Name :Amol Balasaheb Wadekar

ASSIGNMENT 1:PCA

PRINCIPAL COMPONENT ANALYSIS

We have Educational data of 107 countries. In which there are 9 different variables

And we want to study which variables are contributing more in this data set.

So instead of analysing by 9 different dimensions we can analyse the data in less dimensions by reducing the dimensions of data by PCA technique

Country Name	PPT	GDP	PRPE	OOCP	ESE
Albania	2893654	13219857459	0.73	7097	333291
Arab World	384222592	2.89E+12	6.42	6827541	30972246
United Arab Emirates	9086139	4.02E+11	0.22	14611	411040
Azerbaijan	9535079	75198010965	0.16	22821	949294
Burundi	10816860	3093647227	24.25	69246	583308
Belgium	11231213	5.32E+11	2.49	6538	1210112
Benin	10598482	9707432016	10.84	70228	896763
Burkina Faso	17589198	12257141798	6.87	956718	841886
Bulgaria	7223938	56732006512	0.1	9514	518914
Bosnia and Herzegovina	3817554	18523988467	0.13	1668	297460
Belarus	9483000	76103961203	0.03	22130	648541
Belize	351706	1717861730	6.87	397	37095
Bhutan	765008	1958819915	6.25	10967	74227
Central Europe	103496179	1.46E+12	1.04	218150	7879276
Switzerland	8188649	7.03E+11	1.24	1671	616127
Chile	17762647	2.59E+11	4.05	101222	1556372
Cote d'Ivoire	22157107	34217693110	15.01	878850	1418361
Cameroon	22773014	32050817633	12.73	193251	2000076
Colombia	47791393	3.78E+11	1.57	309168	4827962
Cabo Verde	513906	1881496592	9.02	1041	59823
Costa Rica	4757606	50167623290	4.18	15638	460235
Cyprus	1153658	23308212817	0.13	1535	58634
Germany	80982500	3.88E+12	0.41	4282	7201072
Denmark	5643475	3.46E+11	0.96	7401	553791
Dominican Republic	10405943	65231032303	6.98	178919	931068
E. Asia & Pacific(-hi)	202065603 7	1.27E+13	0.75	6301808	139756448
Early-demographic	307900194	1.03E+13	3.4	2086879	272891456

	9			6	
	226406435				
East Asia & Pacific	1	2.16E+13	0.74	6488666	154212112
Europe&C. Asia(-hi)	408984246	3.87E+12	0.76	1110253	36413404
Europe &Central Asia	903987814	2.34E+13	1.23	1423031	78143392
Ecuador	15902916	1.02E+11	1.35	63225	1877666
Euro area	338617134	1.34E+13	2.23	189071	28055416
Spain	46480882	1.38E+12	2.32	28642	3288424
European Union	508344735	1.86E+13	1.59	409889	42786516
				1673666	
Conflict affected	474395380	8.10E+11	11.07	9	29457496
Finland	5461512	2.72E+11	0.33	1721	536925
Ghana	26786598	38616536132	2.65	413314	2265692
Guinea	12275527	6624068016	13.64	416632	715702
Guatemala	16015494	58722087392	9.31	257525	1165624
	118059744				
High income	5	5.01E+13	1.44	2309701	90841792
Hong Kong SAR, China	7241700	2.91E+11	0.53	1913	415971
Honduras	7961680	19380958759	1.64	53956	619832
				2258805	
Poor countries (HIPC)	701714118	6.40E+11	10.05	2	38850524
Croatia	4238389	57136241867	0.19	2676	370356
Hungary	9866468	1.39E+11	1.73	15945	857807
	448065659			1562890	
IBRD only	0	2.68E+13	2.51	8	373634784
	610446102			5866428	
IDA & IBRD total	4	2.92E+13	4.1	0	478212864
	162380443			4303537	
IDA total	4	2.37E+12	7.05	2	104578088
				1659803	
IDA blend	573991143	1.28E+12	2.15	2	39283708
Indonesia	254454778	8.90E+11	2.21	2008214	22586956
	104981329			2643734	
IDA only	1	1.08E+12	9.1	0	65294380
Iran, Islamic Rep.	78143644	4.25E+11	1.25	44743	5794537
Israel	8215700	3.09E+11	1.38	24274	767729
Kazakhstan	17289224	2.21E+11	0.04	5110	1661586
Kyrgyz Republic	5835500	7468096567	0.03	8148	650516
Latin America(-hi)	598846559	5.75E+12	4.79	3480099	61868840
Lao PDR	6689300	11739027121	6.92	36499	600747
Liberia	4396554	2013000000	6.04	442455	222857
Latin					
America&Caribbean	626285603	6.23E+12	4.77	3633176	64141980
Least developed				2451226	
countries	931955068	9.55E+11	9.59	0	55732712
				1942607	
Low income	621322015	4.06E+11	10.24	2	35371836

