The CTIS R package

Anna-Carolina Haensch, Yue Xiong

15 7 2021

The package

This package offers an R interface for the APIs connected to the Global COVID-19 Trends and Impact Survey, formerly known as COVID-19 World Symptoms Survey.

The Global COVID-19 Trends and Impact Survey

The University of Maryland Social Data Science Center Global COVID-19 Trends and Impact Survey, in partnership with Facebook, is available in 56 languages. A representative sample of Facebook users is invited on a daily basis to report on topics including, for example, symptoms, social distancing behavior, vaccine acceptance, mental health issues, and financial constraints. Facebook provides weights to reduce nonresponse and coverage bias.

Data for the United States is not included in the Global COVID-19 Trends and Impact Survey but is available at covidcast.cmu.edu.

Open Data and Microdata API

Country and region-level statistics are published daily via the opendata API [2] and dashboards, and microdata is available for researchers via data use agreements. Over half a million responses are collected daily. For more information on the data, see https://covidmap.umd.edu/. While the open data is available without an account, the microdata API requires an account that can be requested here: https://dataforgood.fb.com/docs/covid-19-symptom-survey-request-for-data-access/.

Working example

Install and load the package

As a first step, we install and load the package. At the moment, the package is only on Github, but we also plan to publish it on CRAN.

```
# Installation (required only once)
# remotes::install_github("CaroHaensch/CTIS")

# Loading (always required)
library(CTIS)
```

Check availability

We will first check for which dates data is available for a specific country.

```
CTIS_Germany_available_dates <- CTIS_available_country(country = "Germany")

# Number of available dates
nrow(CTIS_Germany_available_dates)
```

```
#> [1] 445

# Some example dates
head(CTIS_Germany_available_dates)
#>   data.country data.survey_date  status
#> 1      Germany         20200423 success
#> 2      Germany         20200424 success
#> 3      Germany         20200425 success
#> 4      Germany         20200426 success
#> 5      Germany         20200427 success
#> 6      Germany         20200428 success
```

We can do the same check (available dates) for specific regions in a country. Compared to the country names (in English), the names for the regions are usually specified in the native language. As an example, we check the available dates for the region Bavaria ("Bayern") in Germany.

```
CTIS_Bavaria_available_dates <- CTIS_available_region(country = "Germany",
                                                      region = "Bayern")

# Number of available dates
nrow(CTIS_Bavaria_available_dates)
#> [1] 436

# Some example dates
head(CTIS_Bavaria_available_dates)
#>   data.country data.region data.survey_date  status
#> 1      Germany      Bayern         20200501 success
#> 2      Germany      Bayern         20200502 success
#> 3      Germany      Bayern         20200503 success
#> 4      Germany      Bayern         20200504 success
#> 5      Germany      Bayern         20200505 success
#> 6      Germany      Bayern         20200506 success
```

Load data from the Open Data API

As already mentioned, country and region-level statistics are published daily via the opendata API [2] and dashboards.

A full list of indicators is available here: https://gisumd.github.io/COVID-19-API-Documentation/docs/indicators/indicators.html

We will look up the indicator Vaccine Acceptance for three dates in May 2021 for Germany. We obtain different weighted and unweighted point and variance estimates.

```
CTIS_open_data_country(indicator = "vaccine_acpt",
                       type = "daily",
                       country = "Germany",
                       daterange = "20210501-20210503")
#>   data.percent_vu data.vu_se data.percent_vu_unw data.vu_se_unw
#> 1        0.741662   0.011080            0.739149       0.008971
#> 2        0.755268   0.010899            0.759745       0.008521
#> 3        0.767383   0.010731            0.756687       0.008912
#>   data.sample_size data.country data.iso_code data.gid_0 data.survey_date
#> 1             2396      Germany           DEU        DEU         20210501
#> 2             2514      Germany           DEU        DEU         20210502
#> 3             2318      Germany           DEU        DEU         20210503
```

```
#>    status
#> 1 success
#> 2 success
#> 3 success
```

Again, we can also receive data for a specific region in a country.

```
CTIS_open_data_region(indicator = "vaccine_acpt",
                      type = "daily",
                      country = "Germany",
                      region = "Bayern",
                      daterange = "20210501-20210503")
#>   data.percent_vu data.vu_se data.percent_vu_unw data.vu_se_unw
#> 1        0.730857   0.030113            0.740484       0.025786
#> 2        0.737634   0.031187            0.738516       0.026122
#> 3        0.781430   0.031339            0.773504       0.027362
#>   data.sample_size data.country data.region data.iso_code data.gid_0 data.gid_1
#> 1              289      Germany      Bayern           DEU        DEU    DEU.2_1
#> 2              283      Germany      Bayern           DEU        DEU    DEU.2_1
#> 3              234      Germany      Bayern           DEU        DEU    DEU.2_1
#>   data.survey_date  status
#> 1         20210501 success
#> 2         20210502 success
#> 3         20210503 success
```

Microdata API

While the open data is available without an account, the microdata API requires an account that can be requested here by researchers: https://dataforgood.fb.com/docs/covid-19-symptom-survey-request-for-data-access/

The following function can then be used to download, save as a .csv and load all at once the microdata for a specific date.

```
username <- "Specify your username here."
password <- "Specify your password here."

# Once you have received access, uncomment the following lines
# and specify the date for which you wish to receive the microdata.

# CTIS_microdata(username = username,
#                password = password,
#                date = "YYYY-MM-DD")
```

And that's it. We thank the University of Maryland and the Facebook team for providing us with this fantastic data source! We also want to note that we used the tutorials provided on the https://covidmap.umd.edu/ site to create the core of the functions.

References

[1] Junchuan Fan, Yao Li, Kathleen Stewart, Anil R. Kommareddy, Andres Garcia, Jinyi Ma, Zheng Liu, Joe O'Brien, Adrianne Bradford, Xiaoyi Deng, Samantha Chiu, Frauke Kreuter, Neta Barkay, Alyssa Bilinski, Brian Kim, Tal Galili, Daniel Haimovich, Sarah LaRocca, Stanley Presser, Katherine Morris, Joshua A Salomon, Elizabeth A. Stuart, Ryan Tibshirani, Tali Alterman Barash, Curtiss Cobb, Andi Gros, Ahmed Isa, Alex Kaess, Faisal Karim, Roee Eliat, Ofir Eretz Kedosha, Shelly Matskel, Roee Melamed, Amey Patankar, Irit Rutenberg, Tal Salmona, David Vannette (2020). The University of Maryland Social Data Science Center Global COVID-19 Trends and Impact Survey, in partnership with Facebook. https://covidmap.umd.edu/api.html