

Iris Versicolor

## Iris data set

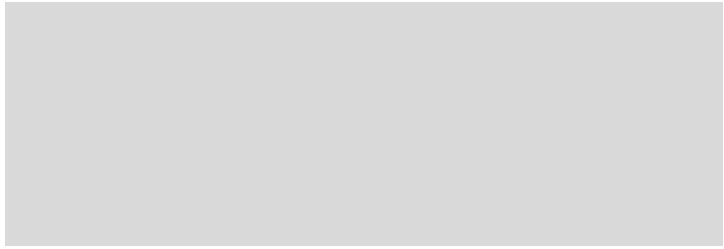
Species name	Petal width	Petal length	Sepal width	Sepal length
Setosa	0.2	1.4	3.5	5.1
Setosa	0.2	1.4	3	4.9
Setosa	0.2	1.3	3.2	4.7
Setosa	0.2	1.5	3.1	4.6
Setosa	0.2	1.4	3.6	5
Setosa	0.4	1.7	3.9	5.4
Setosa	0.3	1.4	3.4	4.6
Setosa	0.2	1.5	3.4	5
Setosa	0.2	1.4	2.9	4.4
Setosa	0.1	1.5	3.1	4.9
Setosa	0.2	1.5	3.7	5.4
Setosa	0.2	1.6	3.4	4.8
Setosa	0.1	1.4	3	4.8
Setosa	0.1	1.1	3	4.3
Setosa	0.2	1.2	4	5.8
Setosa	0.4	1.5	4.4	5.7
Setosa	0.4	1.3	3.9	5.4
Setosa	0.3	1.4	3.5	5.1

Setosa	0.3	1.7	3.8	5.7
Setosa	0.3	1.5	3.8	5.1
Setosa	0.2	1.7	3.4	5.4
Setosa	0.4	1.5	3.7	5.1
Setosa	0.2	1	3.6	4.6
Setosa	0.5	1.7	3.3	5.1
Setosa	0.2	1.9	3.4	4.8
Setosa	0.2	1.6	3	5
Setosa	0.4	1.6	3.4	5
Setosa	0.2	1.5	3.5	5.2
Setosa	0.2	1.4	3.4	5.2
Setosa	0.2	1.6	3.2	4.7
Setosa	0.2	1.6	3.1	4.8
Setosa	0.4	1.5	3.4	5.4
Setosa	0.1	1.5	4.1	5.2
Setosa	0.2	1.4	4.2	5.5
Setosa	0.2	1.5	3.1	4.9
Setosa	0.2	1.2	3.2	5
Setosa	0.2	1.3	3.5	5.5
Setosa	0.1	1.4	3.6	4.9
Setosa	0.2	1.3	3	4.4
Setosa	0.2	1.5	3.4	5.1
Setosa	0.3	1.3	3.5	5
Setosa	0.3	1.3	2.3	4.5
Setosa	0.2	1.3	3.2	4.4
Setosa	0.6	1.6	3.5	5
Setosa	0.4	1.9	3.8	5.1
Setosa	0.3	1.4	3	4.8
Setosa	0.2	1.6	3.8	5.1
Setosa	0.2	1.4	3.2	4.6
Setosa	0.2	1.5	3.7	5.3
Setosa	0.2	1.4	3.3	5
Versicolor	1.4	4.7	3.2	7
Versicolor	1.5	4.5	3.2	6.4
Versicolor	1.5	4.9	3.1	6.9
Versicolor	1.3	4	2.3	5.5
Versicolor	1.5	4.6	2.8	6.5

