Seshat

Achille-Laurent

03/06/2020

# Analyse de la base Seshat

Code concernant la première approche, l’ACP, la classification et la régression

# 1. Données Axial

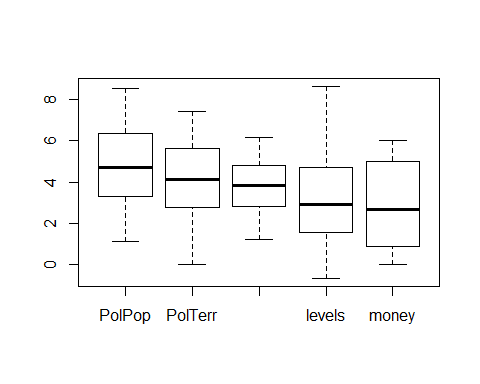
## 1.1 Importation

dfn <- read.csv('../databases/axial\_index.csv',sep=',')  
# Mettre l'Index en index et retirer SPC1  
data <- subset(dfn, select=-c(Index,Time,NGA,PolID))  
rownames(data) <- dfn$Index  
attach(data)  
head(data)

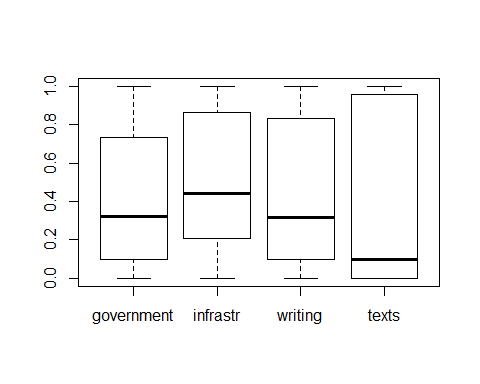
## PolPop PolTerr CapPop levels government infrastr  
## Konya Plain (-9600) 3.293724 4.138263 1.442118 -0.05581711 0 0  
## Konya Plain (-9500) 3.293724 4.138263 1.442118 -0.05581711 0 0  
## Konya Plain (-9400) 3.293724 4.138263 1.442118 -0.05581711 0 0  
## Konya Plain (-9300) 3.293724 4.138263 1.442118 -0.05581711 0 0  
## Konya Plain (-9200) 3.293724 4.138263 1.442118 -0.05581711 0 0  
## Konya Plain (-9100) 3.293724 4.138263 1.442118 -0.05581711 0 0  
## writing texts money  
## Konya Plain (-9600) 0 0 0  
## Konya Plain (-9500) 0 0 0  
## Konya Plain (-9400) 0 0 0  
## Konya Plain (-9300) 0 0 0  
## Konya Plain (-9200) 0 0 0  
## Konya Plain (-9100) 0 0 0

## 1.2. Vue d’ensemble

# Séparer les 2 plots qui sont d'échelle différente  
boxplot(subset(data,select = c(PolPop,PolTerr,CapPop,levels,money)))



boxplot(subset(data,select = c(government,infrastr,writing,texts)))

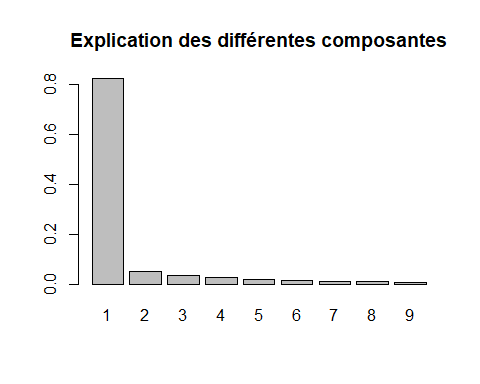


# 2. Analyse en Composante Principale

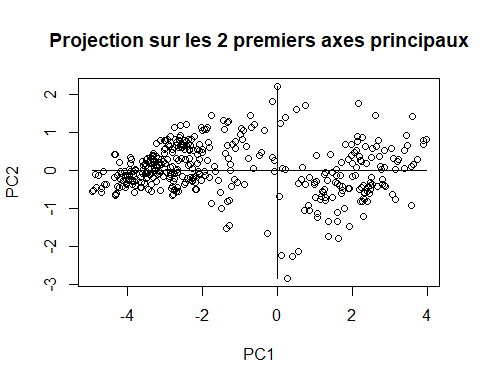
X= scale(data, center=T, scale=T)   
  
