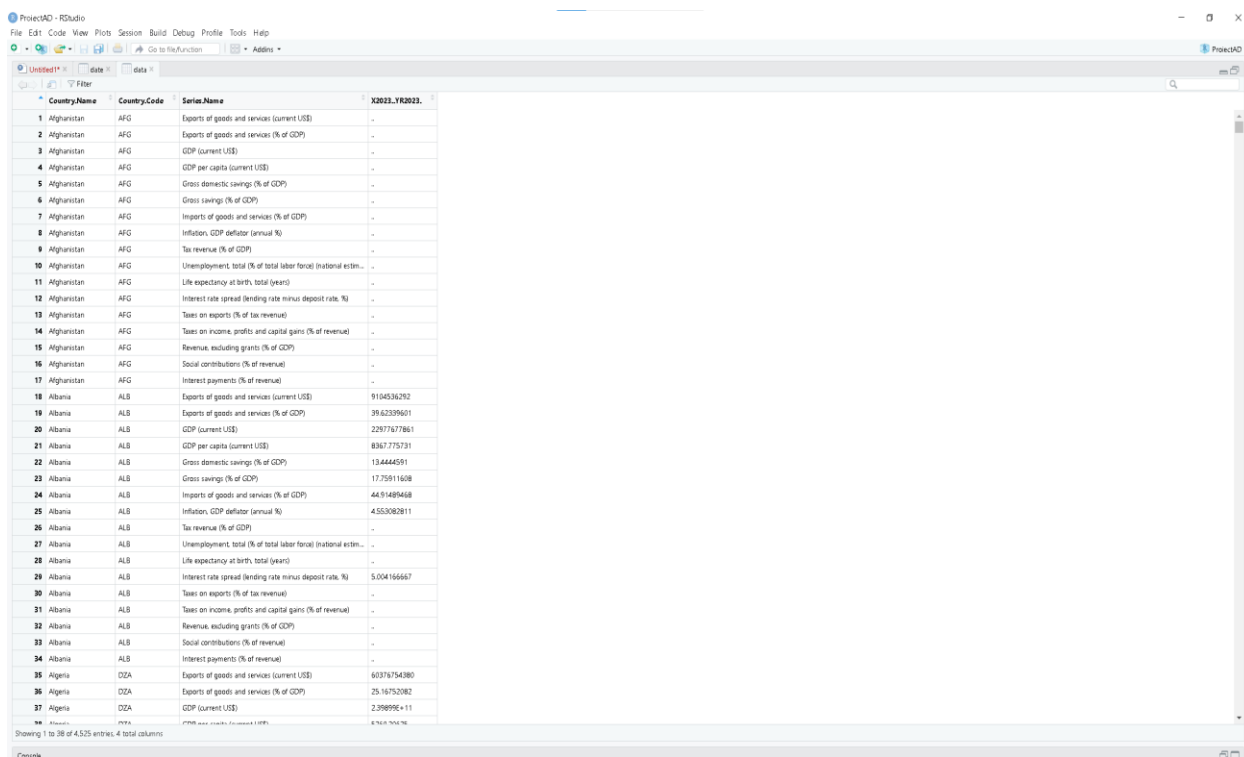


Prezentare Proiect Analiza Datelor (TEMA 3)

- COVLEA RADU-ALIN
Grupa 1080

1. In acest proiect am ales sa analizez relatiile dintre anumiti indicatori economici la nivel global.



The screenshot displays the ProjectAD - RStudio application window. The main area shows a data table with the following columns: CountryName, CountryCode, SeriesName, and X2023..YR2023.. The table lists economic indicators for Afghanistan (AFG), Albania (ALB), and Algeria (DZA). The indicators include Exports of goods and services, GDP, GDP per capita, Gross domestic savings, Gross savings, Imports of goods and services, Inflation, GDP deflator, Tax revenue, Unemployment, Life expectancy, Interest rate spread, Taxes on exports, Taxes on income, profits and capital gains, Revenue, excluding grants, Social contributions, and Interest payments. The table is sorted by CountryName and then by SeriesName. The status bar at the bottom indicates 'Showing 1 to 38 of 4,525 entries. 4 total columns'.

CountryName	CountryCode	SeriesName	X2023..YR2023..
1 Afghanistan	AFG	Exports of goods and services (current US\$)	-
2 Afghanistan	AFG	Exports of goods and services (% of GDP)	-
3 Afghanistan	AFG	GDP (current US\$)	-
4 Afghanistan	AFG	GDP per capita (current US\$)	-
5 Afghanistan	AFG	Gross domestic savings (% of GDP)	-
6 Afghanistan	AFG	Gross savings (% of GDP)	-
7 Afghanistan	AFG	Imports of goods and services (% of GDP)	-
8 Afghanistan	AFG	Inflation, GDP deflator (annual %)	-
9 Afghanistan	AFG	Tax revenue (% of GDP)	-
10 Afghanistan	AFG	Unemployment, total (% of total labor force) (national estimate)	-
11 Afghanistan	AFG	Life expectancy at birth, total (years)	-
12 Afghanistan	AFG	Interest rate spread (lending rate minus deposit rate, %)	-
13 Afghanistan	AFG	Taxes on exports (% of tax revenue)	-
14 Afghanistan	AFG	Taxes on income, profits and capital gains (% of revenue)	-
15 Afghanistan	AFG	Revenue, excluding grants (% of GDP)	-
16 Afghanistan	AFG	Social contributions (% of revenue)	-
17 Afghanistan	AFG	Interest payments (% of revenue)	-
18 Albania	ALB	Exports of goods and services (current US\$)	9104536292
19 Albania	ALB	Exports of goods and services (% of GDP)	39.62339601
20 Albania	ALB	GDP (current US\$)	22977677861
21 Albania	ALB	GDP per capita (current US\$)	8367.775731
22 Albania	ALB	Gross domestic savings (% of GDP)	13.44444591
23 Albania	ALB	Gross savings (% of GDP)	17.75911608
24 Albania	ALB	Imports of goods and services (% of GDP)	44.91489468
25 Albania	ALB	Inflation, GDP deflator (annual %)	4.553082811
26 Albania	ALB	Tax revenue (% of GDP)	-
27 Albania	ALB	Unemployment, total (% of total labor force) (national estimate)	-
28 Albania	ALB	Life expectancy at birth, total (years)	-
29 Albania	ALB	Interest rate spread (lending rate minus deposit rate, %)	5.004166667
30 Albania	ALB	Taxes on exports (% of tax revenue)	-
31 Albania	ALB	Taxes on income, profits and capital gains (% of revenue)	-
32 Albania	ALB	Revenue, excluding grants (% of GDP)	-
33 Albania	ALB	Social contributions (% of revenue)	-
34 Albania	ALB	Interest payments (% of revenue)	-
35 Algeria	DZA	Exports of goods and services (current US\$)	60376754380
36 Algeria	DZA	Exports of goods and services (% of GDP)	25.16752082
37 Algeria	DZA	GDP (current US\$)	2.39899E+11

Asa arata setul de date initial inainte de a fi prelucrat.

2. Am inceput prelucrarea prin eliminarea valorilor duplicate in coloanele “Country.Name”, “Country.Code” si “Series.Name” daca acestea exista. (nu a fost cazul)

Mai departe am aplicat conceptul de pivotare, intrucat setul de date nu respecta regula “fiecare variabila se refera la un singur obiect” si am transformat observatiile din coloana “Series.Name” in variabile, intrucat acestea fac obiectul analizei ulterioare.

	nume_tara	cod_tara	exporturi	exporturi (%gdp)	pib	pib/nr_loc	economie_interna (%gdp)	consum_interna (%gdp)	investit_publice (%gdp)	inflatie_consum	inflatie_pib	investit_straine (%gdp)	rate_soma	urbanizare	investit_entere	NA.1	NA.2
1	Afghanistan	AFG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	Albania	ALB	9104536292	39.62399601	22977677861	8367.775731	13.6644591	17.75911608	44.91489468	4.553082811	-	-	-	5.004166667	-	-	-
35	Algeria	DZA	60376754380	25.16752082	2.39899E+11	5260.29625	41.87642996	41.2259377	21.02837305	-2.258254272	-	-	-	6.25	-	-	-
52	American Samoa	ASM	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
66	Andorra	AND	-	-	3727673593	46544.72072	-	-	-	-	-	-	-	5.859995553	-	-	-
86	Angola	AGO	33781306111	39.8726733	8472295262	2395.52142	-	-	-	-	-	-	-	11.6475693	-	-	-
103	Anguilla and Barbuda	ATG	-	-	2033089185	21660.31939	-	-	-	-	-	-	-	4.85647573	-	-	-
120	Argentina	ARG	82760724362	12.91943661	6.40591E+11	13730.51471	17.04933302	15.12142697	14.05299192	133.8398673	-	6.139	-	0.897421188	-	-	-
137	Armenia	ARM	14086972330	56.18145547	24212134631	8715.765336	20.64503862	19.17896607	58.8615499	2.83212316	-	-	-	4.022573807	-	-	-
154	Aruba	ABW	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
171	Australia	AUS	4.60641E+11	26.72198648	1.72383E+12	64711.7456	29.02766032	24.81177276	21.36069104	6.507375485	-	3.668	-	-	-	-	-
188	Austria	AUT	3.06923E+11	59.47718589	5.16034E+11	56505.96828	27.94707042	27.05366709	56.64198122	7.60957137	-	5.264	-	-	-	-	-
205	Azerbaijan	AZE	35487352941	49.04336789	72356176471	7155.083654	32.77588889	29.81959553	34.57398246	-9.20663945	-	-	-	5.311920499	-	-	-
222	Bahamas, The	BHS	5699000000	39.74613802	14338500000	34749.63829	15.9850333	-	42.9584846	6.344420145	-	8.643	-	3.711666667	-	-	-
239	Bahrain	BHR	-	-	43205000000	29084.30713	-	-	-	-	-	-	-	-	-	-	-
256	Bangladesh	BGD	57551451741	13.15716383	4.37415E+11	25.29.080127	25.75951408	34.90435988	17.82580758	6.89712404	-	-	-	1.051666667	-	-	-
273	Barbados	BBB	-	-	6393564190	22672.61544	-	-	-	-	-	-	-	4.758828343	-	-	-
290	Belarus	BLR	48014396058	66.81901876	71857382746	7829.053137	27.20042499	25.12014494	66.17977856	7.368230536	-	3.461	-	5.110018951	-	-	-
307	Belgium	BEL	5.48013E+11	86.68123029	6.32371E+11	53475.2935	24.39571889	25.26096668	87.61379575	4.079874107	-	5.528	-	-	-	-	-
324	Belize	BLZ	1795950000	53.6187719	3281600000	7987.589165	27.06079337	27.61142969	32.85692519	10.89614793	-	-	-	6.365	-	-	-
341	Benin	BDN	4169446538	21.15345399	1967328468	1434.662834	31.50750787	-	30.15701102	3.385464352	-	-	-	-	-	-	-
358	Bermuda	BMU	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
375	Bhutan	BTN	-	-	-	-	-	-	-	-	-	-	-	3.126	-	-	-
392	Bolivia	BOL	12507803619	27.28012912	45849832956	3700.578338	13.11118716	13.48458453	31.56913918	1.699553215	-	3.034	-	5.46436066	-	-	-
409	Bosnia and Herzegovina	BH	11939049957	44.12899198	27054889363	8426.09111	8.191426509	18.15246839	56.89710957	5.530501969	-	10.668	-	2.936145228	-	-	-
426	Botswana	BWA	61413897614	31.6635491	19395765126	7249.799326	28.1114262	34.48093249	35.71008683	2.125860244	-	23.381	-	4.753325926	-	-	-
443	Brazil	BRA	3.50657E+11	18.11027542	2.17367E+12	10043.62392	18.4525996	15.58974482	15.74366717	4.05894759	-	7.547	-	31.46123433	-	-	-
460	British Virgin Islands	VCB	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
477	Brunei Darussalam	BRN	11578413348	76.53488336	15128292954	33430.91848	49.61833345	45.96239131	60.02817372	-12.89619583	-	5.296	-	5.307729167	-	-	-
494	Bulgaria	BGR	61847387271	60.88276999	1.01584E+11	15797.95558	22.01543135	20.11737728	57.71731775	7.510536602	-	4.319	-	4.35875	-	-	-
511	Burkina Faso	BFA	5864763345	28.85504619	20334617839	874.1212804	15.44611371	-	35.96020997	1.998045499	-	5.348	-	-	-	-	-
528	Burundi	BDI	1396471444	5.333835236	2642161689	195.3807027	4.079996485	11.00740759	24.3798264	11.37639148	-	-	-	-	-	-	-
545	Cabo Verde	CPV	1027691030	39.72133377	2987253076	4331.578864	4.850097671	20.80085899	54.14983739	3.857299077	-	-	-	6.768370992	-	-	-
562	Cambodia	KHM	23241413029	73.14889556	31772759999	1875.071482	50.10875287	57.09754889	40.38951118	2.384035721	-	-	-	-	-	-	-
579	Cameroon	CMR	8774028437	18.3	47945510390	1673.648889	15.4	-	20.3	2.736099551	-	-	-	-	-	-	-
596	Canada	CAN	7.17708E+11	33.53643807	2.14019E+12	53371.69743	23.544330329	21.84062202	33.52683079	1.60524181	-	5.415	-	-	-	-	-
613	Cayman Islands	CYM	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

De asemenea, am schimbat si numele coloanelor pentru ca nu erau deloc lizibile si nu ma puteam referi la ele, dupa cum se poate observa in output-ul de mai jos.

> names(date)	"Country.Name"
[1] "X2023..YR2023..Exports of goods and services (current US\$)"	"Country.Code"
[3] "X2023..YR2023..GDP (current US\$)"	"X2023..YR2023..Exports of goods and services (% of GDP)"
[5] "X2023..YR2023..GDP (current US\$)"	"X2023..YR2023..GDP per capita (current US\$)"
[7] "X2023..YR2023..Gross domestic savings (% of GDP)"	"X2023..YR2023..Gross savings (% of GDP)"
[9] "X2023..YR2023..Imports of goods and services (% of GDP)"	"X2023..YR2023..Inflation, GDP deflator (annual %)"
[11] "X2023..YR2023..Tax revenue (% of GDP)"	"X2023..YR2023..Unemployment, total (% of total labor force) (national estimate)"
[13] "X2023..YR2023..Life expectancy at birth, total (years)"	"X2023..YR2023..Interest rate spread (lending rate minus deposit rate, %)"
[15] "X2023..YR2023..Taxes on exports (% of tax revenue)"	"X2023..YR2023..Taxes on income, profits and capital gains (% of revenue)"
[17] "X2023..YR2023..Revenue, excluding grants (% of GDP)"	"X2023..YR2023..Social contributions (% of revenue)"
[19] "X2023..YR2023..Interest payments (% of revenue)"	"X2023..YR2023.."
> names(date) <- c(
+ "nume_tara",	
+ "cod_tara",	
+ "exporturi",	
+ "exporturi (%pib)",	
+ "pib",	
+ "pib/nr_loc", #pib per numaral de locuitori	
+ "economie_interna (%pib)",	

Am observat ca in setul de date se afla la final si o analiza a indicatorilor la nivel de continent si regiuni, insa eu vreau sa fac o analiza bazata doar pe tari, asadar am eliminat toate acele observatii si am eliminat si coloanele redundante care au fost generate, precum “NA.1”, “NA.2” etc.

