COVID-19 European Regional Tracker

Asjad Naqvi[1,2,*]

[1]Research Scholar, Advanced Systems Analysis, International Institute for Applied Systems Analysis (IIASA), Laxenburg, 2361, Austria
[2]Assistant Professor, Department of Socioeconomics, Vienna University of Economics and Business (WU), Vienna, 1020, Austria
[*]Corresponding author: Asjad Naqvi (naqvi@iiasa.ac.at)

ABSTRACT

This Tracker presents data on daily COVID-19 cases at the sub-national level for 26 European countries from January 2020 till present. Country-level data sources are identified and processed to form a homogenized panel at the NUTS 3 or NUTS 2 level, the two lowest standardized administrative units of Europe. The strengths and weaknesses of each country dataset are discussed in detail. The raw data, spatial layers, the code, and the final homogenized files are provided in an online repository for replication. The data highlights the spatial distribution of cases both within and across countries that can be utilized for a disaggregated analysis on the impacts of the pandemic. The Tracker is updated monthly to expand its coverage.

Background & Summary

The COVID-19 European Regional Tracker (henceforth referred to as the Tracker)[1] collates sub-national information for cumulative and daily reported COVID-19 cases for 26 countries in Europe starting from 15th January 2020 till present. Data sources of each country are discussed in detail, including their strengths and weaknesses, and the raw country-level files are provided in an online repository. Additional effort has been put into homogenizing this data at the NUTS 3 level for all countries in the Tracker. For two countries, Poland and Greece, the complete data is only available at the NUTS 2 level. NUTS stands for Nomenclature of Territorial Units for Statistics and represent standardized administrative units defined by the European Commission for reporting various regional statistics on Europe.[2] The reason for creating a homogenized dataset at the NUTS level is to allow the Tracker to be easily merged with other datasets for analysis. For example, most of the economic, demographic, health and other indicators available on Eurostat[3], the official statistical agency of the European Co mission, is also standardized to NUTS regions. Furthermore, several global datasets that have emerged in the past year, like Google mobility trends[4] or the Facebook Social Connectivity Index[5], also structure their regional data on NUTS regions for European countries.

COVID-19 cases exploded in Europe around early March 2020. At the center of this spread were the regions of North Italy[6] and the ski resort of Ishgl in the western part of Austria.[7] From this point onward, the virus quickly spread across the European continent, resulting in a rapid increase in cases and deaths, and alarming governments that implemented stringent lockdown measures including border closures, reductions in mobility, to shutting down the economy.[8] While the virus was mostly contained during the summer of 2020, a resurgence in cases in the Fall of 2020 resulted in a massive second wave that completely overshadowed the first wave in terms of cases and deaths. In order to contain the virus, a second round of lockdown measures were put in place around October 2020.[9] Currently, in early 2021, the detection of more infectious virus variants, and a slower-than-expected vaccine roll-out in Europe has prompted countries to further extend their lockdown measures. As Figure 1 shows, no other continent has borne the brunt of the virus as Europe which still has a major share in global cases and deaths.

A unique feature of the COVID-19 pandemic is the amount of knowledge and data that is constantly being generated to understand how this event unfolds. For a high-income region like Europe, the quality of information that is generated on a daily basis is exceptionally high for most countries. Furthermore, several innovative datasets have appeared since the start of the pandemic that provide unique information. For example, the Oxford COVID-19 Government Response Tracker[10] and the Complexity Science Hub (CSH) Tracker[11] evaluate daily policy changes for a host of indicators for all countries of Europe with their coverage also extending to the rest of the world. Our World in Data (OWID), a website that curates various global data sets, has produced a tracker on COVID-19 tests performed[12] and is currently leading the efforts to document the vaccine roll-out. Google has released information how mobility is evolving over time[4] and Facebook has released data on how connected regions are with each other.[5] Various other data sources can be viewed on the Oxford COVID-19 Super-Tracker website (https://supertracker.spi.ox.ac.uk/policy-trackers/) that has catalogued over a hundred new and innovative data projects.[13]

In Europe, almost all countries provide information through interactive dashboards, maps, and data visualizations. Before October 2020, COVID-19 information for European countries was collected daily by the European Centre for Disease Prevention and Control (ECDC), and NUTS 2 level maps were regularly released to track regional trends.[14] In November 2020, ECDC decided to stop daily updates and switched to a bi-weekly reporting interval of raw data (https://www.ecdc.europa.eu/en/cases-2019-ncov-eueea). Since ECDC was the official source of COVID-19 related information for Europe, the reduction in the frequency of reporting resulted in a major data gap for tracking how the virus is evolving in the continent. As a result data aggreagator websites like Our World in Data (OWID) switched to other sources to maintain a daily reporting frequency (https://ourworldindata.org/covid-data-switch-jhu). On a positive side, almost all countries in Europe increased their efforts to display and share regional data at a daily frequency on various online platforms.

The aim of this Tracker is to identify, collect, and collate various official regional dataset for European countries. This tracker, while providing raw regional-level data, also combines and homogenizes the data at the NUTS 3 or NUTS 2 level. This homogenized dataset allows us to explore how the virus spreads in terms of cumulative cases, daily cases, and cases per capita in Europe at a daily resolution. In this Tracker, country-level data sources and their strengths and weaknesses are discussed in detail. Country-wise regional data, and the Stata code that compiles the data is released on GitHub approximately every four weeks for public use.[1] Raw and homogenized data files are also provided in a common CSV format which allows users to import the data or replicate the code in other software languages.

Methods

Figure 2 shows the workflow for the Tracker. In the first step, each country's dataset source is identified together with its spatial unit of analysis. The source can either be official or scraped data depending on how open the countries are about sharing data especially in a machine-readable format. These raw files are saved to allow users access to the original set of information. For this Tracker data on cases is extracted from all the raw files since this variable is the lowest common denominator that exists for all the countries. The raw data also contains additional variables like, deaths, tests performed, hospitalization rates, and breakdowns by age groups and gender. These additional variable can be easily extracted from the raw data as well. In the second step, the raw files are homogenized to NUTS 3 2016 boundaries either using official correspondence table or through a crosswalk extracted from spatially merging the administrative boundaries with NUTS 2016 layers. If NUTS 3 is not available then NUTS 2 boundaries are used, for example in the case of Greece and Poland.

Countries in Europe define regions differently, and therefore, making data homogeneous is a challenging task. For consistency, Eurostat, the official agency of the European Union (EU), uses homogeneous units called Nomenclature of Territorial Units for Statistics or NUTS (https://ec.europa.eu/eurostat/web/nuts/history).[2] NUTS 0 represent countries, NUTS 1 are provinces, NUTS 2 are broadly districts, and NUTS 3 are broadly defined as municipalities or other sub-divisions of districts. Each country independently defines its own administrative units that are mapped onto NUTS regions. Regions below NUTS 3, are referred to as Local Administrative Units (LAUs) that were formerly NUTS 4 and NUTS 5 tiers. The documentation of LAUs can be found here: https://ec.europa.eu/eurostat/web/nuts/local-administrative-units.

Table 1 summarizes the regional classifications of countries currently in the Tracker. The table shows the mapping of country-level regions together with the number of administrative units within that regional classification in brackets. The administrative unit at which the data is available is highlighted in bold. For most countries this is at either NUTS 3 level or lower. Two countries, Poland and Greece, are mapped at the NUTS 2 level since data is only available at this resolution for the whole duration of the tracker. United Kingdom (UK) is dealt with as four separate countries: England, North Ireland, Scotland, and Wales. This is because each country has it's own COVID-19 dashboard and the centralized COVID-19 database for the UK has put restrictions on bulk data access (see https://coronavirus.data.gov.uk/details/download). An additional challenge in creating this Tracker was to navigate the different websites of individual countries, most of which are in their native languages. As a result, there was a significant time investment in translating the websites, and identifying the correct files and the variables.

On the data side, two main challenges exist with the regions-to-NUTS mapping exercise. First, NUTS are re-classified every a few years (2003, 2006, 2010, 2013, 2016, 2021) due to demographic changes, boundary shifts, and splits in regions. Since the epidemic started in 2020, 2016 definitions were in place and therefore the tracker also homogenizes the data to NUTS 2016 boundaries. NUTS 2016 are also the definitions currently used by Eurostat for regional data. The boundary data is provided by Eurostat's GISCO (the Geographic Information System of the COmmission) (https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts).[15] Since 1 Jan 2021, the NUTS 2021 definitions have come into effect, and countries might switch to reporting on the new boundaries depending on how long the pandemic lasts. While most of the regions remain unchanged, minor shifts in boundaries can result in imperfect matching for some regions. A good example of this is Italy which already started reporting data at the NUTS 2021 definitions in 2020. These bottom-up aggregation errors are highlighted and discussed in the Technical Validation section

together with the extent of the errors. Regardless of the minor matching issues in the homogenized dataset, raw data is also available in case the users prefer to use the original source of information.

Second, some countries use different types of administrative divisions for regional tracking of COVID-19 cases that do not have an official correspondence to NUTS classifications. For example, Finland uses Hospital Districts, Greece uses Prefectures, Norway uses Kommunes, and the UK reports at the Local Authority Districts (LADs). This issue is resolved by overlaying the different administrative regions with NUTS boundaries to generate a spatial crosswalk or a region-to-NUTS mapping. For Greece, Norway, and the UK the regions map perfectly to NUTS boundaries while for Finland, small errors persist in some regions. These are also highlighted and discussed in the Technical Validation section.

Data Records

The data for the tracker has been made publicly available on Zenodo under the Creative Commons Attribution 4.0 International Licence (CC-BY). The latest version of the repository can be accessed using the Zenodo DOI: https://doi.org/10.5281/zenodo.4244878.[1] The Tracker data shown in this paper shows the evolution of cases in European regions from 15 January 2020 till 9 February 2021. Since the COVID-19 virus started spreading in Europe around March 2020, the panel of the country-regions-date is fairly complete from April 2020 onward. Figure 3 summarizes the exact date range for each country. The GitHub repository, https://github.com/asjadnaqvi/COVID19-European-Regional-Tracker, is updated once a month to expand its temporal coverage, and fix errors and bugs if any.

For each country two sets of data files are provided. The first set contains the raw data files downloaded from various sources. These files include the records as they exist in the original data including the information on the spatial unit at which the information is released. Most countries provided data at a administrative units below NUTS 3 (see Table 1). Furthermore, most of the raw files contain more information than in the final data set homogenized at the NUTS level which contains cumulative and daily cases information, the baseline variables that exist in all country files. Additional variables in the raw data for example, include deaths, recovered, tested, hospitalized, and vaccinated. Some countries also provide age and gender-wise breakdowns. Therefore, users of this Tracker can go deeper with their analysis by compiling a finer resolution and more detailed dataset for analysis. The second set are the processed country files that map and convert the raw data into homogenized datasets at the NUTS level. This mapping is done using Stata scripts or dofiles, which also identify the region-to-NUTS mapping. If Stata is not available, the dofiles can be viewed with any generic text editor. The code structure is fairly straightforward to read and can be easily converted to other programming languages. The files homogenized at the NUTS level are ready for analysis and can be merged with other NUTS-level datasets available on Eurostat[3] for example.

