Assignment 1 for Data Analysis 2 and Coding with R

Pauline Broussolle

2020-11-24

Introduction

The aim of the analysis is to analyse the pattern of association between registered covid-19 cases and registered number of death due to covid-19 on a **15/10/2020**. My dependent variable (y) is the **number of registered death due to covid** and my explanatory variable (x) is the **number of registered cases**. The aim of the analysis is to create a report on the pattern of association, choose and interpret a regression model and refer to robustness checks.

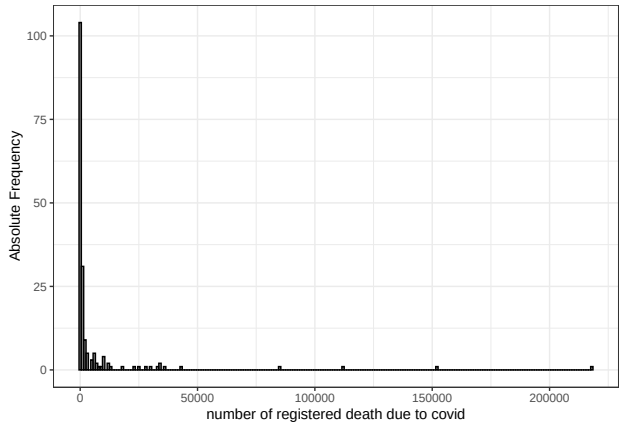
Introduction of the dataset and variables The aim of this assignment is to guide you through in creating a short report on the pattern of association. You will need to chose your final model, interpret the results and ### Executive summary

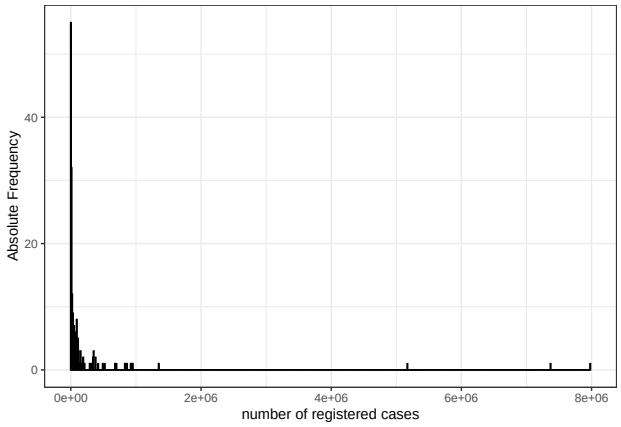
1. Summary statistics and Distribution for x and y

Table 1: Summary for the number of registered death and registered cases

mean	median	min	max	std	variable
213757	16483.0	3	7983919	899396	confirmed cases
6032	281.5	0	217883	22934	nb of death

2-3 sentence, explain the main features and distribution - use histograms and summary statistics table (mean, median, min, max, standard deviation)





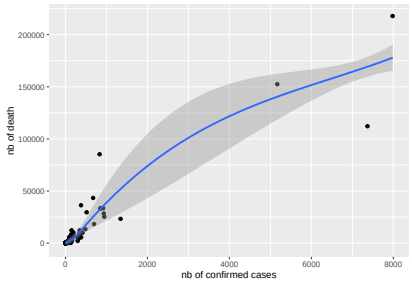
Select or drop observations, checking extreme values

We check countries which have confirmed cases aboves 2 milion and registered number of death above 50,000. These are India, Brazil and United States, which are not measurement errors. We keep these values.

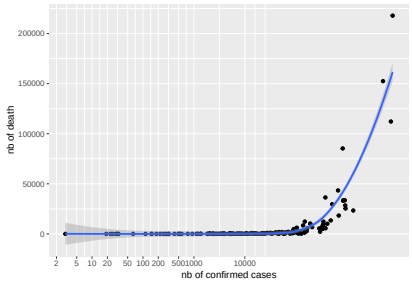
2. Investigate the transformation of your variables

Scaling

```
## `geom_smooth()` using formula 'y ~ x'
```




```
## `geom_smooth()` using formula 'y ~ x'
```

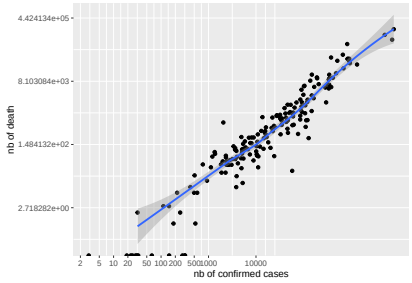



```
## Warning: Transformation introduced infinite values in continuous y-axis
```

```
## Warning: Transformation introduced infinite values in continuous y-axis
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 12 rows containing non-finite values (stat_smooth).
```



Estimating different models

Presentation of model choice

Hypothesis testing on beta (which interacts with x)

Analysis of the residuals

Conclusion