Am facut aceste schimbari prin referinta la index.

ProjectAD - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Add

data

	name_tara	cod_tara	exporturi	exporturi (\$kpb)	pib	pib/nc_jec	economie_interna (\$kpb)	consum_interna (\$kpb)	investit_publica (\$kpb)	inflatie_consum	inflatie_pib	investit_straini (\$kpb)	rata_somaj	urbanizare	investit_externe
1	Afghanistan	AFG	-	-	-	-	-	-	-	-	-	-	-	-	-
2	Albania	ALB	9104536292	3562339601	2297377861	8367.775731	134444591	17.75911608	44.91489468	4.553082811	-	-	-	5.004166667	-
3	Algeria	DZA	60376754380	25.16752082	239899E+11	5260.20625	41.87642996	41.2259377	21.02837305	-2.258254272	-	-	-	6.25	-
4	American Samoa	ASM	-	-	-	-	-	-	-	-	-	-	-	-	-
5	Andorra	AND	-	-	3727673593	46544.72072	-	-	-	5.859905503	-	-	-	-	-
6	Angola	AGO	3378130811	39.8726733	84722957642	2309.52162	39.82419499	30.17871709	26.66147599	17.64775693	-	-	-	10.04341747	-
7	Antigua and Barbuda	ATG	-	-	2033085185	21560.21533	-	-	-	4.805475773	-	-	-	5.337422	-
8	Argentina	ARG	82760724362	12.91942461	640591E+11	13730.51471	17.04993202	15.12142697	14.05239192	133.8398673	-	6.139	-	0.897421188	-
9	Armenia	ARM	14086972330	58.18145547	24212134631	8715.765336	20.64928642	19.17896607	58.86150499	2.83212316	-	-	-	4.022573807	-
10	Aruba	ABW	-	-	-	-	-	-	-	-	-	-	-	-	-
11	Australia	AUS	4.60641E+11	26.72188648	1.72288E+12	64711.7656	29.02756032	24.81177276	21.36069104	6.557375685	-	3.668	-	-	-
12	Austria	AUT	3.06925E+11	59.47718589	5.16034E+11	56505.96828	27.94790742	27.05366709	56.64198122	7.6097137	-	5.264	-	-	-
13	Azerbaijan	AZE	35487352541	49.0436789	72356176471	7155.083604	32.77568889	29.81595953	34.57398246	-9.20663945	-	-	-	5.311920499	-
14	Bahamas, The	BHS	569900000	39.74613802	14398500000	34749.63829	19.580333	-	42.95846846	6.44420145	-	8.643	-	3.711666667	-
15	Bahrain	BHR	-	-	43205000000	29084.30713	-	-	-	-5.011919018	-	-	-	-	-
16	Bangladesh	BGD	57551451741	13.15716383	4.37415E+11	25.29.080127	25.75951408	34.90435988	17.82580758	6.89712404	-	-	-	1.051666667	-
17	Barbados	BRB	-	-	6393564190	22672.61544	-	-	-	4.735828043	-	-	-	-	-
18	Belarus	BLR	4801498058	66.81901876	71857382746	7829.053137	27.20042499	25.12331494	66.17977856	7.368230536	-	3.461	-	5.110018951	-
19	Belgium	BEL	5.48013E+11	66.68123029	6.32217E+11	53475.2935	24.39571889	25.20999608	87.61379575	4.079874107	-	5.528	-	-	-
20	Belize	BLZ	1795900000	53.6187719	3281500000	7987.585955	27.06079537	27.61142959	52.85692519	10.89614793	-	-	-	6.355	-
21	Benin	BDN	4169448538	21.19345399	19673284686	1434.662834	31.50750787	-	30.15701102	3.385454352	-	-	-	-	-
22	Bermuda	BMU	-	-	-	-	-	-	-	-	-	-	-	-	-
23	Bhutan	BTN	-	-	-	-	-	-	-	-	-	-	-	8.52	-
24	Bolivia	BOL	12507893618	27.28012912	45849832906	3700.978038	13.11118716	13.48584653	31.56913918	1.69953215	-	3.024	-	5.46430066	-
25	Bosnia and Herzegovina	BH	11935049957	44.12899198	27054889363	8426.09111	8.91426209	18.15246839	56.89710957	5.530501969	-	-	-	2.936145228	-
26	Botswana	BWA	6141387614	31.6635491	19395765126	7249.799326	28.1114262	34.48993349	35.7109683	2.12584044	-	23.381	-	4.75325926	-
27	Brunei	BRU	339557E+11	18.11027942	2.17367E+12	10043.62392	18.4635996	15.98774482	15.7496717	4.858994759	-	-	-	7.947	-
28	British Virgin Islands	VG	-	-	-	-	-	-	-	-	-	-	-	-	-
29	Burkina Faso	BUR	11578413348	16.53488306	15129752594	33430.91848	49.61833245	45.96239131	60.02817372	-13.89619983	-	5.296	-	5.307729167	-
30	Bulgaria	BGR	61847382771	26.88276999	1015084E+11	15797.99598	22.01543135	20.11737728	57.71717775	73.10536602	-	4.319	-	4.358875	-
31	Burundi Faso	BDI	5864783349	28.8556519	20324017839	874.1212804	15.44611371	-	35.96030997	1.998605499	-	5.348	-	-	-
32	Burundi	BDI	1395817464	3.623985206	2642161460	199.5807602	-6.07898685	11.00104759	24.37862366	11.37633149	-	-	-	-	-
33	Cabo Verde	CPV	1027691033	39.27133377	2587253074	4321.579864	4.95097471	20.80180589	54.14383739	3.85799077	-	-	-	6.748370992	-
34	Cameroon	CMR	8774028347	18.3	47945510090	1672.648889	15.4	-	20.3	2.722609551	-	-	-	-	-
35	Canada	CAN	717778E+11	53.53408007	2.14009E+12	53371.69743	23.54430029	21.84652202	33.92683079	1.60524181	-	5.415	-	-	-
36	Cayman Islands	CYM	-	-	-	-	-	-	-	-	-	-	-	-	-
37	Central African Republic	CAT	3600341704	14.44887571	76551885788	44E.078762	6.705780001	-	10.038744	3.988861705	-	-	-	-	-

Showing 1 to 38 of 217 entries. 15 total columns

Console

Valorile lipsa erau notate implicit in acest set de date prin “.” asa ca am transformat aceasta notare in una tipica (“NA”).

ProjectAD - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Add

data

	name_tara	cod_tara	exporturi	exporturi (\$kpb)	pib	pib/nc_jec	economie_interna (\$kpb)	consum_interna (\$kpb)	investit_publica (\$kpb)	inflatie_consum	inflatie_pib	investit_straini (\$kpb)	rata_somaj	urbanizare	investit_externe
1	Afghanistan	AFG	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2	Albania	ALB	9104536292	3562339601	2297377861	8367.775731	134444591	17.75911608	44.91489468	4.553082811	NA	NA	NA	5.004166667	NA
3	Algeria	DZA	60376754380	25.16752082	239899E+11	5260.20625	41.87642996	41.2259377	21.02837305	-2.258254272	NA	NA	NA	6.25	NA
4	American Samoa	ASM	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
5	Andorra	AND	NA	NA	3727673593	46544.72072	NA	NA	NA	5.859905503	NA	NA	NA	NA	NA
6	Angola	AGO	3378130811	39.8726733	84722957642	2309.52162	39.82419499	30.17871709	26.66147599	17.64775693	NA	NA	NA	10.04341747	NA
7	Antigua and Barbuda	ATG	NA	NA	2033085185	21560.21533	NA	NA	NA	4.805475773	NA	NA	NA	5.337422	NA
8	Argentina	ARG	82760724362	12.91942461	640591E+11	13730.51471	17.04993202	15.12142697	14.05239192	133.8398673	-	6.139	NA	0.897421188	NA
9	Armenia	ARM	14086972330	58.18145547	24212134631	8715.765336	20.64928642	19.17896607	58.86150499	2.83212316	NA	NA	NA	4.022573807	NA
10	Aruba	ABW	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
11	Australia	AUS	4.60641E+11	26.72188648	1.72288E+12	64711.7656	29.02756032	24.81177276	21.36069104	6.557375685	NA	3.668	NA	NA	NA
12	Austria	AUT	3.06925E+11	59.47718589	5.16034E+11	56505.96828	27.94790742	27.05366709	56.64198122	7.6097137	NA	5.264	NA	NA	NA
13	Azerbaijan	AZE	35487352541	49.0436789	72356176471	7155.083604	32.77568889	29.81595953	34.57398246	-9.20663945	NA	NA	NA	5.311920499	NA
14	Bahamas, The	BHS	569900000	39.74613802	14398500000	34749.63829	19.580333	NA	42.95846846	6.44420145	NA	8.643	NA	3.711666667	NA
15	Bahrain	BHR	NA	NA	43205000000	29084.30713	NA	NA	NA	-5.011919018	NA	NA	NA	-	NA
16	Bangladesh	BGD	57551451741	13.15716383	4.37415E+11	25.29.080127	25.75951408	34.90435988	17.82580758	6.89712404	NA	NA	NA	1.051666667	NA
17	Barbados	BRB	NA	NA	6393564190	22672.61544	NA	NA	NA	4.735828043	NA	NA	NA	NA	NA
18	Belarus	BLR	4801498058	66.81901876	71857382746	7829.053137	27.20042499	25.12331494	66.17977856	7.368230536	NA	3.461	NA	5.110018951	NA
19	Belgium	BEL	5.48013E+11	66.68123029	6.32217E+11	53475.2935	24.39571889	25.20999608	87.61379575	4.079874107	NA	5.528	NA	NA	NA
20	Belize	BLZ	1795900000	53.6187719	3281500000	7987.585955	27.06079537	27.61142959	52.85692519	10.89614793	NA	NA	NA	6.355	NA
21	Benin	BDN	4169448538	21.19345399	19673284686	1434.662834	31.50750787	NA	30.15701102	3.385454352	NA	NA	NA	NA	NA

Showing 1 to 21 of 217 entries. 15 total columns

```
> str(date)
'data.frame': 217 obs. of 15 variables:
 $ nume_tara      : chr  "Afghanistan" "Albania" "Algeria" "American Samoa" ...
 $ cod_tara       : chr  "AFG" "ALB" "DZA" "ASM" ...
 $ exporturi      : num  NA 9.10e+09 6.04e+10 NA NA ...
 $ exporturi (%pib) : num  NA 39.6 25.2 NA NA ...
 $ pib            : num  NA 2.30e+10 2.40e+11 NA 3.73e+09 ...
 $ pib/nr_loc     : num  NA 8368 5260 NA 46545 ...
 $ economie_interna (%pib) : num  NA 13.4 41.9 NA NA ...
 $ consum_intern (%pib)  : num  NA 17.8 41.2 NA NA ...
 $ investitii_publice (%pib): num  NA 44.9 21 NA NA ...
 $ inflatie_consum : num  NA 4.55 -2.26 NA 5.86 ...
 $ inflatie_pib    : num  NA NA NA NA NA NA NA NA NA ...
 $ investitii_straini (%pib): num  NA NA NA NA NA ...
 $ rata_somaj     : num  NA NA NA NA NA NA NA NA NA NA ...
 $ urbanizare     : num  NA 5 6.25 NA NA ...
 $ investitii_externe : num  NA NA NA NA NA NA NA NA NA ...

> |
```

ProiectAO - T000											
File Edit Code View Plugins Session Build Debug Profile Tools Help											
Go to RunFunction											
+ Addrs											
proiectAO - T000											
Filter											
name_tara	cod_tara	exporturi	exporturi (Fipb)	pib	pib/hr_joc	economic_interna (Fipb)	consum_intern (Fipb)	investit_publica (Fipb)	infatie_consum	investit_traini (Fipb)	urbanizare
1 Afghanistan	AFG	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2 Albania	ALB	9.104536e+09	39.823396	2.297761e+10	8367.7757	13.6664991	17.599116	44.914895	4.55308281	N/A	5.0541667
3 Algeria	DZA	6.037875e+10	25.167521	2.396901e+11	5260.2063	41.8764300	41.225938	21.028373	-2.25825427	N/A	6.2500000
4 American Samoa	ASM	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
5 Andorra	AND	N/A	N/A	3.727214e+09	46544.7207	N/A	N/A	N/A	5.89590550	N/A	N/A
6 Angola	AGO	3.378131e+10	39.872673	8.472256e+10	2305.5216	39.8241950	30.178717	26.661476	17.56475693	N/A	10.0426175
7 Antigua and Barbuda	ATG	N/A	N/A	2.038085e+09	21560.2153	N/A	N/A	N/A	4.80545757	N/A	5.3392420
8 Argentina	ARG	8.276072e+10	12.919425	6.405910e+11	37035.147	17.0499320	15.121427	14.052392	133.83986730	6.139	0.8914212
9 Armenia	ARM	1.408097e+10	58.181455	2.421213e+10	8715.7653	20.6492846	19.178966	58.861505	2.83212316	N/A	4.0225738
10 Aruba	ABW	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
11 Australia	AUS	4.606410e+11	26.721986	1.728830e+12	64711.7656	29.0275602	24.811773	21.360691	6.55737569	3.668	N/A
12 Austria	AUT	3.060230e+11	59.477186	5.163040e+11	56505.9683	27.9478074	27.053667	56.641981	7.60687137	5.264	N/A
13 Azerbaijan	AZE	3.548735e+10	49.045387	7.256189e+10	7155.0836	32.7758889	29.815960	34.573982	-9.20663945	N/A	5.3119205
14 Bahamas, The	BHS	5.699000e+09	39.746138	1.433805e+10	34749.6383	15.5850333	N/A	42.958468	6.34642014	8.643	3.7116667
15 Bahrain	BHR	N/A	N/A	4.320500e+10	20084.3071	N/A	N/A	N/A	-5.01191902	N/A	N/A
16 Bangladesh	BGD	5.755145e+10	13.157164	4.374150e+11	2529.0801	25.7595141	34.904360	17.825808	6.89712404	N/A	1.0916667
17 Barbados	BRE	N/A	N/A	6.395564e+09	22972.154	N/A	N/A	N/A	4.75823904	N/A	N/A
18 Belarus	BLR	4.8071460e+10	66.819019	7.186738e+10	7829.0351	27.2004250	25.223315	66.179779	7.368202054	3.461	5.1180190
19 Belgium	BEL	5.489130e+11	86.681230	6.322170e+11	53475.2905	34.9571169	25.206997	61.873796	4.01786741	5.528	N/A
20 Belize	BZL	1.759500e+09	53.618772	3.281500e+09	7987.5860	27.0607554	27.811430	52.856425	10.89617493	N/A	6.3550000
21 Benin	BJN	4.161946e+09	21.193454	1.967338e+10	1434.6628	31.5075079	N/A	30.157011	3.38546435	N/A	N/A
22 Bermuda	BMU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
23 Bhutan	BTN	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
24 Bolivia	BOL	1.250789e+10	27.280129	4.584983e+10	3700.8783	13.1111872	13.8484585	31.548139	1.69953021	3.024	5.4364607
25											

