

RESEARCH ARTICLE

Demographic and Socioeconomic Homogeneity among Districts and District Towns in Bangladesh

Md. Belal Hossain^{1,*} and Mohammad Sorowar Hossain¹

¹Biomedical Research Foundation, Dhaka, Bangladesh

*Corresponding author: belal.hossain@brfbd.org; bhossain@isrt.ac.bd

Received: December 10, 2018; revised: January 25, 2019; accepted: January 28, 2019.

Abstract: Exploring the homogeneity (or heterogeneity) at sub-national level is crucial as it is associated with design, budget allocation and implementation of a research project. Since demographic and socioeconomic factors depict the first valuable insight of a community, it is imperative to explore the homogeneity within a country by considering these variables. Yet, the information on this aspect is scarce in Bangladesh. Therefore, the present study aimed to identify the district and district town specific homogeneity in Bangladesh. The data for this study were extracted from the most recent Population and Housing Census of the country, and the multivariate cluster analysis was employed to identify the natural groups or segments. We found that Bangladesh could be classified into three distinct clusters both at district and district town levels based on demographic and socioeconomic characteristics. The findings of this study would provide insights to the policymakers and researchers for designing and implementing community-based research initiatives, particularly in the area of public health and social science as well as market analysis research. The findings could also be helpful in the situation when the national representation of data is required with budget and time constraints.

Keywords: Bangladesh, demographic factors, homogeneity, socioeconomic factors, multi-variate cluster analysis

1 Introduction

For designing intervention studies, especially in the field of health and social science, it is essential to know the demographic and socioeconomic profile of the study population. In

fact, demographic and socioeconomic factors are considered as the primary determinants of health (WHO, 2009), social science (Hirschman and Tolnay, 2005; Hodgson, 1983) and market analysis related activities (Mazilu and Mitroi, 2010). For instance, variables such as age, gender, literacy status, occupation, and household socioeconomic characteristics are naturally associated with different health indicators. Moreover, to understand the community's future needs, resource allocation and other services, it is recommended to explore the demographic and socioeconomic variables as it is the first valuable intuitions about a community (French, 2014). Further, to improve human health, strong national health research systems are required which has a practical association with demography and socioeconomic condition of the people (Baiden *et al.*, 2006; Lansang and Dennis, 2004; Nchinda, 2002). In fact, understanding the homogeneity (or heterogeneity) of the population in terms of demographic and socioeconomic characteristics is the prerequisite for capacity building and generalization of study findings in the population.

Homogeneity within a country can be characterized by multivariate cluster analysis technique based on multiple characteristics of the population (Johnson and Wichern, 2007). Relying on similarity measures, it can classify the data into different meaningful groups or patterns in such a way that the elements within a segment are similar to each other than elements of other groups (Johnson and Wichern, 2007). This method is widely used in health research, social sciences and market analysis research to identify the natural patterns of the data as exact groups are frequently unknown to us (Fonseca, 2013; Glatman-Freedman *et al.*, 2016; Vidden *et al.*, 2016). For instance, identifying the clusters for health indicators of a country is crucial since it is involved with designing public health initiatives and evidence-based health policies.

Bangladesh, a small delta in South Asia, is divided into eight administrative divisions and which further divided into 64 districts. Moreover, for administrative purpose, each district is further divided into smaller administrative units such as thana/upazila or sub-district. The capital of a district is known as district town or sadar. The district town itself is a thana/upazila, but generally is the centre of all activities of a district. Since independence in 1971, the development of each of the districts and district towns are heterogeneous (BBS, 2012). There is a range of variations in terms of demographic and socioeconomic characteristics of people across the country (Etzold and Mallick, 2015). In particular, some districts such as Dhaka, Chittagong, Gazipur, Narayangonj have experienced a high level of development in terms of infrastructure and socioeconomic conditions (BBS, 2012). Similarly, few district towns such as Bogra, Comilla, Gazipur, Jessore, Narayangonj and Sylhet sadar upazilas are also experiencing a high level of development in the same period (BBS, 2012). As geographical disparity varies within the country, it would be very effective for the policymakers and researchers particularly, from social science and public health domains to identify the natural groups among districts and district towns of the country. To our best knowledge, there is no study conducted yet in Bangladesh to identify the homogeneous groups by considering the demographic and socioeconomic characteristics. The present study, therefore, aimed to identify the homogeneity or natural segments among districts and district towns of Bangladesh based on demographic and socioeconomic characteristics using multivariate cluster analysis technique.

