

Cluster and Principal Component Analysis

Applied Empirical Economics I

Fadhil Nadhif Muharam

September 19, 2023

1. Cluster Analysis

1.1. Non-normalized data

1.1.1. Question clustering

Figure 1: Cluster analysis of questions with non-normalized data

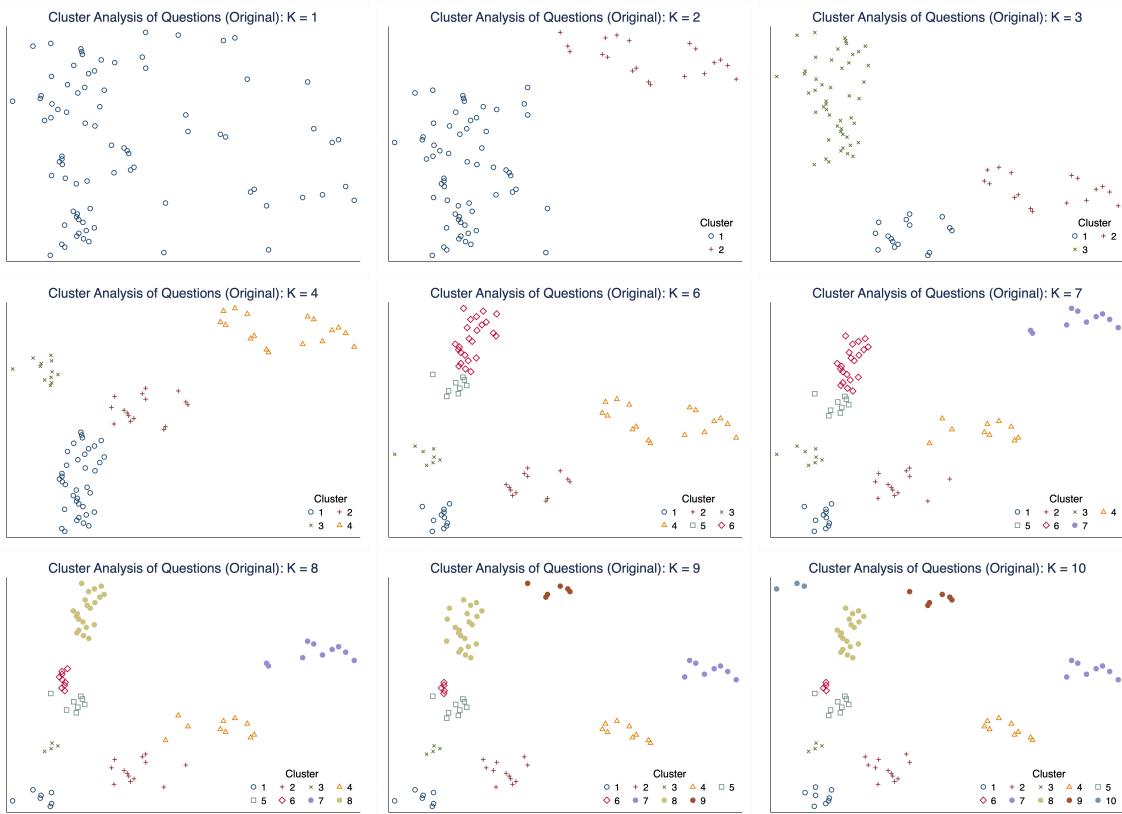


Figure 1 shows the cluster analysis when grouping the questions for different numbers of clusters. The x-axis is the means of each question, and the y-axis is a random number assignment based on groups. The top left figure is when the data has not been clustered. Starting from K=4, we see that the data is well-clustered. My preference for the number of clusters is five depicted in figure 2 below.

With five clusters, the questions are grouped into five clusters, namely, politics, life security, life importance, democracy and freedom, and traditional values explained in table 1. And as we can see, the cluster analysis provided a decent job in clustering questions that naturally belong to each other. However, there are some instances where this could be done better. For example, some of the questions regarding "confidence" are grouped into different categories, and some questions seem to not fit with the group theme. This problem could be made better with a unified scale of answers.