Lower middle income	2885248344	5.89E+12	3.1	3037660	239921488
Low & middle income	6080182833	2.83E+13	4.09	5863375	477174112
Lesotho	2109197	2538831766	8.75	65881	130836
Late-demographic dividend	2235786397	1.96E+13	1.5	5386553	164160592
Lithuania	2932367	48353937110	0.54	561	277349
Latvia	1993782	31322649595	0.76	3395	121580
Middle East & North Africa	416401841	3.54E+12	5.88	2458452	35294308
Mexico	125385833	1.30E+12	0.82	340062	12993322
Middle income	5458860818	2.79E+13	2.88	3920768	441802272
Mali	17086022	14004067516	20.06	1029672	961046
Malta	427364	10737561364	0.48	841	30230
Myanmar	53437159	65574726566	0.2	284278	3191268
M.East&N.Africa(-hi)	356262920	1.57E+12	6.56	2285200	29813136
Mongolia	2909871	12226514746	0.05	10311	285597
Mozambique	27216276	16961127046	6.41	677573	783558
Mauritania	3969625	5442297174	3.49	150634	178968
Mauritius	1260934	12803445934	0.98	3539	133001
Niger	19113728	8245312137	3.85	1233332	515075
Nepal	28174724	19811974852	10.22	173764	3163946
OECD members	1274981875	4.89E+13	1.39	2779138	107619152
Oman	4236057	81035110533	0	5713	391099
Other small states	28142985	4.43E+11	11.51	511619	2021330
Pakistan	185044286	2.44E+11	2.28	5611792	11286628
Peru	30973148	2.01E+11	4.2	161695	2670847
Pre-demographic dividend	826352062	1.48E+12	8.81	3263543 0	46566744
Russian Federation	143819569	2.05E+12	0.38	220011	9061324
South Asia	1721152580	2.58E+12	3.23	1132241 0	153192192
Saudi Arabia	30886545	7.54E+11	0.96	120782	3419441
El Salvador	6107706	25054200000	6.06	40769	625060
Serbia	7130576	44210806366	0.45	3975	548158
Saharan Africa(-hi)	974280491	1.77E+12	8.18	3413398 8	56130096
Sub-Saharan Africa	974371891	1.77E+12	8.18	3413443 6	56136952
Small states	37479355	5.21E+11	9.96	579860	2841358
Slovenia	2061980	49530147016	0.78	2454	145179
Sweden	9696110	5.74E+11	0	1787	826694
East Asia&Pacific(IDA)	199558400	1.27E+13	0.75	6246218	137370112

	7				
Eur &C. Asia(IDA)	451234370	4.47E+12	0.77	1174909	39439656
Latin America(IDA)	610149938	6.02E+12	4.83	3533954	62970044
M. East&N. Africa(IDA)	351968238	1.56E+12	6.62	2252351	29103910
Timor-Leste	1212107	1399000000	17.75	3869	118935
South Asia(IDA)	1721152580	2.58E+12	3.23	11322410	153192192
Saharan Africa(IDA)	974371891	1.77E+12	8.18	34134436	56136952
Ukraine	45362900	1.34E+11	0.06	56464	2713646
Upper middle income	2573612474	2.20E+13	2.55	8831078	201880784
West Bank and Gaza	4294682	12715599989	0.2	32849	709226
World	7260780278	7.84E+13	3.81	60943456	568015872

