

Wkazniki społeczne wojewodztw

Robert Kiermasz

15 października 2017

Dane i standaryzacja

```
setwd("/home/cengupan/wojewodztwa/")
wojewodztwa<-read.csv("wojewodztwa2016.csv", sep=";", dec="," ,stringsAsFactors = FALSE)
rownames(wojewodztwa)<-wojewodztwa[,1]
wojewodztwa<-wojewodztwa[,-1]

####średnia, odchylenie, standaryzacja
srednie<-apply(wojewodztwa,2,mean)
odchylenia<-apply(wojewodztwa,2,sd)

for(i in 1:dim(wojewodztwa)[1]){
  wojewodztwa[i,]<- (wojewodztwa[i,]-srednie)/odchylenia
}
```

korelacje

```
korelacje<-round(cor(wojewodztwa),2)
```

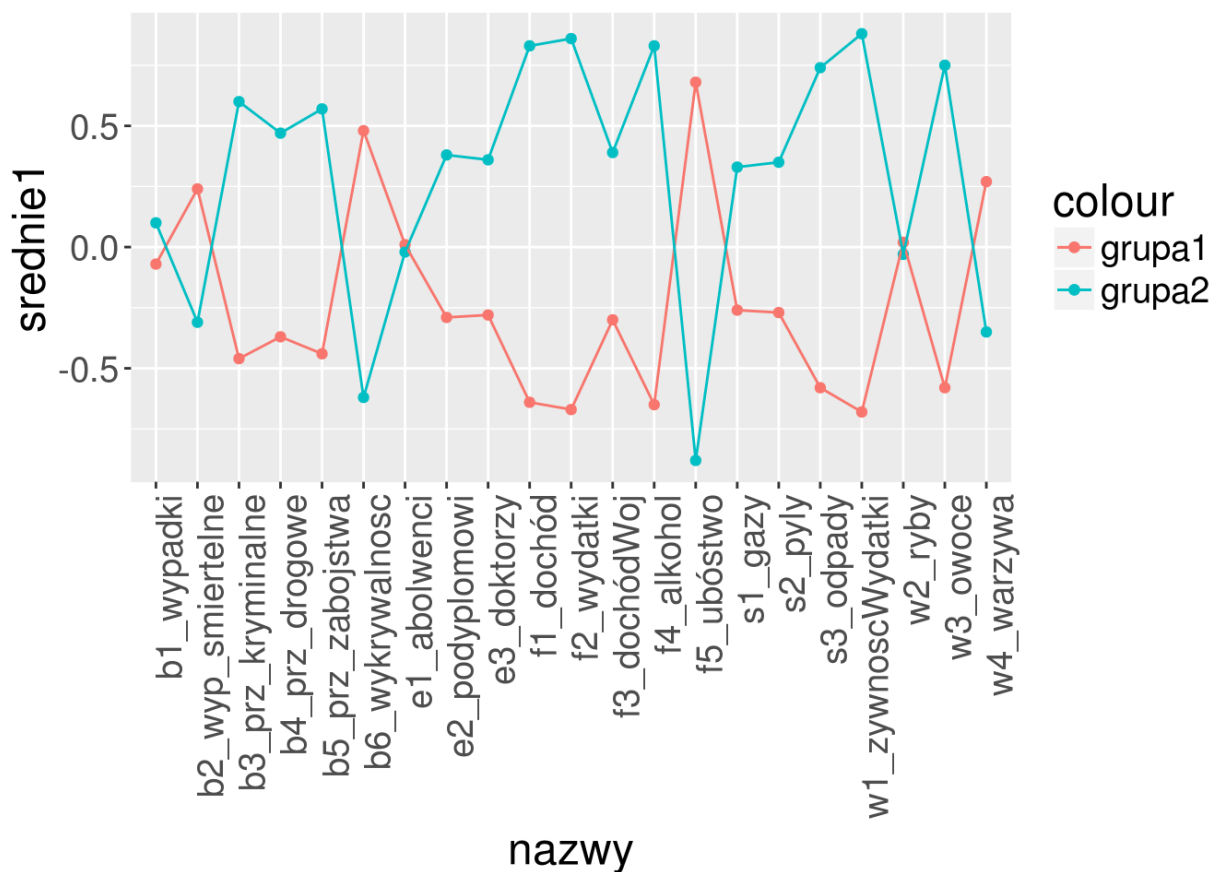
grupowanie k, wykresy średnich

tworzenie tabeli srednich

wykres średnich

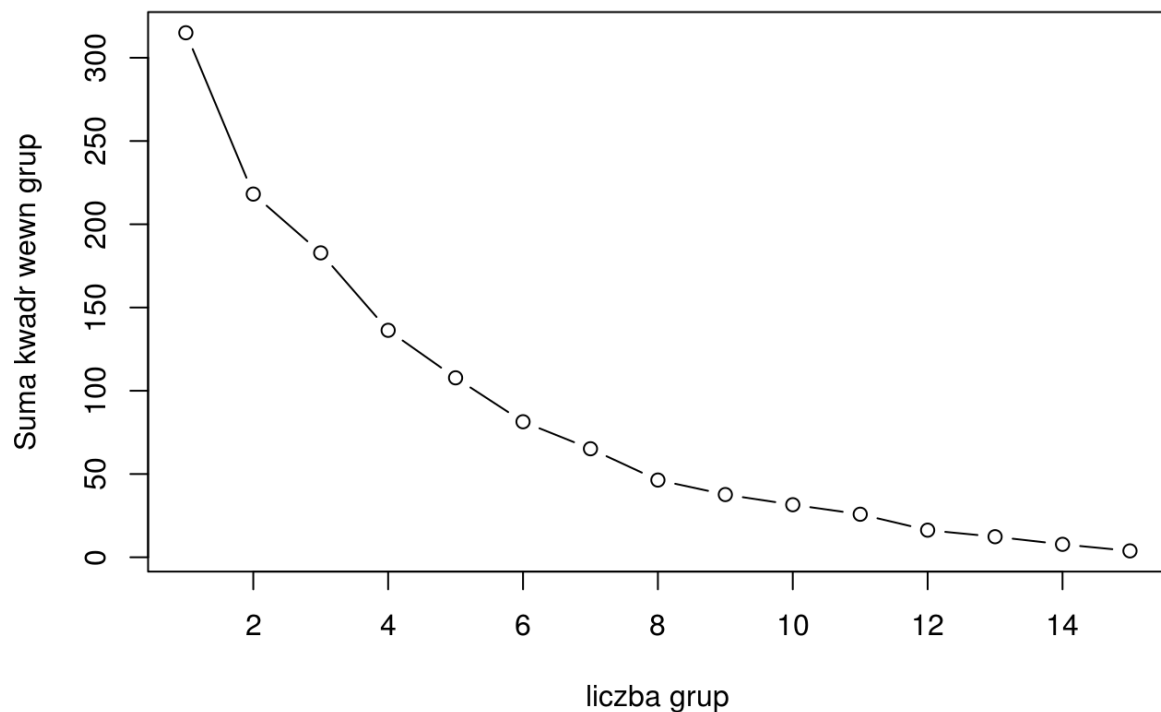
```
wyniki<-kmeans(województwa,2)
województwa$grupa<-wyniki$cluster
srednie1<-round(apply(województwa[which(wyniki$cluster==1),-22],2,mean),2)
srednie2<-round(apply(województwa[which(wyniki$cluster==2),-22],2,mean),2)
nazwy<-names(srednie1)
tabelaSrednie<-as.data.frame(cbind(srednie1,srednie2))
tabelaSrednie<-cbind(tabelaSrednie,nazwy)
tabelaSrednie[,3]<-as.character(tabelaSrednie[,3])
```

```
library("ggplot2")
ggplot(data=tabelaSrednie, aes(nazwy))+
  geom_point(aes(y=srednie1, colour="grupa1")) +
  geom_point(aes(y=srednie2, colour="grupa2"))+
  geom_line(aes(y=srednie1, group=1, colour="grupa1")) +
  geom_line(aes(y=srednie2, group=2, colour="grupa2")) +
  theme(text = element_text(size=18),
        axis.text.x = element_text(angle = 90, hjust = 1))
```