Figure 2: Cluster analysis of questions with non-normalized data (K=5)

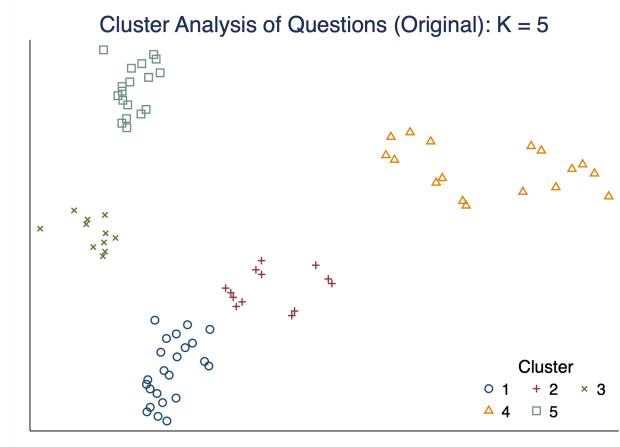
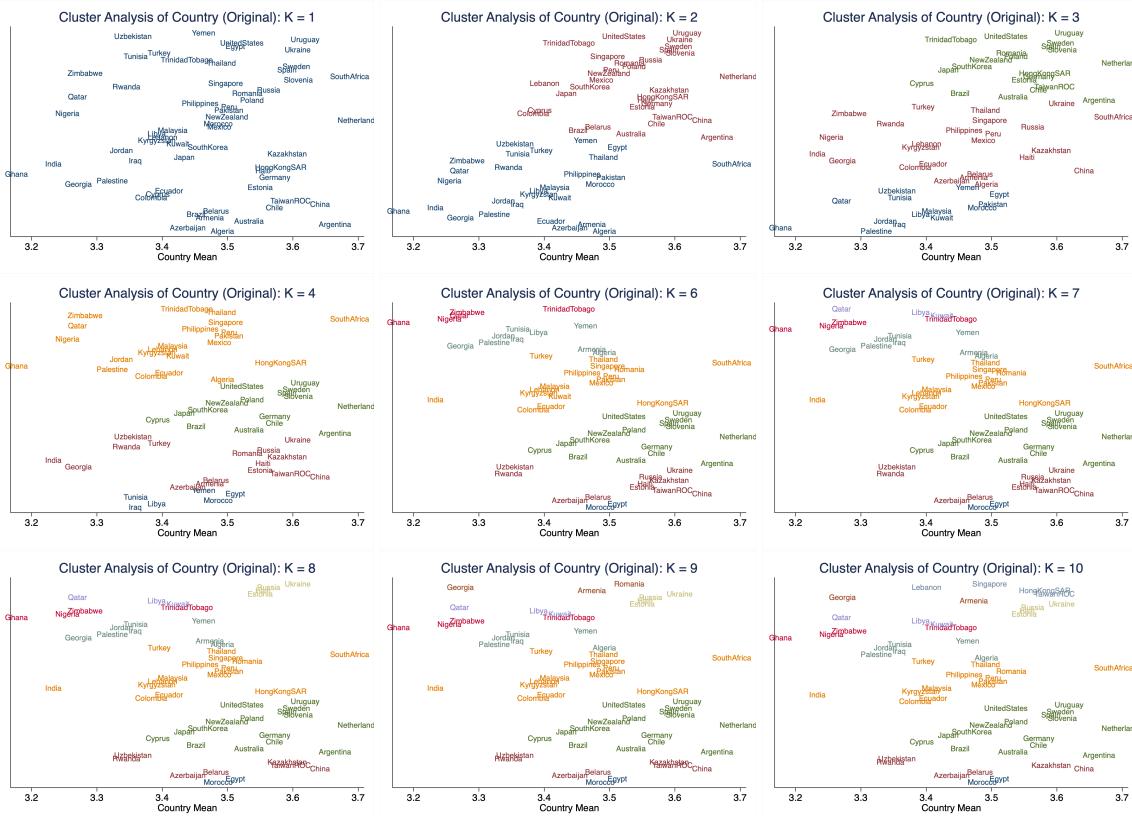


