

COVID Datasets

- Data: COVID Public Datasets
- Data sources & targets:
 - Azure SQL
 - Google BigQuery
 - SQL Server or MySQL
 - csv files
- Data integration tools:
 - Talend Pipeline Manager
 - Stitch
 - Talend Big Data (optional)
 - Alteryx (optional)



COVID Data Engineering Project

- Objectives:
 - Perform source systems analysis on data sources (& compare)
 - Ingest & Integrate from/to various sources to generate BI Dashboards
 - Gain experience with cloud databases & cloud DW
 - Gain experience with data integration & data pipeline tools using data from both cloud and on-premise (notebook) sources
- Deliverables:
 - Dashboards visualizing COVID data tracking US states and counties
 - Microsoft PowerBI
 - Google Data Studio
 - Data should be loaded into:
 - Google BigQuery
 - Azure SQL (note: likely limitations on student credits)
 - SQL Server or MySQL
 - Review & compare data integration tools

COVID Data Engineering Project

Google Cloud Platform Select a project ▾

Search products and resources

Marketplace > Datasets

Datasets for COVID-19 research

Filter by

31 results

Type	Category	Dataset	Description
Datasets	Datasets for COVID-19	About COVID-19 Public Datasets	BigQuery Public Datasets Program
		Getting started with COVID-19 Public Datasets	
		American Community Survey (ACS)	United States Census Bureau
		American Hospital Association Hospital Capacity Dataset	American Hospital Association
		US Hospital Capacity Data	aggregated at the county level
		CDC Births Data Summary	Centers for Disease Control
		Nativity Data from CDC Births	
		BREATHE BioMedical Literature Dataset	BREATHE
		Extensive machine consumable biomed literature dataset	
		Center for Medicare and Medicaid Services - Dual...	U.S. Department of Health & Hum.
		Dual enrollment of Medicare and Medicaid by county	
		COVID-19 Cases by Country	European Centre for Disease Prev...
		ECDC worldwide COVID-19 cases	
		COVID-19 Cases in Italy	Italian Department of Civil Protect...
		Regional case volume from the Department of Civil Protection	
		COVID-19 Data Repository by CSSE at JHU	Johns Hopkins University
		Repository of aggregated coronavirus COVID-19 cases by JHU	
		COVID-19 Dataset of Global Research by Dimensions	Digital Science & Research Solutio...
		Datasets for COVID19	
		COVID-19 Mobility Impact	Geotab
		Impact of COVID-19 on commercial transportation activity	
		COVID-19 Open Data	BigQuery Public Datasets Program
		Daily time-series data related to COVID-19 globally	
		COVID-19 Public Forecasts	BigQuery Public Datasets Program
		COVID-19 forecasts for the next 28 days in the US and Japan	
		COVID-19 Search Trends symptoms dataset	BigQuery Public Datasets Program
		Trends in symptom search volumes due to COVID-19	
		GDP and Income by County	Bureau of Economic Analysis
		US income and GDP at the county level	
		Global Health	The World Bank
		Global healthy	
		Google Community Mobility Reports	BigQuery Public Datasets Program
		Changes in community movement due to COVID-19	
		Health Professional Shortage Areas	U.S. Department of Health & Hum...
		Federally designated areas with a shortage of healthcare workers	
		Hospital General Information	U.S. Department of Health & Hum...
		List of hospitals registered with Medicare	
		International Census Data	United States Census Bureau
		World population estimates 1950 through 2050	
		Low Income Housing Tax Credit Program	US Dept of Housing and Urban De...
		Data on low-income housing development across the US	
		OpenStreetMap Public Dataset	OpenStreetMap
		The OpenStreetMap planet-wide dataset loaded to BigQuery	
		Oxford COVID-19 Government Response Tracker	University of Oxford
		Data on policy responses to the COVID-19 outbreak	
		Point-in-Time Homelessness Count	US Dept of Housing and Urban De...
		Annual Homeless Assessment Report to Congress	
		RxRx19	Recursion
		Cell image dataset of drugs on SARS-CoV-2 infection/inflammation	
		SNAP Enrollment	US Department of Agriculture
		County-level enrollment for supplemental nutrition assistance	
		The Immune Epitope Database (IEDB)	U.S. Department of Health & Hum...
		Experimental data characterizing antibody and T cells epitopes	
		The New York Times US Coronavirus Database	The New York Times
		Data based on reports from state and local health agencies	
		US Census Data	United States Census Bureau
		2000 and 2010 US Census data	
		USAFacts US Coronavirus Database	USAFacts
		USAFacts COVID-19 Case and Death Data by state and county	
		Weather Data for COVID-19 Research	WeatherSource
		Past, present, and forecast weather data relevant to COVID-19	

COVID Data Engineering Project

Google Cloud Platform Select a project ▾

About COVID-19 Public Datasets

BigQuery Public Datasets Program

Getting started with COVID-19 Public Datasets

[VIEW DATASET](#)

[OVERVIEW](#) [SAMPLES](#)

Overview

In an effort to help combat COVID-19, we created a COVID-19 Public Datasets program to make data more accessible to researchers, data scientists and analysts. The program will host a repository of public datasets that relate to the COVID-19 crisis and make them free to access and analyze. These include datasets from the New York Times, European Centre for Disease Prevention and Control, Google, Global Health Data from the World Bank, and OpenStreetMap.

Free hosting and queries of COVID datasets

As with all data in the [Google Cloud Public Datasets Program](#), Google pays for storage of datasets in the program. BigQuery also provides free queries over certain COVID-related datasets to support the response to COVID-19. Queries on COVID datasets will not count against the [BigQuery sandbox free tier](#), where you can query up to 1TB free each month.

Limitations and duration

Queries of COVID data are free. If, during your analysis, you join COVID datasets with non-COVID datasets, the bytes processed in the non-COVID datasets will be counted against the free tier, then charged accordingly, to prevent abuse. Queries of COVID datasets will remain free until Sept 15, 2021.

The contents of these datasets are provided to the public strictly for educational and research purposes only. We are not onboarding or managing PII or PII data as part of the COVID-19 Public Dataset Program. Google has [practices & policies](#) in place to ensure that data is handled in accordance with widely recognized patient privacy and data security policies.

See [the list of all datasets](#) included in the program

Additional details

Type: [Datasets](#)

Last updated: 8/12/20

Category: [Datasets for COVID-19 research](#), [Science & research](#)

Dataset source: Various

Cloud service: BigQuery

Region: US/EU

Update frequency: Daily

Google Cloud Platform Select a project ▾

Navigation menu About COVID-19 Public Datasets
BigQuery Public Datasets Program

[OVERVIEW](#) [SAMPLES](#)

Samples

[More Information - Getting Started - FAQ](#)

How do I get started using COVID-19 Public Datasets?
The public datasets available as part of this program are available for query in BigQuery by default. There's no need to sign up or register for the program specifically. Just log on to BigQuery (see [Sandbox](#) below if you are new) and run your query. Each of the datasets have sample queries linked in the description so you can start there.

I'm new to BigQuery. How can I get started quickly?
BigQuery's [Sandbox](#) allows a new user to store up to 10 GB of data and query up to 1 TB of data each month for free without having to provide a credit card. This can be useful for citizen data scientists and researchers from organizations that don't have agreements with Google Cloud. Since queries over COVID-related datasets will be free, this is a quick and efficient way to get started.

What is the difference between a COVID-19 public dataset and a normal public dataset?
Google Cloud has an extensive [Public Datasets Program](#) that launched over four years ago. Datasets listed as COVID-19 public datasets are hosted by Google so that customers can more easily discover and access them as part of that program. However, we have taken the additional step of making queries over the COVID-19 datasets free until September 15th, 2021 to support response to the COVID-19 pandemic. To see which datasets are part of the COVID-19 public datasets, see [the list here](#).

If queries over COVID-19 datasets are free, why does the UI still tell me how many bytes the query will process?
We've left this feature in place because users find this helpful to help to optimize performance. So you can still see the bytes processed in your queries even though you will not be billed for bytes processed on datasets that are part of the COVID-19 Public Datasets Program. BigQuery reports two different numbers in the query statistics: the bytes processed and the bytes billed. For the most part, these numbers are the same. However, for queries against COVID-19 Public Datasets, the bytes billed will be lower, since they will not include any bytes read from those tables. To see the actual bytes billed, click on "Job Information" on the query results pane. It will show both the bytes scanned and bytes billed metrics.

COVID Data Engineering Project

Google Cloud Platform Select a project ▾

COVID-19 Open Data

BigQuery Public Datasets Program

Daily time-series data related to COVID-19 globally

[VIEW DATASET](#)

[OVERVIEW](#) [SAMPLES](#)

Overview

This repository contains country-level datasets of daily time-series data related to COVID-19 globally. It includes data relating to demographics, economy, epidemiology, geography, health, hospitalizations, mobility, government response, and weather.

This dataset is available in both the US and EU regions of BigQuery at the following links:

- [COVID-19 Open Data: US Region](#)
- [COVID-19 Open Data: EU Region](#)

All data in this dataset is retrieved automatically. When possible, data is retrieved directly from the relevant authorities, like a country's ministry of health.

This dataset has significant public interest in light of the COVID-19 crisis. All bytes processed in queries against this dataset will be zeroed out, making this part of the query free. Data joined with the dataset will be billed at the normal rate to prevent abuse. After September 15, queries over these datasets will revert to the normal billing rate.

This public dataset is hosted in Google BigQuery and is included in BigQuery's 1TB/mo of free tier processing. This means that each user receives 1TB of free BigQuery processing every month, which can be used to run queries on this public dataset. Watch this short video to learn how to get started quickly using BigQuery to access public datasets. [What is BigQuery](#)

Additional details

Type: [Datasets](#)

Last updated: 7/9/20

Category: [Datasets for COVID-19 research](#), [Science & research](#)

Dataset source: <https://github.com/open-covid-19/data>

Cloud service: BigQuery

Region: US/EU

Update frequency: Daily

Google Cloud Platform Select a project ▾

Navigation menu COVID-19 Open Data

BigQuery Public Datasets Program

[OVERVIEW](#) [SAMPLES](#)

Samples

Dataset Migration

If you are migrating from any of the other COVID-19 Public Datasets, such as the [New York Times "US Coronavirus Database"](#), [European Centre for Disease Prevention and Control's "COVID-19 Cases by Country"](#) or other datasets with data on confirmed cases and deaths either within the US or around the globe, the following list should help you remap your queries to the structure of this dataset. For an example of this, check out the [compatibility view](#) created in the US region copy of the dataset or the [compatibility view](#) created in the EU region copy.