S = cov(X)  
acp = eigen(S)  
lambda = acp$values  
vecteurs\_propres = acp$vectors  
Inertie = sum(diag(S))  
part.inertie = lambda/sum(lambda)

## 2.1 Explication

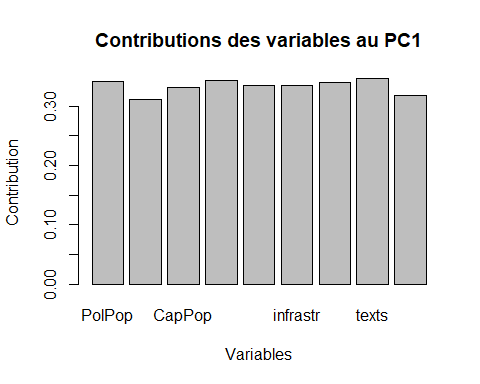
barplot(lambda/sum(lambda),names.arg = 1:length(lambda))  
title(main="Explication des différentes composantes")

 ## 2.2. Projection

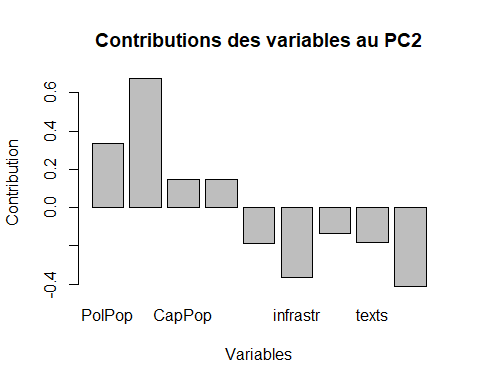
# Les composantes principales :   
C = X %\*% vecteurs\_propres  
# colnames(C) = paste("comp", 1:4)  
plot(C[,1:2],type="p",xlab='PC1',ylab='PC2')  
# text(C[,1:2])  
title(main="Projection sur les 2 premiers axes principaux")  
lines(c(min(C[,1]),max(C[,1])),c(0,0))  
lines(c(0,0),c(min(C[,2]),max(C[,2])))

 ## 2.3 Composantes principales PC1 et PC2

barplot(-vecteurs\_propres[,1],ylab = 'Contribution',xlab = 'Variables',names.arg = names(data),axes = TRUE)  
title(main="Contributions des variables au PC1")

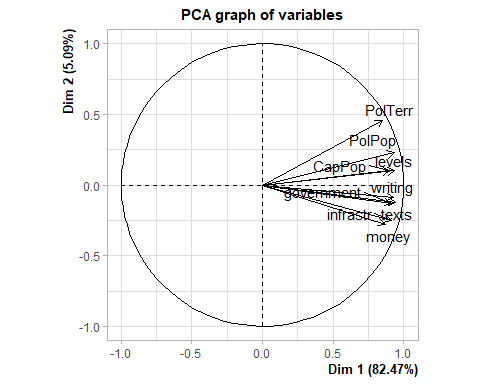
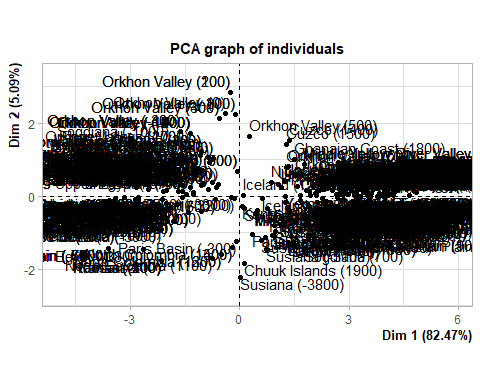


barplot(-vecteurs\_propres[,2],ylab = 'Contribution',xlab = 'Variables',names.arg = names(data),axes = TRUE)  
title(main="Contributions des variables au PC2")



