BODY COMPOSITION FAKTOREN UND LEBERTRANSPLANTATIONS-WARTELISTENMORTALITÄT

DATEN BESCHREIBUNG DES DATENSETS

(Report 2 von 3)

This report was created for an overview quality diagnosis of . data. It was created for **the purpose of judging the validity of variables** before conducting EDA.

Contents

Overview

Data Structures

division	metrics	value
size	observations	479
size	variables	33
size	values	15,807
size	memory size (KB)	0
duplicated	duplicate observation	0
missing	complete observation	479
missing	missing observation	0
missing	missing variables	0
missing	missing values	0

division	metrics	value
data type	numerics	20
data type	integers	0
data type	factors/ordered	13
data type	characters	0
data type	Dates	0
data type	POSIXcts	0
data type	others	0

Table 1: Data structures and types

Job Informations

division	metrics	value
dataset	dataset	.
dataset	dataset type	data.frame
job	samples	479 / 479 (100%)
job	created	2023-08-03 11:59:37.106222
job	created by	Eugen Malamutmann MD

Table 2: Job informations

Warnings

checks	judgements	removes
4	22	0

Table 3: Summary of warnings

warnings	status	recommand
status_changed_to_nt has a low cardinality. 2 (0.4%) distinct values	cardinality	judgement
known_mortality has a low cardinality. 5 (1%) distinct values	cardinality	judgement
status_changed_to_nt has 420 (87.68%) zeros	zero	check
exceptional_meld has 389 (81.21%) zeros	zero	check
mortality_increase has 15 (3.13%) zeros	zero	check
eat has 1 (0.21%) zeros	zero	check
exceptional_meld has 90 (18.79%) outliers	outlier	judgement
time has 71 (14.82%) outliers	outlier	judgement
status_changed_to_nt has 59 (12.32%) outliers	outlier	judgement
inr_listing has 47 (9.81%) outliers	outlier	judgement
mortality_increase has 45 (9.39%) outliers	outlier	judgement
platelets_listing has 25 (5.22%) outliers	outlier	judgement
known_mortality has 19 (3.97%) outliers	outlier	judgement
sodium_listing has 16 (3.34%) outliers	outlier	judgement
pat has 12 (2.51%) outliers	outlier	judgement
bmi has 7 (1.46%) outliers	outlier	judgement
muscle has 6 (1.25%) outliers	outlier	judgement
vat has 6 (1.25%) outliers	outlier	judgement
sat has 5 (1.04%) outliers	outlier	judgement
imat has 4 (0.84%) outliers	outlier	judgement
eat has 4 (0.84%) outliers	outlier	judgement
tat has 4 (0.84%) outliers	outlier	judgement

Table 4: Warnings in dataset and variables

	warnings	status	recommand
	warnings	**status**	**recommand**
23	age_by_listung has 3 (0.63%) outliers	outlier	judgement
24	bone has 3 (0.63%) outliers	outlier	judgement
25	height has 2 (0.42%) outliers	outlier	judgement
26	weight has 1 (0.21%) outliers	outlier	judgement

Table 4: Warnings in dataset and variables (continued)

Variables

variables	types	missing	cardinality	zero	minus	outlier
time	numeric					X
death	factor					
hu_listung	factor					
sex	factor					
blood_type	factor					
primary_diagnosis	factor					
hcc	factor					
child_pugh_score	factor					
dialysis_cat	factor					
icu_cat	factor					
ventilation_cat	factor					
catecholamine_cat	factor					
portal_vein_thrombosis	factor					
meld_score_category	factor					
age_by_listung	numeric					X
height	numeric					X
weight	numeric					X
bmi	numeric					X
inr_listing	numeric					X
sodium_listing	numeric					X
platelets_listing	numeric					X
exceptional_meld	numeric			X		X
bone	numeric					X
muscle	numeric					X
sat	numeric					X

Table 5: List of variables diagnosis

variables	types	missing	cardinality	zero	minus	outlier
variables	types	missing	cardinality	zero	minus	outlier
vat	numeric					X
imat	numeric					X
eat	numeric			X		X
pat	numeric					X
tat	numeric					X
status_changed_to_nt	numeric		< low	X		X
mortality_increase	numeric			X		X
known_mortality	numeric		< low			X

Table 5: List of variables diagnosis (continued)

Missing Values

List of Missing Values

No variables including missing values

Visualization

No variables including missing values

Unique Values

Categorical Vaiables

No variable with a high proportion greater than 0.5

Numerical Vaiables

Variables where the unique cases is less than 5 or unique is 1.

variables	types	unique	unique (%)	status	recommand
status_changed_to_nt	numeric	2	0.4%	low cardinality	Judgment
known_mortality	numeric	5	1%	low cardinality	Judgment

Table 6: Detail warning numerical cardinality

Categorical Variable Diagnosis

Top Ranks

variables	levels	freq	ratio (%)
blood_type	A	202	42.2
blood_type	O	184	38.4
blood_type	B	71	14.8
blood_type	AB	22	4.6
catecholamine_cat	yes	417	87.1
catecholamine_cat	no	62	12.9
child_pugh_score	A	230	48.0
child_pugh_score	B	218	45.5
child_pugh_score	C	31	6.5
death	0	361	75.4
death	1	118	24.6
dialysis_cat	no	409	85.4
dialysis_cat	yes	70	14.6
hcc	no	378	78.9
hcc	yes	101	21.1
hu_listung	no	452	94.4
hu_listung	yes	27	5.6
icu_cat	yes	417	87.1
icu_cat	no	62	12.9
meld_score_category	10 to 19	250	52.2
meld_score_category	20 to 29	90	18.8
meld_score_category	Less than 9	77	16.1
meld_score_category	30 to 39	42	8.8
meld_score_category	Over 40	20	4.2
portal_vein_thrombosis	no	428	89.4

Table 7: Top 10 levels of categorical variables

	variables	levels	freq	ratio (%)
26	portal_vein_thrombosis	partiell	39	8.1
27	portal_vein_thrombosis	yes	12	2.5
28	primary_diagnosis	Hepatocellular carcinoma and cirrhosis	102	21.3
29	primary_diagnosis	Alcoholic cirrhosis	86	18.0
30	primary_diagnosis	Others specify	63	13.2
31	primary_diagnosis	Virus related cirrhosis	57	11.9
32	primary_diagnosis	Primary sclerosing cholangitis	47	9.8
33	primary_diagnosis	NASH	28	5.8
34	primary_diagnosis	Autoimmunhepatitis	22	4.6
35	primary_diagnosis	Primary biliary cirrhosis	19	4.0
36	primary_diagnosis	Polycystic disease	11	2.3
37	primary_diagnosis	Budd Chiari	10	2.1
38	primary_diagnosis	Other levles	34	7.1
39	sex	male	307	64.1
40	sex	female	172	35.9
41	ventilation_cat	yes	414	86.4
42	ventilation_cat	no	65	13.6

Table 7: Top 10 levels of categorical variables (continued)

Numerical Variable Diagnosis

Distributions

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
time	1.00	49.50	440.36	154.00	442.00	5,872.00	0	0	

