REPORT SERIES WITH DLOOKR

Exploratory Data Analysis Report

Author:
dlookr package

Version:
0.3.12

January 9, 2020

Contents

1 Introduction **3**

1.1 Information of Dataset 3

1.2 Information of Variables 3

1.3 About EDA Report 3

2 Univariate Analysis **5**

2.1 Descriptive Statistics 5

2.2 Normality Test of Numerical Variables 21

2.2.1 Statistics and Visualization of (Sample) Data 21

3 Relationship Between Variables **27**

3.1 Correlation Coefficient 27

3.1.1 Correlation Coefficient by Variable Combination 27

3.1.2 Correlation Plot of Numerical Variables 27

4 Target based Analysis **29**

4.1 Grouped Descriptive Statistics 29

4.1.1 Grouped Numerical Variables 29

4.1.2 Grouped Categorical Variables 29

4.2 Grouped Relationship Between Variables 29

4.2.1 Grouped Correlation Coefficient 29

4.2.2 Grouped Correlation Plot of Numerical Variables 29

Chapter 1

Introduction

The EDA Report provides exploratory data analysis information on objects that inherit data.frame and data.frame.

1.1 Information of Dataset

The dataset that generated the EDA Report is an 'data.frame' object. It consists of 5,607 observations and 148 variables.

1.2 Information of Variables

The target variable of the data is 'NULL', and the data type of the variable is NULL(You did not specify a target variable).

1.3 About EDA Report

EDA reports provide information and visualization results that support the EDA process. In particular, it provides a variety of information to understand the relationship between the target variable and the rest of the variables of interest.

Table 1.1: Information of Variables

variables	types	missing_count	missing_percent	unique_cou
subject_id	integer	0	0.0000000	560
sex	factor	222	3.9593365	
age_years	numeric	0	0.0000000	9
bmi	numeric	0	0.0000000	165
height_cm	numeric	0	0.0000000	18
weight_kg	numeric	0	0.0000000	16
country_of_birth	factor	22	0.3923667	9
census_region	factor	1053	18.7800963	
economic_region	factor	1053	18.7800963	
state	factor	791	14.1073658	9
country_residence	logical	5607	100.0000000	
diet_type	factor	59	1.0522561	
multivitamin	factor	88	1.5694667	
probiotic_frequency	factor	2990	53.3261994	
vitamin_b_supplement_frequency	factor	3005	53.5937221	
vitamin_d_supplement_frequency	factor	3010	53.6828964	
other_supplement_frequency	factor	93	1.6586410	
specialized_diet_exclude_dairy	logical	5607	100.0000000	
specialized_diet_exclude_nightshades	logical	5607	100.0000000	
specialized_diet_exclude_refined_sugars	logical	5607	100.0000000	
specialized_diet_fodmap	logical	5607	100.0000000	
specialized_diet_halaal	logical	5607	100.0000000	
specialized_diet_i_do_not_eat_a_specialized_diet	logical	5607	100.0000000	
specialized_diet_kosher	logical	5607	100.0000000	
specialized_diet_modified_paleo_diet	logical	5607	100.0000000	
specialized_diet_other_restrictions_not_described_here	logical	5607	100.0000000	
specialized_diet_paleodiet_or_primal_diet	logical	5607	100.0000000	
specialized_diet_raw_food_diet	logical	5607	100.0000000	
specialized_diet_unspecified	logical	5607	100.0000000	
specialized_diet_westenprice_or_other_lowgrain_low_processed_fo	logical	5607	100.0000000	
consume_animal_products_abx	factor	2979	53.1300161	
drinking_water_source	factor	38	0.6777243	
race	factor	26	0.4637061	
last_move	factor	2968	52.9338327	
last_travel	factor	140	2.4968789	
roommates	factor	4959	88.4430177	
roommates_in_study	factor	3237	57.7314072	
livingwith	factor	89	1.5873016	
dog	factor	83	1.4802925	
cat	factor	128	2.2828607	
pets_other	logical	5607	100.0000000	
dominant_hand	factor	120	2.1401819	
level_of_education	factor	3060	54.5746388	
exercise_frequency	factor	48	0.8560728	
exercise_location	factor	191	3.4064562	
nail_biter	factor	158	2.8179062	
pool_frequency	factor	50	0.8917425	
smoking_frequency	factor	46	0.8204031	
alcohol_consumption	factor	49	0.8739076	
alcohol_frequency	factor	49	0.8739076	
alcohol_types_beercider	factor	0	0.0000000	
alcohol_types_red_wine	factor	0	0.0000000	
alcohol_types_sour_beers	factor	0	0.0000000	
alcohol_types_spiritshard_alcohol	factor	0	0.0000000	
alcohol_types_unspecified	factor	0	0.0000000	

Chapter 2

Univariate Analysis

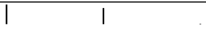
2.1 Descriptive Statistics

edaData
148 Variables 5607 Observations

subject_id

```
       n   missing  distinct    Info    Mean     Gmd     .05     .10     .25     .50     .75     .90     .95
    5607         0      5607       1    4286    2233    1310    1610    2594    4359    5952    6936    7280

lowest : 1000 1001 1002 1003 1004, highest: 7569 7570 7571 7572 7573
```

sex



```
       n   missing  distinct
    5385       222         3

Value      female   male  other
Frequency    2964   2419      2
Proportion  0.550  0.449  0.000
```

age_years

```
       n   missing  distinct    Info    Mean     Gmd   .05   .10   .25   .50   .75   .90   .95
    5607         0        91       1   46.52   19.81    10    23    35    48    60    67    71

lowest :   1   2   3   4   5, highest:  88  89  92  94 101
```

bmi

```
       n   missing  distinct    Info    Mean     Gmd      .05      .10      .25      .50      .75      .90      .95
    5607         0      1650       1    2561    5033    16.94    18.75    20.82    23.31    26.57    32.83    46.65

lowest :       0.00       0.21       0.22       0.23       0.24
highest:  760000.00  830000.00  970000.00 1050000.00 1130000.00
```

height_cm

```
       n   missing  distinct     Info    Mean    Gmd   .05   .10   .25   .50   .75   .90   .95
    5607         0       185    0.998   168.9   44.2    81   147   162   170   178   184   188

lowest :     1     2     3     4     5, highest:  1800  1820  1867  1918 16540

Value          0   200   400   600   800  1000  1200  1400  1600  1800  2000 16600
Frequency    343  5226    18     1     2     2     3     1     3     6     1     1
Proportion 0.061 0.932 0.003 0.000 0.000 0.000 0.001 0.000 0.001 0.001 0.000 0.000

For the frequency table, variable is rounded to the nearest 200
```

weight_kg

```
       n   missing  distinct     Info    Mean    Gmd    .05    .10    .25    .50    .75    .90    .95
    5607         0       168    0.999   69.77     23   39.3   50.0   58.0   68.0   79.0   90.0  100.0

lowest :   2   3   4   5   6, highest: 419 463 515 659 932
```

country_of_birth

```
         n   missing   distinct
      5585        22         89

lowest : Algeria                              Argentina                             Australia                             Austria
highest: United States Minor Outlying Islands Venezuela                             Viet Nam                              Zambia
```

---

census_region

```
         n   missing   distinct
      4554      1053          4

Value        Midwest Northeast     South      West
Frequency        595       987      1100      1872
Proportion     0.131     0.217     0.242     0.411
```

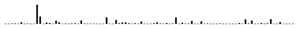
---

economic_region

```
         n   missing   distinct
      4554      1053          8

lowest : Far West       Great Lakes    Mideast        New England    Plains
highest: New England    Plains         Rocky Mountain Southeast      Southwest

Value            Far West    Great Lakes        Mideast    New England         Plains
Frequency            1317            411            739            481            184
Proportion          0.289          0.090          0.162          0.106          0.040

Value      Rocky Mountain      Southeast      Southwest
Frequency             445            638            339
Proportion          0.098          0.140          0.074
```

