REPORT SERIES WITH DLOOKR

Data Quality Diagnosis Report

Author:
dlookr package

Version:
0.3.6

March 11, 2019

Contents

Chapter 1

Diagnose Data

1.1 Overview of Diagnosis

1.1.1 List of all variables quality

Table 1.1: Data quality overview table

variables	type	missing (n)	missing (%)	unique (n)	unique (n/N)
sex	character	0	0.000	2	0.000
age	numeric	0	0.000	65	0.010
race	factor	0	0.000	6	0.001
juv_fel_count	numeric	0	0.000	11	0.002
juv_misd_count	numeric	0	0.000	10	0.002
juv_other_count	numeric	0	0.000	11	0.002
priors_count	numeric	0	0.000	33	0.005
c_charge_degree	character	0	0.000	12	0.002
c_charge_desc	character	4	0.061	391	0.060
is_recid	factor	0	0.000	2	0.000
jail_days	numeric	0	0.000	384	0.059
age_at_offense	numeric	849	12.988	4,523	0.692

1.1.2 Diagnosis of missing data

Table 1.2: Variables that include missing values

variables	type	missing (n)	missing (%)	unique (n)	unique (n/N)
age_at_offense	numeric	849	12.988	4,523	0.692
c_charge_desc	character	4	0.061	391	0.060

1.1.3 Diagnosis of unique data(Text and Category)

No variable with a high proportion greater than 0.5

1.1.4 Diagnosis of unique data(Numerical)

Table 1.3: Variables where the proportion of unique data is less than 0.1

variables	type	missing (n)	missing (%)	unique (n)	unique (n/N)
jail_days	numeric	0	0	384	0.059
age	numeric	0	0	65	0.010
priors_count	numeric	0	0	33	0.005
juv_fel_count	numeric	0	0	11	0.002
juv_other_count	numeric	0	0	11	0.002
juv_misd_count	numeric	0	0	10	0.002

1.2 Detailed data diagnosis

1.2.1 Diagnosis of categorical variables

Table 1.4: Categorical variable level top 10

variables	levels	N	freq	ratio(%)	rank
sex	Male	6,537	5,195	79.471	1
sex	Female	6,537	1,342	20.529	2
race	African-American	6,537	3,257	49.824	1
race	Caucasian	6,537	2,256	34.511	2
race	Hispanic	6,537	535	8.184	3
race	Other	6,537	441	6.746	4
race	Asian	6,537	33	0.505	5
race	Native American	6,537	15	0.229	6
c_charge_degree	(F3)	6,537	3,535	54.077	1
c_charge_degree	(M1)	6,537	1,840	28.147	2
c_charge_degree	(F2)	6,537	544	8.322	3
c_charge_degree	(M2)	6,537	452	6.914	4
c_charge_degree	(F1)	6,537	73	1.117	5
c_charge_degree	(F7)	6,537	50	0.765	6
c_charge_degree	(MO3)	6,537	34	0.520	7
c_charge_degree	(F5)	6,537	3	0.046	8
c_charge_degree	(NI0)	6,537	3	0.046	9
c_charge_degree	(CO3)	6,537	1	0.015	10
c_charge_degree	(TCX)	6,537	1	0.015	11
c_charge_degree	(X)	6,537	1	0.015	12
c_charge_desc	Battery	6,537	1,302	19.917	1
c_charge_desc	arrest case no charge	6,537	849	12.988	2
c_charge_desc	Grand Theft in the 3rd Degree	6,537	457	6.991	3
c_charge_desc	Possession of Cocaine	6,537	411	6.287	4
c_charge_desc	Driving While License Revoked	6,537	166	2.539	5
c_charge_desc	Felony Battery (Dom Strang)	6,537	124	1.897	6
c_charge_desc	Felony Driving While Lic Suspd	6,537	122	1.866	7
c_charge_desc	Driving Under The Influence	6,537	101	1.545	8
c_charge_desc	Pos Cannabis W/Intent Sel/Del	6,537	97	1.484	9
c_charge_desc	Grand Theft (Motor Vehicle)	6,537	90	1.377	10
is_recid	0	6,537	4,317	66.039	1
is_recid	1	6,537	2,220	33.961	2

