

EDA REPORT

IRIS

Report Overview

This report was created for the EDA of *iris* 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	5
Normality Test	7
Bivariate Analysis	12
Compare Numerical Variables	12
Compare Categorical Variables	19
Multivariate Analysis	20
Correlation Analysis	20
Correlation Coefficient Matrix	20
Correlation Plot	21

Overview

Data Structures

division	metrics	value	division	metrics	value
size	observations	150	data type	numerics	4
size	variables	5	data type	integers	0
size	values	750	data type	factors/ordered	1
size	memory size (KB)	0	data type	characters	0
duplicated	duplicate observation	1	data type	Dates	0
missing	complete observation	150	data type	POSIXcts	0
missing	missing observation	0	data type	others	0
missing	missing variables	0			
missing	missing values	0			

Table 1: Data structures and types

Job Informations

division	metrics	value
dataset	dataset	iris
dataset	dataset type	data.frame
dataset	target	not defied
job	samples	150 / 150 (100%)
job	created	2021-07-30 10:33:50
job	created by	dlookr

Table 2: Job informations

Univariate Analysis

Descriptive Statistics

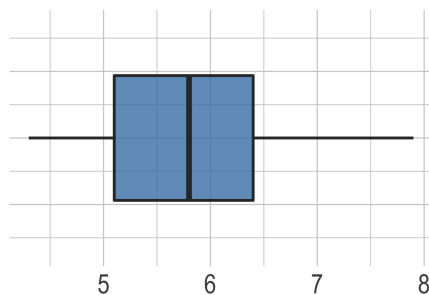
Numerical Variables

variables	missing	mean	sd	min	Q1	median	Q3	max
Sepal.Length	0	5.84	0.83	4.3	5.1	5.80	6.4	7.9
Sepal.Width	0	3.06	0.44	2.0	2.8	3.00	3.3	4.4
Petal.Length	0	3.76	1.77	1.0	1.6	4.35	5.1	6.9
Petal.Width	0	1.20	0.76	0.1	0.3	1.30	1.8	2.5

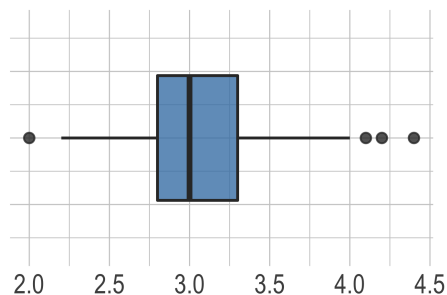
Table 3: Descriptive statistics of numerical variables

