

## Public 12 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: 9/6/2022

## Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_counts Double	quasi_fields_pct Double
1	total_rows	86,008,515	NaN%	86,008,515	NaN%
2	total_columns	12	NaN%	3	NaN%
3	total_cells	1,032,102,180	100.0%	258,025,545	100.0%
4	suppressed_fields ...	89	0.0%	89	0.0%
5	missing_fields	235,798,738	22.8%	3,590,199	1.4%
6	unknown_fields	70,659,864	6.8%	28,420,475	11.0%
7	non_blank_fields	725,643,489	70.3%	226,014,782	87.6%

## Field Level Utility Summary

	variable String	suppressed Long	suppressed_percent String	missing Long	missing_percent String	unknown Long	unknown_percent String
1	sex	11	0.0%	478,102	0.6%	1,229,663	1.4%
2	age_group	74	0.0%	779,667	0.9%	0	0.0%
3	race_ethnicity_combined	4	0.0%	2,332,430	2.7%	27,190,812	31.6%
4	records_with_any_quasi_identifier	74	0.0%	3,299,402	3.8%	27,664,809	32.2%

