<다변량분석 리포트 2>

2014170852 산업경영공학부 조영관

College data에서 private column을 제외하고 군집화를 수행하였다.

<K-MEANS Clustering>

Scale 함수를 사용해서 모든 변수의 평균을 0, 표준편차를 1로 만드는 정규화를 수행하였다.

	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	PUndergrad	Outstate
Abilene Christian University	-0.346658529	-0.320998691	-0.06346802	-0.25841635	-0.19170393	-0.16800756	-0.209072463	-0.745875453
Adelphi University	-0.210748292	-0.038678077	-0.28839845	-0.65523355	-1.35303987	-0.20965344	0.244149784	0.457201895
Adrian College	-0.406603728	-0.376075689	-0.47781355	-0.31510452	-0.29268967	-0.54921162	-0.496770063	0.201175104
Agnes Scott College	-0.667830438	-0.681243057	-0.69198176	1.83904599	1.67653214	-0.65765583	-0.520416441	0.626229291
Alaska Pacific University	-0.725708562	-0.764062544	-0.78023197	-0.65523355	-0.59564687	-0.71146560	0.008999690	-0.716047089
Albertson College	-0.623905076	-0.628205948	-0.66938109	0.59190622	0.31322473	-0.62301966	-0.534867005	0.760456929
Albertus Magnus College	-0.684367045	-0.684914857	-0.72857331	-0.59854538	-0.54515400	-0.67703559	-0.410723520	0.708257292
Albion College	-0.284904638	-0.121905541	-0.31315156	0.53521805	0.61618194	-0.43417005	-0.540778600	0.851930578
Albright College	-0.507373676	-0.481333953	-0.59512176	0.13840085	0.36371760	-0.56220019	-0.360803389	1.281210450
Alderson-Broaddus College	-0.625196999	-0.620454370	-0.65431398	-0.37179269	-0.59564687	-0.59807337	-0.510563783	0.006793599
Alfred University	-0.328054846	-0.242258982	-0.33144734	0.53521805	0.96963200	-0.38551448	-0.489544781	1.518097374
Allegheny College	-0.090341124	-0.048469544	-0.31853267	0.93203525	1.07061774	-0.41087310	-0.532896474	1.650336454
Allentown Coll. of St. Francis de Sales	-0.470941464	-0.505404641	-0.52731977	0.59190622	0.41421047	-0.52983186	-0.142731236	-0.186593628
Alma College	-0.448203630	-0.383011311	-0.42507866	0.93203525	0.86864627	-0.49354635	-0.543405975	0.529784247
Alverno College	-0.647934833	-0.695930257	-0.67045731	-0.25841635	-0.49466114	-0.49127850	0.249404534	-0.519179886

다음과 같이 모든 값들이 정규화 된 상태임을 확인할 수 있다. (위는 일부만 나타냄)

clValid 함수를 이용해서 k-means clustering의 군집 수를 변화시키며 internal, stability 타당성 지표를 산출한 후 최적의 군집 수를 판별해보았다.

```
Clustering Methods:
 kmeans

Cluster sizes:
 2 3 4 5 6 7 8 9 10

Validation Measures:
                                 2        3        4        5        6        7        8        9       10

kmeans APN              0.1287   0.0436   0.1344   0.1815   0.2321   0.1748   0.1632   0.2372   0.3118
       AD               5.0040   4.3162   4.2363   4.1467   4.1325   3.9275   3.8215   3.8121   3.8180
       ADM              0.7046   0.1872   0.6012   0.8891   1.1200   0.6406   0.5745   0.8429   1.1216
       FOM              0.9627   0.8144   0.7938   0.7653   0.7755   0.7569   0.7439   0.7389   0.7334
       Connectivity   100.0270 194.8433 279.2139 258.9845 300.7508 283.7972 398.0944 434.5683 458.6734
       Dunn             0.0842   0.0611   0.0439   0.0481   0.0481   0.0577   0.0577   0.0679   0.0947
       Silhouette       0.3201   0.2421   0.1966   0.1883   0.1840   0.1908   0.1682   0.1500   0.1343

Optimal Scores:

             Score    Method Clusters
APN            0.0436 kmeans 3
AD             3.8121 kmeans 9
ADM            0.1872 kmeans 3
FOM            0.7334 kmeans 10
Connectivity 100.0270 kmeans 2
Dunn           0.0947 kmeans 10
Silhouette     0.3201 kmeans 2
```

군집 수를 2에서 10까지 변화시켰고, 위와 같은 결과를 얻을 수 있었다.

시간은 약 30초 정도 소요되었다.

Dunn index와 Silhouette index 기준으로 각각의 최적 군집 수는 10개와 2개임을 알 수 있다.

K=3 (군집을 3으로 지정했을 때)으로 군집화를 10번 반복 수행하였다.

그리고 각 회차마다 각 군집의 Center와 Size를 확인하였다.

그 결과 1번을 제외하고 9번은 동일한 결과를 얻을 수 있었다.

```
> college_kmc$centers
         Apps     Accept     Enroll  Top10perc  Top25perc F.Undergrad P.Undergrad   Outstate
1 -0.03489665 -0.1165637 -0.2330693  0.8552988  0.8370369  -0.3018808  -0.3693397  1.0480076
2  1.83946866  2.0074927  2.2024888  0.2298334  0.3990083   2.2662920   1.5787435 -0.5368952
3 -0.37097263 -0.3607812 -0.3367498 -0.5291736 -0.5548376  -0.3116495  -0.1277753 -0.4746087
  Room.Board      Books    Personal         PhD   Terminal  S.F.Ratio perc.alumni      Expend
1  0.7172261  0.0590233 -0.37425038  0.7370637  0.7356034 -0.6350344   0.8164997  0.80516666
2 -0.1695508  0.3262545  0.81162753  0.6898820  0.6782520  0.5844663  -0.5945498 -0.05849565
3 -0.3668251 -0.1024233  0.03786354 -0.5604491 -0.5571596  0.2325641  -0.3323420 -0.43979658
   Grad.Rate
1  0.7674493
2 -0.3637427
3 -0.3538001
> college_kmc$size
[1] 246  93 438
```

