CDC Centers for Disease Control and Prevention
CDC 24/7: Saving Lives. Protecting People.™ Data.CDC.gov

Menu

COVID-19 Case Surveillance Public Use Data with Geography Case Surveillance

View Data Visualize
Export
API

Access this Dataset via OData

Use OData to open the dataset in tools like Excel or Tableau. This provides a direct connection to the data that can be refreshed on-demand within the connected application.

Tableau users should select the OData v2 endpoint option.

Socrata OData Documentation

OData Endpoint

https://data.cdc.gov/api/odata/v4/n8mc-b4w4	OData V4 ⌄	Copy

Done
Updated April 8, 2022
Data Provided by CDC Data, Analytics and Visualization Task Force

About this Dataset

Mute Dataset

Mute this asset to stop receiving notifications. To resume notifications, you can unmute the asset at any time.

Updated
April 8, 2022

Data Last Updated April 6, 2022
Metadata Last Updated April 8, 2022
Date Created February 3, 2021

Views	Downloads
256K	**26.8K**

Data Provided by CDC Data, Analytics and Visualization Task Force
Dataset Owner Brian Lee

Contact Dataset Owner

Common Core

Contact Name	Surveillance Review and Response Group
Contact Email	eocevent394@cdc.gov
Publisher	CDC
Public Access Level	Data asset is publicly available to all without restrictions (public)
References	https://github.com/CDCgov/covid_case_privacy_review/blob/master/analysis/utility_summary_public_geo.pdf
Bureau Code	Centers for Disease Control and Prevention (009:20)
Program Code	HHS:Program Management (009:020)
Geographic Coverage	US
Temporal Applicability	2020-01-01/2022-04-04
Data Dictionary	https://data.cdc.gov/api/views/n8mc-b4w4/files/5245e8f8-d328-4bad-8dc3-4c3a80d73e36?download=true&filename=data_dictionary_covid_cases_public_geo.xlsx
Update Frequency	Monthly

Data Quality

Suggested Citation	Centers for Disease Control and Prevention, COVID-19 Response. COVID-19 Case Surveillance Public Use Data with Geography (version date: April 04, 2022).
Geospatial Resolution	State, County
Analytical Methods Reference	https://github.com/CDCgov/covid_case_privacy_review

Attachments

faq_covid_cases_public_geo.pdf
utility_summary_public_geo.pdf

Topics

Category	Case Surveillance
Tags	covid-19, covid19, coronavirus, cases, surveillance, county, state, microdata

Licensing and Attribution

License	Public Domain U.S. Government
Source Link	https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/index.html

Show Less

What's in this Dataset?

Rows
69.7M
Columns
19
Each row is a
Deidentified Patient Case

Columns in this Dataset

Column Name	Description	Type
case_month	The earlier of month the Clinical Date (date related to the illness or specimen collection) or the Date Received by CDC	Plain Text
Data Type		
Text		
API Field Name		
case_month		
res_state	State of residence	Plain Text
Data Type		
Text		
API Field Name		
res_state		
state_fips_code	State FIPS code	Plain Text
Data Type		
Text		
API Field Name		
state_fips_code		
res_county	County of residence	Plain Text

April 7, 2022: An adjustment was made to CDC's cleaning algorithm for COVID-19 line level case notification data. An assumption in CDC's algorithm led to misclassifying deaths that were not COVID-19 related. The algorithm has since been revised, and this dataset update reflects corrected individual level information about death status for all cases collected to date.

This case surveillance public use dataset has 19 elements for all COVID-19 cases shared with CDC and includes demographics, geography (county and state of residence), any exposure history, disease severity indicators and outcomes, and presence of any underlying medical conditions and risk behaviors.

Currently, CDC provides the public with three versions of COVID-19 case surveillance line-listed data: this 19 data element dataset with geography, a 12 data element public use dataset, and a 32 data element restricted access dataset.

