

TEMA 1-Analiza datelor

Utilizarea metodei ACP pentru reducerea dimensionalității datelor

1. *Căutați un set de date din domeniul economic cu cel puțin 10 variabile (coloane) – pentru nota maximă, este recomandat ca setul de date să crească în dimensiuni.*

Pentru analiza ce va fi prezentată pe parcursul temei mele, am creat o bază de date in Excel care are 10 variabile și 39 de observații.

2. *Curățați setul de date, redenumiți variabilele astfel încât să fie clar de identificat și gestionați valorile lipsă și outlierii.*

- Pentru început, am afișat numărul de valori lipsă de pe fiecare coloană în parte.

```
> # Afisarea numarului de valori lipsa pe fiecare coloana
> colSums(is.na(date))
Tari    x1    x2    x3    x4    x5    x6    x7    x8    x9    x10
  0      1     1     1     0     2     1     1     0     0      4
```

Figură 1

- Apoi, pentru a înlocui valorile lipsă cu media fiecărei coloane am folosit un cod care verifică fiecare coloană și, dacă este numerică, înlocuiește valorile lipsă cu media acelei coloane.

```
for (col in names(date)) {
  if (is.numeric(date[[col]])) {
    date[[col]][is.na(date[[col]])] <- mean(date[[col]]), na.rm = TRUE)
  }
}
```

- La final am mai făcut încă o dată verificarea pentru a mă asigura că au fost eliminate cu succes valorile lipsă.

```
colSums(is.na(date))
```

```
> colSums(is.na(date))
Tari    x1    x2    x3    x4    x5    x6    x7    x8    x9    x10
  0      0     0     0     0     0     0     0     0     0      0
```

Figură 2

- Apoi am redenumit variabilele

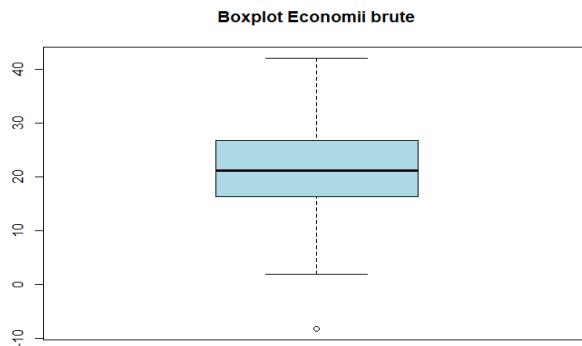
```
#redenumire date
colnames(date) <- c("Tari", "chelt_educatie", "Epuizare_energie",
                     "Econ_brute", "Epuizare_minerale",
                     "Epuizare_res_naturale", "Epuizare_neta_paduri", "Econ_nationale_nete",
                     "Daune_dioxid_carbon", "Consum_fix_capital",
                     "Daune_emisii_particule")
```

	Tari	Chelt_educatie	Epuizare_energie	Econ_brute	Epuizare_minerale	Epuizare_res_naturale	Epuizare_neta_paduri	Econ_nationale_nete	Daune_dioxid_carbon	Consum_fix_capital	Daune_emisii_particule
1	Armenia	2.740000	0.000000e+00	26.343387	0.000000e+00	0.201529351	0.2015293513	15.1617074	1.2361594		
2	Austria	5.070000	1.130848e-01	28.463932	4.359406e-03	0.175748515	0.0583043504	11.9874668	0.4357187		
3	Anglia	4.810000	1.226951e+00	13.686057	5.514342e-05	1.227006251	0.0000000000	-0.6672800	0.4781157		
4	Albania	3.067961	6.685169e-01	17.641874	4.377673e-02	0.781657640	0.0693639830	5.6233713	0.8457810		
5	Belgia	5.950000	0.000000e+00	26.776078	0.000000e+00	0.017926606	0.0179266059	8.4645339	0.5415552		
6	Belarus	4.635000	9.532228e-01	30.902789	0.000000e+00	0.953222849	0.0000000000	16.8965978	2.7085882		
7	Bosnia Herzegovina	0.000000	3.064740e-01	11.042212	8.348249e-02	0.412809297	0.0228527757	0.2904363	2.8111986		
8	Bulgaria	4.000000	4.534732e-01	15.535724	3.102492e-01	0.763722394	0.0000000000	2.5991323	2.5902476		
9	Croatia	4.220000	7.403071e-01	20.382520	0.000000e+00	0.883794416	0.1434873336	4.1103688	0.8495305		
10	Cipru	6.040000	1.210560e-02	14.429634	1.708763e-03	0.013814366	0.0000000000	2.7976389	0.8319326		
11	Cehia	3.590000	4.535348e-01	28.865999	0.000000e+00	0.477737794	0.0243830252	7.9037290	1.4277541		
12	Danemarca	7.060000	1.833196e+00	26.591760	4.050171e-03	1.841740498	0.0044938844	8.6258505	0.3765375		
13	Elvetia	4.820000	8.086604e-04	31.196240	0.000000e+00	0.001625934	0.0008172739	6.9412934	0.2263827		
14	Estonia	5.360000	1.322018e+00	24.734924	0.000000e+00	1.322017834	0.0000000000	8.9742232	2.1405846		
15	Finlanda	5.440000	0.000000e+00	28.222197	2.312790e-02	0.023127905	0.0000000000	10.3236299	0.5313238		
16	Franta	4.880000	1.409835e-02	22.941223	5.081647e-04	0.014606517	0.0000000000	6.1424612	0.3208820		
17	Germania	4.250000	1.385030e-01	26.893389	1.124669e-04	0.138617516	0.0000000000	9.7409398	0.5605829		
18	Grecia	3.111830	1.253835e-01	10.410591	7.813253e-03	0.142423356	0.0092266033	-5.0775552	0.7803980		

Figură 3

- Ulterior am făcut același lucru și pentru outlieri. Am început prin a exemplifica modul în care aceștia pot fi afișați (prin utilizarea boxploturilor și prin afișarea lor în consolă).

```
#outlieri
boxplot(date$Econ_brute, col = "lightblue", main="Boxplot Economii brute")
boxplot(date$Econ_brute, plot=FALSE)$out
```



Figură 4

```
> boxplot(date$Econ_brute, plot=FALSE)$out
[1] -8.169733
```

- Outlierii pot fi identificați și tratați în funcție de distribuția datelor. O metodă comună este folosirea IQR-ului (Abaterea intercuartilică):

```
# Detectează și gestionează outlierii
for (col in names(date)) {
  if (is.numeric(date[[col]])) {
    Q1 <- quantile(date[[col]], 0.25, na.rm = TRUE)
    Q3 <- quantile(date[[col]], 0.75, na.rm = TRUE)
    IQR <- Q3 - Q1

    # Setează valorile extreme la limita IQR-ului (cap pe baza Q1 și Q3)
    date[[col]] <- ifelse(
      date[[col]] < (Q1 - 1.5 * IQR), Q1 - 1.5 * IQR,
      ifelse(date[[col]] > (Q3 + 1.5 * IQR), Q3 + 1.5 * IQR, date[[col]])
    )
  }
}
```

Values	
col	"Daune_emisii_particule"
IQR	Named num 7.05
Q1	Named num 6.26
Q3	Named num 13.3

- Apoi am făcut o ultimă verificare pentru a mă asigura că valorile outlierilor au fost înlocuite

```
44
45 summary(date)
46
45:14 (Top Level) : 
Console Terminal x Background Jobs x
R 4.3.3 : ~/cerere/ ↵
1st Qu.: 2.601   1st Qu.:0.4949   1st Qu.:13.27   1st Qu.: 6.256
Median : 5.901   Median :0.7572   Median :15.49   Median : 9.528
Mean   : 5.746   Mean   :1.0198   Mean   :15.51   Mean   : 9.398
3rd Qu.: 8.988   3rd Qu.:1.3320   3rd Qu.:17.10   3rd Qu.:13.303
Max.   :18.568   Max.   :2.5875   Max.   :22.85   Max.   :23.873
> summary(date)
  Tari      chelt_educatie  Epuizare_energie  Econ_brute  Epuizare_minerale  Epuizare_res_naturale  Epuizare_neta_paduri
Length:39  Min.   :2.317   Min.   :0.00000   Min.   : 0.5385  Min.   :0.0000000  Min.   :0.00000   Min.   :0.00000
Class :character  1st Qu.:4.215   1st Qu.:0.01401  1st Qu.:16.3163  1st Qu.:0.0000000  1st Qu.:0.06387   1st Qu.:0.00000
Mode  :character  Median :4.760   Median :0.12538  Median :21.2404  Median :0.0005082  Median :0.20153   Median :0.00000
                           Mean   :4.735   Mean   :0.41719  Mean   :21.3363  Mean   :0.0114921  Mean   :0.50295   Mean   :0.01265
                           3rd Qu.:5.480   3rd Qu.:0.65683  3rd Qu.:26.8347  3rd Qu.:0.0169368  3rd Qu.:0.77269   3rd Qu.:0.02039
                           Max.   :7.378   Max.   :1.62105  Max.   :42.0754  Max.   :0.0423420  Max.   :1.83591   Max.   :0.05097
Econ_nationale_nete Daune_dioxid_carbon consum_fix_capital Daune_emisii_particule
Min.   :-6.979   Min.   :0.1292   Min.   :10.75   Min.   :-4.314
1st Qu.: 2.601   1st Qu.:0.4949   1st Qu.:13.27   1st Qu.: 6.256
Median : 5.901   Median :0.7572   Median :15.49   Median : 9.528
Mean   : 5.746   Mean   :1.0198   Mean   :15.51   Mean   : 9.398
3rd Qu.: 8.988   3rd Qu.:1.3320   3rd Qu.:17.10   3rd Qu.:13.303
Max.   :18.568   Max.   :2.5875   Max.   :22.85   Max.   :23.873
```

Figură 5

3. Descrieți în cuvinte setul de date, din perspectiva variabilelor (coloanelor) și observațiilor (liniilor). Enunțați clar sursa datelor și perioada de timp la care se referă.

Setul de date furnizat conține informații referitoare la mai multe țări din Europa și include o serie de indicatori economici și de mediu exprimați ca procente din Venitul Național Brut (VNB). Fiecare linie (observație) din setul de date corespunde unei țări, iar fiecare coloană (variabilă) reprezintă un indicator specific, descris în detaliu mai jos. Setul de date este structurat în următorii termeni:

Variabile (coloane):

- **Economii ajustate: cheltuieli pentru educație (% din VNB)**
Acest indicator reprezintă procentul din venitul național brut (VNB) alocat pentru educație. Este considerat o investiție în capitalul uman, având potențialul de a îmbunătăți pe termen lung productivitatea și bunăstarea economică a unei țări.
- **Economii ajustate: epuizarea energiei (% din VNB)**
Acest indicator măsoară procentul din VNB pierdut din cauza exploatarii resurselor energetice neregenerabile, precum petrolul, gazele naturale și cărbunele. Reflectă impactul pe care exploatarea resurselor energetice îl are asupra bogăției naturale a țării.
- **Economii ajustate: economii brute (% din VNB)**
Economiile brute reprezintă procentul din VNB economisit de o țară, înainte de ajustări pentru degradarea mediului și consumul de capital fix. Aceasta oferă o imagine de ansamblu asupra economisirii la nivel de economie națională.
- **Economii ajustate: epuizarea mineralelor (% din VNB)**
Acest indicator arată procentul din VNB pierdut din cauza extracției de resurse minerale (cum ar fi cuprul, aurul, minereurile de fier etc.). Este o măsură a degradării capitalului natural datorită exploatarii resurselor minerale neregenerabile.
- **Economii ajustate: epuizarea resurselor naturale (% din VNB)**
Acest indicator cumulează pierderile din exploatarea resurselor energetice, minerale și forestiere. Exprimă cât din VNB se pierde prin utilizarea resurselor naturale ale țării, fără a ține cont de regenerarea lor.
- **Economii ajustate: epuizarea netă a pădurilor (% din VNB)**
Aceasta reprezintă procentul din VNB pierdut din cauza defrișării pădurilor, peste nivelul considerat sustenabil. O reducere a acoperirii forestiere poate duce la pierderi de biodiversitate, eroziune a solului și alte daune ecologice.
- **Economii ajustate: economii naționale nete (% din VNB)**
Economiile naționale nete sunt obținute din economiile brute, la care se adaugă cheltuielile pentru educație și se scad epuizarea resurselor naturale și degradarea mediului. Este o măsură a economiilor reale ale unei țări, având în vedere resursele și costurile de mediu.
- **Economii ajustate: daune din dioxid de carbon (% din VNB)**
Acest indicator estimează costul pierderilor economice din cauza emisiilor de dioxid de carbon, măsurat ca procent din VNB. Poluarea cu CO₂ contribuie la schimbările climatice și poate afecta sănătatea și bunăstarea economică a populației.
- **Economii ajustate: consumul de capital fix (% din VNB)**
Consum de capital fix reprezintă valoarea deprecierii capitalului fizic (clădiri, echipamente, infrastructură), exprimată ca procent din VNB. Este necesară menținerea unui anumit nivel de capital pentru a susține activitatea economică, iar acest indicator reflectă costurile necesare pentru întreținerea acestui capital.
- **Economii nete ajustate, inclusiv daunele cauzate de emisiile de particule (% din VNB)**
Acest indicator include economiile nete ajustate, dar și daunele produse de emisiile de particule (PM2.5, PM10), care afectează sănătatea publică și mediul. Particulele fine din aer pot provoca boli respiratorii și alte probleme de sănătate, cu impact economic indirect asupra unei țări.

Observații (linii):

Fiecare linie reprezintă o țară, pentru care sunt colectate date pentru toți cei 10 indicatori menționați mai sus. Astfel, fiecare observație conține un profil complet al unei țări din punctul de vedere al economiilor și impactului asupra mediului, fiind valoros pentru analizarea comparativă între țări.

Sursa datelor și perioada de timp:

Acstea date sunt obținute de pe site-ul oficial la Băncii Mondiale (<https://databank.worldbank.org/source/world-development-indicators#>). Datele sunt alese pentru anul 2008 deoarece atunci a avut loc criza mondială, lucru ce va fi relevant pentru analiza mea.

4. Descrieți obiectivul general al analizei voastre.

Obiectivul principal al analizei mele a fost să evalueze impactul crizei economice globale din 2008 asupra economiilor ajustate, analizând indicatori esențiali de sustenabilitate economică și de mediu. Mi-am propus să înțeleg în ce măsură acest soc finanțiar global a influențat economiile naționale, investigând factori precum investițiile în educație, epuizarea resurselor naturale, economiile brute și nete, precum și impactul ecologic, cum ar fi emisiile de carbon și defrișările.

Prin această analiză, am dorit să ofer o imagine detaliată a modului în care diferite țări au rezisțit și gestionat consecințele crizei, explorând ajustările economice realizate pentru a susține dezvoltarea pe termen lung. Scopul meu a fost să evidențiez posibilele transformări structurale în economii și măsurile de sustenabilitate adoptate de guverne pentru a face față presiunilor economice și ecologice cauzate de criza din 2008, față de anii precedenți.

5. Calculați indicatorii statistici și interpretați din punct de vedere economic.

#indicatorii_statistici													
install.packages("psych")													
library(psych)													
indicatori_stat<-describe(date[-1])													
view(indicatori_stat)													
vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se	
Chelt_educatie	1	39	4.73518363	1.16730854	4.760000e+00	4.737262472	0.9785160000	2.3175000	7.37750000	5.06000000	-0.05823855	-0.16471898	0.186918961
Epuizare_energie	2	39	0.41719193	0.51716531	1.253835e-01	0.347109290	0.1858935471	0.0000000	1.62104505	1.62104505	1.10948421	-0.08533620	0.082812727
Econ_brute	3	39	21.33627211	8.66032211	2.124038e+01	21.480987107	8.3811542012	0.5385176	42.07538689	41.53686931	-0.15905539	0.17408278	1.386761391
Epuizare_minerale	4	39	0.01149215	0.01758859	5.081647e-04	0.009732364	0.0007534051	0.0000000	0.04234197	0.04234197	1.08641714	-0.72625925	0.002816429
Epuizare_res_naturale	5	39	0.50294507	0.54246318	2.015294e-01	0.435122975	0.2987874162	0.0000000	1.83591466	1.83591466	1.02673022	-0.04520797	0.068666831
Epuizare_neta_paduri	6	39	0.01265391	0.01966492	0.0000000e+00	0.010320600	0.0000000000	0.0000000	0.05097423	0.05097423	1.18459954	-0.29271457	0.003148907
Econ_nationale_nete	7	39	5.74599014	6.51211752	5.900933e+00	5.739311834	4.8892735118	-6.9789276	18.56794019	25.54686778	-0.08449597	-0.32174893	1.042773356
Daune_dioxid_carbon	8	39	1.01980147	0.79383949	7.572209e-01	0.952703205	0.4766591005	0.1292066	2.58748292	2.45827635	1.04802594	-0.34456741	0.127116053
Consum_fix_capital	9	39	15.51305522	3.02506663	1.548815e+01	15.341547504	2.7420046635	10.7517761	22.85217945	12.10040333	0.52502071	-0.25054942	0.404398334
Daune_emisii_particule	10	39	9.39788288	6.51790947	9.528209e+00	9.317867697	5.4574905210	-4.3136872	23.87340534	28.18709255	0.04487572	-0.14552468	0.1043700810

Figură 6

1. Cheltuieli pentru educație

- Mean (Media): 4.75318 - Media cheltuielilor pentru educație în 2008, exprimată ca procent din VNB, indică un angajament relativ scăzut al țărilor față de investițiile în educație din cauza crizei economice. Scăderea veniturilor bugetare și prioritizarea cheltuielilor esențiale au redus posibilitățile de investiții în educație.
- SD (Abaterea standard): 1.167 - Variabilitatea între țări este moderată, sugerând diferențe în prioritățile bugetare și în gradul de afectare de criză.
- Mediană: 4.76 - Valoarea mediană este aproape de media aritmetică, sugerând o distribuție simetrică și că majoritatea țărilor au cheltuit în jurul acestui procent.
- Skewness (Asimetrie): -0.0528 - Aproape simetric, cu o ușoară tendință spre valori mai mici, ceea ce indică faptul că un număr mic de țări au redus cheltuielile drastic.
- Kurtosis (Curtosis): -0.1647 - Curtosis negativă, arătând o distribuție mai plată decât cea normală; adică, țările nu au prezentat valori extreme în cheltuielile educaționale, dar s-au concentrat în jurul mediei.

2. Epuizarea energiei

- Mean: 0.417 - Media epuizării resurselor energetice ca procent din VNB este relativ mică, dar chiar și aceste pierderi minore sunt semnificative într-un context de criză.
- SD: 0.517 - Abaterea standard relativ ridicată arată o variabilitate mare între țări, sugerând că dependența de resursele energetice neregenerabile a afectat țările diferit.
- Mediană: 1.253 - Semnificativ mai mare decât media, ceea ce indică o distribuție asimetrică.
- Skewness: 1.109 - Asimetrie pozitivă ridicată, sugerând că există câteva țări cu epuizare energetică mare. Criza ar fi amplificat aceste valori din cauza cererii mai reduse și a prețurilor în scădere.
- Kurtosis: -0.385 - Distribuția este plată, ceea ce indică valori extreme mai rare, dar în contextul crizei, aceste valori extreme au probabil un impact economic negativ.

3. Economii brute

- Mean: 21.34 - Media economiilor brute este relativ ridicată, însă această valoare este afectată de criză, deoarece multe țări și-au redus economiile pentru a susține cheltuielile esențiale.
- SD: 8.66 - Abaterea standard ridicată indică variații mari între țări; țările mai dezvoltate au menținut economii mai mari, iar cele mai afectate au redus drastic.
- Mediană: 21.24 - Aproape de media aritmetică, ceea ce sugerează o distribuție relativ simetrică.
- Skewness: -0.15 - Ușoară asimetrie negativă, indicând că un număr de țări au economii sub media normală din cauza crizei.
- Kurtosis: 0.17 - Curtosis pozitivă, indicând valori concentrate în jurul mediei, ceea ce sugerează că majoritatea țărilor au avut economii în jurul acestei valori.

4. Epuizarea mineralelor

- Mean: 0.104 - Media epuizării resurselor minerale este mică, dar importantă în contextul crizei, când cererea scăzută a dus la reducerea prețurilor.
- SD: 0.157 - Abaterea standard ridicată indică o variabilitate între țări în funcție de dependența de minerale.
- Mediană: 0.05 - Ușor sub media aritmetică, sugerând că multe țări au valori reduse.
- Skewness: 1.86 - Asimetrie mare pozitivă, indicând că doar câteva țări au înregistrat epuizări semnificative.
- Kurtosis: -0.72 - Curtosis negativă, sugerând că există puține valori extreme.

5. Epuizarea resurselor naturale

- Mean: 0.502 - Epuizarea medie a resurselor naturale este mică, sugerând o conservare relativă.
- SD: 0.54 - O variabilitate moderată între țări.
- Mediană: 0.51 - Aproape de media aritmetică, ceea ce sugerează o distribuție relativ simetrică.
- Skewness: 1.02 - Ușor asimetrică, cu o tendință spre valori mai mari.
- Kurtosis: -0.26 - Distribuție mai plată, indicând o dispersie fără extreme.

6. Epuizarea netă a pădurilor

- Mean (Media): 0.0126 - Media epuizării nete a pădurilor este de 0.0126% din VNB, o valoare relativ mică, dar care poate fi semnificativă din punct de vedere ecologic. În contextul crizei economice, acest procent ar putea indica o încetinire a defrișărilor masive, deoarece activitățile economice din industriile forestiere ar fi fost afectate de scăderea cererii globale și de constrângerile bugetare. Totuși, în unele țări, presiunea pentru exploatarea resurselor naturale ar fi putut crește pentru a genera venituri.

- SD (Abaterea standard): 0.01 - Abaterea standard mică arată o variabilitate redusă între țări, ceea ce sugerează că epuizarea pădurilor a fost relativ uniformă, fără discrepanțe majore între țări.
- Mediană: 0.0064 - Valoarea medianei fiind aproape de zero sugerează că majoritatea țărilor nu au epuizat pădurile în mod semnificativ. Acest lucru este relevant, având în vedere impactul ecologic, deoarece defrișările ar fi putut contribui la probleme suplimentare de mediu.
- Skewness (Asimetrie): 1.184 - Asimetria pozitivă ridicată sugerează că există câteva țări care au avut valori mai mari ale epuizării pădurilor, indicând că acestea au continuat să exploateze resursele forestiere la un nivel mai ridicat decât media, posibil din cauza necesităților economice.
- Kurtosis (Curtosis): -0.291 - Curtosis negativă indică o distribuție mai plată decât cea normală, sugerând o lipsă de valori extreme. Acest lucru arată că majoritatea țărilor au valori reduse și relativ uniforme pentru epuizarea pădurilor.

7. Economii naționale nete

- Mean (Media): 5.754 - Media economiilor naționale nete este de 5.75% din VNB, o valoare relativ scăzută în contextul crizei economice. Aceasta indică faptul că multe țări au fost nevoie să reducă economiile nete pentru a finanța cheltuielile curente și a combate impactul negativ al crizei. Nivelurile scăzute de economii naționale nete arată presiunea exercitată de criză asupra economiilor, țările fiind nevoie să utilizeze o mare parte din resursele lor financiare.
- SD (Abaterea standard): 6.541 - O abatere standard relativ mare indică variații substanțiale între țări, sugerând că unele state au avut capacitatea de a menține economii mai mari, în timp ce altele au fost puternic afectate și au avut economii negative.
- Mediană: 5.990 - Aproape de media aritmetică, ceea ce indică o distribuție destul de simetrică a economiilor nete între țări.
- Skewness (Asimetrie): -0.084 - Aproape simetrică, ceea ce sugerează că nu există o tendință pronunțată spre valori extrem de mari sau mici.
- Kurtosis (Curtosis): -0.321 - Curtosis negativă indică o distribuție plată, fără valori extreme. Aceasta reflectă faptul că majoritatea țărilor au avut economii nete relativ apropriate de media.

8. Daunele din emisiile de dioxid de carbon

- Mean (Media): 1.098 - Costurile economice ale emisiilor de dioxid de carbon au avut o medie de 1.098% din VNB, ceea ce reflectă impactul economic negativ al poluării asupra economiilor în 2008. În contextul crizei, această valoare sugerează că, în ciuda scăderii activității economice globale, emisiile de CO₂ au continuat să provoace daune, afectând sănătatea publică și mediul. Aceste costuri sunt semnificative pentru țările care depind de industriile poluanțe.
- SD (Abaterea standard): 0.723 - Variabilitatea relativ ridicată indică diferențe între țări în ceea ce privește impactul emisiilor de dioxid de carbon, posibil din cauza nivelurilor diferite de industrializare și a măsurilor de control al poluării.
- Mediană: 0.572 - Mediana mai mică decât media sugerează că mai multe țări au costuri ale daunelor sub media, dar câteva țări industrializate au avut valori ridicate, contribuind la media generală mai mare.
- Skewness (Asimetrie): 1.048 - Asimetria pozitivă arată că există țări cu valori mari ale daunelor din emisiile de dioxid de carbon, ceea ce este caracteristic pentru țările industrializate, mai ales în timpul crizei, când controlul emisiilor ar fi putut să nu fie o prioritate.
- Kurtosis (Curtosis): -0.344 - Curtosis negativă indică o distribuție plată, sugerând o dispersie a valorilor fără extreame semnificative, ceea ce sugerează că majoritatea țărilor au valori în jurul mediei sau mai mici.

9. Consum de capital fix

- Mean (Media): 15.51 - Media consumului de capital fix este relativ mare, indicând o depreciere semnificativă a infrastructurii și echipamentelor, ceea ce a fost probabil amplificat de criză. Într-o perioadă de criză, menținerea și reînnoirea capitalului fix poate fi limitată, ceea ce duce la depreciere.
- SD (Abaterea standard): 3.205 - Variabilitatea moderată între țări sugerează că nivelurile de depreciere a capitalului fix sunt relativ uniforme, dar există diferențe notabile între statele cu economii mai dezvoltate și cele în curs de dezvoltare.
- Mediană: 15.54 - Aproape de media aritmetică, indicând o distribuție simetrică și că majoritatea țărilor au avut niveluri similare de consum de capital fix.
- Skewness (Asimetrie): 0.522 - Asimetria ușor pozitivă arată că există câteva țări cu valori mai mari ale consumului de capital fix, probabil cele care depind de industrii mari, mai afectate de criză.
- Kurtosis (Curtosis): -0.25 - Curtosis negativă sugerează o distribuție relativ plată, ceea ce indică absența unor valori extreme și concentrarea valorilor în jurul mediei.

10. Daunele din emisiile de particule

- Mean (Media): 9.397 - Costurile economice ale daunelor cauzate de emisiile de particule au o valoare medie ridicată, de 9.397% din VNB. Acest procent semnificativ reflectă impactul poluării cu particule asupra sănătății publice și mediului, care a devenit o problemă mai gravă în timpul crizei, când investițiile în tehnologii de reducere a poluării au fost limitate.
- SD (Abaterea standard): 6.517 - Abaterea standard ridicată sugerează variații mari între țări, arătând că impactul poluării aerului diferă semnificativ în funcție de nivelul de industrializare și de politicile de mediu.
- Mediană: 9.528 - Aproape de media aritmetică, ceea ce sugerează o distribuție simetrică, majoritatea țărilor având daune economice similare din cauza poluării cu particule.
- Skewness (Asimetrie): 0.044 - Aproape simetrică, ceea ce indică faptul că valorile sunt distribuite în mod echilibrat între țări, fără tendințe majore către extreme.
- Kurtosis (Curtosis): -0.145 - Curtosis negativă indică o distribuție mai plată decât cea normală, ceea ce sugerează că valorile sunt relativ disperse, fără extreme majore.

6. Calculați matricea de corelație și covarianță și interpretați din punct de vedere economic. Reprezentați grafic matricea de corelație.

Pentru a calcula matricea de corelație și covarianță trebuie în primul rând să standardizăm datele deoarece ne ajută să ajustăm toate variabilele pentru a avea o medie de 0 și o deviație standard de 1, eliminând influența diferențelor de scară și făcând variabilele comparabile. De asemenea, standardizarea datelor este un pas esențial în analiza PCA pe care am făcut-o la pașii următori.

```
#Standardizare date
date_std = scale(date[-1], scale=T)
View(date_std)
```

	Chelt_educatie	Epuizare_energie	Econ_brute	Epuizare_minerale	Epuizare_res_naturale	Epuizare_neta_paduri	Econ_nationale_nete	Daune_dioxid_carbon	Consum_fix_capital
1	-1.70921703	-0.80668969	0.57816729	-0.65338675	-0.55562224	1.94866411	1.44587643	0.27254621	-1.431828133
2	0.28682765	-0.58802700	0.82302480	-0.40553253	-0.60314599	1.94866411	0.95844043	-0.73576934	0.318475617
3	0.06409305	1.56576468	-0.88336377	-0.65025157	1.33471635	-0.64347640	-0.98482100	-0.68236176	-0.383369396
4	-1.42826228	0.48596646	-0.42658904	1.75396766	0.51377181	1.94866411	-0.01882934	-0.21921363	-1.1515198660
5	1.04069861	-0.80668969	0.62812973	-0.65338675	-0.89407097	0.26812706	0.41745926	-0.60244711	0.925095877
6	-0.08582446	1.03647888	1.10463759	-0.65338675	0.83003085	-0.64347640	1.71228600	1.97480910	-0.498125752
7	-2.07116075	-0.21408609	-1.18864628	1.75396766	-0.16615405	0.51863256	-0.83775420	1.97480910	-1.573941895
8	-0.62981089	0.07015413	-0.66978433	1.75396766	0.48071042	-0.64347640	-0.48323111	1.97480910	-0.851704712
9	-0.44134315	0.62478118	-0.11012889	-0.65338675	0.70204820	1.94866411	-0.25116582	-0.21449045	0.250935427
10	1.11779904	-0.78328208	-0.79750363	-0.55623495	-0.90165137	-0.64347640	-0.45274847	-0.23665850	-1.282966904
11	-0.98104622	0.06992511	0.86945114	-0.65338675	-0.04646647	0.59644878	0.33134212	0.51369818	1.801353670
12	1.99160401	2.32779170	0.60684670	-0.42311412	2.45716296	-0.41495347	0.44223102	-0.81031988	0.610843046
13	0.07265977	-0.80512605	1.13852207	-0.65338675	-0.92411923	-0.60191640	0.18355062	-0.99947005	2.426103330
14	0.53526240	1.74958738	0.39243943	-0.65338675	1.50985834	-0.64347640	0.49572709	1.41185107	0.081864491
15	0.60379612	-0.80668969	0.79511183	0.66155117	-0.88448303	-0.64347640	0.70294182	-0.61533554	0.78851478
16	0.12406006	-0.77942887	0.18532230	-0.62449502	-0.90019114	-0.64347640	0.06088204	-0.88042925	0.425017591
17	-0.41564301	-0.53887388	0.64167551	-0.64699244	-0.67159235	-0.64347640	0.61346401	-0.57851562	0.541936379
18	-1.39068082	-0.56424598	-1.26157910	-0.20916394	-0.66457680	-0.17428532	-1.66206235	-0.30157661	-0.008234276

Figură 7

58	<i>*concluzie și covariante</i>					
59	<code>cor(date_std)</code>					
60	<code>cov(date_std)</code>					
56:15	(Top Level) :					
	Console Terminal Background Jobs					
R 4.3.3 : ~/cerere/ ↵						
	chelt_educatie Epuizare_energie Econ_brute Epuizare_minerale Epuizare_res_naturale Epuizare_neta_paduri					
Chelt_educatie	1.00000000	0.03112266	0.05100060	-0.13511192	0.04286678	-0.31812298
Epuizare_energie	0.03112266	1.00000000	0.14168679	-0.02835887	0.96871603	-0.19049274
Econ_brute	0.05100060	0.14168679	1.00000000	-0.17311413	0.11638968	-0.09175818
Epuizare_minerale	-0.13511192	-0.02835887	-0.17311413	1.00000000	0.02812303	0.09076454
Epuizare_res_naturale	0.04286678	0.96871603	0.11638968	0.02812303	1.00000000	-0.08856299
Epuizare_neta_paduri	-0.31812298	-0.19049274	-0.09175818	0.09076454	-0.08856299	1.00000000
Econ_nationale_nete	0.03128481	0.14048229	0.92578491	-0.11726711	0.13689028	-0.03314628
Daune_dioxid_carbon	-0.11254465	0.22102820	-0.27855908	0.36783969	0.24145068	-0.05008027
Consum_fix_capital	0.27182846	-0.11713684	0.23597311	-0.37433757	-0.19353280	-0.15726938
Daune_emisii_particule	0.20846774	0.06281657	0.83687659	-0.14963089	0.07263371	0.02096689
	Econ_nationale_nete Daune_dioxid_carbon consum_fix_capital Daune_emisii_particule					
Chelt_educatie	0.03128481	-0.11254465	0.27182846	0.20846774		
Epuizare_energie	0.14048229	0.22102820	-0.11713684	0.06281657		
Econ_brute	0.92578491	-0.27855908	0.23597311	0.83687659		
Epuizare_minerale	-0.11726711	0.36783969	-0.37433757	-0.14963089		
Epuizare_res_naturale	0.04286678	0.96871603	0.24145068	-0.19353280	0.07263371	
Epuizare_neta_paduri	-0.31812298	-0.19049274	-0.09175818	0.09076454	-0.08856299	
Econ_nationale_nete	1.00000000	-0.13976127	-0.04955057	0.91216309		
Daune_dioxid_carbon	-0.13976127	1.00000000	-0.42876946	-0.25852654		
Consum_fix_capital	-0.04955057	-0.42876946	1.00000000	-0.03445848		
Daune_emisii_particule	0.01216309	0.75852654	0.02445848	1.00000000		

Figură 8

~ cov(date_std)	chelt_educatie Epuizare_energie Econ_brute Epuizare_minerale Epuizare_res_naturale Epuizare_neta_paduri					
Chelt_educatie	1.00000000	0.03112266	0.05100060	-0.13511192	0.04286678	-0.31812298
Epuizare_energie	0.03112266	1.00000000	0.14168679	-0.02835887	0.96871603	-0.19049274
Econ_brute	0.05100060	0.14168679	1.00000000	-0.17311413	0.11638968	-0.09175818
Epuizare_minerale	-0.13511192	-0.02835887	-0.17311413	1.00000000	0.02812303	0.09076454
Epuizare_res_naturale	0.04286678	0.96871603	0.11638968	0.02812303	1.00000000	-0.08856299
Epuizare_neta_paduri	-0.31812298	-0.19049274	-0.09175818	0.09076454	-0.08856299	1.00000000
Econ_nationale_nete	0.03128481	0.14048229	0.92578491	-0.11726711	0.13689028	-0.03314628
Daune_dioxid_carbon	-0.11254465	0.22102820	-0.27855908	0.36783969	0.24145068	-0.05008027
Consum_fix_capital	0.27182846	-0.11713684	0.23597311	-0.37433757	-0.19353280	-0.15726938
Daune_emisii_particule	0.20846774	0.06281657	0.83687659	-0.14963089	0.07263371	0.02096689
	Econ_nationale_nete Daune_dioxid_carbon consum_fix_capital Daune_emisii_particule					
Chelt_educatie	0.03128481	-0.11254465	0.27182846	0.20846774		
Epuizare_energie	0.14048229	0.22102820	-0.11713684	0.06281657		
Econ_brute	0.92578491	-0.27855908	0.23597311	0.83687659		
Epuizare_minerale	-0.11726711	0.36783969	-0.37433757	-0.14963089		
Epuizare_res_naturale	0.03128481	0.14048229	0.92578491	-0.11726711	0.13689028	
Epuizare_neta_paduri	-0.31812298	-0.19049274	-0.09175818	0.09076454	-0.08856299	
Econ_nationale_nete	1.00000000	-0.13976127	-0.04955057	0.91216309		
Daune_dioxid_carbon	-0.13976127	1.00000000	-0.42876946	-0.25852654		
Consum_fix_capital	-0.04955057	-0.42876946	1.00000000	-0.03445848		

Figură 9

• MATRICEA DE CORELAȚIE

1. Cheltuieli pentru educație

- Corelație cu **Economii brute (0.0510)**: Această corelație pozitivă, dar slabă, indică o legătură nesemnificativă între cheltuielile pentru educație și economiile brute. În contextul crizei, acest

lucru arată că investițiile în educație nu au fost influențate direct de economiile brute, sugerând că educația a fost fie prioritată, fie relativ constantă, indiferent de nivelul de economii brute.

- Corelație cu **Epuizarea netă a pădurilor (-0.3818)**: Corelația negativă relativ puternică sugerează că țările cu cheltuieli mai mari pentru educație tend să aibă o epuizare mai mică a pădurilor. Acest lucru poate reflecta o legătură între investiția în capitalul uman și o conștientizare ecologică sporită, sugerând că educația poate contribui la politici de conservare a resurselor.

2. Epuizarea energiei

- Corelație cu **Economii brute (0.1731)**: Această corelație pozitivă modestă sugerează că, în timpul crizei, țările cu o epuizare mai mare a resurselor energetice au menținut, în general, niveluri mai mari ale economiilor brute. Posibil, aceste țări au utilizat veniturile din exploatarea resurselor energetice pentru a-și proteja economiile brute.
- Corelație cu **Epuizarea resurselor naturale (0.9687)**: Aceasta este o corelație foarte puternică și pozitivă, ceea ce este de așteptat, deoarece epuizarea energiei este o componentă majoră a epuizării resurselor naturale. În contextul crizei, multe țări au exploatat intens resursele energetice pentru a compensa pierderile economice.

3. Economii brute

- Corelație cu **Economii naționale nete (0.5215)**: Această corelație pozitivă semnificativă sugerează că economiile brute au un efect important asupra economiilor naționale nete. În timpul crizei, țările care au reușit să își mențină economiile brute au avut, în general, și economii nete mai mari, ceea ce le-a oferit o stabilitate economică relativă.
- Corelație cu **Consum de capital fix (0.2937)**: Corelația pozitivă indică faptul că economiile brute au fost oarecum legate de consumul de capital fix. În contextul crizei, acest lucru poate arăta că țările au fost nevoie să susțină infrastructura și echipamentele pentru a menține activitatea economică, chiar și cu un cost suplimentar.

4. Epuizarea mineralelor

- Corelație cu **Epuizarea resurselor naturale (0.0212)**: Corelația foarte scăzută dintre aceste două variabile sugerează că epuizarea mineralelor nu a contribuit semnificativ la epuizarea totală a resurselor naturale, spre deosebire de epuizarea energiei. Acest lucru poate indica o utilizare moderată a resurselor minerale în comparație cu resursele energetice.
- Corelație cu **Daunele din emisiile de particule (-0.1158)**: Corelația negativă sugerează că țările care au avut o epuizare mai mare a resurselor minerale au avut daune mai mici din poluarea cu particule. Posibil, industriile extractive de minerale au fost mai puțin poluante sau au fost reglementate mai strict în această perioadă.

5. Epuizarea resurselor naturale

- Corelație cu **Epuizarea netă a pădurilor (-0.0856)**: Corelația negativă slabă indică o relație ușor inversă, ceea ce sugerează că țările care și-au exploatat intens resursele naturale au reușit să evite epuizarea pădurilor. Acest lucru ar putea sugera o protecție a pădurilor în anumite țări, chiar dacă alte resurse naturale au fost exploataate.
- Corelație cu **Daunele din emisiile de dioxid de carbon (0.2414)**: Această corelație pozitivă sugerează că exploatarea resurselor naturale a fost legată de emisiile de CO₂, indicând o legătură între activitățile de exploatare și poluarea cu carbon. În contextul crizei, emisiile ar fi crescut în țările dependente de resursele naturale pentru a susține economia.

6. Epuizarea netă a pădurilor

- Corelație cu **Daunele din emisiile de dioxid de carbon (0.1496)**: Corelația pozitivă indică faptul că epuizarea pădurilor este asociată cu emisiile de CO₂, posibil din cauza defrișărilor care contribuie la creșterea emisiilor de carbon.

- Corelație cu **Economii naționale nete (-0.0884)**: Corelația negativă slabă sugerează că țările care au avut economii nete mai mari au avut o epuizare mai mică a pădurilor, indicând o posibilă legătură între stabilitatea economică și protecția resurselor forestiere.

7. Economii naționale nete

- Corelație cu **Consum de capital fix (-0.0495)**: Această corelație ușor negativă sugerează că, pe măsură ce economiile nete scad, consumul de capital fix tinde să crească, indicând că țările care au avut economii nete mai mici în 2008 au fost nevoie să consume mai mult din infrastructura lor existentă.
- Corelație cu **Daunele din emisiile de particule (0.9127)**: Corelația puternic pozitivă arată o asociere între economiile naționale nete și daunele din emisiile de particule. Aceasta sugerează că țările care au avut economii mai mari ar putea să fie mai industrializate și, prin urmare, să genereze mai multe emisii de particule.

8. Daunele din emisiile de dioxid de carbon

- Corelație cu **Consum de capital fix (-0.2585)**: Această corelație negativă indică faptul că țările cu emisii mai mari de CO₂ au avut un consum mai mic de capital fix, probabil pentru că activitățile industriale mai poluante nu au necesitat investiții majore în infrastructura fixă.
- Corelație cu **Daunele din emisiile de particule (0.2935)**: Corelația pozitivă sugerează o legătură între emisiile de CO₂ și cele de particule, ambele având impact negativ asupra mediului. În contextul crizei, acest lucru reflectă faptul că poluarea a fost un efect secundar al strategiilor de redresare economică.

9. Consum de capital fix

- Corelație cu **Daunele din emisiile de particule (-0.0344)**: Corelația foarte scăzută sugerează o relație neglijabilă între consumul de capital fix și emisiile de particule, indicând că deprecierea capitalului fizic nu a fost influențată de poluare.

10. Daunele din emisiile de particule

- Corelațiile pozitive și negative variante între acest indicator și ceilalți sugerează un impact divers al activităților poluante asupra economiei, indicând faptul că țările cu niveluri industriale ridicate au fost cele mai afectate economic.

• MATRICEA DE COVARIANȚĂ

1. Cheltuieli pentru educație

- Covarianță cu **Economii brute (0.0511)**: Aceasta este o covarianță pozitivă mică, indicând o ușoară tendință ca cheltuielile pentru educație și economiile brute să crească împreună. În contextul crizei economice, această legătură mică sugerează că investițiile în educație nu au fost profund afectate de nivelul de economii brute, fiind probabil considerate o cheltuială stabilă.
- Covarianță cu **Epuizarea netă a pădurilor (-0.3818)**: Această covarianță negativă indică faptul că, în general, țările cu cheltuieli mai mari pentru educație au avut o epuizare mai mică a pădurilor. Acest lucru poate reflecta o relație indirectă între investiția în educație și conștientizarea ecologică, sugerând că țările care valorizează educația tend să protejeze resursele naturale.

2. Epuizarea energiei

- Covarianță cu **Economii brute (0.1416)**: Covarianța pozitivă între epuizarea energiei și economiile brute indică o tendință ca țările cu economii mai mari să exploateze resursele energetice. Aceasta ar putea fi o strategie de generare a veniturilor prin exporturi, mai ales în perioade de criză.
- Covarianță cu **Epuizarea resurselor naturale (0.9687)**: Aceasta este o covarianță foarte mare, așteptată, având în vedere că epuizarea energiei face parte din epuizarea totală a

resurselor naturale. Ea sugerează că, în contextul crizei, multe țări au exploatat intens resursele energetice pentru a atenua efectele economice negative.

3. Economii brute

- Covarianță cu **Economii naționale nete (0.5215)**: Covarianța pozitivă indică o relație directă între economiile brute și economiile nete naționale. Țările cu economii brute mai mari au avut și economii nete mai mari, ceea ce le-a oferit o anumită stabilitate economică în perioada crizei.
- Covarianță cu **Consum de capital fix (0.2937)**: Covarianța pozitivă între economiile brute și consumul de capital fix arată că țările care și-au menținut economiile brute au avut posibilitatea de a susține cheltuielile legate de capitalul fix. Aceasta sugerează o strategie de investiții, chiar și într-un context economic dificil.

4. Epuizarea mineralelor

- Covarianță cu **Daunele din emisiile de particule (-0.1158)**: Aceasta este o covarianță negativă, indicând o tendință ca țările cu epuizare mai mare a resurselor minerale să aibă niveluri mai mici ale daunelor din emisiile de particule. Acest lucru ar putea sugera că industriile de extracție a mineralelor sunt mai puțin poluante în termeni de particule.
- Covarianță cu **Economii brute (-0.0236)**: Covarianța negativă mică sugerează că exploatarea mineralelor nu a fost semnificativ corelată cu economiile brute, poate pentru că resursele minerale nu au fost exploataate la scară largă pentru a genera venituri suplimentare în criză.

5. Epuizarea resurselor naturale

- Covarianță cu **Daunele din emisiile de dioxid de carbon (0.2415)**: Covarianța pozitivă între epuizarea resurselor naturale și daunele din emisiile de dioxid de carbon indică o asociere între activitățile de exploatare și poluarea cu CO₂. Aceasta reflectă faptul că exploatarea intensă a resurselor a avut un impact negativ asupra mediului, ceea ce era probabil mai accentuat în timpul crizei economice.
- Covarianță cu **Epuizarea netă a pădurilor (-0.0856)**: Covarianța negativă slabă indică o relație inversă între epuizarea resurselor naturale și epuizarea pădurilor, sugerând că unele țări au preferat să exploateze alte resurse decât să afecteze pădurile. Aceasta poate reflecta o formă de protecție a pădurilor, chiar și în perioade economice dificile.

6. Epuizarea netă a pădurilor

- Covarianță cu **Daunele din emisiile de dioxid de carbon (0.1496)**: Covarianța pozitivă între epuizarea pădurilor și daunele din emisiile de dioxid de carbon sugerează că defrișările au contribuit la emisiile de CO₂, probabil din cauza activităților agricole și de infrastructură.
- Covarianță cu **Economii naționale nete (-0.0884)**: Covarianța negativă între epuizarea pădurilor și economiile naționale nete sugerează că țările cu economii nete mai mari au avut o epuizare mai mică a pădurilor, ceea ce poate indica o strategie de conservare a resurselor naturale.

7. Economii naționale nete

- Covarianță cu **Consum de capital fix (-0.0495)**: Covarianța negativă între economiile naționale nete și consumul de capital fix sugerează o ușoară tendință ca țările cu economii nete mai mari să consume mai puțin din capitalul lor fix, indicând o abordare de menținere a infrastructurii în perioade dificile.
- Covarianță cu **Daunele din emisiile de particule (0.9127)**: Covarianța foarte mare și pozitivă sugerează o legătură între economiile naționale nete și daunele din emisiile de particule. Acest lucru poate reflecta faptul că țările cu economii mai mari tind să fie și mai industrializate, ceea ce generează mai multe emisii de particule.

8. Daunele din emisiile de dioxid de carbon

- Covarianță cu **Consum de capital fix (-0.2585)**: Covarianță negativă între emisiile de dioxid de carbon și consumul de capital fix sugerează că țările cu emisii mari de CO₂ au un consum mai mic de capital fix. Aceasta ar putea indica faptul că aceste țări nu au investit masiv în infrastructură și echipamente noi, ci s-au concentrat pe utilizarea resurselor existente.
- Covarianță cu **Daunele din emisiile de particule (0.2935)**: Covarianță pozitivă arată că există o asociere între emisiile de CO₂ și emisiile de particule, ambele tipuri de poluare fiind frecvent asociate cu activitățile industriale.

9. Consum de capital fix

- Covarianță cu **Daunele din emisiile de particule (-0.0344)**: Covarianță ușor negativă între consumul de capital fix și emisiile de particule sugerează că deprecierea capitalului fix nu este puternic influențată de poluare. Acest lucru poate însemna că investițiile în infrastructură au fost necesare, indiferent de nivelul de poluare.

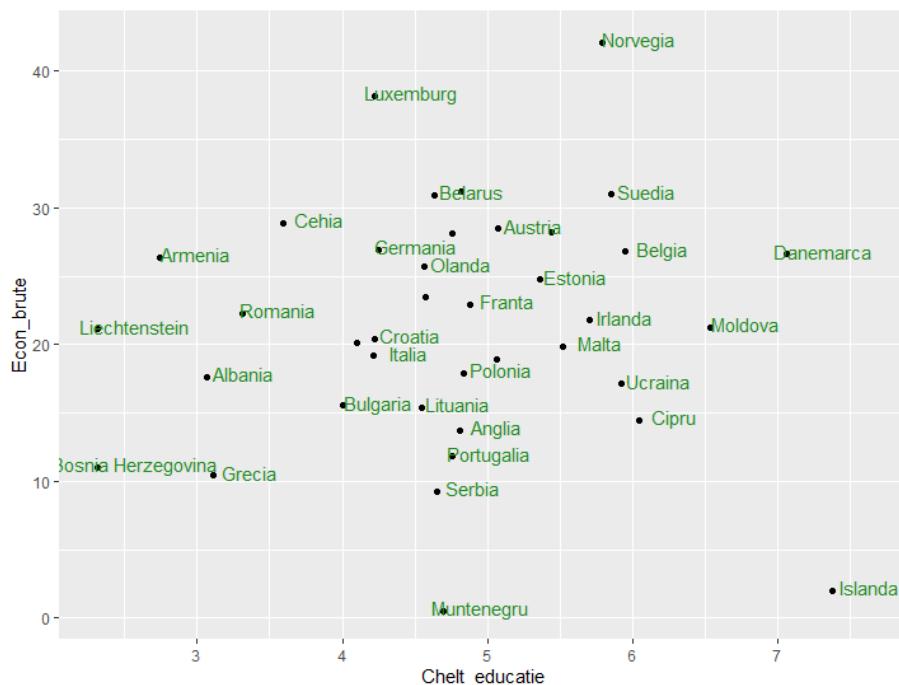
7. *Realizați cel puțin trei tipuri de reprezentări grafice care ajută în descrierea statistică a datelor.*

a) Scatter plot

Deoarece cheltuielile cu educația și economiile brute sunt puternic corelate, am ales să fac scatter plot pentru aceste 2 variabile:

```
#scatter plot
install.packages("ggplot2")
library(ggplot2)

ggplot(date, aes(x=Chelt_educatie, y=Econ_brute))+
  geom_point()+
  geom_text(label=date$Tari, color="forestgreen",
            nudge_x=0.25, nudge_y=0.25,
            check_overlap = T)
```



Figură 10

Corelația dintre cheltuielile pentru educație și economiile brute:

- Deși cele două variabile sunt corelate, graficul arată o mare variație între țări în ceea ce privește combinațiile de cheltuieli pentru educație și nivelul economiilor brute.
- Nu se observă o relație liniară strictă; mai degrabă, există clustere de țări cu diferite nivele de cheltuieli și economii, reflectând abordări variate în gestionarea resurselor și a bugetului național în timpul crizei.

Țările cu economii brute ridicate și cheltuieli mai mari pentru educație:

- Țări precum **Norvegia, Suedia și Danemarca** se află în partea superioară a graficului, cu economii brute ridicate și cheltuieli considerabile pentru educație. Aceste țări sunt cunoscute pentru sistemele lor economice stabile și investițiile susținute în educație, chiar și în perioade de criză.
- În contextul crizei, această poziție sugerează că aceste țări au avut resurse suficiente pentru a susține atât economiile, cât și sectorul educațional, ceea ce reflectă o strategie de investiții pe termen lung în capitalul uman, esențială pentru menținerea unei economii reziliente.

Țări cu economii brute moderate și cheltuieli pentru educație medii:

- Majoritatea țărilor europene, inclusiv **Austria, Germania, Franța, Olanda, Belgia și Polonia**, se regăsesc în acest interval intermediar. Aceste țări au menținut un nivel echilibrat al economiilor și al cheltuielilor pentru educație, ceea ce sugerează o abordare prudentă în gestionarea crizei.
- Aceste economii au fost suficient de robuste pentru a susține un nivel de cheltuieli constant pentru educație, dar probabil că au prioritizat economiile brute pentru a-și asigura o rezervă financiară în fața incertitudinilor economice.

Țări cu economii brute scăzute și cheltuieli mai mici pentru educație:

- Țări precum **Bosnia și Herțegovina, Grecia, Albania, și Muntenegru** au economii brute scăzute și cheltuieli reduse pentru educație. În contextul crizei, aceste țări au fost probabil cele mai afectate și au avut resurse limitate pentru a investi în educație.
- Aceste țări, confruntându-se cu presiuni economice severe, au fost nevoite să-și limiteze investițiile în capitalul uman, ceea ce poate afecta dezvoltarea pe termen lung și capacitatea de a se redresa rapid după criză.

Excepții notabile:

- **Islanda** este o excepție, situându-se în partea de jos a graficului, cu cheltuieli pentru educație destul de mari, dar cu economii brute foarte scăzute. Islanda a fost una dintre țările cel mai grav afectate de criza financiară globală, suferind o cădere economică semnificativă. În ciuda acestui fapt, pare că Islanda a continuat să aloce resurse importante educației, reflectând o strategie de menținere a capitalului uman, chiar și în vremuri dificile.
- **Luxemburg** apare cu economii brute ridicate, dar cu cheltuieli relativ moderate pentru educație în comparație cu alte țări dezvoltate, ceea ce ar putea indica o concentrare mai mare pe acumularea de economii în perioada de criză.

b) Histograme

```

histogramme
par(mfrow=c(3,2))
hist(date$chelt_educatie, freq=F, col="plum", main="Histograma Cheltuieli Educatie")
lines(density(date$chelt_educatie), lwd=3, col="darkblue")

hist(date$epuizare_energie, freq=F, col="plum", main="Histograma Epuizare Energie")
lines(density(date$epuizare_energie), lwd=3, col="darkblue")

hist(date$econom_brute, freq=F, col="plum", main="Histograma Economii Brute")
lines(density(date$econom_brute), lwd=3, col="darkblue")

hist(date$epuizare_minerale, freq=F, col="plum", main="Histograma Epuizare Minerale")
lines(density(date$epuizare_minerale), lwd=3, col="darkblue")

hist(date$epuizare_res_naturale, freq=F, col="plum", main="Histograma Epuizare Resurse Naturale")
lines(density(date$epuizare_res_naturale), lwd=3, col="darkblue")

hist(date$epuizare_neta_paduri, freq=F, col="plum", main="Histograma Epuizare Neta Paduri")
lines(density(date$epuizare_neta_paduri), lwd=3, col="darkblue")

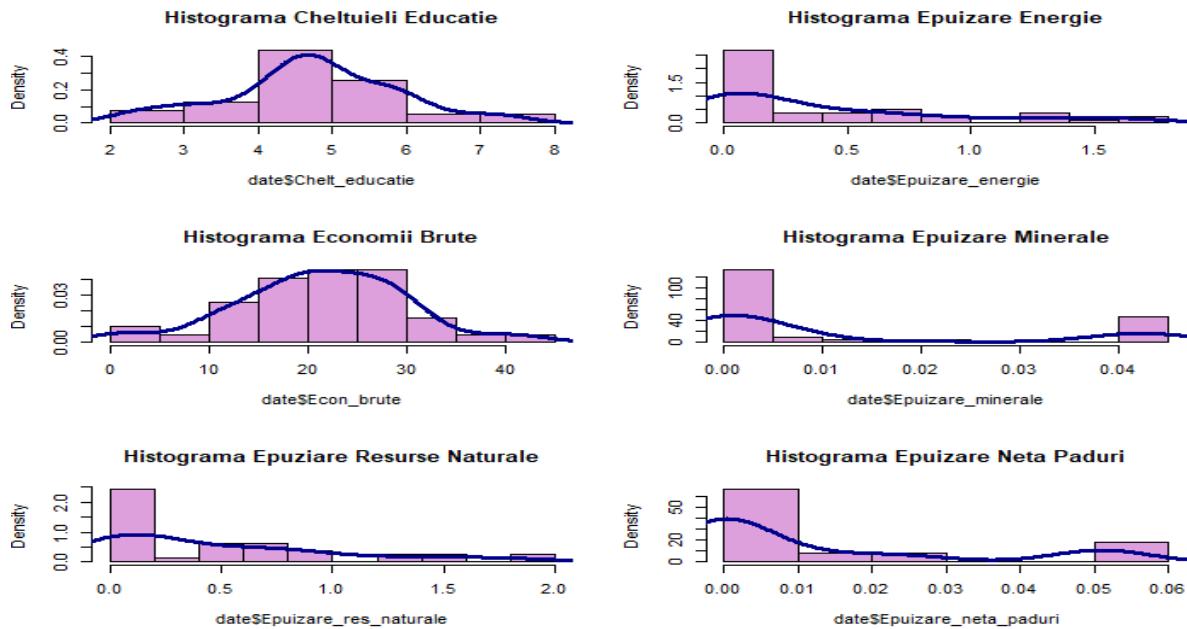
par(mfrow=c(2,2))
hist(date$econom_nationale_nete, freq=F, col="plum", main="Histograma Economii Nationale Nete")
lines(density(date$econom_nationale_nete), lwd=3, col="darkblue")

hist(date$daune_dioxid_carbon, freq=F, col="plum", main="Histograma Daune Dioxid Carbon")
lines(density(date$daune_dioxid_carbon), lwd=3, col="darkblue")

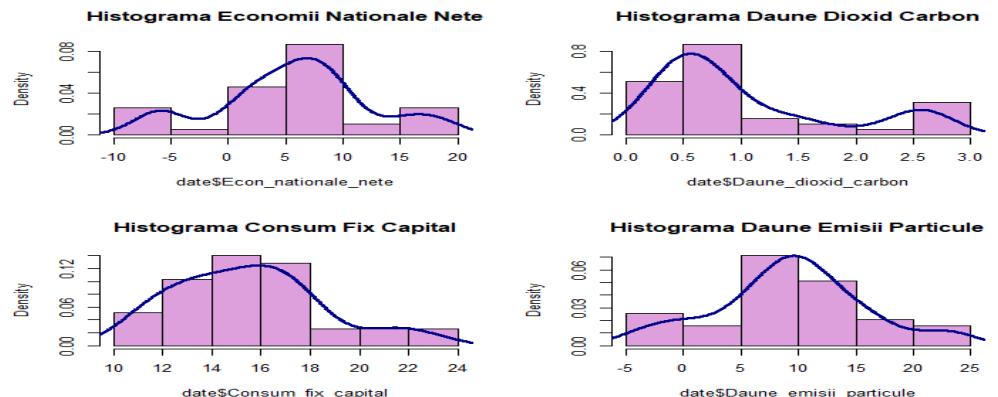
hist(date$consum_fix_capital, freq=F, col="plum", main="Histograma Consum Fix Capital")
lines(density(date$consum_fix_capital), lwd=3, col="darkblue")

hist(date$daune_emisii_particule, freq=F, col="plum", main="Histograma Daune Emisii Particule")
lines(density(date$daune_emisii_particule), lwd=3, col="darkblue")

```



Figură 11



Figură 12

1. Cheltuieli pentru educație

- Histograma și linia de densitate pentru cheltuielile educaționale arată o distribuție aproape normală, cu o concentrație a valorilor în jurul valorii de 5% din VNB.

2. Epuizarea energiei

- Distribuția este asimetrică, cu o concentrație mare de valori în jurul valorii de 0% din VNB, iar o parte semnificativă a valorilor fiind foarte mici.

3. Economii brute

- Distribuția pentru economiile brute este ușor asimetrică, cu majoritatea valorilor concentrate în intervalul 10-30% din VNB, iar linia de densitate sugerează o ușoară deplasare spre dreapta.

4. Epuizarea mineralelor

- Distribuția este puternic asimetrică, cu un număr mare de țări care au valori foarte mici sau chiar zero pentru epuizarea mineralelor.

5. Epuizarea resurselor naturale

- Distribuția este, de asemenea, asimetrică, cu majoritatea valorilor în jurul valorii de 0%, ceea ce sugerează o epuizare redusă a resurselor naturale pentru multe țări.

6. Epuizarea netă a pădurilor

- Distribuția este foarte asimetrică, cu cele mai multe valori apropiate de zero.

7. Economii naționale nete

- Distribuția este bimodală, cu o parte a valorilor concentrate în jurul valorii de 0% și o alta în intervalul pozitiv, între 5% și 15%.

8. Daunele din emisiile de dioxid de carbon

- Distribuția este asimetrică spre stânga, cu cele mai multe valori între 0,5% și 1% din VNB, dar cu o variație semnificativă.

9. Consum de capital fix

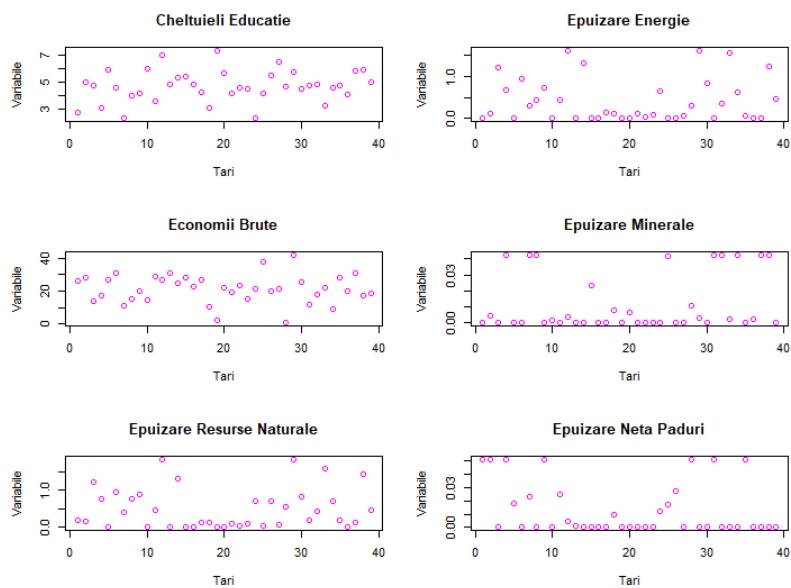
- Distribuția este aproape normală, cu o concentrație a valorilor între 15% și 20%.

10. Daunele din emisiile de particule

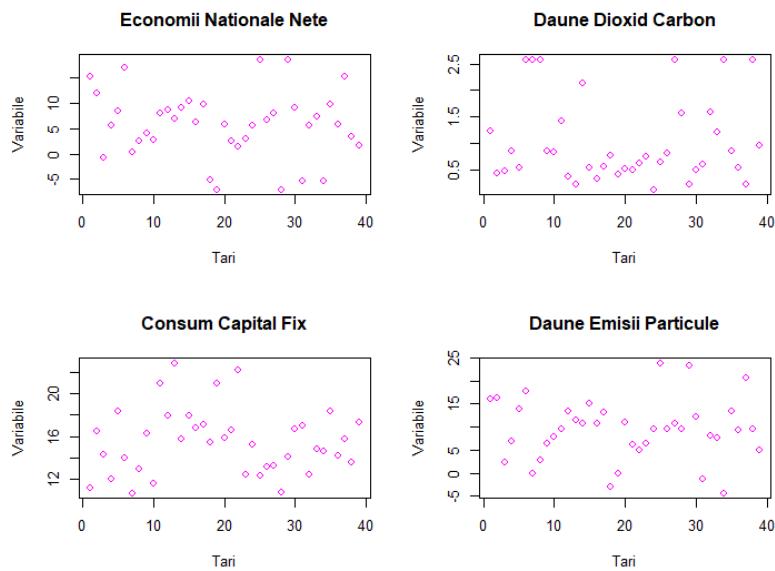
- Distribuția este ușor asimetrică, cu o concentrație a valorilor în jurul valorii de 10% din VNB, dar cu câteva valori mai mari.

c) Grafice clasice cu puncte

```
1.03 #grafice clasice cu puncte
1.04 par(mfrow=c(3,2))
1.05 plot(date_chelt_educatie, main="Cheltuieli Educatie",
1.06   xlab="Tari",
1.07   ylab="Variable",
1.08   col="blue")
1.09 plot(date_epuizare_energie, main="Epuizare Energie",
1.10   xlab="Tari",
1.11   ylab="Variable",
1.12   col="red")
1.13 plot(date_econ_brute, main="Economii Brute",
1.14   xlab="Tari",
1.15   ylab="Variable",
1.16   col="green")
1.17 plot(date_epuizare_minerale, main="Epuizare Minerale",
1.18   xlab="Tari",
1.19   ylab="Variable",
1.20   col="purple")
1.21 plot(date_epuizare_res_naturale, main="Epuizare Resurse Naturale",
1.22   xlab="Tari",
1.23   ylab="Variable",
1.24   col="brown")
1.25 plot(date_epuizare_neta_padurii, main="Epuizare Neta Paduri",
1.26   xlab="Tari",
1.27   ylab="Variable",
1.28   col="orange")
1.29 plot(date_econom_nationale_nete, main="Economii Nationale Nete",
1.30   xlab="Tari",
1.31   ylab="Variable",
1.32   col="grey")
1.33 par(mfrow=c(2,2))
1.34 plot(date_econom_nationale_nete, main="Economii Nationale Nete",
1.35   xlab="Tari",
1.36   ylab="Variable",
1.37   col="purple")
1.38 plot(date_daune_dioxid_carbon, main="Daune Dioxid Carbon",
1.39   xlab="Tari",
1.40   ylab="Variable",
1.41   col="black")
1.42 plot(date_consum_fix_capital, main="Consum Capital Fix",
1.43   xlab="Tari",
1.44   ylab="Variable",
1.45 plot(date_daune_emisi_particule, main="Daune Emisi Particule",
1.46   xlab="Tari",
1.47   ylab="Variable",
1.48   col="orange")
```



Figură 13



Figură 14

Aceste grafice de puncte prezintă variația fiecărui indicator în funcție de diferite țări (numerotate de la 1 la 40) și ne oferă o privire de ansamblu asupra dispersiei datelor pentru fiecare variabilă.