The structure of this list is as follows:
Previous dataset column name: COVID-19 Open Data column name

- province_state : subregion1_name
- country_region : country_name
- date : date
- latitude : latitude
- longitude : longitude
- location_geom : location_geometry
- confirmed : cumulative_confirmed
- deaths : cumulative_deceased
- recovered : cumulative_recovered
- active : current_cases (NOTE: Calculated as [cumulative_confirmed-cumulative_recovered-cumulative_deceased] where none of these values are NULL)
- fips : subregion2_code
- admin_2 : subregion2_name
- combined_key : location_key

Sample Queries

Try the sample queries below in the BigQuery UI.

How many confirmed COVID-19 cases did each country have at the end of June?

This query determines the cumulative number of COVID-19 cases for each county, as well as a normalization of cumulative cases by country population, as of the end of June. [Run this query](#).

How did confirmed COVID-19 cases compare to COVID-19 tests in the US during the month of June, by state?

This query determines the percent of reported COVID-19 tests in each US state that returned positive by dividing the total number of reported tests that returned positive by the total number of COVID-19 tests conducted in the month of June. [Run this query](#).

Google BigQuery Sandbox (or Trail)

The screenshot shows the Google Cloud BigQuery Guides page. The left sidebar has a navigation menu with sections like BigQuery, Overview, Guides (which is selected), Reference, Samples, Support, and Resources. Under the Guides section, there's a sub-menu for BigQuery with items such as Product overview, What is BigQuery?, Quickstarts, Samples, How-to guides, Interacting with BigQuery, and a detailed section for BigQuery sandbox. The main content area displays the 'Using the BigQuery sandbox' guide. The page title is 'Using the BigQuery sandbox'. It includes a brief introduction about the sandbox, a video thumbnail titled 'Setting up the BigQuery sandbox', and a section on getting started with the sandbox. The right sidebar contains a 'Table of contents' with links to 'Getting started with the sandbox', 'Limits', 'Upgrading from the sandbox', 'Troubleshooting the sandbox', and 'What's next'.

BigQuery > Documentation > Guides

Rate and review

Using the BigQuery sandbox

The BigQuery sandbox gives you free access to the power of BigQuery subject to the sandbox's [limits](#). The sandbox lets you use the Cloud Console without providing a credit card. You can use the sandbox without creating a billing account or enabling billing for your project.

The Cloud Console is the graphical interface that you can use to create and manage BigQuery resources and to run SQL queries. See the [Cloud Console quickstart](#) for a working introduction to the Cloud Console.

Getting started with the sandbox

The BigQuery sandbox is available to any Google Cloud customer including Firebase users. To get started with the Google Cloud Free Tier, see [Google Cloud free tier](#). To get started with Firebase, see [the Firebase landing page](#).

If you are a current Firebase user, see [Link BigQuery to Firebase](#) in the Firebase Help for instructions on linking Firebase to BigQuery.

For updates on the BigQuery sandbox, see the [Release notes](#).

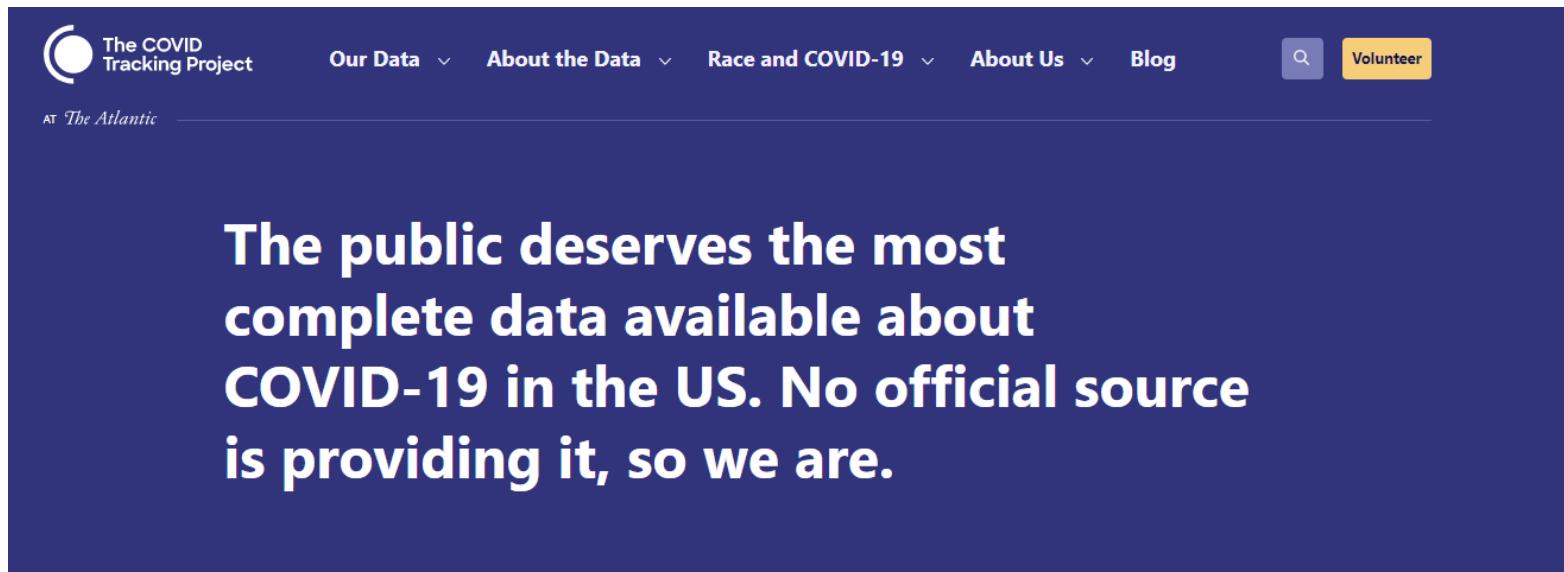
To open the sandbox:

1. Open the Cloud Console.
[Go to the Cloud Console](#)
2. In the navigation pane, click **BigQuery**. You can also open the Cloud Console directly by entering the following URL in your browser.
`https://console.cloud.google.com/bigquery`
3. Accept the terms of service.
4. Before you can use the BigQuery sandbox, you must create a project. Follow the prompts to create your new project.
5. After you create your project, the Cloud Console displays a sandbox banner like the following.

Google BigQuery Public Datasets

The screenshot shows the Google BigQuery Query editor interface. On the left, there is a sidebar with navigation links: Query history, Saved queries, Job history, Transfers, Scheduled queries, Reservations, BI Engine, Resources, and a search bar for tables and datasets. Below this is a list of datasets under the 'Resources' section, with 'covid19_open_data' highlighted. The main area is the 'Query editor' which contains a single row labeled '1'. Below the editor are buttons for Run, Save query, Schedule query, More, CREATE TABLE, SHARE DATASET, AUTHORIZE ROUTINES, COPY DATASET, and DELETE DATASET. The 'covid19_open_data' dataset is described as containing country-level datasets of daily time-series data related to COVID-19 globally. It was created on Jul 8, 2020, at 10:50:10 PM, and last modified on Aug 3, 2020, at 2:01:05 PM. The data location is US. A label 'freebq covid : freebqcovid' is associated with the dataset.

The COVID Tracking Project



The COVID Tracking Project logo and navigation bar are visible at the top. The main headline in the center reads: "The public deserves the most complete data available about COVID-19 in the US. No official source is providing it, so we are."

Every day, our volunteers compile the latest numbers on tests, cases, hospitalizations, and patient outcomes from every US state and territory.

Latest totals:

[See all data →](#)

181,142,529
Total test results

12,230,472
Cases

248,897
Deaths

The COVID Tracking Project

The COVID Tracking Project at The Atlantic

Our Data About the Data Race and COVID-19 About Us Blog

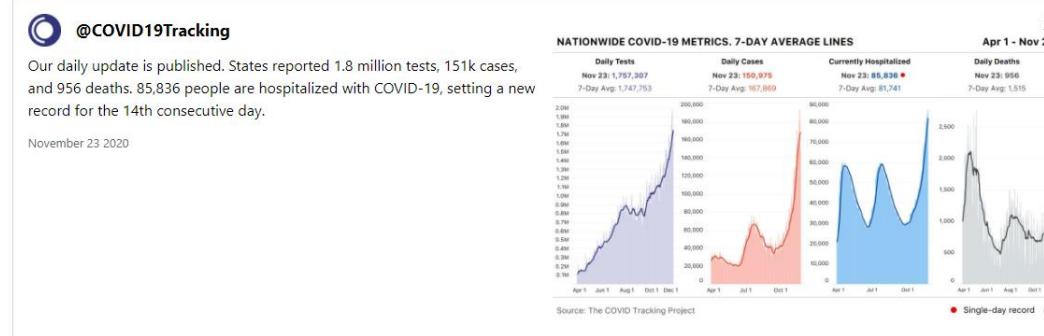
Search Volunteer

Our Data

Totals for the US Racial Data Dashboard Long-Term Care Our Charts Data API Download the Data

We update all our data each day between about 6pm and 7:30pm Eastern Time. We have recently begun publishing a bit later so as to capture [Colorado](#) and [New Mexico](#) data on the same day they update rather than on the following day.

[Sign up for our newsletter](#) to get our latest [articles about this data](#) by email.



Data for November 23, 2020

Get the data as: [CSV](#) [API](#)

Cases	Historical data	Tests	Historical data	Hospitalization	Historical data	Outcomes	Historical data
Total cases	12,230,472	Total tests	181,142,529	Currently hospitalized	85,836	Deaths	248,897
Definition		Definition		Definition		Definition	
New cases today	150,975	New tests today	1,757,307	Currently in ICU	16,811	We have removed recovered data for the US. Here's why.	
(Calculated)		(Calculated)		Definition			
Change over 7 days	10.7%+	Change over 7 days	7.3%+	Currently on ventilator	5,411		
(Calculated)		(Calculated)		Definition			

The COVID Tracking Project - Data API

The COVID Tracking Project - Data API

Our Data About the Data Race and COVID-19 About Us Blog

Our Data

Data API

Latest API changes 0

Our public data API provides access to all of our data at a national and state level. We provide data in both JSON and CSV format. You can also [download CSV data directly](#).

