Secrets Behind the Education Level in Zambia*

Unbalanced education resources distribution is a obstacle in country development

Zhiyue Yu Matthew Ma

10 April 2022

Abstract

The educational level of the population is a strong indicator of national development. From the perspective of Zambia, which has just become a developing country from an underdeveloped country, it is of great significance to study the educational level of the general population in this region ("Zambia" 2022). In this paper, we will use the information from Zambia's 1996 Demographic and Health Survey final paper provided by DHS (Demographic and Health Survey), an official statistic program supported by the United States Agency for International Developments, to extract, clean, and to conclude all the education-related information including local educational penetration, educational attainment by gender, educational attainment by region, and educational attainment by share (Ministry of Health/Zambia and 1997 1996). By reviewing the data, although Zambia was an undeveloped country in 1996, the overall education level is in a relatively better situation than in other undeveloped countries. The majority of the population accepted primary and secondary level education, and many people have accepted higher-level education. Therefore, we concluded that although the proportion of the educated population in the whole society varies greatly in some specific provinces, and there is an obvious inequality in the proportion of the educated population between the sexes, the proportion of the educated population in the whole society is still getting better year by year because of the increase of the total educated population. Lastly, we deduct that the increasing educated population gradually became an important factor that boosted the economic pattern of the whole country in the future, which ultimately indirectly made Zambia become a developing country after more than one decade.

1 Introduction

Education is an important indicator to determine the degree of social development. A country's level of education and its people's general level of education largely determine the development and future of a country.

Generally speaking, the more developed a country is, the more outstanding its education industry is. Take the United States and The United Kingdom, for example; the land of these two recognized as the most developed countries in the world almost accounts for more than 90% of the world's top education resources ("University Ranking" 2022). Moreover, the more developed the country, the higher the share of its population that is generally educated. In developed countries like Canada, people with bachelor's degrees make up 70% of the population ("Study: Youth and Education in Canada" 2021).

On the other hand, In undeveloped countries like Zambia, Cambodia and Mumbai... the proportion of people with higher education is as low as the total number of well-educated population is even less than that in certain city in a developed country. With globalization and expanding economic volume under the global context in recent decades, the difference in social education levels has also become an essential factor that makes many backward countries struggle below the poverty line for decades.

So far, it is tough for the vast majority of underdeveloped countries to become developing countries. However, in recent years, Zambia in southern Africa has broken away from the ranks of underdeveloped countries

*Code and data are available at: https://github.com/gggeorgeeeyu/paper4.

and finally become a developing country, largely due to its people's basic education level ("Zambia" 2022). According to the open data provided by DHS Program (Demographic and Health Survey) founded by United States Agency for International Development, the world's largest official aid agency, we found that Zambia had a better proportion of an educated population than other backward countries as early as 26 years ago in last century (Ministry of Health/Zambia and 1997 1996).

From the DHS's final report of Zambia in 1996, many critical social data, including education, marriage, childbirth, reproduction, infectious disease and mortality rate, were covered (Ministry of Health/Zambia and 1997 1996). The data was based on the sampling survey results and the sampling data extracted from the whole of Zambia in 1996. We found very detailed information on Zambia's educational attainment stratified by region, gender and province in 1996 (Ministry of Health/Zambia and 1997 1996).

Although the information was precisely provided by the DHS program, since this paper was posted more than twenty years, it is still difficult for people to intuitively sum up the information about social education levels efficiently from this 300-page report. Moreover, as the information provided by DHS is a PDF version, a document more likely to be visually read than collected, it will be challenging for most people to collect and sort out data. This form of the file makes it hard to summarize education-related information.

Therefore, in this research paper, we focus on collecting, extracting, and cleaning the education-related information from this report and produce a shorter, more efficient, visually readable report. In addition, the methodology of data cleaning is fundamental; therefore, we will use a more statistically efficient method to summarize and integrate the data into new tables and charts by scanning. The analysis will examine Zambia's 1996 Educational Population from the perspective of multiple variables and extend some generalizations in society. Lastly, an interpretation of how the population's education level eventually benefited it and became a developing country more than one decade later.

2 Data

In the paper, we run our analysis in `R` (R Core Team 2020). To read in pdf file, we used `pdftools` package written by (Ooms 2022). To clean data and visualize data, we used `tidyverse` package written by (Wickham et al. 2019), `ggplot2` written by (Wickham 2016), `purrr` written by (Henry and Wickham 2020), `janitor` written by (Firke 2021), `stringi` written by (Gagolewski 2021).

