```
Hansen@Hansens-MacBook-Pro:~/Desktop/Intern_and_Projects/Python_learning $ " ls
Learning_Python_5th_Edition_Mark_Lutz.pdf
Natural Language Processing with Python.pdf
Notebook_Python
OReilly.Python.for.Finance.2014.12.pdf
Oracle PL_SQL Language Pocket Reference, 5th Edition-O'Reilly(2015).pdf
Programming Python 4th editon.pdf
Python for Data Analysis
Python for Data Analysis.pdf
WebCrawler_Python
functional-programming-python.pdf
lpthw
Hansen@Hansens-MacBook-Pro:~/Desktop/Intern_and_Projects/Python_learning $ " ls -al
total 156672
drwxr-xr-x  14 Hansen  staff       448  9  4  2017 .
drwxr-xr-x  32 Hansen  staff      1024  3 18 15:26 ..
-rw-r--r--@  1 Hansen  staff     14340  3 18 10:53 .DS_Store
-rw-r--r--@  1 Hansen  staff  26015425 12 14  2013 Learning_Python_5th_Edition_Mark_Lutz.pdf
-rw-r--r--@  1 Hansen  staff   3580937  7 16  2017 Natural Language Processing with Python.pdf
drwxr-xr-x   8 Hansen  staff       256  7 25  2016 Notebook_Python
-rw-r--r--@  1 Hansen  staff  11002456  3 28  2015 OReilly.Python.for.Finance.2014.12.pdf
-rw-r--r--@  1 Hansen  staff   1493142  1 17  2016 Oracle PL_SQL Language Pocket Reference, 5th Editi
on-O'Reilly(2015).pdf
-rw-r--r--@  1 Hansen  staff  27410567  6  1  2016 Programming Python 4th editon.pdf
drwxr-xr-x@ 30 Hansen  staff       960  3 17 22:34 Python for Data Analysis
-rw-r--r--@  1 Hansen  staff   8703373  3 18 11:38 Python for Data Analysis.pdf
drwxr-xr-x   6 Hansen  staff       192  5 28  2017 WebCrawler_Python
-rw-r--r--@  1 Hansen  staff   1631017  6  1  2016 functional-programming-python.pdf
drwxr-xr-x   4 Hansen  staff       128  7 16  2015 lpthw
Hansen@Hansens-MacBook-Pro:~/Desktop/Intern_and_Projects/Python_learning $ " cd Python\ for\ Data\ An
ysis
Hansen@Hansens-MacBook-Pro:~/Desktop/Intern_and_Projects/Python_learning/Python for Data Analysis $ "
sl
COPYING                 ch04            ch07.ipynb          ch11.ipynb
My_Note                 ch04.ipynb      ch08                ch12.ipynb
README.md               ch05            ch08.ipynb          ch13
appendix_python.ipynb   ch05.ipynb      ch09                fec_study.ipynb
ch02                    ch06            ch09.ipynb
ch02.ipynb              ch06.ipynb      ch10.ipynb
ch03                    ch07            ch11
Hansen@Hansens-MacBook-Pro:~/Desktop/Intern_and_Projects/Python_learning/Python for Data Analysis $ "
d cho2
-bash: cd: cho2: No such file or directory
Hansen@Hansens-MacBook-Pro:~/Desktop/Intern_and_Projects/Python_learning/Python for Data Analysis $ "
d ch02
Hansen@Hansens-MacBook-Pro:~/Desktop/Intern_and_Projects/Python_learning/Python for Data 2 $ " ls/ch0
movielens                               usagov_bitly_data2012-03-16-1331923249.txt
names
Hansen@Hansens-MacBook-Pro:~/Desktop/Intern_and_Projects/Python_learning/Python for Data 2 $ " cd mov
ielens/
Hansen@Hansens-MacBook-Pro:~/Desktop/Intern_and_Projects/Python_learning/Python for Data Analysis/ch0
2/movielens $ " ls
README     movies.dat  ratings.dat users.dat
Hansen@Hansens-MacBook-Pro:~/Desktop/Intern_and_Projects/Python_learning/Python for Data Analysis/ch0
2/movielens $ " clear


Hansen@Hansens-MacBook-Pro:~/Desktop/Intern_and_Projects/Python_learning/Python for Data Analysis/ch0
2/movielens $ " ipython
Python 2.7.12 |Anaconda custom (x86_64)| (default, Jul  2 2016, 17:43:17)
Type "copyright", "credits" or "license" for more information.

IPython 4.2.0 -- An enhanced Interactive Python.
?         -> Introduction and overview of IPython's features.
%quickref -> Quick reference.
help      -> Python's own help system.
object?   -> Details about 'object', use 'object??' for extra details.

