

# Data Visualization and Analysis of Thailand Women's Contraceptive Methods Preference in 1987

A Data Analysis Study Based on TDHS, 1987

Yichen Duo & Jinli Xu

2022-04-03

## Abstract

This article uses the data from the 1987 Thailand Demographic and Health Survey (hereinafter referred to as TDHS), and extracts Table 4.8–Percent distribution of currently married woman aged 15-44 according to the contraceptive method currently used, by selected characteristics on page 75. , classify and integrate the data in the table through statistical methods, aiming to restore the basis of the research content in the paper, and further combine the social background and cultural environment at that time to analyze the research results and draw conclusions.

## Introduction

Thailand Demographic and Health Survey, 1987 (Chayovan and Knodel 1988) is the first time in the history that the West has conducted a comprehensive health questionnaire in Thailand. The questionnaire itself mainly focuses on health, fertility, family choices, etc., and aims to consider and analyze the health and living conditions of Thai people around the 1980s. In the 1980s, modern medicine and drug research were very different from what we are living in today, so this statistical report has a strong historical reference in our view today. By studying and analyzing the health choices of people in less developed Asian countries decades ago, we can further understand the impact of medical and pharmaceutical advances on human and social changes.

In our study, we specifically extracted TDHS table at page 75 in TDHS(Chayovan and Knodel 1988) (table 4.8) about women's contraceptive measures in that year. Contraception and family planning are unavoidable problems faced by every family, so research on this aspect is of particular social value. We believe that research on this, combined with the social and cultural background at that time, will help us restore some of the social environment and people's living conditions in Thailand at that time. On the other hand, the gap between some data and today's contraceptive habits can also reflect changes brought about by technological advances; for example, advances in materials science have brought thinner artificial latex, making condoms too popular in the past two decades. Ascension, and in the 1980s, birth control pills were more popular than condoms. The research would be conducted using R-code (R Core Team 2020) as a R markdown project and the respected repository will be linked with Github. <sup>1</sup>

Before that, we firstly conducted a simulation analysis of the overall data through the simulator by using tidyverse (Wickham et al. 2019) to draw preliminary conclusions and research. We believe that this method will help us to obtain more helpful and constructive conclusions in combination with our subsequent data operations. In summary, we take a portion of the graphs from the TDHS 1987 study, statistically digitize the graphs, and analyze them on top of that. In this process, we will analyze this data in detail, and combine it with other parts of TDHS 1987, aiming to analyze in detail the social state of Thailand at that time and infer the reasons for the selection of data. The entire research process will be carried out in a reproducible manner and linked in GitHub for further research by interested individuals or groups.

---

<sup>1</sup> Github: <https://github.com/jinlixuokok/STA304-paper4.git>

At the end of the introduction, here we give a general introduction to the different contraceptive methods analyzed in this article. Birth control pills and injections are a common cyclic hormonal drug, first invented around 1950, as a widely used low-risk, high-efficiency drug. After being introduced into several Asian countries in the 1980s, it has created a lot of profits for pharmaceutical companies and local pharmacies. IUD is an intrauterine device, a removable semi-permanent intrauterine implant that releases copper ions to interfere with sperm implantation for contraception; this method is highly reliable, but it is prone to a series of side effects in the medical environment of the last century. Condoms are a common, primitive, and effective contraceptive method, but in the last century when material science was not yet developed, condoms were usually made of rubber, which had a poor user experience and was prone to allergic reactions. Sterilization is a permanent, almost irreversible operation that permanently renders a person infertile by ligating the fallopian tubes or vas deferens.

## Data

The data in our study were themselves produced in the TDHS(Chayovan and Knodel 1988) section of the specific study on contraceptive method choice among married Thai women aged 15-44. The data itself is broken down into several categories for research, including residential area (urban or rural), number of children, residential address, education level, and religious beliefs (mainly Buddhist and Islamic).

Intuitive data analysis can reflect that most of the respondents live in the countryside, have a lower level of education, and the main local religion is Buddhism. These facts honestly reflect some of the local social conditions in Thailand in the 1980s. As a less developed Asian country where Buddhism is prevalent, Thailand's native education in the 1980s was low per capita, and as an agricultural country, the proportion of the urban population was small.