Table 1: List of cluster groups and questions for cluster analysis with K=5

Group ID	Group	Question Code	Question
1	Politics	V110	Confidence: The Press
1	Politics	V111	Confidence: Television
1	Politics	V113	Confidence: The Police
1	Politics	V114	Confidence: Justice System/Courts
1	Politics	V115	Confidence: The Government
1	Politics	V117	Confidence: Parliament
1	Politics	V120	Confidence: Major Companies
1	Politics	V121	Confidence: Banks
1	Politics	V122	Confidence: The Environmental Protection Movement
1	Politics	V123	Confidence: The Women's Movement
1	Politics	V126	Confidence: The United Nations
1	Politics	V155	All religions should be taught in public schools
1	Politics	V171	Frequency in your neighborhood: Robberies
1	Politics	V191	Frequency you/family (last 12 month): Gone without a cash income
1	Politics	V198	Justifiable: Claiming government benefits to which you are not entitled
1	Politics	V199	Justifiable: Avoiding a fare on public transport
1	Politics	V209	Justifiable: Parents beating children
1	Politics	V52	University is more important for a boy than for a girl
1	Politics	V7	Important in life: Politics
1	Politics	V70	Schwartz: It is important to this person to think up new ideas and be creative
1	Politics	V73	Schwartz: It is important to this person to have a good time
1	Politics	V75	Schwartz: It is important to this person being very successful
1	Politics	V78	Schwartz: It is important to this person looking after the environment
2	Life security	V100	Hard work brings success
2	Life security	V132	Democracy: Religious authorities interpret the laws.
2	Life security	V173	Frequency in your neighborhood: Police or military interfere with people's pri
2	Life security	V174	Frequency in your neighborhood: Racist behavior
2	Life security	V188	Frequency you/family (last 12 month): Gone without enough food to eat
2	Life security	V189	Frequency you/family (last 12 month): Felt unsafe from crime in your own home
2	Life security	V190	Frequency you/family (last 12 month): Gone without needed medicine or treatment
2	Life security	V196	It is not important for me to know about science in my daily life
2	Life security	V205	Justifiable: Divorce
2	Life security	V71	Schwartz: It is important to this person to be rich
2	Life security	V76	Schwartz: It is important to this person adventure and taking risks
2	Life security	V98	Government responsibility
2	Life security	V99	Competition good or harmful
3	Life importance	V10	Feeling of happiness
3	Life importance	V170	Secure in neighborhood
3	Life importance	V200	Justifiable: Stealing property
3	Life importance	V202	Justifiable: Someone accepting a bribe
3	Life importance	V208	Justifiable: For a man to beat his wife
3	Life importance	V210	Justifiable: Violence against other people
3	Life importance	V4	Important in life: Family
3	Life importance	V49	One of main goals in life has been to make my parents proud
3	Life importance	V5	Important in life: Friends
3	Life importance	V6	Important in life: Leisure time
3	Life importance	V8	Important in life: Work
4	Democracy and freedom	V101	Wealth accumulation
4	Democracy and freedom	V131	Democracy: Governments tax the rich and subsidize the poor.
4	Democracy and freedom	V133	Democracy: People choose their leaders in free elections.
4	Democracy and freedom	V134	Democracy: People receive state aid for unemployment.
4	Democracy and freedom	V136	Democracy: Civil rights protect people's liberty against oppression.
4	Democracy and freedom	V137	Democracy: The state makes people's incomes equal
4	Democracy and freedom	V138	Democracy: People obey their rulers
4	Democracy and freedom	V139	Democracy: Women have the same rights as men.
4	Democracy and freedom	V192	Science and technology are making our lives healthier, easier, and more comforta
4	Democracy and freedom	V193	Because of science and technology, there will be more opportunities for the next
4	Democracy and freedom	V194	We depend too much on science and not enough on faith
4	Democracy and freedom	V195	One of the bad effects of science is that it breaks down people's ideas of rig
4	Democracy and freedom	V197	The world is better off, or worse off, because of science and technology
4	Democracy and freedom	V55	How much freedom of choice and control
4	Democracy and freedom	V59	Satisfaction with financial situation of household
4	Democracy and freedom	V96	Income equality
4	Democracy and freedom	V97	Private vs state ownership of business
5	Traditional values	V108	Confidence: Churches
5	Traditional values	V11	State of health (subjective)
5	Traditional values	V119	Confidence: Universities
5	Traditional values	V124	Confidence: Charitable or humanitarian organizations
5	Traditional values	V153	Whenever science and religion conflict, religion is always right
5	Traditional values	V156	People who belong to different religions are probably just as moral as those who
5	Traditional values	V181	Worries: Losing my job or not finding a job
5	Traditional values	V182	Worries: Not being able to give one's children a good education
5	Traditional values	V183	Worries: A war involving my country
5	Traditional values	V184	Worries: A terrorist attack
5	Traditional values	V207	Justifiable: Suicide
5	Traditional values	V51	Men make better political leaders than women do
5	Traditional values	V53	Men make better business executives than women do
5	Traditional values	V54	Being a housewife just as fulfilling
5	Traditional values	V72	Schwartz: It is important to this person living in secure surroundings
5	Traditional values	V77	Schwartz: It is important to this person to always behave properly
5	Traditional values	V79	Schwartz: It is important to this person tradition
5	Traditional values	V9	Important in life: Religion

