Final Project

Yigit Tahmisoglu

2023-04-05

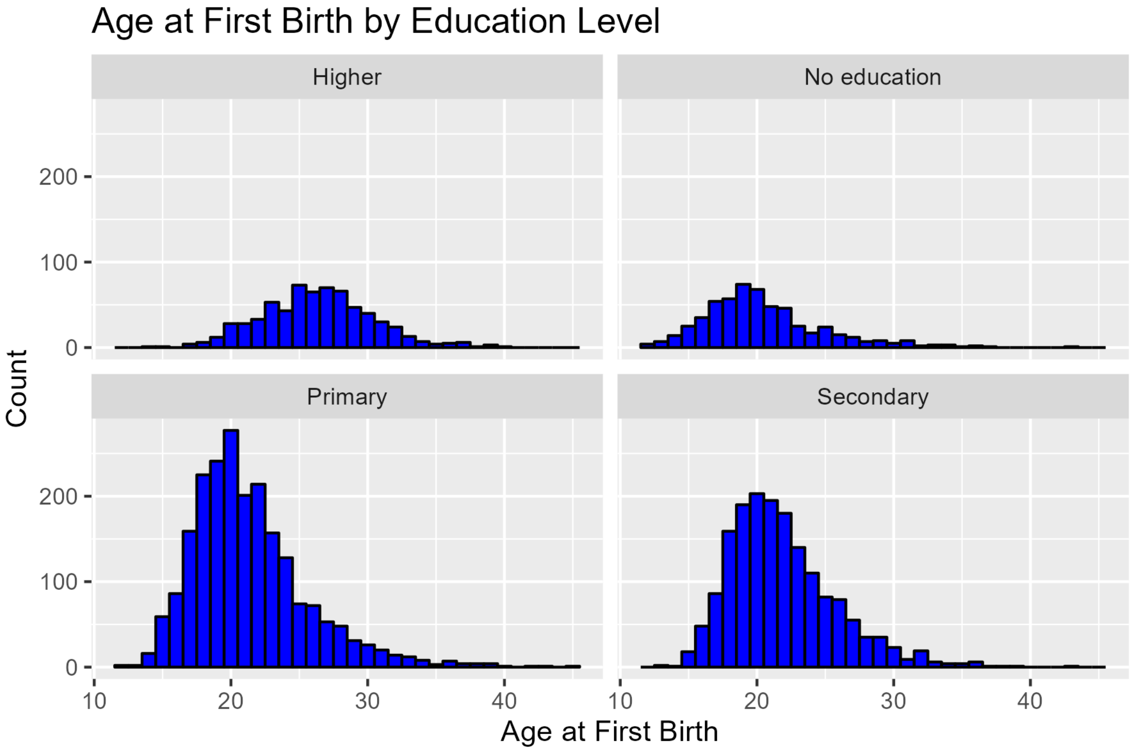
1. Github repository: https://github.com/tahmisoglu-yigit/R_Project.git

