Statistical Epidemiology Project - Deaths due to diarrhoea and rotavirus vaccine coverage in children state wise

Group Project

October 2022

Project Members

1. Parth Sanjay Gandhi, 215280004
2. Saubhagya Kakkar, 215280009
3. Sanjyot Shenoy, 18B030023

1 Background of Study

1.1 Information on Rotavirus

Rotavirus is a virus (more like a class of viruses) that causes diarrhoea and other intestinal symptoms. It's very contagious and is the most common cause of diarrhea in infants and young children worldwide. It has a spherical structure when looked under the microscope, hence earning its name "Rota"virus, which from the Latin word for wheel - *rota*. It is so contagious that nearly every child in the world is infected with a rotavirus at least once by the age of five.

1.2 Rotavirus Vaccination

Despite availability of several proven solutions for diarrhea, no single solution is sufficient for the prevention and treatment of Rotavirus diarrhoea. Vaccination is one of the approaches through which Rotavirus is being fought against in India. Currently the vaccines which are recommended by WHO for India are of 3-dose regimen.

Available evidences suggest that Rotavirus vaccines are most effective at preventing the most severe and life-threatening cases of Rotavirus diarrhea. The efficacy of Rotavirus vaccines against severe Rotavirus diarrhea ranges from 40-60%.

As per WHO position paper (2013), there is also some evidence that Rotavirus vaccination leads to herd protection in unvaccinated older children and adults.

1.3 Aim of the Study

In this study we attempt to gain the following insights

1. Get a preliminary look at the data.
2. Perform some basic spatial analysis and compute spatial auto-correlation indexes.
3. Perform spatial regression to see if the vaccine coverage matters or not.

2 Methods

2.1 Data Collection

The data was obtained through Government of India's NHM Website which is a free service giving wealth of health related information. More data was also obtained through IndiaStat, a paid service.

Census data was obtained through the Census of India's website.

For this study we will be using information on Rotavirus immunization programmes obtained through Ministry of Health and Family Welfare, Govt. of India.

2.2 Assumptions

Since Rotaviral Diarrhoea is usually the main cause of fatal diarrhoea in children, we are making the following assumption in our study -

All Diarrhoea-based deaths in children of ages 0-5 are caused through Rotavirus infections.

Since Rotavirus is extremely contagious, this assumption can be used for statistical inference of diarrhoea deaths.

Regarding the Geospatial data, we will be making the following assumptions -

1. Jammu and Kashmir treated as one entire state. (Ladakh and Jammu not separate)

3 Basic Analysis

To get a preliminary idea of our data, we plotted our raw data on the map to get a look at how the raw children (ages 0-5) diarrhoea counts, death counts and rotavirus immunizations are across all the states. This has been done for three years. So the maps are as follows

3.1 Diarrhoea counts in Children

We plot the following maps in this part of analysis -

1. Diarrhoea in Children (Ages 0-5) across states in the year 2017-18.
2. Diarrhoea in Children (Ages 0-5) across states in the year 2018-19.
3. Diarrhoea in Children (Ages 0-5) across states in the year 2019-20.

3.2 Diarrhoea Death counts in Children

We plot the following maps in this part of analysis -

1. Deaths due to diarrhoea in Children (Ages 0-5) across states in the year 2017-18.
2. Deaths due to diarrhoea in Children (Ages 0-5) across states in the year 2018-19.
3. Deaths due to diarrhoea in Children (Ages 0-5) across states in the year 2019-20.

In this map, both the above prevalence and the death counts are shown together to get a comparative idea.

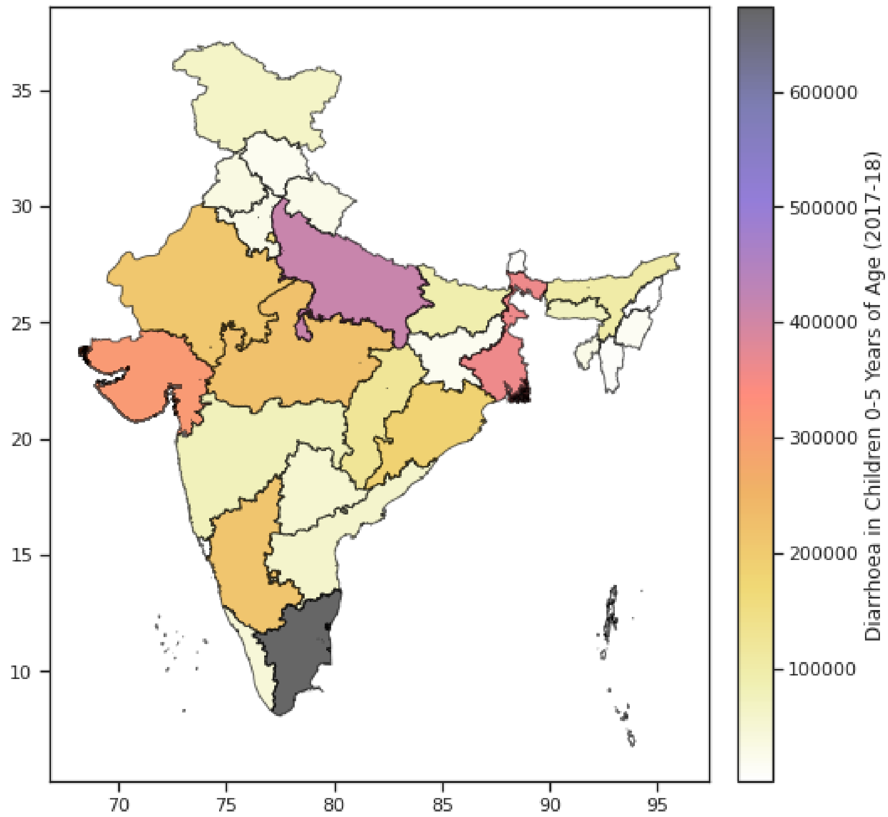

Figure 1: Diarrhoea in Children (Ages 0-5) across states in the year 2017-18.

