Axis

1D array

| 7 | 2 | 9 | 10 |

axis 0

shape: (4,)

2D array

axis 0

| 5.2 | 3.0 | 4.5 |
| 9.1 | 0.1 | 0.3 |

axis 1

shape: (2, 3)

3D array

shape: (4, 3, 2)

ndarray
ndim = 3
shape = (3, 4, 2)

axis = 2
axis = 1
axis = 0

A[0,0,0]   A[0,3,0]

A[2,0,0]   A[2,3,0]

111	121	141	151	171	181
211	221	241	251	271	281
311	321	341	351	371	381
411	421	441	451	471	481

2

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

σ = population standard deviation

N = the size of the population

x_i = each value from the population

μ = the population mean

	observations: X_i	mean σ	obs-mean	(obs-mean)^2
	17	14	3	9
	15	14	1	1
	23	14	9	81
	7	14	-7	49
	9	14	-5	25
	13	14	-1	1
count/len	6			6
total	84			166
mean/average	14			27.66666667 var
				5.259911279 std

```
np.exp(2)
```

$$e^2$$

Logarithms
What is a Logarithm?

Log$_{10}$ X means...	**...the Power to which 10 is raised to get X**
• $\log_{10}1000 = 3$ because	• $10^3 = 1000$
• $\text{Log}_{10}10,000 = 4$ because	• $10^4 = 10000$
• $\log_{10}10,000,000 = 7$	• $10^7 = 10,000,000$
• $\log_{10}30 = 1.477$ because	• $10^{1.477} = 30$
• $\log_{10}300 = 2.477$ because	• $10^{2.477} = 300$

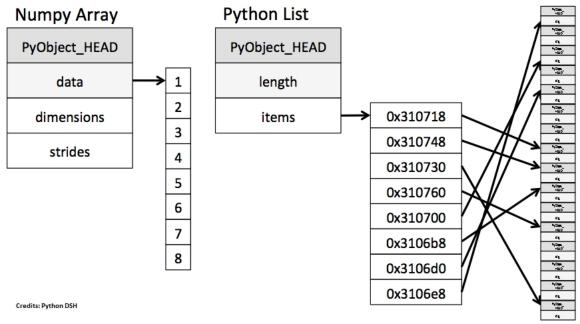So in general ... **$\log_b X = Y$ because $b^Y = X$**
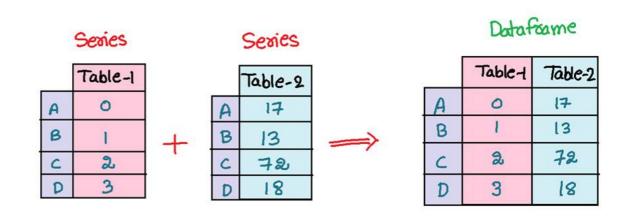
Angle θ		$\sin \theta$	$\cos \theta$	$\tan \theta$
Degrees	Radians			
0	0	0	1	0
30	$\dfrac{\pi}{6}$	$\dfrac{1}{2}$	$\dfrac{\sqrt{3}}{2}$	$\dfrac{1}{\sqrt{3}}$
45	$\dfrac{\pi}{4}$	$\dfrac{1}{\sqrt{2}}$	$\dfrac{1}{\sqrt{2}}$	1
60	$\dfrac{\pi}{3}$	$\dfrac{\sqrt{3}}{2}$	$\dfrac{1}{2}$	$\sqrt{3}$
90	$\dfrac{\pi}{2}$	1	0	undefined
180	π	0	-1	0
270	$\dfrac{3\pi}{2}$	-1	0	undefined
360	2π	0	1	0

Numpy Array

PyObject_HEAD
data
dimensions
strides

1
2
3
4
5
6
7
8

Python List

PyObject_HEAD
length
items

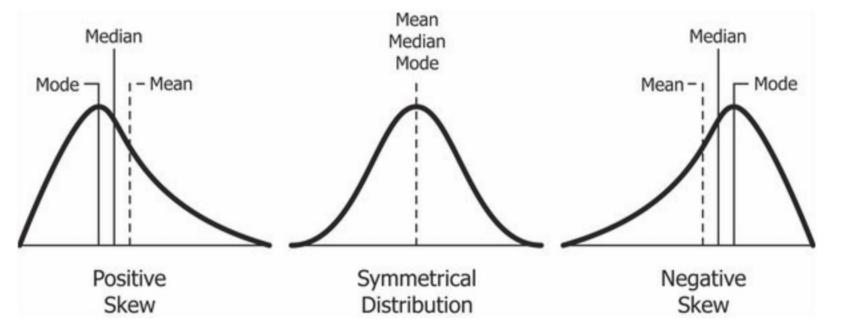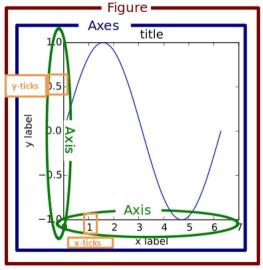| 0x310718 |
|---|
| 0x310748 |
| 0x310730 |
| 0x310760 |
| 0x310700 |
| 0x3106b8 |
| 0x3106d0 |
| 0x3106e8 |