In [1]: import pandas as pd
```

```
In [2]: ls
README*      movies.dat*  ratings.dat* users.dat*

In [3]: unames = ['user_id', 'gender', 'age', 'occupation', 'zip']

In [4]: users = pd.read_table('users.dat', sep = '::', header = None, names = rnames)
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-4-3ed10366e8ac> in <module>()
----> 1 users = pd.read_table('users.dat', sep = '::', header = None, names = rnames)

NameError: name 'rnames' is not defined

In [5]: users = pd.read_table('users.dat', sep = '::', header = None, names = unames)
/Users/Hansen/anaconda/bin/ipython:1: ParserWarning: Falling back to the 'python' engine because the
'c' engine does not support regex separators (separators > 1 char and different from '\s+' are interp
reted as regex); you can avoid this warning by specifying engine='python'.
  #!/bin/bash /Users/Hansen/anaconda/bin/python.app

In [6]: users = pd.read_table('users.dat', sep = '::', header = None, names = unames, engine = 'pytho
n')

In [7]: rnames = ['user_id', 'movie_id', 'rating', 'timestamp']

In [8]: ratings = pd.re
pd.read_clipboard  pd.read_gbq       pd.read_msgpack    pd.read_sql_query  pd.reset_option
pd.read_csv        pd.read_hdf       pd.read_pickle     pd.read_sql_table
pd.read_excel      pd.read_html      pd.read_sas        pd.read_stata
pd.read_fwf        pd.read_json      pd.read_sql        pd.read_table

In [8]: ratings = pd.read_table('ratings.dat', sep = '::', header = None, names = rnames, engine = 'p
ython')

In [9]: mnames = ['movie_id', 'title', 'genres']

In [10]: movies = pd.read_table("movies.dat", sep='::', header = None, names = mnames)
/Users/Hansen/anaconda/bin/ipython:1: ParserWarning: Falling back to the 'python' engine because the
'c' engine does not support regex separators (separators > 1 char and different from '\s+' are interp
reted as regex); you can avoid this warning by specifying engine='python'.
  #!/bin/bash /Users/Hansen/anaconda/bin/python.app

In [11]: movies = pd.read_table("movies.dat", sep='::', header = None, names = mnames, engine = 'pyth
on')

In [12]: users[:5]
Out[12]:
   user_id gender  age  occupation    zip
0        1      F    1          10  48067
1        2      M   56          16  70072
2        3      M   25          15  55117
3        4      M   45           7  02460
4        5      M   25          20  55455

In [13]: ratings
Out[13]:
        user_id  movie_id  rating   timestamp
0             1      1193       5   978300760
1             1       661       3   978302109
2             1       914       3   978301968
3             1      3408       4   978300275
4             1      2355       5   978824291
5             1      1197       3   978302268
6             1      1287       5   978302039
7             1      2804       5   978300719
8             1       594       4   978302268
9             1       919       4   978301368
10            1       595       5   978824268
```

```
11             1       938      4  978301752
12             1      2398      4  978302281
13             1      2918      4  978302124
14             1      1035      5  978301753
15             1      2791      4  978302188
16             1      2687      3  978824268
17             1      2018      4  978301777
18             1      3105      5  978301713
19             1      2797      4  978302039
20             1      2321      3  978302205
21             1       720      3  978300760
22             1      1270      5  978300055
23             1       527      5  978824195
24             1      2340      3  978300103
25             1        48      5  978824351
26             1      1097      4  978301953
27             1      1721      4  978300055
28             1      1545      4  978824139
29             1       745      3  978824268
...          ...       ...    ...        ...
1000179     6040      2762      4  956704584
1000180     6040      1036      3  956715455
1000181     6040       508      4  956704972
1000182     6040      1041      4  957717678
1000183     6040      3735      4  960971654
1000184     6040      2791      4  956715569
1000185     6040      2794      1  956716438
1000186     6040       527      5  956704219
1000187     6040      2003      1  956716294
1000188     6040       535      4  964828734
1000189     6040      2010      5  957716795
1000190     6040      2011      4  956716113
1000191     6040      3751      4  964828782
1000192     6040      2019      5  956703977
1000193     6040       541      4  956715288
1000194     6040      1077      5  964828799
1000195     6040      1079      2  956715648
1000196     6040       549      4  956704746
1000197     6040      2020      3  956715288
1000198     6040      2021      3  956716374
1000199     6040      2022      5  956716207
1000200     6040      2028      5  956704519
1000201     6040      1080      4  957717322
1000202     6040      1089      4  956704996
1000203     6040      1090      3  956715518
1000204     6040      1091      1  956716541
1000205     6040      1094      5  956704887
1000206     6040       562      5  956704746
1000207     6040      1096      4  956715648
1000208     6040      1097      4  956715569

[1000209 rows x 4 columns]

In [14]: ratings.size()
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
<ipython-input-14-6eb952cee7c9> in <module>()
----> 1 ratings.size()

TypeError: 'numpy.int64' object is not callable

In [15]: ra
raise          range          ratings        ratings.dat   raw_input

In [15]: rat
ratings        ratings.dat

In [15]: ratings.
