

project_macro

Autogenerated data summary from dataMaid

2017-08-11 21:24:37

Part 1

Data cleaning summary

The dataset examined has the following dimensions:

Feature	Result
Number of rows	8398
Number of variables	59

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 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?
iso2c	character	165	0.00 %	×
country	character	165	0.00 %	×
year	numeric	56	0.00 %	
gdppc	numeric	7117	15.27 %	×
gdp_ppp	numeric	3817	54.56 %	×
pop_total	numeric	8283	1.30 %	×
pop_dens	numeric	8071	3.87 %	×
oil	numeric	3164	43.53 %	×
natural	numeric	6053	26.79 %	×
health_public	numeric	3081	63.31 %	×
fertility	numeric	4198	1.69 %	
infl	numeric	6822	18.78 %	×
trade	numeric	6862	18.24 %	×
military	numeric	5792	30.72 %	×
suicide	numeric	5400	1.43 %	×
life_exp	numeric	8128	1.85 %	
mobile_1000	numeric	856	75.36 %	×
internet_fix	numeric	1750	77.60 %	×
iso3c	factor	165	1.29 %	×
region	factor	8	1.29 %	
capital	factor	164	1.29 %	×
longitude	factor	165	1.29 %	×
latitude	factor	165	1.29 %	×
income	factor	7	1.29 %	×
lending	factor	5	1.29 %	
reason	numeric	1706	0.65 %	
common	numeric	1493	0.00 %	×
counter_arg	numeric	1793	0.24 %	×
elite_consul	numeric	1861	0.24 %	
engage_soc	numeric	1796	0.24 %	
clientelistic	numeric	1429	0.00 %	×
welfare_working	numeric	1269	0.00 %	×
pol_kill	numeric	1834	0.00 %	
free_tort	numeric	1731	0.00 %	
free_disc	numeric	1947	0.00 %	
rel_infl	numeric	1080	0.00 %	×

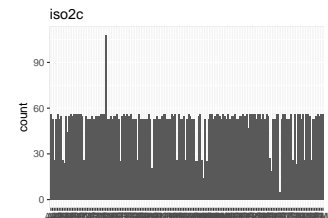
	Variable class	# unique values	Missing observations	Any problems?
gov_cens	numeric	1667	0.00 %	
equal_edu	numeric	1030	0.00 %	×
equal_health	numeric	1076	0.00 %	
duration	integer	209	19.16 %	
free_polity	numeric	148	25.78 %	
pl	integer	8	25.78 %	
rule_law	integer	18	82.22 %	
corruption	numeric	2426	70.33 %	×
gov_effect	numeric	2447	70.33 %	×
pol_stability	numeric	2464	70.28 %	
regime_bi	integer	3	31.35 %	
fh_status	integer	4	25.78 %	
pol_rights	integer	8	25.78 %	
civil_libs	integer	8	25.78 %	
polity_norm	integer	22	25.68 %	
regions	integer	11	23.95 %	×
gov_debt	numeric	3277	60.86 %	×
exports	numeric	6740	17.41 %	×
imports	numeric	6787	17.42 %	×
gini_income	integer	3	33.73 %	
conflict_int	integer	3	16.97 %	
conflict_dom	integer	3	31.75 %	
civil_war	integer	3	23.98 %	

Part 3

Variable list

iso2c

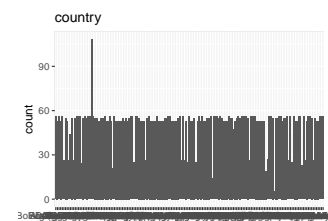
Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	165
Mode	“CV”



- The following suspected missing value codes enter as regular values: "NA".
- Note that the following levels have at most five observations: "SS".

country

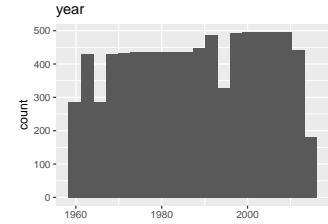
Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	165
Mode	“Cabo Verde”



- Note that the following levels have at most five observations: "South Sudan".

