Assessing the knowledge of AIDS amongst Women in 1997 Vietnam*

Jacob Yoke Hong Si

25 April 2022

Abstract

In what areas of women's and children's health and living conditions can be improved? Utilizing the data from the 1998-1999 India National Family Health Survey provided by the Demographic and Health Survey (DHS) program, we look to depict the demographics of Indian Women and Children in different states of India. third sentence: specify the headline result; and a fourth sentence about implications. We found that most families are small, well-off and in good mental health, and that immigrants make up about one fifth of the population, with a propensity to settle in larger cities and urban centers. We also discuss possible factors that may have influenced this family image, including Canada's investment in family planning and immigration levels planning.

Keywords: AIDS, vietnamese women, source, age group, regions, levels of education

1 Introduction

The Demographic and Health Surveys (DHS) is a nationwide household survey that provides data for a myriad of evaluation indicators with respect to health, nutrition and population. A few examples of topics include education, mortality rate and health conditions. In addition to the topics mentioned, each survey also collects comprehensive socio-demographic data including the respondents' age, marital status, residential area, etc.

The DHS program was established in 1984 and has provided technical assistance to over 300 demographic and health surveys in more than 90 countries with its main objective to enhance and develop data collection and utilization by host countries for monitoring and evaluating demographic indicators in order to devise high quality policy development decisions.

In these regards, the DHS program has been majorly successful with the data collected often serving as crucial assistance to the government's policy making to help improve the well-being of the population. Furthermore, the program data is also accessible to health care providers, researchers and post-secondary institutions that have informed research with respect to the public health of the citizens. The program has thus been an invaluable source for quantitative and statistical methods to further understand the demographics of the population holistically. For these aforementioned reasons, the DHS program serves an important and fundamental role within a country's statistical system. (https://en.wikipedia.org/wiki/Demographic_and_Health_Surveys)

The National Family Health Survey (DHS program in India) on the state findings in India assesses key demographics of women and children including their education, healthcare and living conditions. These factors are of utmost importance in the growth of a human being. Women that are highly educated tends to pay more attention to their own health leading to a happier and healthier lifestyle. Furthermore, these characteristics of a women would be passed down generations, affecting the health conditions of their children.

*Code and data are available at: https://github.com/jacobyokehongsi/Assessing_the_Knowledge_of_AIDS_amongst_Women_in_1997_Vietnam

The DHS on women and children's health aims to provide the International Institute for Population Sciences, which is the organization responsible for coordinating the health survey in India, information that includes demographics and trends. With the available data, it would assist the organization to help the Indian government design social policies that will best inform the Indian female and children population on improving their living standards. Thus, the DHS is one of the key aspects of the foundation in informing Indian women to have a healthy and functional support system as well as ushering new generations of Indians into a prosperous world.

In this paper, we will explore the background and characteristics of Indian women and children using the data from the 1998-1999 India NFHS-2. Specifically, I will address the following research questions:

-

The rest of the paper explores the following: Section 2 explores the origin of the data, methods used to obtain the data and its strengths and weaknesses. Section 3 needs to telegraph the rest of the paper: "Section 2. . . , Section 3. . . ."

2 Data

2.1 Data Source and Methodology

The data was obtained from India's second National Family Health Survey (NFHS-2) on women's and children's health. The survey was conducted in 1998-1999 using the NFHS questionnaire where households would be interviewed. The questionnaire was devised during workshops held in Mumbai where experts in population and health fields would partake in creating the questions. The target population was a sample of more than 90,000 ever-married women age 15–49 as well as children born in the three years preceding the survey. The list of villages is used as a sampling frame and first stratified through several variables. Next, the participants were selected using systematic sampling with equal probability from a household list in each village area. The overall individual response rate was 96% with a total of 89,199 eligible women that are part of the 91,196 interviewed households.

The following is a brief outline of the questionnaire content on the 1998 India NFHS:

- Background characteristics
- Reproductive behavior and intentions
- Quality of care
- Knowledge and use of contraception
- Sources of family planning
- Antenatal, delivery, and postpartum care
- Breastfeeding and health
- Reproductive health
- Status of Women
- Knowledge of AIDS

2.2 Survey Frame and Sampling Method

As mentioned in the previous section, participants of this survey were chosen via stratified sampling and systematic sampling. The survey frame that was created is as follows. The list of villages were first stratified by geographic location, then continuous regions and lastly further stratified using selected variables such as village size, female literacy and so on. Each state in India was stratified individually using these variables with a limit of 6 stratas for small states, 12 stratas for medium size states and 15 stratas for large states. From the list of stratified villages, they are then sampled systematically with probability proportional to the

Table 1: Several Key Features (in percent) Analyzed in the Paper

State	Females Age 6-14 Attending School	Households with Electricity	Infant Mortality Rate
Delhi	90.8	97.7	46.8
Haryana	85.5	89.1	56.8
Himachal Pradesh	97.3	97.2	34.4
Jammu & Kashmir	77.5	90.1	65.0
Punjab	90.0	95.5	57.1
Rajasthan	63.2	64.4	80.4
Madhya Pradesh	70.8	68.1	86.1
Uttar Pradesh	69.4	36.6	86.7
Bihar	54.1	18.2	72.9
Orissa	75.1	33.8	81.0

Table 2: Several Key Features (in percent) Analyzed in the Paper Continued.

State	Mothers >= 1 antenatal check-up	Underweight Children < Age 3	Women with Anaemia
Delhi	83.5	34.7	40.5
Haryana	58.1	34.6	47.0
Himachal Pradesh	86.8	43.6	40.5
Jammu & Kashmir	83.2	34.5	58.7
Punjab	74.0	28.7	41.4
Rajasthan	47.5	50.6	48.5
Madhya Pradesh	61.0	55.1	54.3
Uttar Pradesh	34.6	51.7	48.7
Bihar	36.3	54.4	63.4
Orissa	79.5	54.4	63.0

1991 Census population of the village. The villages that have fewer than five households are not included in the sampling frame. Lastly, the individuals to be interviewed were selected with equal probability from the household list in each village area using systematic sampling.

2.3 Key features

The raw data of the state findings includes 70 variables which were obtained from the questionnaire responses. Key features part of the raw data include education standards, accessibility to electricity and toilets, infant mortality rate, mothers receiving antenatal care, vaccination status of children, anaemia statuses (mothers and children) and weight statuses (mothers and children).

We used R (R Core Team 2020) and R package tidyverse (Wickham et al. 2019) to preprocess and analyze the data. When preprocessing the data, we also use R package pointblank (cite) to put together tests for class and content to ensure that the dataset passes. The key features analyzed in this paper is shown in table 1, which was made using R package knitr (cite).

2.4 Strengths and Weaknesses

2.4.1 Strengths

The primary strengths of this survey is the myriad of variables describing participants and households, as well as the large sample size where more than 99 percent of India's female population is surveyed. Additionally,

all Indian states except for one is surveyed, covering the majority of India. The field data collected by interviewing teams were inputted into microcomputers to produce field-check tables, ensuring that errors are prevented when eliciting information and filling out questionnaires. The large amount of variables allows for thorough data exploration of the relationships between them while the ample amount of data allows concrete conclusions to be drawn. Furthermore, the sampling methodologies is explained in high detail as well as a comprehensive fact sheet put together to display key data from the survey.