The data we used was from the Zambia DHS, 1996 - Final Report (English), which is one of the reports in the DHS Program[Demographic and Health Survey]. The DHS Program is a program that contains reports from various of countries related to demographic and Health Surveys. We converted one of the full-page tables in the report of Zambia in 1996 called "Level of education by background characteristics" into usable table. In table 1, we can see that it describes the distribution of respondents by their highest level of education based on their background characteristics, including their age group, region and province. The table we converted is an aggregated data set, which has shown the aggregated data for analysis. This table includes six sub-tables, dividing the respondents into women and men, it also shows the three characteristics respectively for both men and women. The variables named "No education", "Primary", "Secondary" and "Higher" correspond to the highest level education of the respondents, the data under these variables are calculated in percentage. The column "Total" checks the total percentage of the group, which should be 100 per cent. The last column "number" is the total number of respondents for each group. In order to graph the data for visualization and data analysis, we want to clean and transform the data set before the analysis. We first calculated the exact number of respondents for each education level in each age group, region and province, then we combine the sub-tables together to show the distribution of education levels in each category. After the data transformation, we visualize and analyze the results in the section 3. By doing this practice, we will be able to visualize the data, see the trend and do further analysis. By analyzing this data set, we can also see the nature of the development of an undeveloped country to developing country. We can also provide suggestions for undeveloped countries by emphasizing the importance of education and how they can increase the average highest education level.

Table 1: Level of education by background characteristics(shown in percentage)

background	no_education	primary	secondary	higher	total	number
WOMEN						
Age						
15-19	8.4	61.6	30	0.1	100	2003
20-24	10.9	57.3	29.9	1.9	100	1830
25-29	11.7	57.1	27	4.2	100	1286
30-34	13.1	60.6	21.2	5	100	1081
35-39	14.1	62.9	17.8	5.2	100	758
40-44	20.6	53.5	19.1	6.8	100	568
45-49	37.3	54.3	7.8	0.6	100	494
Residence						
Urban	5.6	48.1	40.7	5.6	100	3604
Rural	19.6	67.6	12.2	0.6	100	4417
Province						
Central	7.9	61.4	26.8	3.8	100	653
Copperbelt	5.8	50.8	40	3.4	100	1588
Eastern	32.7	58.4	8.9	0	100	1075
Luapula	15.6	66.3	17.4	0.7	100	726
Lusaka	7.8	48.1	36.8	7.3	100	1403
Northern	11.8	71.9	15.5	0.8	100	872
North-Western	18.5	60.7	20.1	0.7	100	288
Southern	10	68.5	19.2	2.3	100	816
Western	18.6	61.3	18.1	2	100	600
Total	13.3	58.9	25	2.8	100	8021
MEN						
Age						
15-19	6.9	62.7	30.5	0	100	460
20-24	6.8	50.4	40.9	2	100	404
25-29	6.2	40.1	46.6	7.1	100	255
30-34	8.5	42.7	42	6.8	100	225
35-39	1.4	44.1	42.5	12	100	184
40-44	6.5	42.4	36.3	14.9	100	121
45-49	9.7	51.7	32.4	6.1	100	83
50-54	7.6	60.8	20	11.6	100	65
55-59	18.2	62.7	11.2	7.9	100	52
Residence						
Urban	2.7	34.9	53.2	9.2	100	852
Rural	10.4	64.2	23.4	2	100	997
Province						
Central	7.2	43.5	44.7	4.6	100	157
Copperbelt	2.8	37.6	53.7	5.9	100	396
Eastern	19.2	58.7	21.3	0.8	100	254
Luapula	5.1	64.7	26.6	3.6	100	151
Lusaka	5.2	38.1	43.6	13	100	316
Northern	2.1	64.6	31.2	2.1	100	221
North-Western	3.1	63.6	27.5	58	100	48
Southern	5.6	60.8	32.4	1.2	100	173
Western	11.9	56.5	24.7	7	100	132
Total	6.9	50.7	37.1	53	100	

3 Results

In each data set, the horizontal axis represents different levels of education, while the number axis contains three different variables: age group, region, and province, respectively. Each variable corresponds to two ICONS for gender. Different levels of education are also divided into Categorical so that readers can intuitively see the corresponding amount and proportion of different levels of education under different indicators. In addition, to present a more intuitive visual experience, we also mark different variables in color.

3.1 Age group vs. population education level

From the graph of age group vs. population education level for women(Figure 1), we classified the age group by 5 years from 15 years old to 49 years old. It is evident whether in which age group the majority are the primary education. In addition, the total number of people who have only received primary education will be relatively average regardless of the range, basically accounting for the overall population.

