

COVID-19 JHU "Daily Reports" Data Quality Report

06-02-2020 Update

This automated report describes changes and idiosyncrasies in the structure and content of CSV files maintained in the COVID-19 GitHub "Daily Reports" repository: https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data. This PDF is generated daily and is intended to assist researchers and others who are evaluating raw data for US COVID-19 cases and deaths. Heartfelt thanks to the Johns Hopkins University (JHU) Center for Systems Science and Engineering (CSSE) for creating and continually updating the COVID-19 GitHub repository as well as the COVID-19 international tracker (<https://coronavirus.jhu.edu/map.html>).

Data Coverage

This report references CSV files that span 01-22-2020 to 06-02-2020, comprising 240,312 COVID-19 records; however, only the 212,240 United States records (having "US" in the Country_Region column) are described.

I. CSV File Structure

Throughout the pandemic, COVID-19 data have been reported with increasing consistency and specificity in the GitHub repository. However, several changes have occurred in how the data are represented, including the CSV file columns, column names, and value formats.

Columns included in CSV files dated 01-22-2020 to 03-01-2020:

Province/State, Country/Region, Last
Update, Confirmed, Deaths, Recovered

Columns included in CSV files dated 03-01-2020 to 03-22-2020:

Province/State, Country/Region, Last
Update, Confirmed, Deaths, Recovered, Latitude, Longitude

Columns included in CSV files dated 03-22-2020 to 05-29-2020:

FIPS, Admin2, Province_State, Country_Region, Last_Update, Lat, Long_, Co
nfirmated, Deaths, Recovered, Active, Combined_Key

Columns included in CSV files dated 05-29-2020 to 06-02-2020:

FIPS, Admin2, Province_State, Country_Region, Last_Update, Lat, Long_, Co
nfirmated, Deaths, Recovered, Active, Combined_Key, Incidence_Rate, Case-F
atality_Ratio

Column Mapping

Province/State should be mapped to Province_State

Country/Region should be mapped to Country_Region

Last Update should be mapped to Last_Update

Latitude should be mapped to Lat

Longitude should be mapped to Long_

II. CSV File Content

This section describes variability in the following columns: Province_State, FIPS, and Admin2 (i.e., county).

Province_State Column

The content of Province_State has changed substantially over the course of COVID-19 reporting. For some records, only the state name is referenced, whereas other records additionally include the county or city. On the other hand, some records specify "US" without a reference to a specific state.

Province_State Values with Full State Name

99.5% (N=211,118) of records contain a valid state name (e.g., California). In addition to states, these records also include the following US regions:

- District of Columbia (N=86)
- Puerto Rico (N=81)
- Guam (N=80)
- Virgin Islands (N=77)
- American Samoa (N=9)
- Northern Mariana Islands (N=73)

Province_State Values That Reference Cruise Ships

Cases from the Grand Princess and Diamond Princess cruise ships were originally attributed to states; however, in more recent CSV files, they are attributed only to the US, with Diamond Princess cases assigned the FIPS code 88888 and Grand Princess cases assigned the FIPS code 99999. All cruise-related records include:

- Grand Princess (N=85)
- Diamond Princess (N=85)
- Lackland, TX (From Diamond Princess) (N=17)
- Travis, CA (From Diamond Princess) (N=17)
- Omaha, NE (From Diamond Princess) (N=17)
- Unassigned Location (From Diamond Princess) (N=15)
- Grand Princess Cruise Ship (N=3)

Province_State Values That Include a County

0.2% (N=408) of records reference a county name in the state column:

San Diego County, CA (N=28)
Humboldt County, CA (N=18)
Sacramento County, CA (N=18)
Snohomish County, WA (N=10)
Grafton County, NH (N=8)
Cook County, IL (N=8)
King County, WA (N=8)
Placer County, CA (N=8)
Sonoma County, CA (N=8)
Westchester County, NY (N=7)
Washington County, OR (N=7)
Wake County, NC (N=7)
Fulton County, GA (N=7)
Maricopa County, AZ (N=7)
Orange County, CA (N=6)
Contra Costa County, CA (N=6)
Williamson County, TN (N=5)
Clark County, NV (N=5)
Norfolk County, MA (N=5)
Harris County, TX (N=5)
Grant County, WA (N=5)
San Francisco County, CA (N=5)
Santa Rosa County, FL (N=5)
Fort Bend County, TX (N=5)
Bergen County, NJ (N=5)
Middlesex County, MA (N=4)
Marion County, IN (N=4)
Fayette County, KY (N=4)
Wayne County, PA (N=4)
Yolo County, CA (N=4)
Nassau County, NY (N=4)
Washoe County, NV (N=4)
Suffolk County, MA (N=4)
Denver County, CO (N=4)
Chatham County, NC (N=4)
Ramsey County, MN (N=4)
Delaware County, PA (N=4)
New York County, NY (N=4)

Douglas County, NE (N=4)
Montgomery County, MD (N=4)
Summit County, CO (N=4)
Pinal County, AZ (N=3)
Lee County, FL (N=3)
Broward County, FL (N=3)
Fairfield County, CT (N=3)
Cobb County, GA (N=3)
Alameda County, CA (N=3)
Rockland County, NY (N=3)
El Paso County, CO (N=3)
Kershaw County, SC (N=3)
Providence County, RI (N=3)
Santa Clara County, CA (N=3)
Charleston County, SC (N=3)
Tulsa County, OK (N=3)
Pierce County, WA (N=3)
Clark County, WA (N=3)
Douglas County, CO (N=3)
Montgomery County, TX (N=3)
Davis County, UT (N=3)
Saratoga County, NY (N=3)
Jefferson County, WA (N=3)
Honolulu County, HI (N=3)
Madera County, CA (N=3)
Jackson County, OR (N=3)
Klamath County, OR (N=3)
Plymouth County, MA (N=3)
Davidson County, TN (N=2)
Shelby County, TN (N=2)
Hendricks County, IN (N=2)
Okaloosa County, FL (N=2)
Kittitas County, WA (N=2)
Rockingham County, NH (N=2)
Fresno County, CA (N=2)
Hudson County, NJ (N=2)
Marion County, OR (N=2)
Volusia County, FL (N=2)
St. Louis County, MO (N=2)

Harford County, MD (N=2)
Montgomery County, PA (N=2)
Johnson County, KS (N=2)
Fairfax County, VA (N=2)
Riverside County, CA (N=2)
Ulster County, NY (N=2)
Floyd County, GA (N=2)
Douglas County, OR (N=2)
Suffolk County, NY (N=2)
Spokane County, WA (N=2)
Manatee County, FL (N=2)
Norfolk County, MA (N=2)
Berkshire County, MA (N=2)
Polk County, GA (N=2)
Harrison County, KY (N=1)
Spartanburg County, SC (N=1)
Collin County, TX (N=1)
Johnson County, IA (N=1)
Cherokee County, GA (N=1)
Shasta County, CA (N=1)
Carver County, MN (N=1)
Norwell County, MA (N=1)
Jefferson County, KY (N=1)
Santa Cruz County, CA (N=1)
Bennington County, VT (N=1)
Charlotte County, FL (N=1)
Queens County, NY (N=1)

Province_State Values That Include a City

0.2% (383) of records reference a city name in the Province_State column:

Los Angeles, CA (N=38)
San Benito, CA (N=36)
Tempe, AZ (N=35)
Santa Clara, CA (N=35)
Madison, WI (N=34)
Boston, MA (N=34)
Orange, CA (N=32)
Seattle, WA (N=30)
Chicago, IL (N=30)

San Antonio, TX (N=26)
Hillsborough, FL (N=8)
Sarasota, FL (N=8)
San Mateo, CA (N=8)
Umatilla, OR (N=8)
Providence, RI (N=6)
New York City, NY (N=4)
Berkeley, CA (N=4)
Portland, OR (N=3)
Ashland, NE (N=1)
Jefferson Parish, LA (N=1)
Travis, CA (N=1)
Lackland, TX (N=1)