1. Cheltuieli pentru educație

- Diagrama arată că majoritatea țărilor se situează între 4% și 6% din VNB în cheltuieli pentru educație, cu câteva valori mai scăzute și câteva mai ridicate.

2. Epuizarea energiei

- Majoritatea valorilor sunt apropiate de 0, dar câteva țări au înregistrat valori mai ridicate, indicând o exploatare semnificativă a resurselor energetice.

3. Economii brute

- Variabilitatea între țări este mai pronunțată, iar majoritatea țărilor au economii brute între 10% și 30% din VNB.

4. Epuizarea mineralelor

- Majoritatea țărilor au valori foarte scăzute pentru epuizarea mineralelor, apropiate de zero, cu câteva excepții.

5. Epuizarea resurselor naturale

- Valorile sunt în mare parte foarte scăzute, iar distribuția este concentrată aproape de zero.

6. Epuizarea netă a pădurilor

- Cele mai multe valori sunt, de asemenea, aproape de zero, sugerând un nivel scăzut de epuizare a pădurilor.

7. Economii naționale nete

- Există o variabilitate semnificativă între țări, cu câteva țări înregistrând valori negative și altele cu valori pozitive considerabile.

8. Daunele din emisiile de dioxid de carbon

- Majoritatea valorilor sunt între 0 și 1% din VNB, dar există câteva țări cu daune mai ridicate din emisiile de CO₂.

9. Consum de capital fix

- Valorile sunt distribuite într-un interval destul de larg, între 12% și 24%.

10. Daunele din emisiile de particule

- Există o variabilitate între țări, cu majoritatea valorilor între 5% și 15%, dar și câteva valori extreme.

8. Aplicați metoda ACP. Calculați valorile proprii, procentul de informație reținut de fiecare componentă (procentul de varianță) și procentul cumulat.

```
#analiza ACP
date_std2<-scale(date[-1], scale = TRUE)

acp<-princomp(date_std2, cor = TRUE)

st_dev <-acp$sdev

#val proprietatea standard^2
valp <- st_dev^2

#procent_info= cat la suta din informatie are fiecare componentă
procent_info<-(valp / sum(valp)) * 100
procent_cumulat<-cumsum(procent_info)

rezultate<-data.frame(valp, procent_info, procent_cumulat)
View(rezultate)
```

În urma furnizării codului aferent în Rstudio, am obținut următoarele rezultate:

	valp	procent_info	procent_cumulat
Comp.1	3.02139365	30.2139365	30.21394
Comp.2	2.25566038	22.5566038	52.77054
Comp.3	1.65767608	16.5767608	69.34730
Comp.4	1.08481152	10.8481152	80.19542
Comp.5	0.71729482	7.1729482	87.36836
Comp.6	0.63147167	6.3147167	93.68308
Comp.7	0.51528276	5.1528276	98.83591
Comp.8	0.07187913	0.7187913	99.55470
Comp.9	0.02381859	0.2381859	99.79289
Comp.10	0.02071140	0.2071140	100.00000

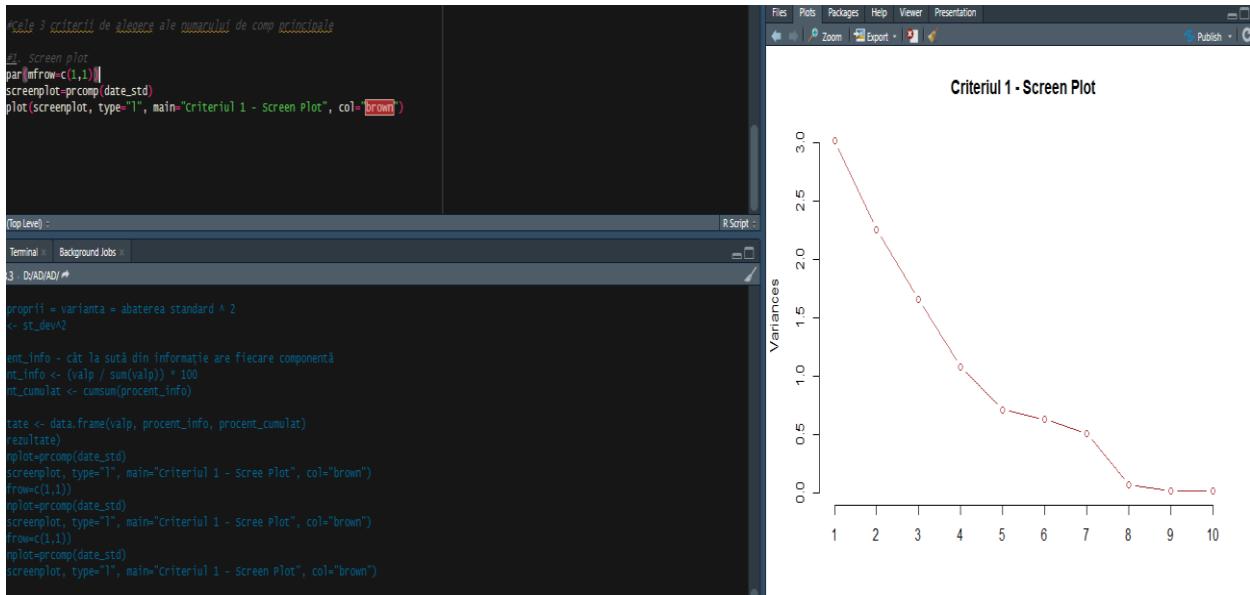
Figură 15

Interpretarea pe componente:

- **Componenta 1 (Comp.1)** explică **30.21%** din variația totală. Aceasta este componenta care reține cea mai mare parte a informației, fiind cea mai semnificativă în reprezentarea variabilității setului de date. Într-o analiză ACP, este de obicei componenta care captează tendințele generale ale datelor, cum ar fi o relație comună între economiile brute, cheltuielile pentru educație sau epuizarea resurselor.
- **Componenta 2 (Comp.2)** explică **22.56%** din variația totală, iar împreună cu Componenta 1 ajunge la un procent cumulat de **52.77%**. Aceste prime două componente acoperă mai mult de jumătate din variația totală a datelor, indicând că o bună parte din structura de bază a datelor poate fi explicată doar cu aceste două componente.
- **Componenta 3 (Comp.3)** explică **16.58%** din variație, ridicând procentul cumulat la **69.35%**. Primele trei componente împreună captează aproape 70% din variația totală, ceea ce înseamnă că putem simplifica analiza la aceste trei componente, pierzând doar o mică parte din informație.
- **Componenta 4 (Comp.4)** explică **10.85%** din variație, ceea ce duce la un procent cumulat de **80.19%**. Primele patru componente capturând deja peste 80% din variația datelor, acestea oferă o imagine destul de completă a structurii variabilelor inițiale.
- **Componenta 5 (Comp.5)** explică **7.17%** din variație, iar procentul cumulat devine **87.37%**. Aceasta indică faptul că, prin adăugarea Componentei 5, se captează încă o parte relevantă a variabilității, însă contribuția sa este semnificativ mai mică decât a primelor componente.
- **Componentele 6-10** explică procente relativ mici din variație, fiecare contribuind cu mai puțin de **6.5%**. Împreună, toate cele 10 componente ajung la 100% din variația totală, dar ultimele componente au o contribuție foarte mică, indicând că acestea captează mai mult zgromadit decât semnificație economică.

9. Aplicați cele trei criterii de alegere a numărului de componente principale. Interpretați.

Criteriul 1: SCREEN PLOT



Figură 16

Screen plot-ul prezentat oferă o imagine a valorilor proprii (varianțele) pentru fiecare componentă principală și este utilizat pentru a determina numărul optim de componente principale care ar trebui păstrate.

Este rezonabil să selectăm primele 3 sau 4 componente principale, deoarece acestea explică o parte semnificativă a variației din date.

Alegerea a 3 componente principale ar capta majoritatea informației esențiale, conform tendinței observate în grafic. Dacă dorim o reprezentare mai completă, primele 4 componente ar fi de asemenea o alegeră bună.

Criteriu 2: CRITERIUL LUI KAISER

```

186  #2. criteriul lui Kaiser
187  nr_comp_principale_kaiser<-sum(valp >= 1)
188  nr_comp_principale_kaiser
189
190
191
192
193
194
195
196
189:1 (Top Level) ◊

```

Console Terminal × Background Jobs ×

R 4.3.3 · D:/AD/AD/ ↗

```

nr_comp_principale_kaiser<-sum(valp >= 1)
nr_comp_principale_kaiser
1] 4

```

Conform criteriului lui Kaiser, ar trebui să reținem primele 4 componente principale, deoarece acestea au valori proprii mai mari sau egale cu 1, fiecare contribuind cu o parte semnificativă din variabilitatea totală a datelor. Astfel, păstrarea acestor 4 componente ne permite să captăm cea mai mare parte din variabilitatea relevantă, oferind o reprezentare eficientă a structurii inițiale a datelor.

Criteriu 3: Criteriul procentului de acoperire (procentul cumulat)

```

190 #3. criteriul procentului de acoperire (procentul cumulat)
191 proc_cumulat<-data.frame(proc_cumulat)
192 proc_cumulat
193
194
195
196
197
198
199 | Top Level |
200 Console Terminal Background Jobs
201 R 4.3.3 - D:/AD/AD/ -->
202 > nr_comp_principale_kaiser<-sum(valp >= 1)
203 > nr_comp_principale_kaiser
204 [1] 4
205 > proc_cumulat <- cumsum(proc_info)
206 > #3. criteriul procentului de acoperire (procentul cumulat)
207 > proc_cumulat<-data.frame(proc_cumulat)
208 > proc_cumulat
209 proc_cumulat
210
211 1 30.21394
2 52.77054
3 69.34730
4 80.19542
5 87.36836
6 93.68308
7 98.83591
8 99.55470
9 99.79289
10 100.00000
212 >

```

Figură 17

Deoarece avem 4 componente principale, procentul cumulat este de 80,19%. De aici reiese că pierderea informațională este de $100\%-80,19\% = 19,81\%$.

10. Calculați vectorii proprii și forma generală pentru componentele principale reținute în analiză.

```

> print(vectori_proprii_retinut)
          PC1        PC2        PC3        PC4
Chelt_educatie 0.13510191 0.0826223056 -0.4329113 -0.49951159
Epuizare_energie 0.12442802 -0.5624210500 -0.3046028 0.24608892
Econ_brute 0.54210049 0.0003300817 0.1215574 -0.01413487
Epuizare_minerale -0.19259805 -0.2176453019 0.3506016 -0.42357199
Epuizare_res_naturale 0.11030041 -0.5803633308 -0.2463610 0.25035910
Epuizare_neta_paduri -0.07898348 0.0628427426 0.4499893 0.52995579
Econ_nationale_nete 0.52710729 -0.0813375022 0.2536449 -0.09380428
Daune_dioxid_carbon -0.21953693 -0.3948455218 0.1359399 -0.33947854
Consum_fix_capital 0.13021734 0.3551445710 -0.4348270 0.15972113
Daune_emisii_particule 0.52335313 -0.0083236424 0.2168715 -0.13125092
>

```

Figură 18

	PC1	PC2	PC3	PC4
1	0.94971065	0.157470082	3.09122955	1.22489659
2	1.53900314	1.232188603	1.27551321	0.82380986
3	-0.93626307	-1.326785391	-1.86837821	1.038833492
4	-1.11841849	-1.267799687	2.16547290	1.20114222
5	1.21195372	1.744691800	-0.24885336	-0.30748601
6	2.04862101	-2.076036746	0.32547984	-0.65025038
7	-3.18260677	-1.562409841	2.13426662	-0.16495879
8	-1.99278526	-1.827525443	0.67748699	-1.25291576
9	-0.28659943	-0.332585245	0.16005064	2.05807405
10	-0.79629753	0.813110605	-0.24451028	-1.12242593
11	0.72231943	0.496363135	-0.06751649	1.15714258
12	2.11081545	-1.937644797	-2.65532824	0.42646673
13	1.39663599	2.338584191	-0.99543236	0.14376979
14	0.91865187	-2.284619256	-1.27961680	-0.06909159
15	1.31108598	1.291205469	0.18558705	-1.20166108
16	0.47207347	1.55861946	-0.32840121	-0.22210749
17	1.15297160	1.125214230	0.03926535	-0.02050103
18	-2.77020602	0.889905711	-0.24002855	0.91074960
19	-2.31300721	2.392027561	-2.99219323	-0.67495734
20	0.34976360	1.331791659	-0.36807170	-0.84264656
21	-0.50079866	1.227294924	-0.46986200	0.22745281
22	-0.203889054	2.033891392	-1.34243790	0.28590751
23	-0.89883154	0.672781693	0.02311371	-0.25699852
24	0.16515962	-0.091354355	0.29582673	1.87614242
25	2.62125983	0.167901589	2.95600590	-1.30235447
26	0.04703620	0.280633317	0.28057050	0.20139240
27	-0.04166052	0.008490802	-0.07449956	-2.04098666
28	-2.84093586	-0.466001273	0.90347774	0.76898817
29	4.42960246	-2.537844038	-0.84914542	0.36788699
30	1.28020149	-0.3777321336	-0.83524627	0.54251568
31	-2.81006740	1.037722770	0.64029144	0.60249822
32	-0.92972342	-0.923074557	0.79172394	-1.55538318
33	0.45313883	-2.631325884	-0.95964965	1.41863197
34	-3.43308819	-1.603648257	-0.52531484	-1.11316967

Figură 19

11. Calculați scorurile pentru componentele principale. Salvați noul set de date.

```

218 # Calculăm scorurile pentru componentele principale
219 scoruri_componente <- as.data.frame(acp$x)
220
221 # salvăm scorurile ca un nou set de date, adăugând numele țărilor pentru referință
222 # Presupunem că prima coloană din setul initial 'date' conține numele țărilor
223 scoruri_componente <- cbind(Tari = date$Tari, scoruri_componente)
224
225 # Afisăm nou set de date cu scorurile componentelor principale
226 print("Noul set de date cu scorurile componentelor principale:")
227 print(scoruri_componente)
228
229 # salvăm nou set de date în Rstudio (de exemplu, ca fișier csv)
230 write.csv(scoruri_componente, "scoruri_componente_principale.csv", row.names = FALSE)
231
232 # Vizualizăm scorurile în Rstudio
233 view(scoruri_componente)
234

```

	Tări	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
1	Armenia	0.94971065	0.157470082	3.09122955	1.22489659	0.145406369	-1.23935823	-0.35110947	0.072490807	-0.181800068	-0.1473245820
2	Austria	1.53900314	1.232188603	1.27551321	0.82380986	-1.274660143	-0.05793021	-0.53102315	0.092732464	-0.137898779	0.1133797615
3	Anglia	-0.93626307	-1.326785391	-1.86837821	1.03883492	0.004093935	-0.29022636	1.12018650	0.187244992	-0.052486115	0.1350527582
4	Albania	-1.11841849	-1.267799687	2.16547290	1.20114222	-0.743402240	1.04851263	0.62786997	0.137709492	-0.275472362	0.0661032217
5	Belgia	1.21195372	1.744691800	-0.24885336	-0.30748601	-0.600434376	-0.12044196	-0.46653133	0.046469364	0.016757528	0.0925917183
6	Belarus	2.04862101	-2.076036746	0.32547984	-0.65025038	0.833086968	-1.02123187	-1.29221030	-0.159142767	-0.186280999	-0.0070137587
7	Bosnia Herzegovina	-3.18260677	-1.562409841	2.13426662	-0.16495879	0.989205114	0.25061272	-0.28185911	0.064312879	-0.044876945	-0.0028800335
8	Bulgaria	-1.99278526	-1.827525443	0.67748699	-1.25291576	0.736532333	0.53340841	-0.39468415	0.093708416	0.084253801	-0.2829257810
9	Croatia	-0.28659945	-0.332585245	0.16005064	2.05807405	-0.911913531	0.01267693	-0.58072566	0.266721825	-0.033696278	0.0668901717
10	Cipru	-0.79629753	0.813110605	-0.24451028	-1.12242593	-0.408549871	-1.51884526	0.89230076	0.287883586	-0.003204353	0.1710872271
11	Cehia	0.72231943	0.496363135	-0.06751649	1.15714258	0.821857791	0.70911160	-1.67770790	-0.141781978	0.042291521	-0.0201685976
12	Danemarca	2.11081545	-1.937644797	-2.65532824	0.42646673	-1.480683832	0.52560488	0.31661717	0.085494262	-0.051143892	-0.0846884546
13	Elvetia	1.39663599	2.338584191	-0.99543236	0.14376979	0.792748549	1.03996932	-0.63932930	-0.139117581	0.207600394	-0.0135625859
14	Estonia	0.91865187	-2.284619256	-1.27961680	-0.06909159	0.182046909	-0.48415932	-0.96112039	-0.009072817	-0.085219354	0.0920431572
15	Finlanda	1.31108598	1.291205469	0.18558705	-1.20166108	0.066218069	0.80199742	0.13027436	-0.178092431	-0.054271245	-0.0280501141
16	Franta	0.47207347	1.558641946	-0.32840121	-0.22210749	0.467401255	-0.23886463	0.42518736	-0.030852834	0.017563082	0.0136228880
17	Germania	1.15297160	1.125214230	0.03926535	-0.02051013	0.922808654	-0.11976882	0.10472152	-0.164868399	-0.033956342	-0.0510037351
18	Grecia	-2.77020602	0.889505711	-0.24002855	0.91074960	1.092538637	0.01890146	0.50502592	0.176534955	0.1118145109	0.0561427907
19	Islanda	-2.31300721	2.392027561	-2.99219323	-0.67495734	-1.185545468	-0.03144516	-0.05845706	-0.297456490	-0.428755984	-0.1694991834
20	Irlanda	0.34976360	1.317191659	-0.36807170	-0.84264656	-0.086426393	-0.33886266	0.42799821	0.048337720	0.007071474	0.0685066110
21	Italia	-0.50079866	1.227294924	-0.46986200	0.22745281	0.845318927	-0.20663638	0.42492141	0.013115033	0.055944632	0.0159811596
22	Letonia	-0.20389054	2.033891392	-1.34243790	0.28590751	0.937062008	0.83190036	-0.70789752	-0.056725657	0.194881470	-0.0123712986
23	Lituania	-0.89883154	0.672781693	0.02311371	-0.25699852	0.515894243	-1.16759588	0.88894002	0.158961903	-0.013568420	0.0790599464
24	Liechtenstein	0.16515962	-0.091354355	0.29582673	1.87614242	1.252508378	0.06571479	0.97840660	-0.388186770	0.004587333	-0.1162836070
25	Luxemburg	2.62125983	0.167901589	2.95600590	-1.30235447	0.021804671	0.75019478	0.70952068	-0.097654243	0.156440026	0.1665984500
26	Malta	0.04703620	0.280633317	0.28057050	0.20139240	-0.999155411	-0.99712048	0.16427264	0.499334667	0.258964798	-0.6424699789
27	Moldova	-0.04166052	0.008490802	-0.07449956	-0.20498666	-0.148304417	-1.72166721	-1.26132591	0.300976294	0.020198345	0.1613954889
28	Muntenegru	-2.84093586	-0.466001273	0.90347774	0.76898817	-1.833455056	-1.38115458	0.12254600	-1.110435036	0.323620774	0.0627498332
29	Norvegia	4.42960246	-2.537844038	-0.84914542	0.36788699	-0.525407420	0.09845930	0.92001691	0.231306149	0.323742493	0.0902149943
30	Olanda	1.28020149	-0.377321336	-0.83524627	0.54251568	0.471287638	0.03004267	0.40904286	-0.117437498	-0.106921407	-0.0017014054
31	Portugalia	-2.81008740	1.037722770	0.64029144	0.60249822	-1.511172720	1.69927189	-0.02107871	0.374183999	0.084008877	0.1151871789
32	Polonia	-0.92972342	-0.923074557	0.79172394	-1.55538318	0.078604874	0.47634487	0.51801901	0.034749786	-0.106118057	-0.0171554015
33	Romania	0.45313883	-2.631325684	-0.95964965	1.41863197	1.065975356	0.05625446	0.30367203	-0.104110775	-0.182776592	-0.0116916356
34	Serbia	-3.43308819	-1.603648257	-0.52531484	-1.11316967	0.356530819	0.84085054	-0.54687772	0.231354043	0.099933046	0.0943168939

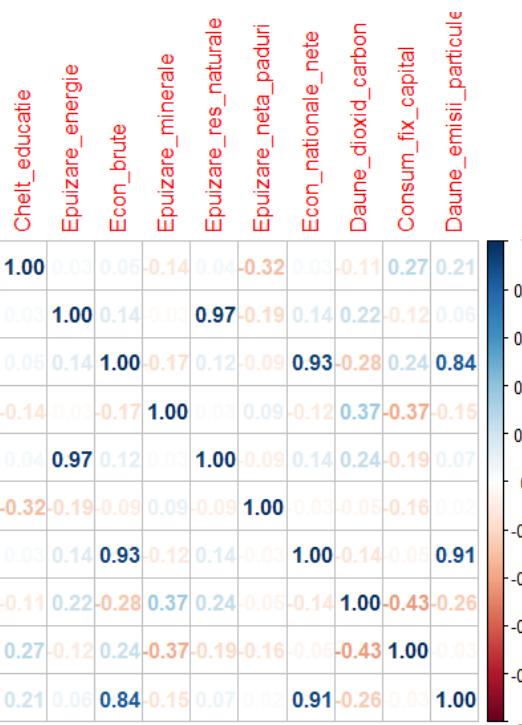
Figură 20

35	Slovenia	1.06716681	1.351812942	0.93512872	1.05577184	-0.877054320	0.08145614	-1.26671275	0.102422193	-0.010416386	0.0428386761
36	Spania	-0.12064688	1.092814689	0.32238688	-0.14403360	0.854238502	-0.62677512	0.72963012	-0.015112511	-0.027232007	-0.0004864063
37	Suedia	2.17131971	0.784076797	1.02354902	-1.95203898	-0.507233151	1.21111598	0.88258077	-0.245973308	-0.183965052	-0.1413936900
38	Ucraina	-0.71457712	-3.209986660	-0.63349349	-1.50017277	-0.723848348	0.61908345	-0.45290987	-0.352167872	0.099407055	-0.0344456726
39	Ungaria	-0.52806694	0.427746857	-1.24844720	0.06207455	0.364076701	-0.13940046	-0.11019050	0.102144138	0.075648879	0.0813529953

Figură 21

12. Calculați și reprezentați grafic matricea factor. Decideți câte componente principale păstrați în analiză și justificați-vă decizia. Explicați apartenența variabilelor inițiale în componentele principale și denumiți-le sugestiv.

> matrice_factor=cor(date_std2, c)						
> matrice_factor						
Chelt_educatie	1.0000000	0.03112266	0.05100060	-0.13511192	0.04286678	-0.31812298
Epuizare_energie	0.03112266	1.0000000	0.14168679	-0.02835887	0.96871603	-0.19049274
Econ_brute	0.05100060	0.14168679	1.0000000	-0.17311413	0.11638968	-0.09175818
Epuizare_minerale	-0.13511192	-0.02835887	-0.17311413	1.0000000	0.02812303	0.09076454
Epuizare_res_naturale	0.04286678	0.96871603	0.11638968	0.02812303	1.0000000	-0.08856299
Epuizare_neta_paduri	-0.31812298	-0.19049274	-0.09175818	0.09076454	-0.08856299	1.0000000
Econ_nationale_nete	0.03128481	0.14048229	0.92578491	-0.11726711	0.13689028	-0.03314628
Daune_dioxid_carbon	-0.11254465	0.22102820	-0.27855908	0.36783969	0.24145068	-0.05008027
Consum_fix_capital	0.27182846	-0.11713684	0.23597311	-0.37433757	-0.19353280	-0.15726938
Daune_emisii_particule	0.20846774	0.06281657	0.83687659	-0.14963089	0.07263371	0.02096689
Econ_nationale_nete	0.03128481	-0.11254465	0.27182846	0.20846774		
Epuizare_neta_paduri	0.14048229	0.22102820	-0.11726711	0.06281657		
Econ_brute	0.92578491	-0.27855908	0.23597311	-0.14963089		
Epuizare_minerale	-0.11726711	0.36783969	-0.37433757	-0.14963089		
Epuizare_res_naturale	0.13689028	0.24145068	-0.19353280	0.07263371		
Epuizare_neta_paduri	-0.03314628	-0.05008027	0.15726938	0.02096689		
Econ_nationale_nete	1.0000000	-0.13976127	0.04955057	-0.03445848		
Daune_dioxid_carbon	-0.13976127	1.0000000	-0.42876946	0.0000000		
Consum_fix_capital	-0.04955057	-0.42876946	1.0000000	-0.03445848		
Daune_emisii_particule	0.91216309	-0.25852654	-0.03445848	1.0000000		



Figură 22

Din graficul matricei factor, putem observa că primele două sau trei componente principale au încărcațuri mari pentru mai multe variabile, ceea ce sugerează că aceste componente explică o parte semnificativă a variației din setul de date. Dacă ne bazăm pe acest lucru, împreună cu valorile proprii și criteriul procentului cumulativ de acoperire, ar fi recomandat să păstrăm 2 sau 3 componente principale.

Primele două componente: Explică majoritatea corelațiilor pozitive puternice între variabile (de exemplu, între Epuizare_energie, Econ_brute, și Epuizare_res_naturale), care au valori ridicate în prima și a doua componentă.

A treia componentă (dacă este necesar): Ar putea explica variația unor variabile cu corelații mai mici cu celelalte componente, însă nu este strict necesar dacă primele două acoperă un procent suficient de mare din varianță.

Pe baza matricei factor și a corelațiilor din grafic, putem observa cum variabilele inițiale se grupează în funcție de componente:

Componenta Principală 1: Aceasta pare să fie asociată cu variabilele legate de **resurse naturale și economii brute** (Epuizare_energie, Epuizare_res_naturale, Econ_brute). Putem denumi această componentă sugestiv drept "**Resurse Naturale și Economii**".

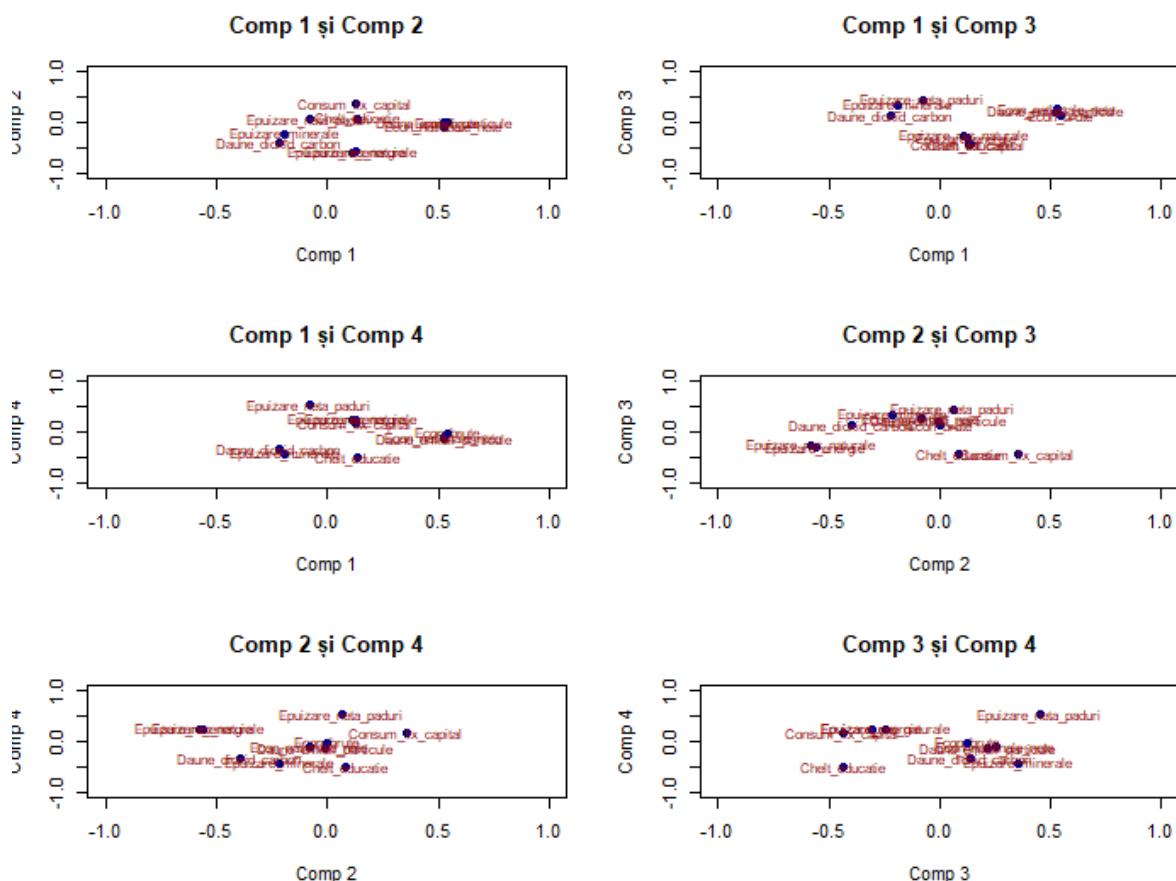
Componenta Principală 2: Aceasta pare să aibă o încărcătură mai mare pentru variabilele care reprezintă **daune și consumuri** (Daune_dioxid_carbon, Consum_fix_capital, Daune_emisii_particule). O denumire sugestivă pentru această componentă ar fi "**Impact Ecologic și Consum**".

Componenta Principală 3 (dacă o păstrăm): Dacă alegem să păstrăm și a treia componentă, aceasta ar putea fi legată de **cheltuieli în educație și epuizare netă a pădurilor** (Chelt_educatie, Epuizare_neta_paduri), sugerând o posibilă denumire precum "**Educație și Conservarea Pădurilor**".

13. Realizați următoarele interpretări grafice: cercul corelațiilor, reprezentarea scorurilor principale pentru diadele de componente principale și biploturile (cel puțin unul din fiecare categorie).

a) Cercul corelațiilor

• **Metoda 1**



Figură 23

1. Comp 1 și Comp 2:

Comp 1 pare să fie influențată de variabile legate de impactul economic și de infrastructură, iar Comp 2 de variabile legate de resurse naturale și conservarea mediului.

2. Comp 1 și Comp 3:

Comp 1 reflectă dimensiunea economică, iar Comp 3 adaugă informații suplimentare legate de resursele naturale.

3. Comp 1 și Comp 4:

Comp 4 pare să adauge informații despre aspectele de mediu, în special daunele asociate activităților industriale și impactul asupra resurselor naturale.

4. Comp 2 și Comp 3:

Aceasta sugerează că Comp 2 și Comp 3 reflectă un echilibru între investițiile în capitalul uman și protecția resurselor naturale.

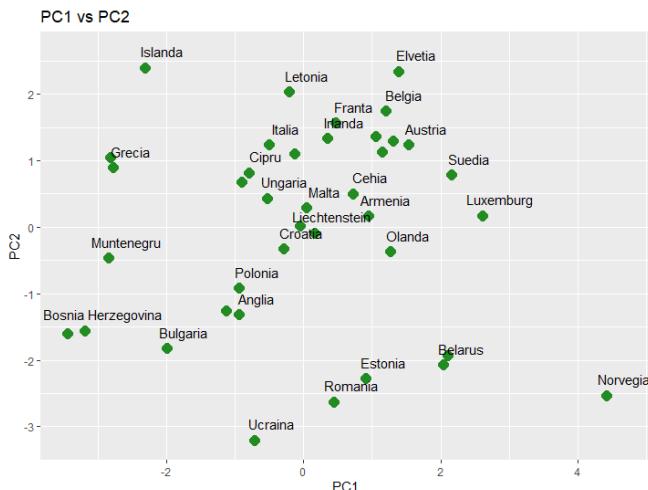
5. Comp 2 și Comp 4:

Această pereche de componente este legată de variația resurselor naturale și a efectelor poluării asupra economiei.

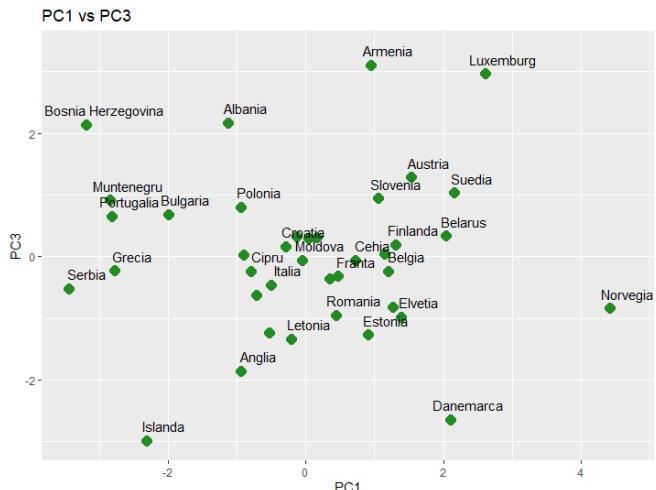
6. Comp 3 și Comp 4:

Comp 3 și Comp 4 sunt componente care au o corelație mai mică cu indicatorii economici clasici și reflectă mai mult aspectele de mediu.

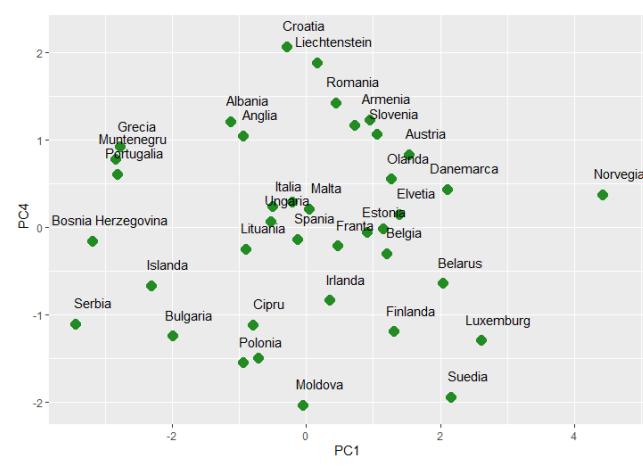
b) Reprezentarea scorurilor principale pentru diadele de componente principale(metoda 2)



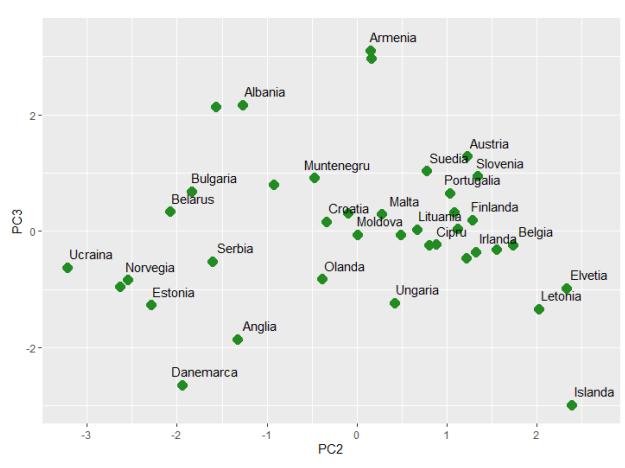
Figură 25



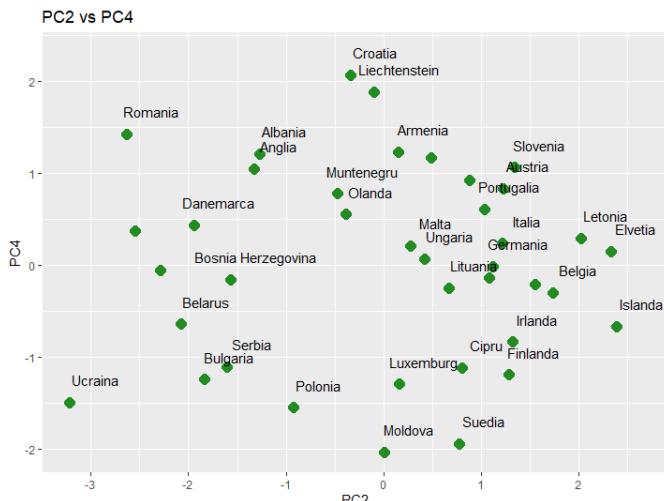
Figură 24



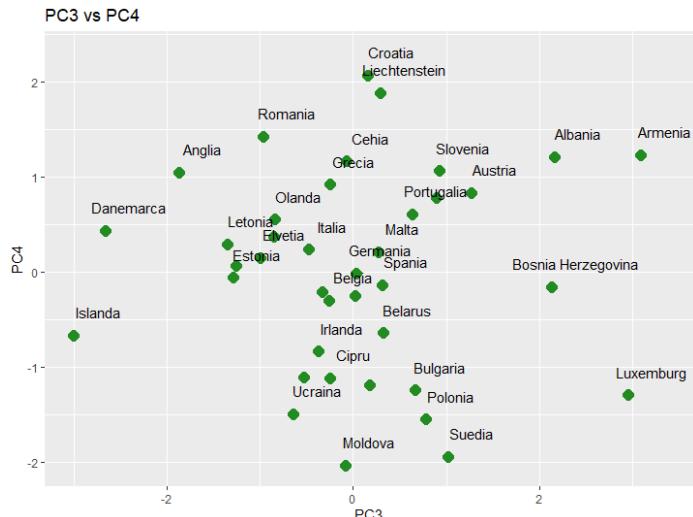
Figură 26



Figură 27



Figură 28



Figură 29

PC1 vs PC2 (Figura 25):

- Acest grafic arată cum sunt distribuite țările în funcție de primele două componente principale, care reprezintă cel mai mare procent din variabilitatea totală a datelor.
- Țările din extrema dreaptă a graficului, cum ar fi **Norvegia și Luxemburg**, au scoruri ridicate pe PC1, ceea ce sugerează că sunt influențate semnificativ de factorii asociați acestei componente (posibil legat de un nivel economic mai avansat sau alte caracteristici distințe).
- Țările din stânga graficului, precum **Grecia, Bosnia și Herțegovina și Ucraina**, au scoruri mai mici pe PC1, ceea ce indică caracteristici opuse față de țările din dreapta.
- Pe PC2, țări precum **Islanda și Grecia** se află în partea de sus, în timp ce **Ucraina și Bosnia și Herțegovina** sunt în partea de jos, sugerând o altă dimensiune de variație (poate o diferență în politici economice, sociale sau mediu).

PC1 vs PC3 (Figura 24):

- În acest grafic, țările sunt distribuite în funcție de PC1 și PC3.
- **Luxemburg și Armenia** au scoruri ridicate pe PC3, în timp ce **Islanda și Danemarca** sunt în partea de jos a PC3, ceea ce sugerează că există o diferență semnificativă între aceste țări pe această componentă (probabil o dimensiune culturală, socială sau legată de resurse).
- PC1 continuă să arate diferențele economice, cu **Norvegia** în extrema dreaptă și **Bosnia și Herțegovina** în extrema stângă.

PC1 vs PC4 (Figura 26):

- PC4 introduce o altă dimensiune de variabilitate. Țări precum **Norvegia și Croația** au scoruri ridicate pe PC4, în timp ce **Moldova și Serbia** sunt la scoruri inferioare, sugerând noi caracteristici relevante pentru aceste țări.
- PC1 continuă să mențină structura observată anterior, unde **Luxemburg și Norvegia** se află pe extrema pozitivă, iar **Grecia și Bosnia și Herțegovina** pe extrema negativă.

PC2 vs PC3 (Figura 27):

- Pe acest grafic, **Armenia și Albania** au scoruri ridicate pe PC3, în timp ce **Danemarca și Islanda** sunt la scoruri mai scăzute pe PC3.

- **Ucraina și Grecia** sunt în extrema stângă pe PC2, în timp ce **Estonia și Luxemburg** se află în partea dreaptă a graficului, ceea ce arată o altă dimensiune de diferență.

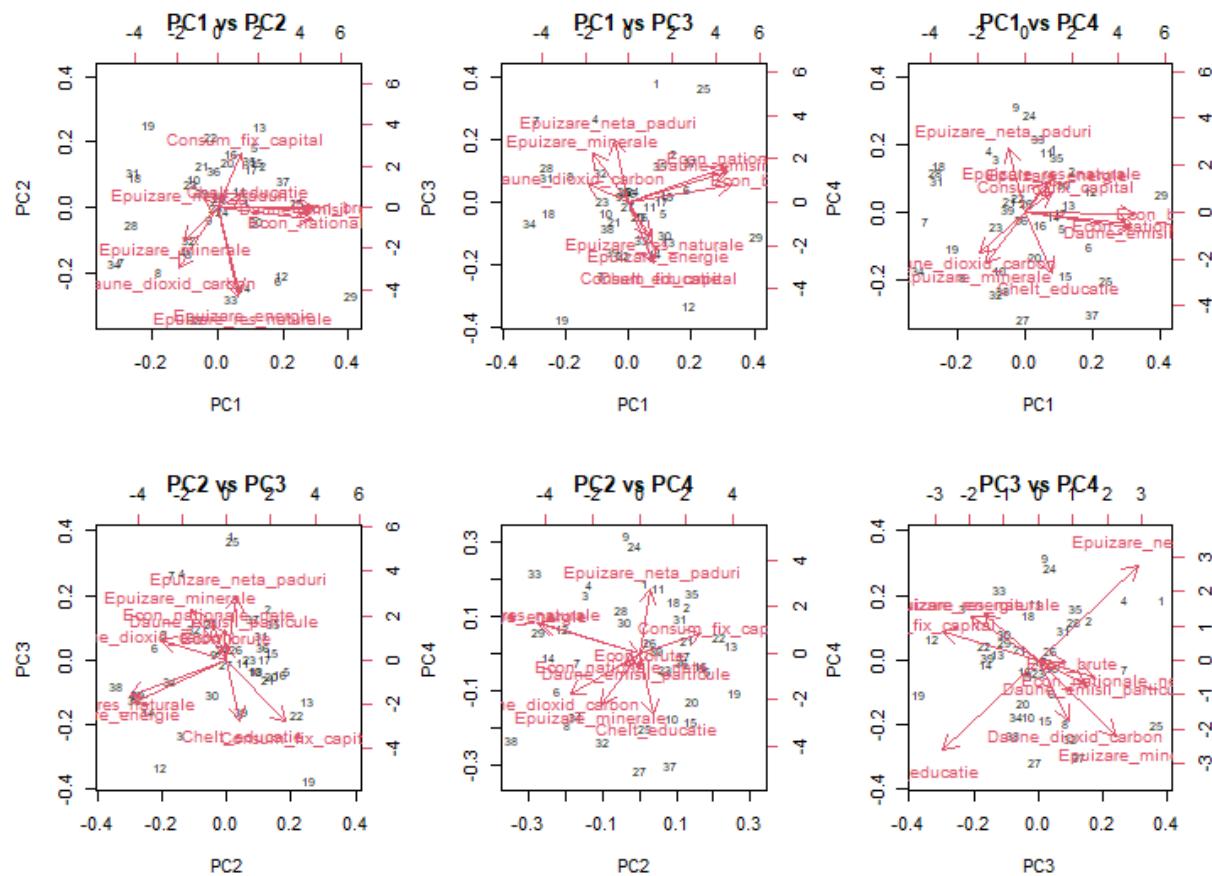
PC2 vs PC4 (Figura 28):

- În această configurație, **România și Albania** au scoruri mai ridicate pe PC4, în timp ce **Moldova și Serbia** sunt în partea inferioară a PC4.
- **Belarus și Estonia** sunt plasate pe partea dreaptă a PC2, sugerând că ele sunt mai apropiate în funcție de caracteristicile captate de această componentă.

PC3 vs PC4 (Figura 29):

- Acest grafic arată distribuția pe ultimele două componente principale, care captează variații mai mici.
- **Croația și Liechtenstein** au scoruri ridicate pe PC4, în timp ce **Moldova și Bosnia și Herțegovina** sunt la scoruri mai scăzute.
- Pe PC3, **Albania și Armenia** au scoruri ridicate, iar **Luxemburg și Suedia** sunt la scoruri inferioare, ceea ce sugerează că aceste țări se diferențiază pe baza acestor componente.

c) Biplot-uri



Figură 30

PC1 vs PC2:

- Vectorii care au o lungime mare și se apropie de axa PC1 sau PC2 indică o contribuție puternică a acestor variabile la componenta respectivă.

- Variabilele precum **Epuizare_minerale** și **Epuizare_neta_paduri** au o contribuție semnificativă atât pe PC1, cât și pe PC2, fiind orientate aproape de direcția diagonalei dintre axele PC1 și PC2.
- **Daune_dioxid_carbon** și **Consum_fix_capital** sunt orientate mai mult spre PC2, indicând o corelație puternică cu această componentă, ceea ce sugerează că țările cu valori ridicate pe PC2 tind să aibă și valori ridicate pentru aceste variabile.

PC1 vs PC3:

- În acest biplot, **Epuizare_minerale** și **Epuizare_neta_paduri** au din nou o influență puternică asupra PC1, dar și asupra PC3.
- Variabilele **Chelt_educatie** și **Daune_dioxid_carbon** sunt orientate mai mult spre PC3, sugerând că această componentă reprezintă o dimensiune diferită, posibil legată de investiții și impactul asupra mediului.

PC1 vs PC4:

- **Epuizare_minerale** și **Epuizare_neta_paduri** rămân importante pe PC1, dar PC4 este dominată de alte variabile, precum **Epuizare_energie** și **Daune_dioxid_carbon**.
- Acest biplot indică faptul că PC4 ar putea captura variația legată de impactul asupra resurselor naturale și al mediului.

PC2 vs PC3:

- **Epuizare_neta_paduri** și **Epuizare_minerale** influențează puternic atât PC2, cât și PC3, ceea ce sugerează că aceste componente sunt importante pentru evaluarea resurselor naturale și a epuizării acestora.
- **Chelt_educatie** are o contribuție semnificativă pe PC3, indicând faptul că țările cu scoruri ridicate pe PC3 ar putea avea niveluri mai ridicate de investiții în educație.

PC2 vs PC4:

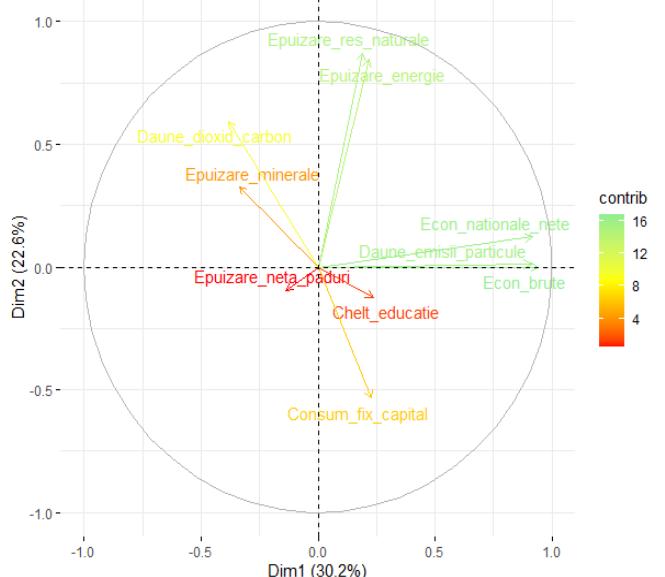
- PC4 este dominată de variabile precum **Epuizare_energie** și **Daune_dioxid_carbon**, ceea ce arată că această componentă capturează variația legată de consumul de energie și impactul ecologic.
- PC2 continuă să fie influențată de variabile precum **Epuizare_minerale** și **Consum_fix_capital**.

PC3 vs PC4:

- Acest biplot arată că variabilele **Epuizare_minerale** și **Epuizare_neta_paduri** influențează ambele componente.
- PC4 este influențată semnificativ de variabilele legate de consumul de energie și impactul ecologic, cum ar fi **Epuizare_energie** și **Daune_dioxid_carbon**.
- Țările situate departe de origine pe axele PC3 și PC4 au valori mai mari la variabilele corespunzătoare direcțiilor acestor axe.

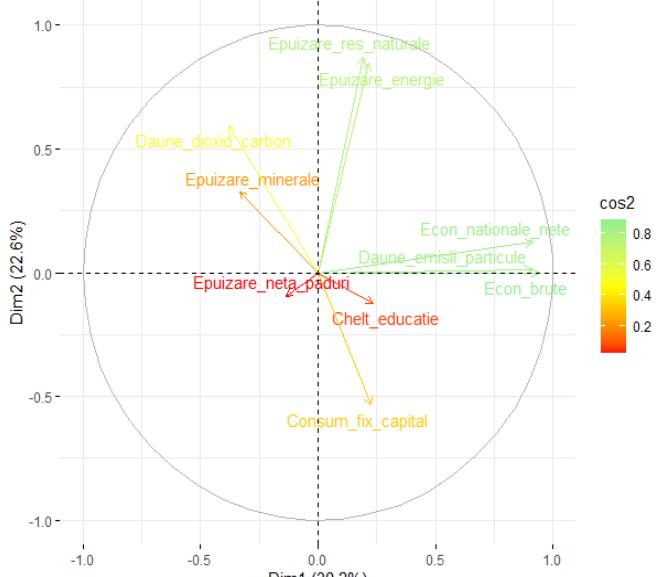
14. Calculați, cu ajutorul funcției PCA din pachetul FactoMineR, contribuțiile variabilelor și observațiilor la construirea componentelor principale (ctr), precum și calitatea reprezentării pentru variabile și indivizi (cos2). Interpretați și reprezentați grafic.

Contribuția variabilelor la componentele principale



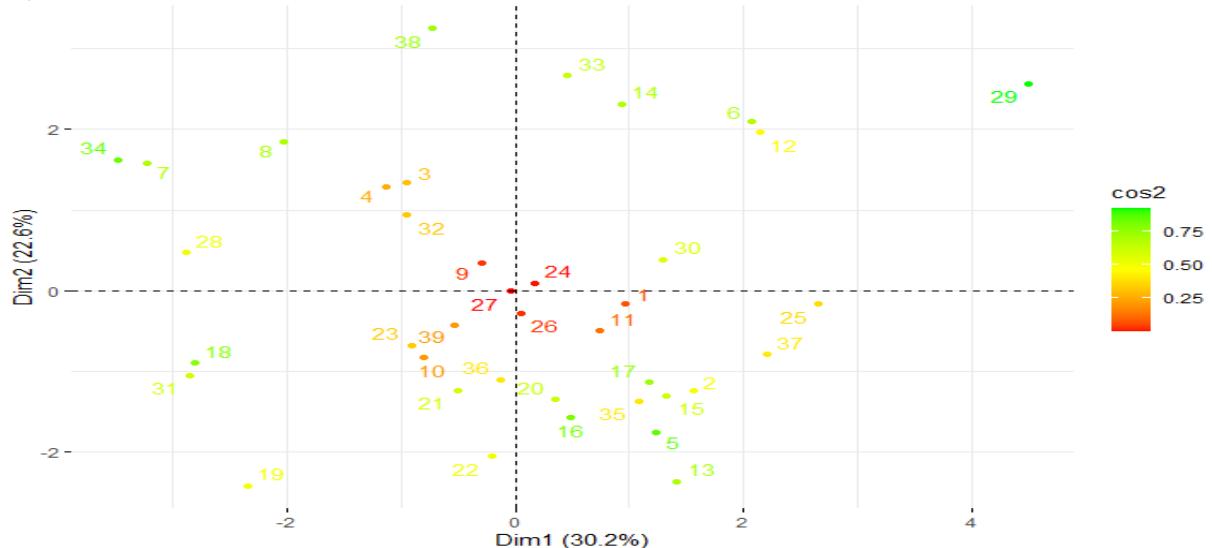
Figură 32

Calitatea reprezentării variabilelor (\cos^2)



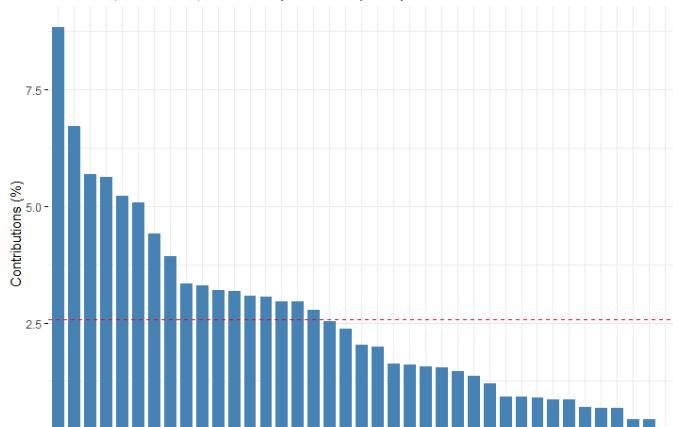
Figură 31

Figură 32 Calitatea reprezentării observațiilor (\cos^2)

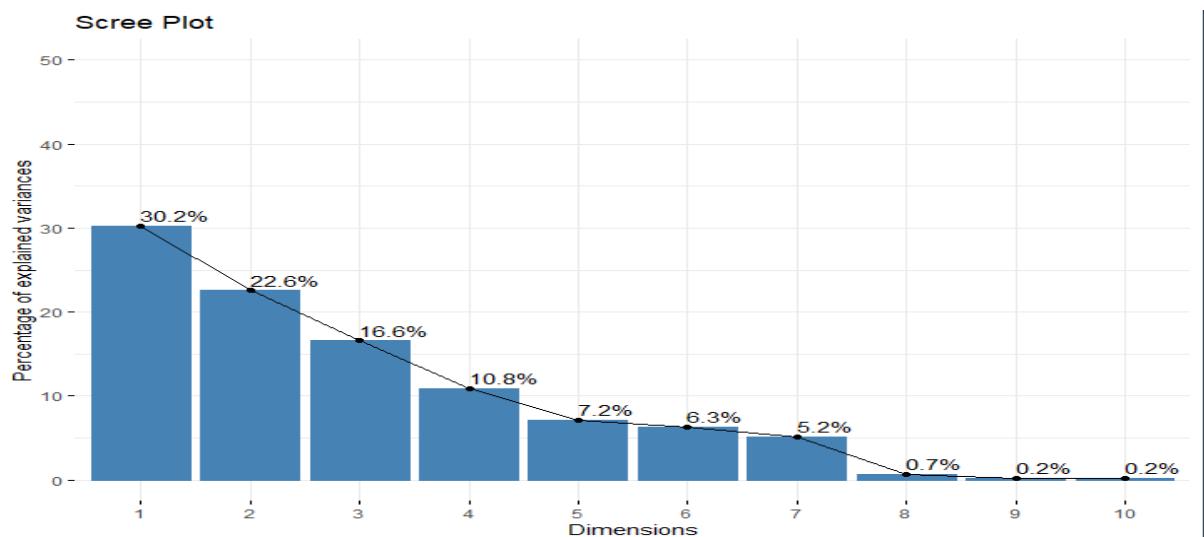


Figură 33

Contribuția observațiilor la componentele principale



Figură 34



Figură 35

1. Contribuția variabilelor la componentele principale (Figura 31)

Acest grafic arată contribuția variabilelor la primele două componente principale (Dim1 și Dim2). În acest caz:

- **Colorile gradientului** (de la roșu la verde) indică nivelul contribuției variabilelor. Variabilele în roșu au contribuția cea mai ridicată la aceste componente, în timp ce variabilele în verde au contribuții mai mici.
- **Vectorii** lungi indică variabilele care au o contribuție mai mare la structura acestor componente. De exemplu, variabile precum **Epuizare_minerale** și **Epuizare_neta_paduri** au vectori mai lunghi și sunt colorate în roșu-portocaliu, indicând o contribuție puternică la prima componentă.
- Orientarea variabilelor în jurul celor două axe ne sugerează grupări ale variabilelor care tind să varieze împreună. De exemplu, variabilele **Chelt_educatie** și **Consum_fix_capital** sunt orientate în direcții apropiate, indicând o posibilă corelație între ele.

2. Calitatea reprezentării variabilelor (\cos^2) (Figura 30)

Acest grafic indică **calitatea reprezentării variabilelor** pe primele două componente principale prin măsura \cos^2 (pătratul cosinusului). \cos^2 indică cât de bine sunt reprezentate variabilele în spațiul definit de aceste componente:

- **COLORILE GRADIENTULUI** (de la roșu la verde) arată calitatea reprezentării fiecărei variabile. Variabilele în verde au o valoare ridicată a \cos^2 , ceea ce înseamnă că sunt bine reprezentate pe Dim1 și Dim2. În schimb, variabilele în roșu au o valoare mai mică a \cos^2 și sunt mai puțin bine reprezentate pe aceste componente.
- De exemplu, **Daune_dioxid_carbon** și **Epuizare_minerale** sunt bine reprezentate în aceste două dimensiuni (au \cos^2 mare), indicând că poziția lor în planul determinat de Dim1 și Dim2 reflectă corect variația lor.

3. Calitatea reprezentării observațiilor (\cos^2) (Figura 32)

Acest grafic arată **calitatea reprezentării fiecărei țări** (observație) pe primele două componente principale, în funcție de \cos^2 . \cos^2 măsoară cât de bine sunt reprezentate țările pe aceste două componente:

- **Culorile** din grafic reflectă valoarea cos2 a fiecărei țări pe Dim1 și Dim2. Observațiile cu cos2 mare (de exemplu, cele colorate în verde) sunt bine reprezentate în acest plan bidimensional.
- Țările din zona centrală, cum ar fi cele cu cos2 mai mic (în culori galben-portocalii sau roșii), sunt mai slab reprezentate în aceste două dimensiuni, ceea ce înseamnă că o parte din variația lor ar putea fi capturată mai bine de alte componente.
- Observațiile situate departe de origine și bine reprezentate indică țările cu caracteristici distincte pentru variabilele respective, sugerând o variație mai mare față de medie.

4. Contribuția observațiilor la componentele principale (Figura 34)

Acest grafic arată contribuția fiecărei țări la variabilitatea explicată de componentele principale:

- **Bare mai lungi** indică o contribuție mai mare a acelei țări la variabilitatea componentei respective.
- Țările cu contribuție mare (barele mai lungi) sunt țările care explică o proporție semnificativă din variația totală a datelor. Aceste țări pot fi considerate ca având caracteristici extreme sau distincte în ceea ce privește variabilele analizate.
- Țările cu contribuție mică la aceste componente explică mai puțin din variabilitatea generală și tind să fie mai apropiate de medie.

5. Scree Plot (Figura 35)

Primele componente principale (PC1 și PC2):

- Primele două componente principale explică împreună aproximativ **52.9%** din variabilitatea totală a datelor (**30.2%** pentru PC1 și **22.6%** pentru PC2).
- Aceste două componente reprezintă cea mai mare parte a variației și, de aceea, sunt considerate cele mai importante pentru descrierea structurii generale a datelor. De obicei, componente care explică mai mult de 10-15% din variație sunt considerate esențiale în interpretarea rezultatelor.

Alte componente importante (PC3 și PC4):

- PC3 explică **16.8%** din variație, iar PC4 explică **10.8%**. Împreună cu PC1 și PC2, aceste patru componente explică un total cumulat de **80.3%** din variabilitatea totală a datelor.
- Aceasta sugerează că, păstrând primele patru componente, captăm o mare parte din informația inițială a datelor, iar restul componentelor contribuie doar marginal.

Componentele inferioare (PC5 - PC10):

- Începând de la PC5 în jos, fiecare componentă explică o proporție din ce în ce mai mică din variație, toate contribuind cu mai puțin de **10%** fiecare. De exemplu, PC5 explică **7.2%**, iar PC6 explică **6.3%** din variație.
- Ultimile componente (PC8 - PC10) explică sub **1%** din variație fiecare, ceea ce sugerează că acestea sunt neesențiale și nu adaugă informație semnificativă la analiza generală. Ele tind să capteze doar zgomotul sau variațiile minore ale datelor.

Punctul de inflexiune:

- În scree plot, punctul de inflexiune apare de obicei după primele componente care explică variația majoră. În acest caz, după PC4 se observă o scădere pronunțată în variabilitatea explicată.
- Acest lucru sugerează că primele patru componente sunt suficiente pentru a descrie datele, iar restul componentelor au o contribuție neglijabilă și pot fi ignorate în interpretări suplimentare.

15. Formulați 3 concluzii principale din analiza voastră.

Relevanța componentelor principale în reducerea dimensionalității datelor

În urma analizei componentelor principale, am reușit să reduc complexitatea setului de date prin identificarea unui număr optim de componente care rețin majoritatea variabilității inițiale. Astfel, primele 4 componente principale au capturat esența informației din variabilele economice și de mediu, oferind o reprezentare eficientă a structurilor existente fără a pierde din semnificația generală. Această reducere a dimensionalității îmi permite să analizez tendințele principale din date într-un mod mai concis și focalizat.

Distincția între variabile economice și variabile de mediu

Analiza componentelor principale a scos în evidență o diferențiere clară între variabilele economice (cum ar fi economiile brute și cheltuielile pentru educație) și variabilele de mediu (precum epuizarea resurselor naturale și daunele din emisiile de carbon). Primele două componente principale au fost în mare parte influențate de variabile economice, în timp ce următoarele componente au capturat impactul asupra mediului. Această separare arată că setul de date reflectă două dimensiuni importante – dezvoltarea economică și sustenabilitatea – permitând o analiză mai precăsă a efectelor crizei din 2008 asupra fiecărui aspect.

Corelarea indicatorilor economici și de mediu în contextul crizei din 2008

Constat că indicatorii economici și de mediu prezintă corelații diferite în contextul crizei, reflectând modul în care diferențele țării au abordat prioritățile naționale. Economiile mai dezvoltate au menținut investiții în educație și infrastructură, reușind în același timp să limiteze exploatarea intensă a resurselor naturale. În schimb, țările cu economii mai mici au arătat o dependență mai mare de resursele naturale, ceea ce le-a expus mai mult la impactul de mediu. Aceste observații îmi permit să înțeleg mai bine cum s-au manifestat diversele strategii de redresare economică și protecție a mediului în funcție de contextul fiecărei țări.