The following apply to the public use datasets and the restricted access dataset:
- Data elements can be found on the COVID-19 case report form located at www.cdc.gov/coronavirus/2019-ncov/downloads/pui-form.pdf.
- Data are considered provisional by CDC and are subject to change until the data are reconciled and verified with the state and territorial data providers.
- Some data are suppressed to protect individual privacy.
- Datasets will include all cases with the earliest date available in each record (date received by CDC or date related to illness/specimen collection) at least 14 days prior to the creation of the previously updated datasets. This 14-day lag allows case reporting to be stabilized and ensure that time-dependent outcome data are accurately captured.
- Datasets are updated monthly.
- Datasets are created using CDC's Policy on Public Health Research and Nonresearch Data Management and Access and include protections designed to protect individual privacy.
- For more information about data collection and reporting, please see wwwn.cdc.gov/nndss/data-collection.html.
- For more information about the COVID-19 case surveillance data, please see www.cdc.gov/coronavirus/2019-ncov/covid-data/faq-surveillance.html.

Overview

The COVID-19 case surveillance database includes patient-level data reported by U.S. states and autonomous reporting entities, including New York City and the District of Columbia (D.C.), as well as U.S. territories and affiliates. On April 5, 2020, COVID-19 was added to the Nationally Notifiable Condition List and classified as "immediately notifiable, urgent (within 24 hours)" by a Council of State and Territorial Epidemiologists (CSTE) Interim Position Statement (Interim-20-ID-01). CSTE updated the position statement on August 5, 2020 to clarify the interpretation of antigen detection tests and serologic test results within the case classification (Interim-20-ID-02). The statement also recommended that all states and territories enact laws to make COVID-19 reportable in their jurisdiction, and that jurisdictions conducting surveillance should submit case notifications to CDC. COVID-19 case surveillance data collected by jurisdictions are shared voluntarily with CDC.
For more information, visit: wwwn.cdc.gov/nndss/conditions/coronavirus-disease-2019-covid-19/case-definition/2020/08/05/.

COVID-19 Case Reports

COVID-19 case reports are routinely submitted to CDC by public health jurisdictions using nationally standardized case reporting forms. On April 5, 2020, CSTE released an Interim Position Statement with national surveillance case definitions for COVID-19. Current versions of these case definitions are available at www.cdc.gov/nndss/conditions/coronavirus-disease-2019-covid-19/. All cases reported on or after were requested to be shared by public health departments to CDC using the standardized case definitions for lab-confirmed or probable cases. On May 5, 2020, the standardized case reporting form was revised. States and territories continue to use this form.

Data are Considered Provisional

- COVID-19 case surveillance data are dynamic; case reports can be modified at any time by the jurisdictions sharing COVID-19 data with CDC. CDC may update prior cases based on any updated information from jurisdictions. For instance, health departments provide updated data to CDC as new information is gathered about previously reported cases. As more information and data become available, analyses might find changes in surveillance data and trends during a previously reported time window. Data may also be shared late with CDC due to the volume of COVID-19 cases.
- Annual finalized data: To create the final NNDSS data used in the annual tables, CDC works carefully with the reporting jurisdictions to reconcile data received during the year until each state or territorial epidemiologist confirms that the data from their area are correct.

Data Limitations

To learn more about the limitations in using case surveillance data, visit FAQ: COVID-19 Data and Surveillance.

Data Quality Assurance Procedures

CDC's Case Surveillance Section routinely performs data quality assurance procedures (i.e., ongoing corrections and logic checks to address data errors). To date, the following data cleaning steps have been implemented:
- Questions that have been left unanswered (blank) on the case report form are reclassified to a Missing value, if applicable to the question. For example, in the question "Was the individual hospitalized?" where the possible answer choices include "Yes," "No," or "Unknown," the blank value is recoded to "Missing" because the case report form did not include a response to the question.
- Logic checks are performed for date data. If an illogical date has been provided, CDC reviews the data with the reporting jurisdiction. For example, if a symptom onset date in the future is reported to CDC, this value is set to null until the reporting jurisdiction updates the date appropriately.
- Additional data quality processing to recode free text data are ongoing. Data on symptoms, race and ethnicity, and healthcare worker status have been prioritized.

