Real-time analysis of predictors of COVID-19 infection spread in the countries of the European Union
User's Manual

24 October, 2021

Table of Contents

Intro

This is a user's manual for the COVID-19 visualization app intended for a public of interest in Epidemiological research and in Statistics and Machine Learning.

You can access our app by opening it on our server at http://www.covidmdmmasterteam.tk:3838/.

1. **What is this COVID-19 app good for?**

The COVID-19 app is an interactive visualization tool of COVID-19 related data for the countries of the European Union.

- You can *explore and compare* the number of COVID-19 infections by countries along with many predictors of the infections like mask usage, direct contact, vaccination, average daily temperature, and restriction measures applied in the selected countries.
- You can check the *effect of the predictors* one by one per country on our model of COVID-19 new confirmed infections (proportionate to population size) on partial dependence graphs.
- You can *compare* the differences between *predictor importance* on COVID-19 new infections between countries on an interactive bump chart.
- You can check the *effect of country characteristics* on predictor importance rankings on a map.
- You can find the *data sources* and background info.

In order to use our app, you can simply click here (server version) and the app opens straight away.

2. **What do you need to use this app?**

You only need

- a computer with internet connection (the app is not fully adapted to phone or tablet)
- Internet browser.

3. How does this app work? What does it show to you?

Just click on this and the app opens in your browser.

On the left-hand side you can choose between the tabs. A menu of the tabs appears when clicking the icon highlighted below (see Fig. 1).
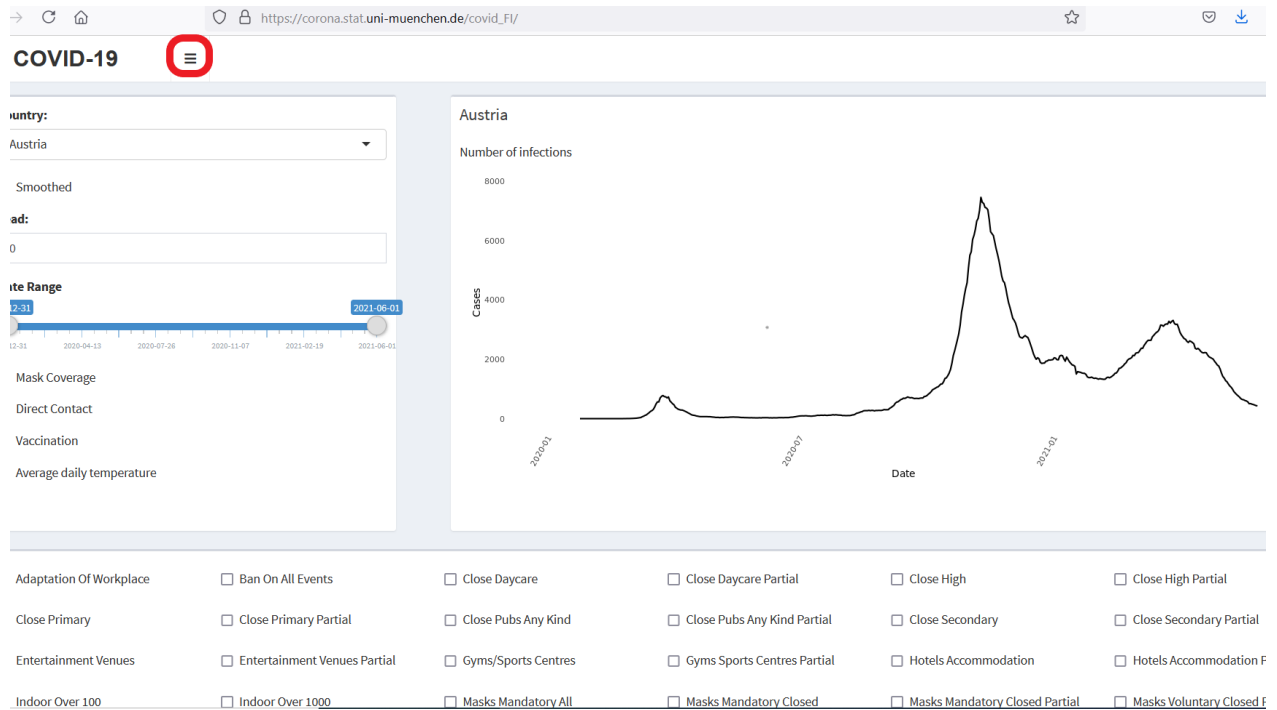

Figure 1. Accessing the tabs of the dashboard

3.1 Exploratory tab

The application starts with the Exploratory tab (see Fig. 1). On this tab, the number of infections is displayed per country for the whole time interval. You can select a country from a dropdown menu. You can choose between the smoothed and the unsmoothed version of the continuous variables (default is smoothed). The visualization of the smoothed version of the variables enables to detect a trend more easily, for example in the case of the number of new COVID-19 cases due to the uneven data provision (like systematically lower number of cases at weekends in many countries). To explore the trends in vaccination we recommend to take a look at the unsmoothed version as well, because due to the high number of days without reported vaccinations in some cases no smoothed versions were calculated for longer

