
knowyourdata Documentation

Release 0.1.0a1

Mubdi Rahman

Mar 15, 2018

Contents:

1	Quickstart	3
2	Installation	5
3	Usage	7
3.1	Returned Parameters	8
4	Indices and tables	9

KnowYourData is a rapid and lightweight module to describe the statistics and structure of data arrays for interactive use. This project was started in 2018 and currently maintained by Mubdi Rahman. This module arose from the regular need to display properties of data arrays while conducting data exploration or diagnostics, for instance, to set min and max values for plotting, or when looking at the first few values in an array don't provide a fair representation of the data.

This module provides a quick way of displaying such information as the mean, median, confidence intervals, and size and shape of the data array.

The simplest way to use KnowYourData is to pass it a numpy array:

```
import numpy as np
from knowyourdata import kyd

# setting x as a numpy array
x = np.random.randn(200)
kyd(x)
```


CHAPTER 1

Quickstart

Installation is most easily done through `pip`, which takes care of all required dependencies:

```
pip install knowyourdata
```

The simplest way to use KnowYourData is to pass it a numpy array:

```
import numpy as np
from knowyourdata import kyd

# setting x as a numpy array
x = np.random.randn(200)
kyd(x)
```


CHAPTER 2

Installation

Installation is most easily done through `pip`, which takes care of all required dependencies:

```
pip install knowyourdata
```


CHAPTER 3

Usage

The simplest way to use KnowYourData is to pass it a numpy array:

```
import numpy as np
from knowyourdata import kyd

# setting x as a numpy array
x = np.random.randn(200)
kyd(x)
```

Basic Statistics			Array Structure	
↪			↪	
↪			↪	
Mean: Min: -2.313 -99 CI: -2.189			Number of Dimensions:	
↪ 1			↪	
0.04288 1Q: -0.6402 -95 CI: -1.969			Shape of Dimensions:	
↪ (200,)			↪	
Median: 0.009476 -68 CI: -0.8657			Array Data Type:	
↪ float64			↪	
Std Dev: 3Q: 0.633 +68 CI: 1.041			Memory Size:	
↪ 1.7KiB			↪	
0.9815 Max: 3.276 +95 CI: 2.075				
↪			↪	
			Number of NaN: 0	
↪			↪	
			Number of Inf: 0	
↪			↪	

or if you are in a jupyter notebook, an HTML version.

The `kyd` function returns a structure that contains the information extracted from the data array. You can access this information through:

```
info_x = kyd(x)
```

```
# The Third Quartile of x:  
print(info_x.thirdquartile)
```

3.1 Returned Parameters

3.1.1 Basic Statistics

All statistics are calculated on a data array filtered for all non-finite elements.

Mean The arithmetic mean as determined by `numpy.mean`:

$$\bar{x} = \frac{1}{N} \sum x$$

Std Dev The standard deviation as determined by `numpy.std`. As is customary in `numpy`:

$$\sigma = \sqrt{\frac{1}{N} \sum (x - \bar{x})^2}$$

Min The minimum of the data array as determined by `numpy.min`.

1Q The first quartile of the data.

Median The median of the data as determined by `numpy.median`.

3Q The third quartile of the data.

Max The maximum of the data as determined by `numpy.max`.

-99 CI, +99 CI The location of the 99% confidence interval.

-95 CI, +95 CI The location of the 95% confidence interval.

-68 CI, +68 CI The location of the 68% confidence interval.

3.1.2 Array Structure

Number of Dimensions The number of dimensions of the data array `numpy.ndim(x)`

Shape of Dimensions The length of each of dimension as determined by `numpy.shape(x)`

Array Data Type The Data Type populating the array.

Memory Size The total size of the array in memory.

Number of NaN The number of Not a Number values in the data array

Number of Inf The number of Infinity values in the data array

CHAPTER 4

Indices and tables

- `genindex`
- `search`