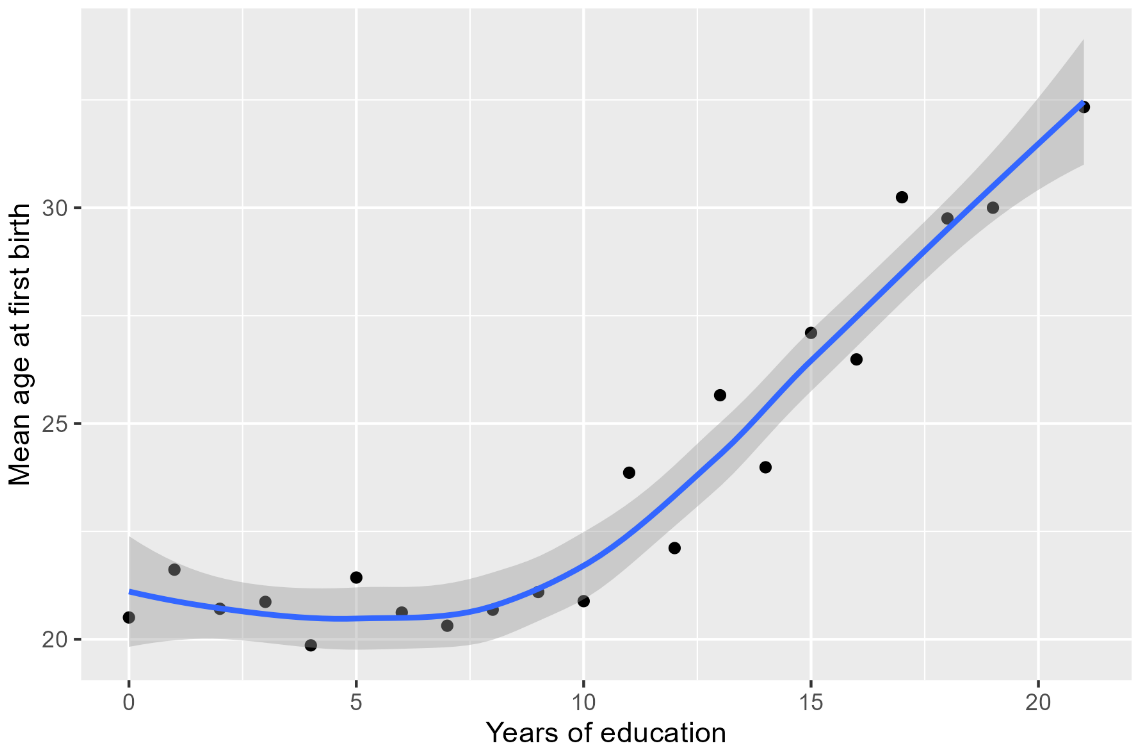
2. Executive summary

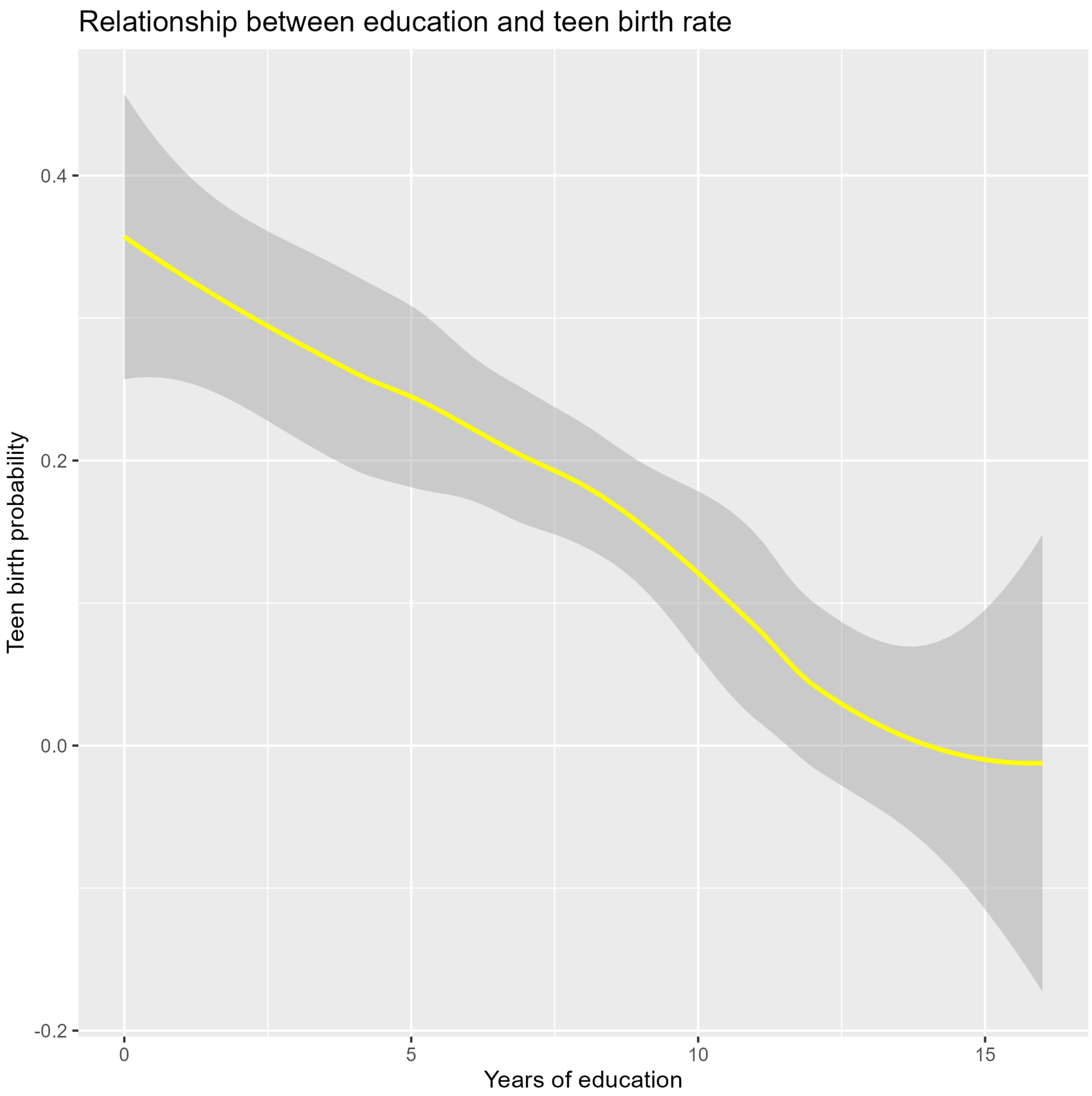
In my analysis, I aim to explore the relationship between years of education and teenage motherhood, as well as other background characteristics that may shed light on this relationship.

To conduct this analysis, I utilize cross-sectional survey data from the 2018 Demographic and Health Surveys for Turkey. This representative household survey provides comprehensive information on birth and individual records for women in each household in developing countries.

Through the use of simple linear regression and descriptive figures, I have discovered that increased years of education appear to be associated with a decreased likelihood of giving birth as a teenager. Moreover, the probability of teenage motherhood is lower among those in higher wealth categories and those residing in urban regions. However, it is essential to approach these observations with caution, as they do not necessarily indicate causality.




Mean age at first birth
30
25
20
0
5
10
15
20
Years of education



3. Summary of raw dataset

There are 7346 observations and 5271 variables in the raw dataset. The reason for such high number of variables is that, since this is a survey dataset there are many categorical variables and each unique value, which are answers by respondents, is counted as a separate variable. Each observation is a surveyed women between the ages 15 and 49.

```
##         CASEID V000 V001 V002 V003 V004    V005 V006 V007 V008
## 1  01010004 02  TR7  101    4    2  101 2356056   10 2018 1426
## 2  01010007 02  TR7  101    7    2  101 2356056   10 2018 1426
## 3  01010011 02  TR7  101   11    2  101 2356056   10 2018 1426
## 4  01010013 03  TR7  101   13    3  101 3133108   10 2018 1426
## 5  01010014 02  TR7  101   14    2  101 2356056   10 2018 1426
## 6  01010016 02  TR7  101   16    2  101 2356056   10 2018 1426
## 7  01010021 02  TR7  101   21    2  101 2356056   10 2018 1426
## 8  01020006 01  TR7  102    6    1  102 2356056   11 2018 1427
## 9  01030004 01  TR7  103    4    1  103 2356056   11 2018 1427
## 10 01030012 02  TR7  103   12    2  103 2356056   12 2018 1428

## Number of columns: 5271 ; Number of rows: 7346
```

4. Data cleaning

Firstly, as can be seen above, the variable names are not very descriptive. So, taking the DHS Manual as reference (https://dhsprogram.com/pubs/pdf/DHSG4/Recode7_Map_31Aug2018_DHSG4.pdf), I renamed the variable of interests to make further analyses easier. Dataset was mostly clean. Some of the steps I did was to check the data type of variables using str(). I also printed boxplots of certain variables to observe whether there are significant outliers. Further in the analysis, I removed NA's by "na.rm() = TRUE". And to make analysis more descriptive, I recoded unique values of categorical variables among my variable of interests.

