



EDA REPORT



Report Overview

This report was created for the EDA of . data. It helps explore data to understand the data and find scenarios for performing the analysis.

Contents

Overview	2
Data Structures	2
Job Informations	2
Univariate Analysis	3
Descriptive Statistics	3
Numerical Variables	3
Categorical Variables	6
Normality Test	9
Bivariate Analysis	17
Compare Numerical Variables	17
Compare Categorical Variables	39
Multivariate Analysis	55
Correlation Analysis	55
Correlation Coefficient Matrix	55
Correlation Plot	56
Target based Analysis	57
Grouped Numerical Variables	57
Grouped Categorical Variables	65
Grouped Correlation	70

Overview

Data Structures

division	metrics	value	division	metrics	value
size	observations	299	data type	numerics	5
size	variables	13	data type	integers	2
size	values	3,887	data type	factors/ordered	6
size	memory size (KB)	0	data type	characters	0
duplicated	duplicate observation	0	data type	Dates	0
missing	complete observation	270	data type	POSIXcts	0
missing	missing observation	29	data type	others	0
missing	missing variables	2			
missing	missing values	30			

Table 1: Data structures and types

Job Informations

division	metrics	value
dataset	dataset	.
dataset	dataset type	data.frame
dataset	target	death_event
job	samples	299 / 299 (100%)
job	created	2021-08-22 05:24:53
job	created by	dlookr

Table 2: Job informations

Univariate Analysis

Descriptive Statistics

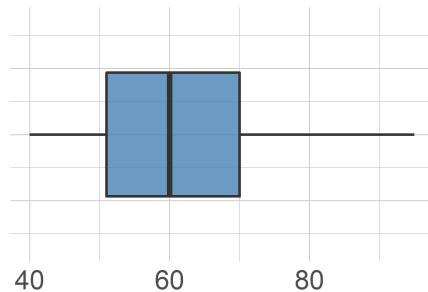
Numerical Variables

variables	missing	mean	sd	min	Q1	median	Q3	max
age	20	60.64	11.87	40.0	51.0	60.0	70.0	95.0
cpk_enzyme	0	581.84	970.29	23.0	116.5	250.0	582.0	7,861.0
ejection_fraction	0	38.08	11.83	14.0	30.0	38.0	45.0	80.0
platelets	0	263,358.03	97,804.24	25,100.0	212,500.0	262,000.0	303,500.0	850,000.0
creatinine	0	1.39	1.03	0.5	0.9	1.1	1.4	9.4
sodium	0	136.63	4.41	113.0	134.0	137.0	140.0	148.0
time	0	130.26	77.61	4.0	73.0	115.0	203.0	285.0

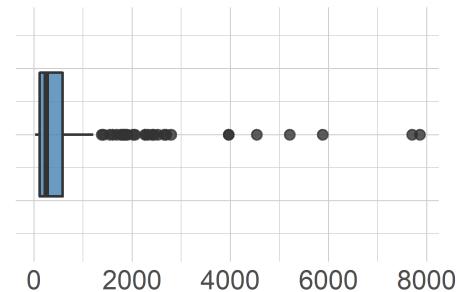
Table 3: Descriptive statistics of numerical variables

Distribution by numerical variables

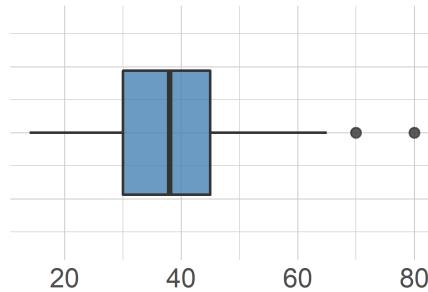
age



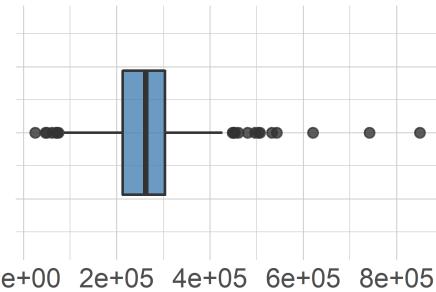
cpk_enzyme



ejection_fraction

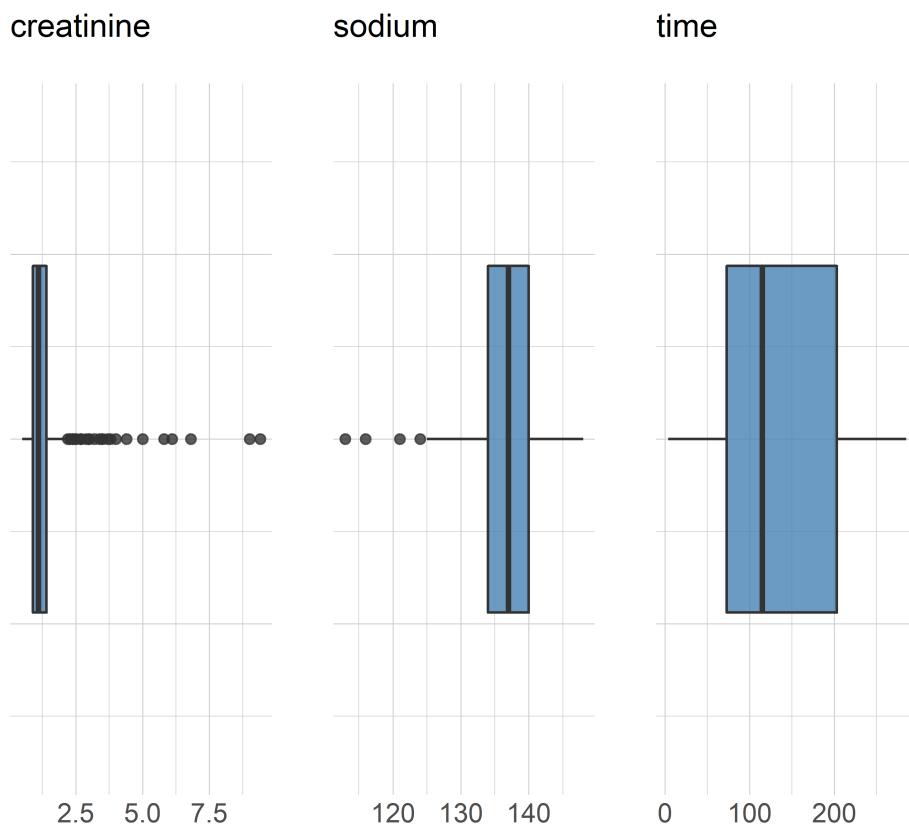


platelets



variables	data types	distinct	skewness	kurtosis	zero	negative	outlier
age	integer	47	0.44	-0.10	0	0	0
cpk_enzyme	numeric	208	4.46	25.15	0	0	29
ejection_fraction	numeric	17	0.56	0.04	0	0	2
platelets	numeric	176	1.46	6.21	0	0	21

Distribution by numerical variables



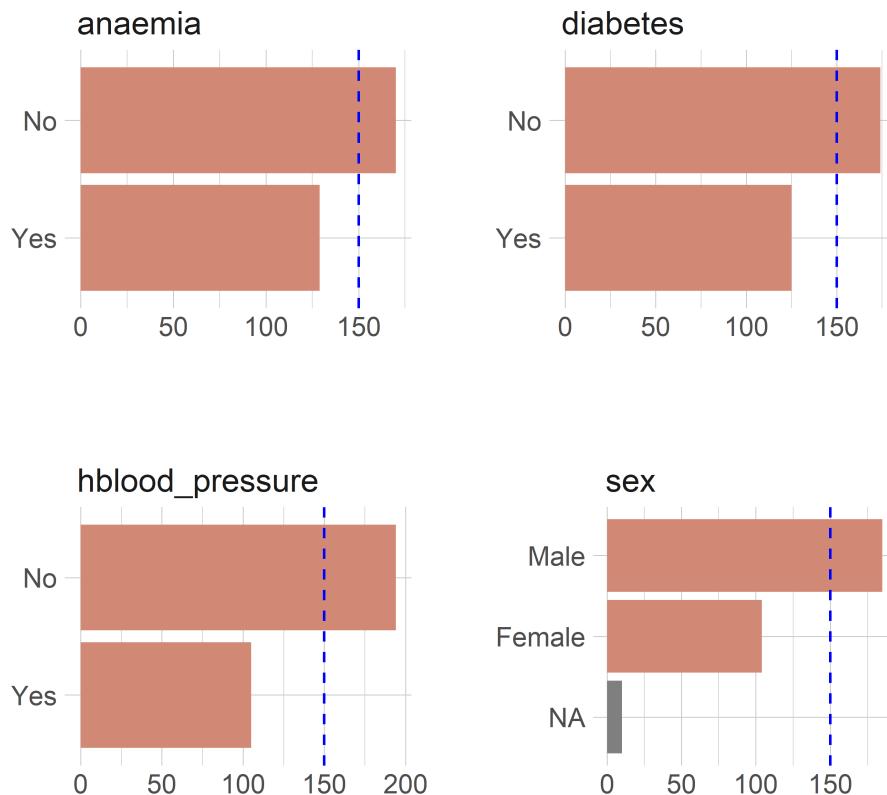
variables	data types	distinct	skewness	kurtosis	zero	negative	outlier
creatinine	numeric	40	4.46	25.83	0	0	29
sodium	numeric	27	-1.05	4.12	0	0	4
time	integer	148	0.13	-1.21	0	0	0

Categorical Variables

variables	levels	observations	frequency	frequency(%)	rank
anaemia	No	299	170	56.86	1
anaemia	Yes	299	129	43.14	2
diabetes	No	299	174	58.19	1
diabetes	Yes	299	125	41.81	2
hblood_pressure	No	299	194	64.88	1
hblood_pressure	Yes	299	105	35.12	2
sex	Male	299	185	61.87	1
sex	Female	299	104	34.78	2
sex	NA	299	10	3.34	3
smoking	No	299	203	67.89	1
smoking	Yes	299	96	32.11	2
death_event	No	299	203	67.89	1
death_event	Yes	299	96	32.11	2

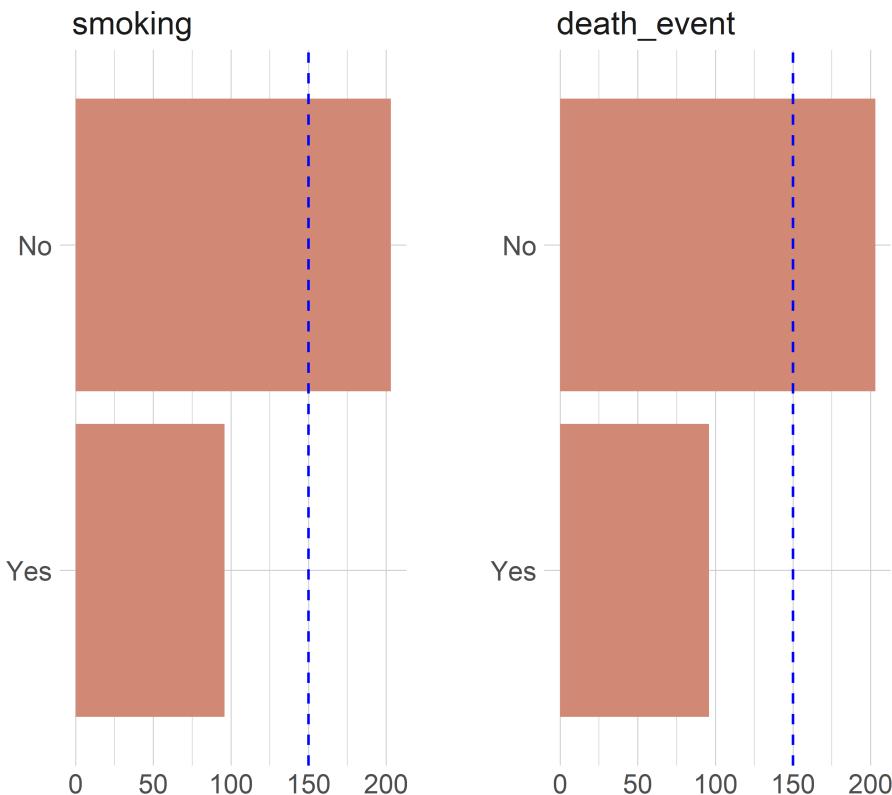
Table 4: Top rank levels of categorical variables

Frequency by levels of category



variables	missing	missing(%)	distinct	distinct ratio
anaemia	0	0.00	2	0.007
diabetes	0	0.00	2	0.007
hblood_pressure	0	0.00	2	0.007
sex	10	3.34	3	0.010

Frequency by levels of category



variables	missing	missing(%)	distinct	distinct ratio
smoking	0	0	2	0.007
death_event	0	0	2	0.007

Normality Test

variable	min	Q1	median	Q3	max	skewness	kurtosis	balance
age	40.0	51.0	60.0	70.0	95.0	0.4	-0.1	Balanced
cpk_enzyme	23.0	116.5	250.0	582.0	7861.0	4.5	25.1	Right-Skewed
ejection_fraction	14.0	30.0	38.0	45.0	80.0	0.6	0.0	Balanced
platelets	25100.0	212500.0	262000.0	303500.0	850000.0	1.5	6.2	Right-Skewed
creatinine	0.5	0.9	1.1	1.4	9.4	4.5	25.8	Right-Skewed
sodium	113.0	134.0	137.0	140.0	148.0	-1.0	4.1	Balanced
time	4.0	73.0	115.0	203.0	285.0	0.1	-1.2	Balanced

Table 5: Descriptive statistics of numerical variables

age

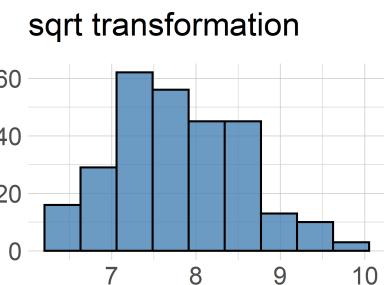
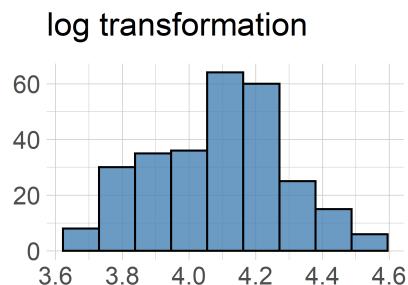
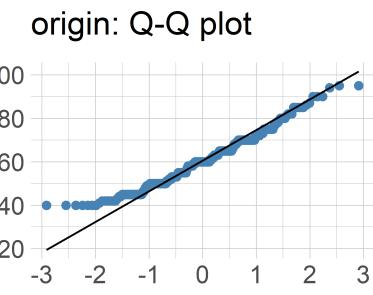
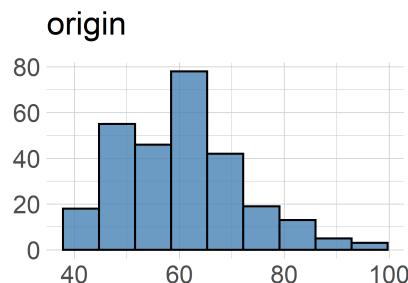
statistic	p_value	remark
0.97466	7.4131e-05	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	0.4415	2.8761
log transformation	-0.0085	2.4761
sqrt transformation	0.2118	2.6089

Table 6: skewness and kurtosis

Normality Diagnosis Plot (x)



cpk_enzyme

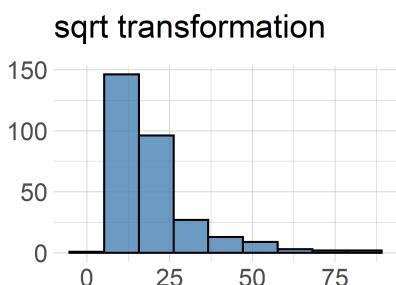
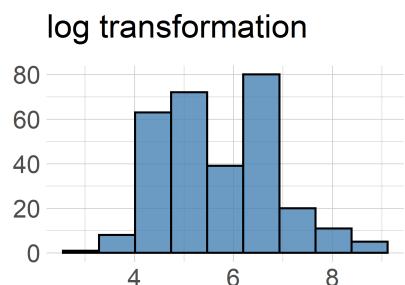
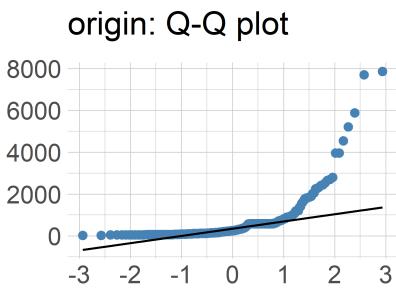
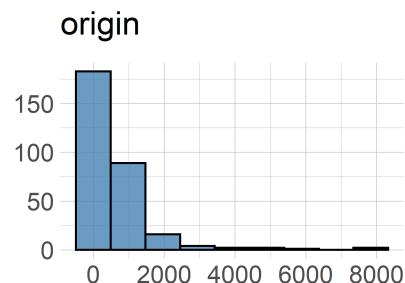
statistic	p_value	remark
0.51426	7.0505e-28	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	4.4407	27.7105
log transformation	0.4119	2.6535
sqrt transformation	2.0967	8.9687

Table 6: skewness and kurtosis

Normality Diagnosis Plot (x)



ejection_fraction

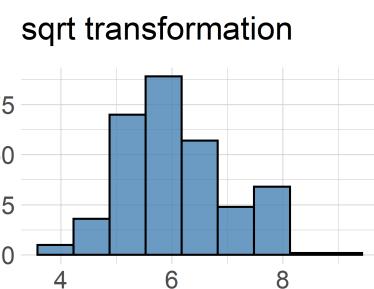
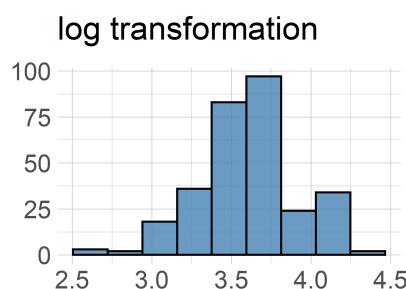
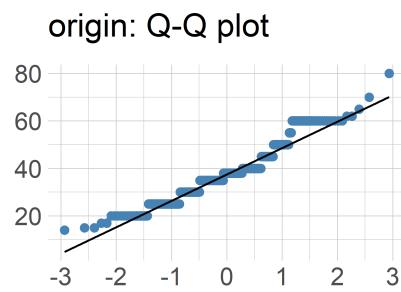
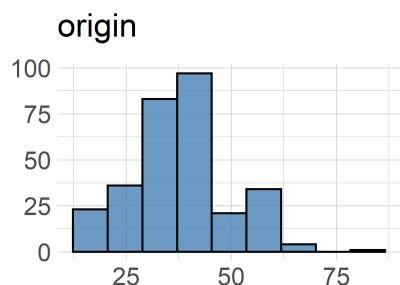
statistic	p_value	remark
0.94732	7.216e-09	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	0.5526	3.0207
log transformation	-0.2707	2.9275
sqrt transformation	0.1553	2.7643

Table 6: skewness and kurtosis

Normality Diagnosis Plot (x)



platelets

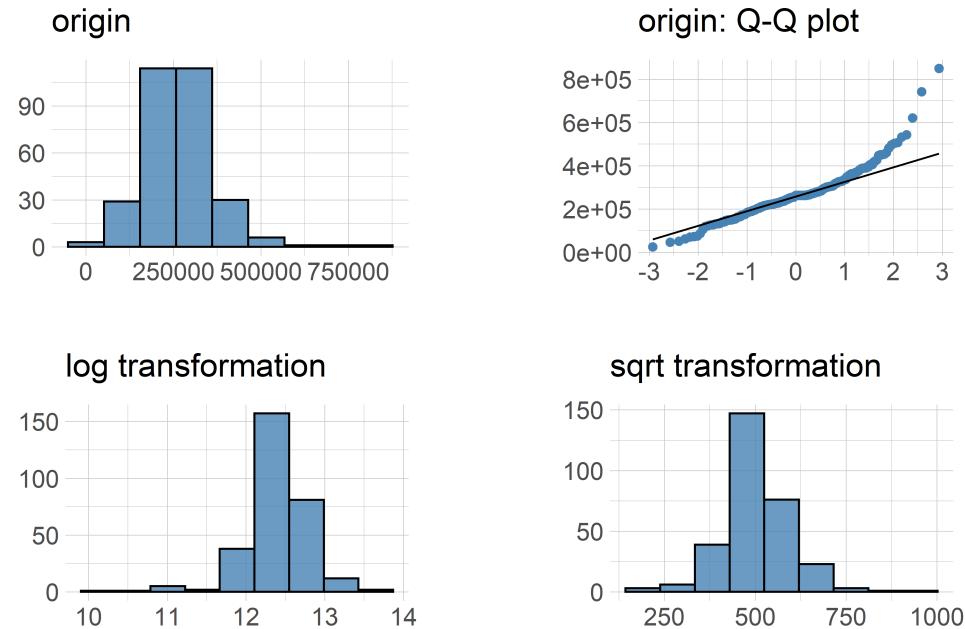
statistic	p_value	remark
0.91151	2.8835e-12	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	1.4550	9.0859
log transformation	-1.2785	8.3256
sqrt transformation	0.1778	5.6143

