

# BIN381 — Milestone 1: Data Understanding

2025-09-09

## Table of Contents

1. Load All Datasets .....	1
2. Dataset-Level Summary.....	3
3. Data Quality Assessment.....	4
3.1 Missing Values (Per Column, All Datasets) .....	4
3.1.1 Missingness Charts (Per Dataset) .....	14
3.2 Duplicates (Row-Level).....	19
3.3 Outliers (Numeric Columns, $ z  > 3$ ) .....	20
3.4 Consolidated Data Quality Issues Log .....	24

## 1. Load All Datasets

```
base <- "."
# Optional: set to TRUE to write CSV exports alongside this HTML
write_exports <- FALSE
# Optional: print full duplicate rows if count <= this threshold (to avoid
# huge output)
dup_print_threshold <- 200L

# Detect files
files_csv <- list.files(base, pattern = "\\.(csv)$", ignore.case = TRUE,
full.names = TRUE)
files_xlsx <- list.files(base, pattern = "\\.(xlsx)$", ignore.case = TRUE,
full.names = TRUE)

# Read helpers
read_csv_clean <- function(path){
  readr::read_csv(path, show_col_types = FALSE) |> janitor::clean_names()
}
read_xlsx_all <- function(path){
  sh <- readxl::excel_sheets(path)
  setNames(
    purrr::map(sh, ~ readxl::read_excel(path, sheet = .x) |>
janitor::clean_names() |> as_tibble() ),
    paste0(tools::file_path_sans_ext(basename(path)), "__", sh)
  )
}
```

```

)
}

# Load data
dfs_csv <- purrr::map(files_csv, read_csv_clean); names(dfs_csv) <-
tools::file_path_sans_ext(basename(files_csv))
dfs_xlsx <- purrr::map(files_xlsx, read_xlsx_all); dfs_xlsx <-
if(length(dfs_xlsx)) purrr::list_flatten(dfs_xlsx) else list()
dfs <- c(dfs_csv, dfs_xlsx)

# Ensure unique names
if(length(dfs)){
  names(dfs) <- make.unique(names(dfs), sep = "_")
}

# Inventory
inventory <- tibble(
  dataset = names(dfs),
  rows = purrr::map_int(dfs, nrow),
  cols = purrr::map_int(dfs, ncol)
) |> arrange(dataset)

if(nrow(inventory) == 0){
  stop("No datasets found. Place this .Rmd in the folder with your CSV/XLSX
files and Knit again.")
}

gt::gt(inventory)

```

dataset	rows	cols
access-to-health-care_national_zaf	276	29
anthropometry_national_zaf	38	29
child-mortality-rates_national_zaf	41	29
covid-19-prevention_national_zaf	35	29
dhs-quickstats_national_zaf	53	29
hiv-behavior_national_zaf	119	29
immunization_national_zaf	117	29
iycf_national_zaf	23	29
literacy_national_zaf	21	29
maternal-mortality_national_zaf	22	29
symptoms-of-acute-respiratory-infection-ari_national_zaf	27	29
toilet-facilities_national_zaf	47	29
water_national_zaf	101	29

## 2. Dataset-Level Summary

```
dataset_summary <- purrr::imap_dfr(dfs, function(df, nm){
  n_rows <- nrow(df); n_cols <- ncol(df)
  dup_rows <- sum(duplicated(df))
  total_cells <- n_rows * n_cols
  miss_cells <- sum(is.na(df))
  miss_pct <- if (total_cells > 0) round(100 * miss_cells / total_cells, 2)
else 0
  num_cols <- df |> dplyr::select(where(is.numeric)) |> ncol()
  tibble(
    dataset = nm,
    rows = n_rows,
    cols = n_cols,
    duplicate_rows = dup_rows,
    missing_cells = miss_cells,
    missing_pct = miss_pct,
    numeric_cols = num_cols,
    categorical_cols = n_cols - num_cols
  )
}) |> arrange(dataset)

gt::gt(dataset_summary)
```

dataset	rows	cols	duplicate_rows	missing_cells	missing_pct	numeric_cols	categorical_cols
access-to-health-care_national_zaf	276	29	0	1199	14.98	8	21
anthropometry_national_zaf	38	29	0	211	19.15	8	21
child-mortality-rates_national_zaf	41	29	0	210	17.66	10	19
covid-19-prevention_national_zaf	35	29	0	192	18.92	8	21
dhs-quickstats_national_zaf	53	29	0	267	17.37	10	19
hiv-behavior_national_zaf	119	29	0	685	19.85	8	21
immunization_national_zaf	117	29	0	554	16.33	8	21
iycf_national_zaf	23	29	0	132	19.79	8	21
literacy_national_zaf	21	29	0	122	20.03	8	21
maternal-mortality_national_zaf	22	29	0	151	23.67	10	19
symptoms-of-acute-respiratory-infection-ari_national_zaf	27	29	0	138	17.62	8	21
toilet-facilities_national_zaf	47	29	0	256	18.78	8	21
water_national_zaf	101	29	0	526	17.96	8	21

```
if (write_exports) {
  readr::write_csv(dataset_summary, "M1_dataset_summary.csv")
}
```

## 3. Data Quality Assessment

### 3.1 Missing Values (Per Column, All Datasets)

```
missingness_by_dataset <- purrr::imap(dfs, function(df, nm){
  tibble(
    dataset = nm,
    column = names(df),
    missing_pct = round(colMeans(is.na(df))*100, 2),
    missing_count = colSums(is.na(df))
  )
})

# Full table (all columns with any missingness first)
missingness_all <- bind_rows(missingness_by_dataset) |>
  arrange(desc(missing_pct), dataset, column)

gt::gt(missingness_all)
```

dataset	column	missing_pct	missing_count
access-to-health-care_national_zaf	ci_high	100.00	276
access-to-health-care_national_zaf	ci_low	100.00	276
access-to-health-care_national_zaf	level_rank	100.00	276
access-to-health-care_national_zaf	region_id	100.00	276
anthropometry_national_zaf	ci_high	100.00	38
anthropometry_national_zaf	ci_low	100.00	38
anthropometry_national_zaf	level_rank	100.00	38
anthropometry_national_zaf	region_id	100.00	38
child-mortality-rates_national_zaf	level_rank	100.00	41
child-mortality-rates_national_zaf	region_id	100.00	41
covid-19-prevention_national_zaf	ci_high	100.00	35
covid-19-prevention_national_zaf	ci_low	100.00	35
covid-19-prevention_national_zaf	level_rank	100.00	35
covid-19-prevention_national_zaf	region_id	100.00	35
dhs-quickstats_national_zaf	level_rank	100.00	53
dhs-quickstats_national_zaf	region_id	100.00	53
hiv-behavior_national_zaf	ci_high	100.00	119
hiv-behavior_national_zaf	ci_low	100.00	119
hiv-behavior_national_zaf	level_rank	100.00	119
hiv-behavior_national_zaf	region_id	100.00	119
immunization_national_zaf	ci_high	100.00	117
immunization_national_zaf	ci_low	100.00	117
immunization_national_zaf	level_rank	100.00	117
immunization_national_zaf	region_id	100.00	117
iycf_national_zaf	ci_high	100.00	23
iycf_national_zaf	ci_low	100.00	23
iycf_national_zaf	level_rank	100.00	23
iycf_national_zaf	region_id	100.00	23
literacy_national_zaf	ci_high	100.00	21
literacy_national_zaf	ci_low	100.00	21
literacy_national_zaf	level_rank	100.00	21
literacy_national_zaf	region_id	100.00	21
maternal-mortality_national_zaf	level_rank	100.00	22
maternal-mortality_national_zaf	region_id	100.00	22
symptoms-of-acute-respiratory-infection-ari_national_zaf	ci_high	100.00	27
symptoms-of-acute-respiratory-infection-ari_national_zaf	ci_low	100.00	27
symptoms-of-acute-respiratory-infection-ari_national_zaf	level_rank	100.00	27
symptoms-of-acute-respiratory-infection-ari_national_zaf	region_id	100.00	27
toilet-facilities_national_zaf	ci_high	100.00	47
toilet-facilities_national_zaf	ci_low	100.00	47
toilet-facilities_national_zaf	level_rank	100.00	47

dataset	column	missing_pct	missing_count
toilet-facilities_national_zaf	region_id	100.00	47
water_national_zaf	ci_high	100.00	101
water_national_zaf	ci_low	100.00	101
water_national_zaf	level_rank	100.00	101
water_national_zaf	region_id	100.00	101
hiv-behavior_national_zaf	by_variable_label	99.16	118
water_national_zaf	by_variable_label	99.01	100
toilet-facilities_national_zaf	by_variable_label	97.87	46
anthropometry_national_zaf	by_variable_label	97.37	37
covid-19-prevention_national_zaf	by_variable_label	97.14	34
iycf_national_zaf	by_variable_label	95.65	22
maternal-mortality_national_zaf	by_variable_label	95.45	21
literacy_national_zaf	by_variable_label	95.24	20
maternal-mortality_national_zaf	denominator_weighted	90.91	20
child-mortality-rates_national_zaf	denominator_unweighted	90.24	37
child-mortality-rates_national_zaf	denominator_weighted	90.24	37
maternal-mortality_national_zaf	ci_high	86.36	19
maternal-mortality_national_zaf	ci_low	86.36	19
dhs-quickstats_national_zaf	ci_high	73.58	39
dhs-quickstats_national_zaf	ci_low	73.58	39
maternal-mortality_national_zaf	denominator_unweighted	72.73	16
dhs-quickstats_national_zaf	by_variable_label	62.26	33
child-mortality-rates_national_zaf	by_variable_label	48.78	20
immunization_national_zaf	by_variable_label	47.86	56
dhs-quickstats_national_zaf	denominator_unweighted	35.85	19
dhs-quickstats_national_zaf	denominator_weighted	35.85	19
hiv-behavior_national_zaf	denominator_weighted	33.61	40
symptoms-of-acute-respiratory-infection-ari_national_zaf	denominator_unweighted	33.33	9
symptoms-of-acute-respiratory-infection-ari_national_zaf	denominator_weighted	33.33	9
hiv-behavior_national_zaf	denominator_unweighted	32.77	39
child-mortality-rates_national_zaf	ci_high	26.83	11
child-mortality-rates_national_zaf	ci_low	26.83	11
literacy_national_zaf	denominator_unweighted	14.29	3
literacy_national_zaf	denominator_weighted	14.29	3
anthropometry_national_zaf	denominator_unweighted	13.16	5
anthropometry_national_zaf	denominator_weighted	13.16	5
iycf_national_zaf	denominator_unweighted	13.04	3
iycf_national_zaf	denominator_weighted	13.04	3
access-to-health-care_national_zaf	denominator_unweighted	12.68	35
access-to-health-care_national_zaf	denominator_weighted	12.68	35
toilet-facilities_national_zaf	denominator_unweighted	10.64	5