All dates and times are in US eastern time (ET). Each state has its own set of caveats, which we have documented on our [data page](#).

Our data is provided under our [project-wide data license](#).

Fields may over time be marked as **deprecated**. These fields may remain static or change to null values after a period of two weeks.

New fields may be added without warning, but will not alter the structure of the data. If you are using CSV files, make sure to use the [column names](#) from the file to lookup data, not the [column numbers](#), which can change.

All URLs in our API, including state codes, should be in lower-case.

You can [view our current uptime and any known issues on our status page](#).

API domain name

All API requests should be made to: <https://api.covidtracking.com>

US Current and Historical Data

[Historic US values](#) ↓

[Current US values](#) ↓

[US historic values for a date](#) ↓

States Current and Historical Data

[State metadata](#) ↓

[Metadata about a specific state](#) ↓

[Historic values for all states](#) ↓

[Current values for all states](#) ↓

[Current values for a single state](#) ↓

[Historic values for a single state](#) ↓

[Values for a single state on a specific date](#) ↓

Current values for all states ↑

The most recent COVID data for every state. The current value may be different than today.

[View data source](#)

JSON format

`/v1/states/current.json`
Example: <https://api.covidtracking.com/v1/states/current.json>

CSV format

`/v1/states/current.csv`
Example: <https://api.covidtracking.com/v1/states/current.csv>

Fields

<code>checkTimeEt</code>	string	Deprecated
<code>commercialScore</code>	integer	Deprecated Returns null if no data is available

Coronavirus numbers by state

[README](#) States Current US current States daily 4 pm ET US daily 4 pm ET States

The COVID Tracking Project, launched from The Atlantic, obtains, organizes, and publishes high-quality data required to understand and respond to the COVID-19 outbreak in the United States. This is the spreadsheet version of the COVID Tracking Project dataset.

This public spreadsheet is not the place to automatically fetch data.
Please use our API at <https://covidtracking.com/api> or our CSV downloads at <https://covidtracking.com/data/download>

As of November 28, this document will no longer be updated.

Website <https://covidtracking.com>
About the project <https://covidtracking.com/about>
About the data <https://covidtracking.com/about-data>
API <https://covidtracking.com/api>
Twitter account [@COVID19Tracking](#)
The original Atlantic <https://www.theatlantic.com/health/archive/2020/03/how-many-americans-have-been-tested-coronavirus/607597/>
<https://www.theatlantic.com/health/archive/2020/03/coronavirus-testing-numbers/607714/>

The COVID Tracking Project

The COVID Tracking Project at The Atlantic

Our Data About the Data Race and COVID-19 About Us Blog

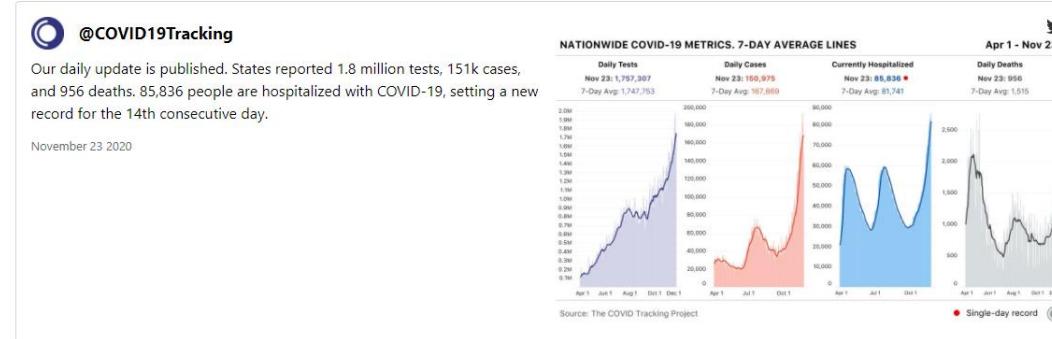
Search Volunteer

Our Data

Totals for the US Racial Data Dashboard Long-Term Care Our Charts Data API Download the Data

We update all our data each day between about 6pm and 7:30pm Eastern Time. We have recently begun publishing a bit later so as to capture [Colorado](#) and [New Mexico](#) data on the same day they update rather than on the following day.

[Sign up for our newsletter](#) to get our latest [articles about this data](#) by email.



Data for November 23, 2020

Get the data as: [CSV](#) [API](#)

Cases	Historical data	Tests	Historical data	Hospitalization	Historical data	Outcomes	Historical data
Total cases	12,230,472	Total tests	181,142,529	Currently hospitalized	85,836	Deaths	248,897
Definition		Definition		Definition		Definition	
New cases today	150,975	New tests today	1,757,307	Currently in ICU	16,811	We have removed recovered data for the US. Here's why.	
(Calculated)		(Calculated)		Definition			
Change over 7 days	10.7%+	Change over 7 days	7.3%+	Currently on ventilator	5,411		
(Calculated)		(Calculated)		Definition			

COVID-19 Public Data (v1) - Stitch

Stitch Docs

All Systems Operational

Getting started

Account management

Security

Destinations

Integrations

Extracting data

Loading data

Analysis tools

Troubleshooting

Developers

COVID-19 Public Data (v1)

- COVID-19 Public Data integration summary
- COVID-19 Public Data feature snapshot
- Connecting COVID-19 Public Data
- COVID-19 Public Data table reference

COVID-19 Public Data is supported by the Singer community
This integration is powered by Singer's COVID-19 tap. For support, visit the GitHub repo or join the Singer Slack.

COVID-19 Public Data integration summary

Stitch's COVID-19 Public Data integration was developed in a collaboration between [Bytecode](#) and [Talend](#). It replicates data from multiple public data sources using the [GitHub REST API v3](#). Refer to the [Schema](#) section for a list of objects available for replication.

COVID-19 Public Data feature snapshot

A high-level look at Stitch's COVID-19 Public Data (v1) integration, including release status, useful links, and the features supported in Stitch.

STITCH	
Release status	Released on April 2, 2020
Supported by	Singer Community
Stitch plan	Standard
API availability	Available
Singer GitHub repository	singer-io/tap-covid-19

master · 1 branch · 9 tags · Go to file · Add file · Code

This branch is 11 commits ahead, 2 commits behind tcgibennett:master. Pull request · Compare

dmosorast Version 0.0.10 and changelog · 5940583 on May 28 · 18 commits

.github	Add PR template	8 months ago
tap_covid_19	Fix Swiss daily files datetime format (#8)	6 months ago
.gitignore	v.0.0.3 New sources (#1)	8 months ago
CHANGELOG.md	Version 0.0.10 and changelog	6 months ago
LICENSE	initial commit	8 months ago
MANIFEST.in	initial commit	8 months ago
README.md	v.0.0.8 row_number, 403 error (#6)	8 months ago
config.json.example	Fix documentation	8 months ago
setup.py	Version 0.0.10 and changelog	6 months ago
state.json.example	v.0.0.6 fix replication strategy (#4)	8 months ago

README.md

tap-covid-19

This is a [Singer](#) tap that produces JSON-formatted data following the [Singer spec](#).

This tap:

- Pulls CSV files from [GitHub v3 API](#).
- Extracts the following resources:
 - CSV Data Files: [Git API Search](#) with [filename](#) and [extension filters](#) from the following COVID-19 Repositories; streaming in new/changed files:
 - Johns Hopkins CSSE Data
 - jh_csse_daily
 - EU Data
 - eu_daily
 - eu_ecdc_daily

COVID-19 Public Data (v1) - Stitch

Stitch Docs

All Systems Operational

- Getting started
- Account management
- Security
- Destinations
- Integrations
- Extracting data
- Loading data
- Analysis tools
- Troubleshooting
- Developers

Connecting COVID-19 Public Data

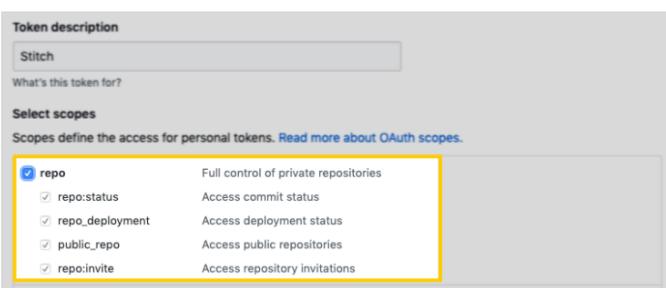
COVID-19 Public Data setup requirements

To set up COVID-19 Public Data in Stitch, you need:

- A regular (free) GitHub account. The GitHub repo for this integration is public - no special access is required.

Step 1: Create a GitHub personal access token

- Sign into your GitHub account.
- Click the User menu (your icon) > Settings.
- Click Developer settings in the navigation on the left side of the page.
- Click Personal access tokens.
- On the Personal access tokens page, click the Generate new token button. If prompted, enter your password.
- In the Description field, enter `stitch`. This will allow you to easily identify what application is using the token.
- In the Select Scopes section, check the `repo` option:



Note: While these are full permissions, Stitch will only ever read your data. The `repo` scope is required due to how GitHub structures permissions.

- Click the Generate token button.
- The new access token will display on the next page. Copy the token before navigating away from the page - GitHub won't display it again.

Stitch Docs

All Systems Operational

- Getting started
- Account management
- Security
- Destinations
- Integrations
- Extracting data
- Loading data
- Analysis tools
- Troubleshooting
- Developers

COVID-19 Public Data table reference

Schemas and versioning

Schemas and naming conventions can change from version to version, so we recommend verifying your integration's version before continuing.

The schema and info displayed below is for **version 1** of this integration.

This is the latest version of the COVID-19 Public Data integration.

Table and column names in your destination

Depending on your destination, table and column names may not appear as they are outlined below.

