Data cleaning

Research Question

My interest in this area came from multiple new information I've received recently. Since I have been taking Asian studies courses and came to learn how unprivileged some countries are, I have been wanting to learn more about them. Another cause of the interest is I have been learning more about woman's birth stories and the amount of preparation and care they need to deliver a baby (or even more) safely. Recently, this topic has been on my mind so I thought it would be a great opportunity to do more research on this area.

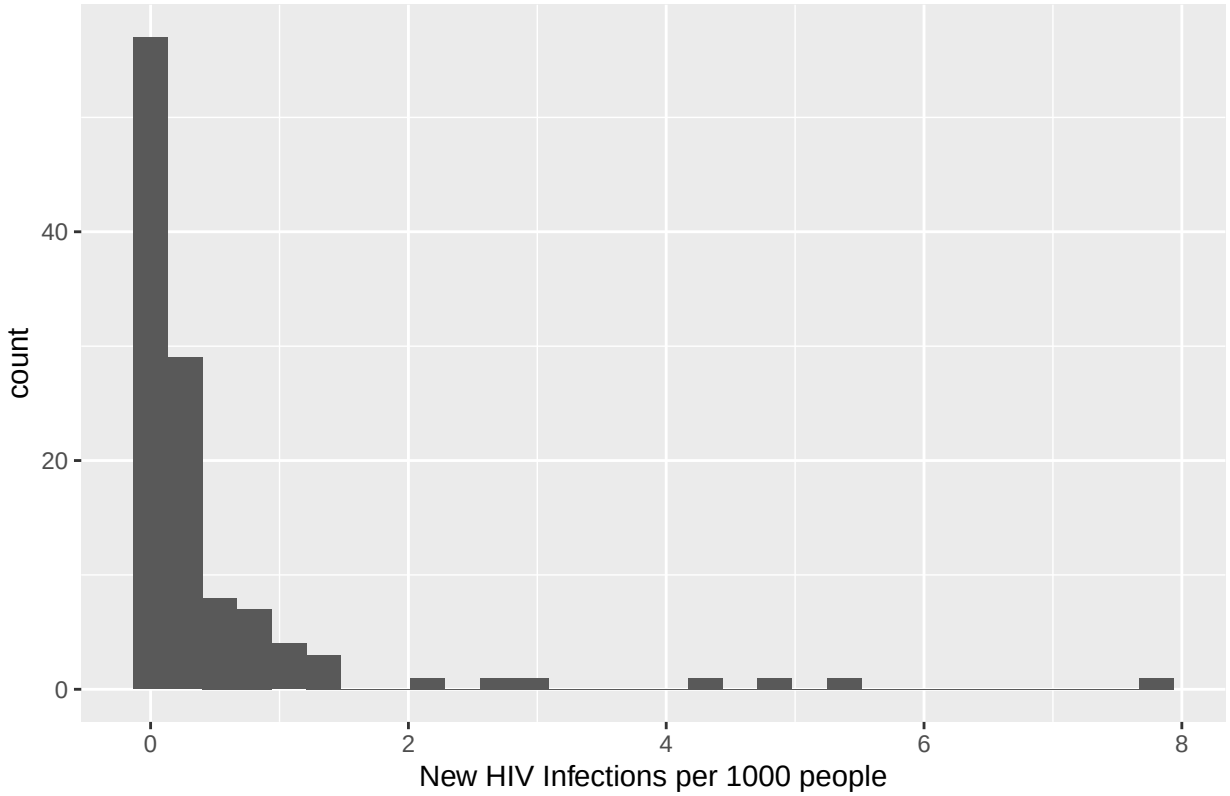
Background

I looked up the World Health Organization website regarding the topic of maternaity morality rate and possible factors to affect it. According to the article, factors including but not limited to: severe bleeding, infection, high blood pressure, complications from delivery, and unsafe abortion, could affect the rate of maternity mortality. World Health Organization. (2019, September 19). Maternal mortality.