2 Methods

2.1 Data sources

The district-specific and district-town-specific data used for this study were collected from the most recent Bangladesh Population and Housing Census 2011 (BBS, 2012). Data on the demographic status of the population and the socioeconomic condition of households in the country were collected in the census. The details can be found in the census report (BBS, 2012).

2.2 Variables

Demographic variables considered for this study were: age-specific distribution, male-female ratio, the percentage of employed people, overall literacy rate, population density, the percentage of the urban population, and annual growth rate. Socioeconomic variables considered for this study were: household size, the percentage of pacca households, percentage of households having access to electricity, percentage of households having access to tube well or tap water for drinking, the percentage of households having sanitary toilet facilities, the percentage of own housing tenancy, and number of institutional households.

The employed people were defined as people who were engaged with income generating activities. Literacy status assessment is made for population 7 years and over. The population density was calculated per square kilometre. Hospitals, clinics, jails, barracks, orphanages, hostels or halls of educational institutions were treated as institutional households. The details can be found elsewhere (BBS, 2012, 2015).

2.3 Method of analysis

To determine the homogeneity among districts and district towns, we used multivariate cluster analysis by considering Ward hierarchical clustering method with Euclidean distance (Johnson and Wichern, 2007). Although single or multiple factors can be used to determine clusters, we utilized multiple factors, and hence it was referred as multivariate cluster analysis. We selected the optimum number of clusters by applying different clustering algorithms such as hierarchical, average, kmeans with Euclidean distance and by considering the number of clusters ranging from 2 to 7. In these case, R package "NbClust" and "factoextra" were used (Charrad *et al.*, 2014; Kassambara and Mundt, 2016). The homogeneity among districts or district towns was shown by using dendrogram and maps produced by using ArcGIS (version 10.0).

3 Results

3.1 Homogeneity among districts

By applying different combinations of the number of clusters, distance measures and clustering methods, we found that the optimum number of clusters among districts were three (Figure S1). Therefore, district-specific homogeneity of Bangladesh in terms of demographic and socioeconomic variables was clustered into three groups. A dendrogram is shown in Fig-

ure S2 to determine this homogeneity. A map produced by using the results obtained from the dendrogram is given in Figure 1.

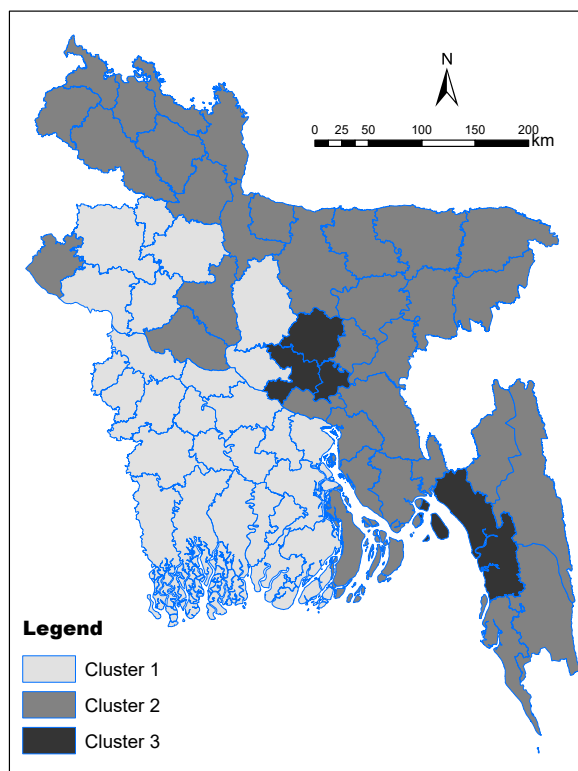


Figure 1: Homogeneous districts in Bangladesh based on demographic and socioeconomic characteristics

Predictably, districts such as Chittagong, Dhaka, Gazipur and Narayanganj were in the same cluster. Districts such as Bagerhat, Barguna, Barisal, Bogra, Chuadanga, Faridpur, Gopalganj, Jessore, Jhalokati, Jhenaidah, Joypurhat, Khulna, Kushtia, Madaripur, Magura, Manikganj, Meherpur, Naogaon, Narail, Natore, Patuakhali, Pirojpur, Rajbari, Rajshahi, Satkhira, Shariatpur and Tangail were grouped into another cluster, while other 33 districts were in another cluster to characterize homogeneous demographic and socioeconomic pattern. The cluster-specific summary statistics of the demographic and socioeconomic variables of the districts are shown in Table 1.