Table 6: skewness and kurtosis

Normality Diagnosis Plot (x)



creatinine

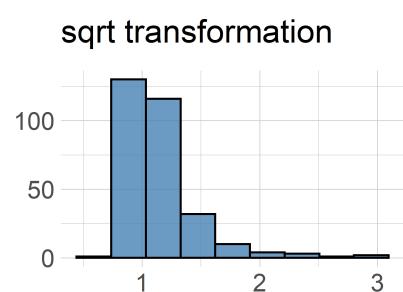
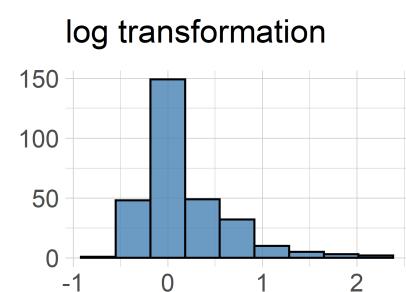
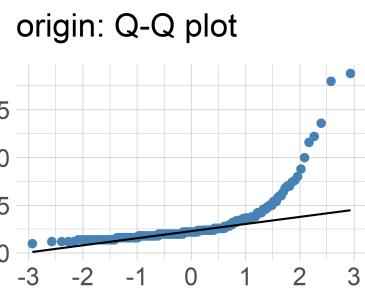
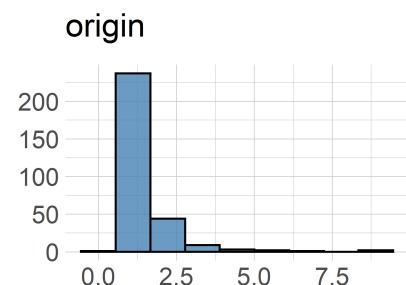
statistic	p_value	remark
0.55147	5.3928e-27	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	4.4336	28.3783
log transformation	1.5760	6.5217
sqrt transformation	2.8069	13.8042

Table 6: skewness and kurtosis

Normality Diagnosis Plot (x)



sodium

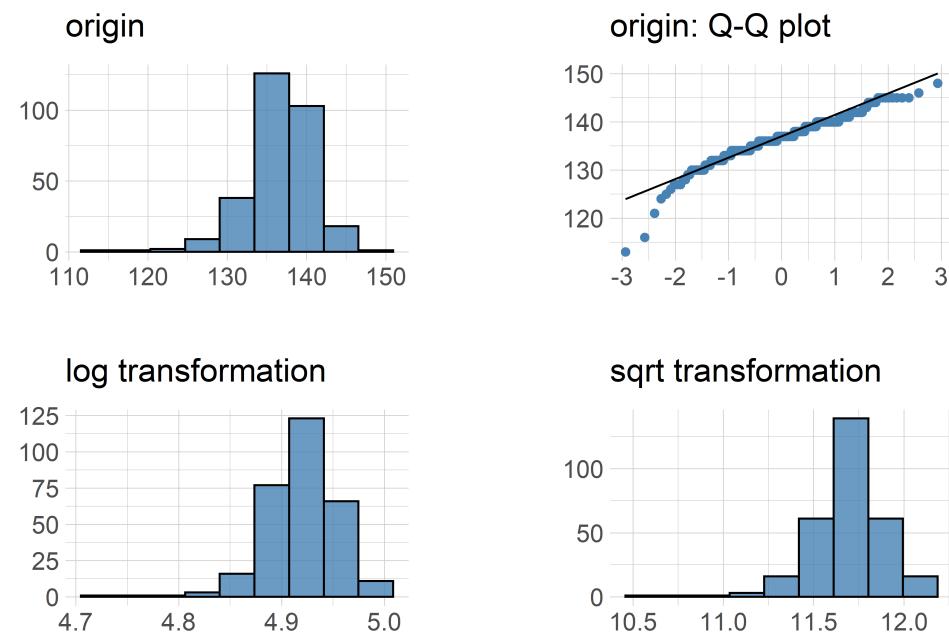
statistic	p_value	remark
0.93903	9.2149e-10	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	-1.0429	7.0311
log transformation	-1.3057	8.3312
sqrt transformation	-1.1706	7.6383

Table 6: skewness and kurtosis

Normality Diagnosis Plot (x)



time

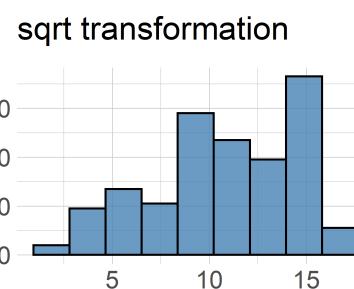
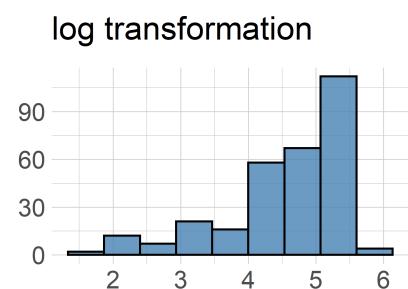
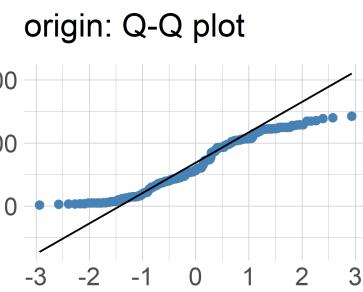
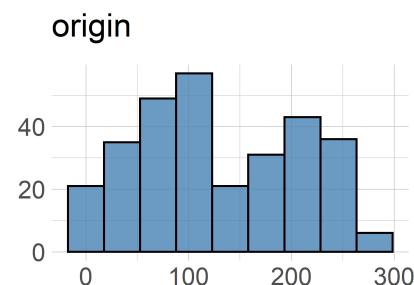
statistic	p_value	remark
0.94678	6.285e-09	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	0.1272	1.7881
log transformation	-1.2253	3.9975
sqrt transformation	-0.3987	2.1741

Table 6: skewness and kurtosis

Normality Diagnosis Plot (x)



Bivariate Analysis

Compare Numerical Variables

first variable	second variable	correlation coefficient
age	cpk_enzyme	-0.07563
age	ejection_fraction	0.06082
age	platelets	-0.01579
age	creatinine	0.16297
age	sodium	-0.05258
age	time	-0.21486
cpk_enzyme	ejection_fraction	-0.04408
cpk_enzyme	platelets	0.02446
cpk_enzyme	creatinine	-0.01641
cpk_enzyme	sodium	0.05955
cpk_enzyme	time	-0.00935
ejection_fraction	platelets	0.07218
ejection_fraction	creatinine	-0.01130
ejection_fraction	sodium	0.17590
ejection_fraction	time	0.04173
platelets	creatinine	-0.04120
platelets	sodium	0.06212
platelets	time	0.01051
creatinine	sodium	-0.18910
creatinine	time	-0.14932
sodium	time	0.08764

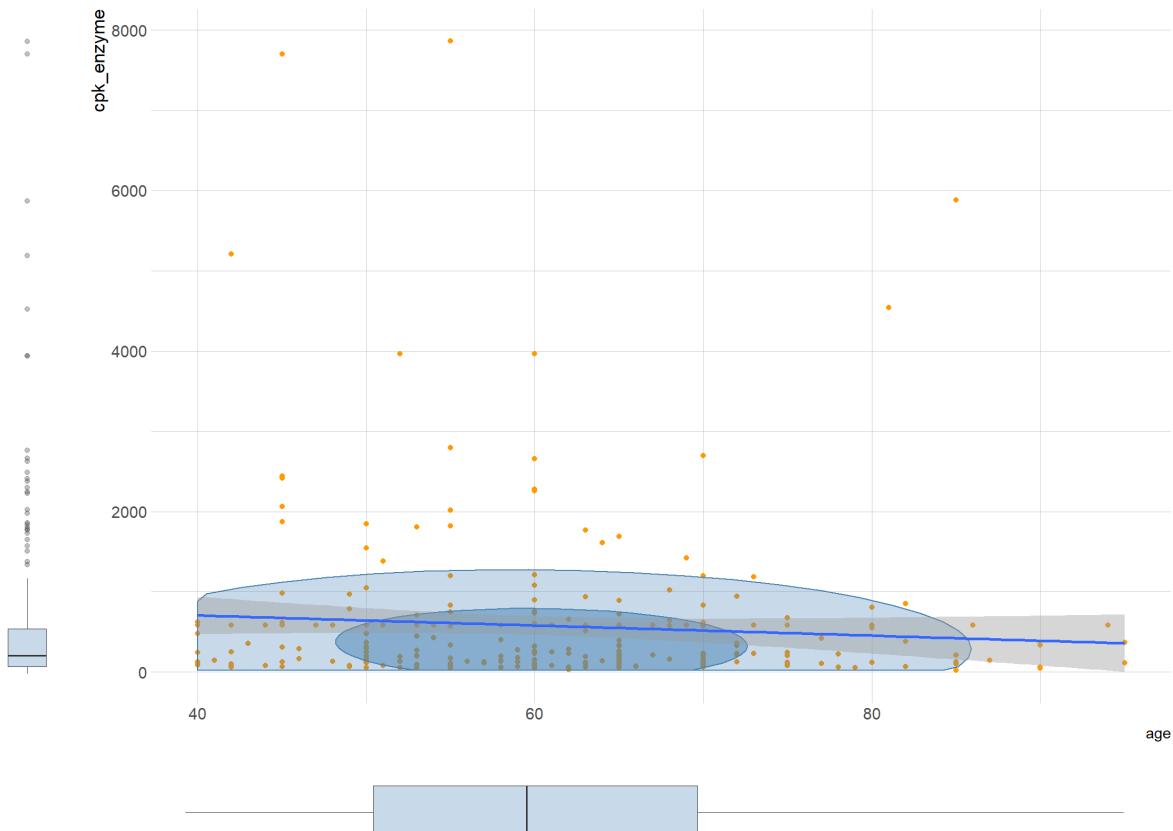
Table 7: Correlation coefficient

'age' vs 'cpk_enzyme'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
age	cpk_enzyme	0.0057197	0.0021303	11.85994	1.593475	0.2078921	1

Table 7: Summary of linear model

Scatterplots with age and cpk_enzyme

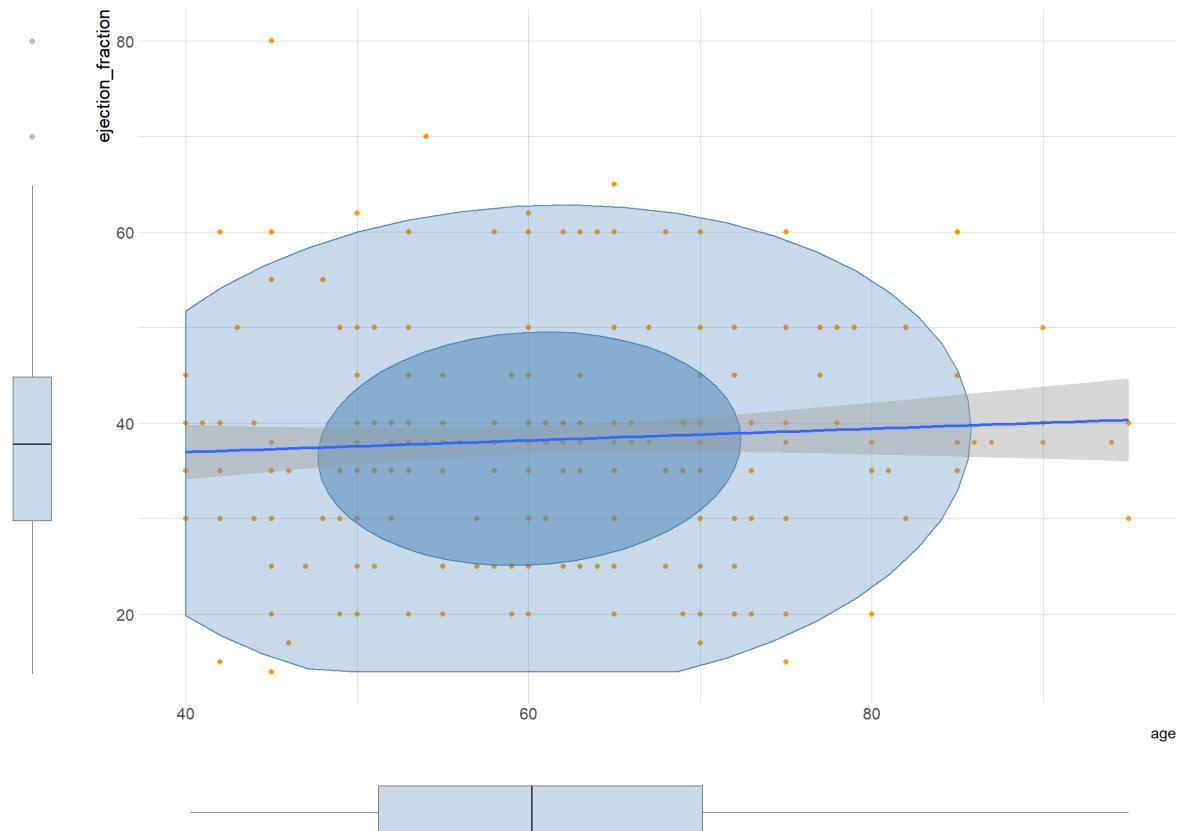


'age' vs 'ejection_fraction'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
age	ejection_fraction	0.0036992	0.0001024	11.87198	1.028474	0.3114024	1

Table 7: Summary of linear model

Scatterplots with age and ejection_fraction

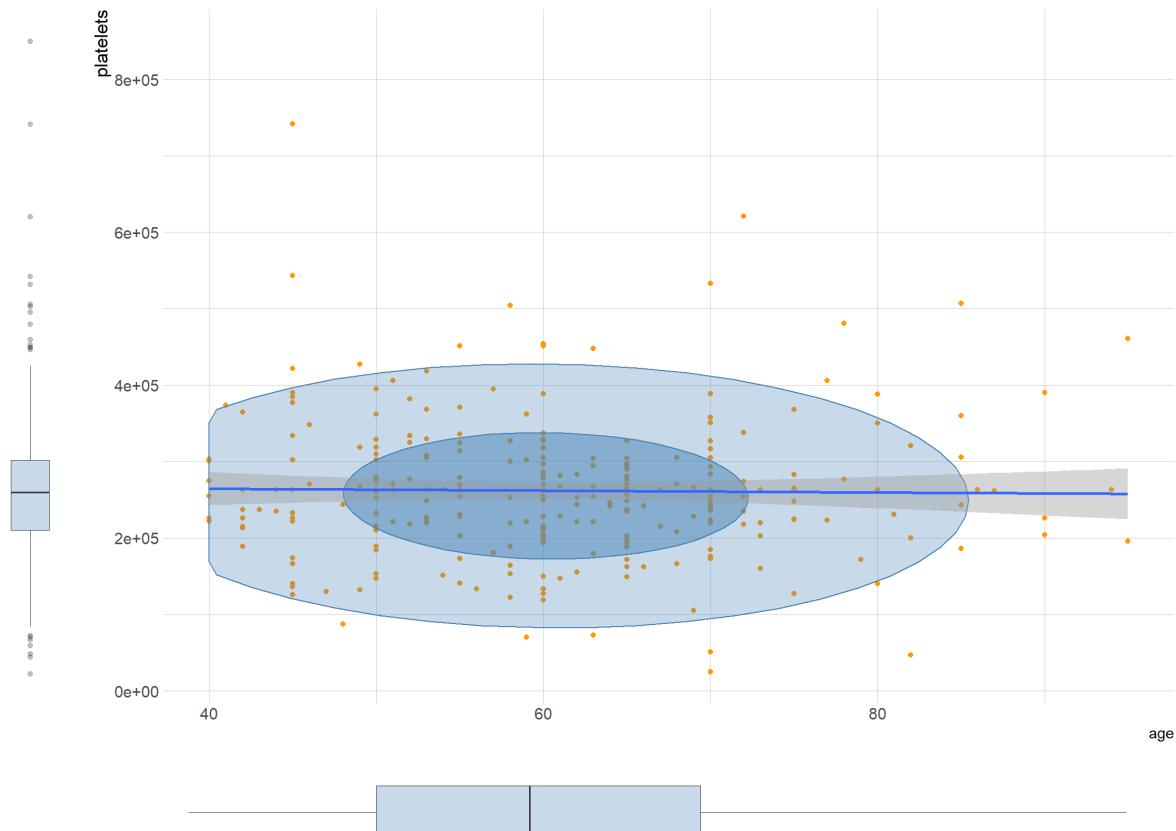


'age' vs 'platelets'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
age	platelets	0.0002493	-0.0033599	11.89252	0.0690846	0.7928696	1

Table 7: Summary of linear model

Scatterplots with age and platelets

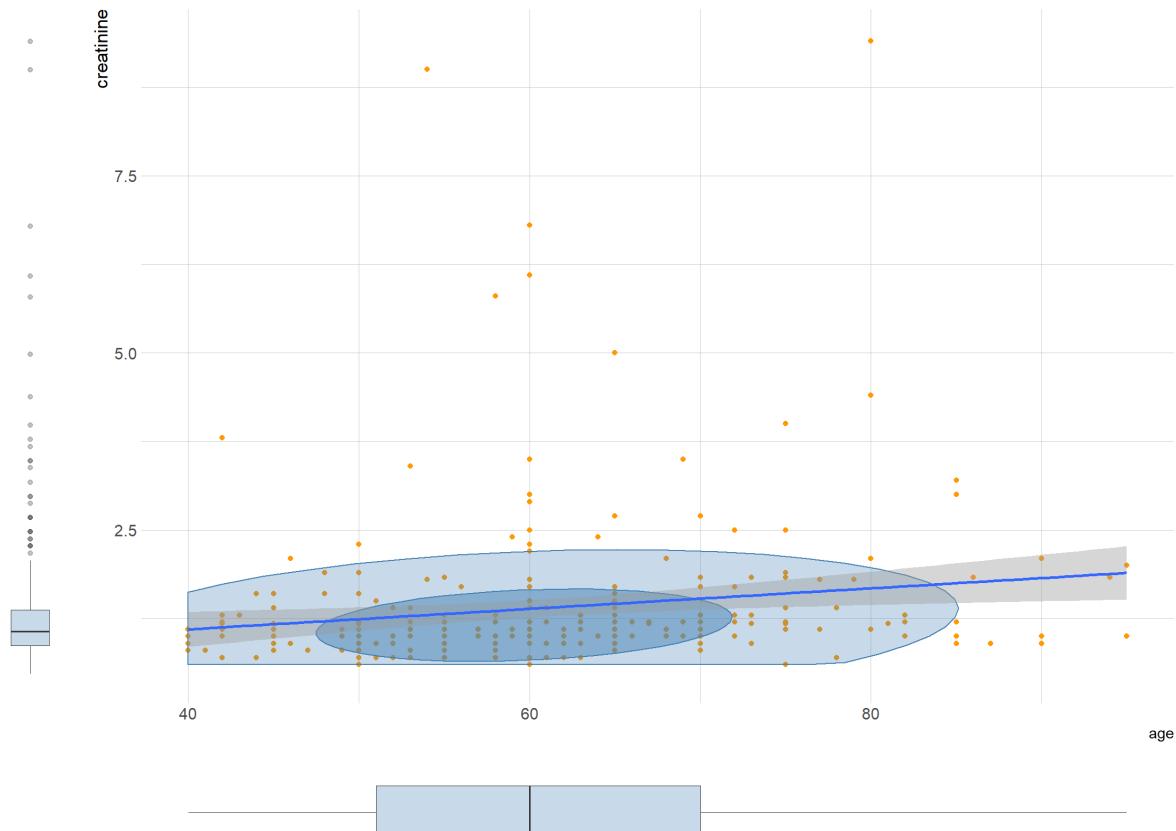


'age' vs 'creatinine'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
age	creatinine	0.026558	0.0230438	11.735	7.557276	0.0063698	1

