Stanford's RA programming assessment

Maria Cruz

Task 2

After reshaping the data base of mothers to births and created the binary variable for infant death, I calculated the total births as the total of rows in the sample that were not missing values: the result was 19,644 total births. For total deaths, I calculated the total of rows in the sample that were not missing and that had a value equal to one for the infant death binary variable: it resulted in 2,018 total deaths.

```
birth_data %>%
  dplyr::filter(is.na(idx) == F) %>%
  nrow(.)
```

```
## [1] 19644
```

```
birth_data %>%
  dplyr::filter(is.na(idx) == F & infant_death == 1) %>%
  nrow(.)
```

```
## [1] 2018
```

Task 3

I ran the linear regression below:

```
birth_data %>%
  fixest::feols(infant_death~v191, vcov = 'hetero') %>%
  fixest::etable(.)
```

```
##                                    .
## Dependent Var.:         infant_death
##
## (Intercept)      0.0162*** (0.0004)
## v191            -0.0068*** (0.0003)
## _______________ ___________________
## S.E. type       Heteroskedast.-rob.
## Observations                126,880
## R2                          0.00299
## Adj. R2                     0.00298
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

First, the results tell us that there is a negative relationship between infant deaths and the wealth score, therefore the poorer is the household the probability for infant death is greater. Although there is a statistically significant relationship between these two variables, the regression can not tell us how much causality there is, if there is some, between infant deaths and the wealth level of the household, and also the linear probability model can predict the magnitude whith deviations.

Task 4

If I found that higher infant mortality was associated with higher wealth score, I could try a nicer specification, controlling for endogeneity with covariates, but if the point was to understand this simple correlation I would try to understand the frequency of the socio-demographic groups in the sample.

Task 5

The regression below shows that the effect of wealth on infant mortality does not depend on the gender of the child, beacuse its interaction is not statistically significant. But the sex of the child impacts infant mortality.

```
##                                    .
```

```
## Dependent Var.:         infant_death
##
## (Intercept)       0.1070*** (0.0030)
## v191             -0.0276*** (0.0030)
## sex              -0.0182*** (0.0042)
## v191 x sex           0.0062 (0.0042)
## _______________ ___________________
## S.E. type        Heteroskedast.-rob.
## Observations                  19,644
## R2                           0.00704
## Adj. R2                      0.00689
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Task 6

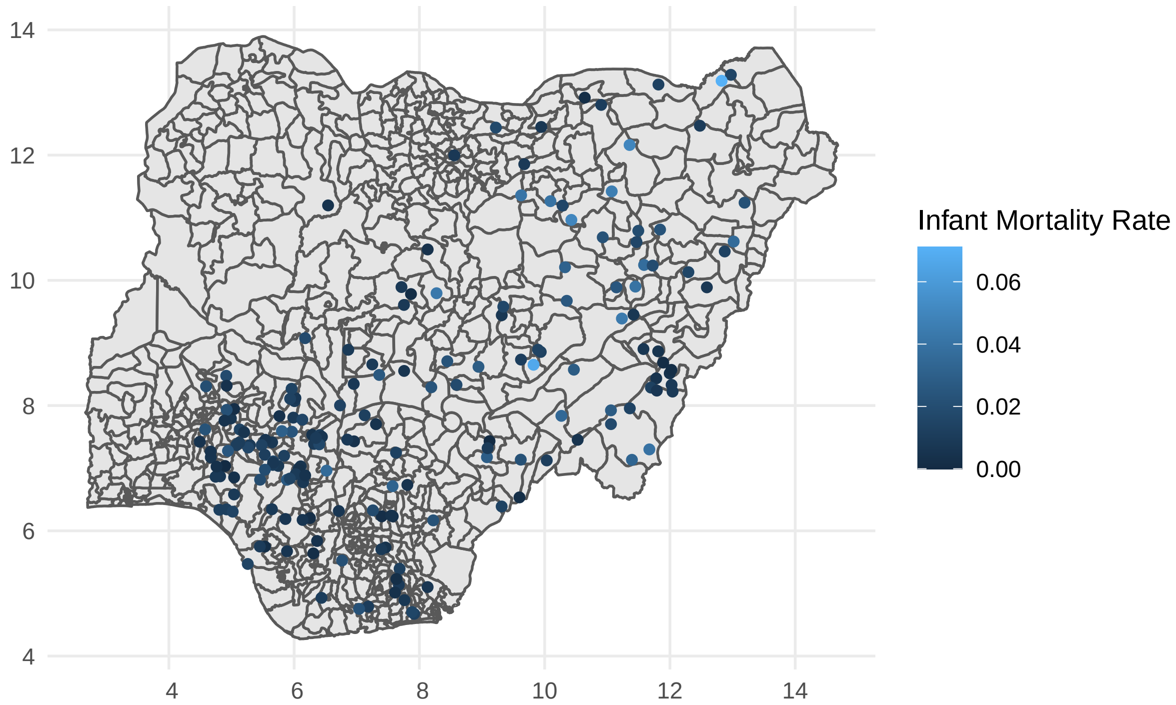
Figure 1: Scatter plot of village−average wealth score versus village−average probability of infant death

The average wealth in the sample is 116.1545 and the average infant mortality in the sample is 0.01590479.