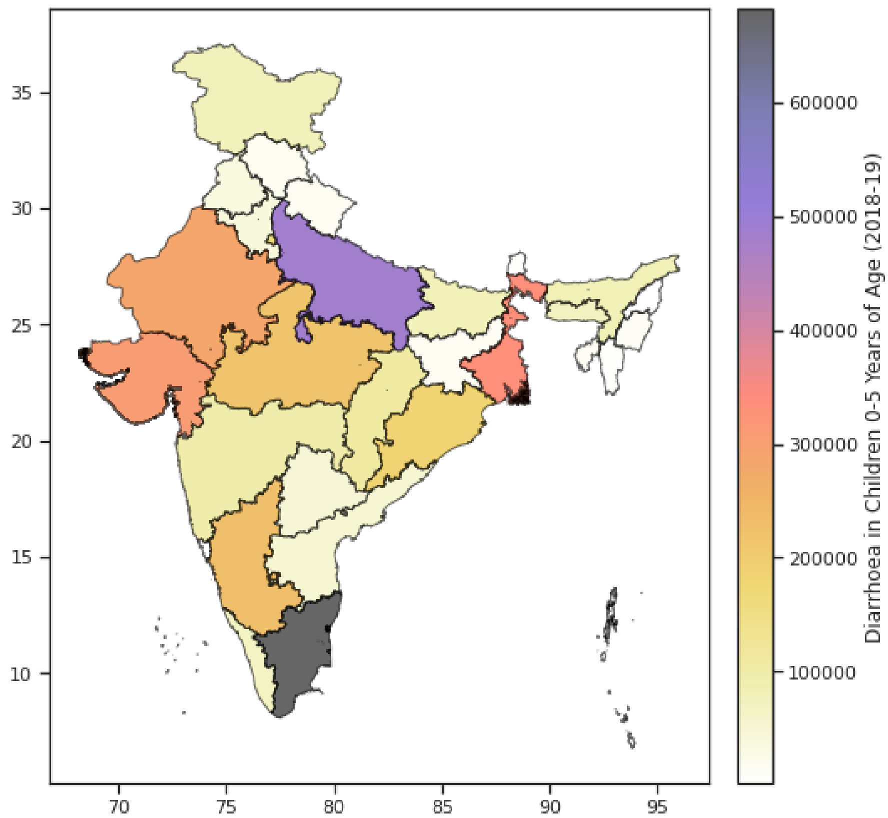

Figure 2: Diarrhoea in Children (Ages 0-5) across states in the year 2018-19.

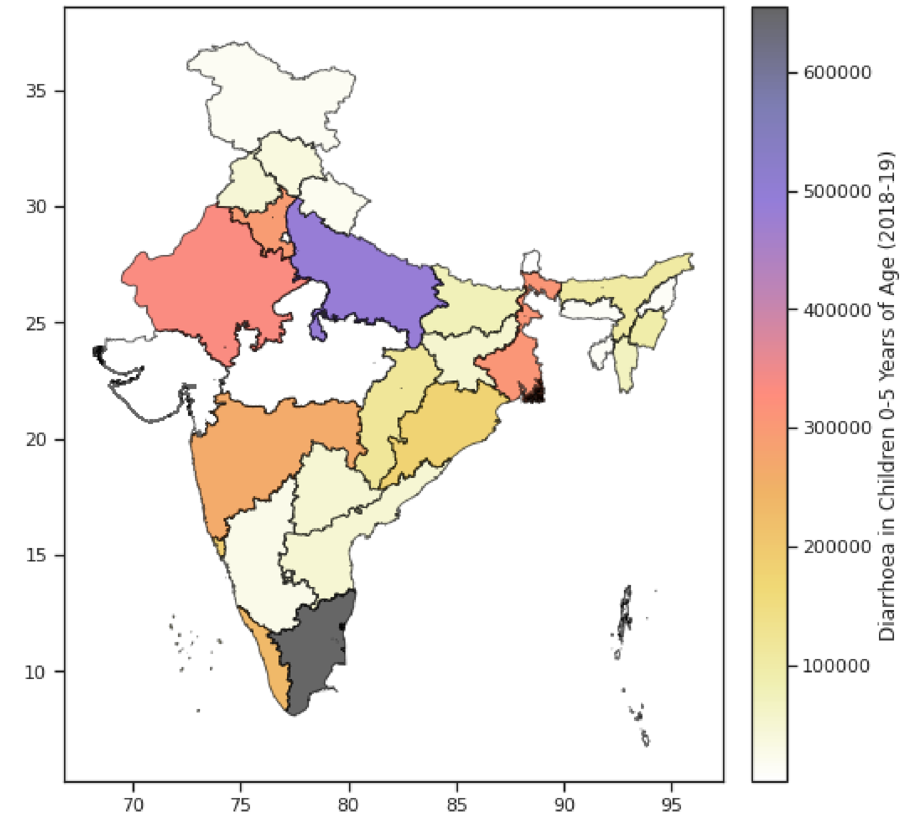

Figure 3: Diarrhoea in Children (Ages 0-5) across states in the year 2019-20.

