



EDA REPORT

SEPSIS

Report Overview

This report was created for the EDA of *sepsis* 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	7
Bivariate Analysis	14
Compare Numerical Variables	14
Compare Categorical Variables	30
Multivariate Analysis	31
Correlation Analysis	31
Correlation Coefficient Matrix	31
Correlation Plot	32

Overview

Data Structures

division	metrics	value	division	metrics	value
size	observations	455	data type	numerics	3
size	variables	9	data type	integers	3
size	values	4,095	data type	factors/ordered	0
size	memory size (KB)	0	data type	characters	3
duplicated	duplicate observation	0	data type	Dates	0
missing	complete observation	143	data type	POSIXcts	0
missing	missing observation	312	data type	others	0
missing	missing variables	3			
missing	missing values	344			

Table 1: Data structures and types

Job Informations

division	metrics	value
dataset	dataset	sepsis
dataset	dataset type	data.frame
dataset	target	not defied
job	samples	455 / 455 (100%)
job	created	2024-06-01 14:43:15.57338
job	created by	dlookr

Table 2: Job informations

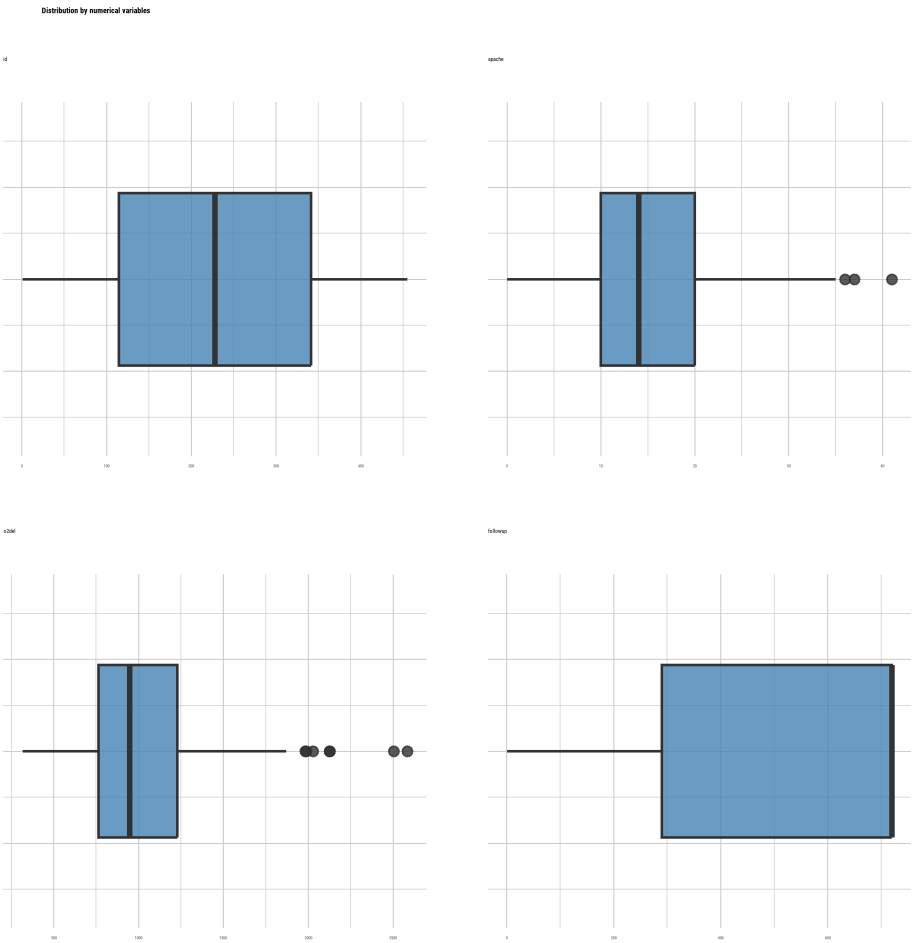
Univariate Analysis

Descriptive Statistics

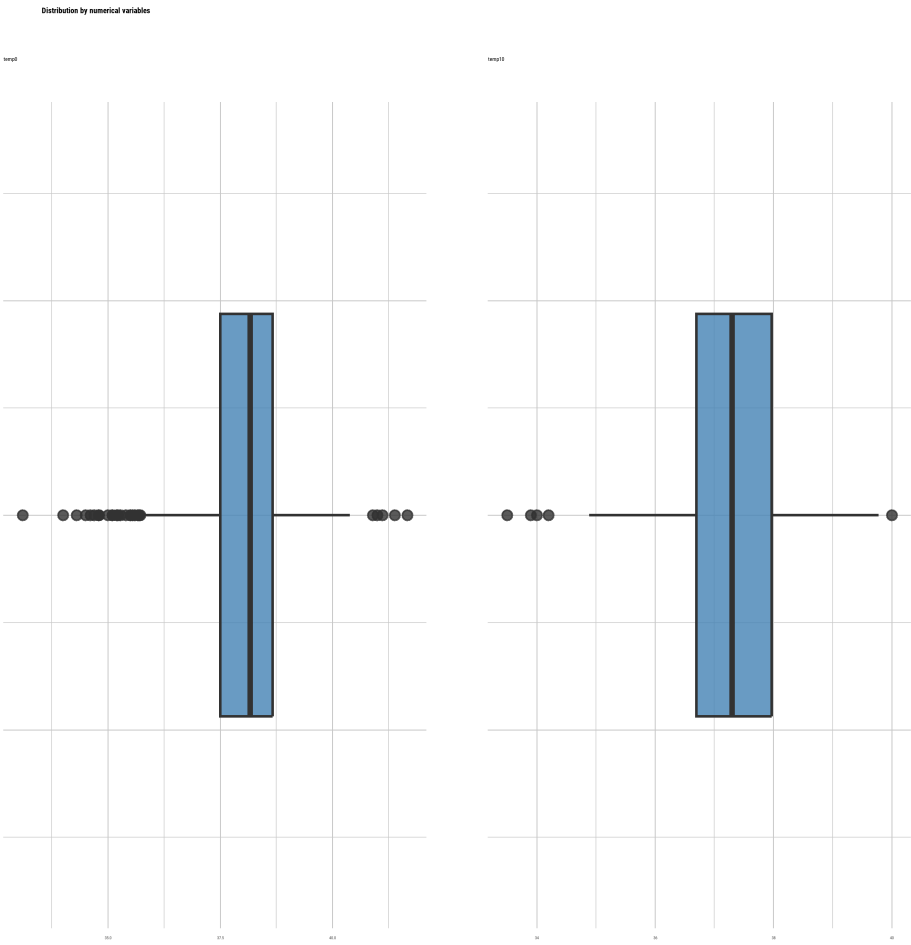
Numerical Variables

variables	missing	mean	sd	min	Q1	median	Q3	max
id	0	228.00	131.49	1.00	114.50	228.00	341.50	455.00
apache	1	15.33	7.09	0.00	10.00	14.00	20.00	41.00
o2del	287	1,023.82	409.44	316.88	765.22	947.20	1,229.57	2,584.34
followup	0	532.58	266.79	1.00	290.00	720.00	720.00	720.00
temp0	0	38.01	1.13	33.10	37.50	38.17	38.67	41.67
temp10	56	37.31	0.97	33.50	36.70	37.30	37.97	40.00

Table 3: Descriptive statistics of numerical variables



variables	data types	distinct	skewness	kurtosis	zero	negative	outlier
id	integer	455	0.00	-1.20	0	0	0
apache	integer	39	0.57	0.01	1	0	3
o2del	numeric	169	1.09	1.70	0	0	7
followup	integer	148	-0.93	-0.85	0	0	0



variables	data types	distinct	skewness	kurtosis	zero	negative	outlier
temp0	numeric	122	-0.78	2.08	0	0	28
temp10	numeric	102	-0.19	0.88	0	0	5

Categorical Variables

variables	levels	observations	frequency	frequency(%)	rank
treat	Placebo	455	231	50.77	1
treat	Ibuprofen	455	224	49.23	2
race	White	455	293	64.40	1
race	African American	455	130	28.57	2
race	Other	455	32	7.03	3
fate	Alive	455	279	61.32	1
fate	Dead	455	176	38.68	2

Table 4: Top rank levels of categorical variables

The number of categorical(factor/ordered) variables is 0.

Normality Test

described_variables	min	Q1	median	Q3	max	skewness	kurtosis	balance
id	1.0	114.5	228.0	341.5	455.0	0.0	-1.2	Balanced
apache	0.0	10.0	14.0	20.0	41.0	0.6	0.0	Balanced
o2del	316.9	765.2	947.2	1229.6	2584.3	1.1	1.7	Balanced
followup	1.0	290.0	720.0	720.0	720.0	-0.9	-0.9	Balanced
temp0	33.1	37.5	38.2	38.7	41.7	-0.8	2.1	Balanced
temp10	33.5	36.7	37.3	38.0	40.0	-0.2	0.9	Balanced

