Education level in the Philippines

Yingxuan Shi, Jingxian Zhai, Wantong Qiu

Abstract

When social problems are being paid more and more attention, education is an indispensable topic, because education affects the development of the whole society. Then this report is to study and analyze the situation of education level through the National Demographic and Health Survey 1998 in the Philippines. We would like to know what will influence a country's education level? Based on the data from the National Demographic and Health Survey 1998 in the Philippines, we find out that the different age groups, different genders, and the differences of regions can influence the education level, through this report, we can have a more comprehensive understanding of the social issue of education level.

Contents

1 Introduction

With the development of society, there is an increasing demand for education, because education can help people improve their speaking skills, develop critical thinking, promote gender equality, and improve their cognitive system after learning knowledge((“Why Is Education Important and How Does It Affect One’s Future?” n.d.)). The education level of the current society is not only for individuals, but also to show the development level of a society and a country. Therefore, we selected the education level of the Philippines among many countries for further analysis. To be specific, we selected some data about education levels from National Demographic and Health Survey 1998 in the Philippines for the following research and analysis.[1]

In this paper, the main idea is about analyzing education levels in the Philippines through the data on age, gender, and region. The first is to observe whether people of each age group participate in their education level relative to what they need to achieve through age data. Because not every age group can receive the corresponding educational conditions in time. Then there is the data about gender, hoping to observe whether men and women receive equal education in this country through the data of gender. As we mentioned above, education is not only about academics, but also about promoting gender equality. Therefore, we hope to analyze whether the education rate is balanced for men and women through gender data. On the other hand, education levels in urban and rural areas. From this data, we can analyze whether there is a big gap between urban and rural education levels.

In the National Demographic and Health Survey 1998, we selected data about education levels for analysis. The first step is to import the data from the Survey. Then it is data cleaning, specifically, the transformation of some wrong typed of data, filtering out missing values and some useless data. Then the data of age, gender, and residence are plotted and analyzed by graphs and tables.In the Results section, we find the relationship between age and respondents’ education levl, schooling time; regions and respondents’ education level; gender and respondents’ education level. In the discussion section, we discuss whether our results match the real cases, and then think about limitations and future.

2 Data section

2.1 Data Cleaning

To gain a better understanding of the Philippines Demographic and Health Survey(DHS) on education, I utilized the data from Philippines DHS 1998, final report in PDF format. The data is collected and provided by the United States Agency for International Development (USAID). The data has been sourced in a reproducible way, with a script for downloading and obtaining the relevant datasets.In this dataset, the raw data includes 8 variables so we cleaned and extracted the important data to start our analysis. In the analysis, we will useR statistical language (R Core Team 2020), tidyverse packages (Wickham et al. 2019), pdftools (Ooms 2022), stringi,tibble (Müller and Wickham 2021), devtools, dplyr (Wickham 2021), fmsb(Nakazawa 2022), janitor(Firke 2021), formattable(Ren and Russell 2021), kableExtra(Zhu 2021).

Firstly, we selected all variables that we will use in analysis from the raw data, they are related to people’s education level, age group, gender and region in the Philippines. Because of the survey using a whole table to describe age groups, regions and gender, we will create four new datasets to start our analysis. The first two datasets are about the relationship between male or female respondents’ age and their education levels. The next two datasets are about relationships between male or female respondents’ regions and their education levels. Some of the variables that we will use in our analysis are background characteristics, elementary education, high school, college or higher, number of males, number of females and median number of years of schooling. Then we use pintblank to check the validation. Then, we will use the select function to delete missing values and total because they are meaningless in our analysis.

[1] https://github.com/JingT13/STA304__paper__4.git

2.2 Methodology

The National Demographic and Health Survey 1998 is about collecting, analyzing, and disseminating population data on fertility, family planning, and maternal and child health in the Philippines. The survey was picked sub-sample from the Integrated Survey of Households (ISH) of the NSO. Data collection is mainly carried out in the form of a questionnaire survey among people in each region. Data for each region, age, and gender are included, and the dataset contains all possible instances.

