df

Autogenerated data summary from dataMaid

2022-04-05 09:10:39

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	28534
Number of variables	21

Checks performed

The following variable checks were performed, depending on the data type of each variable:

	character	factor	labelled	haven labelled	numeric	integer	logical	Date
Identify miscoded missing values	×	×	×	×	×	×		×
Identify prefixed and suffixed whitespace	×	×	×	×				
Identify levels with < 6 obs.	×	×	×	×				
Identify case issues	×	×	×	×				
Identify misclassified numeric or integer variables	×	×	×	×				
Identify outliers					×	×		×

Please note that all numerical values in the following have been rounded to 2 decimals.

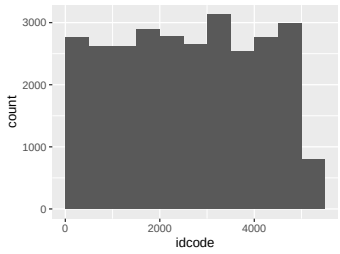
Summary table

	Variable class	# unique values	Missing observations	Any problems?
idcode	numeric	4711	0.00 %	
year	numeric	15	0.00 %	
birth_yr	numeric	14	0.00 %	
age	numeric	34	0.08 %	
race	numeric	3	0.00 %	
msp	numeric	3	0.06 %	
nev_mar	numeric	3	0.06 %	
grade	numeric	20	0.01 %	×
collgrad	numeric	2	0.00 %	
not_smsa	numeric	3	0.03 %	
c_city	numeric	3	0.03 %	
south	numeric	3	0.03 %	
ind_code	numeric	13	1.20 %	×
occ_code	numeric	14	0.42 %	×
union	numeric	3	32.58 %	
wks_ue	numeric	62	19.99 %	×
ttl_exp	numeric	4744	0.00 %	
tenure	numeric	271	1.52 %	
hours	numeric	86	0.23 %	×
wks_work	numeric	106	2.46 %	×
ln_wage	numeric	8173	0.00 %	×

Variable list

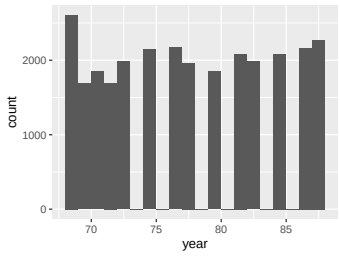
idcode

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	4711
Median	2606
1st and 3rd quartiles	1327; 3881
Min. and max.	1; 5159



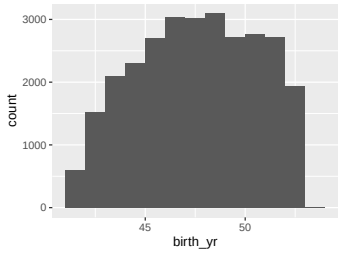
year

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	15
Median	78
1st and 3rd quartiles	72; 83
Min. and max.	68; 88



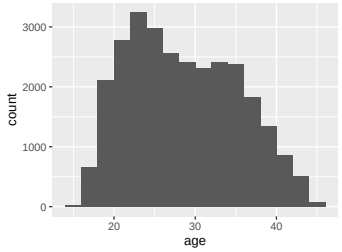
birth_yr

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	14
Median	48
1st and 3rd quartiles	46; 51
Min. and max.	41; 54



age

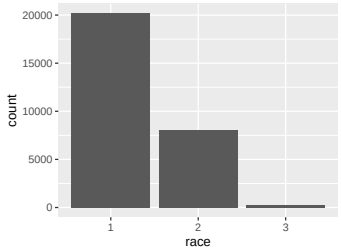
Feature	Result
Variable type	numeric
Number of missing obs.	24 (0.08 %)
Number of unique values	33
Median	28
1st and 3rd quartiles	23; 34
Min. and max.	14; 46



race

- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

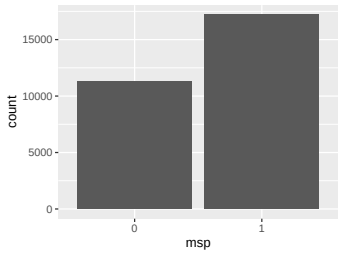
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	3
Mode	“1”
Reference category	1



msp

- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

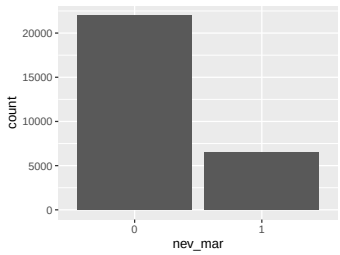
Feature	Result
Variable type	numeric
Number of missing obs.	16 (0.06 %)
Number of unique values	2
Mode	“1”
Reference category	0



nev_mar

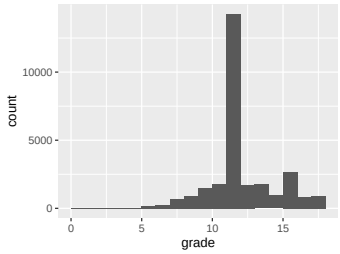
- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