Province_State Values That Are Unassigned

0.0% (7) of records reference Unassigned or Unknown in the Province_State column:

Unassigned Location, WA (N=4)
Unknown Location, MA (N=2)
Unassigned Location, VT (N=1)

Province_State Additional Values

0.0% (N=85) of records have additional values in the Province_State column, some of which are spelling variations that can be manually mapped to state or county FIPS codes:

Recovered (N=71)
United States Virgin Islands (N=4)
US (N=4)
Virgin Islands, U.S. (N=2)
Washington, D.C. (N=2)
Chicago (N=1)
Wuhan Evacuee (N=1)

FIPS Column

The 5-digit FIPS code uniquely describes county and county-level regions within the US. The first two characters describe US states and territories and the last three characters describe counties therein. The US Census maintains a table of FIPS codes at: <https://www.census.gov/geographies/reference-files/2017/demo/popest/2017-fips.html>.

For COVID-19 records having a valid FIPS code, state and county names can be reliably derived by merging the FIPS table, without the need to rely on the Admin2 (i.e., county) or

Province_State columns in the COVID-19 CSV files. However, for the 1,560 records prior to March 22 (when FIPS codes were added), county identification can be difficult, ambiguous, or impossible (so this county mapping is neither attempted nor described herein). In the 210,680 records after March 22, FIPS codes comprise three categories:

- 97.7% (205,882) of US records after March 22 have a valid FIPS code
- 0.5% (1,123) of US records after March 22 are missing a FIPS code
- 1.7% (3,675) of US records after March 22 have an invalid FIPS code

Missing FIPS Values (Unassigned County)

Many missing FIPS values result from a county labeled Unassigned or Unknown:

Province_State	Admin2 (i.e., County)	Count
Louisiana	Unassigned	20
New Jersey	Unassigned	11
New York	Unassigned	11
Nevada	Unassigned	11
Massachusetts	Unassigned	11
Rhode Island	Unassigned	11
Colorado	Unassigned	11
Kentucky	Unassigned	11
Georgia	Unassigned	11
Arkansas	Unassigned	11
Washington	Unassigned	11
Maine	Unassigned	10
Michigan	Unassigned	10
Illinois	Unassigned	10
Tennessee	Unassigned	10
Vermont	Unassigned	10
Hawaii	Unassigned	9
Arizona	Unassigned	9
Wyoming	Unassigned	9
Alaska	Unassigned	8
Florida	Unassigned	8
Wisconsin	Unassigned	6
Minnesota	Unassigned	6
Missouri	Unassigned	6
Maryland	Unassigned	5
New Hampshire	Unassigned	5
California	Unassigned	5
Nebraska	Unassigned	5

Wyoming	unassigned	5
Montana	Unassigned	4
Virginia	Unassigned	3
Connecticut	Unassigned	3
Iowa	Unassigned	3
Utah	unassigned	3
Alabama	Unassigned	2
Delaware	Unassigned	2
North Carolina	Unassigned	2
Wuhan Evacuee	Unassigned	1
Oklahoma	Unassigned	1
Mississippi	Unassigned	1
Tennessee	Unknown	1
North Dakota	Unassigned	1
Indiana	Unassigned	1

Missing FIPS Values (Out of [STATE])

Some missing FIPS values result from a county labeled Out of [state abbreviation]. These records initially did not contain a FIPS value, although later records were attributed an arbitrary FIPS code. For example, the Admin2 value "Out of CO" (i.e., Colorado) is consistently assigned the (arbitrary / invalid) FIPS code 80008.