On the other hand, the portion of those who had experienced higher education is the minor group, which is reasonable because the well-educated group is always a relatively small proportion in most societies, especially in an undeveloped society like Zambia. Regardless of the age group from 15 to 19 years old was too young to take higher education, the most negligible proportion of that appears in 45 to 49 years old, and the highest proportion of that appears in the age group of 40-year-old to 44 years old. Moreover, it shows a gradually decreasing trend with the age group decreasing. Moreover, the portion of secondary education and no education also takes up a significant total. There is little difference between the two in total, but the only difference is that with the decrease in age, fewer and fewer people have no education, while more and more people have received secondary education.

From this graph, we can see that although the higher education portion tends to decrease, the proportion of secondary education and the no education proportion respectively rise and go down significantly. That marks the overall social fundamental education level is increasing.

For the same graph in men(Figure 2), the most dominant portion is still primary education, but the secondary education proportion shows a very fluent trend. It indicates an increasing trend with the decreasing age group from 49 to 25 years, and the peak of that was shown from the age group from 25 to 29 years old. Nevertheless, it decreases when the age decreases from 25-29 to 15-19. The two highest proportion of higher educated group appears from 40-44 and 50-54 years old and there shows no regularity of trend for that in the overall statistics. Regardless of the youngest generation, the non-educated group is the tiniest portion, and that points to the education situation for men being better than women from the perspective of age group.

3.2 Region vs. population education level

In terms of the indicator of the region, the variables on the y-axis are only rural and Urban. Both the two histograms under each variable can be considered as a box plot. Therefore, an analysis of that will be given from the perspective of the box plot.

From the graph of region vs. population education level for women(Figure 3), whether in urban or rural areas, we can find that the primary education proportion occupies the majority of the whole distribution. Rural areas present a normal distribution with primary distribution, with more on both sides and less in the middle, and almost no higher education population. However, for urban areas, the proportion of people with secondary and higher-level education has increased significantly, while those without education have decreased significantly.

From the graph of region vs. population education level for men(Figure 4), it is undeniable that the rural region has poor education level than in urban; this considerable difference is quite typical for most undeveloped countries and developing countries. For that in the rural region, regarding the graph as a box plot, it follows by a normal distribution, the vast majority are the people who take primary education, the secondary education shares the second large part, and the no education and higher education respectively take the third and the tiniest proportion. However, urban areas are different, the secondary education proportion became the most dominative part of the data, and the primary education proportion decreased. Compared

