# BIN381 — Milestone 1: Data Understanding

2025-09-08

## Contents

## 1  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)
  )
}
```

| 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 |

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

# 2  2. Dataset-Level Summary

```r
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 | rows | cols | duplicate_rows | missing_cells |
|---|---|---|---|---|
| access-to-health-care_national_zaf | 276 | 29 | 0 | 1199 |
| anthropometry_national_zaf | 38 | 29 | 0 | 211 |
| child-mortality-rates_national_zaf | 41 | 29 | 0 | 210 |
| covid-19-prevention_national_zaf | 35 | 29 | 0 | 192 |
| dhs-quickstats_national_zaf | 53 | 29 | 0 | 267 |
| hiv-behavior_national_zaf | 119 | 29 | 0 | 685 |
| immunization_national_zaf | 117 | 29 | 0 | 554 |
| iycf_national_zaf | 23 | 29 | 0 | 132 |
| literacy_national_zaf | 21 | 29 | 0 | 122 |
| maternal-mortality_national_zaf | 22 | 29 | 0 | 151 |
| symptoms-of-acute-respiratory-infection-ari_national_zaf | 27 | 29 | 0 | 138 |
| toilet-facilities_national_zaf | 47 | 29 | 0 | 256 |
| water_national_zaf | 101 | 29 | 0 | 526 |

```
    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)

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

# 3  3. Data Quality Assessment

## 3.1  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)
```
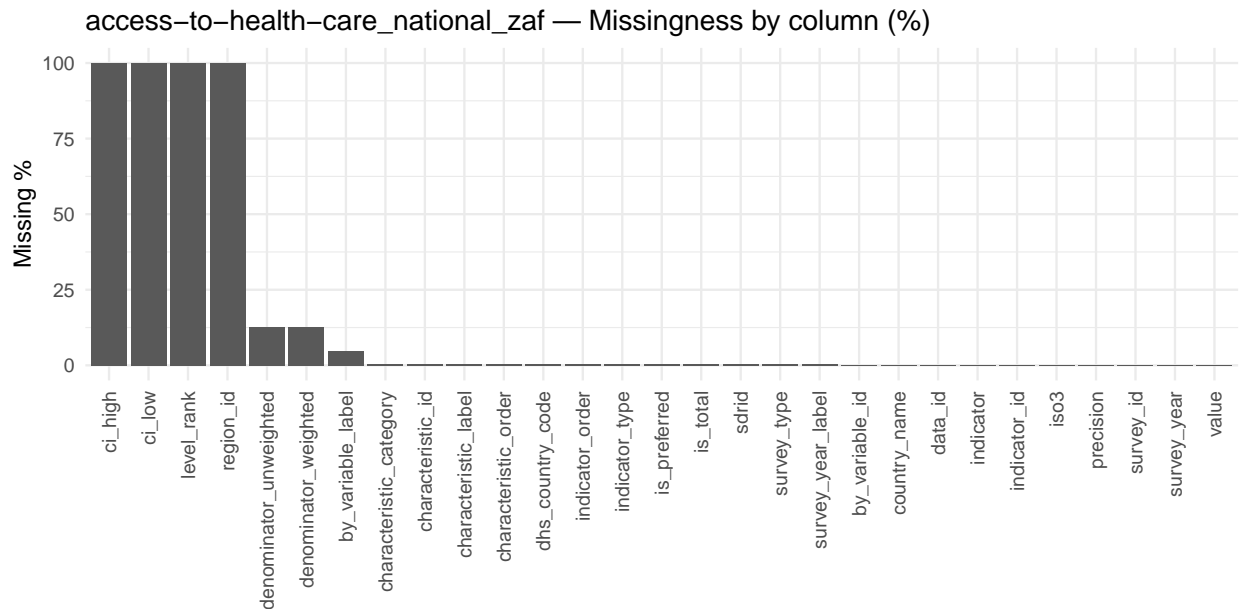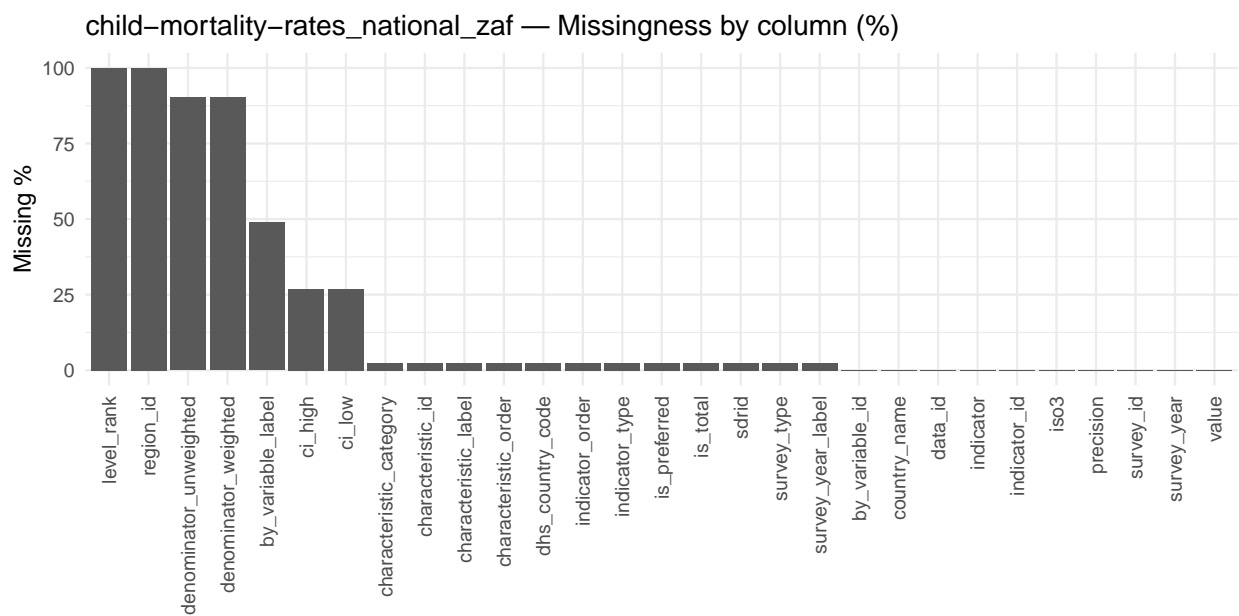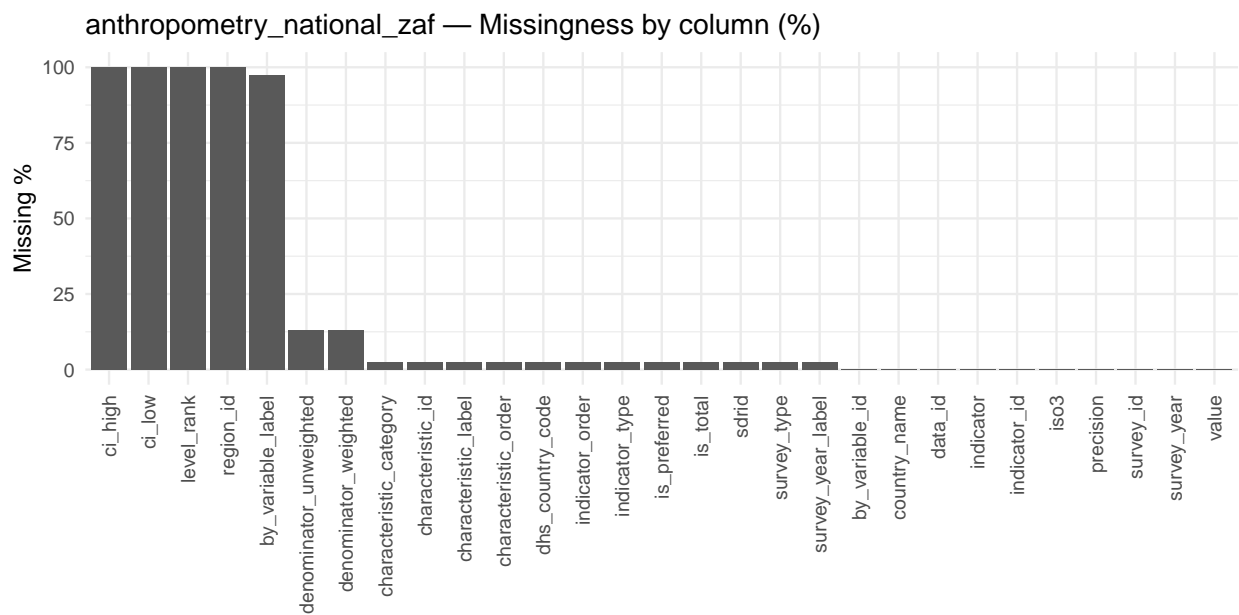
```
gt::gt(missingness_all)
```

```
if (write_exports) {
  readr::write_csv(missingness_all, "M1_missingness_all_columns.csv")
}
```
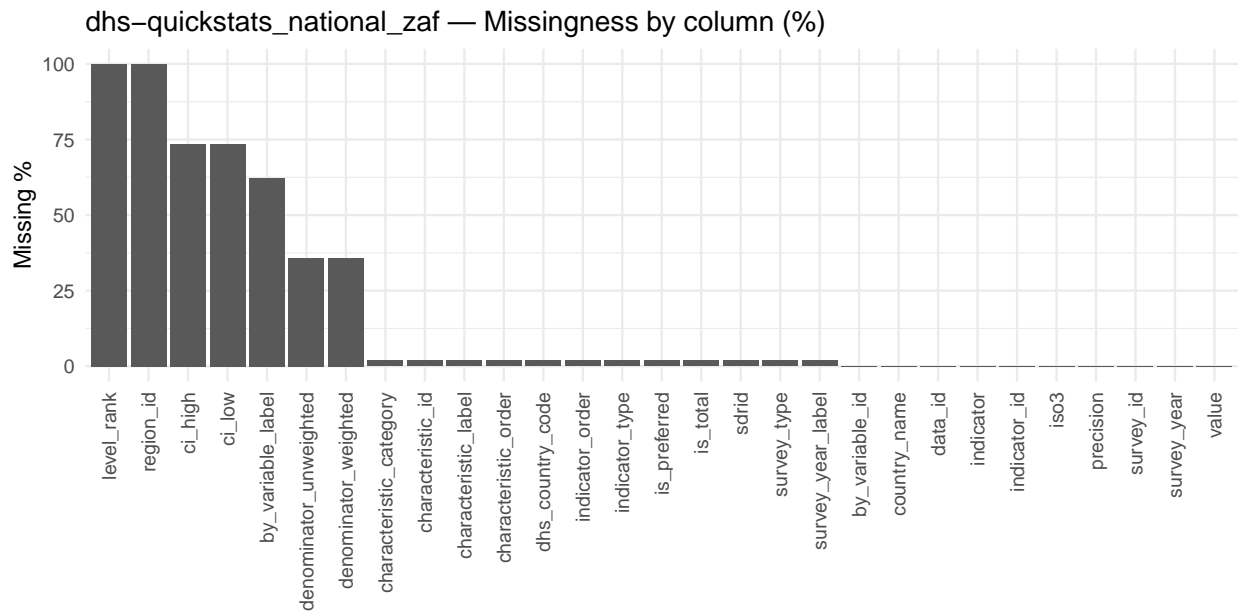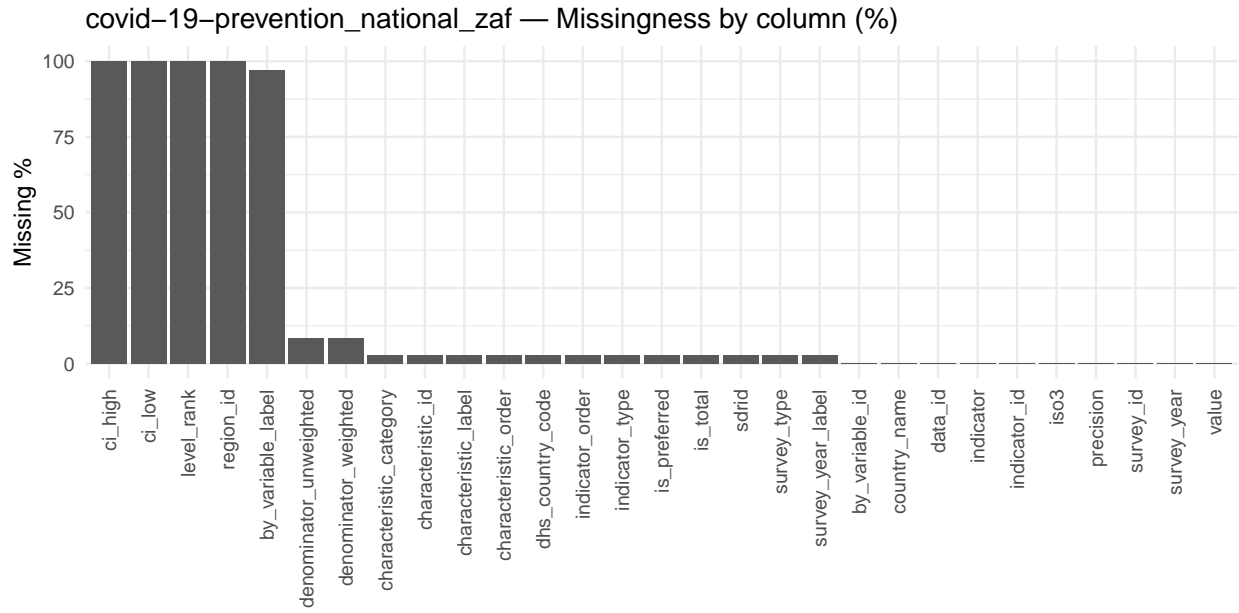
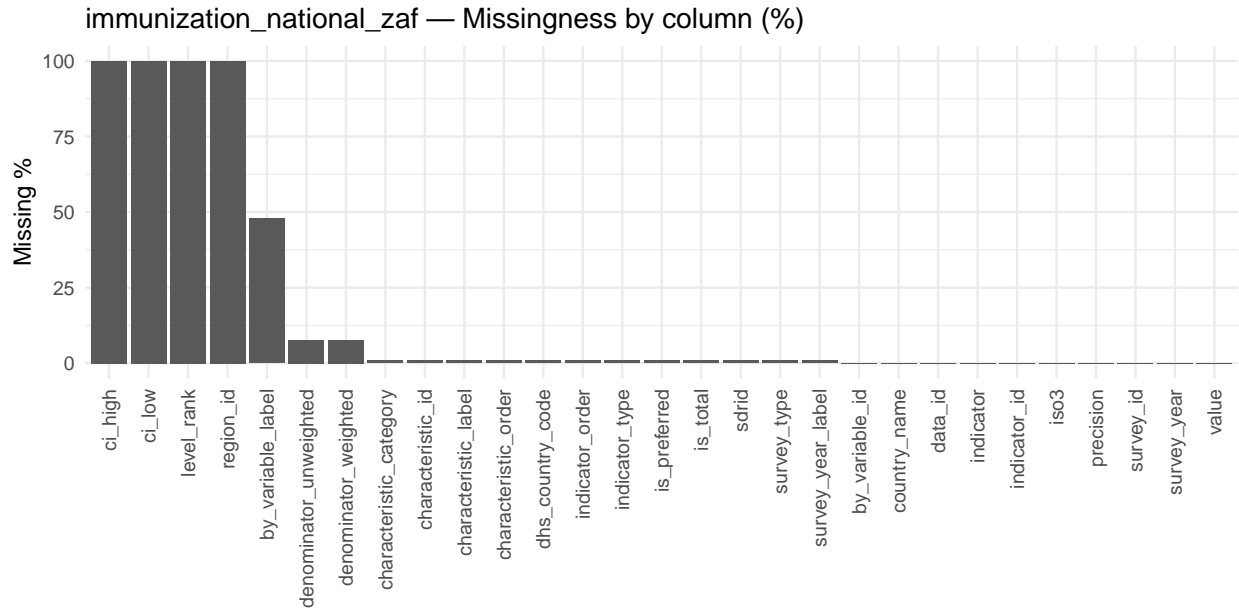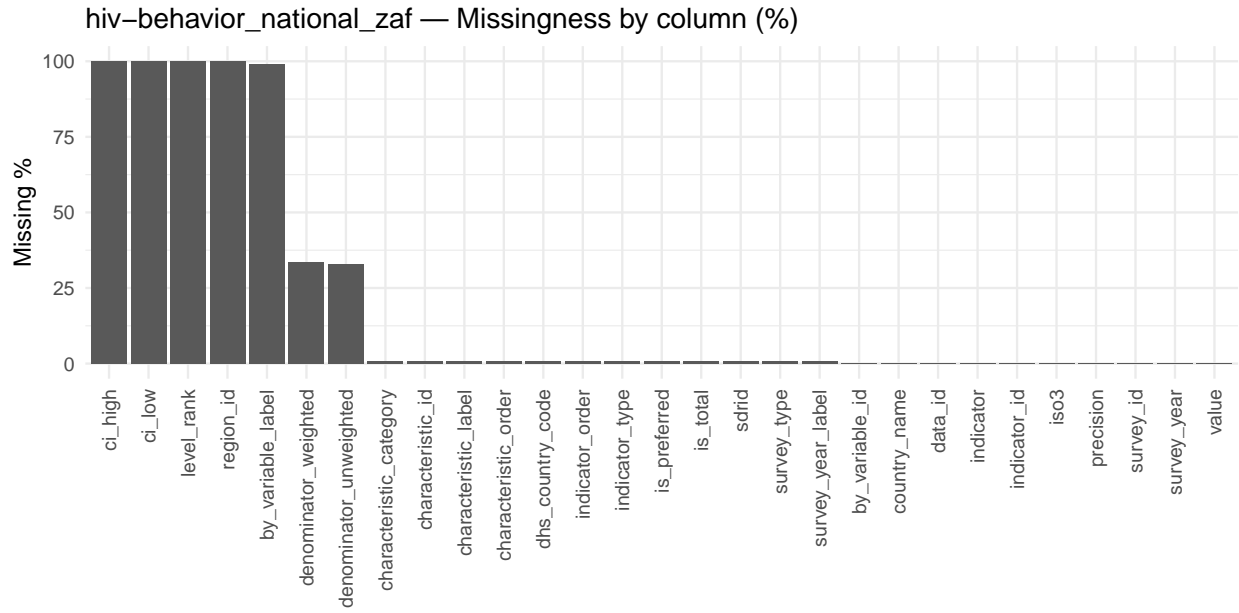### 3.1.1 3.1.1 Missingness Charts (Per Dataset)

```
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)))
```
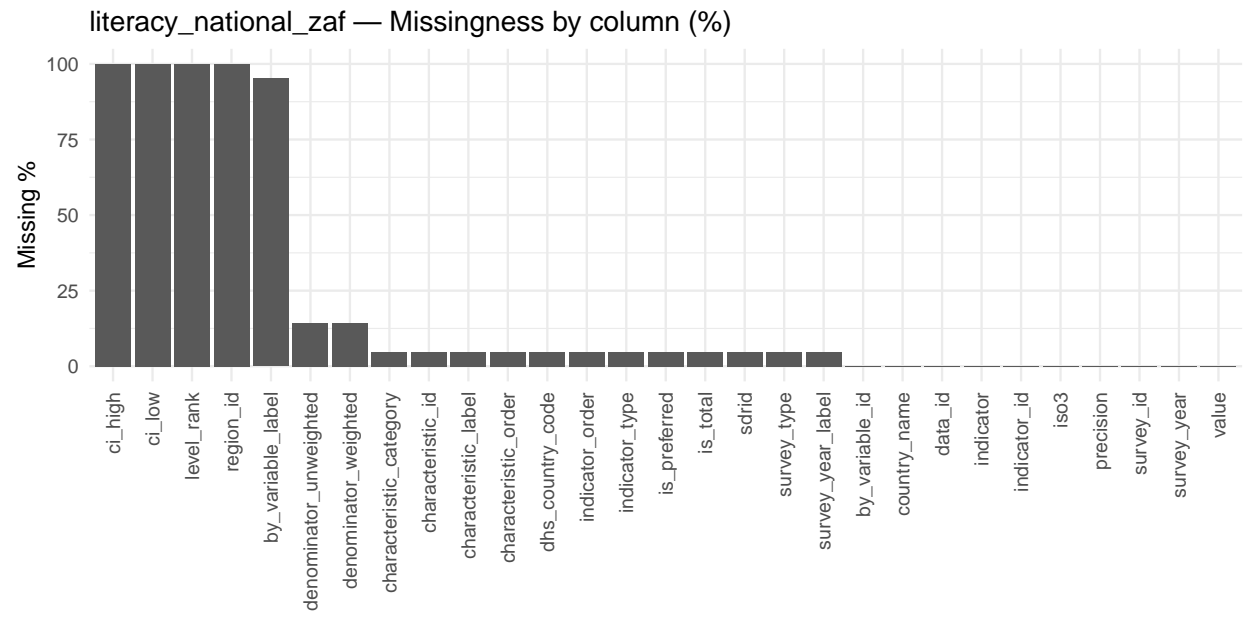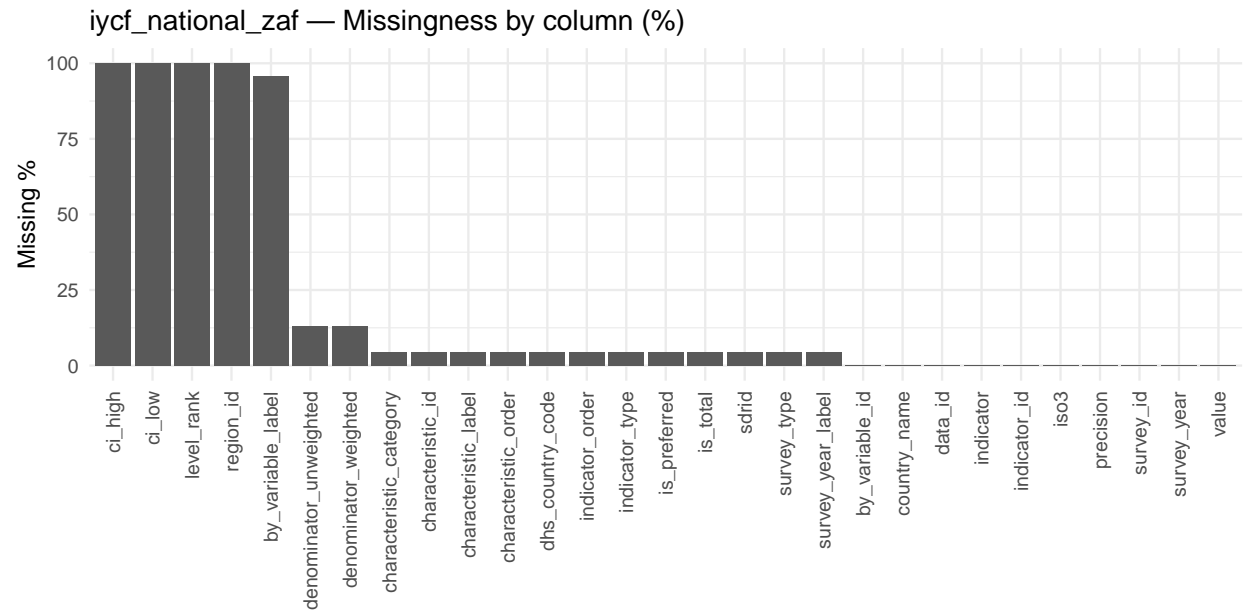