Table 8: General list of numerical diagnosis

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
71									
age_by_listung	18.00	44.00	50.43	53.00	60.00	69.00	0	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
3									
height	140.00	168.00	173.69	174.00	180.00	203.00	0	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
2									
weight	40.00	65.00	79.23	77.00	91.00	133.00	0	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
1									
bmi	15.00	23.00	26.17	25.00	29.00	44.00	0	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
7									
inr_listing	0.80	1.00	1.30	1.08	1.36	7.00	0	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
47									
sodium_listing	116.00	136.00	137.86	138.00	141.00	154.00	0	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
16									
platelets_listing	7.00	67.00	140.43	108.00	175.50	2,017.00	0	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
25									
exceptional_meld	0.00	0.00	4.77	0.00	0.00	40.00	389	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
90									
bone	1,029.81	2,324.24	2,657.73	2,640.30	3,049.43	4,231.98	0	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
3									
muscle	2,840.94	5,854.26	6,935.54	6,934.09	7,940.71	12,557.64	0	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
6									
sat	728.87	4,072.42	8,280.24	6,654.96	11,574.05	34,258.02	0	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
5									
vat	292.75	1,741.64	3,243.55	2,832.01	4,476.48	10,561.94	0	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
6									
imat	153.48	674.63	1,128.15	1,003.16	1,496.46	3,608.21	0	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
4									
eat	0.00	12.37	24.15	21.44	32.44	99.62	1	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
4									
pat	4.98	29.57	63.61	53.96	86.90	294.67	0	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
12									
tat	1,776.77	7,740.98	12,739.85	10,827.70	17,010.69	46,036.93	0	0	

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
4									
status_changed_to_nt	0.00	0.00	0.12	0.00	0.00	1.00	420	0	59
mortality_increase	0.00	0.04	0.10	0.08	0.12	0.40	15	0	45
known_mortality	0.04	0.20	0.31	0.20	0.46	1.00	0	0	19

Zero Values

variables	min	median	max	zero	zero (%)
status_changed_to_nt	0	0.00	1.00	420	87.7
exceptional_meld	0	0.00	40.00	389	81.2
mortality_increase	0	0.08	0.40	15	3.1
eat	0	21.44	99.62	1	0.2

Table 9: List of numerical diagnosis (zero)

Negative Values

No numeric variable with negative value

Outliers

List of Outliers

variables	min	median	max	outlier	outlier (%)
exceptional_meld	0.00	0.00	40.00	90	18.8
time	1.00	154.00	5,872.00	71	14.8
status_changed_to_nt	0.00	0.00	1.00	59	12.3
inr_listing	0.80	1.08	7.00	47	9.8
mortality_increase	0.00	0.08	0.40	45	9.4
platelets_listing	7.00	108.00	2,017.00	25	5.2
known_mortality	0.04	0.20	1.00	19	4.0
sodium_listing	116.00	138.00	154.00	16	3.3
pat	4.98	53.96	294.67	12	2.5
bmi	15.00	25.00	44.00	7	1.5
muscle	2,840.94	6,934.09	12,557.64	6	1.3
vat	292.75	2,832.01	10,561.94	6	1.3
sat	728.87	6,654.96	34,258.02	5	1.0
imat	153.48	1,003.16	3,608.21	4	0.8
eat	0.00	21.44	99.62	4	0.8
tat	1,776.77	10,827.70	46,036.93	4	0.8
age_by_listung	18.00	53.00	69.00	3	0.6
bone	1,029.81	2,640.30	4,231.98	3	0.6
height	140.00	174.00	203.00	2	0.4
weight	40.00	77.00	133.00	1	0.2

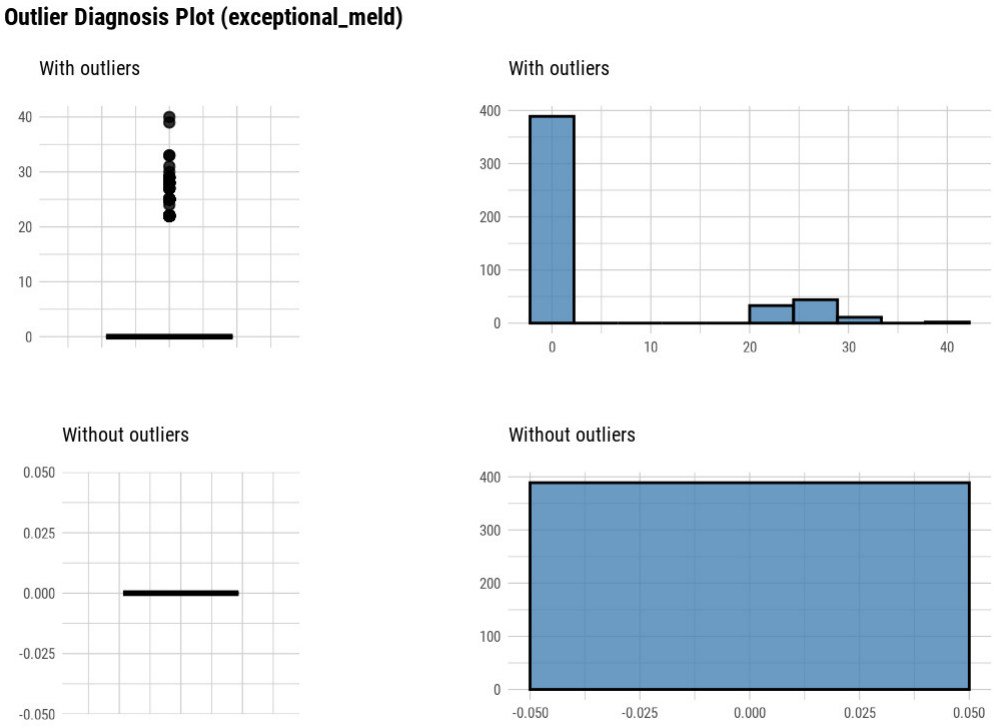
Table 10: Diagnosis of numerical variable outliers

Individual Outliers

variable: exceptional_meld

Measures	Values
Outliers count	90
Outliers ratio (%)	18.79%
Mean of outliers	25.38889
Mean with outliers	4.770355
Mean without outliers	0

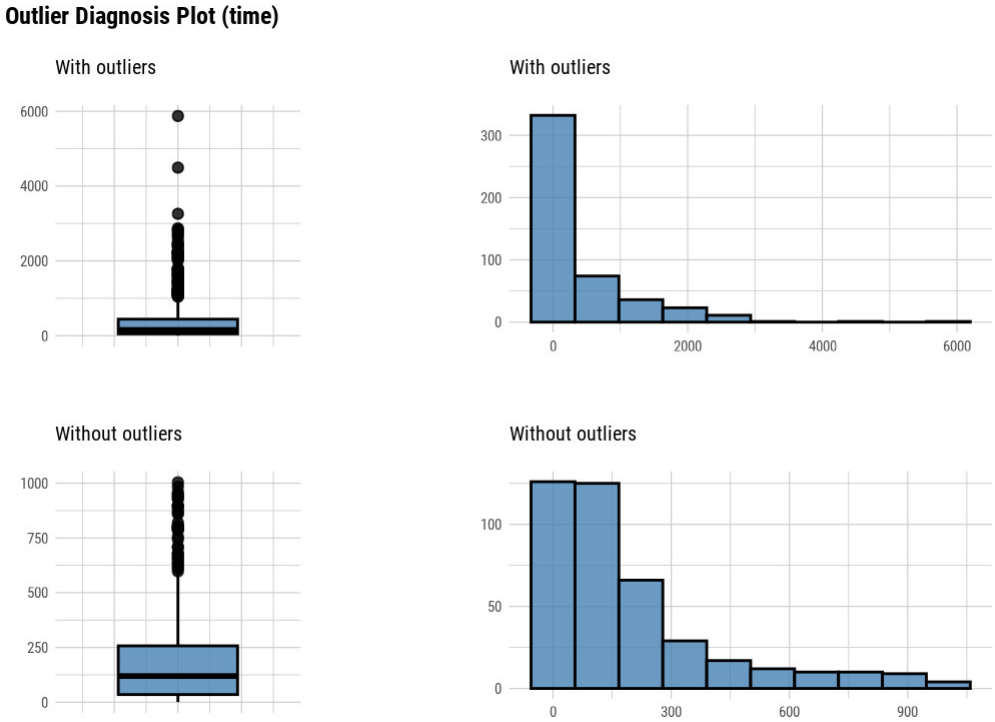
Table 11: exceptional_meld



variable: time

Measures	Values
Outliers count	71
Outliers ratio (%)	14.82%
Mean of outliers	1839.324
Mean with outliers	440.3591
Mean without outliers	196.9118

