REPORT SERIES WITH DLOOKR

Exploratory Data Analysis Report

Author:
dlookr package

Version:
0.4.0

October 19, 2021

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 28,534 observations and 21 variables.

1.2 Information of Variables

Table 1.1: Information of Variables

variables	types	missing_count	missing_percent	unique_count	unique_rate
idcode	numeric	0	0.00	4711	0.165
year	numeric	0	0.00	15	0.001
birth_yr	numeric	0	0.00	14	0.000
age	numeric	24	0.08	34	0.001
race	haven_labelled	0	0.00	3	0.000
msp	numeric	16	0.06	3	0.000
nev_mar	numeric	16	0.06	3	0.000
grade	numeric	2	0.01	20	0.001
collgrad	numeric	0	0.00	2	0.000
not_smsa	numeric	8	0.03	3	0.000
c_city	numeric	8	0.03	3	0.000
south	numeric	8	0.03	3	0.000
ind_code	numeric	341	1.20	13	0.000
occ_code	numeric	121	0.42	14	0.000
union	numeric	9296	32.58	3	0.000
wks_ue	numeric	5704	19.99	62	0.002
ttl_exp	numeric	0	0.00	4744	0.166
tenure	numeric	433	1.52	271	0.009
hours	numeric	67	0.23	86	0.003
wks_work	numeric	703	2.46	106	0.004
ln_wage	numeric	0	0.00	8173	0.286

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

```
Error in proxy[, ..., drop = FALSE]: incorrect number of dimensions
Error in Hmisc::latex(x, file = ""):  object 'x' not found
```

2.2 Normality Test of Numerical Variables

2.2.1 Statistics and Visualization of (Sample) Data

idcode

* normality test : Shapiro-Wilk normality test
- statistic : 0.95505, p-value : 1.04775E-36

Table 2.1: skewness and kurtosis : idcode

type	skewness	kurtosis
original	-0.0221	1.8114
log transformation	-2.1511	9.8380
sqrt transformation	-0.6090	2.4903

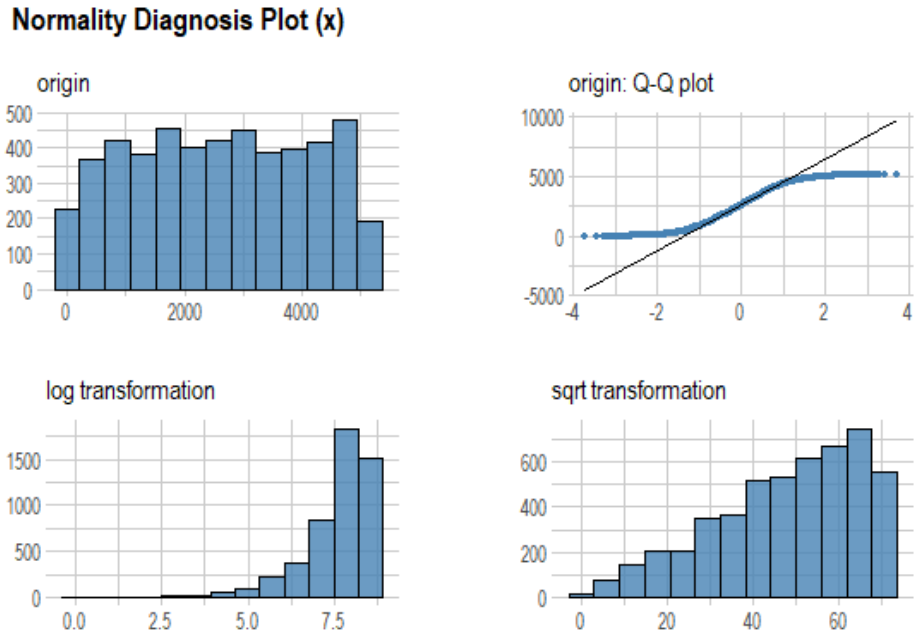

Figure 2.1: idcode

year

* normality test : Shapiro-Wilk normality test
- statistic : 0.93115, p-value : 3.88156E-43

Table 2.2: skewness and kurtosis : year

type	skewness	kurtosis
original	0.0882	1.6983
log transformation	0.0033	1.6930
sqrt transformation	0.0458	1.6937

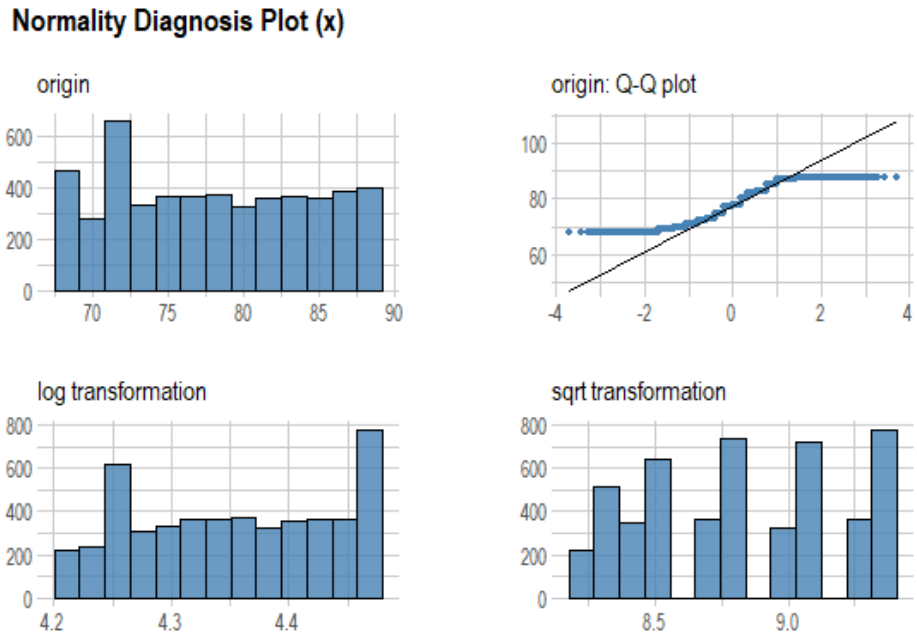

Figure 2.2: year

birth_yr

* normality test : Shapiro-Wilk normality test
- statistic : 0.95878, p-value : 1.824E-35

Table 2.3: skewness and kurtosis : birth_yr

type	skewness	kurtosis
original	-0.0990	1.9730
log transformation	-0.1924	2.0198
sqrt transformation	-0.1455	1.9937

