Child care situation of working mothers in Ghana 1998*

Pascal Lee Slew Yunkyung Park

03 April 2022

Abstract

Working mothers are often faced with the conflict of responsibility towards their children and the role they play at work. The Ghana Demographic and Health Survey (GDHS) forms part of a global initiative to collect data on family, fertility, and child health. In this paper, we used the 1998 GDHS and focused on the child care and working mothers aspect. Geographics and other aspects such as education are used to predict child care preference. We found that families living in Greater Accra are more likely to send their children to school and the majority of working mothers look after their children by themselves. The analysis of the child care and working mothers data provides insight into current trends and has significant implications for public policy and program decisions.

1 Introduction

Child care matters to the development of a child. Previously, women were thought to fit the traditional role of a mother, being at home and looking after her children. This thought is being challenged with a growing presence of women in the workforce, contributing to economic activity. Women are often faced with the conflicting role they have at home and at work. However, sometimes, mothers have to work out of necessity to contribute to the financial situation of their families. Families have to find ways to raise their children. The quality of child care given tends to have positive effects on the development of the child (McCartney 2015). Having an idea of the current child care practices is important since it gives crucial information on the weaknesses of current policies and would give an insight on how to improve the situation around child care.

In this paper, we examined child care decisions of working mothers across regions, the mothers' highest education level attained, their occupation, work and employment status in Ghana. We conducted a descriptive analysis of child status and childcare preferences. We first looked at the number of working mothers from the sample population, then examined child care practices based on employment status, education level among other factors. The differences in child care practices were noticeable in different regions. We found that most working mothers tend to look after their young child by themselves and those living in Greater Accra have a higher tendency to put their children in child care centres compared to other regions. Another observation was the highest level of education attained may have an influence on the number of children they have. The proportion of working mothers with at least secondary education with no child under six years old is lower than those with at least one or more children under six.

The rest of the paper is divided into 4 sections. Section 2 explains the data source, the survey methodology, and some of its key features, strengths, and weaknesses. Section 3 relays the findings through graphs. Section 4 discusses in more detail what was found and some limitations of what we did and Section 5 contains a datasheet for the dataset we used.

*Code and data are available at: https://github.com/Pascal-304/dhs_analysis.

2 Data

2.1 Data Source

This paper uses the Childcare While Working Data (Ghana Statistical Service and Macro International Inc. 1999) obtained from the 1998 Ghana Demographic and Health Survey (GDHS). The dataset was collected by downloading the final report of the 1998 GDHS published in the Demographic and Health Surveys (DHS) program. The package 'pdftools' (Ooms 2022) is used to download and read the content into R by parsing the pdf. The package 'stringi' (Gagolewski 2021) and 'tidyverse' (Wickham et al. 2019) are then used to gather the data.

2.2 Methodology

The 1998 Ghana Demographic and Health Survey was a national-level survey carried out in Ghana. It was conducted by the Ghana Statistical Service collaborated with the Macro International Inc. It was designed to collect data on fertility, family planning, and maternal and child health. The 1998 GDHS was conducted from mid-November 1998 to mid-February 1999.

The survey frame was created using the list of Enumeration Areas (EAs) with population and household information from the 1984 Population Census. The 1998 GDHS used a two-stage stratified sampling method. At the first stage of sampling, 400 EAs were selected using systematic sampling with probability proportional to size where the selected EAs comprised 138 in the urban areas and 262 in the rural areas. A complete household listing operation was then carried out in all the selected EAs to provide a sampling frame for the second stage selection of households. At the second stage of sampling, a systematic sample of 20 households per EA was selected in the Northern, Upper West and Upper East Regions, and 15 households per EA was selected in the other regions.

The target sample size for the 1998 GDHS was 6,000 consisted of 4,500 women of age 15-49 and men of age 15-59. A total of 4,843 women age 15-49 and 1,546 men age 15-59 completed the survey. To account for non-response, 6375 households were selected for interview.

The GDHS was composed of 3 parts: the Household's questionnaire, the Woman's questionnaire and the Man's questionnaire. All the questionnaires were modified to account for the situation in Ghana.

The survey was conducted in person. The interview staff were first trained for a period of three weeks. It was conducted by a total of 14 teams, comprising a team supervisor, a field editor, three interviewers and a driver. The response rate was 97% for individual men and women selected. Non-response was mainly due to inability to find the selected interviewee at home. Repeated callbacks were made to try to reach the selected respondents. The results of the survey did not disclose any confidential information that might identify the respondents. The survey was a voluntary one and participants were informed first.

2.3 Strengths

The questionnaires were made available in the five major local languages (Akan, Ga, Ewe, Hausa, and Dagbani), making the survey accessible to a larger population. After the household listing, spots checks were conducted to make sure the work has been done correctly. Before the actual survey, a test was conducted and adjustments were made to the questionnaires based on the feedback obtained from the test. The questions were comprehensive and detailed to collect information on the socio-economic status of the household. They were also adapted to the laws and norms of the nation.

2.4 Weaknesses

Since the questionnaires were only available in the five major local languages, if the selected household was from a minor ethnic group and know only their own language, then it would be difficult for the interviewee

Table 1: A subset of key features

Background Characteristic	Child under six		Child's caretaker	Number of employed women	Category
	Yes	No	Respondent		
Greater_Accra	39.2	60.8	39.7	564	Region
Volta	57.1	42.9	46.1	384	Region
No_education	62.3	37.7	51.0	1158	Mother's_education
Primary	57.1	42.9	50.4	683	Mother's_education
Middle	50.8	49.2	44.1	1392	Mother's_education
Secondary+	36.6	63.4	39.1	332	Mother's_education
For_family_member	54.5	45.5	56.8	355	Work_status
For_someone_else	36.5	63.5	38.5	491	Work_status
Self-employed	57.7	42.3	47.5	2718	Work_status
Agricultural	64.3	35.7	52.3	1160	Occupation
Total	54.4	45.6	47.6	3564	Total

[a] Decimals are in percentage (%).

to understand and answer the survey. Visitors who stayed a night before in the selected household were eligible for the interview. The visitor may be from a different region, so the targeted sample size for each region may be skewed. There is also the possibility of sampling error.

2.5 Key features

Some key features include background and a percent distribution of currently employed women by whether they have a child under six years of age at home according to selected background characteristics, Ghana 1998. If the respondent answered to have one or more children under six, then they were further asked to indicate the caretaker of the child such as the respondent themselves, their partner, or whether they hired a person for it, etc.

The dataset is processed and analyzed using 'R' (R Core Team 2021) mainly with the 'tidyverse' (Wickham et al. 2019) and 'dplyr' (Wickham et al. 2021) packages. The 'janitor' (Firke 2021) is used to clean data, and the graphs are created in 'ggplot2' (Wickham 2016) and 'ggforce' (Pedersen 2021) packages. The map is created using in 'rnaturalearth' (South 2017) and 'sf' (Pebesma 2018) packages. The package 'stringr' (Wickham 2019) is used to manipulate the character string. The packages 'bookdown' (Xie 2016) and 'knitr' (Xie 2014) are used in generating the R Markdown report. Table 1 created using 'kableExtra' (Zhu 2021) shows a subset of key features that will be discussed in this paper.

2.6 General Overview

We can tell from Table 1 that there are 3,564 working mothers, and over a half of them have a child under six years. Figure 1 illustrates that 47.6% of these mothers look after their own children while they are at work, 22.1% have their relatives other than their husband/partner to take care of the child, 14% send the child to school or other institutional care. On top of that, less than 3 percent of women have husbands/partners to look after the child while they are at work.

It should be noted that 1.9% have not worked since their last birth, and this indicates the percentage of the respondents who were not working at the time of the survey but had been employed at some time during the last 12 months.

