

# **Solution**

## **Numerical Assignment**

### **Multivariate Analysis**

**Submitted by:-**

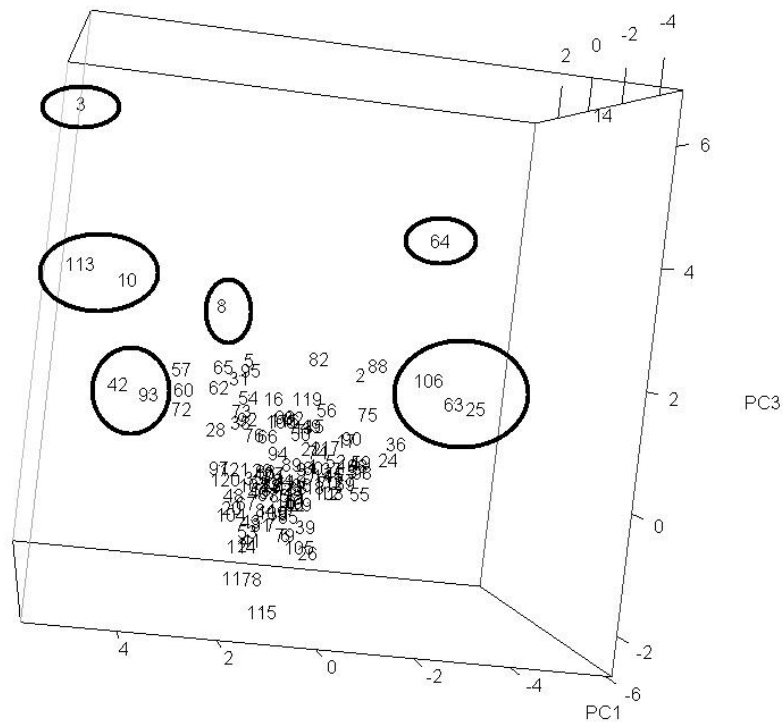
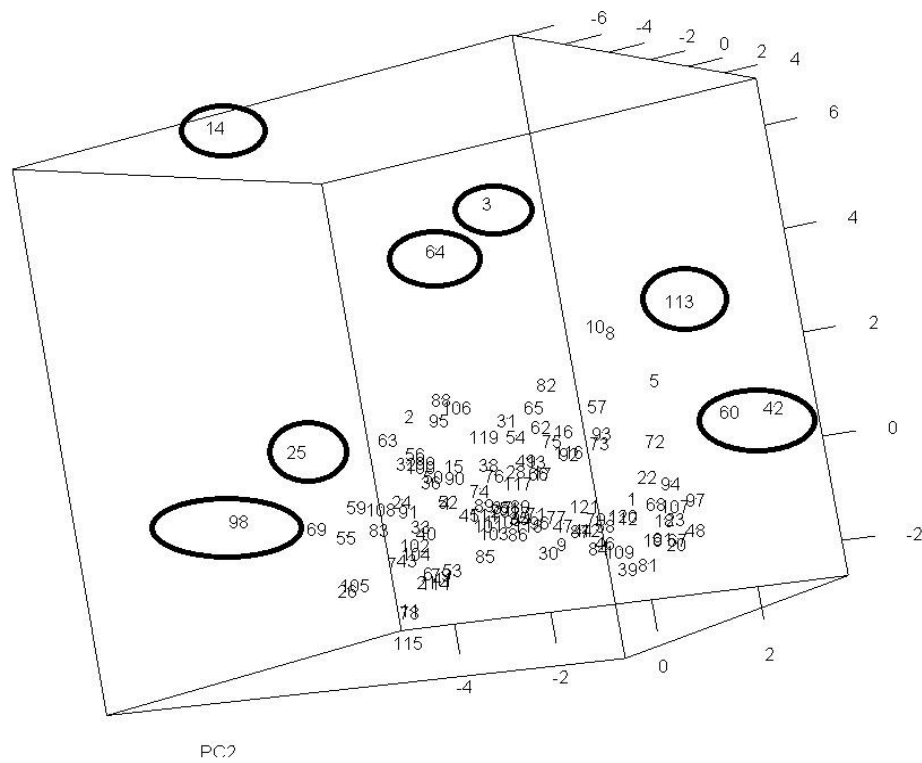
**Deepak Singh**

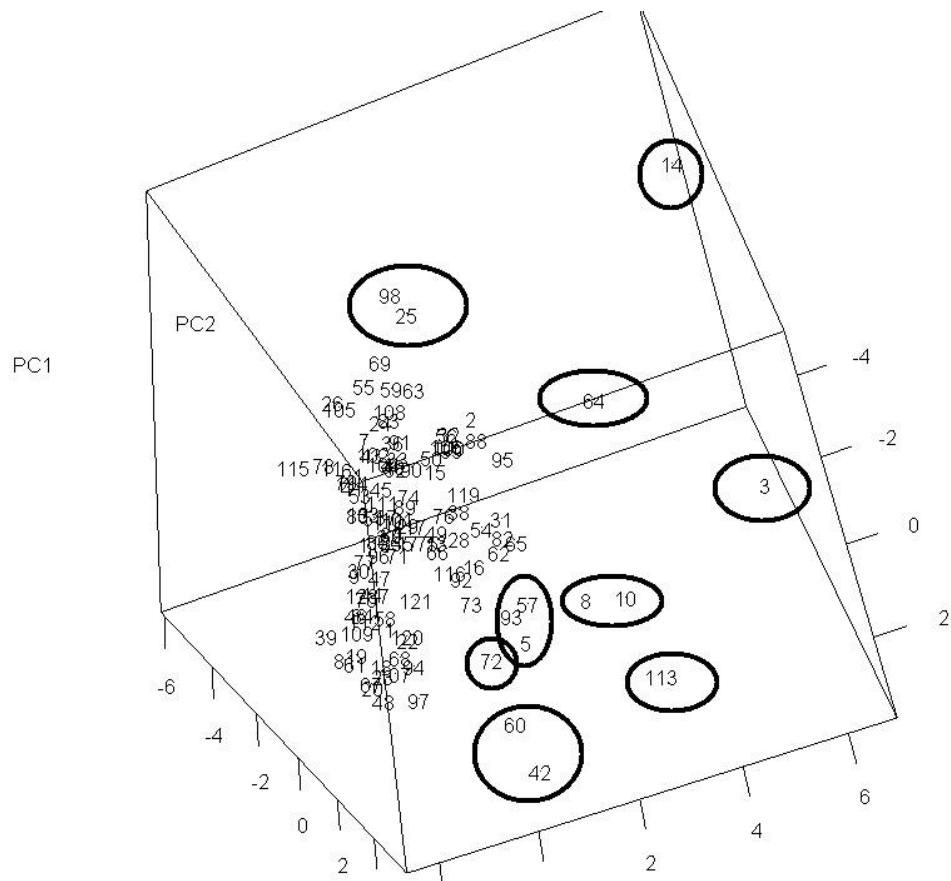
**151091**

### **Answer 1 (a).**

Here we have to detect the multidimensional outliers using PCA.

Below are the visualization of our projected data on the first 3 Principal Component plane.