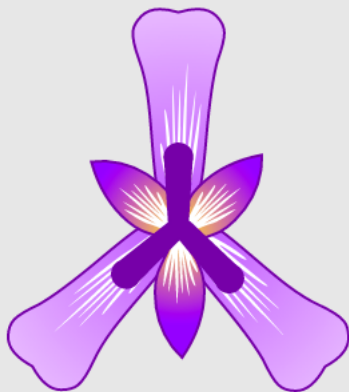
Versicolor	1.3	4.5	2.8	5.7
Versicolor	1.6	4.7	3.3	6.3
Versicolor	1	3.3	2.4	4.9
Versicolor	1.3	4.6	2.9	6.6
Versicolor	1.4	3.9	2.7	5.2
Versicolor	1	3.5	2	5
Versicolor	1.5	4.2	3	5.9
Versicolor	1	4	2.2	6
Versicolor	1.4	4.7	2.9	6.1
Versicolor	1.3	3.6	2.9	5.6
Versicolor	1.4	4.4	3.1	6.7
Versicolor	1.5	4.5	3	5.6
	1	4.1	2.7	5.8
Versicolor				
Versicolor	1.5	4.5	2.2	6.2
Versicolor	1.1	3.9	2.5	5.6
Versicolor	1.8	4.8	3.2	5.9
Versicolor	1.3	4	2.8	6.1
Versicolor	1.5	4.9	2.5	6.3
Versicolor	1.2	4.7	2.8	6.1
Versicolor	1.3	4.3	2.9	6.4
Versicolor	1.4	4.4	3	6.6
Versicolor	1.4	4.8	2.8	6.8
Versicolor	1.7	5	3	6.7
Versicolor	1.5	4.5	2.9	6
Versicolor	1	3.5	2.6	5.7
Versicolor	1.1	3.8	2.4	5.5
Versicolor	1	3.7	2.4	5.5
Versicolor	1.2	3.9	2.7	5.8
Versicolor	1.6	5.1	2.7	6
Versicolor	1.5	4.5	3	5.4
Versicolor	1.6	4.5	3.4	6
Versicolor	1.5	4.7	3.1	6.7
Versicolor	1.3	4.4	2.3	6.3
Versicolor	1.3	4.1	3	5.6
Versicolor	1.3	4	2.5	5.5
Versicolor	1.2	4.4	2.6	5.5
Versicolor	1.4	4.6	3	6.1
Versicolor	1.2	4	2.6	5.8

Versicolor	1	3.3	2.3	5
Versicolor	1.3	4.2	2.7	5.6
Versicolor	1.2	4.2	3	5.7
Versicolor	1.3	4.2	2.9	5.7
Versicolor	1.3	4.3	2.9	6.2
Versicolor	1.1	3	2.5	5.1
Versicolor	1.3	4.1	2.8	5.7
Verginica	2.5	6	3.3	6.3
Verginica	1.9	5.1	2.7	5.8
Verginica	2.1	5.9	3	7.1
Verginica	1.8	5.6	2.9	6.3
Verginica	2.2	5.8	3	6.5
Verginica	2.1	6.6	3	7.6
Verginica	1.7	4.5	2.5	4.9
Verginica	1.8	6.3	2.9	7.3
Verginica	1.8	5.8	2.5	6.7
Verginica	2.5	6.1	3.6	7.2
Verginica	2	5.1	3.2	6.5
Verginica	1.9	5.3	2.7	6.4
Verginica	2.1	5.5	3	6.8
Verginica	2	5	2.5	5.7
Verginica	2.4	5.1	2.8	5.8
Verginica	2.3	5.3	3.2	6.4
Verginica	1.8	5.5	3	6.5
Verginica	2.2	6.7	3.8	7.7
Verginica	2.3	6.9	2.6	7.7
Verginica	1.5	5	2.2	6
Verginica	2.3	5.7	3.2	6.9
Verginica	2	4.9	2.8	5.6
Verginica	2	6.7	2.8	7.7
Verginica	1.8	4.9	2.7	6.3
Verginica	2.1	5.7	3.3	6.7
Verginica	1.8	6	3.2	7.2
Verginica	1.8	4.8	2.8	6.2
Verginica	1.8	4.9	3	6.1
Verginica	2.1	5.6	2.8	6.4
Verginica	1.6	5.8	3	7.2
Verginica	1.9	6.1	2.8	7.4
Verginica	2	6.4	3.8	7.9
Verginica	2.2	5.6	2.8	6.4

