**Part 1: Data Management & Programming**

Observations

* Noticed that in “malaria\_deaths.csv”, 32 countries have no “Code “ column, eg, Andean Latin America, Australasia, Caribbean=> filled them as “NA” labels- following “malaria\_deaths\_age.csv”
* Round death#; instead of having decimals; for readability and easier comparison;

--

* In “malaria\_deaths\_age.csv”, there are some “Age group” formatting issue

A screenshot of a computer

Description automatically generated

--

* In “malaria\_inc.csv”- “Incidence of malaria (per 1,000 population at risk) (per 1,000 population at risk)” column provided easier comparison amongst each entity with regards to malaria risks;

--

A screenshot of a computer

Description automatically generated

--

Interpretation for columns

“malaria\_deaths\_age.csv”

* entity
* code
* year
* age\_group
* deaths
  + presented for respective entity, year& age\_group

--

Other 2 datasets are rather similar; with a different column;

“malaria\_inc.csv”

* Entity
* Code
* Year
  + 5-yearly from year 2000 to 2015
* Incidence of malaria (per 1,000 population at risk) (per 1,000 population at risk)
  + data is reporting the rate at which new cases of malaria are occurring per 1,000 people who are considered at risk of contracting malaria within a specific population and time period.

--

“malaria\_deaths.csv”,

* Entity
* Code
* Year
  + From year 1990 to 2016
* Deaths - Malaria - Sex: Both - Age: Age-standardized (Rate) (per 100,000 people)
  + describing statistics about malaria-related deaths that have been adjusted for age differences and are reported as a rate per 100,000 people for both males and females. This type of data presentation helps in understanding the impact of malaria on a population while accounting for age variations.

--

* the death numbers are presented in decimals;
  + my assumption is that it might be to provide more precise information, especially when dealing with small numbers or when trying to differentiate between rates in different regions or groups; will be rounding them to nearest whole number for for readability and provide easier comparison across entities;

--

* “malaria\_deaths.csv”, and “malaria\_deaths\_age.csv” have 228 entities;
* whereas, “malaria\_inc.csv” has only 127 entities;
* Although,“malaria\_deaths.csv”, and “malaria\_deaths\_age.csv” have the same entity columns, I have decided not to combine them as this might cause confusion since “malaria\_deaths\_age.csv” data is further segregated by individual age group.

--