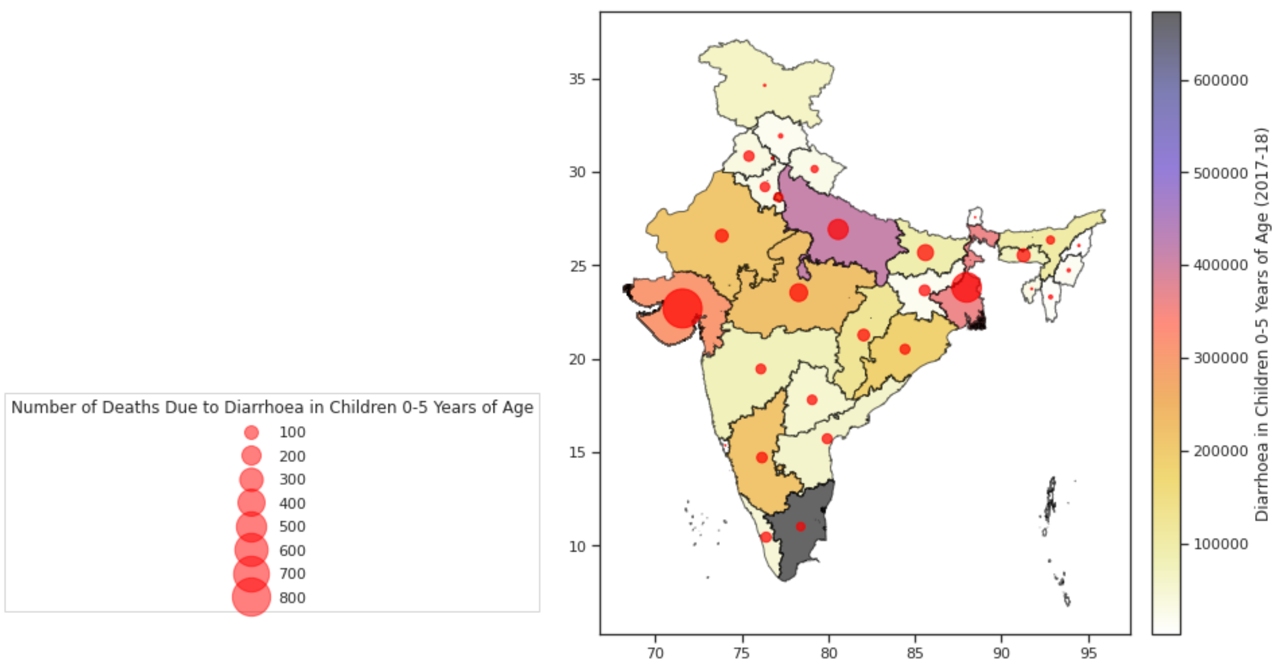

Figure 4: Deaths due to diarrhoea in Children (Ages 0-5) across states in the year 2017-18.

3.3 Rotavirus Immunization in Children

We plot the following maps in this part of analysis -

1. Children (Ages 0-5) who received Rotavirus vaccine first dose across states in the year 2017-18.
2. Children (Ages 0-5) who received Rotavirus vaccine first dose across states in the year 2018-19.

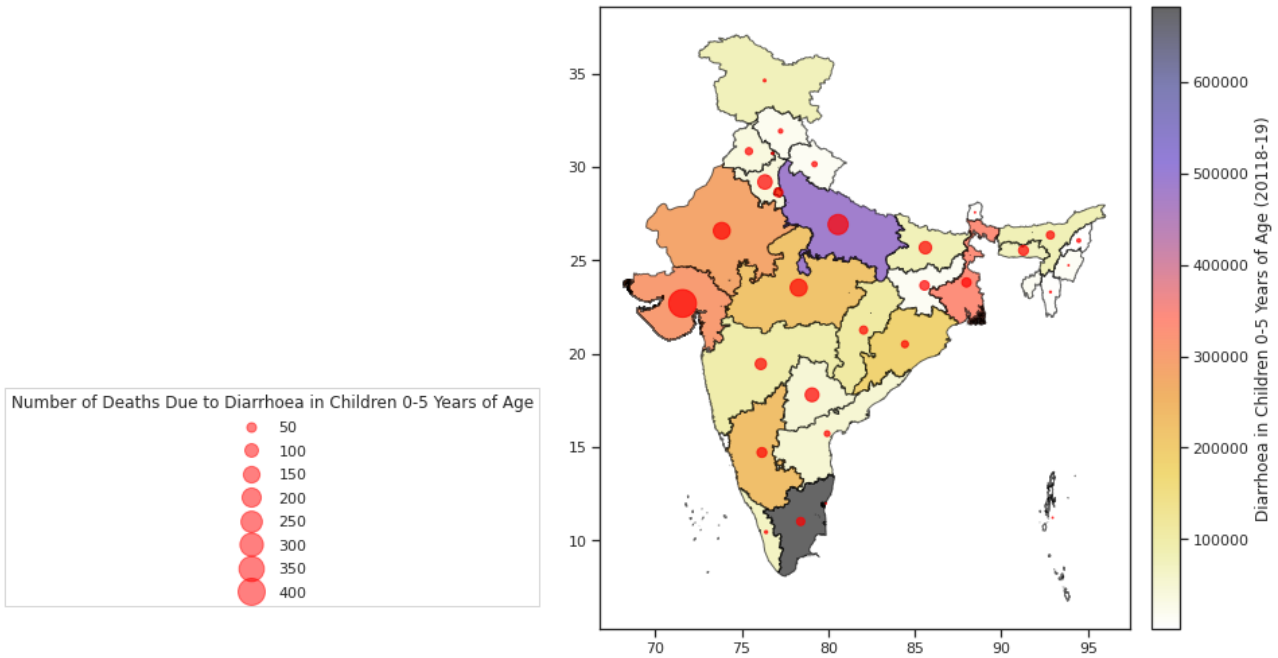

Figure 5: Deaths due to diarrhoea in Children (Ages 0-5) across states in the year 2018-19.