Feature	Result
Variable type	numeric
Number of missing obs.	16 (0.06 %)
Number of unique values	2
Mode	“0”
Reference category	0



grade

Feature	Result
Variable type	numeric
Number of missing obs.	2 (0.01 %)
Number of unique values	19
Median	12
1st and 3rd quartiles	12; 14
Min. and max.	0; 18

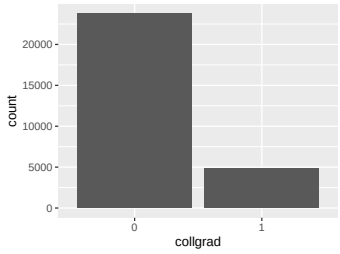


- Note that the following possible outlier values were detected: "0", "1", "2", "3", "4", . . . , "7", "8", "9", "10", "11" (2 values omitted).

collgrad

- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

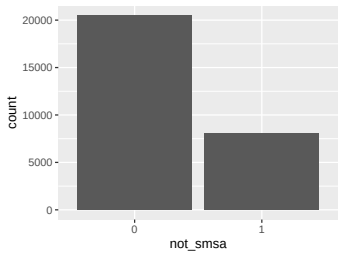
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	“0”
Reference category	0



not_smsa

- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

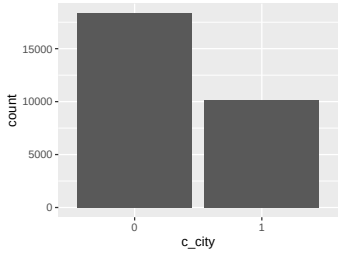
Feature	Result
Variable type	numeric
Number of missing obs.	8 (0.03 %)
Number of unique values	2
Mode	“0”
Reference category	0



c_city

- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

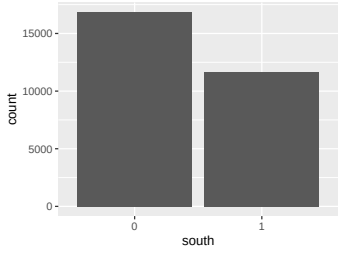
Feature	Result
Variable type	numeric
Number of missing obs.	8 (0.03 %)
Number of unique values	2
Mode	“0”
Reference category	0



south

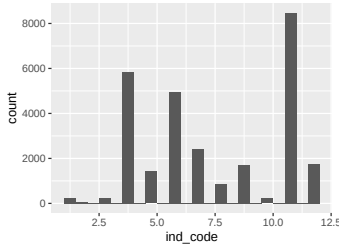
- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

Feature	Result
Variable type	numeric
Number of missing obs.	8 (0.03 %)
Number of unique values	2
Mode	“0”
Reference category	0



ind_code

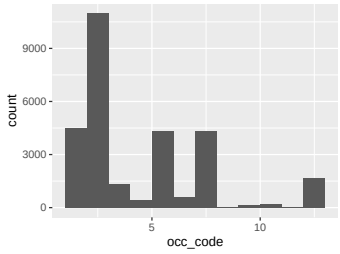
Feature	Result
Variable type	numeric
Number of missing obs.	341 (1.2 %)
Number of unique values	12
Median	7
1st and 3rd quartiles	5; 11
Min. and max.	1; 12



- Note that the following possible outlier values were detected: "1".

occ_code

Feature	Result
Variable type	numeric
Number of missing obs.	121 (0.42 %)
Number of unique values	13
Median	3
1st and 3rd quartiles	3; 6
Min. and max.	1; 13

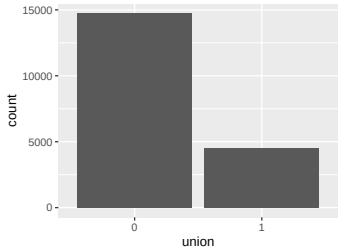


- Note that the following possible outlier values were detected: "1", "2".