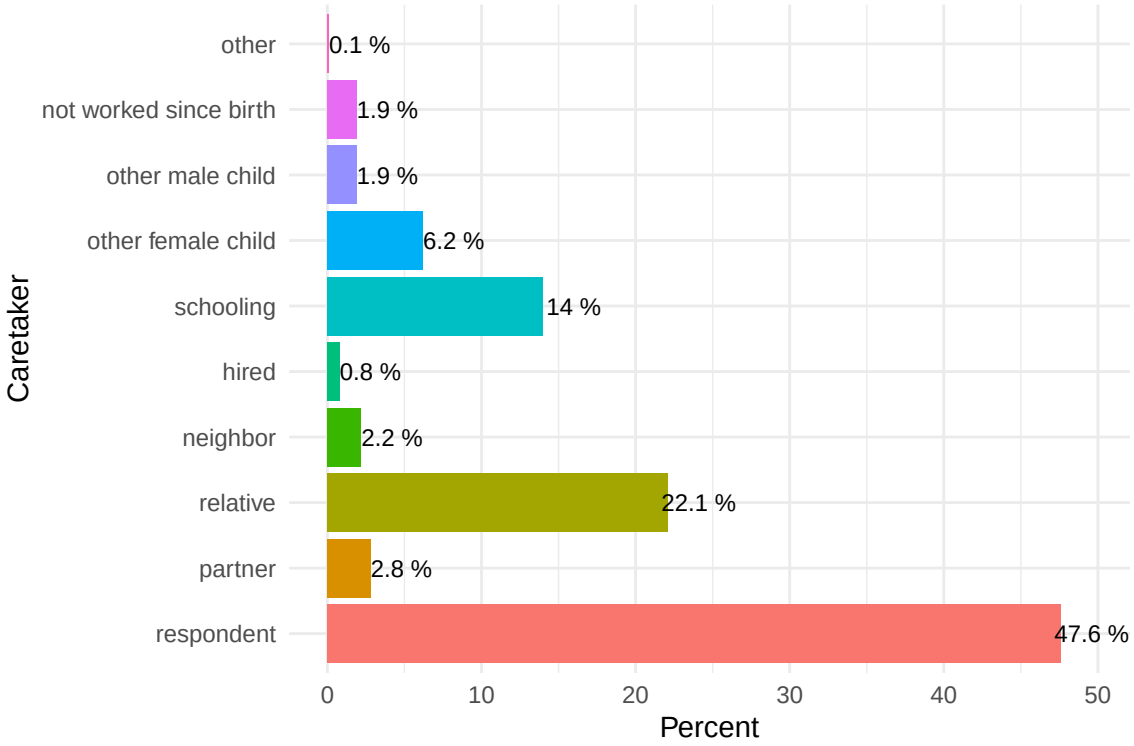

Figure 1: Distribution of the child caretaker for working mothers

3 Results

3.1 Region

Figure 2 shows a map of Ghana and its administrative regions. Ghana has 10 administrative regions, namely, Upper West, Upper East, Northern, Brong Ahato, Western, Ashanti, Central, Eastern, Volta, and Greater Accra. In 1998, about 57.6% of the Ghanese population were living in rural areas (Sasu 2022). Greater Accra is considered the most urbanized region of Ghana.

Figure 3 shows the number of employed mothers in urban and rural areas. There are 1,227 women who were selected in Urban areas while 2,338 mothers were selected in rural areas. There are less women living in urban who are currently working, and this is not surprising as a larger population live in rural areas.

Figure 4 shows the number of employed mothers in the different administrative regions. There were more working mothers in Greater Accra and Ashanti while Upper West had the least number of working mothers. The reason might be because the number of people living in the different administrative regions is different, therefore, the sample size differed from one region to another. Greater Accra is the most urbanized region which implies more work opportunities and therefore, a larger population would be concentrated there.

Figure 5 shows the child status of families in Urban and Rural areas. In rural areas, there is a larger percentage of working mothers having one or more children under six years old. The converse is true in Urban areas. Working mothers tend to have no child under six years old.

Figure 6 shows the Ghanese families' caretaking preferences in both urban and rural areas. The preferred caretaking method in both urban and rural areas is the working mothers looking after their children by themselves. The next preferred method in both areas would be to let the children be looked after by relatives. Other notable observations would be that families living in urban areas tend to put their children

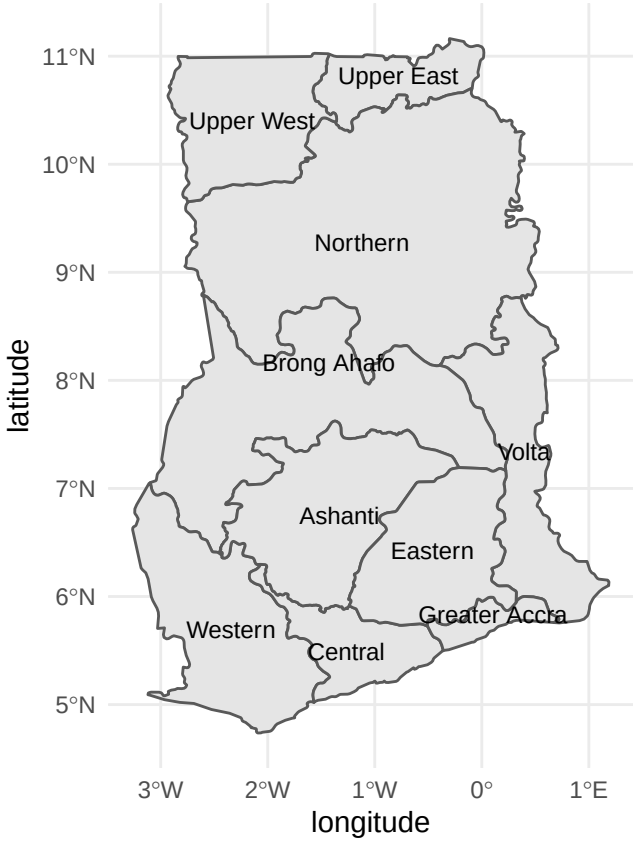

Figure 2: Ghana administrative regions map

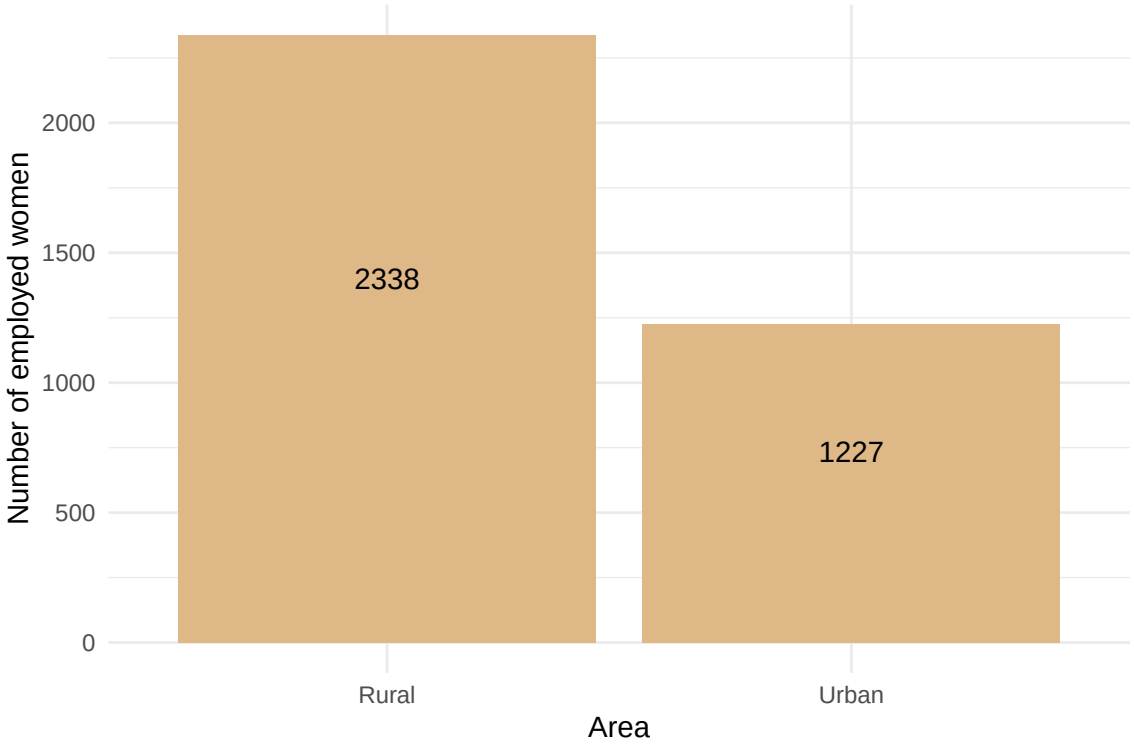

Figure 3: Comparison of the number of working mothers between urban and rural areas

in school and in rural areas, families let the older female child care for the under six years old child much more than in urban areas.

Figure 7 shows the caretaking preference of Ghanese families in the different administrative regions. Working mothers tend to look after their children themselves the most in all regions. In Greater Accra, families choose to put their child in school the most compared to other administrative regions. Another observation is that the proportions of families opting to put their child in school in Northern, Upper East, and Upper West are the lowest, namely 1.3%, 2.7%, and 2.0% respectively. The other popular caretaking options after the mother herself are by a relative or by another older female child.

Figure 8 shows the child status of Ghanese families in the different administrative regions. By child status, we mean whether the families have children under six years old. Ghanese families tend to have one or more children under six years old rather than no child under six with the exception of Greater Accra. The proportion of families with one or more children under six is slightly more pronounced in Northern, Upper West, Brong Ahafo, and Upper East regions.