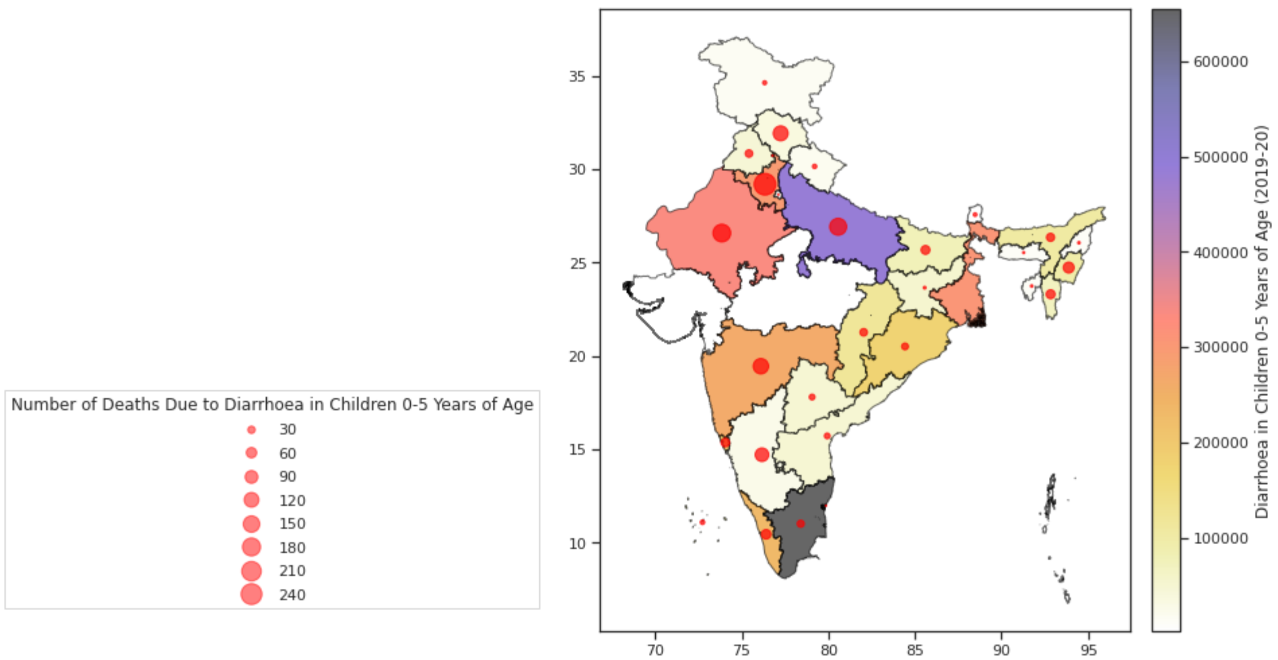

Figure 6: Deaths due to diarrhoea in Children (Ages 0-5) across states in the year 2019-20.

3. Children (Ages 0-5) who received Rotavirus vaccine first dose across states in the year 2019-20.

3.4 Inference from the above three types of maps

First, we note that Tamil Nadu and Uttar Pradesh have a high number of diarrhoea cases while Gujrat has a high number of deaths compared to the rest. Over the three years, maximum rotavirus immunizations have been conducted in Madhya Pradesh and Uttar Pradesh.

Since population and spatial information is not taken into account, these raw counts do not paint the accurate picture of the disease data. For this, we will now compute smoothed rates and probabilities.

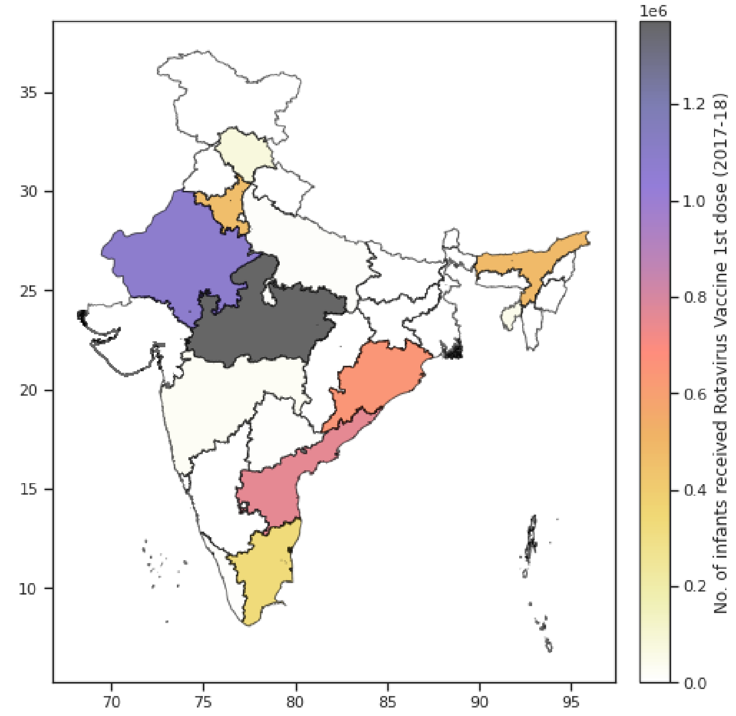

Figure 7: Children (Ages 0-5) who received Rotavirus vaccine first dose across states in the year 2017-18.

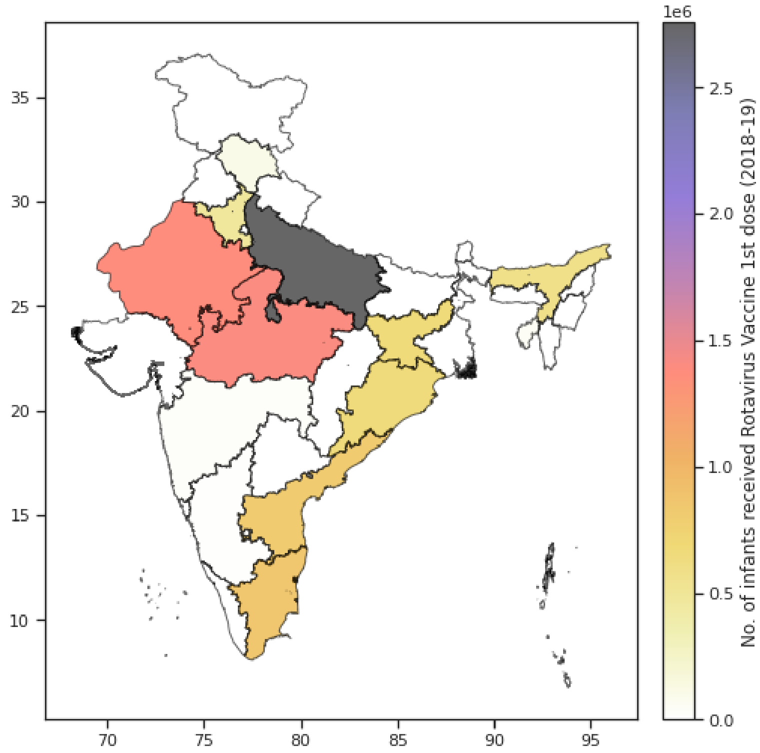

Figure 8: Children (Ages 0-5) who received Rotavirus vaccine first dose across states in the year 2018-19.

4 Basic Spatial Health Analysis

Now we do some basic spatial analysis for this study. First we compute the smoothed prevalence rates using the above data. We followed this up with strategy of probability mapping.