Data Suppression

To prevent release of data that could be used to identify people, data cells are suppressed for low frequency (<11 COVID-19 case records with a given values). Suppression includes low frequency combinations of case month, geographic characteristics (county and state of residence), and demographic characteristics (sex, age group, race, and ethnicity). Suppressed values are re-coded to the NA answer option; records with data suppression are never removed.

Additional COVID-19 Data

COVID-19 data are available to the public as summary or aggregate count files, including total counts of cases and deaths by state and by county. These and other COVID-19 data are available from multiple public locations: COVID Data Tracker; United States COVID-19 Cases and Deaths by State; COVID Data Tracker Weekly Review; COVID-19 Surveillance and Data Analytics; COVID-19 Vaccination Reporting Data Systems; and COVID-19 Death Data and Resources.

Column Name	Description
Data Type Text API Field Name res_county	
county_fips_code	County FIPS code
Data Type Text API Field Name county_fips_code	
age_group	Age group [0 - 17 years; 18 - 49 years; 50 - 64 years; 65 + years; Unknown; Missing; NA, if value suppressed for privacy protection.]
Data Type Text API Field Name age_group	
sex	Sex [Female; Male; Other; Unknown; Missing; NA, if value suppressed for privacy protection.]
Data Type Text API Field Name sex	
race	Race [American Indian/Alaska Native; Asian; Black; Multiple/Other; Native Hawaiian/Other Pacific Islander; White; Unknown; Missing; NA, if
Data Type Text API Field Name race	
ethnicity	Ethnicity [Hispanic; Non-Hispanic; Unknown; Missing; NA, if value suppressed for privacy protection.]
Data Type Text API Field Name ethnicity	
case_positive_specimen_interval	Weeks between earliest date and date of first positive specimen collection
Data Type Number API Field Name case_positive_specimen	
case_onset_interval	Weeks between earliest date and date of symptom onset.
Data Type Number API Field Name case_onset_interval	
process	Under what process was the case first identified? [Clinical evaluation; Routine surveillance; Contact tracing of case patient; Multiple; Other; Unk
Data Type Text API Field Name process	
exposure_yn	In the 14 days prior to illness onset, did the patient have any of the following known exposures: domestic travel, international travel, cruise ship c facility (nursing, assisted living, or long-term care facility), school/university/childcare center, correctional facility, community event/mass gather case? [Yes, Unknown, Missing]
Data Type Text API Field Name exposure_yn	
current_status	What is the current status of this person? [Laboratory-confirmed case, Probable case]
Data Type Text API Field Name current_status	
symptom_status	What is the symptom status of this person? [Asymptomatic, Symptomatic, Unknown, Missing]
Data Type Text API Field Name symptom_status	
hosp_yn	Was the patient hospitalized? [Yes, No, Unknown, Missing]
Data Type Text API Field Name hosp_yn	
icu_yn	Was the patient admitted to an intensive care unit (ICU)? [Yes, No, Unknown, Missing]
Data Type Text API Field Name icu_yn	
death_yn	Did the patient die as a result of this illness? [Yes; No; Unknown; Missing; NA, if value suppressed for privacy protection.]
Data Type Text API Field Name death_yn	
underlying_conditions_yn	Did the patient have one or more of the underlying medical conditions and risk behaviors: diabetes mellitus, hypertension, severe obesity (BMI>4 chronic diseases, immunosuppressive condition, autoimmune condition, current smoker, former smoker, substance abuse or misuse, disability, ps
Data Type Text API Field Name underlying_conditions_yn	

