

# Obtaining Urban Sampling Proportion

To generate urban sampling weights, we need information of urban fraction at Admin-1 level. Such information could usually be obtained with credibility from the country-specific DHS survey reports or national census reports. We need to organize this information in a cleaned table to run our scripts.

Because these sampling weights refer to a specific sampling frame, using a stratified model requires us to only use surveys which use the same sampling frame. Thus, going forward, *we will only use the surveys that have the same sampling frame as our most recent survey*. (The removal of data that comes from surveys with a different frame is already automated in subsequent scripts). To determine which surveys this includes, you can compare `frame_years` with `survey_years` from the `info` file.

## 1. Search for the information about Admin-1 level urban population

It is rare that all the DHS survey reports and national census reports are formatted in the same way, which means that this information might be found at different places in different reports. We provide some general ways to find them and the options are ordered by their priority. We illustrate this process for Zambia.

- a. We recommend the reader to first check out each of the DHS reports for the surveys that will be used (those that use the same sampling frame) until you find one where the admin-1 level urban population fraction is available. You can find these reports at <https://dhsprogram.com/pubs>. Ideally, this information will appear in Appendix A in at least 1 of the DHS reports. For Zambia, we found the information in Table A.1 of the 2013-2014 DHS survey.

**Table A.1 Population distribution by province and by residence from the 2010 Census of Population and Housing, Zambia 2013-14**

Province	Urban	Rural	Total	Percent urban	Percent province
Central	328,537	978,574	1,307,111	25.13	9.99
Copperbelt	1,596,374	376,789	1,973,163	80.90	15.07
Eastern	199,479	1,393,185	1,592,664	12.52	12.17
Luapula	194,744	797,463	992,207	19.63	7.58
Lusaka	1,842,076	336,318	2,178,394	84.56	16.64
Muchinga	123,393	596,469	719,862	17.14	5.50
Northwestern	157,902	569,142	727,044	21.72	5.55
Northern	201,873	903,951	1,105,824	18.26	8.45
Southern	389,215	1,200,761	1,589,976	24.48	12.15
Western	133,090	770,093	903,183	14.74	6.90
Zambia	5,166,683	7,922,745	13,089,428	39.47	100.00

- b. Sometimes, however, the admin-1 level urban population fraction is not available, and we only have the household urban fraction. If this is the case, you can check the census from the frame year of the surveys being used (again you can find this year by checking `frame_years`). The frame year for the Zambia surveys is 2010 so you can google ‘Zambia 2010 census’ to find the census report. In this case, we found the following table.

Province	Total			Rural			Urban		
	Total	Male	Female	Total	Male	Female	Total	Male	Female
<b>Zambia</b>	<b>13,092,666</b>	<b>6,454,647</b>	<b>6,638,019</b>	<b>7,919,216</b>	<b>3,906,636</b>	<b>4,012,580</b>	<b>5,173,450</b>	<b>2,548,011</b>	<b>2,625,439</b>
Central	1,307,111	648,465	658,646	978,574	487,713	490,861	328,537	160,752	167,785
Copperbelt	1,972,317	981,887	990,430	376,861	190,178	186,683	1,595,456	791,709	803,747
Eastern	1,592,661	784,680	807,981	1,392,338	686,577	705,761	200,323	98,103	102,220
Luapula	991,927	488,589	503,338	797,407	393,615	403,792	194,520	94,974	99,546
Lusaka	2,191,225	1,082,998	1,108,227	336,318	169,604	166,714	1,854,907	913,394	941,513
Muchinga	711,657	349,872	361,785	590,575	290,490	300,085	121,082	59,382	61,700
Northern	1,105,824	546,851	558,973	903,208	447,755	455,453	202,616	99,096	103,520
North Western	727,044	358,141	368,903	563,061	277,503	285,558	163,983	80,638	83,345
Southern	1,589,926	779,659	810,267	1,197,751	587,448	610,303	392,175	192,211	199,964
Western	902,974	433,505	469,469	783,123	375,753	407,370	119,851	57,752	62,099

**Source: 2010 Census of Population and Housing**

c. It is possible that the exact numbers or fractions of Admin-1 level urban population cannot be found in either DHS or census report. In this case, we can weight the Admin-1 level urban and rural households by the urban and rural averaged household sizes. This approach is not ideal because averaged household sizes are usually only available at national level and rounding will further compromise the accuracy. Generally we can find this information in Appendix A of the DHS report (any of the applicable DHS reports will suffice). *Note that if this approach is used the total number of urban and rural households must be multiplied by the average number of households in urban and rural areas, respectively.* This information can also be found in the DHS report.

## 2. Organize the information into a clean table

- Open up `ur_prop.R` in from the Rcode folder and run the code up to line 35, making sure to enter in the country of interest.
- Copy the information obtained in Step 1 into an excel file called '(country abbreviation)\_frame\_urb\_prop.xlsx' and save in the newly created 'UR' subfolder within the Results folder. The first column of the excel file should be the name of the Admin-1 area and the second column should be the urban fraction (urban population/total population). Recall that if you are using number of households, you must multiply average household size. The excel file should look like the following:

	A	B	C
1	Admin1	frac	
2	Central	0.277	
3	Copperbelt	0.789	
4	Eastern	0.138	
5	Luapula	0.181	
6	Lusaka	0.821	
7	Muchinga	0.172	
8	Northern	0.184	
9	North Weste	0.222	
10	Southern	0.278	
11	Western	0.143	
12			
13			

- Run the rest of `ur_prop.R` and make sure that the Admin-1 names in the final object, `ref.tab`, match up. In the end, the `ref.tab` object should look like the following:

GADM	Internal	matched_name	urb_frac
Barisal	admin1_1	Barisal	0.1252493
Chittagong	admin1_2	Chittagong	0.2040611
Dhaka	admin1_3	Dhaka	0.2605315
Khulna	admin1_4	Khulna	0.1581063
Rajshahi	admin1_5	Rajshahi	0.1603063
Rangpur	admin1_6	Rangpur	0.1190546
Sylhet	admin1_7	Sylhet	0.1202198