periods. Also, you can set the time interval and add additional variables to the chart (mask coverage, direct contact, vaccination, average daily temperature, and restriction measures applied in the selected country). Since the change in the number of new cases follows the change of the predictors with a delay (Dehning et al. 2020), you can add a lead to the number of infections to make it better comparable with the other variables that can be added to the chart.
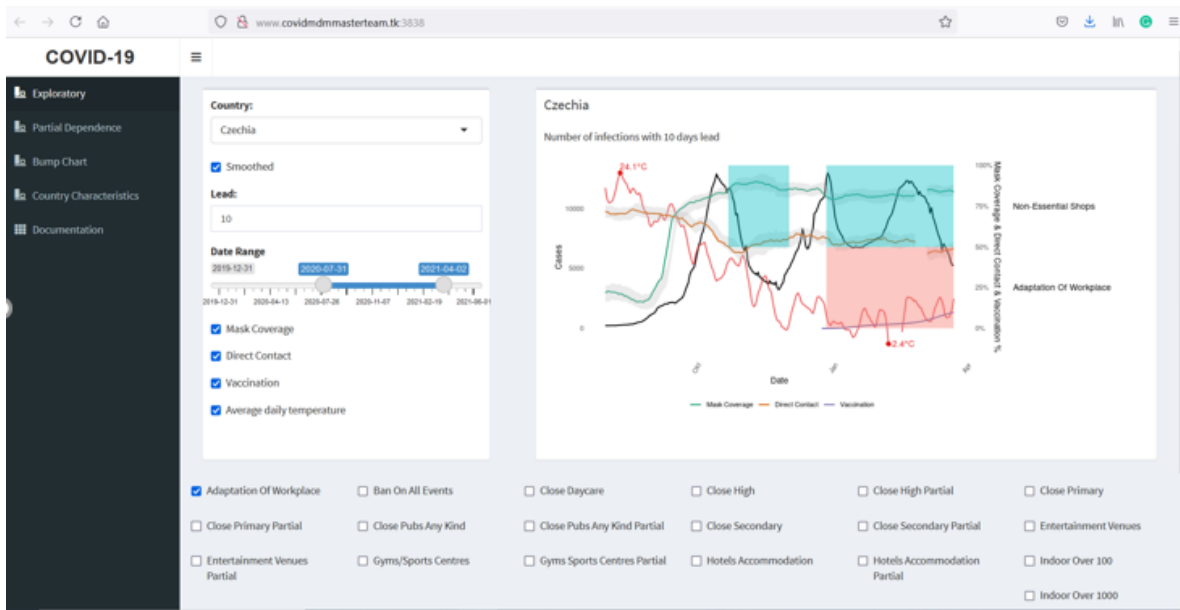

Figure 2. Exploratory tab of the dashboard

Example interpretation

Figure 2 presents an example of the exploratory tab of the application. The number of infections in Czechia between 2020.07.31 and 2021.04.02 is shown on the black line with a 10-day lead. The continuous variables are smoothed. Mask coverage, direct contact, vaccination, and average daily temperature are added to the plot. The time intervals, in which the selected restriction measures were applied are also shown: the non-essential shops were closed in November and again from January 2021 until the end of the selected time interval, the workplaces were adapted also since January, the usage of mask coverage was mandatory in all spaces from autumn 2020 until the end of the selected time interval.

3.2 Partial Dependence tab

This is the second tab on the left-hand side menu of our app. On this tab, you can see different versions of a partial dependence plot based on our country-level Random Forest model predictions[1] (see Fig. 3).

A partial dependence plot illustrates the functional relationship between a predictor and our prediction on new COVID-19 cases, proportionate to country population size. It shows how the prediction partially depends on the values of the selected predictor, it can also show the type of relationship, such as a step function, curvilinear, linear, and so on.

A rug plot is added to the x axis to show the distribution of the selected predictor in order to avoid overinterpretation of some regions of the axis with sparse data.

On this tab, you can select a country and any of the response measures from the relevant drop-down menus, and see the Partial Dependence Plot with a rug function accordingly.
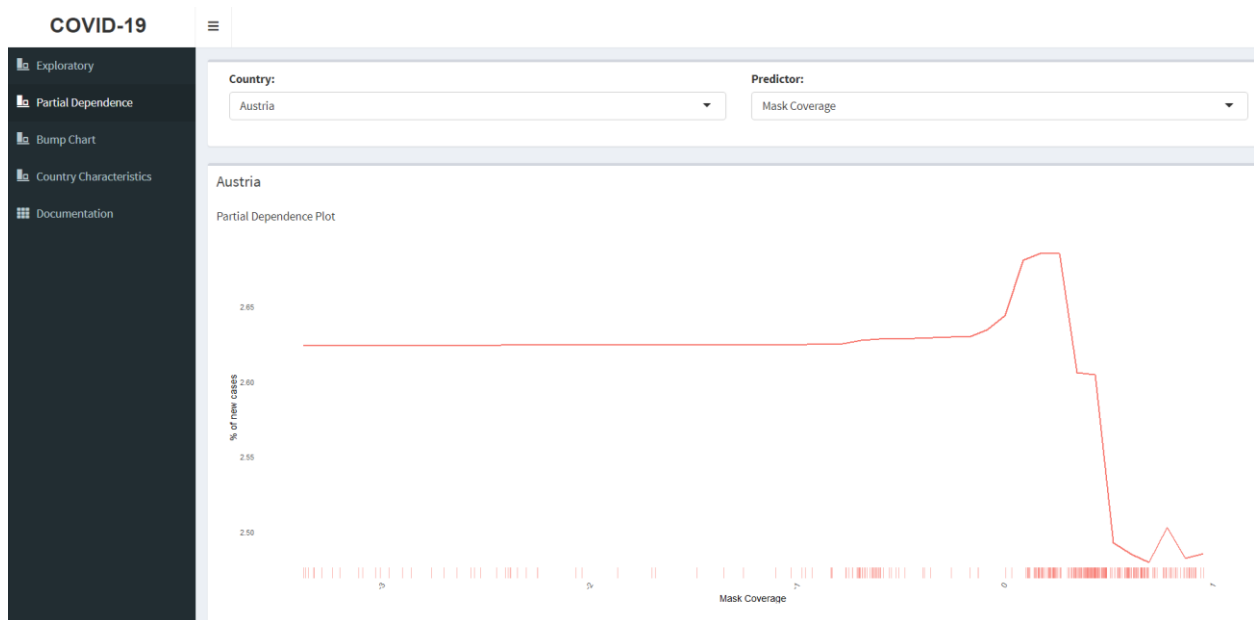

Figure 3. Partial Dependence Plot tab of the dashboard

---

[1] You can find more details in the Methodological background: Modeling.