This basic data analysis will help us to deeply understand the meaning of the data itself and will be further analyzed in the following sections. Besides that, we then need to understand the different methods of listing contraceptive methods. Pills are generally new in the 1980s, while extremely effective for contraception, it was only introduced into Thailand around 1980s as many pharmacy companies were trying to profit off it. As a very primitive but effective means of contraception, condoms were not very popular around 1980 due to material limitations. As a once-for-all contraceptive method, sterilization is not often used due to its irreversibility in the time and the limitation of some religious beliefs. In general, we will conduct specific and detailed analysis in the following data analysis and results section and insert the charts we have built to help data visualization and understanding.

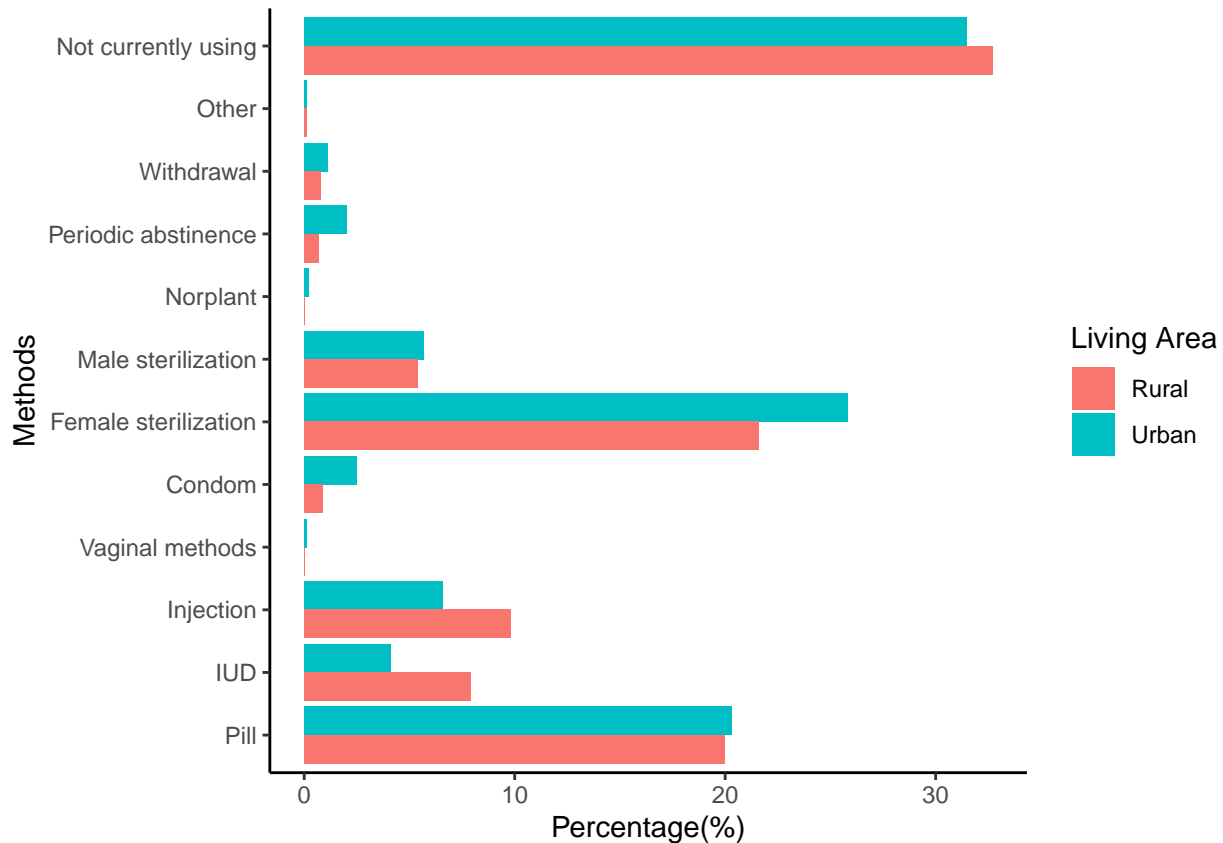
## Methodology

Our method first simulates the random distribution of the data through the simulator by using tidyverse(Wickham et al. 2019), and initially analyzes the data set through this method. Next, we created an RMD script, 01-gather\_data, to collect preliminary information