

# Preprocessing

*Admin*

24 4 2019

## Raw Dataset

## text graphs are displayed; set 'tmp.img.dir' parameter to activate png graphs

## Data Frame Summary

data

Dimensions: 4521 x 17

Duplicates: 0

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
1	age [integer]	Mean (sd) : 41.2 (10.6) min < med < max: 19 < 39 < 87 IQR (CV) : 16 (0.3)	67 distinct values	: : : . : : : : : : : : : . : : : : .	4521 (100%)	0 (0%)
2	job [factor]	1. admin. 2. blue-collar 3. entrepreneur 4. housemaid 5. management 6. retired 7. self-employed 8. services 9. student 10. technician [ 2 others ]	478 (10.6%) 946 (20.9%) 168 ( 3.7%) 112 ( 2.5%) 969 (21.4%) 230 ( 5.1%) 183 ( 4.0%) 417 ( 9.2%) 84 ( 1.9%) 768 (17.0%) 166 ( 3.7%)	II IIII    IIII I  I   III	4521 (100%)	0 (0%)
3	marital [factor]	1. divorced 2. married 3. single	528 (11.7%) 2797 (61.9%) 1196 (26.5%)	II IIIIIIIIII IIII	4521 (100%)	0 (0%)
4	education [factor]	1. primary 2. secondary 3. tertiary 4. unknown	678 (15.0%) 2306 (51.0%) 1350 (29.9%) 187 ( 4.1%)	II IIIIIIII IIII	4521 (100%)	0 (0%)
5	default [factor]	1. no 2. yes	4445 (98.3%) 76 ( 1.7%)	IIIIIIIIIIIIIIIIIIII	4521 (100%)	0 (0%)
6	balance [integer]	Mean (sd) : 1422.7 (3009.6) min < med < max: -3313 < 444 < 71188 IQR (CV) : 1411 (2.1)	2353 distinct values	: : : : : .	4521 (100%)	0 (0%)



No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
17	y [factor]	1. no 2. yes	4000 (88.5%) 521 (11.5%)	IIIIIIIIIIIIIIIIIIII II	4521 (100%)	0 (0%)

## Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.