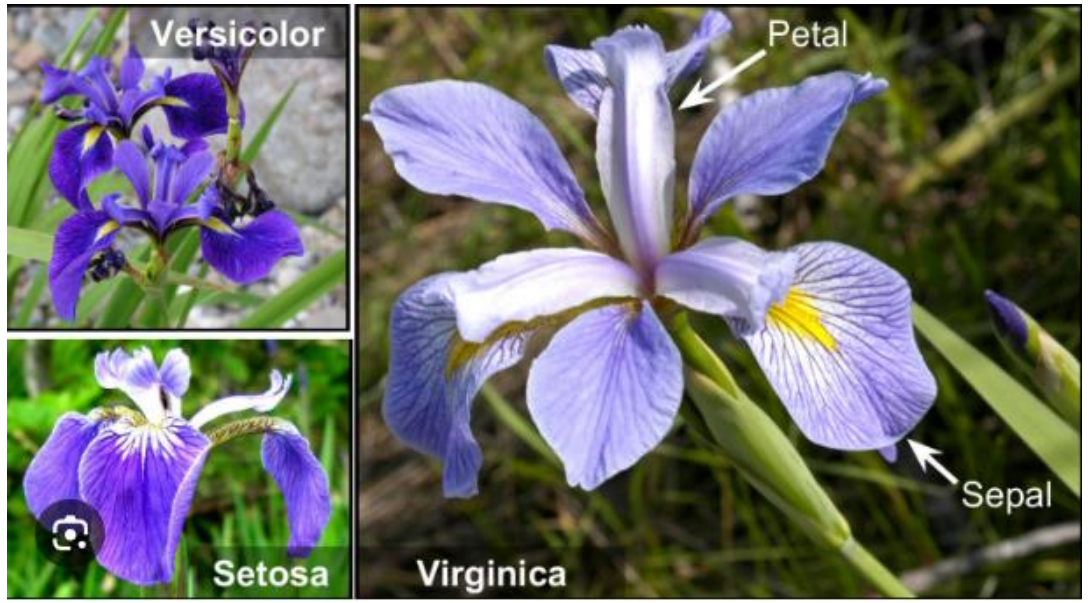
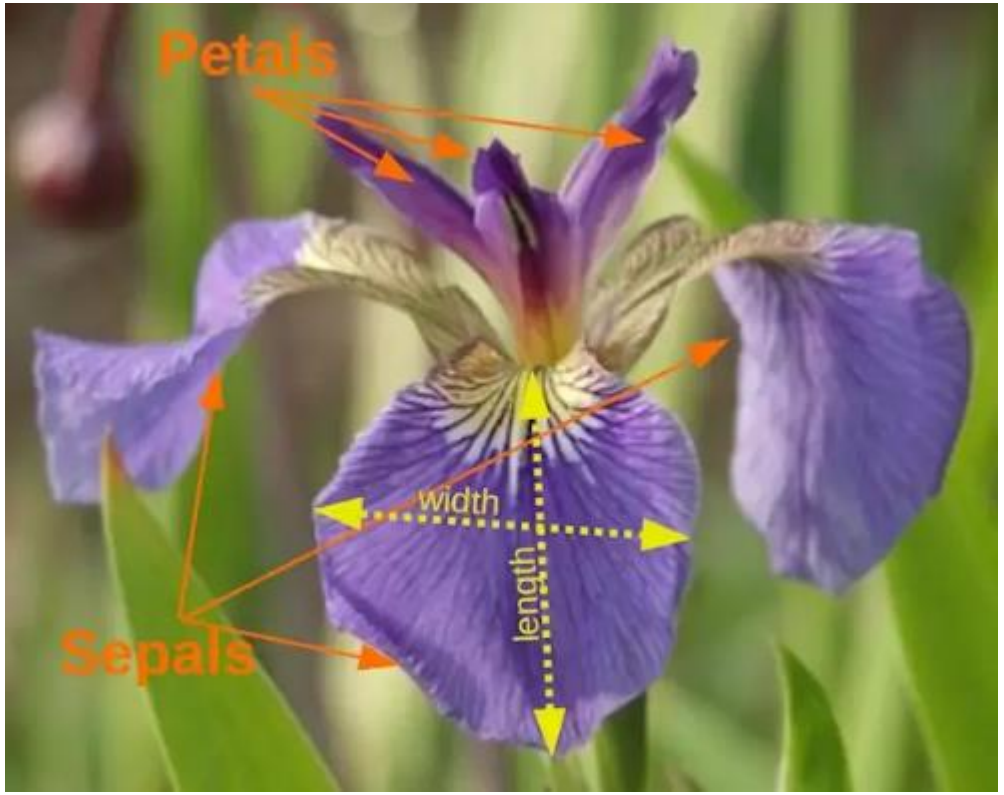
