REPORT SERIES WITH DLOOKR

Transformation Report

Author:
dlookr package

Version:
0.3.12

March 27, 2020

Contents

Chapter 1

Imputation

1.1 Missing Values

1.1.1 Missing values imputation information

The variables for imputate missing values are as follows.

```
There are no variables including missing values.
```

1.2 Outliers

1.2.1 Outliers imputation information

The numerical variables for imputate outliers are as follows.

```
There are no variables including outliers.
```

Chapter 2

Resolving Skewness

2.1 Skewed variables information

The numerical variables for resolving skewness are as follows.

```
There are no variables including skewed.
```

Chapter 3

Binning

3.1 Numerical Variables for Binning

The numerical variables for binning are as follows.

```
displ, year, cyl, cty
```

3.2 Binning

3.2.1 displ

Binning with quantile

Table 3.1: Frequency of Levels : displ with 'quantile'

levels	freq	rate
[1.6,2]	43	0.1838
(2,2.4]	19	0.0812
(2.4,2.5]	20	0.0855
(2.5,3]	26	0.1111
(3,3.5]	24	0.1026
(3.5,4]	31	0.1325
(4,4.7]	33	0.1410
(4.7,5.3]	13	0.0556
(5.3,7]	25	0.1068

Information of Binning

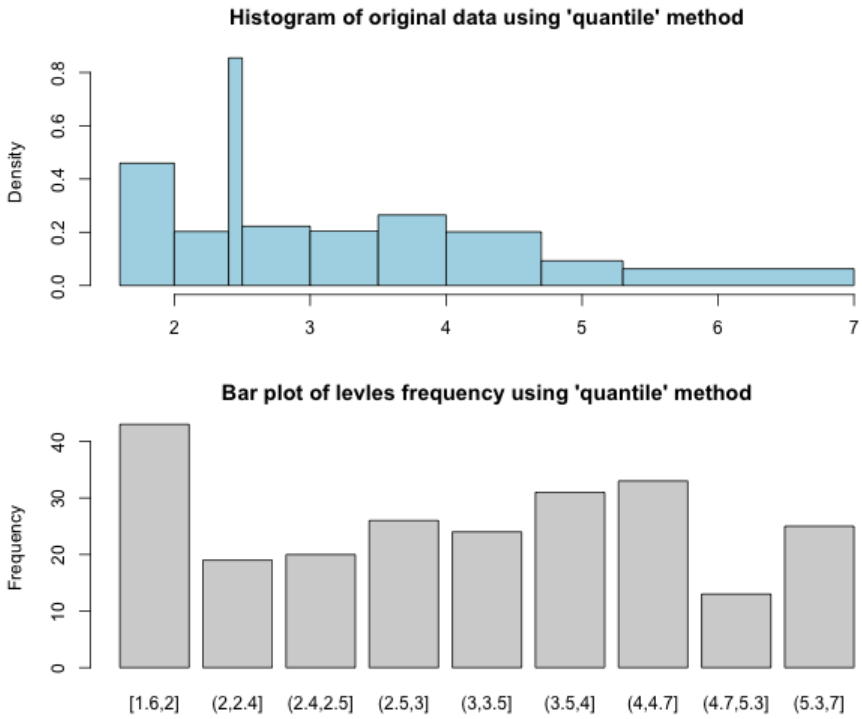

Figure 3.1: displ - quantile

Binning with equal

Table 3.2: Frequency of Levels : displ with 'equal'

levels	freq	rate
[1.6,2.2]	49	0.2094
(2.2,2.8]	51	0.2179
(2.8,3.4]	27	0.1154
(3.4,4]	36	0.1538
(4,4.6]	16	0.0684
(4.6,5.2]	24	0.1026
(5.2,5.8]	23	0.0983
(5.8,6.4]	6	0.0256
(6.4,7]	2	0.0085

Information of Binning

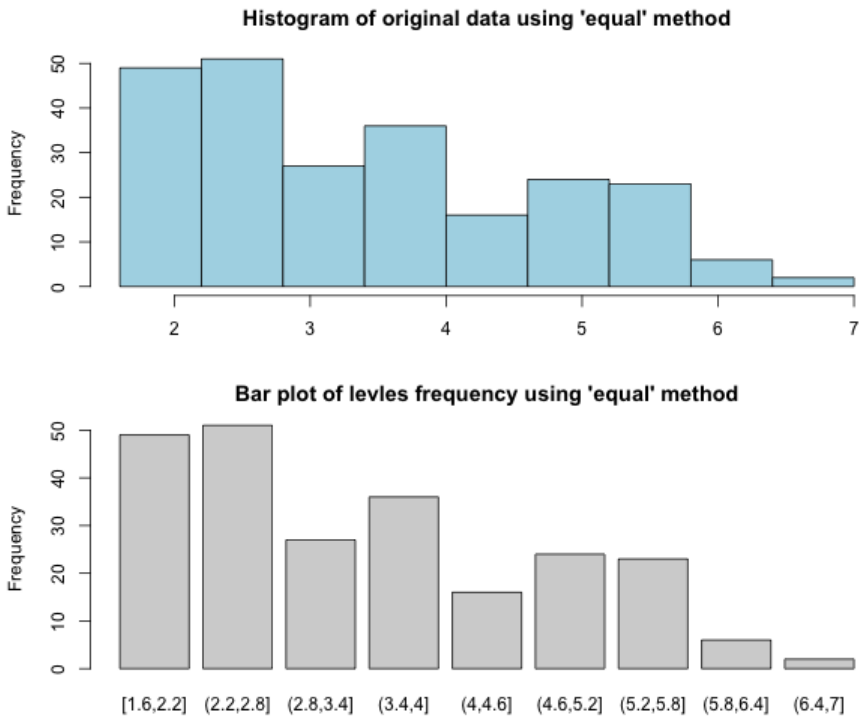

Figure 3.2: displ - equal

Binning with pretty

Table 3.3: Frequency of Levels : displ with 'pretty'

levels	freq	rate
[1.5,2]	43	0.1838
(2,2.5]	39	0.1667
(2.5,3]	26	0.1111
(3,3.5]	24	0.1026
(3.5,4]	31	0.1325
(4,4.5]	5	0.0214
(4.5,5]	30	0.1282
(5,5.5]	19	0.0812
(5.5,6]	12	0.0513
(6,6.5]	4	0.0171
(6.5,7]	1	0.0043