2.3 Variables

We will introduce each variable respectively with plots and tables that are created by kableExtra[(Zhu 2021) and ggplot2. Firstly, three variables of education levels which are elementary education, high school and college are most important in our analysis. In Figure 1, Figure 2 and Figure 3, red bars represent female and blue bars represent male. From Figure 1, we can see the percentage of male and female respondents who just have elementary education at different age groups. From Figure 2, we can see the percentage of male and female respondents who attend high school at different age groups. However, the age group from 6 to 9 has no value because there are no 6-9 years old respondents in high school. From Figure 3, we can see the percentage of male and female respondents who attend college or higher education. Age group from 6 to 14 has no value. Moreover, Figure 4 and Figure 5 both introduce the median number of years of schooling for male and female. At last, the map describes another important variable, regions. The Philippines is consist of Metro Manila, Cordillera Admin, Ilocos, Cagayan Valley, C.Luzon, S.Tagalog, Bicol, W.Visayas, C.Visayas, E.Visayas, W.Mindanao, N.Mindanao, S.Mindanao, C.Mindanao.

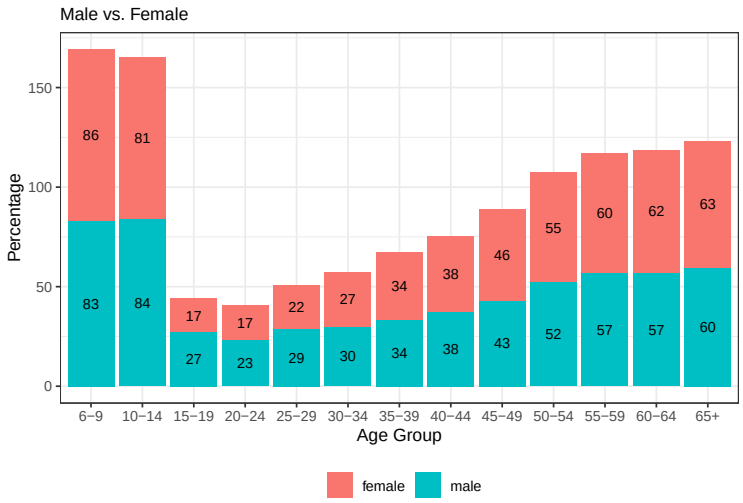

Figure 1: Percent distribution of Elementary Education

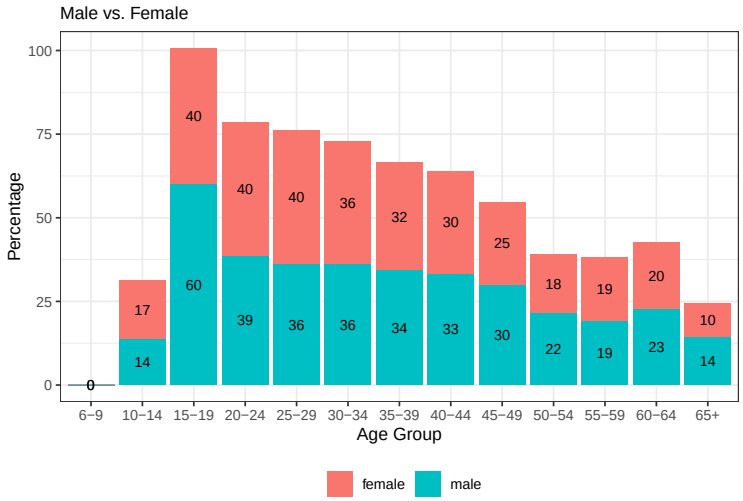

