The spread of COVID-19 in a landlocked country the case of Luxembourg

A Preprint

Bruno Rodrigues *
Statistics department
Ministry of Higher Education and Research, Grand-Duchy of Luxembourg
Luxembourg, 18-20, Montée de la Pétrusse
bruno.rodrigues@mesr.etat.lu

March 5, 2021

Abstract

Enter the text of your abstract here.

***K*eywords** blah · blee · bloo · these are optional and can be removed

1 Introduction

The Grand-Duchy of Luxembourg is a country that is unique in many ways. As its official name quite clearly indicates, it is a grand duchy, the last of its kind on earth. It is a relatively young country, as it became a grand duchy once it gained independence from Napoleonic France in 1815 (but was pretty much still a puppet state of the Kingdom of the Netherlands until the end of the 19th century), is one of the founding members of the European Coal and Steel Community, which evolved to become the European Union and has three official languages: French, German, and Luxembourguish. Luxembourg is also a landlocked country, sandwiched between France, Germany and Belgium. This geographic position has given Luxembourg many advantages. One such advantage is that its labour force, which amounts to 400000 workers and is composed of 50% of French, German and Belgian commuters. This is also the reason why Luxembourg has one of the highest GDPs per capita in the world: half of its riches are produced by foreigners which are not taken into account in the computation of GDP per capita.

Half of Luxembourg's population is also composed of foreigners, the largest community being the Portuguese, followed by the French.

In this article, I posit the following hypothesis: due to its quite unique characteristics, the spread of COVID-19 in a landlocked country like Luxembourg is the exact opposite of the spread of COVID-19 that can be observed on an island country such as New Zealand, or Madagascar. A landlocked country like Luxembourg, which is furthermore highly dependent on foreign workers, has many more difficulties to control the spread of COVID-19 within its borders. Unlike an island country, a landlocked country that is highly tied to its neighbours cannot simply close its borders and put a very hard lockdown in place to control the pandemic. Or if the landlocked country does that, as soon as it opens its borders, the disease will start spreading again. To illustrate this idea, I will discuss how COVID-19 starting spreading, but not only within the borders of Luxembourg, but rather within the so-called Greater Region. The Greater Region *a space for cross-border cooperation in the heart of Europe* and is composed of the Grand-Duchy of Luxembourg, two Belgian Provinces, two French Départements and two German Bundesländer.

*This preprint was written during my free time as a private citizen, and reflects in no manner the views of the Ministry of Higher Education and Research, nor the Government of the Grand-Duchy of Luxembourg.

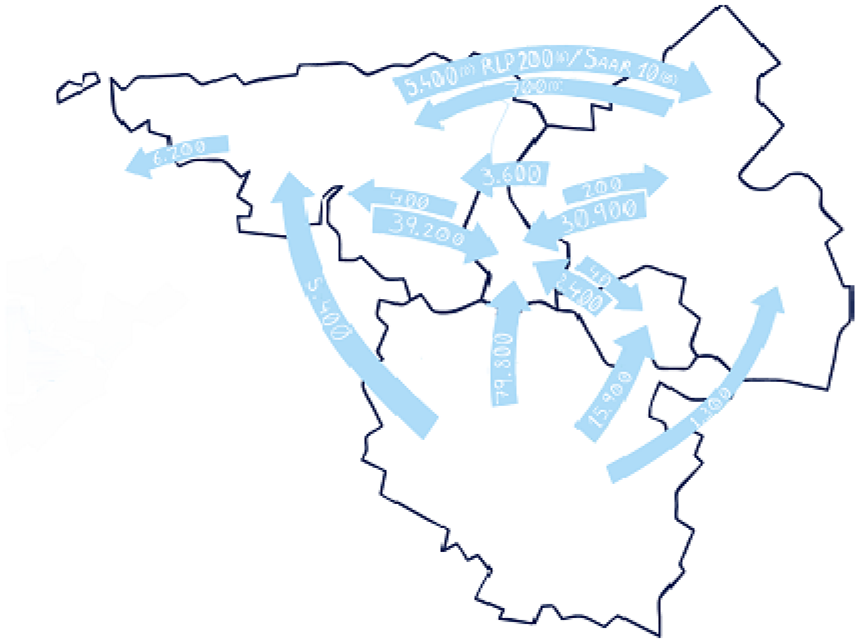


Figure 1: The daily commuters in the Greater Region. Luxembourg absorbs the vast majority. (Source: Bienvenue dans la Grande Région, 2018)

Figure 1 show a map of the Greater Region with the flows of daily commuters between its constituent regions. Every day, according to this map from 2018, more than 150000 commuters go to Luxembourg to work. In 2019, it was reported that this number reached 200000.[2]