Information of Binning

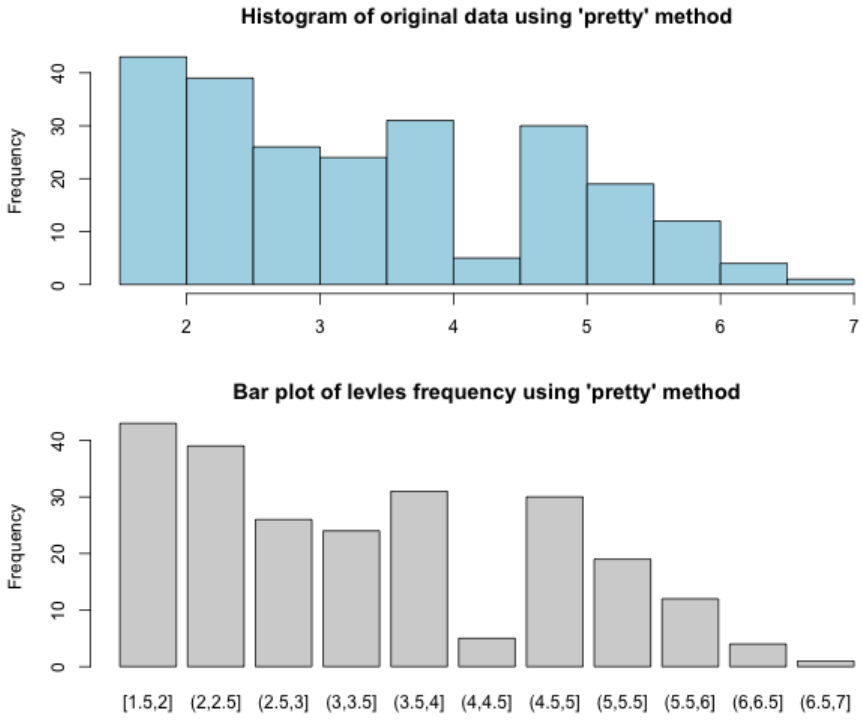

Figure 3.3: displ - pretty

Binning with kmeans

Table 3.4: Frequency of Levels : displ with 'kmeans'

levels	freq	rate
[1.6,1.7]	5	0.0214
(1.7,2.3]	44	0.1880
(2.3,2.9]	51	0.2179
(2.9,3.55]	32	0.1368
(3.55,4.3]	35	0.1496
(4.3,4.85]	29	0.1239
(4.85,5.5]	21	0.0897
(5.5,6.05]	12	0.0513
(6.05,7]	5	0.0214

Information of Binning

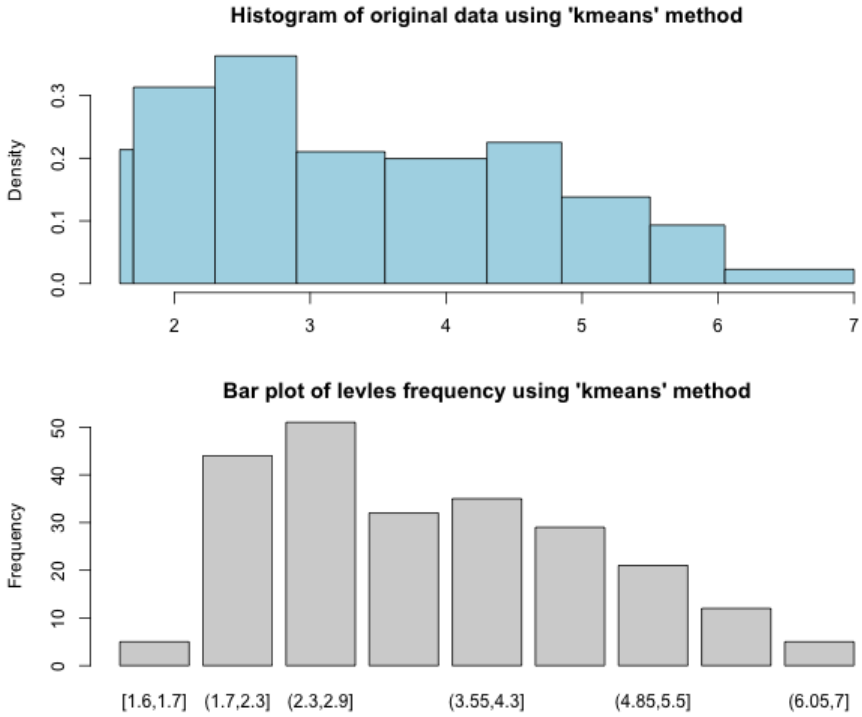

Figure 3.4: displ - kmeans

Binning with bclust

Table 3.5: Frequency of Levels : displ with 'bclust'

levels	freq	rate
[1.6,2.3]	49	0.2094
(2.3,3.2]	65	0.2778
(3.2,3.55]	18	0.0769
(3.55,4.1]	31	0.1325
(4.1,4.85]	33	0.1410
(4.85,5.5]	21	0.0897
(5.5,6.35]	15	0.0641
(6.35,6.75]	1	0.0043
(6.75,7]	1	0.0043