Am vrut sa vad ce procentaj de valori lipsa am in fiecare coloana pentru a stii ce strategie sa aplic pentru a le gestiona.

```
> procentaj_lipsa <- colSums(is.na(date)) / nrow(date) * 100
> print(procentaj_lipsa)
```

	nume_tara	cod_tara	exporturi	exporturi (%pib)	pib	pib/nr_loc	economie_i
ntema (%pib)	consum_intern (%pib)	investitii_publice (%pib)					
	0.00000	0.00000	36.40553	36.40553	14.74654	14.74654	
38.70968	55.29954	36.40553					
	inflatie_consum	investitii_straini (%pib)	urbanizare				
	14.74654	61.75115	61.29032				

Am gestionat valorile lipsa prin inlocuire cu mediana, atunci cand procentajul este mai mic de 55% si prin interpolare folosind package-ul “zoo” atunci cand procentajul este mai mare, intrucat inlocuirea cu mediana ar fi dus la incoerenta datelor.

	nume_tara	cod_tara	exporturi	exporturi (%pib)	pib	pib/nr_loc	economie_interna (%pib)	consum_intern (%pib)	investitii_publice (%pib)	inflatie_consum	investitii_straini (%pib)	urbanizare
1	Afghanistan	AFG	2.69707e+10	35.146212	4.706184e+10	7249.7993	21.1872136	NA	42.170815	5.00172040	NA	NA
2	Albania	ALB	9.10436e+09	39.623396	2.291708e+10	8367.7757	13.4660591	17.359116	44.914895	4.55308281	NA	5.0041667
3	Algeria	DZA	6.93767e+10	25.167521	2.398990e+11	5260.2063	41.8764300	41.255938	21.28373	-2.29825427	NA	6.2500000
4	American Samoa	ASM	2.69707e+10	35.146212	4.706184e+10	7249.7993	21.1872136	37.543531	42.170815	5.00172040	NA	7.5141392
5	Andorra	AND	2.69707e+10	35.146212	3.727674e+09	4654.7207	21.1872136	33.861134	42.170815	5.89909550	NA	8.7787783
6	Angola	AGO	3.78131e+10	39.872673	8.472296e+10	2393.5216	39.87431950	30.178717	26.661476	17.64775693	NA	10.0424175
7	Antigua and Barbuda	ATG	2.69707e+10	35.146212	2.033080e+09	21540.2153	21.1872136	23.650072	42.170815	4.80547577	NA	5.3324220
8	Argentina	ARG	8.276072e+10	12.919425	6.405910e+11	13730.5147	17.0499320	15.121427	14.052392	133.89966730	6.139000	0.8974212
9	Armenia	ARM	1.408697e+10	58.181455	2.421213e+10	8715.7653	20.4493846	15.178966	58.861505	2.83212316	5.315333	4.0225738
10	Aruba	ABW	2.69707e+10	35.146212	4.706184e+10	7249.7993	21.1872136	21.995369	42.170815	5.00172040	4.491167	4.3449105
11	Australia	AUS	4.606410e+11	26.721986	1.723830e+12	64711.7656	29.0275602	24.811773	21.360691	6.55737569	3.668000	4.6672472
12	Austria	AUT	3.069230e+11	59.677186	5.160340e+11	56505.9683	27.9479074	27.053667	56.641981	7.60691137	5.264000	4.9895838
13	Azerbaijan	AZE	3.548735e+10	49.045368	7.235618e+10	7155.0836	32.7758889	29.815960	34.573982	-9.20663945	6.953500	5.3119205
14	Bahamas, The	BHS	5.695000e+09	39.746138	1.433850e+10	34749.6383	19.5850333	31.512093	42.958468	6.34442014	8.643000	3.7116667
15	Bahrain	BHR	2.69707e+10	35.146212	4.320500e+10	29084.3071	21.1872136	33.208226	42.170815	-5.01191902	7.347500	2.3816667
16	Bangladesh	BGD	5.755145e+10	13.157164	4.374150e+11	2529.0801	34.904360	17.825808	6.89712404	6.052000	1.0516667	
17	Barbados	BRB	2.69707e+10	35.146212	6.393544e+09	22672.6154	21.1872136	30.013837	42.170815	4.73582804	4.756500	3.0898428
18	Belarus	BLR	4.801440e+10	66.819019	7.185738e+10	7829.0531	27.204250	25.123315	66.179779	7.36823054	3.461000	5.1100190
19	Belgium	BEL	5.480130e+11	86.481230	6.322170e+11	53475.2935	34.3957189	25.269997	87.613794	4.07987411	5.528000	5.7320595
20	Belize	BLZ	1.795000e+09	53.619772	3.281500e+09	7987.5860	27.060750	27.611430	52.856925	10.89614793	4.927500	6.3500000
21	Benin	BJN	4.184649e+09	21.159404	1.947030e+10	1408.6628	31.5075079	34.079718	30.157011	3.98946405	4.337000	7.0766667
22	Bermuda	BMU	2.69707e+10	35.146212	4.706184e+10	7249.7993	21.1872136	20.548007	42.170815	5.00172040	3.726500	7.7983333
23	Bhutan	BTN	2.69707e+10	35.146212	4.706184e+10	7249.7993	21.1872136	17.016796	42.170815	5.00172040	3.126000	8.3200000
24	Bolivia	BOL	1.250789e+10	27.280129	4.584980e+10	3700.9783	13.111872	13.484585	51.691339	1.69953221	2.040000	5.4643007
25	Bosnia and Herzegovina	BH	1.139050e+10	44.128992	2.705490e+10	8426.0911	8.1914265	18.152468	56.897110	5.53530197	10.688000	2.9361452
26	Botswana	BWA	6.141388e+09	31.663545	1.919577e+10	7249.7993	28.1114262	34.480932	35.710087	2.12586034	23.381000	4.7532058
27	Brazil	BRA	3.936370e+11	18.110279	2.172670e+12	10543.6239	18.4520996	15.589745	15.743667	4.65895476	7.947000	31.4612343
28	British Virgin Islands	VGB	2.69707e+10	35.146212	4.706184e+10	7249.7993	21.1872136	30.776068	42.170815	5.00172040	6.621500	18.8848917
29	Brunei Darussalam	BRN	1.157841e+10	75.534830	1.512820e+10	33430.9185	49.6183334	45.962391	60.028174	-12.89619583	5.296000	5.3072292
30	Bulgaria	BGR	6.184739e+10	60.882770	1.015840e+11	15977.5956	22.0154314	20.117377	57.717318	7.51050660	4.319000	4.3358750
31	Burkina Faso	BFA	5.864783e+09	28.805565	2.033462e+10	874.1213	15.4461137	15.562212	35.960031	1.99804950	5.348000	5.1467070
32	Burundi	BDI	1.390671e+08	5.263385	2.642162e+09	199.5808	-6.079087	11.007048	24.327863	11.37603149	5.361400	5.9575390
33	Cabo Verde	CPV	1.027691e+09	39.721334	2.587252e+09	4321.5799	4.9500977	20.800836	54.143837	3.85729908	5.374800	6.7683710
34	Cambodia	KHM	2.234142e+10	73.148896	3.177276e+10	1875.0715	50.1087529	57.097549	40.389511	2.38456072	5.388200	6.3329664
35	Cameroon	CMR	8.774028e+09	18.300000	4.794551e+10	1673.6489	15.4000000	35.469085	20.300000	2.72609955	5.401600	5.8916219
36	Canada	CAN	7.177080e+11	33.536408	2.140090e+12	53371.6974	23.5443003	21.840622	33.926831	1.60524181	5.415000	5.4622473
37	Cayman Islands	CYM	2.69707e+10	35.146212	4.706184e+10	7249.7993	21.1872136	21.340137	42.170815	5.00172040	6.134600	5.038728

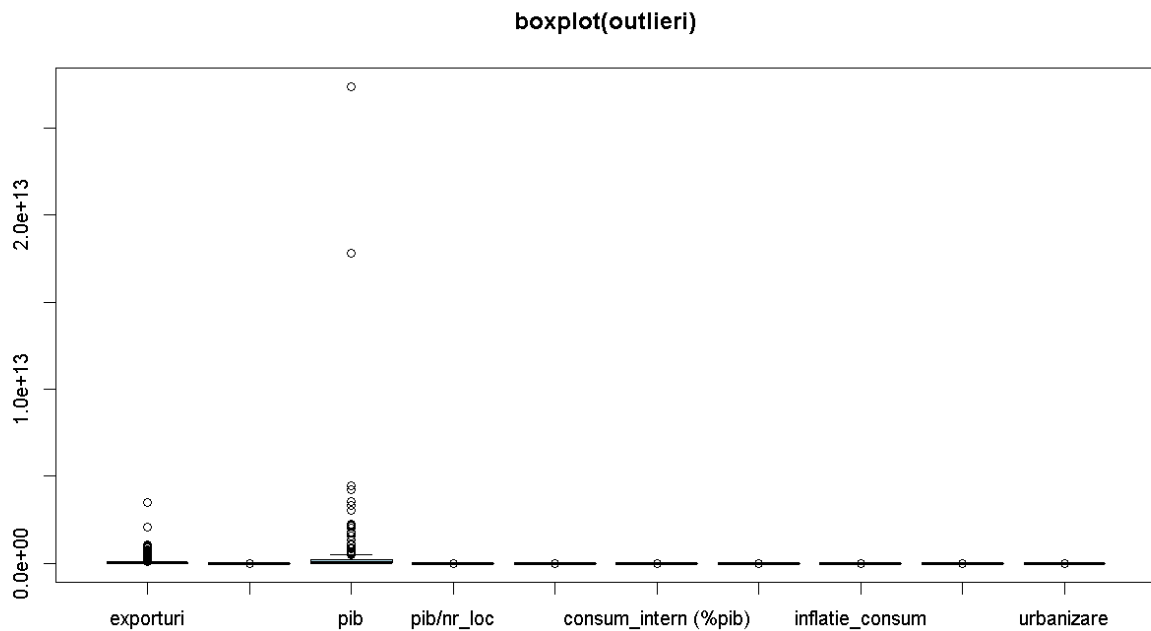
Au ramas cateva valori lipsa in coloanele in care am populat prin interpolare, asadar am folosit inlocuirea prin mediana si in acest caz pentru a scapa definitiv de valorile nule.

```
> procentaj_lipsa <- colSums(is.na(date)) / nrow(date) * 100
> print(procentaj_lipsa)
```

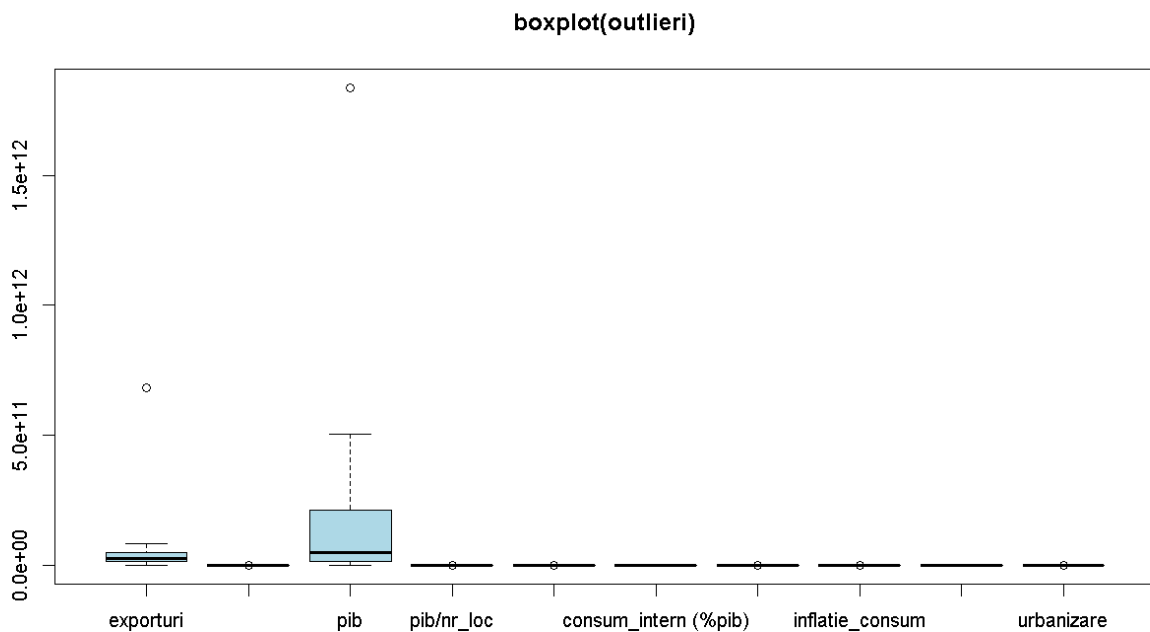
	nume_tara	cod_tara	exporturi	exporturi (%pib)	pib	pib/nr_loc	economie_interna (%pib)	consum_intern (%pib)	investitii_publice (%pib)
	0	0	0	0	0	0	0	0	0
	inflatie_consum	investitii_straini (%pib)	urbanizare						
	0	0	0						

Observam acum ca nu mai exista valori lipsa in setul de date.

Mai departe, in ceea ce priveste outlierii, am ales sa fac un boxplot pentru a avea o idee generala despre prezenta acestora.



Se observa ca avem 2 variabile problematice din acest punct de vedere ("exporturi si "pib"). Am gestionat situatia prin calcularea IQR pentru fiecare coloana in parte si inlocuirea outlierilor cu percentilele 5 si 95 pentru valorile aflate sub limita inferioara, respective peste limita superioara pentru a reduce semnificativ impactul outlierilor asupra setului de date.



Acesta este rezultatul final in ceea ce priveste numarul de outlieri, observam ca este redus semnificativ si am considerat ca outlierii ramasi nu pot aduce un impact negativ prea mare in ceea ce priveste distributia datelor.

3. Setul de date contine informatii despre anumiti indicatori economici relevanti pentru 217 tari din intreaga lume, oferind o imagine detaliata asupra acestor indicatori pentru anul 2023. Datele provin de pe site-ul World Bank, sectiunea World Development Indicators (<https://databank.worldbank.org/source/world-development-indicators>) si cuprind o serie de variabile care reflecta aspect cheie ale economiilor fiecarei tari, dupa cum urmeaza:

- * numele si codul tarilor
- * exporturi
- * exporturi (%pib) (procent raportat la PIB)
- * pib
- * pib/nr_loc
- * economie_interna(%pib) (procent piata interna raportat la PIB)
- * consum_intern (%pib) (procent raportat la PIB)
- * investitii_publice (%pib)

- * inflatie_consum
- * investitii_straini (%pib)
- * urbanizare

De specificat este faptul ca fiecare valoare prezenta in setul de date este exprimata in dolari americani(\$).