Table 7: Summary of linear model

Scatterplots with age and creatinine

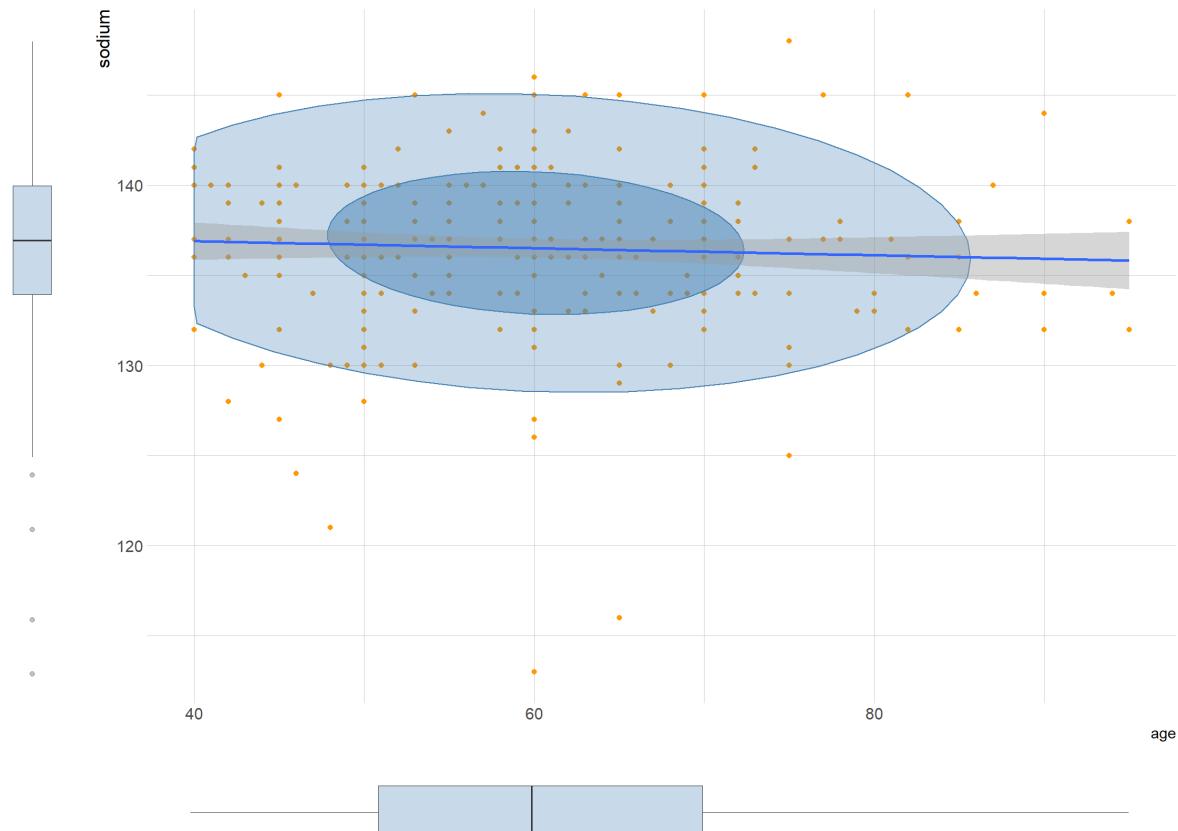


'age' vs 'sodium'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
age	sodium	0.0027644	-0.0008357	11.87755	0.7678659	0.3816373	1

Table 7: Summary of linear model

Scatterplots with age and sodium

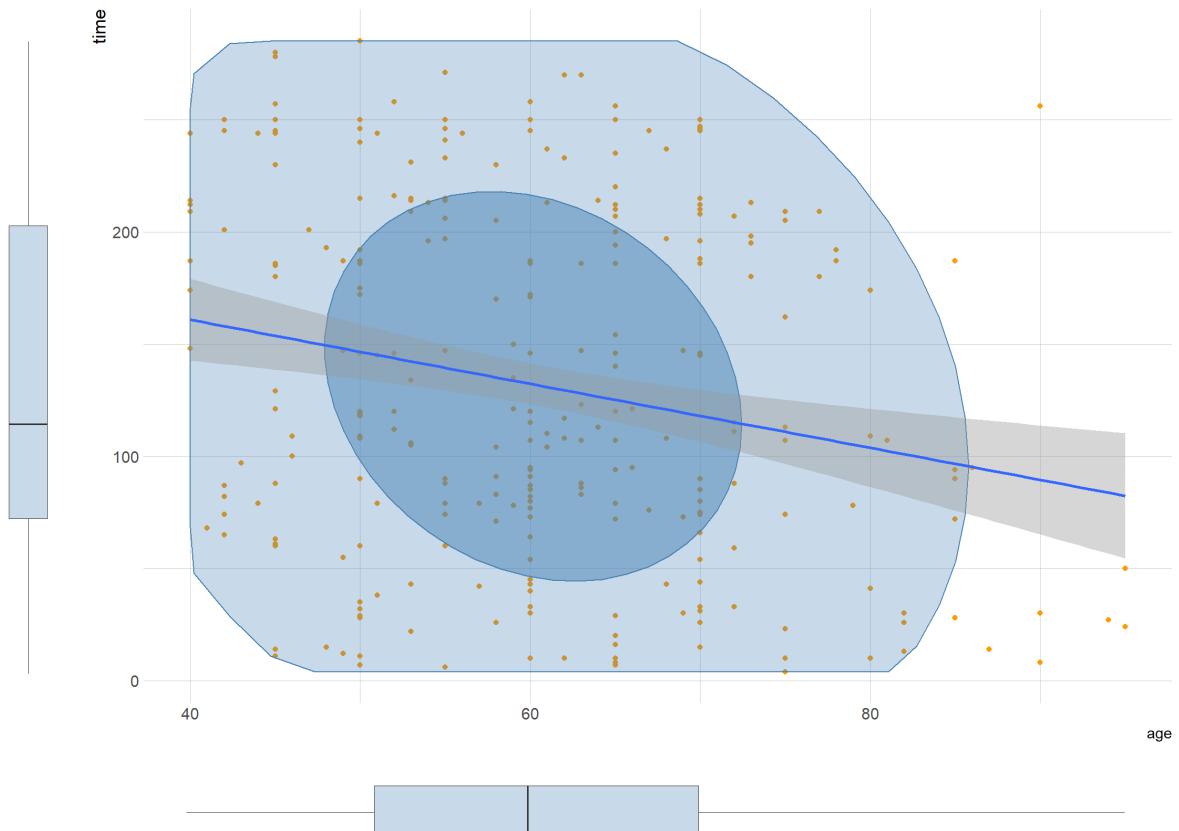


'age' vs 'time'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
age	time	0.0461639	0.0427204	11.61622	13.40628	0.0003003	1

Table 7: Summary of linear model

Scatterplots with age and time

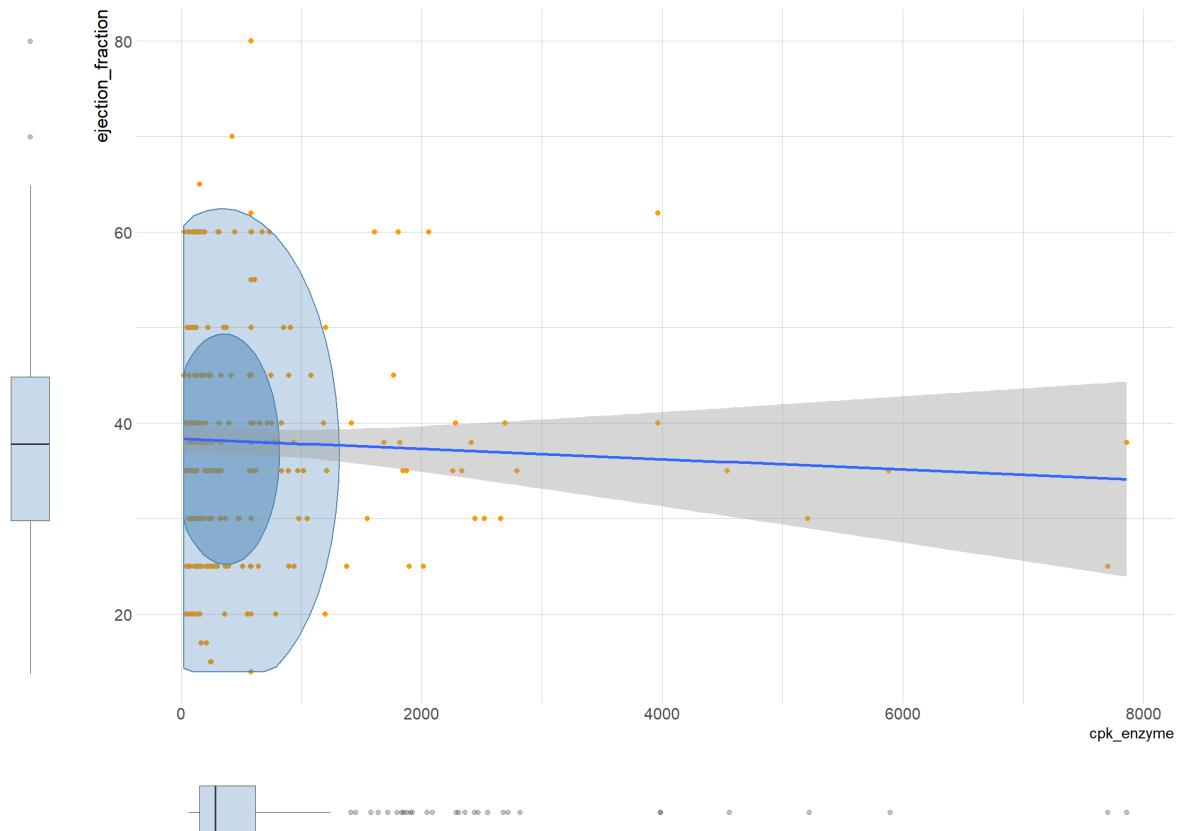


'cpk_enzyme' vs 'ejection_fraction'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
cpk_enzyme	ejection_fraction	0.001943	-0.0014175	970.9753	0.5781966	0.4476233	1

Table 7: Summary of linear model

Scatterplots with cpk_enzyme and ejection_fraction

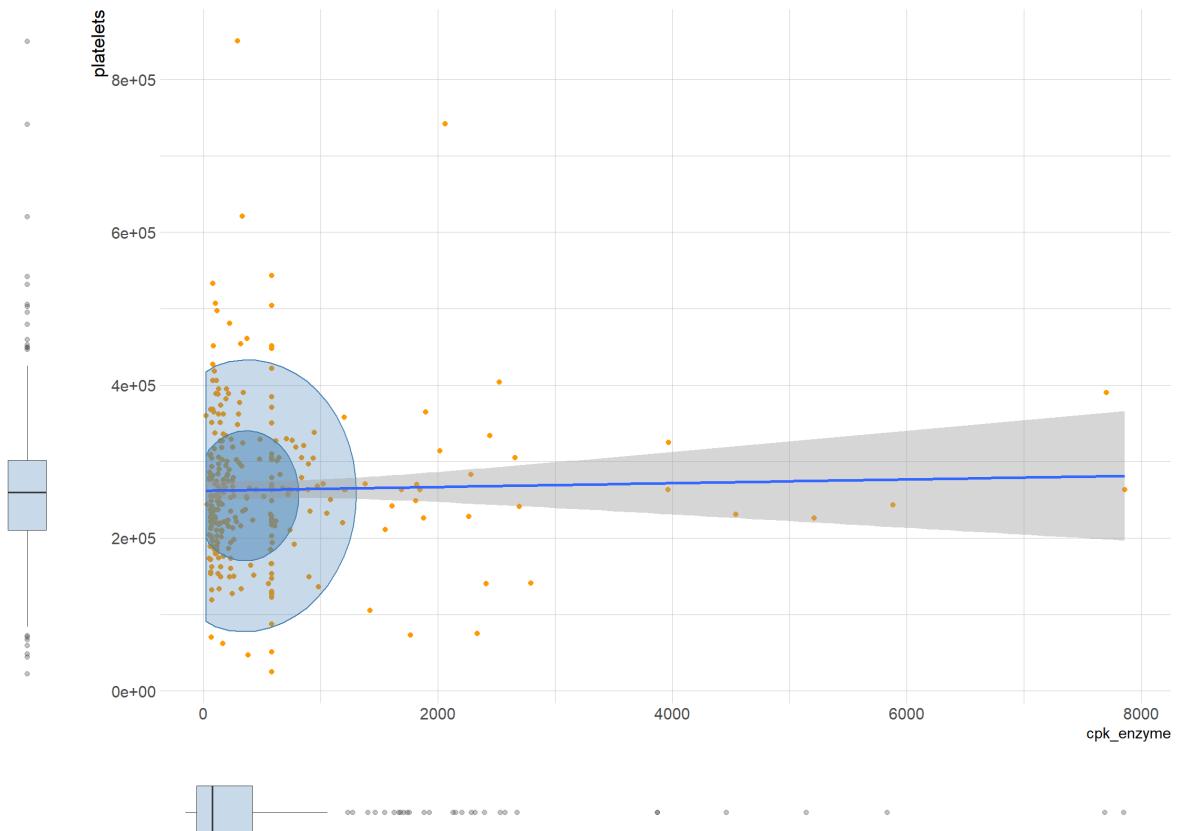


'cpk_enzyme' vs 'platelets'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
cpk_enzyme	platelets	0.0005985	-0.0027665	971.6291	0.1778483	0.673534	1

Table 7: Summary of linear model

Scatterplots with cpk_enzyme and platelets

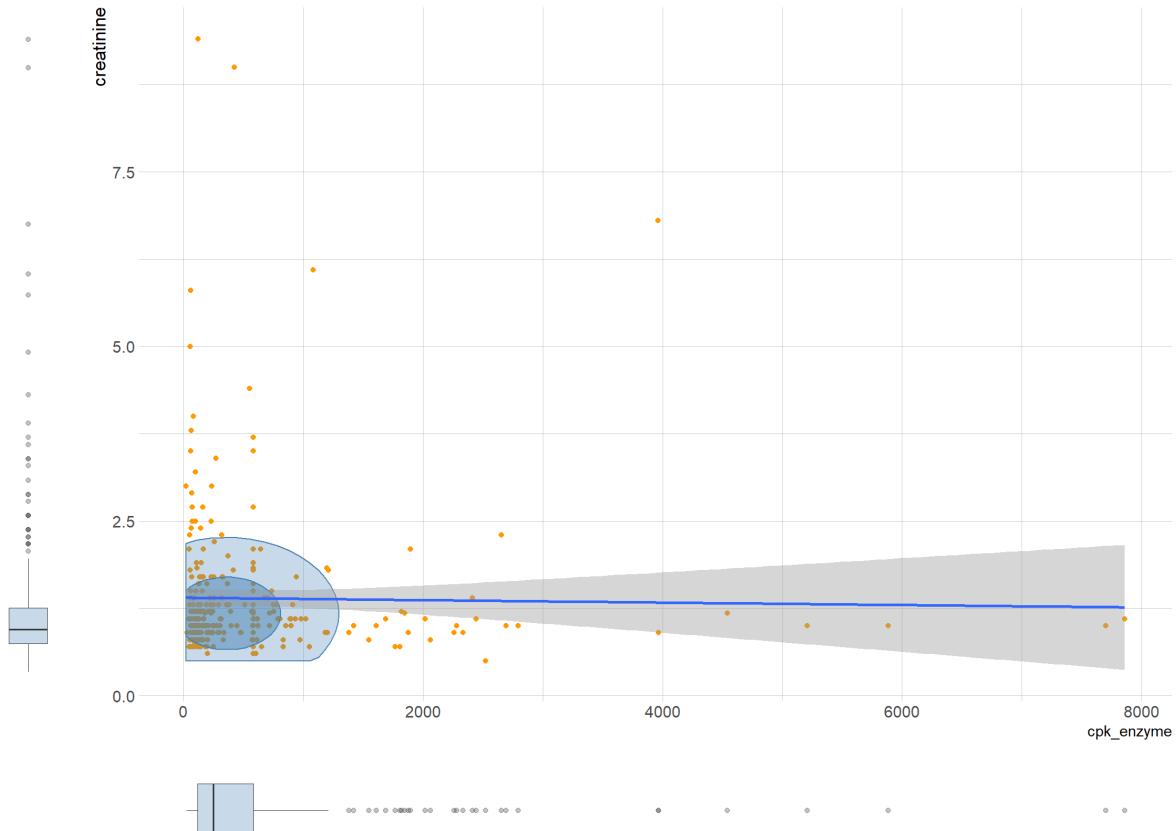


'cpk_enzyme' vs 'creatinine'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
cpk_enzyme	creatinine	0.0002692	-0.0030969	971.7891	0.0799853	0.7775144	1

Table 7: Summary of linear model

Scatterplots with cpk_enzyme and creatinine

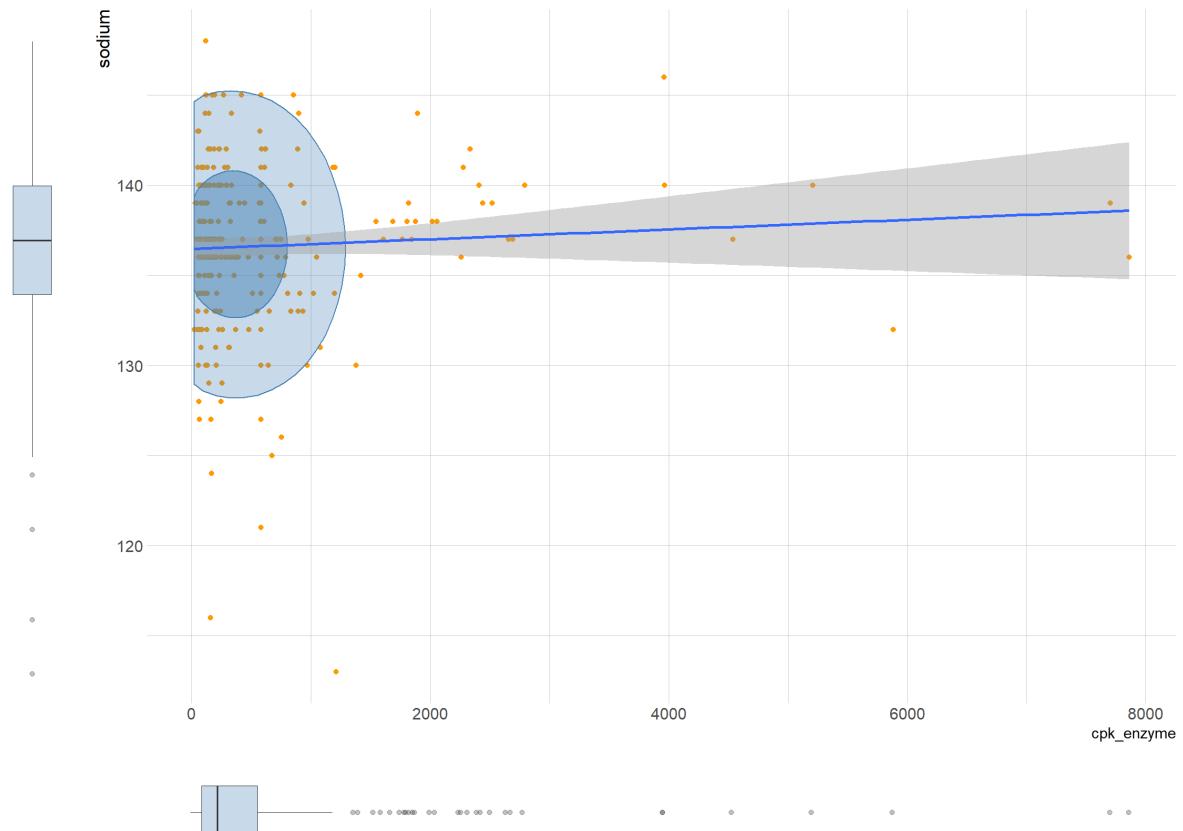


'cpk_enzyme' vs 'sodium'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
cpk_enzyme	sodium	0.0035462	0.0001912	970.1951	1.056976	0.3047425	1