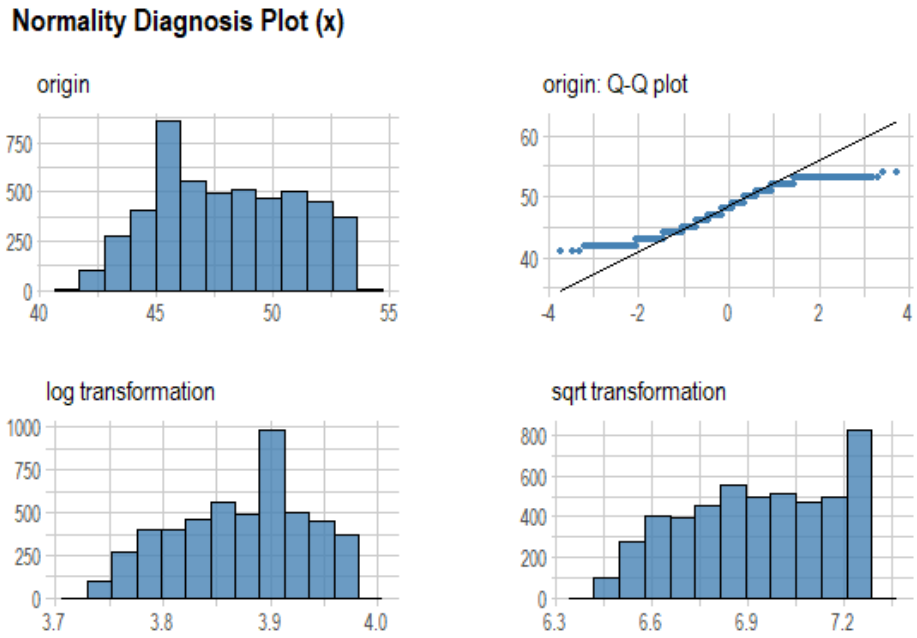

Figure 2.3: birth_yr

age

* normality test : Shapiro-Wilk normality test
- statistic : 0.96654, p-value : 1.43172E-32

Table 2.4: skewness and kurtosis : age

type	skewness	kurtosis
original	0.2957	2.1086
log transformation	-0.0536	2.0189
sqrt transformation	0.1221	2.0228

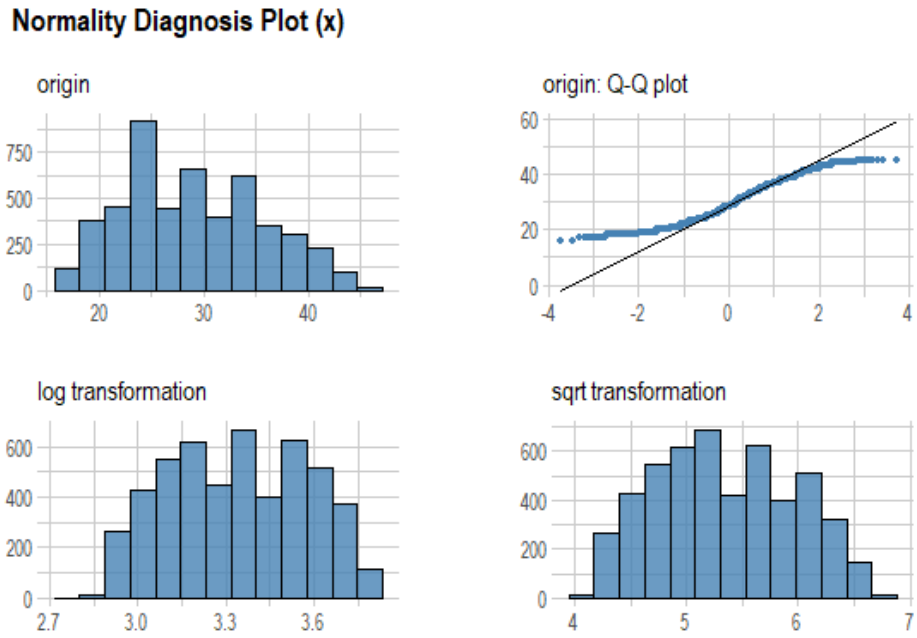

Figure 2.4: age

msp

* normality test : Shapiro-Wilk normality test
- statistic : 0.62274, p-value : 3.54355E-74

Table 2.5: skewness and kurtosis : msp

type	skewness	kurtosis
original	-0.3964	1.1571
log+1 transformation	-0.3964	1.1571
sqrt transformation	-0.3964	1.1571

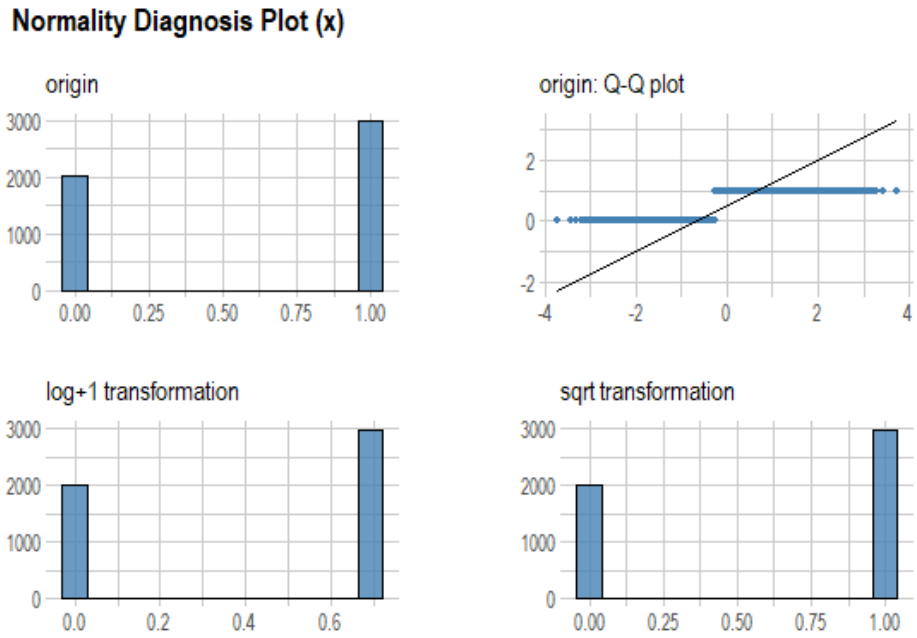

Figure 2.5: msp

nev_mar

* normality test : Shapiro-Wilk normality test
- statistic : 0.51109, p-value : 1.15226E-79