For example: Object names are lowercased in Redshift (`CuStoMERS` > `customers`), while case is maintained in PostgreSQL destinations (`CuStoMERS` > `CuStoMERS`). Refer to the [Loading Guide](#) for your destination for more info.

c19_trk_us_daily

[View table schema on GitHub](#)

Replication Method: Key-based Incremental Replication Key: `git_last_modified`

Primary Key: `_sdc_row_number` Official docs: [Official Docs](#)

The `c19_trk_us_daily` table contains statistics for the United States, aggregated by day.

Note: The source file for this table is a single file that updates on a daily basis. When Stitch replicates this table, it will replicate the entire contents of the file, but only if the file has been modified since the integration's last replication job.

c19_trk_us_population_counties

[View table schema on GitHub](#)

Replication Method: Key-based Incremental Replication Key: `git_last_modified`

Primary Key: `_sdc_row_number` Official docs: [Official Docs](#)

Talend Pipeline Manager

The screenshot displays the Talend Pipeline Manager interface. At the top, a banner greets the user "Welcome richard!" and informs them that their trial expires in 7 days, with a "SUBSCRIBE NOW" button. Below the banner, the "Ingest streaming and cloud data into pipelines" section features a "TALEND PIPELINE DESIGNER" tool. This tool has three steps: 1. CREATE CONNECTIONS, 2. RETRIEVE DATASETS, 3. DESIGN PIPELINES. It includes icons for AWS, Google Cloud, and various data sources, along with "Learn more", "Watch tutorials", and "Launch" buttons. The "Build data integration" section shows the "TALEND MANAGEMENT CONSOLE" with three steps: 1. DESIGN, 2. SCHEDULE, 3. DEPLOY. It features a compass icon and similar "Learn more", "Watch tutorials", and "Launch" buttons.

Talend Stitch

STITCH DATA LOADER [Integrations](#) [Destination](#) [Notifications](#)

Your free trial expires in 8 days. [Buy Now](#)

All your integrations are paused. Data won't replicate to your destination unless you have an active integration.

Find an integration [Add Integration](#)

31,961,186
Total rows replicated this billing period
For details, see your [billing page](#).

Name ▾	Status	Last Sync Completed	Rows This Billing Period
covid_7374_covid	Paused		21,287
covid_public	Paused		0
covid_publicdata	Paused		0
covid2	Paused		0
hubspot_test	Paused		384
info7370_sql	Paused		31,939,515
sheets1	Paused		0
sheets2	Paused		0
SQL_local	Paused		0

Talend Stitch – initially setup for BigQuery

STITCH DATA LOADER Integrations **Destination** Notifications

Your free trial expires in 8 days. [Buy Now](#)

Current Destination

 Google BigQuery
[Test your connection](#)

Connection Settings
Check our documentation for help with your destination settings

Your Key File [Upload new key](#) BigQuery Project Name carbon-airlock-294817

Google Cloud Storage Location

Loading Behavior
This destination is set to **upsert**. Rows in your destination are **updated** as changes in your source data occur.

[Update Google BigQuery Settings](#)

Post-load hooks ENTERPRISE FEATURE
When Stitch successfully loads data into your destination, send a summary webhook to the URL you provide.
Configure post-load hooks in your account's [Notification Settings](#).

Change Destination
Need to change (or delete) your current destination? We'll walk you through the process. You'll be asked to confirm this on the next screen.
We currently support [Amazon Redshift](#), [Google BigQuery](#), [Postgres](#), [Snowflake](#), [S3](#), and more.

[Change Destination](#)

Talend Stitch – initially setup for BigQuery

The screenshot shows the Google Cloud Platform BigQuery interface. On the left, there is a sidebar with navigation links: Query history, Saved queries, Job history, Transfers, Scheduled queries, Reservations, BI Engine, Resources (with a '+ ADD DATA' button), and a search bar for tables and datasets. The main area is titled 'Unsaved query' and contains a single line of SQL: '1: SELECT * FROM `carbon-airlock-294817.covid_publicdata.c19_trk_us_states_current` LIMIT 1000'. Below the query is a validation message: 'Valid.' followed by a green checkmark icon. There are buttons for 'Run', 'Save query', 'Save view', 'Schedule query', and 'More'. To the right of the query, it says 'This query will process 29.5 KB when run.' Below the query editor is a table titled 'c19_trk_us_states_current' with three tabs: Schema, Details, and Preview. The Schema tab is selected, showing a list of fields with their types and modes:

Field name	Type	Mode	Policy tags	Description
total	INTEGER	NULLABLE		
git_last_modified	TIMESTAMP	NULLABLE		
score	INTEGER	NULLABLE		
total_test_results	INTEGER	NULLABLE		
positive	INTEGER	NULLABLE		
hospitalized	INTEGER	NULLABLE		
git_sha	STRING	NULLABLE		
grade	STRING	NULLABLE		
_sdc_row_number	INTEGER	NULLABLE		
positive_score	INTEGER	NULLABLE		
negative	INTEGER	NULLABLE		
date_modified	TIMESTAMP	NULLABLE		
git_owner	STRING	NULLABLE		
_sdc_table_version	INTEGER	NULLABLE		
git_repository	STRING	NULLABLE		
git_url	STRING	NULLABLE		
death	INTEGER	NULLABLE		
git_html_url	STRING	NULLABLE		
_sdc_received_at	TIMESTAMP	NULLABLE		
git_path	STRING	NULLABLE		
_sdc_sequence	INTEGER	NULLABLE		

Below the schema, there are buttons for 'QUERY TABLE', 'SHARE TABLE', 'COPY TABLE', 'DELETE TABLE', and 'EXPORT'.

Talend Stitch – initially setup for BigQuery

The screenshot shows the Google Cloud Platform BigQuery interface. On the left, there's a sidebar with navigation links like 'Query history', 'Saved queries', 'Job history', 'Transfers', 'Scheduled queries', 'Reservations', 'BI Engine', and 'Resources'. Below this is a search bar for tables and datasets. The main area is divided into 'Query editor' and 'Query results'.

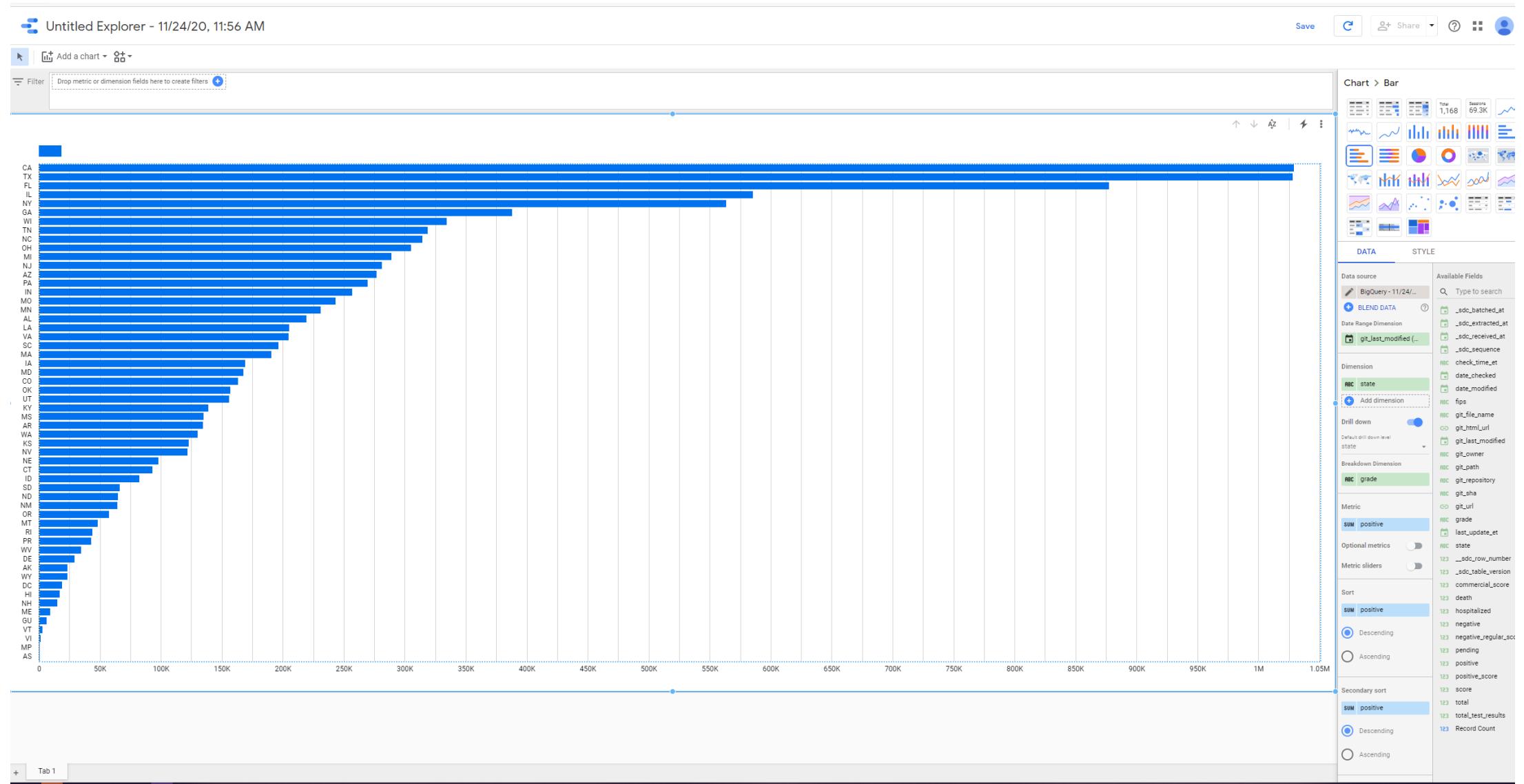
Query editor: Contains a code editor with the following SQL query:

```
1 SELECT * FROM `carbon-airlock-294817.covid_publicdata.c19_trk_us_states_current` LIMIT 1000
```

The status bar indicates the query is valid. Below the editor are buttons for 'Run', 'Save query', 'Schedule query', and 'More'. A note says 'This query will process 29.5 KB when run.'

