

Pandas



What's **pandas** ?

pandas ?

- 2008: [Wes McKinney](#) at [AQR Capital Management](#) - had a need for a high performance, flexible tool to perform [quantitative analysis](#) on stock data
- Panda is derived from "[panel data](#)", a term for data sets that include observations over multiple time periods for the same individuals

pandas?

- <http://pandas.pydata.org>
- Rich relational data tool built on top of NumPy
 - Like R's `data.frame` on steroids
- Excellent performance
- Easy-to-use, highly consistent API
- A foundation for data analysis in Python

Panda Data Types

- DataFrame (2D)
 - most widely used
 - rows and columns like a spreadsheet
 - easy load from CSV files
- Series (1D)
 - each row/col is a Series
 - widely used in financial calculations

DataFrame

	columns	foo	bar	baz	qux
Index					
A	→	0	x	2.7	True
B	→	4	y	6	True
C	→	8	z	10	False
D	→	-12	w	NA	False
E	→	16	a	18	False

- NumPy array-like
- Each column can have a different type
- Row and column index
- Size mutable: insert and delete columns

Series

index		values
A	→	5
B	→	6
C	→	12
D	→	-5
E	→	6.7

- Subclass of `numpy.ndarray`
- Data: any type
- Index labels need not be ordered
- Duplicates are possible (but result in reduced functionality)

DataFrame

```
In [10]: tips[:10]
Out[10]:
```

	total_bill	tip	sex	smoker	day	time	size
1	16.99	1.01	Female	No	Sun	Dinner	2
2	10.34	1.66	Male	No	Sun	Dinner	3
3	21.01	3.50	Male	No	Sun	Dinner	3
4	23.68	3.31	Male	No	Sun	Dinner	2
5	24.59	3.61	Female	No	Sun	Dinner	4
6	25.29	4.71	Male	No	Sun	Dinner	4
7	8.770	2.00	Male	No	Sun	Dinner	2
8	26.88	3.12	Male	No	Sun	Dinner	4
9	15.04	1.96	Male	No	Sun	Dinner	2
10	14.78	3.23	Male	No	Sun	Dinner	2

DataFrame attributes

Attributes	
<code>T</code>	Transpose index and columns.
<code>at</code>	Access a single value for a row/column label pair.
<code>axes</code>	Return a list representing the axes of the DataFrame.
<code>blocks</code>	(DEPRECATED) Internal property, property synonym for <code>as_blocks()</code>
<code>columns</code>	The column labels of the DataFrame.
<code>dtypes</code>	Return the dtypes in the DataFrame.
<code>empty</code>	Indicator whether DataFrame is empty.
<code>ftypes</code>	Return the ftypes (indication of sparse/dense and dtype) in DataFrame.
<code>iat</code>	Access a single value for a row/column pair by integer position.
<code>iloc</code>	Purely integer-location based indexing for selection by position.
<code>index</code>	The index (row labels) of the DataFrame.
<code>ix</code>	A primarily label-location based indexer, with integer position fallback.
<code>loc</code>	Access a group of rows and columns by label(s) or a boolean array.
<code>ndim</code>	Return an int representing the number of axes / array dimensions.
<code>shape</code>	Return a tuple representing the dimensionality of the DataFrame.
<code>size</code>	Return an int representing the number of elements in this object.
<code>style</code>	Property returning a Styler object containing methods for building a styled HTML representation to the DataFrame.
<code>values</code>	Return a Numpy representation of the DataFrame.

See docs

<https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.html>

GroupBy