access–to–health–care_national_zaf — Missingness by column (%)

## anthropometry_national_zaf — Missingness by column (%)



## child−mortality−rates_national_zaf — Missingness by column (%)

## covid−19−prevention_national_zaf — Missingness by column (%)



## dhs−quickstats_national_zaf — Missingness by column (%)

hiv−behavior_national_zaf — Missingness by column (%)



immunization_national_zaf — Missingness by column (%)

## iycf_national_zaf — Missingness by column (%)



## literacy_national_zaf — Missingness by column (%)

maternal–mortality_national_zaf — Missingness by column (%)



symptoms–of–acute–respiratory–infection–ari_national_zaf — Missingness by column (%)

## toilet–facilities_national_zaf — Missingness by column (%)



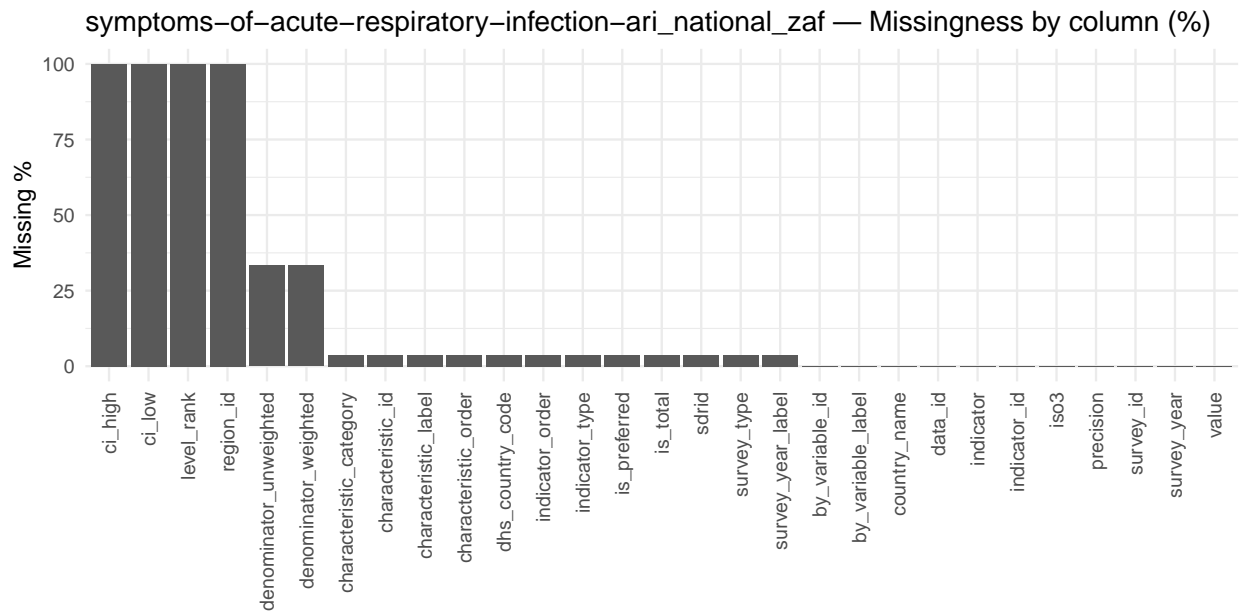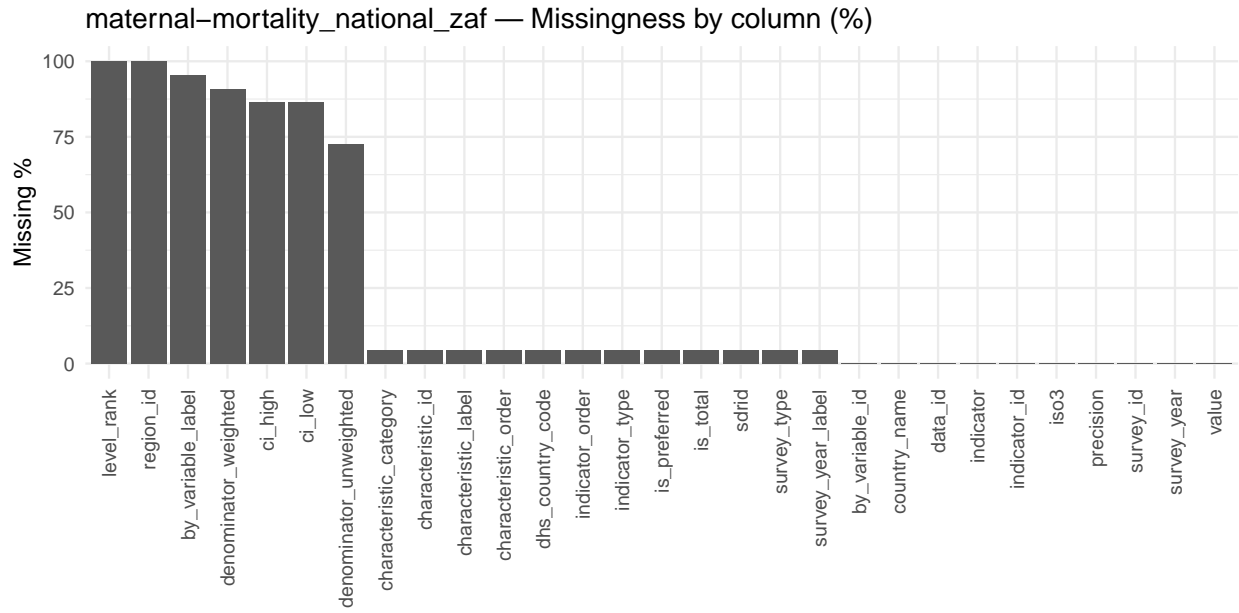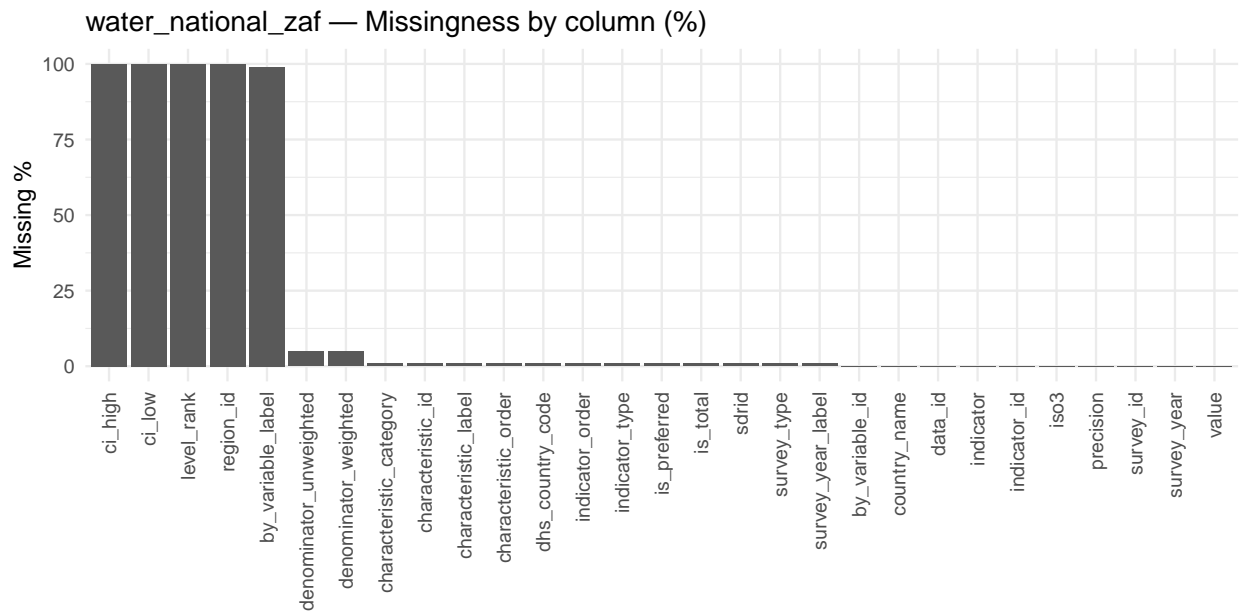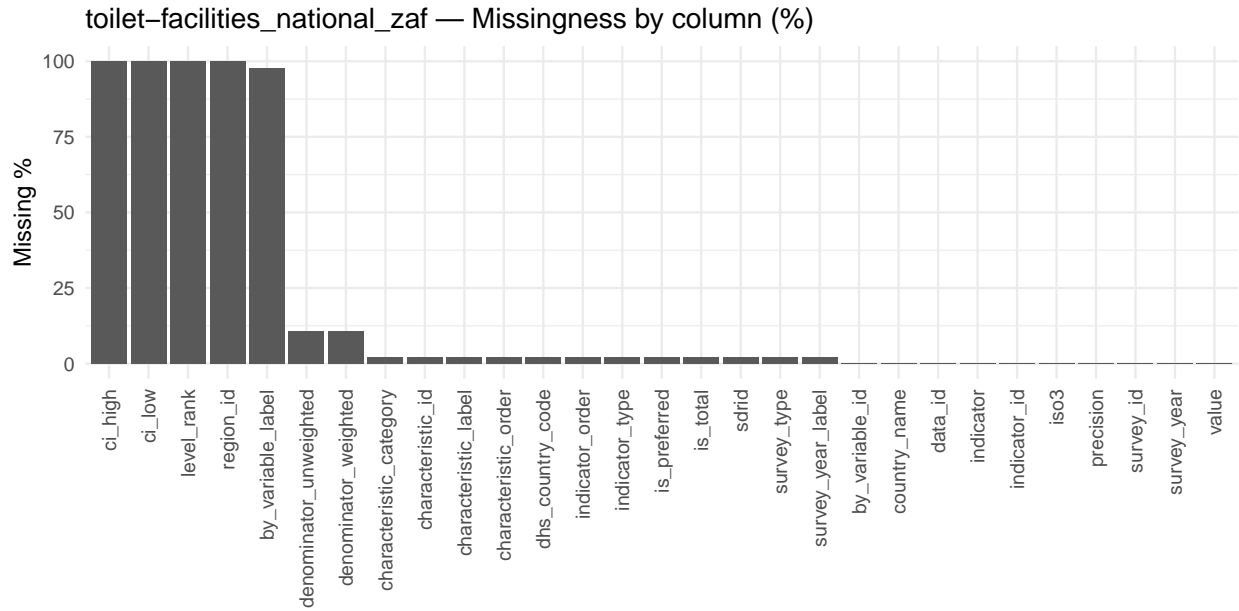## water_national_zaf — Missingness by column (%)



## 3.2   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)

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

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

```r
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.")
}
```

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

## 3.4   3.4 Consolidated Data Quality Issues Log

```r
# 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,
```

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

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

gt::gt(issues_log)
```

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

# 4   4. Preliminary Visualizations
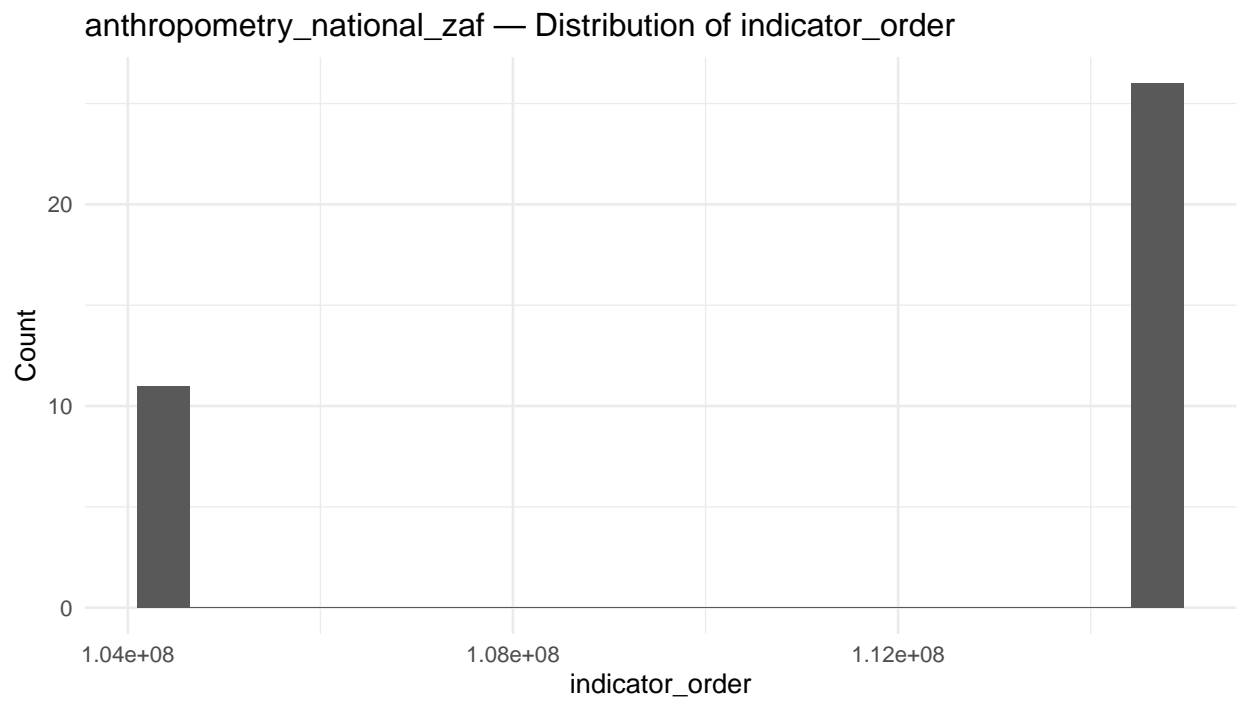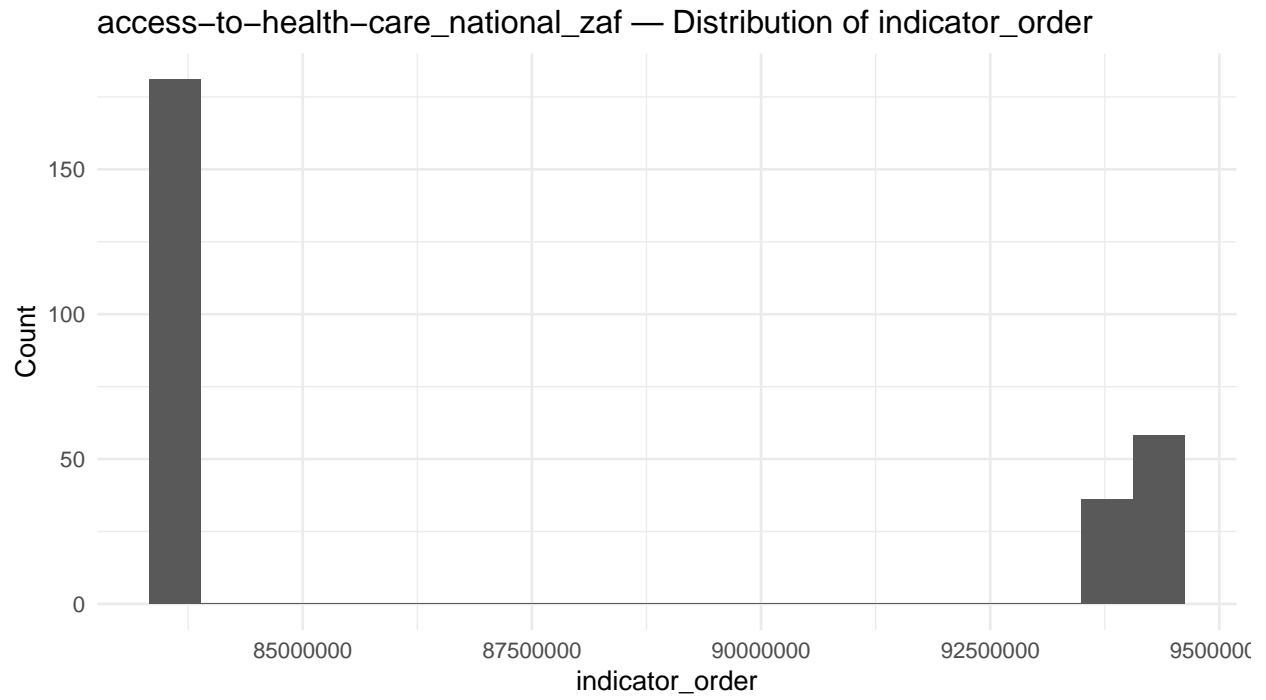
## 4.1   4.1 Numeric Distributions (Highest-Variance Column per Dataset)