2.4.2 Weaknesses

One of the drawbacks of the survey is that survey findings were obtained in all states but one, that is Tripura, due to the delay in fieldwork based on a local problem. Additionally, data is unable to be obtained from areas with union territories. Nonsampling errors are also present where surveyors were unable to locate and interview the correct household and the misunderstanding of the questionnaire by the interviewer or respondent. In addition, the collection of anaemia data was hindered since some female participants refused to partake in the anaemia testing and/or have their children tested as well. These aforementioned weaknesses results in missing data thus, the inference we make could be less diverse.

details of the methodology used by the DHS you used, and its key features, strengths, and weaknesses.

3 Model

4 Results

The figures were made using R (R Core Team 2020), R package tidyverse (Wickham et al. 2019) and Lucidchart (cite). The variables explored provides numeric values that varies between the states of India.

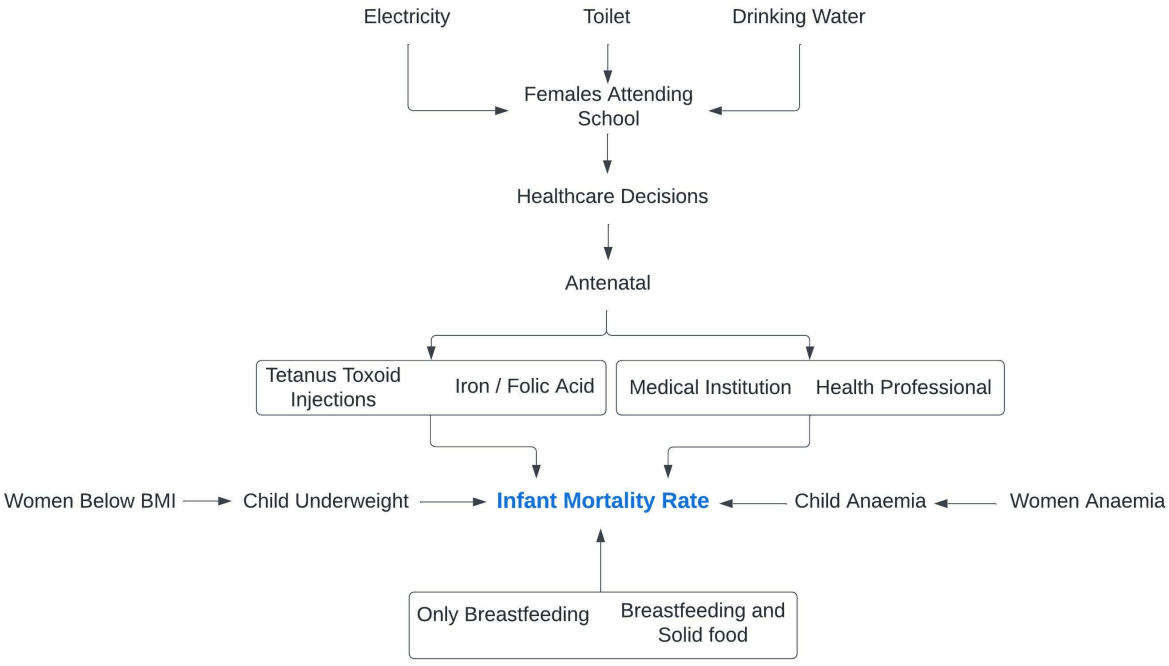

Figure 1: Supplementary Survey Questions

4.1 Household conditions that affect females attending school

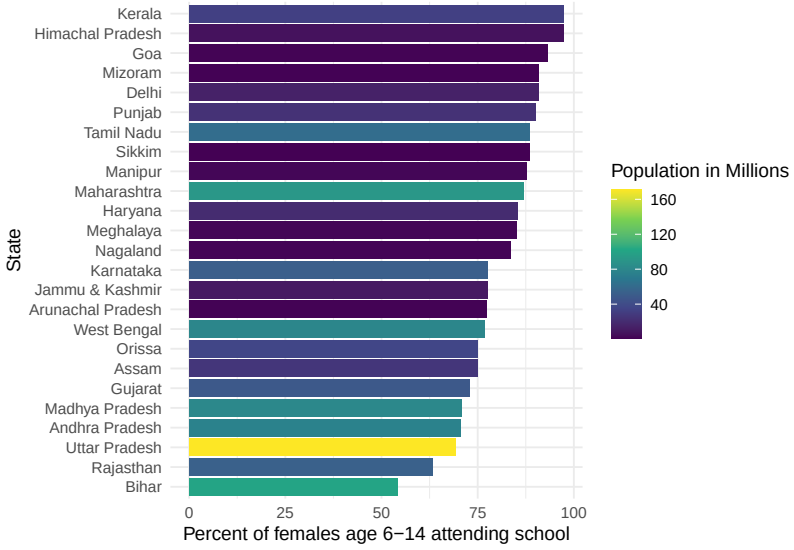

Figure 2: Distribution of the Percent of females age 6-14 attending school in various Indian States

Figure 2 depicts the distribution of females age 6-14 attending school (in percent) across the different states of India. Here, we observe that majority of females attending school, are from states including Kerala, Himachal Pradesh and Goa whereas majority of females not attending school are from Bihar, Rajasthan and Uttar Pradesh. In the states where the majority of females are not attending school, we can infer that these states tend to have a larger population, especially Uttar Pradesh.

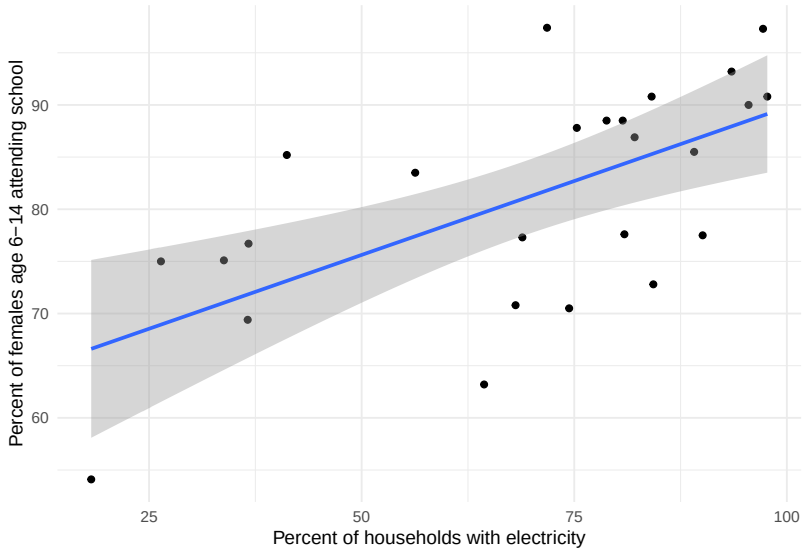

Figure 3: Relationship between females age 6-14 attending school and Percent of households with electricity (in percent)

Figure 3 depicts the relationship between households with electricity and females age 6-14 attending school (in percent). Here, we observe that as the percent of households with electricity increases, the percent of females age 6-14 attending school increases. This implies that households that can afford electricity tend to be more well-off and are able to send their children to school.

