1000    Dim.1    ctr   cos2    Dim.2    ctr   cos2
## GE        |     8.658 | -0.306 14.379  0.739 |  0.182 32.404  0.261 |
## UK        |     3.726 | -0.247  4.967  0.593 | -0.205 21.737  0.407 |
## NL        |     6.568 | -0.224 14.207  0.962 | -0.044  3.548  0.038 |
## SW        |     2.149 | -0.214  2.638  0.546 | -0.195 13.976  0.454 |
## IS        |    15.699 |  0.515 35.282  1.000 |  0.004  0.011  0.000 |
## JP        |    12.179 |  0.498 26.912  0.983 | -0.065  2.949  0.017 |
## CH        |     2.426 |  0.192  1.503  0.276 |  0.312 25.185  0.724 |
## US        |     0.063 |  0.009  0.112  0.790 |  0.005  0.190  0.210 |
##
## Columns
##             Iner*1000    Dim.1    ctr   cos2    Dim.2    ctr   cos2
## Avoidant  |    19.674 | -0.281 37.798  0.855 |  0.116 40.945  0.145 |
## Secure    |     2.444 | -0.003  0.011  0.002 | -0.061 34.963  0.998 |
## Resistant |    29.350 |  0.449 62.190  0.943 |  0.111 24.091  0.057 |
```

```r
#scree plot
fviz_screeplot(VIJ_CA, addlabels = TRUE)
```



From the Correspondence analysis we performed above, we can conclude that most of the variance can be explained in a two-dimensional scale. We can tell that UK and SW; IS and JP are relatively similar to each other in its baby's response when the caregiver is out of sight. This matches the row profiles that are given, the distribution of baby's response of secure, resistant, and avoidant are similar between UK and SW; IS and JP. Also, take NL as an example, we can tell by its distance with origin and the angle it forms with the column variable that NL is strongly associated with the response secure. All of the above indicates that it is consistent with the row profiles.

d)

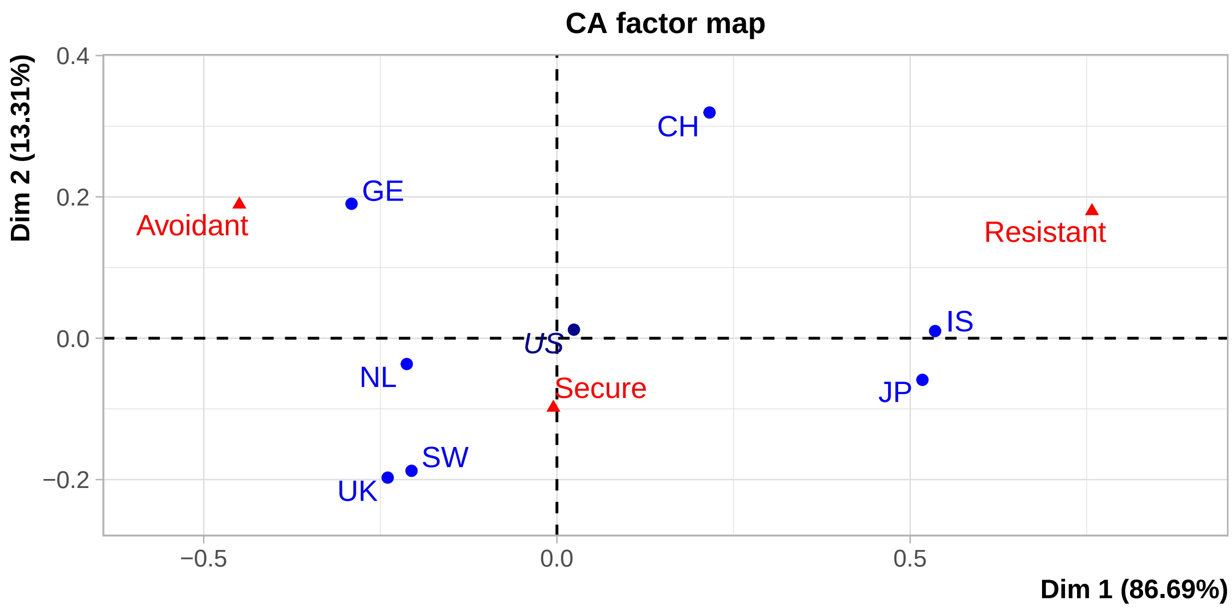
```
get_eigenvalue(VIJ_CA)
```