3.2 Mother's education

Figure 9 shows that there are significantly fewer employed women who studied at secondary school or higher. Also, it can be observed that the working mothers who did not continue their education until secondary school are more likely to have a child under six, whereas there are more employed women who studied beyond secondary school and have no child under six. In addition, when we take a look at the number of employed women who did not get any education, we can see that there are a lot more women with a child under six than without, and the difference between them is notably big. However, as the level of working mothers' education gets higher, the gap between having a child under six and no child under six decreases, and it gets reversed when the women complete secondary school.

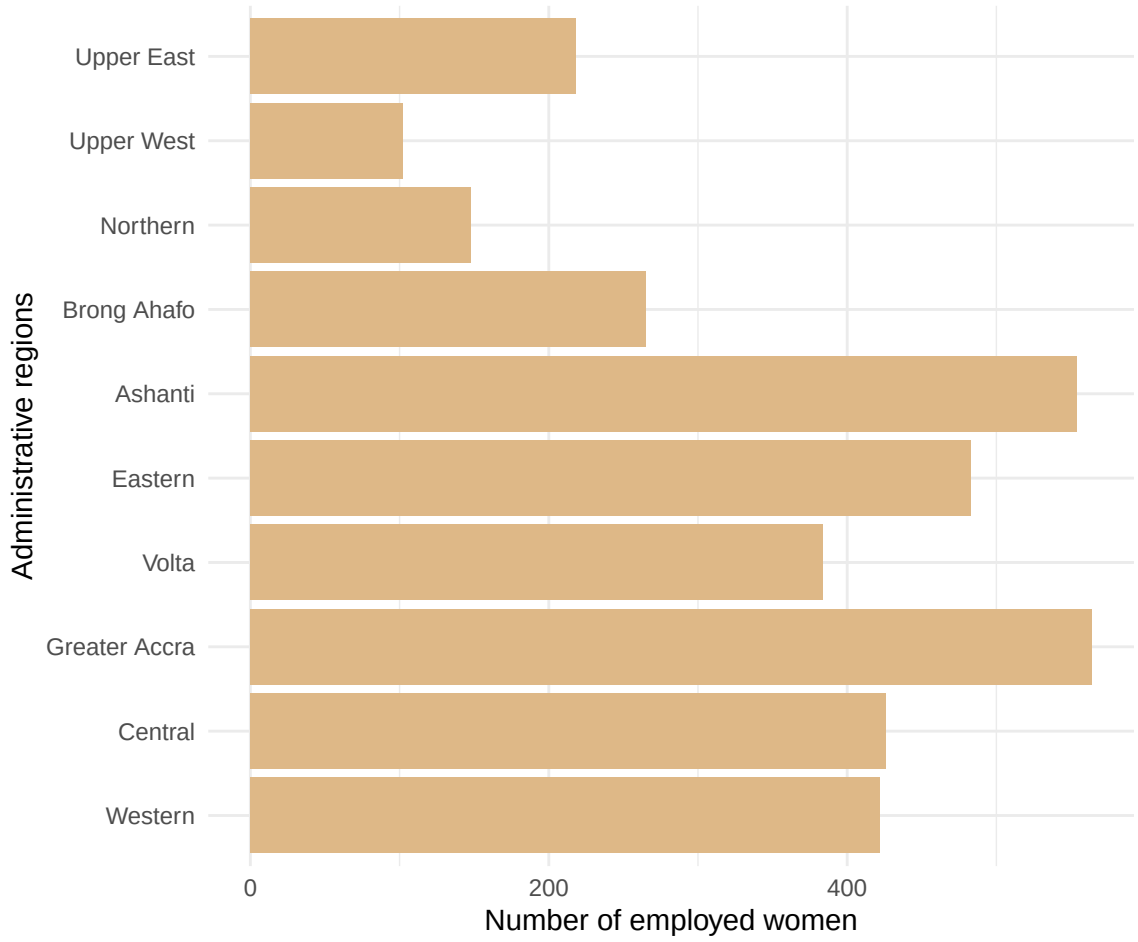

Figure 4: Comparison of the number of working mothers selected in the different administrative regions

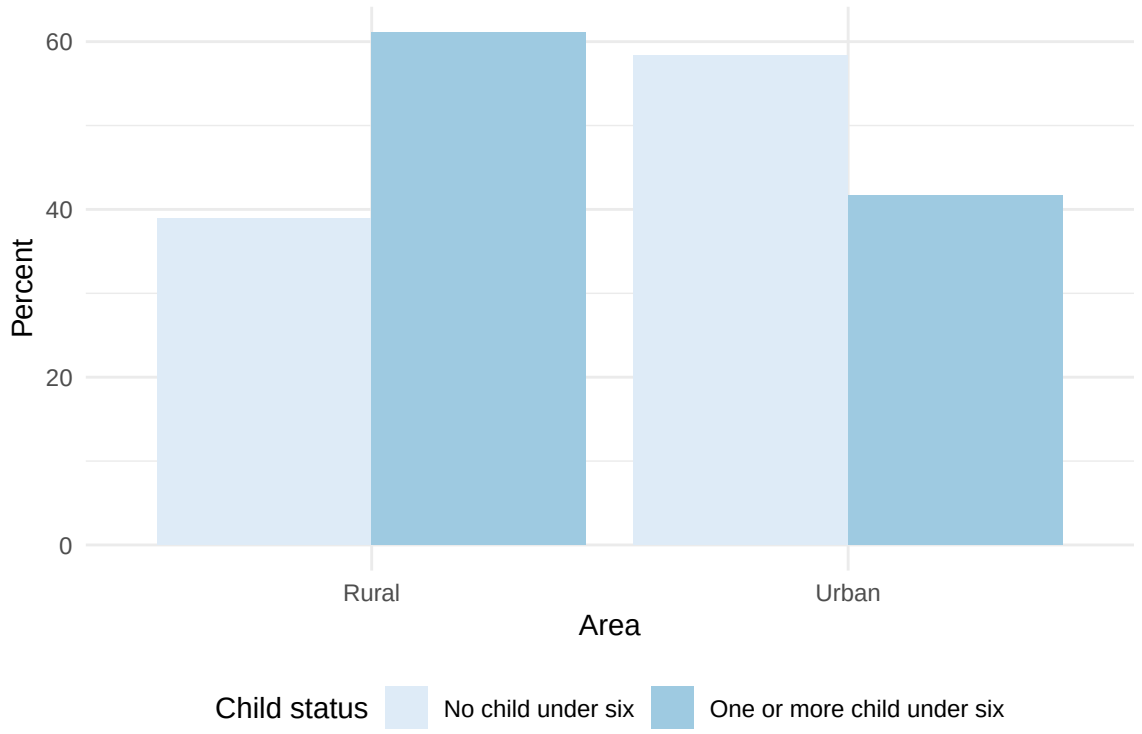

Figure 5: Comparing the percentage of mothers having no child under six to one or more in both Urban and Rural areas