EPE	UNEMP	LEB	TDP
195720	16.1	77.83	5
45126932	11.52	70.57	6
409776	3.6	77.37	5
517708	5.2	70.76	4
2046794	6.9	56.69	6
773568	8.5	80.59	6
2133330	1	59.51	6
2594024	3.1	58.59	6
258840	11.6	75.41	4
161023	27.9	76.43	5
368783	5.9	72.98	4
51978	11.5	70.08	6
101667	2.8	69.47	7
5122451	9.06	76.56	5
483886	4.5	82.85	6
1469099	6.4	81.5	6
3176874	4	51.56	6
4142775	4.3	55.49	6
4542644	10.1	73.99	5
67023	9.2	73.15	6
475731	8.3	79.4	6
53129	15.6	80.13	6
2862690	5	80.84	4
467484	6.6	80.55	7
1267930	15	73.5	6
167582400	4.58	73.99	6

355815200	5.66	69.04	6
180388848	4.52	74.93	6
21083246	7.62	71.89	4
49350704	9.01	76.89	5
2067545	4.6	75.87	6
18148330	11.53	81.65	6
2960626	24.7	83.08	6
28606016	10.21	80.67	6
71616608	7.89	61.72	6
351766	8.6	81.13	6
4117152	2.4	61.31	6
1729630	1.8	58.73	6
2417429	2.9	71.72	6
75883424	7.37	80.58	6
324171	3.2	83.98	6
1150042	3.9	73.14	6
118997400	6.6	60.79	6
160819	16.7	77.33	4
393020	7.8	75.87	4
419335712	5.61	72.05	6
644414208	5.67	69.75	6
225078480	5.87	63.41	6
66395416	5.25	63.32	6
29838440	6.2	68.89	6
158683072	6.19	63.46	6
7441078	12.8	75.39	6
861699	6.1	82.15	6
1122282	4.1	71.62	4
434642	8.1	70.4	4
63043688	6.54	74.72	6
870893	1.4	66.12	5
683977	3.8	60.83	6
65342340	6.58	74.94	6
145810064	6.13	63.19	6
105538408	5.98	61.27	6
331002304	5.25	67.25	6
643171328	5.64	69.68	6
366048	26.2	49.7	7
158910240	5.1	75.12	6
108115	11.3	73.97	4
114623	10	74.19	6
47857856	11.26	72.85	6
14627368	4.9	76.72	6
537632896	5.6	70.64	6

2181617	8.1	57.99	6
24072	5.9	81.75	6
5177276	3.3	65.86	5
42118076	12.88	72.27	6
239289	4.8	69.46	5
5669591	22.6	55.03	7
592249	31	63.02	6
105447	7.7	74.19	6
2277021	5.1	61.46	6
4401780	2.7	69.6	5
89315032	7.34	80.12	6
233306	7.2	77.09	4
3619640	10.99	66.45	6
19431564	5.2	66.18	5
3495614	4.2	74.53	6
136351696	7.78	58.36	6
5725528	5.1	70.37	4
191118304	3.9	68.12	5
3737083	5.6	74.34	6
777043	6.2	72.75	6
284754	22.2	75.53	4
158225632	7.97	58.56	6
158234432	7.97	58.56	6
4699990	10.72	67.84	6
111648	9.5	80.52	6
757164	8	81.96	6
165744688	4.59	74.03	6
23412538	7.83	72.4	4
64228640	6.6	74.83	6
41675600	12.76	72.26	6
245847	4.7	68.26	6
191118304	3.9	68.12	5
158234432	7.97	58.56	6
1685030	7.7	71.19	4
206630624	5.9	74.44	6
442476	26.2	72.9	4
719054784	5.93	71.45	6

Principal Component Analysis By Using Covariance Matrix

Principal Component Analysis: PPT, GDP, PRPE, OOCP, ESE, EPE, UNEMP, LEB, TDP

Eigen analysis of the Covariance Matrix

Eigenvalue	1.38974E+26	7.83098E+17	3.87239E+14	7.09948E+13	3.84865E+12	115
Proportion	1.000	0.000	0.000	0.000	0.000	0.000
Cumulative	1.000	1.000	1.000	1.000	1.000	1.000

