

Project 3: Literacy Rates and Mean Time Spent in School for Men and Women

Eileen Hertwig

October 18, 2018

1 Introduction

With data from *gapminder*, (www.gapminder.org), literacy rates and the average time spent in school are analyzed. Data for various countries is available for a few decades. The countries are divided into different groups based on their current literacy rates to make the analysis more concise.

The following questions will be answered:

1. What are the differences in literacy rates and the average time spent in school based on gender?
2. How did literacy rates and average time spent in school develop over time in the different groups?
3. What are the main differences between the trends in literacy rates and in average time in school for men and women?

2 Data

2.1 Data files used

In this data analysis project data from *gapminder* is used. The following data files (including short description and source url taken from <https://www.gapminder.org/data/>) are used:

1. Literacy rate, adult total (% of people ages 15 and above):
 - Adult literacy rate is the percentage of people ages 15 and above who can, with understanding, read and write a short, simple statement on their everyday life.
 - source: <http://stats.uis.unesco.org/unesco/TableViewer/tableView.aspx?ReportId=210>
2. Literacy rate, adult female (% of females ages 15 and above):
 - Adult literacy rate is the percentage of people ages 15 and above who can, with understanding, read and write a short, simple statement on their everyday life.
 - source: http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=136&IF_Language=eng&BR_Topic=0

3. Literacy rate, adult male (% of males ages 15 and above):
 - Adult literacy rate is the percentage of people ages 15 and above who can, with understanding, read and write a short, simple statement on their everyday life.
 - source: http://stats.uis.unesco.org/unesco/TableView/document.aspx?ReportId=136&IF_Language=eng&BR_Topic=0
4. Mean years in school (men 35 to 44 years):
 - The average number of years in school attended by all people in the age and gender group specified, including primary, secondary, and tertiary education.
 - source: <http://ghdx.healthdata.org/record/global-educational-attainment-1970-2015>
5. Mean years in school (women 35 to 44 years):
 - The average number of years in school attended by all people in the age and gender group specified, including primary, secondary, and tertiary education.
 - source: <http://ghdx.healthdata.org/record/global-educational-attainment-1970-2015>

2.2 Preparing the data

This analysis is based on the literacy rates of all adults (*file: Literacy rate, adult total*). Countries are grouped based on their total literacy rates. Therefore this file will be explored first (see python script *explore_literacy_total.py*):

The data is loaded and explored. It contains 38 years of data for 150 countries. However, there are a lot of missing data. To see if there are any countries without any data, all missing values per row are counted, but each country has at least two years of data. Looking at the counts over all countries for the individual years, it is found that most missing values occur in the years from 1975 to 1999. The temporal evolution is (at the moment) not of interest, because countries will be classed based on the current average literacy rate. It seems appropriate to use the last 12 years of data (2000 to 2011), because there are too many missing values before that. Three countries do not have any data in these 12 years and are dropped, which is not problematic because there are still data for 147 countries.

The mean literacy rate for each country from 2000 to 2011 is computed. A new data frame with the countries and the mean literacy rate is written to a new file to be used in the next part.

To prepare the other data files, the script *prepare_data.py* is used. Data for literacy rates of men and women is provided (*files: Literacy rate, adult male; Literacy rate, adult female*). A similar data exploration as for the total literacy rate is conducted. The dataframes consist of the same years and countries as the one for the total literacy rate. Missing values also follow a similar distribution.

Furthermore, data on how many years men and women, who are between 35 and 44 years old in the specified year, attended school is explored (primary, secondary, and tertiary education is all included, *files: Mean years in school (men/women)*). These dataframes include 47 years (from 1970 to 2015) and 187 countries, but no missing values. All countries from the original list (total literacy rates) are contained in the 4 new data frames. All other countries are dropped.

