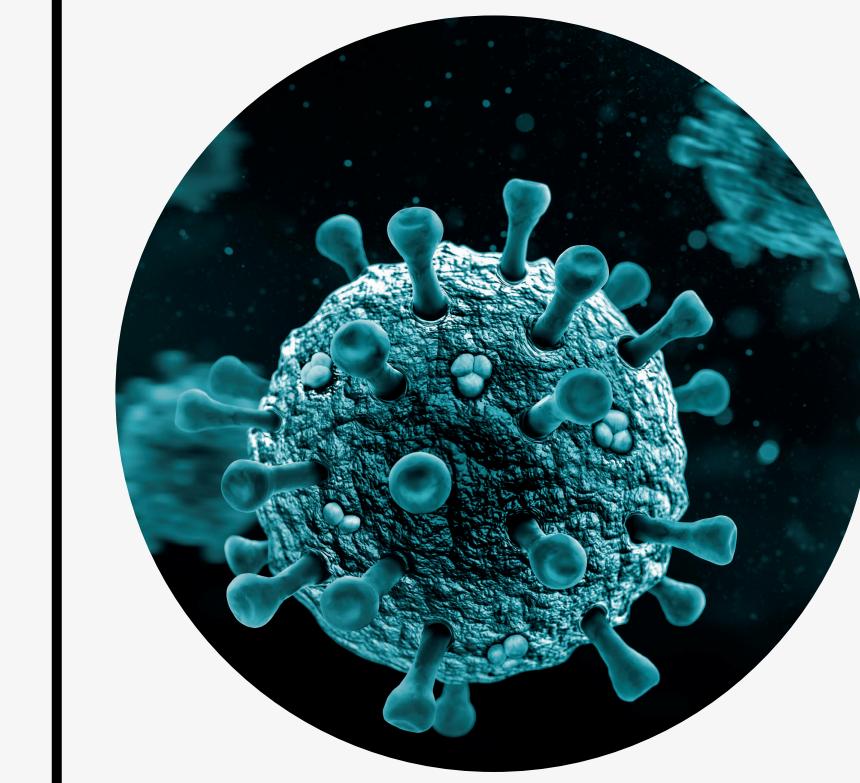


Estimating the basic reproduction number (R_0) of COVID-19 using a Monte Carlo method on the SIR model

The objective of this project is to estimate the basic reproduction number R_0 (I) of the COVID-19 pandemic in Germany using a Monte-Carlo simulation method (IV) on data from the first case peak in late March 2020 (II). The simulation is based on fitting randomly sampled parameters of the well-understood SIR model for infectious diseases (III) to case, recovery and death data. The obtained estimates (V) are presented and juxtaposed with similar estimates for the countries South Africa, India, Peru and Russia. Since in contrast to the data for Germany, the data for the other four countries is affected by a lockdown, this gives insight into the effectiveness of social distancing measures.

