REPORT SERIES WITH DLOOKR

Exploratory Data Analysis Report

Author:
dlookr package

Version:
0.4.0

March 2, 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 400 observations and 11 variables.

1.2 Information of Variables

Table 1.1: Information of Variables

variables	types	missing_count	missing_percent	unique_count	unique_rate
Sales	numeric	0	0.00	336	0.840
CompPrice	numeric	0	0.00	73	0.182
Income	numeric	20	5.00	99	0.248
Advertising	numeric	0	0.00	28	0.070
Population	numeric	0	0.00	275	0.688
Price	numeric	0	0.00	101	0.252
ShelveLoc	factor	0	0.00	3	0.007
Age	numeric	0	0.00	56	0.140
Education	numeric	0	0.00	9	0.022
Urban	factor	5	1.25	3	0.007
US	factor	0	0.00	2	0.005

The target variable of the data is 'US', and the data type of the variable is factor.

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
11 Variables 400 Observations

Sales

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
400	0	336	1	7.496	3.192	3.149	4.119	5.390	7.490	9.320	11.300	12.442

```
lowest :  0.00  0.16  0.37  0.53  0.91, highest: 13.91 14.37 14.90 15.63 16.27
```

CompPrice

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
400	0	73	0.999	125	17.3	98	106	115	125	135	145	150

```
lowest :  77  85  86  88  89, highest: 157 159 161 162 175
```

Income

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
380	20	98	1	68.12	32.47	25.95	30.00	42.00	68.50	90.00	107.00	115.00

```
lowest :  21  22  23  24  25, highest: 116 117 118 119 120
```

Advertising

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
400	0	28	0.952	6.635	7.337	0	0	0	5	12	16	19

```
lowest :  0  1  2  3  4, highest: 23 24 25 26 29
```

Population

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
400	0	275	1	264.8	170.3	29.0	58.9	139.0	272.0	398.5	467.0	493.1

```
lowest :  10  12  13  14  16, highest: 503 504 507 508 509
```

Price

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
400	0	101	1	115.8	26.52	77	87	100	117	131	146	155

```
lowest :  24  49  53  54  55, highest: 166 171 173 185 191
```

ShelveLoc

n	missing	distinct
400	0	3

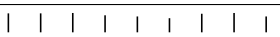
Value	Bad	Good	Medium
Frequency	96	85	219
Proportion	0.240	0.212	0.547

Age

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
400	0	56	1	53.32	18.71	27.00	30.00	39.75	54.50	66.00	76.00	79.00

```
lowest : 25 26 27 28 29, highest: 76 77 78 79 80
```

Education



n	missing	distinct	Info	Mean	Gmd
400	0	9	0.987	13.9	3.009

```
lowest : 10 11 12 13 14, highest: 14 15 16 17 18
```

Value	10	11	12	13	14	15	16	17	18
Frequency	48	48	49	43	40	36	47	49	40
Proportion	0.120	0.120	0.122	0.108	0.100	0.090	0.117	0.122	0.100

Urban

n	missing	distinct
395	5	2

Value	No	Yes
Frequency	116	279
Proportion	0.294	0.706

US

n	missing	distinct
400	0	2

Value	No	Yes
Frequency	142	258
Proportion	0.355	0.645

2.2 Normality Test of Numerical Variables

2.2.1 Statistics and Visualization of (Sample) Data

Sales

* normality test : Shapiro-Wilk normality test
- statistic : 0.9952, p-value : 0.253975

Table 2.1: skewness and kurtosis : Sales

type	skewness	kurtosis
original	0.1849	2.9052
log+1 transformation	-1.4133	7.4162
sqrt transformation	-0.7389	4.9166

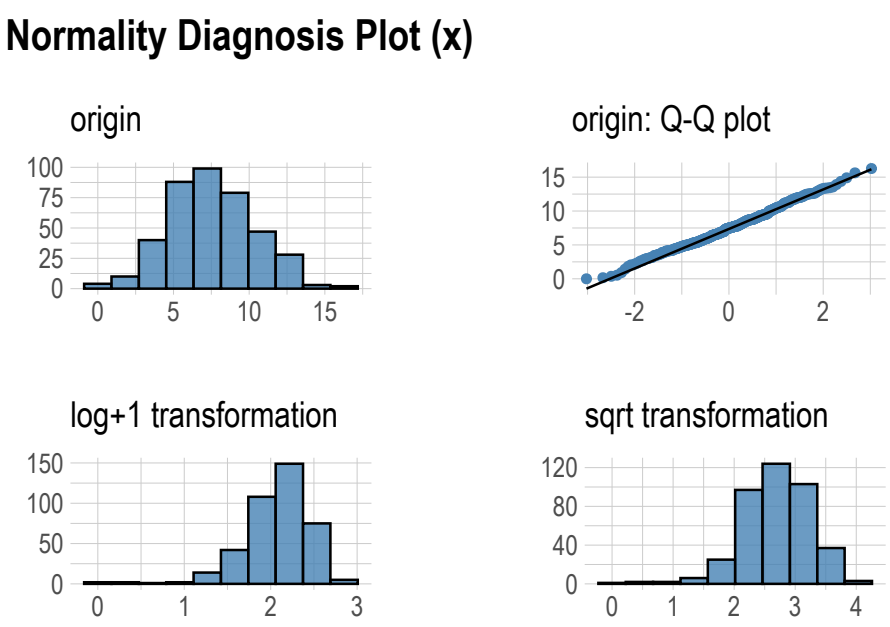


Figure 2.1: Sales

CompPrice

* normality test : Shapiro-Wilk normality test
- statistic : 0.99843, p-value : 0.977151

Table 2.2: skewness and kurtosis : CompPrice

type	skewness	kurtosis
original	-0.0426	3.0262
log transformation	-0.4347	3.3671
sqrt transformation	-0.2347	3.1280

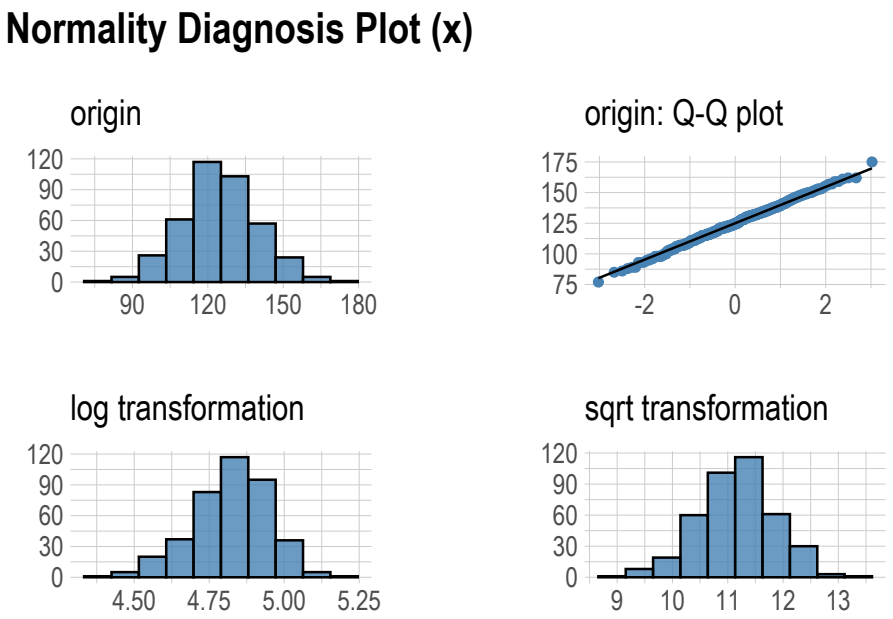


Figure 2.2: CompPrice

Income

* normality test : Shapiro-Wilk normality test
- statistic : 0.95995, p-value : 1.14495E-08