```

```
Display all 224 possibilities? (y or n)
ratings.T                      ratings.get_ftype_counts    ratings.replace
ratings.abs                    ratings.get_value           ratings.resample
ratings.add                    ratings.get_values          ratings.reset_index
ratings.add_prefix             ratings.groupby             ratings.rfloordiv
ratings.add_suffix             ratings.gt                  ratings.rmod
ratings.align                  ratings.head                ratings.rmul
ratings.all                    ratings.hist                ratings.rolling
ratings.any                    ratings.iat                 ratings.round
ratings.append                 ratings.icol                ratings.rpow
ratings.apply                  ratings.idxmax              ratings.rsub
ratings.applymap               ratings.idxmin              ratings.rtruediv
ratings.as_blocks              ratings.iget_value          ratings.sample
ratings.as_matrix              ratings.iloc                ratings.select
ratings.asfreq                 ratings.index               ratings.select_dtypes
ratings.assign                 ratings.info                ratings.sem
ratings.astype                 ratings.insert              ratings.set_axis
ratings.at                     ratings.interpolate         ratings.set_index
ratings.at_time                ratings.irow                ratings.set_value
ratings.axes                   ratings.is_copy             ratings.shape
ratings.between_time           ratings.isin                ratings.shift
ratings.bfill                  ratings.isnull              ratings.size
ratings.blocks                 ratings.iteritems           ratings.skew
ratings.bool                   ratings.iterkv              ratings.slice_shift
ratings.boxplot                ratings.iterrows            ratings.sort
ratings.clip                   ratings.itertuples          ratings.sort_index
ratings.clip_lower             ratings.ix                  ratings.sort_values
ratings.clip_upper             ratings.join                ratings.sortlevel
ratings.columns                ratings.keys                ratings.squeeze
ratings.combine                ratings.kurt                ratings.stack
ratings.combineAdd             ratings.kurtosis            ratings.std
ratings.combineMult            ratings.last                ratings.style
ratings.combine_first          ratings.last_valid_index    ratings.sub
ratings.compound               ratings.le                  ratings.subtract
ratings.consolidate            ratings.loc                 ratings.sum
ratings.convert_objects        ratings.lookup              ratings.swapaxes
ratings.copy                   ratings.lt                  ratings.swaplevel
ratings.corr                   ratings.mad                 ratings.tail
ratings.corrwith               ratings.mask                ratings.take
ratings.count                  ratings.max                 ratings.timestamp
ratings.cov                    ratings.mean                ratings.to_clipboard
ratings.cummax                 ratings.median              ratings.to_csv
ratings.cummin                 ratings.memory_usage        ratings.to_dense
ratings.cumprod                ratings.merge               ratings.to_dict
ratings.cumsum                 ratings.min                 ratings.to_excel
ratings.dat                    ratings.mod                 ratings.to_gbq
ratings.describe               ratings.mode                ratings.to_hdf
ratings.diff                   ratings.movie_id            ratings.to_html
ratings.div                    ratings.mul                 ratings.to_json
ratings.divide                 ratings.multiply            ratings.to_latex
ratings.dot                    ratings.ndim                ratings.to_msgpack
ratings.drop                   ratings.ne                  ratings.to_panel
ratings.drop_duplicates        ratings.nlargest            ratings.to_period
ratings.dropna                 ratings.notnull             ratings.to_pickle
ratings.dtypes                 ratings.nsmallest           ratings.to_records
ratings.duplicated             ratings.pct_change          ratings.to_sparse
ratings.empty                  ratings.pipe                ratings.to_sql
ratings.eq                     ratings.pivot               ratings.to_stata
ratings.equals                 ratings.pivot_table         ratings.to_string
ratings.eval                   ratings.plot                ratings.to_timestamp
ratings.ewm                    ratings.pop                 ratings.to_wide
ratings.expanding              ratings.pow                 ratings.to_xarray
ratings.ffill                  ratings.prod                ratings.transpose
ratings.fillna                 ratings.product             ratings.truediv
ratings.filter                 ratings.quantile            ratings.truncate
ratings.first                  ratings.query               ratings.tshift
ratings.first_valid_index      ratings.radd                ratings.tz_convert
ratings.floordiv               ratings.rank                ratings.tz_localize
```

```
ratings.from_csv          ratings.rating          ratings.unstack
ratings.from_dict         ratings.rdiv            ratings.update
ratings.from_items        ratings.reindex         ratings.user_id
ratings.from_records      ratings.reindex_axis    ratings.values
ratings.ftypes            ratings.reindex_like    ratings.var
ratings.ge                ratings.rename          ratings.where
ratings.get               ratings.rename_axis     ratings.xs
ratings.get_dtype_counts  ratings.reorder_levels

In [15]: ratings.size
Out[15]: 4000836

In [16]: data = pd.merge(pd.merge(ratings, users), movies)

In [17]: data
Out[17]:
         user_id  movie_id  rating   timestamp gender  age  occupation   zip  \
0              1      1193       5   978300760      F    1          10  48067
1              2      1193       5   978298413      M   56          16  70072
2             12      1193       4   978220179      M   25          12  32793
3             15      1193       4   978199279      M   25           7  22903
4             17      1193       5   978158471      M   50           1  95350