```
plot_numeric_hist <- function(df, nm){
  nums <- df |> dplyr::select(where(is.numeric))
  if(ncol(nums) == 0) return(NULL)
  var_tbl <- summarize(nums, across(everything(), function(y) var(y, na.rm = TRUE)))
  var_col <- var_tbl |> pivot_longer(everything(), names_to="col", values_to="v") |>
    arrange(desc(v)) |> slice(1) |> pull(col)
  ggplot(df, aes(x = .data[[var_col]])) +
    geom_histogram(bins = 20) +
    labs(title = paste0(nm, " - Distribution of ", var_col), x = var_col, y = "Count")
}
purrr::iwalk(dfs, function(df, nm){ p <- plot_numeric_hist(df, nm); if(!is.null(p)) print(p) })
```
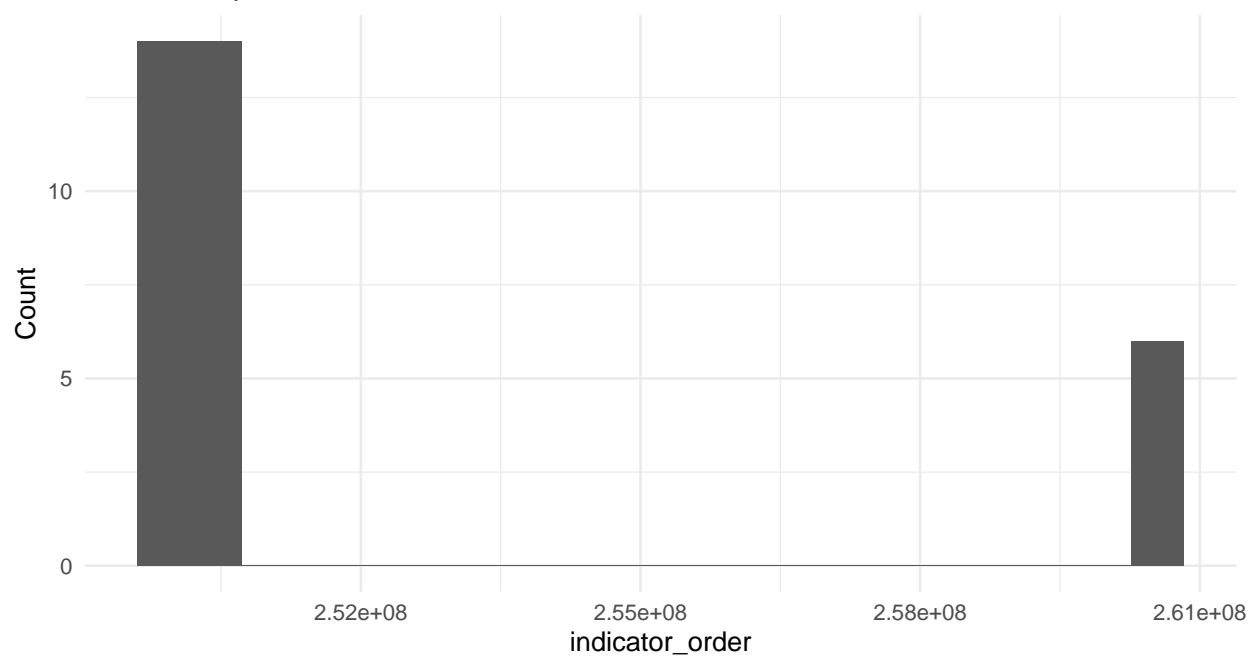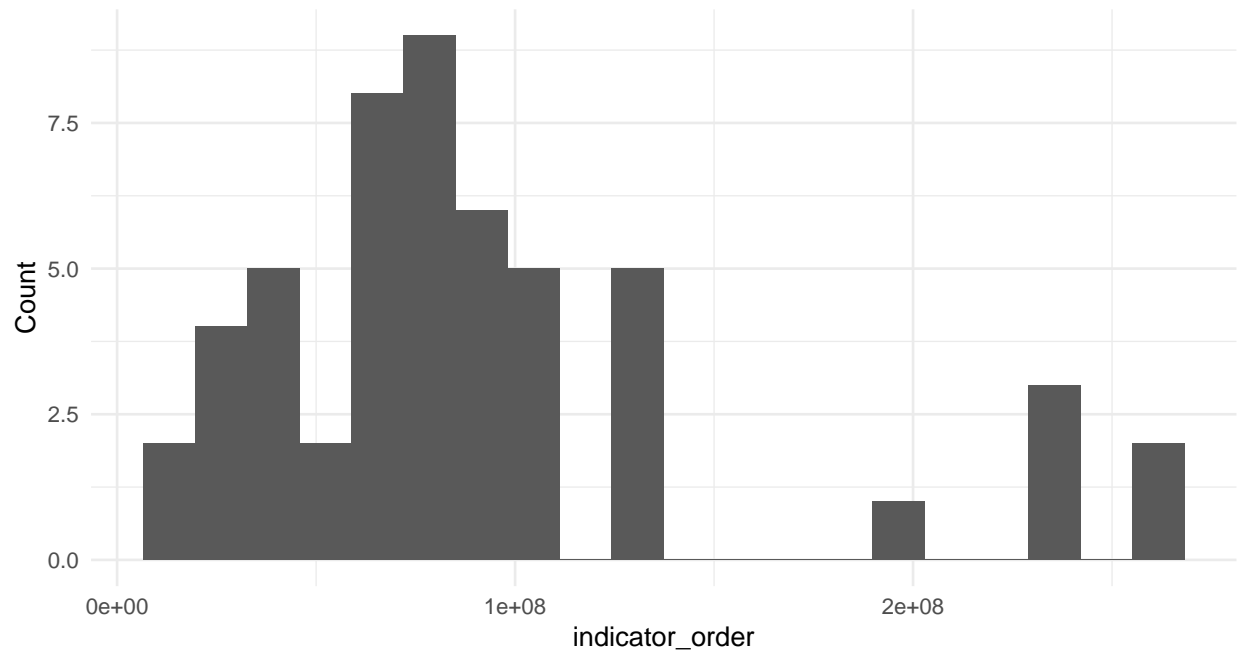
access–to–health–care_national_zaf — Distribution of indicator_order



anthropometry_national_zaf — Distribution of indicator_order

13

## child−mortality−rates_national_zaf — Distribution of characteristic_order
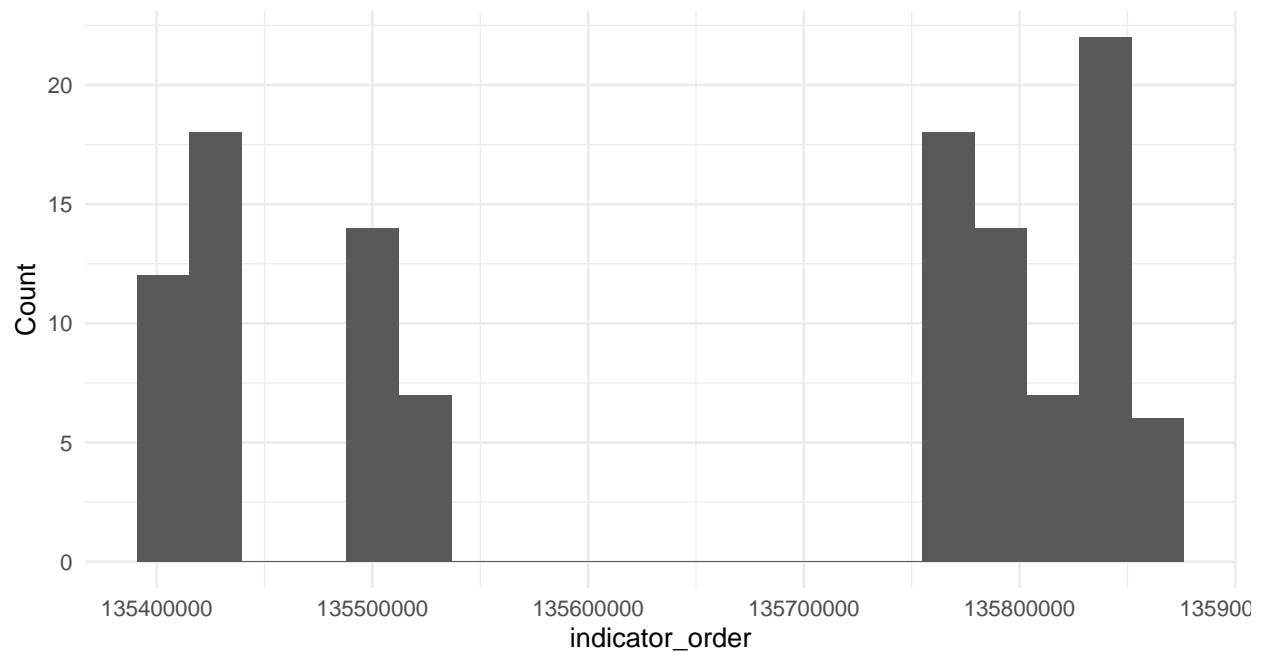


## covid−19−prevention_national_zaf — Distribution of indicator_order
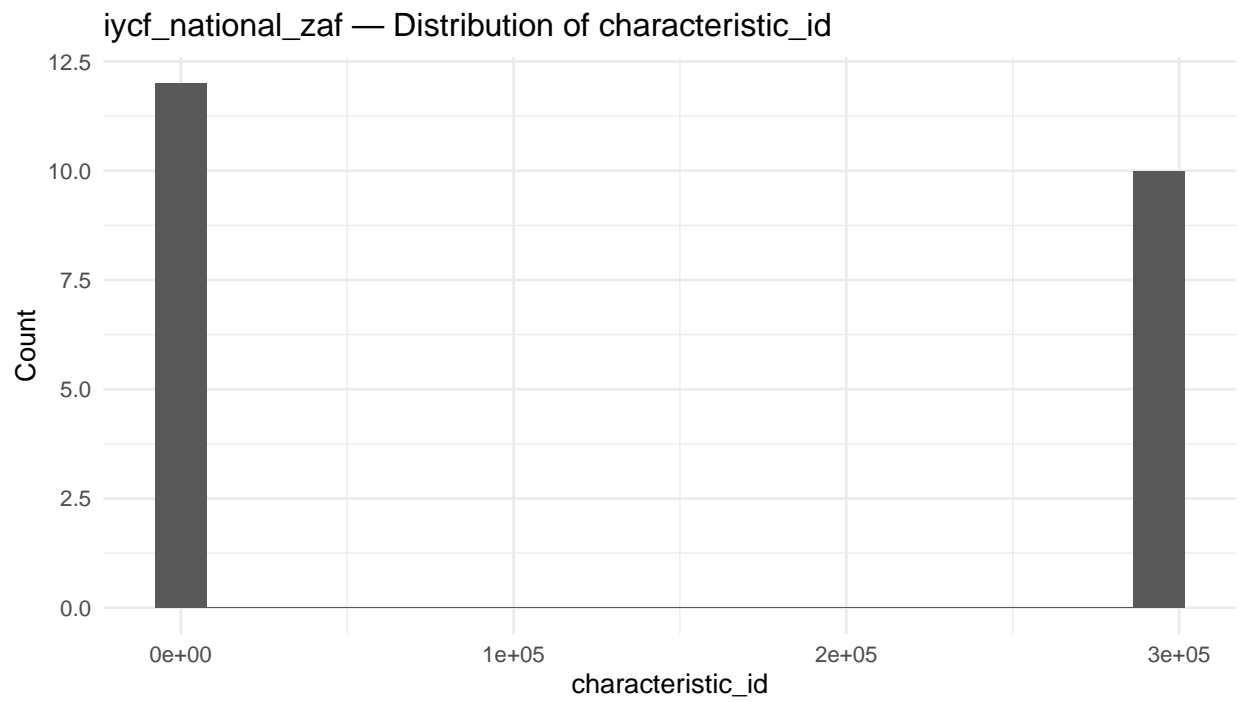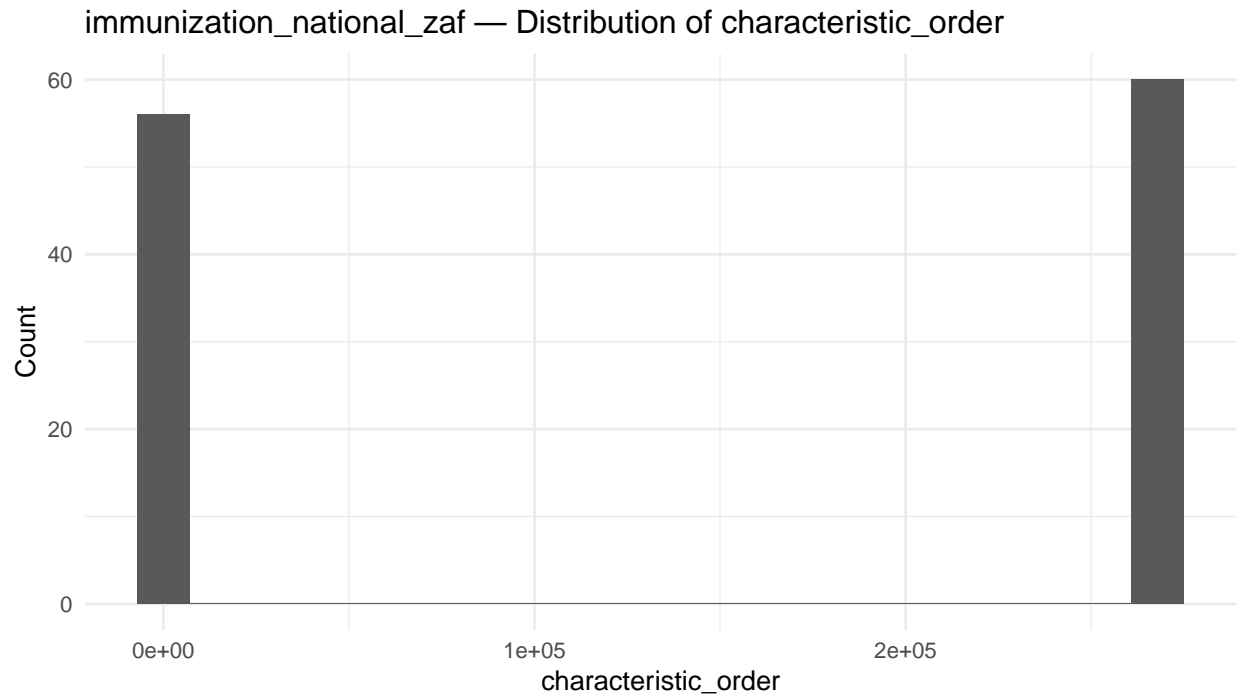
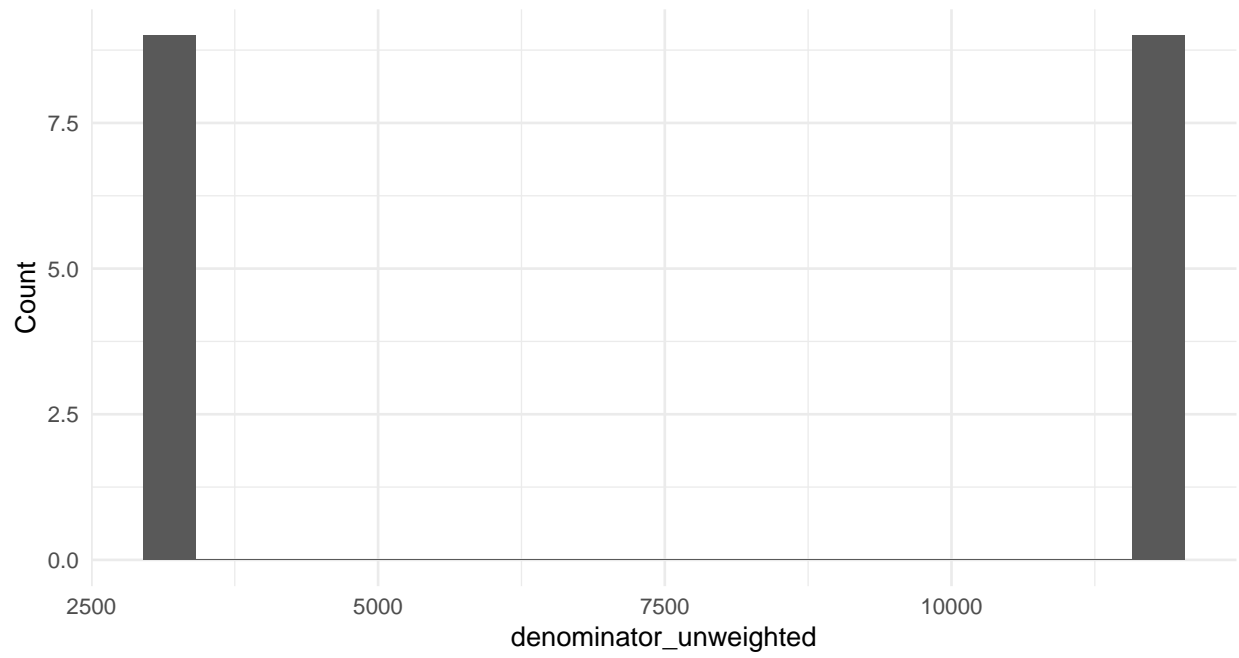dhs−quickstats_national_zaf — Distribution of indicator_order



hiv−behavior_national_zaf — Distribution of indicator_order

## immunization_national_zaf — Distribution of characteristic_order



## iycf_national_zaf — Distribution of characteristic_id
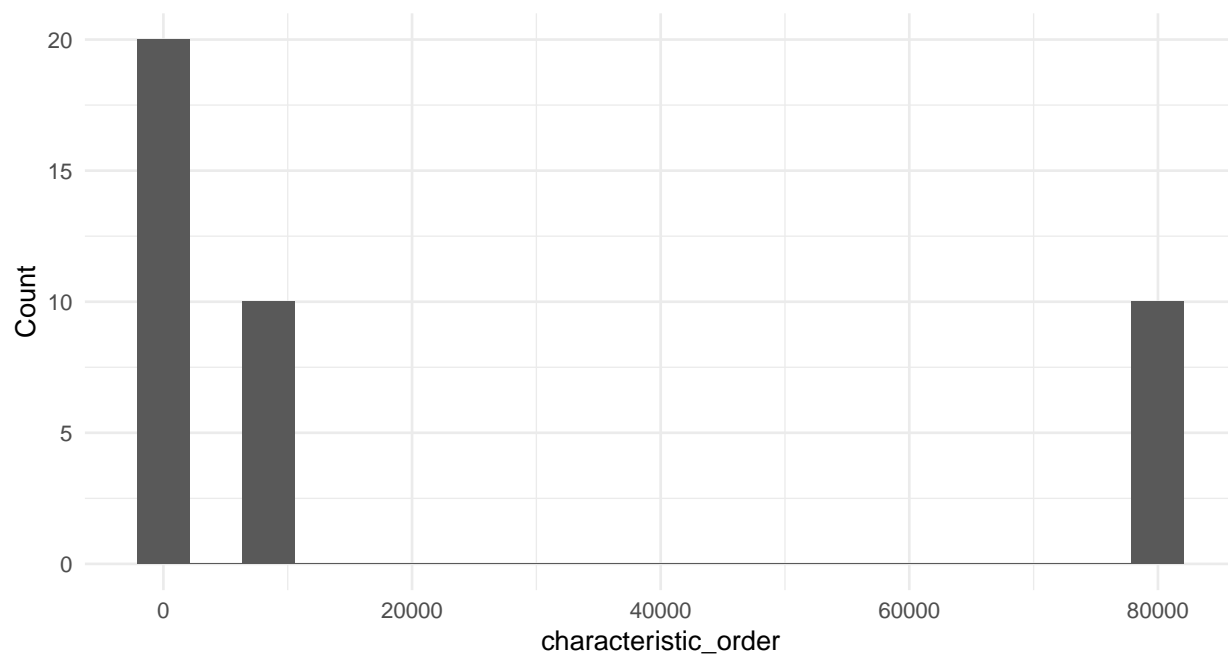
## literacy_national_zaf — Distribution of denominator_unweighted



## maternal−mortality_national_zaf — Distribution of denominator_weighted

symptoms-of-acute-respiratory-infection-ari_national_zaf — Distribution of deno...



toilet-facilities_national_zaf — Distribution of denominator_unweighted

water_national_zaf — Distribution of indicator_order
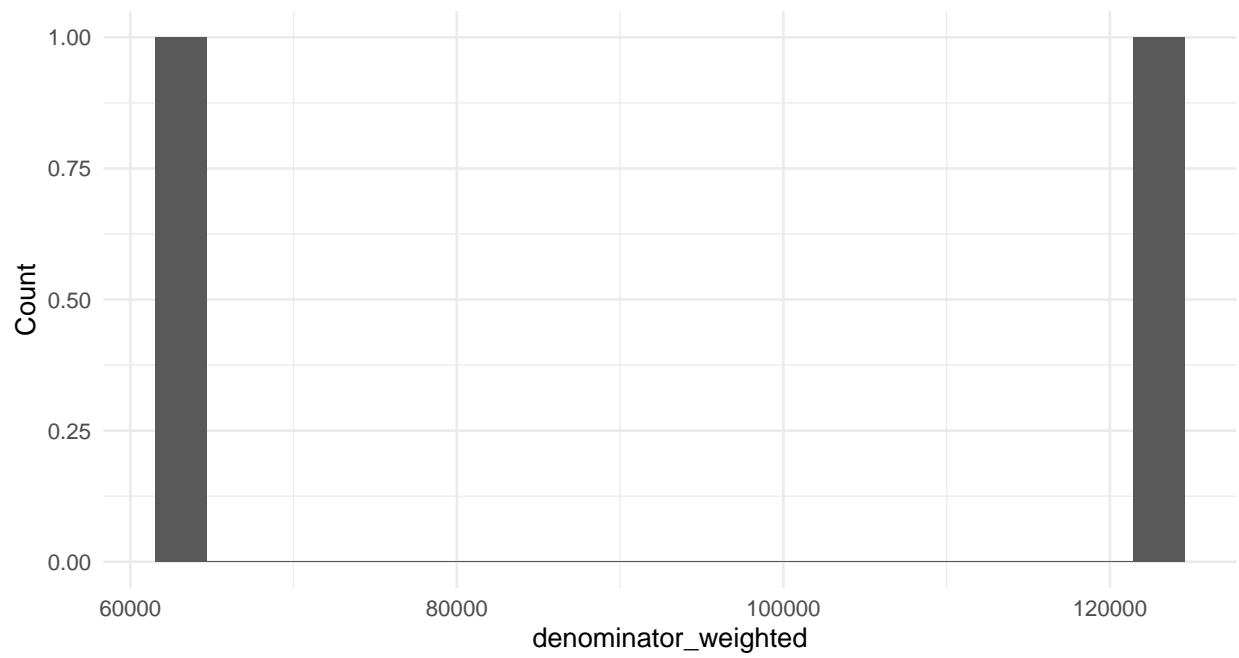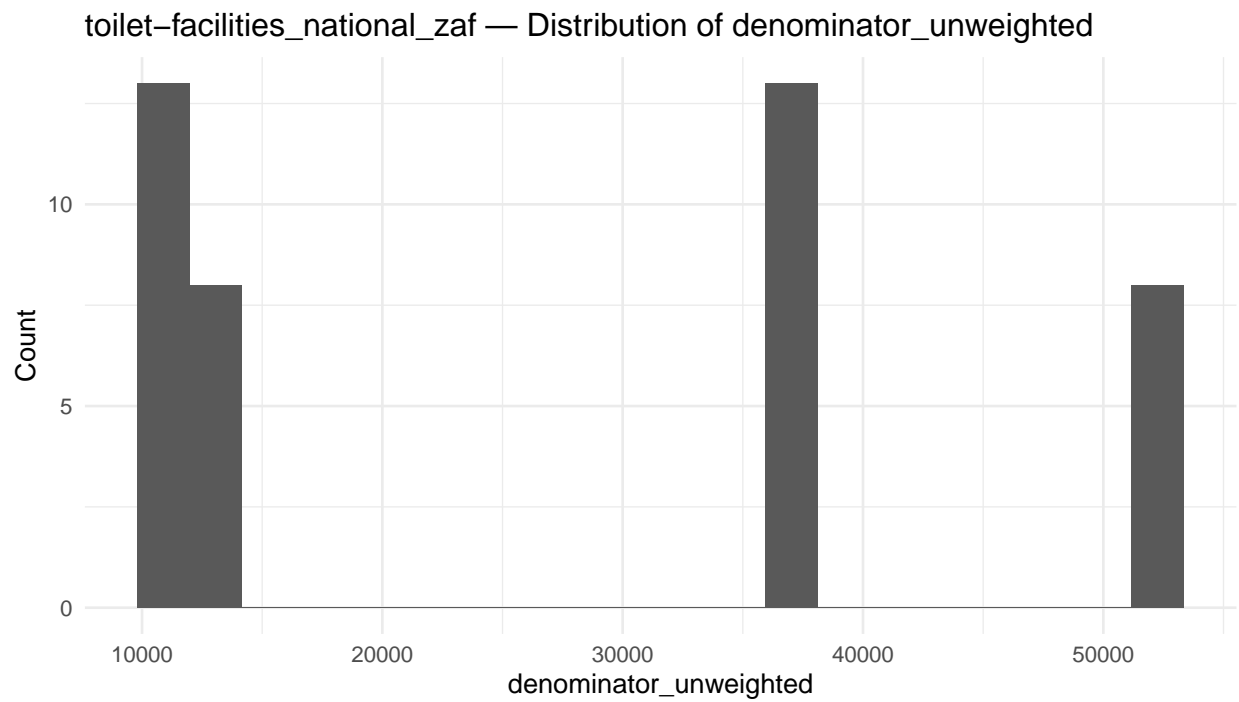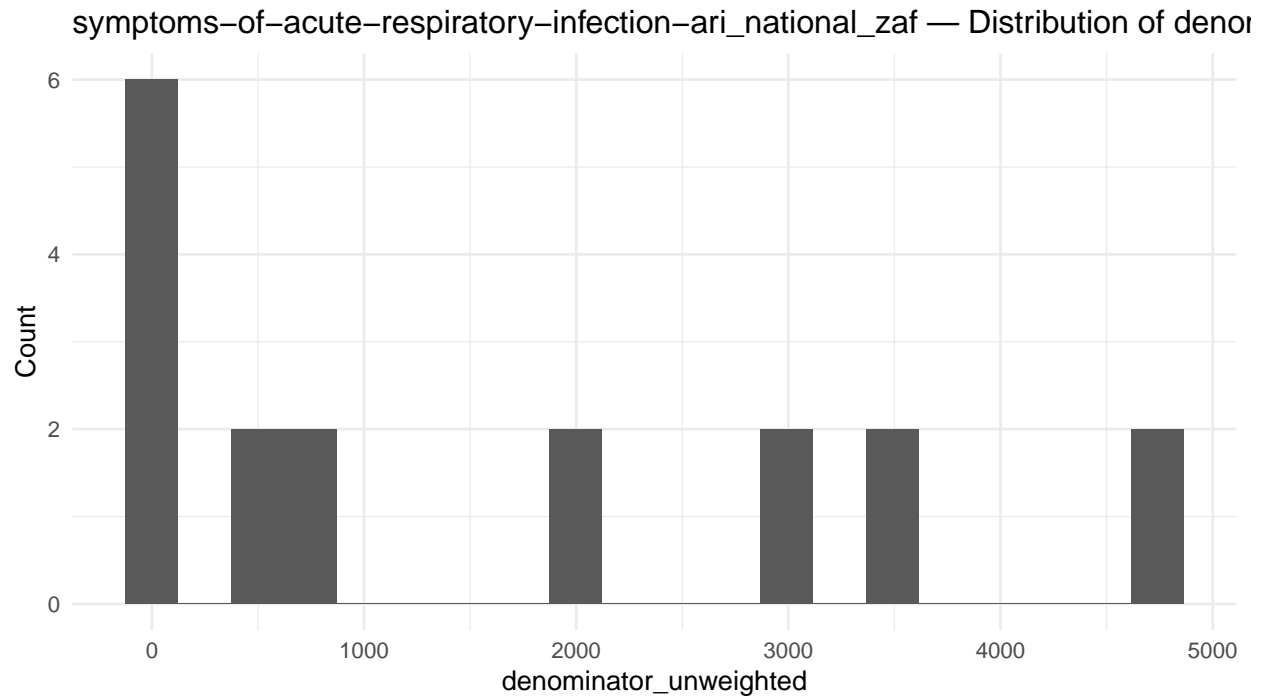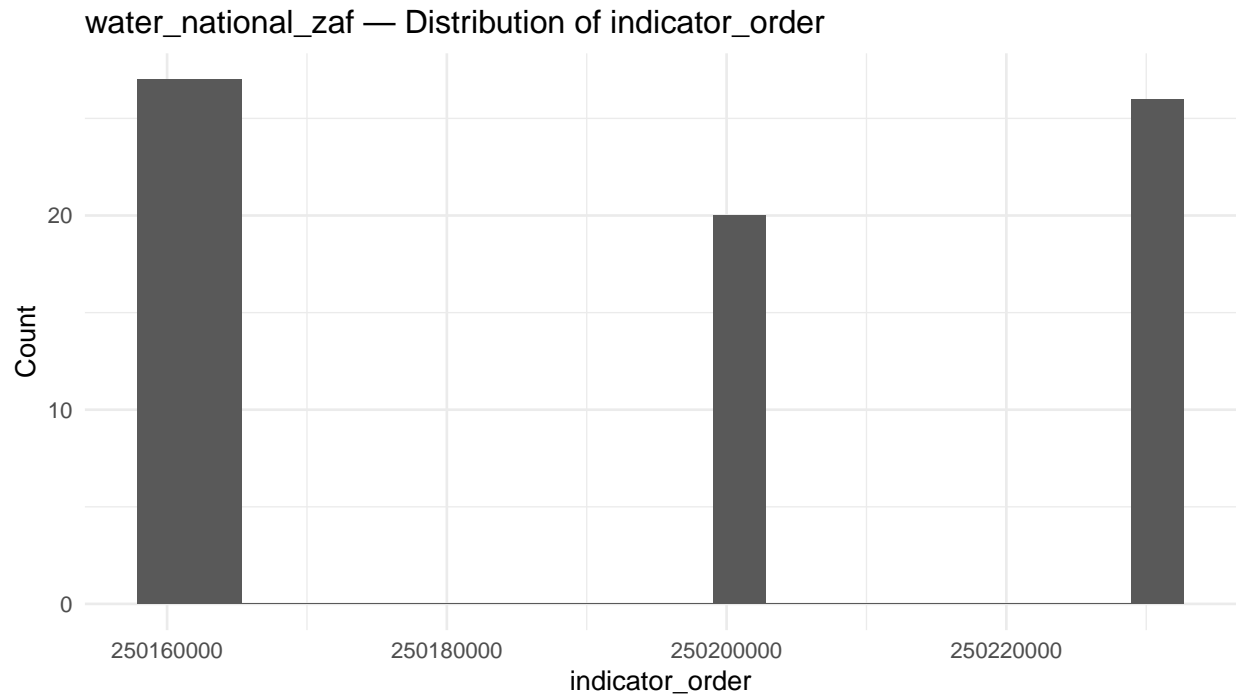


## 4.2   4.2 Top Categories (First Categorical Column)

