Pandas

- Pandas is one of those packages, and makes importing and analyzing data much easier.
- Pandas builds on packages like NumPy and matplotlib to give you a single, convenient, place to do most of your data analysis and visualization work.

DataFrame

	A	B	C
1	Country	City	Population
2	England	London	8615246
3	Germany	Berlin	3562166
4	Spain	Madrid	3165235
5	Italy	Rome	2874038
6	France	Paris	2273305
7	Austria	Vienna	1805681
8	Romania	Bucharest	1803425
9	Germany	Hamburg	1760433
10	Hungary	Budapest	1754000
11	Poland	Warsaw	1740119
12	Spain	Barcelona	1602386
13	Germany	Munich	1493900
14	Italy	Milan	1350680

Read data

- import pandas as pd

  m = pd.read_csv("movies.csv")

Head and Tail

- Once we read in a DataFrame, Pandas gives us
- two methods that make it fast to print out the data. These functions are:
- – pandas.DataFrame.head – prints the first N rows of a DataFrame. By default 5.
- – pandas.DataFrame.tail – prints the last N rows of aDataFrame. By default 5.
- We’ll use the head method to see what’s in movies:
- m.head()

Find number of rows and columns

```
>>> m.shape
(10, 5)
>>> x = m.shape
>>> type(x)
<type 'tuple'>
>>> x[0]
10
>>> x[1]
5
```

Indexing

- m.iloc[:5,:] – the first 5 rows, and all of the columns for those rows.
- m.iloc[:,:] – the entire DataFrame.
- m.iloc[5:,5:] – rows from position 5 onwards, and columns from position 5 onwards.
- m.iloc[:,0] – the first column, and all of the rows for the column.
- m.iloc[9,:] – the 10th row, and all of the columns for that row.