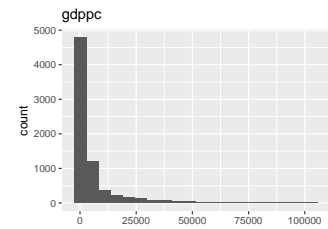
year

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	56
Median	1989
1st and 3rd quartiles	1974; 2002
Min. and max.	1960; 2015



gdppc

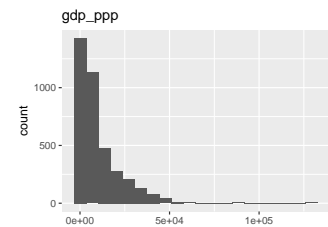
Feature	Result
Variable type	numeric
Number of missing obs.	1282 (15.27 %)
Number of unique values	7116
Median	1151.93
1st and 3rd quartiles	377.41; 4155.38
Min. and max.	37.52; 102910.44



- Note that the following possible outlier values were detected: "46437.07", "46510.39", "46511.6", "46523.27", "46530.91", "46586.65", "46596.34", "46610.06", "46657.06", "46810.33" (113 additional values omitted).

gdp_ppp

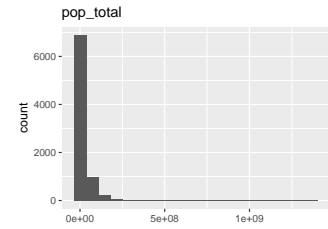
Feature	Result
Variable type	numeric
Number of missing obs.	4582 (54.56 %)
Number of unique values	3816
Median	5681.19
1st and 3rd quartiles	1961.66; 14270.79
Min. and max.	242; 129349.92



- Note that the following possible outlier values were detected: "90348.99", "91083.79", "97767.19", "100426.55", "108831.25", "111959.86", "113312.85", "115012.45", "122609.41", "126913.04" (4 additional values omitted).

pop_total

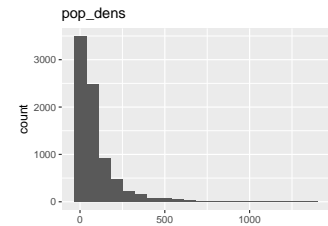
Feature	Result
Variable type	numeric
Number of missing obs.	109 (1.3 %)
Number of unique values	8282
Median	7585914
1st and 3rd quartiles	2716659; 21645535
Min. and max.	41700; 1364270000



- Note that the following possible outlier values were detected: "170560182", "171728917", "171829303", "172759243", "174184265", "175000916", "175287587", "176460502", "177750670", "177911533" (209 additional values omitted).

pop_dens

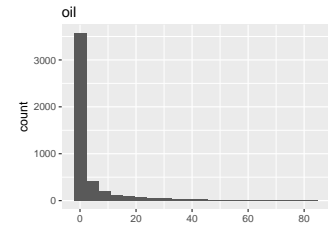
Feature	Result
Variable type	numeric
Number of missing obs.	325 (3.87 %)
Number of unique values	8070
Median	45.91
1st and 3rd quartiles	15.7; 110.75
Min. and max.	0.63; 1363.88



- Note that the following possible outlier values were detected: "652.56", "654.52", "654.85", "657", "659.03", "660.73", "672.24", "677.08", "678.64", "692.61" (60 additional values omitted).

oil

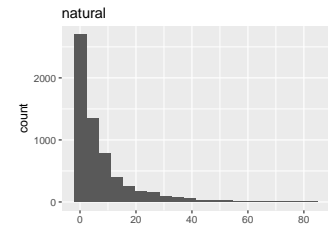
Feature	Result
Variable type	numeric
Number of missing obs.	3656 (43.53 %)
Number of unique values	3163
Median	0.07
1st and 3rd quartiles	0; 2.07
Min. and max.	0; 82.52



- Note that the following possible outlier values were detected: "54.52", "54.94", "55.44", "55.54", "57.4", "57.59", "58.28", "58.96", "59.21", "60.21" (12 additional values omitted).

natural

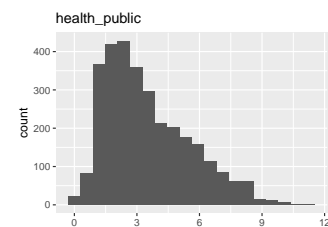
Feature	Result
Variable type	numeric
Number of missing obs.	2250 (26.79 %)
Number of unique values	6052
Median	2.97
1st and 3rd quartiles	0.63; 9.27
Min. and max.	0; 82.62



- Note that the following possible outlier values were detected: "77.37", "80.63", "80.92", "81.35", "82.49", "82.59", "82.62".

health_public

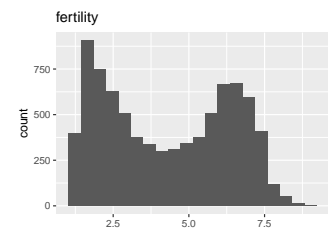
Feature	Result
Variable type	numeric
Number of missing obs.	5317 (63.31 %)
Number of unique values	3080
Median	2.98
1st and 3rd quartiles	1.93; 4.79
Min. and max.	0.04; 11.28



- Note that the following possible outlier values were detected: "0.04", "0.11", "0.14", "0.15", "0.16", "0.16", "0.17", "0.17", "0.18", "0.18" (23 additional values omitted).