```
##        eigenvalue variance.percent cumulative.variance.percent
## Dim.1 0.044492190         86.44618                    86.44618
## Dim.2 0.006975892         13.55382                   100.00000
```

From the CA we run, we can find out that there only exists two eigenvalue after the decomposition using the scaling data. It is essentially project the high dimensional space into two-dimensional subspace. Also, the percentage of variance can be completely explained by only two dimension. Therefore, the data are accounted by two dimensions in the correspondence analysis.

e)

```
#make USA supplementary
USA_sup_CA<- CA(VIJ_tab, ncp= 5, row.sup = c(8), col.sup = NULL, graph = T)
```




```
USA_sup_CA
```

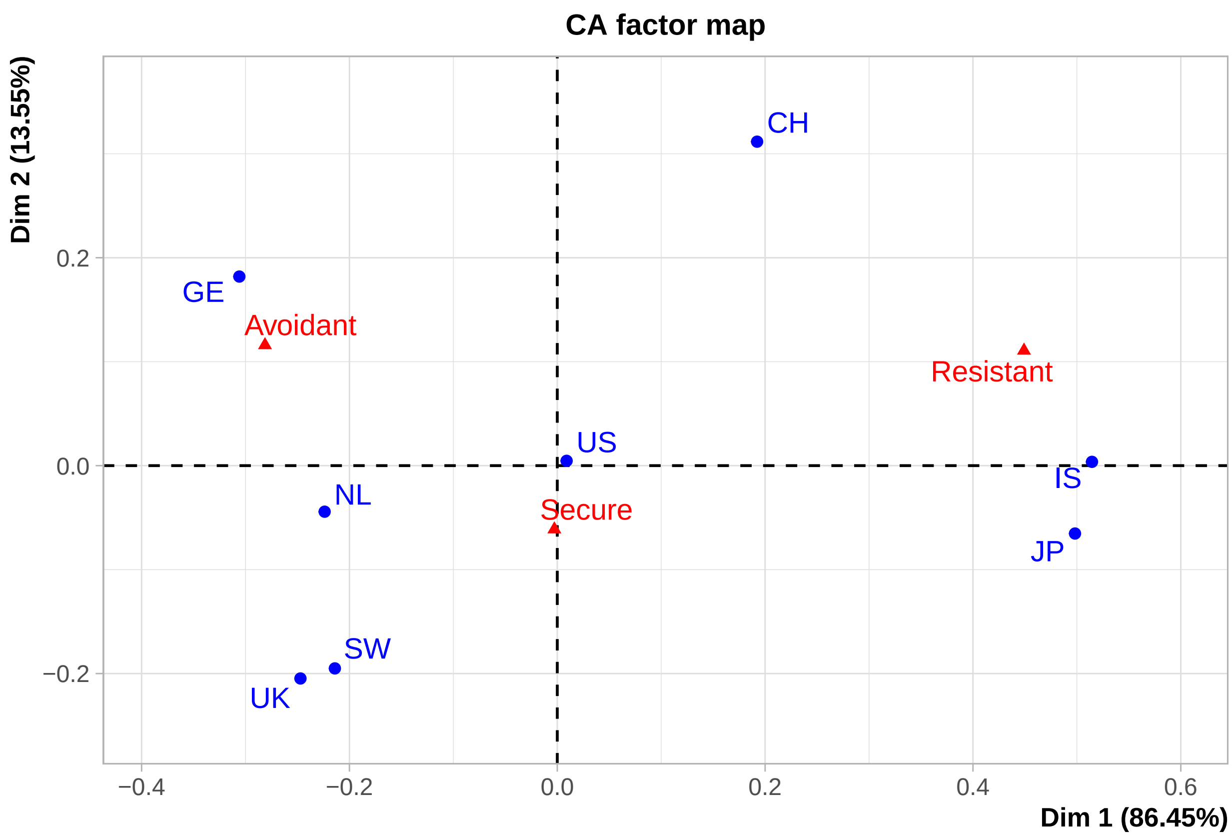
```
## **Results of the Correspondence Analysis (CA)**
## The row variable has  7  categories; the column variable has 3 categories
## The chi square of independence between the two variables is equal to 104.1937 (p-value =  8.363526e-
## *The results are available in the following objects:
##
##    name                 description
## 1  "$eig"               "eigenvalues"
## 2  "$col"               "results for the columns"
## 3  "$col$coord"         "coord. for the columns"
## 4  "$col$cos2"          "cos2 for the columns"
## 5  "$col$contrib"       "contributions of the columns"
## 6  "$row"               "results for the rows"
## 7  "$row$coord"         "coord. for the rows"
## 8  "$row$cos2"          "cos2 for the rows"
## 9  "$row$contrib"       "contributions of the rows"
## 10 "$row.sup$coord"     "coord. for supplementary rows"
## 11 "$row.sup$cos2"      "cos2 for supplementary rows"
## 12 "$call"              "summary called parameters"
## 13 "$call$marge.col"    "weights of the columns"
## 14 "$call$marge.row"    "weights of the rows"
```