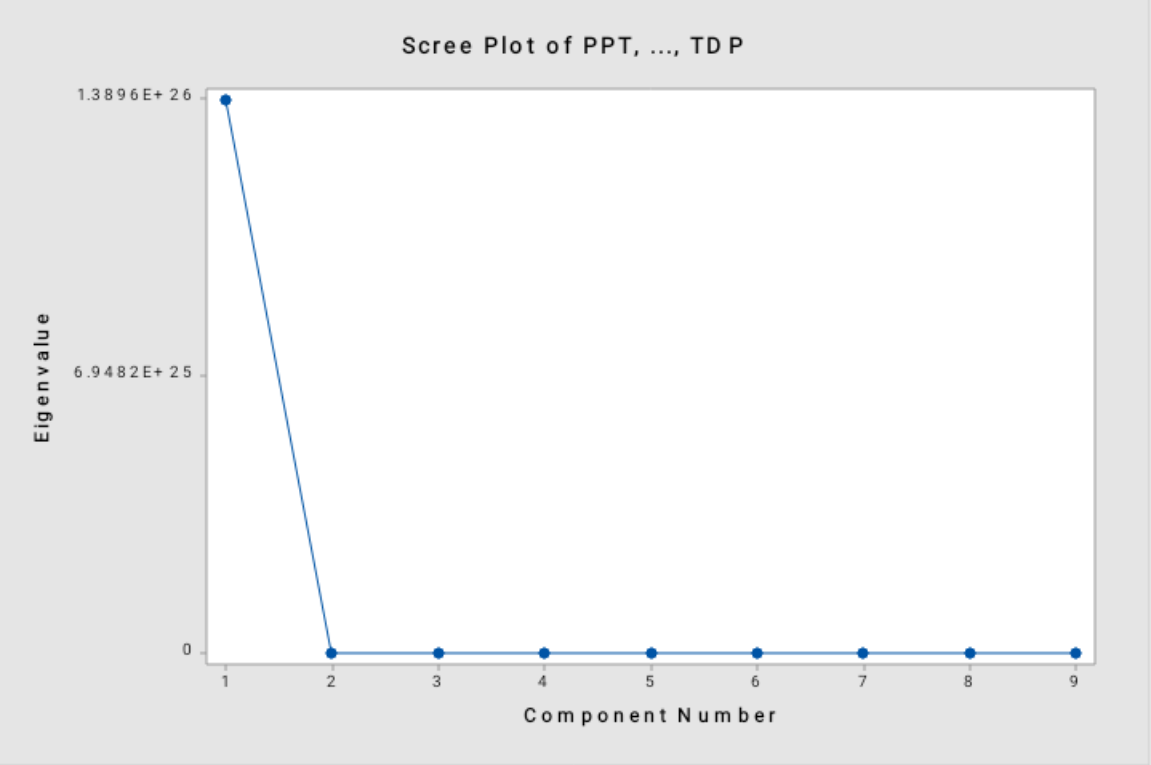
Eigenvalue	19	0	-6103011388
Proportion	0.000	0.000	-0.000
Cumulative	1.000	1.000	1.000

By proportion of eigen value we select only 1 components to explain the Educational data. This one component explain 100% variation of the multivariate data.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
PPT	0.000	0.991	0.093	0.100	0.001	-0.000	0.000	0.000	-0.000
GDP	1.000	-0.000	-0.000	-0.000	0.000	0.000	-0.000	-0.000	-0.000
PRPE	-0.000	0.000	-0.000	0.000	0.000	0.000	-0.014	0.997	-0.070
OOCP	0.000	0.010	-0.345	0.226	-0.911	-0.000	0.000	0.000	0.000
ESE	0.000	0.078	0.176	-0.935	-0.298	0.000	-0.000	0.000	0.000
EPE	0.000	0.111	-0.917	-0.255	0.286	-0.000	-0.000	-0.000	-0.000
UNEMP	-0.000	-0.000	-0.000	-0.000	0.000	-0.000	1.000	0.014	-0.000
LEB	0.000	-0.000	0.000	-0.000	0.000	-1.000	-0.000	0.000	-0.000
TDP	0.000	0.000	-0.000	0.000	0.000	0.000	-0.001	0.070	0.998

