EDA_Immunization

Group K

2025-09-09

1) Filter to Immunization indicators

```
## # A tibble: 6 x 29
##   ISO3  DataId Indicator   Value Precision DHS_CountryCode CountryName SurveyYear
##   <chr> <chr>  <chr>       <dbl> <chr>     <chr>           <chr>       <fct>
## 1 ZAF   330965 Bcg vacci~   96.8 1         ZA              South Afri~ 1998
## 2 ZAF   139796 Bcg vacci~   94.9 1         ZA              South Afri~ 1998
## 3 ZAF   330966 Dpt 1 vac~   93.3 1         ZA              South Afri~ 1998
## 4 ZAF   139797 Dpt 1 vac~   93.1 1         ZA              South Afri~ 1998
## 5 ZAF   330967 Dpt 2 vac~   86.2 1         ZA              South Afri~ 1998
## 6 ZAF   139798 Dpt 2 vac~   82.4 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 <dbl>, CIHigh <dbl>, LevelRank <lgl>
```

#2) Data overview

```
## Rows: 100
## Columns: 29
## $ ISO3                   <chr> "ZAF", "ZAF", "ZAF", "ZAF", "ZAF", "ZAF", "ZAF"~
## $ DataId                 <chr> "330965", "139796", "330966", "139797", "330967~
## $ Indicator              <chr> "Bcg vaccination received", "Bcg vaccination re~
## $ Value                  <dbl> 96.8, 94.9, 93.3, 93.1, 86.2, 82.4, 76.4, 73.8,~
## $ Precision              <chr> "1", "1", "1", "1", "1", "1", "1", "1", "1", "1~
## $ DHS_CountryCode        <chr> "ZA", "ZA", "ZA", "ZA", "ZA", "ZA", "ZA", "ZA",~
## $ CountryName            <chr> "South Africa", "South Africa", "South Africa",~
## $ SurveyYear             <fct> 1998, 1998, 1998, 1998, 1998, 1998, 1998, 1998,~
## $ SurveyId               <chr> "ZA1998DHS", "ZA1998DHS", "ZA1998DHS", "ZA1998D~
## $ IndicatorId            <chr> "CH_VACS_C_BCG", "CH_VACS_C_BCG", "CH_VACS_C_DP~
## $ IndicatorOrder         <dbl> 93886010, 93886010, 93886020, 93886020, 9388603~
## $ IndicatorType          <chr> "I", "I", "I", "I", "I", "I", "I", "I", "I", "I~
## $ CharacteristicId       <dbl> 268002, 268002, 268002, 268002, 268002, 268002,~
## $ CharacteristicOrder    <dbl> 268002, 268002, 268002, 268002, 268002, 268002,~
## $ CharacteristicCategory <chr> "Source of vaccination information", "Source of~
## $ CharacteristicLabel    <chr> "Either source", "Either source", "Either sourc~
## $ ByVariableId           <chr> "258001", "258002", "258001", "258002", "258001~
```

```
## $ ByVariableLabel        <chr> "12-23", "24-35", "12-23", "24-35", "12-23", "2~
## $ IsTotal                <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
## $ IsPreferred            <dbl> 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0,~
## $ SDRID                  <chr> "CHVACSCBCG", "CHVACSCBCG", "CHVACSCDP1", "CHVA~
## $ RegionId               <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ SurveyYearLabel        <dbl> 1998, 1998, 1998, 1998, 1998, 1998, 1998, 1998,~
## $ SurveyType             <chr> "DHS", "DHS", "DHS", "DHS", "DHS", "DHS", "DHS"~
## $ DenominatorWeighted    <dbl> 973, 933, 973, 933, 973, 933, 973, 933, 973, 93~
## $ DenominatorUnweighted  <dbl> 971, 951, 971, 951, 971, 951, 971, 951, 971, 95~
## $ CILow                  <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ CIHigh                 <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ LevelRank              <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~


##      ISO3              DataId            Indicator             Value
##  Length:100         Length:100         Length:100         Min.   : 2.20
##  Class :character   Class :character   Class :character   1st Qu.:64.90
##  Mode  :character   Mode  :character   Mode  :character   Median :78.05
##                                                           Mean   :72.74
##                                                           3rd Qu.:89.22
##                                                           Max.   :96.80
##
##   Precision         DHS_CountryCode    CountryName         SurveyYear
##  Length:100         Length:100         Length:100         1998:45
##  Class :character   Class :character   Class :character   2016:55
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##    SurveyId           IndicatorId        IndicatorOrder     IndicatorType
##  Length:100         Length:100         Min.   :93886010   Length:100
##  Class :character   Class :character   1st Qu.:93886070   Class :character
##  Mode  :character   Mode  :character   Median :93886250   Mode  :character
##                                        Mean   :93902298
##                                        3rd Qu.:93906230
##                                        Max.   :93936260
##
## CharacteristicId CharacteristicOrder CharacteristicCategory
##  Min.   : 1000   Min.   :     0     Length:100
##  1st Qu.: 1000   1st Qu.:     0     Class :character
##  Median :268002  Median :268002     Mode  :character
##  Mean   :139841  Mean   :139361
##  3rd Qu.:268002  3rd Qu.:268002
##  Max.   :268002  Max.   :268002
##
## CharacteristicLabel ByVariableId      ByVariableLabel       IsTotal
##  Length:100          Length:100         Length:100         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
##
##   IsPreferred        SDRID              RegionId        SurveyYearLabel
```

```
##  Min.   :0.00   Length:100        Mode:logical   Min.   :1998
##  1st Qu.:0.00   Class :character   NA's:100       1st Qu.:1998
##  Median :1.00   Mode  :character                  Median :2016
##  Mean   :0.74                                      Mean   :2008
##  3rd Qu.:1.00                                      3rd Qu.:2016
##  Max.   :1.00                                      Max.   :2016
##
##   SurveyType         DenominatorWeighted DenominatorUnweighted    CILow
##  Length:100         Min.   : 660        Min.   : 670          Min.   : NA
##  Class :character   1st Qu.: 677        1st Qu.: 670          1st Qu.: NA
##  Mode  :character   Median : 933        Median : 951          Median : NA
##                     Mean   :1201        Mean   :1207          Mean   :NaN
##                     3rd Qu.: 973        3rd Qu.: 971          3rd Qu.: NA
##                     Max.   :3734        Max.   :3761          Max.   : NA
##                                                               NA's   :100
##      CIHigh    LevelRank
##  Min.   : NA   Mode:logical
##  1st Qu.: NA   NA's:100
##  Median : NA
##  Mean   :NaN
##  3rd Qu.: NA
##  Max.   : NA
##  NA's   :100
```