Table 5: Descriptive statistics of numerical variables

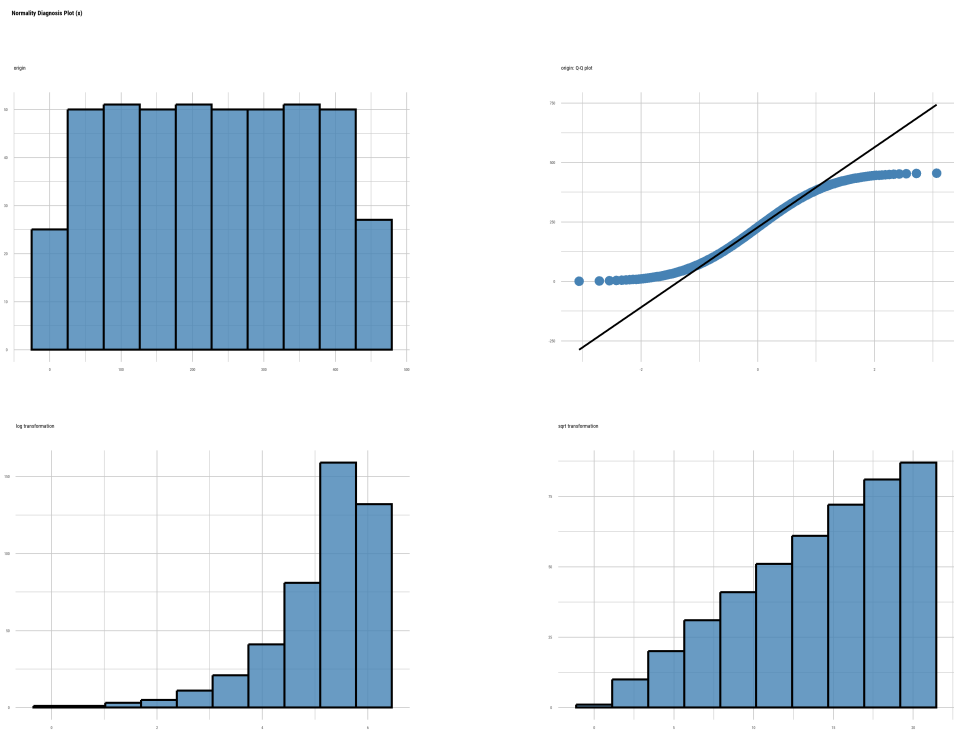
id

statistic	p_value	remark
0.95471	1.341e-10	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	0.0000	1.8000
log transformation	-1.7754	6.9132
sqrt transformation	-0.5586	2.3809

Table 6: skewness and kurtosis



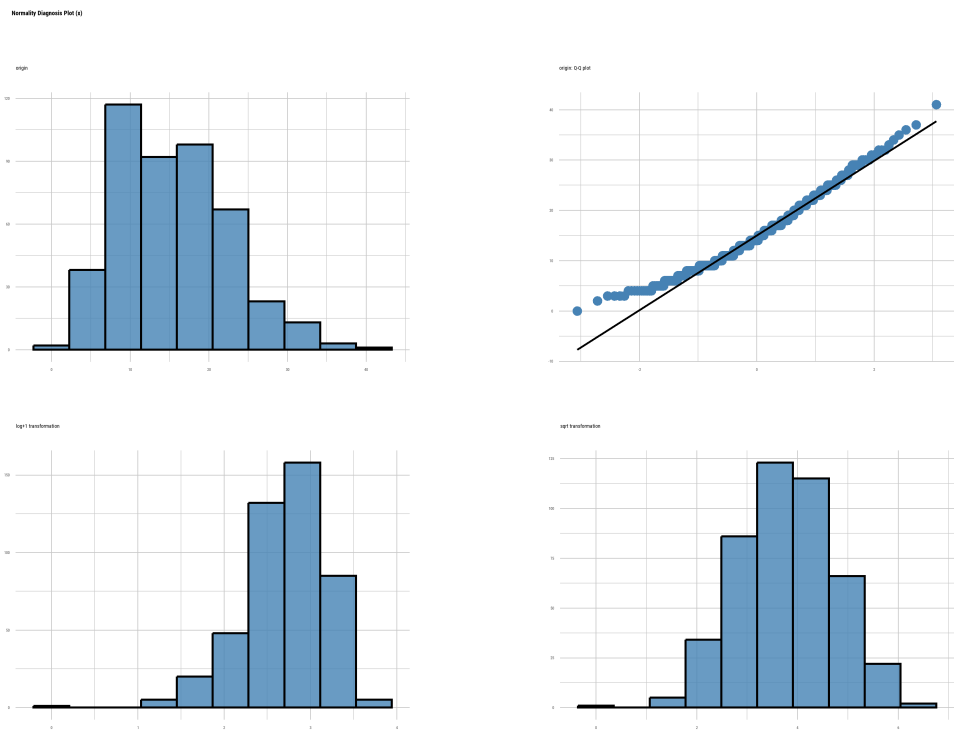
apache

statistic	p_value	remark
0.97356	2.5431e-07	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	0.5692	3.0009
log+1 transformation	-0.7955	4.7303
sqrt transformation	-0.0994	3.0324

Table 6: skewness and kurtosis



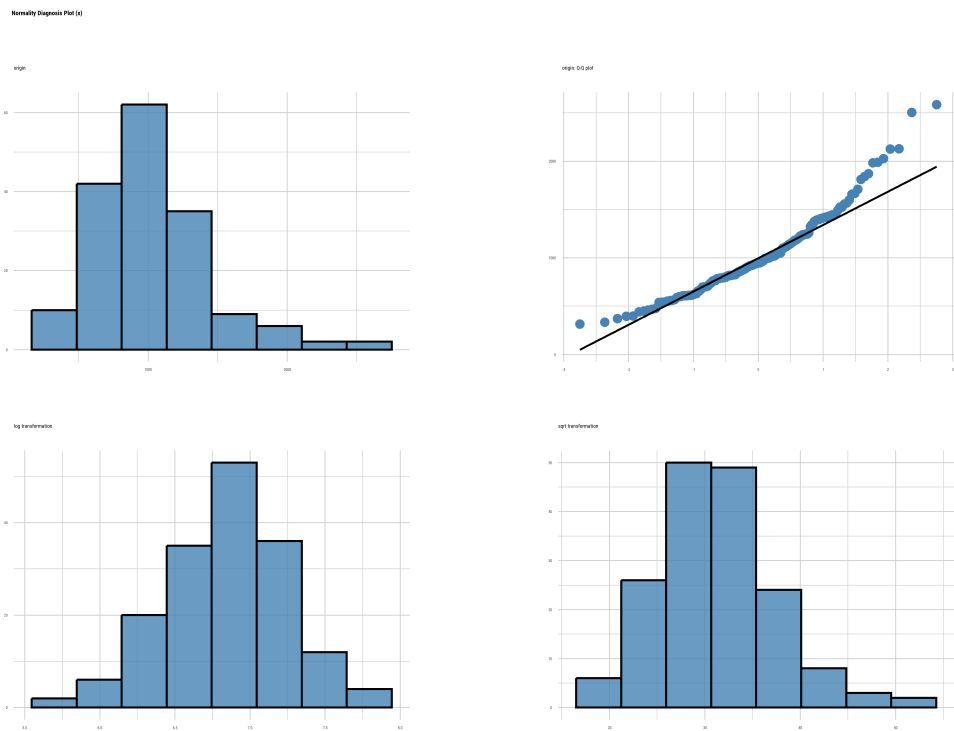
o2del

statistic	p_value	remark
0.93634	8.3998e-07	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	1.0801	4.6148
log transformation	-0.1382	3.1064
sqrt transformation	0.4705	3.3556

Table 6: skewness and kurtosis



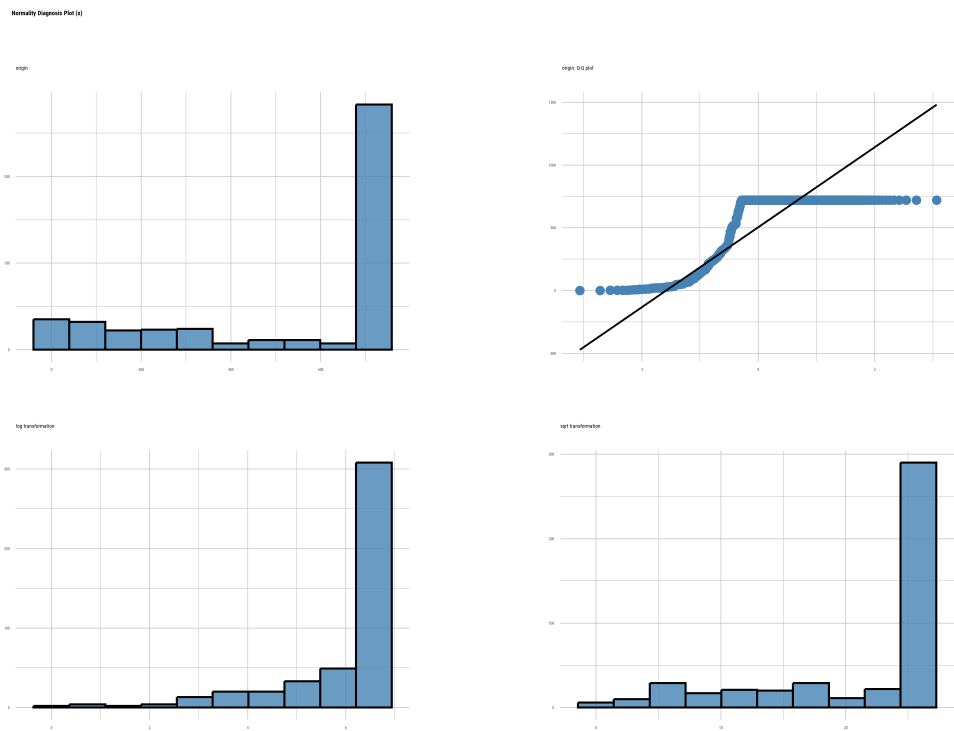
followup

statistic	p_value	remark
0.6911	5.0691e-28	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	-0.9312	2.1431
log transformation	-2.2538	8.1969
sqrt transformation	-1.2418	3.0854

Table 6: skewness and kurtosis



temp0

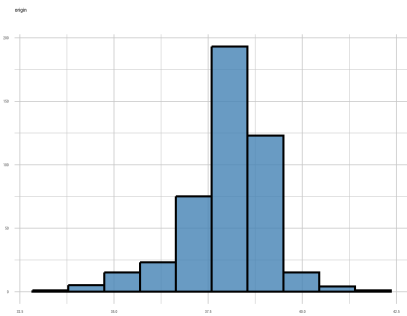
statistic	p_value	remark
0.94704	1.1132e-11	No sample