Table 2.6: skewness and kurtosis : nev_mar

type	skewness	kurtosis
original	1.3504	2.8237
log+1 transformation	1.3504	2.8237
sqrt transformation	1.3504	2.8237

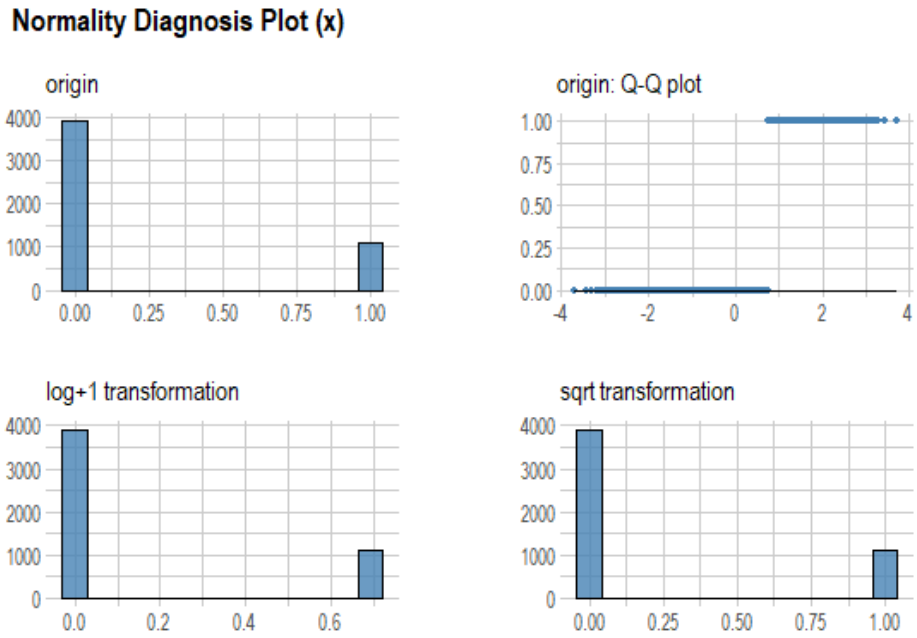

Figure 2.6: nev_mar

grade

* normality test : Shapiro-Wilk normality test
- statistic : 0.88955, p-value : 6.73868E-51

Table 2.7: skewness and kurtosis : grade

type	skewness	kurtosis
original	0.0530	4.4800
log+1 transformation	-2.9854	36.5645
sqrt transformation	-1.2842	14.7455

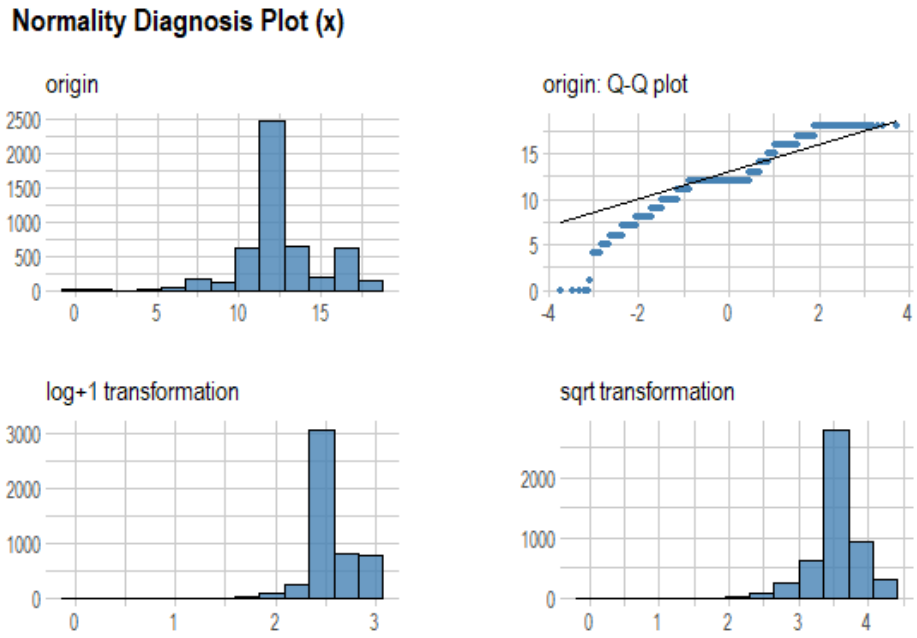

Figure 2.7: grade

collgrad

* normality test : Shapiro-Wilk normality test
- statistic : 0.45485, p-value : 4.97625E-82

Table 2.8: skewness and kurtosis : collgrad

type	skewness	kurtosis
original	1.7495	4.0608
log+1 transformation	1.7495	4.0608
sqrt transformation	1.7495	4.0608

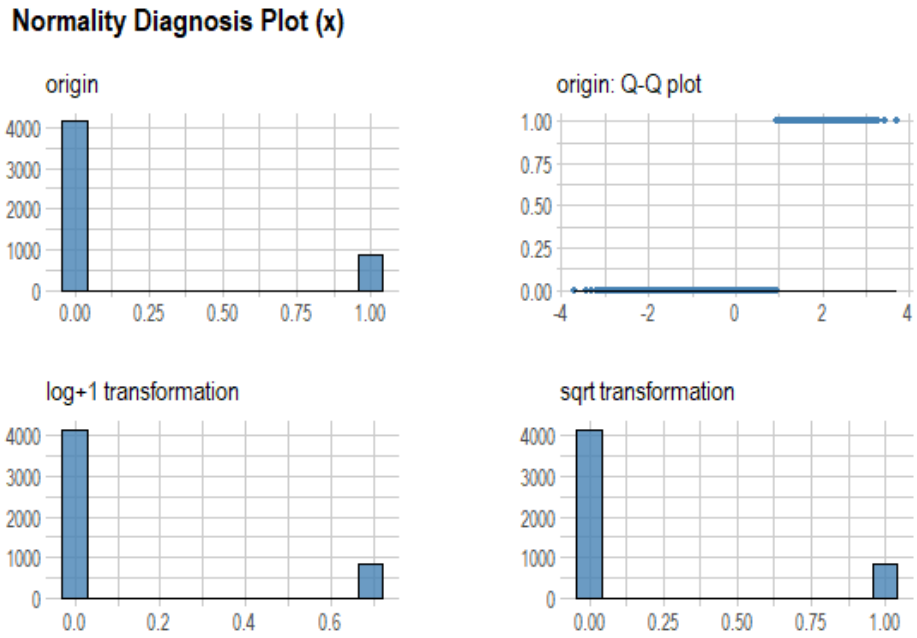

Figure 2.8: collgrad

not_smsa

* normality test : Shapiro-Wilk normality test
- statistic : 0.56308, p-value : 2.92344E-77

