

# 08\_Maternal Mortality\_eda

Group K

2025-09-08

## 1. Loading the dataset

```
## # A tibble: 6 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    91409 Female d~ 5.5    1        ZA           South Afri~ 1998
## 3 ZAF    91377 Number o~ 19     0        ZA           South Afri~ 1998
## 4 ZAF    768646 Years of~ 1227~ 0        ZA           South Afri~ 1998
## 5 ZAF    768647 Years of~ 1237~ 0        ZA           South Afri~ 1998
## 6 ZAF    535566 Pregnanc~ 0.15   2        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 <dbl>, CIHigh <dbl>, LevelRank <lgl>
```

## 2. Data Overview

### Glimpse and summary statistics

```
## Rows: 22
## Columns: 29
## $ ISO3          <chr> "#country+code", "ZAF", "ZAF", "ZAF", "ZAF", "Z~
## $ DataId        <chr> "#meta+id", "91409", "91377", "768646", "768647~
## $ Indicator     <chr> "#indicator+name", "Female deaths that are preg~
## $ Value         <chr> "#indicator+value+num", "5.5", "19", "122701", ~
## $ Precision     <chr> "#indicator+precision", "1", "0", "0", "0", "2"~
## $ DHS_CountryCode <chr> NA, "ZA", "ZA", "ZA", "ZA", "ZA", "ZA", "ZA", "Z~
## $ CountryName   <chr> "#country+name", "South Africa", "South Africa"~
## $ SurveyYear    <chr> "#date+year", "1998", "1998", "1998", "1998", "~
## $ SurveyId      <chr> "#survey+id", "ZA1998DHS", "ZA1998DHS", "ZA1998~
## $ IndicatorId   <chr> "#indicator+code", "MM_MMRT_W_FDP", "MM_MMRT_W_~
## $ IndicatorOrder <dbl> NA, 77003010, 77003020, 77003030, 77003040, 770~
## $ IndicatorType <chr> NA, "I", "N", "D", "U", "I", "I", "I", "C", "C"~
## $ CharacteristicId <dbl> NA, 10000, 10000, 10000, 10000, 10000, 1000, 10~
## $ CharacteristicOrder <dbl> NA, 10000, 10000, 10000, 10000, 10000, 0, 0, 0,~
```

```

## $ CharacteristicCategory <chr> NA, "Total 15-49", "Total 15-49", "Total 15-49"~
## $ CharacteristicLabel <chr> NA, "Total 15-49", "Total 15-49", "Total 15-49"~
## $ ByVariableId <chr> "#indicator+label+code", "0", "0", "0", "0", "0"~
## $ ByVariableLabel <chr> "#indicator+label", NA, NA, NA, NA, NA, NA, NA,~
## $ IsTotal <dbl> NA, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ IsPreferred <dbl> NA, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ SDRID <chr> NA, "MMMRTWFDP", "MMMRTWPD", "MMMRTWEXP", "~
## $ RegionId <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ SurveyYearLabel <dbl> NA, 1998, 1998, 1998, 1998, 1998, 1998, 1998, 1~
## $ SurveyType <chr> NA, "DHS", "DHS", "DHS", "DHS", "DHS", "DHS", "~
## $ DenominatorWeighted <dbl> NA, NA, NA, NA, NA, 122701, NA, NA, NA, NA, NA,~
## $ DenominatorUnweighted <dbl> NA, NA, NA, 123738, 123738, 123738, NA, NA, NA,~
## $ CILow <dbl> NA, NA, NA, NA, NA, NA, NA, NA, 77, NA, NA, NA, NA,~
## $ CIHigh <dbl> NA, NA, NA, NA, NA, NA, NA, NA, 223, NA, NA, NA, NA~
## $ LevelRank <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~

```

```

##      IS03      DataId      Indicator      Value
## Length:22      Length:22      Length:22      Length:22
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##
##
##

```

```

##      Precision      DHS_CountryCode      CountryName      SurveyYear
## Length:22      Length:22      Length:22      Length:22
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##
##
##

```

```

##      SurveyId      IndicatorId      IndicatorOrder      IndicatorType
## Length:22      Length:22      Min. :77003010      Length:22
## Class :character Class :character      1st Qu.:77003030      Class :character
## Mode  :character Mode  :character      Median :77033010      Mode  :character
##                                     Mean  :77018746
##                                     3rd Qu.:77033030
##                                     Max.  :77033060
##                                     NA's   :1

```

```

##      CharacteristicId CharacteristicOrder CharacteristicCategory
## Min. : 1000      Min. : 0      Length:22
## 1st Qu.: 1000      1st Qu.: 0      Class :character
## Median : 1000      Median : 0      Mode  :character
## Mean : 5286      Mean : 4762
## 3rd Qu.:10000      3rd Qu.:10000
## Max. :10000      Max. :10000
## NA's :1      NA's :1

```

```

##      CharacteristicLabel ByVariableId      ByVariableLabel      IsTotal
## Length:22      Length:22      Length:22      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:22      Mode:logical Min.      :1998
## 1st Qu.:1      Class :character NA's:22      1st Qu.:1998
## Median :1      Mode  :character      Median :2016
## Mean    :1                                     Mean    :2007
## 3rd Qu.:1                                     3rd Qu.:2016
## Max.    :1                                     Max.    :2016
## NA's    :1                                     NA's    :1
##      SurveyType      DenominatorWeighted DenominatorUnweighted      CILow
## Length:22      Min.      : 62768      Min.      : 63523      Min.      : 77.0
## Class :character 1st Qu.: 77751      1st Qu.: 63523      1st Qu.:155.5
## Mode  :character Median : 92735      Median : 93631      Median :234.0
##                                     Mean    : 92735      Mean    :193.7
##                                     3rd Qu.:107718     3rd Qu.:123738     3rd Qu.:252.0
##                                     Max.      :122701     Max.      :123738     Max.      :270.0
##                                     NA's      :20      NA's      :16      NA's      :19
##      CIHigh      LevelRank
## Min.      :223.0 Mode:logical
## 1st Qu.:469.0 NA's:22
## Median :715.0
## Mean    :580.0
## 3rd Qu.:758.5
## Max.    :802.0
## NA's    :19

```

### 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      21
##      IsTotal      IsPreferred      SDRID
##      1      1      1
##      RegionId      SurveyYearLabel      SurveyType
##      22      1      1
##      DenominatorWeighted DenominatorUnweighted      CILow
##      20      16      19
##      CIHigh      LevelRank
##      19      22