---

state



```
         n   missing   distinct
      4816       791         92

lowest : AB  ACT AK  AL  AR , highest: WI  WV  WY  YT  ZH
```

---

diet_type

```
         n   missing   distinct
      5548        59          5

lowest : Omnivore                           Omnivore but do not eat red meat Vegan                            Vegetarian
highest: Omnivore                           Omnivore but do not eat red meat Vegan                            Vegetarian

Omnivore (4541, 0.818), Omnivore but do not eat red meat (385, 0.069), Vegan (127, 0.023),
Vegetarian (213, 0.038), Vegetarian but eat seafood (282, 0.051)
```

---

multivitamin

```
         n   missing   distinct
      5519        88          2

Value      false  true
Frequency   3593  1926
Proportion 0.651 0.349
```

---

probiotic_frequency

```
         n   missing   distinct
      2617      2990          5

lowest : Daily                         Never                         Occasionally (1-2 times/week) Rarely (a few times/month)
highest: Daily                         Never                         Occasionally (1-2 times/week) Rarely (a few times/month)

Daily (549, 0.210), Never (1000, 0.382), Occasionally (1-2 times/week) (267, 0.102), Rarely (a
few times/month) (523, 0.200), Regularly (3-5 times/week) (278, 0.106)
```

---

vitamin_b_supplement_frequency

```
         n   missing   distinct
      2602      3005          5

lowest : Daily                         Never                         Occasionally (1-2 times/week) Rarely (a few times/month)
highest: Daily                         Never                         Occasionally (1-2 times/week) Rarely (a few times/month)

Daily (507, 0.195), Never (1413, 0.543), Occasionally (1-2 times/week) (182, 0.070), Rarely (a
few times/month) (274, 0.105), Regularly (3-5 times/week) (226, 0.087)
```

---

vitamin_d_supplement_frequency

```
       n   missing   distinct
    2597      3010          5
```

```
lowest : Daily                         Never                         Occasionally (1-2 times/week) Rarely (a few times/month)
highest: Daily                         Never                         Occasionally (1-2 times/week) Rarely (a few times/month)
```

```
Daily (722, 0.278), Never (961, 0.370), Occasionally (1-2 times/week) (253, 0.097), Rarely (a
few times/month) (319, 0.123), Regularly (3-5 times/week) (342, 0.132)
```

other_supplement_frequency

```
       n   missing   distinct
    5514        93          2
```

```
Value      false  true
Frequency   2069  3445
Proportion 0.375 0.625
```

consume_animal_products_abx

```
       n   missing   distinct
    2628      2979          3
```

```
Value            No Not sure      Yes
Frequency       726      734     1168
Proportion    0.276    0.279    0.444
```

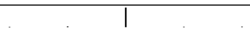
drinking_water_source

```
       n   missing   distinct
    5569        38          5
```

```
lowest : Bottled  City     Filtered Not sure Well    , highest: Bottled  City     Filtered Not sure Well
```

```
Value       Bottled     City Filtered Not sure     Well
Frequency       490     2451     2088       41      499
Proportion    0.088    0.440    0.375    0.007    0.090
```

race



```
       n   missing   distinct
    5581        26          5
```

```
lowest : African American          Asian or Pacific Islander Caucasian                 Hispanic                  Other
highest: African American          Asian or Pacific Islander Caucasian                 Hispanic                  Other
```

```
Value             African American Asian or Pacific Islander                 Caucasian
Frequency                       47                       208                      5081
Proportion                   0.008                     0.037                     0.910

Value                     Hispanic                     Other
Frequency                      107                       138
Proportion                   0.019                     0.025
```

last_move

```
       n   missing   distinct
    2639      2968          5
```

```
lowest : I have lived in my current state of residence for more than a year. Within the past 3 months
highest: I have lived in my current state of residence for more than a year. Within the past 3 months
```

```
I have lived in my current state of residence for more than a year. (2417, 0.916), Within the
past 3 months (33, 0.013), Within the past 6 months (54, 0.020), Within the past month (33,
0.013), Within the past year (102, 0.039)
```

last_travel

```
       n   missing   distinct
    5467       140          5
```

```
lowest : 1 year                                                         3 months
highest: 1 year                                                         3 months
```

```
1 year (865, 0.158), 3 months (634, 0.116), 6 months (605, 0.111), I have not been outside of
my country of residence in the past year. (2916, 0.533), Month (447, 0.082)
```

roommates

```
        n   missing  distinct
      648      4959         4

Value      More than three              One            Three             Two
Frequency               54              450               61              83
Proportion           0.083            0.694            0.094           0.128
```

roommates_in_study

```
        n   missing  distinct
     2370      3237         3

Value            No Not sure      Yes
Frequency      2052       35      283
Proportion    0.866    0.015    0.119
```

livingwith

```
        n   missing  distinct
     5518        89         3

Value            No Not sure      Yes
Frequency      3006      294     2218
Proportion    0.545    0.053    0.402
```

dog

```
        n   missing  distinct
     5524        83         2

Value      false  true
Frequency   3751  1773
Proportion 0.679 0.321
```

cat

```
        n   missing  distinct
     5479       128         2

Value      false  true
Frequency   3856  1623
Proportion 0.704 0.296
```

dominant_hand

```
        n   missing  distinct
     5487       120         3

Value      I am ambidextrous  I am left handed I am right handed
Frequency                129               505              4853
Proportion             0.024             0.092             0.884
```

level_of_education

```
        n   missing  distinct
     2547      3060         7

lowest : Associate's degree               Bachelor's degree                Did not complete high school         Graduat
highest: Did not complete high school     Graduate or Professional degree  High School or GED equilivant        Some col

Associate's degree (62, 0.024), Bachelor's degree (634, 0.249), Did not complete high school
(100, 0.039), Graduate or Professional degree (1218, 0.478), High School or GED equilivant
(111, 0.044), Some college or technical school (216, 0.085), Some graduate school or
professional (206, 0.081)
```

exercise_frequency

```
        n   missing  distinct
     5559        48         5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (a few times/month)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (a few times/month)

Daily (1281, 0.230), Never (151, 0.027), Occasionally (1-2 times/week) (1260, 0.227), Rarely
(a few times/month) (597, 0.107), Regularly (3-5 times/week) (2270, 0.408)
```

exercise_location

```
       n   missing  distinct
    5416       191         5

lowest : Both                  Depends on the season Indoors               None of the above     Outdoors
highest: Both                  Depends on the season Indoors               None of the above     Outdoors

Value                      Both Depends on the season               Indoors
Frequency                  1926                   633                  1184
Proportion                0.356                 0.117                 0.219

Value         None of the above              Outdoors
Frequency                   161                  1512
Proportion                0.030                 0.279
```

nail_biter

```
       n   missing  distinct
    5449       158         2

Value      false  true
Frequency   4380  1069
Proportion 0.804 0.196
```

pool_frequency

```
       n   missing  distinct
    5557        50         5

lowest : Daily                           Never                           Occasionally (1-2 times/week) Rarely (a few times/month)
highest: Daily                           Never                           Occasionally (1-2 times/week) Rarely (a few times/month)

Daily (33, 0.006), Never (2672, 0.481), Occasionally (1-2 times/week) (460, 0.083), Rarely (a
few times/month) (2211, 0.398), Regularly (3-5 times/week) (181, 0.033)
```

smoking_frequency

```
       n   missing  distinct
    5561        46         5

lowest : Daily                           Never                           Occasionally (1-2 times/week) Rarely (a few times/month)
highest: Daily                           Never                           Occasionally (1-2 times/week) Rarely (a few times/month)

Daily (59, 0.011), Never (5283, 0.950), Occasionally (1-2 times/week) (35, 0.006), Rarely (a
few times/month) (159, 0.029), Regularly (3-5 times/week) (25, 0.004)
```