9번의 경우 위와 같은 center와 size가 나왔다.

```
> college_kmc$centers
         Apps     Accept     Enroll  Top10perc  Top25perc F.Undergrad P.Undergrad    Outstate
1 -0.3159592 -0.3243645 -0.3378374 -0.5312282 -0.6051143  -0.3389592 -0.19622296 -0.09311589
2 -0.3262037 -0.3027779 -0.2488333 -0.4824584 -0.4830089  -0.2198444 -0.03923668 -0.45283503
3  0.5784262  0.5386644  0.4467651  0.8591505  0.8643183   0.3970303  0.07861575  0.78322680
  Room.Board      Books    Personal         PhD   Terminal   S.F.Ratio perc.alumni       Expend
1  0.2887409  3.7381073  0.64909366 -1.4643560 -0.7439503 -0.03845138  -0.5997019 -0.08834099
2 -0.3782806 -0.2134495  0.02916773 -0.4388389 -0.4496251  0.23390775  -0.2788057 -0.42639283
3  0.6333326  0.1530696 -0.08719706  0.8375400  0.8149032 -0.39962287   0.5132171  0.73753002
   Grad.Rate
1 -0.1725096
2 -0.3497567
3  0.6106897
> college_kmc$size
[1]  16 481 280
```

1번의 경우 위와 같은 center와 size가 나왔다.

K=10(군집이 10개일 때)으로 군집화를 수행하고 각 군집의 center와 size를 살펴보았다.

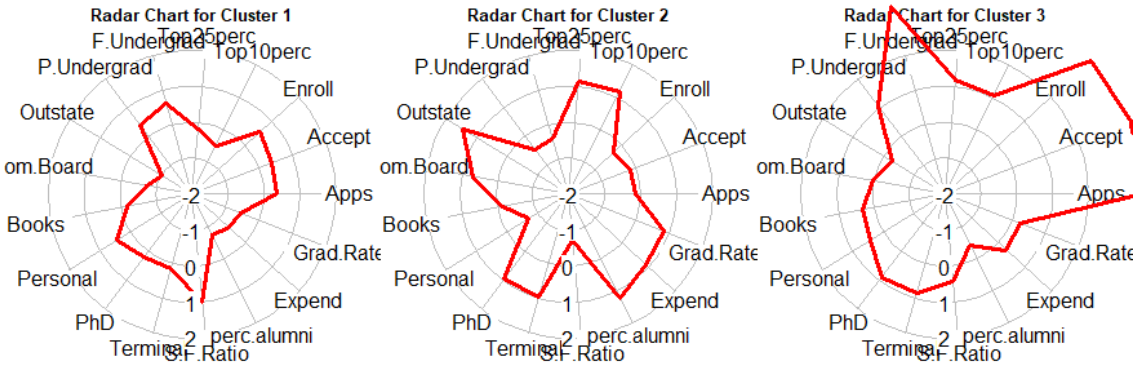
```
> college_kmc$size
 [1]  72 120  20  82  29 135 110  77  79  53
> # k=10일 때
> college_kmc <- kmeans(college_x_scaled,10)
> college_kmc$size
 [1]  28  82  12  95 129  14 127 133 130  27
> # k=10일 때
> college_kmc <- kmeans(college_x_scaled,10)
> college_kmc$size
 [1]  28  27  78 102   9 107 118 171  54  83
> # k=10일 때
> college_kmc <- kmeans(college_x_scaled,10)
> college_kmc$size
 [1]  29  86  54  90 188 118   4  89  95  24

> college_kmc <- kmeans(college_x_scaled,10)
> college_kmc$size
 [1]  20  95  14  65 154 186  74  51  17 101
> # k=10일 때
> college_kmc <- kmeans(college_x_scaled,10)
> college_kmc$size
 [1] 173  14  80  10 143  91 140  74  24  28
> # k=10일 때
> college_kmc <- kmeans(college_x_scaled,10)
> college_kmc$size
 [1]  95 142  27  37  53  24 110 121  86  82
> # k=10일 때
> college_kmc <- kmeans(college_x_scaled,10)
> college_kmc$size
 [1]  74  92  27  52 146 116  24  53 112  81
```

위의 결과를 통해 알 수 있듯이 모두 다른 형태로 군집이 형성되었다.

각 군집의 center와 size가 매 iteration을 돌릴 때마다 다른 결과가 나왔다.

K=3으로 군집화를 수행한 후 RADAR CHART를 도시해보았다.



위 결과를 눈으로 확인해 보면 1번 CLUSTER와 2번 CLUSTER가 상대적으로 3에 비해 유사한 것으로 보인다. 반대로 1번 CLUSTER와 3번 CLUSTER가 가장 다른 군집 쌍이라 보인다.

보다 정확한 결과를 얻기 위해 T-TEST 검정을 진행한다.

실제로 수치적으로 확인해보기 위해 T-TEST 검정을 수행해보았다.

	v1	v2	v3
1	6.508494e-10	3.254247e-10	1.000000e+00
2	2.707715e-14	1.353857e-14	1.000000e+00
3	6.145997e-25	3.072999e-25	1.000000e+00
4	8.662172e-36	1.000000e+00	4.331086e-36
5	4.499404e-25	1.000000e+00	2.249702e-25
6	1.935472e-30	9.677361e-31	1.000000e+00
7	2.569720e-18	1.284860e-18	1.000000e+00
8	1.072409e-69	1.000000e+00	5.362043e-70
9	7.990904e-23	1.000000e+00	3.995452e-23
10	8.684370e-01	5.657815e-01	4.342185e-01
11	8.082045e-16	4.041022e-16	1.000000e+00
12	1.077062e-18	1.000000e+00	5.385312e-19
13	2.947919e-18	1.000000e+00	1.473959e-18
14	7.317220e-41	3.658610e-41	1.000000e+00
15	8.632100e-38	1.000000e+00	4.316050e-38
16	8.218473e-33	1.000000e+00	4.109236e-33
17	9.949542e-26	1.000000e+00	4.974771e-26

CLUSTER 1과 2에 대해 T-TEST를 진행해보았다. 왼쪽부터 TWO-SIDED, GREATER, LESS 순이다.