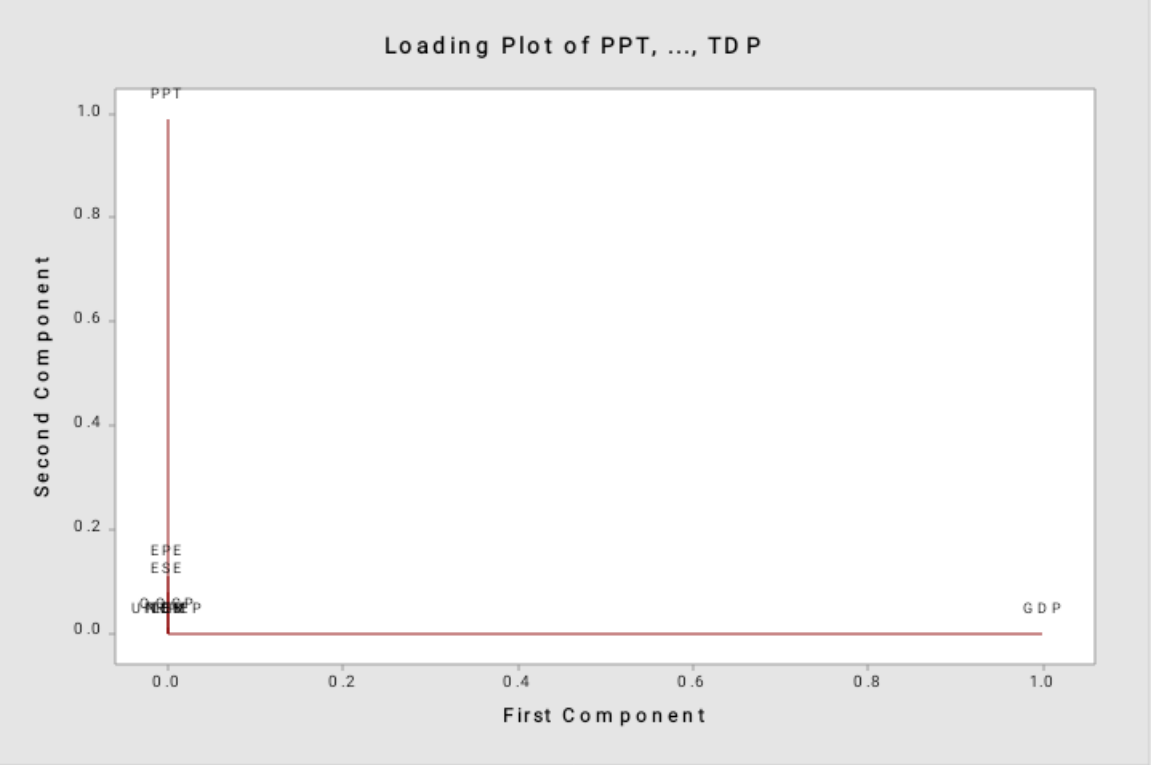
From above table, in 1st component the PC1 Variate contributes all variation in given educational data which is contribute totally by GDP and it is positive.

Name :Amol Balasaheb Wadekar



Interpretation: From scree plot we can see the elbow point at component number 2. So we select 1 principal components.

