Differences between urban housholds and rural households in Vietnam in various aspects[*]

Jiayi Du   Siyi Liu   Cheng chen

01 April 2022

Abstract

In this paper, we conducted the research on the various aspects of households in urban and rural areas, which may have an impact on fertility index about life based on the 1997 Viemam Demographic and Health Survey (VNDHS-II). We mainly analyzed the possible influence by comparison between urban households and rural households. After screening and comparison, We found that in all aspects, urban housing has obvious advantages over rural housing in terms of facilities, and is more conducive to people's health and thus beneficial to the improvement of fertility indicators. Moreover, we analyzed the reason for these factors causing effect and discussed the possible bias in the research and the solutions to it.

1 Introduction

Different aspects of the Vietnam society has entered a relatively stable stage in the past few decades, such as its economy and politics. However, it doesn't mean that the Vietnamese haven't been experiencing changes that drastically affect their conception. Factors including social policies and technology advancements have greatly shift the Vietnamese's ideology, values, and one of the most common change is their attitude towards fertility.

The 1997 Viemam Demographic and Health Survey (VNDHS-II) conducted a survey on the topic of fertility in Vietnamese Families in 1997 (Ha Noi 1997). I chose the table 2.7 and stimulated the data according to the table. It showed the housing characteristic of each respondent of the survey. By utilizing the data, we can find that urban dwellers enjoy better living conditions in all respects than those in rural areas, which is consistent with many people's perception about households. In order to make the difference more intuitive, this paper aims to explore the different possible aspects in which urban households are different from rural ones, such as electricity, source of drinking water, time to water source, sanitation facility, flooring and persons per sleeping room. The survey data is thoroughly discussed in the section of 2. In the sub-section 2.1, we presented an overview of the original survey data, and explained our cleaned dataset that we'll for exploration. The methods used to collect the VNDHS-II data as well as the strengths and weaknesses associated with these methods are outlined in the sub-sections 2.2, 2.3, and 2.4. Section 3 presents a series of findings in the difference between the urban households and rural ones.

2 Data

2.1 Dataset of interest

The survey we utilized in this paper comes from the 1997 Viemam Demographic and Health Survey (VNDHS-II) (Ha Noi 1997). The 1997 Vietnam Demographic and Health Survey (VNDHS-II) was carried out according to a subcontract between the National Committee for Population and Family Planning (NCPFP) and the Futures Group International Inc. in USA within the frame of the Population and Family Health Project. The Macro International Inc. in USA provided technical assistance to the survey. This paper selected the table

---

which focused on the topic of households characteristic of families. This questionnaire has a total of 7001 samples and 7 variables. Its target respondents are ever-married women age 15-49 selected from 205 sampling clusters throughout Vietnam. (Ha Noi 1997).

This paper focused on investigating 5 of these variables: Characteristic, Electricity, Sourceofdrinkingwater, Flooring and Sanitation facility. R (R Core Team 2021), and R packages "tidyverse" (Wickham et al. 2019), "janitor" (Firke 2021), "knitr" (Xie 2021), "dplyr" (Wickham et al. 2021), and "kableExtra" (Zhu 2021) are utilized to create an extract of the cleaned dataset (Table 1).

Table 1: Extracting the first ten rows from the cleaned VNDHS-II survey dataset

ID	Characteristic	Electricity	Sourceofdrinkingwater	Timetowatersource	Sanitationfacility	Flooring	Personspersleepingroom
1	Urban	Yes	Well in residence	Less than 15 minutes	No facility bush	Finished floor	3-4
2	Urban	Yes	Piped into residence	Less than 15 minutes	Own flush toilet	Finished floor	1-2
3	Urban	Yes	Piped into residence	Less than 15 minutes	Own flush toilet	Earth sand	3-4
4	Urban	Yes	Rain water	Less than 15 minutes	Own flush toilet	Finished floor	1-2
5	Urban	Yes	Piped into residence	Less than 15 minutes	Own flush toilet	Finished floor	3-4
6	Urban	Yes	Piped into residence	Less than 15 minutes	Own flush toilet	Finished floor	1-2
7	Urban	Yes	Well in residence	Less than 15 minutes	Own flush toilet	Finished floor	1-2
8	Urban	Yes	Piped into residence	Less than 15 minutes	Own flush toilet	Finished floor	3-4
10	Urban	Yes	Well in residence	Less than 15 minutes	Vent. imp. pit toilet	Finished floor	5-6
11	Urban	Yes	Piped into residence	Less than 15 minutes	No facility bush	Finished floor	1-2

