

dn_sub

Autogenerated data summary from dataMaid

2020-10-19 12:36:05

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	19193
Number of variables	16

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.

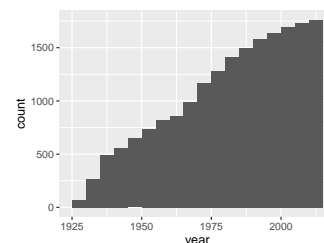
Summary table

	Variable class	# unique values	Missing observations	Any problems?
year	numeric	87	0.00 %	
gnis_id	character	363	0.00 %	×
geometry	character	2	0.00 %	
metadata	character	315	12.46 %	×
number_of_records	numeric	1	0.00 %	×
parkname	character	323	10.15 %	×
region	character	8	0.00 %	
state	factor	50	0.00 %	
unit_code	character	363	0.00 %	×
unit_name	character	371	0.00 %	×
unit_type	character	28	0.00 %	×
visitors	numeric	17100	0.00 %	
pop	numeric	3455	0.00 %	×
gas_current	numeric	52	0.00 %	×
gas_constant	numeric	70	0.00 %	×
state_int	integer	50	0.00 %	

Variable list

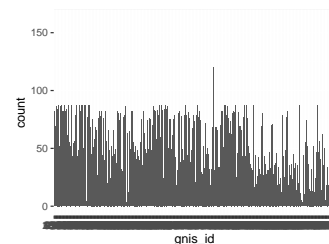
year

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	87
Median	1986
1st and 3rd quartiles	1967; 2002
Min. and max.	1929; 2015



gnis_id

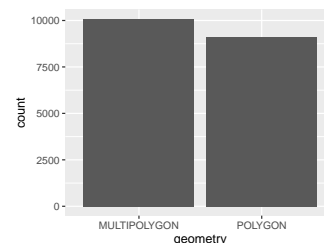
Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	363
Mode	"2733443"



- Note that the following levels have at most five observations: "1888944", "1967105", "2002666", "2026588", "255952", "2733433", "2733434", "2736075", "2738287", "2754548" (3 additional values omitted).
- Note: The variable consists exclusively of numbers and takes a lot of different values. Is it perhaps a misclassified numeric variable?

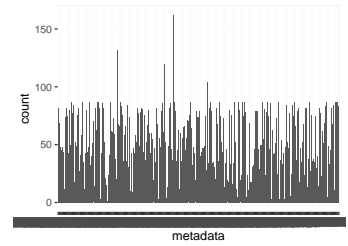
geometry

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	2
Mode	"MULTIPOLYGON"



metadata

Feature	Result
Variable type	character
Number of missing obs.	2391 (12.46 %)
Number of unique values	314
Mode	"http://nrdana.nps.gov/programs/Lands/GRSA_METADATA.xml"



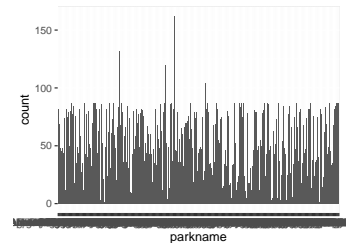
- Note that the following levels have at most five observations: "http://nrdana.nps.gov/programs/Lands/CECH_metadata.xml/", "http://nrdana.nps.gov/programs/Lands/CHYO_METADATA.xml/", "http://nrdana.nps.gov/programs/Lands/GATE_metadata.xml/", "http://nrdana.nps.gov/programs/Lands/GOGA_metadata.xml/", "http://nrdana.nps.gov/programs/Lands/MLKM_metadata.xml/", "http://nrdana.nps.gov/programs/Lands/NACA_metadata.xml/", "http://nrdana.nps.gov/programs/Lands/POCH_metadata.xml/", "http://nrdana.nps.gov/programs/Lands/RIRA_metadata.xml/", "http://nrdana.nps.gov/programs/Lands/RORI_metadata.xml/", "http://nrdana.nps.gov/programs/Lands/SACR_metadata.xml/" (2 additional values omitted).

number_of_records

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

parkname

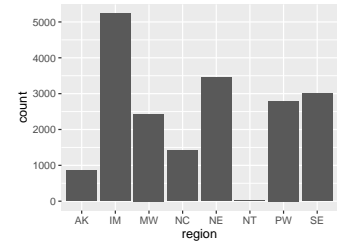
Feature	Result
Variable type	character
Number of missing obs.	1949 (10.15 %)
Number of unique values	322
Mode	"Great Sand Dunes"



- Note that the following levels have at most five observations: "Cesar E. Chavez", "Charles Young Buffalo Soldier", "Gateway", "Golden Gate", "Martin Luther King Jr Memorial", "National Capital Parks", "Port Chicago Naval Magazine", "President William Jefferson Clinton Birthplace Home", "River Raisin", "Rosie the Riveter/World War II Home Front" (2 additional values omitted).