Table 6: Shapiro-Wilk normality test

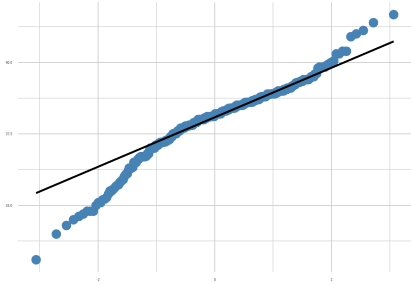
type	skewness	kurtosis
original	-0.7724	5.0448
log transformation	-0.9277	5.3254
sqrt transformation	-0.8500	5.1749

Table 6: skewness and kurtosis

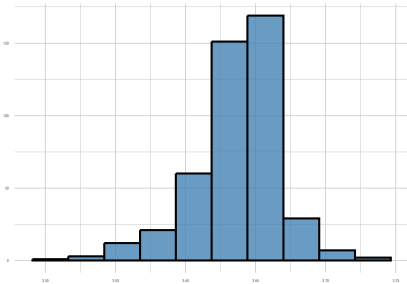
Normality Diagnostic Plot (x)



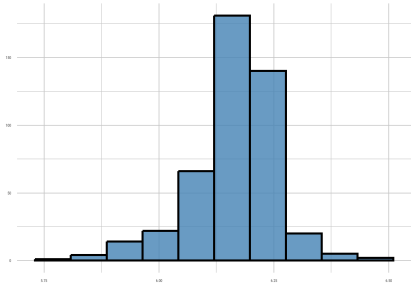
orign QQ plot



log transformation



sqrt transformation



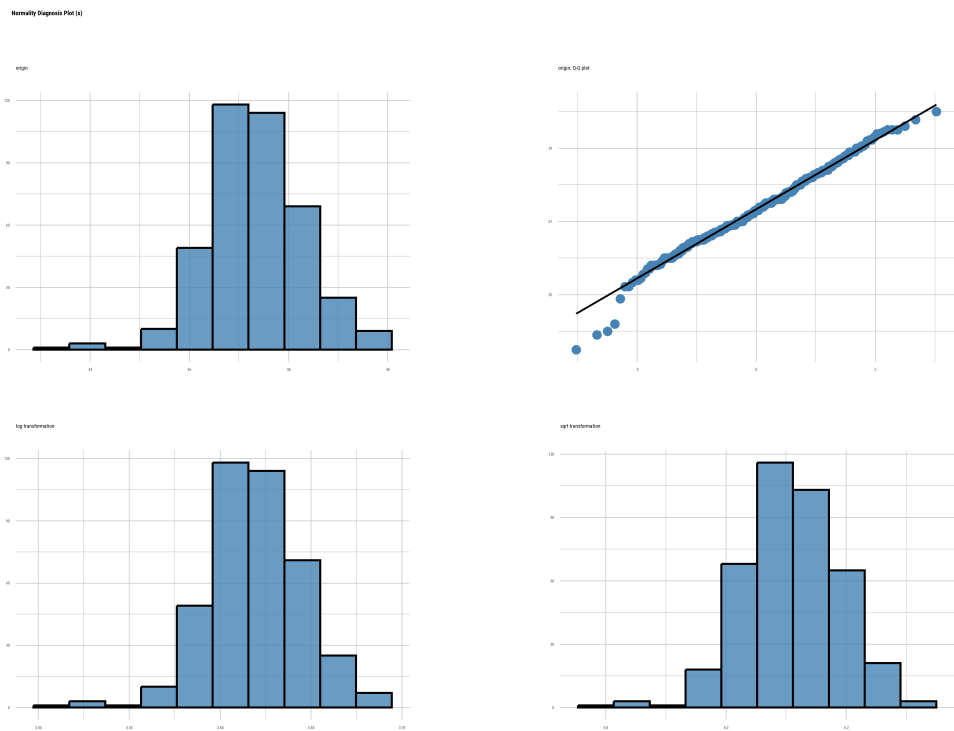
temp10

statistic	p_value	remark
0.9881	0.002412	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	-0.1864	3.8499
log transformation	-0.2995	4.0694
sqrt transformation	-0.2423	3.9528

Table 6: skewness and kurtosis



Bivariate Analysis

Compare Numerical Variables

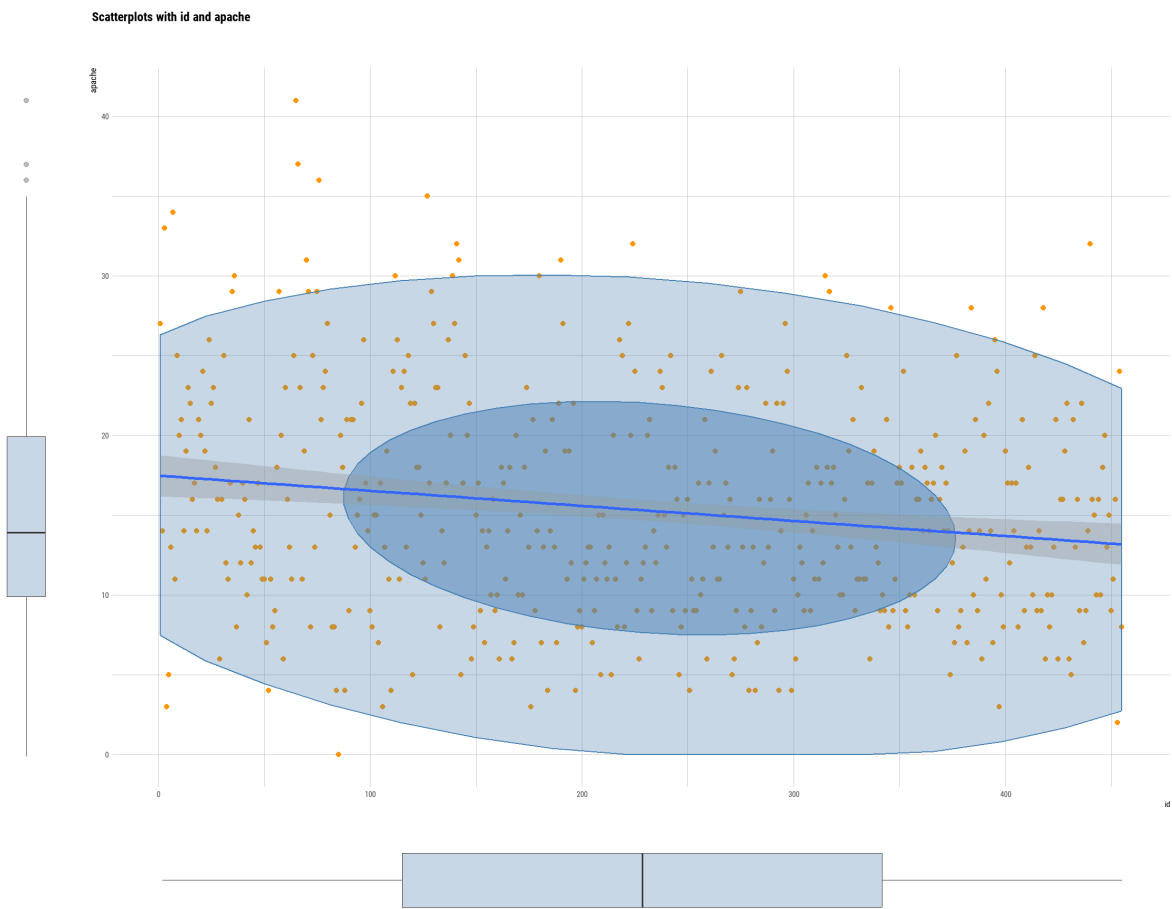
first variable	second variable	correlation coefficient
id	apache	-0.17533
id	o2del	-0.05117
id	followup	-0.03909
id	temp0	-0.05705
id	temp10	-0.09823
apache	o2del	-0.18148
apache	followup	-0.36922
apache	temp0	-0.02899
apache	temp10	-0.10602
o2del	followup	0.15496
o2del	temp0	0.14742
o2del	temp10	0.19034
followup	temp0	0.20451
followup	temp10	0.19510
temp0	temp10	0.39531

Table 7: Correlation coefficient

'id' vs 'apache'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
id	apache	0.030741	0.0285966	129.475	14.3356	0.0001736	1