Table 2.9: skewness and kurtosis : not_smsa

type	skewness	kurtosis
original	0.9722	1.9453
log+1 transformation	0.9722	1.9453
sqrt transformation	0.9722	1.9453

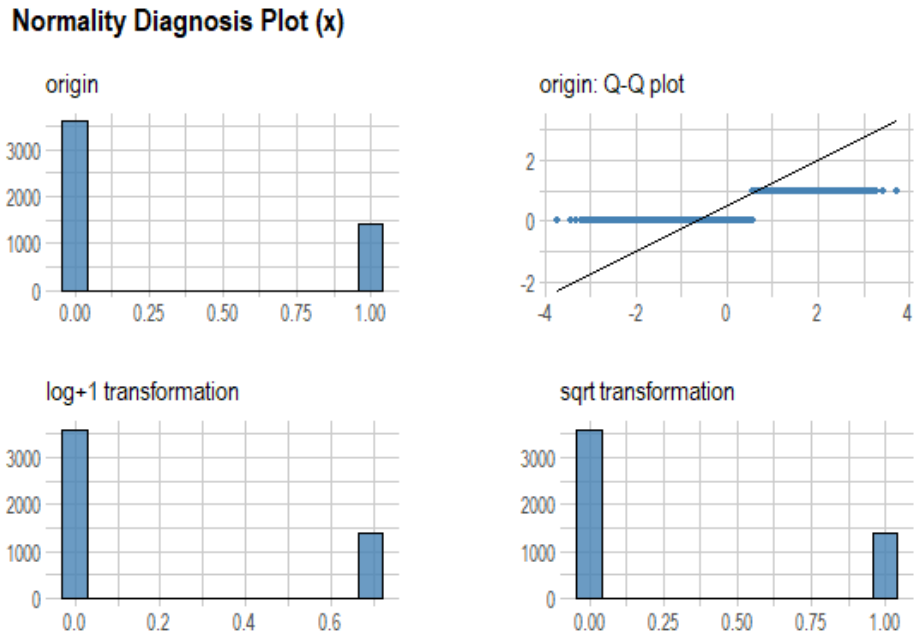

Figure 2.9: not_smsa

c_city

* normality test : Shapiro-Wilk normality test
- statistic : 0.60799, p-value : 5.62563E-75

Table 2.10: skewness and kurtosis : c_city

type	skewness	kurtosis
original	0.5779	1.334
log+1 transformation	0.5779	1.334
sqrt transformation	0.5779	1.334

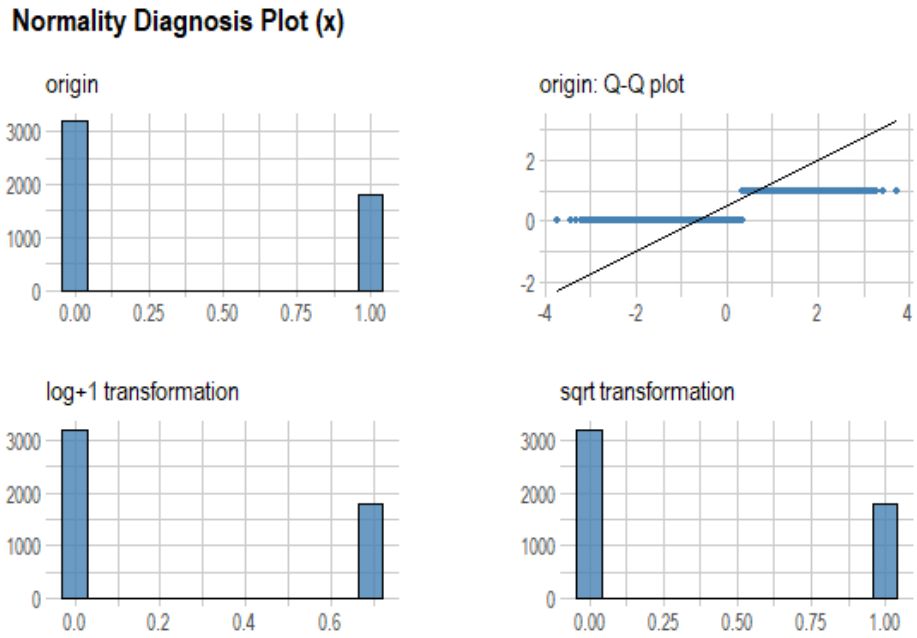

Figure 2.10: c_city

south

* normality test : Shapiro-Wilk normality test
- statistic : 0.62522, p-value : 4.85386E-74

Table 2.11: skewness and kurtosis : south

type	skewness	kurtosis
original	0.3584	1.1285
log+1 transformation	0.3584	1.1285
sqrt transformation	0.3584	1.1285