I. Background - R_0

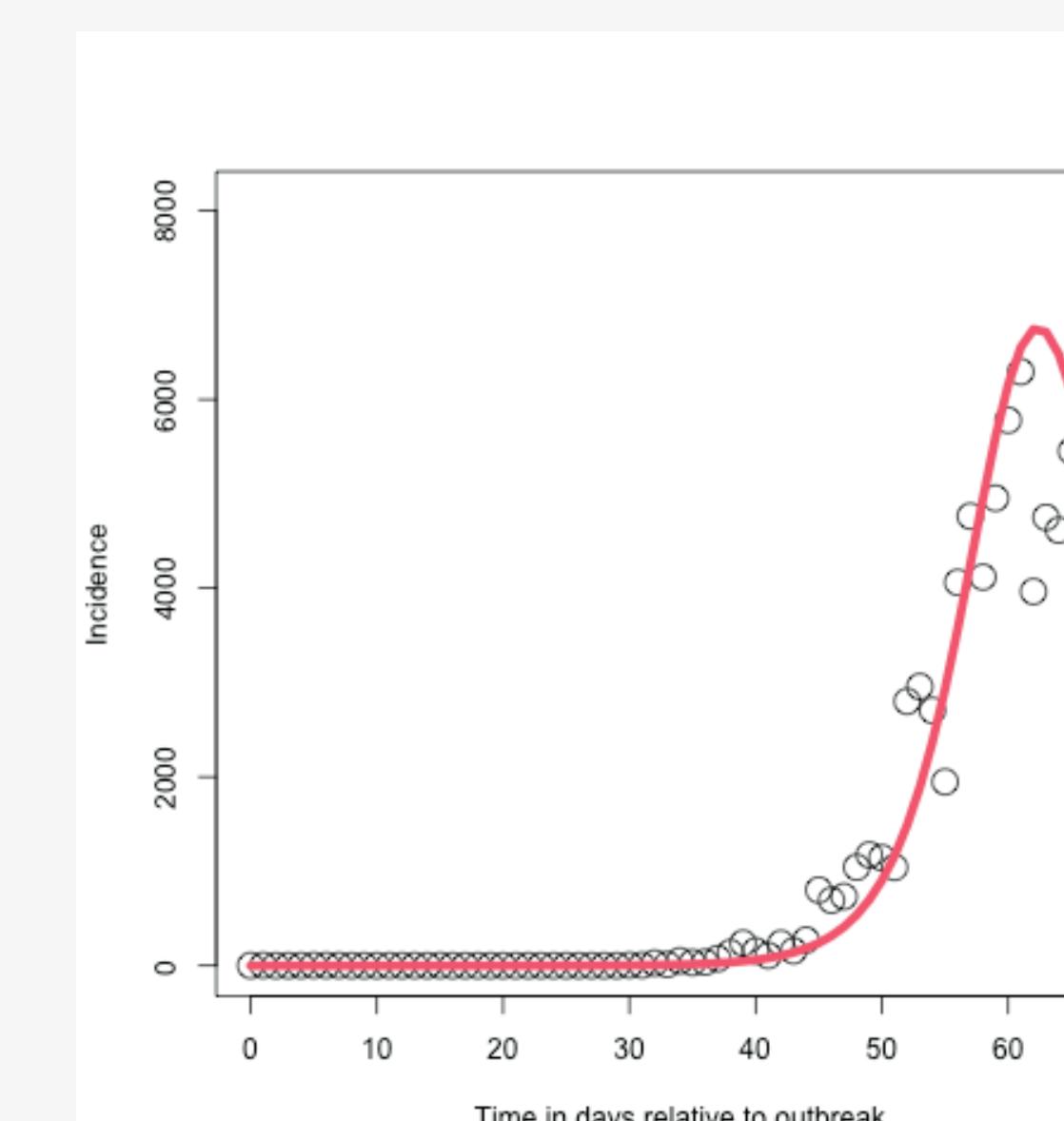


The basic reproduction number R_0 is "the number of secondary cases which one case would produce in a completely susceptible population." (Dietz, 1993) As such, R_0 is an epidemiological measure for a disease derived from how many new infections an infected individual is expected to cause on average. Hence, the number contains information on the spread of a disease (with $R_0=1$ being a cutoff value) and enables evidence-based policy-making and public communication.

Dietz, K. (1993). The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research*, 2 (1), 23–41. <https://doi.org/10.1177/096228029300200103>

II. Data

We retrieve global COVID-19 data compiled by the Johns Hopkins University from the official GitHub page "CSSEGISandData", maintained by the Johns Hopkins Whiting School of Engineering (see QR code). In particular, the dataset comprises COVID-19 infection, recovery and death data for most days of the pandemic and a large set of countries. We focus on the data for Germany from January to March 2020, and then extend our analysis to data from South Africa, Peru, Russia and India.



III. Background - SIR

SIR is a compartmental model consisting of three compartments:

Susceptible are individuals who can contract the disease

Infected are those infected with and thus capable of transmitting the disease

Removed are individuals who are not susceptible any more, either due to recovery or death.



A system of non-linear differential equations relates the three compartments, defining transition rates depending on the contact rate β and the removal parameter γ .

Note: We define $R_0 = \beta/\gamma$.