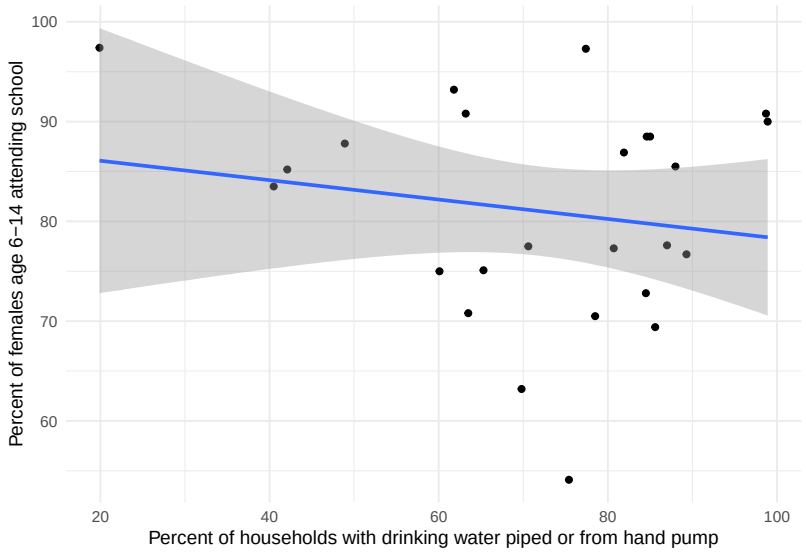

Figure 4: Graph of Illiteracy vs Immunization Rate

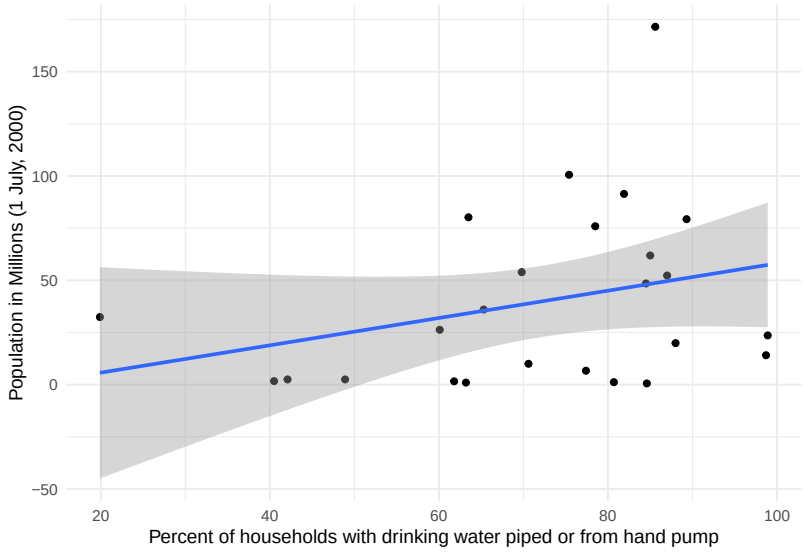

Figure 5: Graph of Illiteracy vs Immunization Rate

Figure 4 depicts the relationship between households with drinking water piped or from hand pump and females age 6-14 attending school (in percent). Here, we observe that as the percent of households drinking water piped or from hand pump increases, the percent of females age 6-14 attending school decreases at a moderate level. If we were to also look at Figure 5, as the percent of households with drinking water piped or from hand pump increases, the population increases. Thus, this aligns with the fact that in Figure 2, states with higher populations are also states where the majority of females are not attending school.

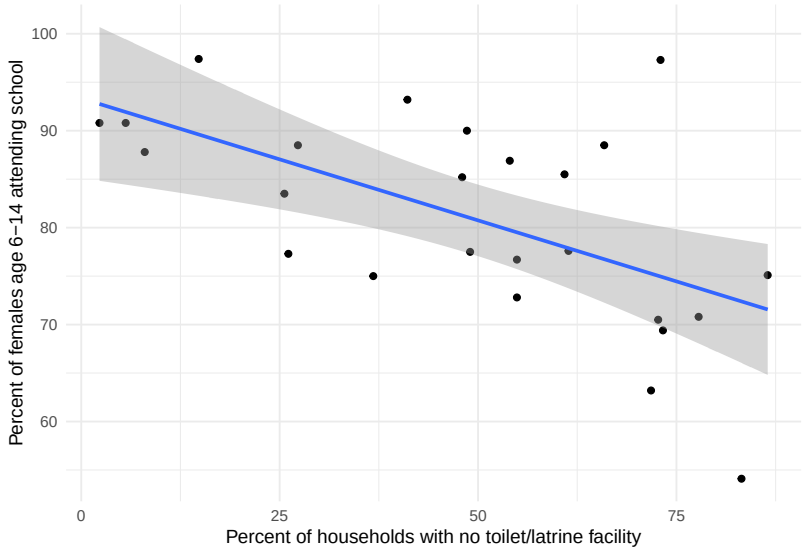

Figure 6: Graph of Illiteracy vs Immunization Rate

Lastly, figure 6 depicts the relationship between households with no toilet/latrine facility and females age 6-14 attending school (in percent). Here, we observe that as the percent of households with no toilet/latrine facility increases, the percent of females age 6-14 attending school decreases This implies that households that have sanitation areas contributes to a better living environment leading to educated female individuals in the household.

4.2 Females attending school and Women involved in decisions regarding personal healthcare

Figure 7 depicts the relationship between females age 6-14 attending school and women involved in decisions about their own healthcare (in percent). Here, we observe that as the percent of females age 6-14 attending school increases, the percent of women involved in decisions about their own healthcare increases. This suggests that female individuals that are educated are more likely to take care of their own health and live a healthier lifestyle.

4.3 Women involved in decisions regarding personal healthcare and mothers receiving at least one antenatal check-up

Figure 8 depicts the relationship between women involved in their own healthcare decisions and mothers receiving at least one antenatal check-up for births in the three years preceding the survey (in percent). Here, we observe that as the percent of women involved in own healthcare decisions increases, the percent of mothers receiving at least one antenatal check-up for births increases. This suggests that female individuals who are informed of their personal health care are also wary of having antenatal check-up conducted.