Table 7: Summary of linear model

Scatterplots with cpk_enzyme and sodium

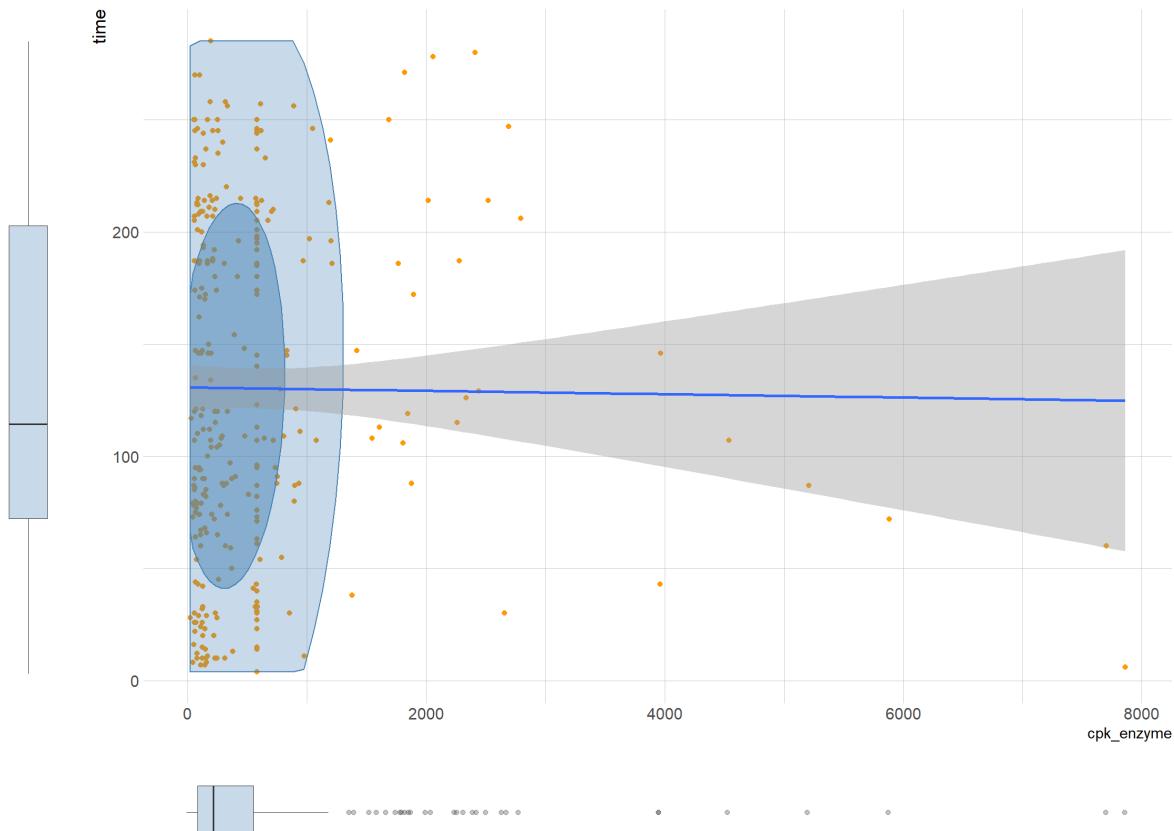


'cpk_enzyme' vs 'time'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
cpk_enzyme	time	8.73e-05	-0.0032794	971.8775	0.0259426	0.8721501	1

Table 7: Summary of linear model

Scatterplots with cpk_enzyme and time

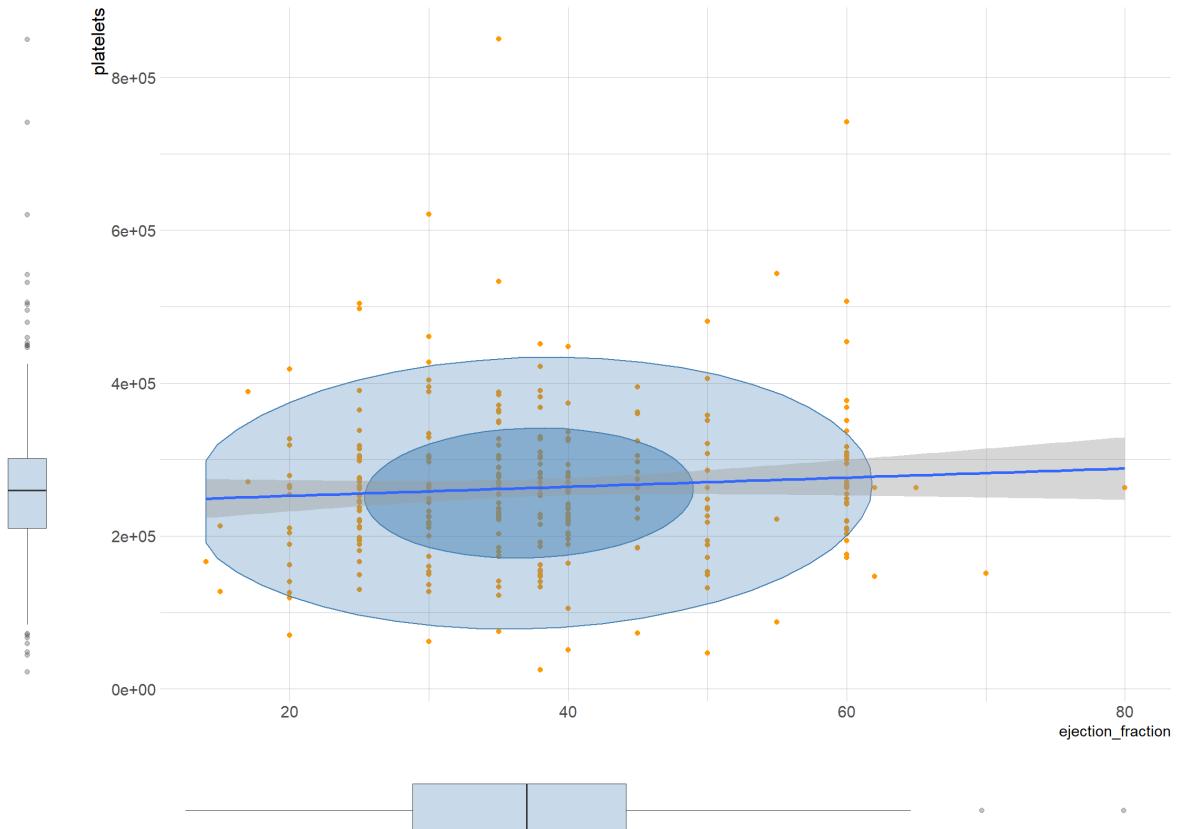


'ejection_fraction' vs 'platelets'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
ejection_fraction	platelets	0.0052096	0.0018601	11.82383	1.55535	0.2133299	1

Table 7: Summary of linear model

Scatterplots with ejection_fraction and platelets

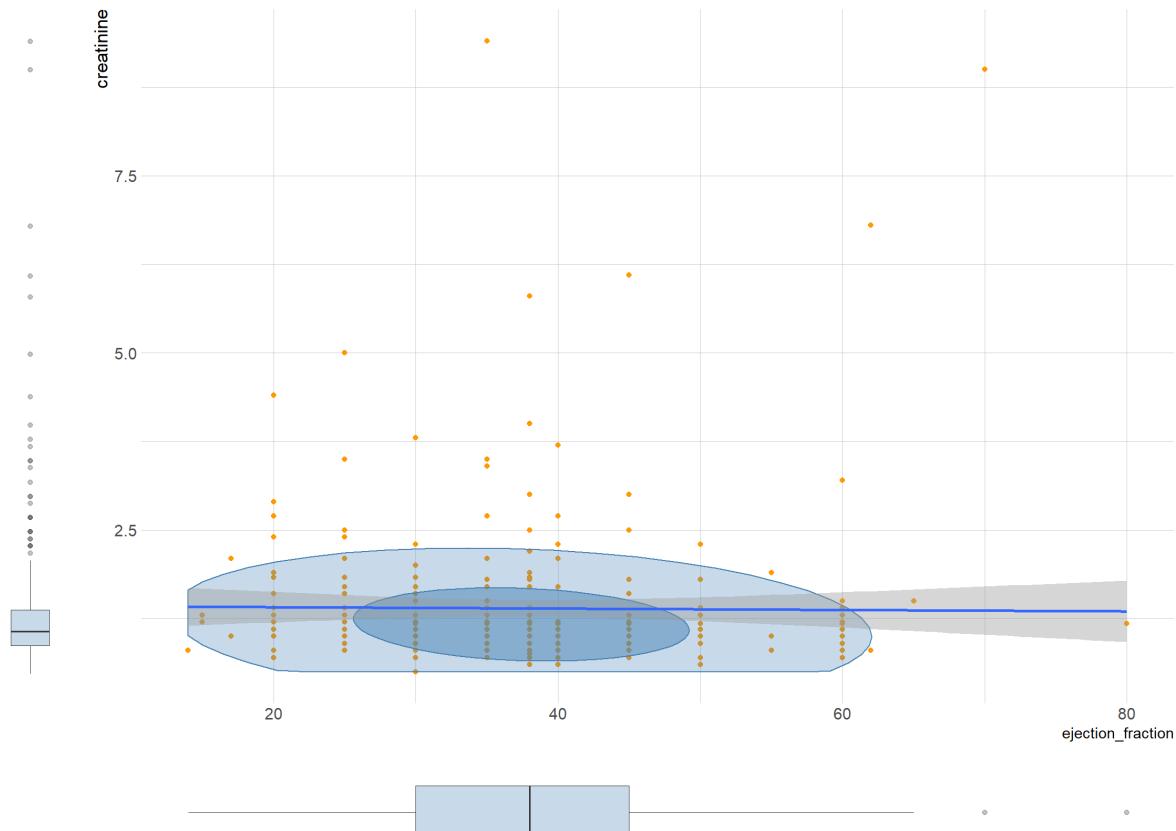


'ejection_fraction' vs 'creatinine'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
ejection_fraction	creatinine	0.0001277	-0.0032388	11.85399	0.0379454	0.845686	1

Table 7: Summary of linear model

Scatterplots with ejection_fraction and creatinine

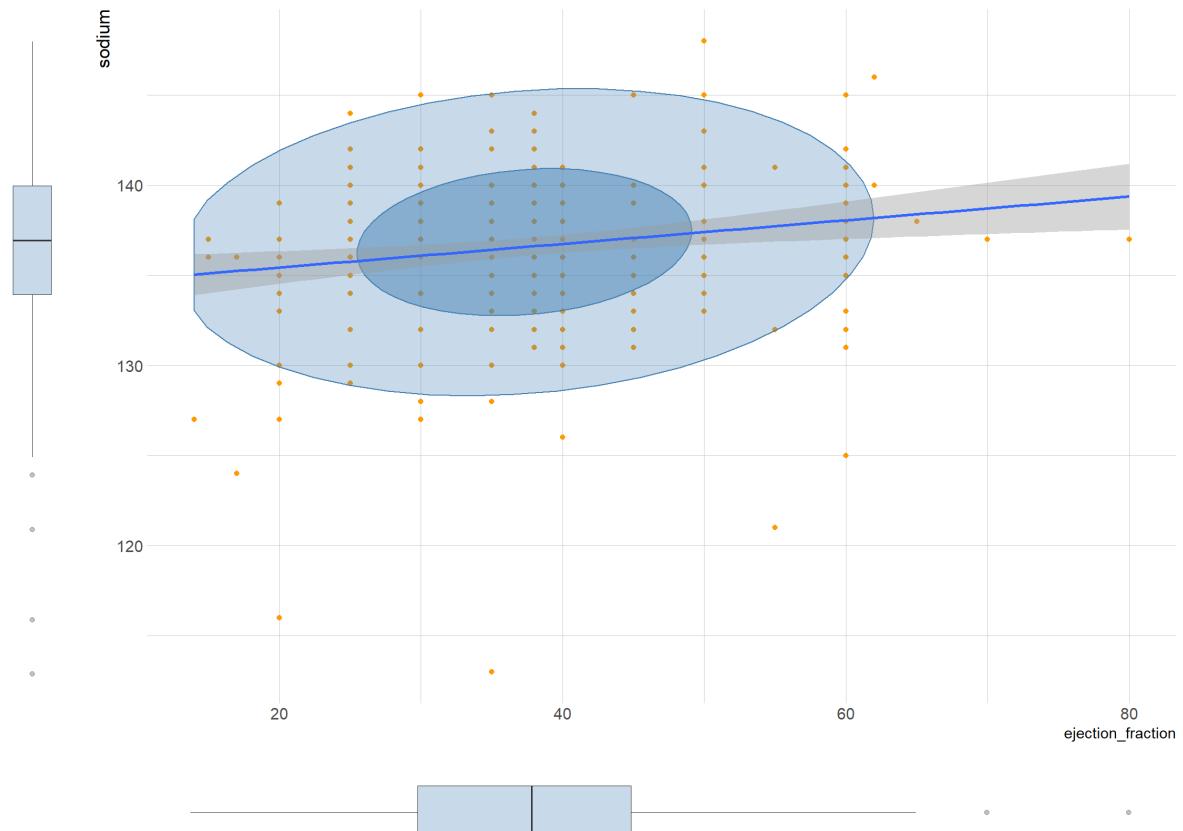


'ejection_fraction' vs 'sodium'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
ejection_fraction	sodium	0.0309416	0.0276788	11.6699	9.483081	0.0022677	1

Table 7: Summary of linear model

Scatterplots with ejection_fraction and sodium

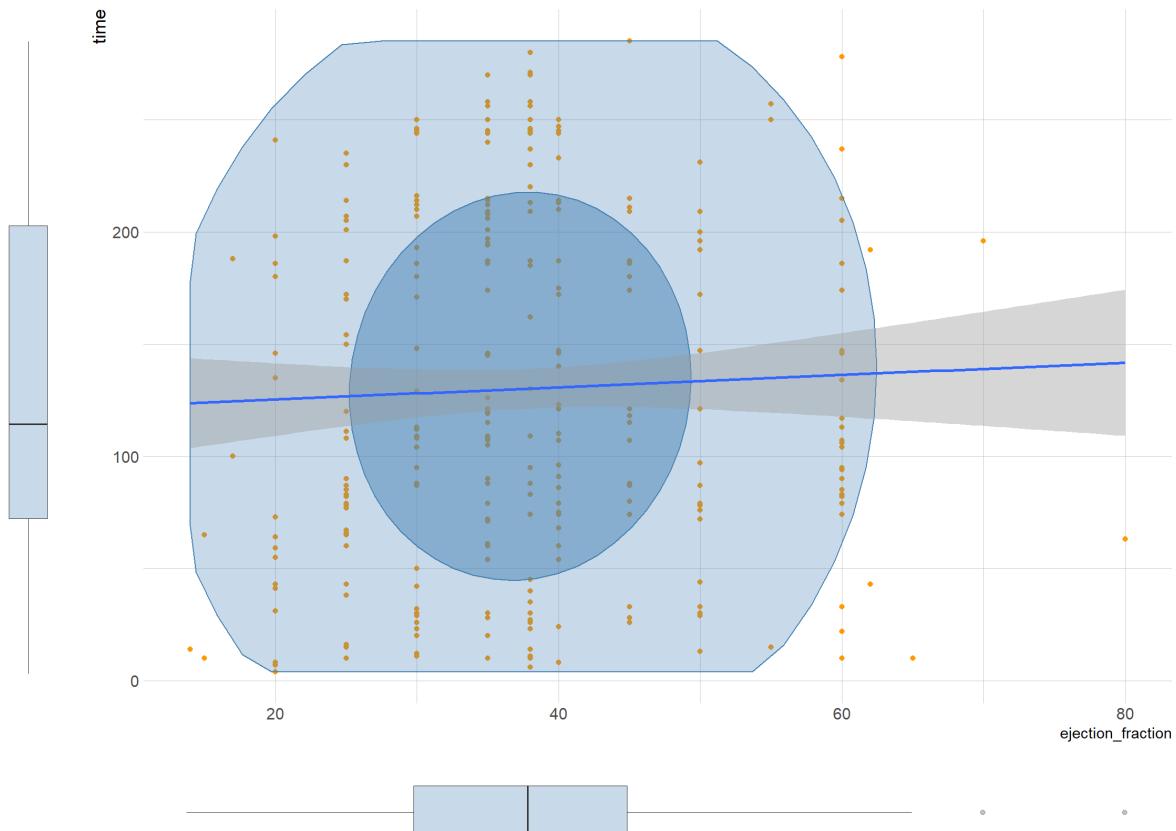


'ejection_fraction' vs 'time'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
ejection_fraction	time	0.0017413	-0.0016198	11.84442	0.5180769	0.4722292	1

Table 7: Summary of linear model

Scatterplots with ejection_fraction and time

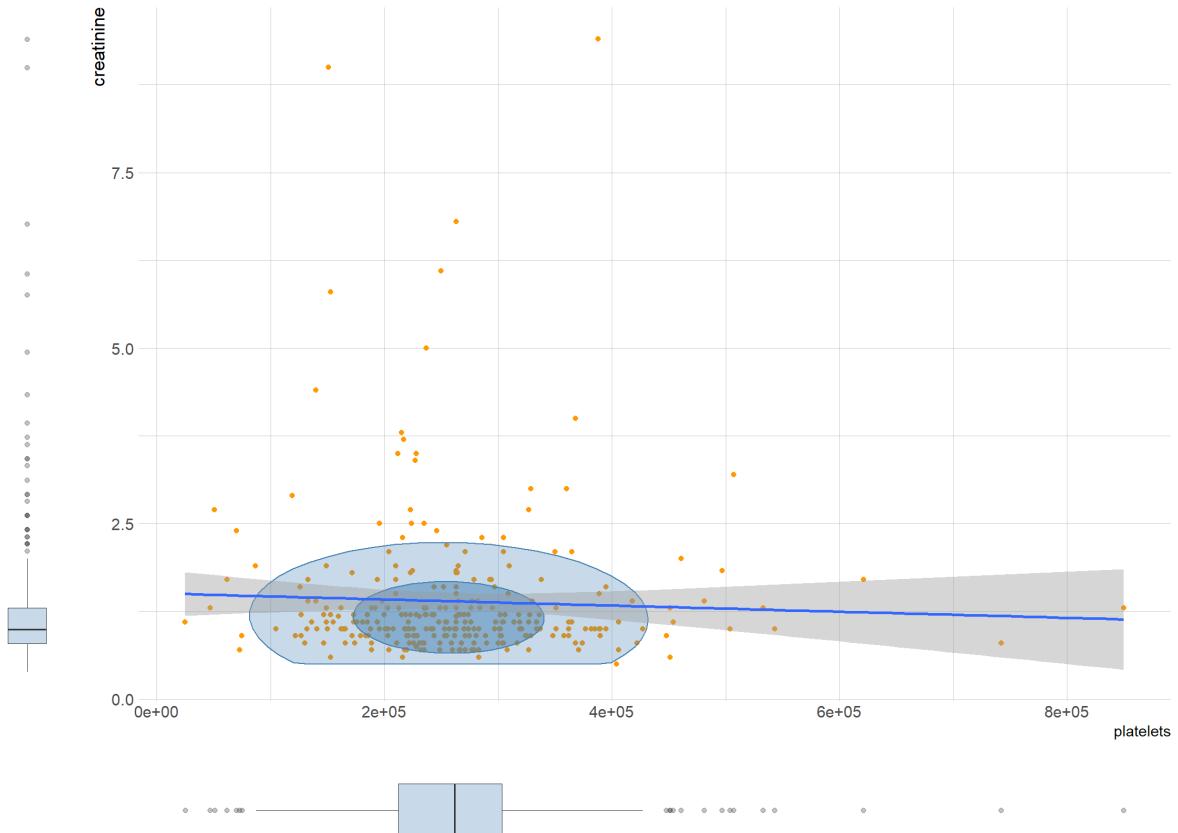


'platelets' vs 'creatinine'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
platelets	creatinine	0.0016973	-0.001664	97885.58	0.5049497	0.4778909	1