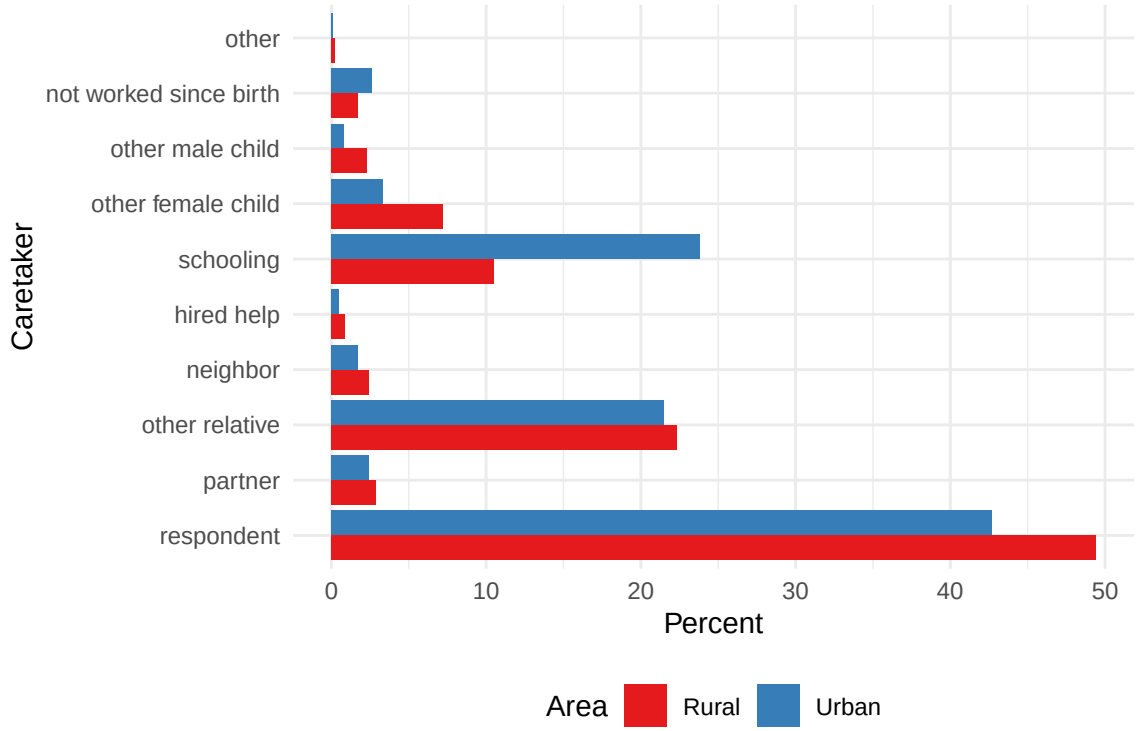

Figure 6: Comparing child caretaker preferences in Urban and Rural areas

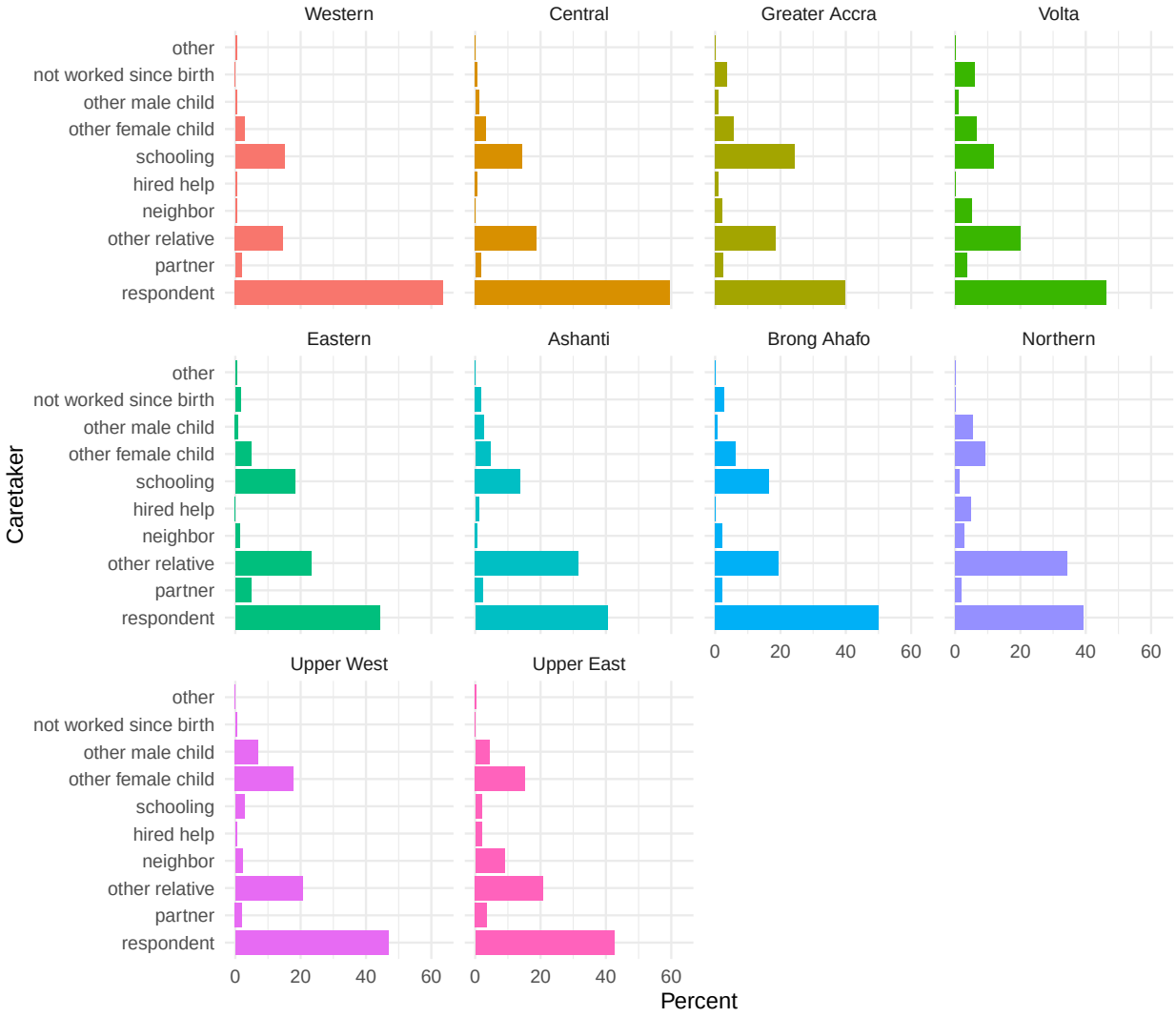

Figure 7: Caretaking behaviour in the different administrative regions

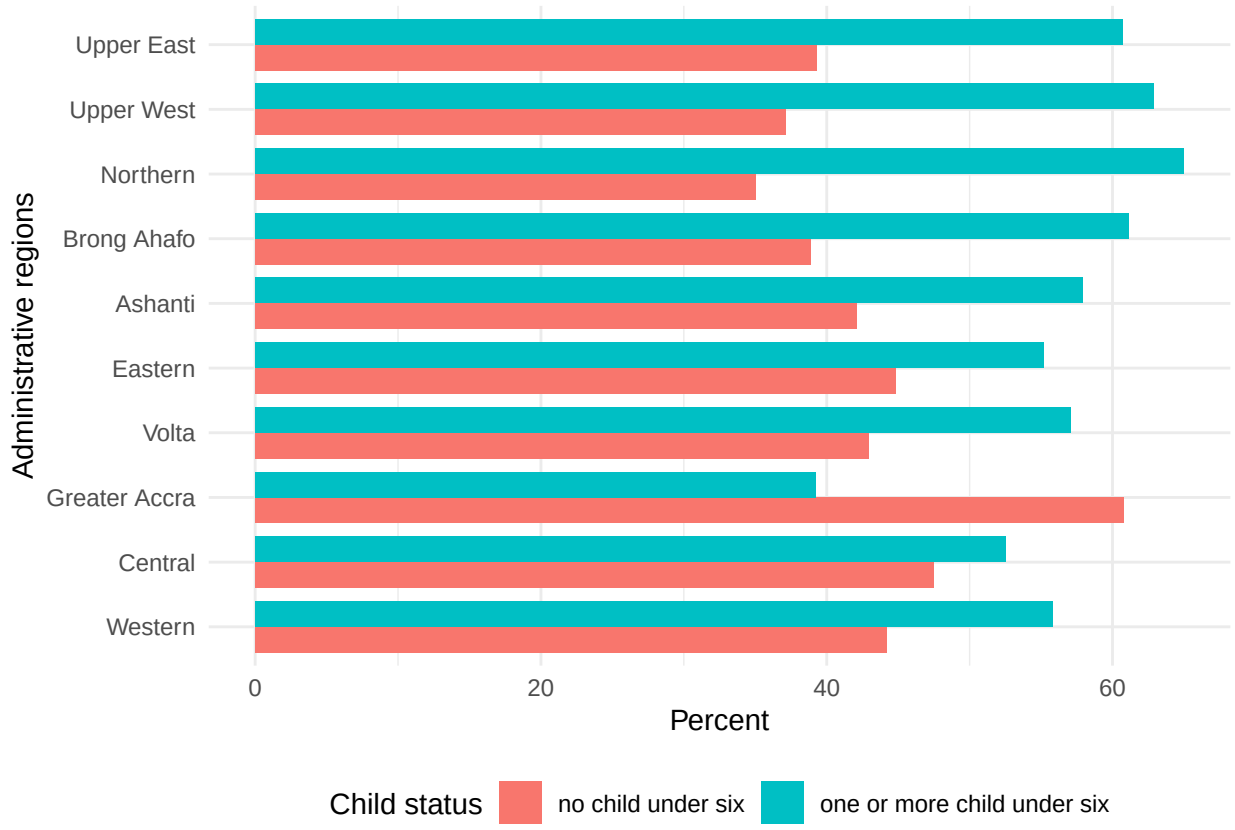

Figure 8: Comparing child status in different administrative regions

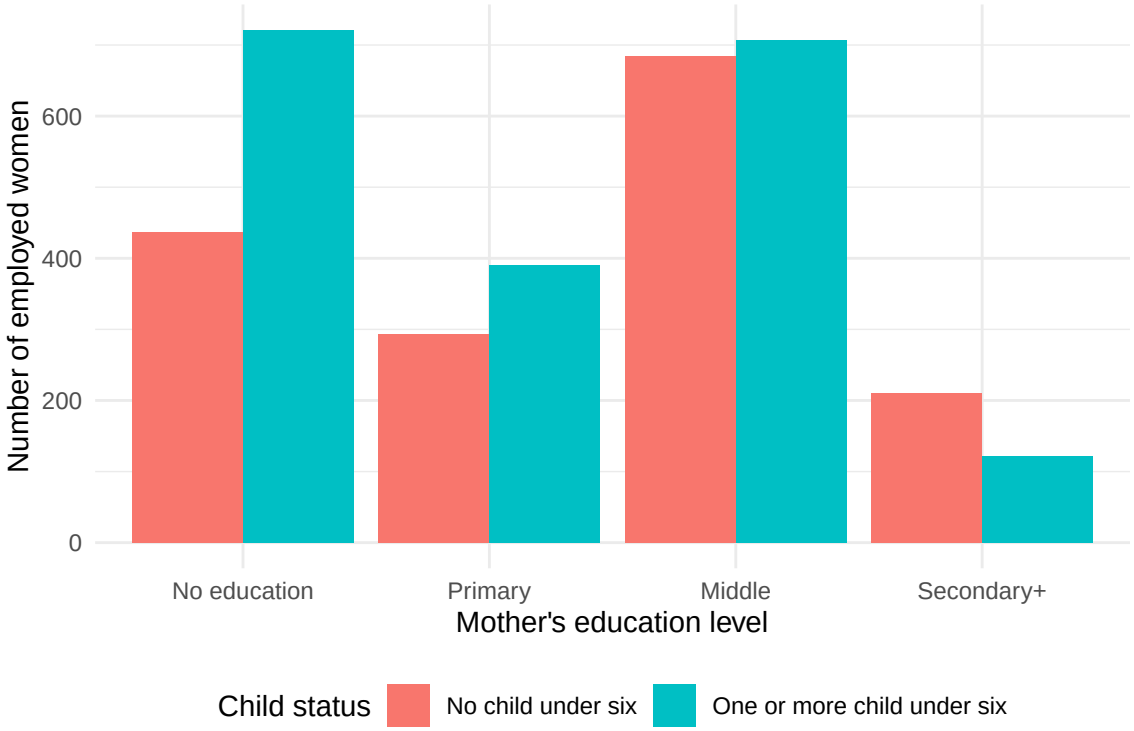

Figure 9: Comparing the child status of working mothers based on their education level

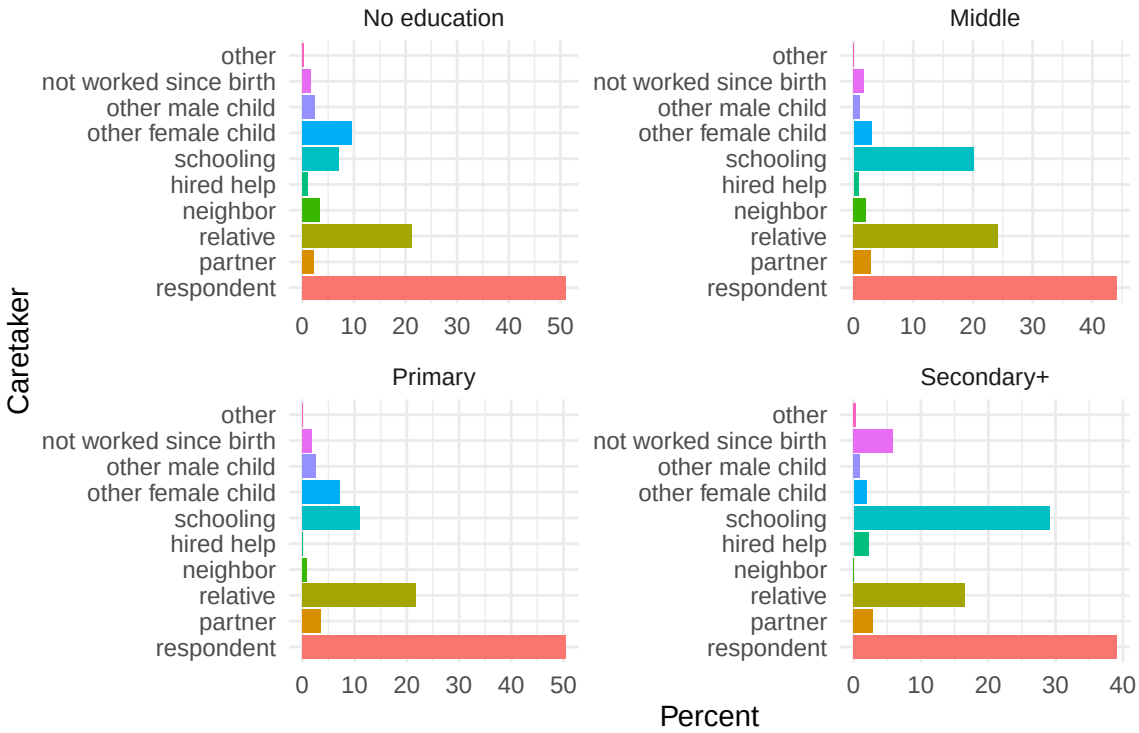

Figure 10: Distribution of child's caretaker by mother's education level

Figure 10 demonstrates that no matter what their education level is, it is mostly the employed women who take care of their children while working. When the level of education is lower, then it is more likely for the respondent to look after their own children under six years. Over 50% of the women with no education take care of their own children whereas less than 40% of the women with secondary or higher levels of education do. As the level of education gets higher, then the percentage of schooling their children gets higher as well. Moreover, it is worthwhile to mention that the women who completed high school are more likely to stop working after having a birth compared to the women whose education level is lower than secondary school.