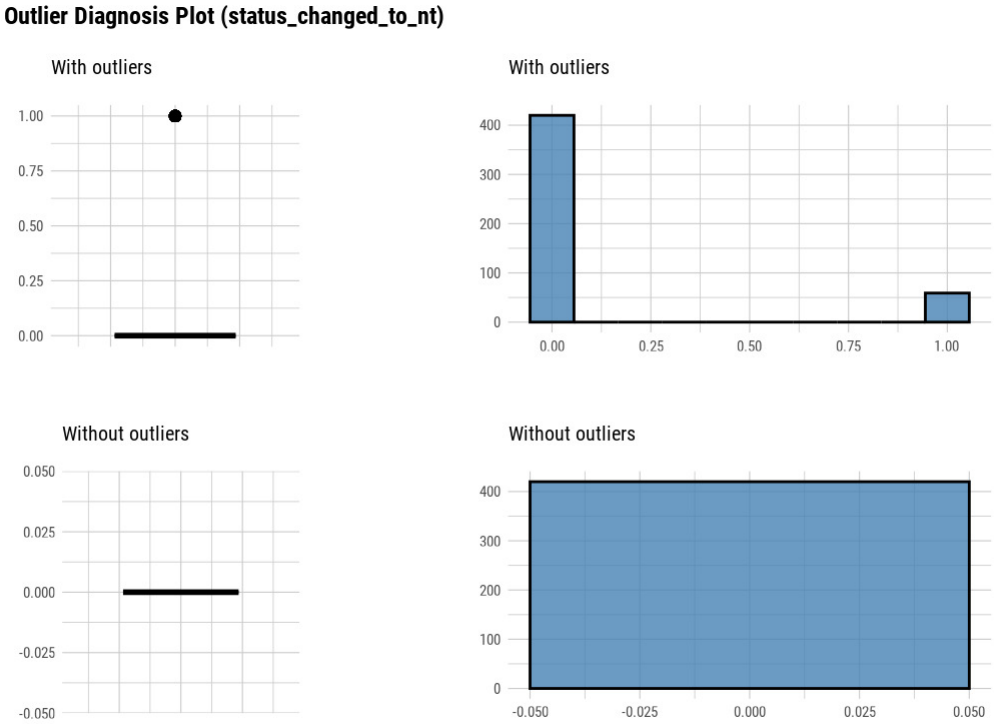
Table 11: time



variable: status_changed_to_nt

Measures	Values
Outliers count	59
Outliers ratio (%)	12.32%
Mean of outliers	1
Mean with outliers	0.1231733
Mean without outliers	0

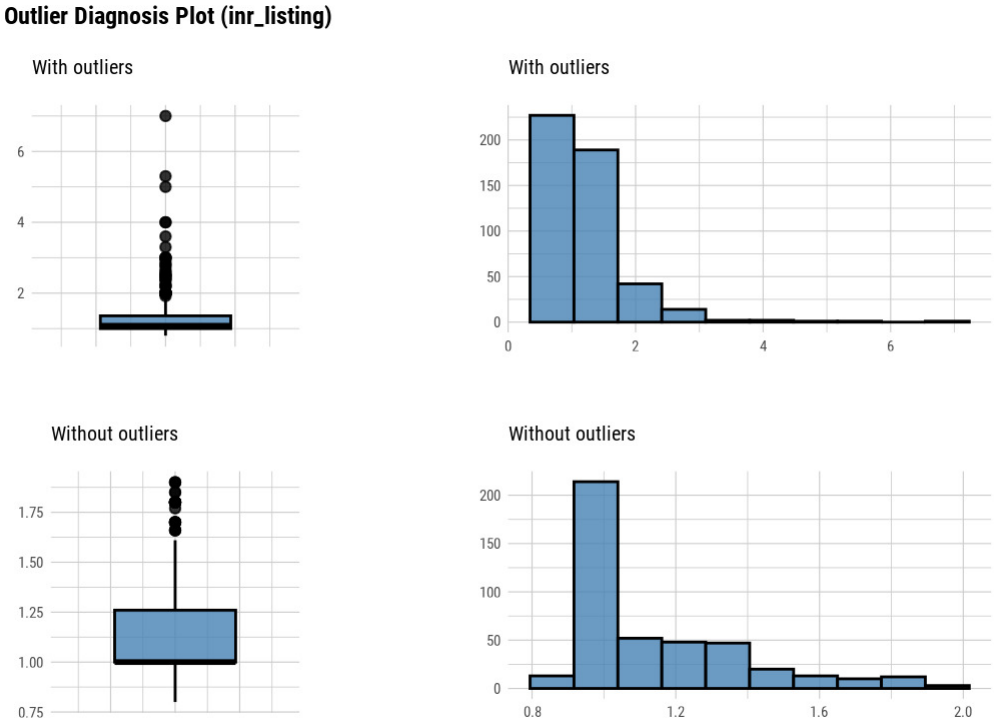
Table 11: status_changed_to_nt



variable: inr_listing

Measures	Values
Outliers count	47
Outliers ratio (%)	9.81%
Mean of outliers	2.651277
Mean with outliers	1.299562
Mean without outliers	1.1525

