SDG INDICATORS CALCULATIONS - Zambia Example

INTRODUCTION

This R markdown template will take through computation of some SDG 3, 5 and 8 key indicators namely;

- SDG Indicator 3.7.2: Adolescent birth rate (aged 10–14 years; aged 15–19 years) per 1,000 women in that age group.
- SDG INDICATOR 5.3.1: Proportion of women aged 20–24 years who were married or in a union before age 15 and before age 18.
- SDG INDICATOR 8.6.1: Proportion of youth (aged 15-24 years) not in education, employment or training.

LOAD LIBRARIES

```
if(!require("pacman")) install.packages("pacman")
```

```
## Loading required package: pacman
```

```
pacman::p_load(tidyverse, haven, forcats, gt, sf)
```

LOAD DATA

This represent 10% from the Zambia 2010 population and houses census data whose sole purpose is for this exercise.

```
demographics <- haven::read_sav("input/DataExercise/SPSS files/DemographicsDIST.sav")

sf_feature <- sf::read_sf("input/TUC/zambia_constituency_reproj_pop_GHS-DU-TUC.shp")
```

SDG 3 : Good Health and Well Being

SDG TARGET 3.7: By 2030, ensure universal access to sexual and reproductive health-care services, including for family planning, information and education, and the integration of reproductive health into national strategies and programmes.

SDG Indicator 3.7.2: Adolescent birth rate (aged 10–14 years; aged 15–19 years) per 1,000 women in that age group adolescent_birth_rate_10_14 = (number of women aged 10 - 14 with live birth in the past 12 months / number of women aged 10 - 14)*1000

adolescent_birth_rate_14_19 = (number of women aged 14 - 19 with live birth in the past 12 months / number of women aged 14 - 19)*1000

Input Variables Definition:

- *P4_SEX_P = SexP4_SEX_P*
- *P5_AGE_P = Age*
- *P42_LAST_12_MON_P = Live births in the past 12 months*
- *CONST_P = Constituency (Admin 4)*

Output Variables Definition:

- *total.ado.10_14 = Total female adolescents aged 10 to 14*
- *total.ado.15_19 = Total female adolescents aged 15 to 19*
- *total.ado.birth.10_14 = Total female adolescents aged 10 to 14 with live birth in the past 12 months*
- *total.ado.birth.15_19 = Total female adolescents aged 15 to 19 with live birth in the past 12 months*
- *abr.10_14 = Adolescent aged 10 to 14 birth rate*
- *abr.15_19 = Adolescent aged 15 to 19 birth rate*

Methodology:

1- Filter the Sex to Female by using the P4_SEX_P == 2

2- Group by the constituency

3- Compute summary statistics

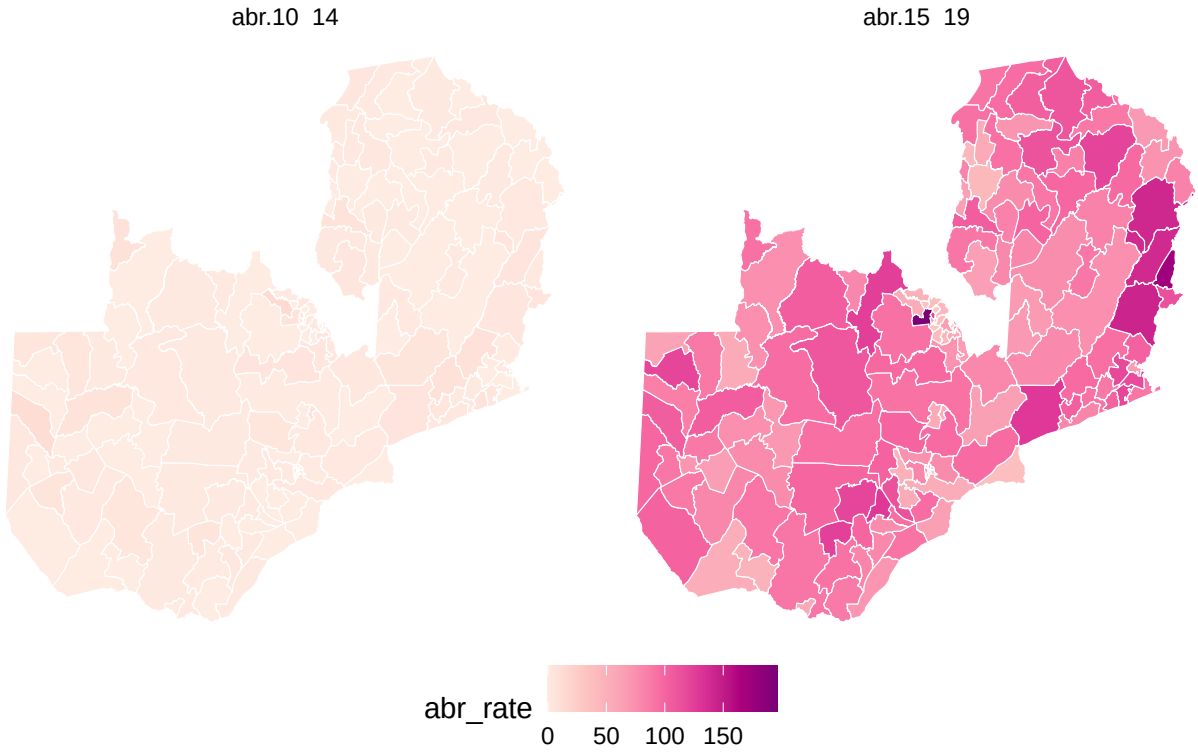
```
adolescent_birth_rate <-
  demographics |>
  dplyr::filter(P4_SEX_P == 2) |>
  dplyr::group_by(CONST_P) |>
  dplyr::summarise(
    total.ado.10_14 = sum(dplyr::between(P5_AGE_P, 10, 14), na.rm = TRUE),
    total.ado.15_19 = sum(dplyr::between(P5_AGE_P, 15, 19), na.rm = TRUE),
    total.ado.birth.10_14 = sum(dplyr::between(P5_AGE_P, 10, 14) &
                                  P42_LAST_12_MON_P == 1, na.rm = TRUE),
    total.ado.birth.15_19 = sum(dplyr::between(P5_AGE_P, 15, 19) &
                                  P42_LAST_12_MON_P == 1, na.rm = TRUE),
    abr.10_14 = round((total.ado.birth.10_14 / total.ado.10_14) * 1000,2),
    abr.15_19 = round((total.ado.birth.15_19 / total.ado.15_19) * 1000,2)
  ) |>
  dplyr::mutate(
    CONST_P_name = forcats::as_factor(CONST_P)
  ) |>
  dplyr::select(
    CONST_P_name,
    everything()
  )

# Print table
adolescent_birth_rate |>
  head() |>
  gt::gt()
```

Visualization

Constituency	Constituency	total.ado.10_14	total.ado.15_19	total.ado.birth.10_14	total.ado.birt
Chisamba	1	753	622	2	
Katuba	2	613	523	0	
Keembe	3	883	740	4	
Bwacha	4	588	559	0	
Kabwe Central	5	881	950	6	
Kapiri Mposhi	6	1828	1507	2	

```
sf_feature |>
  dplyr::select(NAME1_, geometry) |>
  dplyr::left_join(
    adolescent_birth_rate,
    by = c("NAME1_" = "CONST_P_name")
  ) |>
  tidyr::pivot_longer(
    cols = c("abr.10_14", "abr.15_19"),
    names_to = "abr_cat",
    values_to = "abr_rate"
  ) |>
  ggplot2::ggplot() +
  ggplot2::geom_sf(aes(fill = abr_rate), color = "white") +
  ggplot2::scale_fill_distiller(palette = "RdPu", direction = 1) +
  ggplot2::theme_void() +
  ggplot2::theme(
    legend.position = "bottom"
  ) +
  ggplot2::facet_grid(~abr_cat)
```



SDG 5 : GENDER EQUALITY

SDG TARGET 5.3 : Eliminate all harmful practices, such as child, early and forced marriage and female genital mutilations

SDG INDICATOR 5.3.1: Proportion of women aged 20–24 years who were married or in a union before age 15 and before age 18 child_marriage_before_15 = (number of women aged 20-24 who were married before age 15 / total number of women aged 20-24)* 100

child_marriage_before_18 = (number of women aged 20-24 who were married before age 18 / total number of women aged 20-24)* 100

Input Variables Definition:

- *P4_SEX_P = Sex (1 - Male, 2 - Female)*
- *P5_AGE_P = Age*
- *CONST_P = Constituency*

Output Variable Definition:

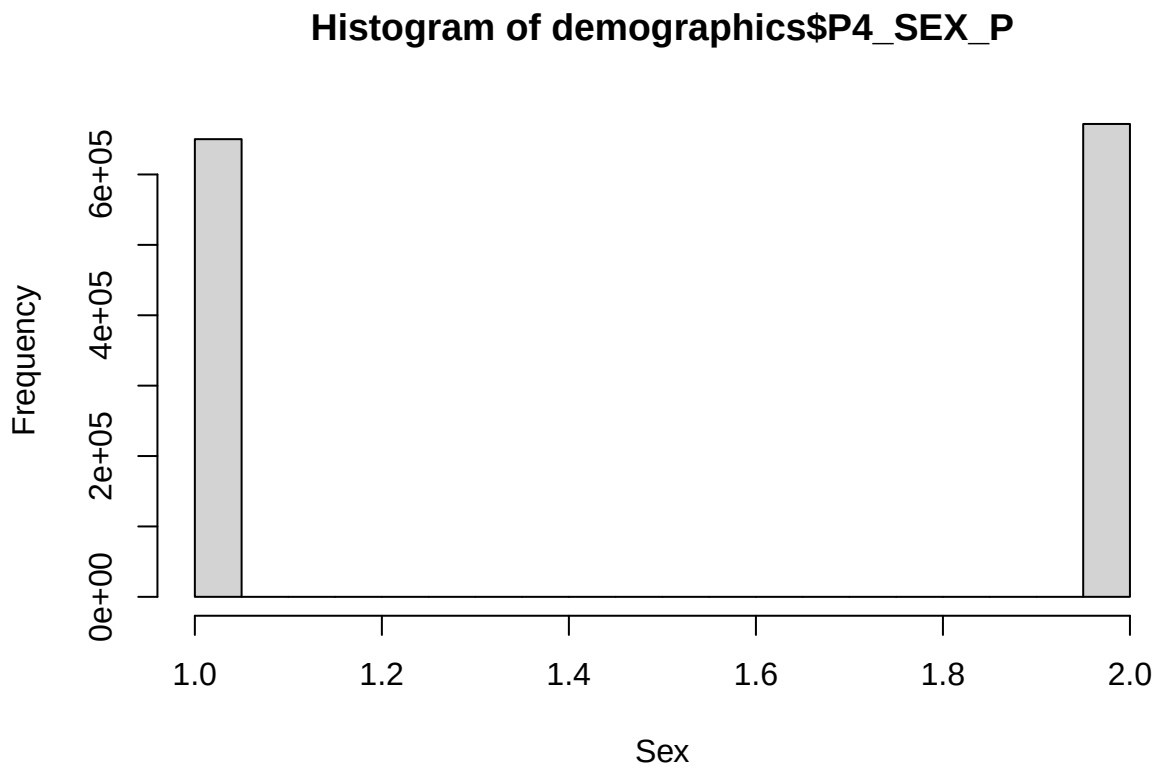
- *total.girls.20_24 = Total number of females aged 20 to 24*
- *married.before.15 = Total number of females married before 15*
- *married.before.18 = Total number of females married before 18*

- *prop.cm.before.15 = Proportion of child marriages before 15*
- *prop.cm.before.18 = Proportion of child marriages before 18*

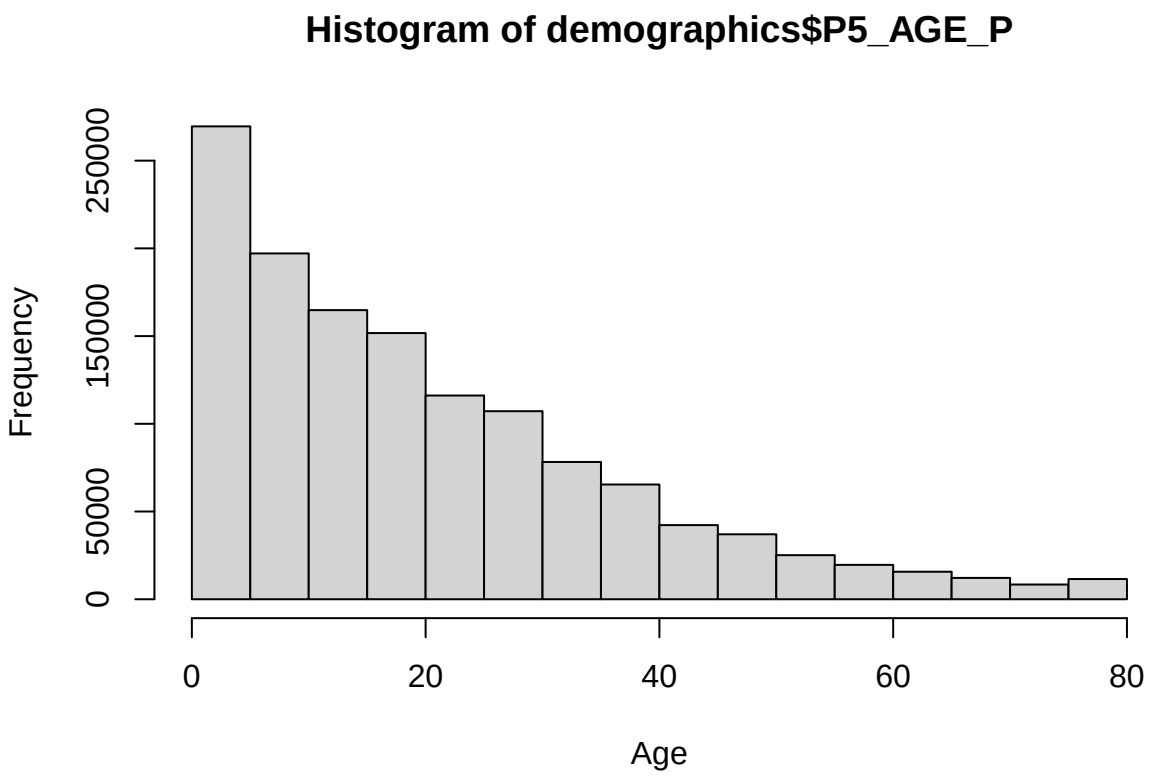
Methodolgy :

