localcovid19now: processing and mapping COVID-19 case data at subnational scales

Stephen J. Beckett, Freyja A. Brandel-Tanis, Quan Nguyen, Aroon T. Chande, Lavanya Rishishwar, Clio Andris, and Joshua S. Weitz

1 School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, United States of America
2 School of City and Regional Planning, Georgia Institute of Technology, Atlanta, GA, United States of America
3 School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, United States of America
4 School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA, United States of America
5 Applied Bioinformatics Laboratory, Georgia Institute of Technology, Atlanta, GA, United States of America
6 School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, United States of America
7 School of Physics, Georgia Institute of Technology, Atlanta, GA, United States of America
8 Institut de Biologie, École Normale Supérieure, Paris, France

Corresponding author

DOI: 10.21105/joss.04898

Software

• Review
• Repository
• Archive

Summary

The localcovid19now R package provides functionality to load, unify and visualize recent COVID-19 case data at subnational scales in order to provide localized situational reports and improve understanding of the scale of local COVID-19 transmission. The package loads data from a variety of data sources and returns the most recent estimate of recorded per capita active COVID-19 infections, the date of the most recent report, and the geometry of each region. These data can then be visualized via mapping documented per capita active infections. We also provide functionality to visualize the risk of exposure to COVID-19, given a particular event size.

Statement of need

The COVID-19 global pandemic has led to hundreds of millions of recorded cases, over six million recorded deaths, and critical pressures on health and economic systems (World Health Organization, 2022). A key challenge in mitigating ongoing spread is that disease transmission and human behavior are coupled (Bavel et al., 2020; Funk et al., 2010; Weitz et al., 2020). Coupled viral and human dynamics have limited the predictive window of infectious disease models (including those of COVID-19), while also catalyzing efforts to develop awareness campaigns that leverage localized real-time information to inform the public and mitigate transmission. Beginning in May 2020, our team launched the COVID-19 Event Risk Tool (https://covid19risk.biosci.gatech.edu/), a situational dashboard with maps estimating exposure risk to COVID-19 across counties in the US (Chande et al., 2020). We use the number of active cases per capita and an assumed case ascertainment bias to estimate current local COVID-19 prevalence $p$. We then quantify and visualize COVID-19 exposure risk $\kappa$ (%) within a group of size $n$ as $\kappa = 100 \times (1 - (1 - p)^n)$. Since launch, more than 16M individuals have accessed the site and several similar international efforts have been initiated, e.g., for Italy (Eventi e Covid-19, 2020), Spain (Eventos y Covid-19, 2020), Mexico (COVID-19 México, 2020), Argentina (Cassese, 2020). Here, in the localcovid19now R package, we extend our initial release (Chande et al., 2020) to develop a pipeline to combine population estimates, COVID-19 case data, and spatial geometry such that risk estimates can...
be performed at localized scales across multiple countries. Our package aims to facilitate the process of downloading and visualizing recent COVID-19 data, in support of public health research and campaigns to inform and alert the public of the ongoing risk of transmission – and is of particular interest to public health scientists and researchers, epidemiologists, community leaders, and policymakers.

Our work is complementary to existing efforts to provide localized COVID-19 data across multiple countries, e.g., (Dong et al., 2020; Guidotti & Ardia, 2020; Palmer et al., 2021; Waltitinez & others, 2020). Via localcovid19now, we focus on aggregating recent case data rather than assembling a timeseries of all different COVID-19 data metrics. We use recent case data as the basis for providing localized real-time estimates of active cases. Additionally, spatial coverage differs among existing aggregate efforts to harmonize COVID-19 data. In order to maximize geographic coverage, the localcovid19now package pulls in data collated by John Hopkins U CSSE (Dong et al., 2020), from Google COVID-19 Open Data (Waltitinez & others, 2020), the Humanitarian Data Exchange, the COVID-19 Data Hub Guidotti (2022), and the WHO European Region COVID19 Subnational Explorer (WHO-EURO COVID19 Subnational Explorer, 2020). We also include additional governmental/volunteer curated subnational datasets, e.g., (Berry et al., 2021; Cota, 2020; Jefferies et al., 2020), (a full list of sources is provided with the package). Individual sources store data in different file formats, languages and file structures requiring tailored data retrieval strategies. Given that as time progresses and needs or efforts of those maintaining COVID-19 datasets change, the sourced datasets may change formats (e.g., Czechia added an extra identifier column to their dataset in November 2021 (Komenda et al., 2020)), change reporting areas (e.g., administrative border changes in the Netherlands), become deprecated (e.g., Cuba which last updated in September 2021), or otherwise be forced offline (e.g., Ireland (McNamee, 2021)). For this reason we have modularized our data reading process such that errors in one part of the pipeline do not prevent the loading of other datasets. An example of the number of active cases per 100,000 individuals as read and visualized via the localcovid19now pipeline is shown in Figure 1.

![Figure 1: Active cases per 100,000 people (as of 9 December 2022). Greyscale polygons denote areas where data estimates are over 30 days old, which we show here to show the current package scope. This map was created using the localcovid19now package by running the code: GlobalMap = LoadData(dropNAcountries=FALSE); PerCapitaMap_tmap(GlobalMap,100000). More examples are shown on the packages homepage.](image)

Our motivation in developing this project is to provide data-driven localized epidemic information that connects case data to potential exposure risk. While individual perception of COVID-19 exposure risk is linked to the utilization of risk-reducing actions, individual assessment of risk has been found to be generally uncorrelated to estimated exposure risk (Sinclair, Hakimi, et al., 2021). In contrast, communicating infection exposure risk estimates, as we do at the COVID-19
Event Risk Tool, has been shown as an effective intervention in realigning perceptions of localized event risks (Sinclair, Hakimi, et al., 2021; Sinclair, Stanley, et al., 2021; Sinclair et al., 2022). The package additionally includes functionality to help users visualize data (via leaflet (Cheng et al., 2021) or tmap (Tennekes, 2018)), including that of per capita active documented infections and estimated event-based risk.

Acknowledgements

We thank Seolah Lee for her contributions to developing the COVID-19 Event Risk Tool. We thank Kaitlyn Johnson, Robel Kassa, and Leo Wolansky for constructive feedback. Development of the `localcovid19now` package was supported by funds from the Centers for Disease Control and Prevention (75D30121P10600), The Rockefeller Foundation (11356), the Charities in Aid Foundation, the Marier Cunningham Foundation, and the Chaire Blaise Pascal funds of the Île-de-France region.

Author contributions

SJB led the project. SJB, FAB, QN and ATC contributed to dataset collection and coding methods for data curation. SJB, FAB and ATC contributed to package development. LR, CA, and JSW provided substantial contributions via decision and design guidance, and project management. SJB wrote the manuscript with contributions from all authors.

References