alcohol_consumption

```
       n   missing  distinct
    5558        49         2

Value      false  true
Frequency   1345  4213
Proportion 0.242 0.758
```

alcohol_frequency

```
       n   missing  distinct
    5558        49         5

lowest : Daily                           Never                           Occasionally (1-2 times/week) Rarely (a few times/month)
highest: Daily                           Never                           Occasionally (1-2 times/week) Rarely (a few times/month)

Daily (631, 0.114), Never (1345, 0.242), Occasionally (1-2 times/week) (1177, 0.212), Rarely
(a few times/month) (1350, 0.243), Regularly (3-5 times/week) (1055, 0.190)
```

alcohol_types_beercider

```
       n   missing  distinct
    5607         0         4

Value      false    No  true   Yes
Frequency   4501     7  1097     2
Proportion 0.803 0.001 0.196 0.000
```

alcohol_types_red_wine

```
       n   missing  distinct
    5607         0         2

Value      false  true
Frequency   4144  1463
Proportion 0.739 0.261
```

alcohol_types_sour_beers

n	missing	distinct
5607	0	2

```
Value      false  true
Frequency   5509    98
Proportion 0.983 0.017
```

---

alcohol_types_spiritshard_alcohol

n	missing	distinct
5607	0	2

```
Value      false  true
Frequency   4742   865
Proportion 0.846 0.154
```

---

alcohol_types_unspecified

n	missing	distinct
5607	0	2

```
Value      false  true
Frequency   2026  3581
Proportion 0.361 0.639
```

---

alcohol_types_white_wine

n	missing	distinct
5607	0	2

```
Value      false  true
Frequency   4567  1040
Proportion 0.815 0.185
```

---

teethbrushing_frequency

n	missing	distinct
5549	58	5

```
lowest : Daily                         Never                          Occasionally (1-2 times/week) Rarely (a few times/month)
highest: Daily                         Never                          Occasionally (1-2 times/week) Rarely (a few times/month)
```

```
Daily (5275, 0.951), Never (20, 0.004), Occasionally (1-2 times/week) (41, 0.007), Rarely (a
few times/month) (20, 0.004), Regularly (3-5 times/week) (193, 0.035)
```

---

flossing_frequency

n	missing	distinct
5567	40	5

```
lowest : Daily                         Never                          Occasionally (1-2 times/week) Rarely (a few times/month)
highest: Daily                         Never                          Occasionally (1-2 times/week) Rarely (a few times/month)
```

```
Daily (2295, 0.412), Never (537, 0.096), Occasionally (1-2 times/week) (875, 0.157), Rarely (a
few times/month) (898, 0.161), Regularly (3-5 times/week) (962, 0.173)
```

---

cosmetics_frequency

n	missing	distinct
5552	55	5

```
lowest : Daily                         Never                          Occasionally (1-2 times/week) Rarely (a few times/month)
highest: Daily                         Never                          Occasionally (1-2 times/week) Rarely (a few times/month)
```

```
Daily (1056, 0.190), Never (3104, 0.559), Occasionally (1-2 times/week) (328, 0.059), Rarely
(a few times/month) (528, 0.095), Regularly (3-5 times/week) (536, 0.097)
```

---

deodorant_use

n	missing	distinct
5551	56	4

```
I do not use deodorant or an antiperspirant (1659, 0.299), I use an antiperspirant (1189,
0.214), I use deodorant (2036, 0.367), Not sure, but I use some form of
deodorant/antiperspirant (667, 0.120)
```

---

sleep_duration

```
       n   missing   distinct
    5573        34          5

lowest : 5-6 hours         6-7 hours         7-8 hours         8 or more hours   Less than 5 hours
highest: 5-6 hours         6-7 hours         7-8 hours         8 or more hours   Less than 5 hours

Value              5-6 hours         6-7 hours         7-8 hours   8 or more hours
Frequency                497              1654              2328              1018
Proportion             0.089             0.297             0.418             0.183

Value      Less than 5 hours
Frequency                 76
Proportion             0.014
```

softener

```
       n   missing   distinct
    5450       157          4

Value      false    No  true   Yes
Frequency   3681     5  1761     3
Proportion 0.675 0.001 0.323 0.001
```

bowel_movement_frequency

```
       n   missing   distinct
    2586      3021          6

lowest : Five or more  Four          Less than one One           Three
highest: Four          Less than one One           Three         Two

Value       Five or more          Four Less than one           One         Three           Two
Frequency             43            68           283          1278           211           703
Proportion         0.017         0.026         0.109         0.494         0.082         0.272
```

bowel_movement_quality

```
       n   missing   distinct
    2574      3033          7

lowest : I don't know, I do not have a point of reference                        I tend to be constipated (have difficulty pass
highest: I tend to be constipated (have difficulty passing stool) - Type 1 and 2 I tend to have diarrhea (watery stool)
```

antibiotic_history

```
       n   missing   distinct
    5536        71          5

lowest : 6 months                                        I have not taken antibiotics in the past year. Month
highest: 6 months                                        I have not taken antibiotics in the past year. Month

6 months (719, 0.130), I have not taken antibiotics in the past year. (3752, 0.678), Month
(165, 0.030), Week (113, 0.020), Year (787, 0.142)
```

flu_vaccine_date

```
       n   missing   distinct
    5506       101          5

lowest : 6 months                                             I have not gotten the flu vaccine in the past year. Month
highest: 6 months                                             I have not gotten the flu vaccine in the past year. Month

6 months (1104, 0.201), I have not gotten the flu vaccine in the past year. (3076, 0.559),
Month (152, 0.028), Week (26, 0.005), Year (1148, 0.208)
```

contraceptive

```
       n   missing   distinct
    3969      1638          6

lowest : No                                            Yes, I am taking the pill                    Yes, I use a contraceptive
highest: Yes, I am taking the pill                     Yes, I use a contraceptive patch (Ortho-Evra) Yes, I use a hormonal IUD

No (3687, 0.929), Yes, I am taking the pill (188, 0.047), Yes, I use a contraceptive patch
(Ortho-Evra) (6, 0.002), Yes, I use a hormonal IUD (Mirena) (53, 0.013), Yes, I use an
injected contraceptive (DMPA) (9, 0.002), Yes, I use the NuvaRing (26, 0.007)
```

pregnant

```
         n   missing   distinct
      3448      2159          3

Value          false Not sure     true
Frequency       3406       12       30
Proportion     0.988    0.003    0.009
```

---

weight_change

```
         n   missing   distinct
      5516        91          3

Value      Decreased more than 10 pounds Increased more than 10 pounds
Frequency                            524                           398
Proportion                         0.095                         0.072

Value                    Remained stable
Frequency                           4594
Proportion                         0.833
```

---

tonsils_removed

```
         n   missing   distinct
      5481       126          3

Value          false Not sure     true
Frequency       4007       26     1448
Proportion     0.731    0.005    0.264
```

---

appendix_removed

```
         n   missing   distinct
      5493       114          3

Value          false Not sure     true
Frequency       4934       14      545
Proportion     0.898    0.003    0.099
```

---

chickenpox

```
         n   missing   distinct
      5514        93          3

Value             No Not sure      Yes
Frequency        812      149     4553
Proportion     0.147    0.027    0.826
```

---

acne_medication

```
         n   missing   distinct
      5494       113          2

Value      false  true
Frequency   5362   132
Proportion 0.976 0.024
```

---

acne_medication_otc

```
         n   missing   distinct
      5516        91          2

Value      false  true
Frequency   5080   436
Proportion 0.921 0.079
```

---

csection

```
         n   missing   distinct
      5529        78          3

Value          false Not sure     true
Frequency       4775      206      548
Proportion     0.864    0.037    0.099
```