Distribution by numerical variables

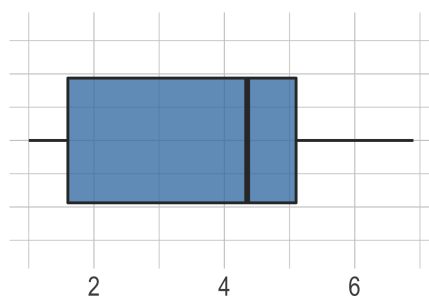
Sepal.Length



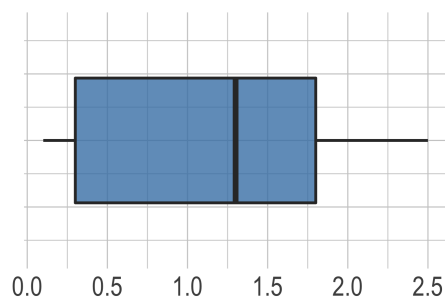
Sepal.Width



Petal.Length



Petal.Width



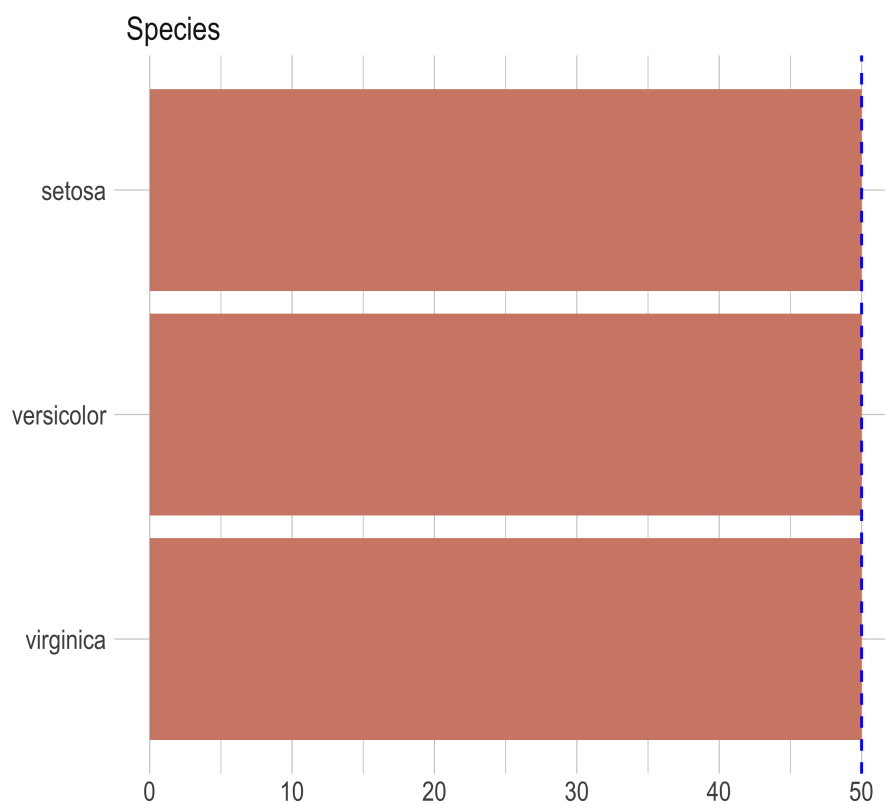
variables	data types	distinct	skewness	kurtosis	zero	negative	outlier
Sepal.Length	numeric	35	0.32	-0.55	0	0	0
Sepal.Width	numeric	23	0.32	0.23	0	0	4
Petal.Length	numeric	43	-0.28	-1.40	0	0	0
Petal.Width	numeric	22	-0.10	-1.34	0	0	0

Categorical Variables

variables	levels	observations	frequency	frequency(%)	rank
Species	setosa	150	50	33.33	1
Species	versicolor	150	50	33.33	1
Species	virginica	150	50	33.33	1

Table 4: Top rank levels of categorical variables

Frequency by levels of category



variables	missing	missing(%)	distinct	distinct ratio
Species	0	0	3	0.02

Normality Test

variable	min	Q1	median	Q3	max	skewness	kurtosis	balance
Sepal.Length	4.3	5.1	5.8	6.4	7.9	0.3	-0.6	Balanced
Sepal.Width	2.0	2.8	3.0	3.3	4.4	0.3	0.2	Balanced
Petal.Length	1.0	1.6	4.3	5.1	6.9	-0.3	-1.4	Balanced
Petal.Width	0.1	0.3	1.3	1.8	2.5	-0.1	-1.3	Balanced

Table 5: Descriptive statistics of numerical variables

Sepal.Length

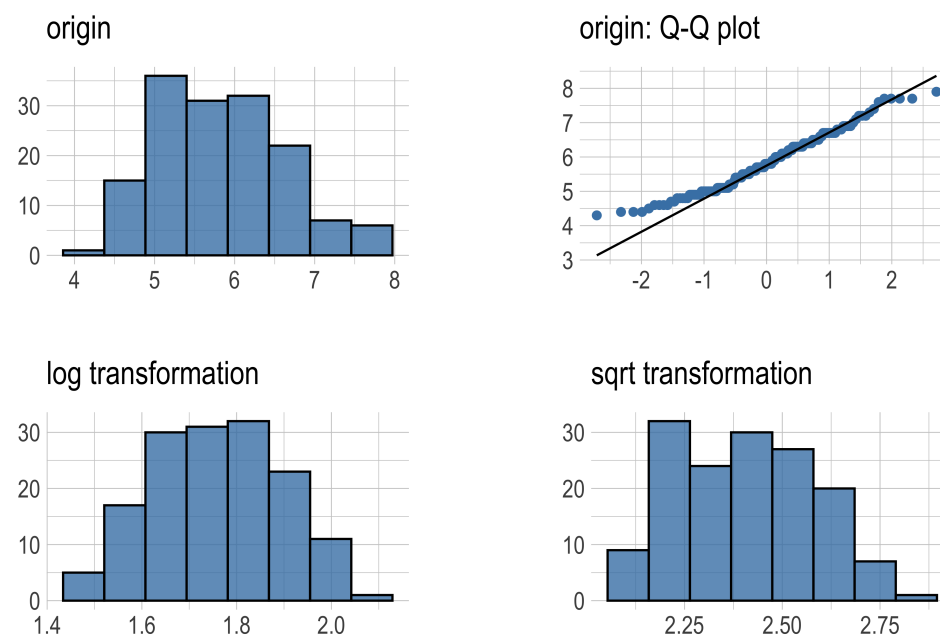
statistic	p_value	remark
0.97609	0.010181	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	0.3118	2.4264
log transformation	0.0432	2.2490
sqrt transformation	0.1761	2.3128