변수들 간에 유의미한 차이가 있음을 확인할 수 있다.

	v1	v2	v3
1	1.947797e-09	0.999999999	9.738984e-10
2	7.823469e-10	1.000000000	3.911734e-10
3	8.325635e-12	1.000000000	4.162818e-12
4	1.565618e-07	0.999999922	7.828091e-08
5	2.328916e-09	0.999999999	1.164458e-09
6	6.467254e-12	1.000000000	3.233627e-12
7	1.063231e-03	0.999468384	5.316157e-04
8	3.168884e-04	0.999841556	1.584442e-04
9	1.426841e-04	0.999928658	7.134207e-05
10	2.716477e-03	0.998641762	1.358238e-03
11	8.921914e-01	0.446095684	5.539043e-01
12	3.425664e-11	1.000000000	1.712832e-11
13	7.702427e-10	1.000000000	3.851214e-10
14	8.934232e-03	0.004467116	9.955329e-01
15	1.752993e-02	0.991235036	8.764964e-03
16	3.838549e-08	0.999999981	1.919274e-08
17	5.324870e-06	0.999997338	2.662435e-06

CLUSTER 1과 3에 대해 T-TEST를 진행해보았다. 거의 모든 변수들이 유의미한 차이를 보이고 있다. 그런데 11번째 변수의 P-VALUE를 살펴보면 기각할 수 없음을 확인할 수 있다. 따라서 11번째 변수는 차이가 유의미하지 않다고 결론 내릴 수 있다.

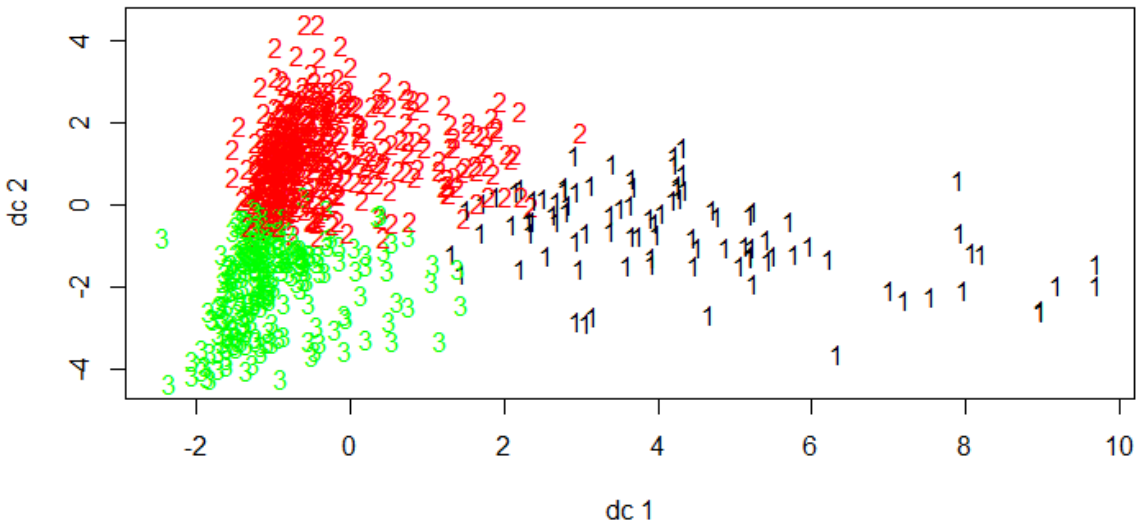
	v1	v2	v3
1	1.030221e-10	1.000000e+00	5.151103e-11
2	3.072440e-11	1.000000e+00	1.536220e-11
3	3.467881e-14	1.000000e+00	1.733941e-14
4	5.649715e-01	2.824857e-01	7.175143e-01
5	8.746910e-01	5.626545e-01	4.373455e-01
6	9.186640e-15	1.000000e+00	4.593320e-15
7	2.712406e-09	1.000000e+00	1.356203e-09
8	8.380321e-11	4.190160e-11	1.000000e+00
9	1.356249e-03	6.781246e-04	9.993219e-01
10	1.726795e-03	9.991366e-01	8.633974e-04
11	2.314495e-07	9.999999e-01	1.157247e-07
12	4.957207e-01	2.478604e-01	7.521396e-01
13	2.246219e-01	1.123110e-01	8.876890e-01
14	6.210353e-06	9.999969e-01	3.105177e-06
15	1.388362e-17	6.941811e-18	1.000000e+00
16	8.965915e-05	4.482957e-05	9.999552e-01
17	1.163225e-04	5.816127e-05	9.999418e-01

마지막으로 CLUSTER2와 CLUSTER3를 비교해보겠다.

4, 5, 12, 13 번째 변수가 차이가 유의미하지 않음을 확인할 수 있다.

따라서 상대적으로 CLUSTER2와 CLUSTER3가 유사하며, CLUSTER1과 CLUSTER2가 제일 다르다고 할 수 있다.

K-MEANS CLUSTERING을 통해 나온 결과물을 시각화 하면 다음과 같다.



각각의 군집이 잘 분리되어 있음을 확인할 수 있다.

<Hierarchical Clustering>

다음으로 계층적 군집 분석을 시행하였다.

clValid 함수를 이용해서 hierarchical clustering의 군집 수를 2부터 10까지 변화시켜 가면서 internal, stability 관련 타당성 지표를 산출하였고, 최적의 군집 수는 몇 개인지를 확인해보았다.

```
Clustering Methods:
 hierarchical

Cluster sizes:
 2 3 4 5 6 7 8 9 10

Validation Measures:
                                2       3       4       5       6       7       8       9      10

hierarchical APN            0.0003  0.0003  0.0042  0.0090  0.0176  0.0210  0.0251  0.0377  0.0729
             AD             5.3248  5.2914  5.2689  5.1870  5.1639  5.1406  5.1171  5.0968  5.0776
             ADM            0.0075  0.0074  0.0425  0.1195  0.1569  0.1938  0.2021  0.2913  0.4527
             FOM            0.9988  0.9944  0.9941  0.9812  0.9660  0.9658  0.9646  0.9575  0.9502
             Connectivity   2.9290  5.8579  8.7869 22.0956 25.0246 31.3833 34.3123 46.5571 46.5571
             Dunn           0.4033  0.4463  0.4393  0.1718  0.1718  0.1718  0.1718  0.1826  0.1826
             Silhouette     0.6777  0.6464  0.5802  0.4806  0.4291  0.3481  0.3015  0.2422  0.2125

Optimal Scores:

             Score  Method       Clusters
APN          0.0003 hierarchical 3
AD           5.0776 hierarchical 10
ADM          0.0074 hierarchical 3
FOM          0.9502 hierarchical 10
Connectivity 2.9290 hierarchical 2
Dunn         0.4463 hierarchical 3
Silhouette   0.6777 hierarchical 2
```

결과는 다음과 같다. 군집을 2에서 10까지 살펴본 결과 각각의 타당성 지표를 산출하였다.

