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: 7/6/2023

Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_counts Double	quasi_fields_pct Double	
1	total_rows	98,952,628	NaN%	98,952,628	NaN%	
2	total_columns	12	NaN%	3	NaN%	
3	total_cells	1,187,431,536	100.0%	296,857,884	100.0%	
4	suppressed_fields	84	0.0%	84	0.0%	
5	missing_fields	269,417,255	22.7%	4,041,923	1.4%	
6	unknown_fields	81,307,654	6.8%	31,077,535	10.5%	
7	non_blank_fields	836,706,543	70.5%	261,738,342	88.2%	

Field Level Utility Summary

	variable String	suppressed Long	suppressed_pct String	missing Long	missing_pct String	unknown Long	unknown_pct String
1	sex	11	0.0%	491,401	0.5%	1,010,854	1.0%
2	age_group	62	0.0%	965,452	1.0%	0	0.0%
3	race_ethnicity_combined	11	0.0%	2,573,405	2.6%	30,045,189	30.5%
4	records_with_any_quasi_identifier	62	0.0%	3,723,227	3.8%	30,358,580	30.8%