REPORT SERIES WITH DLOOKR

Exploratory Data Analysis Report

Author:
dlookr package

Version:
0.3.12

November 12, 2020

Contents

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 218 observations and 12 variables.

1.2 Information of Variables

Table 1.1: Information of Variables

variables	types	missing_count	missing_percent	unique_count	unique_rate
countrycode	character	0	0.000000	218	1.0000000
logGDPpc2000	numeric	24	11.009174	195	0.8944954
logGDPpc2015	numeric	18	8.256881	201	0.9220183
growthGDPpc	numeric	26	11.926606	193	0.8853211
imalaria2000	numeric	119	54.587156	100	0.4587156
imalaria2015	numeric	119	54.587156	86	0.3944954
change_malar	numeric	119	54.587156	98	0.4495413
educ_sec	numeric	119	54.587156	100	0.4587156
life2000	numeric	17	7.798165	201	0.9220183
trade2000	numeric	40	18.348624	179	0.8211009
gov2000	numeric	25	11.467890	190	0.8715596
invest_growth	numeric	91	41.743119	128	0.5871560

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.

Chapter 2

Univariate Analysis

2.1 Descriptive Statistics

edaData
12 Variables 218 Observations

---

countrycode : Country Code Format:%9s

n	missing	distinct
218	0	218

```
lowest : ABW AFG AGO ALB AND, highest: XKX YEM ZAF ZMB ZWE
```

---

logGDPpc2000 : Log GDPpc 2000 Format:%9.0g

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
194	24	194	1	8.39	1.8	6.035	6.285	7.195	8.222	9.618	10.601	10.810

```
lowest :  5.285396  5.434622  5.667314  5.669834  5.711302
highest: 11.289637 11.310238 11.410585 11.445320 11.840078
```

---

logGDPpc2015 : Log GDPpc 2015 Format:%9.0g

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
200	18	200	1	8.719	1.69	6.353	6.618	7.609	8.745	9.794	10.732	10.917

```
lowest :  5.431241  5.847467  5.956626  6.018643  6.089357
highest: 11.245743 11.307289 11.407891 11.586531 12.151471
```

---

growthGDPpc : Av. GDPpc growth 2015-2000 Format:%9.0g

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50
192	26	192	1	2.276	2.355	-0.7001	-0.1279	0.8492	1.9916
.75	.90	.95							
3.5759	5.2073	5.8420							

```
lowest : -3.013770 -2.788296 -2.009163 -1.677084 -1.348502
highest:  6.849905  6.890720  8.641517  8.664331  9.077437
```

---

imalaria2000 : Malaria incidence rate in 2000 Format:%9.0g

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
99	119	99	1	159.9	201	0.628	2.594	8.105	57.310	305.990	434.374	500.148

```
lowest :   0.01   0.30   0.31   0.38   0.61, highest: 514.62 589.59 607.11 623.24 796.38
```

---

imalaria2015 : Malaria incidence rate in 2015 Format:%9.0g

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50
99	119	85	0.997	87.99	126.4	0.000	0.000	0.205	9.970
.75	.90	.95							
153.215	329.172	372.290							

```
lowest :   0.00   0.03   0.04   0.05   0.06, highest: 376.88 377.65 386.75 391.18 429.04
```

---

change_malar : Change in incidence rate 2015-2000 Format:%9.0g

```
       n  missing distinct     Info     Mean      Gmd      .05      .10      .25      .50
      99      119       97        1   -71.93    113.7 -266.853 -189.738 -105.440  -16.770
     .75      .90      .95
  -3.315   -0.570    0.944

lowest : -796.380005 -588.920044 -554.609985 -367.669983 -282.720001
highest:    9.529999   11.460001   55.079987   59.889996  195.129990
```

---

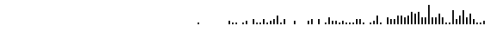
educ_sec : % Population 25+ with lower sec. school Format:%9.0g

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
99	119	99	1	61.4	29.69	16.86	26.23	44.35	61.70	86.11	93.01	97.17

```
lowest :  5.98647  8.57250  9.95948 11.51888 14.31745, highest: 97.32744 97.48644 99.63876 99.82408 99.86594
```

---

life2000 : Life expectancy 2000 Format:%9.0g

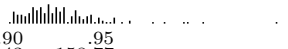


n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
201	17	200	1	67.01	11.19	47.71	50.49	60.06	70.18	74.40	77.95	79.04

```
lowest : 39.44100 44.00000 44.19000 44.64900 45.09000, highest: 79.68049 79.77805 80.40700 80.87805 81.07610
```

---

trade2000 : Trade openess in % GDP 2000 Format:%9.0g



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
178	40	178	1	84.15	49.68	27.00	33.98	52.65	74.77	104.91	137.43	158.77

```
lowest :   1.165696  19.819653  22.553724  22.622444  22.639761
highest: 220.406784 245.862244 247.654022 271.950958 364.364532
```

---

gov2000 : Government Effectiveness Índex 2000 Format:%9.0g

```
       n  missing distinct     Info     Mean      Gmd      .05      .10      .25      .50
     193       25      189        1 -0.01016    1.131  -1.3713  -1.1556  -0.7374  -0.1861
     .75      .90      .95
  0.6253   1.7004   1.9300

lowest : -2.231651 -2.195506 -1.977284 -1.884151 -1.836376
highest:  2.030386  2.042147  2.070121  2.118074  2.199069
```

---

invest_growth : Mean growth gross capital formation Format:%9.0g


```
       n  missing distinct     Info     Mean      Gmd      .05      .10      .25      .50
     127       91      127        1     5.16    4.795  -0.5900   0.2674   2.1334   4.6793
     .75      .90      .95
  7.5833  10.9392  12.2072

lowest : -5.363957 -2.987212 -2.567656 -2.147890 -1.331013
highest: 13.047798 13.729401 14.678500 18.315086 25.512703
```

2.2 Normality Test of Numerical Variables

2.2.1 Statistics and Visualization of (Sample) Data

logGDPpc2000

normality test : Shapiro-Wilk normality test
statistic : 0.97382, p-value : 0.00107378

type	skewness	kurtosis
original	0.1427	2.0782
log transformation	-0.1674	2.1642
sqrt transformation	-0.0094	2.0926