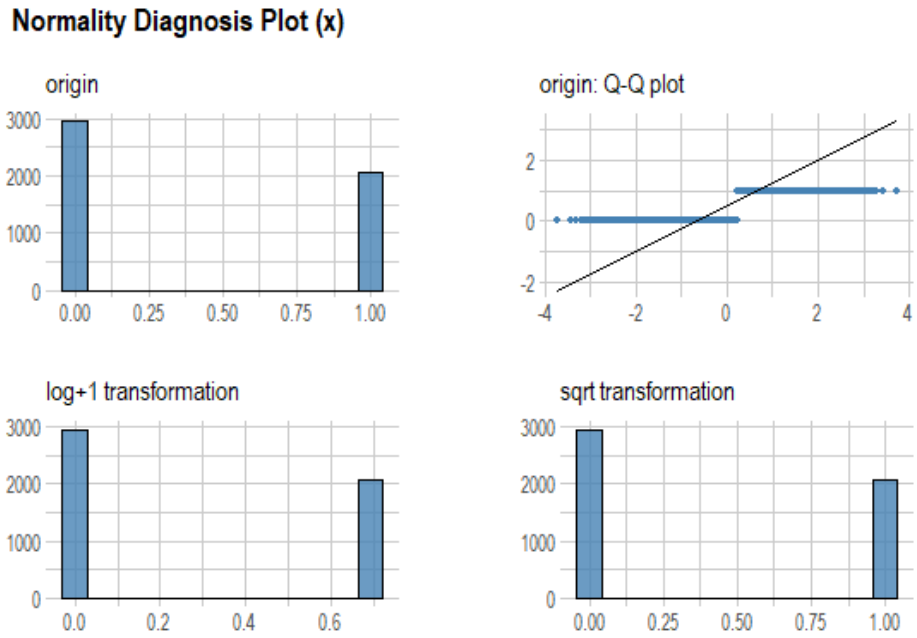

Figure 2.11: south

ind_code

* normality test : Shapiro-Wilk normality test
- statistic : 0.87227, p-value : 2.03798E-53

Table 2.12: skewness and kurtosis : ind_code

type	skewness	kurtosis
original	-0.0292	1.5558
log transformation	-0.9022	4.4623
sqrt transformation	-0.3044	2.0958

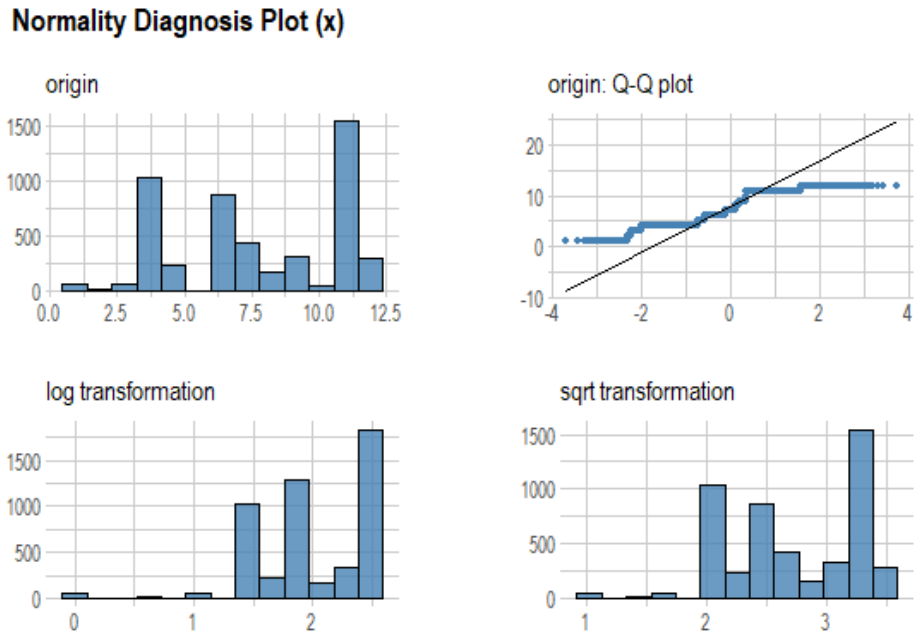

Figure 2.12: ind_code

occ_code

* normality test : Shapiro-Wilk normality test
- statistic : 0.84955, p-value : 2.48497E-56

Table 2.13: skewness and kurtosis : occ_code

type	skewness	kurtosis
original	1.0872	3.6729
log transformation	-0.2926	2.6586
sqrt transformation	0.4529	2.6251

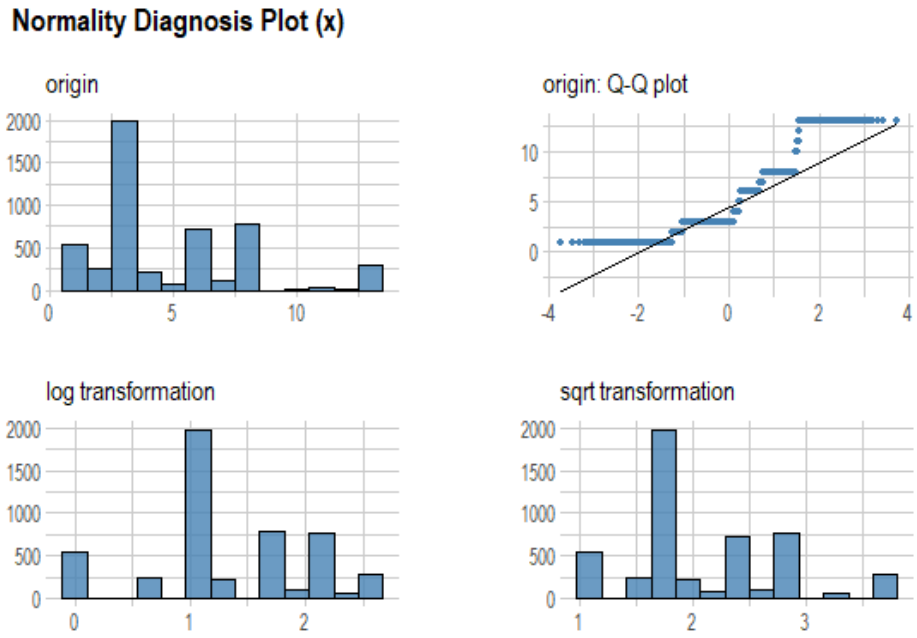

Figure 2.13: occ_code

union

* normality test : Shapiro-Wilk normality test
- statistic : 0.5291, p-value : 7.39059E-79

Table 2.14: skewness and kurtosis : union

type	skewness	kurtosis
original	1.2227	2.495
log+1 transformation	1.2227	2.495
sqrt transformation	1.2227	2.495