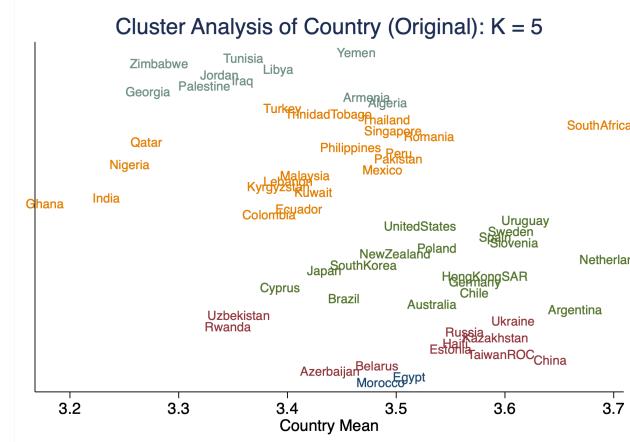
1.1.2. Country clustering

Figure 3: Cluster analysis of country with non-normalized data



Similar to the question clustering, figure 3 constitutes the clustering of countries with various numbers of clusters. The preferred number of cluster is five and is shown in figure 4.

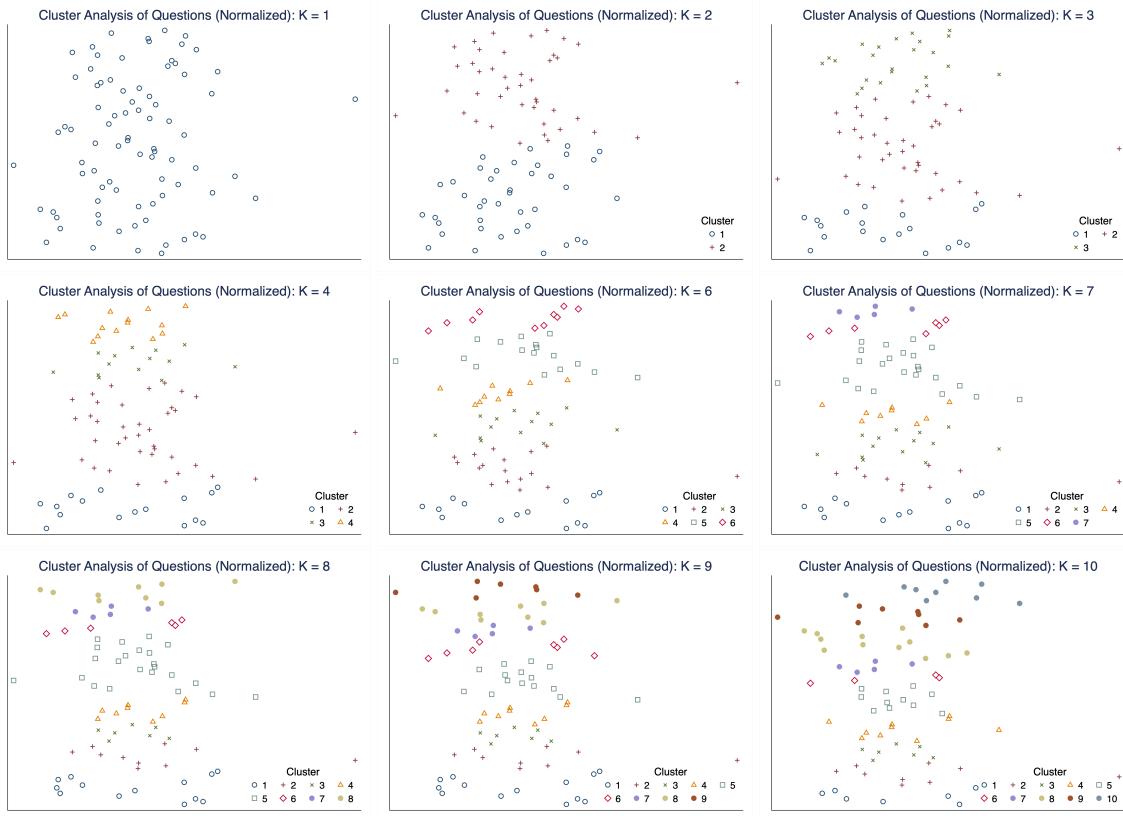
Figure 4: Cluster analysis of country with non-normalized data ($K=5$)



