Web Trafic Time Series Forecasting

library(factoextra)

## Loading required package: ggplot2

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

library(fastDummies)  
library(psych)

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

library(GPArotation)  
library(datasets)  
library(readr)

data <- read.csv("C:/Users/HP/OneDrive/Desktop/africa.csv")  
  
data <- na.omit(data)

# Keep/drop columns (TYPE IN THE COL NO.)

data <- data[-c(18)]

# Dummy coding

# data <- dummy\_cols(data, remove\_most\_frequent\_dummy = TRUE, remove\_selected\_columns = TRUE)

# scale the data

data <- scale(data)

# a. Kaiser-Meyer-Olkin Measure of Sampling Adequacy

KMO(data)

## Kaiser-Meyer-Olkin factor adequacy  
## Call: KMO(r = data)  
## Overall MSA = 0.63  
## MSA for each item =   
## pop emp emp\_to\_pop\_ratio   
## 0.68 0.66 0.41   
## hc ccon cda   
## 0.69 0.73 0.71   
## cn ck ctfp   
## 0.90 0.86 0.59   
## cwtfp rconna rdana   
## 0.53 0.72 0.71   
## rnna rkna rtfpna   
## 0.86 0.90 0.48   
## rwtfpna labsh delta   
## 0.67 0.72 0.62   
## xr pl\_con pl\_da   
## 0.48 0.79 0.79   
## pl\_gdpo csh\_c csh\_i   
## 0.87 0.31 0.35   
## csh\_g csh\_x csh\_m   
## 0.24 0.37 0.29   
## csh\_r pl\_c pl\_i   
## 0.11 0.82 0.73   
## pl\_g pl\_x pl\_m   
## 0.72 0.85 0.86   
## pl\_n total excl\_energy   
## 0.64 0.62 0.62   
## energy metals\_minerals forestry   
## 0.61 0.62 0.50   
## agriculture fish total\_change   
## 0.65 0.70 0.67   
## excl\_energy\_change energy\_change metals\_minerals\_change   
## 0.45 0.58 0.38   
## forestry\_change agriculture\_change fish\_change   
## 0.38 0.40 0.41   
## growthbucket   
## 0.72

# b. Bartlett’s Test of Sphericity

cortest.bartlett(data, n=NULL, diag=TRUE)

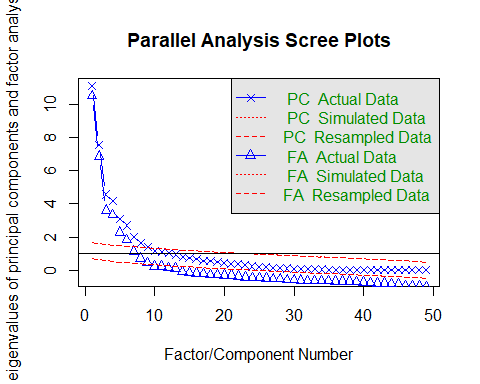
## R was not square, finding R from data

## $chisq  
## [1] 64527.35  
##   
## $p.value  
## [1] 0  
##   
## $df  
## [1] 1176

# Parallel Analysis (TYPE IN: factoring method)

fa.parallel(data, fm="pa", fa="both", n.iter=100)

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :  
## The estimated weights for the factor scores are probably incorrect. Try a  
## different factor score estimation method.

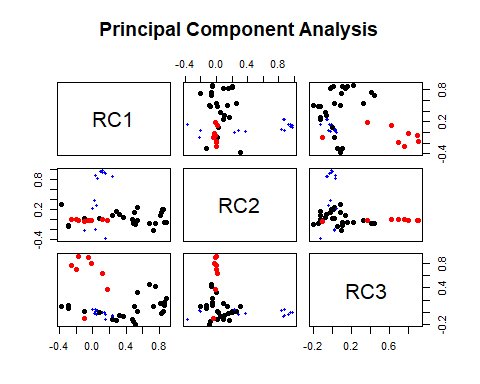


## Parallel analysis suggests that the number of factors = 9 and the number of components = 9

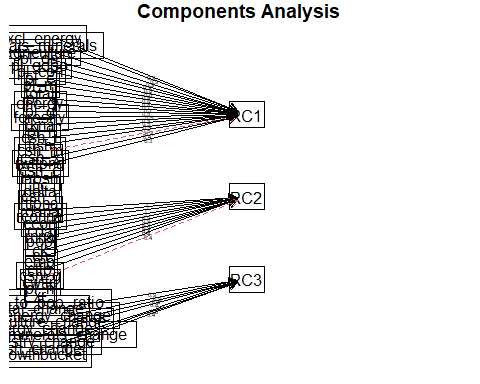
# Pricipal Components Analysis (TYPE IN: pc no./rotation method)  
  
fit <- principal(data, 3, rotate="varimax")  
print(fit$loadings, cutoff=.3)