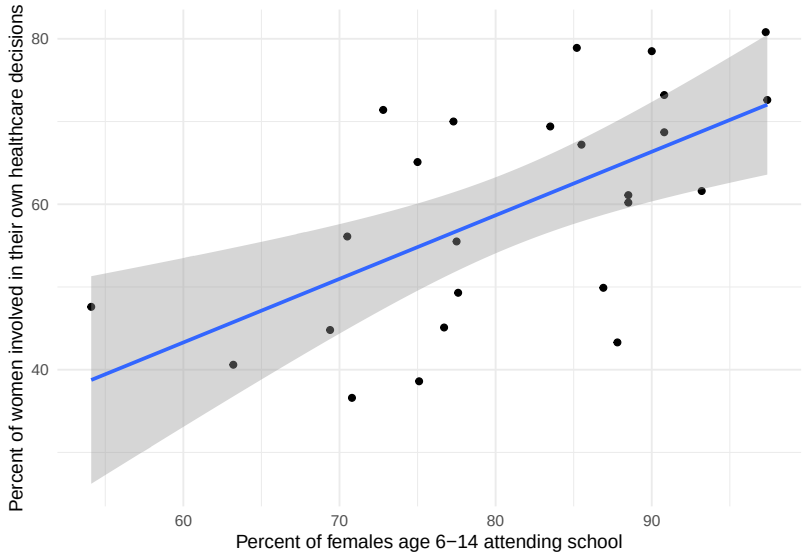

Figure 7: Graph of Illiteracy vs Immunization Rate

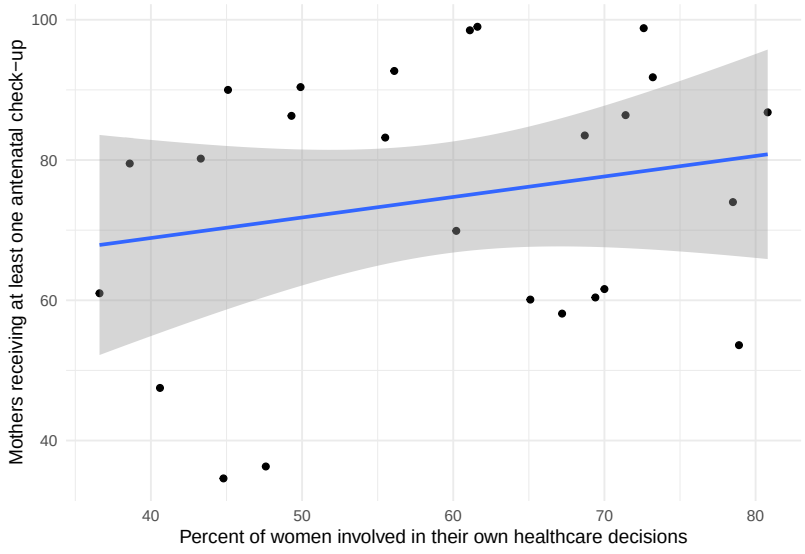

Figure 8: Graph of Illiteracy vs Immunization Rate

4.4 Mothers that receive antenatal check-up and the antenatal care they receive

Here, we explore how mothers that have antenatal check-up affect the antenatal care they receive for births three years preceding the survey.

4.4.1 Antenatal Injections and Supplements

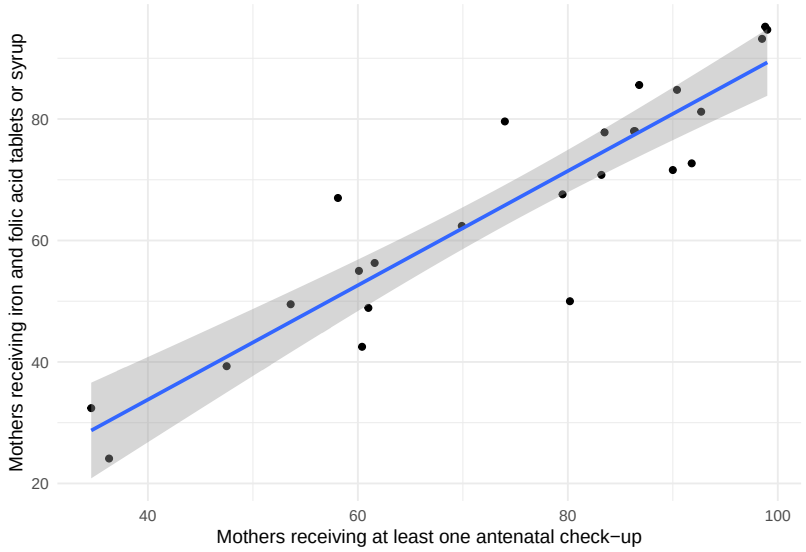

Figure 9: Graph of Illiteracy vs Immunization Rate

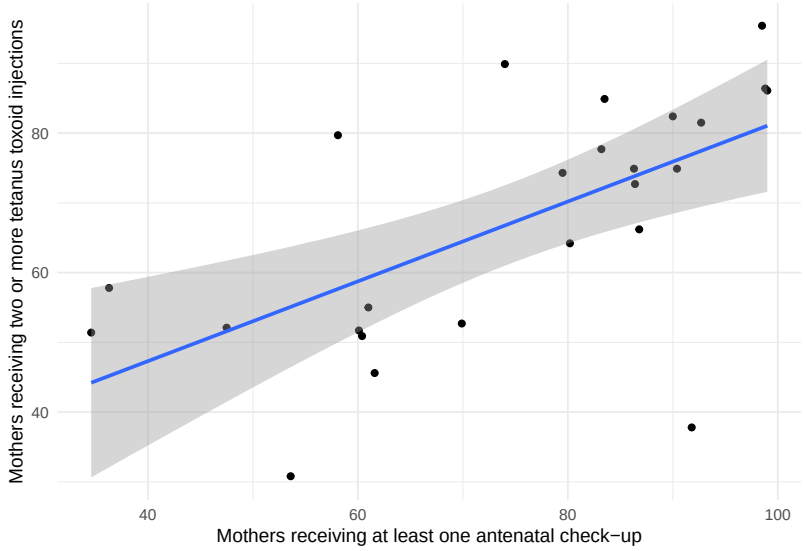

Figure 10: Graph of Illiteracy vs Immunization Rate

Figures 9 and 10 depict the relationships between mothers receiving at least one antenatal check-up with mothers receiving iron and folic acid tablets or syrup as well as two or more tetanus toxoid injections (in percent) respectively. We observe that as the percent of mothers receiving at least one antenatal check-up increases, the percent of mothers receiving two or more tetanus toxoid injections as well as iron and folic acid

tablets or syrup increases. This suggests that as mothers attend antenatal check-ups, doctors are likely to prescribe the mothers with tetanus toxoid injections as well as iron and folic acid supplements which boosts the health of the mother for a good birth delivery.

4.4.2 Birth Delivery

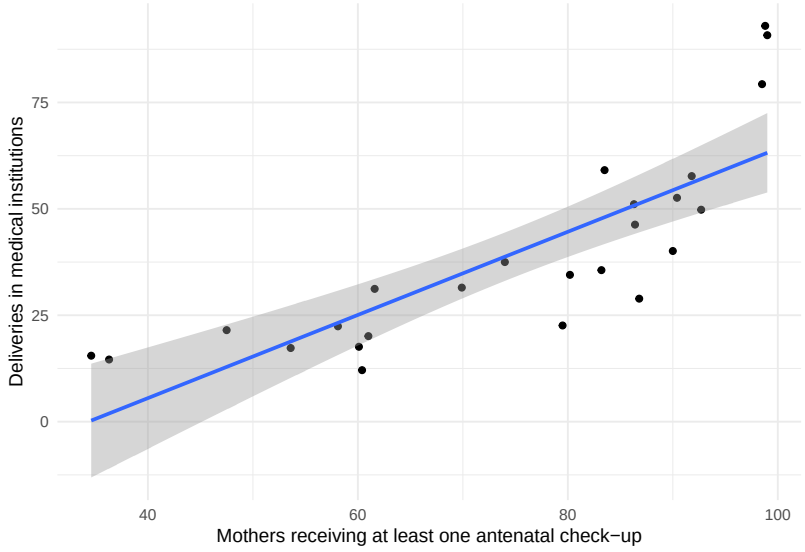

Figure 11: Graph of Illiteracy vs Immunization Rate

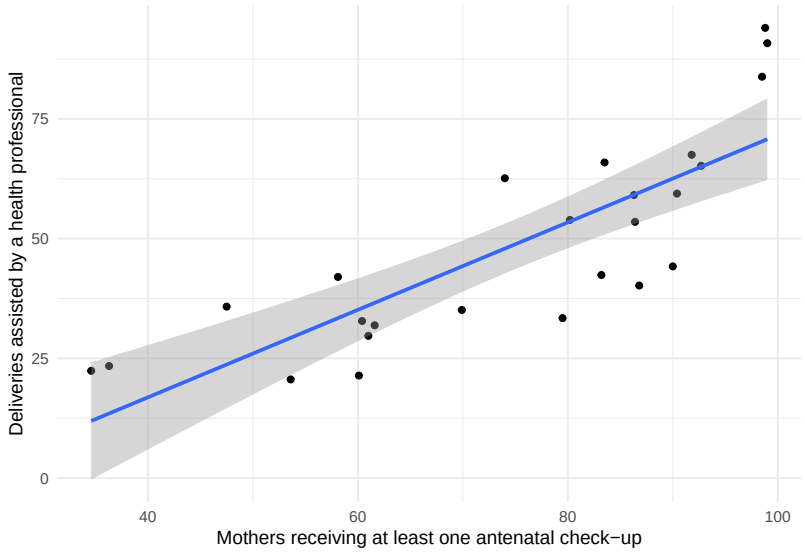

Figure 12: Graph of Illiteracy vs Immunization Rate

Figures 11 and 12 depict the relationships between mothers receiving at least one antenatal check-up with deliveries in medical institutions as well as deliveries assisted by a health professional (in percent) respectively. We observe that as the percent of mothers receiving at least one antenatal check-up increases, the percent of deliveries in medical institutions as well as deliveries assisted by a health professional increases. This suggests that when mothers go for their antenatal check-up, they likely have to approach a medical institution and have their check-up conducted by a health professional.