Information of Binning

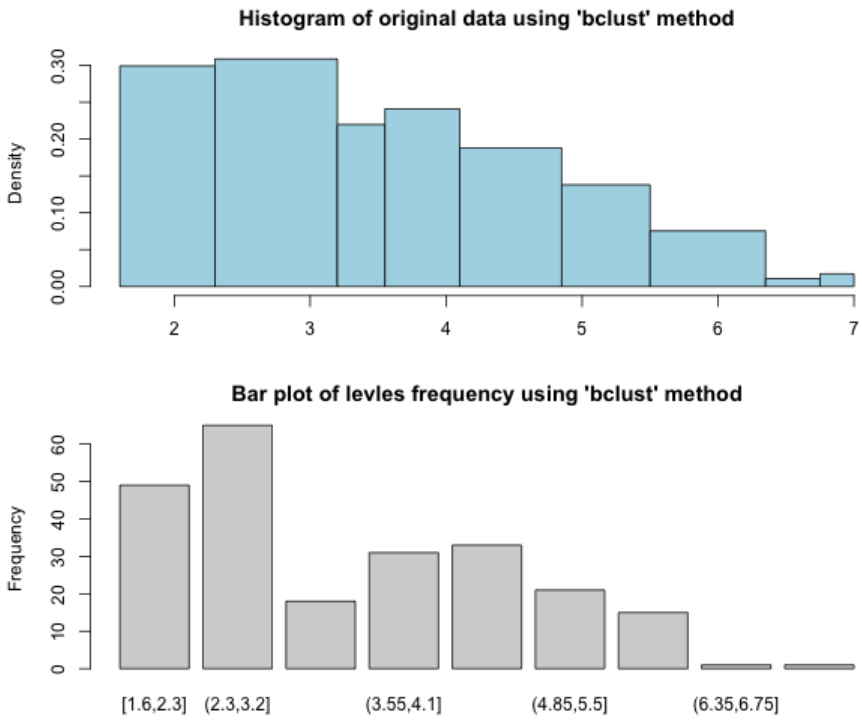

Figure 3.5: displ - bclust

3.2.2 year

Binning with quantile

Table 3.6: Frequency of Levels : year with 'quantile'

levels	freq	rate
[1.99e+03,2e+03]	117	0.5
(2e+03,2.01e+03]	117	0.5

Information of Binning

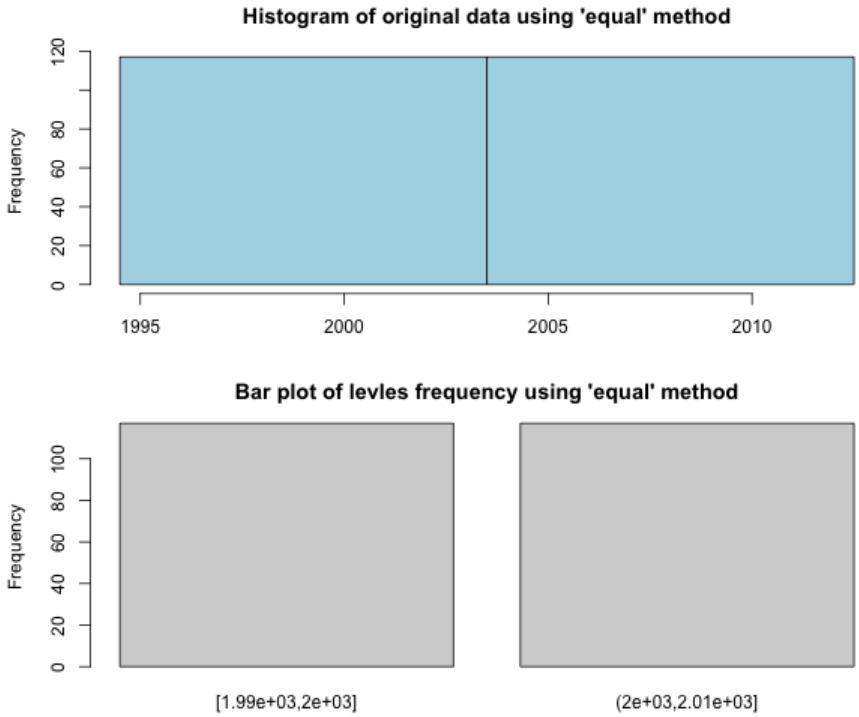

Figure 3.6: year - quantile

Binning with equal

Table 3.7: Frequency of Levels : year with 'equal'

levels	freq	rate
[1.99e+03,2e+03]	117	0.5
(2e+03,2.01e+03]	117	0.5

Information of Binning

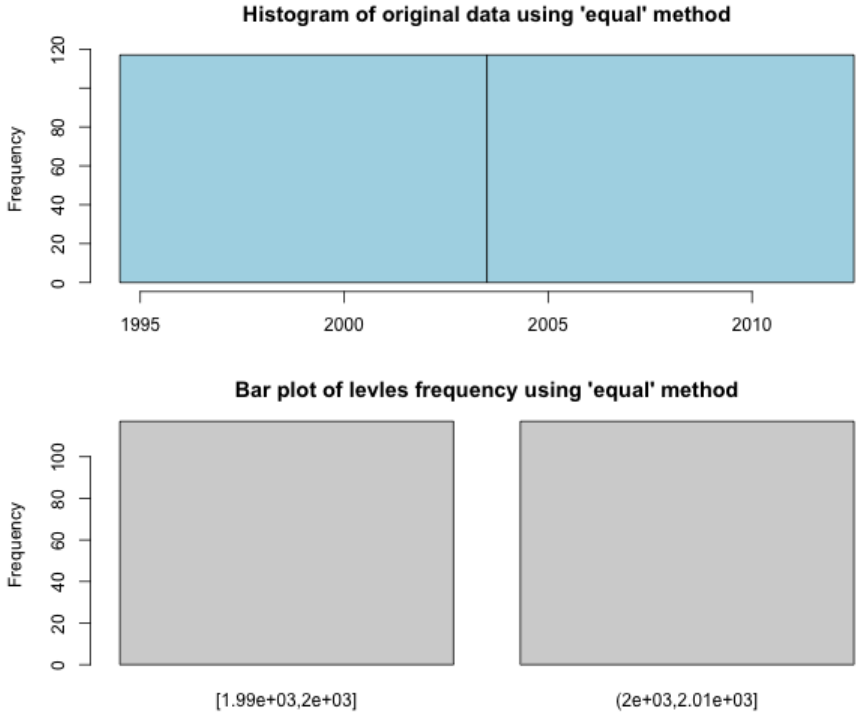

Figure 3.7: year - equal

Binning with pretty

Table 3.8: Frequency of Levels : year with 'pretty'