Figure 2: Percent distribution of High school

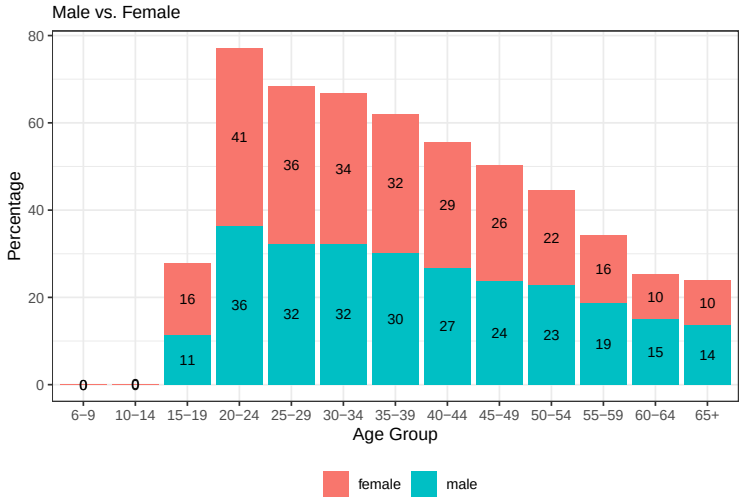

Figure 3: Percent distribution of College or higher

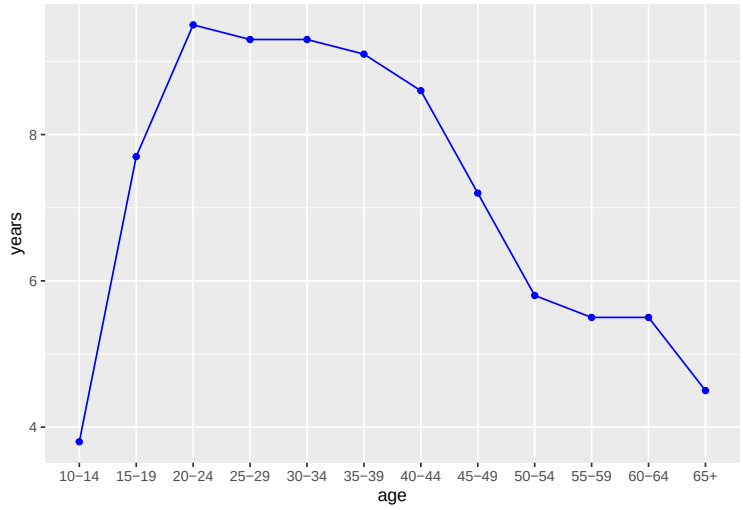

Figure 4: Male respondents' schooling years and age groups

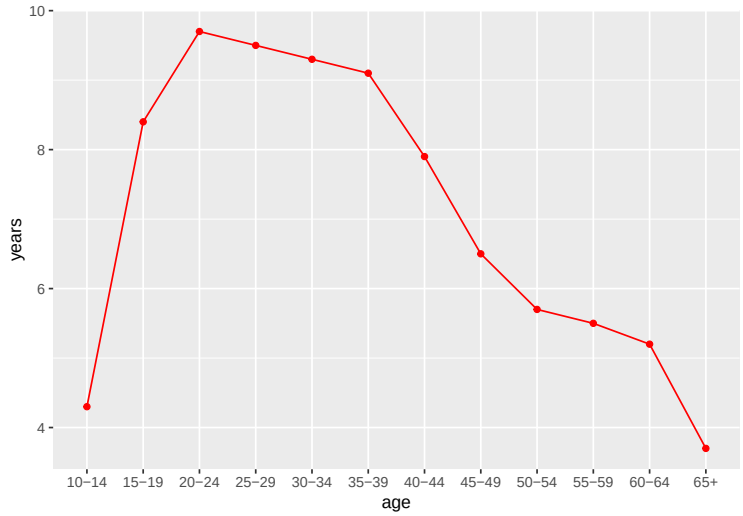

Figure 5: Male respondents' schooling years and age groups

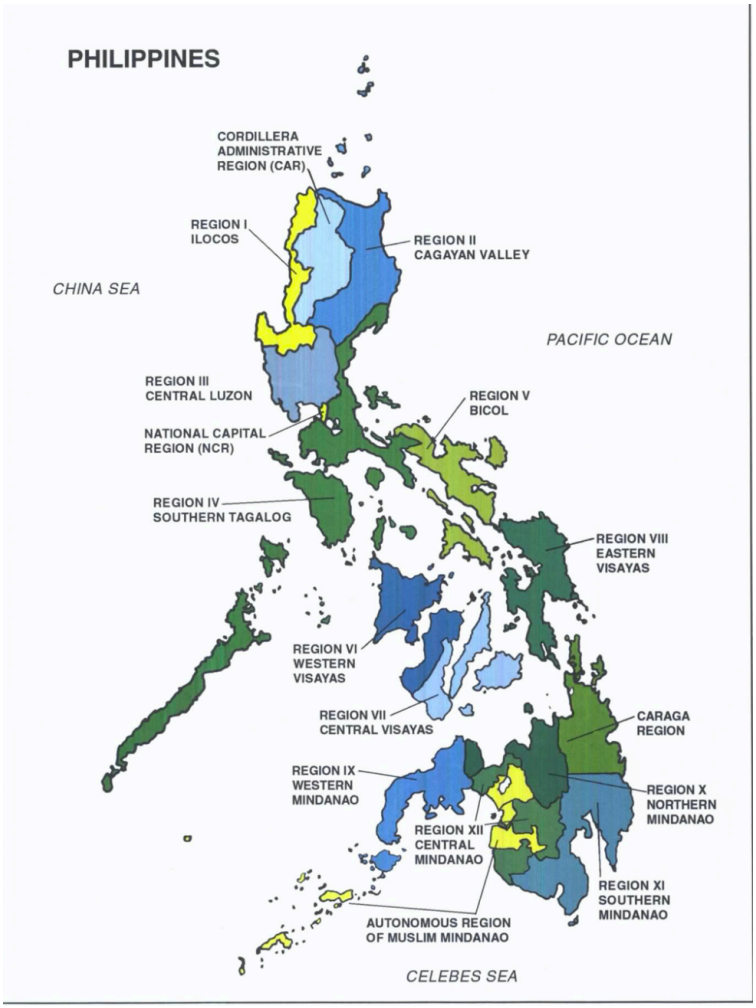

Figure 6: The map of Philippines

3 Result Section

3.1 Age

From line graphs at the first row, a blue line which represents male and a red line which represents female respectively. They both describe the relationship between the number of respondents who are educated and their age. Generally, two lines show a decreasing slope from 15-24 to 65+. Most number of respondents who are education attended is at 15-24 age group(approximate 5500) and least number of respondents who are education attended is at 65+ age group(approximate 1000). With increasing age, the number of respondents who are educated decreases sharply. From line graphs at the second row, they describe the years of male and female respondents schooling. We can see the center at age from 25 to 44, which means respondents who are in the 25-44 age group experience the longest schooling time, longer than 9 years. 65+ years old respondents have the shortest schooling time, less than 4 years.

