We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

5,800
Open access books available

142,000
International authors and editors

180M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Semantic Map: Bringing Together Groups and Discourses

Theodore Chadjipadelis and Georgia Panagiotidou

Abstract

This chapter presents a multivariate analysis method which is developed in two steps using a combination of Hierarchical cluster analysis (HCA) and Factorial Correspondence Analysis (AFC). To explain and describe the steps of the method, we use an application example on a survey dataset from young students in Thessaloniki trying to investigate their behavioral profiles in terms of political characteristics and how these may be affected about their attendance to a civic education course offered by the Political Science department in the Aristotle University of Thessaloniki. The method is explained step by step on this example serving as a manual of its application to the researcher. HCA assigns subjects into cluster membership variables and in the next stage, these new variables are jointly analyzed with AFC. Correspondence analysis manages to extract the dimensions of the phenomenon in the study, explaining the inner antithesis between the categories but also giving the opportunity to visualize the information in a two-dimensional space, a semantic map, making interpretation more comprehensive. HCA is then applied again to the AFC's coordinates of the categories constructing profiles of subjects, assigning them to the categories of the variables.

Keywords: hierarchical cluster analysis, correspondence analysis, political analysis, multivariate methods, data analysis

1. Introduction

This chapter presents a multivariate analysis method, using a combination of Hierarchical Cluster Analysis (HCA) [1] and Factorial Correspondence Analysis (AFC) in two steps [2]. The method provides the advantage of jointly handling multiple variables with many levels. The approach exploits HCA in reducing many variables into fewer ones that represent the individuals within them and then with Correspondence analysis it manages to reduce the information even further and express it upon dimensions.

These dimensions not only organize the information within the data to be explained more thoroughly but also visualizes the inner relationships among categories of the variables. By analyzing the antagonism of the clusters on different sets of dimensions, as we can also have a three-dimensional or more system of axes [3], we can understand further the behavior of the variables and their categories, as well as the associations among them.

Clustering in the final step of the coordinates of the categories on the dimensions we link the initial clusters with the categories, creating a semantic map [4] that can visualize the phenomenon in a Cartesian field or a three-dimensional space [3]. In this chapter, we present the application of the method in a specific case, which works only as an example.

The sample consists of students in Thessaloniki, Greece measuring specifically their political attitudes and their views on democracy, on moral values and the way they are informed in general about politics. In the example that is developed through the chapter we describe the application of the method and the interpretation of the results step by step.

2. Methodology

Our data analysis is based on Hierarchical Cluster Analysis (HCA) and Factorial Correspondence Analysis (AFC) in two steps [5]. The dataset is analyzed using advanced multivariate methods (Hierarchical Cluster analysis, Factorial correspondence Analysis (Analyse factorielle des correspondences AFC) [2]. Using this mixed-method approach, enables the detection of profiles of similar behavior, the association of each profile to the distinct categories that compose it and the detection of the dimension which describes the dynamics of the phenomenon, enabling the visualization of these dynamics in its final output.

In the first step, HCA assigns subjects into distinct groups according to their response patterns [2]. The main output of HCA is a group or cluster membership variable, which reflects the partitioning of the subjects into groups. Furthermore, for each group, the contribution of each question (variable) to the group formation is investigated [2], to reveal a typology of behavioral patterns. To determine the number of clusters, we use the empirical criterion of the change in the ratio of between-cluster inertia to total inertia, when moving from a partition with r clusters to a partition with r-1 clusters [6]. The metric used is chi-square. Analysis was conducted with the software M.A.D. (Methodes de l' Analyse des Donnees) [7]. In the second step, the group membership variable, obtained from the first step, is jointly analyzed with the existing variables via Multiple Correspondence Analysis on the so-called Burt table [8]. At this stage, correspondence extracts the dimensions that constitute the overall phenomenon, explaining the inner inertia between all subjects. To determine the number of factors, the empirical criterion of Benzecri was used. According to the empirical criterion of Benzecri [2], two specific sub-criteria should be fulfilled.

COR > 200 and CTR value >1000/(n + 1).

where n = total number of categories.

We proceed by applying again HCA for the coordinates of the categories on the dimensions. Bringing these two analyses steps together, we can construct a semantic map that can visualize the behavioral structure of the variables and the subjects, creating behavioral patterns and abstract discourses [4].

3. An application example in political analysis

To demonstrate the method of HCA and MCA in two steps, an example was selected to be described in the following sections. This example refers to the analysis of data collected during a survey in Thessaloniki, Greece in the period 2019–2020. The topic of the survey is to collect data about the political characteristics of young

students who participated in a civic education course offered by the Department of Political Sciences in the Aristotle University of Thessaloniki. The sample consists of 1618 participants, allocated into four groups:

Group 1: random university students within the campus of the university who were not part of the civic education course.

Group 2: university students who attended the course in-classroom.

Group 3: university students who attended the course through e-learning, due to covid-19 restrictions and measures.

Group 4: high-school students who attended the course.

The tool of the survey was a questionnaire, structured in three sections: 1) demographics, 2) political behavior, 3) information means, views on democracy and moral context.