Table 1 shows the first ten rows of the VNDHS-II dataset of interest .Variable "Characteristic" indicates if the household is in the urban area or rural area. Variable "Electricity" indicates if the household has electricity. Variable "Source of drinking water" indicates the source of drinking water for residents. Variable "Time to water source" gives information on time for residents to obtain the water source. Variable "Sanitation facility" shows the type of toilet in the household. Variable "Flooring" represents the type of floor in the household. Variable "Persons per sleeping room" provides the number of persons per sleeping room.

2.2 Methodology

The sample for the VNDHS-II is a subsample of the 1996 Multi-Round Demographic Survey (1996 MRS). The latter is a semi-annual survey of about 243,000 households that the General Statistical Office (GSO) regularly undertakes to collect information on population changes in households. The sample households are located in 1,590 sample areas, or enumeration areas (EAs), spread throughout the 53 provinces of Vietnam, with 30 EAs in each province. On average, an EA comprises about 150 households.

For the VNDHS-II, a subsample of 205 EAs was selected and 26 households from each urban EA and 39 households from each rural EA were selected. The sample households were selected from a listing of households that was conducted as part of the 1996 MRS. A total of 7,150 households were selected.

The VNDHS-II was designed to provide separate estimates for the whole cotmtry, for urban and rural areas, for 18 project provinces, and for the remaining non-project provinces.

In the VNDHS-II, information was collected for usual residents of the selected households and visitors who had spent the previous night in the households. A household was defined as a person living alone or a group of persons who live and eat together.

2.3 Strengths

From a total 7,150 households selected in the sample, 7,031 households were occupied at the time of the interview, and 7,001 were successfully interviewed, giving a household response rate of 99.6 percent. The household response rate was high in both urban (99.2 percent) and rural (99.7 percent) areas. It is a pretty high number. To make sure that the samples were representative of the target population, the survey data were applied by a weighing factor so that the outcome of the survey was more accurate. Confidentiality rules prevented the respondents' personal information to be disclosed without permission (Ha Noi 1997).

2.4 Weaknesses

Some questions which appeared to be quite sensitive to many respondents resulted in some non-response situations. For instance, respondents were generally unwilling to provide detailed information about their households. Because the situation of the households can often reflect a family's economic level, many people do not want their economic level to be exposed.

Some bias were associated with the VNDHS data. For instance, the variable of Time to water source is less than 15 minutes for most people, so the research on this variable is of little significance. It will be better if it is divided into less than 10 minutes or more than 10 minutes. And it potentially increased the bias of the survey data. Non-response bias was resulted from the group of sampled populations who did not respond to the survey (Ha Noi 1997).
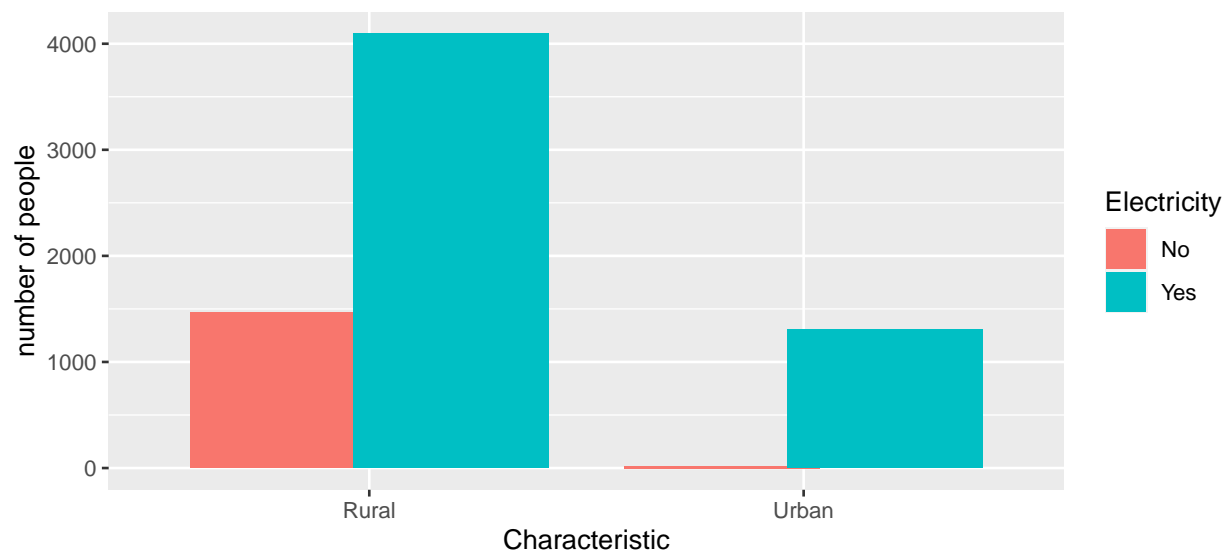
3 Results

