Analysis of registered covid-19 cases and numbers of death

Fanni Kiss

2020 11 29

Introduction

Current report aims to analyse the pattern of association between registered COVID-19 cases per capita and registered numbers of death per capita due to COVID-19 on 22 September 2020. GitHub https://github.com/fanni-k/DA2.

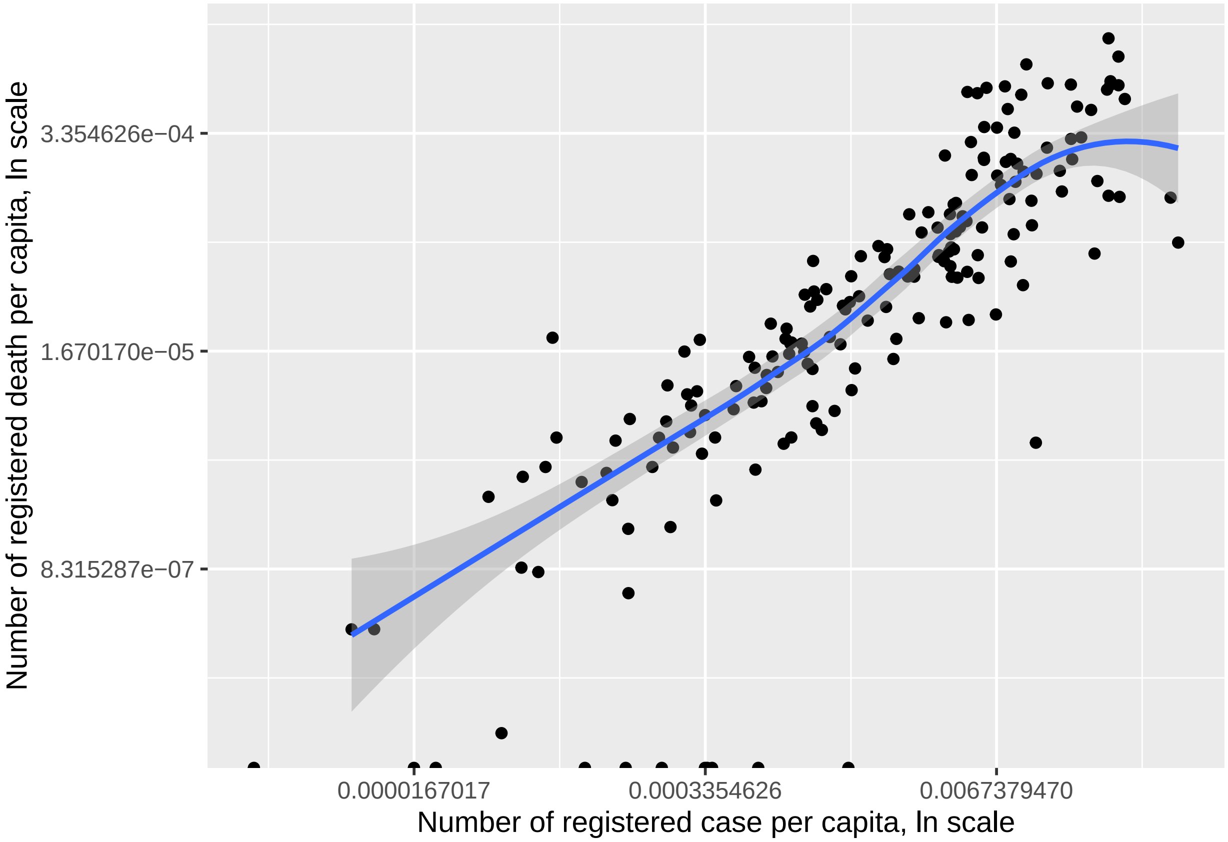
Exploring the data

The distribution of each variables are all skewed with a long right tale. In the summary table (Appendix, Teble 1), we can observe that there are some extreme values, which are presumably not measurement errors. Thus, extreme values are involved into the further analysis. The analysis focuses on two variables: the number of cases per capita (dependent variable) and the number of death per capita (explanatory variable). Both of them skewed with a long right tale and contains only positive numbers.

Choosing a model

As the dependent variable and the explanatory variable both skewed with long right tale and does not contain negative numbers, the log transformation could be applied on the values. We compared four different models: * level-level model * log-level model * level-log model * log-log model

On the plots, we can observe that the log-log model makes the association close to linear between the both variable. This model shows us, that how many percent higher number of death per capita associated with 1% higher number of cases per capita by country, which is a meaningful interpretation. Furhtermore, both of the variables are skewed with a long right tale, so the log transformation balance the distribution for the further analysis. So, the log-log regression is going to be analysed in the further analysis.