The objective of the research is to investigate the students' levels of political knowledge, political interest, preferable way of political mobilization and distinguish the different profiles among the four groups of participants. The variables of the research -associated with each one of the questions- correspond to: a) political interest, c) political knowledge, b) political mobilization, c) their self-positioning

Code	Variable	Categories				
group	**group**	**1: random students**	**2: students in-class**	**3: students e-learning**	**4: high school students**	
lr_c	ideology	1: left	2: left-left	3: left	4: left-right	5: right
PM	political mobilization (nominal)	1: I personally address the authorities	2: I participate with others in collective mobilizations	3: I take action through Social Media	4: I let the authorities to do their job	5: I do not know / I do not answer
PI	political interest (ordinal)	1: very much	2: quite	3: a little	4: not at all	
PK	political knowledge (ordinal)	1: low	2: moderate	3: high		
E13	political info source (categorical, binary 0–1)	1: TV-Radio	2: Online newspapers-Internet	3: Social media	4: Family-relatives	5: Friends 6: Newspapers
E14	perception of democracy (categorical, binary 0–1)	12 pictures which visualize concepts for how they perceive democracy				
E15	personal values (categorical, binary 0–1)	12 pictures which visualize concepts of moral values				

Table 1.
Coding and categories of the variables used in the analysis.

on the ideological left–right axis, d) sources of information on politics e) structure of the "political" and f) "moral" self [9, 10].

More specifically, the respondents are asked directly for their level of political interest (ordinal scale) and the way they prefer to mobilize themselves on political issues which may arise (nominal scale). The variable of political knowledge (ordinal scale) is composed through the answers of the respondents on basic questions about politics, many correct answers produce a high score of political knowledge. Next, the respondents are asked to position themselves on a scale of 0 to 10 resembling the left–right ideological axis.

In the last section of the questionnaire, the questions on information sources, democratic and moral self are found. Regarding the preferable source of information, the respondents are asked to choose the two sources they use more often to get informed about politics. Moving on to the variable of "democratic self" [10], the respondent finds a set of 12 pictures, which conceptualize different versions of democracy. They are asked to choose three of them that symbolize in the best way how they understand democracy. Same wise, in the next question they are asked again to choose 3 pictures from a new set of 12 pictures, representing attitudes and views on life and moral values in general. These two sets of pictures construct symbolic representations of democratic institutions and of their personal moral compass (**Table 1**) [9].

3.1 First step of the analysis: clustering subjects into distinct groups

In this step of the analysis, we select the three variables of the last section, these are the sources of information (E13), the understanding of democracy (E14) and the moral values (E15). For these variables, we have a dataset comprising of 0–1 values, where 0 equals to a not selected picture or source and 1 to a selected one. For each one of these three sets of variables, we apply HCA, aiming to summarize the information. HCA's output is the dendrogram in **Figure 1** visualizing the clusters created in each step.

Initially, we cluster the variables to see patterns of categories. In the example below, we cluster the pictures for democracy, getting 5 clusters (38, 40, 41, 46 and 44). As seen in **Figure 2**, cluster 38 is created by the selection of pictures 3, 10 and 11, cluster 40 consists of selecting picture 1 etc.

Processing the same HCA analysis, to cluster the variables for each one of the three selected variables, we get 5 clusters for E14, 5 clusters for E15 and 4 clusters for E13, as shown in the **Table 2**.

We proceed by clustering now the subjects. Instead of having 12 binary variables to represent the democratic self, we produce clusters of similar choices and assign each one of the respondents to the clusters he is closer to according to this profile of answers. HCA again produces a dendrogram with the steps of the clustering process (**Figure 3**).

In the example shown in **Figure 4** we see how the answers on the 12 pictures on democratic self are transformed into one clustering variable (gr_dem), assigning each respondent into one of the clusters of HCA. Following the same method, a separate application of HCA for information sources and for the moral self we get the clustering variables (gr_inf) and (gr_val).

