obesity

Autogenerated data summary from dataMaid

2021-01-12 23:06:11

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	16380
Number of variables	9

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?
country	character	195	0.00 %	
year	integer	42	0.00 %	
obesity	character	11638	0.00 %	×
sex	character	2	0.00 %	
literacy_rate_adult	numeric	1367	91.47 %	
alcohol_cons_per_cap	numeric	858	92.32 %	
edu_primary_rate	numeric	1222	91.89 %	
life_exp_birth	numeric	9652	19.67 %	×
unemployment_rate	numeric	6342	51.43 %	×

Variable list

country

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	195
Mode	"Afghanistan"
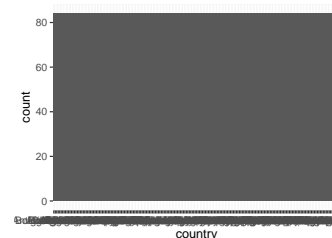
year

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	42
Median	1995.5
1st and 3rd quartiles	1985; 2006
Min. and max.	1975; 2016
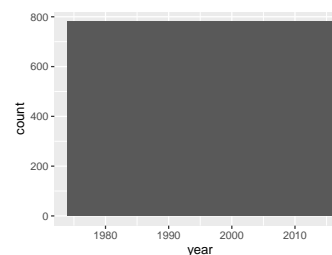
obesity

Feature	Result
Variable type	character
Number of missing obs.	0 (0 %)
Number of unique values	11638
Mode	"No data"
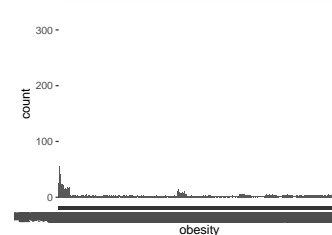
- Note that the following levels have at most five observations: "0.1 [0.0-0.5]", "0.1 [0.1-0.3]", "0.2 [0.0-0.8]", "0.2 [0.1-0.3]", "0.2 [0.1-0.7]", "0.3 [0.0-0.8]", "0.3 [0.0-0.9]", "0.3 [0.1-1.1]", "0.3 [0.2-0.5]", "0.3 [0.2-0.6]" (11458 additional values omitted).
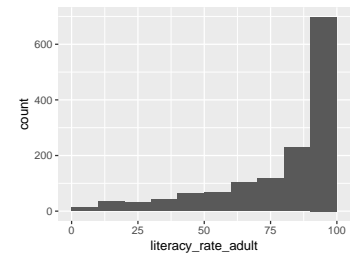
sex

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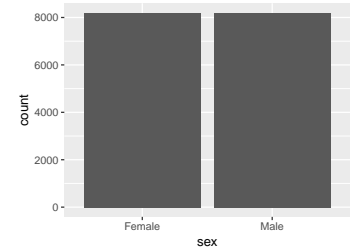
literacy_rate_adult

Feature	Result
Variable type	numeric
Number of missing obs.	14982 (91.47 %)
Number of unique values	1366
Median	90
1st and 3rd quartiles	69.78; 96.14
Min. and max.	3.39; 100

alcohol_cons_per_cap

Feature	Result
Variable type	numeric
Number of missing obs.	15122 (92.32 %)
Number of unique values	857
Median	3.79
1st and 3rd quartiles	1.49; 8.88
Min. and max.	0; 28.57
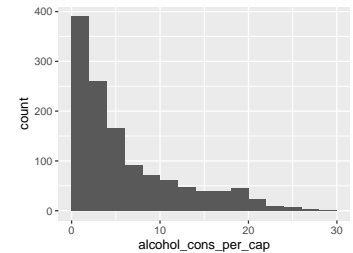
edu_primary_rate

Feature	Result
Variable type	numeric
Number of missing obs.	15052 (91.89 %)
Number of unique values	1221
Median	84.42
1st and 3rd quartiles	67.12; 97.59
Min. and max.	0.59; 100
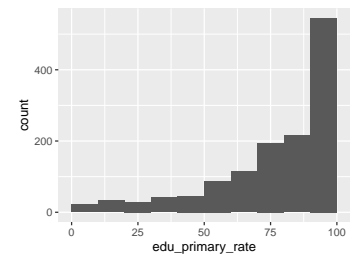
life_exp_birth

Feature	Result
Variable type	numeric
Number of missing obs.	3222 (19.67 %)
Number of unique values	9651
Median	67.49
1st and 3rd quartiles	58.04; 74.11
Min. and max.	16.29; 87.14
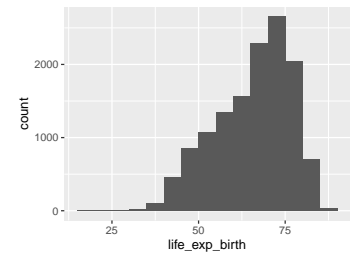
- Note that the following possible outlier values were detected: "85.3", "85.33", "85.4", "85.49", "85.5", "85.59", "85.6", "85.7", "85.8", "85.81" (13 additional values omitted).

---

unemployment_rate

Feature	Result
Variable type	numeric
Number of missing obs.	8424 (51.43 %)
Number of unique values	6341
Median	6.16
1st and 3rd quartiles	3.61; 10.91
Min. and max.	0.06; 47.65
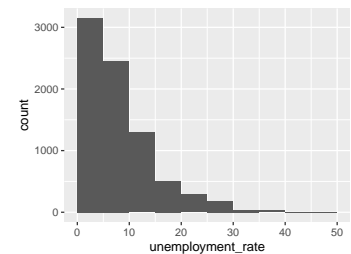
- Note that the following possible outlier values were detected: "0.06", "0.07", "0.09", "0.09", "0.1", "0.12", "0.12", "0.13", "0.14", "0.14" (218 additional values omitted).

Report generation information:

- Created by: Dustin Burnham (username: dusty).

- Report creation time: Tue Jan 12 2021 23:06:13

- Report was run from directory: /Users/dusty/Desktop/UBC_2020/block_4/dsci_532/dash

- dataMaid v1.4.0 [Pkg: 2019-12-10 from CRAN (R 4.0.2)]

- R version 4.0.2 (2020-06-22).

- Platform: x86_64-apple-darwin17.0 (64-bit)(macOS Catalina 10.15.6).

- Function call: dataMaid::makeDataReport(data = obesity, file = here::here("eda", "dustin", "images", "profile_world_bank.Rmd"), replace = TRUE)