

Household characteristics in urban and rural area in Ghana*

Xuetong Tang, Yilin Lan, Lida Liu

04/10/2022

Abstract

The Worldwide Demographic and Health Survey Program would like to know the economic and environmental household characteristics in Ghana in 1993. Therefore, with the support of U.S. Agency and the Ghana government, the Ghana Statistical Service has surveyed the specific economic and environmental household characteristics about the distribution of the source of water in regions, the sanitation facilities conditions, the use of flooring materials in different regions and the population densities in different region. In a reproducible way, we obtained a table that containing the results of the survey, and we analyze the specific conditions of the household characteristics in the past.

Contents

1	Introduction	2
2	Data	3
2.1	Methodology	3
2.2	Data cleaning	3
2.3	Data visualization	4
3	Result	7
3.1	Linear relationship between urban and rural area	7
3.2	Linear relationship between urban and rural area	8
4	Discussion	8
4.1	Overview and result	8
4.2	Future development	9
4.3	Limitations	9
	Reference	10

*Code and data are available at: [LINK](#).

1 Introduction

According to the report of the findings of the 1993 Ghana Demographic and Health Survey conducted by Ghana Statistical Service, the economic and environmental characteristics of households are illustrated in one table. To be more specific, the table reveals the distribution of the source of water in regions, the sanitation facilities conditions, the use of flooring materials in different regions and the population densities in different region (Daasebre 1993).

From the table that the Ghana Statistical Service has created, about 35 percent of the households used piped water as their main source of water, while approximately 31 percent of the households used well water as their main source of water. Moreover, there were about 27 percent of the households getting water from nature, like springs, river, stream, and rainwater. From another point of view, Upper East region and Upper West region's main major source of water were depending on well water, while the piped water was being used by the Greater Accra and Central regions (Daasebre 1993).

For the sanitation facilities, there were only 6 percent of the households that had flush toilets, and the majority of households were using latrines. In urban areas, there were about 8 percent of the households that do not have toilet facilities, while there were about 31 percent of the household that do not have toilet facilities. Moreover, in the three northern regions, about 75 to 89 percent of people had no access to private or public toilet facilities at all (Daasebre 1993).

For the living part, cement is the main material for flooring in house units, and the second most important flooring material is vinyl and linoleum in urban area, while in rural area the second important area is the earth and dung. To be more specific, only region of Upper West had over 50 percent of flooring that is natural. For the population density, there are 2.5 persons per sleeping room in Ghana no matter they were from rural or urban areas. Region of Upper East has the lowest population density and Brong-Ahafo has the highest population density (Daasebre 1993).

The paper would be consisted of mainly three parts, which are data part, result part and the discussion part. In the data part, the analysis of the data about some specific economic and environmental characteristics of Ghana in 1993 would be conducted, and the results of the analysis would be shown in the result part. Ultimately, the discussion about the phenomenon we analyzed before is going to be revealed in the discussion part at the end of the paper.

All the data analysis in this paper uses R studio (R Core Team 2021) with tidyverse (Wickham et al. 2019), janitor (Firke 2021), ggplot2 (Wickham 2016), kableExtra (Zhu 2021), purrr (Henry and Wickham 2020) and pdftools (Ooms 2022) packages.

2 Data

2.1 Methodology

The economy of Ghana is mixed and contained lot of small, unregulated business between 1975 to 1982. The GDP is growing at a rate of -0.5% due to poor management of fiscal and monetary policy. To improve the economy the government started the Economy Recover Program (ERP). The data we choose collect the household characteristics in Ghana in 1993. The 1993 GDHS is a stratified, representative sample that chose from 400 Enumeration Area. Then, the frame was divided into three ecological zones, including urban and rural areas. The EA was selected with probability proportional to the number of households. The survey was design to collect data from 5400 females aged 15-49 and 5400 males aged 15-49. The way that they collect data is to use questionnaire. The questionnaire was first developed in English and then translated into five local languages. People were required to provide information about the education level, household information, health etc. The questionnaires were filled and sent to Head office. Then, the data was edit by the data entry operator. At last, three professional data analyzation staff are performing second editing for the data and the final table was generate at 1994, March. For the response rate, 6161 household were selected and 5919 household were able to complete the interview. For the rest of the household, the staff had visited their house for several times but they were not at home.