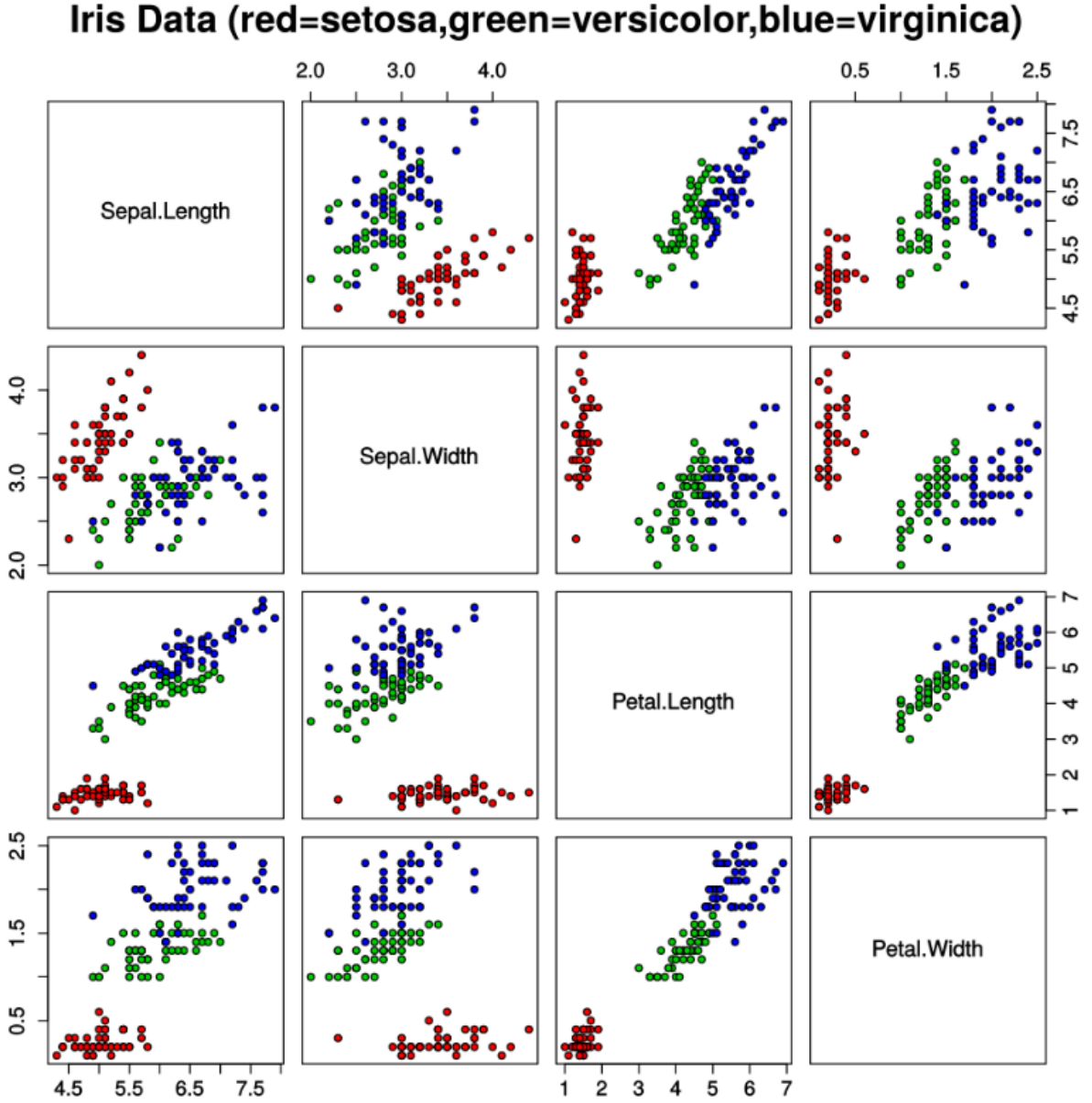