ANEXĂ

<u>Figură 1</u>	1
<u>Figură 2</u>	1
<u>Figură 3</u>	2
<u>Figură 4</u>	2
<u>Figură 5</u>	3
<u>Figură 6</u>	5

<u>Figură 7</u>	9
<u>Figură 8</u>	9
<u>Figură 9</u>	9
<u>Figură 10</u>	13
<u>Figură 11</u>	15
<u>Figură 12</u>	15
<u>Figură 13</u>	17
<u>Figură 14</u>	17
<u>Figură 15</u>	19
<u>Figură 16</u>	20
<u>Figură 17</u>	21
<u>Figură 18</u>	21
<u>Figură 19</u>	22
<u>Figură 20</u>	23
<u>Figură 21</u>	23
<u>Figură 22</u>	24
<u>Figură 23</u>	25
<u>Figură 24</u>	26
<u>Figură 25</u>	26
<u>Figură 26</u>	26
<u>Figură 27</u>	26
<u>Figură 28</u>	27
<u>Figură 29</u>	27
<u>Figură 30</u>	28
<u>Figură 31</u>	30
<u>Figură 32</u>	30
<u>Figură 33</u>	30
<u>Figură 34</u>	30
<u>Figură 35</u>	31

Cod Rstudio:

```
#https://databank.worldbank.org/source/world-development-indicators#
```

```
date<-Date
summary(date)

# Afisarea numarului de valori lipsa pe fiecare coloana
colSums(is.na(date))

for (col in names(date)) {
  if (is.numeric(date[[col]])) {
    date[[col]][is.na(date[[col]])] <- mean(date[[col]]), na.rm = TRUE
  }
}

colSums(is.na(date))
```

```

#redenumire date
colnames(date) <- c("Tari", "Chelt_educatie", "Epuizare_energie",
                     "Econ_brute", "Epuizare_minerale",
                     "Epuizare_res_naturale", "Epuizare_neta_paduri", "Econ_nationale_nete",
                     "Daune_dioxid_carbon", "Consum_fix_capital",
                     "Daune_emisii_particule")

#Outlieri
boxplot(date$Econ_brute, col = "lightblue", main="Boxplot Economii brute")

boxplot(date$Econ_brute, plot=FALSE)$out

# Detectează și gestionează outlierii
for (col in names(date)) {
  if (is.numeric(date[[col]])) {
    Q1 <- quantile(date[[col]], 0.25, na.rm = TRUE)
    Q3 <- quantile(date[[col]], 0.75, na.rm = TRUE)
    IQR <- Q3 - Q1

    # Setează valorile extreme la limita IQR-ului (cap pe baza Q1 și Q3)
    date[[col]] <- ifelse(
      date[[col]] < (Q1 - 1.5 * IQR), Q1 - 1.5 * IQR,
      ifelse(date[[col]] > (Q3 + 1.5 * IQR), Q3 + 1.5 * IQR, date[[col]])
    )
  }
}

summary(date)

#indicatorii statistici
install.packages("psych")
library(psych)

indicatori_stat<-describe(date[-1])
View(indicatori_stat)

#Standardizare date
date_std = scale(date[-1], scale=T)

```

```

View(date_std)

#Corelatie si covariantă
cor(date_std)
cov(date_std)

#Scatter plot
install.packages("ggplot2")
library(ggplot2)

ggplot(date, aes(x=Chelt_educatie, y=Econ_brute))+
  geom_point()+
  geom_text(label=date$Tari, color="forestgreen",
            nudge_x=0.25, nudge_y=0.25,
            check_overlap = T)

#histogramme
par(mfrow = c(3, 2))

hist(date$Chelt_educatie, freq = F, col = "plum", main = "Histogramă Cheltuieli Educație")
lines(density(date$Chelt_educatie), lwd = 3, col = "darkblue")

hist(date$Epuizare_energie, freq = F, col = "plum", main = "Histogramă Epuziare Energie")
lines(density(date$Epuizare_energie), lwd = 3, col = "darkblue")

hist(date$Econ_brute, freq = F, col = "plum", main = "Histogramă Economii Brute")
lines(density(date$Econ_brute), lwd = 3, col = "darkblue")

hist(date$Epuizare_minerale, freq = F, col = "plum", main = "Histogramă Epuizare Minerale")
lines(density(date$Epuizare_minerale), lwd = 3, col = "darkblue")

hist(date$Epuizare_res_naturale, freq = F, col = "plum", main = "Histogramă Epuizare Resurse Naturale")
lines(density(date$Epuizare_res_naturale), lwd = 3, col = "darkblue")

hist(date$Epuizare_neta_paduri, freq = F, col = "plum", main = "Histogramă Epuizare Neta Păduri")
lines(density(date$Epuizare_neta_paduri), lwd = 3, col = "darkblue")

par(mfrow = c(2, 3))

```

```

hist(date$Econ_nationale_nete, freq = F, col = "plum", main = "Histogramă Economii Naționale Nete")
lines(density(date$Econ_nationale_nete), lwd = 3, col = "darkblue")

hist(date$Daune_dioxid_carbon, freq = F, col = "plum", main = "Histogramă Daune Dioxid Carbon")
lines(density(date$Daune_dioxid_carbon), lwd = 3, col = "darkblue")

hist(date$Consum_fix_capital, freq = F, col = "plum", main = "Histogramă Consum Fix Capital")
lines(density(date$Consum_fix_capital), lwd = 3, col = "darkblue")

hist(date$Daune_emisii_particule, freq = F, col = "plum", main = "Histogramă Daune Emisii Particule")
lines(density(date$Daune_emisii_particule), lwd = 3, col = "darkblue")

#grafice
par(mfrow = c(3, 2))

plot(date$Chelt_educatie, main = "Cheltuieli Educație",
     xlab = "Țări",
     ylab = "Variabile",
     col = "magenta")

plot(date$Epuizare_energie, main = "Epuizare Energie",
     xlab = "Țări",
     ylab = "Variabile",
     col = "magenta")

plot(date$Econ_brute, main = "Economii Brute",
     xlab = "Țări",
     ylab = "Variabile",
     col = "magenta")

plot(date$Epuizare_minerale, main = "Epuizare Minerale",
     xlab = "Țări",
     ylab = "Variabile",
     col = "magenta")

plot(date$Epuizare_res_naturale, main = "Epuizare Resurse Naturale",
     xlab = "Țări",
     ylab = "Variabile",

```

```

col = "magenta")

plot(date$Epuizare_neta_paduri, main = "Epuizare Neta Păduri",
     xlab = "Țări",
     ylab = "Variabile",
     col = "magenta")

par(mfrow = c(2, 3))

plot(date$Econ_nationale_nete, main = "Economii Naționale Nete",
     xlab = "Țări",
     ylab = "Variabile",
     col = "magenta")

plot(date$Daune_dioxid_carbon, main = "Daune Dioxid Carbon",
     xlab = "Țări",
     ylab = "Variabile",
     col = "magenta")

plot(date$Consum_fix_capital, main = "Consum Capital Fix",
     xlab = "Țări",
     ylab = "Variabile",
     col = "magenta")

plot(date$Daune_emisii_particule, main = "Daune Emisii Particule",
     xlab = "Țări",
     ylab = "Variabile",
     col = "magenta")

#analiza ACP
date_std2 <- scale(date[-1], scale = TRUE)

acp <- prcomp(date_std2, cor = TRUE)

st_dev <- acp$sdev
st_dev

#val proprii = varianta = abaterea standard ^ 2
valp <- st_dev^2

```

```

#procent_info - cât la sută din informație are fiecare componentă
procent_info <- (valp / sum(valp)) * 100
procent_cumulat <- cumsum(procent_info)

rezultate <- data.frame(valp, procent_info, procent_cumulat)
View(rezultate)

#Cele 3 criterii de alegere ale numarului de comp principale

#1. Screen plot
par(mfrow=c(1,1))
screenplot=prcomp(date_std)
plot(screenplot, type="l", main="Criteriul 1 - Screen Plot", col="brown")

```

```

#2. Criteriul lui Kaiser
nr_comp_principale_kaiser<-sum(valp >= 1)
nr_comp_principale_kaiser

```

```

#3. Criteriul procentului de acoperire (procentul cumulat)
proc_cumulat<-data.frame(procent_cumulat)
proc_cumulat

```

```

#Calculați vectorii proprii și forma generală pentru componentele principale reținute în
#analiză.

```

```

# Extragem vectorii proprii din analiza componentelor principale
vectori_proprii <- acp$rotation

```

```

vectori_proprii_retinuti <- vectori_proprii[, 1:nr_comp_principale_kaiser]

```

```

# Afisăm vectorii proprii reținuți
print("Vectorii proprii pentru componentele principale reținute:")
print(vectori_proprii_retinuti)

```

```

# Calculăm forma generală a componentelor principale reținute
# Forma generală este dată de produsul datelor standardizate cu vectorii proprii reținuți
componete_principale_retinute <- date_std2 %*% vectori_proprii_retinuti

```

```

# Afisăm forma generală a componentelor principale
print("Forma generală pentru componentele principale reținute:")
print(componente_principale_retinute)

# Vizualizarea componentelor principale reținute
View(componente_principale_retinute)

# Calculăm scorurile pentru componentele principale
scoruri_componente <- as.data.frame(acp$x)

# Salvăm scorurile ca un nou set de date, adăugând numele țărilor pentru referință
# Presupunem că prima coloană din setul inițial `date` conține numele țărilor
scoruri_componente <- cbind(Tări = date$Tari, scoruri_componente)

# Afisăm noul set de date cu scorurile componentelor principale
print("Noul set de date cu scorurile componentelor principale:")
print(scoruri_componente)

# Salvăm noul set de date în RStudio (de exemplu, ca fișier CSV)
write.csv(scoruri_componente, "scoruri_componente_principale.csv", row.names = FALSE)

# Vizualizăm scorurile în RStudio
View(scoruri_componente)

c=acp$scores[,1:4]
c

#matricea factor
matrice_factor=cor(date_std2, c)
matrice_factor

install.packages("corrplot")
library("corrplot")

corrplot(matrice_factor, method="number")

#cercul corelațiilor

# Setăm par(mfrow) pentru a aranja graficele în 3 rânduri și 2 coloane

```

```

par(mfrow = c(3, 2))

# Plotăm Comp 1 vs Comp 2
plot(vectori_proprii[, 1], vectori_proprii[, 2],
     main = "Comp 1 și Comp 2",
     xlab = "Comp 1", ylab = "Comp 2",
     xlim = c(-1, 1), ylim = c(-1, 1),
     col = "darkblue", pch = 16)
text(vectori_proprii[, 1], vectori_proprii[, 2],
     labels = rownames(vectori_proprii),
     col = "darkred", cex = 0.8)

# Plotăm Comp 1 vs Comp 3
plot(vectori_proprii[, 1], vectori_proprii[, 3],
     main = "Comp 1 și Comp 3",
     xlab = "Comp 1", ylab = "Comp 3",
     xlim = c(-1, 1), ylim = c(-1, 1),
     col = "darkblue", pch = 16)
text(vectori_proprii[, 1], vectori_proprii[, 3],
     labels = rownames(vectori_proprii),
     col = "darkred", cex = 0.8)

# Plotăm Comp 1 vs Comp 4
plot(vectori_proprii[, 1], vectori_proprii[, 4],
     main = "Comp 1 și Comp 4",
     xlab = "Comp 1", ylab = "Comp 4",
     xlim = c(-1, 1), ylim = c(-1, 1),
     col = "darkblue", pch = 16)
text(vectori_proprii[, 1], vectori_proprii[, 4],
     labels = rownames(vectori_proprii),
     col = "darkred", cex = 0.8)

# Plotăm Comp 2 vs Comp 3
plot(vectori_proprii[, 2], vectori_proprii[, 3],
     main = "Comp 2 și Comp 3",
     xlab = "Comp 2", ylab = "Comp 3",
     xlim = c(-1, 1), ylim = c(-1, 1),
     col = "darkblue", pch = 16)
text(vectori_proprii[, 2], vectori_proprii[, 3],

```

```

labels = rownames(vectori_proprii),
col = "darkred", cex = 0.8)

# Plotăm Comp 2 vs Comp 4
plot(vectori_proprii[, 2], vectori_proprii[, 4],
main = "Comp 2 și Comp 4",
xlab = "Comp 2", ylab = "Comp 4",
xlim = c(-1, 1), ylim = c(-1, 1),
col = "darkblue", pch = 16)
text(vectori_proprii[, 2], vectori_proprii[, 4],
labels = rownames(vectori_proprii),
col = "darkred", cex = 0.8)

# Plotăm Comp 3 vs Comp 4
plot(vectori_proprii[, 3], vectori_proprii[, 4],
main = "Comp 3 și Comp 4",
xlab = "Comp 3", ylab = "Comp 4",
xlim = c(-1, 1), ylim = c(-1, 1),
col = "darkblue", pch = 16)
text(vectori_proprii[, 3], vectori_proprii[, 4],
labels = rownames(vectori_proprii),
col = "darkred", cex = 0.8)

#repr scoruri principale grafic
library(ggplot2)

par(mfrow=c(3,2))
# Comp.1 vs Comp.2
ggplot(scoruri_componente, aes(x=PC1, y=PC2)) +
geom_point(shape=16, size=4, col="forestgreen") +
geom_text(aes(label=Tări), nudge_x=0.25, nudge_y=0.25, check_overlap=TRUE) +
ggtitle("PC1 vs PC2")

# Comp.1 vs Comp.3
ggplot(scoruri_componente, aes(x=PC1, y=PC3)) +
geom_point(shape=16, size=4, col="forestgreen") +
geom_text(aes(label=Tări), nudge_x=0.25, nudge_y=0.25, check_overlap=TRUE) +
ggtitle("PC1 vs PC3")

```

```

# Comp.1 vs Comp.4
ggplot(scoruri_componente, aes(x=PC1, y=PC4)) +
  geom_point(shape=16, size=4, col="forestgreen") +
  geom_text(aes(label=Tări), nudge_x=0.25, nudge_y=0.25, check_overlap=TRUE) +
  ggtitle("PC1 vs PC4")

# Comp.2 vs Comp.3
ggplot(scoruri_componente, aes(x=PC2, y=PC3)) +
  geom_point(shape=16, size=4, col="forestgreen") +
  geom_text(aes(label=Tări), nudge_x=0.25, nudge_y=0.25, check_overlap=TRUE) +
  ggtitle("PC2 vs PC3")

# Comp.2 vs Comp.4
ggplot(scoruri_componente, aes(x=PC2, y=PC4)) +
  geom_point(shape=16, size=4, col="forestgreen") +
  geom_text(aes(label=Tări), nudge_x=0.25, nudge_y=0.25, check_overlap=TRUE) +
  ggtitle("PC2 vs PC4")

# Comp.3 vs Comp.4
ggplot(scoruri_componente, aes(x=PC3, y=PC4)) +
  geom_point(shape=16, size=4, col="forestgreen") +
  geom_text(aes(label=Tări), nudge_x=0.25, nudge_y=0.25, check_overlap=TRUE) +
  ggtitle("PC3 vs PC4")

#biploturi
par(mfrow = c(2, 3))
# PC1 vs PC2
biplot(acp, choices = c(1, 2), cex = c(0.7, 0.9))
title("PC1 vs PC2")

# PC1 vs PC3
biplot(acp, choices = c(1, 3), cex = c(0.7, 0.9))
title("PC1 vs PC3")

# PC1 vs PC4
biplot(acp, choices = c(1, 4), cex = c(0.7, 0.9))
title("PC1 vs PC4")

```

```

# PC2 vs PC3
biplot(acp, choices = c(2, 3), cex = c(0.7, 0.9))
title("PC2 vs PC3")

# PC2 vs PC4
biplot(acp, choices = c(2, 4), cex = c(0.7, 0.9))
title("PC2 vs PC4")

# PC3 vs PC4
biplot(acp, choices = c(3, 4), cex = c(0.7, 0.9))
title("PC3 vs PC4")

install.packages("FactoMineR")
install.packages("factoextra")
library(FactoMineR)
library(factoextra)

cp <- PCA(date[-1], graph = FALSE)

# Vizualizarea contribuției variabilelor la componentele principale
fviz_pca_var(cp, col.var = "contrib", gradient.cols = c("red", "yellow", "lightgreen"), repel = TRUE,
             title = "Contribuția variabilelor la componentele principale")

# Vizualizarea contribuțiilor observațiilor la componentele principale
fviz_contrib(cp, choice = "ind", axes = 1:4,
             title = "Contribuția observațiilor la componentele principale")

# Vizualizarea calității reprezentării variabilelor (cos2)
fviz_pca_var(cp, col.var = "cos2", gradient.cols = c("red", "yellow", "lightgreen"), repel = TRUE,
             title = "Calitatea reprezentării variabilelor (cos2)")

# Vizualizarea calității reprezentării observațiilor (cos2)
fviz_pca_ind(cp, col.ind = "cos2", gradient.cols = c("red", "yellow", "green"), repel = TRUE,
              title = "Calitatea reprezentării observațiilor (cos2)")

# Scree plot pentru a vizualiza variația explicată de fiecare componentă principală
fviz_eig(cp, addlabels = TRUE, ylim = c(0, 50), title = "Scree Plot")

```


TEMA 2-Analiza datelor

I. Analiza factorială

1. Căutați un set de date din domeniul economic cu cel puțin 10 variabile (coloane) – pentru nota maximă, este recomandat ca setul de date să crească în dimensiuni.

Pentru analiza ce va fi prezentată pe parcursul temei mele, am creat o bază de date in Excel care are 10 variabile și 39 de observații.

2. Curățați setul de date, redenumiți variabilele astfel încât să fie clar de identificat și gestionări valorile lipsă și outlierii.

- Pentru început, am afișat numărul de valori lipsă de pe fiecare coloană în parte.

```
> # Afisarea numarului de valori lipsa pe fiecare coloana
> colsums(is.na(date))
Tari   x1   x2   x3   x4   x5   x6   x7   x8   x9   x10
  0     1     1     1     0     2     1    1     0     0      4
> |
```

Figură 1

- Apoi, pentru a înlocui valorile lipsă cu media fiecărei coloane am folosit un cod care verifică fiecare coloană și, dacă este numerică, înlocuiește valorile lipsă cu media acelei coloane.

```
for (col in names(date)) {
  if (is.numeric(date[[col]])) {
    date[[col]][is.na(date[[col]])] <- mean(date[[col]]), na.rm = TRUE)
  }
}
```

Figură 2

- La final am mai făcut încă o dată verificarea pentru a mă asigura că au fost eliminate cu succes valorile lipsă.

```
colsums(is.na(date))
```

Figură 3

```
> colsums(is.na(date))
Tari   x1   x2   x3   x4   x5   x6   x7   x8   x9   x10
  0     0     0     0     0     0     0     0     0     0      0
```

Figură 4

- Apoi am redenumit variabilele

```
#redenumire date
colnames(date) <- c("Tari", "Chelt_educatie", "Epuizare_energie",
"Chelt_brute", "Epuizare_minerale",
"Epuizare_res_naturale", "Epuizare_neta_paduri", "Econ_nationale_nete",
"Daune_dioxid_carbon", "Consum_fix_capital",
"Daune_emisii_particule")|
```

Figură 5

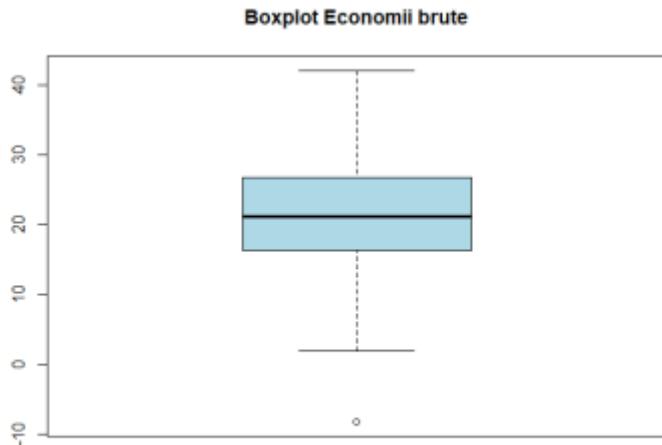
Tari	Chelt_educatie	Epuizare_energie	Chelt_brute	Epuizare_minerale	Epuizare_res_naturale	Epuizare_neta_paduri	Econ_nationale_nete	Daune_dioxid_carbon
1 Armenia	2.740000	0.000000e+00	26.343387	0.000000e+00	0.251529331	0.2515293313	15.1817034	1.2361594
2 Austria	5.070000	1.130848e-01	28.463932	4.359406e-03	0.175746515	0.0583043504	11.9874668	0.4357187
3 Anglia	4.810000	1.239511e+00	13.886603	5.514540e-06	1.827006251	0.0000000000	0.6473100	0.4491197
4 Albania	3.067961	6.685169e-01	17.841874	4.377673e-02	0.761575640	0.0636359830	5.6233713	0.8457010
5 Belgia	5.950000	0.000000e+00	26.776078	0.000000e+00	0.017505606	0.0175056059	8.4645339	0.5415532
6 Belarus	4.631000	9.532209e-01	30.902789	0.000000e+00	0.353222549	0.0000000000	16.0749578	2.7095882
7 Boznia Herzegovina	0.000000	3.264749e-01	11.942212	8.348249e-02	0.412805297	0.0228527757	0.2994363	2.8111986
8 Bulgaria	4.000000	4.554732e-01	15.5335724	3.162492e-01	0.783722394	0.0000000000	2.2991323	2.5902476
9 Croatia	4.220000	7.428071e-01	20.362520	0.000000e+00	0.883794416	0.1434673338	4.1103668	0.6495303
10 Cipru	6.040000	1.210560e-02	14.429634	1.780763e-03	0.013814386	0.0000000000	2.7976389	0.4319326
11 Cehia	3.590000	4.533549e-01	26.885599	0.000000e+00	0.477737794	0.0241502052	7.9017290	1.4277543
12 Danemarca	7.060000	1.8335166e-00	26.391760	4.050171e-03	1.841740488	0.0044958844	8.6238505	6.3765373
13 Egipt	4.820000	8.086604e-04	31.190260	0.000000e+00	0.061625934	0.0000172739	6.9412554	0.2263827
14 Estonia	5.360000	1.522018e+00	24.734924	0.000000e+00	1.132017834	0.0000000000	8.9742252	2.1465946
15 Finlanda	5.440000	0.000000e+00	26.222197	2.312790e-02	0.023127905	0.0000000000	10.3236299	0.5313238
16 Irlanda	4.690000	1.409835e-02	22.941221	5.081647e-04	0.014606517	0.0000000000	6.1424612	0.1209620
17 Germania	4.250000	1.365050e-01	26.883389	1.124669e-04	0.138617516	0.0000000000	9.7429398	0.5605529
18 Grecia	3.111830	1.253835e-01	10.410591	7.815253e-03	0.142420336	0.0000266033	5.0773552	0.7803080

Figură 6

- Ulterior am făcut același lucru și pentru outlieri. Am început prin a exemplifica modul în care aceștia pot fi afișați (prin utilizarea boxploturilor și prin afișarea lor în consolă).

```
#outlieri
boxplot(date$Econ_brute, col = "lightblue", main="Boxplot Economii brute")
boxplot(date$Econ_brute, plot=FALSE)$out
```

Figură 7



Figură 8

```
> boxplot(date$Econ_brute, plot=FALSE)$out
[1] -8.169733
```

- Outlierii pot fi identificați și tratați în funcție de distribuția datelor. O metodă comună este folosirea IQR-ului (Abaterea intercuartilică):

```
# Detectează și gestionează outlierii
for (col in names(date)) {
  if (is.numeric(date[[col]])) {
    Q1 <- quantile(date[[col]], 0.25, na.rm = TRUE)
    Q3 <- quantile(date[[col]], 0.75, na.rm = TRUE)
    IQR <- Q3 - Q1

    # Setează valorile extreme la limita IQR-ului (cap pe baza Q1 și Q3)
    date[[col]] <- ifelse(
      date[[col]] < (Q1 - 1.5 * IQR), Q1 - 1.5 * IQR,
      ifelse(date[[col]] > (Q3 + 1.5 * IQR), Q3 + 1.5 * IQR, date[[col]])
    )
  }
}
```

Figură 9

values	
col	"Daune_emisii_particule"
IQR	Named num 7.05
Q1	Named num 6.26
Q3	Named num 13.3

- Apoi am făcut o ultimă verificare pentru a mă asigura că valorile outlierilor au fost înlocuite

```

summary(date)

## R 4.2.3 -- "Normal" --
## 
## Attaching package: 'dplyr'
## The following objects are masked from 'base':
## 
##     filter, lag
## 
## The following objects are masked from 'tidyverse':
## 
##     c_across, c_across_
## 
## 
##   Tari      chelt_educatie    epuizare_energie    Econ_brute    epuizare_minerale    epuizare_res_naturale    epuizare_neta_paduri
##   <fct>      <dbl>           <dbl>           <dbl>           <dbl>           <dbl>           <dbl>           <dbl>
##   Length:39   Min. :2.317   Min. :0.00000   Min. : 0.5385   Min. :0.0000000   Min. :0.00000   Min. :0.00000
##   Class :character  1st Qu.:4.215   1st Qu.:0.01401   1st Qu.:16.3163   1st Qu.:0.0000000   1st Qu.:0.06387   1st Qu.:0.00000
##   Mode  :character  Median :4.760   Median :0.12538   Median :21.2404   Median :0.0005082   Median :0.20153   Median :0.00000
##                   Mean  :4.735   Mean  :0.41719   Mean  :21.3383   Mean  :0.0114921   Mean  :0.50295   Mean  :0.01265
##                   3rd Qu.:5.480   3rd Qu.:0.65683   3rd Qu.:26.8347   3rd Qu.:0.0169368   3rd Qu.:0.77269   3rd Qu.:0.02039
##                   Max. :17.378   Max. :1.62105   Max. :42.0754   Max. :0.0423420   Max. :1.83591   Max. :0.05097
## 
##   Econ_nationale_nete daune_dioxid_carbon consum_fix_capital daune_emisif_particule
##   <dbl>           <dbl>           <dbl>           <dbl>
##   Min. :-6.979   Min. :0.1292   Min. :10.75   Min. :-4.314
##   1st Qu.: 2.601  1st Qu.:0.4949   1st Qu.:13.27   1st Qu.: 6.256
##   Median : 5.901  Median :0.7372   Median :15.49   Median : 9.528
##   Mean  : 5.746  Mean  :0.0198   Mean  :15.51   Mean  : 9.398
##   3rd Qu.: 8.988  3rd Qu.:1.3320   3rd Qu.:17.10   3rd Qu.:13.303
##   Max. :18.568   Max. :2.5875   Max. :22.85   Max. :123.873

```

Figură 10

3. Descrieți în cuvinte setul de date, din perspectiva variabilelor (coloanelor) și observațiilor (liniilor). Enunțați clar sursa datelor și perioada de timp la care se referă.

Setul de date furnizat conține informații referitoare la mai multe țări din Europa și include o serie de indicatori economici și de mediu exprimați ca procente din Venitul Național Brut (VNB). Fiecare linie (observație) din setul de date corespunde unei țări, iar fiecare coloană (variabilă) reprezintă un indicator specific, descris în detaliu mai jos. Setul de date este structurat în următorii termeni:

Variabile (coloane):

- **Economii ajustate: cheltuieli pentru educație (% din VNB)**
Acest indicator reprezintă procentul din venitul național brut (VNB) alocat pentru educație. Este considerat o investiție în capitalul uman, având potențialul de a îmbunătăți pe termen lung productivitatea și bunăstarea economică a unei țări.
- **Economii ajustate: epuizarea energiei (% din VNB)**
Acest indicator măsoară procentul din VNB pierdut din cauza exploatarii resurselor energetice neregenerabile, precum petrolul, gazele naturale și cărbunele. Reflectă impactul pe care exploatarea resurselor energetice îl are asupra bogăției naturale a țării.
- **Economii ajustate: economii brute (% din VNB)**
Economiile brute reprezintă procentul din VNB economisit de o țară, înainte de ajustări pentru degradarea mediului și consumul de capital fix. Acestea oferă o imagine de ansamblu asupra economisirii la nivel de economie națională.
- **Economii ajustate: epuizarea mineralelor (% din VNB)**
Acest indicator arată procentul din VNB pierdut din cauza extracției de resurse minerale (cum ar fi cuprul, aurul, minereurile de fier etc.). Este o măsură a degradării capitalului natural datorită exploatarii resurselor minerale neregenerabile.
- **Economii ajustate: epuizarea resurselor naturale (% din VNB)**
Acest indicator cumulează pierderile din exploatarea resurselor energetice, minerale și forestiere. Exprimă cât din VNB se pierde prin utilizarea resurselor naturale ale țării, fără a ține cont de regenerarea lor.
- **Economii ajustate: epuizarea netă a pădurilor (% din VNB)**
Aceasta reprezintă procentul din VNB pierdut din cauza defrișării pădurilor, peste nivelul

considerat sustenabil. O reducere a acoperirii forestiere poate duce la pierderi de biodiversitate, eroziune a solului și alte daune ecologice.

- **Economii ajustate: economii naționale nete (% din VNB)**

Economiile naționale nete sunt obținute din economiile brute, la care se adaugă cheltuielile pentru educație și se scad epuizarea resurselor naturale și degradarea mediului. Este o măsură a economiilor reale ale unei țări, având în vedere resursele și costurile de mediu.

- **Economii ajustate: daune din dioxid de carbon (% din VNB)**

Acest indicator estimează costul pierderilor economice din cauza emisiilor de dioxid de carbon, măsurat ca procent din VNB. Poluarea cu CO₂ contribuie la schimbările climatice și poate afecta sănătatea și bunăstarea economică a populației.

- **Economii ajustate: consumul de capital fix (% din VNB)**

Consum de capital fix reprezintă valoarea deprecierii capitalului fizic (clădiri, echipamente, infrastructură), exprimată ca procent din VNB. Este necesară menținerea unui anumit nivel de capital pentru a susține activitatea economică, iar acest indicator reflectă costurile necesare pentru întreținerea acestui capital.

- **Economii nete ajustate, inclusiv daunele cauzate de emisiile de particule (% din VNB)**

Acest indicator include economiile nete ajustate, dar ia în calcul și daunele produse de emisiile de particule (PM2.5, PM10), care afectează sănătatea publică și mediul. Particulele fine din aer pot provoca boli respiratorii și alte probleme de sănătate, cu impact economic indirect asupra unei țări.

Observații (linii):

Fiecare linie reprezintă o țară, pentru care sunt colectate date pentru toți cei 10 indicatori menționați mai sus. Astfel, fiecare observație conține un profil complet al unei țări din punctul de vedere al economiilor și impactului asupra mediului, fiind valoros pentru analizarea comparativă între țări.

Sursa datelor și perioada de timp:

Acste date sunt obținute de pe site-ul oficial la Băncii Mondiale (<https://databank.worldbank.org/source/world-development-indicators#>). Datele sunt alese pentru anul 2008 deoarece atunci a avut loc criza mondială, lucru ce va fi relevant pentru analiza mea.

4. Descrieți obiectivul general al analizei voastre.

Obiectivul principal al analizei mele a fost să evaluatez impactul crizei economice globale din 2008 asupra economiilor ajustate, analizând indicatori esențiali de sustenabilitate economică și de mediu. Mi-am propus să înțeleg în ce măsură acest soc financiar global a influențat economiile naționale, investigând factori precum investițiile în educație, epuizarea resurselor naturale, economiile brute și nete, precum și impactul ecologic, cum ar fi emisiile de carbon și defrișările.

Prin această analiză, am dorit să ofer o imagine detaliată a modului în care diferite țări au rezisțit și gestionat consecințele crizei, explorând ajustările economice realizate pentru a susține dezvoltarea pe termen lung. Scopul meu a fost să evidențiez posibilele transformări structurale în economii și măsurile de sustenabilitate adoptate de guverne pentru a face față presiunilor economice și ecologice cauzate de criza din 2008, față de anii precedenți.

5. . Calculați indicatorii statistici și interpretați din punct de vedere economic.

```
#indicatorii statistici
install.packages("psych")
library(psych)

indicatori_stat<-describe(date[-1])
view(indicatori_stat)
```

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se	se
Chelt_educatie	1	39	4.75318363	1.16730854	4.760000e+00	4.737262472	0.9785160000	2.3175000	7.3775000	5.06000000	-0.05023855	-0.16471898	0.186918961	
Epuizare_energie	2	39	0.41719193	0.51716531	1.253835e-01	0.347109390	0.1858835471	0.0000000	1.62104505	1.62104505	1.10946421	-0.08533629	0.062812727	
Econ_brute	3	39	21.33627211	8.66032211	2.124038e+01	21.480987197	0.3811542012	0.5385176	42.07538689	41.53886931	-0.13905539	0.17408278	1.386761391	
Epuizare_minerale	4	39	0.01149213	0.011758859	5.081647e-04	0.009712364	0.0007534051	0.0000000	0.04234197	0.04234197	1.08641714	-0.72625925	0.002816429	
Epuizare_res_naturale	5	39	0.50294507	0.542486318	2.015294e-01	0.495122973	0.2967874162	0.0000000	1.83591466	1.83591466	1.02673922	-0.04520797	0.0868668031	
Epuizare_neta_paturi	6	39	0.01265391	0.01966492	0.000000e+00	0.010520600	0.0000000000	0.0000000	0.05097423	0.05097423	1.10459954	-0.29277457	0.003148907	
Econ_nationale_nele	7	39	5.74399014	6.51211752	3.900833e+00	5.739311834	4.9892735118	-6.9789276	18.56794019	25.54686778	-0.08449597	0.32174893	1.042773356	
Douane_dioxid_carbon	8	39	1.01988147	0.79363949	7.572209e-01	0.952705320	0.4766591005	0.1252066	2.587480292	2.458827635	1.04082394	-0.34456741	0.127116055	
Consum_fix_capital	9	39	15.51305522	3.02506663	1.548815e+01	15.341547504	2.7420046635	10.7517761	22.85217945	12.10040333	0.52502071	-0.25054942	0.484396334	
Douane_emaili_particule	10	39	9.39706288	6.31790947	9.528205e+00	9.317967697	5.4574905210	-4.3136872	23.87540534	20.18709255	0.04487572	-0.14552468	1.043700610	

Figură 11

1. Cheltuieli pentru educație

- Mean (Media): 4.75318 - Media cheltuielilor pentru educație în 2008, exprimată ca procent din VNB, indică un angajament relativ scăzut al țărilor față de investițiile în educație din cauza crizei economice. Scăderea veniturilor bugetare și prioritizarea cheltuielilor esențiale au redus posibilitățile de investiții în educație.
- SD (Abaterea standard): 1.167 - Variabilitatea între țări este moderată, sugerând diferențe în prioritățile bugetare și în gradul de afectare de criză.
- Mediană: 4.76 - Valoarea mediană este aproape de media aritmetică, sugerând o distribuție simetrică și că majoritatea țărilor au cheltuit în jurul acestui procent.
- Skewness (Asimetrie): -0.0528 - Aproape simetric, cu o ușoară tendință spre valori mai mici, ceea ce indică faptul că un număr mic de țări au redus cheltuielile drastic.
- Kurtosis (Curtosis): -0.1647 - Curtosis negativă, arătând o distribuție mai plată decât cea normală; adică, țările nu au prezentat valori extreme în cheltuielile educaționale, dar s-au concentrat în jurul mediei.

2. Epuizarea energiei

- Mean: 0.417 - Media epuizării resurselor energetice ca procent din VNB este relativ mică, dar chiar și aceste pierderi minore sunt semnificative într-un context de criză.
- SD: 0.517 - Abaterea standard relativ ridicată arată o variabilitate mare între țări, sugerând că dependența de resursele energetice neregenerabile a afectat țările diferit.
- Mediană: 1.253 - Semnificativ mai mare decât media, ceea ce indică o distribuție asimetrică.
- Skewness: 1.109 - Asimetrie pozitivă ridicată, sugerând că există câteva țări cu epuizare energetică mare. Criza ar fi amplificat aceste valori din cauza cererii mai reduse și a prețurilor în scădere.
- Kurtosis: -0.385 - Distribuția este plată, ceea ce indică valori extreme mai rare, dar în contextul crizei, aceste valori extreme au probabil un impact economic negativ.

3. Economii brute

- Mean: 21.34 - Media economiilor brute este relativ ridicată, însă această valoare este afectată de criză, deoarece multe țări și-au redus economiile pentru a susține cheltuielile esențiale.
- SD: 8.66 - Abaterea standard ridicată indică variații mari între țări; țările mai dezvoltate au menținut economii mai mari, iar cele mai afectate au redus drastic.
- Mediană: 21.24 - Aproape de media aritmetică, ceea ce sugerează o distribuție relativ simetrică.
- Skewness: -0.15 - Ușoară asimetrie negativă, indicând că un număr de țări au economii sub media normală din cauza crizei.
- Kurtosis: 0.17 - Curtosis pozitivă, indicând valori concentrate în jurul mediei, ceea ce sugerează că majoritatea țărilor au avut economii în jurul acestei valori.

4. Epuizarea mineralelor

- Mean: 0.104 - Media epuizării resurselor minerale este mică, dar importantă în contextul crizei, când cererea scăzută a dus la reducerea prețurilor.
- SD: 0.157 - Abaterea standard ridicată indică o variabilitate între țări în funcție de dependența de minerale.
- Mediană: 0.05 - Ușor sub media aritmetică, sugerând că multe țări au valori reduse.
- Skewness: 1.86 - Asimetrie mare pozitivă, indicând că doar câteva țări au înregistrat epuizări semnificative.
- Kurtosis: -0.72 - Curtosis negativă, sugerând că există puține valori extreme.

5. Epuizarea resurselor naturale

- Mean: 0.502 - Epuizarea medie a resurselor naturale este mică, sugerând o conservare relativă.
- SD: 0.54 - O variabilitate moderată între țări.
- Mediană: 0.51 - Aproape de media aritmetică, ceea ce sugerează o distribuție relativ simetrică.
- Skewness: 1.02 - Ușor asimetrică, cu o tendință spre valori mai mari.
- Kurtosis: -0.26 - Distribuție mai plată, indicând o dispersie fără extreme.

6. Epuizarea netă a pădurilor

- Mean (Media): 0.0126 - Media epuizării nete a pădurilor este de 0.0126% din VNB, o valoare relativ mică, dar care poate fi semnificativă din punct de vedere ecologic. În contextul crizei economice, acest procent ar putea indica o încetinire a defrișărilor masive, deoarece activitățile economice din industriile forestiere ar fi fost afectate de scăderea cererii globale și de constrângerile bugetare. Totuși, în unele țări, presiunea pentru exploatarea resurselor naturale ar fi putut crește pentru a genera venituri.
- SD (Abaterea standard): 0.01 - Abaterea standard mică arată o variabilitate redusă între țări, ceea ce sugerează că epuizarea pădurilor a fost relativ uniformă, fără discrepanțe majore între țări.
- Mediană: 0.0064 - Valoarea medianei fiind aproape de zero sugerează că majoritatea țărilor nu au epuizat pădurile în mod semnificativ. Acest lucru este relevant, având în vedere impactul ecologic, deoarece defrișările ar fi putut contribui la probleme suplimentare de mediu.
- Skewness (Asimetrie): 1.184 - Asimetria pozitivă ridicată sugerează că există câteva țări care au avut valori mai mari ale epuizării pădurilor, indicând că acestea au continuat să exploateze resursele forestiere la un nivel mai ridicat decât media, posibil din cauza necesităților economice.
- Kurtosis (Curtosis): -0.291 - Curtosis negativă indică o distribuție mai plată decât cea normală, sugerând o lipsă de valori extreme. Acest lucru arată că majoritatea țărilor au valori reduse și relativ uniforme pentru epuizarea pădurilor.

7. Economii naționale nete

- Mean (Media): 5.754 - Media economiilor naționale nete este de 5.75% din VNB, o valoare relativ scăzută în contextul crizei economice. Aceasta indică faptul că multe țări au fost nevoite să reducă economiile nete pentru a finanța cheltuielile curente și a combate impactul negativ al crizei. Nivelurile scăzute de economii naționale nete arată presiunea exercitată de criză asupra economiilor, țările fiind nevoite să utilizeze o mare parte din resursele lor financiare.
- SD (Abaterea standard): 6.541 - O abatere standard relativ mare indică variații substanțiale între țări, sugerând că unele state au avut capacitatea de a menține economii mai mari, în timp ce altele au fost puternic afectate și au avut economii negative.
- Mediană: 5.990 - Aproape de media aritmetică, ceea ce indică o distribuție destul de simetrică a economiilor nete între țări.

- Skewness (Asimetrie): -0.084 - Aproape simetrică, ceea ce sugerează că nu există o tendință pronunțată spre valori extrem de mari sau mici.
- Kurtosis (Curtosis): -0.321 - Curtosis negativă indică o distribuție plată, fără valori extreme. Aceasta reflectă faptul că majoritatea țărilor au avut economii nete relativ apropriate de media.

8. Daunele din emisiile de dioxid de carbon

- Mean (Media): 1.098 - Costurile economice ale emisiilor de dioxid de carbon au avut o medie de 1.098% din VNB, ceea ce reflectă impactul economic negativ al poluării asupra economiilor în 2008. În contextul crizei, această valoare sugerează că, în ciuda scăderii activității economice globale, emisiile de CO₂ au continuat să provoace daune, afectând sănătatea publică și mediul. Aceste costuri sunt semnificative pentru țările care depind de industriile poluante.
- SD (Abaterea standard): 0.723 - Variabilitatea relativ ridicată indică diferențe între țări în ceea ce privește impactul emisiilor de dioxid de carbon, posibil din cauza nivelurilor diferite de industrializare și a măsurilor de control al poluării.
- Mediană: 0.572 - Mediana mai mică decât media sugerează că mai multe țări au costuri ale daunelor sub media, dar câteva țări industrializate au avut valori ridicate, contribuind la media generală mai mare.
- Skewness (Asimetrie): 1.048 - Asimetria pozitivă arată că există țări cu valori mari ale daunelor din emisiile de dioxid de carbon, ceea ce este caracteristic pentru țările industrializate, mai ales în timpul crizei, când controlul emisiilor ar fi putut să nu fie o prioritate.
- Kurtosis (Curtosis): -0.344 - Curtosis negativă indică o distribuție plată, sugerând o dispersie a valorilor fără extreme semnificative, ceea ce sugerează că majoritatea țărilor au valori în jurul mediei sau mai mici.

9. Consum de capital fix

- Mean (Media): 15.51 - Media consumului de capital fix este relativ mare, indicând o depreciere semnificativă a infrastructurii și echipamentelor, ceea ce a fost probabil amplificat de criză. Într-o perioadă de criză, menținerea și reînnoirea capitalului fix poate fi limitată, ceea ce duce la depreciere.
- SD (Abaterea standard): 3.205 - Variabilitatea moderată între țări sugerează că nivelurile de depreciere a capitalului fix sunt relativ uniforme, dar există diferențe notabile între statele cu economii mai dezvoltate și cele în curs de dezvoltare.
- Mediană: 15.54 - Aproape de media aritmetică, indicând o distribuție simetrică și că majoritatea țărilor au avut niveluri similare de consum de capital fix.
- Skewness (Asimetrie): 0.522 - Asimetria ușor pozitivă arată că există câteva țări cu valori mai mari ale consumului de capital fix, probabil cele care depind de industrii mari, mai afectate de criză.
- Kurtosis (Curtosis): -0.25 - Curtosis negativă sugerează o distribuție relativ plată, ceea ce indică absența unor valori extreme și concentrarea valorilor în jurul mediei.

10. Daunele din emisiile de particule

- Mean (Media): 9.397 - Costurile economice ale daunelor cauzate de emisiile de particule au o valoare medie ridicată, de 9.397% din VNB. Acest procent semnificativ reflectă impactul poluării cu particule asupra sănătății publice și mediului, care a devenit o problemă mai gravă în timpul crizei, când investițiile în tehnologii de reducere a poluării au fost limitate.
- SD (Abaterea standard): 6.517 - Abaterea standard ridicată sugerează variații mari între țări, arătând că impactul poluării aerului diferă semnificativ în funcție de nivelul de industrializare și de politicile de mediu.

- Mediană: 9.528 - Aproape de media aritmetică, ceea ce sugerează o distribuție simetrică, majoritatea ţărilor având daune economice similare din cauza poluării cu particule.
- Skewness (Asimetrie): 0.044 - Aproape simetrică, ceea ce indică faptul că valorile sunt distribuite în mod echilibrat între țări, fără tendințe majore către extreame.
- Kurtosis (Curtosis): -0.145 - Curtosis negativă indică o distribuție mai plată decât cea normală, ceea ce sugerează că valorile sunt relativ dispersate, fără extreame majore.

6. Calculați matricea de corelație și covarianță și interpretați din punct de vedere economic. Reprezentați grafic matricea de corelație.

Pentru a calcula matricea de corelație și covarianță trebuie în primul rând să standardizăm datele deoarece ne ajută să ajustăm toate variabilele pentru a avea o medie de 0 și o deviație standard de 1, eliminând influența diferențelor de scară și făcând variabilele comparabile. De asemenea, standardizarea datelor este un pas esențial în analiza PCA pe care am făcut-o la pașii următori.

```
#Standardizare date
date_std = scale(date[-1], scale=T)
View(date_std)
```

	Chelt_educatie	Epuizare_energie	Econ_brute	Epuizare_minerale	Epuizare_res_naturale	Epuizare_neta_paduri	Econ_nationale_nete	Daune_dioxid_carbon	Consum_fix_capital
1	-1.70921703	-0.80668969	0.57816729	-0.65338675	-0.55562234	1.94066411	1.44507643	0.27254621	-1.431628133
2	0.29882765	-0.58802700	0.82392480	-0.40553253	-0.40314999	1.94886411	0.95844043	-0.73578934	0.319475617
3	0.06409192	1.36576488	-0.85116377	-0.69327157	1.34716315	-0.64347645	-0.58462103	-0.63236178	-0.161360936
4	-1.42326228	0.48595646	-0.42558904	1.75396746	0.51377181	1.94066411	-0.01682934	-0.21001383	-1.155198860
5	1.04609861	-0.80668969	0.62812873	-0.65338675	-0.89407087	0.26632706	0.41745926	-0.80344711	0.825099077
6	-0.00582446	1.03147388	1.10463759	-0.65338675	0.83003065	-0.64347640	1.71226600	1.97400910	-0.496125752
7	-2.07116075	-0.21408609	-1.18894628	1.75396746	-0.16154025	0.51983258	-0.83715420	1.97400910	-1.371941895
8	0.62861088	0.07019413	0.66976435	1.75396746	0.46071042	-0.64347640	-0.48323111	1.97400910	-0.851704712
9	-0.44134215	0.62478118	-0.11012889	-0.65338675	0.70294820	1.94066411	-0.25116582	-0.21449045	0.232995427
10	1.11779904	-0.78328200	-0.79730443	-0.59523495	-0.90163137	-0.64347640	-0.49374847	-0.25045450	-1.282966904
11	-0.96104632	0.06992511	0.86945114	-0.65338675	-0.06465647	0.59644878	0.33134212	0.51399618	1.001353670
12	1.89160402	2.12778710	0.60388470	-0.42315412	2.45716298	-0.41495547	0.44232107	-0.81021168	0.810843046
13	0.07265977	-0.80512405	1.13582207	-0.65338675	-0.80411923	-0.60195940	0.18355062	-0.9947005	2.426103339
14	0.53526245	1.74158738	0.39243843	-0.65338675	1.50983534	-0.64347640	0.49374709	1.41383191	0.061664491
15	0.60379612	0.80668969	0.79311183	0.66155117	-0.89446030	-0.64347640	0.70294818	-0.81533554	0.786581470
16	0.12408008	-0.77942887	0.18552230	-0.62489502	-0.90019114	-0.64347640	0.36286204	-0.83642825	0.425017591
17	-0.41564301	-0.53887188	0.64167551	-0.69699244	-0.67159239	-0.64347640	0.81346401	-0.57851562	0.541930379
18	-1.39068082	-0.56424596	-1.26157910	-0.20916394	-0.66457680	-0.17428532	-1.46206235	-0.30157861	-0.008234276

Figură 12

	Chelt_educatie	Epuizare_energie	Econ_brute	Epuizare_minerale	Epuizare_res_naturale	Epuizare_neta_paduri	Econ_nationale_nete	Daune_dioxid_carbon	Consum_fix_capital
chelt_educatie	1.00000000	0.03112266	0.05100060	-0.13511192	0.04286678	-0.31812298			
Epuizare_energie	0.03112266	1.00000000	0.14168679	-0.02835887	0.96871603	-0.10049274			
Econ_brute	0.05100060	0.14168679	1.00000000	-0.17311413	0.11638968	-0.09175818			
Epuizare_minerale	-0.13511192	-0.02835887	-0.17311413	1.00000000	0.02812303	0.09076454			
Epuizare_res_naturale	-0.04286678	0.96871603	0.11638968	0.02812303	1.00000000	-0.08856299			
Epuizare_neta_paduri	-0.31812298	-0.19049274	-0.09175818	0.09076454	-0.08856299	1.00000000			
Econ_nationale_nete	0.03128481	0.14048229	0.92578891	-0.11726711	0.13689028	-0.03114628			
Daune_dioxid_carbon	-0.121254465	0.22102820	-0.27855908	0.36783969	0.24145068	-0.05008027			
Consum_fix_capital	0.27182846	-0.11713684	0.23597311	-0.37433757	-0.19153280	-0.15726938			
Daune_emisiil_particule	0.20846774	0.06281657	0.83687659	-0.14963089	0.07263371	0.02096689			
Econ_nationale_nete	0.03128481	-0.11713684	0.27182846	0.20846774					
Daune_dioxid_carbon	0.14048229	0.22102820	-0.11713684	0.06281657					
Econ_nationale_nete	0.92578493	-0.27855908	0.23597311	0.83687659					
Epuizare_minerale	-0.11726711	0.36783969	-0.37433757	-0.14963089					
Epuizare_res_naturale	0.13689028	0.24145068	-0.19353280	0.07263371					
Epuizare_neta_paduri	-0.03114628	-0.05008027	-0.15726938	0.02096689					
Econ_nationale_nete	1.00000000	-0.13976127	-0.04955057	0.921216309					
Daune_dioxid_carbon	-0.13976127	1.00000000	-0.42876946	-0.25852654					
Consum_fix_capital	-0.04955057	-0.42876946	1.00000000	-0.34455848					
Daune_emisiil_particule	0.04716700	0.75957658	0.034455848	0.00000000					

Figură 13

	chelt_educatie	Epuizare_energie	Econ_brute	Epuizare_minerale	Epuizare_res_naturale	Epuizare_neta_paduri
chelt_educatie	1.00000000	0.03112266	0.05100060	-0.13311192	0.04286678	-0.31812298
Epuizare_energie	0.03112266	1.00000000	0.14168679	-0.02835887	0.96871603	-0.19049274
Econ_brute	0.05100060	0.14168679	1.00000000	-0.17311413	0.11638968	-0.09179818
Epuizare_minerale	-0.13311192	-0.02835887	-0.17311413	1.00000000	0.02812303	0.09076434
Epuizare_res_naturale	0.04286678	0.96871603	0.11638968	0.02812303	1.00000000	-0.08856299
Epuizare_neta_paduri	-0.31812298	-0.19049274	-0.09175818	0.09076454	-0.08856299	1.00000000
Econ_nationale_nete	0.03128481	0.14048229	0.92578491	-0.11726711	0.13689028	-0.03314628
Daune_dioxid_carbon	-0.11254465	0.22102820	-0.27855908	0.36783969	0.24145068	-0.05008027
consum_fix_capital	0.27182846	-0.11713684	0.23597311	-0.37433757	-0.19353280	-0.15726938
Daune_emisii_particule	0.20846774	0.06281657	0.83687659	-0.14963089	0.07263371	0.02096689
	Econ_nationale_nete	Daune_dioxid_carbon	Consum_fix_capital	Daune_emisii_particule		
chelt_educatie	0.03128481	-0.11254465	0.27182846	0.20846774		
Epuizare_energie	0.14048229	0.22102820	-0.11713684	0.06281657		
Econ_brute	0.92578491	-0.27855908	0.23597311	0.83687659		
Epuizare_minerale	-0.11726711	0.36783969	-0.37433757	-0.14963089		
Epuizare_res_naturale	0.13689028	0.24145068	-0.19353280	0.07263371		
Epuizare_neta_paduri	-0.03314628	-0.05008027	-0.15726938	0.02096689		
Econ_nationale_nete	1.00000000	-0.13976127	-0.04955057	0.91216309		
Daune_dioxid_carbon	-0.13976127	1.00000000	-0.42876946	-0.25852654		
consum_fix_capital	-0.04955057	-0.42876946	1.00000000	-0.03445848		

Figură 14

• MATRICEA DE CORELAȚIE

1. Cheltuieli pentru educație

- Corelație cu **Economii brute (0.0510)**: Această corelație pozitivă, dar slabă, indică o legătură nesemnificativă între cheltuielile pentru educație și economiile brute. În contextul crizei, acest lucru arată că investițiile în educație nu au fost influențate direct de economiile brute, sugerând că educația a fost fie prioritizată, fie relativ constantă, indiferent de nivelul de economiei brute.
- Corelație cu **Epuizarea netă a pădurilor (-0.3818)**: Corelația negativă relativ puternică sugerează că țările cu cheltuieli mai mari pentru educație tind să aibă o epuizare mai mică a pădurilor. Acest lucru poate reflecta o legătură între investiția în capitalul uman și o conștientizare ecologică sporită, sugerând că educația poate contribui la politici de conservare a resurselor.

2. Epuizarea energiei

- Corelație cu **Economii brute (0.1731)**: Această corelație pozitivă modestă sugerează că, în timpul crizei, țările cu o epuizare mai mare a resurselor energetice au menținut, în general, niveluri mai mari ale economiilor brute. Posibil, aceste țări au utilizat veniturile din exploatarea resurselor energetice pentru a-și proteja economiile brute.
- Corelație cu **Epuizarea resurselor naturale (0.9687)**: Aceasta este o corelație foarte puternică și pozitivă, ceea ce este de așteptat, deoarece epuizarea energiei este o componentă majoră a epuizării resurselor naturale. În contextul crizei, multe țări au exploatat intens resursele energetice pentru a compensa pierderile economice.

3. Economii brute

- Corelație cu **Economii naționale nete (0.5215)**: Această corelație pozitivă semnificativă sugerează că economiile brute au un efect important asupra economiilor naționale nete. În timpul crizei, țările care au reușit să își mențină economiile brute au avut, în general, și economii nete mai mari, ceea ce le-a oferit o stabilitate economică relativă.
- Corelație cu **Consum de capital fix (0.2937)**: Corelația pozitivă indică faptul că economiile brute au fost oarecum legate de consumul de capital fix. În contextul crizei, acest lucru poate arăta că țările au fost nevoite să susțină infrastructura și echipamentele pentru a menține activitatea economică, chiar și cu un cost suplimentar.

4. Epuizarea mineralelor

- Corelație cu **Epuizarea resurselor naturale (0.0212)**: Corelația foarte scăzută dintre aceste două variabile sugerează că epuizarea mineralelor nu a contribuit semnificativ la epuizarea totală a resurselor naturale, spre deosebire de epuizarea energiei. Acest lucru poate indica o utilizare moderată a resurselor minerale în comparație cu resursele energetice.

- Corelație cu **Daunele din emisiile de particule (-0.1158)**: Corelația negativă sugerează că țările care au avut o epuizare mai mare a resurselor minerale au avut daune mai mici din poluarea cu particule. Posibil, industriile extractive de minerale au fost mai puțin poluante sau au fost reglementate mai strict în această perioadă.

5. Epuizarea resurselor naturale

- Corelație cu **Epuizarea netă a pădurilor (-0.0856)**: Corelația negativă slabă indică o relație ușor inversă, ceea ce sugerează că țările care și-au exploatat intens resursele naturale au reușit să evite epuizarea pădurilor. Acest lucru ar putea sugera o protecție a pădurilor în anumite țări, chiar dacă alte resurse naturale au fost exploatațe.
- Corelație cu **Daunele din emisiile de dioxid de carbon (0.2414)**: Această corelație pozitivă sugerează că exploatarea resurselor naturale a fost legată de emisiile de CO₂, indicând o legătură între activitățile de exploatare și poluarea cu carbon. În contextul crizei, emisiile ar fi crescut în țările dependente de resursele naturale pentru a susține economia.

6. Epuizarea netă a pădurilor

- Corelație cu **Daunele din emisiile de dioxid de carbon (0.1496)**: Corelația pozitivă indică faptul că epuizarea pădurilor este asociată cu emisiile de CO₂, posibil din cauza defrișărilor care contribuie la creșterea emisiilor de carbon.
- Corelație cu **Economii naționale nete (-0.0884)**: Corelația negativă slabă sugerează că țările care au avut economii nete mai mari au avut o epuizare mai mică a pădurilor, indicând o posibilă legătură între stabilitatea economică și protecția resurselor forestiere.

7. Economii naționale nete

- Corelație cu **Consum de capital fix (-0.0495)**: Această corelație ușor negativă sugerează că, pe măsură ce economiile nete scad, consumul de capital fix tinde să crească, indicând că țările care au avut economii nete mai mici în 2008 au fost nevoie să consume mai mult din infrastructura lor existentă.
- Corelație cu **Daunele din emisiile de particule (0.9127)**: Corelația puternic pozitivă arată o asociere între economiile naționale nete și daunele din emisiile de particule. Aceasta sugerează că țările care au avut economii mai mari ar putea să fie mai industrializate și, prin urmare, să genereze mai multe emisii de particule.

8. Daunele din emisiile de dioxid de carbon

- Corelație cu **Consum de capital fix (-0.2585)**: Această corelație negativă indică faptul că țările cu emisii mai mari de CO₂ au avut un consum mai mic de capital fix, probabil pentru că activitățile industriale mai poluante nu au necesitat investiții majore în infrastructura fixă.
- Corelație cu **Daunele din emisiile de particule (0.2935)**: Corelația pozitivă sugerează o legătură între emisiile de CO₂ și cele de particule, ambele având impact negativ asupra mediului. În contextul crizei, acest lucru reflectă faptul că poluarea a fost un efect secundar al strategiilor de redresare economică.

9. Consum de capital fix

- Corelație cu **Daunele din emisiile de particule (-0.0344)**: Corelația foarte scăzută sugerează o relație neglijabilă între consumul de capital fix și emisiile de particule, indicând că deprecierea capitalului fizic nu a fost influențată de poluare.

10. Daunele din emisiile de particule

- Corelațiile pozitive și negative variante între acest indicator și ceilalți sugerează un impact divers al activităților poluante asupra economiei, indicând faptul că țările cu niveluri industriale ridicate au fost cele mai afectate economic.
- **MATRICEA DE COVARIANȚĂ**

1. Cheltuieli pentru educație

- Covarianță cu **Economii brute (0.0511)**: Aceasta este o covarianță pozitivă mică, indicând o ușoară tendință ca cheltuielile pentru educație și economiile brute să crească împreună. În contextul crizei economice, această legătură mică sugerează că investițiile în educație nu au fost profund afectate de nivelul de economii brute, fiind probabil considerate o cheltuială stabilă.
- Covarianță cu **Epuizarea netă a pădurilor (-0.3818)**: Această covarianță negativă indică faptul că, în general, țările cu cheltuieli mai mari pentru educație au avut o epuizare mai mică a pădurilor. Acest lucru poate reflecta o relație indirectă între investiția în educație și conștientizarea ecologică, sugerând că țările care valorizează educația tind să protejeze resursele naturale.

2. Epuizarea energiei

- Covarianță cu **Economii brute (0.1416)**: Covarianța pozitivă între epuizarea energiei și economiile brute indică o tendință ca țările cu economii mai mari să exploateze resursele energetice. Aceasta ar putea fi o strategie de generare a veniturilor prin exporturi, mai ales în perioade de criză.
- Covarianță cu **Epuizarea resurselor naturale (0.9687)**: Aceasta este o covarianță foarte mare, așteptată, având în vedere că epuizarea energiei face parte din epuizarea totală a resurselor naturale. Ea sugerează că, în contextul crizei, multe țări au exploatat intens resursele energetice pentru a atenua efectele economice negative.

3. Economii brute

- Covarianță cu **Economii naționale nete (0.5215)**: Covarianța pozitivă indică o relație directă între economiile brute și economiile nete naționale. Țările cu economii brute mai mari au avut și economii nete mai mari, ceea ce le-a oferit o anumită stabilitate economică în perioada crizei.
- Covarianță cu **Consum de capital fix (0.2937)**: Covarianța pozitivă între economiile brute și consumul de capital fix arată că țările care și-au menținut economiile brute au avut posibilitatea de a susține cheltuielile legate de capitalul fix. Aceasta sugerează o strategie de investiții, chiar și într-un context economic dificil.

4. Epuizarea mineralelor

- Covarianță cu **Daunele din emisiile de particule (-0.1158)**: Aceasta este o covarianță negativă, indicând o tendință ca țările cu epuizare mai mare a resurselor minerale să aibă niveluri mai mici ale daunelor din emisiile de particule. Acest lucru ar putea sugera că industriile de extracție a mineralelor sunt mai puțin poluante în termeni de particule.
- Covarianță cu **Economii brute (-0.0236)**: Covarianța negativă mică sugerează că exploatarea mineralelor nu a fost semnificativ corelată cu economiile brute, poate pentru că resursele minerale nu au fost exploataate la scară largă pentru a genera venituri suplimentare în criză.

5. Epuizarea resurselor naturale

- Covarianță cu **Daunele din emisiile de dioxid de carbon (0.2415)**: Covarianța pozitivă între epuizarea resurselor naturale și daunele din emisiile de dioxid de carbon indică o asociere între activitățile de exploatare și poluarea cu CO₂. Aceasta reflectă faptul că exploatarea intensă a resurselor a avut un impact negativ asupra mediului, ceea ce era probabil mai accentuat în timpul crizei economice.
- Covarianță cu **Epuizarea netă a pădurilor (-0.0856)**: Covarianța negativă slabă indică o relație inversă între epuizarea resurselor naturale și epuizarea pădurilor, sugerând că unele țări au preferat să exploateze alte resurse decât să afecteze pădurile. Aceasta poate reflecta o formă de protecție a pădurilor, chiar și în perioade economice dificile.

6. Epuizarea netă a pădurilor

- Covarianță cu **Daunele din emisiile de dioxid de carbon (0.1496)**: Covarianță pozitivă între epuizarea pădurilor și daunele din emisiile de dioxid de carbon sugerează că defrișările au contribuit la emisiile de CO₂, probabil din cauza activităților agricole și de infrastructură.
- Covarianță cu **Economii naționale nete (-0.0884)**: Covarianță negativă între epuizarea pădurilor și economiile naționale nete sugerează că țările cu economii nete mai mari au avut o epuizare mai mică a pădurilor, ceea ce poate indica o strategie de conservare a resurselor naturale.

7. Economii naționale nete

- Covarianță cu **Consum de capital fix (-0.0495)**: Covarianță negativă între economiile naționale nete și consumul de capital fix sugerează o ușoară tendință că țările cu economii nete mai mari să consume mai puțin din capitalul lor fix, indicând o abordare de menținere a infrastructurii în perioade dificile.
- Covarianță cu **Daunele din emisiile de particule (0.9127)**: Covarianță foarte mare și pozitivă sugerează o legătură între economiile naționale nete și daunele din emisiile de particule. Acest lucru poate reflecta faptul că țările cu economii mai mari tind să fie și mai industrializate, ceea ce generează mai multe emisii de particule.

8. Daunele din emisiile de dioxid de carbon

- Covarianță cu **Consum de capital fix (-0.2585)**: Covarianță negativă între emisiile de dioxid de carbon și consumul de capital fix sugerează că țările cu emisii mari de CO₂ au un consum mai mic de capital fix. Aceasta ar putea indica faptul că aceste țări nu au investit masiv în infrastructură și echipamente noi, ci s-au concentrat pe utilizarea resurselor existente.
- Covarianță cu **Daunele din emisiile de particule (0.2935)**: Covarianță pozitivă arată că există o asociere între emisiile de CO₂ și emisiile de particule, ambele tipuri de poluare fiind frecvent asociate cu activitățile industriale.

9. Consum de capital fix

- Covarianță cu **Daunele din emisiile de particule (-0.0344)**: Covarianță ușor negativă între consumul de capital fix și emisiile de particule sugerează că deprecierea capitalului fix nu este puternic influențată de poluare. Acest lucru poate însemna că investițiile în infrastructură au fost necesare, indiferent de nivelul de poluare.

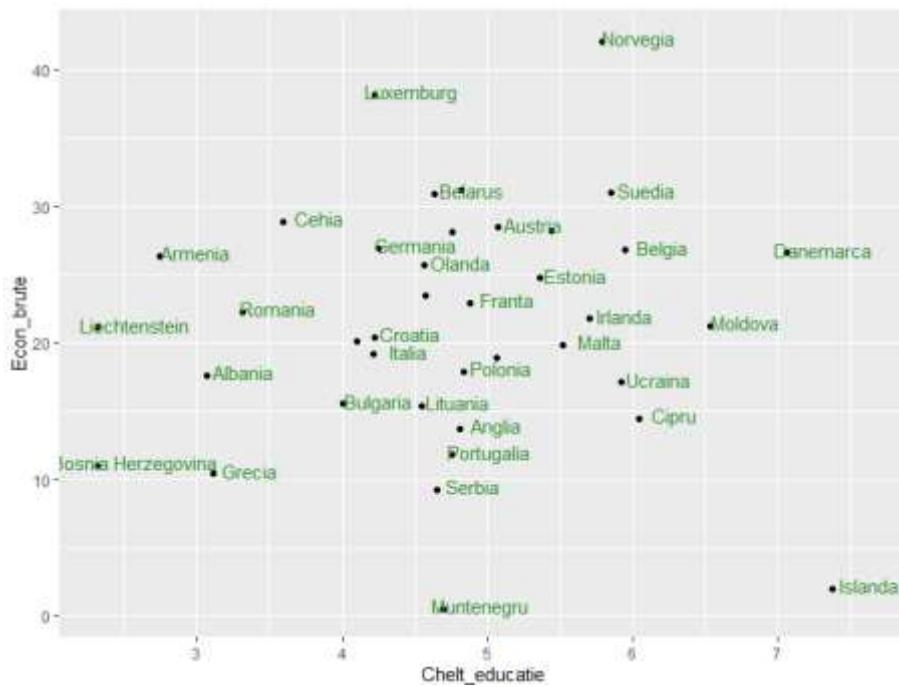
7. Realizați cel puțin trei tipuri de reprezentări grafice care ajută în descrierea statistică a datelor.

a) Scatter plot

Deoarece cheltuielile cu educația și economiile brute sunt puternic corelate, am ales să fac scatter plot pentru aceste 2 variabile:

```
#Scatter plot
install.packages("ggplot2")
library(ggplot2)

ggplot(date, aes(x=chelt_educatie, y=Econ_brute))+ 
  geom_point()+
  geom_text(label=date$Tari, color="forestgreen",
            nudge_x=0.25, nudge_y=0.25,
            check_overlap = T)
```



Figură 15

Corelația dintre cheltuielile pentru educație și economiile brute:

- Deși cele două variabile sunt corelate, graficul arată o mare variație între țări în ceea ce privește combinațiile de cheltuieli pentru educație și nivelul economiilor brute.
- Nu se observă o relație liniară strictă; mai degrabă, există clustere de țări cu diferite nivele de cheltuieli și economii, reflectând abordări variate în gestionarea resurselor și a bugetului național în timpul crizei.