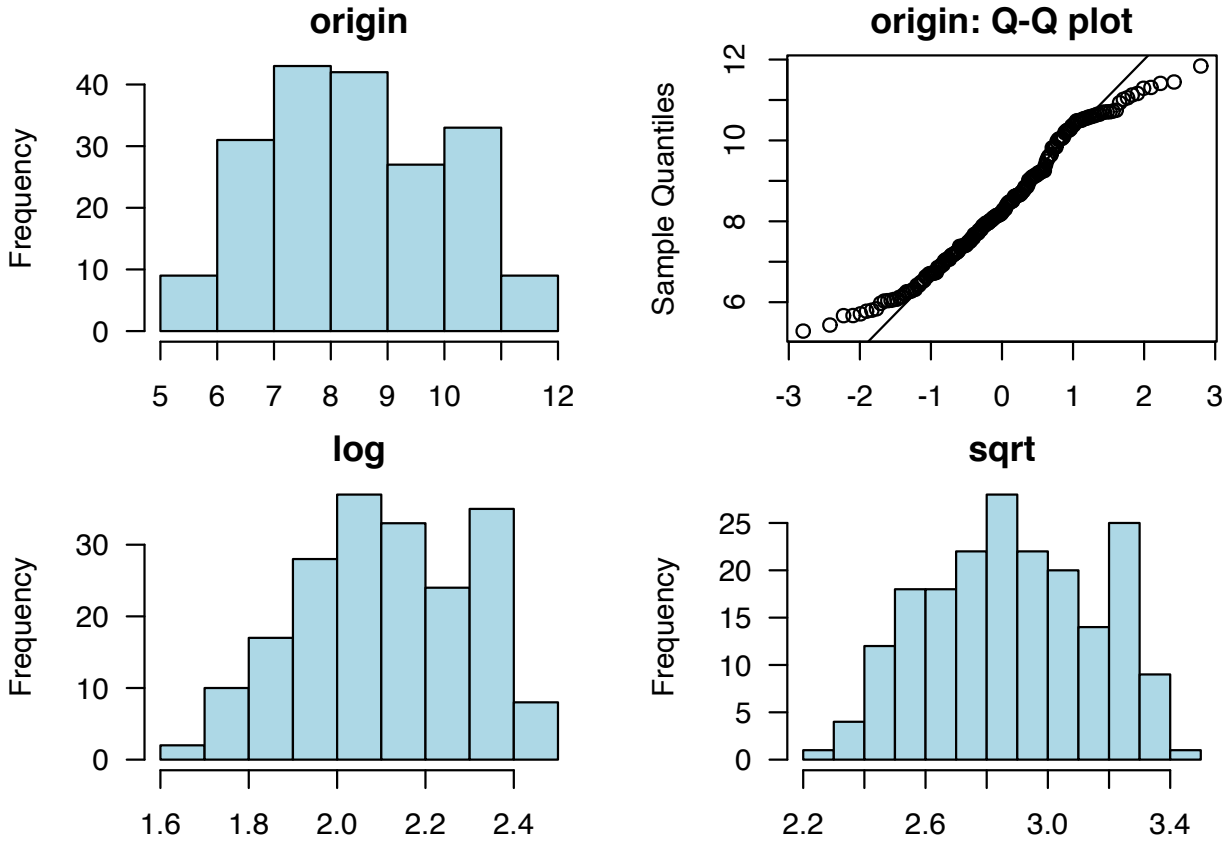


Figure 2.1: logGDPpc2000

logGDPpc2015

normality test : Shapiro-Wilk normality test
statistic : 0.98126, p-value : 0.0089949

type	skewness	kurtosis
original	-0.0268	2.1397
log transformation	-0.3263	2.2943
sqrt transformation	-0.1746	2.1871

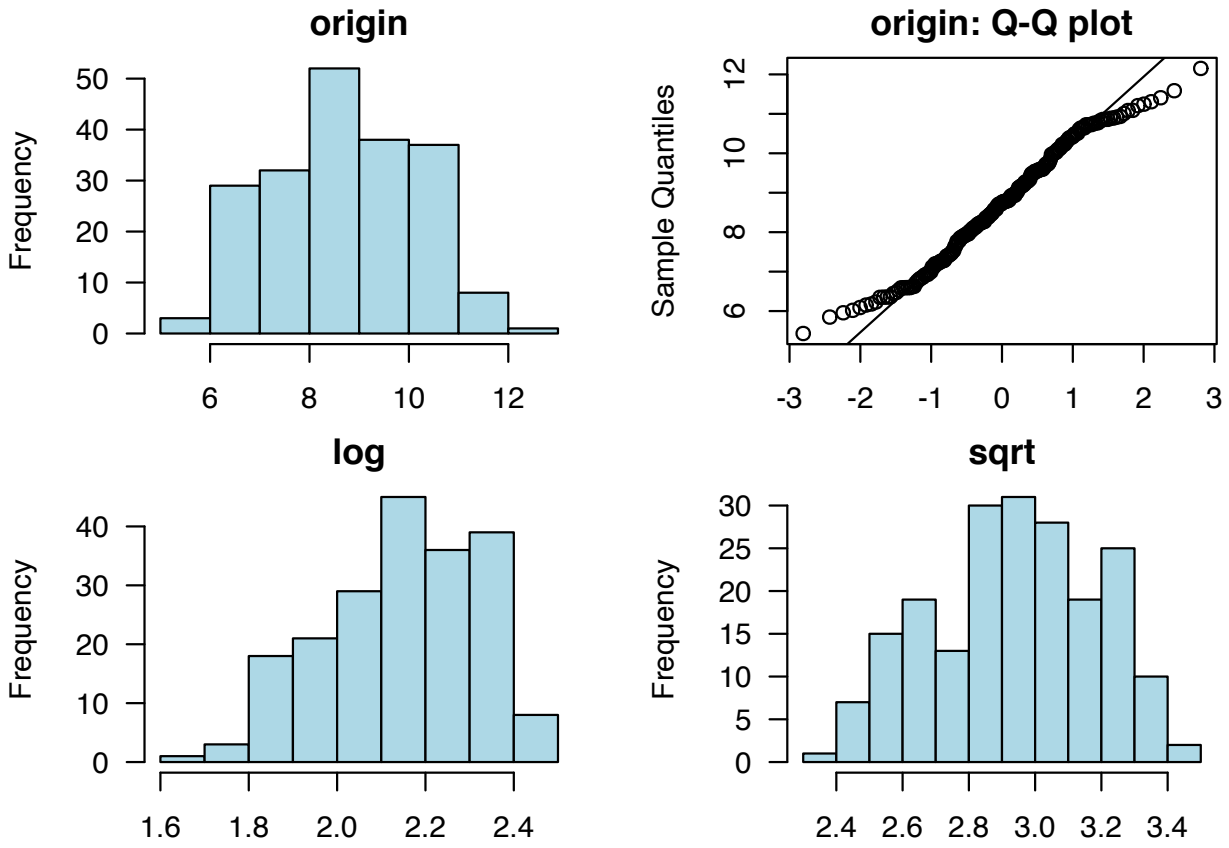


Figure 2.2: logGDPpc2015

growthGDPpc

normality test : Shapiro-Wilk normality test
statistic : 0.98337, p-value : 0.0225057

type	skewness	kurtosis
original	0.4715	3.4139
log transformation	-1.4207	6.0170
sqrt transformation	0.0311	2.5905