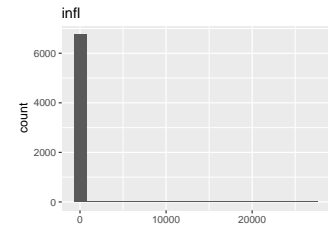
fertility

Feature	Result
Variable type	numeric
Number of missing obs.	142 (1.69 %)
Number of unique values	4197
Median	4.21
1st and 3rd quartiles	2.27; 6.24
Min. and max.	1.08; 8.87



infl

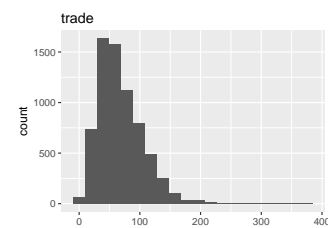
Feature	Result
Variable type	numeric
Number of missing obs.	1577 (18.78 %)
Number of unique values	6821
Median	6.56
1st and 3rd quartiles	2.66; 13.56
Min. and max.	-98.7; 26762.02



- Note that the following possible outlier values were detected: "-98.7", "-29.17", "-29.17", "-27.63", "-27.52", "-27.05", "-26.87", "-26.87", "-26.3", "-25.31" (645 additional values omitted).

trade

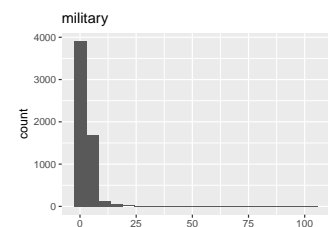
Feature	Result
Variable type	numeric
Number of missing obs.	1532 (18.24 %)
Number of unique values	6861
Median	61.02
1st and 3rd quartiles	41.6; 89.22
Min. and max.	0.02; 375.38



- Note that the following possible outlier values were detected: "0.02", "0.03", "0.1", "0.17", "0.18", "0.18", "0.2", "0.22", "0.25", "0.27" (65 additional values omitted).

military

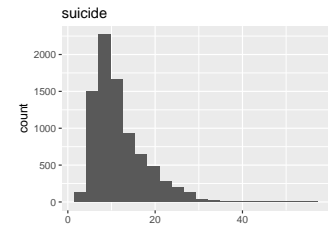
Feature	Result
Variable type	numeric
Number of missing obs.	2580 (30.72 %)
Number of unique values	5791
Median	1.99
1st and 3rd quartiles	1.27; 3.17
Min. and max.	0; 102.9



- Note that the following possible outlier values were detected: "0", "0", "0", "0", "0", "0.02", "0.02", "0.03", "0.05", "0.08" (248 additional values omitted).

suicide

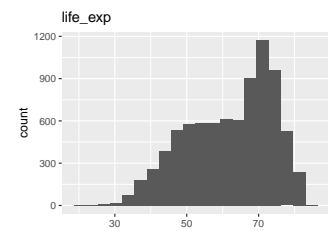
Feature	Result
Variable type	numeric
Number of missing obs.	120 (1.43 %)
Number of unique values	5399
Median	10
1st and 3rd quartiles	7.43; 14.34
Min. and max.	1.48; 54.44



- Note that the following possible outlier values were detected: "1.48", "1.48", "1.49", "1.5", "1.52", "1.55", "1.6", "1.65", "1.71", "1.78" (154 additional values omitted).

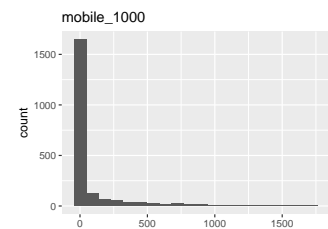
life_exp

Feature	Result
Variable type	numeric
Number of missing obs.	155 (1.85 %)
Number of unique values	8127
Median	64.32
1st and 3rd quartiles	52.56; 71.76
Min. and max.	19.27; 83.59



mobile_1000

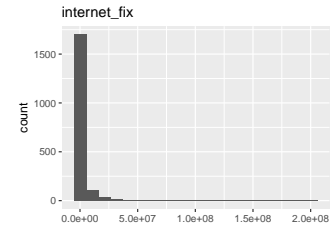
Feature	Result
Variable type	numeric
Number of missing obs.	6329 (75.36 %)
Number of unique values	855
Median	0
1st and 3rd quartiles	0; 15.25
Min. and max.	0; 1715.15