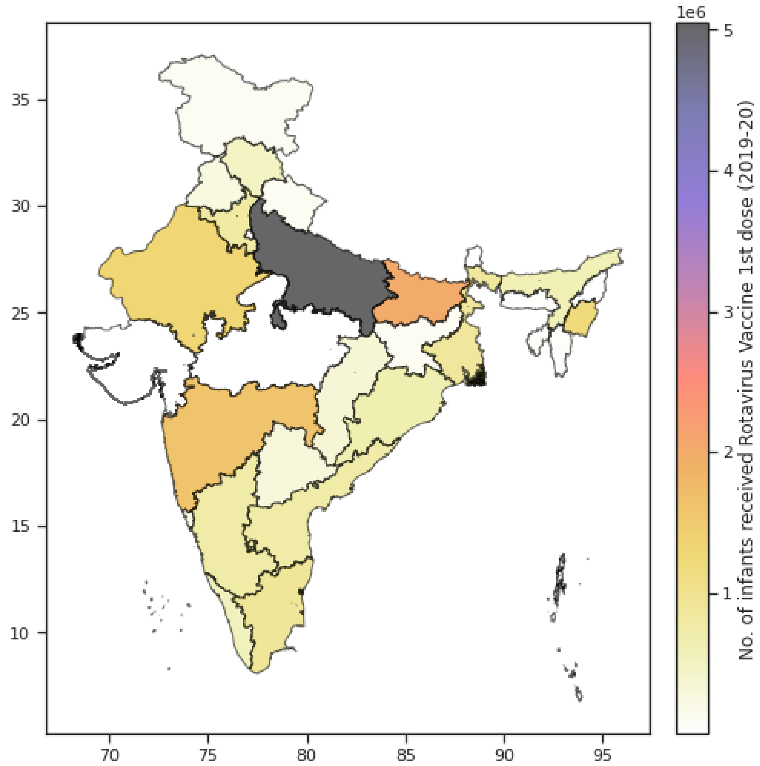

Figure 9: Children (Ages 0-5) who received Rotavirus vaccine first dose across states in the year 2019-20.

4.1 Smoothed Rates

We will compute a estimator which is a smoothened estimator for risk based on raw disease count and weights which are assigned as follows

$$w_{ij} = \begin{cases} 1 & \text{, if } i-\text{th state and } j-\text{th state border each other} \\ 0 & \text{, else} \end{cases}$$

Again, we are assuming that disease counts are Poisson r.v.s, such that, disease count for the i-th state $Y_i \sim Poisson(n_i\xi)$, where ξ is the probability of any person contacting the disease.

For this case, we have n_i =population of the i-th state, ξ would be the prevalence rate of diarrhoea in the entire India.

Then the rate we will estimate is $r_i = \dfrac{Y_i}{n_i}$.

Then locally smoothed estimate of ξ for the i-th region, $\widetilde{r_i}$ is given by

$$\widetilde{r_i} = \frac{\sum_{j=1}^{N} w_{ij} Y_j}{\sum_{j=1}^{N} w_{ij} n_j}$$

4.1.1 Implementation

We obtained the smoothed estimator based on the above strategy for the year 2011 (year of census) for an accurate analysis. The map is as follows:

Note. The states which have 0 are those whose data is not available because either the state was not formed in 2011 (Telangana)

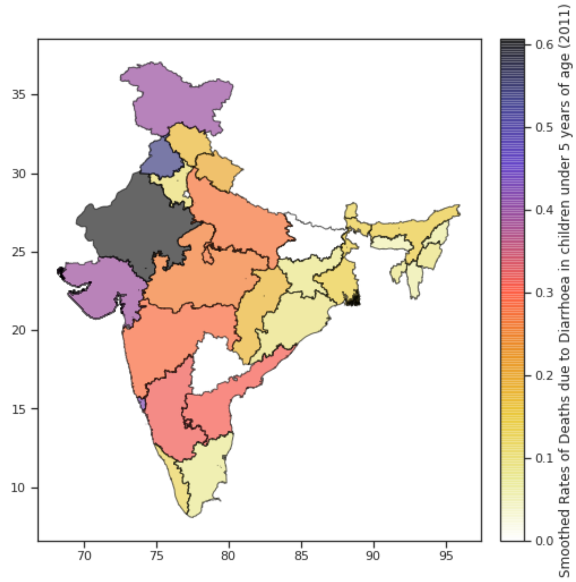

Figure 10: Smoothed rates for Diarrhoea related deaths in the year 2011

State/UTs	Smoothed Rates of Deaths due to Diarrhoea in children under 5 years of age (2011)
Andaman & Nicobar Island	0
Andhra Pradesh	0.312155691
Arunachal Pradesh	0.065792888
Assam	0.151159709
Bihar	0
Chandigarh	0
Chhattisgarh	0.180814365
Daman & Diu	0
Goa	0.423307512
Gujarat	0.391918007
Haryana	0.104498994
Himachal Pradesh	0.176977726
Jammu & Kashmir	0.397239065
Jharkhand	0.087584388
Karnataka	0.316457169
Kerala	0.127535732
Lakshadweep	0
Madhya Pradesh	0.258949943
Maharashtra	0.283254248
Manipur	0.096043188
Meghalaya	0.054752301
Mizoram	0.06276501
Nagaland	0.067772231
NCT of Delhi	0
Puducherry	0
Punjab	0.541156106
Rajasthan	0.606361747
Sikkim	0.131449931
Tamil Nadu	0.076844068
Telangana	0
Tripura	0.087193835
Uttar Pradesh	0.271888355
Uttarakhand	0.197485665
West Bengal	0.146257103
Odisha	0.094314389

Figure 11: Smoothed rates for Diarrhoea related deaths in the year 2011

4.2 Probability Mapping

We ignore the UTs because they will skew the average population of a state, as also affect the inference of the study.

Here we are assuming that the disease counts of each states are Poisson r.v.s

$$Y_1, Y_2, \ldots, Y_n \sim Poisson(\lambda = np)$$

where n =average population of children aged 0 to 4 years in a state, p would be the prevalence rate of diarrhoea in the entire india among children aged 0 to 4 years

We first compute the overall mean estimate as follows:

$$\hat{\xi} = \frac{\sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} n_i}$$

Then, under the constant risk hypothesis (risk is same in all states), the expected counts for each state will be

$$\widehat{E_i} = n_i \hat{\xi}$$

Now we will compute the following probabilities (based on the $Poisson(\lambda = np)$ distribution). If the observed disease counts for Y_i are y_i

$$p_i = \begin{cases} \mathbb{P}\left[Y_i \geq y_i | E(Y_i) = \widehat{E_i}\right] & , \text{ if } y_i > \widehat{E_i} \\ \mathbb{P}\left[Y_i \leq y_i | E(Y_i) = \widehat{E_i}\right] & , \text{ if } y_i < \widehat{E_i} \end{cases}$$

4.2.1 Implementation

We tried implementing the above to our data, but some states had very large values, which were insignificant. Only few states had un-skewed data which was fit for mapping. Since the map was turning out to have uneven because of the extremely skewed data, we have decided to not include it in this report.