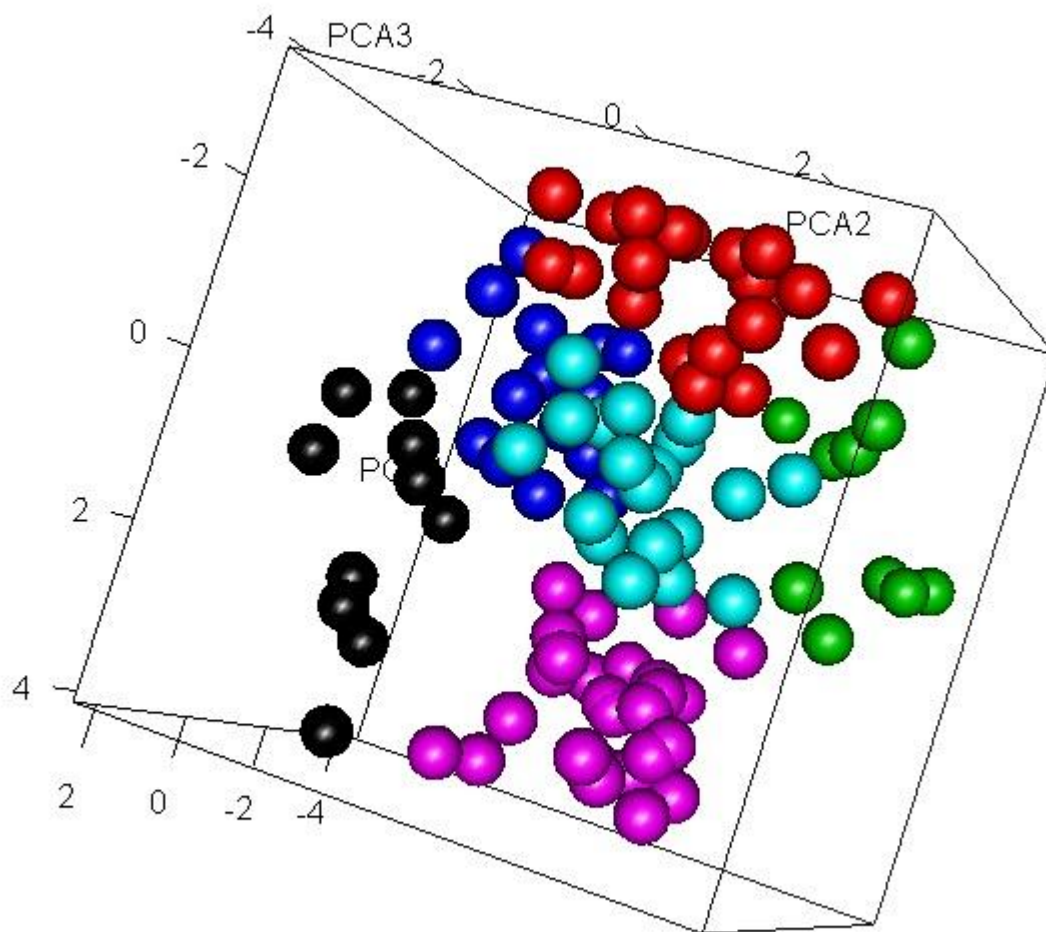


In the above graph we can see that countries having tag (3, 5, 8, 10, 14, 25, 42, 37, 60, 63, 64, 72, 93, 98, 106, 113) seems to be multidimensional outliers. The countries corresponding to these tags:

## [1] Angola	Armenia	Azerbaijan
## [4] Belarus	Botswana	China
## [7] Georgia	Kazakhstan	KyrgyzRepublic
## [10] Lebanon	Lesotho	Moldova
## [13] RussianFederation	Singapore	SyrianArabRepublic
## [16] Ukraine		

### 1(b).

We have to detect rough clusters of world economies from the PCA projection. So after detection of Multi-Dimensional outliers I have removed them and again used PCA on remaining countries data and obtained the first 3 PCs and projected the original data to the PCs plane and plotted that projected data in 3D and got the following graph:



In the above graph we can easily see that there are roughly 6 clusters.

In the above plot the countries belonging to same clusters are:

#### **1<sup>st</sup> cluster**

## [1]	"Australia"	"Austria"	"Belgium"	"Canada"
## [5]	"HongKongChina"	"CongoRep"	"CzechRepublic"	"Denmark"
## [9]	"Finland"	"France"	"Germany"	"Hungary"
## [13]	"Italy"	"Jordan"	"Netherlands"	"NewZealand"
## [17]	"Norway"	"Portugal"	"SaudiArabia"	"SlovakRepublic"
## [21]	"Slovenia"	"SouthAfrica"	"Spain"	"Sweden"
## [25]	"Switzerland"	"UnitedKingdom"	"UnitedStates"	

## 2<sup>nd</sup> Cluster

## [1]	"Albania"	"Burundi"	"Cambodia"	"Cameroon"	"Chad"
## [6]	"Ethiopia"	"Ghana"	"Guatemala"	"Haiti"	"Kenya"
## [11]	"LaoPDR"	"Madagascar"	"Malawi"	"Niger"	"Pakistan"
## [16]	"Rwanda"	"SierraLeone"	"Tanzania"	"Togo"	"Zambia"

## 3<sup>rd</sup> Cluster

## [1]	"Benin"	"Bolivia"
## [3]	"BurkinaFaso"	"CentralAfricanRepublic"
## [5]	"Colombia"	"ElSalvador"
## [7]	"Greece"	"Honduras"
## [9]	"India"	"Mali"
## [11]	"Morocco"	"Mozambique"
## [13]	"Nepal"	"Nigeria"
## [15]	"Poland"	"Turkey"
## [17]	"Uganda"	"Uruguay"
## [19]	"Venezuela"	"Yemen"

## 4<sup>th</sup> cluster

## [1]	"Bangladesh"	"CostaRica"	"CtedIvoire"
## [4]	"DominicanRepublic"	"Ecuador"	"Guinea"
## [7]	"Macedonia"	"Mexico"	"Namibia"
## [10]	"Nicaragua"	"Panama"	"PapuaNewGuinea"
## [13]	"Paraguay"	"Philippines"	"Senegal"
## [16]	"SriLanka"	"Tunisia"	"Vietnam"
## [19]	"Zimbabwe"		

## 5<sup>th</sup> Cluster

## [1]	"Bulgaria"	"Croatia"	"Estonia"	"Jamaica"	"Latvia"	"Lithuania"
## [7]	"Mongolia"	"Romania"				

## 6<sup>th</sup> cluster

## [1]	"Algeria"	"Argentina"	"Chile"	"Egypt"	"Indonesia"
## [6]	"Japan"	"KoreaRep"	"Malaysia"	"Peru"	"Thailand"

### 1(c).

In this we are supposed to do the ranking of the countries.

For the ranking of the countries I have used the projected data along the maximum variance direction i.e. 1<sup>st</sup> PC.

For ranking purpose we have to decide that what will be our criteria *Lower the better or Higher the better*. For this we have to look at elements of Eigen vector corresponding to highest Eigen value.

For the given data I have the following as Eigen vector corresponding to highest Eigen value:

```
## [1] 0.432389652 0.005297265 0.202104840 -0.171820352 -0.445948240
## [6] 0.248509700 0.247461671 0.292299455 0.313341147 0.408970703
## [11] -0.025186651 -0.038590399 -0.260338389
```

In this we can observe that the coefficients corresponding to variables for which *Better the values better will be Economy* are positive. So we have to *Higher the better Rule* for the ranking Purpose.