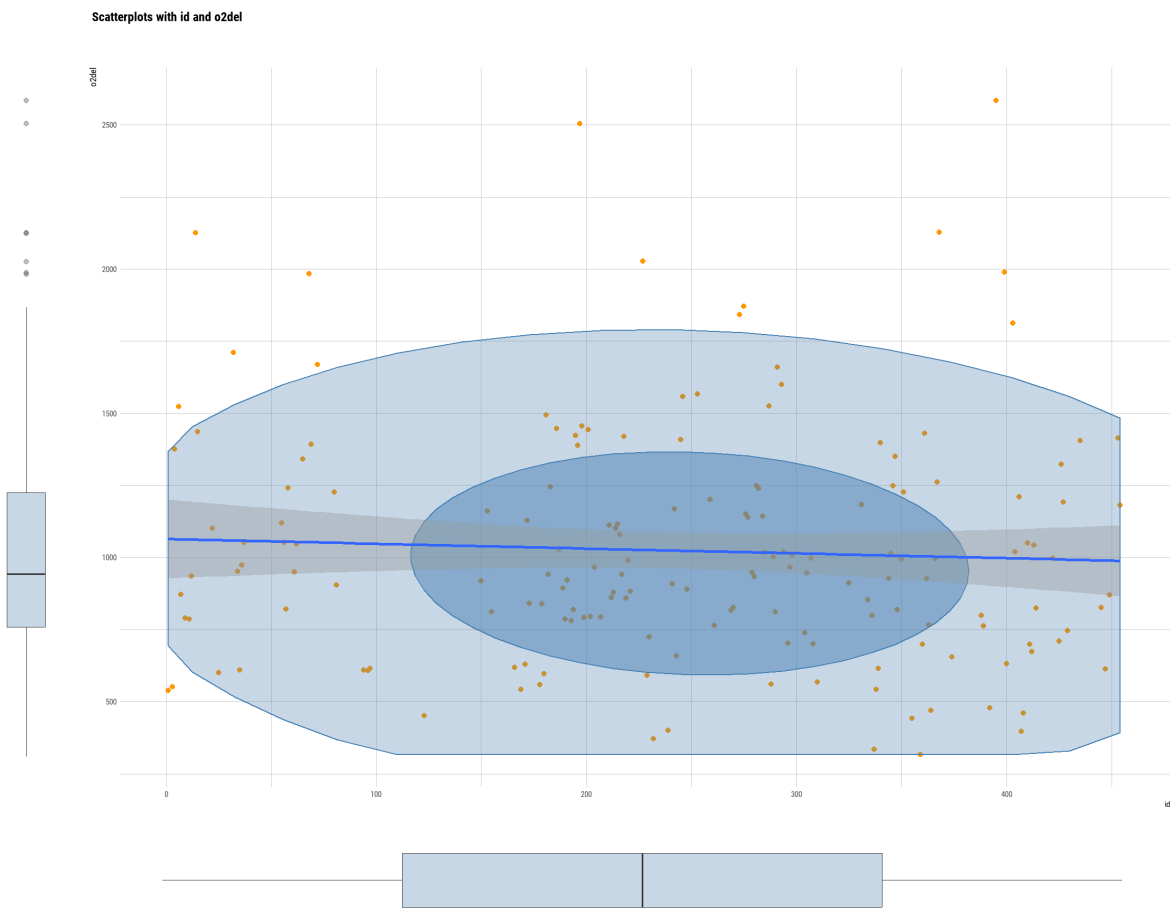
Table 7: Summary of linear model



‘id’ vs ‘o2del’

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
id	o2del	0.0026186	-0.0033897	125.8752	0.4358345	0.5100549	1

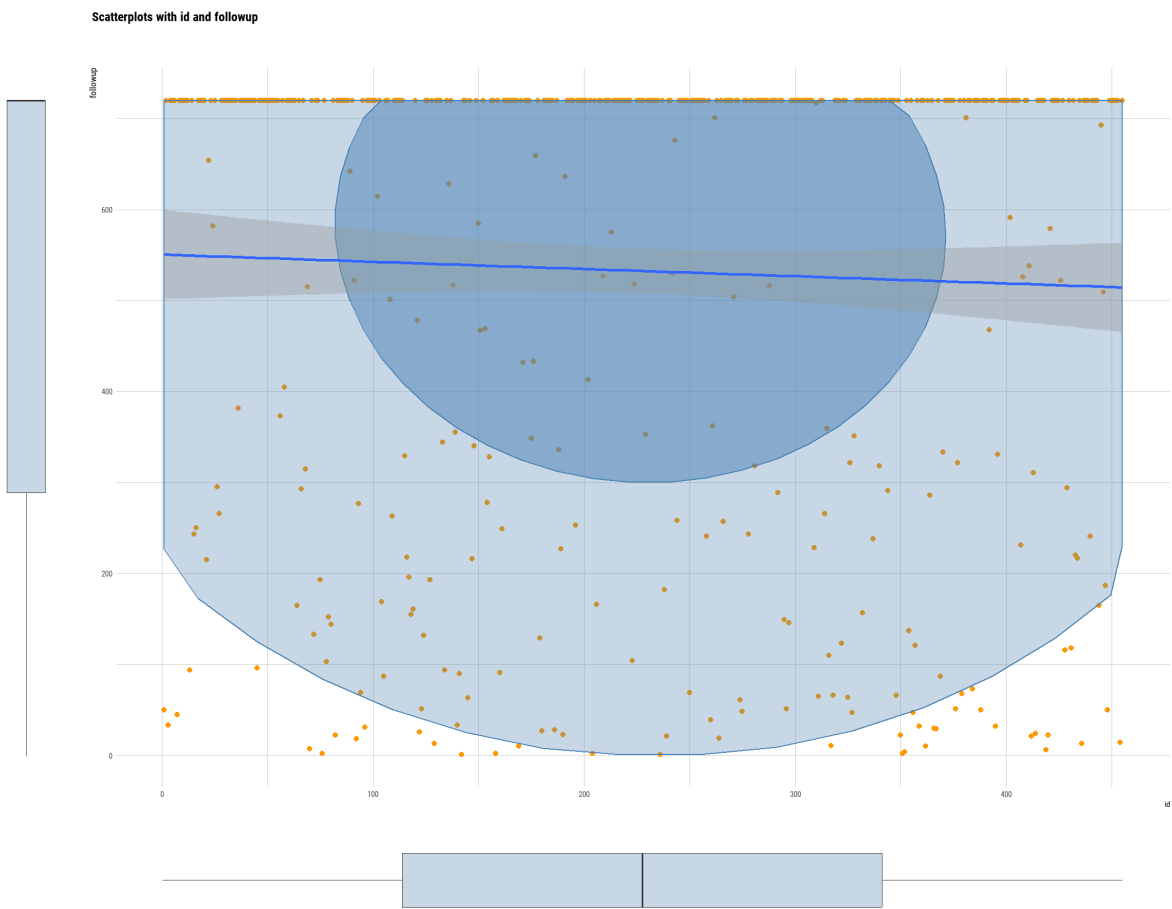
Table 7: Summary of linear model



'id' vs 'followup'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
id	followup	0.0015278	-0.0006763	131.5359	0.6931672	0.4055279	1

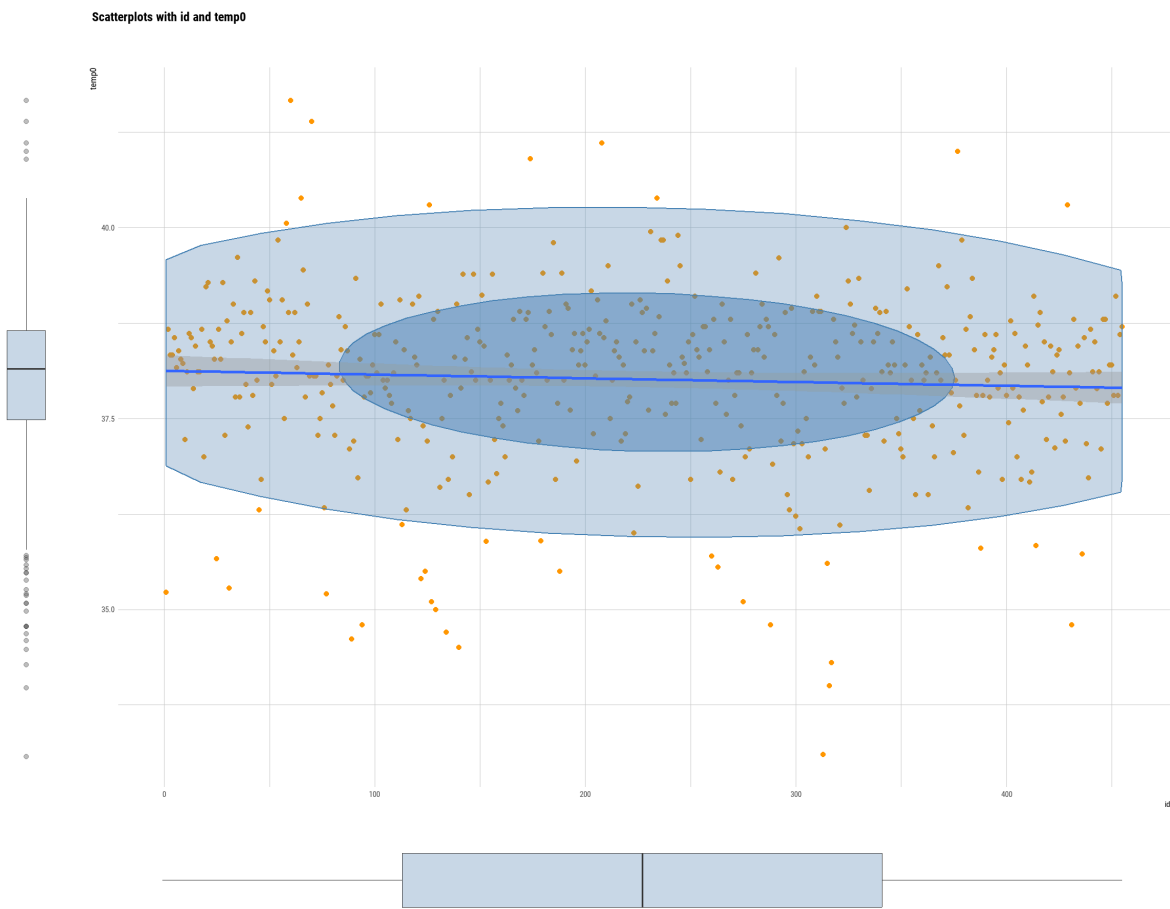
Table 7: Summary of linear model



'id' vs 'temp0'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
id	temp0	0.0032546	0.0010542	131.4221	1.479131	0.2245438	1

