CL_Symptoms_Of_Respiratory_Infection_10

Group K

2025-09-19

Display Dataset content

#check for unique values

```
## # A tibble: 29 x 3
##    column          n_unique sample_values
##    <chr>              <int> <chr>
##  1 ISO3                   1 ZAF
##  2 DataId                26 598577, 397915, 598578
##  3 Indicator              7 Children with symptoms of ARI, Number of children b~
##  4 Value                 26 21.9, 19.3, 2912
##  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            7 CH_ARIS_C_ARI, CH_ARIS_C_NUM, CH_ARIS_C_UNW
## # i 19 more rows
```

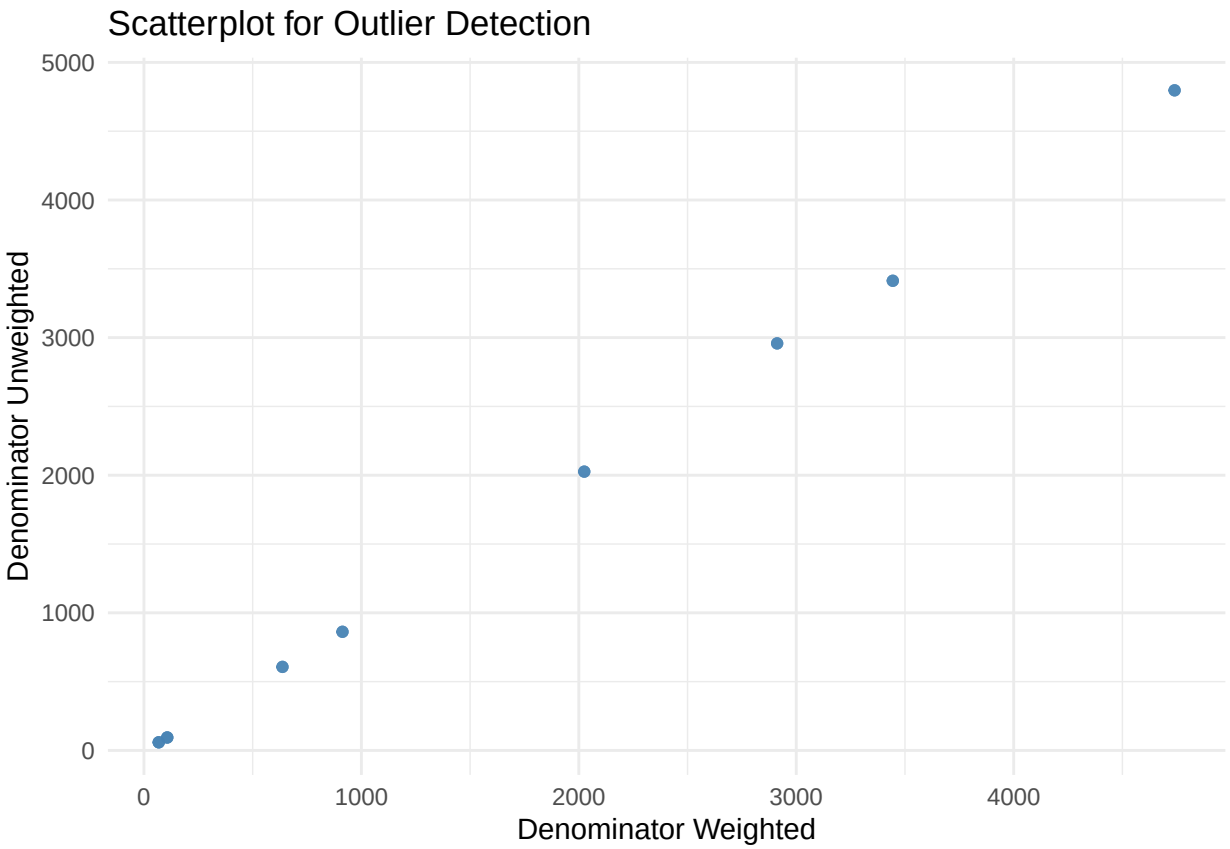
Drop the countries only one unqiue value: reason, there is no useful information - county is also always za