1.2.2 Diagnosis of numerical variables

Table 1.5: General list of numerical diagnosis

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
age	18.000	25.00	34.681	31.000	42.000	96.000	0	0	59
juv_fel_count	0.000	0.00	0.056	0.000	0.000	20.000	6,335	0	202
juv_misd_count	0.000	0.00	0.074	0.000	0.000	12.000	6,219	0	318
juv_other_count	0.000	0.00	0.100	0.000	0.000	11.000	6,102	0	435
priors_count	0.000	0.00	2.842	1.000	4.000	36.000	2,455	0	423
jail_days	29.000	365.00	1,142.105	366.000	1,096.000	10,592.000	0	0	873
age_at_offense	15.044	23.29	32.817	29.679	40.474	94.518	0	0	49

1.2.3 List of numerical diagnosis (zero)

Table 1.6: List of numerical diagnosis (zero)

variables	min	median	max	zero	zero ratio(%)
juv_fel_count	0	0	20	6,335	96.910
juv_misd_count	0	0	12	6,219	95.135
juv_other_count	0	0	11	6,102	93.346
priors_count	0	1	36	2,455	37.555

1.2.4 List of numerical diagnosis (minus)

No numeric variable with negative value

Chapter 2

Diagnose Outliers

2.1 Overview of Diagnosis

2.1.1 Diagnosis of numerical variable outliers

Table 2.1: Diagnosis of numerical variable outliers

variables	min	median	max	outlier	outlier ratio(%)
jail_days	29.000	366.000	10,592.000	873	13.355
juv_other_count	0.000	0.000	11.000	435	6.654
priors_count	0.000	1.000	36.000	423	6.471
juv_misd_count	0.000	0.000	12.000	318	4.865
juv_fel_count	0.000	0.000	20.000	202	3.090
age	18.000	31.000	96.000	59	0.903
age_at_offense	15.044	29.679	94.518	49	0.750

2.2 Detailed outliers diagnosis

variable : jail_days

Table 2.2: Outliers information of jail_days

Measures	Values
Outliers count	873.00
Outliers ratio (%)	13.35
Mean of outliers	4,683.88
Mean with outliers	1,142.10
Mean without outliers	596.21

