

C: Pickling and stack, unstack, pivot and melt

a290

1. Pickling (feature of Python in general):

- Saves an **internal copy of a data object to disk**
- Faster to reload than rebuilding from text
- Not human-readable

```
data.to_pickle("filename.pkl")    Standard pickle  
data = pd.read_pickle("filename.pkl")
```

```
data.to_pickle("filename.zip")    Adds compression  
data = pd.read_pickle("filename.zip")
```

- Pickling can be done for other Python objects as well, not just Pandas

2. Unstacking and stacking (feature of Pandas):

- Converts an **index level** to a **set of columns**

Example:

- Series `expend_long`
- Indexed by year and state

Year	State	
2018	NY	335
2019	NY	400
2020	NY	290
2018	CA	349
2019	CA	326
2020	CA	228
2018	FL	174

2019	FL	290
2020	FL	176

```
expend_wide = expend_long.unstack('State')
```

- Dataframe expend_wide
- Indexed by year
- Columns for states

Year	NY	CA	FL
2018	335	349	174
2019	400	326	290
2020	290	228	176

- To go the other direction use .stack():

```
new_long = expend_wide.stack()
```

3. Pivoting (feature of Pandas):

- Converts long to wide form using columns only, not the index

Example:

raw_data:

Index	Year	State	Alpha	Obs
0	2018	NY	335	1
1	2019	NY	400	2
2	2020	NY	290	3
3	2018	CA	349	1
4	2019	CA	326	2

5	2020	CA	228	3
6	2018	FL	174	1
7	2019	FL	290	2
8	2020	FL	176	3

```
pivoted = raw_data.pivot(index=["Obs","Year"],columns="State", values="Alpha")
```

pivoted:

Obs	Year	CA	FL	NY
1	2018	349	174	335
2	2019	326	290	400
3	2020	228	176	290

- Note:
Does not allow aggregation like Excel pivot tables (will cover later)
For that, use `.pivot_table()` instead.

4. Melting (feature of Pandas):

- Converts a wide-form dataset into long variable-value form

Example:

raw_data

Index	Name	Month	A	B	C
0	Hans	10	10	98	green
1	Alice	2	11	97	blue
2	Juan	9	13	95	blue
3	Sarah	3	15	93	yellow

```
melted = raw_data.melt( id_vars=["Name"] )
```

melted:

Index	Name	variable	value
0	Hans	Month	10
1	Alice	Month	2
2	Juan	Month	9
3	Sarah	Month	3
4	Hans	A	10
5	Alice	A	11
6	Juan	A	13
7	Sarah	A	15
8	Hans	B	98
9	Alice	B	97
10	Juan	B	95
11	Sarah	B	93
12	Hans	C	green
13	Alice	C	blue
14	Juan	C	blue
15	Sarah	C	yellow