ex7

jacopo

2023-12-01

Exercise 1

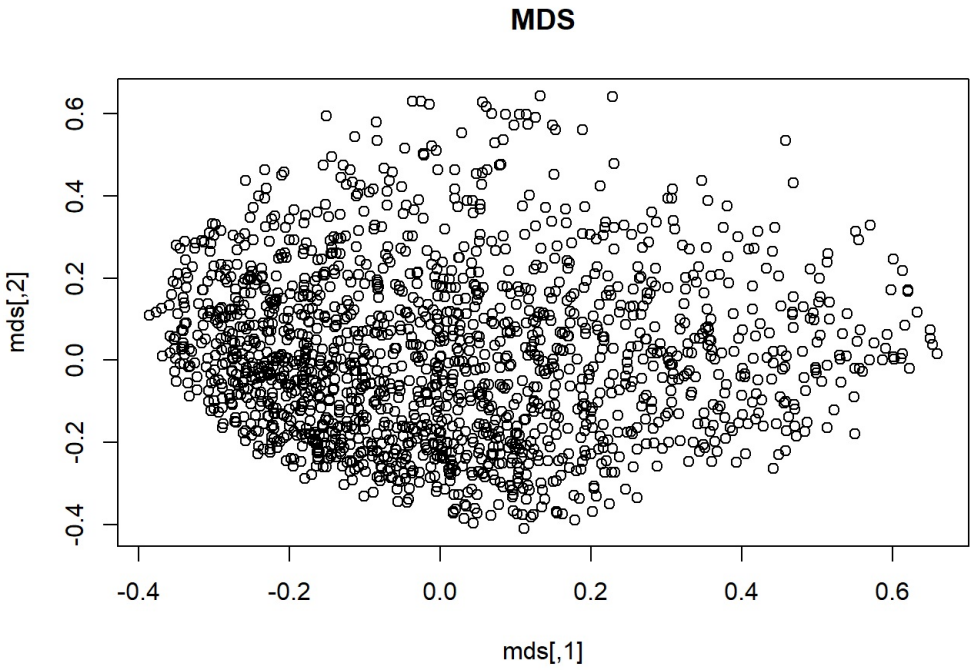
Survey data from the 2000 American National Election Study. Two sets of six questions with four responses each, asking respondents' opinions of how well various traits (moral, caring, knowledgable, good leader, dishonest, intelligent) describe presidential candidates Al Gore and George W. Bush. The responses are (1) Extremely well; (2) Quite well; (3) Not too well; (4) Not well at all. Many respondents have varying numbers of missing values on these variables. The data set also includes potential covariates VOTE3, the respondent's 2000 vote choice (when asked); AGE, the respondent's age; EDUC, the respondent's level of education; GENDER, the respondent's gender; and PARTY, the respondent's Democratic-Republican partisan identification.

VOTE3 is coded as (1) Gore; (2) Bush; (3) Other.

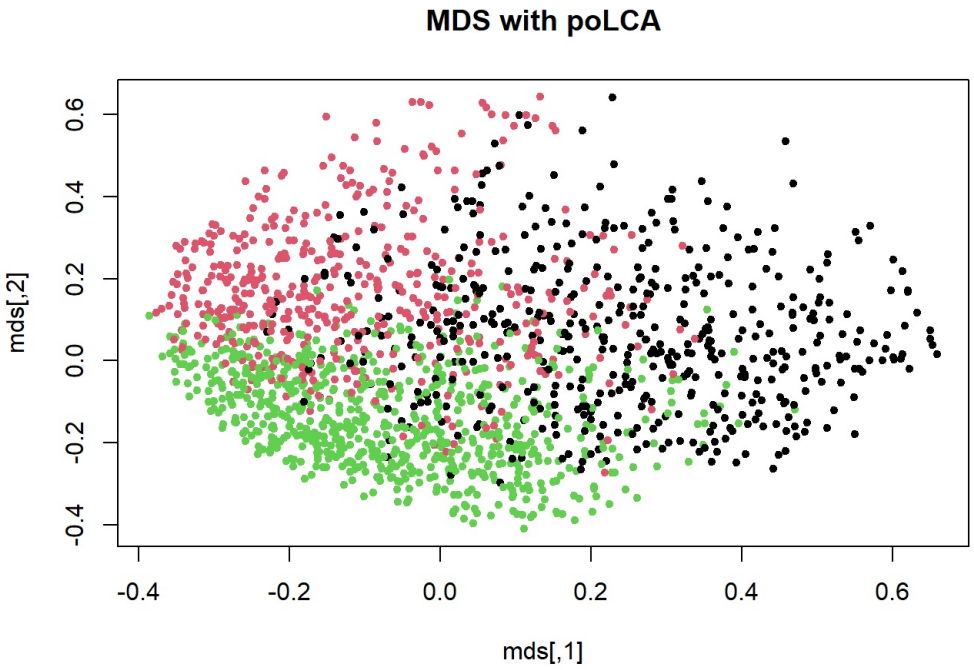
EDUC is coded as (1) 8 grades or less; (2) 9-11 grades, no further schooling; (3) High school diploma or equivalency; (4) More than 12 years of schooling, no higher degree; (5) Junior or community college level degree; (6) BA level degrees, no advanced degree; (7) Advanced degree.