union

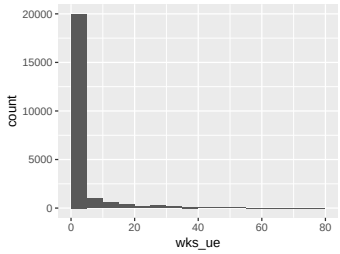
- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

Feature	Result
Variable type	numeric
Number of missing obs.	9296 (32.58 %)
Number of unique values	2
Mode	“0”
Reference category	0



wks_ue

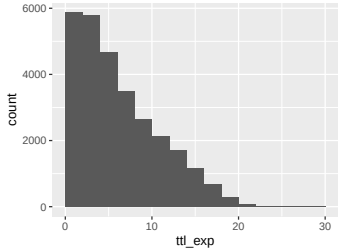
Feature	Result
Variable type	numeric
Number of missing obs.	5704 (19.99 %)
Number of unique values	61
Median	0
1st and 3rd quartiles	0; 0
Min. and max.	0; 76



- Note that the following possible outlier values were detected: "1", "2", "3", "4", "5", . . . , "56", "62", "73", "75", "76" (50 values omitted).

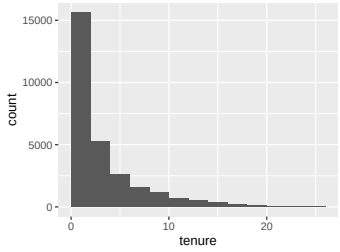
ttl_exp

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	4744
Median	5.06
1st and 3rd quartiles	2.46; 9.13
Min. and max.	0; 28.88



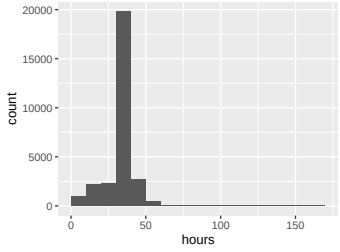
tenure

Feature	Result
Variable type	numeric
Number of missing obs.	433 (1.52 %)
Number of unique values	270
Median	1.67
1st and 3rd quartiles	0.5; 4.17
Min. and max.	0; 25.92



hours

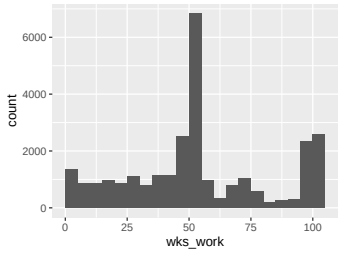
Feature	Result
Variable type	numeric
Number of missing obs.	67 (0.23 %)
Number of unique values	85
Median	40
1st and 3rd quartiles	35; 40
Min. and max.	1; 168



- Note that the following possible outlier values were detected: "41", "42", "43", "44", "45", ..., "99", "100", "105", "112", "168" (35 values omitted).

wks_work

Feature	Result
Variable type	numeric
Number of missing obs.	703 (2.46 %)
Number of unique values	105
Median	52
1st and 3rd quartiles	36; 72
Min. and max.	0; 104

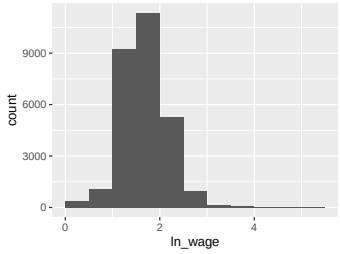


- Note that the following possible outlier values were detected: "0", "1".

---

ln_wage

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	8173
Median	1.64
1st and 3rd quartiles	1.36; 1.96
Min. and max.	0; 5.26



- Note that the following possible outlier values were detected: "0", "0", "0", "0.01", "0.02", . . . , "4.35", "4.35", "4.5", "4.83", "5.26" (448 values omitted).

Report generation information:

- Created by: Miguel Portela (username: `miguelportela`).
- Report creation time: Ter Abr 05 2022 09:10:41
- Report was run from directory: `/Users/miguelportela/Documents/GitHub/R_Training/EDA`
- dataMaid v1.4.1 [Pkg: 2021-10-08 from CRAN (R 4.1.0)]
- R version 4.1.3 (2022-03-10).
- Platform: x86_64-apple-darwin17.0 (64-bit)(macOS Monterey 12.3.1).
- Function call: `makeDataReport(data = df, output = "dM_nlswork.html", replace = TRUE)`