EDA_Literacy_07

Group K

2025-09-09

1. Loading the dataset

```
## # 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    563770 Women wi~ 11.8  1         ZA              South Afri~ 2016
## 3 ZAF    563771 Women wh~ 76.2  1         ZA              South Afri~ 2016
## 4 ZAF    563772 Women wh~ 8.2   1         ZA              South Afri~ 2016
## 5 ZAF    563773 Women wh~ 3.5   1         ZA              South Afri~ 2016
## # 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

```
## Rows: 21
## Columns: 29
## $ ISO3                   <chr> "#country+code", "ZAF", "ZAF", "ZAF", "ZAF", "Z~
## $ DataId                 <chr> "#meta+id", "563770", "563771", "563772", "5637~
## $ Indicator              <chr> "#indicator+name", "Women with secondary or hig~
## $ Value                  <chr> "#indicator+value+num", "11.8", "76.2", "8.2", ~
## $ Precision              <chr> "#indicator+precision", "1", "1", "1", "1", "1"~
## $ DHS_CountryCode        <chr> NA, "ZA", "ZA", "ZA", "ZA", "ZA", "ZA", "ZA", "~
## $ CountryName            <chr> "#country+name", "South Africa", "South Africa"~
## $ SurveyYear             <chr> "#date+year", "2016", "2016", "2016", "2016", "~
## $ SurveyId               <chr> "#survey+id", "ZA2016DHS", "ZA2016DHS", "ZA2016~
## $ IndicatorId            <chr> "#indicator+code", "ED_LITR_W_SCH", "ED_LITR_W_~
## $ IndicatorOrder         <dbl> NA, 231233010, 231233020, 231233030, 231233040,~
## $ IndicatorType          <chr> NA, "I", "I", "I", "I", "I", "I", "T", "I", "D"~
## $ CharacteristicId       <dbl> NA, 10000, 10000, 10000, 10000, 10000, 10000, 1~
## $ CharacteristicOrder    <dbl> NA, 10000, 10000, 10000, 10000, 10000, 10000, 1~
## $ 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, "EDLITRWSCH", "EDLITRWRDW", "EDLITRWRDP", "~
## $ RegionId               <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ SurveyYearLabel        <dbl> NA, 2016, 2016, 2016, 2016, 2016, 2016, 2016, 2~
## $ SurveyType             <chr> NA, "DHS", "DHS", "DHS", "DHS", "DHS", "DHS", "~
## $ DenominatorWeighted    <dbl> NA, 8514, 8514, 8514, 8514, 8514, 8514, 8514, 8~
## $ DenominatorUnweighted  <dbl> NA, 11805, 11805, 11805, 11805, 11805, 11805, 1~
## $ CILow                  <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ CIHigh                 <lgl> 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:21          Length:21          Length:21          Length:21         
##  Class :character   Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character  
##                                                                             
##                                                                             
##                                                                             
##                                                                             
##   Precision         DHS_CountryCode    CountryName         SurveyYear       
##  Length:21          Length:21          Length:21          Length:21         
##  Class :character   Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character  
##                                                                             
##                                                                             
##                                                                             
##                                                                             
##    SurveyId         IndicatorId        IndicatorOrder      IndicatorType     
##  Length:21          Length:21          Min.   :231233010   Length:21         
##  Class :character   Class :character   1st Qu.:231233058   Class :character  
##  Mode  :character   Mode  :character   Median :231233615   Mode  :character  
##                                        Mean   :231233614                     
##                                        3rd Qu.:231234162                     
##                                        Max.   :231234220                     
##                                        NA's   :1                             
##  CharacteristicId CharacteristicOrder CharacteristicCategory
##  Min.   :10000    Min.   :10000       Length:21             
##  1st Qu.:10000    1st Qu.:10000       Class :character      
##  Median :10000    Median :10000       Mode  :character      
##  Mean   :10000    Mean   :10000                             
##  3rd Qu.:10000    3rd Qu.:10000                             
##  Max.   :10000    Max.   :10000                             
##  NA's   :1        NA's   :1                                 
##  CharacteristicLabel ByVariableId       ByVariableLabel       IsTotal  
##  Length:21           Length:21          Length:21          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:21          Mode:logical   Min.   :2016   
##  1st Qu.:1    Class :character   NA's:21        1st Qu.:2016   
##  Median :1    Mode  :character                  Median :2016   
##  Mean   :1                                      Mean   :2016   
##  3rd Qu.:1                                      3rd Qu.:2016   
##  Max.   :1                                      Max.   :2016   
##  NA's   :1                                      NA's   :1      
##   SurveyType        DenominatorWeighted DenominatorUnweighted  CILow        
##  Length:21          Min.   :3202        Min.   : 3179        Mode:logical  
##  Class :character   1st Qu.:3202        1st Qu.: 3179        NA's:21       
##  Mode  :character   Median :5858        Median : 7492                      
##                     Mean   :5858        Mean   : 7492                      
##                     3rd Qu.:8514        3rd Qu.:11805                      
##                     Max.   :8514        Max.   :11805                      
##                     NA's   :3           NA's   :3                          
##   CIHigh        LevelRank     
##  Mode:logical   Mode:logical  
##  NA's:21        NA's:21       
##                               
##                               
##                               
##                               
##                               
```

3. Data Quality

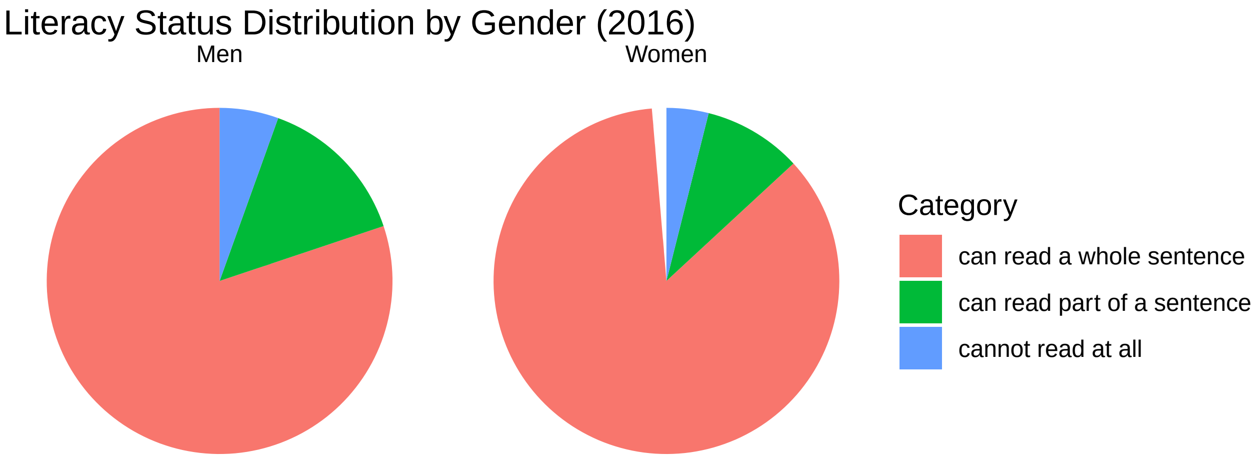
Missing values per column

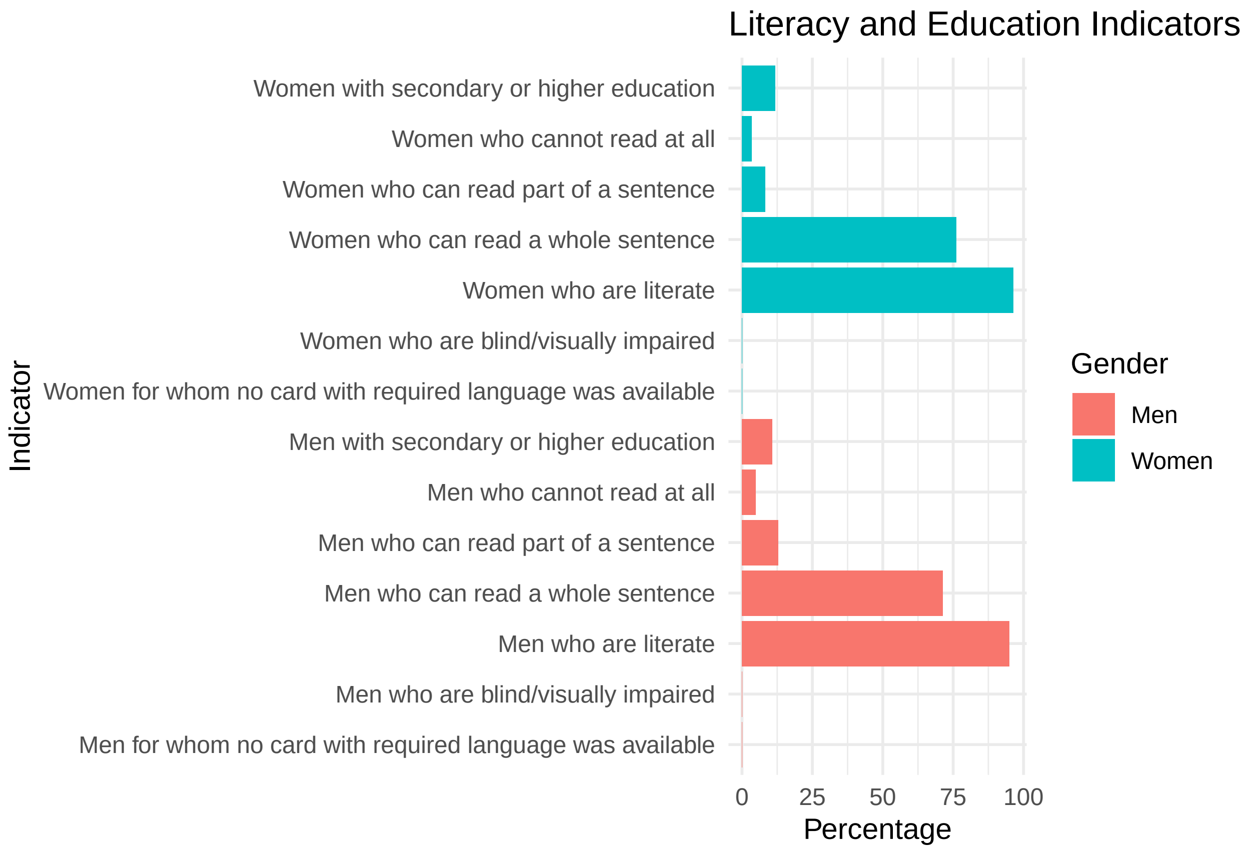
```
##                   ISO3                 DataId              Indicator 
##                      0                      0                      0 
##                  Value              Precision        DHS_CountryCode 
##                      0                      0                      1 
##            CountryName             SurveyYear               SurveyId 
##                      0                      0                      0 
##            IndicatorId         IndicatorOrder          IndicatorType 
##                      0                      1                      1 
##       CharacteristicId    CharacteristicOrder CharacteristicCategory 
##                      1                      1                      1 
##    CharacteristicLabel           ByVariableId        ByVariableLabel 
##                      1                      0                     20 
##                IsTotal            IsPreferred                  SDRID 
##                      1                      1                      1 
##               RegionId        SurveyYearLabel             SurveyType 
##                     21                      1                      1 
##    DenominatorWeighted  DenominatorUnweighted                  CILow 
##                      3                      3                     21 
##                 CIHigh              LevelRank 
##                     21                     21
```

Dupilcates

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

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 Literacy Dataset

Variable	Statistic	Value
CIHigh	mean	NaN
CIHigh	sd	NA
CIHigh	min	Inf
CIHigh	max	-Inf
CILow	mean	NaN
CILow	sd	NA
CILow	min	Inf
CILow	max	-Inf
DenominatorUnweighted	mean	7492.000
DenominatorUnweighted	sd	4438.040
DenominatorUnweighted	min	3179.000
DenominatorUnweighted	max	11805.000
DenominatorWeighted	mean	5858.000
DenominatorWeighted	sd	2733.001
DenominatorWeighted	min	3202.000
DenominatorWeighted	max	8514.000
Value	mean	1364.555
Value	sd	3203.736
Value	min	0.000
Value	max	11805.000