4.5 Factors that affect Infant Mortality Rate

In this subsection, we will explore the factors that will affect the infant mortality rate per 1,000 live births for the five years preceding the survey.

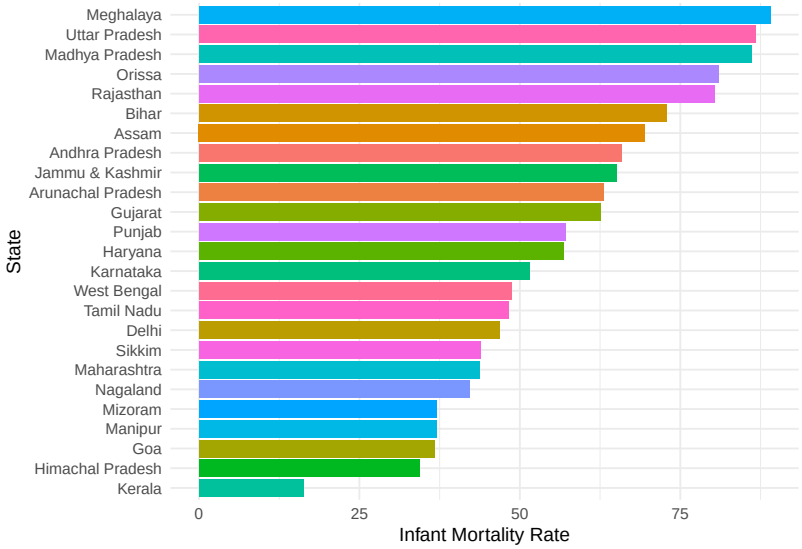

Figure 13: Distribution of Infant Mortality Rate in various Indian States

Figure 13 depicts the distribution of infant mortality rate across the different states of India. Here, we observe that states with the highest infant mortality rates include Meghalaya, Uttar Pradesh and Madhya Pradesh whereas states with the lowest infant mortality rates include Kerala, Himachal Pradesh and Goa.

4.5.1 Health Profile and Delivery Methods of Mothers prior to giving birth

Here, we explore how different health profile factors and delivery methods affect the infant mortality rate for births in the three years preceding the survey.

Figures 14 and 15 depicts the relationship between mothers receiving iron and folic acid tablets or syrup, and two or more tetanus toxoid injections (in percent) with infant mortality rate respectively. Here, we observe that as the percent of mothers receiving iron and folic acid tablets or syrup and two or more tetanus toxoid injections increases, the infant mortality rate decreases. This suggests that mothers that pay more attention to their own health and maintain a healthy regimen will lead to a reduction in deaths of infants when they are born.

Figures 16 and 17 depicts the relationship between deliveries in medical institutions and deliveries assisted by a health professional (in percent) with the infant mortality rate respectively. Here, we observe that as the percent of deliveries in medical institutions and deliveries assisted by a health professional increases, the infant mortality rate decreases. This implies that mothers that receive proper care during birth deliveries will also have a smooth child birth.

4.6 Children that receive breastmilk and/or both breastmilk and solid food

Figures 18 and 19 portrays the relationship between children age 0-3 months exclusively breastfed, and children age 6-9 months receiving breast milk and solid/mushy food (in percent) with the infant mortality rate respectively. Here, we observe that the percent of children age 0-3 months that are exclusively breastfed does not have a significant impact on infant mortality rate. On the other hand, as the percent of children age