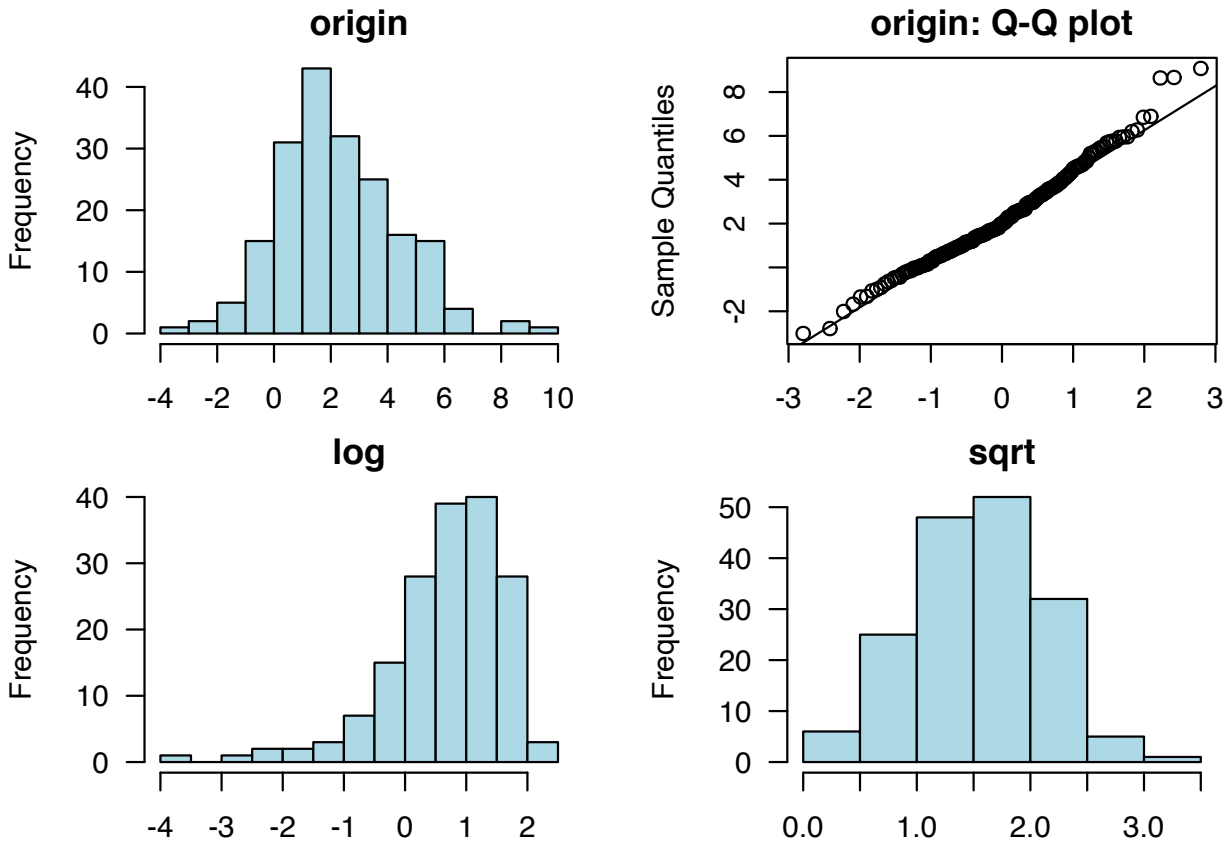


Figure 2.3: growthGDPpc

imalaria2000

normality test : Shapiro-Wilk normality test
statistic : 0.80376, p-value : 3.71679E-10

type	skewness	kurtosis
original	1.0566	3.1271
log transformation	-0.7556	3.3193
sqrt transformation	0.4679	1.7609