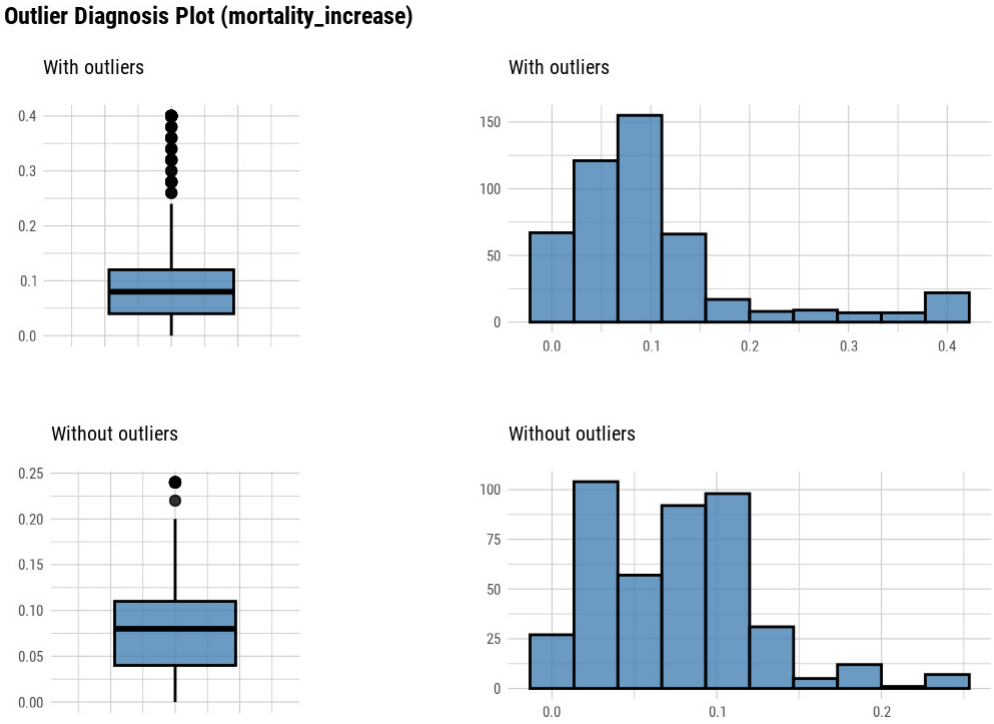
Table 11: inr_listing



variable: mortality_increase

Measures	Values
Outliers count	45
Outliers ratio (%)	9.39%
Mean of outliers	0.3528889
Mean with outliers	0.1038413
Mean without outliers	0.07801843

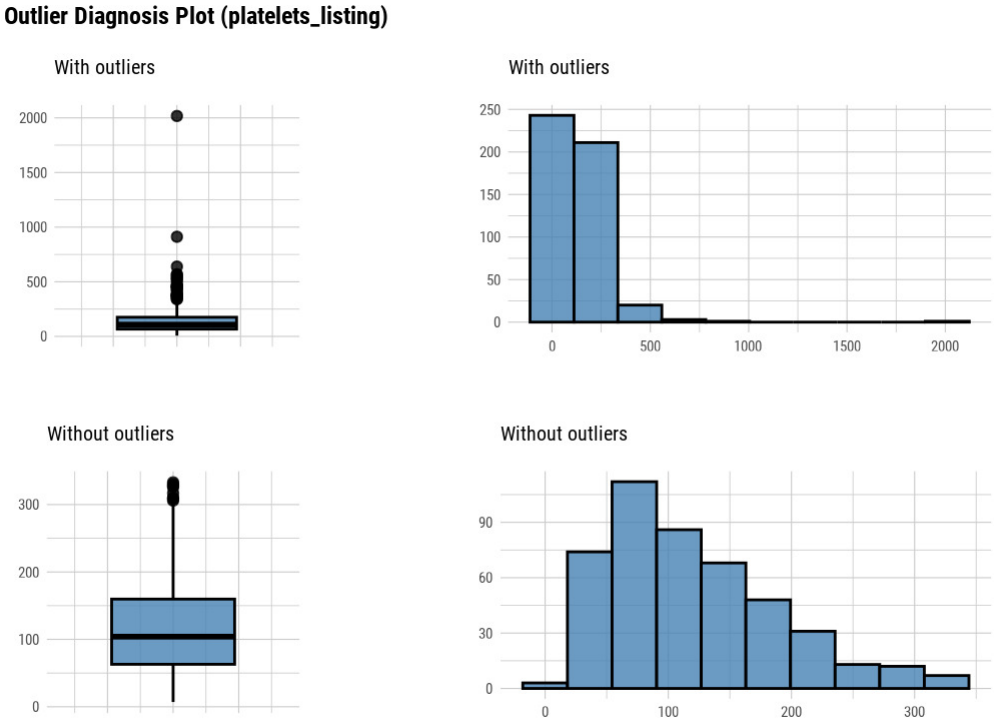
Table 11: mortality_increase



variable: platelets_listing

Measures	Values
Outliers count	25
Outliers ratio (%)	5.22%
Mean of outliers	528.52
Mean with outliers	140.4301
Mean without outliers	119.0595

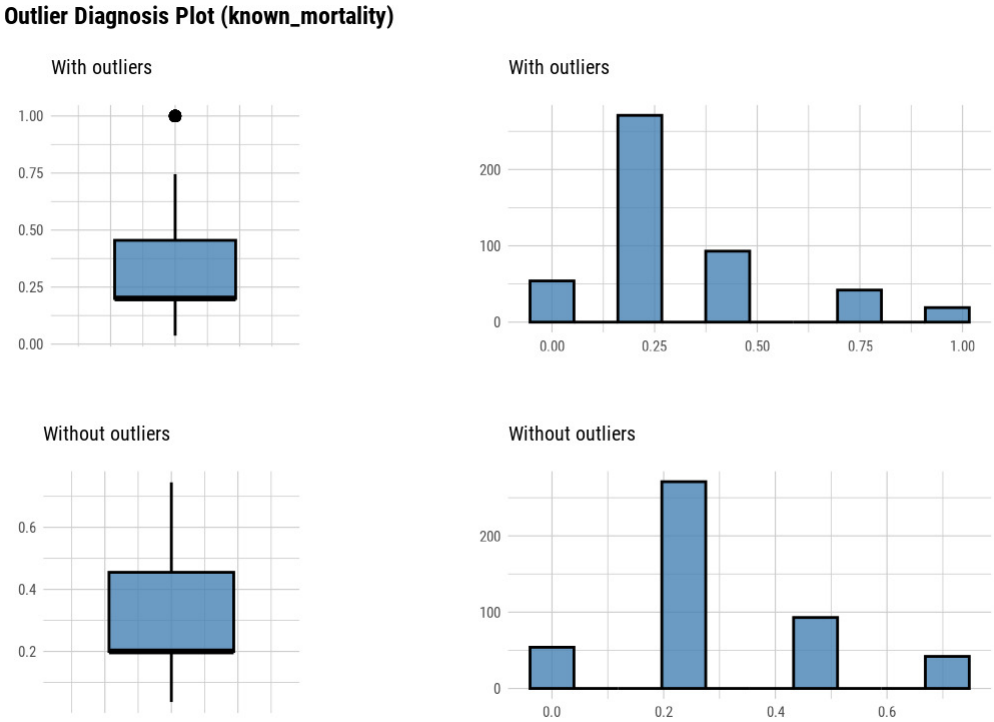
Table 11: platelets_listing



variable: known_mortality

Measures	Values
Outliers count	19
Outliers ratio (%)	3.97%
Mean of outliers	1
Mean with outliers	0.3106534
Mean without outliers	0.2821804