Table 7: Summary of linear model

Scatterplots with platelets and creatinine

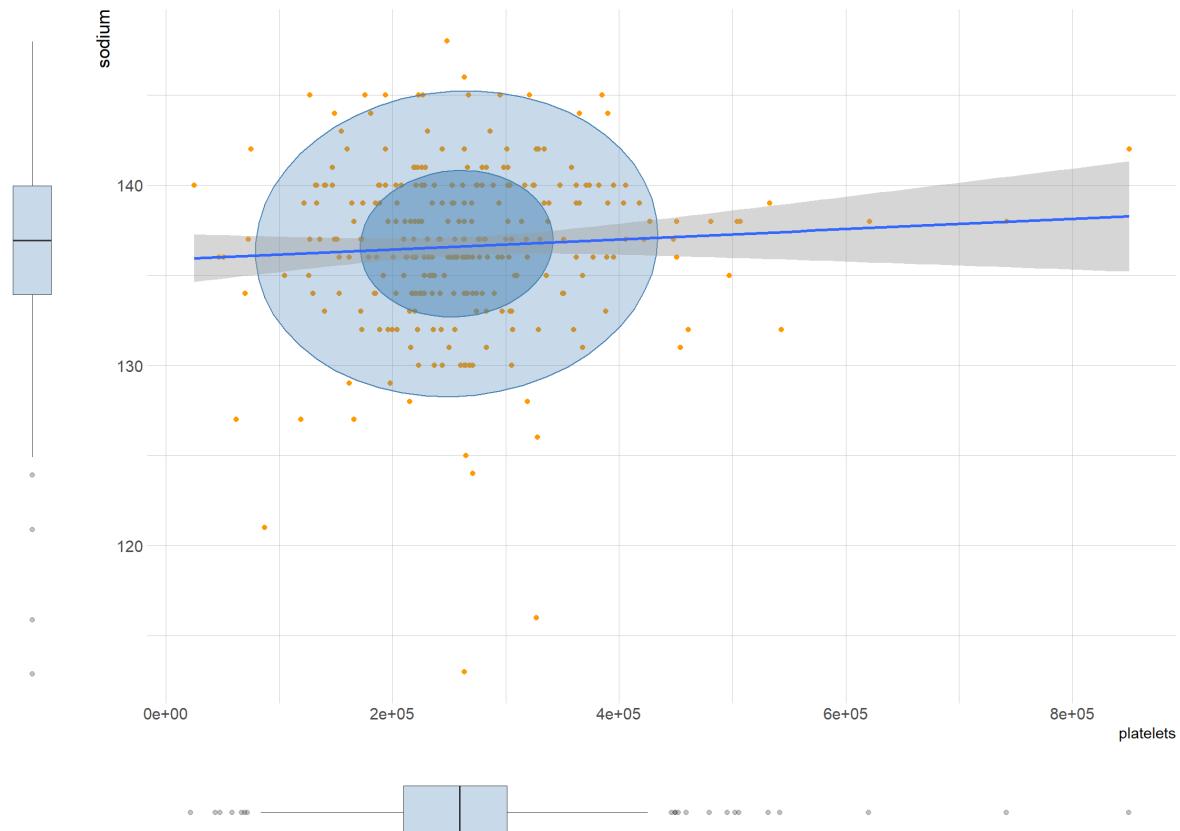


'platelets' vs 'sodium'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
platelets	sodium	0.0038595	0.0005055	97779.52	1.150703	0.2842728	1

Table 7: Summary of linear model

Scatterplots with platelets and sodium

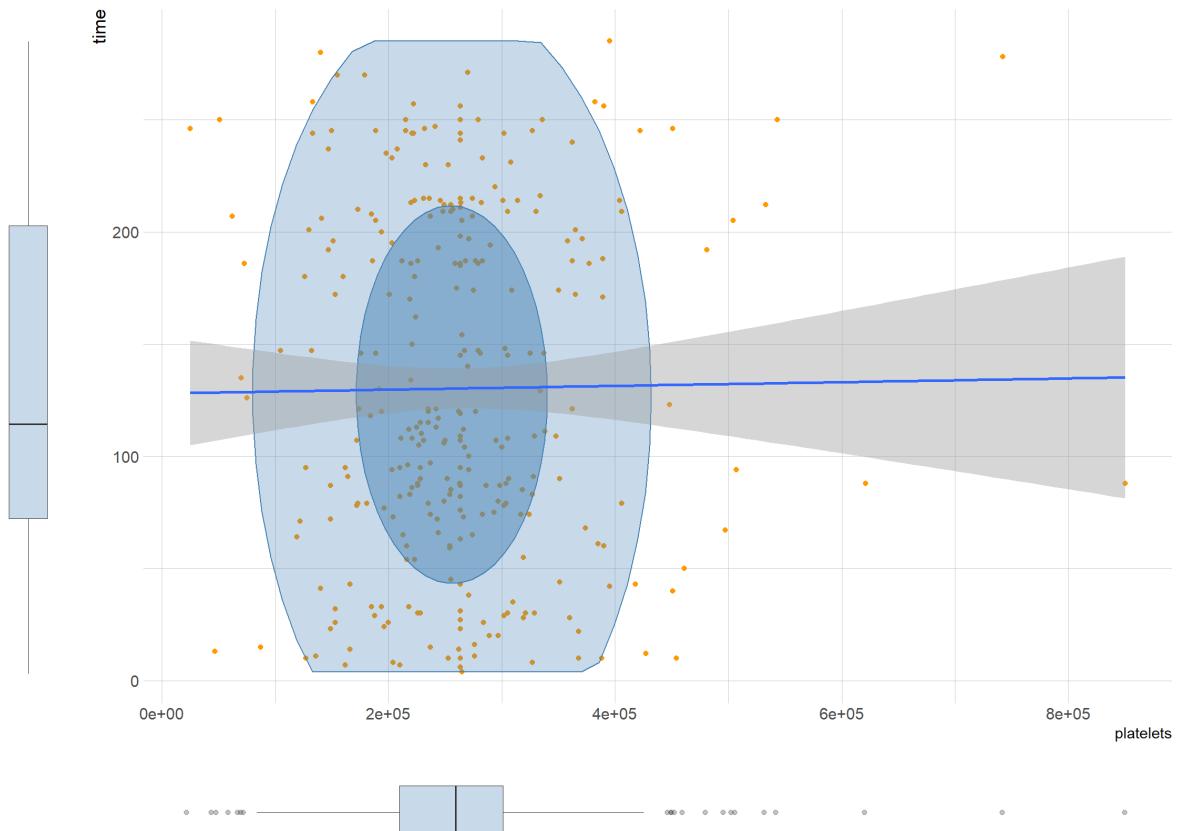


'platelets' vs 'time'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
platelets	time	0.0001105	-0.0032561	97963.34	0.0328347	0.8563315	1

Table 7: Summary of linear model

Scatterplots with platelets and time

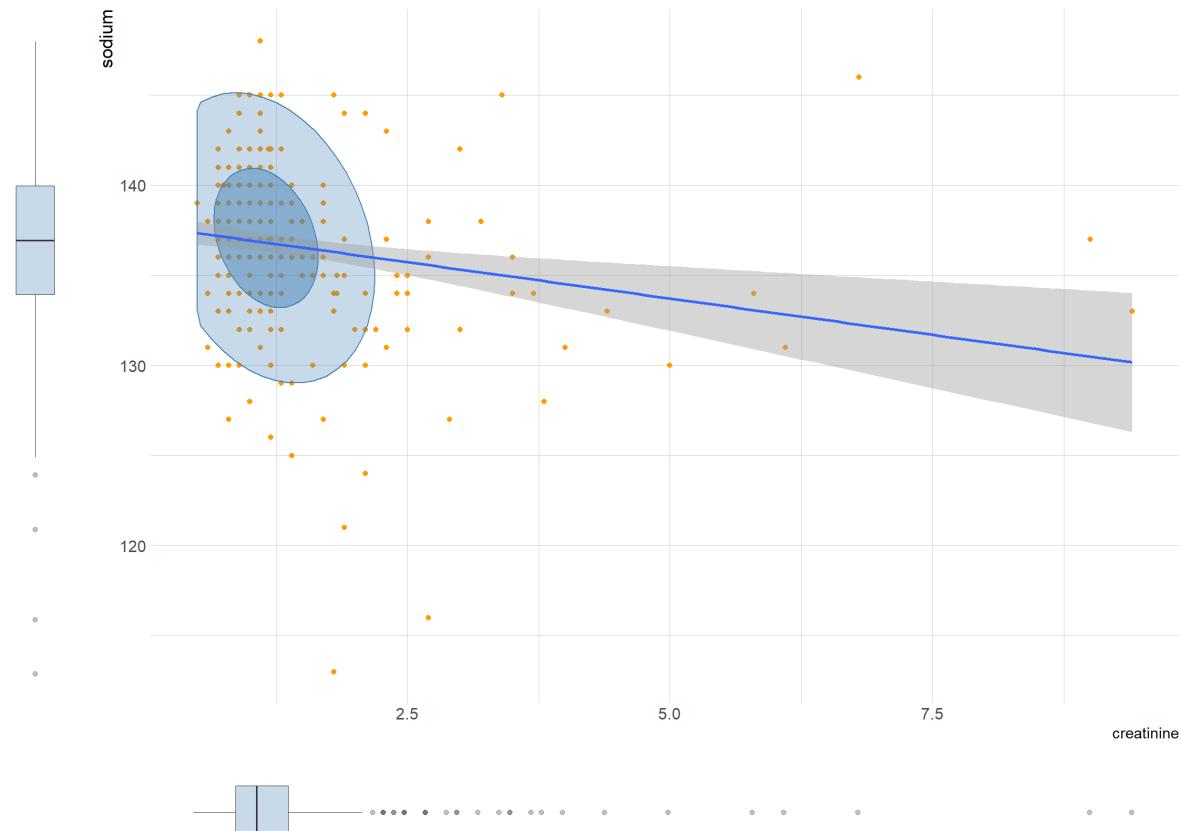


'creatinine' vs 'sodium'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
creatinine	sodium	0.035757	0.0325104	1.017555	11.01364	0.0010171	1

Table 7: Summary of linear model

Scatterplots with creatinine and sodium

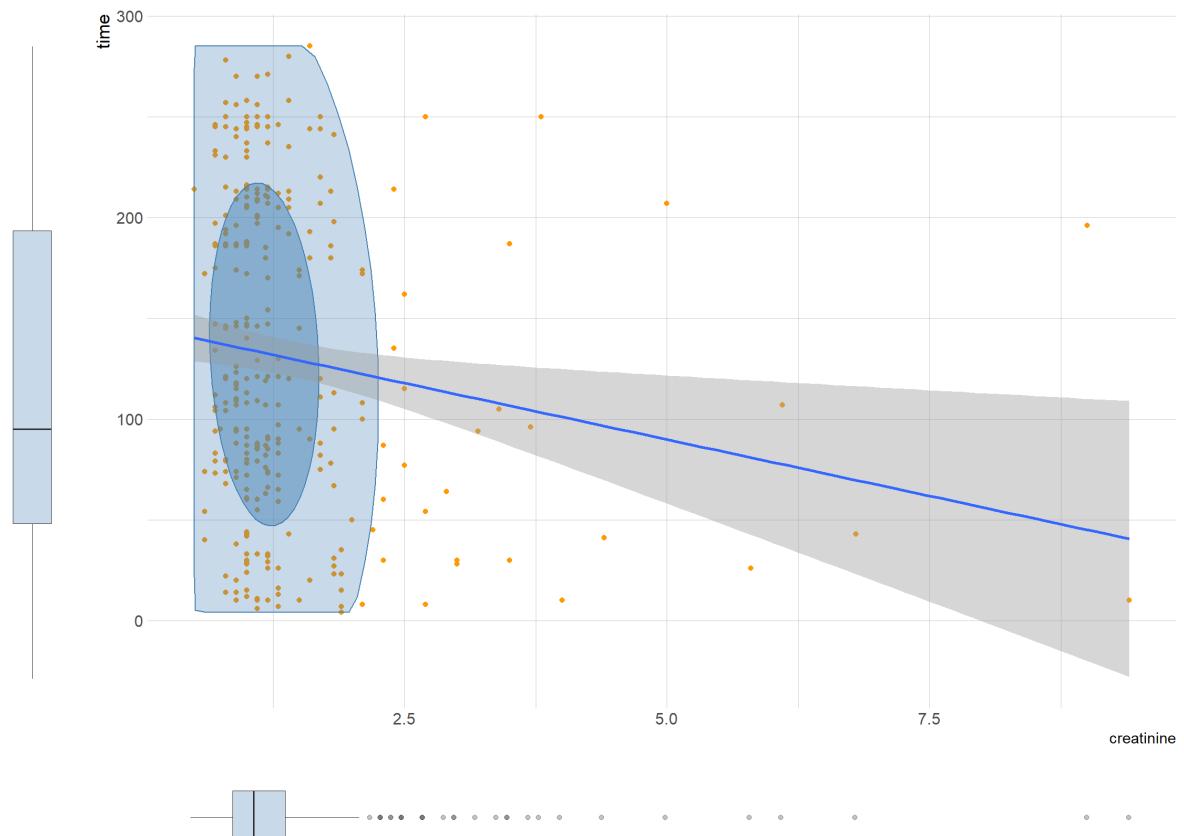


'creatinine' vs 'time'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
creatinine	time	0.0222951	0.0190032	1.024633	6.77264	0.0097209	1

Table 7: Summary of linear model

Scatterplots with creatinine and time

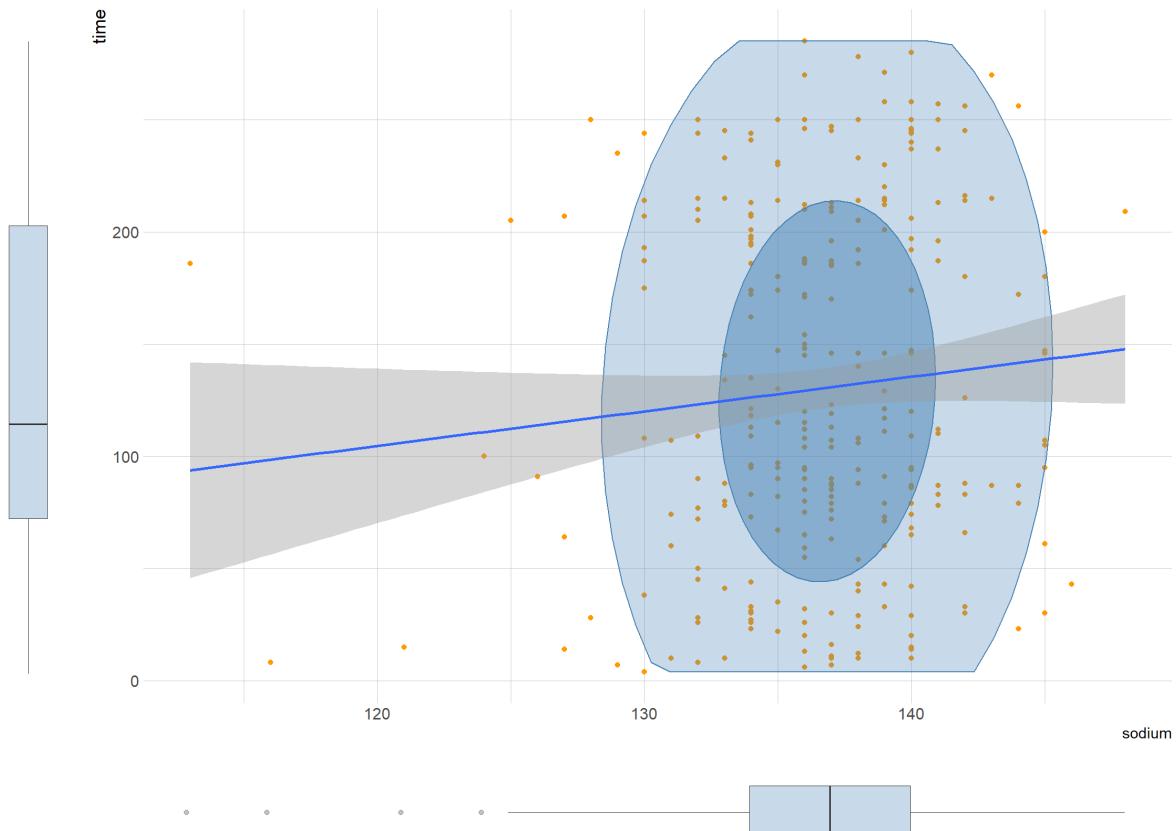


'sodium' vs 'time'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
sodium	time	0.0076808	0.0043396	4.402893	2.298846	0.1305338	1

Table 7: Summary of linear model

Scatterplots with sodium and time



Compare Categorical Variables

first variable	second variable	degree of freedom	statistic	p-value
anaemia	diabetes	4	0.04845	0.99971
anaemia	hblood_pressure	4	0.43590	0.97943
anaemia	sex	6	3.44266	0.75158
anaemia	smoking	4	3.44182	0.48678
anaemia	death_event	4	1.31313	0.85914
diabetes	hblood_pressure	4	0.04847	0.99971
diabetes	sex	6	10.56187	0.10290
diabetes	smoking	4	6.47634	0.16629
diabetes	death_event	4	0.00113	1.00000
hblood_pressure	sex	6	5.71400	0.45598
hblood_pressure	smoking	4	0.92802	0.92051
hblood_pressure	death_event	4	1.88268	0.75733
sex	smoking	6	59.22212	0.00000
sex	death_event	6	0.69776	0.99454
smoking	death_event	4	0.04764	0.99972

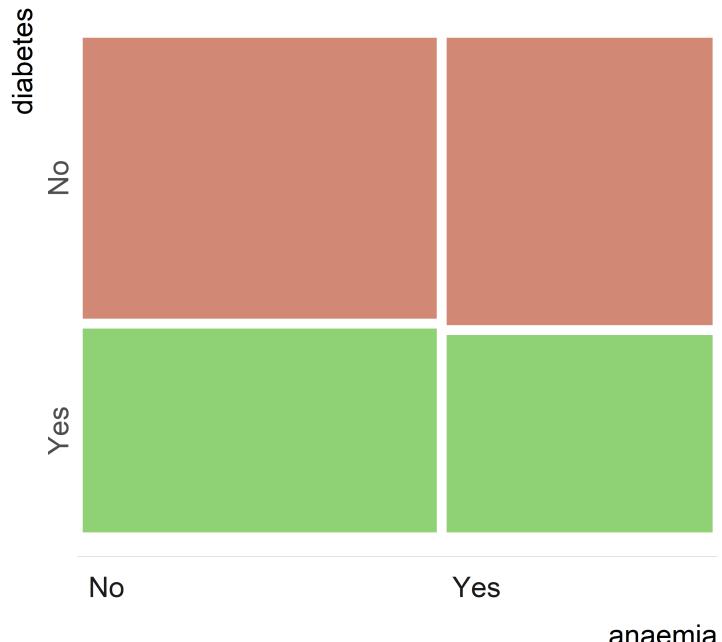
Table 8: Chisquare-test

'anaemia' vs 'diabetes'

	anaemia	diabetes		<Total>
		No	Yes	
No	98	72		170
Yes	76	53		129
<Total>	174	125		299

	anaemia	diabetes		<Total>
		No	Yes	
No	32.78	24.08		56.86
Yes	25.42	17.73		43.14
<Total>	58.19	41.81		100.00

Mosaics plot by 'anaemia' vs 'diabe

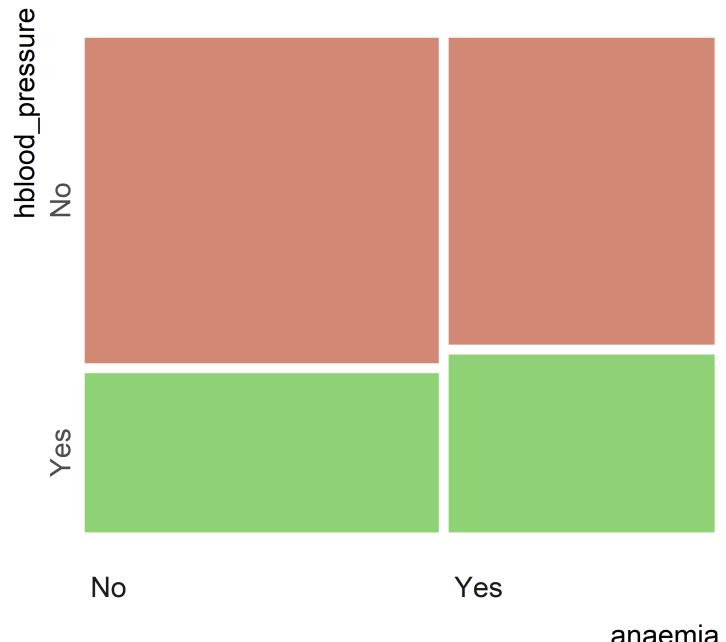


'anaemia' vs 'hblood_pressure'

anaemia	hblood_pressure		
	No	Yes	<Total>
No	113	57	170
Yes	81	48	129
<Total>	194	105	299

anaemia	hblood_pressure		
	No	Yes	<Total>
No	37.79	19.06	56.86
Yes	27.09	16.05	43.14
<Total>	64.88	35.12	100.00

Mosaics plot by 'anaemia' vs 'hblood_pressure'