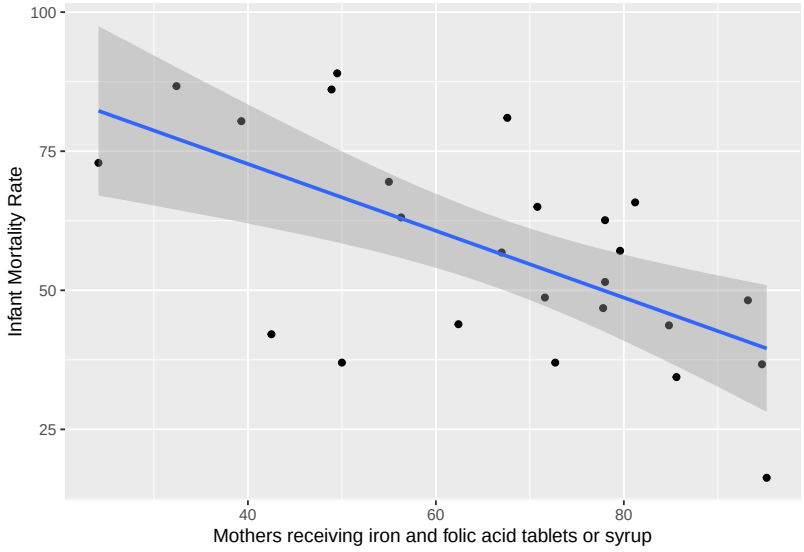

Figure 14: Graph of Illiteracy vs Immunization Rate

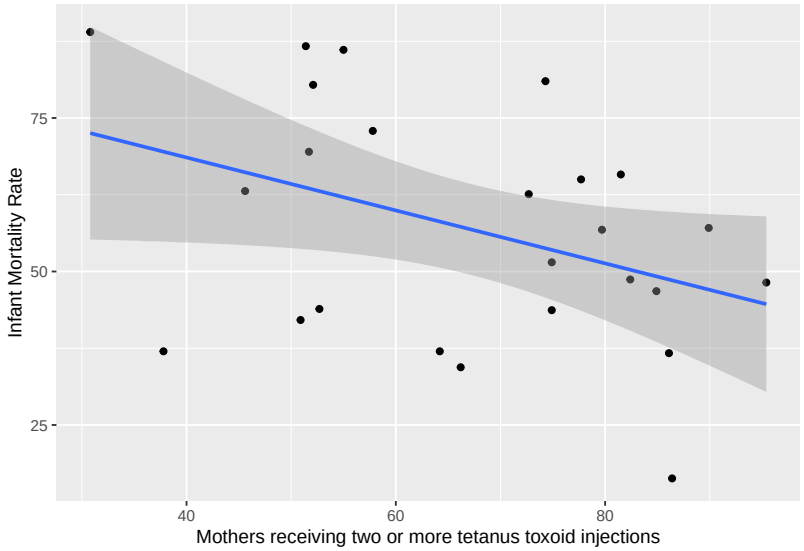

Figure 15: Graph of Illiteracy vs Immunization Rate

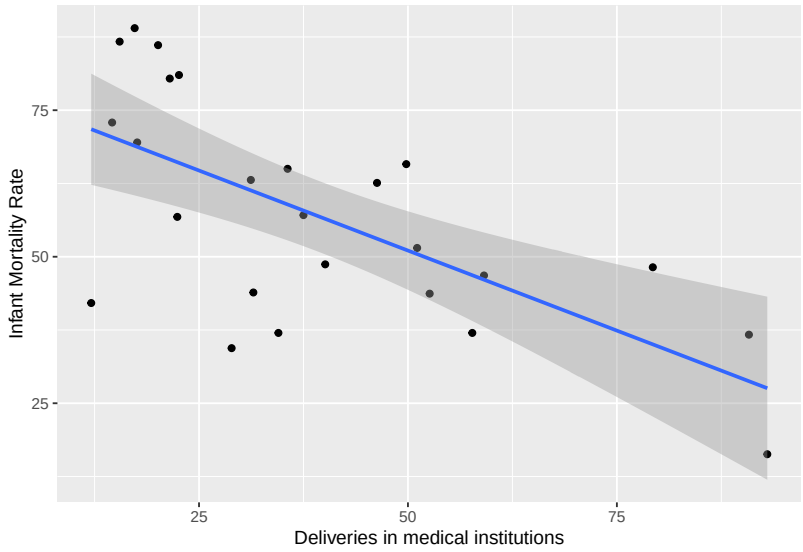

Figure 16: Graph of Illiteracy vs Immunization Rate

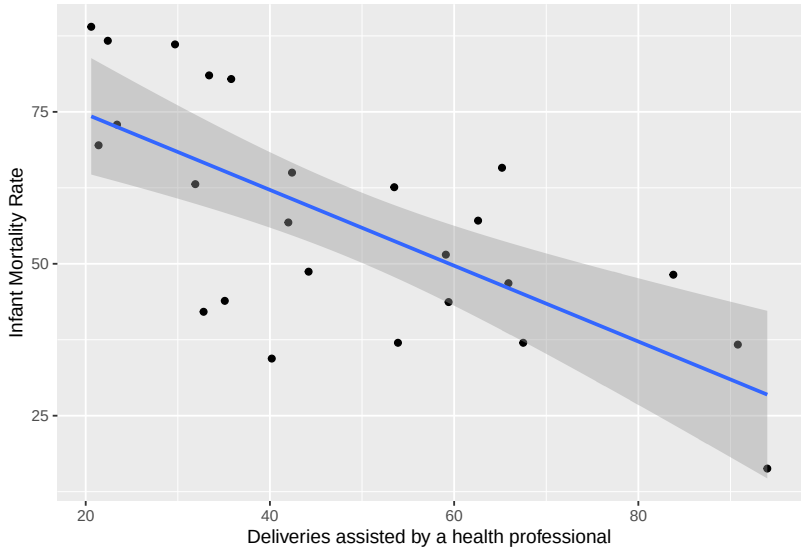

Figure 17: Graph of Illiteracy vs Immunization Rate

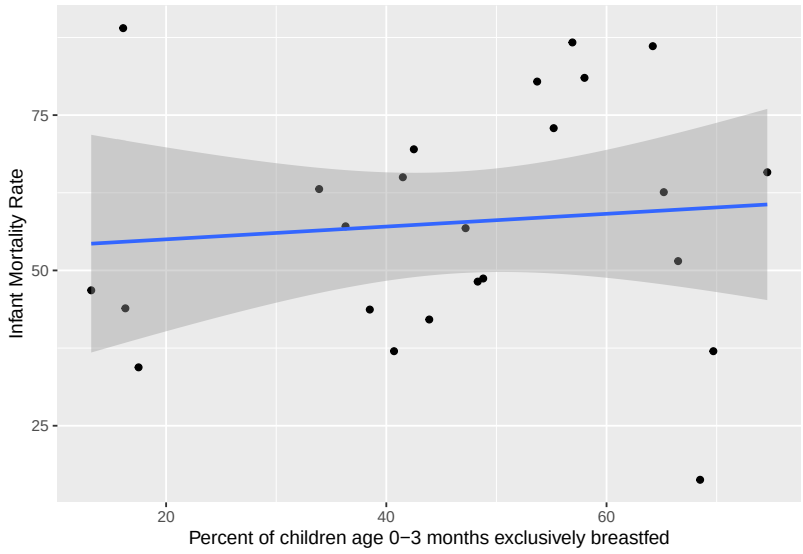

Figure 18: Graph of Illiteracy vs Immunization Rate

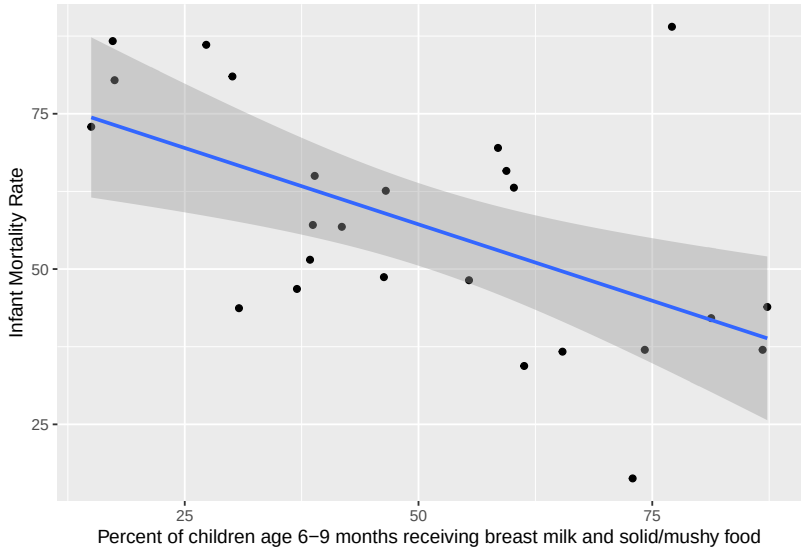

Figure 19: Graph of Illiteracy vs Immunization Rate

6-9 months receiving breast milk and solid/mushy food increases, the infant mortality rate decreases. This indicates that regardless of breast milk or formula milk fed to children at age 0-3 months, infant mortality rate is not affected. However, as children grow to the age of 6-9 months, they should be provided solid/mushy food on top of breast milk to support their growth, reducing their mortality.

4.7 Relationship between a women's body mass index (BMI), child's weight and infant mortality rate

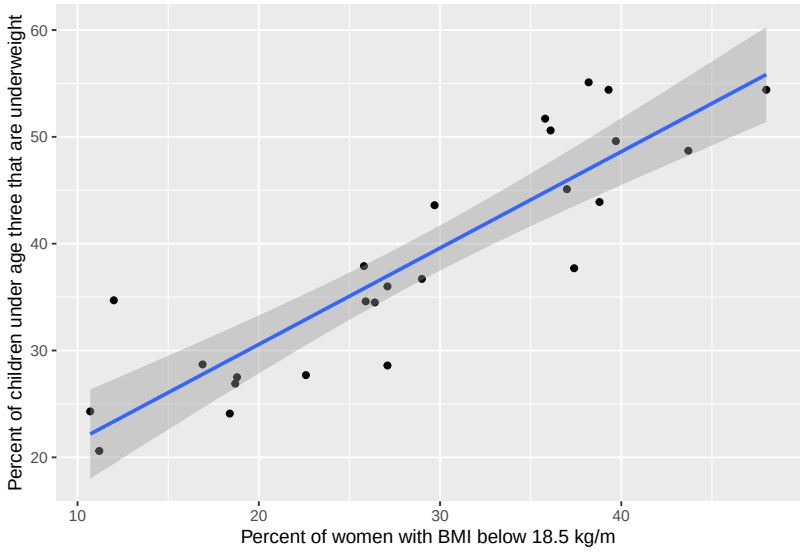

Figure 20: Graph of Illiteracy vs Immunization Rate

Figure 20 depicts the relationship between women with BMI below 18.5 kg/m and children under age three years that are underweight (in percent). Here, we observe that as the percent of women with BMI below 18.5 kg/m increases, the percent of children under age three years that are underweight increases. This implies that mothers that are skinny and malnutritioned would also lead to offsprings that are underweight.