## 2.4 Avec FactoMineR

library(FactoMineR)  
pca <- PCA(data, scale.unit = TRUE, ncp = 11, graph = TRUE)

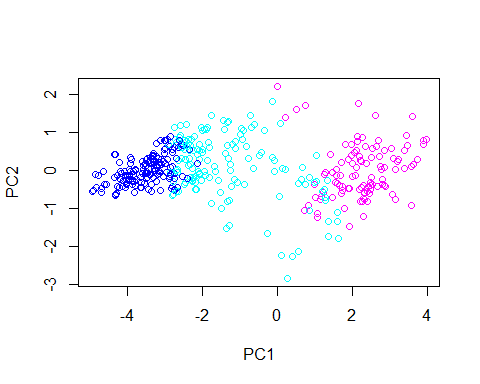


# 3 k-means

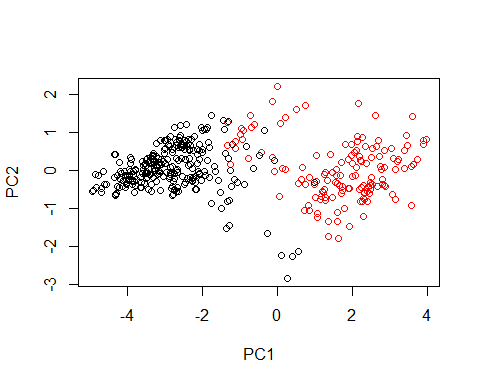
On a vu sur l’ACP qu’on pouvait distinguer environ deux groupes. On essaie des k-means :

## 3.1 Projection

kmeans.result = kmeans(data,3)  
plot(C[,1:2],type="p",xlab='PC1',ylab='PC2',col = kmeans.result$cluster+3)

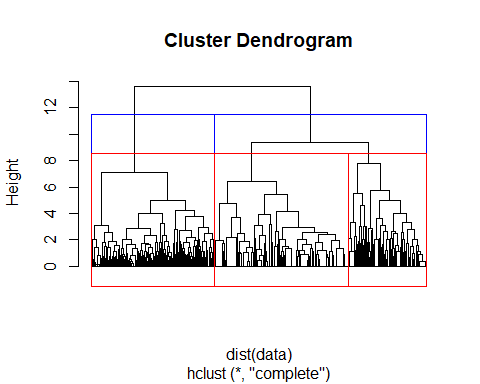


kmeans.result = kmeans(data,2)  
plot(C[,1:2],type="p",xlab='PC1',ylab='PC2',col =kmeans.result$cluster)

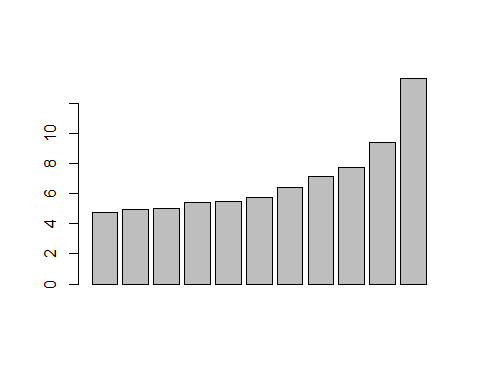


Pour k = 3, on a un groupe qui contient les points dispersés du milieu. ## 3.2 CAH Pour savoir quelle classification est la plus pertinente, essayons un CAH :

hc <- hclust(dist(data))  
plot(hc,hang=-1,labels = FALSE)  
rect.hclust(hc,k=2,border = 4)  
rect.hclust(hc,k=3)



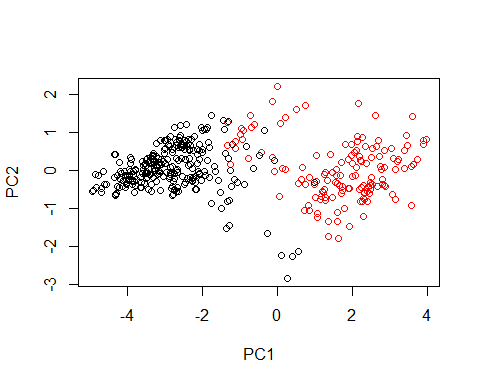
barplot(hc$height[(length(hc$height)-10):(length(hc$height))])