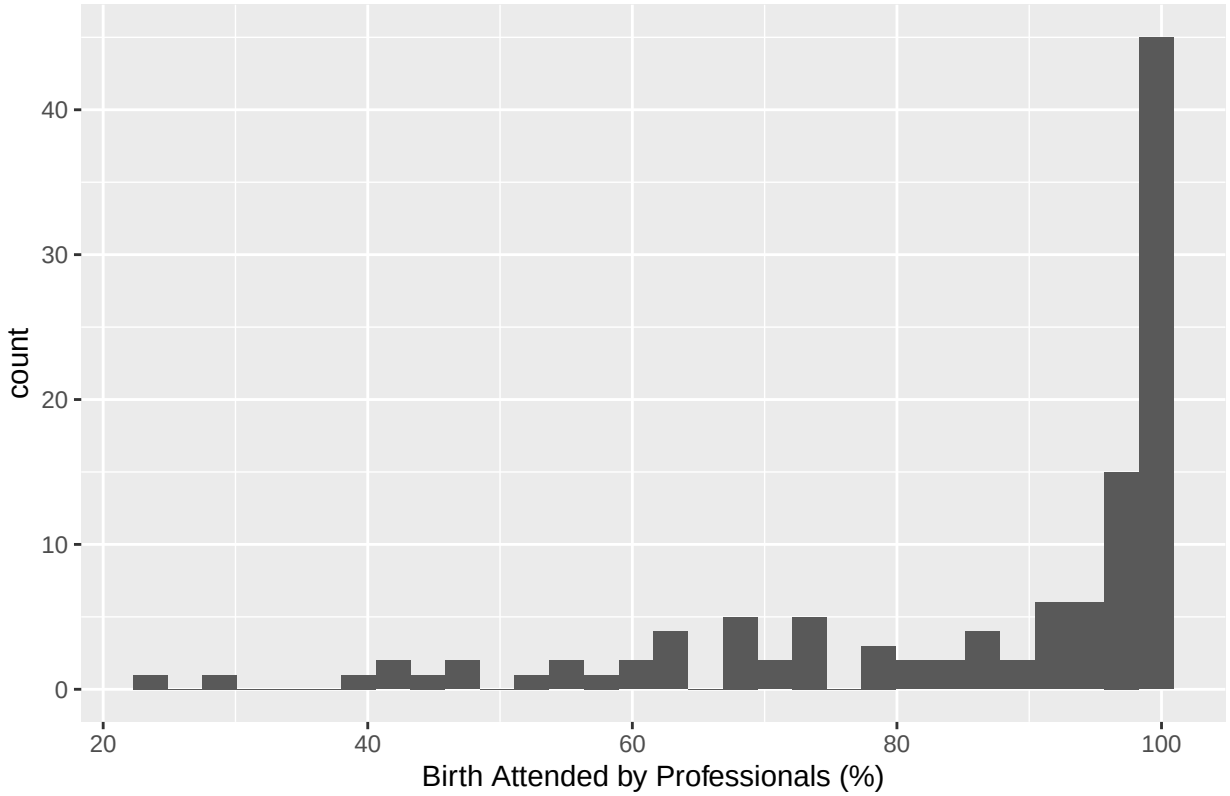
Data Source

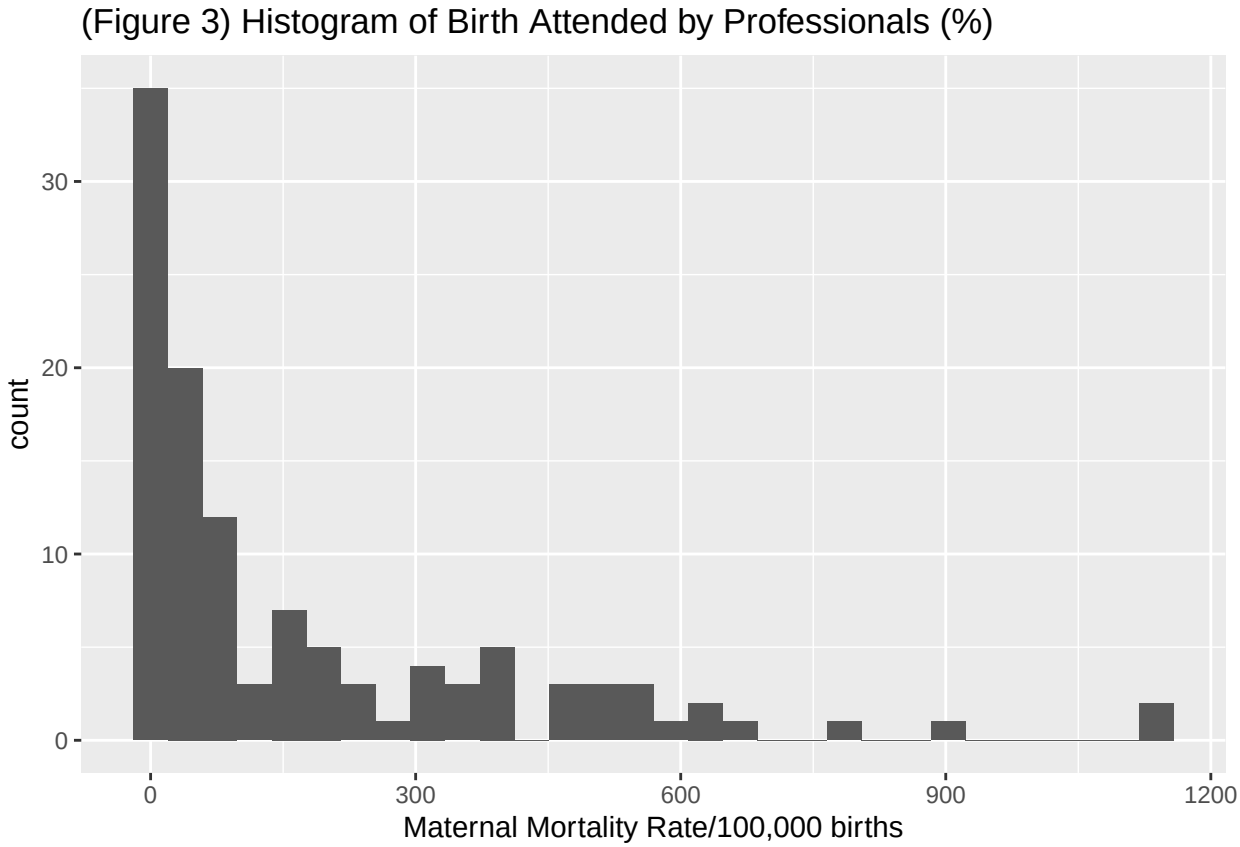
The original dataset has multiple tabs included. We're only focusing on the ICPD_health tab in this project. In the dataset, there are multiple columns with various factors included. We're only keeping a few to conduct the research. From the WHO website I visited, the factors included in the dataset that are related to the factors mentioned are: Percentage of births attended by skilled health personnel and new HIV infections.