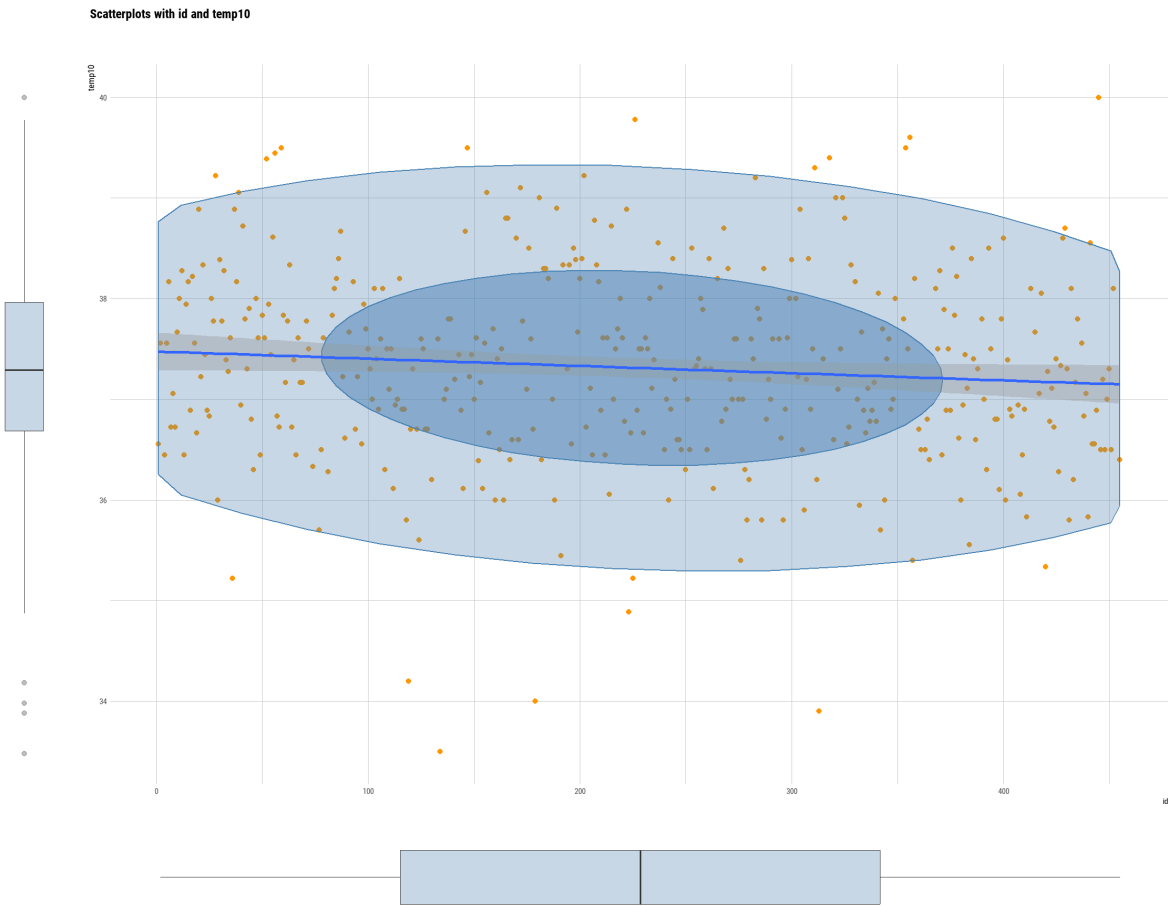
Table 7: Summary of linear model



'id' vs 'temp10'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
id	temp10	0.0096492	0.0071546	131.5787	3.868038	0.0499101	1

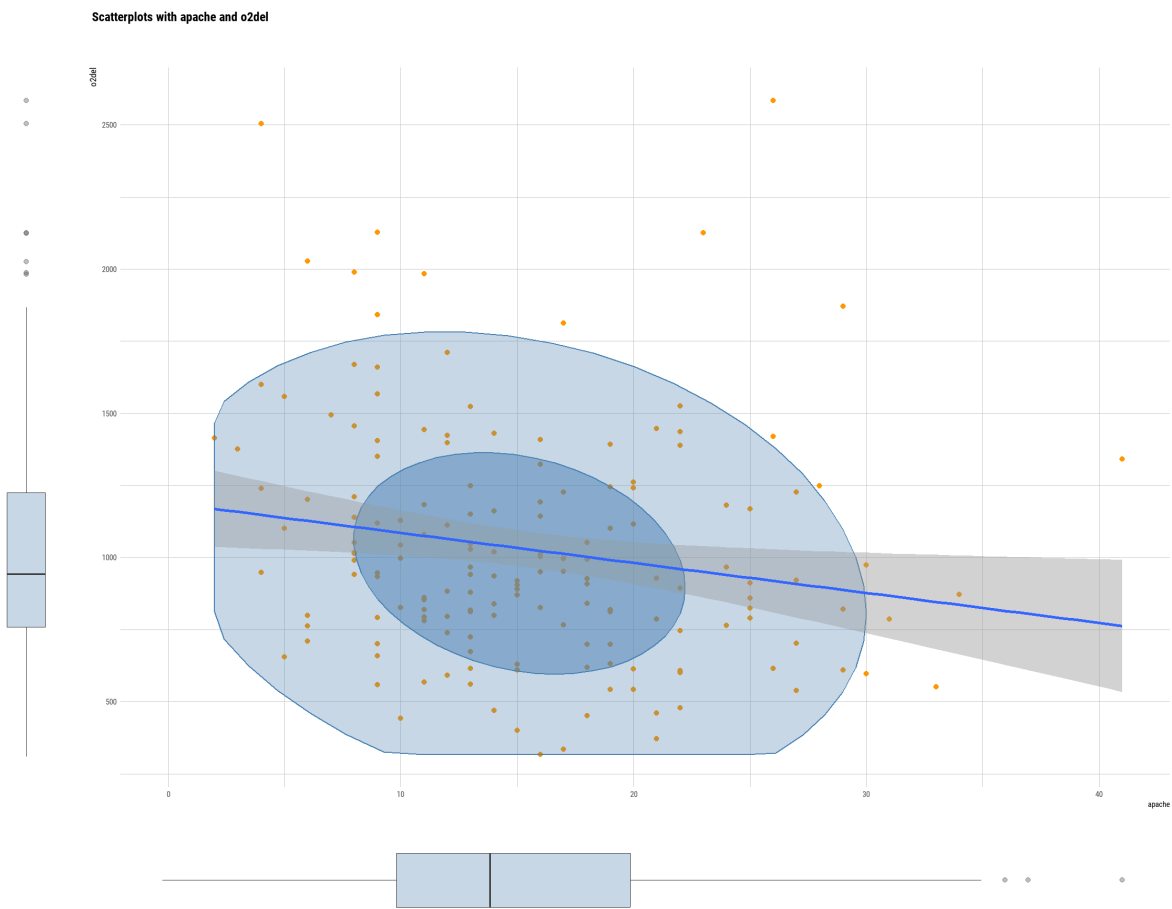
Table 7: Summary of linear model



'apache' vs 'o2del'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
apache	o2del	0.032934	0.027073	7.017089	5.619168	0.0189196	1

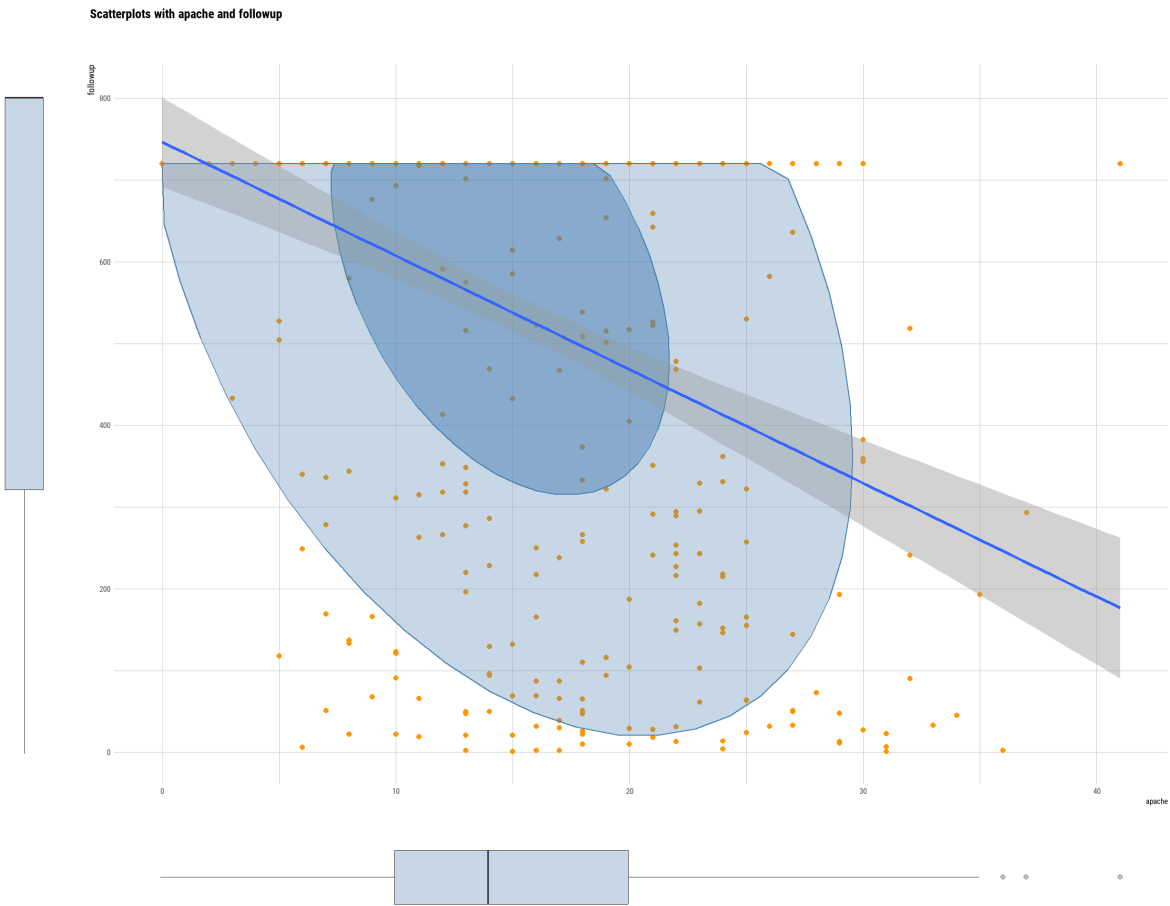
Table 7: Summary of linear model



'apache' vs 'followup'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
apache	followup	0.1363269	0.1344161	6.592393	71.34616	0	1