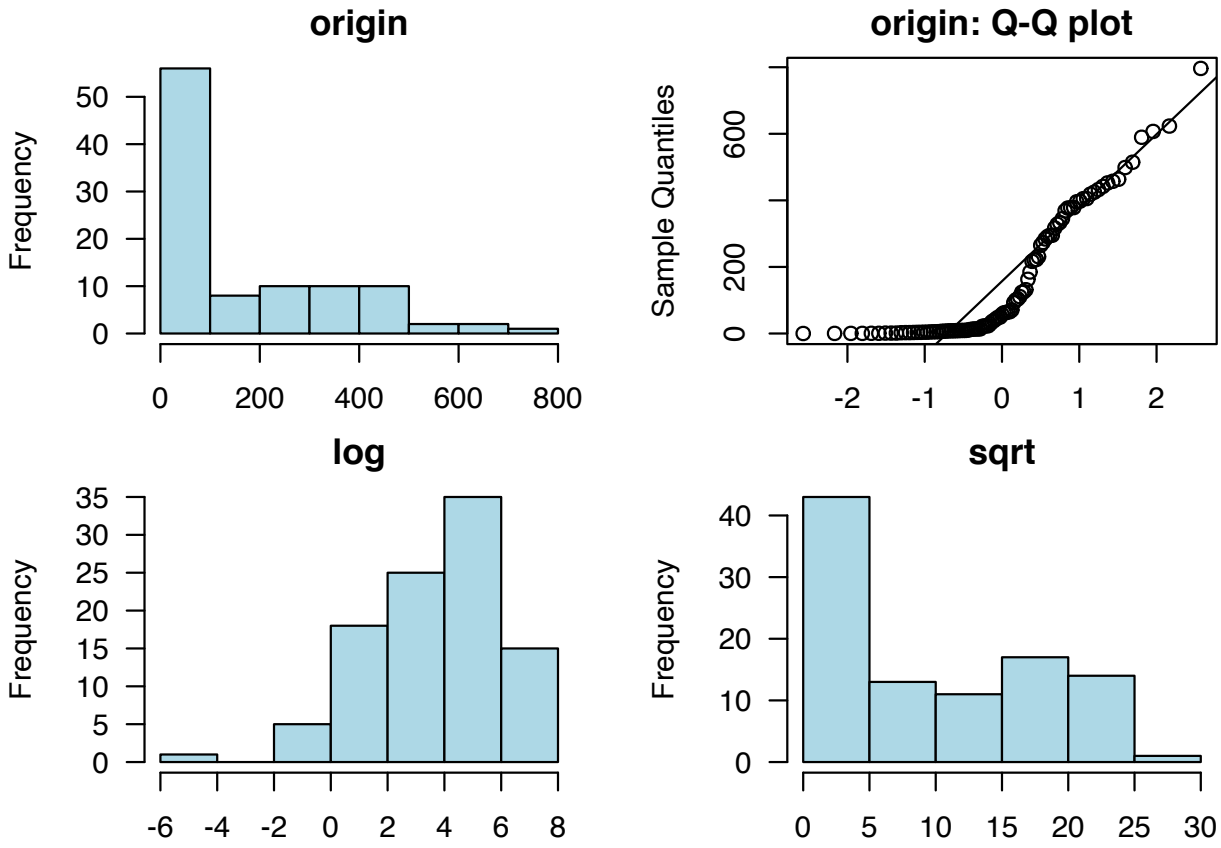


Figure 2.4: imalaria2000

imalaria2015

normality test : Shapiro-Wilk normality test
statistic : 0.71267, p-value : 1.26255E-12

type	skewness	kurtosis
original	1.2888	3.2152
log transformation		
sqrt transformation	0.7632	2.0458

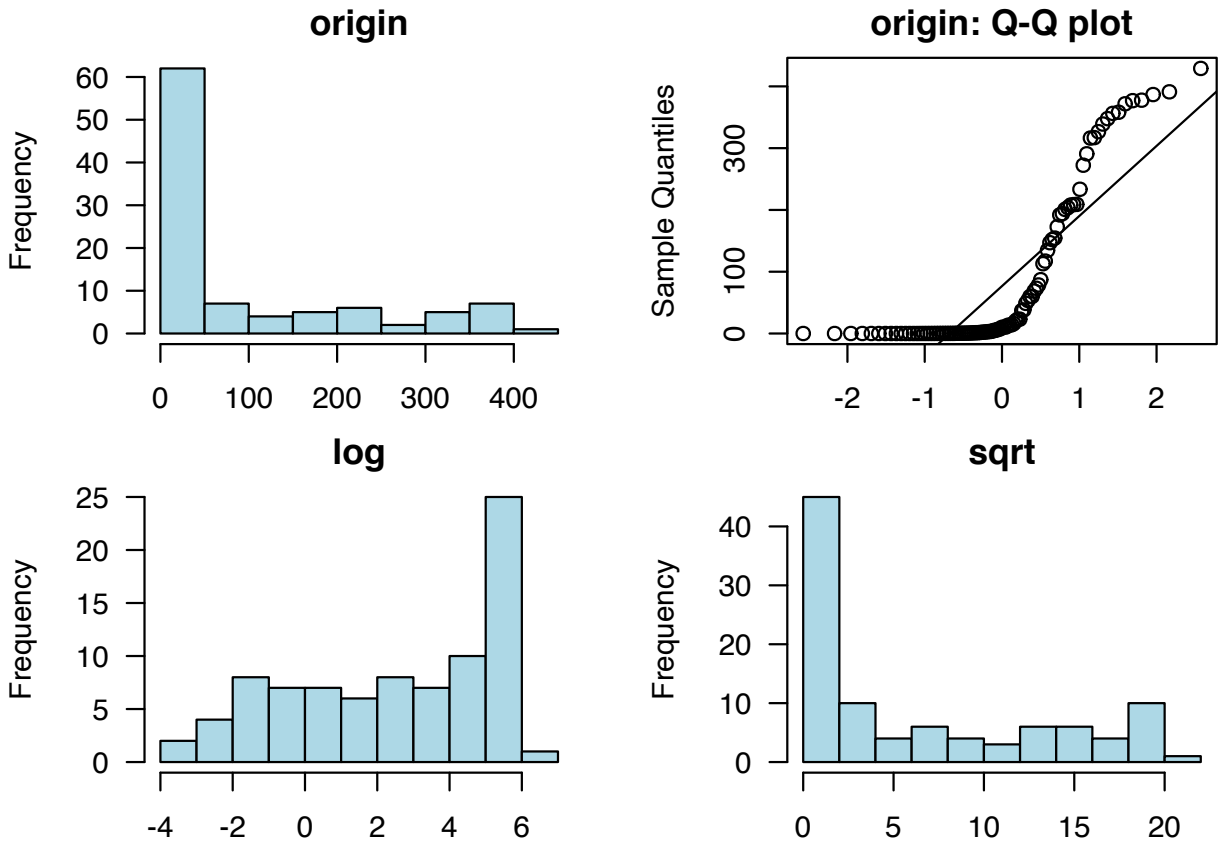


Figure 2.5: imalaria2015

change_malar

normality test : Shapiro-Wilk normality test
statistic : 0.67302, p-value : 1.56668E-13

type	skewness	kurtosis
original	-2.9067	14.1567
log transformation	0.0853	1.6011
sqrt transformation	0.6666	2.2084

