

Dirty president data - cleaned  
Autogenerated data summary from dataMaid

2020-03-31 13:32:31

## 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	haven 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
- complex is treated as numeric

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

## Part 2

# Summary table

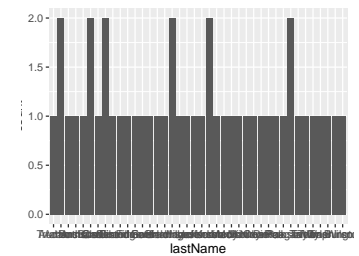
	Variable class	# unique values	Missing observations	Any problems?
lastName	Name	40	0.00 %	×
firstName	Name	31	0.00 %	×
orderOfPresidency	factor	46	0.00 %	×
birthday	Date	45	0.00 %	×
stateOfBirth	character	23	0.00 %	×
assassinationAttempt	numeric	2	0.00 %	
sex	factor	1	0.00 %	×
ethnicity	factor	2	0.00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

## Part 3

# Variable list

### lastName

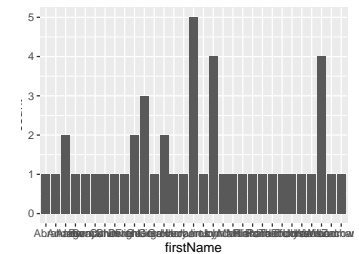
Feature	Result
Variable type	Name
Number of missing obs.	0 (0 %)
Number of unique values	40
Mode	“Adams”



- 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”



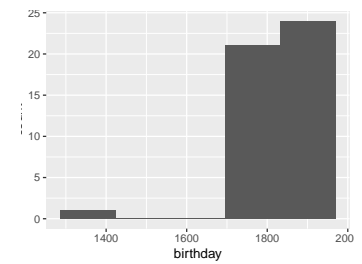
- The following suspected missing value codes enter as regular values: ".".

### orderOfPresidency

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

## birthday

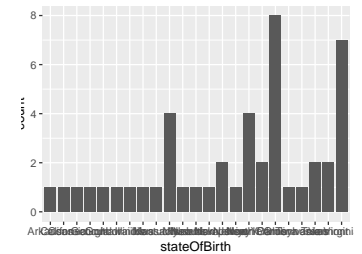
Feature	Result
Variable type	Date
Number of missing obs.	0 (0 %)
Number of unique values	45
Mode	"1837-03-18"
Min. and max.	1300-03-01; 1961-08-04
1st and 3rd quartiles	1790-03-29; 1890-10-14



- 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"

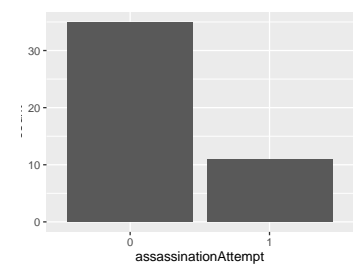


- Note that there might be case problems with the following levels: "New york", "New York".

## 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"
Reference category	0

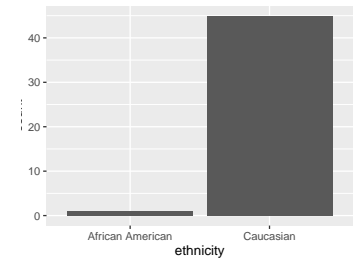


## sex

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

## ethnicity

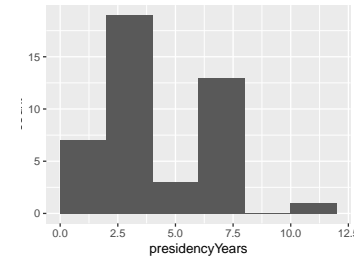
Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"Caucasian"
Reference category	African American



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

## presidencyYears

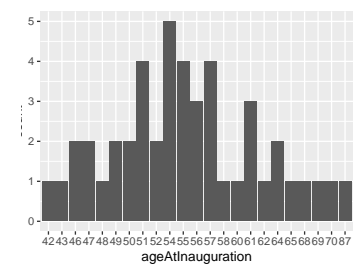
Feature	Result
Variable type	numeric
Number of missing obs.	2 (4.35 %)
Number of unique values	10
Median	4
1st and 3rd quartiles	3.75; 8
Min. and max.	0; Inf



- The following suspected missing value codes enter as regular values: "Inf".
- Note that the following possible outlier values were detected: "0", "1", "2", "Inf".

## ageAtInauguration

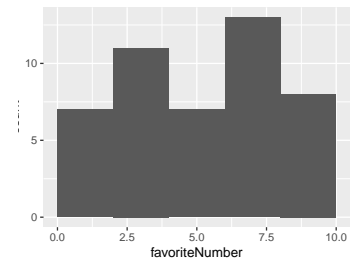
Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	23
Mode	"54"



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

## favoriteNumber

Feature	Result
Variable type	complex
Number of missing obs.	0 (0 %)
Number of unique values	11
Median	6
1st and 3rd quartiles	3.25; 8
Min. and max.	0; 10



---

### Report generation information:

- Created by: Piotr Piątysek (username: piotrp).
- Report creation time: Tue Mar 31 2020 13:32:32
- Report was run from directory: /home/piotrp/projects/warszt/pd1/pd1-pp
- dataMaid v1.4.0 [Pkg: 2019-12-10 from CRAN (R 3.6.2)]
- R version 3.6.2 (2019-12-12).
- Platform: x86\_64-suse-linux-gnu (64-bit)(openSUSE Leap 15.0).
- Function call: `makeDataReport(data = presidentData, replace = TRUE, vol = "_cleaned", checks = setChecks(character = defaultCharacterChecks(remove = "identifyLoners")), reportTitle = "Dirty president data - cleaned", treatXasY = list(Name = "character", complex = "numeric"))`