After we have completed a separate HCA, to classify the subjects (respondents) for each one of the selected variables (E14, E15 and E13) we get 8 clusters of respondents for E14 (renamed to gr_dem), 9 clusters for E15 (renamed to gr_val) and 8 clusters for E13 (renamed to gr_inf). **Table 3** shows a summary of the clusters of subjects for each one of the three variables we get the following table including the clusters and their relative frequency.
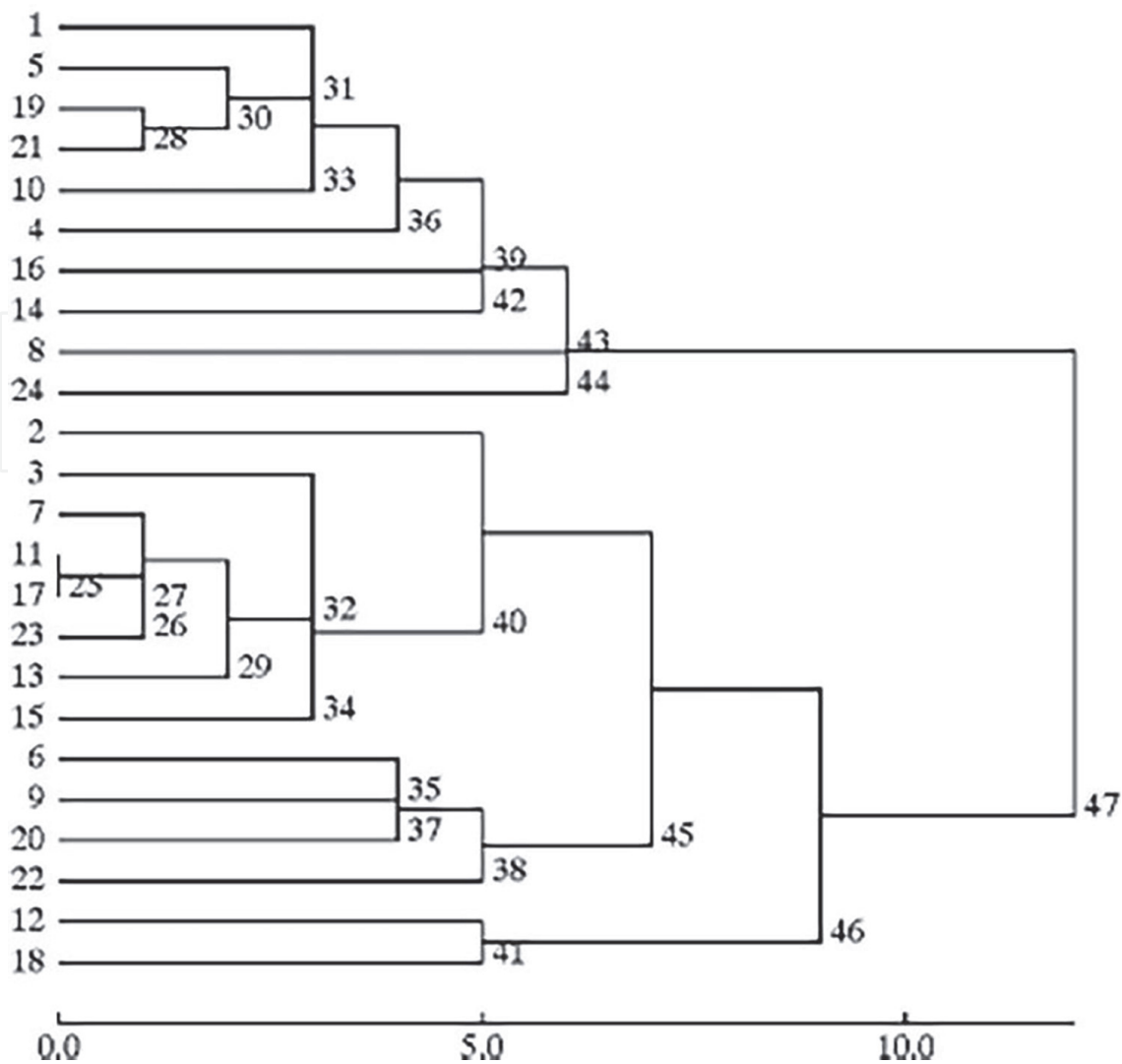
Figure 1.
Dendrogram (HCA) indicating the clusters for E14 variable.

Cluster	A(I)	B(I)	Βάρος										
25	11	17	0,1542	E1460	E1490								
26	25	23	0,2261	E1460	E1490	E14120							
27	7	26	0,29546	E1440	E1460	E1490	E14120						
28	19	21	0,12652	E14100	E14110								
29	27	13	0,36042	E1440	E1460	E1490	E14120	E1470					
30	5	28	0,18704	E1430	E14100	E14110							
31	1	30	0,24394	E1410	E1430	E14100	E14110						
32	3	29	0,411	E1420	E1440	E1460	E1490	E14120	E1470				
33	31	10	0,28227	E1410	E1430	E14100	E14110	E1451					
34	32	15	0,46164	E1420	E1440	E1460	E1490	E14120	E1470	E1480			
35	6	9	0,06781	E1431	E1450								
36	33	4	0,31501	E1410	E1430	E14100	E14110	E1451	E1421				
37	35	20	0,09113	E1431	E1450	E14101							
38	37	22	0,10794	E1431	E1450	E14101	E14111						
39	36	16	0,3477	E1410	E1430	E14100	E14110	E1451	E1421	E1481			
40	2	34	0,48807	E1411	E1420	E1440	E1460	E1490	E14120	E1470	E1480		
41	12	18	0,01246	E1461	E1491								
42	39	14	0,36606	E1410	E1430	E14100	E14110	E1451	E1421	E1481	E1471		
43	42	8	0,38002	E1410	E1430	E14100	E14110	E1451	E1421	E1481	E1471	E1441	
44	43	24	0,39145	E1410	E1430	E14100	E14110	E1451	E1421	E1481	E1471	E1441	E14121
45	40	38	0,59601	E1411	E1420	E1440	E1460	E1490	E14120	E1470	E1480	E1431	E1450
46	45	41	0,60846	E1411	E1420	E1440	E1460	E1490	E14120	E1470	E1480	E1431	E1450
47	44	46	0,99992	E1410	E1430	E14100	E14110	E1451	E1421	E1481	E1471	E1441	E14121

Figure 2.
Classification process of the 12 pictures-variables of E14 (from E141 to E1412).

We investigate further the profile of each cluster for the variable E14. Each cluster is associated with selecting a set of pictures. As shown in **Table 4** cluster 3201 consists of the respondents who are more likely to select picture number 12,

Clusters for democracy	38	40	41	44	46
pictures selected	E1431	E1411	E1461	E1451	
	E14101		E1491	E1421	
	E14111			E1481	
				E1471	
				E1441	
				E14121	
Clusters for values	**37**	**38**	**39**	**40**	**41**
pictures selected	E1531	E1591	E1511	E1571	E1541
	E1581		E1521	E15121	E1551
				E1561	
clusters for information	23	24	25	26	
pictures selected	e1311	e1361	e1321	e1381	
	e1331	e1351			
		e1341			