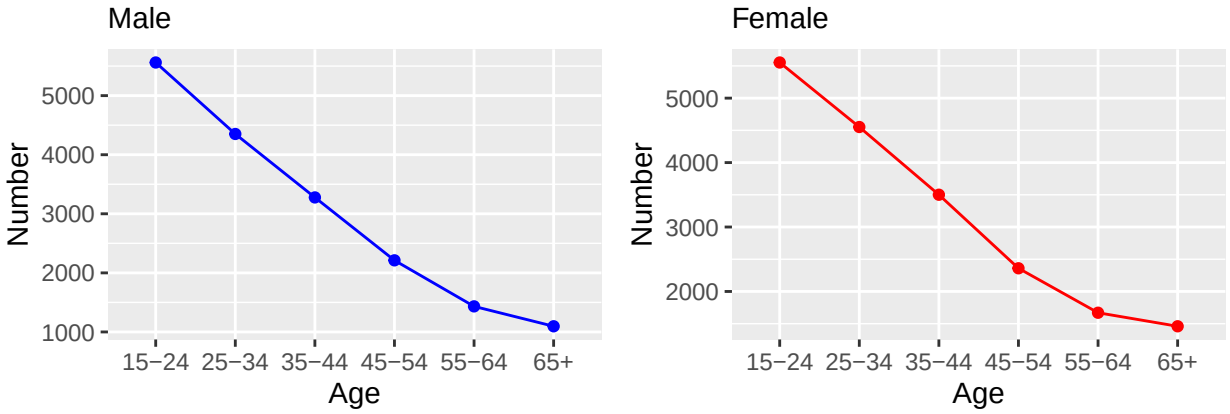

Figure 7: The number of respondents who have education in different age groups

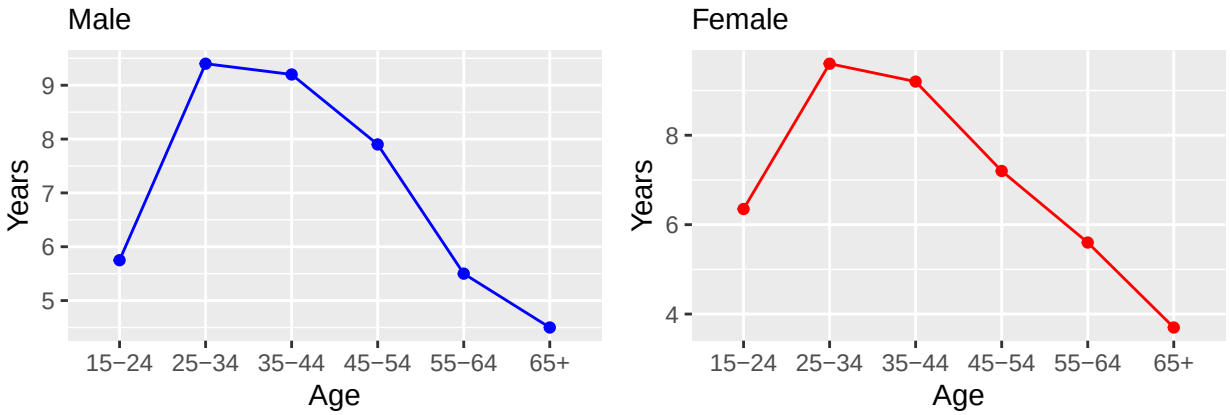

Figure 8: The median number of years of schooling in differnent age groups

3.2 Region

From table 1, we can see most respondents have college or higher education in the region of Metro Manila and their schooling time is the longest(9.3 years). Thus, Metro Manila has a strong educational force. Only ten percent of respondents have college or higher education in the region of E.Visayas and most respondents just attend elementary education with the shortest schooling time(4.5 years). Respondents who live in regions of Mindanao and Visayas have weak educational forces. From radar charts, black lines mean rural education and red lines mean urban education. We can see the number of rural respondents who just have elementary is more than the number of urban respondents significantly because the black point is above the red point at the Elementary. Furthermore, the number of urban respondents who have college or higher education is more than the number of rural respondents.

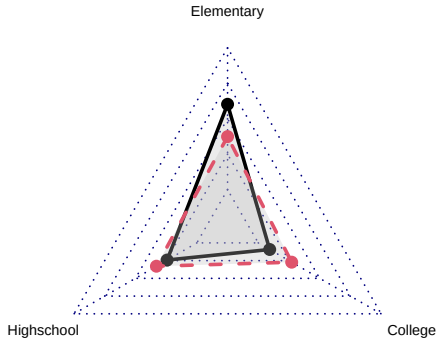

Figure 9: The percent of male's education level in Rural and Urban

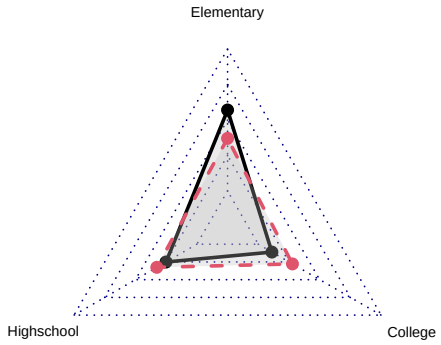

Figure 10: The percent of female's education level in Rural and Urban

Table 1: The percentage of education level of male and female in different regions

	Elementary	HighSchool	College	Years
Metro Manila	27.6(29.2)	35.0(36.8)	35.7(32.5)	9.4(9.2)