3.2 Homogeneity among district towns

We found three optimum number of clusters among district towns (Figure S1). Therefore, district-town-specific homogeneity of Bangladesh in terms of demographic and socioeconomic variables was clustered into three groups (Figure S3). A map produced by using the results obtained from the dendrogram is given in Figure 2.

Table 1: Cluster-specific summary statistics of the demographic and socioeconomic variables of the 64 districts in Bangladesh

Characteristics	Cluster 1 ^a		Cluster 2 ^b		Cluster 3 ^c	
	Mean	SD	Mean	SD	Mean	SD
Age (years)						
<5	9.5	0.8	11.6	1.1	9.2	0.8
5-9	11.8	1.2	13.9	1.2	10.6	1.2
10-14	11.5	1.0	12.1	0.8	10.4	1.3
15-19	8.3	0.5	8.8	1.0	10.2	0.6
20-24	8.5	0.7	8.5	0.5	12.1	1.3
25-29	9.0	0.8	8.7	0.6	11.7	1.5
30-49	25.7	2.0	22.8	2.0	25.0	1.6
50-59	7.0	0.3	6.1	0.6	5.4	0.2
60-64	3.1	0.2	2.7	0.3	2.2	0.2
≥65	5.6	0.5	4.7	0.6	3.2	0.4
Sex ratio	98.1	2.7	98.8	4.7	109.3	7.1
% of employed people						
Male	20.0	6.5	22.6	5.4	14.0	2.1
Female	1.2	0.4	2.3	2.1	4.8	1.5
Literacy rate	53.3	6.1	46.1	6.2	62.3	6.0
Population density	924.6	169.9	993.1	414.0	3965.8	3108.9
% of the urban population	16.3	6.6	16.5	5.5	45.7	21.6
Annual growth rate	0.7	0.5	1.5	0.5	3.3	1.6
Household size	4.2	0.2	4.6	0.5	4.4	0.4
% of pacca households	7.9	5.9	5.8	4.7	25.7	14.4
% of households having access to						
Electricity	50.2	9.6	45.2	17.2	88.8	8.7
Tube well or tap water for drinking	93.8	7.2	90.5	12.6	96.8	1.6
Sanitary toilet facilities	67.7	13.3	52.8	15.8	82.6	7.5
% of own housing tenancy	92.4	4.4	90.4	5.4	49.4	16.5
Number of institutional households	241.3	164.9	292.8	203.3	2625.5	2246.9
N (number of districts)	27		33		4	

^aBagerhat, Barguna, Barisal, Bogra, Chuadanga, Faridpur, Gopalganj, Jessore, Jhalokati, Jhenaidah, Joypurhat, Khulna, Kushtia, Madaripur, Magura, Manikganj, Meherpur, Naogaon, Narail, Natore, Patuakhali, Pirojpur, Rajbari, Rajshahi, Satkhira, Shariatpur, Tangail; ^bBandarban, Bhola, Brahmanbaria, Chandpur, Comilla, Cox's Bazar, Dinajpur, Feni, Gaibandha, Habiganj, Jamalpur, Khagrachhari, Kishoreganj, Kurigram, Lakshmipur, Lalmonirhat, Maulvibazar, Munshiganj, Mymensingh, Narsingdi, Nawabganj, Netrakona, Nilphamari, Noakhali, Pabna, Panchagarh, Rangamati, Rangpur, Sherpur, Sirajganj, Sunamganj, Sylhet, Thakurgaon; ^cChittagong, Dhaka, Gazipur, Narayanganj

Clearly, thana/upazila sadar of the districts such as Bandarban, Barisal, Bogra, Comilla, Dinajpur, Feni, Gazipur, Jessore, Khagrachhari, Khulna, Narayanganj, Rangamati, Rangpur and Sylhet grouped into a cluster. District town of Bhola, Brahmanbaria, Chandpur, Cox's Bazar, Gopalganj, Habiganj, Kishoreganj, Lakshmipur, Madaripur, Maulvibazar, Nawabganj, Netrakona, Noakhali, Panchagarh, Shariatpur and Sunamganj were grouped into another cluster, while other 31 district towns were in another cluster to characterize homogeneous demographic and socioeconomic pattern. The cluster-specific summary statistics of the demographic and socioeconomic variables of the district towns are shown in Table 2.