on charts 4-8 by using tools of pdftools(Ooms 2022), janitor(Firke 2021); Function of Stringi(Gagolewski 2021) was also used to further manually adjust the read data from THDS. In this process, we have carried out preliminary sorting and manual screening of the text conversion itself to ensure the readability and recognition of the data conversion. The next step is to create an RMD script for clean and prepare data, 02-clean\_and\_prepare\_data. In this step, we used function of readr(Wickham, Hester, and Bryan 2022) to read the raw data generated by the last step; tidyverse(Wickham et al. 2019) and blankplank(Iannone and Vargas 2022) were also used in this part. The main function of this script is to collect and further organize the data identified in the first step to create a chart.

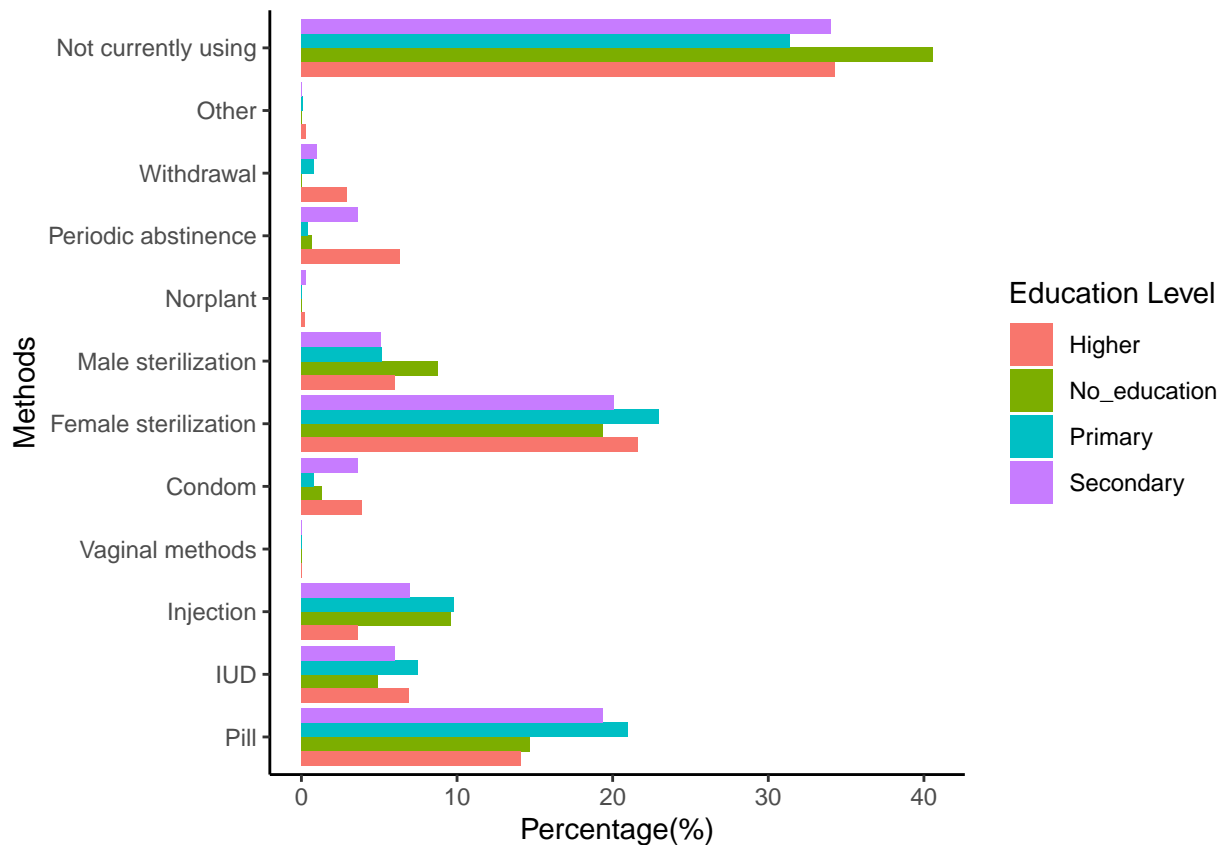
After that, we produced a raw data called raw\_data.csv, which has been sorted, cleaned, and aggregated to fully reflect the data set in Figures 4-8 in TDHS, which we will use for further research. and analysis and use this dataset for data visualization. In this step, we used an important tool of Reshape(Wickham 2007) to shift the columns of the data into rows for graph generating step.