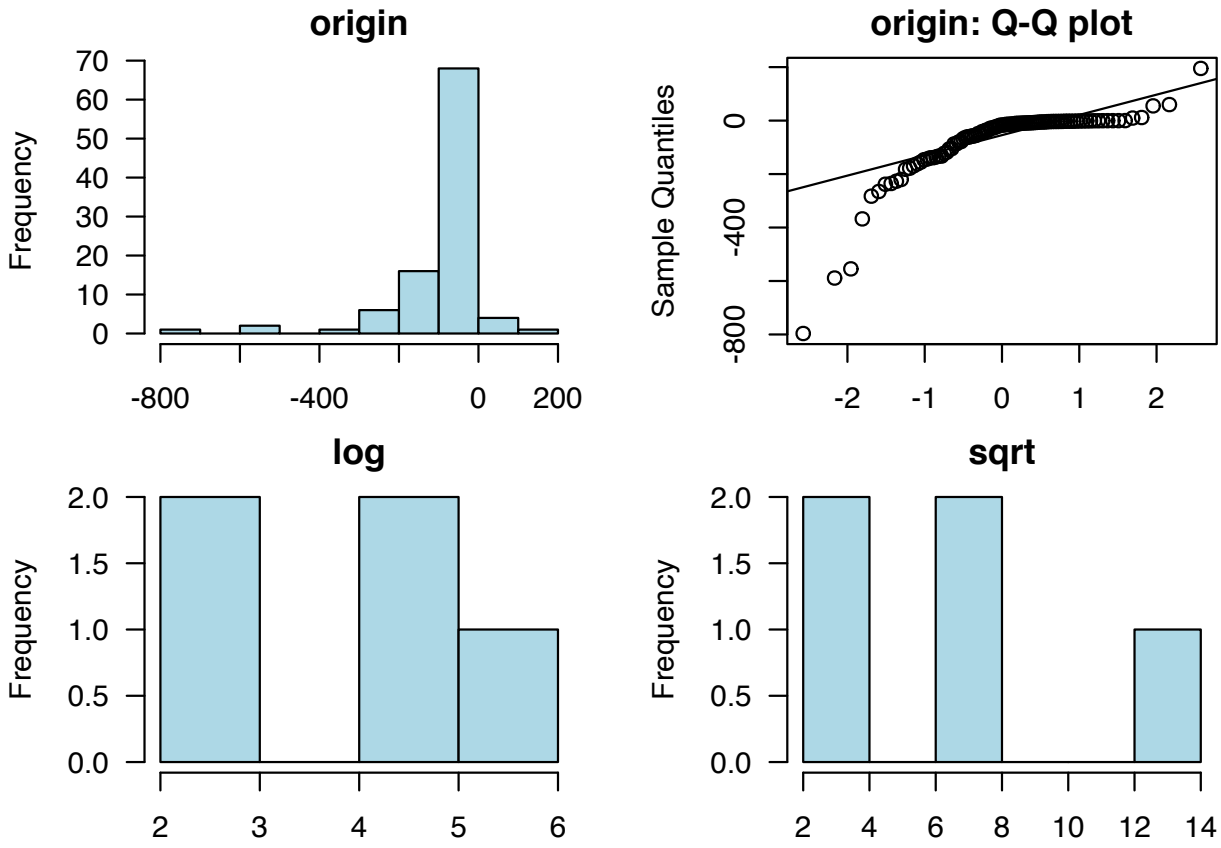


Figure 2.6: change_malar

educ_sec

normality test : Shapiro-Wilk normality test
statistic : 0.95245, p-value : 0.00129114

type	skewness	kurtosis
original	-0.2543	2.0154
log transformation	-1.4991	5.3899
sqrt transformation	-0.7400	2.9183

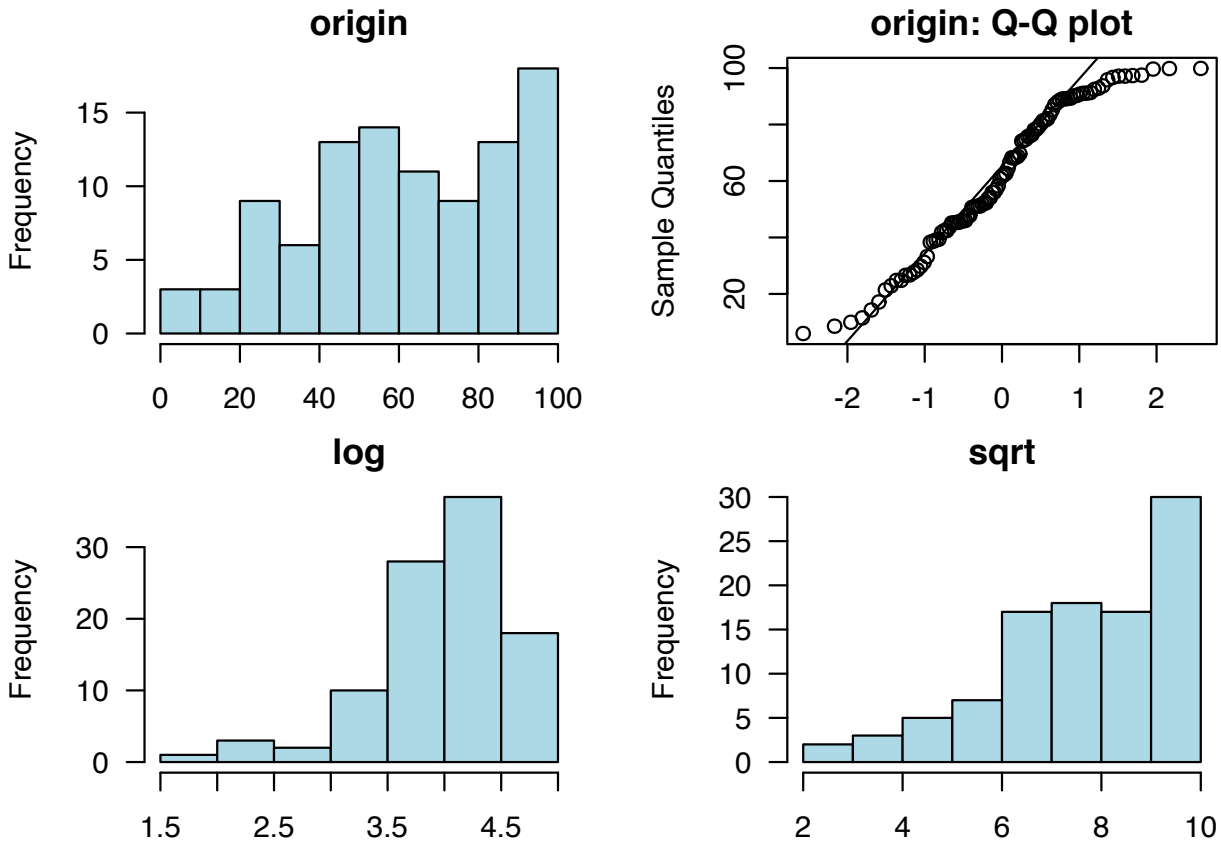


Figure 2.7: educ_sec

life2000

normality test : Shapiro-Wilk normality test
statistic : 0.91018, p-value : 1.0976E-09

type	skewness	kurtosis
original	-0.7728	2.4996
log transformation	-1.0014	2.9791
sqrt transformation	-0.8848	2.7134