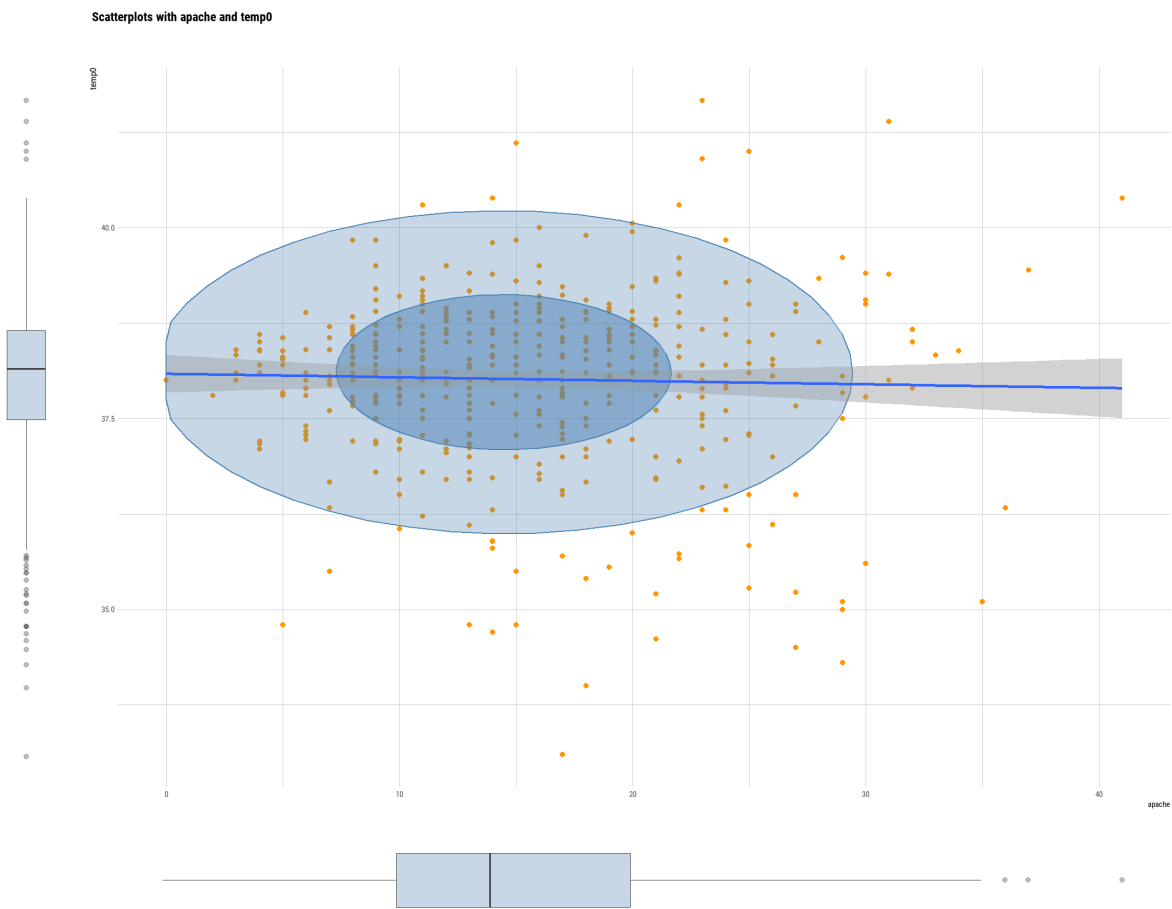
Table 7: Summary of linear model



'apache' vs 'temp0'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
apache	temp0	0.0008405	-0.00137	7.090646	0.3802328	0.5377891	1

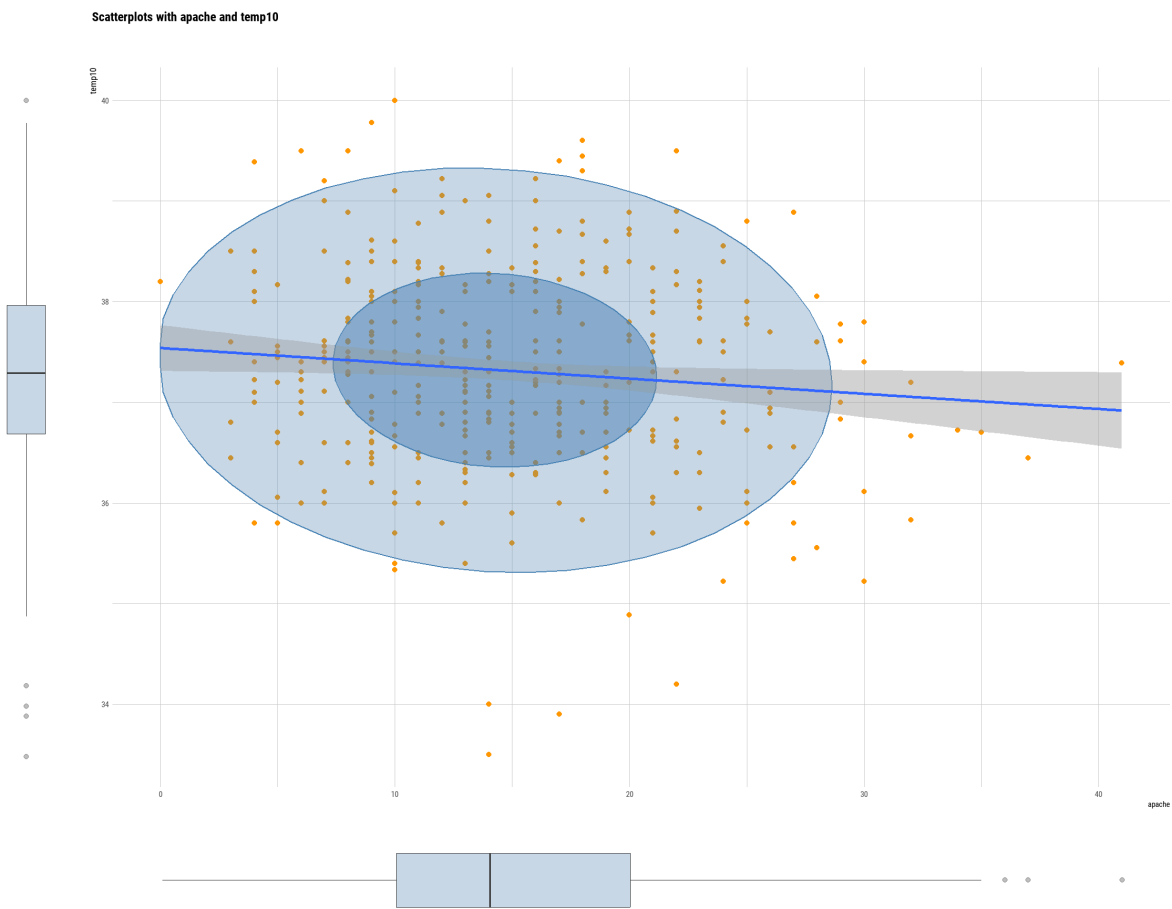
Table 7: Summary of linear model



'apache' vs 'temp10'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
apache	temp10	0.0112392	0.0087424	6.7501	4.501333	0.034489	1

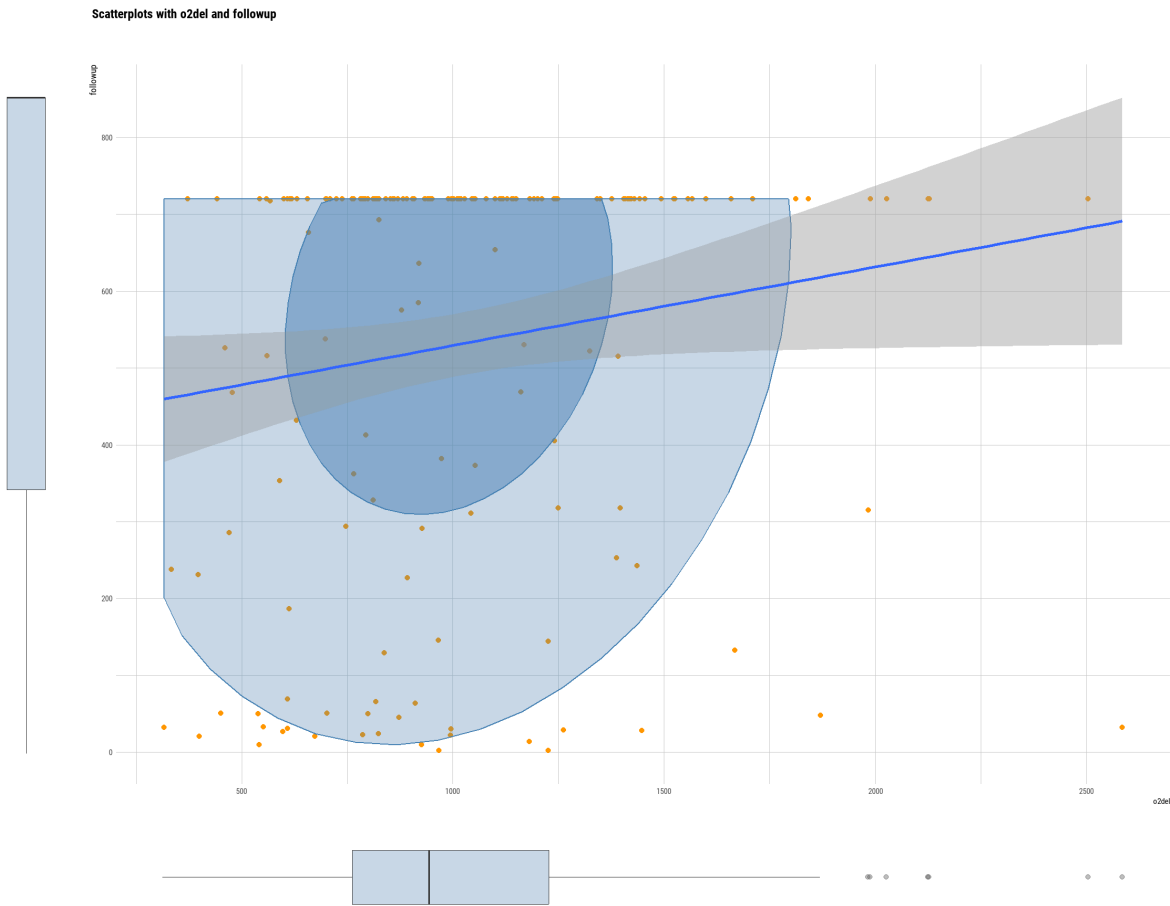
Table 7: Summary of linear model



'o2del' vs 'followup'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
o2del	followup	0.0240113	0.0181319	405.7136	4.083937	0.0448998	1