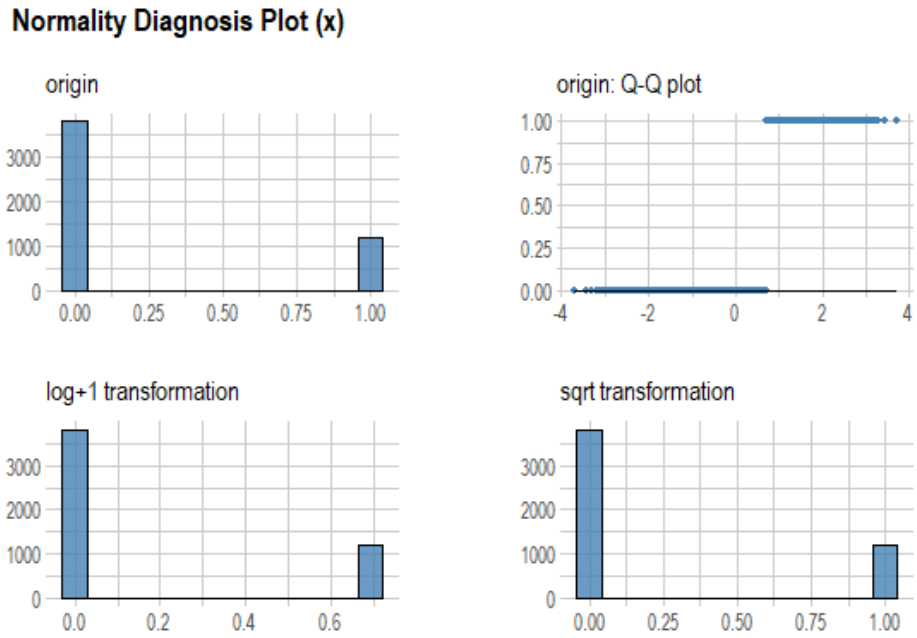

Figure 2.14: union

wks_ue

* normality test : Shapiro-Wilk normality test
- statistic : 0.41202, p-value : 1.0809E-83

Table 2.15: skewness and kurtosis : wks_ue

type	skewness	kurtosis
original	3.9513	20.4943
log+1 transformation	1.8849	5.3525
sqrt transformation	2.2685	7.6753

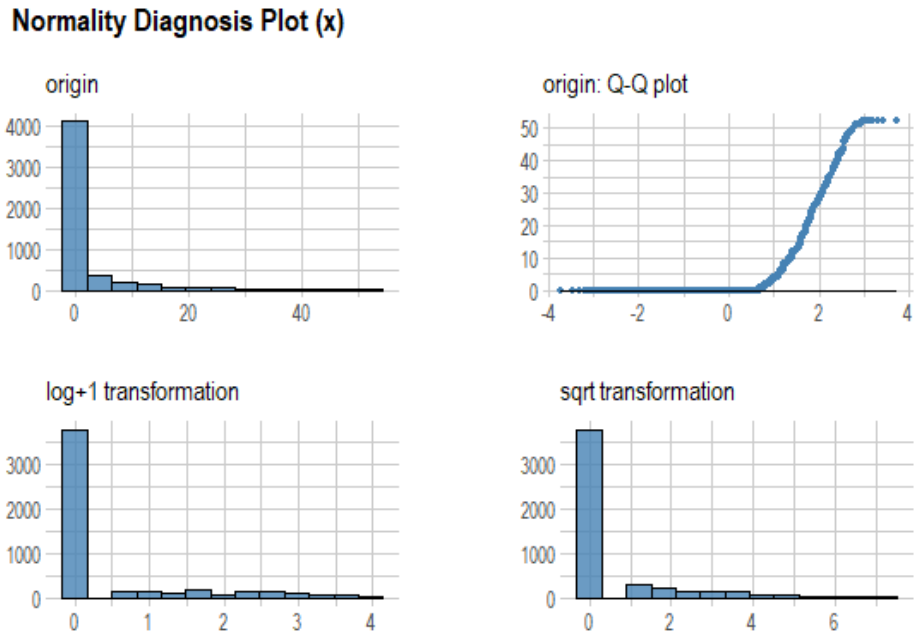

Figure 2.15: wks_ue

ttl_exp

* normality test : Shapiro-Wilk normality test
- statistic : 0.9248, p-value : 1.55071E-44

Table 2.16: skewness and kurtosis : ttl_exp

type	skewness	kurtosis
original	0.8629	3.0808
log+1 transformation	-0.2884	2.2615
sqrt transformation	0.1377	2.2589

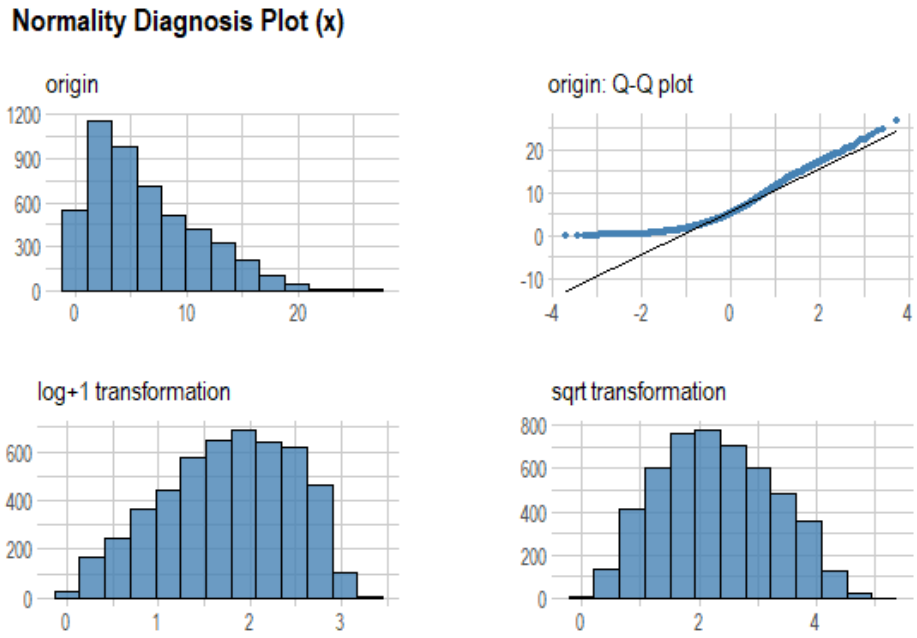

Figure 2.16: ttl_exp

tenure

* normality test : Shapiro-Wilk normality test
- statistic : 0.7701, p-value : 2.9984E-64

Table 2.17: skewness and kurtosis : tenure

type	skewness	kurtosis
original	1.9180	6.7779
log+1 transformation	0.4796	2.3005
sqrt transformation	0.7565	3.0801