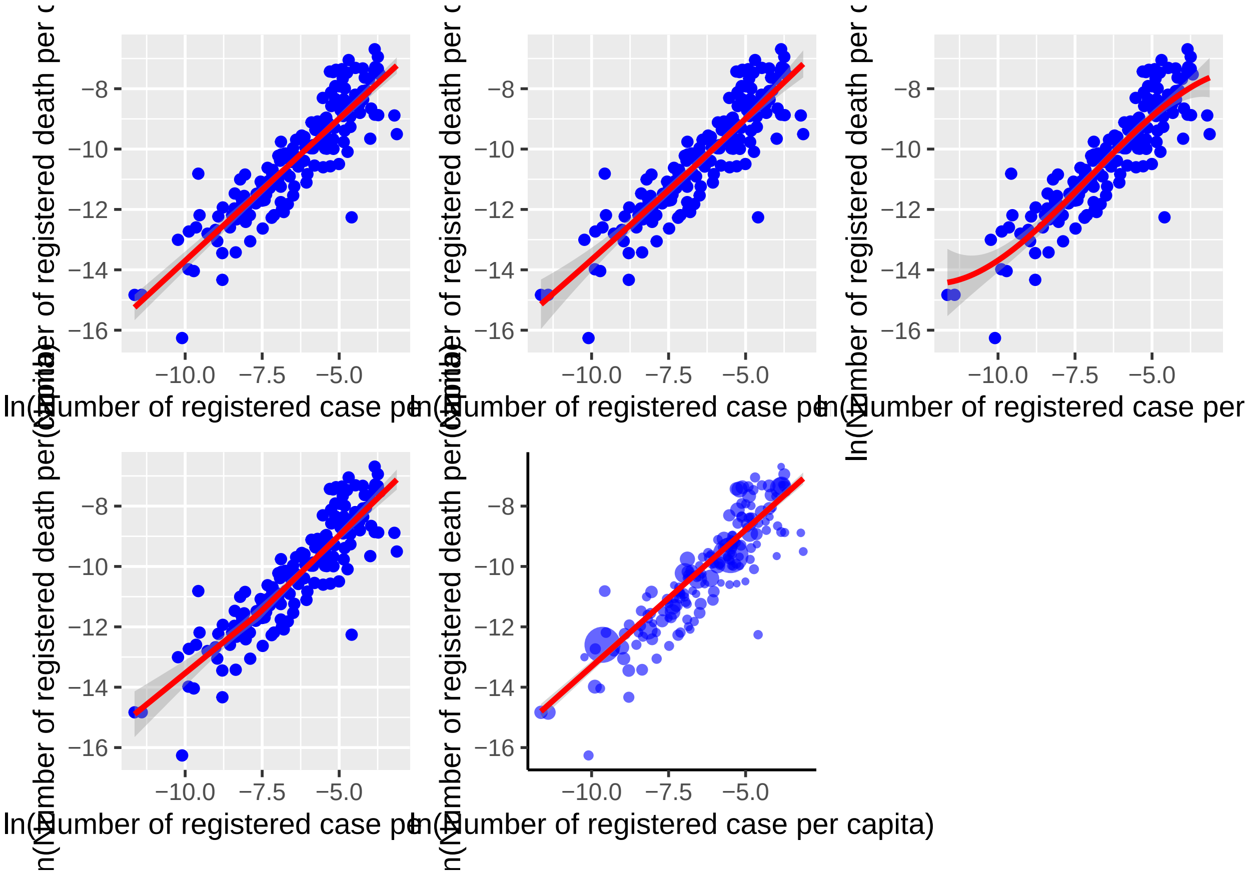
Regression

In the analysis, we make five regression:

- Simple linear regression (log-log):
  - reg1: ln_death_per_capita = alpha + beta * ln_case_per_capita
- Quadratic linear regression:
  - reg2: ln_death_per_capita = alpha + beta_1 * ln_case_per_capita + beta_2 * ln_case_per_capitaˆ2
- Cubic liear regression:
  - reg3: ln_death_per_capita = alpha + beta_1 * ln_case_per_capita + beta_2 * ln_case_per_capitaˆ2 + beta_3 * ln_case_per_capitaˆ3
- Piecewise linear spline regression
  - reg4: ln_death_per_capita = alpha + beta_1 * ln_case_per_capita * 1(case_per_capita < 0.00091) + beta_2 * ln_case_per_capita * 1(case_per_capita >= 0.00091)
- Weighted linear regression, using population as weights
  - reg5: ln_death_per_capita = alpha + beta * ln_case_per_capita, weights: population

For the further analysis, we pick the most reliable model. Based on the adjusted R-squared, it is the weighted linear regression model, which uses population as weights. In this case, the R-squared is 0.89, which could be accepted as a relatively high R-squared value.