Query results: Shows a table with 21 rows of data from the 'covid_publicdata.c19_trk_us_states_current' table. The columns are: Row, total, git_last_modified, score, total_test_results, positive, hospitalized, git_sha, grade, _sdc_row_number, positive_score, negative, date_modified, git_owner, _sdc_table_version, git_repository, and git_url. The data includes various COVID-19 tracking metrics for US states, such as positive cases, hospitalizations, and testing results, along with their corresponding GitHub repository URLs.

Talend Stitch – initially setup for BigQuery



Google BigQuery API Client Libraries

The screenshot shows the Google Cloud BigQuery API Client Libraries documentation page. The navigation bar includes links for Google Cloud, Why Google, Solutions, Products, Pricing, Getting Started, BigQuery, Overview, Guides, Reference (which is underlined), Samples, Support, and Resources. The left sidebar has a tree view of API references, with 'BigQuery client libraries' expanded and 'Java' selected. The main content area displays the 'BigQuery API Client Libraries' page, which provides instructions for getting started with the Java client library. It includes sections for 'Installing the client library', Java code snippets for Maven and Gradle, and SBT dependencies. A right sidebar contains a 'Table of contents' with links to 'Installing the client library', 'Setting up authentication', 'Using the client library', 'Additional resources', 'Third-party BigQuery API client libraries', and 'What's next?'. The bottom of the page features a 'Rate and review' button and social sharing icons.

BigQuery > Documentation > Reference

BigQuery API Client Libraries

This page shows how to get started with the Cloud Client Libraries for the BigQuery API. Read more about the client libraries for Cloud APIs, including the older Google APIs Client Libraries, in [Client Libraries Explained](#).

Installing the client library

C# Go Java Node.js PHP Python Ruby

For more information, see [Setting Up a Java Development Environment](#).

If you are using Maven, add the following to your `pom.xml` file. For more information about BOMs, see [The Google Cloud Platform Libraries BOM](#).

```
<!-- Using libraries-bom to manage versions.  
See https://github.com/GoogleCloudPlatform/cloud-opensource-java/wiki/The-Google-Cloud-Platform-Lib  
<dependencyManagement>  
  <dependencies>  
    <dependency>  
      <groupId>com.google.cloud</groupId>  
      <artifactId>libraries-bom</artifactId>  
      <version>16.1.0</version>  
      <type>pom</type>  
      <scope>import</scope>  
    </dependency>  
  </dependencies>  
</dependencyManagement>  
  
<dependencies>  
  <dependency>  
    <groupId>com.google.cloud</groupId>  
    <artifactId>google-cloud-bigquery</artifactId>  
  </dependency>  
</dependencies>
```

If you are using Gradle, add the following to your dependencies:

```
compile 'com.google.cloud:google-cloud-bigquery:1.125.0'
```

If you are using sbt, add the following to your dependencies:

```
libraryDependencies += "com.google.cloud" % "google-cloud-bigquery" % "1.125.0"
```

GCP NYT US Coronavirus Database

Google Cloud Platform Select a project ▾

The New York Times US Coronavirus Database

The New York Times

Data based on reports from state and local health agencies

[VIEW DATASET](#)

[OVERVIEW](#) [SAMPLES](#)

Overview

This is the US Coronavirus data repository from [The New York Times](#). This data includes COVID-19 cases and deaths reported by state and county. The New York Times compiled this data based on reports from state and local health agencies. More information on the data repository is available [here](#). For additional reporting and data visualizations, see The New York Times' [U.S. coronavirus interactive site](#).

This public dataset is hosted in Google BigQuery and is included in BigQuery's 1TB/mo of free tier processing. This means that each user receives 1TB of free BigQuery processing every month, which can be used to run queries on this public dataset. Watch this short video to learn how to get started quickly using BigQuery to access public datasets. [What is BigQuery](#)

This dataset has significant public interest in light of the COVID-19 crisis. All bytes processed in queries against this dataset will be zeroed out, making this part of the query free. Data joined with the dataset will be billed at the normal rate to prevent abuse. After September 15, queries over these datasets will revert to the normal billing rate.

Users of The New York Times public-use data files must comply with data use restrictions to ensure that the information will be used solely for noncommercial purposes.

Additional details

Type: [Datasets](#)
Last updated: 5/13/20
Category: [Datasets for COVID-19 research](#), [Science & research](#)
Dataset source: [NYT COVID-19 US Dataset](#)
Cloud service: BigQuery
Region: US/EU
Update frequency: Daily

Google Cloud Platform My First Project

BigQuery [FEATURES & INFO](#) [SHORTCUT](#)

Query history
Saved queries
Job history
Transfers
Scheduled queries
Reservations
BI Engine
Resources [+ ADD DATA](#)

Search for your tables and datasets

- covid19_nyt
- covid19_italy
- covid19_italy_eu
- covid19_jhu_csse
- covid19_jhu_csse_eu
- covid19_nyt
- excess_deaths
- mask_use_by_county
- us_counties
- us_states
- covid19_open_data
- covid19_open_data_eu

[Run](#) [Save query](#) [Save view](#) [Schedule query](#) [More](#)

bigquery-public-data:covid19_nyt

Description [Edit](#)
None

Dataset info [Edit](#)

Dataset ID	bigquery-public-data:covid19_nyt
Created	Apr 9, 2020, 6:54:26 PM
Default table expiration	Never
Last modified	Apr 9, 2020, 7:13:23 PM
Data location	US

GCP NYT US Coronavirus Database

Google Cloud Platform Select a project ▾

The New York Times US Coronavirus Database

The New York Times

Data based on reports from state and local health agencies

[VIEW DATASET](#)

[OVERVIEW](#) [SAMPLES](#)

Overview

This is the US Coronavirus data repository from [The New York Times](#). This data includes COVID-19 cases and deaths reported by state and county. The New York Times compiled this data based on reports from state and local health agencies. More information on the data repository is available [here](#). For additional reporting and data visualizations, see The New York Times' [U.S. coronavirus interactive site](#).

This public dataset is hosted in Google BigQuery and is included in BigQuery's 1TB/mo of free tier processing. This means that each user receives 1TB of free BigQuery processing every month, which can be used to run queries on this public dataset. Watch this short video to learn how to get started quickly using BigQuery to access public datasets. [What is BigQuery](#)

This dataset has significant public interest in light of the COVID-19 crisis. All bytes processed in queries against this dataset will be zeroed out, making this part of the query free. Data joined with the dataset will be billed at the normal rate to prevent abuse. After September 15, queries over these datasets will revert to the normal billing rate.

Users of The New York Times public-use data files must comply with data use restrictions to ensure that the information will be used solely for noncommercial purposes.

Additional details

Type: [Datasets](#)
Last updated: 5/13/20
Category: [Datasets for COVID-19 research](#), [Science & research](#)
Dataset source: [NYT COVID-19 US Dataset](#)
Cloud service: BigQuery
Region: US/EU
Update frequency: Daily

Google Cloud Platform My First Project

BigQuery [FEATURES & INFO](#) [SHORTCUT](#)

Query history
Saved queries
Job history
Transfers
Scheduled queries
Reservations
BI Engine
Resources [+ ADD DATA](#)

Search for your tables and datasets

- covid19_nyt
- covid19_italy
- covid19_italy_eu
- covid19_jhu_csse
- covid19_jhu_csse_eu
- covid19_nyt
- excess_deaths
- mask_use_by_county
- us_counties
- us_states
- covid19_open_data
- covid19_open_data_eu

[Run](#) [Save query](#) [Save view](#) [Schedule query](#) [More](#)

bigquery-public-data:covid19_nyt

Description [edit](#)
None

Dataset info [edit](#)

Dataset ID	bigquery-public-data:covid19_nyt
Created	Apr 9, 2020, 6:54:26 PM
Default table expiration	Never
Last modified	Apr 9, 2020, 7:13:23 PM
Data location	US

NYT US Coronavirus Database

nytimes / covid-19-data

Code Issues 74 Pull requests Projects 1 Wiki Security Insights

master 1 branch 0 tags Go to file Add file Code

nyt-covid-19-bot	Updating data.	b3e3280 1 hour ago	1,426 commits
.github/ISSUE_TEMPLATE	New data for 6/26.	5 months ago	
colleges	11/19 update of colleges and universities case data	12 days ago	
excess-deaths	Update excess deaths data with 4 additional countries and more recent...	2 months ago	
live	Updating data.	1 hour ago	
mask-use	*NEW DATASET*: Estimates of mask-usage by county from a nationwide ...	4 months ago	
.gitignore	excess deaths	7 months ago	
LICENSE	Initial public release.	8 months ago	
NEW-YORK-DEATHS-METHODOLOGY...	Add special methodology note explaining New York deaths.	4 months ago	
PROBABLE-CASES-NOTE.md	Adding note on our methodology change to include probable cases.	7 months ago	
README.md	Change Puerto Rico municipio deaths 0s to blanks. A few other data re...	3 days ago	
us-counties.csv	Updating data.	1 hour ago	
us-states.csv	Updating data.	1 hour ago	
us.csv	Updating data.	1 hour ago	

README.md

About

An ongoing repository of data on coronavirus cases and deaths in the U.S.

www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html

covid-19

Readme

View license

Releases

No releases published

Packages

No packages published

Contributors 8



NYT US Coronavirus Database

README.md

Coronavirus (Covid-19) Data in the United States

NEW: We are publishing the data behind our [survey of mask usage](#) in the United States in order to provide researchers a way to understand the role of mask wearing in the course of the pandemic. See the data and documentation in the [mask-use/](#) directory.

NEW: We are publishing the data behind our [excess deaths tracker](#) in order to provide researchers and the public with a better record of the true toll of the pandemic. This data is compiled from official national and municipal data for 24 countries. See the data and documentation in the [excess-deaths/](#) directory.

[[U.S. Data \(Raw CSV\)](#) | [U.S. State-Level Data \(Raw CSV\)](#) | [U.S. County-Level Data \(Raw CSV\)](#)]

The New York Times is releasing a series of data files with cumulative counts of coronavirus cases in the United States, at the state and county level, over time. We are compiling this time series data from state and local governments and health departments in an attempt to provide a complete record of the ongoing outbreak.

Since late January, The Times has tracked cases of coronavirus in real time as they were identified after testing. Because of the widespread shortage of testing, however, the data is necessarily limited in the picture it presents of the outbreak.