(Figure 1) Histogram of New HIV Infections per 1000 people



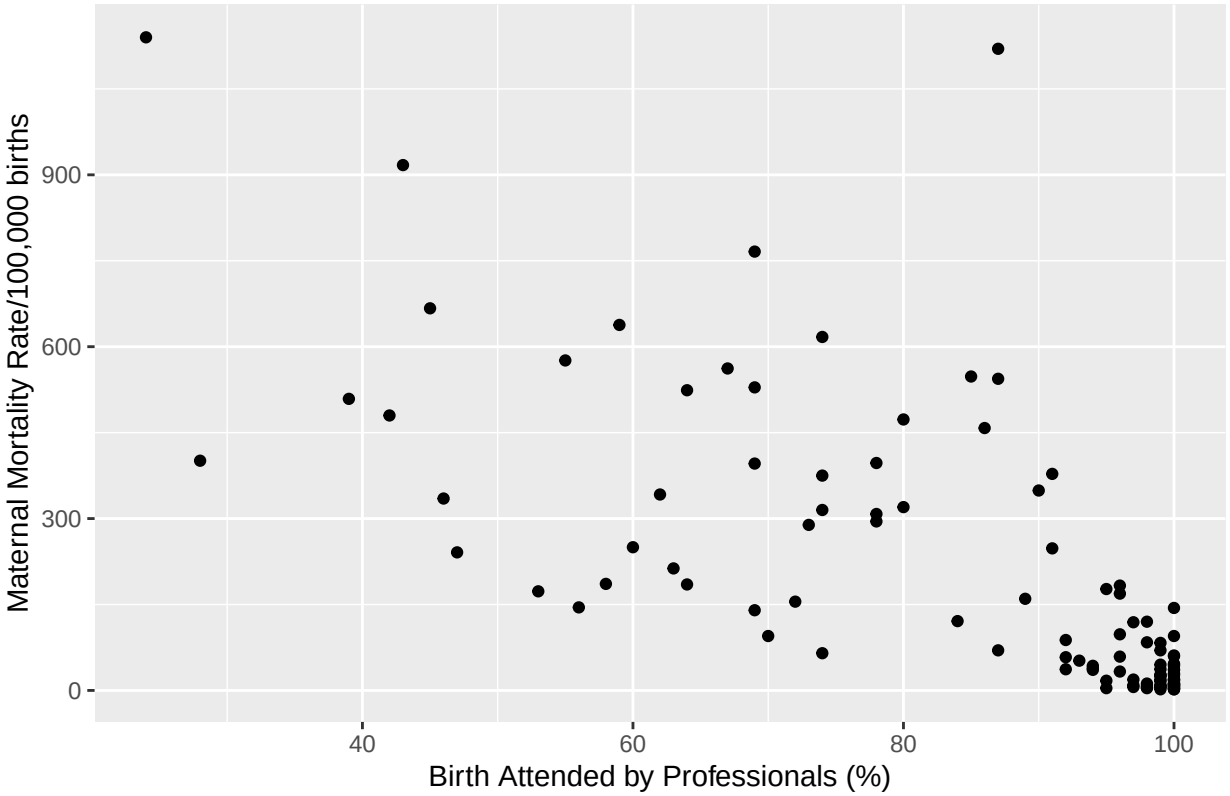
(Figure 2) Histogram of Birth Attended by Professionals (%)