4 Discussion

We found district-specific heterogeneity and these districts were divided into three clusters based on demographic and socioeconomic characteristics. Percentage of employed people, literacy rate, population density, percentage of the urban population, annual growth rate, percentage of pacca households, percentage of households having access to electricity, percentage of households having access to tube well or tap water for drinking, percentage of households having sanitary toilet facilities, and number of institutional households- all were higher in Cluster 3 (Chittagong, Dhaka, Gazipur and Narayanganj districts). This is due to the high level of development in the port city Chittagong, the main city Dhaka and its two surrounding cities Gazipur and Narayanganj (BBS, 2012; Kabir and Parolin, 2012). It was also seen that the percentage of people aged 50 years or more was considerably lesser in these districts than other districts, despite the higher number of health facilities.

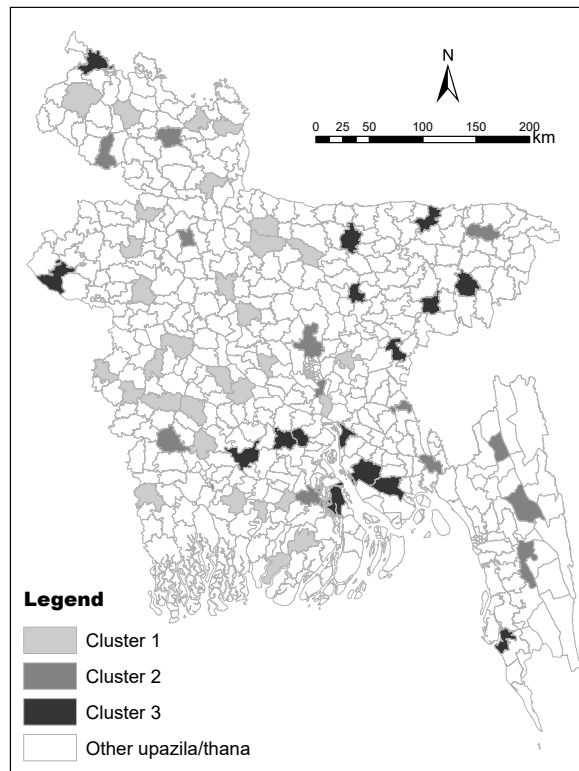


Figure 2: Homogeneous district towns in Bangladesh based on demographic and socioeconomic characteristics

Table 2: Cluster-specific summary statistics of the demographic and socioeconomic variables of the 61 district towns in Bangladesh (three districts have no specified thana/upazila sadar)

Characteristics	Cluster 1 ^a		Cluster 2 ^b		Cluster 3 ^c	
	Mean	SD	Mean	SD	Mean	SD
Age (years)						
<5	9.6	1.0	9.2	0.9	11.5	1.0
5-9	11.8	1.2	11.0	1.1	14.0	1.0
10-14	11.3	0.6	11.1	1.0	12.7	0.6
15-19	8.5	0.6	10.1	0.6	9.5	0.7
20-24	8.9	0.6	10.7	1.6	8.7	0.6
25-29	9.3	0.5	10.4	1.4	8.4	0.5
30-49	25.8	1.6	25.4	1.7	21.9	1.3
50-59	6.9	0.5	6.0	0.7	5.9	0.6
60-64	2.9	0.3	2.3	0.3	2.7	0.3
≥65	5.2	0.7	3.8	0.8	4.7	0.7
Sex ratio	99.6	2.8	107.7	5.2	98.5	4.7
% of employed people						
Male	20.3	6.0	13.6	3.6	18.8	4.1
Female	1.6	0.6	3.6	2.3	1.5	0.5
Literacy rate	54.1	7.3	63.5	5.8	50.7	5.2
Population density	1250.5	519.7	2679.0	3247.7	1291.6	505.3
% of the urban population	23.1	6.7	50.4	18.6	23.3	8.2
Annual growth rate	1.1	0.9	1.7	2.4	0.4	1.7
Household size	4.2	0.2	4.4	0.4	4.9	0.4
% of pacca households	9.6	6.1	21.6	11.4	9.3	5.9
% of households having access to						
Electricity	56.7	15.6	78.6	15.4	56.8	13.8
Tube well or tap water for drinking	97.2	1.1	92.3	11.0	94.9	2.7
Sanitary toilet facilities	61.9	17.9	73.0	13.5	66.5	14.1
% of own housing tenancy	88.3	5.3	61.0	19.6	87.5	5.6
Number of institutional households	80.5	51.8	232.3	161.4	67.3	35.2
N (number of district towns)	31		14		16	