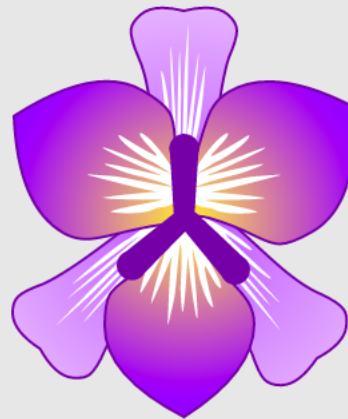
Verginica	1.5	5.1	2.8	6.3
Verginica	1.4	5.6	2.6	6.1
Verginica	2.3	6.1	3	7.7
Verginica	2.4	5.6	3.4	6.3
Verginica	1.8	5.5	3.1	6.4
Verginica	1.8	4.8	3	6
Verginica	2.1	5.4	3.1	6.9
Verginica	2.4	5.6	3.1	6.7
Verginica	2.3	5.1	3.1	6.9
Verginica	1.9	5.1	2.7	5.8
Verginica	2.3	5.9	3.2	6.8
Verginica	2.5	5.7	3.3	6.7
Verginica	2.3	5.2	3	6.7
Verginica	1.9	5	2.5	6.3
Verginica	2	5.2	3	6.5
Verginica	2.3	5.4	3.4	6.2
Verginica	1.8	5.1	3	5.9



petal



**Iris Setosa**



**Iris Virginica**

The Iris Dataset contains four dependent variables:

1. sepal length
2. sepal width
3. petal length
4. petal width

of 50 samples of three species of Iris (Iris setosa, Iris virginica and Iris versicolor).

In this case, your independent variable is species, of which there are three categories

To summarize:

**Your dependent variables are:**

- |                 |            |
|-----------------|------------|
| 1. sepal length | continuous |
| 2. sepal width  | continuous |
| 3. petal length | continuous |
| 4. petal width  | continuous |

**Your independent variable is:**

plant species:

- |               |             |
|---------------|-------------|
| 1. setosa     | categorical |
| 2. virginica  | categorical |
| 3. versicolor | categorical |

## Your replicates include

50 samples of each dependent variable  
for each species

The experiment looks like this:

Species name	Petal width	Petal length	Sepal width	Sepal length	
setosa	50	50	50	50	# of samples/ measurements/ replicates
virginica	50	50	50	50	
versicolor	50	50	50	50	

Breaking down your data to reflect the experimental design can show  
how to design a data table

What if we fill in the mean for each combination of independent  
dependent variables in place of the replicate number?

Species name	Petal width	Petal length	Sepal width	Sepal length
setosa	0.25	1.46	3.43	5.01
virginica	2.03	5.55	2.97	6.59
versicolor	1.33	4.26	2.77	5.94

How-to:

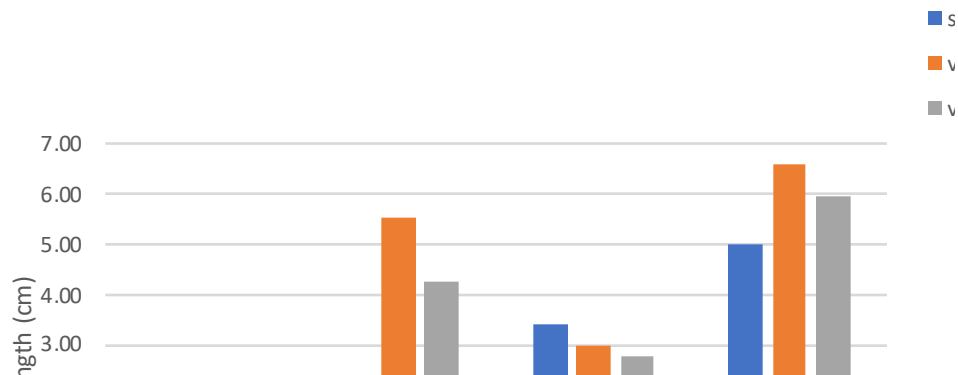
excel form  
=average

Poof! You now have a summary of your data!

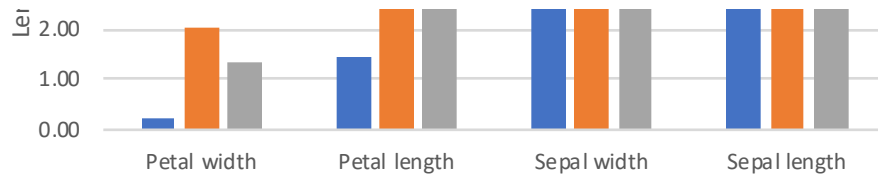
Let's graph these data!