Below is the subsample with all other variables dropped.

```
##         CASEID birth_month birth_year age age_5bin region residence educ_level
## 1  01010004 02           9       1990  28    25-29   West     Urban    Primary
## 2  01010007 02           9       1988  30    30-34   West     Urban     Higher
## 3  01010011 02          10       1989  29    25-29   West     Urban     Higher
## 4  01010014 02           2       1981  37    35-39   West     Urban     Higher
## 5  01010016 02           9       1983  35    35-39   West     Urban     Higher
## 6  01010021 02           7       1983  35    35-39   West     Urban     Higher
## 7  01030004 01           1       1974  44    40-44   West     Urban     Higher
## 8  01030012 02           9       1970  48    45-49   West     Urban     Higher
## 9  01040003 02          10       1986  32    30-34   West     Urban  Secondary
## 10 01040004 02           2       1981  37    35-39   West     Urban  Secondary
##    educ wealthindex
## 1     5      Richer
## 2    15     Richest
## 3    17     Richest
## 4    15     Richest
## 5    17     Richest
## 6    15     Richest
## 7    15     Richest
## 8    17     Richest
## 9    11      Middle
## 10   11      Richer

## Number of columns: 11 ; Number of rows: 5074
```

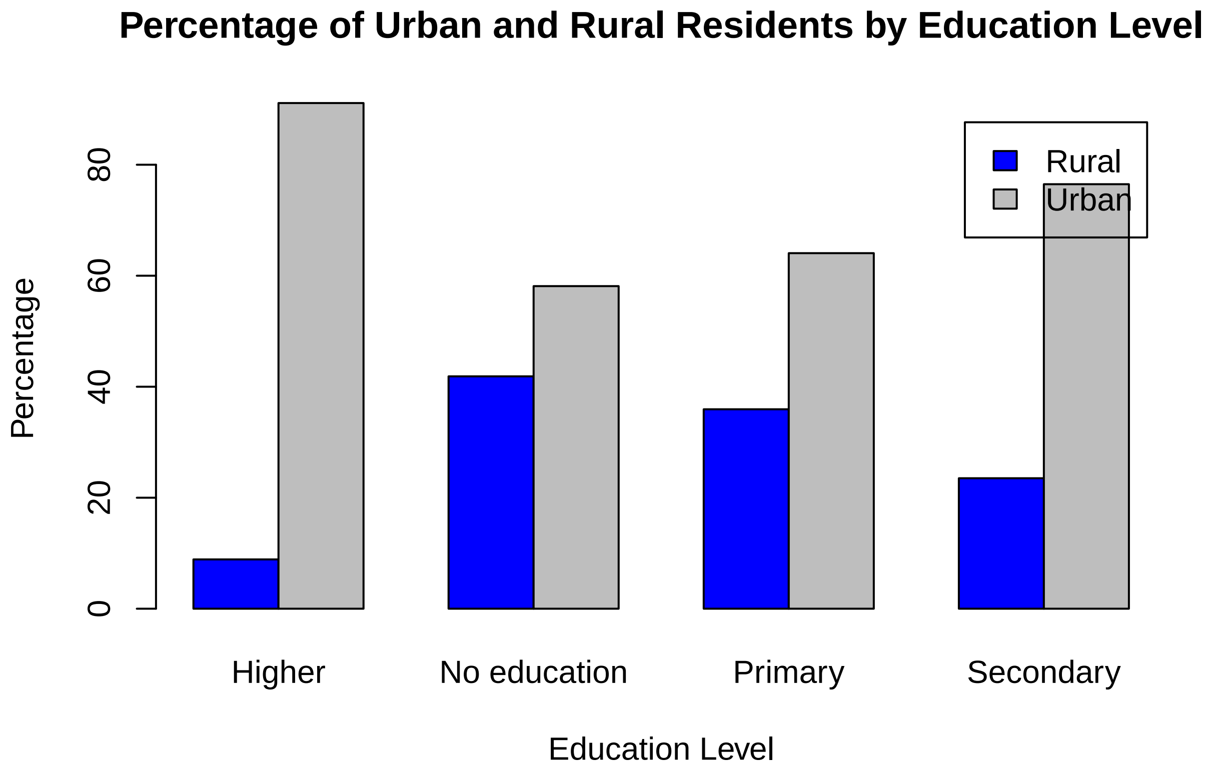
5. Data exploration, questions you tried to answer, interesting things.