## Result and Anaylize

After cleaning and summarizing the data in to raw\_data.csv, we then proceed to analyze the data by grouping them into different graphs, depending on their scopes of focus. Several conclusions on graphs' topics would be drawn in this section.



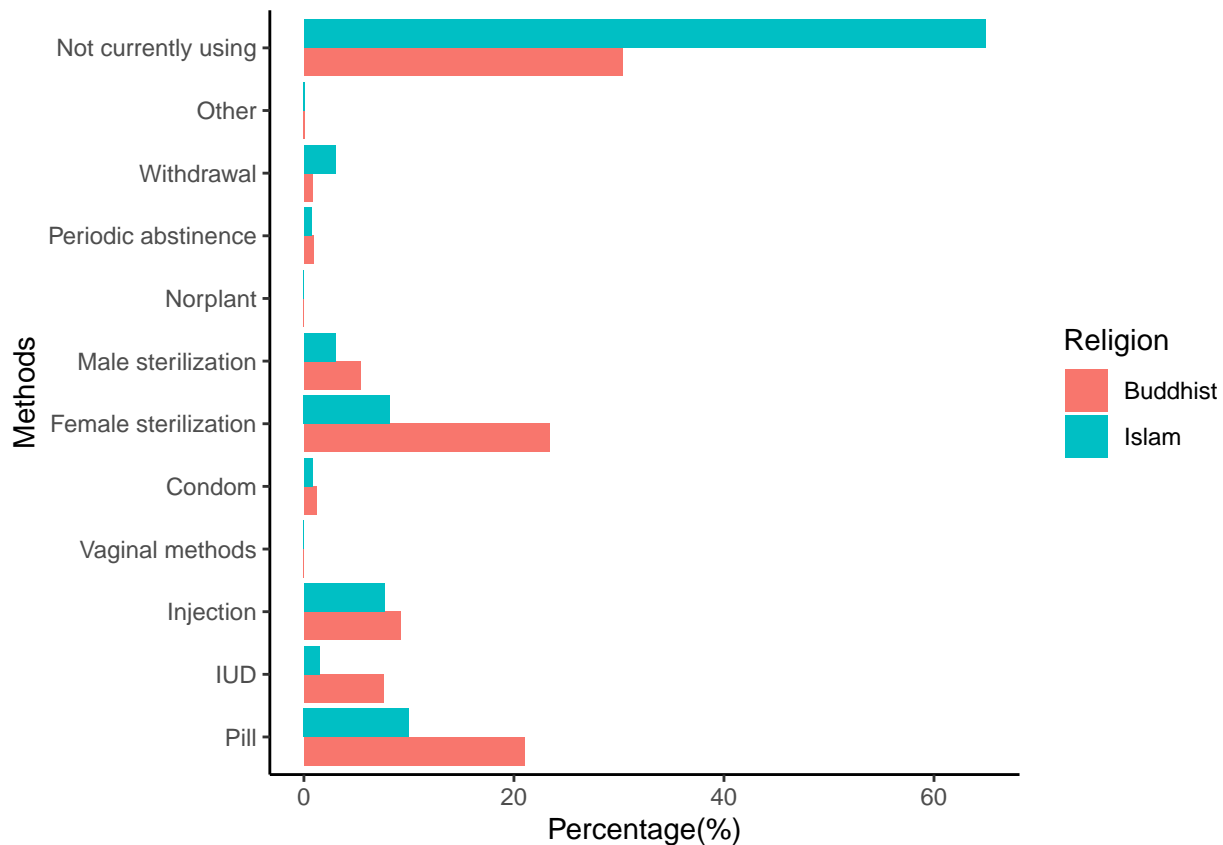
This graph is visualizing on the differences of percentage of contraceptive methods used by women in Thailand between people living in urban and rural area. While most of the population in Thailand lived in Rural area, as we discussed before in the Data section, the preference of choices between different contraceptive methods used is significant and visible, and shown clearly in the graph above. For people who are not using any contraceptive method and using withdraw, the difference is minor; this is believed to mainly caused by the fact that withdraw method requires minimum resources—it is naturally taught and requires no material assistance. While interestingly, condoms are much more widely used in urban area than it is used in the rural area, and this is caused by that fact that in a less developed country like Thailand, acquiring condoms could have been difficult in its vast rural area. This assumption is further supported by the fact that semi-permanent method like Injection and IUD was more popular in rural area, as those method would be easily performed in a local clinic or pharmacy and requires no future attendance.