Step1: Explore Variables

```
# Frequency distribution by sex and age
demographics$P4_SEX_P |> hist(xlab = "Sex", ylab = "Frequency")
```

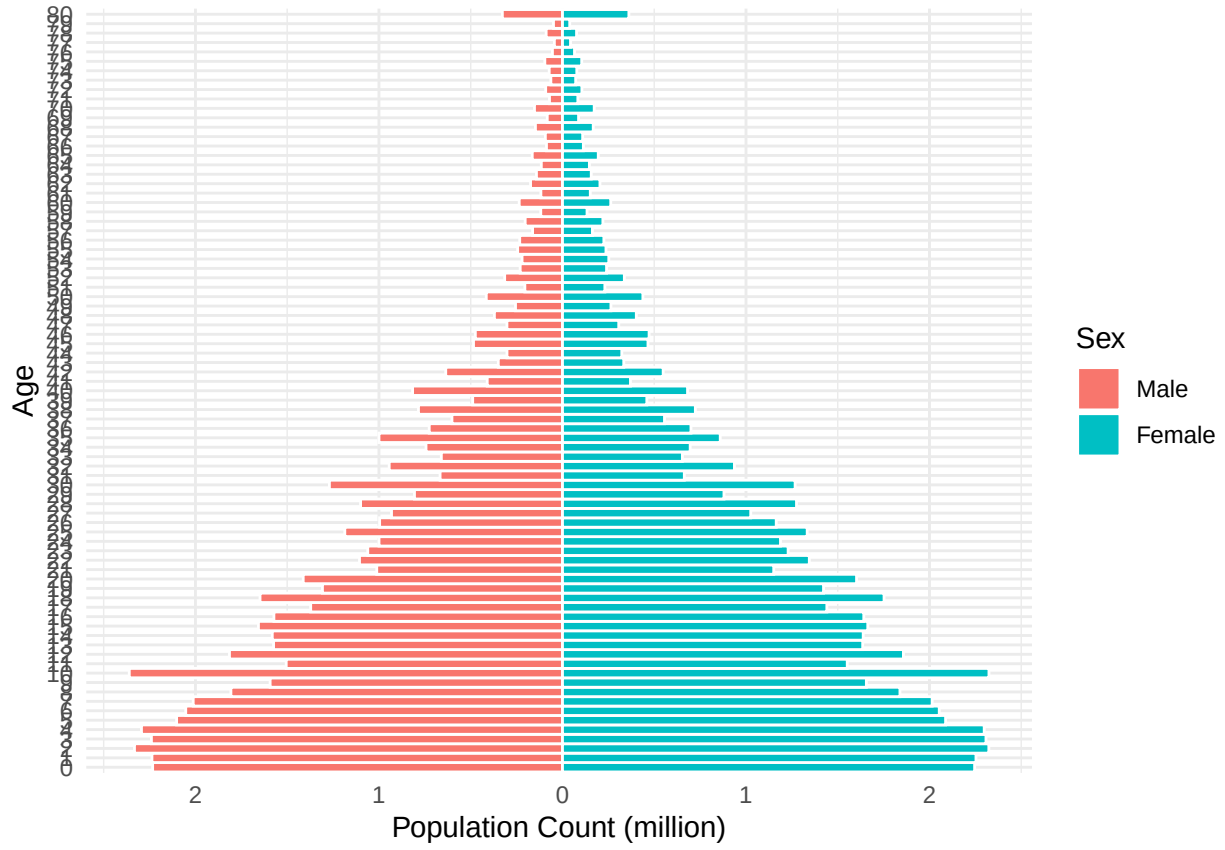



```
demographics$P5_AGE_P |> hist(xlab = "Age", ylab = "Frequency")
```




```
# Age distribution by Sex
demographics |>
  group_by(P4_SEX_P, P5_AGE_P) |>
  summarise(pop_count = n()/10000) |>
  mutate(
    pop_count = case_when(
      P4_SEX_P == 1 ~ -pop_count,
      P4_SEX_P == 2 ~ pop_count
    ),
    sex_label = factor(P4_SEX_P, levels = c(1, 2), labels = c("Male", "Female")),
    age_factor = factor(P5_AGE_P, levels = sort(unique(P5_AGE_P)))
  ) |>
  ggplot2::ggplot(aes(x = pop_count, y = age_factor, fill = sex_label)) +
  ggplot2::geom_bar(stat = "identity", width = 1, , color = "white") +
  ggplot2::scale_x_continuous(labels = abs) +
  ggplot2::labs(x = "Population Count (million)", y = "Age", fill = "Sex") +
  ggplot2::theme_minimal()
```

```
## `summarise()` has grouped output by 'P4_SEX_P'. You can override using the
## `.groups` argument.
```



