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: 6/6/2022

Quick Summary

	summary	all_fields_counts	all_fields_pct	quasi_fields_counts	quasi_fields_pct
1	total_rows	74,849,225	NaN%	74,849,225	NaN%
2	total_columns	12	NaN%	3	NaN%
3	total_cells	898,190,700	100.0%	224,547,675	100.0%
4	suppressed_fields	108	0.0%	108	0.0%
5	missing_fields	201,682,678	22.5%	2,836,517	1.3%
6	unknown_fields	62,200,564	6.9%	24,988,007	11.1%
7	non_blank_fields	634,307,350	70.6%	196,723,043	87.6%

Field Level Utility Summary

	variable	suppressed	suppressed_percent	missing	missing_percent	unknown	unknown_percent
1	sex	14	0.0%	203,757	0.3%	1,433,509	1.9%
2	age_group	86	0.0%	750,557	1.0%	0	0.0%
3	race_ethnicity_combined	8	0.0%	1,882,203	2.5%	23,554,498	31.5%
4	records_with_any_quasi_identifier	86	0.0%	2,657,528	3.6%	24,179,698	32.3%