Table 2.3: skewness and kurtosis : Income

type	skewness	kurtosis
original	0.0797	1.9065
log transformation	-0.5412	2.2170
sqrt transformation	-0.2222	1.9480

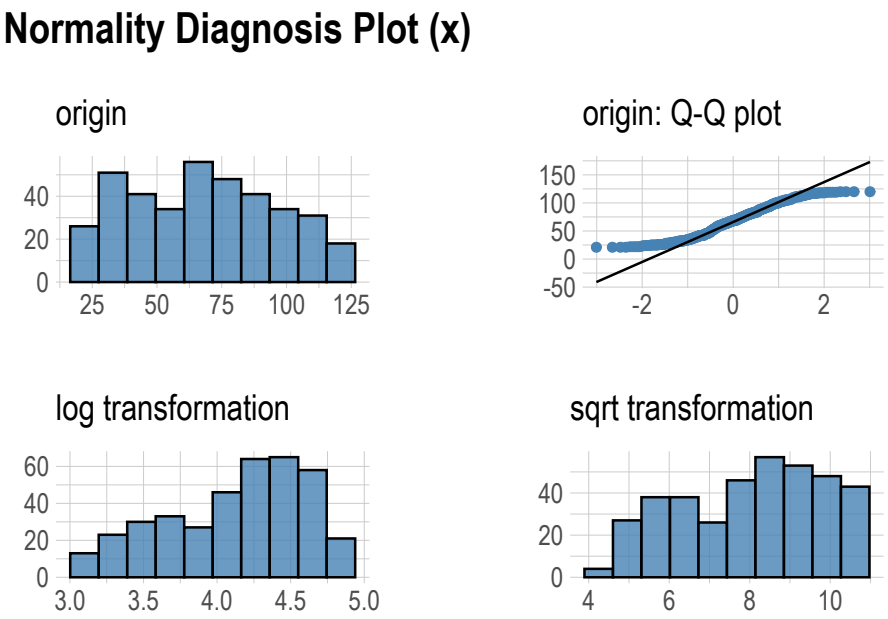


Figure 2.3: Income

Advertising

* normality test : Shapiro-Wilk normality test
- statistic : 0.87354, p-value : 1.49183E-17

Table 2.4: skewness and kurtosis : Advertising

type	skewness	kurtosis
original	0.6372	2.4467
log+1 transformation	-0.1978	1.3423
sqrt transformation	-0.0565	1.4653

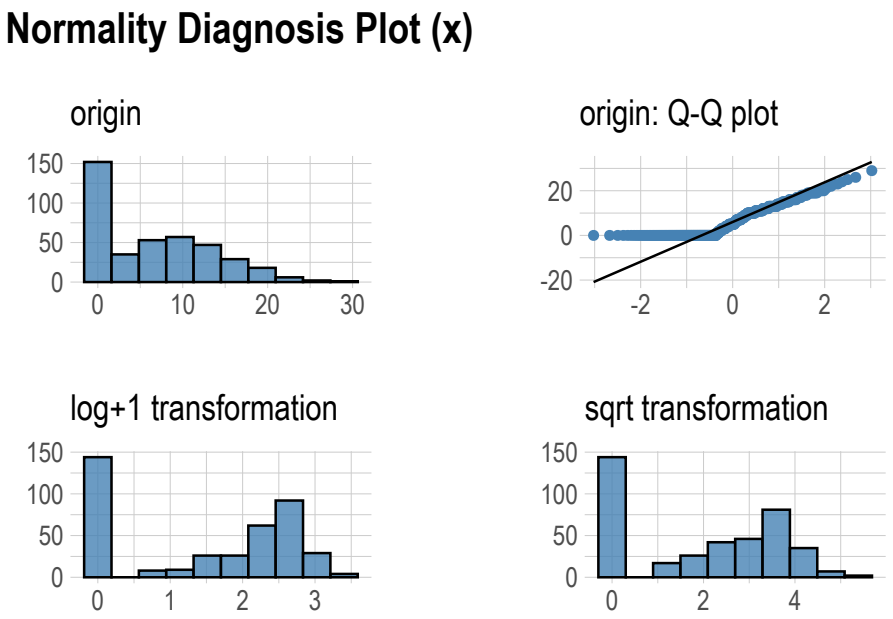


Figure 2.4: Advertising

Population

* normality test : Shapiro-Wilk normality test
- statistic : 0.95201, p-value : 4.08085E-10

Table 2.5: skewness and kurtosis : Population

type	skewness	kurtosis
original	-0.0510	1.7977
log transformation	-1.2945	4.1336
sqrt transformation	-0.5427	2.2584

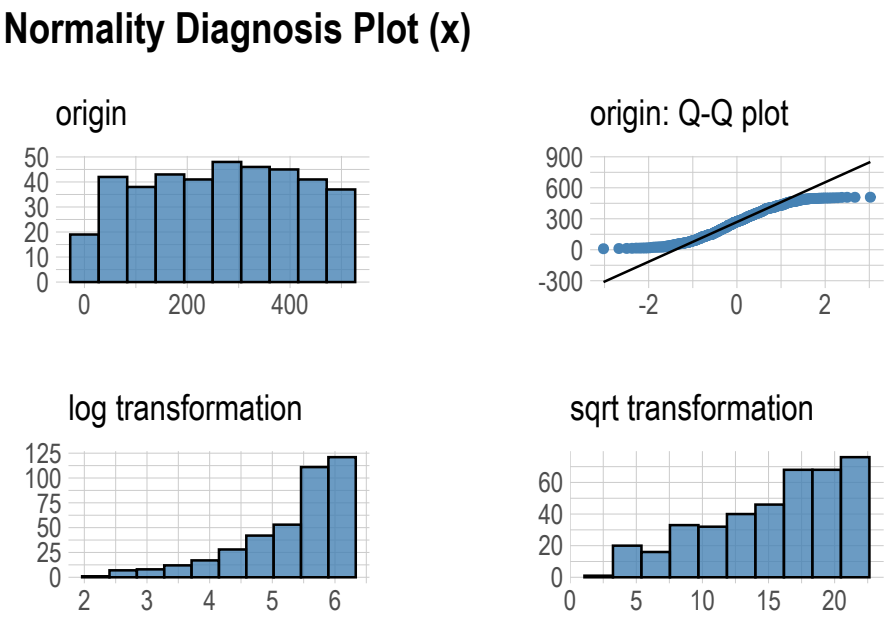


Figure 2.5: Population

Price

* normality test : Shapiro-Wilk normality test
- statistic : 0.99592, p-value : 0.390213

Table 2.6: skewness and kurtosis : Price

type	skewness	kurtosis
original	-0.1248	3.4313
log transformation	-1.3589	8.6448
sqrt transformation	-0.6083	4.5887