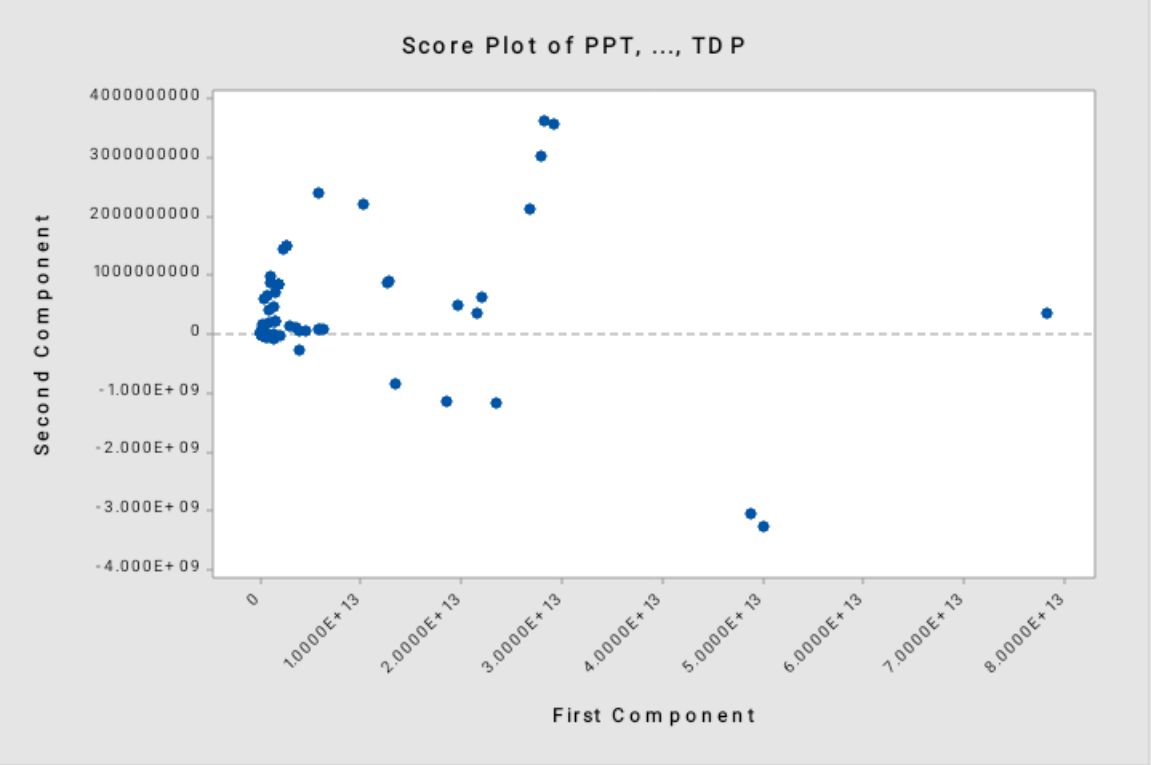
Name :Amol Balasaheb Wadekar



From loading plot we come to know that GDP contributing more in 1st component.

PPT ,EPE,ESE loading on 2nd component.

Name :Amol Balasaheb Wadekar



INFERENCE

After PCA we come to know that the GDP (PC1) variates we should study to analyse the data .

Principal Component Analysis By Using Correlation Matrix

Principal Component Analysis: PPT, GDP, PRPE, OOCP, ESE, EPE, UNEMP, LEB, TDP

Eigen analysis of the Correlation Matrix

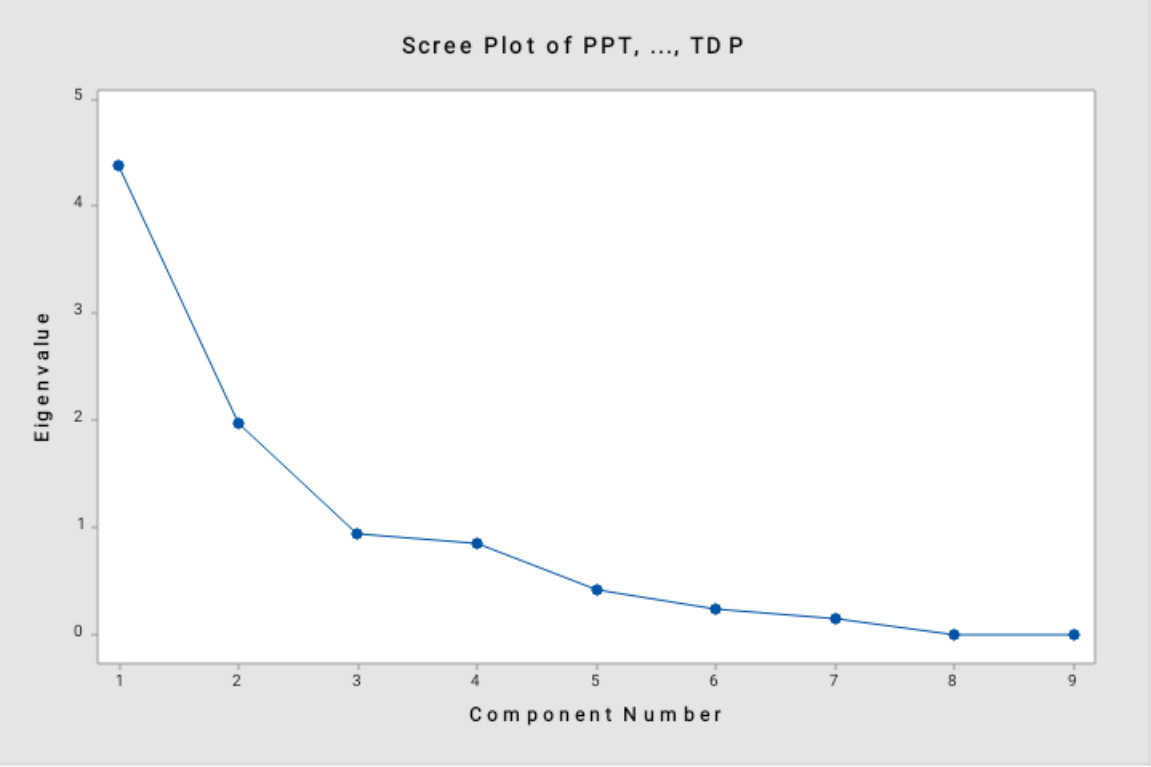
Eigenvalue	4.3895	1.9744	0.9536	0.8557	0.4202	0.2464	0.1568	0.0022	0.0012
Proportion	0.488	0.219	0.106	0.095	0.047	0.027	0.017	0.000	0.000
Cumulative	0.488	0.707	0.813	0.908	0.955	0.982	1.000	1.000	1.000