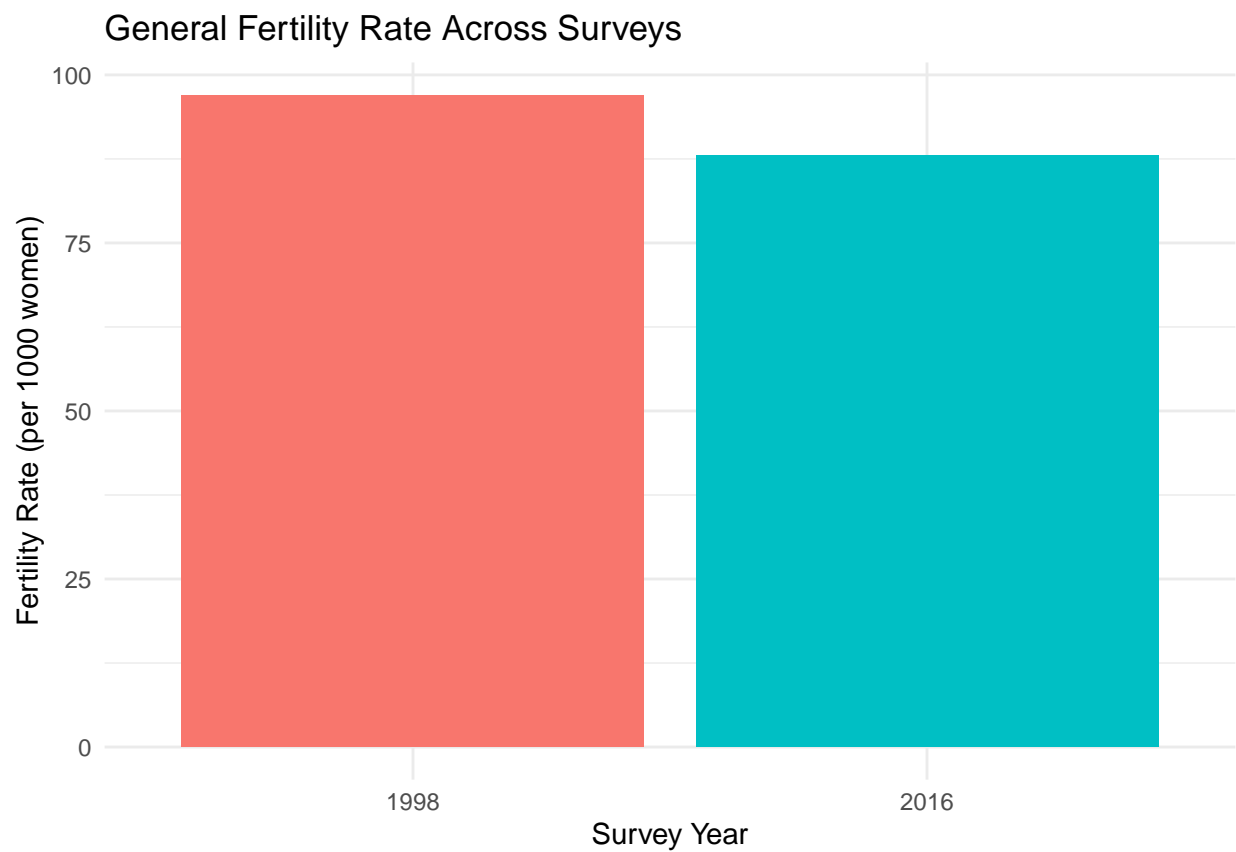
```

Dupilcates

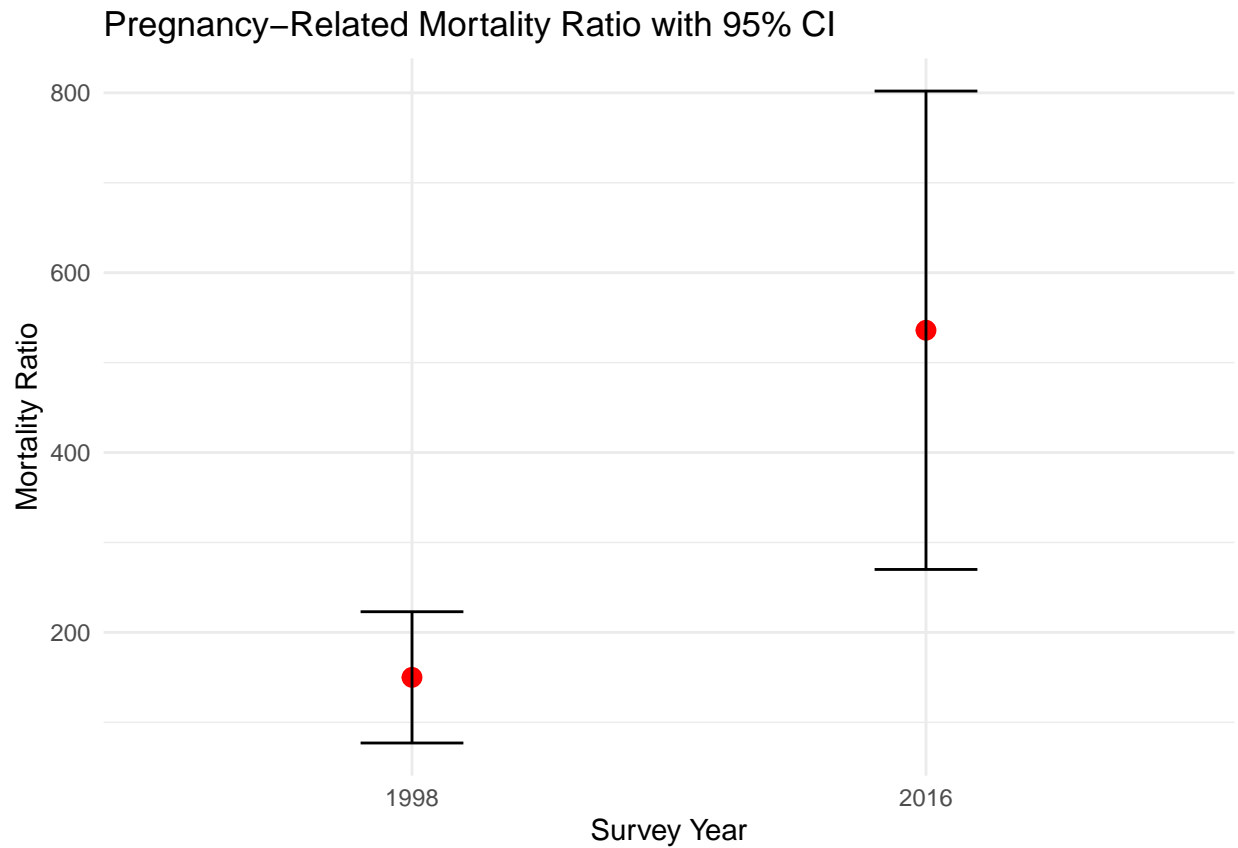
```
## [1] 0
```

## 4. Visualizations

Bar plot showing the distribution of fertility-related indicators



Point plot with error bars to visualize the Pregnancy-related mortality ratio along with its lower and upper bounds.



## 5. Summary Tables