We have used this data to power our [maps](#) and [reporting](#) tracking the outbreak, and it is now being made available to the public in response to requests from researchers, scientists and government officials who would like access to the data to better understand the outbreak.

The data begins with the first reported coronavirus case in Washington State on Jan. 21, 2020. We will publish regular updates to the data in this repository.

Live and Historical Data

We are providing two sets of data with cumulative counts of coronavirus cases and deaths: one with our most current numbers for each geography and another with historical data showing the tally for each day for each geography.

The historical data files are at the top level of the directory and contain data up to, but not including the current day. The live data files are in the [live/](#) directory.

A key difference between the historical and live files is that the numbers in the historical files are the final counts at the end of each day, while the live files have figures that may be a partial count released during the day but cannot necessarily be considered the final, end-of-day tally.

The historical and live data are released in three files, one for each of these geographic levels: U.S., states and counties.

Each row of data reports the cumulative number of coronavirus cases and deaths based on our best reporting up to the moment we publish an update. Our counts include both laboratory confirmed and probable cases using [criteria](#) that were developed by states and the federal government. Not all geographies are reporting probable cases and yet others are providing confirmed and probable as a single total. Please [read here](#) for a full discussion of this issue.

We do our best to revise earlier entries in the data when we receive new information. If a county is not listed for a date, then there were zero reported confirmed cases and deaths.

State and county files contain [FIPS codes](#), a standard geographic identifier, to make it easier for an analyst to combine this data with other data sets like a map file or population data.

Historical Data

U.S. National-Level Data

The daily number of cases and deaths nationwide, including states, U.S. territories and the District of Columbia, can be found in the [us.csv](#) file. ([Raw CSV file here](#).)

```
date,cases,deaths  
2020-01-21,1,0
```

...

State-Level Data

State-level data can be found in the [states.csv](#) file. ([Raw CSV file here](#).)

```
date,state,fips,cases,deaths  
2020-01-21,Washington,53,1,0
```

...

County-Level Data

County-level data can be found in the [counties.csv](#) file. ([Raw CSV file here](#).)

```
date,county,state,fips,cases,deaths  
2020-01-21,Snohomish,Washington,53061,1,0
```

...

In some cases, the geographies where cases are reported do not map to standard county boundaries. See the list of [geographic exceptions](#) for more detail on these.

NYT US Coronavirus Database

Live Data

The files in the [live/](#) directory are also available at three geographic levels and contain all the fields the historical data files have, but with only data for the current day. We try to update these files multiple times per day.

Because these are updated throughout the day, they can have inconsistent counts, are more likely to contain errors, and should be considered less reliable than the historical data. Different areas of the country update at different times and our data collection process can move at a different pace as well.

In addition to the columns that are in the historical files, these files also include new columns that include detail on the number of confirmed and probable cases, separately.

In the live files, the case and death fields have the following definitions:

- **cases:** The total number of cases of Covid-19, including both confirmed and probable.
- **deaths:** The total number of deaths from Covid-19, including both confirmed and probable.
- **confirmed_cases:** The number of laboratory confirmed Covid-19 cases only, or blank if not available.
- **confirmed_deaths:** The number of laboratory confirmed Covid-19 deaths only, or blank if not available.
- **probable_cases:** The number of probable Covid-19 cases only, or blank if not available.
- **probable_deaths:** The number of probable Covid-19 deaths only, or blank if not available.

We understand this breakout would also be valuable historically, and are working toward providing that. Please bear with us as we roll out this new and more complicated data.

The live data can be found in files at the U.S. level in the [us.csv](#) file, at the state level in the [states.csv](#) file, and at the county level in the [counties.csv](#) file.

Methodology and Definitions

The data is the product of dozens of journalists working across several time zones to monitor news conferences, analyze data releases and seek clarification from public officials on how they categorize cases.

It is also a response to a fragmented American public health system in which overwhelmed public servants at the state, county and territorial level have sometimes struggled to report information accurately, consistently and speedily. On several occasions, officials have corrected information hours or days after first reporting it. At times, cases have disappeared from a local government database, or officials have moved a patient first identified in one state or county to another, often with no explanation. In those instances, which have become more common as the number of cases has grown, our team has made every effort to update the data to reflect the most current, accurate information while ensuring that every known case is counted.

When the information is available, we count patients where they are being treated, not necessarily where they live.

In most instances, the process of recording cases has been straightforward. But because of the patchwork of reporting methods for this data across more than 50 state and territorial governments and hundreds of local health departments, our journalists sometimes had to make difficult interpretations about how to count and record cases.

For those reasons, our data will in some cases not exactly match with the information reported by states and counties. Those differences include these cases: When the federal government arranged flights to the United States for Americans exposed to the coronavirus in China and Japan, our team recorded those cases in the states where the patients subsequently were treated, even though local health departments generally did not. When a resident of Florida died in Los Angeles, we recorded her death as having occurred in California rather than Florida, though officials in Florida counted her case in their own records. And when officials in some states reported new cases without immediately identifying where the patients were being treated, we attempted to add information about their locations later, once it became available.

- "Probable" and "Confirmed Cases and Deaths"

Cases and deaths can be reported as either "confirmed" or "probable." Our total cases and deaths include both. The number of cases includes all cases, including those who have since recovered or died.

On April 5, the Council of State and Territorial Epidemiologists [advised states](#) to include both confirmed cases, based on confirmatory laboratory testing, and probable cases, based on specific criteria for testing, symptoms and exposure. The Centers for Disease Control adopted these definitions and national CDC data began including confirmed and probable cases on April 14.

Some governments continue to report only confirmed cases, while others are reporting both confirmed and probable numbers. And there is also another set of governments that is reporting the two types of numbers combined without providing a way to separate the confirmed from the probable.

The Geographic Exceptions section below has more details on specific areas. The methodology of individual states changes frequently.

NYT US Coronavirus Database

WORLD COUNTRIES ▾ | U.S.A. STATES ▾ COLLEGES

Coronavirus in the U.S.: Latest Map and Case Count

By The New York Times Updated December 1, 2020, 7:44 A.M. E.T.

Leer en español



TOTAL REPORTED	ON NOV. 30	14-DAY CHANGE
Cases 13.6 million+	167,759	+3% →
Deaths 268,023	1,265	+28% ↗
Hospitalized	96,039	+34% →

Day with data reporting anomaly.
Hospitalization data from the Covid Tracking Project; 14-day change trends
use 7-day averages.

Map Cases by state Hot spots Clusters

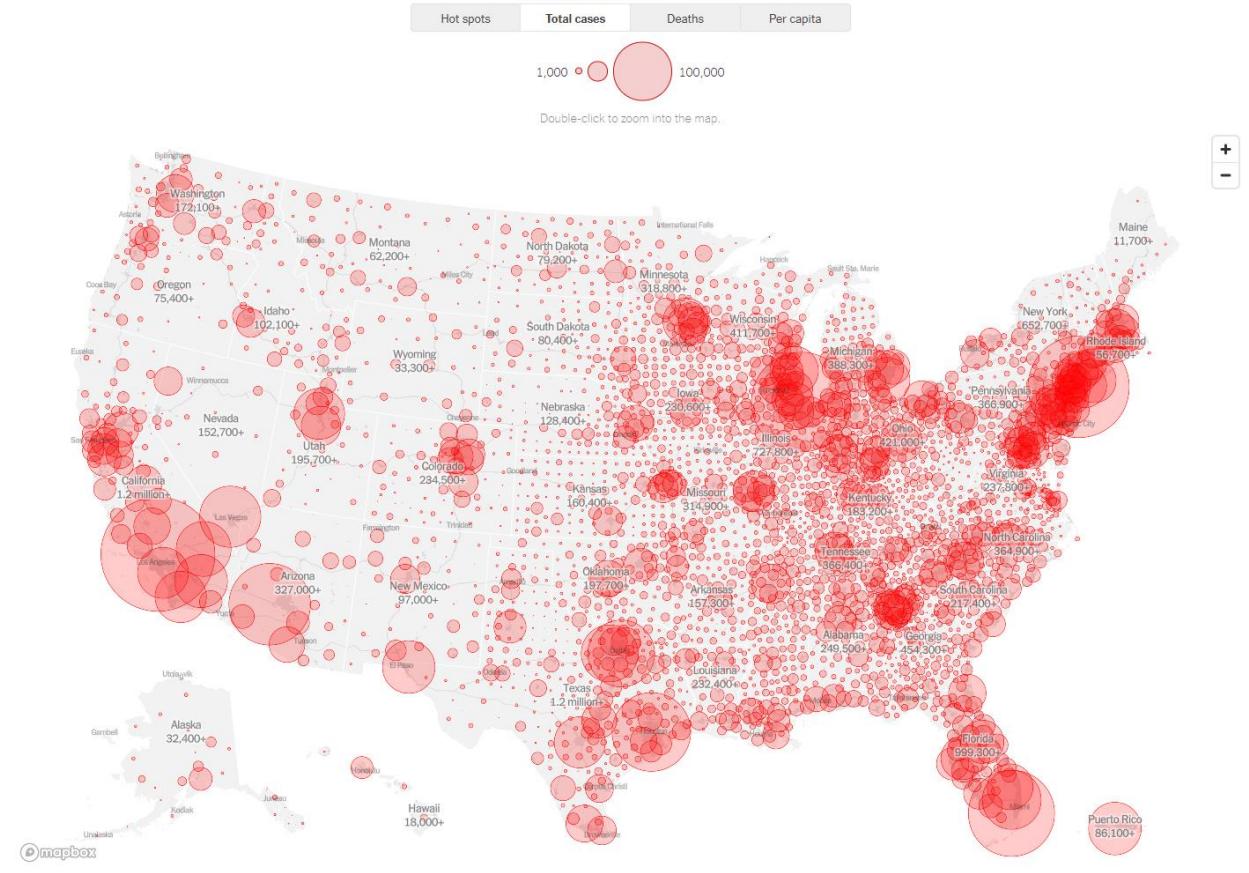
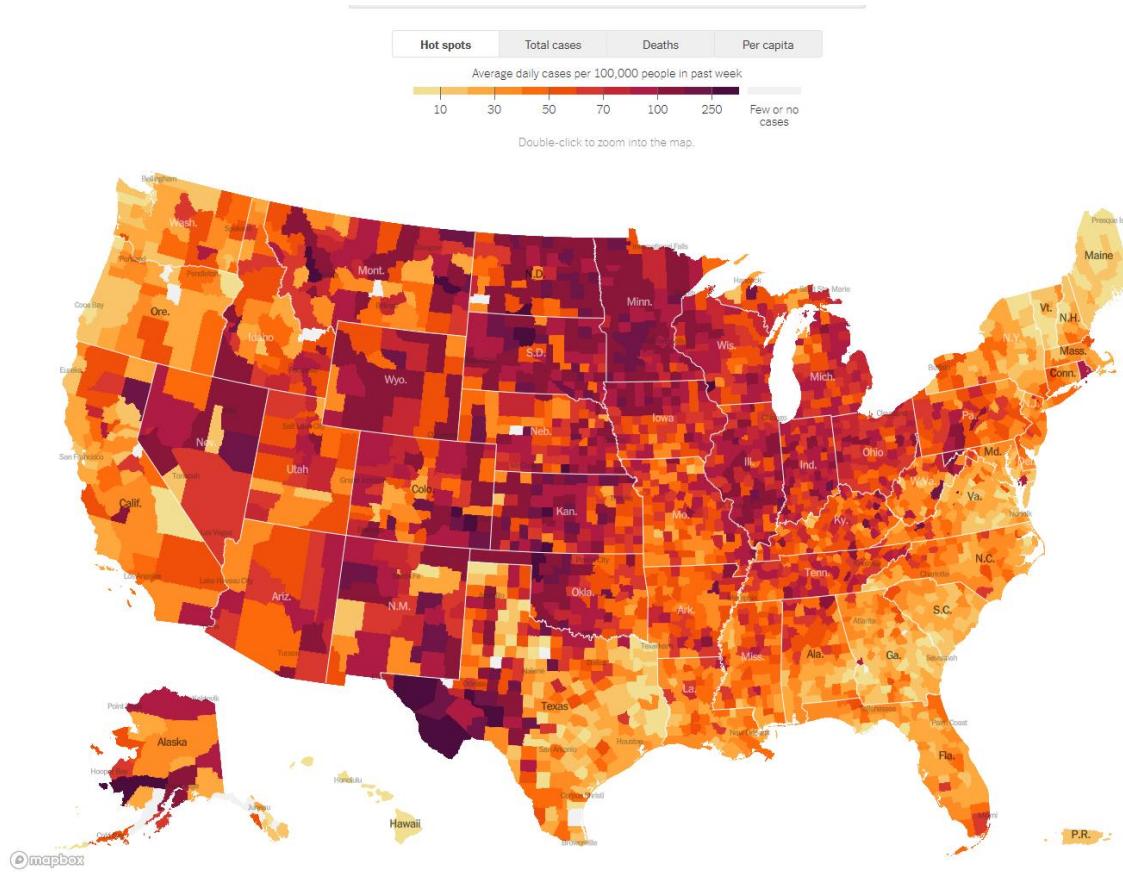
At least 1,265 new coronavirus deaths and 167,759 new cases were reported in the United States on Nov. 30. Over the past week, there has been an average of 160,387 cases per day, an increase of 3 percent from the average two weeks earlier.