The result of the weighted linear regression model is ln_death_per_capita = -4.25 + 0.91 * ln_case_per_capita, which means that the number of death is 0.91% higher on average for observations with one percent higher number of cases due to COVID.



Hypothesis testing

In the hypothesis testing, we test if there is a significant linear relationship between the number of death and number of cases on a 0.05 level of significance.

- H0: beta = 0, or the slope of the regression line is zero
- H1: beat != 0, or the slope of the regression line is not equal to zero. If the relationship between the death and cases is significant, the slope will not equal to zero.

The p-value of the model is 2.2e-16, which is less than the significance level (0.05), we cannot accept the null hypothesis.

```
## Linear hypothesis test
##
## Hypothesis:
## ln_case_per_capita = 0
##
## Model 1: restricted model
## Model 2: ln_death_per_capita ~ ln_case_per_capita
##
##   Res.Df Df Chisq Pr(>Chisq)
## 1    169
## 2    168  1 119.5  < 2.2e-16 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

Residual analysis

Based on the residual analysis, we can observe, which countries lost and save relatively the most people. The residual analysis shows, which observations are the farest from the predicted value. Countries, where there are less death due to COVID-19 compared to the number of cases, are shown Table 1. While those countries, where there are more death due to COVID-19 compared to the number of cases, are shown Table 2.

Table 1: The TOP5 countries, who saved relatively the most people

country	ln_death_per_capita	reg5_y_pred	reg5_res
Burundi	-16.260513	-13.403346	-2.857167
Maldives	-9.656068	-7.867324	-1.788744
Qatar	-9.504659	-7.086783	-2.417876
Singapore	-12.260766	-8.415050	-3.845716
Sri Lanka	-14.332609	-12.218580	-2.114028

Table 2: The TOP5 countries, who saved relatively the most people

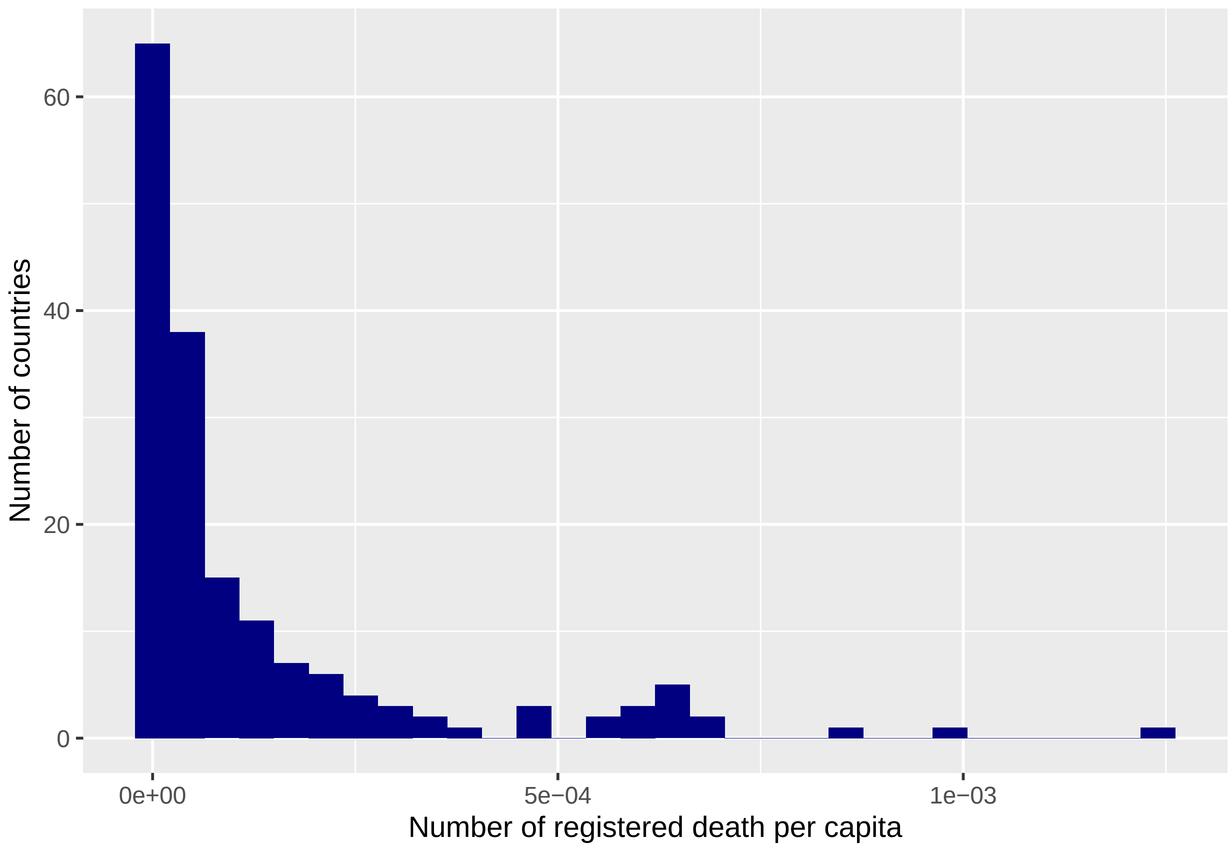
country	ln_death_per_capita	reg5_y_pred	reg5_res
Belgium	-7.050640	-8.503632	1.452992
Italy	-7.430830	-9.054322	1.623493
Mexico	-7.447707	-8.961540	1.513833
United Kingdom	-7.373471	-8.875977	1.502506
Yemen	-10.815055	-12.926908	2.111853

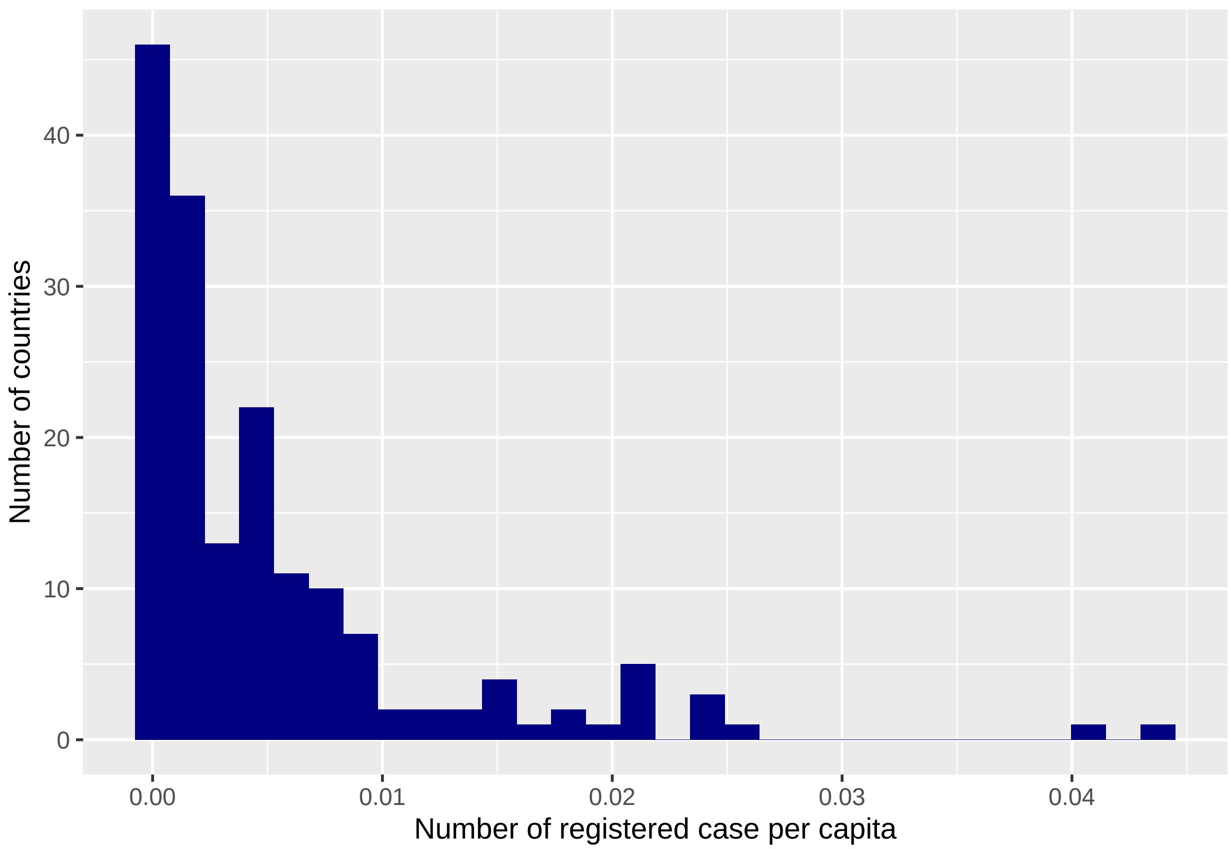
Appendix

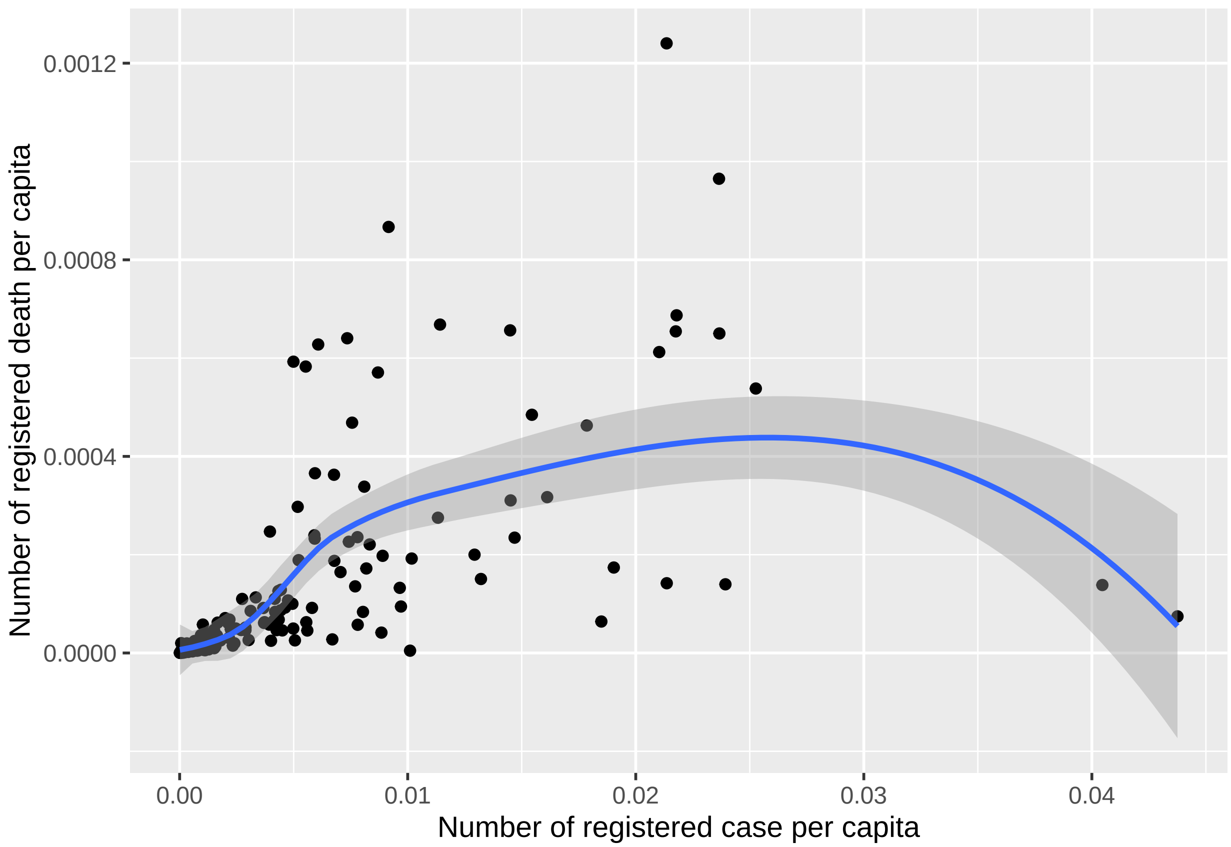