2.2 Data cleaning

For the data cleaning part, we first read the table from the PDF to RStudio. Then, we change the variable names and remove some unnecessary spaces between the words. Then, we turn the table into a dataset. For the dataset, we compare it with the table on the PDF and fixed some small problem in the data. For example, we fixed “6,5” into “6.5.” At last, we only select the row and column we needed for the data analyzation. All the variable counts for total was removed.

2.3 Data visualization

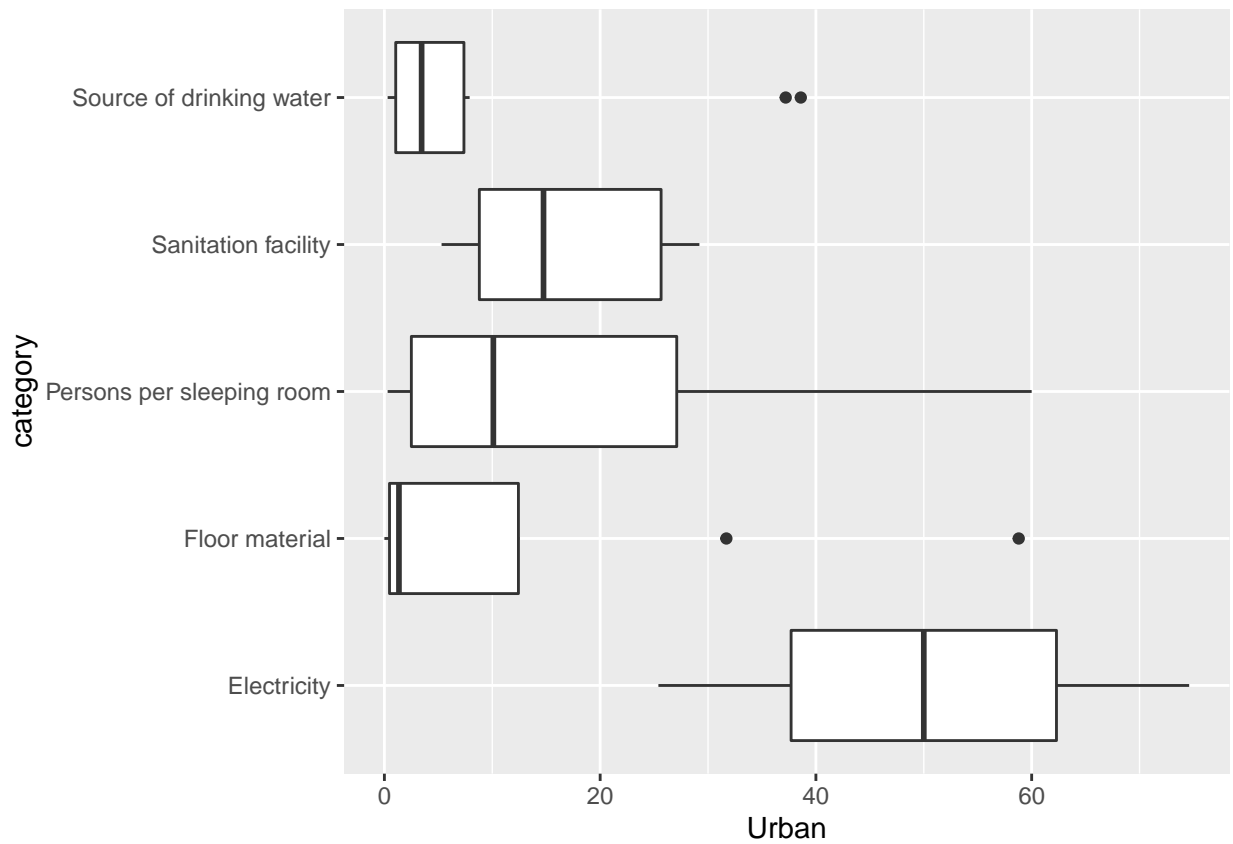


Figure 1: Boxplots showing the median difference between percentage of rural residence and different categorical variables

The boxplot above 1 aims to analysis the median of the percentage in each catogory in our table of interest. First, from the plot we find that the median of source of drinking water is very low with a relative small variance, which represents that the water source of each urban households is very different. There's no evident preference or usage of one kind of water source.

