

# Pivoting DataFrames

MANIPULATING DATAFRAMES WITH PANDAS



**Anaconda**  
Instructor

# Clinical trials data

```
import pandas as pd
trials = pd.read_csv('trials_01.csv')
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

```
trials.pivot(index='treatment',  
              columns='gender',  
              values='response')
```

gender	F	M
treatment		
A	5	3
B	8	9

# Pivoting multiple columns

```
trials.pivot(index='treatment', columns='gender')
```

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

# Let's practice!

MANIPULATING DATAFRAMES WITH PANDAS

# Stacking & unstacking DataFrames

MANIPULATING DATAFRAMES WITH PANDAS



**Anaconda**  
Instructor

# Creating a multi-level index

```
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
```

```
trials = trials.set_index(['treatment', 'gender'])
print(trials)
```

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

# Unstacking a multi-index

```
print(trials)
```

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

```
trials.unstack(level='gender')
```

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



# Unstacking a multi-index

```
print(trials)
```

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

```
trials.unstack(level=1)
```

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

# Stacking DataFrames

```
trials_by_gender = trials.unstack(level='gender')
trials_by_gender
```

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

```
trials_by_gender.stack(level='gender')
```

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

# Stacking DataFrames

```
stacked = trials_by_gender.stack(level='gender')  
stacked
```

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

# Swapping levels

```
swapped = stacked.swaplevel(0, 1)
print(swapped)
```

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

# Sorting rows

```
sorted_trials = swapped.sort_index()  
print(sorted_trials)
```

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

# Let's practice!

MANIPULATING DATAFRAMES WITH PANDAS

# Melting DataFrames

MANIPULATING DATAFRAMES WITH PANDAS



**Anaconda**  
Instructor

# Clinical trials data

```
import pandas as pd
trials = pd.read_csv('trials_01.csv')
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

```
trials.pivot(index='treatment',  
              columns='gender',  
              values='response')
```

gender	F	M
treatment		
A	5	3
B	8	9

# Clinical trials data

```
new_trials = pd.read_csv('trials_02.csv')  
print(new_trials)
```

	treatment	F	M
0	A	5	3
1	B	8	9

# Melting DataFrame

```
pd.melt(new_trials)
```

	variable	value
0	treatment	A
1	treatment	B
2	F	5
3	F	8
4	M	3
5	M	9

# Specifying id\_vars

```
pd.melt(new_trials, id_vars=['treatment'])
```

	treatment	variable	value
0	A	F	5
1	B	F	8
2	A	M	3
3	B	M	9

# Specifying value\_vars

```
pd.melt(new_trials, id_vars=['treatment'],  
        value_vars=['F', 'M'])
```

	treatment	variable	value
0	A	F	5
1	B	F	8
2	A	M	3
3	B	M	9

# Specifying value\_name

```
pd.melt(new_trials, id_vars=['treatment'],  
        var_name='gender',  
        value_name='response')
```

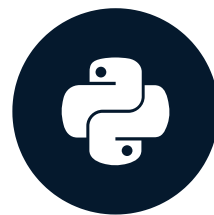
	treatment	gender	response
0	A	F	5
1	B	F	8
2	A	M	3
3	B	M	9

# Let's practice!

MANIPULATING DATAFRAMES WITH PANDAS

# Pivot tables

MANIPULATING DATAFRAMES WITH PANDAS



**Anaconda**  
Instructor



# More clinical trials data

```
import pandas as pd
more_trials = pd.read_csv('trials_03.csv')
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

```
more_trials.pivot(index='treatment',  
                  columns='gender',  
                  values='response')
```

```
ValueError: Index contains duplicate entries, cannot reshape
```

# Pivot table

```
more_trials.pivot_table(index='treatment',  
                          columns='gender',  
                          values='response')
```

gender	F	M
treatment		
A	7.000000	5.5
B	3.666667	8.0

# Other aggregations

```
more_trials.pivot_table(index='treatment',  
                          columns='gender',  
                          values='response',  
                          aggfunc='count')
```

gender	F	M
treatment		
A	2	2
B	3	1

# Let's practice!

MANIPULATING DATAFRAMES WITH PANDAS