Show All (19)Show Less

Table Preview

View Data Create Visualization

case_month	res_state	state_fips_code	res_county	county_fips_code	age_group	sex	race	ethnicity	case_positive_specimen_interval	case_onset_interval	process	exposure_yn	current_status	symptom_status	hosp_yn	icu_yn	death_yn	underlying_conditions_yn
2021-10	NE	31	ADAMS	31001	18 to 49 years	NA	NA	NA			Missing	Missing	Laboratory-confirmed case	Missing	Missing	Missing	Missing	
2021-01	ME	23	LINCOLN	23015	65+ years	NA	NA	NA		0	Missing	Missing	Laboratory-confirmed case	Symptomatic	Yes	Missing	No	
2021-08	OK	40	GARVIN	40049	0 - 17 years	Female	NA	NA			Missing	Missing	Laboratory-confirmed case	Unknown	No	Missing	NA	
2022-01	IN	18	LAWRENCE	18093	0 - 17 years	Female	NA	NA	0		Missing	Missing	Laboratory-confirmed case	Missing	No	Missing	No	
2020-07	AL	01	SHELBY	01117	0 - 17 years	Female	Missing	Missing			Missing	Missing	Laboratory-confirmed case	Missing	Missing	Missing	Missing	
2020-09	IL	17	CLINTON	17027	18 to 49 years	Female	Missing	Missing		0	Missing	Missing	Laboratory-confirmed case	Missing	No	Missing	Missing	
2021-03	GA	13	DAWSON	13085	18 to 49 years	Female	Unknown	Missing		0	Missing	Missing	Probable Case	Symptomatic	Missing	Missing	Missing	
2021-09	OH	39	FULTON	39051	18 to 49 years	Female	NA	NA	0		Missing	Missing	Laboratory-confirmed case	Missing	Missing	Missing	Unknown	
2021-12	OH	39	PERRY	39127	18 to 49 years	Female	NA	NA	0	0	Missing	Missing	Laboratory-confirmed case	Symptomatic	Missing	Missing	Unknown	
2020-11	MO	29	RANDOLPH	29175	18 to 49 years	Female	NA	NA	0	0	Missing	Missing	Laboratory-confirmed case	Symptomatic	Unknown	Missing	Unknown	
2021-03	MI	26	BAY	26017	50 to 64 years	Female	NA	NA		0	Missing	Missing	Laboratory-confirmed case	Symptomatic	No	Missing	NA	
2021-04	OK	40	LINCOLN	40081	50 to 64 years	Female	NA	NA			Missing	Missing	Probable Case	Unknown	Unknown	Missing	NA	
2021-07	NV	32	NYE	32023	50 to 64 years	Female	NA	NA	0	0	Routine surveillance	Missing	Laboratory-confirmed case	Symptomatic	No	Missing	NA	Yes

Previous Next Showing Deidentified Patient Cases 1 to 13 out of 69,664,983
Click and drag to pan the chart
View Source Data ➔

Related Content Using this Data

Public **COVID-19 Case Surveillance Public Use Data with Geog…**

April 8, 2022

5 Views

Data profile with summary of variables designed to help support exploratory analysis. This case surveillance public u…

Open this dataset in Carto?

Your browser will be redirected to Carto, would you like to proceed?
Note that Carto has a file size limit of 150MB or 500,000 rows.
Learn about more external integrations at
https://support.socrata.com/hc/en-us/articles/115010730868
CloseOpen

Open this dataset in Plot.ly?

Your browser will be redirected to Plot.ly, would you like to proceed?
Note that Plot.ly has a file size limit of 5MB.
Learn about more external integrations at
https://support.socrata.com/hc/en-us/articles/115010730868
Close Open

Send a Message to the Owner of this Dataset

Send a short message to the dataset owner to ask a question, make a comment, or point out something about the data.

Subject
Message
Your Email (Your email address will be shared with the dataset owner so they can get back to you)

I'm not a robot
reCAPTCHA
Privacy - Terms

Cancel Send

Manage Featured Content

Featured Content

What would you like people to see when they first view this dataset? Promote up to 3 Socrata assets that use this data, or choose to feature any relevant external resources. Think of this section as the face of your data.

Add…
Add…
Add…
Done

sex