Step2: Statistics by Constituency (Admin 4 Zambia)

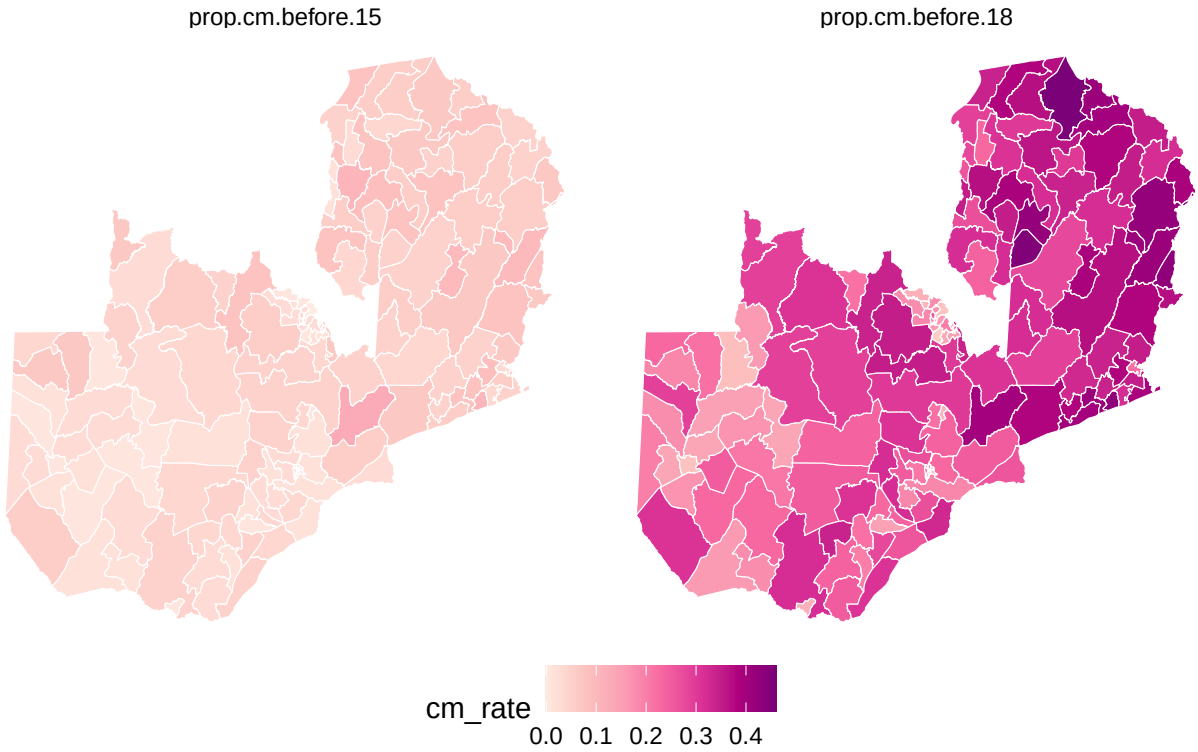
```
child_marriage <-
  demographics |>
  dplyr::filter(P5_AGE_P >= 20, P5_AGE_P <= 24, P4_SEX_P == 2) |>
  dplyr::group_by(CONST_P) |>
  dplyr::summarise(
    total.girls.20_24 = n(),
    married.before.15 = sum(P37_AGE_FIRST_MARRAIGE_P < 15, na.rm = T),
    married.before.18 = sum(P37_AGE_FIRST_MARRAIGE_P < 18, na.rm = T),
    prop.cm.before.15 = round(married.before.15 / total.girls.20_24, 2),
    prop.cm.before.18 = round(married.before.18 / total.girls.20_24, 2)
  ) |>
  dplyr::mutate(
    CONST_P_name = forcats::as_factor(CONST_P)
  ) |>
  dplyr::select(
    CONST_P_name,
    everything()
  )

# print table
child_marriage |>
  head() |>
  gt::gt()
```