4. Obiectivul general al analizei mele este sa evaluez si sa compar performanta economica a fiecarei tari in anul 2023, utilizand indicatorii disponibili in setul de date. Prin aceasta analiza, imi doresc sa identific diferentele majore intre economiile tarilor si sa inteleg ce factori sunt predominanti in structura economica si nivelul de dezvoltare al fiecarei tari.

5. Output-urile indicatorilor statistici sunt urmatoarele:

```
> date_numerice <- date[, sapply(date, is.numeric)]
> stat_descriptive <- summary(date_numerice)
> print(stat_descriptive)
```

exporturi	exporturi (%pib)	pib	pib/nr_loc	conomie_interna (%pib)	consum_intern (%pib)	investitii_publice (%pib)	inflatie_consum
Min. : 2.064e+07	Min. : 12.95	Min. : 6.228e+07	Min. : 199.6	Min. : 0.3242	Min. : 8.944	Min. : 15.74	Min. : -7.371
1st Qu.: 1.345e+10	1st Qu.: 32.17	1st Qu.: 1.584e+10	1st Qu.: 3367.1	1st Qu.: 19.2082	1st Qu.: 18.782	1st Qu.: 35.71	1st Qu.: 2.866
Median : 2.697e+10	Median : 39.15	Median : 4.706e+10	Median : 7249.8	Median : 21.1872	Median : 22.134	Median : 42.17	Median : 5.002
Mean : 1.506e+11	Mean : 41.98	Mean : 3.096e+11	Mean : 14886.8	Mean : 20.4160	Mean : 23.115	Mean : 45.08	Mean : 6.129
3rd Qu.: 4.801e+10	3rd Qu.: 44.00	3rd Qu.: 2.124e+11	3rd Qu.: 20626.2	3rd Qu.: 23.9993	3rd Qu.: 28.330	3rd Qu.: 49.58	3rd Qu.: 7.393
Max. : 6.821e+11	Max. : 85.16	Max. : 1.835e+12	Max. : 57712.1	Max. : 39.9698	Max. : 41.226	Max. : 86.57	Max. : 26.779

investitii_straini (%pib)	urbanizare
Min. : 0.733	Min. : 0.8974
1st Qu.: 3.742	1st Qu.: 4.0218
Median : 5.368	Median : 5.5757
Mean : 5.729	Mean : 6.3108
3rd Qu.: 6.954	3rd Qu.: 7.5141
Max. : 11.608	Max. : 15.6657

Din acest output putem intelege ca exista diferente mari intre medie si mediana la variabilele “exporturi”, “pib” si “pib/nr_loc”, ceea ce indica outliers mari si sugereaza ca exista cateva tari foarte performante cum ar fi SUA si China care ridica media globala.

```
> descriere <- describe(date_numerice)
> print(descriere)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
exporturi	1	217	1.505954e+11	2.613862e+11	2.697067e+10	1.047182e+11	2.144251e+10	20644623.84	6.821336e+11	6.821130e+11	1.53	0.36	1.774405e+10
exporturi (%pib)	2	217	4.198000e+01	1.955000e+01	3.915000e+01	4.010000e+01	9.080000e+00	12.95	8.516000e+01	7.221000e+01	1.03	0.60	1.330000e+00
pib	3	217	3.096006e+11	5.991418e+11	4.706184e+10	1.635288e+11	5.736648e+10	62280311.59	1.835396e+12	1.835334e+12	2.06	2.46	4.067240e+10
pib/nr_loc	4	217	1.488680e+04	1.751463e+04	7.249800e+03	1.144201e+04	7.828030e+03	199.58	5.771212e+04	5.751254e+04	1.53	1.07	1.188970e+03
conomie_interna (%pib)	5	217	2.042000e+01	1.112000e+01	2.119000e+01	2.048000e+01	3.810000e+00	0.32	3.997000e+01	3.965000e+01	-0.27	-0.01	7.500000e-01
consum_intern (%pib)	6	217	2.312000e+01	7.030000e+00	2.213000e+01	2.293000e+01	6.990000e+00	8.94	4.123000e+01	3.228000e+01	0.26	-0.55	4.800000e-01
investitii_publice (%pib)	7	217	4.508000e+01	1.724000e+01	4.217000e+01	4.331000e+01	1.063000e+01	15.74	8.657000e+01	7.082000e+01	0.95	0.69	1.170000e+00
inflatie_consum	8	217	6.130000e+00	7.770000e+00	5.000000e+00	5.280000e+00	3.270000e+00	-7.37	2.678000e+01	3.415000e+01	1.22	2.12	5.300000e-01
investitii_straini (%pib)	9	217	5.730000e+00	2.440000e+00	5.370000e+00	5.530000e+00	2.380000e+00	0.73	1.161000e+01	1.088000e+01	0.64	-0.27	1.700000e-01
urbanizare	10	217	6.310000e+00	3.490000e+00	5.580000e+00	5.780000e+00	2.490000e+00	0.90	1.567000e+01	1.477000e+01	1.37	1.59	2.400000e-01

Diferentele majore de economie sunt reprezentate si de catre deviatia standard si interval (range), ceea ce indica o dispersie semnificativa a valorilor.

```
> sapply(date_numerice, sd, na.rm = TRUE)
```

	exporturi	exporturi (%pib)	pib	pib/nr_loc	conomie_interna (%pib)	consum_intern (%pib)
	2.613862e+11	1.955438e+01	5.991418e+11	1.751463e+04	1.112179e+01	7.031510e+00
investitii_publice (%pib)	1.723602e+01	7.765653e+00	2.435057e+00	3.485321e+00		


```
> coef_var <- sapply(date_numerice, function(x) sd(x, na.rm = TRUE) / mean(x, na.rm = TRUE) * 100)
> print(coef_var)
```

exporturi	exporturi (%pib)	pib	pib/nr_loc	economie_interna (%pib)	consum_intern (%pib)
173.56851	46.58513	193.52089	117.65208	54.47586	30.41958
investitii_publice (%pib)	inflatie_consum	investitii_straini (%pib)	urbanizare		
38.23494	126.71221	42.50523	55.22751		

```
> |
```

Coeficientul de variatie ne arata unde sunt cele mai mari diferente intre tari (“exporturi” si “PIB”) si unde exista mai multa uniformitate (“exporturi (%pib)”, “investitii_publice (%pib)”, “investitii_straini (%pib)”). Astfel, este pusa in evidenta structura economica variata, dar cu elemente comune, a tarilor analizate.

```
> varianta <- sapply(date_numerice, var, na.rm = TRUE)
> print(varianta)
```

exporturi	exporturi (%pib)	pib	pib/nr_loc	economie_interna (%pib)	consum_intern (%pib)
6.832272e+22	3.823736e+02	3.589709e+23	3.067622e+08	1.236942e+02	4.944213e+01
investitii_publice (%pib)	inflatie_consum	investitii_straini (%pib)	urbanizare		
2.970802e+02	6.030536e+01	5.929505e+00	1.214746e+01		

```
> |
```

Observam ca variabilele exprimate ca procente din PIB au variante mai mici, sugerand echilibru. Varianta inflatiei arata fluctuatii influentate de politici monetare si stabilitate economica, iar urbanizarea arata niveluri diferite de dezvoltare urbana, insa diferentele sunt mai moderate decat in cazul PIB-ului.

```
> skewness <- sapply(date_numerice, skewness, na.rm = TRUE)
> print(skewness)
```

exporturi	exporturi (%pib)	pib	pib/nr_loc	economie_interna (%pib)	consum_intern (%pib)
1.5403648	1.0356247	2.0733568	1.5377225	-0.2752608	0.2622236
investitii_publice (%pib)	inflatie_consum	investitii_straini (%pib)	urbanizare		
0.9608080	1.2326397	0.6471670	1.3784512		

```
> |
```

Valorile pozitive ale coeficientului de asimetrie la majoritatea variabilelor sugereaza o distributie asimetrica spre dreapta, cu cateva valori mari care afecteaza distributia. In ceea ce priveste economia interna in raport cu PIB, observam o asimetrie usoara la stanga.

```
> kurtosis <- sapply(date_numerice, kurtosis, na.rm = TRUE)
> print(kurtosis)
```

exporturi	exporturi (%pib)	pib	pib/nr_loc	economie_interna (%pib)	consum_intern (%pib)
3.390594	3.628601	5.512209	4.103371	3.014333	2.469427
investitii_publice (%pib)	inflatie_consum	investitii_straini (%pib)	urbanizare		
3.723602	5.170287	2.751785	4.635524		

```
> |
```

Coeficientul de boltire ne arata ca in cazul variabilelor precum “pib” si “inflatie_consum” avem parte de un numar de outlieri mai mare decat ar avea o distributie normala, iar curba este leptocurtica. In cazul celorlalte variabile, valorile sunt mai mult concentrate in jurul mediei si nu avem o tendinta catre valori extreme, intrucat coeficientul de boltire se invarte in jurul valorii 3.

6. Matrice corelatie:

	exporturi	exporturi (%pib)	pib	pib/nr_loc	economie_interna (%pib)	consum_intern (%pib)	investitii_publice (%pib)	inflatie_consum	investitii_straini (%pib)	urbanizare
exporturi	1.00000000	0.2370976	0.75794332	0.5569202	0.33466460	0.178564953	0.076591182	-0.02783680	-0.09706921	-0.148552585
exporturi (%pib)	0.23709764	1.00000000	0.03909500	0.4335968	0.39598208	0.124094526	0.752857669	-0.16014909	-0.12013289	-0.062141197
pib	0.75794332	0.0390950	1.00000000	0.5154577	0.28021833	0.147486822	-0.102097499	-0.01128977	-0.14464516	-0.125331004
pib/nr_loc	0.55692020	0.4335968	0.51545772	1.00000000	0.35709709	0.178964685	0.273862389	-0.20123923	-0.15179104	-0.168619601
economie_interna (%pib)	0.33466460	0.3959821	0.28021833	0.3570971	1.00000000	0.427316670	-0.013148359	-0.21482210	-0.15462155	-0.069776681
consum_intern (%pib)	0.17856495	0.1240945	0.14748682	0.1789647	0.42731667	1.000000000	-0.003766772	-0.09503387	-0.10952015	-0.119474183
investitii_publice (%pib)	0.07659118	0.7528577	-0.10209750	0.2738624	-0.01314836	-0.003766772	1.000000000	-0.05465885	-0.11321564	-0.005760445
inflatie_consum	-0.02783680	-0.1601491	-0.01128977	-0.2012392	-0.21482210	-0.095033870	-0.054658852	1.00000000	0.07803365	0.159883909
investitii_straini (%pib)	-0.09706921	-0.1201329	-0.14464516	-0.1517910	-0.15462155	-0.109520145	-0.113215639	0.07803365	1.00000000	0.052224710
urbanizare	-0.14855258	-0.0621412	-0.12533100	-0.1686196	-0.06977668	-0.119474183	-0.005760445	0.15988391	0.05222471	1.000000000

Observam o corelatie pozitiva puternica intre PIB si exporturi (0.76), ceea ce sugereaza ca, pe masura ce exporturile unei tari cresc, PIB-ul tinde si el sa creasca, indicand o dependenta a economiilor de exporturi.

De asemenea, corelatia moderata intre PIB si PIB/nr. loc sugereaza, pe langa influenta PIB-ului asupra PIB-ului pe cap de locuitor, ca PIB-ul pe cap de locuitor poate varia in functie de populatie si distributia raselor.

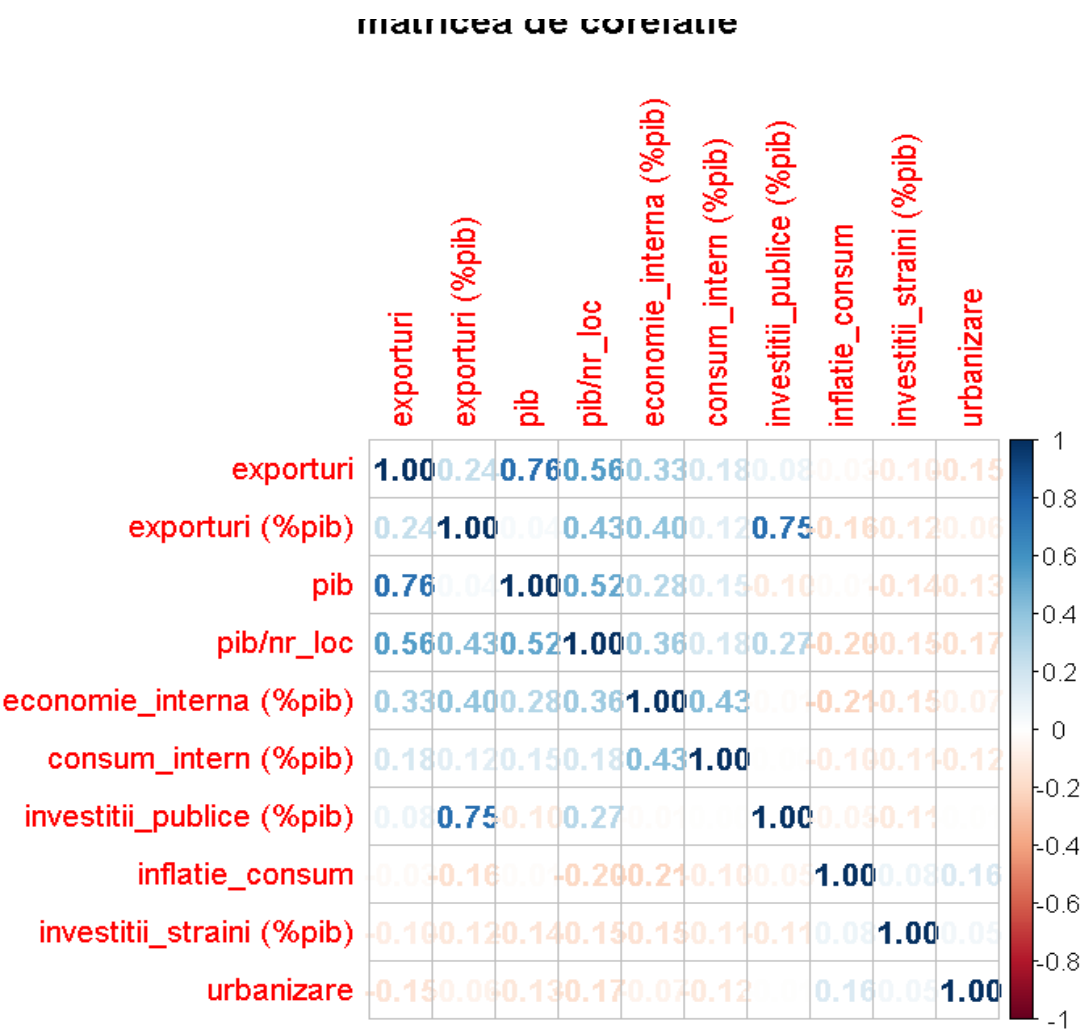
Corelatia negativa intre urbanizare si unele variabile economice, cum ar fi consumul intern si PIB-ul, ne arata ca economiile urbanizate pot avea distributii diferite ale consumului si productiei fata de cele cu o pondere rurala mai mare.

Matrice covarianta:

	exporturi	exporturi (%pib)	pib	pib/nr_loc	economie_interna (%pib)	consum_intern (%pib)	investitii_publice (%pib)	inflatie_consum	investitii_straini (%pib)	urbanizare
exporturi	6.832272e+22	1.211864e+12	1.186995e+23	2.549626e+15	9.728974e+11	3.281915e+11	3.450629e+11	-5.650409e+10	-6.178361e+10	-1.353336e+11
exporturi (%pib)	1.211864e+12	3.823736e+02	4.580310e+11	1.485015e+05	8.611805e+01	1.706260e+01	2.537428e+02	-2.431904e+01	-5.720251e+00	-4.235126e+00
pib	1.186995e+23	4.580310e+11	3.589709e+23	5.409082e+15	1.867243e+12	6.213431e+11	-1.054342e+12	-5.252824e+10	-2.110293e+11	-2.617164e+11
pib/nr_loc	2.549626e+15	1.485015e+05	5.409082e+15	3.067622e+08	6.956037e+04	2.204027e+04	8.267423e+04	-2.737105e+04	-6.473754e+03	-1.029323e+04
economie_interna (%pib)	9.728974e+11	8.611805e+01	1.867243e+12	6.956037e+04	1.236942e+02	3.341744e+01	-2.520479e+00	-1.855375e+01	-4.187491e+00	-2.704754e+00
consum_intern (%pib)	3.281915e+11	1.706260e+01	6.213431e+11	2.204027e+04	3.341744e+01	4.944213e+01	-4.565147e-01	-5.189255e+00	-1.875218e+00	-2.927962e+00
investitii_publice (%pib)	3.450629e+11	2.537428e+02	-1.054342e+12	8.267423e+04	-2.520479e+00	-4.565147e-01	2.970802e+02	-7.316028e+00	-4.751738e+00	-3.460475e-01
inflatie_consum	-5.650409e+10	-2.431904e+01	-5.252824e+10	-2.737105e+04	-1.855375e+01	-5.189255e+00	-7.316028e+00	6.030536e+01	1.475601e+00	4.327385e+00
investitii_straini (%pib)	-6.178361e+10	-5.720251e+00	-2.110293e+11	-6.473754e+03	-4.187491e+00	-1.875218e+00	-4.751738e+00	1.475601e+00	5.929505e+00	4.432288e-01
urbanizare	-1.353336e+11	-4.235126e+00	-2.617164e+11	-1.029323e+04	-2.704754e+00	-2.927962e+00	-3.460475e-01	4.327385e+00	4.432288e-01	1.214746e+01

Covarianta negativa intre inflatie si alte variabile, cum ar fi investitiile publice, sugereaza ca, in general, inflatia crescuta este asociata cu o scadere a anumitor tipuri de investitii.

Grafic pentru reprezentarea matricei de corelatie:



7. Distanța dintre observații (forme) prin metoda euclidiană:

```
> print("Distanțe euclidiene pentru primele 10 observații:")
[1] "Distanțe euclidiene pentru primele 10 observații:"
> print(matrice_eucl[1:10, 1:10])
```

	1	2	3	4	5	6	7	8	9	10
1	0.000000e+00	2.998743e+10	1.957093e+11	1.553123e+01	4.333417e+10	3.827198e+10	4.502876e+10	1.789204e+12	2.623164e+10	1.517298e+00
2	2.998743e+10	0.000000e+00	2.228984e+11	2.998743e+10	2.626331e+10	6.649378e+10	2.752952e+10	1.813914e+12	5.133084e+09	2.998743e+10
3	1.957093e+11	2.228984e+11	0.000000e+00	1.957093e+11	2.385222e+11	1.574386e+11	2.402002e+11	1.595654e+12	2.205982e+11	1.957093e+11
4	1.553123e+01	2.998743e+10	1.957093e+11	0.000000e+00	4.333417e+10	3.827198e+10	4.502876e+10	1.789204e+12	2.623164e+10	1.589206e+01
5	4.333417e+10	2.626331e+10	2.385222e+11	4.333417e+10	0.000000e+00	8.128112e+10	1.694588e+09	1.832518e+12	2.419923e+10	4.333417e+10
6	3.827198e+10	6.649378e+10	1.574386e+11	3.827198e+10	8.128112e+10	0.000000e+00	8.296987e+10	1.751358e+12	6.363510e+10	3.827198e+10
7	4.502876e+10	2.752952e+10	2.402002e+11	4.502876e+10	1.694588e+09	8.296987e+10	0.000000e+00	1.834212e+12	2.564956e+10	4.502876e+10
8	1.789204e+12	1.813914e+12	1.595654e+12	1.789204e+12	1.832518e+12	1.751358e+12	1.834212e+12	0.000000e+00	1.812485e+12	1.789204e+12
9	2.623164e+10	5.133084e+09	2.205982e+11	2.623164e+10	2.419923e+10	6.363510e+10	2.564956e+10	1.812485e+12	0.000000e+00	2.623164e+10
10	1.517298e+00	2.998743e+10	1.957093e+11	1.589206e+01	4.333417e+10	3.827198e+10	4.502876e+10	1.789204e+12	2.623164e+10	0.000000e+00

Am afișat output-ul numai pentru primele 10 observații pentru a putea fi observat mai ușor.

Valorile din matricea distanțelor indică similaritatea/disimilaritatea între observații bazată pe distanța euclidiană. Diagonala principală va avea mereu valoarea 0, iar valorile din afara diagonalei arată distanțele reale. De exemplu, distanța dintre obs. 1 și obs. 2 este $2.998e+10$, iar distanța dintre observația 1 și observația 3 este $1.957e+11$, ceea ce sugerează că obs. 3 este mult mai diferită de obs. 1 decât obs. 2.

Distanța dintre observații prin metoda Manhattan:

```
> print("Distanțe Manhattan pentru primele 10 observații:")
[1] "Distanțe Manhattan pentru primele 10 observații:"
> print(matrice_manh[1:10, 1:10])
```

	1	2	3	4	5	6	7	8	9	10
1	0.000000e+00	4.195030e+10	2.262432e+11	1.734821e+01	4.333422e+10	4.447176e+10	4.502877e+10	1.844124e+12	3.573341e+10	2.245617e+00
2	4.195030e+10	0.000000e+00	2.681935e+11	4.195030e+10	3.711619e+10	8.642206e+10	3.881074e+10	1.886075e+12	6.216893e+09	4.195030e+10
3	2.262432e+11	2.681935e+11	0.000000e+00	2.262432e+11	2.695775e+11	1.817715e+11	2.712720e+11	1.617881e+12	2.619767e+11	2.262432e+11
4	1.734821e+01	4.195030e+10	2.262432e+11	0.000000e+00	4.333422e+10	4.447176e+10	4.502877e+10	1.844124e+12	3.573341e+10	1.959382e+01
5	4.333422e+10	3.711619e+10	2.695775e+11	4.333422e+10	0.000000e+00	8.780598e+10	1.694625e+09	1.887458e+12	3.336821e+10	4.333422e+10
6	4.447176e+10	8.642206e+10	1.817715e+11	4.447176e+10	8.780598e+10	0.000000e+00	8.950053e+10	1.799652e+12	8.020517e+10	4.447176e+10
7	4.502877e+10	3.881074e+10	2.712720e+11	4.502877e+10	1.694625e+09	8.950053e+10	0.000000e+00	1.889153e+12	3.506276e+10	4.502877e+10
8	1.844124e+12	1.886075e+12	1.617881e+12	1.844124e+12	1.887458e+12	1.799652e+12	1.889153e+12	0.000000e+00	1.879858e+12	1.844124e+12
9	3.573341e+10	6.216893e+09	2.619767e+11	3.573341e+10	3.336821e+10	8.020517e+10	3.506276e+10	1.879858e+12	0.000000e+00	3.573341e+10
10	2.245617e+00	4.195030e+10	2.262432e+11	1.959382e+01	4.333422e+10	4.447176e+10	4.502877e+10	1.844124e+12	3.573341e+10	0.000000e+00

Formula de calcul a distanței Manhattan este: $(x, y) = \sum |x_i - y_i|$

Valorile din matrice sunt calculate pe baza distanței absolute dintre fiecare pereche de observații, ceea ce face această metodă mai puțin sensibilă la valorile extreme. Observațiile cu distanțe mici observăm că sunt similare între cele două metode (de ex. observațiile 1 și 4).

8. Algoritmul de CLUSTERIZARE IERARHICĂ:

8.1 Încărcarea datelor (facut)

8.2 Standardizarea datelor (cu asta am început rezolvarea acestei cerințe)

8.3 Distanța dintre forme (facută la punctul anterior)

8.4 Analiza cluster prin 2 metode:

8.4.1 Metoda agregare completa(complete linkage): favorizeaza clustere mai compacte, dar poate fi sensibila la outliers

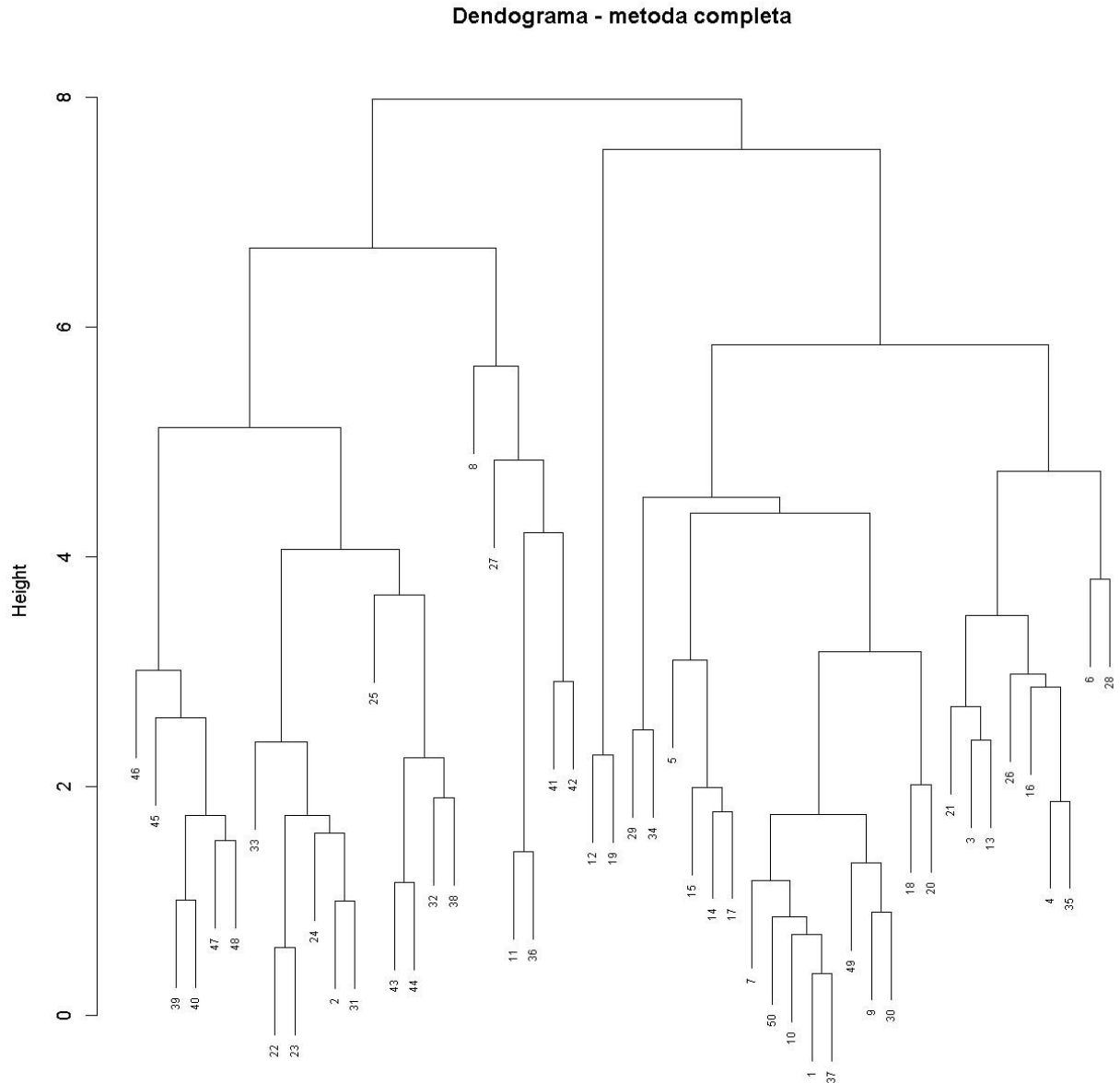
Name	Type	Value
clusterizare_completa	list [7] (S3: hclust)	List of length 7
merge	integer [49 x 2]	-1 -22 -10 -50 -9 -2 -37 -23 1 3 -30 -31 ...
height	double [49]	0.370 0.597 0.709 0.861 0.899 0.998 ...
order	integer [50]	46 45 39 40 47 48 ...
labels	NULL	Pairlist of length 0
method	character [1]	'complete'
call	language	hclust(d = dist(date_subset, method = "euclidean"), method = "complete")
dist.method	character [1]	'euclidean'

8.4.2 Metoda Ward (Ward linkage): minimizeaza varianta, avem grupuri mai compacte.

Name	Type	Value
clusterizare_ward	list [7] (S3: hclust)	List of length 7
merge	integer [49 x 2]	-1 -22 -10 -50 -9 -2 -37 -23 1 3 -30 -31 ...
height	double [49]	0.370 0.597 0.678 0.873 0.899 0.998 ...
order	integer [50]	11 36 12 19 8 27 ...
labels	NULL	Pairlist of length 0
method	character [1]	'ward.D2'
call	language	hclust(d = dist(date_subset, method = "euclidean"), method = "ward.D2")
dist.method	character [1]	'euclidean'

8.5 Dendrograma (arborele de clusterizare)

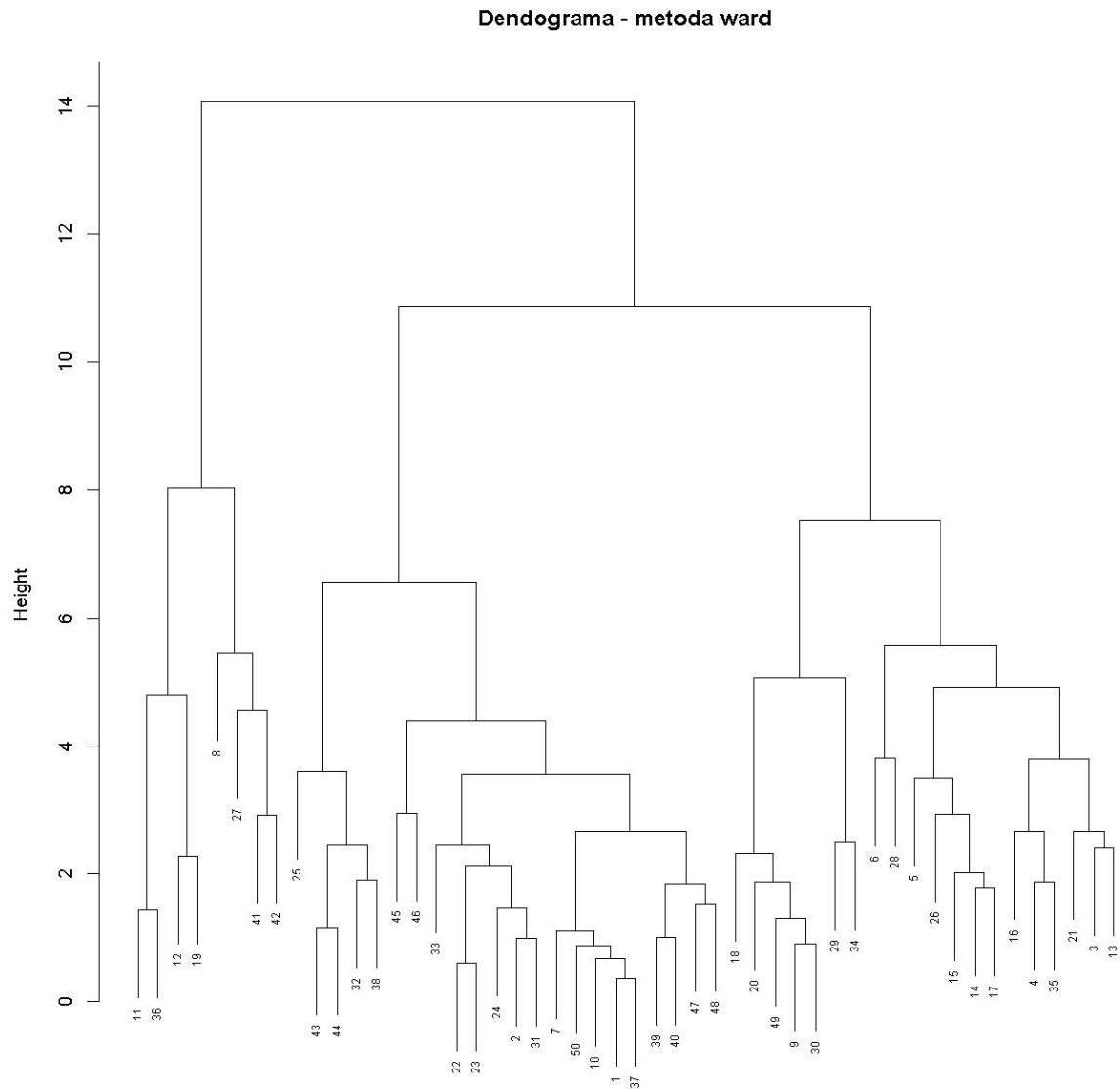
Metoda agregare completa(complete linkage):



Am lucrat pe un subset care reprezinta primele 50 de observatii pentru a putea pune in evidenta diferentele dintre grupuri.

Structura ierarhica arata cum sunt grupate observatiile, incepand de la nivel individual pana la un singur cluster. Ramurile lungi de la nivelele superioare sugereaza diferente mari intre grupuri.

Metoda Ward (Ward linkage):

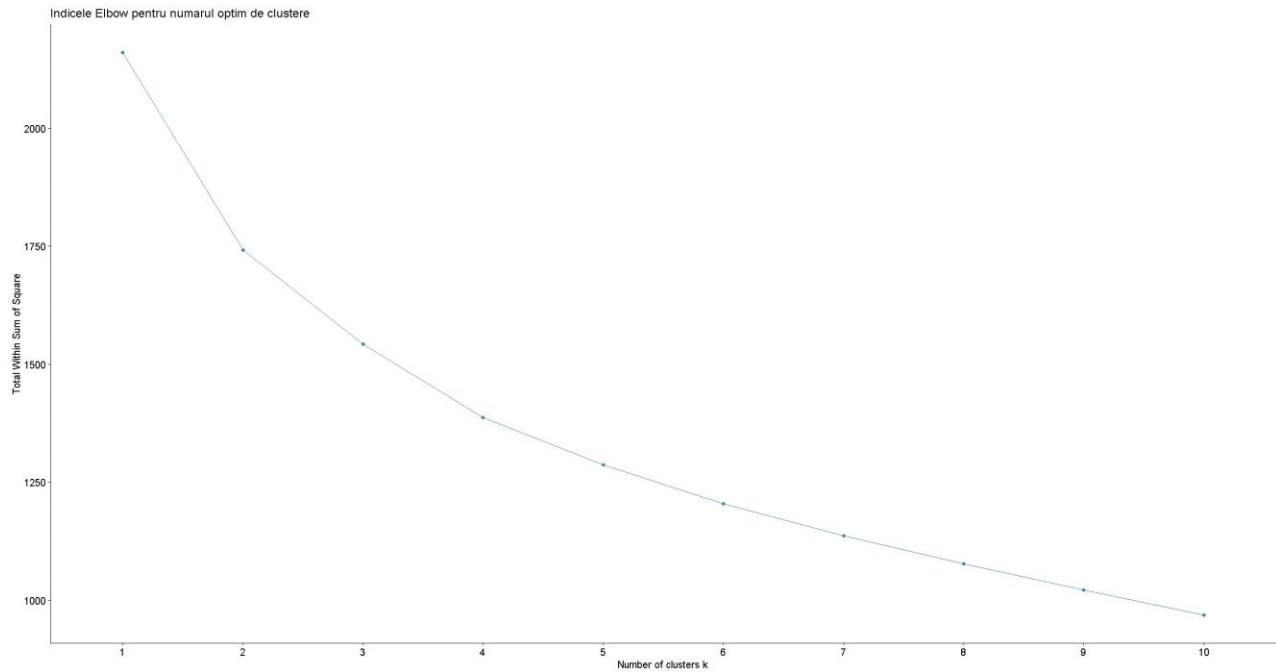


Comparativ cu metoda completa, Ward produce o structura mai echilibrata si mai omogena a clusterelor. Taierea la un anumit nivel (de exemplu 8-10) ar genera mai multe clusterse diferite.

STOP Algoritm

9. Pentru metoda Ward, variantele de decizie pentru numarul optim de clustere sunt:

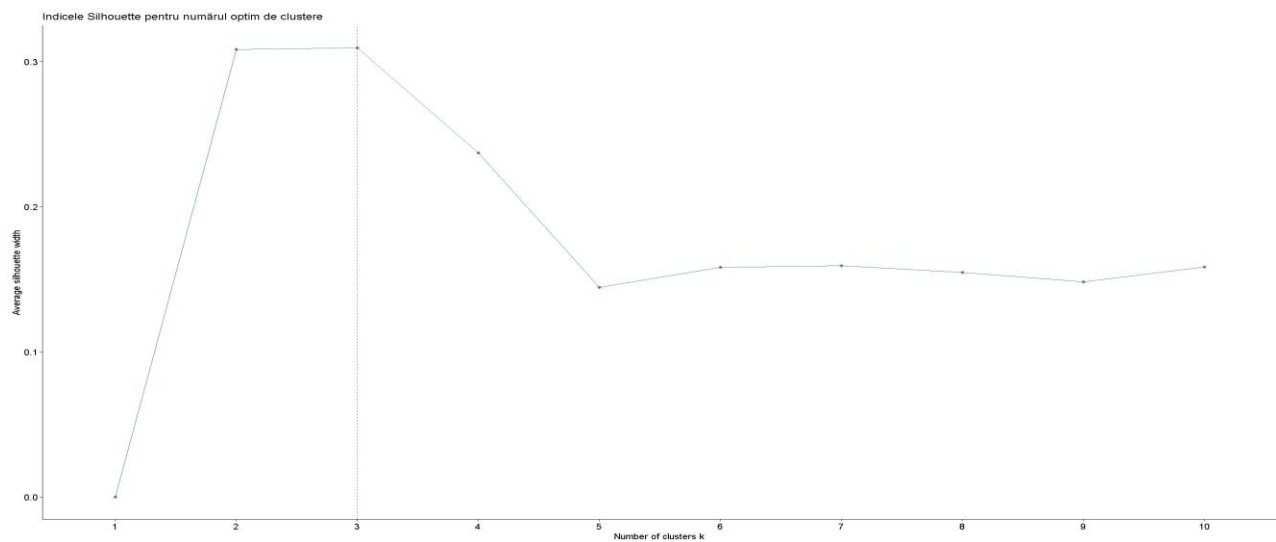
9.1 Indicele Elbow



Se observa o descrestere semnificativa a valorilor sumei totale a distantei patratelor intra-cluster pe masura ce numarul de clustere creste.

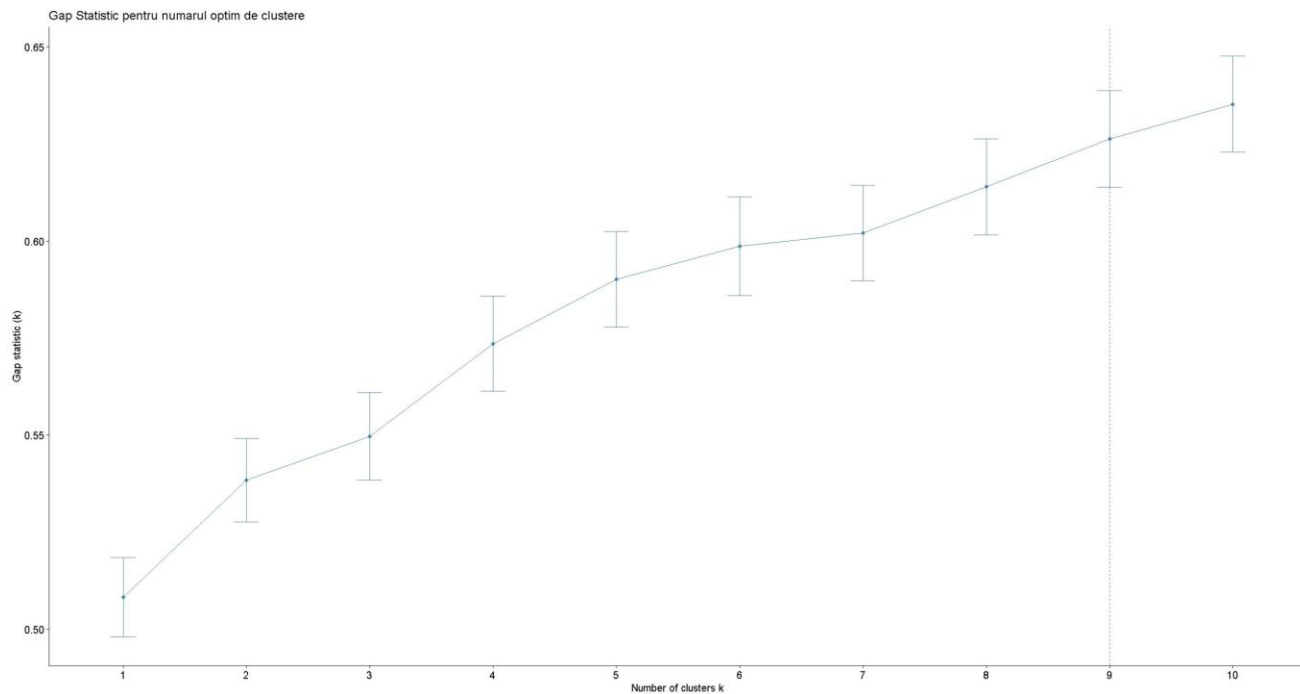
In acest grafic, numarul optim de clustere pare sa fie 4, deoarece scaderea devine mai lenta dupa acest punct.

9.2 Indicele Silhouette:



In acest grafic, valoarea maxima a indicelui Silhouette este atinsa pentru k=3. Asta inseamna ca numarul optim de clustere este 3 conform acestei metode.

9.3 Calculul Gap Statistic:



Din calculul variabilitatii intra-cluster cu o referinta de date aleatorii, numarul optim de clustere pare sa fie k = 4, intrucat cresterile ulterioare sunt mai lente.

Din informatiile pe care le-au oferit toate cele 3 variante, putem trage concluzia ca numarul optim de clustere este 4, intrucat acesta a fost rezultatul majoritar al metodelor.

10. Dupa ce am aplicat metoda Ward cu hclust(), am taiat dendograma pentru a obtine 4 clustere si am adaugat informatia despre clustere la datele originale.

cod_tara	exporturi	exporturi (%pib)	pib	pib/nr_loc	economie_interna (%pib)	consum_intern (%pib)	investitii_publice (%pib)	inflatie_consum	investitii_straine (%pib)	urbanizare	Cluster
AFG	26970668447	39.14621	4.706184e+10	7249.7993	21.1872136	22.133735	42.17081	5.00172040	5.368100	5.5757284	1
ALB	9104536292	39.62340	2.297768e+10	8367.7757	13.4444591	17.759116	44.91489	4.55308281	5.368100	5.0041667	1
DZA	60376754380	25.16752	2.398990e+11	5260.2063	39.9698352	41.225938	21.02837	-2.25825427	5.368100	6.2500000	1
ASM	26970668447	39.14621	4.706184e+10	7249.7993	21.1872136	37.543531	42.17081	5.00172040	5.368100	7.5141392	1
AND	26970668447	39.14621	3.727674e+09	57712.1210	21.1872136	33.861124	42.17081	5.85990550	5.368100	8.7782783	1
AGO	33781308111	39.87267	8.472296e+10	2309.5216	39.9698352	30.178717	26.66148	26.77868896	5.368100	10.0424175	2
ATG	26970668447	39.14621	2.033085e+09	21560.2153	21.1872136	22.650072	42.17081	4.80547577	5.368100	5.3374220	1
ARG	82760724362	12.94857	1.835396e+12	13730.5147	17.0499320	15.121427	19.88806	26.77868896	6.139000	0.8974212	2
ARM	14086972330	58.18146	2.421213e+10	8715.7653	20.6492846	19.178966	58.86150	2.83212316	5.315333	4.0225738	1
ABW	26970668447	39.14621	4.706184e+10	7249.7993	21.1872136	21.995369	42.17081	5.00172040	4.491667	4.3449105	1
AUS	682133600000	26.72199	1.835396e+12	57712.1210	29.0275602	24.811773	21.36069	6.55737569	3.668000	4.6672472	3
AUT	682133600000	59.47719	1.835396e+12	57712.1210	27.9479074	27.053667	56.64198	7.60697137	5.264000	4.9895838	3
AZE	35487352941	49.04537	7.235618e+10	7155.0836	39.9698352	29.815960	34.57398	-7.37137225	6.953500	5.3119205	1
BHS	5699000000	39.74614	1.433850e+10	34749.6383	19.5850333	31.512093	42.95847	6.34442014	8.643000	3.7116667	1
BHR	26970668447	39.14621	4.320500e+10	29084.3071	21.1872136	33.208226	42.17081	-7.37137225	7.347500	2.3816667	1
BGD	57551451741	12.94857	4.374150e+11	2529.0801	25.7595141	34.904360	17.82581	6.89712404	6.052000	1.0516667	1
BRB	26970668447	39.14621	6.393564e+09	22672.6154	21.1872136	30.013837	42.17081	4.73582804	4.756500	3.0808428	1
BLR	48014398058	85.16075	7.185738e+10	7829.0531	27.2004250	25.123315	66.17978	7.36823054	3.461000	5.1100190	4
BEL	682133600000	85.16075	1.835396e+12	57712.1210	24.3957189	25.206997	86.56720	4.07987411	5.528000	5.7325095	4
BLZ	1759500000	53.61877	3.281500e+09	7987.5860	27.0607954	27.611430	52.85693	10.89674793	4.927500	6.3550000	1

Am ales sa calculez indicatorii statistici per cluster doar pentru variabilele “pib”, “exporturi” si “urbanizare”, intrucat daca as fi facut asta pentru toate variabilele rezultatele erau foarte greu de urmarit si interpretat.

	Cluster	pib_medie	pib_mediana	pib_sd	exporturi_medie	exporturi_mediana	exporturi_sd	urbanizare_medie	urbanizare_mediana	urbanizare_sd
1	1	9.509685e+10	4.645584e+10	230980835065	53349376110	26970668447	142285375223	5.613713	5.503618	2.230103
2	2	2.180149e+11	4.706184e+10	479999555176	74883482656	26970668447	176077447890	10.895570	11.575058	5.109590
3	3	1.835396e+12	1.835396e+12	0	682133600000	682133600000	0	5.806799	4.479874	4.062512
4	4	4.460361e+11	1.092745e+11	654724761146	384402586718	682133600000	330873669297	5.813927	5.553456	3.367611

Pentru Cluster 1, PIB-ul si exporturile au valori medii si mediane relativ scazute, cu o deviatie standard mare, iar urbanizarea are un nivel mediu-scazut (5.6) cu o variabilitate redusa, ceea ce sugereaza un grup mai omogen din punct de vedere urban.

Pentru Cluster 2, PIB-ul si exporturile au deviatie standrd mare, ceea ce implica o diversitate considerabila. Este cel mai urbanizat cluster (media 10.9), deci implicit cel mai dezvoltat.

Pentru Cluster 3, PIB si exporturi au valori medii si mediane extrem de ridicate (fara variatie), iar asta indica un nivel economic foarte bun si exporturi concentrate. Urbanizare usor scazuta fata de celelalte, dar constanta.

In ceea ce priveste Cluster 4, valorile pentru indicatorii statistici sunt mari, mai ales pentru variatie, ceea ce indica economii active cu o pondere mare a exporturilor. Urbanizarea medie este mai mare decat la Cluster 1 si 2 (5.81), cu o deviatie standard redusa (3.36).

11. Tabel conclusiv:

```
> print(tabel_conclisiv)
```

	Cluster	nr_observatii	pib_medie	pib_mediana	pib_sd	exporturi_medie	exporturi_mediana	exporturi_sd	urbanizare_medie	urbanizare_mediana	urbanizare_sd
1	1	146	9.509685e+10	4.645584e+10	230980835065	53349376110	26970668447	142285375223	5.613713	5.503618	2.230103
2	2	27	2.180149e+11	4.706184e+10	479999555176	74883482656	26970668447	176077447890	10.895570	11.575058	5.109590
3	3	20	1.835396e+12	1.835396e+12	0	682133600000	682133600000	0	5.806799	4.479874	4.062512
4	4	24	4.460361e+11	1.092745e+11	654724761146	384402586718	682133600000	330873669297	5.813927	5.553456	3.367611

```
>
```

Numarul de observatii clasificate in fiecare cluster:

- Cluster 1 – 146 de observatii
- Cluster 2 – 27 de observatii
- Cluster 3 – 20 de observatii
- Cluster 4 – 24 de observatii

Cluster 1 si 2 includ economii mici si medii, cu o diversitate mare in PIB si exporturi, dar mai omogene in urbanizare.

Cluster 3 reprezinta un grup unic, cu economii foarte mari si valori extrem de omogene.

Cluster 4 indica economii mari, dar variate, cu un grad ridicat de urbanizare.

12. Denumirea clusterelor:

- * Cluster 1: Economii tinere si variate
- * Cluster 2: Economii urbane in crestere
- * Cluster 3: Giganti economici centralizati
- * Cluster 4: Centre economice globale

Analiza clusterizarii a permis identificarea a 4 grupuri economice distincte, fiecare cu caracteristici unice care in functie de PIB, exporturi si urbanizare (variabilele pe care am ales sa le analizez pentru a putea face o analiza mai usoara si reprezentativa). Rezultatele analizei sunt evidentiate mai sus.

13. Algoritmul de partitionare K-MEANS CLUSTERING include toti pasii facuti la algoritmul de CLUSTERIZARE IERARHICA, iar in plus avem rezultatele k-means (output) si vizualizarea clusterelor.

Algoritmul K-Means pentru 4 cluster identificate anterior (output):

```
K-means clustering with 4 clusters of sizes 26, 29, 104, 58

Cluster means:
  exporturi (%pib)      pib pib/nr_loc economie_interna (%pib) consum_intern (%pib) investitii_publice (%pib) inflatie_consum
1  0.7846490      2.2084663  0.2590259  1.0754366      0.7758198      0.3248918      1.8790254      -0.09324200
2  1.7816037      -0.3374832  2.0371960  1.0448635      0.5879534      0.2245838      -0.6447400      -0.07769154
3  -0.4836264      -0.1845203  -0.4130330  -0.2232725      0.1642386      0.1454953      -0.2243427      -0.39847860
4  -0.3753489      -0.4903966  -0.3941022  -0.6041734      -0.9362547      -0.5188212      -0.1176820      0.71746572
  investitii_straini (%pib) urbanizare
1  -0.42525398  -0.1723219
2  -0.14672971  -0.3120039
3  -0.06294176  -0.2728491
4   0.37685703  0.7224964

Clustering vector:
[1] 3 3 3 3 3 4 3 2 2 3 3 3 3 1 1 3 3 3 3 3 4 3 2 4 1 3 3 4 4 3 3 2 3 4 3 3 2 2 4 4 4 3 3 3 3 3 1 1 1 1 3 3 3 3 4 4 3 3 1 3 4 3 4 2 2 3 3 4 3 2 4 3 4 3
[79] 3 3 4 3 4 3 4 4 1 1 3 2 2 2 3 1 3 2 2 3 3 3 3 3 2 3 3 4 4 1 4 3 3 3 3 1 1 1 4 4 1 3 4 1 3 3 3 2 3 4 3 1 4 3 4 4 4 3 3 1 3 3 4 3 4 1 3 2 3 4 3 3 3 3 4
[157] 2 2 3 3 2 2 4 4 3 4 2 3 4 1 4 1 3 1 1 4 4 4 4 2 4 3 3 3 3 4 4 2 1 3 3 3 1 3 4 3 3 4 2 3 4 4 4 4 3 2 2 3 3 3 3 3 4 4 4 4 4

within cluster sum of squares by cluster:
[1] 238.0747 233.1062 423.4392 426.6942
(between_SS / total_SS = 38.8 %)

Available components:
[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"    "size"         "iter"         "ifault"
> |
```

Algoritmul a creat 4 cluster de dimensiuni diferite:

- Cluster 1: 26 observatii
- Cluster 2: 29 observatii
- Cluster 3: 104 observatii
- Cluster 4: 58 observatii

Valorile pentru fiecare variabila sunt standardizate, ceea ce inseamna ca sunt exprimate in devieri standard fata de media intregului set de date.

Cluster 1 ale cele mai ridicate valori ale mediei in “exporturi (% pib)” si “investitii_publice (%pib)”, Cluster 2 in “exporturi” si “pib”, Cluster 3 in “economie_interna” si “consum_intern”, insa majoritatea valorilor negative sugereaza economii mai subdezvoltate, iar Cluster 4 este dominat de “inflatie_consum” si “urbanizare”

Suma patratelor distantelor intra-cluster (within cluster sum of squares) indica variabilitatea in interiorul fiecarui cluster. Cluster 3 si Cluster 4 au cele mai mari valori (~ 423 si 426), ceea ce indica o diversitate interna mai mare.

Variabilitatea inter-cluster ($\text{between_SS} / \text{total_SS} = 38.8\%$) arata cat de bine sunt separate clusterurile. O valoare mai mare ar fi indicat o separare mai clara, insa acest procent sugereaza o separare moderata.

14. Indicatorii variabilitatii (BSS, WSS, TSS) si calitatea partiei:

```
> tss <- sum((date_standardizate - rowMeans(date_standardizate))^2)
> wss <- sum(kmeans_result$withinss)
> bss <- tss - wss
> calitate_partitie <- bss / tss
> print(tss)
[1] 1776.243
> print(wss)
[1] 1321.314
> print(bss)
[1] 454.9289
> print(calitate_partitie)
[1] 0.2561186
> |
```

Calitatea partiei ne arata ca aproximativ 25.6% din variabilitatea totala este explicata de gruparea in 4 cluster, ceea ce indica o separare moderata intre cluster.

Indicatorii variabilitatii (BSS, WSS, TSS) si calitatea partiei pentru un interval de 2-6 cluster:

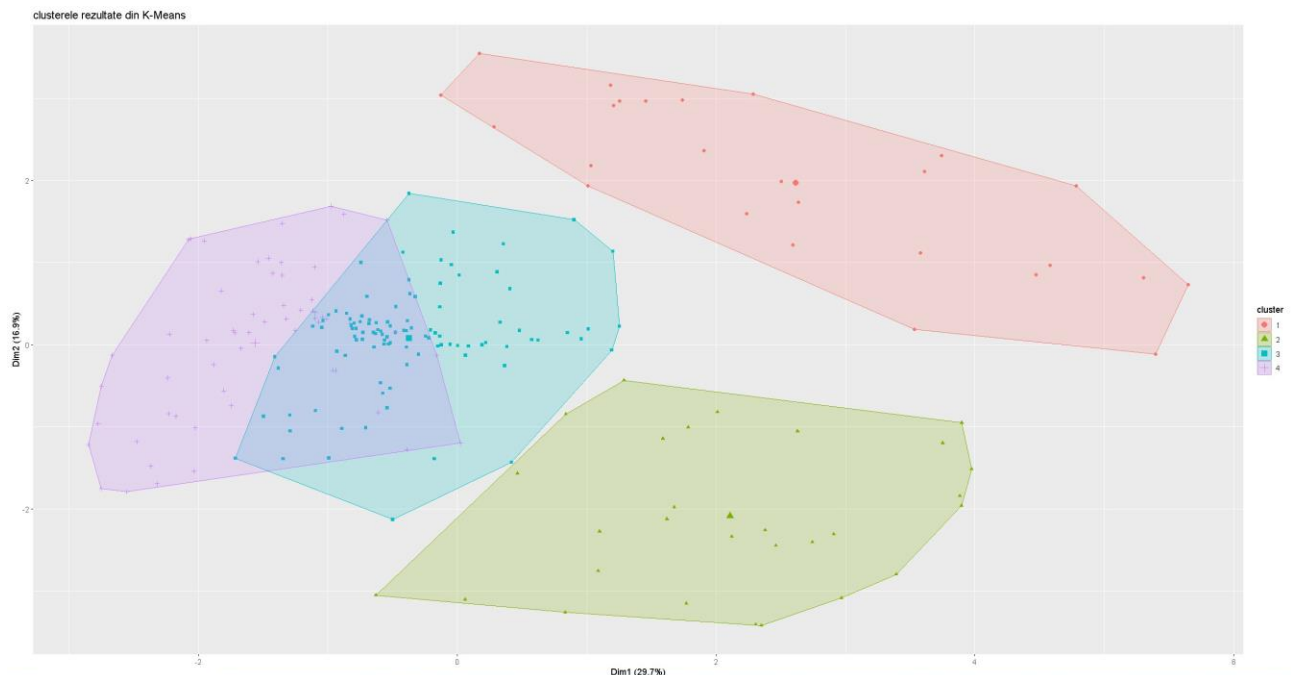
```
> valori_k <- 2:6
> rezultate_k <- data.frame(K = valori_k, BSS_TSS = numeric(length(valori_k)))
> for (k in valori_k) {
+   set.seed(123)
+   test_kmeans <- kmeans(date_standardizate, centers = k, nstart = 25)
+   wss_test <- sum(test_kmeans$withinss)
+   bss_test <- tss - wss_test
+   rezultate_k[rezultate_k$K == k, "BSS_TSS"] <- bss_test / tss
+ }
> print(tss)
[1] 1776.243
> print(wss_test)
[1] 1136.035
> print(bss_test)
[1] 640.2085
> print(rezultate_k)
  K    BSS_TSS
1 2 0.05067518
2 3 0.16645023
3 4 0.25611862
4 5 0.31745796
5 6 0.36042843
> |
```

Pentru un numar diferit de cluster, WSS scade si BSS creste.

Se observa ca pe masura ce numarul de cluster creste, variabilitatea total explicata de model creste de asemenea. Calitatea partiei isi are cea mai buna valoare in k=6 (36%). Cu toate acestea, o valoare prea mare pentru k poate duce la cluster prea mici si nesemnificative.

Alegerea k=4 este o solutie echilibrata intre interpretabilitate si calitatea partiei (BSS/TSS=25.6%)

15. Reprezentarea grafica a clusterlor cu package-ul "factoextra":



16. Am calculat indicatorii statistici (medie, mediana si deviatia standard) pentru k=4 variabile referindu-ma la 5 variabile diferite fata de cele pentru care am calculat indicatorii la Clusterizarea Ierarhica. De data asta pentru a interpreta mai usor output-urile am ales variabilele "economie_interna", "consum_intern", "investitii_publice", "inflatie" si "investitii_straini", deoarece am considerat ca au mare legatura intre ele.