Cordillera Admin	51.9(48.2)	24.6(25.7)	16.5(16.7)	5.3(5.5)
Ilocos	44.6(46.2)	34.2(30.7)	16.6(18.9)	6.3(6.1)
Cagayan Valley	54.2(54.1)	25.9(23.9)	14.9(17.4)	5.5(5.6)
C.Luzon	45.1(47.1)	35.6(31.4)	16.5(19.1)	6.5(6.2)
S.Tagalog	47.2(47.5)	30.0(29.0)	19.7(20.6)	6.0(6.0)
Bicol	61.2(58.6)	25.2(24.8)	10.4(12.8)	5.4(5.5)
W.Visayas	55.8(50.2)	24.3(26.5)	16.4(19.4)	5.3(5.8)
C.Visayas	57.3(52.4)	24.0(28.0)	14.6(15.6)	5.3(5.6)
E.Visayas	65.4(61.7)	19.3(20.6)	9.0(10.7)	4.1(5.0)
W.Mindanao	54.1(49.6)	21.3(24.9)	15.5(15.3)	5.0(5.3)
N.Mindanao	56.1(51.9)	26.6(28.6)	13.4(15.9)	5.3(5.7)
S.Mindanao	50.9(46.0)	30.3(30.3)	14.4(18.9)	5.6(6.0)
C.Mindanao	55.4(45.4)	24.9(29.6)	15.1(18.4)	5.4(5.9)
Caraga	58.7(56.5)	27.6(26.1)	11.2(14.3)	5.2(5.5)

Note:

Male(Female) This is a comparison between male and female

3.3 Gender

From the above line graphs, we can make two comparisons between male and female. The first comparison is the number of respondents who are educated between male and female and the second comparison is the schooling time of respondents who are educated between male and female. We can see that the line graphs at the first row have a similar slope, it means that the number of male and females who are educated are almost equal at different age groups. Moreover, the line graphs at the second row have similar slopes as well, it means that male and female experience an equal schooling time. The radar charts from the last result show male and female respondents have a similar situation as well in rural or urban areas.