Table 1: Data summary

Name	immun
Number of rows	100
Number of columns	29
Column type frequency:	
character	15
factor	1
logical	2
numeric	11
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ISO3	0	1.00	3	3	0	1	0
DataId	0	1.00	5	6	0	100	0
Indicator	0	1.00	24	51	0	28	0
Precision	0	1.00	1	1	0	1	0
DHS_CountryCode	0	1.00	2	2	0	1	0
CountryName	0	1.00	12	12	0	1	0
SurveyId	0	1.00	9	9	0	2	0
IndicatorId	0	1.00	13	13	0	46	0
IndicatorType	0	1.00	1	1	0	1	0
CharacteristicCategory	0	1.00	5	33	0	2	0
CharacteristicLabel	0	1.00	5	13	0	2	0

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ByVariableId	0	1.00	1	6	0	3	0
ByVariableLabel	48	0.52	5	5	0	2	0
SDRID	0	1.00	10	10	0	46	0
SurveyType	0	1.00	3	3	0	1	0

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
SurveyYear	0	1	FALSE	2	201: 55, 199: 45

Variable type: logical

skim_variable	n_missing	complete_rate	mean	count
RegionId	100	0	NaN	:
LevelRank	100	0	NaN	:

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Value	0	1	72.74	23.28	2.2	64.9	78.05	89.22	96.8	
IndicatorOrder	0	1	93902297.90	10889.40	93886010.0	93886070.0	93886250.0	93906230.0	93936260.0	
CharacteristicId	0	1	139841.04	134066.17	1000.0	1000.0	268002.00	268002.00	268002.0	
CharacteristicOrder	0	1	139361.04	134568.29	0.0	0.0	268002.00	268002.00	268002.0	
IsTotal	0	1	1.00	0.00	1.0	1.0	1.00	1.00	1.0	
IsPreferred	0	1	0.74	0.44	0.0	0.0	1.00	1.00	1.0	
SurveyYearLabel	0	1	2007.90	9.00	1998.0	1998.0	2016.00	2016.00	2016.0	
DenominatorWeighted	0	1	1200.57	962.63	660.0	677.0	933.00	973.00	3734.0	
DenominatorUnweighted	0	1	1206.55	970.76	670.0	670.0	951.00	971.00	3761.0	
CILow	100	0	NaN	NA	NA	NA	NA	NA	NA	
CIHigh	100	0	NaN	NA	NA	NA	NA	NA	NA	

#3) Data quality checks

```
##             ISO3              DataId             Indicator
##                0                   0                     0
##            Value           Precision        DHS_CountryCode
##                0                   0                     0
##      CountryName          SurveyYear              SurveyId
##                0                   0                     0
##      IndicatorId      IndicatorOrder          IndicatorType
##                0                   0                     0
##   CharacteristicId  CharacteristicOrder CharacteristicCategory
##                0                   0                     0
##  CharacteristicLabel       ByVariableId        ByVariableLabel
##                0                   0                    48
##          IsTotal         IsPreferred                 SDRID
##                0                   0                     0
```

```
##                RegionId      SurveyYearLabel               SurveyType
##                     100                    0                        0
##     DenominatorWeighted  DenominatorUnweighted                    CILow
##                       0                    0                      100
##                  CIHigh            LevelRank
##                     100                  100


## [1] 0


## # A tibble: 1 x 2
##   out_of_bounds pct_rows_violating
##           <int>              <dbl>
## 1               0                  0
```


Immunization Coverage by Vaccine (faceted b