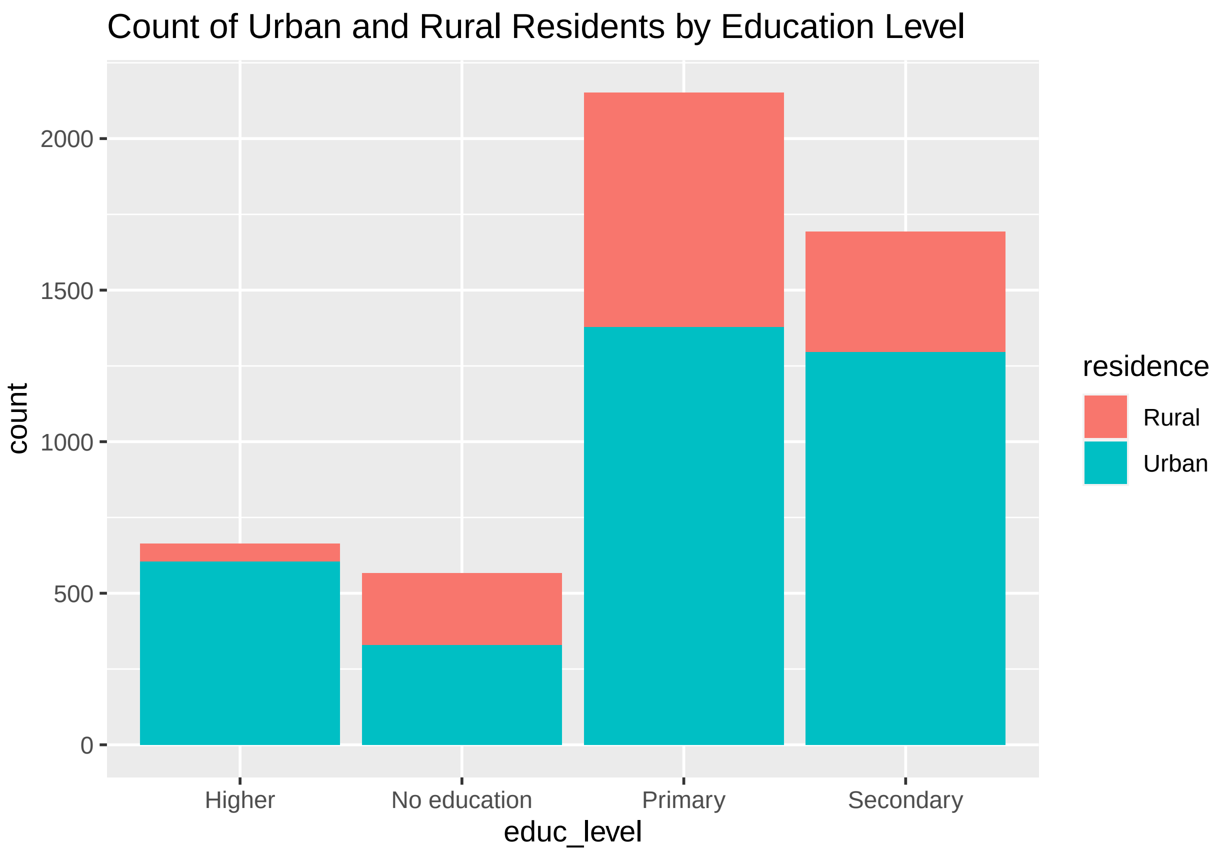
In order to identify women who gave birth as a teenager, I create a binary variable "teenbirth" =1 if age of women at first birth is below 18, and =0 otherwise. Using that, I try to answer the relationship between teenage motherhood and residence type, education level, years of education, and wealth.

Percentage of Urban and Rural Residents by Education Level

Below table and figure show that percentage of women who live in urban parts increase with the level of education.

```
##
##              Higher No education   Primary Secondary
##   Rural  8.885542    41.872792 35.936774 23.508565
##   Urban 91.114458    58.127208 64.063226 76.491435
```

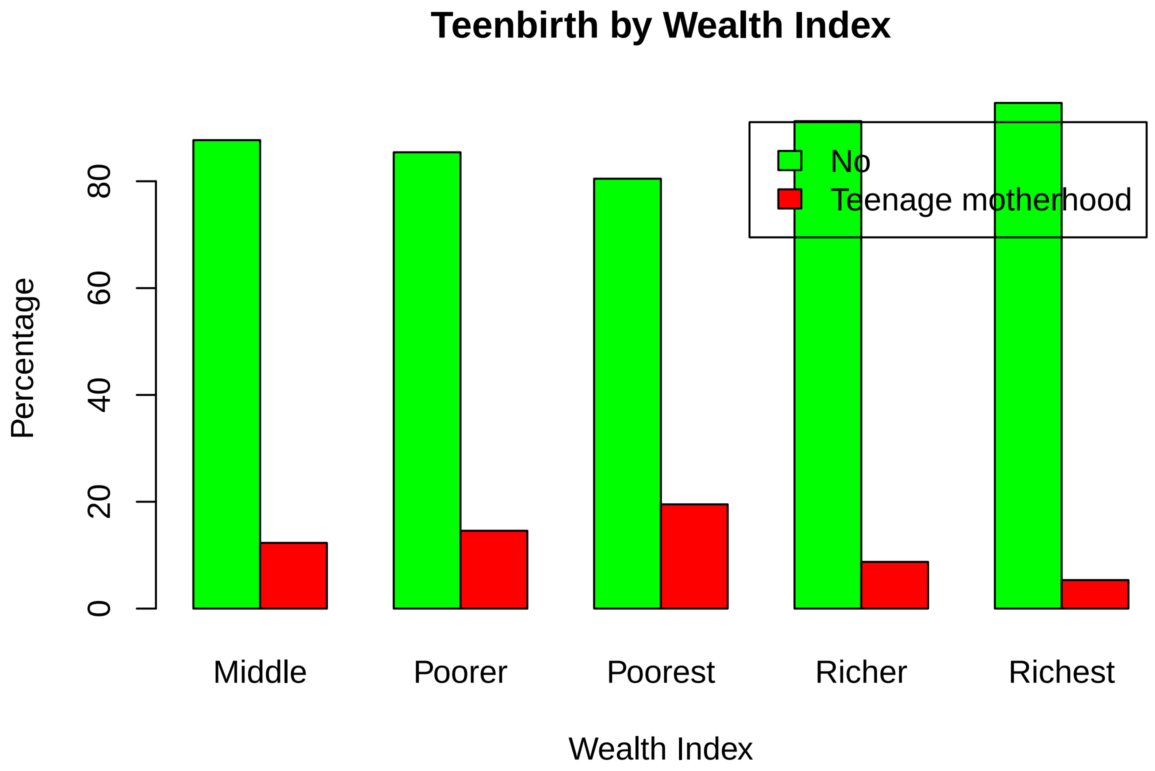




Teenage motherhood by Wealth Index

Below barplot show us that in Turkey, percentage of women who gave birth as a teenager increases as their economic status decreases.