1.2. Normalized data

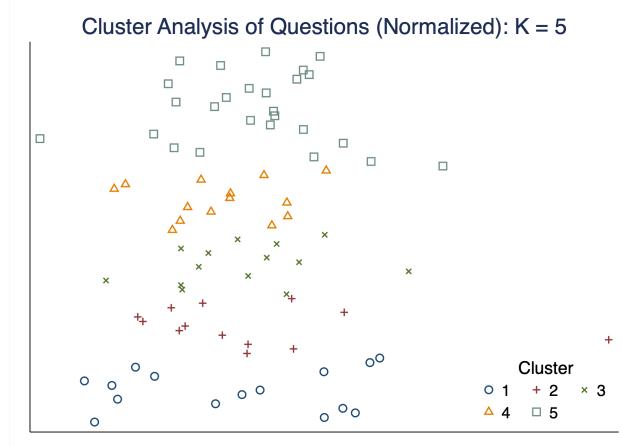
1.2.1. Question clustering

Figure 5: Cluster analysis of questions with normalized data



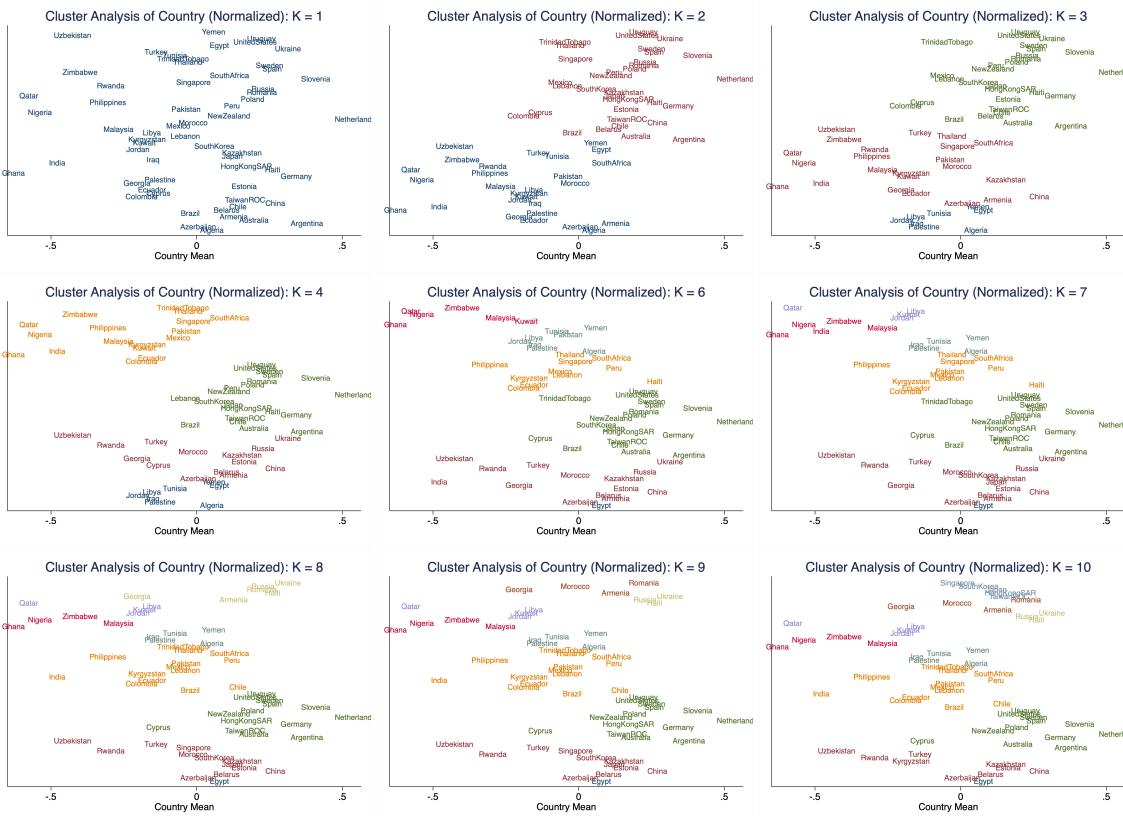
By clustering the questions for the normalized data, we can see that the resulting groups seem to have a similar dispersion of means. This is because, the way we normalized the data is that it all centered around the mean zero and standard deviation one. However, the cluster analysis made it possible to define the pattern and do a good job in clustering it similarly to the ones with non-normalized data, which