3.3 Work status

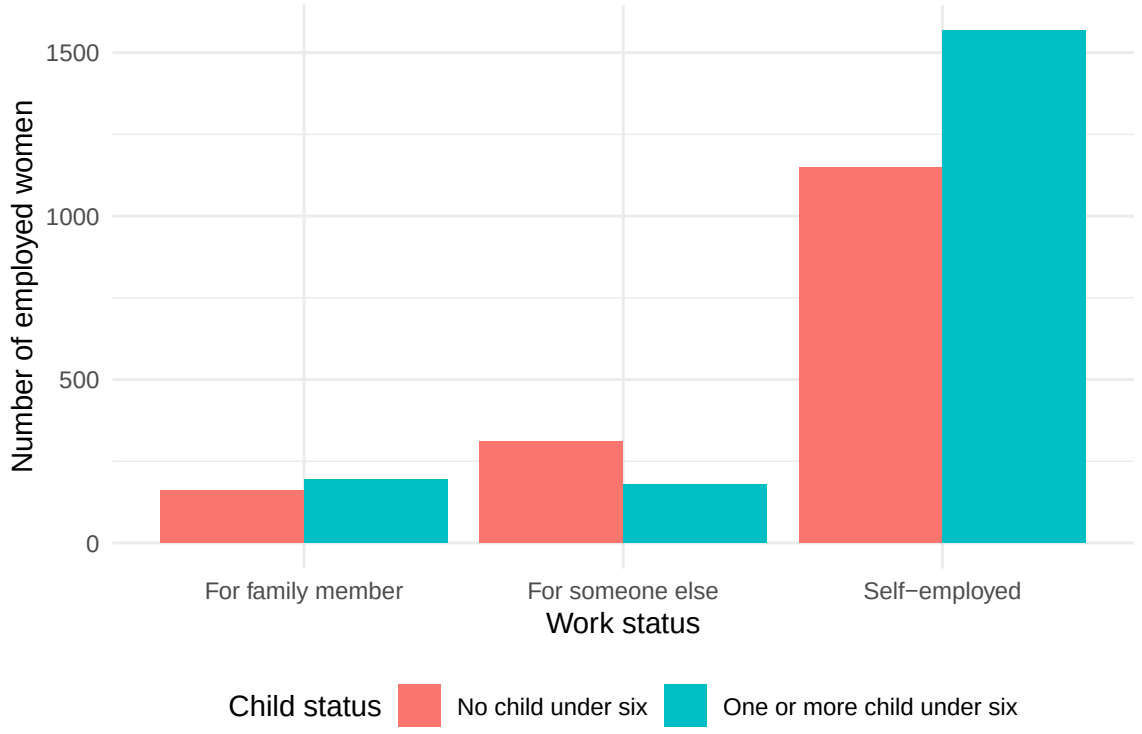

Figure 11: Comparing the child status of working mothers based on work status

Figure 11 shows that most employed mothers are self-employed. If the women work for themselves or for their family, then it is more likely to have one or more than one child under six years old, whereas when they work for someone else, there are more women who do not have a child under six years old. It should be noted that no child under six does not mean that they have no child.

It can be observed from Figure 12 that most employed women take care of their children by themselves regardless of whom they work for. It is their relatives who take care of their children while they are at work. We can tell that if they are self-employed or work for the family, then it is more likely to have a family member taking care of their children under six years old, whereas the percentage of families opting for schooling option is relatively higher when they work for someone else.