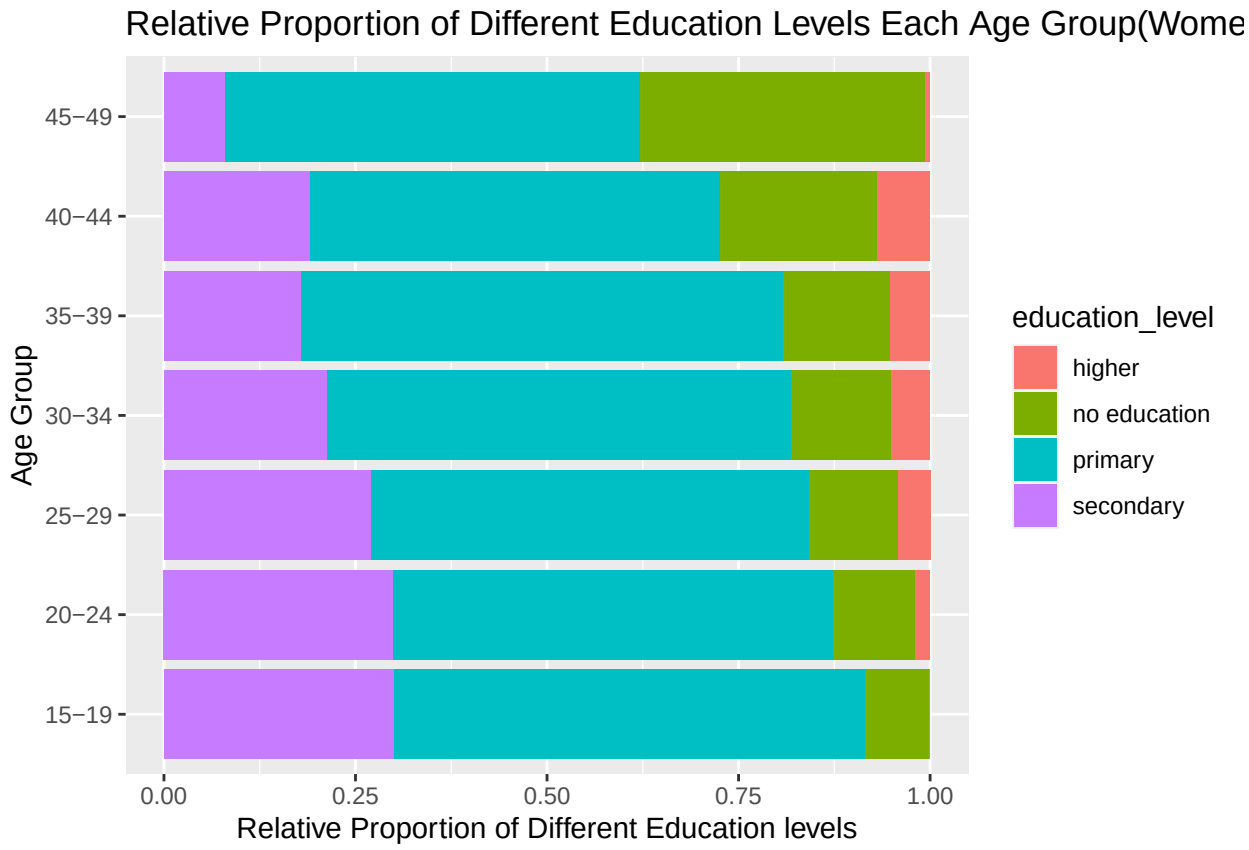

Figure 1: Relative Proportion of Education Levels Each Age Group(Women)

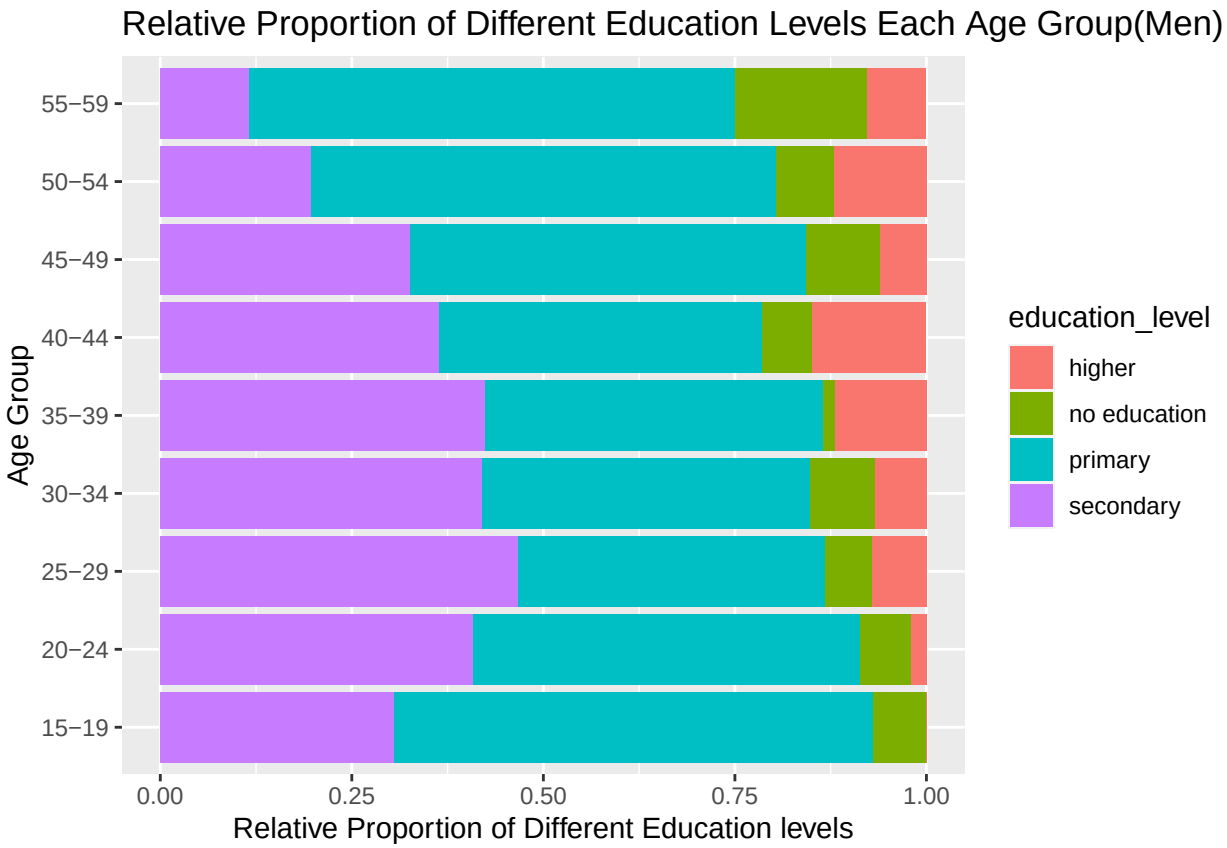

Figure 2: Relative Proportion of Different Education Levels Each Age Group(Men)

with rural areas, the proportion of higher education and no education in the urban area shows an opposite trend, occupying the third and fourth place respectively, and the number of people with higher education has increased significantly.