wykres osypiska

```
wojewodztwa<-wojewodztwa[,-22]
wss <- (nrow(wojewodztwa)-1)*sum(apply(wojewodztwa,2,var))
for (i in 2:15) wss[i] <- sum(kmeans(wojewodztwa, centers = i)$withinss)
plot(1:15, wss, type = "b", xlab = "liczba grup", ylab = "Suma kwadr wewn
grup")
```

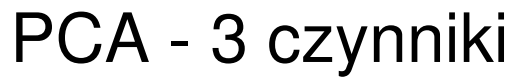


PCA - 2 czynniki

```
library(psych)
fit<- principal(wojewodztwa, nfactors=2, rotate="varimax")
fit$Structure
```

```
##
## Loadings:
##          RC1    RC2
## f1_dochód    0.518  0.788
## f2_wydatki    0.442  0.845
## f3_dochódWoj  0.942  0.172
## f4_alkohol    0.117  0.928
## f5_ubóstwo   -0.152 -0.914
## w1_zywnoscWydatki      0.885
## w2_ryby        0.234
## w3_owoce        0.789
## w4_warzywa    -0.532 -0.187
## e1_abolwenci    0.711 -0.141
## e2_podyplomowi  0.826  0.243
## e3_doktorzy     0.905  0.175
## b1_wypadki     -0.403  0.276
## b2_wyp_smiertelne -0.297 -0.251
## b3_prz_kryminalne  0.916  0.340
## b4_prz_drogowe   0.961  0.173
## b5_prz_zabojstwa  0.835  0.272
## b6_wykrywalnosc -0.352 -0.599
## s1_gazy         0.456  0.138
## s2_pyly         0.590
## s3_odpady       0.286  0.733
##
##          RC1    RC2
## SS loadings    7.158  5.950
## Proportion Var 0.341  0.283
## Cumulative Var 0.341  0.624
```