3.4 Occupation

Figure 13 shows that there are more women working in non-agricultural occupations. When we compare the mothers working in the agricultural sector, the number of women with a child under six is twice as many as

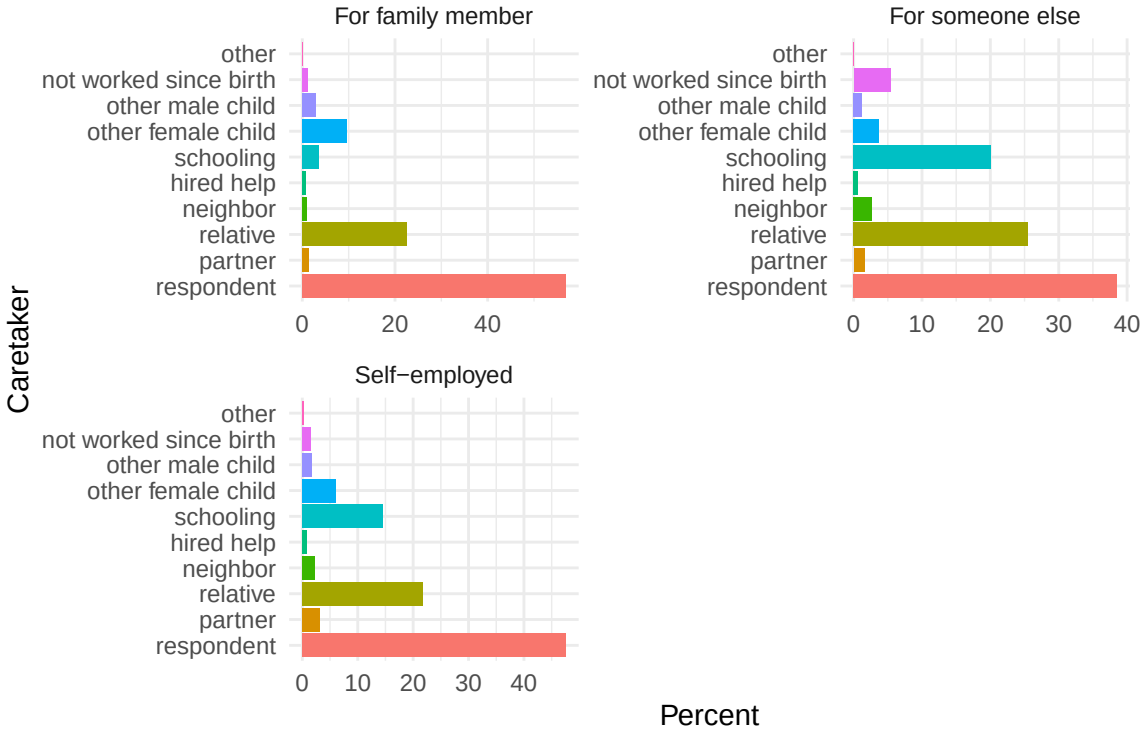

Figure 12: Distribution of child's caretaker by work status

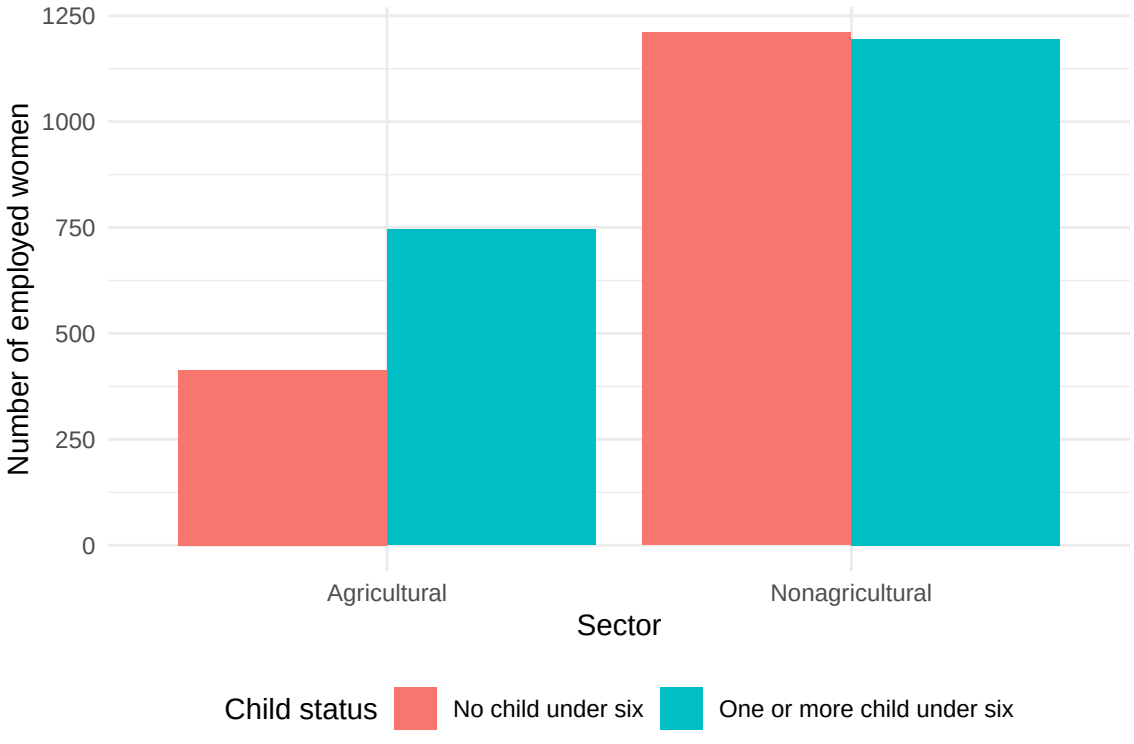

Figure 13: Comapring child status based on occupation

the number of the ones without. On the contrary, the number of women with no child under six is slightly higher than that of the ones with a child in non-agricultural occupations.

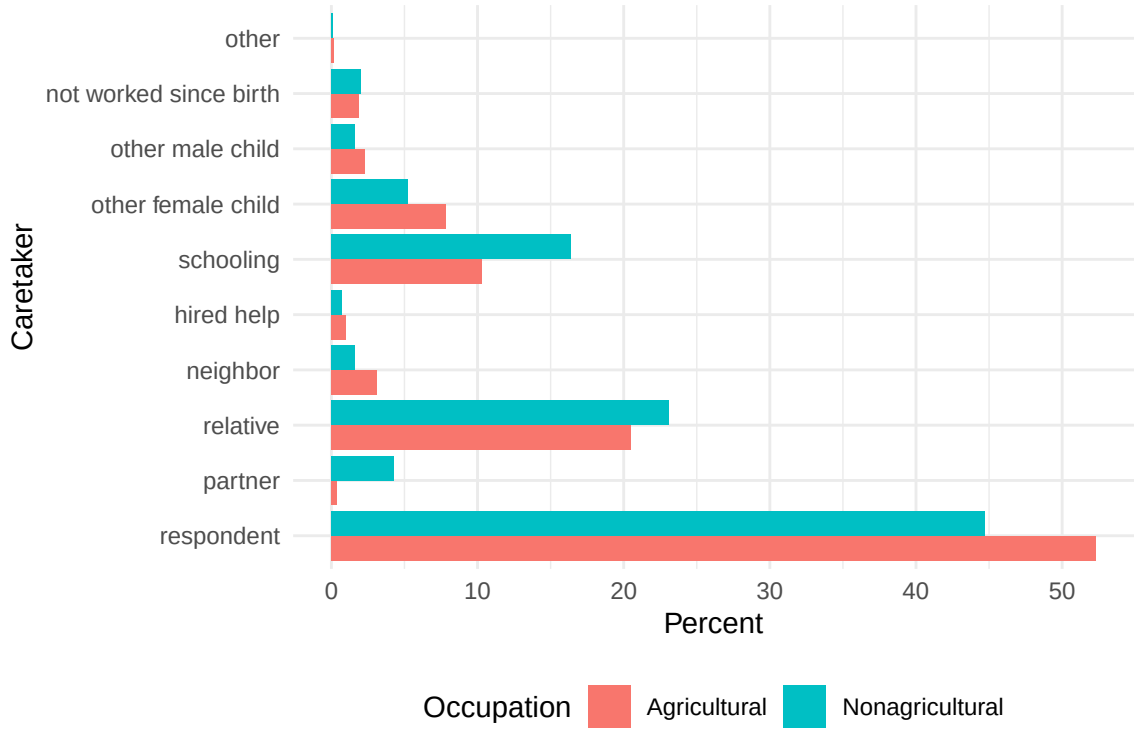

Figure 14: Distribution of child's caretaker by occupation

Figure 14 illustrates that the percentage of the child's caretaker is highest among the respondent followed by the relative and schooling regardless of what type of jobs they have. This means that it is mostly the case that the employed women take care of their children under six while they are working. Meanwhile, women who work in the agricultural sector are more likely to care for a child themselves than women in the non-agricultural occupations. In particular, more than a half of the employed women working in the agricultural industry are the caretaker of the child under six whereas around 45% of the women working in non-agricultural sector take care of a child themselves while working. Furthermore, we can see that it is more likely for the female siblings to take care of the kids under six than the male siblings.