Example interpretation

The plot below shows the relationship (according to the model that we trained) between the new COVID-19 infections proportionate to population country size and the weighted percentage of respondents that have reported using a mask in Austria. Here, we see that the number of new COVID-19 infections sharply drops as the percentage of respondents with direct contact increases. The rug plot orientates us to the region of the x axis, where most of the cases occur.

3.3 Bump Chart tab

Bump chart is a particular kind of line chart that usually focuses on the comparison of positions of observations related to one another, and is often used to express changes in rank over time (R-bloggers 2018). However, instead of time, in this project we use a bump chart for the visualization of differences in the rank of predictors by their permutation feature importance[2] over countries (see Fig. 4).

On the right side of the tab, you can select predictors and countries for the visualization. Also, all predictors and countries can be selected or unselected with the action buttons under the chart. The predictors are ordered according to the sum of the reciprocal values of their ranks by their permutation feature importance in the countries in which they were applied. In this way the lower ranks have a larger weight in selecting the predictors with the overall highest feature importance.

---

[2] The feature importance measures the contribution of a predictor in predicting the response. "Permutation feature importance measures the increase in the prediction error of the model after we permuted the feature's values, which breaks the relationship between the feature and the true outcome.", as Molnar states (Molnar 2021).
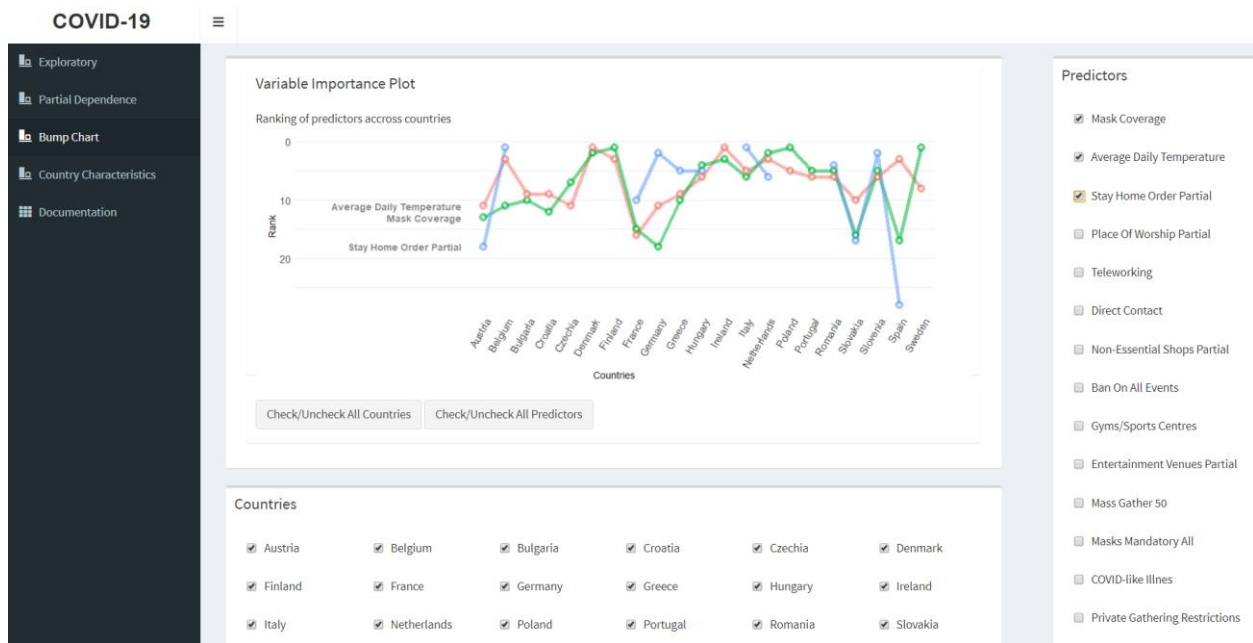
Figure 3. Bump Chart tab of the dashboard

Example interpretation

All countries' and the predictors' average daily temperature, percentage of the population using mask coverage, and application of the restriction measure stay-at-home order for the general population (also referred as "lockdown") are selected. As we see, the variable importance of the same predictors can vary widely between the countries.

3.4 Country Characteristics Tab

On this tab you can compare the effect of predictors on the number of confirmed new COVID-19 cases in countries with different characteristics (like age, gender distribution, health expenditure and cultural participation) and see the relation between country characteristics and the predictor importance rank of the COVID-19 new infections (see Fig. 4). The countries were assigned to groups with hierarchical clustering[3]. The predictors are ranked by their feature importance (variable with highest feature importance in a country has rank 1, variable with second highest variable importance got rank 2, and so on.), where the variable importance ranks were created by country-level and cluster-level Random Forest

---

[3] More details about the hierarchical clustering can be found in the Methodological Background: Hierarchical cluster.

modeling, respectively. The similarity of the rank of the predictors between countries and their clusters are visualized here.

Two maps can be found on this tab:

- You can see the country groups (clusters with similar country characteristics) on the first map at the top. This is a static map.
- The second map is an interactive one: you can select a cluster (group of countries) and the app shows how its countries' predictor ranks are typical for their cluster, based on their country characteristics: you can also select a 4-category indicator showing overall/high/medium/low rank correlation between countries' predictor importance rank and their cluster's predictor importance rank.