- Note that the following possible outlier values were detected: "478.09", "482.36", "483.78", "483.84", "484.09", "490.71", "491.68", "497.35", "497.36", "502.95" (117 additional values omitted).

internet_fix

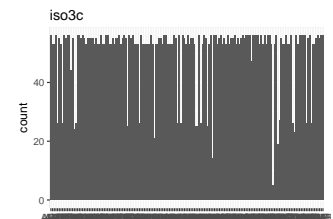
Feature	Result
Variable type	numeric
Number of missing obs.	6517 (77.6 %)
Number of unique values	1749
Median	99316
1st and 3rd quartiles	7458; 1090000
Min. and max.	0; 200483000



- Note that the following possible outlier values were detected: "22749000", "23301105", "23588919", "23745346", "23980000", "24939630", "24944252", "24950550", "24964600", "26161950" (37 additional values omitted).

iso3c

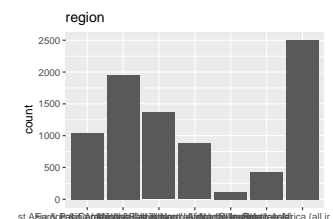
Feature	Result
Variable type	factor
Number of missing obs.	108 (1.29 %)
Number of unique values	164
Mode	"AFG"



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

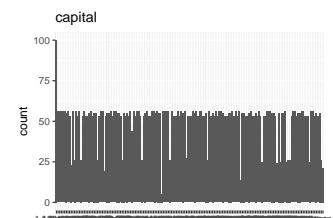
region

Feature	Result
Variable type	factor
Number of missing obs.	108 (1.29 %)
Number of unique values	7
Mode	"Sub-Saharan Africa (all income levels)"



capital

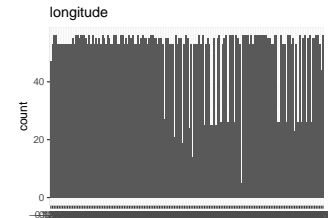
Feature	Result
Variable type	factor
Number of missing obs.	108 (1.29 %)
Number of unique values	163
Mode	" "



- The following suspected missing value codes enter as regular values: "".
- Note that the following levels have at most five observations: "Funafuti".

longitude

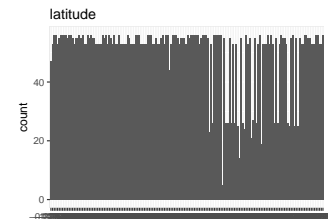
Feature	Result
Variable type	factor
Number of missing obs.	108 (1.29 %)
Number of unique values	164
Mode	"-0.20795"



- The following suspected missing value codes enter as regular values: "".
- Note that the following levels have at most five observations: "171.135".

latitude

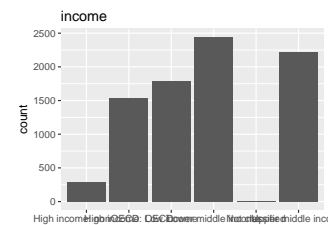
Feature	Result
Variable type	factor
Number of missing obs.	108 (1.29 %)
Number of unique values	164
Mode	"-1.27975"



- The following suspected missing value codes enter as regular values: "".
- Note that the following levels have at most five observations: "27.5768".

income

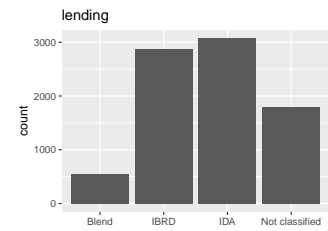
Feature	Result
Variable type	factor
Number of missing obs.	108 (1.29 %)
Number of unique values	6
Mode	"Lower middle income"



- Note that the following levels have at most five observations: "Lower middle income".