^aBagerhat, Barguna, Chuadanga, Faridpur, Gaibandha, Jamalpur, Jhalokati, Jhenaidah, Joypurhat, Kurigram, Kushtia, Lalmonirhat, Magura, Manikganj, Meherpur, Munshiganj, Mymensingh, Naogaon, Narail, Narsingdi, Natore, Nilphamari, Pabna, Patuakhali, Pirojpur, Rajbari, Satkhira, Sherpur, Sirajganj, Tangail, Thakurgaon;

^bBandarban, Barisal, Bogra, Comilla, Dinajpur, Feni, Gazipur, Jessore, Khagrachhari, Khulna, Narayanganj, Rangamati, Rangpur, Sylhet; ^cBhola, Brahmanbaria, Chandpur, Cox's Bazar, Gopalganj, Habiganj, Kishoreganj, Lakshmipur, Madaripur, Maulvibazar, Nawabganj, Netrakona, Noakhali, Panchagarh, Shariatpur, Sunamganj

We also found district towns were heterogeneous in terms of demographic and socioeconomic characteristics and were grouped into three clusters. It is notable that districts such as Dhaka, Chittagong and Rajshahi have no specified district town. However, for the cluster 2 (thana/upazila sadar of Bandarban, Barisal, Bogra, Comilla, Dinajpur, Feni, Gazipur, Jessore, Khagrachhari, Khulna, Narayanganj, Rangamati, Rangpur, Sylhet), the literacy rate, population density, percentage of the urban population, annual growth rate, percentage of pacca households, percentage of households having access to electricity, percentage of households having access to tube well or tap water for drinking, percentage of households

having sanitary toilet facilities, and number of institutional households- all were noticeably higher than other two clusters. Moreover, people aged 50 years or more were lesser in cluster 2 than other clusters. However, it is believed that if a district is in a given cluster, the district town should be in that same cluster. We also found approximately similar findings, although some district towns scattered into a different segment. This might be due to inequality in the development across the district.

It is broadly acknowledged that multivariate cluster analysis is an effective analytical tool to carry out regional segmentation (Johnson and Wichern, 2007; Shubat *et al.*, 2007). It is also evident that regional segmentation, i.e., identifying homogeneous (or heterogeneous) groups is an effective and need-based approach for designing and implementing the community-based research. It helps the policymakers to focus on the right people and allocate the budget on the sub-national or further disaggregated level for adjusting better research plan. Moreover, as the demographic and socioeconomic conditions are the primary valuable understandings about a community, finding out the clusters that are similar (or dissimilar) in terms of demographic and socioeconomic characteristics would be a prodigious way to generalize and allocate proper resources. For instance, taking a few elements from each of the three segments can reduce the time and budget constraints without hampering the national representation of the data, particularly in public health research. Therefore, the findings of the present study would be an effective elucidation to the policymakers or research practitioners to design and implement community-based research initiatives especially public health and social science research.

Acknowledgments

The authors would like to thank and acknowledge the contribution of the Bangladesh Bureau of Statistics team for their efforts in providing open access to the data set. The authors would also like to thank the referees for their critical readings and most helpful comments which improve the paper significantly.

Declarations

Funding: There were no specific funding for this study.

Conflict of interest: The authors declare no conflict of interest.

Ethical approval: Not applicable.

References

- Baiden F, Hodgson A, Binka FN (2006). "Demographic surveillance sites and emerging challenges in international health." *Bulletin of the World Health Organization*, **84**(3), 163–163. doi:S0042-96862006000300003.
- BBS (2012). "National population and housing census 2011." URL <http://catalog.ihsn.org/index.php/catalog/4376>.