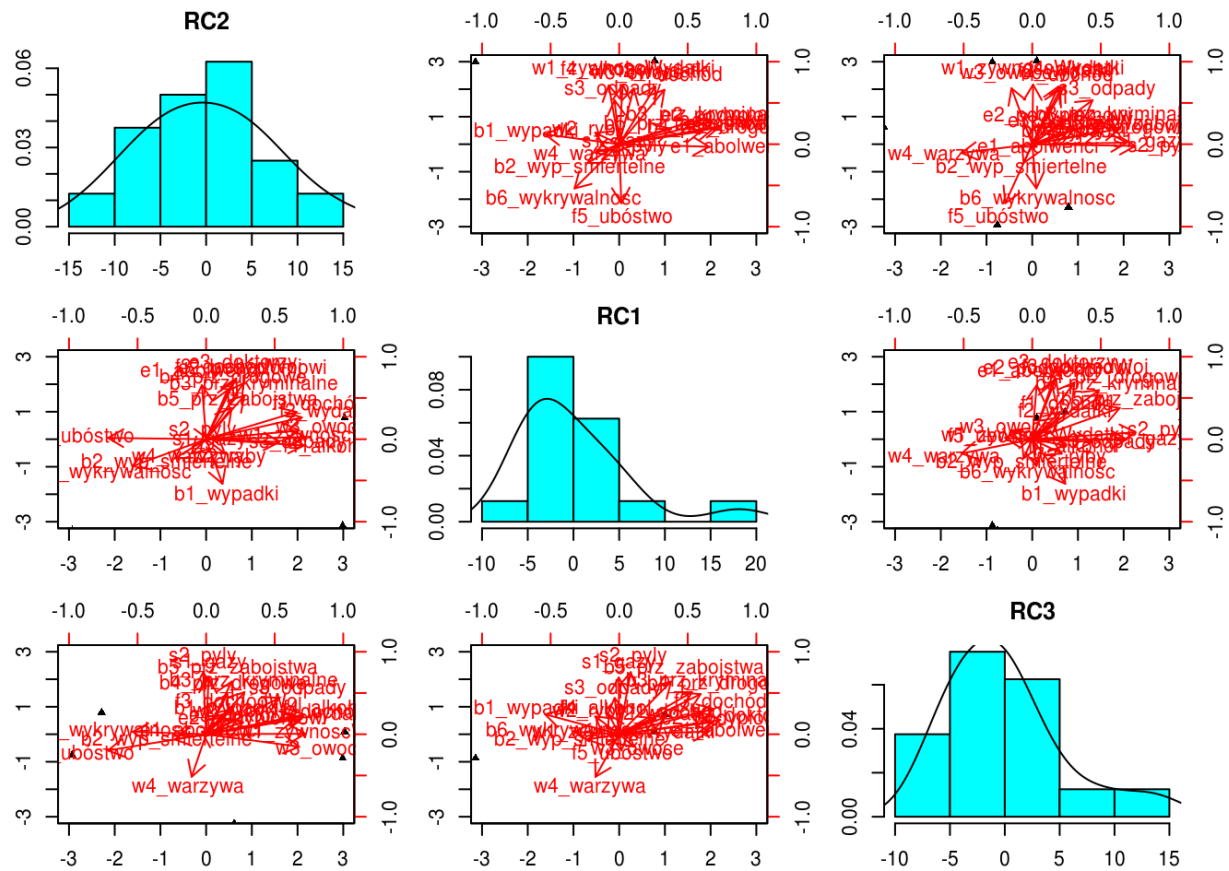
```
biplot(fit)
```



15.10.2017, 23:05

```
##
## Loadings:
##          RC2    RC1    RC3
## f1_dochód      0.824  0.413  0.254
## f2_wydatki      0.873  0.330  0.239
## f3_dochódWoj    0.239  0.861  0.385
## f4_alkohol      0.892          0.300
## f5_ubóstwo     -0.895          -0.246
## w1_zywnoscWydatki 0.904
## w2_ryby         0.172 -0.218  0.269
## w3_owoce        0.847  0.133 -0.189
## w4_warzywa     -0.132 -0.216 -0.644
## e1_abolwenci          0.819
## e2_podyplomowi   0.337  0.846  0.182
## e3_doktorzy      0.274  0.922  0.216
## b1_wypadki       0.149 -0.675  0.302
## b2_wyp_smiertelne -0.284 -0.288
## b3_prz_kryminalne 0.349  0.664  0.638
## b4_prz_drogowe   0.194  0.743  0.609
## b5_prz_zabojstwa 0.233  0.473  0.798
## b6_wykrywalnosc -0.670 -0.407
## s1_gazy          0.860
## s2_pyly          0.105  0.940
## s3_odpady        0.667          0.550
##
##          RC2    RC1    RC3
## SS loadings    6.077  5.312  4.462
## Proportion Var 0.289  0.253  0.212
## Cumulative Var 0.289  0.542  0.755
```

```
biplot(pc2o,main="",labels=paste0(1:12))
```



Analiza czynników głównych

```
fit2 <- fa(województwa,3, rotate = "varimax")
fit2$loadings
```