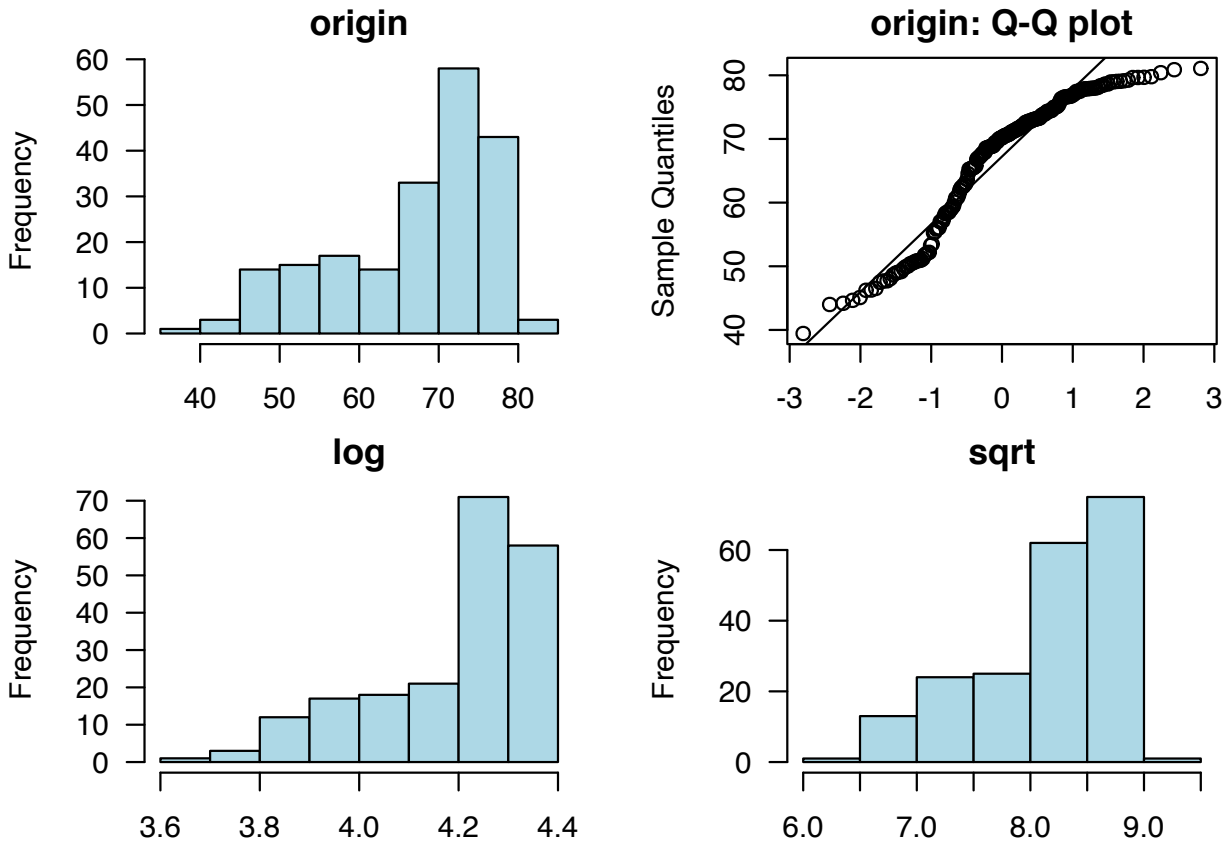


Figure 2.8: life2000

trade2000

normality test : Shapiro-Wilk normality test
statistic : 0.85759, p-value : 6.95547E-12

type	skewness	kurtosis
original	1.9988	9.9260
log transformation	-1.6383	12.9331
sqrt transformation	0.6646	4.8121