The approach I will be using in this paper is thus as follows: I will train a machine learning model to predict the spread of COVID-19 in Luxembourg using openly available data on the weekly positive cases of COVID-19. However, because of the very tight economic and social integration of Luxembourg to its neighbours I will use as features weekly positive cases in the border regions as well as Google Mobility data[3] for Luxembourg to proxy for hard, and soft, lockdowns. I will show that lags of weekly cases in the neighbouring regions predict cases for Luxembourg. The end goal however, is *not* to build a model to predict how many weekly positive cases will be detected in Luxembourg. This would be a fools errand; in order to predict the future, the future has to look like the past, but in the case of this pandemic there is absolutely no guarantee that the future will look like the past, and there are many reasons for this. First of all, people are constantly adapting their behaviour, and public health policies are also constantly being tuned, and getting sometimes more restrictive, sometimes more relaxed. Secondly, vaccines have started being administrated and it would be impossible to predict the effect on weekly positive cases using the approach I'm using. Finally, there's also the threat of the variants. Here again, it is impossible to predict which new variants could arise and how much more contagious -and deadly- these could be. So then, why bother with this paper? The end goal is not prediction, but explainability. Once the model is trained, I will use explainability methods to show which variables, and their interaction with each other, predict positive cases for Luxembourg. This will be a clear illustration of the hypothesis that I posited at the beginning; that a landlocked country like Luxembourg which is very tightly economically and socially integrated with its neighbours cannot fight a pandemic on its own, but must cooperate with its neighbours. This argument can also be applied to any other country in the same situation as Luxembourg or even to the constituent states of a federal nation. Unfortunately, the virus does not respect the borders of sovereign nations.

2 The COVID-19 pandemic in the Greater Region

COVID-19 started to spread in the Greater Region around end of February. Two french *départements*, the *Bas-Rhin* and the *Haut-Rhin*, which together form the historical region of Alsace, were hit very hard. A

[2] `https://paperjam.lu/article/plus-200-000-frontaliers-sur-m`
[3] `https://www.google.com/covid19/mobility/`

