

df_clean

Autogenerated data summary from dataReporter

2021-12-03 10:00:02

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	4133
Number of variables	15

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.

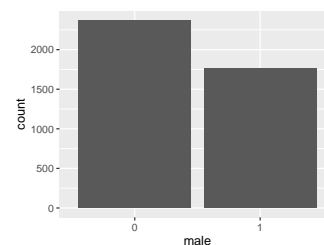
Summary table

	Variable class	# unique values	Missing observations	Any problems?
male	factor	2	0.00 %	
age	numeric	39	0.00 %	
education	factor	2	0.00 %	
currentSmoker	factor	2	0.00 %	
cigsPerDay	numeric	33	0.00 %	
prevalentStroke	factor	2	0.00 %	
prevalentHyp	factor	2	0.00 %	
diabetes	factor	2	0.00 %	
totChol	numeric	247	0.00 %	×
sysBP	numeric	234	0.00 %	×
diaBP	numeric	146	0.00 %	×
BMI	numeric	1355	0.00 %	×
heartRate	numeric	73	0.00 %	×
glucose	numeric	140	0.00 %	×
TenYearCHD	factor	2	0.00 %	

Variable list

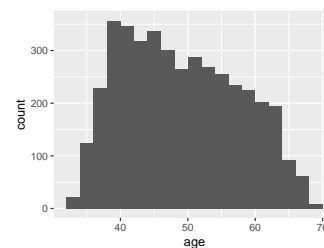
male

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



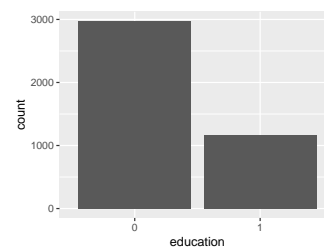
age

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	39
Median	49
1st and 3rd quartiles	42; 56
Min. and max.	32; 70



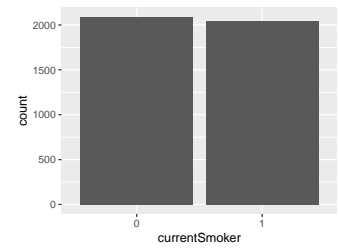
education

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



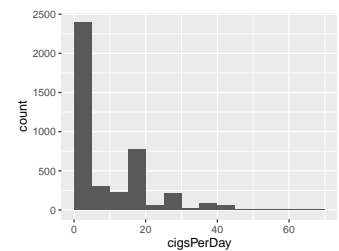
currentSmoker

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"0"
Reference category	0



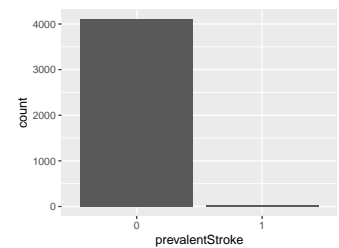
cigsPerDay

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	33
Median	0
1st and 3rd quartiles	0; 20
Min. and max.	0; 70



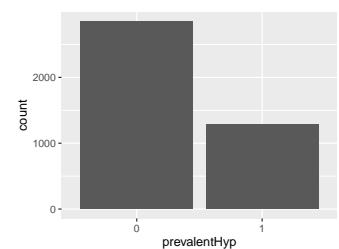
prevalentStroke

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"0"
Reference category	0



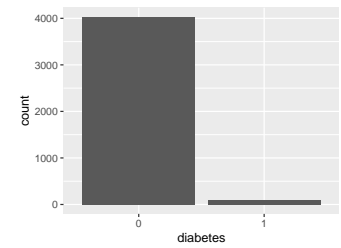
prevalentHyp

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"0"
Reference category	0



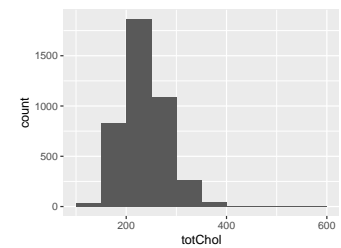
diabetes

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"0"
Reference category	0



totChol

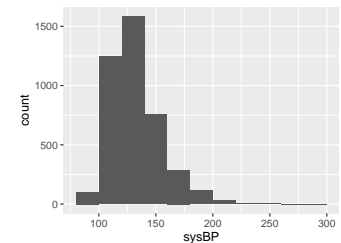
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	247
Median	234
1st and 3rd quartiles	206; 262
Min. and max.	107; 600



