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: 5/3/2022

Quick Summary

	summary	all_fields_counts	all_fields_pct	quasi_fields_counts	quasi_fields_pct
1	total_rows	71,387,132	NaN%	71,387,132	NaN%
2	total_columns	12	NaN%	3	NaN%
3	total_cells	856,645,584	100.0%	214,161,396	100.0%
4	suppressed_fields	101	0.0%	101	0.0%
5	missing_fields	191,478,752	22.4%	2,715,244	1.3%
6	unknown_fields	59,892,695	7.0%	23,669,841	11.1%
7	non_blank_fields	605,274,036	70.7%	187,776,210	87.7%

Field Level Utility Summary

	variable	suppressed	suppressed_percent	missing	missing_percent
1	sex	8	0.0%	202085	0.3%
2	age_group	85	0.0%	753501	1.1%
3	race_ethnicity_combined	8	0.0%	1759658	2.5%
4	records_with_any_quasi_identifier	85	0.0%	2534600	3.6%