Public 19 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: 2/2/2023

Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co... Double	quasi_fields_pct Double	
1	total_rows	95,607,907	NaN%	95,607,907	NaN%	
2	total_columns	19	NaN%	8	NaN%	
3	total_cells	1,816,550,233	100.0%	764,863,256	100.0%	
4	suppressed_fields •••	49,813,149	2.7%	43,207,056	5.6%	
5	missing_fields	419,363,409	23.1%	65,659,385	8.6%	
6	unknown_fields	87,236,120	4.8%	44,631,925	5.8%	
7	non_blank_fields	1,260,137,555	69.4%	611,364,890	79.9%	

Field Level Utility Summary

	variable String	suppressed Long	suppressed_pct String	missing Long	missing_pct String	unknown Long	unknown_pct String
1	case_month	9	0.0%	0	0.0%	0	0.0%
2	res_state	1,261	0.0%	0	0.0%	0	0.0%
3	res_county	6,604,832	6.9%	0	0.0%	0	0.0%
4	sex	2,386,582	2.5%	421,424	0.4%	1,182,348	1.2%
5	age_group	800,304	0.8%	900,407	0.9%	0	0.0%
6	ethnicity	16,082,300	16.8%	6,024,846	6.3%	18,211,182	19.0%
7	race	14,367,887	15.0%	7,784,773	8.1%	12,811,111	13.4%
8	death_yn	2,963,881	3.1%	50,527,935	52.8%	12,427,284	13.0%
9	records_with_any_quasi_identifier	23,652,155	24.7%	53,502,397	56.0%	31,652,343	33.1%