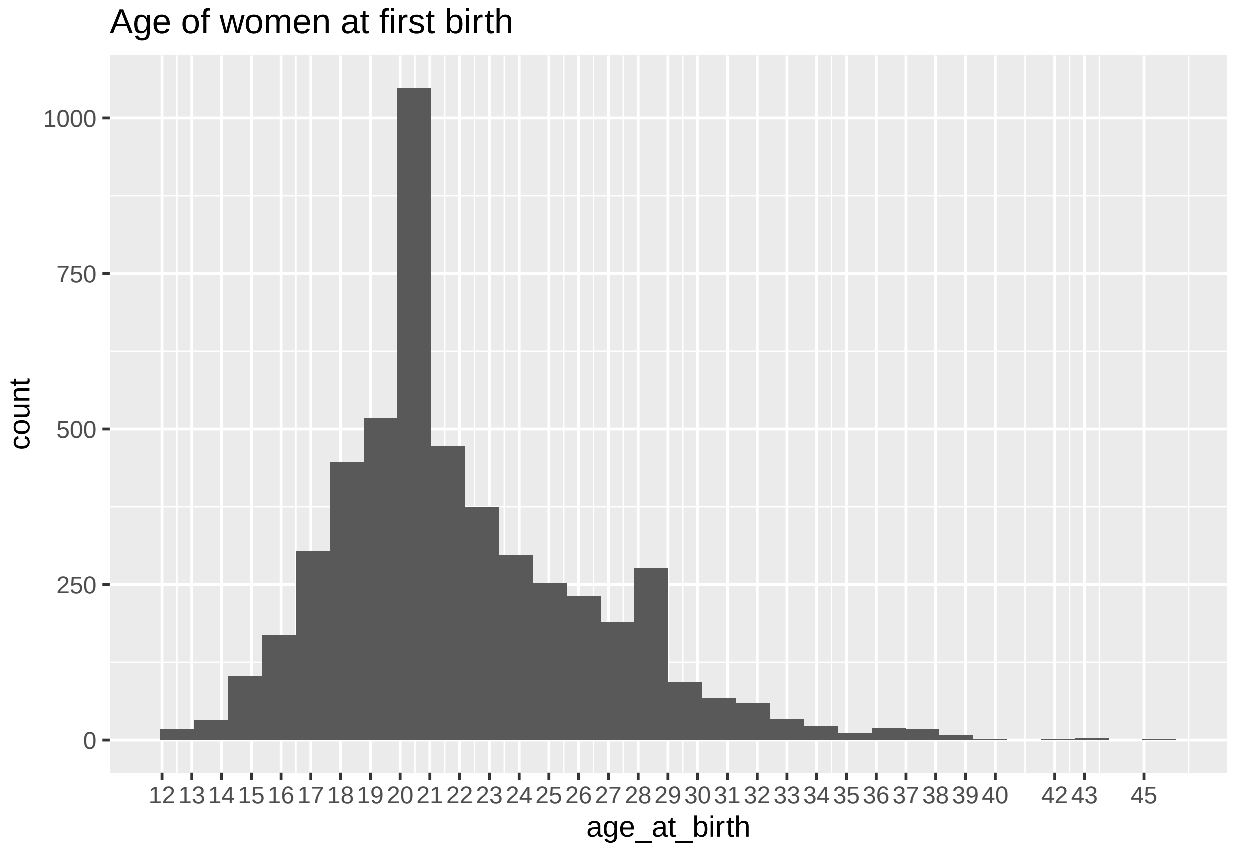
```
##
##        Middle    Poorer   Poorest    Richer   Richest
##   0 87.696850 85.428051 80.480769 91.251272 94.663821
##   1 12.303150 14.571949 19.519231  8.748728  5.336179
```