5 Spatial Autocorrelation

A global index of spatial auto-correlation provides a summary over the entire study area of the level of spatial similarity observed among neighbouring observations.

We will be computing the Moran's I for India based on the diarrhoea case counts.

We obtained the following Moran's I over three years for India and also compared it to the Moran's I which implies perfect spatial auto-correlation.

Moran's I		
Year	Computed Spatial Autocorrelation	Perfect measure for no Autocorrelation
2017-18	-0.0448	-0.03704
2018-19	-0.0121	-0.03704
2019-20	-0.0927	-0.03704

Figure 12: Moran's I Values

Thus there the death count pattern is very regular and there is very little to no spatial clustering

6 Spatial Regression

We now perform spatial regression to see whether vaccine coverage affects diarrhoea death counts. This spatial regression was performed on R using the `gstat` package. It is done for three years - 2017-18, 2018-19, and 2019-20

6.1 Variogram Fitting

First we find the best fit for the variogram using the data that we had.

6.1.1 Year 2017-2018

The best fit variogram for this year was found to be gaussian with the following parameters:

	model	psill	range
1	Nug	0.00	0.000000
2	Gau	27715.39	1.918232

Figure 13: Year 2017-18 Variogram Fitting Parameters

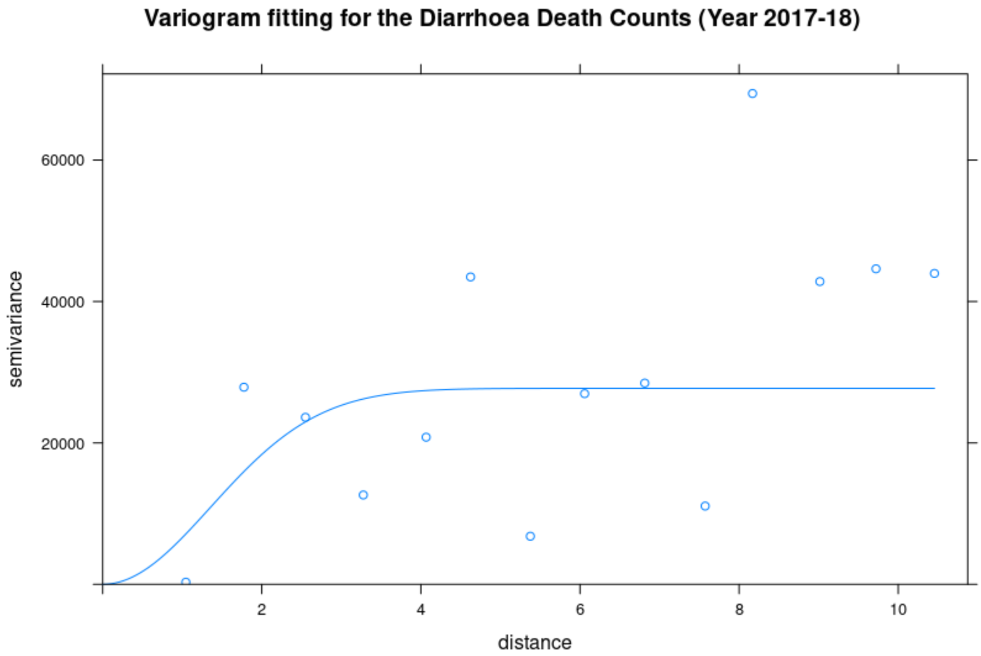

Figure 14: Year 2017-18 Variogram Fitting

6.1.2 Year 2018-2019

The best fit variogram for this year was found to be exponential with the following parameters:

	model	psill	range
1	Nug	3078.122	0.0000
2	Exp	76614.768	153.0025

Figure 15: Year 2018-19 Variogram Fitting Parameters

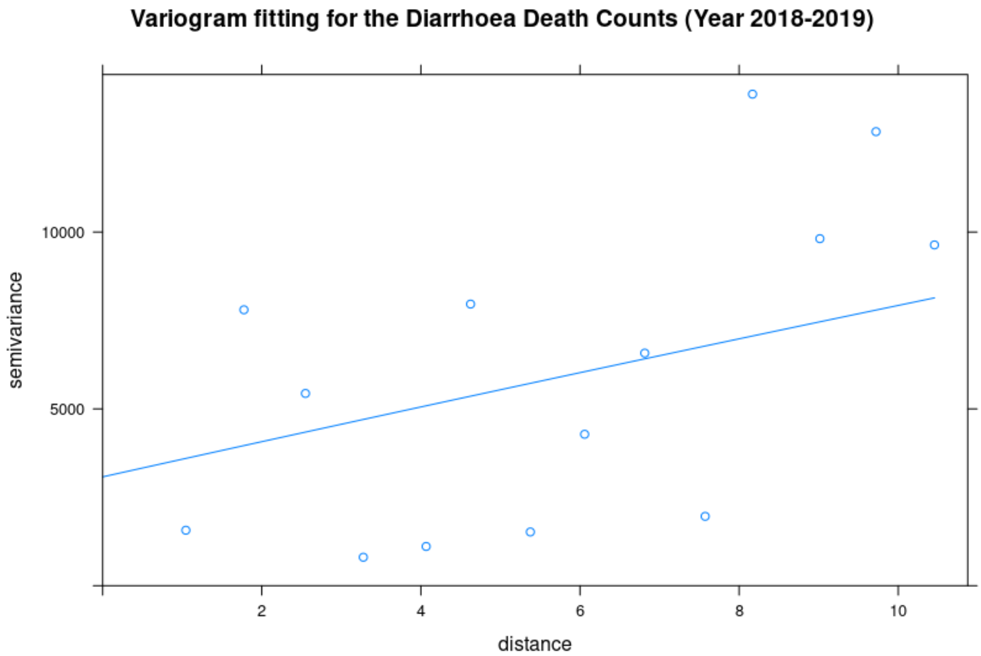

Figure 16: Year 2018-19 Variogram Fitting

6.1.3 Year 2019-2020

The best fit variogram for this year was found to be exponential with the following parameters:

	model	psill	range
1	Nug	0.000	0.000000
2	Exp	2669.046	1.702381

Figure 17: Year 2019-2020 Variogram Fitting Parameters

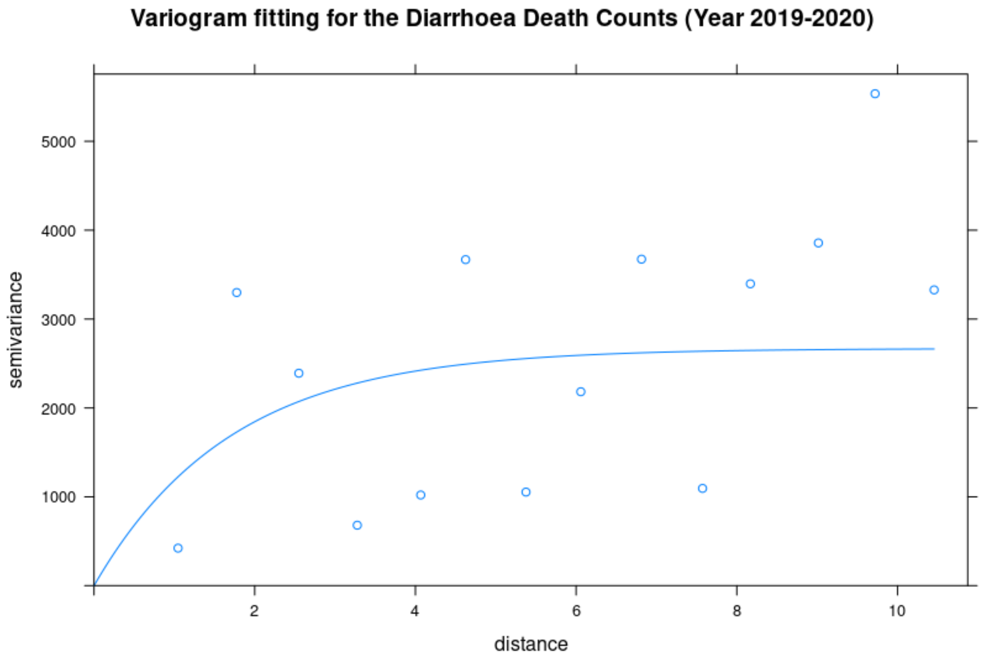

Figure 18: Year 2019-2020 Variogram Fitting

6.2 Regression

After finding the best-fit variogram, we transformed the variables (diarrhoea death counts is the response variable and rotavirus vaccine dose 1 counts are the explanatory variables)

6.2.1 Year 2017-2018

The regression summary and plot are as follows:

```
Residuals:
     Min       1Q   Median       3Q      Max 
-0.57909 -0.23836 -0.19657  0.07763  2.29502 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)  
(Intercept)     2.428e-01  9.004e-02   2.697   0.0108 *
explanatory_new 5.519e-05  4.068e-05   1.357   0.1837  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5012 on 34 degrees of freedom
Multiple R-squared:  0.05137,	Adjusted R-squared:  0.02347 
F-statistic: 1.841 on 1 and 34 DF,  p-value: 0.1837
```

Figure 19: Year 2017-18 Regression Summary

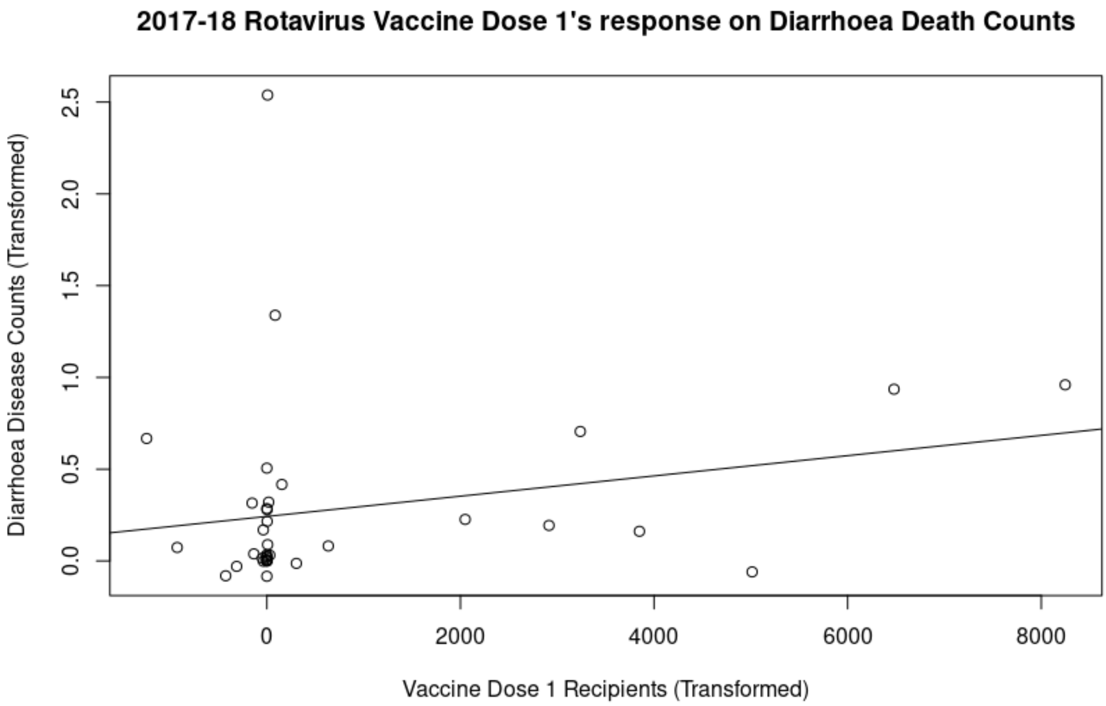

Figure 20: Year 2017-18 Regression Plot

6.2.2 Year 2018-2019

The regression summary and plot are as follows:

```
Residuals:
    Min      1Q  Median      3Q     Max
-2.3940 -0.3952 -0.1929  0.1565  5.5250

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     2.567e-02  2.039e-01   0.126    0.901
explanatory_new 3.160e-05  2.696e-05   1.172    0.249

Residual standard error: 1.223 on 34 degrees of freedom
Multiple R-squared:  0.03882,   Adjusted R-squared:  0.01055
F-statistic: 1.373 on 1 and 34 DF,  p-value: 0.2494
```