As a next step the countries are divided into 4 groups based on the mean literacy rates computed before:

1. *low*: minimum to q_{25} (19.0% to 68.1%)
 - 'Afghanistan', 'Bangladesh', 'Benin', 'Bhutan', 'Burkina Faso', 'Central African Republic', 'Chad', 'Congo, Dem. Rep.', 'Cote d'Ivoire', 'Eritrea', 'Ethiopia', 'Gambia', 'Ghana', 'Guinea', 'Guinea-Bissau', 'Haiti', 'India', 'Liberia', 'Madagascar',

'Malawi', 'Mali', 'Mauritania', 'Morocco', 'Mozambique', 'Nepal', 'Nigeria', 'Pakistan', 'Papua New Guinea', 'Rwanda', 'Senegal', 'Sierra Leone', 'Sudan', 'Timor-Leste', 'Togo', 'Yemen', 'Zambia'

2. *medium: q_{25} to q_{50} (68.1% to 88.7%)*

- *'Algeria', 'Angola', 'Botswana', 'Burundi', 'Cambodia', 'Cameroon', 'Cape Verde', 'Comoros', 'Dominican Republic', 'Ecuador', 'Egypt', 'El Salvador', 'Gabon', 'Guatemala', 'Guyana', 'Honduras', 'Iran', 'Iraq', 'Jamaica', 'Kenya', 'Lao', 'Lesotho', 'Libya', 'Mauritius', 'Namibia', 'Nicaragua', 'Oman', 'Peru', 'Sao Tome and Principe', 'Saudi Arabia', 'Swaziland', 'Syria', 'Tanzania', 'Tunisia', 'Uganda', 'Vanuatu', 'Zimbabwe'*

3. *moderate: q_{50} to q_{75} (88.7% to 97.4%)*

- *'Albania', 'Bahrain', 'Bolivia', 'Bosnia and Herzegovina', 'Brazil', 'Brunei', 'China', 'Colombia', 'Costa Rica', 'Equatorial Guinea', 'Greece', 'Indonesia', 'Jordan', 'Kuwait', 'Lebanon', 'Macedonia, FYR', 'Malaysia', 'Maldives', 'Malta', 'Mexico', 'Myanmar', 'Palestine', 'Panama', 'Paraguay', 'Philippines', 'Portugal', 'Qatar', 'Seychelles', 'Singapore', 'South Africa', 'Sri Lanka', 'Suriname', 'Thailand', 'Turkey', 'United Arab Emirates', 'Venezuela', 'Vietnam'*

4. *high: q_{75} to maximum (97.4% to 100%)*

- *'Antigua and Barbuda', 'Argentina', 'Armenia', 'Azerbaijan', 'Belarus', 'Bulgaria', 'Chile', 'Croatia', 'Cuba', 'Cyprus', 'Estonia', 'Georgia', 'Hungary', 'Italy', 'Kazakhstan', 'Kyrgyz Republic', 'Latvia', 'Lithuania', 'Moldova', 'Mongolia', 'Montenegro', 'North Korea', 'Poland', 'Romania', 'Russia', 'Samoa', 'Serbia', 'Slovenia', 'Spain', 'Tajikistan', 'Tonga', 'Trinidad and Tobago', 'Turkmenistan', 'Ukraine', 'Uruguay', 'Uzbekistan'*

The mean over all countries in each category is taken and four new dataframes are created, combining the information on male/female literacy rates and years in school and written to new files. (Note that the new and ordered data still contains missing values, which will be considered in the analysis.)

3 Literacy rates and average schooling time for men and women

3.1 Proportions (women/men)

To get a first look at gender differences for literacy rates and time in school a bar chart is plotted which shows the proportion for women/men (see fig. 1). Only the most recent data (the year 2011) is used in this plot. For equal literacy rates and time in school for men and women all values would have to be one (see upper gray dashed line in fig. 1). If values are less than one, women have lower literacy rates or spent less time in school and vice versa.

From the bar chart it is obvious that in regions with generally lower literacy rates, the gender gap is the greatest. The lower the total literacy rates the bigger is the difference in literacy rates between men and women (with women having lower rates). Only in regions with overall high literacy rates, the ratio between men and women is equal. However, while gender differences in literacy rates are great in regions with lower literacy, it is even worse for the average time spent in school. In the regions with the lowest overall literacy rates, women spent only little more than half the time in school than men (see lower gray dashed line in fig. 1). In the “medium”

and “moderate” groups, women also spent significantly less time in school and the gender gap is even greater than for the literacy rates. Only in regions with a very high overall literacy, women spent slightly more time in school.

