HW5-Codebook

Caroll Rodriguez

February 13, 2018

## Homework 5 Codebook

##### Backstory: Your client is expecting a baby soon. However, he is not sure what to name the child. Being out of the loop, he hires you to help him figure out popular names. He provides for you raw data in order to help you make a decision.

### Original Dataset yob2016.txt (df)

Note: This file is a series of popular children’s names born in the year 2016 in the United States. It consists of three columns with a first name, a gender, and the amount of children given that name. However, the data is raw and will need cleaning to make it tidy and usable. FirstName Gender Amount  
Length:32869 Length:32869 Min. : 5.0  
Class :character Class :character 1st Qu.: 7.0  
Mode :character Mode :character Median : 12.0  
Mean : 110.7  
3rd Qu.: 30.0  
Max. :19414.0  
‘data.frame’: 32869 obs. of 3 variables: $ FirstName: chr “Emma” “Olivia” “Ava” “Sophia” … $ Gender : chr “F” “F” “F” “F” … $ Amount : int 19414 19246 16237 16070 14722 14366 13030 11699 10926 10733 …

#### Original Dataset yob2015.txt

Note: This file is similar to yob2016, but contains names, gender, and total children given that name for the year 2015.

#### $ FirstName: chr “Emma” “Olivia” “Sophia” “Ava” …

#### $ Gender : chr “F” “F” “F” “F” …

#### $ Amount : int 20415 19638 17381 16340 15574 14871 12371 11766 11381 10283 …

FirstName Gender Amount  
Length:33063 Length:33063 Min. : 5.0  
Class :character Class :character 1st Qu.: 7.0  
Mode :character Mode :character Median : 11.0  
Mean : 111.4  
3rd Qu.: 30.0  
Max. :20415.0 str(y2015) ‘data.frame’: 33063 obs. of 3 variables: $ FirstName: chr “Emma” “Olivia” “Sophia” “Ava” … $ Gender : chr “F” “F” “F” “F” … $ Amount : int 20415 19638 17381 16340 15574 14871 12371 11766 11381 10283 …

#### Merged Dataset (final)

Note: Final dataset is a merged table consisting of ‘data.frame’: 39381 obs. of 4 variables: $ FirstName: chr “Aaban” “Aabha” “Aabid” “Aabir” … $ Gender : chr “M” “F” “M” “M” … $ Amount.x : num 9 7 5 5 11 0 18 0 11 194 … $ Amount.y : num 15 7 0 0 5 5 22 10 15 297 …

FirstName Gender Amount.x Amount.y  
Length:39381 Length:39381 Min. : 0.00 Min. : 0.00  
Class :character Class :character 1st Qu.: 5.00 1st Qu.: 5.00  
Mode :character Mode :character Median : 9.00 Median : 9.00  
Mean : 92.36 Mean : 93.54  
3rd Qu.: 24.00 3rd Qu.: 24.00  
Max. :19414.00 Max. :20415.00

#### Top10PopNames

Note: Top10PopNames <- Top 10 popular names from Finals dataset.

#### Top10GirlsNames

Note: Top10GirlsNames <- Top 10 Girls names from Finals dataset.