Cases and deaths by state and county

This table is sorted by places with the most cases per 100,000 residents in the last seven days. Charts are colored to reveal when outbreaks emerged.

	Cases	Deaths	Search counties	WEEKLY CASES PER CAPITA		
	TOTAL CASES	PER 100,000	DAILY AVG. IN LAST 7 DAYS	▼ PER 100,000	FEWER	MORE
+ North Dakota MAP »	79,258	10,400	836.4	109.8		March 1 Nov. 30
+ Minnesota MAP »	318,822	5,653	6,038	107.1		
+ South Dakota MAP »	80,464	9,095	945.1	106.8		
+ Wyoming MAP »	33,305	5,755	553.4	95.6		
+ Nebraska MAP »	128,407	6,638	1,783.7	92.2		
+ New Mexico MAP »	97,095	4,631	1,849.6	88.2		
+ Alaska MAP »	32,407	4,430	611.3	83.6		
- Rhode Island MAP »	56,723	5,354	878.6	82.9		
Providence	35,980	5,631	638.3	99.9		
Kent	5,180	3,153	146.9	89.4		
Bristol	1,244	2,566	28.6	58.9		
Washington	2,223	1,770	47.4	37.8		
Newport	1,217	1,483	24.4	29.8		
Unknown	10,879	—	876.2	—		
+ Indiana MAP »	341,694	5,076	5,445.6	80.9		
+ Wisconsin MAP »	411,730	7,071	4,576.7	78.6		
Show all						

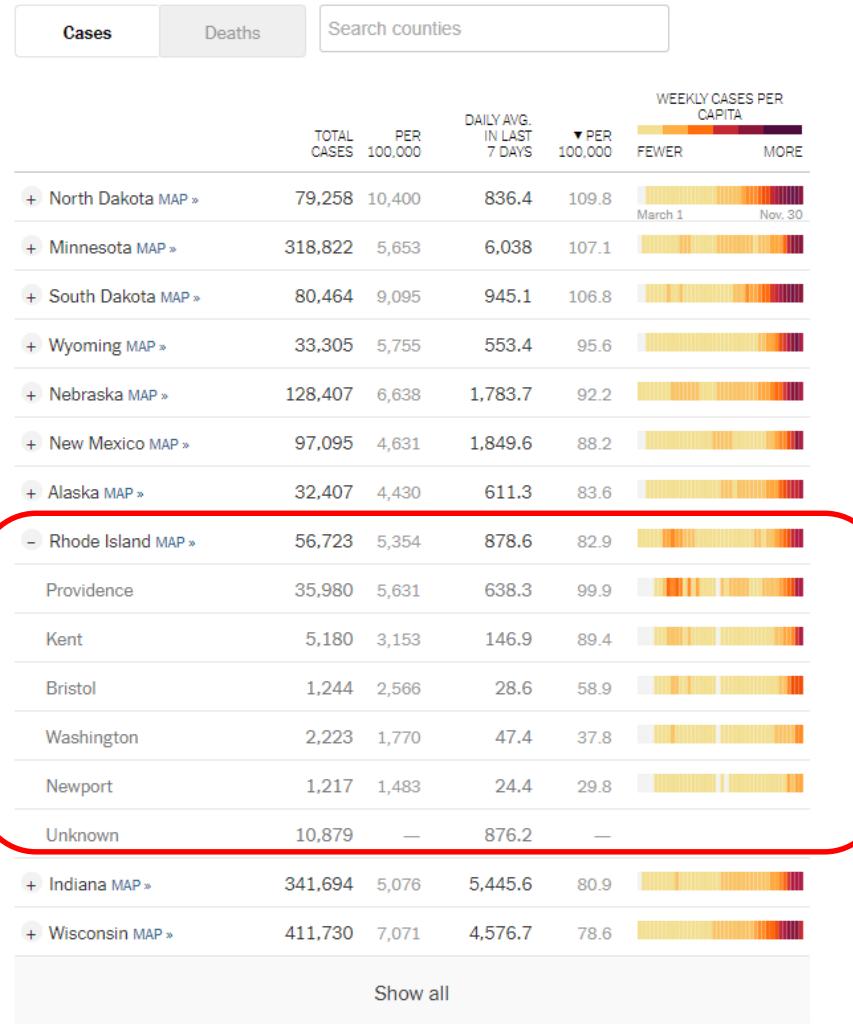
NYT US Coronavirus Database



NYT US Coronavirus Database

Cases and deaths by state and county

This table is sorted by places with the most cases per 100,000 residents in the last seven days. Charts are colored to reveal when outbreaks emerged.



Google Cloud Platform My First Project

BigQuery FEATURES & INFO SHORTCUT

Query history Query editor

```
1 SELECT * FROM `bigquery-public-data.covid19_nyt.us_counties`  
2 WHERE state_name = 'Rhode Island' AND county NOT IN ('Unknown')  
3 -- LIMIT 1000
```

Saved queries Job history Transfers Scheduled queries Reservations BI Engine

Resources + ADD DATA

Search for your tables and datasets

- covid19_italy_eu
- covid19_jhu_csse
- covid19_jhu_csse_eu
- covid19_nyt
 - excess_deaths
 - mask_use_by_county
 - us_counties
- us_states
- covid19_open_data
- covid19_open_data_eu
- covid19_public_forecasts
- covid19_public_forecasts_asian
- covid19_rxrx19

Run Save query Save view Schedule query More

Query results SAVE RESULTS EXPLORE DATA

Query complete (0.7 sec elapsed, 37.5 MB processed)

Row	Date	County	State Name	County FIPS Code	Confirmed Cases	Deaths
1	2020-08-04	Bristol	Rhode Island	44001	304	15
2	2020-08-05	Bristol	Rhode Island	44001	316	15
3	2020-08-06	Bristol	Rhode Island	44001	316	15
4	2020-08-07	Bristol	Rhode Island	44001	316	15
5	2020-08-08	Bristol	Rhode Island	44001	316	15
6	2020-08-09	Bristol	Rhode Island	44001	316	15
7	2020-08-10	Bristol	Rhode Island	44001	316	15
8	2020-08-11	Bristol	Rhode Island	44001	316	15

NYT US Coronavirus Database

Google Cloud Platform • My First Project ▾

BigQuery FEATURES & INFO SHORTCUT

Query history

Saved queries

Job history

Transfers

Scheduled queries

Reservations

BI Engine

Resources + ADD DATA ▾

Search for your tables and datasets

- covid19_italy_eu
- covid19_jhu_csse
- covid19_jhu_csse_eu
- covid19_nyt
 - excess_deaths
 - mask_use_by_county
 - us_counties
 - us_states
- covid19_open_data
- covid19_open_data_eu
- covid19_public_forecasts
- covid19_public_forecasts_asia...
- covid19_rxrx19

Query editor

```
1 SELECT * FROM `bigquery-public-data.covid19_nyt.us_counties`  
2 WHERE state_name = 'Rhode Island' AND county >< 'Unknown'  
3 -- LIMIT 1000
```