This way you can check for each country whether its most important predictors of COVID-19 daily new infections are typical for countries with similar country characteristics or not, thus *evaluating the underlying hypothesis that country characteristics (listed above) are related to the most important predictors of new COVID-19 infections* and *countries with similar country characteristics have similar rank of COVID-19 new infection predictors*.
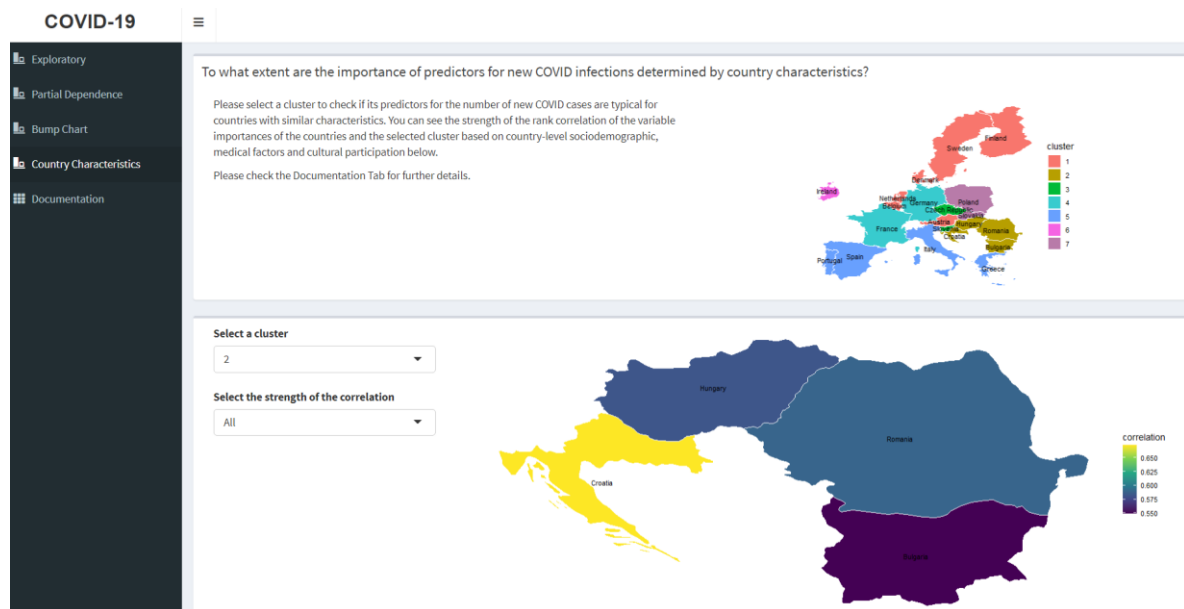

Figure 4. Country Characteristics tab of the dashboard

Example interpretation

The tab above, on the first map shows the clusters (country groups with similar country characteristics) created with hierarchical clustering[4]. The second, interactive map shows the strength of the correlation of COVID-19 predictor importance ranks between the selected second cluster consisting of Croatia, Hungary, Romania and Bulgaria and the single countries within the cluster. The map shows that the cluster-level COVID-19 predictor variable importance rank, - i.e. the list of the variable importance of predictors on COVID-19 infections sorted by their relevance (most important predictors are at the top of the list, the least important are at the bottom) -, is quite similar to the country-level COVID-19 predictor variable importance ranks with a correlation varying from 0.55 to 0.67 by countries of the cluster. The order of the COVID-19 predictor variable importance rank in Croatia resembles the most to the same measure at the cluster level with a rank correlation over 0.65, while Bulgaria's order of the COVID-19 predictor variable importance rank is the least similar to the cluster-level order of the COVID-19 predictor variable importance rank.

Additionally, you can select low, middle or high rank correlation, this way the app shows only the relevant countries of the selected cluster with the selected level of rank correlation.

3.5 Documentation Tab

On the last tab you can find background information of our app (see Fig. 5). You can find the data sources and User's Manual (the document you are reading) there, which contains detailed description of how the app works. After the desired acceptance of our paper, the data collection process and modeling behind, as well as bibliography references and all the R codes to reproduce data collection, cleansing, data manipulation, modeling and building of the app will also be added here.

---

[4] Detailed description of the clustering process and the clusters can be found in the Methodological background: Hierarchical cluster.
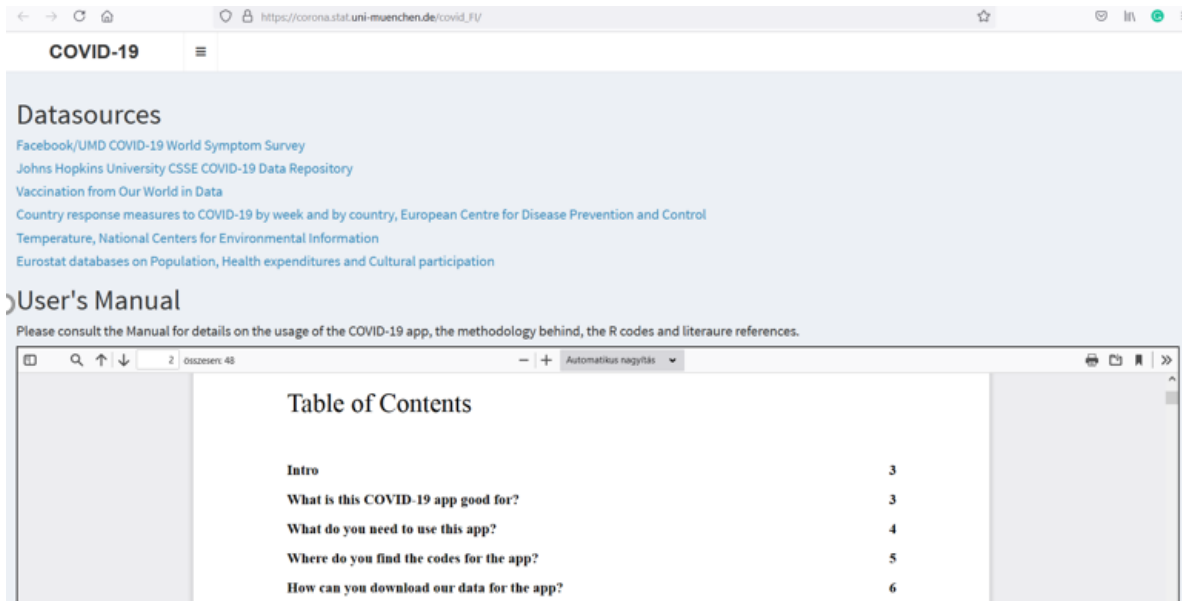
Figure 5. Documentation tab of the dashboard