GENDER is coded as (1) Male; (2) Female.

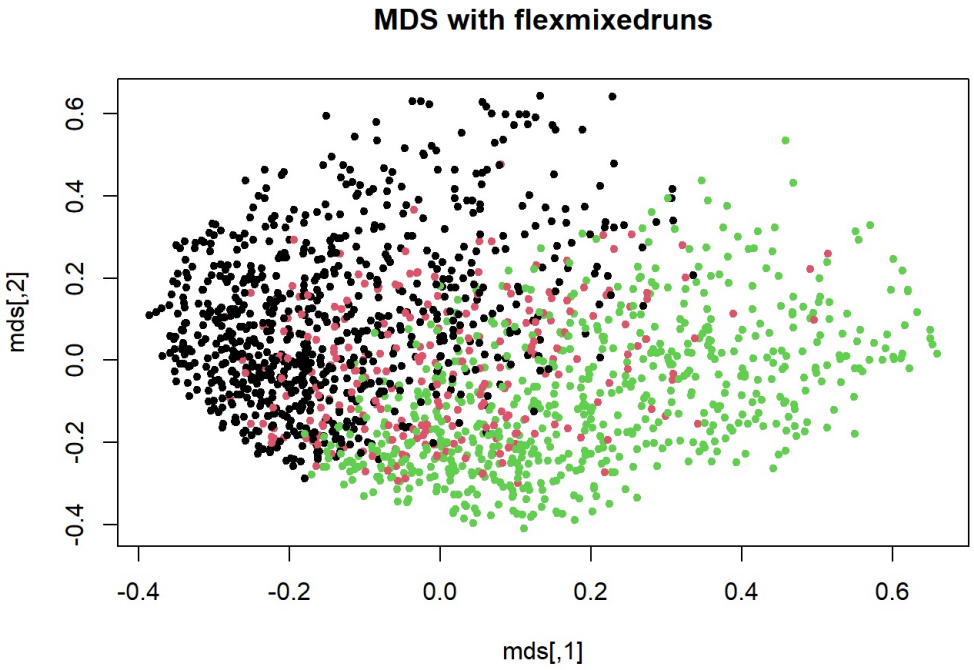
PARTY is coded as (1) Strong Democrat; (2) Weak Democrat; (3) Independent-Democrat; (4) Independent-Independent; (5) Independent-Republican; (6) Weak Republican; (7) Strong Republican.