3.1 Electrcity

Figure 1: Comparing the proportion of households with electricity in urban and rural areas

Figure 1 indicates the different proportion of households with electricity. It's obvious that the proportion of households with no electricity in rural area is much larger than that in urban area. In addition, the proportion of households with no electricity in urban area is quite close to 0.

3.2 Source of drinking water

Figure 2 demonstrates the different kinds of source of drinking water for respondents in their households. It's not hard to find that most of the households have piped water in urban area and well water in rural water. In addition, there are also lots of people who have to drink water from the spring and rain water in rural area.
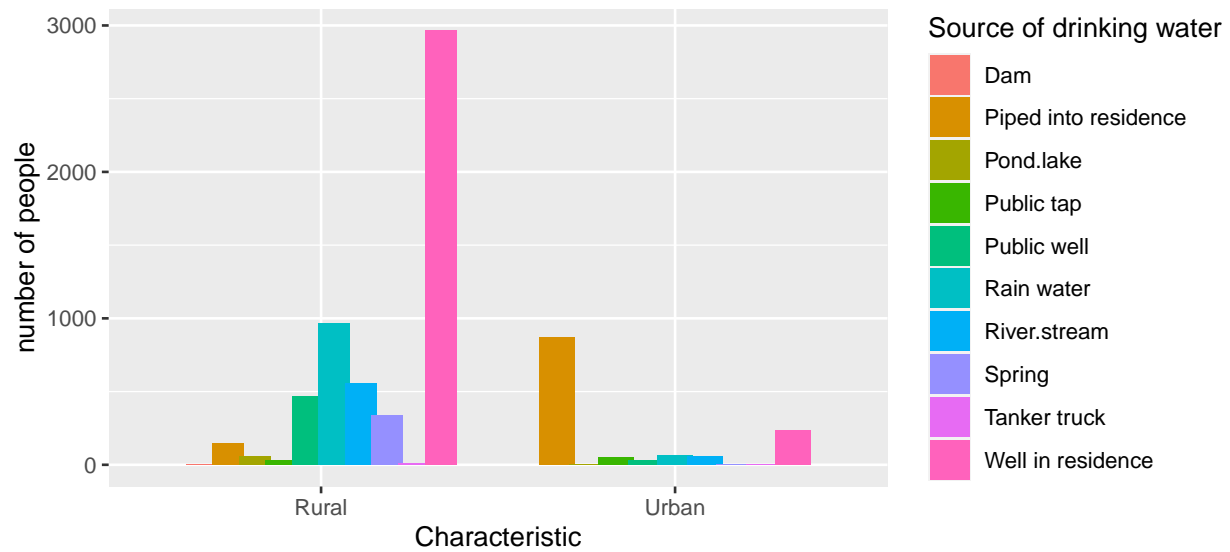
Figure 2: Comparing the different source of drinking water of households in urban and rural areas
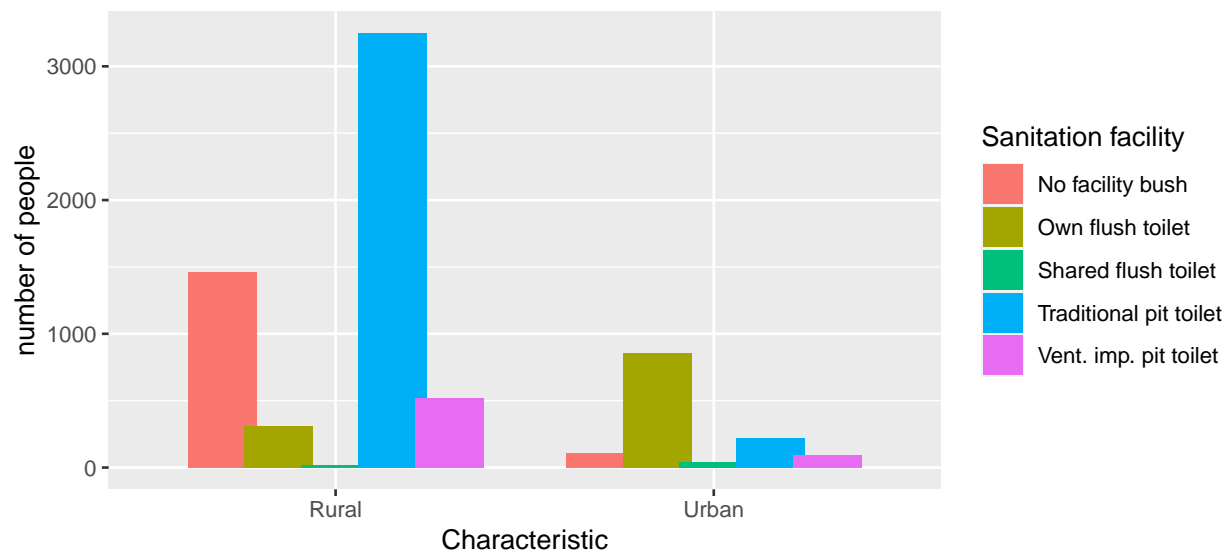
3.3 Sanitation facility