4. Where do you find more details on our data and methodology?

Our data is collected (web-scraped) and merged from various sources. After the desired acceptance of our full paper the list and description of available variables, the various methods of web scraping will be added in the Appendix: Data Sources and Database Description. The reference for the data sources can be found in the Appendix.

The Bump chart and the Partial Dependence chart are based on a group of country-level Random Forest models. The Country Characteristics Tab uses hierarchical clustering, cluster-level Random Forest models and rank correlation.

After the desired acceptance of our full paper you will find the methodology behind the Partial Dependence Tab, the Bump Chart Tab and the Country Characteristics Tab in the Methodological Background chapter.

5. Methodological background

After the desired acceptance of our full paper, this chapter will be added to the User's Manual.

5.1 Data collection and preparation

5.2 Modeling

5.2.1 Times series methods revisited

5.2.2 Model selection

5.2.3 Our model

5.2.4 Time series cross-validation

5.2.5 Future plans to improve our predictions

5.3 Hierarchical cluster

5. 3.1. Results

5.4 Rank correlations

5.5 Shiny Dashboard – description

5.6. Main findings

Appendix

1. Data sources and Database Description
2. Code structure *(will be added upon acceptance of our paper)*
3. Bibliography *(will be added upon acceptance of our paper)*
4. List of figures *(will be added upon acceptance of our paper)*
5. R packages

1. Data sources and Database Description

1.1. Time-constant country characteristics from Eurostat sources

Name of database: country_char

Sources and variables:

- Data source: eurostat, https://ec.europa.eu/eurostat
- Population: population of 1. January 2019 by age group and sex
    - Variables:
        - geo: Geopolitical entity
        - Y_LT5: Less than 5 years
        - Y5-9: From 5 to 9 years
        - Y10-14: From 10 to 14 years
        - Y15-19: From 15 to 19 years
        - Y20-24: From 20 to 24 years
        - Y25-29: From 25 to 29 years
        - Y30-34: From 30 to 34 years
        - Y35-39: From 35 to 39 years
        - Y40-44: From 40 to 44 years
        - Y45-49: From 45 to 49 years
        - Y50-54: From 50 to 54 years
        - Y55-59: From 55 to 59 years
        - Y60-64: From 60 to 64 years
        - Y65-69: From 65 to 69 years
        - Y70-74: From 70 to 74 years
        - Y75-79: From 75 to 79 years
        - Y_GE75: 75 years or over
        - Y80-84: From 80 to 84 years
        - Y_GE80: 80 years or over
        - Y_GE85: 85 years or over
        - T: Total
        - M: Males
        - F: Females
- Health expenditures: Total health care expenditure amount in millions of euro, 2018
    - Variables:
        - health_expenditures: Total health expenditures
- Cultural participation: Frequency of participation in cultural activities in the last 12 months by age, 2015. Percentage of those who did not attend any cultural event (cinema, live performances or cultural sites) in the last 12 months by age groups
    - Variables:
        - cult_Y_GE16 : 16 years or over

- cult_Y_GE75 : 75 years or over
- cult_Y16-24: From 16 to 24 years
- cult_Y16-29: From 16 to 29 years
- cult_Y25-34: From 25 to 34 years
- cult_Y25-64: From 25 to 64 years
- cult_Y35-49: From 35 to 49 years
- cult_Y50-64: From 50 to 64 years
- cult_Y65-74: From 65 to 74 years

1.2. Time-varying COVID-19 related variables from various sources

Name of database: tdata

Sources and variables:

- Testing:
  - Data source: European Centre for Disease Prevention and Control, Data on testing for COVID-19 by week and country, https://www.ecdc.europa.eu/en/publications-data/covid-19-testing
  - Testing volume and positivity rate by week
  - Variables:
    - country
    - country_code: 2-letter ISO country code
    - testing_new_cases: Number of new confirmed cases
    - tests_done: Number of tests done
    - testing_population
    - testing_rate: Testing rate per 100 000 population
    - testing_positivity_rate: Weekly test positivity (%): 100 x Number of new confirmed cases/number of tests done per week
- Response measures:
  - Data source: European Centre for Disease Prevention and Control, Data on country response measures to COVID-19 by week and country, https://www.ecdc.europa.eu/en/publications-data/download-data-response-measures-covid-19
  - Non-pharmaceutical interventions taken by countries in response to the pandemics
  - Variables:
    - Country
    - date
    - year
    - week