The files on GitHub and Zenodo (https://doi.org/10.5281/zenodo.4244878) are sorted in the following folder structure:

```
/root/
├── 01_raw
│   ├── <country folders>
│   ├── LAU
│   ├── Eurostat
│   └── OWID
├── 02_dofiles
│   ├── <country>_v*.do
│   ├── COUNTRY_SETUP.do
│   ├── COUNTRY_MERGE.do
│   ├── COUNTRY_GIS_setup.do
│   ├── COUNTRY_GIS_map.do
│   └── COUNTRY_VALIDATION.do
├── 03_GIS
│   └── <various raw and processed shapefiles>
└── 04_master
    ├── <country>_data.dta
    ├── <country>_data_original.dta
    ├── EUROPE_COVID19_master.dta
    ├── csv_nuts
    │   ├── <country>_data.csv
    │   └── EUROPE_COVID19_master.csv
    └── csv_original
```

```
    └─ <country>_data_original.csv
└─ 05_figures
   ├─ COVID19_EUROPE_cases.png
   ├─ COVID19_EUROPE_casespop.png
   ├─ COVID19_<countrycode>.png
   └─ <other graphs>
```

The folders are described as follows:

- **01_raw** contains miscellaneous country-level files. Each country has its own sub-folder with all the files necessary to generate a clean version of the raw and homogenized data. These folders also include various files that help map region identifiers to NUTS classifications. The raw data itself is saved in a Stata .dta format and the generic .csv file format in the 04_master folder. The LAU folder contains the LAU 2019-to-NUTS 2016 correspondence file[2] from which several country-level files are extracted where necessary. The Eurostat folder contains the cleaned regional population file, downloaded from https://ec.europa.eu/eurostat/databrowser/view/demo_r_pjangrp3/. OWID contains data from the Our World in Data[12] GitHub repository, which is used for validating the Tracker.
- **02_dofiles** contains the Stata scripts called dofiles for each country, and five additional dofiles that compile, merge, map, and validate the data (see the Code Availability section for details). The dofiles are also version controlled (_v1, _v2, _v3 etc.) to track changes in data sources and data structures. Only the latest dofile for each country is uploaded to the repository. Each country dofile saves the raw data, and processes the raw data to created the homogenized file, both of which are saved in the 04_master folder. See the Code Availability section for details on how to run these files.
- **03_GIS** contains the raw and processed GIS files. The raw NUTS 0 to NUTS 3 2016 shapefiles are downloaded from the Eurostat GISCO website https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts.[15] and processed using the COUNTRY_GIS_setup.do in the **02_dofiles** folder.
- **04_master** contains the raw and the homogenized country files, and the final dataset **EUROPE_COVID19_master.dta**. The .csv versions of the the raw and homogenized files are saved in the csv_original and csv_nuts folders respectively.
- **05_figures** contains the maps and figures generated from various dofiles.

The main data file "**COVID19_master.dta**" and it's .csv version are given in the 04_master folder. The master files contain the following variables with their description given in brackets:

```
nuts0_id              (country code)
nuts2_id              (for Greece and Poland only)
nuts3_id
nuts_id               (combined NUTS 2 and NUTS 3)
date
cases                 (cumulative cases)
population            (extracted from Eurostat population file)
cases_daily           (daily new cases)
cases_daily_pop       (daily new cases per 10,000 population)
```

Table 2 summarizes the individual country sources of raw data. For the 26 countries currently in the Tracker, the table lists their respective departments that collect and disseminate COVID-19 data, links to official COVID-19 dashboards, and links to data repositories that are used to pull the data for this Tracker. The very precise paths are given in the Stata dofile of each country and if Stata is not available, the they can be viewed in any text editor. Each country's raw data files are also saved in the 04_master folder with the suffix _original.dta and _original.csv. Please note that these links are also subject to change and for the latest information check the information on Zenodo or GitHub.[1]

Data for individual countries

Each country is briefly discussed below and data challenges are highlighted:

- **Austria**: Austria currently provides daily updates at the district (Bezirk) level, a tier below NUTS 3. The data can be downloaded directly from the official website as a zipped file from which the relevant file is extracted and processed.

- **Belgium**: Belgium provides daily updates and the data file can be read directly from the website. Regions with cases 5 or below are labeled as $\leq$5 for privacy reasons. As a judgement call, to keep the variable numeric, these values have been replaced with 1. Users can refer to the original data or change the dofiles if a different way of dealing with ranges is preferred.

- **Croatia**: Croatia provides official data at the NUTS 3 level in a JSON format which is processed directly in Stata.

- **Czechia**: Czechia data is imported directly from the official website for processing.

- **Denmark**: Denmark releases a zipped file daily which is imported and processed directly in Stata. Data is split across several files from which only the one containing information on cases is used.

- **Estonia**: Estonia officially releases the data on COVID-19 but in order to maintain privacy, data is only provided in ranges of 10s, for example, 10-20, 20-30, 100-110, etc. In order to make this variable numeric, the mid-points of these ranges are taken, and aggregated at the NUTS 3 level. Using mid-points is a judgement call to make the cases variable numeric. The raw data file contains the original information which can be processed differently if needed.

- **Finland**: The data for Finland is taken from a GitHub repository that scrapes the data and provides it in a JSON format. This is read and parsed in Stata. This data is at the hospital district level which does not perfectly correspond to NUTS 3 boundaries. Therefore errors persist in the final data file. See Technical validation section for detailed notes.

- **France**: France provides a comprehensive range of COVID-19 related indicators on its official website. France also switched methodologies for improving the data quality in May 2020. Old data is available online and has been merged with the current data but the quality is poor. Data of territories outside of mainland Europe have been dropped. The file is downloaded manually from the website for processing.

- **Germany**: Germany's data, provided by the Robert Koch Institute (RKI), has been of really high quality data since the early days of the pandemic. Several repositories of this data exist on GitHub, of which, one has been selected for the Tracker.

- **Greece**: Greece only releases PDFs which contain information at the prefecture level. This information is downloaded from a GitHub repository that scrapes the PDFs. The data which is at the prefecture level, matches perfectly to NUTS 2 boundaries. Autonomous islands of Greece are dropped from the homogenized dataset but they exist in the raw file.

- **Hungary**: Hungary does not officially release any data. A map is uploaded in an image format on the official website https://koronavirus.gov.hu/, which is scrapped daily by independent users. This information is retrieved from a GitHub repository.

- **Ireland**: Ireland releases regional data on ArcGIS Hub which is downloaded manually for processing. The links are provided in the Stata script.

- **Italy**: Italy has one of the best data-sharing setup among the European countries. Italy currently provides full documentation and access to their data on their official GitHub page https://github.com/pcm-dpc/COVID-19. The regions defined in the data perfectly match NUTS 2021 definitions but the aggregation to NUTS 2016 has to be approximated for three small island regions since their boundaries were modified. See Technical Validation sector for details.

- **Latvia**: Latvia updates the data daily which can be downloaded directly from the official website. Regions with cases under 5 are displayed with a range of 1-5 for privacy reasons. This has been replaced with a value of one to keep the variable numeric. The raw file containing the original information can used to process these ranges differently.

- **Netherlands**: Netherlands data is downloaded directly from the official website for processing.

- **Norway**: Since Norway is outside of the EU, no consistent information at the NUTS 3 level is available. The Norwegian data, which is provided at the Kommune level which is one tier below NUTS 3. This data is spatially joined with the NUTS 2016 boundaries to create a Kommune-to-NUTS 3 crosswalk which results in a perfect mapping. See section on Technical Validation for details.

- **Poland**: Poland data is extracted from a GitHub repository that extracts NUTS 2 level data. Poland launched its data sharing service relatively recently which currently also provides NUTS 3 level, but it does not have the temporal extent to fit in this Tracker.

- **Portugal**: The official regional data for Portugal is off poor quality. Data on regions was released on a daily frequency till July 2020, and then it was switched to a weekly frequency. Currently information is only available on a bi-weekly basis. Out of all the countries in the Tracker, Portugal's data is the least usable but is kept in the dataset since the information is provided at the NUTS 3 level.

- **Romania**: The official data for Romania is available in JSON format which is converted into a CSV format for processing in Stata. The links are provided in the Stata script for Romania.

- **Slovak Republic**: The data for Slovak Republic is downloaded from a GitHub repository which processes and cleans the official data.

- **Slovenia**: The data for Slovenia is downloaded from a GitHub repository which processes and cleans the official data.

- **Spain**: Spain's data is downloaded directly from the official website.

- **Sweden**: Sweden releases a file daily which is manually downloaded. The link is provided in Table 3 and in the data scripts.

- **Switzerland**: Switzerland does not have country-wide information available in a centralized place. Instead, a group of independent researchers collate this information from the official websites of the various Cantons, which are also NUTS 3 regions. Therefore the very latest data is not immediately available.

- **United Kingdom**: The data for UK is not easily accessible, neither it is well documented. Despite shifting to a new website in the summer of 2020, information is only released for 7-day intervals for regions below NUTS 3 in the form of an interactive map, `https://coronavirus.data.gov.uk/details/interactive-map`. Furthermore, the data portal for the UK `https://coronavirus.data.gov.uk/details/download` has put significant restrictions on bulk downloading. Therefore, UK is dealt with as four separate countries; England, North Ireland, Scotland, and Wales.

  - **England**: In the early days of the pandemic, Tom White put an enormous amount of effort to collate UK information from various sources. He efforts were later picked up by ODI Leeds (`https://github.com/odileeds/covid-19-uk-datasets`) that now update the dataset for England. The data is available at the Local Authority Districts (LAD) using April 2019 definitions which can be found here. LADs aggregate up perfectly to NUTS 3 2016 boundaries. More recent mapping of LADs-to-NUTS 3 can also be done but this would require a significant effort to map the old identifiers to the new ones. See the section Technical Validation for details.

  - **Scotland**: Data for Scotland is available on their official website which is also processed by ODI Leeds (`https://github.com/odileeds/covid-19-uk-datasets`). Scotland's data is processed following the same routine as England where LAD April 2019 boundaries are mapped on to NUTS 3 using a spatial merge. See the section Technical Validation for details.

  - **North Ireland**: Data for North Ireland has not been found at a consistent regional-day resolution in a machine-readable format.

  - **Wales**: Data for Wales has not been found at a consistent regional-day resolution in a machine-readable format.

The following countries – Albania, Bosnia, Bulgaria, Serbia, Lithuania, North Ireland (UK), North Macedonia, Turkey, and Wales (UK), have official NUTS 3 correspondence tables but are not currently in the Tracker since their regional data has not been located. If these countries have to be included with the current data version, they can be replaced with country-level indicators to complete the map. For analysis this is still useful as cases normalized by population allow for comparison of regions of various sizes. As a note of caution, the above information and descriptions are subject to change as countries evolve their COVID-19 data sharing strategies. Please check the GitHub page for the latest updates.

Overall data trends

Figure 3 shows the data points available for each country. Here one can note major gaps in Portugal, Greece, and Estonia daily cases data but on average the remaining countries are fairly complete. For the data used there are over 402,000 data points. Individual NUTS-level cases per 10,000 population are plotted in Figure 4. In this figure we can also observe the difference in normalized daily new cases between the first and the second waves. The spread goes up significantly in the second wave.

Figure 6 shows the cumulative distribution of cases and cases per 10,000 population for the complete date ranging from 15 January 2020 till 9 February 2021. Here we can immediately observe that larger units in terms of area have more cases overall than smaller ones. If we control for population size then a different picture emerges. For example, it is clear from the cumulative cases per 10,000 population map that Germany managed to insulate itself very well from neighboring countries by enforcing strict border controls. Despite this, regions on the east side show a higher incidence rate than the west and north Germany. Similarly the western part of Austria, north Italy, Switzerland, and eastern part of France have a much higher level of cases per capita relative to the remaining regions in these countries. Sweden with its lack of strict lockdown policies also stands out among the Nordic countries. Explanations for these trends are left as research questions. Furthermore, given the daily resolution of the data in the Tracker, these maps can also be checked for variations in the first and the second waves.

Technical Validation

Three steps have been taken to ensure that veracity and accuracy of the information and to document errors that might occur in the homogenized file.

First, since the country-level datasets are official records provided by different government departments of each country, they can be compared with various online dashboards highlighted in Table 2. It is also important to point out that not all countries release the latest data at the regional level. A good example of this is France, that releases a file daily but the regional information is usually two to three days old. Similarly the crowd-sourced data for Switzerland, is back-filled data as information for the different regions (Cantons) is updated. Regardless of these lags, comparison of values with online dashboards is possible for most countries in the Tracker.

