

testData

Autogenerated data summary from cleanR

2016-12-29

Part 1

Data cleaning summary

The dataset examined has the following dimensions:

Feature	Result
Number of rows	15
Number of variables	15

Checks performed

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

	character	factor	labelled	numeric	integer	logic
Identify miscoded missing values	×	×	×	×	×	
Identify prefixed and suffixed whitespace	×	×	×			
Identify levels with < 6 obs.	×	×	×			
Identify case issues	×	×	×			
Identify misclassified numeric or integer variables	×	×	×			
Identify non-suffixed nor -prefixed colons	×	×	×			
Identify outliers				×	×	

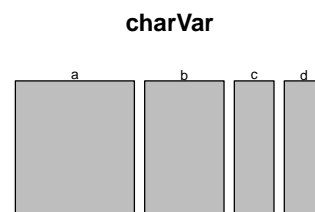
Please note that all numerical values in the following have been rounded to 2 decimals.

Part 2

Variable list

charVar

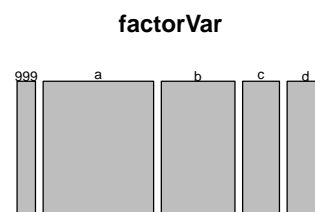
Feature	Result
Variable type	character
Number of missing obs.	1 (6.67 %)
Number of unique values	4
Mode	"a"
No. zeros	0



- Note that the following levels have at most five observations: "b", "c", "d".

factorVar

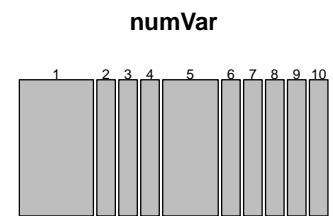
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	5
Mode	"a"
No. zeros	0



- The following suspected missing value codes enter as regular values: "999".
 - Note that the following levels have at most five observations: "999", "b", "c", "d".
-

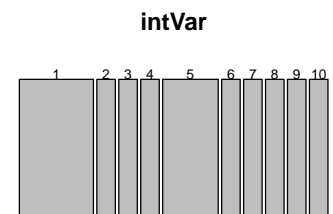
numVar

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	10
Median	5
1st and 3rd quartiles	1.5; 6.5
Min. and max.	1; 10
No. zeros	0



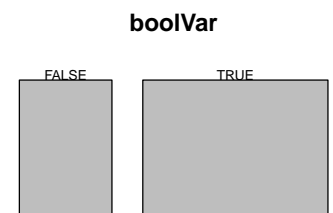
intVar

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	10
Median	5
1st and 3rd quartiles	1.5; 6.5
Min. and max.	1; 10
No. zeros	0



boolVar

Feature	Result
Variable type	logical
Number of missing obs.	3 (20 %)
Number of unique values	2
Mode	"TRUE"



keyVar

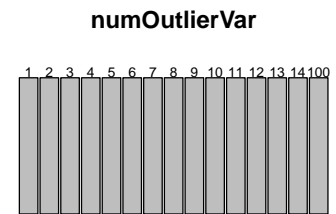
- The variable is a key (distinct values for each observation).

emptyVar

- The variable only takes one (non-missing) value: "1". The variable contains 0 % missing observations.

numOutlierVar

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	15
Median	8
1st and 3rd quartiles	4.5; 11.5
Min. and max.	1; 100
No. zeros	0

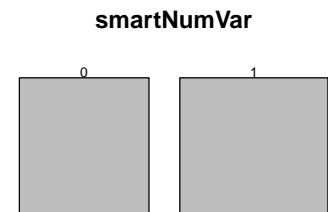


- Note that the following possible outlier values were detected: "100".

smartNumVar

- 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	"1"
No. zeros	7



cprVar

- Warning: This variable seems to contain ID codes!

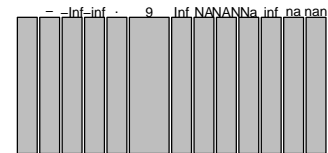
cprKeyVar

- The variable is a key (distinct values for each observation).
 - Warning: This variable seems to contain ID codes!
-

miscodedMissingVar

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	14
Mode	"9"
No. zeros	0

miscodedMissingVar

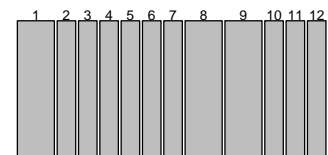


- The following suspected missing value codes enter as regular values: "", "-", "-Inf", "-inf", ".", "9", "Inf", "NA", "NaN", "Na", "NaN", "inf", "na", "nan".
- Note that the following levels have at most five observations: "", "-", "-Inf", "-inf", ".", "9", "Inf", "NA", "NaN", "Na", "NaN", "inf", "na".
- Note that there might be case problems with the following levels: "-Inf", "-inf", "Inf", "NA", "NaN", "Na", "NaN", "inf", "na", "nan".

misclassifiedNumVar

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	12
Mode	"1"
No. zeros	0

misclassifiedNumVar

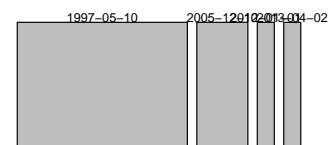


- Note that the following levels have at most five observations: "1", "10", "11", "12", "2", "3", "4", "5", "6", "7", "8", "9".
- Note: The variable consists exclusively of numbers and takes a lot of different values. Is it perhaps a misclassified numeric variable?

dateVar

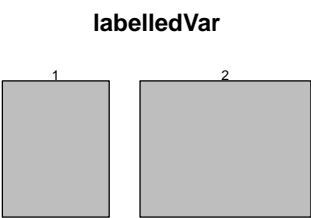
Feature	Result
Variable type	Date
Number of missing obs.	0 (0 %)
Number of unique values	4
Mode	"1997-05-10"
Min. and max.	1997-05-10; 2013-04-02
1st and 3rd quartiles	1997-05-10; 2005-12-10

dateVar



labelledVar

Feature	Result
Variable type	labelled
Number of missing obs.	2 (13.33 %)
Number of unique values	2
Mode	“2”
No. zeros	0



This report was created by cleanR v0.6.2.