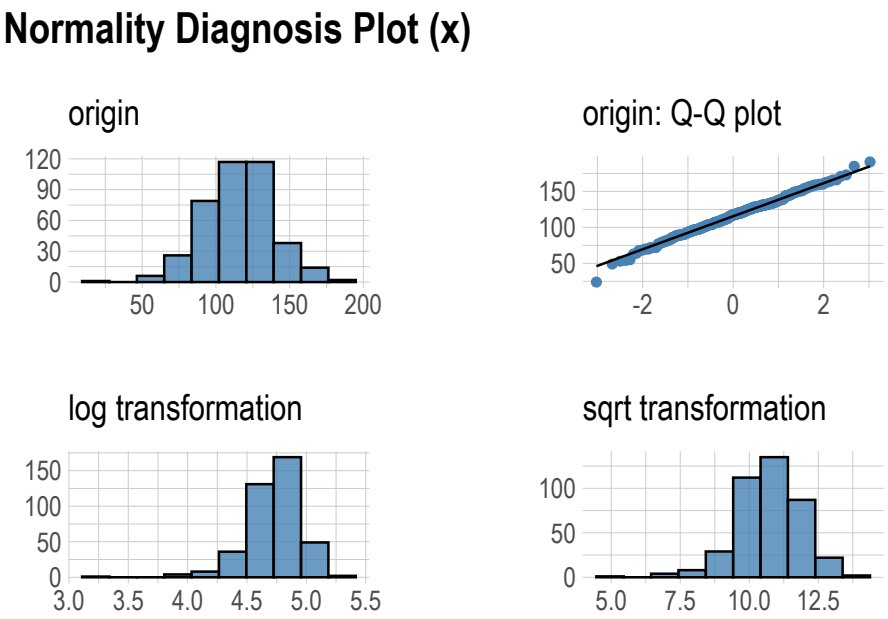


Figure 2.6: Price

Age

* normality test : Shapiro-Wilk normality test
- statistic : 0.95672, p-value : 1.86455E-09

Table 2.7: skewness and kurtosis : Age

type	skewness	kurtosis
original	-0.0769	1.8648
log transformation	-0.5112	2.1718
sqrt transformation	-0.2890	1.9631

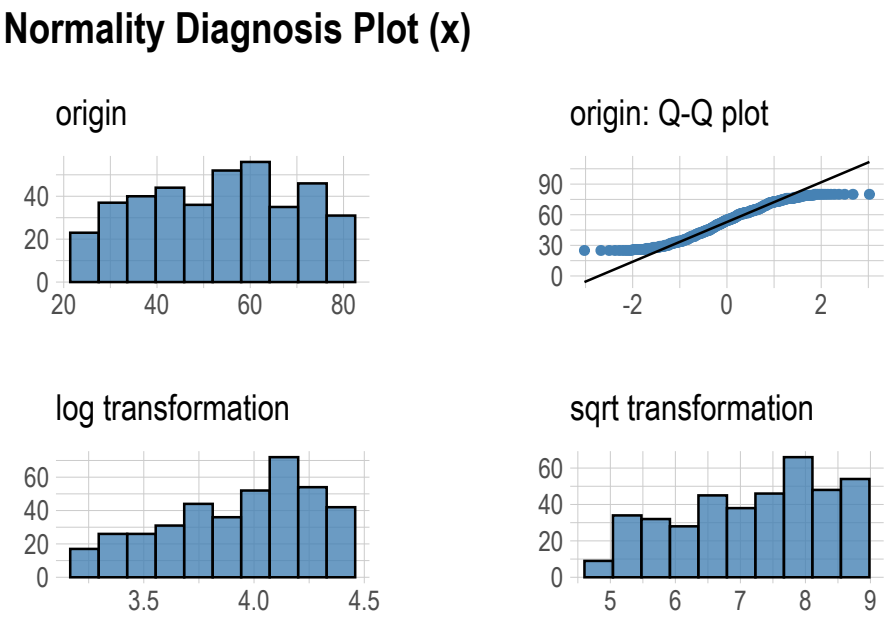


Figure 2.7: Age

Education

* normality test : Shapiro-Wilk normality test
- statistic : 0.9242, p-value : 2.42693E-13

Table 2.8: skewness and kurtosis : Education

type	skewness	kurtosis
original	0.0438	1.7029
log transformation	-0.1599	1.7434
sqrt transformation	-0.0572	1.7118