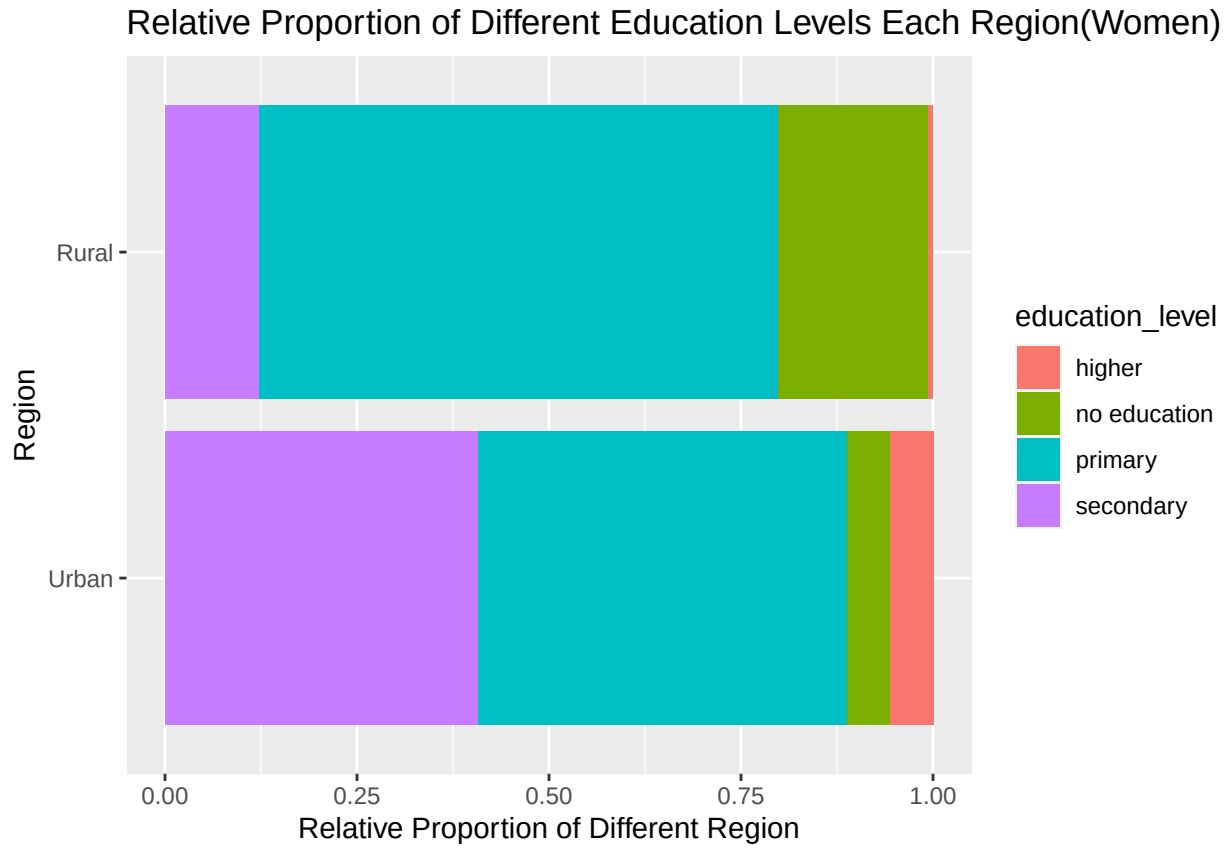

Figure 3: Relative Proportion of Education Levels Each Region(Women)

3.3 Province vs. population education level

From the graph of province vs. population education level for women(Figure 5), primary education still takes the biggest proportion in all the provinces. In the Central, Lusaka and Copperbelt provinces, the education ratio is relatively high, with a relatively high proportion of people receiving higher education and secondary education. Among them, the Copperbelt has the largest proportion of people receiving secondary education. Lusaka has the highest percentage of higher education. In addition, for all the other provinces, there is basically no difference between them. The number of people with higher education is very small. In many provinces, there are almost no people with higher education, even in Eastern region.

There appears to be a tremendous difference between women and men on this indicator. As can be seen from the chart(Figure 6), the most significant changes are seen in the North-Western region, where the high proportion of male education occupies the second largest proportion and is only slightly lower than the most significant proportion of primary education level. The overall status of the other regions is similar to that of women in the provinces. Central, Lusaka and Copperbelt provinces still have the highest secondary education level.