5             18      1193       4   978156168      F   18           3  95825
6             19      1193       5   982730936      M    1          10  48073
7             24      1193       5   978136709      F   25           7  10023
8             28      1193       3   978125194      F   25           1  14607
9             33      1193       5   978557765      M   45           3  55421
10            39      1193       5   978043535      M   18           4  61820
11            42      1193       3   978038981      M   25           8  24502
12            44      1193       4   978018995      M   45          17  98052
13            47      1193       4   977978345      M   18           4  94305
14            48      1193       4   977975061      M   25           4  92107
15            49      1193       4   978813972      M   18          12  77084
16            53      1193       5   977946400      M   25           0  96931
17            54      1193       5   977944039      M   50           1  56723
18            58      1193       5   977933866      M   25           2  30303
19            59      1193       4   977934292      F   50           1  55413
20            62      1193       4   977968584      F   35           3  98105
21            80      1193       4   977786172      M   56           1  49327
22            81      1193       5   977785864      F   25           0  60640
23            88      1193       5   977694161      F   45           1  02476
24            89      1193       5   977683596      F   56           9  85749
25            95      1193       5   977626632      M   45           0  98201
26            96      1193       3   977621789      F   25          16  78028
27            99      1193       2   982791053      F    1          10  19390
28           102      1193       5  1040737607      M   35          19  20871
29           104      1193       2   977546620      M   25          12  00926
...          ...       ...     ...         ...    ...  ...         ...    ...
1000179     4933      3084       3   962757020      M   25          15  94040
1000180     4802      2218       2  1014866656      M   56           1  40601
1000181     4812      2308       2   962932391      M   18          14  25301
1000182     4874       624       4   962781918      F   25           4  70808
1000183     5059      1434       4   962484364      M   45          16  22652
1000184     5947      1434       4   957190428      F   45          16  97215
1000185     5077      1868       3   962417299      M   25           2  20037
1000186     5944      1868       1   957197520      F   18          10  27606
1000187     5105       404       3   962337582      M   50           7  18977
1000188     5185       404       4   963402617      F   35           4  44485
1000189     5532       404       5   959619841      M   25          17  27408
1000190     5543       404       3   960127592      M   25          17  97401
1000191     5220      2543       3   961546137      M   25           7  91436
1000192     5754      2543       4   958272316      F   18           1  60640
1000193     5227       591       3   961475931      M   18          10  64050
1000194     5795       591       1   958145253      M   25           1  92688
1000195     5313      3656       5   960920392      M   56           0  55406
1000196     5328      2438       4   960838075      F   25           4  91740
1000197     5334      3323       3   960796159      F   56          13  46140
1000198     5334       127       1   960795494      F   56          13  46140
```

```
1000199    5334    3382    5    960796159    F    56         13    46140
1000200    5420    1843    3    960156505    F     1         19    14850
1000201    5433     286    3    960240881    F    35         17    45014
1000202    5494    3530    4    959816296    F    35         17    94306
1000203    5556    2198    3    959445515    M    45          6    92103
1000204    5949    2198    5    958846401    M    18         17    47901
1000205    5675    2703    3    976029116    M    35         14    30030
1000206    5780    2845    1    958153068    M    18         17    92886
1000207    5851    3607    5    957756608    F    18         20    55410
1000208    5938    2909    4    957273353    M    25          1    35401

                                                    title  \
0                 One Flew Over the Cuckoo's Nest (1975)
1                 One Flew Over the Cuckoo's Nest (1975)
2                 One Flew Over the Cuckoo's Nest (1975)
3                 One Flew Over the Cuckoo's Nest (1975)
4                 One Flew Over the Cuckoo's Nest (1975)
5                 One Flew Over the Cuckoo's Nest (1975)
6                 One Flew Over the Cuckoo's Nest (1975)
7                 One Flew Over the Cuckoo's Nest (1975)
8                 One Flew Over the Cuckoo's Nest (1975)
9                 One Flew Over the Cuckoo's Nest (1975)
10                One Flew Over the Cuckoo's Nest (1975)
11                One Flew Over the Cuckoo's Nest (1975)
12                One Flew Over the Cuckoo's Nest (1975)
13                One Flew Over the Cuckoo's Nest (1975)
14                One Flew Over the Cuckoo's Nest (1975)
15                One Flew Over the Cuckoo's Nest (1975)
16                One Flew Over the Cuckoo's Nest (1975)
17                One Flew Over the Cuckoo's Nest (1975)
18                One Flew Over the Cuckoo's Nest (1975)
19                One Flew Over the Cuckoo's Nest (1975)
20                One Flew Over the Cuckoo's Nest (1975)
21                One Flew Over the Cuckoo's Nest (1975)
22                One Flew Over the Cuckoo's Nest (1975)
23                One Flew Over the Cuckoo's Nest (1975)
24                One Flew Over the Cuckoo's Nest (1975)
25                One Flew Over the Cuckoo's Nest (1975)
26                One Flew Over the Cuckoo's Nest (1975)
27                One Flew Over the Cuckoo's Nest (1975)
28                One Flew Over the Cuckoo's Nest (1975)
29                One Flew Over the Cuckoo's Nest (1975)
...                                                  ...