I have obtained the following Ranking for countries:

##	<u>Country Name</u>	<u>PC1 Value</u>	<u>Rank</u>
##	Malaysia	3.70715955	1
##	HongKongChina	3.47120615	2
##	Switzerland	3.44213440	3
##	Belgium	3.34009417	4
##	Sweden	3.09888332	5
##	Netherlands	3.09711010	6
##	Austria	3.05149011	7
##	Germany	3.01659877	8
##	UnitedStates	2.89789183	9
##	Norway	2.86114910	10
##	Canada	2.69790649	11
##	Japan	2.60989878	12
##	UnitedKingdom	2.53508087	13
##	Finland	2.40260668	14
##	Denmark	2.33288171	15
##	KoreaRep	2.26615240	16
##	France	2.25255712	17
##	CzechRepublic	2.19358293	18
##	Spain	2.18734946	19
##	SaudiArabia	2.10660244	20
##	Italy	2.07429150	21
##	SlovakRepublic	2.00874800	22
##	Australia	1.99396279	23
##	Portugal	1.93415369	24
##	Slovenia	1.92353184	25
##	Thailand	1.92294175	26
##	NewZealand	1.79592293	27
##	Hungary	1.47149881	28
##	CongoRep	1.34946970	29
##	Jordan	1.28616513	30

##	SouthAfrica	1.11209068	31
##	Panama	0.97376168	32
##	Algeria	0.90862122	33
##	Estonia	0.88129628	34
##	CostaRica	0.82883015	35
##	Namibia	0.81785283	36
##	Chile	0.61830687	37
##	Philippines	0.60034146	38
##	Indonesia	0.58056989	39
##	PapuaNewGuinea	0.53526303	40
##	Jamaica	0.42719032	41
##	Tunisia	0.42373815	42
##	Morocco	0.32139894	43
##	Ecuador	0.27372692	44
##	Argentina	0.20874521	45
##	Greece	0.06918419	46
##	Yemen	0.06402538	47
##	Latvia	0.03347706	48
##	SriLanka	-0.04296881	49
##	Poland	-0.06586108	50
##	DominicanRepublic	-0.07104237	51
##	Croatia	-0.09491463	52
##	Lithuania	-0.09835614	53
##	Egypt	-0.17738367	54
##	CtedIvoire	-0.24213274	55
##	Mexico	-0.32081440	56
##	Turkey	-0.33962485	57
##	Macedonia	-0.36266026	58
##	Peru	-0.48603041	59
##	Vietnam	-0.50067873	60
##	Colombia	-0.51815843	61
##	Senegal	-0.61611235	62
##	ElSalvador	-0.66945636	63
##	Zimbabwe	-0.67668889	64
##	Nicaragua	-0.73098234	65
##	Guinea	-0.81834630	66
##	Honduras	-0.83619180	67
##	India	-0.92220465	68
##	Paraguay	-0.94230993	69
##	Bolivia	-0.96380347	70
##	Bangladesh	-1.12538928	71
##	Venezuela	-1.13858859	72
##	Pakistan	-1.26505210	73
##	Romania	-1.34268017	74
##	Mongolia	-1.34445993	75
##	Kenya	-1.42628323	76
##	BurkinaFaso	-1.47893788	77
##	Ghana	-1.52020246	78
##	Zambia	-1.60138442	79
##	Togo	-1.65259167	80
##	Bulgaria	-1.67603584	81
##	Nepal	-1.69773569	82
##	Uruguay	-1.71873961	83
##	Nigeria	-1.77055888	84

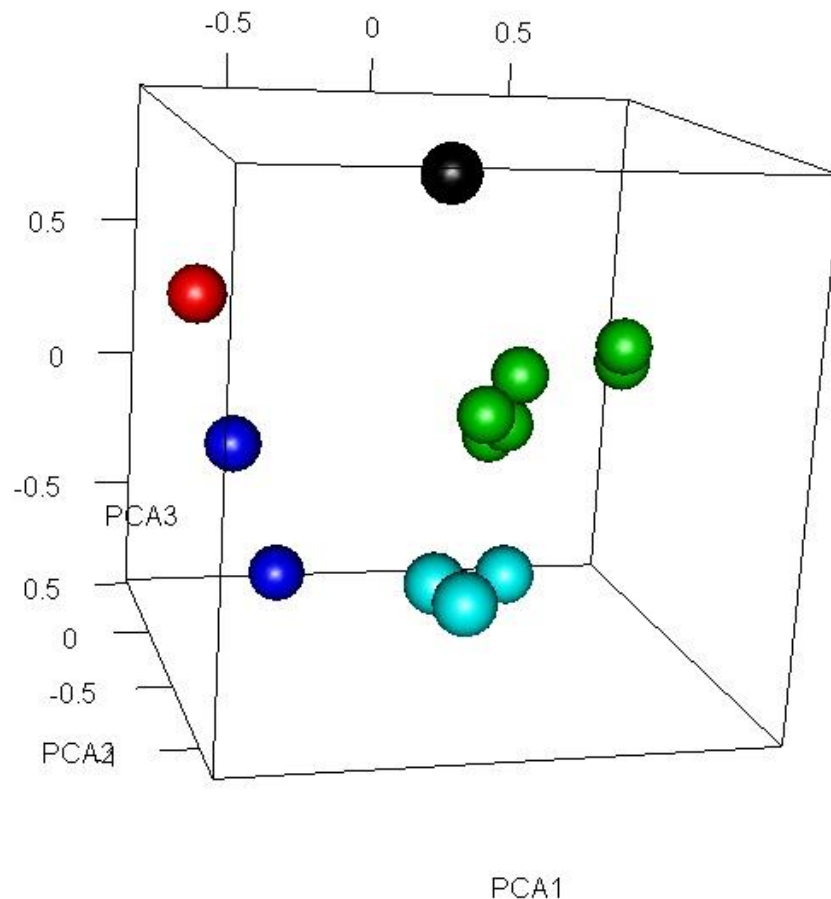
##	Guatemala	-1.96605615	85
##	Cameroon	-1.99632803	86
##	Niger	-2.01603947	87
##	Benin	-2.10742113	88
##	Mali	-2.13236105	89
##	Ethiopia	-2.26885575	90
##	Mozambique	-2.37556829	91
##	Burundi	-2.52769747	92
##	Madagascar	-2.64422764	93
##	SierraLeone	-2.67455323	94
##	Malawi	-2.68601171	95
##	Cambodia	-2.84405251	96
##	LaoPDR	-2.84955723	97
##	Uganda	-2.86857701	98
##	Tanzania	-2.88422024	99
##	Haiti	-2.89832350	100
##	Chad	-2.91473752	101
##	CentralAfricanRepublic	-2.99463526	102
##	Rwanda	-3.00934471	103
##	Albania	-3.09351248	104

From the above Ranking we can say that ranking is looking like realistic as United States, China, Malaysia, Japan etc. got good ranks.



**1(d).**

In this we have to do Variable clustering. I have done the variable clustering and got the following 3D plot:

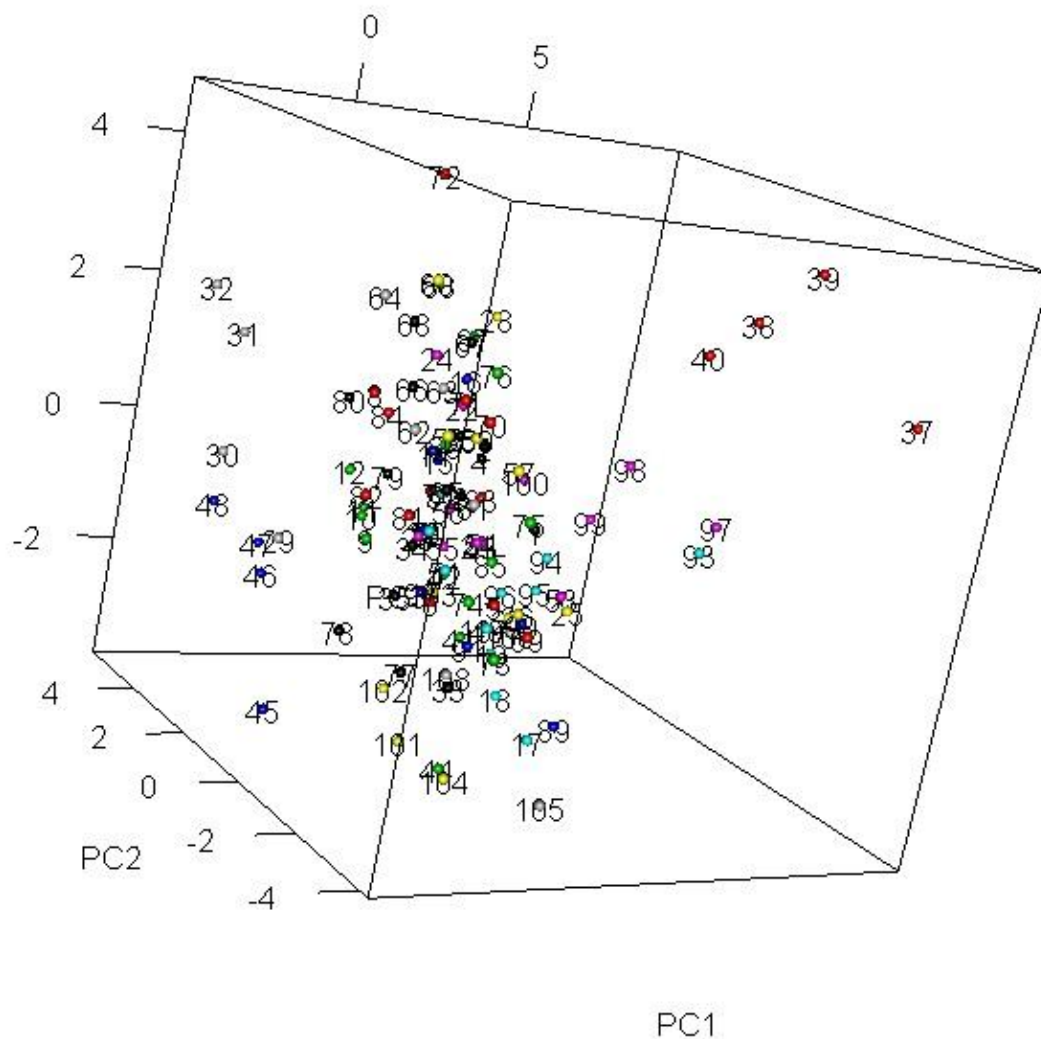


Variables ("gnpper" "dominv" "govcon" "exp" "resbl" "domcrdt") ,("indvlad" "girimp" "impcov"), ("gdpdf1" "intsprd") are in one cluster while "agrvlad" and "gdpgr" seemed to be alone (It seems that they are not clustering with none of the other variables).

## 2(a).

In this we have to trace a significant trajectory movement on the projected plane using PCA of financial ratio data over the years.

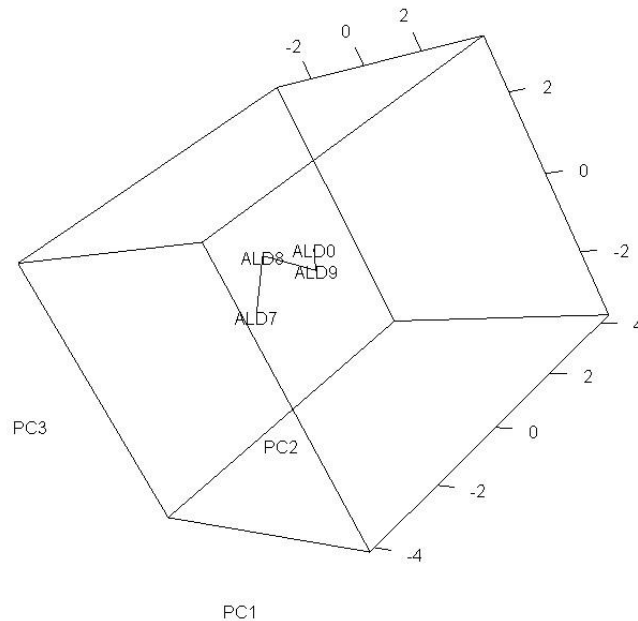
Before tracing we have to remove the outlier countries in order to compare the trajectory movement of the banks. I have plotted the data to detect the outlier banks



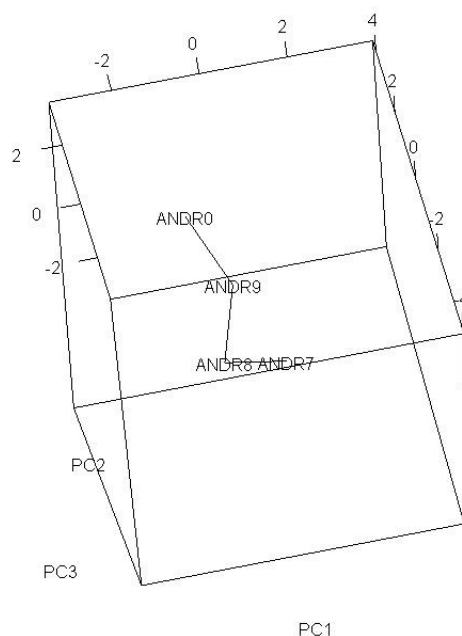
From the above plot I have detected IND, UBI, UCO, OBC, COPR, SBIN and SBSAU.

I have detected the trajectory movement of all banks in R and the results in the format of 3D plots are as follows:

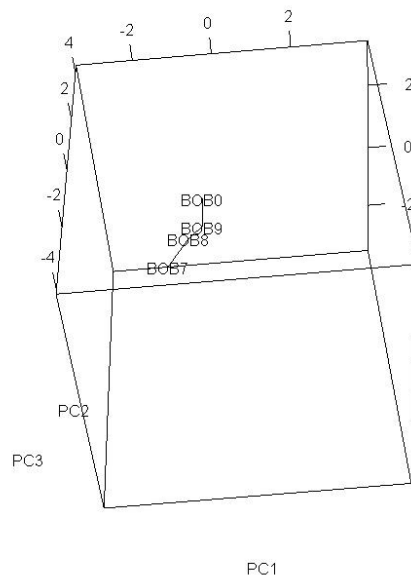
**ALD Bank:**



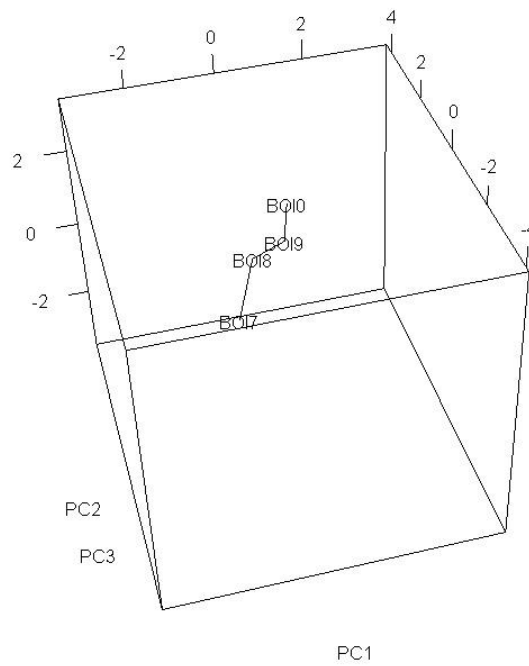
By observing the above plot we cannot say that ALD bank has significant trajectory movement. As it has got better from 1997 to 1998 but after that we are not seeing any significant movement.