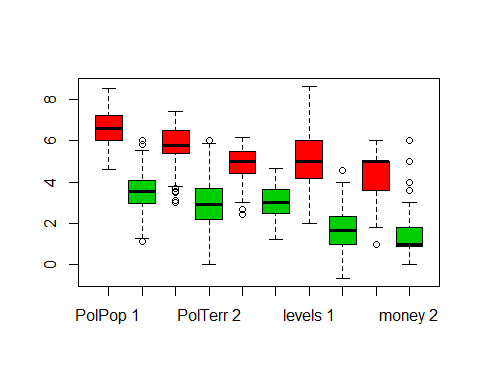
## 3.3 Comparaisons des deux clusters

Une fois les groupes réunis, il serait intéressant de comprendre ce qui les distingue, les deux dimensions n’y suffisent pas entièrement :

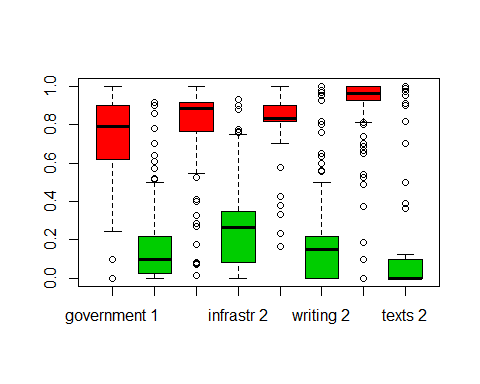
# Kmeans pour k choisi  
k = 2   
kmeans.result = kmeans(data,k)  
plot(C[,1:2],type="p",xlab='PC1',ylab='PC2',col = kmeans.result$cluster)



# Affichage des grandes variables  
data.grand = subset(data,select = c(PolPop,PolTerr,CapPop,levels,money))  
data.grand.all = list()  
for(nom in names(data.grand))  
{  
 for(i in 1:k)  
 {  
 data.grand.all[[paste(nom,i)]] <- data.grand[kmeans.result$cluster==i,nom]  
 }  
}  
boxplot(data.grand.all,col = rep(c(2:(k+1)),length(data.grand)))



# Affichage des petites variables  
data.petit = subset(data,select = c(government,infrastr,writing,texts))  
data.petit.all = list()  
for(nom in names(data.petit))  
{  
 for(i in 1:k)  
 {  
 data.petit.all[[paste(nom,i)]] <- data.petit[kmeans.result$cluster==i,nom]  
 }  
}  
boxplot(data.petit.all,col = rep(c(2:(k+1)),length(data.petit)))



# 4. Analyse sur la base Morale

On utilise les nouveaux groupes ppur tenter de classifier avec des forêts sur les variables morales

detach(data)

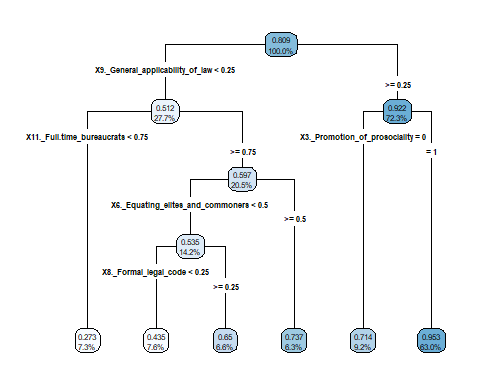
## 4.1 Création des données

dfm <- read.csv('../databases/morale\_index\_inaxe.csv',sep=',')  
  
# Mettre l'Index en index et retirer SPC1  
datam <- subset(dfm, select=-c(Index,Time,NGA,sum))  
rownames(datam) <- dfm$Index  
attach(datam)  
  
# On associe ensuite à chaque individu dans la base morale son cluster (1 ou 2) correspondant  
rreess <- kmeans.result$cluster  
datam$cluster <- (rreess[(dfn$Index)%in%(dfm$Index)] ==1)\*1  
  
head(datam)