```r
plot_top_category <- function(df, nm){
  cats <- df |> dplyr::select(where(negate(is.numeric)))
  if(ncol(cats) == 0) return(NULL)
  col1 <- names(cats)[1]
  top5 <- df |> mutate(across(all_of(col1), as.character)) |>
    count(.data[[col1]], sort = TRUE) |> slice_head(n = 5)
  ggplot(top5, aes(x = reorder(.data[[col1]], -n), y = n)) +
    geom_col() +
    labs(title = paste0(nm, " - Top 5 of ", col1), x = col1, y = "Count")
}
purrr::iwalk(dfs, function(df, nm){ p <- plot_top_category(df, nm); if(!is.null(p)) print(p) })
```

## access–to–health–care_national_zaf — Top 5 of iso3



## anthropometry_national_zaf — Top 5 of iso3

## child–mortality–rates_national_zaf — Top 5 of iso3



## covid–19–prevention_national_zaf — Top 5 of iso3
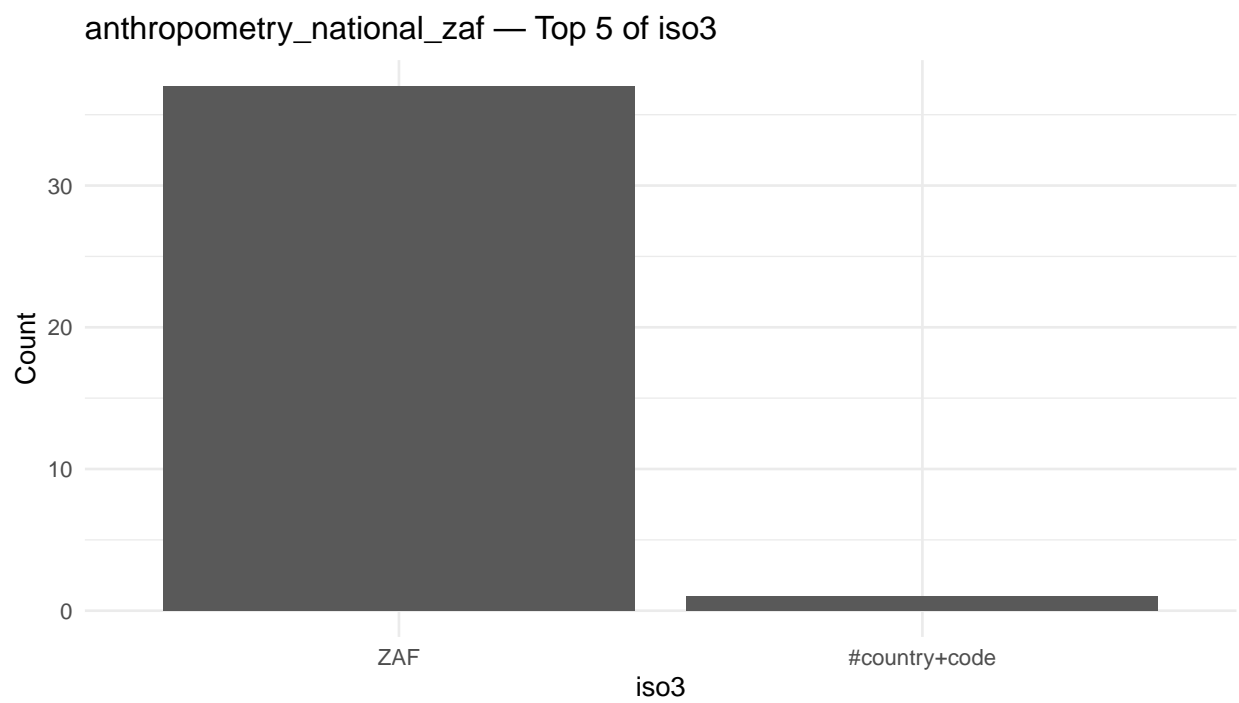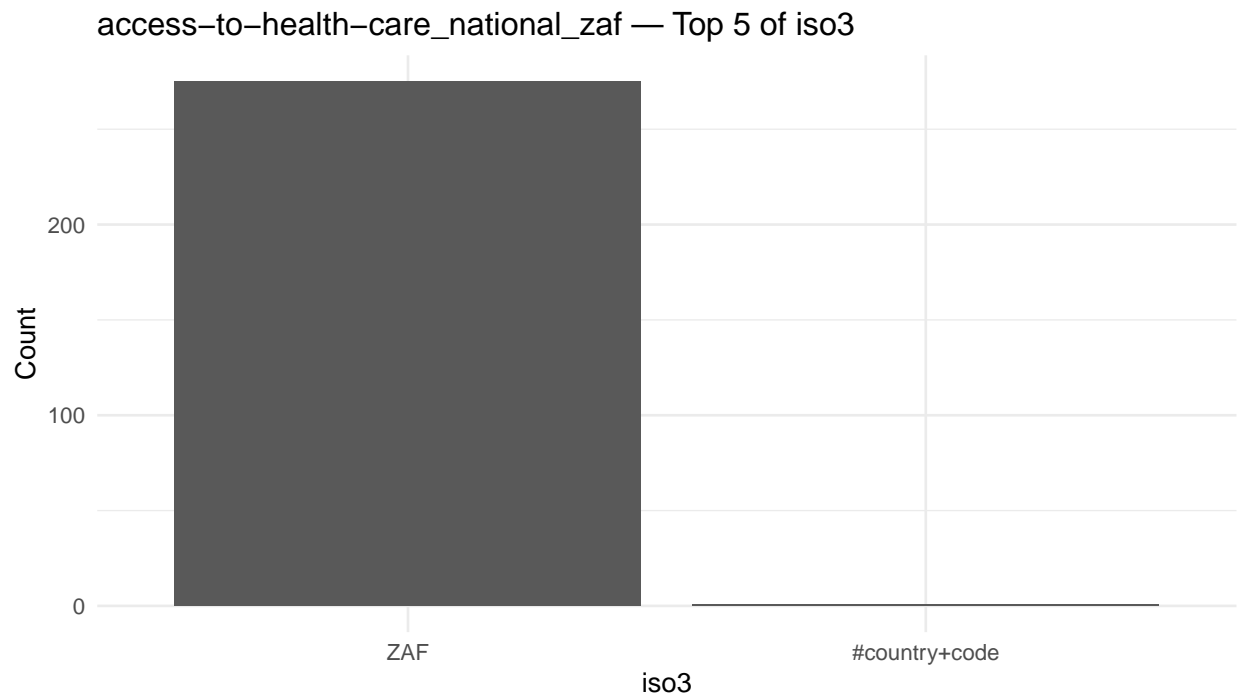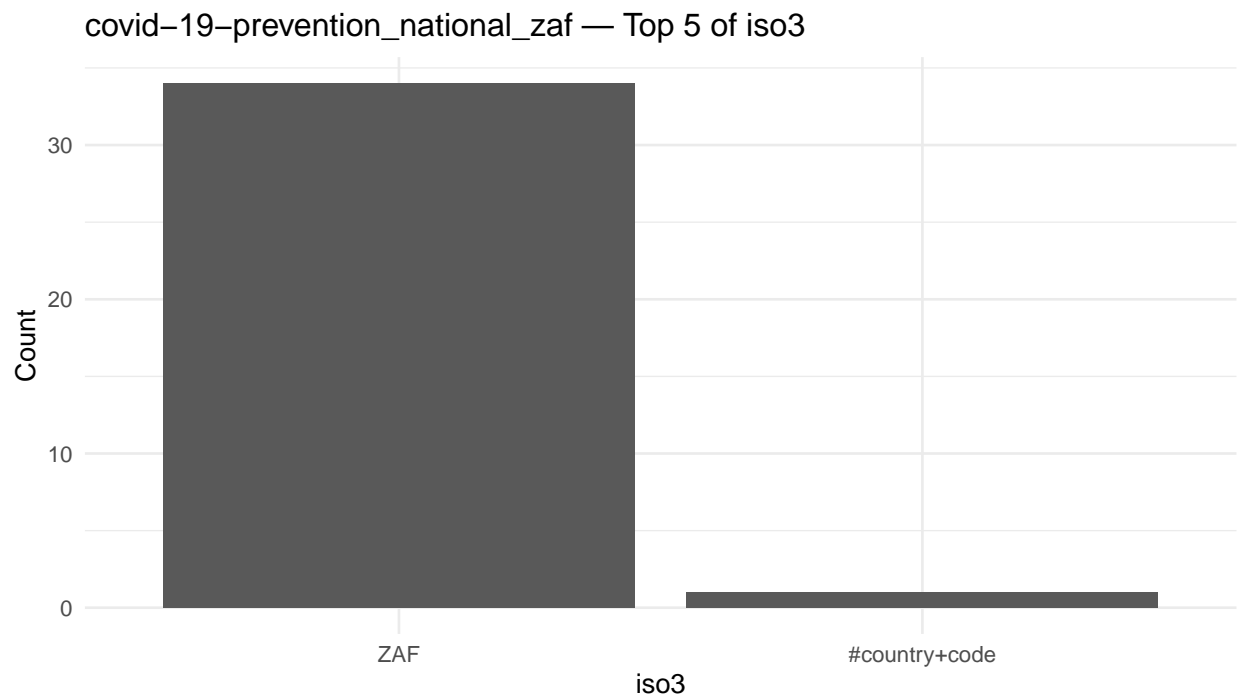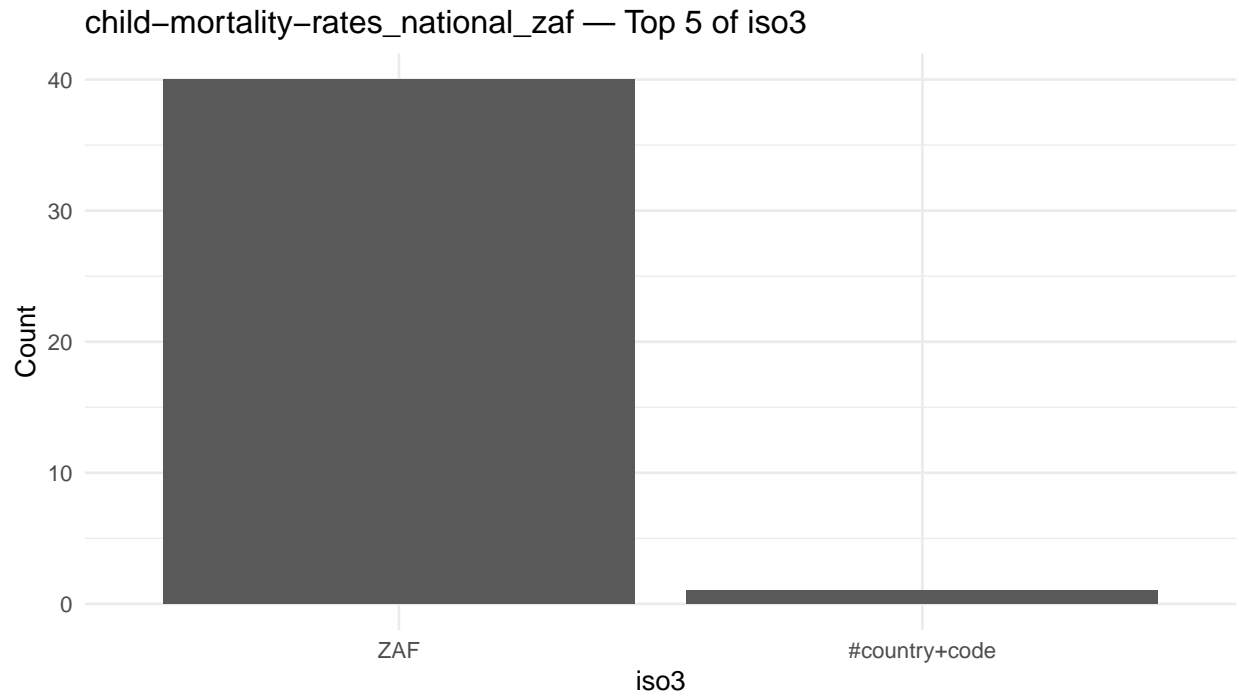