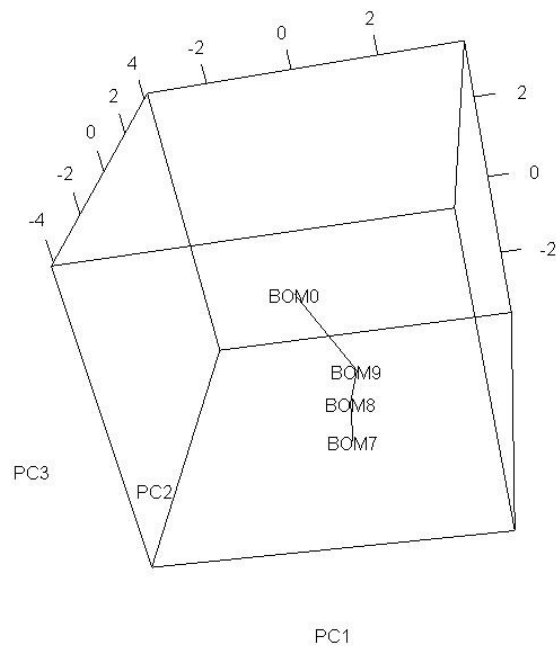
**ANDR Bank:** Here we can detect significant trajectory movement along 3<sup>rd</sup> PC but not in 1<sup>st</sup> and 2<sup>nd</sup> PC. So we can infer that this bank is doing good but not so good.



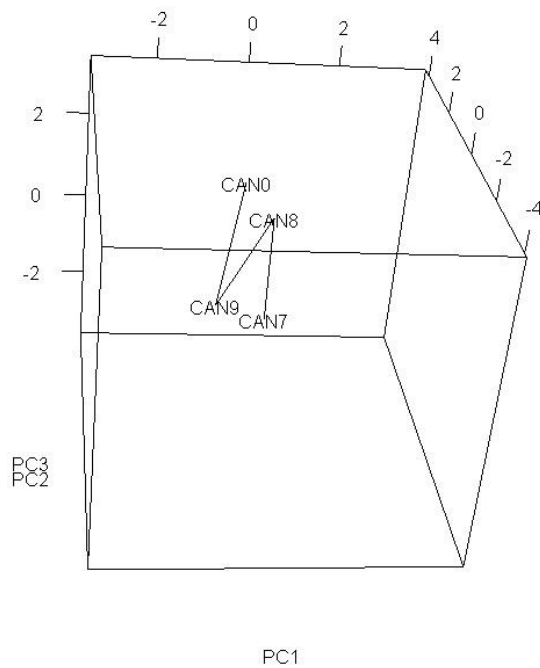
**BOB Bank:** Here we are not able to detect any significant movement.

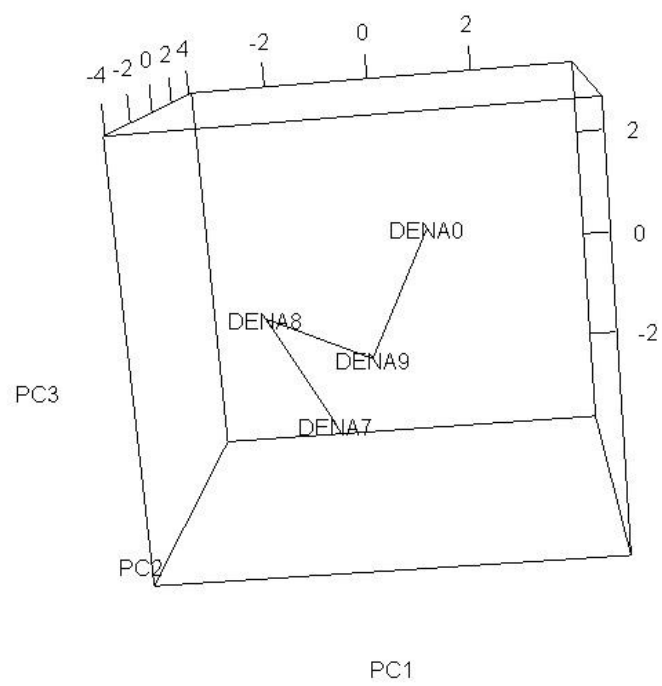
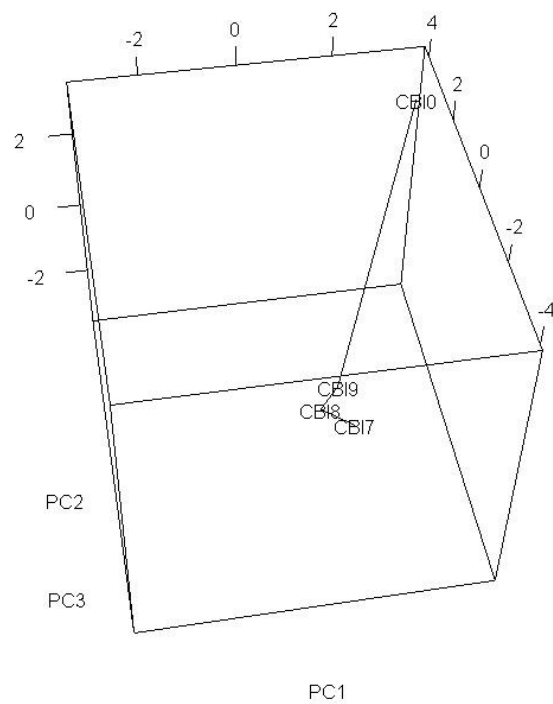


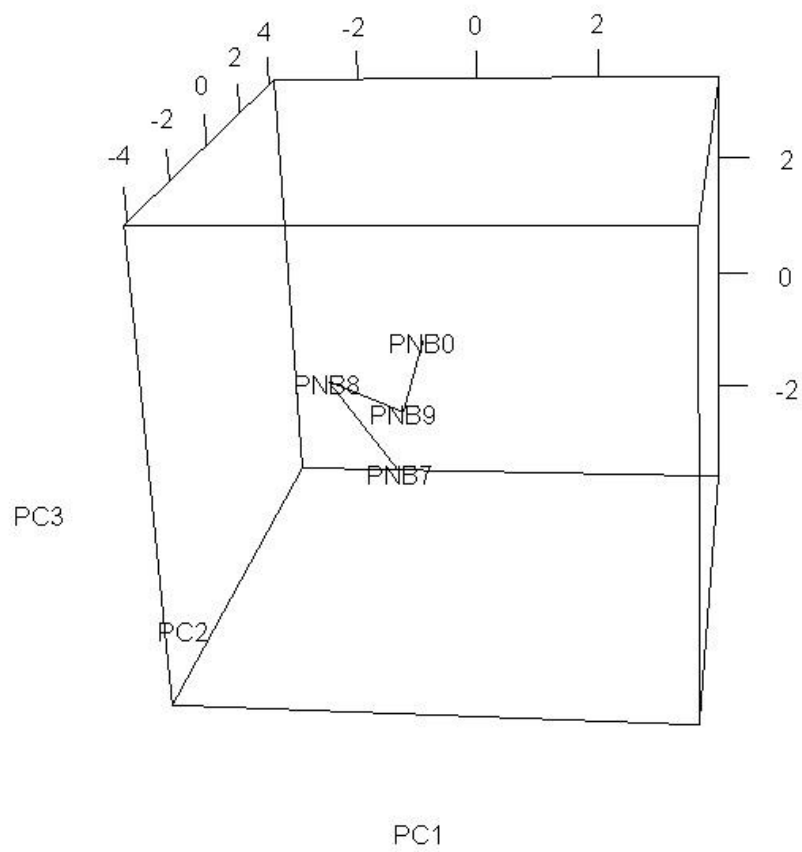
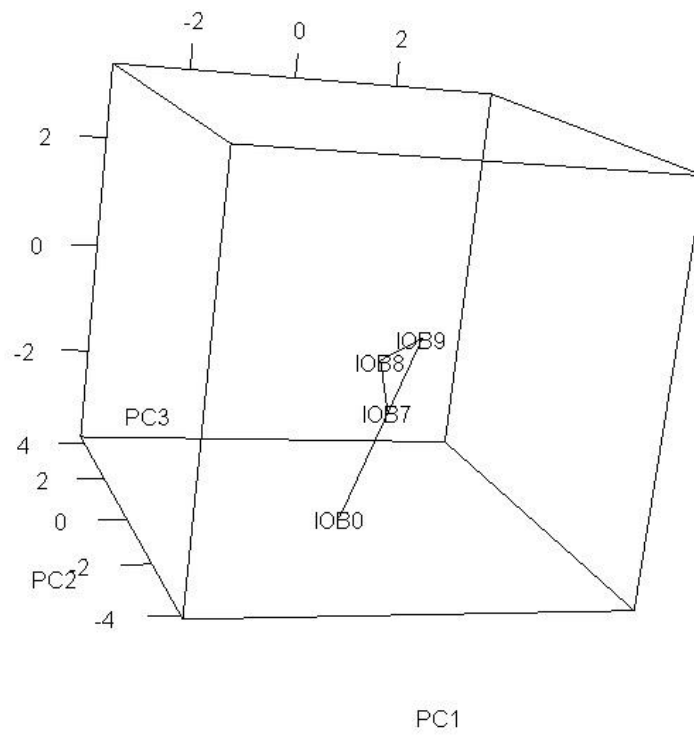
**BOI Bank:** This bank also has no significant trajectory movement.