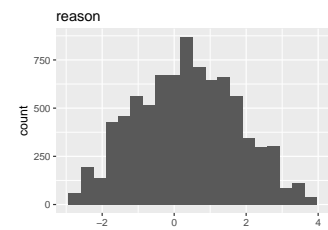
lending

Feature	Result
Variable type	factor
Number of missing obs.	108 (1.29 %)
Number of unique values	4
Mode	"IDA"



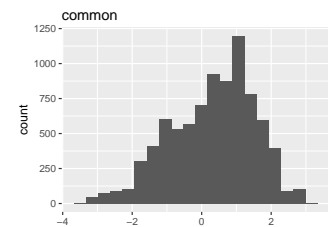
reason

Feature	Result
Variable type	numeric
Number of missing obs.	55 (0.65 %)
Number of unique values	1705
Median	0.39
1st and 3rd quartiles	-0.68; 1.32
Min. and max.	-2.82; 3.73



common

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1493
Median	0.47
1st and 3rd quartiles	-0.56; 1.2
Min. and max.	-3.41; 3.25

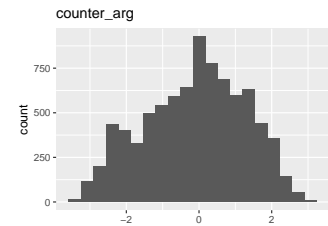


- Note that the following possible outlier values were detected: "2.66", "2.66", "2.68", "2.69", "2.7", "2.78", "2.81", "2.81", "2.82", "2.86" (3 additional values omitted).

counter_arg

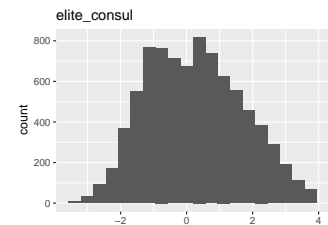
Feature	Result
Variable type	numeric
Number of missing obs.	20 (0.24 %)
Number of unique values	1792
Median	0
1st and 3rd quartiles	-1.13; 0.94
Min. and max.	-3.25; 3.24

- Note that the following possible outlier values were detected: "3.24".



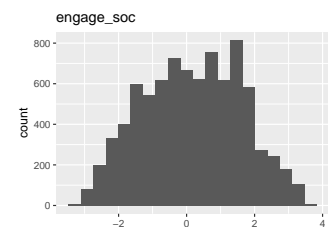
elite_consul

Feature	Result
Variable type	numeric
Number of missing obs.	20 (0.24 %)
Number of unique values	1860
Median	0.21
1st and 3rd quartiles	-0.89; 1.27
Min. and max.	-3.21; 3.94



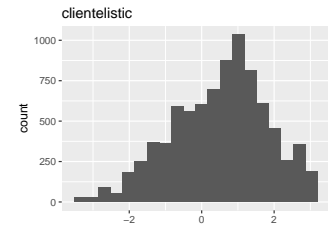
engage_soc

Feature	Result
Variable type	numeric
Number of missing obs.	20 (0.24 %)
Number of unique values	1795
Median	0.19
1st and 3rd quartiles	-0.96; 1.32
Min. and max.	-3.24; 3.7



clientelistic

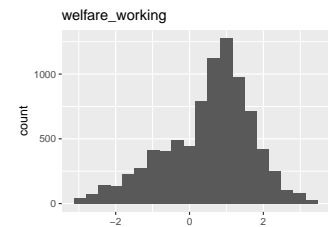
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1429
Median	0.64
1st and 3rd quartiles	-0.39; 1.39
Min. and max.	-3.23; 3.15



- Note that the following possible outlier values were detected: "3.04", "3.07", "3.15".

welfare_working

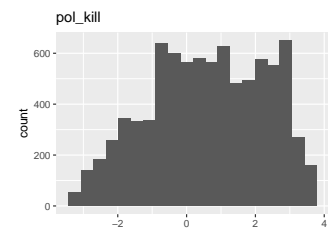
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1269
Median	0.73
1st and 3rd quartiles	-0.23; 1.32
Min. and max.	-3.03; 3.23



- Note that the following possible outlier values were detected: "2.25", "2.27", "2.28", "2.29", "2.29", "2.31", "2.31", "2.32", "2.36", "2.36" (18 additional values omitted).

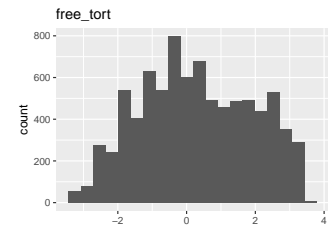
pol_kill

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1834
Median	0.68
1st and 3rd quartiles	-0.64; 2.08
Min. and max.	-3.17; 3.69



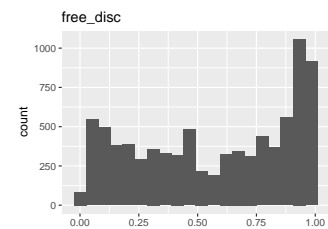
free_tort

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1731
Median	0.21
1st and 3rd quartiles	-1.01; 1.64
Min. and max.	-3.34; 3.53