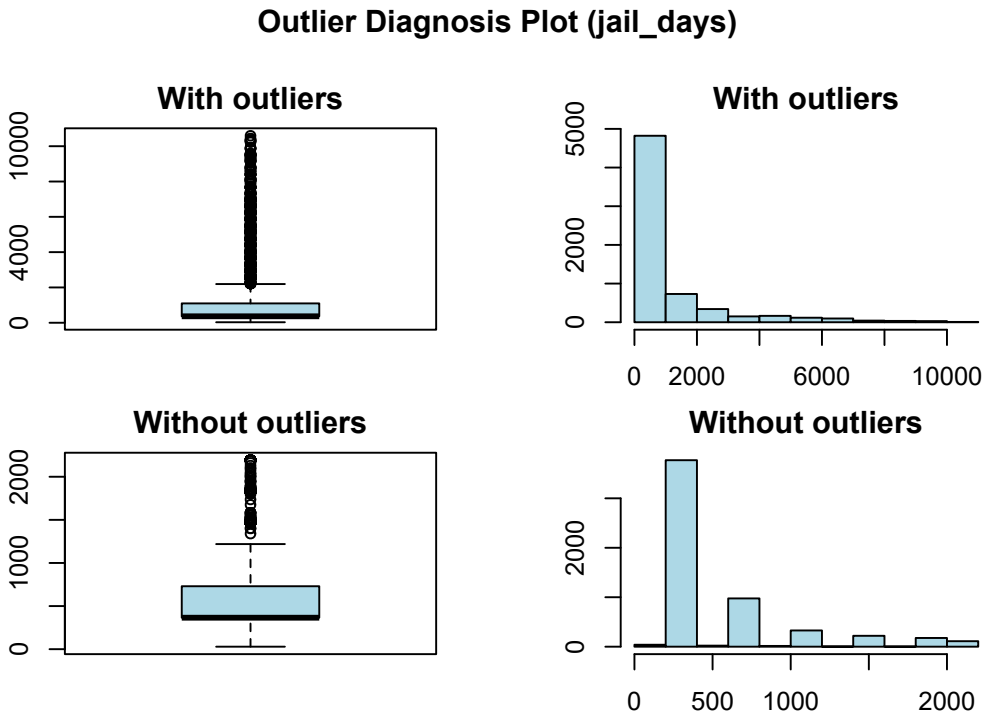

Figure 2.1: Distribution of jail_days

variable : juv_other_count

Table 2.3: Outliers information of juv_other_count

Measures	Values
Outliers count	435.00
Outliers ratio (%)	6.65
Mean of outliers	1.50
Mean with outliers	0.10
Mean without outliers	0.00

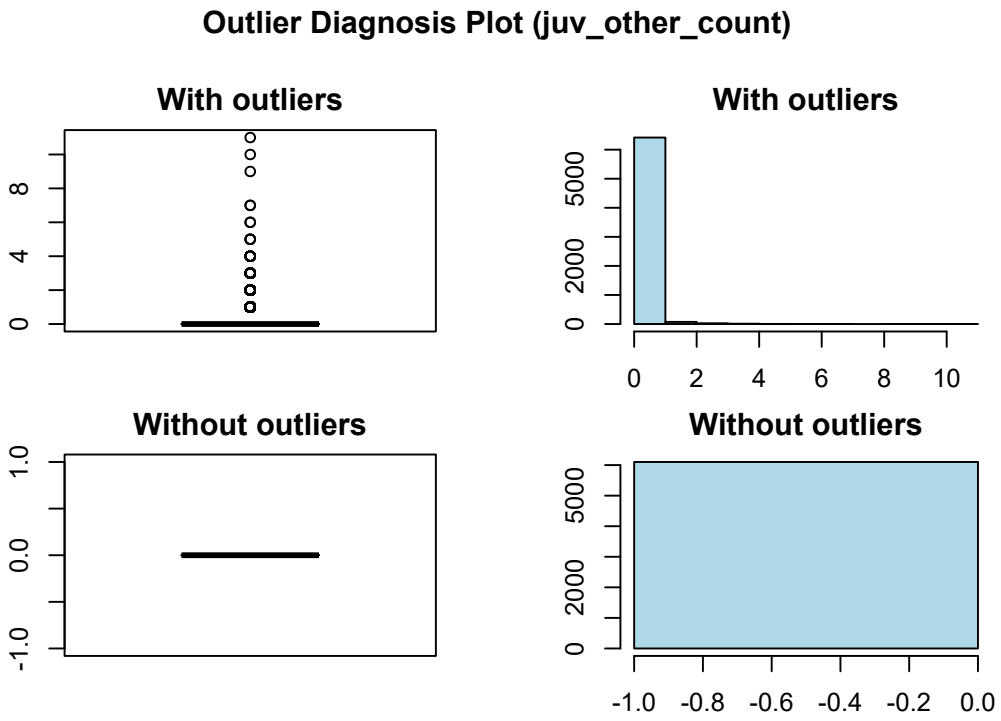

Figure 2.2: Distribution of juv_other_count

variable : priors_count

Table 2.4: Outliers information of priors_count

Measures	Values
Outliers count	423.00
Outliers ratio (%)	6.47
Mean of outliers	15.70
Mean with outliers	2.84
Mean without outliers	1.95