##   
## Loadings:  
## RC1 RC2 RC3   
## pop 0.873   
## emp 0.832   
## emp\_to\_pop\_ratio   
## hc   
## ccon 0.951   
## cda 0.950   
## cn 0.858   
## ck 0.870   
## ctfp 0.378   
## cwtfp   
## rconna 0.970   
## rdana 0.975   
## rnna 0.927   
## rkna 0.512   
## rtfpna   
## rwtfpna 0.319   
## labsh   
## delta   
## xr   
## pl\_con 0.841   
## pl\_da 0.852   
## pl\_gdpo 0.849   
## csh\_c   
## csh\_i 0.493   
## csh\_g -0.372   
## csh\_x 0.364   
## csh\_m -0.378 0.305   
## csh\_r   
## pl\_c 0.829   
## pl\_i 0.504   
## pl\_g 0.542   
## pl\_x 0.731   
## pl\_m 0.819   
## pl\_n   
## total 0.751 0.415  
## excl\_energy 0.884   
## energy 0.686 0.449  
## metals\_minerals 0.860   
## forestry 0.534 0.328  
## agriculture 0.857   
## fish 0.480   
## total\_change 0.910  
## excl\_energy\_change 0.898  
## energy\_change 0.764  
## metals\_minerals\_change 0.692  
## forestry\_change 0.627  
## agriculture\_change 0.800  
## fish\_change 0.372  
## growthbucket   
##   
## RC1 RC2 RC3  
## SS loadings 10.036 8.493 4.656  
## Proportion Var 0.205 0.173 0.095  
## Cumulative Var 0.205 0.378 0.473

factor.plot(fit)



fa.diagram(fit)



# Factor Analysis (TYPE IN: factor no./rotation method/factoring method)

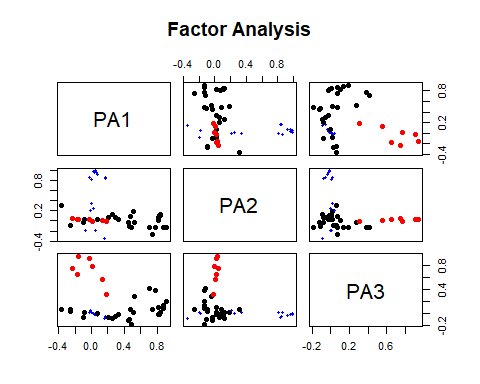
fit <- fa(data, 3, rotate="promax", fm="pa")

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :  
## The estimated weights for the factor scores are probably incorrect. Try a  
## different factor score estimation method.

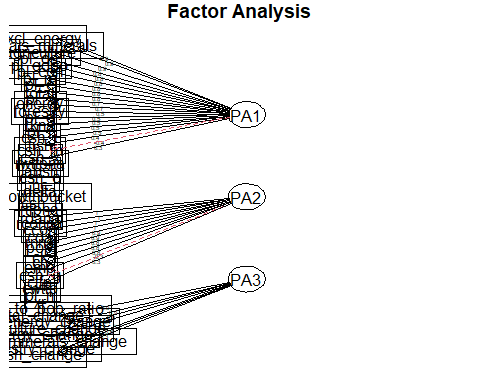
print(fit$loadings, cutoff=.3)

##   
## Loadings:  
## PA1 PA2 PA3   
## pop 0.861   
## emp 0.812   
## emp\_to\_pop\_ratio   
## hc   
## ccon 0.955   
## cda 0.954   
## cn 0.835   
## ck 0.849   
## ctfp 0.339   
## cwtfp   
## rconna 0.981   
## rdana 0.986   
## rnna 0.922   
## rkna 0.487   
## rtfpna   
## rwtfpna   
## labsh   
## delta   
## xr   
## pl\_con 0.829   
## pl\_da 0.843   
## pl\_gdpo 0.838   
## csh\_c   
## csh\_i 0.463   
## csh\_g -0.347   
## csh\_x 0.330   
## csh\_m -0.369 0.312   
## csh\_r   
## pl\_c 0.815   
## pl\_i 0.479   
## pl\_g 0.505   
## pl\_x 0.742   
## pl\_m 0.819   
## pl\_n   
## total 0.775 0.381  
## excl\_energy 0.911   
## energy 0.705 0.411  
## metals\_minerals 0.878   
## forestry 0.521   
## agriculture 0.872   
## fish 0.448   
## total\_change 0.944  
## excl\_energy\_change 0.908  
## energy\_change 0.745  
## metals\_minerals\_change 0.653  
## forestry\_change 0.557  
## agriculture\_change 0.770  
## fish\_change 0.310  
## growthbucket   
##   
## PA1 PA2 PA3  
## SS loadings 9.706 8.240 4.283  
## Proportion Var 0.198 0.168 0.087  
## Cumulative Var 0.198 0.366 0.454

factor.plot(fit)

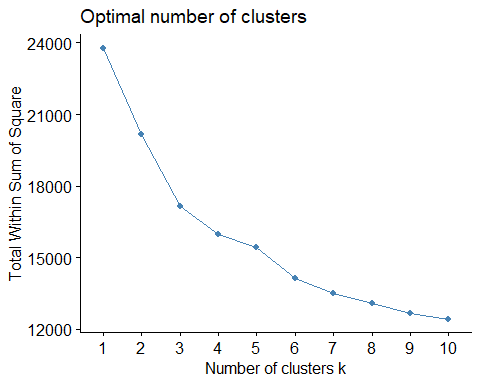


fa.diagram(fit)



# Cluster Analysis based on PCA/FA scores (TYPE IN: cluster no./method)

fviz\_nbclust(data, kmeans, method = "wss")



cluster <- kmeans(data, 3, nstart = 24)  
  
fviz\_cluster(cluster, data = data, ellipse.type = "euclid", star.plot = TRUE, repel = TRUE, ggtheme = theme\_minimal())

## Warning: ggrepel: 440 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps