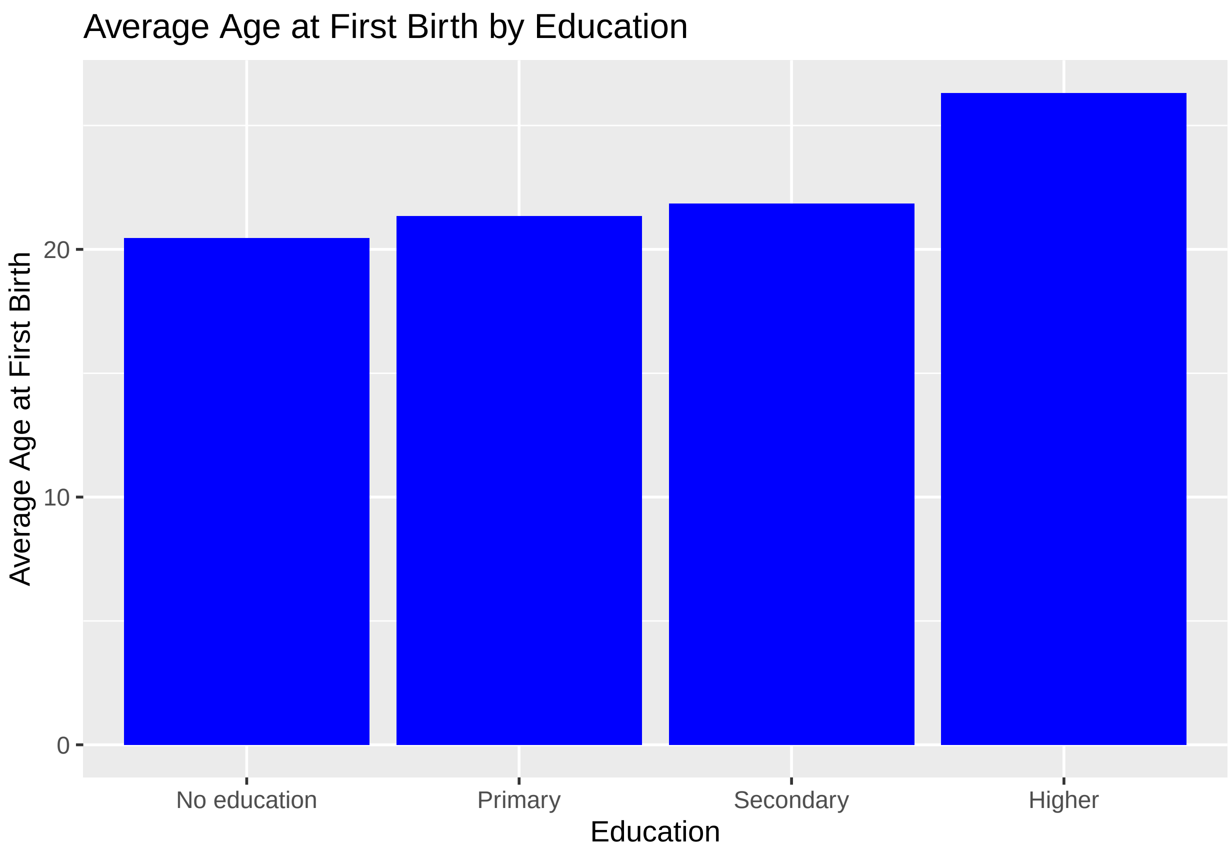
```
## pdf
##   2
```

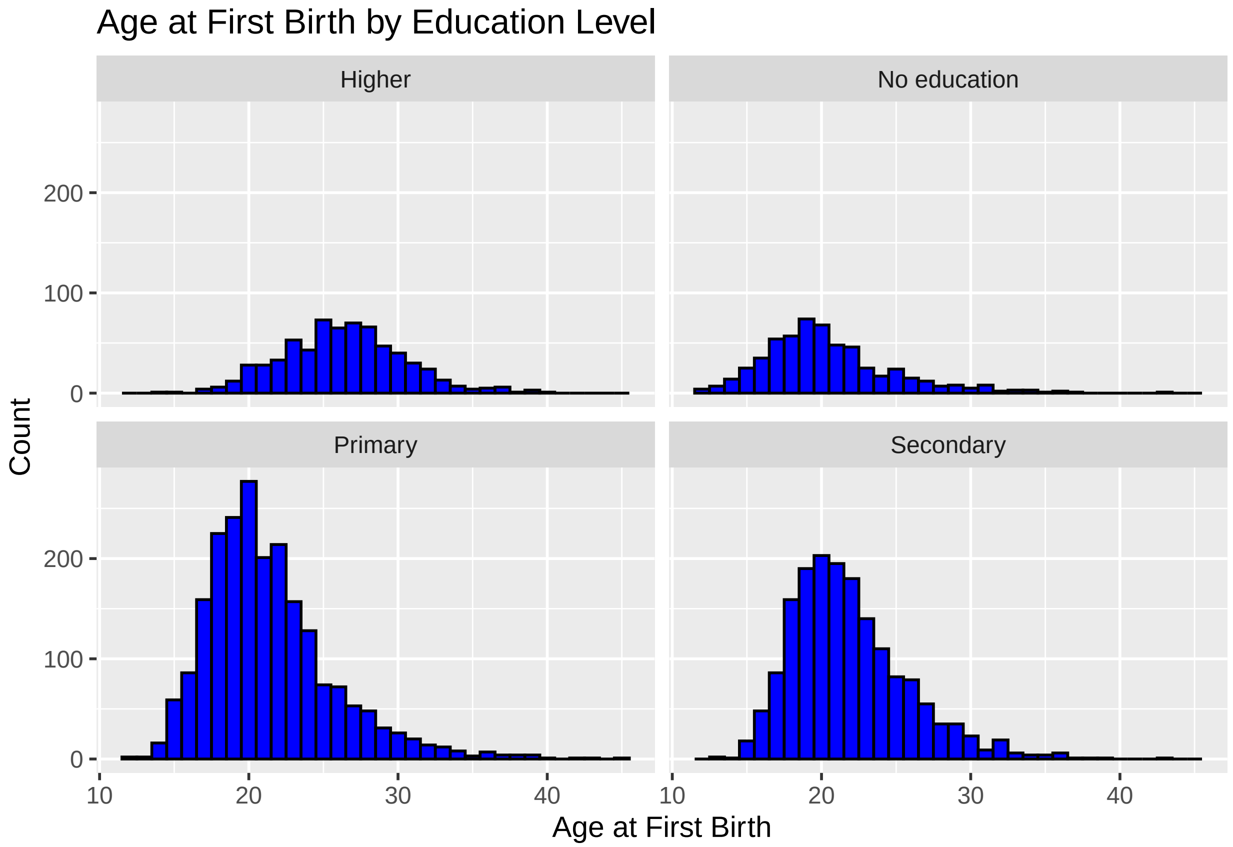
Age of women at first birth

The distribution of women by their age at first birth is skewed towards left, mainly accumulated between the ages 18 and 23. There is a sharp spike for women at the age of 21.

