

# EDA\_toilet\_facilities\_09

Group K

2025-09-09

## 1. Loading the dataset

```
## # A tibble: 5 x 29
##   IS03    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    795762 Househol~ 50.1  1        ZA           South Afri~ 1998
## 3 ZAF    795768 Househol~ 38.3  1        ZA           South Afri~ 1998
## 4 ZAF    795760 Househol~ 31.2  1        ZA           South Afri~ 1998
## 5 ZAF    795764 Househol~ 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>
```

## 2. Data Overview

### Glimpse and summary statistics

```
##      IS03              DataId              Indicator              Value
## Length:47          Length:47          Length:47          Length:47
## Class :character    Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
##
## Precision          DHS_CountryCode      CountryName          SurveyYear
## Length:47          Length:47          Length:47          Length:47
## Class :character    Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
##
## SurveyId            IndicatorId            IndicatorOrder        IndicatorType
```

```

## Length:47          Length:47          Min.   :250261010   Length:47
## Class :character   Class :character   1st Qu.:250261182   Class :character
## Mode  :character   Mode  :character   Median :250261630   Mode  :character
##                                     Mean  :250261648
##                                     3rd Qu.:250262178
##                                     Max.   :250262250
##                                     NA's    :1
## CharacteristicId CharacteristicOrder CharacteristicCategory
## Min.   :1000      Min.   :0           Length:47
## 1st Qu.:1000      1st Qu.:0           Class :character
## Median :1000      Median :0           Mode  :character
## Mean   :1000      Mean   :0
## 3rd Qu.:1000      3rd Qu.:0
## Max.   :1000      Max.   :0
## NA's    :1        NA's    :1
## CharacteristicLabel ByVariableId      ByVariableLabel      IsTotal
## Length:47          Length:47          Length:47            Min.   :1
## Class :character    Class :character    Class :character     1st Qu.:1
## Mode  :character    Mode  :character    Mode  :character     Median :1
##                                     Mean   :1
##                                     3rd Qu.:1
##                                     Max.   :1
##                                     NA's    :1
## IsPreferred      SDRID          RegionId      SurveyYearLabel
## Min.   :1        Length:47      Mode:logical   Min.   :1998
## 1st Qu.:1        Class :character   NA's:47        1st Qu.:1998
## Median :1        Mode  :character           Median :2016
## Mean   :1                                     Mean  :2009
## 3rd Qu.:1                                     3rd Qu.:2016
## Max.   :1                                     Max.   :2016
## NA's    :1                                     NA's    :1
## SurveyType      DenominatorWeighted DenominatorUnweighted CILow
## Length:47       Min.   :11083      Min.   :11083      Mode:logical
## Class :character 1st Qu.:11083      1st Qu.:11083      NA's:47
## Mode  :character Median :24726      Median :25086
##                                     Mean  :27185      Mean  :27495
##                                     3rd Qu.:37205    3rd Qu.:37925
##                                     Max.   :52007     Max.   :52465
##                                     NA's    :5        NA's    :5
## CIHigh          LevelRank
## Mode:logical     Mode:logical
## NA's:47          NA's:47
##
##
##
##