a



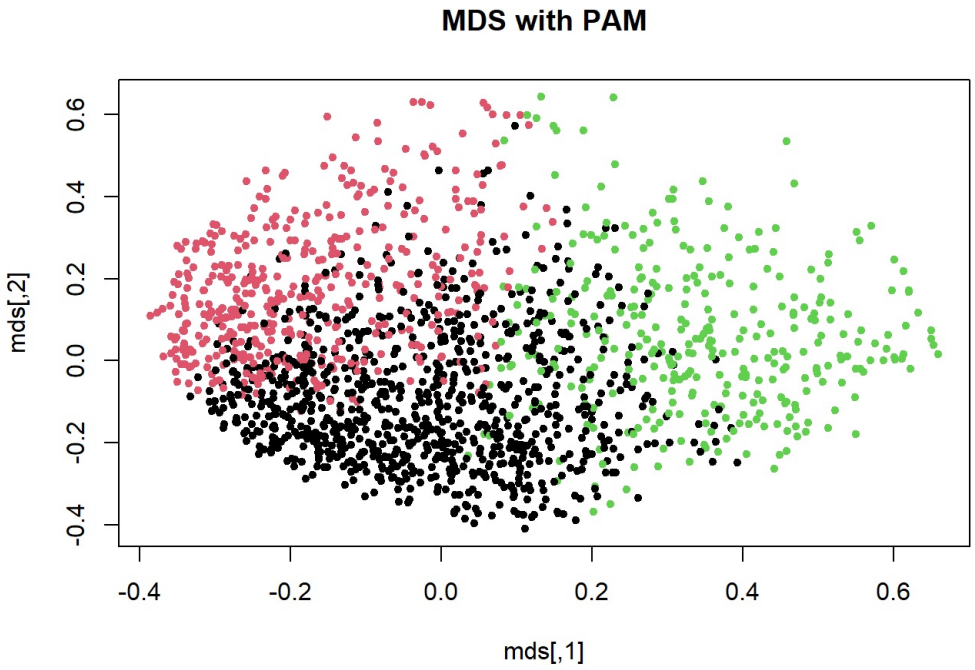
b

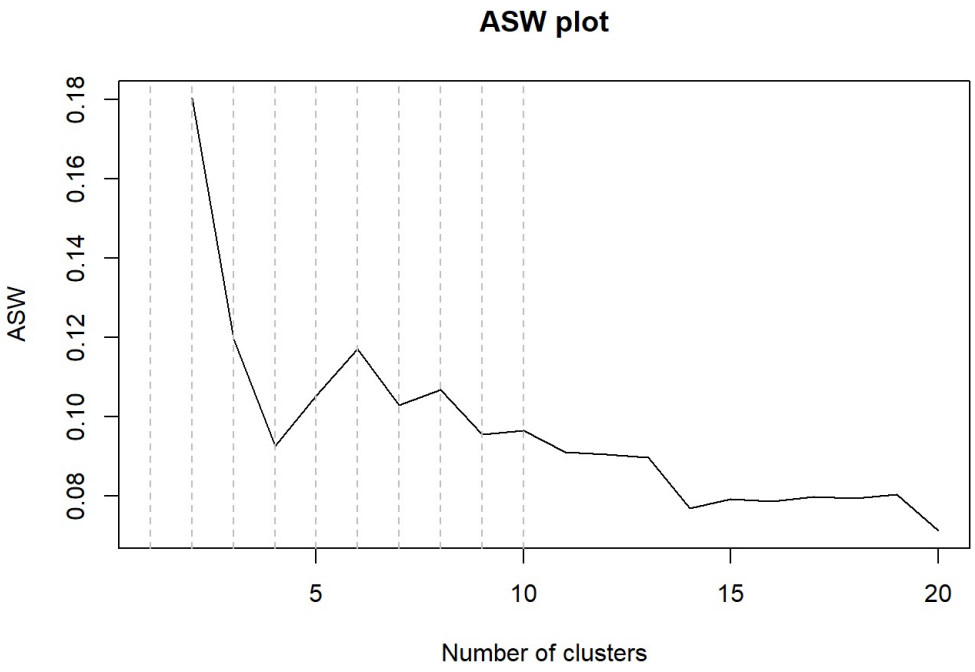


BIC comparison

```
##   Model      BIC
## 2    Ma 55802.96
## 3    Mb 52863.75
```

c





2, 3 and 6 are all good levels for k

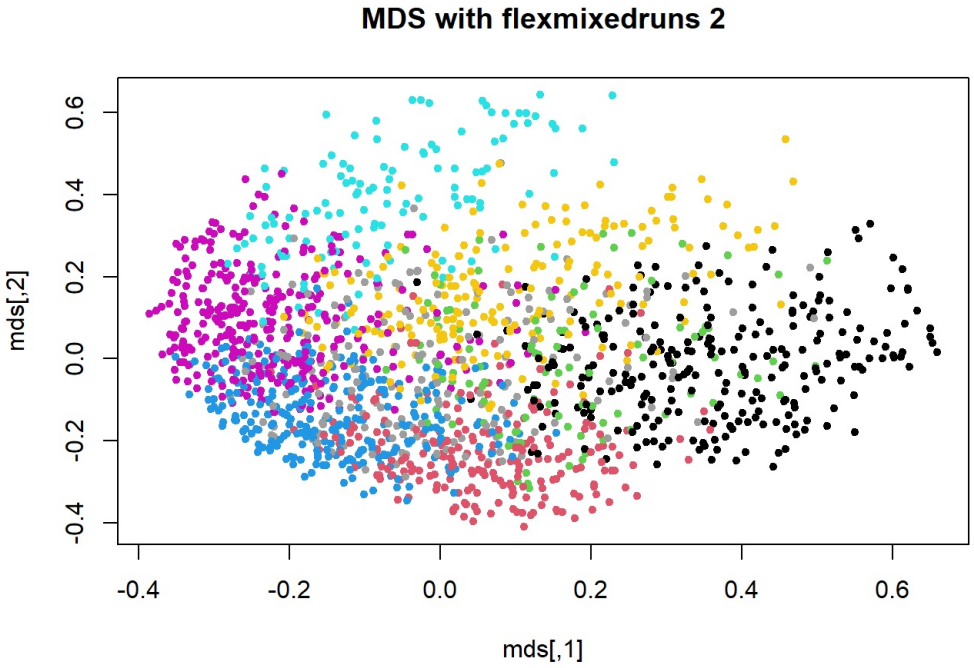
d

```
Md$optimalk
```

```
## [1] 8
```

```
Md$bicvals[8]
```