---

fed_as_infant

```
       n  missing  distinct
    2621     2986         4

Value      A mixture of breast milk and formula                           Not sure
Frequency                                   436                                357
Proportion                                0.166                              0.136

Value                     Primarily breast milk            Primarily infant formula
Frequency                                  1077                                751
Proportion                                0.411                              0.287
```

add_adhd

```
       n  missing  distinct
    2599     3008         4
```

```
Diagnosed by a medical professional (doctor, physician assistant) (121, 0.047), Diagnosed by
an alternative medicine practitioner (8, 0.003), I do not have this condition (2377, 0.915),
Self-diagnosed (93, 0.036)
```

alzheimers

```
       n  missing  distinct
    2638     2969         3
```

```
Diagnosed by a medical professional (doctor, physician assistant) (2, 0.001), I do not have
this condition (2635, 0.999), Self-diagnosed (1, 0.000)
```

lung_disease

```
       n  missing  distinct
    5220      387         4
```

```
Diagnosed by a medical professional (doctor, physician assistant) (329, 0.063), Diagnosed by
an alternative medicine practitioner (3, 0.001), I do not have this condition (4876, 0.934),
Self-diagnosed (12, 0.002)
```

asd

```
       n  missing  distinct
    2623     2984         4
```

```
Diagnosed by a medical professional (doctor, physician assistant) (73, 0.028), Diagnosed by an
alternative medicine practitioner (2, 0.001), I do not have this condition (2524, 0.962),
Self-diagnosed (24, 0.009)
```

autoimmune

```
       n  missing  distinct
    2597     3010         4
```

```
Diagnosed by a medical professional (doctor, physician assistant) (295, 0.114), Diagnosed by
an alternative medicine practitioner (18, 0.007), I do not have this condition (2265, 0.872),
Self-diagnosed (19, 0.007)
```

fungal_overgrowth

```
       n  missing  distinct
    2555     3052         4
```

```
Diagnosed by a medical professional (doctor, physician assistant) (122, 0.048), Diagnosed by
an alternative medicine practitioner (186, 0.073), I do not have this condition (2118, 0.829),
Self-diagnosed (129, 0.050)
```

cdiff

```
       n  missing  distinct
    2579     3028         4
```

```
Diagnosed by a medical professional (doctor, physician assistant) (50, 0.019), Diagnosed by an
alternative medicine practitioner (5, 0.002), I do not have this condition (2517, 0.976),
Self-diagnosed (7, 0.003)
```

cardiovascular_disease

```
       n  missing  distinct
    2630     2977         4
```

```
Diagnosed by a medical professional (doctor, physician assistant) (105, 0.040), Diagnosed by
an alternative medicine practitioner (1, 0.000), I do not have this condition (2518, 0.957),
Self-diagnosed (6, 0.002)
```

diabetes

n	missing	distinct
5442	165	4

```
Diagnosed by a medical professional (doctor, physician assistant) (62, 0.011), Diagnosed by an
alternative medicine practitioner (1, 0.000), I do not have this condition (5369, 0.987),
Self-diagnosed (10, 0.002)
```

---

epilepsy_or_seizure_disorder

n	missing	distinct
2623	2984	4

```
Diagnosed by a medical professional (doctor, physician assistant) (33, 0.013), Diagnosed by an
alternative medicine practitioner (2, 0.001), I do not have this condition (2584, 0.985),
Self-diagnosed (4, 0.002)
```

---

ibs

n	missing	distinct
2581	3026	4

```
Diagnosed by a medical professional (doctor, physician assistant) (369, 0.143), Diagnosed by
an alternative medicine practitioner (30, 0.012), I do not have this condition (2000, 0.775),
Self-diagnosed (182, 0.071)
```

---

ibd

n	missing	distinct
5278	329	4

```
Diagnosed by a medical professional (doctor, physician assistant) (121, 0.023), Diagnosed by
an alternative medicine practitioner (11, 0.002), I do not have this condition (5100, 0.966),
Self-diagnosed (46, 0.009)
```

---

migraine

n	missing	distinct
5140	467	4

```
Diagnosed by a medical professional (doctor, physician assistant) (336, 0.065), Diagnosed by
an alternative medicine practitioner (5, 0.001), I do not have this condition (4665, 0.908),
Self-diagnosed (134, 0.026)
```

---

kidney_disease

n	missing	distinct
2616	2991	4

```
Diagnosed by a medical professional (doctor, physician assistant) (41, 0.016), Diagnosed by an
alternative medicine practitioner (3, 0.001), I do not have this condition (2567, 0.981),
Self-diagnosed (5, 0.002)
```

---

liver_disease

n	missing	distinct
2611	2996	4

```
Diagnosed by a medical professional (doctor, physician assistant) (45, 0.017), Diagnosed by an
alternative medicine practitioner (5, 0.002), I do not have this condition (2557, 0.979),
Self-diagnosed (4, 0.002)
```

---

pku

n	missing	distinct
5458	149	4

```
Diagnosed by a medical professional (doctor, physician assistant) (1, 0.000), Diagnosed by an
alternative medicine practitioner (1, 0.000), I do not have this condition (5455, 0.999),
Self-diagnosed (1, 0.000)
```

---

sibo

n	missing	distinct
2528	3079	4

```
Diagnosed by a medical professional (doctor, physician assistant) (91, 0.036), Diagnosed by an
alternative medicine practitioner (50, 0.020), I do not have this condition (2280, 0.902),
Self-diagnosed (107, 0.042)
```

---

skin_condition

n	missing	distinct
4989	618	4

```
Diagnosed by a medical professional (doctor, physician assistant) (702, 0.141), Diagnosed by
an alternative medicine practitioner (10, 0.002), I do not have this condition (4201, 0.842),
Self-diagnosed (76, 0.015)
```

---

thyroid

n	missing	distinct
2611	2996	4

```
Diagnosed by a medical professional (doctor, physician assistant) (354, 0.136), Diagnosed by
an alternative medicine practitioner (29, 0.011), I do not have this condition (2210, 0.846),
Self-diagnosed (18, 0.007)
```

---

seasonal_allergies

n	missing	distinct
5436	171	4

```
Value      false    No  true   Yes
Frequency   3189     7  2238     2
Proportion 0.587 0.001 0.412 0.000
```

---

non_food_allergies_beestings

n	missing	distinct
5607	0	2

```
Value      false  true
Frequency   5529    78
Proportion 0.986 0.014
```

---

non_food_allergies_drug_eg_penicillin

n	missing	distinct
5607	0	2

```
Value      false  true
Frequency   5126   481
Proportion 0.914 0.086
```

---

non_food_allergies_pet_dander

n	missing	distinct
5607	0	2

```
Value      false  true
Frequency   5279   328
Proportion 0.942 0.058
```

---

non_food_allergies_poison_ivyoak

n	missing	distinct
5607	0	2

```
Value      false  true
Frequency   5236   371
Proportion 0.934 0.066
```

---

non_food_allergies_sun

n	missing	distinct
5607	0	2

```
Value      false  true
Frequency   5533    74
Proportion 0.987 0.013
```

---

non_food_allergies_unspecified

n	missing	distinct
5607	0	2

```
Value      false  true
Frequency    977  4630
Proportion 0.174 0.826
```

---

lactose

```
       n   missing  distinct
    5464       143         2

Value      false  true
Frequency   4427  1037
Proportion  0.81  0.19
```

---

gluten

```
       n   missing  distinct
    4809       798         4
```

```
I do not eat gluten because it makes me feel bad (511, 0.106), I was diagnosed with celiac
disease (38, 0.008), I was diagnosed with gluten allergy (anti-gluten IgG), but not celiac
disease (115, 0.024), No (4145, 0.862)
```

---

allergic_to_i_have_no_food_allergies_that_i_know_of

```
       n   missing  distinct
    5607         0         2

Value      false  true
Frequency   3732  1875
Proportion 0.666 0.334
```