levels	freq	rate
[1.99e+03,2e+03]	117	0.5
(2e+03,2.01e+03]	117	0.5

Information of Binning

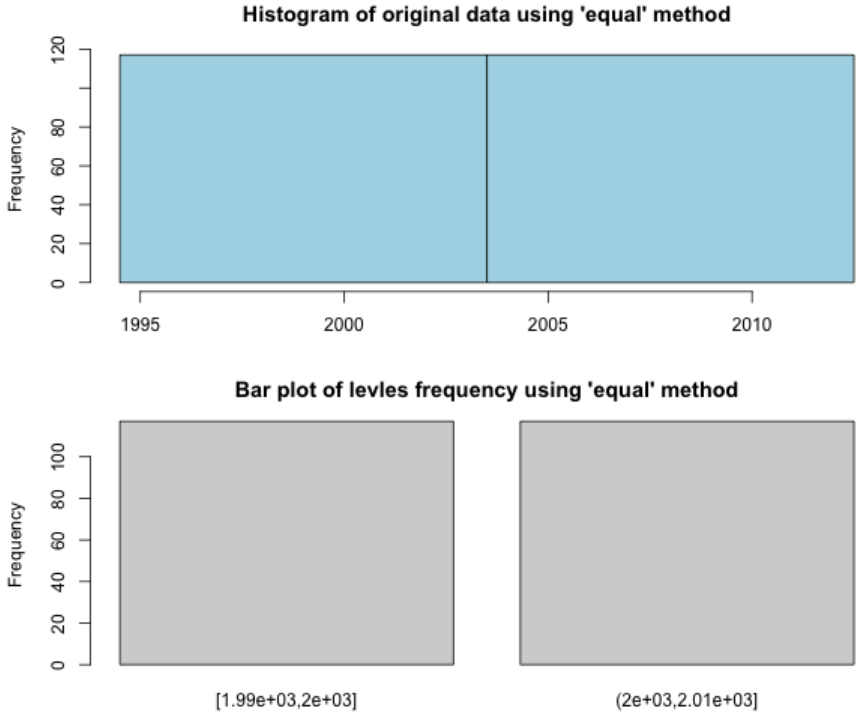

Figure 3.8: year - pretty

Binning with kmeans

Table 3.9: Frequency of Levels : year with 'kmeans'

levels	freq	rate
[1.99e+03,2e+03]	117	0.5
(2e+03,2.01e+03]	117	0.5

Information of Binning

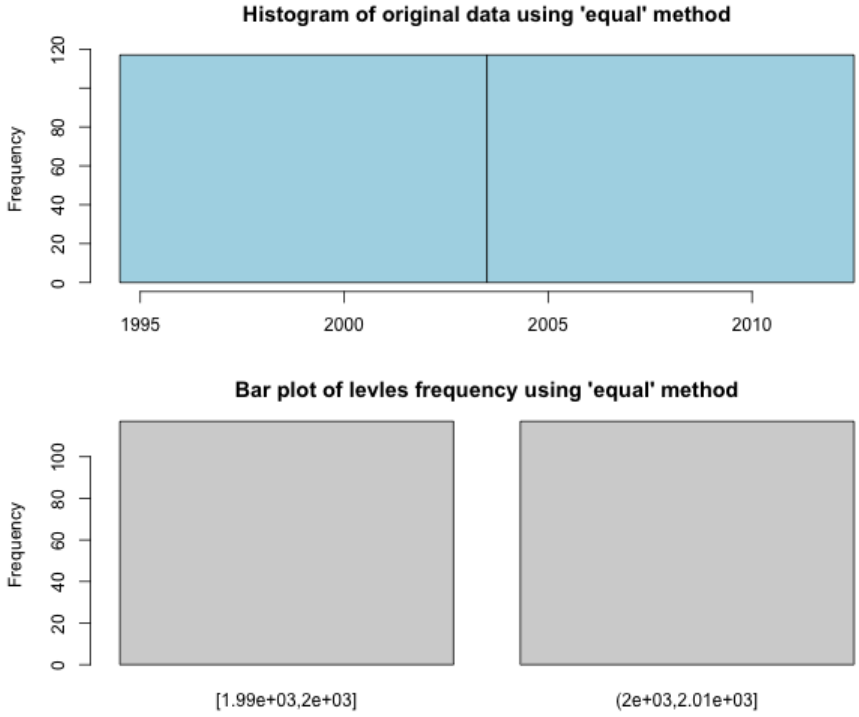

Figure 3.9: year - kmeans

Binning with bclust

Table 3.10: Frequency of Levels : year with 'bclust'

levels	freq	rate
[1.99e+03,2e+03]	117	0.5
(2e+03,2.01e+03]	117	0.5

Information of Binning

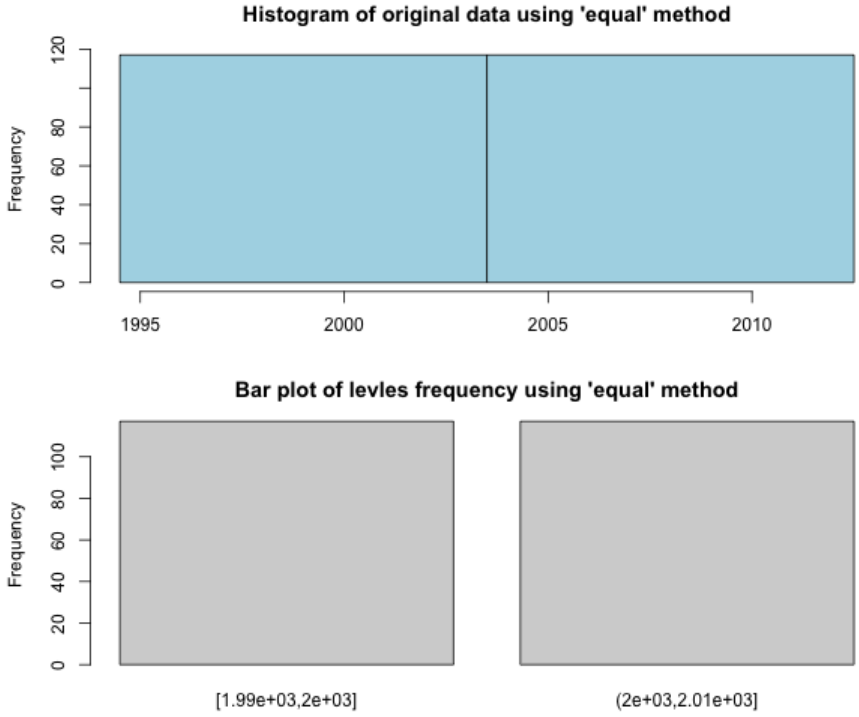

Figure 3.10: year - bclust

3.2.3 cyl

Binning with quantile

Table 3.11: Frequency of Levels : cyl with 'quantile'

levels	freq	rate
[3.5,4.5]	81	0.3462
(4.5,5.5]	4	0.0171
(5.5,7]	79	0.3376
(7,9]	70	0.2991

