

CL_Access_To_Healthcare_01

Group K

2025-09-19

#Loading Libraries

Display Dataset content

```
## # A tibble: 5 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    751751 Antenata~ 28.5  1        ZA           South Afri~ 1998
## 3 ZAF    567476 Antenata~ 30    1        ZA           South Afri~ 1998
## 4 ZAF    205488 Antenata~ 27.3  1        ZA           South Afri~ 1998
## 5 ZAF    751748 Antenata~ 66.6  1        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 <lgl>, CIHigh <lgl>, LevelRank <lgl>
```

Remove the first row(meta data)

```
acc_df <- acc_df[-1, ]
```

Check data types

```
##           Column paste0.$apply.$acc_df$.typeof..
## 1           ISO3                character
## 2          DataId                character
## 3          Indicator              character
## 4            Value                character
## 5          Precision              character
## 6   DHS_CountryCode              character
## 7          CountryName            character
## 8          SurveyYear              character
## 9            SurveyId              character
## 10         IndicatorId              character
## 11        IndicatorOrder              double
```

## 12	IndicatorType	character
## 13	CharacteristicId	double
## 14	CharacteristicOrder	double
## 15	CharacteristicCategory	character
## 16	CharacteristicLabel	character
## 17	ByVariableId	character
## 18	ByVariableLabel	character
## 19	IsTotal	double
## 20	IsPreferred	double
## 21	SDRID	character
## 22	RegionId	logical
## 23	SurveyYearLabel	double
## 24	SurveyType	character
## 25	DenominatorWeighted	double
## 26	DenominatorUnweighted	double
## 27	CILow	logical
## 28	CIHigh	logical
## 29	LevelRank	logical

Convert Data Types

```
acc_df <- acc_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),
  )
```

Check for unique values

```
## # A tibble: 29 x 3
##   column      n_unique sample_values
##   <chr>      <int> <chr>
## 1 ISO3          1 ZAF
## 2 DataId       239 751751, 567476, 205488
## 3 Indicator     68 Antenatal care provider: Doctor, Antenatal care pro-
## 4 Value        190 28.5, 30, 27.3
## 5 Precision      2 1, 0
## 6 DHS_CountryCode 1 ZA
## 7 CountryName    1 South Africa
## 8 SurveyYear      2 1998, 2016
## 9 SurveyId        2 ZA1998DHS, ZA2016DHS
## 10 IndicatorId    70 RH_ANCP_W_DOC, RH_ANCP_W_NRS, RH_ANCP_W_TBA
```

```
## # i 19 more rows
```

#Drop the countries only one unique value: reason, there is no useful information - county is also always za

#Assumed pattern, the missing values can be filled with the previous non missing value in the opposite attribute