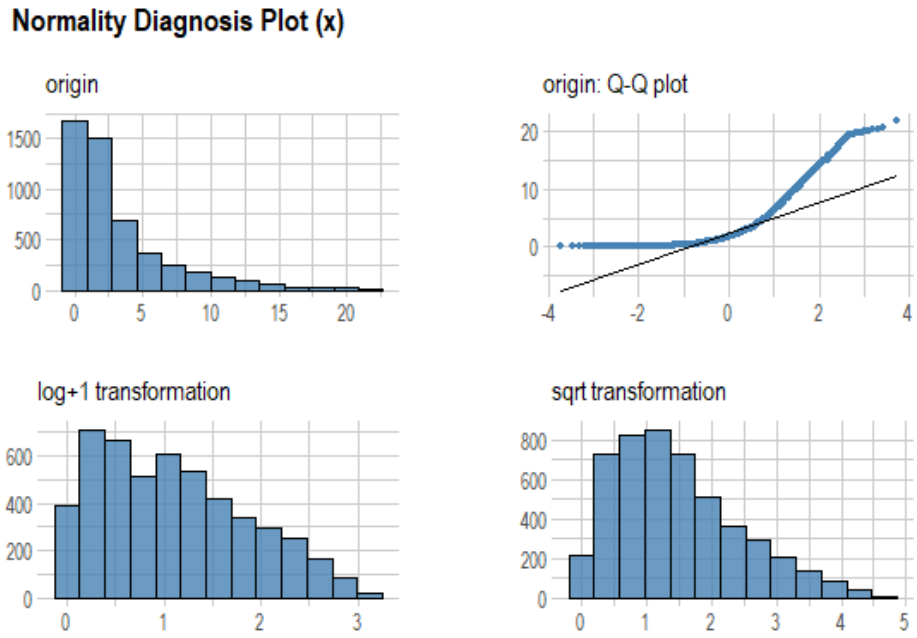

Figure 2.17: tenure

hours

* normality test : Shapiro-Wilk normality test
- statistic : 0.79234, p-value : 2.6665E-62

Table 2.18: skewness and kurtosis : hours

type	skewness	kurtosis
original	-0.8231	6.5721
log transformation	-2.9944	14.9207
sqrt transformation	-1.8007	7.5291

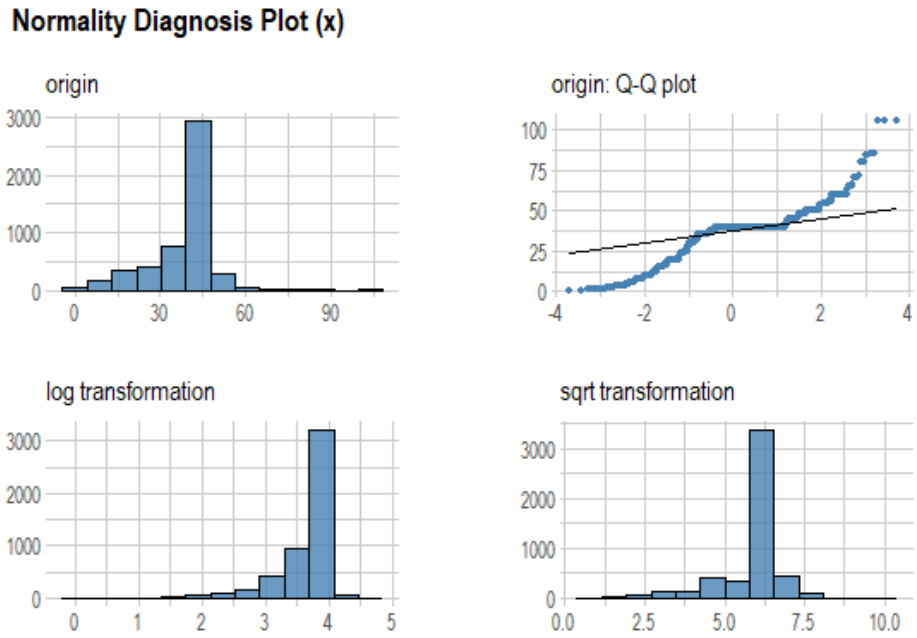

Figure 2.18: hours

wks_work

* normality test : Shapiro-Wilk normality test
- statistic : 0.93838, p-value : 2.06203E-41

Table 2.19: skewness and kurtosis : wks_work

type	skewness	kurtosis
original	0.2017	2.3113
log+1 transformation	-2.0552	8.2239
sqrt transformation	-0.7276	3.4835

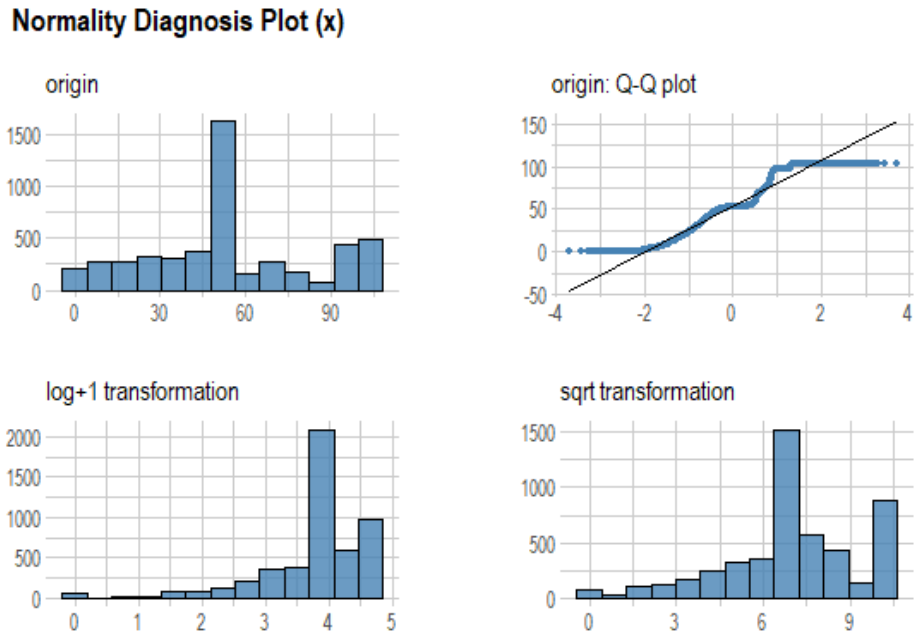

Figure 2.19: wks_work

ln_wage

* normality test : Shapiro-Wilk normality test
- statistic : 0.98366, p-value : 1.31903E-23