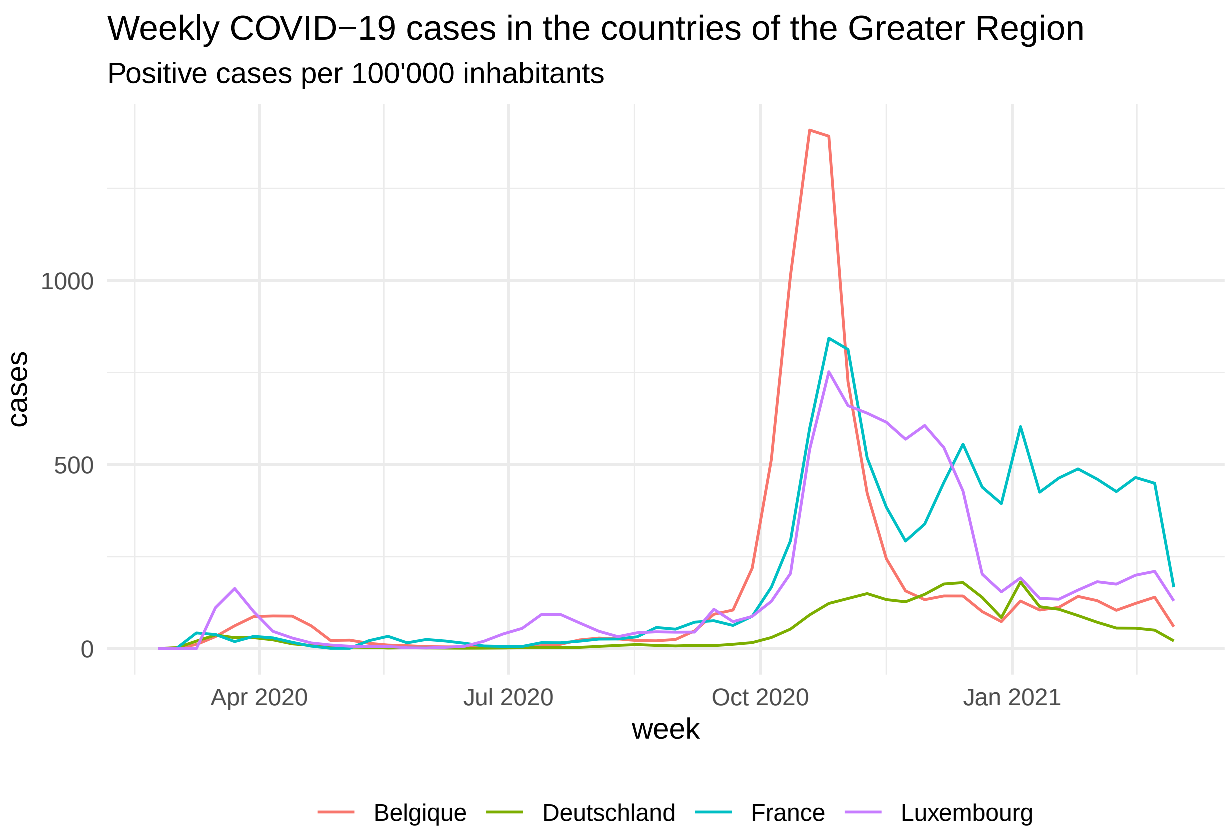


Figure 2: Test

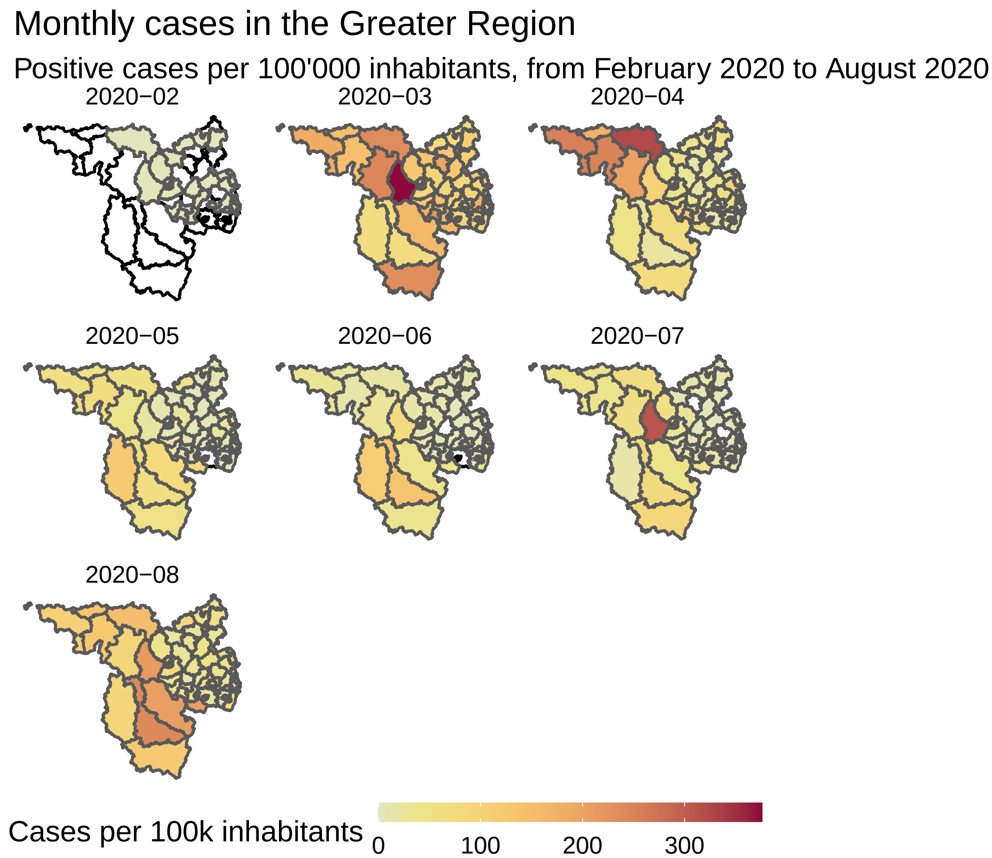
religious gathering in the *Haut-Rhin*, with more than 2000 worshipers is the very likely start of the spread in Alsace, and then to the neighbouring Lorraine administrative region. Alsace is part of the *Grand Est* region, as is Lorraine. Two *départements* of *Lorraine* are part of the Greater Region, and the disease spread there quickly as well.

In Luxembourg, the first identified case was announced during a press conference on the 29th of February 2020. However, in the publicly available data, cases start on the week of the 16th. This was never, as far as I know, explained. However, in August 2020 it was announced that positive cases detected among non-residents would not be counted anymore. The only logical conclusion is that the first cases, up until the 16th of March, were of non-residents.