Figure () shows male respondents' education levels which are elementary, high school and college or higher. We can see elementary takes the biggest share(48.8%), most male respondents only experience elementary education. College or higher takes the smallest share(18.2%). Figure () shows female respondents' education level. It has a similar situation with male's, elementary education is the biggest share (46.6%) and college or higher is the smallest share(19.8).

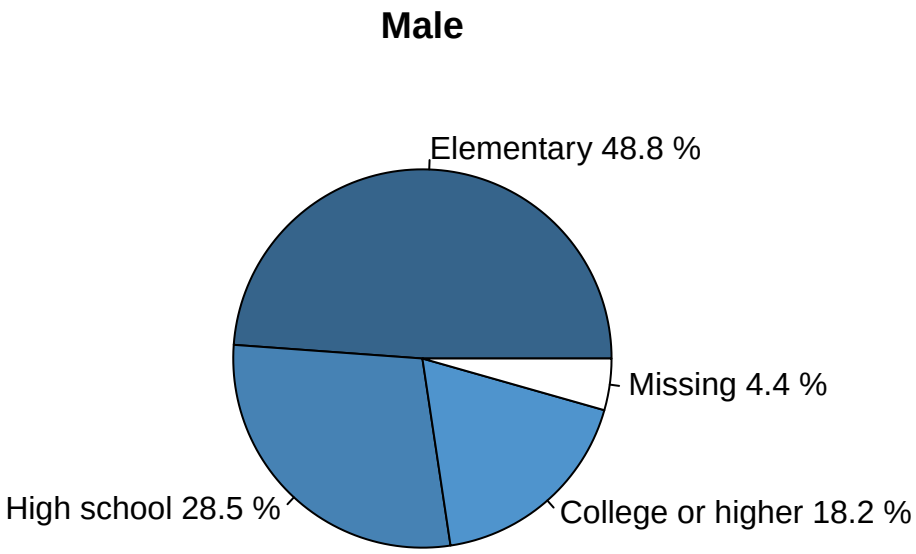

Figure 11: The share of different education levels

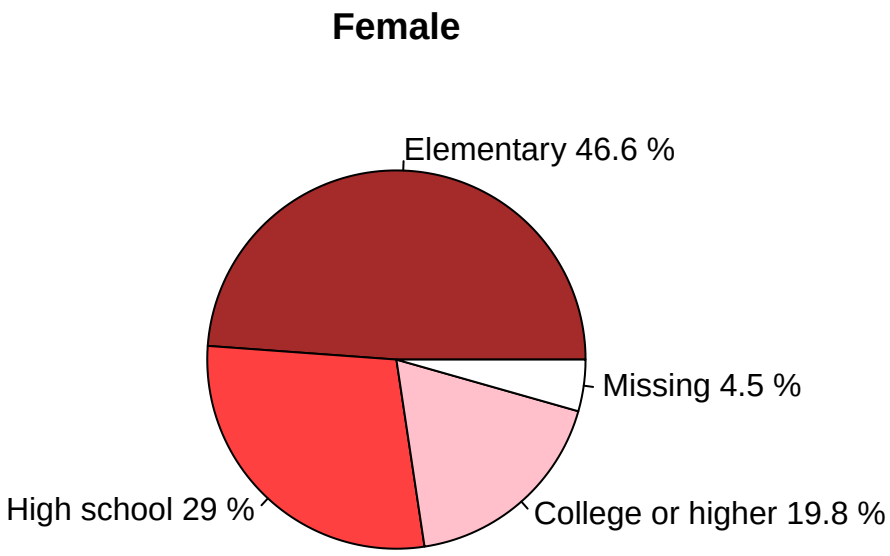

Figure 12: The share of different education levels

4 Discussion

The above analysis indicates how age, gender, and region impact people's education level in The Philippines. Start by analyzing the educational attainment of different age groups in the Philippines and the number of years they have been in education. After that, the education level in the Philippines is compared and analyzed through the proportion of males and females in the education level in graphs of different genders. Then it explores the differences in education levels between urban and rural men and women at different levels. The following paragraphs will talk about more specific data feedback on these three aspects and whether further discussion on educational level can be carried out in the future.