Constituency	Constituency	total.girls.20_24	married.before.15	married.before.18	prop.cm.before
Chisamba	1	459	9	112	0
Katuba	2	376	13	101	0
Keembe	3	567	29	176	0
Bwacha	4	414	15	91	0
Kabwe Central	5	716	8	85	0
Kapiri Mposhi	6	1153	57	344	0

Visualization

```
sf_feature |>
  dplyr::select(NAME1_, geometry) |>
  dplyr::left_join(
    child_marriage,
    by = c("NAME1_" = "CONST_P_name")
  ) |>
  tidyr::pivot_longer(
    cols = c("prop.cm.before.15", "prop.cm.before.18"),
    names_to = "child_marriage",
    values_to = "cm_rate"
  ) |>
  ggplot2::ggplot() +
  ggplot2::geom_sf(aes(fill = cm_rate), color = "white") +
  ggplot2::scale_fill_distiller(palette = "RdPu", direction = 1) +
  ggplot2::theme_void() +
  ggplot2::theme(
    legend.position = "bottom"
  ) +
  ggplot2::facet_grid(~child_marriage)
```



SDG 8 : DECENT WORK AND ECONOMIC GROWTH

SDG TARGET 8.6 : By 2020, substantially reduce the proportion of youth not in employment, education or training

SDG INDICATOR 8.6.1: Proportion of youth (aged 15-24 years) not in education, employment or training NEET rate (%) = (Youth – Youth in employment – Youth not in employment but in education or training) / Youth *100

Input Variables Definition :

- *P32_ACTIVITY_LAST_12_MONTHS_P = Activity in the last 12 months*
- *P5_AGE_P = Age*

Output Variables Definition :

- *total.youth.15_24 = Total number of people aged 15 to 24 (Youth)*
- *total.neet = Number of youths not employed, not in education or training*
- *rate.need = proportion of youths not employed, not in education or training*

Step1: Explore Variables

Step2: Statistics by Constituency (Admin 4 Zambia) & Sex

CONST_P_name	Constituency	total.youth.15_24_Male	total.youth.15_24_Female	total.neet_M
Chisamba	1	1078	1081	1
Katuba	2	834	899	2
Keembe	3	1227	1307	2
Bwacha	4	903	973	2
Kabwe Central	5	1444	1666	3
Kapiri Mposhi	6	2571	2660	3