We have a categorical independent variable (species)  
and continuous dependent variables (flower measurements)  
A bar graph will give us a nice overview of our dataset

Flower measurements of 3 different Iris species







### What does this chart tell us?

- The average measurements (y-axis) of each flower part (x-axis) for each species
- We can make comparisons between each species with respect to its average
- Sepal and petal length is longer in virginica than the other species
- The petal width and length is smallest in setosa

and so on...

But we don't know how variable the data is

For example, the average sepal length of virginica could be driven (or made larger) by the measurement of several unusually large individuals

.....>

.....>

.....>

example

average

Given this example, maybe we can get more descriptive than the mean

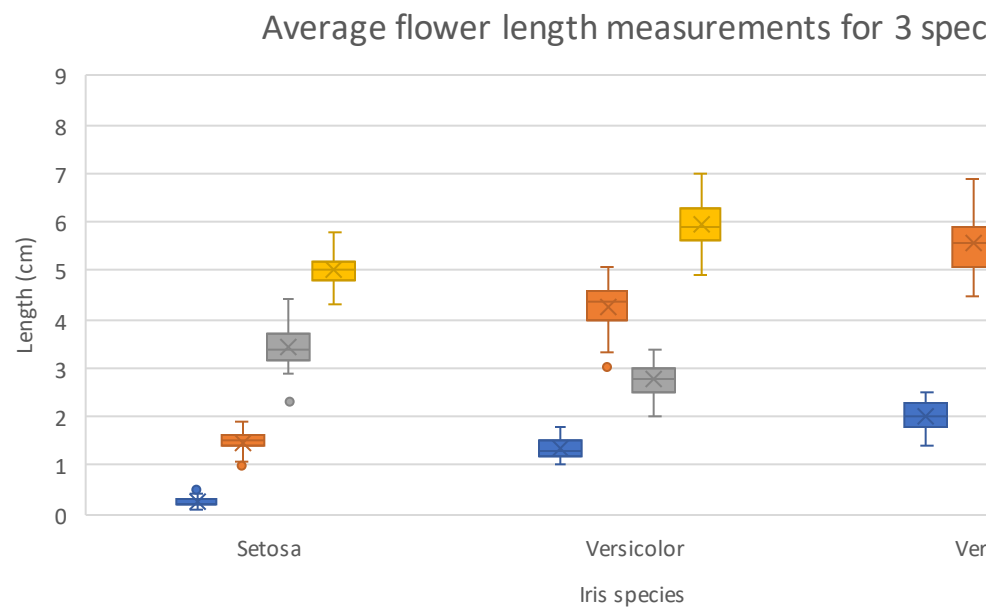
**An outlier is an observation that lies an abnormal distance from other**

	Set 1	Set 2	Set 3
	7	5.5	4
	4	6	2
	6	7	4
	3	7	15
	5.5	7	16
	18	6	1.75
	2	7	3
Average	6.5	6.5	6.5
Variance	28.8	0.4	38.3
25th percentile	3.5	6	2.5

50th percentile	5.5	7	4
75th percentile	6.5	7	9.5
Median	5.5	7	4

The above statistics not only give a good description of the center (=median) but also indicates how much spread/variability the data has.

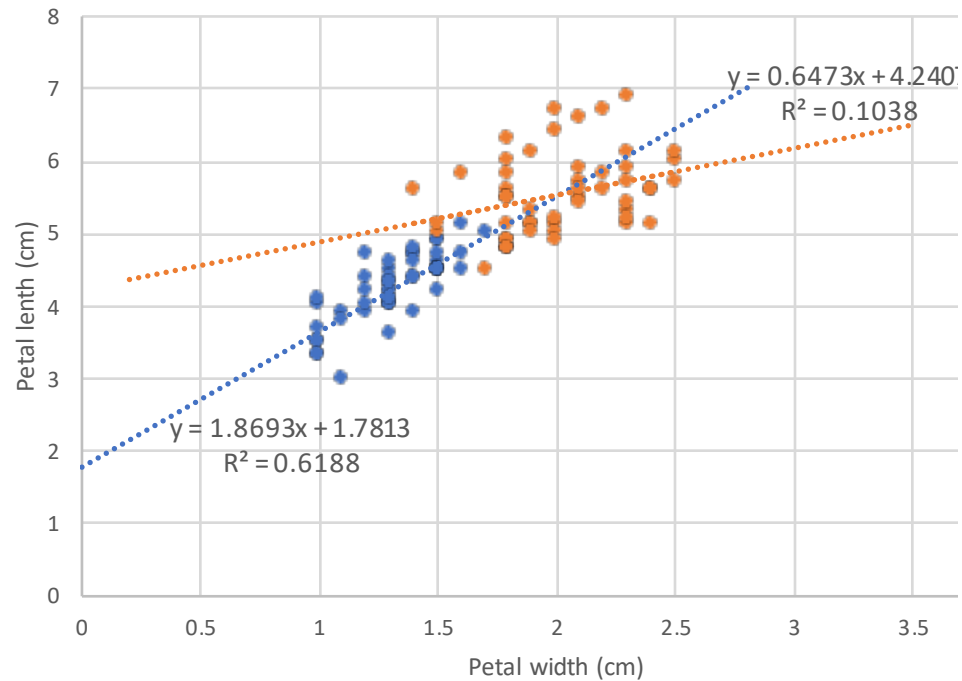
Let's try to make a more descriptive figure than the one previous