Province_State	Admin2 (i.e., County)	Count
Tennessee	Out of TN	24
Michigan	Out of MI	24
Utah	Out of UT	11
Oklahoma	Out of OK	3
Tennessee	Out-of-state	1

Missing FIPS Values (Other Values)

The remaining records missing a FIPS value include the following, with some counties having been merged into larger regions (e.g., Central Utah) and other Admin2 (i.e., county) values referencing a borough (e.g., Matanuska-Susitna) that can be manually mapped to its respective county-level FIPS value:

Province_State	Admin2 (i.e., County)	Count
Massachusetts	Dukes and Nantucket	73
Missouri	Kansas City	72
Recovered		71
Michigan	Federal Correctional Institution (FCI)	51
Michigan	Michigan Department of Corrections (MDOC)	51

Utah	TriCounty	45
Utah	Bear River	45
Utah	Southeast Utah	45
Utah	Weber-Morgan	45
Utah	Central Utah	45
Utah	Southwest Utah	21
Virginia	Manassas	16
Louisiana	LaSalle	16
Tennessee	DeKalb	16
Northern Mariana Islands		16
Utah	Weber	16
Georgia	McIntosh	16
Georgia	McDuffie	16
Illinois	LaSalle	14
Minnesota	Fillmore	14
Minnesota	Lac qui Parle	14
Puerto Rico		11
Pennsylvania	McKean	8
Florida	DeSoto	7
New Mexico	Dona Ana	6
Alaska	Matanuska-Susitna	3
Texas	Harris	2
Alaska	Kenai Peninsula	2
Virgin Islands		2
Alaska	Soldotna	1
Alaska	Sterling	1
Virginia	Staunton	1
Florida	Desoto	1
Massachusetts	Brockton	1
Minnesota	LeSeur	1
New Hampshire	Nashua	1
Tennessee	Unknown	1
Tennessee	Out-of-state	1

Invalid FIPS Values

Invalid FIPS values commonly result from records for which the Admin2 (i.e., county) value is Unassigned or Out of [state abbreviation], in which case an arbitrary (yet unique by state) FIPS value is assigned to the record. This enables these records to be attributed to individual states and may be sufficient for state-level analysis; however, for geospatial or other county-level

analysis in which a valid FIPS code is required, COVID-19 records that lack county distinction will underrepresent the pandemic because the sum of county-level data will fail to fully represent state-level data.

FIPS	Admin2 (i.e., County)	Province_State	Count
00066		Guam	62
00069		Northern Mariana Islands	48
00072		Puerto Rico	62
00078		Virgin Islands	62
80008	Out of CO	Colorado	39
80013	Out of GA	Georgia	50
80015	Out of HI	Hawaii	59
80017	Out of IL	Illinois	28
80026	Out of MI	Michigan	48
80040	Out of OK	Oklahoma	61
80047	Out of TN	Tennessee	48
88888		Diamond Princess	73
90001	Unassigned	Alabama	52
90002	Unassigned	Alaska	62
90004	Unassigned	Arizona	62
90005	Unassigned	Arkansas	62
90006	Unassigned	California	61
90008	Unassigned	Colorado	53
90009	Unassigned	Connecticut	62
90010	Unassigned	Delaware	62
90012	Unassigned	Florida	62
90013	Unassigned	Georgia	62
90015	Unassigned	Hawaii	62
90016	Unassigned	Idaho	60
90017	Unassigned	Illinois	62
90018	Unassigned	Indiana	58
90019	Unassigned	Iowa	62
90020	Unassigned	Kansas	60
90021	Unassigned	Kentucky	62
90022	Unassigned	Louisiana	53
90023	Unassigned	Maine	62
90024	Unassigned	Maryland	52
90025	Unassigned	Massachusetts	62
90026	Unassigned	Michigan	57
90027	Unassigned	Minnesota	62

90028	Unassigned	Mississippi	56
90029	Unassigned	Missouri	60
90030	Unassigned	Montana	62
90031	Unassigned	Nebraska	62
90032	Unassigned	Nevada	62
90033	Unassigned	New Hampshire	62
90034	Unassigned	New Jersey	62
90035	Unassigned	New Mexico	62
90036	Unassigned	New York	62
90037	Unassigned	North Carolina	46
90038	Unassigned	North Dakota	50
90039	Unassigned	Ohio	59
90040	Unassigned	Oklahoma	62
90041	Unassigned	Oregon	54
90042	Unassigned	Pennsylvania	57
90044	Unassigned	Rhode Island	62
90045	Unassigned	South Carolina	41
90046	Unassigned	South Dakota	62
90047	Unassigned	Tennessee	62
90048	Unassigned	Texas	41
90049	Southwest	Utah	9
90049	Southwest Utah	Utah	24
90049	Unassigned	Utah	56
90050	Unassigned	Vermont	62
90051	Unassigned	Virginia	62
90053	Unassigned	Washington	62
90054	Unassigned	West Virginia	57
90055	Unassigned	Wisconsin	62
90056	Unassigned	Wyoming	48
99999		Grand Princess	73