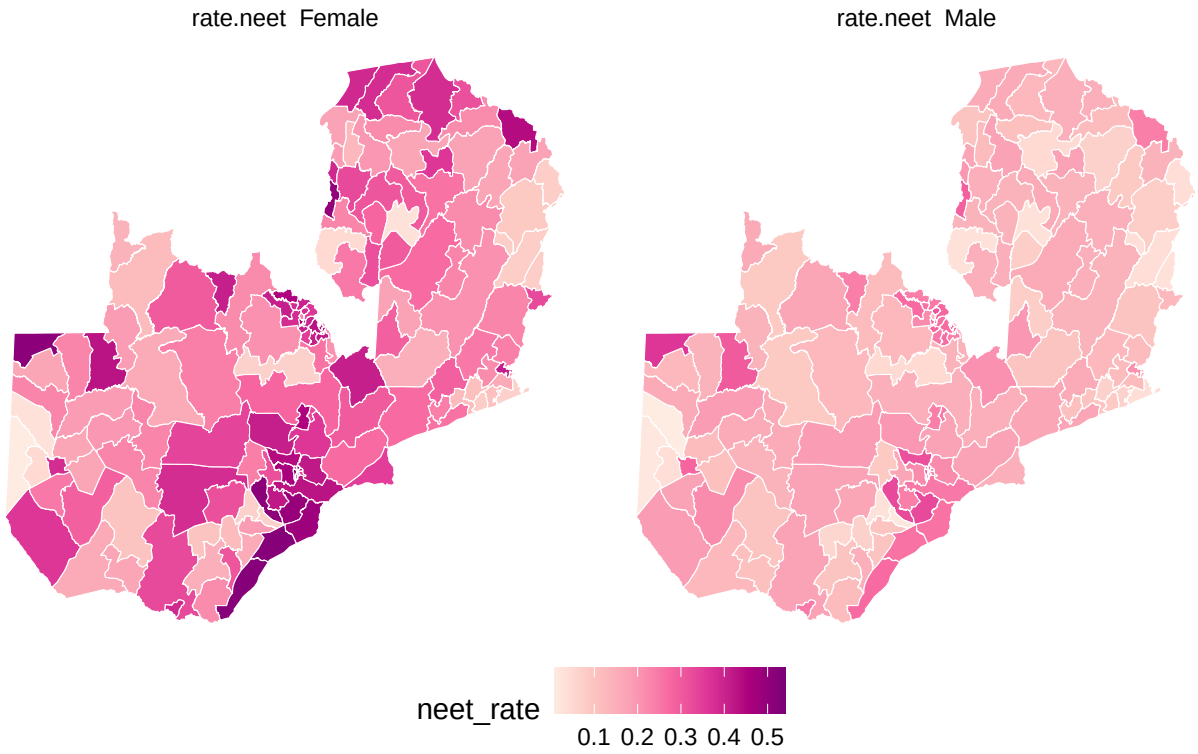
```
youth_umemplyment <-
  demographics |>
  dplyr::mutate(
    Employ = case_when(
      P32_ACTIVITY_LAST_12_MONTHS_P %in% 1:6 ~ 1,
      P32_ACTIVITY_LAST_12_MONTHS_P %in% 7:9 ~ 2,
      P32_ACTIVITY_LAST_12_MONTHS_P == 11 ~ 2,
      P32_ACTIVITY_LAST_12_MONTHS_P == 10 ~ 3
    )
  ) |>
  dplyr::filter(P5_AGE_P >= 15, P5_AGE_P <= 24) |>
  dplyr::group_by(CONST_P, P4_SEX_P) |>
  summarise(
    total.youth.15_24 = n(),
    total.neet = sum(Employ == 2, na.rm = TRUE),
    rate.neet = round(total.neet / total.youth.15_24, 2)
  ) |>
  dplyr::mutate(
    CONST_P_name = forcats::as_factor(CONST_P),
    P4_SEX_P = forcats::as_factor(P4_SEX_P)
  ) |>
  dplyr::select(
    CONST_P_name,
    everything()
  ) |>
  tidyr::pivot_wider(
    names_from =  P4_SEX_P,
    values_from = 4:6
  )
```

```
## `summarise()` has grouped output by 'CONST_P'. You can override using the
## `.groups` argument.
```

```
# print table
youth_umemplyment |>
  tibble::tibble() |>
  head() |>
  gt::gt()
```

Visualization

```r
sf_feature |>
  dplyr::select(NAME1_, geometry) |>
  dplyr::left_join(
    youth_umemplyment,
    by = c("NAME1_" = "CONST_P_name")
  ) |>
  tidyr::pivot_longer(
    cols = c("rate.neet_Male", "rate.neet_Female"),
    names_to = "neet",
    values_to = "neet_rate"
  ) |>
  ggplot2::ggplot() +
  ggplot2::geom_sf(aes(fill = neet_rate), color = "white") +
  ggplot2::scale_fill_distiller(palette = "RdPu", direction = 1) +
  ggplot2::theme_void() +
  ggplot2::theme(
    legend.position = "bottom"
  ) +
  ggplot2::facet_grid(~neet)
```



EXPORT RESULTS

1- Merge all results into one data frame

CONST_P_name	Constituency	total.girls.20_24	married.before.15	married.before.18	prop.cm.be
Chisamba	1	459	9	112	
Katuba	2	376	13	101	
Keembe	3	567	29	176	
Bwacha	4	414	15	91	
Kabwe Central	5	716	8	85	
Kapiri Mposhi	6	1153	57	344	

```
indicators_df <-
  child_marriage |>
  dplyr::left_join(
    adolescent_birth_rate,
    by = c("CONST_P_name", "CONST_P")
  ) |>
  dplyr::left_join(
    youth_umemplyment,
    by = c("CONST_P_name", "CONST_P")
  )
# print table
indicators_df |>
  head() |>
  gt::gt()
```

2- Export the results into a CSV file for further analysis

```
readr::write_csv(x = indicators_df, file = "output/indicators.csv")
```

Optional : Application of DEGURBA to SDG Indicators

In this section, we are Going to classify our admin data by DEGURBA classfication generated from the application of the DEGURBA methodology.