In contrast, from the boxplot for the category, persons per sleeping room, through it also has a low median, its variance is much larger. This shows that the majority of urban residence may usually have a certain range of number of persons per sleeping room.

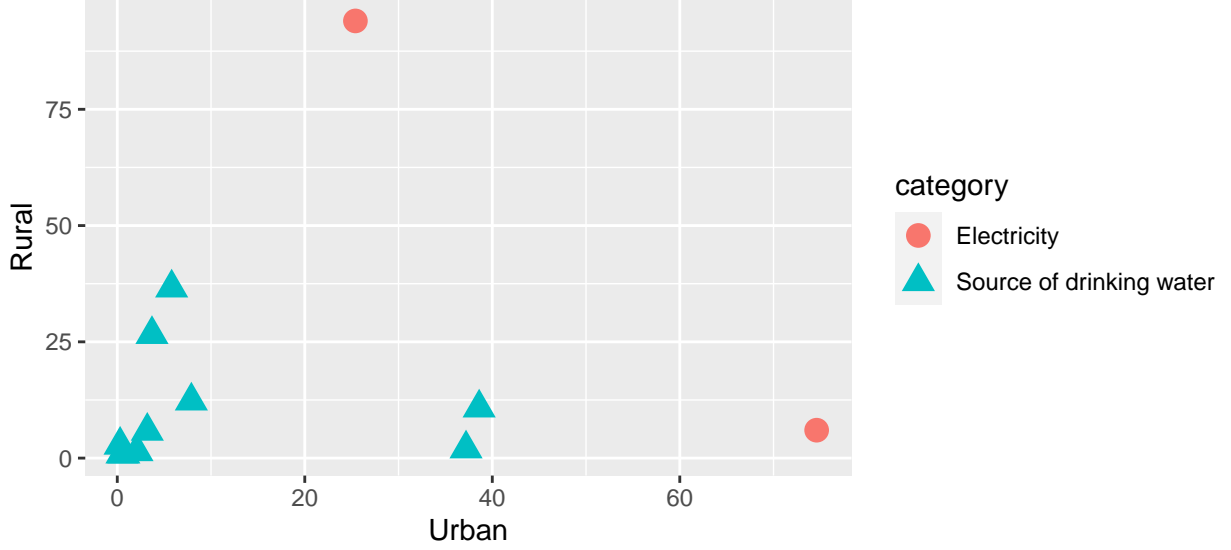


Figure 2: Scatterplots showing the distribution between urban and rural area with different categorical variables

Table 1: The data summary in electricity and source of drinking water in urban and rural area

category	Characteristics	Urban	Rural
Electricity	Yes	74.6	6.0
Electricity	No	25.4	94.0
Source of drinking water	Piped-into-residence	37.2	2.0
Source of drinking water	Public-tap/Neighbour	38.6	10.8
Source of drinking water	Well-in-residence	2.1	1.4
Source of drinking water	Public-well	7.9	12.3
Source of drinking water	Borehole	3.7	26.6
Source of drinking water	Spring	0.5	0.9
Source of drinking water	River/Stream	5.8	36.6
Source of drinking water	Rainwater	0.7	0.9
Source of drinking water	Other	3.2	5.8
Source of drinking water	Missing/Don't-know	0.3	2.8

The scatterplot 2 and table 1 above discusses the difference in residents' source of drinking water and electricity in urban and rural area. From the plot, we find that there's much difference in use of electricity in urban and rural area. The percentage of having electricity in urban area is 74.6% while in rural area is only 6%. This significant difference shows that the infrastructure of utility in rural area is very insufficient.