Variability in US Territory FIPS Values

Since March 22, COVID-19 cases have been reported at the county level for all US states; however, US territories continue to report cases only at the territory level. Note that FIPS values for US territories can be either two or five characters in length, including the following:

FIPS	Province_State	Count
	Guam	7
	Northern Mariana Islands	16
	Puerto Rico	19

	United States Virgin Islands	4
	Virgin Islands	6
	Virgin Islands, U.S.	2
00066	Guam	62
00069	Northern Mariana Islands	48
00072	Puerto Rico	62
00078	Virgin Islands	62
60000	American Samoa	9
66000	Guam	11
69000	Northern Mariana Islands	9
78000	Virgin Islands	9

Admin2 (i.e., County) Column

County name was infrequently reported prior to March 22, and was commingled with both city and state names in the Province_State column. After March 22, county appears in the Admin2 column. Because a valid FIPS value allows a record to be merged with US Census data (to generate county name), this section discusses only records for which the FIPS value is missing.

Manually Mapping State and County to FIPS

In some cases, county and state are provided but FIPS is missing from the record. This can occur when counties are misspelled, when two or more counties are combined into a larger region, or for unknown reasons. Records with the following state+county profiles are missing FIPS, with manually mapped FIPS values provided where they can be generated (and otherwise indicated as unknown):

State	Admin2 (i.e., County)	FIPS	Count
Alaska	Sterling	02122	1
Alaska	Soldotna	02122	1
Alaska	Kenai Peninsula	02122	73
Alaska	Matanuska-Susitna	02170	73
Florida	DeSoto	12027	73
Florida	Desoto	12027	1
Georgia	McDuffie	13189	73
Georgia	McIntosh	13191	73
Illinois	LaSalle	17099	73
Louisiana	LaSalle	22059	73
Massachusetts	Brockton	25023	1
Minnesota	Fillmore	27045	73
Minnesota	Lac qui Parle	27073	73

Minnesota	LeSeur	27079	1
Missouri	Kansas City	29095	72
New Hampshire	Nashua	33011	1
New Mexico	Dona Ana	35013	73
Pennsylvania	McKean	42083	73
Tennessee	DeKalb	47041	73
Texas	Harris	48201	73
Utah	Bear River	49003	45
Utah	Weber	49057	28
Virginia	Manassas	51683	73
Virginia	Staunton	51790	70
Utah	Southwest Utah	UNK	45
Utah	Southeast Utah	UNK	45
Utah	Weber-Morgan	UNK	45
Utah	Central Utah	UNK	45
Utah	TriCounty	UNK	45
Massachusetts	Dukes and Nantucket	UNK	73

Unmapped State and County Columns (Additional Information)

Some context is provided for these manually mapped records, including those that cannot be mapped due to ambiguity:

Sterling, Alaska:

Sterling is located in the Kenai Peninsula Borough.

Soldotna, Alaska:

Soldotna is located in the Kenai Peninsula Borough.

Kenai Peninsula, Alaska:

The FIPS value was omitted.

Matanuska-Susitna, Alaska:

The FIPS value was omitted.

DeSoto, Florida:

DeSoto County, Florida, is both misspelled and omitted from some records.

(<https://www.census.gov/quickfacts/desotocountyflorida>)

Desoto, Florida:

DeSoto County, Florida, is both misspelled and omitted from some records.

(<https://www.census.gov/quickfacts/desotocountyflorida>)

McIntosh, Georgia:

The FIPS value was omitted.

McDuffie, Georgia:

The FIPS value was omitted.

LaSalle, Illinois:

The FIPS value was omitted.