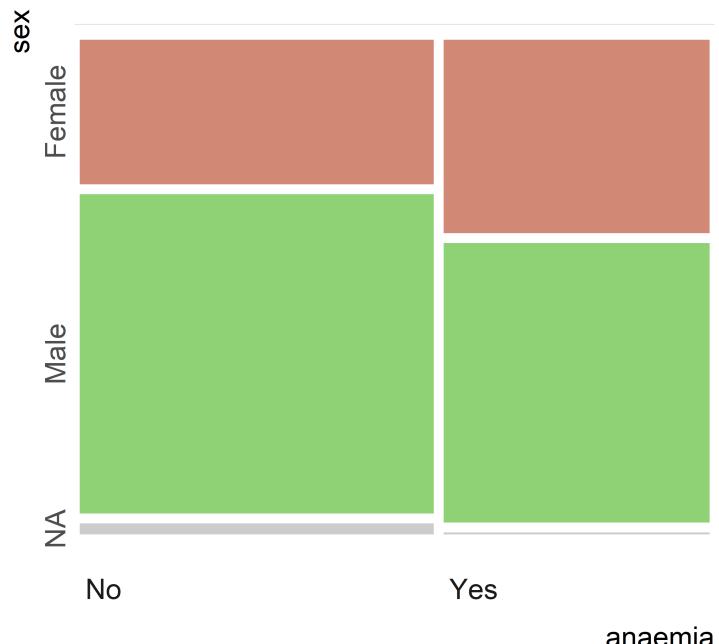


'anaemia' vs 'sex'

anaemia	sex			<Total>
	Female	Male	NA	
No	52	111	7	170
Yes	52	74	3	129
<Total>	104	185	10	299

anaemia	sex			<Total>
	Female	Male	NA	
No	17.39	37.12	2.34	56.86
Yes	17.39	24.75	1.00	43.14
<Total>	34.78	61.87	3.34	100.00

Mosaics plot by 'anaemia' vs 'sex'

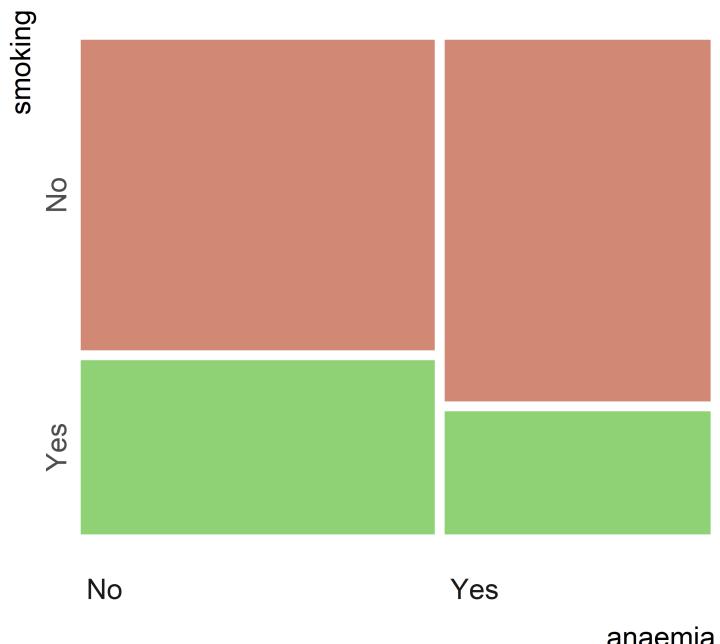


'anaemia' vs 'smoking'

anaemia	smoking		
	No	Yes	<Total>
No	108	62	170
Yes	95	34	129
<Total>	203	96	299

anaemia	smoking		
	No	Yes	<Total>
No	36.12	20.74	56.86
Yes	31.77	11.37	43.14
<Total>	67.89	32.11	100.00

Mosaics plot by 'anaemia' vs 'smok

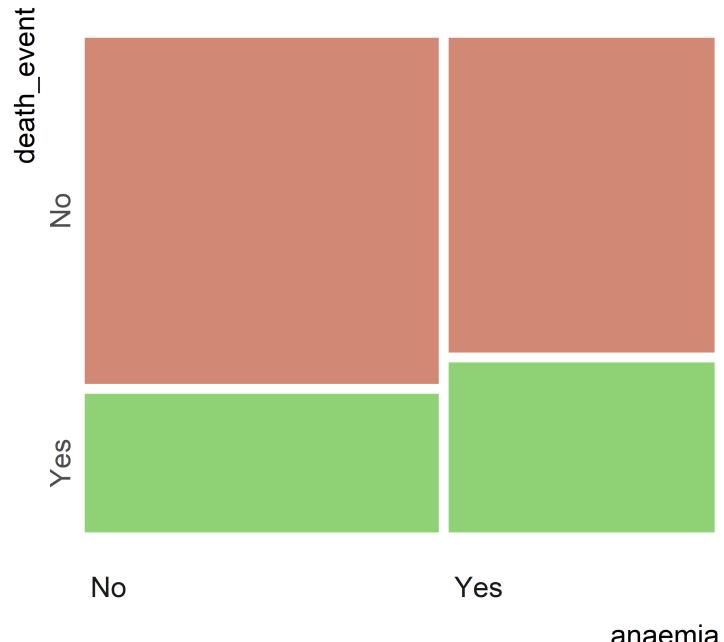


'anaemia' vs 'death_event'

anaemia	death_event		
	No	Yes	<Total>
No	120	50	170
Yes	83	46	129
<Total>	203	96	299

anaemia	death_event		
	No	Yes	<Total>
No	40.13	16.72	56.86
Yes	27.76	15.38	43.14
<Total>	67.89	32.11	100.00

Mosaics plot by 'anaemia' vs 'death

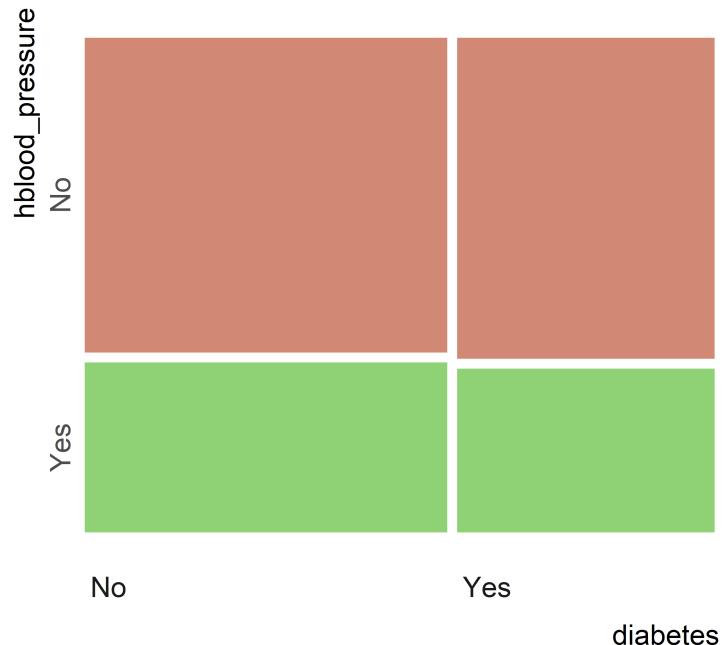


'diabetes' vs 'hblood_pressure'

diabetes	hblood_pressure		
	No	Yes	<Total>
No	112	62	174
Yes	82	43	125
<Total>	194	105	299

diabetes	hblood_pressure		
	No	Yes	<Total>
No	37.46	20.74	58.19
Yes	27.42	14.38	41.81
<Total>	64.88	35.12	100.00

Mosaics plot by 'diabetes' vs 'hblood_pressure'

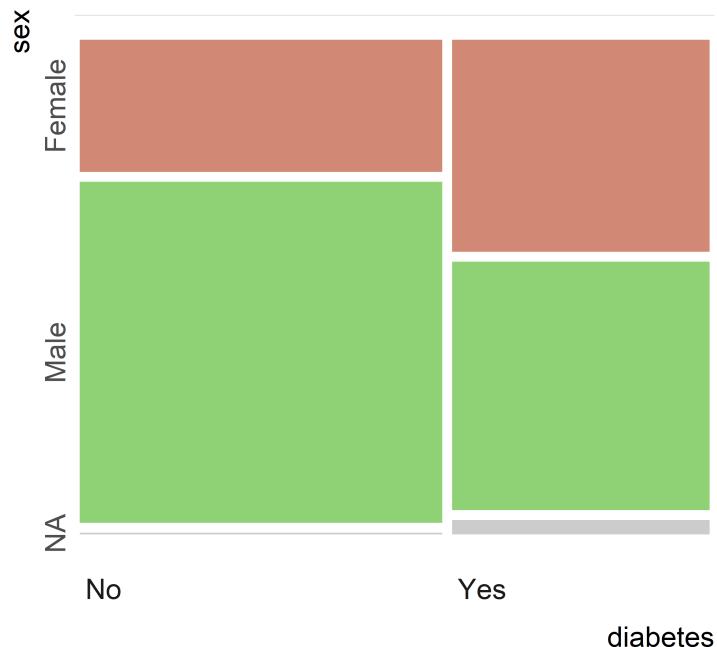


'diabetes' vs 'sex'

diabetes	sex			<Total>
	Female	Male	NA	
No	49	121	4	174
Yes	55	64	6	125
<Total>	104	185	10	299

diabetes	sex			<Total>
	Female	Male	NA	
No	16.39	40.47	1.34	58.19
Yes	18.39	21.40	2.01	41.81
<Total>	34.78	61.87	3.34	100.00

Mosaics plot by 'diabetes' vs 'sex'

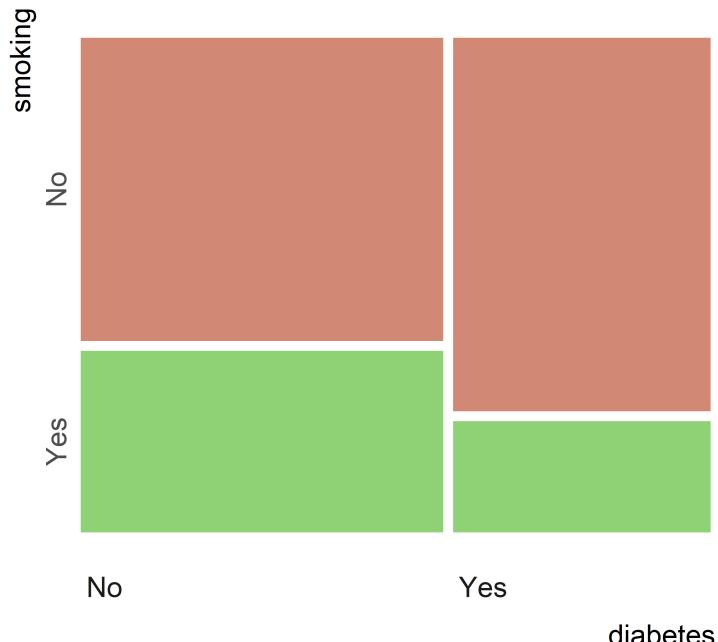


'diabetes' vs 'smoking'

diabetes	smoking		<Total>
	No	Yes	
No	108	66	174
Yes	95	30	125
<Total>	203	96	299

diabetes	smoking		<Total>
	No	Yes	
No	36.12	22.07	58.19
Yes	31.77	10.03	41.81
<Total>	67.89	32.11	100.00

Mosaics plot by 'diabetes' vs 'smok

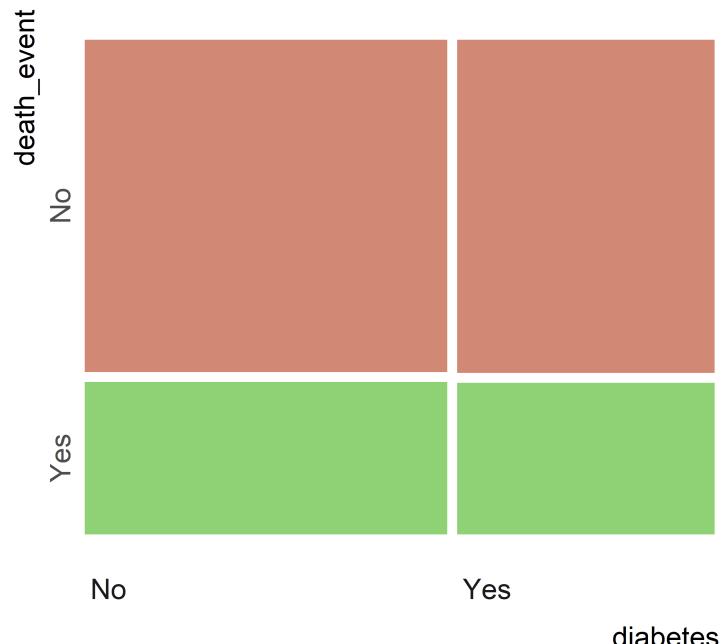


'diabetes' vs 'death_event'

diabetes	death_event		
	No	Yes	<Total>
No	118	56	174
Yes	85	40	125
<Total>	203	96	299

diabetes	death_event		
	No	Yes	<Total>
No	39.46	18.73	58.19
Yes	28.43	13.38	41.81
<Total>	67.89	32.11	100.00

Mosaics plot by 'diabetes' vs 'death'

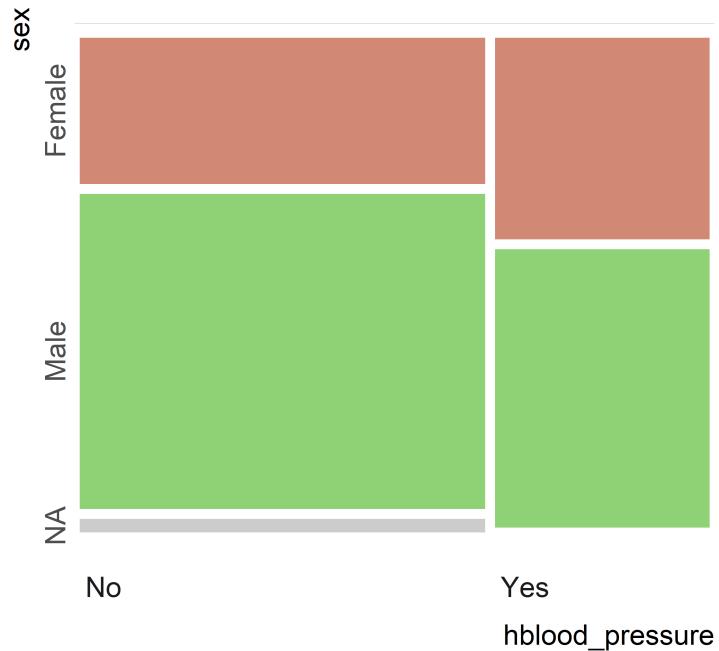


'hblood_pressure' vs 'sex'

hblood_pressure	sex			
	Female	Male	NA	<Total>
No	60	125	9	194
Yes	44	60	1	105
<Total>	104	185	10	299

hblood_pressure	sex			
	Female	Male	NA	<Total>
No	20.07	41.81	3.01	64.88
Yes	14.72	20.07	0.33	35.12
<Total>	34.78	61.87	3.34	100.00

Mosaics plot by 'hblood_pressure'

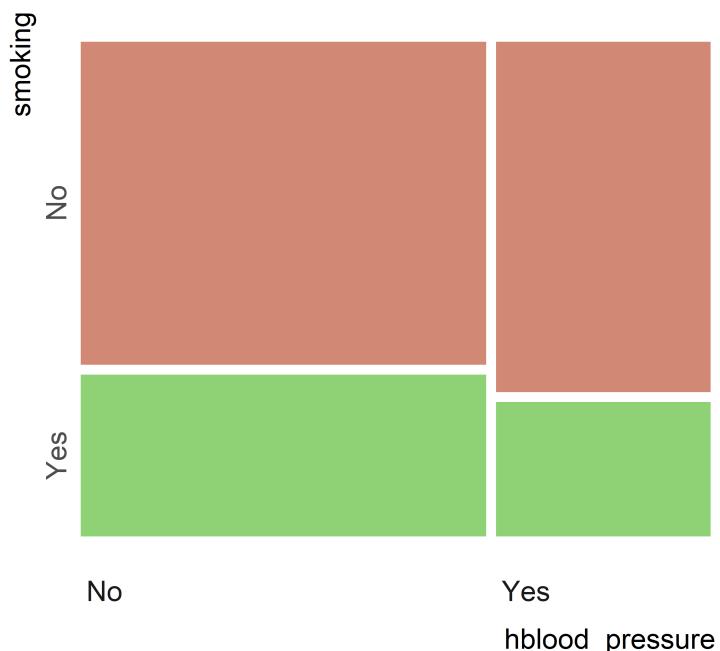


'hblood_pressure' vs 'smoking'

hblood_pressure	smoking		
	No	Yes	<Total>
No	128	66	194
Yes	75	30	105
<Total>	203	96	299

hblood_pressure	smoking		
	No	Yes	<Total>
No	42.81	22.07	64.88
Yes	25.08	10.03	35.12
<Total>	67.89	32.11	100.00

Mosaics plot by 'hblood_pressure'

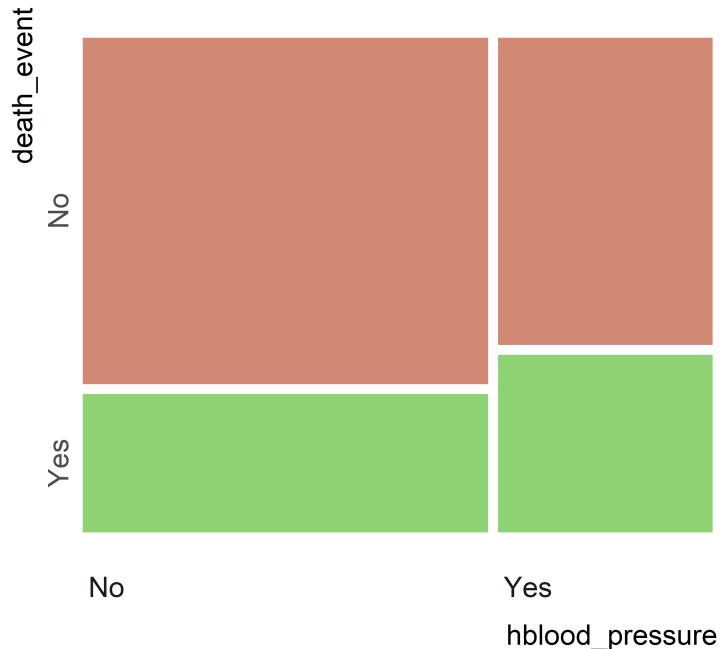


'hblood_pressure' vs 'death_event'

hblood_pressure	death_event		
	No	Yes	<Total>
No	137	57	194
Yes	66	39	105
<Total>	203	96	299

hblood_pressure	death_event		
	No	Yes	<Total>
No	45.82	19.06	64.88
Yes	22.07	13.04	35.12
<Total>	67.89	32.11	100.00

Mosaics plot by 'hblood_pressure'

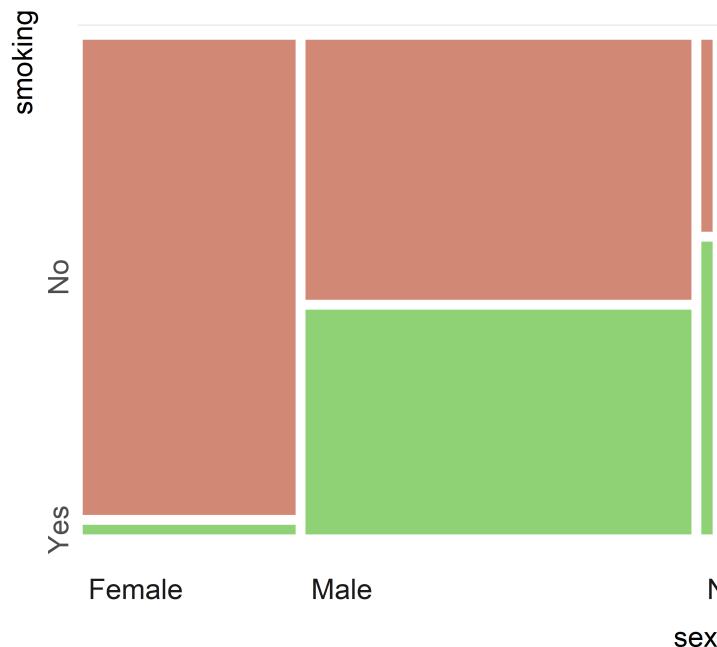


'sex' vs 'smoking'

sex	smoking		<Total>
	No	Yes	
Female	100	4	104
Male	99	86	185
NA	4	6	10
<Total>	203	96	299

sex	smoking		<Total>
	No	Yes	
Female	33.44	1.34	34.78
Male	33.11	28.76	61.87
NA	1.34	2.01	3.34
<Total>	67.89	32.11	100.00

