

hd

Autogenerated data summary from dataMaid

2020-07-02 10:16:58

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	303
Number of variables	14

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.

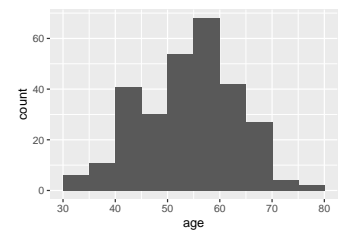
Summary table

	Variable class	# unique values	Missing observations	Any problems?
age	numeric	41	5.94 %	×
sex	character	3	3.96 %	
chest_pain_type	character	5	4.62 %	
resting_bp	numeric	50	4.95 %	×
cholesterol	numeric	151	3.96 %	×
fasting_blood_sugar	character	3	8.25 %	
resting_ecg	numeric	4	3.96 %	×
max_hearttrate	numeric	91	3.96 %	×
exer_angina	character	3	5.28 %	
old_peak	numeric	41	2.64 %	×
slope	character	4	6.27 %	
n_vessels	numeric	6	4.29 %	×
defect	character	4	6.93 %	
heart_disease	numeric	3	5.61 %	

Variable list

age

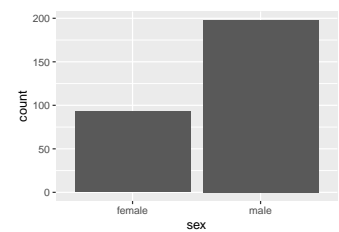
Feature	Result
Variable type	numeric
Number of missing obs.	18 (5.94 %)
Number of unique values	40
Median	56
1st and 3rd quartiles	48; 61
Min. and max.	34; 77



- Note that the following possible outlier values were detected: "71", "74", "76", "77".

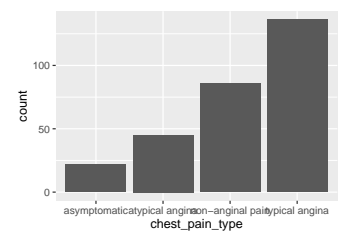
sex

Feature	Result
Variable type	character
Number of missing obs.	12 (3.96 %)
Number of unique values	2
Mode	"male"



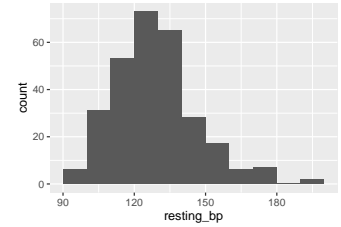
chest_pain_type

Feature	Result
Variable type	character
Number of missing obs.	14 (4.62 %)
Number of unique values	4
Mode	"typical angina"



resting_bp

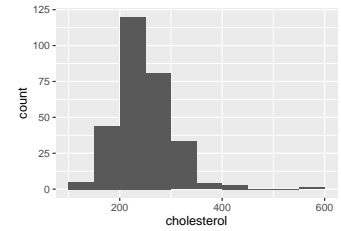
Feature	Result
Variable type	numeric
Number of missing obs.	15 (4.95 %)
Number of unique values	49
Median	130
1st and 3rd quartiles	120; 140
Min. and max.	94; 200



- Note that the following possible outlier values were detected: "94", "178", "180", "192", "200".

cholesterol

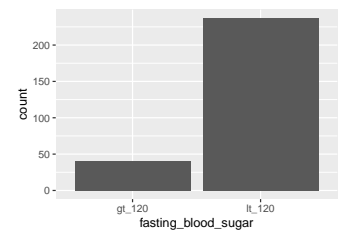
Feature	Result
Variable type	numeric
Number of missing obs.	12 (3.96 %)
Number of unique values	150
Median	242
1st and 3rd quartiles	211; 275
Min. and max.	126; 564



- Note that the following possible outlier values were detected: "126", "131", "394", "407", "409", "417", "564".