Table 3: Descriptive statistics of the variables

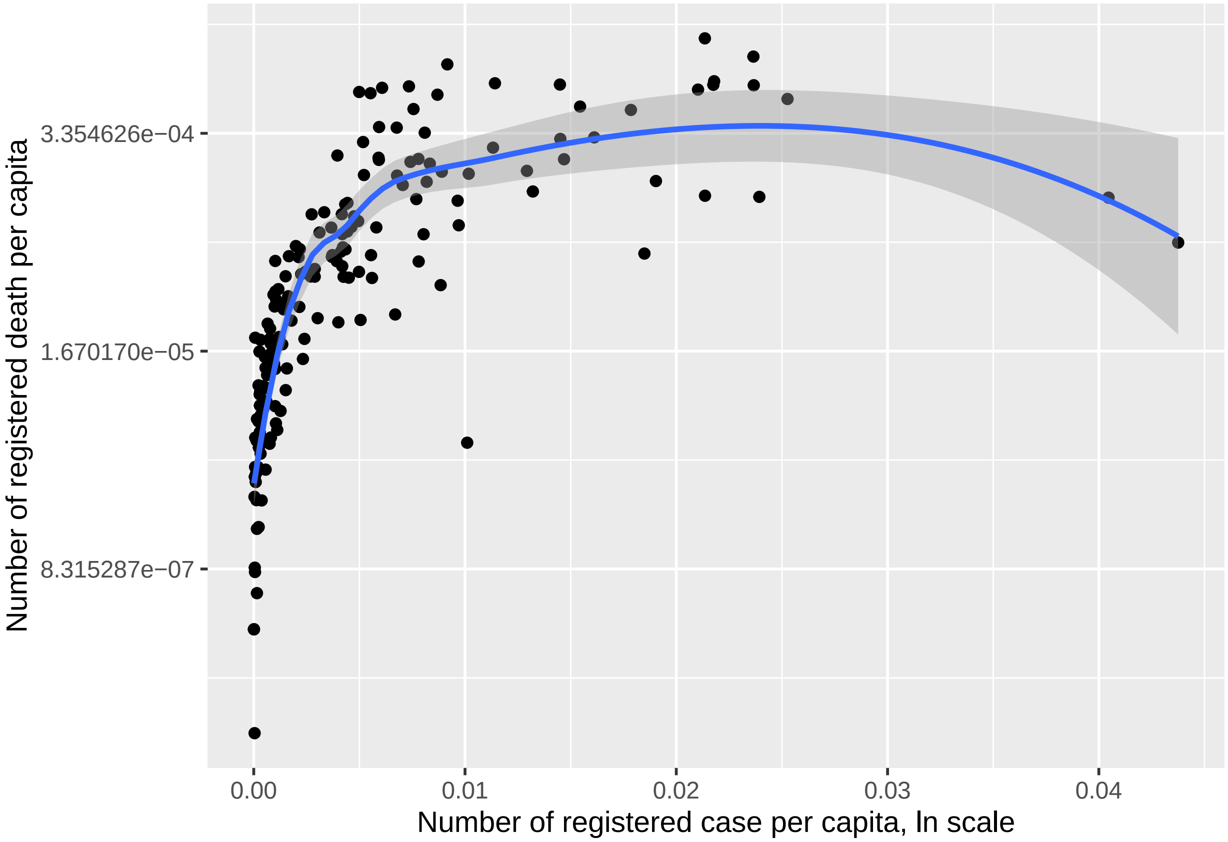
country	confirmed	death	recovered	active	country_code	population	death_per_capita	case_per_capita	ln_death_per_capita	ln_case_per_capita	ln_case_per_capita_squared	ln_case_per_capita_cubed	reg5_y_pred	reg5_res
Length:170	Min. : 32	Min. : 1	Min. : 0	Min. : 0	Length:170	Min. :3.386e+04	Min. :8.470e-08	Min. :8.780e-06	Min. :-16.261	Min. :-11.644	Min. : 9.792	Min. :-1578.56	Min. :-14.803	Min. :-3.8457
Class :character	1st Qu.: 3864	1st Qu.: 64	1st Qu.: 1952	1st Qu.: 535	Class :character	1st Qu.:3.004e+06	1st Qu.:1.035e-05	1st Qu.:6.601e-04	1st Qu.:-11.479	1st Qu.: -7.321	1st Qu.: 24.938	1st Qu.: -392.42	1st Qu.:-10.886	1st Qu.:-0.6789
Mode :character	Median : 15532	Median : 284	Median : 9672	Median : 2619	Mode :character	Median :1.028e+07	Median :4.722e-05	Median :2.579e-03	Median : -9.982	Median : -5.961	Median : 35.540	Median : -211.89	Median : -9.653	Median :-0.1888

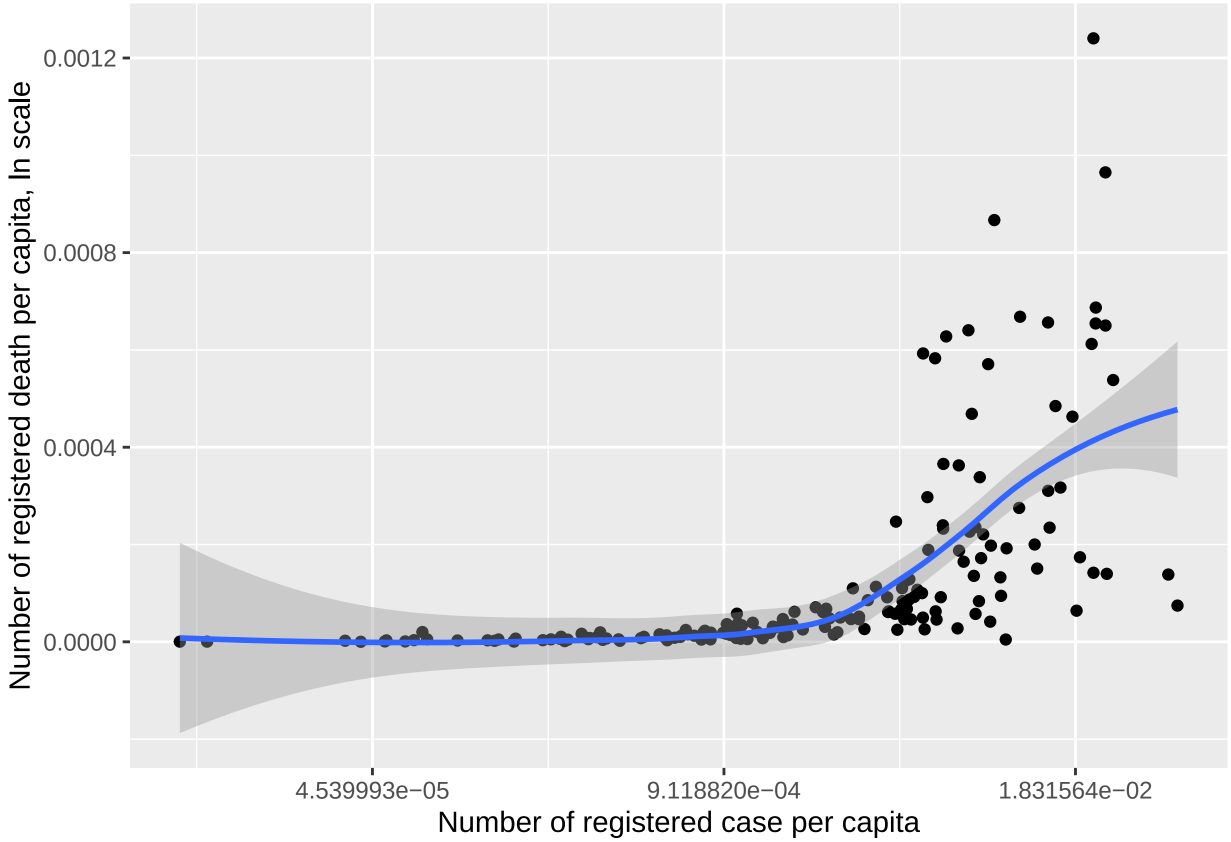
country	confirmed	death	recovered	active	country_code	population	death_per_capita	case_per_capita	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