Second, as discussed earlier, regional data is approximated during the homogenization process for some countries. In order to ensure transparency, Table 3 provides notes for various countries where either data is converted from ranges into unique values (for example, Belgium, Estonia, and Latvia) or boundaries are approximated using a spatial merge (for example, Finland, UK, Norway, and Greece). The *data approximation* error occurs when ranges are converted into unique numerical values. This is a judgement call in order to allow these countries to merge with the remaining files. For Belgium and Latvia, regions with cases less than five are anonymized as <5. These values have been replaced with 1. For Estonia, which only provides data in ranges of 10s, mid values of each range is taken and multiplied with total cases to estimate cases per region, and then the regions are aggregated to NUT 3. Therefore, the data approximation is the highest for Estonia. The raw data and the scripts are available in case users prefer the actual information or have other approximation strategies. For countries which do not have an official correspondence to NUTS regions, a *boundary approximation* is done with a spatial merge by overlaying administrative boundaries with NUTS boundaries. For UK, Norway and Greece, the spatial merge creates a perfect mapping as shown in Figure 6. For Finland, the hospital districts, that are size-wise comparable to NUTS 3 boundaries contains minor errors in some cases since boundaries do not align. In an ideal case scenario, data should be provided at the smallest possible unit that can be aggregated up to various administrative boundaries, as is the case with UK and Norway. The official Finish website does provide COVID-19 data at municipality level, a unit below NUTS 3, but accessing the map is not straightforward nor it is clear whether this data is even openly accessible. On a related note, Italy data is officially released at the NUTS 2021 definitions so mapping to NUTS 2016 definitions do not work perfectly for some regions on an island off mainland Italy. These regions show up as missing on the map. For both the Finnish and Italian regions that do not perfectly match, the number of cases are very small for it to make a significant difference. Regardless, original administrative units and their data in the raw files and can be used for more accuracy.

Third, since the data is at the regional level, it can be aggregated up to generate country-level totals which can be compared with data aggregator websites like the Our World in Data (OWID) COVID-19 tracker. OWID is used and referenced almost daily in scientific research and the media and has a major impact of policy discussions. OWID was utilizing country-level information provided by the European Center for Disease Control (ECDC) till November 2020. In November 2020, ECDC announced that it only do bi-weekly data releases. As a response OWID switched to the John Hopkins University's (JHU) data repository, a major data source for COVID-19 information at the global level.[1] For validation, both this Tracker and OWID data is merged on a country-date combination and the difference between the country-level daily cases is calculated. Figure 7 plots the difference split by countries. Figure 7 shows a how good the match is before 1 October 2020. After October 2020, the mismatch for most countries increases significantly and persists till today. This highlights two points. First, before October 2020, data was provided by ECDC which was taking information directly from European countries. Since this Tracker is also pulling data from the countries directly, the match is exceptionally close with the exception of some outliers. This exercise helps validate the data of this Tracker. Second, since the data source of this Tracker remains the same, while OWID changed its source to a more unverified data set around October 2020, this Tracker provides a more accurate picture of country-level aggregates and also includes regional variation.

[1] See the press release here https://ourworldindata.org/covid-data-switch-jhu.

Usage Notes

The dataset has been compiled in Stata (www.stata.com)[16], a standard statistical software mostly used in the field of economics. All data including raw files, scripts, and the final data set are provided on GitHub[1]. Besides the Stata .dta data format, all information is also stored in the generic .csv format allowing the data to be accessed in any software. Annotated Stata scripts, or dofiles are also provided. Stata has an easy-to-interpret syntax structure that can be easily translated into other programming languages.

As a caveat, there are two features that Stata cannot handle well. First, is the ability to download files that are redirected from clicks or links on websites. This level of web-scraping works much better in other languages designed for such tasks like R or Python. Second data in JSON format has to be parsed either manually using user-written commands, or via third-party JSON-to-csv converters that are available online. Processing complex data structures like JSON or XML are also easier in other languages.

The repository is updated approximately every four weeks including a public release on Zenodo. If users prefer a higher frequency of updates, then the Stata scripts can be used directly to process the files or a more frequent release can be requested as well. Since the aim of the homogenized dataset is to be able to correlated it with existing NUTS level datasets, a monthly frequency is sufficient since European-level regional data takes several weeks or even months to be updated.

This Tracker can be utilized for a host of different research directions. For example, the Tracker can be mapped onto NUTS-level regional data released by Eurostat[3]. This includes weekly or monthly economic indicators, demographic changes, mortality indicators. Individual countries can also be analyzed if their statistical agencies provide regional or micro data for analysis. Other datasets catalogued on the Oxford COVID-19 Supertracker[13] provides a range of interesting information like policy stringency indices, cross-border movement restrictions, health-related indicators etc. The Tracker data can also be combined with several innovative global datasets which also contain NUTS-level information. This, for example, includes Google Mobility trends[4] or the Facebook Social Connectivity Index[5]. Since, the data for the Tracker has a Creative Commons Attribution 4.0 International Licence (CC-BY), anyone can access it at any point in time, and it will get regular updates until the countries stop publishing regional COVID-19 data.

Code Availability

The code to process the files can be downloaded with the Zenodo (https://doi.org/10.5281/zenodo.4244878) repository or accessed from the Github repository. The code is written using Stata versions 15 and 16, which are also recommended for improved functionality with maps and graphs but general data processing can be done in any version.

As specified in the Data Records section, the dofiles are in the **/02 dofiles/** folder. Within this folder dofiles exist for each country plus a set of five dofiles that setup, merge, map, and validate the final data file. These files are explained as follows:

- COUNTRY_SETUP.do initializes the code for running the country files. One can run each country file independently as well, but they need the directory structure and packages to be loaded in order to function correctly. Directory and packages can be initialized using the first few lines marked in the beginning of the COUNTRY_SETUP.do file. This syntax is as follows:

```
clear
global coviddir "<your directory path>/<your directory name>"

* package for maps and correcting map projections
ssc install spmap, replace
ssc install geo2xy, replace

* packages for color schemes
ssc install palettes, replace
ssc install colrspace, replace

* package for plots
net install cleanplots, from("https://tdmize.github.io/data/cleanplots")
set scheme cleanplots, perm
```

Each country .do file is annotated with notes where necessary.

- COUNTRY_MERGE.do combines all the country datasets saved in **04_master** in one file EUROPE_COVID19 master.dta. The master file is also saved in the **04_master** folder.

- COUNTRY_GIS_setup.do sets up the GIS layers in Stata format for the combined NUTS regions and for individual countries. A mixed NUTS3 and NUTS2 shapefile is also created accommodate the data from Poland and Greece. The logic can also be applied to add data at the provincial (NUTS 1) or country (NUTS 0) level if one needs to add other countries not in the dataset. This file also extracts shapefiles for individual countries and generates a file used for labeling the individual country maps.

- COUNTRY_GIS_map.do create the maps that are saved in the **05_figures** folder. See Figure 6 for the overall map. Individual country COVID-19 maps can be viewed on GitHub.

- COUNTRY_validation.do collapses the Tracker to a country-date level and merges it with OWID COVID-19 dataset, for validation. This file also produces Figure 7.

For the very latest version of the code, please also check GitHub. GitHub files will be converted in Zenodo repositories approximately every four weeks. Each Zenodo release is assigned a unique DOIs, but the generic DOI `https://doi.org/10.5281/zenodo.4244878` always link to the latest version.

References

1. Naqvi, A. COVID-19 European Regional Tracker, https://doi.org/10.5281/zenodo.4244878 (2021).
2. Eurostat. Nomenclature of territorial units for statistics (NUTS) (2021). https://ec.europa.eu/eurostat/web/nuts/background.
3. Eurostat. European statistical agency (2021). https://ourworldindata.org/coronavirus.
4. Google. Community mobility reports (2020). Accessed on 1 February 2021, https://www.google.com/covid19/mobility/.
5. Facebook. Data for good: Social Connected Index (2021). https://dataforgood.fb.com/tools/social-connectedness-index/.
6. ECDC. ECDC statement on the rapid increase of COVID-19 cases in italy (23 Feb 2020). https://www.ecdc.europa.eu/en/news-events/ecdc-statement-rapid-increase-covid-19-cases-italy.
7. Eurostat. Coronavirus: Ischgl resort at heart of Europe's outbreak reopens (12 Oct 2020). https://www.bbc.com/news/world-europe-52384572.
8. Deutsche Welle (DW). Coronavirus: What are the lockdown measures across europe? (14 April 2020). https://www.dw.com/en/coronavirus-what-are-the-lockdown-measures-across-europe/a-52905137.
9. Euronews. How hard is the coronavirus second wave hitting in europe? (13 November 2020). https://www.euronews.com/2020/11/13/europe-s-second-wave-of-coronavirus-here-s-what-s-happening-across-the-continent.
10. Oxford. Oxford COVID-19 Government Response Tracker (oxcgrt) (2021). https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker.
11. Desvars-Larrive, A. *et al.* A structured open dataset of government interventions in response to COVID-19. *Sci. Data* **7**, 285, 10.1038/s41597-020-00609-9 (2020).
12. Hasell, J. *et al.* A cross-country database of COVID-19 testing. *Sci. Data* **7**, 345, 10.1038/s41597-020-00688-8 (2020).
13. Oxford. Oxford Supertracker: The global directory for COVID policy trackers and surveys (2021). https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker.
14. ECDC. COVID-19 database (2020). Data retrieved from European Centre for Disease Prevention and Control, https://www.ecdc.europa.eu/en/covid-19/data.
15. GISCO. The Geographic Information System of the Commission (2021). NUTS Statistical Administrative Units Shapefiles, https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts.
16. Stata (2021). https://www.stata.com/.

Acknowledgements

I would like to thank the IIASA directors, Albert van Jaarsveld, and Leena Srivastava, and the Director of Advanced Systems Analysis (ASA) program, Elena Rovenskaya, for their encouragement and continuous support throughout the course of this Tracker. I would also like to thank IIASA for partially funding this project.

Author contributions statement

A.N. set up the Tracker, including identifying the websites and the relevant files, defining the protocols for the workflow and data management, writing the code, and cleaning up the data. A.N. will be responsible for updating the Tracker on a monthly basis till the countries stop reporting their data.

Competing interests

The author declares no competing interests.