3.5 Employment status

Figure 15 illustrates that the majority of the employed women work throughout the entire year. Specifically, the majority of them work full week. There are few women who work occasionally. On top of that, regardless of their employment status, more working mothers tend to have one or more children under six.

We can see from Figure 16 that the percentage of the respondent being the primary caretaker of a child under six is remarkably higher than the others in all four employment statuses. It exceeds 40% in all four cases, and in particular, mothers working full time are more likely to look after their own children than the others. Next, it is the relatives who mainly take care of the child under six. It should be noted that there are more women sending their kids under six to school when they work all year compared to the ones who do not work throughout the entire year.

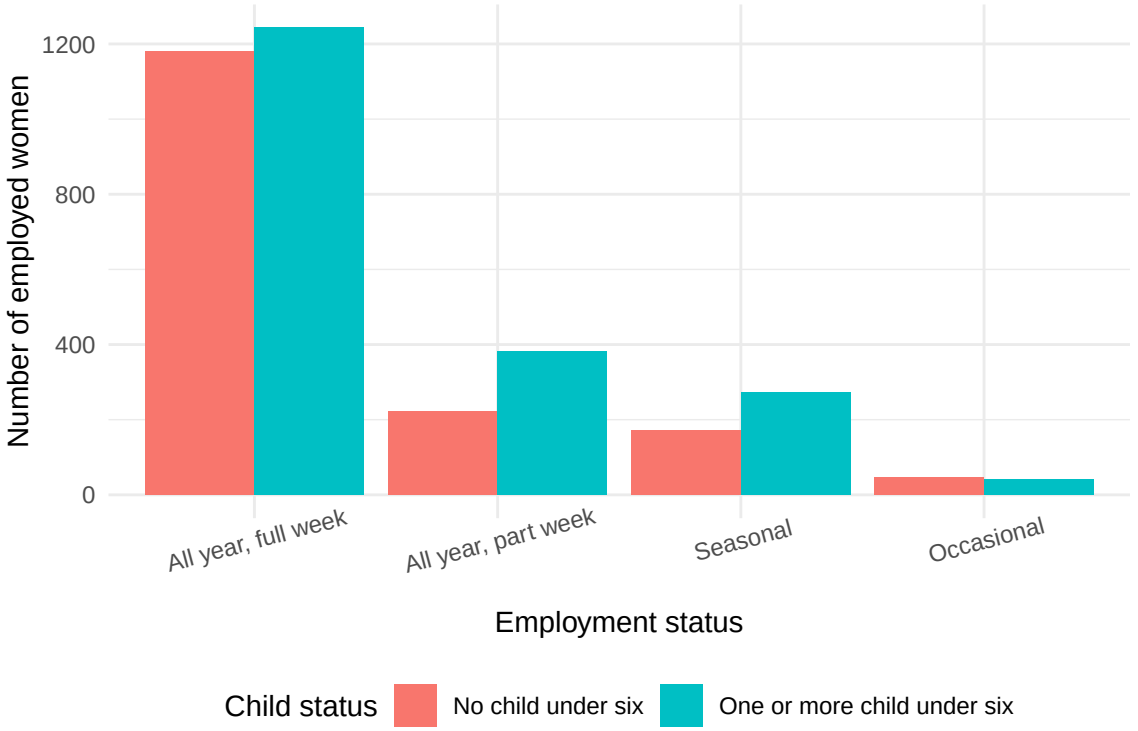

Figure 15: Comparing child status based on employment status

Figure 16: Distribution of child's caretaker by work status

4 Discussion

4.1 Employed Women

Table 1 illustrates that 3,564 out of 4,843 women of age 15-49 who completed the survey are employed. In other words, around three in four women were employed during the 12 months before the survey. Figure 3 shows that there are more employed women in the rural area, and this is because Ghana's population is predominantly rural. 34% of the country's population was reported to be residing in urban areas, and 66% percent were living in rural areas (Ghana Statistical Service and Macro International Inc. 1999).

Next, Figure 4 illustrates that women residing in the Upper West Region are least likely to be employed. Together with Figure 2, we can tell that women are more likely to be employed when they live in the southern part of Ghana. The northern part of Ghana including Upper West, Upper East, and Northern Region are the bottom 3 regions in Figure 4 which demonstrates the distribution of the number of employed women whereas the southern part of Ghana including Greater Accra, Ashanti, and Eastern Region are the top 3.

Additionally, it is surprising to see from Figure 9 that the number of employed women whose education level is a secondary school or higher is remarkably lower than that of the other education levels. Meanwhile, Figure 13 shows that women are approximately twice as likely to be employed in the non-agricultural sector than in the agricultural sector. Lastly, when women are working, then Figure 11 tells that they are most likely to be self-employed and Figure 15 shows that they are most likely to work full-time.

4.2 Presence of Child Under Six Years Old

Table 1 shows that more than a half of the employed women have a child under six years. Figure 5 demonstrates that working women in rural areas are more likely to have at least one child under six years whereas working women in urban areas tend to have no child under six years. Figure 9 shows that as the level of working mother's education gets higher, it is more likely to not have any child under six.

Figure 11 illustrates that when they work for themselves or for their family, then there are more women with a child under six whereas when they work for someone else, then there are more women without. Figure 3.4 shows that when women are working in the agricultural sector, they are more likely to have a child under six whereas when they work in the non-agricultural sector, it is equally likely to have a child under six and without. Lastly, Figure 15 tells that it is more likely to have a child under six years regardless of their employment status. In other words, no matter whether they work full-time or part-time, there are more women who have at least one child under six years.

4.3 Child's Caretaker

Every figure that demonstrates the distribution of the child's caretaker shows that it is most likely to be the mother primarily taking care of their child under six while they are at work. In other words, no matter where they live, what their education level is, and what type of job they have, the majority of the employed women are the ones who look after their children. Therefore, we can deduce that in Ghana it was common that the mother is primarily responsible for the care and raising of a child back in 1998. In 2002, UNICEF conducted a survey from 31 countries, and it asked working women with children under the age of 6 years about the typical childcare arrangement while they are working (Samman and Lombardi 2019). It was reported that 39% of the respondents took care of their children themselves on average, while in poor countries, the percentage rose up to 44% (Samman and Lombardi 2019). Thus, we can tell that the notably high percentage of working mothers looking after the children themselves indicates that Ghana was a low-income country.

Besides the mother being the primary caregiver of a child under six, it can be observed in Figure 10, Figure 12, Figure 14, Figure 16 that the relatives other than husband/partner are likely to look after children under six. Also, there are some families where the older siblings take care of their siblings under six. Thus, we can

tell that if there is a child under six in the family, then it was usual for the family member to take care of them in Ghana, 1998.

In addition, we can observe that the proportion of households hiring people to look after their kids is very low. This is presumably because of poverty, which is discussed above. Instead of hiring a person, there are more cases reported that they sent the kids to school or to institutional care. Figure 6 demonstrates that the percentage of employed women sending their kids to school is higher in urban areas than in rural areas. Figure 10 shows that the number of working mothers sending their kids to school increases as their mother's education level gets higher. Figure 12 illustrates that it is more likely to have their kids in school when they work for someone else than when they work for a family member or when they are self-employed. Figure 14 shows that the percentage of employed women sending their kids to school is higher in non-agricultural sectors than in agricultural sectors. Lastly, we can see from Figure 16 that there are more women sending their kids to school when they work all year compared to the ones who work seasonally and occasionally.

Lastly, we can tell that it is rarely the case that the mother stops working after having birth. However, it is worth noting that the percentage of employed women who stop working after birth is higher in urban areas than in rural areas from Figure 6. In addition, Figure 10 shows that the women who completed high school have more tendency to stop working after having birth than the ones with a lower education level. Furthermore, Figure 16 shows that the percentage of employed women who stop working after birth is highest when they work all year, full week.

4.4 Limitation and weaknesses

4,843 women of age 15-49 completed the survey, and 3,564 of them responded that they are employed. Then, they were further asked to indicate whether they have a child under six. However, it is not mentioned whether their marital status was asked prior to this question. If the question was asked to every woman who participated in the survey, then the respondents who answered that they have no child under six might be single. This dataset is created to identify a woman's overall status, and the percentages of the employed women by their marital status and the presence of a child are considered important indicators. However, it is unclear what they are trying to say by grouping women only by the presence of a child under six years.

Additionally, it would have been better if they were first asked for their marital status and the presence of dependent children. They could have then asked the age of the youngest child in the household or whether they have a child under six in the household. As married life has been reported to have a gendered effect on women and men's labour market outcomes (Gammarano 2020 [Online]), the distribution of the marital status among the employed women would be useful to tell the woman's overall status. In this context, if the respondent answered that they stopped working after the last birth, they could have further asked the reason for stopping working and about the job that they used to have before having birth.