- AdaptationOfWorkplace: Adaptation of workplaces(e.g. to reduce risk of transmission)
- AdaptationOfWorkplacePartial: Adaptation of workplaces (e.g. to reduce risk of transmission)-partially relaxed measure
- BanOnAllEvents: Interventions are in place to limit all indoor/outdoor mass/public gatherings
- BanOnAllEventsPartial: Interventions are in place to limit all indoor/outdoor mass/public gatherings-partially relaxed measure
- ClosDaycare: Closure of educational institutions: daycare or nursery.
- ClosDaycarePartial: Closure of educational institutions: daycare or nursery -partially relaxed measure
- ClosHigh: Closure of educational institutions: higher education.
- ClosHighPartial: Closure of educational institutions: higher education -partially relaxed measure
- ClosPrim: Closure of educational institutions: primary schools.
- ClosPrimPartial: Closure of educational institutions: primary schools -partially relaxed measure
- ClosPubAny: Closure of public spaces of any kind (including restaurants, entertainment venues, non-essential shops, partial or full closure of public transport, gyms and sport centers, etc).
- ClosPubAnyPartial: Closure of public spaces of any kind (including restaurants, entertainment venues, non-essential shops, partial or full closure of public transport, gyms and sport centers etc) -partially relaxed measure
- ClosSec: Closure of educational institutions: secondary schools.
- ClosSecPartial: Closure of educational institutions: secondary schools -partially relaxed measure
- ClosureOfPublicTransport: Closure of public transport
- ClosureOfPublicTransportPartial: Closure of public transport-partially relaxed measure
- EntertainmentVenues: Closure of entertainment venues
- EntertainmentVenuesPartial: Closure of entertainment venues-partially relaxed measure
- GymsSportsCentres: Closure of gyms/sports centres
- GymsSportsCentresPartial: Closure of gyms/sports centres-partially relaxed measure
- HotelsAccommodation: Closure of hotels/accommodation services
- HotelsAccommodationPartial: Closure of hotels/accommodation services-partially relaxed measure
- IndoorOver100: Interventions are in place to limit indoor mass/public gatherings of over 100participants

- IndoorOver1000: Interventions are in place to limit indoor mass/public gatherings of over 1000participants
- IndoorOver1000Partial: Interventions are in place to limit indoor mass/public gatherings of over 1000participants-partially relaxed measure
- IndoorOver100Partial: Interventions are in place to limit indoor mass/public gatherings of over 100participants-partially relaxed measure
- IndoorOver50: Interventions are in place to limit indoor mass/public gatherings of over 50participants
- IndoorOver500: Interventions are in place to limit indoor mass/public gatherings of over 500participants
- IndoorOver500Partial: Interventions are in place to limit indoor mass/public gatherings of over 500participants-partially relaxed measure
- IndoorOver50Partial: Interventions are in place to limit indoor mass/public gatherings of over 50participants-partially relaxed measure
- MasksMandatoryAllSpaces: Protective mask use in all public spaces on mandatory basis (enforced by law)
- MasksMandatoryAllSpacesPartial: Protective mask use in all public spaces on mandatory basis (enforced by law)-partially relaxed measure
- MasksMandatoryClosedSpaces: Protective mask use in closed public spaces/transport on mandatory basis (enforced by law)
- MasksMandatoryClosedSpacesPartial: Protective mask use in closed public spaces/transport on mandatory basis (enforced by law)-partially relaxed measure
- MasksVoluntaryAllSpaces: Protective mask use in all public spaces on voluntary basis (general recommendation not enforced)
- MasksVoluntaryAllSpacesPartial: Protective mask use in all public spaces on voluntary basis (general recommendation not enforced)-partially relaxed measure
- MasksVoluntaryClosedSpaces: Protective mask use in closed public spaces/transport on voluntary basis (general recommendation not enforced)
- MasksVoluntaryClosedSpacesPartial: Protective mask use in closed public spaces/transport on voluntary basis (general recommendation not enforced)-partially relaxed measure

- **MassGatherAll**: Interventions are in place to limit mass/public gatherings (any interventions on mass gatherings up to 1000 participants included)
- **MassGatherAllPartial**: Interventions are in place to limit mass/public gatherings (any interventions on mass gatherings up to 1000 participants included)-partially relaxed measure
- **NonEssentialShops**: Closures of non-essential shops
- **NonEssentialShopsPartial**: Closures of non-essential shops -partially relaxed measure
- **OutdoorOver100**: Interventions are in place to limit outdoor mass/public gatherings of over 100participants
- **OutdoorOver1000**: Interventions are in place to limit outdoor mass/public gatherings of over 1000participants
- **OutdoorOver1000Partial**: Interventions are in place to limit outdoor mass/public gatherings of over 1000participants-partially relaxed measure
- **OutdoorOver100Partial**: Interventions are in place to limit outdoor mass/public gatherings of over 100participants-partially relaxed measure
- **OutdoorOver50**: Interventions are in place to limit outdoor mass/public gatherings of over 50participants
- **OutdoorOver500**: Interventions are in place to limit outdoor mass/public gatherings of over 500participants
- **OutdoorOver500Partial**: Interventions are in place to limit outdoor mass/public gatherings of over 500participants-partially relaxed measure
- **OutdoorOver50Partial**: Interventions are in place to limit outdoor mass/public gatherings of over 50participants-partially relaxed measure
- **PlaceOfWorship**: Closure of places of worship
- **PlaceOfWorshipPartial**: Closure of places of worship-partially relaxed measure
- **PrivateGatheringRestrictions**: Restrictions on private gatherings
- **PrivateGatheringRestrictionsPartial**: Restrictions on private gatherings-partially relaxed measure
- **RegionalStayHomeOrder**: Regional stay-at-home orders for the general population at least in one region(these are enforced and also referred to as 'lockdown')
- **RegionalStayHomeOrderPartial**: Regional stay-at-home orders for the general population at least in one region (these are enforced and also referred to as 'lockdown')-partially relaxed measure

