PMIDs

Autogenerated data summary from dataMaid

2020-05-08 10:01:31

Part 1

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	14488
Number of variables	5

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.

Part 2

Summary table

	Variable class	# unique values	Missing observations	Any problems?
Pmid	numeric	14488	0.00 %	×
Journal	character	2026	0.00 %	×
Year	numeric	42	2.94 %	×
Month	numeric	13	2.94 %	
Pmcid	character	7375	49.10 %	×

Part 3

Variable list

Pmid

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	14488
Median	25859034
1st and 3rd quartiles	17319189.25; 29207036.75
Min. and max.	1284342; 32106637
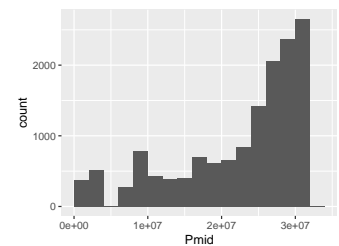
- Note that the following possible outlier values were detected: "31738324", "31740409", "31740932", "31741443", "31742234", "31743375", "31743859", "31743899", "31744045", "31744266" (250 additional values omitted).

Journal

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	2026
Mode	"East African medical journal"
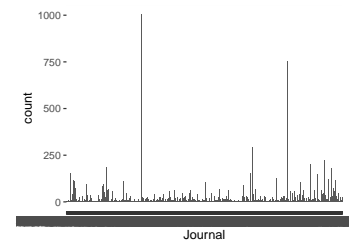
- Note that the following levels have at most five observations: "3 Biotech", "AAS open research", "Academic medicine : journal of the Association of American Medical Colleges", "Academic pediatrics", "Academic psychiatry : the journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry", "Accident; analysis and prevention", "ACS applied materials & interfaces", "ACS infectious diseases", "ACS synthetic biology", "Acta anatomica" (1612 additional values omitted).

Year

Feature	Result
Variable type	numeric
Number of missing obs.	426 (2.94 %)
Number of unique values	41
Median	2015
1st and 3rd quartiles	2006; 2017
Min. and max.	1980; 2020
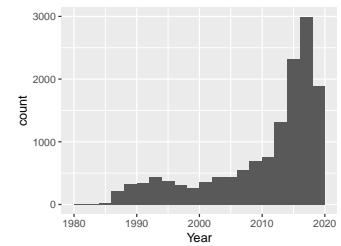
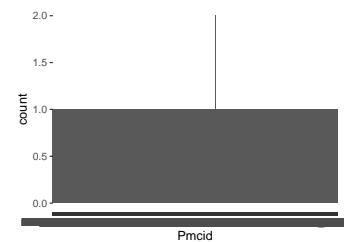- Note that the following possible outlier values were detected: "2019", "2020".

Month

Feature	Result
Variable type	numeric
Number of missing obs.	426 (2.94 %)
Number of unique values	12
Median	5
1st and 3rd quartiles	1; 9
Min. and max.	1; 12
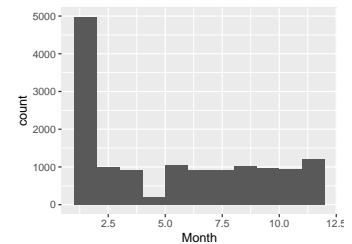
Pmcid

Feature	Result
Variable type	character
Number of missing obs.	7113 (49.1 %)
Number of unique values	7374
Mode	"PMC5521136"

- Note that the following levels have at most five observations: "PMC1029778", "PMC1041405", "PMC1042122", "PMC105474", "PMC108135", "PMC1081565", "PMC1088468", "PMC1117550", "PMC1136262", "PMC1140744" (7364 additional values omitted).

Report generation information:

- Created by: Could not determine from system (username: `kamonde`).

- Report creation time: Fri May 08 2020 10:01:32

- Report was run from directory: `/home/kamonde`

- dataMaid v1.4.0 [Pkg: 2019-12-10 from CRAN (R 3.6.3)]

- R version 3.6.3 (2020-02-29).

- Platform: x86_64-pc-linux-gnu (64-bit)(Ubuntu 20.04 LTS).

- Function call: `makeDataReport(data = PMIDs)`