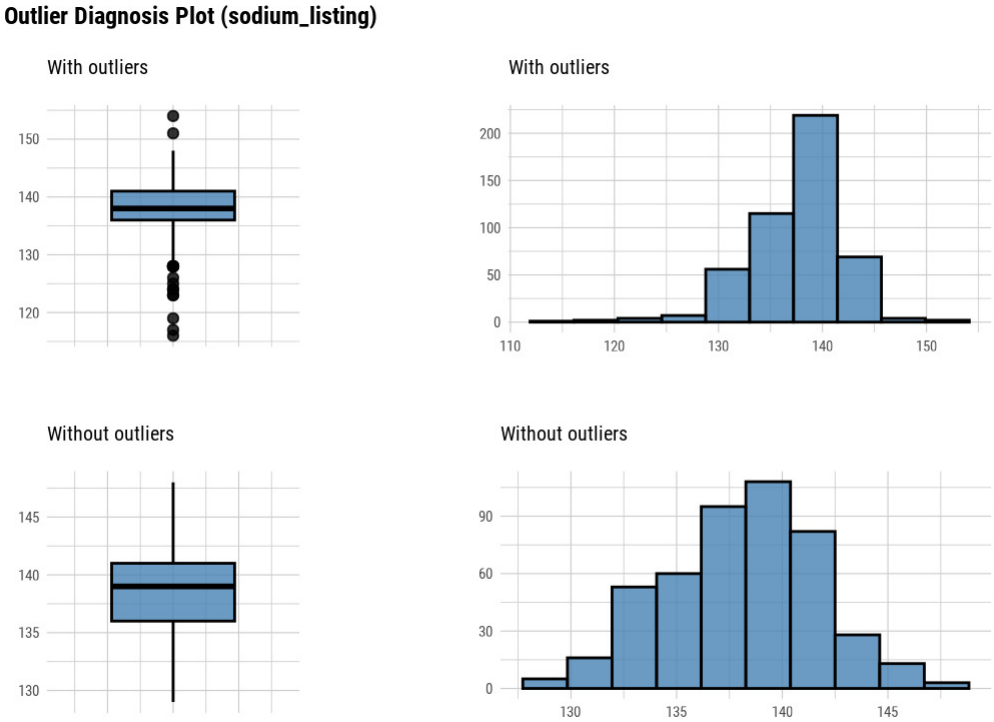
Table 11: known_mortality



variable: sodium_listing

Measures	Values
Outliers count	16
Outliers ratio (%)	3.34%
Mean of outliers	127.625
Mean with outliers	137.8643
Mean without outliers	138.2181

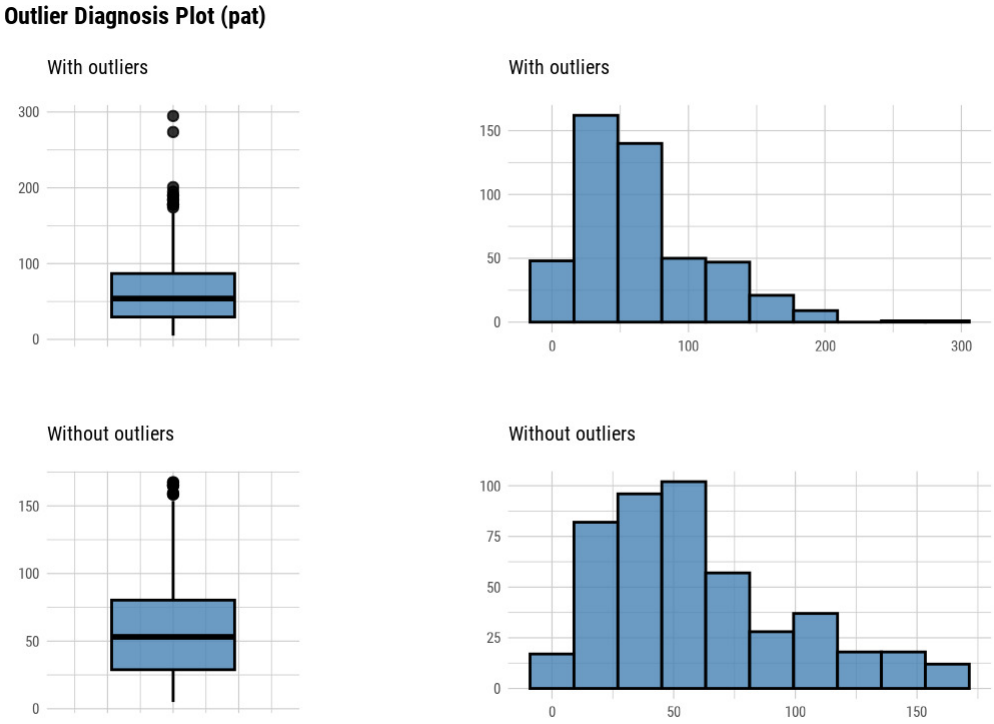
Table 11: sodium_listing



variable: pat

Measures	Values
Outliers count	12
Outliers ratio (%)	2.51%
Mean of outliers	201.6687
Mean with outliers	63.60991
Mean without outliers	60.06236

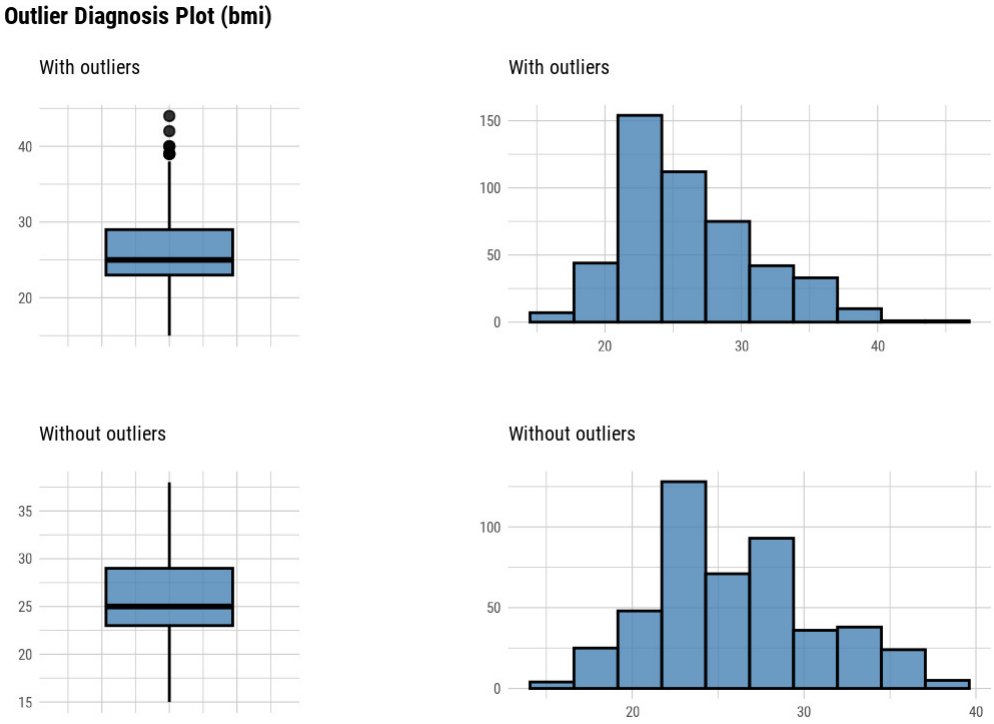
Table 11: pat



variable: bmi

Measures	Values
Outliers count	7
Outliers ratio (%)	1.46%
Mean of outliers	40.42857
Mean with outliers	26.17119
Mean without outliers	25.95975