## X1.\_Moralistic\_punishment X2.\_Moralizing\_norms  
## Upper Egypt (-4400) 0 0  
## Upper Egypt (-4300) 0 0  
## Upper Egypt (-4200) 0 0  
## Upper Egypt (-4100) 0 0  
## Upper Egypt (-4000) 0 0  
## Upper Egypt (-3900) 0 0  
## X3.\_Promotion\_of\_prosociality  
## Upper Egypt (-4400) 0  
## Upper Egypt (-4300) 0  
## Upper Egypt (-4200) 0  
## Upper Egypt (-4100) 0  
## Upper Egypt (-4000) 0  
## Upper Egypt (-3900) 0  
## X4.\_Omniscient\_supernatural\_beings X5.\_Rulers\_not\_gods  
## Upper Egypt (-4400) 0 0  
## Upper Egypt (-4300) 0 0  
## Upper Egypt (-4200) 0 0  
## Upper Egypt (-4100) 0 0  
## Upper Egypt (-4000) 0 0  
## Upper Egypt (-3900) 0 0  
## X6.\_Equating\_elites\_and\_commoners  
## Upper Egypt (-4400) 0  
## Upper Egypt (-4300) 0  
## Upper Egypt (-4200) 0  
## Upper Egypt (-4100) 0  
## Upper Egypt (-4000) 0  
## Upper Egypt (-3900) 0  
## X7.\_Equating\_rulers\_and\_commoners X8.\_Formal\_legal\_code  
## Upper Egypt (-4400) 0 0  
## Upper Egypt (-4300) 0 0  
## Upper Egypt (-4200) 0 0  
## Upper Egypt (-4100) 0 0  
## Upper Egypt (-4000) 0 0  
## Upper Egypt (-3900) 0 0  
## X9.\_General\_applicability\_of\_law  
## Upper Egypt (-4400) 0  
## Upper Egypt (-4300) 0  
## Upper Egypt (-4200) 0  
## Upper Egypt (-4100) 0  
## Upper Egypt (-4000) 0  
## Upper Egypt (-3900) 0  
## X10.\_Constraint\_on\_executive X11.\_Full.time\_bureaucrats  
## Upper Egypt (-4400) 0 0  
## Upper Egypt (-4300) 0 0  
## Upper Egypt (-4200) 0 0  
## Upper Egypt (-4100) 0 0  
## Upper Egypt (-4000) 0 0  
## Upper Egypt (-3900) 0 0  
## X12.\_Impeachment cluster  
## Upper Egypt (-4400) 0 0  
## Upper Egypt (-4300) 0 0  
## Upper Egypt (-4200) 0 0  
## Upper Egypt (-4100) 0 0  
## Upper Egypt (-4000) 0 0  
## Upper Egypt (-3900) 0 0

## 4.2 Arbre

library(rpart)  
arbre=rpart(datam$cluster~.,datam)  
# summary(arbre)  
# print(arbre)  
library(rpart.plot)  
rpart.plot(arbre, type=4, digits=3)

 ## 4.3 Régression ## 4.3.1 Modèle

res <- glm(cluster~.,family=binomial,data=datam)  
summary(res)

##   
## Call:  
## glm(formula = cluster ~ ., family = binomial, data = datam)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.4475 0.0581 0.1950 0.5118 1.6117   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.9801 0.4426 -2.214 0.026803 \*   
## X1.\_Moralistic\_punishment -2.2834 0.8484 -2.691 0.007117 \*\*   
## X2.\_Moralizing\_norms -0.5112 0.6930 -0.738 0.460695   
## X3.\_Promotion\_of\_prosociality 1.4368 0.5674 2.532 0.011331 \*   
## X4.\_Omniscient\_supernatural\_beings 2.4379 1.2431 1.961 0.049853 \*   
## X5.\_Rulers\_not\_gods -0.4511 0.6884 -0.655 0.512308   
## X6.\_Equating\_elites\_and\_commoners -0.2055 0.8437 -0.244 0.807553   
## X7.\_Equating\_rulers\_and\_commoners 2.1845 0.9698 2.253 0.024285 \*   
## X8.\_Formal\_legal\_code 0.9356 0.6282 1.489 0.136373   
## X9.\_General\_applicability\_of\_law 2.4217 0.7191 3.368 0.000758 \*\*\*  
## X10.\_Constraint\_on\_executive -1.5355 0.7283 -2.108 0.035000 \*   
## X11.\_Full.time\_bureaucrats 0.6503 0.5234 1.243 0.214009   
## X12.\_Impeachment 1.0425 0.8235 1.266 0.205531   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 295.89 on 302 degrees of freedom  
## Residual deviance: 190.92 on 290 degrees of freedom  
## AIC: 216.92  
##   
## Number of Fisher Scoring iterations: 7