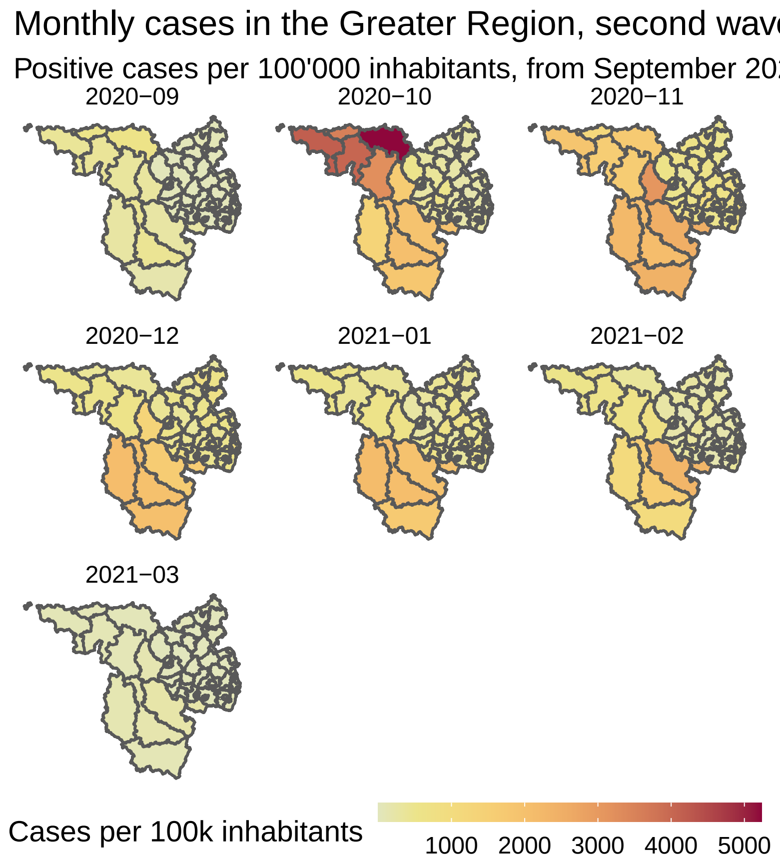
```
tar_read(epid_curves)
```

In Figure @ref(fig:epid_curves)

```
tar_read(epidem_map)$map_first_wave
```



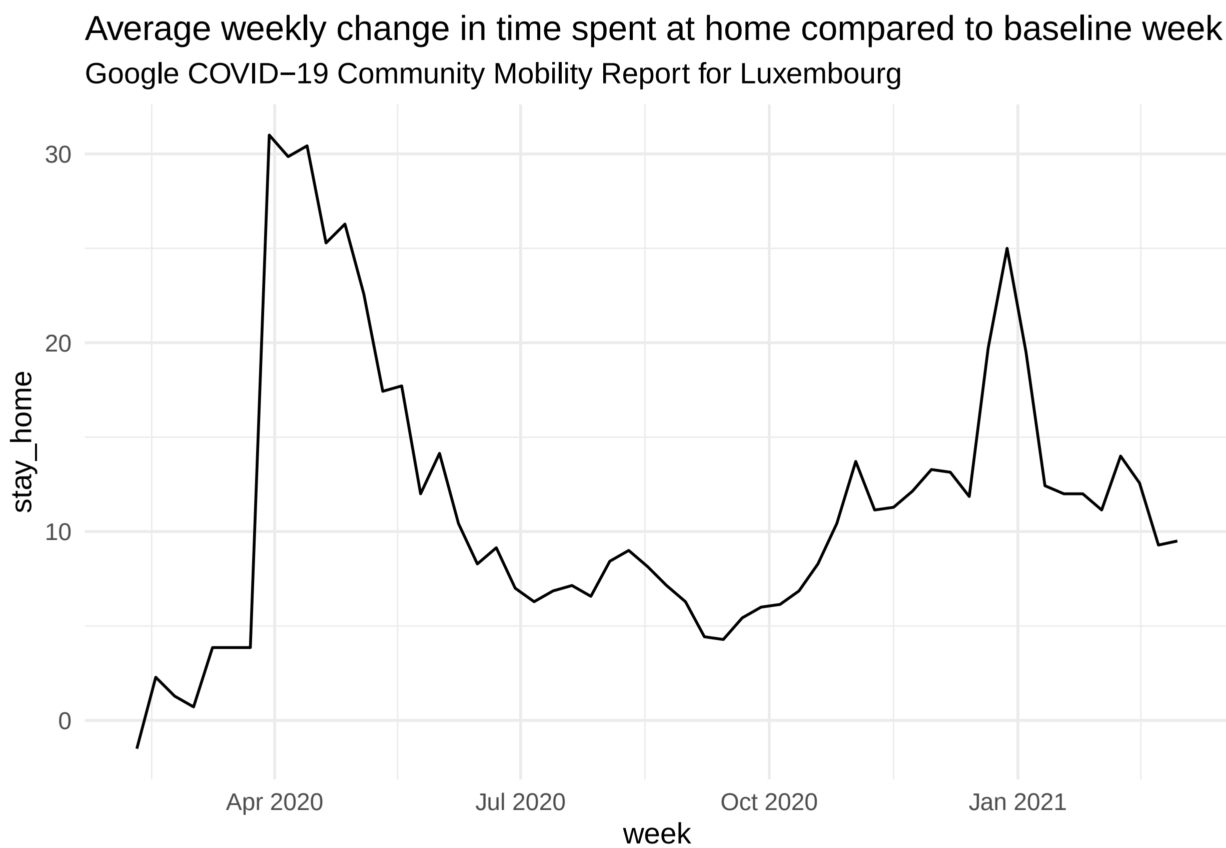

```
tar_read(epidem_map)$map_second_wave
```