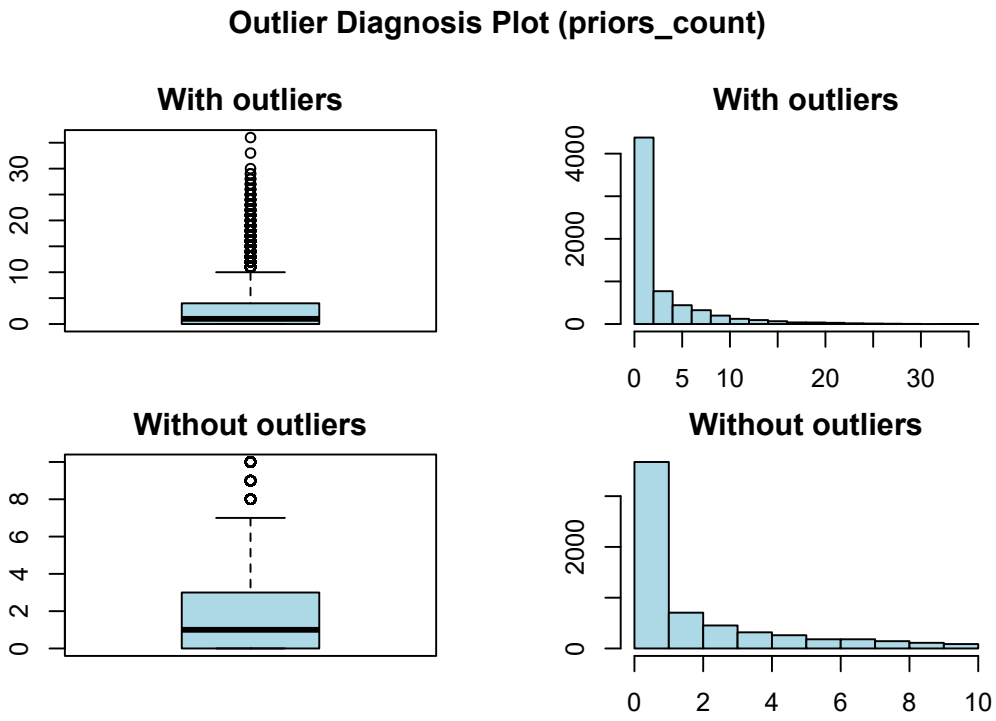

Figure 2.3: Distribution of priors_count

variable : juv_misd_count

Table 2.5: Outliers information of juv_misd_count

Measures	Values
Outliers count	318.00
Outliers ratio (%)	4.86
Mean of outliers	1.52
Mean with outliers	0.07
Mean without outliers	0.00

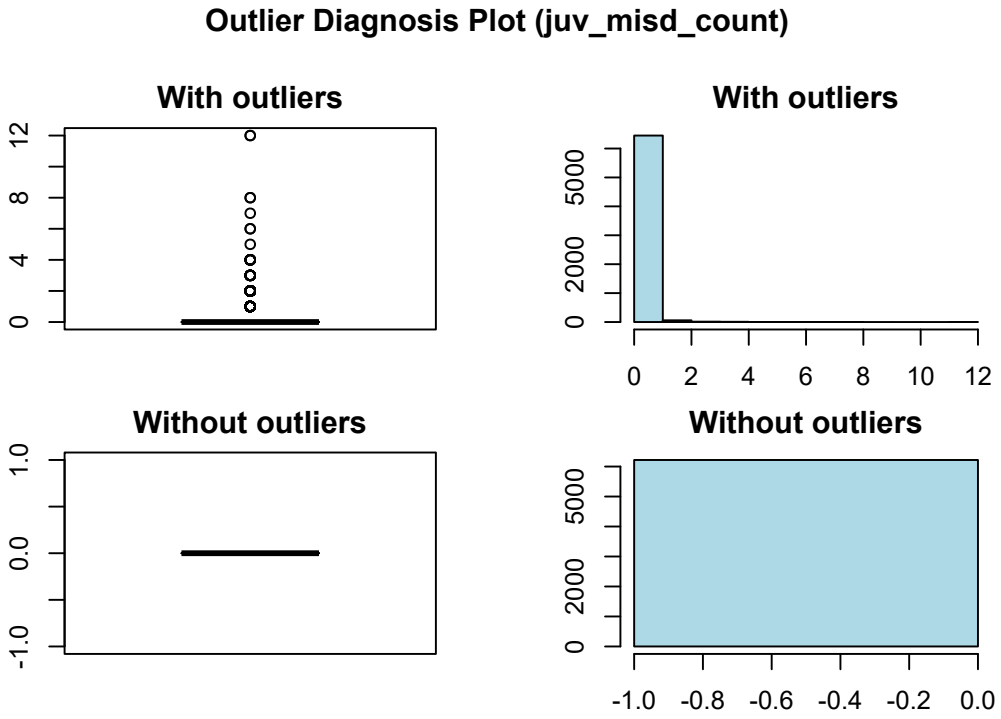

Figure 2.4: Distribution of juv_misd_count

variable : juv_fel_count

Table 2.6: Outliers information of juv_fel_count

Measures	Values
Outliers count	202.00
Outliers ratio (%)	3.09
Mean of outliers	1.83
Mean with outliers	0.06
Mean without outliers	0.00