Credits: Python DSH
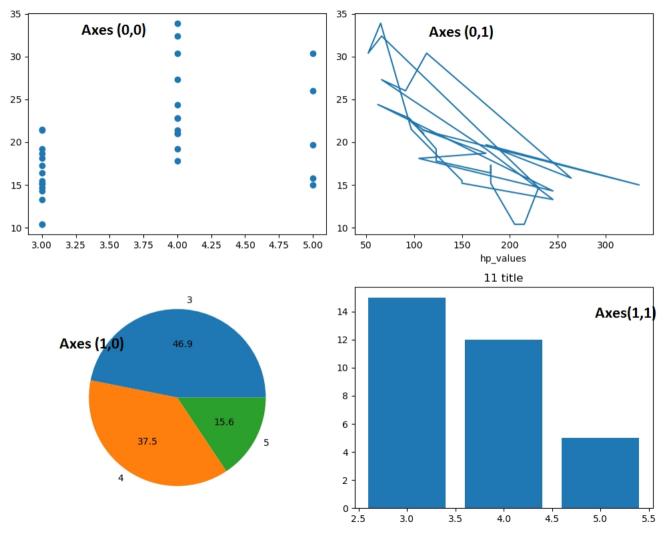
What is in Pandas Package

- Provides Fast, flexible, and expressive data structures.

- Designed to make working with "relational" or "labeled" data both easy and intuitive.

- Aims to be the fundamental high-level building block for doing practical, real world data analysis in Python.

- Series: 1D,   DF : 2D
  - Indices can be Labeled, not just integers
  - Multiple Data types supported.
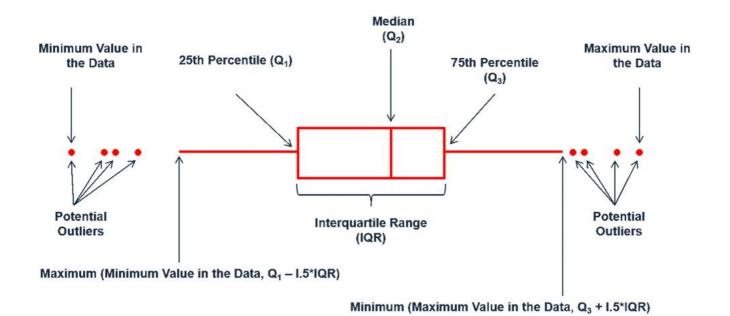
name	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin
volkswagen 1131 deluxe sedan	26.0	4	97.0	46.0	1835	20.5	70	europe
volkswagen super beetle	26.0	4	97.0	46.0	1950	21.0	73	europe
volkswagen rabbit custom diesel	43.1	4	90.0	48.0	1985	21.5	78	europe
vw dasher (diesel)	43.4	4	90.0	48.0	2335	23.7	80	europe
vw rabbit c (diesel)	44.3	4	90.0	48.0	2085	21.7	80	europe
fiat 128	29.0	4	68.0	49.0	1867	19.5	73	europe
chevrolet chevette	29.0	4	85.0	52.0	2035	22.2	76	usa
toyota corona	31.0	4	76.0	52.0	1649	16.5	74	japan
mazda glc deluxe	32.8	4	78.0	52.0	1985	19.4	78	japan
vw pickup	44.0	4	97.0	52.0	2130	24.6	82	europe

| Positive Skew | Symmetrical Distribution | Negative Skew |

Mode — Median — Mean (Positive Skew)

Mean Median Mode (Symmetrical Distribution)

Mean — Median — Mode (Negative Skew)

Axes (0,0)

Axes (0,1)

hp_values

11 title

Axes (1,0)

Axes(1,1)

3

46.9

37.5

15.6

4

5

Box Plot Boundaries

Median
(Q_2)

Minimum Value in the Data

25th Percentile (Q_1)

75th Percentile (Q_3)

Maximum Value in the Data

Potential Outliers

Potential Outliers

Interquartile Range (IQR)

Maximum (Minimum Value in the Data, $Q_1 - 1.5*IQR$)

Minimum (Maximum Value in the Data, $Q_3 + 1.5*IQR$)

13

x Outlier ← Maximum value of the data set

- - - - ← Upper fence = $Q_3 + 1.5 \times IQR$

25% of values

← Maximum value before the upper fence

← Whisker

← 75th percentile, third quartile or Q_3

Interquartile range, IQR

(the box)

$IQR = Q_3 - Q_1$

25% of values

← Median, 50th percentile, second quartile or Q_2

25% of values

← 25th percentile, first quartile or Q_1

← Whisker

25% of values

← Minimum value before the lower fence

- - - - ← Lower fence = $Q_1 - 1.5 \times IQR$

x Outlier

x Outlier ← Minimum value of the data set

Violin Plot

Kurtosis is defined as a measure of 'peakedness'. It is generally measured relative to Normal distribution. (Which means 'excess of kurtosis' is measured)

Mesokurtic

Normal distribution is termed as mesokurtic distribution

Leptokurtic

A leptokurtic distribution has a more acute peak. (positive kurtosis)

Platykurtic

A platykurtic distribution has a flatter peak. (negative kurtosis)

(+) Leptokurtic
(0) Mesokurtic (Normal)
(−) Platykurtic

General Forms of Kurtosis