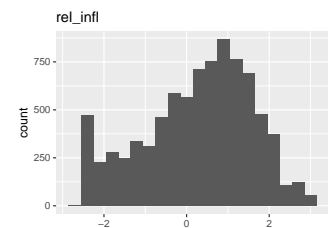
free_disc

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1947
Median	0.61
1st and 3rd quartiles	0.28; 0.89
Min. and max.	0.01; 0.99



rel_infl

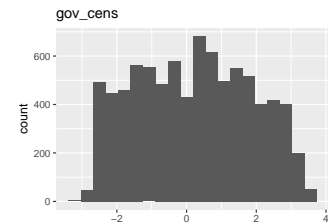
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1080
Median	0.45
1st and 3rd quartiles	-0.58; 1.22
Min. and max.	-2.58; 3.14



- Note that the following possible outlier values were detected: "2.69", "2.7", "2.72", "2.74", "2.74", "2.77", "2.8", "2.81", "2.82", "2.82" (9 additional values omitted).

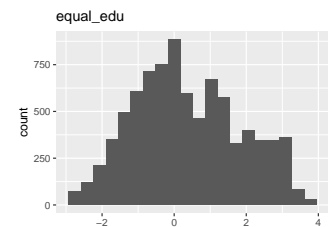
gov_cens

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1667
Median	0.23
1st and 3rd quartiles	-1.16; 1.55
Min. and max.	-3.04; 3.74



equal_edu

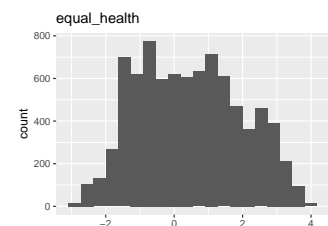
Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1030
Median	0.17
1st and 3rd quartiles	-0.79; 1.39
Min. and max.	-2.9; 3.65



- Note that the following possible outlier values were detected: "-2.9", "-2.82", "-2.71", "-2.7", "-2.67", "-2.66", "-2.66", "-2.62", "-2.58", "-2.55" (2 additional values omitted).

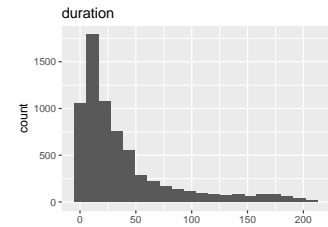
equal_health

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	1076
Median	0.37
1st and 3rd quartiles	-0.77; 1.58
Min. and max.	-3.07; 3.82



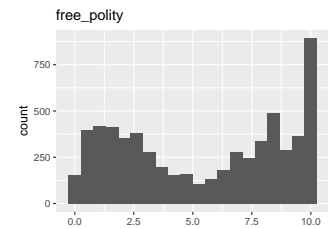
duration

Feature	Result
Variable type	integer
Number of missing obs.	1609 (19.16 %)
Number of unique values	208
Median	21
1st and 3rd quartiles	9; 46
Min. and max.	1; 208



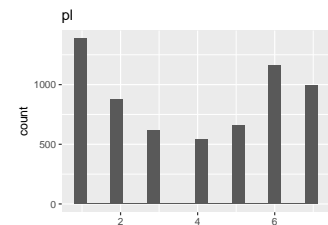
free_polity

Feature	Result
Variable type	numeric
Number of missing obs.	2165 (25.78 %)
Number of unique values	147
Median	5.92
1st and 3rd quartiles	2; 8.67
Min. and max.	0; 10



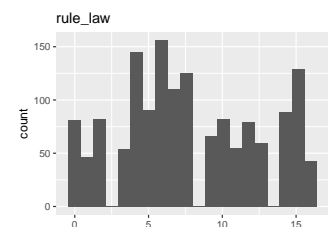
pl

Feature	Result
Variable type	integer
Number of missing obs.	2165 (25.78 %)
Number of unique values	7
Median	4
1st and 3rd quartiles	2; 6
Min. and max.	1; 7