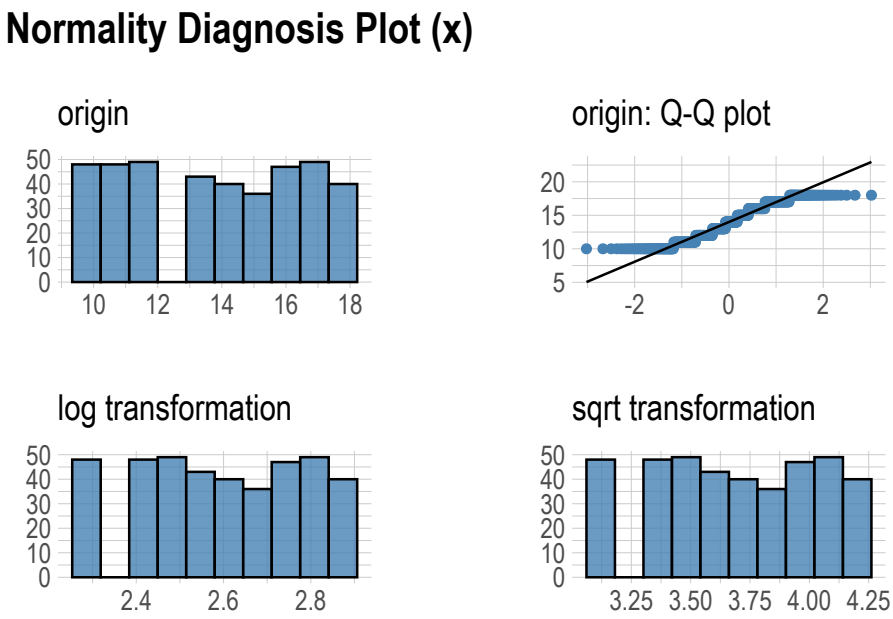


Figure 2.8: Education

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
Price	CompPrice	0.585

3.1.2 Correlation Plot of Numerical Variables

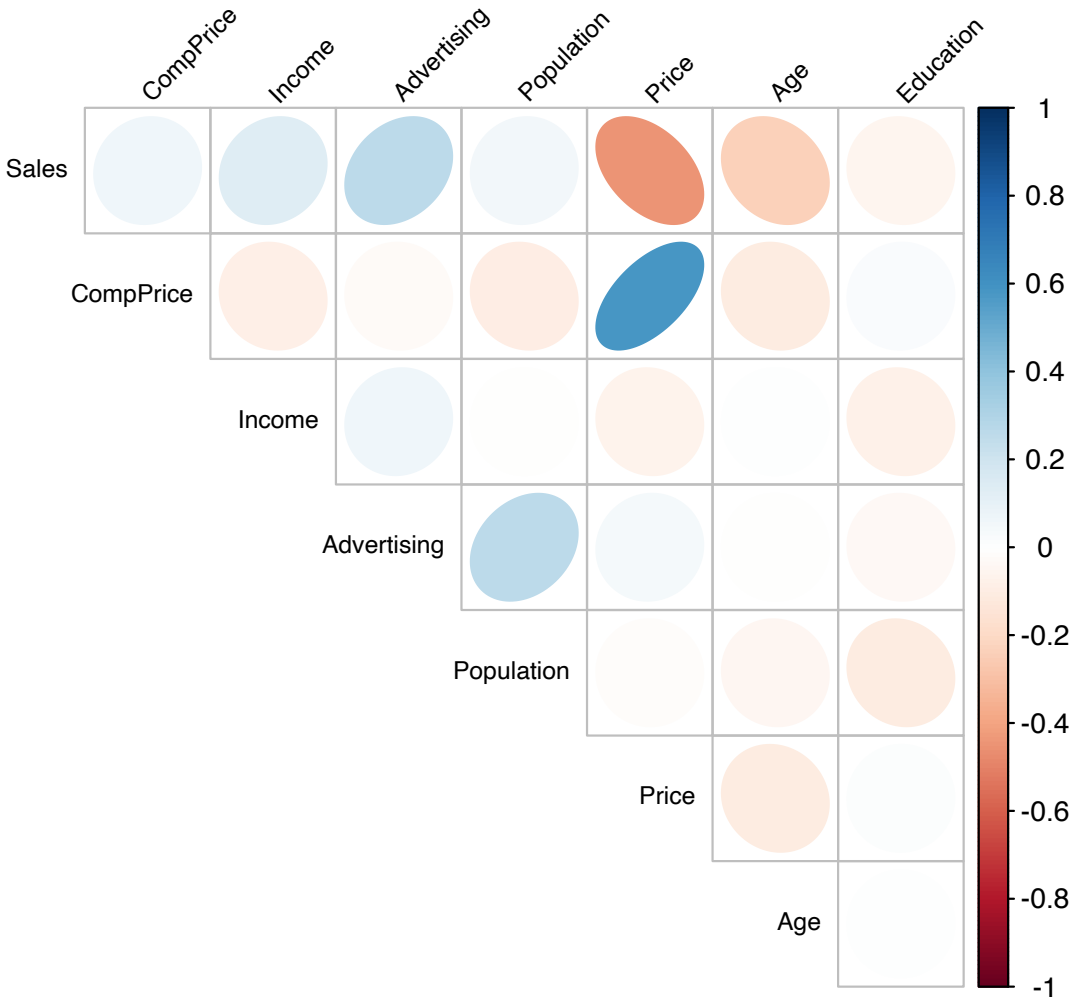


Figure 3.1: The correlation coefficient of numerical variables

Chapter 4

Target based Analysis

4.1 Grouped Descriptive Statistics

4.1.1 Grouped Numerical Variables

Sales

Table 4.1: Sales

	Yes	No
n	258.00	142.00
NA	0.00	0.00
mean	7.87	6.82
sd	2.88	2.60
se(mean)	0.18	0.22
IQR	4.23	3.44
skewness	0.08	0.32
kurtosis	-0.33	0.81
0%	0.37	0.00
1%	1.65	0.47
5%	3.15	3.25
10%	4.18	3.92
20%	5.33	4.75
25%	5.76	5.08
30%	6.15	5.31
40%	6.92	5.99
50%	7.79	6.66
60%	8.65	7.50
70%	9.45	7.96
75%	9.99	8.52
80%	10.46	8.77
90%	11.74	9.35
95%	12.54	11.28
99%	13.64	14.03
100%	16.27	14.90

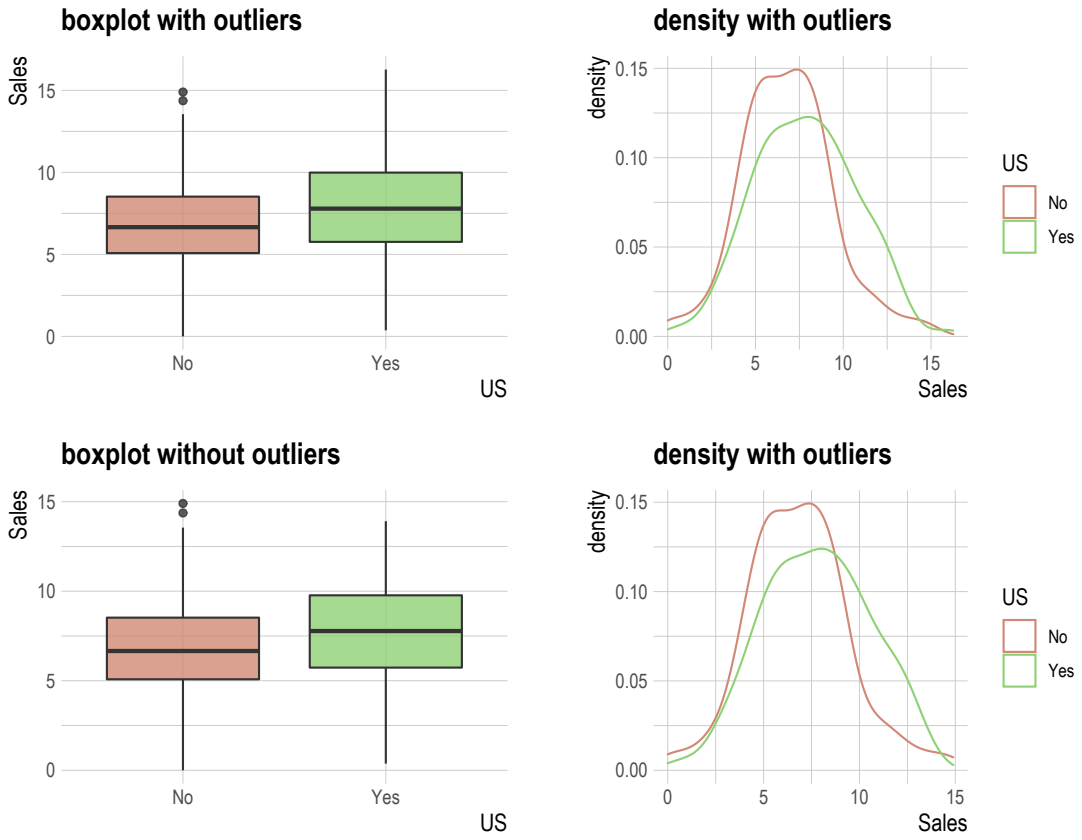


Figure 4.1: Sales

CompPrice

Table 4.2: CompPrice

	Yes	No
n	258.00	142.00
NA	0.00	0.00
mean	125.17	124.63
sd	14.97	16.02
se(mean)	0.93	1.34
IQR	19.75	19.00
skewness	0.01	-0.11
kurtosis	0.06	0.01
0%	85.00	77.00
1%	91.28	87.23
5%	100.00	98.00
10%	106.70	106.00
20%	113.00	112.20
25%	115.25	115.00
30%	117.00	116.00
40%	122.00	121.00
50%	125.00	124.00
60%	130.00	128.60
70%	133.00	132.00
75%	135.00	134.00
80%	137.00	138.00
90%	144.00	145.90
95%	149.00	152.00
99%	161.43	158.18
100%	175.00	159.00