although it is not seem to be shown in the graph, the resulting groups are similar with the ones with non-normalized data.

Figure 6: Cluster analysis of questions with normalized data (K=5)



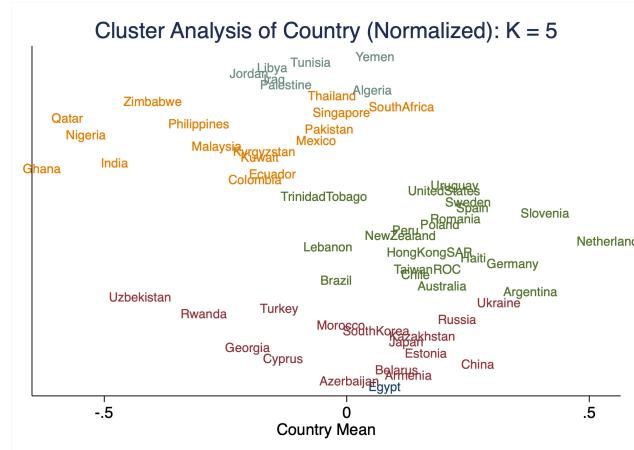
1.2.2. Country clustering

Figure 7: Cluster analysis of country with normalized data



We observe a similar pattern of clustering with the non-normalized data. This, in essence, makes sense due to the normalization that was done at the question level. Hence, unlike the normalized question clustering, we see that each cluster is grouped in a rather concentrated means.