dataset	column	missing_pct	missing_count
toilet-facilities_national_zaf	denominator_weighted	10.64	5
covid-19-prevention_national_zaf	denominator_unweighted	8.57	3
covid-19-prevention_national_zaf	denominator_weighted	8.57	3
immunization_national_zaf	denominator_unweighted	7.69	9
immunization_national_zaf	denominator_weighted	7.69	9
water_national_zaf	denominator_unweighted	4.95	5
water_national_zaf	denominator_weighted	4.95	5
literacy_national_zaf	characteristic_category	4.76	1
literacy_national_zaf	characteristic_id	4.76	1
literacy_national_zaf	characteristic_label	4.76	1
literacy_national_zaf	characteristic_order	4.76	1
literacy_national_zaf	dhs_country_code	4.76	1
literacy_national_zaf	indicator_order	4.76	1
literacy_national_zaf	indicator_type	4.76	1
literacy_national_zaf	is_preferred	4.76	1
literacy_national_zaf	is_total	4.76	1
literacy_national_zaf	sdrid	4.76	1
literacy_national_zaf	survey_type	4.76	1
literacy_national_zaf	survey_year_label	4.76	1
access-to-health-care_national_zaf	by_variable_label	4.71	13
maternal-mortality_national_zaf	characteristic_category	4.55	1
maternal-mortality_national_zaf	characteristic_id	4.55	1
maternal-mortality_national_zaf	characteristic_label	4.55	1
maternal-mortality_national_zaf	characteristic_order	4.55	1
maternal-mortality_national_zaf	dhs_country_code	4.55	1
maternal-mortality_national_zaf	indicator_order	4.55	1
maternal-mortality_national_zaf	indicator_type	4.55	1
maternal-mortality_national_zaf	is_preferred	4.55	1
maternal-mortality_national_zaf	is_total	4.55	1
maternal-mortality_national_zaf	sdrid	4.55	1
maternal-mortality_national_zaf	survey_type	4.55	1
maternal-mortality_national_zaf	survey_year_label	4.55	1
iycf_national_zaf	characteristic_category	4.35	1
iycf_national_zaf	characteristic_id	4.35	1
iycf_national_zaf	characteristic_label	4.35	1
iycf_national_zaf	characteristic_order	4.35	1
iycf_national_zaf	dhs_country_code	4.35	1
iycf_national_zaf	indicator_order	4.35	1
iycf_national_zaf	indicator_type	4.35	1
iycf_national_zaf	is_preferred	4.35	1
iycf_national_zaf	is_total	4.35	1

dataset	column	missing_pct	missing_count
iycf_national_zaf	sdrid	4.35	1
iycf_national_zaf	survey_type	4.35	1
iycf_national_zaf	survey_year_label	4.35	1
symptoms-of-acute-respiratory-infection-ari_national_zaf	characteristic_category	3.70	1
symptoms-of-acute-respiratory-infection-ari_national_zaf	characteristic_id	3.70	1
symptoms-of-acute-respiratory-infection-ari_national_zaf	characteristic_label	3.70	1
symptoms-of-acute-respiratory-infection-ari_national_zaf	characteristic_order	3.70	1
symptoms-of-acute-respiratory-infection-ari_national_zaf	dhs_country_code	3.70	1
symptoms-of-acute-respiratory-infection-ari_national_zaf	indicator_order	3.70	1
symptoms-of-acute-respiratory-infection-ari_national_zaf	indicator_type	3.70	1
symptoms-of-acute-respiratory-infection-ari_national_zaf	is_preferred	3.70	1
symptoms-of-acute-respiratory-infection-ari_national_zaf	is_total	3.70	1
symptoms-of-acute-respiratory-infection-ari_national_zaf	sdrid	3.70	1
symptoms-of-acute-respiratory-infection-ari_national_zaf	survey_type	3.70	1
symptoms-of-acute-respiratory-infection-ari_national_zaf	survey_year_label	3.70	1
covid-19-prevention_national_zaf	characteristic_category	2.86	1
covid-19-prevention_national_zaf	characteristic_id	2.86	1
covid-19-prevention_national_zaf	characteristic_label	2.86	1
covid-19-prevention_national_zaf	characteristic_order	2.86	1
covid-19-prevention_national_zaf	dhs_country_code	2.86	1
covid-19-prevention_national_zaf	indicator_order	2.86	1
covid-19-prevention_national_zaf	indicator_type	2.86	1
covid-19-prevention_national_zaf	is_preferred	2.86	1
covid-19-prevention_national_zaf	is_total	2.86	1
covid-19-prevention_national_zaf	sdrid	2.86	1
covid-19-prevention_national_zaf	survey_type	2.86	1
covid-19-prevention_national_zaf	survey_year_label	2.86	1
anthropometry_national_zaf	characteristic_category	2.63	1
anthropometry_national_zaf	characteristic_id	2.63	1
anthropometry_national_zaf	characteristic_label	2.63	1
anthropometry_national_zaf	characteristic_order	2.63	1
anthropometry_national_zaf	dhs_country_code	2.63	1
anthropometry_national_zaf	indicator_order	2.63	1
anthropometry_national_zaf	indicator_type	2.63	1
anthropometry_national_zaf	is_preferred	2.63	1
anthropometry_national_zaf	is_total	2.63	1
anthropometry_national_zaf	sdrid	2.63	1
anthropometry_national_zaf	survey_type	2.63	1
anthropometry_national_zaf	survey_year_label	2.63	1
child-mortality-rates_national_zaf	characteristic_category	2.44	1
child-mortality-rates_national_zaf	characteristic_id	2.44	1



dataset	column	missing_pct	missing_count
child-mortality-rates_national_zaf	characteristic_label	2.44	1
child-mortality-rates_national_zaf	characteristic_order	2.44	1
child-mortality-rates_national_zaf	dhs_country_code	2.44	1
child-mortality-rates_national_zaf	indicator_order	2.44	1
child-mortality-rates_national_zaf	indicator_type	2.44	1
child-mortality-rates_national_zaf	is_preferred	2.44	1
child-mortality-rates_national_zaf	is_total	2.44	1
child-mortality-rates_national_zaf	sdrid	2.44	1
child-mortality-rates_national_zaf	survey_type	2.44	1
child-mortality-rates_national_zaf	survey_year_label	2.44	1
toilet-facilities_national_zaf	characteristic_category	2.13	1
toilet-facilities_national_zaf	characteristic_id	2.13	1
toilet-facilities_national_zaf	characteristic_label	2.13	1
toilet-facilities_national_zaf	characteristic_order	2.13	1
toilet-facilities_national_zaf	dhs_country_code	2.13	1
toilet-facilities_national_zaf	indicator_order	2.13	1
toilet-facilities_national_zaf	indicator_type	2.13	1
toilet-facilities_national_zaf	is_preferred	2.13	1
toilet-facilities_national_zaf	is_total	2.13	1
toilet-facilities_national_zaf	sdrid	2.13	1
toilet-facilities_national_zaf	survey_type	2.13	1
toilet-facilities_national_zaf	survey_year_label	2.13	1
dhs-quickstats_national_zaf	characteristic_category	1.89	1
dhs-quickstats_national_zaf	characteristic_id	1.89	1
dhs-quickstats_national_zaf	characteristic_label	1.89	1
dhs-quickstats_national_zaf	characteristic_order	1.89	1
dhs-quickstats_national_zaf	dhs_country_code	1.89	1
dhs-quickstats_national_zaf	indicator_order	1.89	1
dhs-quickstats_national_zaf	indicator_type	1.89	1
dhs-quickstats_national_zaf	is_preferred	1.89	1
dhs-quickstats_national_zaf	is_total	1.89	1
dhs-quickstats_national_zaf	sdrid	1.89	1
dhs-quickstats_national_zaf	survey_type	1.89	1
dhs-quickstats_national_zaf	survey_year_label	1.89	1
water_national_zaf	characteristic_category	0.99	1
water_national_zaf	characteristic_id	0.99	1
water_national_zaf	characteristic_label	0.99	1
water_national_zaf	characteristic_order	0.99	1
water_national_zaf	dhs_country_code	0.99	1
water_national_zaf	indicator_order	0.99	1
water_national_zaf	indicator_type	0.99	1

dataset	column	missing_pct	missing_count
water_national_zaf	is_preferred	0.99	1
water_national_zaf	is_total	0.99	1
water_national_zaf	sdrid	0.99	1
water_national_zaf	survey_type	0.99	1
water_national_zaf	survey_year_label	0.99	1
immunization_national_zaf	characteristic_category	0.85	1
immunization_national_zaf	characteristic_id	0.85	1
immunization_national_zaf	characteristic_label	0.85	1
immunization_national_zaf	characteristic_order	0.85	1
immunization_national_zaf	dhs_country_code	0.85	1
immunization_national_zaf	indicator_order	0.85	1
immunization_national_zaf	indicator_type	0.85	1
immunization_national_zaf	is_preferred	0.85	1
immunization_national_zaf	is_total	0.85	1
immunization_national_zaf	sdrid	0.85	1
immunization_national_zaf	survey_type	0.85	1
immunization_national_zaf	survey_year_label	0.85	1
hiv-behavior_national_zaf	characteristic_category	0.84	1
hiv-behavior_national_zaf	characteristic_id	0.84	1
hiv-behavior_national_zaf	characteristic_label	0.84	1
hiv-behavior_national_zaf	characteristic_order	0.84	1
hiv-behavior_national_zaf	dhs_country_code	0.84	1
hiv-behavior_national_zaf	indicator_order	0.84	1
hiv-behavior_national_zaf	indicator_type	0.84	1
hiv-behavior_national_zaf	is_preferred	0.84	1
hiv-behavior_national_zaf	is_total	0.84	1
hiv-behavior_national_zaf	sdrid	0.84	1
hiv-behavior_national_zaf	survey_type	0.84	1
hiv-behavior_national_zaf	survey_year_label	0.84	1
access-to-health-care_national_zaf	characteristic_category	0.36	1
access-to-health-care_national_zaf	characteristic_id	0.36	1
access-to-health-care_national_zaf	characteristic_label	0.36	1
access-to-health-care_national_zaf	characteristic_order	0.36	1
access-to-health-care_national_zaf	dhs_country_code	0.36	1
access-to-health-care_national_zaf	indicator_order	0.36	1
access-to-health-care_national_zaf	indicator_type	0.36	1
access-to-health-care_national_zaf	is_preferred	0.36	1
access-to-health-care_national_zaf	is_total	0.36	1
access-to-health-care_national_zaf	sdrid	0.36	1
access-to-health-care_national_zaf	survey_type	0.36	1
access-to-health-care_national_zaf	survey_year_label	0.36	1