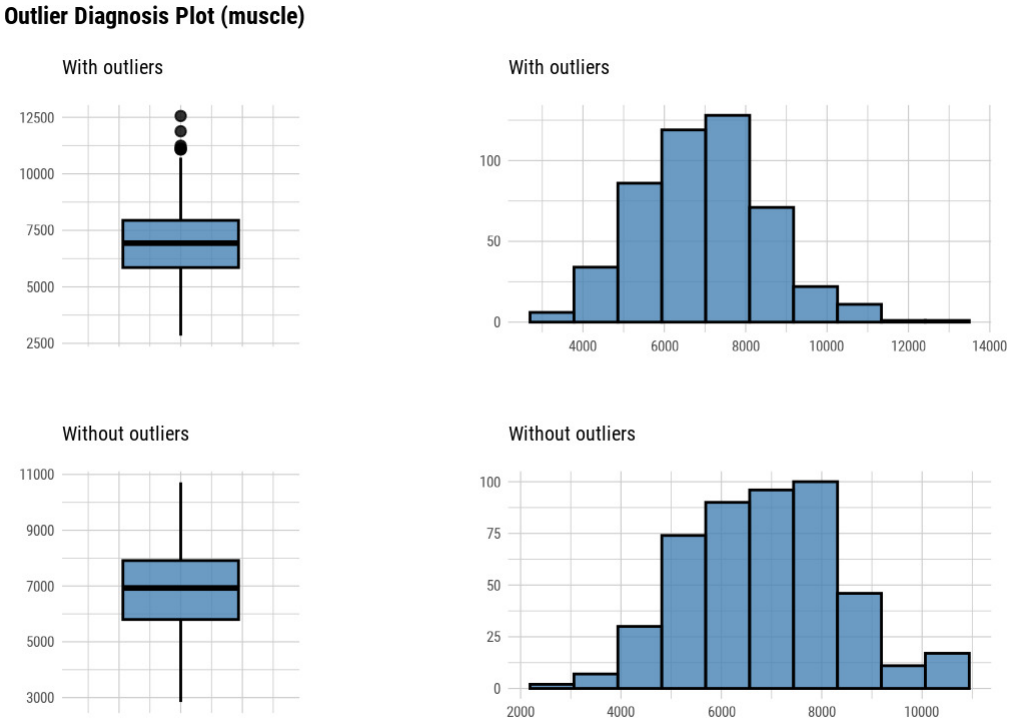
Table 11: bmi



variable: muscle

Measures	Values
Outliers count	6
Outliers ratio (%)	1.25%
Mean of outliers	11496.11
Mean with outliers	6935.539
Mean without outliers	6877.688

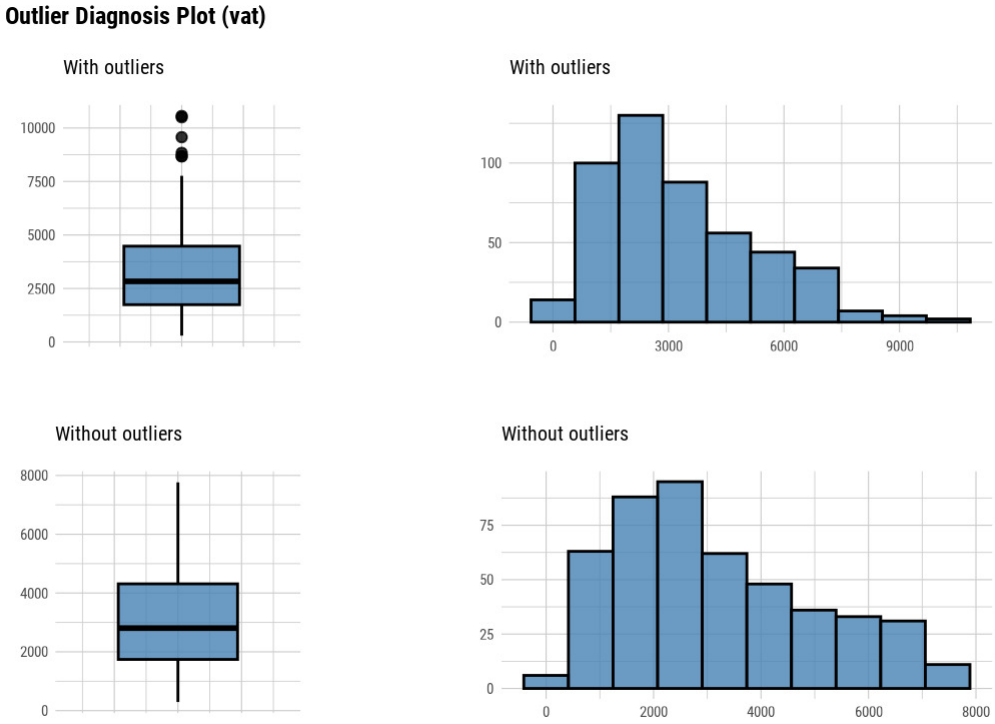
Table 11: muscle



variable: vat

Measures	Values
Outliers count	6
Outliers ratio (%)	1.25%
Mean of outliers	9471.641
Mean with outliers	3243.552
Mean without outliers	3164.549

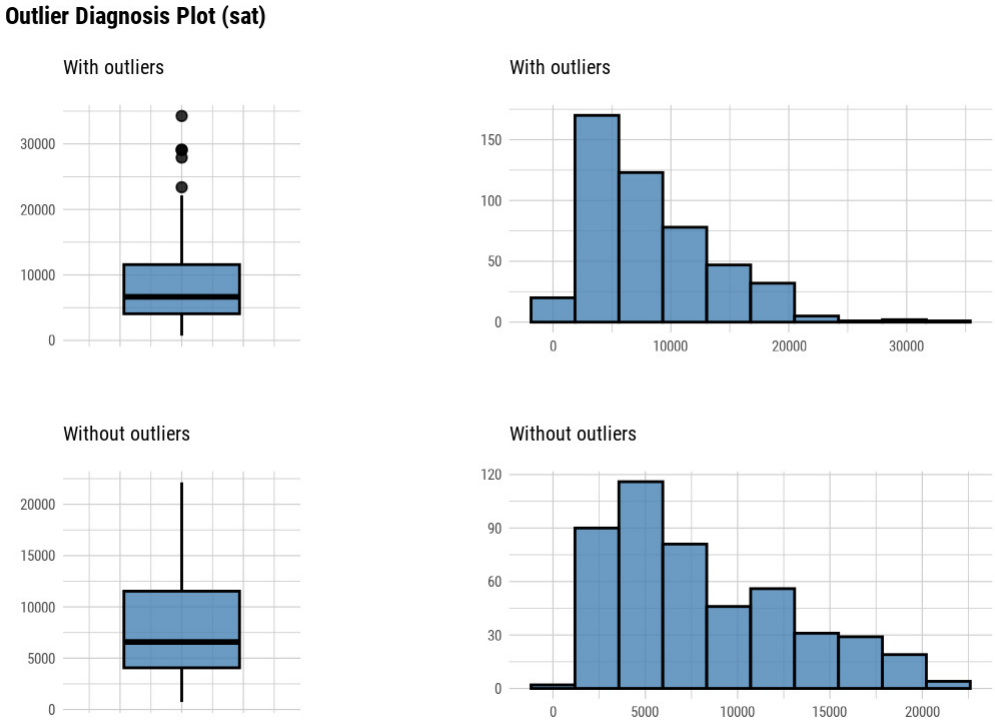
Table 11: vat



variable: sat

Measures	Values
Outliers count	5
Outliers ratio (%)	1.04%
Mean of outliers	28745.95
Mean with outliers	8280.243
Mean without outliers	8064.36