Figures & Tables

Table 1. European regional classifications

Country (NUTS 0)	Code	NUTS 1	NUTS 2	NUTS 3	LAU / Other admin units
Austria	AT	Gruppen von Bundesländern (3)	Bundesländer (9)	Politische Bezirke (35)	**Bezirke (94)**, Gemeniden (2096)
Belgium	BE	Gewesten/ Régions (3)	Provincies/ Provinces (11)	Arrondissementen/ Arrondissements (44)	**Gemeenten/Communes (581)**
Croatia	HR	-	Regija (4)	**Županija (21)**	Gradovi i općine (556)
Czechia	CZ	Území(1)	Regiony soudržnosti (8)	**Kraje (14)**	Obce (6258)
Denmark	DK	-	Regioner (5)	Landsdele (11)	**Kommuner (99)**
Estonia	EE	-	-	Maakondade grupid (5)	**Linn, vald (79)**
Finland*	FI	Manner-Suomi, Ahvenanan-maa/ Fasta Finland, Åland (2)	Suuralueet / Storområden (5)	**Maakunnat / Landskap (19)**	Kunnat / Kommuner (311)
France	FR	Zones d'études et d'aménagement du territoire (14)	Régions (27)	**Départements (101)**	Communes (34970)
Germany	DE	Länder (3)	Regierungsbezirke (38)	**Kreise (401)**	Gemeniden (11087)
Greece	EL	Geografikes Perioches (4)	**Periferies (13)**	Periferiakon Enotiton (52)	Topikes Koinotites (6134)
Hungary	HU	Statisztikai nagyrégiók (3)	Tervezési-statisztikai régiók (8)	**Megyék + Budapest (20)**	Települések (3155)
Ireland	IE	-	Regions (3)	**Regional Authority Regions (8)**	Local Election Areas (166)
Italy	IT	Gruppi di regioni (5)	Regioni (21)	**Provincie (107)**	Comuni (7926)
Latvia	LV	-	-	Statistiskie reģioni (6)	**Republikas pilsētas, novadi (119)**
Netherlands	NL	Landsdelen (4)	Provincies (12)	NUTS3 (40)	**Gemeenten (355)**
Norway*	NO	-	Landsdeler (7)	Fylker (18)	**Kommuner (356)**
Poland	PL	Makroregiony (7)	**Regiony (17)**	Podregiony (73)	Gminy (2478)
Portugal	PT	Continente + Regiões Autónomas (3)	Grupos de Entidades Intermunicipais + Regiões Autónomas (7)	**Entidades Intermunicipais + Regiões Autónomas (25)**	Freguesias (3098)
Romania	RO	Macroregiuni (4)	Regiuni (8)	**Judet + Bucuresti (42)**	Comuni + Municipiu + Orase (3181)
Slovenia	SI	-	Kohezijske regije (2)	Statistične regije (12)	**Občine (212)**
Slovak Republic	SK	-	Oblasti (4)	**Kraje (8)**	Obce (2927)
Spain	ES	Agrupación de comunidades autónomas (7)	Comunidades y ciudades Autónomas (19)	**Provincias + islas + Ceuta, Melilla (59)**	Municipios (8131)
Sweden	SE	Grupper av riksområden (3)	Riksområden (8)	**Län (21)**	Kommuner (290)
Switzerland*	CH	-	Grossregionen (7)	**Kantone (26)**	Gemeinden/Communes (2222)
United Kingdom*	UK	Government Office Regions (12)	Counties (41)	Upper tier authorities (UTAs) (179)	Lower tier authorities (LTAs) (400)

Note: Extracted from the Eurostat NUTS and National Administrative Units Correspondence page. Number of regions are given in brackets. Countries marked with a * are not in the EU and hence are not subject to Eurostat's reporting requirements at the NUTS level. The region at which the data is available is highlighted in bold. UK is dealt with as four different countries: England, North Ireland, Scotland, Wales.

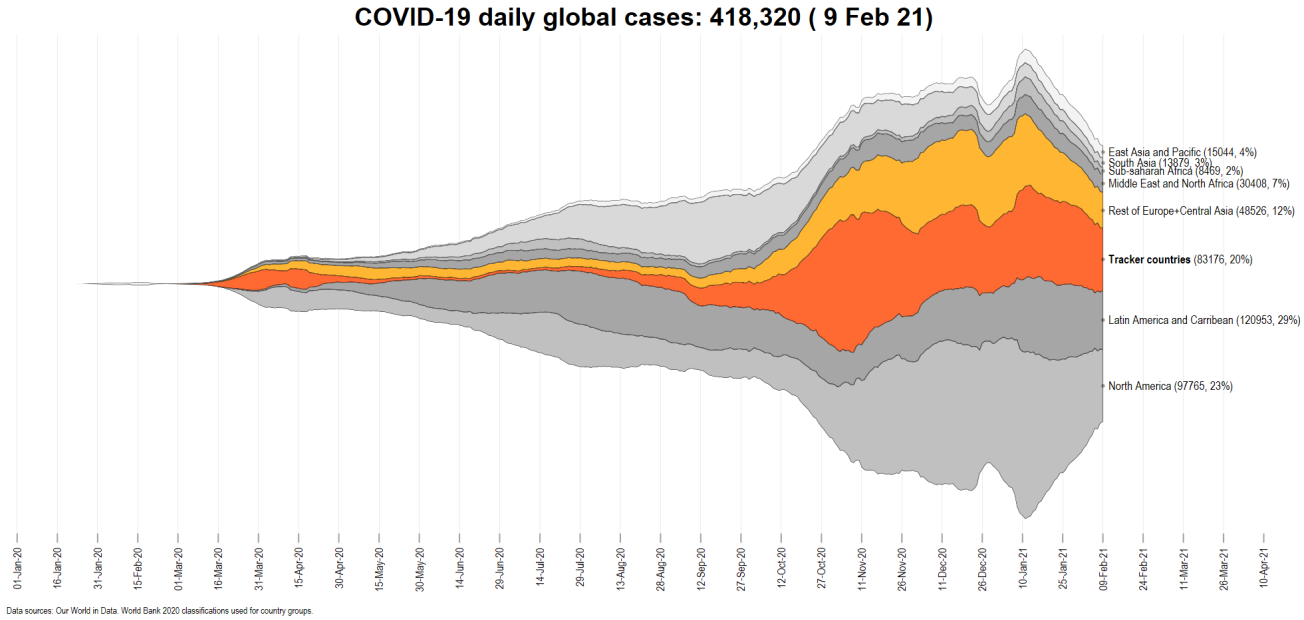
Table 2. Country-specific data sources

Country	Code	Agency responsible for the data	Official site	Data source
Austria	AT	Österreichische Agentur für Gesundheit und Ernährungssicherheit (AGES)	Link	Official website
Belgium	BE	Belgian Institute for Health (sciensano)	Link	Official website
Croatia	HR	Ministry of Health	Link	Official website
Czechia	CZ	Ministry of Health (MZCR)	Link	Official website
Denmark	DK	Statens Serum Institut (SSI)	Link	Official website
Estonia	EE	Estonian Health Board (Terviseamet)	Link	Official website
Finland	FI	Finnish Institute for Health and Welfare (THL)	Link	Official data via GitHub: HS-Datadesk
France	FR	Santé publique France	Link	Official website
Germany	DE	Robert Koch Institute (RKI)	Link	Official data via GitHub: jgehrcke
Greece	EL	National Public Health Organization (EODY)	Link	Scraped data via GitHub: Sandbird
Hungary	HU	The Government of Hungary	Link	Scraped data via GitHub: nickgon
Ireland	IE	The Department of Health	Link	Official data via ArcGIS Hub
Italy	IT	Ministero della Salute	Link	Official data via Github: pcm-dpc
Latvia	LV	The Ministry of Health	Link	Official website
Netherlands	NL	National Institute for Public Health and the Environment (RIVM)	Link	Official data via ArcGIS Hub
Norway	NO	Norwegian Institute of Public Health (NIPH)	Link	Official data via GitHub: thohan88
Poland	PL	The Ministry of Health	Link	Official data via GitHub: covid19-eu-zh
Portugal	PT	The Directorate-General for Health (DSG)	Link	Official data via GitHub: bruno-leal
Romania	RO	The National Institute of Public Health (CNSCBT)	Link	Official website
Slovak Republic	SK	Ministry of Investments, Regional Development and Informatization	Link	Scraped data via GitHub: radoondas
Slovenia	SI	Republic of Slovenia website	Link	Official data via GitHub: sledilnik
Spain	ES	National Center for Epidemiology (isciii)	Link	Official website
Sweden	SE	The Public Health Agency of Sweden	Link	Official data via ArcGIS Hub
Switzerland	CH	Bundesamt für Gesundheit (BAG)	Link	Scrapped data via GitHub: covid19-eu-zh
United Kingdom	UK	The UK Government	Link	
England		National Health Service (NHS)	Link	Scraped data via GitHub: odileeds
Scotland		The Scottish Government	Link	Scraped data via GitHub: odileeds
North Ireland		Department of Health North Ireland	Link	
Wales		The Welsh Government	Link	

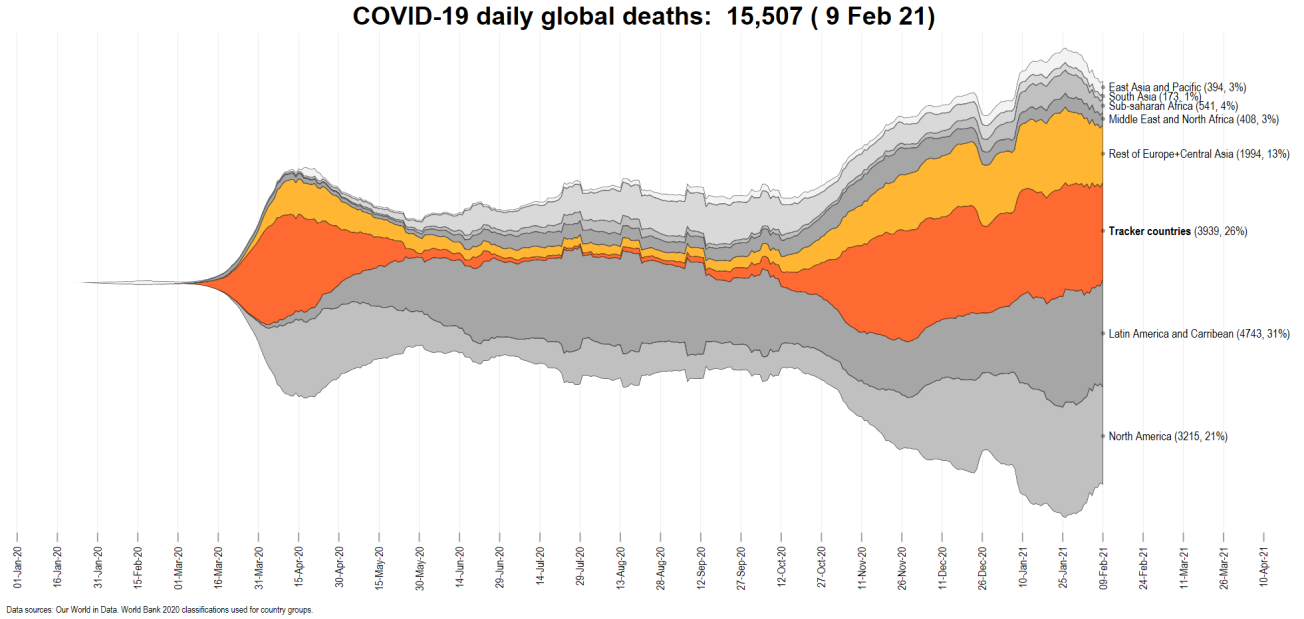
Note: Websites and URLs are subject to change, as links change or get updated, or better resources are found. Please check on the GitHub page for the latest information. UK is treated as four separate countries.