dataset	column	missing_pct	missing_count
access-to-health-care_national_zaf	by_variable_id	0.00	0
access-to-health-care_national_zaf	country_name	0.00	0
access-to-health-care_national_zaf	data_id	0.00	0
access-to-health-care_national_zaf	indicator	0.00	0
access-to-health-care_national_zaf	indicator_id	0.00	0
access-to-health-care_national_zaf	iso3	0.00	0
access-to-health-care_national_zaf	precision	0.00	0
access-to-health-care_national_zaf	survey_id	0.00	0
access-to-health-care_national_zaf	survey_year	0.00	0
access-to-health-care_national_zaf	value	0.00	0
anthropometry_national_zaf	by_variable_id	0.00	0
anthropometry_national_zaf	country_name	0.00	0
anthropometry_national_zaf	data_id	0.00	0
anthropometry_national_zaf	indicator	0.00	0
anthropometry_national_zaf	indicator_id	0.00	0
anthropometry_national_zaf	iso3	0.00	0
anthropometry_national_zaf	precision	0.00	0
anthropometry_national_zaf	survey_id	0.00	0
anthropometry_national_zaf	survey_year	0.00	0
anthropometry_national_zaf	value	0.00	0
child-mortality-rates_national_zaf	by_variable_id	0.00	0
child-mortality-rates_national_zaf	country_name	0.00	0
child-mortality-rates_national_zaf	data_id	0.00	0
child-mortality-rates_national_zaf	indicator	0.00	0
child-mortality-rates_national_zaf	indicator_id	0.00	0
child-mortality-rates_national_zaf	iso3	0.00	0
child-mortality-rates_national_zaf	precision	0.00	0
child-mortality-rates_national_zaf	survey_id	0.00	0
child-mortality-rates_national_zaf	survey_year	0.00	0
child-mortality-rates_national_zaf	value	0.00	0
covid-19-prevention_national_zaf	by_variable_id	0.00	0
covid-19-prevention_national_zaf	country_name	0.00	0
covid-19-prevention_national_zaf	data_id	0.00	0
covid-19-prevention_national_zaf	indicator	0.00	0
covid-19-prevention_national_zaf	indicator_id	0.00	0
covid-19-prevention_national_zaf	iso3	0.00	0
covid-19-prevention_national_zaf	precision	0.00	0
covid-19-prevention_national_zaf	survey_id	0.00	0
covid-19-prevention_national_zaf	survey_year	0.00	0
covid-19-prevention_national_zaf	value	0.00	0
dhs-quickstats_national_zaf	by_variable_id	0.00	0

dataset	column	missing_pct	missing_count
dhs-quickstats_national_zaf	country_name	0.00	0
dhs-quickstats_national_zaf	data_id	0.00	0
dhs-quickstats_national_zaf	indicator	0.00	0
dhs-quickstats_national_zaf	indicator_id	0.00	0
dhs-quickstats_national_zaf	iso3	0.00	0
dhs-quickstats_national_zaf	precision	0.00	0
dhs-quickstats_national_zaf	survey_id	0.00	0
dhs-quickstats_national_zaf	survey_year	0.00	0
dhs-quickstats_national_zaf	value	0.00	0
hiv-behavior_national_zaf	by_variable_id	0.00	0
hiv-behavior_national_zaf	country_name	0.00	0
hiv-behavior_national_zaf	data_id	0.00	0
hiv-behavior_national_zaf	indicator	0.00	0
hiv-behavior_national_zaf	indicator_id	0.00	0
hiv-behavior_national_zaf	iso3	0.00	0
hiv-behavior_national_zaf	precision	0.00	0
hiv-behavior_national_zaf	survey_id	0.00	0
hiv-behavior_national_zaf	survey_year	0.00	0
hiv-behavior_national_zaf	value	0.00	0
immunization_national_zaf	by_variable_id	0.00	0
immunization_national_zaf	country_name	0.00	0
immunization_national_zaf	data_id	0.00	0
immunization_national_zaf	indicator	0.00	0
immunization_national_zaf	indicator_id	0.00	0
immunization_national_zaf	iso3	0.00	0
immunization_national_zaf	precision	0.00	0
immunization_national_zaf	survey_id	0.00	0
immunization_national_zaf	survey_year	0.00	0
immunization_national_zaf	value	0.00	0
iycf_national_zaf	by_variable_id	0.00	0
iycf_national_zaf	country_name	0.00	0
iycf_national_zaf	data_id	0.00	0
iycf_national_zaf	indicator	0.00	0
iycf_national_zaf	indicator_id	0.00	0
iycf_national_zaf	iso3	0.00	0
iycf_national_zaf	precision	0.00	0
iycf_national_zaf	survey_id	0.00	0
iycf_national_zaf	survey_year	0.00	0
iycf_national_zaf	value	0.00	0
literacy_national_zaf	by_variable_id	0.00	0
literacy_national_zaf	country_name	0.00	0

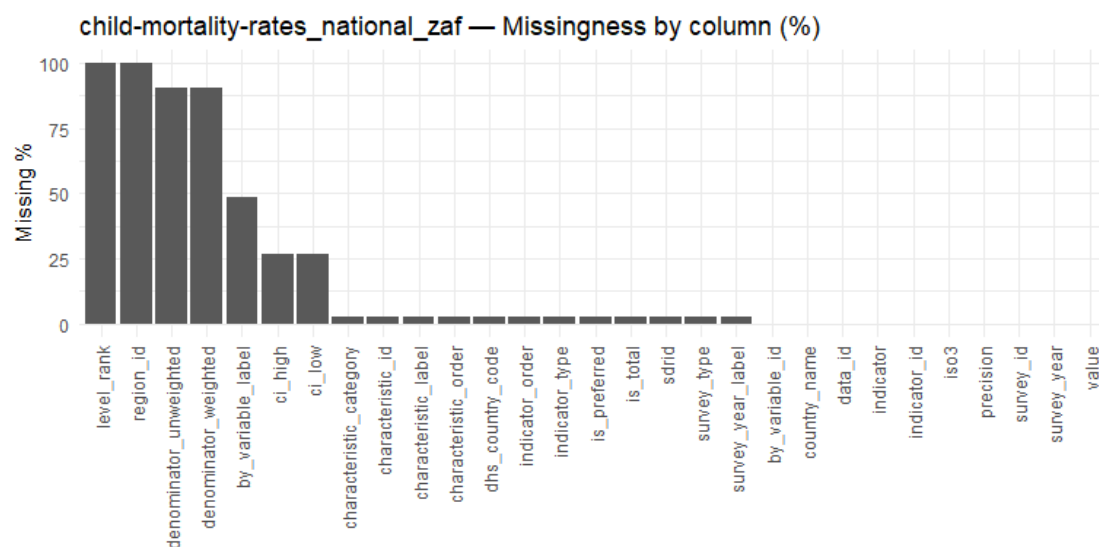
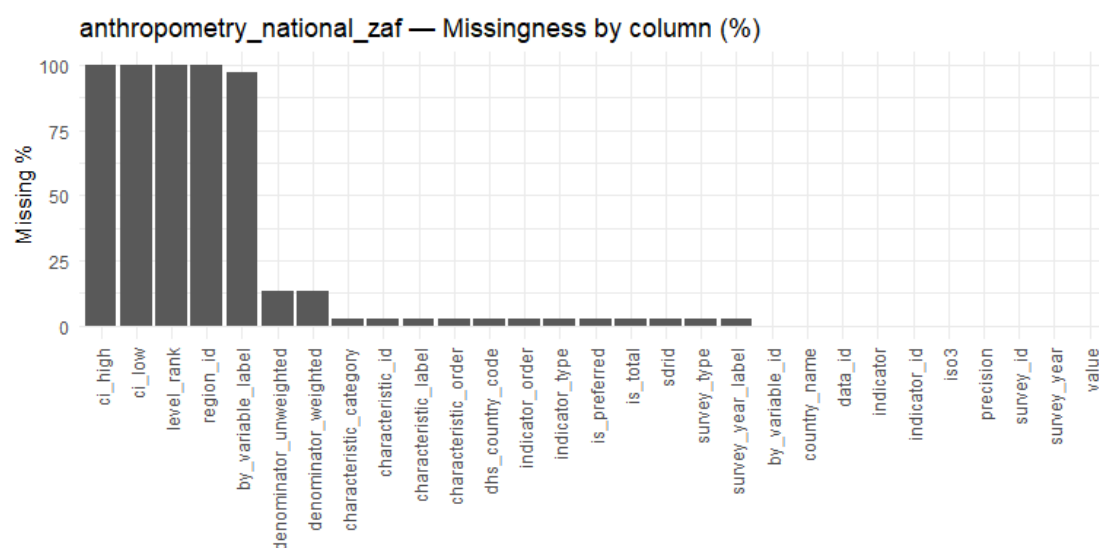
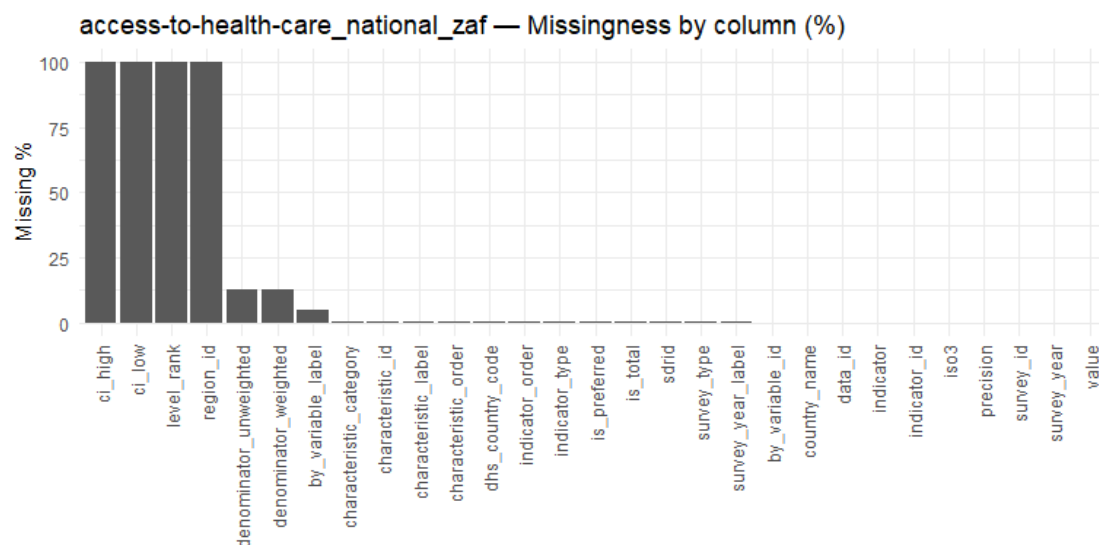
dataset	column	missing_pct	missing_count
literacy_national_zaf	data_id	0.00	0
literacy_national_zaf	indicator	0.00	0
literacy_national_zaf	indicator_id	0.00	0
literacy_national_zaf	iso3	0.00	0
literacy_national_zaf	precision	0.00	0
literacy_national_zaf	survey_id	0.00	0
literacy_national_zaf	survey_year	0.00	0
literacy_national_zaf	value	0.00	0
maternal-mortality_national_zaf	by_variable_id	0.00	0
maternal-mortality_national_zaf	country_name	0.00	0
maternal-mortality_national_zaf	data_id	0.00	0
maternal-mortality_national_zaf	indicator	0.00	0
maternal-mortality_national_zaf	indicator_id	0.00	0
maternal-mortality_national_zaf	iso3	0.00	0
maternal-mortality_national_zaf	precision	0.00	0
maternal-mortality_national_zaf	survey_id	0.00	0
maternal-mortality_national_zaf	survey_year	0.00	0
maternal-mortality_national_zaf	value	0.00	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	by_variable_id	0.00	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	by_variable_label	0.00	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	country_name	0.00	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	data_id	0.00	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	indicator	0.00	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	indicator_id	0.00	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	iso3	0.00	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	precision	0.00	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	survey_id	0.00	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	survey_year	0.00	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	value	0.00	0
toilet-facilities_national_zaf	by_variable_id	0.00	0
toilet-facilities_national_zaf	country_name	0.00	0
toilet-facilities_national_zaf	data_id	0.00	0
toilet-facilities_national_zaf	indicator	0.00	0
toilet-facilities_national_zaf	indicator_id	0.00	0
toilet-facilities_national_zaf	iso3	0.00	0
toilet-facilities_national_zaf	precision	0.00	0
toilet-facilities_national_zaf	survey_id	0.00	0
toilet-facilities_national_zaf	survey_year	0.00	0
toilet-facilities_national_zaf	value	0.00	0
water_national_zaf	by_variable_id	0.00	0
water_national_zaf	country_name	0.00	0