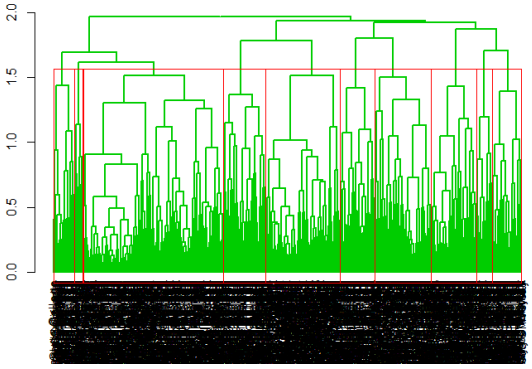
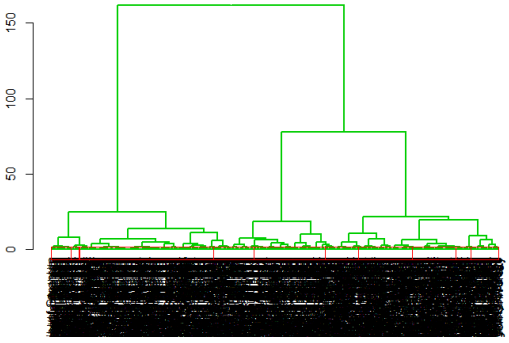
Dunn index에 의하면 최적의 군집 수는 3이며, silhouette index에 의하면 최적의 군집 수는 2이다.

총 소요시간은 약 15초 걸렸다. K-means clustering 보다 소요시간이 적었다.

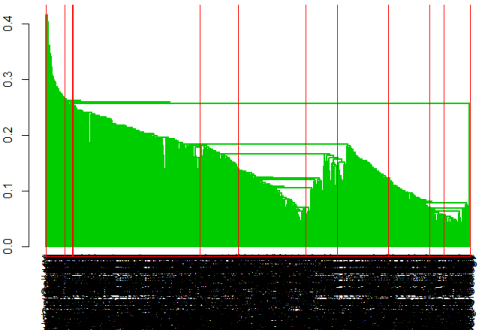
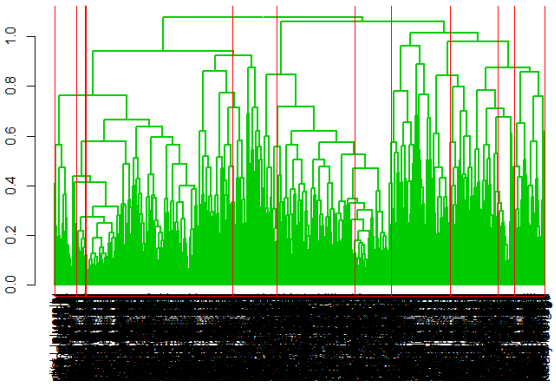
Hclust 함수를 이용하여 (method 옵션 조절하면서) dendrogram을 그려보았다.

Complete 옵션과 wardD 옵션의 경우 다음 그림과 같았다.

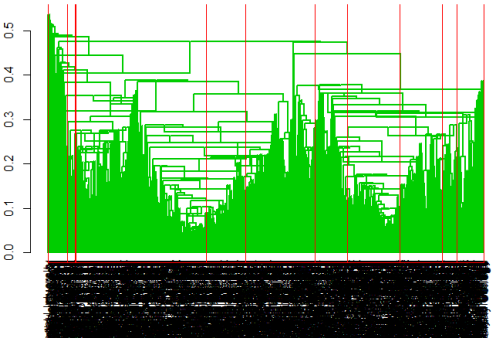
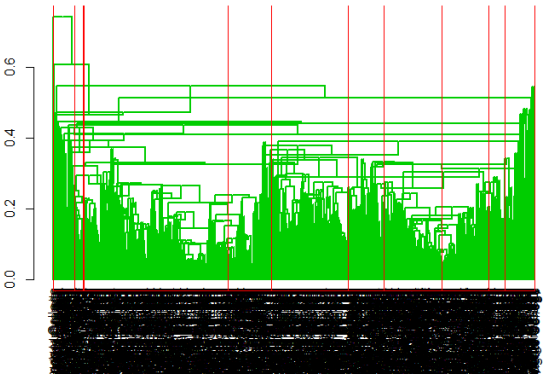
왼쪽이 complete, wardD가 오른쪽이다.

다음 아래는 single과 mcquitty 옵션이다.

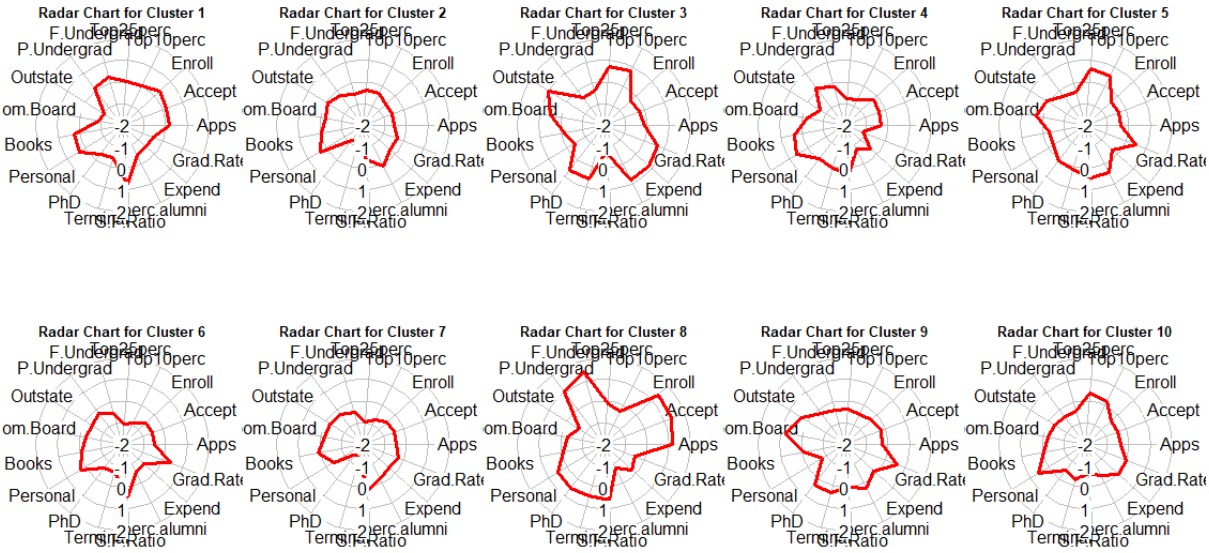
다음 아래는 centroid 와 median 옵션이다.