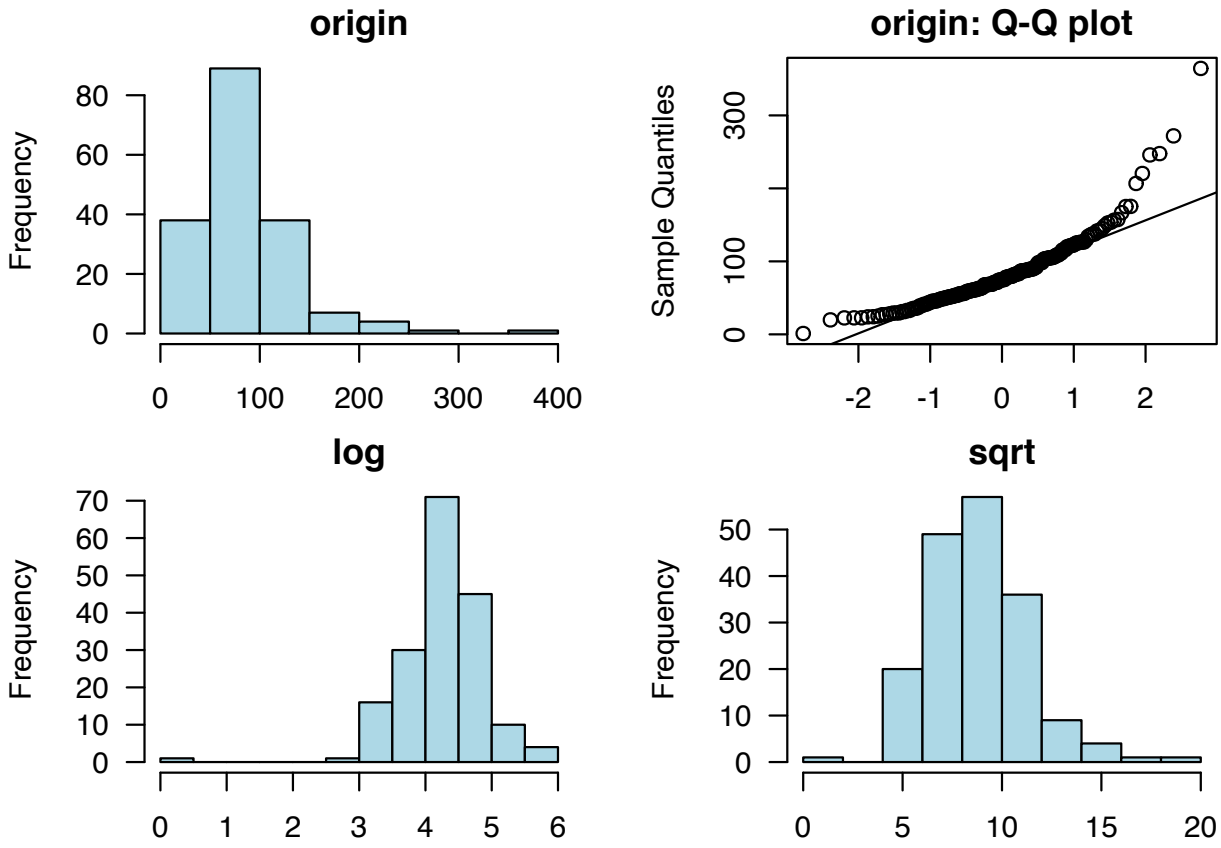


Figure 2.9: trade2000

gov2000

normality test : Shapiro-Wilk normality test
statistic : 0.9637, p-value : 6.99535E-05

type	skewness	kurtosis
original	0.4311	2.5550
log transformation	-1.6500	6.3651
sqrt transformation	-0.2052	2.0131

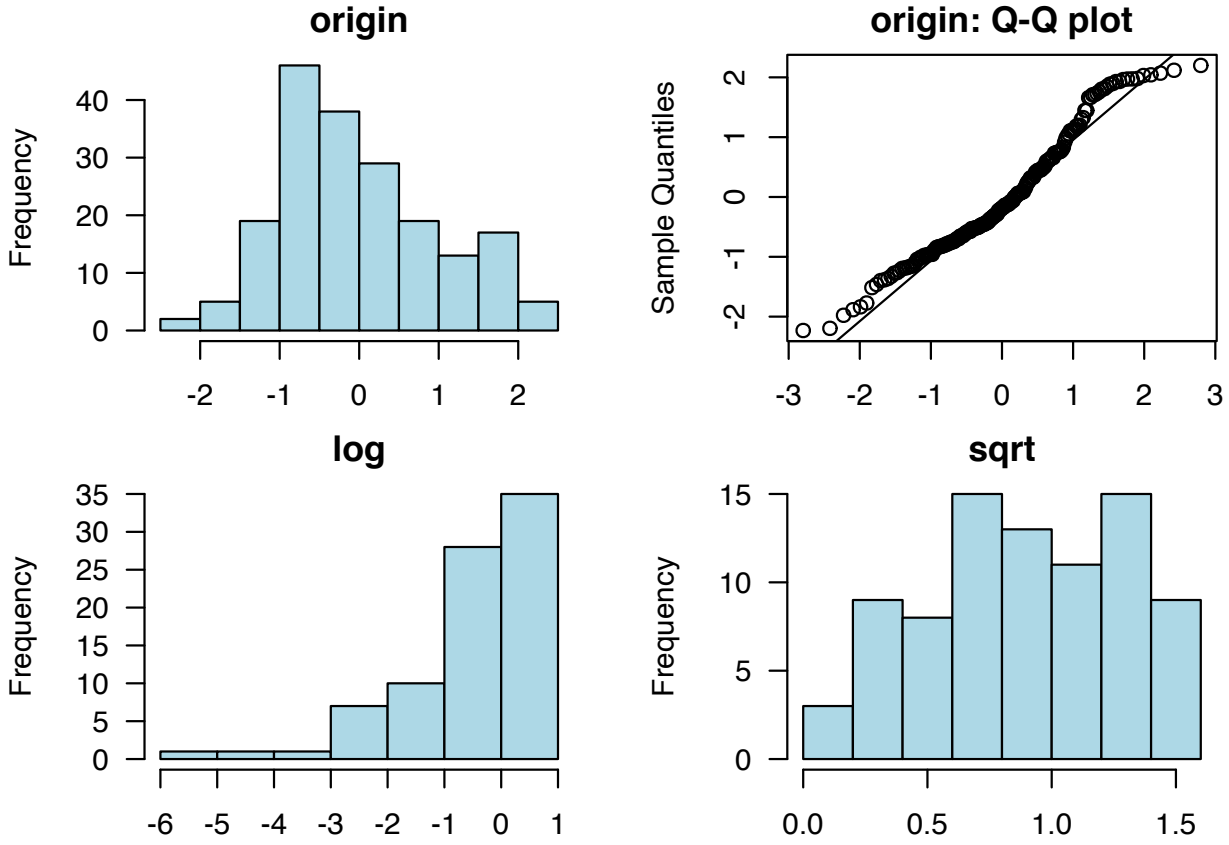


Figure 2.10: gov2000

invest_growth

normality test : Shapiro-Wilk normality test
statistic : 0.95295, p-value : 0.000231148

type	skewness	kurtosis
original	1.0003	5.8176
log transformation	-1.2261	5.6779
sqrt transformation	0.2515	3.2722