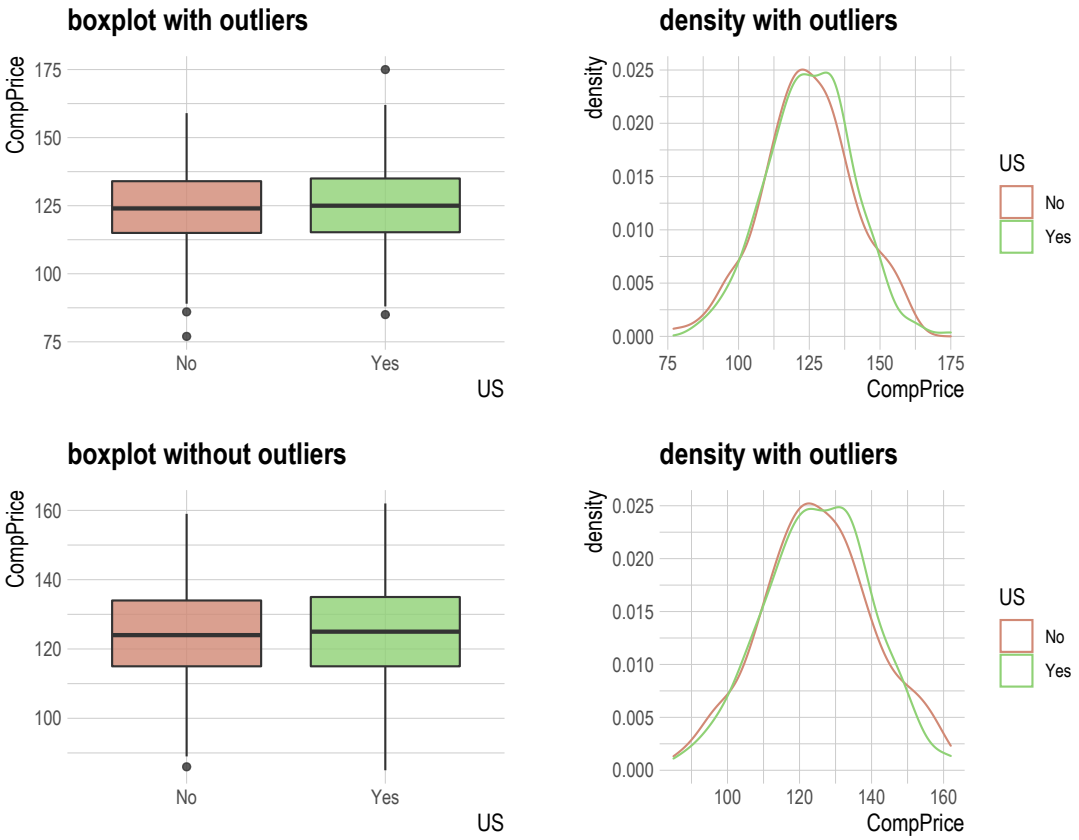


Figure 4.2: CompPrice

Income

Table 4.3: Income

	Yes	No
n	247.00	133.00
NA	11.00	9.00
mean	70.11	64.44
sd	27.95	28.22
se(mean)	1.78	2.45
IQR	47.00	46.00
skewness	0.02	0.21
kurtosis	-1.06	-1.11
0%	21.00	22.00
1%	21.00	22.00
5%	26.00	25.00
10%	32.00	30.00
20%	42.00	33.00
25%	45.00	38.00
30%	51.80	42.00
40%	63.00	55.60
50%	70.00	64.00
60%	78.60	71.20
70%	88.00	81.40
75%	92.00	84.00
80%	99.00	92.00
90%	108.80	105.80
95%	117.00	111.80
99%	119.54	117.68
100%	120.00	120.00

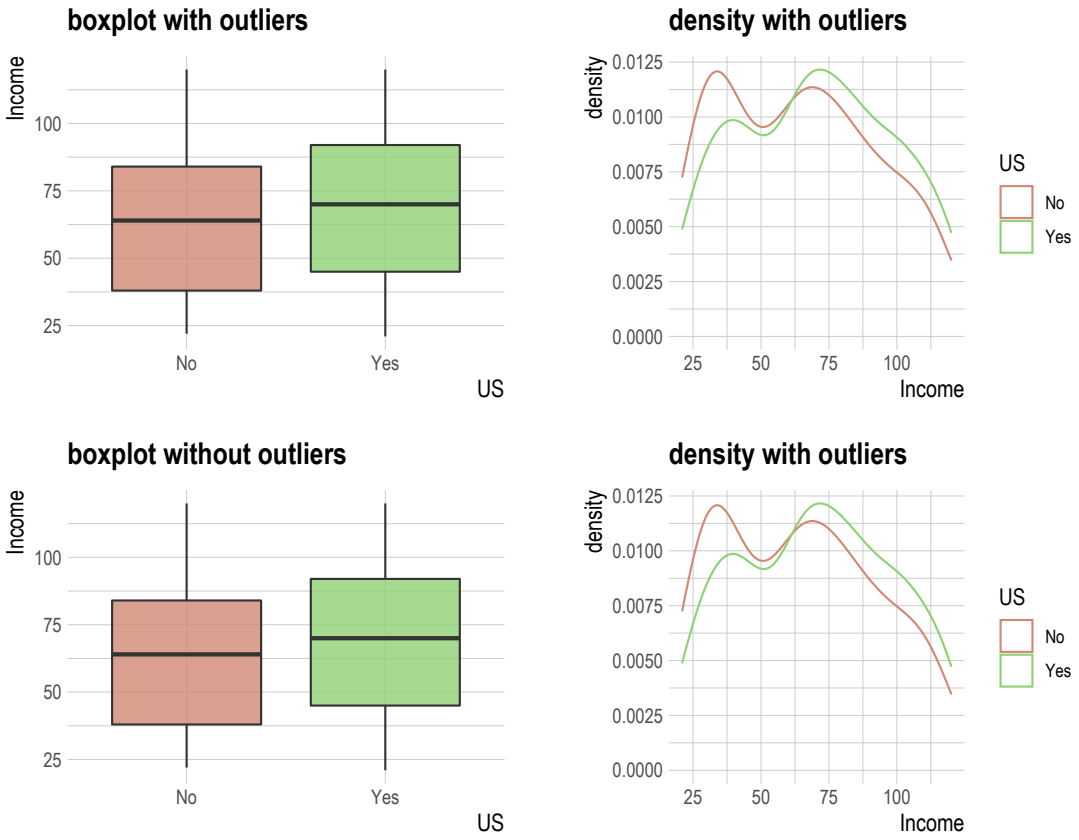


Figure 4.3: Income

Advertising

Table 4.4: Advertising

	Yes	No
n	258.00	142.00
NA	0.00	0.00
mean	10.01	0.51
sd	5.92	1.64
se(mean)	0.37	0.14
IQR	9.00	0.00
skewness	0.21	3.98
kurtosis	-0.23	17.74
0%	0.00	0.00
1%	0.00	0.00
5%	0.00	0.00
10%	2.00	0.00
20%	5.00	0.00
25%	5.00	0.00
30%	7.00	0.00
40%	9.00	0.00
50%	10.00	0.00
60%	11.20	0.00
70%	13.00	0.00
75%	14.00	0.00
80%	15.00	0.00
90%	18.00	1.90
95%	19.15	4.00
99%	24.43	7.77
100%	29.00	11.00