Table 2.
The clusters for each one of the variables (E4, E15 and E13) and the selected pictures they are linked to.
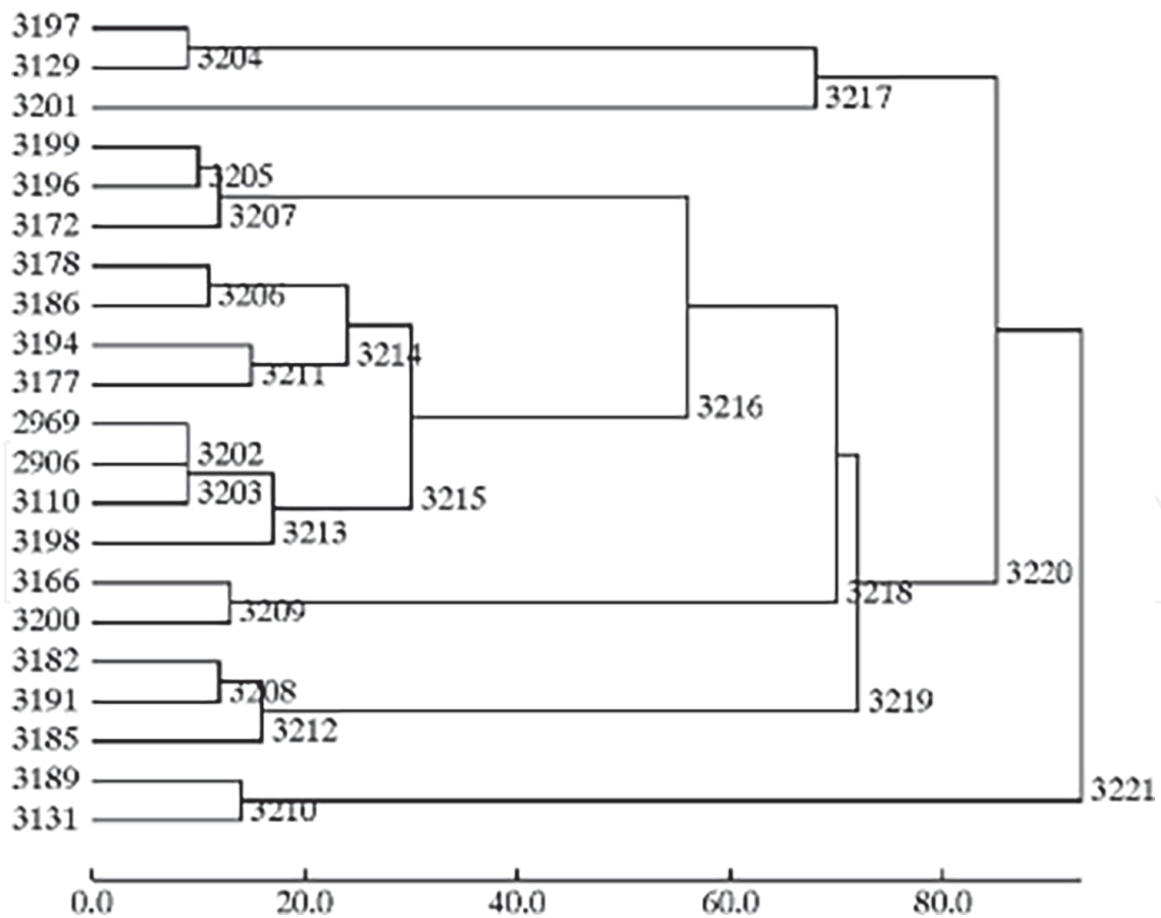

Figure 3.
Dendrogram (HCA) indicating the clusters of subjects for E14 variable.

which corresponds to the symbolic representation for religion (**Table 5**). Cluster 3204 relates to selecting pictures 4,5,9 and 12 (e-democracy, representative, clientelism and religion). The sets of pictures connected to the clusters, depict the

ind	E141	E142	E143	E144	E145	E146	E147	E148	E149	E1410	E1411	E1412
1	1	0	1	0	0	0	0	0	1	0	0	0
2	0	0	0	0	1	0	0	0	0	1	0	1
3	0	1	0	0	1	0	1	0	0	0	0	0
4	0	1	0	0	1	0	1	0	0	0	0	0
5	0	1	1	0	0	0	0	0	0	1	0	0
6	0	0	0	0	1	0	0	0	1	0	0	1
7	0	0	0	0	1	0	1	0	0	0	0	1
8	1	0	0	0	1	0	1	0	0	0	0	0

ind	group	lr_c	PK	PM	PI	gr_inf	gr_dem	gr_val
1	1	5	1	9	2	3213	3204	3200
2	1	4	2	1	3	3215	3201	3206
3	1	4	1	4	3	3215	3207	3191
4	1	3	2	4	2	3208	3207	3203
5	1	3	2	3	2	3208	3214	3192
6	1	5	2	9	2	3213	3204	3206
7	1	3	1	4	3	3213	3201	3191

Figure 4.
Transforming the dataset by replacing the binary E141-E1412 with the cluster membership variable gr_dem.

gr_dem	freq%	gr_val	freq%	gr_inf	freq%
3201	12%	3187	4%	3136	11%
3204	7%	3191	7%	3198	4%
3207	14%	3192	10%	3206	12%
3209	14%	3200	9%	3208	15%
3210	6%	3202	13%	3211	15%
3212	15%	3203	15%	3213	13%
3213	14%	3204	12%	3215	16%
3214	18%	3206	13%	3216	13%
		3207	16%		

Table 3.
Cluster membership variables and their categories for E14, E15 and E13.