Figure 21 depicts the relationship between children under age three years that are underweight (in percent) and infant mortality rate. Here, we observe that as the percent of children under age three that are underweight increases, the infant mortality rate increases. From this we can infer that children who are not underweight tends to be in good health which reduces infant mortality rate.

4.8 Relationship between women with anaemia, children with anaemia and infant mortality rate

Figure 22 depicts the relationship between women age 15-49 with any anaemia and children age 6-35 months with any anaemia (in percent). Here, we observe that as the percent of women age 15-49 with any anaemia increases, the percent of children age 6-35 months with any anaemia increases. This is coherent with the fact that since anaemia such as hemolytic anaemia can be inherited, mothers could pass on the anaemia to their offspring.

Figure 23 depicts the children age 6–35 months with any anaemia (in percent) and infant mortality rate. Here, we observe that as the percent of children age 6–35 months with any anaemia increases, the infant mortality rate increases. From this we can infer that children who have the anaemia disease would have symptoms that lead to critical health conditions and death.

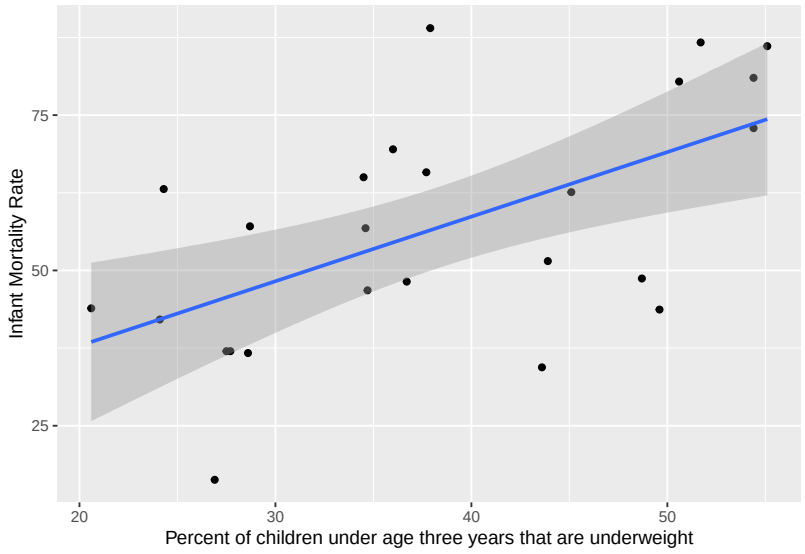

Figure 21: Graph of Illiteracy vs Immunization Rate

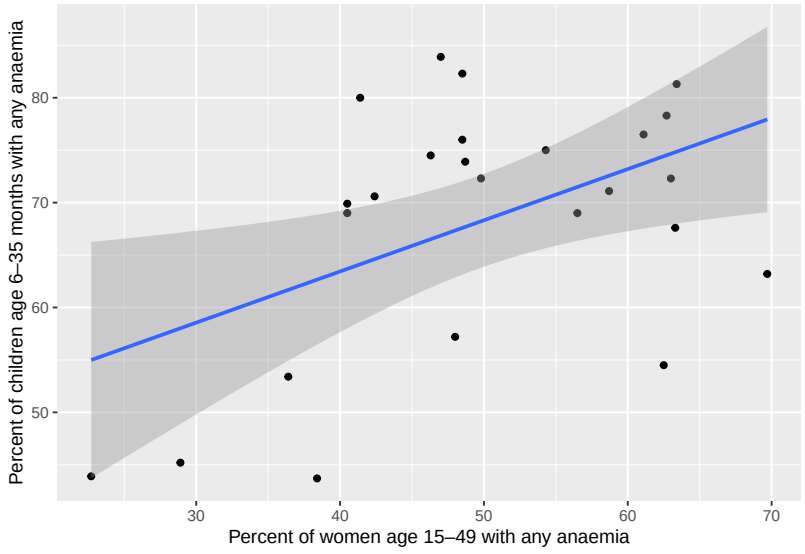

Figure 22: Graph of Illiteracy vs Immunization Rate

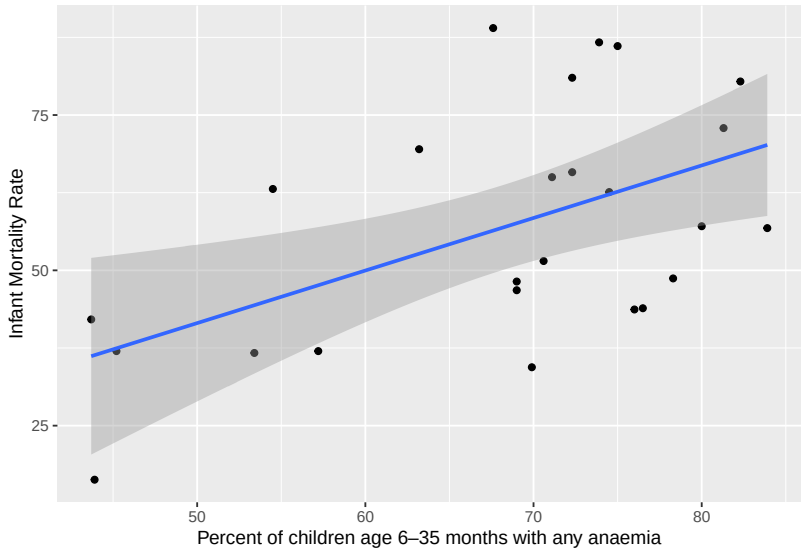

Figure 23: Graph of Illiteracy vs Immunization Rate

5 Discussion

The ultimate goal of the paper is to assist policy makers in implementing appropriate and efficient socio-economic policies to help promote the growth of children and reduce infant mortality rate in India. We will first discuss the root of the issue which is the household living environment, women's education and their healthcare decisions which lead them to attend antenatal check-ups. Then, we examine the numerous factors that is affected by conducting antenatal check-ups which include antenatal supplements and birth delivery methods. These factors are one of the few that directly impact the infant mortality rate. Additionally, we also look at the importance of infant's meals in their growth. Lastly, we examine inheritable traits such as anaemia and body weight that could be passed on from mother to child and how they impact the infant mortality rate.

5.1 Household Living Environment, Women's Education and their Healthcare Decisions

The household living environment of individuals including access to electricity, clean water and sanitation facilities is crucial in providing a comfortable and productive living space. As observed previously, households with better living conditions will also lead to women who are able to attend school and be well educated. In our figures, we could infer that households in India with poorer living conditions also have fewer women attending school. Thus, emphasis should be put in developing the living conditions within households in India, especially states with poorer living conditions such as Bihar. In Bihar, only 18.2% of the households have access to electricity and 54.1% of females attending school. With such poor basic needs and living standards, there is no doubt that the women is unable to attend school. By developing the living conditions, it will in turn improve the well-being of the population and encourage women to attend school. Another proposal could also be to provide direct subsidies for students in hopes to attain a well educated population.