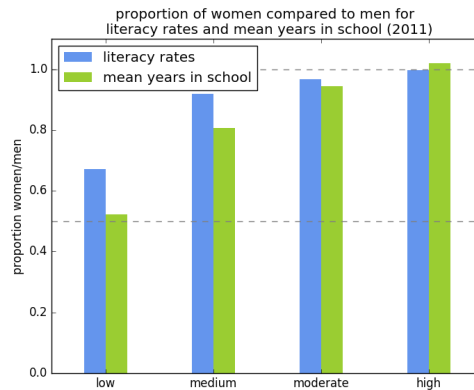


Figure 1: Proportion of women/men for literacy rates and average time spent in school for the four groups

3.2 Trends in literacy rates

Figure 2 shows the time series for literacy rates in the four different groups for men and women. The dots are values for the individual years (but means over the entire group). Especially in the regions with low and medium literacy rates, the fluctuation is relatively high (which is of course due to the method of deviding into groups by quartiles), therefore linear trends are computed (with *numpy.polyfit*). We can see that for both men and women, there is a positive trend in all groups. The trends are stronger in regions with lower literacy rates, but even the group with the highest overall literacy, things have improved over the last few decades.

To better compare trends for men and women, all trendlines are plotted together in figure 3. Here we can again see, that literacy rates for women are overall lower than for men, especially in the 3 groups with lower literacy. For the regions with the lowest literacy rates, the gender gap is not only the greatest, there has also been no improvement over the last decades. In all other groups, the gender gap has narrowed over time and has almost been closed in the “moderate” and “high” groups.

3.3 Trends in school time

The temporal evolution of mean years in school is very smooth. Fluctuations are so low, that trendlines are not computed but the actual values for each year are plotted in a line plot. Trends can be easily deducted from this plot (see figure 4). Trends for the time spent in school follow a different pattern than the trends for literacy rates. Again all trends are positive, but are stronger in the groups with higher than in the ones with lower literacy rates.

In all groups for most of the time, women spent less time in school than men. This difference is the greatest in “low”, where the gender gap has even widened over the last decades. In the “medium” group, the gender gap has more or less remained the same, while it has narrowed in the “moderate” group. In regions with “high” literacy rates, women also spent less time in

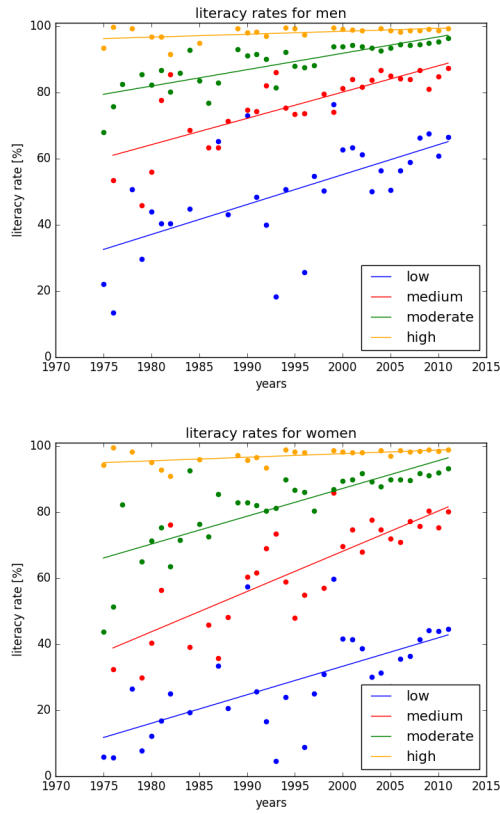


Figure 2: Literacy rates for men and women over time (dots: single years, lines: linear trends)