E14/gr_dem	3201	3204	3207	3209	3210	3212	3213	3214
E1411						40,451	27,1001	
E1421			21,534				18,6383	27,784
E1431						57,626		20,2902
E1441		41,865		82,1471				
E1451		16,437	11,9273				17,5035	
E1461					154,0449			

E14/gr_dem	3201	3204	3207	3209	3210	3212	3213	3214
E1471			67,2476					
E1481							11,8125	11,1089
E1491		122,6539			22,5595			
E14101					30,127	10,2896		34,607
E14111					13,3978	71,9056		
E14121	93,9969	7021						

Table 4.
Weight of selecting each picture to the creation of the clusters for E14.

Democracy		3201	3204	3207	3209	3210	3212	3213	3214
		dem_1	dem_2	dem_3	dem_4	dem_5	dem_6	dem_7	dem_8
Movement	E1411						X	X	
Ancient Greece	E1421			X				X	X
Direct	E1431						X		X
e-Democracy	E1441		X		X				
Representative	E1451		X	X				X	
Riot	E1461					X			
Deliberation	E1471			X					
Volunteerism	E1481							X	X
Clientelism	E1491		X			X			
Rebellion	E14101					X	X		X
Protest	E14111					X	X		
Religion	E14121	X	X						
%Count		11.9%	7.4%	14.0%	13.6%	6.0%	15.0%	14.2%	18.1%

Table 5.
Summarizing the content of each cluster and renaming the clusters for E14.

E15/gr_val	3187	3191	3192	3200	3202	3203	3204	3206	3207
E1511	188,512								
E1521	20,9268	118,983		54,775					
E1531				12,0584				58,0211	
E1541	18,172			121,4029					
E1551				79,232			74,1092		
E1561			73,3846						
E1571			10,2587			48,654			15,5155
E1581								18,882	15,8182
E1591							44,774	82,153	23,8393
E15101						74,8128			
E15111	22,4576	52,603			80,7176				
E15121			23,373						19,8778

Table 6.
Weight of selecting each picture to the creation of the clusters for E15.

different profiles of the respondents according to the way they comprehend democracy.

Similarly, for variable E15, we describe the profiles of the cluster of the respondents regarding the pictures they are more likely to select. In **Table 6** we see that cluster 3187 is connected to the pictures 1, 2, 4 and 11 which correspond to riot, anonymous, army and protest, a representation of expressivist moral values (**Table 7**). In contrast, we see cluster 3207 having a completely naturalist moral values as it is connected to pictures 7, 8, 9, 12 (mountain, family, intimacy and concert).

Once more, we investigate the content of each cluster for the variable E13, regarding sources of information. Cluster 3136 includes those respondents who answer 1 and 3 (**Table 8**) which translates into preferring to get informed about politics by TV-radio and family (**Table 9**).

3.2 Second step: joint analysis of the cluster membership variables

In the second step of the analysis, we jointly analyze the initial variables together with the new cluster membership variables gr_dem, gr_var and gr_inf. We repeat the steps as in the early stages of the analysis applying HCA which produced the following clusters for the subjects, as w result 8 clusters of respondents are detected (**Table 10**).

These clusters relate to the categories of the variables creating a behavioral profile for each one of the clusters of the respondents, in which they have been assigned accordingly. In **Table 11** the profiles of the clusters are given in full detail, e.g., cluster 3155 consists of respondents who belong to group 4, are men [sex1], they characterize themselves as center-left [lr_c2], have moderate political knowledge [PK2], they choose to mobilize by personally addressing the authorities, take action through social media and/or let the authorities to do their job [PM1, PM3 and/ or PM4],have a little political interest [PI3]. Furthermore, respondents in this cluster belong also in cluster 3136, 3208 and 3216 on how they get informed on politics, they belong to clusters 3207,3209, 3213 and 3214 regarding their views on democracy, and finally they belong in cluster 3192 regarding their set of moral values.

In the same way, we continue to examine each one of the clusters of the respondents to understand their behavioral profile, considering the total number of the variables used in our analysis.

In the next step, with the application of correspondence analysis, we extract the dimensions of the analysis and a set of coordinates for each one of the dimensions for each one of the variable categories (**Table 12**).

An extra but final step of HCA is applied this time on the coordinates of the categories classifying them into groups (**Figure 5**).

The analysis highlights the existence of 10 distinct discourses of behavior (**Table 13**):

a. Clusters 51, 62, 87 which is a later step are unified in one cluster 98. This cluster reflects the profile of group 2 and 3 (university students who undertook the civic education course either in-class either online). They are characterized as far left, with high political interest, collective political mobilization, get informed by social media, internet or the newspapers. They see democracy as direct and think of it as rebellion and protest, while in their moral set of values they choose protest (expressivists).

b. Clusters 72, 84, 85, 86 which in later classification stage merge into cluster 94, including the random sample of students who were not part of the civic education course. These participants are characterized as center-left/center-

Values	Picture		3187	3191	3192	3200	3202	3203	3204	3206	3207
			val_1	val_2	val_3	val_4	val_5	val_6	val_7	val_8	val_9
Expressivist	Riot	E1511	X								
Expressivist	Anonymous	E1521	X	X		X					
Christian	Christ	E1531				X				X	
Army	Army	E1541	X			X					
Naturalist	Money	E1551				X			X		
Moon exploration	Astronaut	E1561			X						
Spirituality	Mountain	E1571			X			X			P
Naturalist	Family	E1581								X	P
Naturalist	Intimacy	E1591							X	X	P
Spirituality	Meditation	E15101						X			
Expressivist	Protest	E15111	X	X			X				
Naturalist	Concert	E15121			X						P
%Count			4.2%	7.2%	9.8%	8.6%	13.4%	15.4%	12.4%	13.5%	15.6%

Table 7.
Summarizing the content of each cluster and renaming the clusters for E15.

