

testData

Autogenerated data summary from cleanR

2016-12-01

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 | logical |
|---|-----------|--------|----------|---------|---------|---------|
| 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 | | | | × | × | |

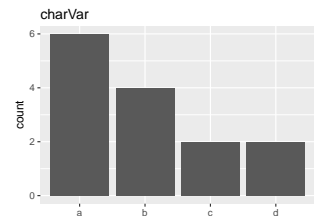
Please note that all numerical values in the following have been rounded to 2 decimals.

Part 2

Variable list

charVar

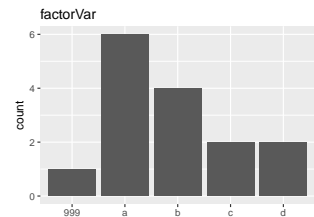
| Feature | Result |
|-------------------------|------------|
| Variable type | character |
| Number of missing obs. | 1 (6.67 %) |
| Number of unique values | 4 |
| Mode | "a" |



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

factorVar

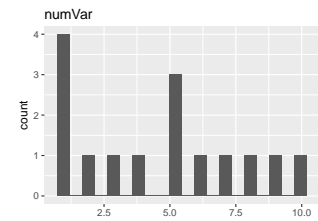
| Feature | Result |
|-------------------------|---------|
| Variable type | factor |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 5 |
| Mode | "a" |



- 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".

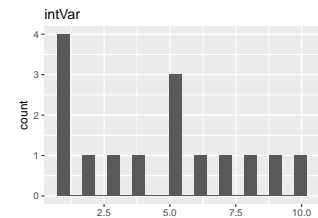
numVar

| 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 |



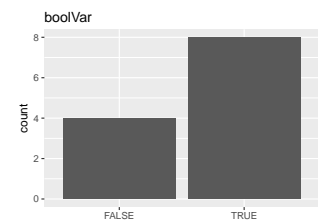
intVar

| 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

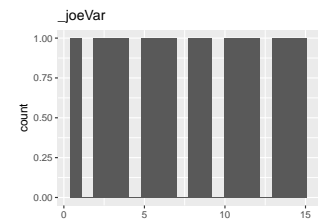
- The variable is a key (distinct values for each observation).

emptyVar

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

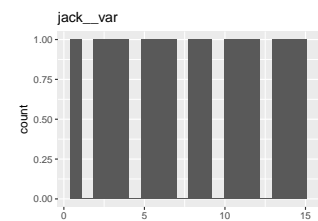
`_joeVar`

| Feature | Result |
|-------------------------|-----------|
| Variable type | integer |
| 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; 15 |



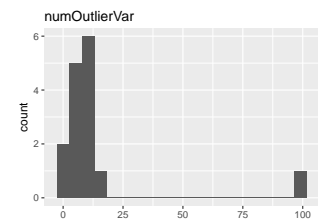
`jack__var`

| Feature | Result |
|-------------------------|-----------|
| Variable type | integer |
| 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; 15 |



`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 |

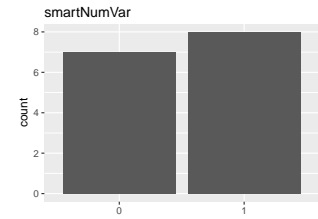


- `identifyOutliers`
- Note that the following possible outlier values were detected: "100".

smartNumVar

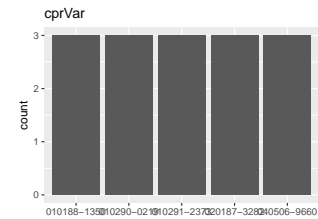
- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

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



cprVar

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



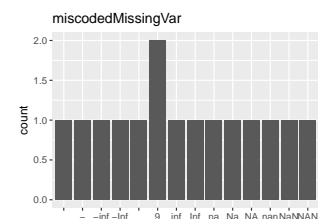
- Note that the following levels have at most five observations: "010188-1350", "010290-0219", "010291-2373", "020187-3282", "040506-9660".

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", "NA", "nan", "NaN", "NAN".
- Note that the following levels have at most five observations: "", "-", "-inf", "-Inf", ".", "9", "inf", "Inf", "na", "Na", "NA", "nan", "NaN".
- Note that there might be case problems with the following levels: "-inf", "-Inf", "inf", "Inf", "na", "Na", "NA", "nan", "NaN", "NAN".

misclassifiedNumVar

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