LaSalle, Louisiana:

The FIPS value was omitted.

Brockton, Massachusetts:

Brockton is located in Plymouth County.

Dukes and Nantucket, Massachusetts:

JHU FAQ describes: "Dukes and Nantucket Counties, Massachusetts: The state reports confirmed cases by county, but aggregates deaths for these two counties together."

(<https://coronavirus.jhu.edu/us-map-faq>)

Lac qui Parle, Minnesota:

The FIPS value was omitted.

LeSeur, Minnesota:

Le Sueur County is misspelled

Fillmore, Minnesota:

The FIPS value was omitted.

Kansas City, Missouri:

Kansas City, Missouri, does not have a US Census-recognized 5-digit FIPS code, but it can be binned into its predominant county, Jackson County (FIPS) 29095, consistent with how other organizations have binned these data.

(<https://source.opennews.org/articles/comparison-four-major-covid-19-data-sources/>)

Nashua, New Hampshire:

Nashua is in Hillsborough County

Dona Ana, New Mexico:

The FIPS value was omitted.

McKean, Pennsylvania:

McKean Borough is located in Erie County, FIPS 42083.

(<https://www.census.gov/quickfacts/eriecountypennsylvania>)

DeKalb, Tennessee:

The FIPS value was omitted.

Harris, Texas:

The FIPS value was omitted.

Weber, Utah:

The FIPS value was omitted.

Southeast Utah, Utah:

JHU FAQ describes: "Utah: Utah is reporting county data somewhat differently than many other states. The larger-population counties are reporting confirmed cases and deaths at the county level. However, the smaller counties are banded together into county groups.

This is in an effort to protect identities of individuals."

(<https://coronavirus.jhu.edu/us-map-faq>). The Southeast Utah Health Department

(<https://www.seuhealth.com/covid-19>) includes Carbon County, Emery County, and Grand County.

Weber-Morgan, Utah:

JHU FAQ describes: "Utah: Utah is reporting county data somewhat differently than many other states. The larger-population counties are reporting confirmed cases and deaths at the county level. However, the smaller counties are banded together into county groups.

This is in an effort to protect identities of individuals."

(<https://coronavirus.jhu.edu/us-map-faq>). The Weber-Morgan Health Department

(<http://www.webermorganhealth.org/coronavirus/>) includes Weber County and Morgan County.

Central Utah, Utah:

JHU FAQ describes: "Utah: Utah is reporting county data somewhat differently than many other states. The larger-population counties are reporting confirmed cases and deaths at the county level. However, the smaller counties are banded together into county groups.

This is in an effort to protect identities of individuals."

(<https://coronavirus.jhu.edu/us-map-faq>). The Central Utah Health Department (<https://centralutahpublichealth.org/>) includes Juab County, Millard County, Piute County, Sanpete County, Sevier County, and Wayne County.

TriCounty, Utah:

JHU FAQ describes: "Utah: Utah is reporting county data somewhat differently than many other states. The larger-population counties are reporting confirmed cases and deaths at the county level. However, the smaller counties are banded together into county groups. This is in an effort to protect identities of individuals."

(<https://coronavirus.jhu.edu/us-map-faq>). The TriCounty Utah Health Department (<https://tricityhealth.com/local-covid-19-situational-update/>) includes Uintah County, Duchesne County, and Daggett County.

Bear River, Utah:

Bear River, Utah, is located in Box Elder County.

(<https://www.census.gov/quickfacts/fact/table/boxeldercountyutah/INC110218>)

Southwest Utah, Utah:

JHU FAQ describes: "Utah: Utah is reporting county data somewhat differently than many other states. The larger-population counties are reporting confirmed cases and deaths at the county level. However, the smaller counties are banded together into county groups. This is in an effort to protect identities of individuals."

(<https://coronavirus.jhu.edu/us-map-faq>). The Southwest Utah Health Department (<https://swuhealth.org/covid/>) includes Washington County, Iron County, Kane County, Beaver County, and Garfield County.

Manassas, Virginia:

The FIPS value was omitted.

Staunton, Virginia:

The FIPS value was omitted.