Lastly, only the employed women who have a child under six were further asked to indicate who takes care of their children while they are at work. The dataset included the percentage of the missing values, and we purposely excluded them in plotting the graphs as they do not tell anything about the data itself and there were only a few missing values. However, when we add up the percentage of all possible items including the missing values, there are several background characteristics where the numbers did not add up to 100. As only the percentage is provided in the report, there might be a round-off error. Particularly, since the numbers were rounded to one decimal place, there is a high possibility that there was a rounding error. To make the dataset more precise, they could have rounded to two decimal places.

On top of that, the survey asked about the employment situation during the 12-month reference period preceding the survey. However, the respondents were allowed to answer using their previous experiences. In particular, this applies to the respondents who had a job before having birth but stopped working after the last birth. Since we do not know the age of the youngest child in their household, they might have answered the questionnaire based on their employment situation prior to the 12-month reference period. Therefore, the time period does not match among the questions.

5 Appendix

5.1 Datasheet

The questions were extracted from Gebru et al. (2021)

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*
    - The dataset was created to enable a descriptive analysis of the child care situation of working mothers in Ghana. The dataset would be used to look at the trends concerning child caretaking across different factors such as region, work status or occupation area. We were unable to find a publicly available dataset.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
    - The dataset was created by Yunkyung Park and Pascal Lee Slew from the University of Toronto.
3. *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*
    - No funding was obtained/required for the creation of the dataset.
4. *Any other comments?*
    - None.

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*
    - The dataset consists of categories: residence (urban or rural), regions of Ghana, mothers' education level, work status, occupation and employment status. For each category, we have different features: child status (whether families have children under six years old), childcare preferences and the number of employed women.
2. *How many instances are there in total (of each type, if appropriate)?*
    - There are 26 rows and 17 columns in the dataset. The residence category has 2 rows, regions has 10 rows, mothers' education level has 4 rows, work status has 3 rows, occupation has 2 rows and employment status has 4 rows.
3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*
    - The dataset contains all the data possible.
4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*
    - Each category consists of already processed data. The observations are mostly percentages which represent the proportions of categories.

5. *Is there a label or target associated with each instance? If so, please provide a description.*
   - No
6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*
   - Yes, some participants did not respond to all survey questions, leading to some missing data in the data collection process.
7. *Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*
   - No
8. *Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*
   - No
9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*
   - There may be possible sampling error.
10. *Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*
    - The dataset was extracted from a publicly available pdf at https://dhsprogram.com/publications/publication-FR106-DHS-Final-Reports.cfm
11. *Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*
    - No, the results of the survey used to create the dataset were made confidential so that we are unable to identify the participants.
12. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*
    - None
13. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*
    - Yes, the population would be the working mothers. Sub-populations would be identified by the categories mentionned above such as employment status, occupation, mothers' education, region, residence and work status.
14. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*
    - No
15. *Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*

- No

16. *Any other comments?*
    - None

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.*
    - No, the data are percentages of categories based on results from the 1998 Ghana Demographic and Health Survey responses.
2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*
    - The data was collected by extracting data from a pdf. In our case, we extracted data from a table located at page 45 in the 1998 Ghana Demographic and Health Survey pdf available at the link: https://dhsprogram.com/publications/publication-FR106-DHS-Final-Reports.cfm
    - The data collection was done using R (R Core Team 2021) and R packages tidyverse (Wickham et al. 2019), pdftools (Ooms 2022) and stringi (Gagolewski 2021).
    - We used the R package pointblank (Iannone and Vargas 2022) to set up tests and check on our dataset.
3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*
    - No
4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*
    - The raw survey data was collected by teams consisting of a team supervisor, a field editor, three interviewers and a driver. They were trained for 3 weeks. No mention of compensation was made.
5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.*
    - The raw GDHS data was collected from mid-November 1998 to mid-February 1999.
6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*
    - The GDHS procedures and survey questionnaires were reviewed by the ICF Institutional Review Board (IRB) of Ghana.
7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*
    - The GDHS raw data was collected from individuals directly while the dataset created was created from a publicly available pdf on the DHS website.
8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*

- Yes, the individuals in questions were informed since data collection of the GDHS survey was in person. However, for the "raw" data from the pdf, it did not involve individuals directly since their responses were already processed together.

9. *Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.*

    - For the GDHS survey, the participants consented to the collection and use of their data. Their data would be made anonymous and public but they were not explicitly informed that it would be used in this way.

10. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).*

    - N/A

11. *Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.*

    - N/A

12. *Any other comments?*

    - None

Preprocessing/cleaning/labeling

1. *Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.*

    - Yes. The "raw" data obtained from the pdf was cleaned using the tidyverse r package (Wickham et al. 2019) in R (R Core Team 2021).

2. *Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.*

    - Yes, both the clean and "raw" data are saved and available at the link: https://github.com/Pascal-304/dhs_analysis/tree/main/outputs/data

3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*

    - We used the statistical software R (R Core Team 2021).

4. *Any other comments?*

    - None

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*

    - The dataset has been used to conduct a descriptive analysis of the child care situation of working mothers in Ghana. We used the dataset to create graphs.

2. *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*
    - The repository containing all the files pertaining to the dataset is available at https://github.com/Pascal-304/dhs_analysis
3. *What (other) tasks could the dataset be used for?*
    - None since the dataset pertained to the childcare situation of working mothers in Ghana only at one time point(1998), so it cannot be used to generalize in other countries or other time points.
4. *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?*
    - N/A
5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*
    - N/A
6. *Any other comments?*
    - None

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*
    - Yes, the dataset is publicly available on the internet.
2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*
    - The dataset can be obtained by running the r script "01-gather_data.R" available at the link: https://github.com/Pascal-304/dhs_analysis/blob/main/scripts/01-gather_data.R
3. *Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/ or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*
    - No
4. *Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.*
    - Unknown to the authors of the dataset
5. *Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.*
    - Unknown to the authors of the dataset

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*

- Pascal Lee Slew and Yunkyung Park

2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
    - If anyone want to reach out to us about the dataset, you can contact us at pascal.leeslew@mail.utoronto.ca or clara.park@mail.utoronto.ca
3. *Is there an erratum? If so, please provide a link or other access point.*
    - No
4. *Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)?*
    - If there is any error made in the creation of the dataset, the dataset will be updated by either Yunkyung Park or Pascal Lee Slew. Any updates will be communicated as a note in the Readme file at the link below: https://github.com/Pascal-304/dhs_analysis/blob/main/README.md
5. *If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.*
    - N/A
6. *Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.*
    - The dataset has just been created. If there are any updates, older versions will still be kept for consistency.
7. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.*
    - Others may do so and should contact the original authors about incorporating fixes/extensions.
8. *Any other comments?*

- None

References

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

Gagolewski, Marek. 2021. *Stringi: Fast and Portable Character String Processing in r*. https://stringi.gagolewski.com/.

Gammarano, Rosina. 2020 [Online]. "International Day of Families: How Marital Status Shapes Labour Market Outcomes." *International Labour Organization*. https://ilostat.ilo.org/international-day-of-families-how-marital-status-shapes-labour-market-outcomes/.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92.

Ghana Statistical Service, and Macro International Inc. 1999. "Ghana Demographic and Health Survey 1998." Calverton, Maryland: GSS; MI.

Iannone, Richard, and Mauricio Vargas. 2022. *Pointblank: Data Validation and Organization of Metadata for Local and Remote Tables*. https://CRAN.R-project.org/package=pointblank.

McCartney, Kathleen. 2015. *What Do We Know about the Effects of Early Child Care?* https://www.purdue.edu/hhs/hdfs/fii/wp-content/uploads/2015/07/s_mifis01c03.pdf.

Ooms, Jeroen. 2022. *Pdftools: Text Extraction, Rendering and Converting of PDF Documents*.

Pebesma, Edzer. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal* 10 (1): 439–46. https://doi.org/10.32614/RJ-2018-009.

Pedersen, Thomas Lin. 2021. *Ggforce: Accelerating 'Ggplot2'*.

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

Samman, Emma, and Joan Lombardi. 2019. "Childcare and Working Families: New Opportunity or Missing Link?" UNICEF.

Sasu, Doris Dokua. 2022. *Rural and Urban Population of Ghana from 1990 to 2020*. Statista. https://www.statista.com/statistics/1170371/distribution-of-rural-and-urban-population-in-ghana/.

South, Andy. 2017. *Rnaturalearth: World Map Data from Natural Earth*. https://github.com/ropenscilabs/rnaturalearth.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2019. *Stringr: Simple, Consistent Wrappers for Common String Operations*.

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*.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

———. 2016. *Bookdown: Authoring Books and Technical Documents with R Markdown*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/bookdown.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*.