Complete 옵션의 경우가 군집의 크기가 가장 극단적으로 차이가 날 것으로 예상된다.

위의 빨간 선은 10개의 군집으로 나눈 것이다.

Radar chart에 도시해보자.



결과는 다음과 같다.

눈으로 확인해보니 CLUSTER2 와 CLUSTER 7이 유사해 보인다. 그리고 6과 8이 가장 다르게 보인다. 따라서 2와7을 그리고 6과8을 T-TEST 검정을 통해 비교해보았다.

```
> nc_t_result
             V1           V2          V3
1  2.856735e-02 9.857163e-01 0.014283673
2  3.479838e-02 9.826008e-01 0.017399190
3  2.855483e-02 9.857226e-01 0.014277413
4  3.475639e-08 1.737819e-08 0.999999983
5  2.716152e-07 1.358076e-07 0.999999864
6  6.351390e-02 9.682430e-01 0.031756951
7  7.232383e-01 6.383809e-01 0.361619140
8  8.996753e-02 4.498377e-02 0.955016234
9  5.675901e-01 7.162049e-01 0.283795052
10 6.057908e-01 6.971046e-01 0.302895393
11 8.322624e-06 4.161312e-06 0.999995839
12 4.355805e-02 2.177903e-02 0.978220974
13 7.156334e-01 3.578167e-01 0.642183324
14 3.228996e-03 9.983855e-01 0.001614498
15 1.630989e-02 8.154944e-03 0.991845056
16 1.363949e-01 6.819743e-02 0.931802571
17 7.203303e-01 6.398349e-01 0.360165127
```

왼쪽부터 TWO-SIDED, GREATER, LESS 순이다.

결과를 확인해보면 6번째 변수, 7번째 변수, 9번째 변수, 10번째 변수, 13번째 변수, 16번째 변수, 17번째 변수가 유의미한 차이를 보이지 않음을 확인할 수 있다. (P-VALUE가 충분히 작지 않음)

따라서 두 클러스터는 대체로 유사함을 확인할 수 있다.

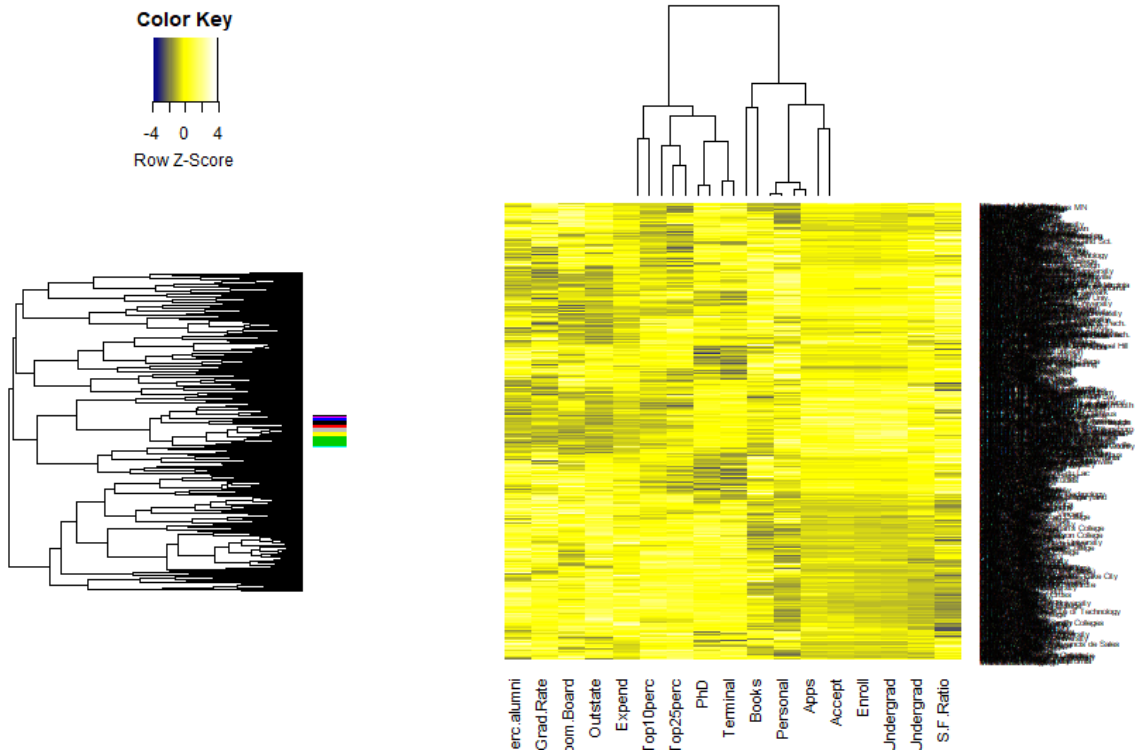
다음으로 6과 8 CLUSTER를 비교해보자.

```
> hc_t_result
             V1           V2           V3
1  1.731324e-23 1.000000e+00 8.656619e-24
2  7.637035e-25 1.000000e+00 3.818518e-25
3  2.474837e-29 1.000000e+00 1.237419e-29
4  1.087017e-10 1.000000e+00 5.435085e-11
5  1.144254e-11 1.000000e+00 5.721271e-12
6  4.324993e-32 1.000000e+00 2.162496e-32
7  5.371576e-11 1.000000e+00 2.685788e-11
8  1.646134e-02 8.230669e-03 9.917693e-01
9  9.218409e-01 5.390795e-01 4.609205e-01
10 9.834163e-01 5.082919e-01 4.917081e-01
11 2.826572e-01 8.586714e-01 1.413286e-01
12 9.902669e-08 1.000000e+00 4.951334e-08
13 5.749652e-07 9.999997e-01 2.874826e-07
14 4.219951e-01 7.890024e-01 2.109976e-01
15 1.903673e-01 9.518366e-02 9.048163e-01
16 2.749934e-06 9.999986e-01 1.374967e-06
17 4.434934e-06 2.217467e-06 9.999978e-01
```

9번째 변수, 10번째 변수, 11번째 변수, 14번째 변수, 15번째 변수가 유의미한 차이를 보이지 않음을 확인할 수 있다.