---

allergic_to_other

```
       n   missing  distinct
    5607         0         2

Value      false  true
Frequency   5218   389
Proportion 0.931 0.069
```

---

allergic_to_peanuts

```
       n   missing  distinct
    5607         0         2

Value      false  true
Frequency   5535    72
Proportion 0.987 0.013
```

---

allergic_to_shellfish

```
       n   missing  distinct
    5607         0         2

Value      false  true
Frequency   5531    76
Proportion 0.986 0.014
```

---

allergic_to_tree_nuts

```
       n   missing  distinct
    5607         0         2

Value      false  true
Frequency   5536    71
Proportion 0.987 0.013
```

---

allergic_to_unspecified

```
       n   missing  distinct
    5607         0         2

Value      false  true
Frequency   2349  3258
Proportion 0.419 0.581
```

---

breastmilk_formula_ensure

```
       n   missing  distinct
    2500      3107         3
```

```
false (2455, 0.982), I eat both solid food and formula/breast milk (19, 0.008), true (26,
0.010)
```

---

meat_eggs_frequency

```
        n  missing  distinct
     2620     2987         5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (871, 0.332), Never (129, 0.049), Occasionally (1-2 times/week) (444, 0.169), Rarely
(less than once/week) (154, 0.059), Regularly (3-5 times/week) (1022, 0.390)
```

homecooked_meals_frequency

```
        n  missing  distinct
     2630     2977         5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (1414, 0.538), Never (59, 0.022), Occasionally (1-2 times/week) (161, 0.061), Rarely
(less than once/week) (75, 0.029), Regularly (3-5 times/week) (921, 0.350)
```

ready_to_eat_meals_frequency

```
        n  missing  distinct
     2630     2977         5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (11, 0.004), Never (1554, 0.591), Occasionally (1-2 times/week) (265, 0.101), Rarely
(less than once/week) (724, 0.275), Regularly (3-5 times/week) (76, 0.029)
```

prepared_meals_frequency

```
        n  missing  distinct
     2633     2974         5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (65, 0.025), Never (226, 0.086), Occasionally (1-2 times/week) (952, 0.362), Rarely
(less than once/week) (1022, 0.388), Regularly (3-5 times/week) (368, 0.140)
```

whole_grain_frequency

```
        n  missing  distinct
     2618     2989         5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (433, 0.165), Never (396, 0.151), Occasionally (1-2 times/week) (569, 0.217), Rarely
(less than once/week) (495, 0.189), Regularly (3-5 times/week) (725, 0.277)
```

fruit_frequency

```
        n  missing  distinct
     2612     2995         5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (748, 0.286), Never (157, 0.060), Occasionally (1-2 times/week) (545, 0.209), Rarely
(less than once/week) (365, 0.140), Regularly (3-5 times/week) (797, 0.305)
```

vegetable_frequency

```
        n  missing  distinct
     2614     2993         5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (1342, 0.513), Never (31, 0.012), Occasionally (1-2 times/week) (240, 0.092), Rarely
(less than once/week) (72, 0.028), Regularly (3-5 times/week) (929, 0.355)
```

types_of_plants

```
       n    missing    distinct
    3023       2584           5

lowest : 11 to 20     21 to 30     6 to 10      Less than 5  More than 30
highest: 11 to 20     21 to 30     6 to 10      Less than 5  More than 30

Value         11 to 20     21 to 30      6 to 10  Less than 5 More than 30
Frequency          959          599          786          270          409
Proportion       0.317        0.198        0.260        0.089        0.135
```

fermented_plant_frequency

```
       n    missing    distinct
    2549       3058           5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (286, 0.112), Never (676, 0.265), Occasionally (1-2 times/week) (425, 0.167), Rarely
(less than once/week) (859, 0.337), Regularly (3-5 times/week) (303, 0.119)
```

milk_cheese_frequency

```
       n    missing    distinct
    2623       2984           5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (461, 0.176), Never (497, 0.189), Occasionally (1-2 times/week) (528, 0.201), Rarely
(less than once/week) (515, 0.196), Regularly (3-5 times/week) (622, 0.237)
```

milk_substitute_frequency

```
       n    missing    distinct
    2623       2984           5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (429, 0.164), Never (1197, 0.456), Occasionally (1-2 times/week) (241, 0.092), Rarely
(less than once/week) (468, 0.178), Regularly (3-5 times/week) (288, 0.110)
```

frozen_dessert_frequency

```
       n    missing    distinct
    2631       2976           5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (31, 0.012), Never (783, 0.298), Occasionally (1-2 times/week) (350, 0.133), Rarely
(less than once/week) (1356, 0.515), Regularly (3-5 times/week) (111, 0.042)
```

red_meat_frequency

```
       n    missing    distinct
    2615       2992           5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (74, 0.028), Never (398, 0.152), Occasionally (1-2 times/week) (965, 0.369), Rarely
(less than once/week) (588, 0.225), Regularly (3-5 times/week) (590, 0.226)
```

high_fat_red_meat_frequency

```
       n    missing    distinct
    2614       2993           5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (49, 0.019), Never (575, 0.220), Occasionally (1-2 times/week) (693, 0.265), Rarely
(less than once/week) (1006, 0.385), Regularly (3-5 times/week) (291, 0.111)
```

poultry_frequency

```
        n  missing  distinct
     2638     2969         5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (91, 0.034), Never (314, 0.119), Occasionally (1-2 times/week) (1073, 0.407), Rarely
(less than once/week) (359, 0.136), Regularly (3-5 times/week) (801, 0.304)
```

seafood_frequency

```
        n  missing  distinct
     2628     2979         5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (26, 0.010), Never (350, 0.133), Occasionally (1-2 times/week) (1040, 0.396), Rarely
(less than once/week) (850, 0.323), Regularly (3-5 times/week) (362, 0.138)
```

salted_snacks_frequency

```
        n  missing  distinct
     2633     2974         5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (152, 0.058), Never (397, 0.151), Occasionally (1-2 times/week) (744, 0.283), Rarely
(less than once/week) (907, 0.344), Regularly (3-5 times/week) (433, 0.164)
```

sugary_sweets_frequency

```
        n  missing  distinct
     2639     2968         5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (432, 0.164), Never (345, 0.131), Occasionally (1-2 times/week) (689, 0.261), Rarely
(less than once/week) (592, 0.224), Regularly (3-5 times/week) (581, 0.220)
```

olive_oil

```
        n  missing  distinct
     2617     2990         5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (596, 0.228), Never (213, 0.081), Occasionally (1-2 times/week) (541, 0.207), Rarely
(less than once/week) (332, 0.127), Regularly (3-5 times/week) (935, 0.357)
```

whole_eggs

```
        n  missing  distinct
     2639     2968         5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (332, 0.126), Never (251, 0.095), Occasionally (1-2 times/week) (934, 0.354), Rarely
(less than once/week) (462, 0.175), Regularly (3-5 times/week) (660, 0.250)
```

sugar_sweetened_drink_frequency

```
        n  missing  distinct
     2612     2995         5

lowest : Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                        Never                        Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (32, 0.012), Never (1948, 0.746), Occasionally (1-2 times/week) (118, 0.045), Rarely
(less than once/week) (470, 0.180), Regularly (3-5 times/week) (44, 0.017)
```

one_liter_of_water_a_day_frequency

```
     n  missing  distinct
  2603     3004         5
```

```
lowest : Daily                         Never                         Occasionally (1-2 times/week) Rarely (less than once/week)
highest: Daily                         Never                         Occasionally (1-2 times/week) Rarely (less than once/week)

Daily (1302, 0.500), Never (111, 0.043), Occasionally (1-2 times/week) (318, 0.122), Rarely
(less than once/week) (234, 0.090), Regularly (3-5 times/week) (638, 0.245)
```

