UNICEF Consultancy Assessment

2025-07-27

Install packages for these liberaries first

```
#install.packages(c("readxl", "readr", "dplyr", "ggplot2", "countrycode", "ggplot2"))
#install.packages("tinytex")
#tinytex::install_tinytex()
```

Install the R libraries

```
library(readxl)
library(readr)
library(dplyr)
library(ggplot2)
library(fuzzyjoin)
library(countrycode)
```

Step 1. Data Preparation

Import and clean datasets from all three sources, ensuring consistent country identifiers

- When importing data from **WPP2022_GEN_F01_DEMOGRAPHIC_INDICATORS_COMPACT_REV1** be sure to **skip the first 16 rows** to read the data correctly.

```
# load health and wealth of nations data
gdf<-read_excel("GLOBAL_DATAFLOW_2018-2022.xlsx", sheet = "Unicef data")
track<-read_excel("On-track and off-track countries.xlsx", sheet = "Sheet1")

#estimates <- read_excel("WPP2022_GEN_F01_DEMOGRAPHIC_INDICATORS_COMPACT_REV1.xlsx", sheet = "Estimates

col_names <- names(read_excel("WPP2022_GEN_F01_DEMOGRAPHIC_INDICATORS_COMPACT_REV1.xlsx", sheet = "Esti
# Default all columns to "guess"
col_types <- rep("guess", length(col_names))
# Identify which columns to force as "text"
col_types[which(col_names == "ISO3 Alpha-code")] <- "text"
col_types[which(col_names == "ISO2 Alpha-code")] <- "text"

Projections<- read_excel("WPP2022_GEN_F01_DEMOGRAPHIC_INDICATORS_COMPACT_REV1.xlsx", sheet = "Projectio
```

Filter the UN World Population Prospects, for 2022 and Type="Country/Area" only

```
projections_2022 <- Projections%>%
  filter(Year == 2022, Type == "Country/Area")
projections_2022 <- projections_2022 %>% rename(ISO3Code = `ISO3 Alpha-code`)
projections_2022 <- projections_2022 %>% rename(Births2022 = `Births (thousands)`)
projections_2022$Births2022<-as.numeric(projections_2022$Births2022)*1000
```

Map Country Names to ISO3 Codes

Since the `ISO3Code` column is missing in the `GLOBAL_DATAFLOW_2018-2022` dataset, we use the `countrycode` package to populate the `ISO3 Alpha-code`. This ensures consistency in country identifiers across all datasets.

```
# Load required library
library(countrycode)
library(dplyr)

# Convert columns to numeric

# Step 1: Clean the 'Geographic area' column
gdf <- gdf %>%
  mutate(Country = gsub("^\\(.*?\\)\\s*", "", `Geographic area`))

# Step 2: Map country names to ISO3 codes
gdf <- gdf %>%
  mutate(`ISO3 Alpha-code` = countrycode(Country, origin = "country.name", destination = "iso3c"))

# Step 3: Manually fill in missing ISO3 codes for known regions or organizations
custom_codes <- c(
  "African Union" = "AFU",
  "Americas" = "AME",
  "Arab States" = "ARB",
  "Asia and the Pacific" = "ASP",
  "Caribbean" = "CAR",
  "Central Africa" = "CAF",
  "Central America" = "CAM",
  "Central Asia" = "CAZ",
  "Eastern Africa" = "EAF",
  "Eastern Asia" = "EAS",
  "Eastern Europe and Central Asia" = "EECA",
  "Europe" = "EUR",
  "Latin America and the Caribbean" = "LAC",
  "Middle East and North Africa" = "MENA",
  "Northern Africa" = "NAF",
  "Northern America" = "NAM",
  "Oceania" = "OCE",
  "South America" = "SAM",
  "South Asia" = "SAS",
  "Sub-Saharan Africa" = "SSA",
  "Western Africa" = "WAF",
  "Western Asia" = "WAS",
  "World Bank (high income)" = "WBH",
```

```
  "World Bank (low income)" = "WBL",
  "World Bank (lower middle income)" = "WBML",
  "World Bank (upper middle income)" = "WBMU"
)

gdf <- gdf %>%
  mutate(`ISO3 Alpha-code` = ifelse(is.na(`ISO3 Alpha-code`),
                                    custom_codes[Country],
                                    `ISO3 Alpha-code`))


# Rename column for consistency
gdf <- gdf %>%
  rename(ISO3Code = `ISO3 Alpha-code`)
```

Filter Most Recent ANC4 and SBA Coverage Data (2018–2022)

Filter the most recent global coverage estimates for each country from **2018 to 2022** for the following indicators:

- **ANC4**: *"Antenatal care 4+ visits - percentage of women (aged 15–49 years) attended at least four times during pregnancy by any provider"*
- **SBA**: *"Skilled birth attendant - percentage of deliveries attended by skilled health personnel"*

```
anc4_most_recent <- gdf %>%
  filter(Indicator == "Antenatal care 4+ visits - percentage of women (aged 15-49 years) attended at lea
  group_by(Country) %>%
  filter(TIME_PERIOD == max(TIME_PERIOD, na.rm = TRUE)) %>%
  ungroup()


sbc_most_recent <- gdf %>%
  filter(Indicator == "Skilled birth attendant - percentage of deliveries attended by skilled health per
    group_by(Country) %>%
  filter(TIME_PERIOD == max(TIME_PERIOD, na.rm = TRUE))
```

Merge the above datasets together

```
anc4_merged <- left_join(track, anc4_most_recent, by = c("ISO3Code" = "ISO3Code"))
anc4<-left_join(anc4_merged, projections_2022, by = c("ISO3Code" =  "ISO3Code"))


sbc_merged<-left_join(track, sbc_most_recent,  by = c("ISO3Code" = "ISO3Code"))
sbc<-left_join(sbc_merged, projections_2022, by = c("ISO3Code" =  "ISO3Code"))
```

Step 2: Calculate Population-Weighted Coverage

Objective:
Compute population-weighted coverage for **ANC4** and **SBA** indicators by **track status** ("On-track"

vs. "Off-track") using **projected 2022 births** as weights. - For each track status group (`On-track` and `Off-track`), calculate the **weighted average** for ANC4 and SBA. - Use **projected 2022 births** as the weighting factor for each country within the group.

```
# Prepare ANC4 data with On-track/Off-track grouping
anc4_clean <- anc4 %>%
  filter(!is.na(OBS_VALUE), !is.na(Births2022), !is.na(Status.U5MR)) %>%
  mutate(
     OBS_VALUE = as.numeric(OBS_VALUE),
    ANC4_prop = OBS_VALUE / 100,
    Track = ifelse(Status.U5MR %in% c("Achieved", "On Track"), "On-track", "Off-track")
  )

# Prepare SBA data with On-track/Off-track grouping
sbc_clean <- sbc %>%
  filter(!is.na(OBS_VALUE), !is.na(Births2022), !is.na(Status.U5MR)) %>%
  mutate(
     OBS_VALUE = as.numeric(OBS_VALUE),
    SBA_prop = OBS_VALUE / 100,
    Track = ifelse(Status.U5MR %in% c("Achieved", "On Track"), "On-track", "Off-track")
  )

# Function to calculate weighted average for a group
calculate_weighted_average <- function(data, value_col, weight_col) {
  sum(data[[value_col]] * data[[weight_col]]) / sum(data[[weight_col]])
}

# Calculate and print weighted ANC4 by Track
cat("Population-weighted ANC4 by Track:\n")
```

```
## Population-weighted ANC4 by Track:
```

```
for (grp in unique(anc4_clean$Track)) {
  sub_data <- anc4_clean %>% filter(Track == grp)
  weighted_ANC4 <- calculate_weighted_average(sub_data, "ANC4_prop", "Births2022")
  cat("  ", grp, ":", round(weighted_ANC4 * 100, 1), "%\n")
}
```

```
##    Off-track : 55.4 %
##    On-track : 72.8 %
```

```
# Calculate and print weighted SBA by Track
cat("\nPopulation-weighted SBA by Track:\n")
```

```
##
## Population-weighted SBA by Track:
```

```
for (grp in unique(sbc_clean$Track)) {
  sub_data <- sbc_clean %>% filter(Track == grp)
  weighted_SBA <- calculate_weighted_average(sub_data, "SBA_prop", "Births2022")
  cat("  ", grp, ":", round(weighted_SBA * 100, 1), "%\n")
}
```