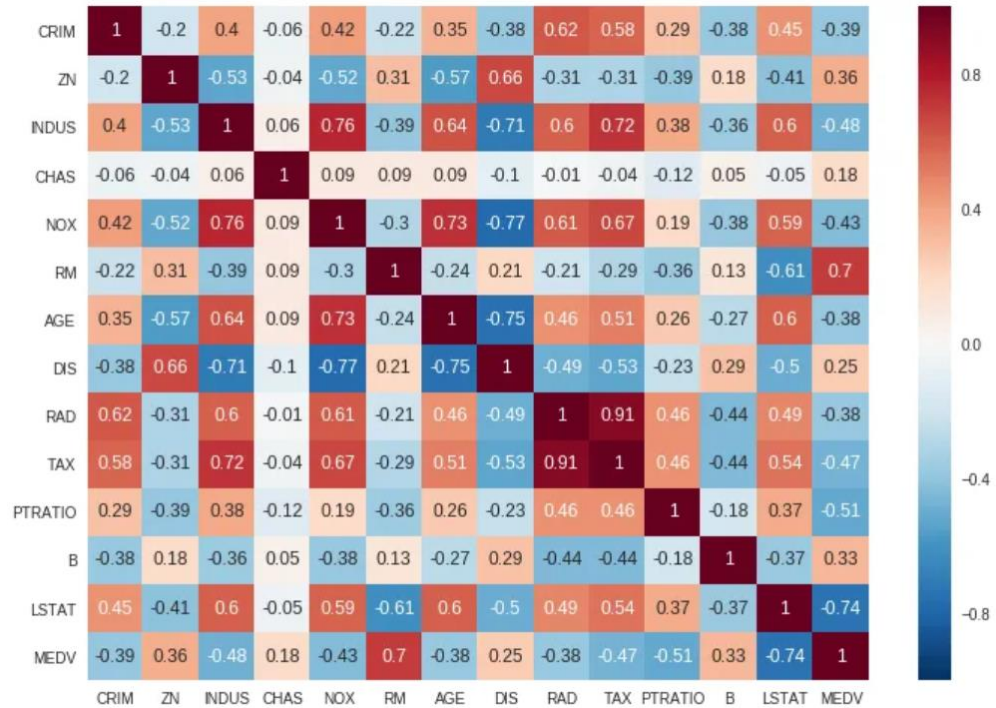
iris

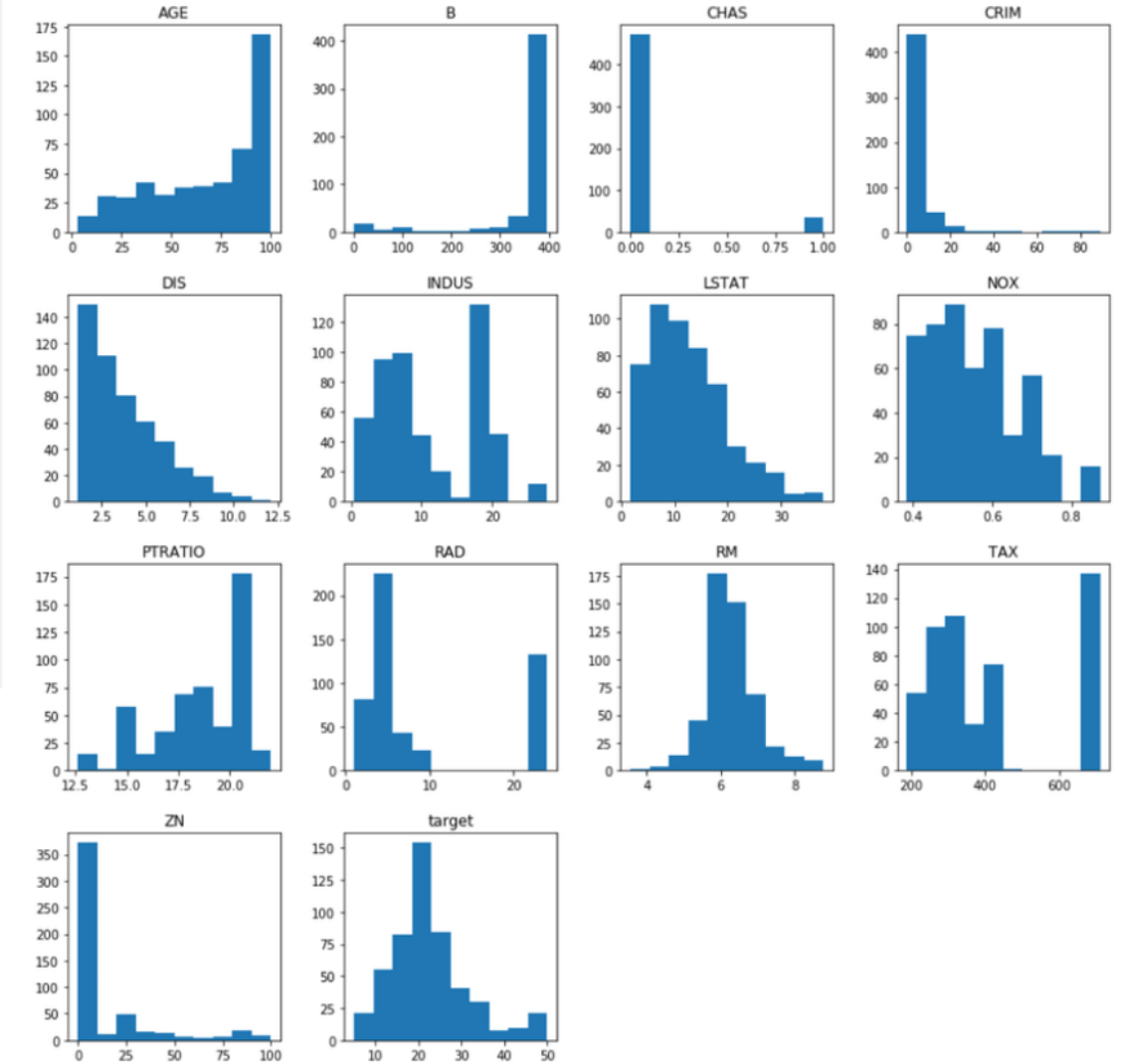


link Wikipedia

(Boston) housing

CRIM: Per capita crime rate by town
ZN: Proportion of residential land zoned for lots over 25,000 sq. ft
INDUS: Proportion of non-retail business acres per town
CHAS: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
NOX: Nitric oxide concentration (parts per 10 million)
RM: Average number of rooms per dwelling
AGE: Proportion of owner-occupied units built prior to 1940
DIS: Weighted distances to five Boston employment centers
RAD: Index of accessibility to radial highways
TAX: Full-value property tax rate per $10,000
PTRATIO: Pupil-teacher ratio by town
B: $1000(Bk - 0.63)^2$, where Bk is the proportion of [people of African American descent] by town
LSTAT: Percentage of lower status of the population
MEDV: Median value of owner-occupied homes in $1000s





Boston Housing Data: histograms of all variables.

link kaggle

Credit: Credit Card Balance Data

In ISLR: Data for an Introduction to Statistical Learning with Applications in R

Source

Simulated data, with thanks to Albert Kim for pointing out that this was omitted, and supplying the data and man documentation page on Oct 19, 2017

Format

A data frame with 400 observations on the following 12

`ID`
: Identification

`Income`
: Income in $1,000's

`Limit`
: Credit limit

`Rating`
: Credit rating

`Cards`
: Number of credit cards

`Age`
: Age in years

`Education`
: Number of years of education

`Gender`
: A factor with levels `Male` and `Female`

`Student`
: A factor with levels `No` and `Yes` indicating whether the individual was a student

`Married`
: A factor with levels `No` and `Yes` indicating whether the individual was married

`Ethnicity`
: A factor with levels `African American`, `Asian`, and `Caucasian` indicating the individual's ethnicity

`Balance`
: Average credit card balance in $.