E13/gr_inf	3136	3198	3206	3208	3211	3213	3215	3216
e1311	29,2252			21,2416	29,2252			
e1321						97,3186		
e1331	52,5758							40,4936
e1341							81,3803	
e1351			38,546		26,4659			78,426
e1361			34,7882	36,0722				
e1381		181,9963						

Table 8.
Weight of selecting each source of information to the creation of the clusters for E13.

Info Source		3136	3198	3206	3208	3211	3213	3215	3216
		inf_1	inf_2	inf_3	inf_4	inf_5	inf_6	inf_7	inf_8
TV-Radio	e1311	X			X	X			
Newspapers	e1321						X		
Family	e1331	X							X
Friends	e1341							X	
Social Media	e1351			X		X			X
internet	e1361			X	X				
No information	e1381		X						
%Count		11.0%	4.5%	12.3%	15.1%	14.6%	13.0%	15.9%	13.5%

Table 9.
Summarizing the content of each cluster and renaming the clusters for E13.

Cluster	Freq%
3155	5%
3170	6%
3174	6%
3177	8%
3185	38%
3187	11%
3192	17%
3194	8%

Table 10.
Clustering for the subjects using all the variables together with the new cluster membership variables, produced in the first step.

	3155	3170	3174	3177	3187	3194	3185	3192
group1			11,8463			25,592	82,319	26,596
group2		148,5301						
group3				125,4921				
group4	10,9198				14,5687	34,229		55,459

	3155	3170	3174	3177	3187	3194	3185	3192
sex1	76,276					29,353		14,1511
sex2		23,639	71,799	75,205	6207		20,234	
lr_c1		30,833	38,7565	90,106		95,594		
lr_c2	93,414	83,845	10,519					
lr_c3					48,308		62,839	
lr_c4		5919			85,217			
lr_c5					2274	79,067		14,2434
PK0						79,321	53,899	
PK1			98,862		33,421	16,697	25,422	27,604
PK2	15,5399	76,055			3605			26,049
PK3		18,3509			50,077			19,997
PK9				125,4921				
PM1	34,843	78,408					45,907	
PM2		73,697	34,2217	30,939		58,304		22,388
PM3	92,603				21,341			
PM4	38,361			21,751	69,315			29,301
PM9						18,0682	25,506	
PI1		21,7371	99,778	23,305				57,641
PI2		86,385	47,958	87,025				16,576
PI3	81,295				68,028		61,683	
PI4						31,3818	1344	
gr_inf3136	35,692		17,137		40,6787			40,057
gr_inf3198						95,479		
gr_inf3206		17,9617	72,514	18,0647				
gr_inf3208	12,1097	13,512		72,484			3135	
gr_inf3211					60,759		16,943	35,994
gr_inf3213		10,4056	20,2326					17,819
gr_inf3215							31,174	
gr_inf3216	68,826						68,826	25,316
gr_dem3201					45,4001			
gr_dem3204								36,4704
gr_dem3207	23,595	16,3147		38,131			36,193	
gr_dem3209	54,348			60,254			45,489	
gr_dem3210						36,1486		17,7273
gr_dem3212			48,5952					
gr_dem3213	36,193						76,896	
gr_dem3214	14,6412	61,185	54,366	58,628				
gr_val3187						92,596		
gr_val3191								38,8079
gr_val3192	114,2029			20,676				

	3155	3170	3174	3177	3187	3194	3185	3192
gr_val3200								35,4314
gr_val3202		22,3301	82,6908	36,965				
gr_val3203		25,968		61,137			12,4995	
gr_val3204							46,966	58,266
gr_val3206				20,116	39,9724			
gr_val3207					5277		11,5083	

Table 11.
Association between the clusters produced in the second step and the categories of the analysis.

categories	x	y
group1	−135	18
group2	415	−276
group3	1192	865
group4	−179	−221
sex1	−22	−160
sex2	15	112
lr_c1	726	−466
lr_c2	78	−13
lr_c3	−235	156
lr_c4	−75	141
lr_c5	−38	−308
PK0	−208	79
PK1	−165	−27
PK2	−94	−86
PK3	−18	−236
PK9	1192	865
PM1	−45	139
PM2	494	−432
PM3	−12	1
PM4	−186	193
PM9	−211	−41
PI1	712	−262
PI2	113	9
PI3	−201	85
PI4	−414	−17
inf_1	−270	−24
inf_2	−518	−167
inf_3	381	181
inf_4	160	112
inf_5	−219	25

categories	x	y
inf_6	283	−206
inf_7	15	19
inf_8	−198	−70
dem_1	−217	31
dem_2	−269	−105
dem_3	−44	215
dem_4	−1	186
dem_5	32	−703
dem_6	371	−316
dem_7	−140	140
dem_8	78	98
val_1	234	−661
val_2	86	−257
val_3	−39	194
val_4	−149	−285
val_5	550	−387
val_6	−31	309
val_7	−137	44
val_8	−160	134
val_9	−192	202

Table 12.
Coordinates for each one of the categories on two main dimensions (x,y).