NA	Mean : 185830	Mean : 5708	Mean : 127664	Mean : 29004	NA	Mean :4.452e+07	Mean :1.312e-04	Mean :5.251e-03	Mean :- 10.191	Mean : - 6.272	Mean : 42.261	Mean : - 304.89	Mean : - 9.935	Mean :- 0.2561
NA	3rd Qu.: 85010	3rd Qu.: 1696	3rd Qu.: 58969	3rd Qu.: 13492	NA	3rd Qu.:3.237e+07	3rd Qu.:1.304e-04	3rd Qu.:6.780e-03	3rd Qu.: - 8.878	3rd Qu.: - 4.994	3rd Qu.: 53.600	3rd Qu.: - 124.54	3rd Qu.: - 8.777	3rd Qu.: 0.2576
NA	Max. :6902414	Max. :200974	Max. :4587613	Max. :3968377	NA	Max. :1.398e+09	Max. :1.240e-03	Max. :4.375e-02	Max. : - 6.692	Max. : - 3.129	Max. :135.573	Max. : - 30.64	Max. : - 7.087	Max. : 2.1119
NA	NA	NA	NA	NA's :3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA


Number of countries
60
40
20
0
0e+00
5e−04
1e−03
Number of registered death per capita


Number of countries
40
30
20
10
0
0.00
0.01
0.02
0.03
0.04
Number of registered case per capita


0.0012
0.0008
0.0004
0.0000
0.00
0.01
0.02
0.03
0.04
Number of registered death per capita
Number of registered case per capita


3.354626e−04
1.670170e−05
8.315287e−07
Number of registered death per capita
0.00
0.01
0.02
0.03
0.04
Number of registered case per capita, ln scale