Information of Binning

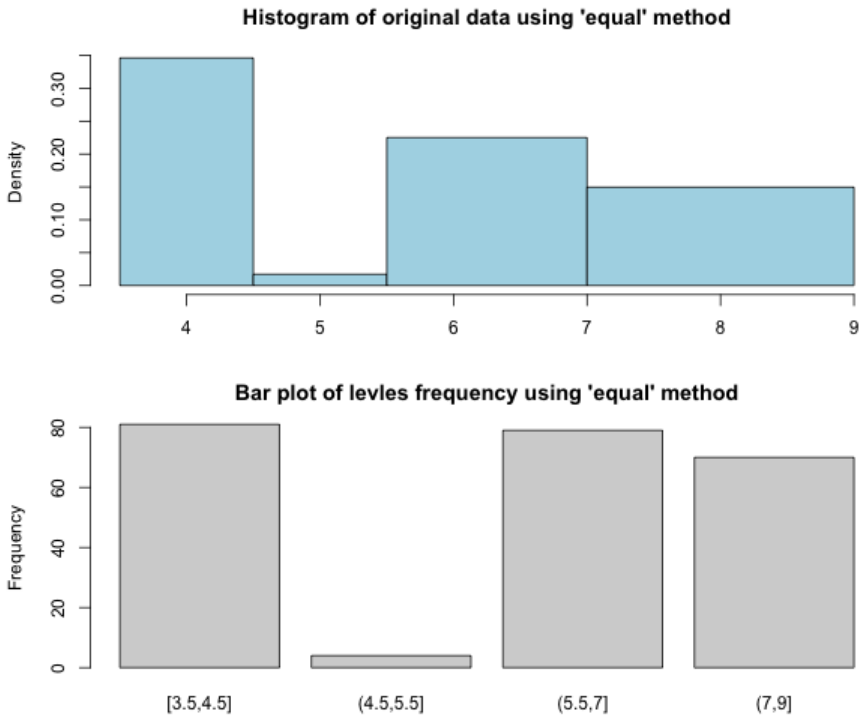

Figure 3.11: cyl - quantile

Binning with equal

Table 3.12: Frequency of Levels : cyl with 'equal'

levels	freq	rate
[3.5,4.5]	81	0.3462
(4.5,5.5]	4	0.0171
(5.5,7]	79	0.3376
(7,9]	70	0.2991

Information of Binning

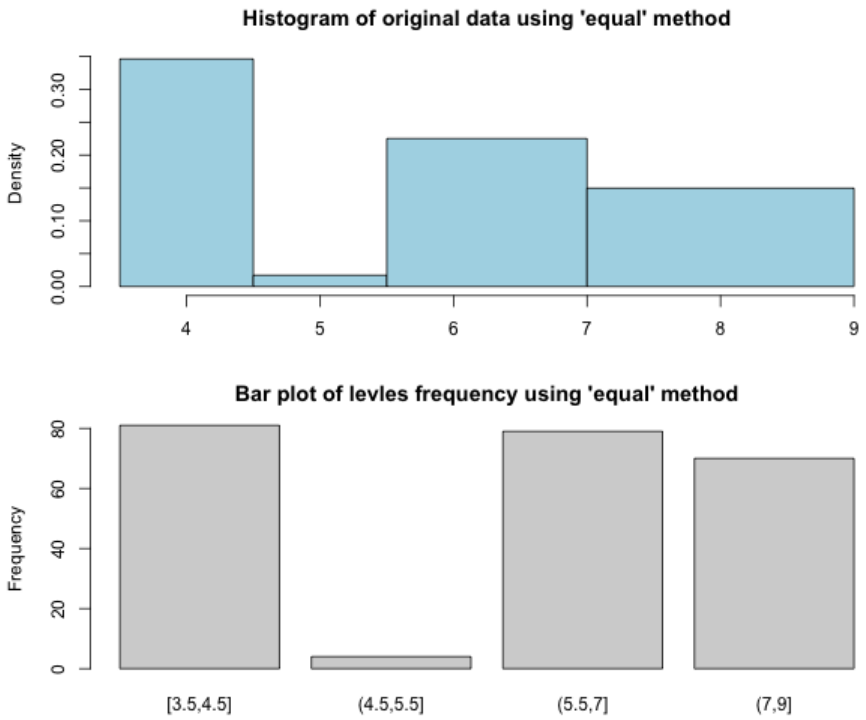

Figure 3.12: cyl - equal

Binning with pretty

Table 3.13: Frequency of Levels : cyl with 'pretty'

levels	freq	rate
[3.5,4.5]	81	0.3462
(4.5,5.5]	4	0.0171
(5.5,7]	79	0.3376
(7,9]	70	0.2991