By proportion of eigen value we select first 2 principal components to explain the Educational DATA. These 2 components explain 70.07% variation of the multivariate data.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
PPT	0.470	0.071	-0.026	-0.041	-0.007	-0.039	0.326	0.804	-0.131
GDP	0.372	0.202	-0.052	0.228	0.720	0.294	-0.396	-0.051	-0.037
PRPE	0.014	-0.645	-0.008	0.013	0.442	-0.622	0.033	0.009	0.002
OOCP	0.415	-0.139	-0.054	-0.234	-0.408	-0.178	-0.711	0.066	0.207
ESE	0.466	0.085	-0.024	-0.034	0.014	-0.049	0.436	-0.351	0.677
EPE	0.471	0.013	-0.034	-0.099	-0.147	-0.109	0.173	-0.472	-0.694
UNEMP	-0.111	0.103	-0.973	-0.152	0.048	-0.064	0.027	0.003	0.000
LEB	-0.053	0.605	0.043	0.422	-0.053	-0.663	-0.094	0.007	0.000
TDP	0.120	-0.367	-0.208	0.824	-0.304	0.189	0.016	-0.001	0.006

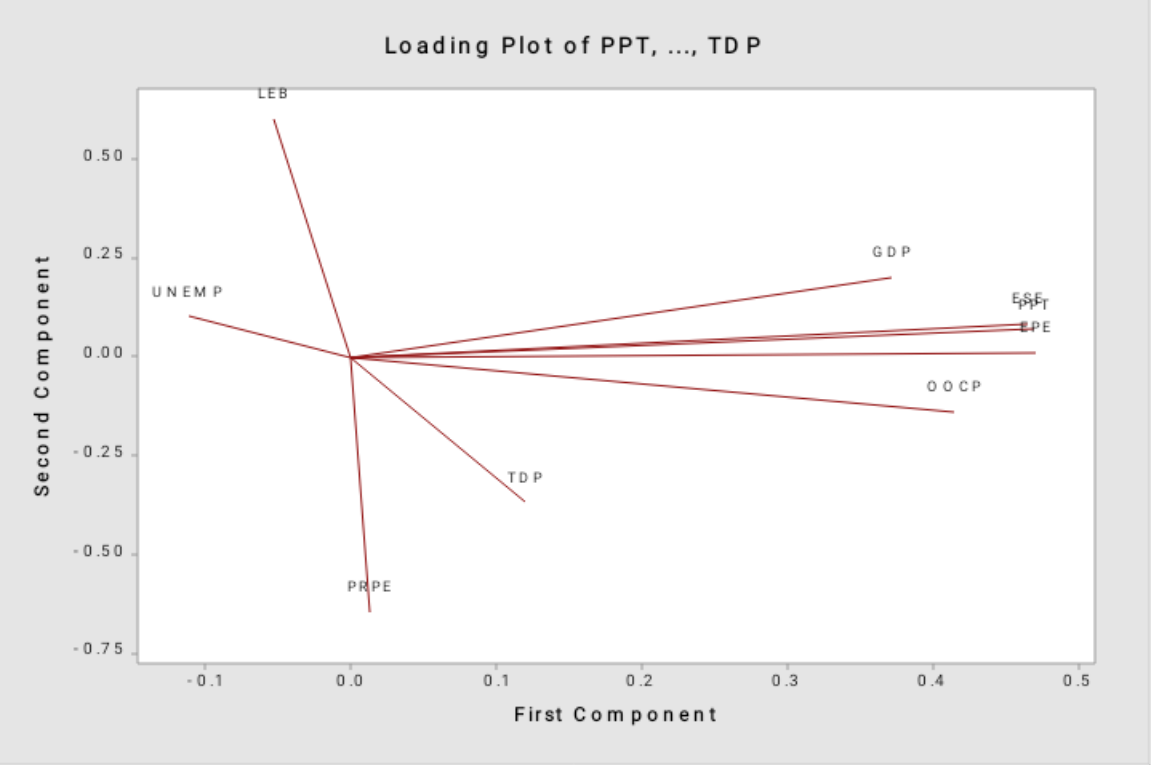
From above table, in 1st component the PPT, GDP,PREP,ESE And EPE variate contributes more.