Here, we can see the median value (which is the horizontal line inside each box) for each species/trait combination. Additionally, we can see how much spread is in each trait by observing the quartiles. This is a very descriptive figure that allows for easy visual analysis of trends within the data.

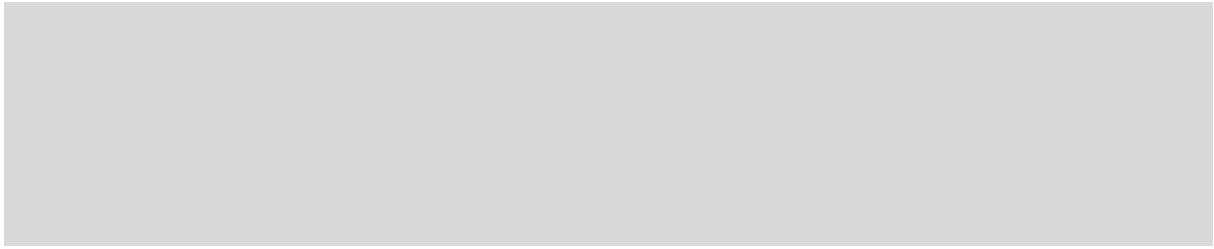
What if we have to graph 2 variables which are both continuous again? Start with a scatter plot!

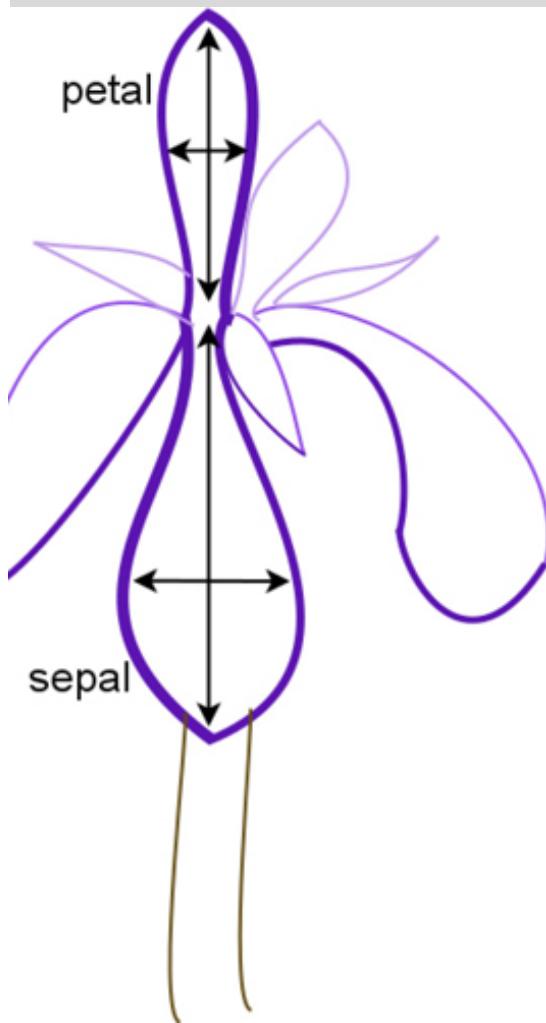
Relationship between petal length and petal width for three species of Iris



Generally, if a line fitted through a series of points represents the relationship between the X-axis and Y-axis variables well, that line will have a high R-squared value

For example, the R-squared value of the Versicolor line (regression I) is a fairly high R-squared value, and indicates that the relationship between petal length and petal width is likely to be occurring by more than sheer chance.





view you

and

formula for average:

(select data)

setosa

virginica

versicolor

### How-to:

1. To create this graph, select the entire data table of averages
  2. click the insert menu
  3. and select recommended charts
- which one is most appropriate for the data set?