- BBS (2015). "Changing Patterns of Urbanization in Bangladesh: An Analysis of Census Data." URL <http://203.112.218.65:8008/WebTestApplication/userfiles/Image/PopMonographs/Volume-12.UM.pdf>.
- Charrad M, Ghazzali N, Boiteau V, Niknafs A, Charrad MM (2014). "NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set." *Journal of Statistical Software*, **61**(6), 1–36.
- Etzold B, Mallick B (2015). "Bangladesh at a Glance." URL https://www.researchgate.net/publication/285371220_Bangladesh_Country_Profile.
- Fonseca JR (2013). "Clustering in the field of social sciences: that is your choice." *International Journal of Social Research Methodology*, **16**(5), 403–428. doi:10.1080/13645579.2012.716973.
- French C (2014). "Why Demographic Data Matters." URL https://extension.unh.edu/resources/files/Resource004765_Rep6784.pdf.
- Glatman-Freedman A, Kaufman Z, Kopel E, Bassal R, Taran D, Valinsky L, Agmon V, Shpriz M, Cohen D, Anis E, Shohat T (2016). "Near real-time space-time cluster analysis for detection of enteric disease outbreaks in a community setting." *Journal of Infection*, **73**(2), 99–106. doi:10.1016/j.jinf.2016.04.038.
- Hirschman C, Tolnay SE (2005). *Population and the Social Sciences: Social Demography*. Springer, Switzerland AG.
- Hodgson D (1983). "Demography as Social Science and Policy Science." *Population and Development Review*, **9**(1), 1–34. doi:10.2307/1972893.
- Johnson RA, Wichern DW (2007). *Applied Multivariate Statistical Analysis*. 6 edition. Pearson Prentice Hall, Upper Saddle River, New Jersey.
- Kabir A, Parolin B (2012). "Planning and Development of Dhaka - A Story of 400 Years." *15th International Planning History Society Conference*, p. 20. doi:10.4103/1319-6308.180154.
- Kassambara A, Mundt F (2016). "Factoextra: extract and visualize the results of multivariate data analyses." *R package version*, **1**(3). URL <https://cran.r-project.org/web/packages/factoextra/factoextra.pdf>.
- Lansang MA, Dennis R (2004). "Building capacity in health research in the developing world." *Bulletin of the World Health Organization*, **82**(10), 764–770.
- Mazilu M, Mitroi S (2010). "Demographic, Social, Economic and Geographic Features-Shaping Factors of the Tourist Market." *Romanian Economic and Business Review*, **5**(1), 159.
- Nchinda TC (2002). "Research capacity strengthening in the South." *Social science & medicine*, **54**(11), 1699–1711. doi:10.1016/S0277-9536(01)00338-0.
- Shubat O, Bagirova A, Ivlev A (2007). "The use of cluster analysis for demographic policy development: evidence from Russia." *30th European Conference on Modelling and Simulation*, pp. 159–165. doi:10.7148/2016-0159.

Vidden C, Vriens M, Chen S (2016). “Comparing clustering methods for market segmentation: A simulation study.” *Applied Marketing Analytics*, 2(3), 225–238.

WHO (2009). “Demographic and socioeconomic statistics.” URL <https://www.who.int/whosis/whostat/EN-WHS09-Table9.pdf>.