```
> print(indicatorii_statistici_kmeans)
Cluster economie_medie economie_mediana economie_ds consum_medie consum_mediana consum_ds investitii_publice_medie investitii_publice_mediana
1 1 29.04450 27.4971739 11.744302 25.39956 25.06430 7.549686 77.46613 86.56720
2 2 26.95509 24.2601704 7.681676 24.69424 23.01126 7.016376 33.96647 33.92683
3 3 22.24263 21.1872136 6.939090 24.13813 22.90188 6.730572 41.21245 42.17081
4 4 10.00317 0.3242509 10.959596 19.46698 19.34075 6.100632 43.05085 42.17081
investitii_publice_ds inflatie_medie inflatie_mediana inflatie_ds investitii_straini_medie investitii_straine_mediana investitii_straini_ds
1 12.23640 5.404490 4.645024 7.595881 4.693325 4.587857 1.932194
2 10.98190 6.731900 5.472333 7.898567 5.371547 4.172000 2.454997
3 10.78394 3.034128 4.903598 4.598545 5.575576 5.368100 2.207429
4 14.76385 11.700164 8.682477 9.212933 6.646511 6.092125 2.759107
>
```

Cluster 1 sugerează o economie predominant intern (medie 29 și mediană 27.49) cu o variabilitate redusă, cu o dependență puternică de consumul intern (medie 25.40) în care investițiile publice au cel mai ridicat nivel (medie 77.46 și mediană 86.56), ceea ce indică o economie sustinută în general de investiții guvernamentale cu un nivel moderat al investițiilor din partea străinilor.

Cluster 2 indică o economie mai echilibrată între producție și consum (economia internă cu medie și mediană mai scăzută, 29 respectiv 27.49), în care investițiile publice au nivel redus (medie 33.97) și cu un nivel moderat al inflației (medie 6.73), cu variație redusă.

Cluster 3 pune accentul pe piața externă (consum intern scăzut, medie 24.14), în care investițiile publice au un rol important (medie 41.21). De asemenea, se observă un nivel foarte scăzut al inflației (medie 3.03) ceea ce indică stabilitate economică. Investițiile străine au un nivel moderat, deci este o economie deschisă la investiții externe.

Cluster 4 are cel mai scăzut nivel de economie internă (medie 10), ceea ce înseamnă că se bazează foarte mult pe import. Investițiile publice au un nivel ridicat (medie 43.05), indicând o economie susținută în mare parte de stat. Cel mai ridicat nivel al inflației (medie 11.7) și cea mai mare variație a inflației, deci se confruntă cu instabilitate economică, dar este și o economie atractivă pentru investițiile din partea străinilor (medie 6.09), dar și cea mai mare variație a investițiilor (2.75)

16. Tabel concludiv:

cluster	numar_observatii	economie_medie	economie_mediana	economie_ds	consum_medie	consum_mediana	consum_ds	investitii_publice_medie	investitii_publice_mediana	investitii_publice_ds
1	26	29.04450	27.4971739	11.744302	25.39956	25.06430	7.549686	77.46613	86.56720	12.23640
2	29	26.95509	24.2601704	7.681676	24.69424	23.01126	7.016376	33.96647	33.92683	10.98190
3	104	22.24263	21.1872136	6.939090	24.13813	22.90188	6.730572	41.21245	42.17081	10.78394
4	58	10.00317	0.3242509	10.959596	19.46698	19.34075	6.100632	43.05085	42.17081	14.76385

Am obținut un număr de 4 clustere care includ:

- Cluster 1: 26 observatii
- Cluster 2: 29 observatii
- Cluster 3: 104 observatii
- Cluster 4: 58 observatii

Observatii:

Cluster 1 reprezintă economii auto-susținute, Cluster 2 indică economii echilibrate între consumul intern și piața externă, Cluster 3 este orientat spre piața externă cu un nivel ridicat de deschidere, iar Cluster 4 reflectă economii globalizate și urbanizate cu mare intervenție din partea statului și o atracție puternică a investitorilor străini, dar cu riscuri economice mai mari (inflație ridicată).

18. Denumirea clusterlor:

- * Cluster 1: Economii interne consolidate
- * Cluster 2: Economii echilibrate si stabile
- * Cluster 3: Economii orientate spre export
- * Cluster 4: Economii globale urbanizate

Analiza clusterizarii evidentiaza patru tipare economice distincte care variaza de la economii interne consolidate si echilibrate, la economii orientate spre export si economii globalizate urbane si ofera o perspectiva clara asupra diversitatii strategiilor economice.

Cod folosit in R pentru realizarea temei (incepand cu cerinta 7):

```
#am incarcato setul de date prelucrat
```

```
date <- read.csv('set_de_date_prelucrat.csv')
```

```
colnames(date) <- c("nume_tara", "cod_tara", "exporturi", "exporturi (%pib)", "pib",  
  "pib/nr_loc", "economie_interna (%pib)", "consum_intern (%pib)",  
  "investitii_publice (%pib)", "inflatie_consum",  
  "investitii_straini (%pib)", "urbanizare")
```

```
# lucrez doar cu coloanele numerice
```

```
coloane_numerice <- date[, c("exporturi", "exporturi (%pib)", "pib", "pib/nr_loc",  
  "economie_interna (%pib)", "consum_intern (%pib)",  
  "investitii_publice (%pib)", "inflatie_consum",
```

```
"investitii_straini (%pib)", "urbanizare"]]
```

```
#distanța dintre observații prin metoda euclidiană
```

```
distanțe_euclidiene <- dist(coloane_numerice, method = "euclidean")
```

```
#metoda Manhattan
```

```
distanțe_manhattan <- dist(coloane_numerice, method = "manhattan")
```

```
#matricea distanțelor
```

```
matrice_eucl <- as.matrix(distanțe_euclidiene)
```

```
matrice_manh <- as.matrix(distanțe_manhattan)
```

```
print("Distanțe euclidiene pentru primele 10 observații:")
```

```
print(matrice_eucl[1:10, 1:10])
```

```
print("Distanțe Manhattan pentru primele 10 observații:")
```

```
print(matrice_manh[1:10, 1:10])
```

```
date_standardizate <- scale(coloane_numerice)
```

```
#clusterizare completă
```

```
clusterizare_completa <- hclust(dist(date_subset, method = "euclidean"), method = "complete")
```

```
print(clusterizare_completa)
```

```
#clusterizare prin metoda Ward
```



```
clusterizare_ward <- hclust(dist(date_subset, method = "euclidean"), method = "ward.D2")
print(clusterizare_ward)

date_subset <- date_standardizate[1:50, ]

#dendograme
par(mfrow = c(1, 2))

#dendograma metoda completa
plot(clusterizare_completa, main = "Dendograma - metoda completa", xlab = "", sub = "", cex = 0.6)
par(mfrow = c(1, 2))

#dendograma metoda ward
plot(clusterizare_ward, main = "Dendograma - metoda ward", xlab = "", sub = "", cex = 0.6)

#numarul optim de clustere

#metoda elbow
install.packages("factoextra")
library(factoextra)

fviz_nbclust(date_standardizate, FUN = hcut, method = "wss") +
  ggtitle("Indicele Elbow pentru numarul optim de clustere")

#indicele Silhouette

fviz_nbclust(date_standardizate, FUN = hcut, method = "silhouette") +
  ggtitle("Indicele Silhouette pentru numărul optim de clustere")

#calcul gap statistic
install.packages("cluster")
```

```
library(cluster)

set.seed(123)

gap_stat <- clusGap(date_standardizate, FUN = hcut, nstart = 25, K.max = 10, B = 50)

fviz_gap_stat(gap_stat) +
  ggtitle("Gap Statistic pentru numarul optim de clustere")


#adaugam numarul de 4 clustere la date

hclust_ward <- hclust(dist(date_standardizate, method = "euclidean"), method = "ward.D2")

clustere <- cutree(hclust_ward, k = 4)

date_cu_clustere <- cbind(date, Cluster = clustere)


install.packages("dplyr")

library(dplyr)

#indicatori statistici clustere


pib_medie <- tapply(date_cu_clustere$pib, date_cu_clustere$Cluster, mean, na.rm = TRUE)

pib_mediana <- tapply(date_cu_clustere$pib, date_cu_clustere$Cluster, median, na.rm = TRUE)

pib_sd <- tapply(date_cu_clustere$pib, date_cu_clustere$Cluster, sd, na.rm = TRUE)


exporturi_medie <- tapply(date_cu_clustere$exporturi, date_cu_clustere$Cluster, mean, na.rm = TRUE)

exporturi_mediana <- tapply(date_cu_clustere$exporturi, date_cu_clustere$Cluster, median, na.rm =
TRUE)

exporturi_sd <- tapply(date_cu_clustere$exporturi, date_cu_clustere$Cluster, sd, na.rm = TRUE)


urbanizare_medie <- tapply(date_cu_clustere$urbanizare, date_cu_clustere$Cluster, mean, na.rm =
TRUE)
```

```
urbanizare_mediana <- tapply(date_cu_clustere$urbanizare, date_cu_clustere$Cluster, median, na.rm = TRUE)
```

```
urbanizare_sd <- tapply(date_cu_clustere$urbanizare, date_cu_clustere$Cluster, sd, na.rm = TRUE)
```

```
rezultate <- data.frame(  
  Cluster = 1:4,  
  pib_medie = pib_medie,  
  pib_mediana = pib_mediana,  
  pib_sd = pib_sd,  
  exporturi_medie = exporturi_medie,  
  exporturi_mediana = exporturi_mediana,  
  exporturi_sd = exporturi_sd,  
  urbanizare_medie = urbanizare_medie,  
  urbanizare_mediana = urbanizare_mediana,  
  urbanizare_sd = urbanizare_sd  
)
```

```
#tabel conclusiv
```

```
numar_observatii = table(date_cu_clustere$Cluster)
```

```
tabel_conclusiv <- data.frame(  
  Cluster = names(numar_observatii),  
  nr_observatii = as.numeric(numar_observatii),  
  pib_medie = pib_medie,  
  pib_mediana = pib_mediana,  
  pib_sd = pib_sd,  
  exporturi_medie = exporturi_medie,
```

```
exporturi_mediana = exporturi_mediana,  
exporturi_sd = exporturi_sd,  
urbanizare_medie = urbanizare_medie,  
urbanizare_mediana = urbanizare_mediana,  
urbanizare_sd = urbanizare_sd  
)
```

```
print(tabel_conclusiv)
```

```
#algoritmul k-means
```

```
set.seed(123)
```

```
kmeans_result <- kmeans(date_standardizate, centers = 4, nstart = 25)
```

```
print(kmeans_result)
```

```
data_cu_kmeans <- cbind(date, Cluster = kmeans_result$cluster)
```

```
fviz_cluster(kmeans_result, data = date_standardizate, geom = "point", ellipse.type = "convex") +  
  ggtitle("clusterelor rezultate din K-Means")
```

```
#tss
```

```
tss <- sum((date_standardizate - rowMeans(date_standardizate))^2)
```

```
wss <- sum(kmeans_result$withinss)
```

```
bss <- tss - wss
```

```
calitate_partitie <- bss / tss
```

```
print(tss)
```

```

print(wss)

print(bss)

print(calitate_partitie)


#alegem un alt numar de clustere

valori_k <- 2:6

rezultate_k <- data.frame(K = valori_k, BSS_TSS = numeric(length(valori_k)))

for (k in valori_k) {

  set.seed(123)

  test_kmeans <- kmeans(date_standardizate, centers = k, nstart = 25)

  wss_test <- sum(test_kmeans$withinss)

  bss_test <- tss - wss_test

  rezultate_k[rezultate_k$K == k, "BSS_TSS"] <- bss_test / tss

}

print(tss)

print(wss_test)

print(bss_test)

print(rezultate_k)

```

```

#indicatori statistici (media, mediana si deviatie standard)

```

```

#economie interna

```

```

economie_medie <- tapply(data_cu_kmeans$`economie_interna (%pib)`, data_cu_kmeans$Cluster,
mean, na.rm = TRUE)

```

```

economie_mediana <- tapply(data_cu_kmeans$`economie_interna (%pib)`, data_cu_kmeans$Cluster,
median, na.rm = TRUE)

```

```

economie_ds <- tapply(data_cu_kmeans$`economie_interna (%pib)`, data_cu_kmeans$Cluster, sd,
na.rm = TRUE)

```

```
#consum intern
```

```
consum_medie <- tapply(data_cu_kmeans$`consum_intern (%pib)` , data_cu_kmeans$Cluster, mean,  
na.rm = TRUE)
```

```
consum_mediana <- tapply(data_cu_kmeans$`consum_intern (%pib)` , data_cu_kmeans$Cluster,  
median, na.rm = TRUE)
```

```
consum_ds <- tapply(data_cu_kmeans$`consum_intern (%pib)` , data_cu_kmeans$Cluster, sd, na.rm =  
TRUE)
```

```
#investitii publice
```

```
investitii_publice_medie <- tapply(data_cu_kmeans$`investitii_publice (%pib)` ,  
data_cu_kmeans$Cluster, mean, na.rm = TRUE)
```

```
investitii_mediana <- tapply(data_cu_kmeans$`investitii_publice (%pib)` , data_cu_kmeans$Cluster,  
median, na.rm = TRUE)
```

```
investitii_ds <- tapply(data_cu_kmeans$`investitii_publice (%pib)` , data_cu_kmeans$Cluster, sd, na.rm =  
TRUE)
```

```
#inflatie
```

```
inflatie_medie <- tapply(data_cu_kmeans$inflatie_consum, data_cu_kmeans$Cluster, mean, na.rm =  
TRUE)
```

```
inflatie_mediana <- tapply(data_cu_kmeans$inflatie_consum, data_cu_kmeans$Cluster, median, na.rm  
= TRUE)
```

```
inflatie_ds <- tapply(data_cu_kmeans$inflatie_consum, data_cu_kmeans$Cluster, sd, na.rm = TRUE)
```

```
#investitii straini
```

```
investitii_straini_medie <- tapply(data_cu_kmeans$`investitii_straini (%pib)` , data_cu_kmeans$Cluster,  
mean, na.rm = TRUE)
```

```
investitii_straini_mediana <- tapply(data_cu_kmeans$`investitii_straini (%pib)` ,  
data_cu_kmeans$Cluster, median, na.rm = TRUE)
```

```
investitii_straini_ds <- tapply(data_cu_kmeans$`investitii_straini (%pib)`, data_cu_kmeans$Cluster, sd,  
na.rm = TRUE)
```

```
indicatori_statistici_kmeans <- data.frame(  
  Cluster = names(economie_medie),  
  economie_medie = economie_medie,  
  economie_mediana = economie_mediana,  
  economie_ds = economie_ds,  
  consum_medie = consum_medie,  
  consum_mediana = consum_mediana,  
  consum_ds = consum_ds,  
  investitii_publice_medie = investitii_publice_medie,  
  investitii_publice_mediana = investitii_mediana,  
  investitii_publice_ds = investitii_ds,  
  inflatie_medie = inflatie_medie,  
  inflatie_mediana = inflatie_mediana,  
  inflatie_ds = inflatie_ds,  
  investitii_straini_medie = investitii_straini_medie,  
  investitii_straine_mediana = investitii_straini_mediana,  
  investitii_straini_ds = investitii_straini_ds  
)
```

```
print(indicatori_statistici_kmeans)
```

```
#tabel conclusiv
```

```
num_observatii <- table(data_cu_kmeans$Cluster)
```

```
tabel_conclusiv <- data.frame(  
  cluster = names(num_observatii),  
  numar_observatii = as.numeric(num_observatii),  
  economie_medie = economie_medie,  
  economie_mediana = economie_mediana,  
  economie_ds = economie_ds,  
  consum_medie = consum_medie,  
  consum_mediana = consum_mediana,  
  consum_ds = consum_ds,  
  investitii_publice_medie = investitii_publice_medie,  
  investitii_publice_mediana = investitii_mediana,  
  investitii_publice_ds = investitii_ds,  
  inflatie_medie = inflatie_medie,  
  inflatie_mediana = inflatie_mediana,  
  inflatie_ds = inflatie_ds,  
  investitii_straini_medie = investitii_straini_medie,  
  investitii_straine_mediana = investitii_straini_mediana,  
  investitii_straini_ds = investitii_straini_ds  
)
```

```
View(tabel_conclusiv)
```