3 The data used

Data on positive cases from the regions of the Greater Region was collected through each of the countries' open data portal. The levels of detail were heterogeneous, with Belgium and Germany providing a high level of detail (daily cases by sex, age group, Province in the case of Belgium, and Land- and Stadtkreise in the case of Germany), followed by France (daily cases by department and age group), with Luxembourg providing the least amount of details; only daily cases at the national level. In order to simplify the process of getting the data from all these sources, we wrote an R package called `{covidGrandeRegion}` which can be found on the following github repository. This R package provides several functions to download daily or weekly data, either for one single country or for the whole of the Greater Region as well as a function to call an interactive map of the region with a timeline, making it easy to visualise the spread of the disease through the region. It is also possible to normalize the data by dividing the daily or weekly cases by the size of the population in each sub-region. Another variable that was included comes from the Google Mobility website. This data shows on a daily basis how communities move since the start of the pandemic. This data is used here as proxy for lockdowns. The plot below shows the daily percentage change in time spent at home in Luxembourg:

```
tar_read(plot_mobility)
```



The hard lockdown from March and April can be clearly seen in the data. Weekly positive cases where divided by the population size in each region and all the variable were lagged up to four times. The goal of lagging these variables is to see if cases that were detected in neighbouring countries in the past 4 weeks have predictive power. The same was done with the mobility variable (time spent at home in Luxembourg). The following table shows the structure of data used in this study:

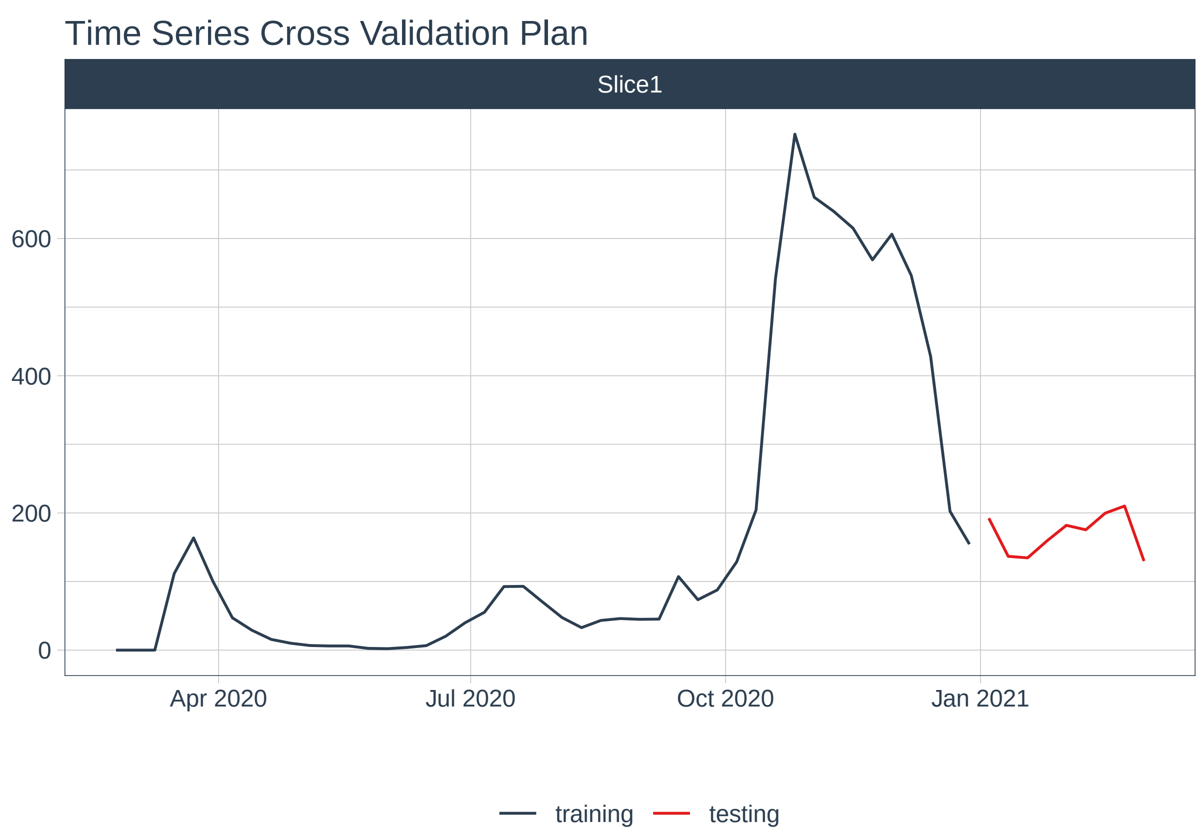
```
print_data <- tar_read(data_for_model) %>%
  mutate(across(where(is.numeric), ~round(., 0))) %>%
  t() %>%
  as.data.frame() %>%
  tibble::rownames_to_column(var="Variable names") %>%
  select(seq(1, 6)) %>%
  unite("First 5 observations", 2:6, sep = ", ")

kbl(print_data, booktabs = TRUE, position = "center")
```