Task 7

In this task I used the shapefile available in the World Bank Data website. Some of the village points in the locations.csv file where outside of the shapefile for Nigeria, so I preferred to use the intersection of these, therefore, there is less villages in the Figure 2 compared to the number of villages in the survey.

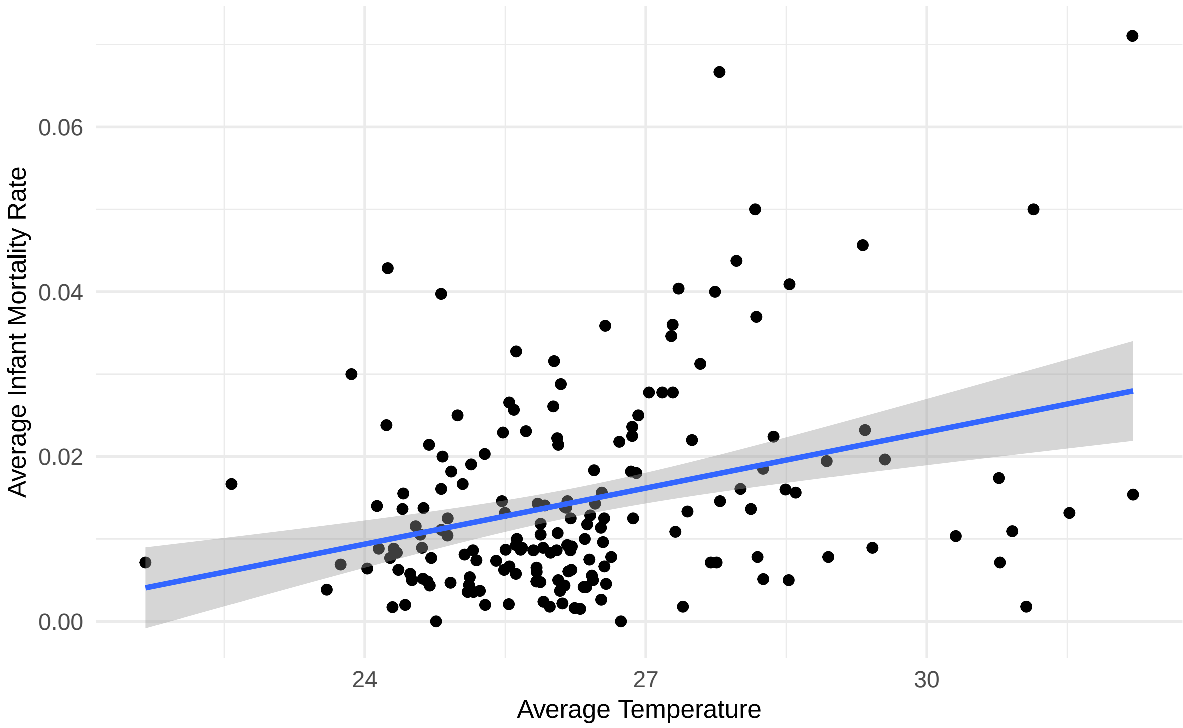
Figure 2: Infant Mortality Rate of villages in Nigeria



Task 8

I downloaded the data of average temperature in 10m and the website returned a zip file with TIFF files for each month, then I selected the month 6 (June). The locations of these file were not the same as the locations for the villages, so I created a variable that calculates the spatial distance of all points and then I used the point with minimum distance of the WorldClim data as the point for the village.

Figure 3: Relationship between average temperature on June and average infant mortality rate in each village of Nigeria



It is possible to see in Figure 3 that higher average temperatures are associated with higer average infant mortality rates, which is something possibly already accepted in literature and could be associated with natural causes from great heat or even lack of access of resources – like clean water.

Task 9

The scatter plot by itself can not show if this is exactly a linear relationship, but it can be tested using a Loess Regression to see if the specification of a local regression can describe a better asscociation between the variables. Very low average temperatures do not seem to be necessarily better for infant mortality rate. Median average temperatures are better for average infant mortality rate.

Figure 4: Loess Regression between average temperature on June and average infant mortality rate in each village of Nigeria