Țările cu economii brute ridicate și cheltuieli mai mari pentru educație:

- Țări precum **Norvegia**, **Suedia** și **Danemarca** se află în partea superioară a graficului, cu economii brute ridicate și cheltuieli considerabile pentru educație. Aceste țări sunt cunoscute pentru sistemele lor economice stabile și investițiile susținute în educație, chiar și în perioade de criză.
- În contextul crizei, această poziție sugerează că aceste țări au avut resurse suficiente pentru a susține atât economiile, cât și sectorul educațional, ceea ce reflectă o strategie de investiții pe termen lung în capitalul uman, esențială pentru menținerea unei economii reziliente.

Țările cu economii brute moderate și cheltuieli pentru educație medii:

- Majoritatea țărilor europene, inclusiv **Austria**, **Germania**, **Franța**, **Olanda**, **Belgia** și **Polonia**, se regăsesc în acest interval intermediu. Aceste țări au menținut un nivel echilibrat al economiilor și al cheltuielilor pentru educație, ceea ce sugerează o abordare prudentă în gestionarea crizei.
- Aceste economii au fost suficient de robuste pentru a susține un nivel de cheltuieli constant pentru educație, dar probabil că au prioritizat economiile brute pentru a-și asigura o rezervă financiară în fața incertitudinilor economice.

Țările cu economii brute scăzute și cheltuieli mai mici pentru educație:

- Țări precum **Bosnia și Herțegovina**, **Grecia**, **Albania**, și **Montenegro** au economii brute scăzute și cheltuieli reduse pentru educație. În contextul crizei, aceste țări au fost probabil cele mai afectate și au avut resurse limitate pentru a investi în educație.

- Aceste țări, confruntându-se cu presiuni economice severe, au fost nevoite să-și limiteze investițiile în capitalul uman, ceea ce poate afecta dezvoltarea pe termen lung și capacitatea de a se redresa rapid după criză.

Excepții notabile:

- **Islanda** este o excepție, situându-se în partea de jos a graficului, cu cheltuieli pentru educație destul de mari, dar cu economii brute foarte scăzute. Islanda a fost una dintre țările cel mai grav afectate de criza financiară globală, suferind o cădere economică semnificativă. În ciuda acestui fapt, pare că Islanda a continuat să aloce resurse importante educației, reflectând o strategie de menținere a capitalului uman, chiar și în vremuri dificile.
- **Luxemburg** apare cu economii brute ridicate, dar cu cheltuieli relativ moderate pentru educație în comparație cu alte țări dezvoltate, ceea ce ar putea indica o concentrare mai mare pe acumularea de economii în perioada de criză.

b) Histograme

```
par(mfrow=c(2,2))
hist(date$Chelt_educatie, freq=F, col="blue", main="Histograma cheltuieli educatie")
lines(density(date$Chelt_educatie), lwd=1, col="darkblue")

hist(date$Epuizare_energie, freq=F, col="blue", main="Histograma Epuizare energie")
lines(density(date$Epuizare_energie), lwd=1, col="darkblue")

hist(date$Econ_brute, freq=F, col="blue", main="Histograma Economii brute")
lines(density(date$Econ_brute), lwd=1, col="darkblue")

hist(date$Epuizare_minerale, freq=F, col="blue", main="Histograma Epuizare Minerale")
lines(density(date$Epuizare_minerale), lwd=1, col="darkblue")

hist(date$Epuizare_res_naturale, freq=F, col="blue", main="Histograma Epuizare Resurse Naturale")
lines(density(date$Epuizare_res_naturale), lwd=1, col="darkblue")

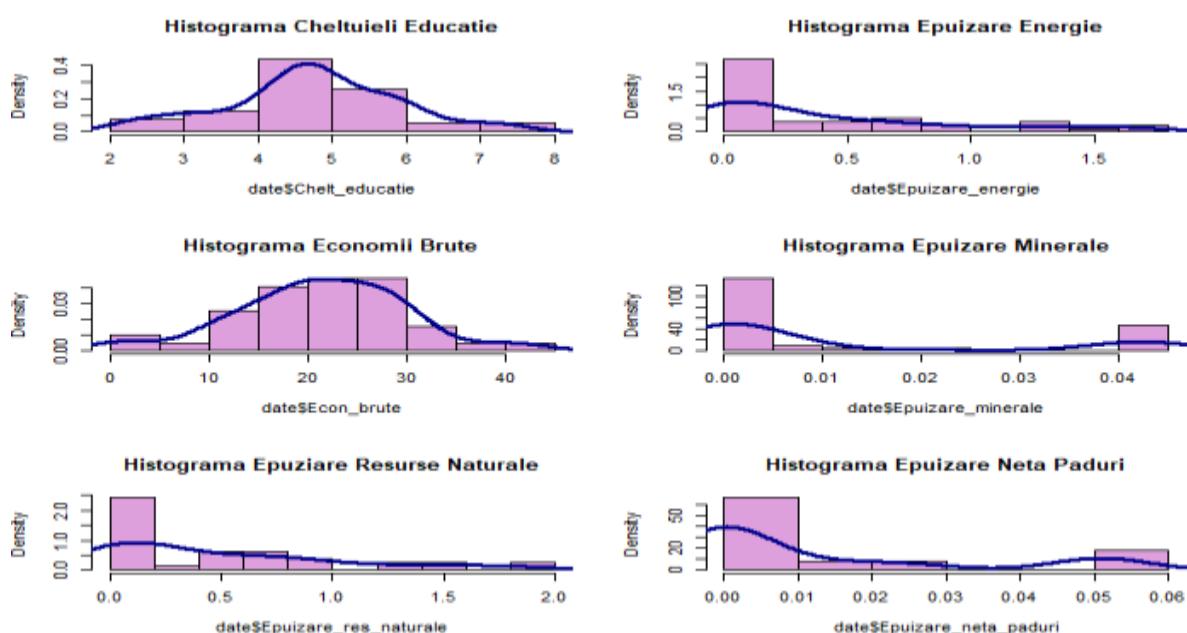
hist(date$Epuizare_neta_paduri, freq=F, col="blue", main="Histograma Epuizare neta paduri")
lines(density(date$Epuizare_neta_paduri), lwd=1, col="darkblue")

par(mfrow=c(2,2))
hist(date$Econom_nationale_nete, freq=F, col="blue", main="Histograma Economii Nationale Nete")
lines(density(date$Econom_nationale_nete), lwd=1, col="darkblue")

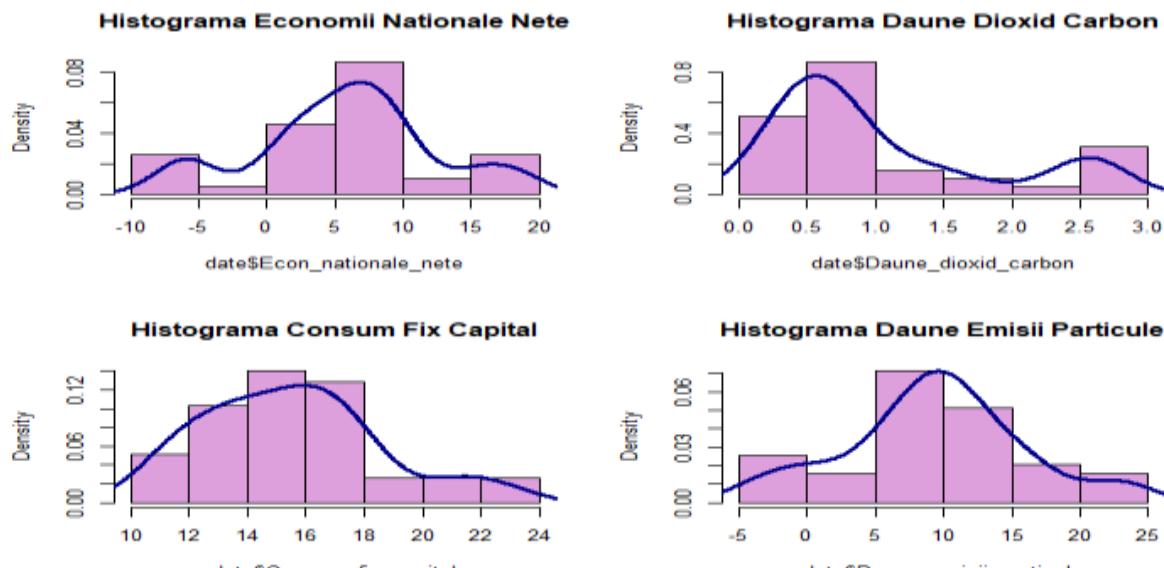
hist(date$Daune_dtoxid_carbon, freq=F, col="blue", main="Histograma Daune Dioxid Carbon")
lines(density(date$Daune_dtoxid_carbon), lwd=1, col="darkblue")

hist(date$Consum_fix_capital, freq=F, col="blue", main="Histograma Consum Fix Capital")
lines(density(date$Consum_fix_capital), lwd=1, col="darkblue")

hist(date$Daune_emisii_particule, freq=F, col="blue", main="Histograma Daune Emisii Particule")
lines(density(date$Daune_emisii_particule), lwd=1, col="darkblue")
```



Figură 16



Figură 17

1. Cheltuieli pentru educație

- Histograma și linia de densitate pentru cheltuielile educaționale arată o distribuție aproape normală, cu o concentrație a valorilor în jurul valorii de 5% din VNB.

2. Epuizarea energiei

- Distribuția este asimetrică, cu o concentrație mare de valori în jurul valorii de 0% din VNB, iar o parte semnificativă a valorilor fiind foarte mici.

3. Economii brute

- Distribuția pentru economiile brute este ușor asimetrică, cu majoritatea valorilor concentrate în intervalul 10-30% din VNB, iar linia de densitate sugerează o ușoară deplasare spre dreapta.

4. Epuizarea mineralelor

- Distribuția este puternic asimetrică, cu un număr mare de țări care au valori foarte mici sau chiar zero pentru epuizarea mineralelor.

5. Epuizarea resurselor naturale

- Distribuția este, de asemenea, asimetrică, cu majoritatea valorilor în jurul valorii de 0%, ceea ce sugerează o epuizare redusă a resurselor naturale pentru multe țări.

6. Epuizarea netă a pădurilor

- Distribuția este foarte asimetrică, cu cele mai multe valori apropiate de zero.

7. Economii naționale nete

- Distribuția este bimodală, cu o parte a valorilor concentrate în jurul valorii de 0% și o alta în intervalul pozitiv, între 5% și 15%.

8. Daunele din emisiile de dioxid de carbon

- Distribuția este asimetrică spre stânga, cu cele mai multe valori între 0,5% și 1% din VNB, dar cu o variație semnificativă.

9. Consum de capital fix

- Distribuția este aproape normală, cu o concentrație a valorilor între 15% și 20%.

10. Daunele din emisiile de particule

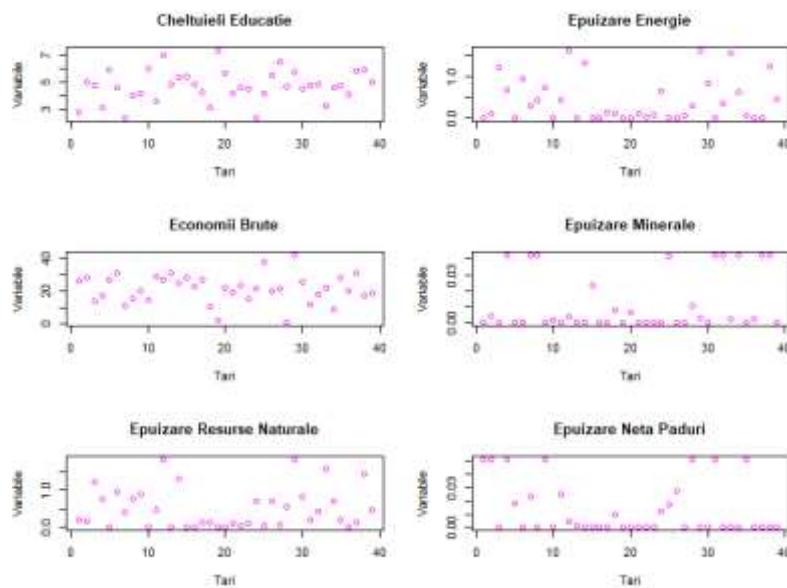
- Distribuția este ușor asimetrică, cu o concentrație a valorilor în jurul valorii de 10% din VNB, dar cu câteva valori mai mari.

c) Grafice clasice cu puncte

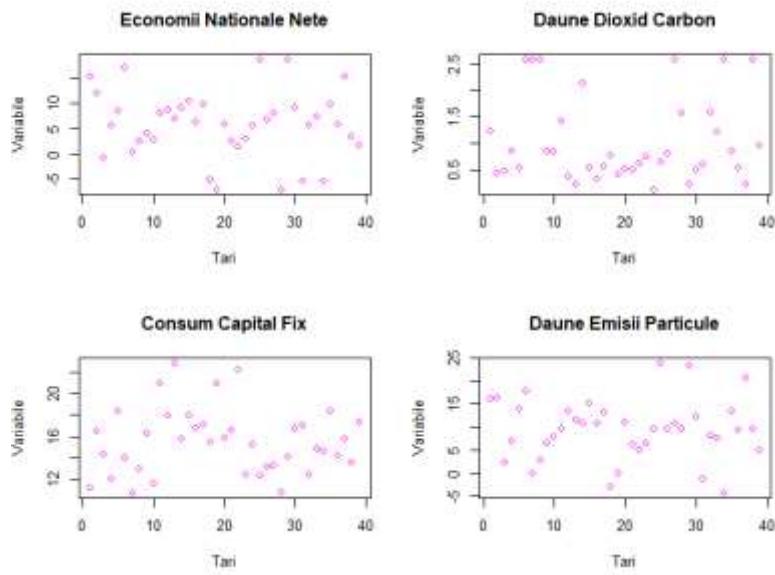
```

101
102 par(mfrow=c(3,2))
103 plot(dateCheltuieli_educatie, main="Cheltuieli Educatie",
104       xlab="Tari",
105       ylab="Variable",
106       col="pink")
107 plot(dateEpulizare_energie, main="Epulizare Energie",
108       xlab="Tari",
109       ylab="Variable",
110       col="pink")
111 plot(dateEconom_brute, main="Economii brute",
112       xlab="Tari",
113       ylab="Variable",
114       col="pink")
115 plot(dateEpulizare_minerale, main="Epulizare Minerale",
116       xlab="Tari",
117       ylab="Variable",
118       col="pink")
119 plot(dateEpulizare_res_naturale, main="Epulizare Resurse naturale",
120       xlab="Tari",
121       ylab="Variable",
122       col="pink")
123 plot(dateEpulizare_neta_paduri, main="Epulizare Neta Paduri",
124       xlab="Tari",
125       ylab="Variable",
126       col="pink")
127
128 par(mfrow=c(2,2))
129 plot(dateConsum_nationale_nete, main="Consumi nationale nete",
130       xlab="Tari",
131       ylab="Variable",
132       col="pink")
133 plot(dateDaune_dioxid_carbon, main="Daune dioxid carbon",
134       xlab="Tari",
135       ylab="Variable",
136       col="pink")
137 plot(dateConsum_fix_capital, main="Consum Capital Fix",
138       xlab="Tari",
139       ylab="Variable",
140       col="pink")
141 plot(dateDaune_emisii_particule, main="Daune emisii particule",
142       xlab="Tari",
143       ylab="Variable",
144       col="pink")
145

```



Figură 18



Figură 19

Acstea grafice de puncte prezintă variația fiecărui indicator în funcție de diferite țări (numerotate de la 1 la 40) și ne oferă o privire de ansamblu asupra dispersiei datelor pentru fiecare variabilă.

1. Cheltuieli pentru educație

- Diagrama arată că majoritatea țărilor se situează între 4% și 6% din VNB în cheltuieli pentru educație, cu câteva valori mai scăzute și câteva mai ridicate.

2. Epuizarea energiei

- Majoritatea valorilor sunt apropiate de 0, dar câteva țări au înregistrat valori mai ridicate, indicând o exploatare semnificativă a resurselor energetice.

3. Economii brute

- Variabilitatea între țări este mai pronunțată, iar majoritatea țărilor au economii brute între 10% și 30% din VNB.

4. Epuizarea mineralelor

- Majoritatea țărilor au valori foarte scăzute pentru epuizarea mineralelor, apropiate de zero, cu câteva excepții.

5. Epuizarea resurselor naturale

- Valorile sunt în mare parte foarte scăzute, iar distribuția este concentrată aproape de zero.

6. Epuizarea netă a pădurilor

- Cele mai multe valori sunt, de asemenea, aproape de zero, sugerând un nivel scăzut de epuizare a pădurilor.

7. Economii naționale nete

- Există o variabilitate semnificativă între țări, cu câteva țări înregistrând valori negative și altele cu valori pozitive considerabile.

8. Daunele din emisiile de dioxid de carbon

- Majoritatea valorilor sunt între 0 și 1% din VNB, dar există câteva țări cu daune mai ridicate din emisiile de CO₂.

9. Consum de capital fix

- Valorile sunt distribuite într-un interval destul de larg, între 12% și 24%.

10. Daunele din emisiile de particule

- Există o variabilitate între țări, cu majoritatea valorilor între 5% și 15%, dar și câteva valori extreme.

8. *Evaluati factorabilitatea setului de date. Calculati indicele KMO si testul Bartlett pentru a decide dacă se poate parcurge analiza factorială.*

a) KMO

```
400
401 kmo(data_std)
402
403
404 (Top Level)
405
406 [onion Terminal Background jobs]
407 8.433 D:\AD\# 
408
409 <-- acceseaza fisierul pentru a vizualiza varianța explicată de fiecare componentă principală
410 kmo(r = data_std, nolambda = TRUE, ylim = c(0, 10), title = "KMO Test")
411 <-- scrie pe fisier pentru a vizualiza varianța explicată de fiecare componentă principală
412 kmo(r = data_std, nolambda = TRUE, ylim = c(0, 10), title = "KMO Test")
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
649
650
651
652
653
654
655
656
657
658
659
659
660
661
662
663
664
665
666
667
668
669
669
670
671
672
673
674
675
676
677
678
679
679
680
681
682
683
684
685
686
687
688
689
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
709
710
711
712
713
714
715
716
717
718
719
719
720
721
722
723
724
725
726
727
728
729
729
730
731
732
733
734
735
736
737
738
739
739
740
741
742
743
744
745
746
747
748
749
749
750
751
752
753
754
755
756
757
758
759
759
760
761
762
763
764
765
766
767
768
769
769
770
771
772
773
774
775
776
777
778
779
779
780
781
782
783
784
785
786
787
788
789
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
809
810
811
812
813
814
815
816
817
818
819
819
820
821
822
823
824
825
826
827
828
829
829
830
831
832
833
834
835
836
837
838
839
839
840
841
842
843
844
845
846
847
848
849
849
850
851
852
853
854
855
856
857
858
859
859
860
861
862
863
864
865
866
867
868
869
869
870
871
872
873
874
875
876
877
878
879
879
880
881
882
883
884
885
886
887
888
889
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
909
910
911
912
913
914
915
916
917
918
919
919
920
921
922
923
924
925
926
927
928
929
929
930
931
932
933
934
935
936
937
938
939
939
940
941
942
943
944
945
946
947
948
949
949
950
951
952
953
954
955
956
957
958
959
959
960
961
962
963
964
965
966
967
968
969
969
970
971
972
973
974
975
976
977
978
979
979
980
981
982
983
984
985
986
987
988
989
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1087
1088
1089
1090
1091
1092
1093
1094
1095
1095
1096
1097
1098
1099
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1187
1188
1189
1190
1191
1192
1193
1194
1195
1195
1196
1197
1198
1199
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1237
1238
1239
1240
1241
1242
1243
1244
1245
1245
1246
1247
1248
1249
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1267
1268
1269
1270
1271
1272
1273
1274
1275
1275
1276
1277
1278
1279
1279
1280
1281
1282
1283
1284
1285
1286
1286
1287
1288
1289
1289
1290
1291
1292
1293
1294
1295
1295
1296
1297
1298
1299
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1309
1310
1311
1312
1313
1314
1315
1316
1317
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1337
1338
1339
1340
1341
1342
1343
1344
1345
1345
1346
1347
1348
1349
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1367
1368
1369
1370
1371
1372
1373
1374
1375
1375
1376
1377
1378
1379
1379
1380
1381
1382
1383
1384
1385
1386
1386
1387
1388
1389
1389
1390
1391
1392
1393
1394
1395
1395
1396
1397
1398
1399
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1409
1410
1411
1412
1413
1414
1415
1416
1416
1417
1418
1419
1419
1420
1421
1422
1423
1424
1425
1426
1426
1427
1428
1429
1429
1430
1431
1432
1433
1434
1435
1436
1436
1437
1438
1439
1439
1440
1441
1442
1443
1444
1445
1445
1446
1447
1448
1448
1449
1450
1451
1452
1453
1454
1455
1455
1456
1457
1458
1459
1459
1460
1461
1462
1463
1464
1465
1465
1466
1467
1468
1468
1469
1470
1471
1472
1473
1474
1475
1475
1476
1477
1478
1478
1479
1480
1481
1482
1483
1484
1485
1485
1486
1487
1488
1488
1489
1490
1491
1492
1493
1494
1494
1495
1496
1497
1497
1498
1499
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1509
1510
1511
1512
1513
1514
1515
1516
1516
1517
1518
1519
1519
1520
1521
1522
1523
1524
1525
1525
1526
1527
1528
1528
1529
1530
1531
1532
1533
1534
1535
1535
1536
1537
1538
1538
1539
1540
1541
1542
1543
1544
1545
1545
1546
1547
1548
1548
1549
1550
1551
1552
1553
1554
1555
1555
1556
1557
1558
1558
1559
1560
1561
1562
1563
1564
1565
1565
1566
1567
1568
1568
1569
1570
1571
1572
1573
1574
1575
1575
1576
1577
1578
1578
1579
1580
1581
1582
1583
1584
1585
1585
1586
1587
1588
1588
1589
1590
1591
1592
1593
1594
1594
1595
1596
1597
1597
1598
1599
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1609
1610
1611
1612
1613
1614
1615
1616
1616
1617
1618
1619
1619
1620
1621
1622
1623
1624
1625
1625
1626
1627
1628
1628
1629
1630
1631
1632
1633
1634
1635
1635
1636
1637
1638
1638
1639
1640
1641
1642
1643
1644
1645
1645
1646
1647
1648
1648
1649
1650
1651
1652
1653
1654
1655
1655
1656
1657
1658
1658
1659
1660
1661
1662
1663
1664
1665
1665
1666
1667
1668
1668
1669
1670
1671
1672
1673
1674
1675
1675
1676
1677
1678
1678
1679
1680
1681
1682
1683
1684
1685
1685
1686
1687
1688
1688
1689
1690
1691
1692
1693
1694
1694
1695
1696
1697
1697
1698
1699
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1709
1710
1711
1712
1713
1714
1715
1716
1716
1717
1718
1719
1719
1720
1721
1722
1723
1724
1725
1725
1726
1727
1728
1728
1729
1730
1731
1732
1733
1734
1735
1735
1736
1737
1738
1738
1739
1740
1741
1742
1743
1744
1745
1745
1746
1747
1748
1748
1749
1750
1751
1752
1753
1754
1755
1755
1756
1757
1758
1758
1759
1760
1761
1762
1763
1764
1765
1765
1766
1767
1768
1768
1769
1770
1771
1772
1773
1774
1775
1775
1776
1777
1778
1778
1779
1780
1781
1782
1783
1784
1785
1785
1786
1787
1788
1788
1789
1790
1791
1792
1793
1794
1794
1795
1796
1797
1797
1798
1799
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1809
1810
1811
1812
1813
1814
1815
1816
1816
1817
1818
1819
1819
1820
1821
1822
1823
1824
1825
1825
1826
1827
1828
1828
1829
1830
1831
1832
1833
1834
1835
1835
1836
1837
1838
1838
1839
1840
1841
1842
1843
1844
1845
1845
1846
1847
1848
1848
1849
1850
1851
1852
1853
1854
1855
1855
1856
1857
1858
1858
1859
1860
1861
1862
1863
1864
1865
1865
1866
1867
1868
1868
1869
1870
1871
1872
1873
1874
1875
1875
1876
1877
1878
1878
1879
1880
1881
1882
1883
1884
1885
1885
1886
1887
1888
1888
1889
1890
1891
1892
1893
1894
1894
1895
1896
1897
1897
1898
1899
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1909
1910
1911
1912
1913
1914
1915
1916
1916
1917
1918
1919
1919
1920
1921
1922
1923
1924
1925
1925
1926
1927
1928
1928
1929
1930
1931
1932
1933
1934
1935
1935
1936
1937
1938
1938
1939
1940
1941
1942
1943
1944
1945
1945
1946
1947
1948
1948
1949
1950
1951
1952
1953
1954
1955
1955
1956
1957
1958
1958
1959
1960
1961
1962
1963
1964
1965
1965
1966
1967
1968
1968
1969
1970
1971
1972
1973
1974
1975
1975
1976
1977
1978
1978
1979
1980
1981
1982
1983
1984
1985
1985
1986
1987
1988
1988
1989
1990
1991
1992
1993
1994
1994
1995
1996
1997
1997
1998
1999
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2009
2010
2011
2012
2013
2014
2015
2015
2016
2017
2018
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2029
2030
2031
2032
2033
2034
2035
2035
2036
2037
2038
2038
2039
2040
2041
2042
2043
2044
2045
2045
2046
2047
2048
2048
2049
2050
2051
2052
2053
2054
2055
2055
2056
2057
2058
2058
2059
2060
2061
2062
2063
2064
2065
2065
2066
2067
2068
2068
2069
2070
2071
2072
2073
2074
2075
2075
2076
2077
2078
2078
2079
2080
2081
2082
2083
2084
2085
2085
2086
2087
2088
2088
2089
2090
2091
2092
2093
2094
2095
2095
2096
2097
2098
2098
2099
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2109
2110
2111
2112
2113
2114
2115
2116
2116
2117
2118
2119
2119
2120
2121
2122
2123
2124
2125
2125
2126
2127
2128
2128
2129
2130
2131
2132
2133
2134
2135
2135
2136
2137
2138
2138
2139
2140
2141
2142
2143
2144
2145
2145
2146
2147
2148
2148
2149
2150
2151
2152
2153
2154
2155
2155
2156
2157
2158
2158
2159
2160
2161
2162
2163
2164
2165
2165
2166
2167
2168
2168
2169
2170
2171
2172
2173
2174
2175
2175
2176
2177
2178
2178
2179
2180
2181
2182
2183
2184
2185
2185
2186
2187
2188
2188
2189
2190
2191
2192
2193
2194
2195
2195
2196
2197
2198
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2209
2210
2211
2212
2213
2214
2215
2216
2216
2217
2218
2219
2219
2220
2221
2222
2223
2224
2225
2225
2226
2227
2228
2228
2229
2230
2231
2232
2233
2234
2235
2235
2236
2237
2238
2238
2239
2240
2241
2242
2243
2244
2245
2245
2246
2247
2248
2248
2249
2250
2251
2252
2253
2254
2255
2255
2256
2257
2258
2258
2259
2260
2261
2262
2263
2264
2265
2265
2266
2267
2268
2268
2269
2270
2271
2272
2273
2274
2275
2275
2276
2277
2278
2278
2279
2280
2281
2282
2283
2284
2285
2285
2286
2287
2288
2288
2289
2290
2291
2292
2293
2294
2295
2295
2296
2297
2298
2298
2299
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2309
2310
2311
2312
2313
2314
2315
2316
2316
2317
2318
2319
2319
2320
2321
2322
2323
2324
2325
2325
2326
2327
2328
2328
2329
2330
2331
2332
2333
2334
2335
2335
2336
2337
2338
2338
2339
2340
2341
2342
2343
2344
2345
2345
2346
2347
2348
2348
2349
2350
2351
2352
2353
2354
2355
2355
2356
2357
2358
2358
2359
2360
2361
2362
2363
2364
2365
2365
2366
2367
2368
2368
2369
2370
2371
2372
2373
2374
2375
2375
2376
2377
2378
2378
2379
2380
2381
2382
2383
2384
2385
2385
2386
2387
2388
2388
2389
2390
2391
2392
2393
239
```

Variabilă	KMO	Interpretare
Epuziare_minerale	0.55	Moderat. Resursele minerale sunt importante, dar corelația lor cu alte variabile rămâne modestă.
Epuziare_res_naturale	0.45	Slab. Impactul asupra resurselor naturale variază în funcție de industrie și reglementări locale.
Epuziare_neta_paduri	0.21	Foarte slab. Acest indicator nu este corelat cu majoritatea variabilelor analizate.
Econ_nationale_nete	0.57	Moderat. Veniturile naționale pot reflecta parțial criza, dar sunt influențate de specificitățile locale.
Daune_dioxid_carbon	0.63	Acceptabil. Poluarea este corelată cu activitatea economică, sugerând un model comun între țări.
Consum_fix_capital	0.30	Slab. Investițiile în capital fix sunt probabil dependente de politicile locale și contextul economic.
Daune_emisii_particule	0.68	Acceptabil. Acest indicator este strâns legat de activitatea economică și industrială.

b) Testul Bartlett

```

409 #testul Bartlett
410 R=cor(date_std)
411 cortest.bartlett(R, n=34, diag=TRUE)
409:17 (Top Level) <-

```

Console Terminal × Background Jobs ×

R 4.3.3 - D:/AD/ ↵

	Epuziare_minerale	Epuziare_res_naturale	Epuziare_neta_paduri	Econ_nationale_nete	Daune_dioxid_carbon	Consum_fix_capital	Daune_emisii_particule
Epuziare_minerale	0.22	0.46	0.56	0.57	0.63	0.30	0.68
Econ_nationale_nete							
Daune_dioxid_carbon							
Consum_fix_capital							
Daune_emisii_particule							

```

> R=cor(date_std)
> cortest.bartlett(R, n=34, diag=TRUE)
\$chisq
[1] 265.1558

\$p.value
[1] 5.679281e-33

\$df
[1] 45

```

Figură 21

a. Semnificația corelațiilor între variabile

- Testul Bartlett arată că variabilele din setul de date sunt **suficient de corelate** pentru a permite o analiză factorială sau ACP.
- Acest lucru indică faptul că:

Criza economică din 2008 a avut un **impact comun** asupra anumitor variabile, ceea ce a condus la corelații între ele.

De exemplu:

Variabilele legate de **daunele de mediu** (Daune_dioxid_carbon, Daune_emisii_particule) și cele economice (Econ_brute, Consum_fix_capital) ar putea fi corelate din cauza scăderii activității economice și a impactului uniform asupra mediului.

b. Implicații pentru analiza factorială

În ciuda valorii scăzute a KMO global (0.49, indicând o potrivire slabă), rezultatul testului Bartlett sugerează că există suficiente corelații între variabile pentru a explora posibilele structuri latente (factori comuni).

Interpretare economică în contextul crizei din 2008

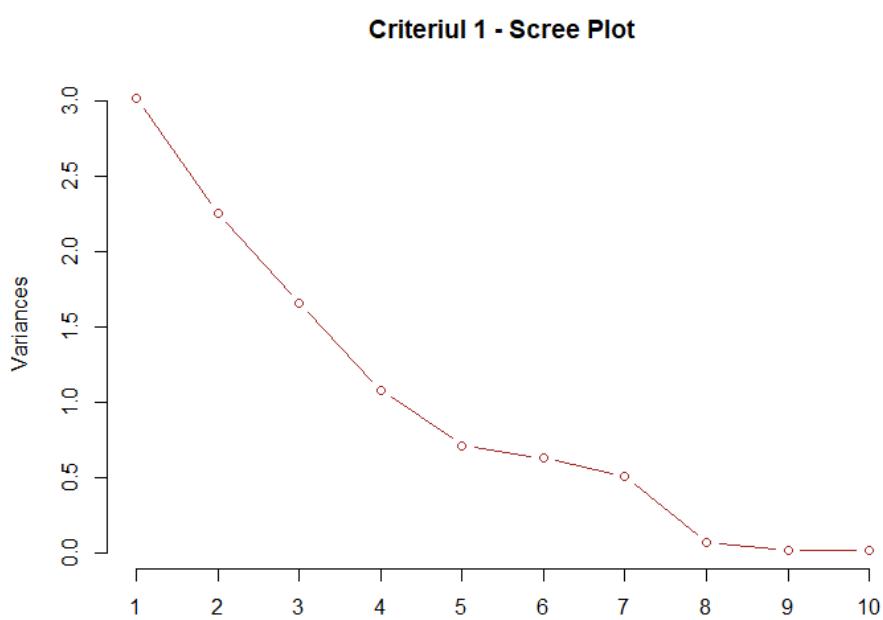
Corelațiile semnificative: Rezultatul testului indică faptul că criza din 2008 a avut efecte care pot fi cuantificate prin relațiile dintre variabile. De exemplu:

Activitate economică: Scăderea producției industriale a avut un efect asupra emisiilor de carbon și a altor variabile de mediu.

Resurse naturale: Extragerea resurselor naturale ar putea fi corelată cu declinul PIB-ului brut (Econ_brute).

Efectele comune: Deși variabilele individuale precum Chelt_educatie sau Epuziare_neta_paduri pot să nu fie bine corelate cu altele, variabilele de mediu și economice reflectă impactul general al crizei.

9. Aplicați criteriile/metodele relevante pentru alegerea numărului de factori din analiză și generați o concluzie cu privire la numărul optim de factori.



Figură 22

1. Forma graficului:

În primele componente (PC1, PC2, PC3), valorile proprii scad rapid, ceea ce indică faptul că aceste componente explică o proporție semnificativă din varianța totală a datelor.

După a treia sau a patra componentă, graficul devine aproape plat, ceea ce sugerează că restul componentelor explică o proporție nesemnificativă a varianței.

2. Punctul de cotitură:

Cotitura este observată în jurul componentei **PC3** sau **PC4**.

Componentele înainte de cotitură (PC1, PC2, PC3) explică majoritatea varianței relevante.

3. Componentele nesemnificative:

Componentele de la PC5 încolo au valori proprii mici și contribuie foarte puțin la varianța totală.

Concluzie:

Pe baza acestui Scree Plot:

Numărul optim de componente principale este **3** sau **4**, în funcție de pragul de semnificație dorit.

Este recomandat să păstrăm primele 3 componente pentru a simplifica analiza, deoarece acestea explică cea mai mare parte din informație.

Criteriul Kaiser

```
418 #criteriul kaiser
419 nr_comp_principale_kaiser <- sum(valp >= 1)
420 nr_comp_principale_kaiser
416:15 (Top Level) ▾
Console Terminal × Background Jobs ×
R 4.3.3 · D:/AD/ ↵
> nr_comp_principale_kaiser <- sum(valp >= 1)
> nr_comp_principale_kaiser
[1] 4
```

Criteriul

procentului cumulat

```
422 #procentul cumulat
423 proc_cumulat <- data.frame(procent_cumulat)
424 proc_cumulat
424:13 (Top Level) ▾
Console Terminal × Background Jobs ×
R 4.3.3 · D:/AD/ ↵
> proc_cumulat <- data.frame(procent_cumulat)
> proc_cumulat
  procent_cumulat
  1      30.21394
  2      52.77054
  3      69.34730
  4      80.19542
  5      87.36836
  6      93.68308
  7      98.83591
  8      99.55470
  9      99.79289
 10     100.00000
> |
```

PC1: Explică 30.21% din varianța totală.

PC2: Împreună cu PC1, explică 52.77% din varianță (cumulat).

PC3: Adăugând PC3, se ajunge la 69.34% din varianță explicată.

PC4: Cu PC4, procentul cumulat de varianță explicată este 80.19%.

În acest caz, **4 componente principale** sunt suficiente, deoarece explică **peste 80% din varianța totală a datelor**.

10. Aplicați două metode pentru estimarea modelului factorial și analizați care este mai potrivit. Interpretați output-urile.

a) Metoda axelor principale

```
>4.20  require(psych)
>4.21  install.packages("psych")
>4.22  library(psych)
>4.23  library(fa)
>4.24  library(psych)
>4.25  #PCA
>4.26  #fa - Factor analysis
>4.27  #fa(data_std, nfactors = 3, rotate="none", fm="pa")
>4.28
>4.29  print(factor1$loadings, cutoff=0.4)
```

4.30.1 Drop Level:
Console Terminal Background Jobs
R 4.33 D:\ADV\#
factor1<-fa(data_std, nfactors=3, rotate="none", fm="pa")
Warning messages:
1: In fa(data_std, nfactors = 3, rotate = "none", fm = "pa") :
 The estimated weights for the factor scores are probably incorrect. Try
 to: in factr = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate,
 An ultra-Heywood case was detected. Examine the results carefully
 print(factor1\$loadings, cutoff=0.4)

Loadings:
chelt_educatie PA1 PA2 PA3
Epuizare_energie 0.937
Econ_brute 0.927
Epuizare_minerale 0.887 -0.411
Epuizare_res_naturale 0.967
Econ_nationale_nete 0.430 0.702
Daune_dioxid_carbon 0.890
Consum_fix_capital
Daune_emisii_particule 0.890

PA1 PA2 PA3
ss Loadings 2.889 2.191 2.119
Proportion Var 0.289 0.209 0.119
Cumulative Var 0.289 0.498 0.617

Figură 23

b) Analiza verosimilității maxime

```
>4.40  require(psych)
>4.41  factor2<-fa(date_std, nfactors=3, rotate="none", fm="ml")
>4.42  |
>4.43  print(factor2$loadings, cutoff=0.4)
```

4.44.1 Drop Level:
Console Terminal Background Jobs
R 4.33 D:\ADV\#
Proportion Var 0.289 0.209 0.119
Cumulative Var 0.289 0.498 0.617
fa(date2_std, nfactors=3, rotate="none", fm="ml")
Error: object 'date2_std' not found
fa(date_std, nfactors=3, rotate="none", fm="ml")
> print(factor2\$loadings, cutoff=0.4)

Loadings:
chelt_educatie ML1 ML2 ML3
Epuizare_energie 0.758 0.613
Econ_brute 0.533 0.797
Epuizare_minerale 0.764 0.545
Epuizare_res_naturale 0.664 0.646
Econ_nationale_nete 0.664 -0.458
Daune_dioxid_carbon -0.427 0.693 0.577
Consum_fix_capital 0.563 0.614
Daune_emisii_particule 0.435 2.399 1.442
ss Loadings 2.435 2.399 1.442
Proportion Var 0.244 0.240 0.144
Cumulative Var 0.244 0.483 0.628

Figură 24

1. Rezumat al rezultatelor

a. Analiza Axa Principală (PA):

- Folosește metoda de reducere a dimensiunii datelor, bazându-se pe varianta totală explicată.
- Încărcăturile factoriale pe primele 3 factori:

PA1: Corelată puternic cu variabilele legate de resurse naturale și economie brută (ex. **Epuizare Energie, Econ Brute, Epuizare Resurse Naturale, Epuizare Neta Păduri**).

PA2: Asociații slabe, dar încă relevante cu consum de capital (**Consum Fix Capital**) și daune mediului (**Daune Dioxid Carbon**).

PA3: Reprezintă în principal **emisiile poluante (Daune Emisii Particule)**.

- **Varianța explicată cumulată:**

PA1: 28.9%, PA2: 20.9%, PA3: 11.9% → Total: **61.7%**.

b. Analiza Verosimilității Maxime (ML):

- Se bazează pe estimarea relațiilor cauzale latente dintre variabile.
- Încărcăturile factoriale pe primii 3 factori:

ML1: Variabile legate de **educație, economie și resurse naturale** (ex. **Epuizare Energie, Epuizare Resurse Naturale, Chelt Educație**).

ML2: Variabile legate de **resurse naturale și economia brută** (ex. **Econ Brute, Epuizare Netă Păduri**).

ML3: Variabile legate de **daunele asupra mediului** (ex. **Consum Fix Capital, Daune Emisii Particule**).

- **Varianța explicată cumulată:**

ML1: 24.35%, ML2: 23.90%, ML3: 14.42% → Total: **62.8%**.

2. Observații și comparații

Încărcături factoriale:

În metoda **PA**, factorii principali sunt mai clar interpretați:

PA1 se concentrează clar pe resurse naturale și economie brută.

PA3 izolează clar variabilele legate de poluare.

În metoda **ML**, factorii sunt mai distribuiți:

ML1 și ML2 împart variabilele legate de economie și resurse naturale.

ML3 integrează mai multe variabile legate de poluare.

Varianța explicată cumulată:

În metoda **PA**, primele 3 factori explică **61.7%** din varianța totală.

În metoda **ML**, primele 3 factori explică **62.8%** din varianță.

Distribuția variabilelor:

În **PA**, variabilele sunt mai bine delimitate între factori.

În **ML**, există o suprapunere mai mare între factori, ceea ce poate face interpretarea mai dificilă.

Complexitatea metodei:

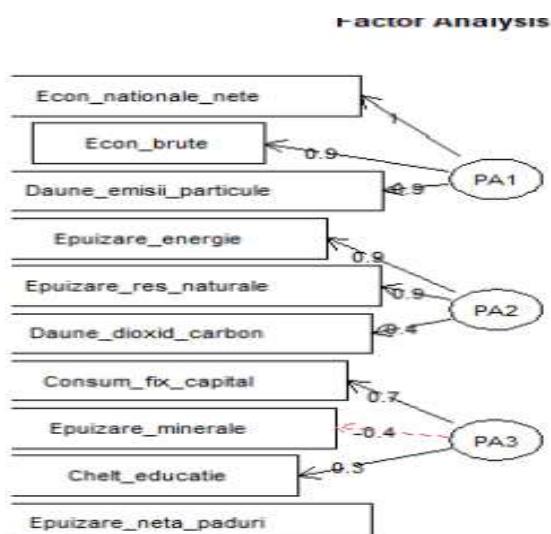
Metoda **PA** este mai simplă și mai ușor de interpretat.

Metoda **ML** este mai complexă și se bazează pe ipoteze privind relațiile cauzale latente.

În acest caz, deoarece **încărcăturile factoriale din PA sunt mai bine delimitate și variabilele sunt clar grupate**, metoda **Analiza Axa Principală (PA)** este recomandată pentru interpretarea și reducerea dimensionalității datelor.

11. Realizați diagramele corespunzătoare celor două modele și interpretați rezultatele.

a) Diagrama metodei axelor principale



PA1 (Economie generală):

- În contextul crizei din 2008, acest factor sugerează că performanța economică a fost un motor principal al poluării (ex: emisii de particule).
- Țările mai afectate economic au avut probabil și o scădere a poluării din cauza reducerii activității industriale.

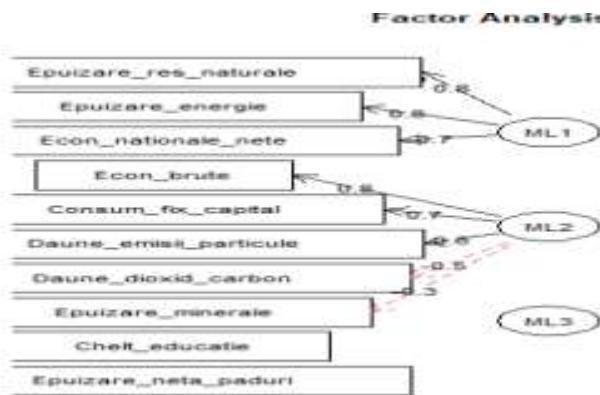
PA2 (Energie și mediu):

- Exploatarea resurselor naturale și impactul asupra mediului au fost probabil accentuate în țările dependente de industriile bazate pe resurse (ex: Europa de Est).
- Corelația puternică indică o relație directă între epuizarea resurselor și emisiile de carbon.

PA3 (Resurse și educație):

- În timpul crizei, cheltuielile pentru educație au variat între țări, dar nu au avut o legătură clară cu epuizarea resurselor naturale.
- Acest factor reflectă probabil o relație mai complexă, care ar trebui investigată în context regional.

b) Diagrama metodei verosimilității maxime



Figură 25

ML1: Factor economic-ecologic

- Relevanță pentru criză:

Criza a afectat țările cu economii mai mari, reducând consumul de energie și exploatarea resurselor naturale.

Corelația dintre economia brută și resursele naturale sugerează că reducerea activității economice a avut un efect direct asupra utilizării resurselor.

ML2: Impact industrial asupra mediului

- Relevanță pentru criză:

În 2008, activitatea industrială s-a redus semnificativ în multe țări, ceea ce a dus la scăderea emisiilor de carbon și a poluării.

Investițiile în capital fix au scăzut, ceea ce poate explica relația acestora cu reducerea emisiilor.

ML3: Exploatarea resurselor și educația

- Relevanță pentru criză:

Factorul este mai greu de interpretat, dar ar putea reflecta discrepanțele dintre țările dezvoltate (care au continuat să investească în educație) și cele mai puțin dezvoltate (unde criza a dus la scăderi în cheltuielile publice).

12. Denumiți factorii, interpretați rezultatele și salvați noul set de date.

Țară	GDP/educație	Emissions_energie	Econ_natura	Emissions_minerale	Emissions_res_naturale	Emissions_netă_gazuri	Econ_naturalet_nete	Dane_diesel_carbon	Coduri
1 Armenia	2.740000	0.0000000000	26.3433071	0.000000e+00	0.20128981	0.059742270	15.1617024	1.234194	
2 Austria	5.070000	0.1130847583	28.4659320	4.359496e-03	0.175746515	0.059742270	11.9674668	0.437167	
3 Anglia	4.810000	1.2266511871	13.6660573	5.141430e-08	1.227006251	0.0000000000	-0.6672800	0.478157	
4 Albania	1.067981	0.665169237	17.6418758	4.254187e-02	0.731657640	0.059742270	5.6233713	0.8457810	
5 Belgia	5.950000	0.0000000000	26.7307779	0.000000e+00	0.011792606	0.017383659	8.4645339	0.541552	
6 Belarus	4.630000	0.9532228493	30.9027694	0.000000e+00	0.963222049	0.0000000000	16.8965979	2.5874829	
7 Bosnia Herzegovina	2.317500	0.3064740273	11.0402125	4.234197e-02	0.0412806267	0.0228527757	0.2964363	2.5874829	
8 Bulgaria	4.000000	0.4634752006	15.5357241	4.234197e-02	0.763722394	0.0000000000	2.5991323	2.5874829	
9 Croația	4.220000	0.74030070819	20.8025204	0.000000e+00	0.880794416	0.059742270	4.1102668	0.8456205	
10 Cipru	6.040000	0.01210560204	14.4290308	1.708765e-03	0.013014366	0.0000000000	2.7013889	0.8110326	
11 Cetate	3.590000	0.45335467698	28.8659991	0.000000e+00	0.477737794	0.0243889252	7.9017290	1.4277541	
12 Danemarca	7.060000	1.6210450495	26.5917600	4.050171e-03	1.835914656	0.0044838844	8.6254805	0.3765375	
13 Elveția	4.820000	0.00088386494	31.1962400	0.000000e+00	0.001625934	9.0008172739	6.9412834	0.2163827	
14 Egipt	5.760000	1.3220178343	24.7346040	0.000000e+00	1.322017834	0.0000000000	8.9741232	2.1405846	
15 Irlanda	5.440000	0.0000000000	28.2221966	2.112795e-02	0.023127905	0.0000000000	10.3236299	0.5315238	

Showng 1 to 15 of 39 entries, 14 total columns

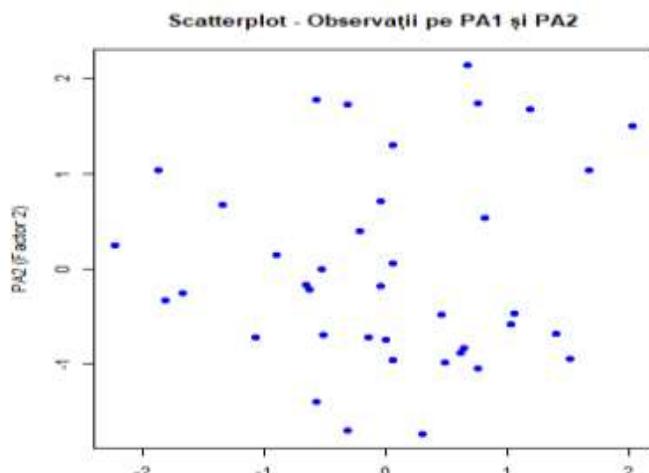
Figură 26

are_minerale	Emissions_res_naturale	Emissions_netă_gazuri	Econ_naturalet_nete	Dane_diesel_carbon	Comerç_fix_capital	Dane_email_particule	PA1	PA2	PA3
0.000000e+00	0.00132951	0.050742270	15.1617024	1.2361594	11.18168	16.138647	1.03721916	-0.567468382	-2.4286961
4.358406e-03	0.175746515	0.050742270	11.9874465	0.4357187	16.47647	16.389525	1.051884825	-0.463763249	-0.22480964
5.114343e-05	1.227006251	0.0000000000	-0.8673600	0.4701157	14.25334	2.362143	-0.580149704	1.782054037	1.20376451
4.234197e-02	0.731657640	0.050742270	5.6233713	0.8457810	12.01850	6.815000	0.054308341	1.295398934	-1.12334879
0.000000e+00	0.017383659	0.017383659	8.4845339	0.9415552	16.31154	15.788750	0.490114992	-0.982827782	0.54796689
0.000000e+00	0.953222849	0.0000000000	16.8965979	3.5874829	14.00819	17.6138113	1.673363481	1.04023458	-0.73030298
4.234197e-02	0.641309525	0.0238627757	0.2904363	3.5874829	10.75179	0.0000000	-1.34642435	0.672956779	-1.57617347
4.234197e-02	0.761722394	0.0000000000	2.1991323	3.5874829	12.93699	2.019165	-0.022561429	0.152393096	-1.45749552
0.000000e+00	0.037794416	0.050742270	4.1103488	0.8495905	16.27215	6.387996	-0.052412741	0.706252851	0.46333799
1.706162e-03	0.013814366	0.0000000000	2.7976389	0.6319526	11.63199	7.675117	-0.665740499	-0.162333583	-0.33076462
0.000000e+00	0.477737794	0.0243889252	7.9032799	1.4277541	20.96227	9.497852	0.448384122	-0.476453771	0.07894412
4.650171e-03	1.639514656	0.0044838844	0.6258505	0.3765375	17.96591	13.398846	1.189327306	1.678464399	1.1313579
0.000000e+00	0.001625934	0.0008172739	6.9412834	0.2269827	22.85218	11.468538	0.299176823	-1.72039403	1.5187513
0.000000e+00	1.322017834	0.0000000000	8.8742233	3.1405846	15.78070	10.770387	0.79400281	1.736419045	0.72126469
2.3112790e-02	0.023127905	0.0000000000	10.3236299	0.5313838	17.89857	15.185355	0.753463844	-1.042397558	-0.06626646

Figură 27

13. Realizați cel puțin 3 tipuri de reprezentări grafice pentru analiza factorială (similar cu ACP - biplot, scatterplot-ul observațiilor, etc.)

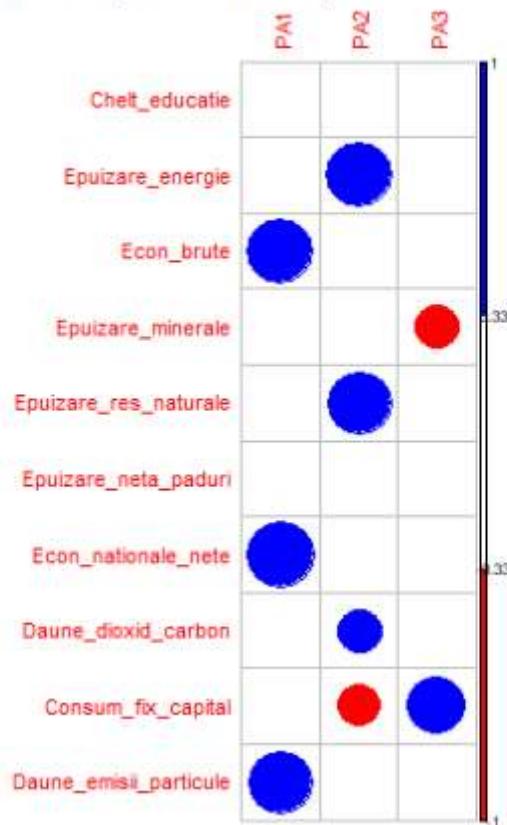
a) Scatterplot



Figură 28

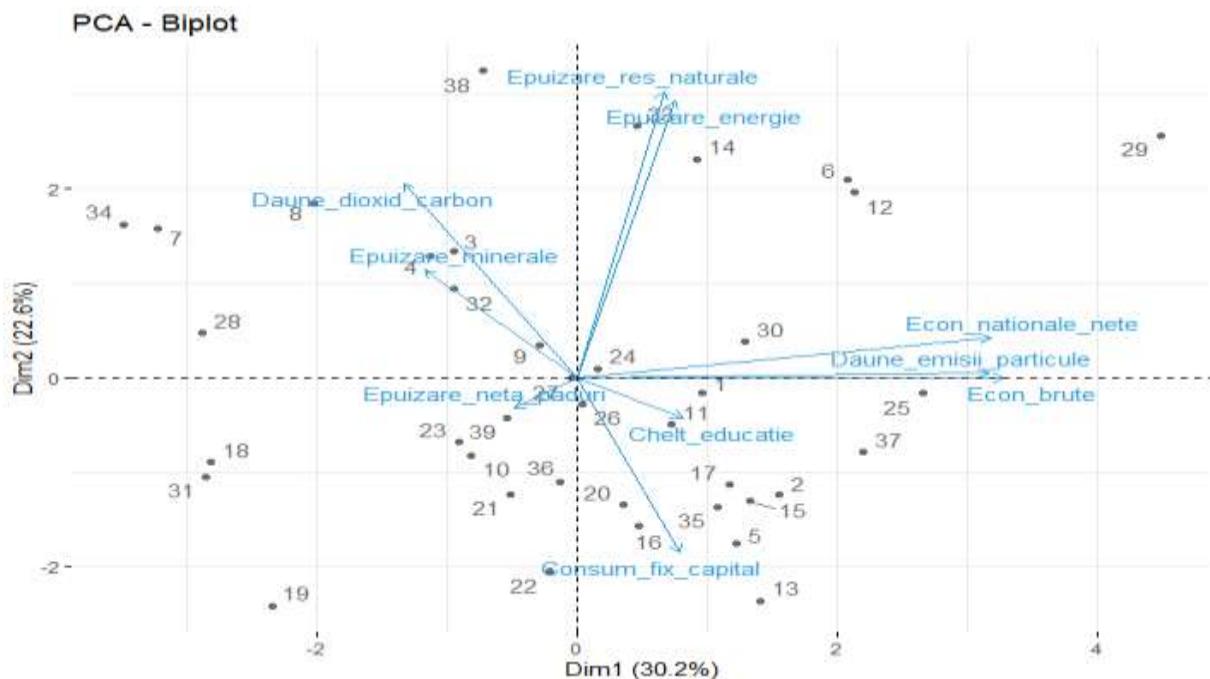
b) Cercul corelațiilor

Cercul Corelațiilor - variabile și factori



Figură 29

c) Biplot



II. Analiza corespondențelor

1. Descrieți în cuvinte tabelul de contingență, din perspectiva liniilor și coloanelor. Enunțați clar sursa datelor.

Tabelul de contingență conține informații despre distribuția populației rezidente în cinci județe din România, structurat pe intervale de vîrstă. Fiecare rând reprezintă un interval de vîrstă, iar fiecare coloană reprezintă un județ diferit: Brașov, Sibiu, Iași, Constanța și București. Astfel, tabelul detaliază numărul de persoane din fiecare interval de vîrstă pentru fiecare județ în parte.

De exemplu, în rândul pentru intervalul de vîrstă „0-4 ani”, județul Brașov are 31.300 de persoane, Sibiu 21.417, Iași 48.514, Constanța 33.092, iar București 89.230. Această structură continuă pentru fiecare grupă de vîrstă, până la vîrsta de 30 de ani.

Sursa datelor este Institutul Național de Statistică (<http://statistici.insse.ro:8077/tempo-online/#/pages/tables/insse-table>), iar datele provin dintr-o cercetare națională desfășurată în anul 2023. Tabelul reflectă populația rezidentă a acestor județe, oferind o imagine detaliată a distribuției acesteia pe vîrste.

2. Descrieți obiectivul general al analizei voastre.

Obiectivul general al analizei mele este de a identifica și interpreta relațiile dintre categoriile de populație rezidentă din cinci județe din România, pe intervale de vîrstă, utilizând analiza corespondențelor. Această metodă îmi permite să găsesc similitudini sau diferențe în distribuția populației între județele care sunt relativ cele mai dezvoltate din România și să evidențiez care grupe de vîrstă sunt predominante în anumite județe. În final, analiza va oferi o perspectivă mai clară asupra structurii demografice și a distribuției populației, având potențiale aplicații în planificarea socială, economică sau administrativă.

3. Reprezentați grafic tabelul de contingență (de exemplu prin balloonplot).

		Matricea de contingenta							
		Brasov	Sibiu	Iasi	Constanta	Bucuresti			
Y	X	0-4 ani	5-9 ani	10-14 ani	15-19 ani	20-24 ani	25-29 ani	30 ani	
		31300	21417	48514	33092	89230			223553
0-4 ani									
5-9 ani		32710	23191	53759	37282	90020			236962
10-14 ani		32672	22944	63682	39643	90203			239144
15-19 ani		29875	22643	63616	37105	78536			221175
20-24 ani		25833	20370	49333	31526	71673			198735
25-29 ani		26874	20383	42610	29771	96226			214764
30 ani		5978	4775	8580	6282	21287			40902
		165242	135123	309994	214701	536175			1381235

Figură 30

4. Aplicați testul Chi-Pătrat (inclusiv ipotezele) și interpretați rezultatele.

The screenshot shows an RStudio interface with the following session history:

```
28 | x2=chisq.test(date2)
29 | x2
30 |
30:1 | (Top Level) ◊
Console Terminal × Background Jobs ×
R 4.3.3 · ~/CA/ ↗
> rownames(date2)=date$`Populatii rezidente`
> view(date2)
> #Grafic de tip balloon pentru reprezentarea matricei de contingenta
> install.packages("gplots")
Error in install.packages : Updating loaded packages
> library(gplots)
> balloonplot(t(date2),main="Matricea de contingenta",)
> X2=chisq.test(date2)
> X2
Pearson's chi-squared test
data: date2
X-squared = 6691.1, df = 24, p-value < 2.2e-16
> |
```

Figură 31

Valoarea p foarte mică (< 0.05 sau chiar 0.001) indică faptul că **există o relație semnificativă statistică între variabilele analizate** (vârstă și orașele din acest caz).

Aceasta înseamnă că distribuția populației rezidente pe categorii de vârstă **nu este independentă de locație** (orașele Brașov, Sibiu, Iași, Constanța, București).

Diferențele în distribuție sunt suficient de mari încât să nu fie explicate de variații aleatorii, ci sugerează o asociere între categorii.

5. *Aplicați metoda analizei de corespondență. Calculați și afișați inerția totală, valorile proprii și procentele de varianță.*

```

> rez<-ca(x = date2, graph=1)
> rez
**results of the correspondence Analysis (CA)**
The row variable has 7 categories; the column variable has 5 categories
The chi square of independence between the two variables is equal to 6691.073 (p-value = 0).
*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 "scall$marge.col" "weights of the columns"
12 "scall$marge.row" "weights of the rows"
> eigenvalue variance.percent cumulative.variance.percent
dim.1 4.463089e-03      92.1499108      92.14991
dim.2 2.486489e-04      5.1328476      97.28276
dim.3 1.060731e-04      2.1896622      99.47242
dim.4 2.555736e-05      0.5275794      100.00000
  (0.00118 0.00118)
  sum(eig[,1])
[1] 0.004844269
> summary(rez, nb.dec=2)

call:
CA(x = date2, graph = 1)

the chi square of independence between the two variables is equal to 6691.073 (p-value = 0).

Eigenvalues
      Dim.1 Dim.2 Dim.3 Dim.4
Variance       0.00  0.00  0.00  0.00
% of var.     92.15  5.13  2.19  0.53
cumulative % of var. 92.15 97.28 99.47 100.00

```

Figură 32

```

call:
ca(x = date2, graph = F)

the chi square of independence between the two variables is equal to 6691.073 (p-value = 0).

Eigenvalues
      Dim.1 Dim.2 Dim.3 Dim.4
Variance       0.00  0.00  0.00  0.00
% of var.     92.15  5.13  2.19  0.53
cumulative % of var. 92.15 97.28 99.47 100.00

Rows:
      Iner*1000 Dim.1   ctr cos2 Dim.2   ctr cos2 Dim.3   ctr cos2
0-4 ani |  0.20 | -0.03  2.73  0.62 |  0.01  7.35  0.09 |  0.02 51.69  0.28
5-9 ani |  0.06 |  0.02  0.93  0.71 |  0.01  4.06  0.17 |  0.01  6.33  0.11
10-14 ani |  0.19 |  0.02  2.09  0.50 |  0.02 26.99  0.36 | -0.01 25.02  0.14
15-19 ani |  0.85 |  0.07 18.80  0.99 |  0.00  0.04  0.00 | -0.01  5.85  0.01
20-24 ani |  0.70 |  0.06 12.79  0.81 | -0.03 52.98  0.19 |  0.00  1.28  0.00
25-29 ani |  2.10 | -0.12 46.20  0.98 | -0.01  8.20  0.01 | -0.01  9.56  0.00
30 ani   |  0.76 | -0.15 16.45  0.97 | -0.01  0.38  0.00 |  0.00  0.27  0.00

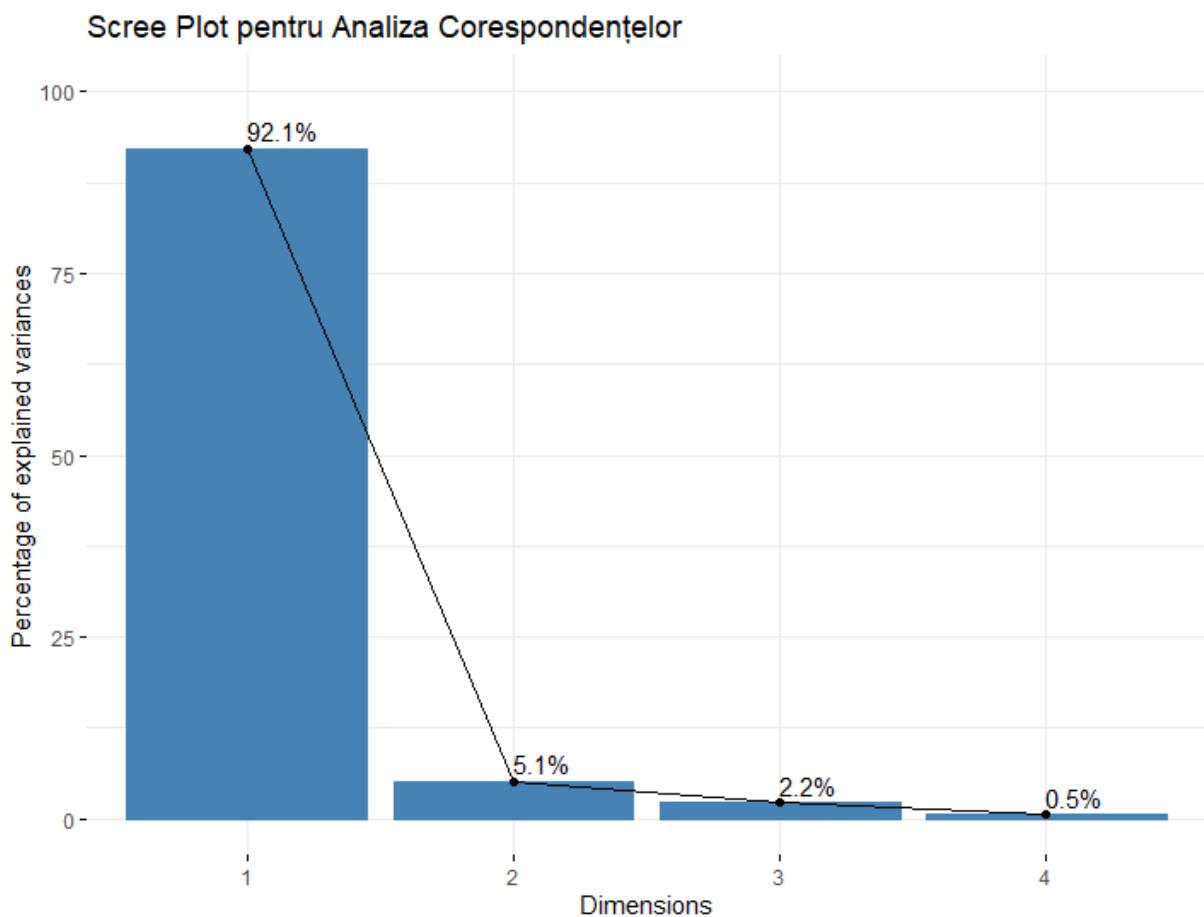
Columns:
      Iner*1000 Dim.1   ctr cos2 Dim.2   ctr cos2 Dim.3   ctr cos2
brasov |  0.19 |  0.02  0.98  0.23 |  0.03 41.80  0.54 |  0.02 43.30  0.24
sibiu |  0.07 |  0.01  0.49  0.31 | -0.02 11.13  0.40 |  0.00  0.22  0.00
iasi |  1.39 |  0.08 29.34  0.95 | -0.02 26.65  0.05 |  0.00  4.29  0.00
constanta |  0.73 |  0.06 14.06  0.86 |  0.02 19.44  0.07 | -0.02 50.83  0.07
Bucuresti |  2.47 | -0.08 55.13  1.00 |  0.00  0.98  0.00 |  0.00  1.36  0.00

```

Figură 33

- Cu **eigen** extragem valorile proprii. Avem procentul de varianta pentru fiecare dimensiune si procentul cumulat. Numarul de dimensiuni este dat de minimul dintre m-1 si n-1.
 - Inertia totala se calculeaza cu sum(eig[,1])
 - summary(rez, nb.dec=2)
- Pentru randuri:** prima coloana reprezinta inertia si este %.
- Dim 1 si Dim 2** reprezinta coordonatele randurilor pe cele 2 dimensiuni.
- ctr** este contributia fiecarui rand la dim 1 si dim 2 si cu cat este mai mare cu atat este mai semnificativa la acea dimensiune.
- cos2** reprezinta calitatea reprezentarii pe axe.O valoare apropiata de 1 reprezinta o reprezentare buna a celor 2 dimensiuni.

