2023 under-5 mortality targets

Load packages and read datasets

```
library(tidyverse)
library(readxl)
library(kableExtra)

df_1 = read_excel("01_rawdata/GLOBAL_DATAFLOW_2018-2022.xlsx")
## this dataset includes the ANC4 and SAB indicators
df_2 =
read_excel("01_rawdata/WPP2022_GEN_F01_DEMOGRAPHIC_INDICATORS_COMPACT_REV1.xlsx" , sheet = 2 )
## This dataset includes the projected births
df_3 = read_excel("01_rawdata/On-track and off-track countries.xlsx")
## This dataset includes the countries classification (In/off track)
```

1. Data prep

1. selecting key variables and filtering the datasets
2. renaming and standardizing the variables names
3. replace missing value of pop size with the most recent reported
4. data type conversion
5. merging all datasets in one data source

```
#### df_1

#names(df_1)
#unique(df_1$`Geographic area`)
#unique(df_1$TIME_PERIOD)
df_1 = df_1 %>% select (`Geographic area` , Indicator, TIME_PERIOD , OBS_VALUE) ## select variables of interest

## Rename the variables of interest
df_1 = df_1 %>% rename(Country = `Geographic area`, Year = TIME_PERIOD)
kable( str(df_1) , format = "markdown" )

## tibble [448 × 4] (S3: tbl_df/tbl/data.frame)
##  $ Country  : chr [1:448] "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
##  $ Indicator: chr [1:448] "Antenatal care 4+ visits - percentage of women (aged 15-49 years) attended at least four times during pregnancy"| __truncated__ "Antenatal care 4+ visits - percentage of women (aged 15-49 years) attended at least four times during pregnancy"| __truncated__ "Skilled birth attendant - percentage of deliveries attended by skilled health personnel" "Skilled birth attendant - percentage of deliveries attended by skilled health personnel" ...
##  $ Year     : chr [1:448] "2018" "2020" "2018" "2020" ...
##  $ OBS_VALUE: chr [1:448] "20.9" "27.6" "58.8" "61.8" ...
```

```r
df_1$Year = as.numeric(df_1$Year) ## convert the data type into numeric
df_1$OBS_VALUE = as.numeric(df_1$OBS_VALUE)

## Lower the letters of the country names ( except the first letter)
df_1$Country = str_to_title(df_1$Country)
df_1$Country = trimws(df_1$Country) ## remove white spaces

###df_2
#names(df_2)


## select  2022 projections

df_2 = df_2 [df_2$Year == 2022,]



## rename variables
df_2 =df_2 %>%  rename(Country =  `Region, subregion, country or area *` ,
Births =  `Births (thousands)`)
df_2 = df_2 %>% select(Country , Projected_births = Births)

kable(str(df_2) , format = "markdown")
```

```
## tibble [290 × 2] (S3: tbl_df/tbl/data.frame)
##  $ Country         : chr [1:290] "WORLD" NA "Sub-Saharan Africa" "Northern
Africa and Western Asia" ...
##  $ Projected_births: chr [1:290] "133990.59899999999" NA
"39949.955000000002" "11522.254000000001" ...
```

```r
df_2$Projected_births = as.numeric(df_2$Projected_births) ## convert to
numeric
df_2 = df_2 [! is.na(df_2$Country) ,]
df_2$Country [df_2$Country == "Asia And The Pacific"] = "Asia"

## Lower the letters of the country names ( except the first letter)
df_2$Country = str_to_title(df_2$Country)
df_2$Country = trimws(df_2$Country) ## remove white spaces



## merge df_1 with df_2
df_all = df_1 %>% full_join(df_2 , by = c('Country' ))



## df_3
#names(df_3)
```

```
df_3 = df_3 %>% rename( Country = OfficialName) ## Rename

df_all = df_all %>% full_join(df_3 , by = "Country") ## merge with dataset
kable (str(df_all) , format = "markdown")

## tibble [598 × 7] (S3: tbl_df/tbl/data.frame)
##  $ Country         : chr [1:598] "Afghanistan" "Afghanistan" "Afghanistan"
"Afghanistan" ...
##  $ Indicator       : chr [1:598] "Antenatal care 4+ visits - percentage of
women (aged 15-49 years) attended at least four times during pregnancy"|
__truncated__ "Antenatal care 4+ visits - percentage of women (aged 15-49
years) attended at least four times during pregnancy"| __truncated__ "Skilled
birth attendant - percentage of deliveries attended by skilled health
personnel" "Skilled birth attendant - percentage of deliveries attended by
skilled health personnel" ...
##  $ Year            : num [1:598] 2018 2020 2018 2020 2022 ...
##  $ OBS_VALUE       : num [1:598] 20.9 27.6 58.8 61.8 56.8 71 77.8 99.8
69.8 98.8 ...
##  $ Projected_births: num [1:598] 1447 1447 1447 1447 45839 ...
##  $ ISO3Code        : chr [1:598] "AFG" "AFG" "AFG" "AFG" ...
##  $ Status.U5MR     : chr [1:598] "Acceleration Needed" "Acceleration
Needed" "Acceleration Needed" "Acceleration Needed" ...

df_all$Projected_births = as.numeric(df_all$Projected_births)

df_all = df_all [! is.na(df_all$Status.U5MR) ,] ## remove cases where status
is not reported
df_all = df_all [! is.na(df_all$OBS_VALUE) , ]


## filter for only the most recent estimate for each indicator

df_all = df_all %>%
    group_by(Country , Indicator) %>%
    arrange(desc(Year)) %>%
    slice(1)

## Export the final dataset
writexl::write_xlsx(df_all , 'datasets_cleaned_merged.xlsx')
```