dhs−quickstats_national_zaf — Top 5 of iso3

hiv−behavior_national_zaf — Top 5 of iso3

immunization_national_zaf — Top 5 of iso3



iycf_national_zaf — Top 5 of iso3

## literacy_national_zaf — Top 5 of iso3
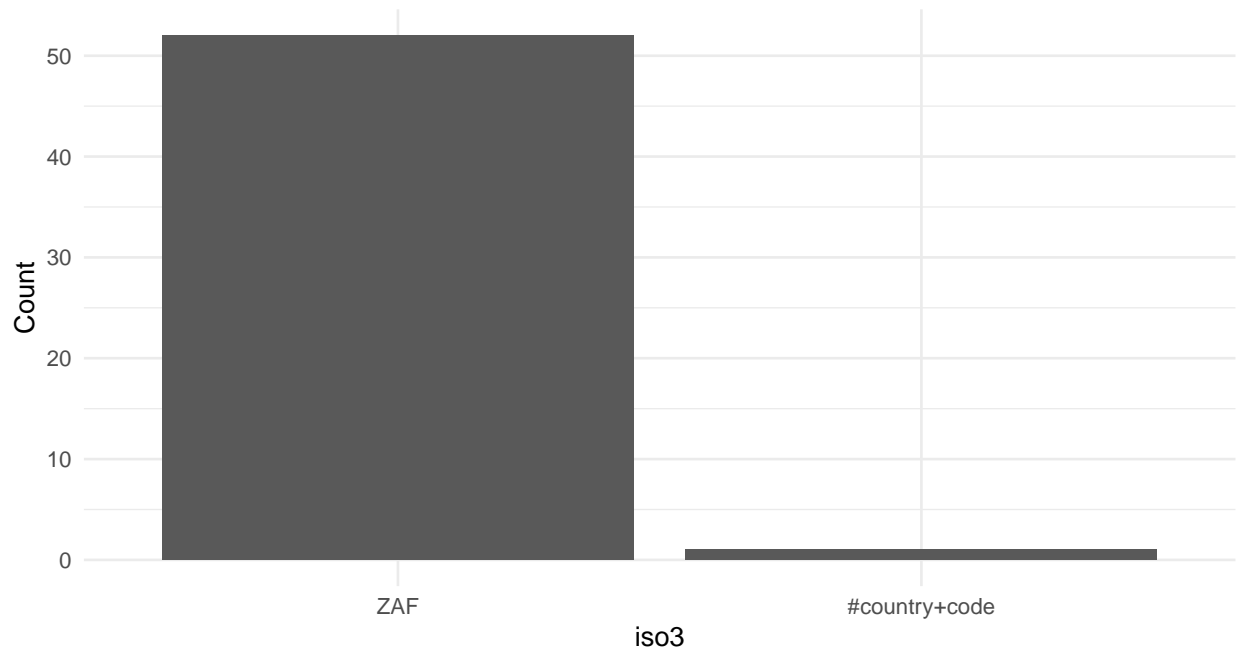


## maternal−mortality_national_zaf — Top 5 of iso3
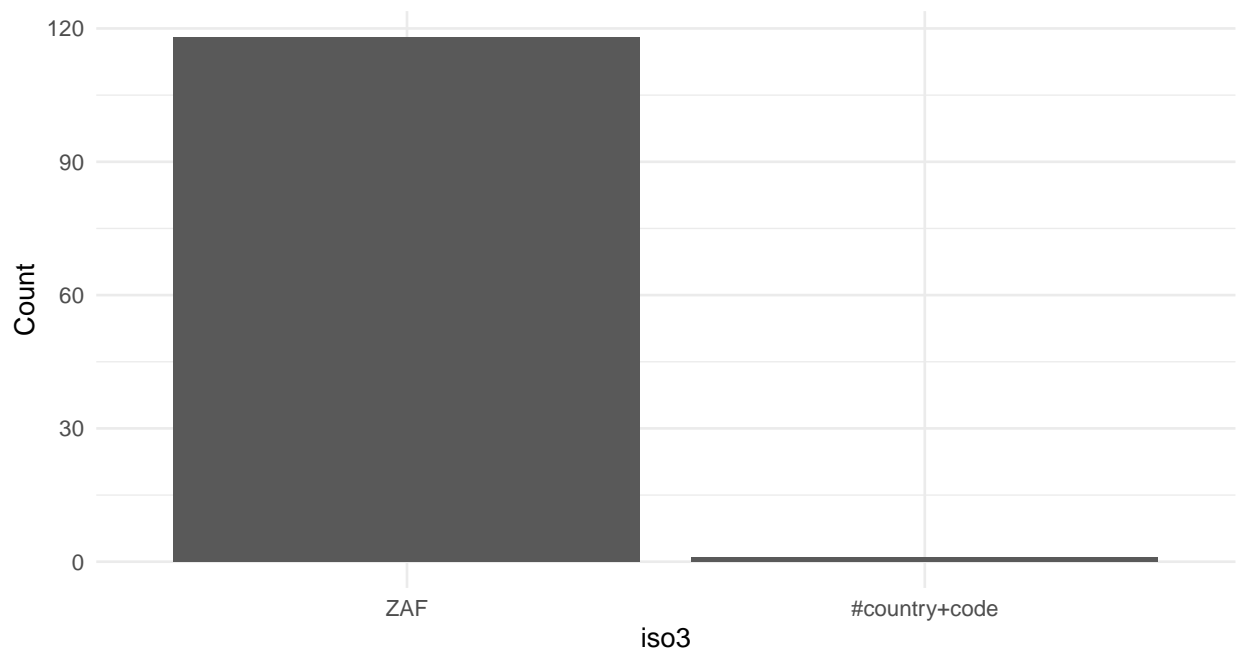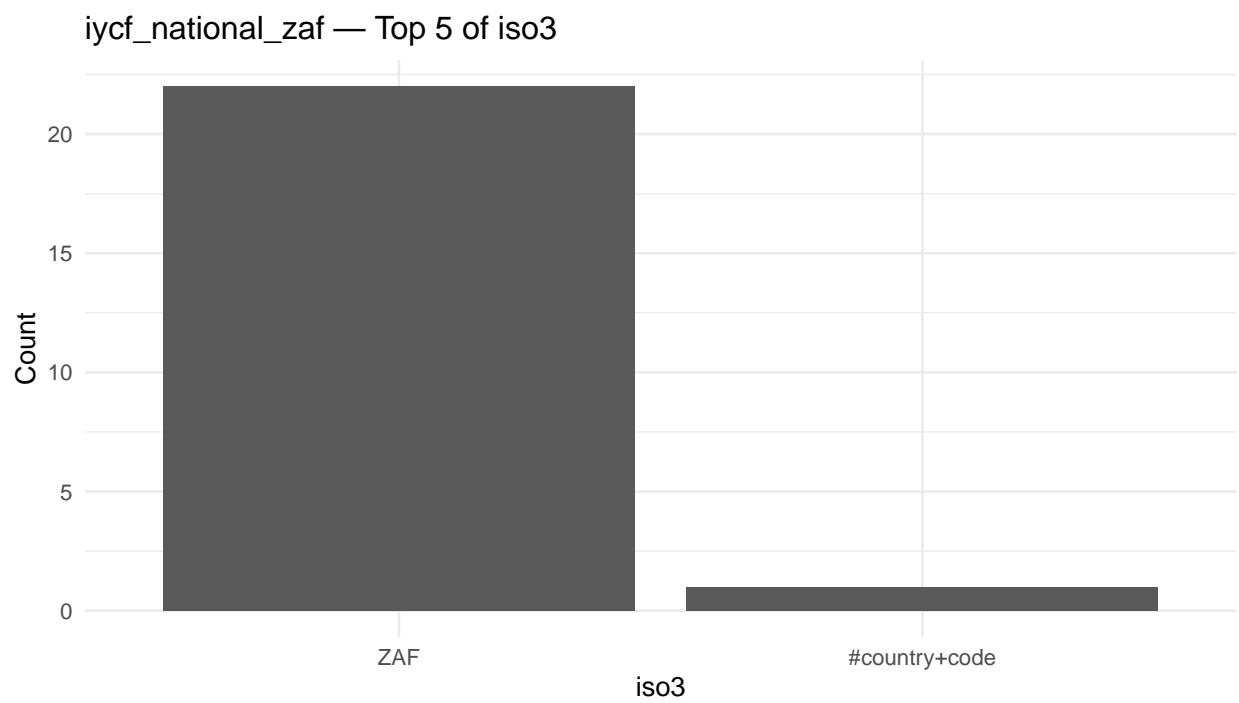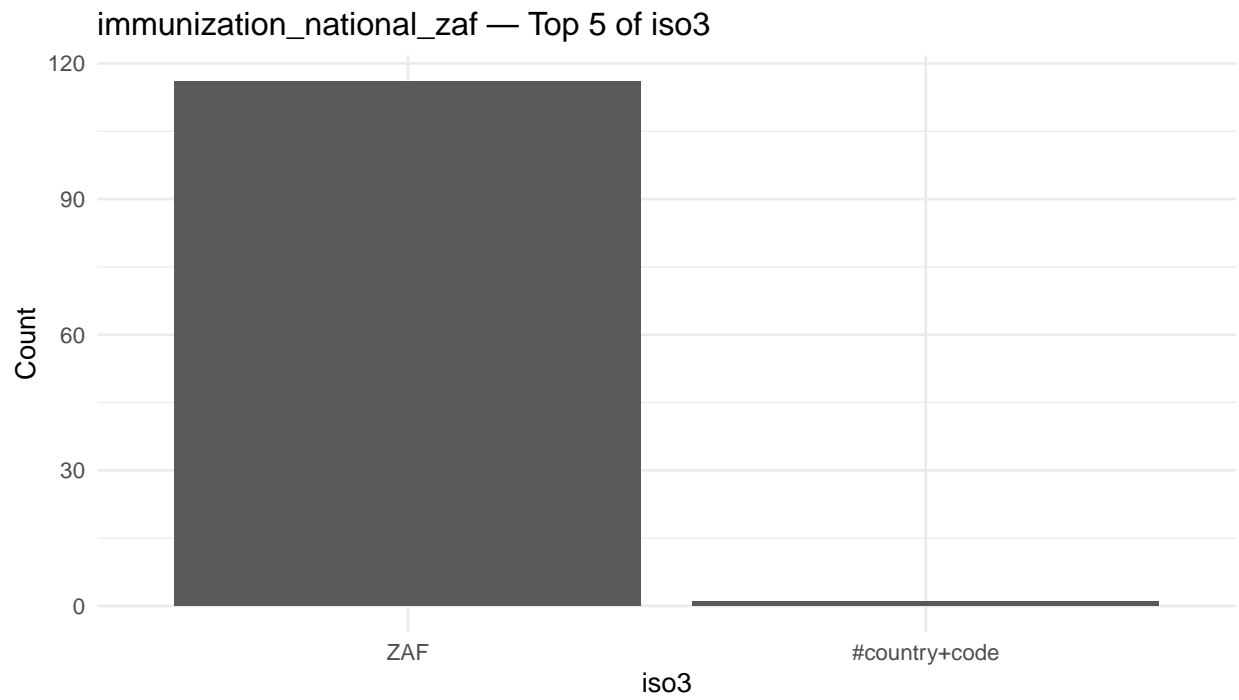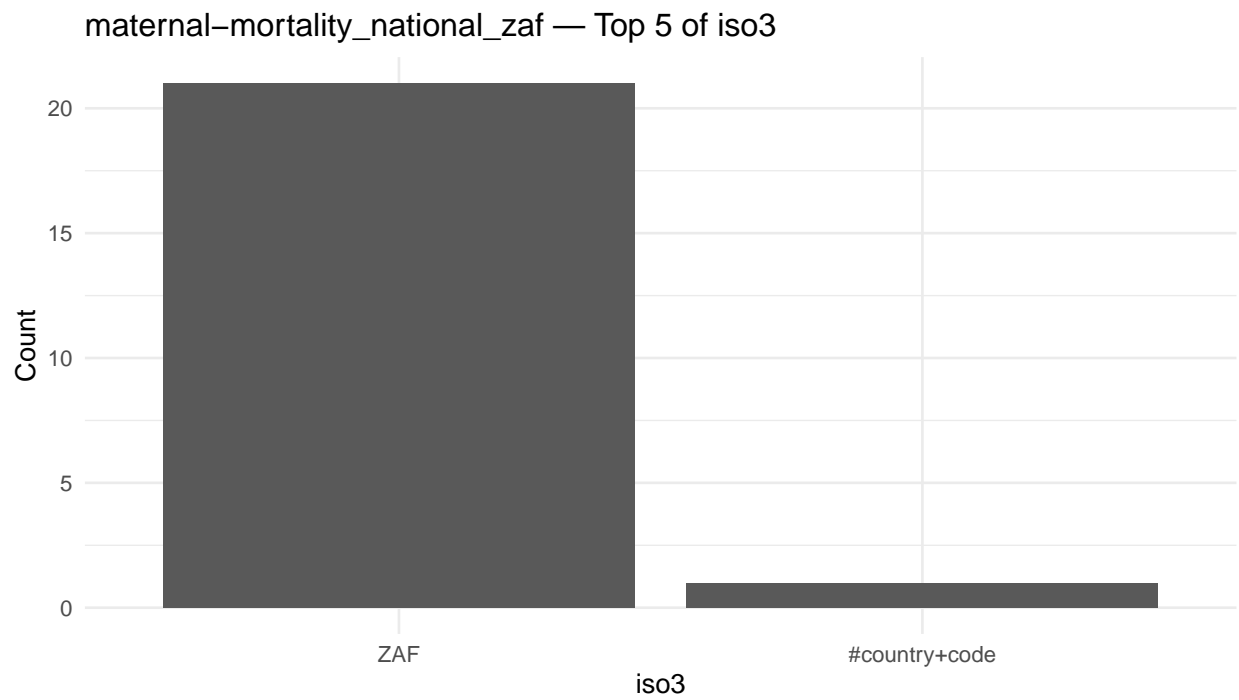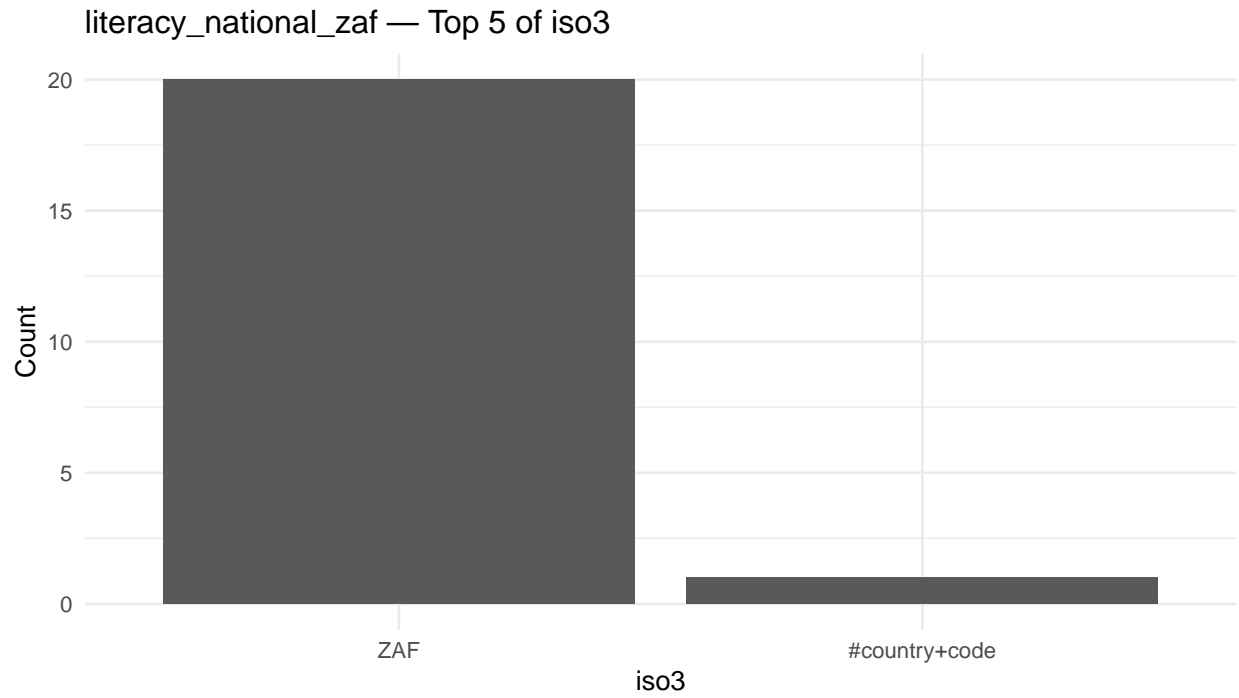
## symptoms−of−acute−respiratory−infection−ari_national_zaf — Top 5 of iso3



## toilet−facilities_national_zaf — Top 5 of iso3
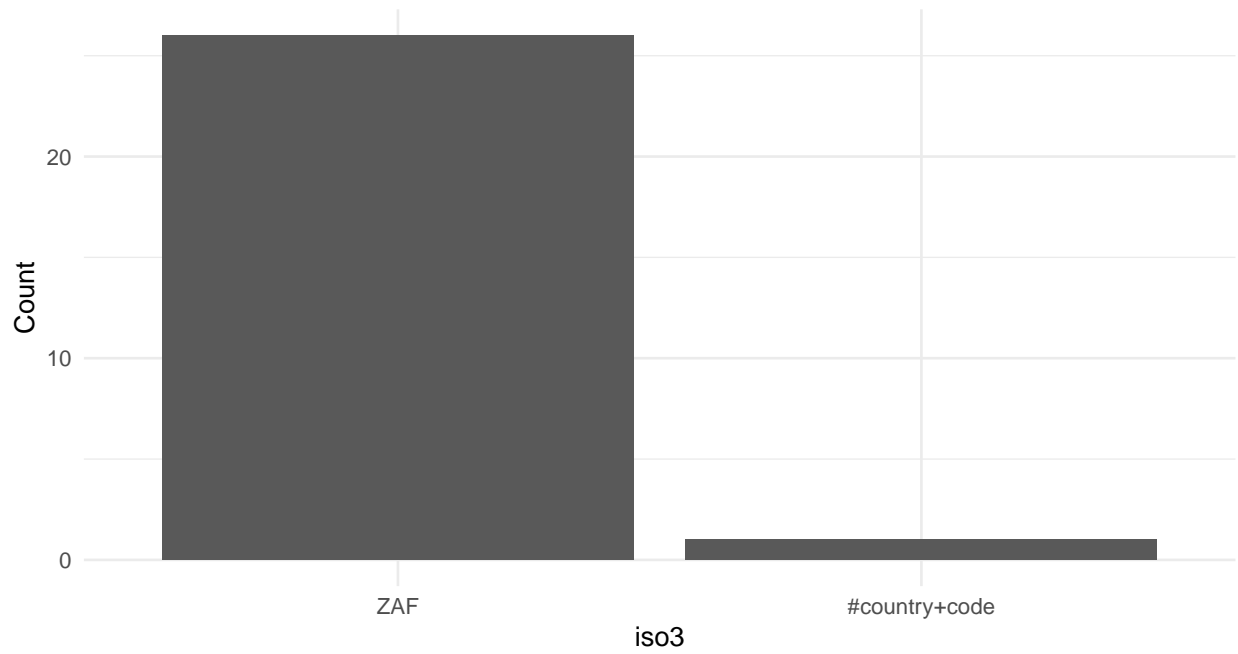
## water_national_zaf — Top 5 of iso3
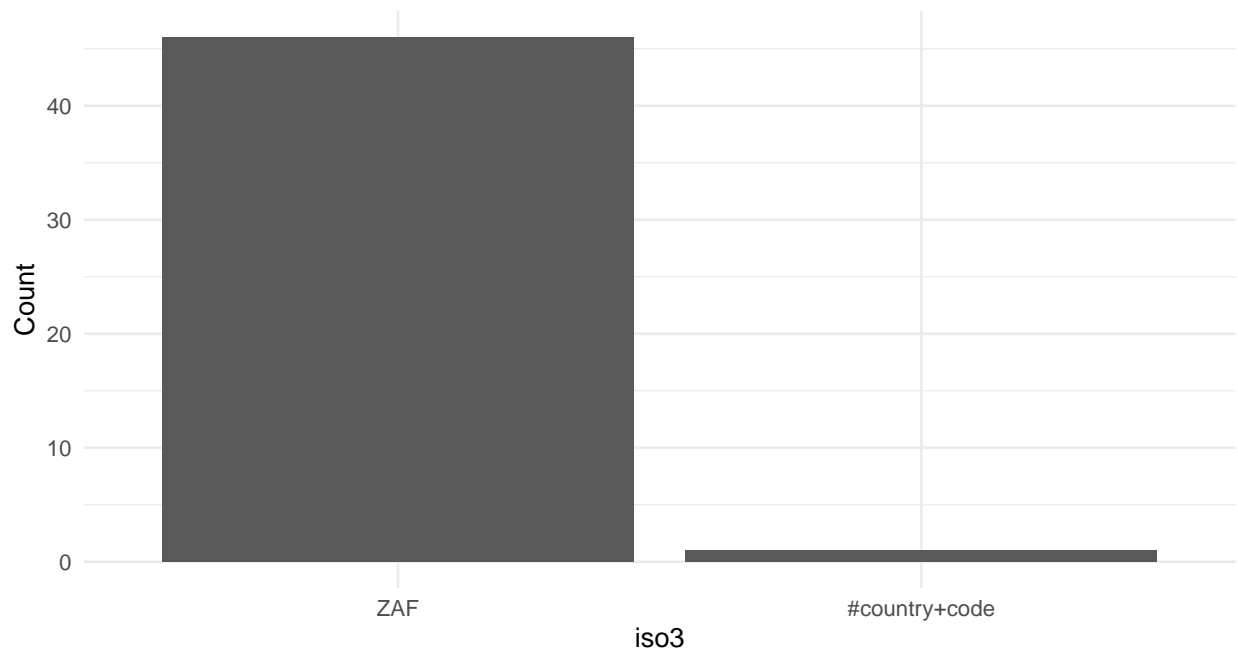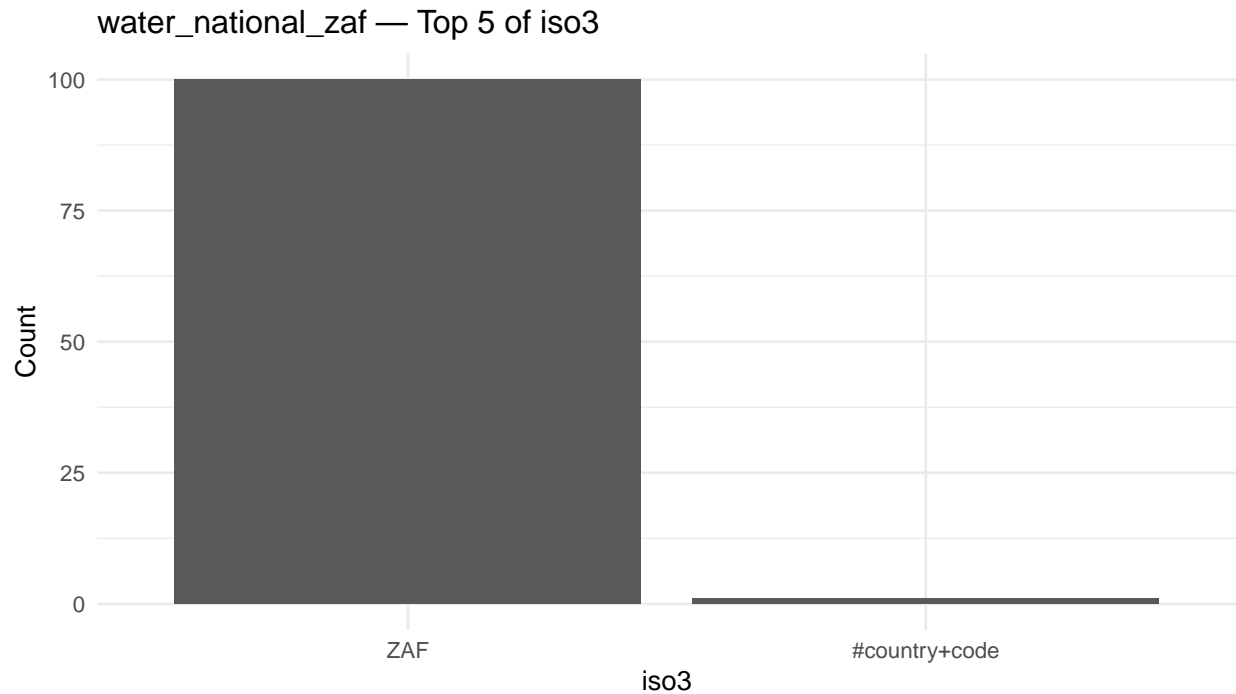


# 5 5. Correlations (All Numeric Columns)

