Codebook for JColeman\_

Livesession5assignment

# Data Frames

**df:** Contains a series of popular children’s names born in the year 2016 in the United States from the file yob2016.txt.

The dataset examined has the following dimensions:

|  |  |
| --- | --- |
| Feature | Result |
| Number of observations | 32869 |
| Number of variables | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Class | # unique values | Missing |
| **[First\_Name]** | factor | 30295 | 0.00 % |
| **[Gender]** | factor | 2 | 0.00 % |
| **{[Amount\_of\_Children](#amount_of_children)]** | integer | 1385 | 0.00 % |

**y2016:** Contents of dataframe df omitting row 212 which contained the name Fionayyy.

**y2015:** Contains a series of popular children’s names born in the year 2016 in the United States from the file yob2015.txt.

The dataset examined has the following dimensions:

|  |  |
| --- | --- |
| Feature | Result |
| Number of observations | 33063 |
| Number of variables | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Class | # unique values | Missing |
| **[First\_Name]** | factor | 30553 | 0.00 % |
| **[Gender]** | factor | 2 | 0.00 % |
| **{[Amount\_of\_Children](#amount_of_children)]** | integer | 1397 | 0.00 % |

**final:** Merging of data frames y2015 and y2016 by [First\_Name}. Column added to sum totals for each name’s use in 2015 and 2016.

The dataset examined has the following dimensions:

|  |  |
| --- | --- |
| Feature | Result |
| Number of observations | 31641 |
| Number of variables | 6 |

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Class | # unique values | Missing |
| **[First\_Name]** | factor | 24636 | 0.00 % |
| **[[Gender.x](#gender.x)]** | factor | 2 | 0.00 % |
| **[[Amount\_of\_Children.x](#amount_of_children.x)]** | integer | 1385 | 0.00 % |
| **[[Gender.y](#gender.y)]** | factor | 2 | 0.00 % |
| **[[Amount\_of\_Children.y](#amount_of_children.y)]** | integer | 1397 | 0.00 % |
| **[**[**Total**](#total)**]** | integer | 2558 | 0.00 % |

**finalFemale:** Contents of dataframe final omitting rows where Gender.x or Gender.y is Male.

**Top10Female:** Top 10 rows of dataframe finalFemale ordered by Total column.

Report table generation information:

* Created by Jasmine Coleman (username: Jazzy).
* Report creation time: Tue Feb 12 2019 19:53:27
* Report was run from directory: C:/Users/Jazzy/OneDrive/Documents/SMU/Spring '19/Doing Data Science/Unit 5
* dataMaid v1.2.0 [Pkg: 2018-10-03 from CRAN (R 3.5.2)]
* R version 3.5.2 (2018-12-20).
* Platform: x86\_64-w64-mingw32/x64 (64-bit)(Windows >= 8 x64 (build 9200)).
* Function call: makeDataReport(data = final, mode = c("summarize", "visualize", "check"), smartNum = FALSE, file = "codebook\_final.Rmd", checks = list( character = "showAllFactorLevels", factor = "showAllFactorLevels", labelled = "showAllFactorLevels", haven\_labelled = "showAllFactorLevels", numeric = NULL, integer = NULL, logical = NULL, Date = NULL), listChecks = FALSE, maxProbVals = Inf, codebook = TRUE, reportTitle = "2015-2016 Data Table")