CL_Child_Mortality_Rates_02

Group K

2025-09-19

Disdplay Dataset content

```
## # A tibble: 3 x 29
##   ISO3   DataId Indicator Value Precision DHS_CountryCode CountryName SurveyYear
##   <chr>  <chr>  <chr>     <chr> <chr>     <chr>           <chr>       <chr>
## 1 #coun~ #meta~ #indicat~ #ind~ #indicat~ <NA>            #country+n~ #date+year
## 2 ZAF    85995  Neonatal~ 20    0         ZA              South Afri~ 1998
## 3 ZAF    794581 Postneon~ 26    0         ZA              South Afri~ 1998
## # i 21 more variables: SurveyId <chr>, IndicatorId <chr>, IndicatorOrder <dbl>,
## #   IndicatorType <chr>, CharacteristicId <dbl>, CharacteristicOrder <dbl>,
## #   CharacteristicCategory <chr>, CharacteristicLabel <chr>,
## #   ByVariableId <chr>, ByVariableLabel <chr>, IsTotal <dbl>,
## #   IsPreferred <dbl>, SDRID <chr>, RegionId <lgl>, SurveyYearLabel <dbl>,
## #   SurveyType <chr>, DenominatorWeighted <dbl>, DenominatorUnweighted <dbl>,
## #   CILow <dbl>, CIHigh <dbl>, LevelRank <lgl>
```

Remove the first row(meta data)

```
cmr_df <- cmr_df[-1, ]
```

Inspect Duplicated rows

```
## # A tibble: 2 x 29
## # Groups:   Indicator, SurveyYear, CharacteristicId, Value [1]
##   ISO3  DataId Indicator  Value Precision DHS_CountryCode CountryName SurveyYear
##   <chr> <chr>  <chr>      <chr> <chr>     <chr>           <chr>       <chr>
## 1 ZAF   799142 Child mor~ 15    0         ZA              South Afri~ 1998
## 2 ZAF   66518  Child mor~ 15    0         ZA              South Afri~ 1998
## # i 21 more variables: SurveyId <chr>, IndicatorId <chr>, IndicatorOrder <dbl>,
## #   IndicatorType <chr>, CharacteristicId <dbl>, CharacteristicOrder <dbl>,
## #   CharacteristicCategory <chr>, CharacteristicLabel <chr>,
## #   ByVariableId <chr>, ByVariableLabel <chr>, IsTotal <dbl>,
## #   IsPreferred <dbl>, SDRID <chr>, RegionId <lgl>, SurveyYearLabel <dbl>,
## #   SurveyType <chr>, DenominatorWeighted <dbl>, DenominatorUnweighted <dbl>,
## #   CILow <dbl>, CIHigh <dbl>, LevelRank <lgl>
```

Removed newer duplicated row

```r
cmr_df <- cmr_df %>% filter(DataId != "799142")
```

Convert Data Types

```r
cmr_df <- cmr_df %>%
  mutate(
        Value = as.numeric(Value),
    Precision = as.numeric(Precision),
    SurveyYear = as.integer(SurveyYear),
    IndicatorOrder = as.integer(IndicatorOrder),
    CharacteristicId = as.integer(CharacteristicId),
    CharacteristicOrder = as.integer(CharacteristicOrder),
    IsTotal = as.logical(as.integer(IsTotal)),
    IsPreferred = as.logical(as.integer(IsPreferred)),
    SurveyYearLabel = as.integer(SurveyYearLabel),
    DenominatorWeighted = as.numeric(DenominatorWeighted),
    DenominatorUnweighted = as.numeric(DenominatorUnweighted),
  )
```

Summary table: column name, number of unique values, sample of unique values

```
## # A tibble: 29 x 3
##    column          n_unique sample_values
##    <chr>              <int> <chr>
##  1 ISO3                   1 ZAF
##  2 DataId                39 85995, 794581, 785930
##  3 Indicator             15 Neonatal mortality rate (5 year periods), Postneona~
##  4 Value                 27 20, 26, 45
##  5 Precision              1 0
##  6 DHS_CountryCode        1 ZA
##  7 CountryName            1 South Africa
##  8 SurveyYear             2 1998, 2016
##  9 SurveyId               2 ZA1998DHS, ZA2016DHS
## 10 IndicatorId           15 CM_ECMT_C_NNR, CM_ECMT_C_PNR, CM_ECMT_C_IMR
## # i 19 more rows
```

Drop the countries only one unqiue value: reason, there is no useful information - country is also always za