4 Discussion

Based on the above results and graphs(Section 3), we will first discuss the specific generalization of different educational proportions in Zambia society in 1996 that we learned from the survey. In addition, we will

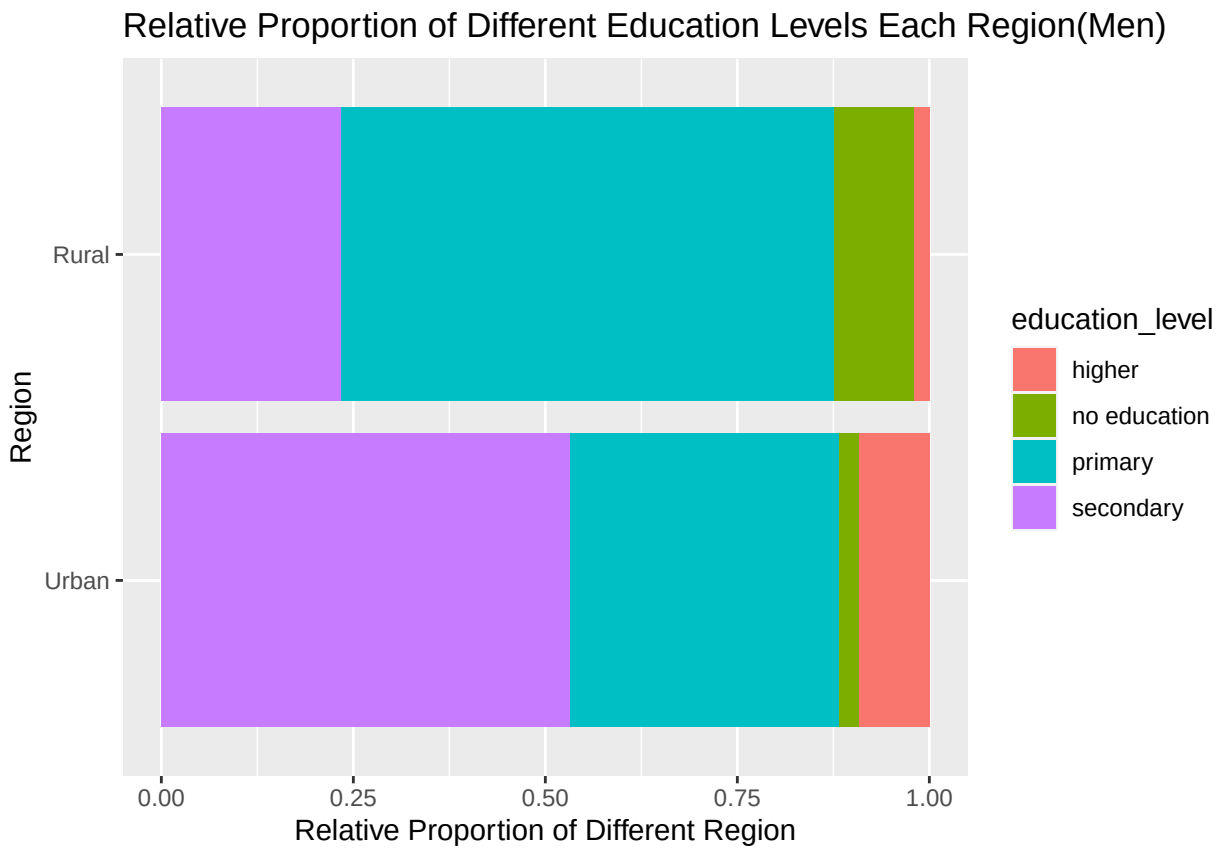

Figure 4: Relative Proportion of Different Education Levels Each Region(Men)