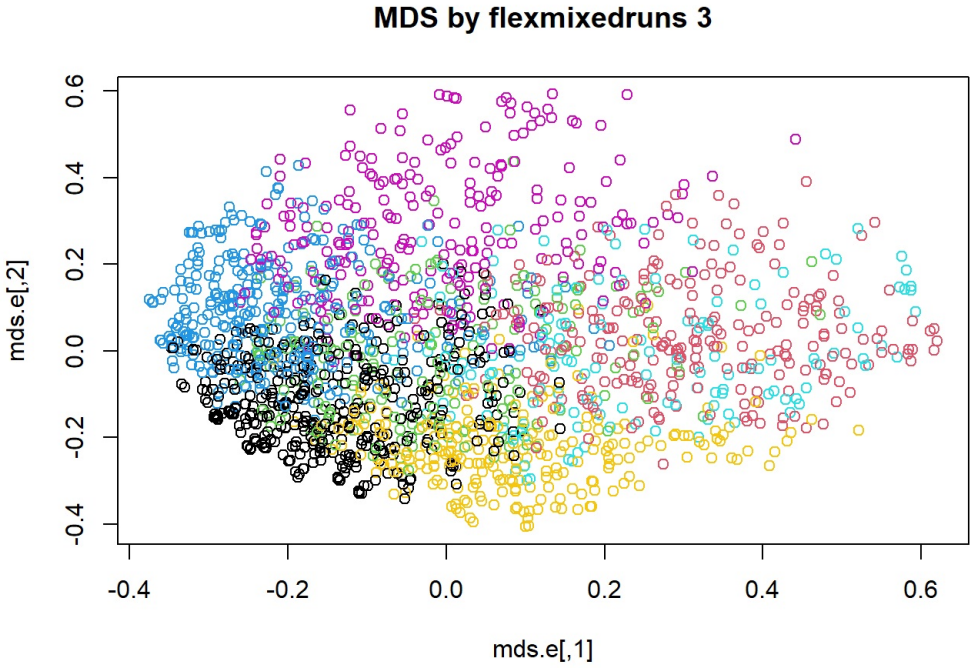
```
## [1] 50744.06
```



e

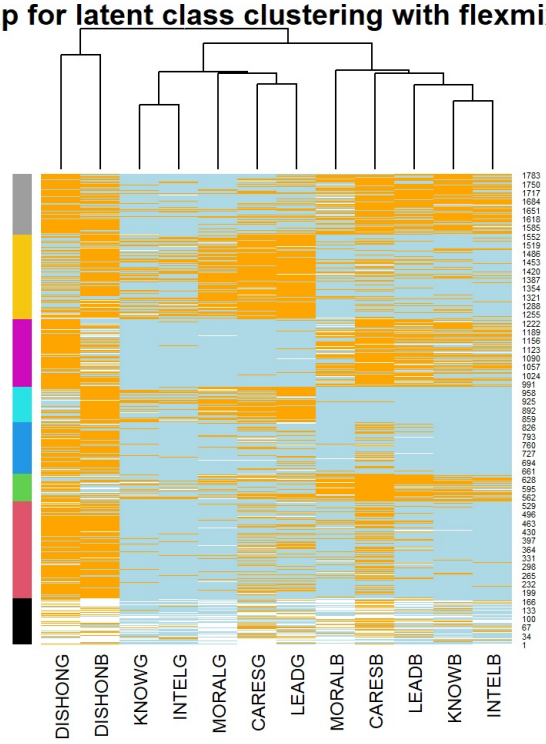
```
Me$optimalk
```

```
## [1] 8
```



Exercise 2

Heatmap for latent class clustering with flexmixedruns

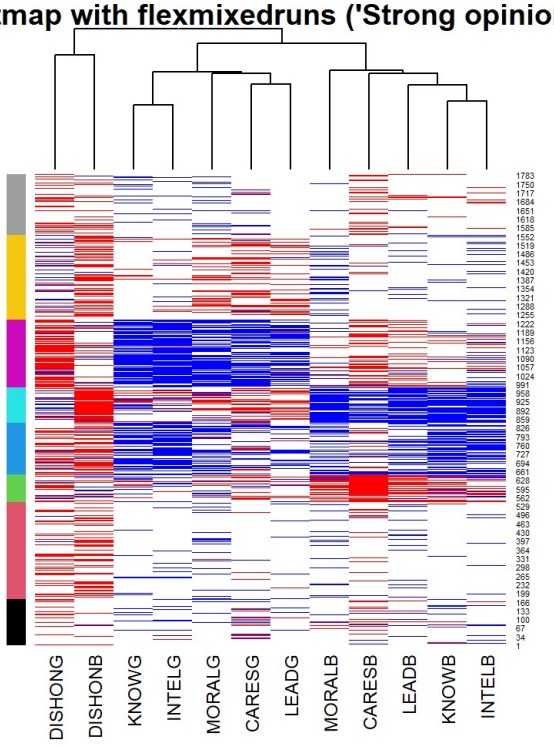


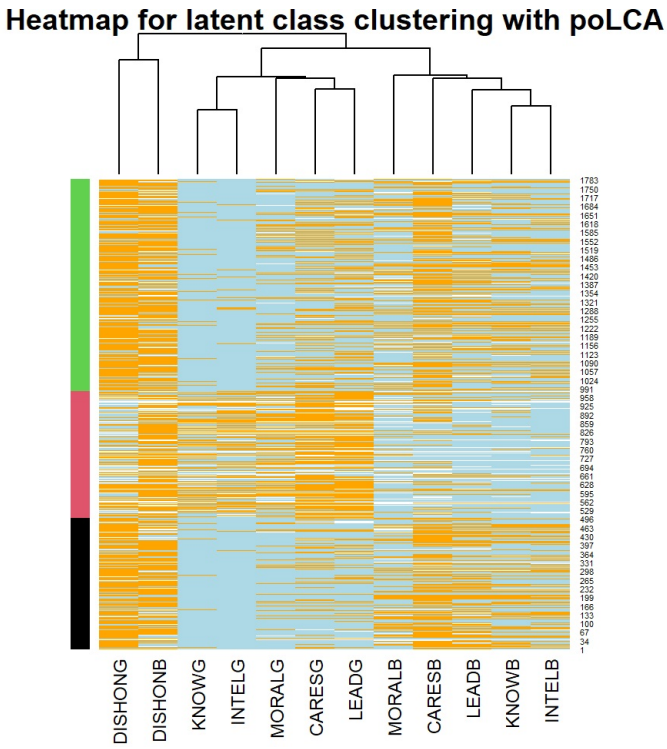
Positive Judgment
Negative Judgment
No Answer