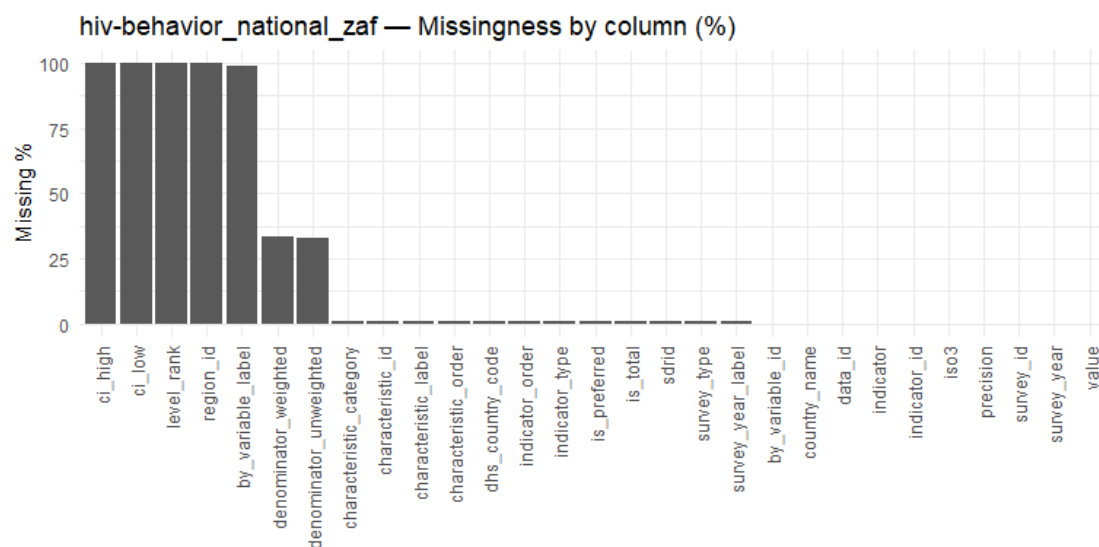
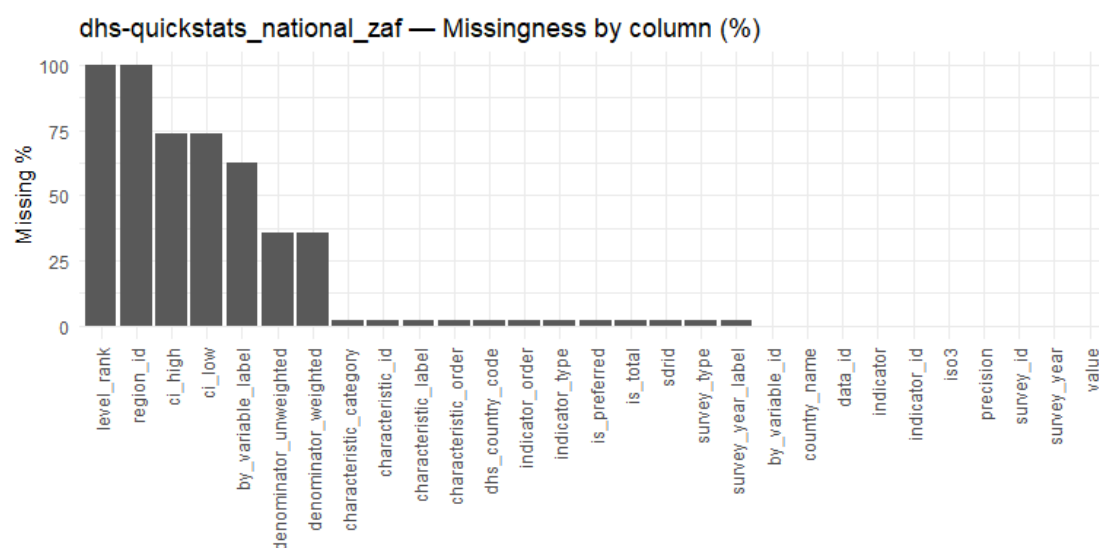
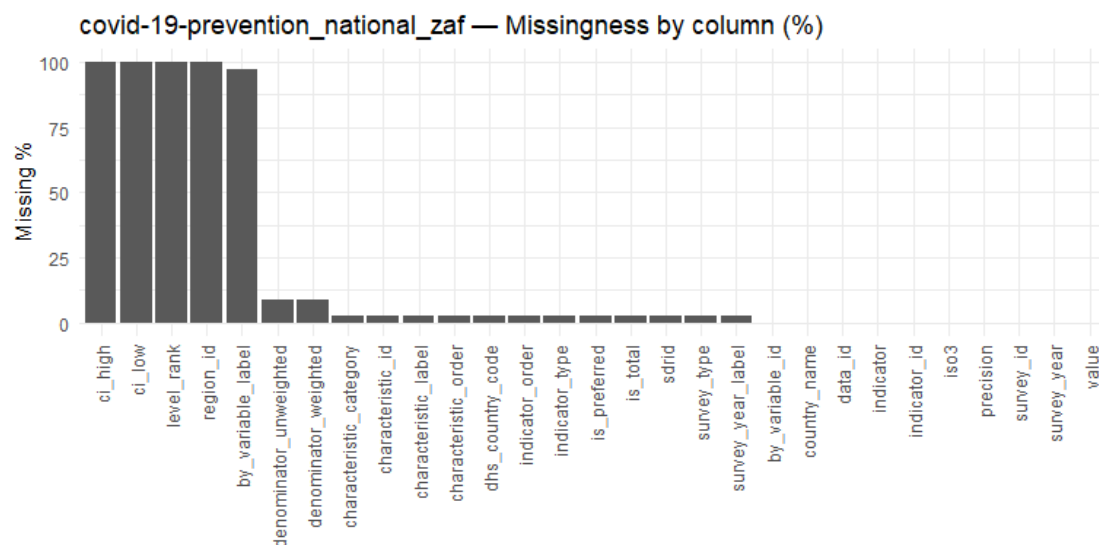
dataset	column	missing_pct	missing_count
water_national_zaf	data_id	0.00	0
water_national_zaf	indicator	0.00	0
water_national_zaf	indicator_id	0.00	0
water_national_zaf	iso3	0.00	0
water_national_zaf	precision	0.00	0
water_national_zaf	survey_id	0.00	0
water_national_zaf	survey_year	0.00	0
water_national_zaf	value	0.00	0

```
if (write_exports) {
  readr::write_csv(missingness_all, "M1_missingness_all_columns.csv")
}
```