fasting_blood_sugar

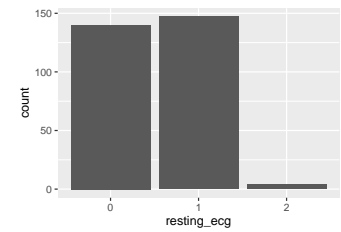
Feature	Result
Variable type	character
Number of missing obs.	25 (8.25 %)
Number of unique values	2
Mode	"lt_120"



resting_ecg

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

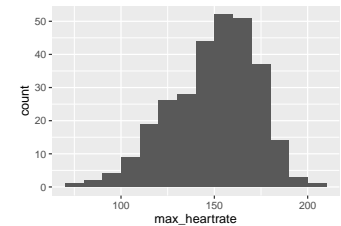
Feature	Result
Variable type	numeric
Number of missing obs.	12 (3.96 %)
Number of unique values	3
Mode	"1"
Reference category	0



- Note that the following levels have at most five observations: "2".

max_hearttrate

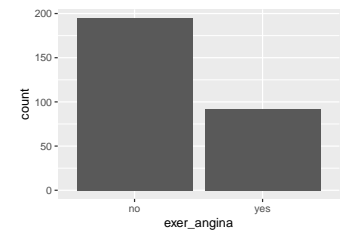
Feature	Result
Variable type	numeric
Number of missing obs.	12 (3.96 %)
Number of unique values	90
Median	153
1st and 3rd quartiles	133; 166
Min. and max.	71; 202



- Note that the following possible outlier values were detected: "194", "195", "202".

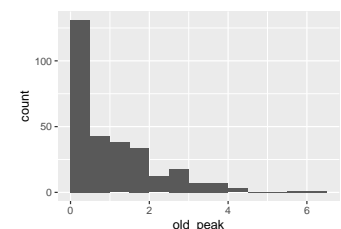
exer_angina

Feature	Result
Variable type	character
Number of missing obs.	16 (5.28 %)
Number of unique values	2
Mode	"no"



old_peak

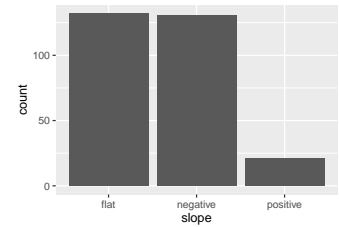
Feature	Result
Variable type	numeric
Number of missing obs.	8 (2.64 %)
Number of unique values	40
Median	0.8
1st and 3rd quartiles	0; 1.6
Min. and max.	0; 6.2



- Note that the following possible outlier values were detected: "6.2".

slope

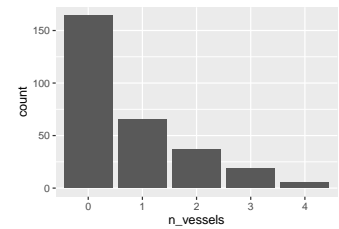
Feature	Result
Variable type	character
Number of missing obs.	19 (6.27 %)
Number of unique values	3
Mode	"flat"



n_vessels

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

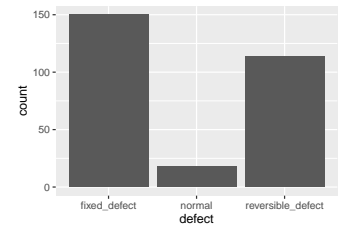
Feature	Result
Variable type	numeric
Number of missing obs.	13 (4.29 %)
Number of unique values	5
Mode	"0"
Reference category	0



- Note that the following levels have at most five observations: "4".

defect

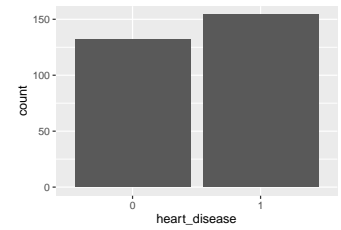
Feature	Result
Variable type	character
Number of missing obs.	21 (6.93 %)
Number of unique values	3
Mode	"fixed_defect"



heart__disease

- 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.	17 (5.61 %)
Number of unique values	2
Mode	"1"
Reference category	0



Report generation information:

- Created by: Michael Clark (username: micl).
- Report creation time: Thu Jul 02 2020 10:16:58
- Report was run from directory: /Users/micl/Documents/Stats/Repositories/Workshops/workshops-2020/explorator
- dataMaid v1.4.0 [Pkg: 2019-12-10 from CRAN (R 4.0.0)]
- R version 4.0.0 (2020-04-24).
- Platform: x86_64-apple-darwin17.0 (64-bit)(macOS Catalina 10.15.2).
- Function call: `makeDataReport(data = hd, file = "other_docs/datamaid_report.Rmd", replace = TRUE, ouput = "html")`