Step 1: Join the Constituency Shapes and The SDG Indicators

```
sdg_degurba <-
  sf_feature |>
  dplyr::select(NAME1_, DEGURBA_L1, DEGURBA_L2) |>
  dplyr::left_join(
    indicators_df,
    by = c("NAME1_" = "CONST_P_name")
  ) |>
  tibble::as.tibble() |>
  dplyr::select(-geometry)
```

```
## Warning: `as.tibble()` was deprecated in tibble 2.0.0.
## i Please use `as_tibble()` instead.
## i The signature and semantics have changed, see `?as_tibble`.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

NAME1_	DEGURBA_L1	DEGURBA_L2	Constituency	total.girls.20_24	married.before.15	r
Itezhi Tezhi	1	12	130	310	13	
Mulobezi	1	12	148	135	4	
Sesheke	1	12	150	213	4	
Sinjembela	1	12	147	454	26	
Katombola	1	12	121	463	21	
Mwandi	1	12	149	128	2	

```
# print table
sdg_degurba |>
  head() |>
  gt::gt()
```

Step 2 : SDG dissagregation by DEGURBA

```
sdg_degurba_diss <-
  sdg_degurba |>
  tidyr::pivot_longer(
    cols = c("DEGURBA_L1", "DEGURBA_L2"),
    names_to = "degurba_level",
    values_to = "degurba_class"
  ) |>
  dplyr::group_by(degurba_level, degurba_class) |>
  dplyr::summarise(
    total.girls.20_24 = sum(total.girls.20_24),
    married.before.15   = sum(married.before.15),
    married.before.18 = sum(married.before.18),
    prop.cm.before.15   = round((married.before.15 / total.girls.20_24), 2),
    prop.cm.before.18 = round((married.before.18 / total.girls.20_24), 2),
    total.ado.10_14 = sum(total.ado.10_14),
    total.ado.15_19 = sum(total.ado.15_19),
    total.ado.birth.10_14   = sum(total.ado.birth.10_14),
    total.ado.birth.15_19   = sum(total.ado.birth.15_19),
    abr.10_14   = round((total.ado.birth.10_14/total.ado.10_14)*1000, 2),
    abr.15_19   = round((total.ado.birth.15_19/total.ado.15_19)*1000, 2),
    total.youth.15_24_Male = sum(total.youth.15_24_Male),
    total.youth.15_24_Female = sum(total.youth.15_24_Female),
    total.neet_Male = sum(total.neet_Male),
    total.neet_Female   = sum(total.neet_Female),
    rate.neet_Male = round((total.neet_Male / total.youth.15_24_Male), 2),
    rate.neet_Female = round((total.youth.15_24_Female / total.neet_Female), 2)
  )
```

```
## `summarise()` has grouped output by 'degurba_level'. You can override using the
## `.groups` argument.
```

```
sdg_degurba_diss <-
  sdg_degurba_diss |>
  dplyr::mutate(
    degurba_class = factor(degurba_class,
```

```
                            levels = c(1,2,3,11,12,13,21,23,30),
                            labels = c("Rural Area",
                                       "Town or Semi-dense Area",
                                       "City",
                                       "Very Disperded Rural Area",
                                       "Disperse Rural Area",
                                       "Village",
                                       "Suburban or Peri-urban Area",
                                       "Dense Town",
                                       "city"
                                       )
                            )
    )
```

Step 3: Visualization

```
sdg_degurba_diss |>
  gt::gt() |>
  gt::tab_header(title = "SDG INDICATOR BY DEGURBA",
                 subtitle = "Zambia 2010 Pop & House Census") |>
  gt::data_color(
    columns = c("prop.cm.before.15",
                "prop.cm.before.18"
                ),
    method = "numeric",
    palette = "OrRd",
    domain = c(min(sdg_degurba_diss$prop.cm.before.15),
               max(sdg_degurba_diss$prop.cm.before.18))
  ) |>
  gt::data_color(
    columns = c("abr.10_14",
                "abr.15_19"
                ),
    method = "numeric",
    palette = "OrRd",
    domain = c(min(sdg_degurba_diss$abr.10_14),
               max(sdg_degurba_diss$abr.15_19))
  ) |>
  gt::data_color(
    columns = c("rate.neet_Male",
                "rate.neet_Female"
                ),
    method = "numeric",
    palette = "OrRd",
    domain = c(min(sdg_degurba_diss$rate.neet_Male),
               max(sdg_degurba_diss$rate.neet_Female))
  )
```

1

[1] *UNFPA - SDG DISSAGREGATION BY DEGURBA | by Derrick DEMEVENG | Tanzania DEGURBA Workshop - May*

SDG INDICATOR BY DEGURBA

Zambia 2010 Pop & House Census

degurba_class	total.girls.20_24	married.before.15	married.before.18	prop.cm.before.
DEGURBA_L1				
Rural Area	37060	1822	11142	0.
Town or Semi-dense Area	6963	284	1628	0.
City	21146	423	3272	0.
DEGURBA_L2				
Very Disperded Rural Area	265	27	102	0.
Disperse Rural Area	29834	1412	8907	0.
Village	6961	383	2133	0.
Suburban or Peri-urban Area	5663	254	1428	0.
Dense Town	1300	30	200	0.
city	21146	423	3272	0.