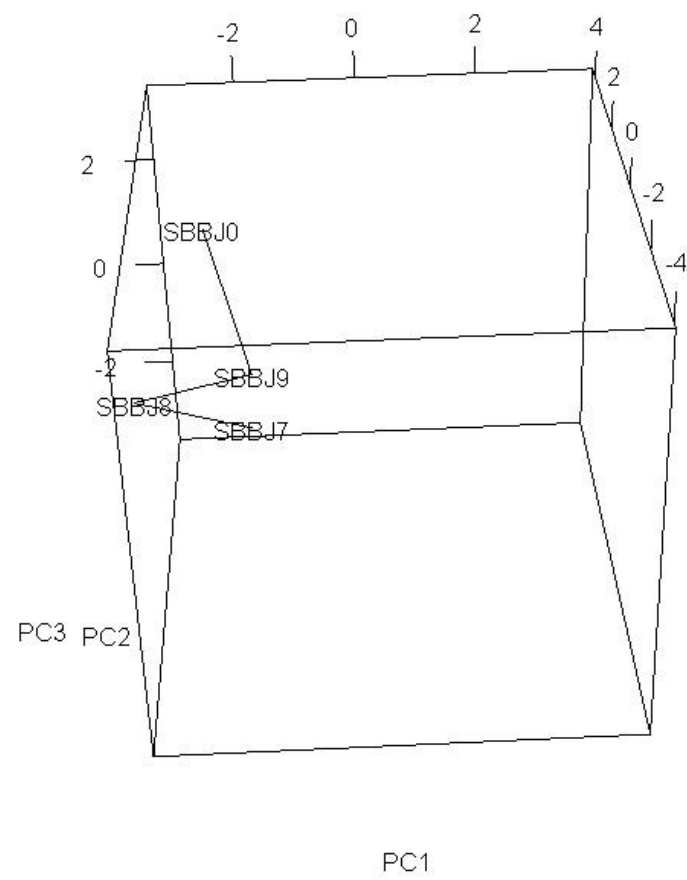
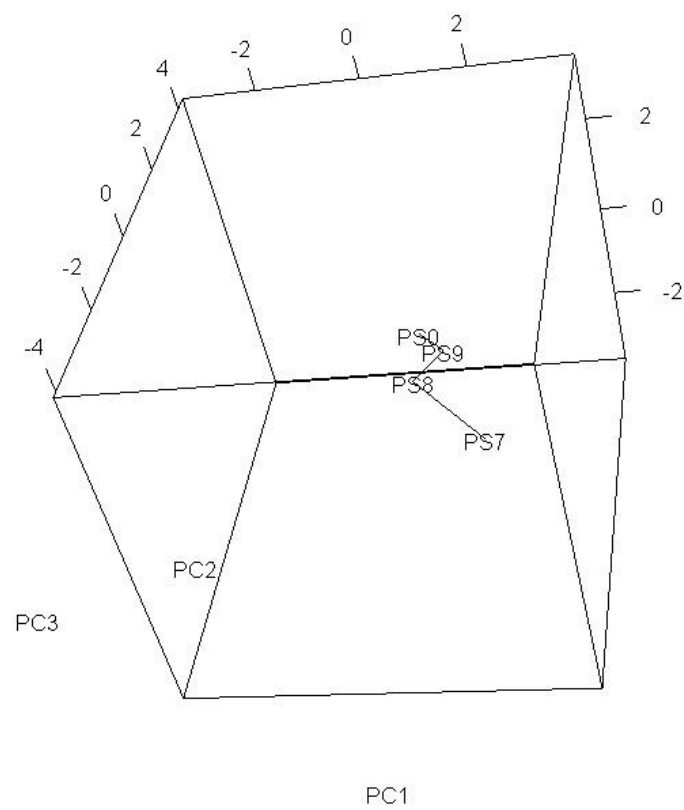


**BOM Bank:** This bank has no significant trajectory movement from 1996-1999 but in financial year it has got significant movement.

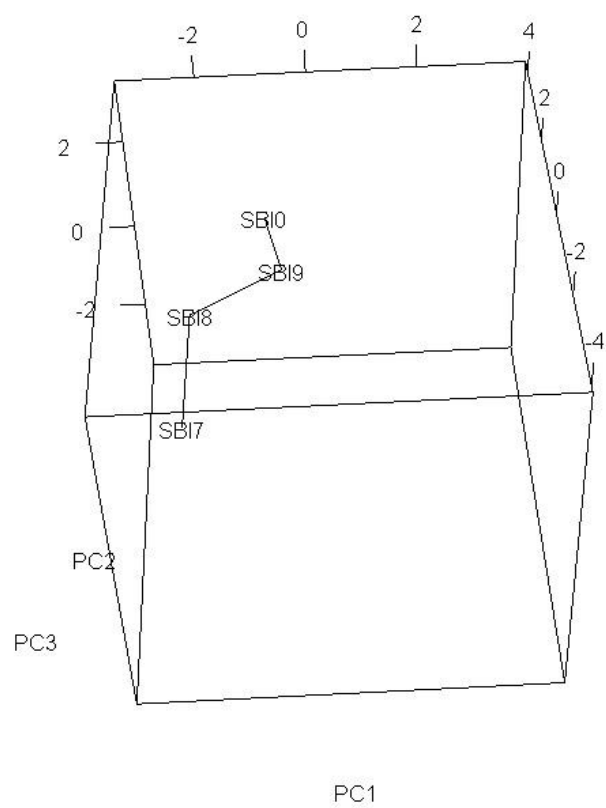
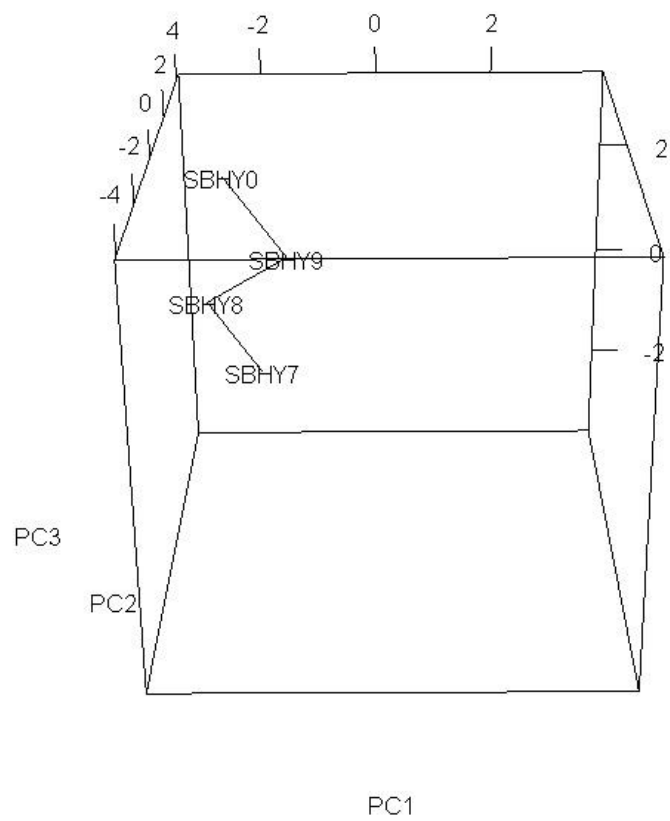