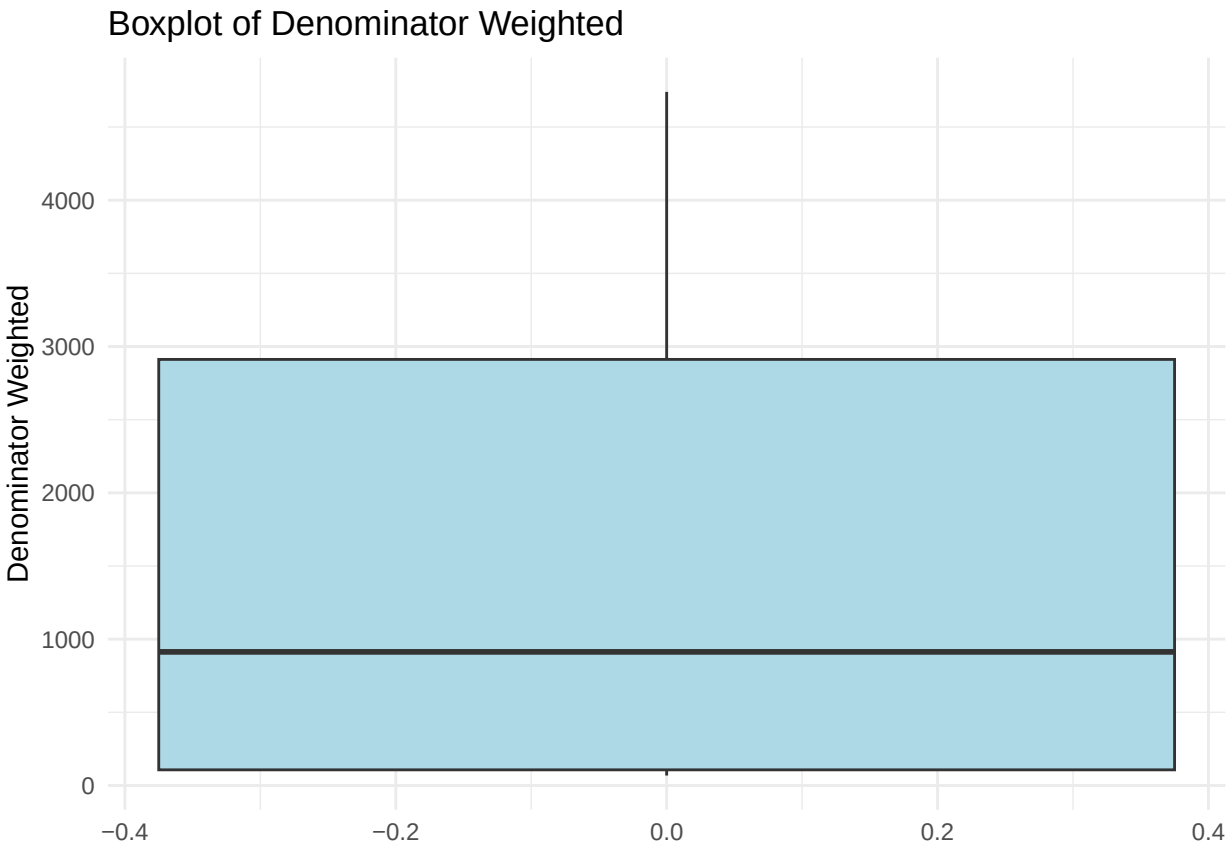
Assumed pattern, the missing values can be filled with the non missing value in the opposite attribute

```
ari_df <- ari_df %>%
   mutate(
    # 4740 <-> 4797
    DenominatorUnweighted = if_else(
      is.na(DenominatorUnweighted) & DenominatorWeighted == 4740,
      4797,
      DenominatorUnweighted
    ),
    DenominatorWeighted = if_else(
      is.na(DenominatorWeighted) & DenominatorUnweighted == 4797,
      4740,
      DenominatorWeighted
    ),
```

```
  # 2912 <-> 2958
  DenominatorUnweighted = if_else(
    is.na(DenominatorUnweighted) & DenominatorWeighted == 2912,
    2958,
    DenominatorUnweighted
  ),
  DenominatorWeighted = if_else(
    is.na(DenominatorWeighted) & DenominatorUnweighted == 2958,
    2912,
    DenominatorWeighted
  ),

  # 2025 <-> 2026
  DenominatorUnweighted = if_else(
    is.na(DenominatorUnweighted) & DenominatorWeighted == 2025,
    2026,
    DenominatorUnweighted
  ),
  DenominatorWeighted = if_else(
    is.na(DenominatorWeighted) & DenominatorUnweighted == 2026,
    2025,
    DenominatorWeighted
  ),

  # 3444 <-> 3413
  DenominatorUnweighted = if_else(
    is.na(DenominatorUnweighted) & DenominatorWeighted == 3444,
    3413,
    DenominatorUnweighted
  ),
  DenominatorWeighted = if_else(
    is.na(DenominatorWeighted) & DenominatorUnweighted == 3413,
    3444,
    DenominatorWeighted
  ),

  # 68 <-> 59
  DenominatorUnweighted = if_else(
    is.na(DenominatorUnweighted) & DenominatorWeighted == 68,
    59,
    DenominatorUnweighted
  ),
  DenominatorWeighted = if_else(
    is.na(DenominatorWeighted) & DenominatorUnweighted == 59,
    68,
    DenominatorWeighted
  ),

  # 107 <-> 94
  DenominatorUnweighted = if_else(
    is.na(DenominatorUnweighted) & DenominatorWeighted == 107,
    94,
    DenominatorUnweighted
```

```
    ),
    DenominatorWeighted = if_else(
      is.na(DenominatorWeighted) & DenominatorUnweighted == 94,
      107,
      DenominatorWeighted
    ),

    # 637 <-> 607
    DenominatorUnweighted = if_else(
      is.na(DenominatorUnweighted) & DenominatorWeighted == 637,
      607,
      DenominatorUnweighted
    ),
    DenominatorWeighted = if_else(
      is.na(DenominatorWeighted) & DenominatorUnweighted == 607,
      637,
      DenominatorWeighted
    ),

    # 913 <-> 862
    DenominatorUnweighted = if_else(
      is.na(DenominatorUnweighted) & DenominatorWeighted == 913,
      862,
      DenominatorUnweighted
    ),
    DenominatorWeighted = if_else(
      is.na(DenominatorWeighted) & DenominatorUnweighted == 862,
      913,
      DenominatorWeighted
    )
  )



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


Scatterplot for Outlier Detection
Denominator Unweighted
Denominator Weighted
5000
4000
3000
2000
1000
0
0
1000
2000
3000
4000


Boxplot of Denominator Weighted
Denominator Weighted
4000
3000
2000
1000
0
−0.4
−0.2
0.0
0.2
0.4