In 2nd component the PREP and LEB contributes more.



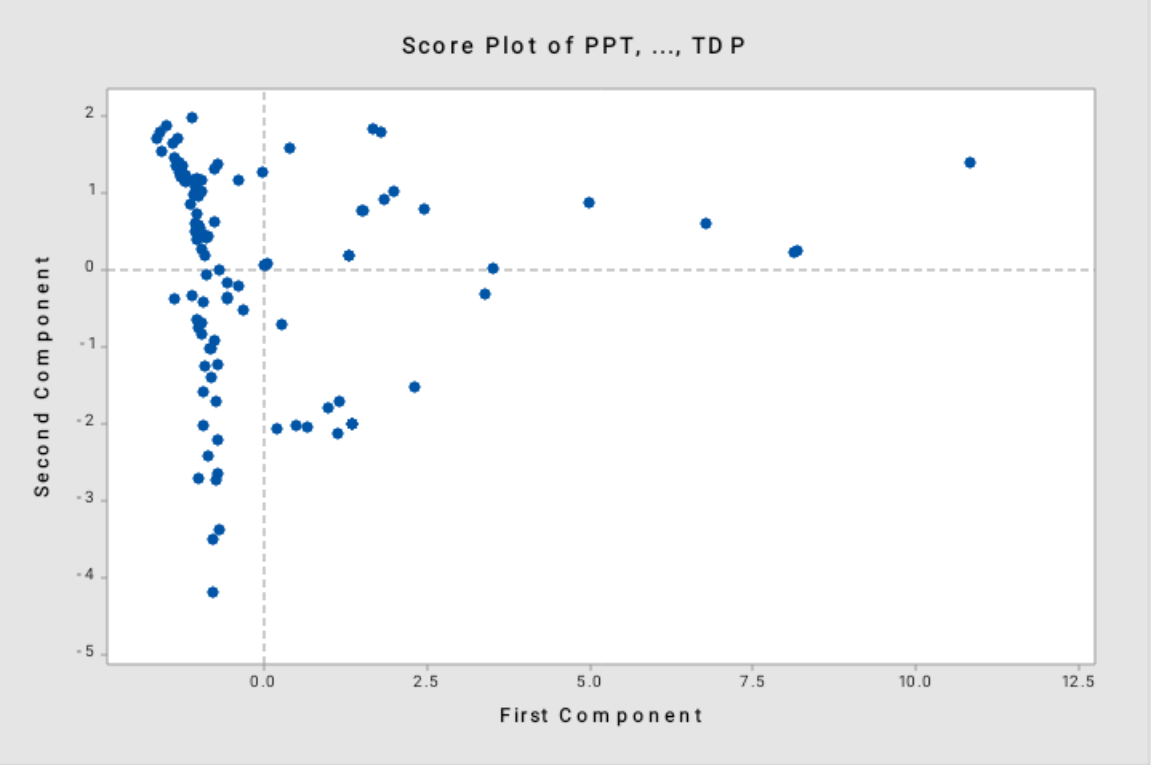
Interpretation:

From scree plot we can see the elbow point at component number 3. So we select first 2 principal components.



From loading plot we come to know that GDP,ESE,EPE,PPT contributing more in 1st component.

PRPE and LEB variates has large loading on 2nd component.



INFERENCE

After PCA we come to know that the PPT, GDP,ESE ,EPE, PREP and LEB (PC1 and PC2) variates we should study to analyse the data and then to get more information we can look at 3nd and 4rd component to make more reliable estimates.

Conclusion :Here principal component analysis by using covariance matrix method is better than correlation matrix method because in PCA by covariance method we get (one PC) single variable GDP which contibute all 100% variation in data so all other variables have no significance but in PCA by correlation method we get 2 PC's (6 variables) which contibute all 70.07% variation in data .