```
##
## Call:
## lm(formula = ln_death_per_capita ~ ln_case_per_capita, data = df)
##
## Coefficients:
##        (Intercept)  ln_case_per_capita
##            -4.2882              0.9411
```

```
##
## Call:
## lm(formula = ln_death_per_capita ~ ln_case_per_capita, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6481 -0.3962  0.0791  0.4922  2.4833
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -4.28825    0.24669  -17.38   <2e-16 ***
## ln_case_per_capita  0.94114    0.03795   24.80   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8465 on 168 degrees of freedom
## Multiple R-squared:  0.7855, Adjusted R-squared:  0.7842
## F-statistic: 615.1 on 1 and 168 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm_robust(formula = ln_death_per_capita ~ ln_case_per_capita,
##     data = df, se_type = "HC2")
##
## Standard error type:  HC2
##
## Coefficients:
##                    Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)         -4.2882    0.29961  -14.31 6.811e-31   -4.880   -3.697 168
## ln_case_per_capita   0.9411    0.04566   20.61 7.142e-48    0.851    1.031 168
##
## Multiple R-squared:  0.7855 ,    Adjusted R-squared:  0.7842
## F-statistic: 424.9 on 1 and 168 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm_robust(formula = ln_death_per_capita ~ ln_case_per_capita +
##     ln_case_per_capita_sq, data = df)
##
## Standard error type:  HC2
##
## Coefficients:
##                       Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper
## (Intercept)           -4.06315    0.96580 -4.2070 4.214e-05 -5.96990 -2.15640
## ln_case_per_capita     1.01221    0.28173  3.5928 4.301e-04  0.45599  1.56843
## ln_case_per_capita_sq  0.00522    0.02009  0.2598 7.953e-01 -0.03445  0.04489
##                        DF
## (Intercept)           167
## ln_case_per_capita    167
## ln_case_per_capita_sq 167
##
## Multiple R-squared:  0.7856 ,    Adjusted R-squared:  0.783
## F-statistic: 210.2 on 2 and 167 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm_robust(formula = ln_death_per_capita ~ ln_case_per_capita +
##     ln_case_per_capita_sq + ln_case_per_capita_cb, data = df)
##
## Standard error type:  HC2
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)  CI Lower  CI Upper
## (Intercept)            -8.2243    3.01617 -2.7268 0.007083 -14.17933 -2.269348
## ln_case_per_capita     -0.9500    1.32032 -0.7195 0.472820  -3.55680  1.656760
## ln_case_per_capita_sq  -0.2856    0.18368 -1.5550 0.121841  -0.64827  0.077018
## ln_case_per_capita_cb  -0.0136    0.00813 -1.6725 0.096304  -0.02965  0.002454
##                        DF
## (Intercept)           166
## ln_case_per_capita    166
## ln_case_per_capita_sq 166
## ln_case_per_capita_cb 166
##
```