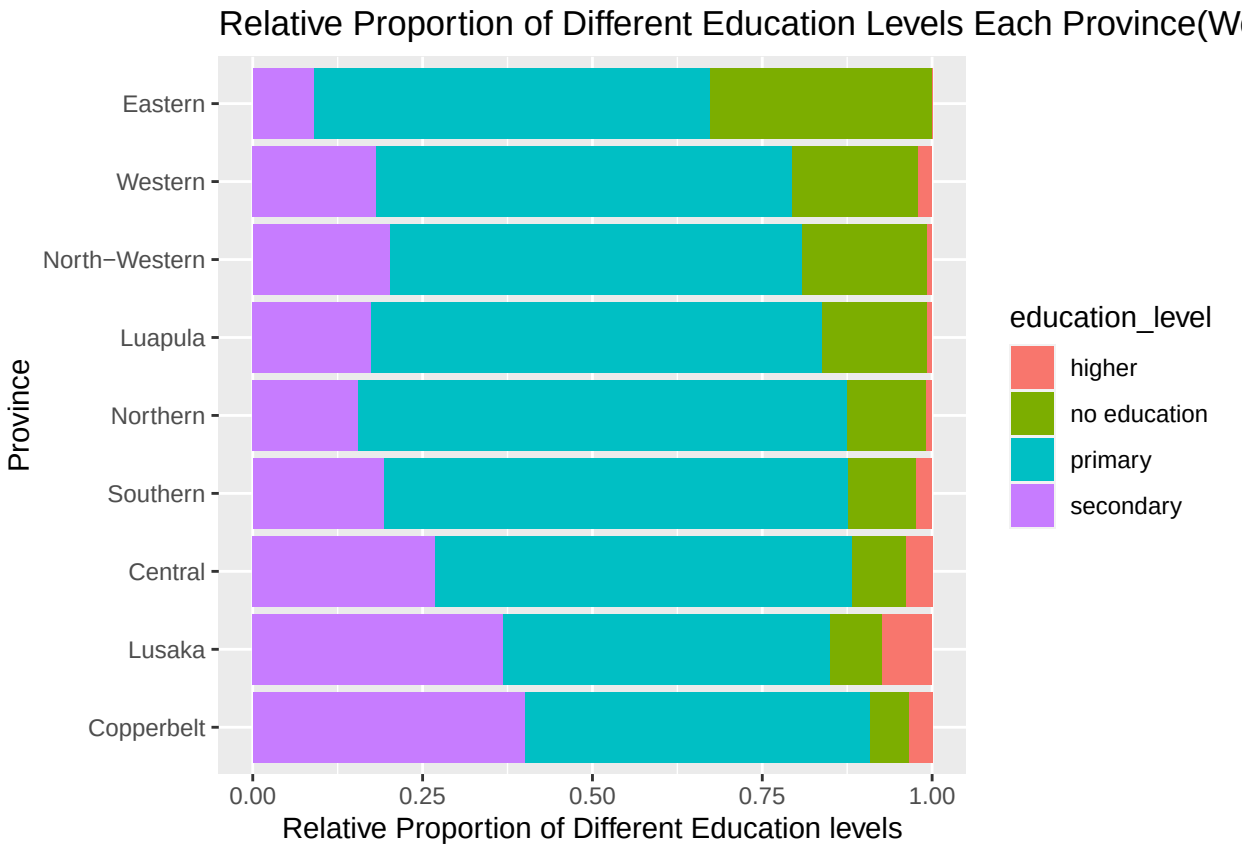

Figure 5: Relative Proportion of Education Levels Each Province(Women)