CLUSTER의 경우 7개의 변수가 유의미한 차이를 보이지 않았지만(다시 말해, 7개의 변수는 비슷함) 6번과 8번 CLUSTER는 5개의 변수에서 유의미한 차이를 보이지 않았으므로 상대적으로 6번과 8번 CLUSTER는 차이가 크다. 반면에 2번과 7번 클러스터는 상대적으로 유사하다.

HEATMAP을 사용해 도시해보자.



결과는 다음과 같이 나온다.

<DBSCAN>

PERSONAL LOAN.CSV 파일을 불러와 COLUMN을 제외하고, 평균=0, 표준편차=1로 정규화 하였다.

이제 DBSCAN을 활용해 군집을 생성해보자.

EPS	MinPts	군집 수	NOISE 수
1	5	37	1122
1	8	22	1481
1	11	20	1798
1	14	11	2117
1	17	6	2276
1.5	5	12	363
1.5	8	5	503
1.5	11	5	618
1.5	14	7	726
1.5	17	4	871
2	5	11	131
2	8	9	198
2	11	7	263
2	14	6	300
2	17	6	330
2.5	5	4	35
2.5	8	4	56
2.5	11	5	71
2.5	14	5	107
2.5	17	4	143
3	5	4	8
3	8	4	11
3	11	4	15
3	14	4	19
3	17	4	33

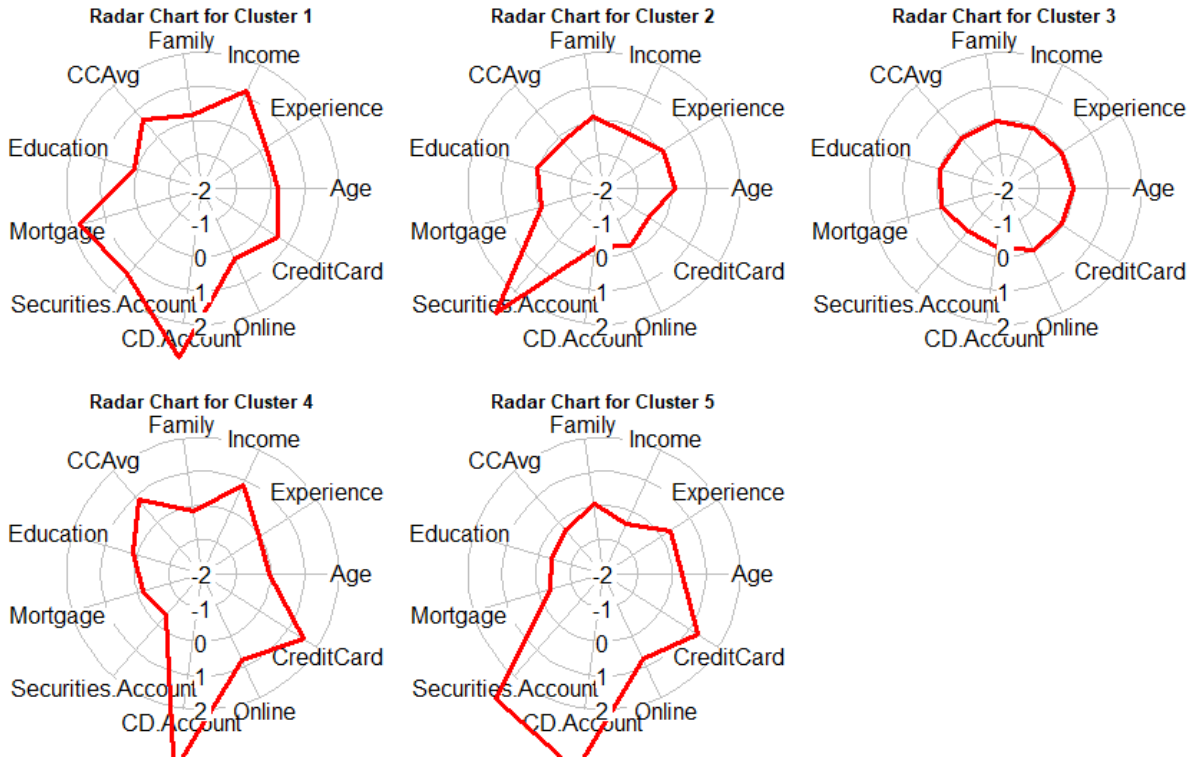
Eps/minpts 조절함에 따라 변화하는 군집 수와 noise 수를 위의 표에 정리하였다.

위의 목록 중 eps=2.5, minpts=11의 경우를 보자.

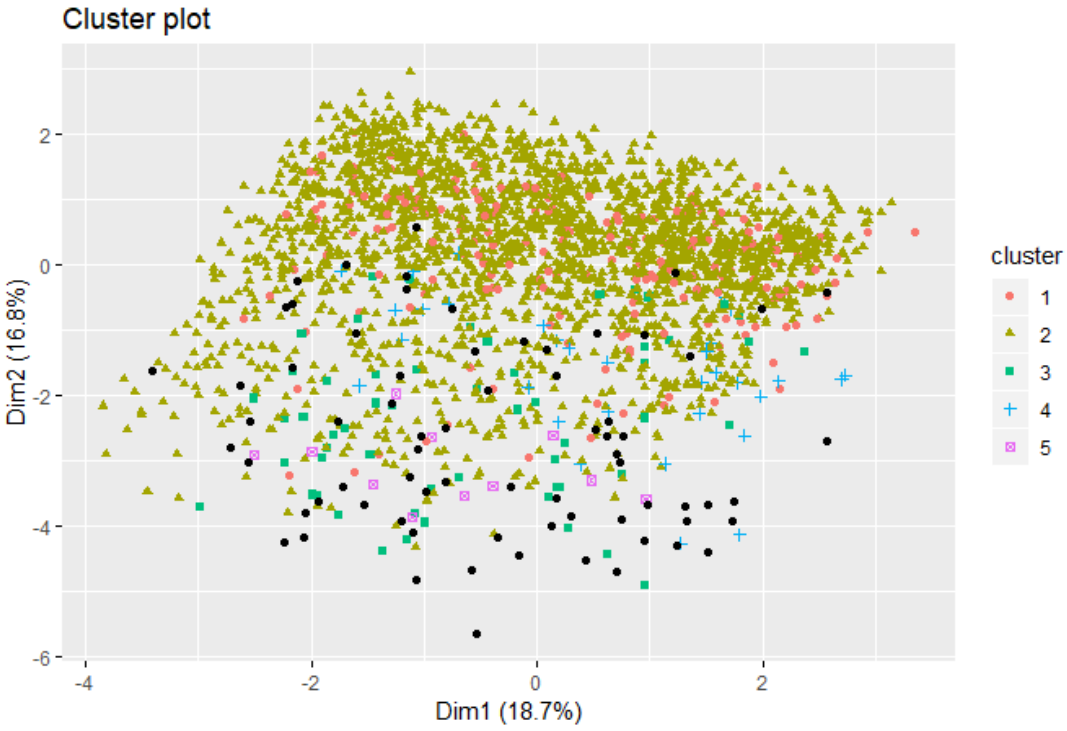
시각화 해보면 다음과 같이 cluster가 5개로 나뉘고 점들이 분포되어 있다.