Variables with all observations missing: country_residence, specialized_diet_exclude_dairy, specialized_diet_exclude_nightshad
specialized_diet_exclude_refined_sugars, specialized_diet_fodmap, specialized_diet_halaal, specialized_diet_i_do_not_eat_a
specialized_diet_kosher, specialized_diet_modified_paleo_diet, specialized_diet_other_restrictions_not_described_here,
specialized_diet_paleodiet_or_primal_diet, specialized_diet_raw_food_diet, specialized_diet_unspecified, specialized_diet
pets_other, drinks_per_session, mental_illness, mental_illness_type_anorexia_nervosa, mental_illness_type_bipolar_disorder
mental_illness_type_bulimia_nervosa, mental_illness_type_depression, mental_illness_type_ptsd_posttraumatic_stress_disor
mental_illness_type_schizophrenia, mental_illness_type_substance_abuse, mental_illness_type_unspecified, diabetes_type,
ibd_diagnosis, ibd_diagnosis_refined, vivid_dreams, artificial_sweeteners

2.2 Normality Test of Numerical Variables

2.2.1 Statistics and Visualization of (Sample) Data

subject_id

normality test : Shapiro-Wilk normality test
statistic : 0.94829, p-value : 9.28396E-39

type	skewness	kurtosis
original	0.0065	1.7460
log transformation	-0.6394	2.3179
sqrt transformation	-0.2937	1.8895

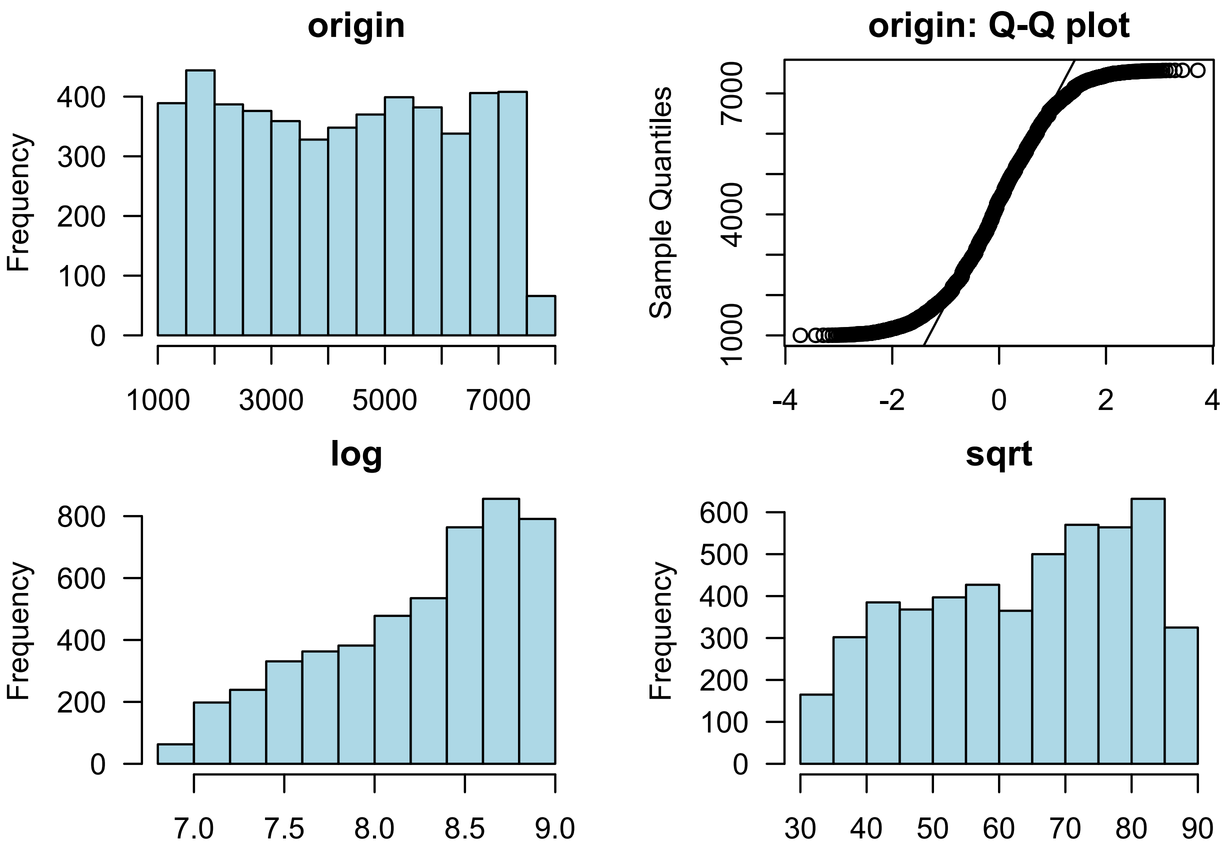


Figure 2.1: subject_id

age_years

normality test : Shapiro-Wilk normality test
statistic : 0.9729, p-value : 8.93971E-30

type	skewness	kurtosis
original	-0.5156	2.8372
log transformation	-2.7523	12.2237
sqrt transformation	-1.3302	4.9328

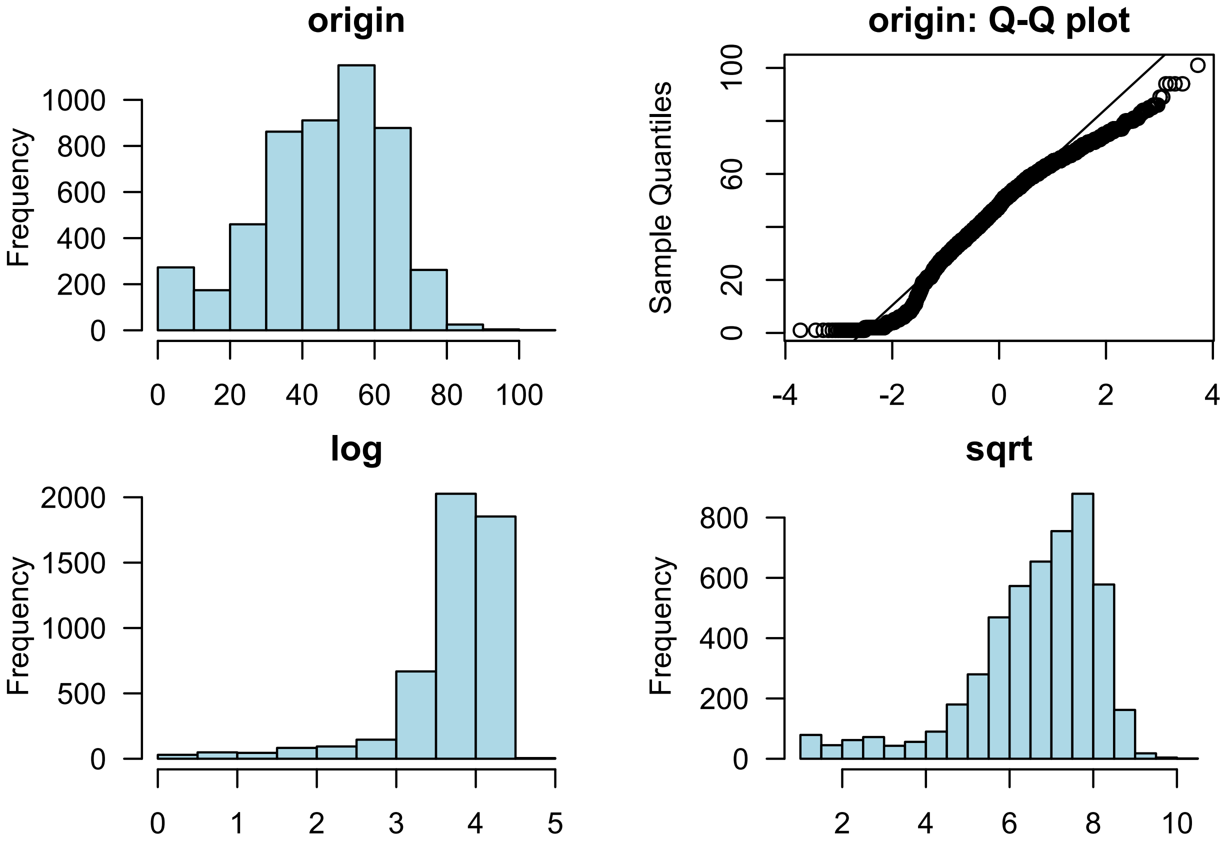


Figure 2.2: age_years

bmi

normality test : Shapiro-Wilk normality test
statistic : 0.04341, p-value : 8.36222E-95

type	skewness	kurtosis
original	23.6524	638.5686
log transformation		
sqrt transformation	12.5037	203.8279