4.1 Age

The two figures in the first row show the analysis of the different numbers of people in different age groups in education. As we can see the age of 15-24 owns the highest number of educated people. And the 65+ group has the lowest number of educated people. And the two line groups in the second line show that people of different age groups have different years of education. It can be seen that people aged 25-34 have the highest number of years of education, and people aged 65+ have the least number of years of education. In fact, people's social activities are determined by the response to the development stage of our brain. This includes the different education levels of people of different ages in this paper. For example, the most educated people are between the ages of 15 to 24, because our brains complete brain development at the age of 25, and until then our brains are not fully developed, which means that we are not fully in control of and able to determine our social behavior (Changrong (2017)). which means people need to improve the way of thinking and make some contributions to the development of the brain through continuous learning at the age of 15 to 24. So in line graph above, the group of people aged 15 to 24 who are studying is the largest. Compared to the pre-colonial period, the education system in the Philippines changed significantly after 1990, from only vocational training for all children to the 1991 REPORT of the US Congressional Committee on Education (EDCOM), which recommended DECS be divided into three parts. That is, academic education to vocational education have a more perfect system. The second row of line graphs shows that the people aged 25-34 have the longest number of years of education, and the age group from 15 to 24 is still in school. The education system in the Philippines is constantly improving so that more people can receive a better education. In this paper, we could not make a more comprehensive analysis of the reasons for people to receive education at different ages and their choices through data. In the future analysis, we can further explore the factors of different age groups in the Philippines for their choice of study(Connecticut (n.d.)).

4.2 Region

The pie charts above showed the educational level of people in both urban and rural areas. To be specific, these bar charts all have a common feature that elementary accounts for more than 37% and the proportion is the largest, while college or higher accounts for the least. And then the proportion of elementary part in the rural is almost twice as high as that in the urban. The proportion of college or higher students in rural areas is nearly three times lower than that in urban areas.

There are many impacts that affect the education level in urban and rural areas.Since urban and rural are different from each other in terms of economic development, vocational demand, or education development system, they play an important role in urban and rural people's education levels.For example, some places in urban will attract more people with higher education through higher salaries and superior consumption facilities. In fact, another way of understanding is that such an incentive method in urban can make more people pursue more education ((“The Urban–Rural Education Gap: Do Cities Indeed Make Us Smarter?” n.d.)). Therefore, according to the pie chart, we can see why the proportion of college or higher in cities is so much higher than that in urban and rural areas. Because the speed of urban development and social needs continue to spur people. This is just one example, and there are many, many more reasons. The data in the paper only tells us the proportion of different education levels in urban and rural, so that we can explore how urban and rural have an impact on education levels in a more detailed way in the future

study. Further research can be carried out through the economic development, education system, and policy mentioned above to start the analysis.

4.3 Gender

In most developing countries, such as China and Thailand, men are more educated than women. In the Philippines, however, this data comparison of educational level is the exact opposite. In most cases, Filipino women are more educated than men. In the data section, we also just discussed the educational level of men and women in the Philippines, and it does confirm this(National Statistics Office 1999). Based on this information, we know that maybe these are related to the cost and benefit of education (Elizabeth M. King 1993). In terms of school construction, if it is in some village areas, it provides a "short distance". While female enrolment in schools increases by 3%, male enrolment increases by only 1%. Another reason is that schools have more female teachers than male teachers. In some traditional families, parents would expect their daughter's teacher to be a woman. So schools with a higher proportion of female teachers have many girls enrolled. This is one of the reasons why female enrolment rates are higher than males. ##Limitations and Future

Apendix

A Datasheet

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 aims to collect, analyze and disseminate population data on fertility, family planning and maternal and child health in the Philippines.
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 National Demographic and Health Survey (NDHS), alongside the National Statistics Office in collaboration with the Department of Health (DOH), the University of the Philippines Population Institute, and other concerned agencies in the Philippine government.
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 creation was funded by the Government of Philippine.
4. *Any other comments?*
    - TBD

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 represent the educational level of Philippine. The types are: sex, age groups and regions.

2. *How many instances are there in total (of each type, if appropriate)?*

- There are 32 instances

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 1998 NDHS sample is a sub-sample of the new master sample of the Integrated Survey of Households (ISH) of the NSO. In this data, every region; age group and sexes was included. We could say the dataset does contain all possible instances.

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.*

- In the raw data, the instance consists of 32 variables.

5. *Is there a label or target associated with each instance? If so, please provide a description.*

- None.

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.*

- There is no missing individual instances.

