MANIPULATING DATAFRAMES WITH PANDAS

Pivoting DataFrames

Clinical trials data

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

In [2]: trials = pd.read_csv('trials_01.csv')

In [3]: print(trials)
   id treatment gender  response
0   1         A      F         5
1   2         A      M         3
2   3         B      F         8
3   4         B      M         9
```

Reshaping by pivoting

```
In [4]: trials.pivot(index='treatment',
   ...:              columns='gender',
   ...:              values='response')

Out[4]:
gender       F  M
treatment
A            5  3
B            8  9
```

Pivoting multiple columns

```
In [5]: trials.pivot(index='treatment', columns='gender')
Out[5]:
             id          response
gender       F   M          F   M
treatment
A            1   2          5   3
B            3   4          8   9
```

MANIPULATING DATAFRAMES WITH PANDAS

Let’s practice!

MANIPULATING DATAFRAMES WITH PANDAS

Stacking & unstacking DataFrames

Creating a multi-level index

```
In [1]: print(trials)
   id treatment gender  response
0   1         A      F         5
1   2         A      M         3
2   3         B      F         8
3   4         B      M         9

In [2]: trials = trials.set_index(['treatment', 'gender'])

In [3]: print(trials)
                   id  response
treatment gender
A         F         1         5
          M         2         3
B         F         3         8
          M         4         9
```

Unstacking a multi-index (1)

```
In [4]: print(trials)
                    id  response
treatment gender
A         F          1         5
          M          2         3
B         F          3         8
          M          4         9

In [5]: trials.unstack(level='gender')
Out[5]:
           id    response
gender      F  M        F  M
treatment
A           1  2        5  3
B           3  4        8  9
```

Unstacking a multi-index (2)

```
In [6]: print(trials)
                    id  response
treatment gender
A         F          1         5
          M          2         3
B         F          3         8
          M          4         9

In [7]: trials.unstack(level=1)
Out[7]:
           id    response
gender      F  M        F  M
treatment
A           1  2        5  3
B           3  4        8  9
```

Stacking DataFrames

```
In [8]: trials_by_gender = trials.unstack(level='gender')

In [9]: trials_by_gender
Out[9]:
           id    response
gender      F  M        F  M
treatment
A           1  2        5  3
B           3  4        8  9

In [10]: trials_by_gender.stack(level='gender')
Out[10]:
                   id  response
treatment gender
A         F         1         5
          M         2         3
B         F         3         8
          M         4         9
```

Stacking DataFrames

```
In [11]: stacked = trials_by_gender.stack(level='gender')

In [12]: stacked
Out[12]:
                    id  response
treatment gender
A         F          1         5
          M          2         3
B         F          3         8
          M          4         9
```

Swapping levels

```
In [13]: swapped = stacked.swaplevel(0, 1)

In [14]: print(swapped)

                  id  response
gender treatment
F      A           1         5
M      A           2         3
F      B           3         8
M      B           4         9
```

Sorting rows

```
In [15]: sorted_trials = swapped.sort_index()

In [16]: print(sorted_trials)
                    id  response
gender treatment
F      A             1         5
       B             3         8
M      A             2         3
       B             4         9
```

MANIPULATING DATAFRAMES WITH PANDAS

Let’s practice!

MANIPULATING DATAFRAMES WITH PANDAS

Melting DataFrames

Clinical trials data

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

In [2]: trials = pd.read_csv('trials_01.csv')

In [3]: print(trials)
   id treatment gender  response
0   1         A      F         5
1   2         A      M         3
2   3         B      F         8
3   4         B      M         9
```

Clinical trials after pivoting

```
In [4]: trials.pivot(index='treatment',
   ...:              columns='gender',
   ...:              values='response')
Out[4]:
gender      F  M
treatment
A           5  3
B           8  9
```

Clinical trials data

```
In [5]: new_trials = pd.read_csv('trials_02.csv')

In [6]: print(new_trials)
  treatment  F  M
0         A  5  3
1         B  8  9
```

Melting DataFrame

```
In [7]: pd.melt(new_trials)
Out[7]:
    variable value
0  treatment     A
1  treatment     B
2          F     5
3          F     8
4          M     3
5          M     9
```

Specifying id_vars

```
In [8]: pd.melt(new_trials, id_vars=['treatment'])
Out[8]:
  treatment variable  value
0         A        F      5
1         B        F      8
2         A        M      3
3         B        M      9
```

Specifying value_vars

```
In [9]: pd.melt(new_trials, id_vars=['treatment'],
   ...:         value_vars=['F', 'M'])
Out[9]:
  treatment variable  value
0         A        F      5
1         B        F      8
2         A        M      3
3         B        M      9
```

Specifying value_name

```
In [10]: pd.melt(new_trials, id_vars=['treatment'],
    ...:          var_name='gender', value_name='response')
Out[10]:
  treatment gender  response
0         A      F         5
1         B      F         8
2         A      M         3
3         B      M         9
```

MANIPULATING DATAFRAMES WITH PANDAS

Let’s practice!

MANIPULATING DATAFRAMES WITH PANDAS

Pivot tables

More clinical trials data

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

In [2]: more_trials = pd.read_csv('trials_03.csv')

In [3]: print(more_trials)
   id treatment gender  response
0   1         A      F         5
1   2         A      M         3
2   3         A      M         8
3   4         A      F         9
4   5         B      F         1
5   6         B      M         8
6   7         B      F         4
7   8         B      F         6
```

Rearranging by pivoting

```
In [4]: more_trials.pivot(index='treatment',
   ...:                   columns='gender',
   ...:                   values='response')
---------------------------------------------------------------------------
ValueError: Index contains duplicate entries, cannot reshape
```

Pivot table

```
In [5]: more_trials.pivot_table(index='treatment',
   ...:                         columns='gender',
   ...:                         values='response')
Out[5]:
gender                F    M
treatment
A             7.000000  5.5
B             3.666667  8.0
```

Other aggregations

```
In [6]: more_trials.pivot_table(index='treatment',
   ...:                         columns='gender',
   ...:                         values='response',
   ...:                         aggfunc='count')
Out[6]:
gender      F  M
treatment
A           2  2
B           3  1
```

MANIPULATING DATAFRAMES WITH PANDAS

Let’s practice!