Variable names	First 5 observations
week	2020-02-24, 2020-03-02, 2020-03-09, 2020-03-16, 2020-03-23
Luxembourg	0, 0, 0, 112, 164
lag_Belgique_01	0, 0, 3, 11, 33
lag_Belgique_02	0, 0, 0, 3, 11
lag_Belgique_03	0, 0, 0, 0, 3
lag_Belgique_04	0, 0, 0, 0, 0
lag_Deutschland_01	0, 1, 3, 21, 37
lag_Deutschland_02	0, 0, 1, 3, 21
lag_Deutschland_03	0, 0, 0, 1, 3
lag_Deutschland_04	0, 0, 0, 0, 1
lag_France_01	0, 0, 2, 43, 39
lag_France_02	0, 0, 0, 2, 43
lag_France_03	0, 0, 0, 0, 2
lag_France_04	0, 0, 0, 0, 0
lag_stay_home_01	0, 1, 1, 4, 4
lag_stay_home_02	0, 0, 1, 1, 4
lag_stay_home_03	0, 0, 0, 1, 1
lag_stay_home_04	0, 0, 0, 0, 1

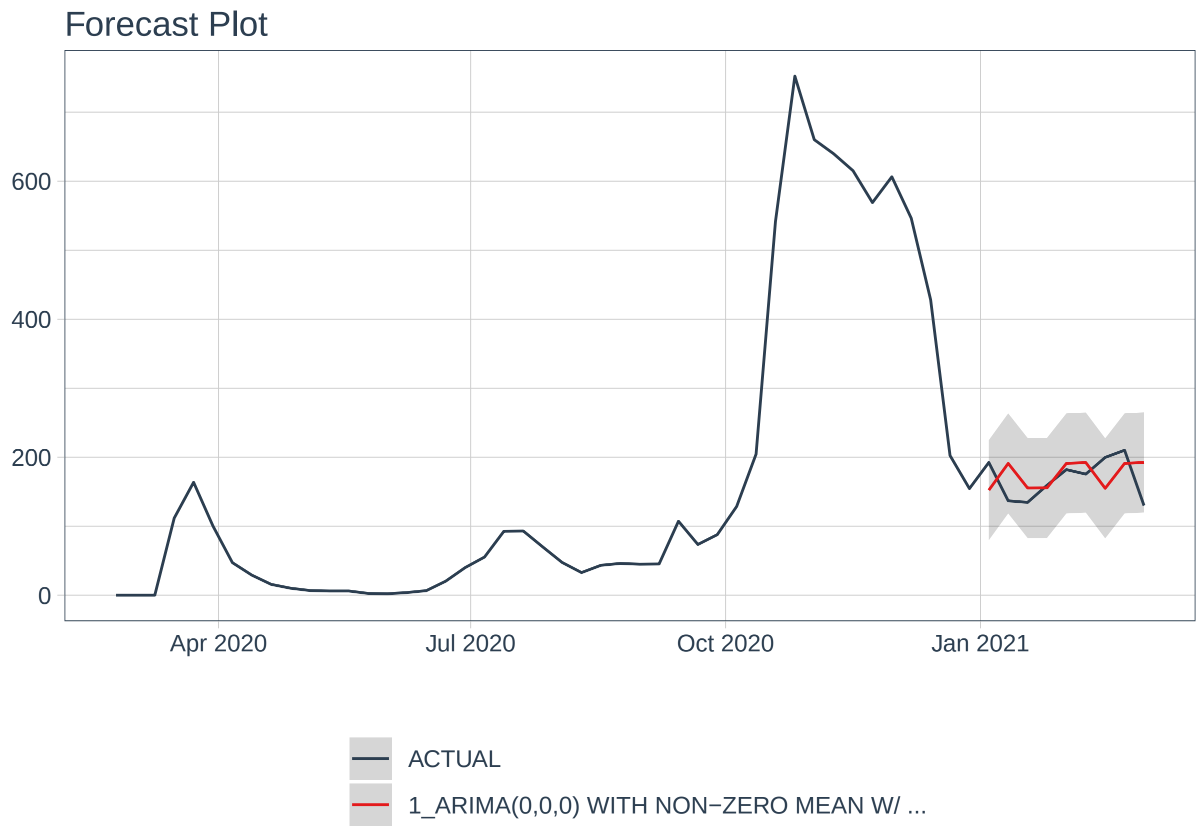
`Luxembourg` is the target variable.

```
tar_read(cv_plan)
```

