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.

Versions:

Dataset version: 7/1/2022

Quick Summary

	summary	all_fields_counts	all_fields_pct	quasi_fields_counts	quasi_fields_pct
1	total_rows	77,544,202	NaN%	77,544,202	NaN%
2	total_columns	12	NaN%	3	NaN%
3	total_cells	930,530,424	100.0%	232,632,606	100.0%
4	suppressed_fields	85	0.0%	85	0.0%
5	missing_fields	209,572,687	22.5%	3,068,244	1.3%
6	unknown_fields	63,814,915	6.9%	25,634,422	11.0%
7	non_blank_fields	657,142,737	70.6%	203,929,855	87.7%

Field Level Utility Summary

	variable	suppressed	suppressed_percent	missing	missing_percent	unknown	unknown_percent
1	sex	10	0.0%	321,683	0.4%	1,112,374	1.4%
2	age_group	71	0.0%	755,733	1.0%	0	0.0%
3	race_ethnicity_combined	4	0.0%	1,990,828	2.6%	24,522,048	31.6%
4	records_with_any_quasi_identifier	71	0.0%	2,883,287	3.7%	24,939,534	32.2%