Run Save query Save view Schedule query More

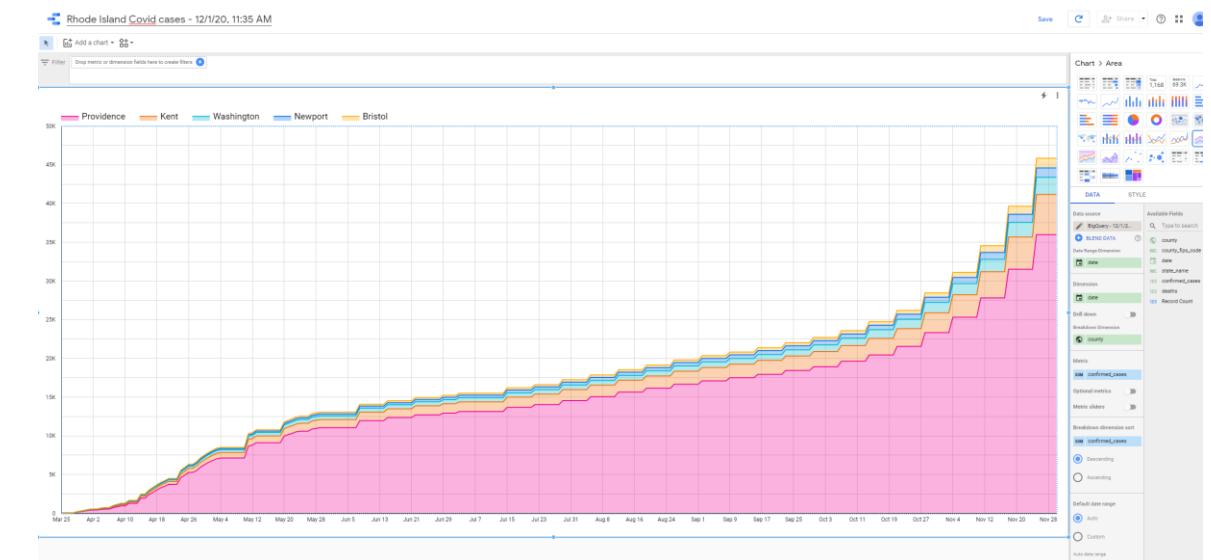
Query results SAVE RESULTS EXPLORE DATA

Query complete (0.7 sec elapsed, 37.5 MB processed)

Job information Results JSON Execution details

Row	date	county	state_name	county_fips_code	confirmed_cases	deaths
1	2020-08-04	Bristol	Rhode Island	44001	304	15
2	2020-08-05	Bristol	Rhode Island	44001	316	15
3	2020-08-06	Bristol	Rhode Island	44001	316	15
4	2020-08-07	Bristol	Rhode Island	44001	316	15
5	2020-08-08	Bristol	Rhode Island	44001	316	15
6	2020-08-09	Bristol	Rhode Island	44001	316	15
7	2020-08-10	Bristol	Rhode Island	44001	316	15
8	2020-08-11	Bristol	Rhode Island	44001	316	15

info7374_2020_Fall\COVID
Project\sample_Rhode_Island_results-20201201-110122.csv



Rhode Island Covid cases - 12/1/20, 11:35 AM

Add a chart ▾

Filter date: Nov 1, 2020 - Nov 30, 2020

county	deaths	confirmed_cases
1. Providence	29,777	878,243
2. Kent	3,886	111,774
3. Washington	2,371	51,860
4. Bristol	744	27,037
5. Newport	240	28,579

COVID Open Data

Google Cloud Platform Select a project ▾

COVID-19 Open Data

BigQuery Public Datasets Program

Daily time-series data related to COVID-19 globally

[VIEW DATASET](#)

[OVERVIEW](#) [SAMPLES](#)

Overview

This repository contains country-level datasets of daily time-series data related to COVID-19 globally. It includes data relating to demographics, economy, epidemiology, geography, health, hospitalizations, mobility, government response, and weather.

This dataset is available in both the US and EU regions of BigQuery at the following links:

- [COVID-19 Open Data: US Region](#)
- [COVID-19 Open Data: EU Region](#)

All data in this dataset is retrieved automatically. When possible, data is retrieved directly from the relevant authorities, like a country's ministry of health.

This dataset has significant public interest in light of the COVID-19 crisis. All bytes processed in queries against this dataset will be zeroed out, making this part of the query free. Data joined with the dataset will be billed at the normal rate to prevent abuse. After September 15, queries over these datasets will revert to the normal billing rate.

This public dataset is hosted in Google BigQuery and is included in BigQuery's 1TB/mo of free tier processing. This means that each user receives 1TB of free BigQuery processing every month, which can be used to run queries on this public dataset. Watch this short video to learn how to get started quickly using BigQuery to access public datasets. [What is BigQuery](#)

Additional details

Type: [Datasets](#)

Last updated: 7/9/20

Category: [Datasets for COVID-19 research](#), [Science & research](#)

Dataset source: <https://github.com/open-covid-19/data>

Cloud service: BigQuery

Region: US/EU

Update frequency: Daily

Google Cloud Platform Select a project ▾

Navigation menu COVID-19 Open Data

BigQuery Public Datasets Program

[OVERVIEW](#) [SAMPLES](#)

Samples

Dataset Migration

If you are migrating from any of the other COVID-19 Public Datasets, such as the [New York Times "US Coronavirus Database"](#), [European Centre for Disease Prevention and Control's "COVID-19 Cases by Country"](#) or other datasets with data on confirmed cases and deaths either within the US or around the globe, the following list should help you remap your queries to the structure of this dataset. For an example of this, check out the [compatibility view](#) created in the US region copy of the dataset or the [compatibility view](#) created in the EU region copy.

The structure of this list is as follows:
Previous dataset column name: COVID-19 Open Data column name

- province_state : subregion1.name
- country_region : country_name
- date : date
- latitude: latitude
- longitude : longitude
- location_geom : location_geometry
- confirmed : cumulative_confirmed
- deaths : cumulative_deceased
- recovered : cumulative_recovered
- active : current_cases (NOTE: Calculated as [cumulative_confirmed-cumulative_recovered-cumulative_deceased] where none of these values are NULL)
- fips : subregion2_code
- admin_2 : subregion2.name
- combined_key : location_key

Sample Queries

Try the sample queries below in the BigQuery UI.

How many confirmed COVID-19 cases did each country have at the end of June?

This query determines the cumulative number of COVID-19 cases for each county, as well as a normalization of cumulative cases by country population, as of the end of June. [Run this query](#).

How did confirmed COVID-19 cases compare to COVID-19 tests in the US during the month of June, by state?

This query determines the percent of reported COVID-19 tests in each US state that returned positive by dividing the total number of reported tests that returned positive by the total number of COVID-19 tests conducted in the month of June. [Run this query](#).

COVID Data Engineering Project

Google Cloud Platform • My First Project

BigQuery FEATURES & INFO SHORTCUT

Query history

Saved queries

Job history

Transfers

Scheduled queries

Reservations

BI Engine

Resources + ADD DATA

Search for your tables and datasets

covid19_jhu_csse

covid19_jhu_csse_eu

covid19_nyt

covid19_open_data

compatibility_view

covid19_open_data

covid19_open_data_eu

covid19_public_forecasts

covid19_public_forecasts_asia

covid19_rxrx19

covid19_symptom_search

covid19_usafacts

covid19_weathersource_com

crypto_bitcoin

crypto_bitcoin_cash

crypto_dash

crypto_dogecoin

crypto_ethereum

NYT COVID 19: How do I calculate the number of new COVID-19 cases per day

```
1 SELECT
2     b.state_name,
3     b.date,
4     MAX(b.confirmed_cases - a.confirmed_cases) AS daily_confirmed_cases
5 FROM
6     (SELECT
7         state_name AS state,
8         state_fips_code ,
9         confirmed_cases,
10        DATE_ADD(date, INTERVAL 1 day) AS date_shift
11    FROM
12        `bigquery-public-data.covid19_nyt.us_states`
13    WHERE
14        confirmed_cases + deaths > 0) a
15 JOIN
16    `bigquery-public-data.covid19_nyt.us_states` b ON
17        a.state_fips_code = b.state_fips_code
18        AND a.date_shift = b.date
19 GROUP BY
20     b.state_name, date
21 ORDER BY
22     date desc
```

Run Save query Save view Schedule query More

Query results SAVE RESULTS EXPLORE DATA

Query complete (0.9 sec elapsed, 572.1 KB processed)

Job information	Results	JSON	Execution details
35 Tennessee	2020-11-30	7603	
36 Wisconsin	2020-11-30	2676	
37 California	2020-11-30	17696	
38 New Jersey	2020-11-30	3190	
39 New Mexico	2020-11-30	1678	
40 Washington	2020-11-30	3842	
41 Connecticut	2020-11-30	4714	
42 Mississippi	2020-11-30	1485	
43 Puerto Rico	2020-11-30	843	

Open Data by State - 12/1/20

Add a chart

Filter

subregion1_name	Nov_confirmed...	Nov_tested
1. Illinois	316,040	2,188,842
2. California	302,257	3,602,964
3. Texas	251,187	2,319,365
4. Ohio	206,719	650,482
5. Florida	194,826	2,129,351
6. Michigan	185,542	1,670,879
7. Wisconsin	173,861	1,192,302
8. Minnesota	170,211	1,761,460
9. Indiana	160,143	1,162,678
1...	154,368	1,462,205
1... Pennsylvania	140,418	3,959,781
1... New York	125,864	1,158,769
1... Colorado	121,251	360,866
1... Missouri	107,434	508,099
1... Tennessee	101,252	455,191
1... Iowa	97,398	1,308,403
1... New Jersey	89,636	1,147,330
1... North Carolina	81,100	752,869
1... Arizona	80,679	450,206
2... Utah	74,916	173,794
2... Oklahoma	73,666	171,747
2... Kansas	72,770	357,620
2... Kentucky	63,427	2,148,929
2... Massachusetts	59,963	674,752
2... Georgia	59,393	459,823
2... Nebraska	57,200	249,308
2... Virginia	56,657	561,583
2... Alabama	53,850	370,807
2... Maryland	53,274	905,409
3... Nevada	51,566	280,519
3... Louisiana	50,000	1,054,454