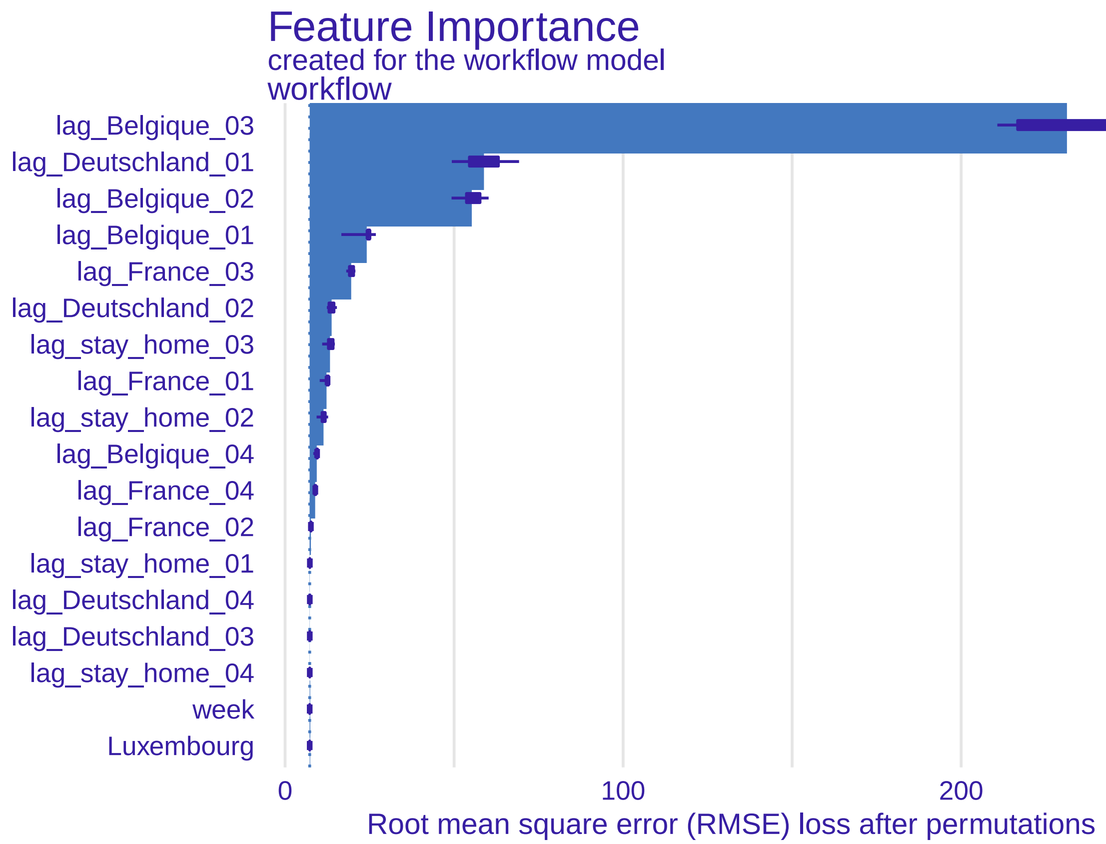



```
tar_read(forecast_plot)
```

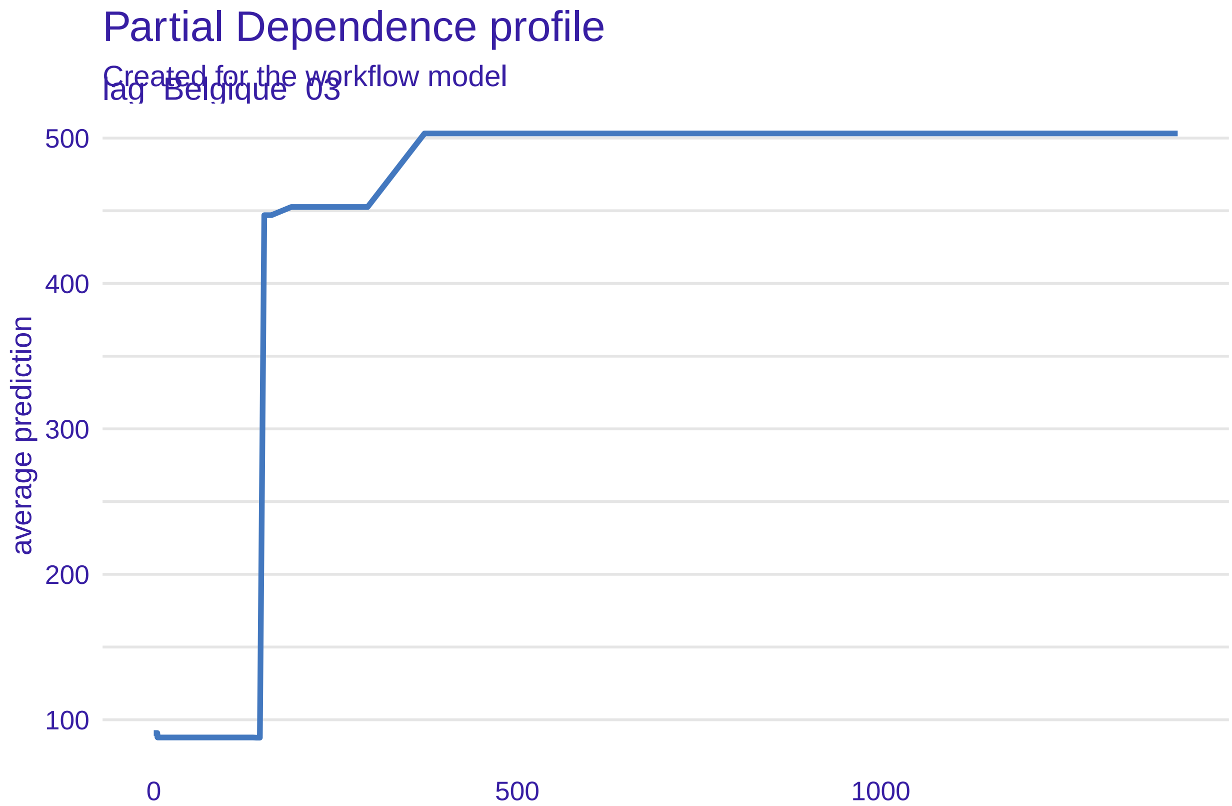
```
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```



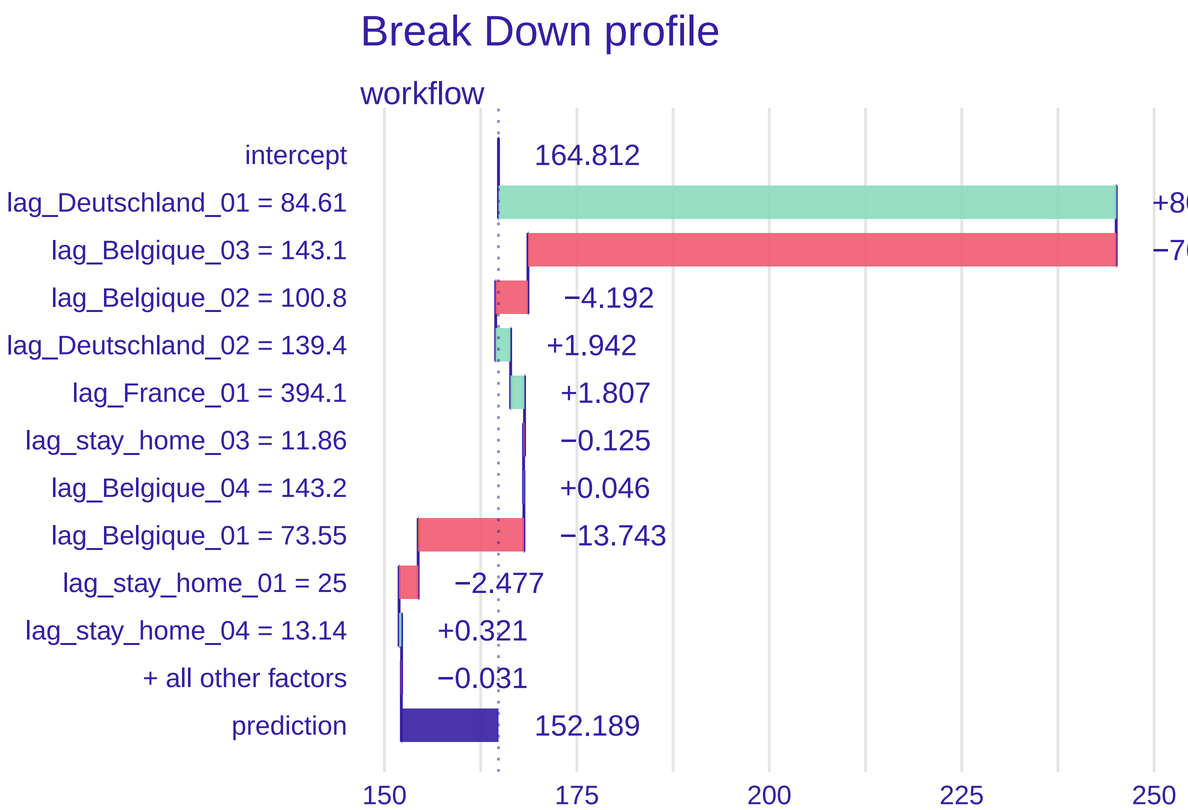

```
tar_read(plot_var_imp)
```




```
tar_read(plot_var_resp)
```




```
tar_read(plot_pred_contributions)
```


Break Down profile
workflow
intercept
lag_Deutschland_01 = 84.61
lag_Belgique_03 = 143.1
lag_Belgique_02 = 100.8
lag_Deutschland_02 = 139.4
lag_France_01 = 394.1
lag_stay_home_03 = 11.86
lag_Belgique_04 = 143.2
lag_Belgique_01 = 73.55
lag_stay_home_01 = 25
lag_stay_home_04 = 13.14
+ all other factors
prediction
164.812
−4.192
+1.942
+1.807
−0.125
+0.046
−13.743
−2.477
+0.321
−0.031
152.189
150
175
200
225
250

4 Appendix