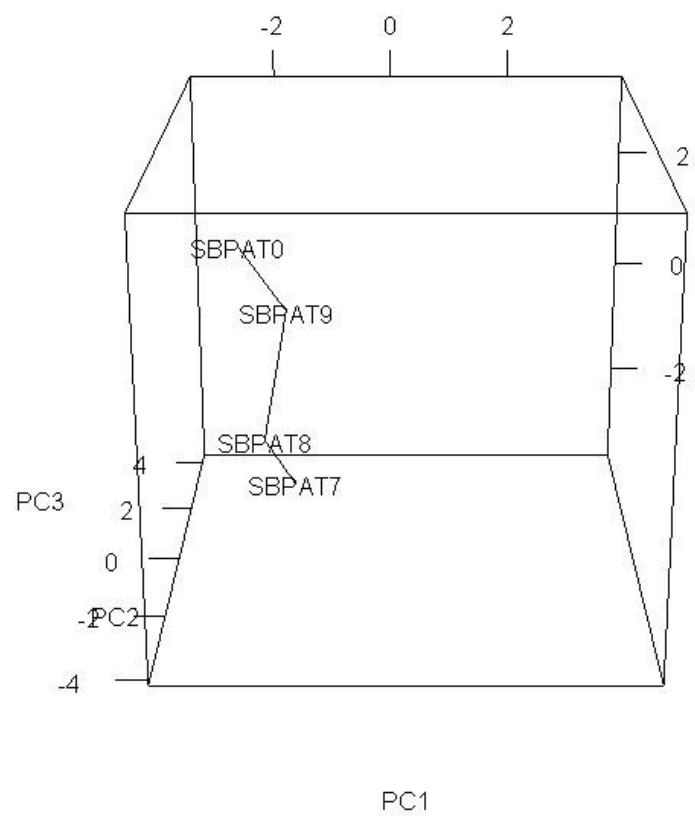
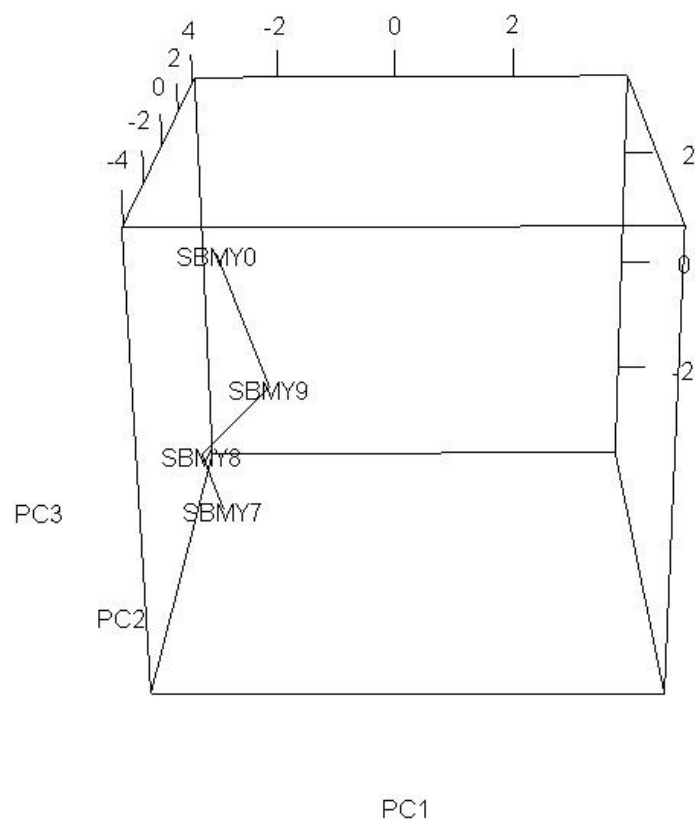


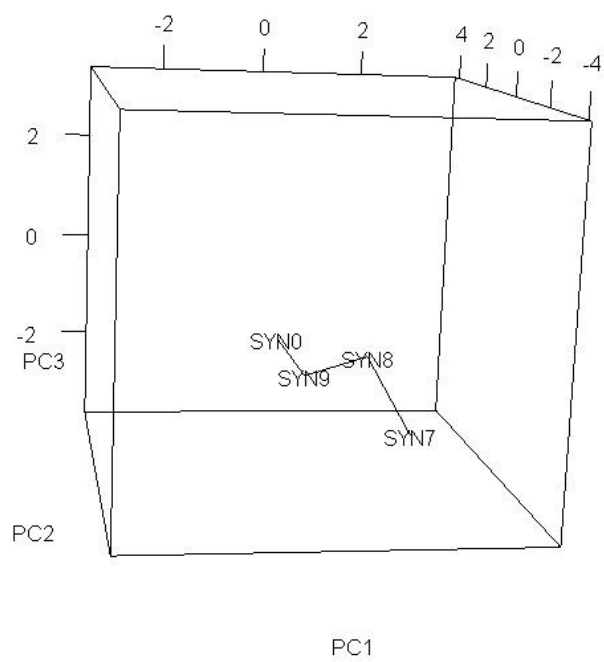
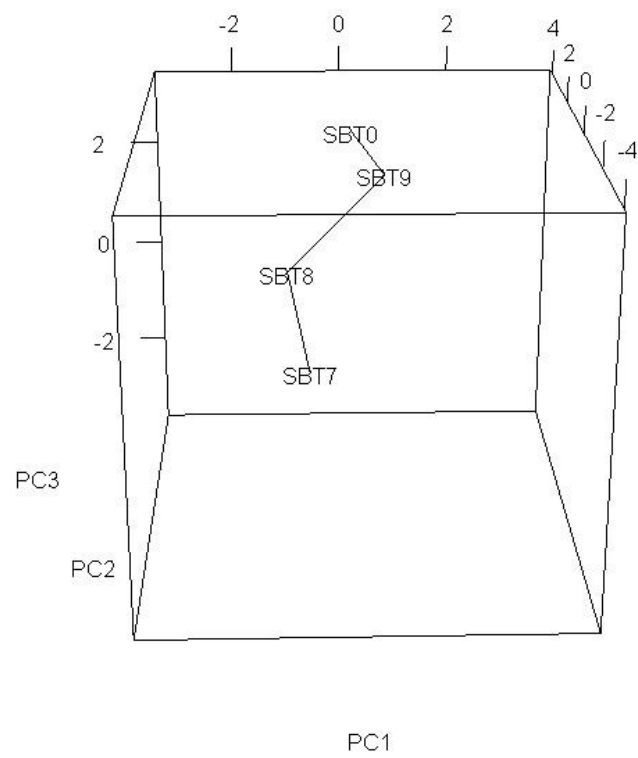