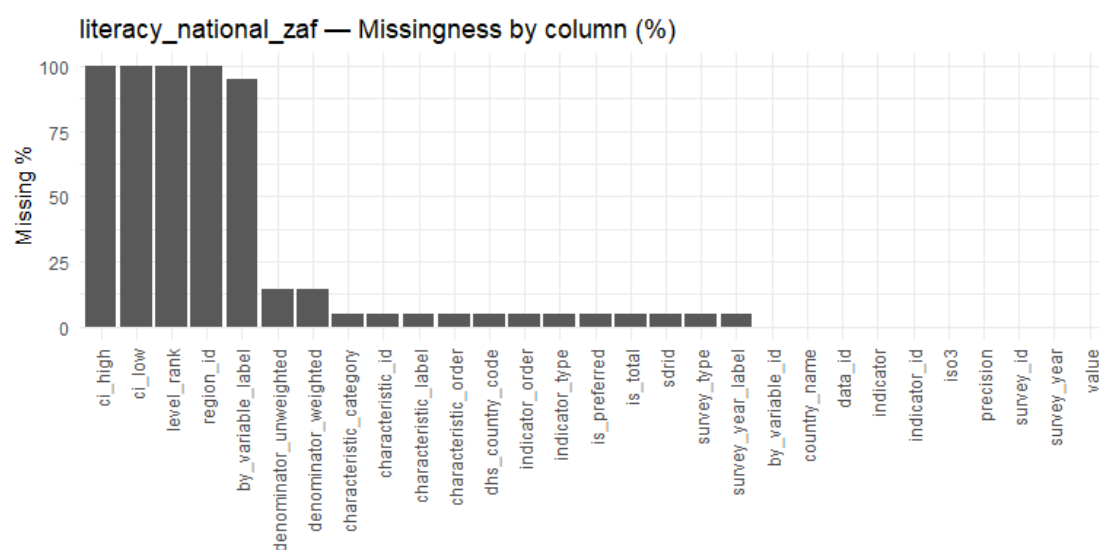
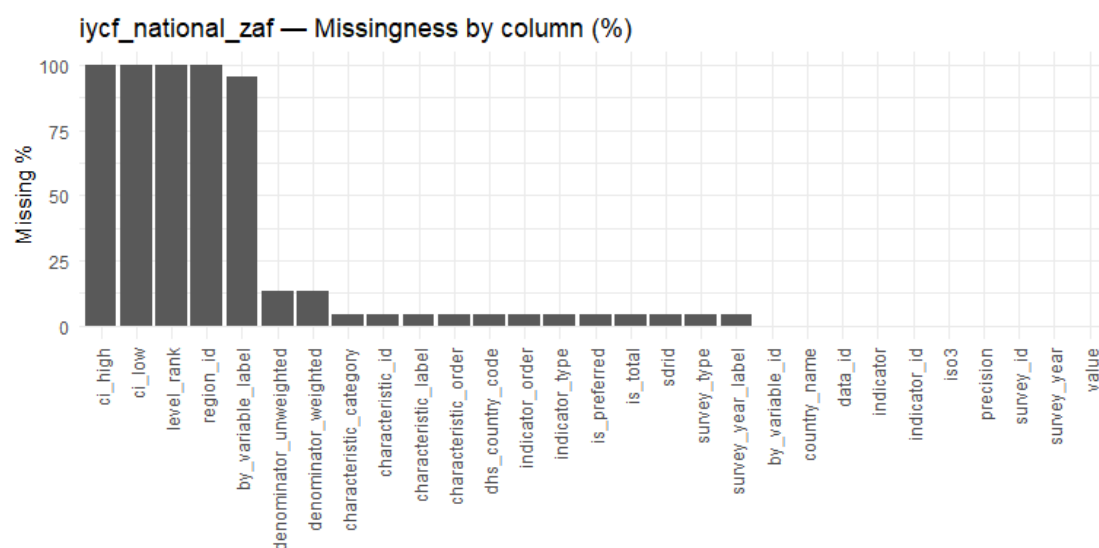
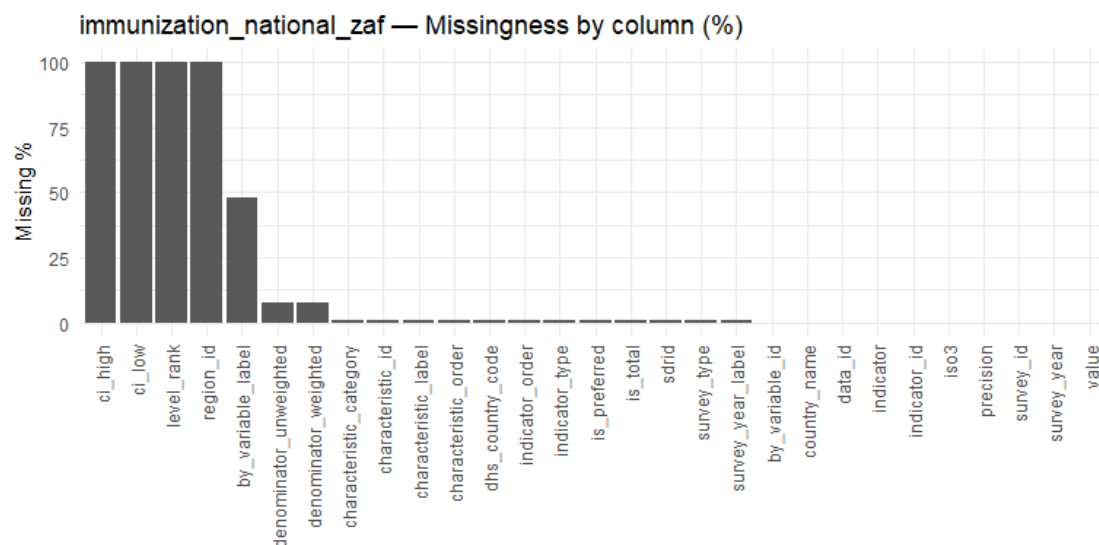
### 3.1.1 Missingness Charts (Per Dataset)

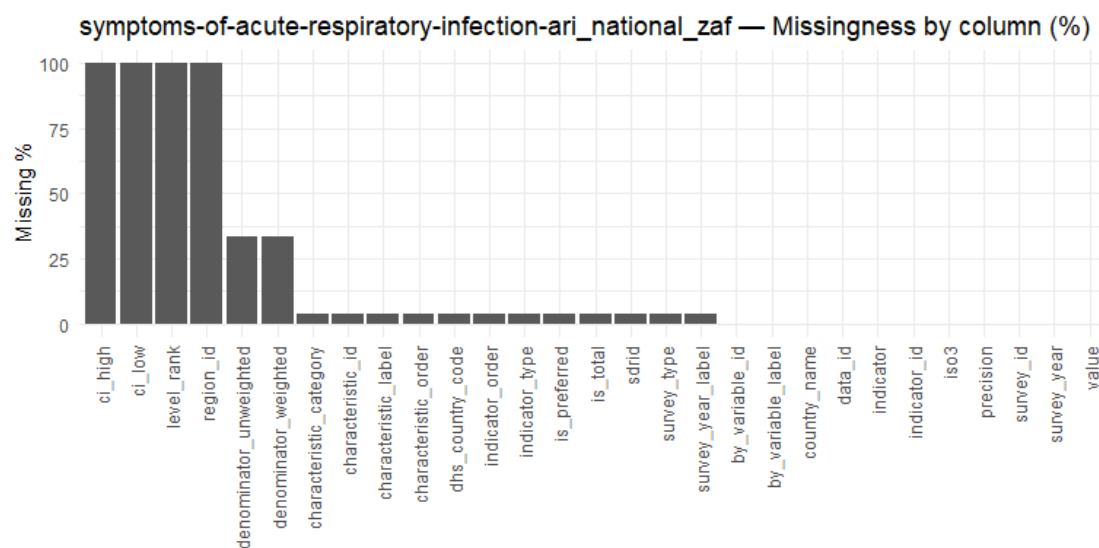
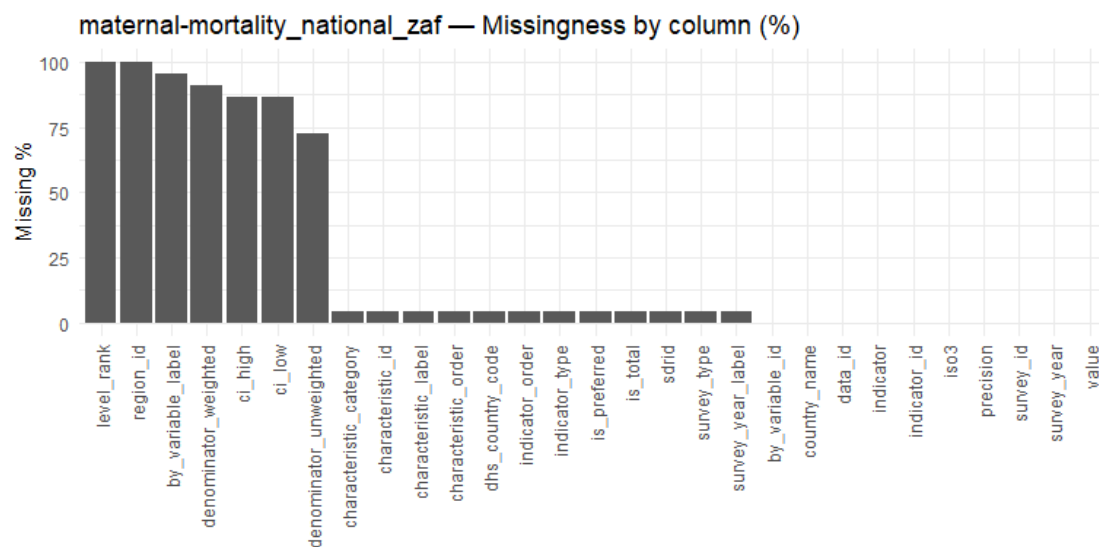
```
plot_missingness <- function(df, nm){
  miss <- tibble(column = names(df), missing_pct =
round(colMeans(is.na(df))*100, 2))
  ggplot(miss, aes(x = reorder(column, -missing_pct), y = missing_pct)) +
    geom_col() +
    coord_cartesian(ylim = c(0, 100)) +
    labs(title = paste0(nm, " - Missingness by column (%)", x = NULL, y =
"Missing %") +
    theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
}
purrr::iwalk(dfs, function(df, nm) print(plot_missingness(df, nm)))
```