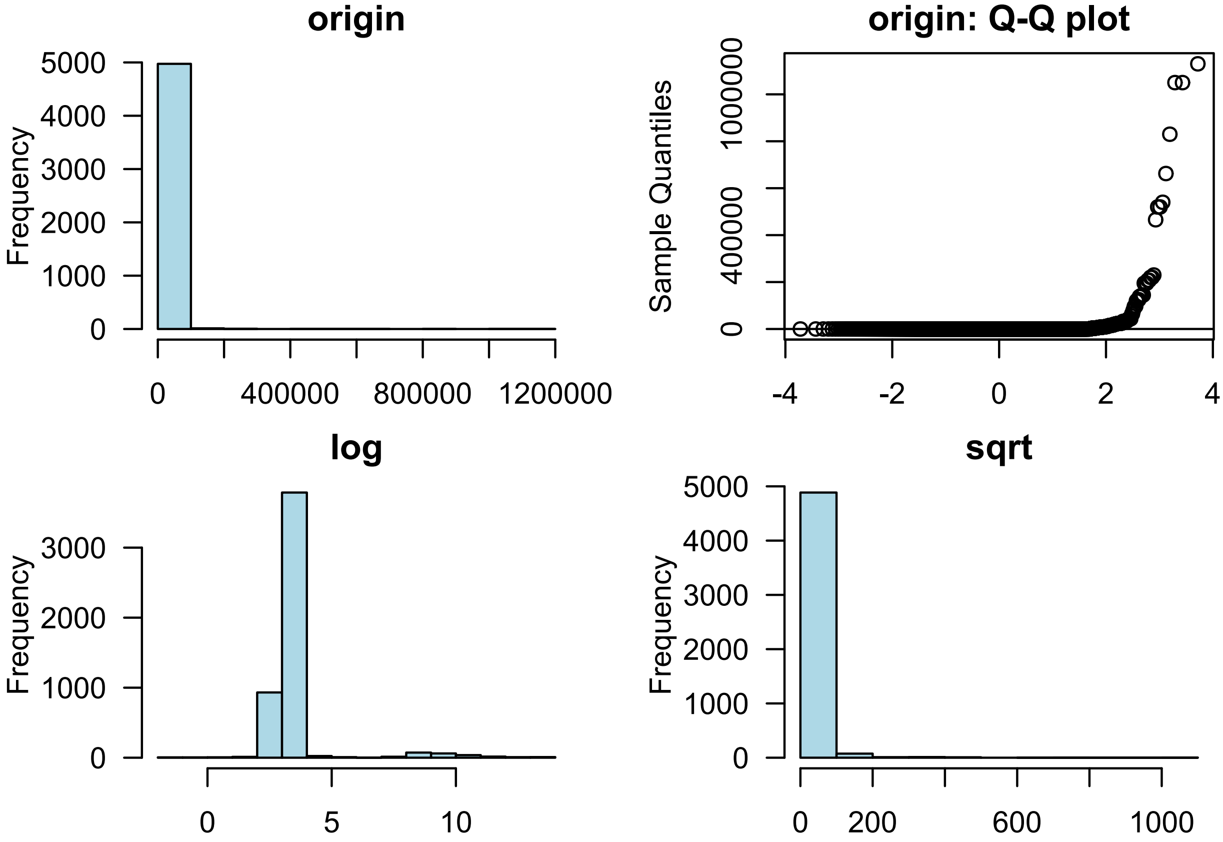


Figure 2.3: bmi

height_cm

normality test : Shapiro-Wilk normality test
statistic : 0.05416, p-value : 1.54261E-94

type	skewness	kurtosis
original	55.1508	3523.4326
log transformation	-3.9082	21.3513
sqrt transformation	9.7591	325.9280

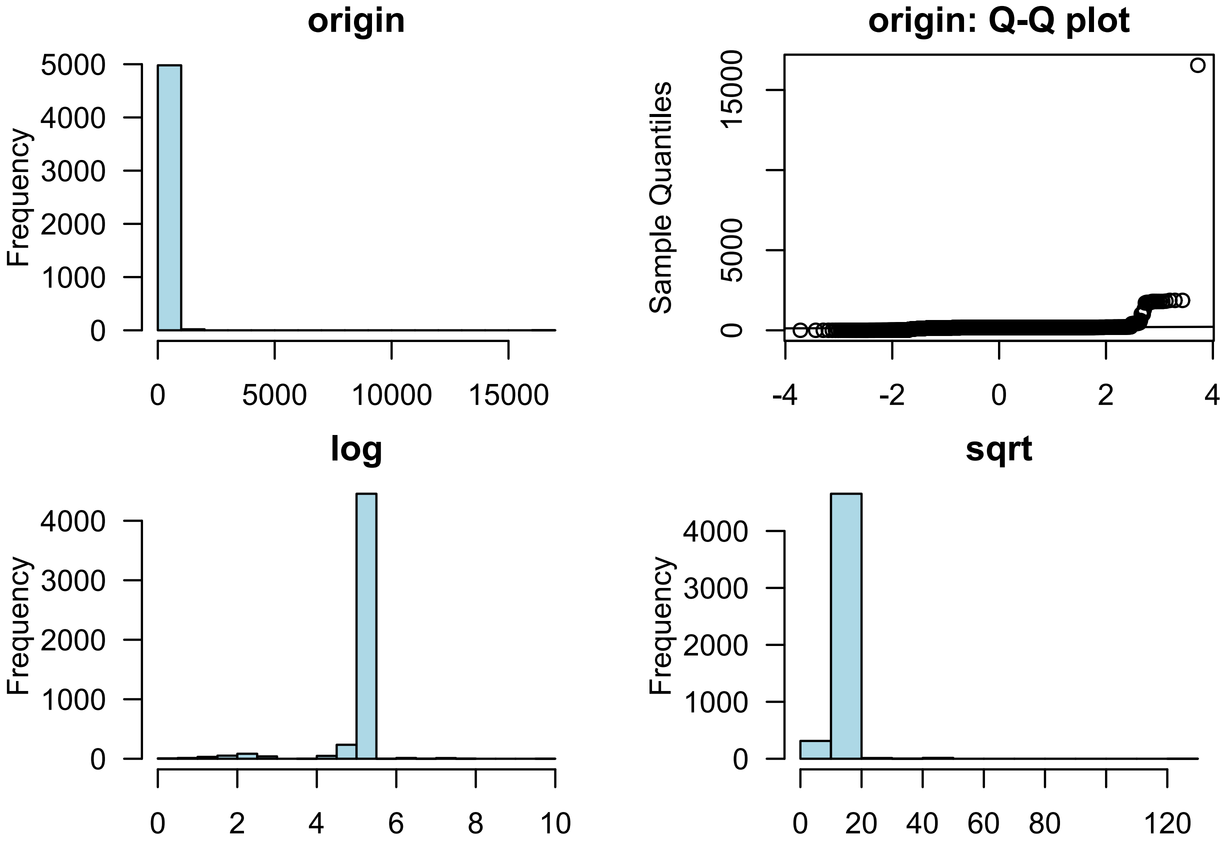


Figure 2.4: height_cm

weight_kg

normality test : Shapiro-Wilk normality test
statistic : 0.50726, p-value : 7.82391E-80

type	skewness	kurtosis
original	11.1069	233.0895
log transformation	-1.5857	12.8931
sqrt transformation	2.3268	35.0952