```
plot_corr_heatmap_all <- function(df, nm){
  nums <- df |> dplyr::select(where(is.numeric))
  if(ncol(nums) < 2) return(NULL)
  mat <- cor(nums, use = "pairwise.complete.obs")
  tidy <- as_tibble(mat, rownames = "row") |>
    pivot_longer(-row, names_to = "col", values_to = "corr")
  ggplot(tidy, aes(x = row, y = col, fill = corr)) +
    geom_tile() +
    scale_fill_gradient2(limits = c(-1, 1)) +
    labs(title = paste0(nm, " - Correlation heatmap (all numeric columns)"),
         x = NULL, y = NULL, fill = "r") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
}
purrr::iwalk(dfs, function(df, nm){ p <- plot_corr_heatmap_all(df, nm); if(!is.null(p)) print(p) })
```

access–to–health–care_national_zaf — Correlation heatmap (all nur
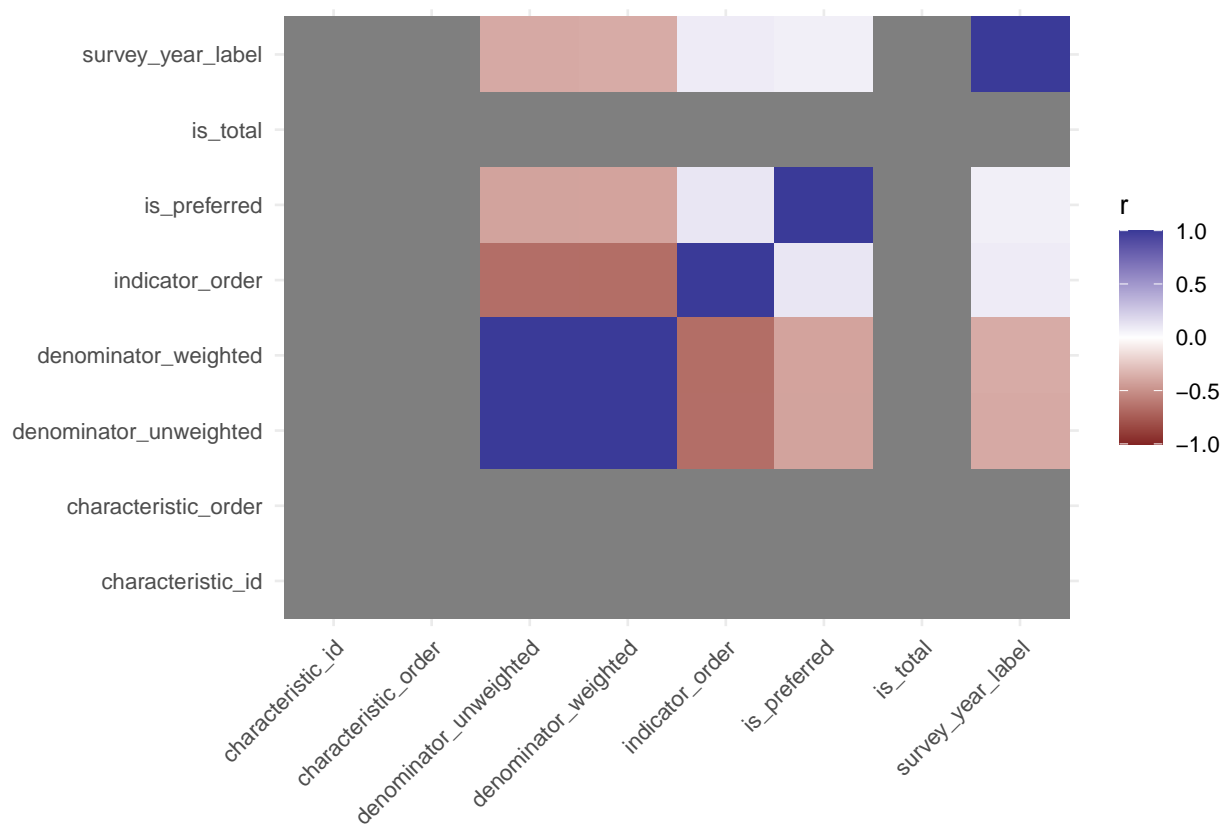
anthropometry_national_zaf — Correlation heatmap (all numeric col
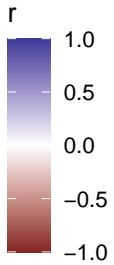
child−mortality−rates_national_zaf — Correlation heatmap (all nume
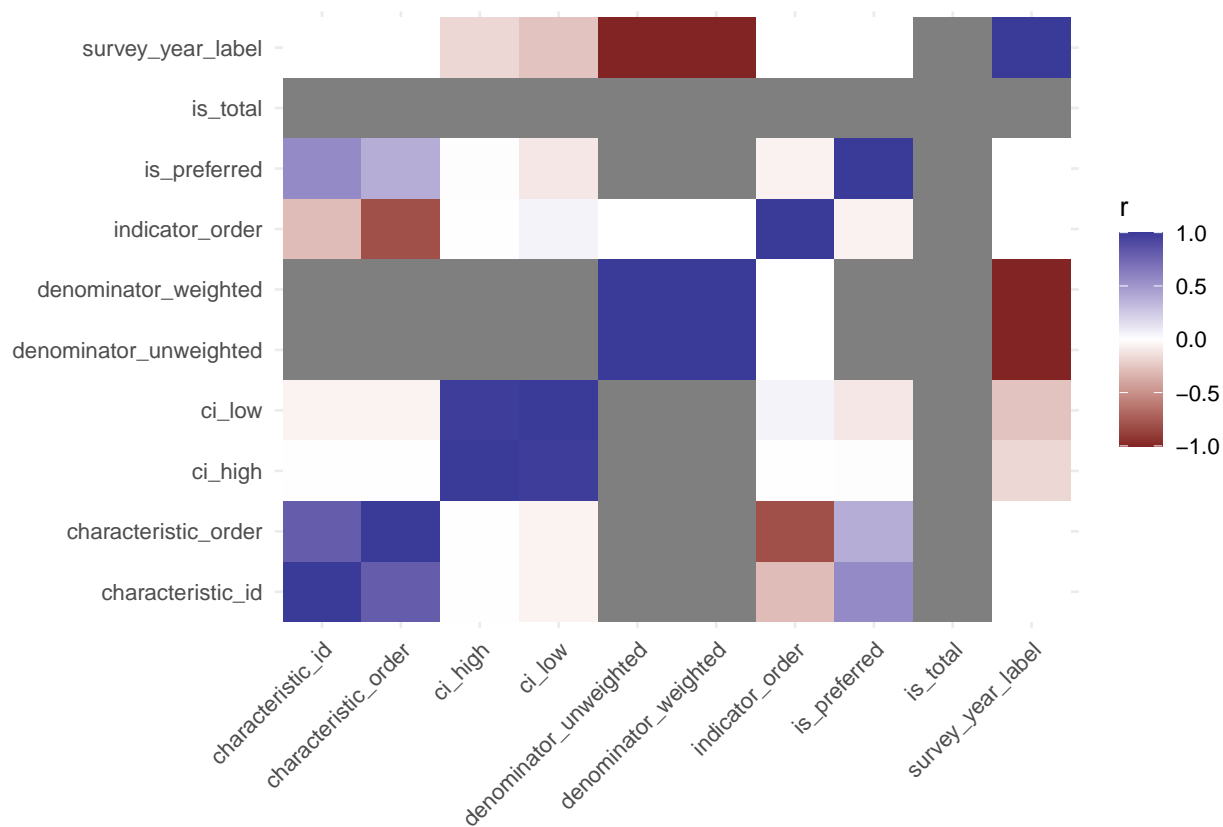
covid−19−prevention_national_zaf — Correlation heatmap (all nume