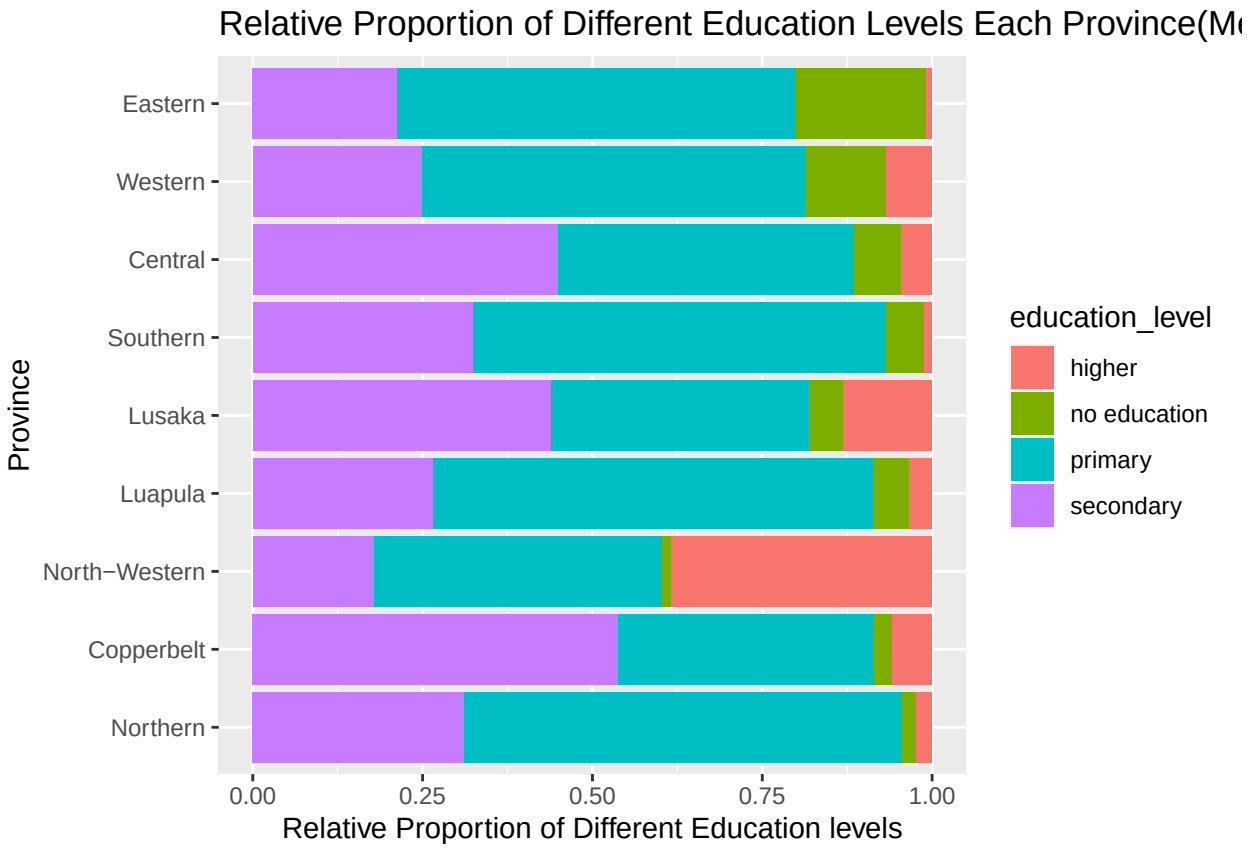

Figure 6: Relative Proportion of Different Education Levels Each Province(Men)

discuss how the reorganization of the data model will affect the country's future and find some shortcomings in the experiment of this group.

4.1 Research Conclusion

Back to the above three variables and the six different charts for analysis: From the graph of age group vs. population education level, we found that the overall education situation is that men's education level is generally better than women's, and the trend for women is getting better with age decreasing. We can also see that in terms of Region vs. population education level, the gender situation is the same. There is a similar normal distribution trend in rural areas centered on primary education. However, in urban, we can find that the proportion of no education decreases rapidly. The proportion of secondary education and higher education level is much higher than that of rural. From the point of province vs. population education level, From the chart, we can see that of the two genders, the most significant changes are found in the North-Western region. There is almost no proportion of higher education for women in North Western, which probably indicates severe gender discrimination in this region, and the distribution pattern is very different from that of a backward country. No matter in which province, the number of males who have received secondary education is much higher than that of females, and the number of females who have not received education is also much higher than that of males.

To sum up, in Zambia in 1996, the distribution of education levels in different regions was very unreasonable. Even though there were significantly more cities than villages, it was reasonable, but cities had absolute educational advantages over rural regions under the premise of such huge differences. In particular provinces, there is a massive gap in the education level that men and women can receive. Even though men have a better distribution of education status than women in different age groups, the extremely unbalanced ratio of men and women in education in a particular province magnifies the difference in education level between men and women.

4.2 Social Relevance

From the point of view that Zambia had finally left the underdeveloped country to become a developing country after more than one decades, or from the point of view that looked ahead in 1996, the data showed that the first factor that ultimately enabled Zambia to progress was the gradual increase in educational attainment with lower age (Ministry of Health/Zambia and 1997 1996). We see in the restructuring data that regardless of gender, people are increasingly educated, and young people are getting better and better education so that they can build their countries more effectively. In addition, from this set of data, we can find that even though Zambia was still an underdeveloped country at that time, the overall education level in urban areas was relatively good, especially for men. It can be seen from the data that there are very few uneducated males in urban areas, and most of them have received education above secondary level, which can promote the rapid development of cities and thus promote the development of the whole country. Similarly, even though there may be gender inequality in education levels in provinces like the North-Western, there are a large number of highly educated people in the region, and the aggregation of these highly educated people may be crucial to the scientific and technological development of a society. Therefore, the above three factors may lead Zambia eventually became a developing country after a period of development.

4.3 Weakness and limitation

Even if GHS provides this data set, the data collection methods more than 20 years ago may not be diverse, which may lead to the problem of insufficient samples. On the other hand, since the READ PDF file scans the file, there are still a few problems with data loss when doing data cleansing.

References

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

Gagolewski, Marek. 2021. "Stringi: Fast and Portable Character String Processing in r." *Journal of Statistical Software*.

Henry, Lionel, and Hadley Wickham. 2020. *Purrr: Functional Programming Tools*.

Ministry of Health/Zambia, Central Statistical Office at the, and Macro International. 1997. 1996. "Zambia Demographic and Health Survey, 1996." *Calverton, Maryland: Central Statistical Office and Macro International Inc.* https://dhsprogram.com/pubs/pdf/FR86/FR86.pdf.

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

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

"Study: Youth and Education in Canada." 2021. Statistics Canada. https://www150.statcan.gc.ca/n1/daily-quotidien/211004/dq211004c-eng.htm.

"University Ranking." 2022. QS TOPUNIVERSITIES. https://www.topuniversities.com/university-rankings.

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

"Zambia." 2022. USAID FROM THE AMERICAN PEOPLE. https://www.usaid.gov/zambia?page=2.