Table 3. Notes on countries with data or mapping issues

Country	Machine readable data?	NUTS 3 2016 mapping?	Additional comments
Belgium	Yes	Yes	Regions with cases less than five are provided as (<5). For the homogenized dataset this is replaced with 1 as a judgement call. Please see raw file for original data.
Estonia	Yes	Yes	Cases are only provided in ranges of 10. These are averaged to the middle value of each range and averaged over NUTS regions. Please see the raw file for actual micro data or visit the official website for notes
Finland	No	No	The data is provided at the Hospital District level. These do not correspond perfectly to NUTS 3 regions. See Figure 5b for reference and the raw file for the original data. An official spatial map of hospital districts is not readily available. The official website also shows municipality level data but this has not been found collated in any repository.
Greece	No	No	Data is available at the county level only and this corresponds perfectly to NUTS 2 2016. See Figure 5d.
Italy		No	Data perfectly merges with NUTS 3 2021 boundaries and 2016 boundaries are approximated for some regions. See Figure 5c.
Latvia	Yes	Yes	Regions with cases less than five are provided as (<5). For the homogenized dataset this is replaced with 1 as a judgement call. Please see original raw file or modify the dofile as necessary.
Poland	No	No	The data correspondence to NUTS 2 2016 definitions for all regions except Warsaw (PL90) which is manually duplicated to PL91 and PL92. The values are d. The data file switches from cumulative to daily cases in the same column after 23 November 2020. This is manually fixed in the NUTS 2 file. Data errors exist for 13th June 2020 which are manually fixed by taking the average of the previous and the next date. See raw file for original data. The official website now provides NUTS 3 level data but the range in limited.
Portugal	No	Yes	Portugal switched to releasing regional data at weekly intervals after 7th July 2020. From 26 October 2020 the reporting range was increased to 14 day intervals. Therefore daily cases data is mostly limited to pre-July 2020 values. Only the daily cases reported in the weekly or bi-weekly data are added to the homogenized tracker. Please see the raw data and the `https://github.com/dssg-pt/covid19pt-data` for detailed documentation.
Switzerland	No	Yes	Data for each Swiss canton, that match perfectly to NUTS 3 regions, is released separately and volunteer groups aggregate this information. Therefore, there are variations on the frequency of updated across the cantons. See original file for details.
United Kingdom			UK now has an official API server for data queries but there are limits to how much data can be extracted. Currently the data aggregator website Our World in Data is the only source providing LAU level information for UK, but this is only available at a 7-day rolling average normalized by population (`https://ourworldindata.org/local-covid-uk`). Bulk micro regional data has not been publicly released.
England	No	No	NUTS 3 correspondence is extracted using a spatial merge between the Local Authority Districts (LADs) using April 2019 definitions and NUTS 3 2016 boundaries. See Figure 5a for details and see raw file for the original data.
Scotland	No	No	NUTS 3 correspondence is extracted using a spatial merge between the local authority districts and NUTS 3 2016 boundaries. See map for details and see raw file for the original data.
North Ireland	No		Country skipped. See notes above for UK
Wales	No		Country skipped. See notes above for UK



(a)



(b)

Figure 1. Global distribution of COVID-19 cases and deaths. Countries in the tracker are highlighted in dark orange, while the rest of Europe is shown in light orange color.

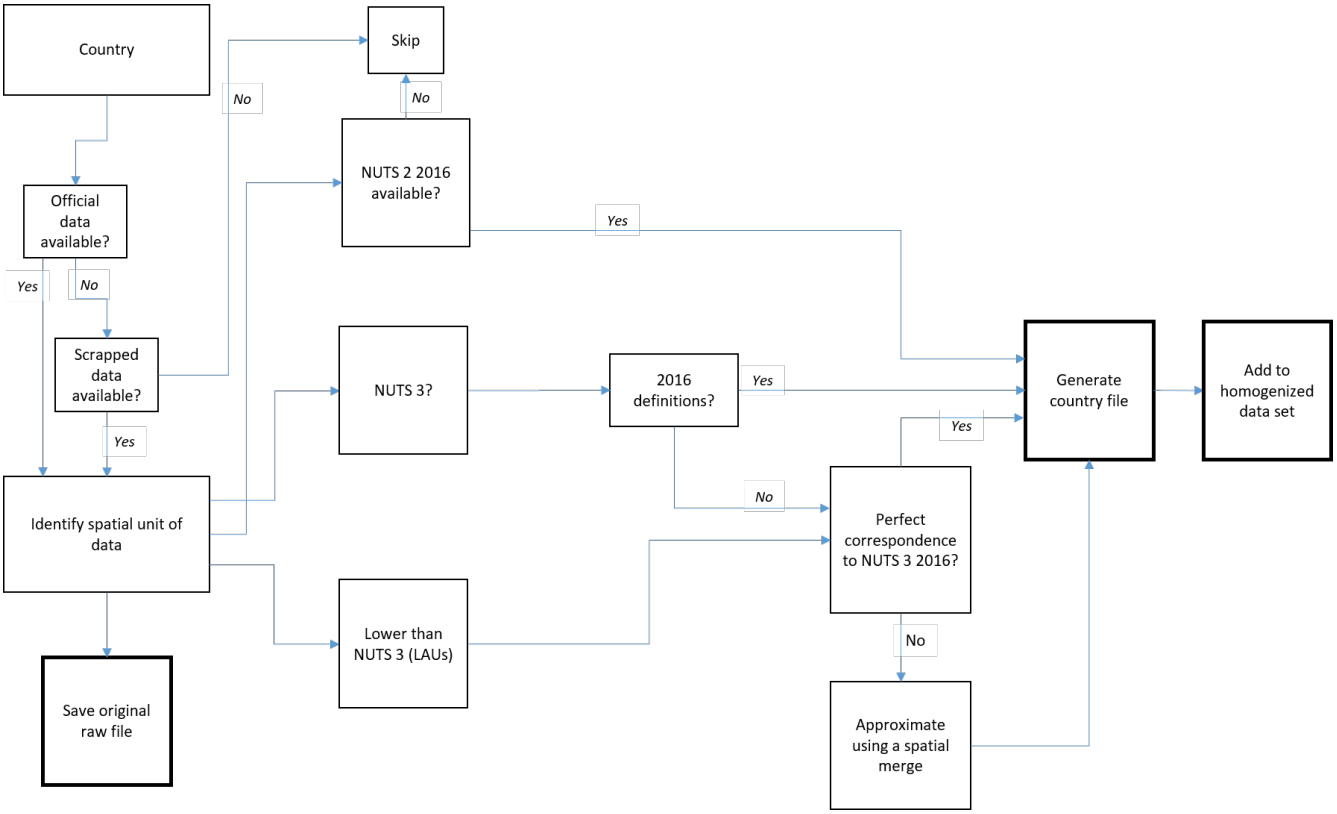


Figure 2. Workflow for the Tracker

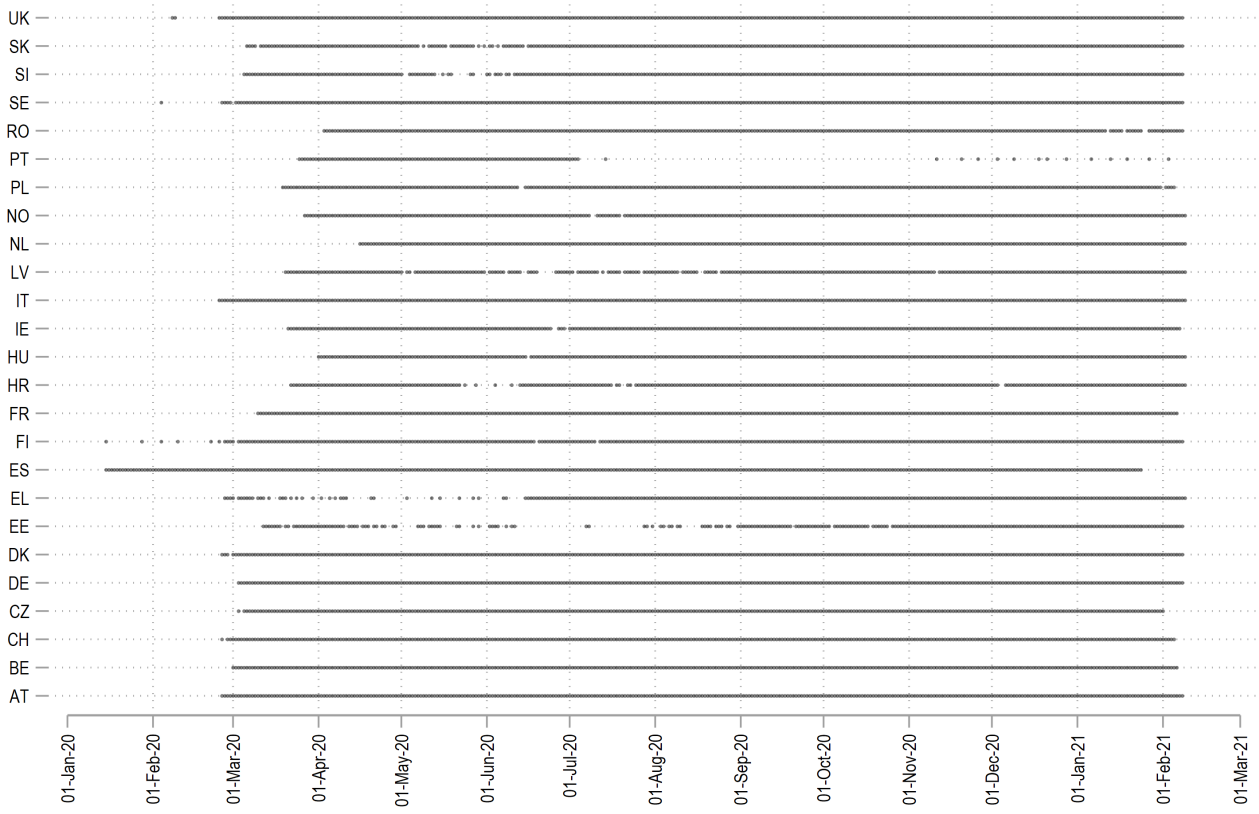


Figure 3. Data range of countries in the Tracker.

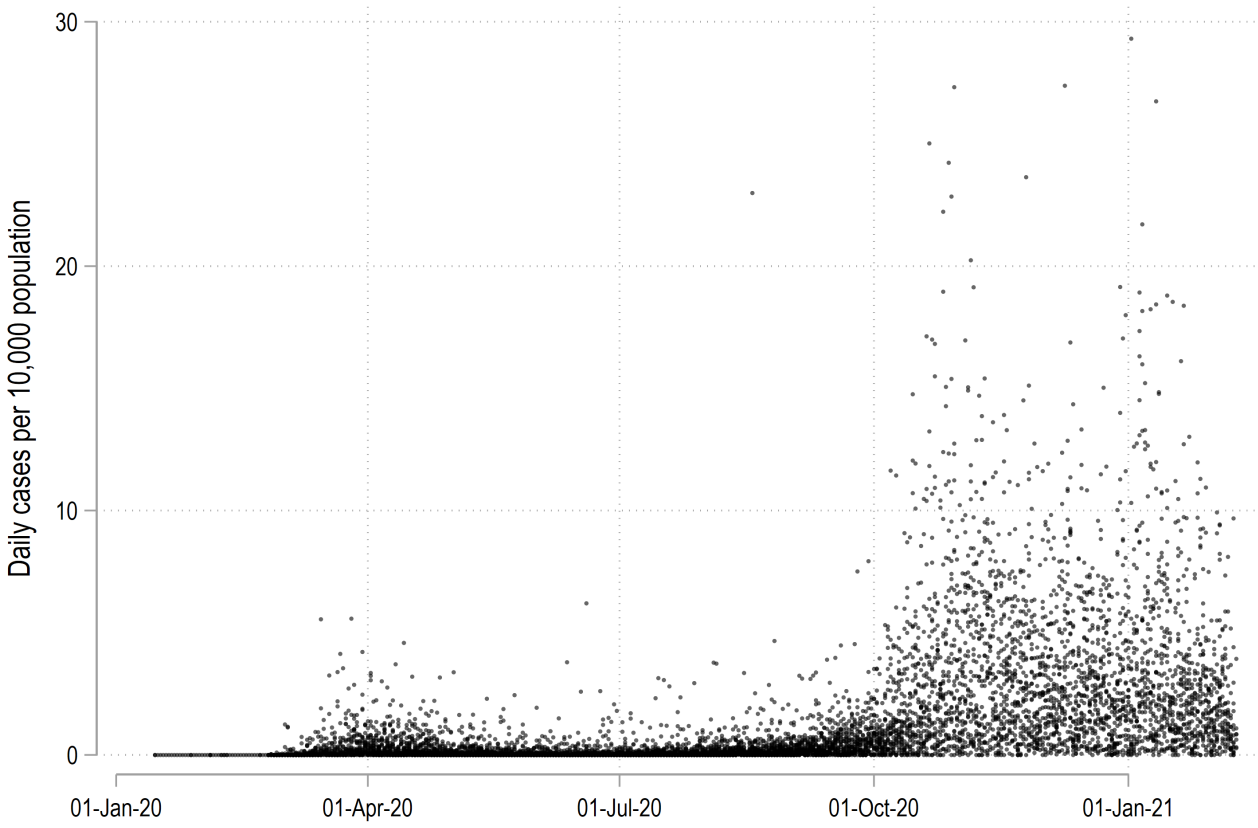
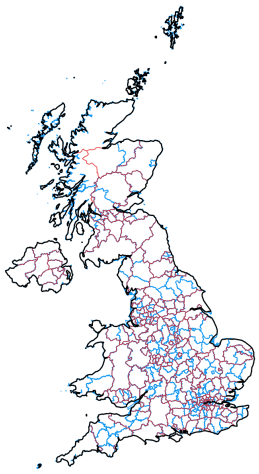
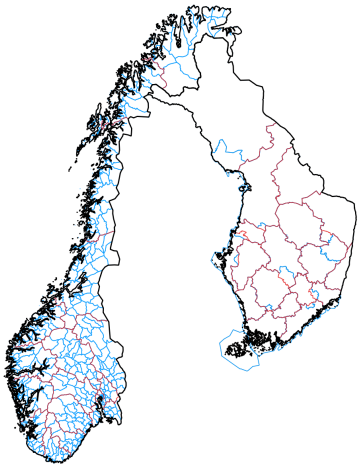


Figure 4. Distribution of daily cases per 10,000 population.



