df

Autogenerated data summary from dataMaid

2021-09-18 11:28:36

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	1338
Number of variables	7

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?
age	numeric	47	0.00 %	
sex	character	2	0.00 %	
bmi	numeric	548	0.00 %	×
children	numeric	6	0.00 %	
smoker	character	2	0.00 %	
region	character	4	0.00 %	
charges	numeric	1337	0.00 %	

Variable list

age

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	47
Median	39
1st and 3rd quartiles	27; 51
Min. and max.	18; 64
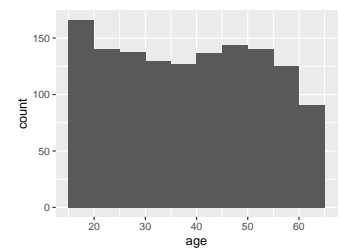
sex

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"male"
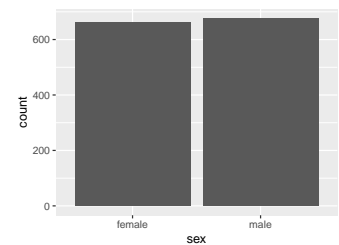
bmi

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	548
Median	30.4
1st and 3rd quartiles	26.3; 34.69
Min. and max.	15.96; 53.13
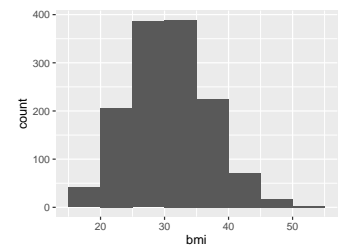
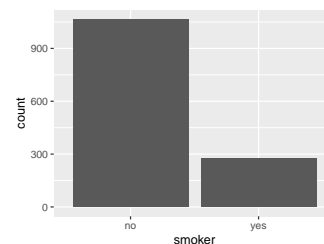- Note that the following possible outlier values were detected: "49.06", "50.38", "52.58", "53.13".

children

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	6
Median	1
1st and 3rd quartiles	0; 2
Min. and max.	0; 5
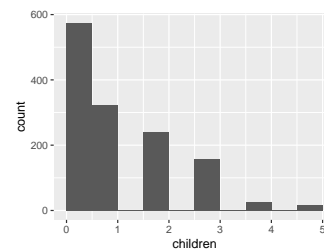
smoker

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"no"

region

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	4
Mode	"southeast"
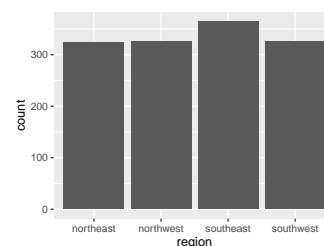
charges

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1337
Median	9382.03
1st and 3rd quartiles	4740.29; 16639.91
Min. and max.	1121.87; 63770.43
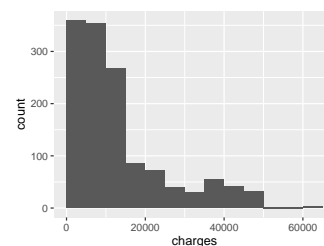
Report generation information:

- Created by: Julio Waissman Vilanova (username: `juliowaissman`).

- Report creation time: Sat Sep 18 2021 11:28:36

- Report was run from directory: `/Users/juliowaissman/Documents/cursos/IC-MCD-Unison/2021-2/ing-caract/ejempl`

- dataMaid v1.4.0 [Pkg: 2019-12-10 from CRAN (R 4.0.2)]

- R version 4.0.3 (2020-10-10).

- Platform: x86_64-apple-darwin17.0 (64-bit)(macOS Big Sur 10.16).

- Function call: `makeDataReport(data = df, render = FALSE, file = "insurance_dataMaid.Rmd",` `replace = TRUE)`