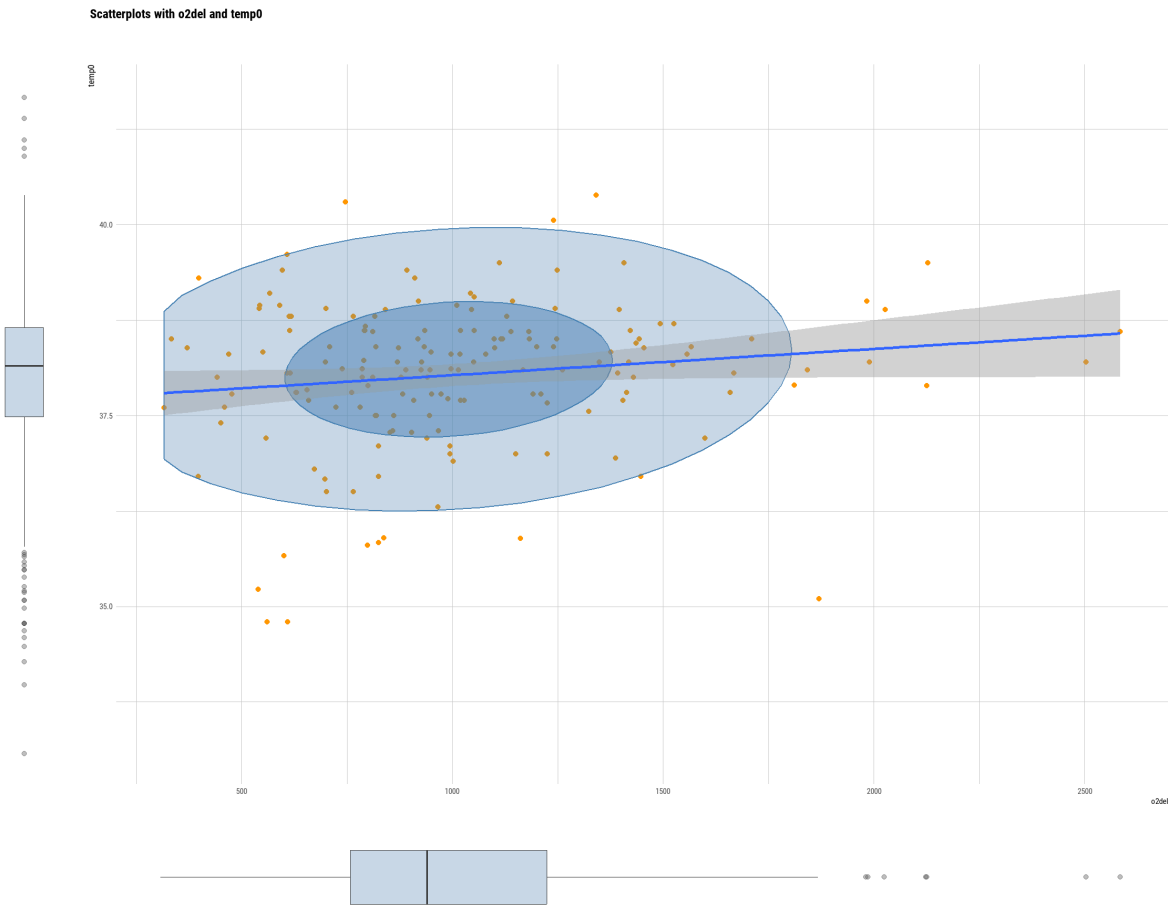
Table 7: Summary of linear model



'o2del' vs 'temp0'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
o2del	temp0	0.0217319	0.0158387	406.1871	3.687636	0.0565316	1

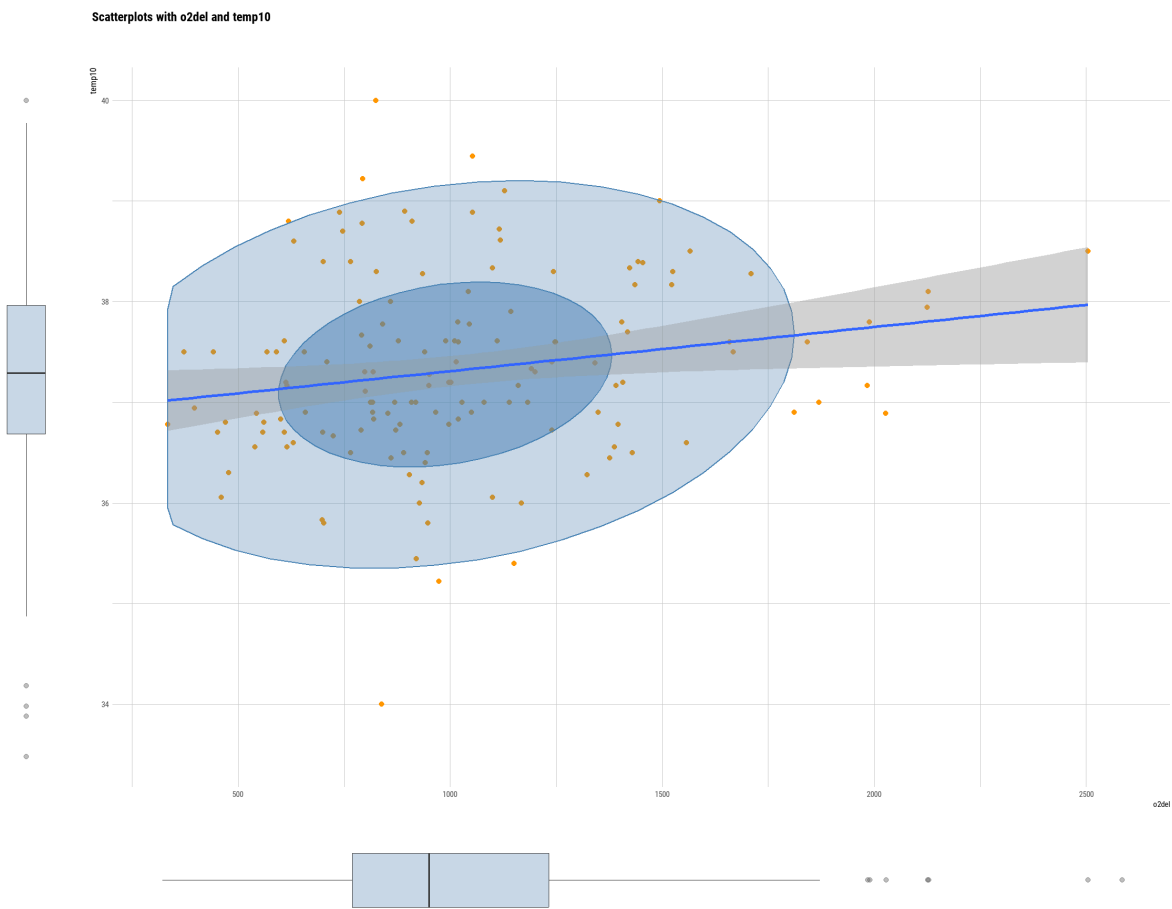
Table 7: Summary of linear model



'o2del' vs 'temp10'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
o2del	temp10	0.0362308	0.0294437	391.9693	5.338184	0.0223038	1