Mosaics plot by 'sex' vs 'smoking'

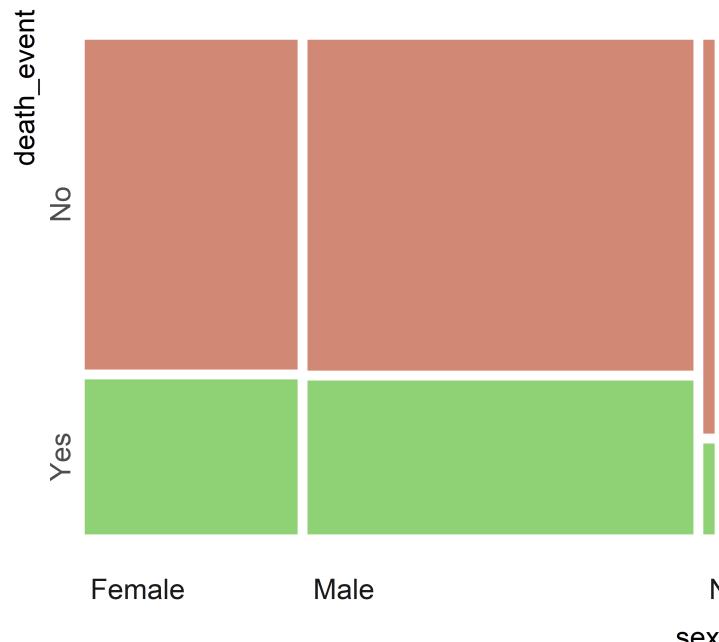


'sex' vs 'death_event'

sex	death_event		
	No	Yes	<Total>
Female	70	34	104
Male	125	60	185
NA	8	2	10
<Total>	203	96	299

sex	death_event		
	No	Yes	<Total>
Female	23.41	11.37	34.78
Male	41.81	20.07	61.87
NA	2.68	0.67	3.34
<Total>	67.89	32.11	100.00

Mosaics plot by 'sex' vs 'death_event'

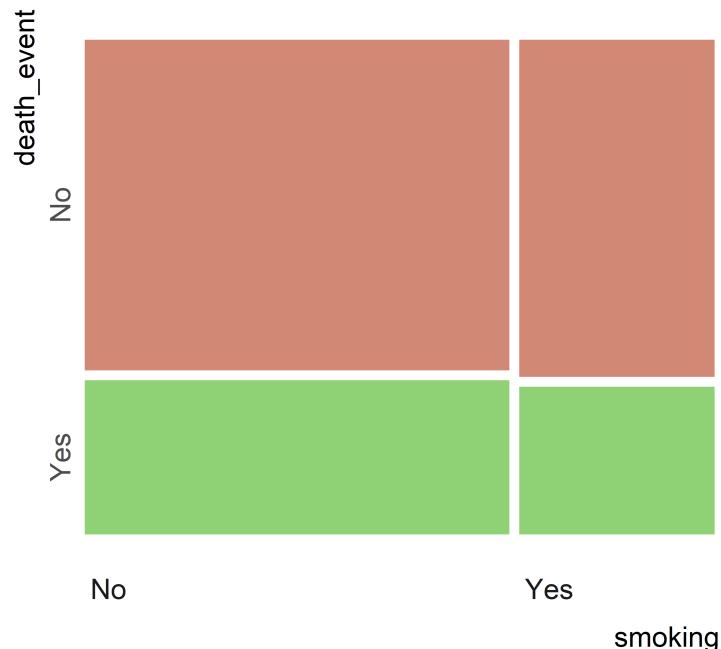


'smoking' vs 'death_event'

smoking	death_event		
	No	Yes	<Total>
No	137	66	203
Yes	66	30	96
<Total>	203	96	299

smoking	death_event		
	No	Yes	<Total>
No	45.82	22.07	67.89
Yes	22.07	10.03	32.11
<Total>	67.89	32.11	100.00

Mosaics plot by 'smoking' vs 'death



Multivariate Analysis

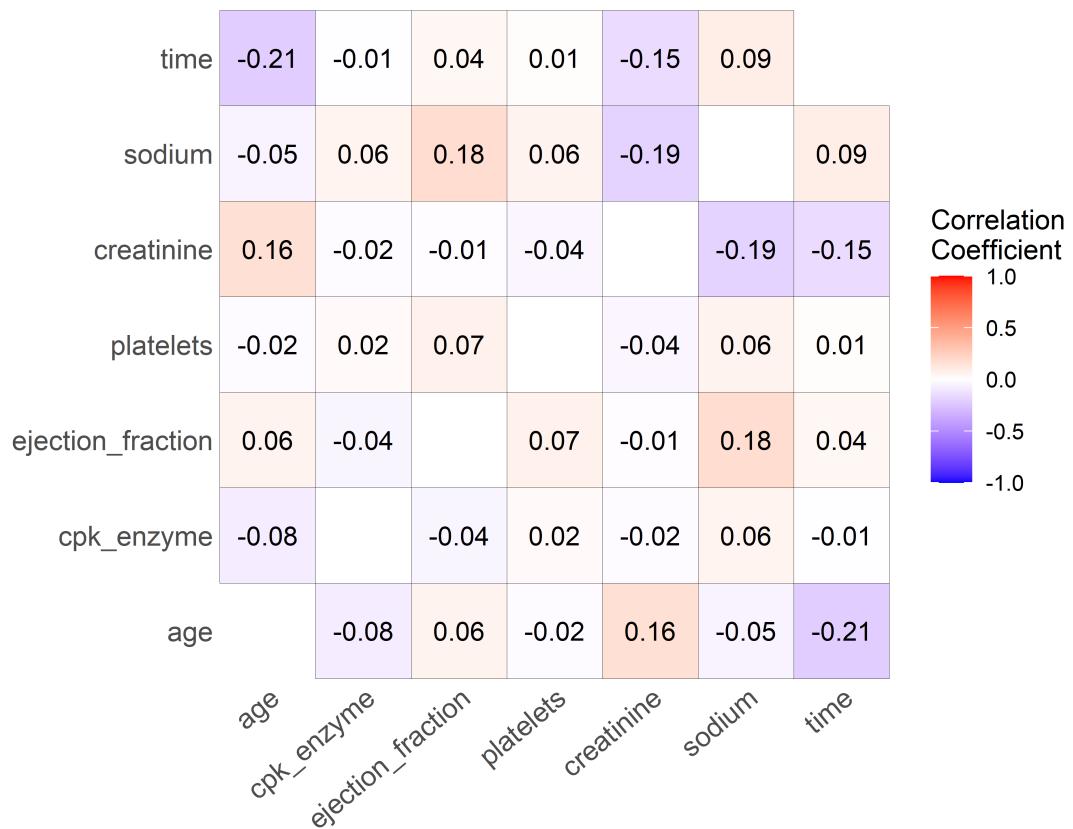
Correlation Analysis

Correlation Coefficient Matrix

first variable	second variable						
	age	cpk_enzyme	ejection_fraction	platelets	creatinine	sodium	time
age	NA	-0.076	0.061	-0.016	0.163	-0.053	-0.215
cpk_enzyme	-0.076	NA	-0.044	0.024	-0.016	0.060	-0.009
ejection_fraction	0.061	-0.044	NA	0.072	-0.011	0.176	0.042
platelets	-0.016	0.024	0.072	NA	-0.041	0.062	0.011
creatinine	0.163	-0.016	-0.011	-0.041	NA	-0.189	-0.149
sodium	-0.053	0.060	0.176	0.062	-0.189	NA	0.088
time	-0.215	-0.009	0.042	0.011	-0.149	0.088	NA

Table 9: Matrix table of correlation coefficient

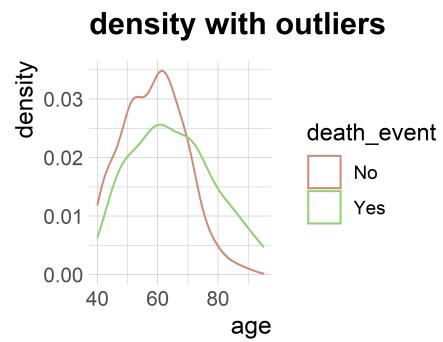
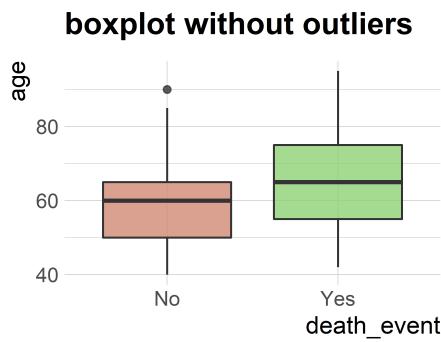
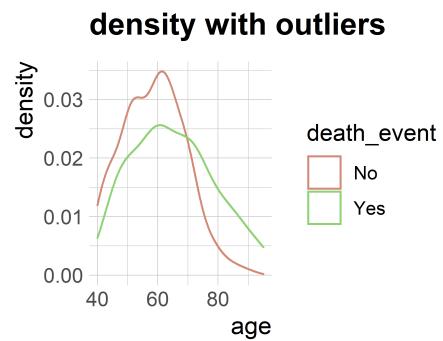
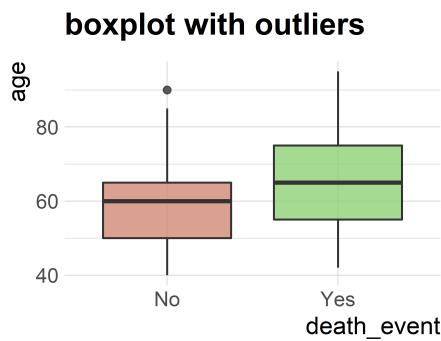
Correlation Plot



Target based Analysis

Grouped Numerical Variables

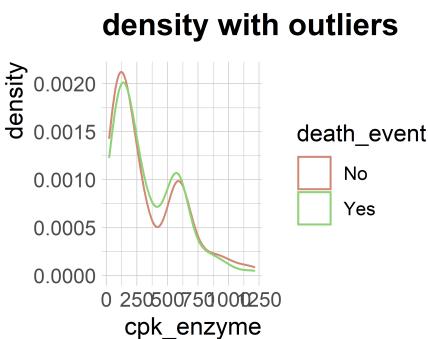
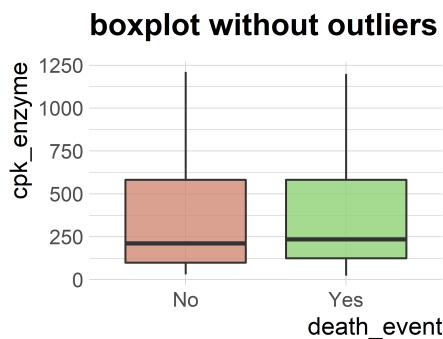
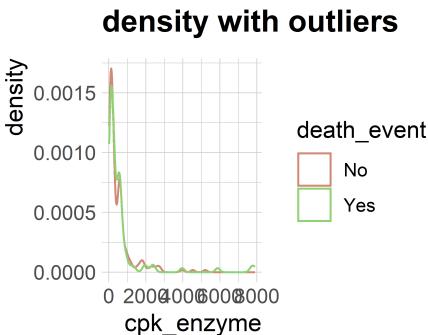
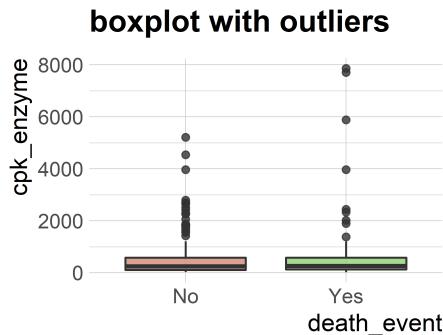
age



	death_event		
	No	Yes	<Total>
N	192	87	279
Missing	11	9	20
Mean	58.60938	65.11494	60.63799
Standard Deviation	10.45101	13.55044	11.87259
IQR	15	20	19
Skewness	0.2164947	0.3204182	0.4438566
Kurtosis	-0.3353744	-0.6786882	-0.1043259
Min	40	42	40
1%	40.00	44.58	40.00
5%	42	45	42
Q1	50	55	51
Median	60	65	60
Q3	65	75	70
95%	75.0	89.1	82.0
99%	85.00	95.00	90.88
Max	90	95	95

Table 10: Descriptive statistics with levels of target variable

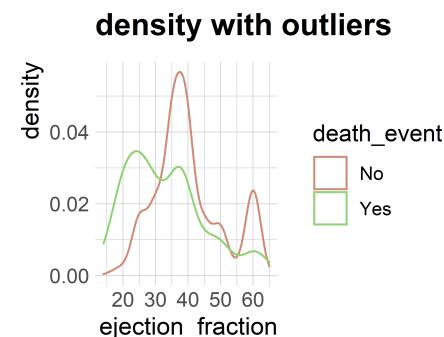
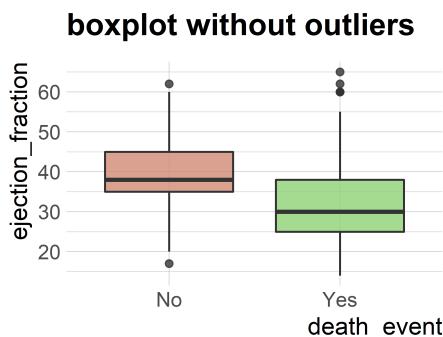
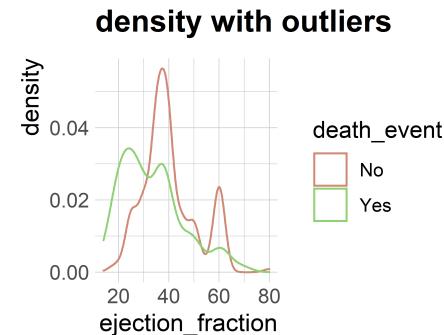
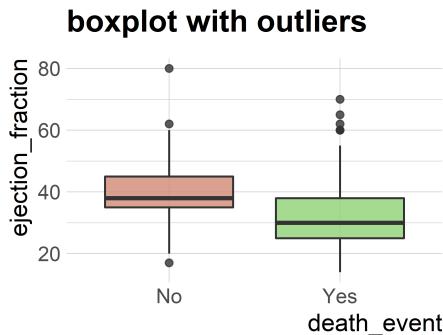
cpk_enzyme



	No	Yes	<Total>
N	203	96	299
Missing	0	0	0
Mean	540.0542	670.1979	581.8395
Standard Deviation	753.7996	1316.5806	970.2879
IQR	473.00	453.25	465.50
Skewness	3.194872	4.257468	4.463110
Kurtosis	12.91090	19.30083	25.14905
Min	30	23	23
1%	52.02	45.80	47.00
5%	59.0	64.5	59.0
Q1	109.00	128.75	116.50
Median	245	259	250
Q3	582	582	582
95%	2041.6	2361.0	2263.0
99%	3942.56	7709.95	5222.46
Max	5209	7861	7861

Table 10: Descriptive statistics with levels of target variable

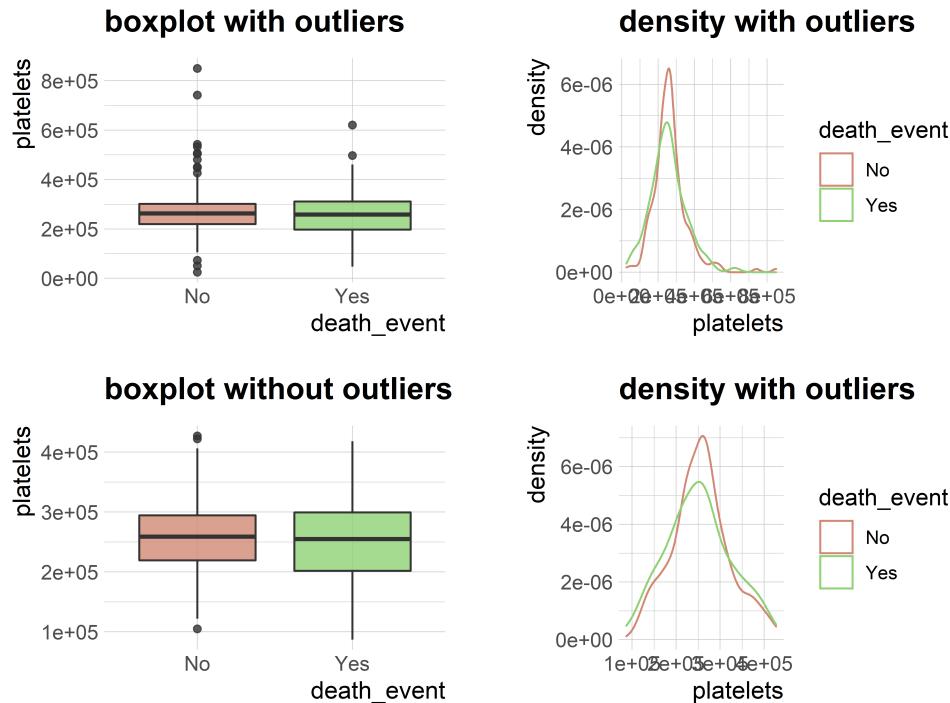
ejection_fraction



	death_event		
	No	Yes	<Total>
N	203	96	299
Missing	0	0	0
Mean	40.26601	33.46875	38.08361
Standard Deviation	10.85996	12.52530	11.83484
IQR	10	13	15
Skewness	0.7022167	0.8222937	0.5553828
Kurtosis	0.22572291	0.20578061	0.04140936
Min	17	14	14
1%	20.10	14.95	16.96
5%	25	20	20
Q1	35	25	30
Median	38	30	38
Q3	45	38	45
95%	60	60	60
99%	60.00	65.25	62.06
Max	80	70	80

Table 10: Descriptive statistics with levels of target variable

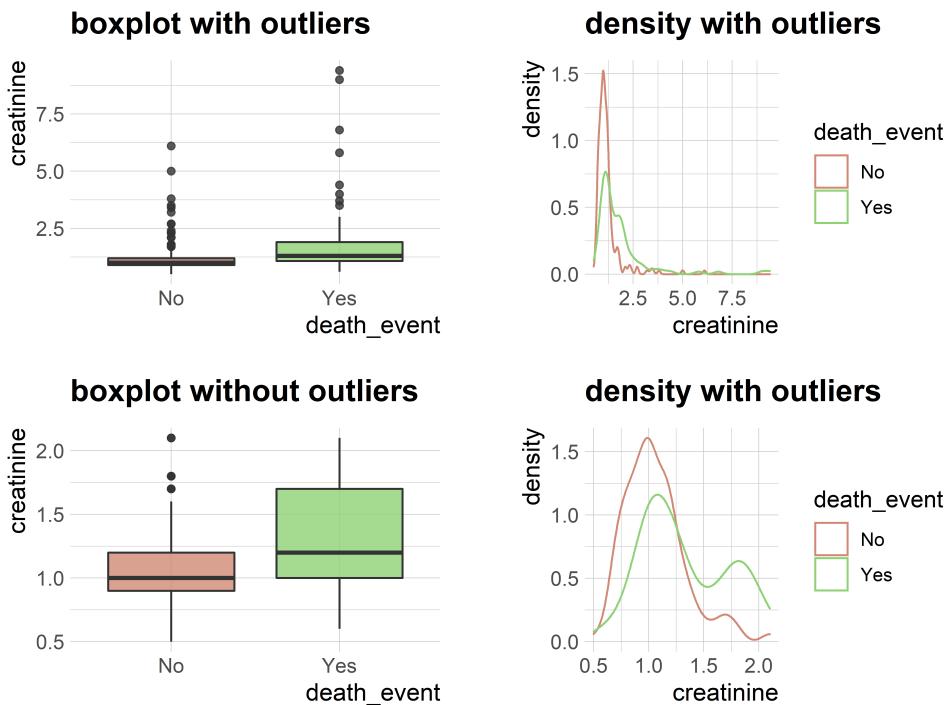
platelets



	death_event		
	No	Yes	<Total>
N	203	96	299
Missing	0	0	0
Mean	266657.5	256381.0	263358.0
Standard Deviation	97531.20	98525.68	97804.24
IQR	82500	113500	91000
Skewness	1.8842727	0.6252768	1.4623208
Kurtosis	8.560857	1.369982	6.209255
Min	25100	47000	25100
1%	73640	61250	61780
5%	140100	111000	131800
Q1	219500	197500	212500
Median	263000	258500	262000
Q3	302000	311000	303500
95%	420400	426250	422500
99%	542800	503200	544560
Max	850000	621000	850000