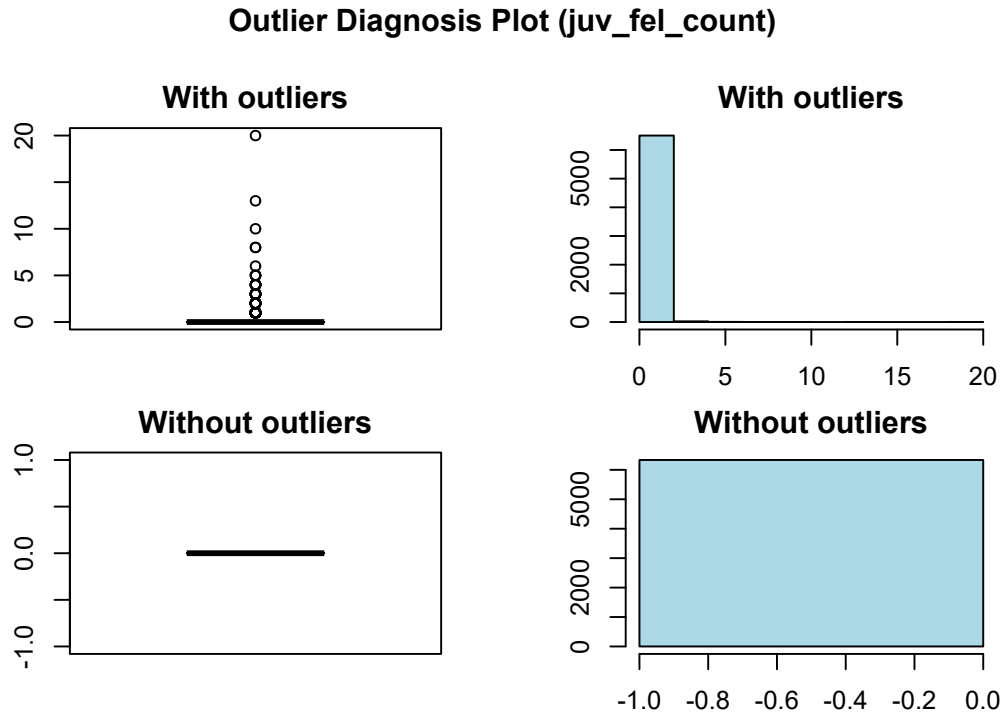

Figure 2.5: Distribution of juv_fel_count

variable : age

Table 2.7: Outliers information of age

Measures	Values
Outliers count	59.00
Outliers ratio (%)	0.90
Mean of outliers	71.68
Mean with outliers	34.68
Mean without outliers	34.34