This graph focuses on preference of contraceptive methods used by different group of women, grouping by their education level, from no-education to higher than secondary. We first need to understand in Thailand around 1980s, female rights were not widely recognized, and educations were scarce resources that hardly distributed to female individuals. In general, it is rare to find an educated woman in Thailand back then.

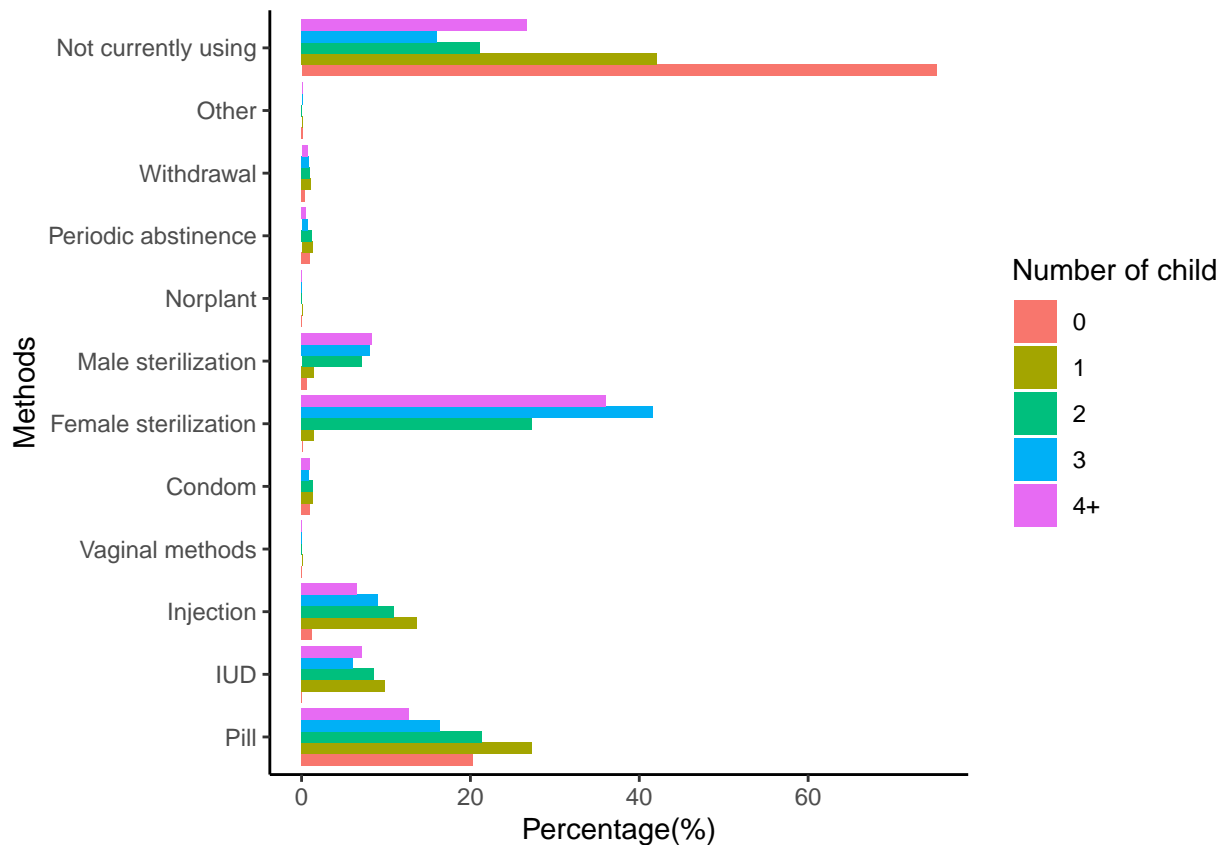
Back to the graph itself, we see a vast difference in the contraceptive method of Periodic Abstinence, a method required the noting and understanding of female repopulation system and menstrual period, used between educated and uneducated individuals. This is mainly caused by the complicated nature of this method, and it proves that lower level of education in Thailand back then provided only limited knowledge of female repopulation system.