What's more, we also can see much information regarding to the source of water of residents in rural and urban area. From the table, we see that there's only 2% residents in rural area who have water directly piped into residence. The majority, 36% residents in rural area still use river and stream water while 26% residents in rural area use water from borehole. On the other hand, the water source of residents in urban area is very different. Majority of residents in urban area use water with public-tap or water piped directly into their residence. The two ways of water use covers over 70% of urban residents. 37% of residents have water piped into residence while 38.6% of residents use public tap.

3 Result

3.1 Linear relationship between urban and rural area

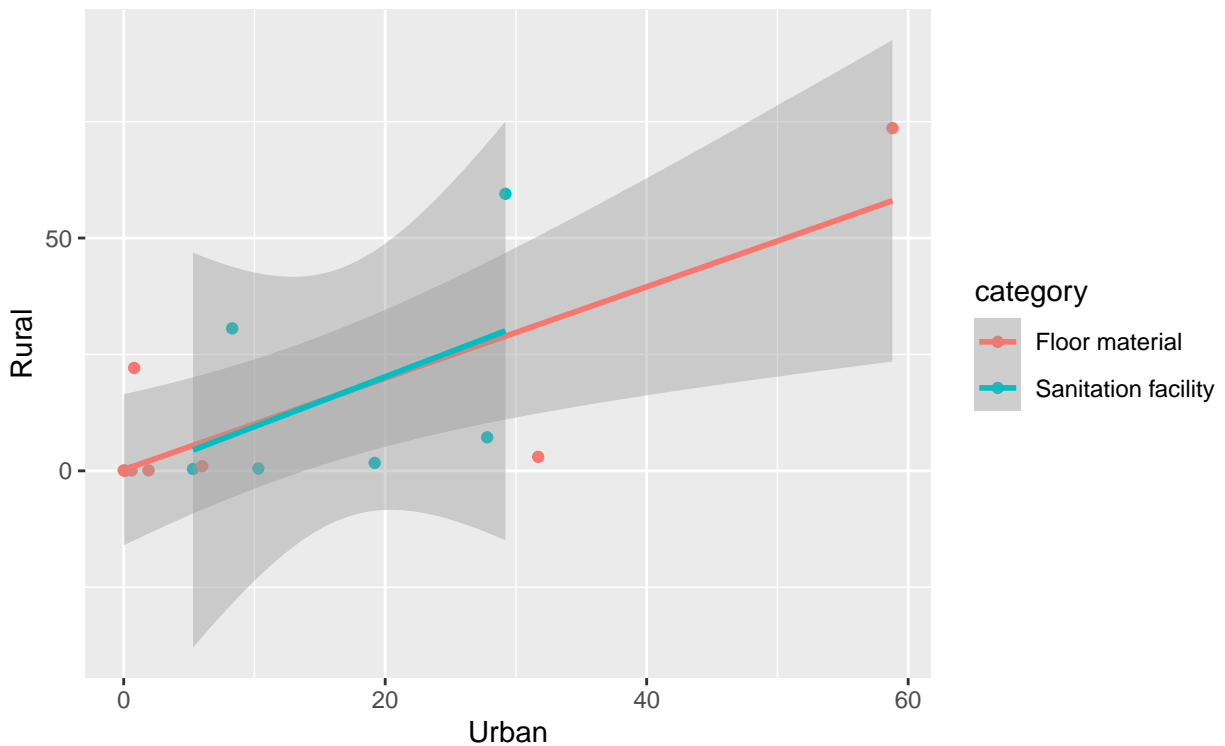


Figure 4: Regression model showing the relationship between rural and urban area with different categorical variables

To track the relationship between urban resident's housing characteristics and rural resident's housing characteristics, we generate the simple regression model 4 above, with red line represents floor material and blue line represents sanitation facility. The regression line shows the relationship of percentage of housing characteristics between urban and rural area. From the regression lines, we find that they both show positive relationship between the percentages of housing characteristics of urban residents and rural residents. This finding implies that the facility in rural area follow the development in urban area. It is saying that for one kind of advanced floor material sanitation facility, as more households in urban area use it, more households in rural area also follow the trend to use it.