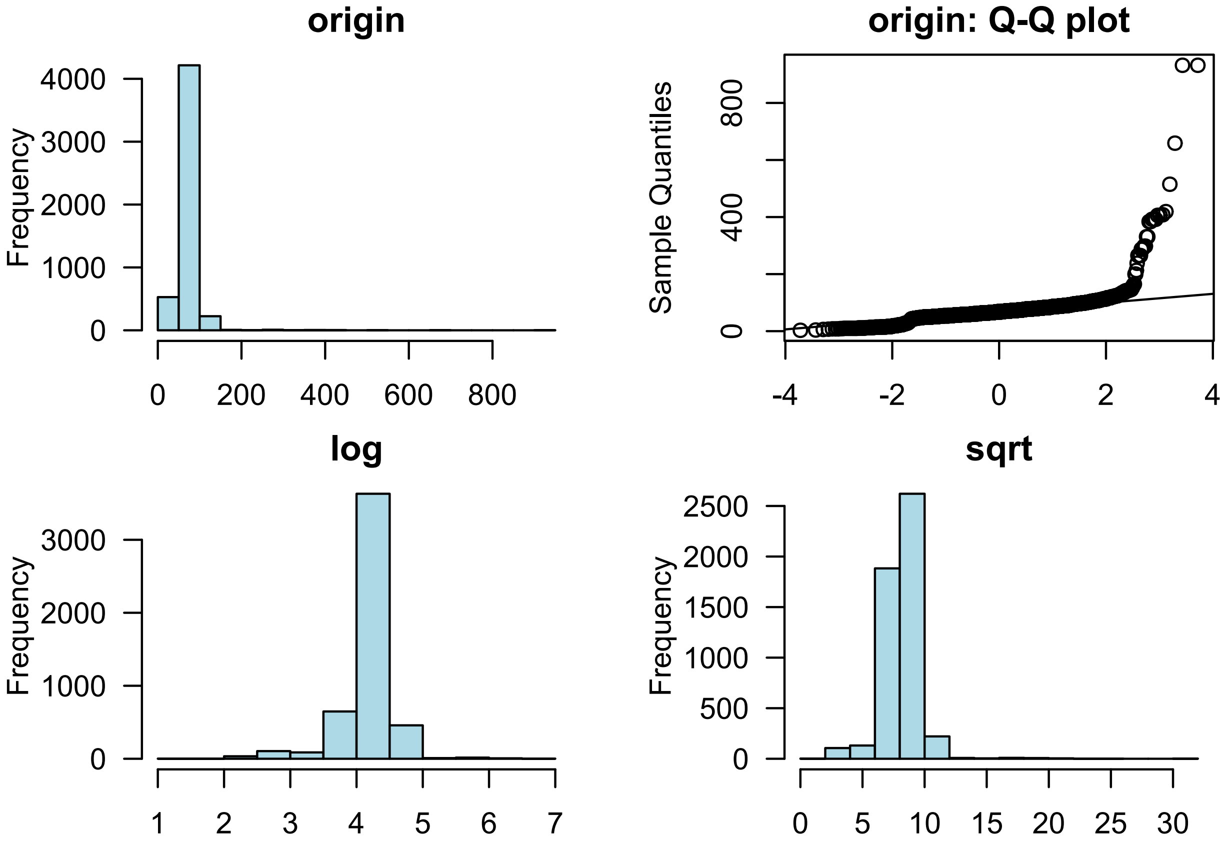


Figure 2.5: weight_kg

Chapter 3

Relationship Between Variables

3.1 Correlation Coefficient

3.1.1 Correlation Coefficient by Variable Combination

No correlation coefficient is greater than 0.5.

3.1.2 Correlation Plot of Numerical Variables

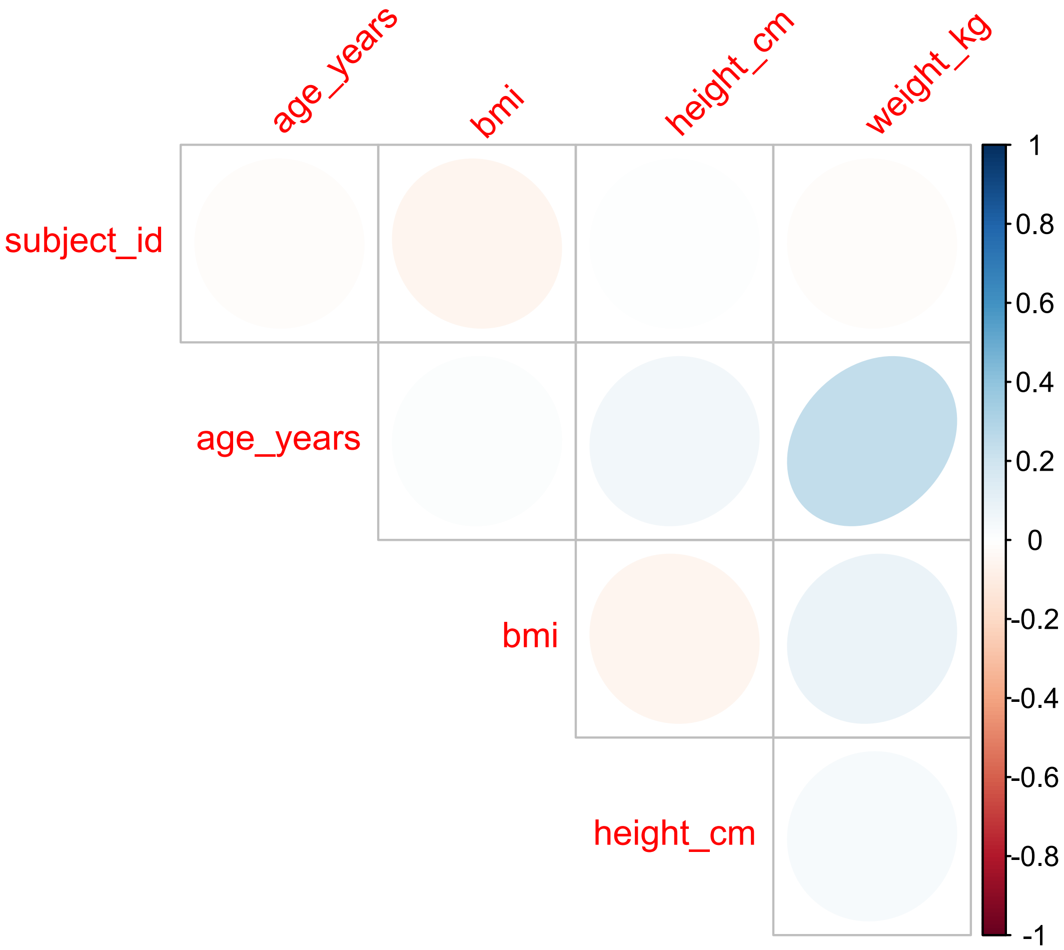


Figure 3.1: The correlation coefficient of numerical variables

Chapter 4

Target based Analysis

4.1 Grouped Descriptive Statistics

4.1.1 Grouped Numerical Variables

There is no target variable.

4.1.2 Grouped Categorical Variables

There is no target variable.

4.2 Grouped Relationship Between Variables

4.2.1 Grouped Correlation Coefficient

There is no target variable.

4.2.2 Grouped Correlation Plot of Numerical Variables

There is no target variable.