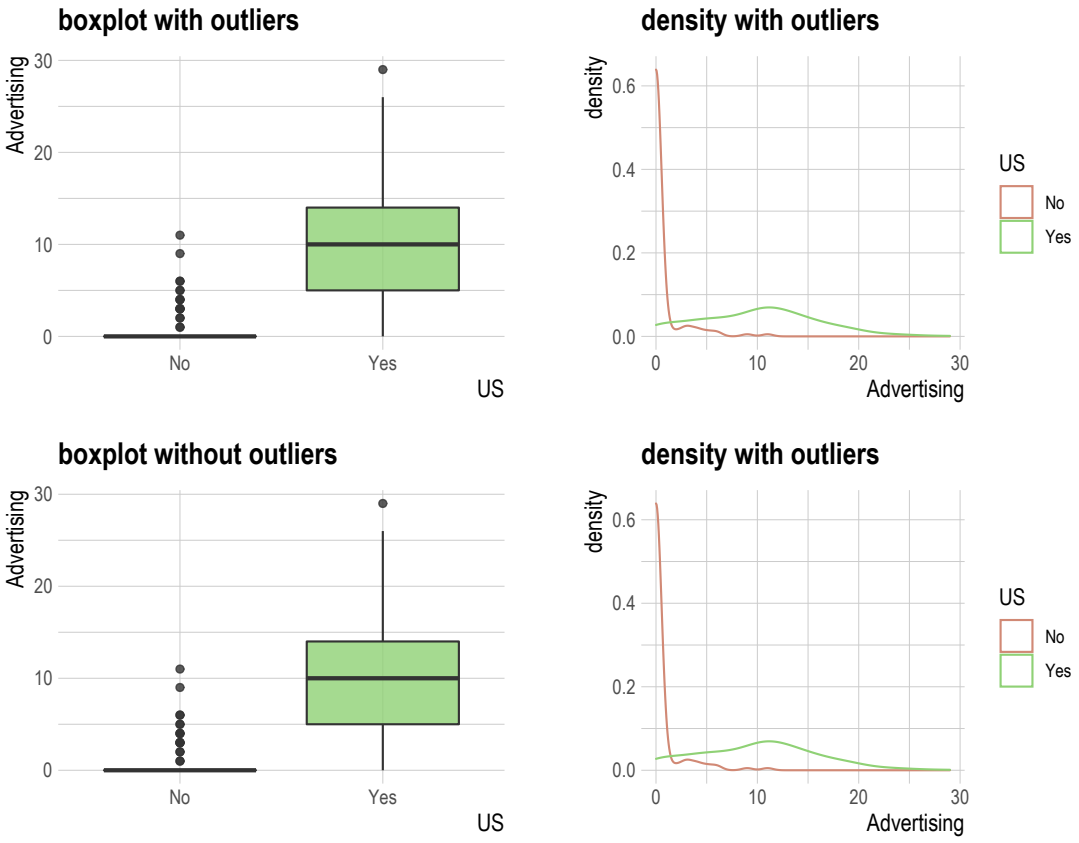


Figure 4.4: Advertising

Population

Table 4.5: Population

	Yes	No
n	258.00	142.00
NA	0.00	0.00
mean	271.45	252.82
sd	144.44	152.36
se(mean)	8.99	12.79
IQR	249.25	284.50
skewness	-0.15	0.13
kurtosis	-1.13	-1.26
0%	12.00	10.00
1%	16.57	13.41
5%	29.00	38.10
10%	60.00	57.20
20%	127.20	95.40
25%	148.25	113.75
30%	176.20	142.60
40%	237.80	193.40
50%	281.50	244.00
60%	326.00	295.60
70%	367.90	355.30
75%	397.50	398.25
80%	412.60	412.00
90%	464.60	472.00
95%	489.45	496.80
99%	501.43	507.59
100%	509.00	508.00

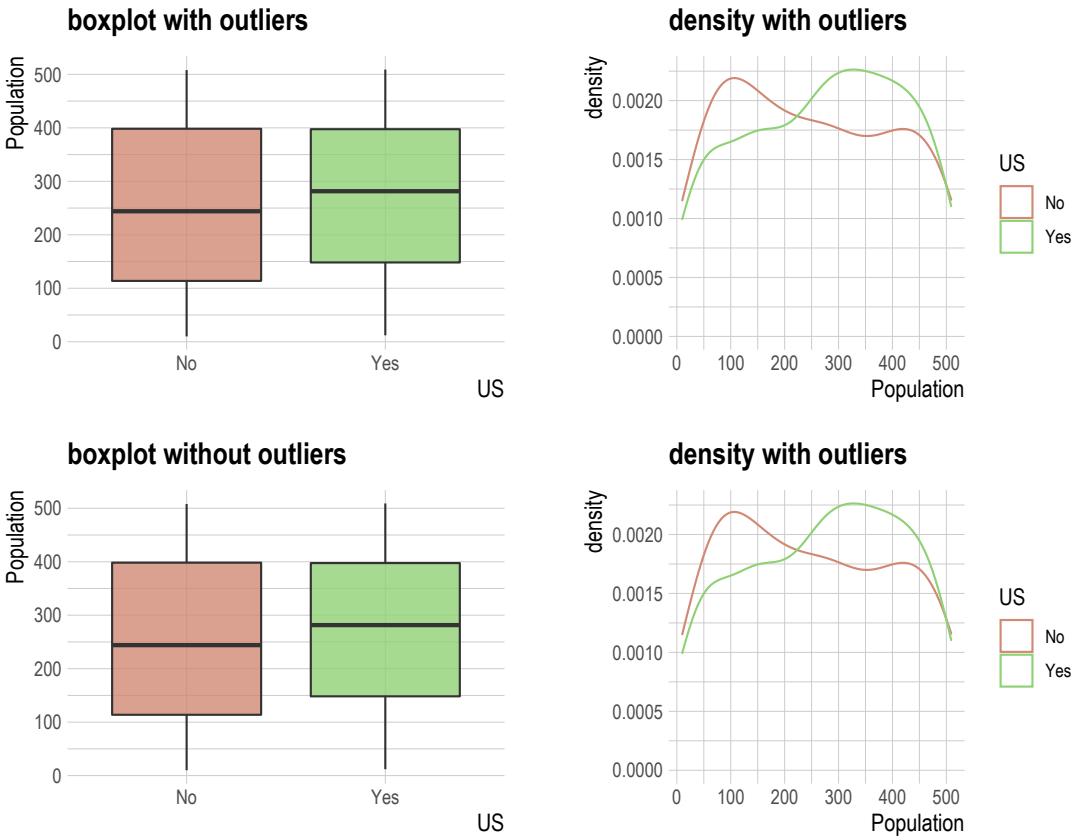


Figure 4.5: Population

Price

Table 4.6: Price

	Yes	No
n	258.00	142.00
NA	0.00	0.00
mean	116.81	113.95
sd	22.59	25.51
se(mean)	1.41	2.14
IQR	30.00	31.75
skewness	0.09	-0.35
kurtosis	-0.03	0.83
0%	55.00	24.00
1%	70.00	50.64
5%	79.00	69.05
10%	87.70	86.30
20%	97.00	94.00
25%	101.00	98.00
30%	104.00	102.00
40%	110.00	108.00
50%	118.00	116.50
60%	123.20	121.60
70%	129.00	126.00
75%	131.00	129.75
80%	133.00	134.00
90%	147.00	144.00
95%	155.15	153.85
99%	168.15	165.18
100%	191.00	185.00