- ■ RestaurantsCafes: Closure of restaurants and cafes/bars
- ■ RestaurantsCafesPartial: Closure of restaurants and cafes/bars- partially relaxed measure
- ■ SocialCircle: Social circle/bubble to limit social contacts e.g. to limited number of households
- ■ SocialCirclePartial: Social circle/bubble to limit social contacts e.g. to limited number of households-partially relaxed measure
- ■ StayHomeGen: Stay-at-home recommendations for the general population (which are voluntary or not enforced)
- ■ StayHomeGenPartial: Stay-at-home recommendations for the general population (which are voluntary or not enforced) -partially relaxed measure
- ■ StayHomeOrder: Stay-at-home orders for the general population (these are enforced and also referred to as 'lockdown')
- ■ StayHomeOrderPartial: Stay-at-home orders for the general population (these are enforced and also referred to as 'lockdown') - partially relaxed measure
- ■ StayHomeRiskG: Stay-at-home recommendations for risk groups or vulnerable populations (such as the elderly, people with underlying health conditions, physically disabled people, etc.)
- ■ StayHomeRiskGPartial: Stay-at-home recommendations for risk groups or vulnerable populations (such as the elderly, people with underlying health conditions, physically disabled people, etc.) - partially relaxed measure
- ■ Teleworking: Teleworking recommendation
- ■ TeleworkingPartial: Teleworking recommendation or workplace closures -partially relaxed measure
- ■ WorkplaceClosures: Closures of workplaces
- ■ WorkplaceClosuresPartial: Closures of workplaces-partially relaxed measure
- ● Weather:
  - ○ Data source: National Centers for Environmental Information, https://www.ncdc.noaa.gov/
  - ○ Daily average temperatures measured in the capitals of the countries involved in the analysis
  - ○ In a few cases some values are missing for shorter periods for some countries, for example: values are missing for Spain between 03.11.2020 and 03.14.2020.
  - ○ Variables:
    - ■ country_code: 2-letter ISO country code
    - ■ date

- ■ tavg: average daily temperature
- Vaccination:
  - Our World in Data, https://ourworldindata.org/coronavirus
  - Variables:
    - iso_code: ISO country code
    - country
    - date
    - total_vaccinations: total number of doses administered
    - people_vaccinated: total number of people who received at least one vaccine dose. If a person receives the first dose of a 2-dose vaccine, this metric goes up by 1. If they receive the second dose, the metric stays the same.
    - people_fully_vaccinated: total number of people who received all doses prescribed by the vaccination protocol. If a person receives the first dose of a 2-dose vaccine, this metric stays the same. If they receive the second dose, the metric goes up by 1.
    - new_vaccinations: daily change in the total number of doses administered
    - new_vaccinations_smoothed: new doses administered per day (7-day smoothed (for countries that don't report data on a daily basis, the daily changes on doses assumed to be equal over the period in which no data was reported)
    - total_vaccinations_per_hundred: people vaccinated per 100 people in the total population of the country
    - people_vaccinated_per_hundred: people vaccinated per 100 people in the total population of the country.
    - people_fully_vaccinated_per_hundred: people fully vaccinated per 100 people in the total population of the country.
    - new_vaccinations_smoothed_per_million: daily vaccinations per 1,000,000 people in the total population of the country

- Covid cases:
  - https://github.com/RamiKrispin/coronavirus
  - The coronavirus package provides a tidy format dataset of the 2019 Novel Coronavirus COVID-19 (2019-nCoV) epidemic. The raw data is pulled from the Johns Hopkins University Center for Systems Science and Engineering (JHU CCSE) Coronavirus repository.
  - Variables:
    - cases_new: Confirmed daily new cases
    - deaths_new: Daily number of deaths
    - recovered_new: Daily number of the recovered

- UMD/Facebook World Symptom survey:
  - https://covidmap.umd.edu/api/
  - variables:
    - fb_data.iso_code: ISO country codes
    - fb_data.percent_cli: weighted percentage of respondents that have reported Covid Like Illness
    - fb_data.cli_se: standard error of percent_cli
    - fb_data.percent_cli_unw: unweighted percentage of respondents that have reported CLI
    - fb_data.cli_se_unw: standard error of percent_cli_unw
    - fb_data.sample_size_cli: sample size for calculating CLI
    - fb_data.smoothed_cli: smoothed percentage of respondents that have reported Covid Like Illness
    - fb_data.smoothed_cli_se: standard error of smoothed percent_cli
    - fb_data.sample_size_smoothed_cli: sample size for calculating smoothed CLI
    - fb_data.percent_mc: weighted percentage of respondents that have reported using a mask
    - fb_data.mc_se: standard error of percent_mc
    - fb_data.percent_mc_unw: unweighted percentage of respondents that have reported use mask cover
    - fb_data.mc_se_unw: standard error of percent_mc_unw
    - fb_data.sample_size_mc: sample size for calculating mask coverage
    - fb_data.smoothed_mc: smoothed percentage of respondents that have reported use mask cover
    - fb_data.smoothed_mc_se_ standard error of smoothed percent_mc
    - fb_data.sample_size_mc_smoothed: sample size for calculating smoothed mc
    - fb_data.percent_dc: weighted percentage of respondents that have reported had direct contact (longer than one minute) with people not staying with them in last 24 hours
    - fb_data.mc_se_dc: standard error of fb_data.percent_mc
    - fb_data.percent_dc_unw: unweighted percentage of respondents that have reported use have direct contact with people not staying with them
    - fb_data.dc_se_unw: standard error of percent_dc_unw
    - fb_data.sample_size_dc: sample size for calculating direct contact
    - fb_data.smoothed_dc: smoothed percentage of respondents that have reported direct contact
    - fb_data.smoothed_dc_se: standard error of smoothed percent_dc