Info
400 instances
12 features (no missing values)
Data has no target variable.
0 meta attributes

Columns (Double click to edit)

	Name	Type	Role	Values
1	Feature 1	N numeric	feature	
2	Income	N numeric	feature	
3	Limit	N numeric	feature	
4	Rating	N numeric	feature	
5	Cards	N numeric	feature	
6	Age	N numeric	feature	
7	Education	N numeric	feature	
8	Gender	C categorical	feature	Female, Male
9	Student	C categorical	feature	No, Yes
10	Married	C categorical	feature	No, Yes
11	Ethnicity	C categorical	feature	African American, Asian, Caucasian
12	Balance	N numeric	feature	

Credit

CreditCat

Info
400 instances
12 features (no missing values)
Data has no target variable.
0 meta attributes

Columns (Double click to edit)

	Name	Type	Role	Values
1	Feature 1	N numeric	feature	
2	Income	N numeric	feature	
3	Limit	N numeric	feature	
4	Rating	N numeric	feature	
5	Cards	N numeric	feature	
6	Age	N numeric	feature	
7	Education	N numeric	feature	
8	Gender	C categorical	feature	Female, Male
9	Student	C categorical	feature	No, Yes
10	Married	C categorical	feature	No, Yes
11	Ethnicity	C categorical	feature	African American, Asian, Caucasian
12	Balance	C categorical	feature	alto, basso, medio

case NY-SF

Info

492 instances
8 features (no missing values)
Data has no target variable.
0 meta attributes

Columns (Double click to edit)

	Name	Type	Role	Values
1	**in_sf**	**C categorical**	**target**	**0, 1**
2	beds	N numeric	feature	
3	bath	N numeric	feature	
4	price	N numeric	feature	
5	year_built	N numeric	feature	
6	sqft	N numeric	feature	
7	price_per_sqft	N numeric	feature	
8	elevation	N numeric	feature	

Zoo Animal Classification

kaggle

zoo

Use Machine Learning Methods to Correctly Classify Animals Based Upon Attributes

#	Name	Type	Domain
1	hair	N	{yes, no}
2	feathers	N	{yes, no}
3	eggs	N	{yes, no}
4	milk	N	{yes, no}
5	airborne	N	{yes, no}
6	aquatic	N	{yes, no}
7	predator	N	{yes, no}
8	toothed	N	{yes, no}
9	backbone	N	{yes, no}
10	breathes	N	{yes, no}
11	venomous	N	{yes, no}
12	fins	N	{yes, no}
13	legs	I	{0,2,4,5,6,8}
14	tail	N	{yes, no}
15	domestic	N	{yes, no}
16	catsize	N	{yes, no}

Description of the attributes in the Zoo dataset.

This dataset consists of 101 animals from a zoo.

There are 16 variables with various traits to describe the animals.

The 7 Class Types are: Mammal, Bird, Reptile, Fish, Amphibian, Bug and Invertebrate

The purpose for this dataset is to be able to predict the classification of the animals, based upon the variables.

It is the perfect dataset for those who are new to learning Machine Learning.

1 -- (41) aardvark, antelope, bear, boar, buffalo, calf, cavy, cheetah, deer, dolphin, elephant, fruitbat, giraffe, girl, goat, gorilla, hamster, hare, leopard, lion, lynx, mink, mole, mongoose, opossum, oryx, platypus, polecat, pony, porpoise, puma, pussycat, raccoon, reindeer, seal, sealion, squirrel, vampire, vole, wallaby,wolf
2 -- (20) chicken, crow, dove, duck, flamingo, gull, hawk, kiwi, lark, ostrich, parakeet, penguin, pheasant, rhea, skimmer, skua, sparrow, swan, vulture, wren
3 -- (5) pitviper, seasnake, slowworm, tortoise, tuatara
4 -- (13) bass, carp, catfish, chub, dogfish, haddock, herring, pike, piranha, seahorse, sole, stingray, tuna
5 -- (4) frog, frog, newt, toad
6 -- (8) flea, gnat, honeybee, housefly, ladybird, moth, termite, wasp
7 -- (10) clam, crab, crayfish, lobster, octopus, scorpion, seawasp, slug, starfish, worm

16 feature + 1 risposta categorica (7 classi): ***type***

link kaggle