3.2 Linear relationship between urban and rural area

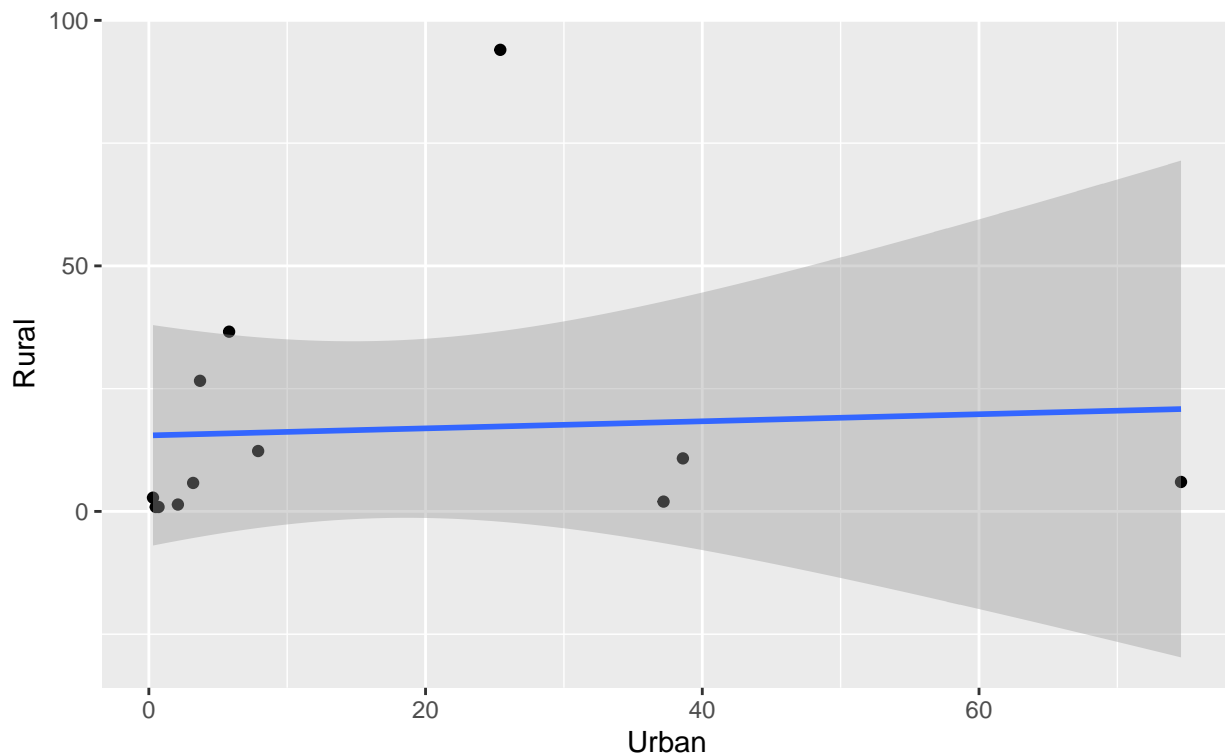


Figure 5: Regression model showing the relationship between rural and urban area in water quality

In the second result graph, we build linear regression 5 of the percentage of different source of drinking water between urban and rural area. From the plot, we find that there's no significant relationship. This shows that the use of water in rural and urban area is completely different. While households in urban area tend to use more advanced water facility, households in rural area are still using some old ways to get water. Regarding to what we have discussed in previous data section, we know that majority of residents in rural area tend to use water from river and stream. Borehole is also a common choice for residents in rural area. On the other hand, it is common for residents in urban area to have water piped into residence.