- fb_data.sample_size_dc_smoothed: sample size for calculating smoothed dc

5 R packages

- caret, Max Kuhn et al (2020).
  caret: Classification and Regression Training. R package version 6.0-86.
- compare, Paul Murrell (2015). compare: Comparing Objects for Differences. R package version 0.2-6. https://CRAN.R-project.org/package=compare
- coronavirus, Rami Krispin and Jarrett Byrnes (2021). coronavirus: The 2019 Novel Coronavirus COVID-19 (2019-nCoV) Dataset. R package version 0.3.1. https://CRAN.R-project.org/package=coronavirus
- Cowplot, Claus O. Wilke (2020). cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.1.1. https://CRAN.R-project.org/package=cowplot
- data.table, Matt Dowle and Arun Srinivasan (2020). data.table: Extension of `data.frame`. R package version 1.13.6. https://CRAN.R-project.org/package=data.table
- dplyr, Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.2. https://CRAN.R-project.org/package=dplyr,
- eurostat, (C) Leo Lahti, Janne Huovari, Markus Kainu, Przemyslaw Biecek. Retrieval and analysis of Eurostat open data with the eurostat package. R Journal 9(1):385-392, 2017. Version 3.6.84 Package URL: http://ropengov.github.io/eurostat Manuscript URL: https://journal.r-project.org/archive/2017/RJ-2017-019/index.html
- factoextra, Alboukadel Kassambara and Fabian Mundt (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7. https://CRAN.R-project.org/package=factoextra
- googlesheets4, Jennifer Bryan (2020). googlesheets4: Access Google Sheets using the Sheets API V4. R package version 0.2.0. https://CRAN.R-project.org/package=googlesheets4
- gridextra, Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra
- Hmisc, Frank E Harrell Jr, with contributions from Charles Dupont and many others. (2020). Hmisc: Harrell Miscellaneous. R package version 4.4-2. https://CRAN.R-project.org/package=Hmisc

  https://CRAN.R-project.org/package=caret
- httr, Hadley Wickham (2020). httr: Tools for Working with URLs and HTTP. R package version 1.4.2. https://CRAN.R-project.org/package=httr
- iml, Molnar C, Bischl B, Casalicchio G (2018). "iml: An R package for Interpretable Machine Learning." *JOSS*, **3**(26), 786.

https://joss.theoj.org/papers/10.21105/joss.00786. R package version 0.10.1
https://CRAN.R-project.org/package=iml

● jsonlite, Jeroen Ooms (2014). The jsonlite Package: A Practical and Consistent
Mapping Between JSON Data and R Objects. arXiv:1403.2805 [stat.CO] URL
https://arxiv.org/abs/1403.2805.

● lubridate, Garrett Grolemund, Hadley Wickham (2011). Dates and Times Made
Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL
https://www.jstatsoft.org/v40/i03/.

● maps, Original S code by Richard A. Becker, Allan R. Wilks. R version by Ray
Brownrigg. Enhancements by Thomas P Minka and Alex Deckmyn. (2018). maps:
Draw Geographical Maps. R package version 3.3.0. https://CRAN.R-
project.org/package=maps

● pdp, Brandon M. Greenwell (2017). pdp: An R Package for Constructing Partial
Dependence Plots. The R Journal, 9(1), 421--436. URL https://journal.r-
project.org/archive/2017/RJ-2017-016/index.html

● randomForest, Leo Breiman, Adele Cutler, Andy Liaw , Matthew Wiener (2018).
 randomForest: Breiman and Cutler's Random Forests for Classification and
Regression. R package version 4.6-14. https://CRAN.R-
project.org/package=randomForest

● ranger, Marvin N. Wright, Andreas Ziegler (2017). ranger: A Fast Implementation
of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical
Software, 77(1), 1-17. doi:10.18637/jss.v077.i01

● RColorBrewer, Erich Neuwirth (2014). RColorBrewer: ColorBrewer Palettes. R
package version 1.1-2. https://CRAN.R-project.org/package=RColorBrewer

● readr, Hadley Wickham and Jim Hester (2020). readr: Read Rectangular Text Data.
R package version 1.4.0. https://CRAN.R-project.org/package=readr

● reshape2, Hadley Wickham (2007). Reshaping Data with the reshape Package.
Journal of Statistical Software, 21(12), 1-20. URL http://www.jstatsoft.org/v21/i12/

● rnoaa, Scott Chamberlain (2021). rnoaa: 'NOAA' Weather Data from R. R package
version 1.3.0. https://CRAN.R-project.org/package=rnoaa

● rvest, Hadley Wickham (2020). rvest: Easily Harvest (Scrape) Web Pages. R
package version 0.3.6. https://CRAN.R-project.org/package=rvest

● shiny, Winston Chang, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke,
Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert and Barbara Borges
(2021). shiny: Web Application Framework for R. R package version 1.6.0.
https://CRAN.R-project.org/package=shiny

● shinyBS, Eric Bailey (2015). shinyBS: Twitter Bootstrap Components for Shiny. R
package version 0.61. https://CRAN.R-project.org/package=shinyBS

● shinydashboard, Winston Chang and Barbara Borges Ribeiro (2018).
shinydashboard: Create Dashboards with 'Shiny'. R package version 0.7.1.
https://CRAN.R-project.org/package=shinydashboard

- sjlabelled, Lüdecke D (2020). _sjlabelled: Labelled Data Utility Functions (Version 1.1.7)_. doi: 10.5281/zenodo.1249215 (URL: https://doi.org/10.5281/zenodo.1249215), <URL: https://CRAN.R-project.org/package=sjlabelled>
- Stats, R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/
- stringr, Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. https://CRAN.R-project.org/package=stringr
- tidyverse, Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
- zoo, Achim Zeileis and Gabor Grothendieck (2005). zoo: S3 Infrastructure for Regular and Irregular Time Series. Journal of Statistical Software, 14(6), 1-27. doi:10.18637/jss.v014.i06