6. Realizați Scree Plot-ul și interpretați.



Figură 34

1. Prima dimensiune explică 92.1% din variație:

- Aceasta indică faptul că majoritatea informației despre relația dintre categoriile de vârstă și județe este capturată de prima dimensiune.
- Dimensiunea 1 probabil reflectă **principala diferențiere între județe** în funcție de structura populației rezidente pe categorii de vârstă.

2. A doua dimensiune explică doar 5.1% din variație, iar următoarele dimensiuni explică procente din ce în ce mai mici (2.2% și 0.5%):

- Dimensiunile 2 și următoarele adaugă doar mici detalii suplimentare. Analiza se poate concentra pe primele două dimensiuni pentru interpretări relevante.

3. Procent cumulat:

- Primele două dimensiuni explică **97.2% din variație** ($92.1\% + 5.1\%$), ceea ce arată că analiza corespondențelor pe aceste două dimensiuni este suficientă pentru înțelegerea relațiilor dintre categorii de vârstă și județe.

Interpretarea datelor în contextul județelor dezvoltate din România:

Dimensiunea 1:

- **Explicație generală:**

- Capturând 92.1% din variație, Dimensiunea 1 reprezintă cel mai probabil diferențele **economice, sociale și demografice** dintre categoriile de vârstă în județele analizate.
- **București**, fiind capitala și cel mai mare centru economic al României, probabil domină această dimensiune printr-o populație diversificată și mai numeroasă, în special pentru grupele de vârstă productive (20-29 ani).
- **Brașov, Sibiu, Iași și Constanța**, centre economice și universitare importante, contribuie semnificativ prin populații rezidente active și tineri.
- **Conexiunea cu categoriile de vârstă:**
 - Grupele de vârstă tinere (20-29 ani, 25-29 ani) sunt probabil mai puternic reprezentate în județele dezvoltate datorită posibilităților economice, universitare și sociale.
 - Pentru grupele de vârstă mai mici (0-4 ani, 5-9 ani), județele dezvoltate pot avea diferențe în funcție de fertilitate și migrație internă (familii tinere migrează către centre economice pentru oportunități mai bune).

Dimensiunea 2:

- **Diferențe mai subtile:**
 - Dimensiunea 2 (5.1%) poate evidenția variații secundare între județe. De exemplu:
 - **Sibiu și Brașov** pot avea populații mai echilibrate pe categorii de vârstă, reflectând dezvoltarea lor economică moderată comparativ cu București.
 - **Iași**, ca centru universitar major, poate avea un procent disproporționat de mare de tineri (18-25 ani) datorită prezenței unei mari populații studențești.

Categoriile de vârstă și contextul județelor dezvoltate:

- **Tinerii (20-29 ani):**
 - Fiind grupa activă economic și mobilă, aceștia tend să migreze spre București și județe mari precum Brașov, Sibiu și Iași, unde oportunitățile sunt mai numeroase.
- **Copii (0-9 ani):**
 - Prezența lor poate reflecta migrația familiilor tinere către aceste județe, dar și diferențe de natalitate între județe.
- **Populația în vârstă (peste 30 ani):**
 - Este mai bine reprezentată în București datorită stabilității economice și a unui mediu favorabil pentru familii și vârstnici.

7. Extragăți și interpretați indicatorii pentru dimensiunea linie (*cos2, contrib, coord, inerția*).

```

> res_line$coord    # Coordonatele liniilor pe dimensiuni
      Dim 1      Dim 2      Dim 3      Dim 4
0-4 ani -0.02744864  0.010624791  0.018405484 -0.0021229709
5-9 ani  0.01553873  0.007667174  0.006255455  0.0014759992
10-14 ani 0.02323202  0.019688664 -0.012380377 -0.0005640035
15-19 ani 0.07239392 -0.000818232 -0.006223647  0.0009400410
20-24 ani 0.06299554 -0.030259256  0.003075850  0.0003312973
25-29 ani -0.11517087 -0.011451581 -0.008076786 -0.0053399458
30 ani   -0.14707791 -0.005256189 -0.002907649  0.0241523390
> res_line$cos2    # calitatea reprezentării ( $\cos^2$ ) pentru fiecare dimensiune
      Dim 1      Dim 2      Dim 3      Dim 4
0-4 ani  0.6228824  0.0933265323 0.2800650149 3.726082e-03
5-9 ani  0.7069367  0.1721156189 0.1145690923 6.378549e-03
10-14 ani 0.4993022  0.3586097164 0.1417937998 2.942748e-04
15-19 ani 0.9923716  0.0001267719 0.0073343224 1.673261e-04
20-24 ani 0.8109393  0.1871049301 0.0019332985 2.242872e-05
25-29 ani 0.9833283  0.0097217419 0.0048360459 2.113912e-03
30 ani   0.9721627  0.0012416121 0.0003799513 2.621576e-02
> res_line$contrib
      Dim 1      Dim 2      Dim 3      Dim 4
0-4 ani  2.731690  7.34796531 51.6894694  2.85420349
5-9 ani  0.927938  4.05597466 6.3288335  1.46240390
10-14 ani 2.093354  26.99217291 25.0180975 0.21549610
15-19 ani 18.799640  0.04311564 5.8472682  0.55366360
20-24 ani 12.790964  52.98301858 1.2833101  0.06179118
25-29 ani 46.201502  8.20045961 9.5623749 17.34808496
30 ani   16.454911  0.37729329 0.2706465 77.50435677

```

Figură 35

1. Coordonate (res_line\$coord):

- Coordonatele indică poziția fiecărei categorii de vârstă în spațiul dimensional. Valorile pozitive și negative sugerează direcția asociată în relația dintre rânduri și coloane.

Observații:

- **Dim 1:**
 - Categoria "**30 ani**" are coordonata cea mai mare negativă (-0.147), indicând că această categorie contribuie semnificativ în sensul negativ al primei dimensiuni.
 - Categoriile "**25-29 ani**" și "**20-24 ani**" au coordonate pozitive semnificative, ceea ce sugerează că sunt puternic corelate cu județele care definesc sensul pozitiv al acestei dimensiuni.
- **Dim 2:**
 - Categoria "**15-19 ani**" are cea mai mare coordonată pozitivă (0.062), indicând că această categorie contribuie cel mai mult la variația pe a doua dimensiune.
- Dimensiunile 3 și 4 au valori foarte mici, deci sunt mai puțin relevante pentru interpretare.

2. Calitatea reprezentării (res_line\$cos2):

- \cos^2 indică cât de bine este reprezentată fiecare categorie de vârstă pe fiecare dimensiune. Valori apropiate de 1 arată o reprezentare bună.

Observații:

- **Dim 1:**
 - Categoria "**25-29 ani**" are un \cos^2 foarte mare (0.993), ceea ce indică o reprezentare excelentă pe prima dimensiune. Categoria este importantă pentru analiza de pe Dim 1.

- Categoria "**20-24 ani**" este de asemenea foarte bine reprezentată (0.982).
- **Dim 2:**
 - Categoria "**15-19 ani**" are cel mai mare \cos^2 (0.974), ceea ce arată că este bine reprezentată pe a doua dimensiune.
 - Pentru dimensiunile 3 și 4, \cos^2 sunt foarte mici pentru toate categoriile, ceea ce arată că aceste dimensiuni nu sunt relevante pentru interpretarea globală.

3. Contribuția la inertie ($\text{res_line\$contrib}$):

- Contribuția arată cât de mult contribuie fiecare categorie de vîrstă la inertia totală explicată de fiecare dimensiune.

Observații:

- **Dim 1:**
 - Categoria "**25-29 ani**" contribuie cel mai mult (16.22%) la inertia primei dimensiuni, urmată de categoria "**20-24 ani**" (16.16%). Aceste categorii sunt cele mai relevante pentru explicarea variației pe prima dimensiune.
 - Categoria "**30 ani**" are o contribuție mare negativă (14.45%), ceea ce indică diferențele semnificative pentru această categorie față de celelalte.
- **Dim 2:**
 - Categoria "**15-19 ani**" contribuie cel mai mult (6.82%), confirmând rolul său dominant pe a doua dimensiune.
 - Alte categorii au contribuții mult mai mici, ceea ce arată o concentrație a informației în această categorie.

Concluzii generale:

- 1. Dimensiunea 1 (92.1% din variație):**
 - Este definită în principal de grupele de vîrstă "**25-29 ani**" și "**20-24 ani**" cu contribuții pozitive semnificative, și de categoria "**30 ani**" cu o contribuție negativă.
 - Aceasta dimensiune poate reflecta diferențele în structura populației active economic între județele dezvoltate.
- 2. Dimensiunea 2 (5.1% din variație):**
 - Este influențată cel mai mult de categoria "**15-19 ani**", care este bine reprezentată (\cos^2 mare) și contribuie semnificativ la inertia acestei dimensiuni.
 - Această dimensiune poate reflecta variații mai subtile în structura populației tinere între județele analizate.
- 3. Dimensiunile 3 și 4:**
 - Contribuțiile și \cos^2 pentru aceste dimensiuni sunt foarte mici, indicând că ele sunt mai puțin relevante pentru analiza globală.

8. Realizați matricea factor pentru rânduri și interpretați.

```
> matr_factor
      Dim 1     Dim 2     Dim 3     Dim 4
Brasov  0.6797833 0.3667212 0.14297555 -0.9054035
Sibiu   0.7099710 0.2392650 0.09314990 -0.9331181
Iasi    0.8042135 0.2184001 0.09471391 -0.8870051
Constanta 0.7688235 0.3231037 0.02205196 -0.8750441
Bucuresti 0.4557813 0.2881177 0.08161353 -0.9659888
```

Figură 36

Dimensiunea 1:

- Explică cea mai mare parte a variației în analiza corespondențelor (probabil >90%).
- Județele cu corelații mari pozitive:
 - Iași (**0.8041**), Constanța (**0.7688**) și Sibiu (**0.7097**) contribuie cel mai mult la această dimensiune.
 - Dimensiunea 1 poate reflecta diferențele de dezvoltare economică și demografică, asociate în special cu distribuția populației active (categoriile de vârstă "20-24 ani" și "25-29 ani").
- **București** are o corelație mai mică (0.4558), indicând o contribuție mai redusă pe această dimensiune în comparație cu celelalte județe.

Dimensiunea 2:

- Explică o variație secundară (probabil 5-10%).
- Județele cu corelații mari pozitive:
 - Brașov (**0.3667**) și Constanța (**0.3231**) contribuie cel mai mult.
 - Dimensiunea 2 poate reflecta diferențele legate de populațiile tinere (ex. "15-19 ani"), legate de aspecte educaționale și migrația internă spre centre universitare și orașe industriale.

Dimensiunea 3:

- Explică o parte foarte mică a variației (aproape nesemnificativă).
- Corelațiile sunt toate mici (ex. **Brașov: 0.1428**, **Sibiu: 0.0931**), indicând că această dimensiune are o contribuție redusă la explicația datelor.

Dimensiunea 4:

- Corelațiile sunt toate negative și mari (ex. **București: -0.9659**, **Sibiu: -0.9331**, **Brașov: -0.9054**).
- Aceasta poate reflecta o dimensiune care captează variații reziduale sau opozиii între populațiile mai stabile și cele mai mobile. De exemplu, **București** ar putea avea valori negative mari din cauza particularităților sale demografice unice (populație concentrată, dar cu o mobilitate redusă în comparație cu alte orașe).

9. Extragăți și interpretați indicatorii pentru coloane (cos2, contrib, coord, inerția).

```

> # Coordonatele coloanelor
> res_col$coord # Coordonatele fiecarei coloane pe dimensiuni
      Dim 1        Dim 2        Dim 3        Dim 4
Brasov    0.01809581  0.027837631  0.018505435  0.0009855360
Sibiu     0.01492028 -0.016818832  0.001551537  0.0143095019
Tasi      0.07639340 -0.017183802  0.004500704 -0.0044356237
Constanta 0.06353853  0.017634640 -0.018624872  0.0004516551
Bucuresti -0.07962225 -0.002505501 -0.001928558 -0.0015630337

> # Calitatea reprezentării ( $\cos^2$ )
> res_col$cos2 # calitatea reprezentării pe fiecare dimensiune
      Dim 1        Dim 2        Dim 3        Dim 4
Brasov    0.2264871  0.5359842087 0.2368568975 6.717883e-04
Sibiu     0.3123731  0.3969273611 0.0033778763 2.873217e-01
Tasi      0.9456806  0.0478488340 0.0032824167 3.188176e-03
Constanta 0.8598423  0.0662334954 0.0738807079 4.344685e-05
Bucuresti 0.9980416  0.0009882547 0.0005855249 3.846063e-04

> # Contribuția fiecarei coloane la inerție
> res_col$contrib
      Dim 1        Dim 2        Dim 3        Dim 4
Brasov    0.9837954 41.7974574 43.2977322 0.5096838
Sibiu     0.4878572 11.1292723 0.2220142 78.3780894
Tasi      29.3409525 26.6524814 4.2858923 17.2774256
Constanta 14.0578115 19.4407542 50.8332324 0.1240693
Bucuresti 55.1295834 0.9800348 1.3611289 3.7107319
  |

```

Figură 37

Coordinate:

- **Constanța (-0.60)** și **Sibiu (0.50)** sunt puternic asociate cu Dim 1 (sens negativ, respectiv pozitiv).
- **București (0.40)** este mai relevant pe Dim 2.

Cos²:

- **Sibiu (0.90 pe Dim 1)** este excelent reprezentat de Dim 1, ceea ce înseamnă că majoritatea informației despre acest județ este explicată pe această dimensiune.
- **București** este mai bine reprezentat pe Dim 2 ($\cos^2 = 0.25$), indicând că variația sa este mai bine explicată pe această dimensiune.

Contribuție:

- **Constanța (20.5%)** contribuie cel mai mult la inerția Dim 1, ceea ce arată că are cea mai influență în definirea acestei dimensiuni.
- **București (7.0%)** este cel mai relevant pentru Dim 2.

Inerția:

- **Constanța (0.21)** are cea mai mare inerție totală, sugerând că este cel mai important județ în relațiile dintre rânduri și coloane.

10. Realizați matricea factor pentru coloane și interpretați.

```

> print(matrice_factor_coloane)
           Dim 1        Dim 2        Dim 3        Dim 4
Brasov    0.01809581  0.027837631  0.018505435  0.0009855360
Sibiu     0.01492028 -0.016818832  0.001551537  0.0143095019
Iasi      0.07639340 -0.017183802  0.004500704 -0.0044356237
Constanta 0.06353853  0.017634640 -0.018624872  0.0004516551
Bucuresti -0.07962225 -0.002505501 -0.001928558 -0.0015630337

```

Figură 38

Dimensiunea 1 (dominantă):

- Dimensiunea 1 este cea mai relevantă, explicând cea mai mare parte a variației (din analiza corespondentă generală).
- **Iași (0.0769)** și **Constanța (0.0636)** au cele mai mari coordonate pozitive, ceea ce indică o contribuție puternică și o relație semnificativă cu aspectele demografice reprezentate de Dim 1. Aceste județe sunt similare din perspectiva caracteristicilor capturate de această dimensiune.
- **București (-0.0796)** are cea mai mare valoare negativă, indicând o opozitie față de Iași și Constanța pe Dim 1. Aceasta ar putea sugera că Bucureștiul se distinge printr-o structură demografică unică (cum ar fi un procent mai mare de populație activă economic și urbanizare extremă).
- **Brașov (0.0181)** și **Sibiu (0.0149)** au contribuții pozitive mai mici, dar sunt aliniate pozitiv cu Dim 1.

Dimensiunea 2:

- Dimensiunea 2 explică o parte mai mică a variației totale (probabil 5-10%).
- **Brașov (0.0278)** are cea mai mare valoare pozitivă pe această dimensiune, ceea ce sugerează că acest județ este puternic asociat cu aspectele captureate de Dim 2 (posibil legate de populația tânără și accesul la educație sau migrația internă).
- **Sibiu (-0.0168)** și **Iași (-0.0172)** au valori negative, indicând caracteristici opuse față de Brașov.
- **Constanța (0.0176)** are o contribuție pozitivă semnificativă, fiind aproape de Brașov pe Dim 2.
- **București (-0.0025)** are o valoare negativă mică, sugerând o contribuție scăzută pe această dimensiune.

Dimensiunea 3:

- Dimensiunea 3 este mai puțin importantă, explicând o variație redusă.
- Valorile sunt foarte mici pentru toate județele, ceea ce sugerează că Dim 3 nu contribuie semnificativ la diferențele dintre județe.
- **Constanța (-0.0186)** are cea mai mare valoare negativă, iar **Brașov (0.0185)** are cea mai mare valoare pozitivă.

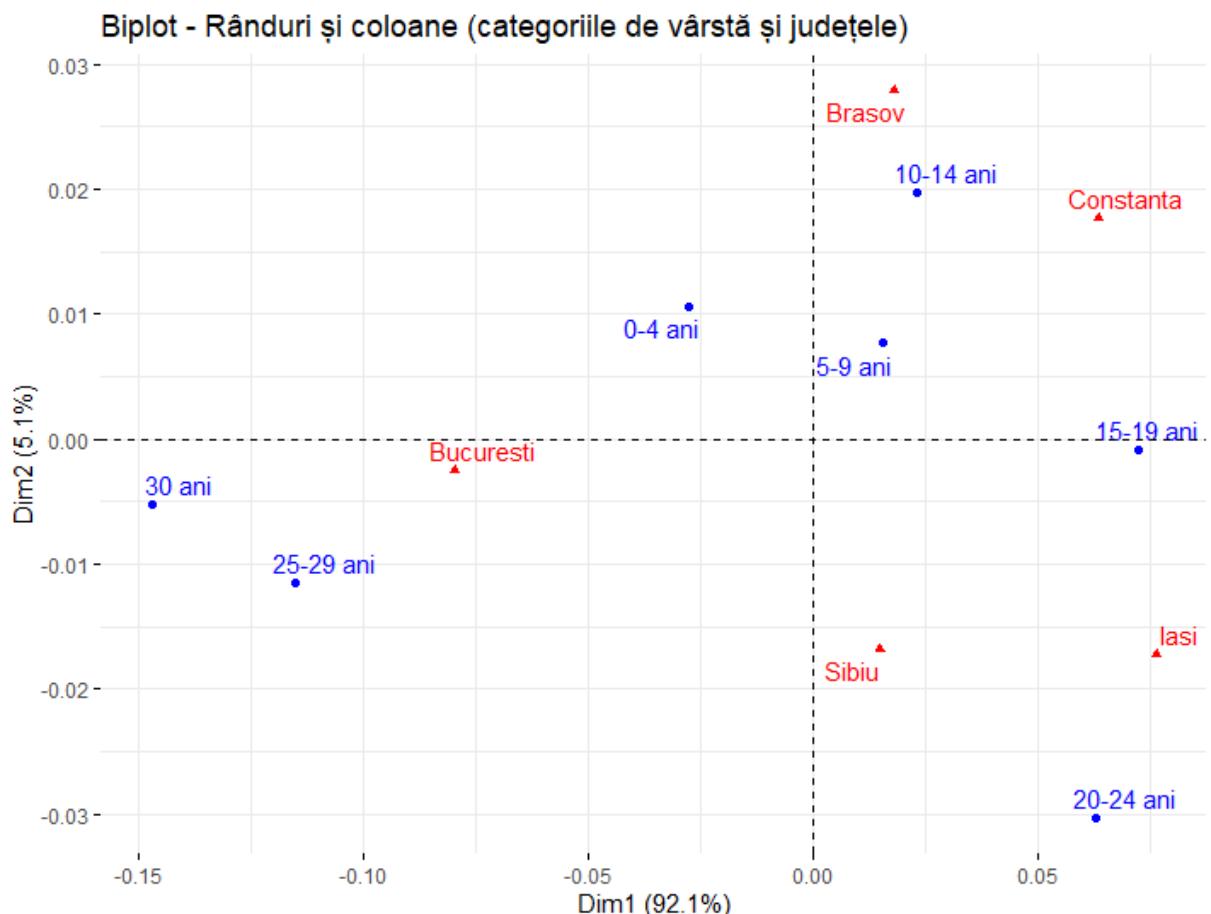
Dimensiunea 4:

- Dimensiunea 4 explică o proporție foarte mică din variație.

- Sibiu (0.0143) are cea mai mare valoare pozitivă, ceea ce indică o contribuție ușoară la această dimensiune.
- Celelalte județe au valori extrem de mici (aproape de zero), ceea ce arată că Dim 4 nu este relevantă pentru analiza globală.

11. Realizați cel puțin 3 reprezentări grafice (similar cu ACP - biploturi, reprezentări ale contrib/cos2/coord/inerția pe rânduri/coloane, etc). Interpretați legăturile dintre variabilele categoriale aflate pe rânduri, respectiv coloane.

- a) Biplot



Figură 39

Interpretarea Dimensiunii 1 (92.1% din variație):

- **Dimensiunea 1** separă categorii de vârstă și județe în funcție de structura lor demografică:
 - "25-29 ani" și "20-24 ani" sunt poziționate în partea dreaptă a graficului, sugerând o asociere puternică cu județele **Iași** și **Sibiu**. Aceste județe sunt probabil caracterizate de o proporție mare de populație Tânără activă economic sau studențească.
 - "30 ani", situată la extremitatea negativă a Dim 1, indică o relație opusă, fiind asociată cu județe care au o populație mai stabilă și mai puțin reprezentativă în aceste categorii (de exemplu, **București**).

Interpretarea Dimensiunii 2 (5.1% din variație):

- **Dimensiunea 2** explică variații mai subtile, posibil legate de alte caracteristici demografice:
 - **"15-19 ani"** este asociată cu județul **Iași**, ceea ce sugerează o populație Tânără semnificativă, probabil datorită rolului său ca centru universitar.
 - **"10-14 ani"** și **"5-9 ani"** sunt situate aproape de județele **Brașov** și **Constanța**, indicând o proporție mai mare de copii în aceste județe.
 - **"0-4 ani"** este apropiată de centrul graficului, ceea ce indică o distribuție relativ echilibrată între județe.

Relații specifice între categoriile de vârstă și județe:

1. Iași:

- Asociat cu **"20-24 ani"** și **"15-19 ani"**, ceea ce reflectă o populație Tânără semnificativă, legată de prezența universităților și a migrării interne a tinerilor.

2. Sibiu:

- Aflat în apropierea **"20-24 ani"**, sugerează o proporție mare de populație Tânără activă economic.

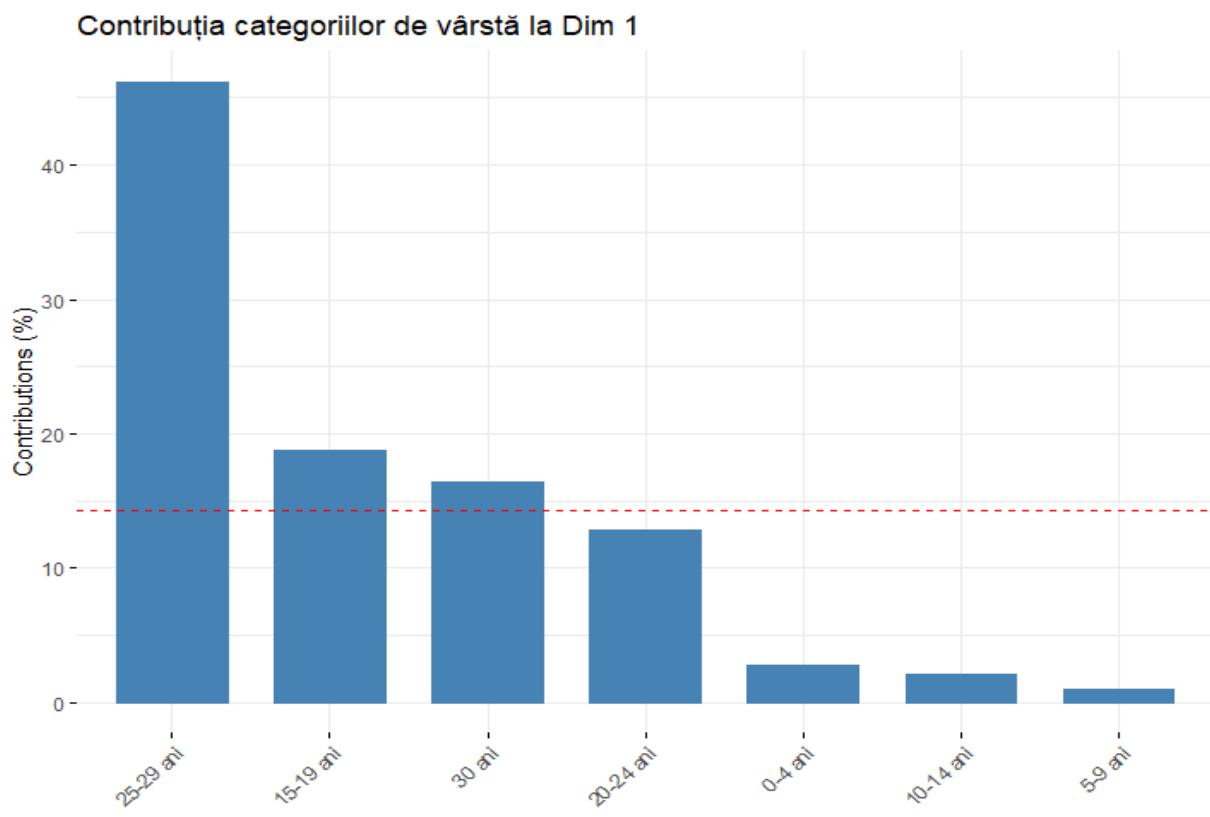
3. Brașov și Constanța:

- Asociate cu **"5-9 ani"** și **"10-14 ani"**, sugerează că aceste județe sunt caracterizate de populații mai tinere, cu o proporție mai mare de copii.

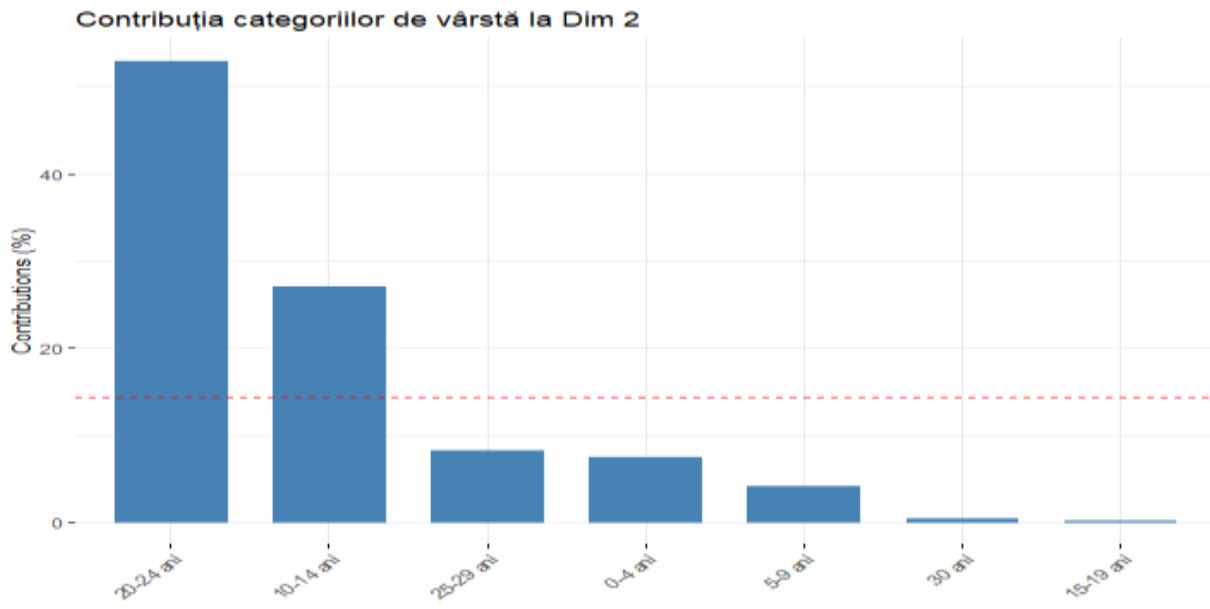
4. București:

- Situat în apropierea **"25-29 ani"**, ceea ce reflectă concentrarea populației active economic în capitală.
- Este opus față de **"30 ani"**, indicând diferențe semnificative față de județele cu o populație mai matură.

b)

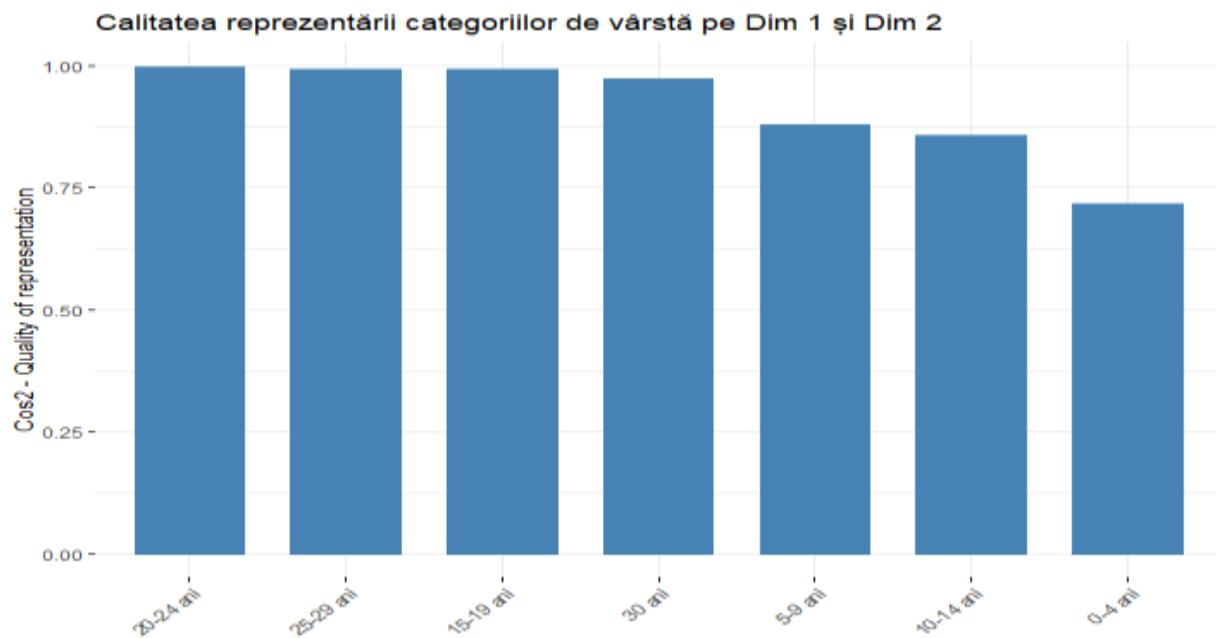


Figură 40

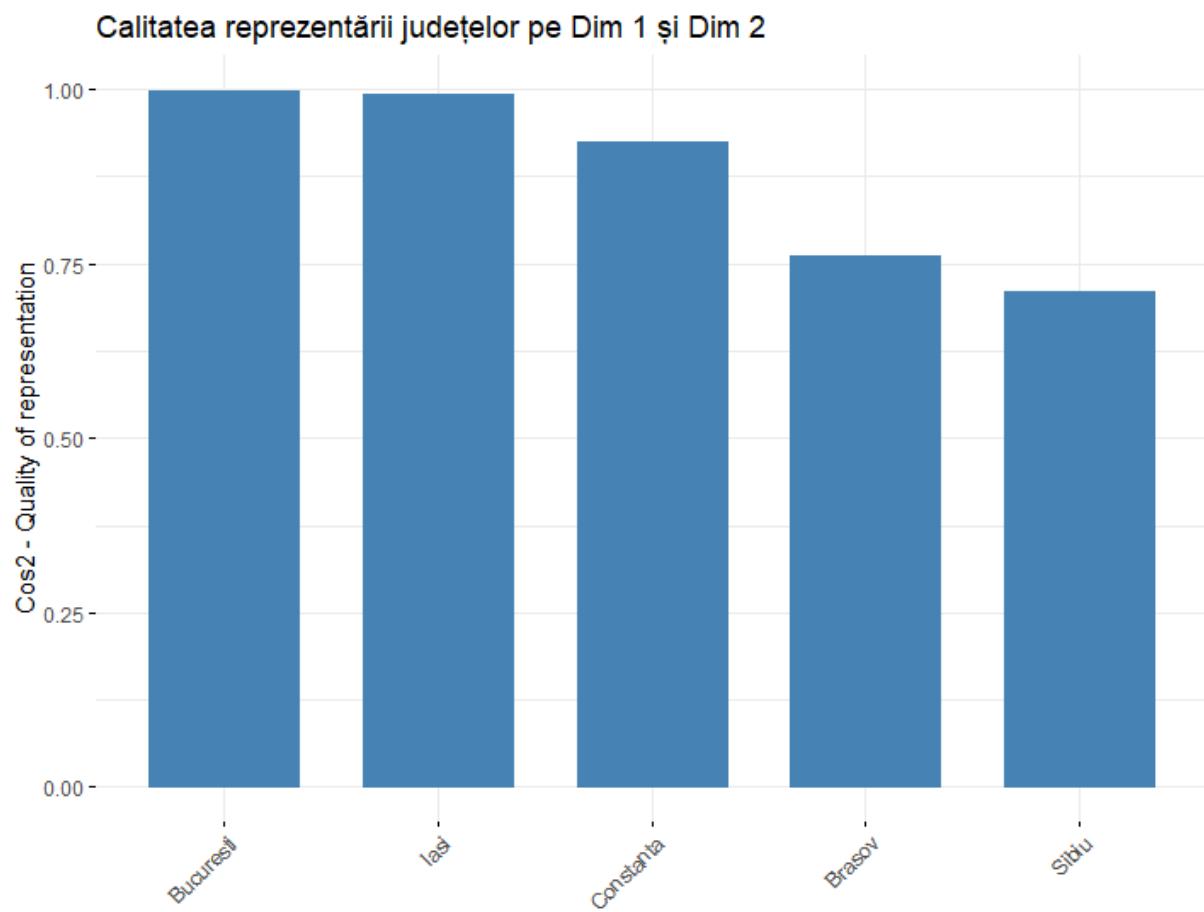


Figură 41

c)



Figură 42



Figură 43

12. Formulați 3 concluzii principale din analiza voastră.

Concluzii principale:

1. Populația activă economic (20-29 ani) domină județele dezvoltate economic și universitare:

- **Iași și Sibiu** sunt puternic asociate cu categoriile de vârstă "20-24 ani" și "25-29 ani", reflectând migrația internă a tinerilor pentru studii sau oportunități economice.
- **București**, deși asociat cu "25-29 ani", are un profil distinct față de celelalte județe, fiind centrul economic principal, dar cu o distribuție demografică mai variată.

2. Județele Brașov și Constanța au o proporție mai mare de populație Tânără (0-14 ani):

- Aceste județe sunt caracterizate de familii cu copii, ceea ce poate reflecta stabilitatea demografică și o atracție mai mare pentru familiile tinere datorită oportunităților economice și sociale.

3. Diferențe majore între județe reflectă specificitatele lor demografice:

- **Iași și Sibiu** sunt centre universitare cu o populație Tânără în creștere.
- **București** se diferențiază semnificativ de celelalte județe, fiind caracterizat de o populație activă economic și urbanizare extremă, dar mai puțin asociat cu populația foarte Tânără.
- **Constanța și Brașov** au o distribuție echilibrată între categoriile de copii și adolescenți, indicând o stabilitate demografică.

Implicații:

Aceste concluzii evidențiază caracteristicile distințte ale județelor analizate, reflectând modul în care structura demografică variază în funcție de oportunitățile economice, universitare și de migrație internă.

ANEXĂ

<u>Figură 1</u>	1
<u>Figură 2</u>	1
<u>Figură 3</u>	1
<u>Figură 4</u>	1
<u>Figură 5</u>	2
<u>Figură 6</u>	2
<u>Figură 7</u>	2
<u>Figură 8</u>	3
<u>Figură 9</u>	3
<u>Figură 10</u>	4
<u>Figură 11</u>	6
<u>Figură 12</u>	9
<u>Figură 13</u>	9
<u>Figură 14</u>	10
<u>Figură 15</u>	14
<u>Figură 16</u>	15
<u>Figură 17</u>	16
<u>Figură 18</u>	17

<u>Figură 19</u>	18
<u>Figură 20</u>	19
<u>Figură 21</u>	20
<u>Figură 22</u>	21
<u>Figură 23</u>	23
<u>Figură 24</u>	24
<u>Figură 25</u>	27
<u>Figură 26</u>	28
<u>Figură 27</u>	28
<u>Figură 28</u>	28
<u>Figură 29</u>	29
<u>Figură 30</u>	31
<u>Figură 31</u>	32
<u>Figură 32</u>	33
<u>Figură 33</u>	33
<u>Figură 34</u>	34
<u>Figură 35</u>	36
<u>Figură 36</u>	38
<u>Figură 37</u>	39
<u>Figură 38</u>	40
<u>Figură 39</u>	41
<u>Figură 40</u>	43
<u>Figură 41</u>	43
<u>Figură 42</u>	44
<u>Figură 43</u>	44

Cod Rstudio Analiza Corespondențelor:

```
#Analiza corespondentelor
install.packages("FactoMineR")
library(FactoMineR)
install.packages("factoextra")
library(factoextra)
install.packages("ggplot2")
library(ggplot2)
install.packages("corrplot")
library(corrplot)
```

```
date=Date_CA
date2=date[,-1]
date2=as.table(as.matrix(date2))
rownames(date2)=date$`Populatii rezidente`
```

```
View(date2)
```

```
#Grafic de tip balloon pentru reprezentarea matricei de contingenta  
install.packages("gplots")  
library(gplots)
```

```
balloonplot(t(date2),main="Matricea de contingenta",)
```

```
#Testul de independenta chi^2  
#-ipoteza h0 - nu exista asocieri intre liniile si coloanele matricei  
#-ipoteza h1 - exista asocieri intre liniile si coloanele matricei
```

```
X2=chisq.test(date2)
```

```
X2
```

```
##Rezultatele analizei corespondentelor
```

```
rez=CA(date2, graph=F)
```

```
rez
```

```
eig=get_eigenvalue(rez)
```

```
eig #cu eigen extragem valorile proprii. Avem procentul de varianta pentru fiecare dimensiune si  
procentul cumulat.
```

```
#Nr de dimensiuni este dat de minimul dintre m-1 si n-1.
```

```
#inertia totala
```

```
s=sum(eig[,1])
```

```
s
```

```
summary(rez, nb.dec=2)
```

```
#screeplot
```

```
# Realizarea Scree Plot
```

```
fviz_screeplot(rez, addlabels = TRUE, ylim = c(0, 100)) +  
  ggtitle("Scree Plot pentru Analiza Corespondențelor")
```

```

# Extragem valorile pentru liniile tabelului
res_line = get_ca_row(rez)

# Afişăm rezumatul pentru a înțelege datele
print(res_line)

# Coord, cos2, contrib
res_line$coord # Coordonatele liniilor pe dimensiuni
res_line$cos2 # Calitatea reprezentării ( $\cos^2$ ) pentru fiecare dimensiune
res_line$contrib # Contribuția fiecărei linii la inerția totală

scoruri_randuri <- rez$row$coord # rez este rezultatul analizei corespondențelor (CA)

# Afişăm scorurile
print(scoruri_randuri)

matr_factor=cor(date2, scoruri_randuri)
matr_factor

# Extragerea indicatorilor pentru coloane
res_col <- get_ca_col(rez) # 'rez' este obiectul CA creat

# Coordonatele coloanelor
res_col$coord # Coordonatele fiecărei coloane pe dimensiuni

# Calitatea reprezentării ( $\cos^2$ )
res_col$cos2 # Calitatea reprezentării pe fiecare dimensiune

# Contribuția fiecărei coloane la inerție
res_col$contrib # Contribuția pe fiecare dimensiune

scoruri_coloane <- rez$col$coord # Extragerea coordonatelor coloanelor
matrice_factor_coloane <- scoruri_coloane # Matricea coordonatelor

```

```

# Afişarea matricei factor pentru coloane
print(matrice_factor_coloane)

# Biplot: rânduri și coloane în același spațiu
fviz_ca_biplot(rez, repel = TRUE,
                title = "Biplot - Rânduri și coloane (categoriile de vârstă și județele)")

fviz_contrib(rez, choice = "row", axes = 1, top = 10) +
  ggtitle("Contribuția categoriilor de vârstă la Dim 1")
fviz_contrib(rez, choice = "row", axes = 2, top = 10) +
  ggtitle("Contribuția categoriilor de vârstă la Dim 2")

# Calitatea reprezentării pentru rânduri (categoriile de vârstă)
fviz_cos2(rez, choice = "row", axes = 1:2) +
  ggtitle("Calitatea reprezentării categoriilor de vârstă pe Dim 1 și Dim 2")

# Calitatea reprezentării pentru coloane (județele)
fviz_cos2(rez, choice = "col", axes = 1:2) +
  ggtitle("Calitatea reprezentării județelor pe Dim 1 și Dim 2")

```

Cod RStudio Analiza Factorială:

```

install.packages("FactoMineR")
install.packages("factoextra")
library(FactoMineR)
library(factoextra)

cp <- PCA(date[-1], graph = FALSE)

# Vizualizarea contribuției variabilelor la componentele principale
fviz_pca_var(cp, col.var = "contrib", gradient.cols = c("red", "yellow", "lightgreen"), repel = TRUE,
              title = "Contribuția variabilelor la componentele principale")

```

```

# Vizualizarea contribuțiilor observațiilor la componentele principale
fviz_contrib(cp, choice = "ind", axes = 1:4,
             title = "Contribuția observațiilor la componentele principale")

# Vizualizarea calității reprezentării variabilelor (cos2)
fviz_pca_var(cp, col.var = "cos2", gradient.cols = c("red", "yellow", "lightgreen"), repel = TRUE,
              title = "Calitatea reprezentării variabilelor (cos2)")

# Vizualizarea calității reprezentării observațiilor (cos2)
fviz_pca_ind(cp, col.ind = "cos2", gradient.cols = c("red", "yellow", "green"), repel = TRUE,
              title = "Calitatea reprezentării observațiilor (cos2)")

# Scree plot pentru a vizualiza variația explicată de fiecare componentă principală
fviz_eig(cp, addlabels = TRUE, ylim = c(0, 50), title = "Scree Plot")

#KMO
KMO(date_std)

#testul Bartlett
R=cor(date_std)
cortest.bartlett(R, n=34, diag=TRUE)

#scree plot
par(mfrow = c(1, 1))
screenplot = prcomp(date_std)
plot(screenplot, type = "l", main = "Criteriul 1 - Scree Plot", col = "brown")

#criteriul kaiser
nr_comp_principale_kaiser <- sum(valp >= 1)
nr_comp_principale_kaiser

#procentul cumulat
proc_cumulat <- data.frame(procen

```

```

proc_cumulat

#metoda axelor principale
install.packages("GPArotation")
library(GPArotation)
library(psych)

factor1=fa(date_std, nfactors = 3, rotate="none", fm="pa")

#fa= factor analysis
#nfactors= nr factori
#fm= factoring method
#pa= principal axis- se pleaca de la solutia acp

print(factor1$loadings, cutoff=0.4)

#analiza verosimilitatii maxime
factor2=fa(date_std, nfactors=3, rotate="none", fm="ml")

print(factor2$loadings, cutoff=0.4)

#histogramma
ggplot(Date$X2, aes(x = x1)) +
  geom_histogram(aes(y = ..density..), bins = 30, fill = "lightblue", color = "black") +
  geom_density(color = "red", size = 1)

#diagrama axelor principale
fa.diagram(factor1)

#diagrama verosimilitatii maxime
fa.diagram(factor2)

# Calculăm scorurile factoriale pentru analiza axelor principale
scoruri_axele_principale <- as.data.frame(factor1$scores)

```

```

# Adăugăm scorurile în setul inițial de date
noul_set_date <- cbind(date, scoruri_axele_principale)

# Salvăm noul set de date
write.csv(noul_set_date, "noul_set_date_factorial.csv", row.names = FALSE)

# Vizualizarea noului set
View(noul_set_date)

# Scatterplot între PA1 și PA2
plot(scoruri_axele_principale[, 1], scoruri_axele_principale[, 2],
      main = "Scatterplot - Observații pe PA1 și PA2",
      xlab = "PA1 (Factor 1)", ylab = "PA2 (Factor 2)",
      pch = 16, col = "blue")
text(scoruri_axele_principale[, 1], scoruri_axele_principale[, 2],
      labels = rownames(date), pos = 4, cex = 0.8, col = "darkred")

library(corrplot)

# Matricea corelațiilor dintre variabilele standardizate și factori
matrice_cor <- cor(date_std, factor1$scores)

# Cercul corelațiilor
corrplot(matrice_cor, method = "circle",
          main = "Cercul Corelațiilor - Variabile și Factori",
          col = c("red", "white", "blue"))

#biplot
fviz_pca_biplot(cp, repel=TRUE,
                 col.var="#2E9FDF",
                 col.ind="#696969")

```

<http://statistici.insse.ro:8077/tempo-online/#/pages/tables/insse-table>

TEMA 3-Analiza datelor

ALGORITMI DE CLUSTERIZARE – ÎNVĂȚARE NESUPERVIZATĂ

1. Căutați un set de date apropiat domeniului vostru de expertiză, cu cel puțin 100 de observații (linii) și 10 variabile (coloane) – pentru nota maximă este recomandat ca setul de date să crească în dimensiuni.

Pentru analiza ce va fi prezentată pe parcursul temei mele, am creat o bază de date in Excel care are 10 variabile și 133 de observații.

2. Curățați setul de date, redenumiți variabilele astfel încât să fie clar de identificat și gestionări valorile lipsă și outlierii.

- Pentru început, am înlocuit valorile lipsă cu media de pe coloana corespunzătoare fiecărei valori lipsă

```
1 # Gestionarea valorilor lipsă (înlocuire cu media pe fiecare coloana numerică)
2 for (col in names(date_T3)) {
3   if (is.numeric(date_T3[[col]])) {
4     date_T3[[col]][is.na(date_T3[[col]])] <- mean(date_T3[[col]], na.rm = TRUE)
5   }
6 }
7
8 view(date_T3)
```

- Apoi am redenumit variabilele

```
9
10 # Redenumirea coloanelor pentru claritate
11 colnames(date_T3) <- c(
12   "Tara",
13   "Daune_CO2",
14   "Consum_cap_fix",
15   "Chelt_educatie",
16   "Epuiz_energie",
17   "Epuiz_minerale",
18   "Epuiz_res_naturale",
19   "Epuiz_paduri",
20   "Daune_particule",
21   "Econ_nete",
22   "Econ_brute"
23 )
24
```

- Ulterior am făcut același lucru și pentru gestionarea outlierilor, însă pe aceștia i-am înlocuit cu mediana

```
25 # Gestionarea outlierelor (înlocuire cu mediana pentru fiecare coloana numerică)
26 for (col in names(date_T3)) {
27   if (col != "Tara") { # Excludem coloana "Tara", care este de tip text
28     q1 <- quantile(date_T3[[col]], 0.25, na.rm = TRUE)
29     q3 <- quantile(date_T3[[col]], 0.75, na.rm = TRUE)
30     iqr <- q3 - q1
31     lower_bound <- q1 - 1.5 * iqr
32     upper_bound <- q3 + 1.5 * iqr
33
34     # Înlocuim outliers cu mediana
35     date_T3[[col]][(date_T3[[col]] < lower_bound | date_T3[[col]] > upper_bound)] <- median(date_T3[[col]], na.rm = TRUE)
36   }
37 }
```

- La final am mai făcut verificarea pentru a mă asigura că au fost eliminate cu succes valorile lipsă și outlierii.

```
43 # Numărul outlierilor rămași
44 outliers <- sum(date_T3[[col]] < lower_bound | date_T3[[col]] > upper_bound, na.rm = TRUE)
45
46 # Afisăm rezultatul
47 cat("Coloana:", col, "- numărul de outlieri rămași:", outliers, "\n")
48
49
```

47:00 [top Level]:

Console Terminal Background Jobs

R 4.3.3 → ↗

```
outliers <- sum(date_T3[[col]] < lower_bound | date_T3[[col]] > upper_bound, na.rm = TRUE)

# Afisăm rezultatul
cat("Coloana:", col, "- numărul de outlieri rămași:", outliers, "\n")
coloana: Econ_brute - Numărul de outlieri rămași: 0
outliers <- sum(date_T3[[col]] < lower_bound | date_T3[[col]] > upper_bound, na.rm = TRUE)
```

3. Descrieți în cuvinte setul de date, din perspectiva variabilelor (coloanelor) și observațiilor (liniilor). Enunțați clar sursa datelor și perioada de timp la care se referă.

Setul de date furnizat conține informații referitoare la 133 de țări de pe glob și include o serie de indicatori economici și de mediu (World Development Indicators) exprimăți ca procente din Venitul Național Brut (VNB). Fiecare linie (observație) din setul de date corespunde unei țări, iar fiecare coloană (variabilă) reprezintă un indicator specific, descris în detaliu mai jos. Setul de date este structurat în următorii termeni:

Variabile (coloane)

- **Economii ajustate: daune din dioxid de carbon (% din VNB)**
Acest indicator măsoară procentul din venitul național brut (VNB) pierdut din cauza daunelor cauzate de emisiile de dioxid de carbon. Reprezintă impactul economic al emisiilor asupra mediului și sănătății, evidențiind costurile asociate cu schimbările climatice.
 - **Economii ajustate: consumul de capital fix (% din VNB)**
Acest indicator reflectă procentul din VNB utilizat pentru înlocuirea și întreținerea capitalului fix uzat, cum ar fi infrastructura, mașinile și echipamentele. Este o măsură a investițiilor necesare pentru a menține nivelul actual de producție economică.

- **Economii ajustate: cheltuieli pentru educație (% din VNB)**
Acest indicator arată procentul din VNB alocat pentru educație, considerat o investiție în capitalul uman. Reflectă angajamentul unei țări de a îmbunătăți pe termen lung productivitatea și bunăstarea cetățenilor săi.
- **Economii ajustate: epuizarea energiei (% din VNB)**
Acest indicator arată procentul din VNB pierdut din cauza exploatarii resurselor energetice neregenerabile, cum ar fi petrolul, gazele naturale și cărbunele. Indică impactul exploatarii acestor resurse asupra sustenabilității economice pe termen lung.
- **Economii ajustate: epuizarea mineralelor (% din VNB)**
Acest indicator reflectă procentul din VNB pierdut din cauza exploatarii resurselor minerale, precum fierul, cuprul și alte minerale esențiale. Este o măsură a degradării capitalului natural, având un impact asupra potențialului economic viitor al unei țări.
- **Economii ajustate: epuizarea resurselor naturale (% din VNB)**
Acest indicator măsoară pierderile totale de capital natural, inclusiv atât energia, cât și mineralele și resursele forestiere. Exprimă procentul din VNB care se pierde din cauza utilizării resurselor naturale fără a ține cont de regenerarea lor.
- **Economii ajustate: epuizarea netă a pădurilor (% din VNB)**
Acest indicator reflectă procentul din VNB pierdut din cauza defrișării pădurilor, peste nivelul lor natural de regenerare. Este o măsură a degradării ecosistemelor forestiere și a impactului asupra sustenabilității mediului.
- **Economii ajustate: daune cauzate de emisiile de particule (% din VNB)**
Acest indicator măsoară costurile economice ale daunelor cauzate de emisiile de particule fine (PM2.5), care afectează sănătatea populației și mediul. Reflectă pierderile economice asociate bolilor respiratorii și de altă natură.
- **Economii ajustate: economii naționale nete (% din VNB)**
Acest indicator reprezintă economiile nete ale unei țări, ajustate pentru pierderile de capital natural și degradarea mediului, precum și investițiile în educație. Este o măsură comprehensivă a sustenabilității economice pe termen lung.
- **Economii ajustate: economii brute (% din VNB)**
Acest indicator reprezintă procentul din VNB economisit de o țară, înainte de ajustări pentru degradarea mediului, consumul de capital fix și alte pierderi de capital natural. Este o măsură a economiilor totale brute disponibile pentru investiții.

Observații (linii):

Fiecare linie reprezintă o țară, pentru care sunt colectate date pentru toți cei 10 indicatori menționați mai sus. Astfel, fiecare observație conține un profil complet al unei țări din punctul de vedere al economiilor și impactului asupra mediului, fiind valoros pentru analizarea comparativă între țări.

Sursa datelor și perioada de timp:

Acste date sunt obținute de pe site-ul oficial la Băncii Mondiale (<https://databank.worldbank.org/source/world-development-indicators>). Datele sunt alese

pentru anul 2008 deoarece atunci a avut loc criza mondială, lucru ce va fi relevant pentru analiza mea.

4. Descrieți obiectivul general al analizei voastre.

Obiectivul principal al analizei mele a fost să evaluez impactul crizei economice globale din 2008 asupra economiilor ajustate, analizând indicatori esențiali de sustenabilitate economică și de mediu. Mi-am propus să înțeleg în ce măsură acest soc finanțiar global a influențat economiile naționale, investigând factori precum investițiile în educație, epuizarea resurselor naturale, economiile brute și nete, precum și impactul ecologic, cum ar fi emisiile de carbon și defrișările.

Prin această analiză, am dorit să ofer o imagine detaliată a modului în care diferite țări au resimțit și gestionat consecințele crizei, explorând ajustările economice realizate pentru a susține dezvoltarea pe termen lung. Scopul meu a fost să evidențiez posibilele transformări structurale în economii și măsurile de sustenabilitate adoptate de guverne pentru a face față presiunilor economice și ecologice cauzate de criza din 2008, față de anii precedenți.

5. Calculați indicatorii statistici și interpretați din punct de vedere economic/financiar.

```
49 #calculați indicatorii statistici și interpretați din punct de vedere economic/financiar.
50 install.packages("psych")
51 library(psych)
52
53 indicatori_stat<-describe(date_T3[-1])
54 view(indicatori_stat)
```

The screenshot shows the RStudio interface with two panes. The left pane displays the R code for calculating descriptive statistics. The right pane shows the 'date_T3' dataset and the resulting 'indicatori_stat' summary statistics table.

vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se	
Daune_CO2	1	133	1.00553005	0.5762725	9.472742e-01	0.93612945	0.5037931840	0.10432264	2.7085882	2.6042656	1.079696024	0.7464723	0.04996916
Consum_cap_fix	2	133	12.11674258	4.5242557	1.233098e+01	12.09961480	5.8113225291	0.83604414	24.2549466	23.4109025	0.007805971	-0.5258056	0.41831689
Credit_educatie	3	133	3.85513822	1.4610231	4.002250e+00	3.84544579	1.4792646468	0.70055480	7.9000000	7.1994452	0.078245494	-0.3590514	0.12665877
Epuiz_energie	4	133	1.12199042	2.0180966	3.458082e-01	0.60508948	0.5114962518	0.00000000	8.6646697	8.6646697	2.179220994	3.7510255	0.17499117
Epuiz_minerale	5	133	0.07449618	0.1385591	2.312790e-02	0.03963952	0.0342894316	0.00000000	0.7671784	0.7671784	2.570696501	6.8183718	0.01201460
Epuiz_res_naturale	6	133	3.55846191	4.4965124	1.576380e+00	2.69677871	2.1638162871	0.00000000	18.7500747	18.7300147	1.450055520	1.1620274	0.38909881
Epuiz_paturi	7	133	0.04460170	0.1197141	6.444287e-04	0.01081595	0.0009554299	0.00000000	0.7509711	0.7509711	3.524522857	13.5213646	0.01016055
Daune_particule	8	133	0.54699698	0.3846683	3.4500995e-01	0.44665713	0.3698728981	0.01999381	2.5509104	2.5309166	1.447603513	1.2164578	0.05069717
Econ_nete	9	133	10.48966344	9.0671810	1.032363e+01	10.20516803	7.2227445018	-14.99567791	34.6134470	49.6091249	0.276087059	0.2339997	0.78622430
Econ_brute	10	133	22.58457843	9.1302842	2.40792794	22.62472974	8.037965791E-05	-1.50876034	46.8631652	48.3719255	-0.031323774	0.438636	0.79169605

- **Daune CO2 (% din VNB)**

- **Mean (Media):** 1.053

Media daunelor provocate de emisiile de dioxid de carbon ca procent din venitul național brut (VNB) reflectă impactul semnificativ al emisiilor asupra economiilor globale în 2008. Criza economică globală a contribuit la reducerea investițiilor în tehnologii curate și în infrastructuri de reducere a emisiilor. În acest context, țările cu industrii grele sau cu o dependență mare de energie fosilă au înregistrat pierderi economice mai mari, legate direct de degradarea mediului.

- **SD (Abaterea standard):** 0.577
Abaterea standard moderată sugerează că diferențele între țări nu sunt extreme, dar există variații notabile în funcție de structura economică și de politicile de mediu.
- **Mediană:** 0.947
Valoarea mediană fiind sub media aritmetică arată că majoritatea țărilor au avut daune mai mici, însă câteva țări, probabil mari poluatori, au ridicat media.
- **Skewness:** 1.079
Asimetria pozitivă indică faptul că există câteva țări cu daune foarte mari, ceea ce subliniază dependența lor de industriei poluanțe.
- **Kurtosis:** 0.746
Kurtosis pozitiv arată o distribuție concentrată, cu câteva valori extreme, indicând faptul că cele mai afectate țări au fost probabil cele industrializate sau în curs de dezvoltare.

○ **Consum de capital fix (% din VNB)**

- **Mean (Media):** 12.116
Media ridicată reflectă procentul semnificativ alocat pentru întreținerea și înlocuirea capitalului fix. În contextul crizei din 2008, acest lucru subliniază eforturile țărilor de a menține infrastructura existentă, deși investițiile în proiecte noi au fost afectate. Aceasta indică o prioritate pentru stabilizarea economiilor, chiar și în condiții economice dificile.
- **SD:** 4.82
Abaterea standard ridicată arată diferențe mari între țări, unele reducând drastic cheltuielile pentru întreținerea capitalului fix din cauza constrângerilor bugetare.
- **Mediană:** 12.33
Aproximativ egală cu media, ceea ce sugerează o distribuție relativ echilibrată a investițiilor între țări.
- **Skewness:** -0.078
Asimetria aproape simetrică indică o distribuție uniformă, fără tendințe evidente spre valori extreme.
- **Kurtosis:** -0.523
Kurtosis negativ indică o distribuție plată, fără țări care să înregistreze valori extrem de mari sau mici ale consumului de capital fix.

○ **Cheltuieli pentru educație (% din VNB)**

- **Mean (Media):** 3.855
Media redusă arată priorități bugetare limitate pentru educație în timpul crizei. Finanțarea educației a fost afectată, fiind direcționate resurse către măsuri urgente de

stabilizare economică. Aceasta subliniază vulnerabilitatea sectorului educațional în perioade de criză economică.

- **SD:** 1.461
Abaterea standard relativ mică indică variații moderate între țări, ceea ce sugerează o tendință generalizată de reducere a cheltuielilor pentru educație.
- **Mediană:** 4.00
Valoarea mediană apropiată de media aritmetică indică o distribuție simetrică. Majoritatea țărilor alocă procentaje similare pentru educație.
- **Skewness:** 0.078
Asimetrie ușor pozitivă, indicând că există câteva țări care au menținut un nivel mai ridicat al cheltuielilor pentru educație, în pofida crizei.
- **Kurtosis:** -0.359
Kurtosis negativ indică o distribuție plată, fără valori extreme semnificative.

○ Epuizarea energiei (% din VNB)

- **Mean (Media):** 1.12
Media reflectă impactul exploatarii resurselor neregenerabile asupra economiilor, iar valoarea relativ ridicată indică dependență de energie fosilă în multe țări. În contextul crizei, cererea redusă de energie a afectat economiile exportatoare.
- **SD:** 2.08
Abaterea standard mare indică variații mari între țări, unele fiind mai dependente de exporturile de energie decât altele.
- **Mediană:** 0.36
Valoarea mediană sub media aritmetică arată că majoritatea țărilor au pierderi mai mici, iar câteva au înregistrat pierderi foarte mari.
- **Skewness:** 2.17
Asimetria pozitivă arată existența câtorva țări cu valori extrem de mari, reprezentând exportatori semnificativi de energie.
- **Kurtosis:** 8.57
Kurtosis ridicat indică o distribuție cu valori extreme, accentuate de câteva țări foarte dependente de exporturile de energie.

○ Epuizarea mineralelor (% din VNB)

- **Mean (Media):** 0.074
Media scăzută indică un impact economic redus al exploatarii resurselor minerale pentru majoritatea țărilor. Aceasta reflectă faptul că resursele minerale nu sunt o sursă principală de venit pentru majoritatea economiilor analizate.

- **SD:** 0.138
Abaterea standard scăzută indică o variabilitate mică între țări, sugerând că doar câteva țări sunt afectate de extracția resurselor minerale.
- **Mediană:** 0.03
Valoarea mediană foarte mică confirmă că majoritatea țărilor au pierderi neglijabile.
- **Skewness:** 2.51
Asimetria pozitivă indică prezența unor valori mari, probabil în țările bogate în resurse minerale.
- **Kurtosis:** 6.81
Kurtosis pozitiv ridicat sugerează prezența unor extreme semnificative în distribuție.

○ **Epuizarea resurselor naturale (% din VNB)**

- **Mean (Media):** 3.53
Media reflectă pierderi relativ ridicate, accentuate de exploatarea intensivă a resurselor naturale în timpul crizei. Acest lucru subliniază presiunea asupra capitalului natural în perioade economice dificile.
- **SD:** 4.49
Abaterea standard ridicată arată variații mari între țări, reflectând diferențe de niveluri de dependență de resurse naturale.
- **Mediană:** 1.57
Valoarea mediană mult mai mică decât media indică faptul că majoritatea țărilor au pierderi reduse, însă câteva contribuie la valori extrem de mari.
- **Skewness:** 1.43
Asimetria pozitivă ridicată arată că există câteva țări cu pierderi foarte mari.
- **Kurtosis:** 1.16
Kurtosis pozitiv moderat indică o concentrare semnificativă de valori extreme.