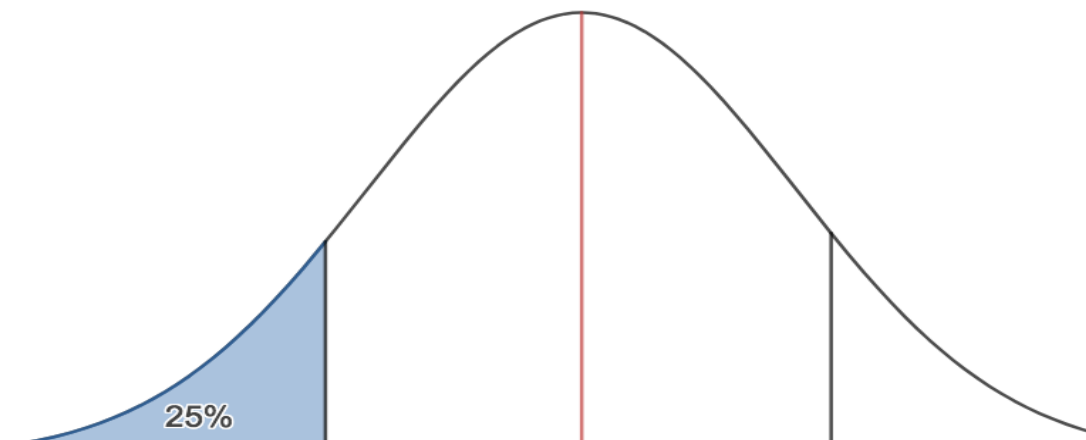
species (legend)

age flower part measurements:

The mean sepal length of virginica is approximately 6.5.  
You can get a mean of 6.5 many different ways! For  
example, all three sets of data below have a mean of 6.5.

Set 1	Set 2	Set 3
7	5.5	4
4	6	2
6	7	4
3	7	15
5.5	7	16
18	6	1.75
2	7	3
6.5	6.5	6.5

mean, since it can be influenced heavily by outliers. **NOTE:**  
these values in a random sample from a population.



Q1 (25th Percentile)

Q2 (Median)

Q3 (75th Percentile)

= 50th percentile

tion of  
uch

### How-to:

excel formula for variance:

=VAR(select cells)

excel formula for 25th percentile:

=PERCENTILE(select\_cells, 0.25)

excel formula for 50th percentile:

=PERCENTILE(select\_cells, 0.50)

excel formula for 75th percentile:

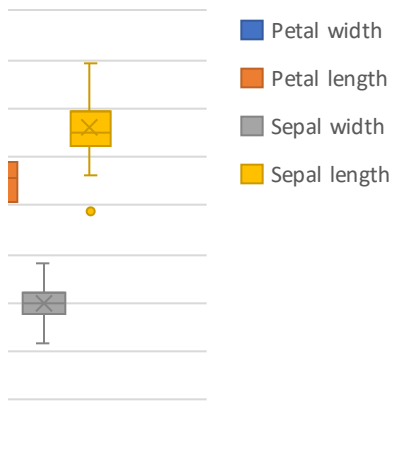
=PERCENTILE(select\_cells, 0.75)

excel formula for median:

=MEDIAN(select cells)

### esented above

ies of Iris



ginica

l line within the box) for  
ow much variability there  
scriptive figure, and

nst eachother?

or two





### How-to

Select data as above

Click insert

Click X Y scatter as your chart type

It is likely you'll need to double check if the variable that you want on your x-axis/y-axis ends up in the right place.

To do this, right click on your chart and select "select data"  
The cells selected in "X values" should be the correct variable for the x-axis

Repeat the previous step for the cells selected for "Y values"

You can add line equations by right clicking on the points of a particular subset (i.e. Versicolor); the same applies to  $R^2$ .

relationship  
gh R-squared

ine) is 0.62. This  
etween Petal  
ance.