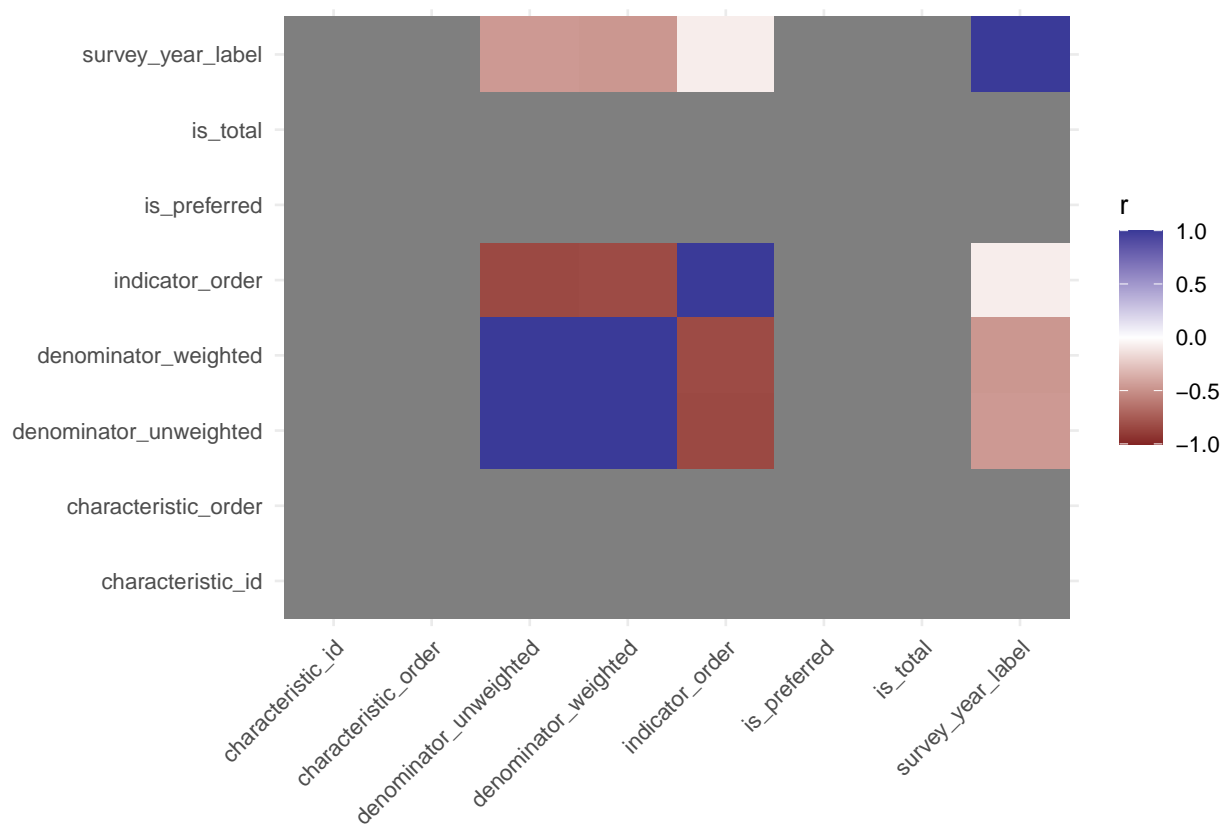
dhs−quickstats_national_zaf — Correlation heatmap (all numeric col
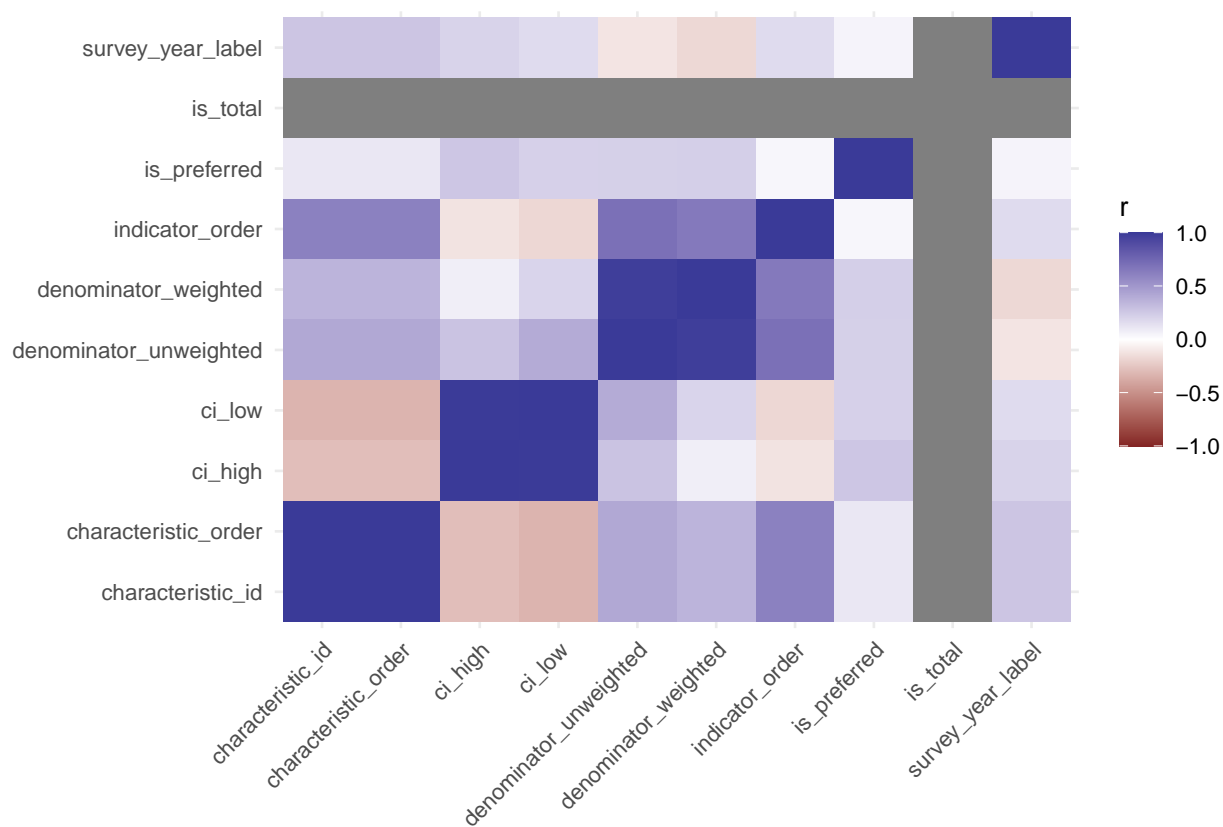
hiv−behavior_national_zaf — Correlation heatmap (all numeric colum
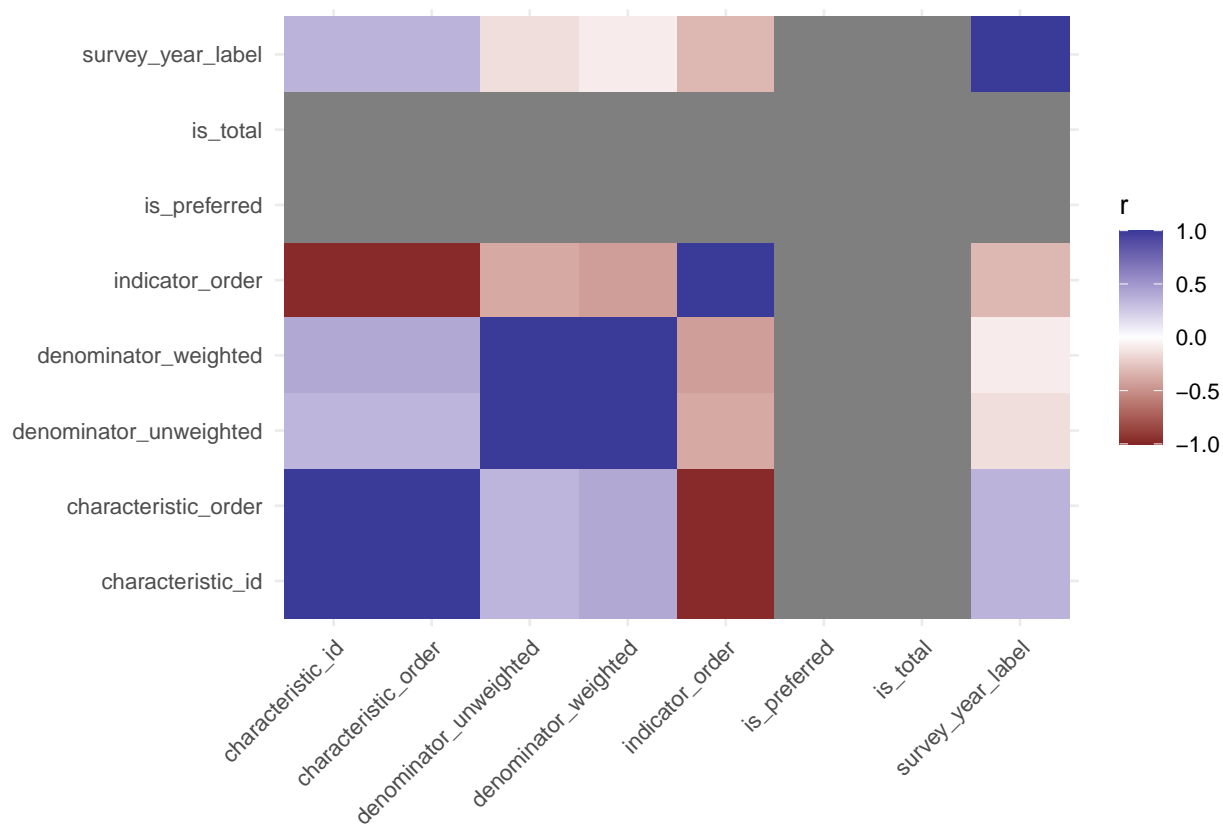
immunization_national_zaf — Correlation heatmap (all numeric colur
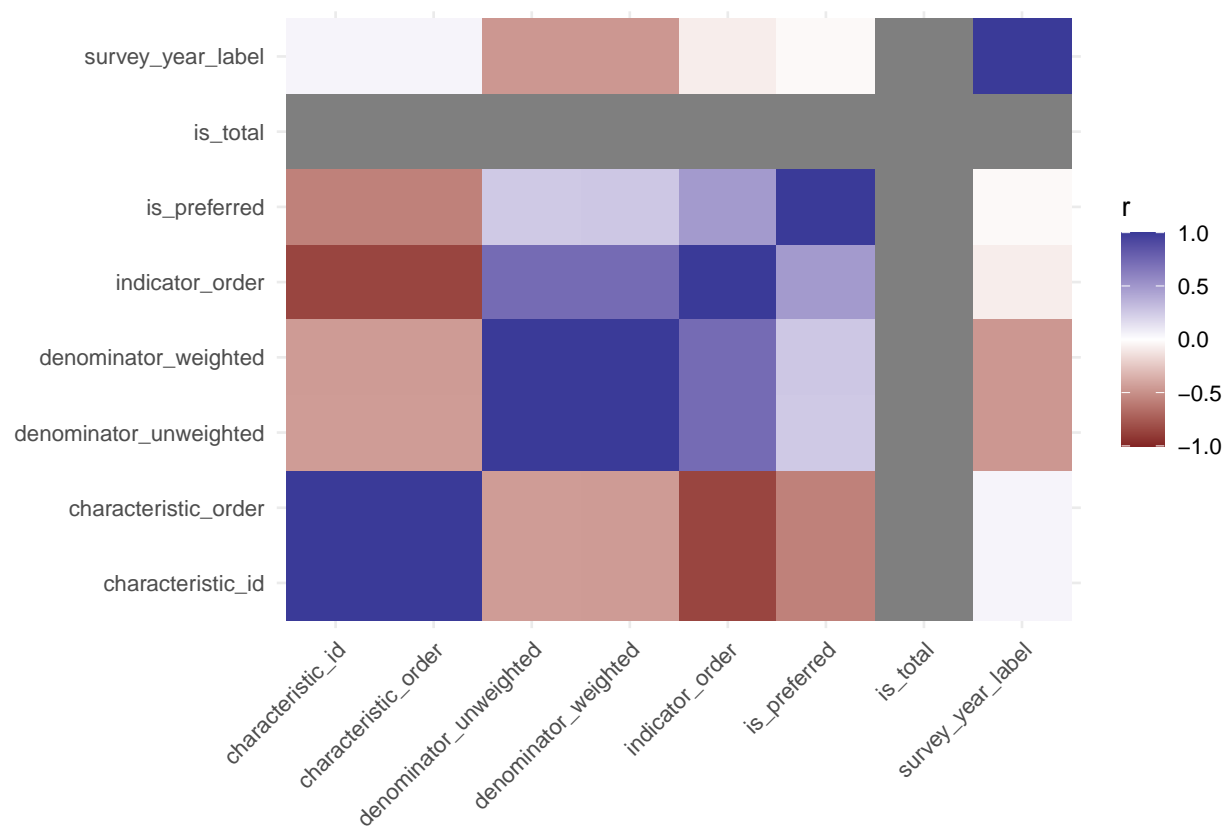
iycf_national_zaf — Correlation heatmap (all numeric columns)
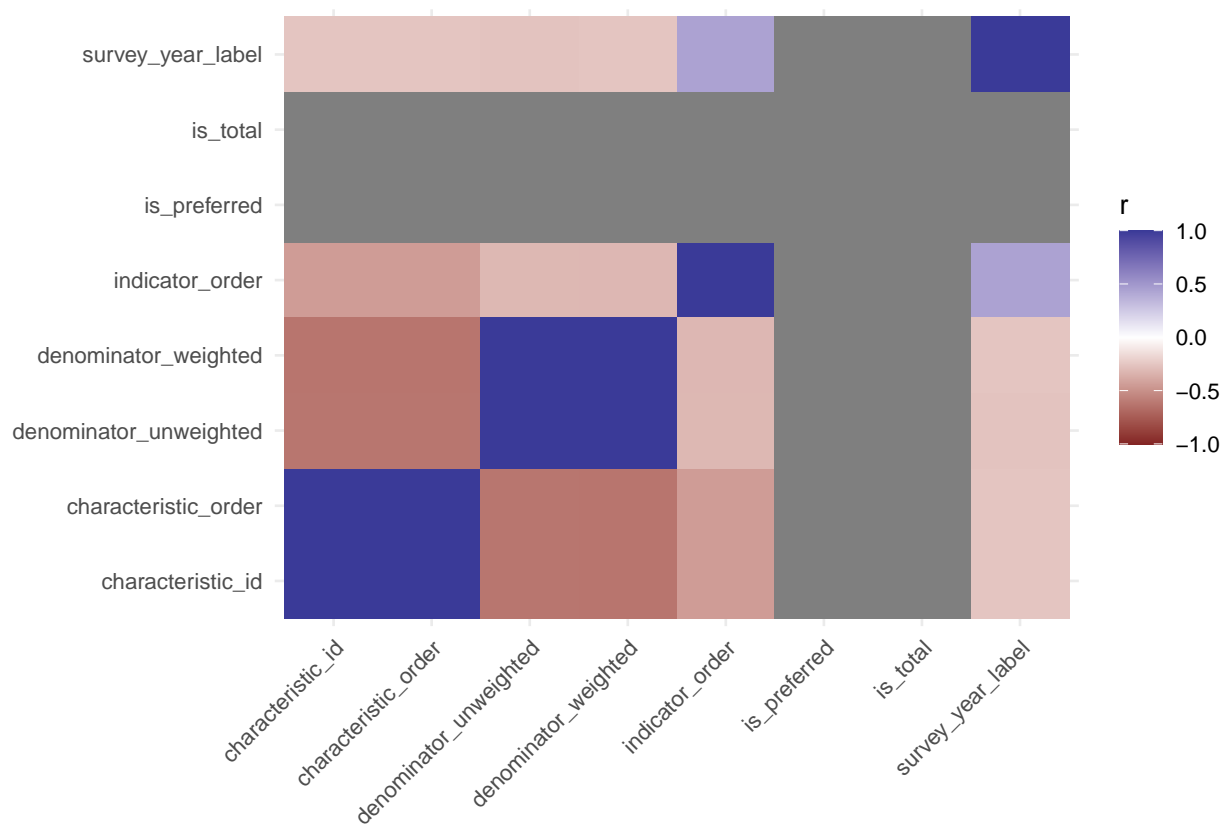
# literacy_national_zaf — Correlation heatmap (all numeric columns)
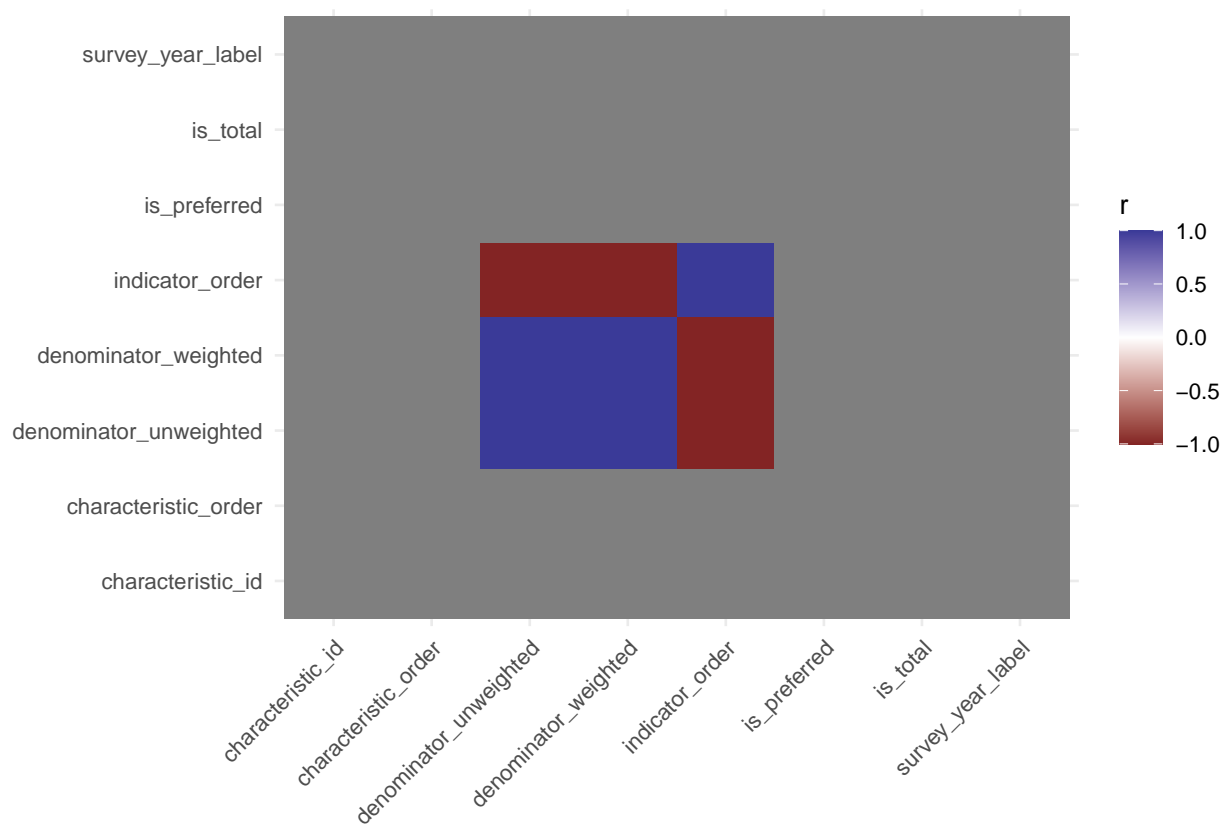
maternal−mortality_national_zaf — Correlation heatmap (all numeric)
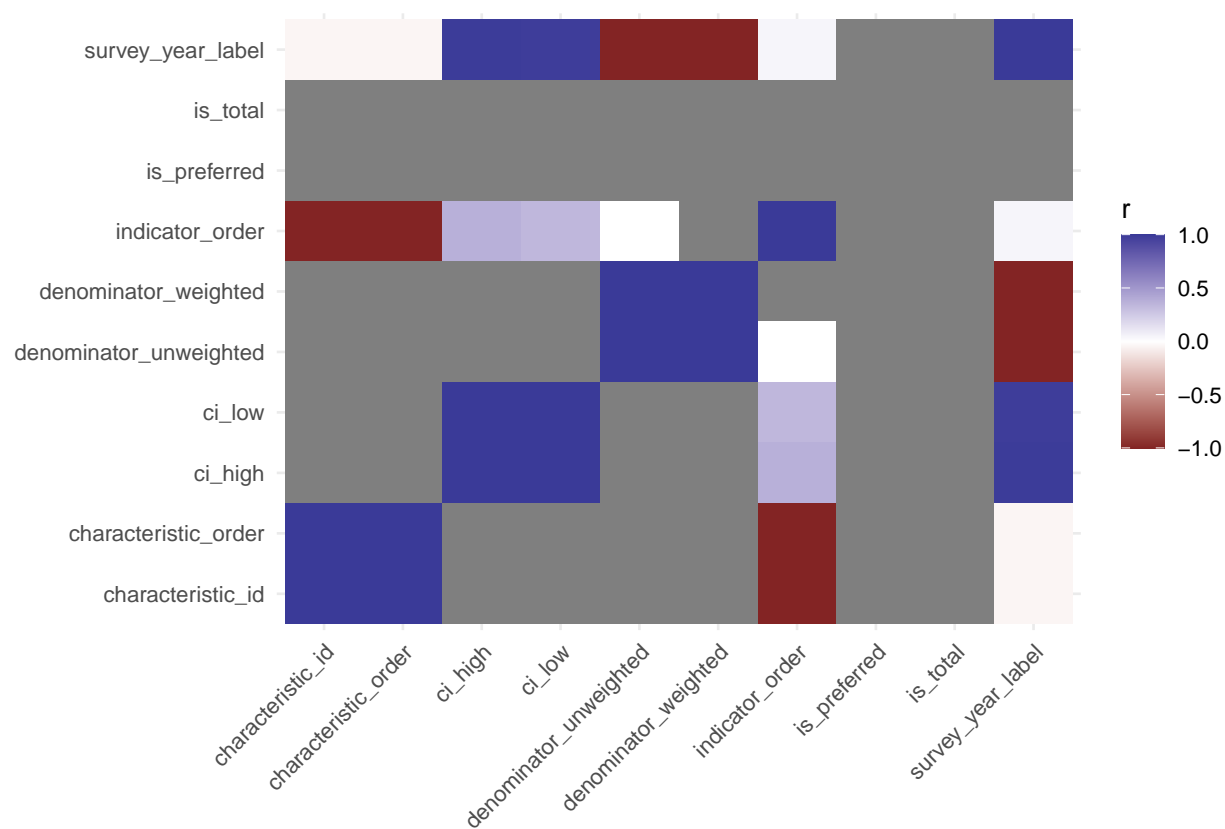
# symptoms−of−acute−respiratory−infection−ari_national_zaf — Corre

toilet–facilities_national_zaf — Correlation heatmap (all numeric colu
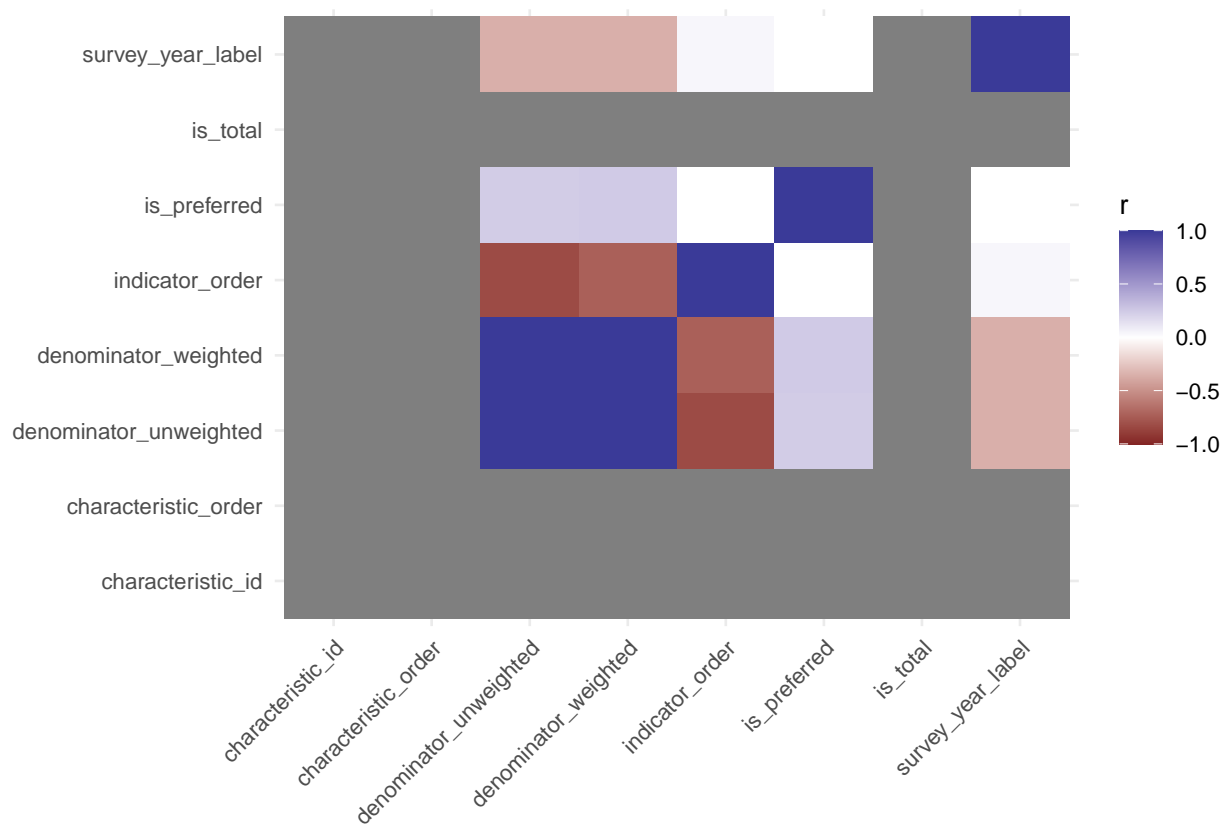
water_national_zaf — Correlation heatmap (all numeric columns)
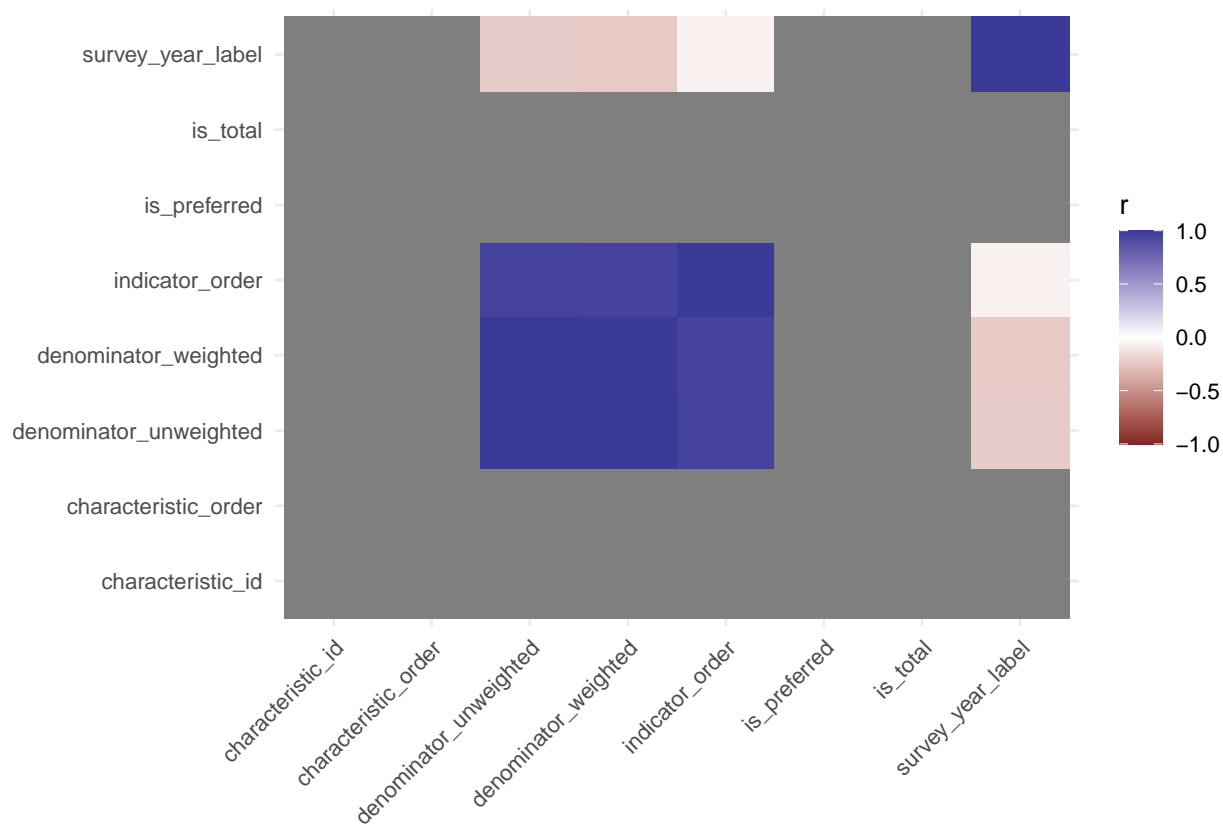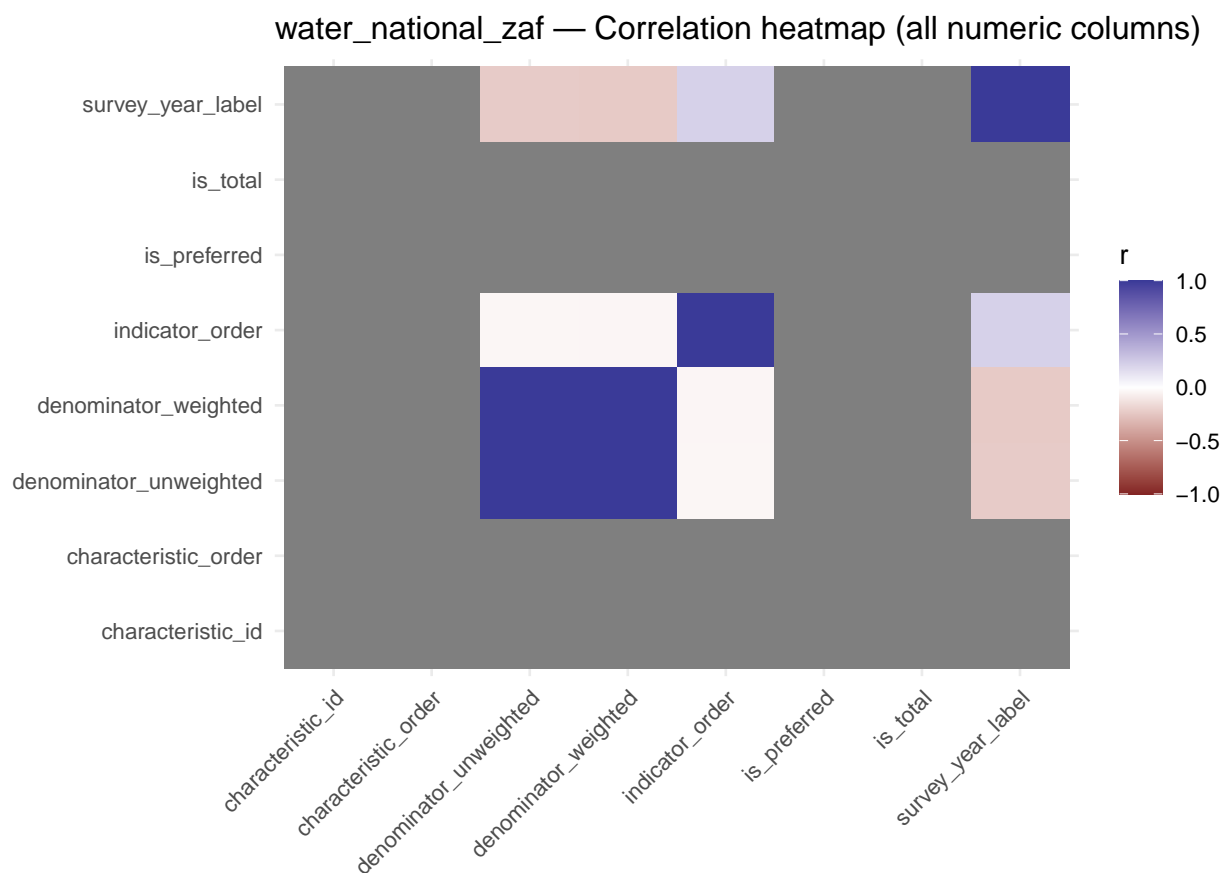
# 6 Appendix. Session Info