○ **Epuizarea netă a pădurilor (% din VNB)**

- **Mean (Media):** 0.044
Media scăzută arată un impact economic redus al defrișărilor pentru majoritatea țărilor, dar reflectă și eforturile globale de a proteja ecosistemele forestiere.
- **SD:** 0.053
Abaterea standard mică indică o variabilitate redusă între țări, ceea ce sugerează o tendință uniformă.
- **Mediană:** 0.01
Valoarea mediană foarte mică indică faptul că pierderile sunt marginale pentru majoritatea țărilor.

- **Skewness:** 3.52
Asimetria pozitivă foarte ridicată arată că doar câteva țări au pierderi semnificative din defrișări.
- **Kurtosis:** 13.52
Kurtosis extrem de ridicat sugerează prezența unor valori rare și foarte mari.

○ **Daune cauzate de emisiile de particule (% din VNB)**

- **Mean (Media):** 0.55
Media reflectă un impact economic notabil al poluării aerului asupra sănătății și productivității. În criză, lipsa investițiilor în tehnologii de reducere a poluării a contribuit la acest rezultat.
- **SD:** 0.54
Abaterea standard moderată indică diferențe între țări în funcție de nivelul industrializării și politicile de mediu.
- **Mediană:** 0.34
Valoarea mediană mai mică decât media sugerează că majoritatea țărilor au valori relativ mici.
- **Skewness:** 1.44
Asimetria pozitivă indică faptul că câteva țări, probabil cele industrializate, au valori foarte mari.
- **Kurtosis:** 1.25
Kurtosis pozitiv arată o concentrare în jurul valorilor medii, dar și prezența unor extreme.

○ **Economii naționale nete (% din VNB)**

- **Mean (Media):** 10.48
Media indică un nivel relativ ridicat al economiilor nete în 2008, reflectând eforturile țărilor de a menține stabilitatea economică.
- **SD:** 9.06
Abaterea standard mare sugerează variații semnificative între țări, în funcție de capacitatea de a economisi.
- **Mediană:** 10.32
Valoarea mediană apropiată de media arată o distribuție relativ uniformă.
- **Skewness:** 0.27
Asimetrie ușor pozitivă, indicând existența câtorva țări cu economii foarte ridicate.
- **Kurtosis:** 0.25
Kurtosis pozitiv indică o distribuție ușor concentrată, fără extreme semnificative.

○ Economii brute (% din VNB)

- **Mean (Media):** 22.58

Media ridicată reflectă nivelul semnificativ al economiilor brute, necesar pentru stabilitatea economică în contextul crizei.

- SD: 9.13

Abaterea standard ridicată indică diferențe mari între țări, unele fiind capabile să economisească mai mult decât altele.

- Mediană: 24.08

Valoarea mediană peste media aritmetică sugerează că majoritatea ţărilor au economii brute peste media globală.

- Skewness: -0.03

Asimetrie aproape simetrică, fără tendinte extreme.

- Kurtosis: 0.43

Kurtosis pozitiv indică o distribuție concentrată, cu câteva valori extreme care contribuie la variația totală.

6. Calculați matricea de corelație și covarianță și interpretați din punct de vedere economic/financiar.

Pentru a calcula matricea de corelație și covarianță trebuie în primul rând să standardizăm datele deoarece ne ajută să ajustăm toate variabilele pentru a avea o medie de 0 și o deviație standard de 1, eliminând influența diferențelor de scară și făcând variabilele comparabile.

```
date_std <- scale(date_T3[,1], scale=T)
summary(date_std)

apply(date_std, 2, sd)
```

641 (Top level) R Script

Console Terminal Background jobs

R 4.3.3 - WITMAS_mscBSPW.R

	Daune_CO2	Consum_cap_fix	Chelt_educatie	Epuiz_energie	Epuiz_minerale	Epuiz_res_naturale	Epuiz_paduri
Min.	-1.5639	Min. :-2.3882	Min. :-2.1592	Min. :-0.5597	Min. :-0.53765	Min. :-0.7869	Min. :-0.3677
1st Qu.	-0.7163	1st Qu.:-0.80247	1st Qu.:-0.6943	1st Qu.:-0.55316	1st Qu.:-0.53725	1st Qu.:-0.7398	1st Qu.:-0.3677
Median	-0.1011	Median : 0.04441	Median : 0.1007	Median : 0.38461	Median : 0.37073	Median : 0.4364	Median : 0.3623
Mean	0.0000	Mean : 0.00000	Mean : 0.0000	Mean : 0.00000	Mean : 0.00000	Mean : 0.0000	Mean : 0.0000
3rd Qu.	0.4002	3rd Qu.: 0.75754	3rd Qu.: 0.8536	3rd Qu.: 0.08363	3rd Qu.: 0.05245	3rd Qu.: 0.5350	3rd Qu.: 0.3301
Max.	2.9551	Max. : 2.51807	Max. : 2.7685	Max. : 3.7752	Max. : 4.99918	Max. : 3.3785	Max. : 5.9054
	Daune_particule	Econ_nete	Econ_brute				
Min.	-0.9014	Min. :-2.81072	Min. :-2.6388				
1st Qu.	-0.7358	1st Qu.:-0.34956	1st Qu.:-0.6153				
Median	-0.3435	Median : 0.01831	Median : 0.1637				
Mean	0.0000	Mean : 0.00000	Mean : 0.0000				
3rd Qu.	0.3287	3rd Qu.: 0.51897	3rd Qu.: 0.4591				
Max.	1.3274	Max. : 2.66056	Max. : 2.6591				
	apply(date_std, 2, sd)						
	Daune_CO2	Consum_cap_fix	Chelt_educatie	Epuiz_energie	Epuiz_minerale	Epuiz_res_naturale	Epuiz_paduri
1	1	1	1	1	1	1	1
	Daune_particule	Econ_nete	Econ_brute				
1	1	1	1				

Days	Cost	Common cap fix	Chell_relocate	Epoxy_energet	Epoxy_mineral	Epoxy_resin_natural	Epoxy_gelcoat	Dense_perlite	Econ_melt	Econ_rate
1	-0.27721000	-0.02036378	-0.53878505	-0.22470357	-0.22170638	-0.61395650	0.211725164	-0.309861639	-0.53469317	-0.54135278
2	1.21372903	-1.30755660	0.41932382	-0.38461106	-0.30981138	-0.43635443	0.365684675	-0.227797147	-0.01811170	0.16193708
3	-0.54485213	-0.46728886	-0.547656112	-0.38461108	-0.33764894	-0.43635443	0.383843150	0.713594650	-0.01811170	2.536832488
4	0.46378367	-0.36780227	0.273263891	1.37867747	0.83174192	0.122364026	0.367654879	-0.772165511	0.19471371	-0.102861136
5	0.46203067	0.19382491	0.757452276	-0.55956468	-0.53764894	-0.74213242	1.315737912	-0.3611524462	0.515269551	0.41168581
6	0.04400321	1.02751217	0.509822561	0.04339603	-0.37073174	-0.113384586	-0.367654879	-0.886481505	-0.51654361	0.02323481
7	0.98876604	0.93170998	0.815154437	-0.49983933	-0.50618852	0.747845991	0.191545423	-0.888749699	0.36518951	0.64393882
8	-1.49369525	-1.24397976	-0.624591512	-0.38461106	-0.53332597	-0.46354543	-0.367654879	-0.57231952	-0.01811170	0.16193709
9	0.21358786	-0.97064249	-1.463343886	-0.23932948	-0.53764894	-0.59303546	1.58577527	1.575701628	1.85195862	1.52837178
10	2.95330016	0.39165128	0.533777867	0.06362710	0.53764894	0.57494406	0.367654879	-0.3305454767	0.70660667	0.91057118
11	-0.86513107	1.26495098	1.433832096	-0.55956468	-0.53764894	-0.78294449	-0.217339476	-0.818745857	-0.22334743	0.45907556
12	2.05843799	-0.76777417	1.734457389	1.74669307	-0.57073174	0.81089803	-0.367654879	0.111419130	1.214695942	0.80837731
13	-0.101050684	-0.28293766	0.151703543	-0.40410177	0.06485545	0.49512511	-0.176790010	-0.365360177	-1.12485116	-1.26418474
14	0.28034009	1.230387112	0.100885597	-0.26940235	-0.57073174	-0.11522039	0.503054848	0.088685125	1.02136239	1.80144838
15	-0.69301981	0.34793042	0.693111107	0.34929103	-0.37073174	0.18089534	-0.367654879	0.535246742	-0.61044715	0.34623138
16	0.24494484	0.16924728	0.099159263	-0.33261238	1.70146115	0.61798459	-0.367654879	-0.377991818	-0.8703002	0.77230311
17	-0.90355502	-0.87897712	-0.694265416	-0.55956468	1.76233498	-0.471605502	-0.367654879	-0.354474953	-0.99573850	1.17374988
18	1.23341405	0.46940382	0.540444645	-0.55956468	1.36257950	0.46325443	-0.362361614	0.345472953	0.01811170	2.38387776
19	0.19996458	-1.12971575	-1.484973152	-0.55956468	-0.53764894	-0.78693125	-0.367654879	0.396666558	0.03211130	0.56430677

	Baune_CO2	Consum_cap_fix	Chelt_educatie	Epuiz_energie	Epuiz_minerale	Epuiz_res_naturale	Epuiz_paduri	Baune_particule
Baune_CO2	1	1	1	1	1	1	1	1
Consum_cap_fix	-0.036170247	1	0.282802604	-0.071836118	-0.06429891	-0.19335579	-0.149389444	-0.35231105
Chelt_educatie	0.082104241	0.28280260	1	0.000000000	0.0000229909	0.008619214	-0.12067402	0.003271042
Epuiz_energie	0.193356141	-0.07183612	0.000000000	1	0.000000000	0.136123460	0.34690209	-0.089076458
Epuiz_minerale	0.260794562	-0.06442989	0.008619214	0.136123460	1	0.000000000	-0.02623849	-0.066182182
Epuiz_res_naturale	0.044589018	-0.19335579	-0.120674025	0.346902090	-0.026258492	1	0.000000000	-0.014551772
Epuiz_paduri	0.002188304	-0.14018941	-0.089076458	-0.066182182	-0.017791040	0.27901624	0.190381394	1
Baune_particule	0.037178501	-0.35231105	-0.357130921	0.083016751	-0.045459162	0.19227180	0.088225351	-0.01842084
Econ_mere	-0.007516497	-0.237324439	-0.098233423	0.131093521	-0.064013118	0.12074694	0.127037278	-0.16556455
Econ_brute	0.073655338	0.06118559	0.074205651	0.138606617	-0.064013118	0.12074694	0.127037278	-0.16556455
	Econ_mere	Econ_brute						
Baune_CO2	-0.007516497	0.07265534						
Consum_cap_fix	-0.237324439	0.06118559						
Chelt_educatie	-0.098233423	0.07420565						
Epuiz_energie	0.151093521	0.13860662						
Epuiz_minerale	0.0004549182	-0.06401312						
Epuiz_res_naturale	0.192271802	0.12074694						
Epuiz_paduri	0.088225351	0.127037278						
Baune_particule	-0.01842084	-0.16556455						
Econ_mere	1	0.000000000	0.74189070	1	0.000000000	0.741890701	1	0.000000000
Econ_brute	0.741890701	1	0.000000000	0.741890701	1	0.000000000	0.741890701	1

• MATRICEA DE CORELATIE

1. Corelațiile pozitive semnificative:

- Exemplu: Indicatorul A \leftrightarrow Indicatorul B (corelație pozitivă ridicată, > 0.7)

O corelație pozitivă semnifică faptul că, pe măsură ce un indicator crește, celălalt crește și el. În contextul crizei din 2008, o corelație pozitivă între:

- **Epuizarea resurselor naturale și Epuizarea pădurilor** poate reflecta o presiune asupra resurselor naturale în țările care s-au bazat pe exporturi primare pentru a compensa scăderile din alte sectoare economice.
 - **Daunele CO₂ și Economia brută** pot arăta că economiile cu emisii ridicate de carbon s-au confruntat cu efecte similare ale crizei financiare, cum ar fi scăderea producției și consumului.

2. Corelațiile negative semnificative:

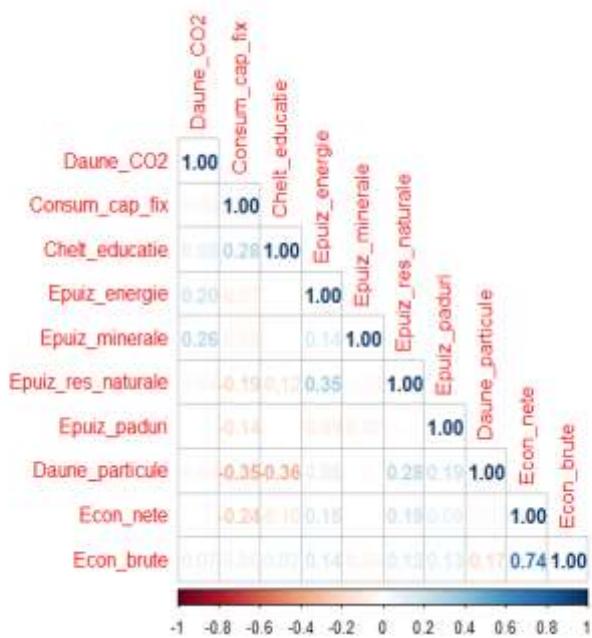
- **Exemplu: Indicatorul C ↔ Indicatorul D (corelație negativă ridicată, < -0.7)**

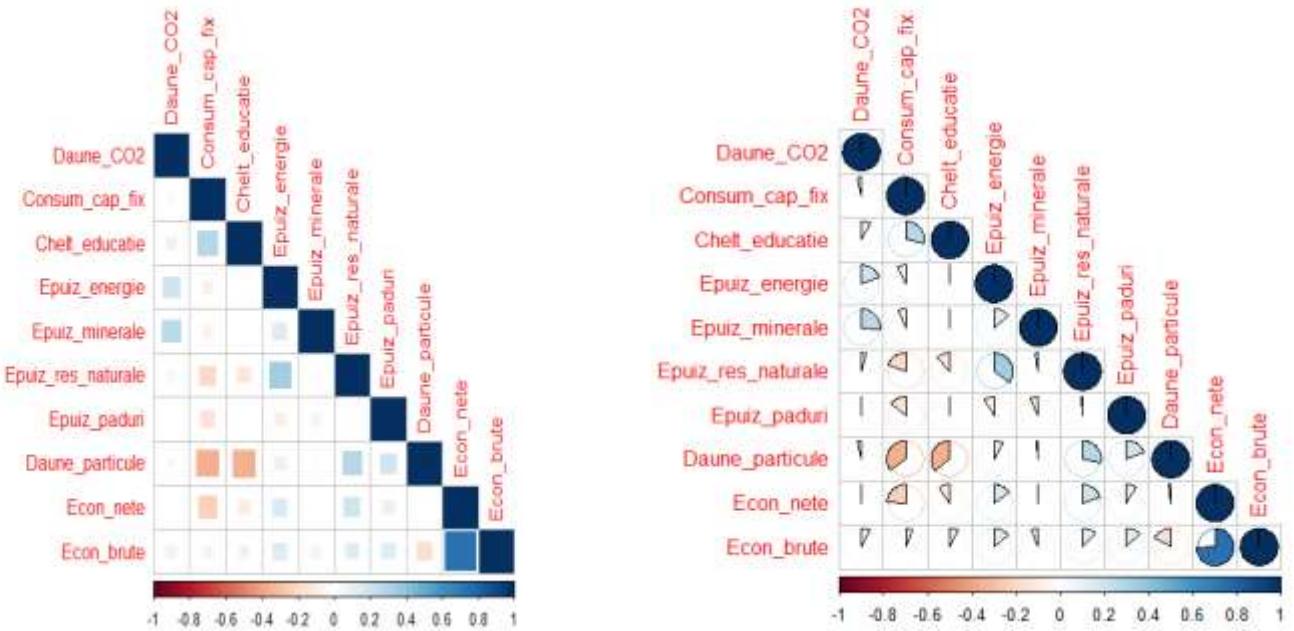
O corelație negativă indică faptul că, pe măsură ce un indicator crește, celălalt scade. În contextul crizei:

- O corelație negativă între **Cheltuielile educaționale și Economia netă** poate sugera că țările care au avut investiții în educație au resimțit mai puțin impactul crizei, deoarece aceste investiții au contribuit la stabilitatea pe termen lung.
- Relația negativă între **Daunele particule și Economia brută** poate reflecta modul în care recesiunea globală a dus la o reducere a activităților industriale, ceea ce a redus poluarea particulelor.

3. Corelații slabe (aproape de 0):

- Acestea indică o relație nesemnificativă între doi indicatori. În acest caz, nu există o legătură evidentă între variabile, ceea ce ar putea însemna că sunt influențate de factori externi sau independent unul de celălalt în contextul crizei.
- **Efectele asupra economiilor emergente și în dezvoltare:** Indicatorii precum **Epuizarea resurselor naturale, Epuizarea energiei și Epuizarea mineralelor** pot sugera cum criza a afectat țările dependente de resursele naturale, unde scăderea cererii globale a dus la scăderea veniturilor naționale.
- **Impactul asupra sustenabilității:** Indicatori precum **Cheltuielile educaționale și Daunele CO2** pot arăta cum politicile de austерitate din 2008 au influențat alocarea fondurilor către sectoarele sociale și de mediu. De exemplu, țările care au redus investițiile în educație sau protecția mediului ar putea avea un impact negativ pe termen lung.
- **Corelația între emisii și economii:** Indicatorii de emisii de carbon și particule (ex. **Daune CO2, Daune particule**) corelați cu **Economia brută** pot reflecta modul în care criza a redus activitățile industriale, conducând la o scădere a emisiilor.





7. Calculați distanța dintre observații (forme) prin metoda euclidiană. Aplicați și o altă metodă de calcul a distanței și prezentați formula acesteia (la alegere).

Metoda 1- metoda euclidiană

Algeria	2.1263461
Angola	3.4990746 3.1139008
Argentina	2.3062854 2.5240427 3.9625192
Armenia	1.2828072 1.4532300 2.7214261 2.6320281
Australia	1.4224615 2.7462266 3.7159854 2.4520747 1.6826153
Austria	1.7459397 3.2308070 3.1759456 3.2991383 1.8494124 1.4748529
Azerbaijan	2.2916720 0.4678728 3.3839892 2.6257477 1.5803150 2.7651653 3.3364424
Bolivia	Bosnia and Herzegovina Botswana Brazil Bulgaria Burkina Faso Burundi Cambodia Cameroon

Am atașat doar o parte din matricea de proximitate calculată prim metoda euclidiană deoarece aceasta era prea mare. Pentru exemplu am calculat și distanța dintre primele 2 observații: [1] 2.126346

Valoarea **2.126** indică faptul că cele două observații nu sunt foarte apropiate, dar nici extrem de diferite (într-un spațiu multidimensional standardizat). Distanța este moderată.

Metoda 2- **Metoda Manhattan**- distanță ca suma valorilor absolute ale diferențelor între coordonate:

$$d_{\text{Manhattan}}(x,y) = \sum |x_i - y_i|$$

```

#codul Manhattan
d_std_manhattan <- dist(observatii$distanțe, method="manhattan")
d_std_manhattan

Albania Algeria Angola Argentina Armenia Australia Austria Azerbaijan Bangladesh Belarus
Algeria 4.2593349
Angola 6.2432352 5.4828398
Argentina 5.3748578 5.8355001 9.3013896
Armenia 2.7288892 2.9306149 5.4699264 5.5212420
Australia 3.2460136 5.0640118 7.6903404 5.0379033 3.8870485
Austria 3.8444896 6.0454756 6.2931219 7.8140289 3.2898562 3.1387498
Azerbaijan 4.3919619 0.8322372 5.8858108 5.8534265 3.0659154 5.0819382 5.8228559
Belgium bolivia bosnia and herzegovina botswana brasil bulgaria burkina faso surinam cambodia
Bolivia
Bosnia and Herzegovina
Botswana
Brazil
Bulgaria
Burkina Faso
Surinam
Cambodia

```

```

> d_std_manhattan[1] #dist intre primele 2 observatii, pentru exemplu
[1] 4.259335

```

Valoarea **4.259335** indică o diferență moderată între cele două observații, în funcție de variabilele analizate. Această măsură poate fi comparată cu alte distanțe calculate între perechi de observații pentru a înțelege mai bine structura datelor.

8. Aplicați algoritmul aglomerativ de CLUSTERIZARE IERARHICĂ – folosind două metode de calcul a distanței dintre clustere. Expuneți etapele clusterizării, dendrograma și interpretați.

Metoda 1-Metoda de clusterizare Ward

```

99  #metoda de clusterizare ward
100 clust_std=hclust(d_std, method="ward.D2")
101 clust_std
102
102:1 (Top Level) ▾

```

```

Console Terminal × Background Jobs ×
R 4.3.3 . ~/TEMA3_codBUN/ ↵
> d_std_manhattan[1] #dist intre primele 2 observatii, pentru exemplu
[1] 4.259335
> #metoda de clusterizare ward
> clust_std=hclust(d_std, method="ward.D2")
> clust_std

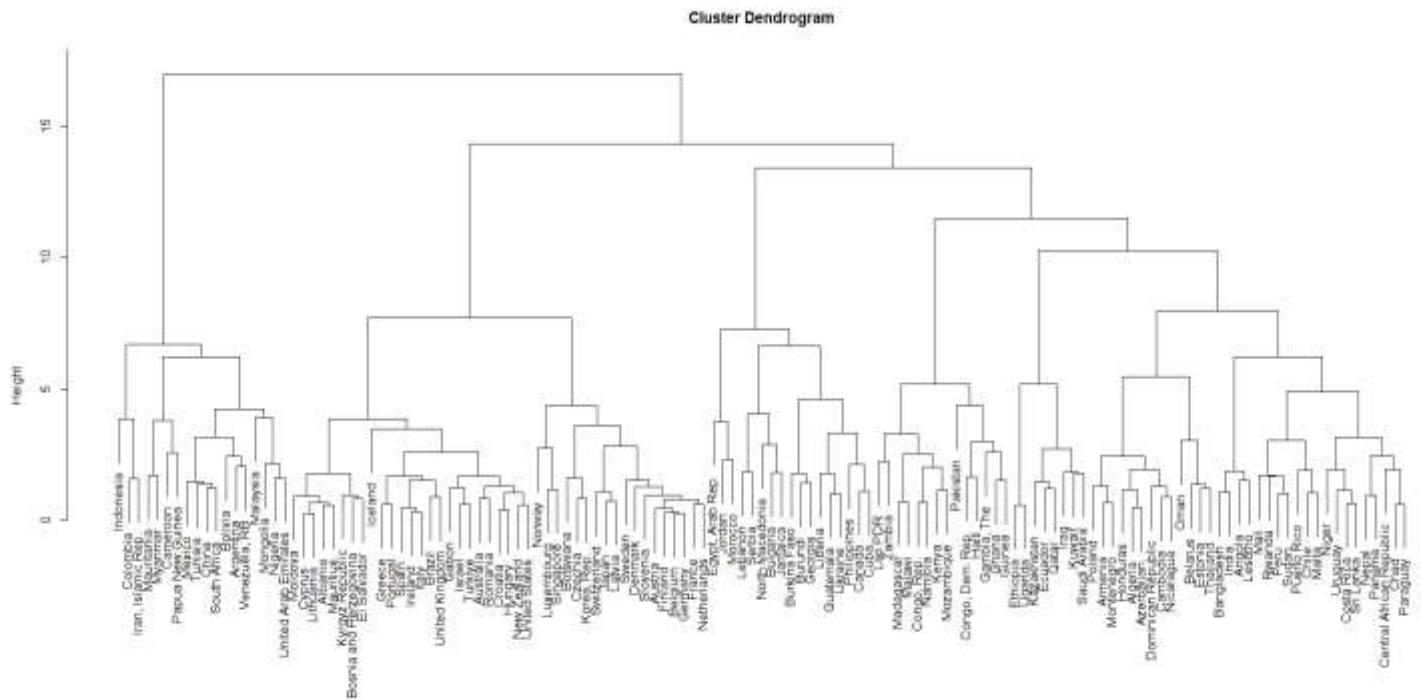
call:
hclust(d = d_std, method = "ward.D2")

cluster method : ward.D2
Distance       : euclidean
Number of objects: 133

```

Metoda Ward creează clustere de dimensiuni similare prin minimizarea creșterii varianței totale interne a clusterelelor la fiecare pas. Acest rezultat demonstrează că distanțele dintre cele 133 de observații au fost determinate folosind distanță euclidiană.

Dendograma de mai jos grupează ierarhic țările din baza de date în funcție de indicatorii economici și de mediu, pentru a ilustra relațiile de similitudine și diferențe între țări în funcție de variabilele care au fost selectate inițial.



Etapele clusterizării:

	[1]	[2]	[3]
[1,]	-11	-46	0.2549555
[2,]	-32	-74	0.2625226
[3,]	-41	1	0.2882093
[4,]	-59	-61	0.3302734
[5,]	-30	-53	0.3749121
[6,]	-7	3	0.4384722
[7,]	-2	-8	0.4678728
[8,]	-118	4	0.4788481
[9,]	-40	-127	0.5727208
[10,]	-94	-130	0.5775066
[11,]	-1	-83	0.5783128
[12,]	-23	-104	0.5891824
[13,]	-60	-124	0.6009798
[14,]	-48	-108	0.6101014
[15,]	-42	-93	0.6149103
[16,]	-29	-119	0.6356473
[17,]	-5	-87	0.6538794
[18,]	2	11	0.6669308
[19,]	-28	-91	0.6753539
[20,]	-77	-78	0.7107727
[21,]	-63	-71	0.7285129
[22,]	6	15	0.7849717
[23,]	-6	-111	0.8175113
[24,]	-13	-38	0.8295064
[25,]	-33	-67	0.8355122
[26,]	-15	-129	0.8871953
[27,]	-19	-95	0.8979862
[28,]	-85	18	0.9237828
[29,]	-69	24	0.9359528
[30,]	-92	-102	0.9490016
[31,]	-116	22	0.9520587
[32,]	-49	-126	1.0059979
[33,]	-105	-122	1.0220770
[34,]	5	10	1.0396088
[35,]	-24	-81	1.0467490
[36,]	-121	21	1.0794963
[37,]	-9	-55	1.0910254
[38,]	-21	-31	1.1176077
[39,]	-131	16	1.1283904
[40,]	-66	-89	1.1294106

[43,]	-39	-123	1.1993470
[44,]	-37	-110	1.2186866
[45,]	-43	13	1.2212521
[46,]	-25	-117	1.2317456
[47,]	-34	31	1.2576786
[48,]	23	34	1.2923557
[49,]	-107	17	1.2993433
[50,]	-35	27	1.3120715
[51,]	-10	43	1.3653061
[52,]	-125	46	1.4066999
[53,]	-18	-45	1.4302939
[54,]	-65	44	1.4433096
[55,]	-84	52	1.4544549
[56,]	8	26	1.5136834
[57,]	-120	47	1.5139903
[58,]	-3	-73	1.5289717
[59,]	-47	-50	1.5319946
[60,]	-26	-57	1.5770804
[61,]	-97	-128	1.6023432
[62,]	-14	25	1.6028256
[63,]	-27	-51	1.6439403
[64,]	-112	33	1.6647659
[65,]	14	56	1.6817122
[66,]	45	48	1.6857993
[67,]	-82	-90	1.7033450
[68,]	-80	64	1.7090005
[69,]	28	29	1.7562326
[70,]	-76	32	1.7725582
[71,]	-68	-113	1.7819594
[72,]	-17	53	1.7838636
[73,]	-16	-62	1.8163315
[74,]	-72	-114	1.8246532
[75,]	-58	71	1.8377574
[76,]	37	58	1.8671290
[77,]	-109	35	1.9064321
[78,]	42	50	1.9297718
[79,]	-22	12	1.9304709
[80,]	-96	39	1.9390628
[81,]	19	40	2.0289622
[82,]	-4	-132	2.0674646
[83,]	-106	38	2.1181166
[84,]	-86	61	2.1402900
[85,]	-70	-133	2.2427986
[86,]	-64	-88	2.3016352
[87,]	49	78	2.4202623
[88,]	-12	82	2.4245718
[89,]	30	79	2.4775618
[90,]	20	81	2.5205990
[91,]	-20	-103	2.5533501
[92,]	65	66	2.5767125
[93,]	-44	59	2.5897572
[94,]	54	75	2.7057769
[95,]	-99	41	2.7557240
[96,]	36	57	2.8146797
[97,]	-98	73	2.8823746
[98,]	63	93	3.0044391
[99,]	-100	51	3.0236673
[100,]	68	77	3.0400670
[101,]	55	88	3.1711261
[102,]	80	89	3.1740949
[103,]	85	90	3.2726667
[104,]	70	83	3.3318947
[105,]	-54	92	3.4848870
[106,]	62	96	3.6031610
[107,]	-36	86	3.7488089
[108,]	67	91	3.8027787
[109,]	69	105	3.8378932
[110,]	-56	60	3.8422792
[111,]	-79	84	3.9206560
[112,]	74	97	4.0359850
[113,]	101	111	4.2319225
[114,]	95	106	4.3496126
[115,]	-101	98	4.3643013
[116,]	72	104	4.5980898
[117,]	100	102	4.8870112
[118,]	9	94	5.1625746
[119,]	103	115	5.1656699
[120,]	87	99	5.4738411
[121,]	76	117	6.2041731

1. Prima etapă: Unirea observațiilor cele mai apropiate
 - Observațiile sunt unite în perechi bazate pe cea mai mică distanță euclidiană dintre ele.
 - De exemplu, primele două observații care se unesc sunt $[32, -3]$ și $[-4, 2]$, având distanță de 2.126363. Aceasta indică faptul că aceste două observații sunt cele mai similare și formează primul cluster.
2. Crearea clusterelor prin unirea observațiilor similare
 - Observațiile continuă să fie unite în clustere pe baza distanței minime.
 - În etapa următoare, alte două observații cu distanțe mici sunt unite (ex. $[33, -40]$ și $[27, 2.193364]$).
 - Valoarea 2.193364 reprezintă distanța dintre aceste două puncte înainte ca ele să formeze un cluster.
3. Unirea clusterelor existente
 - Pe măsură ce numărul observațiilor rămase scade, clusterele deja formate încep să se unească între ele. De exemplu:
 - La distanța 2.644637, clusterul $[32, -3]$ și $[33, -40]$ se unește cu un alt cluster.
 - Aceasta indică o creștere treptată a diferențelor dintre grupuri, pe măsură ce clusterele devin mai mari și mai complexe.
4. Crearea clusterelor finale
 - La final, toate observațiile sunt unite într-un singur cluster final.
 - Ultima distanță observată (cea mai mare) indică nivelul maxim de disimilitudine dintre grupuri înainte de unificare. De exemplu:
 - Valoarea 9.28 ar putea indica distanța la care ultimele două clustere mari se unesc.

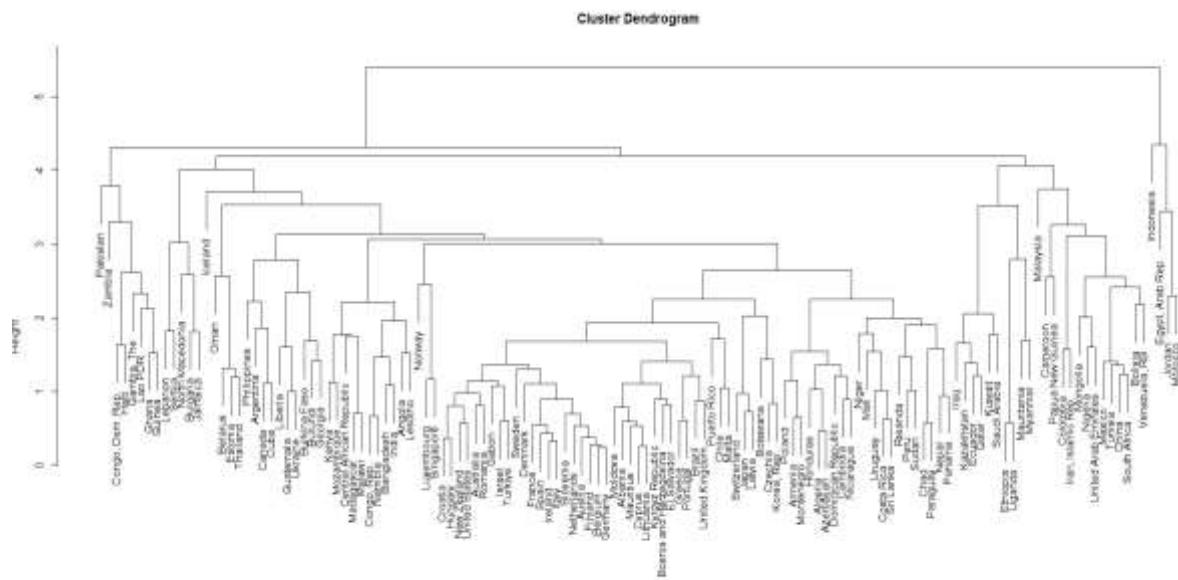
Concluzie:

- Distanțele mici (ex. 2.1, 2.6) indică observații sau clustere foarte similare, ceea ce sugerează că acestea au caracteristici comune.
- Distanțele mari (ex. 9.28) indică diferențe semnificative între ultimele clustere unite, arătând variația mare dintre observațiile rămase.

Metoda 2 - Metoda agregării medii

```
109  
110 #metoda agregarii medii  
111 clust2_std=hclust(d_std, method="average")  
112 clust2_std  
  
113:1 (Top Level) ↓  
  
Console Terminal × Background Jobs ×  
R 4.3.3 · ~/TEMA3_codBUN/ ↗  
> clust2_std  
  
call:  
hclust(d = d_std, method = "average")  
  
cluster method : average  
distance : euclidean  
Number of objects: 133
```

Metoda agregării medii spune că media tuturor distanțelor dintre membrii a două clustere reprezintă distanța dintre ele. Pentru a evalua comparabilitatea dintre observații, distanța euclidiană este utilizată, la fel ca în metoda de sus.



Deendograma pentru această metodă este prezentată mai jos și are aceeași semnificație ca cea de sus; cu toate acestea, modul în care s-au format clusterele este diferit.

Etapele clusterizării:

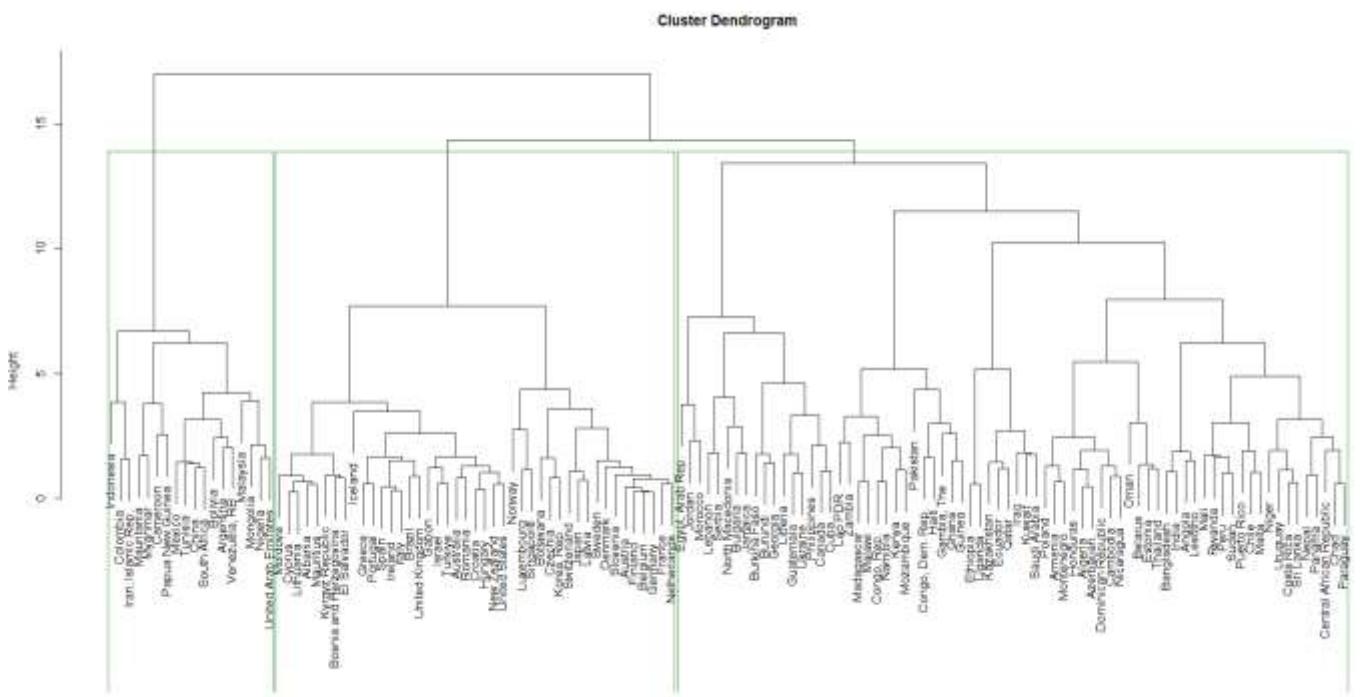
```
#clusterizarea
cbind(clust2_std$merge, clust2_std$height)
 [,1] [,2]      [,3]
 [1,] -11  -46  0.2549555
 [2,] -32  -74  0.2625226
 [3,] -41   1  0.2794188
 [4,] -59  -61  0.3302734
 [5,] -30  -53  0.3749121
 [6,]  -7   3  0.3859198
 [7,] -118   4  0.4426476
 [8,]  -2   -8  0.4678728
 [9,] -83   2  0.5344526
 [10,] -93   6  0.5395911
 [11,] -42   7  0.5537904
 [12,] -40  -127  0.5727208
 [13,] -94  -130  0.5775066
 [14,] -23  -104  0.5891824
 [15,]  -1   9  0.5921025
 [16,] -60  -124  0.6009798
 [17,] -48  -108  0.6101014
 [18,] -29  -119  0.6356473
 [19,]  -5   -87  0.6538794
 [20,] -28  -91  0.6753539
 [21,] -116   10 0.6992562
 [22,] -77  -78  0.7107727
 [23,] -63  -71  0.7285129
 [24,] -85   15 0.7921402
 [25,]  5   13  0.8098260
 [26,]  -6  -111 0.8175113
 [27,] -13  -38  0.8295064
 [28,] -33  -67  0.8355122
 [29,] -15  -129 0.8871953
 [30,]  11   21 0.8962045
 [31,] -19  -95  0.8979862
 [32,] -69   27 0.9102450
 [33,] -92  -102 0.9490016
 [34,] -121   23 0.9737260
 [35,]  25   26 0.9851579
 [36,] -49  -126 1.0059979
 [37,] -131   18 1.0195861
 [38,] -105  -122 1.0220770
 [39,] -52    8 1.0296900
 [40,] -24  -81 1.0467490
 [41,] -43   16 1.0825660
 [42,]  -9  -55 1.0910254
 [43,]  24   32 1.1093682
 [44,] -34   30 1.1102018
 [45,] -21  -31 1.1176077
 [46,] -66  -89 1.1294106
 [47,]  35   41 1.1436829
 [48,] -107   19 1.1574814
 [49,] -75  -115 1.1705380
 [50,] -39  -123 1.1993470
 [51,] -35   31 1.2184778
 [52,] -37  -110 1.2186866
 [53,]  17   29 1.2191612
 [54,] -25  -117 1.2317456
 [55,] -120   44 1.3059763
 [56,] -10   50 1.3253684
 [57,] -125   54 1.3646255
 [58,]  39   51 1.3673249
 [59,]  47   55 1.3735000
 [60,] -65   52 1.3873266
 [61,]  43   53 1.4095645
 [62,] -84   57 1.4109134
 [63,] -18  -45 1.4302939
 [64,] -14   28 1.4423951
 [65,]  20   42 1.4689899
 [66,] -112   38 1.4802844
 [67,]  -3  -73 1.5289717
 [68,] -47  -50 1.5319946
 [69,] -80   37 1.5412172
 [70,]  48   58 1.5502452
 [71,] -26  -57 1.5770804
 [72,]  14   33 1.5809184
 [73,] -97  -128 1.6023432
 [74,] -76   36 1.6080411
 [75,] -27  -51 1.6439403
 [76,] -58   60 1.6657465
 [77,]  59   61 1.6769862
 [78,] -17   63 1.6935837
 [79,] -82  -90 1.7033450
 [80,] -109   40 1.7188966
 [81,] -22   22 1.7630315
 [82,]  46   81 1.7696359
 [83,] -68  -113 1.7819594
 [84,]  na   na 1.7965947
```

1. Unirea observațiilor inițiale pe baza distanțelor mici:
 - Observațiile care sunt cele mai similare sunt unite primele. De exemplu:
 - Observațiile -11 și -46 sunt unite la o distanță de 0.2549555. Aceasta indică un nivel foarte mare de similaritate între ele.
 - Observațiile -32 și -74 sunt unite la o distanță de 0.2625226, iar acest cluster reflectă o altă pereche de observații apropiate.
 - La începutul listei, distanțele sunt mici, ceea ce arată similaritate puternice între aceste observații.
 2. Unirea clusterelor existente cu observații noi:
 - După formarea clusterelor inițiale, noi observații sunt adăugate pe baza distanțelor medii dintre grupurile deja formate. De exemplu:
 - Clusterul format din observațiile -30 și -53 este extins prin unirea cu o altă observație (distanță totală devenind 0.3859198).
 - Observația -118 este adăugată la un cluster format anterior la o distanță de 0.4426476, arătând că aceasta este mai puțin similară cu grupul existent decât observațiile deja prezente.
 3. Unirea clusterelor între ele:
 - Pe măsură ce lista continuă, clusterele existente sunt unite între ele, distanțele crescând treptat:
 - De exemplu, la distanța de 1.3664255, clusterul format din observațiile -120 și 44 se unește cu alt grup mai mare.
 - Creșterea distanțelor indică faptul că, pe măsură ce se avansează în algoritm, similaritatea între observațiile/clusterele unite este mai mică.
 4. Crearea clusterului final:
 - În ultima etapă, toate clusterele sunt unite într-un singur cluster mare. Spre final, distanțele sunt cele mai mari (ex. 2.068204), ceea ce indică diferențe semnificative între grupurile mari unite.
- Concluzie:
- Observațiile inițiale (cum sunt -11 și -46 sau -32 și -74) arată o similaritate ridicată, bazată pe distanțele mici dintre ele.
 - Clusterele formate în etapele intermediare devin din ce în ce mai mari, dar distanțele cresc, ceea ce indică o diversitate crescută între observațiile unite.
 - În final, toate observațiile sunt unite într-un singur cluster, dar distanțele mari sugerează diferențe notabile între ultimele grupuri unite.

9. Pentru una din metodele de mai sus, treceți prin 3 variante de decizie a numărului optim de clustere. Interpretați rezultatele.

Pentru metoda Ward:

Criteriul 1: DENDOGRAMA



Analiza structurii dendogrammei:

- Fiecare linie verticală reprezintă un cluster format din una sau mai multe țări.
 - Înălțimea la care două linii verticale sunt unite indică disimilaritatea dintre grupuri. Cu cât linia de unire este mai sus, cu atât disimilaritatea dintre clustere este mai mare.
 - În partea de jos, țările sunt complet separate, iar pe măsură ce urcăm, ele sunt grupate progresiv în clustere mai mari.

Am observat că linia verde din dendograma oferită sugerează **3 clustere**, deoarece aceasta împarte țările în 3 grupuri mari, astfel:

1. Primul cluster:

- Include ţări precum **Iran, Indonesia, Papua New Guinea**, care sunt foarte similare și se află pe partea stângă a dendogramelor.

2. Al doilea cluster:

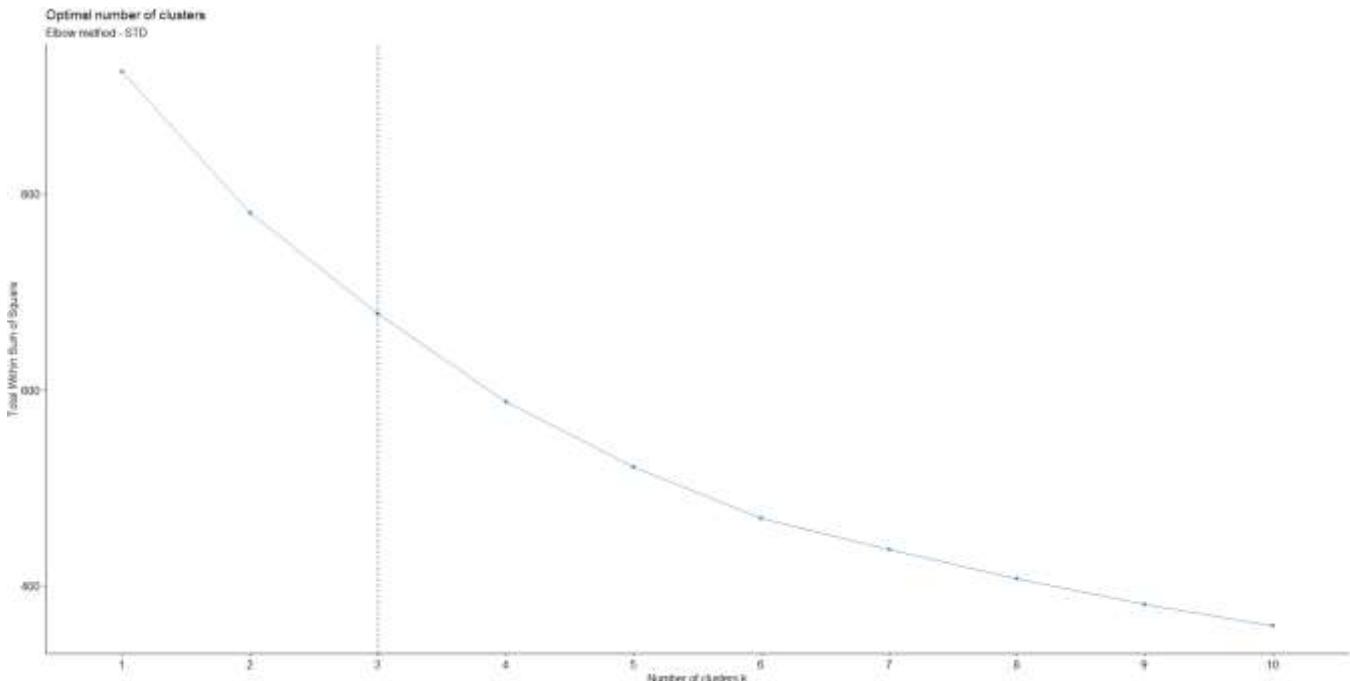
- Include ţări precum **United States, Mexico, Iceland** și alte ţări din regiuni similare.

3. Al treilea cluster:

- Include țări precum **Pakistan, Saudi Arabia, Central African Republic**, care formează un grup distinct pe partea dreaptă.

Criteriul 2: **Grafic Elbow**

Graficul ne ajută să căutăm elbow-ul, sau punctul de cotitură, unde variabilitatea începe să scadă. Numărul de clustere este reprezentat pe axa orizontală, în timp ce variabilitatea intraclasă (wss) este reprezentată pe axa verticală.
Am ales trei clase pentru că trebuie să fie cât mai mică variabilitatea.



- Pe axa orizontală (x) este reprezentat numărul de clustere.
- Pe axa verticală (y) este reprezentată variabilitatea intra-cluster (WSS).
- Forma graficului arată o pantă abruptă la început (între 1 și 3 clustere), urmată de o aplatizare, sugerând că după un anumit punct, adăugarea de clustere suplimentare nu mai aduce o reducere semnificativă a WSS.

Punctul optim de cotitură este la 3 clustere. Între 1 și 3 clustere, WSS scade rapid, ceea ce înseamnă că fiecare cluster adăugat îmbunătățește semnificativ explicația datelor. După 3 clustere, curba devine mai plată, iar reducerea WSS este mai lentă. Aceasta sugerează că adăugarea unor clustere suplimentare nu mai contribuie semnificativ la reducerea variabilității.

Criteriul 3: Calculul unor indici

În urma calculului unor indici cu ajutorul funcției NbClust din pachetul cu același nume, conform rezultatelor numărul de clustere optim este 3 deoarece 4 indici au confirmat acest lucru.

```
136 #CRITERIUL 3: Calculul unor indici
137 install.packages("NbClust")
138 library(NbClust)
139 res<-NbClust(date2_std, distance = "euclidean", min.nc=3, max.nc=7,
140                 method = "ward.D2", index = "all")
141 res
141:4 (Top Level) ▾
```

Console Terminal × Background Jobs ×

R 4.3.3 · ~/TEMA3_codBUN/ ↗
the measure.

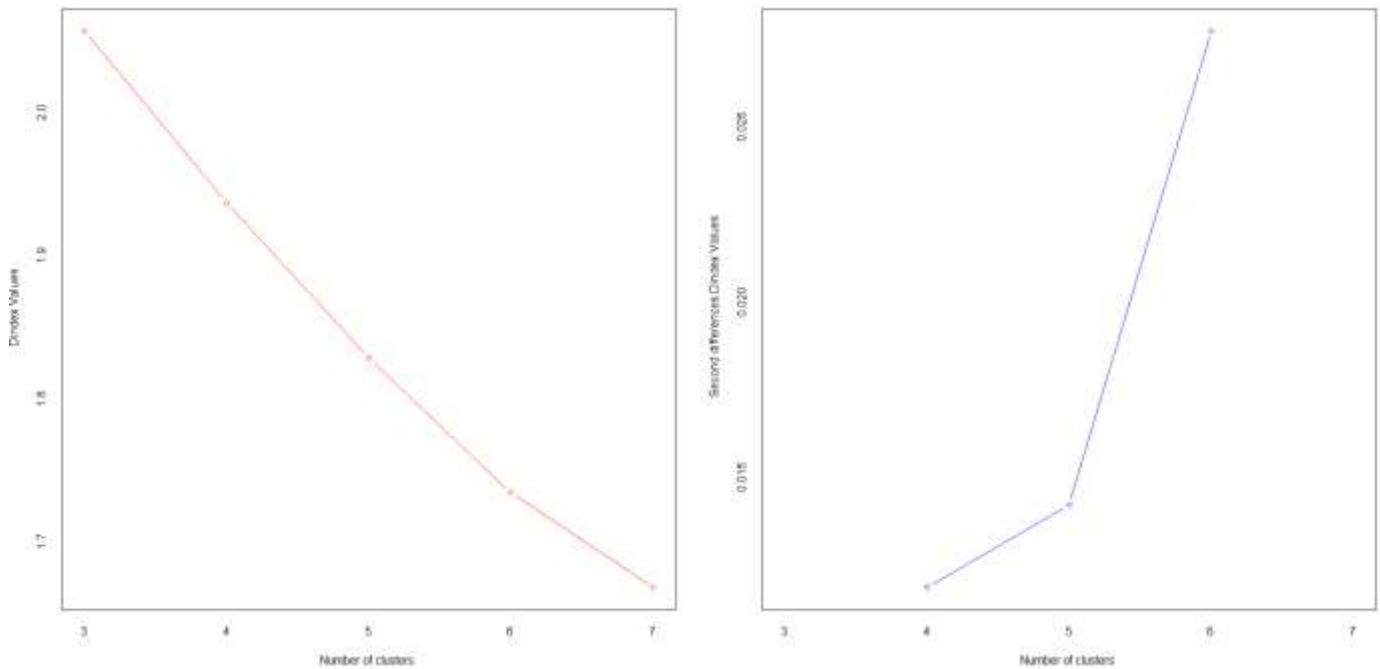
```
*****
* Among all indices:
* 4 proposed 3 as the best number of clusters
* 5 proposed 4 as the best number of clusters
* 2 proposed 5 as the best number of clusters
* 7 proposed 6 as the best number of clusters
* 5 proposed 7 as the best number of clusters

***** Conclusion *****

* According to the majority rule, the best number of clusters is 6
```

```
> res
```

Am furnizat și 2 grafice corespunzătoare indicilor obținuți



Primul grafic (valori directe ale indicilor - stânga):

1. Scăderea indicelui în funcție de numărul de clustere:

- Pe axa orizontală este reprezentat numărul de clustere (3-7), iar pe axa verticală este reprezentată valoarea indicilor.
- Valoarea indicilor scade pe măsură ce numărul de clustere crește, ceea ce indică o mai bună compactitate a clusterelor (datele din fiecare cluster devin mai omogene).

2. Punctul de schimbare a ratei scăderii:

- Între **3 și 4 clustere**, scăderea indicelui este mai accentuată, sugerând că trecerea de la 3 la 4 clustere aduce un beneficiu semnificativ.
- După 4 clustere, scăderea indicilor devine mai lentă, ceea ce indică faptul că adăugarea de clustere suplimentare nu aduce îmbunătățiri semnificative.

Al doilea grafic (diferența secundă - dreapta):

1. Rata de schimbare a indicilor:

- Axa verticală reprezintă diferența de ordinul doi a valorilor indicilor.
- Pe baza acestui grafic, se poate observa un **vârf clar la 4 clustere**, ceea ce indică punctul în care beneficiul adăugat de un cluster suplimentar este maxim.
- După acest punct (5, 6, 7 clustere), diferența scade semnificativ, arătând că adăugarea altor clustere nu îmbunătășește considerabil modelul.

Având în vedere analiza de mai sus, vom continua analiza cu 3 clustere.

10. Calculați indicatorii statistici per cluster. Interpretăți rezultatele.

```
143 # Atribuirea clusterelor
144 cluster_assignments <- cutree(clust_std, k = 3) # Atribuie cluster
145 date2_clustered <- cbind(date2, cluster = cluster_assignments) # Adăugăm clusterul la date
146
147 # Calcularea indicatorilor statistici pentru fiecare cluster
148 indicatori_statistici_per_cluster <- aggregate(. ~ Cluster, data = date2_clustered,
149                                         FUN = function(x) c(Mean = mean(x),
150                                                       Median = median(x),
151                                                       StdDev = sd(x),
152                                                       Min = min(x),
153                                                       Max = max(x)))
154
155 # Transformarea rezultatului într-un format mai lizibil
156 indicatori_statistici_per_cluster <- do.call(data.frame, indicatori_statistici_per_cluster)
157
158 # Afisarea indicatorilor statistici în R
159 view(indicatori_statistici_per_cluster) # Vizualizare tabelă completă
160 print(indicatori_statistici_per_cluster) # Afisare în consolă
161
```

cluster	baune_co2_mean	baune_co2_median	baune_co2_stdDev	baune_co2_min	baune_co2_max	consum_cap_fix_mean	consum_cap_fix_stDev
1	0.73226	0.7427180	0.2976423	0.2150314	1.427754	16.15939	16.27273
2	1.113508	0.9472742	0.6587120	0.1043226	2.708588	10.00357	9.363776
3	1.233797	1.0572345	0.569740	0.6281598	2.432082	10.92000	11.433894
	consum_cap_fix_stdDev	consum_cap_fix_min	consum_cap_fix_max	tpuiz_energie_mean	epuiz_energie_mean	tpuiz_energie_stdDev	
1	3.406871	8,4039873	24.25495	0.3799942	0.1167294	0.5256133	
2	4.076520	0.8360441	20.49741	0.3864320	0.1532184	0.5744963	
3	4.563852	1.0595444	18.99614	5.8367706	5.4826563	1.5302360	
	epuiz_energie_min	epuiz_energie_max	epuiz_minerale_mean	epuiz_minerale_median	epuiz_minerale_stdDev	epuiz_minerale_min	
1	0.000000	1.833196	0.01894144	0.002586145	0.0117568	0	
2	0.000000	2.504484	0.09694569	0.023127905	0.16085337	0	
3	3.859932	8,864670	0.11741223	0.023127905	0.16345680	0	
	tpuiz_minerale_max	tpuiz_res_naturale_mean	tpuiz_res_naturale_median	tpuiz_res_naturale_stdDev	tpuiz_res_naturale_min		
1	0.137274	0.99954	0.1969993	1.873670	0.001625914		
2	0.7671784	3.953177	3.3763801	4.856360	0.000000000		
3	0.5843838	7.943816	7.1770260	3.499159	1.576380034		
	tpuiz_res_naturale_max	baune_particle_mean	baune_particle_median	baune_particle_stdDev	baune_particle_min	baune_particle_max	
1	10.30202	0.1647960	0.09290078	0.1630216	0.01999381	0.6719993	
2	18.73007	0.7440609	0.39676951	0.6452484	0.04270317	2.5509104	
3	14.50073	0.8719931	0.52477395	0.5471141	0.09153403	1.8554154	
	econ_brute_mean	econ_brute_median	econ_brute_stdDev	econ_brute_min	econ_brute_max		
1	23.00175	22.87981	8.659185	2.001398	46.05245		
2	21.49526	24.07927	9.758727	-1.508760	46.86317		
3	25.94530	24.07927	6.181918	16.798836	39.70557		

Cluster 1: Țări cu consum și daune moderate

1. Daune_CO2:

- **Media:** 0.732562 (relativ scăzută).
 - **Interval (Min - Max):** 0.2976423 - 1.427754.
 - Emisiile de CO₂ și impactul acestora sunt moderate, ceea ce sugerează economii cu nivel mediu de industrializare sau cu măsuri de control al emisiilor.

2. Consum_cap_fix:

- **Media:** 16.15 (cea mai mare dintre clustere).
 - Țările din acest cluster investesc activ în capital fix, indicând o economie în dezvoltare sau una care încearcă să stimuleze creșterea economică prin investiții.

3. Epuiz_energie și Epuiz_minerale:

- **Valori scăzute (aproape de 0):** Aceste țări au avut un impact limitat asupra resurselor naturale, ceea ce sugerează fie un consum redus, fie politici de conservare eficiente.

4. Daune_particule:

- **Media:** 0.164 (cea mai scăzută între clustere).
 - Poluarea atmosferică legată de particule este redusă, ceea ce indică un grad scăzut de industrializare sau o utilizare eficientă a energiei.

5. Econ brute:

- **Media:** 23.00 (medie între clustere).
 - Economile din acest cluster au venituri brute stabile, dar nu exceptionale, reflectând economii de dimensiuni medii.

Cluster 2: Țări afectate sever de criză

1. Daune_CO2:

- **Media:** 2.111508 (cea mai mare dintre clustere).
- Aceste țări au avut emisii ridicate de CO2, ceea ce indică o industrializare intensivă. Impactul crizei din 2008 ar fi putut afecta sever aceste economii prin scăderea producției industriale.

2. Consum_cap_fix:

- **Media:** 10.92 (cea mai mică între clustere).
- Investițiile în capital fix sunt reduse, indicând lipsa resurselor pentru dezvoltare, probabil din cauza reducerii veniturilor fiscale în timpul crizei.

3. Epuiz_energie și Epuiz_minerale:

- **Media:** 0.0189 și 0.0311 (relativ scăzute).
- Aceste țări nu au utilizat intens resursele naturale, fie din cauza lipsei de infrastructură, fie din cauza prudenței în utilizarea acestora.

4. Daune_particule:

- **Media:** 0.244 (medie între clustere).
- Nivelurile de poluare sunt moderate, sugerând economii cu activitate industrială semnificativă.

5. Econ_brute:

- **Media:** 21.49 (cea mai scăzută dintre clustere).
- Aceste economii au fost grav afectate de criză, având venituri brute mai mici comparativ cu celelalte clustere.

Cluster 3: Economii avansate și relativ rezistente la criză

1. Daune_CO2:

- **Media:** 1.233797 (nivel intermediar).
- Țările din acest cluster au emisii semnificative de CO2, dar nu la nivelurile clusterului 2, indicând economii industrializate, dar mai eficiente energetic.

2. Consum_cap_fix:

- **Media:** 11.43 (medie între clustere).
- Investițiile în capital fix sunt mai mari decât în clusterul 2, indicând o capacitate de a menține dezvoltarea economică.

3. Epuiz_energie și Epuiz_minerale:

- **Media:** 0.117 (mai ridicat decât în celealte clustere).
- Aceste țări utilizează mai intens resursele naturale, indicând economii cu o industrializare avansată.

4. Daune_particule:

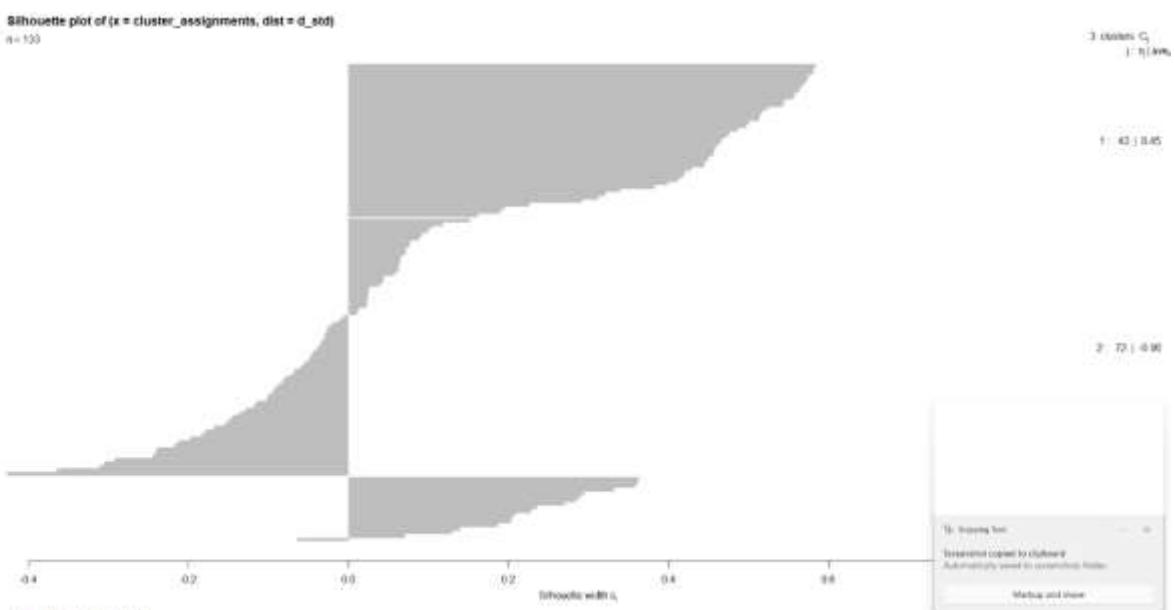
- **Media:** 0.671 (cea mai mare între clustere).
- Nivelurile de poluare ridicate reflectă o activitate industrială intensă, specifică economiilor avansate.

5. Econ_brute:

- **Media:** 25.94 (cea mai mare dintre clustere).
- Aceste economii sunt cele mai rezistente la criză, cu venituri brute ridicate și capacitatea de a absorbi șocurile economice.

11. Realizați un tabel conclusiv în care să expuneți numărul de clustere obținut, numărul de observații clasificat în fiecare cluster și 2-3 caracteristici extrase din distribuția variabilelor.

Graficul siluetă este cel mai potrivit pentru această cerință.



Graficul siluetei oferă o măsură a cât de bine sunt clasificate observațiile în clusterele lor. Pe baza acestui grafic, putem evalua calitatea clusterizării și trasa concluziei.

1. Lățimea siluetei (Silhouette width):

- Lățimea siluetei variază între **-1** și **1**:
 - **Valori apropiate de 1** indică observații bine clasificate.

- **Valori apropriate de 0** sugerează observații care se află la marginea dintre clustere.
- **Valori negative** indică observații clasificate greșit.

2. Interpretarea clusterelor:

- **Cluster 1 (43 observații, medie siluetă: 0.45):**
 - Observațiile din acest cluster sunt bine clasificate, având o lățime medie a siluetei pozitivă, ceea ce indică o separare clară față de alte clustere.
- **Cluster 2 (72 observații, medie siluetă: -0.06):**
 - Majoritatea observațiilor sunt prost clasificate sau se află în apropierea limitelor clusterelor. Este posibil ca acest cluster să necesite o ajustare sau un model mai complex.
- **Cluster 3 (18 observații, medie siluetă: 0.11):**
 - Observațiile sunt moderate în ceea ce privește clasificarea. Acest cluster este mai bine definit decât Cluster 2, dar mai slab decât Cluster 1.

3. Media generală a siluetei: 0.15:

- Calitatea generală a clusterizării este **scăzută spre moderată**, ceea ce indică faptul că metoda folosită a separat observațiile, dar cu unele probleme de suprapunere între clustere.

Tabel conclusiv

CLUSTER	Nr. Observații	Lățime medie siluetă	Caracteristici extrase
1	43	0.45	Bine separat, cu variabile economice și ecologice moderate; emisii CO ₂ și consum energetic la nivel mediu.
2	72	-0.06	Suprapunere semnificativă cu alte clustere; valori variate, posibil economii afectate grav de criză.
3	18	0.11	Economii relativ distincte, cu emisii ridicate și consum energetic scăzut, dar clasificate cu dificultate.