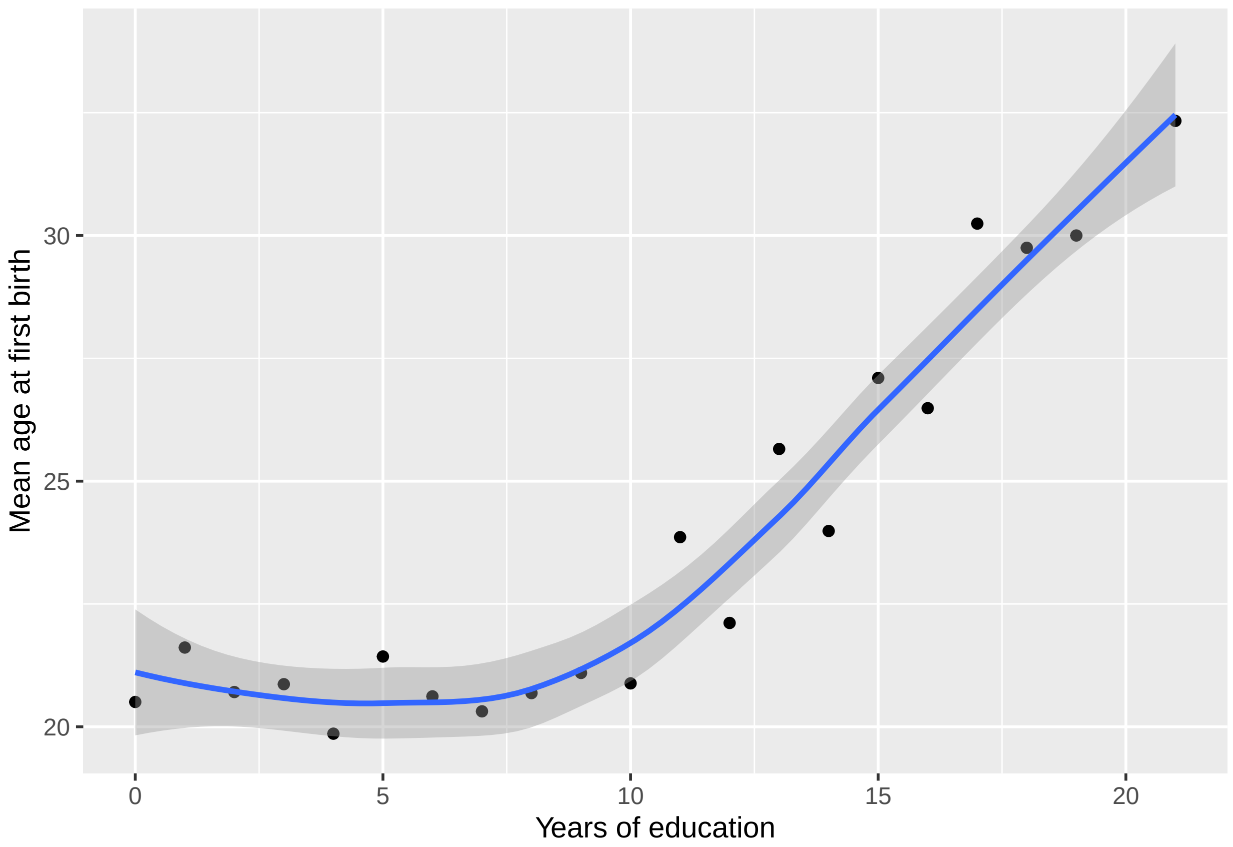


Also we can observe that as level of education increases, age of women at first birth decrease.

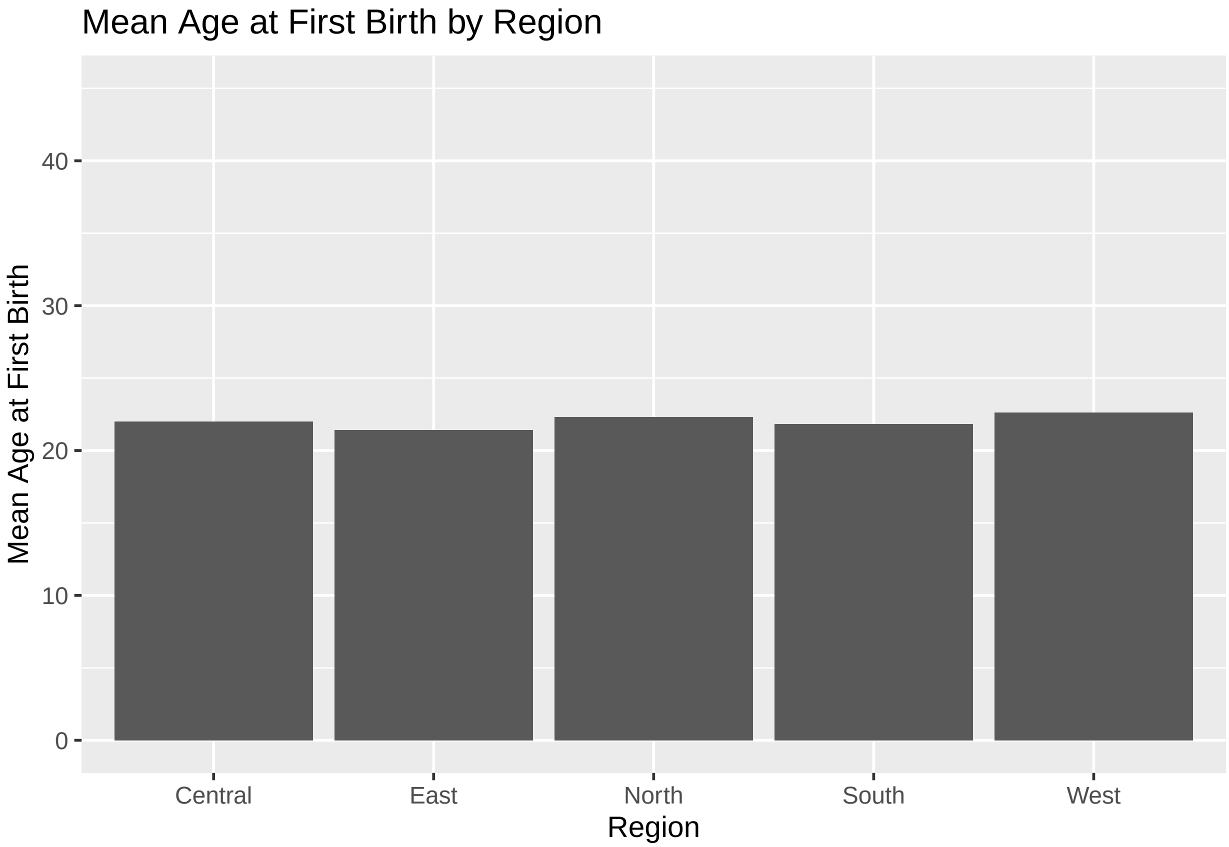

Average Age at First Birth by Education
Average Age at First Birth
0
10
20
No education
Primary
Secondary
Higher
Education



As years of education increase, per year of education the mean age of women at their first birth also increase.