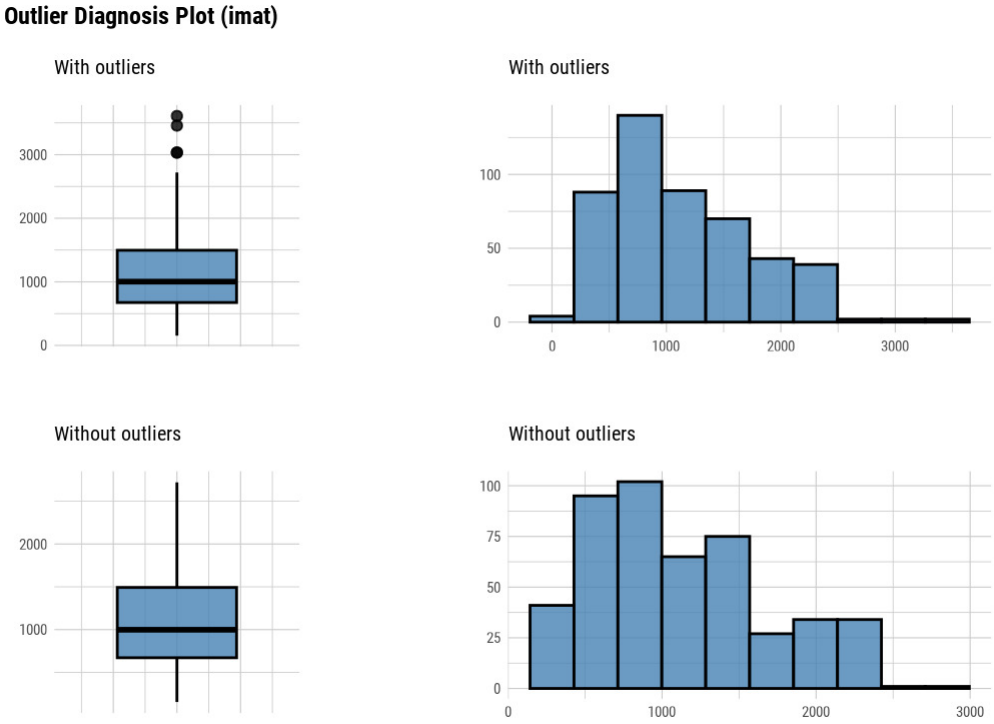
Table 11: sat



variable: imat

Measures	Values
Outliers count	4
Outliers ratio (%)	0.84%
Mean of outliers	3283.13
Mean with outliers	1128.154
Mean without outliers	1110.007

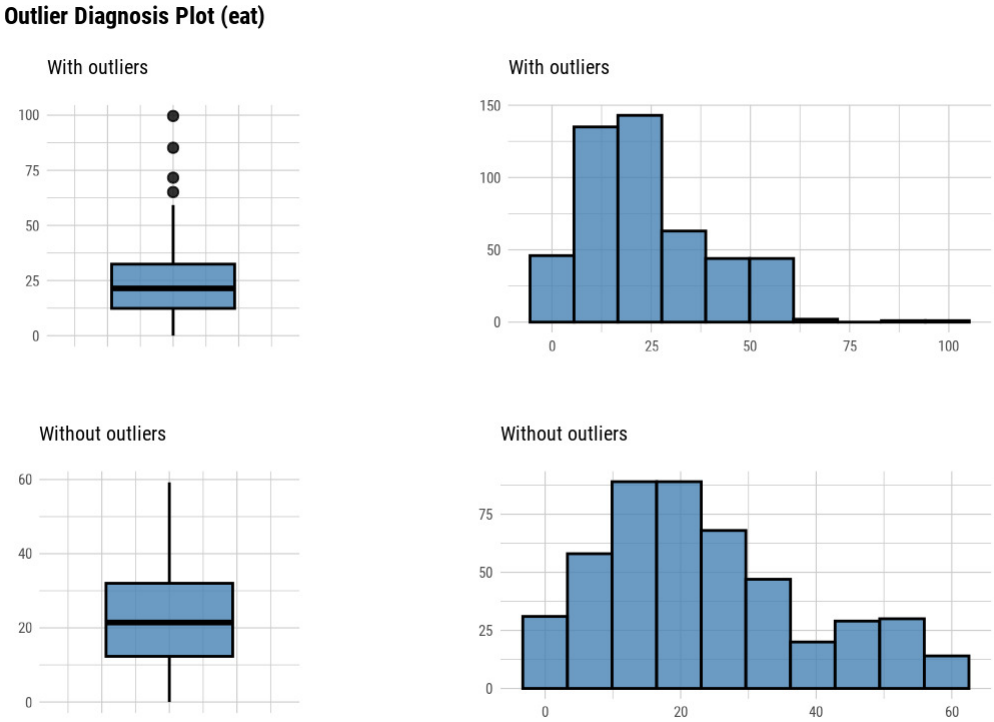
Table 11: imat



variable: eat

Measures	Values
Outliers count	4
Outliers ratio (%)	0.84%
Mean of outliers	80.40111
Mean with outliers	24.15385
Mean without outliers	23.68019

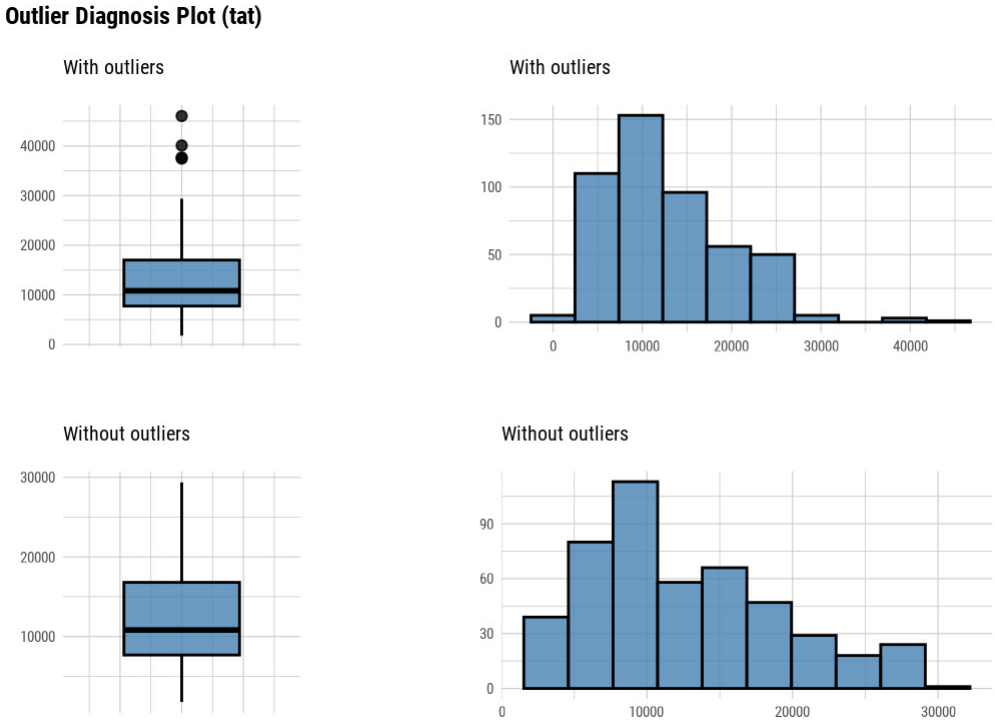
Table 11: eat



variable: tat

Measures	Values
Outliers count	4
Outliers ratio (%)	0.84%
Mean of outliers	40314.37
Mean with outliers	12739.85
Mean without outliers	12507.65