Plot seems to underline how dishonesty isn't considered a characteristic that describes one of the two candidates, exception made for the first cluster, the black one, that thinks Gore is enough dishonest. We also notice that lightblue cluster (number 5) contains almost all the NA observations. While yellow and blue cluster have, in general, a good opinion of both the candidates, violet one has a bad opinion of both. To conclude, black and green clusters seem to prefer Gore and, while red and grey ones seem to prefer Bush.

I personally find the clustering convincing, because the distinctions are pretty clear. Data really seem locally dependent

Heatmap with flexmixedruns ('Strong opinions')





Exercise 3

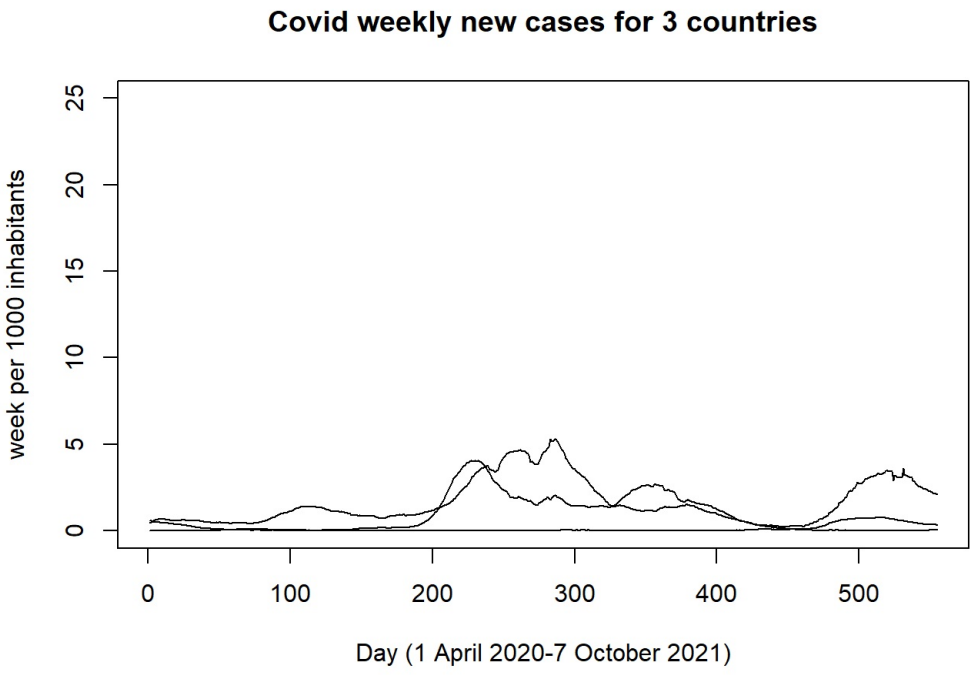
```
5+6+8
```

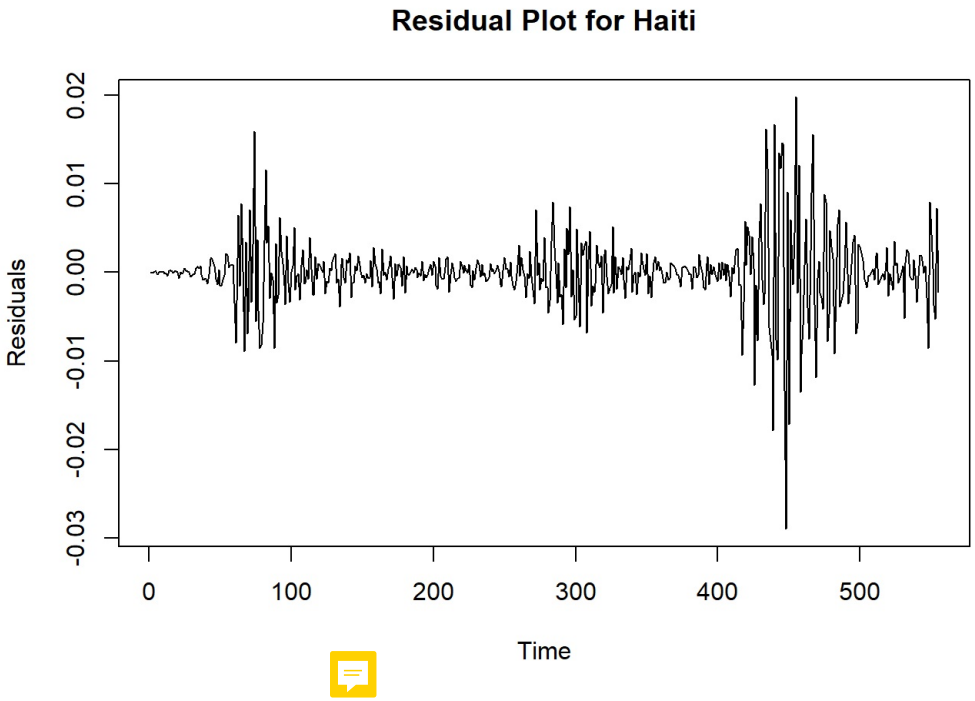
```
## [1] 19
```

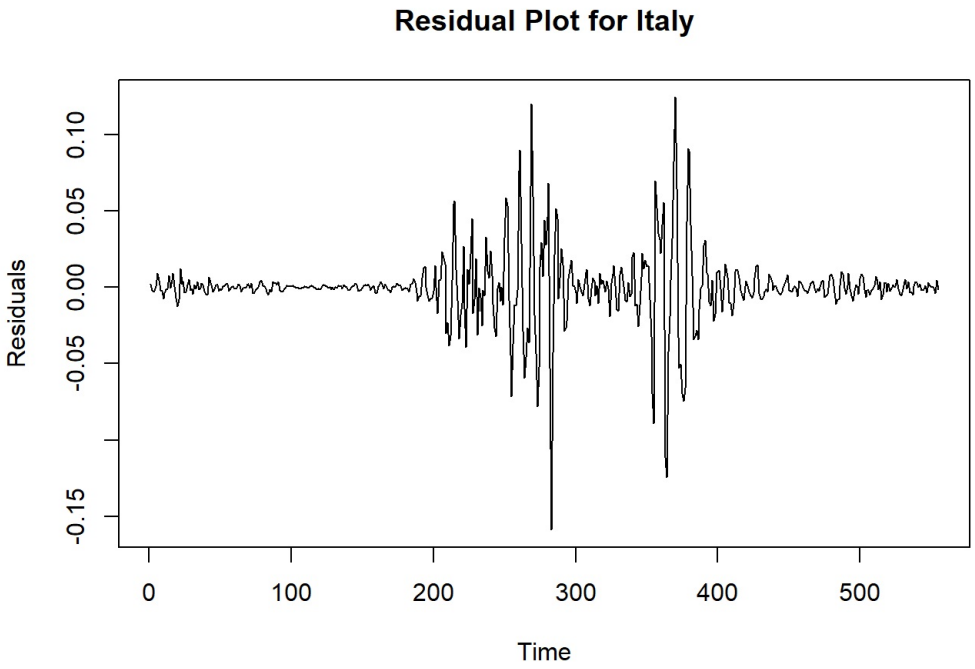
```
3+(1*3+2*3+4*3)
```

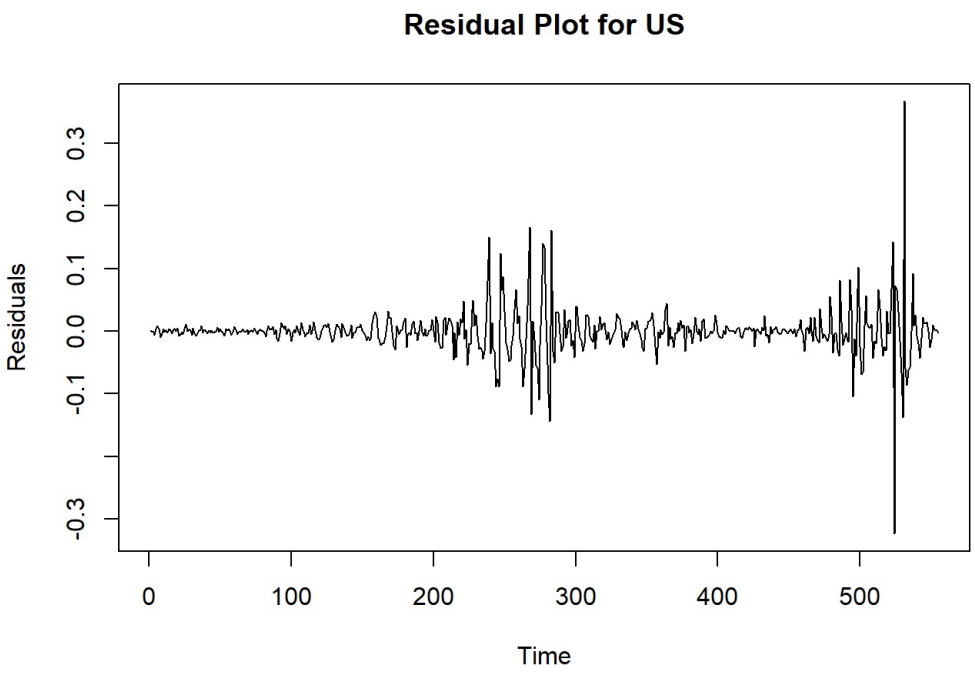
```
## [1] 24
```

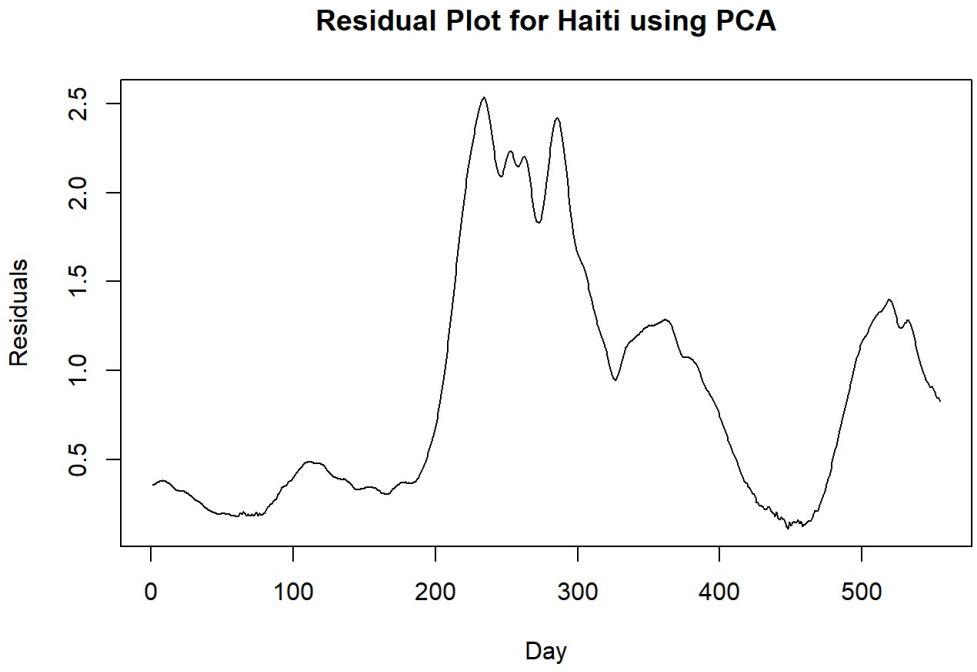
Exercise 4

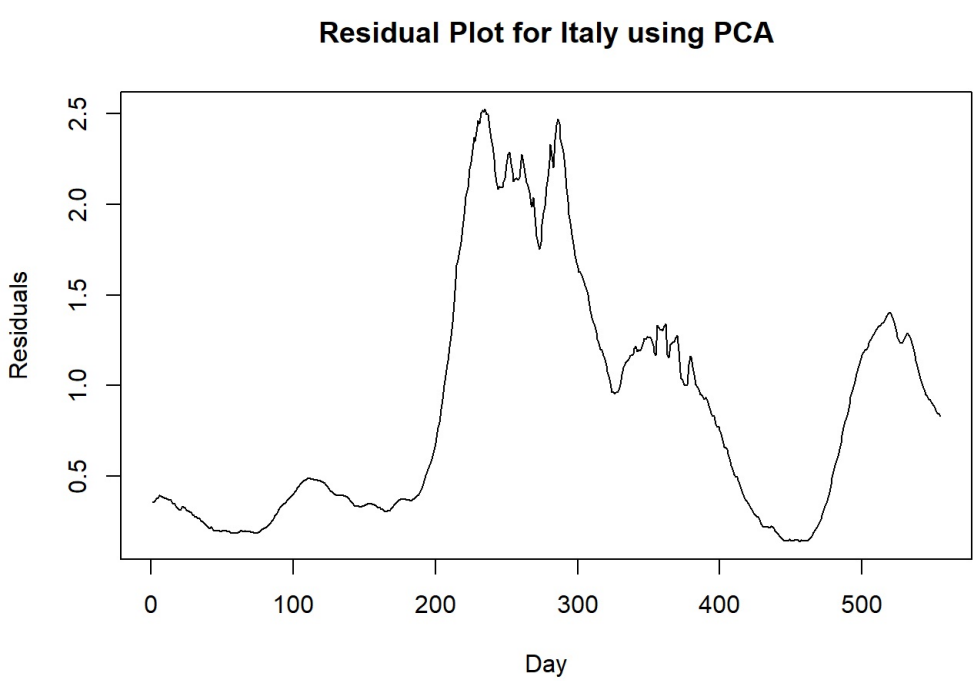


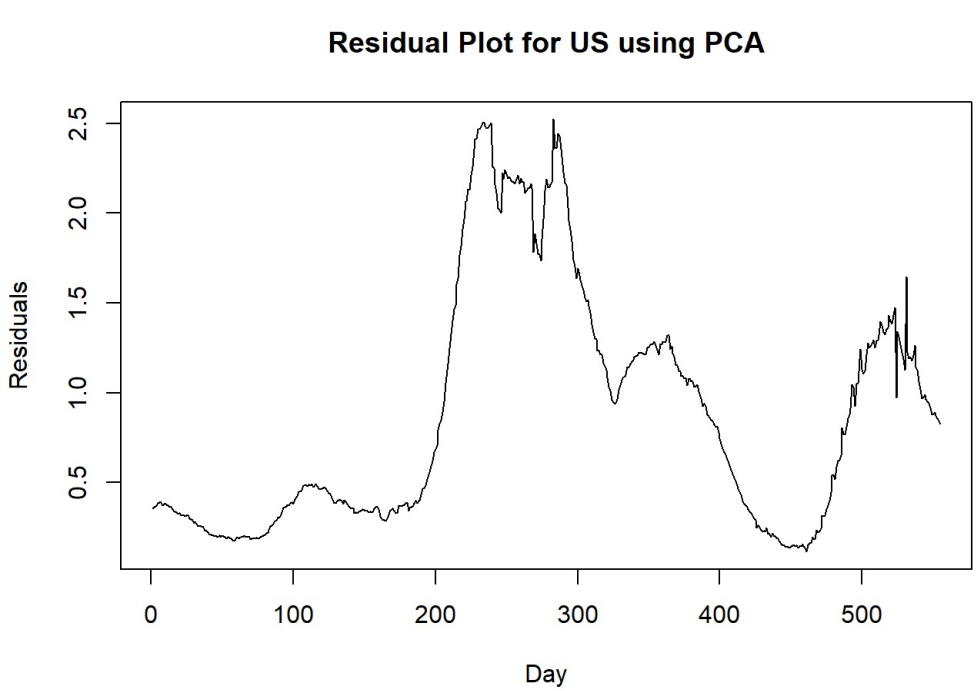

Residual Plot for Haiti
Residuals
0.02
0.01
0.00
-0.01
-0.02
-0.03
0
100
200
300
400
500
Time


Residual Plot for Italy
Residuals
0.10
0.05
0.00
-0.05
-0.15
0
100
200
300
400
500
Time


Residual Plot for US
Residuals