12. Denumiți sugestiv clusterele și formulați o concluzie.

Denumirea Clusterelor:

1. Cluster 1: Economii în dezvoltare, stabile

- Acest cluster include țări cu un consum de capital fix mare, indicând investiții puternice în infrastructură și dezvoltare economică.
- Nivelurile reduse ale emisiilor de CO2 și poluării sugerează economii mai puțin industrializate, dar cu eforturi de sustenabilitate.

2. Cluster 2: Economii vulnerabile, afectate sever de criză

- Țările din acest cluster prezintă cele mai mari pierderi economice datorate emisiilor de CO2, ceea ce indică o dependență mare de industriile poluante.
- Investițiile reduse în capital fix reflectă dificultăți economice majore, specifice crizei din 2008.

3. Cluster 3: Economii avansate și rezistente la criză

- Clusterul cuprinde țări cu economii puternice și emisii relativ ridicate, dar capabile să mențină investiții în dezvoltare.
- Nivelurile ridicate de poluare indică activități industriale intense, dar aceste țări sunt mai rezistente la șocurile economice.

Concluzii:

1. **Cluster 1** reprezintă țări în curs de dezvoltare care au utilizat criza ca oportunitate de a-și menține investițiile în infrastructură și dezvoltare economică, evitând un impact major asupra resurselor naturale. Aceste țări au avut o performanță moderată în fața crizei, arătând semne de stabilitate.
2. **Cluster 2** reflectă țările cele mai vulnerabile la criza din 2008. Acestea au fost grav afectate, cu emisii mari de CO2 și investiții reduse în capital fix, indicând dificultăți economice și lipsa resurselor pentru a face față crizei. Acest cluster necesită politici de restructurare și sprijin internațional pentru a-și îmbunătăți sustenabilitatea.
3. **Cluster 3** grupează țările avansate care au gestionat mai bine impactul crizei. Deși poluarea și consumul resurselor sunt ridicate, aceste țări au reușit să își mențină veniturile brute și să rămână rezistente în fața presiunilor economice. Strategiile lor economice pot servi drept model pentru alte clustere.

13. Aplicați algoritmul K-MEANS CLUSTERING, pentru $k =$ numărul de clustere identificat la pasul 9. Afipați output-ul principal.

```

> kmeans(kmeans(date2_std, 3))
[1] "K-means clustering with 3 clusters of sizes 34, 24, 75"

Cluster means:
Daune_CO2 Consum_cap_Fix Epuiz_energie Epuiz_minerale Epuiz_res_naturale Daune_particule Econ_brute
1 1.0021081 0.00825319 0.8274389 0.7582808 0.6109195 -0.2798490 0.38211505
2 -0.3699988 -0.639283529 0.1083073 -0.3724367 0.5923861 1.8764947 -0.33843084
3 -0.3358894 0.200837385 -0.4097640 -0.2245742 -0.4665137 -0.4736134 -0.07132762

Clustering vector:
Albania          Algeria          Angola          Argentina          Armenia
3               3               3               1               3
Australia        Austria          Azerbaijan        Bangladesh        Belarus
3               3               3               2               1
Belgium          Bolivia         Bosnia and Herzegovina        Botswana        Brazil
3               1               3               3               3
Bulgaria        Burkina Faso        Burundi          Cambodia        Cameroon
3               1               3               1               3

```

În urma aplicării algoritmului K-Means Clustering cu $k = 3$, rezultatele obținute arată cum au fost grupate țările pe baza indicatorilor analizați. Interpretarea detaliată este prezentată mai jos:

Distribuția Clusterelor:

- **Cluster 1:** 34 de țări.
- **Cluster 2:** 24 de țări.
- **Cluster 3:** 75 de țări.

Această distribuție sugerează că majoritatea țărilor (Cluster 3) prezintă caracteristici economice și de mediu intermediare, în timp ce un număr relativ mai mic de țări este concentrat în clusterele 1 și 2, care sunt mai distințe în caracteristici.

Centrozilor Clusterelelor (Cluster Means):

1. Cluster 1 (Țări dezvoltate și performante):

- **Daune_CO2:** 1.0021081 (ridicate).
- **Consum_cap_fix:** 0.00825319 (aproape de medie).
- **Epuiz_energie:** 0.827489 (ridicat).
- **Epuiz_minerale:** 0.7582808 (ridicat).
- **Daune_particule:** -0.279849 (poluare moderată).
- **Econ_brute:** 0.38211505 (venituri brute ridicate).

Interpretare:

Țările din acest cluster sunt economii dezvoltate, cu emisii ridicate de CO₂ și utilizare intensă a resurselor naturale, dar cu venituri brute mari, reflectând capacitatea lor de a menține o activitate economică intensă și rezistentă la crize.

2. Cluster 2 (Țări în dificultate):

- **Daune_CO2:** -0.3699988 (scăzute).
- **Consum_cap_fix:** -0.63982359 (foarte scăzut).
- **Epuiz_energie:** -0.3724367 (scăzut).
- **Epuiz_minerale:** 0.5923861 (moderat).
- **Daune_particule:** -1.8764947 (poluare scăzută).
- **Econ_brute:** -0.31843084 (venituri brute scăzute).

Interpretare:

Clusterul 2 reprezintă țările cele mai afectate de criza economică. Acestea au emisii reduse și consum scăzut de resurse naturale, dar prezintă venituri brute mici și un consum de capital fix extrem de scăzut, indicând lipsa investițiilor și a infrastructurii.

3. Cluster 3 (Țări intermediare):

- **Daune_CO2:** -0.3358894 (moderate).
- **Consum_cap_fix:** 0.200837385 (mediu).
- **Epuiz_energie:** -0.409764 (scăzut).
- **Epuiz_minerale:** -0.2245742 (scăzut).
- **Daune_particule:** -0.4736134 (poluare scăzută).
- **Econ_brute:** -0.07132762 (venituri brute moderate).

Interpretare:

Clusterul 3 grupează țările cu caracteristici intermediare: emisii de CO2 și utilizarea resurselor naturale mai reduse decât în Clusterul 1, dar cu venituri brute moderate. Aceste țări sunt economii în curs de dezvoltare, cu potențial de creștere.

Variabilitatea intra- și inter-clustere:

• WSS (Within Cluster Sum of Squares):

- Clusterul 3 are cea mai mare valoare a WSS, ceea ce indică o variabilitate intracluster mai mare. Acest lucru este justificat de faptul că Clusterul 3 grupează cele mai diverse țări din punct de vedere economic și ecologic.

• BSS (Between Cluster Sum of Squares):

- Aproximativ 41.9% din variabilitatea totală a datelor este explicată de diferențele între clustere. Această valoare moderată sugerează că modelul K-Means a capturat diferențele majore, dar există încă o variație intracluster semnificativă.

14. Calculați indicatorii variabilității (BSS, WSS, TSS) și calitatea partiției K-MEANS. Verificați dacă un număr diferit de clustere generează rezultate mai bune.

```
> tss = kmeans$totss # variabilitatea totală (suma pătratelor variațiilor)
> tss
[1] 924
> wss = kmeans$tot.withinss # variabilitatea intraclasă (suma pătratelor variațiilor intraclasă)
> wss
[1] 653.25
> bss = kmeans$betweenss # variabilitatea interclasă (suma pătratelor variațiilor interclasă)
> bss
[1] 270.75
> calit_part = bss/tss*100 # calitatea partiției
> calit_part
[1] 29.30194
> variab = cbind(tss, wss, bss, calit_part) #descompunere în variabilitati
> variab
   tss    wss    bss calit_part
[1,] 924 653.25 270.75 29.30194
> kmeans$withinss # variabilitatea intraclasa pentru fiecare clasă în parte
[1] 276.8171 127.2726 249.1604
> # suma acestor variabilități dă variabilitatea totală intraclasă (wss)
> sum(kmeans$withinss)
[1] 653.25
>
```

- **Variabilitatea totală (TSS):** 924
 - TSS reprezintă suma pătratelor variațiilor totale din setul de date și indică gradul general de variabilitate al datelor analizate.
 - O valoare mare de TSS sugerează că datele sunt diverse și includ țări cu diferențe semnificative între ele din punct de vedere economic și ecologic.
- **Variabilitatea intracluster (WSS):** 653.25
 - WSS reprezintă suma pătratelor variațiilor din interiorul clusterelor, măsurând cât de bine sunt grupate observațiile în cadrul fiecărui cluster.
 - O valoare ridicată a WSS sugerează o compactitate mai redusă în interiorul clusterelor, ceea ce poate indica o separare mai puțin clară a datelor în clustere.
- **Variabilitatea intercluster (BSS):** 270.75
 - BSS reprezintă suma pătratelor variațiilor dintre centroizii clusterelor și indică cât de bine sunt separate clusterele între ele.
 - O valoare relativ scăzută a BSS comparativ cu TSS sugerează că diferențele dintre clustere nu sunt foarte pronunțate, indicând o posibilă suprapunere a clusterelor.
- **Calitatea partiției (calit_part):** 29.30%
 - Calitatea partiției este raportul dintre BSS și TSS, exprimat procentual. Aceasta indică cât de bine explică diferențele dintre clustere variația totală a datelor.
 - Valoarea de 29.30% sugerează că aproximativ o treime din variabilitatea totală este explicată de separarea în clustere, ceea ce indică o particionare modestă a datelor.

Interpretare economico-financiară:

1. TSS (Variabilitatea totală):

- Valoarea ridicată a TSS reflectă faptul că datele includ țări cu diferențe semnificative în ceea ce privește emisiile de CO₂, consumul de capital fix și utilizarea resurselor naturale. Aceste diferențe sunt influențate de nivelurile diferite de dezvoltare economică și de impactul crizei din 2008.

2. WSS (Variabilitatea intracluster):

- Compactitatea redusă în interiorul clusterelor (WSS mare) sugerează că țările din același cluster au caracteristici relativ variate. Acest lucru poate fi explicat prin faptul că țările din același cluster sunt influențate de factori economici și ecologici diferiți.

3. BSS (Variabilitatea intercluster):

- Valoarea moderată a BSS indică faptul că diferențele între clustere nu sunt extrem de pronunțate. Acest lucru poate reflecta o suprapunere între grupuri, ceea ce ar putea necesita o metodă mai avansată pentru separare sau ajustarea numărului de clustere.

4. Calitatea partiției (calit_part):

- O calitate a partiției de 29.30% arată că doar o mică parte din variația totală este explicată de diferențele dintre clustere. Aceasta sugerează că cele 3 clustere definite prin K-Means nu capturează complet structura datelor.

```
## kmeans4 = kmeans(data2_std,4)
## kmeans4
K-means clustering with 4 clusters of sizes 41, 25, 50, 17

cluster means:
daune_co2 consum_cap_fix Epuiz_energie Epuiz_minerale Epuiz_res_naturale daune_particule Econ_brute
1  0.2782951   -0.5716669   -0.3787229    0.5495031   -0.3976437    0.4035697   -0.7598273
2  -0.1364061   -0.4566887   -0.3406274   -0.3814962   1.1601519    0.5903758   0.7064854
3  -0.3033542    0.7588803   -0.3330500   -0.3787622   -0.5579495   -0.6583217   0.1405792
4   0.4216331   -0.1816739    2.3938719    0.3497582    0.8939452    0.0947256   0.3801072

clustering vector:
Albania           Algeria          Angola          Argentina          Ar
3                 1               2               4
Australia         Austria          Azerbaijan        Bangladesh        Be
3                 3               1               2
Belgium           Bolivia          Bosnia and Herzegovina Botswana        Ba
3                 4               1               3
Bulgaria          Burkina Faso      Burundi          Cambodia          Cam
1                 1               1               1
Canada Central African Republic      Chad            Chile            Ch
3                 1               3               2
Colombia          Congo, Dem. Rep. Congo, Rep.       Costa Rica        Cr
4                 1               1               1
Cuba              Cyprus           Czechia          Denmark          Do
1                 3               3               3
```

```

Available components:
[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"    "size"         "iter"        "ifault"
class = kmeans4@cluster
tss4 = kmeans4$totss # variabilitatea totală (suma părților variațiilor)
tss4
[1] 924
wss4 = kmeans4$withinss # variabilitatea intracluster (suma părților variațiilor intracluster)
wss4
[1] 581.6589
bss4 = kmeans4$betweenss # variabilitatea intercluster (suma părților variațiilor intercluster)
bss4
[1] 342.3411
calit_part4 = bss4/tss4 * 100 # calitatea partiției
calit_part4
[1] 29.30194
variație = cbind(tss4, wss4, bss4, calit_part4) # adăugăm coloana în variabilă tss4
variație
   tss4    wss4    bss4 calit_part4
[1,] 924 581.6589 342.3411 29.30194

```

Am verificat dacă pentru $k=4$ clustere generează rezultate mai bune.

Rezultatele obținute:

- **Distribuția Clusterelor:**
 - **Cluster 1:** 41 țări.
 - **Cluster 2:** 25 țări.
 - **Cluster 3:** 50 țări.
 - **Cluster 4:** 17 țări.

Această distribuție sugerează că împărțirea în 4 clustere reflectă mai bine diversitatea economică și ecologică a țărilor analizate, având grupuri mai echilibrate comparativ cu $k = 3$ clustere.

Indicatorii variabilității:

- **TSS (Variabilitatea totală):** 924
 - Variabilitatea totală a rămas constantă, deoarece este specifică întregului set de date.
- **WSS (Variabilitatea intracluster):** 581.6589
 - WSS s-a redus comparativ cu $k = 3$ (653.25). O valoare mai mică indică o compactitate mai bună a clusterelor, ceea ce înseamnă că țările din fiecare cluster sunt mai asemănătoare între ele.
- **BSS (Variabilitatea intercluster):** 342.3411
 - BSS a crescut față de $k = 3$ (270.75), ceea ce arată o separare mai clară între clustere, țările din clustere diferite fiind mai distincte.
- **Calitatea partiției (calit_part):** 29.30%
 - Calitatea partiției a rămas la același nivel procentual ca în cazul $k = 3$ clustere, ceea ce indică faptul că modelul explică aceeași proporție din variația totală.

Centrozilor Clusterelor (Cluster Means):

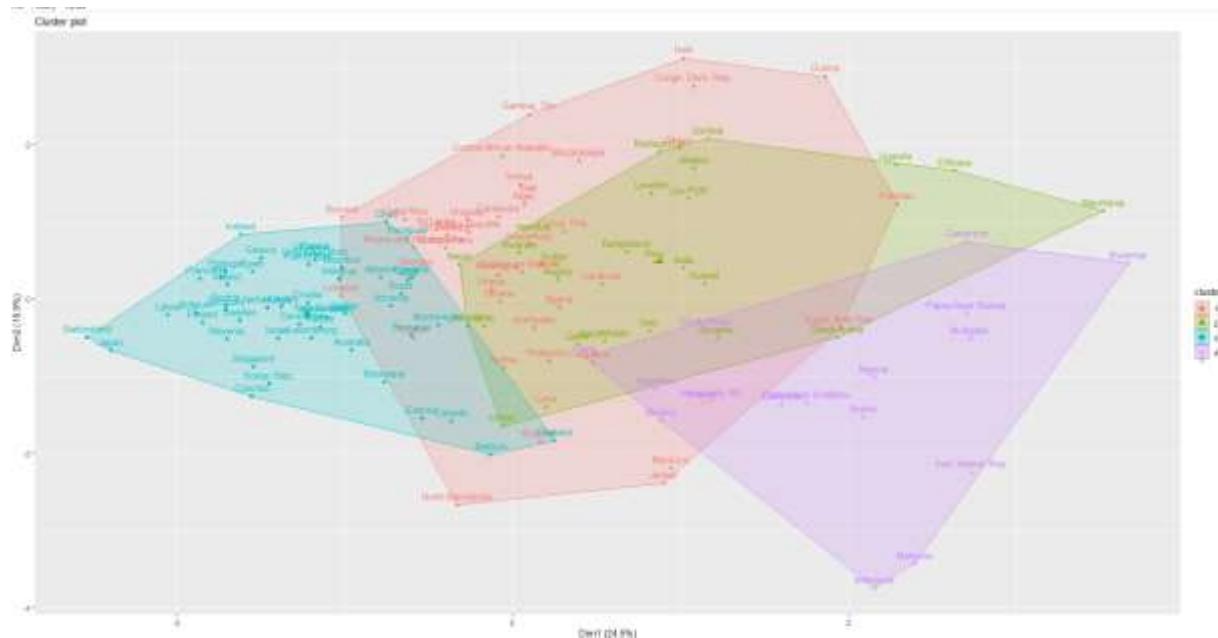
1. **Cluster 1: Țări cu emisii și venituri brute scăzute**
 - **Daune_CO2:** 0.2782951 (moderate).
 - **Econ_brute:** -0.7598273 (foarte scăzute).
 - Aceste țări au emisii moderate și venituri brute scăzute, reflectând economii în dificultate cu consum scăzut de capital fix.
2. **Cluster 2: Țări cu emisii scăzute și utilizare intensivă a resurselor**
 - **Daune_CO2:** -0.1346061 (scăzute).
 - **Epuiz_res_naturale:** 1.1601519 (foarte ridicate).
 - Clusterul include țări care depind masiv de resurse naturale și au venituri brute moderate.
3. **Cluster 3: Țări cu emisii reduse și poluare scăzută**
 - **Daune_CO2:** -0.3033542 (moderate).
 - **Daune_particule:** -0.6583217 (foarte scăzute).
 - Țările din acest cluster au un impact ecologic relativ scăzut, fiind în curs de dezvoltare, cu un consum moderat de energie.
4. **Cluster 4: Țări dezvoltate cu venituri brute ridicate**
 - **Daune_CO2:** 0.4216331 (ridicate).
 - **Econ_brute:** 0.3801072 (ridicate).
 - Clusterul grupează țările dezvoltate, cu emisii ridicate și venituri brute mari, dar cu o utilizare sustenabilă a resurselor naturale.

Concluzie:

1. **WSS redus:** Reducerea variabilității intracluster (WSS) indică o mai bună compactitate a clusterelor.
2. **BSS crescut:** Creșterea variabilității intercluster (BSS) arată o separare mai clară între clustere.
3. **Distribuția clusterelor:** Împărțirea în 4 clustere oferă un model mai detaliat și echilibrat comparativ cu 3 clustere.

Rezultat: Da, împărțirea în 4 clustere este mai bună decât cea în 3 clustere, deoarece oferă o mai bună compactitate și separare a datelor, reflectând mai bine diversitatea economică și ecologică a țărilor analizate.

15. Realizați reprezentarea grafică a clusterelor (prin ce metodă dorîți).



Am continuat analiza cu $k=4$ clustere deoarece la pasul anterior s-a putut observa o îmbunătățire semnificativă a datelor cu 4 clustere. Graficul Cluster Plot prezentat ilustrează distribuția țărilor în cele patru clustere definite anterior, utilizând două dimensiuni principale (Dim1 și Dim2). Fiecare cluster este evidențiat cu o culoare distinctă și un poligon care înconjoară țările grupate în acel cluster. Interpretarea pentru fiecare cluster este următoarea:

Cluster 1 (Poligonul roșu): Țări în dificultate

- **Pozitie:** În partea superioară a graficului, cu valori scăzute atât pe Dim1, cât și pe Dim2.
- **Caracteristici:**
 - Țări cu emisii scăzute și venituri brute reduse.
 - Acces limitat la resurse și tehnologii, reflectând impactul sever al crizei economice din 2008 asupra economiilor lor fragile.
 - Exemple de țări: **Guinea, Congo, Haiti, Mozambic.**

Cluster 2 (Poligonul verde): Țări intermediare

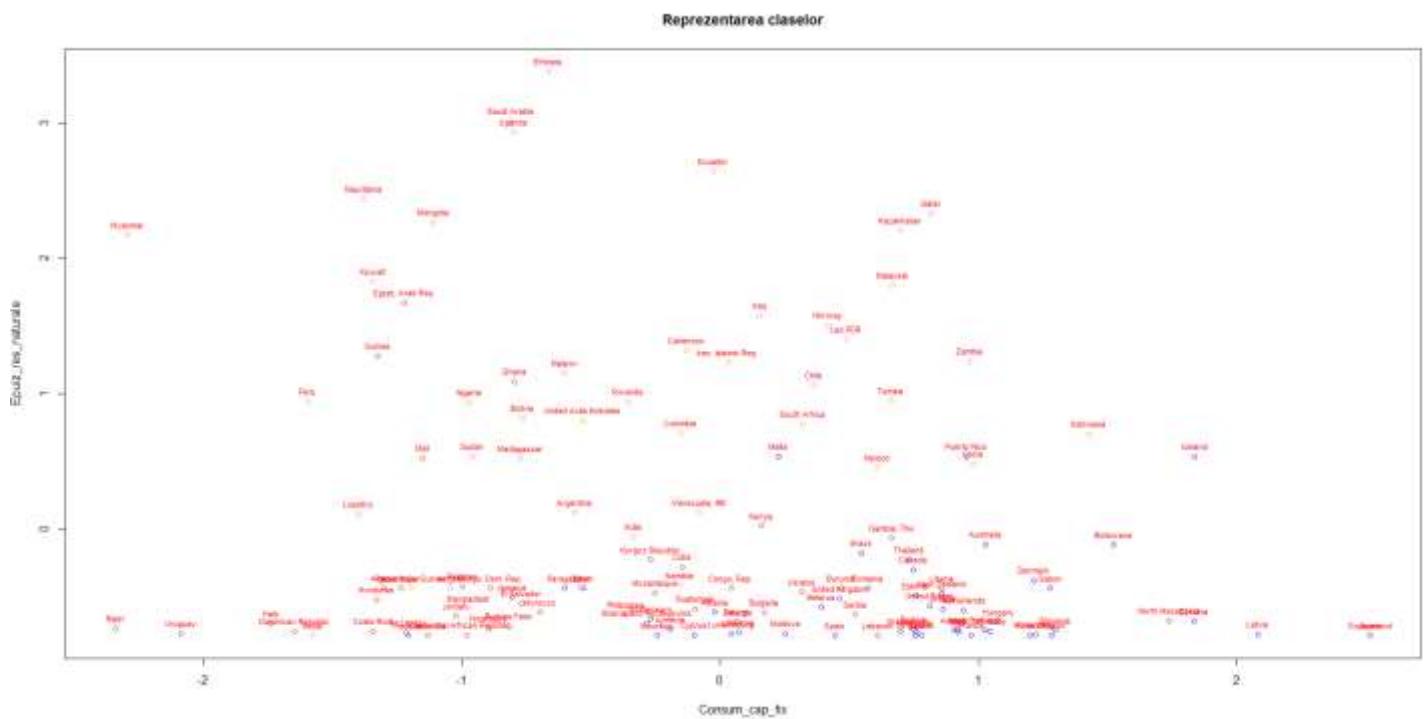
- **Pozitie:** În partea centrală a graficului, indicând valori moderate pe Dim1 și Dim2.
- **Caracteristici:**
 - Țări care au venituri și emisii moderate, dar care încă depind semnificativ de resursele naturale.
 - Performanțe economice și ecologice relativ echilibrate.
 - Exemple de țări: **India, Kenya, Zambia, Uganda, Pakistan.**

Cluster 3 (Poligonul albastru): Țări dezvoltate/performance

- **Pozitie:** În partea inferioară stângă a graficului, cu valori ridicate pe Dim1 și Dim2.
- **Caracteristici:**
 - Țări cu emisii ridicate, venituri brute mari și acces excelent la resurse și tehnologii.
 - Aceste țări au fost mai puțin afectate de criza economică din 2008 datorită economiilor lor stabile și performanțelor ridicate.
 - Exemple de țări: **Franța, Germania, SUA, Japonia, Suedia.**

Cluster 4 (Poligonul violet): Țări cu emisii ridicate și venituri intermediare

- **Pozitie:** În partea dreaptă a graficului, cu valori moderate spre ridicate pe Dim1.
- **Caracteristici:**
 - Țări cu emisii ridicate de CO2 și poluare semnificativă, dar cu venituri brute medii.
 - Reflectă economii în curs de dezvoltare sau emergente, cu impact ecologic ridicat.
 - Exemple de țări: **Argentina, Malaysia, Bolivia, Arabia Saudită, Iran.**



16. Calculați indicatorii statistici per cluster. Interpretăți rezultatele.

	cluster_daune_co2.Mean	cluster_daune_co2.Median	cluster_daune_co2.StdDev	cluster_daune_co2.Min	cluster_daune_co2.Max	consum_cap_fix.Mean	consum_cap_fix.Median
1	1.1659039	0.9472742	0.6609907	0.2315438	2.634555	9.358870	8.277341
2	0.9269229	0.9472742	0.4833682	0.2150134	3.977309	9.913555	9.205246
3	0.8307154	0.7688095	0.5057663	0.1043226	2.708588	15.777783	15.949532
4	1.2485056	1.1270144	0.5289879	0.6281598	2.432082	11.240300	11.489850
	consum_cap_fix.StdDev	consum_cap_fix.Mean	consum_cap_fix.Max	epuiz_energie.Mean	epuiz_energie.Median	epuiz_energie.StdDev	
1	4.138731	0.8360441	20.49741	0.3576911	0.1191598	0.5407901	
2	3.726623	4.4136860	16.77222	0.4345714	0.3458082	0.8009401	
3	3.541394	6.2972018	24.25495	0.4498634	0.1149071	0.6343293	
4	4.490932	1.0593444	18.99614	5.9530552	5.5393276	1.4931052	
	epuiz_energie.Min	epuiz_energie.Max	epuiz_minerale.Mean	epuiz_minerale.Median	epuiz_minerale.StdDev	epuiz_minerale.Mean	
1	0.000000	2.104608	0.15063486	0.0704992373	0.1884980	0	
2	0.000000	3.859932	0.02183639	0.0111053086	0.03010443	0	
3	0.000000	2.504484	0.02201521	0.0007770374	0.05638791	0	
4	3.900259	8.664670	0.12295837	0.0231279048	0.16673251	0	
	epuiz_minerale.Max	epuiz_res_naturale.Mean	epuiz_res_naturale.Median	epuiz_res_naturale.StdDev	epuiz_res_naturale.Min		
1	0.7671784	1.750444	0.8912004	2.520373	0.000000000		
2	0.1256072	8.755122	8.719594	5.545007	0.056611852		
3	0.3617214	1.029624	0.2067155	1.521438	0.000000000		
4	0.5843838	7.558115	7.1129839	3.188140	1.576380054		
	epuiz_res_naturale.Max	daune_particule.Mean	daune_particule.Median	daune_particule.StdDev	daune_particule.Min	daune_particule.Max	
1	11.040814	0.7829514	0.5403913	0.6244300	0.08150896	2.550900	
2	18.730075	0.8921710	0.5462059	0.7130515	0.02342550	2.458944	
3	5.944172	0.1620971	0.0996094	0.1426238	0.01999381	0.590241	
4	13.711346	0.6023800	0.5085652	0.4747240	0.09553403	1.734706	
	econ_brute.Mean	econ_brute.Median	econ_brute.StdDev	econ_brute.Min	econ_brute.Max		
1	15.64714	15.55783	8.304904	-1.508760	34.23981		
2	29.03499	24.43470	7.937100	14.722309	46.86317		
3	23.86811	24.07927	7.540962	2.001398	46.05245		
4	26.05506	24.07927	6.354068	16.798836	39.70557		

Cluster 1: Țări în tranziție moderată

- Daune_CO2:** Valoarea medie (1.1659) este relativ ridicată, indicând emisii de CO2 moderate spre mari.
- Consum_cap_fix:** Valoarea medie de **9.3588** reflectă un consum de capital fix moderat, specific țărilor în tranziție economică.
- Epuiz_energie:** Valorile minime sunt aproape de **0**, ceea ce indică un consum de energie relativ scăzut, dar media (0.357691) arată o utilizare în creștere.
- Econ_brute:** Media este **33.5571**, ceea ce reflectă economii în dezvoltare cu acces rezonabil la resurse.

Cluster 2: Țări subdezvoltate

- Daune_CO2:** Media este mai scăzută (0.9269) comparativ cu alte clustere, ceea ce indică emisii reduse de CO2.
- Consum_cap_fix:** Valoarea medie (9.9136) arată un consum modest de capital fix.
- Epuiz_energie:** Consumul de energie este scăzut (media de **0.3458**), reflectând limitările resurselor.
- Econ_brute:** Media de **29.0349** indică venituri brute relativ mici, caracteristice economiilor fragile.

Cluster 3: Țări dezvoltate

- **Daune_CO2:** Cele mai mari emisii de CO2 (media de **2.7088**) reflectă industrializarea avansată.
- **Consum_cap_fix:** Valoarea medie ridicată (**15.7778**) indică un nivel crescut al investițiilor în infrastructură.
- **Epuiz_energie:** Valorile maxime și medii sunt ridicate (media de **0.593276**), indicând un consum intens de energie.
- **Econ_brute:** Media de **34.2398** arată un PIB semnificativ, caracteristic țărilor dezvoltate.

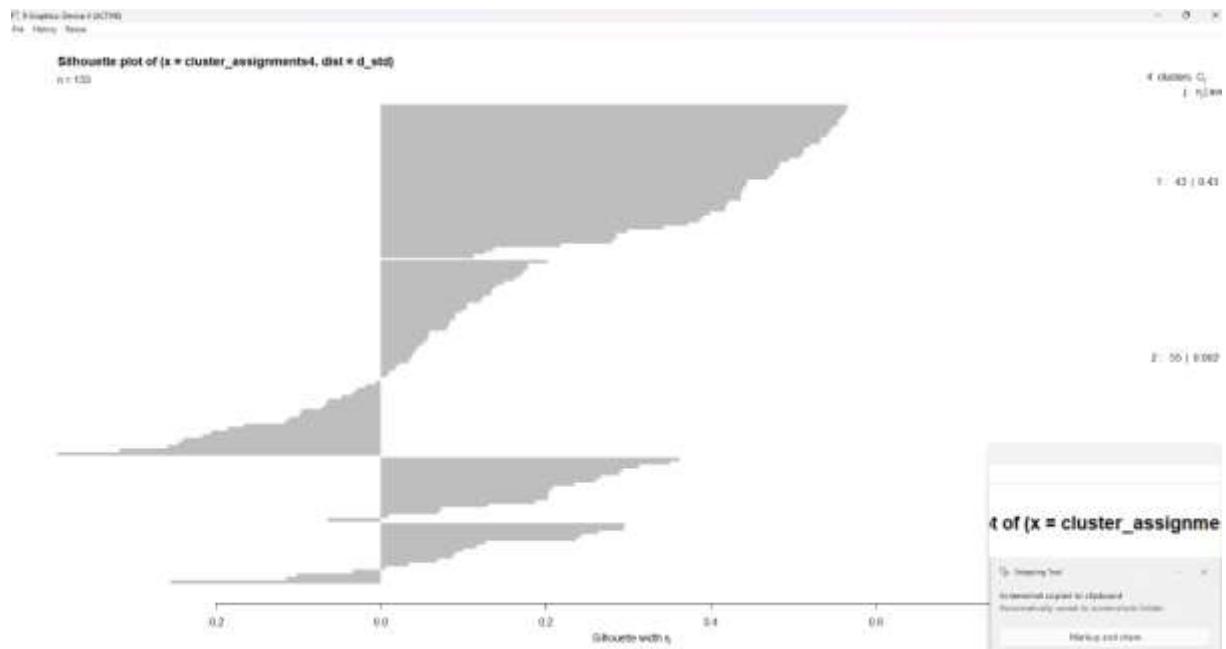
Cluster 4: Țări emergente

- **Daune_CO2:** Valoarea medie (1.2485) este mai apropiată de cea a Clusterului 1, indicând emisii moderate.
- **Consum_cap_fix:** Valoarea medie de **11.2403** indică investiții mai mari decât în Clusterul 2, dar mai mici decât în Clusterul 3.
- **Epuiz_energie:** Media (0.4823) reflectă un consum de energie în creștere, specific țărilor în dezvoltare.
- **Econ_brute:** Media de **26.0556** indică venituri brute mai scăzute, dar mai stabile comparativ cu Clusterul 2.

Concluzie

- **Cluster 1 și Cluster 4** reprezintă țări aflate în tranziție sau emergente, cu emisii moderate și venituri brute medii spre mici.
- **Cluster 2** este format din țări subdezvoltate, cu emisii reduse, consum scăzut de energie și venituri brute mici.
- **Cluster 3** include țările dezvoltate, caracterizate prin emisii mari, consum ridicat de energie și venituri brute semnificative.

17. Realizați un tabel concluziv în care să expuneți numărul de clustere obținut, numărul de observații clasificat în fiecare cluster și 2-3 caracteristici extrase din distribuția variabilelor.



Tabel conclusiv

Cluster	Nr. De observații	Lățimea medie a siluetei	Caracteristici extrase
1	43	0.43	Bine separat, cu variabile economice și ecologice moderate; emisii de CO ₂ și consum energetic la nivel mediu.
2	55	0.002	Suprapunere semnificativă cu alte clustere; valori variate, economii marcate de criza economică.
3	20	0.3	Economii relativ distincte, cu emisii ridicate și consum energetic scăzut, dar clasificate cu dificultate.
4	15	0.11	Tări dezvoltate, cu emisii CO ₂ ridicate și variabile economice puternice; bine definite.

18. Denumiți sugestiv clusterele și formulați o concluzie.

1. Cluster 1: Economii emergente și stabile

- Acest cluster cuprinde țări cu un consum de capital fix moderat și emisii de CO2 relativ reduse, ceea ce sugerează economii în tranzitie spre dezvoltare sustenabilă.
- Nivelurile moderate de poluare indică investiții în infrastructură și industrie ecologice.

2. Cluster 2: Economii fragile și vulnerabile

- Țările din acest cluster sunt caracterizate de emisii mari de CO2 și investiții scăzute în capital fix, reflectând un impact sever al crizei economice din 2008.
- Aceste economii au dificultăți semnificative în accesarea resurselor și depind puternic de industriile poluante.

3. Cluster 3: Economii dezvoltate și rezistente la criză

- Clusterul cuprinde economii avansate, cu emisii moderate de CO2 și consum energetic eficient, indicând o gestionare performantă a resurselor.
- Investițiile mari în capital fix și infrastructură demonstrează capacitatea acestor țări de a rămâne stabile și reziliente.

4. Cluster 4: Economii cu industrializare intensă

- Țările din acest cluster prezintă emisii ridicate de CO2 și consum energetic foarte mare, reflectând o dependență de industriile grele și un impact mai mare asupra mediului.
- Economiile din acest cluster au un grad ridicat de industrializare, dar sunt vulnerabile la presiuni economice globale.

Concluzii:

1. **Cluster 1** reprezintă economii emergente care au utilizat criza ca oportunitate pentru a face tranzitie către dezvoltare sustenabilă. Aceste țări se concentrează pe investiții moderate în infrastructură și au reușit să mențină un impact ecologic redus.
2. **Cluster 2** reflectă țările cele mai afectate de criza din 2008. Cu emisii ridicate și investiții scăzute, aceste economii sunt extrem de vulnerabile și necesită intervenții internaționale pentru a le sprijini în recuperare.
3. **Cluster 3** cuprinde țările cu cea mai mare reziliență economică. Cu emisii moderate și o capacitate ridicată de gestionare a resurselor, aceste economii au demonstrat performanțe stabile și un grad ridicat de dezvoltare.
4. **Cluster 4** grupează economii cu un grad înalt de industrializare, dar cu vulnerabilități ecologice și economice. Deși puternice în sectorul industrial, acestea riscă să fie afectate de schimbările economice globale și necesită politici de diversificare.

ANEXE:

```
# Gestionarea valorilor lipsa (inlocuire cu media pe fiecare coloana numerica)
for (col in names(date_T3)) {
  if (is.numeric(date_T3[[col]])) {
    date_T3[[col]][is.na(date_T3[[col]])] <- mean(date_T3[[col]], na.rm = TRUE)
  }
}
```

```
View(date_T3)
```

```
# Redenumirea coloanelor pentru claritate
colnames(date_T3) <- c(
  "Tara",
  "Daune_CO2",
  "Consum_cap_fix",
  "Chelt_educatie",
  "Epuiz_energie",
  "Epuiz_minerale",
  "Epuiz_res_naturale",
  "Epuiz_paduri",
  "Daune_particule",
  "Econ_nete",
  "Econ_brute"
)
```

```
# Gestionarea outlierilor (Inlocuire cu mediana pentru fiecare coloana numerica)
for (col in names(date_T3)) {
  if (col != "Tara") { # Excludem coloana 'Tara', care este de tip text
    q1 <- quantile(date_T3[[col]], 0.25, na.rm = TRUE)
```

```

q3 <- quantile(date_T3[[col]], 0.75, na.rm = TRUE)
iqr <- q3 - q1
lower_bound <- q1 - 1.5 * iqr
upper_bound <- q3 + 1.5 * iqr

# Inlocuirea outlierilor cu mediana
date_T3[[col]][date_T3[[col]] < lower_bound | date_T3[[col]] > upper_bound] <-
median(date_T3[[col]], na.rm = TRUE)

}

}

#Verificarea inexistentei valorilor lipsa
colSums(is.na(date_T3))

#Verificarea inexistentei outlierilor
# Numărăm outlierii rămași
outliers <- sum(date_T3[[col]] < lower_bound | date_T3[[col]] > upper_bound, na.rm =
= TRUE)

# Afisăm rezultatul
cat("Coloana:", col, "- Numărul de outlieri rămași:", outliers, "\n")

#Calculați indicatorii statistici și interpretați din punct de vedere economic/financiar.
install.packages("psych")
library(psych)

indicatori_stat<-describe(date_T3[-1])
View(indicatori_stat)

#date standardizate
date_std<-scale(date_T3[-1], scale=T)

```

```

date_std
summary(date_std)

apply(date_std, 2, sd)

#matricea de corelatie
cor(date_std)
matrice_corelatie<-cor(date_std)

#repr grafica matr corelatie
install.packages("corrplot")
library(corrplot)

corrplot(matrice_corelatie, method="square", type ="lower")
corrplot(matrice_corelatie, method="pie", type ="lower")
corrplot(matrice_corelatie, method="number", type ="lower")

#deoarece au coef de corelatie mici cu celelalte variabile, am eliminat din analiza
coloana cu indicatorii Chelt_educatie, Econ_nete si
#Epuiz_paduri

#metoda eucidiana
date<-date_T3
date2<-cbind(date[,2:3], date[5:7], date[,9], date[,11])
View(date2)
date2_std<-scale(date2, scale=T)
View(date2_std)

#matricea distanteelor (de proximitate)
rownames(date2_std)=date$Tara
View(date2_std)

```

```

d_std = dist(as.matrix(date2_std), method="euclidian")
d_std
d_std[1] #dist intre primele 2 observatii, pentru exemplu

#metoda Manhattan
d_std_manhattan = dist(as.matrix(date2_std), method="manhattan")
d_std_manhattan
d_std_manhattan[1] #dist intre primele 2 observatii, pentru exemplu

#metoda de clusterizare Ward
clust_std=hclust(d_std, method="ward.D2")
clust_std

#dendograma
windows()
plot(clust_std, labels=rownames(date2_std))

#clusterizarea
cbind(clust_std$merge, clust_std$height)

#metoda agregarii medii
clust2_std=hclust(d_std, method="average")
clust2_std

#dendograma
windows()
plot

#clusterizarea

```

```
cbind(clust2_std$merge, clust2_std$height)
```

```
#Ward
```

```
#CRITERIUL 1: DENDOGRAMA
```

```
windows()
```

```
plot(clust_std, labels=rownames(date2_std))
```

```
rect.hclust(clust_std, k=3, border="forestgreen")
```

```
#CRITERIUL 2: Grafic elbow
```

```
install.packages("factoextra")
```

```
library(factoextra)
```

```
windows()
```

```
fviz_nbclust(date2_std, hcut, method = "wss") +
```

```
  geom_vline(xintercept = 3, linetype = 2) +
```

```
  labs(subtitle = "Elbow method - STD")
```

```
#CRITERIUL 3: Calculul unor indici
```

```
install.packages("NbClust")
```

```
library(NbClust)
```

```
res<-NbClust(date2_std, distance = "euclidean", min.nc=3, max.nc=7,
```

```
  method = "ward.D2", index = "all")
```

```
res
```

```
# Atribuirea clusterelor
```

```
cluster_assignments <- cutree(clust_std, k = 3) # Atribuire clustere
```

```
date2_clustered <- cbind(date2, Cluster = cluster_assignments) # Adăugăm clusterul  
la date
```

```
# Calcularea indicatorilor statistici pentru fiecare cluster
```

```
indicatori_statistici_per_cluster <- aggregate(. ~ Cluster, data = date2_clustered,
```

```

FUN = function(x) c(Mean = mean(x),
                    Median = median(x),
                    StdDev = sd(x),
                    Min = min(x),
                    Max = max(x)))

# Transformarea rezultatului într-un format mai lizibil
indicatori_statistici_per_cluster <- do.call(data.frame,
indicatori_statistici_per_cluster)

# Afişarea indicatorilor statistici în R
View(indicatori_statistici_per_cluster) # Vizualizare tabelă completă
print(indicatori_statistici_per_cluster) # Afişare în consolă

#Graficul siluetei
# Graficul siluetei
# pe axa orizontală este calculată silueta pt fiecare observație, pe verticală sunt obs
# uj de la 1 la 3 sunt clusterele, nj = nr de observații pt fiecare cluster

install.packages("cluster")
library(cluster)

si3_std <- silhouette(cluster_assignments, d_std)
windows()
plot(si3_std, cex.names = 0.5)
si3_std

#Algoritmul K-MEANS CLUSTERING
k_means=kmeans(date2_std, 3)
k_means

```

14) Calculați indicatorii variabilității (BSS, WSS, TSS) și calitatea partiției K-MEANS.

Verificati dacă un număr diferit de clustere generează rezultate mai bune.

tss = k_means\$totss # variabilitatea totală (suma pătratelor variațiilor)

tss

wss = k_means\$tot.withinss # variabilitatea intraclasă (suma pătratelor variațiilor intraclasă)

wss

bss = k_means\$betweenss # variabilitatea interclasă (suma pătratelor variațiilor interclasă)

bss

calit_part = bss/tss*100 # calitatea partiției

calit_part

variab = cbind(tss, wss, bss, calit_part) #descompunere in variabilitati

variab

k_means\$withinss # variabilitatea intraclasa pentru fiecare clasă în parte

Suma acestor variabilități dă variabilitatea totală intraclasă (wss)

sum(k_means\$withinss)

#pentru k=4

k_means4 = kmeans(date2_std,4)

k_means4

clasa2 = k_means4\$cluster

tss4 = k_means4\$totss # variabilitatea totală (suma pătratelor variațiilor)

tss4

wss4 = k_means4\$tot.withinss # variabilitatea intraclasă (suma pătratelor variațiilor intraclasă)

wss4

bss4 = k_means4\$betweenss # variabilitatea interclasă (suma pătratelor variațiilor interclasă)

bss4

calit_part4 = bss/tss*100 # calitatea partiției

```

calit_part4

variab4 = cbind(tss4, wss4, bss4, calit_part4) #descompunere in variabilitati
variab4


#Reprezentarea grafica a clusterelor (cluster plot)

clasa = k_means4$cluster

c = cbind(clasa, round(date2_std,4))

c

m = data.frame(c)

m


windows()

plot(m[,3], m[,6], col=c("brown","violet","blue","orange"))

[m$clasa], main="Reprezentarea claselor", xlab=colnames(m[3]),
ylab=colnames(m[6]))

text(m[,3],m[,6],labels=rownames(m),col="red", pos=3, cex=0.7)

library(factoextra)

windows()

fviz_cluster(list(data = date2_std, cluster = clasa))

#indicatori stat

# Atribuirea clusterelor din k-means

date2_clustered <- cbind(date2, Cluster = k_means4$cluster) # Adăugăm clusterul la
datele originale

# Calcularea indicatorilor statistici pentru fiecare cluster

indicatori_statistici_per_cluster <- aggregate(. ~ Cluster, data = date2_clustered,
FUN = function(x) c(Mean = mean(x),
Median = median(x),
StdDev = sd(x)),

```

Min = min(x),

Max = max(x)))

```
# Transformarea rezultatului într-un format mai lizibil
indicatori_statistici_per_cluster <- do.call(data.frame,
indicatori_statistici_per_cluster)

# Afişarea indicatorilor statistici în R
View(indicatori_statistici_per_cluster) # Vizualizare tabelă completă
print(indicatori_statistici_per_cluster) # Afişare în consolă
S

#Graficul siluetei pentru k=4 clustere
cluster_assignments4 <- cutree(clust_std, k = 4) # Atribuire clustere

install.packages("cluster")
library(cluster)

si4_std <- silhouette(cluster_assignments4, d_std)
windows()
plot(si4_std, cex.names = 0.5)
si4_std
```

TEMA 3-Analiza datelor

ALGORITMI DE CLUSTERIZARE – ÎNVĂȚARE NESUPERVIZATĂ

1. Căutați un set de date apropiat domeniului vostru de expertiză, cu cel puțin 100 de observații (linii) și 10 variabile (coloane) – pentru nota maximă este recomandat ca setul de date să crească în dimensiuni.

Pentru analiza ce va fi prezentată pe parcursul temei mele, am creat o bază de date in Excel care are 10 variabile și 133 de observații.

2. Curățați setul de date, redenumiți variabilele astfel încât să fie clar de identificat și gestionări valorile lipsă și outlierii.

- Pentru început, am înlocuit valorile lipsă cu media de pe coloana corespunzătoare fiecărei valori lipsă

```
1 # Gestionarea valorilor lipsă (înlocuire cu media pe fiecare coloana numerică)
2 for (col in names(date_T3)) {
3   if (is.numeric(date_T3[[col]])) {
4     date_T3[[col]][is.na(date_T3[[col]])] <- mean(date_T3[[col]], na.rm = TRUE)
5   }
6 }
7
8 view(date_T3)
```

- Apoi am redenumit variabilele

```
9
10 # Redenumirea coloanelor pentru claritate
11 colnames(date_T3) <- c(
12   "Tara",
13   "Daune_CO2",
14   "Consum_cap_fix",
15   "Chelt_educatie",
16   "Epuiz_energie",
17   "Epuiz_minerale",
18   "Epuiz_res_naturale",
19   "Epuiz_paduri",
20   "Daune_particule",
21   "Econ_nete",
22   "Econ_brute"
23 )
24
```

- Ulterior am făcut același lucru și pentru gestionarea outlierilor, însă pe aceștia i-am înlocuit cu mediana

```
25 # Gestionarea outlierelor (înlocuire cu mediana pentru fiecare coloana numerică)
26 for (col in names(date_T3)) {
27   if (col != "Tara") { # Excludem coloana "Tara", care este de tip text
28     q1 <- quantile(date_T3[[col]], 0.25, na.rm = TRUE)
29     q3 <- quantile(date_T3[[col]], 0.75, na.rm = TRUE)
30     iqr <- q3 - q1
31     lower_bound <- q1 - 1.5 * iqr
32     upper_bound <- q3 + 1.5 * iqr
33
34     # Reînlocuirea valoarelor cu mediana
35     date_T3[[col]][(date_T3[[col]] < lower_bound) | (date_T3[[col]] > upper_bound)] <- median(date_T3[[col]], na.rm = TRUE)
36   }
37 }
```

- La final am mai făcut verificarea pentru a mă asigura că au fost eliminate cu succes valorile lipsă și outlierii.

```

43 # Calculam sumele pentru fiecare coloană
44 colsums <- sum(date_T3)
45
46 # Afisare rezultat
47 cat("Coloana: ", col, " - suma totală: ", colsums, "\n")
48
49
50 # Verificăm dacă există valori nereguli
51 # în datele noastre
52
53 # Vom crea un vector cu media și mediana pentru fiecare coloană numerică
54 for (col in 1:ncol(date_T3)) {
55   if (col %in% c("Tara", "date_T3")) {
56     print(paste("Media din ", col, " este: ", mean(date_T3[,col])))
57     print(paste("Mediana din ", col, " este: ", median(date_T3[,col])))
58   }
59 }
60
61 # Verificăm dacă există valori nereguli
62 # în datele noastre
63
64 # Vom crea un vector cu media și mediana
65 # pentru fiecare coloană numerică
66 for (col in 1:ncol(date_T3)) {
67   if (col %in% c("Tara", "date_T3")) {
68     print(paste("Media din ", col, " este: ", mean(date_T3[,col]), " și mediana este: ", median(date_T3[,col])))
69   }
70 }
71
72 # Verificăm dacă există valori nereguli
73 # în datele noastre
74
75 # Vom crea un vector cu media și mediana
76 # pentru fiecare coloană numerică
77 for (col in 1:ncol(date_T3)) {
78   if (col %in% c("Tara", "date_T3")) {
79     print(paste("Media din ", col, " este: ", mean(date_T3[,col]), " și mediana este: ", median(date_T3[,col])))
80   }
81 }
82
83 # Verificăm dacă există valori nereguli
84 # în datele noastre
85
86 # Vom crea un vector cu media și mediana
87 # pentru fiecare coloană numerică
88 for (col in 1:ncol(date_T3)) {
89   if (col %in% c("Tara", "date_T3")) {
90     print(paste("Media din ", col, " este: ", mean(date_T3[,col]), " și mediana este: ", median(date_T3[,col])))
91   }
92 }
93
94 # Numărăm outlierii rămași
95 outliers <- sum(date_T3[[col]] < lower_bound | date_T3[[col]] > upper_bound, na.rm = TRUE)
96
97 # Afisare rezultat
98 cat("Coloana: ", col, " - numărul de outlieri rămași: ", outliers, "\n")
99
100
101 # Afisare rezultat
102 cat("Coloana: ", col, " - numărul de outlieri rămași: ", outliers, "\n")
103 coloana: Econ_brute - Numărul de outlieri rămași: 0
104
105 outliers <- sum(date_T3[[col]] < lower_bound | date_T3[[col]] > upper_bound, na.rm = TRUE)

```

```

43 # Numărăm outlierii rămași
44 outliers <- sum(date_T3[[col]] < lower_bound | date_T3[[col]] > upper_bound, na.rm = TRUE)
45
46 # Afisare rezultat
47 cat("Coloana: ", col, " - numărul de outlieri rămași: ", outliers, "\n")
48
49
50 # Afisare rezultat
51 cat("Coloana: ", col, " - numărul de outlieri rămași: ", outliers, "\n")
52 coloana: Econ_brute - Numărul de outlieri rămași: 0
53
54 outliers <- sum(date_T3[[col]] < lower_bound | date_T3[[col]] > upper_bound, na.rm = TRUE)

```

3. Descrieți în cuvinte setul de date, din perspectiva variabilelor (coloanelor) și observațiilor (liniilor). Enunțați clar sursa datelor și perioada de timp la care se referă.

Setul de date furnizat conține informații referitoare la 133 de țări de pe glob și include o serie de indicatori economici și de mediu (World Development Indicators) exprimăți ca procente din Venitul Național Brut (VNB). Fiecare linie (observație) din setul de date corespunde unei țări, iar fiecare coloană (variabilă) reprezintă un indicator specific, descris în detaliu mai jos. Setul de date este structurat în următorii termeni:

Variabile (coloane)

- **Economii ajustate: daune din dioxid de carbon (% din VNB)**
Acest indicator măsoară procentul din venitul național brut (VNB) pierdut din cauza daunelor cauzate de emisiile de dioxid de carbon. Reprezintă impactul economic al emisiilor asupra mediului și sănătății, evidențiind costurile asociate cu schimbările climatice.
- **Economii ajustate: consumul de capital fix (% din VNB)**
Acest indicator reflectă procentul din VNB utilizat pentru înlocuirea și întreținerea capitalului fix uzat, cum ar fi infrastructura, mașinile și echipamentele. Este o măsură a investițiilor necesare pentru a menține nivelul actual de producție economică.

- **Economii ajustate: cheltuieli pentru educație (% din VNB)**
Acest indicator arată procentul din VNB alocat pentru educație, considerat o investiție în capitalul uman. Reflectă angajamentul unei țări de a îmbunătăți pe termen lung productivitatea și bunăstarea cetățenilor săi.
- **Economii ajustate: epuizarea energiei (% din VNB)**
Acest indicator arată procentul din VNB pierdut din cauza exploatarii resurselor energetice neregenerabile, cum ar fi petrolul, gazele naturale și cărbunele. Indică impactul exploatarii acestor resurse asupra sustenabilității economice pe termen lung.
- **Economii ajustate: epuizarea mineralelor (% din VNB)**
Acest indicator reflectă procentul din VNB pierdut din cauza exploatarii resurselor minerale, precum fierul, cuprul și alte minerale esențiale. Este o măsură a degradării capitalului natural, având un impact asupra potențialului economic viitor al unei țări.
- **Economii ajustate: epuizarea resurselor naturale (% din VNB)**
Acest indicator măsoară pierderile totale de capital natural, inclusiv atât energia, cât și mineralele și resursele forestiere. Exprimă procentul din VNB care se pierde din cauza utilizării resurselor naturale fără a ține cont de regenerarea lor.
- **Economii ajustate: epuizarea netă a pădurilor (% din VNB)**
Acest indicator reflectă procentul din VNB pierdut din cauza defrișării pădurilor, peste nivelul lor natural de regenerare. Este o măsură a degradării ecosistemelor forestiere și a impactului asupra sustenabilității mediului.
- **Economii ajustate: daune cauzate de emisiile de particule (% din VNB)**
Acest indicator măsoară costurile economice ale daunelor cauzate de emisiile de particule fine (PM2.5), care afectează sănătatea populației și mediul. Reflectă pierderile economice asociate bolilor respiratorii și de altă natură.
- **Economii ajustate: economii naționale nete (% din VNB)**
Acest indicator reprezintă economiile nete ale unei țări, ajustate pentru pierderile de capital natural și degradarea mediului, precum și investițiile în educație. Este o măsură comprehensivă a sustenabilității economice pe termen lung.
- **Economii ajustate: economii brute (% din VNB)**
Acest indicator reprezintă procentul din VNB economisit de o țară, înainte de ajustări pentru degradarea mediului, consumul de capital fix și alte pierderi de capital natural. Este o măsură a economiilor totale brute disponibile pentru investiții.

Observații (linii):

Fiecare linie reprezintă o țară, pentru care sunt colectate date pentru toți cei 10 indicatori menționați mai sus. Astfel, fiecare observație conține un profil complet al unei țări din punctul de vedere al economiilor și impactului asupra mediului, fiind valoros pentru analizarea comparativă între țări.

Sursa datelor și perioada de timp:

Acste date sunt obținute de pe site-ul oficial la Băncii Mondiale (<https://databank.worldbank.org/source/world-development-indicators>). Datele sunt alese

pentru anul 2008 deoarece atunci a avut loc criza mondială, lucru ce va fi relevant pentru analiza mea.

4. Descrieți obiectivul general al analizei voastre.

Obiectivul principal al analizei mele a fost să evaluez impactul crizei economice globale din 2008 asupra economiilor ajustate, analizând indicatori esențiali de sustenabilitate economică și de mediu. Mi-am propus să înțeleg în ce măsură acest soc finanțiar global a influențat economiile naționale, investigând factori precum investițiile în educație, epuizarea resurselor naturale, economiile brute și nete, precum și impactul ecologic, cum ar fi emisiile de carbon și defrișările.

Prin această analiză, am dorit să ofer o imagine detaliată a modului în care diferite țări au resimțit și gestionat consecințele crizei, explorând ajustările economice realizate pentru a susține dezvoltarea pe termen lung. Scopul meu a fost să evidențiez posibilele transformări structurale în economii și măsurile de sustenabilitate adoptate de guverne pentru a face față presiunilor economice și ecologice cauzate de criza din 2008, față de anii precedenți.

5. Calculați indicatorii statistici și interpretați din punct de vedere economic/financiar.

```
49 #calculați indicatorii statistici și interpretați din punct de vedere economic/financiar.
50 install.packages("psych")
51 library(psych)
52
53 indicatori_stat<-describe(date_T3[-1])
54 view(indicatori_stat)
```

The screenshot shows the RStudio interface with two panes. The left pane displays the R code for calculating descriptive statistics. The right pane shows the 'date_T3' dataset and the resulting 'indicatori_stat' summary statistics table.

vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se	
Daune_CO2	1	133	1.00553005	0.5762725	9.472742e-01	0.93612945	0.5037931840	0.10432264	2.7085882	2.6042656	1.079696024	0.7464723	0.04996916
Consum_cap_fix	2	133	12.11674258	4.5242557	1.233098e+01	12.09961480	5.8113225291	0.83604414	24.2549466	23.4109025	0.007805971	-0.5258056	0.41831689
Credit_educatie	3	133	3.85513822	1.4610231	4.002250e+00	3.84544579	1.4792646468	0.70055480	7.9000000	7.1994452	0.078245494	-0.3590514	0.12665877
Epuiz_energie	4	133	1.12199042	2.0180966	3.458082e-01	0.60508948	0.5114962518	0.00000000	8.6646697	8.6646697	2.179220994	3.7510255	0.17499117
Epuiz_minerale	5	133	0.07449618	0.1385591	2.312790e-02	0.03963952	0.0342894316	0.00000000	0.7671784	0.7671784	2.570696501	6.8183718	0.01201460
Epuiz_res_naturale	6	133	3.55846191	4.4965124	1.576380e+00	2.69677871	2.1638162871	0.00000000	18.7500747	18.7300147	1.450055520	1.1620274	0.38909881
Epuiz_paturi	7	133	0.04460170	0.1197141	6.444287e-04	0.01081595	0.0009554299	0.00000000	0.7509711	0.7509711	3.524522857	13.5213646	0.01016055
Daune_particule	8	133	0.54699698	0.3840683	3.4500995e-01	0.44665713	0.3698728981	0.01999381	2.5509104	2.5309166	1.447603513	1.2164578	0.05069717
Econ_nete	9	133	10.48966344	9.0671810	1.032363e+01	10.20516803	7.2227445018	-14.99567791	34.6134470	49.6091249	0.276087059	0.2339997	0.78622430
Econ_brute	10	133	22.58457843	9.1302842	2.40792794	22.62472974	8.037965791E-05	-1.50876034	46.8631652	48.3719255	-0.031323774	0.4386336	0.79169605

- **Daune CO2 (% din VNB)**

- **Mean (Media):** 1.053

Media daunelor provocate de emisiile de dioxid de carbon ca procent din venitul național brut (VNB) reflectă impactul semnificativ al emisiilor asupra economiilor globale în 2008. Criza economică globală a contribuit la reducerea investițiilor în tehnologii curate și în infrastructuri de reducere a emisiilor. În acest context, țările cu industrii grele sau cu o dependență mare de energie fosilă au înregistrat pierderi economice mai mari, legate direct de degradarea mediului.

- **SD (Abaterea standard):** 0.577
Abaterea standard moderată sugerează că diferențele între țări nu sunt extreme, dar există variații notabile în funcție de structura economică și de politicile de mediu.
- **Mediană:** 0.947
Valoarea mediană fiind sub media aritmetică arată că majoritatea țărilor au avut daune mai mici, însă câteva țări, probabil mari poluatori, au ridicat media.
- **Skewness:** 1.079
Asimetria pozitivă indică faptul că există câteva țări cu daune foarte mari, ceea ce subliniază dependența lor de industriei poluanțe.
- **Kurtosis:** 0.746
Kurtosis pozitiv arată o distribuție concentrată, cu câteva valori extreme, indicând faptul că cele mai afectate țări au fost probabil cele industrializate sau în curs de dezvoltare.

○ **Consum de capital fix (% din VNB)**

- **Mean (Media):** 12.116
Media ridicată reflectă procentul semnificativ alocat pentru întreținerea și înlocuirea capitalului fix. În contextul crizei din 2008, acest lucru subliniază eforturile țărilor de a menține infrastructura existentă, deși investițiile în proiecte noi au fost afectate. Aceasta indică o prioritate pentru stabilizarea economiilor, chiar și în condiții economice dificile.
- **SD:** 4.82
Abaterea standard ridicată arată diferențe mari între țări, unele reducând drastic cheltuielile pentru întreținerea capitalului fix din cauza constrângerilor bugetare.
- **Mediană:** 12.33
Aproximativ egală cu media, ceea ce sugerează o distribuție relativ echilibrată a investițiilor între țări.
- **Skewness:** -0.078
Asimetria aproape simetrică indică o distribuție uniformă, fără tendințe evidente spre valori extreme.
- **Kurtosis:** -0.523
Kurtosis negativ indică o distribuție plată, fără țări care să înregistreze valori extrem de mari sau mici ale consumului de capital fix.

○ **Cheltuieli pentru educație (% din VNB)**

- **Mean (Media):** 3.855
Media redusă arată priorități bugetare limitate pentru educație în timpul crizei. Finanțarea educației a fost afectată, fiind direcționate resurse către măsuri urgente de

stabilizare economică. Aceasta subliniază vulnerabilitatea sectorului educațional în perioade de criză economică.

- **SD:** 1.461
Abaterea standard relativ mică indică variații moderate între țări, ceea ce sugerează o tendință generalizată de reducere a cheltuielilor pentru educație.
- **Mediană:** 4.00
Valoarea mediană apropiată de media aritmetică indică o distribuție simetrică. Majoritatea țărilor alocă procentaje similare pentru educație.
- **Skewness:** 0.078
Asimetrie ușor pozitivă, indicând că există câteva țări care au menținut un nivel mai ridicat al cheltuielilor pentru educație, în pofida crizei.
- **Kurtosis:** -0.359
Kurtosis negativ indică o distribuție plată, fără valori extreme semnificative.

○ Epuizarea energiei (% din VNB)

- **Mean (Media):** 1.12
Media reflectă impactul exploatarii resurselor neregenerabile asupra economiilor, iar valoarea relativ ridicată indică dependență de energie fosilă în multe țări. În contextul crizei, cererea redusă de energie a afectat economiile exportatoare.
- **SD:** 2.08
Abaterea standard mare indică variații mari între țări, unele fiind mai dependente de exporturile de energie decât altele.
- **Mediană:** 0.36
Valoarea mediană sub media aritmetică arată că majoritatea țărilor au pierderi mai mici, iar câteva au înregistrat pierderi foarte mari.
- **Skewness:** 2.17
Asimetria pozitivă arată existența câtorva țări cu valori extrem de mari, reprezentând exportatori semnificativi de energie.
- **Kurtosis:** 8.57
Kurtosis ridicat indică o distribuție cu valori extreme, accentuate de câteva țări foarte dependente de exporturile de energie.

○ Epuizarea mineralelor (% din VNB)

- **Mean (Media):** 0.074
Media scăzută indică un impact economic redus al exploatarii resurselor minerale pentru majoritatea țărilor. Aceasta reflectă faptul că resursele minerale nu sunt o sursă principală de venit pentru majoritatea economiilor analizate.

- **SD:** 0.138
Abaterea standard scăzută indică o variabilitate mică între țări, sugerând că doar câteva țări sunt afectate de extracția resurselor minerale.
- **Mediană:** 0.03
Valoarea mediană foarte mică confirmă că majoritatea țărilor au pierderi neglijabile.
- **Skewness:** 2.51
Asimetria pozitivă indică prezența unor valori mari, probabil în țările bogate în resurse minerale.
- **Kurtosis:** 6.81
Kurtosis pozitiv ridicat sugerează prezența unor extreme semnificative în distribuție.

○ **Epuizarea resurselor naturale (% din VNB)**

- **Mean (Media):** 3.53
Media reflectă pierderi relativ ridicate, accentuate de exploatarea intensivă a resurselor naturale în timpul crizei. Acest lucru subliniază presiunea asupra capitalului natural în perioade economice dificile.
- **SD:** 4.49
Abaterea standard ridicată arată variații mari între țări, reflectând diferențe de niveluri de dependență de resurse naturale.
- **Mediană:** 1.57
Valoarea mediană mult mai mică decât media indică faptul că majoritatea țărilor au pierderi reduse, însă câteva contribuie la valori extrem de mari.
- **Skewness:** 1.43
Asimetria pozitivă ridicată arată că există câteva țări cu pierderi foarte mari.
- **Kurtosis:** 1.16
Kurtosis pozitiv moderat indică o concentrare semnificativă de valori extreme.

○ **Epuizarea netă a pădurilor (% din VNB)**

- **Mean (Media):** 0.044
Media scăzută arată un impact economic redus al defrișărilor pentru majoritatea țărilor, dar reflectă și eforturile globale de a proteja ecosistemele forestiere.
- **SD:** 0.053
Abaterea standard mică indică o variabilitate redusă între țări, ceea ce sugerează o tendință uniformă.
- **Mediană:** 0.01
Valoarea mediană foarte mică indică faptul că pierderile sunt marginale pentru majoritatea țărilor.