- Note that the following possible outlier values were detected: "107", "113", "119", "124", "126", ..., "432", "439", "453", "464", "600" (28 values omitted).

sysBP

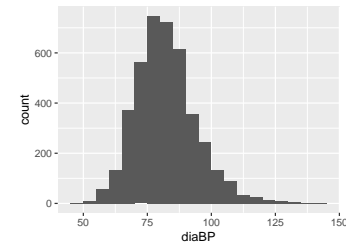
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	234
Median	128
1st and 3rd quartiles	117; 144
Min. and max.	83.5; 295



- Note that the following possible outlier values were detected: "83.5", "85", "85.5", "90", "92", ..., "235", "243", "244", "248", "295" (15 values omitted).

diaBP

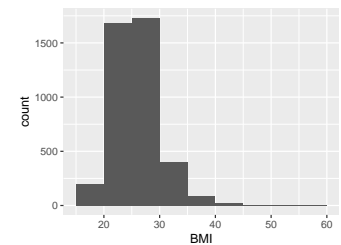
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	146
Median	82
1st and 3rd quartiles	75; 89.5
Min. and max.	48; 142.5



- Note that the following possible outlier values were detected: "48", "50", "51", "52", "53", ..., "133", "135", "136", "140", "142.5" (22 values omitted).

BMI

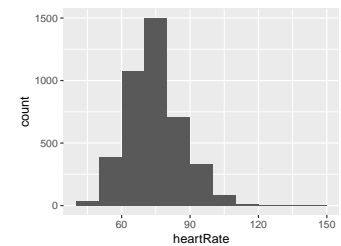
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1355
Median	25.38
1st and 3rd quartiles	23.06; 27.99
Min. and max.	15.54; 56.8



- Note that the following possible outlier values were detected: "15.54", "15.96", "16.48", "16.59", "16.61", ..., "44.71", "45.79", "45.8", "51.28", "56.8" (73 values omitted).

heartRate

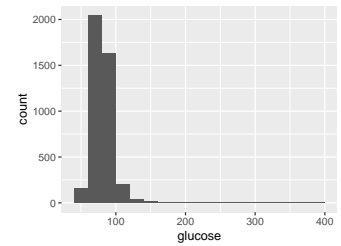
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	73
Median	75
1st and 3rd quartiles	68; 83
Min. and max.	44; 143



- Note that the following possible outlier values were detected: "44", "45", "46", "47", "48", ..., "122", "125", "130", "140", "143" (5 values omitted).

glucose

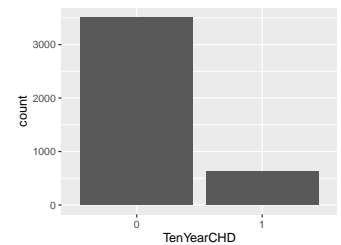
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	140
Median	80
1st and 3rd quartiles	72; 85
Min. and max.	40; 394



- Note that the following possible outlier values were detected: "40", "43", "44", "45", "47", ..., "348", "368", "370", "386", "394" (81 values omitted).

TenYearCHD

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"0"
Reference category	0



Report generation information:

- Created by: mccurcio (username: mcc).
- Report creation time: Fri Dec 03 2021 10:00:02
- Report was run from directory: /home/mcc/Projects/Fram_logit
- dataReporter v1.0.2 [Pkg: 2021-11-11 from CRAN (R 3.6.3)]
- R version 3.6.3 (2020-02-29).
- Platform: x86_64-pc-linux-gnu (64-bit)(Linux Mint 20.2).
- Function call: `makeDataReport(data = df_clean)`