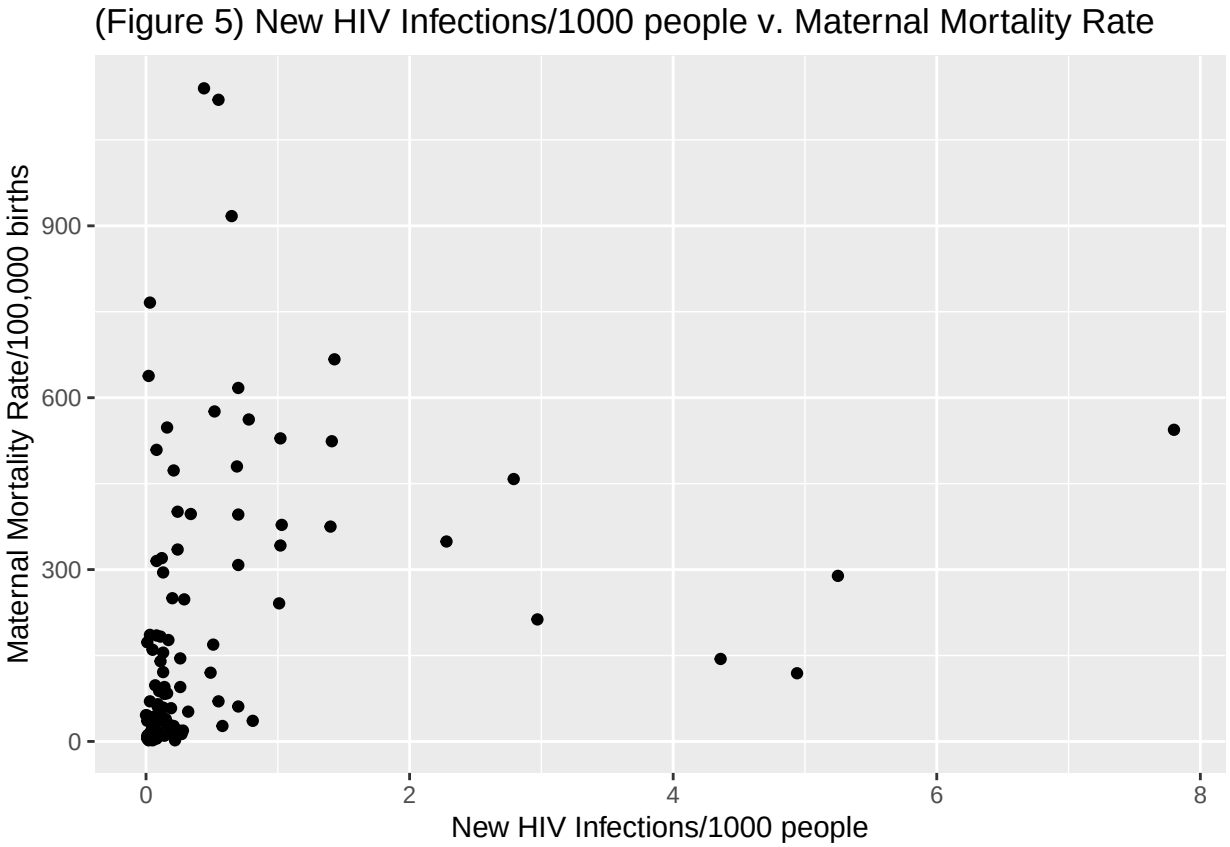




In Figure 1 we can see the graph is skewed to the right with possible outliers on the right side of the graph. This figure has a downward sloping trend. In Figure 2 we can see the graph is skewed to the left with possible outliers around 20% of birth attended by professionals. In figure 3, the graph is again skewed to the right with possible outlier on the right side of the graph, around 1200 mortality at birth per 100,000 births. We also want to note that locations without the presence of all three data we need are deleted. Therefore, the skewedness in these graphs could be caused by the absence of those data.

(Figure 4) Birth Attended by Professionals (%) v. Maternal Mortality Rate





In figure 4, there is a weak negative linear relationship between the two factors. The scatterplot clutters around 100% birth attended by professionaals, In figure 5, there's a weak linear relationship between the two factors with a clutter around 0 new HIV infections per people in relation to maternal mortality rate.

My research question can be used using linear models because there are linear relationships (but weak) between variables. But some issues worth mentioning is the data we have is only from one year. Perha[s if we increase the number of years in the data the linear relationship would be stronger.

Reference

World Health Organization. *Maternal mortality* Retrieved October 22, 2021, from [https://www.who.int/ news-room/fact-sheets/detail/maternal-mortality#:~:text=The% 20major%20complications%20that%20account,(pre%2Dec //www.who.int/ news-room/fact-sheets/detail/maternal-mortality#:~:text=The%20major%20complications%20that%20acc (pre%2Declampsia%20and%20eclampsia)