```
sessionInfo()
```

```
## R version 4.5.1 (2025-06-13 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 11 x64 (build 26100)
##
## Matrix products: default
##   LAPACK version 3.12.1
##
## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## time zone: Africa/Johannesburg
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
```

```
## other attached packages:
##  [1] gt_1.0.0         janitor_2.2.1   readxl_1.4.5    lubridate_1.9.4
##  [5] forcats_1.0.0   stringr_1.5.1   dplyr_1.1.4     purrr_1.1.0
##  [9] readr_2.1.5     tidyr_1.3.1     tibble_3.3.0    ggplot2_3.5.2
## [13] tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##  [1] bit_4.6.0         gtable_0.3.6     crayon_1.5.3    compiler_4.5.1
##  [5] tidyselect_1.2.1  xml2_1.4.0       parallel_4.5.1  snakecase_0.11.1
##  [9] scales_1.4.0      yaml_2.3.10      fastmap_1.2.0   R6_2.6.1
## [13] labeling_0.4.3    generics_0.1.4   knitr_1.50      pillar_1.11.0
## [17] RColorBrewer_1.1-3 tzdb_0.5.0      rlang_1.1.6     stringi_1.8.7
## [21] xfun_0.52         bit64_4.6.0-1    timechange_0.3.0 cli_3.6.5
## [25] withr_3.0.2       magrittr_2.0.3   digest_0.6.37   grid_4.5.1
## [29] vroom_1.6.5       rstudioapi_0.17.1 hms_1.1.3       lifecycle_1.0.4
## [33] vctrs_0.6.5       evaluate_1.0.5   glue_1.8.0      cellranger_1.1.0
## [37] farver_2.1.2      rmarkdown_2.29   tools_4.5.1     pkgconfig_2.0.3
## [41] htmltools_0.5.8.1
```

| dataset | | column | missing_pct | mis |
|---|---|---|---|---|
| access-to-health-care_national_zaf | | ci_high | 100.00 | |
| access-to-health-care_national_zaf | | ci_low | 100.00 | |
| access-to-health-care_national_zaf | | level_rank | 100.00 | |
| access-to-health-care_national_zaf | | region_id | 100.00 | |
| anthropometry_national_zaf | | ci_high | 100.00 | |
| anthropometry_national_zaf | | ci_low | 100.00 | |
| anthropometry_national_zaf | | level_rank | 100.00 | |
| anthropometry_national_zaf | | region_id | 100.00 | |
| child-mortality-rates_national_zaf | | level_rank | 100.00 | |
| child-mortality-rates_national_zaf | | region_id | 100.00 | |
| covid-19-prevention_national_zaf | | ci_high | 100.00 | |
| covid-19-prevention_national_zaf | | ci_low | 100.00 | |
| covid-19-prevention_national_zaf | | level_rank | 100.00 | |
| covid-19-prevention_national_zaf | | region_id | 100.00 | |
| dhs-quickstats_national_zaf | | level_rank | 100.00 | |
| dhs-quickstats_national_zaf | | region_id | 100.00 | |
| hiv-behavior_national_zaf | | ci_high | 100.00 | |
| hiv-behavior_national_zaf | | ci_low | 100.00 | |
| hiv-behavior_national_zaf | | level_rank | 100.00 | |
| hiv-behavior_national_zaf | | region_id | 100.00 | |
| immunization_national_zaf | | ci_high | 100.00 | |
| immunization_national_zaf | | ci_low | 100.00 | |
| immunization_national_zaf | | level_rank | 100.00 | |
| immunization_national_zaf | | region_id | 100.00 | |
| iycf_national_zaf | | ci_high | 100.00 | |
| iycf_national_zaf | | ci_low | 100.00 | |
| iycf_national_zaf | | level_rank | 100.00 | |
| iycf_national_zaf | | region_id | 100.00 | |
| literacy_national_zaf | | ci_high | 100.00 | |
| literacy_national_zaf | | ci_low | 100.00 | |
| literacy_national_zaf | | level_rank | 100.00 | |
| literacy_national_zaf | | region_id | 100.00 | |
| maternal-mortality_national_zaf | | level_rank | 100.00 | |
| maternal-mortality_national_zaf | | region_id | 100.00 | |
| symptoms-of-acute-respiratory-infection-ari_national_zaf | | ci_high | 100.00 | |
| symptoms-of-acute-respiratory-infection-ari_national_zaf | | ci_low | 100.00 | |
| symptoms-of-acute-respiratory-infection-ari_national_zaf | | level_rank | 100.00 | |
| symptoms-of-acute-respiratory-infection-ari_national_zaf | | region_id | 100.00 | |
| toilet-facilities_national_zaf | | ci_high | 100.00 | |
| toilet-facilities_national_zaf | | ci_low | 100.00 | |
| toilet-facilities_national_zaf | | level_rank | 100.00 | |
| toilet-facilities_national_zaf | | region_id | 100.00 | |
| water_national_zaf | | ci_high | 100.00 | |
| water_national_zaf | | ci_low | 100.00 | |
| water_national_zaf | | level_rank | 100.00 | |
| water_national_zaf | 41 | region_id | 100.00 | |
| hiv-behavior_national_zaf | | by_variable_label | 99.16 | |
| water_national_zaf | | by_variable_label | 99.01 | |
| toilet-facilities_national_zaf | | by_variable_label | 97.87 | |

| 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 |

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

| dataset | | issue_type | column | deta |
|---|---|---|---|---|
| access-to-health-care_national_zaf | | missing | by_variable_label | 4.719 |
| access-to-health-care_national_zaf | | missing | denominator_unweighted | 12.68 |
| access-to-health-care_national_zaf | | missing | denominator_weighted | 12.68 |
| access-to-health-care_national_zaf | | missing | ci_high | 100% |
| access-to-health-care_national_zaf | | missing | ci_low | 100% |
| access-to-health-care_national_zaf | | missing | level_rank | 100% |
| access-to-health-care_national_zaf | | missing | region_id | 100% |
| access-to-health-care_national_zaf | | missing | characteristic_category | 0.369 |
| access-to-health-care_national_zaf | | missing | characteristic_id | 0.369 |
| access-to-health-care_national_zaf | | missing | characteristic_label | 0.369 |
| access-to-health-care_national_zaf | | missing | characteristic_order | 0.369 |
| access-to-health-care_national_zaf | | missing | dhs_country_code | 0.369 |
| access-to-health-care_national_zaf | | missing | indicator_order | 0.369 |
| access-to-health-care_national_zaf | | missing | indicator_type | 0.369 |
| access-to-health-care_national_zaf | | missing | is_preferred | 0.369 |
| access-to-health-care_national_zaf | | missing | is_total | 0.369 |
| access-to-health-care_national_zaf | | missing | sdrid | 0.369 |
| access-to-health-care_national_zaf | | missing | survey_type | 0.369 |
| access-to-health-care_national_zaf | | missing | survey_year_label | 0.369 |
| anthropometry_national_zaf | | missing | by_variable_label | 97.37 |
| anthropometry_national_zaf | | missing | characteristic_category | 2.639 |
| anthropometry_national_zaf | | missing | characteristic_id | 2.639 |
| anthropometry_national_zaf | | missing | characteristic_label | 2.639 |
| anthropometry_national_zaf | | missing | characteristic_order | 2.639 |
| anthropometry_national_zaf | | missing | dhs_country_code | 2.639 |
| anthropometry_national_zaf | | missing | indicator_order | 2.639 |
| anthropometry_national_zaf | | missing | indicator_type | 2.639 |
| anthropometry_national_zaf | | missing | is_preferred | 2.639 |
| anthropometry_national_zaf | | missing | is_total | 2.639 |
| anthropometry_national_zaf | | missing | sdrid | 2.639 |
| anthropometry_national_zaf | | missing | survey_type | 2.639 |
| anthropometry_national_zaf | | missing | survey_year_label | 2.639 |
| anthropometry_national_zaf | | missing | denominator_unweighted | 13.16 |
| anthropometry_national_zaf | | missing | denominator_weighted | 13.16 |
| anthropometry_national_zaf | | missing | ci_high | 100% |
| anthropometry_national_zaf | | missing | ci_low | 100% |
| anthropometry_national_zaf | | missing | level_rank | 100% |
| anthropometry_national_zaf | | missing | region_id | 100% |
| child-mortality-rates_national_zaf | | missing | denominator_unweighted | 90.24 |
| child-mortality-rates_national_zaf | | missing | denominator_weighted | 90.24 |
| child-mortality-rates_national_zaf | | missing | by_variable_label | 48.78 |
| child-mortality-rates_national_zaf | | missing | ci_high | 26.83 |
| child-mortality-rates_national_zaf | | missing | ci_low | 26.83 |
| child-mortality-rates_national_zaf | | missing | characteristic_category | 2.44 |
| child-mortality-rates_national_zaf | | missing | characteristic_id | 2.44 |
| child-mortality-rates_national_zaf | 44 | missing | characteristic_label | 2.44 |
| child-mortality-rates_national_zaf | | missing | characteristic_order | 2.44 |
| child-mortality-rates_national_zaf | | missing | dhs_country_code | 2.44 |
| child-mortality-rates_national_zaf | | missing | indicator_order | 2.44 |