testData

Autogenerated data summary from dataMaid

2019-09-27 04:09:53

# Data report overview

The dataset examined has the following dimensions:

|  |  |
| --- | --- |
| Feature | Result |
| Number of observations | 15 |
| 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.

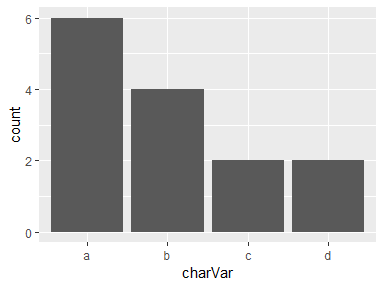
# Summary table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Variable class | # unique values | Missing observations | Any problems? |
| [charVar](#charvar) | character | 5 | 6.67 % | × |
| [factorVar](#factorvar) | factor | 5 | 0.00 % | × |
| [numVar](#numvar) | numeric | 10 | 0.00 % |  |
| [intVar](#intvar) | integer | 10 | 0.00 % |  |
| [boolVar](#boolvar) | logical | 3 | 20.00 % |  |
| [keyVar](#keyvar) | character | 15 | 0.00 % | × |
| [emptyVar](#emptyvar) | numeric | 1 | 0.00 % | × |
| [numOutlierVar](#numoutliervar) | numeric | 15 | 0.00 % | × |
| [smartNumVar](#smartnumvar) | numeric | 2 | 0.00 % |  |
| [cprVar](#cprvar) | character | 5 | 0.00 % | × |
| [cprKeyVar](#cprkeyvar) | character | 15 | 0.00 % | × |
| [miscodedMissingVar](#miscodedmissingvar) | character | 14 | 0.00 % | × |
| [misclassifiedNumVar](#misclassifiednumvar) | factor | 12 | 0.00 % | × |
| [dateVar](#datevar) | Date | 4 | 0.00 % |  |
| [labelledVar](#labelledvar) | labelled | 3 | 13.33 % | × |

# Variable list

## charVar

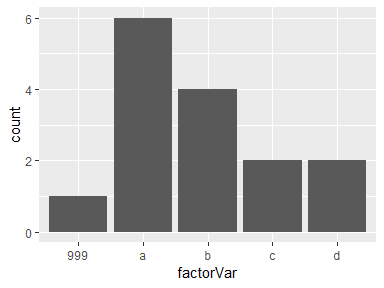
|  |  |
| --- | --- |
| Feature | Result |
| Variable type | character |
| Number of missing obs. | 1 (6.67 %) |
| Number of unique values | 4 |
| Mode | “a” |



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

## factorVar

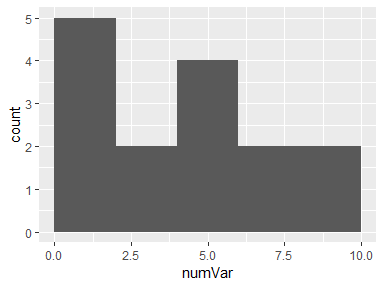
|  |  |
| --- | --- |
| Feature | Result |
| Variable type | factor |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 5 |
| Mode | “a” |
| Reference category | 999 |



* 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".

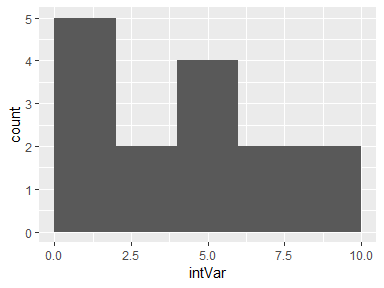
## numVar

|  |  |
| --- | --- |
| 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 |



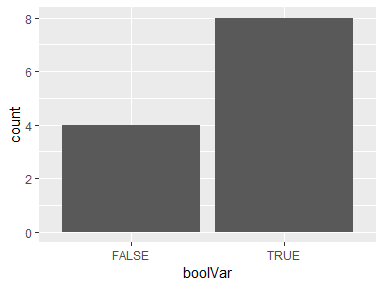
## intVar

|  |  |
| --- | --- |
| 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 |



## boolVar

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



## keyVar

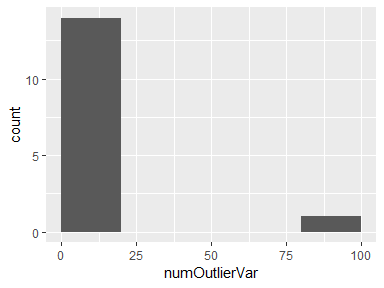
* The variable is a key (distinct values for each observation).