1000179                                 Home Page (1999)
1000180                          Juno and Paycock (1930)
1000181                              Detroit 9000 (1973)
1000182                             Condition Red (1995)
1000183                             Stranger, The (1994)
1000184                             Stranger, The (1994)
1000185                                Truce, The (1996)
1000186                                Truce, The (1996)
1000187   Brother Minister: The Assassination of Malcolm...
1000188   Brother Minister: The Assassination of Malcolm...
1000189   Brother Minister: The Assassination of Malcolm...
1000190   Brother Minister: The Assassination of Malcolm...
1000191                         Six Ways to Sunday (1997)
1000192                         Six Ways to Sunday (1997)
1000193                          Tough and Deadly (1995)
1000194                          Tough and Deadly (1995)
1000195                                     Lured (1947)
1000196                             Outside Ozona (1998)
1000197                             Chain of Fools (2000)
1000198   Silence of the Palace, The (Saimt el Qusur) (1...
1000199                            Song of Freedom (1936)
1000200                   Slappy and the Stinkers (1998)
1000201                         Nemesis 2: Nebula (1995)
1000202                         Smoking/No Smoking (1993)
1000203                               Modulations (1998)
```

```
1000204                               Modulations (1998)
1000205                            Broken Vessels (1998)
1000206                               White Boys (1999)
1000207                        One Little Indian (1973)
1000208    Five Wives, Three Secretaries and Me (1998)

                               genres
0                               Drama
1                               Drama
2                               Drama
3                               Drama
4                               Drama
5                               Drama
6                               Drama
7                               Drama
8                               Drama
9                               Drama
10                              Drama
11                              Drama
12                              Drama
13                              Drama
14                              Drama
15                              Drama
16                              Drama
17                              Drama
18                              Drama
19                              Drama
20                              Drama
21                              Drama
22                              Drama
23                              Drama
24                              Drama
25                              Drama
26                              Drama
27                              Drama
28                              Drama
29                              Drama
...                               ...
1000179                   Documentary
1000180                         Drama
1000181                  Action|Crime
1000182         Action|Drama|Thriller
1000183                        Action
1000184                        Action
1000185                     Drama|War
1000186                     Drama|War
1000187                   Documentary
1000188                   Documentary
1000189                   Documentary
1000190                   Documentary
1000191                        Comedy
1000192                        Comedy
1000193         Action|Drama|Thriller
1000194         Action|Drama|Thriller
1000195                         Crime
1000196                Drama|Thriller
1000197                  Comedy|Crime
1000198                         Drama
1000199                         Drama
1000200              Children's|Comedy
1000201         Action|Sci-Fi|Thriller
1000202                        Comedy
1000203                   Documentary
1000204                   Documentary
1000205                         Drama
1000206                         Drama
1000207         Comedy|Drama|Western
1000208                   Documentary
```

```
[1000209 rows x 10 columns]

In [18]: data.ix[0]
Out[18]:
user_id                                        1
movie_id                                    1193
rating                                         5
timestamp                              978300760
gender                                         F
age                                            1
occupation                                    10
zip                                        48067
title          One Flew Over the Cuckoo's Nest (1975)
genres                                     Drama
Name: 0, dtype: object

In [19]: data.ix?
Type:        property
String form: <property object at 0x112e0c788>
Docstring:
A primarily label-location based indexer, with integer position
fallback.

``.ix[]`` supports mixed integer and label based access. It is
primarily label based, but will fall back to integer positional
access unless the corresponding axis is of integer type.

``.ix`` is the most general indexer and will support any of the
inputs in ``.loc`` and ``.iloc``. ``.ix`` also supports floating
point label schemes. ``.ix`` is exceptionally useful when dealing
with mixed positional and label based hierachical indexes.

However, when an axis is integer based, ONLY label based access
and not positional access is supported. Thus, in such cases, it's
usually better to be explicit and use ``.iloc`` or ``.loc``.

See more at :ref:`Advanced Indexing <advanced>`.