Information of Binning

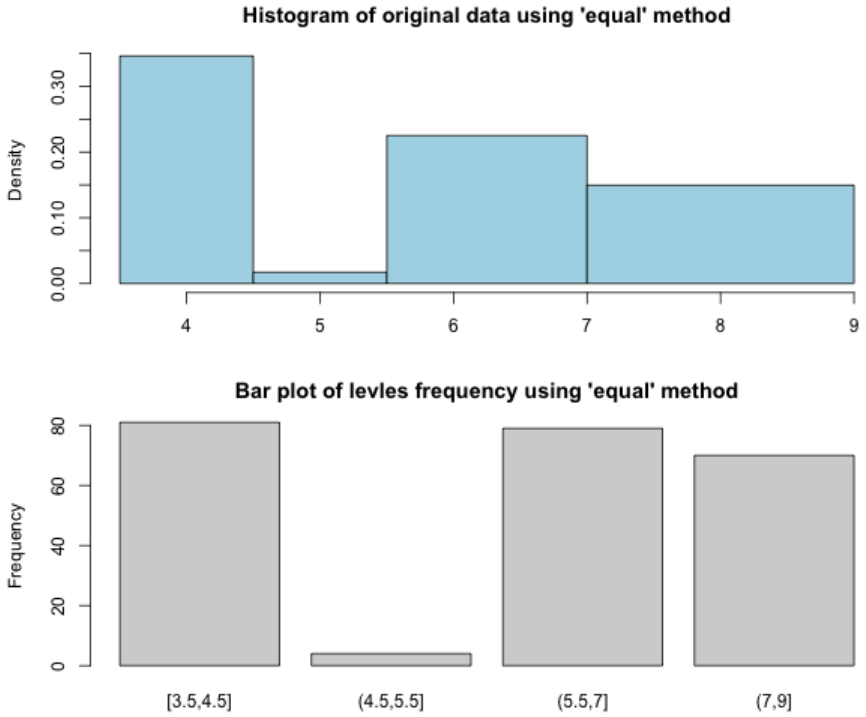

Figure 3.13: cyl - pretty

Binning with kmeans

Table 3.14: Frequency of Levels : cyl with 'kmeans'

levels	freq	rate
[3.5,4.5]	81	0.3462
(4.5,5.5]	4	0.0171
(5.5,7]	79	0.3376
(7,9]	70	0.2991

Information of Binning

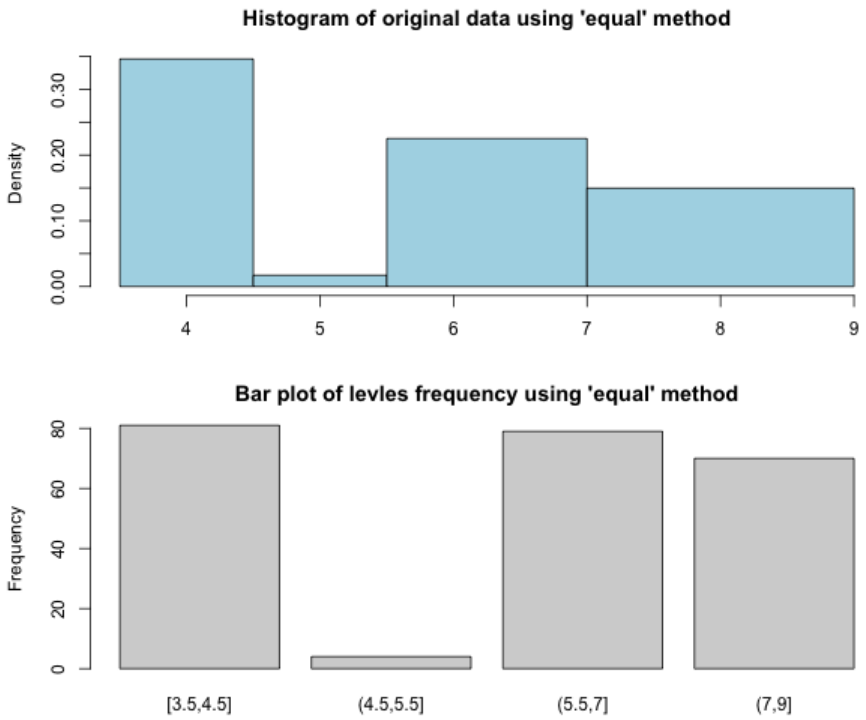

Figure 3.14: cyl - kmeans

Binning with bclust

Table 3.15: Frequency of Levels : cyl with 'bclust'

levels	freq	rate
[3.5,4.5]	81	0.3462
(4.5,5.5]	4	0.0171
(5.5,7]	79	0.3376
(7,9]	70	0.2991

Information of Binning

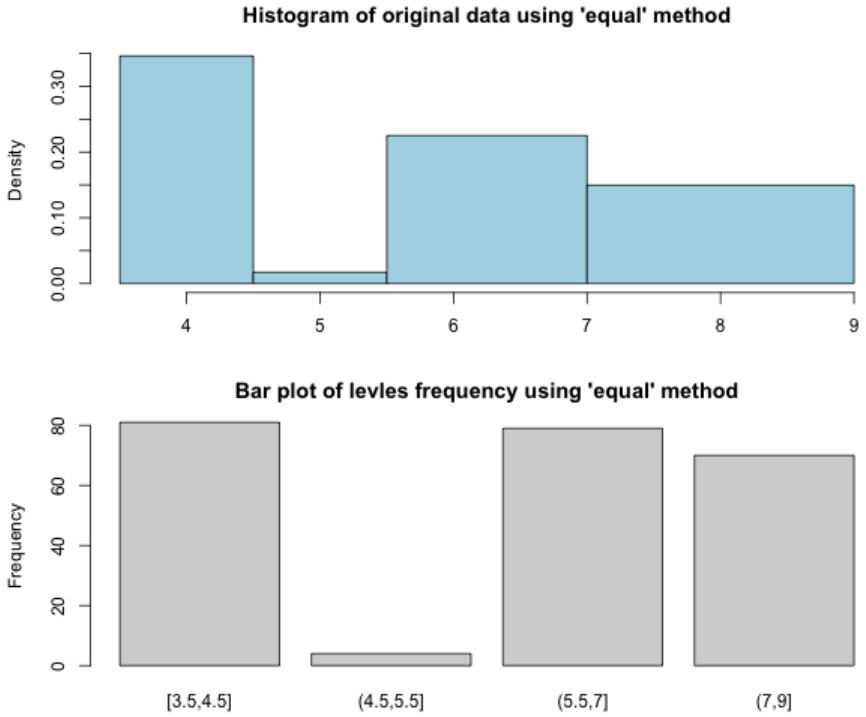

Figure 3.15: cyl - bclust

3.2.4 cty

Binning with quantile

Table 3.16: Frequency of Levels : cty with 'quantile'

levels	freq	rate
[9,12]	33	0.1410
(12,13]	21	0.0897
(13,15]	43	0.1838
(15,16]	19	0.0812
(16,17]	16	0.0684
(17,18]	26	0.1111
(18,20]	31	0.1325
(20,21]	23	0.0983
(21,35]	22	0.0940