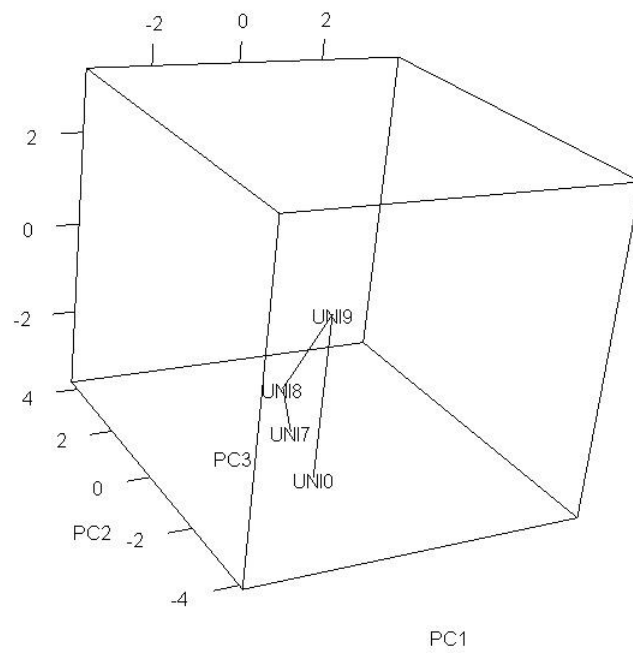








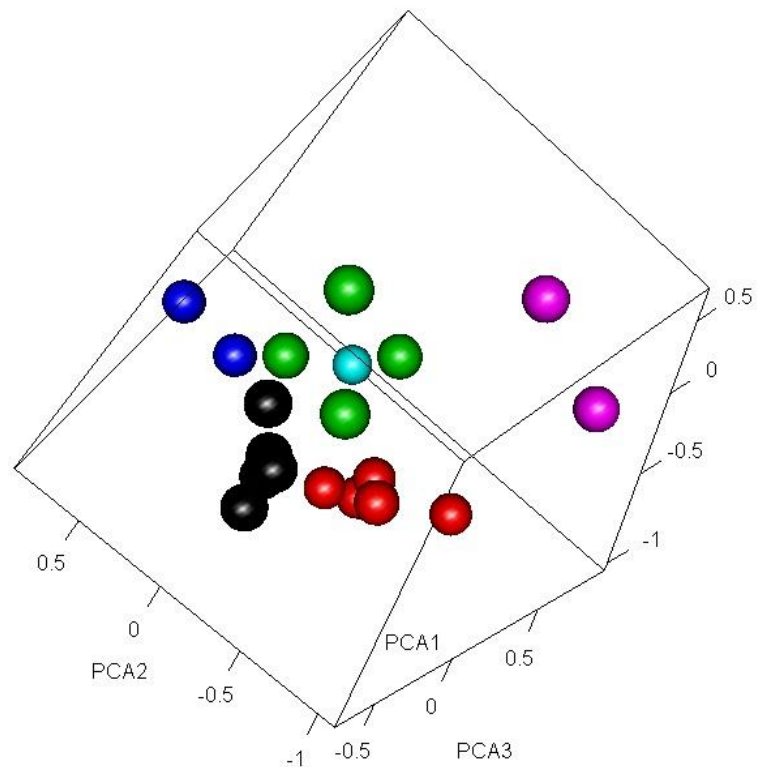




From the above graphs we can conclude that **ANGR, DENA, PNB, SBBJ, SBHY, SBI, SBMY, SBPAT, SBT** has significant trajectory movement.

## 2(b).

Here we have to do PCA based clustering of Variables.



Variables belonging to same clusters are:

"ROE" "ROA" "OPAST" "PRVTAST" "PFTEMP"  
"OTHINTOT" "COMINTOT" "NETNPA" "CRAR"  
"COD" "ROAD" "ROI" "NIITAST" "SPRD"  
"STEXTX" "INTRMTAST"  
"COBR" "BUSEMP"  
"INTINTOT"

**2(c):**

In this we have to do performance ranking of banks for each year:

**For 1996-1997**

Bank Name	PC_1_Values	Rank
COPR7	-4.5908	1
OBC7	-4.0524	2
SBSAU7	-2.6166	3
BOB7	-2.3192	4
SBI7	-2.3073	5
SBMY7	-2.2631	6
SBPAT7	-2.1624	7
SBHY7	-1.9975	8
SBIN7	-1.4681	9
SBBJ7	-1.2836	10
SBT7	-1.1249	11
DENA7	-0.7841	12
UNI7	-0.7460	13
BOI7	-0.3595	14
PNB7	-0.2683	15
CAN7	-0.1539	16
ALD7	0.6435	17
IOB7	1.0392	18
ANDR7	1.2323	19
BOM7	1.6215	20
CBI7	2.1063	21
VIJ7	2.2697	22
PS7	2.3385	23
SYN7	3.1357	24
UBI7	6.6516	25
UCO7	7.4596	26

**For 1997-1998**

Bank Name	PC_1_Values	Rank
SBSAU7	-4.7821	1
SBBJ7	-4.1587	2
SBM7	-4.0712	3
SBHY7	-2.8645	4
SBIN7	-2.8026	5
SBPAT7	-2.0163	6
DENA7	-1.9064	7
PNB7	-1.5153	8
SBI7	-1.4944	9
ANDR7	-0.1824	10
BOB7	0.1305	11
ALD7	0.3253	12
UNI7	0.9799	13
CBI7	1.1444	14
BOM7	1.3123	15
BOI7	1.3906	16
PS7	1.4504	17
CAN7	1.6426	18
SYN7	2.2569	19
IOB7	2.8682	20
UBI7	3.3095	21
VIJ7	3.4229	22
UCO7	5.5603	23

**For 1998-1999**

Bank Name	PC_1_Values	Rank
SBIN7	-5.556	1
SBBJ7	-3.699	2
SBHY7	-2.875	3
SBSAU7	-2.598	4
SBPAT7	-2.204	5
PNB7	-1.533	6
SBI7	-1.125	7
CAN7	-0.583	8
ANDR7	-0.468	9
BOB7	-0.234	10
SYN7	-0.066	11
ALD7	0.233	12
BOM7	0.816	13
DENA7	0.902	14
CBI7	1.243	15
BOI7	1.661	16

VIJ7	1.766	17
PS7	2.251	18
UNI7	3.000	19
IOB7	3.388	20
UBI7	5.679	21

### For 1999-2000

Bank Name	PC_1_Values	Rank
SBIN7	-4.200	1
SBBJ7	-3.703	2
SBPAT7	-2.648	3
SBHY7	-2.599	4
SBSAU7	-2.506	5
SBMY7	-2.088	6
SBI7	-1.716	7
ANDR7	-1.475	8
SBT7	-0.713	9
BOB7	-0.077	10
PNB7	0.330	11
CAN7	0.373	12
BOM7	0.827	13
SYN7	0.968	14
ALD7	1.248	15
BOI7	1.339	16
PS7	2.030	17
DENA7	2.030	18
VIJ7	3.279	19
UNI7	4.377	20
IOB7	4.925	21

Note: If any particular bank is not looking in any of the rankings then it means that that bank has been declared as outlier for that financial year.

For 3D figures and all the codes please go to my GitHub Repository

<https://github.com/sdeepak09/MTH-514-Multivariate-Analysis-Assignments>