Interestingly, Pills were popular among females with intermediate to high level of education, while it is not popular among people with lower or higher education. This could be caused by that people with low education would hardly receive any knowledge about contraceptive pills while highly educated individuals received certain knowledge that contraceptive pills back then could be somewhat harmful and has some serious side effects. Contraceptive pills were massively introduced into Thailand by pharmacies and local clinics as a significant way of gaining profits back then; though proven effective, its massive spread in Thailand and some other less developed countries in Asia like India did cause some side effects.



This image focuses primarily on visualizing differences in contraceptive choices among women of different religious beliefs. First of all, we need to understand that Thailand is a monarchical country where religion has great influence in society. Buddhism is of great importance in Thailand. In Buddhist scriptures, abortion is an extremely serious crime, equivalent to killing. So, for Buddhists, the majority of Thailand's population, abortion is not an acceptable option; in this case, contraception is especially important. On the other hand, Thailand also has many Muslim groups, and in more traditional Muslim teachings, women bear a very important or inescapable reproductive responsibility. In traditional Muslim families, women have a low status and usually have no voice. As a result, Muslim women are generally less likely to use effective contraception.

In this picture we can see multiple examples that support the above point. First of all, the rate of Muslim women not using any contraceptives is much higher than that of Buddhists. Among Muslim women who use contraceptives, the withdrawal method, which has a higher risk of pregnancy, is widely used due to women's lack of voice. On the other hand, among Buddhist women, many women use almost irreversible sterilization methods to eliminate the risk of pregnancy; at the same time, contraceptive methods with low pregnancy risks, such as the pill and injection, are also widely used.



This graph simply and intuitively shows the relationship between contraceptive use and the number of children. We hope to use this picture to roughly analyze the fertility desire of Thai women around 1980, or the number of children they expect. As shown, most women who did not use any contraception were childless; the majority of women who used effective contraception had already had three children. From this, we can conclude that the number of children expected by Thai women at that time was three. Interestingly, women with four or more children were also more likely to avoid contraception. We believe this was due to the fact that Thailand was a de-industrialized agricultural country at the time, and the rural population was much larger than the urban population; in this case, the population was an important labor resource, and for most rural populations, households More children means more family productivity in the future, so many women would like to have more than three offspring.

## Discussion

In this article, we visualize and analyze the contraceptive-choice habits of married women aged 15-44 in Thailand during the 1980s by extracting some data from TDHS 1987(Chayovan and Knodel 1988). We first extracted Table 4.8 from the TDHS report, first used the simulator method to simulate the data, then converted the graphs into processable data and transformed it into a visual data graph `raw_data.cvs` by cleaning and sorting. By classifying and comparing the contraceptive habits of different groups of women, we have drawn several valid conclusions: Thailand in the 1980s, as a sub-developed country, was a sub-developed and agricultural country with a lot of rural land; Thailand at that time was based on Buddhism. is the predominant religion and the majority of the population has a low level of education. Under this circumstance, with the entry and influence of the advanced market economy in the West, a large number of new private clinics and pharmacies were born in Thailand, which promoted the local popularity of contraceptives, injections, IUD and other periodic contraceptive methods. As a sub-developed country, Thailand's local cultural environment and religious beliefs determine that contraception is a demand that cannot be ignored in the social environment of the country. Therefore, the promotion of contraceptive methods must have had a certain impact on the local economy at that time. On the other hand, as an agricultural country, some women also consider producing a large number of offspring to supplement the family's labor force.