(a) UK

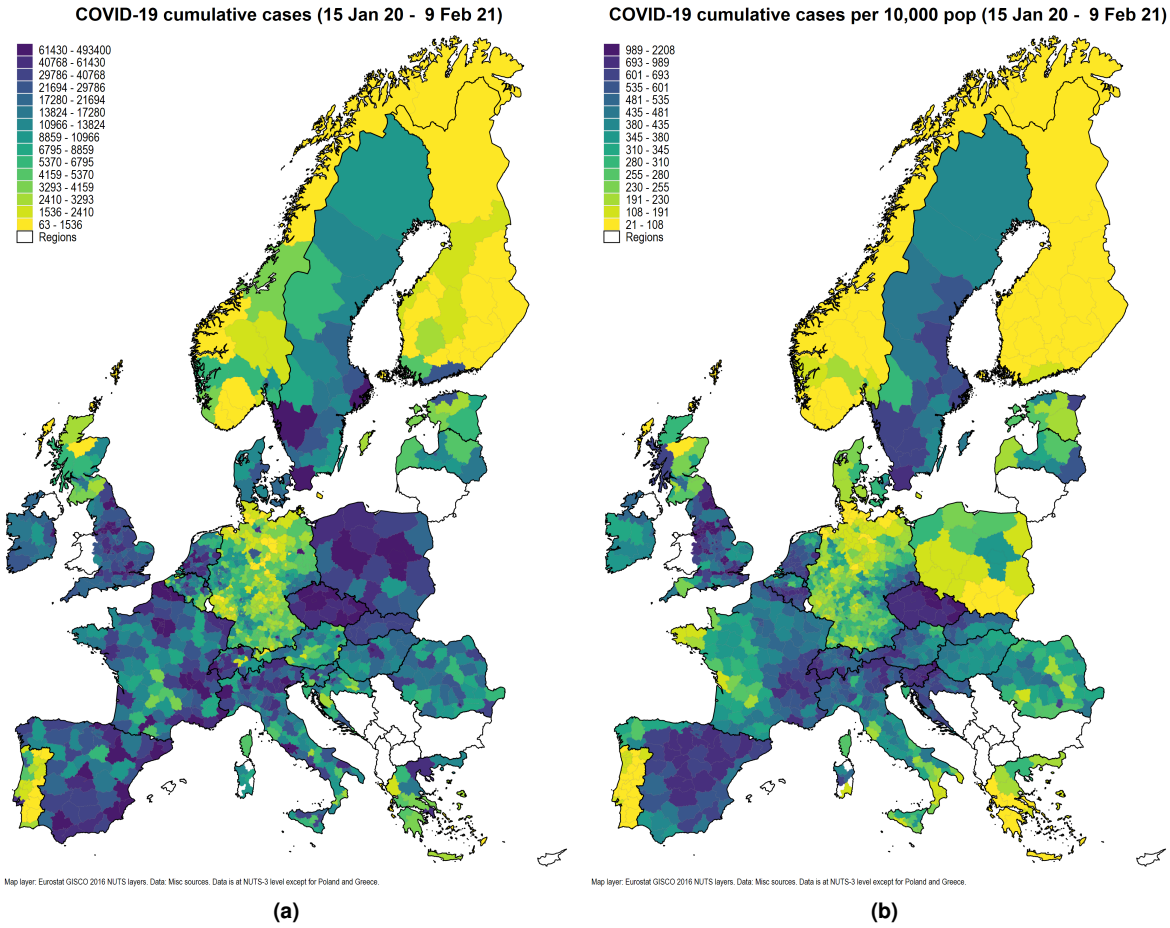


(b) Norway and Finland

(c) Italy

(d) Greece

Figure 5. Countries where region to NUTS boundaries are approximated.



(a) **(b)**

Figure 6. Spatial distribution of cumulative COVID-19 cases.

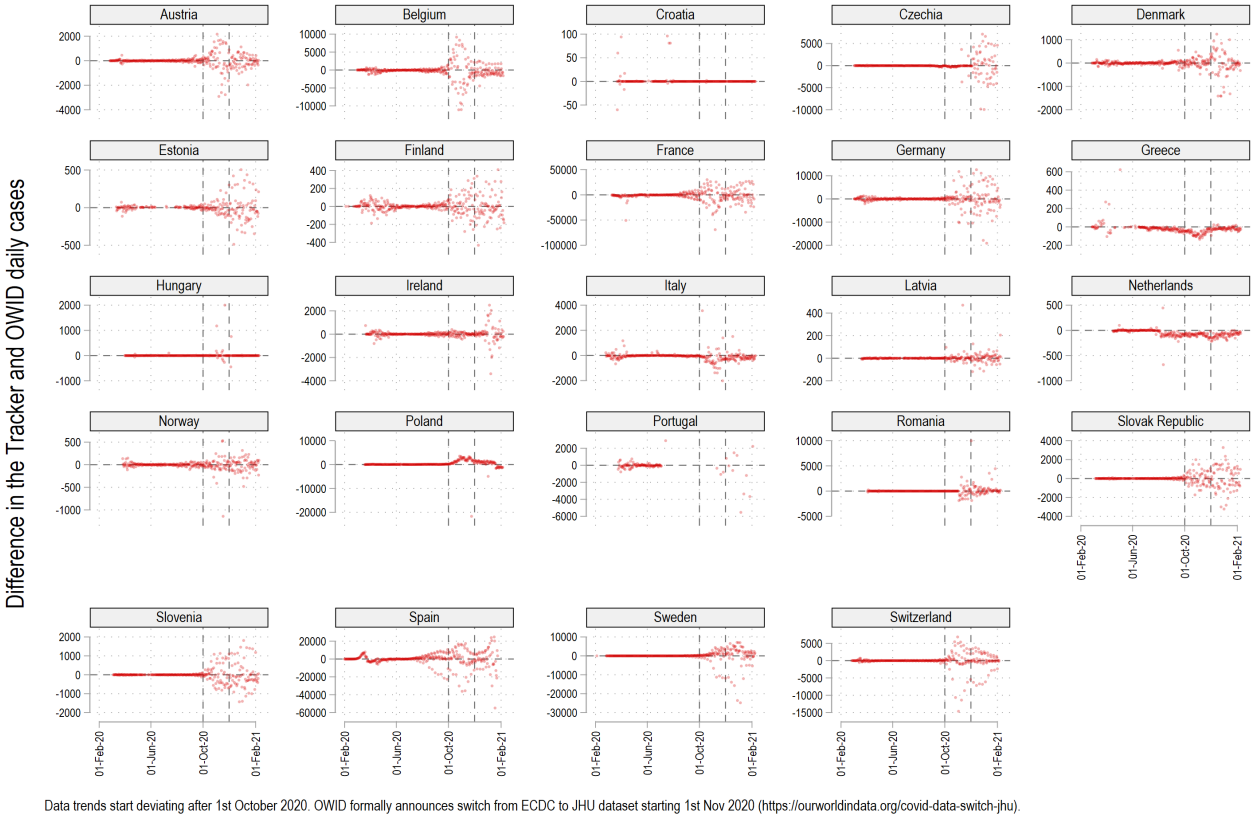
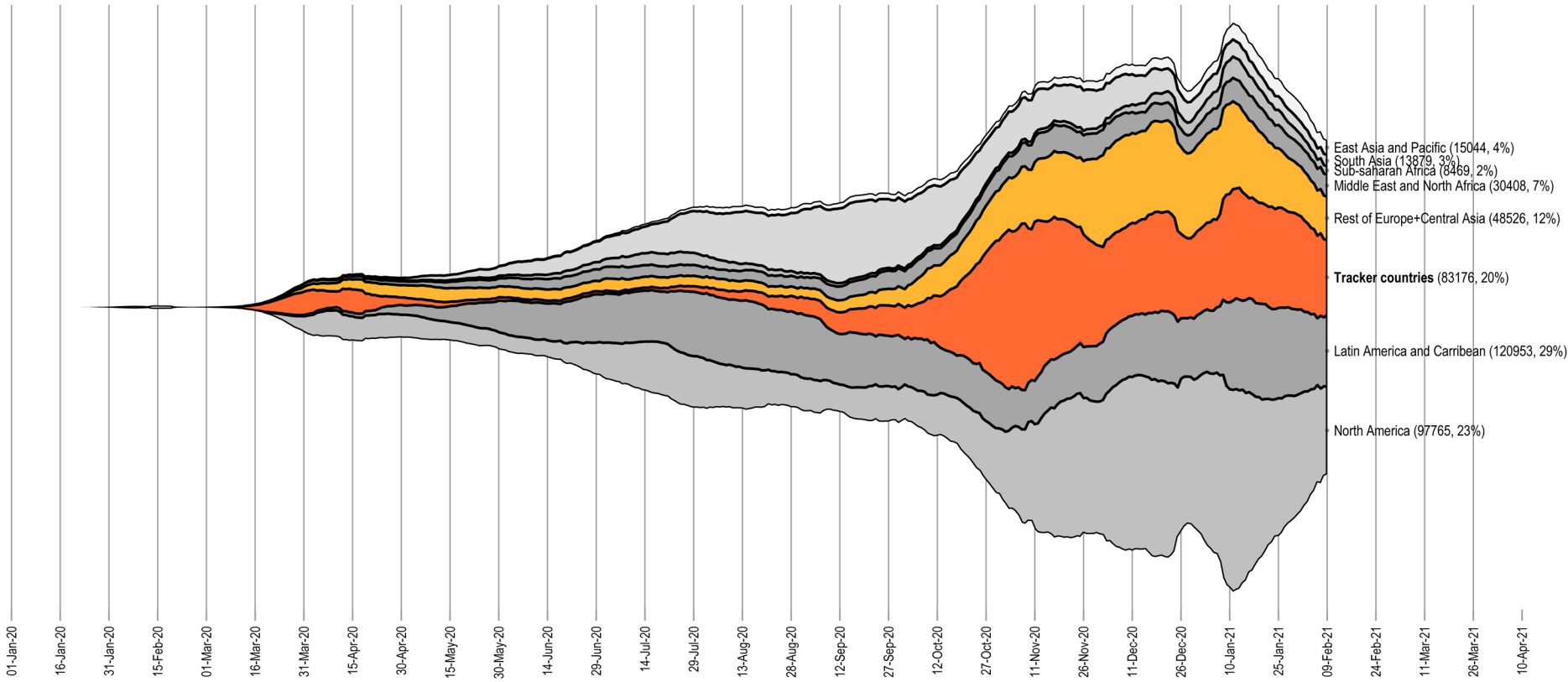


Figure 7. Difference between this Tracker and the Our World in Data (OWID) values for daily cases at the country level.