### 4.3.2 Odds Ratios

exp(res$coefficients)

## (Intercept) X1.\_Moralistic\_punishment   
## 0.3752810 0.1019414   
## X2.\_Moralizing\_norms X3.\_Promotion\_of\_prosociality   
## 0.5997592 4.2074037   
## X4.\_Omniscient\_supernatural\_beings X5.\_Rulers\_not\_gods   
## 11.4491608 0.6369282   
## X6.\_Equating\_elites\_and\_commoners X7.\_Equating\_rulers\_and\_commoners   
## 0.8142326 8.8862247   
## X8.\_Formal\_legal\_code X9.\_General\_applicability\_of\_law   
## 2.5488605 11.2649718   
## X10.\_Constraint\_on\_executive X11.\_Full.time\_bureaucrats   
## 0.2153397 1.9161609   
## X12.\_Impeachment   
## 2.8362184

### 4.3.3 Pertinence

res0 <- glm(cluster~1,family=binomial,data=datam)  
anova(res0,res,test="Chisq")

## Analysis of Deviance Table  
##   
## Model 1: cluster ~ 1  
## Model 2: cluster ~ X1.\_Moralistic\_punishment + X2.\_Moralizing\_norms +   
## X3.\_Promotion\_of\_prosociality + X4.\_Omniscient\_supernatural\_beings +   
## X5.\_Rulers\_not\_gods + X6.\_Equating\_elites\_and\_commoners +   
## X7.\_Equating\_rulers\_and\_commoners + X8.\_Formal\_legal\_code +   
## X9.\_General\_applicability\_of\_law + X10.\_Constraint\_on\_executive +   
## X11.\_Full.time\_bureaucrats + X12.\_Impeachment  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 302 295.89   
## 2 290 190.92 12 104.97 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### 4.3.4 AIC

library(MASS)  
res\_AIC <- step(res)