## 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 |

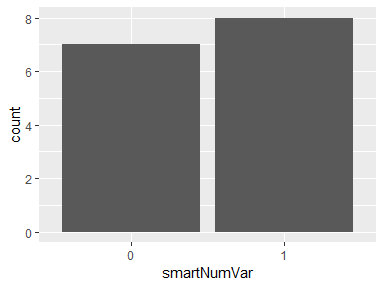


* Note that the following possible outlier values were detected: "100".

## smartNumVar

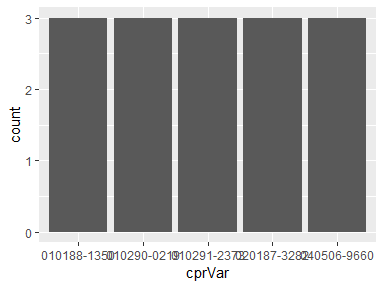
* Note that this variable is treated as a factor variable below, as it only takes a few unique values.

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



## cprVar

|  |  |
| --- | --- |
| Feature | Result |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 5 |
| Mode | “010188-1350” |



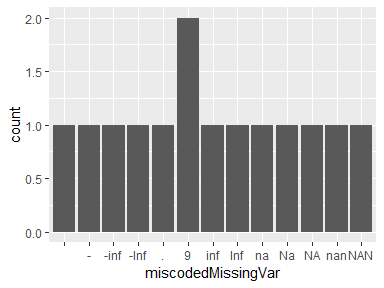
* Note that the following levels have at most five observations: "010188-1350", "010290-0219", "010291-2373", "020187-3282", "040506-9660".

## cprKeyVar

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

## miscodedMissingVar

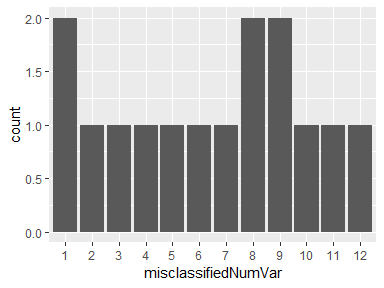
|  |  |
| --- | --- |
| Feature | Result |
| Variable type | character |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 14 |
| Mode | “9” |



* The following suspected missing value codes enter as regular values: "", "-", "-inf", "-Inf", ".", "9", "inf", "Inf", "na", "Na" (4 additional values omitted).
* Note that the following levels have at most five observations: "", "-", "-inf", "-Inf", ".", "9", "inf", "Inf", "na", "Na" (4 additional values omitted).
* Note that there might be case problems with the following levels: "-inf", "-Inf", "inf", "Inf", "na", "Na", "NA", "nan", "NaN", "NAN".

## misclassifiedNumVar

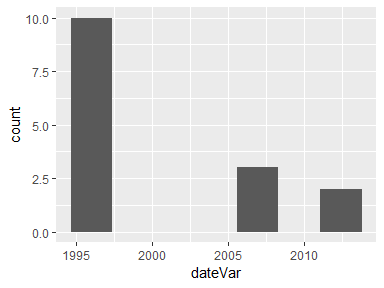
|  |  |
| --- | --- |
| Feature | Result |
| Variable type | factor |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 12 |
| Mode | “1” |
| Reference category | 1 |



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

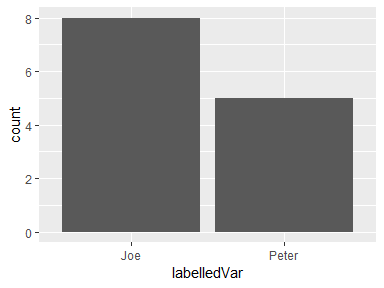
## dateVar

|  |  |
| --- | --- |
| 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 |



## labelledVar

|  |  |
| --- | --- |
| Feature | Result |
| Variable type | labelled |
| Number of missing obs. | 2 (13.33 %) |
| Number of unique values | 2 |
| Mode | “Joe” |



* Note that the following levels have at most five observations: "Peter".

Report generation information:

* Created by Nikhil Gupta (username: Nikhil).
* Report creation time: Fri Sep 27 2019 04:09:54
* Report was run from directory: C:/Users/Nikhil/Google Drive Work/gre/evaluations/r/eda
* dataMaid v1.3.2 [Pkg: 2019-07-27 from CRAN (R 3.5.3)]
* R version 3.5.1 (2018-07-02).
* Platform: x86\_64-w64-mingw32/x64 (64-bit)(Windows >= 8 x64 (build 9200)).
* Function call: makeDataReport(data = testData, replace = TRUE)