CL_Literacy_03

Group K

2025-09-19

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    563770 Women wi~ 11.8  1         ZA              South Afri~ 2016
## 3 ZAF    563771 Women wh~ 76.2  1         ZA              South Afri~ 2016
## 4 ZAF    563772 Women wh~ 8.2   1         ZA              South Afri~ 2016
## 5 ZAF    563773 Women wh~ 3.5   1         ZA              South Afri~ 2016
## # 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)

Inspect Duplicated rows

```
## # A tibble: 0 x 29
## # Groups:   Indicator, SurveyYear, CharacteristicId, Value [0]
## # i 29 variables: ISO3 <chr>, DataId <chr>, Indicator <chr>, Value <chr>,
## #   Precision <chr>, DHS_CountryCode <chr>, CountryName <chr>,
## #   SurveyYear <chr>, 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>, ...
```

Convert Data Types

```
lit_df <- lit_df %>%
  mutate(
        Value = as.numeric(Value),
    Precision = as.numeric(Precision),
```

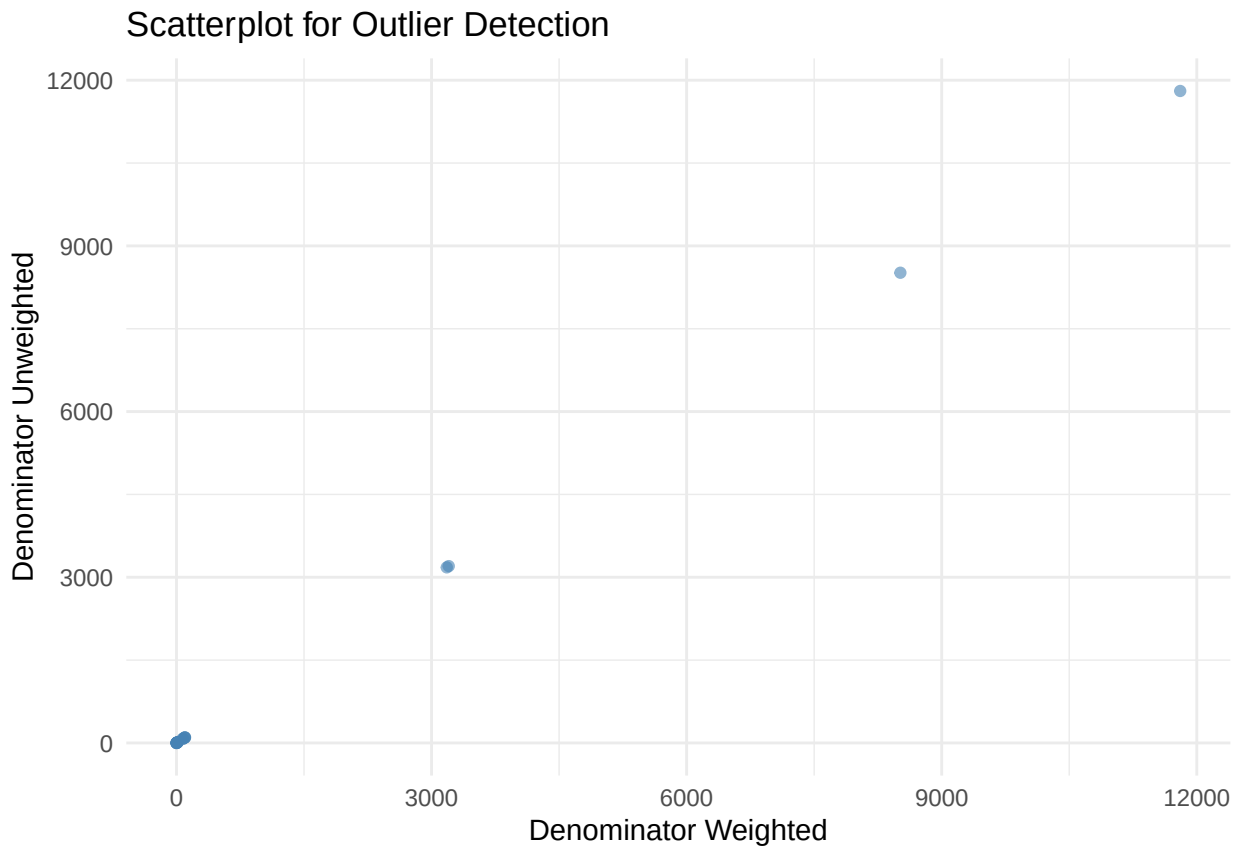
```
    SurveyYear = as.integer(SurveyYear),
    IndicatorOrder = as.integer(IndicatorOrder),
    CharacteristicId = as.integer(CharacteristicId),
    CharacteristicOrder = as.integer(CharacteristicOrder),
    IsTotal = as.logical(as.integer(IsTotal)),
    IsPreferred = as.logical(as.integer(IsPreferred)),
    SurveyYearLabel = as.integer(SurveyYearLabel),
    DenominatorWeighted = as.numeric(DenominatorWeighted),
    DenominatorUnweighted = as.numeric(DenominatorUnweighted),
  )
```