Supplementary Material

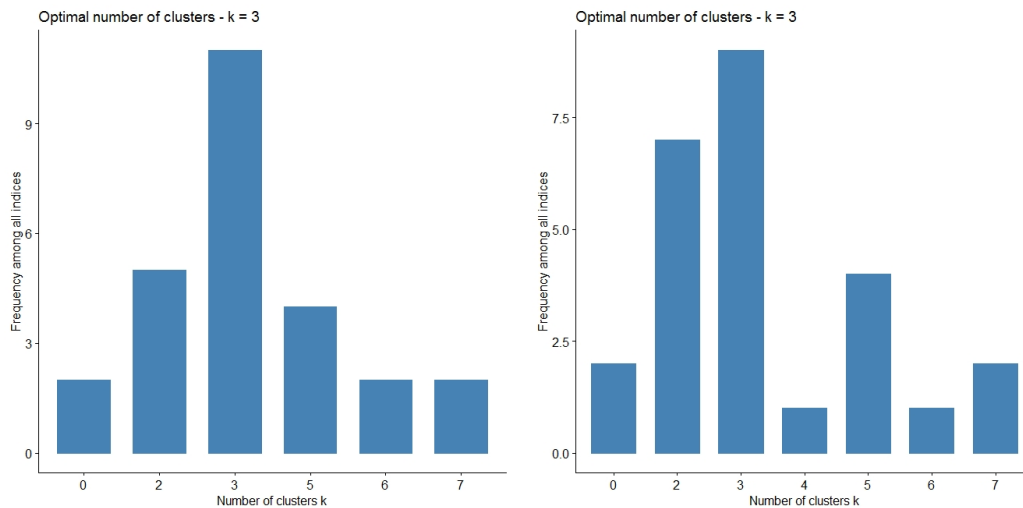


Figure S1: Determination of optimum number of clusters of districts (left panel) and district towns (right panel)

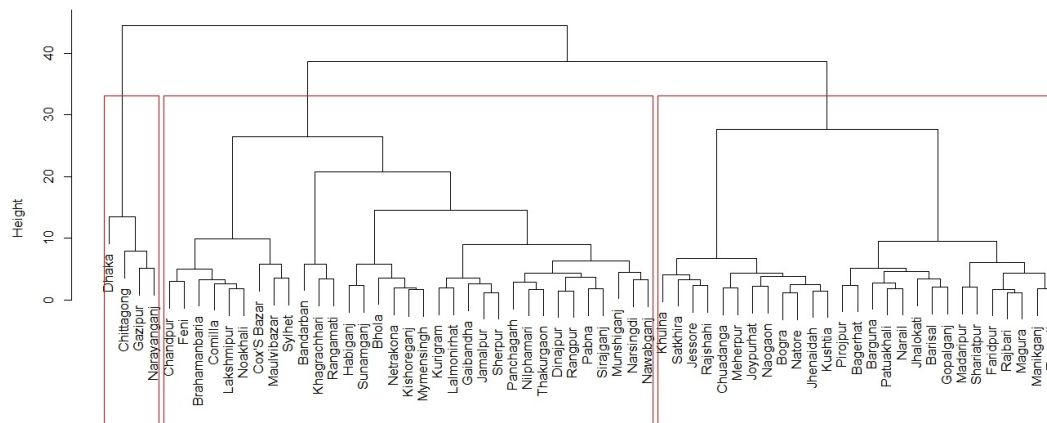


Figure S2: Three distinct clusters among 64 districts in Bangladesh based on demographic and socioeconomic variables

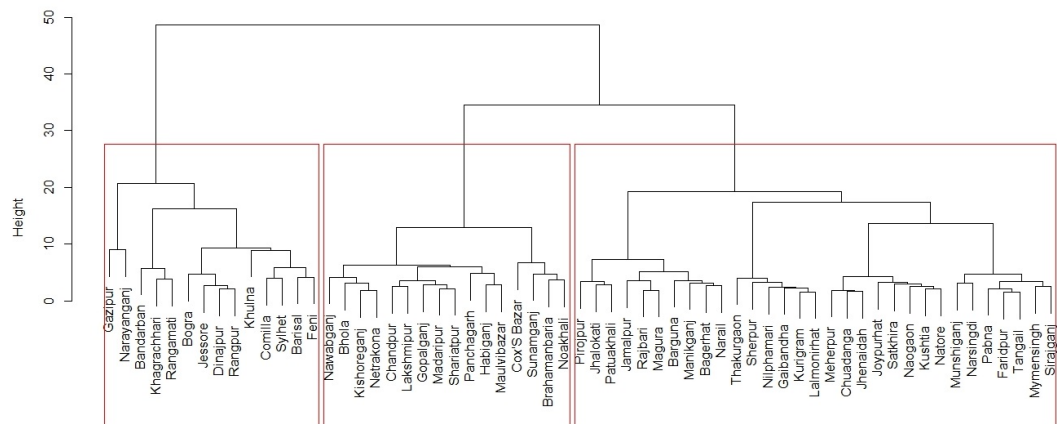


Figure S3: Three distinct clusters among 61 district towns in Bangladesh based on demographic and socioeconomic variables (Three districts Dhaka, Chittagong and Rajshahi have no specified district town)