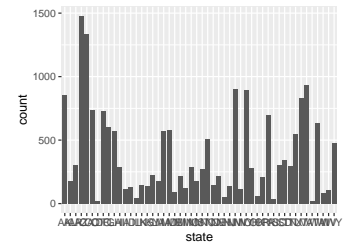
region

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	8
Mode	“IM”



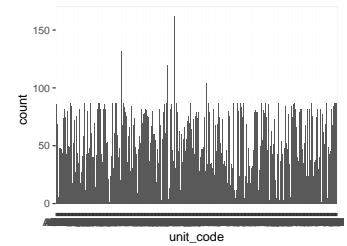
state

Feature	Result
Variable type	factor
Number of missing obs.	0 (0 %)
Number of unique values	50
Mode	“AZ”
Reference category	AK



unit_code

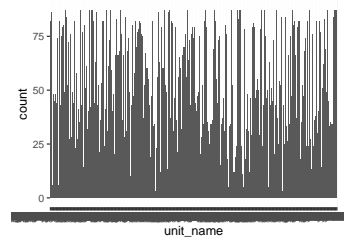
Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	363
Mode	“GRSA”



- Note that the following levels have at most five observations: "CECH", "CHYO", "GATE", "GOGA", "MISS", "MLKM", "NACA", "POCH", "RIRA", "RORI" (3 additional values omitted).

unit_name

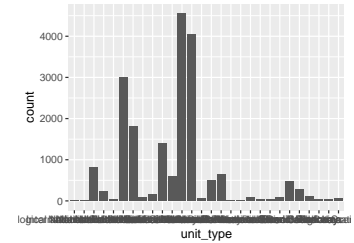
Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	371
Mode	“Aztec Ruins National Monument”



- Note that the following levels have at most five observations: "Casar E. Chpvez National Monument", "Charles Young Buffalo Soldiers National Monument", "Gateway National Recreation Area", "Golden Gate National Recreation Area", "Martin Luther King, Jr., Memorial", "Mississippi National River and Recreation Area", "National Capital Parks", "Port Chicago Naval Magazine National Memorial", "President William Jefferson Clinton Birthplace Home National Historic Site", "River Raisin National Battlefield Park" (3 additional values omitted).

unit__type

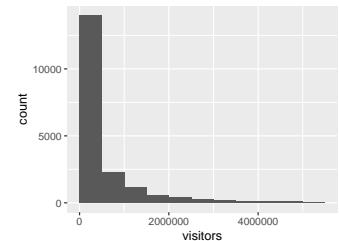
Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	28
Mode	"National Monument"



- Note that the following levels have at most five observations: "International Historic Site".

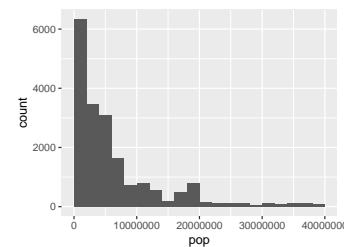
visitors

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	17100
Median	158600
1st and 3rd quartiles	43300; 557217
Min. and max.	613; 5192600



pop

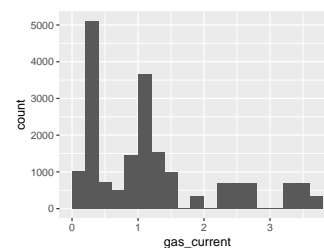
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	3455
Median	3867333
1st and 3rd quartiles	1309400; 7622159
Min. and max.	98000; 39144818



- Note that the following possible outlier values were detected: "36961229", "37336011", "37701901", "38062780", "38431393", "38802500", "39144818".

gas__current

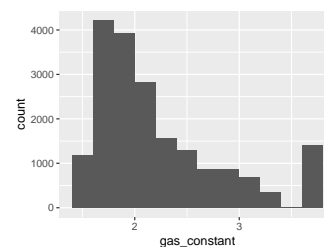
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	52
Median	1.11
1st and 3rd quartiles	0.33; 1.51
Min. and max.	0.17; 3.64



- Note that the following possible outlier values were detected: "2.59", "2.79", "2.8", "3.27", "3.37", "3.53", "3.64".

gas__constant

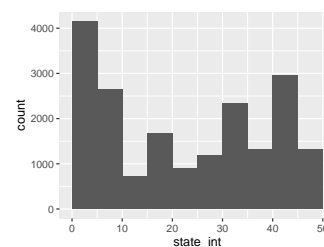
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	70
Median	2.01
1st and 3rd quartiles	1.76; 2.58
Min. and max.	1.47; 3.8



- Note that the following possible outlier values were detected: "1.47".

state__int

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	50
Median	23
1st and 3rd quartiles	6; 38
Min. and max.	1; 50



Report generation information:

- Created by: error: invalid key: Spyrison/.gitconfig (username: Nicholas Spyrison).
- Report creation time: Mon Oct 19 2020 12:36:07
- Report was run from directory: C:/Users/Nicholas Spyrison/Documents/R/tidy_tuesday_national_parks
- dataMaid v1.4.0 [Pkg: 2019-12-10 from CRAN (R 4.0.3)]

- R version 4.0.2 (2020-06-22).
- Platform: x86_64-w64-mingw32/x64 (64-bit)(Windows 10 x64 (build 18362)).
- Function call: `dataMaid::makeDataReport(data = dn_sub)`