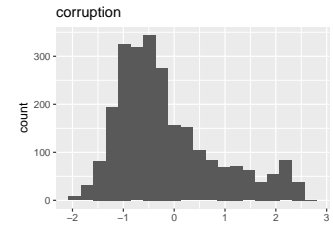
rule_law

Feature	Result
Variable type	integer
Number of missing obs.	6905 (82.22 %)
Number of unique values	17
Median	7
1st and 3rd quartiles	4; 12
Min. and max.	0; 16



corruption

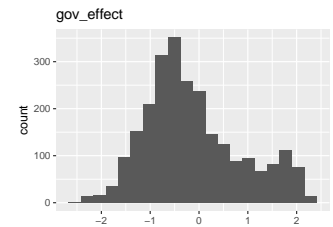
Feature	Result
Variable type	numeric
Number of missing obs.	5906 (70.33 %)
Number of unique values	2425
Median	-0.4
1st and 3rd quartiles	-0.87; 0.34
Min. and max.	-2.06; 2.59



- Note that the following possible outlier values were detected: "-2.06", "-1.92", "-1.91", "-1.9", "-1.9", "-1.85", "-1.84", "-1.84", "-1.83", "-1.82" (60 additional values omitted).

gov_effect

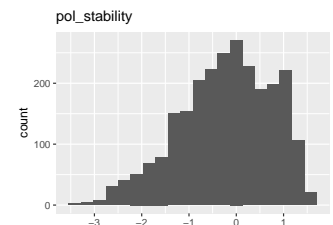
Feature	Result
Variable type	numeric
Number of missing obs.	5906 (70.33 %)
Number of unique values	2446
Median	-0.33
1st and 3rd quartiles	-0.81; 0.47
Min. and max.	-2.48; 2.36



- Note that the following possible outlier values were detected: "-2.48", "-2.45", "-2.34", "-2.32", "-2.31", "-2.28", "-2.27", "-2.26", "-2.26", "-2.25" (99 additional values omitted).

pol_stability

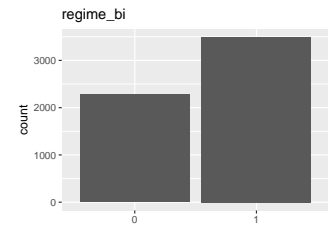
Feature	Result
Variable type	numeric
Number of missing obs.	5902 (70.28 %)
Number of unique values	2463
Median	-0.15
1st and 3rd quartiles	-0.87; 0.55
Min. and max.	-3.32; 1.67



regime_bi

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

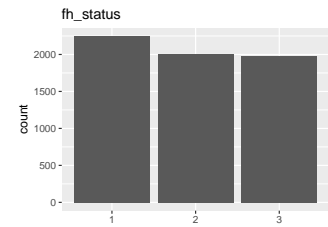
Feature	Result
Variable type	integer
Number of missing obs.	2633 (31.35 %)
Number of unique values	2
Mode	"1"



fh_status

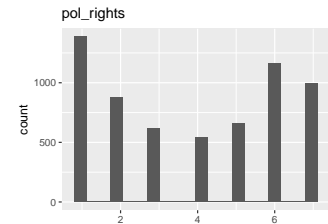
- Note that this variable is treated as a factor variable below, as it only takes a few unique values.

Feature	Result
Variable type	integer
Number of missing obs.	2165 (25.78 %)
Number of unique values	3
Mode	"1"



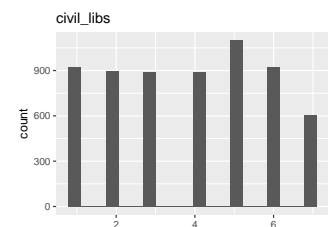
pol_rights

Feature	Result
Variable type	integer
Number of missing obs.	2165 (25.78 %)
Number of unique values	7
Median	4
1st and 3rd quartiles	2; 6
Min. and max.	1; 7



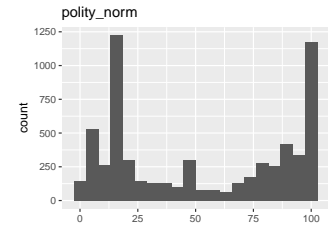
civil_libs

Feature	Result
Variable type	integer
Number of missing obs.	2165 (25.78 %)
Number of unique values	7
Median	4
1st and 3rd quartiles	2; 5
Min. and max.	1; 7