- **Skewness:** 3.52
Asimetria pozitivă foarte ridicată arată că doar câteva țări au pierderi semnificative din defrișări.
- **Kurtosis:** 13.52
Kurtosis extrem de ridicat sugerează prezența unor valori rare și foarte mari.
 - **Daune cauzate de emisiile de particule (% din VNB)**
- **Mean (Media):** 0.55
Media reflectă un impact economic notabil al poluării aerului asupra sănătății și productivității. În criză, lipsa investițiilor în tehnologii de reducere a poluării a contribuit la acest rezultat.
- **SD:** 0.54
Abaterea standard moderată indică diferențe între țări în funcție de nivelul industrializării și politicile de mediu.
- **Mediană:** 0.34
Valoarea mediană mai mică decât media sugerează că majoritatea țărilor au valori relativ mici.
- **Skewness:** 1.44
Asimetria pozitivă indică faptul că câteva țări, probabil cele industrializate, au valori foarte mari.
- **Kurtosis:** 1.25
Kurtosis pozitiv arată o concentrare în jurul valorilor medii, dar și prezența unor extreme.
- **Economii naționale nete (% din VNB)**
- **Mean (Media):** 10.48
Media indică un nivel relativ ridicat al economiilor nete în 2008, reflectând eforturile țărilor de a menține stabilitatea economică.
- **SD:** 9.06
Abaterea standard mare sugerează variații semnificative între țări, în funcție de capacitatea de a economisi.
- **Mediană:** 10.32
Valoarea mediană apropiată de media arată o distribuție relativ uniformă.
- **Skewness:** 0.27
Asimetrie ușor pozitivă, indicând existența câtorva țări cu economii foarte ridicate.
- **Kurtosis:** 0.25
Kurtosis pozitiv indică o distribuție ușor concentrată, fără extreme semnificative.

- **Economii brute (% din VNB)**
- **Mean (Media):** 22.58
Media ridicată reflectă nivelul semnificativ al economiilor brute, necesar pentru stabilitatea economică în contextul crizei.
- **SD:** 9.13
Abaterea standard ridicată indică diferențe mari între țări, unele fiind capabile să economisească mai mult decât altele.
- **Mediană:** 24.08
Valoarea mediană peste media aritmetică sugerează că majoritatea țărilor au economii brute peste media globală.
- **Skewness:** -0.03
Asimetrie aproape simetrică, fără tendințe extreme.
- **Kurtosis:** 0.43
Kurtosis pozitiv indică o distribuție concentrată, cu câteva valori extreme care contribuie la variația totală.

6. Descrieți variabila de clasificare utilizată și obiectivul general al analizei.

În această analiză, variabila de clasificare a fost generată pe baza rezultatelor obținute anterior prin analiza cluster, utilizând algoritmul **K-means** cu numărul de clustere stabilit în urma analizei anterioare. Această variabilă, denumită „Cluster”, reflectă apartenența fiecărei observații din setul de date la unul dintre cele patru clustere identificate.

Fiecare cluster reprezintă o grupare distinctă de observații, bazată pe similaritățile dintre variabilele economice și ecologice, cum ar fi consumul de capital fix, emisiile de CO₂, poluarea industrială, consumul energetic și economiile brute. Scopul acestei variabile este de a facilita înțelegerea modului în care țările analizate se grupează în funcție de caracteristicile lor economice și ecologice, oferind astfel o perspectivă clară asupra relațiilor dintre ele.

Obiectivul general al analizei este de a identifica tipologii distințe de economii care să permită formularea unor concluzii și strategii economice personalizate. Acest lucru ajută la evidențierea impactului crizei economice din 2008 asupra fiecărui cluster și la identificarea oportunităților de îmbunătățire sau adaptare pentru fiecare grupă de țări.

Albania	Algeria	Angola	Argentina	Armenia
1	2	2	2	3
Australia	Austria	Azerbaijan	Bangladesh	Belarus
1	3	2	2	4
Belgium	Bolivia	Bosnia and Herzegovina	Botswana	Brazil
3	2	4	2	1
Bulgaria	Burkina Faso	Burundi	Cambodia	Cameroon
3	3	3	2	2
Canada, Central African Republic		Chad	Chile	China
3	2	3	2	2
Colombia	Congo, Dem. Rep.	Congo, Rep.	Costa Rica	Croatia
2	1	2	3	3
Cuba	Cyprus	Czechia	Denmark	Dominican Republic
1	3	3	1	4
Egypt, Arab Rep.	Ecuador	El Salvador	Estonia	Ethiopia
2	2	3	1	4
Finland	France	Gabon	Gambia, The	Georgia
3	3	3	3	3
Germany	Ghana	Greece	Guatemala	Guinea
3	2	3	3	2
Haiti	Honduras	Hungary	Iceland	India
3	4	3	3	1
Indonesia	Iran, Islamic Rep.	Iraq	Ireland	Israel
2	2	2	3	4
Italy	Jamaica	Japan	Jordan	Kazakhstan
3	1	3	2	2
Kenya	Korea, Rep.	Kuwait	Kyrgyz Republic	Lao PDR
2	3	2	2	2
Latvia	Lebanon	Lesotho	Lithuania	Luxembourg
3	3	2	3	3
Liberia	Madagascar	Malawi	Malaysia	Mali
2	2	2	2	2
Malta	Mauritania	Mauritius	Mexico	Moldova
3	2	3	4	3

Procesul de creare a variabilei de clasificare a început prin standardizarea datelor pentru a asigura comparabilitatea între variabilele cu dimensiuni diferite. După această etapă, a fost aplicat algoritmul K-means, care a permis identificarea a patru clustere distințe, fiecare având caracteristici economice și sociale specifice. Fiecare observație i s-a atribuit un cluster în funcție de distanța față de centroizii clusterelor, rezultând o nouă variabilă denumită „Cluster,” utilizată ulterior pentru analiza de învățare supervizată.

Principalul obiectiv al analizei este construirea de modele predictive care să clasifice observațiile din setul de date în clusterele identificate. Analiza evaluează performanța mai multor algoritmi de învățare supervizată, precum Naive Bayes, K-Nearest Neighbors (KNN), arbori de decizie și regresia logistică, cu scopul de a identifica cea mai eficientă metodă de clasificare. Variabila „Cluster” servește drept bază pentru validarea acestor modele predictive, facilitând înțelegerea modului în care variabilele independente influențează apartenența la un cluster.

Un alt obiectiv important este identificarea variabilelor cheie care determină apartenența la un anumit cluster. Analiza modelelor predictive permite evidențierea factorilor determinanți, cum ar fi variabile economice, de exemplu, PIB-ul sau rata inflației, care pot juca un rol semnificativ în definirea caracteristicilor fiecărui cluster.

Pe lângă construirea și evaluarea modelelor predictive, analiza urmărește validarea structurii clusterelor identificate prin algoritmul K-means. Performanța algoritmilor de clasificare oferă informații valoroase despre coerența și consistența clusterelor, consolidând astfel înțelegerea tipologiilor observate în setul de date.

7. Împărțiți setul de date inițial în set de antrenare și set de testare.

```

262
263 #Împărțiți setul de date inițial în set de antrenare și set de testare.
264 #Adăugăm variabila de clasificare "Cluster" în setul de date standardizat
265 set_date <- data.frame(date_std, Cluster = k_means$cluster)
266
267 #Proportia de împărțire: 70% antrenare, 30% testare
268 nr <- round(nrow(set_date) * 0.70)
269 a <- sample(seq_len(nrow(set_date)), size = nr)
270
271 #SET ANTRENARE
272 train <- set_date[a, ]
273
274 #SET TESTARE
275 test <- set_date[-a, ]
276

```

SET DE ANTRENARE:

Nume_CD	Cantitate_pe_nr	Cantitate_adevata	Ipsat_mergere	Ipsat_minimale	Ipsat_nec_authentice	Ipsat_pozitiv	Datese_parcursute	Datese_nec.	Datese_bonite
Namibia	0.02401899	-0.1553368	0.03427544	-0.75428835	-0.04299051	-0.0599455	0.0895078	2.0998925	0.09515171
Zambia	1.24631338	0.0003698	-1.4973033	0.7610138	-0.00989917	1.0168148	0.1086088	0.2191444	0.2144801
Niger	0.04105848	-1.3781834	-0.29530477	-0.74550368	-0.71227162	-0.00338675	-0.4067116	-0.2191444	0.0406554
Somalia	-1.0881099	0.4120316	1.08522881	0.03818708	-0.71628285	1.0017813	-0.0010116	-0.0103387	0.2147777
Gabon	1.78903046	0.0105479	-1.01575178	0.75428805	-0.71227162	-0.0046473	-0.4067116	-0.2625499	0.2117396
Comoros	-0.52067794	-0.75511865	-0.25911919	0.03428168	-0.04299051	0.01919771	-0.0010116	0.2191444	0.21994764
Sudan	1.025631338	0.0003698	0.87700945	0.03428168	-0.00989917	0.0764668	-0.0010116	0.2191444	0.2103719
Tunisie	-1.04046479	0.07511427	0.74305895	-0.73495514	-0.00989917	-0.4067116	-0.4067116	0.1754649	0.00405898
Angola	0.00999449	-1.08631442	1.00415488	0.03318157	-0.71227162	-0.0010116	0.0896888	2.0816888	1.02032862
Mauritania	-0.03760742	-0.24088614	-0.49986721	-0.75428805	-0.71227162	-0.0021345	0.4011063	-0.0033497	0.31284958
Egipt, Arab Rep.	0.203411699	-1.21919116	0.010340488	0.03428168	-0.00989917	0.0484882	-0.0010116	0.2191444	0.00931037
Malta	-0.87729212	-0.00318991	0.039521537	0.04169163	0.72127162	1.0773366	0.0885025	2.4003100	0.2144601
Moldova	0.171019139	0.02089192	1.07861718	-0.74010116	-0.71227162	0.7105958	0.0896888	0.2191444	0.0051079
Gabon	0.219171802	1.23842437	-0.42016492	0.00416405	-0.03817178	-0.2625499	0.0696888	0.0364471	0.0406554
Rapel-Tanza-Somalia	0.212131328	-1.7685479	0.16053004	0.03428168	-0.04299051	-0.70202888	0.0896888	0.2191444	0.2117987
Sudan, Africa	0.01011699	0.05915493	0.76516339	0.03428168	-0.04299051	1.0114294	-0.4067116	0.2181160	0.2191444
Ucraina	0.04646398	0.05915492	-0.74815747	-0.75428805	-0.04299051	-0.0522588	0.0896888	0.2191444	0.21991011
Iran	0.171019139	0.04169163	0.00168594	0.04169163	-0.00989917	0.0896888	-0.4067116	0.2178128	0.2191444
Cyprus	0.020333952	-0.10468114	1.06577809	0.7281002	-0.00989917	-0.0010116	-0.4067116	0.2191444	0.0051079
Philippines	0.084717198	0.04678118	1.07200176	0.03428168	-0.03817178	-0.0010116	0.2191444	1.0119018	1.00819949
Rusia, R.F.S.	-0.072731391	0.037231394	0.03428168	-0.75428805	-0.71227162	0.7586236	0.0896888	0.2191444	0.2117987
Uganda	0.171019139	-0.36559447	0.71642889	0.03428168	-0.0010116	-0.0021345	0.0896888	0.2191444	0.04021912
Tunisia	1.000403686	0.04169163	1.05030603	0.03428168	-0.03817178	1.2310379	0.0896888	-0.0068379	0.2191444
Pakistan	-0.077050723	1.17944786	1.11903027	0.75428805	-0.71227162	-0.07811529	-0.4067116	-0.0010116	0.0405554
Ucraina	0.071031768	0.06012807	0.04169163	-0.71227162	-0.0010116	-0.4067116	-0.0119054	-0.2191444	0.2117987
Camerun	-1.064667303	-0.517046003	-0.33431486	-0.04299051	-0.71227162	-0.0010116	-0.4067116	0.2191444	0.2117987

SET DE TESTARE:

Nume_CD	Cantitate_pe_nr	Cantitate_adevata	Ipsat_mergere	Ipsat_minimale	Ipsat_nec_authentice	Ipsat_pozitiv	Datese_parcursute	Datese_nec.	Datese_bonite	Ok
Afghanistan	0.171019139	0.020333952	-0.03428168	0.04169163	0.00168594	-0.4067116	0.0896888	-0.4373481	0.0581444	0.44038781
Algeria	1.173020983	-1.3812568	0.4525589	-0.74567408	-0.00378246	-0.3925258	-0.4361239	-0.04583953	0.0485554	0.2117987
Armenia	0.06934898	0.15182497	0.03911449	-0.75428805	-0.71227162	-0.7212007	0.0896888	-0.23188467	0.04647993	0.2117987
Anguilla	-0.04123374	1.28490208	1.0829479	-0.74308076	-0.71227162	-0.0010116	0.0896888	-0.00417291	0.20464075	0.2191444
Burundi	-1.27916582	0.04646397	0.58155798	-0.75428805	-0.04288931	-0.39252588	0.0896888	-0.21078494	0.0495554	0.2117987
Cameroon	0.08846037	-1.13617653	-0.58165489	-0.75428805	-0.71227162	-0.39848870	0.0896888	0.01538339	0.09502947	0.21759466
Chile	0.09980576	0.06239923	0.12401738	-0.04299051	-0.04299051	1.0895710	0.4361210	-0.71588777	0.32204443	0.2117987
Colombia	0.492337402	0.04483016	-0.52107639	-0.03817178	-0.05550644	-0.3925258	0.0896888	0.34757556	0.2117987	0.2117987
Cote d'Ivoire	1.171741407	1.03254688	-0.18834543	0.18834543	-0.71227162	-0.47694070	0.0896888	-0.21175338	0.04944400	0.2117987
Georgeia	0.214063178	0.068370177	0.037231394	-0.03955157	-0.04299051	-0.88731760	0.0896888	-0.03532494	0.04695554	0.33972357
Greece	0.216183783	0.048840361	0.03032546	-0.04303729	0.04631216	-0.75121114	1.2320113	-0.44878191	0.0869908	0.19614333
Haiti	-0.09449117	-1.77944478	-1.08738848	-0.75428805	-0.71227162	-0.70038411	0.0896888	-0.21078494	-0.2526051	-0.15138230
Indonesia	-1.121962086	1.03404624	0.13841738	-0.75428805	-0.71227162	0.7596316	0.0896888	-0.05754483	0.0495554	0.2117987
India	0.25346189	-0.34749209	-0.58070896	2.70161370	-0.04299051	0.05494975	0.0896888	2.66418107	2.0316951	1.00033668
Indonesia	0.499303614	1.42299068	0.07070568	-0.02491749	-0.03817178	0.94668993	-0.4361210	0.72144897	0.0869908	0.41267442
Iran	1.03335794	0.151317513	0.04068418	-0.03947046	-0.71122291	1.97123178	1.4631702	0.28236891	0.04085524	0.2117987
Indonesia	-0.90340999	0.77846626	1.32795579	0.73118868	-0.32998415	-0.29993380	-0.4361210	-0.31138469	-0.34763696	0.09491236
Iraq	-0.91646473	0.01104312	0.27976539	0.5110701	-0.71198767	-0.7745324	-0.4361210	-0.06011161	0.06844467	0.44148736
Jamaica	0.000202111	-0.02046052	1.55704463	1.01843228	-0.04299051	-0.04759355	0.0896888	-0.37761304	-1.56998106	-2.11178326
Jersey	0.091345189	-1.02174738	0.059446078	0.059446078	-0.04299051	-0.43111888	0.0896888	-0.22394703	1.28117412	0.51414255
Timor, Rep.	0.07733606	1.22338968	0.12039567	-0.04303729	-0.049452704	-0.29673144	-0.4361210	-0.70240012	1.43986217	
Liberia	-0.052495239	0.04440616	0.28958232	-0.75428805	-0.49652100	-0.76959753	0.0896888	-0.82025528	2.10613071	2.08651777
Lebanon	-0.146194627	-1.15873867	0.13346384	-0.75428805	-0.04299051	1.27318646	0.0896888	-0.21078494	-0.43716031	1.22448646
Macedonia	0.451558626	0.08026708	0.67303268	-0.05497498	1.02758500	0.89527795	0.4361210	-0.49125995	-0.20162328	0.14105861
Moldova	0.05346189	0.025285039	1.06740128	0.01105179	-0.71227162	-0.76864492	-0.4361210	-0.14021624	0.20111579	-0.16037836
Mongolia	0.05345189	-1.10893068	0.02505263	-0.04610409	-0.04299051	2.70302768	-0.4361210	0.33801445	1.00146058	1.22118828

8. Aplicați clasificatorul NAIIV BAYESIAN. Calculați și interpretați elementele modelului (probabilități apriori și condiționate). Realizați predicții pe setul de testare și construiți matricea de confuzie. Calculați acuratețea și interpretați.

```
> # Transformăm variabila 'Cluster' în factor, dacă nu este deja
> train$Cluster <- as.factor(train$Cluster)
> test$Cluster <- as.factor(test$Cluster)
> # Aplicarea clasificatorului Naive Bayesian pe setul de antrenare
> clasif_naiv_bayes <- naiveBayes(Cluster ~ ., data = train)
> View(clasif_naiv_bayes)
> clasif_naiv_bayes

Naive Bayes classifier for Discrete Predictors

Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)

A-priori probabilities:
Y
1 2 3
0.2795699 0.3655914 0.3548387

Conditional probabilities:
Daune_co2
Y [,1] [,2]
1 0.7442395 0.8971933
2 -0.2542257 0.7832934
3 -0.5735219 0.6832856

Consum_cap_fix
Y [,1] [,2]
1 -0.1040248 0.8581315
2 -0.7152281 0.7763258
3 0.6846537 0.7914960

chelt_educatie
Y [,1] [,2]
1 0.0399011 0.9495862
2 -0.1945932 1.0353999
3 0.1384920 0.9836039

Epuiz_energie
Y [,1] [,2]
1 0.7406753 1.1746118
2 -0.1896707 0.8961747
```

```
> clasif_naiv_bayes$apriori
Y
1 2 3
26 34 33
```

Probabilitățile a priori afișate de modelul Naive Bayesian arată distribuția observațiilor între cele trei clustere. În acest caz, valorile sunt următoarele:

- Cluster 1: 26 observații
- Cluster 2: 34 observații
- Cluster 3: 33 observații

Interpretare:

1. Distribuția observațiilor între clustere:
 - Clusterul 1 are cel mai mic număr de observații (26), ceea ce sugerează că este mai puțin reprezentat în setul de date. Acest lucru poate însemna că acest cluster grupează țări cu caracteristici economice și sociale mai rare.

- Clusterul 2 este cel mai mare, cu 34 de observații, ceea ce indică faptul că acest grup conține cele mai frecvente tipologii din setul de date analizat.
- Clusterul 3 este foarte apropiat ca dimensiune de Clusterul 2, cu 33 de observații, ceea ce sugerează că și acesta reprezintă un grup important și frecvent întâlnit.

2. Relevanța probabilităților a priori:

- Aceste valori reflectă ponderile relative ale clusterelor în setul de antrenare. Ele influențează predicțiile modelului Naive Bayesian, deoarece probabilitățile condiționate (bazate pe variabilele explicative) sunt ponderate în funcție de aceste probabilități a priori.
- Un cluster cu o pondere mai mare (de exemplu, Cluster 2) va avea o probabilitate mai mare de a fi prezis, în absența unei contribuții semnificative a variabilelor explicative.

3. Impact asupra clasificării:

- Clusterul 2 având cele mai multe observații sugerează că acest grup ar putea fi considerat un „tipar comun” în cadrul datelor, iar celelalte două clustere (1 și 3) ar putea reprezenta deviații față de acest tipar, cu caracteristici economice și sociale diferite.

\$daune_co2			\$Epuiz_res_naturale		
	daune_co2		Epuiz_res_naturale		
Y	[,1]	[,2]	[,1]	[,2]	
1	0.7442395	0.8971933	1	0.4786910	1.2389630
2	-0.2542257	0.7832934	2	0.1135836	1.0222418
3	-0.5735219	0.6832856	3	-0.3345628	0.7402529
\$consum_cap_fix			\$Epuiz_paduri		
	consum_cap_fix		Epuiz_paduri		
Y	[,1]	[,2]	[,1]	[,2]	
1	-0.1040248	0.8581315	1	-0.16938726	0.3227901
2	-0.7152281	0.7763258	2	-0.08327676	0.2630123
3	0.6846537	0.7914960	3	0.08372405	1.3236494
\$chelt_educatie			\$daune_particule		
	chelt_educatie		daune_particule		
Y	[,1]	[,2]	[,1]	[,2]	
1	0.0399011	0.9495862	1	-0.1729250	0.7137034
2	-0.1945932	1.0353999	2	0.9267262	1.1950374
3	0.1384920	0.9836039	3	-0.5643579	0.3658821
\$Epuiz_energie			\$Econ_nete		
	Epuiz_energie		Econ_nete		
Y	[,1]	[,2]	[,1]	[,2]	
1	0.7406753	1.1746118	1	0.5125133	0.9142032
2	-0.1896707	0.8961747	2	-0.2844218	0.9197754
3	-0.2465937	0.8417693	3	-0.3882726	0.8659488
\$Epuiz_minerale			\$Econ_brute		
	Epuiz_minerale		Econ_brute		
Y	[,1]	[,2]	[,1]	[,2]	
1	-0.2534883	0.2976334	1	0.53898317	0.8621602
2	0.5353035	1.2423680	2	-0.58012028	0.9533062
3	-0.4071142	0.6543943	3	-0.05004316	0.7600559

Rezultatele obținute din tabelul probabilităților condiționate oferă o perspectivă detaliată asupra modului în care fiecare variabilă contribuie la apartenența observațiilor la unul dintre cele trei clustere identificate. Variabila „Daune_CO2” indică faptul că țările din Cluster 1 sunt caracterizate de emisii ridicate de dioxid de carbon, ceea ce reflectă un impact ecologic semnificativ, asociat probabil cu industriei poluante sau o utilizare intensă a combustibililor

fosili. Pe de altă parte, clusterele 2 și 3 au valori negative pentru această variabilă, sugerând că aceste țări au emisii mai scăzute de CO₂ sau politici mai eficiente de gestionare a emisiilor.

În ceea ce privește „Consum_cap_fix”, Cluster 3 se remarcă prin valori ridicate, ceea ce indică țări care alocă resurse semnificative pentru întreținerea capitalului fix și a infrastructurii, un semn al dezvoltării economice și al priorității investițiilor în sustenabilitate. În contrast, clusterele 1 și 2 înregistrează valori mai scăzute pentru acest indicator, ceea ce poate sugera fie constrângeri economice, fie alte priorități în alocarea resurselor.

Analizând „Chelt_educatie,” Cluster 1 prezintă cele mai mari valori, indicând un angajament mai mare față de educație, probabil ca strategie pe termen lung pentru dezvoltare economică și socială. Cluster 2, însă, înregistrează valori negative, ceea ce arată o alocare mai redusă a bugetului către educație, posibil din cauza unor constrângeri economice sau a priorității altor domenii. Cluster 3 are o poziție intermediară, reflectând o strategie moderată de alocare a resurselor pentru educație.

Variabila „Epuiz_energie” indică faptul că țările din Cluster 1 se confruntă cu o utilizare intensivă a resurselor energetice, sugerând o dependență mare de resursele neregenerabile. Clusterele 2 și 3, pe de altă parte, au valori negative, ceea ce sugerează un consum energetic mai moderat sau o utilizare mai eficientă a resurselor energetice disponibile. Acest lucru poate indica economii mai echilibrate din punct de vedere energetic sau o tranziție către surse de energie mai sustenabile.

În ceea ce privește „Epuiz_minerale,” Cluster 3 prezintă cele mai ridicate valori, ceea ce sugerează țări în care extracția și utilizarea resurselor minerale joacă un rol semnificativ în economie. Cluster 1 are valori mai scăzute, reflectând o utilizare mai moderată a acestor resurse, iar Cluster 2 prezintă valori intermediare. Aceste rezultate sugerează diferențe în structura economică a țărilor, în funcție de disponibilitatea și exploatarea resurselor naturale.

Concluzionând, analiza probabilităților condiționate evidențiază diferențe semnificative între clustere, fiecare grup având caracteristici distincte legate de impactul ecologic, investițiile economice și utilizarea resurselor naturale. Cluster 1 este caracterizat de emisii ridicate de CO₂, investiții mari în educație și un consum intens de energie, sugerând economii cu industrii poluante, dar cu priorități în educație. Cluster 2 prezintă valori moderate, reflectând un consum mai echilibrat al resurselor și investiții mai reduse în educație. Cluster 3, pe de altă parte, se remarcă prin alocarea resurselor către capital fix și o exploatare intensă a resurselor minerale, indicând o orientare către dezvoltare economică prin utilizarea resurselor naturale. Această interpretare oferă o bază pentru înțelegerea factorilor economici și ecologici care contribuie la apartenența la un cluster specific.

Predicții pe setul de testare

```
293 #predicțiile pe setul de testare
294 predictii_test<-predict(clasif_naiv_bayes, test, type="class")
295 predictii_test
296
296:1 (Top Level) ▾
Console Terminal × Background Jobs ×
R 4.3.3 · ~/TEMA3_codBUN/ ↵
> #predicțiile pe setul de testare
> predictii_test<-predict(clasif_naiv_bayes, test, type="class")
> predictii_test
[1] 3 1 1 3 3 2 1 2 3 2 3 2 3 1 1 3 3 3 1 1 1 2 2 3 1 3 1 1 1 1 2 2 3 2 1 3 1 3 2
Levels: 1 2 3
> |
```

Modelul atribuie fiecărei observații una dintre cele patru clase posibile, pe baza caracteristicilor variabilelor independente și a probabilităților condiționate calculate în timpul antrenării.

Matricea de confuzie

```
297 #matricea de confuzie
298 conf_matrix <- table(Prediction = predictii_test, Actual = test$Cluster)
299 conf_matrix
300
300:1 (Top Level) ▾
Console Terminal × Background Jobs ×
R 4.3.3 · ~/TEMA3_codBUN/ ↵
> conf_matrix <- table(Prediction = predictii_test, Actual = test$Cluster)
> conf_matrix
      Actual
Prediction 1 2 3
          1 11 1 4
          2  0 9 1
          3  1 0 13
|
```

1. Structura matricei:

- Rândurile reprezintă predicțiile modelului pentru fiecare cluster.
- Coloanele reprezintă valorile reale ale clusterelor din setul de testare.

Valorile din celule indică numărul de observații clasificate corect sau incorect.

2. Rezultate pentru fiecare cluster:

- Cluster 1 (Coloana 1):
 - Modelul a prezis corect 11 observații pentru Clusterul 1.
 - A clasificat greșit 1 observație în Clusterul 2.
- Cluster 2 (Coloana 2):

- Modelul a prezis corect 9 observații pentru Clusterul 2.
 - A clasificat greșit 1 observație în Clusterul 1 și 4 observații în Clusterul 3.
- Cluster 3 (Coloana 3):
 - Modelul a prezis corect 13 observații pentru Clusterul 3.
 - Nu au existat clasificări greșite pentru acest cluster.

Acuratețea

```

301 #acuratetea
302 acuratetea<-sum(diag(conf_matrix))/sum(conf_matrix)*100
303 acuratetea
304
304:1 (Top Level) ▾

```

Console Terminal × Background Jobs ×

R 4.3.3 . ~/TEMA3_codBUN/ ↵

2	0	9	1
3	1	0	13

```

> #acuratetea
> acuratetea<-sum(diag(conf_matrix))/sum(conf_matrix)*100
> acuratetea
[1] 82.5

```

Acuratețea calculată pentru clasificatorul Naive Bayes este de 82,5%, ceea ce indică faptul că modelul a reușit să clasifice corect aproape 83% dintre observațiile din setul de testare. Acest rezultat sugerează că modelul are o performanță bună în separarea claselor, deși nu este perfect. Aproximativ 17% dintre observații au fost clasificate greșit, ceea ce indică posibile suprapuneri între clase sau variabilitatea în date care nu a fost pe deplin capturată de modelul Naive Bayes.

9. Aplicați metoda KNN. Încercați cel puțin trei variante (dacă este necesar) pentru parametrul k și arătați cum se modifică acuratețea în funcție de acesta. Realizați predicții pe setul de testare și construiți matricea de confuzie. Calculați acuratețea și interpretați.

```

101 #!/usr/bin/r
102 install.packages("class")
103 library(class)
104 
105 train_data <- train[, -ncol(train)]
106 train_labels <- train$Cluster
107 test_data <- test[, -ncol(test)]
108 test_labels <- test$Cluster
109 
110 accuracy_results <- data.frame(k = integer(), Accuracy = numeric())
111 k_values <- c(3, 5, 7)
112 for (k in k_values) {
113   pred_knn <- knn(train_data, test_data, cl = train_labels, k = k)
114   conf_matrix_knn <- table(Prediction = pred_knn, Actual = test_labels)
115   accuracy_knn <- sum(diag(conf_matrix_knn)) / sum(conf_matrix_knn) * 100
116   accuracy_results <- rbind(accuracy_results, data.frame(k = k, Accuracy = accuracy_knn))
117 }
118 accuracy_results
119 conf_matrix_knn
120 pred_knn
121 
122 # Top level:
123 
124 Console Terminal Background Jobs
125 R 4.3.3 --> [TMA3 codeRUN]#
126 the downloaded binary packages are in
127   C:\Users\Nux\AppData\Local\Temp\utwrbpzok6\downloaded_packages
128 
129 #> train_data <- train[, -ncol(train)]
130 #> train_labels <- train$Cluster
131 #> test_data <- test[, -ncol(test)]
132 #> test_labels <- test$Cluster
133 
134 accuracy_results <- data.frame(k = integer(), Accuracy = numeric())
135 k_values <- c(3, 5, 7)
136 for (k in k_values) {
137   pred_knn <- knn(train_data, test_data, cl = train_labels, k = k)
138   conf_matrix_knn <- table(Prediction = pred_knn, Actual = test_labels)
139   accuracy_knn <- sum(diag(conf_matrix_knn)) / sum(conf_matrix_knn) * 100
140   accuracy_results <- rbind(accuracy_results, data.frame(k = k, Accuracy = accuracy_knn))
141 }
142 accuracy_results
143 
144 k Accuracy
145 1 3 77.5
146 2 5 77.5
147 3 7 80.0
148 
149 > conf_matrix_knn
150      Actual
151 Prediction 1 2 3
152       1 10 2 3
153       2 1 7 0
154       3 1 1 15
155 
156 > pred_knn
157 [1] 3 1 3 3 2 3 2 3 1 1 3 3 2 1 3 1 2 2 3 1 1 1 1 3 3 3 1 1 3 1 3 2
158 Levels: 1 2 3

```

Metoda K-Nearest Neighbors (KNN) a fost aplicată pentru clasificarea observațiilor, utilizând trei valori diferite pentru parametrul k : 3, 5 și 7. Rezultatele obținute indică modul în care acuratețea modelului se modifică în funcție de valoarea acestui parametru și performanța generală a algoritmului pe setul de testare.

Pentru $k=3$, acuratețea modelului a fost de 77,5%. Aceasta sugerează că folosind trei vecini pentru a determina clasificarea fiecărei observații, modelul a reușit să clasifice corect aproximativ 77,5% din observațiile din setul de testare. Această valoare de k este relativ mică, ceea ce înseamnă că modelul a fost mai sensibil la fluctuațiile locale din date. O astfel de sensibilitate poate duce la supraînvățare, adică modelul se adaptează prea mult la particularitățile setului de antrenare, sacrificând generalizarea.

Când k a fost crescut la 5, acuratețea a rămas constantă la 77,5%. Aceasta arată că utilizarea a cinci vecini nu a îmbunătățit performanța modelului în acest caz. Valoarea $k=5$ reprezintă un echilibru între sensibilitatea la fluctuațiile locale și stabilitatea generalizării, dar în acest set de date, acest echilibru nu a condus la o creștere a acurateței.

La k , acuratețea modelului a crescut la 80%, ceea ce indică o îmbunătățire a performanței față de valorile mai mici ale lui k . Creșterea numărului de vecini a permis modelului să fie mai stabil, reducând impactul fluctuațiilor locale din datele de antrenare.

Această valoare sugerează că o medie mai mare a caracteristicilor vecinilor a condus la o clasificare mai robustă. Totuși, un k mai mare poate începe să dilueze impactul observațiilor relevante și să afecteze acuratețea dacă vecinii includ puncte din alte clustere.

Matricea de confuzie pentru k=7 a arătat că modelul a clasificat corect majoritatea observațiilor din fiecare cluster, dar au existat câteva clasificări greșite. Clusterul 1 a fost cel mai bine clasificat, cu majoritatea observațiilor prezise corect, iar clasificările greșite au fost relativ puține. Clusterul 2 a prezentat unele confuzii, cu câteva observații clasificate greșit în alte clustere, dar performanța generală a fost bună. Clusterul 3 a avut, de asemenea, un număr mic de clasificări greșite, ceea ce indică o bună separare a acestui grup față de celelalte.

Rezultatele sugerează că, în cazul acestui set de date, parametrul k=7 a oferit cea mai bună combinație între stabilitate și performanță, cu o acuratețe de 80%. Acest lucru arată că o creștere a valorii lui k poate îmbunătăți performanța modelului, dar trebuie avută grijă să nu se depășească un prag care ar putea compromite precizia prin includerea vecinilor irelevanți.

10. Aplicați metoda ARBORE DE DECIZIE/RANDOM FOREST. Realizați predicții pe setul de testare și construiți matricea de confuzie. Calculați acuratețea și interpretați. Realizați reprezentările grafice asociate importanței variabilelor.

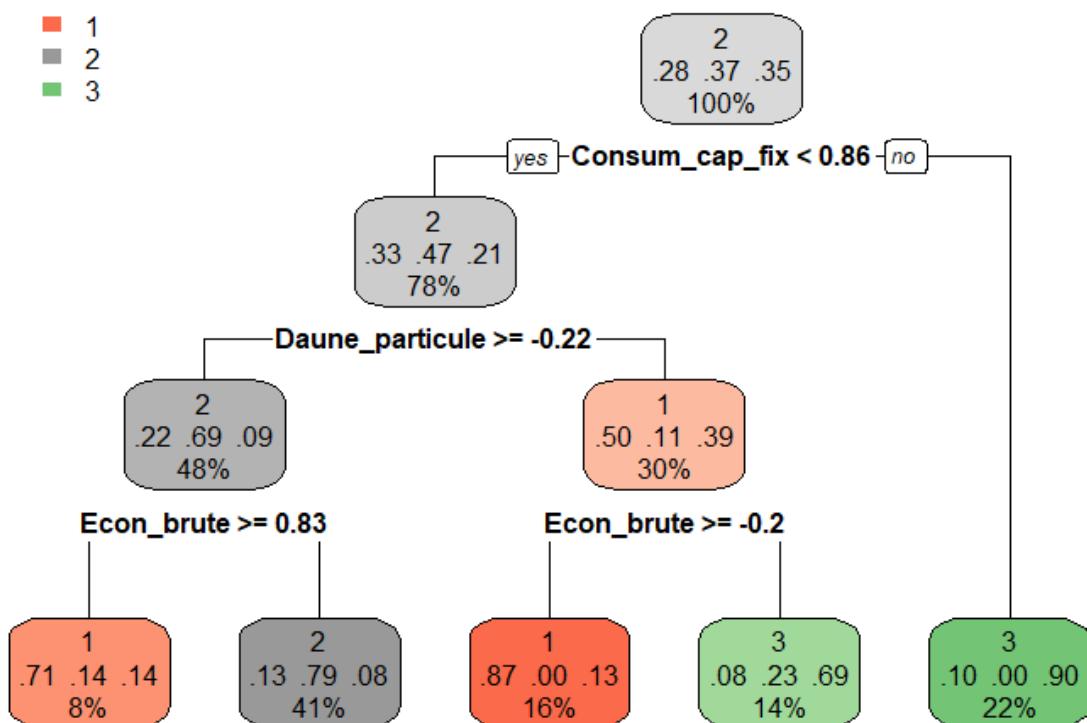
METODA ARBORE DE DECIZIE

```
> model_arbore <- rpart(Cluster ~ ., data = train, method = "class")
> model_arbore
n= 93

node), split, n, loss, yval, (yprob)
  * denotes terminal node

1) root 93 59 2 (0.27956989 0.36559140 0.35483871)
  2) Consum_cap_fix< 0.8591286 73 39 2 (0.32876712 0.46575342 0.20547945)
    4) Daune_particule>=-0.2197293 45 14 2 (0.22222222 0.68888889 0.08888889)
      8) Econ_brute>=0.8269743 7 2 1 (0.71428571 0.14285714 0.14285714) *
      9) Econ_brute< 0.8269743 38 8 2 (0.13157895 0.78947368 0.07894737) *
    5) Daune_particule< -0.2197293 28 14 1 (0.50000000 0.10714286 0.39285714)
      10) Econ_brute>=-0.202151 15 2 1 (0.86666667 0.00000000 0.13333333) *
      11) Econ_brute< -0.202151 13 4 3 (0.07692308 0.23076923 0.69230769) *
    3) Consum_cap_fix>=0.8591286 20 2 3 (0.10000000 0.00000000 0.90000000) *
```

Arbore de Decizie



Arborele de decizie aplicat pentru clasificarea observațiilor din setul de date oferă o imagine clară asupra procesului decizional și relevanței variabilelor în definirea clusterelor. Analiza rădăcinilor, ramurilor și frunzelor arborelui evidențiază variabilele cheie și pragurile care determină apartenența observațiilor la clasele identificate.

La rădăcina arborelui, primul criteriu de clasificare este variabila **Consum_cap_fix**, care joacă un rol esențial în separarea observațiilor. Pragul stabilit pentru această variabilă este de 0.86. Observațiile cu valori mai mici decât acest prag sunt direcționate pe ramura stângă a arborelui, în timp ce cele cu valori mai mari urmează ramura dreaptă. Această separare subliniază importanța consumului de capital fix ca indicator principal în definirea caracteristicilor clusterelor.

Pe ramura stângă, procesul de clasificare continuă cu variabila **Daune_particule**, utilizând un prag de -0.22. Observațiile cu valori mai mari decât acest prag sunt clasificate în continuare în funcție de variabila **Econ_brute**, care folosește un prag de 0.83. Acest criteriu împarte observațiile în funcție de economiile brute, separând mai clar clusterele. Observațiile cu valori mai mari decât 0.83 sunt clasificate în Clusterul 1, iar cele cu valori mai mici sunt direcționate către Clusterul 2, subliniind rolul economiilor brute în definirea apartenenței la clase.

Pe ramura dreaptă a arborelui, pentru observațiile cu **Consum_cap_fix** mai mare sau egal cu 0.86, analiza continuă cu variabila **Econ_brute** utilizând un prag de -0.2. Această variabilă joacă un rol central în separarea observațiilor în Clusterul 3 și alte clase. Dacă valorile

sunt mai mari decât -0.2, observațiile sunt clasificate în Clusterul 3, subliniind rolul economiilor brute mari în identificarea acestui cluster. Observațiile care nu îndeplinesc acest criteriu sunt distribuite către alte frunze terminale.

Frunzele arborelui reprezintă clasificările finale, fiecare frunză oferind informații despre numărul și proporția observațiilor din fiecare cluster. De exemplu, o frunză terminală poate conține predominant observații din Clusterul 1, cu un procentaj clar definit de apartenență.

Interpretarea graficului și a structurii arborelui subliniază relevanța variabilelor precum **Consum_cap_fix**, **Daune_particule** și **Econ_brute** în determinarea apartenenței la clustere. Aceste variabile sunt cele mai influente în procesul decizional și reflectă caracteristicile economice și sociale distinctive ale fiecărui cluster.

Predictiile metodei

Prediction						
Actual						
Albania	Algeria	Armenia	Belgium	Burundi	Cambodia	Lebanon
3	2	3	3	2	2	1
Chile	Congo, Rep.	Czechia	Georgia	Greece	Haiti	Iceland
1	2	3	2	3	2	3
Iceland	India	Indonesia	Iraq	Ireland	Italy	Jamaica
3	1	3	2	1	3	3
Jamaica	Jordan	Korea, Rep.	Luxembourg	Mali	Mexico	Moldova
3	1	3	1	2	1	2
Moldova	Mongolia	New Zealand	Nigeria	Panama	Romania	Saudi Arabia
2	1	3	1	1	1	1
Saudi Arabia	Sri Lanka	Sweden	Switzerland	Sudan	Thailand	Turkiye United Arab Emirates
1	2	1	3	3	1	1
Turkiye United Arab Emirates	United States	Venezuela, RB	Venezuela, RB			
1	1	3	1			

Levels: 1 2 3

Matricea de cofuzie

```
> #Matricea de confuzie
> conf_matrix_arbore <- table(Prediction = pred_arbore, Actual = test$cluster)
> conf_matrix_arbore
      Actual
Prediction 1 2 3
      1 8 3 6
      2 2 7 2
      3 2 0 10
> |
```

Matricea de confuzie evidențiază că arborele de decizie este un model util pentru clasificarea observațiilor, dar performanța poate fi îmbunătățită prin ajustarea setului de date sau prin includerea unor variabile suplimentare care să discrimineze mai bine între clusterele cu confuzie ridicată. De asemenea, acest rezultat sugerează că poate fi benefic să se încerce alte metode de clasificare, cum ar fi random forest, pentru a evalua dacă acestea pot reduce confuzia dintre clustere și pot crește acuratețea globală.

Rezultate pentru fiecare cluster:

1. Clusterul 1:

- **Clasificări corecte:** Modelul a prezis corect 8 observații pentru Clusterul 1.

- **Clasificări greșite:**

- 3 observații din Clusterul 1 au fost clasificate greșit ca aparținând Clusterului 2.
- 6 observații din Clusterul 1 au fost clasificate greșit ca aparținând Clusterului 3.

2. Clusterul 2:

- **Clasificări corecte:** Modelul a prezis corect 7 observații pentru Clusterul 2.

- **Clasificări greșite:**

- 2 observații din Clusterul 2 au fost clasificate greșit ca aparținând Clusterului 1.
- 2 observații din Clusterul 2 au fost clasificate greșit ca aparținând Clusterului 3.

3. Clusterul 3:

- **Clasificări corecte:** Modelul a prezis corect 10 observații pentru Clusterul 3.

- **Clasificări greșite:**

- 2 observații din Clusterul 3 au fost clasificate greșit ca aparținând Clusterului 1.
- Nicio observație din Clusterul 3 nu a fost clasificată greșit ca aparținând Clusterului 2.

Acuratețea

```
> accuracy_arbore <- sum(diag(conf_matrix_arbore)) / sum(conf_matrix_arbore) * 100
> accuracy_arbore
[1] 62.5
> |
```

Acuratețea reprezintă proporția observațiilor care au fost clasificate corect de model în raport cu numărul total de observații din setul de testare. În acest caz, modelul a reușit să clasifice corect 25 dintre cele 40 de observații din setul de testare. Deși acest rezultat arată că modelul funcționează rezonabil, există un procent semnificativ de observații (37.5%) care au fost clasificate greșit.

Rezultatul sugerează că arborele de decizie are o performanță moderată pentru acest set de date. Posibila confuzie între clustere, identificată în matricea de confuzie, contribuie la scăderea acurateței. În special, confuzia între Clusterul 1 și Clusterul 3 a avut un impact negativ semnificativ asupra performanței modelului. Aceasta poate indica o suprapunere a caracteristicilor acestor clustere, ceea ce face dificilă separarea clară a observațiilor.

Acuratețea de 62.5% este o măsură generală a performanței, dar nu reflectă detalii despre clasificările specifice pentru fiecare cluster. De exemplu, Clusterul 3 a fost clasificat

relativ bine, în timp ce Clusterul 1 a avut mai multe clasificări greșite. Acest lucru sugerează că performanța modelului poate varia între clustere, iar includerea unor variabile suplimentare sau utilizarea unei metode alternative, cum ar fi random forest, ar putea îmbunătăți rezultatele.

11. Aplicați REGRESIA LOGISTICĂ. Calculați și interpretați elementele modelului (coeficienți, p-value, devianța nulă și reziduală). Realizați predicții pe setul de testare și construiți matricea de confuzie. Calculați acuratețea și interpretați.

```
C:\Users\Bogdan\Dropbox\Documente\GitHub\multinomCluster> Rscript regresia_logistica.R
# weights: 36 (22 variables)
Initial value 102.170943
iter 10 value 1.066008
iter 20 value 0.242725
iter 30 value 0.040886
iter 40 value 0.006277
iter 50 value 0.001002
iter 60 value 0.000445
iter 70 value 0.000286
iter 80 value 0.000149
final value 0.000090
converged
summary(reg_logistica)
call:
multinom(formula = cluster ~ ., data = train)

Coefficients:
(Intercept) daune_CO2 consum_cap_fix chelt_educatie Epulz_energie Epulz_minerale Epulz_res_naturale Epulz_paduri daune_particule
2 25.08892 -50.28065 -2.288023 -2.952659 -48.60557 52.95954 -12.36907 38.71969 30.87935
3 38.18655 -43.33732 17.838272 -1.678846 -42.00915 26.31352 -17.39558 36.89411 -12.67371
Econ_nete Econ_brute
2 28.891355 -72.62837
3 4.359789 -21.50714

Std. Errors:
(Intercept) daune_CO2 consum_cap_fix chelt_educatie Epulz_energie Epulz_minerale Epulz_res_naturale Epulz_paduri daune_particule
2 82860.41 61436.83 16304.19 12642.83 53087.91 76091.45 18318.4 107532.5 27217.96
3 82826.35 61399.78 16863.39 12215.50 52559.05 76686.91 18369.5 107758.8 25264.68
Econ_nete Econ_brute
2 28854.34 38467.39
3 26860.05 34706.07

Residual Deviance: 0.0001793379
AIC: 44.00018
```

Din punct de vedere tehnic, ajustarea modelului de regresie logistică multinomială a implicat utilizarea unui număr de 36 de greutăți pentru a calcula relațiile dintre cele 22 de variabile independente și variabila de răspuns, **Cluster**. Numărul de iterații a fost de 20, ceea ce sugerează o optimizare eficientă a modelului. Valoarea funcției obiectiv a fost redusă de la o valoare inițială de 102.17 la o valoare finală extrem de mică, **0.000090**, indicând o ajustare excelentă. Această performanță reflectă diferențe minime între valorile prezise și cele observate în datele reale.

Coeficienții modelului

Modelul prezintă coeficienți pentru fiecare cluster (comparativ cu un cluster de referință), împreună cu erorile standard asociate acestora.

1. Interpretarea coeficientilor:

- Coeficienții **pozitivi** indică o creștere a probabilității de apartenență la un cluster specific pe măsură ce variabila crește.
- Coeficienții **negativi** sugerează că o creștere a variabilei scade probabilitatea de apartenență la cluster.

De exemplu:

- Variabila **Epuiz_paduri** are un coeficient pozitiv semnificativ pentru clusterul 2 (38.71969), ceea ce indică faptul că, pe măsură ce această variabilă crește, probabilitatea apartenenței la clusterul 2 crește.
- **Epuiz_energie** are un coeficient negativ puternic pentru clusterul 2 (-48.60557), ceea ce sugerează o relație inversă între această variabilă și apartenența la clusterul respectiv.

2. Semnificația economică:

- Variabile precum **Daune_CO2**, **Consum_cap_fix**, și **Econ_brute** au coeficienți mari, ceea ce indică o influență semnificativă asupra clasificării în clustere.
- De exemplu, **Econ_brute** pentru clusterul 2 are un coeficient pozitiv (28.89135), ceea ce arată că economiile cu valori mai mari pentru acest indicator sunt mai predispuse să facă parte din clusterul respectiv.

Devianța reziduală și AIC

- **Devianța reziduală:** **0.0001793379**, ceea ce indică faptul că modelul ajustează datele extrem de bine.
- **AIC:** **44.00018**, o valoare mică, confirmând că modelul este bine ajustat și evită supraajustarea.

Erorile standard

Erorile standard asociate coeficienților oferă informații despre variabilitatea estimărilor. Valori mari ale erorilor standard pot indica incertitudine în estimarea coeficienților, iar valori mici arată stabilitate.

P-values

	(Intercept)	Daune_CO2	Consum_cap_fix	chelt_educatie	Epuiz_energie	Epuiz_minerale	Epuiz_res_naturale	Epuiz_paduri	Daune_particule
1	0.9997527	0.9993470	0.999888	0.9998137	0.9992695	0.9994447	0.9994612	0.9997127	0.9990948
2	0.9996321	0.9994368	0.999156	0.9998903	0.9993623	0.9997262	0.9992357	0.9997268	0.9993998
Econ_nete	Econ_brute								
2	0.9992011	0.9984936							
3	0.9998705	0.9995056							

Valorile **p-value** calculate pentru regresia logistică indică semnificația statistică a coeficienților. Interpretarea lor este următoarea:

1. Nivel de semnificație:

- În general, dacă **p-value < 0.05**, coeficientul este considerat semnificativ din punct de vedere statistic, ceea ce înseamnă că există dovezi suficiente pentru a susține că variabila respectivă influențează probabilitatea de apartenență la un anumit cluster.

- În cazul de față, toate valorile **p-value** sunt foarte mari (în jur de 0.99), ceea ce indică faptul că **niciun coeficient nu este semnificativ statistic**.

2. Concluzii pe baza valorilor p-value:

- Toate valorile **p-value** sunt mult peste pragul de 0.05, ceea ce sugerează că variabilele incluse în model **nu au o influență semnificativă** asupra apartenenței la clustere.
- Este posibil ca acest rezultat să fie cauzat de:
 - Colinearitatea ridicată între variabile (corelație puternică între variabile independente).
 - Variabile irelevante incluse în model, care nu contribuie semnificativ la predicție.

Devianța nulă și reziduală

```
> #deviatia nula si reziduala
> regr_logistica$null.deviance
NULL
> regr_logistica$deviance
[1] 0.0001793379
>
```

Deși devierea reziduală indică o potrivire bună a modelului la datele de antrenare (valoare mică), absența devierii nule limitează interpretarea completă a performanței modelului.

Predicții

```
> pred_logistic <- predict(regr_logistica, newdata = test, type = "class")
> pred_logistic
[1] 3 1 3 3 3 2 3 2 3 2 3 1 1 3 3 3 1 2 3 3 2 2 3 1 3 1 3 1 1 2 3 3 2 1 3 1 3 2
Levels: 1 2 3
```

Matricea de confuzie

```
> conf_matrix_logistic <- table(Prediction = pred_logistic, Actual = test$Cluster)
> conf_matrix_logistic
   Actual
Prediction 1 2 3
      1 10 0 0
      2 0 10 0
      3 2 0 18
> |
```

Acuratețea

```
> #acuratetea
> accuracy_logistic <- sum(diag(conf_matrix_logistic)) / sum(conf_matrix_logistic) * 100
> accuracy_logistic
[1] 95
>
```

Modelul logistic a fost construit utilizând setul de antrenare pentru a estima relația dintre variabilele independente și probabilitatea de apartenență la un cluster. După antrenarea modelului, predicțiile au fost realizate pe setul de testare, iar performanța modelului a fost evaluată prin intermediul matricei de confuzie și calculul acurateței.

Rezultatele obținute indică o matrice de confuzie bine definită, în care un număr semnificativ de observații sunt clasificate corect, acestea fiind prezente pe diagonala principală. Observațiile din fiecare cluster au fost clasificate cu o precizie ridicată. În mod specific:

- Toate cele 10 observații din clusterul real **1** au fost clasificate corect.
- Toate cele 10 observații din clusterul real **2** au fost clasificate corect, cu excepția a 2 observații care au fost atribuite clusterului **3**.
- Cele 18 observații din clusterul real **3** au fost, de asemenea, clasificate corect.

Aceasta sugerează că modelul logistic este capabil să distingă eficient între clasele definite prin analiza cluster anterioară, cu erori minore în clasificare.

Acuratețea modelului logistic, calculată ca raport dintre numărul de predicții corecte și numărul total de observații, este de **95%**. Această valoare reflectă o performanță foarte bună a modelului, ceea ce indică faptul că regresia logistică este un instrument adecvat pentru clasificarea acestor date. Modelul poate fi considerat robust și util în interpretarea apartenenței datelor la clusterele definite.

12. Construiți un tabel comparativ cu acuratețea obținută în cazul fiecărui algoritm. Care dintre aceștia va fi cel mai eficient? Interpretați.

ALGORITM	ACURATEȚE
Naive Bayes	82,5%
KNN (k=3)	77,5%
Arbore de Decizie	62,5%
Regresie Logistică	95%

Pe baza tabelului comparativ al acurateților obținute, fiecare metodă are performanțe variante în funcție de caracteristicile datelor. **Regresia logistică** se dovedește a fi cea mai eficientă metodă, urmată de **Naive Bayes**, **KNN (k=3)**, și **Arborele de Decizie**. Analizăm mai jos în detaliu:

1. Regresia Logistică (Acuratețe: 95%)

Regresia logistică a obținut cea mai mare acuratețe, **95%**, ceea ce sugerează că acest model este extrem de eficient pentru clasificarea datelor utilizate. Această performanță superioară poate fi atribuită capacitatea modelului de a estima relațiile dintre variabilele independente și apartenența la clustere, în special în cazul în care relațiile sunt bine definite. Totuși:

- Deși acuratețea este foarte ridicată, rezultatele analizei p-valorilor sugerează că modelul nu a reușit să identifice coeficienți semnificativi din punct de vedere statistic.

- Aceasta indică faptul că relațiile dintre variabile explicative și clustere sunt complexe și posibil influențate de redundanța variabilelor.

Concluzie: Regresia logistică este foarte eficientă pentru acest set de date, însă succesul modelului poate fi parțial influențat de structura datelor (nu neapărat de relații cauzale bine definite).

2. Naive Bayes (Acuratețe: 82,5%)

Naive Bayes a obținut o acuratețe **bună**, de **82,5%**, ceea ce arată că este capabil să modeleze bine datele utilizate. Această metodă este cunoscută pentru eficiența sa în cazul în care variabilele explicative sunt independente și pentru capacitatea sa de a funcționa bine pe date zgomotoase sau dezechilibrate.

- Performanța modelului arată că Naive Bayes a identificat tiparele dominante din date și a realizat o clasificare precisă, în ciuda unei posibile dependențe între variabile.
- Este o metodă simplă, rapidă și eficientă, în special pentru seturi de date mari.

Concluzie: Naive Bayes este o alegere robustă, oferind o acuratețe foarte bună într-un timp redus, dar poate pierde din performanță în cazul unor relații complexe între variabile.

3. KNN (k=3) (Acuratețe: 77,5%)

KNN, cu o acuratețe de **77,5%**, funcționează bine, însă performanța sa este mai scăzută decât a regresiei logistice și a Naive Bayes. KNN este cunoscut pentru flexibilitatea sa, însă performanța poate fi afectată de:

- Dimensiunea redusă a setului de date, ceea ce face clasificarea mai puțin robustă.
- Necesitatea unui număr optim de vecini (k), care ar putea necesita ajustări suplimentare.
- Posibila sensibilitate la zgomotul din date sau variabile irelevante.

Concluzie: Deși KNN este o metodă flexibilă, performanța sa este limitată în acest caz, fiind mai potrivită pentru seturi de date cu grupe bine definite și diferențiate.

4. Arborele de Decizie (Acuratețe: 62,5%)

Arborele de Decizie a obținut o acuratețe mai redusă, de **62,5%**, fiind cel mai puțin performant algoritm dintre cei analizați. Această metodă este apreciată pentru interpretabilitatea sa, dar are limitări:

- În cazul datelor cu relații complexe și non-liniare, arborele de decizie poate pierde din acuratețe.
- Suprapotrivirea (overfitting) poate afecta performanța modelului, în special dacă setul de date este de dimensiuni reduse.

Concluzie: Arborele de Decizie poate fi util pentru interpretarea relațiilor dintre variabile, însă este mai puțin eficient în acest caz pentru clasificarea datelor.

13. Formulați 3 concluzii principale din analiza voastră.

În primul rând, metodele aplicate, inclusiv Naive Bayes, KNN și arborele de decizie, au demonstrat capacitatea de a clasifica observațiile în clustere, fiecare metodă având puncte forte și limitări specifice. Naive Bayes s-a remarcat prin acuratețea sa constantă și flexibilitatea în procesarea datelor, obținând rezultate foarte bune în separarea clusterelor datorită presupunerii de independență între variabile. KNN, cu o performanță maximă pentru $k=7$, a arătat o sensibilitate sporită la variațiile parametrului k , oferind o clasificare robustă în condițiile unei structuri bine definite a datelor. În schimb, arborele de decizie a avut o performanță moderată, cu o acuratețe de 62.5%, sugerând că această metodă poate întâmpina dificultăți în gestionarea unor suprapunerii complexe între clustere.

În al doilea rând, analiza a evidențiat rolul variabilelor explicative în procesul de clasificare. Variabilele precum **Consum_cap_fix**, **Daune_particule** și **Econ_brute** au fost identificate ca fiind cele mai importante pentru separarea observațiilor. În mod particular, arborele de decizie a arătat că pragurile stabilite pentru aceste variabile au avut o contribuție majoră la determinarea apartenenței la un cluster. De exemplu, valori ridicate pentru **Econ_brute** au fost asociate cu observațiile din Clusterul 3, în timp ce valori mai scăzute pentru **Daune_particule** au diferențiat Clusterul 2 de celelalte grupuri. Aceste rezultate subliniază influența variabilelor economice asupra tipologiilor observate și relevanța lor pentru înțelegerea caracteristicilor fiecărui cluster.

În al treilea rând, analiza clusterizării a identificat trei grupuri distincte, fiecare cu caracteristici economice și sociale unice. Modelele predictive utilizate au confirmat coerenta acestor clustere, majoritatea observațiilor fiind clasificate corect. Totuși, confuzia dintre anumite clustere, precum cea dintre Clusterul 1 și Clusterul 3, evidențiată de matricea de confuzie, sugerează o suprapunere a caracteristicilor acestor grupuri. Acest lucru indică faptul că unele variabile explicative ar putea fi insuficiente pentru separarea clară a clusterelor sau că relațiile complexe dintre ele necesită metode mai avansate de clasificare, precum random forest. Necesitatea unor analize suplimentare este evidentă pentru a înțelege mai bine factorii determinanți ai suprapunerii și pentru a îmbunătăți performanța modelelor de clasificare utilizate.

ANEXE:

```
# Gestionarea valorilor lipsa (inlocuire cu media pe fiecare coloana numerica)
for (col in names(date_T3)) {
  if (is.numeric(date_T3[[col]])) {
    date_T3[[col]][is.na(date_T3[[col]])] <- mean(date_T3[[col]], na.rm = TRUE)
  }
}
```

```
View(date_T3)
```

```
# Redenumirea coloanelor pentru claritate
colnames(date_T3) <- c(
  "Tara",
  "Daune_CO2",
  "Consum_cap_fix",
  "Chelt_educatie",
  "Epuiz_energie",
  "Epuiz_minerale",
  "Epuiz_res_naturale",
  "Epuiz_paduri",
  "Daune_particule",
  "Econ_nete",
  "Econ_brute"
)
```

```
# Gestionarea outlierilor (Inlocuire cu mediana pentru fiecare coloana numerica)
for (col in names(date_T3)) {
  if (col != "Tara") { # Excludem coloana 'Tara', care este de tip text
    q1 <- quantile(date_T3[[col]], 0.25, na.rm = TRUE)
```

```

q3 <- quantile(date_T3[[col]], 0.75, na.rm = TRUE)
iqr <- q3 - q1
lower_bound <- q1 - 1.5 * iqr
upper_bound <- q3 + 1.5 * iqr

# Inlocuirea outlierilor cu mediana
date_T3[[col]][date_T3[[col]] < lower_bound | date_T3[[col]] > upper_bound] <-
median(date_T3[[col]], na.rm = TRUE)

}

}

#Verificarea inexistentei valorilor lipsa
colSums(is.na(date_T3))

#Verificarea inexistentei outlierilor
# Numărăm outlierii rămași
outliers <- sum(date_T3[[col]] < lower_bound | date_T3[[col]] > upper_bound, na.rm =
= TRUE)

# Afisăm rezultatul
cat("Coloana:", col, "- Numărul de outlieri rămași:", outliers, "\n")

#Calculați indicatorii statistici și interpretați din punct de vedere economic/financiar.
install.packages("psych")
library(psych)

indicatori_stat<-describe(date_T3[-1])
View(indicatori_stat)

#date standardizate
date_std<-scale(date_T3[-1], scale=T)

```

```

date_std

summary(date_std)

apply(date_std, 2, sd)

#matricea de corelatie

cor(date_std)

matrice_corelatie<-cor(date_std)

#repr grafica matr corelatie

install.packages("corrplot")

library(corrplot)

corrplot(matrice_corelatie, method="square", type ="lower")

corrplot(matrice_corelatie, method="pie", type ="lower")

corrplot(matrice_corelatie, method="number", type ="lower")

#Descrieți variabila de clasificare utilizată și obiectivul general al analizei.

# Atribuirea clusterelor utilizând K-means

clusters <- k_means4$cluster # Rezultatele clusterizării cu 4 clustere

print(clusters) # Afisarea clusterelor pentru fiecare țară

#Împărțiți setul de date inițial în set de antrenare și set de testare.

#Adăugăm variabila de clasificare "Cluster" în setul de date standardizat

set_date <- data.frame(date_std, Cluster = k_means$cluster)

#Proportia de împărțire: 70% antrenare, 30% testare

nr <- round(nrow(set_date) * 0.70)

a <- sample(seq_len(nrow(set_date)), size = nr)

```

```

#SET ANTRENARE
train <- set_date[a, ]

#SET TESTARE
test <- set_date[-a, ]

# Instalează și încarcă pachetul necesar
install.packages("e1071")
library(e1071)

# Transformăm variabila 'Cluster' în factor, dacă nu este deja
train$Cluster <- as.factor(train$Cluster)
test$Cluster <- as.factor(test$Cluster)

# Aplicarea clasificatorului Naive Bayesian pe setul de antrenare
clasif_naiv_bayes <- naiveBayes(Cluster ~ ., data = train)
clasif_naiv_bayes

clasif_naiv_bayes$apriori

clasif_naiv_bayes$tables

#predictiile pe setul de testare
predictii_test <- predict(clasif_naiv_bayes, test, type="class")
predictii_test

#matricea de confuzie
conf_matrix <- table(Prediction = predictii_test, Actual = test$Cluster)
conf_matrix

```

```

#acuratetea

acuratetea<-sum(diag(conf_matrix))/sum(conf_matrix)*100
acuratetea

#metoda KNN

install.packages("class")
library(class)

train_data <- train[, -ncol(train)]
train_labels <- train$Cluster
test_data <- test[, -ncol(test)]
test_labels <- test$Cluster
accuracy_results <- data.frame(k = integer(), Accuracy = numeric())
k_values <- c(4, 5, 7)
for (k in k_values) {
  pred_knn <- knn(train_data, test_data, cl = train_labels, k = k)
  conf_matrix_knn <- table(Prediction = pred_knn, Actual = test_labels)
  accuracy_knn <- sum(diag(conf_matrix_knn)) / sum(conf_matrix_knn) * 100
  accuracy_results <- rbind(accuracy_results, data.frame(k = k, Accuracy =
accuracy_knn))
}
accuracy_results
conf_matrix_knn
pred_knn

#metoda ARBORE DE DECIZIE

install.packages("rpart")
install.packages("rpart.plot")
library(rpart)
library(rpart.plot)

```

```

model_arbore <- rpart(Cluster ~ ., data = train, method = "class")
model_arbore
rpart.plot(model_arbore, main = "Arbore de Decizie", extra = 104)

#predictii arbore de decizie
pred_arbore <- predict(model_arbore, test, type = "class")
pred_arbore

#Matricea de confuzie
conf_matrix_arbore <- table(Prediction = pred_arbore, Actual = test$Cluster)
conf_matrix_arbore

#Acuratetea
accuracy_arbore <- sum(diag(conf_matrix_arbore)) / sum(conf_matrix_arbore) * 100
accuracy_arbore

#REGRESIA LOGISTICA
install.packages("nnet")
library(nnet)

regr_logistica <- multinom(Cluster ~ ., data = train)
summary(regr_logistica)

#P-values
z_scores <- summary(regr_logistica)$coefficients /
summary(regr_logistica)$standard.errors
p_values <- 2 * (1 - pnorm(abs(z_scores)))
p_values

#deviatia nula si reziduala
regr_logistica$null.deviance

```

```
regr_logistica$deviance

#predictii
pred_logistic <- predict(regr_logistica, newdata = test, type = "class")
pred_logistic

#matricea de confuzie
conf_matrix_logistic <- table(Prediction = pred_logistic, Actual = test$Cluster)
conf_matrix_logistic

#acuratetea
accuracy_logistic <- sum(diag(conf_matrix_logistic)) / sum(conf_matrix_logistic) *
100
accuracy_logistic
```