(a)

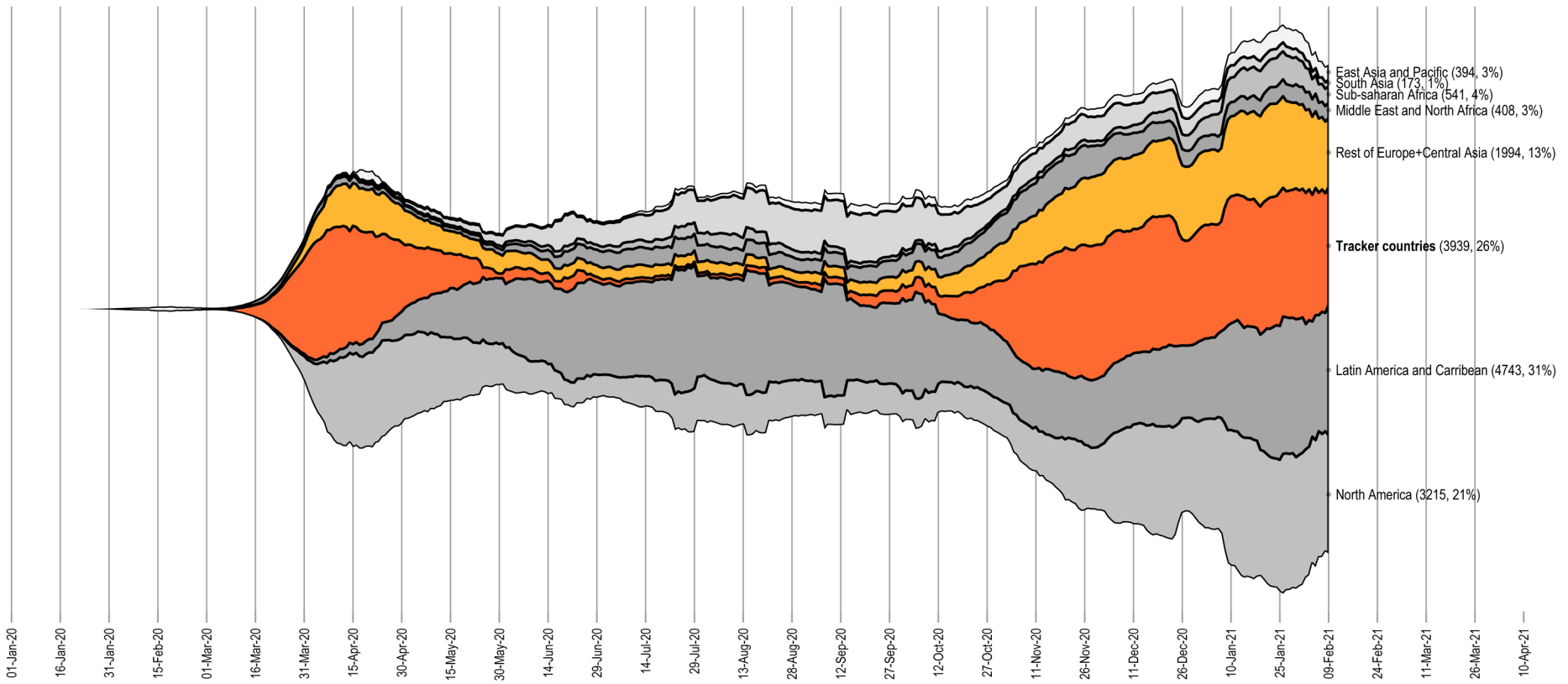
COVID-19 daily global cases: 418,320 (9 Feb 21)



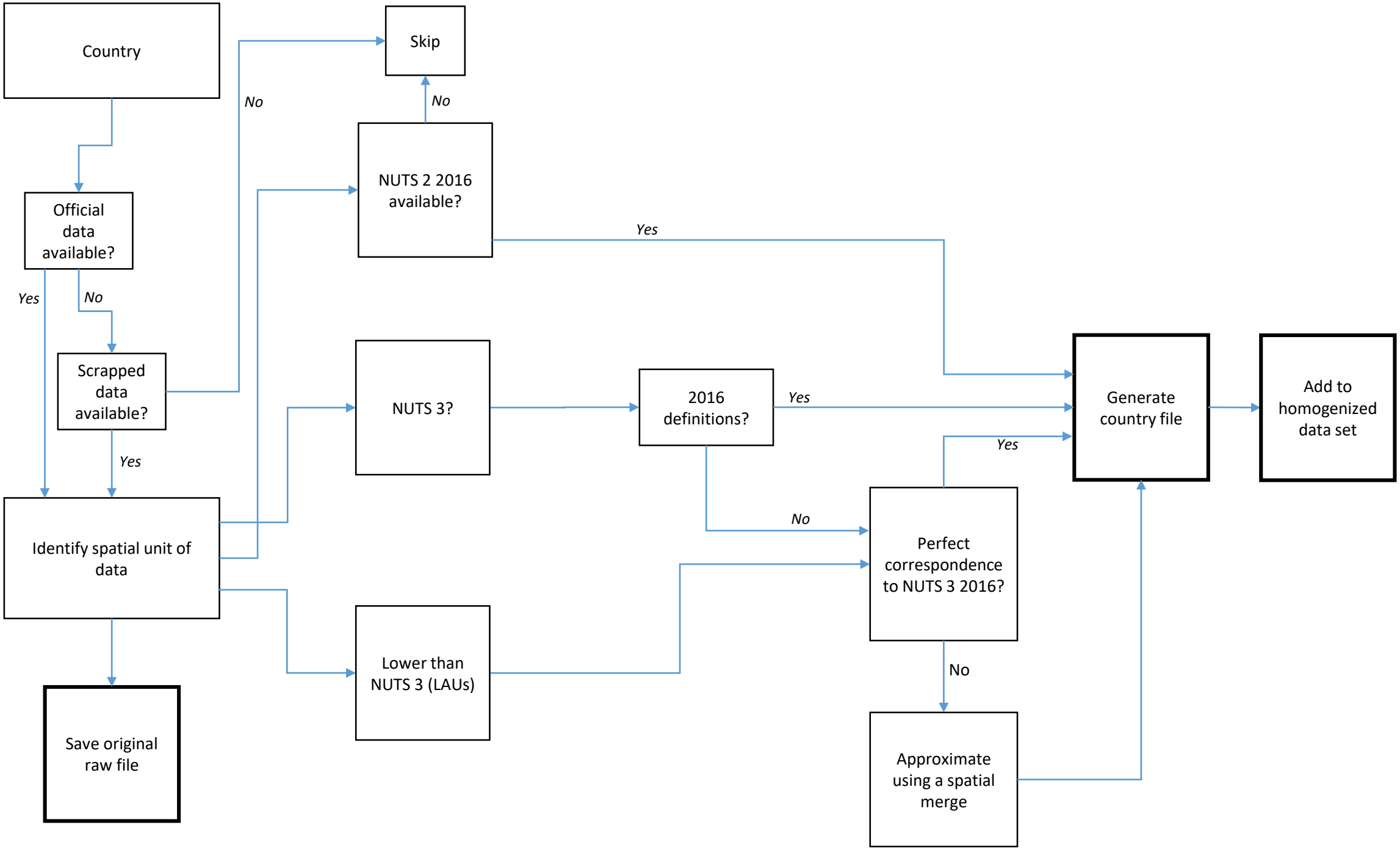
Data sources: Our World in Data. World Bank 2020 classifications used for country groups.

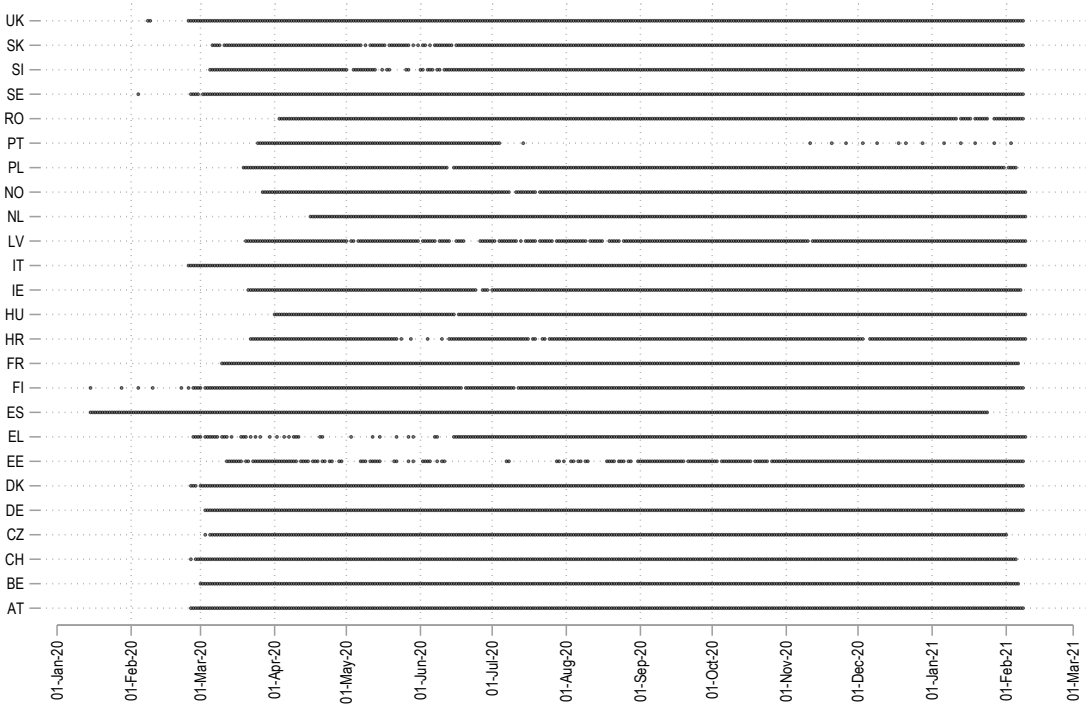
(b)

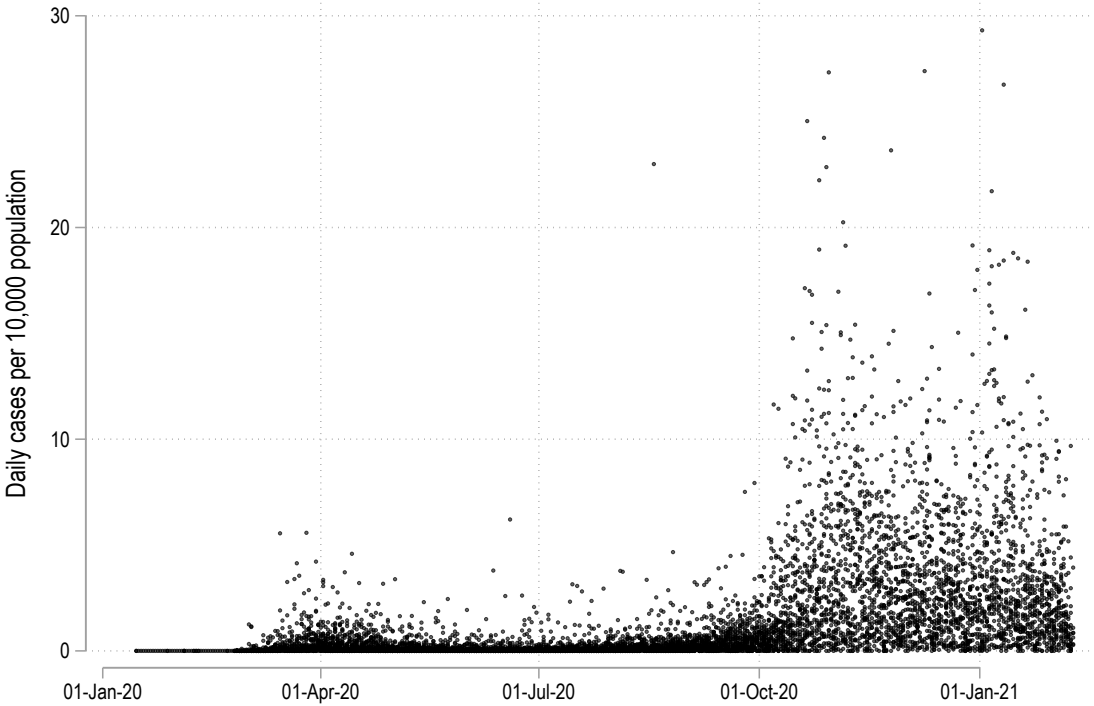
COVID-19 daily global deaths: 15,507 (9 Feb 21)

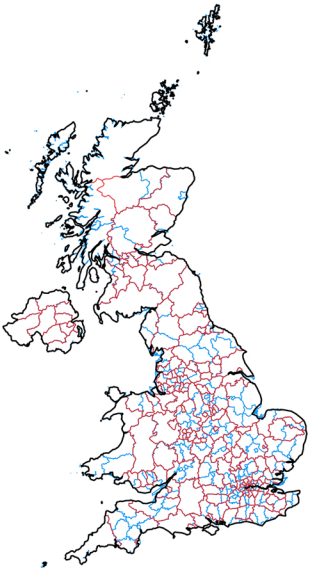


Data sources: Our World in Data. World Bank 2020 classifications used for country groups.