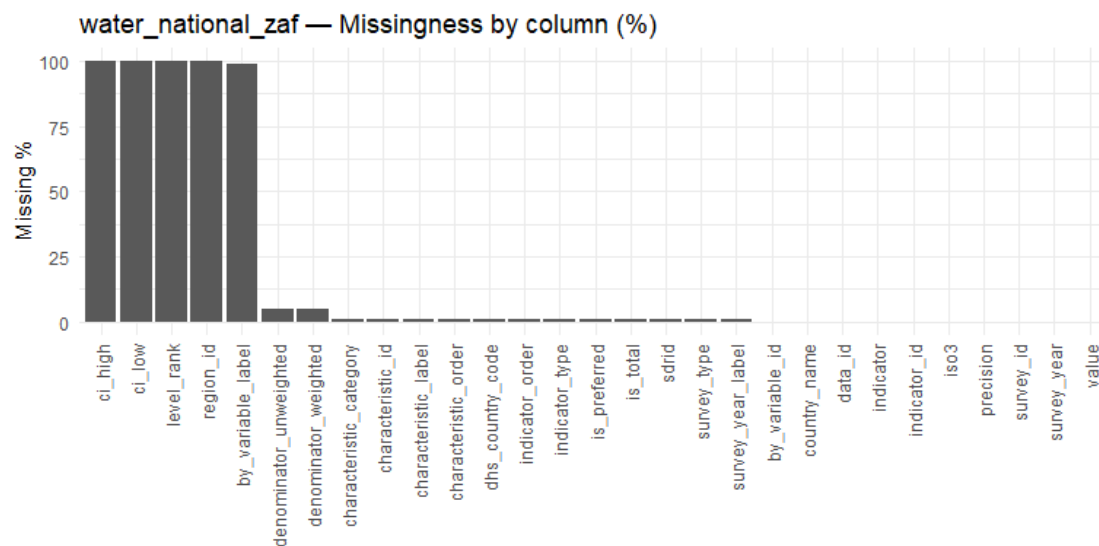
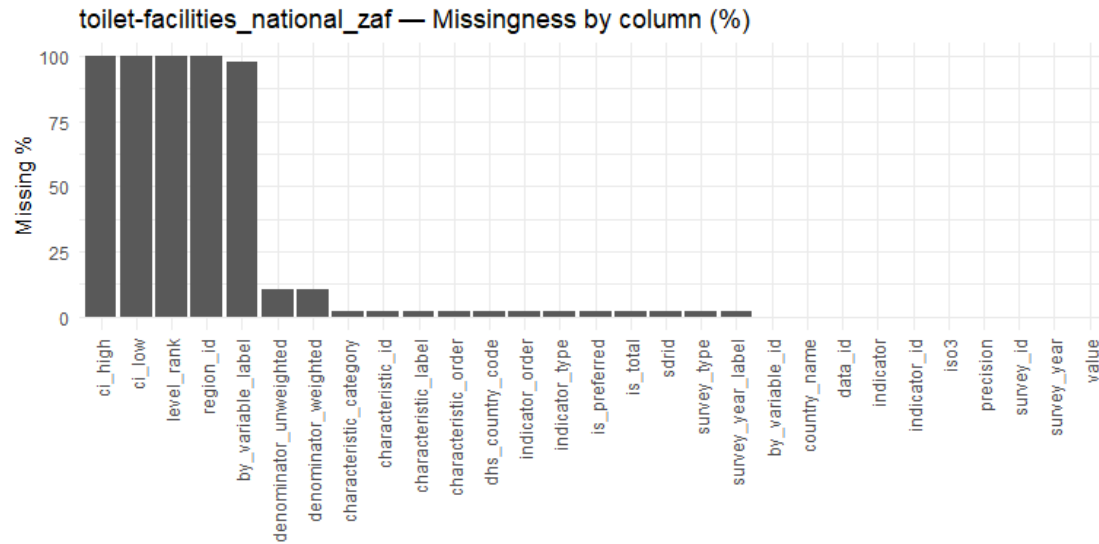












## 3.2 Duplicates (Row-Level)

```
# Count duplicates per dataset (across all columns)
dup_summary <- purrr::imap_dfr(dfs, function(df, nm){
  tibble(dataset = nm, duplicate_rows = sum(duplicated(df)))
}) |> arrange(desc(duplicate_rows))

gt::gt(dup_summary)
```

dataset	duplicate_rows
access-to-health-care_national_zaf	0
anthropometry_national_zaf	0
child-mortality-rates_national_zaf	0
covid-19-prevention_national_zaf	0
dhs-quickstats_national_zaf	0
hiv-behavior_national_zaf	0
immunization_national_zaf	0
iycf_national_zaf	0
literacy_national_zaf	0
maternal-mortality_national_zaf	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	0
toilet-facilities_national_zaf	0
water_national_zaf	0

```

if (write_exports) {
  readr::write_csv(dup_summary, "M1_duplicates_by_dataset.csv")
}

# Optionally show duplicate rows when counts are manageable
purrr::iwalk(dfs, function(df, nm){
  dup_ct <- sum(duplicated(df))
  if (dup_ct > 0 && dup_ct <= dup_print_threshold) {
    cat("\n\n### Duplicate Rows -", nm, "(showing all duplicates because
count <=", dup_print_threshold, ")\n\n")
    dups <- df[duplicated(df) | duplicated(df, fromLast = TRUE), , drop =
FALSE]
    print(dups)
    if (write_exports) readr::write_csv(dups, paste0("M1_duplicates__", nm,
".csv"))
  } else if (dup_ct > dup_print_threshold) {
    cat("\n\n### Duplicate Rows -", nm, "(too many to print; exporting if
write_exports=TRUE)\n\n")
    if (write_exports) {
      dups <- df[duplicated(df) | duplicated(df, fromLast = TRUE), , drop =
FALSE]
      readr::write_csv(dups, paste0("M1_duplicates__", nm, ".csv"))
    }
  }
})

```

### 3.3 Outliers (Numeric Columns, $|z| > 3$ )

```

outlier_counts <- function(df){
  nums <- df |> dplyr::select(where(is.numeric))
  if(ncol(nums) == 0) return(tibble(column=character(),
outliers_abs_z_gt_3=integer()))
  purrr::map_dfr(names(nums), function(col){

```

```

    v <- nums[[col]]
    v <- v[!is.na(v)]
    if(length(v) < 5 || sd(v) == 0) return(tibble(column = col,
outliers_abs_z_gt_3 = 0L))
    z <- (v - mean(v)) / sd(v)
    tibble(column = col, outliers_abs_z_gt_3 = as.integer(sum(abs(z) > 3)))
  }) |> arrange(desc(outliers_abs_z_gt_3))
}

outliers_by_dataset <- purrr::imap(dfs, function(df, nm){
  oc <- outlier_counts(df) |> mutate(dataset = nm, .before = 1)
  oc
})
outliers_all <- bind_rows(outliers_by_dataset)

if(nrow(outliers_all) > 0){
  gt::gt(outliers_all)
} else {
  cat("No numeric columns suitable for outlier analysis were found.")
}

```

dataset	column	outliers_abs_z_gt_3
access-to-health-care_national_zaf	indicator_order	0
access-to-health-care_national_zaf	characteristic_id	0
access-to-health-care_national_zaf	characteristic_order	0
access-to-health-care_national_zaf	is_total	0
access-to-health-care_national_zaf	is_preferred	0
access-to-health-care_national_zaf	survey_year_label	0
access-to-health-care_national_zaf	denominator_weighted	0
access-to-health-care_national_zaf	denominator_unweighted	0
anthropometry_national_zaf	indicator_order	0
anthropometry_national_zaf	characteristic_id	0
anthropometry_national_zaf	characteristic_order	0
anthropometry_national_zaf	is_total	0
anthropometry_national_zaf	is_preferred	0
anthropometry_national_zaf	survey_year_label	0
anthropometry_national_zaf	denominator_weighted	0
anthropometry_national_zaf	denominator_unweighted	0
child-mortality-rates_national_zaf	indicator_order	0
child-mortality-rates_national_zaf	characteristic_id	0
child-mortality-rates_national_zaf	characteristic_order	0
child-mortality-rates_national_zaf	is_total	0
child-mortality-rates_national_zaf	is_preferred	0
child-mortality-rates_national_zaf	survey_year_label	0
child-mortality-rates_national_zaf	denominator_weighted	0
child-mortality-rates_national_zaf	denominator_unweighted	0
child-mortality-rates_national_zaf	ci_low	0
child-mortality-rates_national_zaf	ci_high	0
covid-19-prevention_national_zaf	indicator_order	0
covid-19-prevention_national_zaf	characteristic_id	0
covid-19-prevention_national_zaf	characteristic_order	0
covid-19-prevention_national_zaf	is_total	0
covid-19-prevention_national_zaf	is_preferred	0
covid-19-prevention_national_zaf	survey_year_label	0
covid-19-prevention_national_zaf	denominator_weighted	0
covid-19-prevention_national_zaf	denominator_unweighted	0
dhs-quickstats_national_zaf	indicator_order	0
dhs-quickstats_national_zaf	characteristic_id	0
dhs-quickstats_national_zaf	characteristic_order	0
dhs-quickstats_national_zaf	is_total	0
dhs-quickstats_national_zaf	is_preferred	0
dhs-quickstats_national_zaf	survey_year_label	0
dhs-quickstats_national_zaf	denominator_weighted	0

dataset	column	outliers_abs_z_gt_3
dhs-quickstats_national_zaf	denominator_unweighted	0
dhs-quickstats_national_zaf	ci_low	0
dhs-quickstats_national_zaf	ci_high	0
hiv-behavior_national_zaf	indicator_order	0
hiv-behavior_national_zaf	characteristic_id	0
hiv-behavior_national_zaf	characteristic_order	0
hiv-behavior_national_zaf	is_total	0
hiv-behavior_national_zaf	is_preferred	0
hiv-behavior_national_zaf	survey_year_label	0
hiv-behavior_national_zaf	denominator_weighted	0
hiv-behavior_national_zaf	denominator_unweighted	0
immunization_national_zaf	indicator_order	0
immunization_national_zaf	characteristic_id	0
immunization_national_zaf	characteristic_order	0
immunization_national_zaf	is_total	0
immunization_national_zaf	is_preferred	0
immunization_national_zaf	survey_year_label	0
immunization_national_zaf	denominator_weighted	0
immunization_national_zaf	denominator_unweighted	0
iycf_national_zaf	indicator_order	0
iycf_national_zaf	characteristic_id	0
iycf_national_zaf	characteristic_order	0
iycf_national_zaf	is_total	0
iycf_national_zaf	is_preferred	0
iycf_national_zaf	survey_year_label	0
iycf_national_zaf	denominator_weighted	0
iycf_national_zaf	denominator_unweighted	0
literacy_national_zaf	indicator_order	0
literacy_national_zaf	characteristic_id	0
literacy_national_zaf	characteristic_order	0
literacy_national_zaf	is_total	0
literacy_national_zaf	is_preferred	0
literacy_national_zaf	survey_year_label	0
literacy_national_zaf	denominator_weighted	0
literacy_national_zaf	denominator_unweighted	0
maternal-mortality_national_zaf	indicator_order	0
maternal-mortality_national_zaf	characteristic_id	0
maternal-mortality_national_zaf	characteristic_order	0
maternal-mortality_national_zaf	is_total	0
maternal-mortality_national_zaf	is_preferred	0
maternal-mortality_national_zaf	survey_year_label	0

dataset	column	outliers_abs_z_gt_3
maternal-mortality_national_zaf	denominator_weighted	0
maternal-mortality_national_zaf	denominator_unweighted	0
maternal-mortality_national_zaf	ci_low	0
maternal-mortality_national_zaf	ci_high	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	indicator_order	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	characteristic_id	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	characteristic_order	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	is_total	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	is_preferred	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	survey_year_label	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	denominator_weighted	0
symptoms-of-acute-respiratory-infection-ari_national_zaf	denominator_unweighted	0
toilet-facilities_national_zaf	indicator_order	0
toilet-facilities_national_zaf	characteristic_id	0
toilet-facilities_national_zaf	characteristic_order	0
toilet-facilities_national_zaf	is_total	0
toilet-facilities_national_zaf	is_preferred	0
toilet-facilities_national_zaf	survey_year_label	0
toilet-facilities_national_zaf	denominator_weighted	0
toilet-facilities_national_zaf	denominator_unweighted	0
water_national_zaf	indicator_order	0
water_national_zaf	characteristic_id	0
water_national_zaf	characteristic_order	0
water_national_zaf	is_total	0
water_national_zaf	is_preferred	0
water_national_zaf	survey_year_label	0
water_national_zaf	denominator_weighted	0
water_national_zaf	denominator_unweighted	0

```
if (write_exports) {
  readr::write_csv(outliers_all, "M1_outliers_all_numeric_columns.csv")
}
```