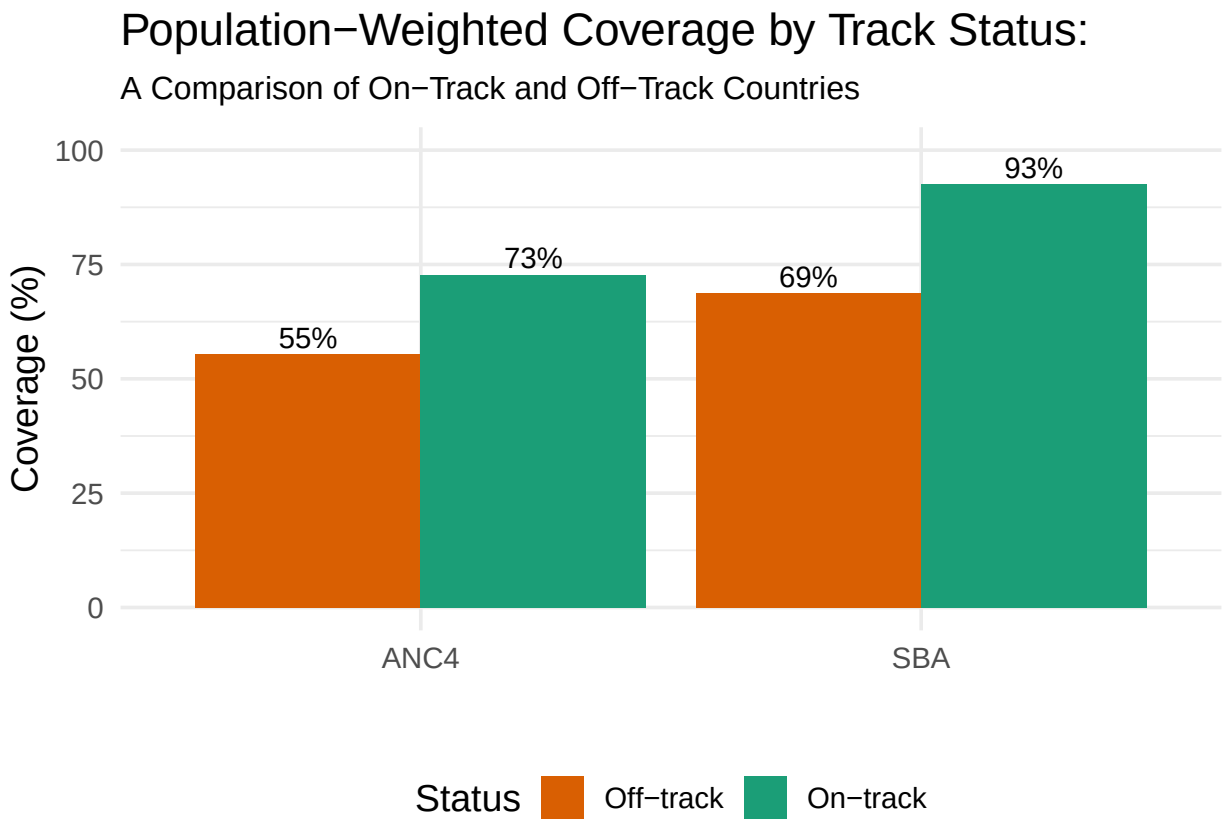
```
##    Off-track : 68.8 %
##    On-track : 92.5 %
```

Step 3: Reporting

Create a visualization to compare **population-weighted coverage** between **on-track** and **off-track** countries.

- A clear visualization comparing coverage across the two groups.

```r
library(ggplot2)

# Calculate weighted values using existing data
anc4_summary <- anc4_clean %>%
  group_by(Track) %>%
  summarise(Coverage = sum(ANC4_prop * Births2022) / sum(Births2022)) %>%
  mutate(Indicator = "ANC4")

sbc_summary <- sbc_clean %>%
  group_by(Track) %>%
  summarise(Coverage = sum(SBA_prop * Births2022) / sum(Births2022)) %>%
  mutate(Indicator = "SBA")

# Combine summaries for plotting
coverage_summary <- bind_rows(anc4_summary, sbc_summary)

# Plot using ggplot2
library(scales)  # for percentage formatting
library(ggtext)  # needed for rich text formatting

ggplot(coverage_summary, aes(x = Indicator, y = Coverage * 100, fill = Track)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.9)) +
  geom_text(aes(label = paste0(round(Coverage * 100, 0), "%")),
            position = position_dodge(width = 0.9),
            vjust = -0.3,
            size = 4) +
  labs(
    title = "Population-Weighted Coverage by Track Status:<br><span style='font-size:12pt'>A Comparison
    y = "Coverage (%)",
    x = "",
    fill = "Status"
  ) +
  scale_y_continuous(breaks = seq(0, 100, 25), limits = c(0, 100)) +
  scale_fill_manual(values = c("On-track" = "#1b9e77", "Off-track" = "#d95f02")) +
  theme_minimal(base_size = 14) +
  theme(
    legend.position = "bottom",
    plot.title = element_markdown(lineheight = 1.2)  # render HTML in title
  )
```



Disparities in Maternal Health Coverage Between On-Track and Off-Track Countries: A Population-Weighted Analysis

The chart shows population-weighted coverage rates for antenatal care (ANC4) and skilled birth attendance (SBA) across on-track and off-track countries. On-track countries have notably higher coverage, with **72.8%** for ANC4 and **92.5%** for SBA, compared to **55.4%** and **68.8%** respectively in off-track countries.

These figures highlight persistent gaps in maternal health service access between the two groups. The analysis uses **2022 birth data** to weight coverage estimates, providing a more population-representative view. However, it assumes consistent data quality and comparability across countries and does not capture subnational disparities or potential reporting limitations.