```
## Multiple R-squared:   0.79 , Adjusted R-squared:  0.7862
## F-statistic: 145.5 on 3 and 166 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm_robust(formula = ln_death_per_capita ~ lspline(ln_case_per_capita,
##     cutoff_ln_1, cutoff_ln_2), data = df)
##
## Standard error type:  HC2
##
## Coefficients:
##                                                         Estimate Std. Error
## (Intercept)                                              -5.2564     0.9736
## lspline(ln_case_per_capita, cutoff_ln_1, cutoff_ln_2)1   0.8278     0.1213
## lspline(ln_case_per_capita, cutoff_ln_1, cutoff_ln_2)2   0.1618     0.1555
##                                                         t value  Pr(>|t|)
## (Intercept)                                              -5.399 2.272e-07
## lspline(ln_case_per_capita, cutoff_ln_1, cutoff_ln_2)1   6.822 1.573e-10
## lspline(ln_case_per_capita, cutoff_ln_1, cutoff_ln_2)2   1.040 2.996e-01
##                                                         CI Lower CI Upper  DF
## (Intercept)                                              -7.1786  -3.3342 167
## lspline(ln_case_per_capita, cutoff_ln_1, cutoff_ln_2)1   0.5883   1.0674 167
## lspline(ln_case_per_capita, cutoff_ln_1, cutoff_ln_2)2  -0.1452   0.4688 167
##
## Multiple R-squared:  0.787 , Adjusted R-squared:  0.7845
## F-statistic: 212.6 on 2 and 167 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm_robust(formula = ln_death_per_capita ~ ln_case_per_capita,
##     data = df, weights = population)
##
## Weighted, Standard error type:  HC2
##
## Coefficients:
##                    Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)         -4.2511     0.5224  -8.137 8.630e-14  -5.2825    -3.22 168
## ln_case_per_capita   0.9062     0.0829  10.932 2.377e-21   0.7426     1.07 168
##
## Multiple R-squared:  0.8939 ,    Adjusted R-squared:  0.8933
## F-statistic: 119.5 on 1 and 168 DF,  p-value: < 2.2e-16
```