아래에서 확인해보자.

일단 각각의 군집들을 잘 살펴보기 위해 radar chart에 도시해보자.



각 군집의 특성을 살펴보자. Cluster 1의 경우 Mortgage와 CDACCOUNT의 밀도가 특히 높음을 확인할 수 있다. CLUSTER2의 경우 SECURITIES ACCOUNT가 큰 값을 가진다. CLUSTER 3는 대체로 모든 변수가 비슷한 분포를 띄고 있다. CLUSTER 4의 경우 CDACCOUNT와 CREDIT CARD 변수가 두드러진 경향을 보인다. 마지막으로 CLUSTER 5의 경우 CDACCOUNT와 SECURITIES ACCOUNT 두 변수 모두에 밀도가 큰 값을 가짐을 알 수 있다. CLUSTER 1,4,5의 경우 비슷해 보이지만 변수 밀도에 조금씩 차이가 있음을 파악할 수 있다.

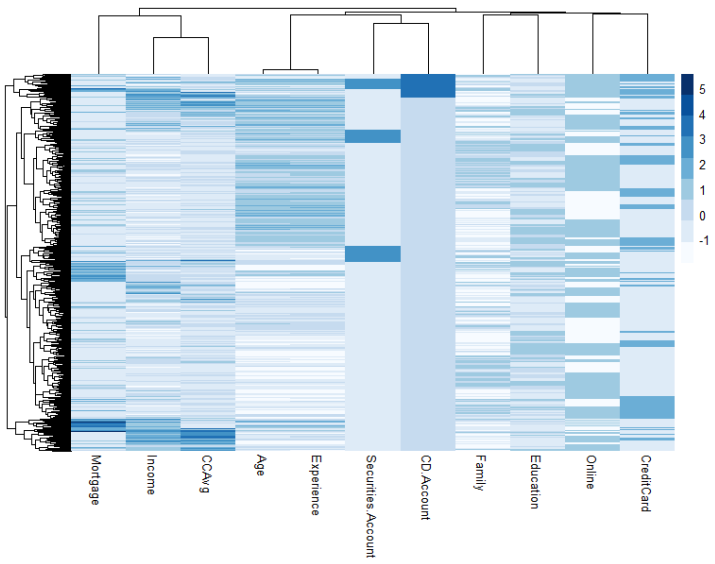


PCA를 통해 2차원으로 축소하여, 제일 영향을 크게 미치는 변수 두개를 뽑고, 선택한 군집 수를 활용해 군집과 NOISE POINT를 2차원 평면에 도시했다.

다른 CLUSTERING 기법인 KMEANS와 달리 원형의 형태로 군집이 존재하는 것이 아닌 밀도를 중심으로 퍼져 있는 상태임을 확인할 수 있다.

<Extra>

Pheatmap을 통해 hierarchical clustering 시각화 해보기



Pca를 통해 2차원으로 차원 축소한 후 다른 방식으로 2차원 평면에 도시해보기

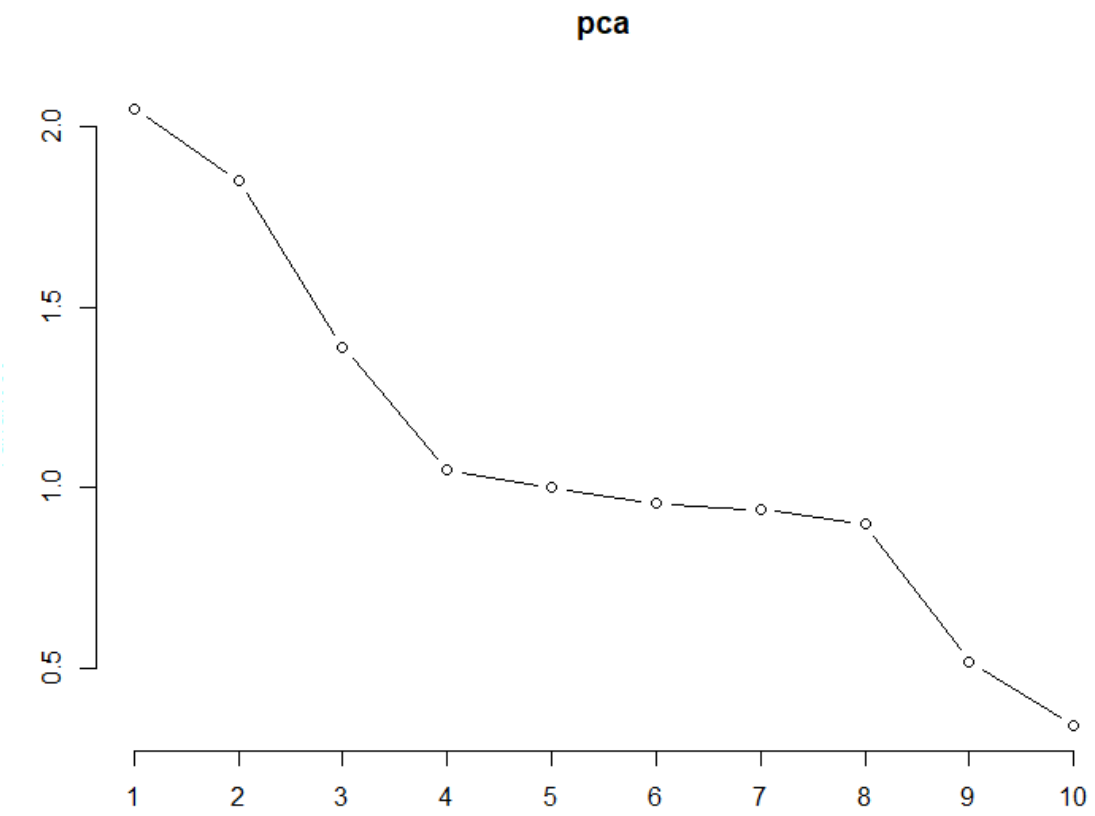
```
Importance of components:
                          PC1    PC2    PC3     PC4     PC5     PC6     PC7     PC8
Standard deviation     1.4323 1.3604 1.1789 1.02423 1.00065 0.97796 0.96932 0.94793
Proportion of Variance 0.1865 0.1683 0.1263 0.09537 0.09103 0.08695 0.08542 0.08169
Cumulative Proportion  0.1865 0.3548 0.4811 0.57647 0.66750 0.75444 0.83986 0.92155
                           PC9    PC10    PC11
Standard deviation     0.71894 0.58379 0.07285
Proportion of Variance 0.04699 0.03098 0.00048
Cumulative Proportion  0.96853 0.99952 1.00000
```