Analysis and research on contraceptive measures, infant mortality and other data can effectively analyze the local cultural environment and economic situation from the side. However, the report only conducted analysis and research on contraceptive measures, so the evidence for some of the views in the article is based on common-sense inferences. Such inferences are often incomplete under more rigorous academic discussions. We believe that the research on the local human environment and economic situation will be more rigorous and convincing if the research on infant mortality and women's marriage rate is added.

# Appendix

Extract of the questions 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 help we have a better further plan on the demographic and health in Thailand. This is the first time to have a large scale surveys about health in Thailand.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
  - Institute of Population Studies/Thailand and Institute for Resource Development
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.*
  - The dataset, or TDHS itself, was funded by the Population Council of US
4. *Any other comments?*
  - N/A

## 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 instances that comprise the dataset were women from age 15-44 in Thailand who took the TDHS survey about their current status.
2. *How many instances are there in total (of each type, if appropriate)?*
  - For the dataset we created based on table 4.8 in TDHS 1987, a total of 5561 instances are there in total.
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).*
  - N/A
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.*
  - The data was unprocessed text of a data table.
5. *Is there a label or target associated with each instance? If so, please provide a description.*
  - Yes, the dataset labeled instances by grouping collected data based on several categories.
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.*
  - No
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, the relationships between individual instances are not explicit
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.*
  - Yes, the text of dataset is somewhat blurry and requires manual adjustment when cleaning.
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.*

- Yes, it is self-contained
- 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
- 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.*
  - No
- 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.*
  - No
- 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?*
  - N/A

## 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.*
  - The data was acquired through survey conducted on behalf of TDHS, 1987.
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?*
  - Surveys.
3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*
  - N/A
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)?*
  - Population Council of United States, a NGO.
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.*
  - Yes
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.*
  - Yes, the research and original data collecting process was based on complete voluntary and used for research purposes only; all answers to the survey were anonymous.
7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other*

*sources (for example, websites)?*

- No
- 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 collection was by surveys and notices were delivered with the survey.
- 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.*
  - Yes, all individuals who answered the survey consent to the collection.
- 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).*
  - No
- 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.*
  - No
- 12. *Any other comments?*
  - N/A

### 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 paper is conducted by preprocessing and cleaning the data from TDHS report by using text reading method and several data cleaning script.
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.
3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*
  - Yes. (Add references here.)
4. *Any other comments?*
  - N/A

### Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*
  - The dataset was originally part of TDHS, 1987 research and was used for determine the local social and population healthy status.
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.*
  - Yes. (Add github repository here)
3. *What (other) tasks could the dataset be used for?*
  - The dataset can be used as a reference as local economic and education status analysis project.
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?*
  - The composition of this dataset is very straight forward and the cleaned raw data is frankly labeled; generally the data processed should provide easier access to the data in future.
5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*

- No
6. *Any other comments?*
- N/A

## 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.*
  - No.
2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*
  - N/A
3. *When will the dataset be distributed?*
  - N/A
4. *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.*
  - N/A
5. *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.*
  - N/A
6. *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.*
  - N/A
7. *Any other comments?*
  - N/A

## Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
  - The original dataset requires no maintaince and the processed data of table 4.8 would be aviliable on Github following the linked repository.
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
  - jinli.xu@mail.utoronto.ca; mark.duo@mail.utoronto.ca
3. *Is there an erratum? If so, please provide a link or other access point.*
  - N/A
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)?*
  - N/A
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.*
  - N/A
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.*

- N/A
8. *Any other comments?*
- N/A

## References

- Chayovan, Peerasit Kamnuansilpa, Napaporn, and John Knodel. 1988. *Thailand DHS, 1987-Final Report (english)*. Maryland, USA: Institute of Population Studies/Thailand; Institute for Resource Development. <http://dhsprogram.com/pubs/pdf/FR37/FR37.pdf>.
- 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/>.
- 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.
- Iannone, Richard, and Mauricio Vargas. 2022. *Pointblank: Data Validation and Organization of Metadata for Local and Remote Tables*.
- Ooms, Jeroen. 2022. *Pdftools: Text Extraction, Rendering and Converting of PDF Documents*.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley. 2007. “Reshaping Data with the reshape Package.” *Journal of Statistical Software* 21 (12): 1–20. <http://www.jstatsoft.org/v21/i12/>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2022. *Readr: Read Rectangular Text Data*.