Cluster	A(I)	B(I)	Weight											
51	3	16	0,01903	group3	PK9									
62	57	39	0,06357	lr_c1	PM2	gr_val3202	gr_dem3212							
72	66	29	0,06336	PM1	gr_dem3207	gr_inf3208								
84	82	43	0,12708	group1	PK1	PK0	gr_dem3209	gr_val3191						
85	83	77	0,31495	sex2	PI3	lr_c3	PK2	gr_val3207	gr_dem3213	gr_val3203	gr_inf3215	lr_c4	PM4	gr_inf3211
86	79	75	0,17293	sex1	PI2	PM3	gr_inf3216	lr_c2	gr_dem3214	gr_val3192				
87	78	65	0,05044	group2	gr_inf3206	PI1	gr_inf3213							
89	60	63	0,04217	lr_c5	gr_val3200	gr_dem3204	gr_val3204							
92	81	56	0,09893	group4	PK3	gr_inf3136	gr_dem3201	gr_val3206						
93	69	53	0,04737	PM9	PI4	gr_inf3198	gr_dem3210	gr_val3187						

Figure 5.
Clustering the variables using their coordinates on the dimension as input.

right, have a moderate to low political interest, little to none political knowledge, low political mobilization (letting others do their job) or social media, they get informed by tv-radio, social media, friends and family. They view democracy as movement, representative, direct and they see a strong connection to ancient Greece. Their moral values are mainly naturalist, focusing on entertainment, family or spirituality.

c. In clusters 89, 92 which meet later in cluster 95, we find the younger high school students, who also attended the civic education course. This cluster is characterized as closer to the righter positions of the left–right axis. They demonstrate high political knowledge, they get informed by TV-radio and family and they see democracy as e-democracy, representative and connected

10 clusters	51	62	87	72	84	85	86	89	92	93
6 clusters	51	91	91	88	90	90	88	95	95	93
4 clusters	98	98	98	94	94	94	94	95	95	93
				96	96	96	96	96	96	93
group	group3		group2		group1				group4	
Left–Right		far left				center-left/center-right	center-left	far right		
Political Interest			Very			Not very	Somewhat			Not at all
Political Knowledge	No Data				None/Little	Adequate		High		
Political Mobilization		Collective		Personal		Let others to do their job	Social Media			N/A
Gender						Female	Male			
Information Source		Social media, Internet/ Newspapers	TV-Radio, Internet			TV-Radio, social media/ Friends	Family, social media		TV-Radio, Family	No information
Democracy	Movement, Direct, Rebellion, Protest		Ancient Greece, Representative, Deliberation	e-Democracy		Movement, Ancient Greece, Representative, Volunteerism	Ancient Greece, Direct, Volunteerism, Rebellion	e-Democracy, Representative, Corruption	Religion	Riot, Corruption, Rebellion, Protest
Values	Protest			Anonymous, Protest		Spirituality, meditation/ Mountain, family, intimacy, concert	Astronaut, mountain, concert	Anonymous, Christ, money, army/ Money intimacy	Christ, family, intimacy	Riot, anonymous, army, protest

Table 13.
Summarizing the association between the categories and the clusters.

to corruption and religion. Their moral setting is a mixture of expressivist and naturalistic values, including a set of nationalist symbolism including army, Christ, and family.

d. Cluster 93 concentrates respondents of no political interest, or information who understand democracy as rebellion or corruption and are closer to expressivist values such as riot, protest but also army.

4. Final output: the semantic map

Utilizing the coordinates of the points on the two first axes which were obtained from the correspondence analysis, we construct a system of 2 axes on which we place all these points [3]. The output resembles a simple Cartesian field where x is the first dimension (horizontal), and y is the second dimension (vertical). A third dimension can be brought into the analysis by using a three-dimensional space, visualizing the objects within a cube, or by presenting the different sets of the dimension by two.

The output is a semantic map, where all objects can be seen altogether, and their positioning on the field can be explained in terms of the object's proximity or opposition on each one of the dimensions.

In our example (**Figure 6**), we make the following observations:

The first axis is created by the opposing objects of: 1) group 1 (random students) and group 4 (high school students), followed by characteristics such as low political interest, getting informed by V-radio or friend and family, center left\center right, naturalistic values, choosing not to be mobilized or act on an individual level if needed and 2) group 2 and group 3 (university students of the civic education course)
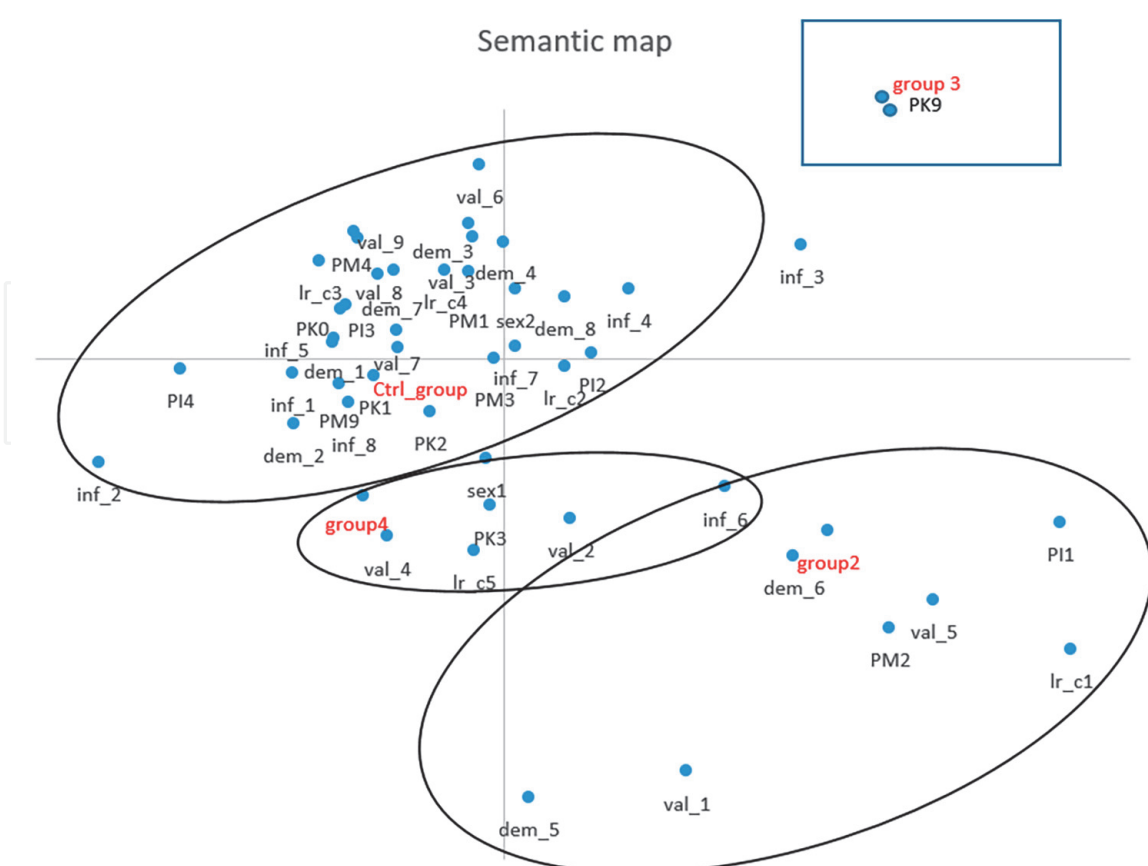