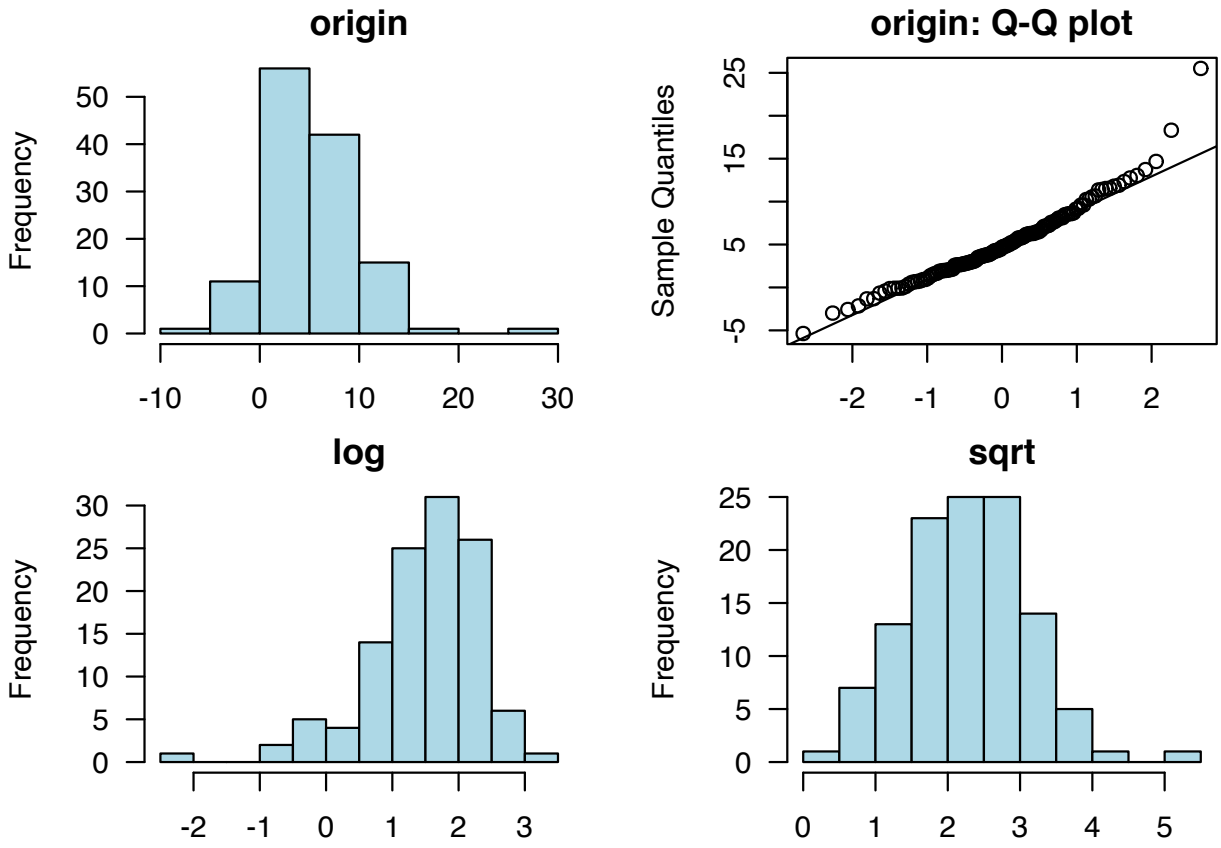


Figure 2.11: invest_growth

Chapter 3

Relationship Between Variables

3.1 Correlation Coefficient

3.1.1 Correlation Coefficient by Variable Combination

Table 3.1: The correlation coefficients (0.5 or more)

Variable1	Variable2	Correlation Coefficient
logGDPpc2015	logGDPpc2000	0.980
gov2000	logGDPpc2000	0.815
life2000	logGDPpc2015	0.812
gov2000	logGDPpc2015	0.808
life2000	logGDPpc2000	0.790
change_malar	imalaria2000	-0.743
imalaria2015	imalaria2000	0.726
life2000	imalaria2015	-0.702
gov2000	life2000	0.673
invest_growth	logGDPpc2000	-0.566
life2000	imalaria2000	-0.566
invest_growth	life2000	-0.541
imalaria2015	logGDPpc2015	-0.532
invest_growth	logGDPpc2015	-0.514
educ_sec	imalaria2015	-0.514
invest_growth	imalaria2015	0.503

3.1.2 Correlation Plot of Numerical Variables

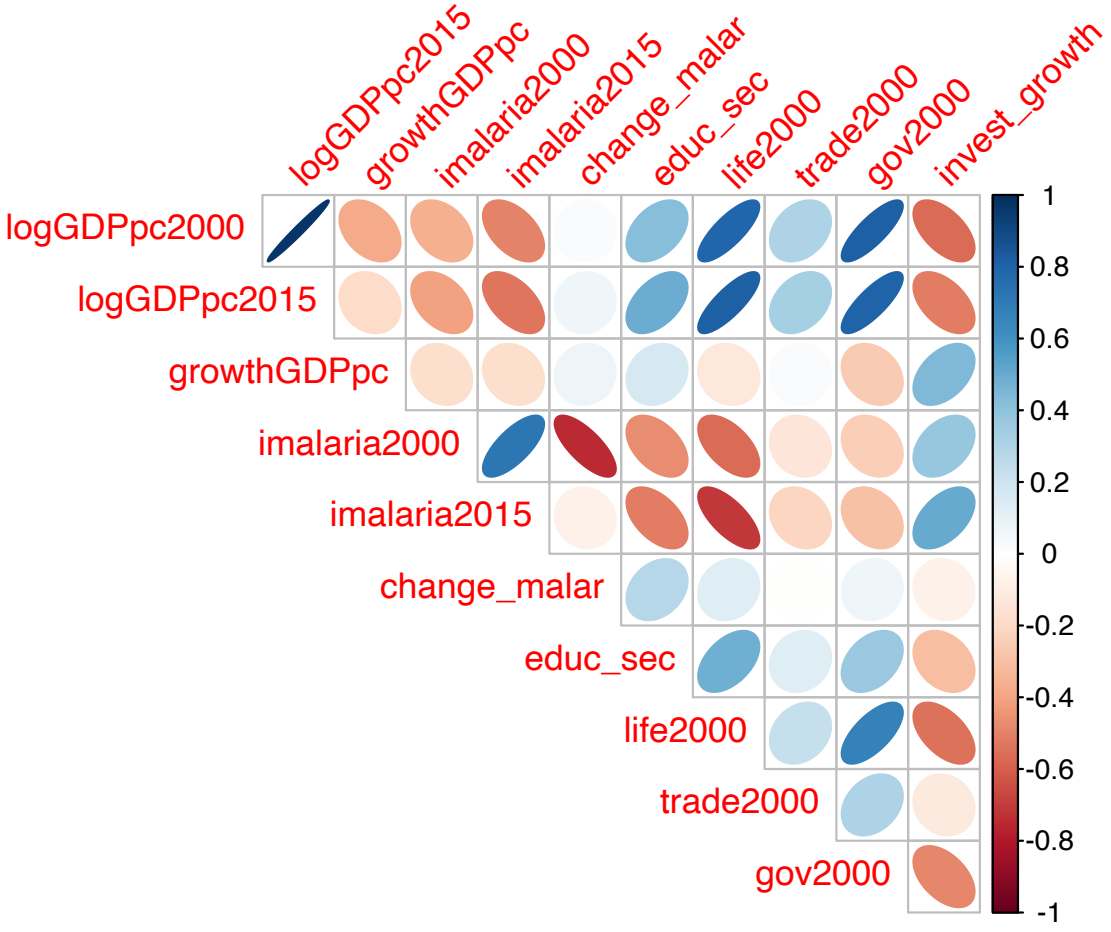


Figure 3.1: The correlation coefficient of numerical variables

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.