Figure 8: Cluster analysis of country with normalized data (K=5)

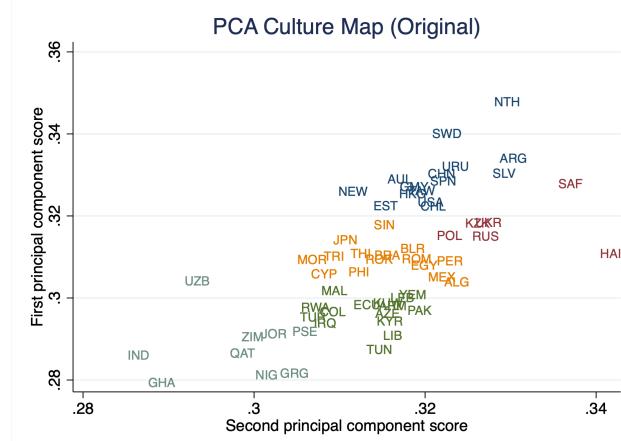


2. Principal Component Analysis

For the PCA, the figure depicted below is more of a biplot rather than a culture map. But this is interesting given how the interpretation could be different from the culture map depicted in the lecture note. The first component put more weight in the traditional values whereas the second component put more weight in the democracy and freedom. Table 2 and 3 presents how each questions are weighted in each component. I used the cluster analysis to group the countries, rather than area.

2.1. Non-normalized data

Figure 9: PCA culture map with non-normalized data



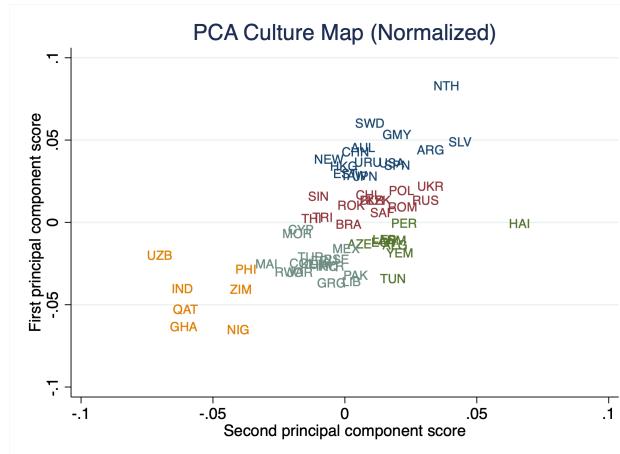
Seeing Figure 9, the result seem to show that countries with high democracy and freedom also upholds a higher traditional values. Which is quite contrasting to the result shown in the lecture note.

Table 2: PCA factor loading for two components (non-normalized data)

Question Code	Factor Loading PC1	Factor Loading PC2
V49	.2078106	.0658256
V9	.2017511	.0373875
V153	.2016577	.0314327
V132	.199843	.0211471
V205	.1874759	.079644
V53	.1823724	.0495819
V139	.1805832	.0364045
V75	.1788874	.0352874
V182	.1774179	.0284611
V51	.1764657	.0915946
V52	.1737652	.0117149
V71	.1696302	.004705
V183	.1689363	.0242158
V184	.1667707	.0295192
V77	.1618756	.0688607
V138	.1602611	.1039313
V8	.1559734	.0437046
V108	.1557414	.1039401
V79	.1544669	.0982276
V181	.1501807	.0503627
V72	.1487002	.0325913
V208	.1458341	.090243
V190	.1418065	.1053625
V76	.1339771	.0371005
V188	.1331382	.1314552
V209	.1327176	.0259985
V207	.1290189	.0856401
V6	.1284723	.0459342
V195	.1181935	.0424774
V133	.1179397	.0268561
V136	.1068502	.0415142
V191	.1066016	.0818459
V189	.105217	.1394171
V96	.0965359	.0665168
V173	.096088	.1549147
V200	.0926921	.1472956
V70	.0900791	.06666295
V202	.090005	.1505664
V171	.0879594	.1521657
V210	.0873032	.1249943
V126	.0837561	.1363711
V4	.0832893	.1597116
V156	.0831289	.0105251
V174	.0816748	.1234134
V97	.0771694	.02775
V121	.0719786	.1525045
V78	.0700914	.1473734
V192	.0679175	.1678085
V98	.0647198	.0564803
V194	.0646833	.0449819
V73	.063028	.0154043
V59	.0600082	.0839329
V137	.0584312	.0506936
V11	.0584005	.0921803
V198	.0575535	.0963539
V111	.0554647	.1431794
V99	.0517413	.1275601
V5	.0515413	.0975953
V119	.0505839	.1894179
V134	.0486164	.0669977
V100	.0474854	.1168805
V199	.046895	.1595382
V113	.0432354	.1974874
V7	.0420467	.0563234
V120	.0413593	.1927005
V123	.0391502	.1555381
V196	.0390704	.0715778
V122	.0388508	.1572322
V110	.0320941	.1464552
V193	.0296571	.1585311
V55	.0276206	.1280686
V54	.0271649	.0957843
V155	.0243402	.0559897
V115	.0243188	.2031389
V170	.0142834	.1656278
V197	.0139431	.1389542
V101	.0129306	.1452573
V124	.0074677	.2014915
V10	.0074069	.1022216
V114	.0046042	.2051757
V131	.0036992	.0142737
V117	.0018599	.1779269