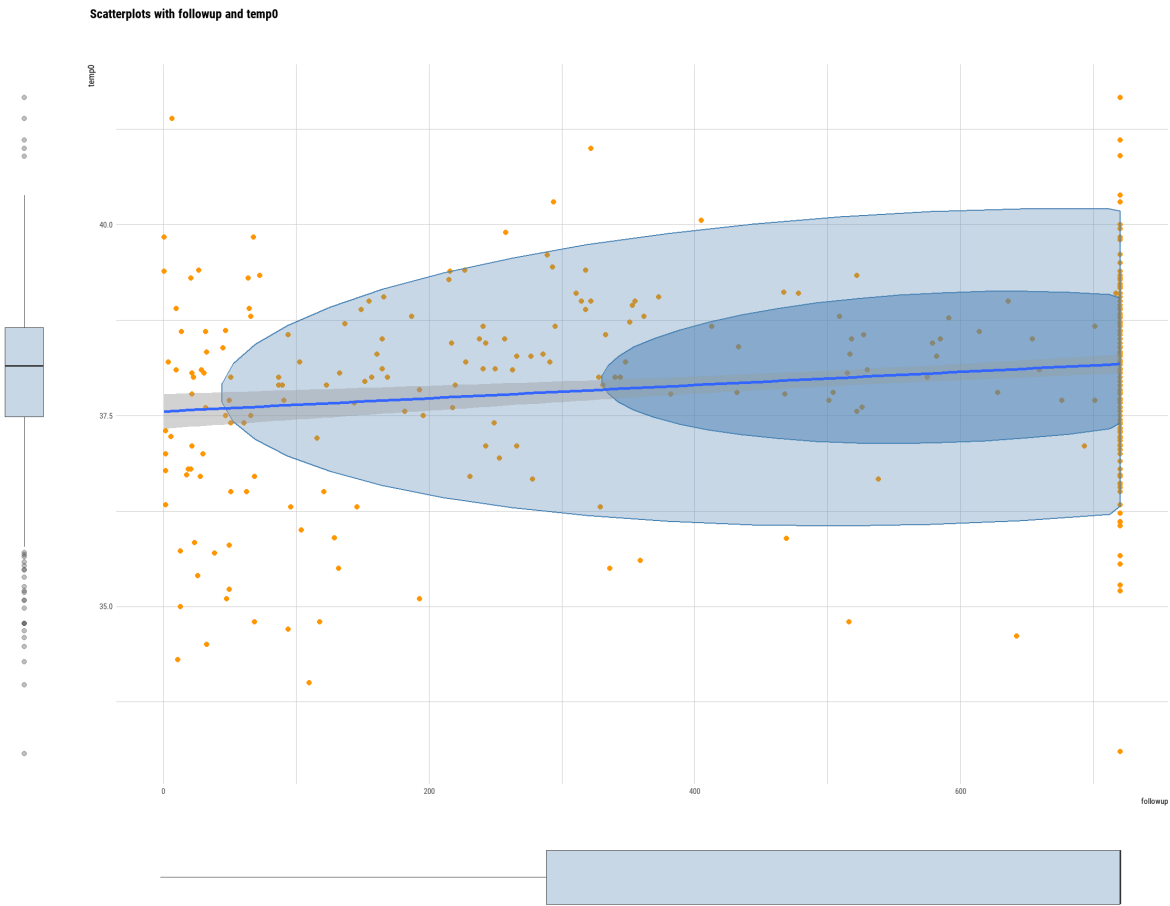
Table 7: Summary of linear model



'followup' vs 'temp0'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
followup	temp0	0.0418256	0.0397104	261.4439	19.77407	1.1e-05	1

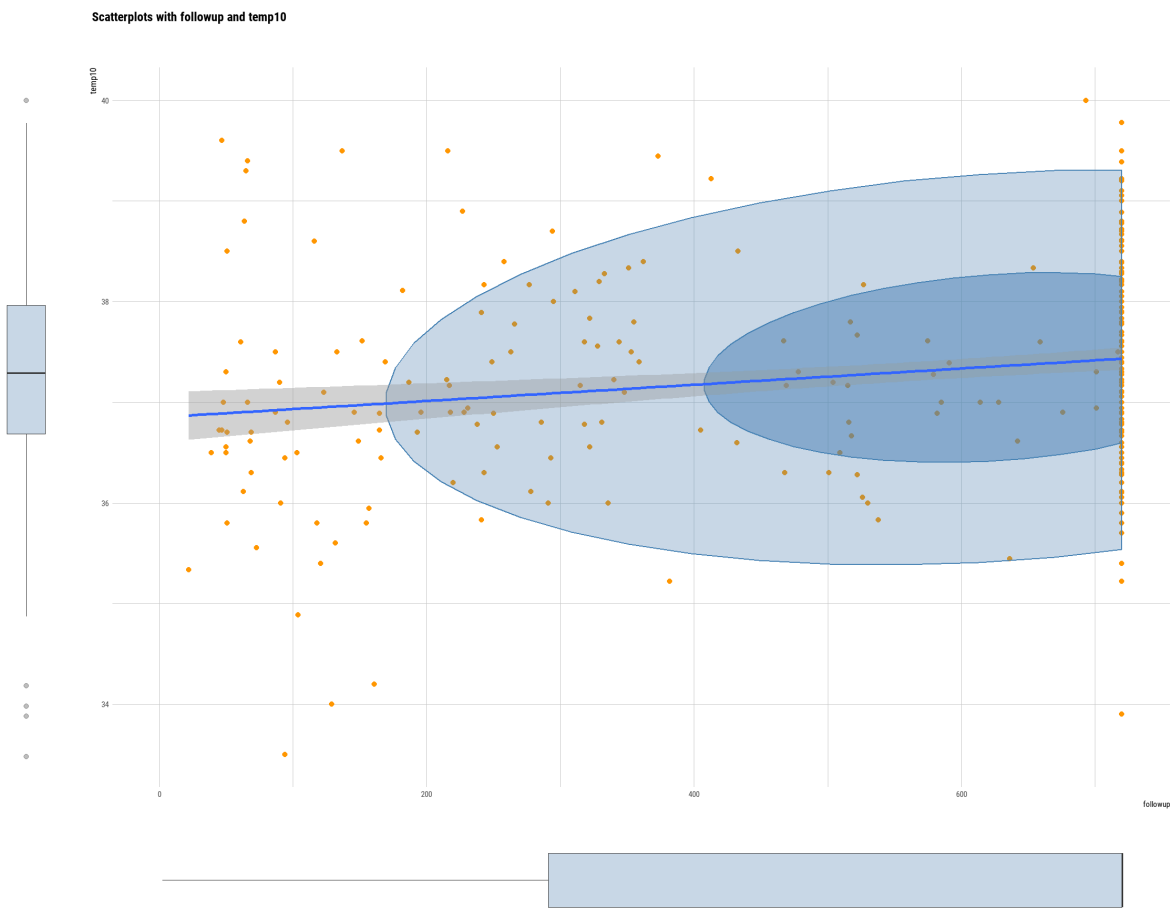
Table 7: Summary of linear model



'followup' vs 'temp10'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
followup	temp10	0.0380626	0.0356396	229.1805	15.70878	8.76e-05	1

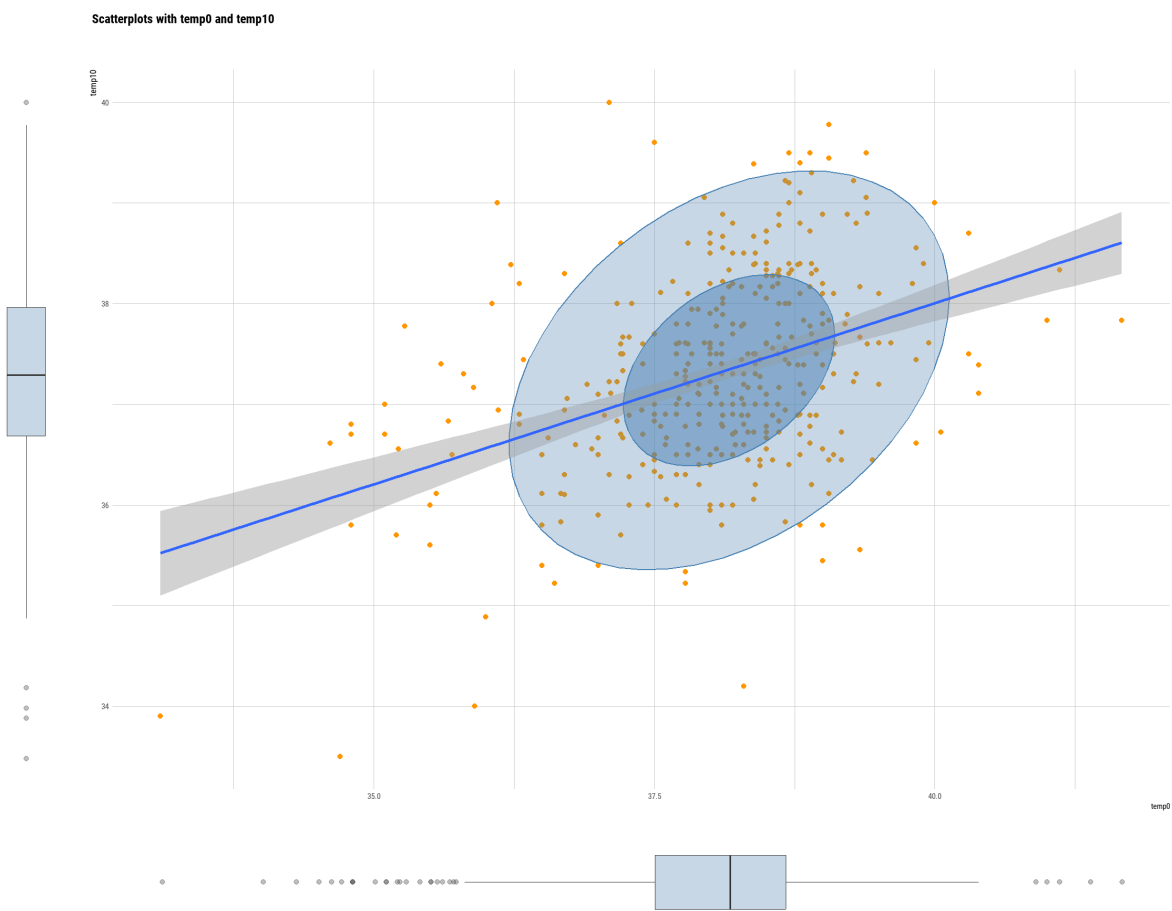
Table 7: Summary of linear model



'temp0' vs 'temp10'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
temp0	temp10	0.1562676	0.1541423	0.9754064	73.52834	0	1

Table 7: Summary of linear model



Compare Categorical Variables

The number of categorical variables is less than 2.

Multivariate Analysis

Correlation Analysis

Correlation Coefficient Matrix

first variable	second variable					
	id	apache	o2del	followup	temp0	temp10
id	NA	-0.175	-0.051	-0.039	-0.057	-0.098
apache	-0.175	NA	-0.181	-0.369	-0.029	-0.106
o2del	-0.051	-0.181	NA	0.155	0.147	0.190
followup	-0.039	-0.369	0.155	NA	0.205	0.195
temp0	-0.057	-0.029	0.147	0.205	NA	0.395
temp10	-0.098	-0.106	0.190	0.195	0.395	NA

Table 8: Matrix table of correlation coefficient

Correlation Plot