```
summary(USA_sup_CA)
```

```
##
## Call:
```

```
## CA(X = VIJ_tab, ncp = 5, row.sup = c(8), col.sup = NULL, graph = T)
##
## The chi square of independence between the two variables is equal to 104.1937 (p-value =  8.363526e-
##
## Eigenvalues
##                        Dim.1   Dim.2
## Variance               0.119   0.018
## % of var.             86.687  13.313
## Cumulative % of var.  86.687 100.000
##
## Rows
##             Iner*1000    Dim.1    ctr   cos2    Dim.2    ctr   cos2
## GE        |    21.601 | -0.291 12.724  0.700 |  0.190 35.499  0.300 |
## UK        |     9.121 | -0.240  4.574  0.596 | -0.197 20.189  0.404 |
## NL        |    15.357 | -0.213 12.552  0.971 | -0.037  2.411  0.029 |
## SW        |     5.204 | -0.206  2.392  0.546 | -0.188 12.937  0.454 |
## IS        |    44.520 |  0.535 37.447  1.000 |  0.010  0.088  0.000 |
## JP        |    34.251 |  0.517 28.451  0.987 | -0.059  2.398  0.013 |
## CH        |     7.044 |  0.216  1.861  0.314 |  0.319 26.478  0.686 |
##
## Columns
##             Iner*1000    Dim.1    ctr   cos2    Dim.2    ctr   cos2
## Avoidant  |    51.027 | -0.450 36.474  0.849 |  0.189 42.079  0.151 |
## Secure    |     6.330 | -0.005  0.013  0.003 | -0.098 34.592  0.997 |
## Resistant |    79.740 |  0.757 63.513  0.947 |  0.180 23.329  0.053 |
##
## Supplementary row
##             Dim.1  cos2   Dim.2  cos2
## US        | 0.024 0.798 | 0.012 0.202 |
```

```
#original ca
CA(VIJ_tab, ncp=5,graph=TRUE)
```




```
## **Results of the Correspondence Analysis (CA)**
## The row variable has  8  categories; the column variable has 3 categories
## The chi square of independence between the two variables is equal to 102.4215 (p-value =  1.626285e-
## *The results are available in the following objects:
##
##    name              description
## 1  "$eig"            "eigenvalues"
## 2  "$col"            "results for the columns"
## 3  "$col$coord"      "coord. for the columns"
## 4  "$col$cos2"       "cos2 for the columns"
## 5  "$col$contrib"    "contributions of the columns"
## 6  "$row"            "results for the rows"
## 7  "$row$coord"      "coord. for the rows"
## 8  "$row$cos2"       "cos2 for the rows"
## 9  "$row$contrib"    "contributions of the rows"
## 10 "$call"           "summary called parameters"
## 11 "$call$marge.col" "weights of the columns"
## 12 "$call$marge.row" "weights of the rows"
```

After making USA supplementary in the correspondence analysis, we can conclude that there are not major changes to the configuration plot. The similarity between countries does not affected by making USA a supplementary row. However, the column variable(response) are shifted away from the origin. This implies that the column varible now has a greater impact on the correspondence analysis. And the association of the column variable(response) with the row variable(countries) are stronger than before. However, one thing that seems odd is the fact that GE have a stronger association with "Avoidant" while the row profiles suggest otherwise. The row profiles indicates that more response are chosen as "Secure".