## Start: AIC=216.92  
## cluster ~ X1.\_Moralistic\_punishment + X2.\_Moralizing\_norms +   
## X3.\_Promotion\_of\_prosociality + X4.\_Omniscient\_supernatural\_beings +   
## X5.\_Rulers\_not\_gods + X6.\_Equating\_elites\_and\_commoners +   
## X7.\_Equating\_rulers\_and\_commoners + X8.\_Formal\_legal\_code +   
## X9.\_General\_applicability\_of\_law + X10.\_Constraint\_on\_executive +   
## X11.\_Full.time\_bureaucrats + X12.\_Impeachment  
##   
## Df Deviance AIC  
## - X6.\_Equating\_elites\_and\_commoners 1 190.98 214.98  
## - X5.\_Rulers\_not\_gods 1 191.35 215.35  
## - X2.\_Moralizing\_norms 1 191.48 215.48  
## - X11.\_Full.time\_bureaucrats 1 192.48 216.48  
## - X12.\_Impeachment 1 192.60 216.60  
## <none> 190.92 216.92  
## - X8.\_Formal\_legal\_code 1 193.14 217.14  
## - X10.\_Constraint\_on\_executive 1 195.71 219.71  
## - X7.\_Equating\_rulers\_and\_commoners 1 196.23 220.23  
## - X4.\_Omniscient\_supernatural\_beings 1 196.53 220.53  
## - X3.\_Promotion\_of\_prosociality 1 197.63 221.63  
## - X1.\_Moralistic\_punishment 1 197.65 221.65  
## - X9.\_General\_applicability\_of\_law 1 203.75 227.75  
##   
## Step: AIC=214.98  
## cluster ~ X1.\_Moralistic\_punishment + X2.\_Moralizing\_norms +   
## X3.\_Promotion\_of\_prosociality + X4.\_Omniscient\_supernatural\_beings +   
## X5.\_Rulers\_not\_gods + X7.\_Equating\_rulers\_and\_commoners +   
## X8.\_Formal\_legal\_code + X9.\_General\_applicability\_of\_law +   
## X10.\_Constraint\_on\_executive + X11.\_Full.time\_bureaucrats +   
## X12.\_Impeachment  
##   
## Df Deviance AIC  
## - X5.\_Rulers\_not\_gods 1 191.36 213.36  
## - X2.\_Moralizing\_norms 1 191.75 213.75  
## - X11.\_Full.time\_bureaucrats 1 192.50 214.50  
## <none> 190.98 214.98  
## - X12.\_Impeachment 1 192.99 214.99  
## - X8.\_Formal\_legal\_code 1 193.20 215.20  
## - X4.\_Omniscient\_supernatural\_beings 1 196.54 218.54  
## - X10.\_Constraint\_on\_executive 1 196.66 218.66  
## - X1.\_Moralistic\_punishment 1 197.65 219.65  
## - X3.\_Promotion\_of\_prosociality 1 197.99 219.99  
## - X7.\_Equating\_rulers\_and\_commoners 1 198.72 220.72  
## - X9.\_General\_applicability\_of\_law 1 205.06 227.06  
##   
## Step: AIC=213.36  
## cluster ~ X1.\_Moralistic\_punishment + X2.\_Moralizing\_norms +   
## X3.\_Promotion\_of\_prosociality + X4.\_Omniscient\_supernatural\_beings +   
## X7.\_Equating\_rulers\_and\_commoners + X8.\_Formal\_legal\_code +   
## X9.\_General\_applicability\_of\_law + X10.\_Constraint\_on\_executive +   
## X11.\_Full.time\_bureaucrats + X12.\_Impeachment  
##   
## Df Deviance AIC  
## - X2.\_Moralizing\_norms 1 192.73 212.73  
## - X12.\_Impeachment 1 193.03 213.03  
## - X8.\_Formal\_legal\_code 1 193.32 213.32  
## - X11.\_Full.time\_bureaucrats 1 193.33 213.33  
## <none> 191.36 213.36  
## - X4.\_Omniscient\_supernatural\_beings 1 196.86 216.86  
## - X10.\_Constraint\_on\_executive 1 197.11 217.11  
## - X3.\_Promotion\_of\_prosociality 1 198.10 218.10  
## - X1.\_Moralistic\_punishment 1 198.12 218.12  
## - X7.\_Equating\_rulers\_and\_commoners 1 200.00 220.00  
## - X9.\_General\_applicability\_of\_law 1 207.63 227.63  
##   
## Step: AIC=212.73  
## cluster ~ X1.\_Moralistic\_punishment + X3.\_Promotion\_of\_prosociality +   
## X4.\_Omniscient\_supernatural\_beings + X7.\_Equating\_rulers\_and\_commoners +   
## X8.\_Formal\_legal\_code + X9.\_General\_applicability\_of\_law +   
## X10.\_Constraint\_on\_executive + X11.\_Full.time\_bureaucrats +   
## X12.\_Impeachment  
##   
## Df Deviance AIC  
## - X8.\_Formal\_legal\_code 1 193.72 211.72  
## - X12.\_Impeachment 1 194.20 212.20  
## - X11.\_Full.time\_bureaucrats 1 194.25 212.25  
## <none> 192.73 212.73  
## - X3.\_Promotion\_of\_prosociality 1 198.88 216.88  
## - X10.\_Constraint\_on\_executive 1 199.16 217.16  
## - X1.\_Moralistic\_punishment 1 199.32 217.32  
## - X4.\_Omniscient\_supernatural\_beings 1 199.47 217.47  
## - X7.\_Equating\_rulers\_and\_commoners 1 200.47 218.47  
## - X9.\_General\_applicability\_of\_law 1 207.69 225.69  
##   
## Step: AIC=211.72  
## cluster ~ X1.\_Moralistic\_punishment + X3.\_Promotion\_of\_prosociality +   
## X4.\_Omniscient\_supernatural\_beings + X7.\_Equating\_rulers\_and\_commoners +   
## X9.\_General\_applicability\_of\_law + X10.\_Constraint\_on\_executive +   
## X11.\_Full.time\_bureaucrats + X12.\_Impeachment  
##   
## Df Deviance AIC  
## <none> 193.72 211.72  
## - X12.\_Impeachment 1 196.24 212.24  
## - X11.\_Full.time\_bureaucrats 1 196.39 212.39  
## - X1.\_Moralistic\_punishment 1 200.10 216.10  
## - X10.\_Constraint\_on\_executive 1 200.20 216.20  
## - X4.\_Omniscient\_supernatural\_beings 1 200.43 216.43  
## - X7.\_Equating\_rulers\_and\_commoners 1 200.51 216.51  
## - X3.\_Promotion\_of\_prosociality 1 203.89 219.89  
## - X9.\_General\_applicability\_of\_law 1 210.74 226.74