Table 6: skewness and kurtosis

Normality Diagnosis Plot (x)



Sepal.Width

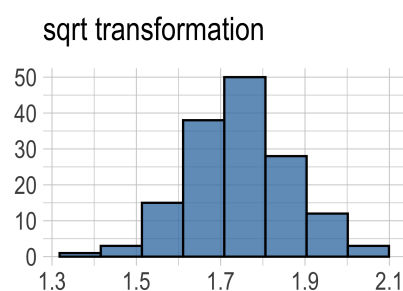
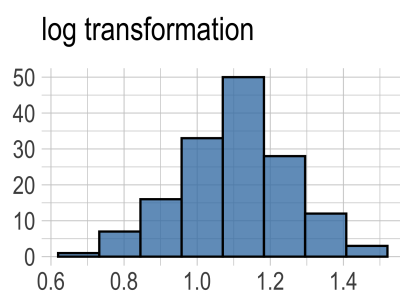
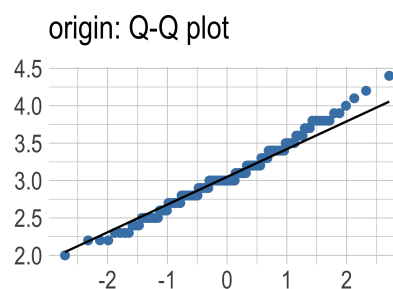
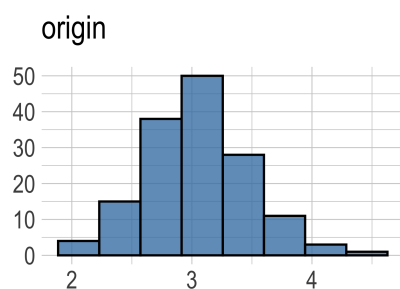
statistic	p_value	remark
0.98492	0.10115	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	0.3158	3.1810
log transformation	-0.1233	3.0861
sqrt transformation	0.0970	3.0686

Table 6: skewness and kurtosis

Normality Diagnosis Plot (x)



Petal.Length

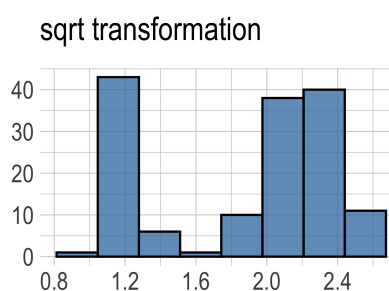
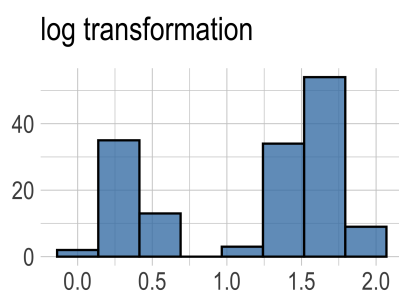
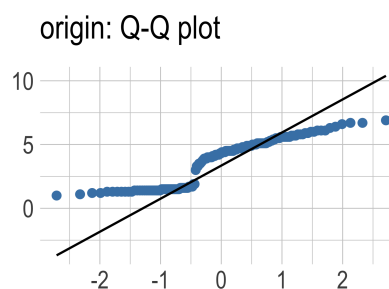
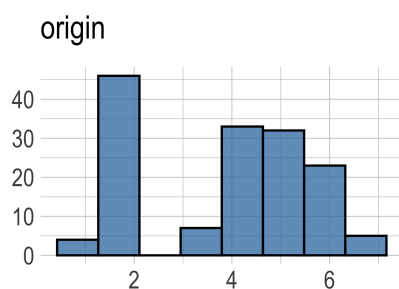
statistic	p_value	remark
0.87627	7.4123e-10	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	-0.2721	1.6045
log transformation	-0.5819	1.6188
sqrt transformation	-0.4445	1.5731

Table 6: skewness and kurtosis

Normality Diagnosis Plot (x)



Petal.Width

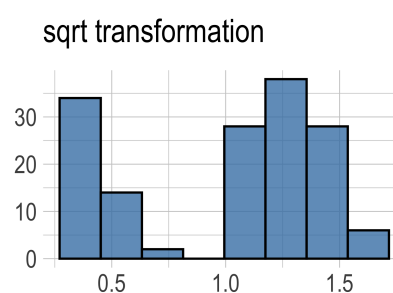
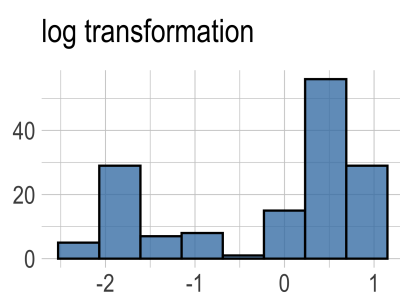
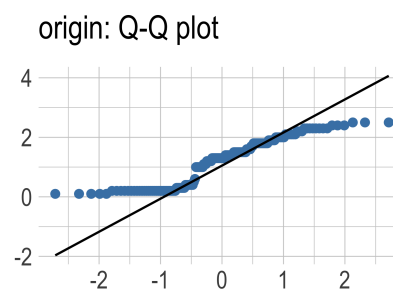
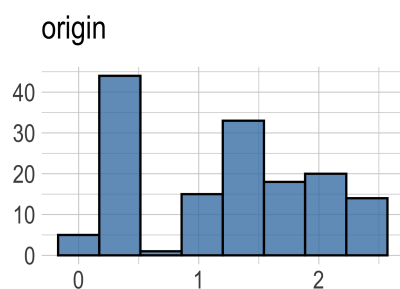
statistic	p_value	remark
0.90183	1.6805e-08	No sample

Table 6: Shapiro-Wilk normality test

type	skewness	kurtosis
original	-0.1019	1.6639
log transformation	-0.7380	1.9736
sqrt transformation	-0.4315	1.6619

Table 6: skewness and kurtosis

Normality Diagnosis Plot (x)



Bivariate Analysis

Compare Numerical Variables

first variable	second variable	correlation coefficient
Sepal.Length	Sepal.Width	-0.11757
Sepal.Length	Petal.Length	0.87175
Sepal.Length	Petal.Width	0.81794
Sepal.Width	Petal.Length	-0.42844
Sepal.Width	Petal.Width	-0.36613
Petal.Length	Petal.Width	0.96287

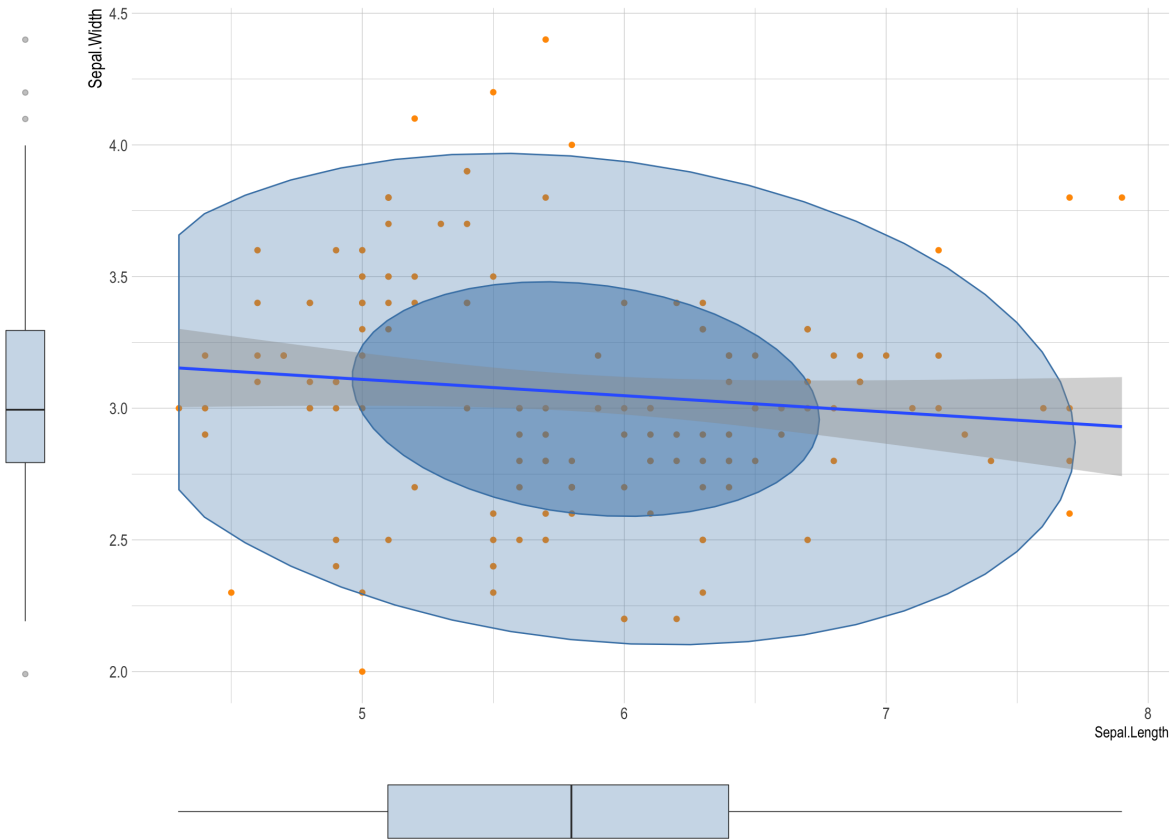
Table 7: Correlation coefficient

'Sepal.Length' vs 'Sepal.Width'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
Sepal.Length	Sepal.Width	0.0138227	0.0071593	0.8250966	2.074427	0.1518983	1

Table 7: Summary of linear model

Scatterplots with Sepal.Length and Sepal.Width

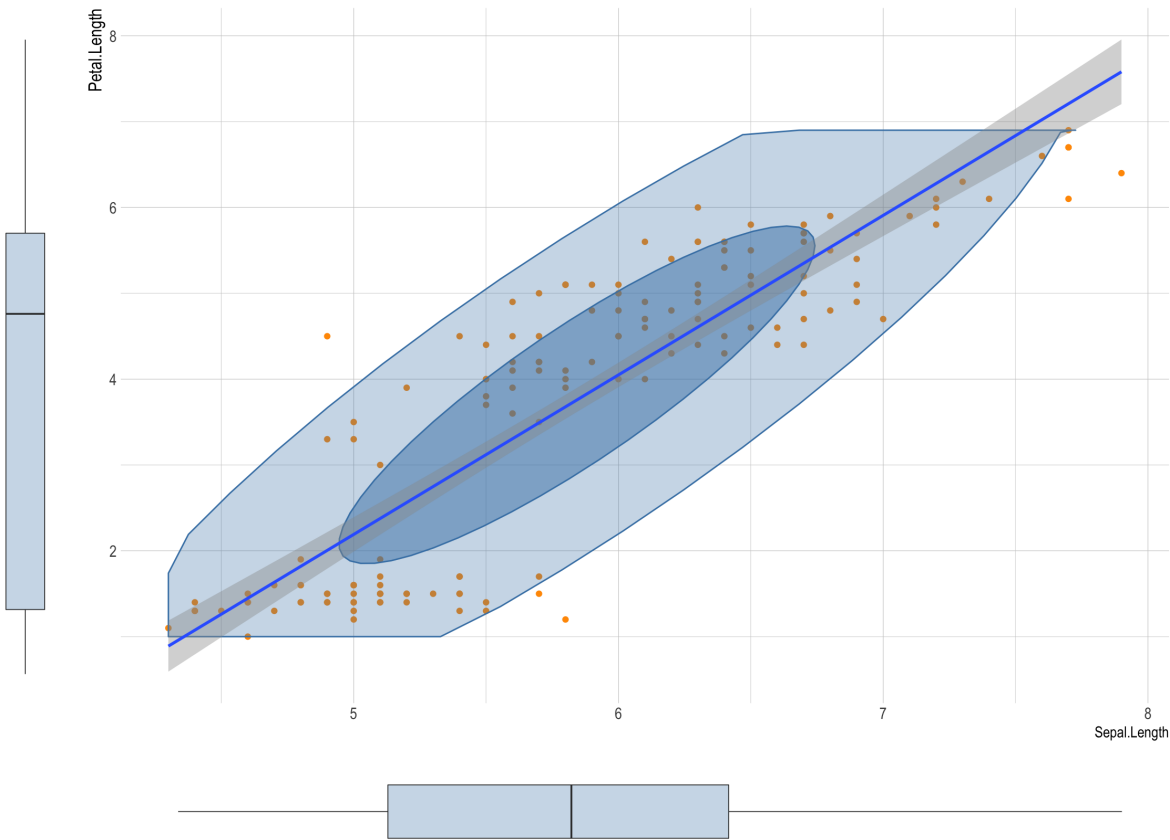


'Sepal.Length' vs 'Petal.Length'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
Sepal.Length	Petal.Length	0.7599546	0.7583327	0.4070745	468.5502	0	1

Table 7: Summary of linear model

Scatterplots with Sepal.Length and Petal.Length

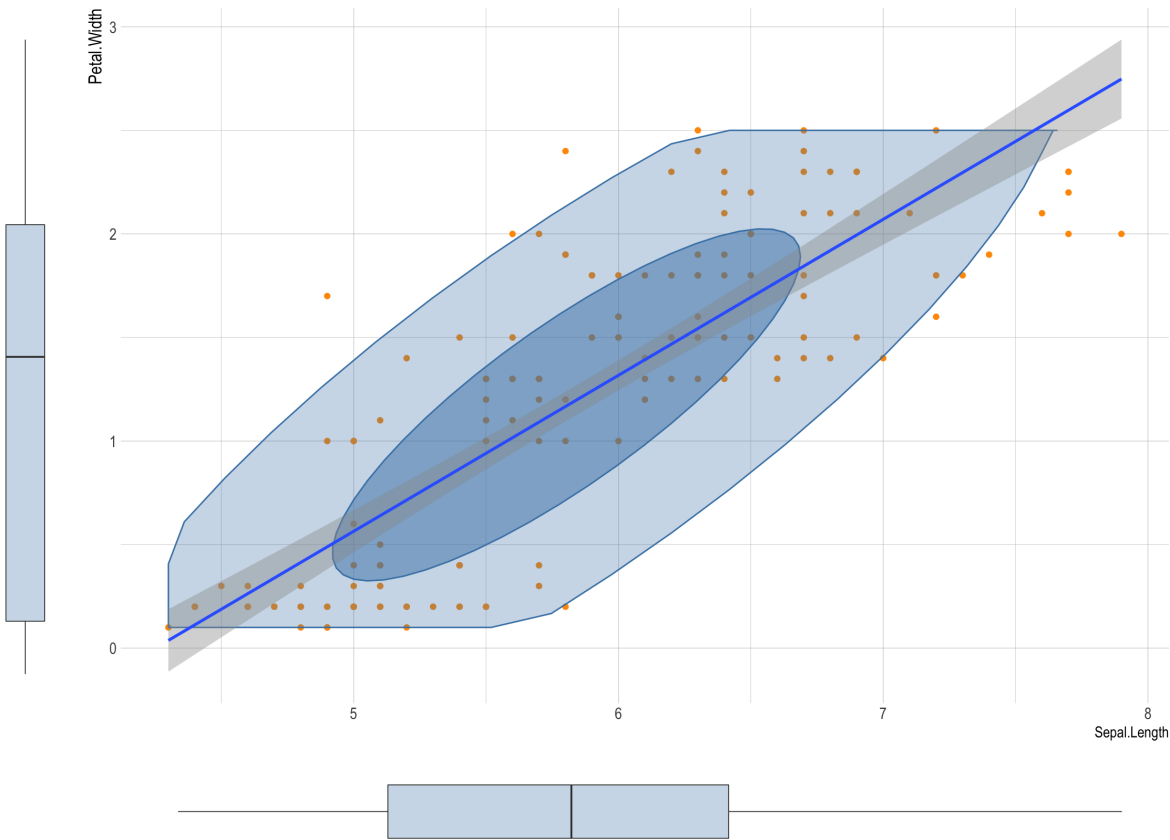


'Sepal.Length' vs 'Petal.Width'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
Sepal.Length	Petal.Width	0.6690277	0.6667914	0.4779948	299.1673	0	1

Table 7: Summary of linear model

Scatterplots with Sepal.Length and Petal.Width

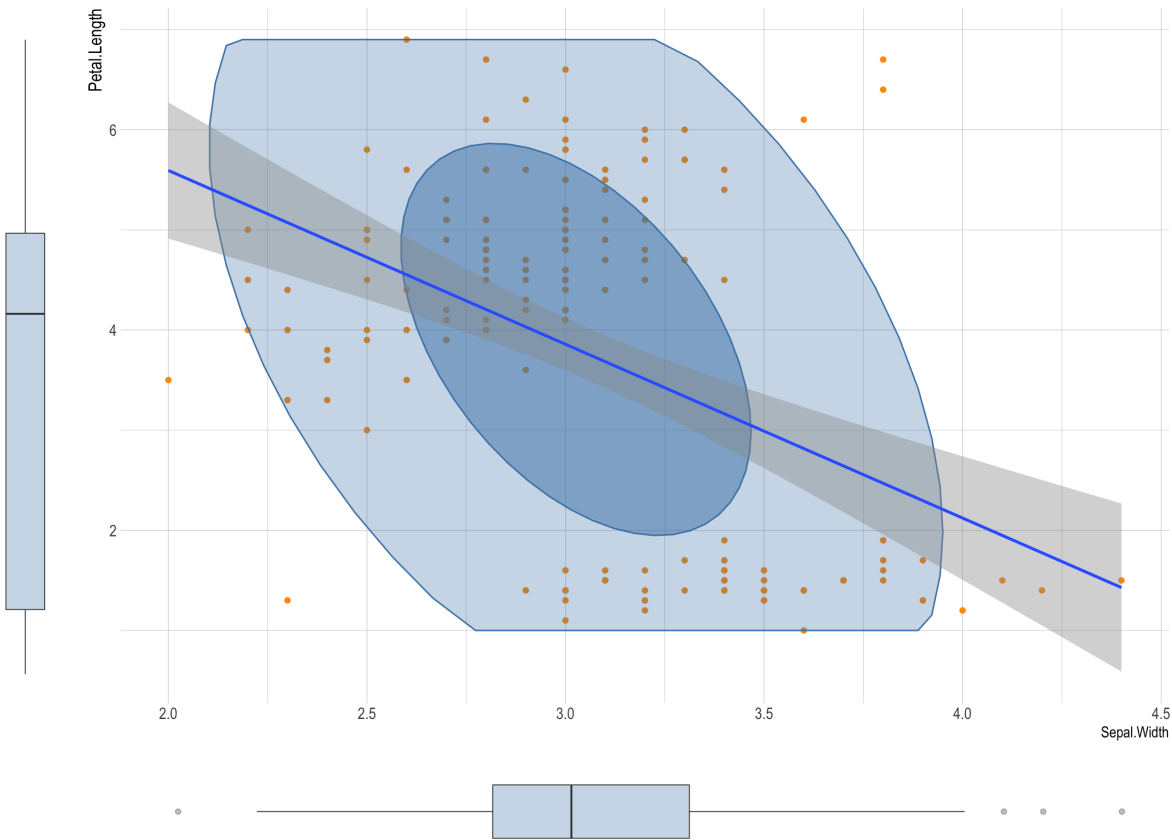


'Sepal.Width' vs 'Petal.Length'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
Sepal.Width	Petal.Length	0.1835609	0.1780444	0.3951641	33.27501	0	1

Table 7: Summary of linear model

Scatterplots with Sepal.Width and Petal.Length

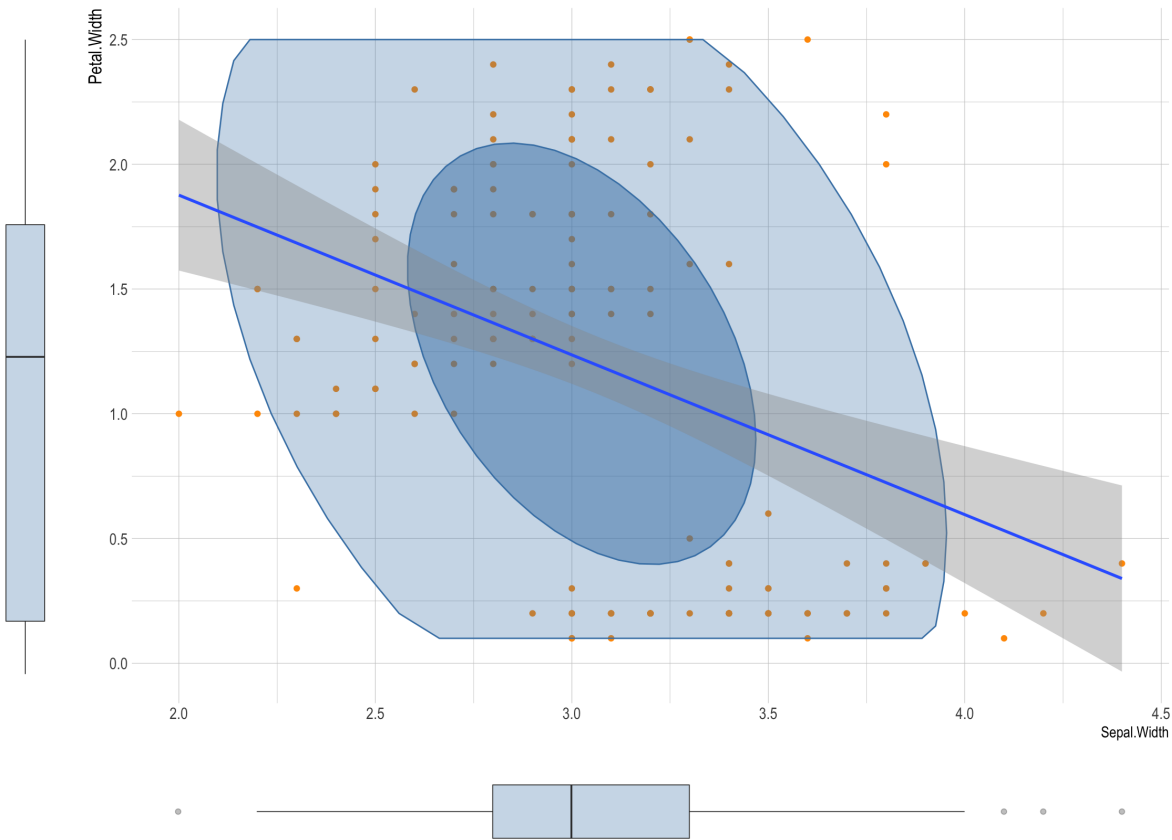


'Sepal.Width' vs 'Petal.Width'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
Sepal.Width	Petal.Width	0.1340482	0.1281972	0.40697	22.91021	4.1e-06	1