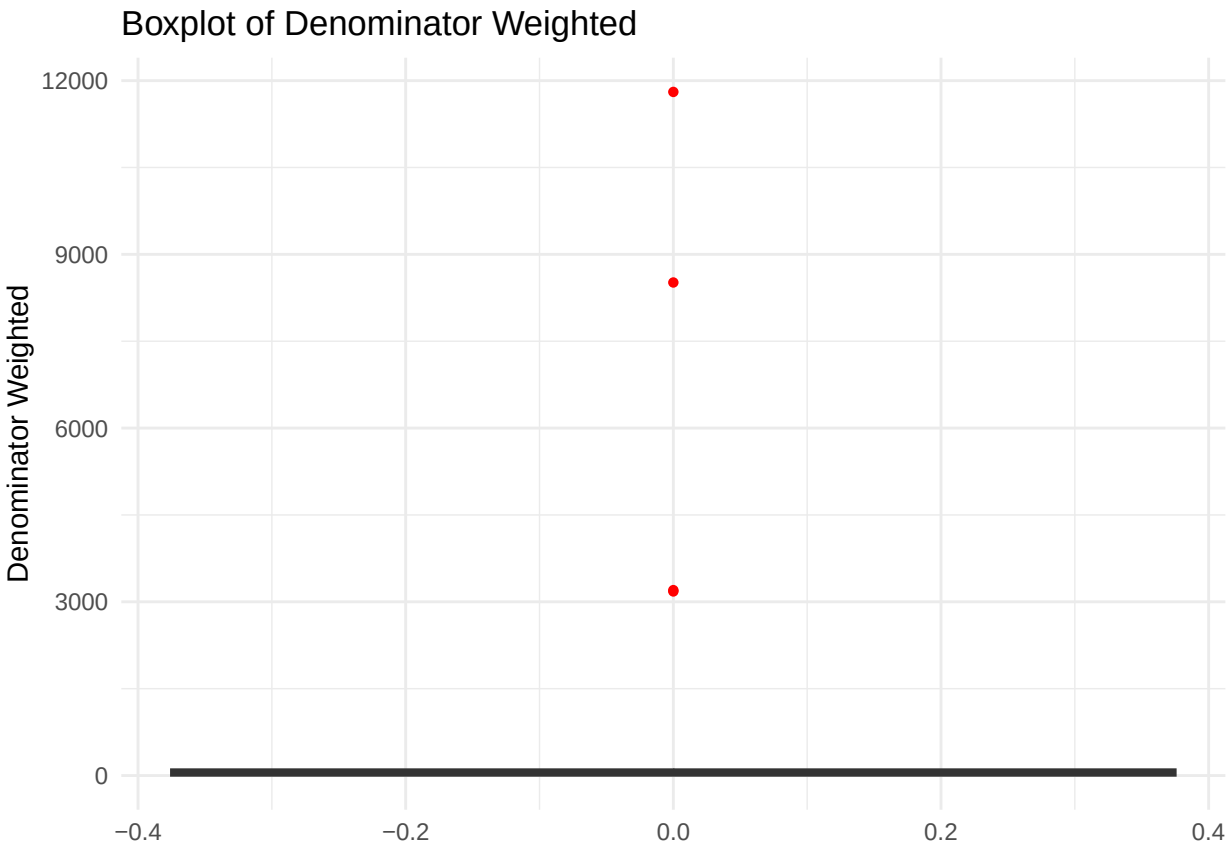
Drop the countries only one unqiue value: reason, there is no useful information - county is also always za

Missing Value Handling

```
lit_df <- lit_df %>%
  fill(DenominatorWeighted, DenominatorUnweighted, .direction = "downup")

lit_df[
       c("DenominatorWeighted", "DenominatorUnweighted")]
```




Boxplot of Denominator Weighted
Denominator Weighted
12000
9000
6000
3000
0
−0.4
−0.2
0.0
0.2
0.4