### 3.4 Consolidated Data Quality Issues Log

```
# Build a tidy issues log: one row per issue instance
issues_missing <- missingness_all |>
  filter(missing_count > 0) |>
  transmute(dataset, issue_type = "missing", column, detail =
paste0(missing_pct, "% (", missing_count, " cells)"))

issues_dup <- dup_summary |>
  filter(duplicate_rows > 0) |>
  transmute(dataset, issue_type = "duplicates", column = NA_character_,
detail = paste0(duplicate_rows, " duplicate rows"))
```



```
issues_outliers <- outliers_all |>
  filter(outliers_abs_z_gt_3 > 0) |>
  transmute(dataset, issue_type = "outliers", column, detail =
paste0(outliers_abs_z_gt_3, " outliers (|z|>3)"))

issues_log <- bind_rows(issues_missing, issues_dup, issues_outliers) |>
  arrange(dataset, issue_type, desc(detail))

gt::gt(issues_log)
```

dataset	issue_type	column	detail
access-to-health-care_national_zaf	missing	by_variable_label	4.71% (13 cells)
access-to-health-care_national_zaf	missing	denominator_unweighted	12.68% (35 cells)
access-to-health-care_national_zaf	missing	denominator_weighted	12.68% (35 cells)
access-to-health-care_national_zaf	missing	ci_high	100% (276 cells)
access-to-health-care_national_zaf	missing	ci_low	100% (276 cells)
access-to-health-care_national_zaf	missing	level_rank	100% (276 cells)
access-to-health-care_national_zaf	missing	region_id	100% (276 cells)
access-to-health-care_national_zaf	missing	characteristic_category	0.36% (1 cells)
access-to-health-care_national_zaf	missing	characteristic_id	0.36% (1 cells)
access-to-health-care_national_zaf	missing	characteristic_label	0.36% (1 cells)
access-to-health-care_national_zaf	missing	characteristic_order	0.36% (1 cells)
access-to-health-care_national_zaf	missing	dhs_country_code	0.36% (1 cells)
access-to-health-care_national_zaf	missing	indicator_order	0.36% (1 cells)
access-to-health-care_national_zaf	missing	indicator_type	0.36% (1 cells)
access-to-health-care_national_zaf	missing	is_preferred	0.36% (1 cells)
access-to-health-care_national_zaf	missing	is_total	0.36% (1 cells)
access-to-health-care_national_zaf	missing	sdrid	0.36% (1 cells)
access-to-health-care_national_zaf	missing	survey_type	0.36% (1 cells)
access-to-health-care_national_zaf	missing	survey_year_label	0.36% (1 cells)
anthropometry_national_zaf	missing	by_variable_label	97.37% (37 cells)
anthropometry_national_zaf	missing	characteristic_category	2.63% (1 cells)
anthropometry_national_zaf	missing	characteristic_id	2.63% (1 cells)
anthropometry_national_zaf	missing	characteristic_label	2.63% (1 cells)
anthropometry_national_zaf	missing	characteristic_order	2.63% (1 cells)
anthropometry_national_zaf	missing	dhs_country_code	2.63% (1 cells)
anthropometry_national_zaf	missing	indicator_order	2.63% (1 cells)
anthropometry_national_zaf	missing	indicator_type	2.63% (1 cells)
anthropometry_national_zaf	missing	is_preferred	2.63% (1 cells)
anthropometry_national_zaf	missing	is_total	2.63% (1 cells)
anthropometry_national_zaf	missing	sdrid	2.63% (1 cells)
anthropometry_national_zaf	missing	survey_type	2.63% (1 cells)
anthropometry_national_zaf	missing	survey_year_label	2.63% (1 cells)
anthropometry_national_zaf	missing	denominator_unweighted	13.16% (5 cells)
anthropometry_national_zaf	missing	denominator_weighted	13.16% (5 cells)
anthropometry_national_zaf	missing	ci_high	100% (38 cells)
anthropometry_national_zaf	missing	ci_low	100% (38 cells)
anthropometry_national_zaf	missing	level_rank	100% (38 cells)
anthropometry_national_zaf	missing	region_id	100% (38 cells)
child-mortality-rates_national_zaf	missing	denominator_unweighted	90.24% (37 cells)
child-mortality-rates_national_zaf	missing	denominator_weighted	90.24% (37 cells)
child-mortality-rates_national_zaf	missing	by_variable_label	48.78% (20 cells)

dataset	issue_type	column	detail
child-mortality-rates_national_zaf	missing	ci_high	26.83% (11 cells)
child-mortality-rates_national_zaf	missing	ci_low	26.83% (11 cells)
child-mortality-rates_national_zaf	missing	characteristic_category	2.44% (1 cells)
child-mortality-rates_national_zaf	missing	characteristic_id	2.44% (1 cells)
child-mortality-rates_national_zaf	missing	characteristic_label	2.44% (1 cells)
child-mortality-rates_national_zaf	missing	characteristic_order	2.44% (1 cells)
child-mortality-rates_national_zaf	missing	dhs_country_code	2.44% (1 cells)
child-mortality-rates_national_zaf	missing	indicator_order	2.44% (1 cells)
child-mortality-rates_national_zaf	missing	indicator_type	2.44% (1 cells)
child-mortality-rates_national_zaf	missing	is_preferred	2.44% (1 cells)
child-mortality-rates_national_zaf	missing	is_total	2.44% (1 cells)
child-mortality-rates_national_zaf	missing	sdrid	2.44% (1 cells)
child-mortality-rates_national_zaf	missing	survey_type	2.44% (1 cells)
child-mortality-rates_national_zaf	missing	survey_year_label	2.44% (1 cells)
child-mortality-rates_national_zaf	missing	level_rank	100% (41 cells)
child-mortality-rates_national_zaf	missing	region_id	100% (41 cells)
covid-19-prevention_national_zaf	missing	by_variable_label	97.14% (34 cells)
covid-19-prevention_national_zaf	missing	denominator_unweighted	8.57% (3 cells)
covid-19-prevention_national_zaf	missing	denominator_weighted	8.57% (3 cells)
covid-19-prevention_national_zaf	missing	characteristic_category	2.86% (1 cells)
covid-19-prevention_national_zaf	missing	characteristic_id	2.86% (1 cells)
covid-19-prevention_national_zaf	missing	characteristic_label	2.86% (1 cells)
covid-19-prevention_national_zaf	missing	characteristic_order	2.86% (1 cells)
covid-19-prevention_national_zaf	missing	dhs_country_code	2.86% (1 cells)
covid-19-prevention_national_zaf	missing	indicator_order	2.86% (1 cells)
covid-19-prevention_national_zaf	missing	indicator_type	2.86% (1 cells)
covid-19-prevention_national_zaf	missing	is_preferred	2.86% (1 cells)
covid-19-prevention_national_zaf	missing	is_total	2.86% (1 cells)
covid-19-prevention_national_zaf	missing	sdrid	2.86% (1 cells)
covid-19-prevention_national_zaf	missing	survey_type	2.86% (1 cells)
covid-19-prevention_national_zaf	missing	survey_year_label	2.86% (1 cells)
covid-19-prevention_national_zaf	missing	ci_high	100% (35 cells)
covid-19-prevention_national_zaf	missing	ci_low	100% (35 cells)
covid-19-prevention_national_zaf	missing	level_rank	100% (35 cells)
covid-19-prevention_national_zaf	missing	region_id	100% (35 cells)
dhs-quickstats_national_zaf	missing	ci_high	73.58% (39 cells)
dhs-quickstats_national_zaf	missing	ci_low	73.58% (39 cells)
dhs-quickstats_national_zaf	missing	by_variable_label	62.26% (33 cells)
dhs-quickstats_national_zaf	missing	denominator_unweighted	35.85% (19 cells)
dhs-quickstats_national_zaf	missing	denominator_weighted	35.85% (19 cells)
dhs-quickstats_national_zaf	missing	level_rank	100% (53 cells)

dataset	issue_type	column	detail
dhs-quickstats_national_zaf	missing	region_id	100% (53 cells)
dhs-quickstats_national_zaf	missing	characteristic_category	1.89% (1 cells)
dhs-quickstats_national_zaf	missing	characteristic_id	1.89% (1 cells)
dhs-quickstats_national_zaf	missing	characteristic_label	1.89% (1 cells)
dhs-quickstats_national_zaf	missing	characteristic_order	1.89% (1 cells)
dhs-quickstats_national_zaf	missing	dhs_country_code	1.89% (1 cells)
dhs-quickstats_national_zaf	missing	indicator_order	1.89% (1 cells)
dhs-quickstats_national_zaf	missing	indicator_type	1.89% (1 cells)
dhs-quickstats_national_zaf	missing	is_preferred	1.89% (1 cells)
dhs-quickstats_national_zaf	missing	is_total	1.89% (1 cells)
dhs-quickstats_national_zaf	missing	sdrid	1.89% (1 cells)
dhs-quickstats_national_zaf	missing	survey_type	1.89% (1 cells)
dhs-quickstats_national_zaf	missing	survey_year_label	1.89% (1 cells)
hiv-behavior_national_zaf	missing	by_variable_label	99.16% (118 cells)
hiv-behavior_national_zaf	missing	denominator_weighted	33.61% (40 cells)
hiv-behavior_national_zaf	missing	denominator_unweighted	32.77% (39 cells)
hiv-behavior_national_zaf	missing	ci_high	100% (119 cells)
hiv-behavior_national_zaf	missing	ci_low	100% (119 cells)
hiv-behavior_national_zaf	missing	level_rank	100% (119 cells)
hiv-behavior_national_zaf	missing	region_id	100% (119 cells)
hiv-behavior_national_zaf	missing	characteristic_category	0.84% (1 cells)
hiv-behavior_national_zaf	missing	characteristic_id	0.84% (1 cells)
hiv-behavior_national_zaf	missing	characteristic_label	0.84% (1 cells)
hiv-behavior_national_zaf	missing	characteristic_order	0.84% (1 cells)
hiv-behavior_national_zaf	missing	dhs_country_code	0.84% (1 cells)
hiv-behavior_national_zaf	missing	indicator_order	0.84% (1 cells)
hiv-behavior_national_zaf	missing	indicator_type	0.84% (1 cells)
hiv-behavior_national_zaf	missing	is_preferred	0.84% (1 cells)
hiv-behavior_national_zaf	missing	is_total	0.84% (1 cells)
hiv-behavior_national_zaf	missing	sdrid	0.84% (1 cells)
hiv-behavior_national_zaf	missing	survey_type	0.84% (1 cells)
hiv-behavior_national_zaf	missing	survey_year_label	0.84% (1 cells)
immunization_national_zaf	missing	denominator_unweighted	7.69% (9 cells)
immunization_national_zaf	missing	denominator_weighted	7.69% (9 cells)
immunization_national_zaf	missing	by_variable_label	47.86% (56 cells)
immunization_national_zaf	missing	ci_high	100% (117 cells)
immunization_national_zaf	missing	ci_low	100% (117 cells)
immunization_national_zaf	missing	level_rank	100% (117 cells)
immunization_national_zaf	missing	region_id	100% (117 cells)
immunization_national_zaf	missing	characteristic_category	0.85% (1 cells)