```
## # A tibble: 20 x 3
##   Variable      Statistic      Value
##   <chr>         <chr>         <dbl>
## 1 CIHigh       mean          580
## 2 CIHigh       sd            312.
## 3 CIHigh       min           223
## 4 CIHigh       max           802
## 5 CILow        mean          194.
## 6 CILow        sd            103.
## 7 CILow        min            77
## 8 CILow        max           270
## 9 DenominatorUnweighted mean    93630.
## 10 DenominatorUnweighted sd      32981.
## 11 DenominatorUnweighted min     63523
## 12 DenominatorUnweighted max    123738
## 13 DenominatorWeighted mean    92734.
## 14 DenominatorWeighted sd      42379.
## 15 DenominatorWeighted min     62768
## 16 DenominatorWeighted max    122701
```

## 17 Value	mean	17882.
## 18 Value	sd	39767.
## 19 Value	min	0.005
## 20 Value	max	123738

## 6. Data dictionary

Table 1: Data Dictionary for the DHS Maternal Mortality Dataset

Column	DataType	Description
ISO3	character	ISO3 country code
DataId	character	Unique ID for the dataset row
Indicator	character	Name of the health indicator
Value	character	Numeric value of the indicator (currently character)
Precision	character	Number of decimal places or precision
DHS_CountryCode	character	Country code from DHS dataset
CountryName	character	Full country name
SurveyYear	character	Year survey was conducted
SurveyId	character	Unique ID for the survey
IndicatorId	character	DHS indicator code
IndicatorOrder	numeric	Order of indicator in dataset
IndicatorType	character	Type of indicator (I,N,D,U,C)
CharacteristicId	numeric	ID for the population characteristic
CharacteristicOrder	numeric	Order of the characteristic
CharacteristicCategory	character	Category of the population (e.g., Total 15-49)
CharacteristicLabel	character	Label describing the population characteristic
ByVariableId	character	ID for any breakdown variable
ByVariableLabel	character	Label for breakdown variable
IsTotal	numeric	1 if row represents a total, 0 otherwise
IsPreferred	numeric	1 if this is the preferred row for this indicator, 0 otherwise
SDRID	character	DHS short code for the indicator
RegionId	logical	Region identifier (mostly NA)
SurveyYearLabel	numeric	Year label for survey (redundant with SurveyYear)
SurveyType	character	Type of survey (e.g., DHS)
DenominatorWeighted	numeric	Weighted denominator used to calculate indicator
DenominatorUnweighted	numeric	Unweighted denominator used to calculate indicator
CILow	numeric	Lower bound of the confidence interval
CIHigh	numeric	Upper bound of the confidence interval
LevelRank	logical	Level rank (mostly NA; optional hierarchy indicator)