다음은 pca를 통해 각각의 변수가 얼마나 중요한 비율을 차지하는지를 보여주는 지표이다.

Ploan 데이터에서는 PC1부터 PC7까지의 변수들이 유의미한 비중을 차지함을 알 수 있다.

그 이외의 변수들은 제거해도 좋다.

SCREEPLOT을 통해서도 확인해볼 수 있다.

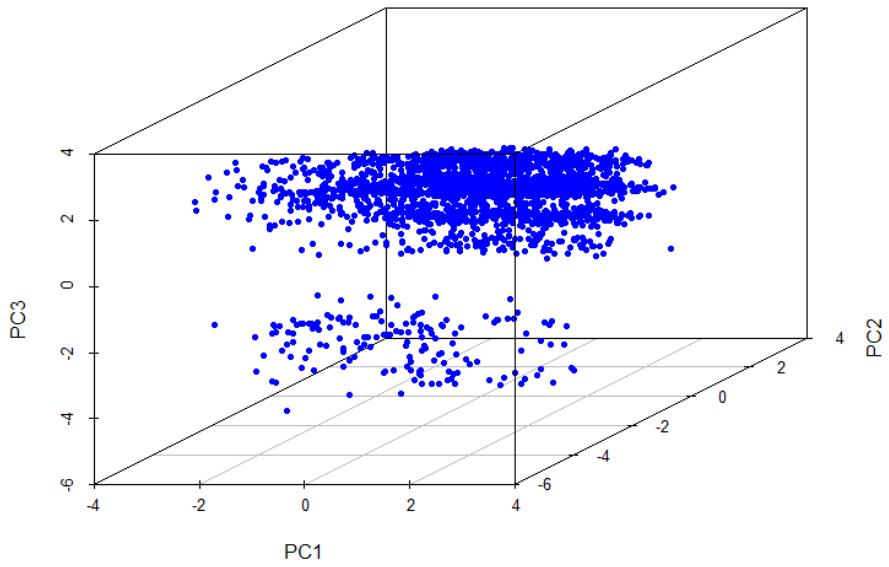


SCREEPLOT을 통해서도 각 변수들의 유의미한 정도를 시각화 해서 확인할 수 있다.

보통 언덕이 꺾이는 지점까지의 변수들이 유의미하다고 볼 수 있다.

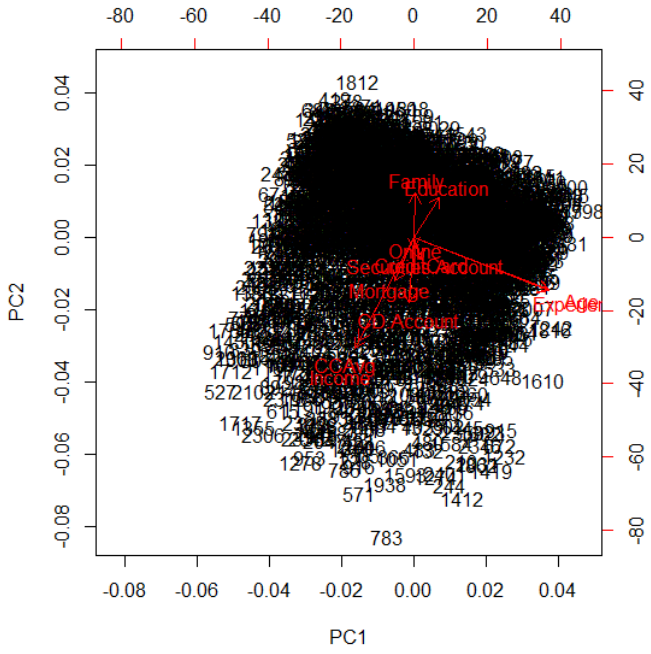
즉 위의 PLOT을 살펴보면, 짧게 끊으면 4개까지로 끊을 수 있고 길게 끊으면 8개까지 끊을 수 있다. 상황에 따라 맞는 것을 선택한다.

4개 이상의 데이터들은 시각화 할 수 없으니 가장 유의미한 변수들 3개만을 선택하여 3차원 산점도(SCATTERPLOT3D)로 도시해보자.



3차원으로 보면 위 아래로도 군집이 구분될 수 있음을 확인할 수 있다.

3차원이 아닌 두개의 변수만 선택해 2차원으로 도식화하면 다음과 같다.



화살표가 같은 방향으로 뻗어 나가는 것들이 변수와 동일한 상관관계를 가진다는 것을 의미한다.

예를 들어, 상단의 빨간 화살표 두개 FAMILY 와 EDUCATION은 PC2 변수와 강한 양의 상관관계를 갖고 있다. 하단의 빨간 화살표 두개 CCAVG와 INCOME은 PC2 변수와 강한 음의 상관관계를 가진다. 좌우로도 마찬가지다. AGE는 PC1 변수와 강한 양의 상관관계를 가진다.