2.2. Normalized data

Figure 10: PCA culture map with normalized data



The result for the normalized data is similar to that of the non-normalized ones, but a bit more centered towards the mean zero. When seeing Table 2 and 3, there seem to be not much of a difference in terms how the PCA results constituted for the normalized and non-normalized data.

Table 3: PCA factor loading for two components (normalized data)

Question Code	Factor Loading PC1	Factor Loading PC2
V49	.2078106	.0658256
V9	.2017511	.0373875
V153	.2016577	.0314327
V132	.199843	.0211471
V205	.1874759	.079644
V53	.1823724	.0495819
V139	.1805832	.0364045
V75	.1788874	.0352874
V182	.1774179	.0284611
V51	.1764657	.0915946
V52	.1737652	.0117149
V71	.1696302	.004705
V183	.1689363	.0242158
V184	.1667707	.0295192
V77	.1618756	.0688607
V138	.1602611	.1039313
V8	.1559734	.0437046
V108	.1557414	.1039401
V79	.1544669	.0982276
V181	.1501807	.0503627
V72	.1487002	.0325913
V208	.1458341	.090243
V190	.1418065	.1053625
V76	.1339771	.0371005
V188	.1331382	.1314552
V209	.1327176	.0259985
V207	.1290189	.0856401
V6	.1284723	.0459342
V195	.1181935	.0424774
V133	.1179397	.0268561
V136	.1068502	.0415142
V191	.1066016	.0818459
V189	.105217	.1394171
V96	.0965359	.0665168
V173	.096088	.1549147
V200	.0926921	.1472956
V70	.0900791	.06666295
V202	.090005	.1505664
V171	.0879594	.1521657
V210	.0873032	.1249943
V126	.0837561	.1363711
V4	.0832893	.1597116
V156	.0831289	.0105251
V174	.0816748	.1234134
V97	.0771694	.02775
V121	.0719786	.1525045
V78	.0700914	.1473734
V192	.0679175	.1678085
V98	.0647198	.0564803
V194	.0646833	.0449819
V73	.063028	.0154043
V59	.0600082	.0839329
V137	.0584312	.0506936
V11	.0584005	.0921803
V198	.0575535	.0963539
V111	.0554647	.1431794
V99	.0517413	.1275601
V5	.0515413	.0975953
V119	.0505839	.1894179
V134	.0486164	.0669977
V100	.0474854	.1168805
V199	.046895	.1595382
V113	.0432354	.1974874
V7	.0420467	.0563234
V120	.0413593	.1927005
V123	.0391502	.1555381
V196	.0390704	.0715778
V122	.0388508	.1572322
V110	.0320941	.1464552
V193	.0296571	.1585311
V55	.0276206	.1280686
V54	.0271649	.0957843
V155	.0243402	.0559897
V115	.0243188	.2031389
V170	.0142834	.1656278
V197	.0139431	.1389542
V101	.0129306	.1452573
V124	.0074677	.2014914
V10	.0074069	.1022216
V114	.0046042	.2051757
V131	.0036992	.0142737
V117	.0018599	.1779269