In [20]: mean_ratings = data.pivot_table('rating', row = 'title', cols = 'gender', aggfunc='mean')
---------------------------------------------------------------------
TypeError                              Traceback (most recent call last)
<ipython-input-20-7a3b42bb1510> in <module>()
----> 1 mean_ratings = data.pivot_table('rating', row = 'title', cols = 'gender', aggfunc='mean')

TypeError: pivot_table() got an unexpected keyword argument 'row'

In [21]: mean_ratings = data.pivot_table('rating', rows = 'title', cols = 'gender', aggfunc='mean')
----------------------------------------------------------------------
TypeError                              Traceback (most recent call last)
<ipython-input-21-818478d17263> in <module>()
----> 1 mean_ratings = data.pivot_table('rating', rows = 'title', cols = 'gender', aggfunc='mean')

TypeError: pivot_table() got an unexpected keyword argument 'rows'

In [22]: mean_ratings = data.pivot_table('rating', index = 'title', columns = 'gender', aggfunc='mean')

In [23]: mean_ratings[:5]
Out[23]:
gender                          F         M
title
$1,000,000 Duck (1971)       3.375000  2.761905
'Night Mother (1986)         3.388889  3.352941
'Til There Was You (1997)    2.675676  2.733333
'burbs, The (1989)           2.793478  2.962085
...And Justice for All (1979) 3.828571  3.689024
```

```
In [24]: ratings_by_title = data.groupby('title').size()

In [25]: ra
raise              ratings             ratings_by_title
range              ratings.dat         raw_input

In [25]: ratings_by_title[:10]
Out[25]:
title
$1,000,000 Duck (1971)                37
'Night Mother (1986)                  70
'Til There Was You (1997)             52
'burbs, The (1989)                   303
...And Justice for All (1979)        199
1-900 (1994)                           2
10 Things I Hate About You (1999)    700
101 Dalmatians (1961)                565
101 Dalmatians (1996)                364
12 Angry Men (1957)                  616
dtype: int64

In [26]: active_titles = ratings_by_title.index[ratings_by_title >= 250]

In [27]: active_titles[:10]
Out[27]:
Index([u''burbs, The (1989)', u'10 Things I Hate About You (1999)',
       u'101 Dalmatians (1961)', u'101 Dalmatians (1996)',
       u'12 Angry Men (1957)', u'13th Warrior, The (1999)',
       u'2 Days in the Valley (1996)', u'20,000 Leagues Under the Sea (1954)',
       u'2001: A Space Odyssey (1968)', u'2010 (1984)'],
      dtype='object', name=u'title')

In [28]: active_titles.size
Out[28]: 1216

In [29]: active_titles
Out[29]:
Index([u''burbs, The (1989)', u'10 Things I Hate About You (1999)',
       u'101 Dalmatians (1961)', u'101 Dalmatians (1996)',
       u'12 Angry Men (1957)', u'13th Warrior, The (1999)',
       u'2 Days in the Valley (1996)', u'20,000 Leagues Under the Sea (1954)',
       u'2001: A Space Odyssey (1968)', u'2010 (1984)',
       ...
       u'X-Men (2000)', u'Year of Living Dangerously (1982)',
       u'Yellow Submarine (1968)', u'You've Got Mail (1998)',
       u'Young Frankenstein (1974)', u'Young Guns (1988)',
       u'Young Guns II (1990)', u'Young Sherlock Holmes (1985)',
       u'Zero Effect (1998)', u'eXistenZ (1999)'],
      dtype='object', name=u'title', length=1216)

In [30]: mean_ratings = mean_ratings.ix[active_titles]

In [31]: mean_ratings
Out[31]:
gender                                       F        M
title
'burbs, The (1989)                    2.793478  2.962085
10 Things I Hate About You (1999)     3.646552  3.311966
101 Dalmatians (1961)                 3.791444  3.500000
101 Dalmatians (1996)                 3.240000  2.911215
12 Angry Men (1957)                   4.184397  4.328421
13th Warrior, The (1999)              3.112000  3.168000
2 Days in the Valley (1996)           3.488889  3.244813
20,000 Leagues Under the Sea (1954)   3.670103  3.709205
2001: A Space Odyssey (1968)          3.825581  4.129738
2010 (1984)                           3.446809  3.413712
28 Days (2000)                        3.209424  2.977707
39 Steps, The (1935)                  3.965517  4.107692
```

```
54 (1998)                                             2.701754  2.782178
7th Voyage of Sinbad, The (1958)                      3.409091  3.658879
8MM (1999)                                            2.906250  2.850962
About Last Night... (1986)                            3.188679  3.140909
Absent Minded Professor, The (1961)                   3.469388  3.446809
Absolute Power (1997)                                 3.469136  3.327759
Abyss, The (1989)                                     3.659236  3.689507
Ace Ventura: Pet Detective (1994)                     3.000000  3.197917
Ace Ventura: When Nature Calls (1995)                 2.269663  2.543333
Addams Family Values (1993)                           3.000000  2.878531
Addams Family, The (1991)                             3.186170  3.163498
Adventures in Babysitting (1987)                      3.455782  3.208122
Adventures of Buckaroo Bonzai Across the 8th Di...    3.308511  3.402321
Adventures of Priscilla, Queen of the Desert, T...    3.989071  3.688811
Adventures of Robin Hood, The (1938)                  4.166667  3.918367
African Queen, The (1951)                             4.324232  4.223822
Age of Innocence, The (1993)                          3.827068  3.339506
Agnes of God (1985)                                   3.534884  3.244898
...                                                        ...       ...