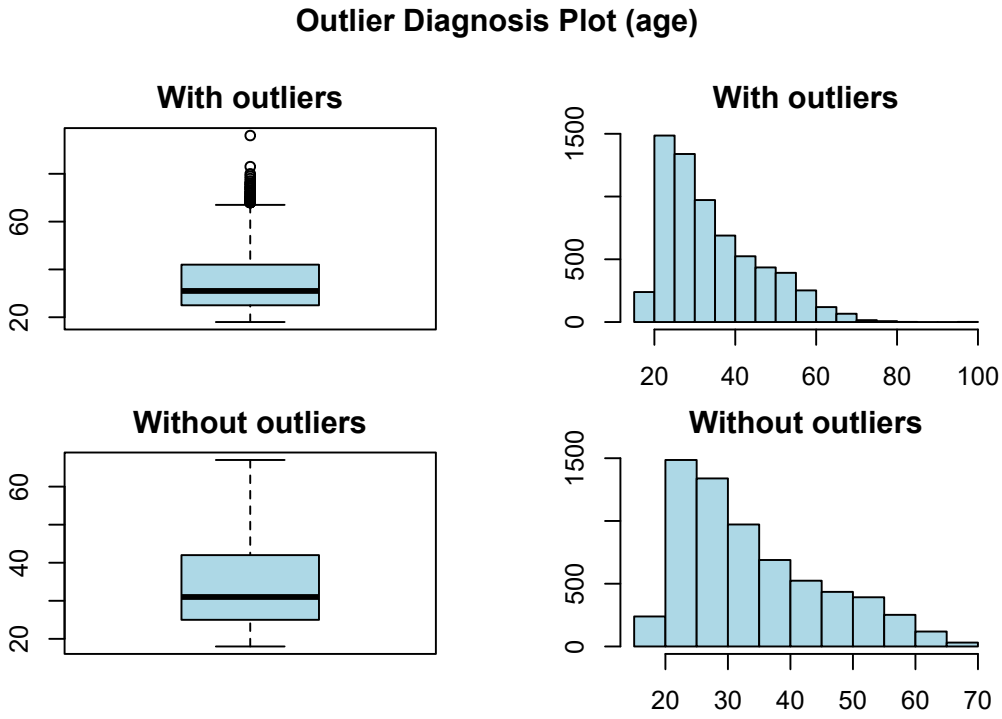

Figure 2.6: Distribution of age

variable : age_at_offense

Table 2.8: Outliers information of age_at_offense

Measures	Values
Outliers count	49.00
Outliers ratio (%)	0.75
Mean of outliers	70.16
Mean with outliers	32.82
Mean without outliers	32.49

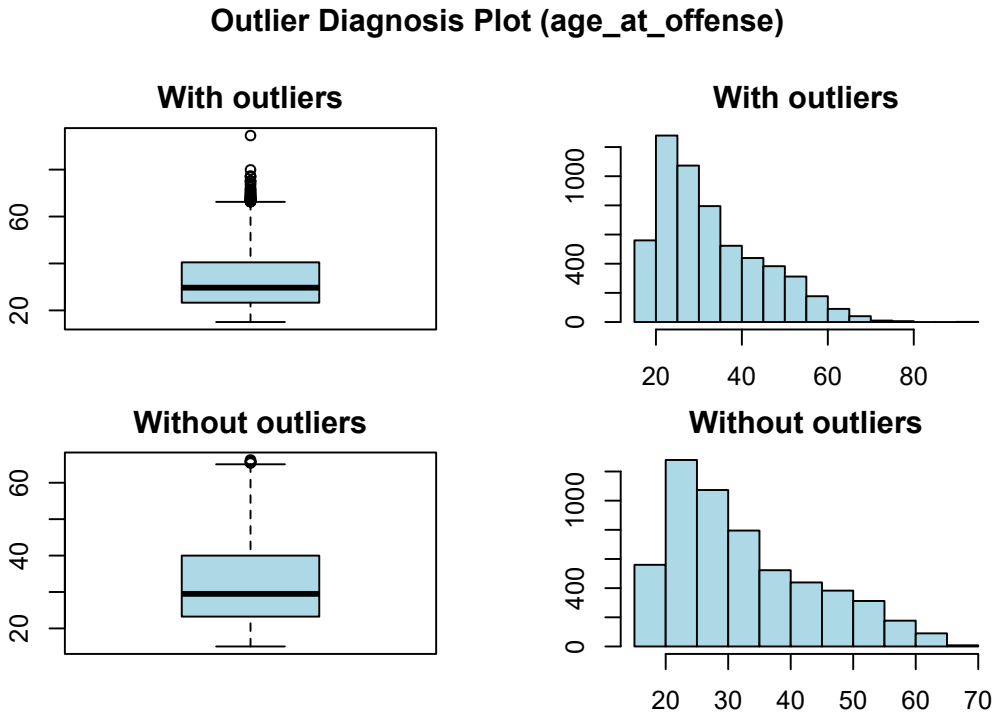

Figure 2.7: Distribution of age_at_offense