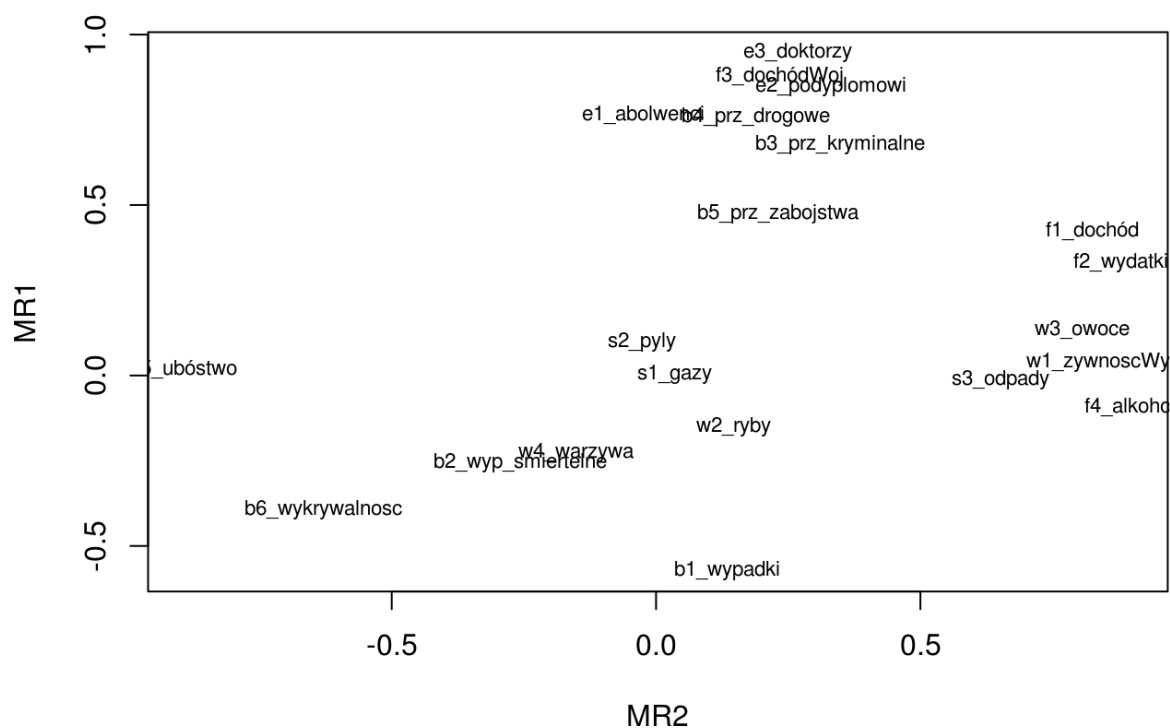
```
##
## Loadings:
##
```

	MR2	MR1	MR3
## f1_dochód	0.825	0.422	0.240
## f2_wydatki	0.879	0.332	0.234
## f3_dochódWoj	0.234	0.878	0.368
## f4_alkohol	0.896		0.299
## f5_ubóstwo	-0.889		-0.249
## w1_zywnoscWydatki	0.877		
## w2_ryby	0.146	-0.151	0.213
## w3_owoce	0.805	0.134	-0.159
## w4_warzywa	-0.152	-0.225	-0.549
## e1_abolwenci		0.764	
## e2_podyplomowi	0.331	0.846	0.166
## e3_doktorzy	0.269	0.946	0.192
## b1_wypadki	0.134	-0.573	0.234
## b2_wyp_smiertelne	-0.256	-0.255	
## b3_prz_kryminalne	0.348	0.676	0.635
## b4_prz_drogowe	0.188	0.757	0.605
## b5_prz_zabojstwa	0.231	0.473	0.804
## b6_wykrywalnosc	-0.630	-0.394	
## s1_gazy			0.825
## s2_pyly			0.963
## s3_odpady	0.651		0.512

```
##
##
```