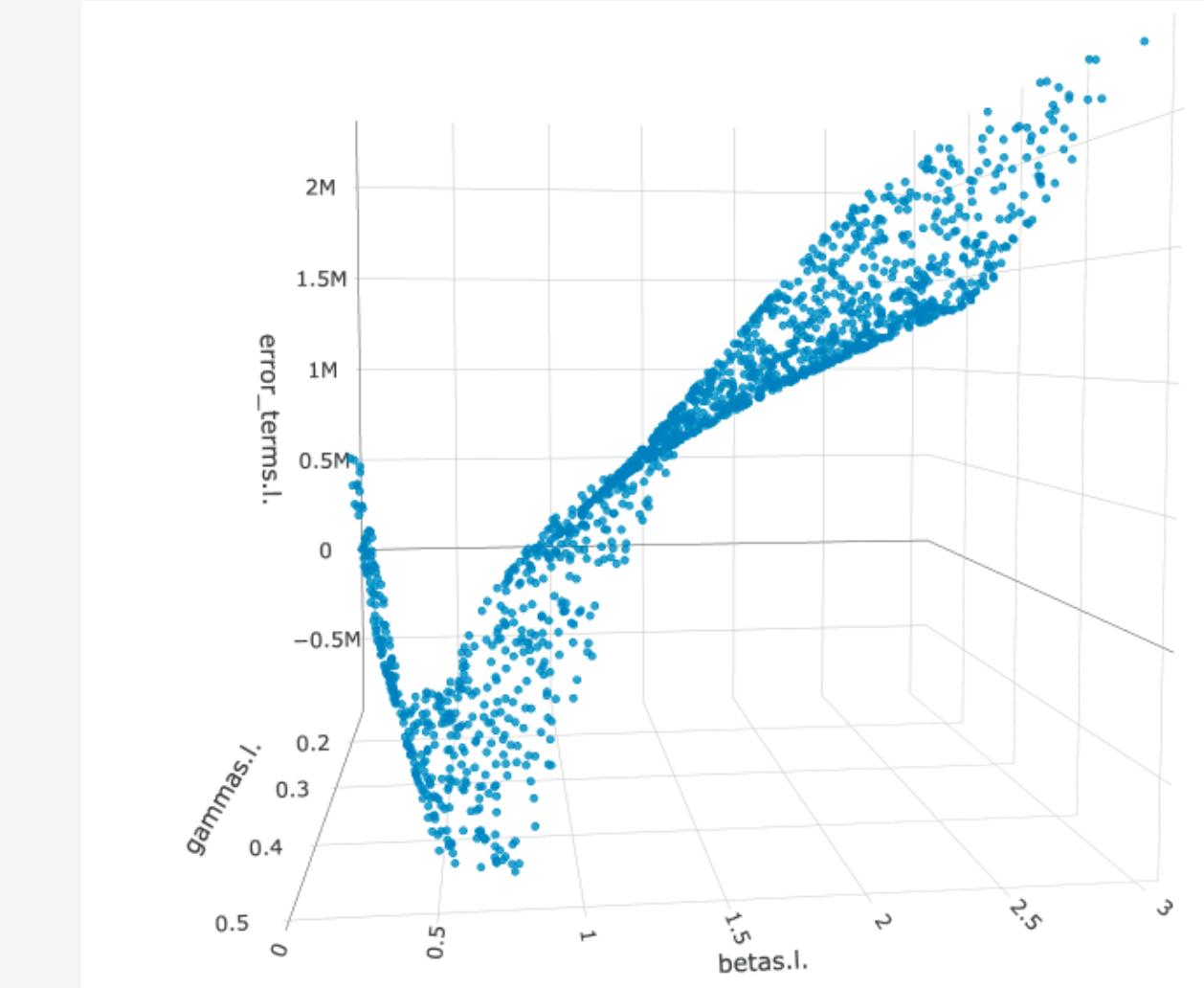
$$dS = -\beta S \frac{I}{N}$$

$$dI = +\beta I \frac{S}{N} - \gamma I$$

$$dR = +\gamma I$$

IV. Monte Carlo simulation

We use a bottom-up approach agnostic to the model parameters, and proceed to subset each country data around the peak date. Then we run a Monte Carlo simulation with 1500 iterations. For each iteration, a pair (β, γ) is sampled from specified ranges, and provided to a system of difference equations derived from (III) to simulate the spread of the pandemic. The resulting simulations for infected and removed individuals over each time frame are compared to the actual data, and the goodness-of-fit is quantified using a Poisson Negative Log-Likelihood function. The best percent of estimates is extracted for each time step, and the resulting pairs of parameters are stored in order to estimate R_0 .