dataset	issue_type	column	detail
immunization_national_zaf	missing	characteristic_id	0.85% (1 cells)
immunization_national_zaf	missing	characteristic_label	0.85% (1 cells)
immunization_national_zaf	missing	characteristic_order	0.85% (1 cells)
immunization_national_zaf	missing	dhs_country_code	0.85% (1 cells)
immunization_national_zaf	missing	indicator_order	0.85% (1 cells)
immunization_national_zaf	missing	indicator_type	0.85% (1 cells)
immunization_national_zaf	missing	is_preferred	0.85% (1 cells)
immunization_national_zaf	missing	is_total	0.85% (1 cells)
immunization_national_zaf	missing	sdrid	0.85% (1 cells)
immunization_national_zaf	missing	survey_type	0.85% (1 cells)
immunization_national_zaf	missing	survey_year_label	0.85% (1 cells)
iycf_national_zaf	missing	by_variable_label	95.65% (22 cells)
iycf_national_zaf	missing	characteristic_category	4.35% (1 cells)
iycf_national_zaf	missing	characteristic_id	4.35% (1 cells)
iycf_national_zaf	missing	characteristic_label	4.35% (1 cells)
iycf_national_zaf	missing	characteristic_order	4.35% (1 cells)
iycf_national_zaf	missing	dhs_country_code	4.35% (1 cells)
iycf_national_zaf	missing	indicator_order	4.35% (1 cells)
iycf_national_zaf	missing	indicator_type	4.35% (1 cells)
iycf_national_zaf	missing	is_preferred	4.35% (1 cells)
iycf_national_zaf	missing	is_total	4.35% (1 cells)
iycf_national_zaf	missing	sdrid	4.35% (1 cells)
iycf_national_zaf	missing	survey_type	4.35% (1 cells)
iycf_national_zaf	missing	survey_year_label	4.35% (1 cells)
iycf_national_zaf	missing	denominator_unweighted	13.04% (3 cells)
iycf_national_zaf	missing	denominator_weighted	13.04% (3 cells)
iycf_national_zaf	missing	ci_high	100% (23 cells)
iycf_national_zaf	missing	ci_low	100% (23 cells)
iycf_national_zaf	missing	level_rank	100% (23 cells)
iycf_national_zaf	missing	region_id	100% (23 cells)
literacy_national_zaf	missing	by_variable_label	95.24% (20 cells)
literacy_national_zaf	missing	characteristic_category	4.76% (1 cells)
literacy_national_zaf	missing	characteristic_id	4.76% (1 cells)
literacy_national_zaf	missing	characteristic_label	4.76% (1 cells)
literacy_national_zaf	missing	characteristic_order	4.76% (1 cells)
literacy_national_zaf	missing	dhs_country_code	4.76% (1 cells)
literacy_national_zaf	missing	indicator_order	4.76% (1 cells)
literacy_national_zaf	missing	indicator_type	4.76% (1 cells)
literacy_national_zaf	missing	is_preferred	4.76% (1 cells)
literacy_national_zaf	missing	is_total	4.76% (1 cells)
literacy_national_zaf	missing	sdrid	4.76% (1 cells)

dataset	issue_type	column	detail
literacy_national_zaf	missing	survey_type	4.76% (1 cells)
literacy_national_zaf	missing	survey_year_label	4.76% (1 cells)
literacy_national_zaf	missing	denominator_unweighted	14.29% (3 cells)
literacy_national_zaf	missing	denominator_weighted	14.29% (3 cells)
literacy_national_zaf	missing	ci_high	100% (21 cells)
literacy_national_zaf	missing	ci_low	100% (21 cells)
literacy_national_zaf	missing	level_rank	100% (21 cells)
literacy_national_zaf	missing	region_id	100% (21 cells)
maternal-mortality_national_zaf	missing	by_variable_label	95.45% (21 cells)
maternal-mortality_national_zaf	missing	denominator_weighted	90.91% (20 cells)
maternal-mortality_national_zaf	missing	ci_high	86.36% (19 cells)
maternal-mortality_national_zaf	missing	ci_low	86.36% (19 cells)
maternal-mortality_national_zaf	missing	denominator_unweighted	72.73% (16 cells)
maternal-mortality_national_zaf	missing	characteristic_category	4.55% (1 cells)
maternal-mortality_national_zaf	missing	characteristic_id	4.55% (1 cells)
maternal-mortality_national_zaf	missing	characteristic_label	4.55% (1 cells)
maternal-mortality_national_zaf	missing	characteristic_order	4.55% (1 cells)
maternal-mortality_national_zaf	missing	dhs_country_code	4.55% (1 cells)
maternal-mortality_national_zaf	missing	indicator_order	4.55% (1 cells)
maternal-mortality_national_zaf	missing	indicator_type	4.55% (1 cells)
maternal-mortality_national_zaf	missing	is_preferred	4.55% (1 cells)
maternal-mortality_national_zaf	missing	is_total	4.55% (1 cells)
maternal-mortality_national_zaf	missing	sdrid	4.55% (1 cells)
maternal-mortality_national_zaf	missing	survey_type	4.55% (1 cells)
maternal-mortality_national_zaf	missing	survey_year_label	4.55% (1 cells)
maternal-mortality_national_zaf	missing	level_rank	100% (22 cells)
maternal-mortality_national_zaf	missing	region_id	100% (22 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	denominator_unweighted	33.33% (9 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	denominator_weighted	33.33% (9 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	characteristic_category	3.7% (1 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	characteristic_id	3.7% (1 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	characteristic_label	3.7% (1 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	characteristic_order	3.7% (1 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	dhs_country_code	3.7% (1 cells)

dataset	issue_type	column	detail
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	indicator_order	3.7% (1 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	indicator_type	3.7% (1 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	is_preferred	3.7% (1 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	is_total	3.7% (1 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	sdrid	3.7% (1 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	survey_type	3.7% (1 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	survey_year_label	3.7% (1 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	ci_high	100% (27 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	ci_low	100% (27 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	level_rank	100% (27 cells)
symptoms-of-acute-respiratory-infection-ari_national_zaf	missing	region_id	100% (27 cells)
toilet-facilities_national_zaf	missing	by_variable_label	97.87% (46 cells)
toilet-facilities_national_zaf	missing	characteristic_category	2.13% (1 cells)
toilet-facilities_national_zaf	missing	characteristic_id	2.13% (1 cells)
toilet-facilities_national_zaf	missing	characteristic_label	2.13% (1 cells)
toilet-facilities_national_zaf	missing	characteristic_order	2.13% (1 cells)
toilet-facilities_national_zaf	missing	dhs_country_code	2.13% (1 cells)
toilet-facilities_national_zaf	missing	indicator_order	2.13% (1 cells)
toilet-facilities_national_zaf	missing	indicator_type	2.13% (1 cells)
toilet-facilities_national_zaf	missing	is_preferred	2.13% (1 cells)
toilet-facilities_national_zaf	missing	is_total	2.13% (1 cells)
toilet-facilities_national_zaf	missing	sdrid	2.13% (1 cells)
toilet-facilities_national_zaf	missing	survey_type	2.13% (1 cells)
toilet-facilities_national_zaf	missing	survey_year_label	2.13% (1 cells)
toilet-facilities_national_zaf	missing	ci_high	100% (47 cells)
toilet-facilities_national_zaf	missing	ci_low	100% (47 cells)
toilet-facilities_national_zaf	missing	level_rank	100% (47 cells)
toilet-facilities_national_zaf	missing	region_id	100% (47 cells)
toilet-facilities_national_zaf	missing	denominator_unweighted	10.64% (5 cells)
toilet-facilities_national_zaf	missing	denominator_weighted	10.64% (5 cells)
water_national_zaf	missing	by_variable_label	99.01% (100 cells)

dataset	issue_type	column	detail
water_national_zaf	missing	denominator_unweighted	4.95% (5 cells)
water_national_zaf	missing	denominator_weighted	4.95% (5 cells)
water_national_zaf	missing	ci_high	100% (101 cells)
water_national_zaf	missing	ci_low	100% (101 cells)
water_national_zaf	missing	level_rank	100% (101 cells)
water_national_zaf	missing	region_id	100% (101 cells)
water_national_zaf	missing	characteristic_category	0.99% (1 cells)
water_national_zaf	missing	characteristic_id	0.99% (1 cells)
water_national_zaf	missing	characteristic_label	0.99% (1 cells)
water_national_zaf	missing	characteristic_order	0.99% (1 cells)
water_national_zaf	missing	dhs_country_code	0.99% (1 cells)
water_national_zaf	missing	indicator_order	0.99% (1 cells)
water_national_zaf	missing	indicator_type	0.99% (1 cells)
water_national_zaf	missing	is_preferred	0.99% (1 cells)
water_national_zaf	missing	is_total	0.99% (1 cells)
water_national_zaf	missing	sdrid	0.99% (1 cells)
water_national_zaf	missing	survey_type	0.99% (1 cells)
water_national_zaf	missing	survey_year_label	0.99% (1 cells)

```

if (write_exports) {
  readr::write_csv(issues_log, "M1_data_quality_issues_log.csv")
}

```