```

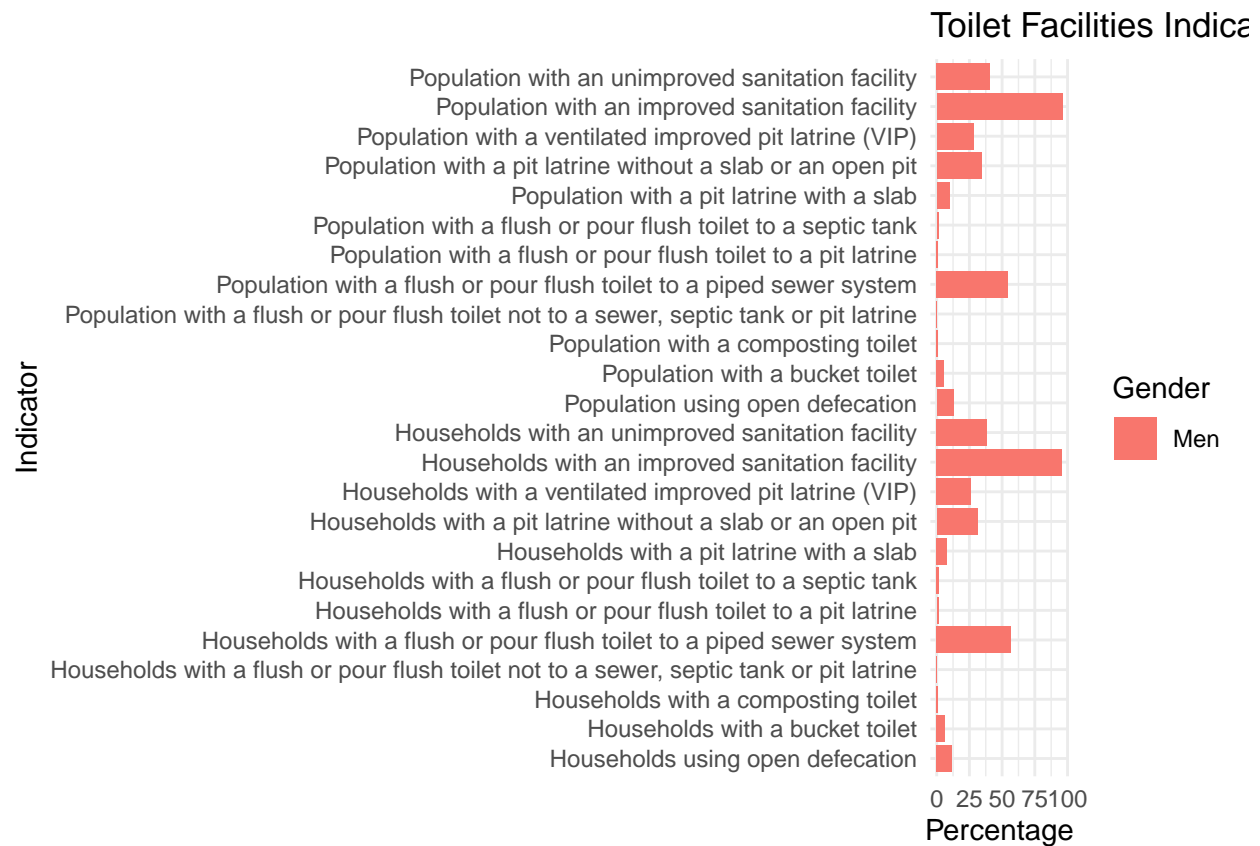
### 3. Data Quality

#### Missing values per column

##	IS03	DataId	Indicator
##	0	0	0
##	Value	Precision	DHS_CountryCode
##	0	0	1
##	CountryName	SurveyYear	SurveyId
##	0	0	0
##	IndicatorId	IndicatorOrder	IndicatorType
##	0	1	1
##	CharacteristicId	CharacteristicOrder	CharacteristicCategory
##	1	1	1
##	CharacteristicLabel	ByVariableId	ByVariableLabel
##	1	0	46
##	IsTotal	IsPreferred	SDRID
##	1	1	1
##	RegionId	SurveyYearLabel	SurveyType
##	47	1	1
##	DenominatorWeighted	DenominatorUnweighted	CILow
##	5	5	47
##	CIHigh	LevelRank	
##	47	47	

Duplicates

## 4. Visualizations



## 5. Summary Tables

```
## Warning: There were 4 warnings in 'summarise()'.
## The first warning was:
## i In argument: 'across(...)'.
## Caused by warning in 'min()':
## ! no non-missing arguments to min; returning Inf
## i Run 'dplyr::last_dplyr_warnings()' to see the 3 remaining warnings.
```

Table 1: Summary Statistics for Key Numeric Variables in the toilet\_facilities Dataset

Variable	Statistic	Value
CIHigh	mean	NaN
CIHigh	sd	NA
CIHigh	min	Inf
CIHigh	max	-Inf
CILow	mean	NaN

Variable	Statistic	Value
CILow	sd	NA
CILow	min	Inf
CILow	max	-Inf
DenominatorUnweighted	mean	27495.238
DenominatorUnweighted	sd	16938.786
DenominatorUnweighted	min	11083.000
DenominatorUnweighted	max	52465.000
DenominatorWeighted	mean	27185.143
DenominatorWeighted	sd	16667.463
DenominatorWeighted	min	11083.000
DenominatorWeighted	max	52007.000
Value	mean	4942.057
Value	sd	13082.679
Value	min	0.400
Value	max	52465.000