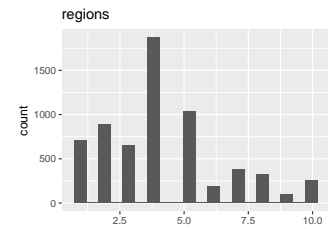
polity_norm

Feature	Result
Variable type	integer
Number of missing obs.	2157 (25.68 %)
Number of unique values	21
Median	50
1st and 3rd quartiles	15; 90
Min. and max.	0; 100



regions

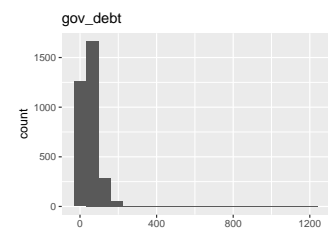
Feature	Result
Variable type	integer
Number of missing obs.	2011 (23.95 %)
Number of unique values	10
Median	4
1st and 3rd quartiles	2; 5
Min. and max.	1; 10



- Note that the following possible outlier values were detected: "10".

gov_debt

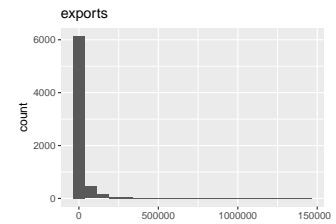
Feature	Result
Variable type	numeric
Number of missing obs.	5111 (60.86 %)
Number of unique values	3276
Median	41.28
1st and 3rd quartiles	22.15; 66.52
Min. and max.	1.94; 1209.3



- Note that the following possible outlier values were detected: "185.93", "189.14", "196.04", "196.93", "201.45", "203.68", "206.15", "206.95", "207.5", "207.8" (32 additional values omitted).

exports

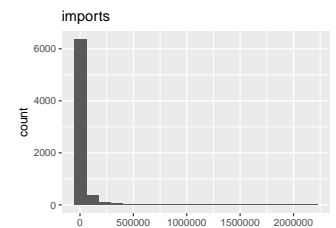
Feature	Result
Variable type	numeric
Number of missing obs.	1462 (17.41 %)
Number of unique values	6739
Median	1148.98
1st and 3rd quartiles	235.36; 8652.2
Min. and max.	0.1; 1429340



- Note that the following possible outlier values were detected: "158899", "160664", "161235", "161269", "162308", "163030.09", "163807", "164892", "165025.8", "165204" (248 additional values omitted).

imports

Feature	Result
Variable type	numeric
Number of missing obs.	1463 (17.42 %)
Number of unique values	6786
Median	1626.8
1st and 3rd quartiles	401.77; 9576.31
Min. and max.	0.3; 2166020

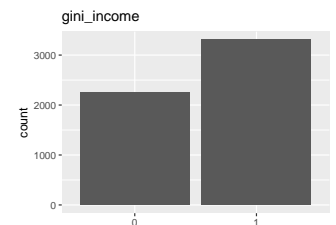


- Note that the following possible outlier values were detected: "153143", "153430", "154392", "154756", "155373", "155549", "156172", "156932", "157292", "157682" (244 additional values omitted).

gini_income

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

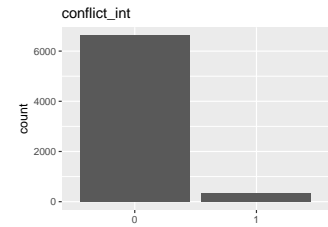
Feature	Result
Variable type	integer
Number of missing obs.	2833 (33.73 %)
Number of unique values	2
Mode	"1"



conflict_int

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

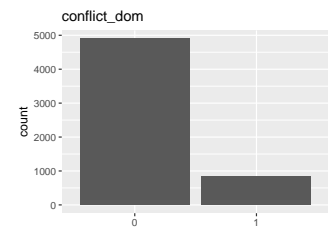
Feature	Result
Variable type	integer
Number of missing obs.	1425 (16.97 %)
Number of unique values	2
Mode	"0"



conflict_dom

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

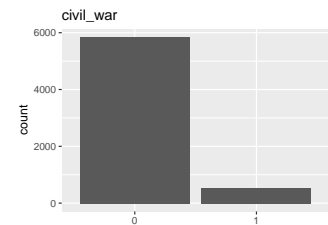
Feature	Result
Variable type	integer
Number of missing obs.	2666 (31.75 %)
Number of unique values	2
Mode	"0"



civil_war

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

Feature	Result
Variable type	integer
Number of missing obs.	2014 (23.98 %)
Number of unique values	2
Mode	"0"



Report generation information:

- Created by Simon Roth.
- Report creation time: Fr Aug 11 2017 21:24:33
- dataMaid v0.9.7 [Pkg: 2017-08-11 from Github (ekstroem/dataMaid@15c0983)]
- R version 3.4.0 (2017-04-21).
- Platform: x86_64-apple-darwin15.6.0 (64-bit)(macOS Sierra 10.12.5).
- Function call: `clean(data = project_macro)`