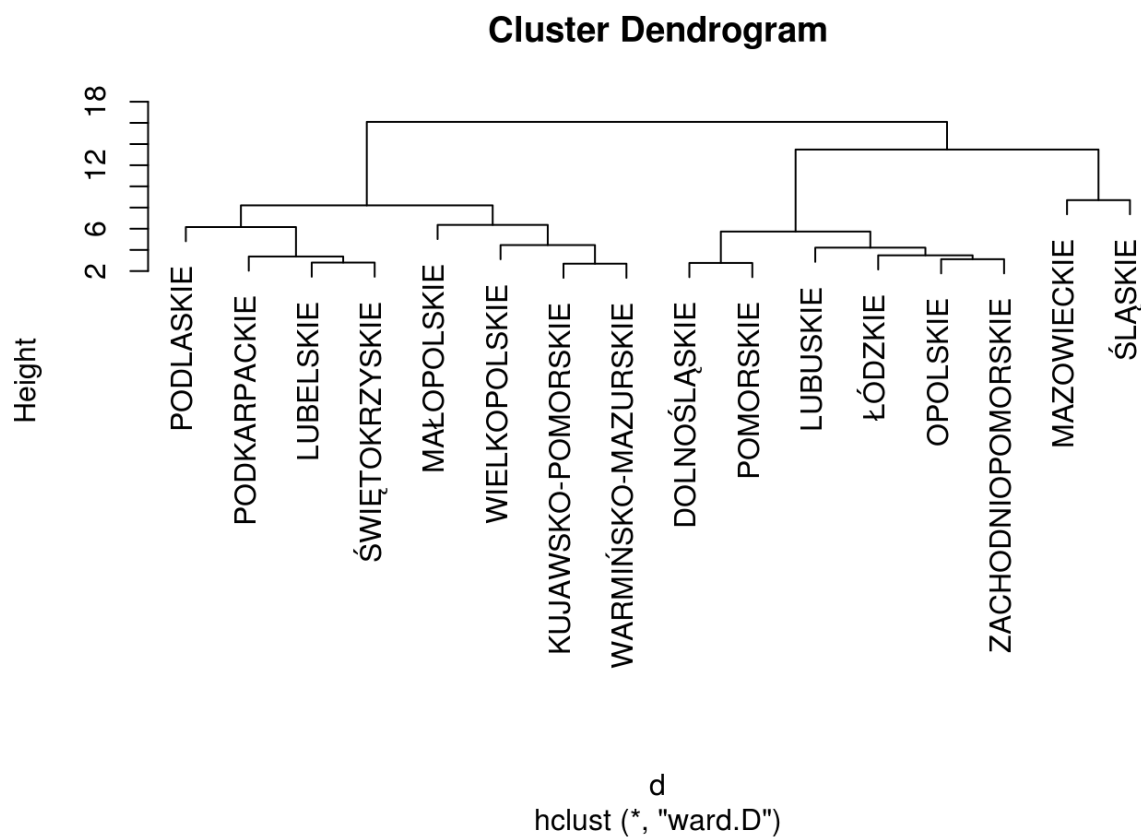
	MR2	MR1	MR3
## SS loadings	5.861	5.169	4.184
## Proportion Var	0.279	0.246	0.199
## Cumulative Var	0.279	0.525	0.725

```
load <- fit2$loadings[,1:2]
plot(load,type="n") # set up plot
text(load,labels=names(wojewodztwa),cex=.7) # add variable names
```

grupowanie metoda Warda

```
d <- dist(wojewodztwa,method = "euclidean")
fit <- hclust(d,method = "ward")
plot(fit, labels = rownames(wojewodztwa))
```



pam

```
library(cluster)
pam = pam(województwa, 4)
pam$clustering
```

```
##      DOLNOŚLĄSKIE  KUJAWSKO - POMORSKIE      LUBELSKIE
##              1              2              3
##      LUBUSKIE      ŁÓDZKIE      MAŁOPOLSKIE
##              1              1              1
##      MAZOWIECKIE      OPOLSKIE      PODKARPACKIE
##              4              2              3
##      PODLASKIE      POMORSKIE      ŚLĄSKIE
##              2              1              1
##      ŚWIĘTOKRZYSKIE  WARMIŃSKO - MAZURSKIE  WIELKOPOLSKIE
##              3              2              2
##  ZACHODNIOPOMORSKIE
##              1
```