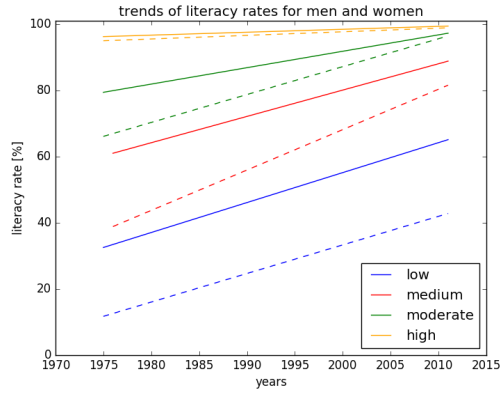


Figure 3: Trendlines for literacy rates for men (solid lines) and women (dashed lines)

school until approximately the year 2000. The positive trend is stronger for women than for men in this group and in the last 1-2 decades, women have progressively spent more time in school.

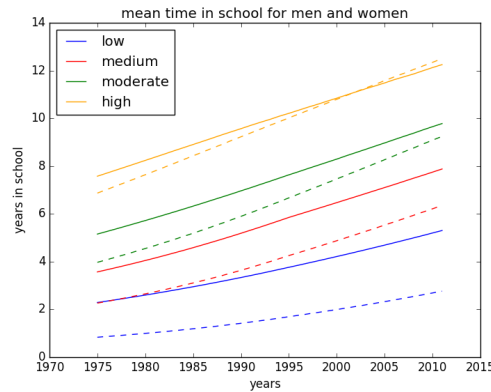


Figure 4: Trendlines for average time spent in school for men (solid lines) and women (dashed lines)

4 Conclusion

In the introduction the following questions have been asked:

1. What are the differences in literacy rates and the average time spent in school based on gender?
2. How did literacy rates and average time spent in school develop over time in the different groups?
3. What are the main differences between the trends in literacy rates and in average time in school for men and women?

With a bar chart showing the proportions of women/men, the first question could be answered for each group for the most recent year (2011). It could be seen that the gender gap is still present in many regions of the world. Only in countries with very high overall literacy rates, it has been closed (or even overturned in case of time spent in school).

The second question has been answered showing time series and linear trends over the past few decades. While all variables have positive trends, there are some differences. In regions where there is the most room for improvement, literacy rates have the strongest positive trends. However, the second-to-last group (“medium”) seems to have the strongest positive trends, while improvements in the “low” group are a bit slower. For the mean time in school, trends have the strongest positive tendency in regions with the highest literacy rates.

The third question again addresses trends, but looks into differences between men and women. For the “medium”, “moderate”, and “high” regions, literacy rates as well as time spent in school improve faster for women than for men (even though in most cases, the gap is not closed at the end of the time series). In countries with the lowest literacy rates, the gender gap has not narrowed at all over the last decades and even worsened in case of time in school.

The average time spent in school is, of course, related to the literacy rates in a country, because reading and writing is taught in school. It appears, however, that improvement is faster in the time spent in school than in literacy rates. Partly, this is due to the fact that there is an upper limit for literacy rates (100%), while there is (almost) always room for improvement in time

spent in school. At least in the “moderate” and “high” group, which have literacy rates close to 100% in the year 2011, this can explain the differences in the trends.

In the “medium”, literacy rates have improved strongly, especially for women, over the last few decades. While time in school has also improved, the stronger trends in literacy rates can be explained by the fact that the mean years in school in this group are now (2011) higher than 6 years (for women and even higher for men), which seems more than sufficient to learn how to read and write. Secondary and tertiary educations should not be related to literacy anymore.

While literacy rates and time in school in the “low” group have improved, the improvement is not as great as hoped and the gender gap has not narrowed. It might be worth looking into individual countries in this group. It is very possible that the developments are very different in the individual countries, because countries with overall mean literacy rates between 19.0% and 68.1% are included in this group. In the “high” group the range is only from 97.4% to 100.0%, so much less differences are expected. This is a disadvantage of the method of deviding into groups based on quartiles (all groups have and equal amount of memebers) and needs to be taken into consideration when interpreting the data.