Figure 3 shows that flush toilet is the most common in urban areas while pit toilet is the most common in rural areas. However, not only in rural area, but also in urban area, there are still many people whose house has no facility or only bush. And the proportion of this in rural area is not low.
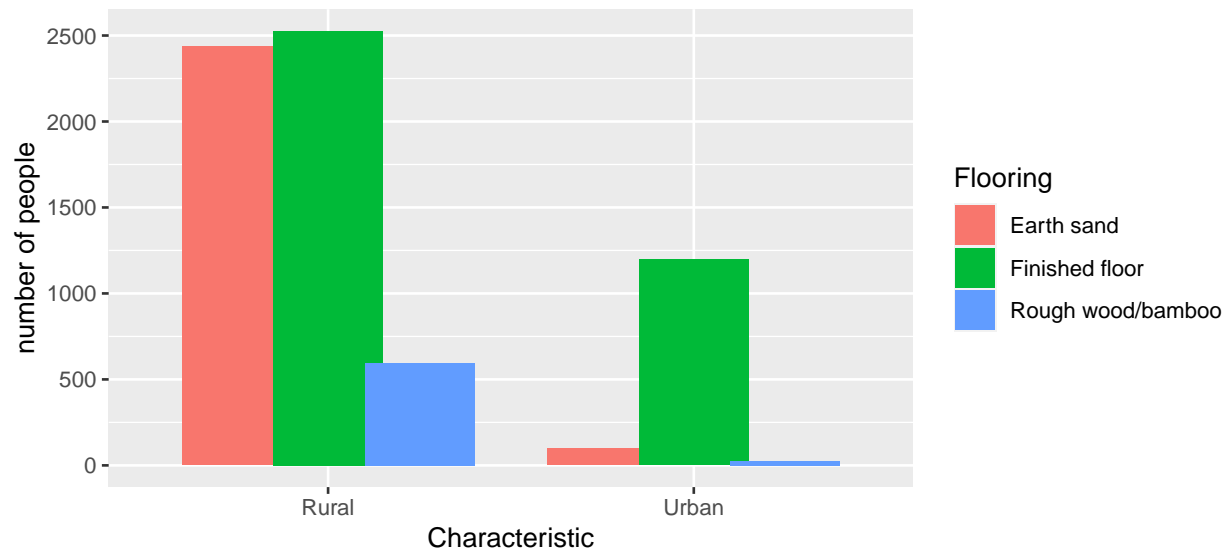
3.4   Flooring

Figure 4 shows that although the proportion of finished floor is the highest in both urban and rural areas, the ratio of earth sand and rough wood/bamboo are quite high in rural area.