Table 2.20: skewness and kurtosis : ln_wage

type	skewness	kurtosis
original	0.2905	4.6636
log transformation	-3.0011	24.9327
sqrt transformation	-0.6959	6.2322

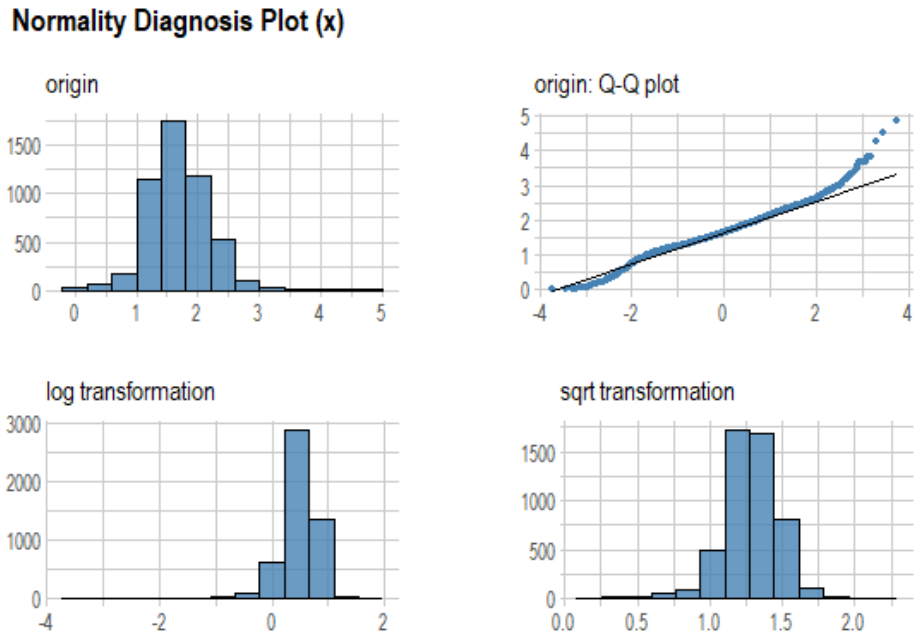

Figure 2.20: ln_wage

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
age	year	0.895
ttl_exp	year	0.777
collgrad	grade	0.757
ttl_exp	age	0.756
tenure	ttl_exp	0.674
nev_mar	msp	-0.673
wks_work	ttl_exp	0.630
wks_work	year	0.565
wks_work	age	0.525

3.1.2 Correlation Plot of Numerical Variables

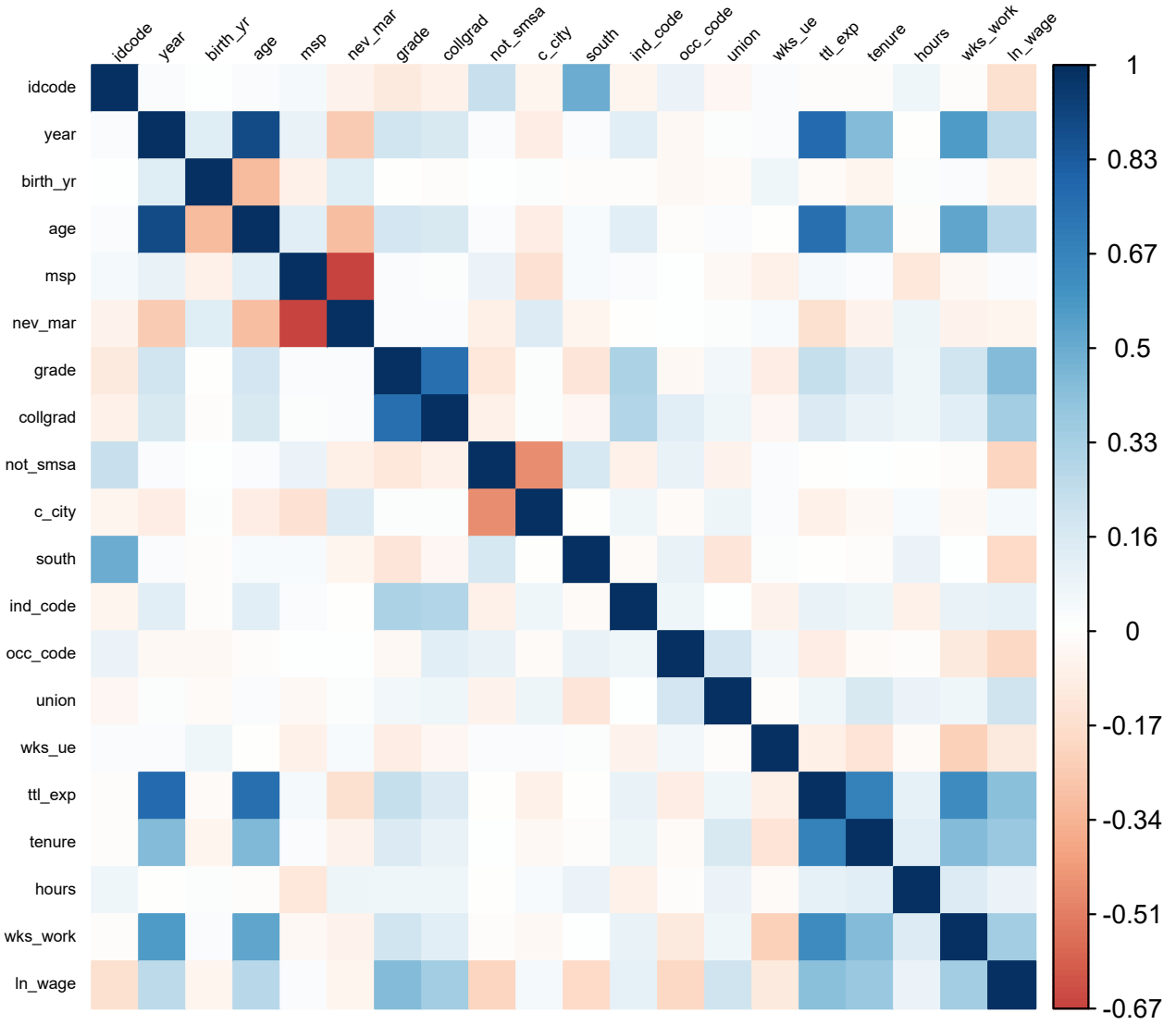

Figure 3.1: The correlation coefficient of numerical variables

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.