

# CL\_Immunization\_06

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    330965 BCG vacc~ 96.8 1        ZA            South Afri~ 1998
## 3 ZAF    139796 BCG vacc~ 94.9 1        ZA            South Afri~ 1998
## 4 ZAF    330966 DPT 1 va~ 93.3 1        ZA            South Afri~ 1998
## 5 ZAF    139797 DPT 1 va~ 93.1 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)

```
imm_df <- imm_df[-1, ]
```

## Convert Data Types

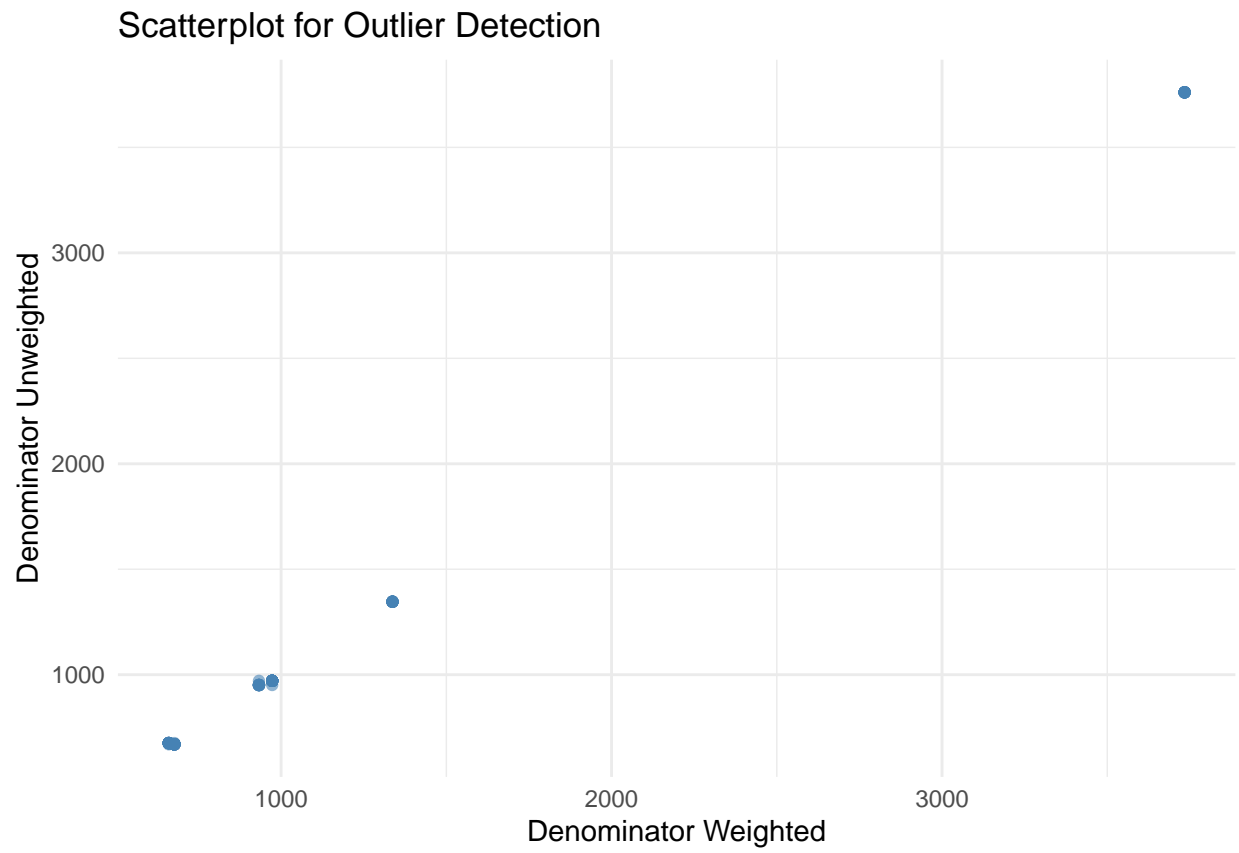
#check for unique values

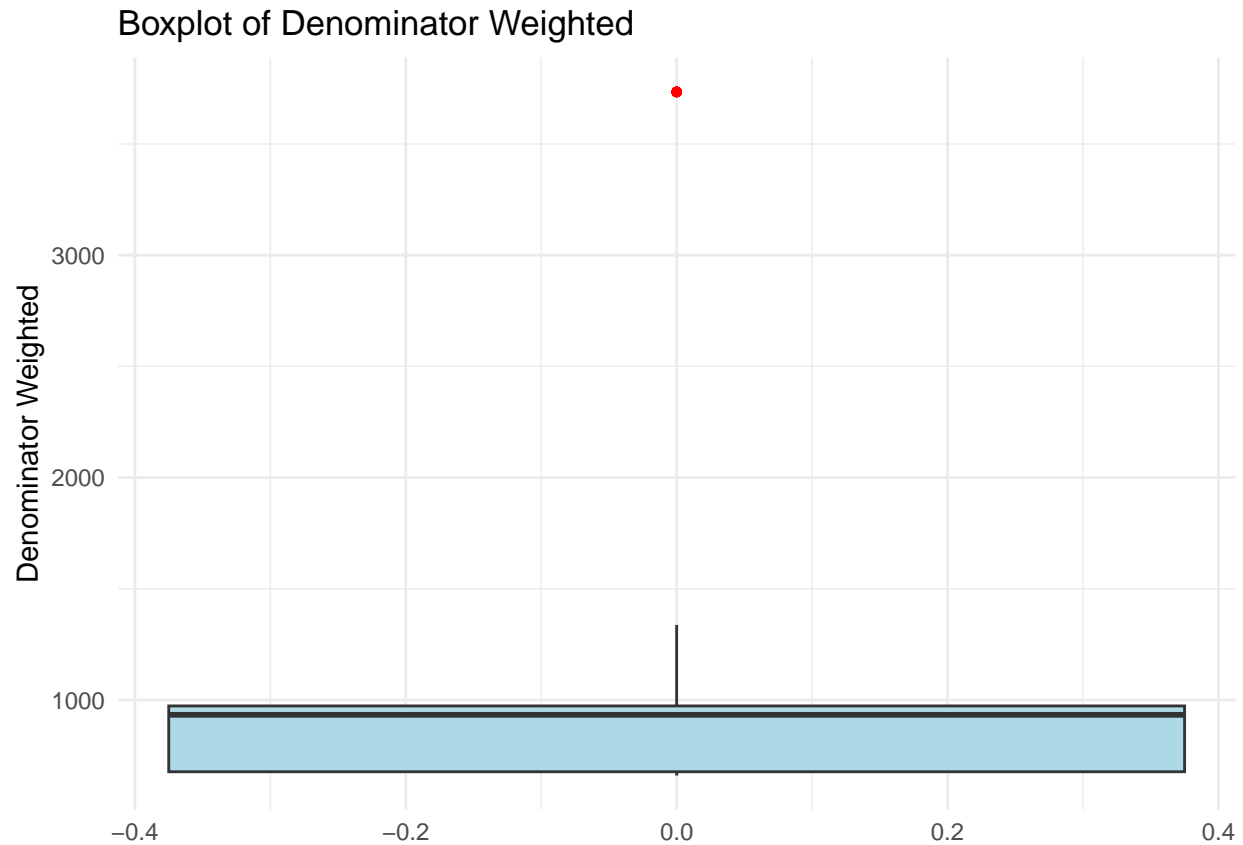
#Drop the countries only one unique value: reason, there is no useful information - county is also always za

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

```
imm_df <- imm_df %>%
  fill(DenominatorWeighted, DenominatorUnweighted, .direction = "down")

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





## Outlier Handling

```
# Calculate IQR boundaries
Q1_w <- quantile(imm_df$DenominatorWeighted, 0.25, na.rm = TRUE)
Q3_w <- quantile(imm_df$DenominatorWeighted, 0.75, na.rm = TRUE)
IQR_w <- Q3_w - Q1_w
lower_w <- Q1_w - 1.5 * IQR_w
upper_w <- Q3_w + 1.5 * IQR_w

Q1_uw <- quantile(imm_df$DenominatorUnweighted, 0.25, na.rm = TRUE)
Q3_uw <- quantile(imm_df$DenominatorUnweighted, 0.75, na.rm = TRUE)
IQR_uw <- Q3_uw - Q1_uw
lower_uw <- Q1_uw - 1.5 * IQR_uw
upper_uw <- Q3_uw + 1.5 * IQR_uw

# Cap values to the IQR limits
imm_df <- imm_df %>%
  mutate(
    DenominatorWeighted = pmin(pmax(DenominatorWeighted, lower_w), upper_w),
    DenominatorUnweighted = pmin(pmax(DenominatorUnweighted, lower_uw), upper_uw)
  )
```