Dirty president data

Autogenerated data summary from dataMaid

2017-10-20 11:48:27

Part 1

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	46
Number of variables	11

Checks performed

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

	character	factor	labelled	numeric	integer	logical	Date
Identify miscoded missing values	×	×	×	×	×		
Identify prefixed and suffixed whitespace	×	×	×				
Identify case issues	×	×	×				
Identify misclassified numeric or integer variables	×	×	×				
Identify levels with < 6 obs.		×	×				
Identify outliers				×	×		×

Non-supported variable types were set to be handled in the following way:

- Name is treated as character

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

Part 2

Summary table

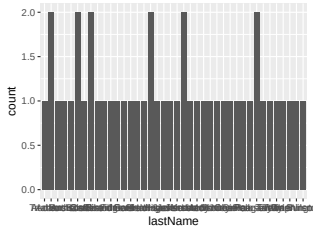
	Variable class	# unique values	Missing observations	Any problems?
lastName	Name	40	0.00 %	×
firstName	Name	31	0.00 %	×
orderOfPresidency	factor	46	0.00 %	×
birthday	Date	45	0.00 %	×
stateOfBirth	character	23	0.00 %	×
assassinationAttempt	numeric	2	0.00 %	
sex	factor	1	0.00 %	×
ethnicity	factor	2	0.00 %	×
presidencyYears	numeric	11	4.35 %	×
ageAtInauguration	character	23	0.00 %	×
favoriteNumber	complex	11	0.00 %	×

Part 3

Variable list

lastName

Feature	Result
Variable type	Name
Number of missing obs.	0 (0 %)
Number of unique values	40
Mode	“Adams”

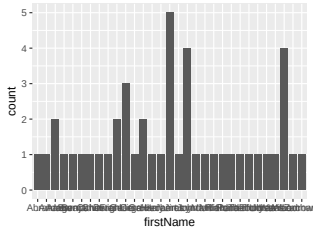


- The following values appear with prefixed or suffixed white space: " Truman".

---

firstName

Feature	Result
Variable type	Name
Number of missing obs.	0 (0 %)
Number of unique values	31
Mode	“James”



- The following suspected missing value codes enter as regular values: ".".

---

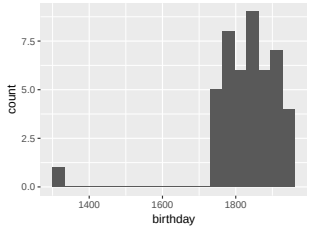
orderOfPresidency

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

---

birthday

Feature	Result
Variable type	Date
Number of missing obs.	0 (0 %)
Number of unique values	45
Mode	“1837-03-18”
Min. and max.	1300-03-01; 1961-08-04
1st and 3rd quartiles	1790-03-29; 1890-10-14

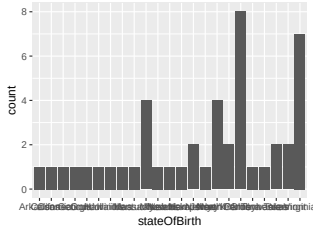


- Note that the following possible outlier values were detected: "1300-03-01".

---

stateOfBirth

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	23
Mode	“Ohio”



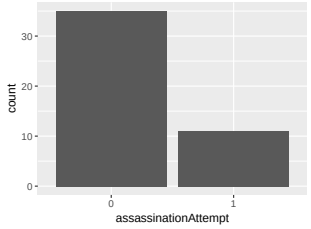
- Note that there might be case problems with the following levels: "New york", "New York".

---

assassinationAttempt

- 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	“0”



---

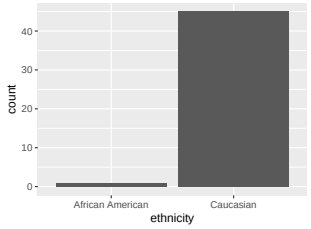
sex

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

---

ethnicity

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	“Caucasian”

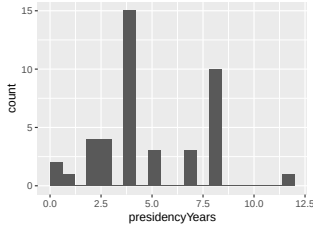


- Note that the following levels have at most five observations: "African American".

---

presidencyYears

Feature	Result
Variable type	numeric
Number of missing obs.	2 (4.35 %)
Number of unique values	10
Median	4
1st and 3rd quartiles	3.75; 8
Min. and max.	0; Inf

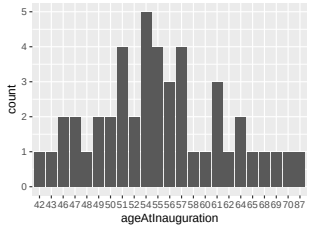


- The following suspected missing value codes enter as regular values: "Inf".
- Note that the following possible outlier values were detected: "0", "1", "Inf".

---

ageAtInauguration

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	23
Mode	“54”



- Note: The variable consists exclusively of numbers and takes a lot of different values. Is it perhaps a misclassified numeric variable?

---

favoriteNumber

- The variable has class complex which is not supported by dataMaid.

Report generation information:

- Created by Anne Helby Petersen.
- Report creation time: fr okt 20 2017 11:48:26
- dataMaid v0.9.7.9000 [Pkg: 2017-10-19 from local (ekstroem/dataMaid@NA)]
- R version 3.3.2 (2016-10-31).
- Platform: x86_64-w64-mingw32/x64 (64-bit)(Windows 7 x64 (build 7601) Service Pack 1).
- Function call: `makeDataReport(data = presidentData, replace = TRUE, checks = setChecks(character = defaultCharacterChecks(remove = "identifyLoners")), reportTitle = "Dirty president data", treatXasY = list(Name = "character"))`