2. Calculate weighted averages for on-track and off-track countries

```
summary_tbl = df_all %>% group_by(Status.U5MR , Indicator) %>%
summarise(weighted_avg = sum(OBS_VALUE * Projected_births , na.rm = T) /
sum(Projected_births , na.rm = T) )

summary_tbl$Indicator [summary_tbl$Indicator == "Antenatal care 4+ visits -
percentage of women (aged 15-49 years) attended at least four times during
pregnancy by any provider"] = "ANC4"
```

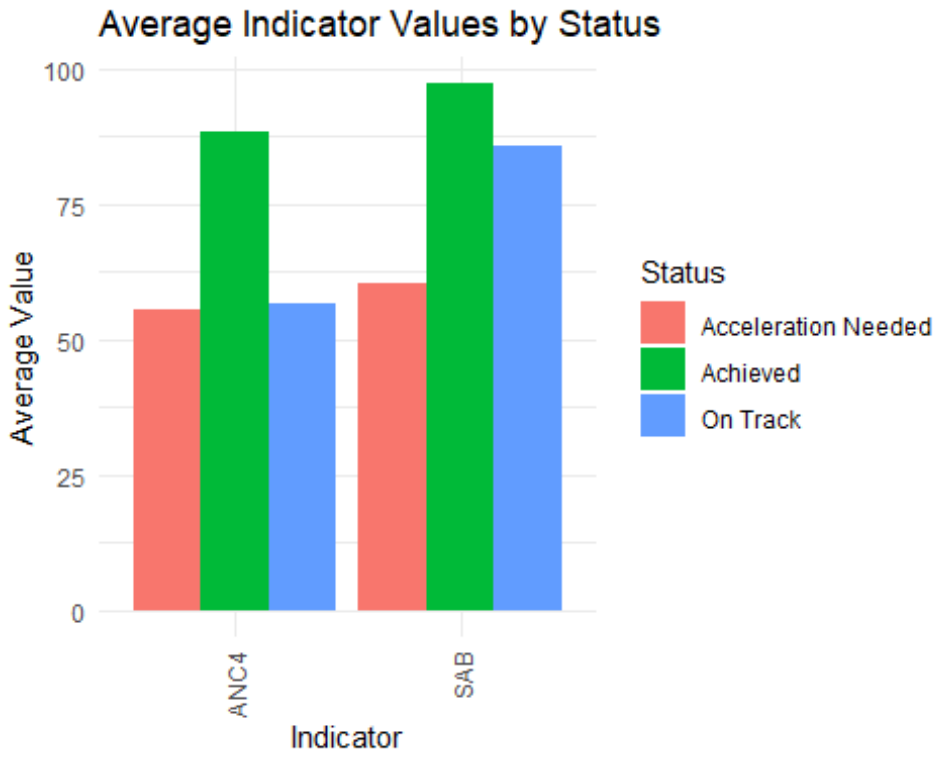
```r
summary_tbl$Indicator [summary_tbl$Indicator == "Skilled birth attendant - percentage of deliveries attended by skilled health personnel"] = "SAB"

kable(summary_tbl , format =  "markdown")
```

Status.U5MR	Indicator	weighted_avg
Acceleration Needed	ANC4	55.59325
Acceleration Needed	SAB	60.42495
Achieved	ANC4	88.40933
Achieved	SAB	97.17840
On Track	ANC4	56.53749
On Track	SAB	85.71095

```r
#Create a visualization of your choice comparing population-weighted coverage estimates for on-track and off-track countries for each indicator, with a short paragraph on interpretation and caveats.


# Create the bar chart
ggplot(summary_tbl, aes(x = Indicator, y = weighted_avg, fill = Status.U5MR))
+
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Average Indicator Values by Status",
    x = "Indicator",
    y = "Average Value",
    fill = "Status"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```



This graph demonstrates that countries that have achieved the targets of mortality rate among children below 5 years tend to have higher percentages of at least 4 antenatal care visits and higher percentages of deliveries attended by skilled health personnel. This suggests that improving access to antenatal care and skilled birth attendance may be crucial factors in reducing under-5 mortality rates.

Caveat: Not all countries reported recent values of the two indicators. [The data points shown may not represent the most current situation in each country.]