Table 10: Descriptive statistics with levels of target variable

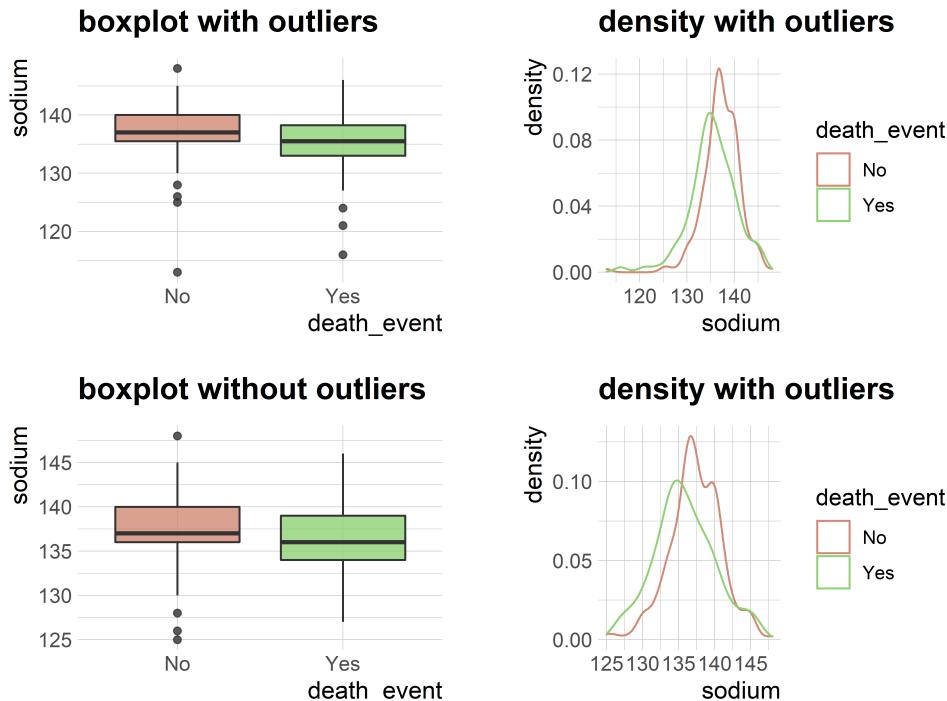
creatinine



	death_event		
	No	Yes	<Total>
N	203	96	299
Missing	0	0	0
Mean	1.184877	1.835833	1.393880
Standard Deviation	0.6540827	1.4685615	1.0345101
IQR	0.300	0.825	0.500
Skewness	4.170620	3.404452	4.455996
Kurtosis	22.91954	13.48084	25.82824
Min	0.5	0.6	0.5
1%	0.602	0.600	0.600
5%	0.7	0.9	0.7
Q1	0.900	1.075	0.900
Median	1.0	1.3	1.1
Q3	1.2	1.9	1.4
95%	2.28	4.10	3.00
99%	3.794	9.020	6.114
Max	6.1	9.4	9.4

Table 10: Descriptive statistics with levels of target variable

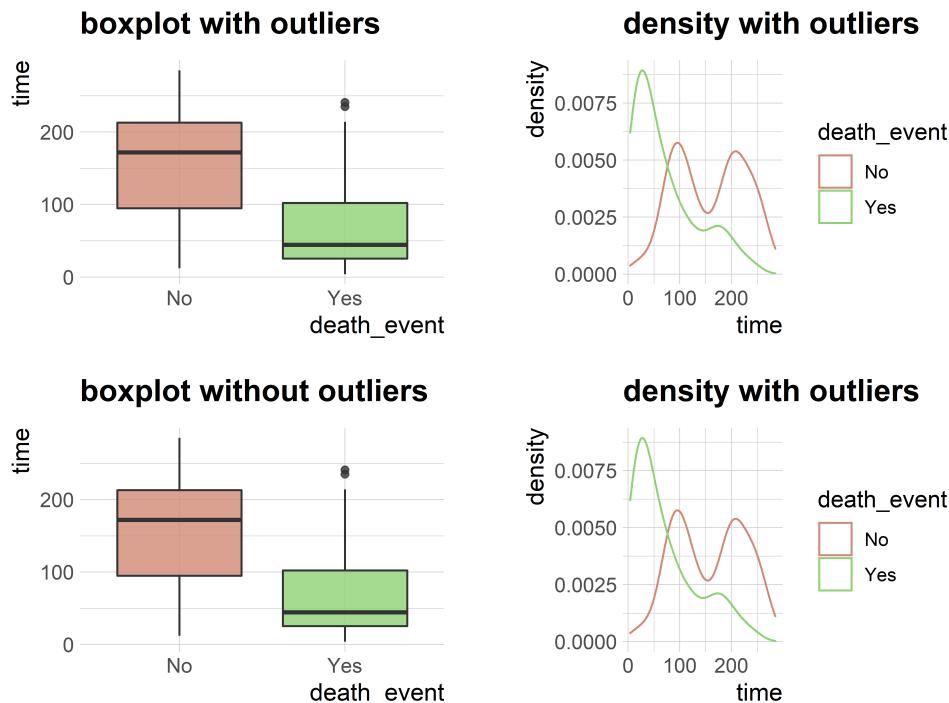
sodium



	death_event		
	No	Yes	<Total>
N	203	96	299
Missing	0	0	0
Mean	137.2167	135.3750	136.6254
Standard Deviation	3.982923	5.001579	4.412477
IQR	4.50	5.25	6.00
Skewness	-1.2189038	-0.6765943	-1.0481360
Kurtosis	6.487565	2.080837	4.119712
Min	113	116	113
1%	126.04	120.75	123.94
5%	131	127	130
Q1	135.5	133.0	134.0
Median	137.0	135.5	137.0
Q3	140.00	138.25	140.00
95%	143	144	144
99%	145.00	145.05	145.00
Max	148	146	148

Table 10: Descriptive statistics with levels of target variable

time

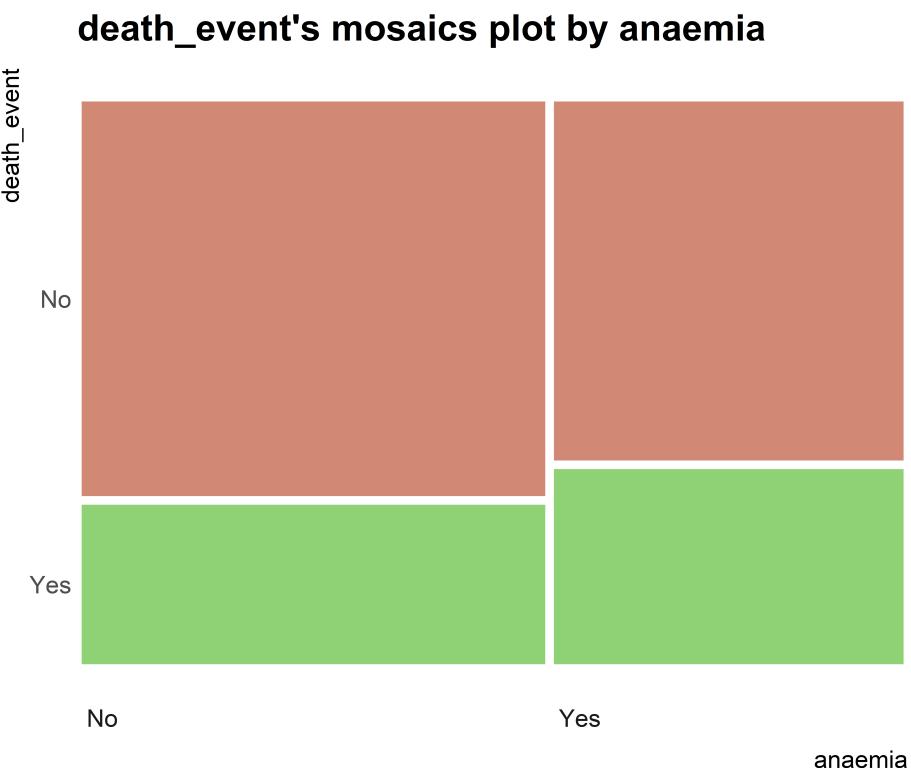


	death_event		
	No	Yes	<Total>
N	203	96	299
Missing	0	0	0
Mean	158.33990	70.88542	130.26087
Standard Deviation	67.74287	62.37828	77.61421
IQR	118.00	76.75	130.00
Skewness	-0.04818204	1.06481945	0.12780265
Kurtosis	-1.23212265	0.09318104	-1.21204797
Min	12	4	4
1%	22.14	5.90	7.00
5%	68.3	8.0	12.9
Q1	95.0	25.5	73.0
Median	172.0	44.5	115.0
Q3	213.00	102.25	203.00
95%	255.4	196.5	250.0
99%	277.86	235.30	271.14
Max	285	241	285

Table 10: Descriptive statistics with levels of target variable

Grouped Categorical Variables

anaemia

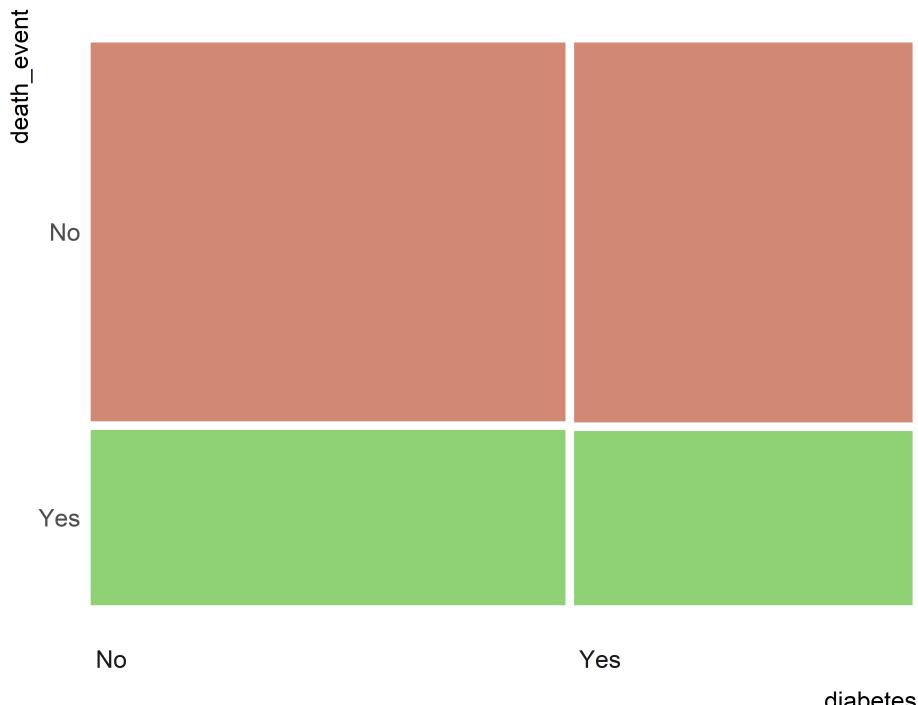


anaemia	death_event		<Total>
	No	Yes	
No	120	50	170
Yes	83	46	129
<Total>	203	96	299

Table 11: Contingency table with target variable

diabetes

death_event's mosaics plot by diabetes

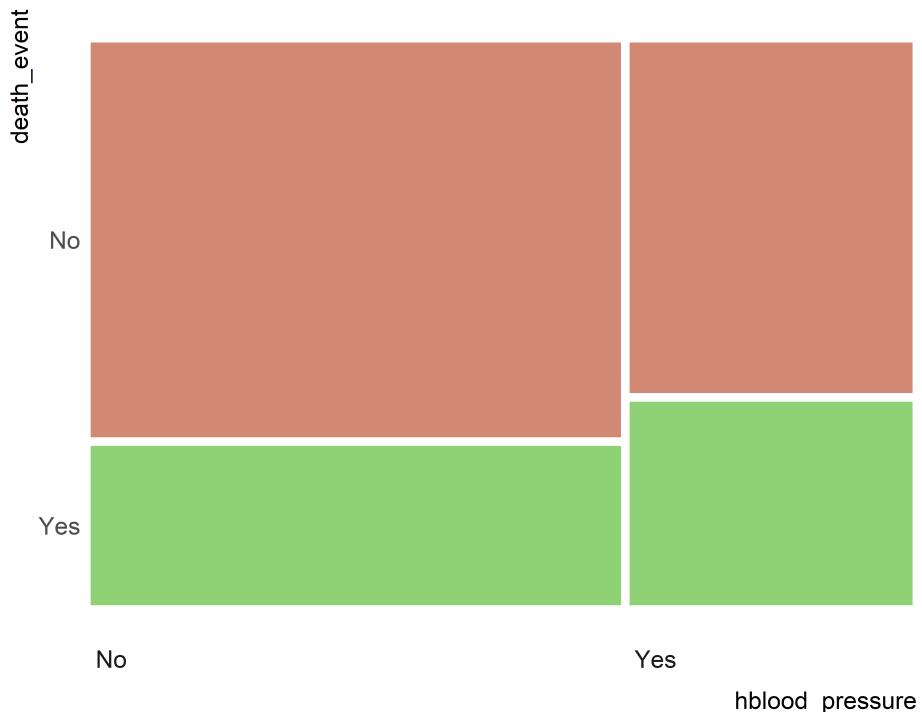


diabetes	death_event		
	No	Yes	<Total>
No	118	56	174
Yes	85	40	125
<Total>	203	96	299

Table 11: Contingency table with target variable

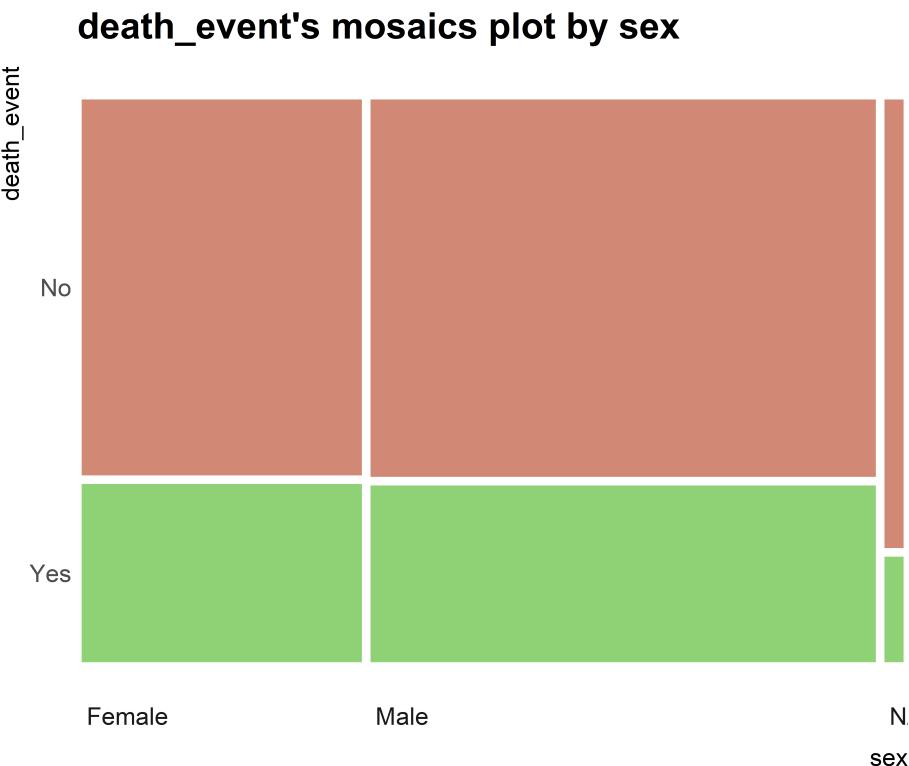
hblood_pressure

death_event's mosaics plot by hblood_pressure



hblood_pressure	death_event			<Total>
	No	Yes	<Total>	
No	137	57	194	
Yes	66	39	105	
<Total>	203	96	299	

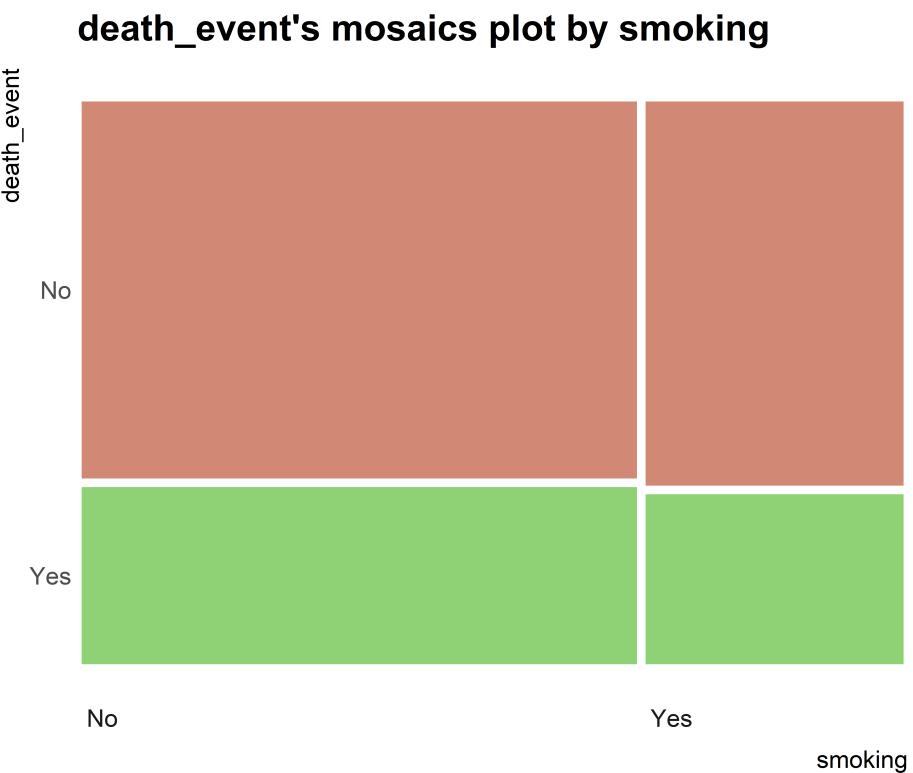
Table 11: Contingency table with target variable

sex

sex	death_event		<Total>
	No	Yes	
Female	70	34	104
Male	125	60	185
<Total>	203	96	299
NA	8	2	10

Table 11: Contingency table with target variable

smoking



smoking	death_event		<Total>
	No	Yes	
No	137	66	203
Yes	66	30	96
<Total>	203	96	299

Table 11: Contingency table with target variable

Grouped Correlation

death_event : Yes

Correlation Coefficient Matrix

first variable	second variable						
	age	cpk_enzyme	ejection_fraction	platelets	creatinine	sodium	time
age	NA	-0.145	0.219	0.072	0.072	0.026	-0.175
cpk_enzyme	-0.145	NA	0.022	0.079	-0.033	0.149	0.024
ejection_fraction	0.219	0.022	NA	0.017	0.230	0.173	-0.122
platelets	0.072	0.079	0.017	NA	-0.029	0.141	-0.093
creatinine	0.072	-0.033	0.230	-0.029	NA	-0.094	0.016
sodium	0.026	0.149	0.173	0.141	-0.094	NA	0.021
time	-0.175	0.024	-0.122	-0.093	0.016	0.021	NA

Table 12: Matrix table of correlation coefficient

death_event : Yes

Correlation Plot

death_event : No

Correlation Coefficient Matrix

first variable	second variable						
	age	cpk_enzyme	ejection_fraction	platelets	creatinine	sodium	time
age	NA	-0.044	0.079	-0.059	0.137	-0.016	-0.054
cpk_enzyme	-0.044	NA	-0.077	-0.013	-0.043	-0.002	0.034
ejection_fraction	0.079	-0.077	NA	0.086	-0.107	0.102	-0.123
platelets	-0.059	-0.013	0.086	NA	-0.031	0.002	0.015
creatinine	0.137	-0.043	-0.107	-0.031	NA	-0.216	0.000
sodium	-0.016	-0.002	0.102	0.002	-0.216	NA	-0.040
time	-0.054	0.034	-0.123	0.015	0.000	-0.040	NA

Table 12: Matrix table of correlation coefficient

death_event : No

Correlation Plot