

# Public 19 Reporting

## COVID-19 Case Surveillance Public Use Data Utility Summary

Users should consider the level of completeness, including suppression levels when planning their analyses and use of public datasets. Privacy protections will suppress field values to reduce reidentification risks. Completeness varies by jurisdiction (i.e., state, local, and territorial) and time period. Variables are consistently coded to the value “Unknown” when jurisdictions specify in the case data submitted to CDC that the value is unknown, the value “Missing” when jurisdictions do not provide a value, and the value “NA” when the value is suppressed as part of privacy protections.

Dataset version: 12/1/2022

## Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co... Double	quasi_fields_pct Double	
1	total_rows	91,690,933	NaN%	91,690,933	NaN%	
2	total_columns	19	NaN%	8	NaN%	
3	total_cells	1,742,127,727	100.0%	733,527,464	100.0%	
4	suppressed_fields   ...	47,223,032	2.7%	40,888,414	5.6%	
5	missing_fields	403,421,451	23.2%	63,423,061	8.6%	
6	unknown_fields	83,777,044	4.8%	43,254,344	5.9%	
7	non_blank_fields	1,207,706,200	69.3%	585,961,645	79.9%	

## Field Level Utility Summary

	variable String	suppressed Long	suppressed_pct String	missing Long	missing_pct String	unknown Long	unknown_pct String
1	case_month	9	0.0%	0	0.0%	0	0.0%
2	res_state	1,279	0.0%	0	0.0%	0	0.0%
3	res_county	6,333,339	6.9%	0	0.0%	0	0.0%
4	sex	2,202,445	2.4%	416,651	0.5%	1,139,559	1.2%
5	age_group	748,007	0.8%	761,485	0.8%	0	0.0%
6	ethnicity	15,177,106	16.6%	6,228,877	6.8%	17,865,076	19.5%
7	race	13,551,093	14.8%	7,657,100	8.4%	12,557,223	13.7%
8	death_yn	2,875,136	3.1%	48,358,948	52.7%	11,692,486	12.8%
9	records_with_any_quasi_identifier	22,471,549	24.5%	51,314,686	56.0%	30,540,994	33.3%