Information of Binning

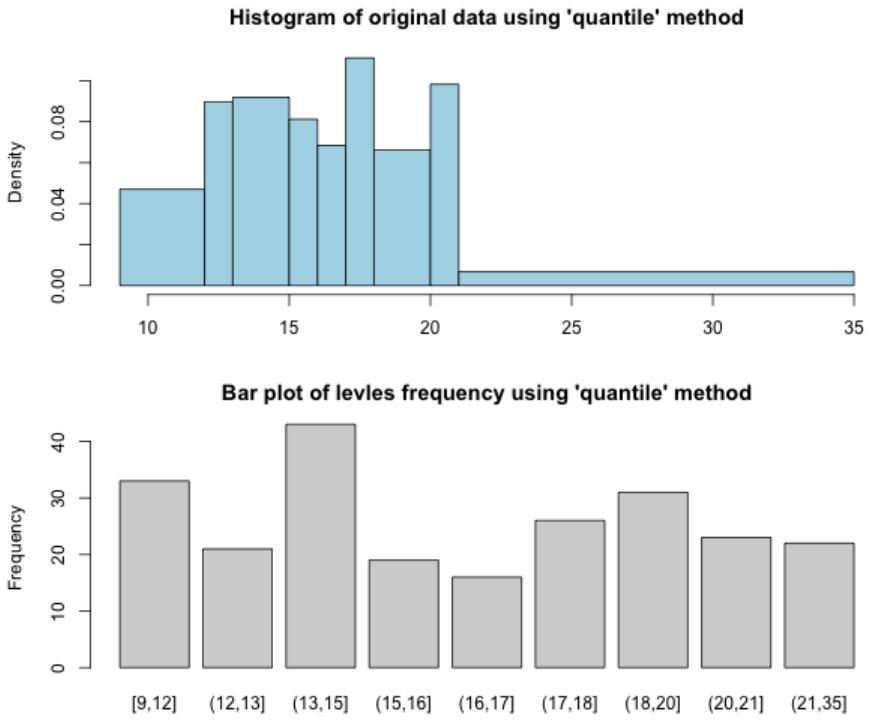

Figure 3.16: cty - quantile

Binning with equal

Table 3.17: Frequency of Levels : cty with 'equal'

levels	freq	rate
[9,11.9]	25	0.1068
(11.9,14.8]	48	0.2051
(14.8,17.7]	59	0.2521
(17.7,20.6]	57	0.2436
(20.6,23.4]	30	0.1282
(23.4,26.3]	10	0.0427
(26.3,29.2]	3	0.0128
(32.1,35]	2	0.0085

Information of Binning

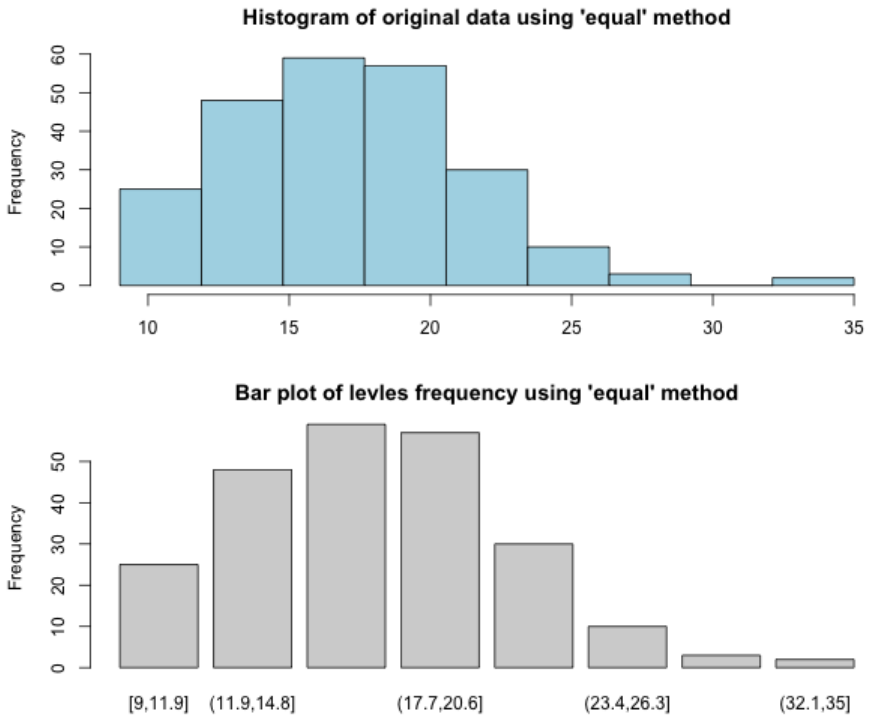

Figure 3.17: cty - equal

Binning with pretty

Table 3.18: Frequency of Levels : cty with 'pretty'

levels	freq	rate
[5,10]	5	0.0214
(10,15]	92	0.3932
(15,20]	92	0.3932
(20,25]	37	0.1581
(25,30]	6	0.0256
(30,35]	2	0.0085

Information of Binning

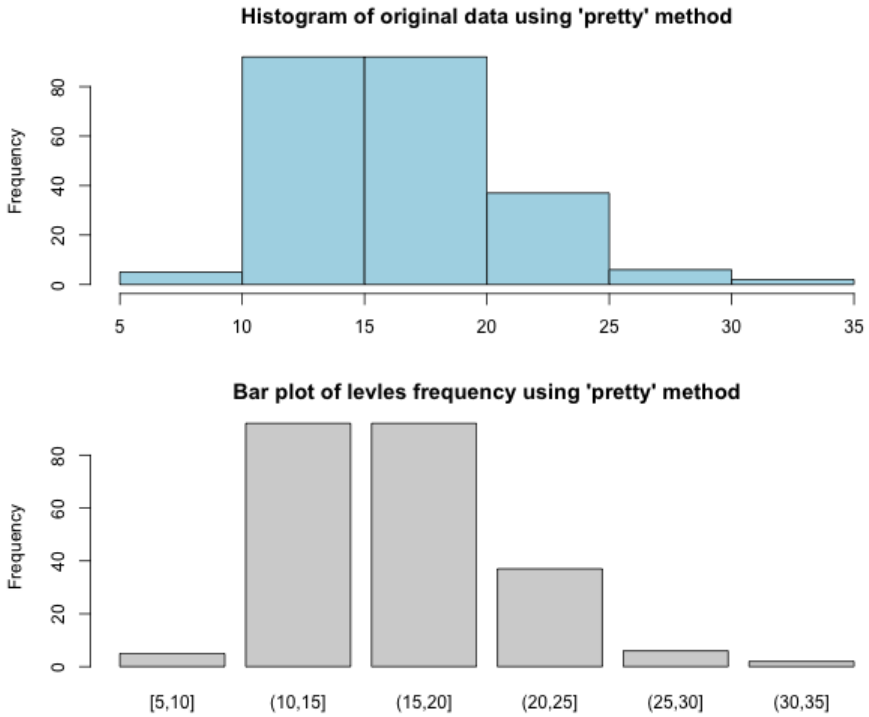

Figure 3.18: cty - pretty

Binning with kmeans

Table 3.19: Frequency of Levels : cty with 'kmeans'