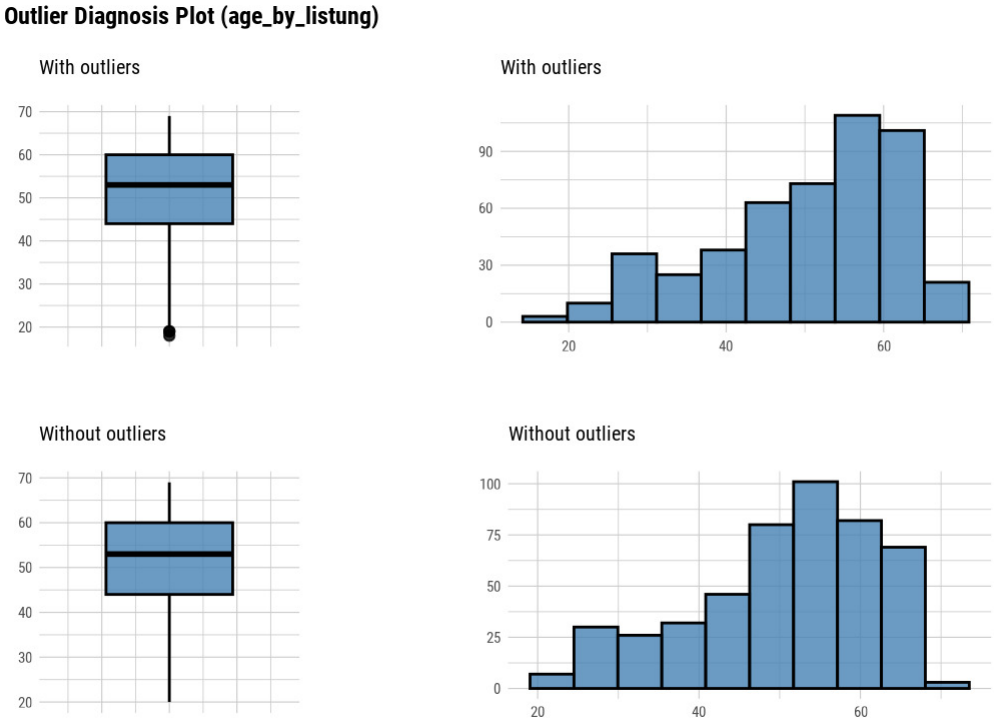
Table 11: tat



variable: age_by_listung

Measures	Values
Outliers count	3
Outliers ratio (%)	0.63%
Mean of outliers	18.66667
Mean with outliers	50.42797
Mean without outliers	50.62815

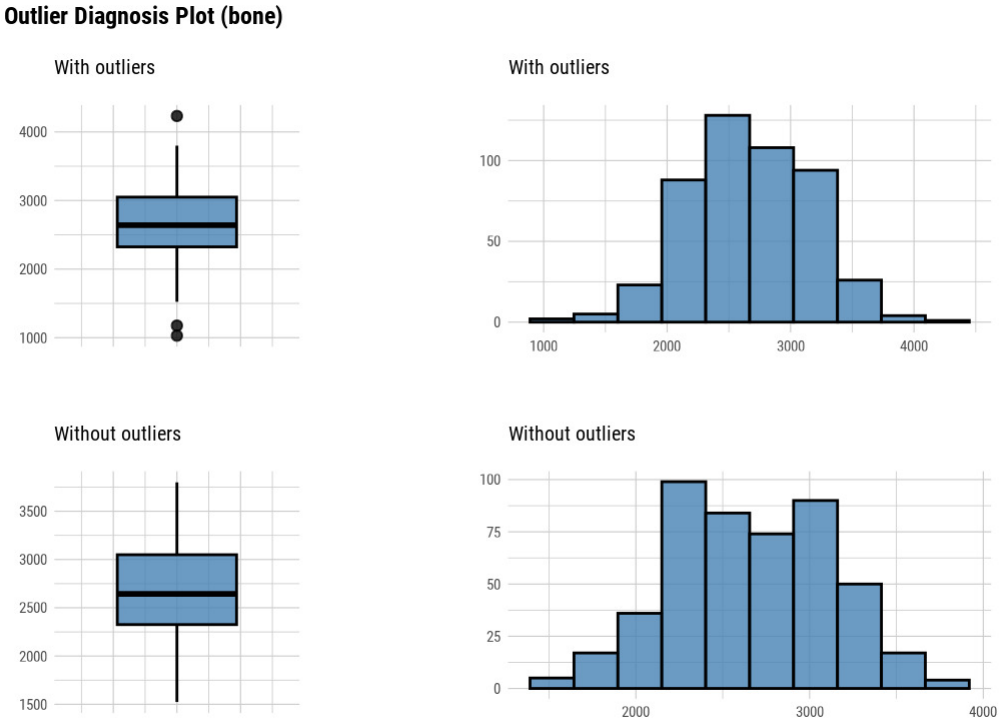
Table 11: age_by_listung



variable: bone

Measures	Values
Outliers count	3
Outliers ratio (%)	0.63%
Mean of outliers	2146.062
Mean with outliers	2657.726
Mean without outliers	2660.95

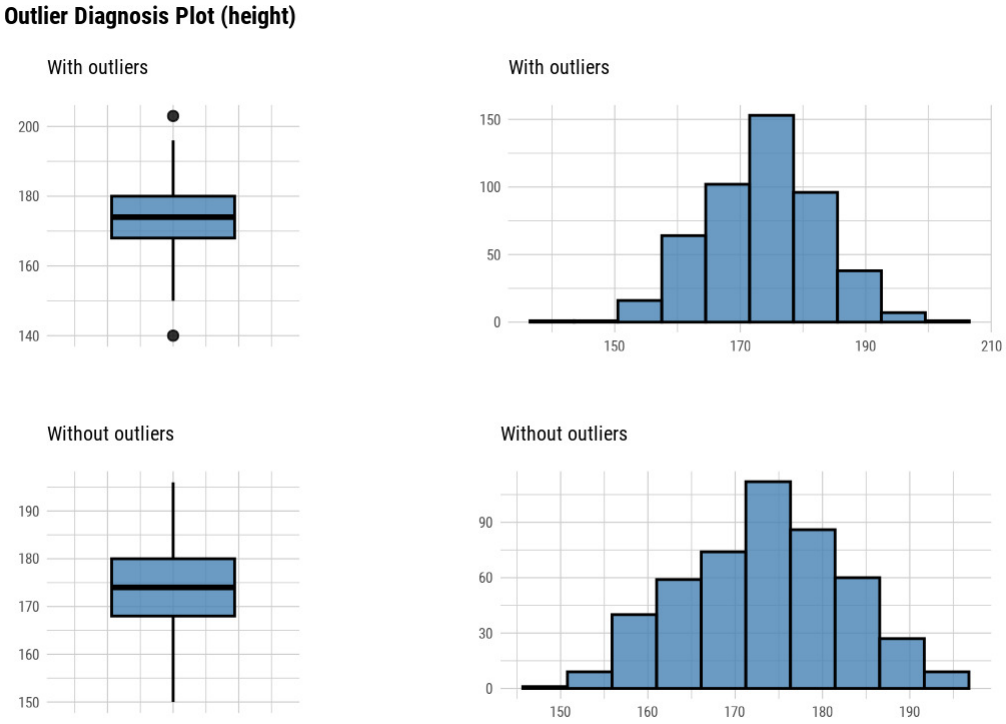
Table 11: bone



variable: height

Measures	Values
Outliers count	2
Outliers ratio (%)	0.42%
Mean of outliers	171.5
Mean with outliers	173.691
Mean without outliers	173.7002

Table 11: height



variable: weight

Measures	Values
Outliers count	1
Outliers ratio (%)	0.21%
Mean of outliers	133
Mean with outliers	79.22965
Mean without outliers	79.11715

Table 11: weight