Having an educated female population, we observe that women are more involved in making decisions regarding their own health care. By paying a higher attention to their personal health, women would devote more time and effort in ensuring a healthy lifestyle. In more developed states such as Himachal Pradesh, 97.3% of females age 6-14 are attending school. As these females become mothers, their personal knowledge on healthcare would be carried forward to having antenatal care which is shown as 80.8% of women is involved in decisions about their own health care in Himachal Pradesh.

5.2 Antenatal check-ups, injections and supplements to promote mothers and infant health

Attending antenatal check-ups is key for protecting the health of women and their unborn children. During antenatal check-ups, it is likely that doctors prescribe supplements and procedures needed in order for a smooth child birth. This includes iron and folic acid supplements that supports the development of the placenta and fetus as well as enabling the red blood cells in the body to supply oxygen to the child (https://www.mayoclinic.org/healthy-lifestyle/pregnancy-week-by-week/in-depth/prenatal-vitamins/art-20046945#:~:text=Ideally%2C%20you'll%20begin%20taking,of%20healthy%20red%20blood%20cells.), and tetanus toxoid injections to prevent neonatal tetanus (https://www.cdc.gov/pertussis/pregnant/hcp/vaccine-safety.html#:~:text=Pregnant%20women%20have%20been%20getting,1960s%20to%20prevent%20neonatal%20tetan Furthermore, partaking in check-ups will also increase the likelihood of having child deliveries in medical institutions and/or by a health professional since the check-ups are conducted in a similar fashion. Child deliveries in this manner are safe and sanitary, ensuring that the baby is delivered without any infections or accidents. In our data, we observe that in states where mothers often receive at least one antenatal check-up such as Goa (99.0% of mothers receiving at least one antenatal check-up), they also pay more attention to antenatal care where mothers receive two or more tetanus toxoid injections (86.1%), iron and folic acid tablets or syrups (94.7%) as well as having their infant delivered in medical institutions (90.8%) and/or by a health professional (90.8%).

5.3 Meals of infants and their effects on infant mortality rate

The diet of an infant is one of the most important fundamentals in a child's health. An infant has to be well-fed in order to receive the nutrients and vitamins required for a healthy body. To recap, we observe that for children age 0-3 months, being exclusively breastfed does not have a significant impact on the infant mortality rate. On the other hand, for children age 6-9 months, it is important that they receive solid/mushy food to supplement the breast milk they are consuming. States such as Rajasthan with low percent of children age 6-9 months receiving breast milk and solid/mushy food (17.5%) also have high infant mortality rates (80.4%). As a child grows, the nutrients supplied by the breast milk is simply insufficient to develop a healthy immune system. Thus, it is vital that children are supplied with ample amount of nutritious solid foods on top of breast/formula milk to promote a healthy growth.

5.4 Body weight of mothers and children and their effect on infant mortality rate

5.5 Anaemia of mothers and children and their effect on infant mortality rate

5.6 Weaknesses and next steps

- combination of all these factors

Appendix

A Data Sheet

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 analysis of policy makers and program administrators in India responsible for improving health and family welfare programs in the states.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
   - The International Institute for Population Sciences (IIPS) was designated as the agency for creating the dataset on behalf of the Ministry of Health and Family Welfare, Government of India, New Delhi.
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.*
   - It was funded by the Government of India.
4. *Any other comments?*
   - This second National Family Health Survey (NFHS-2) is created to further bolster the existing database from the first NFHS and further update the required health programs that should be implemented in the country.

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 represent the states of India. This includes North, Central, East, Northeast, West and South states of India.
2. *How many instances are there in total (of each type, if appropriate)?*
   - North: 6
   - Central: 2
   - East: 3
   - Northeast: 7
   - West: 3
   - South: 4
   - Total: 25 + 1 (India) = 26
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 contain all possible instances (states of India).
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 instance consists of 1 categorical variable representing the state and 18 continuous variables representings variables such as school attendence rate, percent of households with electricity, infant mortality rate etc.

5. *Is there a label or target associated with each instance? If so, please provide a description.*
    - Theres is no target associated with each instance.
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.*
    - Under the instance Goa and variable percent of children age 0-3 months exclusively breastfed, information is missing because the data collected is based on fewer than 25 unweighted cases.
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.*
    - The dataset does not contain data that might be considered confidential.
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.*
    - The dataset does not contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety.
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 identifies women and children below age three.
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 (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset.
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.*

- The dataset does not contain data that might be considered sensitive in any way.

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 associated with each instance is acquired through interviews conducted in 25 Indian states.
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?*
    - Questionnaires are first filled out using manual human curation then entered into microcomputers to produce field-check tables which ensures that the data is free from errors.
3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*
    - The dataset is not a sample from a larger set.
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 fieldwork in each state was carried out by a number of interviewing teams, each team consisting of one field supervisor, one female field editor, four female interviewers, and one health investigator. Compensation data is not provided.
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 collection was carried out in two phases, starting in November 1998 and March 1999. The timeframe matches the creation timeframe of the data associated with the instances.
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)?*
    - The data is obtained from the Demographic and Health Survey program website. https://dhsprogram.com/publications/publication-frind2-dhs-final-reports.cfm?cssearch=467922_1
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 in question were notified as they voluntarily provide data to the interviewers. The exact language of the notification is not available.

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 in question consent to the collection and use of their data. For example, the health investigator read a detailed informed consent statement to the respondent, informing her about anaemia, describing the procedure to be followed for the test, and emphasizing the voluntary nature of the test. The exact language is as follows. "May I ask you now to give your consent to have the test(s) done. If you decide not to have the test(s), it is your right, and we will respect your decision. Now please tell me whether you agree to have the test(s)(and allow me to test your child)."
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).*
    - Yes. In the questionnaire, they are allowed to revoke their consent.
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.*
    - Analysis of the potential impact of the dataset and its use on data subjects has not been conducted.
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.*
    - Preprocessing/cleaning/labeling of the data was done. The original format of the file is in pdf and the data was extracted using R package pdftools and R in order for data analysis to be conducted.
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 is saved in inputs/data/raw_data1.csv, inputs/data/raw_data2.csv and inputs/data/raw_data3.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 available at https://www.r-project.org/
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 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?*
   - Tasks could include analyzing the demographics of women and children in India during 1998-1999.
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 preprocessing is curated just for this specific dataset and cannot be applied to other datasets.
5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*
   - N/A
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.*
   - The dataset will not be distributed to third parties outside of the entity on behalf of which the dataset was created.
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 on Github.
3. *When will the dataset be distributed?*
   - The dataset will be distributed during 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 distributed 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.*
   - No third parties have imposed IP-based or other restrictions on the data associated with the instances.
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.*
   - No export controls or other regulatory restrictions apply to the dataset or to individual instances.
7. *Any other comments?*
   - N/A

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
   - Jacob Yoke Hong Si
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
   - Through email at jacobyh.si@mail.utoronto.ca.
3. *Is there an erratum? If so, please provide a link or other access point.*
   - There is no erratum.
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)?*
   - There is no plan to update the dataset. Any updates will be communicated to dataset consumers via Github.
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.*
   - There are no applicable limits on the retention of the data associated with the instances.
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.*
   - Older versions will not be hosted. Its obsolenscence is communicated via commit history on 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.*
   - The repository on Github can be cloned to extend/augment/build on/contribute to the dataset. Contributions will not be validated/verified since I am not responsible for any other dataset extensions.
8. *Any other comments?*
   - N/A

References

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, 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.