Table 7: Summary of linear model

Scatterplots with Sepal.Width and Petal.Width

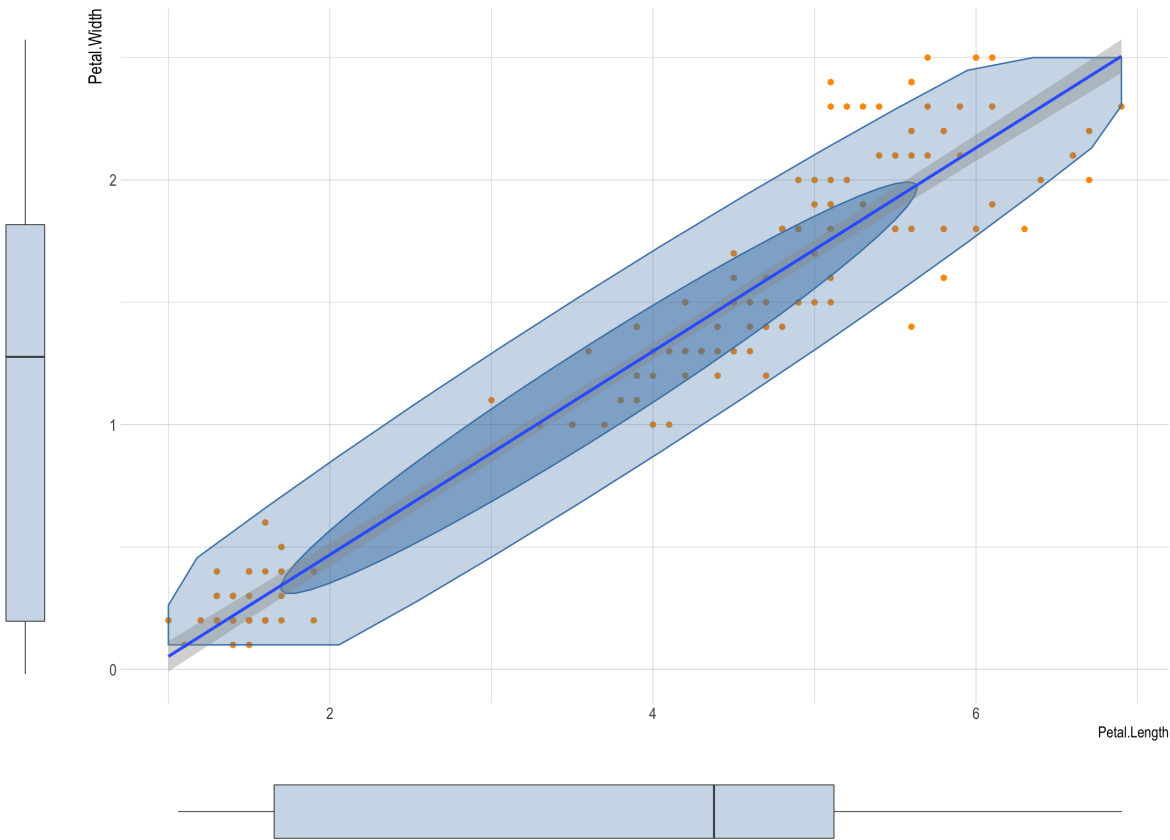


'Petal.Length' vs 'Petal.Width'

first variable	second variable	r.squared	adj.r.squared	sigma	statistic	p.value	df
Petal.Length	Petal.Width	0.9271098	0.9266173	0.4782058	1882.452	0	1

Table 7: Summary of linear model

Scatterplots with Petal.Length and Petal.Width



Compare Categorical Variables

The number of categorical variables is less than 2.

Multivariate Analysis

Correlation Analysis

Correlation Coefficient Matrix

first variable	second variable			
	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	NA	-0.118	0.872	0.818
Sepal.Width	-0.118	NA	-0.428	-0.366
Petal.Length	0.872	-0.428	NA	0.963
Petal.Width	0.818	-0.366	0.963	NA

Table 8: Matrix table of correlation coefficient

Correlation Plot