Figure 21: Year 2018-19 Regression Summary

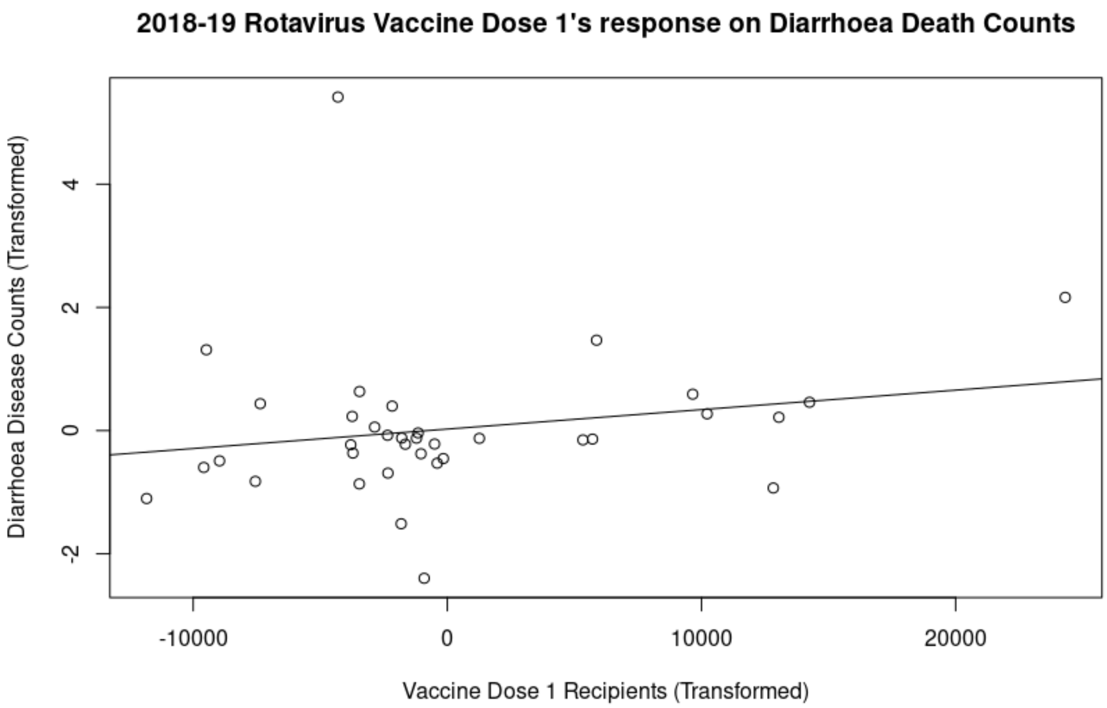

Figure 22: Year 2018-19 Regression Plot

6.2.3 Year 2019-2020

The regression summary and plot are as follows:

```
Residuals:
    Min      1Q  Median      3Q     Max
-1.1166 -0.4991 -0.4139 -0.0651  7.0695

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     4.863e-01  2.583e-01   1.882 0.068354 .
explanatory_new 5.194e-05  1.360e-05   3.819 0.000542 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.37 on 34 degrees of freedom
Multiple R-squared:  0.3002,    Adjusted R-squared:  0.2796
F-statistic: 14.59 on 1 and 34 DF,  p-value: 0.0005418
```

Figure 23: Year 2019-20 Regression Summary

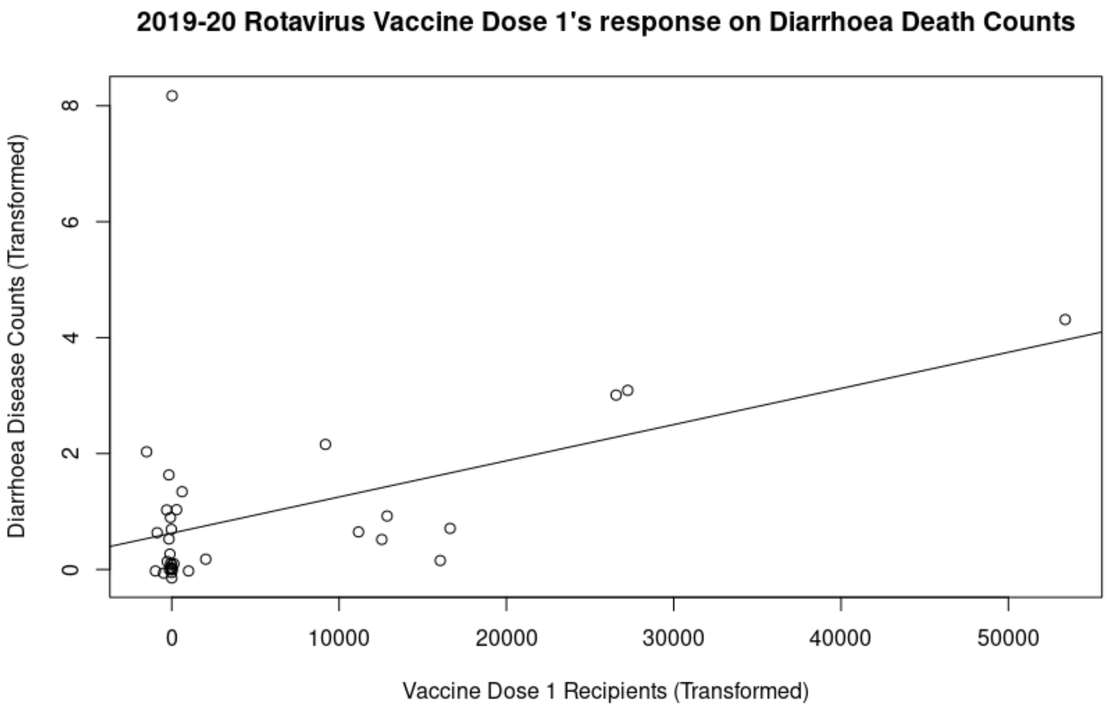

Figure 24: Year 2019-20 Regression Plot

7 Acknowledgements

Information about Rotavirus and its immunization has been obtained through the Government of India's Rotavirus Immunization Guidelines. More information about the same was obtained through WikiPedia.

Thanks to the course instructor and the TA for providing essential resources for doing this project!