4 Discussion

4.1 Overview and result

In this paper, we analyze the distribution of different housing characteristics in urban and rural area. We first extract the table of housing characteristics from the report of demographic and health survey 1993. By reproduce the table, we are able to conduct further analysis. In Data section, we basically analysis the data of electricity, source of drinking water, floor material and sanitation facility with scatterplot and summary tables. In result section, we do simpler linear regression on urban and rural areas' percentage of different kind of housing facilities to have further discussion on the possible relationship.

In data section, by comparing the percentage of different kind of housing characteristics, we find that urban area has more advanced facility comparing to rural area. For instance, the source of drinking water in urban area is by piping directly into residence while in some rural area, residents are still using river or stream water. As for the other important housing characteristic, electricity, 94% of residents in rural area don't have electricity. This show that the supply of electricity and infrastructure of electricity in rural area is very insufficient. For sanitation facility, a larger proportion of households in rural area are using traditional pit latrine compared to rural area. About 10% of residents in urban area are even using own-flush-toilet, which is a kind of advanced toilet. What's more, for floor material, majority

of residents in both rural and urban area are using cement floor. There's not much difference in floor material selection. In result section, we generate regression model to discuss the relationship of different housing characteristics between rural and urban area. The positive relationship between the percentages of housing characteristics of urban residents and rural residents indicates that for floor material and sanitation facility, rural area is following the infrastructure development trend of urban area. As urban area uses more advanced flooring material and sanitation facility, rural area also began to use these advanced facilities. On the other hand, we don't recognize any significant relationship of source of drinking water between urban and rural areas. This is quite reasonable since the environment and construction in urban and rural area is very different. In rural area, people can get river water and stream water directly since they are closer to nature. In contrast, urban area is full of high buildings which need piped water.

4.2 Future development

According to the content of the paper, it widely expands our horizon about the world. Honestly, we are shocked by the immature infrastructures of the developing country in Africa in the past. The households of the African developing country were terrible in 1993, and nowadays, some of the fundamental facilities in Ghana still need to be improved, and some of the Ghanaian are still live in poverty. For example, 43 percent of the total population are still living in rural areas, and 29 percent of the jobs are within agriculture (n.d.). Moreover, 43 percent of Ghanaian do not have accounts in financial institutions like bank and 13.3 percent of people live on less than 1.9 dollars per day. Overall, the phenomenon of the Ghanaian condition is still not optimistic. Therefore, we could develop some method for them to improve their economic condition. For instance, the Ghana government could provide loan to the farmers to invest their agricultural business, especially providing them with skills of planting and harvesting cocoa trees, because the cocoa trees are appropriate to live based on their local environment (n.d.). Secondly, government should provide the parents in the country with more income, so that their children could have a better educational environment, because education always means less poverty (n.d.). In addition, the government could pay more attention to the rural areas and provide people there with mobile technology, in order to let them be accessible to the bank system. The modern technology is a efficient way of boosting economy in Ghana (n.d.).

4.3 Limitations

Though our paper discusses housing characteristics and aims to give further comments on Ghana's infrastructure, the data table we have is from year 1993, which is very old. The situation in the country may change much nowadays. Thus, our comments may be not very useful in today.

The table we got only describe the percentage not the exact number. We are not able to know the total number of households that have electricity and the total number of households that use spring as their drinking water. We can't get a general sense of the total number.

There are limited number of data when we convert the PDF table into dataset so that when we draw the scatterplot, it is hard for us to investigate the trend between different variables.

Reference

- n.d. <https://opportunity.org/our-impact/where-we-work/ghana-facts-about-poverty>.
- Daasebre. 1993. *Ghana Demographic and Health Survey 1993*. <https://dhsprogram.com/pubs/pdf/FR59/FR59.pdf>.
- Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Henry, Lionel, and Hadley Wickham. 2020. *Purrr: Functional Programming Tools*. <https://CRAN.R-project.org/package=purrr>.
- Ooms, Jeroen. 2022. *Pdftools: Text Extraction, Rendering and Converting of PDF Documents*. <https://CRAN.R-project.org/package=pdfutils>.
- 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. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.