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.*

- There are no relationships between individual instances.

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.*

- There are no recommended data splits.

9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*

- There are no errors, sources of noise, or redundancies in the dataset.

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 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.*

- There is no confidential data, and the dataset is publicly available.

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.*

- Columns that might cause anxiety include: the education level in different regions are different, this may lead to some inferiority complex.

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.*

- The dataset entirely comprises differnet age groups and differnt sex (men and women) and different regions.

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.*

- It is not possible to identify individuals in any way.

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.*

- Sensitive columns may include but are not limited to: different education levels for men and women.

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.*

- The 1998 National Demographic and Health Survey (NDHS) was the seventh in a series of population surveys conducted every five years since 1968. Using the way of asking people to fill questionnaires.

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?*

- Manual human curation.

3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*

- Households were systematically sampled in urban areas to distribute the NDHS sample throughout the sampling area, while compact clustering was used in rural areas to facilitate field operations.

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 1998 NDHS was undertaken as part of the worldwide Demographic and Health Survey (DHS) program. "It was conducted by the National Statistics Office (NSO), in collaboration with the Department of Health (DOH).

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 data was collected in 1998.

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.*

- Ethical review processes were not conducted.

7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*

- We obtained the data via the Demographic and Health Surveys website: dhsprogram.com.

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.*

- The individuals voluntarily fill out the questionnaire.

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.*

- The individuals consented to the collection and use of their data. The exact language to which consent was granted is not available.

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).*

- The mechanism to revoke their consent was not provided.

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.*

- The analysis of the potential impact of the dataset and its use on data subjects has not conducted.

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.*

- The data was appearing as a table originally in the PDF report. We obtain the data information in this PDF table and convert it into the data frame in R by pdftools package for R.

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.*

- The raw data obtained from the PDF is saved in inputs/data/just_page_i.csv and inputs/data/just_page_i1.csv.

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

- R Software is avalaible at https://www.R-project.org/

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 not been used for other tasks yet.

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.*

- TBD

3. *What (other) tasks could the dataset be used for?*

- The dataset can be used for analyze the different education level status.

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 process of cleaning data is specific to only this table in the original PDF report. This is not suitable in other tables.

5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*

- The dataset is not suitable for any other purposes except the education level in the aspects of age group, sex and regions.

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.*

- No, this dataset if openly available.

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 will be distributed using Github.

3. *When will the dataset be distributed?*

- The dataset will be distributed in April 2022.

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.*

- The dataset will be released under the MIT license.

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.*

- There are no restrictions.

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.*

- None.

7. *Any other comments?*

- None.

Maintenance

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

- Jingxian Zhsai, Yingxuan Shi and Wantong Qiu.

2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*

- This can be contacted by Github.

3. *Is there an erratum? If so, please provide a link or other access point.*

- There is no erratum available.

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)?*

- No, the dataset will not be updated.

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.*

- This dataset was collected by the quastionnaires which were filled by people in Phillipine in 1998. There are no applicable limits.

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 older versions of the dataset are not hosted. The dataset somsumers could be able to check the dataset by github.

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.*

- None.

8. *Any other comments?*

- None.

Reference

Changrong, Liu. 2017. "The Human Brain Can Only Develop Completely at the Age of 25." *China Youth Daily.*

Connecticut, Easton. n.d. "History of Education in the Philippines." *Easton Country Day School.*

Elizabeth M. King, M. Anne Hill. 1993. "Women's Education in Developing Countries: Barriers, Benefits, and Policies."

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://CRAN.R-project.org/package=janitor.

Müller, Kirill, and Hadley Wickham. 2021. *Tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

Nakazawa, Minato. 2022. *Fmsb: Functions for Medical Statistics Book with Some Demographic Data*. https://CRAN.R-project.org/package=fmsb.

National Statistics Office, Department of Health. 1999. "National Demographic and Health Survey."

Ooms, Jeroen. 2022. *Pdftools: Text Extraction, Rendering and Converting of Pdf Documents*. https://CRAN.R-project.org/package=pdftools.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ren, Kun, and Kenton Russell. 2021. *Formattable: Create 'Formattable' Data Structures*. https://CRAN.R-project.org/package=formattable.

"The Urban–Rural Education Gap: Do Cities Indeed Make Us Smarter?" n.d.

"Why Is Education Important and How Does It Affect One's Future?" n.d.

Wickham, Hadley. 2021. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

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.

Zhu, Hao. 2021. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.