White Men Can't Jump (1992)                           3.028777  3.231061
Who Framed Roger Rabbit? (1988)                       3.569378  3.713251
Who's Afraid of Virginia Woolf? (1966)                4.029703  4.096939
Whole Nine Yards, The (2000)                          3.296552  3.404814
Wild Bunch, The (1969)                                3.636364  4.128099
Wild Things (1998)                                    3.392000  3.459082
Wild Wild West (1999)                                 2.275449  2.131973
William Shakespeare's Romeo and Juliet (1996)         3.532609  3.318644
Willow (1988)                                         3.658683  3.453543
Willy Wonka and the Chocolate Factory (1971)          4.063953  3.789474
Witness (1985)                                        4.115854  3.941504
Wizard of Oz, The (1939)                              4.355030  4.203138
Wolf (1994)                                           3.074074  2.899083
Women on the Verge of a Nervous Breakdown (1988)      3.934307  3.865741
Wonder Boys (2000)                                    4.043796  3.913649
Working Girl (1988)                                   3.606742  3.312500
World Is Not Enough, The (1999)                       3.337500  3.388889
Wrong Trousers, The (1993)                            4.588235  4.478261
Wyatt Earp (1994)                                     3.147059  3.283898
X-Files: Fight the Future, The (1998)                 3.489474  3.493797
X-Men (2000)                                          3.682310  3.851702
Year of Living Dangerously (1982)                     3.951220  3.869403
Yellow Submarine (1968)                               3.714286  3.689286
You've Got Mail (1998)                                3.542424  3.275591
Young Frankenstein (1974)                             4.289963  4.239177
Young Guns (1988)                                     3.371795  3.425620
Young Guns II (1990)                                  2.934783  2.904025
Young Sherlock Holmes (1985)                          3.514706  3.363344
Zero Effect (1998)                                    3.864407  3.723140
eXistenZ (1999)                                       3.098592  3.289086

[1216 rows x 2 columns]

In [32]: top_female_ratings = mean_ratings.sort_index(by = 'F', ascending = False)
/Users/Hansen/anaconda/bin/ipython:1: FutureWarning: by argument to sort_index is deprecated, pls use
  .sort_values(by=...)
  #!/bin/bash /Users/Hansen/anaconda/bin/python.app

In [33]: top_female_ratings = mean_ratings.sort_values(by = 'F', ascending = False)

In [34]: top_female_ratings[:10}
  File "<ipython-input-34-f39e8306b1fb>", line 1
    top_female_ratings[:10}
                          ^
SyntaxError: invalid syntax


In [35]: top_female_ratings[:10]
Out[35]:
```

```
gender                                                  F        M
title
Close Shave, A (1995)                            4.644444  4.473795
Wrong Trousers, The (1993)                       4.588235  4.478261
Sunset Blvd. (a.k.a. Sunset Boulevard) (1950)    4.572650  4.464589
Wallace & Gromit: The Best of Aardman Animation...  4.563107  4.385075
Schindler's List (1993)                          4.562602  4.491415
Shawshank Redemption, The (1994)                 4.539075  4.560625
Grand Day Out, A (1992)                          4.537879  4.293255
To Kill a Mockingbird (1962)                     4.536667  4.372611
Creature Comforts (1990)                         4.513889  4.272277
Usual Suspects, The (1995)                       4.513317  4.518248


In [36]: mean_ratings['diff'] = mean_ratings['M'] - mean_ratings['F']

In [37]: sorted_by_diff = mean_ratings.sort_value(by = 'diff')
---------------------------------------------------------------------
AttributeError                         Traceback (most recent call last)
<ipython-input-37-1136462f41db> in <module>()
----> 1 sorted_by_diff = mean_ratings.sort_value(by = 'diff')

/Users/Hansen/anaconda/lib/python2.7/site-packages/pandas/core/generic.pyc in __getattr__(self, name)
   2670             if name in self._info_axis:
   2671                 return self[name]
-> 2672             return object.__getattribute__(self, name)
   2673
   2674     def __setattr__(self, name, value):

AttributeError: 'DataFrame' object has no attribute 'sort_value'

In [38]: sorted_by_diff = mean_ratings.sort_values(by = 'diff')

In [39]: sorted_by_diff
Out[39]:
gender                                   F         M       diff
title
Dirty Dancing (1987)                  3.790378  2.959596 -0.830782
Jumpin' Jack Flash (1986)             3.254717  2.578358 -0.676359
Grease (1978)                         3.975265  3.367041 -0.608224
Little Women (1994)                   3.870588  3.321739 -0.548849
Steel Magnolias (1989)                3.901734  3.365957 -0.535777
Anastasia (1997)                      3.800000  3.281609 -0.518391
Rocky Horror Picture Show, The (1975) 3.673016  3.160131 -0.512885
Color Purple, The (1985)              4.158192  3.659341 -0.498851
Age of Innocence, The (1993)          3.827068  3.339506 -0.487561
Free Willy (1993)                     2.921348  2.438776 -0.482573
French Kiss (1995)                    3.535714  3.056962 -0.478752
Little Shop of Horrors, The (1960)    3.650000  3.179688 -0.470312
Guys and Dolls (1955)                 4.051724  3.583333 -0.468391
Mary Poppins (1964)                   4.197740  3.730594 -0.467147
Patch Adams (1998)                    3.473282  3.008746 -0.464536
Grease 2 (1982)                       2.243478  1.792553 -0.450925
Sound of Music, The (1965)            4.233677  3.783418 -0.450259
Never Been Kissed (1999)              3.452174  3.002538 -0.449636
Casper (1995)                         3.291139  2.844444 -0.446695
Jewel of the Nile, The (1985)         3.494253  3.048096 -0.446157
South Pacific (1958)                  3.887850  3.445087 -0.442764
Gone with the Wind (1939)             4.269841  3.829371 -0.440471
Sommersby (1993)                      3.482270  3.046667 -0.435603
Skulls, The (2000)                    3.053333  2.621145 -0.432188
Corrina, Corrina (1994)               3.523490  3.092593 -0.430897
Iron Eagle (1986)                     2.968750  2.543933 -0.424817
Hope Floats (1998)                    3.206186  2.782383 -0.423802
Beverly Hills Cop III (1994)          2.865385  2.443333 -0.422051
Sense and Sensibility (1995)          4.233333  3.819277 -0.414056
Parent Trap, The (1961)               3.805556  3.393333 -0.412222
...                                        ...       ...       ...