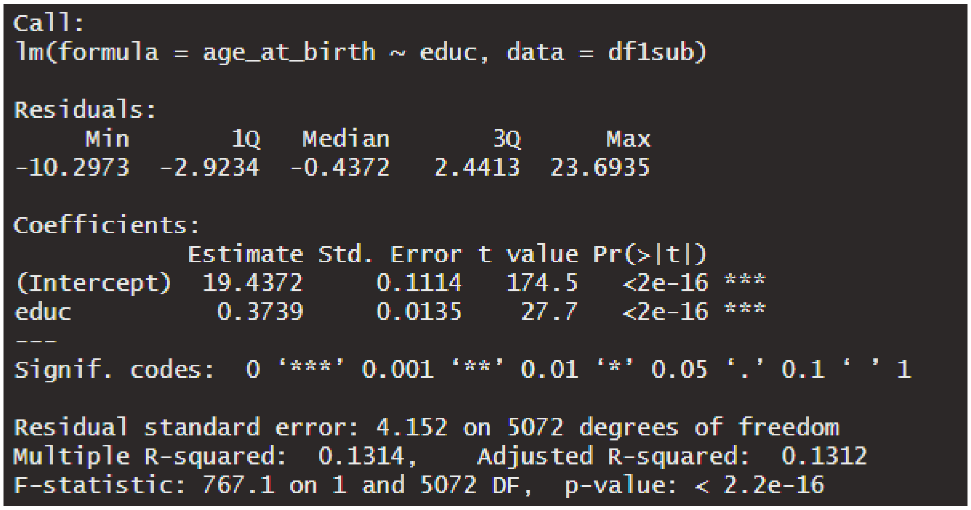
# Regional distribution: There are no significant differences between regions for mean age of women at first



birth.

Regression results:

Models show positive significant relationship between years of education and age at first birth. As we add more controls, R-squared value improve slightly.

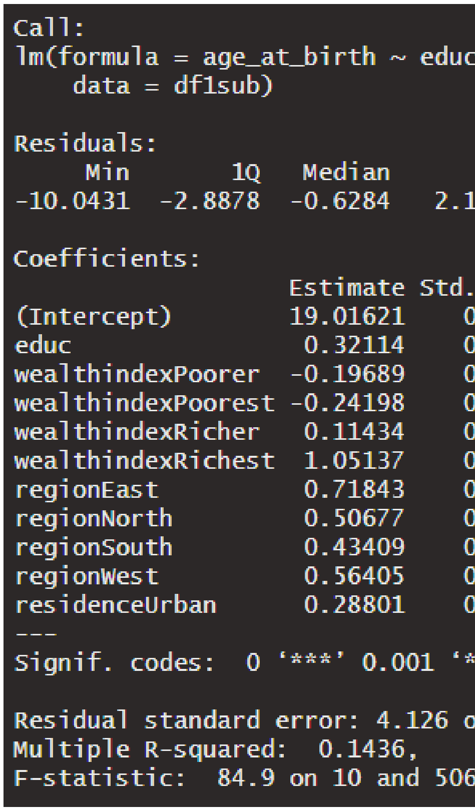


```
Call:
lm(formula = age_at_birth ~ educ, data = df1sub)

Residuals:
     Min       1Q   Median       3Q      Max 
-10.2973  -2.9234  -0.4372   2.4413  23.6935 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  19.4372     0.1114   174.5   <2e-16 ***
educ          0.3739     0.0135    27.7   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.152 on 5072 degrees of freedom
Multiple R-squared:  0.1314,	Adjusted R-squared:  0.1312 
F-statistic: 767.1 on 1 and 5072 DF,  p-value: < 2.2e-16
```



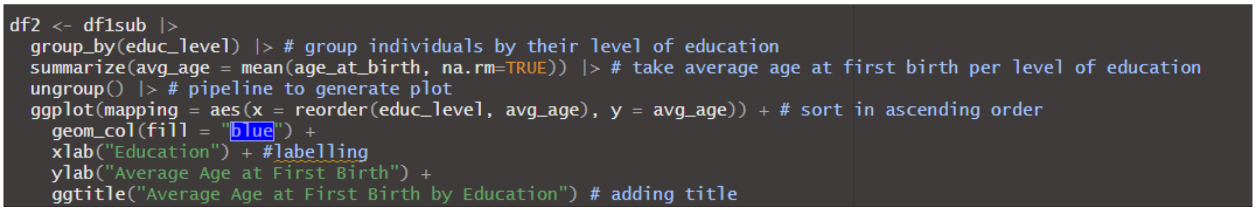
```
Call:
lm(formula = age_at_birth ~ educ
    data = df1sub)

Residuals:
     Min       1Q   Median       
-10.0431  -2.8878  -0.6284   2.1

Coefficients:
                   Estimate Std.
(Intercept)        19.01621    0
educ                0.32114    0
wealthindexPoorer  -0.19689    0
wealthindexPoorest -0.24198    0
wealthindexRicher   0.11434    0
wealthindexRichest  1.05137    0
regionEast          0.71843    0
regionNorth         0.50677    0
regionSouth         0.43409    0
regionWest          0.56405    0
residenceUrban      0.28801    0
---
Signif. codes:  0 ‘***’ 0.001 ‘*

Residual standard error: 4.126 o
Multiple R-squared:  0.1436,	
F-statistic:  84.9 on 10 and 506
```

6. Explain part of code:



```
df2 <- df1sub |>
  group_by(educ_level) |> # group individuals by their level of education
  summarize(avg_age = mean(age_at_birth, na.rm=TRUE)) |> # take average age at first birth per level of education
  ungroup() |> # pipeline to generate plot
  ggplot(mapping = aes(x = reorder(educ_level, avg_age), y = avg_age)) + # sort in ascending order
    geom_col(fill = "blue") +
    xlab("Education") + #labelling
    ylab("Average Age at First Birth") +
    ggtitle("Average Age at First Birth by Education") # adding title
```