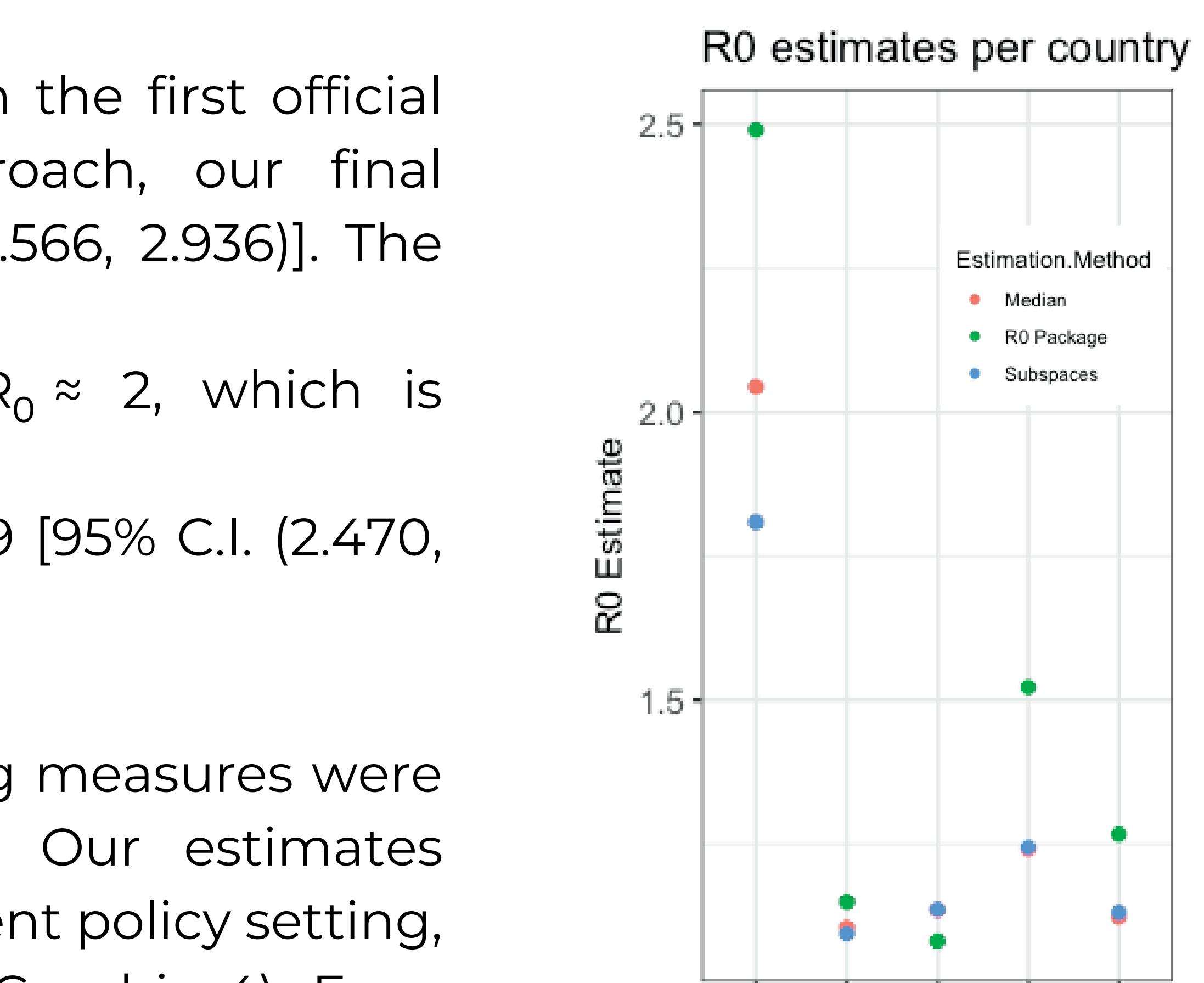
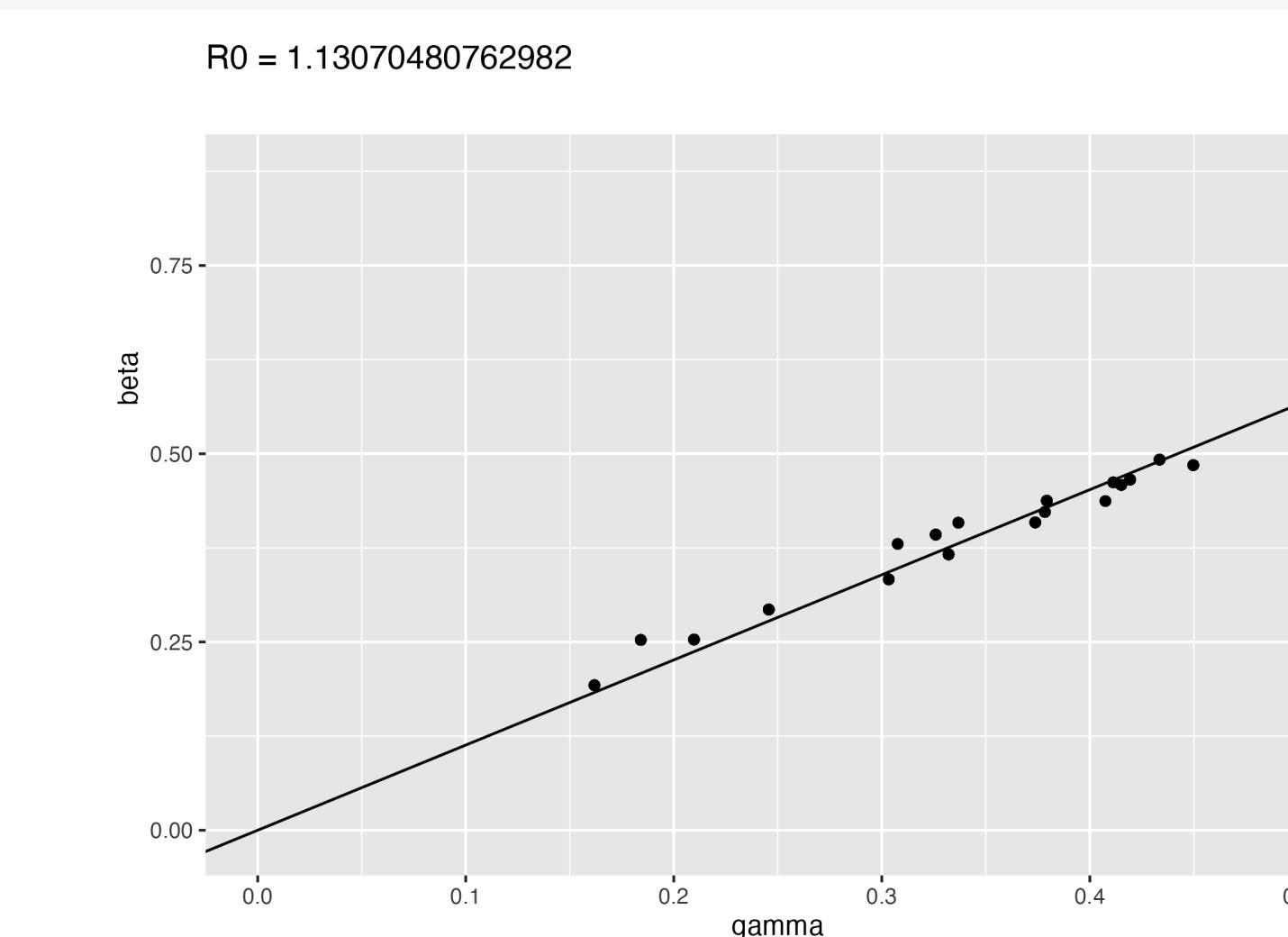
Graphic 2: 3D plot of error function with respect to parameters β and γ

V. Results

ESTIMATION METHODS

We estimate R_0 in two different ways, using 1. the median of the ratios of best fit pairs (β_i, γ_i) , and 2. the best-fit one-dimensional subspace calculated by a regression through the origin.

We validate these estimates with the R package "R0".



VI. Conclusion

Estimating metrics of infectious diseases will always remain a challenging yet imperative task in combating outbreaks. Broad engagement in the discussion around appropriate metrics and estimates is needed to mitigate the threat of such diseases to human populations around the world. In this context, the COVID-19 pandemic attested an increased importance of epidemiological measures for disease spread, but also for public communication.

We were able to demonstrate the ability of our approach to obtain a reasonably good estimate on the basic reproduction number. Furthermore, we were able to quantify the effectiveness of social distancing measures. However, the results were not perfectly stable, and indicated that the method might work better in lower ranges of R_0 .

Overall, the estimation method provided in this project is rudimentary, but shows how to engage with some of the inherent questions posed by epidemiological metrics.