Tommy Boy (1995)                      3.170732  3.585366  0.414634
```

```
Eyes Wide Shut (1999)                              2.891213  3.305949  0.414736
South Park: Bigger, Longer and Uncut (1999)  3.422481  3.846686  0.424206
Godzilla (Gojira) (1954)                           2.948718  3.374046  0.425328
Raging Bull (1980)                                 3.831933  4.263441  0.431508
Reservoir Dogs (1992)                              3.769231  4.213873  0.444642
Where Eagles Dare (1969)                           3.384615  3.833333  0.448718
Kingpin (1996)                                     3.092308  3.557480  0.465173
Me, Myself and Irene (2000)                        2.629310  3.096847  0.467537
Akira (1988)                                       3.511111  3.980344  0.469233
Wild Bunch, The (1969)                             3.636364  4.128099  0.491736
From Dusk Till Dawn (1996)                         2.735714  3.232558  0.496844
Big Trouble in Little China (1986)                 2.987952  3.485030  0.497078
Beavis and Butt-head Do America (1996)             2.637363  3.135417  0.498054
Rocky II (1979)                                    2.741379  3.242894  0.501515
Barb Wire (1996)                                   1.585366  2.100386  0.515020
Fright Night (1985)                                2.973684  3.500000  0.526316
Exorcist, The (1973)                               3.537634  4.067239  0.529605
Animal House (1978)                                3.628906  4.167192  0.538286
Porky's (1981)                                     2.296875  2.836364  0.539489
For a Few Dollars More (1965)                      3.409091  3.953795  0.544704
Caddyshack (1980)                                  3.396135  3.969737  0.573602
Rocky III (1982)                                   2.361702  2.943503  0.581801
Hidden, The (1987)                                 3.137931  3.745098  0.607167
Evil Dead II (Dead By Dawn) (1987)                 3.297297  3.909283  0.611985
Cable Guy, The (1996)                              2.250000  2.863787  0.613787
Longest Day, The (1962)                            3.411765  4.031447  0.619682
Dumb & Dumber (1994)                               2.697987  3.336595  0.638608
Kentucky Fried Movie, The (1977)                   2.878788  3.555147  0.676359
Good, The Bad and The Ugly, The (1966)             3.494949  4.221300  0.726351

[1216 rows x 3 columns]

In [40]: ratings_std_by_title = data.groupby('title')['rating'].std()

In [41]: ratings_std_by_title = ratings_
ratings_by_title     ratings_std_by_title

In [41]: ratings_std_by_title = ratings_std_by_title.ix[active_titles]

In [42]: ra
raise                ratings              ratings_by_title      raw_input
range                ratings.dat          ratings_std_by_title

In [42]: ratings_std_by_title.order(ascending=False)[:10]
/Users/Hansen/anaconda/bin/ipython:1: FutureWarning: order is deprecated, use sort_values(...)
  #!/bin/bash /Users/Hansen/anaconda/bin/python.app
Out[42]:
title
Dumb & Dumber (1994)                    1.321333
Blair Witch Project, The (1999)         1.316368
Natural Born Killers (1994)             1.307198
Tank Girl (1995)                        1.277695
Rocky Horror Picture Show, The (1975)   1.260177
Eyes Wide Shut (1999)                   1.259624
Evita (1996)                            1.253631
Billy Madison (1995)                    1.249970
Fear and Loathing in Las Vegas (1998)   1.246408
Bicentennial Man (1999)                 1.245533
Name: rating, dtype: float64

In [43]:
```