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	

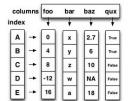
pandas?

- 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



- 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

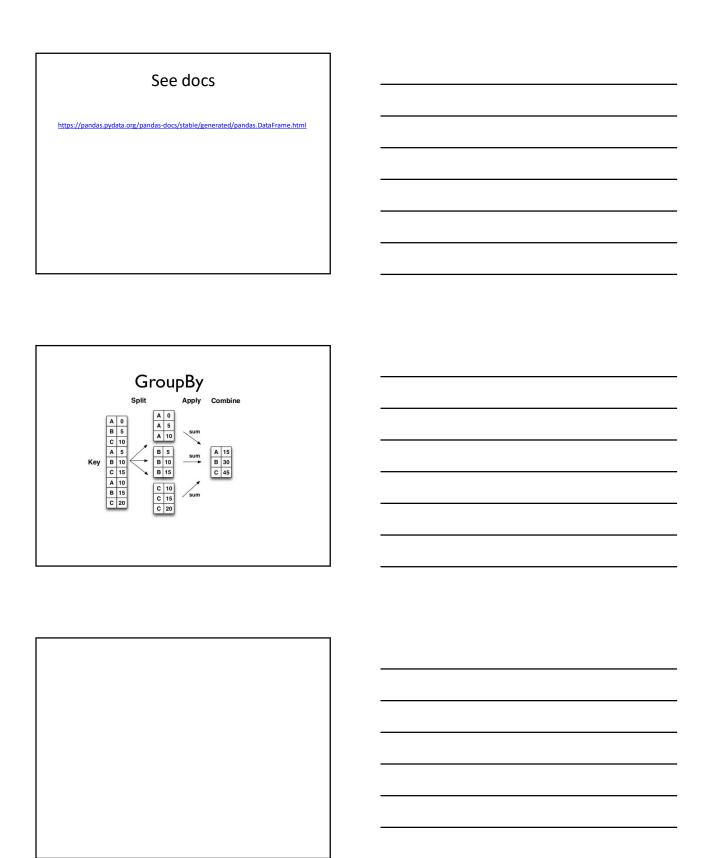
```
In [10]: tips[:10]

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 2
6 25.29 4.71 Male No Sun Dinner 4
7 8.770 2.00 Male No Sun Dinner 4
8 26.88 3.12 Male No Sun Dinner 4
9 15.04 1.95 Male No Sun Dinner 2
10 14.78 3.23 Male No Sun Dinner 2
```

DataFrame attributes

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



-
•

-	
_	
_	