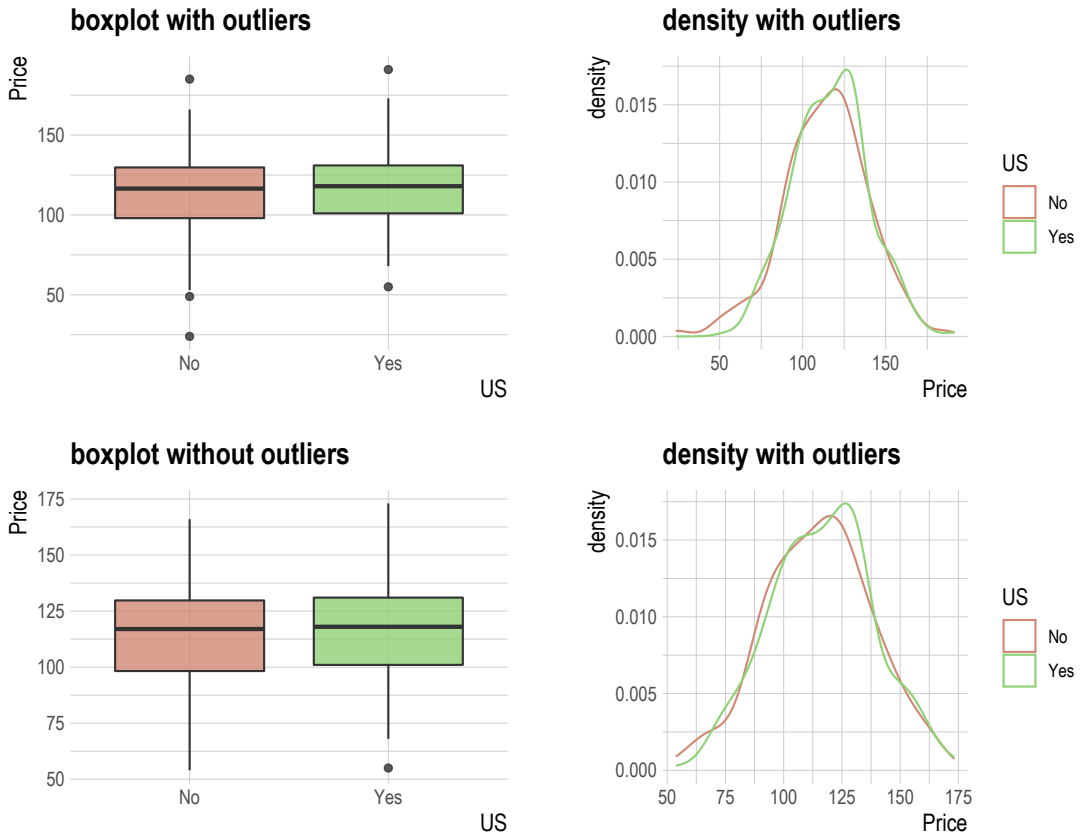


Figure 4.6: Price

Age

Table 4.7: Age

	Yes	No
n	258.00	142.00
NA	0.00	0.00
mean	53.43	53.13
sd	15.57	17.34
se(mean)	0.97	1.46
IQR	24.75	27.75
skewness	-0.08	-0.06
kurtosis	-1.07	-1.26
0%	25.00	25.00
1%	25.00	25.00
5%	28.00	26.00
10%	31.70	28.10
20%	37.00	34.00
25%	41.25	38.00
30%	44.00	41.00
40%	49.00	46.80
50%	54.50	54.50
60%	59.00	60.60
70%	63.00	64.70
75%	66.00	65.75
80%	69.00	71.80
90%	74.30	76.00
95%	77.15	79.00
99%	80.00	80.00
100%	80.00	80.00

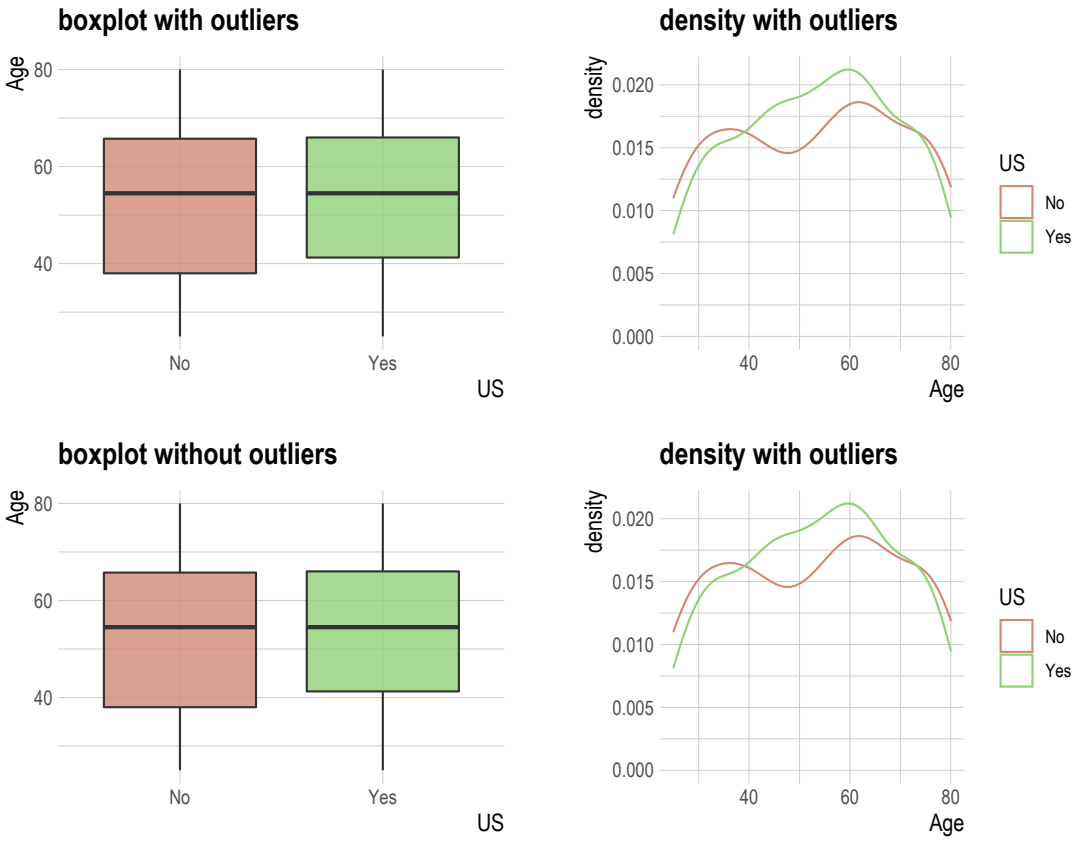


Figure 4.7: Age

Education

Table 4.8: Education

	Yes	No
n	258.00	142.00
NA	0.00	0.00
mean	13.75	14.18
sd	2.67	2.52
se(mean)	0.17	0.21
IQR	5.00	4.00
skewness	0.10	-0.04
kurtosis	-1.33	-1.23
0%	10.00	10.00
1%	10.00	10.00
5%	10.00	10.00
10%	10.00	11.00
20%	11.00	12.00
25%	11.00	12.00
30%	12.00	12.00
40%	13.00	13.00
50%	14.00	14.00
60%	15.00	15.00
70%	16.00	16.00
75%	16.00	16.00
80%	17.00	17.00
90%	17.00	18.00
95%	18.00	18.00
99%	18.00	18.00
100%	18.00	18.00

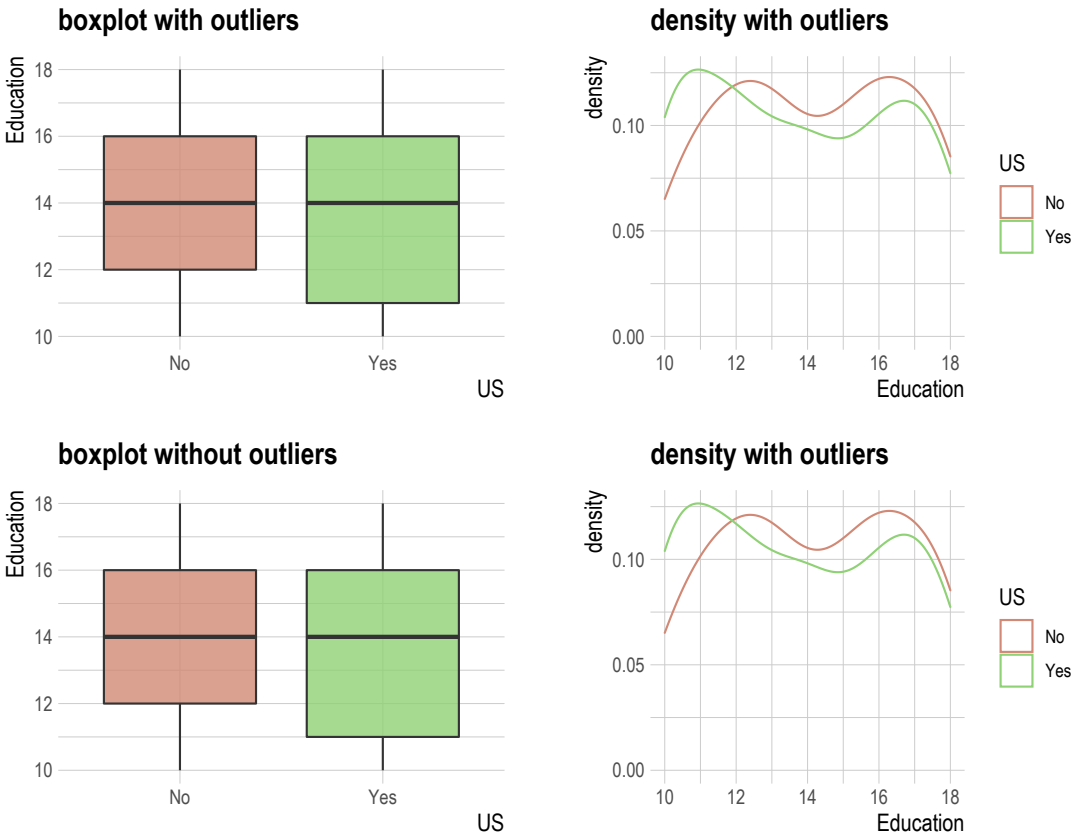


Figure 4.8: Education

4.1.2 Grouped Categorical Variables

ShelveLoc

	No	Yes	Sum
Bad	34	62	96
Good	24	61	85
Medium	84	135	219
Sum	**142**	**258**	**400**