summary(res\_AIC)

##   
## Call:  
## glm(formula = cluster ~ X1.\_Moralistic\_punishment + X3.\_Promotion\_of\_prosociality +   
## X4.\_Omniscient\_supernatural\_beings + X7.\_Equating\_rulers\_and\_commoners +   
## X9.\_General\_applicability\_of\_law + X10.\_Constraint\_on\_executive +   
## X11.\_Full.time\_bureaucrats + X12.\_Impeachment, family = binomial,   
## data = datam)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.5239 0.0701 0.2327 0.5169 1.6513   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.0680 0.4301 -2.483 0.013030 \*   
## X1.\_Moralistic\_punishment -2.2065 0.8322 -2.651 0.008014 \*\*   
## X3.\_Promotion\_of\_prosociality 1.4964 0.4820 3.104 0.001907 \*\*   
## X4.\_Omniscient\_supernatural\_beings 2.6113 1.2348 2.115 0.034455 \*   
## X7.\_Equating\_rulers\_and\_commoners 1.4504 0.5903 2.457 0.014006 \*   
## X9.\_General\_applicability\_of\_law 2.3946 0.6167 3.883 0.000103 \*\*\*  
## X10.\_Constraint\_on\_executive -1.6501 0.6786 -2.432 0.015032 \*   
## X11.\_Full.time\_bureaucrats 0.7727 0.4756 1.625 0.104248   
## X12.\_Impeachment 1.1235 0.7434 1.511 0.130725   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 295.89 on 302 degrees of freedom  
## Residual deviance: 193.72 on 294 degrees of freedom  
## AIC: 211.72  
##   
## Number of Fisher Scoring iterations: 7

anova(res\_AIC,res,test="Chisq")

## Analysis of Deviance Table  
##   
## Model 1: cluster ~ X1.\_Moralistic\_punishment + X3.\_Promotion\_of\_prosociality +   
## X4.\_Omniscient\_supernatural\_beings + X7.\_Equating\_rulers\_and\_commoners +   
## X9.\_General\_applicability\_of\_law + X10.\_Constraint\_on\_executive +   
## X11.\_Full.time\_bureaucrats + X12.\_Impeachment  
## Model 2: cluster ~ X1.\_Moralistic\_punishment + X2.\_Moralizing\_norms +   
## X3.\_Promotion\_of\_prosociality + X4.\_Omniscient\_supernatural\_beings +   
## X5.\_Rulers\_not\_gods + X6.\_Equating\_elites\_and\_commoners +   
## X7.\_Equating\_rulers\_and\_commoners + X8.\_Formal\_legal\_code +   
## X9.\_General\_applicability\_of\_law + X10.\_Constraint\_on\_executive +   
## X11.\_Full.time\_bureaucrats + X12.\_Impeachment  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)  
## 1 294 193.72   
## 2 290 190.92 4 2.795 0.5927

par(mfrow=c(2,2))  
plot(res\_AIC)