levels	freq	rate
[9,11.5]	25	0.1068
(11.5,13.5]	29	0.1239
(13.5,15.5]	43	0.1838
(15.5,17.5]	35	0.1496
(17.5,19.5]	46	0.1966
(19.5,20.5]	11	0.0470
(20.5,23.5]	30	0.1282
(23.5,31]	13	0.0556
(31,35]	2	0.0085

Information of Binning

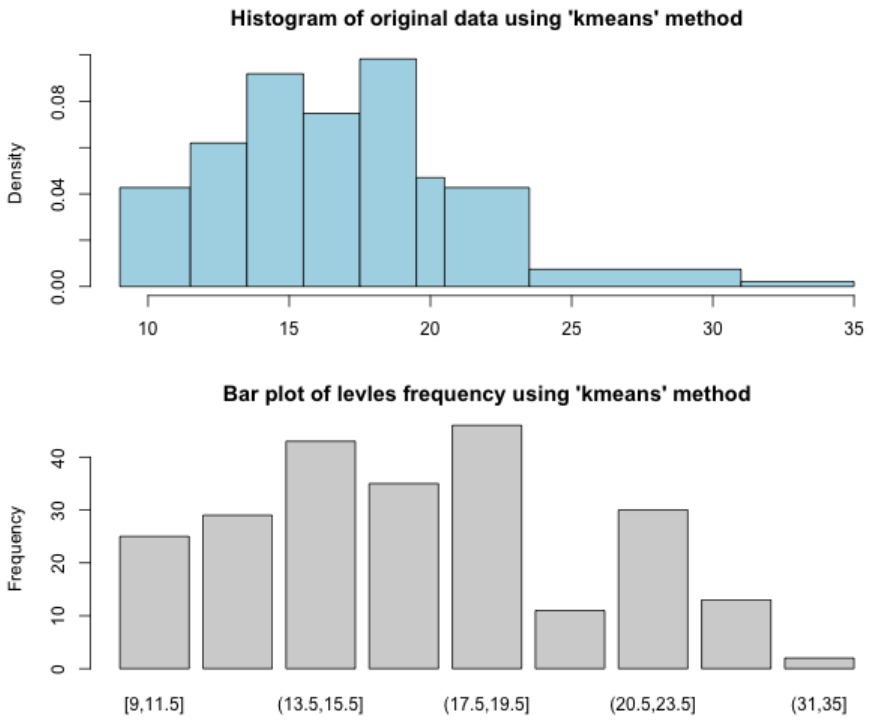

Figure 3.19: cty - kmeans

Binning with bclust

Error in bclust(x = var, centers = n, ...) : Could not find valid cluster solution in 20 replications

Table 3.20: Frequency of Levels : cty with 'bclust'

levels	freq	rate
[8.82,9.99]	5	0.0214
(9.99,13.5]	49	0.2094
(13.5,15.5]	43	0.1838
(15.5,18.5]	61	0.2607
(18.5,23.2]	61	0.2607
(23.2,26]	10	0.0427
(26,28.5]	2	0.0085
(28.5,31]	1	0.0043
(31,35.1]	2	0.0085

Information of Binning

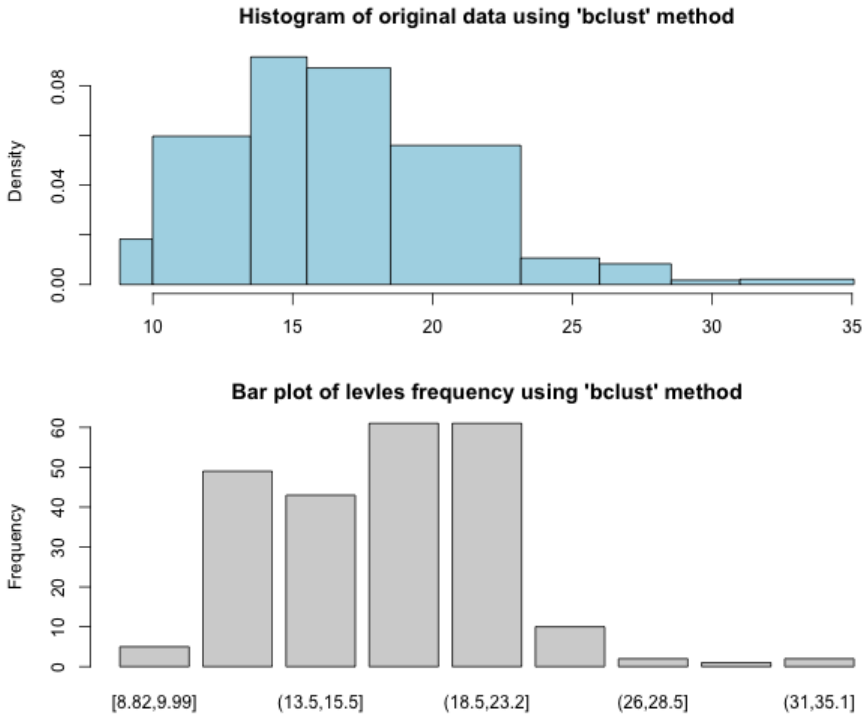

Figure 3.20: cty - bclust

3.3 Optimal Binning

For the scoring modeling, optimal binning is performed considering the characteristics of numerical data. This work uses the recursive partitioning technique known as supervised discretization when categorizing numeric data as categorical data.

The target variable is not a binary class.