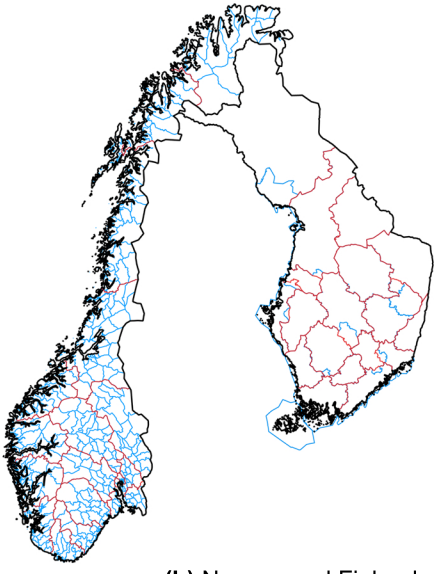

Country
Official data available?
Yes
No
Scrapped data available?
No
Yes
Skip
No
NUTS 2 2016 available?
Yes
Identify spatial unit of data
NUTS 3?
2016 definitions?
Yes
No
Lower than NUTS 3 (LAUs)
Perfect correspondence to NUTS 3 2016?
Yes
No
Approximate using a spatial merge
Generate country file
Add to homogenized data set
Save original raw file


UK
SK
SI
SE
RO
PT
PL
NO
NL
LV
IT
IE
HU
HR
FR
FI
ES
EL
EE
DK
DE
CZ
CH
BE
AT
01-Jan-20
01-Feb-20
01-Mar-20
01-Apr-20
01-May-20
01-Jun-20
01-Jul-20
01-Aug-20
01-Sep-20
01-Oct-20
01-Nov-20
01-Dec-20
01-Jan-21
01-Feb-21
01-Mar-21


Daily cases per 10,000 population
30
20
10
0
01-Jan-20
01-Apr-20
01-Jul-20
01-Oct-20
01-Jan-21



(a) UK

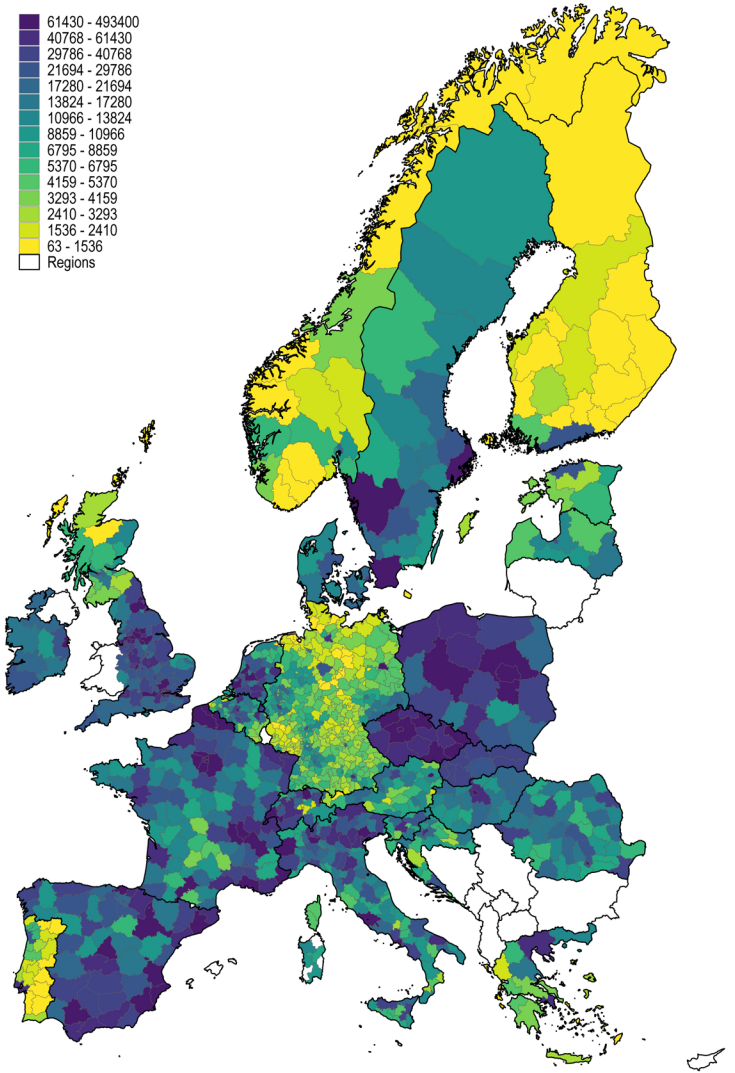


(b) Norway and Finland

(c) Italy

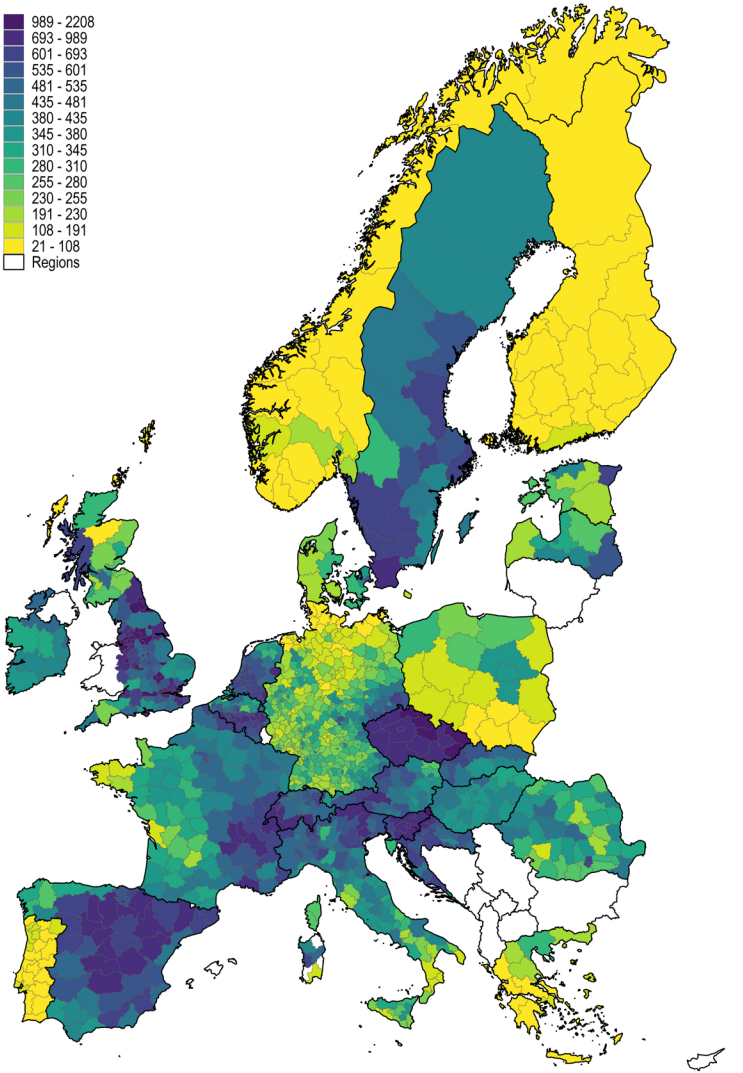
(d) Greece

(a) COVID-19 cumulative cases (15 Jan 20 - 9 Feb 21)

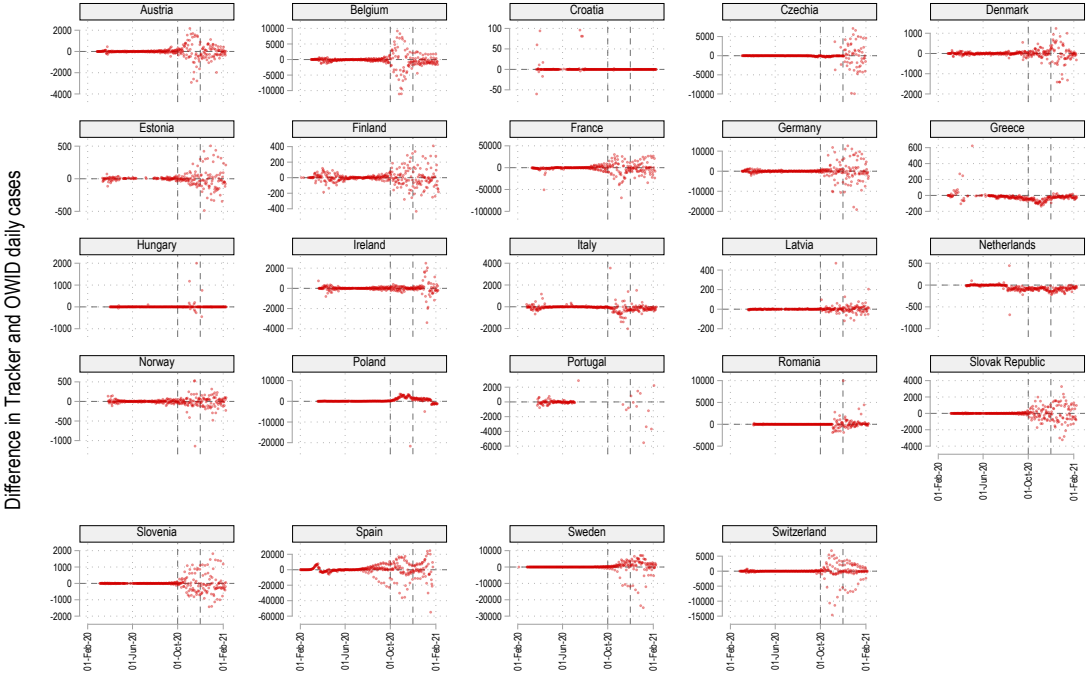


Map layer: Eurostat GISCO 2016 NUTS layers. Data: Misc sources. Data is at NUTS-3 level except for Poland and Greece.

(b) COVID-19 cumulative cases per 10,000 pop (15 Jan 20 - 9 Feb 21)



Map layer: Eurostat GISCO 2016 NUTS layers. Data: Misc sources. Data is at NUTS-3 level except for Poland and Greece.



Data trends start deviating after 1st October 2020. OWID formally announces switch from ECDC to JHU dataset starting 1st Nov 2020 (https://ourworldindata.org/covid-data-switch-jhu).