4   Discussion

4.1   Electrcity

It is not difficult to find from the results that most urban households have access to electricity, but a large part of households in rural areas have not yet access to electricity. So the development level of Vietnam's rural areas was still lagging behind at that time, which had a great impact on people's fertility rate and the survival rate of newborns.

4.2   Source of drinking water

Most Vietnamese were not short of water at that time, but many families in rural areas had to rely on well water, and even a large number of them had to rely on rain water. The water is often dirty with impurities and harmful to human health, which naturally leads to a low survival rate for newborns.

4.3   Sanitation facility

The majority of urban households have flush toilet, while the majority of households in rural area have pit toilet, which is reasonable considering the circumstances. However it's quite shocking that 26 percent of households in rural area don't have toilet. When there is no toilet at home, it is necessary to go outside to go to the toilet, which is dangerous and uncertain for both pregnant women and newborns, and it will also have an impact on fertility indicators.

4.4   Flooring

Although most urban families have finished floors, a large part of urban and rural families have earth or sand in their homes, which is easy for all kinds of insects to enter the house, thus exerting a great influence on the living environment and health.

4.5   Weakness and limitations

Our results are more directly reflected by images, and then we make subjective summaries and think about the reasons behind them. We do not specifically point out the influence of different aspects of households that affect fertility index. With more study, we can use more statistical knowledge to conduct modeling and analysis to obtain the specific influence of different factors, so as to exclude the deviation caused by some subjective factors.

5   Appendix

5.1   Datasheet

5.1.1   Motivation

The dataset was created to provide a descriptive summary of some characteristics of the households and the individual respondents of the second Vietnam Demographic and Health Survey (VNDHS-II). It examines some environmental conditions, such as water sources, toilet facilities and so on. I created the dataset according to the table in the DHS program Demographic and Health Survey 1997 in Vietnam. The program is available at https://dhsprogram.com/publications/publication-fr104-dhs-final-reports.cfm?cssearch=454695_1 .

The VNDHS-II was designed to provide information on levels of fertility, family planning knowledge and use, infant and child mortality, and indicators of maternal and child health. And this dataset includes information which is very important for interpreting the survey findings and can provide an approximate indication of the representativeness of the survey.

5.1.2   Composition

The variables in the dataset are as below:

Characteristic: If the household is in the urban area or rural area.

Electricity: If the household has electricity

Sourceofdrinkingwater: The source of drinking water for residents(Piped into residence, Public tap, Well in residence, Public well, Spring, River. stream, Pond. lake, Dam, Rain water and Tanker truck)

Timetowatersource: Time for residents to obtain the water source(Less than 15 minutes and More than 15 minutes)

Sanitationfacility: the type of toilet in the household(Own flush toilet, Shared flush toilet, Traditional pit toilet, Vent. imp. pit toilet,No facility bush)

Flooring: the type of floor in the household(Earth sand, Rough wood/bamboo, Finished floor)

Personspersleepingroom: the number of persons per sleeping room(1-2, 3-4, 5-6, 7+, Dont know/missing)

There are 7 variables in total and 7001 observations. And the dataset contains all possible instances.

In the VNDHS-II, information was collected for usual residents of the selected households and visitors who had spent the previous night in the households. A household was defined as a person living alone or a group of persons who live and eat together.

However, there exist some missing values in the dataset. Some of the Personspersleepingroom of respondents are "Don't know/missing".

In addition, the dataset doesn't contain data that, might be offensive, insulting, threatening, or might otherwise cause anxiety. The dataset also doesn't contain data that might be considered confidential. Moreover, there is no errors, sources of noise, or redundancies in the dataset.

Reference

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://github.com/sfirke/janitor.

Ha Noi, Vietnam. 1997. *VietNam Demographic and Health Survey.* http://dhsprogram.com/pubs/pdf/FR104/FR104.pdf.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation.* https://dplyr.tidyverse.org,%20https://github.com/tidyverse/dplyr.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in R.* https://yihui.org/knitr/.

Zhu, Hao. 2021. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* http://haozhu233.github.io/kableExtra/,%0Ahttps://github.com/haozhu233/kableExtra.