Figure 6.
The semantic map, visualizing in a Cartesian field (x,y) the categories of all variables positioned according to their coordinates from AFC.

with high political interest, left, getting informed by newspapers and social media, expressivists choosing collective ways of mobilization.

The second axis depicts the antithesis between group 3 (online students of the civic education course) who are connected to the online information about politics, in contrast to the in-class students of group 2 who are linked to collective ways of mobilization. Additionally, the second axis is described by the antithesis between the set val_1 (Riot, Anonymous, Army, Protest), dem_5 (Riot, Deliberation, Volunteerism, Clientelism, Rebellion, Protest) and the set val_6/val_9 (Mountain, Family/ Mountain, Family, Intimacy) and dem_3/dem_4 (Ancient Greece, Representative, Deliberation /e-Democracy). This polarization is explained as the difference between the democratic and moral discourses which were detected in the analysis.

5. Conclusion

The method presented in this chapter, as applied in the example of a survey among universities and high school in Thessaloniki, follows the application of HCA and MCA (or AFC) in two steps.

The added value of the presented methodological approach lies in its competence to utilize an advanced clustering method that incorporates the dimension reduction function of correspondence analysis. Clustering in multiple stages of the analysis, produces summarized variables that can describe the overall behavior or profile of the subjects. Then these new cluster membership variables can be associated with the categories of the variables used in the clustering analysis, therefore we can associate each cluster not only with its subjects but with the categories as well. In the second step, the joint analysis of the cluster membership variables together with the rest of the variables of the analysis, produces a comprehensive clustering of all items together, associating them again with the categories of the variables. This procedure allows the researcher to have a full and comprehensive overview of the profiles of each cluster.

Moreover, correspondence analysis brings forward the inner competition of the phenomenon, extracting multiple dimensions that explain the dynamics within it. The coordinates of each object give a better understanding of the distances between them, and when analyzed again with HCA we get the final fully described clusters. The coordinates can visualize the phenomenon in a simple two-dimensional space or even of more dimensions, where the observer can comprehend in more detail the revealed inner relationships or oppositions among the subjects and the objects of the analysis.

Author details

Theodore Chadjipadelis and Georgia Panagiotidou
Aristotle University of Thessaloniki, Greece

*Address all correspondence to: chadji@polsci.auth.gr

References

[1] Galbraith JI, Bartholomew DJ, Moustaki I, Steele F. The Analysis and Interpretation of Multivariate Data for Social Scientists. London: Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences; 2002

[2] Benzècri JP. L'analyse des donnees. Tome 2: L'analyse des correspondances. Paris: Dunod; 1973

[3] Greenacre M. Biplots in Practice. Bilbao: Fundación BBVA; 2010

[4] Panagiotidou G, Chadjipadelis T. First-time voters in Greece: Views and attitudes of youth on Europe and democracy. In: Chadjipadelis T, Lausen B, Markos A, Lee TR, Montanari A, Nugent R, editors. Studies in Classification, Data Analysis and Knowledge Organization. Springer, Cham; 2020. pp. 415-429

[5] Chadjipadelis T. Parties, Candidates, Issues: The Effect of Crisis, Correspondence Analysis and Related Methods. Napoli, Italy: CARME; 2015

[6] Papadimitriou G, Florou G. Contribution of the Euclidean and chi-square metrics to determining the most ideal clustering in ascending hierarchy (in Greek). In: Annals in Honor of Professor I Liakis. Thessaloniki: University of Macedonia; 1996. pp. 546-581

[7] Karapistolis D. Software Method of Data Analysis MAD [Internet]. 2010. Available from: http://www.pylimad.gr/ [Accessed: January 25, 2022]

[8] Greenacre M. Correspondence Analysis in Practice. Boca Raton: Chapman and Hall/CRC Press; 2007

[9] Marangudakis M, Chadjipadelis T. The Greek Crisis and its Cultural Origins. New York: Palgrave-Macmillan; 2019

[10] Taylor C. Sources of the Self. Cambridge, MA: Harvard University Press; 1991