	No	Yes	Sum
Bad	23.94	24.03	24.00
Good	16.90	23.64	21.25
Medium	59.15	52.33	54.75
Sum	**100.00**	**100.00**	**100.00**

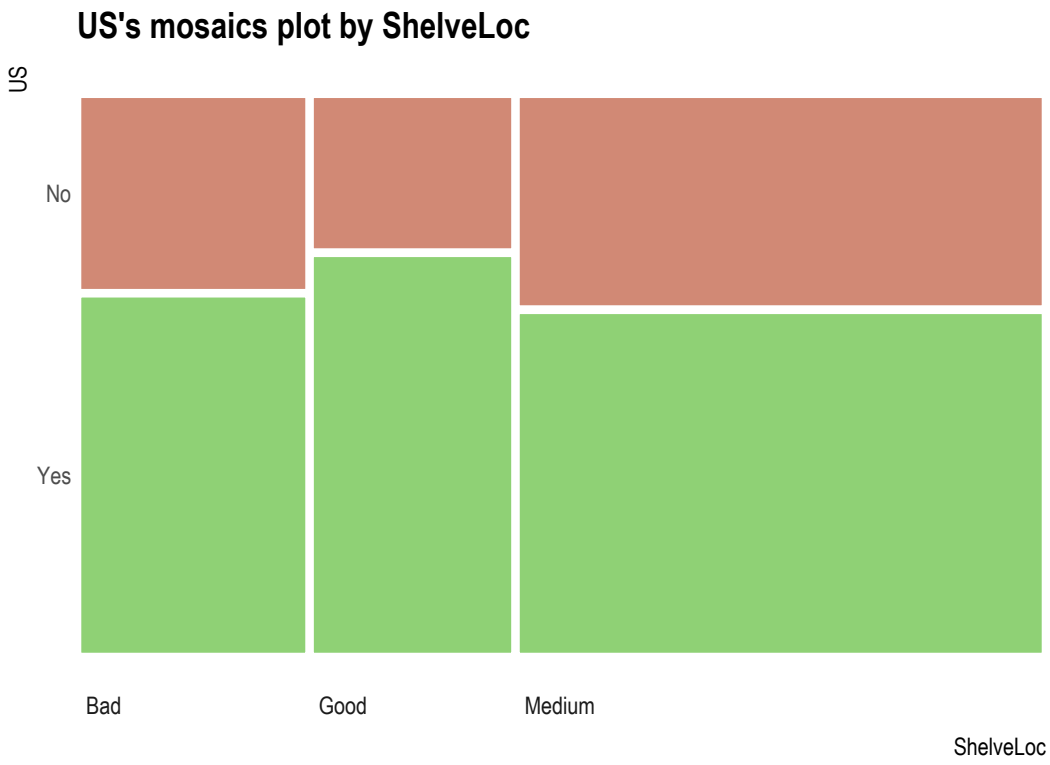


Figure 4.9: ShelveLoc

Urban

	No	Yes	Sum
No	46	70	116
Yes	94	185	279
NA	2	3	5
Sum	**142**	**258**	**400**

	No	Yes	Sum
No	32.39	27.13	29.00
Yes	66.20	71.71	69.75
NA	1.41	1.16	1.25
Sum	**100.00**	**100.00**	**100.00**

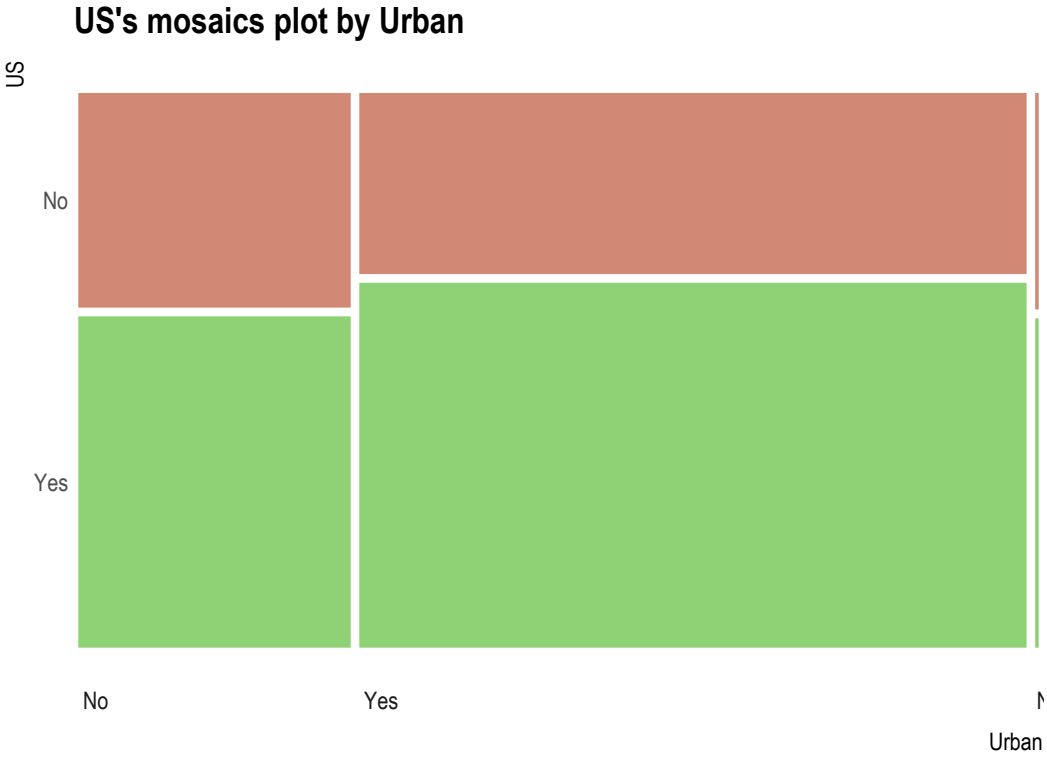


Figure 4.10: Urban

4.2 Grouped Relationship Between Variables

4.2.1 Grouped Correlation Coefficient

Table 4.9: The correlation coefficients (0.5 or more)

US	Variable1	Variable2	Correlation Coefficient
No	Price	CompPrice	0.638
No	Price	Sales	-0.529
Yes	Price	CompPrice	0.550

4.2.2 Grouped Correlation Plot of Numerical Variables

- Grouped Correlation Case of (US == No)

Figure 4.11: Correlation Matrix Plot (US == No)

- Grouped Correlation Case of (US == Yes)

Figure 4.12: Correlation Matrix Plot (US == Yes)