0.3
0.2
0.1
0.0
-0.1
-0.3
0
100
200
300
400
500
Time

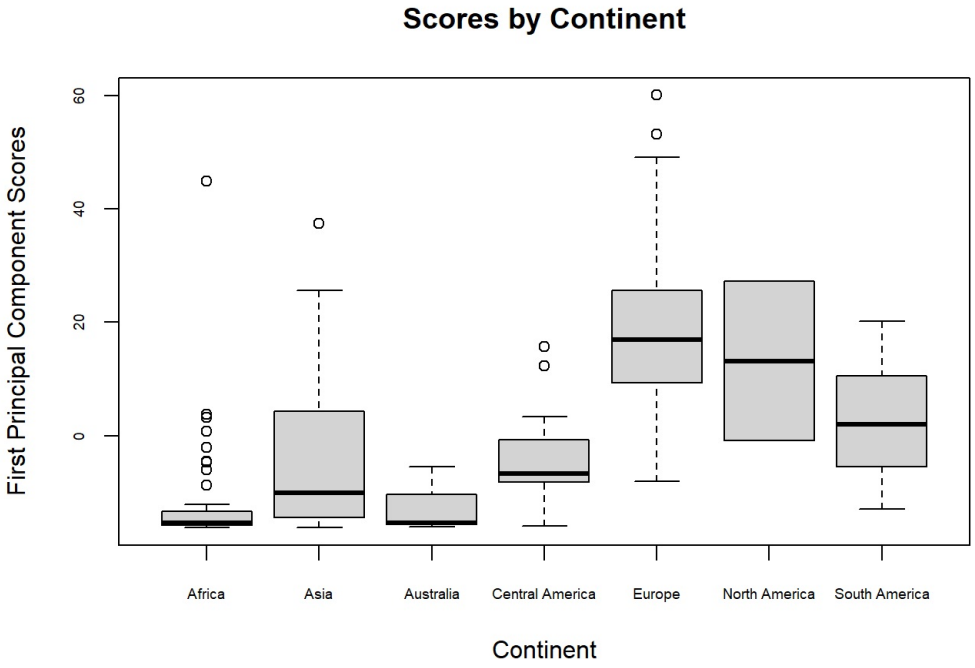

Residual Plot for Haiti using PCA
Residuals
0.5
1.0
1.5
2.0
2.5
0
100
200
300
400
500
Day


Residual Plot for Italy using PCA
Residuals
0.5
1.0
1.5
2.0
2.5
0
100
200
300
400
500
Day


Residual Plot for US using PCA
Residuals
0.5
1.0
1.5
2.0
2.5
0
100
200
300
400
500
Day

Exercise 5

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Continent    6  25062    4177   27.62 <2e-16 ***
## Residuals  172  26008     151
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



scores interpretation

```
##        Continents Scores Averages
## 1          Africa      -11.984971
## 2            Asia       -3.486471
## 3       Australia      -12.277853
## 4 Central America       19.006286
## 5          Europe       13.196160
## 6   North America        2.919687
## 7   South America       -4.288695
```

We notice visible differences between the continents, in particular we notice a sort of clustering between , we could say, average poorer ones and average richer ones. The ANOVA test confirms the result and also looking at the average scores in the last table we see a sort of trend that confirm this thesis. I don’t know if the richness factor is the clue one, but it should be something at least related with it.